data quality and availability tests of public seismometer data in europe

F. Eckel[1] | J. Stampa[1] | M. Timkó[2] | L. Vecsey[3] | P. Kolínský[3]

[1]Institut für Geowissenschaften, Christian-Albrechts-Universität, Kiel, Germany
[2]EPSS Kövesligethy Radó Seismological Observatory, Budapest, Hungary
[3]Institute of Geophysics, Czech Academy of Sciences, Prague, Czech Republic

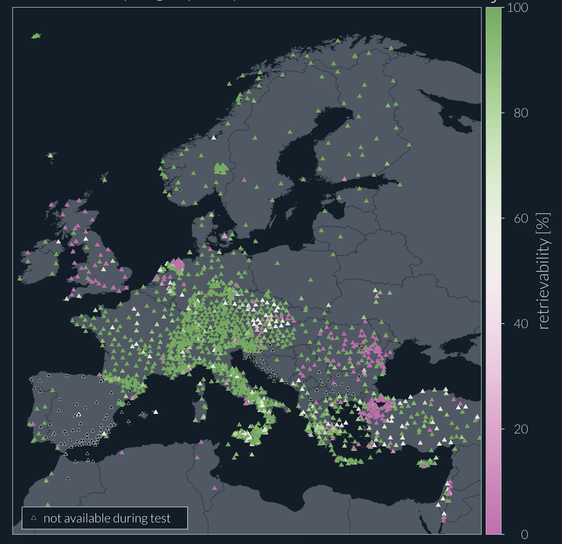
felix.eckel@ifg.uni-kiel.de
@FelixEckel

Many European countries agreed to share their seismic waveform data via the European Integrated Data Archive infrastructure (**EIDA**) where data from **thousands of seismic stations** is available. With these **large data sets**, manual data quality checks become more and more unfeasible.

Due to **rising traffic** at the data centers, errors are more likely to occur. Identifying stations and networks that are **more likely to be not available** upon requests or provide **erroneous data** is important in terms of quality control.

performed: 29.06 / 20.07.2022
locations: Kiel | Prague | Budapest

data retrievability



Data retrievability tests are performed by requesting **random** hourly data chunks within one year for all stations at the data centers (**EIDA node**). The request is forwarded by a routing client.

For every station in EIDA the amount of data that can be **successfully downloaded** is protocolled.

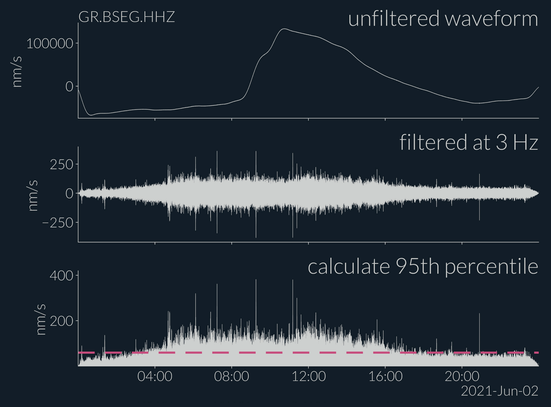
The test is simultaneously performed from **three different locations** (Kiel, Prague, Budapest). For each station the **retrievability** is calculated as the ratio of the requested data from all indivdiual tests and the downloaded amount of data.

Stations that are known to EIDA but **cannot be downloaded** appear in **magenta** colors. Stations for which every requested data chunk could be **successfully downloaded** appear in **green**.

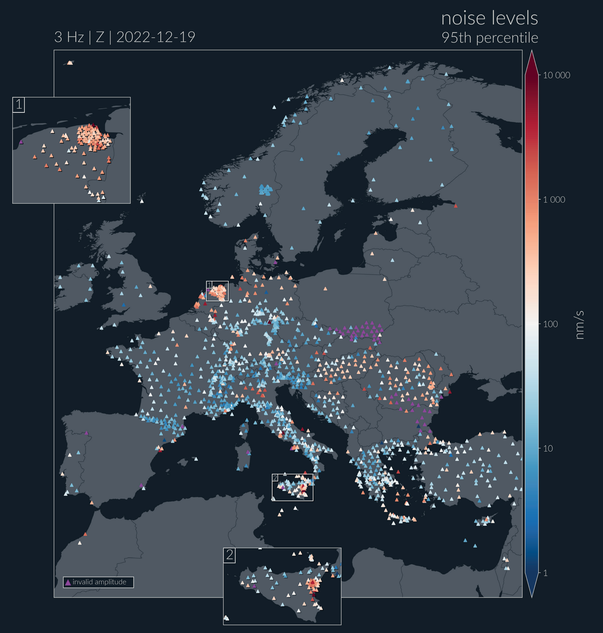
Data quality checks are vital to avoid processing of false data. Criteria for data quality depend on the methods being used. Here, a **noise level analysis** is used to focus on reported amplitudes.

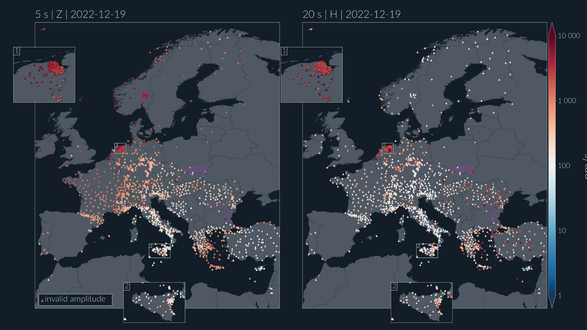
The average noise level at each station is calculated as the **95th percentile** of the **filtered absolute amplitudes**.

Comparing both the absolute amplitude values and the noise levels at neighboring stations, **false sensitivity values** or other **metadata problems** can be identified.



noise levels
95th percentile



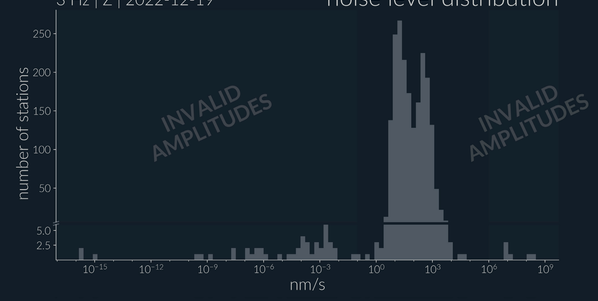


Average noise levels indicate structural variations over Europe, e.g. sedimentary basins (Po basin, molasse basin) or orogens (Alps, Apennines).

Stations with problems are identifiable by **outliering amplitudes**. These areas are marked as "invalid" in the histogram and colored in magenta on the map. Reasons for these outliers might be false metadata reportings or site effects.

3 Hz | Z | 2022-12-19

noise level distribution



Filtering in **frequency bands** with decreasing central frequency can reveal stations that **perform poorly** at long periods. Additionally the effect of the **microseism** band on the data quality becomes intutively visible. Watch this **video** to see the effect.

https://doi.org/10.5194/egusphere-egu23-7852

This presentation participates in OSPP.
Outstanding Student & PhD candidate Presentation contest

Sharing is encouraged

Christian-Albrechts-Universität zu Kiel
Mathematisch-Naturwissenschaftliche Fakultät