Colombian Seismic Monitoring Using Advanced Machine-Learning Algorithms

Emmanuel Castillo*[1], Daniel Siervo[2], and Germán A. Prieto[1]

Abstract

Seismic networks worldwide are designed to monitor seismic ground motion. This process includes identifying seismic events in the signals, picking and associating seismic phases, determining the event's location, and calculating its magnitude. Although machine-learning (ML) methods have shown significant improvements in some of these steps individually, there are other stages in which traditional non-ML algorithms outperform ML approaches. We introduce SeisMonitor, a Python open-source package to monitor seismic activity that uses ready-made ML methods for event detection, phase picking and association, and other well-known methods for the rest of the steps. We apply these steps in a totally automated process for almost 7 yr (2016–2022) in three seismic networks located in Colombian territory, the Colombian seismic network and two local and temporary networks in northern South America: the Middle Magdalena Valley and the Caribbean-Mérida Andes seismic arrays. The results demonstrate the reliability of this method in creating automated seismic catalogs, showcasing earthquake detection capabilities and location accuracy similar to standard catalogs. Furthermore, it effectively identifies significant tectonic structures and emphasizes local crustal faults. In addition, it has the potential to enhance earthquake processing efficiency and serve as a valuable supplement to manual catalogs, given its ability at detecting minor earthquakes and aftershocks.

Cite this article as Castillo, E., D. Siervo, and G. A. Prieto (2024). Colombian Seismic Monitoring Using Advanced Machine-Learning Algorithms, *Seismol. Res. Lett.* **XX**, 1–15, doi: 10.1785/0220240036.

Supplemental Material

Introduction

The expansion of seismological networks, persistent monitoring efforts, and the establishment of local or temporary networks for specific investigations have contributed to a significant augmentation in the quantity of seismological data (Ringler *et al.*, 2019). This has also led to an increase in the number of identified earthquakes, unveiling seismic activity that had previously gone unnoticed.

Because of the volume of data, there is a need to develop computational tools capable of processing it. These tools have been studied, even before the recent increase in the use of machine-learning (ML) algorithms, and address the different steps of the earthquake monitoring workflow, which mainly includes event detection (Allen, 1978; Gibbons and Ringdal, 2006), phase picking (Allen, 1978; Sleeman and van Eck, 1999; Saragiotis *et al.*, 2002; Ross and Ben-Zion, 2014), phase association (McBrearty, Gomberg, *et al.*, 2019; Yeck *et al.*, 2019; Sheen and Friberg, 2021), earthquake location (Lahr, 1999; Lomax *et al.*, 2000; Klein, 2002; Pavlis *et al.*, 2004), and magnitude estimation (Rengifo and Carriazo, 2004; Lopez *et al.*, 2020).

Event detection and phase picking can be particularly time-consuming tasks. This is due to the need to process continuous data from multiple stations to identify earthquake signals and record the arrival times for both *P* and *S* waves. This process must contend with nonstationary background noise, which can originate from various sources, including instrumentation, ambient factors, and human-generated disturbances. Furthermore, it is crucial to consider various factors that can result in different types of earthquake signals. These factors may include the source's distance from the recording station, the signal-to-noise ratio, and the occurrence of multiple earthquake signals in close temporal proximity, often stemming from a sequence of aftershocks. Each of these examples can manifest at any station in the seismological network, presenting a formidable challenge in seismology. Existing non-ML picking algorithms have often fallen short in achieving results comparable to manual methods. Moreover, configuring these algorithms can be laborious and time-intensive, requiring the definition of parameters such as filters and thresholds for each station. As a result, it is crucial to explore novel approaches that enable the training of algorithms to generalize detections across various signal types.

1. Departamento de Geociencias, Universidad Nacional de Colombia, Bogotá, Colombia, U.S.A., https://orcid.org/0000-0002-9799-9775 (EC); https://orcid.org/0000-0001-8538-7379 (GAP); 2. Bureau of Economic Geology, Jackson School of Geosciences, The University of Texas, Austin, Texas, U.S.A., https://orcid.org/0000-0003-4797-2942 (DS)

*Corresponding author: ecastillot@unal.edu.co

© Seismological Society of America

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

Recently, because seismology is a data-rich science, several recent results show that ML algorithms have the potential to carry on some of the steps in the earthquake monitoring workflow (Woollam *et al.*, 2022). Deep learning, a subset of ML known for its ability to learn complex relationships within vast data sets through layered neurons and nonlinear transformations (deep neural networks), has significantly influenced seismological tasks (Mousavi and Beroza, 2022). These algorithms can be implemented using supervised learning or unsupervised learning. In supervised learning, algorithms learn from labeled data to make predictions on new, unlabeled data sets. On the other hand, unsupervised learning algorithms classify data by identifying hidden patterns without the need for labeled data. As a result of the large, quality-controlled, and labeled data sets in seismology (Mousavi, Sheng, *et al.*, 2019; Yeck *et al.*, 2019; Magrini *et al.*, 2020), supervised learning has succeeded in generating seismic catalogs with unprecedented detail. Unlike supervised learning, which depends on labeled data sets created by humans, unsupervised learning is not biased by predetermined labels. Consequently, it has the potential to reveal patterns that are not normally easily identified by supervised methods. This characteristic establishes unsupervised learning as a promising approach to complete the higher expression of seismicity (Beroza *et al.*, 2021). Both supervised and unsupervised learning have been used in the state-of-the-art ML models in the earthquake monitoring workflow.

Because of the extensive archives of seismic data and the availability of handpicked labels from earthquake catalogs, event detection and phase picking can be considered supervised learning tasks (Ross *et al.*, 2018; Dokht *et al.*, 2019; Pardo *et al.*, 2019; Woollam *et al.*, 2019; Mousavi and Beroza, 2022; Saad *et al.*, 2023). These tasks can be approached either independently or simultaneously (Mousavi, Zhu, *et al.*, 2019; Zhu and Beroza, 2019; Mousavi *et al.*, 2020). In addition, much more robust and accurate data sets have emerged over time (Mousavi, Sheng, *et al.*, 2019), and more than one million labeled signals recorded around the world have been used to train state-of-the-art algorithms (Mousavi *et al.*, 2020).

After identifying seismic phases across various stations, the next step involves their association. This process entails linking phase arrival times observed at multiple stations to each shared source that generated these phases. Although the phase-picking step has been significantly improved using ML methods, phase association has not received the same attention. This task primarily relies on grid-search and back-projection algorithms (Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences and gempa GmbH, 2018; McBrearty, Gomberg, *et al.*, 2019; Yeck *et al.*, 2019; Sheen and Friberg, 2021). Despite these algorithms being robust and effective, they have demonstrated limitations when associating phases that are close in time and come from different sources or by ignoring false phases poorly picked by the automatic picking algorithms. Therefore, other approaches are emerging using random sample consensus (Woollam *et al.*, 2019), Bayesian Gaussian mixture models (BGMMs; Zhu *et al.*, 2022), and supervised deep-learning algorithms (Dickey *et al.*, 2019; McBrearty, Delorey, *et al.*, 2019; Ross *et al.*, 2019; McBrearty and Beroza, 2023).

In this study, we present an automatic seismic monitoring workflow using some of the state-of-the-art ML approaches for the three first main steps of the earthquake monitoring workflow. For event detection and phase picking, we use the two original pretrained deep-learning models: PhaseNet (Zhu and Beroza, 2019) and EQTransformer (Mousavi *et al.*, 2020). Both models use the three-component seismic waveforms as input, but they differ in neural network architecture, output results, and the number of training samples.

PhaseNet employs several stacked convolutional layers in an adapted version of the U-Net architecture (Ronneberger *et al.*, 2015) to generate Gaussian probability distributions of *P* and *S* arrivals. A pick value of 1 indicates arrival detection, whereas 0 signifies noise. The original model was trained with approximately six hundred thousand labeled samples from the Northern California Earthquake Data Center Catalog. In contrast, EQTransformer incorporates several convolutional, recurrent, and residual stacked layers. It features a hierarchical attention mechanism (Vaswani *et al.*, 2017), comprising a deep encoder directing attention to the earthquake signal and three separate decoders generating triangular probability distributions to predict the earthquake signal's detection and the *P* and *S* phases. EQTransformer was trained with the STanford Earthquake Data set (Mousavi, Sheng, *et al.*, 2019), a large-scale global data set of labeled earthquake and nonearthquake signals. In addition, data augmentation, an artificial technique used to create new data from the existing data, was employed during training to enhance the model's performance.

Although both deep-learning models were developed using different neural network approaches and training data sets, they demonstrated outstanding performance in generalizing the task of picking seismic phases across diverse sources and regions not included in their training data (Park *et al.*, 2020; Jiang *et al.*, 2021; Chin *et al.*, 2022; Gong and Fan, 2022; Münchmeyer *et al.*, 2022; Miller *et al.*, 2023). This suggests that these models are applicable to various environments, regardless of geological context or source phenomena. We applied both models for event detection and phase picking in seismic waveforms recorded and generated by various stations and sources within the Colombian territory. Because these models were already pretrained, they were run in a prediction mode with an efficient and lightweight processing. Upon comparing the results obtained from both models using a subset of the data from the CM network, we preferred EQTransformer to detect and pick seismic phases simultaneously at each station.

For phase association, we use Gaussian mixture model associator (GaMMA; Zhu *et al.*, 2022), an unsupervised ML algorithm that uses a BGMM for earthquake phase association, determining preliminary earthquake locations while optimizing

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

the maximum-likelihood criterion. It has shown positive results with automatic phase-picking algorithms; due to its unsupervised nature, it does not require any training and works for any station geometry. Although GaMMA can generate preliminary earthquake locations, these locations are not robust. Therefore, we use the NonLinLoc (NLLoc) algorithm (Lomax *et al.*, 2000) to obtain more reliable hypocentral locations. Finally, we estimate the local magnitude of each localized event.

This implementation was done to get an automatic catalog for almost 7 yr of data (January 2016–September 2022) in seismic networks located in northern South America. The Colombian national seismic network (period analyzed: January 2017–September 2022) and two temporary local networks in northern South America: the Middle Magdalena Valley (VMM, abbreviation in Spanish) and the CARibbean-Mérida Andes (CARMA) seismic arrays (period analyzed for VMM: January 2016–September 2022; period analyzed for CARMA: January 2016–January 2018). The result is an automatic catalog with good quality in terms of the event location errors, produced in much less time than it could have taken to do it manually, and which defines the major tectonic structures and illuminates some crustal faults in northern South America.

Data and Methods

We downloaded all publicly available seismological data and metadata from January 2016 to September 2022 of three-component multichannel stations of the three seismic networks shown in Figure 1. The Colombian seismic network (CM) is operated by Servicio Geológico Colombiano (1993) (SGC), the Colombian Geological Service. It has a regional geometry that seeks to cover all the national seismic activity, mainly located in the Andean zone and on the Pacific and Caribbean coasts. Using the International Federation of Digital Seismograph Networks Web Services (FDSNWS), CM network only has available data since 2017, then the period analyzed for CM network was January 2017–September 2022. VMM array is a local portable seismic network, also operated by the SGC, and was installed in 2014 in the VMM due to the interest to develop a seismic baseline catalog due to shale gas and unconventional oil exploration. Although there was only FDSNWS data available from 2017, for this local network we requested an additional year of data from the SGC. Then, the period analyzed for VMM network was January 2016–September 2022. And finally, CARMA array is an experimental local temporary network installed for two years (2016–2018) and designed to explore the dynamics of flat slab subduction and plate-edge tectonics in northern South America (Levander, 2016). The period analyzed for CARMA network was January 2016–January 2018.

