LLNL-TR-834057

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Waveform Data Quality Assessment

G. Ichinose

April 15, 2022

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Waveform Data Quality Assessment

RV/DC Seismoacoustic Network Working Group & Data Management Working Groups

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL Release# XXXXX

Purpose

Healthy seismoacoustic sensors and accurate metadata are critical to all the science objectives of both the Source Physics Experiment (SPE) Rock Valley Direct Comparison (RV/DC) and Low-Yield Physics Experiment-1 (LYNM PE-1). Both projects share similar seismoacoustic stations. We provide a plan on how to move forward with waveform data quality assessments (QA) to the working groups. The identification of stations with instrument response errors will be useful to researchers using amplitudes to study source effects and Earth attenuation. The data QA can also be useful for field technicians to help identify and correct problem sites.

Introduction

Mechanical seismometers and even stable modern broadband electronic feedback seismometers can drift over time (*e.g., Ekstrom et al*., 2006) due to temperature and humidity variations which can affect the true instrument responses (*e.g., Hutt and Ringler*, 2011). Ambient noise power levels are useful tool in monitoring time-dependent response changes (*e.g., Ringler et al*., 2010) based on power spectral density-probability density function (PSD-PDF) methods (*e.g., Peterson*, 1993; *McNamara and Buland*, 2004). Figure 1 illustrates the concept of the PSD/PDF method. Background noise is energy recorded by sensors typically comprised mostly of interactions between ocean and atmosphere with the solid Earth at lower frequency. At very low frequency there is a constant noise level termed the "Earth hum" that varies as a function of sensor emplacement. At higher frequencies, instrument self-noise can be a bigger factor than emplacement conditions. Background noise is continuous and best recorded during times when there is an absence of seismic signals typically during quiet nighttime hours (e.g., Phillips and Stead, 2020). Seismic signals from earthquakes or explosions are transient seismic signals that are short in time duration and occur randomly. Most human generated noise sources are transient (e.g., vehicle traffic, generators). This is mostly the cause in variations in seismic amplitudes between day and night. There are also seasonal variations "winter-summer" due to weather changes, for example, winter windstorms, ocean surf, and freezing and defrosting of permafrost can affect the noise levels over days-to-months.

Methods

One method to compute waveform data QA metrics is to use the IRIS System for Portable Assessment of Quality (ISPAQ). From IRIS: ISPAQ can calculate over 40 MUSTANG-style data quality metrics using either local miniSEED files or data available through any Data Center that supports FDSN web services. These metrics include basic trace statistics, metrics based on miniSEED state-of-health flags (if available), metrics based on event arrivals, and metrics derived from Power Spectral Densities (PSD). PSDs and Probability Density Functions (PDF) can be output as text files. PDFs can also be output as plots. All results are computed and stored on the user's local machine. The final decision on what software package is used depends upon the decision of who is assigned this role by the project managers and working groups.

Plan Recommendations and Requirements

1. We recommend using ISPAQ to compute QA metrics (or any comparable software).
2. We recommend that the management teams delegate roles and responsibilities so that a weekly or monthly QA status reports of problem sites be distributed based on a set of agreed upon QA metrics.
3. We recommend that a requirement is made so that PSDs can be labeled. This labeling is necessary in developing training datasets for any future machine learning algorithms based on neural networks.
4. An optional website is recommended so that researchers, managers, and field techs can look at QA metrics.
5. We recommend spot checks for sensor orientation errors using cross correlation of co-located sensors for any deployed small aperture arrays like gradiometers (e.g., Langston, 2018) based on signals from earthquakes (see co-located sensor requirement below).
6. We also recommend checking channel polarities and amplitude gains using regional long-period moment tensor inversion from regional earthquakes (e.g., Figure 4).

