Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Brazilian Seismic Acceleration Database and GMPE Weighting for PSHA Studies

Bruno Collaço

bruno@iag.usp.br

Universidade de São Paulo

Marcelo Assumpção

Universidade de São Paulo

Marcelo Bianchi

Universidade de São Paulo

Marcelo Rocha

University of Brasília

Luciano Assis

Vale SA

Maximiliano Simão

Vale SA

Research Article

Keywords: Peak Ground Acceleration, PSHA, GMPE, NGA-East, Brazil, Strong-Motion Database.

Posted Date: June 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9760942/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Brazilian Seismic Acceleration Database and GMPE Weighting for PSHA Studies

Bruno Collaço[1,2*], Marcelo Assumpção[1], Marcelo Bianchi[2], Marcelo Rocha[2], Luciano Assis[4], Maximiliano Simão[4]

[1*]IEE/USP, University of São Paulo, Av. Prof. Luciano Gualberto, 1289, Sao Paulo, 05508-900, Sao Paulo, Brazil.
[2]Observatório Sismológico, University of Brasília, Prédio SG-13, Campus Universitário Darcy Ribeiro, Asa Norte, SGAN, Plano Piloto, Brasilia, 70910-900, Distrito Federal, Brazil.
[3]Centro de Sismologia, University of São Paulo, Rua do Matão 1226, Cid. Universitária, São Paulo, 05508-090, São Paulo, Brazil.
[4]Diretoria de Serviços Geotécnicos, Vale/SA, Av. Dr. Marco Paulo Simon Jardim, 3580, Belo Horizonte, 34006-270, Minas Gerais, Brazil.

*Corresponding author(s). E-mail(s): bruno@iag.usp.br;
Contributing authors: marcelo.iagusp@gmail.com; m.bianchi@iag.usp.br; marcelorocha@unb.br; luciano.assis@vale.com; max.simão@vale.com;

Abstract

Ground-Motion Prediction Equations (GMPEs) are a fundamental component of Probabilistic Seismic Hazard Analysis (PSHA). In low-seismicity regions such as Brazil, the lack of locally calibrated GMPEs has led to the adoption of models developed for tectonically analogous regions, particularly Central and Eastern North America (CENA). Previous Brazilian PSHA studies based on foreign GMPEs produced high peak ground acceleration (PGA) estimates, raising concerns regarding their engineering applicability. In this study, the full suite of 17 NGA-East USGS models is empirically evaluated using Brazilian strong-motion data. A database comprising approximately 150 PGA records from 27 regional earthquakes ($3.0 < Mw < 5.6$), combined with site characterization surveys measuring Vs30, is used for validation. Model performance is quantified through complementary metrics, namely normalized residuals (Z-scores), root-mean-square error (RMS), log-likelihood (LLH), and the stochastic area metric (SA), and examined across magnitude, distance, and site-condition subgroups to detect systematic biases. These metrics are integrated through a

multicriteria normalization scheme to derive epistemic weights for PSHA logic trees. Results indicate a systematic tendency of NGA-East models to overestimate PGA in Brazil, attributed to differences in crustal attenuation, stress-drop, and focal-depth distributions between CENA and the Brazilian stable platform. USGS_13 exhibits the best overall performance (RMS = 1.26, LLH = 2.47, SA = 0.16), while the NGA-East Graizer model provides complementary lower-bound predictions. A case study for Southeastern Brazil shows that the resulting BR26 hazard maps are more spatially homogeneous than earlier Brazilian PSHA results, primarily due to the inclusion of lower-magnitude seismicity (Mw 3 to 5) in the activity rate estimation, consistent with GEM2019. The empirically weighted GMPE ensemble contributes independently by reducing systematic PGA overestimation. The compiled dataset is made publicly available, and the proposed framework offers a robust, reproducible methodology transferable to other low-seismicity intraplate settings.

Keywords: Peak Ground Acceleration, PSHA, GMPE, NGA-East, Brazil, Strong-Motion Database.

1 Introduction

Ground Motion Prediction Equations (GMPE) are key elements of Probabilistic Seismic Hazard Analysis (PSHA). GMPE relate earthquake source parameters, wave propagation effects, and local site conditions to expected ground-motion intensity measures, such as peak ground acceleration (PGA). In many countries, however, the development of regionally calibrated GMPE is constrained by limited instrumental data and low seismicity rates.

Brazil exemplifies a typical intraplate, low-seismicity environment characterized by infrequent moderate-to-large earthquakes and sparse strong-motion records, i.e., conditions that fundamentally limit the development of regionally calibrated GMPE. As a result, PSHA studies in Brazil have historically relied on GMPEs developed for other stable continental regions, particularly the Central and Eastern North America (CENA) ([1], [2], [3], [4]). Although tectonic similarities exist, the direct transfer of such models may introduce epistemic bias and inflate hazard estimates

Early PSHA applications in Brazil reported unexpectedly high PGA values exceeding 0.30 g for 10,000-year return periods across multiple regions [5]. Engineering practitioners received these estimates with considerable skepticism, particularly in critical applications such as tailings dam safety assessments. This skepticism motivated the present systematic evaluation of foreign GMPE applicability to Brazilian conditions.

A database of peak ground acceleration measured at seismic stations in Brazil for small to moderate earthquake (magnitudes 3.0 to 5.6 Mw) was prepared, including corrections for sampling rate and Vs30 local site conditions, and used to rank NGA-East GMPEs for application in Brazil.

2 Database and Site Characterization

The validation of foreign GMPEs requires observational data representative of Brazilian seismic conditions. A database compiled by the University of São Paulo Seismological Center in collaboration with the University of Brasília, University of Rio Grande do Norte, and the Brazilian Seismographic Network (RSBR) was used in this study. The dataset comprises 150 PGA measurements recorded at RSBR stations from 27 regional earthquakes (red circles in Fig. 1). We used the geometric mean of the maximum acceleration from two horizontal components, after converting the original records to acceleration. All earthquake information is available in Table 1.

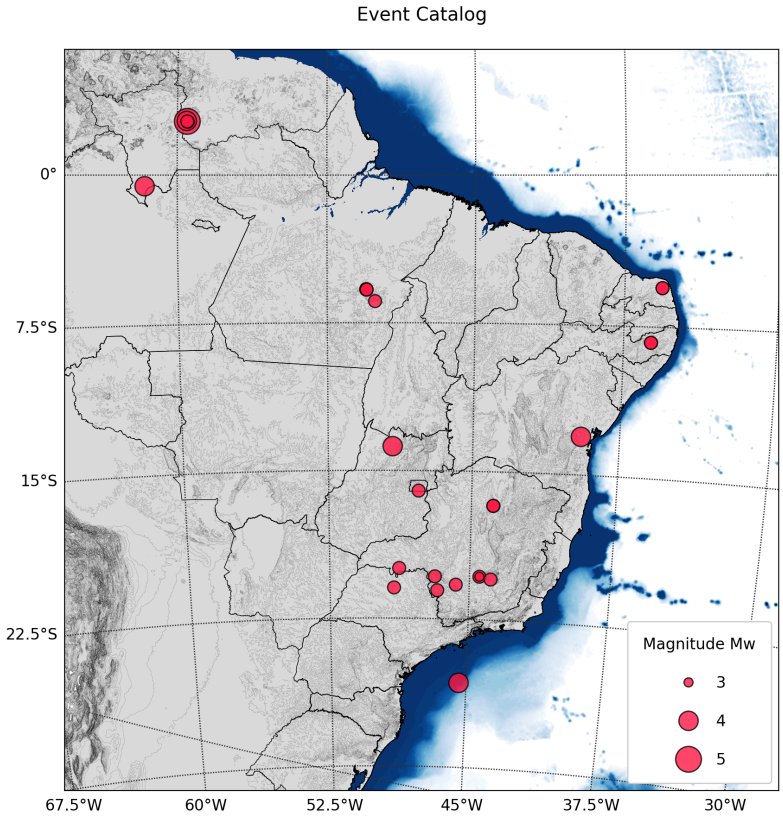


Fig. 1 Map of epicenters of the 27 regional earthquakes used in this study. Red circles show the locations of the analyzed events, with symbol size proportional to Mw magnitude. All earthquake information is available in the supplementary material.

Since PGA estimates are sensitive to the sampling rate of the recording instrument, and broadband stations typically operate at lower sampling rates than standard accelerographs, a correction was applied to account for potential underestimation of PGA values, as described in the following section.

As shown in Table 1, a significant portion of the analyzed events have moment magnitudes (Mw) below 4.0. However, most of ground motion prediction equations were developed for magnitudes Mw 4.0 or above, which limits a direct comparison between observed Brazilian data and predicted values. To overcome this limitation, a normalization procedure was applied: observed PGA values from events with Mw < 4.0 were scaled upward to a reference magnitude of Mw 4.0, following the methodology described in the next sections. This approach allows for a consistent comparison between observed ground motions and GMPE predictions within their intended magnitude range.

Table 1 Brazilian seismic events catalog.

ID	Origin Time (UTC)	Lat (°)	Lon (°)	Depth (km)	Mw
2021031_GuyanaMain	2021-01-31 19:05:13	2.70	−59.52	3.0	5.64*
2008114_SVicente	2008-04-23 00:00:48	−25.70	−45.41	17.0	4.69*
2010281_MaraRosa	2010-10-08 20:16:54	−13.77	−49.16	1.4	4.35*
20210428_Barcelos	2021-04-28 03:26:41	−0.55	−61.63	17.0	4.32
2021085_GuyanaAft	2021-03-26 15:57:43	2.73	−59.55	2.2	4.25
2020243_Amargosa	2020-08-30 10:44:28	−13.07	−39.47	1.2	4.10*
20250403_Salobo	2025-04-03 07:02:16	−5.78	−50.52	0.0	3.85
20250710_Salobo	2025-07-10 00:03:50	−5.80	−50.52	0.0	3.80
2021093_GuyanaAft3	2021-04-03 14:41:01	2.70	−59.52	1.1	3.75
1996292_Jacuih	1996-10-18 21:44:01	−21.04	−46.74	0.0	3.70
2010011_Taipu	2010-01-11 15:53:56	−5.55	−35.65	5.5	3.70
20250710_Salobo2	2025-07-10 09:46:10	−5.79	−50.51	0.0	3.70
2021090_GuyanaAft2	2021-03-31 21:24:09	2.69	−59.50	1.2	3.67
2019140_Delfinopolis	2019-05-20 19:46:34	−20.34	−46.86	20.0	3.64
20240618_Frutal	2024-06-18 13:08:01	−19.95	−48.80	0.0	3.60
2006140_SCaetano	2006-05-20 04:26:05	−8.26	−36.16	5.0	3.59
20230123_Carajas	2023-01-23 02:23:49	−6.40	−50.07	0.0	3.54
1997321_Guapeh	1997-11-17 17:27:01	−20.75	−45.75	0.0	3.38
2000325_Brasilia	2000-11-20 09:36:33	−16.01	−47.79	0.0	3.32
1993361_Itaguara	1993-12-27 22:48:11	−20.32	−44.47	4.0	3.26
2012354_MClaros_04h	2012-12-19 04:54:38	−16.70	−43.88	1.8	3.26
2012354_MClaros_05h	2012-12-19 05:31:16	−16.70	−43.88	1.5	3.26
2007079_SCaetano	2007-03-20 13:38:52	−8.26	−36.16	5.0	3.15
2010009_Taipu	2010-01-09 18:18:05	−5.55	−35.65	5.5	3.15
2019329_Congonhas	2019-11-25 23:27:26	−20.43	−43.89	0.6	3.15
2000272_Catiguah	2000-09-28 16:04:07	−20.93	−49.05	0.0	3.04
1993362_Itaguara	1993-12-28 21:32:06	−20.32	−44.47	4.0	2.82

* Mw measured directly; other magnitudes converted from mb.