This combined dataset was used as input for our automatic monitoring workflow. For event detection and phase picking, we tested the performance of the original pretrained PhaseNet and EQTransformer deep-learning models on 1 yr and 1 month (December 2019–January 2021) of data of the CM network. This period was considered to cover several types of seismic sources, such as cultural and instrumental noise, explosions, several earthquakes with different hypocentral locations and magnitude values, including Bucaramanga nest events at a depth of 150 km, induced seismic events, and a set of aftershocks.

PhaseNet and EQTransformer were configured with default processing window settings. PhaseNet used a processing window of 3000 samples with a 50% overlap, whereas EQTransformer used a processing window of 6000 samples with a 30% overlap. Regarding the *P*- and *S*-phase picking probability thresholds, PhaseNet was set to 0.3, whereas EQTransformer was set to 0.01, considering that its event detection threshold was set to 0.3.

Based on the results obtained, EQTransformer was identified as the preferred method to detect and pick seismic phases simultaneously at each station during the entire period of the present study (January 2016–September 2022). The automatic catalog of picks was associated with GaMMA, an unsupervised deep-learning algorithm that is fast and easy to implement and the output of which also provides a preliminary location and event origin time. GaMMA was executed using BGMM, considering 4 *P* and 2 *S* phases as minimum for the associated earthquakes, 2 s for maximum phase residual, 7 km/s for *P* velocity to be able to consider deep events, 1.75 for V_P/V_S ratio, and no amplitude information. GaMMA allows the use of the Density-based spatial clustering of applications with noise (DBSCAN) algorithm (Schubert *et al.*, 2017) to improve initialization strategies by dividing picks into sub-windows for association. This approach enhances computational efficiency and increases association performance. In this work, we defined the following DBSCAN hyperparameters: 20 s for epsilon and 4 for the minimum number of samples, which correspond to the maximum time between two picks for one to be considered a neighbor of the other and the number of samples in a neighborhood for a point to be considered a core point.

Although preliminary GaMMA locations are good enough to allow us to visualize the main seismic activity, we use NLLoc to further improve earthquake locations (Lomax *et al.*, 2000) by interpolating a 1D velocity model to 3D, as defined by Ojeda and Havskov (2001) for Colombia. In this case, the simplicity of the velocity model was chosen to preserve as many events as possible for earthquake location. Although using more robust velocity models, such as the 3D anisotropy velocity model (Poveda *et al.*, 2018), could be interesting, it might not be ideal for associations that have not been manually checked. This approach could decrease the number of localized earthquakes by not converging to the earthquake location required by robust velocity models. We employed NLLoc with the Gaussian analytic location method, which employs the nonlinear probabilistic inversion approach (Tarantola and Valette, 1982), along with the oct-tree sampling algorithm. This algorithm uses a recursive 8-cell subdivision to find the maximum-likelihood point in 3D space based on the selection of the cell with the lower misfit between the observed arrival times and theoretical travel times.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

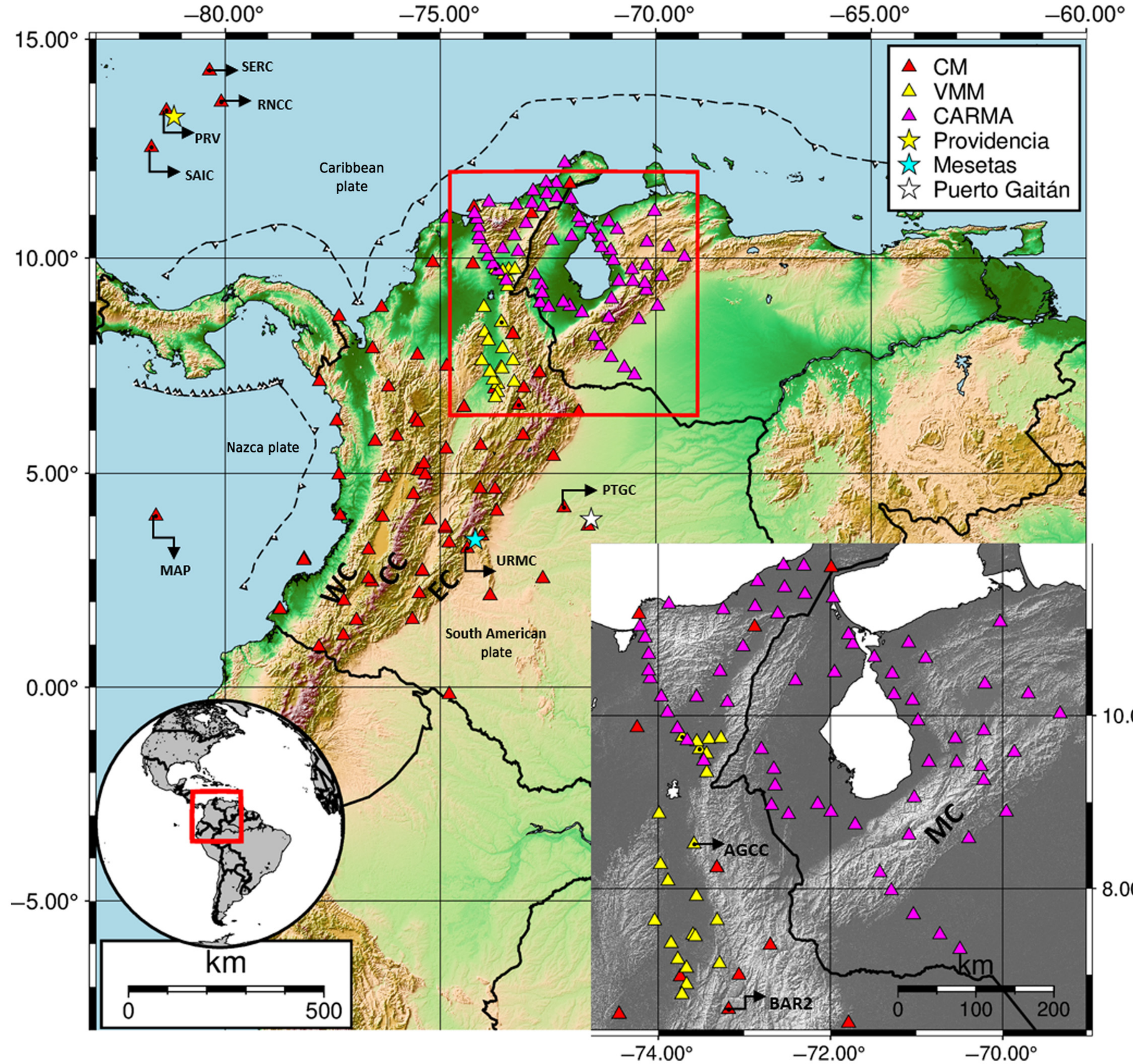

Figure 1. Permanent seismic network: the Colombian seismic network with red triangles. **Temporary seismic networks**: (1) the Middle Magdalena seismic array with yellow triangles and (2) the CARibbean-Mérida Andes seismic array with magenta triangles. **Isolated stations**: 1. MAP in Malpelo Island. 2. PRV, RNCC, SAIC, and SERC in the San Andres Archipielago. Other highlighted stations are referenced in the analysis. **Mainshocks**: 25 November 2018 M_w 6.0 Providencia, 24 December 2019 M_w 6.0 Mesetas, and 22 March 2021 M_w 5.1 Puerto Gaitán. CC, central Cordillera; EC, eastern Cordillera; MC, Cordillera de Mérida; WC, western Cordillera. The color version of this figure is available only in the electronic edition.

This approach provides an accurate, efficient, and complete mapping of the earthquake location. NLLoc program was applied exhaustively until several quality parameters were met; otherwise, the event was removed. The exhaustive process consisted of gradually eliminating picks from distant stations until the quality parameters were met. The quality parameters are the following: epicentral errors less than 0.15°, depth errors less than 10 km, no negative depths, and root mean square less than 2.5 s.

Finally, we estimated the local magnitude M_L for each event calculated using the following equation, which is not considering the station-correction function but is considering the maximum amplitude A in mm on the two horizontal components simulating a Wood–Anderson seismograph, a distance-correction function in terms of the geometrical spreading, and the anelastic attenuation of the medium:

$$M_L = \log A + a \times \log(r/r_{ref}) + b \times \log(r - r_{ref}) + K(r_{ref}). \quad (1)$$

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

We used the traditional calibration parameters for the distance-correction function in the Colombian territory (Rengifo and Carriazo, 2004), with a = 1.019 related to the geometrical spreading, b = 0.0016 related to the anelastic attenuation factor, $r_{\mathrm{ref}} = 140$ km is the reference distance that establishes a common point with the traditional Richter scale in California, and $K(r_{\mathrm{ref}})$ is the base level or the reference magnitude when M_{L} 3. These values serve as a general and approximate means to estimate the local magnitude for the entire Colombian territory, which we find ideal for simplifying the magnitude estimation process. This approach proves effective enough to produce M_{L} values comparable to those reported by the SGC, which specifies magnitudes by zones (Lopez *et al.*, 2020). We avoid this level of detail because it would require more analysis in our workflow, including consideration of mapped lithologies and updated attenuation values.

Results and Discussion

Event detection and phase picking

We conducted a comparative analysis between PhaseNet, EQTransformer, and manual picks reported by SGC. The SGC picks database corresponds to different classifications of events: local and international earthquakes, as well as non-locatable earthquakes and explosions. All manual picks of these events were used in this analysis. Our analysis focuses on a subset of data spanning approximately one year, encompassing various signal types originating from diverse sources, including both noise and earthquake events. Figure 2 illustrates pick results using records from a set of aftershocks in Mesetas, Colombia, and a signal affected by electronic issues. Regarding event detection, both PhaseNet and EQTransformer outperform SGC. PhaseNet exhibits the ability to pick numerous phases and detect very small events, but it also generates numerous false picks in stations with instrumental issues. In contrast, EQTransformer demonstrates slightly lower recall than PhaseNet but offers higher accuracy, resulting in a reduced number of false positives and assigning low probabilities to potential mistakes.

In general, we found that PhaseNet has difficulty in evading some spike noise signals that can happen very often in the seismic record due to different types of noise that influence the stations, such as cultural noise or noise due to equipment malfunction.

Figure 3 depicts that the number of picks obtained by PhaseNet is consistently higher for all stations compared to the manual catalog. In contrast, EQTransformer shows a similar number of picks to the manual catalog for several stations. EQTransformer tends to pick a higher number of phases than the manual method, particularly for stations that are isolated or distant from each other.

This observation aligns with the typical manual processing approach, in which analysts tend to allocate more attention to stations in close proximity to each other. This strategy is motivated by the expectation of obtaining a sufficient number of picks for an accurate event location. In contrast, isolated stations may receive comparatively less attention from analysts.

