

Ground-Motion Dataset for Shallow Earthquakes in Colombia

Daniel Martinez-Jaramillo^{1,2}, Sreeram-Reddy Kotha³, F. Ramón Zúñiga⁴, Pierre Lacan⁴.

¹Posgrado en Ciencias de la Tierra, Instituto de Geociencias, Universidad Nacional Autónoma de México. Querétaro, 76230, México.

²Now at Servicio Geológico Colombiano. Observatorio Vulcanológico y Sismológico de Pasto, 520002, Nariño, Colombia.

~~³Institut des Sciences de la Terre, Université Grenoble Alpes, Grenoble, 38400, France.~~

³[University Grenoble Alpes, University Savoie Mont Blanc, CNRS, IRD, University Gustave Eiffel, ISTerre, Grenoble, France.](#)

⁴Instituto de Geociencias, Universidad Nacional Autónoma de México. Juriquilla, 76230, Querétaro, México.

~~Correspondence to: Daniel Martinez-Jaramillo (dmartinezj@sgc.gov.co)~~

[Correspondence to: Daniel Martinez-Jaramillo \(danmartinezjar@geociencias.unam.mx\)](mailto:danmartinezjar@geociencias.unam.mx)

Abstract. The tectonics of Colombia is characterised by the interaction of multiple oceanic plates subducting beneath the continent, generating active transpressional deformation and a wide range of subduction-related and crustal earthquakes. We present a flatfile containing ground-motion intensity measurements (IMs) from 667 shallow earthquakes. [Magnitudes were homogenised using S-wave corner-frequency-based estimates where a reliable agency-reported moment magnitude was unavailable.](#) The dataset includes 80867550 three-component acceleration records that were uniformly processed. The computed IMs include peak ground velocity, peak ground acceleration, and 5%-damped spectral accelerations for 31 periods ranging from 0.01 to 8 s. Furthermore, the Fourier Amplitude Spectra for each component is provided in the frequency range from 0.04 to 3050.00 Hz. Epicentral and hypocentral distances are reported for all events, while finite-fault distance metrics are estimated for earthquakes with $M > 5.5$. Site conditions are characterised using V_{s30} and the horizontal-to-vertical spectral ratio (H/V). Hypocentre locations are constrained using the ISC-EHB catalogue and the Integrated Seismological Catalogue of the Colombian Geological Survey. For validation and consistency checks, the dataset was compared to a global and region-adapted Ground Motion Prediction Model. This flatfile constitutes a valuable resource for seismic hazard analysis, ground-motion modelling, risk assessment, earthquake engineering applications, seismic-site response and source physics characterisation in Colombia, surrounding regions, and other comparable tectonic environments.

1 Introduction

The subduction of the Nazca and Caribbean plates, together with the indentation (or possible additional subduction) of the Panamá-Chocó block against the northwestern South American plate, generates moderate to large earthquakes in the region. As a result, Colombia experiences a complex mixture of interface, intraslab,

shallow continental, lithospheric continental, and volcanic-related earthquakes. Several significant earthquakes have strongly affected Colombia, for example:

- 18 May 1875, **M** 6.8, in Cúcuta, Norte de Santander, occurred near the Colombia–Venezuela border, causing approximately 1,000 fatalities. The likely source is the Aguas Calientes Fault System, part of the Boconó Fault System in the Mérida Andes (Rodríguez et al., 2018).
- 31 January 1906, **M** 8.8 in the Colombia–Ecuador border, a megathrust event along the Nazca subduction zone that generated a devastating tsunami and caused around 600 fatalities - one of the largest earthquakes recorded globally.
- 12 December 1979, **M** 8.2 (Tumaco, Nariño): Another major subduction earthquake that caused a tsunami and a death toll of more than 450.
- 31 March 1983, **M** 5.5 (Popayán, Cauca): Caused approximately 300 fatalities; the source is attributed to one of the structures belonging to the Romeral Fault System (RFS) (Marín-Árias et al., 2006).
- 18 October 1992, **M** 7.2, Murindó, Antioquia, caused more than 10 fatalities and widespread environmental effects, including soil liquefaction. This earthquake is associated with the Murindó Fault (Paris et al., 2000; Marín et al., 2009; Mosquera-Machado et al., 2009).
- 25 January 1999, **M** 6.1 (Eje Cafetero): Produced more than 2,000 deaths and has been linked to the Córdoba Fault of the RFS (Paris et al., 2000; Vargas-Jiménez & Monsalve-Jaramillo, 2009).

This brief set of examples highlights the importance of characterising and quantifying the intensity caused by shallow crustal earthquakes in Colombia. Globally, several major initiatives have undertaken the compilation and processing of seismograms to make ground-motion IMs widely available. These include the Next Generation Attenuation (NGA) projects for Western (NGA-West2) and Eastern U.S. (Ancheta et al., 2012; Goulet et al., 2021), the Near-Source Ground-motion Flatfile (Pacor et al., 2018; Sgobba et al., 2021), and regional efforts covering Europe (Akkar et al., 2014; Lanzano, 2020), Italy (Oliveti et al., 2021), France (Buscetti et al. 2024), Belgium (Vanneste et al. 2026), India (Sharma et al., 2025), Japan (Dawood et al., 2016), and Chile (Bastías & Montalva, 2016) among others.

For Northern South America, Arteta et al. (2021) developed a Ground Motion Prediction Model (GMPM) for subduction earthquakes, Arteta et al. (2023) developed a GMPM for shallow crustal earthquakes with 709 records of 56 earthquakes, also Pájaro et al. (2024) developed a GMPM for the Bucaramanga Nest.

The accessibility of IMs for a wide [**M**, **R**] ranges is essential for the development of GMPMs, probabilistic seismic hazard assessment (PSHA), engineering seismology, and earthquake engineering. The development of GMPMs, particularly the partially non-ergodic kind, relies heavily on the quantity and quality of ground-motion IMs, and associated event, path, and site metadata. Such datasets are often referred to as *flatfiles*. Here, we present a flatfile generated from consistently processed ground-motion recordings from the shallow earthquakes across Colombia and tested with a global and a region-adapted GMPM for consistency check purposes.

2 Compilation of the Data

The flatfile is arranged as a table that contains verified and reliable metadata and IMs of processed waveforms recorded by the Colombian Geological Survey (SGC, per its Spanish acronym). SGC provided, upon request,

more than 10,000 quality-checked acceleration time series. Earthquake criteria selection to the flatfile from SGC database were:

- Events in the latitude range 0° to 14°N ;
- ~~Events in the longitude range 69°W , 82°W ;~~
- Events in the longitude range 69°W to 82°W ;
- Events with depth estimation < 50 km;
- Any magnitude > 4 ;
- Records with $R_{\text{epi}} < 400$ km.

After applying these earthquake selection criteria, we obtained more than 700 earthquakes. After filtering out volcanic events and spikes in the signals, we ultimately retained 667 events and 7550 records.

SGC has a national network consisting of 506 seismological and accelerometric stations operating since 1993 installed around Colombia; some of them were removed or operated in temporary networks. The FDSN code for this network is CM. In addition, there are local networks within the country to specifically monitor volcanoes, mining districts, and oil and gas fields; some of these stations were shared with SGC and contributed to this flatfile. The sampling frequency of accelerometric stations is predominantly 200 Hz, with a few stations (3) at 100 Hz.

2.1 Flatfile structure

The fields of flatfile are grouped as: (1) Event metadata; (2) Site metadata; (3) Metrics of event-to-site distances; (4) Intensity measures.

The 667 events are defined by their location (Latitude, Longitude, depth), origin time (date), and moment magnitude (**M**). Magnitudes were taken from the Integrated Seismic Catalog for Colombia (Montejo et al., 2023). This catalog was reviewed for locations and homogenised in **M**. For the events not included in the dataset of Montejo et al. (2023), we followed the same criteria, prioritizing locations made by ISC-EHB (Weston et al., 2018) and SGC. **M** was compiled from the Global CMT catalog (Ekström et al., 2012), USGS-NEIC, ISC, and SGC. Focal mechanisms were taken from GCMT and SGC. Geometry of the rupture plane is a simple description from the nodal plane preferred from the focal mechanism selected by the authors of this dataset from the geological context, with strike, dip, and rake for slip kinematics.

