ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1743

Unravelling Seismic-Tsunami Complexity in Northern Chile Through Stochastic Scenario Clustering

Javier Araneda[1], Mailiu Díaz[2] & Juan Gonzalez[1,3]
[1] Escuela de Ciencias de la Tierra, Facultad de Ingeniería, Universidad Andrés Bello, Chile
[2] Facultad de Ingeniería, Universidad Andrés Bello, Chile
[3] Center for Interdisciplinary Research on Disaster Risk, Resilience and Recovery (CIGIDEN R+), Santiago, Chile
j.aranedaherrera1@uandresbello.edu; mailiu.diaz@unab.cl; juan.gonzalez.c@unab.cl

Abstract*– Accurate tsunami hazard assessment requires the consideration of large ensembles of earthquake scenarios to represent the source variability and associated uncertainties adequately. However, simulating hundreds or thousands of tsunami scenarios is computationally demanding, limiting its applicability in operational and planning contexts. This study presents a stochastic clustering and classification framework aimed at reducing and organizing tsunami generating earthquake scenarios while preserving their physical and geological representativeness. A total of 257 synthetic megathrust earthquake scenarios with magnitude Mw 9.0 were generated for the northern Chilean subduction margin (19°S–25°S) and propagated through numerical tsunami simulations to estimate inundation-related impact metrics for the coastal city of Mejillones. Multivariate K-Means clustering was applied to group scenarios with similar tsunamigenic behavior based on key variables such as rupture depth, maximum slip, inundation height, inundated area, and distance to the site. Subsequently, interpretable decision tree models (CART) were used to identify the dominant variables and thresholds controlling cluster differentiation. Results indicate that tsunami impact variability in Mejillones is primarily governed by rupture depth, tectonic domain, and distance to the site, with shallow ruptures located within the megathrust Domain B producing the highest inundation levels. The proposed approach enables a substantial reduction of scenarios without loss of physical consistency, providing an efficient and robust framework for tsunami hazard assessment and risk-informed coastal planning along the northern Chilean margin.

Keywords-- Tsunami hazard; stochastic scenarios; megathrust earthquakes; clustering analysis; Mejillones, Chile

I. INTRODUCTION

Numerical tsunami modelling requires adequately representing the uncertainty associated with the seismic source, especially in highly coupled subduction margins where rupture scenarios can generate significant inundation. Its application in operational contexts continues to face challenges due to the need to simulate large numbers of possible configurations, with high computational costs and execution times that limit its use for risk analysis and rapid response. In recent years, several studies have incorporated advanced numerical models, data assimilation techniques, and machine learning-based approaches to improve the predictive capacity and efficiency of tsunami hazard assessment systems [1,2,3]. However, most of these approaches still depend on the computation of hundreds or thousands of simulated scenarios, which restricts their operational applicability.

An alternative corresponds to the use of stochastic clustering techniques, which allow organizing large sets of seismic scenarios into representative groups while preserving the physical variability of the system [4,5]. These methods reduce the dimensionality of the original dataset and decrease the demand for numerical simulations without compromising predictive quality, explicitly incorporating the uncertainty and statistical structure of the source parameters [6]. Their adoption has shown potential to advance toward faster simulation schemes that are compatible with coastal planning and risk management tasks.

The northern Chile margin constitutes a relevant natural laboratory for this type of approach due to the active convergence between the Nazca and South American plates, where persistent seismic activity has generated a significant number of historical tsunamis [7,8].

The study area (Fig. 1) corresponds to the northern Chilean margin between 19°S and 25°S, where the Nazca plate subducts beneath the South American plate at an approximate rate of 6.5 cm/yr [9]. This segment includes portions of the Arica and Parinacota, Tarapacá, and Antofagasta regions, and coincides with one of the best documented historical seismic gaps along the Andean margin, associated with tsunamigenic events such as the 1868 and 1877 earthquakes and the 2014 Iquique earthquake, the latter linked to a partial release of accumulated interseismic coupling [10, 11, 12].

In this study, the scenarios were characterized according to the seismo-tectonic segmentation proposed by [13], who identified 17 unitary segments of the Chilean megathrust through a multivariate analysis of frictional proxies. Within the 19°S–25°S sector, the Iquique segments (I1, I2, I3) and the Antofagasta segment (A1) dominate, exhibiting differentiated mechanical behaviors within the potential rupture zone. **Fig. 1** shows the spatial distribution of these segments together with the location of Mejillones, allowing the direct relationship with the simulated seismic sources to be visualized.