Requirements to consider

1. **Continuous Data** – Continuous data collection was accomplished in SPE and DAG experiments. We anticipate that all RV/DC seismoacoustic data from "far-field" network and "near-field" sensors will be telemetered to designated data storage site. FDSN standard naming conventions will also be supplied along with metadata.
2. **Deployment Durations** – There was at least 1 year of data collected before the first DAG experiment. This provided seasonal variations in background noise however at least a few months of continuous data collection will be adequate for PSD/PDF. Deployments of a year or longer can provide enough time for moderate to large magnitude regional and teleseismic seismicity to be recorded for additional spot checks using earthquake signals.
3. **QA needed for sites**: TBD. Primary sites have high priority for QA and will be included in regular reporting while secondary sites may be QA less often.
    a. **Primary Priority:** seismoacoustic sites near Rock Valley and others within the NNSS boundary including gradiometer arrays (not Large-N or DAS).
    b. **Secondary Priority:** Entire SN and NN network not included as primary?
    c. *Other sites?*
4. **Co-located Sensors** – A few sites (e.g., 5-10) should contain 2 or more sensor types (e.g., 3-C broadband and geophone). The direct comparison of amplitudes provides validation of instrument response correction metadata. The co-location of 3C sensors also provides validation of the three sensor component orientations. The actual number of doubled up

sensors needs to be agreed upon by WG because there is a conflict with achieving the maximum experimental coverages and trusting the absolute ground motion amplitudes.

5. **Data formats**
    a. Waveform: miniSEED is the best file format saved consistently in 1-day or hourly intervals.
    b. Metadata: Dataless SEED, XML, EVALRESP, CSS
    c. *Other formats?*
6. **Reporting**. A seismoacoustic network QA report can be made available on either a weekly or monthly basis. More frequent reporting may be necessary prior to major experiments. Reports include:
    a. Mustang metrics provided by ISPAQ, including deviations of amplitude created by PSD/PDF outside of thresholds (e.g., mode outside of the 95% probabilities or LNM and HNM). Threshold may need to be adjusted to prevent the repeated reporting of known noisy sites (e.g., near highways or construction sites)
    b. Basic Data availability chart and/or Helicorder
    c. In addition to metrics, we should consider human review and labeling for future use as a training dataset for machine learning.

References

Barzilai, A., T. VanZandt, and T. Kenney (1998). Technique for measurement of the noise of a senor in the presence of large background signals, Rev. Sci. Instruments 69(7), 2767-2772.

Cessaro, R. K. (1994). Sources of Primary and Secondary Microseisms, Bull. Seismol. Soc. Am., 84(1), 142-148.

Cooley, J. W., and J. W. Tukey (1959). An algorithm for machine calculation of complex Fourier series, Math. Comp. 19, 297-301.

Ekstrom, G. C. A. Dalton, and M. Nettles (2006). Observations of time-dependent errors in long-period instrument gain at global seismic stations, Seismo. Res. Lett., 77(1), 12-22.

Hutt, C. R., and A. T. Ringler (2011). Some possible causes of and corrections for STS-1 response changes in the global seismograph network, Seismol. Res. Lett., 82(4), 560-571, doi: 10.1785/gssrl.82.4.560

ISPAQ https://github.com/iris-edu/ispaq

Langston, C. A. (2018). Calibrating Dense Spatial Arrays for Amplitude Statics and Orientation Errors, Journal Geophysical Research, 123(8), 3849-3870. https://doi.org/10.1002/2017JB015098

McNamara, D. E., and R. Buland (2004). Ambient noise levels in the continental United States, Bull. Seismol. Soc. Am., 94(4), 1517-1527, doi:10.1785/012003001

McNamara, D. E., R. P. Buland, H. M. Benz, and W. R. Leith (2005). An assessment of seismic noise levels for the Advanced National Seismic System backbone network and selected regional broadband stations, U.S. Geol. Surv. Open File Report 2005-1077, 29pp.

Peterson, J. R. (1993). Observations and modeling of seismic background noise, U.S. Geol. Surv. Open-File Report 93-322, 94pp.

