Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-2197

A flatfile for processed seismic records in South Korea using an automated protocol

Bongseok Ryu, Inhyeok Choi
Dept. of Civil, Environmental and System Engineering, Hanyang University, Seoul, South Korea

Dongyoup Kwak
Dept. of Civil and Environmental Engineering, Hanyang University ERICA, Ansan, South Korea, dkwak@hanyang.ac.kr

Jae-Kwang Ahn
Dept. of Fire & Emergency Management, Kangwon National University, Samcheok, South Korea

ABSTRACT: South Korea, classified as a low-to-moderate seismic region, has experienced significant social and economic damage from recent earthquakes, including the 2016 Gyeongju (M_L 5.8), the 2017 Pohang (M_L 5.4), and the 2024 Buan earthquake (M_L 4.8). A flatfile is a strong motion database supplying a crucial foundation for evaluating seismic hazards and developing reliable ground motion prediction equations (GMPEs). This study proposes an automated protocol for seismic wave preprocessing and classification of anomalous seismic waves, aiming to construct a region-specific flatfile tailored to South Korea. The flatfile includes information on earthquake events, seismic stations, signal processing, and intensity measures (IMs) such as PGA, PGV, and SA. We conducted the research steps as follows: 1) collect seismic records from the Korean National Seismic Network for local earthquake events with $M_L \geq$ 2.5 between 2000 and 2024, 2) perform seismic pre-processing and instrument response removal to remove noise in the signal and enhance the quality of the data, 3) classify anomalous seismic records based on the signal-to-noise ratio (SNR), and 4) create the flatfile for the IMs of each processed record. The generated flatfile is expected to serve as a high-quality and reliable dataset, contributing to future research on seismic characteristics in South Korea.

KEYWORDS: Flatfile, Intensity measures, Automated protocol, Seismic wave processing, Anomalous seismic wave classification

1 INTRODUCTION

South Korea is generally considered a low-to-moderate seismic region; however, recent damaging events— the 2016 Gyeongju (M_L 5.8), the 2017 Pohang (M_L 5.4), and the 2024 Buan earthquake (M_L 4.8)—highlight the necessity for enhanced hazard characterization and ground-motion modeling on the Korean Peninsula. A flatfile is a standardized strong-motion database that integrates event and station metadata, processing parameters, and intensity measures (IMs). The flatfile serves as the fundamental resource for GMPE development, site-effect analyses, and hazard assessments (Ancheta et al., 2014; Boore & Bommer, 2005). International efforts such as NGA-West2, NGA-East, the European Strong-Motion (ESM) flatfile, and automated processing of KiK-net records demonstrate that consistency and automation are important for reliability and reuse (Ancheta et al., 2014; Dawood et al., 2016).

In South Korea, the expanding station networks, mainly operated by the Korea Meteorological Administration (KMA) and the Korea Institute of Geoscience and Mineral Resources (KIGAM), are producing rapidly growing volumes of strong-motion records, making the manual processing impractical. This study develops an automated preprocessing and quality-assurance pipeline, implements rule-based screening for anomalous records, and creates a Korean-specific flatfile to support regional ground-motion research.

2 DATA COLLECTION

2.1 *Earthquake events*

For the flatfile construction, we collected earthquake events with local magnitude (M_L) ≥ 2.5 that occurred on the Korean Peninsula between 2000 and 2024. Event information, including origin time, epicentral coordinates, focal depth, and magnitude, was obtained from the National Earthquake Comprehensive Information System (NECIS) provided by the KMA. Figure 1 shows the distribution of these events. The size of the circle corresponds to the event magnitude, and its color indicates the magnitude range.

