This is a review of the preprint paper **"Ground-Motion Dataset for Shallow Earthquakes in Colombia"** by Daniel Martinez-Jaramillo, Sreeram-Reddy Kotha, F. Ramón Zúñiga, Pierre Lacan (2026).

The paper presents the development of a flatfile with uniformly processed, high-quality ground-motions derived from waveforms recorded by the Servicio Geológico Colombiano (SGC) and partner local monitoring arrays. The dataset is intended to fill a critical gap in Northern South America's engineering seismology infrastructure by supporting the calibration of regional Ground-Motion Prediction Models (GMPMs).

The waveforms processing pipeline strictly follows standard international protocols (baseline correction, cosine tapering, linear detrending, and a 2nd-order acausal Butterworth filter). The authors perform residual decomposition to compute total (δ_{total}), between-event (δB_e), and site-to-site ($\delta S2S_s$) residuals. Comparing the data against a classic global model, a region-adapted option, and a FAS model ensures complete structural consistency. Residuals adhere closely to a Gaussian distribution.

The paper is an excellent contribution to earth science data. It provides a dense, clean, and robust ground-motion dataset for a highly complex tectonic zone. The methodology is clear, the figures are high-fidelity, and the verification metrics are statistically sound. There are some minor revisions that are addressed hereinafter.

These are some general remarks and recommendations to improve the quality of the manuscript:

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 m_b and M_L 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.
b. proxy-based V_{S30} limitations - only 28 of the 227 sites feature direct, in-situ shear wave velocity measurements; the remaining site parameters rely heavily on typographic-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.
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. This would increase its usability for practitioners looking to test specific path attenuation characteristics.

Moreover, here are some targeted remarks on the manuscript:

1. Line 56: Define the acronym for the general readership as ground motion prediction model
2. Line 68: Define SGC. Not defined yet in the text, just in the affiliation section.
3. Line 69: It must be stated that given these criteria, 694 seismic events are selected.
4. Line 97: How was the uncertainty (sigma) effect incorporated in the conversion?
5. Line 108: The distances must be defined and/or referenced.
6. Line 169: The terms of rel. (3) shall be explained.
7. Line 237: The significance of statistical data shall be commented.