Fig. 2 shows the distribution of the 150 PGA records as a function of epicentral distance and moment magnitude. The plot highlights the sparse coverage at larger

magnitudes, with only a single event exceeding Mw 5.0 (the 2021 Guyana mainshock, Mw 5.64). Records at short epicentral distances (below 100 km) are predominantly associated with lower-magnitude events, while higher-magnitude events tend to be recorded at greater distances, reflecting the typical detection pattern of the RSBR network. This magnitude–distance distribution underscores the importance of the normalization procedure mentioned above, as a large fraction of the dataset falls below the Mw 4.0 threshold commonly adopted by GMPEs.

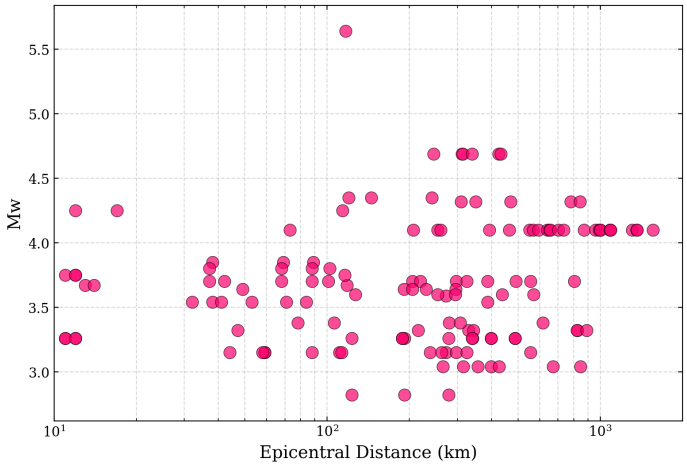


Fig. 2 Distribution of the 150 PGA records used in this study as a function of epicentral distance (km) and moment magnitude (Mw).

Although the full database includes records at rupture distances up to approximately 1500 km, which reflects the broad detection capability of the RSBR network for moderate-to-large events, the residual analysis presented in Section 3 is restricted to records with Rrup $\leq$ 100 km. This distance threshold was chosen because (1) NGA-East GMPEs were primarily developed and validated for near-to-intermediate source distances relevant to engineering applications, and (2) beyond 100 km, post-critical Moho reflections (PmP/SmS phases) increasingly dominate the wavefield in cratonic settings [6], introducing propagation effects not well captured by standard distance-scaling formulations.

2.1 Effect of sampling rate and correction factors

Most velocity sensors of the seismographic stations operate at sampling rates of 50 or 100 samples/s (sps), lower than the usual 1000 sps of a typical accelerograph. This yields PGA values lower than would be recorded in an accelerograph, especially for small (<5 Mw) and near (<100 km) earthquakes. Therefore, a correction is needed to account for the "low sampling rate effect". For stations that operated with two trigger modes, PGA values recorded at 200 sps were used directly. At these high-sampling stations PGA values at 100 and 50 sps were measured from traces downsampled from

the original 200 sps recordings to estimate possible bias in PGA estimates. We assumed that at 200 sps PGA is very close to the value that would be observed if the event had been recorded in a proper accelerograph.

Fig. 3 shows the "attenuation factor", factor = PGA (100Hz) / PGA (200Hz), as a function of distance and mb magnitude (we now use "100Hz" as equivalent to "100 sps"). It is seen that the PGA attenuation factor increases with distance (getting closer to 1.0). That is, longer distances attenuate the high frequencies more rapidly making the PGA (100Hz) closer to the PGA (200Hz).

Fig. 3 also shows a slight trend of increasing ratio with magnitude. Higher magnitudes tend to have lower corner frequencies which decreases the relative contribution of higher frequencies to the PGA. A simple attenuation factor was fitted to the data, as:

$$f = a\,R_{\rm epi} + b\,M + c \tag{1}$$

where,

$$a = 0.22 \pm 0.05, b = 0.07 \pm 0.08, c = 0.20 \pm 0.29, \mathrm{RMS}_{\rm residual} = 0.150 \tag{2}$$

In this model, the dependence on the magnitude (coefficient b) is not significant and there is a large covariance between b and c.

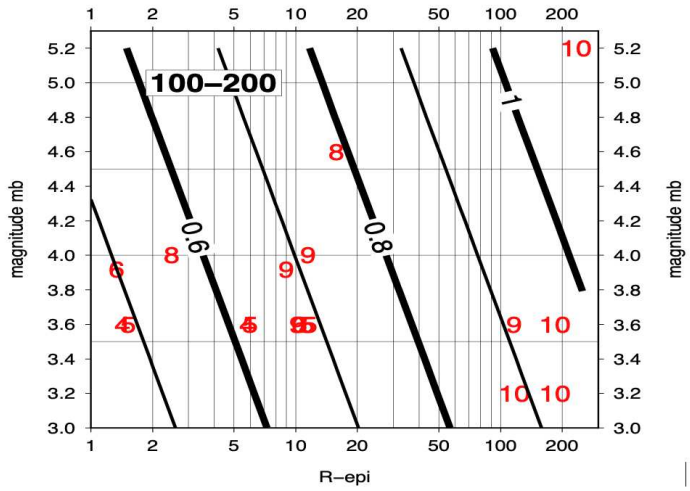


Fig. 3 Attenuation factor for 100 Hz, i.e., ratio of PGA(100Hz)/PGA(200Hz) as a function of epicentral distance and mb magnitude. Red numbers are PGA ratios multiplied by 10 so as not to clutter the figure. The straight lines show the best-fitting function $factor = a\,R_{epi} + b\,mag + c$ (see eq. 1), contoured at 0.1 interval.

Fig. 4 shows the attenuation factor for 50 Hz, with the function factor(dist, mb) fitted only for distances < 100 km. Fitting the 50 Hz data for all distances produces coefficients with large uncertainties and a negative b value, which is unphysical. For near distances (< 100km), the 50 Hz factor is:

$$f = a\,R_{\rm epi} + b\,M + c \tag{3}$$

where,

$$a = 0.28 \pm 0.11, b = 0.10 \pm 0.14, c = -0.13 \pm 0.53, \mathrm{RMS}_{\rm residual} = 0.144 \tag{4}$$

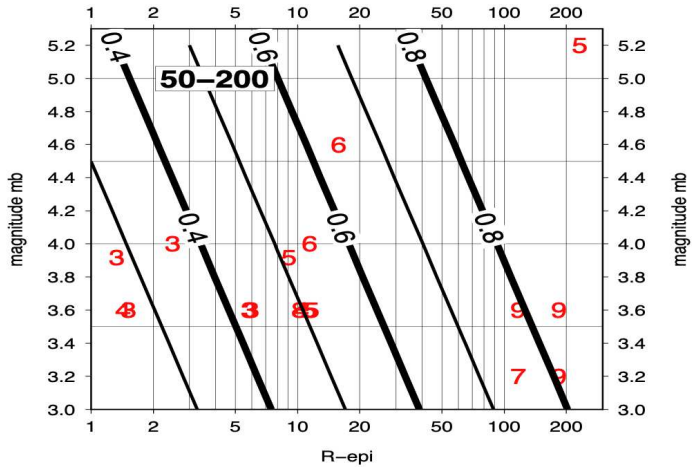


Fig. 4 Attenuation factor, PGA(50)/PGA(200), for 50 Hz. Only data for distances R_{epi} <100 km were fitted with the function $factor = a\,R_{epi} + b\,mag + c$, shown by the straight lines at 0.1 interval.

At distances longer than 100 km, the attenuation factor converges to 1 and fitting a linear function may not be appropriate, especially when only a few data points are available. After several tests, the best attenuation factors to be adopted to correct the PGA values of the Brazilian database were functions fitted to distances shorter than 100 km. For the 100 Hz, a magnitude dependence improves the corrections (Fig. 5). For the 50 Hz, the best attenuation function depends only on the epicentral distance (Fig. 6).

Table 2 shows a summary of corrections applied due to sampling rate. The correction factor is expressed as a linear regression of epicentral distance R_{epi} and magnitude (mag). RMS residual is defined as:

$$\mathrm{RMS} = \sqrt{\frac{\sum r_i^2}{\nu}} \tag{5}$$

where ν is the number of degrees of freedom.

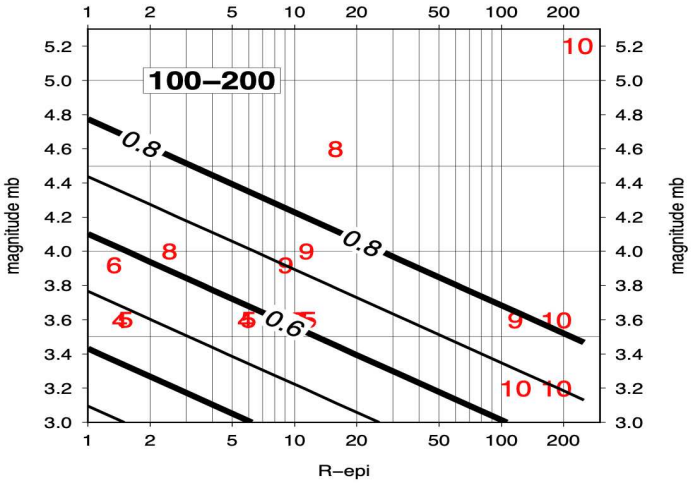


Fig. 5 Attenuation factor for 100Hz, fitted for near distances only (R_{epi} <100km).

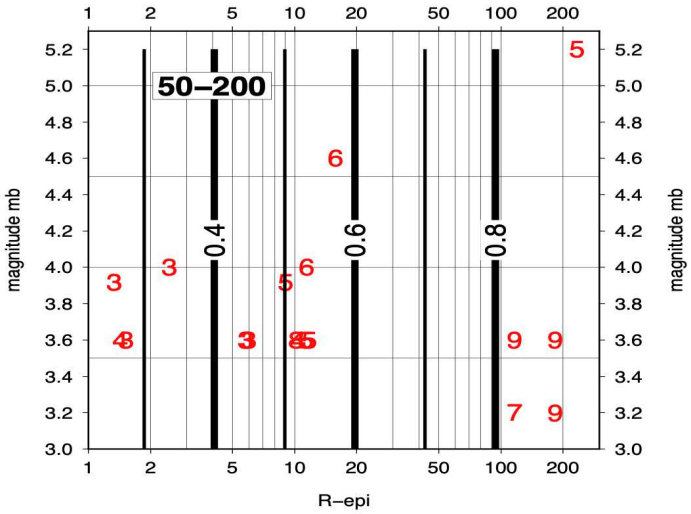


Fig. 6 Attenuation factor for 50Hz, fitted for near distances only (R_{epi} <100km) with no dependence on the magnitude: $factor = a\,R_{epi} + b\,mag + c$.