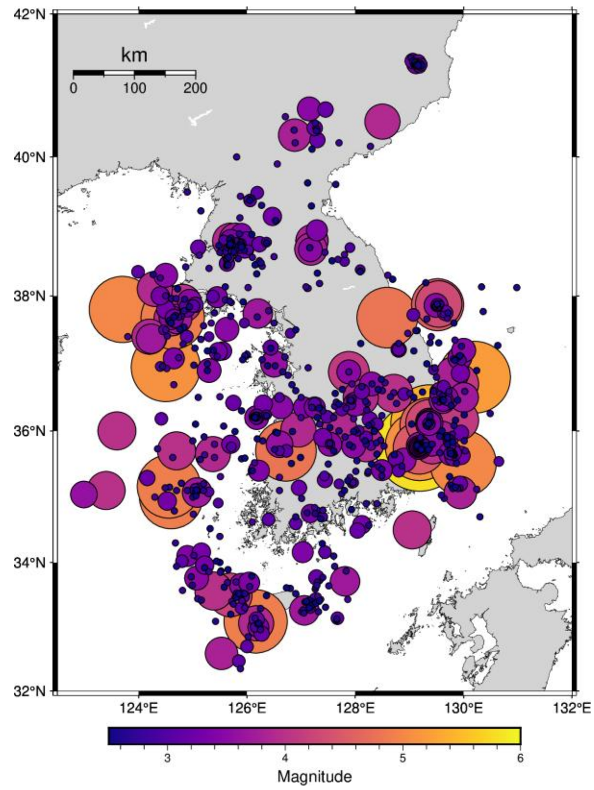

Figure 1. Epicentral distribution of 832 earthquakes ($M_L \geq 2.5$) across the Korean Peninsula during 2000–2024. Circle size is proportional to magnitude; color denotes the magnitude range.

2.2 *Stations and seismic records*

Seismic waveforms and station metadata were obtained from two national repositories: NECIS for the KS network, and the KIGAM Quake (Lee at al. 2022) for the KG network. For each target event, all available three-component recorded data were downloaded in miniSEED format together with the associated station metadata. Figure 2 shows the current distribution of seismic stations in South Korea, comprising the KS and KG network. As of August 2025, the network comprises 364 KS stations operated by the KMA and 61 KG stations operated by the KIGAM, totaling 425 stations. Symbols show individual station locations, and colors represent the hosting network.

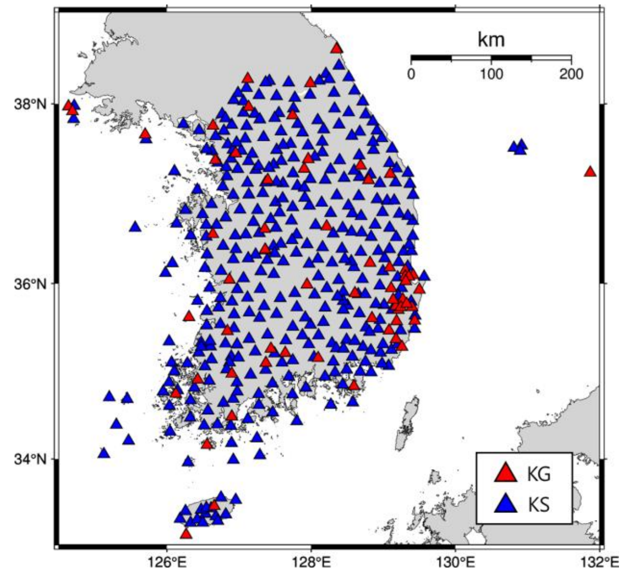

Figure 2. Distribution of seismic stations operated by KMA (KS, 364 stations) and KIGAM (KG, 61 stations) as of August 2025.

3 PREPROCESSING OF SEISMIC RECORDS

Seismic records from the KMA and KIGAM networks were provided in miniSEED format, accompanied by sensor and station metadata. To preprocess the waveforms, the following procedures were applied: instrument response correction, noise removal, and baseline correction.

3.1 *Instrument correction*

Raw miniSEED data are recorded in digital counts and include instrument-related noise, so instrument-response correction is required. For each station, instrument responses were removed using metadata from NECIS (KS) and KIGAM Quake (KG) to convert raw counts to physical units. For KMA stations commissioned before July 2023, instrument responses were taken from the StationXML files compiled by Ahn et al. (2024).