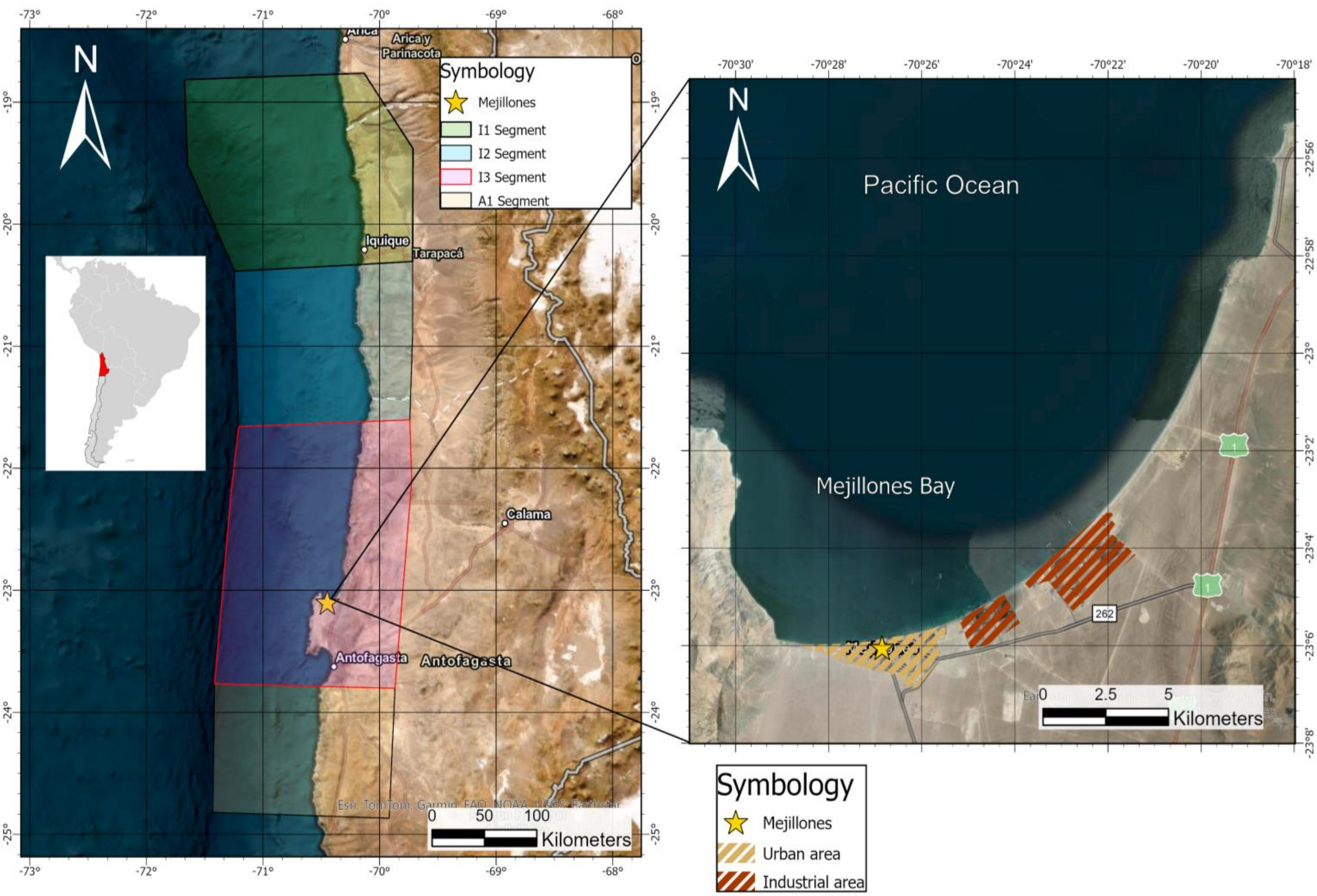

Fig 1. Spatial distribution of the tectonic segments of the subduction margin [13] and cartographic detail of Mejillones Bay, highlighting populated areas, industrial zones, and the reference point used in the analysis.

Furthermore, the tsunamigenic efficiency of this margin is associated with structural variations along the seismic interface, which control both the rupture depth and the deformation style. According to the segmentation proposed by [14], the megathrust can be divided into three domains: a shallow domain (A, 0–15 km) associated with a highly porous and fluid-saturated accretionary prism, where slow and tsunamigenic ruptures may occur; an intermediate domain (B, 15–35 km) where the largest slip tends to concentrate; and a deep domain (C, 35–55 km) characterized by more segmented ruptures and lower contribution to vertical seafloor deformation. These structural differences suggest that not only magnitude, but also rupture depth and the tectonic segment involved, condition the generation and propagation of tsunamis.

In this context, this study uses a stochastic approach based on the simulation of earthquake-tsunami scenarios of magnitude Mw 9.0, which are propagated using tsunami modeling to estimate flood metrics in the city of Mejillones.

Given the large number of scenarios needed to represent the variability of the source, a reduction scheme based on the unsupervised K-Means clustering algorithm is implemented to identify clusters with similar tsunami-generating behaviors. Then, interpretable classification models (CART) are used to identify the variables and thresholds that explain membership in each cluster and validate the internal consistency of the clustering using performance metrics.