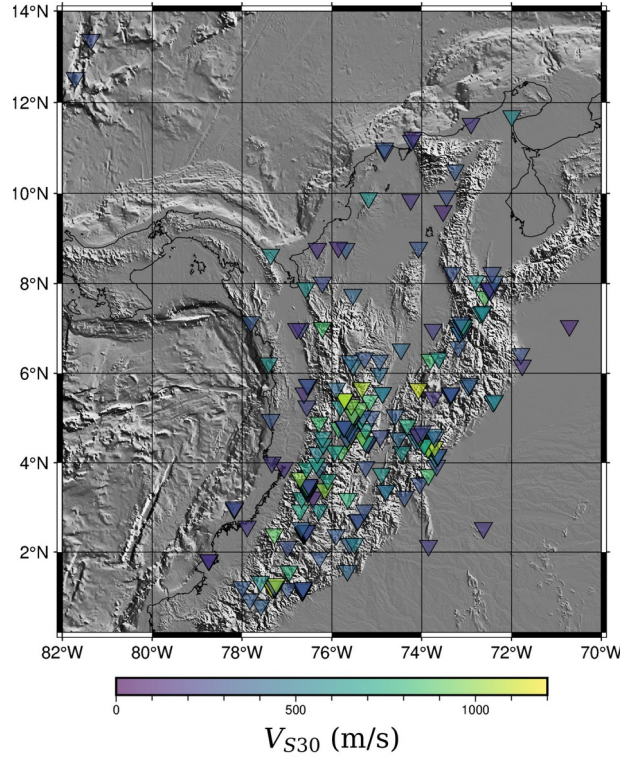


Figure 1. Sites used in this work and V_{S30} .

Relationship	σ
$M = 0.954 m_b + 0.42$ for $3.6 < m_b < 5.7$	0.28
$M = 1.433 m_b - 2.35$ for $5.7 < m_b < 7.7$	0.36
$M = 0.958 M_L + 0.1$ for $2.9 < M_L < 6.1$	0.50

~~GEM (2018):~~

~~Table 1. Equations used to homogenize moment magnitude (M) from SGC & GEM (2018).~~

~~Moment magnitude (M)~~ M is the preferred magnitude in GMPMs and PSHA. Magnitude range is 3.5 to 7.2. When M was not available, m_b or M_L values were taken depending on availability. To homogenize these magnitudes into M , we estimated corner frequencies (f_c) for each station-event recording. A single-corner ω^{-2} Brune (1970) source model was then fit to each corrected spectrum via a coarse grid search followed by non-linear least-squares refinement in log-log space, minimizing the residual between the observed and predicted spectral shape. From the resulting f_c , the source radius r was computed following Brune (1970), $r = 2.34\beta/(2\pi f_c)$, using a shear-wave velocity $\beta = 3.5$ km/s representative of active continental crust. The seismic moment M_0 was then obtained from the Eshelby circular-crack relation, $M_0 = (16/7)(\Delta\sigma)r^3$, under an assumed constant stress drop ($\Delta\sigma$), and converted to M following Hanks and Kanamori (1979), $M = (2/3)\log_{10}(M_0) - 10.7$. Values of $\Delta\sigma = 1, 3$, and 5 MPa were tested and compared against independently reported M (GCMT and other M solutions) to select the value that best reproduced the catalog magnitudes. Ultimately, a stress drop of 5

MPa was chosen as it yielded the best agreement. This calculated magnitude was used when a reliable agency-reported M was unavailable.

The 227 sites featured in this dataset have been installed, maintained, and administered by SGC (Figure 1). Of these, 154 sites are also present in the Mercado et al. (2024) database for north-western South America, where the predominant site period (T_n) and the H/V ratios in the Fourier (HVFSR) and pseudo-spectral acceleration (HVRSR) domains were derived from seismograms, they also reported the amplitude of the HVSR at the peak corresponding to T_n (P^*). Within this subset, 28 sites additionally have in-situ, microtremor-based measured V_{s30} , available from Mercado et al. (2024). For the remaining sites without in-situ measurements, we report an **inferred** V_{s30} taken from the topographic-slope-based V_{s30} map for Colombia (Eraso and Montejo, 2019).

Epicentral (R_{epi}) and hypocentral (R_{hyp}) distances and Azimuth are reported for the entire dataset; ~~whereas extended rupture-based distances (R_{rup} , R_{JB} , R_x , R_{y0}) were computed only for earthquakes with $M > 5.5$.~~ Additionally, finite fault distances were computed from the geometry of the rupture plane derived from the preferred nodal plane for events with $M > 5.5$. Rupture dimensions were computed following Strasser et al. (2010) for subduction interphase-related earthquakes and following Wells and Coppersmith (1994) for continental events. The hypocenter was placed at the midpoint of the fault plane, but for earthquakes with hypocentral depths shallower than 10 km, the midpoint of the rupture plane was placed at 12.5 km depth as the average thickness of the seismogenic depth. Joyner-Boore distance (R_{JB}) was computed as the closest horizontal distance from the site to the surface projection of the rupture plane. R_x represents the horizontal distance from the top edge of the rupture measured perpendicular to the fault strike, with positive and negative values indicating sites on the hanging-wall ($R_x > 0$) and footwall ($R_x < 0$) sides, respectively. R_{y0} is the horizontal distance measured parallel to the strike from the fault midpoint.

For the **waveform** processing, we followed the steps presented by Paolucci et al. (2010) and Lanzano et al. (2019):

1. Baseline correction;
2. Cosine taper;
3. Application of a second order acausal bandpass Butterworth filter to the acceleration time-series;
4. Double integration to obtain displacement time series;
5. Linear detrending of the displacement;
6. Double differentiation to get the corrected acceleration.

For the bandpass filter limits, we used median values from Lanzano et al. (2019). As they pointed out, the low-pass filter frequency (f_{lp}) is magnitude independent, we took the mean value of 32.26 Hz, while for the high pass (f_{hp}) filter we used a magnitude-dependent function as:

$$\begin{aligned}
 f_{hp} &= 0.15 \text{ Hz for } M < 4 \\
 f_{hp} &= -0.05(M) + 0.35 \text{ Hz } 6 \geq M \geq 4 \\
 f_{hp} &= 0.05 \text{ Hz for } M > 6.
 \end{aligned} \tag{1}$$

We test the sensitivity of the parameter RotD50 peak ground acceleration (PGA) with the fixed f_{lp} of the bandpass filter at 32.26 Hz, taking the records and changing f_{lp} in a range of 27 to 37 Hz. The sensitivity analysis shows that most records are weakly affected by the choice of f_{lp} , as the average variation is 2.31% and the 90th quantile is 7.0%. However, 6.4% of the recordings exhibit variation larger than 10% in RotD50 PGA. These records are concentrated around M 4.5 and $R_{epi} > 250$ km (Figure A1). These differences are likely related to high-frequency signal content or noise close to the selected filter corner frequency, which could affect the peak acceleration amplitude.

Fourier Amplitude Spectrum (FAS) from the acceleration records in the three components are computed in the frequency range 0.04 to 3050.00 Hz, applying Konno and Ohmachi (1998) function with smoothing (Lanzano et al. 2019). An example of FAS computed for an M 6.1 earthquake recorded at 13 km is shown in Figure 2.

Usable frequencies are limited to those between the high-pass and low-pass by a safety factor of 1.25 to ensure that the filters do not have a significant effect on the response spectral values (Abrahamson and Silva, 1997). Hence, the usable bandwidth decreases at low frequencies for small magnitudes, and the upper limit for our dataset is 25.808 Hz. Lowest usable frequencies (LUF) are:

$$\begin{aligned} LUF &= 0.1875 \text{ Hz for } M < 4, \\ LUF &= -0.0625(M) + 0.4375 \text{ Hz for } 6 \geq M \geq 4, \\ LUF &= 0.0625 \text{ Hz for } M > 6. \end{aligned} \quad (2)$$

To facilitate direct use of the usable frequency range by the community, we added two columns to the FAS dataset - LUF Hz and UF Hz - reporting the lowest and highest usable frequency for each record, respectively.

A compilation of Effective Amplitude Spectrum (EAS) (Kottke et al. 2018), defined in equation 3, binned by magnitude, is shown in Figure 3.