For example, stations PRV, RNCC, SAIC, and SERC that are located in the San Andres Archipielago, or station MAP in Malpelo Island, a small oceanic island in the Pacific Ocean, are some of the most isolated stations (a few nearby stations are shown in Fig. 3). If the picks of the automatic algorithms are, in fact, real P and S phases, it would mean the manual catalog in these cases is incomplete.

Is the manual catalog incomplete? As an example, we plotted several traces around the picks from the EQTransformer arrivals (Fig. S1, available in the supplemental material to this article) and can visually confirm that many of these picks are valid and true detections. This suggests that the significant increase in detections, especially in isolated stations, can be true. They are likely not useful for locating the earthquakes but shows that the algorithms have in fact the potential to complement the manual catalog for individual stations.

There are also a few specific nonisolated stations, like station AGCC, which has more than 30 stations within a 200 km radius. For these stations, both models pick a higher number of phases. Given the trend of good performance for nonisolated stations, we consider that the station was not completely manually picked.

Verifying the veracity of the picks databases is challenging because there is no certainty about which one constitutes the complete database. Initially, it may seem reasonable to assume that the SGC's picks database is complete, given that SGC analysts review all raw seismic signals rather than relying solely on automatic algorithms. However, ML picking models have been shown to increase the number of manual picks (Zhu and Beroza, 2019; Mousavi *et al.*, 2020). Therefore, we constructed a confusion matrix with the following logic: Iteratively, we assumed one true picks database and analyzed the veracity of the other picks databases (test databases) in relation to it, searching for pick correspondences. Correspondence occurs if the time difference between them is less than 1.5 s, sufficient time to capture P and S phases with reliable probabilities, with the S phase being the most vulnerable to time discrepancies as probability decreases (Fig. S2). Although a 1 s time difference might suffice to capture most phases with higher probabilities, we chose to extend it by an additional 0.5 s to encompass more S phases in our analysis.

Figure 4 shows the confusion matrix for the SGC's manual picks database and EQTransformer's and PhaseNet's picks databases with the initial probability thresholds specified in the Data and Methods section, as well as for probability thresholds greater than 0.5 and 0.7 to assess model performance when restricting probability. Overall, PhaseNet with its respective probability thresholds demonstrates that it can achieve a high percentage of picks from both the SGC and EQTransformer databases (Fig. 4, upper-left confusion matrix for both phases). However, it is also evident that SGC and EQTransformer

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

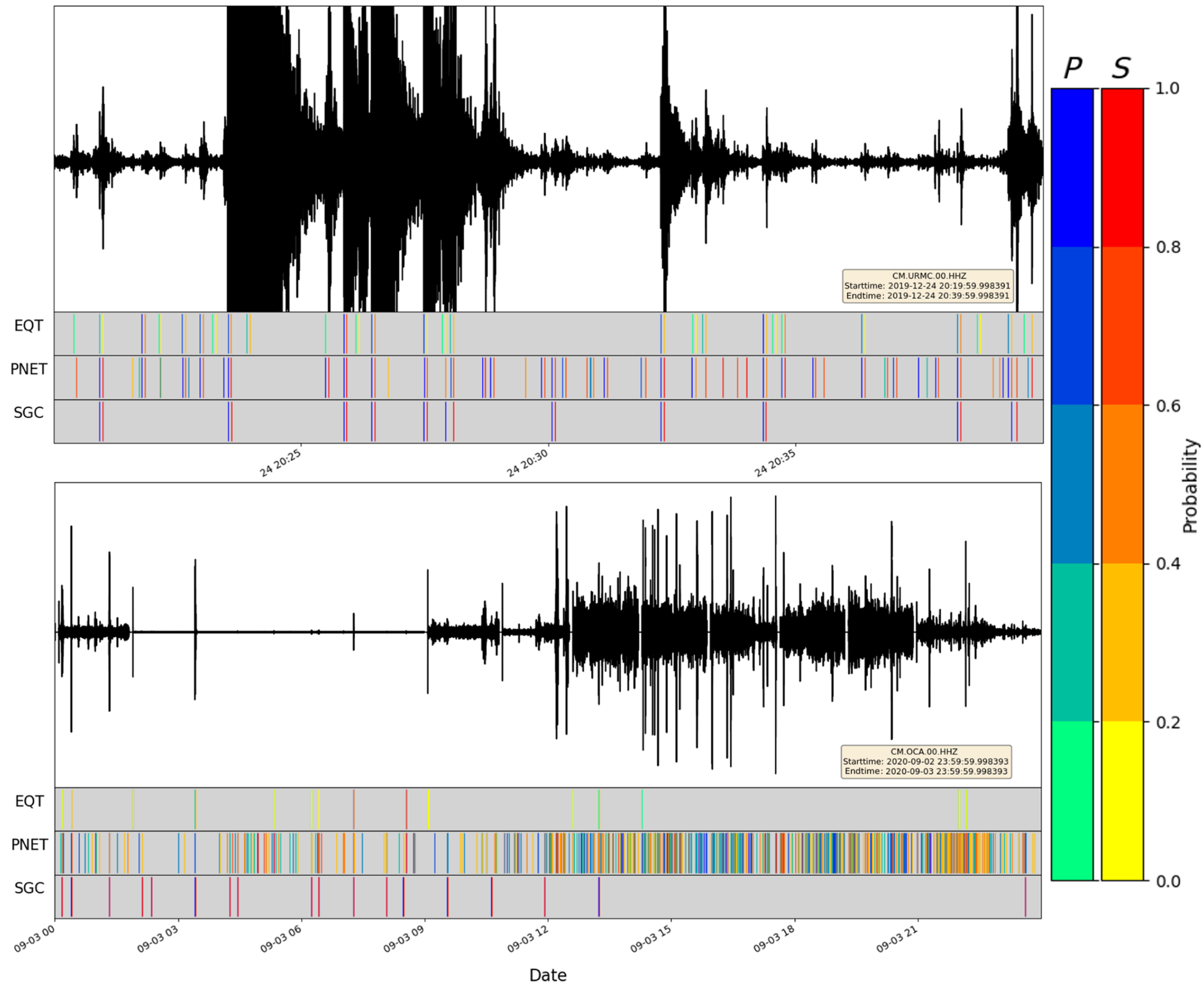

Figure 2. Examples of phase-picking results. Upper panel shows 20 min of a set of aftershocks 1.28 hr after the mainshock in Mesetas (M_w 6.0) at 24 December 2019 19:03:52 UTC recorded by URMC station, the closest station of the epicenter at that time (32 km). Lower panel shows one day of data in a station with technical issues in the digitizer. Vertical lines in color represent the arrival timestamp picked by EQTransformer, PhaseNet, and manual analysts of the Servicio Geológico Colombiano (SGC). Cold and warm colors represent *P* and *S* picks, respectively, and the more intense the color, the higher the probability of the pick. The color version of this figure is available only in the electronic edition.

capture very few picks from PhaseNet's total picks (Fig. 4, bottom-right confusion matrix for both phases). In other words, as mentioned throughout the discussion, PhaseNet manages to obtain a considerable number of picks that EQTransformer and SGC can identify, but even for picks with probability ≥0.7, its number of picks surpasses what EQTransformer and SGC can obtain.

On the other hand, EQTransformer demonstrates its ability to capture 43% of *P* picks from the SGC, which is 20% less than what PhaseNet can achieve (63%). However, SGC can obtain 20% of EQTransformer's total number of *P* picks, which increases to 34% if we restrict *P*-phase probabilities to 0.5 or 0.7. Yet, in these two cases, EQTransformer's ability to detect manual picks decreases significantly (22% and 13%, respectively). Therefore, we prefer to keep the initial EQTransformer probability thresholds, relying on 20% of its picks. In contrast, when considering PhaseNet, we would only have confidence in 2% of its data.

Based on this analysis, we consider that EQTransformer, with the initial probability thresholds, represents the best scenario in our testing. It manages to capture 43% of the manual data, and we can trust at least 20% of its picks (indicated by rectangles with red borders in Fig. 4). Some of the remaining 80% of picks from EQTransformer are likely true picks that are not recorded in the SGC's database, as mentioned for picks from isolated stations. For *S* picks, the analysis

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

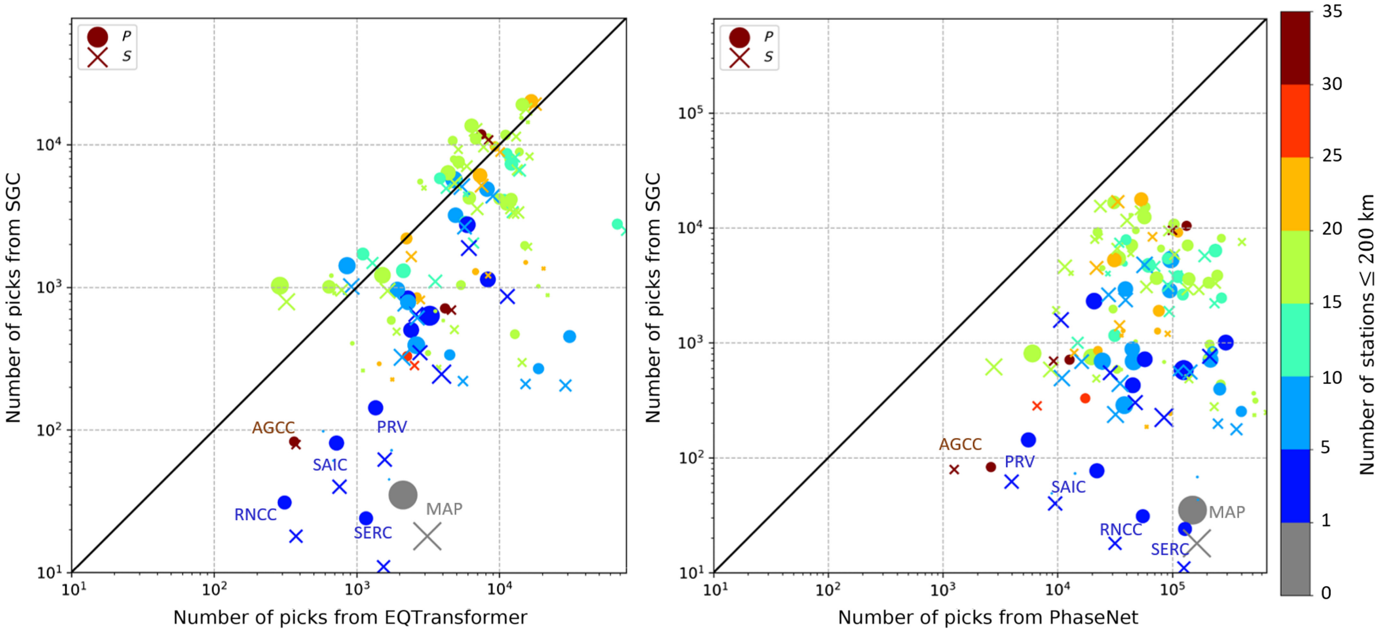

Figure 3. Comparison by station of the number of picks between automatic and manual pickers in one year of data (approximately December 2019–January 2021). The circles and crosses represent the *P* and *S* phases, respectively. The color bar represents the number of stations the station has within a 200 km radius, and the size is proportional to the distance between the station and its closest station. The color version of this figure is available only in the electronic edition.

