Pursuing data completeness in RSBR data archives
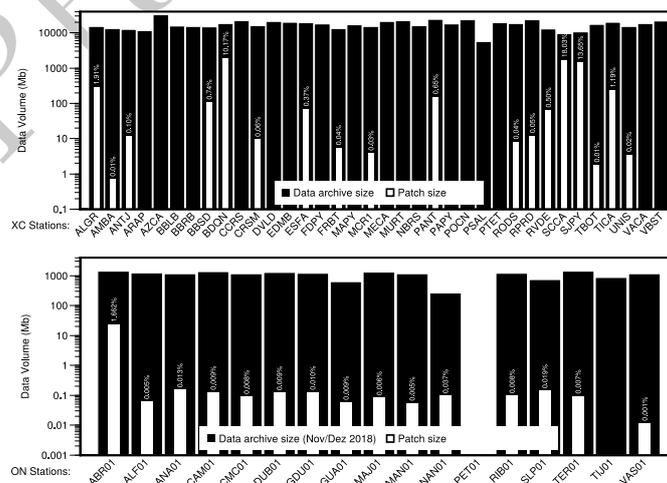
Marcelo Bianchi*[1], Fabio Dias[2], Thiago Moeda[2], and Giovanna Anacleto[2]

[1]*Universidade de São Paulo (USP)*
[2]*Observatório Nacional (ON)*

Abstract

Brazilian seismographic network (RSBR) is a well established distributed seismic network with one central archive (National Observatory/ON) and satellites archives in each node. Each node has its own procedure to acquire data that results in different data archives for the same data span and station. Real-time data is sometimes lost on the original node and can be found on a satellite node, the other way around is also true. This could indicate that SeedLink along with its archive tool is not guaranteeing data integrity. Using IRIS miniseed inspector (msi) we detected different issues that we do not understand the reason, but we could treat the symptom. A tool was implemented using PyRocko Python package and msi that can create a miniseed patch file from one archive to another. The patch is a file with data available from one archive that is not on the other. When ON and University of São Paulo (USP) archives were considered for data being acquired in the period of Nov-Dez/2018 some missing records could be observed. USP archive could contribute to a 24.7 Mb patch file to all 17 ON stations. It is worth mention that ABR01 station, which has shown the largest difference is acquired by USP. This value corresponds to a 0.148% of the total data volume (16.698 Mb) archived at the ON node for the ON stations in this period. When data recovered from instruments are considered a higher entropy is found, specially Nanometrics instruments have software issues that cannot recover all data stored in persistent storage. Again, the tool was tested with the XC temporary experiment operated by USP considering the transmitted and flash-disks-extracted data. A patch with 6.025 Mb was generated for 31 stations that had online transmission in the last 3 years. Again, the patch volume is small (∼1.7%) when compared to the total volume of 361.811 Mb of archived data for those stations, but we were able to fill gaps in the archive extracted from the instruments using transmitted data.