Table 2 Summary of PGA correction factors for low sampling rate effect.

Correction	R_{epi} coeff.	M coeff.	Intercept	RMS Residual
100 → 200 sps (Fig. 5)	0.16 ± 0.12	0.30 ± 0.16	-0.62 ± 0.59	0.161
50 → 200 sps (Fig. 6)	0.29 ± 0.10	—	0.22 ± 0.09	0.140

2.2 Normalizing PGA to Mw = 4.0

Each GMPE presents a distinct functional dependence on magnitude. In several models, the magnitude scaling is not constant but varies with distance, reflecting magnitude–distance interaction effects embedded in the regression formulation.

As a first-order approximation to normalize all observed PGA values to a common reference magnitude (Mw = 4.0), we adopted an average magnitude scaling factor, near Mw=4.0 for all 17 NGA-E GMPEs. Average scaling factors from two representative distances were estimated. Specifically, we considered:

- At $R = 10$ km:

$$\frac{d\left(\log_{10}\mathrm{PGA}\right)}{dM} = 0.50 \tag{6}$$

- At $R = 100$ km:

$$\frac{d\left(\log_{10}\mathrm{PGA}\right)}{dM} = 0.72 \tag{7}$$

Assuming a linear dependence of the magnitude scaling coefficient on $\log_{10}(R)$, we fitted a straight line through these two points. This yields a distance-dependent magnitude correction term:

$$\frac{d\left(\log_{10}\mathrm{PGA}\right)}{dM} = 0.22\ \log_{10}(R_{\mathrm{epi}}) + 0.28 \tag{8}$$

Thus, the normalized PGA at Mw = 4.0 is computed as:

$$\log_{10}(\mathrm{PGA}_{\mathrm{norm}}) = \log_{10}(\mathrm{PGA}) + \left[0.22\ \log_{10}(R_{\mathrm{epi}}) + 0.28\right](4.0 - M_w) \tag{9}$$

where R_{epi} is given in km.

Table 3 shows an extract of the normalized PGA ("PGA_N") together with the original PGA ("PGA_{obs}", either measured or corrected for low sampling rate). The complete dataset, including all 150 records with site characterization parameters, is provided as Table S3 in the supplementary material and is available for download in machine-readable format at https://doi.org/10.5281/zenodo.20287122.

Fig. 7 presents the normalized PGA values (PGA_N) as a function of rupture distance (Rrup) for all 150 records. Blue circles represent events with original Mw $<$ 4.0, whose PGA values were scaled up to a reference magnitude of Mw 4.0, while red

Table 3 Extract of the PGA dataset with observed values (PGA_{obs}) and values normalized to Mw = 4.0 (PGA_N). The complete table is provided as Table S3 in the supplementary material and is available for download at https://doi.org/10.5281/zenodo.20287122.

Station	Rrup (km)	PGA_{obs} (g)	PGA_N (g)	Mw	ID
DAES	244.8	2.64×10^{-3}	7.34×10^{-4}	4.69	2008114_SVicente
SPB	309.3	1.34×10^{-3}	3.60×10^{-4}	4.69	2008114_SVicente
ESAR	314.5	1.06×10^{-3}	4.12×10^{-4}	4.69	2008114_SVicente
VABB	337.4	3.54×10^{-3}	9.37×10^{-5}	4.69	2008114_SVicente
RCLB	421.9	3.57×10^{-3}	9.14×10^{-5}	4.69	2008114_SVicente
SP7	429.8	4.66×10^{-5}	1.19×10^{-5}	4.69	2008114_SVicente
BDFB	60.5	1.08×10^{-5}	3.09×10^{-4}	3.32	2000325_Brasilia
CORB	215.0	2.70×10^{-5}	9.35×10^{-4}	3.32	2000325_Brasilia
PORB	339.3	2.15×10^{-5}	7.97×10^{-5}	3.32	2000325_Brasilia
PAZB	358.5	9.18×10^{-5}	3.43×10^{-4}	3.32	2000325_Brasilia
			⋮		
BB19B	125.7	3.59×10^{-4}	6.91×10^{-4}	3.60	20240618_Frutal
CANS	253.7	8.62×10^{-5}	1.66×10^{-4}	3.60	20240618_Frutal
PMNB	292.7	6.47×10^{-5}	1.24×10^{-4}	3.60	20240618_Frutal
BSCB	435.7	2.10×10^{-5}	4.05×10^{-5}	3.60	20240618_Frutal
DIAM	360.7	2.17×10^{-5}	4.18×10^{-5}	3.60	20240618_Frutal
SLBO	4.7	3.12×10^{-4}	4.78×10^{-4}	3.70	20250710_Salobo2
SRN1	18.9	1.89×10^{-4}	2.90×10^{-3}	3.70	20250710_Salobo2
SRSU	67.7	6.53×10^{-4}	1.00×10^{-3}	3.70	20250710_Salobo2
PRPB	87.7	2.44×10^{-4}	3.98×10^{-5}	3.70	20250710_Salobo2
SRLE	100.7	1.20×10^{-4}	1.97×10^{-5}	3.70	20250710_Salobo2

circles correspond to events with Mw > 4.0, whose PGA values were scaled down to the same reference magnitude.

Although the magnitude normalization used here is a rough approximation, the normalized values from both subsets follow a consistent attenuation trend, suggesting that the normalization procedure effectively removes the magnitude dependence and allows all records to be compared on a common basis.

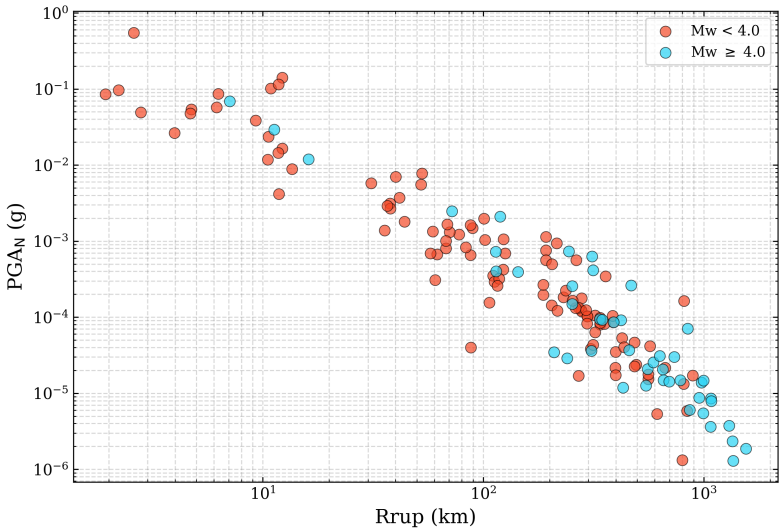


Fig. 7 Normalized PGA (PGA_N) versus rupture distance (Rrup) for the 150 records used in this study. Blue circles indicate events with original Mw < 4.0, scaled up to a reference magnitude of Mw 4.0. Red circles indicate events with Mw > 4.0, scaled down to the same reference magnitude.

2.3 Corrections for Vs30 at station site

Geophysical surveys, conducted in 2023/2024 using Multichannel Analysis of Surface Waves (MASW) and Horizontal-to-Vertical Spectral Ratio (HVSR) techniques, provided direct estimates of the average shear-wave velocity in the upper 30 m (Vs30) for many stations. The MASW surveys were supported by Vale/SA. The Vs30 estimates were obtained mainly by a joint inversion of phase velocity dispersion (from the MASW) and HVSR (Horizontal-to-Vertical Spectral ratios) using Geopsy [7]. The results and estimation methods are listed in Table 4.

Fig. 8 shows the distribution of measured Vs30 values for the 46 stations with geophysical site characterization. The histogram (upper panel) shows that the distribution is concentrated in the 500–1200 m/s range, with a geometric mean of 760 m/s (dashed line). The tail extending toward higher values (some stations exceeding 2000 m/s) reflects stations installed directly on exposed bedrock.

Table 4: Seismic stations with measured V_{S30} used in this study. For stations where in-situ geophysical measurements were unavailable, ad-hoc V_{S30} values were assigned based on lithological analogy and expert judgment, using published reference values for crystalline rock (granite, gneiss: $V_{S30} \approx$ 2000–3000 m/s; NEHRP site class A), consistent with values reported in the literature for similar geological settings (e.g., [8]; [9]).

Station	Lat (°)	Lon (°)	Elev. (m)	Site Type	Vs30 Method	Vs30 (m/s)
AGLP	−19.9381	−43.2278	1085	metasedimentary rock	MASW + HVSR	330
AREB	−21.3625	−46.1234	976	granite/gneiss	MASW + HVSR	800
BB19B	−19.5902	−45.7479	571	hard soil	MASW + HVSR	562
BDFB	−15.6418	−48.0148	1193	metasedimentary rock	MASW + HVSR	700
BOAV	−2.3953	−60.5225	114	granite/gneiss	HVSR	858
BSCB	−20.0945	−44.7635	935	granite/gneiss	MASW	978
BSFB	−18.5313	−40.8165	185	—	HVSR	1190
CAN3	−13.0993	−48.2898	330	—	MASW + HVSR	690
CANS	−20.2896	−46.3921	1018	—	HVSR	690
CHCB	−19.6120	−43.2402	752	metasedimentary rock	MASW + HVSR	700
CMC01	−15.3601	−39.5191	169	granite/gneiss	MASW + HVSR	391
DAES	−23.5453	−45.9239	828	granite/gneiss	ad-hoc	2700
DIAM	−18.2952	−43.6648	1280	—	MASW + HVSR	565
ESAB	−23.0229	−44.4162	8	—	HVSR	5000
FABR	−20.1182	−43.9133	1454	metasedimentary rock	MASW + HVSR	460
FJAO	−20.1001	−44.1281	1380	metasedimentary rock	MASW + HVSR	373
FRTB	−29.1876	−55.6589	340	limestone	MASW	500
FRTB	−23.3139	−49.5610	518	—	HVSR	500
FZND	−20.1445	−43.4230	620	metasedimentary rock	MASW + HVSR	400
GDU01	−13.7200	−39.3753	251	granite/gneiss	MASW	1500
GN3.6	−20.4000	−43.6060	890	metasedimentary rock	MASW + HVSR	265
GNG6	−19.9698	−43.5929	970	metasedimentary rock	MASW + HVSR	1000
GUY1	2.7953	−59.5401	135	sediment	HVSR	543
GUY2	2.6527	−59.6202	153	granite/gneiss	HVSR	635
GUY3	2.6584	−59.4356	160	hard soil	HVSR	541
GUY4	2.7010	−59.4982	128	granite/gneiss	HVSR	635
JANB	−15.0581	−44.3112	693	—	HVSR	694
LEVE	−16.8971	−43.8986	730	hard soil	MASW + HVSR	393
MCI7	−16.7753	−43.8213	571	hard soil	MASW + HVSR	412
MCI8	−16.6597	−43.9201	696	hard soil	MASW + HVSR	422
MCI9	−16.8981	−43.8933	679	hard soil	MASW	622
MNNB	−11.3521	−40.4356	570	granite/gneiss	MASW + HVSR	1001
NBPN	−10.8470	−40.1990	486	granite/gneiss	MASW + HVSR	1001
PMNB	−18.5473	−46.4380	950	—	HVSR	614
PRPB	−6.1724	−49.8150	265	granite/gneiss	MASW	709
RCLB	−21.1994	−47.5321	575	metasedimentary rock	MASW + HVSR	2011