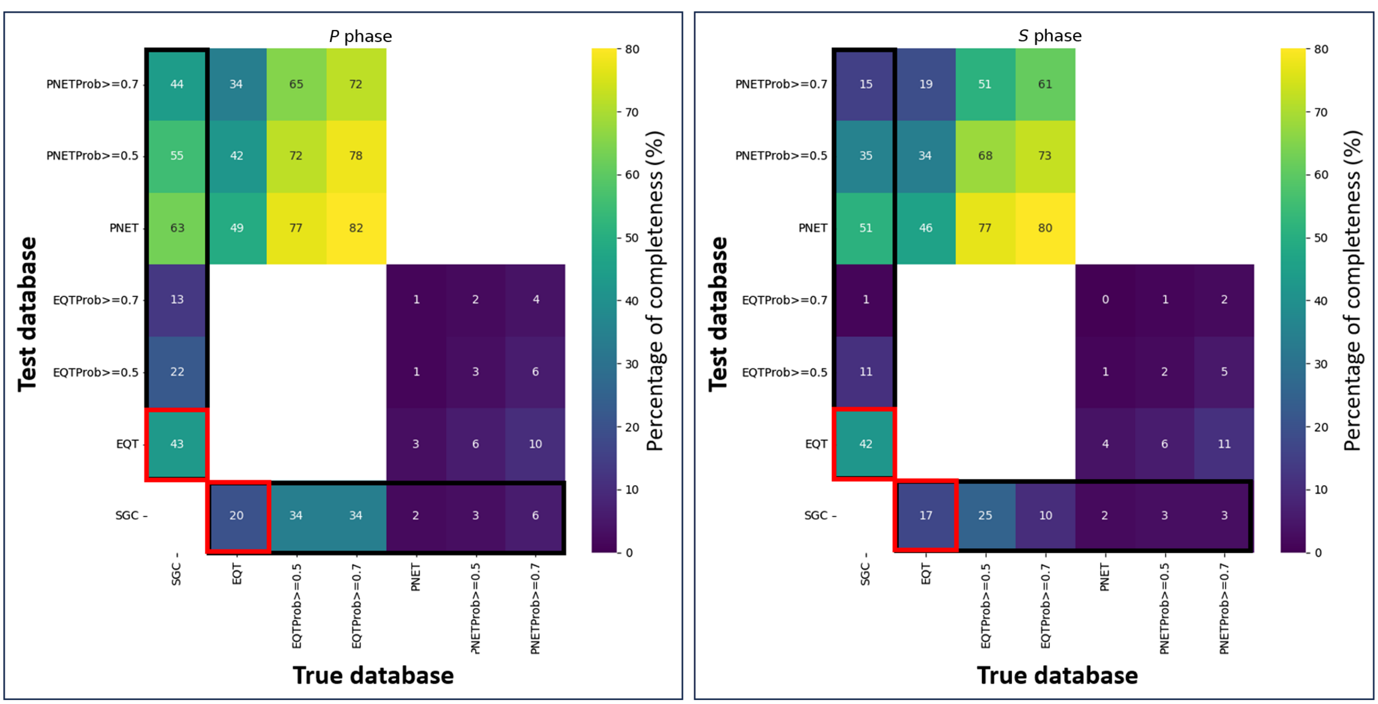

Figure 4. Confusion matrix for both *P* and *S* picks databases for 1 yr and 1 month (December 2019–January 2021). PNET and EQT are using the initial probability thresholds defined in the Data and Methods section. Other thresholds are also shown to review the model's performance when the probability is restricted. Rectangles with black borders represent the model's performance compared with the SGC database picks. Rectangles with red borders indicate the best scenario to verify the veracity of the picks. The color version of this figure is available only in the electronic edition.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

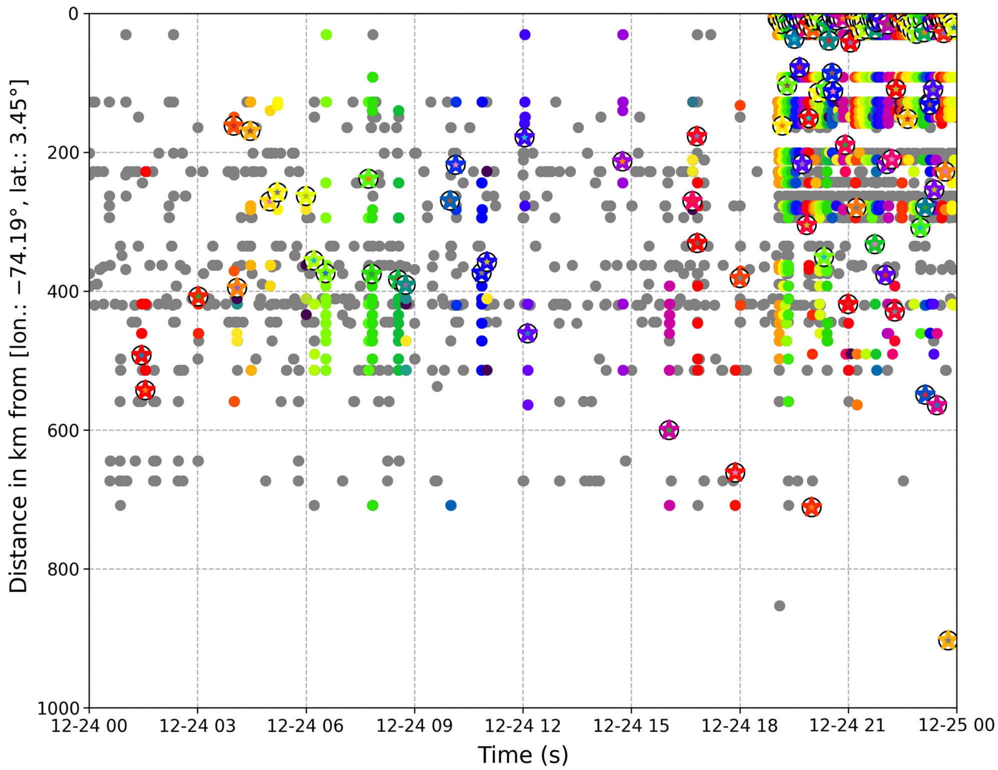

Figure 5. GaMMA phase association on 24 December 2019. Stars within circles represent events based on their origin time and distance from the epicenter to the coordinates: latitude 3.45° and longitude −74.19°, situated in the Municipality of Mesetas. Colored circles depict associated phases corresponding to their respective sources, whereas circles without color indicate unassociated phases. Around 19:00, there are associations related to the series of aftershocks following the main earthquake in Mesetas M_w 6.0 The color version of this figure is available only in the electronic edition.

is quite similar to that presented for the *P* picks. Finally, we strongly believe that these results can be notably improved if both models are retrained or fine-tuned with Colombian data, as shown by previous research in other regions (Chai *et al.*, 2020; Lapins *et al.*, 2021; Ni *et al.*, 2023; Zhu *et al.*, 2023; Niksejel and Zhang, 2024).

In conclusion, the number of picks detected by EQTransformer has a scale similar to that reported by the SGC, whereas PhaseNet causes an overflow. Furthermore, although PhaseNet detects a larger number of arrival times, thus potentially detecting smaller earthquakes, EQTransformer inspires higher confidence in identifying authentic *P* and *S* phases. Based on these comparisons, we will use from here forward EQTransformer as the automatic phase picker to produce our earthquake catalog in this study between January 2016 and September 2022.

Phase association

GaMMA stands out from traditional association algorithms because it does not rely on standard procedures like grid search or supervised training. It effectively associates phases from various sources at different epicenter distances. Figure 5 illustrates its effectiveness in associating phases during the earthquakes that occurred on 24 December 2019, across different locations in the country. Furthermore, GaMMA demonstrates its capability in handling a series of aftershocks following the Mesetas mainshock of M_w 6.0. It not only successfully associates phases that occur close in time and space but also provides approximate yet reasonably accurate locations for these events. For example, Figure 5 indicates that the aftershocks are located a few kilometers away from the coordinates latitude: 3.45° and longitude: −74.19°, which fall within the Mesetas municipality where the mainshock occurred. Therefore, GaMMA's association approach offers preliminary insights into the seismic activity of the region.

Given that the seismic network used in this study is pretty disperse, with a significant station separation, many picks were discarded during the association step (a bottleneck that we wish to address in future work). Figure 6 shows the number of picks associated throughout the entire study period, based on the picks' probability levels. The associated picks for both the *P* and *S* phases were less than 30%. The percentage of *P*-phase association increases as the pick probability increases, whereas for the *S* phase, the maximum probability does not correlate with the larger number of associations. Therefore, even *S* picks with high probabilities, which likely correspond to true picks, are not all associated with an event and are discarded.

There are several reasons why the association algorithm could not associate a large number of picks. First, there are not enough picks to locate the event, so the algorithm does not converge to a reliable solution. Second, there are some false picks taken as true picks, introducing noise to the data. Third, there are several picks close in time and space, which makes the task considerably more difficult. And finally, there are several true picks from stations at greater distances, so the theoretical travel time of the event to the station could be computed in a wrong way, considering the simple velocity model used in this

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

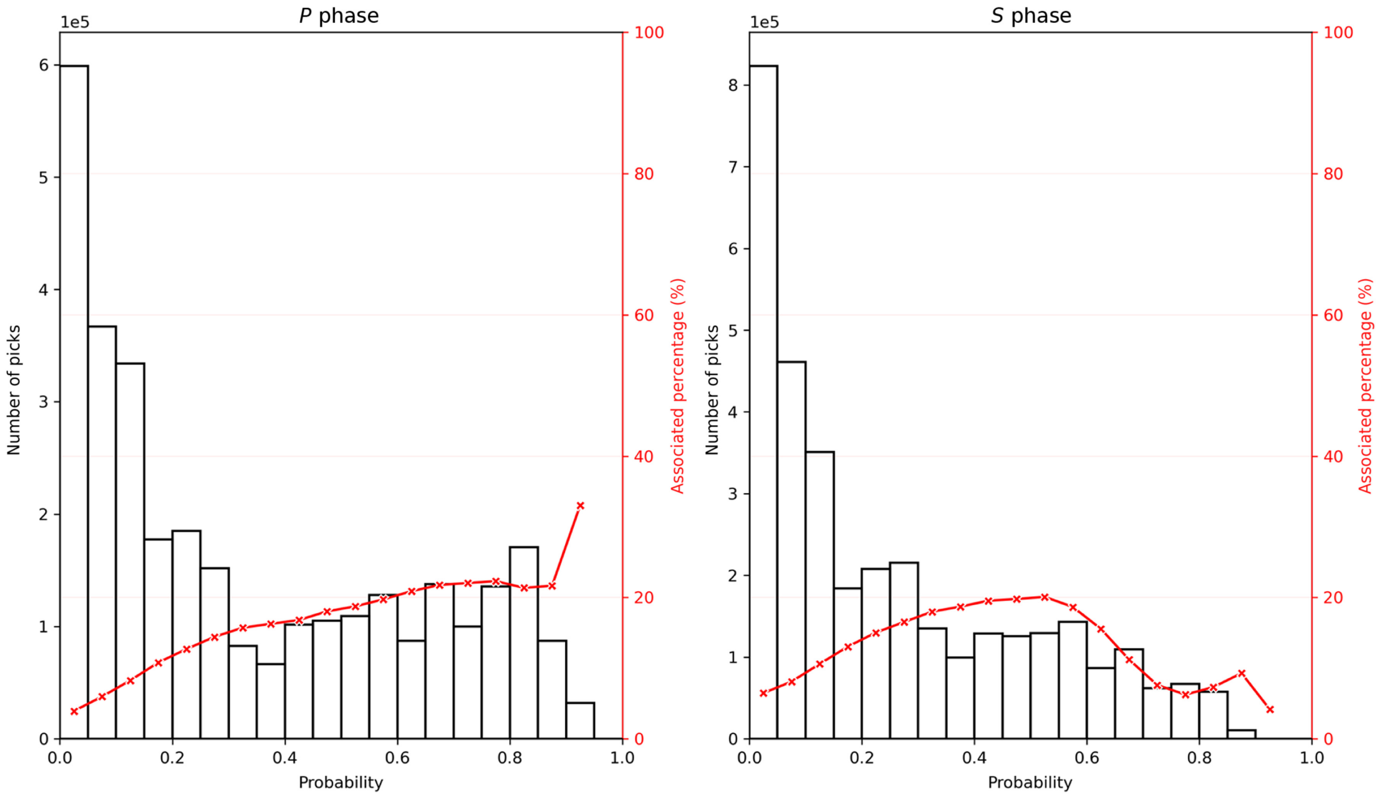

Figure 6. Number of phases and association percentage as a function of EQTransformer probabilities for the entire study period (January 2016–September 2022). The color version of this figure is available only in the electronic edition.

case. It implies the GaMMA loss function would be higher than expected, and then it could not optimize the earthquake parameters at the M-step properly for the Gaussian mixture model.