This methodology allows us to address the following problem: how to identify a reduced and representative subset of seismic scenarios that captures the tsunamigenic variability of the northern Chile margin without compromising the physical fidelity or the geological interpretation of the phenomenon?

The working hypothesis states that the combination of K-Means and CART effectively reduces the number of scenarios required to describe the tsunamigenic impact in Mejillones, while preserving relevant geological patterns and maintaining predictive capability for subsequent applications.

The main objective of this study is to develop a reduction and classification model for tsunamigenic scenarios using clustering techniques, to identify spatial and geological patterns relevant to the flooding response along the northern Chile margin.

II. Methodology

The methodology employed in this study integrates spatial analysis, statistical processing, unsupervised clustering, and supervised classification techniques with the objective of identifying representative subsets of seismic–tsunamigenic scenarios for the northern Chilean margin and characterizing their potential impact on the city of Mejillones.

A. Scenario database and study area

The study area corresponds to the northern segment of the Chilean margin between 19°S and 25°S. For the analysis, 257 synthetic scenarios of magnitude Mw 9.0 were used, including seismological information (rupture depth and maximum slip), geospatial data (source coordinates), and tsunami impact metrics. The scenarios were previously generated and distributed as ready-to-analyze numerical outputs; therefore, this work does not include the direct tsunami simulation stage.

B. Hydrodynamic tsunami impact data

The scenarios incorporate hydrodynamic results derived from numerical tsunami simulations for the city of Mejillones. The variables used as dependent metrics during the analysis were:

- Maximum mean inundation (m)
- Total inundated area (km²)

These metrics constitute indicators of tsunamigenic efficiency and allow quantifying the relative impact of different rupture patterns without the need to perform additional hydrodynamic simulations.

C. Spatial processing and distance calculation

Seismic sources were georeferenced and projected to UTM Zone 19S using ArcGIS Pro. To quantify the geometric relationship between rupture location and coastal impact, the Euclidean distance between each source and a coastal reference point in Mejillones was computed. This parameter was later incorporated as an independent variable in the multivariate analysis.

D. Preprocessing, variable selection, and standardization

The variables considered for the multivariate analysis include:

- rupture depth (km)
- maximum slip (m)
- distance to Mejillones (km)
- maximum mean inundation (m)
- inundated area (km²)

Statistical processing was performed in Python using the libraries pandas, numpy, and scikit-learn. To ensure comparability between variables expressed in different physical units, z-score standardization was applied by subtracting the mean and dividing by the standard deviation. This operation prevents differences in scale from influencing distance computations during clustering.

E. Unsupervised clustering using K-Means

To reduce the total number of scenarios and obtain representative subsets, the K-Means algorithm was applied using the scikit-learn implementation. The optimal number of clusters was determined through the elbow method by evaluating intra-cluster inertia for different values of k. To minimize sensitivity to initial conditions, the algorithm was executed with 10 independent initializations (n_init = 10), selecting the solution with the lowest total inertia. The result consisted of a structured partitioning of the original dataset into groups with similar seismological and hydrodynamic properties.

F. Classification and interpretability using CART

To identify the variables that explain cluster membership and interpret relevant physical thresholds, a supervised CART

(Classification and Regression Trees) model was implemented. The process was carried out in R. The dataset was split into 80% training and 20% validation. Model performance was evaluated using a confusion matrix and standard classification metrics (accuracy, precision, recall, and F1-score), which allowed validating the internal coherence of the clustering and interpreting the resulting multivariate structure through explicit decision rules.

III. RESULTS

A. Descriptive statistics of the Mw 9.0 scenarios

The set of 257 synthetic Mw 9.0 scenarios exhibits significant variability in both seismological parameters and tsunami impact metrics. Rupture depths range from ~9 to ~54 km, with a mean close to 30 km and a positive skewness resulting from the presence of deeper ruptures. Maximum slip values span between 10 and 20 m, with a median around 16 m and a relatively narrow interquartile range, indicating a moderately concentrated distribution around intermediate value. Tsunami impact metrics show mean inundation heights of ~18 m and average inundated areas on the order of ~0.45 km², both displaying upper-tail behavior and controlled dispersion.

Histograms of maximum slip display a unimodal distribution concentrated mainly between 13 and 19 m, with low frequency at the extremes. The shallow domain concentrates the highest slip values within the dataset, whereas intermediate and deeper regions exhibit more balanced distributions with reduced contributions in the upper range.

Overall, the results show that even within a fixed magnitude class (Mw 9.0), the tsunamigenic response displays considerable variability, reflected in both seismological parameters and inundation and exposure metrics.

B. Application of the K-Means Algorithm

The elbow method allowed identifying an inflection point at k = 4, which was used as the criterion for selecting the final number of clusters for the Mw 9.0 scenarios. This value enabled the dataset to be segmented into statistically differentiated and coherent groups in terms of rupture depth, maximum slip, and flooding-related variables.

