Analysis of the Port Townsend Underwater Acoustic Background

Mark L. Stockham
Department of Mechanical Engineering
University of Washington
Seattle, Washington 98195
Email: stockham@uw.edu

Peter H. Dahl
Applied Physics Laboratory
and Department of Mechanical Engineering
University of Washington
Seattle, Washington 98195

Per G. Reinhall
Department of Mechanical Engineering
University of Washington
Seattle, Washington 98195

I. Executive Summary

The regulation of anthropogenic underwater acoustic noise equires that the background levels be measured and undertood as a baseline for comparison. Seven days of data from ort Townsend, WA have been compiled and analyzed with bootstrap-inspired approximation of standard deviation of everal percentile measures as a function of sample size as substitute for a traditional power analysis. Because of the ighly varying nature of the underwater acoustic background t is recommended to get at least three days of data to ensure 5% confidence levels of approximately 2 dB reference 1 icropascal or less.

II. Introduction

It is important to know the background level in an underwater acoustic environment for comparison to many regulated noise-producing activities (pile driving for one). Also of interest is how much data must be collected from a site to be confident that any statistics from that data set are representative of the true values of the statistics at that site. There is an exact formula for the standard error of the mean, given by,

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}}, \tag{1}$$

where s is the sample standard deviation and n is the sample size. As [1] points out, however, there are no such elegant equations for some other statistics of interest, such as the standard error of the sample median or the other percentiles. Bootstrapping, a computationally intensive method described in [1], allows for a single technique to be used to determine the standard error of any estimator and to also infer a confidence interval.

III. Data Collection

More than seven whole days of background data were collected from near the Port Townsend, Washington ferry terminal starting on the afternoon of April 19, 2010. The AMAR, developed by JASCO Research, was tethered to the sea floor with a hydrophone sensitive in the bandwidth of 1 hertz to 10 kilohertz.

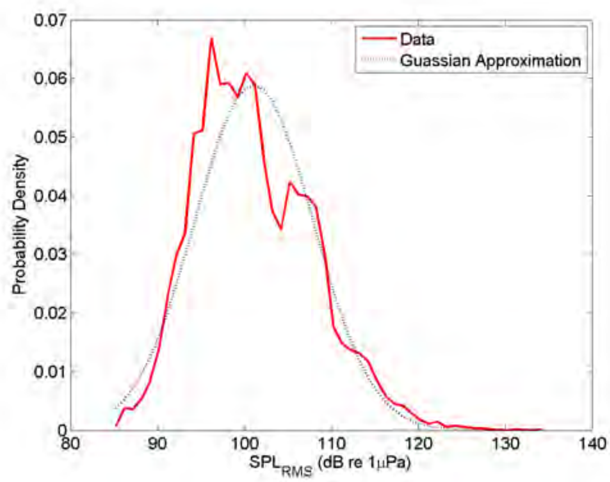

Fig. 1. The empirical probability density function of all seven days compared with a log-normal distribution that has a mean of 101.1 dB and a standard deviation of 6.8 dB.

IV. Data Analysis

The data was digitally filtered through a passband of 1-10 kilohertz designed using MATLAB® and the RMS pressure was calculated for every consecutive 30 seconds of data. Only the seven complete days were used, giving a data set of 20,160 points. The distribution of these RMS data points is approximately log-normal (Fig. 1). However, the daily distributions show large variations (Fig. 2.

A. Sliding Bootstrap

Bootstrapping makes use of resampling a data set to obtain an approximation of a given estimator. From the data set of sample size N a random sample (with replacement) is taken, also of size N. This is done B times and each time the estimator is evaluated. Then the standard deviation of this estimator is taken to calculate the standard error for that estimator in the original sample. Equation (1) shows the exact equation of the standard error of the mean. The bootstrap equation for the standard error of the mean would be given by

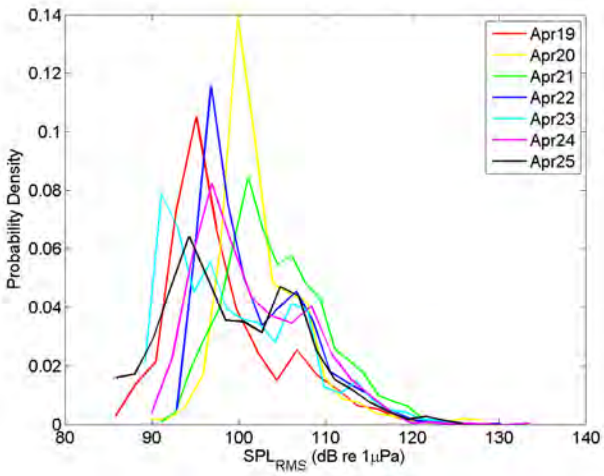

Fig. 2. The approximately log-normal distribution of the data for each day.

