

Response to Reviewers

Manuscript ESSD-2026-202 — "Ground-Motion Dataset for Shallow Earthquakes in Colombia"

We thank both reviewers for their careful reading of the manuscript and the accompanying dataset, and for their constructive comments.

Reviewer 1

Manuscript text

Reviewer comment: *Lines 18-19: the Fourier Amplitude Spectra for each component is provided in the frequency range from 0.04 to 30 Hz. The data file contains frequencies up to 50 Hz (maybe more considering the discrepancies from frequencies — see comments for COL_FAS). Please clarify the actual frequency range provided in the dataset.*

Response: We thank the reviewer for catching this. The stated range was outdated; FAS and EAS are provided from 0.04 to 50 Hz in the published files. We have corrected this throughout the manuscript (Abstract, Compilation of the Data, and the list of provided IMs) to consistently state 0.04–50 Hz.

Reviewer comment: *Line 105: The predominant period of the site (T_0). Does this correspond to the T_n values reported in the Sites_Vs30_HV file? Please use consistent notation throughout the manuscript and the accompanying datasets.*

Response: Confirmed — T_0 and T_n refer to the same quantity (the predominant site period). We have standardized on T_n throughout both the manuscript text and the Sites_Vs30_HV file to avoid any ambiguity.

Reviewer comment: *Abbreviations: All abbreviations should be defined when first introduced, e.g., SGC — for readers who are not familiar with the study region, the meaning of this abbreviation is not obvious.*

Response: We reviewed the manuscript and now define all abbreviations at first use in the main text, including SGC (Servicio Geológico Colombiano / Colombian Geological Survey) and GMPM (Ground Motion Prediction Model).

COL_FAS

Reviewer comment: *The file contains 9330 components instead of the 8086 reported in the manuscript. Please explain this discrepancy. Are some components duplicated?*

Response: Yes — we traced this to (1) a small number of exact duplicate rows introduced by a data-concatenation step, (2) a subset of records generated before the final quality-control filter applied to the ground-motion (COL_GM) flatfile was finalized. We reprocessed COL_FAS so that it now contains exactly the same (eqid, sta) pairs as COL_GM_journal, with matching record counts in both files.

Reviewer comment: *It is unclear what eqid (col. B) represents. If it is an identifier, it should not be stored in scientific notation.*

Response: eqid is a unique event identifier. It is now stored consistently as a text/string field (with an "EQ_" prefix for numeric, date-derived identifiers) to prevent spreadsheet software from reinterpreting long numeric IDs in scientific notation.

Reviewer comment: *date2 (col. B) contains valid dates only up to row 1693. From that point onward, the entries no longer appear to be dates.*

Response: This was a date-format parsing artifact, not an error in the underlying data: entries after that row included fractional seconds, which some spreadsheet/CSV readers do not parse consistently alongside entries without them. We standardized all date2 values to a single ISO 8601 format (YYYY-MM-DD HH:MM:SS) throughout the file.

Reviewer comment: Columns E–H contain latitude and longitude values. Please clarify whether these coordinates correspond to the earthquake epicenter or to the recording station.

Response: We confirmed and clearly separated the two: station coordinates are now labeled sta_Lat/sta_Lon, and epicenter coordinates are labeled eq_Lat/eq_Lon, consistently in COL_GM and COL_FAS.

Reviewer comment: The frequencies for FAS_NS, FAS_EW and EAS appear to be incorrect. For example, the frequency in column CQ is 18.789 Hz, while the adjacent column CR is 2.006 Hz, and column CX is 20.15 Hz. Please verify the frequency labels.

Response: The frequency values themselves were correct, but the columns had been sorted alphabetically as text rather than numerically, producing the apparent disorder the reviewer noticed. We re-sorted all spectral columns by ascending numerical frequency; for each frequency the column order is now FAS_EW, FAS_NS, FAS_VER, EAS.

Reviewer comment: It would be more convenient to provide a separate row containing only the frequency values instead of including both the parameter name and the frequency in each column header.

Response: We considered this suggestion. We opted to retain the frequency value embedded in each column header (e.g., FAS_EW_1.069), consistent with the convention used by other widely used strong-motion flatfiles (e.g., NGA-West2 FAS files), which allows the file to be parsed directly by standard data-analysis tools (pandas, R data.table, etc.) without requiring a separate frequency-row lookup step. We are, however, happy to reconsider this if the editor believes a separate frequency row would better serve the community.