1) Two-Dimensional Representation (PCA)

The partition obtained for Mw 9.0 (**Fig. 2**) shows four well-defined groups when projected using Principal Component Analysis (PCA). Components PC1 (37.9%) and PC2 (26.3%) jointly explain more than 60% of the total variability, allowing an adequate representation of the multivariate structure of the dataset. Cluster separation is clear, with limited overlap at the boundaries, and the centroids appear as midpoints that synthesize the average characteristics of each group.

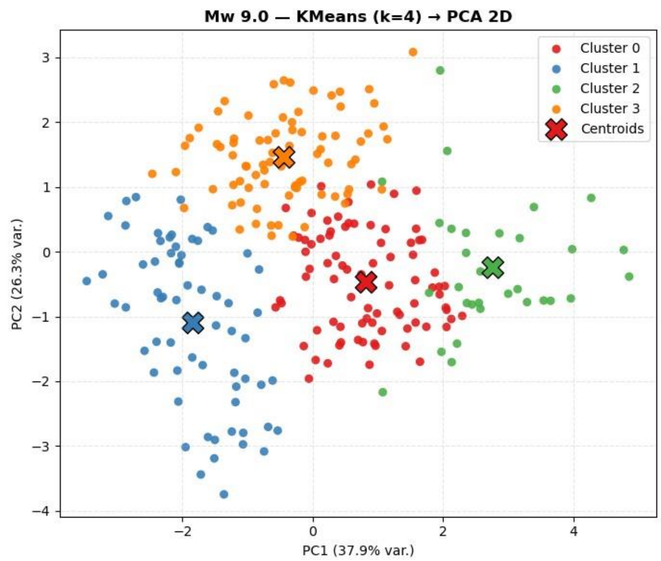

Fig 2. Distribution of Mw 9.0 scenarios in the 2D PCA space and grouping through K-means (k = 4). The four identified clusters and the locations of their centroids are shown.

2) Cluster Statistics

The four clusters identified for Mw 9.0 exhibit systematic differences in both seismological parameters and impact-related variables.

Cluster 0 (n = 82): intermediate rupture depths (~18.4 km) and moderate maximum slip (~15.5 m). It shows moderate flooding heights (~17.8 m) and inundated areas (~0.36 km²), with average distances of ~131 km.

Cluster 1 (n = 64): deeper ruptures (~32.3 km) and similar maximum slip (~15.8 m), but associated with the lowest flooding heights (~16.5 m) and the smallest inundated area (~0.09 km²). It represents the most distant group (~386 km).

Cluster 2 (n = 31): the shallowest group (~17.6 km), characterized by the highest values of maximum slip (~17.9 m), flooding height (~20.5 m), and inundated area (~1.59 km²). It also exhibits the shortest average distances (~122 km).

Cluster 3 (n = 80): deeper scenarios (~43.4 km) with maximum slip of ~15.3 m, intermediate flooding heights (~17.9 m), and inundated areas (~0.40 km²), located at ~162 km.

3) Boxplots by Cluster

The boxplots reinforce the structure observed in the statistical analysis, revealing a consistent ordering among the clusters. Cluster 2 concentrates the highest values in both maximum slip (~18 m) and flooding (~20 m), as well as in affected area (> 1 km²), suggesting a higher tsunamigenic efficiency. In contrast, Cluster 1 exhibits the lowest values for these three variables, together with the largest distance to the site (~400 km), while Clusters 0 and 3 lie in intermediate ranges with similar dispersion patterns. The distance variable shows the clearest separation between groups, distinguishing a distant cluster (1), a proximal cluster (2), and two intermediate ones (0 and 3). Overall, the boxplots emphasize systematic differences in both magnitude and dispersion, providing visual support for the segmentation obtained through K-Means (**Figs. 3 to 6**).

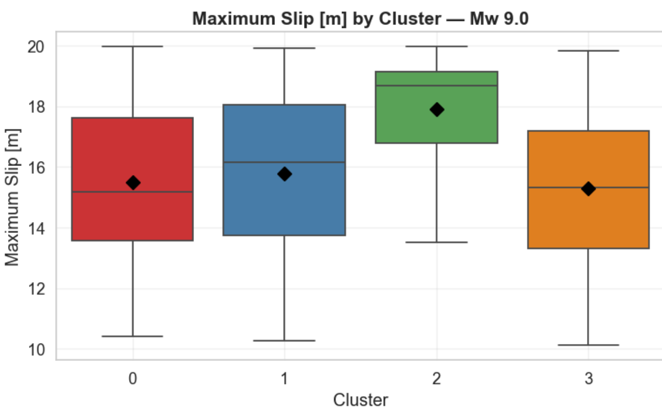

Fig 3. Maximum slip by cluster (Mw 9).