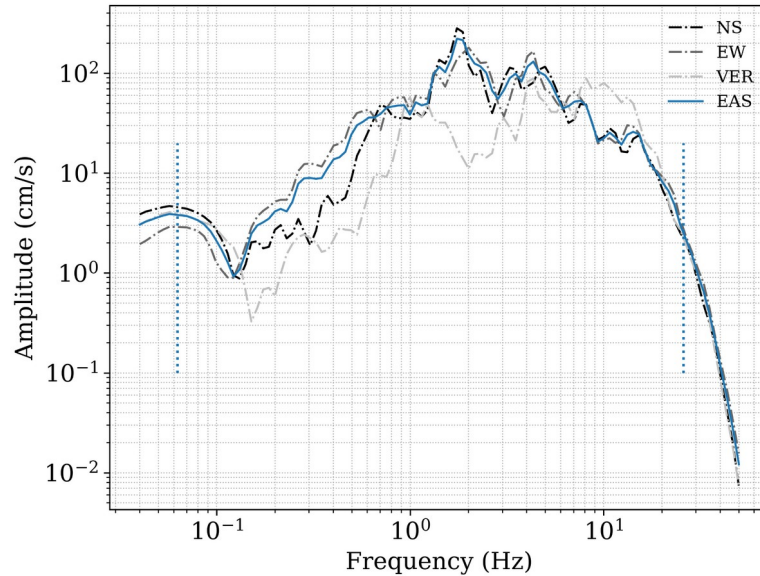


Figure 2. Fourier amplitude spectrum recorded by each component at ARMEC. Earthquake occurred on January 25, 1999 with $M = 6.1$. Dotted vertical blue lines represent bandwidth limits of the usable spectra for this earthquake.

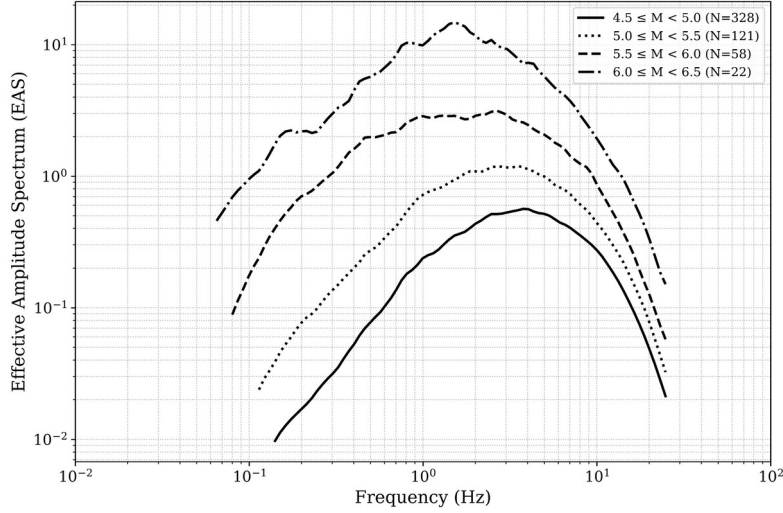


Figure 3. Fourier amplitude spectrum geometric mean of the horizontal components (EAS) is binned by magnitudes for records with $R_{epi} < 100$ km. The limits of usable frequencies in function of magnitude can be appreciated.

The dataset includes IMs that are independent of sensor orientation, calculated using the median and maximum spectral ordinates over all non-redundant horizontal orientations, RotD50 and RotD100 (Boore, 2010), respectively. All spectral values were computed from the corrected accelerograms using standard response-spectrum analysis. Furthermore, the EAS is computed from the two horizontal components of FAS. The following IMs are provided:

- ~~RotD50 Peak Ground Acceleration (PGA) from the horizontal components.~~
- [RotD50 PGA from the horizontal components.](#)
- PGA, PGV and PGD for the three components.
- ~~-5%-damped spectral accelerations for each of the north-south, east-west, and vertical components.~~
- RotD50 spectral accelerations for 5% damping and oscillator periods ranging from 0.010 s to 8.0 s.
- [Amplitudes of the Fourier Spectrum in the range 0.04 to 350.00 Hz for the three components.](#)
- EAS defined as

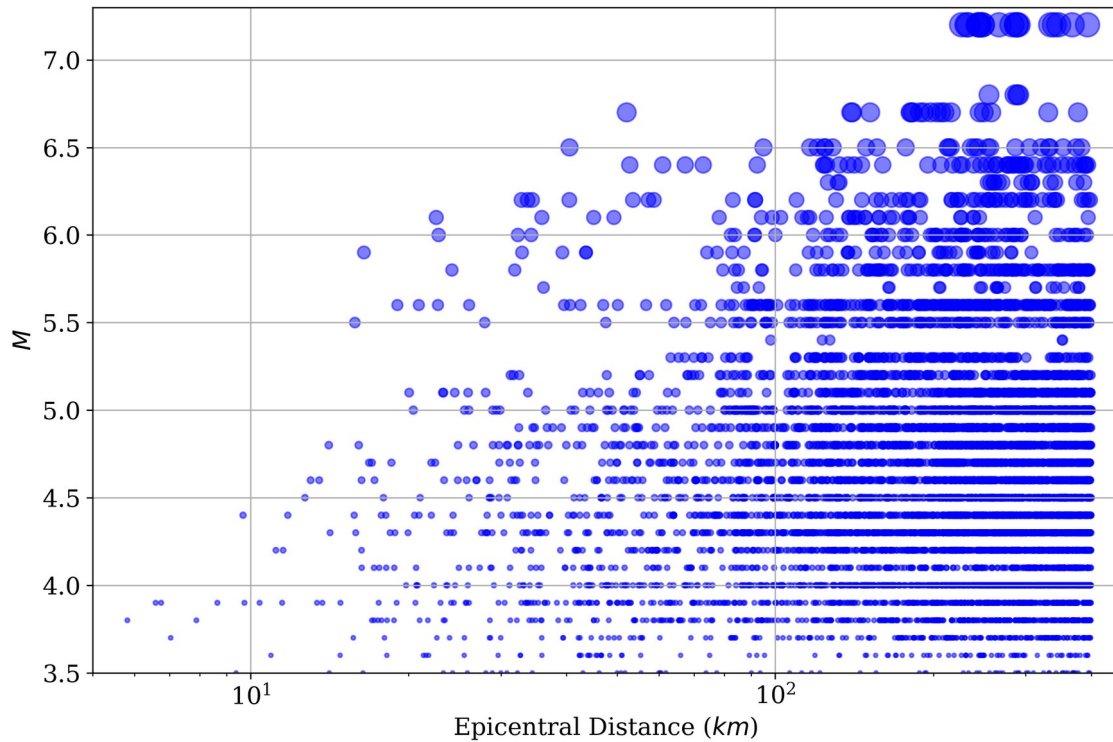
$$EAS(f) = \sqrt{\frac{1}{2} [FAS_{H1}(f)^2 + FAS_{H2}(f)^2]} \quad (3)$$

where H_1 and H_2 are the given horizontal components at specific frequency f of the FAS.

3 Data Statistics

The $[M, R_{epi}]$ distribution of ~~7550~~^{7,992} records from 667 earthquakes in the catalogue is shown in Figure 4. Epicenters of 488 earthquakes are located in the continental territory, whereas 179 are located offshore of the Colombian coastline. The classification between tectonic environments, such as subduction interface and intraslab among others is left for user determination depending on the slab models and classification methods used.

227



228

Figure 4. Distribution of **M** and distance of the 7550 records.

229

230

231

232

233

Data are sampled in the magnitude range 3.5 - 7.2 and for epicentral distances up to 400 km. There is also a significant number of records related to strong events with a magnitude greater than 5.5, corresponding to 9.1% of the records (716) from 42 earthquakes. In addition, there are 15 events with magnitudes larger than 6.0, such as the **M** 6.1 1999 Eje Cafetero earthquake, **M** 7.2 2004 Pizarro earthquake on the Pacific coast, and **M** 6.1 2023 San Juanito earthquake in eastern Colombia.

234

235

236

237

238

239

240

241

~~Figure 5 shows the number of records by R_{epi} , **M**, depth, and V_{s30} in the dataset. Most of the data correspond to distances larger than 100 km (about 85%); 5.3% of the records correspond to distances shorter than 50 km. The distribution of recordings in terms of depth intervals shows that most of the data have focal depths shallower than 20 km, corresponding to about 679% of the total records in the flatfile, thus indicating a predominance of shallow crustal events in the dataset. Indeed, about 2610868 records (331%) correspond to $20 < \text{depth (km)} < 55$. Although the initial catalogue selection required a raw, non-homogenised magnitude greater than 4, recalculating magnitude to a homogeneous **M** widened the final magnitude range downward, with 162 of the 667 events (24%) falling below **M** 4.~~

242

243

244

245

246

247

248

249

Figure 5 shows the number of records by R_{epi} , **M**, depth, and V_{s30} in the dataset. Most of the data correspond to distances larger than 100 km (about 85%); 5.3% of the records correspond to distances shorter than 50 km. The distribution of recordings by of depth shows that most of the data have focal depths shallower than 20 km, corresponding to about 679% of the total records in the flatfile, indicating a predominance of shallow crustal events in the dataset. About 2610 records (33%) correspond to $20 < \text{depth (km)} < 55$. Although the initial catalogue selection required a raw, non-homogenised magnitude greater than 4, recalculating magnitude to a homogeneous **M** widened the final magnitude range downward, with 162 of the 667 events (24%) falling below **M** 4.