GaMMA provides a preliminary hypocentral location that generally illuminates the seismic activity in northern South America, as shown in Figure 7a. The seismic activity provides insight into the significant subduction mechanisms resulting from the relative convergence among the Nazca, Caribbean, and South American plates. There is also shallow seismicity along the Andes mountain chain and some earthquake mainshocks with their respective aftershocks, such as the 25 November 2018 M_w 6.0, Providencia, Colombia (Bishop *et al.*, 2022), and the 24 December 2019 M_w 6.0, Mesetas (Mayorga *et al.*, 2021), Colombia. December 2019 was the month with the highest number of earthquakes due to a big series of aftershocks in Mesetas, followed by other aftershocks in March 2021 induced by massive wastewater injection near Puerto Gaitán, Colombia (Molina *et al.*, 2020; to observe 1 day of recorded aftershocks at the station closest to the event, see Figs. S3 and S4).

Seismicity is expected to increase over time due to the growing number of stations and available data, as illustrated in the early months of 2016, when the CARMA network was included in the analysis between 2016 and 2018 (Fig. 7b,c). During this period, the catalog recorded the highest number of earthquakes, particularly those with magnitudes between M_L 1 and 2. From January 2018 to September 2022, the number of stations varied between 35 and 50, averaging 45 stations. Despite the relatively constant number of stations since 2018, seismic activity seems to have decreased mainly from March to December 2020. During this period, the available stations reached their lowest value, suggesting a notable impact on the association step, particularly for non-dense networks in which some stations are crucial for the association process.

Although GaMMA demonstrates proficiency in illuminating both shallow and deeper seismicity and excels in associating picks within sets of aftershocks, it is essential to underscore the critical role of the association step in the seismic monitoring process. The substantial information loss at this stage represents a bottleneck in advancing earthquake detection and location accuracy. Developing algorithms capable of harnessing the wealth of phase picks provided by neural network algorithms in regional networks would constitute a significant leap forward.

Automatic seismic catalog

GaMMA seismic catalog was relocated using NLLoc, removing events with large uncertainties. Figure 8 shows the seismic catalog after removing events with few arrivals or large uncertainties, leading to a catalog with only half of the earthquakes of the GaMMA preliminary locations. Most of the seismicity in Providencia was removed using this procedure because those

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

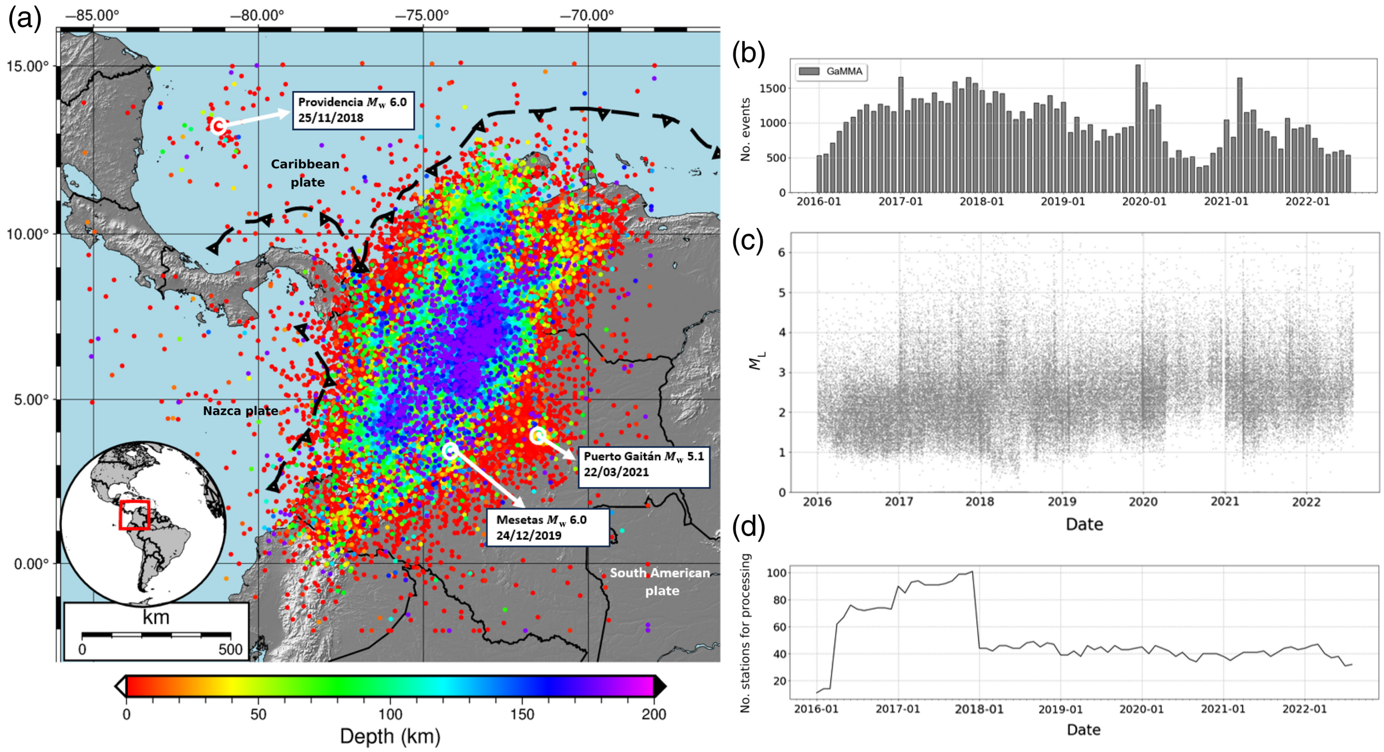

Figure 7. GaMMA seismic catalog. (a) Plan view of the seismicity. (b) Histogram of events by month. (c) Magnitudes as a function of time. (d) The number of stations used for processing as a function of time. The color version of this figure is available only in the electronic edition.

events had high uncertainty due to the low number of stations and the poor azimuthal coverage of the network in that region, or because the velocity model used was not the best to converge the location of the events. Nevertheless, the rest of the seismicity is robust, both in epicentral location and depth.

The primary shallow seismic activity in Colombia is predominantly situated in a north-northeastern direction along the eastern Cordillera frontal fault system. Illustrated in Figure 9a, this system is positioned at the boundary between the eastern Cordillera and the South American shield. It encompasses the Santa Marta-Bucaramanga fault system trending north-northwest, one of the three strike-dip faults that compose the triangular Maracaibo block (Fuenzalida *et al.*, 1998; Londoño *et al.*, 2019). The other two are the Oca fault with an east–west azimuth and the Boconó fault (running parallel to the Cordillera de Mérida) with a northeast azimuth. In addition, it encompasses the Romeral fault system, stretching along the Cauca-Patía Valley and between the central and western Cordillera. Furthermore, there are identifiable seismicity clusters: two associated with a series of aftershocks in Mesetas and Puerto Gaitán, and other clusters linked to detonations in mining areas. Profile A–A′ displays three shallow alignments with depths shallower than 30 km (Fig. 10). The first corresponds to the Romeral fault system, whereas the second and third alignments exhibit a higher frequency of earthquakes due to their association with the Mesetas and Puerto Gaitán aftershocks, respectively. Some events are fixed at 5 km as an artificial artifact due to the change in the layer of the velocity model.

In terms of the intermediate seismicity, Figure 9b shows a lateral shift in intermediate-depth seismicity around 5.5°N (Ojeda and Havskov, 2001). Correspondingly, volcanic activity halts at the same latitude around 5 million years ago (Wagner *et al.*, 2017; Kellogg *et al.*, 2019), leading to diverse interpretations of the subducted slab at depth. A recent viewpoint posits this shift as a tear in the Nazca plate fostering two distinct subduction styles (Vargas and Mann, 2013; Chiarabba *et al.*, 2016; Syracuse *et al.*, 2016): a typical normal subduction creating a volcanic arc to the south and a flat subduction devoid of volcanism to the north. Another perspective suggests the presence of two distinct and complex subduction processes: one involving ocean–continent subduction to the south and the other involving continent–continent subduction to the north (Fuenzalida *et al.*, 1998). Previous studies refer to the southern slab as the Cauca segment and the northern slab as the Bucaramanga segment. It is hypothesized that the latter, attributed primarily to the tectonic processes associated with the Caribbean plate, might also be influenced by the stresses resulting from its interaction and overlap with the Nazca plate (Taboada *et al.*, 2000; Kellogg *et al.*, 2019; Sun *et al.*, 2022). These segments are clearly delineated in profile B–B′ (Fig. 10). The Bucaramanga segment exhibits deeper events and offers a clearer view of the subducting slab. This profile also highlights an intermediate seismic cluster occurring between depths of 140

