A Method to Establish Seismic Noise Baselines for Automated Station Assessment

D. E. McNamara, C. R. Hutt, L. S. Gee, H. M. Benz, and R. P. Buland

INTRODUCTION

Currently, the United States Geological Survey (USGS) National Earthquake Information Center (NEIC) receives more than 4,000 channels of seismic data from roughly 900 seismic stations distributed around the world in near real time. As the NEIC develops increasingly sophisticated methods for automated data processing, it becomes equally important to develop tools for ensuring that only the highest-quality data are used in rapid earthquake products. In this paper, we present methods and applications for computing long- and short-term, station-specific noise baselines for broadband seismic data channels. Long-term baselines (one year or greater) are useful for determining ambient noise conditions, while short-term baselines (one hour to one month) are useful for monitoring changing station performance and noise characteristics. This method is currently in testing and development at the NEIC for evaluation of instrument responses, assessment of data quality for real-time earthquake processing systems, new station site evaluation, and assessment of gross network characteristics.

Real-time earthquake processing systems at the NEIC rely on high-quality seismic data to compute accurate earthquake locations and magnitudes, moment-tensor solutions, finite-fault models, and shaking intensity. The NEIC receives real-time seismic data from a variety of contributors including: global networks (Global Seismographic Network [GSN], GEOSCOPE, GEOFON), national networks (ANSS, Australia, Spain, Canada, Switzerland), regional networks (ANZA, PRSN, CISN, PNWSN, AEIC), government agencies (National Oceanic and Atmospheric Administration [NOAA] tsunami warning centers, the International Monitoring System [IMS]), and temporary deployments such as the EarthScope transportable array (TA). On most days, approximately 10% of contributed instrument responses are either not known or are erroneous and cannot be used for accurate real-time amplitude information. Maintaining accurate and current metadata for roughly 900 contributed stations is difficult when instrumentation upgrades or other changes occur regularly. Automated processing of real-time seismic data necessitates the development of automated tools and procedures to quickly identify changes in instrument response and other systematic changes in station performance that can affect the accuracy of NEIC earthquake products.

Baseline noise models have been commonly used to assist in the design, installation, and operations of broadband seismic stations (Peterson 1993; McNamara and Buland 2004; Berger et al. 2004; Sleeman and Vila 2006; Sleeman et al. 2006; Bahavar and North 2002). In this study, we investigate the spectral characteristics of numerous individual broadband seismic stations (Figure 1) and develop channel-specific baseline models using probability density functions (PDF) of power spectral densities (PSD) (after McNamara and Buland 2004) for seismic station performance assessment.

METHODS

Station noise baselines are determined by computing a spectral envelope for each channel, based on the 10th and 90th percentiles of the PSD PDF. The first step in this process is computing PSDs following the same algorithm used to develop the GSN New Low Noise and High Noise models (NLNM, NHNM; Peterson 1993). PSDs are computed from continuous, overlapping (50%) time series segments (BH channels: one-hour segments sampled at 40 sps or 20 sps; LH channels: three-hour segments sampled at 1 sps). All available data are included; there is no removal of earthquakes, system transients, or data glitches. The instrument transfer function is deconvolved from each time segment, yielding ground acceleration for direct comparison to the NLNM and NHNM. Each time series segment is divided into 13 subsegments (900 s for BH and 2,700 s for LH channels), overlapping by 75%. Each subsegment is processed by: 1) removing the mean, 2) removing the long-period trend, 3) tapering using a 10% sine function, 4) transforming via fast Fourier transform to obtain the amplitude spectrum, and 5) squaring the amplitude spectrum to obtain the power spectrum. In the next step, PSDs are gathered by binning periods in % octave intervals and binning power in one-dB intervals. To construct an empirical PDF, each period-power bin is normalized by the total number of contributing segments. Finally, we define the standard station noise baseline envelope as the 10th and 90th percentile of the PSD distribution. Figure 2 shows the noise baselines and background PDFs for the current instrument epoch at IU.ANMO.00.BHZ. In this example, baselines were computed...
Figure 1. Map of more than 500 real-time broadband stations contributing to the USGS NEIC with noise baselines computed using the methods discussed in the study. Red triangles indicate stations discussed as examples.