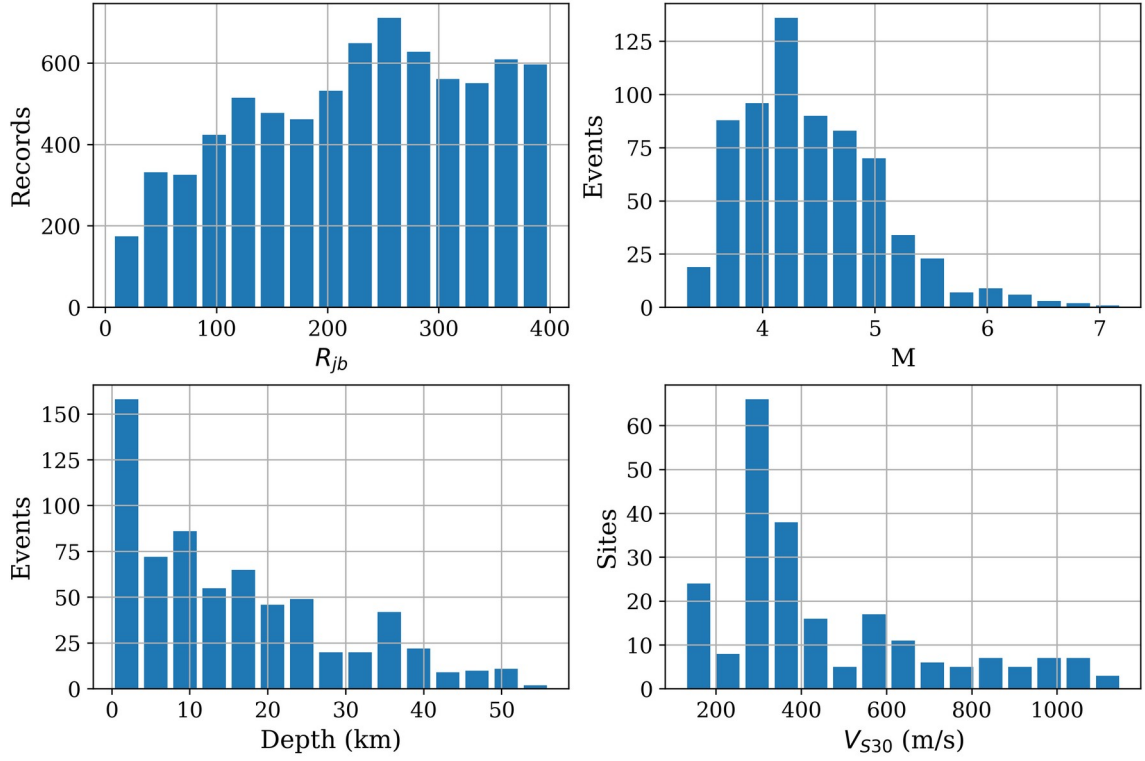


Figure 5. Histograms of the records used in the flatfile showing epicentral distance, V_{s30} , magnitude and focal depth of the earthquakes, records and sites used in the dataset.

3.1 Consistency check: Residual analysis

For validation purposes (Bindi et al. 2019), we predicted PGA, SA (RotD50), and EAS IMs of the continental earthquakes in our dataset (488509 events) – to compute residuals (equation 4) – using the input parameters in the GMPM.

The GMPM used here are the global development of Abrahamson et al. (2014) – made with the NGA-West 2 dataset – and the one of Arteta et al. (2023), which is a regional development for Northern South America of Abrahamson et al. (2014).

The comparison between observed and predicted IMs provides a consistency check for the data quality and processing methodology as shown in Figure 6, where the total residual (δ_{total}) is plotted against R_{epi} and $\mathbf{M}$, without significant biases.

The δ_{total} can be decomposed into between-event (δB_e), between-site ($\delta S2S_s$), and leftover residuals (ϵ) (Al Atik et al., 2010) as:

$$y_{es}(T) = u_{es}(M, R_{es}, V_{s30}, T) + \delta B_e + \delta S2S_s + \epsilon \quad (4)$$

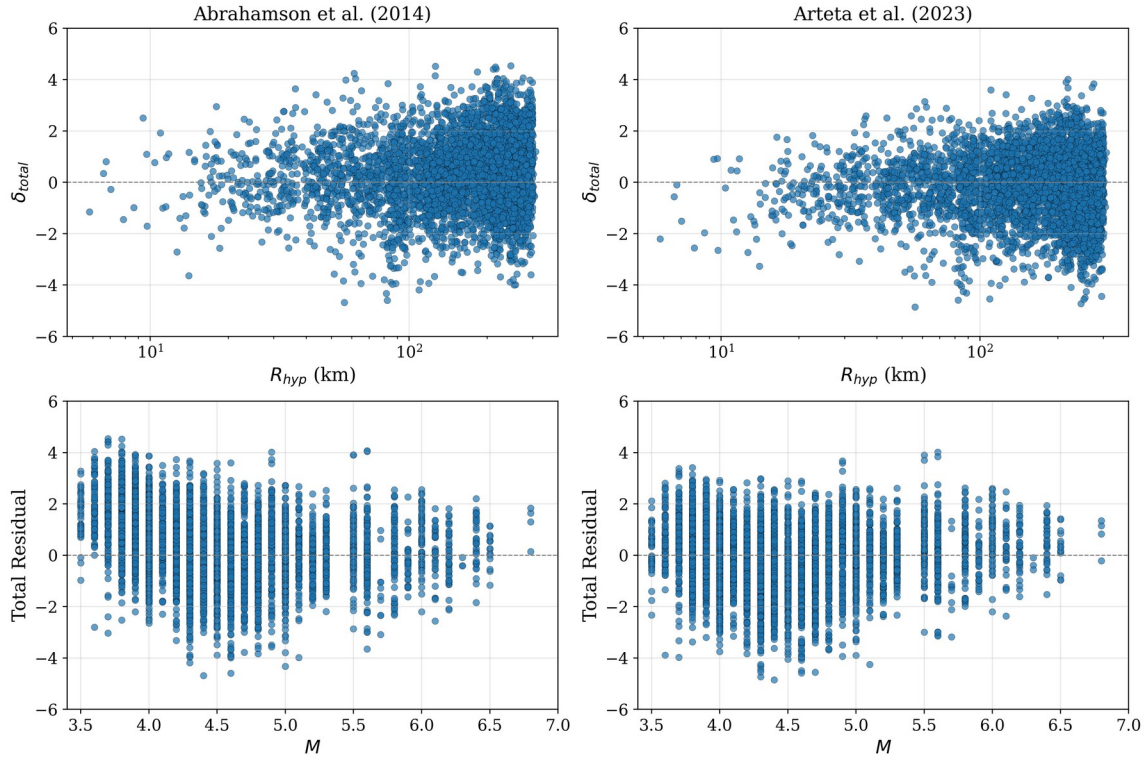
where y is the natural logarithm of the observed IMs from an individual earthquake, e , with magnitude $\mathbf{M}$ observed at site, s , with V_{s30} located at a distance R_{es} . u_{es} is the predicted median IMs from the GMPM.

The total standard deviation of the GMPM (σ) can be written as:

$$\sigma = \sqrt{\tau^2 + \phi_{s2s}^2 + \phi_0^2} \quad (5)$$

where τ , ϕ_{s2s} , ϕ_0 are the standard deviation of δB_e , $\delta S2S_s$ and ϵ respectively.

272 Following residual decomposition of equation (4), we get δB_e and $\delta S2S_s$ to show their distribution,
 273 results are shown in Figures 7 and 8, respectively, and compiled in Table 2.



275 **Figure 6.** Scatterplots of the total residual of the RotD50 PGA of the continental earthquakes in the dataset vs.
 276 R_{rup} and M . The left panel figures are computed from Abrahamson et al. (2014) and right panel figures are
 277 computed from Arteta et al. (2023).