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

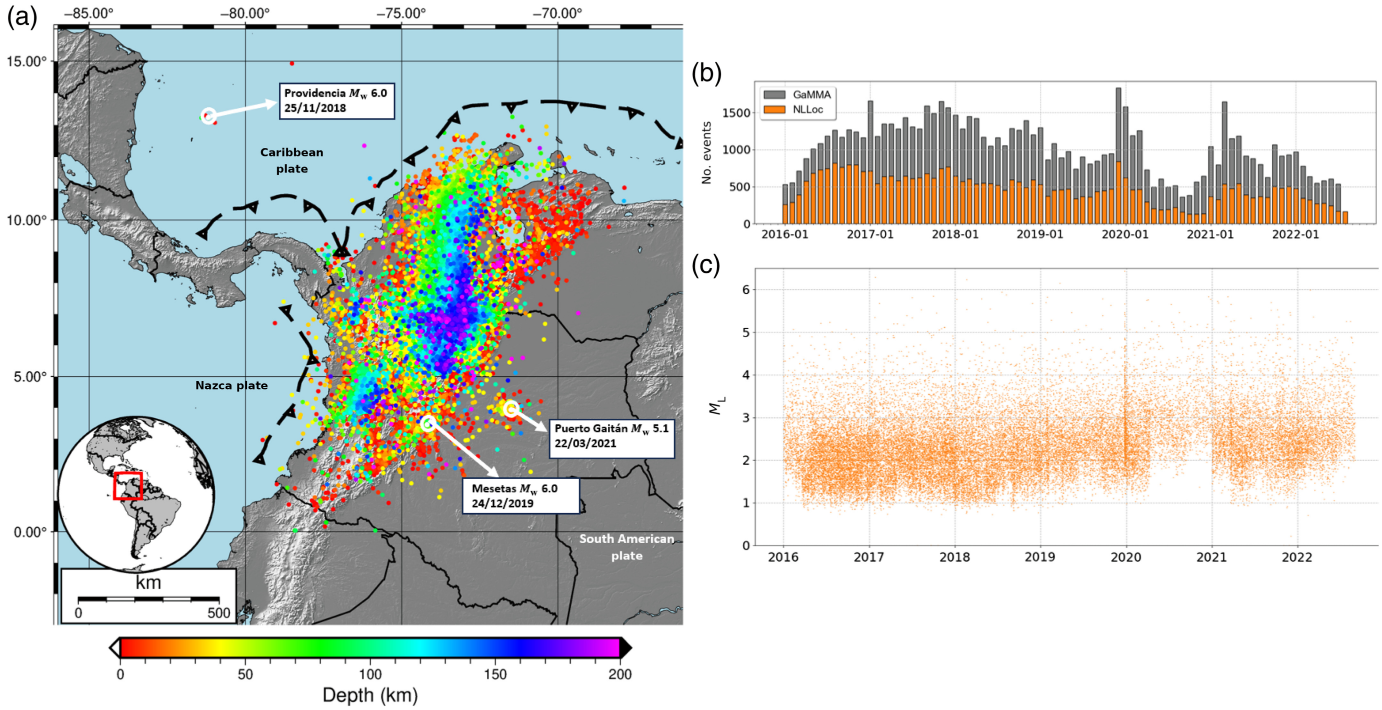

Figure 8. NonLinLoc (NLLoc) seismic catalog. (a) Plan view of the seismicity. (b) Histogram of events by month. (c) Magnitudes as a function of time. The color version of this figure is available only in the electronic edition.

and 160 km within the Bucaramanga segment, located at coordinates 5.31 N and 73.77 W, in proximity to the municipalities of Ubaté and Cucunubá. As of now, there remains limited information available regarding this specific seismic cluster.

Our catalog shows a deep east–west streak of seismicity across the Bucaramanga nest, a region with a high rate of seismic activity within a limited volume (Prieto *et al.*, 2012). However, this streak is an artificial artifact generated during the first two years of analysis (2016–2018) due to the influence of the CARMA network geometry biasing this region. Nevertheless, the alteration in strike and dip is evident at the latitude of the Bucaramanga nest, dividing the Bucaramanga segment into two sections. We approximate that the northern section of the segment demonstrates a strike of approximately ~N5°W, whereas the southern part displays a strike of approximately ~N27°E. Profile C–C′ illustrates how the slab dip varies slightly with depth at the Bucaramanga nest depth (Fig. 10), which assumes an elliptical shape elongating in the down-slip direction, consistent with previous observations (Zarifi *et al.*, 2007). Our findings indicate that this nest is positioned around 6.82 N and 73.10 W, spanning a depth range of roughly 130–145 km. However, it remains an open question regarding the underlying physics driving its behavior (Yarce *et al.*, 2014; Kellogg *et al.*, 2019).

The catalog also shows a north–south intermediate seismicity band at the western edge of the C–C′ cross section, which has been difficult to notice; however, it has been mentioned in previous studies (Sun *et al.*, 2022; Martinez and Prieto, 2023). This catalog includes enough analysis time to illuminate it, which could be evidence of the Coiba microplate located in the northernmost Nazca region, the lithosphere of which is young and may be too warm to host intermediate-depth seismicity (Sun *et al.*, 2022).

Finally, to the north of the Bucaramanga segment, the absence of instrumentation has resulted in various interpretations regarding the presence of the Caribbean subducting slab. These interpretations have primarily relied on regional seismic tomography models, local seismicity, or surface wave data (Vargas and Mann, 2013; Chiarabba *et al.*, 2016; Syracuse *et al.*, 2016). The deployment of the CARMA network in our catalog has illuminated shallow and intermediate seismicity in this region, providing valuable insights into the Caribbean subduction process and complementing recent investigations (Cornthwaite *et al.*, 2021; Sun *et al.*, 2022). Profile D–D′ illustrates the Benioff zone, suggesting flat subduction in which the slab dip was estimated at 28° between depths of 30 and 140 km (Fig. 10). This finding aligns with previous discussions, indicating that the Caribbean slab typically subducts at a shallow angle, generally less than 30°, at shallow depths (Cornthwaite *et al.*, 2021).

Conclusions

Considering the workflow for processing passive seismic data, detection and picking seismic phases could represent the most time-consuming task. In this study, we conducted an extensive evaluation of the performance of two pretrained deep-learning models: PhaseNet and EQTransformer. Both models exhibit high adaptability and operational efficiency, capable of rapidly processing vast data volumes without the need for sophisticated

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

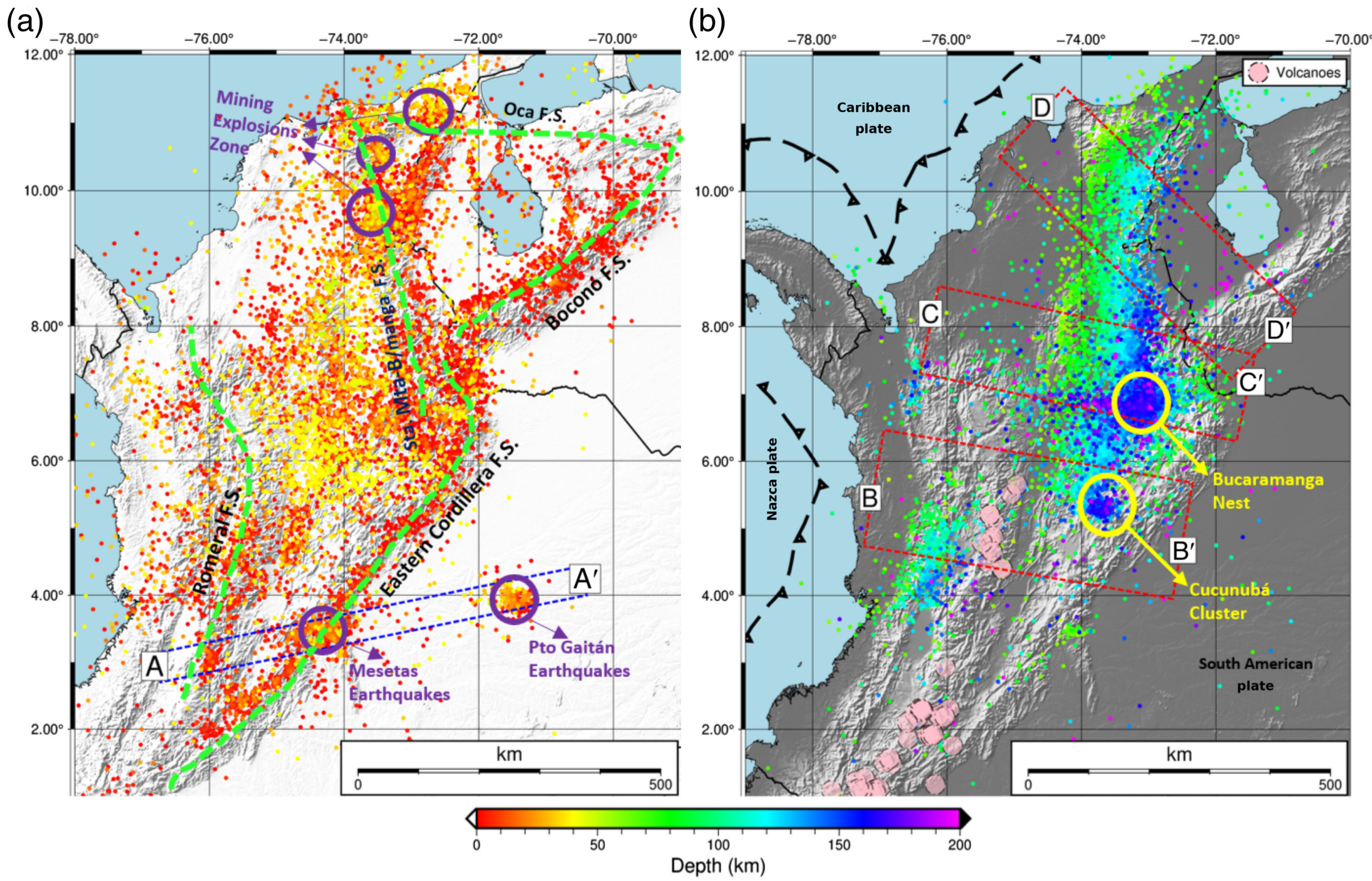

Figure 9. NLLoc seismic catalog illuminating crustal faults, clusters of seismicity, and deeper tectonic structures. Dashed circles represent clusters of seismicity, and dashed rectangles represent seismic profiles at depth. (a) Plan view for depths ≤30 km. (b) Plan view for depths ≥50 km. The color version of this figure is available only in the electronic edition.

hardware. Moreover, they showcase remarkable proficiency in seismic phase detection, producing results comparable to manual picking and even more effective for small earthquakes or aftershocks. Although PhaseNet excels at identifying minor earthquakes, EQTransformer instills higher confidence in the results.

In the phase association step, using the phases detected by EQTransformer (preferred), we found GaMMA to be a highly effective algorithm, easy to implement in any network, and good enough to illuminate the shallow and deeper seismicity, besides providing very good results to associate picks in a set of aftershocks. However, a large number of true picks were removed for different reasons. We consider the most important reasons it is because there were not enough picks to associate with a single event or because there were several true picks from stations far away from the event, making GaMMA unable to optimize the earthquake parameters in M-step properly. GaMMA catalog was relocated using NLLoc, removing events with high uncertainty, which illuminates crustal faults and deeper structures: the main crustal fault systems parallel to the foothills of three Colombian Cordilleras and the marked shift in the intermediate seismicity. Through seismic profiles, we propose the location of both clusters, Bucaramanga nest and Cucunubá cluster, together with other tectonic structures or human-induced activities.

This catalog was produced quickly and was the result from an automatic processing performed mainly by ML models. All the steps carried out were condensed into SeisMonitor, a repository made in Python, allowing its use for any other seismological network. We recommend this workflow to obtain a good seismicity view. For instance, our catalog stands out for its precise event location data, demonstrating a high level of quality that effectively highlights key tectonic features and seismicity trends in northern South America. A similar workflow could potentially be applied to any of the multiple temporary, local arrays that have been deployed around the world.