Reviewer comment: The table headers should contain explicit and self-explanatory parameter names and measurement units.

Response: All column headers in COL_FAS now include explicit units where applicable (e.g., distances in km, coordinates in decimal degrees), and the amplitude spectra are documented in the manuscript as being reported in cm/s.

COL_GM_journal

Reviewer comment: The table headers should contain explicit and self-explanatory parameter names.

Response: Headers were renamed to be self-explanatory, e.g., PGA_ns_g, PGA_ew_g, PGA_v_g, PGA_geomean_g, PGA_rotD50_g, PGA_rotD100_g, PGV_*_cms, PGD_*_cm, and SA_rotD50_g(T).

Reviewer comment: Please clearly distinguish between station and event coordinates in the column headers.

Response: Resolved as above: sta_Lat/sta_Lon for the recording station and eq_Lat/eq_Lon for the epicenter, applied consistently across COL_GM and COL_FAS.

Reviewer comment: Measurement units should be included in column headers.

Response: All intensity measures, distances (rrup, rjb, rhypo, repi, in km), rupture-geometry parameters (width, length, ztor in km; dip, strike, rake in degrees), and V_{S30} (m/s) now carry explicit unit suffixes in their column names.

Reviewer comment: There is no distinction between spectral accelerations (SA) for the EW, NS, and vertical components. Furthermore, SA at 0.01 s should be very close to PGA. This appears to be true only for the NS component (assuming PGA is reported in cm/s^2 and SA in m/s^2). Are spectral accelerations provided only for the NS component? The manuscript states that SA values are available for all components.

Response: We confirm the reviewer's observation: SA values in this dataset are computed as the RotD50 combination of the two horizontal components (Boore, 2010), not per component. We have corrected the manuscript text accordingly and renamed the relevant columns to SA_rotD50_g(T) to make this explicit. PGA is separately reported per component (PGA_ns_g, PGA_ew_g, PGA_v_g) as well as its geometric mean and RotD50/RotD100 combinations.

Sites_Vs30_HV

Reviewer comment: *The table headers should contain explicit and self-explanatory parameter names and measurement units.*

Response: Headers were renamed accordingly: Vs30_ms, Vs30_infered_ms, sta_Lat_deg, sta_Lon_deg, Tn_HVRSR_s, Tn_HVFSR_s, Tn_mHVFSR_s, and P_HVRSR, P_HVFSR, P_mHVFSR (see response on P* below).

Reviewer comment: *I would recommend starting the station numbering from 1 instead of 0.*

Response: We removed the arbitrary numeric row index from the published file entirely; stations are identified solely by their station code (sta), which eliminates any ambiguity related to zero- versus one-based indexing.

Reviewer comment: *The columns lat, lon, lat_y, and lon_y appear to contain similar coordinates. Please clarify the meaning of lat_y and lon_y. Similarly, please explain what is dist_lon and dist_lat.*

Response: lat_y/lon_y were coordinates from an internal cross-referencing/verification step against an external station catalog, and dist_lon/dist_lat were the corresponding coordinate offsets used only for internal quality control. As these are not intended for external use and could confuse users, we removed all four columns from the published file.

Reviewer comment: *Is Tn equivalent to the predominant period T0 referred to in the manuscript? If so, please use a consistent notation throughout the paper and the supplementary files.*

Response: Confirmed and resolved — see response above; Tn is used consistently throughout.

Reviewer comment: *Please define the parameter P*.*

Response: P* is the amplitude of the HVSR curve at the peak corresponding to the identified predominant frequency [HVSR(fn)]. We have added this definition to the manuscript alongside the description of Tn, and cleaned the column names (removing the asterisk, which can cause issues in some CSV readers) to P_HVRSR, P_HVFSR, and P_mHVFSR.