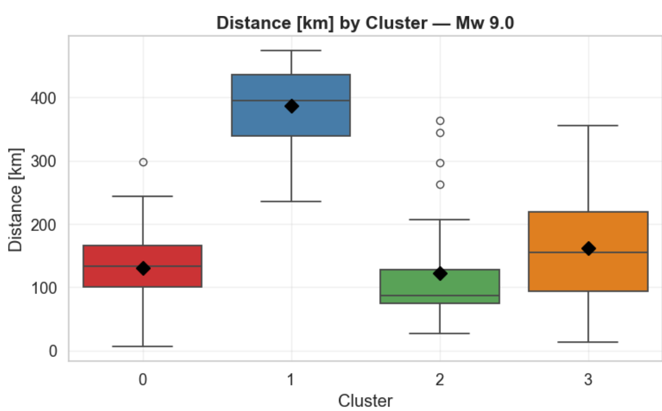

Fig 4. Boxplot of distance variable by cluster (Mw 9).

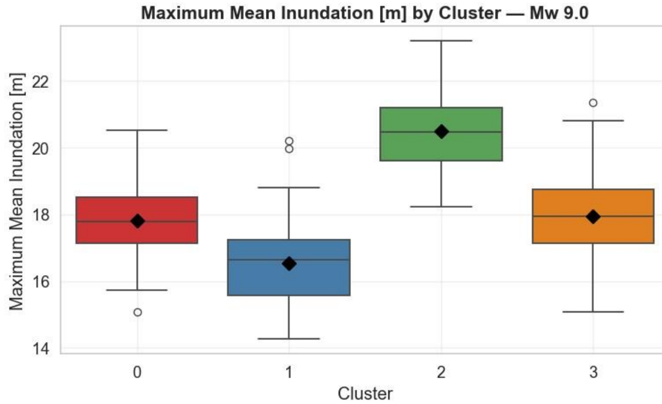

Fig 5. Boxplot of max mean inundation variable by cluster (Mw 9).

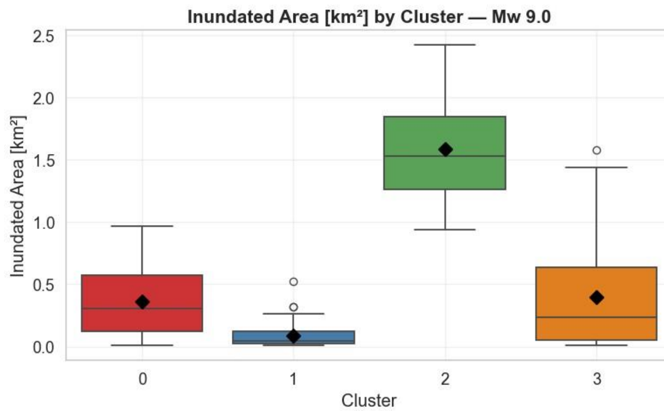

Fig 6. Boxplot of inundated area variable by cluster (Mw 9).

4) Slip–Inundation and Distance–Inundation Relationships

The scatterplots reveal low correlations between maximum slip and inundation variables across the four clusters. For Slip–Inundation, Cluster 0 exhibits weak positive relationships ($r \approx 0.25$), Cluster 1 shows very weak negative trends, and Clusters 2 and 3 display slightly positive tendencies. For Distance–Inundation, mild slopes also dominate, with weak negative correlations in Cluster 0 and near-zero correlations in Clusters 1 and 2. Cluster 3 exhibits the strongest correlation in the set ($r \approx 0.31$ for inundation height), although still low in absolute terms.

Overall, the results indicate that K-Means clustering allows the identification of subsets with systematic differences in rupture depth, slip, distance, and impact variables, whereas simple linear relationships between slip/distance and inundation remain weak within each group.

C. Cluster Characterization and Interpretation

The K-Means partition obtained for Mw 9.0 reveals clear differences among clusters in terms of rupture depth, distance, and tsunamigenic efficiency. In addition to this statistical contrast, the spatial distribution (**Fig. 7**) shows how clusters organize along the subduction margin: shallow clusters from Domain B concentrate offshore of the central segment near Mejillones, whereas deeper clusters from Domain C and the most distal scenarios are distributed toward the lateral ends of the analyzed segment. The centroid locations plotted in Fig. 7 summarize these patterns, reinforcing that the most relevant deformational contributions are situated offshore from the Mejillones Bay.

Based on this spatial arrangement, an operative hierarchy of local relevance can be defined: a high-efficiency cluster associated with shallow and proximal ruptures, a medium-efficiency cluster linked to deep and intermediate ruptures, a moderately efficient but proximal cluster, and a low-relevance cluster controlled by its greater distance.

The Mw 9.0 clusters can be synthesized according to tectonic domain, depth, distance, and relevance to Mejillones as follows:

Cluster 0: Domain B, ~18 km depth, near-to-intermediate distance, with moderate slip and inundation. Relevance: medium–high.

Cluster 1: Domain B, ~32 km depth, very distant, with the smallest inundated area and lowest impact. Relevance: low.