3.2 *Band-pass filtering for noise removal*

An acausal Butterworth band-pass filter was applied to suppress low-frequency drift and high-frequency noise while preserving usable signal content. Corner frequencies were determined by comparing signal-to-noise ratios (SNR) across frequencies. A pre-event noise window and a signal window were used to compute the frequency-dependent SNR. Low corner frequency was set at the first frequency where SNR reaches 3, and the high corner frequency at the last frequency where SNR remains bigger than 3. This procedure removes contaminating bands by noise and stabilizes the waveforms, Fourier amplitude spectra (FAS), and 5%-damped response spectra (SA).

3.3 *Baseline correction*

After integration to displacement, residual drift was removed by fitting a 6th-order polynomial trend excluding the constant and linear terms and subtracting it from the displacement record. This process yields a stable trace without spurious permanent offsets.

4 CLASSIFICATION OF ANOMALOUS RECORDS

An automated and rule-based quality control was applied to each three-component record. We evaluated ten criteria addressing spectral sanity in terms of FAS, passband suitability, displacement stability, edge spikes, and sequence-interference flags. A record was classified as NG (anomalous) if any one of the criteria was violated; otherwise, it was labeled OK (normal). Table 1 defines the ten NG rules (NG1–NG10) in one line each and serves as the reference for the flags.

Table 1. Anomalous-record classification rules

Rule number	Condition
NG1	Peak value of signal FAS for frequencies > 3 Hz is zero
NG2	At the spectral peak > 3 Hz, signal FAS ≤ 3 × noise FAS
NG3	fcHP ≥ 3 Hz or fcLP ≤ 25 Hz
NG4	Start displacement: max |disp| in the first 1% of record > 0.5 × PGD.
NG5	End displacement: max |disp| in the last 10% of record > 0.3 × PGD.
NG6	Spikes near edges: |acc| in first 1% or last 30% of record > 0.5 × PGA.
NG7	Negative FAS slope over the interval fcHP < freq. < freq. at max(FAS)
NG8	Negative FAS slope over 0.1–3 Hz
NG9	Overlap with the preceding earthquake record
NG10	Overlap with the subsequent earthquake record

Figures 3 and 4 show representative outcomes of the classification. Figure 3 shows a record from a larger-magnitude and shorter-distance event with strong amplitudes and high SNR. After preprocessing, spectra and time histories are stable, so the record is classified as OK. In contrast, Figure 4 presents a smaller-magnitude and long-distance case with weak amplitudes and substantial in-band noise; it violates the NG rules and is therefore classified as NG.

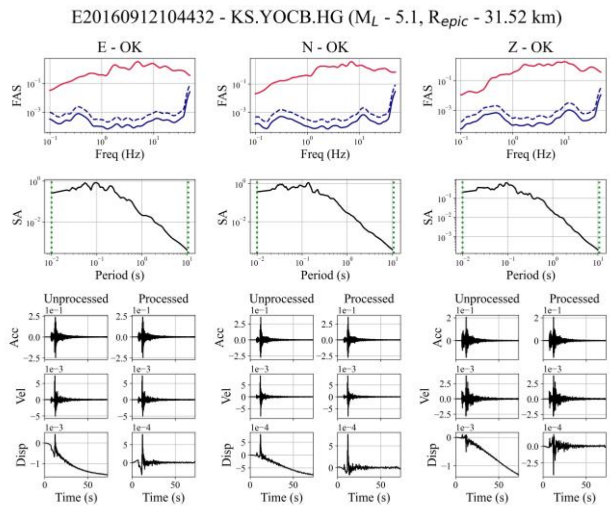

Figure 3. Preprocessing summary for the OK example: event E20160912104432 (M_L 5.1) at station KS.YOCB.HG (epicentral distance 31.52 km). Top: FAS of E/N/Z component signal (solid) vs. pre-event noise (dashed); green dashed lines mark selected band-pass corners. Middle: 5%-damped spectral acceleration from the processed records. Bottom: acceleration, velocity, and displacement time histories before (left) and after (right) processing.