Data and Resources

The SeisMonitor package, scripts, and Jupyter Notebooks to reproduce the examples are available on GitHub (https://github.com/ecastillot/SeisMonitor). The Servicio Geológico Colombiano (SGC) and CARibbean-Mérida Andes (CARMA) data are available at http://sismo.sgc.gov.co:8080 and https://www.fdsn.org/networks/detail/YU_2016/, respectively. SeisMonitor is developed based on the following packages: ObsPy (Beyreuther *et al.*, 2010) to download and manipulate seismological data, PhaseNet (Zhu and Beroza, 2019) and

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

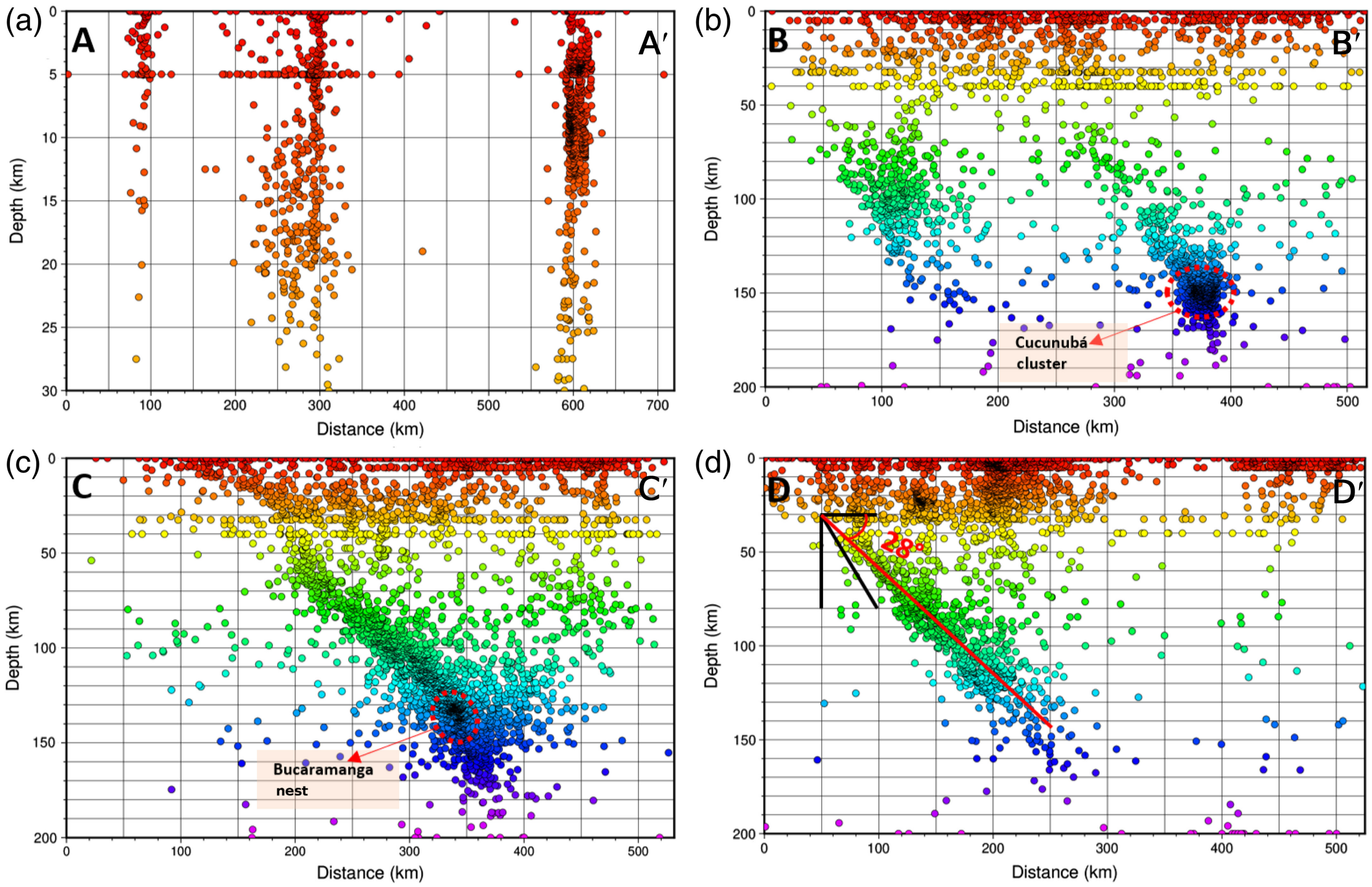

Figure 10. Seismic profiles painted in Figure 9 using the same color bar to represent the depth of the events. (a) Profile A–A′ showing the Romeral fault system, Mesetas, and Puerto Gaitán set of aftershocks. (b) Profile B–B′ showing two slabs associated with the southern slab in the Cauca segment and the northern slab in the Bucaramanga segment, where the Cucunubá cluster is located. (c) Profile C–C′ showing the Benioff zone in the Nazca plate together with the Bucaramanga nest. (d) Profile D–D′ showing the Benioff zone in the Caribbean plate. The color version of this figure is available only in the electronic edition.

EQTransformer (Mousavi *et al.*, 2020) for phase picking, GaMMA (Zhu *et al.*, 2022) for seismic phase association, and NonLinLoc (Lomax *et al.*, 2000) for earthquake location. All websites were last accessed in October 2023. The supplemental material includes a figure that compares the differences in arrival times between manual picks and machine-learning (ML) picking algorithms. In addition, there are other figures that visually represent traces from the MAP, URMC, and PTGC stations. These figures serve to elucidate the accuracy of recorded picks in isolated stations and illustrate the aftershock sequences following the Mesetas and Puerto Gaitán earthquakes.

Declaration of Competing Interests

The author acknowledges that there are no conflicts of interest recorded.

Acknowledgments

Special gratitude to the Colombian Geological Service for authorizing the use of Colombian seismological data and to their seismic analysts and seismologists for their willingness to engage in seismological discussions. The authors also thank two anonymous reviewers for their valuable comments.

References

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces, *Bull. Seismol. Soc. Am.* **68,** 1521–1532.

Beroza, G. C., M. Segou, and S. Mostafa Mousavi (2021). Machine learning and earthquake forecasting-next steps, *Nat. Commun.* **12,** no. 1, 4761.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A Python toolbox for seismology *Seismol. Res. Lett.* **81,** no. 3, 530–533.

Bishop, B., S. Cho, L. Warren, L. Soto-Cordero, P. Pedraza, G. Prieto, and V. Dionicio (2022). Oceanic intraplate faulting as a pathway for deep hydration of the lithosphere: Perspectives from the Caribbean, *Geosphere* **19,** 206–234.

Chai, C., M. Maceira, H. J. Santos-Villalobos, S. V. Venkatakrishnan, M. Schoenball, W. Zhu, G. C. Beroza, C. Thurber, and E. C. Team (2020). Using a deep neural network and transfer learning to bridge scales for seismic phase picking, *Geophys. Res. Lett.* **47,** no. 16, doi: 10.1029/2020GL088651.

Chiarabba, C., P. D. Gori, C. Faccenna, F. Speranza, D. Seccia, V. Dionicio, and G. A. Prieto (2016). Subduction system and flat slab beneath the eastern cordillera of Colombia, *Geochemistry* **17,** 16–27.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

Chin, S.-J., R. Sutherland, M. K. Savage, J. Townend, J. Collot, B. Pelletier, O. Monge, and F. Illsley-Kemp (2022). Earthquakes and seismic hazard in southern new caledonia, southwest pacific, *J. Geophys. Res.* **127,** no. 12, doi: 10.1029/2022JB024207.

Cornthwaite, J., M. J. Bezada, W. Miao, M. Schmitz, G. A. Prieto, V. Dionicio, F. Niu, and A. Levander (2021). Caribbean slab segmentation beneath northwest south america revealed by 3-d finite frequency teleseismic p-wave tomography, *Geochem. Geophys. Geosys.* **22,** 1–19.

Dickey, J., B. Borghetti, W. Junek, and R. Martin (2019). Beyond correlation: A path-invariant measure for seismogram similarity, *Seismol. Res. Lett.* **91,** 356–369.

Dokht, R. M. H., H. Kao, R. Visser, and B. Smith (2019). Seismic event and phase detection using time–frequency representation and convolutional neural networks, *Seismol. Res. Lett.* **90,** 481–490.

Fuenzalida, H., C. Dimate, and A. Taboada (1998). Sismotectonica de colombia: Deformacion continental activa y subduccion, *Fisica de la tierra, ISSN 0214-4557, N 10, 1998 (Ejemplar dedicado a: Sismicidad y sismotectonica de Centro y Sudamerica)*, 111–148 (in Spanish).

Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences and gempa GmbH (2008). The SeisComP seismological software package, GFZ Data Services, doi: 10.5880/GFZ.2.4.2020.003.

Gibbons, S. J., and F. Ringdal (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int.* **165,** 149–166.

Gong, J., and W. Fan (2022). Seismicity, fault architecture, and slip mode of the westernmost gofar transform fault, *J. Geophys. Res.* **127,** no. 11, doi: 10.1029/2022JB024918.

Jiang, C., L. Fang, L. Fan, and B. Li (2021). Comparison of the earthquake detection abilities of phasenet and EQTransformer with the yangbi and maduo earthquakes, *Earthq. Science* **34,** no. 5, 425–435.

Kellogg, J. N., G. B. F. Camelio, and H. Mora-Páez (2019). Chapter 4 - cenozoic tectonic evolution of the north andes with constraints from volcanic ages, seismic reflection, and satellite geodesy, in *Andean Tectonics*, B. K. Horton and A. Folguera (Editors), Elsevier, 69–102, doi: 10.1016/B978-0-12-816009-1.00006-X.

Klein, F. W. (2002). User's guide to hypoinverse-2000, a fortran program to solve for earthquake locations and magnitudes, *U.S. Geol. Surv. Open-File Rept. 02-171*

Lahr, J. C. (1999). HYPOELLIPSE: A computer program for determining local earthquake hypocentral parameters, magnitude, and first motion pattern, *U.S. Geol. Surv. Open-File Rept. 79-431*, CiteSeer.

Lapins, S., B. Goitom, J. Kendall, M. Werner, K. Cashman, and J. Hammond (2021). A little data goes a long way: Automating seismic phase arrival picking at nabro volcano with transfer learning, *J. Geophys. Res.* **126,** doi: 10.1029/2021JB021910.

Levander, A. (2016). Caribbean-merida andes experiment, *International Federation of Digital Seismograph Networks* doi: 10.7914/SN/YU_2016.

Lomax, A., J. Virieux, P. Volant, and C. Berge-Thierry (2000). *Probabilistic Earthquake Location in 3D and Layered Models*, Springer Netherlands, Dordrecht, 101–134.

Londoño, J. M., S. Quintero, K. Vallejo, F. Muñoz, and J. Romero (2019). Seismicity of valle medio del Magdalena basin, Colombia, *J. South Am. Earth Sci.* **92,** 565–585.

Lopez, C. M., L. Velasquez, and V. Dionicio (2020). Calibration of local magnitude scale for Colombia, *Bull. Seismol. Soc. Am.* **110,** 1971–1981.

Magrini, F., D. Jozinović, F. Cammarano, A. Michelini, and L. Boschi (2020). Local earthquakes detection: A benchmark dataset of 3-component seismograms built on a global scale, *Artif. Intell. Geosci.* **1,** 1–10.