Figure 2. Probability density function (PDF) of power spectral density (PSD) for seismic data channel IU.ANMO.00.BHZ. Also shown on the PDF is the station baseline based on the 10th (black line), 50th (yellow line), and 90th (red line) percentiles of the PSD distribution. Annotations indicate sources of noise in the PDF. Gray lines show the high- (NHNM) and low-noise models (NLNM) (Petersen 1993).
from 114,274 hourly PSDs during the period from 19 November 2002 (2002:323) through 14 November 2008 (2008:318). We use the SEED channel naming convention (Ahern et al. 2007): in the case of IU.ANMO.00.BHZ, the network code is IU, the station code is ANMO, the location code is 00 and the channel code is BHZ. Baselines in Figure 2 are shown as the 10th (black line) and 90th (red line) percentiles of the PSD distribution. The 50th percentile (yellow line) and high- and low-noise models (NHNM, NLNM; Peterson 1993) are also shown for reference. Eighty percent of the 114,274 hours sampled from 2002:323 to 2008:318 for IUANMO.00.LHZ fall within the baseline envelope. This includes a large portion of the daily range of cultural noise generated by people working at Albuquerque Seismological Laboratory (ASL), as well as most of the seasonal variation of the oceanic storm microseisms (2-25 s). The remaining 20% of the PSD observations are considered out-of-nominal noise conditions. Below the 10th percentile this includes infrequent anomalously quiet periods of "cultural noise" and oceanic storm microseism events, "dead" sensors, and time segments with multiple short gaps in the data. Above the 90th percentile this includes infrequent local "cultural" noise sources, large oceanic storms, time periods with incorrect instrument responses, system transients, spikes, glitches, offsets, sensor recentering, calibration pulses, excessively noisy sensors, and earthquakes.

**DISCUSSION: AMBIENT NOISE AS A FUNCTION OF TIME AND LOCATION**

Station baseline envelopes vary considerably as a function of time and geographic location. Advances in instrumentation have reduced noise levels in long-period bands (> 10 s) while changing local noise sources such as roads and population density can increase shorter period (< 1 s) noise levels. Ambient noise levels are also strongly affected by geographic location mainly due to proximity to coastlines and population centers (McNamara and Buland 2004).

**Time**

With the deep seismic waveform archives at ASL and the Incorporated Research Institutions for Seismology (IRIS) Data Management Center (DMC), it is possible to examine station noise characteristics over several decades for several
sites around the world. Figure 3 shows the evolution of noise baselines over four generations of seismic instrumentation in Albuquerque using the long-period vertical component (LHZ) (one sample/s) continuous velocity time series from the High-Gain, Long-Period Network (HG.ALQ.--.LHZ; 1972–1978) (Figure 3A), the Seismic Research Observatory (SR.ANMO.--.LHZ; 1974–1989) (Figure 3B), and the current GSN station (IU.ANMO.--.LHZ and IU.ANMO.00.LHZ; 1989–2008) (Figures 3C and 3D). The PDF and noise baselines in Figure 3(A) for HG.ALQ.--.LHZ are computed from 61,678 three-hour PSDs occurring from 31 August 1972 (1972:244) to 17 July 1978 (1978:198). At periods > 100 s, PSDs occurring from 27 October 1998 (1998:300) to 19 May 2008 (2008:140). During this time period, the instrumentation for this channel was the Geotech S-11 long-period vertical seismometer installed in an underground tunnel. Instrumentation and installation differences from this time period are apparent at long periods (> 40 s) by low noise levels and in the secondary microseism period band (< 12 s) that was excluded by analog filtering. It is interesting to note that the 10th percentile long-period noise levels are very low, near the NLNM, and able to resolve the “hum” at periods from 50 to 200 s (Rhein and Romanowicz 2004). However, the separation between the 10th and the 90th percentile can be as much as 25 dB across the entire spectrum. This likely reflects long-period instability due to limitations of the original shallow subsurface vault.