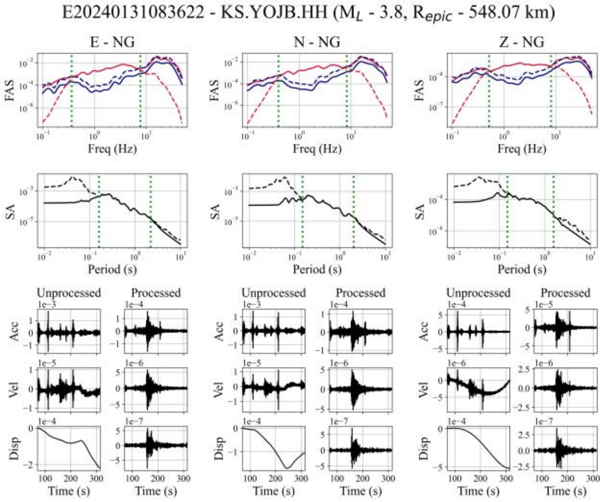

Figure 4. Preprocessing summary for the NG example: event E20240131083622 (M_L 3.8) at station KS.YOJB.HH (epicentral distance 548.07 km).

Figure 5 presents the classification results in terms of local magnitude (M_L) and hypocentral distance (R_{hypo}), illustrating how records were categorized using the NG filters. As shown in Figure 5-B, several records at M_L = 3.5 and M_L = 4.1 within 200 km were flagged NG because the 2022 Goesan earthquakes produced a foreshock (M_L 3.5) and mainshock (M_L 4.1) only 16 second apart, causing waveform overlap that triggered NG9 and NG10. The remaining OK set forms the basis for IM computation and flatfile assembly.

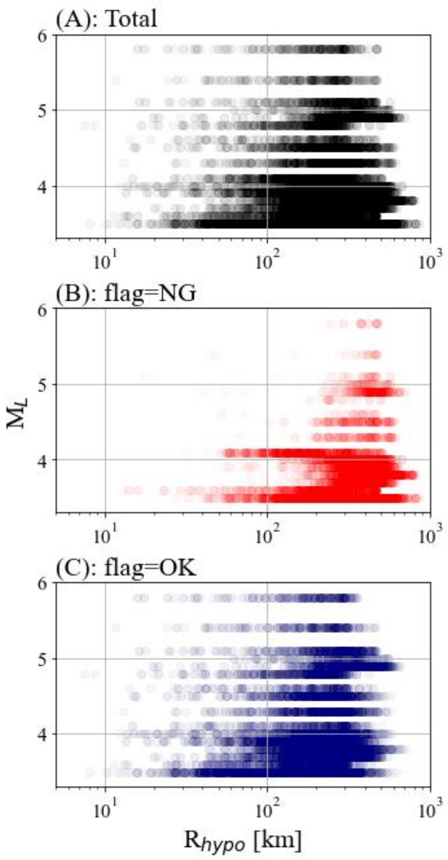

Figure 5. Magnitude (M_L) versus hypocentral distance (R_{hypo}) for (A) all processed three-component records, (B) records flagged as NG (not suitable) by the automated quality-control filters, and (C) records retained as OK for intensity measure computation. Notably, in panel (B), several M_L =3.5 and M_L =4.1 records within 200 km were classified as NG due to waveform overlap from the 2022 Goesan foreshock–mainshock sequence, which occurred only 16 seconds apart and triggered NG filters.

5 FLATFILE CONSTRUCTION

The flatfile integrates four categories of information for each record: (1) earthquake event metadata (origin time, epicenter, depth, magnitude, source-to-site distances), (2) station metadata (location, elevation, network, instrument type), (3) preprocessing parameters, and (4) intensity measures (IMs), including PGA, PGV, and spectral accelerations (SA) at multiple periods. 3-component component, geometric-mean, and RotD50 IMs were computed. The current flatfile covers events with $M_L \geq 3.5$ and future updates will extend coverage to $M_L \geq 2.5$.