Cluster 2: Domain B, ~18 km depth, proximal, with the largest inundation heights and areas. Relevance: very high.

Cluster 3: Domain C, ~43 km depth, intermediate distance, with medium–high inundation. Relevance: medium.

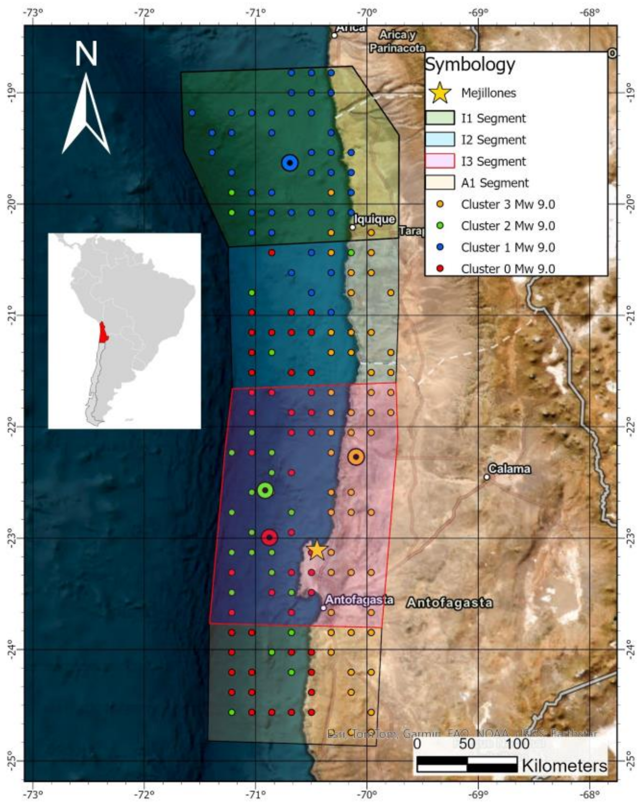

Fig 7. Spatial distribution of Mw 9.0 scenarios by cluster and tectonic segmentation of the margin. The Mw 9.0 clusters are shown superimposed on the seismotectonic segments (I1, I2, I3 and A1), with Mejillones highlighted as a reference point.

D. Decision Trees (CART)

The CART model applied to the Mw 9.0 scenarios sought to evaluate the internal coherence of the K-Means clustering and to identify the variables that best explain the assignment of each scenario to its respective cluster.

1) Tree Structure (Mw 9.0)

The decision tree for Mw 9.0 shows that the distance to Mejillones (Distance_km) constitutes the first partitioning criterion, separating early on the most distant scenarios (≥ 320 km), which are primarily assigned to Cluster 1, characterized by reduced local impact and greater offshore distance. Within the subset of closer scenarios (< 320 km), the second relevant criterion corresponds to rupture depth (Depth_km), distinguishing shallow ruptures (< 31 km), which are mostly grouped into Cluster 0, from deeper ruptures (≥ 31 km) associated with Cluster 3.

Within the shallow branch of the tree, the variable Inundated Area (km²) acts as an additional physical threshold: values exceeding ≈ 0.99 km² trigger classification toward Cluster 2, corresponding to scenarios with the largest inundation extent. Finally, distance reappears at lower tree levels, distinguishing proximal cases (< 240 km) from intermediate cases, which allows refining the assignment within Cluster 0 and avoiding confusion with Cluster 1.

2) Variable Importance

The variable importance analysis shows that Distance_km is the predictor with the greatest influence on class separation, followed by Depth_km. Both variables control the geometric configuration of the ruptures with respect to the impact zone, in agreement with the structure observed in the tree. The variables Inundated Area and Maximum Mean Inundation display intermediate importance, providing hydrodynamic criteria to differentiate scenarios with similar geometric characteristics. In contrast, Maximum Slip exhibits the lowest relative importance, indicating that for Mw 9.0 the maximum slip does not constitute a primary predictor of cluster membership within the CART classification.

3) Validation through Confusion Matrix

The confusion matrix allows evaluating model performance by comparing true and predicted classes. The model correctly classified 47 out of the 50 evaluated scenarios, corresponding to a global accuracy of ≈ 0.94. Misclassifications were concentrated in Cluster 3, which absorbed some scenarios originally belonging to other groups, whereas Clusters 0, 1, and 2 presented lower proportions of false positives.

4) Evaluation Metrics

The quantitative metrics associated with each cluster are summarized in **Table 1**. The results show high precision, recall, and F1-scores for all groups (≥ 0.83), reflecting a good balance between false positives and false negatives. Cluster 3 exhibits the lowest relative precision, consistent with the trend observed in the confusion matrix. The model's performance confirms the internal consistency of the partitions obtained through K-Means and validates that source geometry (distance + depth) constitutes the primary controller of scenario differentiation.