The PDF and noise baselines in Figure 3(B) for SR.ANMO.--.LHZ are computed from 155,994 three-hour PSDs occurring from 28 August 1974 (1974:240) to 9 August 1989 (1989:236). During this time period, the instrumentation for this channel was the Geotech KS-36000 borehole seismometer. Sensor limitations from this time period, relative to current IU.ANMO (Figures 3C and 3D), are apparent as relatively higher noise levels at long-periods (> 40 s).

Recording system improvements are observed in the secondary microseism period band (< 13 s). Originally excluded by analog filtering, the secondary microseism was progressively resolved over time as shorter periods were included (Figure 3B). The PDF and noise baselines in Figure 3(C) for IU.ANMO.--.LHZ are computed from 79,309 three-hour PSDs occurring from 29 August 1989 (1989:224) to 17 August 1998 (1998:299). The PDF and noise baselines in Figure 3(D) for IU.ANMO.00.LHZ are computed from 124,442 three-hour PSDs occurring from 27 October 1998 (1998:300) to 19 May 2008 (2008:140). During this time period (1989:241 to present), the instrumentation for this channel was the Geotech KS-54000 borehole seismometer. Both channels display relatively quiet and stable noise conditions with very low power and only a 15-dB separation between the 10th and 90th percentiles in the microseism band (4–20 s period). It is interesting to note that the peak of the secondary microseism at ANMO is shifted to longer periods (~7–9 s) relative to the NLNM (~4–5 s). This effect is due to the dominance of longer-period Pacific Ocean wave sources at the distant inland Albuquerque site (Aster et al. 2008). At periods > 100 s, PSDs at IU.ANMO.00.LHZ from 1998 to 2008 (Figure 3D) closely follow the NLNM with 5-dB separation between the 10th and 90th percentiles, suggesting that the remote location and borehole installation provide stable long-period conditions.

Station baseline models for current stations as well as stations no longer in operation can serve as a valuable tool for evaluating historical instrument response. This is important for researchers interested in observing absolute ground motion from the older systems. For this reason, an effort is currently underway to compute station baselines for all GSN and ANSS predecessor stations.

**Geographic Location**

Seismic station noise levels also vary with geographic location (Reif et al. 2002; Stutzman et al. 2000, McNamara and Buland 2004). To demonstrate variation in the baselines, we have selected four stations (Figure 4, Figure 1) and described their noise baseline in three bands: short-period (0.1–1.0 s), microseism (2–20 s), and long-period (20–900 s).

Figure 4(A) shows the BHZ and LHZ components for station IU.QSPA.10. This station is located in Antarctica, at the South Pole Remote Earth Science Observatory (Quiet Zone) and is part of the IRIS/USGS component of the GSN. Baseline noise models are computed from one year of continuous 20-sps (BHZ) and 1-sps (LHZ) data recorded using the Guralp CMG3-T borehole seismometer and Quanterra Q680 digitizer.

**Short-period Characteristics (0.1–1.0 s).** Very low short-period noise levels are observed as the 10th percentile is below the NLNM. This observation is due to suppression of 0.2-s (5 Hz) energy by the overlaying ice (Anderson et al. 2003) and is a strong influence on the lower short-period powers in the GSN noise model (GNM; Berget et al. 2004) relative to the NLNM (Petersen 1993). The 20–30 dB separation between the 10th and 90th percentiles reflects a significant diurnal variation in cultural noise due to human activity and machinery at the observatory and suggests noise levels below the NLNM are infrequent and restricted to nonworking hours.

**Microseism Characteristics (2–20 s).** Due to the inland location, the secondary microseism power levels are moderate and relatively stable, with little separation (~10–15 dB) between the 10th and 90th percentiles. Also, the secondary microseism is narrow and peaked at longer periods, 6–7 s, relative to the NLNM. We also observe a strong primary microseism reflected in the 90th percentile at approximately 20-s period.

**Long-period Characteristics (20–900 s).** The tight spread of ~10 dB between the 10th and 90th percentiles, approximately 20–30 dB above the NLNM, indicates moderately quiet and stable long-period conditions; however, not quiet enough to observe the “hum.”