(continued on next page)

(continued from previous page)

Station	Lat (°)	Lon (°)	Elev. (m)	Site Type	Vs30 Method	Vs30 (m/s)
RCBR	−5.8274	−35.9014	292	granite/gneiss	ad-hoc	1500
SAL2	−5.7839	−50.3106	362	metasedimentary rock	HVSR	787
SPB	−18.7029	−41.1847	715	—	HVSR	1500
SLBO	−5.8306	−50.5367	176	metasedimentary rock	HVSR	1100
SPB	−23.5927	−47.4270	647	limestone	ad-hoc	1500
SRN1	−6.0347	−50.2871	695	metasedimentary rock	HVSR	395
SRLE	−6.0344	−49.6167	600	granite/gneiss	MASW + HVSR	542
SRSU	−6.3866	−50.3591	716	granite/gneiss	MASW + HVSR	709
SSV2	−12.8507	−48.2143	315	—	HVSR	609
VABB	−23.0022	−46.9658	866	granite/gneiss	MASW + HVSR	1636

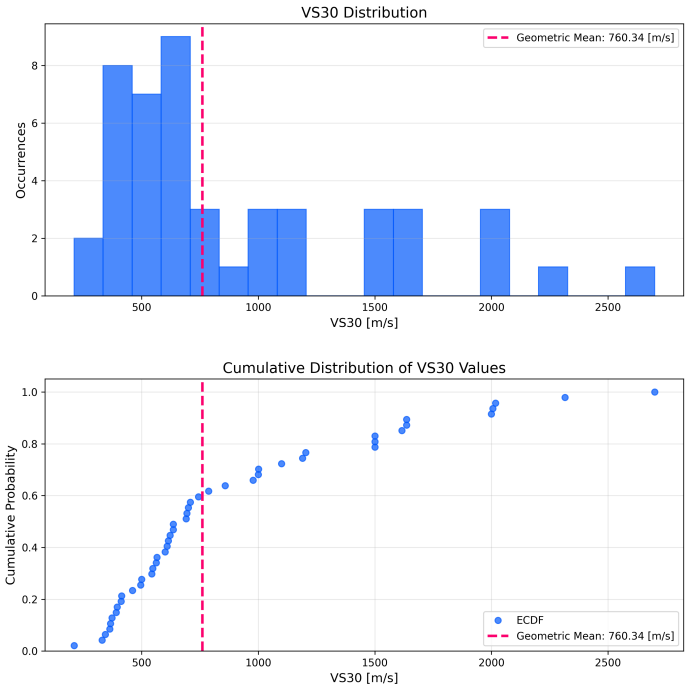


Fig. 8 Distribution of measured Vs30 values for the 46 stations with geophysical site characterization. *Upper panel:* histogram of Vs30 values. *Lower panel:* empirical cumulative distribution function (ECDF). The dashed line indicates the geometric mean of 760 m/s, adopted as the default Vs30 for stations without direct measurements.

The empirical cumulative distribution function (ECDF, lower panel) indicates that approximately 60% of the measured stations have Vs30 below 760 m/s, while the

remaining 40% correspond to stiffer sites, including several stations installed directly on exposed rock with Vs30 exceeding 1500 m/s.

For stations without direct Vs30 measurements, a value of 760 m/s was adopted, corresponding to the geometric mean obtained from the measured subset. This value falls at the boundary between NEHRP site classes B and C. Since most Brazilian seismic stations are installed on rock outcrops, Vs30 values consistent with NEHRP site class A (>1500 m/s) were initially assumed as a reasonable proxy in the absence of direct measurements. However, the geophysical surveys revealed that shallow weathering layers and soil cover over the bedrock significantly reduce the average shear-wave velocity in the upper 30 m, even at sites with exposed rock at the surface.

The complete list of all 82 stations used in this study, including both measured and assumed Vs30 values, is provided as supplementary material. Fig. 9 shows the spatial distribution of all stations, classified according to their Vs30 values.

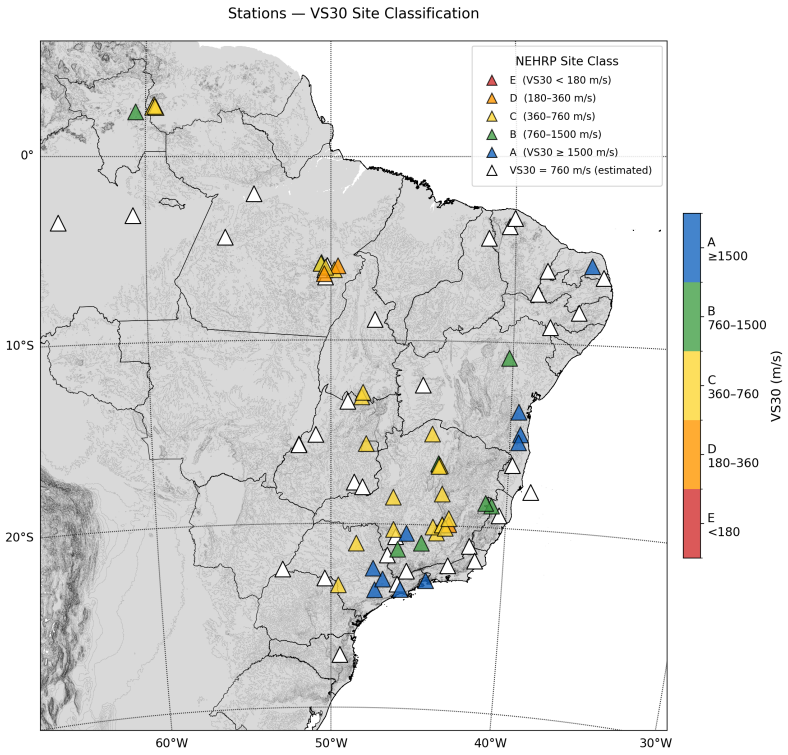


Fig. 9 Map of seismic stations used in this study, classified according to their Vs30 values. Stations with measured Vs30 are shown as colored symbols indicating the corresponding site class, while stations without direct measurements (Vs30 assumed as 760 m/s) are shown as open symbols.

The Vs30 site corrections were applied directly through the OpenQuake Hazard Library [10], as described in the following section. For each PGA record, the corresponding measured Vs30 value was used when available; otherwise, the average value

of 760 m/s was adopted. OpenQuake internally applies the site amplification corrections proposed by [9] and [11], ensuring a consistent treatment of site-effects across all records in the dataset.

3 Assessment of GMPEs

The performance of the USGS suite of NGA-East GMPEs [12], developed for the Central and Eastern United States, was evaluated against Brazilian strong-motion data. These models were selected due to their intended applicability to stable continental regions (SCRs), making them the most physically motivated choice for the Brazilian intraplate environment. The full set of 17 models was tested to preserve the broad epistemic uncertainty coverage of the NGA-East project. Evaluation was carried out by comparing observed and predicted PGA values, analyzing residual distributions, and examining systematic trends through magnitude, distance, and site-condition subgroup analyses. These complementary diagnostics were then integrated into a unified multicriteria framework, providing both quantitative model ranking and physical insight into the transferability of CEUS-based GMPEs to the Brazilian cratonic setting.

Ground-motion predictions were computed using the OpenQuake Hazard Library through its Strong-Motion Toolkit (SMTK) trellis interface, which provides a consistent implementation of all NGA-East models. For each earthquake–station pair in the database, the predicted peak ground acceleration (PGA) was calculated as a function of Mw magnitude, rupture distance (Rrup), and site condition represented by the Vs30 (as described above). For computational efficiency, each GMPE was evaluated over a fixed rupture distance grid (0–101 km, 1 km spacing) for each unique (Mw, Vs30) combination in the database, rather than being called individually for each observation. The predicted median PGA and total standard deviation at each observation's exact Rrup were then obtained by log-linear interpolation along this distance curve, introducing negligible error given the fine grid resolution. Records with Rrup $>$ 100 km were excluded from the residual analysis (the full database extends to distances of approximately 1500 km), as discussed in Section 2.

The evaluation used four complementary analytical components, described in the following subsections.

3.1 Residual Analysis

To quantify the difference between observed and predicted ground motions, residuals were computed in logarithmic scale. For each observation, the residual was normalized by the total variability of the corresponding GMPE, resulting in a normalized residual, or Z-score, defined as:

$$Z_i = \frac{\ln\left(\mathrm{PGA_{obs,i}}\right) - \ln\left(\mathrm{PGA_{pred,i}}\right)}{\sigma_{\mathrm{total},i}} \tag{10}$$

where PGA_{obs} is the observed peak ground acceleration, PGA_{pred} is the PGA predicted by the GMPE, and $\sigma_{\mathrm{total},i}$ is the total logarithmic standard deviation evaluated at the magnitude, distance, and Vs30 of the i-th observation. In modern GMPEs,

σ_{total} is not constant but varies with magnitude and, in some models, with distance – so each observation is normalized by its own aleatory uncertainty.

The use of Z-scores allows residuals from different GMPEs to be compared on a common, dimensionless scale, explicitly accounting for model-specific aleatory uncertainty. If a GMPE is perfectly calibrated for the dataset, the Z-scores should follow a standard normal distribution N(0, 1): zero mean indicates absence of systematic bias (the model neither over- nor underpredicts on average), and unit standard deviation indicates that the model's prescribed variability correctly captures the observed scatter. It should be noted, however, that this normalization has a known limitation: a GMPE with an excessively large σ_{total} will systematically produce low Z-scores regardless of the quality of its median prediction, as the inflated denominator compresses residuals toward zero. For this reason, Z-score statistics are interpreted here in conjunction with the LLH and RMS metrics, which penalize over-dispersed models and large median bias, respectively.

3.2 Global Performance Metrics

Three scalar metrics were computed to quantify the overall fit of each GMPE, described as follows.

Root-Mean-Square (RMS) of the residuals

The RMS of the Z-scores measures the average deviation from a perfect prediction. A value of zero would indicate exact agreement; increasing values reflect growing misfit. Unlike the mean, the RMS penalizes both bias and excess scatter, and is computed as:

$$\text{RMS} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} Z_i^2} \tag{11}$$

where N is the total number of observations and Z_i represents the normalized residual of the i-th record.

Average log-likelihood (LLH)

Following [13], LLH quantifies the statistical compatibility between the observed residual distribution and the distribution implied by each GMPE. It was implemented as misfit measure (negative log-likelihood), i.e., larger values indicate that the residuals deviate from the expected distribution, penalizing both bias and misrepresentation of the observed scatter.

