Evaluating a Machine Learning Workflow for Aftershock Catalog Construction: The 2017 Tehuantepec Mw8.2 Sequence

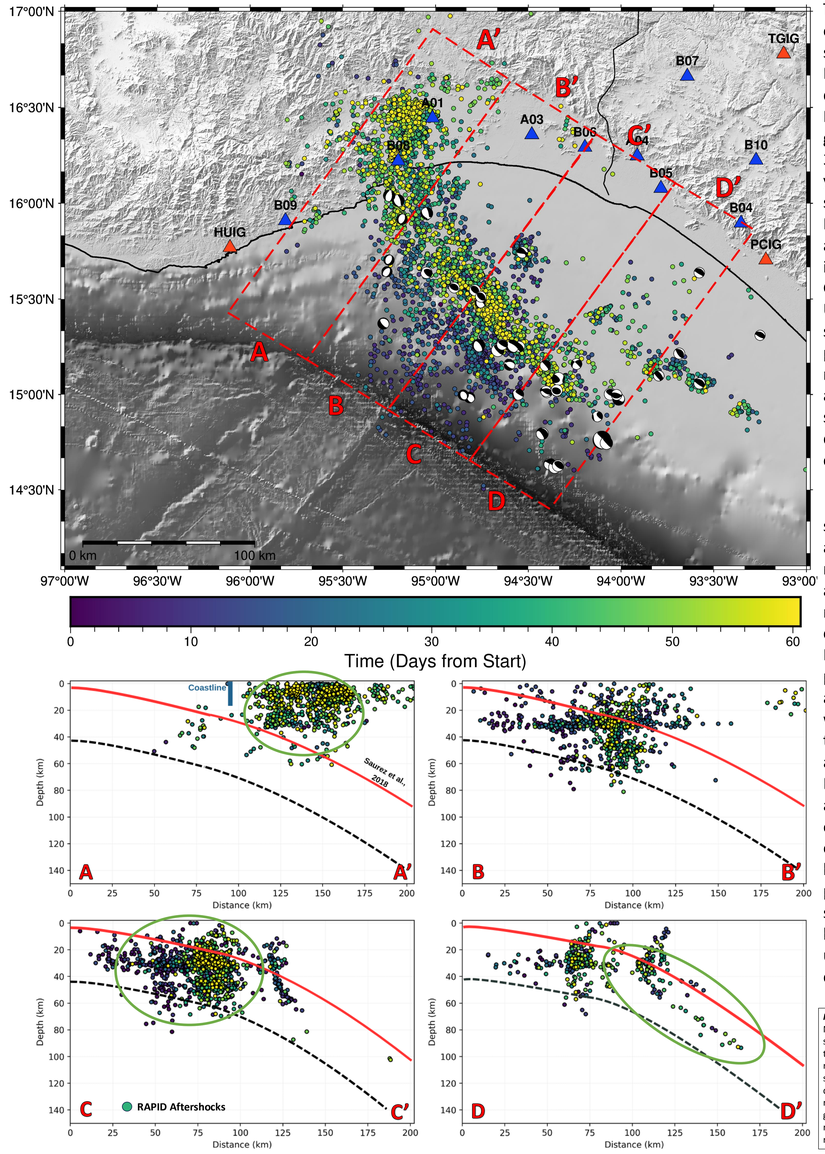
Marc A. Garcia[1], Aaron A. Velasco[1], Chengping Chai[2]

[1]University of Texas at El Paso, Department of Earth, Environmental and Resources Sciences, [2]Oak Ridge National Laboratory

Introduction

The Mw 8.2 Tehuantepec earthquake of September 8, 2017, was one of the strongest intraplate earthquakes in Mexican history, generating over 30,000 aftershocks in the first year. Despite its significance, the dynamics of rupture and the aftershock structure remain poorly understood. In this project, we evaluate a machine learning–assisted workflow for building a high-resolution aftershock catalog by integrating modern arrival-time pickers and association tools with traditional location and relocation techniques. Our approach combines PhaseNet, a convolutional neural network for seismic phase picking; GaMMA, a Bayesian Gaussian Mixture Model–based association algorithm; and the well-established HypoInverse and HypoDD programs for absolute and relative event locations. We incorporate data from both permanent regional stations and a six-month temporary RAPID deployment, enhancing coverage of the offshore rupture zone. The resulting catalog shows significantly improved magnitude completeness, spatial resolution, and consistency, particularly for low-magnitude events. These results demonstrate the value of open-source machine learning tools for earthquake monitoring in complex and data-limited environments, providing a new benchmark catalog for the Tehuantepec sequence.

