

EGU21-6195

<https://doi.org/10.5194/egusphere-egu21-6195>

EGU General Assembly 2021

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Residual analysis of large strong-motion flatfiles as a tool for detecting data error and anomalies

Claudia Mascandola, Giovanni Lanzano, and Francesca Pacor

Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy

The rapid increase of seismic waveforms, due to the increment of seismic stations and continuous real-time streaming to data centres, leads to the need for automatic procedures aimed at supporting data processing and data quality control. In this study, we propose a semi-automatic procedure for the consistency check of large strong-motion datasets, classifying the anomalies observed on the residuals analysis and identifying the possible causes.

The data collected in the strong-motion databases are usually arranged as parametric tables (called flatfiles), used to disseminate the Intensity Measures (IMs) and the associated metadata of the processed waveforms. This is the current practice for the Italian ACcelerometric Archive (ITACA, D'Amico et al., 2020) and Engineering Strong Motion (ESM; Lanzano et al. 2019a) databases. The adopted criteria for flatfile compilation are designed to collect IMs and related metadata in a uniform, updated, and traceable way, with the aim of providing datasets useful to develop Ground Motion Models (GMMs) for Probabilistic Seismic Hazard Assessment (PSHA) and engineering applications. Therefore, the consistency check of the flatfiles is a crucial task to improve the quality of the products provided by the waveform services.

The proposed procedure is based on the residual distributions obtained from ad-hoc ground motion prediction equations for the ordinates of the 5% damped acceleration response spectra. In this study, we focus on the active shallow crust events in ITACA, considering the ITA18 ground motion model (Lanzano et al., 2019b) as a reference for Italy. The total residuals, computed as logarithm difference between observations and predictions, are decomposed in between-event, between-station and event-and-station corrected residuals by applying a mixed-effect regression (Bates et al., 2015). This is the common practice for the (partial) removal of the ergodic assumption in empirical GMMs (e.g., Stafford 2014), where the contribution of the systematic corrective effects of event and station on aleatory variability are identified and shifted to the epistemic uncertainty. Afterward, the proposed procedure is applied to raise a warning in case of anomalous residual values. Warnings are provided when the normalized residuals exceed a certain threshold, in three ranges of periods (i.e., 0.01-0.15 s, 0.15-1 s, 1-5 s). The causes of warnings may be several and may concern the event, the site, the waveform, or a combination of them. Among the possible sources of anomalous trends, the more common are: preliminary or inaccurate event localization or magnitude, wrong soil category assigned based on proxies, misleading tectonic regime assigned to the earthquake, and fault directivity that may cause strong-ground motion amplification in certain

directions. Warnings may also raise for peculiarities in the site-response (e.g., large amplifications/de-amplifications at certain frequency-bands) and to the occurrence of near-source effects in the waveforms (see Pacor et al., 2018). Based on the raised warnings, a decision tree classifier is developed to identify the common anomaly sources and to support the consistency check of the semi-automatic procedure.

This study may help to enhance the waveform services and related products, besides reducing the variability of ground motion models and guiding decisions for site characterization studies and network maintenance.