Stochastic Area (SA)

Following [14], the stochastic area measures the integrated absolute difference between the empirical cumulative distribution function (CDF) of the observed $log_{10}(PGA)$ and the average predictive CDF constructed from the ensemble of individual normal distributions. A stochastic area of zero indicates perfect distributional agreement; larger values reflect systematic discrepancies in either the central tendency or the spread of the predictions.

3.3 Subgroup Analysis

To detect systematic trends that global metrics may mask, the Z-scores were also examined as a function of three key predictor variables: Mw magnitude, rupture distance, and site condition (Vs30).

Each variable was partitioned into six equal-width bins; the mean residual within each non-empty bin was computed; and a binned RMS (denoted RMS_{Mw}, RMS_{Rrup} and RMS_{Vs30}, respectively) was obtained as the root-mean-square of those bin means.

This procedure enables the identification of systematic patterns, such as distance-dependent attenuation biases, magnitude-dependent scaling issues, or site-condition effects, which may not be evident from the global RMS alone.

3.4 Multicriteria Scoring and GMPE Weighting

Because the employed performance metrics (RMS, subgroup RMS, LLH, and SA) have different scales and interpretations, direct comparison is not possible without prior normalization. The following procedure was adopted to combine them into a single composite score and, ultimately, into epistemic weights for the PSHA logic tree.

- **Standardization:** Each metric column was first z-standardized (zero mean, unit variance across all GMPEs), placing all metrics on a comparable scale regardless of their original units.
- **Direction alignment:** Because all six metrics follow a "lower is better" convention, the standardized values were sign-inverted so that higher values correspond to better performance.
- **Normalization:** The adjusted scores were then transformed to a common scale between 0 and 1 using a global min–max normalization as:

$$\hat{m}_{ij} = \frac{(m_{ij} - \min(m_i))}{(\max(m_i) - \min(m_i))} \tag{12}$$

  where m_{ij} is the value of metric i for GMPE j, and $\hat{m}_{ij}$ is the normalized metric value. The resulting matrix is suitable for direct visual comparison as a performance heatmap.

- **Composite score:** A composite performance score for each GMPE was then computed as the weighted mean of its normalized metric values:

$$\mathrm{Score_j} = \frac{1}{N} \sum_{i=1}^{N} \hat{m}_{ij} \tag{13}$$

  where N is the total number of performance metrics considered (6, in our study). In the base case, equal weights were assigned to all metrics ($1/N$ each). However, the framework allows the analyst to assign different weights to reflect domain-specific priorities (e.g., giving greater importance to distance-dependent performance).

- **Weight assignment:** The composite scores were rescaled to the interval [0.5, 1.0] using a min-max transformation, ensuring that the lowest-ranked model retains at least 50% of the score of the best-performing model. This prevents any GMPE from receiving a near-zero weight while still preserving meaningful contrasts between models. The choice of the lower bound is acknowledged to be subjective; the value of 0.5 was adopted as a reasonable threshold reflecting the principle that all selected models carry some epistemic value, but a formal sensitivity analysis of alternative bounds is recommended for future work. The rescaled scores were then normalized so that their sum equals unity, yielding a set of epistemic weights for the GMPEs:

$$w_j = \frac{Score_j}{\sum_{k=1}^{k} Score_k} \tag{14}$$

  where k is the total number of GMPEs evaluated. These weights represent the relative compatibility of each GMPE with the Brazilian strong-motion dataset and can be directly incorporated into PSHA logic trees as objective epistemic weights. The procedure ensures that the final weighting reflects a balanced assessment of each model's ability to reproduce the observed ground motions across the full range of magnitudes, distances and site conditions present in the database.

4 Results and Discussion

The empirical evaluation of the USGS NGA-East GMPE suite reveals systematic model-data discrepancies with important implications for Brazilian PSHA applications. This section presents quantitative performance rankings, identifies physical sources of bias, and discusses the broader significance of these findings for hazard assessment.

Fig. 10 presents the comparison between observed PGA values and model predictions as a function of rupture distance, with the corresponding Z-score RMS values indicated in parentheses in the legend. The PGA values were corrected for Vs30. For events with Mw<4.0, the PGA values were normalized to 4.0. Larger events were not normalized. A general tendency of overestimation is observed for most NGA-East models. This behavior suggests that attenuation characteristics embedded in CEUS-based models may not fully represent wave propagation and geometrical spreading effects in Brazilian crustal conditions. Among all evaluated models, USGS_13 shows the closest alignment with observed values, exhibiting the lowest global RMS and reduced systematic bias.

The observed overestimation pattern likely reflects fundamental differences in source and propagation characteristics between Brazil and CEUS. First, Brazilian cratonic lithosphere may exhibit higher anelastic attenuation than CEUS crustal structures, leading to less efficient wave propagation at intermediate distances. Additionally, systematic differences in stress drop are also expected: earthquakes in the Brazilian SCR tend to occur at shallower focal depths with lower stress-drop distributions [15] compared to CEUS events, which are often characterized by relatively high stress drops [16].

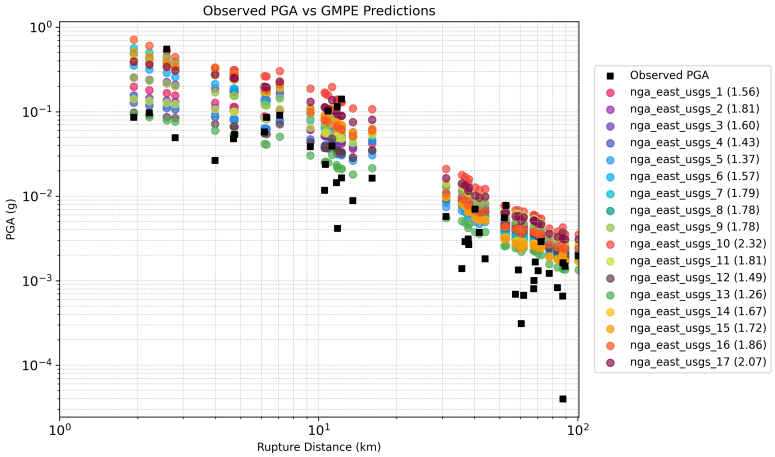


Fig. 10 Comparison between observed PGA values (black points) and USGS NGA-East GMPE predictions (colored symbols) as a function of rupture distance. Each model is identified by color and annotated with the corresponding RMS values indicated in parentheses in the legend.

Since the NGA-East models were calibrated to CEUS stress-drop levels, their application to regions with systematically different source characteristics can introduce magnitude and distance dependent biases. In addition, focal depth affects the partitioning between geometric spreading regimes and the distance at which post-critical Moho reflections (PmP/SmS) begin to dominate the wavefield – a transition that typically occurs around 100 km in cratonic settings [6], consistent with the distance range shown in Fig. 10.

Therefore, direct application of CEUS-calibrated GMPEs to intraplate settings in Brazil introduces systematic bias because the underlying attenuation assumptions may not adequately represent wave propagation in highly efficient cratonic lithosphere. Although both regions are classified as "stable continents", the observed performance differences suggest that finer-scale regionalization accounting for different lithospheric properties and predominant seismogenesis may improve GMPE applicability.

The distribution of normalized residuals further highlights these trends. Kernel density estimates, which provide a smooth continuous approximation of the empirical residual distribution without assuming a fixed parametric form, are shown in Fig. 11 and indicate that most GMPEs produce residual distributions shifted toward negative values, confirming a systematic overprediction of PGA. The USGS_13 model displays a narrower distribution and a mean value closer to zero, indicating both improved accuracy and reduced dispersion. In contrast, models such as USGS_10 present broader distributions with pronounced negative shifts, reflecting poorer statistical compatibility with the Brazilian dataset.

Subgroup analyses provide additional insight into the physical origin of model discrepancies. To illustrate contrasting behavior, Fig. 12 presents the residuals as a function of rupture distance for two representative models: USGS_13, which achieved the lowest overall RMS among all 17 models, and USGS_10, which exhibited the highest.

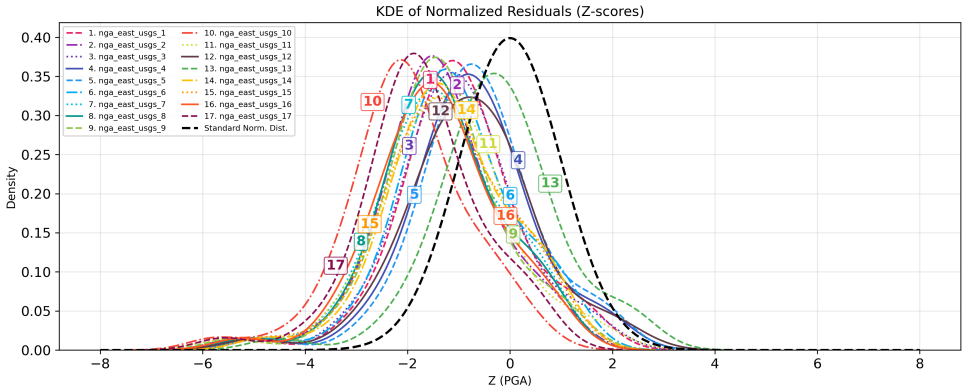


Fig. 11 Kernel density estimation (KDE) of normalized residuals (Z-scores) for the 17 NGA-East USGS models evaluated against the Brazilian PGA dataset. Each colored curve represents one GMPE, identified in the legend. The dashed black curve denotes the standard normal distribution N(0,1), which represents the theoretical target for a perfectly calibrated model.

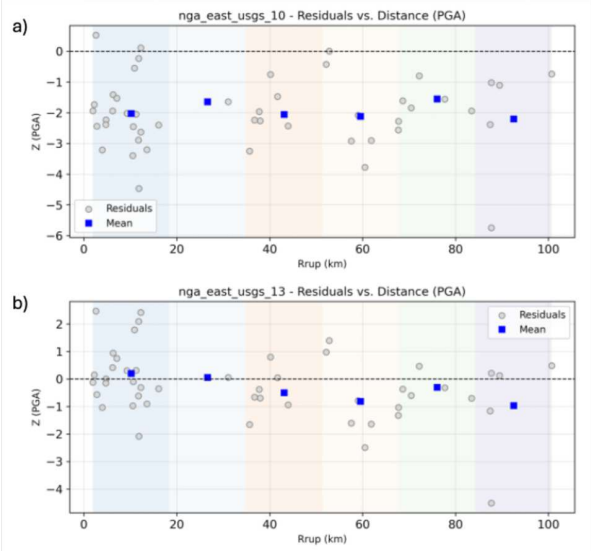


Fig. 12 Normalized residuals (Z-scores) as a function of rupture distance for (a) USGS_10 and (b) USGS_13. Gray circles represent individual residuals, while blue squares indicate mean values within distance bins. Background shading shows the bin ranges used in the subgroup mean RMS value.

The persistent negative bias observed for USGS_10 across nearly all distance ranges indicates a systematic mismatch in distance attenuation behavior, consistent with its poor global ranking. Conversely, USGS_13 maintains near-zero mean residuals and lower dispersion across most bins, suggesting greater robustness over the parameter space. Similar patterns were observed for the magnitude and Vs30 subgroup analyses, reinforcing the conclusion that the superior performance of USGS_13 is not restricted

to a specific subset of the data but reflects a fundamentally better fit to the Brazilian attenuation and source characteristics.