Background

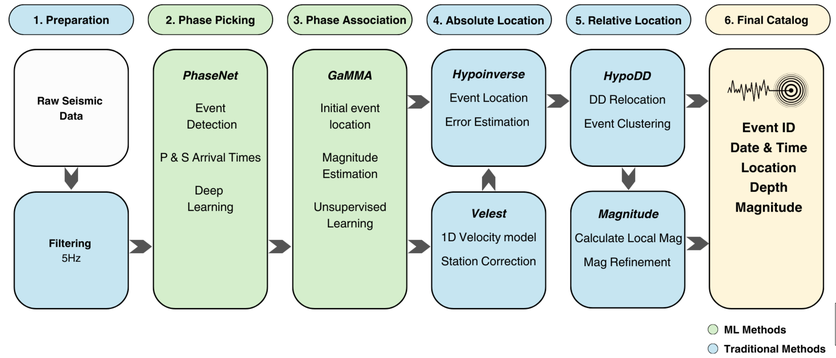


The Mw 8.2 Tehuantepec earthquake, which occurred on September 8, 2017, is one of the strongest recorded intraplate earthquakes in Mexican history. The event ruptured a steeply dipping normal fault within the subducting Cocos Plate at intermediate depth (~45–70 km), generating intense ground motions and more than 30,000 reported aftershocks within the following year. In response, a six-month rapid-response seismic deployment (RAPID), funded by the National Science Foundation, was launched to augment the limited offshore station coverage and improve aftershock monitoring in the Tehuantepec Gap. This collaborative effort between UTEP, UNAM, UACJ, and SSN added ten broadband stations to the region, complementing the permanent seismic network. While the mainshock rupture has been studied, key questions remain about the aftershock distribution, underlying fault structures, and stress evolution within the slab—questions that require a high-resolution earthquake catalog for meaningful analysis.

Building such a catalog involves several processing stages: event detection, phase picking, phase association, and location or relocation. Traditional methods, like STA/LTA-based pickers and grid search association, are often challenged by low signal-to-noise ratios and complex tectonic settings, conditions prevalent in the Tehuantepec region. Machine learning tools have recently emerged as powerful alternatives, capable of improving the accuracy and scalability of each step. In this study, we evaluate a machine learning–assisted workflow that combines PhaseNet for deep learning–based arrival-time picking, GaMMA for unsupervised Bayesian phase association, and the HypoInverse and HypoDD algorithms for absolute and relative event location. This hybrid approach allows us to efficiently process large waveform datasets from both permanent and temporary stations, generate precise hypocenter solutions, and resolve the fine-scale structure of aftershock clusters. The result is a benchmark-quality catalog that enhances our understanding of seismicity in one of Mexico's most complex tectonic regions.

Figure 1: Map and cross-section views of aftershocks following the September 8, 2017, Mw 8.2 Tehuantepec earthquake. (Top) Epicentral distribution of aftershocks overlaid on shaded topography, color-coded by time since the mainshock. Red dashed lines indicate the locations of four cross-section profiles (A–A' through D–D'). Focal mechanisms represent large events from the Global CMT catalog. Temporary and permanent seismic stations are marked by triangles. (Bottom) Cross-sectional views of aftershock depth distributions along each profile, plotted against distance from the coastline. Events recorded by the temporary RAPID deployment are highlighted. Red lines indicate the slab geometry from Suárez et al. (2019), while dashed black lines represent an alternative reference interface. The clustering and variable depth patterns illustrate the complex rupture geometry and slab interaction beneath the Tehuantepec region

Workflow



To build a high-resolution aftershock catalog, we implemented a hybrid workflow that combines machine learning algorithms with traditional seismic processing techniques. PhaseNet, a convolutional neural network, was used to detect and pick P- and S-wave arrivals from continuous waveform data. These picks were passed to GaMMA, an unsupervised phase association algorithm that clusters arrivals into seismic events and estimates initial magnitudes. Event locations were refined using HypoInverse with a 1D velocity model derived from VELEST, followed by relative relocation with HypoDD to reduce spatial uncertainty. Local magnitudes were calculated using amplitude measurements and refined for consistency. This integrated approach allows for efficient processing of large datasets while preserving the accuracy and interpretability of traditional methods.

Figure 2: Workflow for constructing a high-resolution aftershock catalog of the 2017 Tehuantepec earthquake sequence. The process integrates modern machine learning techniques (green) with established seismological methods (blue).