Martinez, D., and G. Prieto (2023). Tectonic setting of the northwestern andes constrained by a high-resolution earthquake catalog: Block kinematics, *J. South Am. Earth Sci.* **134,** 104761.

Mayorga, E., V. Dionicio, M. Lizarazo, P. Pedraza, E. Poveda, O. Mercado, D. Siervo, L. Aguirre, R. Bolaños, F. Garzón, *et al.* (2021). El sismo de mesetas, meta del 24 de diciembre de 2019 aspectos sismológicos, movimiento fuerte y consideraciones geodésicas, Servicio Geológico Colombiano, Bogotá, available at https://www2.sgc.gov.co/Publicaciones/Sismos%20importantes/Informe%20sismo%20Mesetas%20-%20Meta%2024%20de%20diciembre%202019.pdf (last accessed May 2024) (in Spanish).

McBrearty, I. W., and G. C. Beroza (2023). Earthquake phase association with graph neural networks, *Bull. Seismol. Soc. Am.* **113,** no. 2, 524–547.

McBrearty, I. W., A. A. Delorey, and P. A. Johnson (2019). Pairwise association of seismic arrivals with convolutional neural networks, *Seismol. Res. Lett.* **90,** 503–509.

McBrearty, I. W., J. Gomberg, A. A. Delorey, and P. A. Johnson (2019). Earthquake arrival association with backprojection and graph theory, *Bull. Seismol. Soc. Am.* **109,** 2510–2531.

Miller, M. S., R. Pickle, R. Murdie, H. Yuan, T. I. Allen, K. Gessner, B. L. N. Kennett, and J. Whitney (2023). Southwest Australia Seismic Network (SWAN): Recording earthquakes in Australia's most active seismic zone, *Seismol. Res. Lett.* **94,** no. 2A, 999–1011.

Molina, I., J. Velasquez, J. Rubinstein, A. Garcia, and V. Dionicio (2020). Seismicity induced by massive wastewater injection near puerto gaitán, Colombia, *Geophys. J. Int.* **223,** 777–791.

Mousavi, S. M., and G. C. Beroza (2022). Deep-learning seismology, *Science* **377,** eabm4470, doi: 10.1126/science.abm4470.

Mousavi, S. M., W. L. Ellsworth, W. Zhu, L. Y. Chuang, and G. C. Beroza (2020). Earthquake transformer-an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nat. Commun.* **11,** 1–12.

Mousavi, S. M., Y. Sheng, W. Zhu, and G. C. Beroza (2019). Stanford Earthquake Dataset (STEAD): A global data set of seismic signals for ai, *IEEE Access* **7,** 179,464–179,476.

Mousavi, S. M., W. Zhu, Y. Sheng, and G. C. Beroza (2019). Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection, *Sci. Rep.* **9,** 1–14.

Münchmeyer, J., J. Woollam, A. Rietbrock, F. Tilmann, D. Lange, T. Bornstein, T. Diehl, C. Giunchi, F. Haslinger, D. Jozinović, *et al.* (2022). Which picker fits my data? a quantitative evaluation of deep learning based seismic pickers, *J. Geophys. Res.* **127,** no. 1, doi: 10.1029/2021JB023499.

Ni, Y., A. Hutko, F. Skene, M. Denolle, S. Malone, P. Bodin, R. Hartog, and A. Wright (2023). Curated pacific northwest ai-ready seismic dataset, *Seismica* doi: 10.26443/seismica.v2i1.368.

Niksejel, A., and M. Zhang (2024). OBSTransformer: A deep-learning seismic phase picker for OBS data using automated labelling and transfer learning, *Geophys. J. Int.* **237,** no. 1, 485–505.

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto

Ojeda, A., and J. Havskov (2001). Crustal structure and local seismicity in Colombia, *J. Seismol.* **5,** 575–593.

Pardo, E., C. Garfias, and N. Malpica (2019). Seismic phase picking using convolutional networks, *IEEE Trans. Geosci. Remote Sens.* **57,** 7086–7092.

Park, Y., S. M. Mousavi, W. Zhu, W. L. Ellsworth, and G. C. Beroza (2020). Machine-learning-based analysis of the guy-greenbrier, arkansas earthquakes: A tale of two sequences, *Geophys. Res. Lett.* **47,** no. 6, doi: 10.1029/2020GL087032.

Pavlis, G. L., F. Vernon, D. Harvey, and D. Quinlan (2004). The generalized earthquake-location (genloc) package: An earthquake-location library, *Comp. Geosci.* **30,** nos. 9/10, 1079–1091.

Poveda, E., J. Julià, M. Schimmel, and N. Perez-Garcia (2018). Upper and middle crustal velocity structure of the colombian andes from ambient noise tomography: Investigating subduction-related magmatism in the overriding plate, *J. Geophys. Res.* **123,** no. 2, 1459–1485.

Prieto, G. A., G. C. Beroza, S. A. Barrett, G. A. López, and M. Florez (2012). Earthquake nests as natural laboratories for the study of intermediate-depth earthquake mechanics, *Tectonophysics* **570,** 42–56.

Rengifo, F. A., and A. O. Carriazo (2004). Inversion de amplitudes de registros sísmicos para el cálculo de magnitud ml en Colombia, available at http://bdrsnc.sgc.gov.co/publicaRSNC/inversion_reg_sismicos-Rengifo.pdf (last accessed April 2024) (in Spanish).

Ringler, A., J. Steim, D. Wilson, R. Widmer-Schnidrig, and R. Anthony (2019). Improvements in seismic resolution and current limitations in the global seismographic network, *Geophys. J. Int.* **220,** 508–521.

Ronneberger, O., P. Fischer, and T. Brox (2015). U-Net: Convolutional networks for biomedical image segmentation.

Ross, Z. E., and Y. Ben-Zion (2014). Automatic picking of direct p, s seismic phases and fault zone head waves, *Geophys. J. Int.* **199,** 368–381.

Ross, Z. E., M. Meier, E. Hauksson, and T. H. Heaton (2018). Generalized seismic phase detection with deep learning, *Bull. Seismol. Soc. Am.* **108,** 2894–2901.

Ross, Z. E., Y. Yue, M.-A. Meier, E. Hauksson, and T. H. Heaton (2019). Phaselink: A deep learning approach to seismic phase association, *J. Geophys. Res.* **124,** 856–869.

Saad, O. M., Y. Chen, D. Siervo, F. Zhang, A. Savvaidis, G.-C. D. Huang, N. Igonin, S. Fomel, and Y. Chen (2023). Eqcct: A production-ready earthquake detection and phase-picking method using the compact convolutional transformer, *IEEE Trans. Geosci. Remote Sens.* **61,** 1–15.

Saragiotis, C. D., L. J. Hadjileontiadis, and S. M. Panas (2002). Pai-s/k: A robust automatic seismic p phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.* **40,** 1395–1404.

Schubert, E., J. Sander, M. Ester, H. P. Kriegel, and X. Xu (2017). DBSCAN revisited, revisited: Why and how you should (still) use DBSCAN, *ACM Trans. Database Syst.* **42,** no. 3, 1–21.

Servicio Geológico Colombiano (1993). Red sismologica nacional de colombia (in Spanish).

Sheen, D.-H., and P. A. Friberg (2021). Seismic phase association based on the maximum likelihood method, *Front. Earth Sci.* **9,** 699281.

Sleeman, R., and T. van Eck (1999). Robust automatic p-phase picking: An on-line implementation in the analysis of broadband seismogram recordings, *Phys. Earth Planet. In.* **113,** 265–275.

Sun, M., M. J. Bezada, J. Cornthwaite, G. A. Prieto, F. Niu, and A. Levander (2022). Overlapping slabs: Untangling subduction in nw south america through finite-frequency teleseismic tomography, *Earth Planet. Sci. Lett.* **577,** 117253.

Syracuse, E., M. Maceira, G. Prieto, H. Zhang, and C. Ammon (2016). Multiple plates subducting beneath colombia, as illuminated by seismicity and velocity from the joint inversion of seismic and gravity data, *Earth Planet. Sci. Lett.* **444,** 139–149.

Taboada, A., L. Rivera, H. Fuenzalida, A. Cisternas, H. Philip, H. Bijwaard, J. Olaya, and C. Rivera (2000). Geodynamics of the northern andes: Subductions and intracontinental deformation (Colombia), *Tectonics* **19,** 787–813.

Tarantola, A., and B. Valette (1982). Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.* **20,** no. 2, 219–232.

Vargas, C. A., and P. Mann (2013). Tearing and breaking off of subducted slabs as the result of collision of the panama arc-indenter with northwestern South America, *Bull. Seismol. Soc. Am.* **103,** 2025–2046.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin (2017). Attention is all you need, *Proc. of the 31st International Conf. on Neural Information Processing Systems, NIPS'17,* Red Hook, NY, USA, Curran Associates Inc., 6000–6010.

Wagner, L., J. Jaramillo, L. Ramirez-Hoyos, G. Monsalve, A. Cardona, and T. Becker (2017). Transient slab flattening beneath Colombia, *Geophys. Res. Lett.* **44,** 6616–6623.

Woollam, J., J. Münchmeyer, F. Tilmann, A. Rietbrock, D. Lange, T. Bornstein, T. Diehl, C. Giunchi, F. Haslinger, D. Jozinović, *et al.* (2022). SeisBench-A toolbox for machine learning in seismology, *Seismol. Res. Lett.* **93,** no. 3, 1695–1709.

Woollam, J., A. Rietbrock, A. Bueno, and S. D. Angelis (2019). Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network, *Seismol. Res. Lett.* **90,** 491–502.

Yarce, J., G. Monsalve, T. Becker, A. Cardona, E. Poveda, D. Alvira, and O. Ordóñez-Carmona (2014). Seismological observations in northwestern south america: Evidence for two subduction segments, contrasting crustal thicknesses and upper mantle flow, *Tectonophysics* **637,** 57–67.

Yeck, W. L., J. M. Patton, C. E. Johnson, D. Kragness, H. M. Benz, P. S. Earle, M. R. Guy, and N. B. Ambruz (2019). Glass3: A standalone multiscale seismic detection associator, *Bull. Seismol. Soc. Am.* **109,** 1469–1478.

Zarifi, Z., J. Havskov, and A. Hanyga (2007). An insight into the bucaramanga nest, *Tectonophysics* **443,** 93–105.

Zhu, J., Z. Li, and L. Fang (2023). Ustc-pickers: A unified set of seismic phase pickers transfer learned for China, *Earthq. Sci.* **36,** no. 2, 95–112.

Zhu, W., and G. C. Beroza (2019). Phasenet: A deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.* **216,** 261–273.

Zhu, W., I. W. McBrearty, S. M. Mousavi, W. L. Ellsworth, and G. C. Beroza (2022). Earthquake phase association using a Bayesian Gaussian mixture model, *J. Geophys. Res.* **127,** no. 5, doi: 10.1029/2021JB023249.

Manuscript received 18 January 2024
Published online 16 May 2024

Downloaded from http://pubs.geoscienceworld.org/ssa/srl/article-pdf/doi/10.1785/0220240036/6438491/srl-2024036.1.pdf
by Univ Nacional de Colombia, gprieto