The integration of multiple performance metrics is summarized in Fig. 13, which presents the normalized performance matrix as a heatmap. Lighter colors indicate better relative performance. This visualization confirms the consistency of results obtained from individual metrics, highlighting USGS_13 as the most balanced model across RMS, LLH, and stochastic area measures. Importantly, no single GMPE dominates all metrics simultaneously, supporting the use of weighted logic-tree approaches rather than exclusive model selection.

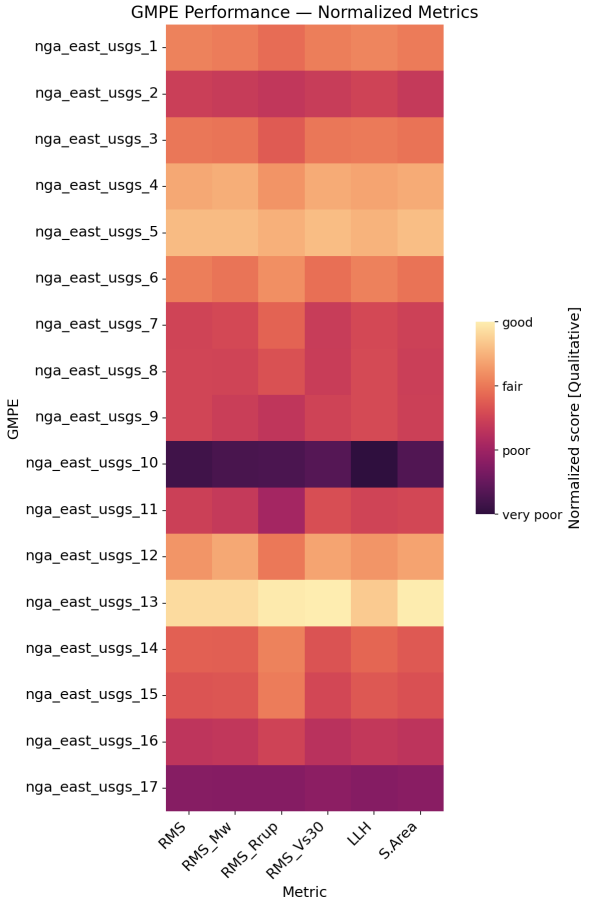


Fig. 13 Heatmap of normalized performance metrics for the USGS NGA-East GMPE suite. Each cell represents the relative performance of a model for a specific metric after min–max normalization. Lighter colors indicate better agreement with Brazilian observations.

The final GMPE weights (Fig. 14) were based on the composite scores derived directly from the normalized performance matrix shown in Fig. 13, following the multicriteria procedure described in Section 3.4. The resulting weights reflect discernible contrasts between models, ranging from 0.038 (USGS_10) to 0.075 (USGS_13), consistent with their relative performance in the multicriteria evaluation. This spread is a direct consequence of rescaling composite scores to the interval [0.5, 1.0] prior to normalization, which preserves score contrasts while ensuring that no model receives a negligible weight. USGS_13 receives the highest weight, whereas USGS_10 is consistently assigned the lowest value, consistent with the heatmap in Fig. 13.

From a broader perspective, the systematic overestimation of PGA by most NGA-East models indicates that the direct transfer of CEUS-calibrated GMPEs to Brazil should be performed with caution. While differences in crustal attenuation properties, stress-drop characteristics, and focal-depth distributions have been identified as likely sources of bias, the relative contribution of each factor remains difficult to isolate with the current dataset. Future studies combining spectral analysis with path specific attenuation inversions could help disentangle these effects and inform the development of region-specific adjustment factors.

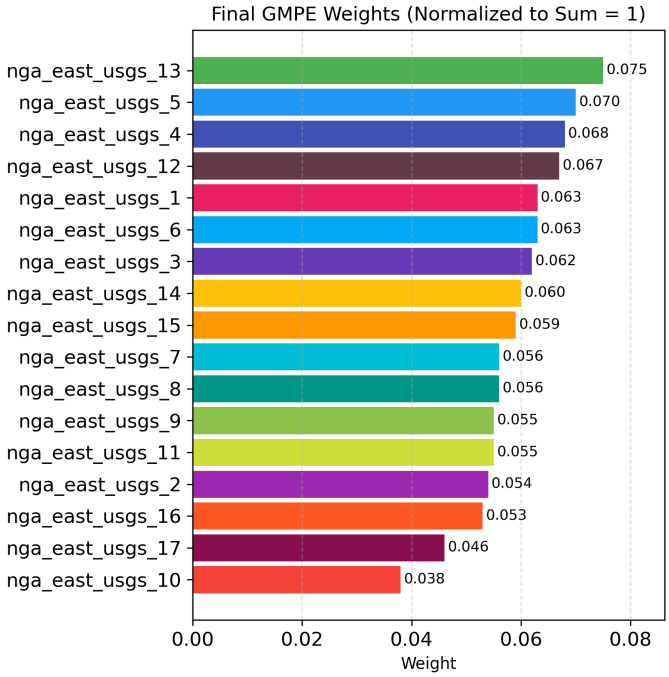


Fig. 14 Final epistemic weights assigned to USGS NGA-East GMPE based on composite multicriteria performance scores. Bars represent normalized weights used for logic-tree implementation in PSHA applications.

The results indicate that Brazilian seismic stations are predominantly located on sites with Vs30 values near the NEHRP B/C boundary (mean Vs30 $\approx$ 760 m/s), which differs substantially from the hard-rock reference conditions assumed by NGA-East models (Vs30 = 3000 m/s). Previous Brazilian PSHA studies often assumed generic rock site conditions in the absence of direct measurements, a simplification that may introduce systematic bias into hazard estimates. The site characterization data compiled here allow GMPE predictions to be evaluated under more realistic site conditions, which is particularly relevant given that NGA-East models require explicit site-response corrections when applied outside their hard-rock reference frame.

These results also suggest that model selection and weighting should not rely exclusively on statistical performance metrics. Adequate coverage of the observed ground-motion parameter space in magnitude, distance, and site condition – ensuring that the selected ensemble envelopes the full range of observed motions – is an additional factor that should be considered alongside goodness-of-fit measures. To illustrate the practical implications of the weighting approach described above, a case-study application for Southeastern Brazil is presented in the following section.

5 Case Study: GMPE Selection for Southeastern Brazil

The Southeastern Brazil region concentrates a large fraction of the country's critical infrastructure, including mining complexes, tailings dams, urban centers, and energy facilities. Consequently, this region represents a priority target for seismic hazard assessment and practical PSHA applications. The case study presented here demonstrates how the GMPE evaluation framework developed in this study can be translated into a consistent model selection strategy for real hazard calculations.

Rather than selecting a single "best" GMPE based exclusively on performance rankings, the adopted strategy follows established PSHA practice, in which a limited set of models is selected to collectively represent epistemic uncertainty and to capture the center, body, and range of expected ground motions in the target region ([17]; [18]). In this context, the selected GMPE ensemble must adequately "envelope" the cloud of observed Brazilian PGA values, reproducing both the upper and lower bounds of recorded motions. By adopting an ensemble-based strategy, hazard estimates become less sensitive to the behavior of any single model and more representative of the plausible range of ground-motion predictions.

As shown in previous sections, most USGS NGA-East models systematically predict higher PGA values relative to Brazilian observations, particularly at short rupture distances. Although the USGS_13 GMPE exhibits the best overall statistical performance in terms of RMS, LLH, and stochastic area metrics, relying exclusively on models that predominantly overestimate PGA would introduce a systematic positive bias into hazard calculations. This observation reinforces the need to complement performance-based ranking with physical and observational coverage criteria.

The NGA-East model proposed by Graizer [19] was incorporated into the GMPE ensemble to ensure representation of the lower bound of observed ground motions.

Unlike the USGS NGA-East models, which were calibrated primarily for the Central and Eastern United States (CEUS), the Graizer model was developed for the broader Central and Eastern North America (CENA) region, encompassing a wider range of crustal and attenuation conditions. This distinction is relevant in the Brazilian context, as CENA-based models may better capture the variability of intraplate ground-motion characteristics beyond the specific tectonic setting of the CEUS. In practice, the Graizer model systematically produces lower PGA predictions than the USGS models and occupies the lower envelope of the Brazilian ground-motion data cloud (Fig. 15), providing a useful counterbalance to the systematic overestimation observed in the remainder of the ensemble.

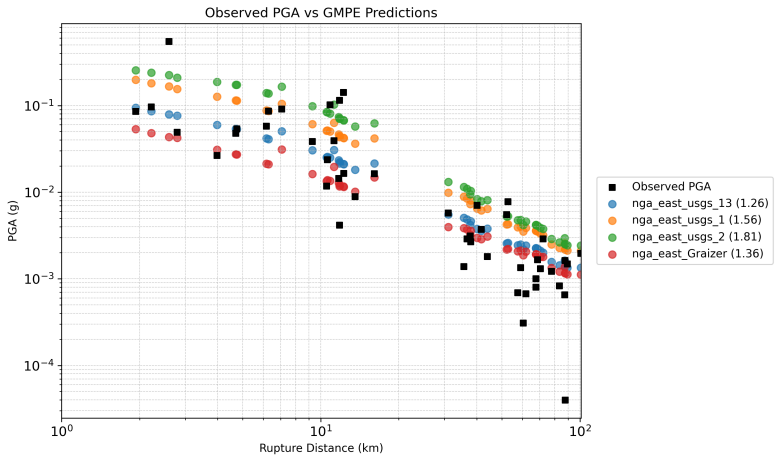


Fig. 15 Comparison between observed PGA values (black points) and NGA-East GMPE predictions (colored symbols) as a function of rupture distance. Each model is identified by color and annotated with the corresponding RMS values indicated in parentheses in the legend. USGS NGA-East models define the upper bound of predicted ground motions, while the Graizer NGA-East model defines the lower bound, collectively enveloping the observed data cloud.

The inclusion of the Graizer model allows the GMPE ensemble to explicitly represent alternative attenuation behaviors and to better capture epistemic uncertainty associated with model formulation. As a result, the final GMPE ensemble used in the Southeastern Brazil case study consists of four models: three NGA-East GMPEs from the USGS models, representing the upper bound of predicted ground motions, and the NGA-East Graizer GMPE, representing the lower bound.

Fig. 16 illustrates the performance of the selected four GMPEs relative to Brazilian PGA observations. While the USGS models tend to overestimate observed values, the Graizer model generally underestimates them. Together, these models collectively span the observed Brazilian PGA range and provide a balanced representation of both overpredicting and underpredicting tendencies, leading to a more robust and defensible hazard assessment framework.