Effectiveness of Machine Learning

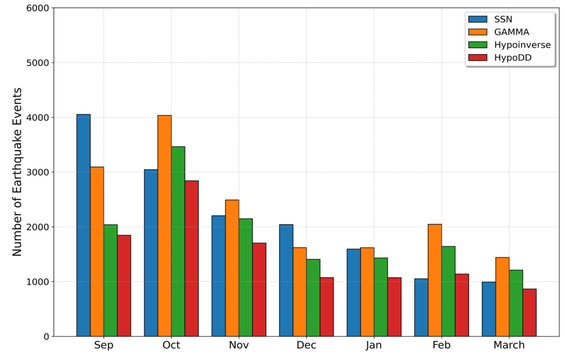


Figure 3: Monthly earthquake event counts from September 2017 to March 2018 comparing four catalogs: SSN, machine learning-based association via GaMMA, absolute locations using HypoInverse, and relocated events using HypoDD. The increase in October reflects the start of the temporary deployment. The gradual reduction in event counts with each processing stage highlights the effects of quality control and relocation filtering.

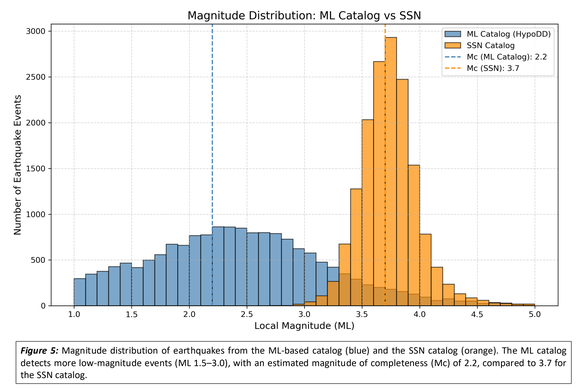


Figure 5: Magnitude distribution of earthquakes from the ML-based catalog (blue) and the SSN catalog (orange). The ML catalog detects more low-magnitude events (ML 1.5–3.0), with an estimated magnitude of completeness (Mc) of 2.2, compared to 3.7 for the SSN catalog.

Detection Algorithm	Precision	Accuracy	Recall	F1 Score
PhaseNet P	91%	88%	55%	65%
PhaseNet S	90%	61%	51%	56%
STA/LTA P	94%	92%	17%	29%
STA/LTA S	94%	74%	21%	30%

Figure 4: Shows the statistical analysis results (precision, accuracy, recall, and F1 score) for the performance of different seismic phase detection algorithms compared to picks generated by the SSN.

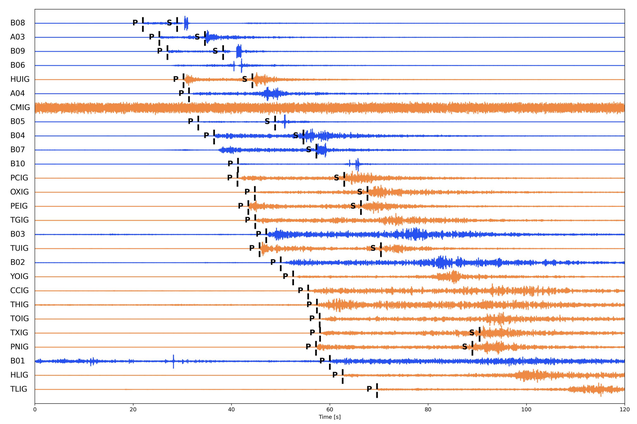


Figure 6: Example waveforms from temporary and SSN stations for an earthquake on October 14, 2017, at 00:21:15.2. The high-quality data from the temporary network (blue) supplements the permanent network (orange), enhancing earthquake detection and characterization at local distances

Conclusions

This study demonstrates that integrating machine learning tools with traditional seismic workflows significantly improves the resolution and completeness of earthquake catalogs in complex tectonic settings. By combining PhaseNet, GaMMA, HypoInverse, and HypoDD, we constructed the most detailed aftershock catalog to date for the 2017 Mw 8.2 Tehuantepec sequence. This enhanced catalog offers new insights into subduction zone processes and highlights the potential of ML-assisted workflows for rapid and scalable seismic analysis.

Acknowledgements

"This material is based upon work supported by the National Science Foundation under Grant No. INTERN-2054442"
"Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

This research used resources of the Compute and Data Environment for Science (CADES) at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725