Figure 4(B) shows the horizontal BHE and LHE channels for station IU.PTCN.00. This station is located on Pitcairn Island in the South Pacific and is part of the IRIS/USGS component of the GSN. Baseline noise models are computed from one year of continuous 20-sps (BHE) and 1-sps (LHE).
data recorded using the Streckeisen STS-2 seismometer and Quanterra Q680 digitizer.

**Short-period Characteristics (0.1–1.0 s).** We observe high short-period noise levels, with the 10th and 90th percentiles above the NHNM, trending into high secondary microseism power levels. While short-period noise levels are high due to local cultural and natural (the ocean) noise sources, we observe a small separation (5–10 dB) between the 10th and 90th percentiles, suggesting little diurnal and seasonal variation in noise levels.

**Microseism Characteristics (2–25 s).** Power levels are high and the peak broader in the microseism band relative to inland stations such as IU.ANMO (Figure 2). High and broad secondary microseism power is observed, peaked at about 5 s due to the proximity of this island site to deep Pacific Ocean sources of secondary microseisms (Aster et al. 2007).

**Long-period Characteristics (20–900 s).** In general, horizontal components have higher noise levels than vertical components.

In this case, the LHE has very high power at or above the high-noise model. This is likely due to thermal instabilities and wind-induced tilt in a surface vault installation.

Figure 4(C) shows the BHN and LHN channels for station CU.BCIP. -- This station is located at the Smithsonian Tropical Research Institute on Isla Barro Colorado in the Panama Canal and is part of the GSN affiliated USGS Caribbean network. Baseline noise models are computed from nearly two years of continuous 40-sps (BHN) and 1-sps (LHN) data recorded using the Streckeisen STS-2 seismometer and Quanterra Q330 high-resolution digitizer.

**Short-period Characteristics (0.1–1.0 s).** Power levels at short periods are relatively high and display a weak diurnal distribution related to activity at the research institute. The station is located in a heavily forested area, so high short-period power levels can also be related to tree motion coupled into the ground through the root systems.
Figure 5. Map showing the geographical distribution of the 2–25 s, band-averaged, PDF medians.

Microseism Characteristics (2–25 s). The secondary microseism peak is high and broad, which is typical of stations located near coastal zones. CU.BCIP is equidistant from both the Pacific Ocean and Caribbean Sea and as a result microseism energy generated by oceanic waves produces high power and has a peak at relatively short periods (2–3 s). The microseism is probably dominated by storms in the Pacific.

Long-period Characteristics (20–900 s). A unique and interesting long-period (100–300 s) signal is apparent in the LHN baseline and appears to be related to canal operations. The signal is in the same period band as the "hum" but does not have the seasonal distribution expected of climate-related seismic signals. Instead it is strongest during daytime and evening hours and absent from roughly 1 to 6 a.m. local time. While the canal is always open, the anomalous long-period signal is dominant when major shipping operations occur (McNamara et al. 2009). Long-period noise levels are very high in general and likely related to vault design. CU.BCIP has a surface vault in the Center for Earthquake Research (CERI) style (Hutt et al. 2006), which is likely affected by tilt induced by the surrounding soils.

Figure 4(D) shows the BHZ and LHZ channels for station G.TAM.00. This station is located in Tamanrasset, Algeria, and is part of the GEOSCOPE network operated by the Institut de Physique du Globe de Paris (Stutzman et al. 2000). Baseline noise models are computed from one year of continuous 40-sps (BHZ) and 1-sps (LHZ) data recorded using the Streckeisen STS-1V/VBB seismometer.

Short-period Characteristics (0.1–1.0 s). Noise levels are high in the shorter periods due to the proximity to the Algerian city of Tamanrasset. The 15-dB separation between the 10th and 90th percentiles represents relatively small diurnal variations of short-period cultural noise.