GMPM	Abrahamson-et al. (2014) (RotD50-PGA)	Arteta-et al. (2023) (RotD50-PGA)	Bayless and Abrahamson (2018) (EAS-5 Hz)
Coverage	Global for SA	Regional for Northern South America for SA	Global for FAS
$\delta_{total-mean}$	0.162	-0.318	0.541
$\delta_{total-std}$	1.316	1.290	1.349
$ \delta_{total} < 1 * std$	69.0%	68.3%	69.1%
$ \delta_{total} < 2 * std$	95.6%	95.3%	95.6%
$ \delta_{total} < 3 * std$	99.7%	99.6%	99.7%
ϵ	0.91	0.74 (0.43)	0.89 (0.38)

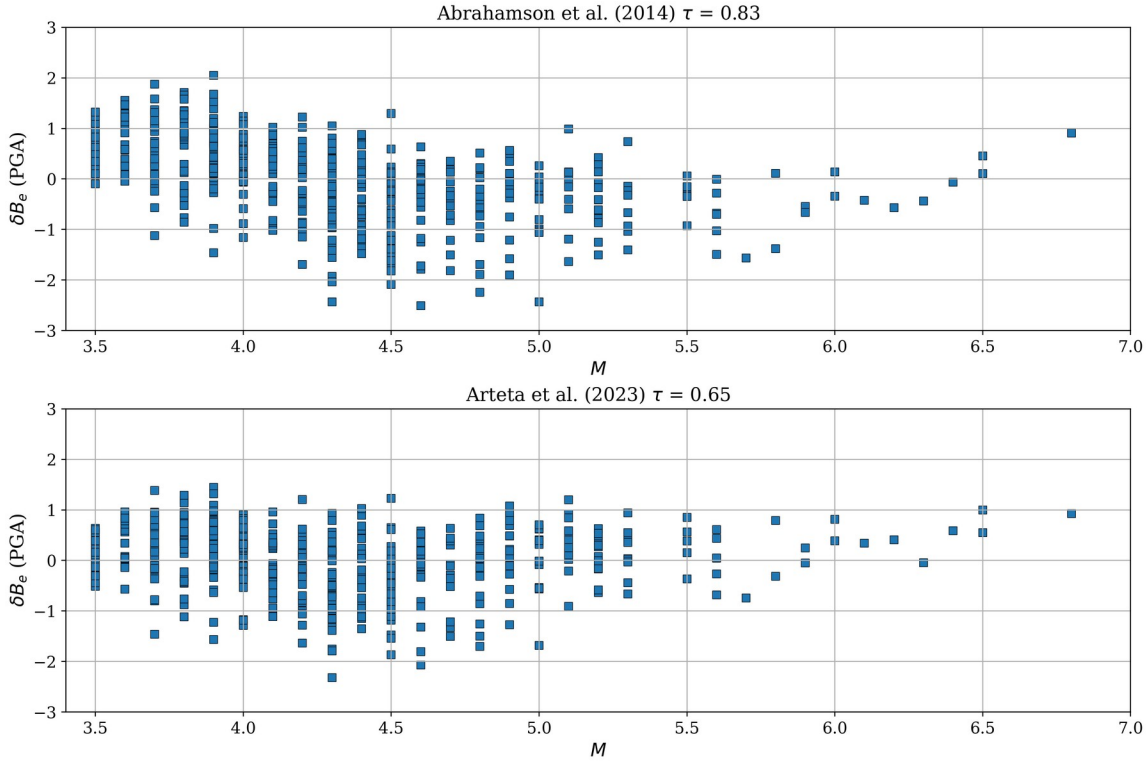
ϕ_{s2s}	0.78	0.79	0.84 (0.40)
ϕ_{θ}	0.72	0.71	0.69
σ	1.40	1.29	1.40

Table 2. Basic statistics and standard deviation of the residuals of the dataset from GMPM of Abrahamson et al. (2014), Arteta et al. (2023), and Bayless and Abrahamson (2018). In parentheses is shown the standard deviation from the publication in the original paper.

<u>GMPM</u>	<u>Abrahamson et al. (2014) (RotD50 PGA)</u>	<u>Arteta et al. (2023) (RotD50 PGA)</u>	<u>Bayless and Abrahamson (2019) (EAS 5 Hz)</u>
<u>Coverage</u>	<u>Global for SA</u>	<u>Regional for Northern South America for SA</u>	<u>Global for FAS</u>
$\delta_{total\ mean}$	<u>0.162</u>	<u>-0.318</u>	<u>0.541</u>
$\delta_{total\ std}$	<u>1.316</u>	<u>1.290</u>	<u>1.349</u>
$ \delta_{total} < 1 * std$	<u>69.0 %</u>	<u>68.3 %</u>	<u>69.1%</u>
$ \delta_{total} < 2 * std$	<u>95.6 %</u>	<u>95.3 %</u>	<u>95.6%</u>
$ \delta_{total} < 3 * std$	<u>99.7 %</u>	<u>99.6%</u>	<u>99.7%</u>
τ	<u>0.91</u>	<u>0.74 (0.43)</u>	<u>0.89 (0.38)</u>
ϕ_{s2s}	<u>0.78</u>	<u>0.79</u>	<u>0.84 (0.40)</u>
ϕ_{θ}	<u>0.72</u>	<u>0.71</u>	<u>0.69</u>
σ	<u>1.40</u>	<u>1.29</u>	<u>1.40</u>

281 [Table 2. Basic statistics and standard deviation of the residuals of the dataset from GMPM of Abrahamson et al.](#)
282 [\(2014\), Arteta et al. \(2023\), and Bayless and Abrahamson \(2019\). In parentheses is shown the standard](#)
283 [deviation from the publication in the original paper.](#)

284



286 **Figure 7.** Between event variability (δB_e) of the continental earthquakes in the dataset and its standar deviation
287 (τ). δB_e in the upper panel is computed with Abrahamson et al. (2014) while in the lower panel is computed
288 with Arteta et al. (2023).

289

290 The trends of δB_e showed a significant scatter for $M < 4.5$, which causes high τ ; this could be related
291 to the empirical relationships used in the conversion of M_L , M_b to M . One way to corroborate this could be the
292 computation of M from the FAS values presented here, from the seismic moment taken from corner frequencies.
293 Another approach could be to update the GMPM in a Bayesian framework with a larger range of small to
294 medium-sized M .

295

296 Trends of δB_e observed in Figure 7 support the consistency of this dataset, where both GMPMs show
297 low variability. As expected, the regional GMPM (Arteta et al., 2023) shows lower overall variability than the
298 global one (Abrahamson et al., 2014). Within the M [4.2, 4.7] range, around M 4.5, δB_e in Arteta et al. (2023)
299 shows a negative trend, suggesting that further magnitude calibration at these levels could help reconcile this
discrepancy.

300

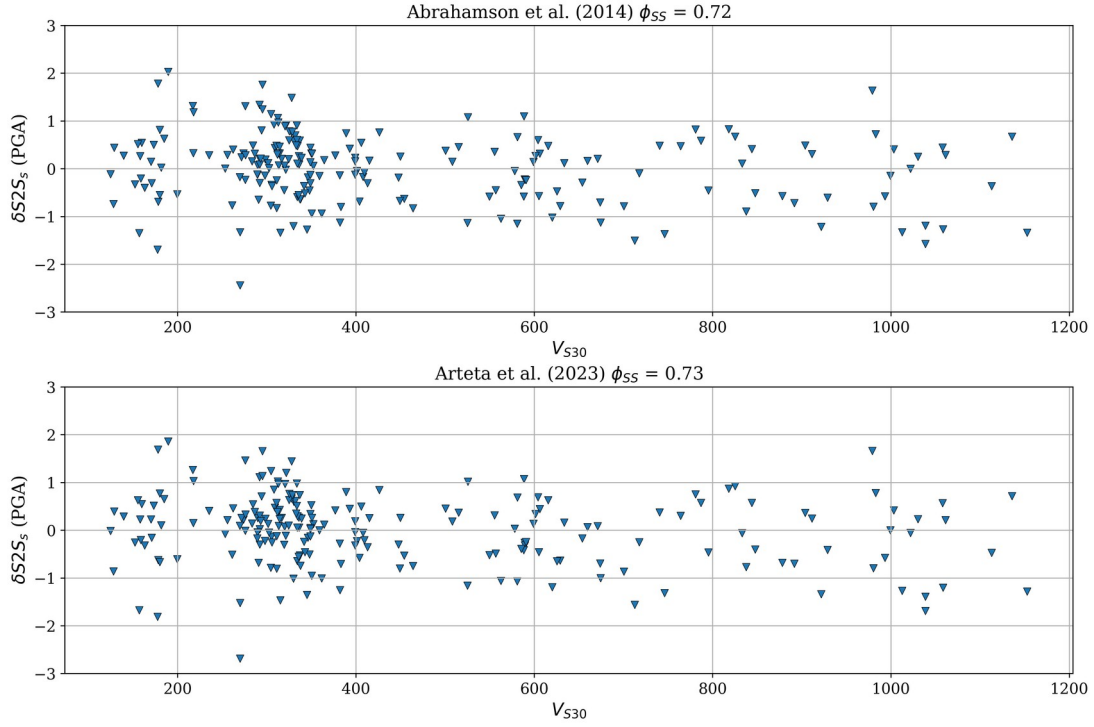


Figure 8. Site-to-site residuals ($\delta S2S_s$) and their standard deviation (ϕ_{s2s}). $\delta S2S_s$ in the upper panel is computed with respect to Abrahamson et al. (2014), while in the lower panel it is computed with respect to Arteta et al. (2023).