Phillips, W. S., and R. J. Stead (2020). Identifying Suspect Instrument Intervals Using Midnight Noise Time Histories. United States: N. p., 2020. Web. doi:10.2172/1659160.

Ringler, A. T., L. S. Gee, C. R. Hutt, and D. E. McNamara (2010). Temporal variations in global seismic station ambient noise power, Seismo. Res. Lett., 81(4), 605-613, doi: 10.1785/gssrl.81.4.605

Ringler, A. T., J. R. Evans, and C. R. Hutt (2015). Self-noise models of five commercial strong-motion accelerometers, Seismo. Res. Lett., 86(5), 1143-1147, doi: 10.1785/0220150027

Figures

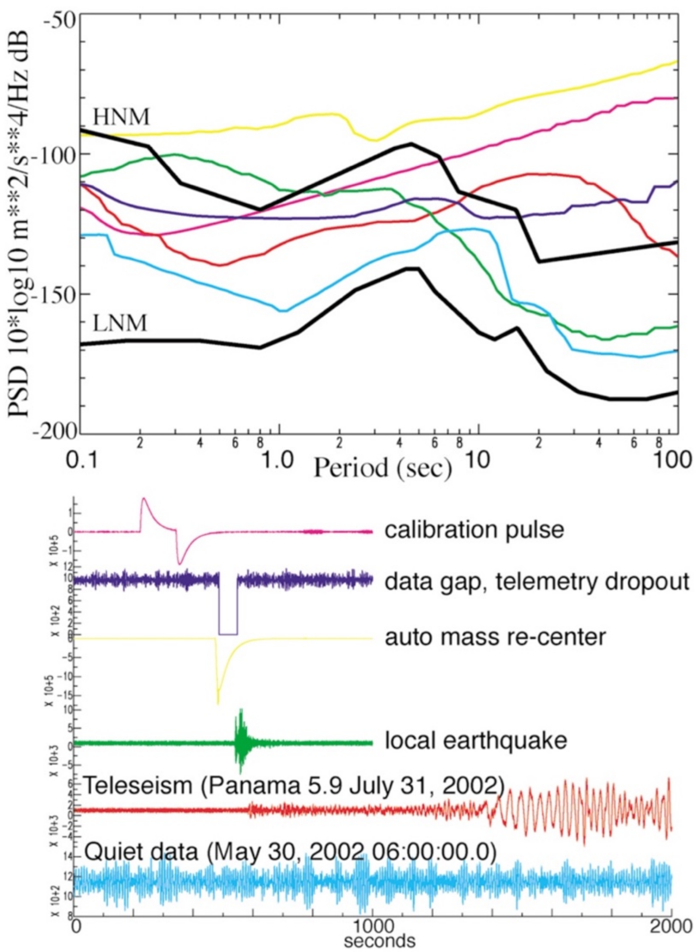

Figure 1. Baseline noise model estimation approach first required the time-consuming removal of transients (*e.g., Peterson*, 1993) but later *McNamara and Buland* (2004) suggested a statistical approach to automatically remove nonstationary parts. The method uses probability density functions (PDF) from many PSDs allowing for the separation of the transients less likely to occur “in the tails of the distribution” and the extraction of the stationary part at the highest frequency of occurrence. The combined PSD and PDF approach therefore requires thousands of records, weeks to years of continuous seismic data to reach stability depending on the variability of ground motions. A stable background noise during quiet data is used to provide a baseline for tracking amplitude and metadata deviations.