Microseism Characteristics (2–25 s). Microseism energy is low, near the NLNM, and the percentiles are tightly grouped due to large distance from ocean coastlines. The remote location distant from the coast has the effect of shifting the secondary microseism peak to slightly longer periods (6 s), relative to the NLNM, than at near-coastal sites.

Long-period Characteristics (20–900 s). The large 20–30-dB separation between the 10th and 90th percentiles in the long-period band is likely due to instability of a surface vault installation versus borehole. The 50th percentile is less than 10 dB from the 10th and very near the NLNM, indicating that nearly half of the long-period noise levels are very low. Noise levels are low enough that at the 10th and 50th percentile levels, the Earth’s "hum" is observable as a subtle increase in power from 60 to 300-s period (Rhie and Romanowicz 2004).

This discussion presents a small sampling of station-specific noise baselines. As demonstrated, baselines can vary considerably as a function of time and geographic location. Figure 5 illustrates the geographic variation of the microseism band using all stations in Figure 1. A power range of over five orders of magnitude is observed in the 2–25 s, band-averaged, median of the individual station PDFs. Also, as expected, we observe...
a clear geographic pattern with higher power in and near the oceans and lower power in the continental interiors. Baselines and PDFs for all GSN, and GSN predecessor networks can be found at the ASL ftp site: ftp://asLftp.cr.usgs.gov/pub/users/McNamara/PDFs, and all networks contributed to the NEIC: ftp://hazards.cr.usgs.gov/web/mcnamara/NEIC/PDFs). A smaller sampling of PSD PDFs and noise baselines is available in the electronic supplement to this article.

**APPLICATION TO THE NEIC**

Figure 5 highlights the importance of individual station baselines for use in quality control tools to identify performance issues. As a first step toward developing a monitoring system, we have implemented a simple review process. On a daily basis, PSDs are computed on all real-time channels using the PQLX software system (for description and download information see McNamara and Boaz 2006 and Boaz and McNamara 2008). The computed PSDs are stored in a database, allowing generation of recent short-term (one hour to one month) PDFs for user-defined time periods. An analyst visually compares the short-term PDFs against the long-term station noise baselines to identify out-of-character noise conditions, such as instrument response changes or system transients.

The visual comparison method is very sensitive to instrument response problems such as incorrect input units and sensitivities. Figure 6 is an example using the ANSS backbone station US.LRAL.--.BHZ. Figure 6(A) shows the PDF and long-term baseline computed using 23,151 PSDs during the current instrument epoch from 2007:078 to 2008:302. In the following panels (Figures 6B–6D) short-time periods with instrument errors are shown compared to the long-term baseline. In Figure 6(B), an incorrect instrument response has been applied to 212 hours of data, from 2008:040 to 2008:044, in order to demonstrate
the sensitivity of this method to a possible error in instrument response units. Units of displacement instead of velocity, expressed as a missing zero in the response file, results in clockwise rotation of the PSDs for the erroneous data time period. This is clearly observed as high power levels at short periods and low power at long periods, relative to the long-term station baseline model. In Figure 6(C), we apply an incorrect instrument response to 213 hours of data from 2008:101 to 2008:106, where units of acceleration are used instead of velocity. This error is implemented by adding an extra zero to the instrument response. The result is a counter-clockwise rotation of the PDF, which is clearly observed and easily detectable as low power at short periods and high power at long periods, relative to the long-term station baseline model. In all three cases, incorrect instrument response causes incorrect amplitudes that can map into station magnitude calculations and other NEIC products that rely on accurate estimates of ground motion.

