Submission code: WA8B8QDWAX

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Development of SSPARQ: a Scalable Seismological Pipeline for Assessment, Review, and Quality of Seismic Data

Diogo Coelho (Observatório Nacional), André Nascimento (Observatório Nacional), Gilberto Leite Neto (Observatório Nacional), Marcelo Bianchi (Universidade de São Paulo), Eveline Sayão (Observatório Nacional), Sergio Fontes (National Observatory (Brazil))

Development of SSPARQ: a Scalable Seismological Pipeline for Assessment, Review, and Quality of Seismic Data

---

Identifying and mapping potential issues over time in stations from large seismographic networks is a challenging task, particularly due to the vast volumes of data involved and the difficulties on tracking the instrumental changes throughout station history. Additionally, networks operated by multiple institutions, like the Brazilian Seismographic Network (RSBR), may face miscommunication that can lead to a lack of awareness regarding problems with specific stations. These issues may include sensor misorientation, clock imprecision, problems related to sensor masses, and background noise. Due to the lack of prior knowledge about potential data issues, seismologists often face significant challenges in identifying the root causes of errors during data processing and analysis. Therefore, they would benefit from an efficient approach for automatically identifying instrumental issues across stations, as it would help to reduce processing time and facilitate making the necessary corrections. To address this need, we are developing SSPARQ (Scalable Seismological Pipeline for Assessment, Review, and Quality), a Python-based solution designed for assessing seismological data quality over time, relying solely on the P-wave energy characteristics of globally distributed earthquakes. SSPARQ processes seismograms to efficiently evaluate sensor orientation, clock stability, and amplification factor (gain) over time for each station in a given network. It conducts an automated analysis, based on Akaike Information Criterion (AIC) and signal-to-noise ratio (SNR) of the vertical component for known events, to select high-quality recorded events prior to metrics estimation, where the P-wave arrival is most prominent. The procedure lies in the formulation and evaluation of a cost function designed to identify the optimal azimuth; the cost function includes signal strength of the tangential component, similarity between vertical and radial components, transverse-to-radial energy ratio, and radial-to-vertical energy ratio. After evaluating the cost function for each angle in the azimuth search space, the algorithm identifies the angle that yields the highest cost value. It also compares the station detection time to the event expected time. For each station, the results are evaluated using a density-based clustering algorithm (DBSCAN), which verifies the consistency of the misorientation estimates over time. The formation of clusters in the misorientation analysis helps reveal persistent or intermittent instrumental issues (e.g., degradation of the seismic signal quality). We tested SSPARQ to analyze nearly 15 years of data from the ON subnetwork of the RSBR. Our analysis revealed that sensors were misoriented at six stations (GDU01, CMC01, NAN01, MAN01, CAM01, and TER01) until 2016/2017. When this issue was identified, the instruments were physically reoriented following field intervention. For example, station TER01 showed a 20º deviation until its correction in 2017. However, GDU01 remains misoriented, indicating that its data still require correction and the sensor needs to be physically rotated. Gain analysis of the sensor components at TRI01 revealed anomalous behavior from late 2015 to late 2017. This issue was solved after recalibration of the masses, back in 2018. Records from ABR01 showed a lag ranging from 5 to 60 s since 2020, which could indicate that the GPS antenna needs to be replaced. Based on these preliminary results, the package has proven valuable in identifying known and unknown instrumental issues within the ON subnetwork, confirming its ability to automatically detect potential problems over the course of network operation.