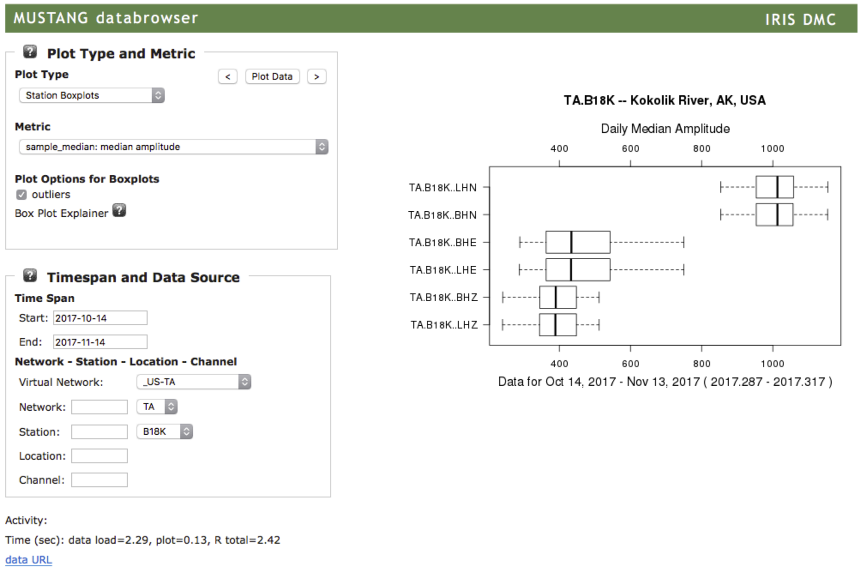

Figure 2. Screen shot from IRIS ISPAQ mustang metrics browser interface

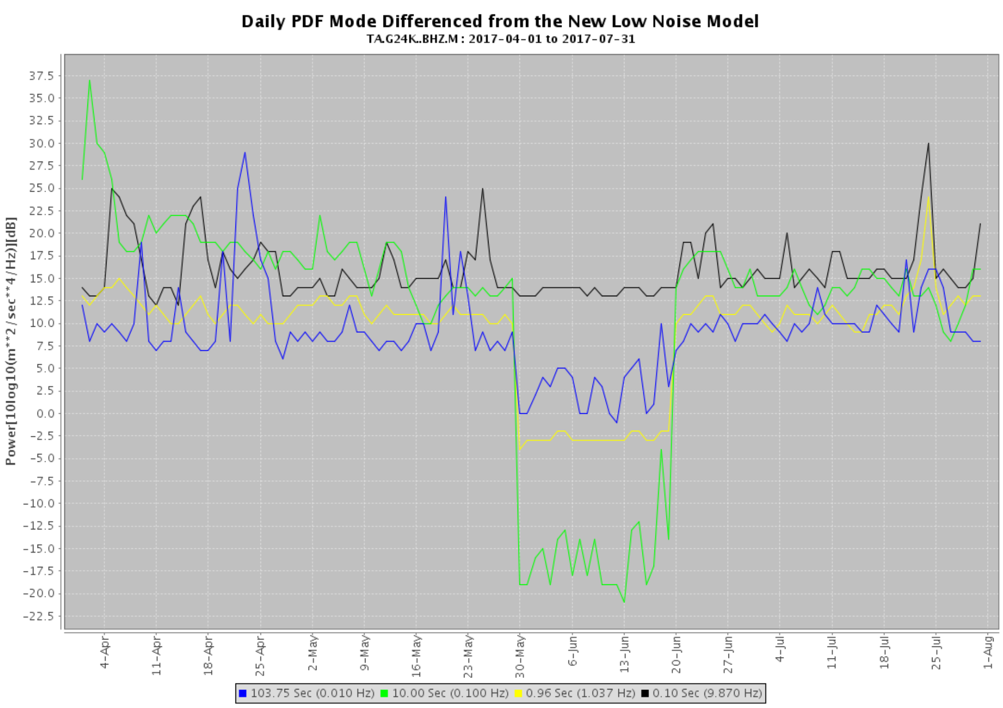

Figure 3. Example time strip chart of amplitude baseline using differences between the PSD/PDF mode and low noise model. Screen shot from IRIS ISPAQ mustang metrics browser interface.