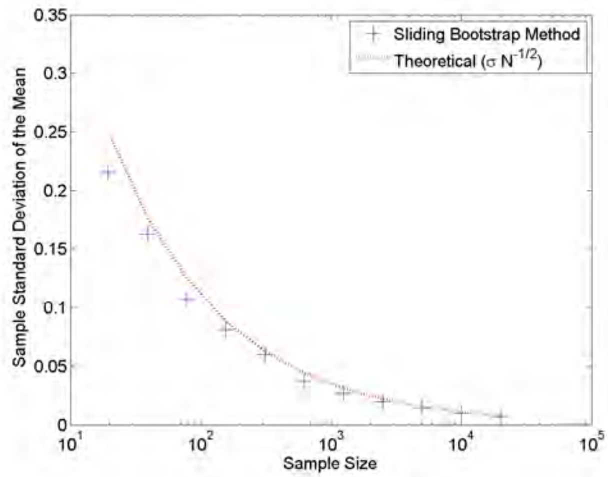

Fig. 3. Comparison of the bootstrap estimate of the standard error of the mean and that given by Eq. (1). This is for computer generated random numbers with μ = 0 and σ = 1.

$$\hat{\sigma}_B = \left(\frac{\sum_{b=1}^{B} \{\hat{\mu}^*(b) - \hat{\mu}^*(\cdot)\}^2}{B-1} \right)^{1/2}, \tag{2}$$

where

$$\hat{\mu}^*(\cdot) = \frac{\sum_{b=1}^{B} \hat{\mu}^*(b)}{B}. \tag{3}$$

$\hat{\mu}$ (b) b

this analysis B = 512 was used, since there was no appreciable difference between results with $B > 512$.

Due to the non-stationarity of the acoustic background at Port Townsend (Fig. 2), it is proposed to use a different method, here called a sliding bootstrap. Out of the seven days of data a single point is chosen using a random number generator in MATLAB®. Starting at this point the statistic of interest (the mean, for example) is calculated for a sample of length N. If a sample of length N will not fit between the random point chosen and the end of the data set, a new random point is chosen. This is done until there have been B calculations made of the statistic. The standard error is then the standard deviation of that statistic, in the same way that it is for bootstrapping. This is an attempt to not understate how much the acoustic environment varies is on a daily basis. That is, if N is equal to two days this method attempts to predict the standard error for any two days, not necassarily two actual days from the data set. As an example, Fig. 3 shows the standard error of the mean of computer generated random numbers as a function of sample size using the sliding bootstrap method with B equal to 512. Because this process is stationary, it agrees with the exact form, Eq. (1). Because the underwater acoustic background is non-stationary, this agreement is not present in the data until a large sample size, which is the near the region where the process approaches stationary (on the order of a week), this is shown in Fig. 4

Inserting the values of the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, 95^{th} percentiles of the data into Eq. (2) gives the standard error on each of those estimates. This is plotted as a function of sample size of the data in Fig. 5.

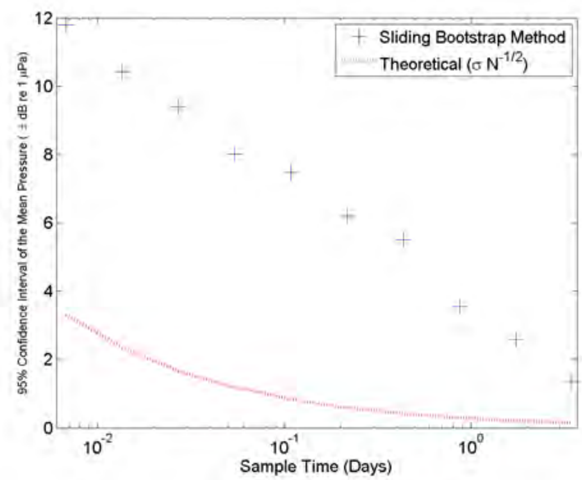

Fig. 4. Comparison of the 95% confidence values calculated using the sliding bootstrap estimate of the standard error of the mean of the data and that given by Eq. (1).