303

Given that only 28 of the 227 sites in this dataset have in-situ, microtremor-based measured V_{s30} (Mercado et al., 2024), the remaining sites rely on inferred V_{s30} from a topographic-slope-based proxy (Eraso and Montejo, 2019). We argue that this proxy-based approach could introduce additional epistemic uncertainty into the site parameter, potentially increasing the ϕ_{s2s} reported in this study. Some authors, such as Mercado et al. (2024), have proposed site classification schemes based on H/V spectral ratio characteristics (HVRSR and HVFSR) as a robust alternative/complement to V_{s30} -based proxies for constraining site response in future applications of this dataset.

For Fourier IMs, we follow the same procedure with the EAS (5 Hz), comparing it with the GMPM of Bayless and Abrahamson (20198). Results are shown in Figure 9. Statistical values computed are shown in Figure 9 and Table 2.

We check the mean, standard deviation, and Gaussian-like shape of ϵ , and the Shapiro-Wilk test for normality. Residuals are Gaussian-like distributed for this dataset with respect to both GMPMs, the global and the regional developed one (Table 2 and Figures A2 and A3). The same result is determined for the Fourier IMs EAS (Table 2 and Figure A4).

319

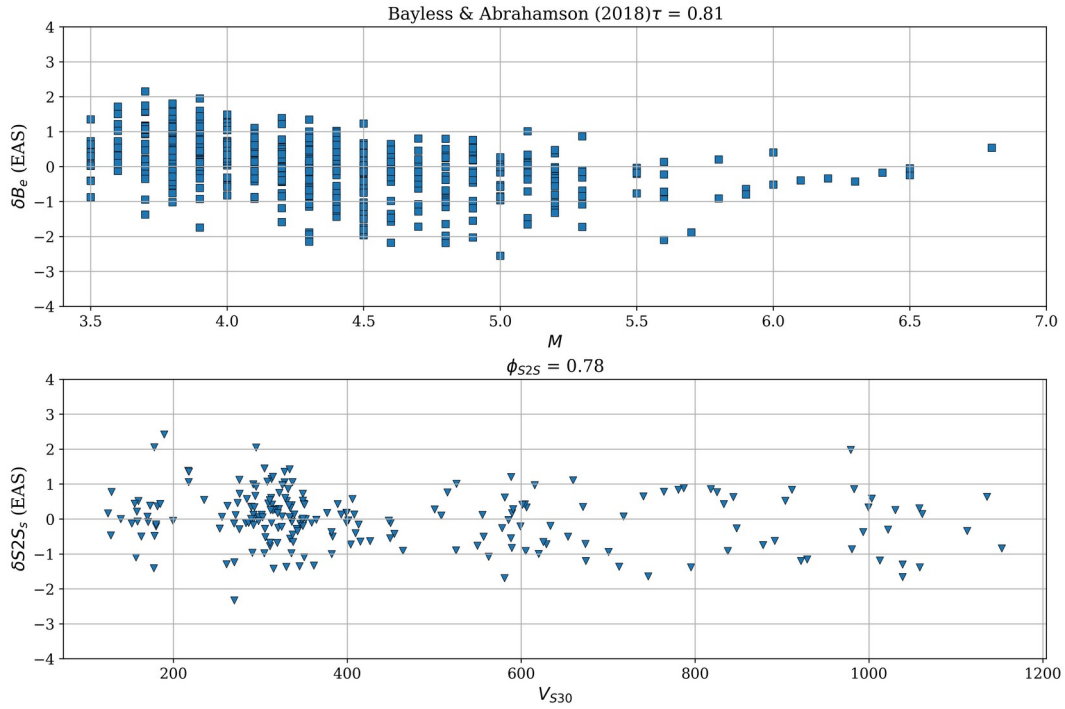
320

321

322

323

324



325 **Figure 9.** δB_e and $\delta S2S_s$ from the EAS (5 Hz) presented in the flatfile and the Bayless and Abrahamson (2019)
 326 GMPM. Upper panel: Between event variability (δB_e) of the continental earthquakes and its standard deviation
 327 (τ). Lower panel: Site-to-site residuals ($\delta S2S_s$) and its standard deviation (ϕ_{S2S}).

328 4 Conclusions

329 A ground motion flatfile is presented after homogeneous preprocessing of the acceleration records following
 330 Pacor et al. (2010) and Lanzano et al. (2019) of the 7550 three-components records of 667 shallow earthquakes
 331 in Colombia, a region with several devastating earthquakes recorded in pre-instrumental and instrumental times.
 332 This flatfile is highly useful for characterising and modeling the ground motion associated with shallow tectonic
 333 structures in the northwestern Andes, as well as to be used in any seismic hazards assessment in Northern South
 334 America or in any comparable tectonic region around the world.

335 A compilation of the site parameters, such as V_{S30} , H/V , and T , was carried out to constrain the
 336 characterisation of the sites and their seismic response.

337 The dataset was validated through residual analysis with a modern and widely used GMPM of
 338 Abrahamson et al. (2014), furthermore, with a region adaptation for Northern South America of Arteta et al.
 339 (2023). For FAS data, we used the Bayless and Abrahamson (2019) GMPM for validation. We showed the
 340 Gaussian-like distribution of the dataset with the aforementioned GMPM.

341 The Fourier Amplitude Spectrum presented here is a powerful tool that can be used to study source
 342 parameters, such as corner frequency, seismic moment, and stress drop. ~~In that way, moment magnitude could
 343 be computed for the whole dataset instead of the empirical relationship.~~

344

345

346

347

Appendix

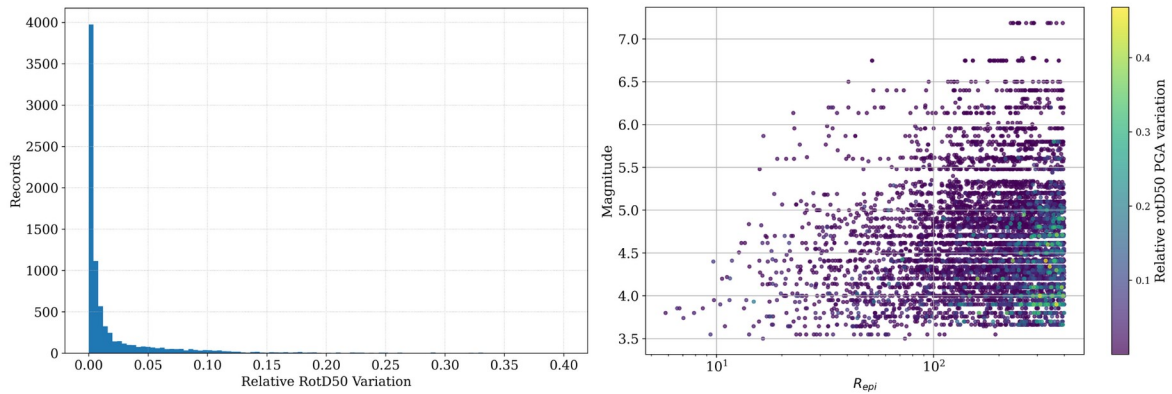


Figure A1. Sensitivity test of the low pass frequency for the bandpass filter used in the preprocessing. L_p was tested in the range 27 to 37 Hz.