Table 1. Performance metrics of the CART model for cluster classification (accuracy, precision, recall, and F1-score) for Mw 9.0.

Cluster	Accuracy	Precision (PPV)	Recall (Sensitivity)	F1-score
0	0.94	1	0.9375	0.9677
1	0.94	1	0.9167	0.9565
2	0.94	1	0.8333	0.9091
3	0.94	0.8421	1	0.9143

Taken together, the results obtained for the Mw 9.0 scenarios reveal an internally consistent structure across the different stages of the analysis. The descriptive statistics show substantial variability in depth, slip, and inundation, justifying the need to reduce the dataset through clustering techniques. The K-Means algorithm allowed identifying four representative groups with differentiated geometric and hydrodynamic patterns which was reinforced through the spatial characterization and physical interpretation of each cluster. Subsequently, the CART model proved capable of reproducing the unsupervised classification with high performance, identifying clear hierarchical relationships among the variables and validating the coherence of the clustering. Distance and depth emerged as the main drivers of inter-cluster separation, whereas inundation variables refined the distinction between scenarios with different potential impact.

IV. Discussion

The results allow consistently linking the patterns identified through K-Means and CART with the geological structure of the northern Chile subduction margin. Tsunamigenic efficiency is primarily explained by the combination of rupture depth, tectonic domain, and distance to Mejillones, with none of these factors alone fully capturing the variability in inundation. This highlights the multifactorial nature of the phenomenon.

From a geological perspective, the observed patterns coincide with the megathrust zonation proposed by [14], in which Domain B (15–35 km) concentrates the most efficient ruptures in terms of slip and vertical deformation. The clusters with the highest impact are located precisely within this domain or in its transition with Domain A, in agreement with studies documenting that maximum tsunami heights originate from shallow and highly coupled ruptures. In contrast, clusters associated with Domain C (>35 km) systematically display lower effective slip and reduced vertical deformation, resulting in diminished inundation levels.

The CART analysis confirmed these tendencies by identifying depth and distance to Mejillones as the most important predictors, yielding simple rules of the form shallow + proximal → higher impact. This result is operationally relevant, as it allows translating the complexity of the multivariate space into explicit criteria for planning and risk assessment.

The inclusion of distance weighted by maximum slip further refined this interpretation, showing that impact does not depend solely on the magnitude of slip, but on where it occurs. Scenarios with high slip nucleated offshore of distant, generate negligible inundation contributions in Mejillones, a phenomenon also observed in real events such as the 2014 Iquique earthquake.

The correlation analyses confirm that neither distance, nor depth, nor slip individually explains inundation, showing low to moderate values of r and ρ. This suggests that tsunamigenic efficiency emerges from complex spatial interactions involving margin morphology, shelf geometry, rupture directivity, slip distribution, and coastal geomorphology. This behavior explains the internal variability observed even within a single tectonic domain and raises questions for future studies regarding the mechanisms that control such heterogeneity.

Methodologically, the results show that clustering techniques allow reducing the number of scenarios without losing physical representativeness, optimizing computational cost in agreement with the findings of [4,5]. The coherence between K-Means and CART supports that the identified patterns do not correspond to statistical artifacts, but rather to physical differences associated with the dynamics of the subduction margin.

Tsunamigenic efficiency in Mejillones depends on the nonlinear interaction among margin geology, rupture depth, maximum slip location, and source-to-site distance, and this interaction can be organized through clustering techniques for operational, risk assessment, and applied communication purposes.

V. CONCLUSION

The results obtained for the Mw 9.0 simulated scenarios demonstrate that the integration of clustering techniques (K-Means) with interpretable classification models (CART) constitutes an effective strategy to reduce the total set of earthquakes–tsunami scenarios and to organize their variability in a physically coherent manner. The statistical, spatial, and multivariate analyses revealed that tsunamigenic efficiency in Mejillones depends primarily on the combination of rupture depth, the relative location of maximum slip, and source-to-coast distance, whereas slip acts as a secondary modulator rather than as a linear predictor of hydrodynamic impact.

The clustering process allowed identifying groups with differentiated behavior, in which the scenarios with the highest potential impact correspond to shallow ruptures associated with depths characteristic of the tectonic domain where slip is most efficient in generating vertical deformation. In contrast, deep or distal scenarios exhibited low tsunamigenic efficiency, even in the presence of elevated slip values. These differences were reinforced by the CART model, which translated the clustering patterns into explicit and quantifiable rules based on depth, distance, and inundated area thresholds, while simultaneously validating the internal consistency of the classification scheme.

The results show that it is possible to significantly reduce the number of scenarios required to represent the spectrum of impacts associated with a large interplate earthquake without sacrificing physical or interpretative fidelity. This approach provides an operational tool for identifying representative and critical scenarios in future hydrodynamic modeling efforts and offers a technical basis for improving computational efficiency in coastal hazard and risk assessment applications along the northern Chile margin.