V. Discussion

The confidence interval for a given parameter, x, is given by

$$x \in \hat{x} \pm \hat{\sigma}_B z^{(\alpha)}, \tag{4}$$

where $z^{(\alpha)}$ is the 100·α percentile point of a standard normal distribution. Table IV gives 95% confidence levels of the parameters discussed above (calculated from N = 3.5 days).

VI. Summary and Conclusions

Several useful statistics of the Port Townsend ferry terminal acoustic background were given with associated confidence

TABLE I

STATISTICS OF PORT TOWNSEND ACOUSTIC BACKGROUND (FOR SAMPLE TIME APPROXIMATELY 3.5 DAYS)

Parameter	Value (dB)	95% Confidence Interval (dB)
Mean	101.7	±1.3
5^{th} Percentile	92.3	±2.4
10^{th} Percentile	94.0	±2.3
25^{th} Percentile	96.9	±1.9
50^{th} Percentile (median)	100.7	±1.7
75^{th} Percentile	106.2	±0.7
90^{th} Percentile	110.4	±0.9
95^{th} Percentile	113.7	±0.7

TABLE II

STATISTICS OF PORT TOWNSEND ACOUSTIC BACKGROUND (FOR SAMPLE TIME APPROXIMATELY 2 DAYS)

Parameter	Value (dB)	95% Confidence Interval (dB)
Mean	101.6	±2.6
5^{th} Percentile	93.2	±4.3
10^{th} Percentile	94.4	±4.2
25^{th} Percentile	97.1	±3.5
50^{th} Percentile (median)	100.6	±2.8
75^{th} Percentile	106.2	±1.9
90^{th} Percentile	110.4	±1.8
95^{th} Percentile	113.6	±1.8

TABLE III

STATISTICS OF PORT TOWNSEND ACOUSTIC BACKGROUND (FOR SAMPLE TIME APPROXIMATELY 1 DAY)

Parameter	Value (dB)	95% Confidence Interval (dB)
Mean	101.3	±3.5
5^{th} Percentile	93.2	±5.1
10^{th} Percentile	94.4	±4.9
25^{th} Percentile	96.6	±4.9
50^{th} Percentile (median)	100.0	±3.9
75^{th} Percentile	105.8	±3.1
90^{th} Percentile	110.3	±2.6
95^{th} Percentile	113.3	±3.0

TABLE IV

STATISTICS OF PORT TOWNSEND ACOUSTIC BACKGROUND (FOR SAMPLE TIME APPROXIMATELY HALF A DAY)

Parameter	Value (dB)	95% Confidence Interval (dB)
Mean	101.1	±5.5
5^{th} Percentile	93.3	±6.4
10^{th} Percentile	94.4	±5.9
25^{th} Percentile	96.6	±5.7
50^{th} Percentile (median)	99.8	±6.4
75^{th} Percentile	104.9	±6.9
90^{th} Percentile	109.4	±6.2
95^{th} Percentile	112.4	±5.9

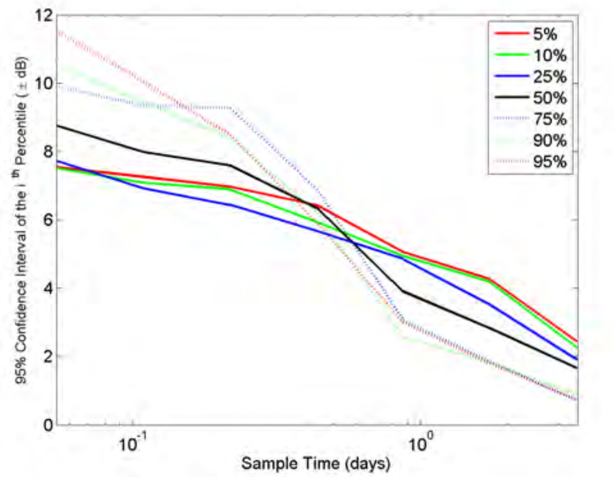

Fig. 5. Sliding bootstrap estimates of the 95% confidence intervals in percentiles as a function of sample size. B = 512.