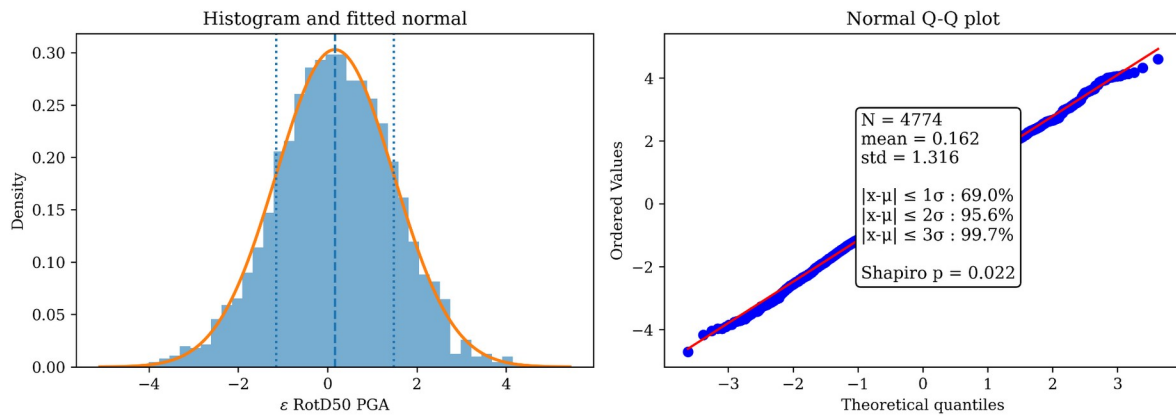


Figure A2. Basic normal analysis of the total residuals of RotD50 PGA of the dataset from the GMPM of Abrahamson et al (2014).

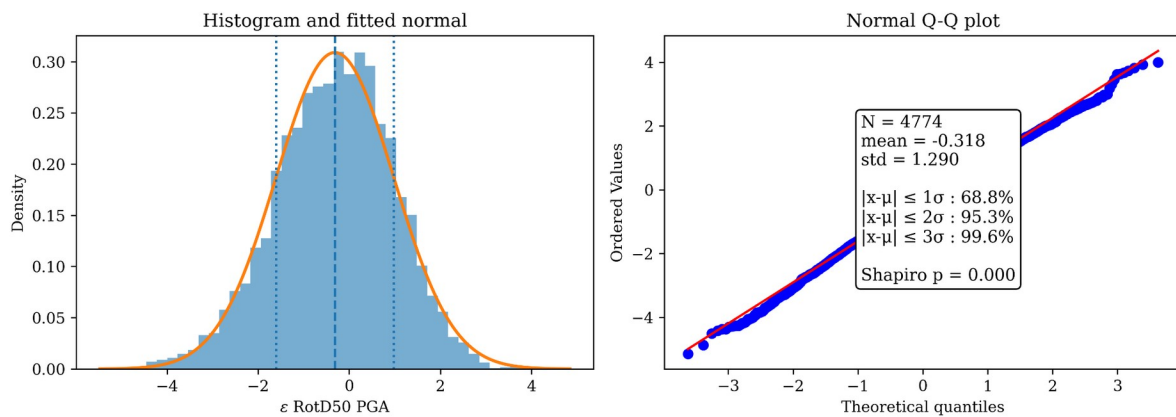


Figure A3. Basic normal analysis of the total residuals of RotD50 PGA of the dataset from the GMPM of Arteta et al (2023).

Figure A3. Basic normal analysis of the total residuals of RotD50 PGA of the dataset from the GMPM of Arteta et al (2023).

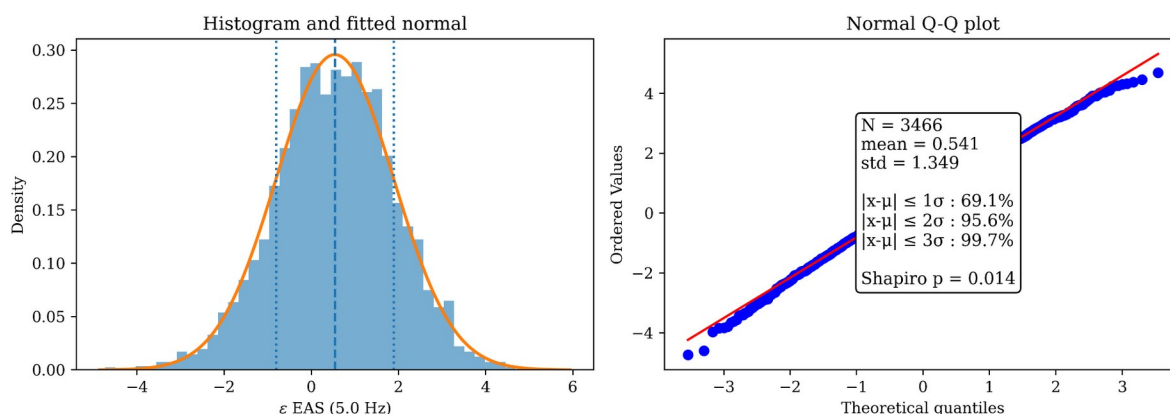


Figure A4. Basic normal analysis of the total residuals of EAS (5 Hz) of the dataset from the GPM of Bayless and Abrahamson et al (2018).

Figure A4. Basic normal analysis of the total residuals of EAS (5 Hz) of the dataset from the GPM of Bayless and Abrahamson et al (2019).

Data availability

Data described in this manuscript can be accessed at the repository under data DOI (10.5281/zenodo.19056074, Martinez-Jaramillo et al. 2026).

Acceleration records are available on the SGC website <https://catalogo-aceleraciones.sgc.gov.co/>, or via FDSN: <https://www.fdsn.org/networks/detail/CM/>.

Author contribution

D Martinez-Jaramillo: Conceptualization, formal analysis, software, validation, writing original draft.
S.R Kotha: Conceptualization, formal analysis, software, validation, writing original draft.
F.R. Zúñiga: Validation, funding acquisition, writing review and editing.
P. Lacan: Validation, funding acquisition, writing review and editing.

Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments: The acceleration time series were given by SGC under request, but also available at: <https://catalogo-aceleraciones.sgc.gov.co/>.

Financial support: Daniel Martinez-Jaramillo is doctoral student of the Earth Sciences Graduate School, Universidad Nacional Autónoma de México and has received a SECIHTI scholarship. Partial support was received from the France-Mexico collaborative project SEP-CONACYT-ANUIES-ECOS N°321193.

References:

- Abrahamson, N. A. and Silva, W. J.: Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismol. Res. Lett.*, 68, 94–127, 1997.
- Abrahamson, N. A., Silva, W. J., and Kamai, R.: Summary of the ASK14 ground motion relation for active crustal regions, *Earthquake Spectra*, 30, 1025–1055, 2014.