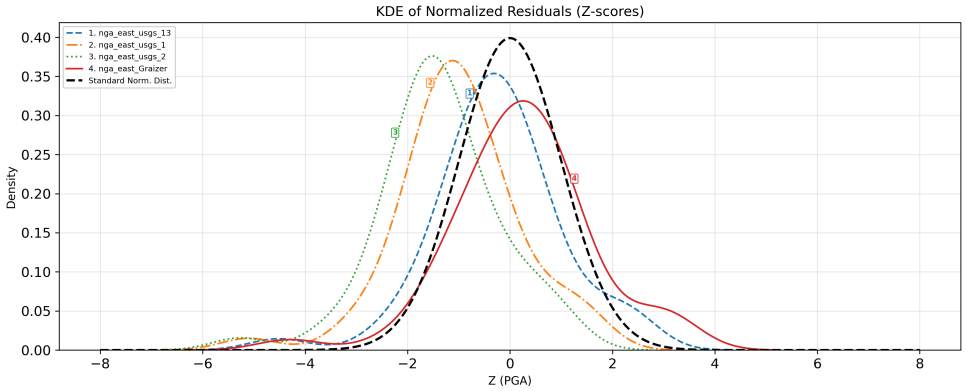


Fig. 16 Kernel density estimates of normalized residual (Z-score) distributions for the final selected NGA-East GMPE applied to the PSHA study in Southeastern Brazil. The dashed black curve denotes the standard normal distribution N(0,1), which represents the theoretical target for a perfectly calibrated model. All colors maintain the relation to GMPE as shown in Fig. 15.

The overall performance of the selected four-GMPE ensemble is summarized in Fig. 17. The combined results confirm that the selected models provide complementary behavior across the observed ground-motion range: while the USGS GMPEs dominate the upper portion of the PGA distribution, the NGA-East Graizer model consistently occupies the lower bound.

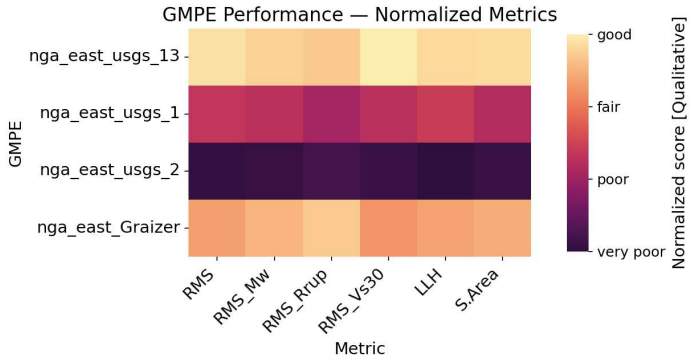


Fig. 17 Heatmap illustrating the performance metrics of the final selected GMPE ensemble for the Southeastern Brazil case study. Each cell represents the relative performance of a model for a specific metric after min–max normalization. Lighter colors indicate better agreement with Brazilian observations.

The objective performance ranking derived in the previous section is preserved within the reduced GMPE set, with USGS_13 remaining the best-performing model. However, instead of discarding lower-performing models entirely, their controlled inclusion allows epistemic uncertainty related to alternative attenuation behaviors to be explicitly propagated into the hazard calculations.

The final epistemic weights assigned to the selected GMPE ensemble are shown in Fig. 18. These weights were derived from the composite performance scores obtained through the multicriteria evaluation framework described earlier.

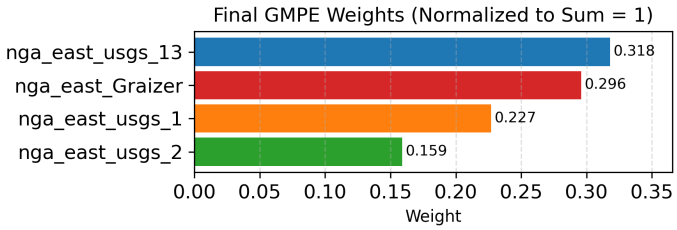


Fig. 18 Final epistemic weights assigned to the selected NGA-East GMPE ensemble for the Southeastern Brazil PSHA application. Weights were derived from normalized composite performance scores and represent the relative contribution of each GMPE within the logic-tree framework.

As expected, the USGS_13 model received the highest weight (w = 0.318), reflecting its superior statistical agreement with Brazilian observations. The NGA-East Graizer model was assigned the second-highest weight (w = 0.296), despite systematically producing lower PGA values than the USGS models. This outcome is intentional: the Graizer model provides essential lower-bound coverage of the observed ground-motion space, and its exclusion or downweighting would bias the ensemble toward systematic overestimation. The remaining USGS models received lower weights (USGS_1: w = 0.227; USGS_2: w = 0.159). The resulting weight distribution reflects discernible contrasts between models, with values ranging from 0.159 to 0.318, following the multicriteria weighting procedure described in Section 3.4.

The impact of the proposed GMPE selection and weighting strategy on seismic hazard estimates is illustrated in Figs. 19 and 20, which compare the PGA hazard maps obtained in this study (BR26) with previously published regional and global PSHA models. This comparison serves two purposes: (1) it demonstrates the practical impact of GMPE selection strategies on final hazard products, and (2) it provides external validation through consistency checks against established global models.

The SBGf2016 map was the first nationwide PSHA attempt for Brazil presented by [5]. That study relied only on the GMPE of [20], without performance-based validation against Brazilian strong-motion data. It also used a low Min = 3.0 for the hazard integration and b-values mostly below 1.0, which over-estimates the hazard. The spatially heterogeneous patterns and localized high-PGA zones (such as USGS2018 [21] and GEM2023 [22]) lack physical justification and reflect the use of only magnitude 5 and above to define seismicity rates.

The USGS2018 model [21] and the GEM2023 model [22] both employ smoothed seismicity approaches but calibrate their occurrence rates using only events with Mw $\geq$ 5.0. In the Brazilian intraplate context, where earthquakes of this magnitude are rare and spatially sparse, this choice leads to poorly constrained rate. The lower-magnitude seismicity (Mw 3–4), which is comparatively well-recorded in Brazil and provides the most reliable spatial information about active seismogenic structures, is effectively excluded from the rate estimation process. It should be noted that GEM2023 also

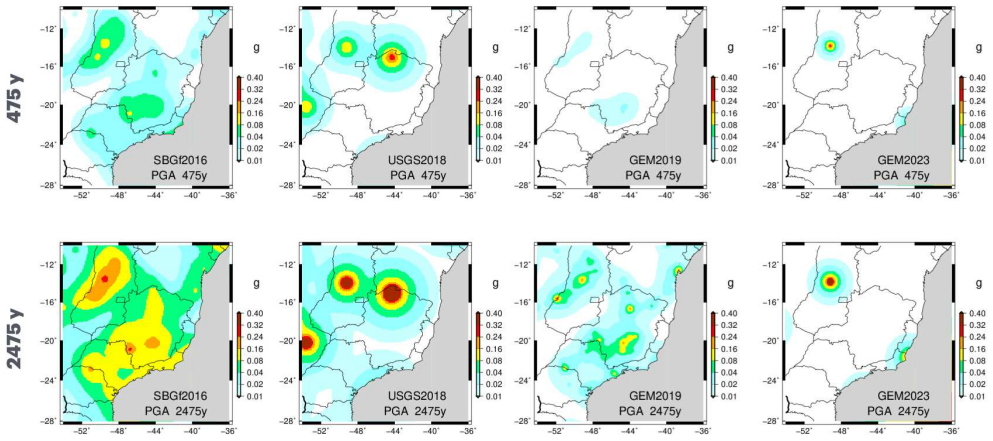


Fig. 19 Comparison of PGA seismic hazard maps for Southeastern Brazil obtained from previous regional and global PSHA models. The upper panels correspond to a return period of 475 years, while the lower panels represent a return period of 2475 years. The maps include results from the first nationwide Brazilian PSHA (SBGF2016; [5]), the global USGS2018 [21], the global GEM2019 [23] and GEM2023 [22] hazard models

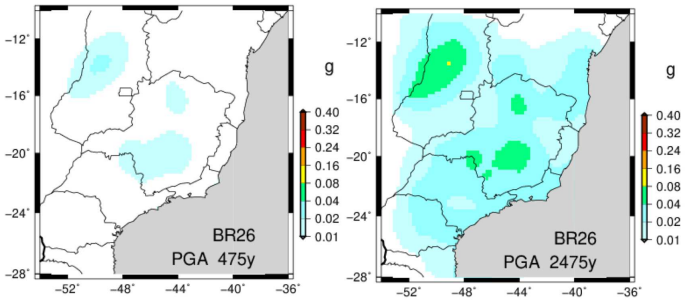


Fig. 20 PGA seismic hazard maps for Southeastern Brazil obtained in the present study (BR26) for return periods of 475 years (left) and 2475 years (right). The maps were computed using the four-GMPE ensemble selected and weighted through the multicriteria framework described in Section 3 combined with a curated regional earthquake catalogue and seismic zones with additional smoothed-seismicity source model approach.

incorporated seismic source zone models alongside the smoothed seismicity component, but the magnitude threshold applied to the catalog constrains the spatial resolution of the resulting hazard in low-seismicity regions such as Brazil.

In contrast, the GEM2019 model [23] incorporated smaller-magnitude events into its smoothed seismicity estimation, resulting in spatially more continuous hazard patterns that better capture the diffuse seismicity characteristic of stable continental regions. This methodological difference explains the closer spatial agreement observed between GEM2019 and the present study (BR26): both approaches draw on the full magnitude range of the regional earthquake catalogue to constrain seismicity rates,

producing smoother spatial gradients and more physically consistent hazard estimates for the Brazilian intraplate environment.

The hazard maps produced in this study (BR26, Fig. 20) are computed for a reference site condition of Vs30 = 760 m/s (NEHRP B/C boundary), consistent with the USGS2018 and GEM2019/23 models and allowing direct visual comparison between all four products. The seismicity model underlying BR26 combines smoothed seismicity kernels following both Frankel [24] and Woo [25] approaches, applied within 10 seismic source zones, each with its own b-value estimated by the maximum likelihood method of Weichert [26]. The resulting hazard patterns are spatially more homogeneous than those of SBGF2016, with reduced extreme PGA values and smoother spatial gradients. The spatial homogeneity reflects primarily the inclusion of lower-magnitude seismicity in the rate estimation, which provides a more complete representation of the diffuse Brazilian seismicity; the reduction in absolute PGA levels reflects, in addition, the difference in reference site conditions (Vs30 = 760 m/s versus $\approx$ 3000 m/s in SBGF2016) and the mitigation of systematic overestimation achieved through the empirically weighted GMPE ensemble.

The closer spatial agreement between BR26 and GEM2019, relative to USGS2018 and GEM2023, is consistent with the interpretation that the inclusion of lower-magnitude seismicity in the rate estimation process leads to more physically representative hazard patterns in low-seismicity intraplate regions such as Brazil.

6 Conclusions

This study presents a comprehensive empirical evaluation of USGS NGA-East GMPEs using Brazilian strong-motion observations and site characterization data. The results demonstrate a systematic tendency of most NGA-East models to overestimate peak ground acceleration in Brazil, reflecting limitations in the direct transfer of CEUS-calibrated GMPEs to the Brazilian intraplate environment. Nevertheless, substantial performance differences were observed among individual models, with USGS_13 exhibiting the best overall statistical agreement and the NGA-East Graizer model providing complementary lower-bound predictions.

Quantitatively, the USGS_13 model demonstrated superior performance with the lowest global RMS (1.26), lowest log-likelihood (2.47), and smallest stochastic area (0.16) among the 17 evaluated NGA-East USGS models. However, the systematic negative bias (mean Z-score = -0.249) across the USGS suite necessitated inclusion of alternative models for a balanced hazard estimation.