levels. The increase in confidence after two to three days of sampling becomes negligible, so (assuming stationarity on the order of a week) sample times longer than this are not recommended. In fact, it is highly likely that the uncertainties in the measurements would dominate over the standard error once the standard error goes below 1 to 2 decibels, which happens for nearly all of the above mentioned statistics after three and a half days of sampling. While several hours of data would be sufficient statistically if the background environment was stationary, it fails in reality because of the dynamic nature of the background (see Appendix for comparison of percentile values from each of the seven days). It is recommended to get at least three full cycles of 24 hours to account for the differences in the background due to the ferries and other traffic, which operate on a daily schedule.

ACKNOWLEDGMENT

Work was funded by the Washington State Department of Transportation.

REFERENCES

[1] B. Efron, R. Tibshirani. "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy". *Statistical Science* vol. 1, no. 1, pp. 54-77, 1986.

APPENDIX

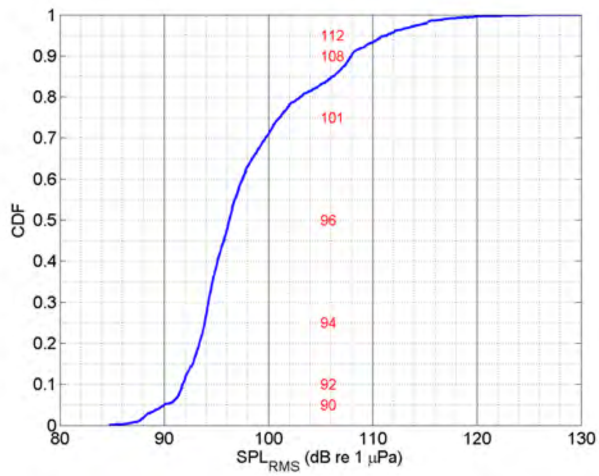

Fig. 6. Cumulative distribution function for the first day of data.

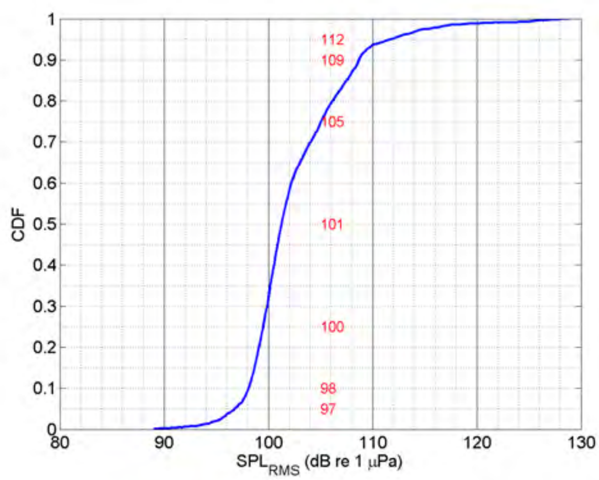

Fig. 7. Cumulative distribution function for the second day of data.

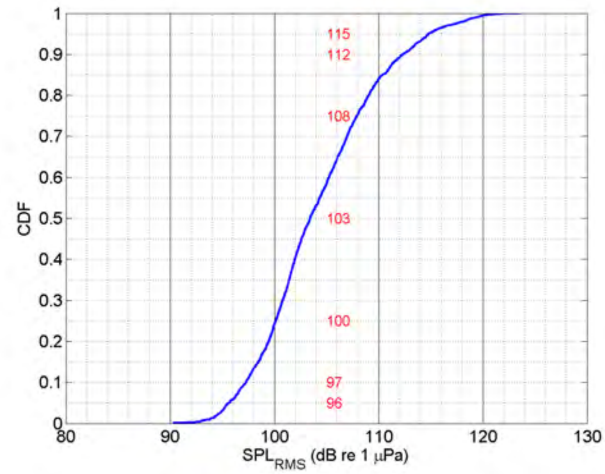

Fig. 8. Cumulative distribution function for the third day of data.