- 389 Akkar, S., Sandıkkaya, M. A., Şenyurt, M., Azari Sisi, A., Ay, B. Ö., Traversa, P., and Godey, S.: Reference
 390 database for seismic ground-motion in Europe (RESORCE), *Bull. Earthquake Eng.*, 12, 311–339, 2014.
- 391 [Al Atik, L., Abrahamson, N., Bommer, J. J., Scherbaum, F., Cotton, F., and Kuehn, N.: The variability of](#)
 392 [ground-motion prediction models and its components, *Seismol. Res. Lett.*, 81, 794–801, 2010.](#)
- 393 Ancheta, T., Bozorgnia, Y., Darragh, R., Silva, W. J., Chiou, B., Stewart, J. P., and Atkinson, G. M.: PEER
 394 NGA-West2 database: A database of ground motions recorded in shallow crustal earthquakes in active tectonic
 395 regions, *Proceedings, 15th world conference on earthquake engineering*, 24–28, 2012.
- 396 Arteta, C. A., Pajaro, C. A., Mercado, V., Montejó, J., Arcila, M., and Abrahamson, N. A.: Ground-motion
 397 model (GMM) for crustal earthquakes in Northern South America (NoSAM crustal GMM), *Bull. Seismol. Soc.*
 398 *Am.*, 113, 186–203, 2023.
- 399 Arteta, C. A., Pajaro, C. A., Mercado, V., Montejó, J., Arcila, M., and Abrahamson, N. A.: Ground-motion
 400 model for subduction earthquakes in northern South America, *Earthquake Spectra*, 37, 2419–2452, 2021.
- 401 [Bastías, N. and Montalva, G. A.: Chile Strong Ground Motion Flatfile, *Earthq. Spectra*, 32, 2549–2566, 2016.](#)
 402 [Bayless, J. and Abrahamson, N. A.: Summary of the BA18 ground-motion model for Fourier amplitude spectra](#)
 403 [for crustal earthquakes in California, *B. Seismol. Soc. Am.*, 109, 2088–2105, 2019.](#)
- 404 Bindi, D., Kotha, S. R., Weatherill, G., Lanzano, G., Luzi, L., and Cotton, F.: The pan-European engineering
 405 strong motion (ESM) flatfile: consistency check via residual analysis, *Bull. Earthquake Eng.*, 17, 583–602,
 406 2019.
- 407 Boore, D. M.: Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal
 408 components of motion, *Bull. Seismol. Soc. Am.*, 100, 1830–1835, 2010.
- 409 Brune, J. N.: Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.*, 75,
 410 4997–5009, 1970.
- 411 Buscetti, M., Traversa, P., Perron, V., Rischette, P., Hollender, F., and Shible, H.: Epos-France Database of
 412 Earthquake Ground Motion for Mainland France: 1996–2021 Updated Version, *Seismol. Res. Lett.*, 96, 1214–
 413 1226, <https://doi.org/10.1785/0220240053>, 2024.
- 414 Dawood, H. M., Rodriguez-Marek, A., Bayless, J., Goulet, C., and Thompson, E.: A flatfile for the KiK-net
 415 database processed using an automated protocol, *Earthquake Spectra*, 32, 1281–1302, 2016.
- 416 Ekström, G., Nettles, M., and Dziewonski, A. M.: The global CMT project 2004-2010: Centroid-moment
 417 tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter.*, 200–201, 1–9,
 418 <https://doi.org/10.1016/j.pepi.2012.04.002>, 2012.
- 419 Eraso, J. and Montejó, J.: Mapa de Vs30 Integrado para Colombia, Servicio Geológico Colombiano, Bogotá,
 420 2019.
- 421 Goulet, C. A., Kishida, T., Ancheta, T. D., Cramer, C. H., Darragh, R. B., Silva, W. J., and Youngs, R. R.:
 422 PEER NGA-east database, *Earthquake Spectra*, 37, 1331–1353, 2021.
- 423 Hanks, T. C. and Kanamori, H.: A moment magnitude scale, *J. Geophys. Res.*, 84, 2348–2350, 1979.
- 424 ~~International Seismological Centre: ISC-EHB dataset [data set], International Seismological Centre, Thatcham,~~
 425 ~~<https://doi.org/10.31905/PY08W6S3> (last access: 12 August 2026).~~
- 426 [International Seismological Centre: ISC-EHB dataset \[data set\], International Seismological Centre, Thatcham,](#)
 427 <https://doi.org/10.31905/PY08W6S3>, 2026 (last access: 12 August 2026).

- 428 Kottke, A., Abrahamson, N. A., Boore, D. M., Bozorgnia, Y., Goulet, C. A., Holleback, J., Kishida, T., Der
 429 Kiureghian, A., Ktenidou, O.-J., Kuehn, N., Rathje, E. M., Silva, W. J., Thompson, E., and Wang, X.: Selection
 430 of random vibration procedures for the NGA East Project, PEER Report No. 2018/05, Pacific Earthquake
 431 Engineering Research Center, University of California, Berkeley, CA, 2018.
- 432 [Konno, K. and Ohmachi, T.: Ground-motion characteristics estimated from spectral ratio between horizontal
 433 and vertical components of microtremor, *B. Seismol. Soc. Am.*, **88**, 228–241, 1998.](#)
- 434 Lanzano, G., Sgobba, S., Luzi, L., Pacor, F., Puglia, R., Felicetta, C., and D’amico, M.: The pan-European
 435 Engineering Strong Motion (ESM) flatfile: comparison with NGA-West2 database, *Boll. Geofis. Teor. Appl.*,
 436 **61**, 2020.
- 437 Marín-Arias, J. P., Valencia Munoz, L. Y., and Rendon Cardona, L. M.: Macroseismic analysis of Murindo
 438 earthquake, Antioquia (Colombia) October 1992, *Boletín de Geología*, **31**, 2009.
- 439 Marín-Arias, J. M., González, H. C., and Hurtado, E. D. J. S.: Modelo geométrico del foco del terremoto de
 440 Popayán (Colombia) a partir de datos macrosísmicos, *Boletín de Geología*, **28**, 93–109, 2006.
- 441 Martinez-Jaramillo, D., Kotha, S. R., Zúñiga, F. R., and Lacan, P.: Colombian Ground Motion: SA and FAS
 442 [data set], Zenodo, <https://doi.org/10.5281/zenodo.19056074>, 2026.
- 443 Minson, S. E. and Dreger, D. S.: Stable inversions for complete moment tensors, *Geophys. J. Int.*, **2**, 585–592,
 444 <https://doi.org/10.1111/j.1365-246X.2008.03797.x>, 2008.
- 445 Montejo, J., Arcila Rivera, M. M. M., and Zornosa, D. G.: Integrated Seismic Catalog for Colombia, *Boletín
 446 Geológico*, **50**, <https://doi.org/10.32685/0120-1425/bol.geol.50.1.2023.665>, 2023.
- 447 Mosquera-Machado, S., Lalinde-Pulido, C., Salcedo-Hurtado, E., and Michetti, A. M.: Ground effects of the 18
 448 October 1992, Murindo earthquake (NW Colombia), using the Environmental Seismic Intensity Scale (ESI
 449 2007) for the assessment of intensity, *Geol. Soc. Spec. Publ.*, **316**, 2009.
- 450 Oliveti, I., Faenza, L., and Michelini, A.: Intensity-ground motion dataset for Italy (INGe), version 2.0 [data
 451 set], Zenodo, <https://doi.org/10.13127/inge.2>, 2021.
- 452 Pacor, F., Felicetta, C., Lanzano, G., Sgobba, S., Puglia, R., D’Amico, M., Russo, E., Baltzopoulos, G., and
 453 Iervolino, I.: NESS v1.0: A worldwide collection of ground-motion data to investigate near source effects,
 454 *Seismol. Res. Lett.*, <https://doi.org/10.1785/0220180149>, 2018.
- 455 Pájaro, C. A., Arteta, C. A., Mercado, V., Montejo, J., Arcila, M., and Abrahamson, N. A.: Partially non-ergodic
 456 ground motion model for the Bucaramanga seismic nest in Northern South America (NoSA Nest GMM), *Bull.
 457 Earthquake Eng.*, **22**, 3677–3702, 2024.
- 458 Paris, G. and United States Geological Survey: Map and database of Quaternary faults and folds in Colombia
 459 and its offshore regions, US Geological Survey, No. 00-0284, 2000.
- 460 Rodríguez, L., Diederix, H., Torres, E., Audemard, F., Hernández, C., Singer, A., and Yezpe, S.: Identification
 461 of the seismogenic source of the 1875 Cucuta earthquake on the basis of a combination of neotectonic,
 462 paleoseismologic and historic seismicity studies, *J. S. Am. Earth Sci.*, **82**, 274–291, 2018.
- 463 Sgobba, S., Felicetta, C., Lanzano, G., Ramadan, F., D’Amico, M., and Pacor, F.: NESS2.0: An Updated
 464 Version of the Worldwide Dataset for Calibrating and Adjusting Ground-Motion Models in Near Source, *Bull.
 465 Seismol. Soc. Am.*, <https://doi.org/10.1785/0120210080>, 2021.
- 466 [Sharma, S., Mannu, U., and Bora, S. S.: A Flat-File Compilation of Strong Ground-Motion Intensity Measures
 467 for Crustal Earthquakes in the Indian Region, *Seismica*, 2025.](#)

- 468 Strasser, F. O., Arango, M. C., and Bommer, J. J.: Scaling of the source dimensions of interface and intraslab
469 subduction-zone earthquakes with moment magnitude, *Seismol. Res. Lett.*, 81, 941–950, 2010.
- 470 Vanneste, K., Onvani, M., Rapagnani, G., and Van Noten, K.: The BELSHAKE Database of Earthquake
471 Ground Motion in Belgium, *Seismol. Res. Lett.*, 97, 918–933, <https://doi.org/10.1785/0220250321>, 2026.
- 472 Vargas-Jiménez, C. A. and Monsalve-Jaramillo, H.: 3D velocity structure around the source area of the Armenia
473 earthquake: 25 January 1999, Mw= 6.2 (Colombia), *Bol. Soc. Geol. Mex.*, 61, 339–351, 2009.
- 474 Wells, D. L. and Coppersmith, K. J.: New empirical relationships among magnitude, rupture length, rupture
475 width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.*, 84, 974–1002, 1994.
- 476 Weston, J., Engdahl, E. R., Harris, J., Di Giacomo, D., and Storchak, D. A.: ISC-EHB: Reconstruction of a
477 robust earthquake dataset, *Geophys. J. Int.*, 214, 474–484, <https://doi.org/10.1093/gji/ggy155>, 2018.