ACKNOWLEDGMENTS

The authors acknowledge the support of the Faculty of Engineering at Universidad Andrés Bello, Viña del Mar campus. JG acknowledges funding from the grant ANID CIGIDEN R+ Grant CIN250023. This research was partially supported by the supercomputing infrastructure of the NLHPC(ECM-02).

REFERENCES

[1] T. Baba, N. Takahashi, Y. Kaneda, K. Ando, D. Matsuoka, and T. Kato, “Parallel Implementation of Dispersive Tsunami Wave Modeling with a Nesting Algorithm for the 2011 Tohoku Tsunami,” Pure and Applied Geophysics, vol. 172, no. 12, pp. 3455–3472, Feb. 2015, doi: https://doi.org/10.1007/s00024-015-1049-2.

[2] Y. Igarashi, T. Hori, S. Murata, K. Sato, T. Baba, and M. Okada, “Maximum tsunami height prediction using pressure gauge data by a Gaussian process at Owase in the Kii Peninsula, Japan,” Marine Geophysical Research, vol. 37, no. 4, pp. 361–370, Dec. 2016, doi: https://doi.org/10.1007/s11001-016-9286-z.

[3] S. Koshimura, Y. Ohta, Y. Sato, M. Kachi, and H. Kobayashi, “Advances of Real-Time Tsunami Inundation And Damage Forecast -Present And Future.” Accessed: May 12, 2026. [Online]. Available: https://wcee.nicee.org/wcee/article/17WCEE/5e-0002.pdf

[4] M. Kamiya, Y. Igarashi, M. Okada, and T. Baba, “Numerical experiments on tsunami flow depth prediction for clustered areas using regression and machine learning models,” Earth, Planets and Space, vol. 74, no. 1, Aug. 2022, doi: https://doi.org/10.1186/s40623-022-01680-9.

[5] A. Williamson, D. Rim, L. M. Adams, R. J. LeVeque, D. Melgar, and F. González, “A Source Clustering Approach for Efficient Inundation Modeling and Regional Scale Probabilistic Tsunami Hazard Assessment,” Frontiers in Earth Science, vol. 8, Aug. 2020, doi: https://doi.org/10.3389/feart.2020.591663.

[6] G. Bolivou, K.G. Reddy and M.G.M. Khan. “A Novel Method of Clustering Using a Stochastic Approach,” IEEE.org, 2021, doi: https://doi.org/10.21227/35et-ph33.

[7] B. K. Horton, “Tectonic Regimes of the Central and Southern Andes: Responses to Variations in Plate Coupling During Subduction,” Tectonics, vol. 37, no. 2, pp. 402–429, Feb. 2018, doi: https://doi.org/10.1002/2017tc004624.

[8] CIGIDEN, “Guía para la Estimación de Peligro de Tsunami.” 2018. Available: https://www.cigiden.cl/wp-content/uploads/2018/12/GUIA-PELIGRO-TSUNAMI-HZ.pdf

[9] D. Angermann, J. Klotz, and C. Reigber, “Space-geodetic estimation of the nazca-south america euler vector,” Earth and Planetary Science Letters, vol. 171, no. 3, pp. 329–334, Sep. 1999, doi: https://doi.org/10.1016/s0012-821x(99)00173-9.

[10] D. Comte and M. Pardo, “Reappraisal of great historical earthquakes in the northern Chile and southern Peru seismic gaps,” Natural Hazards, vol. 4, no. 1, pp. 23–44, 1991, doi: https://doi.org/10.1007/bf00126557.

[11] J. C. Villegas-Lanza et al., “Active tectonics of Peru: Heterogeneous interseismic coupling along the Nazca megathrust, rigid motion of the Peruvian Sliver, and Subandean shortening accommodation,” Journal of Geophysical Research: Solid Earth, vol. 121, no. 10, pp. 7371–7394, Oct. 2016, doi: https://doi.org/10.1002/2016jb013080.

[12] M. Métois, C. Vigny, and A. Socquet, “Interseismic Coupling, Megathrust Earthquakes and Seismic Swarms Along the Chilean Subduction Zone (38°–18°S),” Pure and Applied Geophysics, vol. 173, no. 5, pp. 1431–1449, Apr. 2016, doi: https://doi.org/10.1007/s00024-016-1280-5.

[13] D. Molina, A. Tassara, R. Abarca, D. Melnick, and A. Madella, “Frictional Segmentation of the Chilean Megathrust From a Multivariate Analysis of Geophysical, Geological, and Geodetic Data,” Journal of Geophysical Research: Solid Earth, vol. 126, no. 6, Jun. 2021, doi: https://doi.org/10.1029/2020jb020647.

[14] T. Lay et al., “Depth-varying rupture properties of subduction zone megathrust faults,” Journal of Geophysical Research: Solid Earth, vol. 117, no. B4, p. n/a-n/a, Apr. 2012, doi: https://doi.org/10.1029/2011jb009133.