Reviewer 2

General remarks and recommendations

Reviewer comment: *(a) Magnitude conversion uncertainties ($M < 4.5$) — the authors observe significant scatter in the between-event variability for magnitudes below 4.5. This is attributed to the empirical conversions used to transition from mb and ML to M. It would be beneficial if the authors will include a subsection demonstrating how extracting M from FAS corner frequencies compares directly with the SGC & GEM (2018) empirical equations.*

Response: We implemented a moment magnitude (**M**) estimation based on the S-wave corner frequency (fc), estimated by fitting a Brune (1970) source model to the effective Fourier amplitude spectrum (EAS) of each recording. The corner frequency is converted to a rupture radius (Brune, 1970), and the seismic moment is obtained via the Eshelby circular-crack relation under an assumed constant stress drop, then converted to M following Hanks and Kanamori (1979). We tested stress-drop values of 1, 3, and 5 MPa against independently reported M (GCMT and other agency **M** solutions) and selected 5 MPa as the value that best reproduced the reference magnitudes. This corner-frequency-based **M** is now used preferentially over the original SGC & GEM (2018) mb/ML-to-M empirical conversion whenever a reliable agency-reported **M** is unavailable, since it does not rely on regional empirical conversion relationships. δB_e variability was considerably reduced.

Reviewer comment: *(b) Proxy-based VS30 limitations — only 28 of the 227 sites feature direct, in-situ shear-wave velocity measurements; the remaining site parameters rely heavily on topographic-slope inference maps. The*

authors may add a brief paragraph discussing the potential amplification of the epistemic uncertainty this introduces to the site-to-site standard deviation.

Response: We agree and have added a paragraph noting that only 28 of the 227 sites have in-situ, microtremor-based measured, while the remaining sites rely on a topographic-slope-based VS30 proxy (Eraso and Montejo, 2019). We caution that this introduces additional epistemic uncertainty at proxy-based sites, which may inflate the site-to-site standard deviation ($\phi S2S$) estimated in this study.

Reviewer comment: (c) *Tectonic classification ambiguity — the manuscript leaves the exact tectonic classification (e.g., separating shallow crustal faults from intraplate subduction or interface events at shallow depths) down to the end-user's preference. It is advisable to provide even a preliminary tectonic categorization within the flatfile.*

Response: After consideration, we have chosen not to implement a preliminary tectonic categorization within the flatfile at this stage, as robustly separating shallow crustal, intraplate, and shallow interface events in this tectonically complex region would require a dedicated analysis beyond the scope of the present data-description manuscript, and a preliminary/uncertain classification could itself become a source of misuse if adopted uncritically by end users. We instead provide all the event parameters (depth, location, focal mechanism where available) needed for users to apply their own preferred classification scheme, and we note this limitation explicitly in the manuscript.

Targeted remarks

Reviewer comment: *Line 56: Define the acronym for the general readership as ground motion prediction model.*

Response: Done — GMPM is now defined at first use in the Introduction.

Reviewer comment: *Line 68: Define SGC. Not defined yet in the text, just in the affiliation section.*

Response: Done — SGC is now defined at first use in the main text (Compilation of the Data).

Reviewer comment: *Line 69: It must be stated that given these criteria, 694 seismic events are selected.*

Response: We now state explicitly that more than 700 earthquakes met the initial selection criteria, and that after removing volcanic events, records with signal artifacts, and (following this round of revision) a small number of events with unreliable magnitude estimates or coda-contaminated corner-frequency measurements, 667 events and their associated records were retained in the final flatfile.

Reviewer comment: *Line 97: How was the uncertainty (sigma) effect incorporated in the conversion?*

Response: This comment applied to the original empirical mb/ML-to-M conversion, which we have superseded with the corner-frequency-based approach described above. Uncertainty in the new approach is instead addressed through the sensitivity test across stress-drop values (1, 3, and 5 MPa) and validation against independently reported M.

Reviewer comment: *Line 108: The distances must be defined and/or referenced.*

Response: Distance metrics are now explicitly defined and referenced, including the Joyner-Boore distance (RJB) and standard hypocentral/epicentral/rupture distance definitions.

Reviewer comment: *Line 169: The terms of rel. (3) shall be explained.*

Response: The terms of the sigma decomposition equation are now defined immediately after the equation: τ , $\phi S2S$, and $\phi 0$ are the standard deviations of the between-event (δBe), site-to-site ($\delta S2Ss$), and single-station within-event (ϵ) residuals, respectively.

Reviewer comment: *Line 237: The significance of statistical data shall be commented.*

Response: We expanded the discussion of the δBe trends shown in Figure 7, including the divergence in sign between the global and regional GMPM within the M [4.2, 4.7] range and its possible relation to remaining magnitude calibration uncertainty.