Table 2. Column definitions for the flatfile

Column name	Description	Units
RSN	Record serial number	-
Network	Network code (KS, KG)	-
Station	Station code	-
Channel	Channel code (HG, HH, HA)	-
Direc	Component direction (E, N, Z, GM;Geo-mean, RotD50)	-
Event	Event ID (UTC)	-
ML	Local magnitude	-
E_Depth	Event focal depth	km
Repic	Epicentral distance	km
Azi	Source-to-site azimuth	degrees
E_Lat, E_Lon	Event latitude & longitude	degrees
S_Lat, S_Lon	Station latitude & longitude	degrees
S_Elev	Station elevation	m
S_Depth	Sensor installation depth	m
SPS	Sampling rate	Hz
StartTime, EndTime	Signal time-window	sec
fcHP, fcLP	Corner frequency	Hz
usTH, usTL	Usable spectral period limit	sec
n_fcHP, n_fcLP	Band-pass filter order used for the corner frequency	-
FilterType	Applied band-pass filter type	-
nth_baseline	Polynomial order used for baseline correction	-
flag	NG classification (OK, NG)	-
D5-75	Significant duration between 5% and 75% Arias intensity	sec
D5-95	Significant duration between 5% and 95% Arias intensity	sec
Tm	Mean period (Rathje et al. 1998)	sec
Tp	Predominant period (Rathje et al. 1998)	sec
Pulse	Pulse-like indicator (Shahi & Baker ,2014) (True, False)	-
Tpulse	Pulse period (Shahi & Baker, 2014)	sec
PGA	Peak ground acceleration	m/s^2
PGV	Peak ground velocity	m/s
Tx	Spectral acceleration at period x seconds	m/s^2

6 CONCLUSIONS

This study developed an automated pipeline for preprocessing, quality control, and compilation of South Korea–specific strong-motion records into a standardized flatfile. The proposed rule-based screening successfully identified anomalous records, including those affected by closely spaced events such as the 2022 Goesan foreshock–mainshock sequence. The resulting flatfile provides a consistent and reproducible dataset for ground-motion studies on the Korean Peninsula. Future work will extend the magnitude range to $M_L \geq 2.5$ and incorporate additional parameters, including Fourier amplitude spectra (FAS) from the processed waveforms.

7 ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) (RS-2023-00220751). The authors gratefully acknowledge this support. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors and do not necessarily reflect the views of the NRF.

8 REFERENCES

Ahn, B.S., Kang, T.-S., Jung, J.O. 2024. Station metadata integration of regional seismic networks in the southern Korean Peninsula. *Journal of the Geological Society of Korea*, 60(1), 111–119.

Ancheta, T.D., Darragh, R.B., Stewart, J.P., Seyhan, E., Silva, W.J., Chiou, B.S.-J., et al. 2014. NGA-West2 database. *Earthquake Spectra*, 30(3), 989–1005.

Boore, D.M., and Bommer, J.J. 2005. Processing of strong-motion accelerograms: Needs, options, and consequences. *Seismological Research Letters*, 76(3), 320–331.

Choi, I., Ahn, J.K., and Kwak, D. 2019. A fundamental study on the database of response history for historical earthquake records on the Korean Peninsula, *KSCE Journal of Civil and Environmental Engineering Research,* 39(6), 821-831.

Choi, S.W., Rhie, J., Lee, S.H., and Kang, T.S. 2020. A study on development of an earthquake ground-motion database based on the Korean national seismic network, *Journal of the Earthquake Engineering Society of Korea*, 24(6), 277-283.

Dawood, H.M., Rodriguez-Marek, A., Bayless, J., Goulet, C.A., and Youngs, R.R. 2016. A flatfile for the KiK-net database processed using an automated protocol. *Earthquake Spectra*, 32(2), 1281–1302.

Korea Meteorological Administration (KMA). NECIS: National Earthquake Comprehensive Information System. [Internet]. Available from: http://necis.kma.go.kr/

Lee, M.G., Kim, Y., Cho, H.I., Kim, H.S., Sun, C.G., Seong, Y.J., and Che, I.Y. 2024. KIGAM Quake: An open platform for seismological data and earthquake research information, *Geomechanics and Engineering*, 37(3), 279-291.

Rathje, E.M., Abrahamson, N.A., and Bray, J.D. 1998. Simplified Frequency Content Estimates of Earthquake Ground Motions. *Journal of Geotechnical and Geoenvironmental Engineering (ASCE)*, 124(2), 150–159.

Shahi, S.K., and Baker, J.W. 2014. An Efficient Algorithm to Identify Strong-Velocity Pulses in Multicomponent Ground Motions. *Bulletin of the Seismological Society of America*, 104(5), 2456–2466.