FUTURE DIRECTIONS: STEPS TOWARD AUTOMATION

With a well-defined set of station noise baselines, the next step toward monitoring station performance at the NEIC is to minimize human involvement by automating the visual detection method. As part of this development, we are compiling a set of known noise sources in an attempt to more precisely monitor the overall state of health of a particular station. In addition to incorrect instrument response, out-of-nominal noise conditions currently being characterized include: 1) calibration pulses, 2) gaps or missing data, 3) spikes, and 4) sensor recentering. Figure 7 is an example of a common excursion from the station baseline due to recentering of the Streckeisen STS2 seismometer at the ANSS backbone station US.LRAL.--.BHZ. The PDF and long-term noise baselines in Figure 7(A) for US.LRAL.--.BHZ are computed from 7,737 one-hour PSDs occurring from 1 March 2007 (2007:060) to 31 December 2007 (2007:365). Using the PQLX software, an analyst can select groups of similar PSDs (Figures 7A and 7B) and define a noise-source model by storing the characteristic maximum and minimum (Figure 7B). The STS2 recentering noise source model in Figure 7(B) is composed of 46 PSDs representing 0.6% of the station’s operation time from 1 March 2007 (2007:060) to 31 December 2007 (2007:365). PQLX also allows a user to view the time series segments through a client interface to help determine the source and characteristics of the noise excursion (Figure 7C). Steps described here allow a user to visually define the spectral characteristics of known system transients for comparison against the long-term station baselines.

In the prototype automated system, when an hourly PSD falls outside the long-term station baseline, it is flagged and then compared to the set of known noise source models for that channel (including comparison to the NEIC real-time catalog to identify earthquakes). Figure 8 shows the hourly PSD percentage fits to the baseline (Figure 8A) and several defined noise models (Figures 8B–8D) for US.LRAL.--.BHZ for two months in the year 2007. Noise models include: 1) STS2 mass
Figure 8: This figure demonstrates the fit of hourly PSDs from 2008:200 to 2008:280 to the station baseline (A) and the known noise source models: (B) STS2 re-center, (C) calibration pulse, and (D) negative spikes due to excessive gapping.

In general, the percent fit of the hourly PSDs to the long-term baseline can vary from 100% to 50% due to the occurrence of earthquakes and other complex noise sources. Larger deviations, 0–10% fit to the baseline, are observed to occur when known noise source models have percent fits near 100%. For example, Figure 8(B) shows two clear instances where STS2 recenter fits approach 100% and correspond to where baseline fits drop to zero (Figure 8A). Figure 8(C) shows clear correlations of the weekly recurring calibration pulse to large misfits to the baseline (Figure 8A). Figure 8(D) shows that the system detects gapping problems when there are clear misfits to the baseline (Figure 8A). In the prototype system, if a match of at least 75% occurs, a noise-source detection is declared, stored, and ultimately compiled in a report for further investigation. In many cases the underlying cause of a noise condition cannot be determined and these occurrences are stored for future study.

SUMMARY

We present a method for quantifying station noise baselines and characterizing the spectral shape of out-of-nominal noise sources. Our intent is to automate this method in order to ensure that only the highest-quality data are used in rapid earthquake products at NEIC. In addition, the station noise baselines provide a valuable tool to support the quality control of GSN and ANSS backbone data and metadata. The procedures addressed here are currently in development at the NEIC, and work is underway to understand how quickly changes from nominal can be observed and used within the NEIC processing framework. The spectral methods and software used to compute station baselines and described herein (PQLX) can be useful to both permanent and portable seismic stations operators. Applications include: general seismic station and data quality control (QC), evaluation of instrument responses, assessment of near real-time communication system performance, characterization of site cultural noise conditions, and evaluation of sensor vault design, as well as assessment of gross network capabilities (McNamara et al. 2005). Future PQLX development plans include incorporating station baselines for automated QC methods and automating station status report generation and notification based on user-defined QC parameters. The PQLX software is available through the USGS (http://earthquake.usgs.gov/research/software/pqlx.php) and IRIS (http://www.iris.edu/software/pqlx/).

ACKNOWLEDGMENTS

The authors would like to thank all network operators and staff who contribute data to the NEIC, ASL QC staff for a steady stream of ideas on improving QC techniques, and operational
system support. The authors thank Richard Boaz for his programming and IRIS and the USGS NEIC for supporting PQLX software development. Maps were created using GMT mapping tools (Wessel and Smith 1991).

REFERENCES


USGS/ANSS/NEIC

U.S. Geological Survey

Box 25046, Stop 966

Denver Federal Center

Denver, Colorado 80225, U.S.A.

mcnamara@usgs.gov

(D.E.M.)