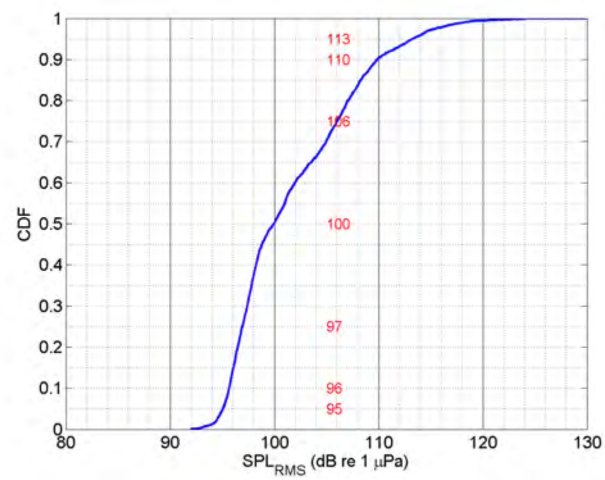

Fig. 9. Cumulative distribution function for the fourth day of data.

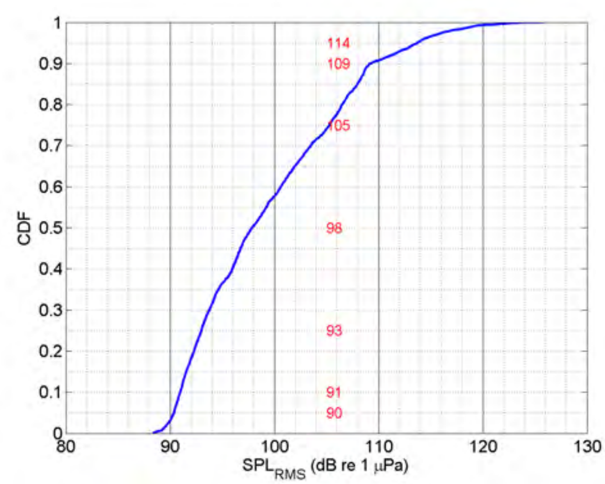

Fig. 10. Cumulative distribution function for the fifth day of data.

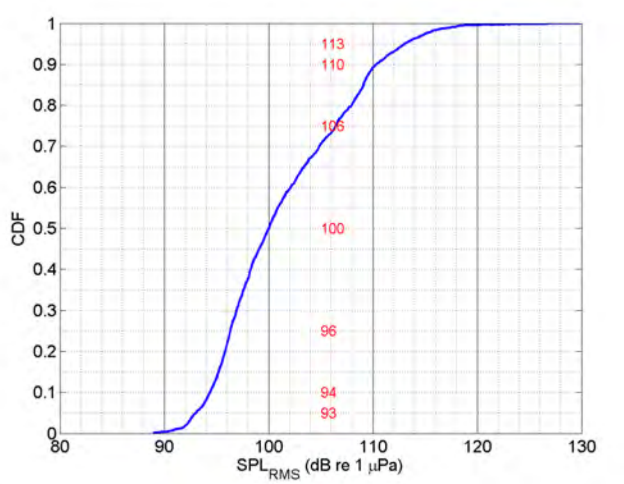

Fig. 11. Cumulative distribution function for the sixth day of data.

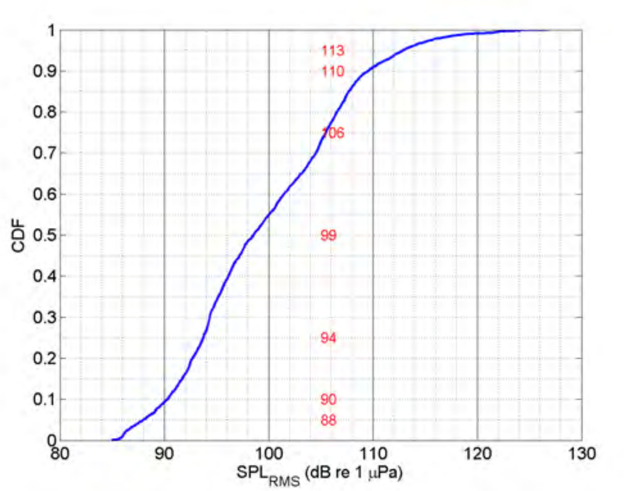

Fig. 12. Cumulative distribution function for the seventh day of data.