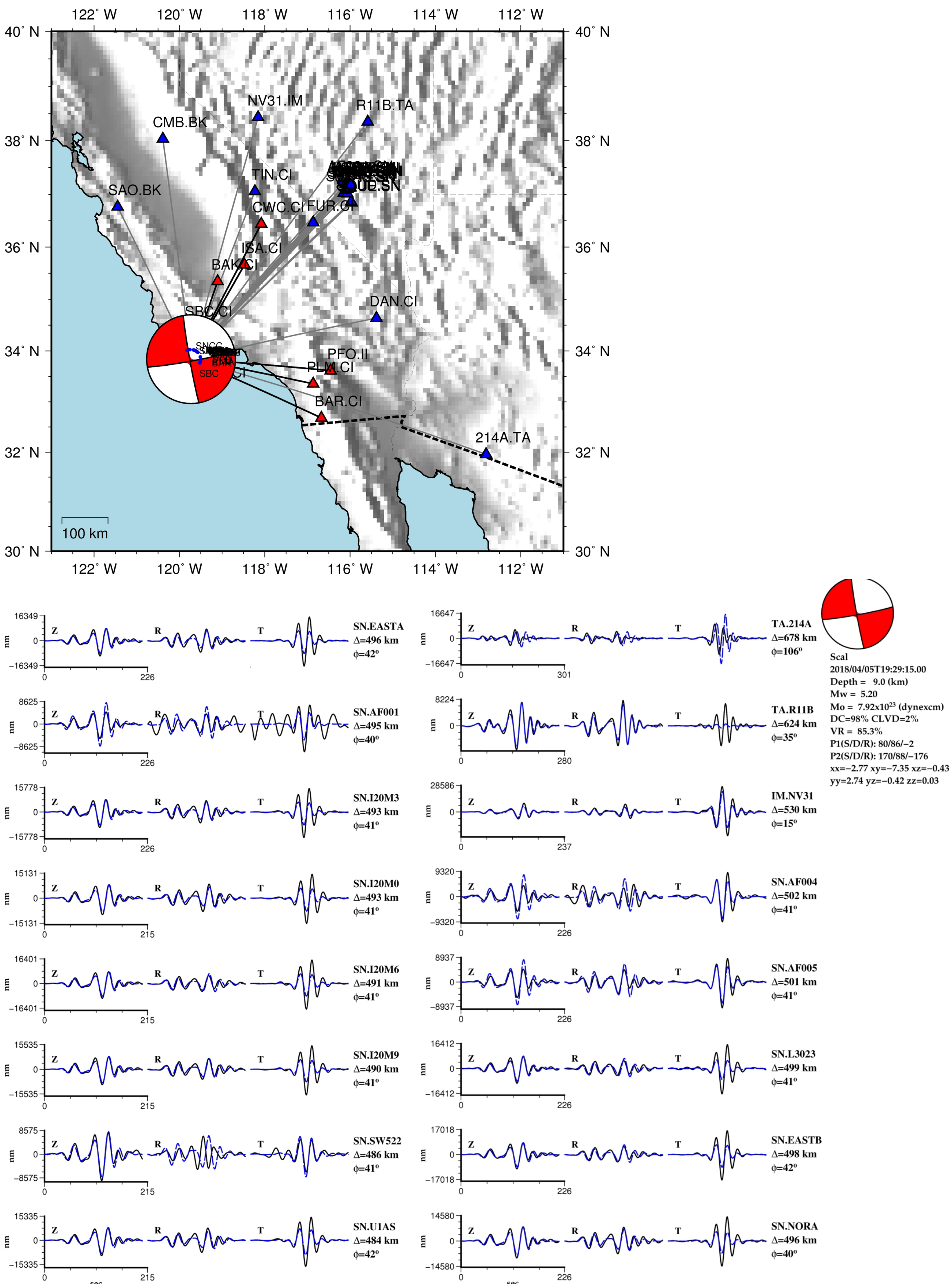

Figure 4. Moment tensor inversion example predicting long-period full waveform amplitude and phase. No polarity or channel orientation issues were found and the misfit for SW522 was found to be transient noise near the sensor or from the recording system.