The application of multiple performance metrics (including RMS, subgroup-based RMS, log-likelihood, and stochastic area) combined within a multicriteria normalization framework, proved effective for deriving objective epistemic weights for PSHA logic trees. This approach allows GMPE selection to explicitly incorporate both statistical performance and physical uncertainty representation.

The Southeastern Brazil case study demonstrated the practical implications of the proposed methodology. The resulting hazard maps (BR26) show more spatially homogeneous patterns compared to previous Brazilian PSHA results. It should be emphasized, however, that this spatial homogeneity is primarily a consequence of

including lower-magnitude seismicity in the activity rate estimation (an approach consistent with GEM2019 [23]) rather than of the GMPE selection and weighting strategy itself. The GMPE ensemble contributes independently to reducing systematic overestimation of hazard levels, a distinct and complementary improvement.

The comparison with USGS2018 [21] and GEM2023 [22] – both of which also use CEUS-based GMPEs, as does BR26 – isolates the effect of the seismicity rate model more directly: the spatially heterogeneous and locally elevated hazard patterns in those models are due to the use of Mw $\geq$ 5.0 minimum-magnitude thresholds in the smoothed seismicity rates, which excludes the moderate and small earthquakes that better represent Brazilian seismicity. The closer agreement between BR26 and GEM2019, which incorporates smaller-magnitude events, is consistent with this interpretation and underscores the importance of accounting for the full magnitude range of the regional catalogue in PSHA for stable continental regions.

As part of this work, the compiled and quality-controlled dataset of peak ground accelerations recorded at Brazilian stations (together with the associated site characterization parameters, vs30) is being made publicly available. This dataset represents the most comprehensive collection of strong-motion observations with measured site conditions for the Brazilian territory. Its availability is intended to support future GMPE calibration, validation, and development efforts specifically targeted at the South American intraplate environment, thereby contributing to reducing the current dependence on foreign models calibrated for other tectonic settings.

More broadly, the proposed GMPE evaluation and weighting framework can serve as a methodological basis for future PSHA studies in Brazil and analogous low-seismicity regions worldwide. By providing a transparent, data-driven, and reproducible strategy for model selection and uncertainty treatment, this approach contributes to improving the robustness and defensibility of seismic hazard assessments where local ground-motion data remain scarce.

Limitations

Despite the validation framework presented here, several limitations warrant acknowledgment. First, the Brazilian strong-motion database, while representing the most comprehensive dataset currently available, remains limited in coverage of large-magnitude events and near-source distances ($<$ 10 km). Second, the validation focused exclusively on PGA; extension to spectral accelerations across multiple periods would strengthen GMPE selection for engineering applications. Third, regional variations in crustal attenuation properties across Brazil's diverse geological provinces remain incompletely characterized.

Data and Code Availability

The Brazilian strong-motion PGA database compiled in this study, comprising 150 records from 27 regional earthquakes with associated site characterization data (Vs30), is publicly available on Zenodo at https://doi.org/10.5281/zenodo.20287122 [27]. The repository includes the complete dataset in CSV format (`brazil_pga_data.csv`), a companion file describing all data fields (`brazil_pga_data_description.txt`), and the Python script used for GMPE evaluation and weight computation

(`evaluate_gmpes.py`). The seismic hazard maps (BR26) were computed using OpenQuake Engine [10], an open-source platform available at https://github.com/gem/oq-engine.

Acknowledgments

We thank National Observatory, University of Brasilia and University of Rio Grande do Norte for facilitating the MASW surveys in some of their stations. The MASW surveys were carried out by AFC Geofísica. The joint inversion of MASW dispersion and HVSR were carried our both by the authors and by Alta Resolução Geofísica.

References

[1] Almeida, A.A.D.d., Assumpção, M., Bommer, J.J., Drouet, S., Riccomini, C., Prates, C.L.M.: Probabilistic seismic hazard analysis for a nuclear power plant site in southeast Brazil. Journal of Seismology **23**, 1273–1301 (2019) https://doi.org/10.1007/s10950-018-9755-8

[2] Borges, R.G., Assumpção, M.S., Almeida, M.C.F., Almeida, M.D.: Seismicity and seismic hazard in the continental margin of southeastern Brazil. Journal of Seismology **24**, 1205–1224 (2020) https://doi.org/10.1007/s10950-020-09941-4

[3] Fonsêca, J.A.S., Lasocki, S., Nascimento, A.F.: Probabilistic estimation of the source component of seismic hazard in North-Eastern Brazil. Heliyon **10**, 30716 (2024) https://doi.org/10.1016/j.heliyon.2024.e30716

[4] Pereira, E.M.V., Assumpção, M.S.d., Goda, K., Siqueira, G.H.: Advancing probabilistic seismic hazard assessment for Northeastern Brazil: Toward a comprehensive treatment of epistemic uncertainties. Earthquake Spectra **42**(1) (2026) https://doi.org/10.1002/esp4.70044

[5] Assumpção, M., et al.: Terremotos no Brasil: preparando-se para eventos raros. Boletim SBGf (96), 25–29 (2016)

[6] Atkinson, G.M., Mereu, R.F.: The shape of ground motion attenuation curves in southeastern Canada. Bulletin of the Seismological Society of America **82**(5), 2014–2031 (1992) https://doi.org/10.1785/BSSA0820052014

[7] Wathelet, M., Chatelain, J.-L., Cornou, C., Di Giulio, G., Guillier, B., Ohrnberger, M., Savvaidis, A.: Geopsy: a user-friendly open-source tool set for ambient vibration processing. Seismological Research Letters **91**(3), 1878–1889 (2020) https://doi.org/10.1785/0220190360

[8] Boore, D.M., Joyner, W.B.: Site amplifications for generic rock sites. Bulletin of the Seismological Society of America **87**(2), 327–341 (1997) https://doi.org/10.1785/BSSA0870020327

[9] Stewart, J.P., Parker, G.A., Harmon, J.A., Atkinson, G.M., Boore, D.M., Darragh, R.B., Silva, W.J., Hashash, Y.M.A.: Expert panel recommendations for ergodic site amplification in central and eastern North America. Earthquake Spectra **36**(1), 42–73 (2020) https://doi.org/10.1177/8755293019878190

[10] Pagani, M., Weatherill, G., Monelli, D.: OpenQuake Engine: An open hazard (and risk) software for the Global Earthquake Model. Seismological Research Letters **85**(3), 692–702 (2019) https://doi.org/10.1785/0220130087

[11] Hashash, Y.M.A., Ilhan, O., Harmon, J.A., Parker, G.A., Stewart, J.P., Rathje, E.M., Campbell, K.W., Silva, W.J.: Nonlinear site amplification model for ergodic

seismic hazard analysis in Central and Eastern North America. Earthquake Spectra **36**(1), 69–109 (2020) https://doi.org/10.1177/8755293019878193

[12] Goulet, C.A., Frankel, A.D., Atkinson, G.M., *et al.*: NGA-East ground-motion characterization model for the Central and Eastern United States. Earthquake Spectra **37**(2), 527–553 (2021) https://doi.org/10.1177/8755293020980570

[13] Scherbaum, F., Cotton, F., Smit, P.: On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic-hazard analysis in regions of moderate seismicity: The case of rock motion. Bulletin of the Seismological Society of America **99**(6), 3234–3247 (2009) https://doi.org/10.1785/0120090116

[14] Sunny, R.C., Kale, Ö., Cotton, F., Stromeyer, D.: A new model-independent method for comparing observed and predicted ground-motion residual distributions: The stochastic area metric. Bulletin of the Seismological Society of America **111**(2), 617–632 (2021) https://doi.org/10.1785/0120200239

[15] Ciardelli, C., Assumpção, M.: Rupture lengths of intraplate earthquakes in Brazil determined by relative location of aftershocks: Evidence for depth dependence of stress drops. Journal of South American Earth Sciences **89**, 246–258 (2019) https://doi.org/10.1016/j.jsames.2018.11.017

[16] Atkinson, G.M., Boore, D.M.: Earthquake ground-motion prediction equations for eastern North America. Bulletin of the Seismological Society of America **96**(6), 2181–2205 (2006) https://doi.org/10.1785/0120050245

[17] Cotton, F., Scherbaum, F., Bommer, J.J., Bungum, H.: Criteria for selecting and adjusting ground-motion models for specific target regions: Application to Central Europe. Bulletin of Earthquake Engineering **4**(4), 341–362 (2006) https://doi.org/10.1007/s10518-006-9006-7

[18] Delavaud, E., Cotton, F., Akkar, S., Scherbaum, F., Danciu, L., Beauval, C., Drouet, S., Douglas, J., Basili, R., Giardini, D.: Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. Bulletin of Earthquake Engineering **10**(2), 409–439 (2012) https://doi.org/10.1007/s10518-011-9303-2

[19] Graizer, V.: Alternative (G-16v2) ground-motion prediction equations for Central and Eastern North America. Bulletin of the Seismological Society of America **107**(2), 869–886 (2017) https://doi.org/10.1785/0120160212

[20] Silva, W.J., Gregor, N., Darragh, R.: Development of regional hard rock attenuation relations for Central and Eastern North America. Technical report, Pacific Engineering and Analysis, El Cerrito (2002)

[21] Petersen, M.D., Harmsen, S.C., Jaiswal, K.S., Rukstales, K.S., Luco, N., Haller, K.M., Mueller, C.S., Shumway, A.M.: Seismic hazard, risk, and design for South

America. Bulletin of the Seismological Society of America **108**(2), 781–800 (2018) https://doi.org/10.1785/0120170002

[22] Johnson, K., Villani, M., Bayliss, K., Brooks, C., Chandrasekhar, S., Chartier, T., Chen, Y., Garcia-Pelaez, J., Gee, R., Styron, R., Rood, A., Simionato, M., Pagani, M.: Global Earthquake Model (GEM) Seismic Hazard Map (version 2023.1) (2023). https://doi.org/10.5281/zenodo.8409647

[23] Pagani, M., Garcia-Pelaez, J., Gee, R., Johnson, K., Poggi, V., Silva, V., Simionato, M., Styron, R., Viganò, D., Danciu, L., Monelli, D., Weatherill, G.: The 2018 version of the Global Earthquake Model: hazard component. Earthquake Spectra **36**(S1), 226–251 (2020) https://doi.org/10.1177/8755293020931866

[24] Frankel, A.: Mapping seismic hazard in the central and eastern United States. Seismological Research Letters **66**(4), 8–21 (1995) https://doi.org/10.1785/gssrl.66.4.8

[25] Woo, G.: Kernel estimation methods for seismic hazard area source modeling. Bulletin of the Seismological Society of America **86**(2), 353–362 (1996)

[26] Weichert, D.H.: Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. Bulletin of the Seismological Society of America **70**(4), 1337–1346 (1980)

[27] Collaço, B., Assumpção, M., Rocha, M., Assis, L., Simão, M.: Brazilian Strong-Motion PGA Database and NGA-East GMPE Evaluation Framework. Zenodo (2026). https://doi.org/10.5281/zenodo.20287122 . https://doi.org/10.5281/zenodo.20287122

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supp.material.pdf
- supp.material.zip