24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

IDENTIFICATION AND ASSESSMENT OF LONG TERM TRENDS IN UNDERWATER NOISE MEASUREMENTS

Kostas Sotirakopoulos, Peter Harris, Stephen Robinson, Lian Wang and Valerie Livina

National Physical Laboratory, Hampton Road, Teddington, UK
email: kostas.sotirakopoulos@npl.co.uk

Increased scientific and societal concern about the effects of underwater sound on marine life have recently been recognised through the introduction of several international initiatives and international regulation. Monitoring of deep-ocean low-frequency sound is challenging, but data have been reported for the Northeast Pacific Ocean and Indian Ocean in recent publications. The CTBTO (Preparatory Commission for the Comprehensive Nuclear-Test-Ban Treaty Organization) has made available data from their deep-ocean hydro-acoustic stations, so that researchers may examine the existence of trends and features in the recorded sound. In the case of some of the stations, the data cover more than ten years of recordings. In this paper, we present trend analysis of data from one CTBTO observatory at Cape Leeuwin (Australia) to examine the rate and magnitude of change in low frequency sound (5-105 Hz) over the period 2003 - 2015. The analysis involves the application of regression to percentile levels in limited frequency bands and employs bootstrap resampling as a non-parametric approach for the necessary quantification of the uncertainties associated with the estimated trends. Results obtained by linear and more complex regression models are compared and the effect of aggregating data over various time intervals is also examined. Finally comparisons are drawn between trends observed in adjacent frequency bands.

Keywords: Trend Analysis, Ocean Noise, CTBTO, Uncertainty Assessment

1. Introduction

In an era where human activity impacts greatly the planet's environment and bio-diversity, monitoring of environmental factors that quantify this effect becomes more necessary than ever. Over the past decades increasing demand for energy together with international commerce and technological growth have dramatically contributed to the increase of off-shore activity and the pollutants it produces.

Human off-shore activity is commonly related to transportation (shipping), geo-physical surveys for gas and oil extraction, and infrastructure construction (pile-driving), all of which contain high-energy low-frequency sound sources. Over the last century shipping traffic has grown substantially [1], [2], [3] and so has the size of the ships themselves [4]. Though each individual ship is not in itself a high-amplitude source, their large and increasing number has the potential to raise ocean noise levels. At the same time, sound produced by oil and gas exploration and production as well as the construction and operation of infrastructure for renewable energy sources is continuously increasing [5], [6], [7].

Since noise is a form of pollution that has an impact on marine life [8], [9], [10], [11], it is reasonable to try to quantify the human contribution to the existing ocean noise levels. Working towards this aim, the European Marine Strategy Framework Directive [12] has identified noise as an important factor for the assessment of the oceans' environmental status. Furthermore, the Commission Decision of 2010 [13] describes indicators for the assessment of impulsive and low frequency continuous sound for the identification of events and trends at low-frequency bands where human activity could potentially have greatest effect. Analysis of trends in ocean noise requires the collection and processing of large datasets over long time periods. Emerging technologies like sensor networks and tools for big data analysis have made possible the acquisition and thorough examination of large data volumes. However, the deployment and maintenance of underwater acoustic equipment still involves prohibitive costs making long term monitoring difficult to implement.

The Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) have created a global network of sensors which includes 11 hydro-acoustic stations located in all the major oceans [14]. CTBTO have made available data from their stations which consist of triads of hydrophones placed in the ocean's deep-sound-channel with inter-separation of two kilometres. The data consist of low frequency (250 Hz sampling rate, 24 bit depth) continuous recordings of sound pressure and in some cases span over more than a decade. Hence they are suitable for long term trend analysis and have been the source of interest for several studies in the recent past [15], [16], [17], [18].

This paper focuses on the examination of data collected by one of the three hydrophones at CTBTO station H01W at Cape Leeuwin, Australia, for the identification of trends over 12 years (2003 - 2015). The sensor depth for station H01W is 1055 m in a water column of 1558 m depth. One of the main questions such trend analysis is intended to address is whether deep-ocean noise has increased or decreased with time. Although this question appears to be straightforward, in practice long-term trends are difficult to quantify accurately because of the presence of large seasonal variations and other effects in the data. In order to identify possible changes and compute the uncertainties associated with the estimates, two different regression models are considered and confidence intervals for each outcome are constructed using bootstrap resampling [19], [20] of the residuals from the regression. In the following sections, the data processing techniques are briefly described and emphasis is given to the comparison of the results obtained by the different models.

2. Method

The period of interest for the presented trend analysis covered 12 years between 01/01/2003 and 01/01/2015. The abundance of data provided allowed great flexibility in the approach for the analysis. However, at the same time the sheer size of the datasets made the extraction from the database, storing, handling and processing of the data a quite challenging task. Hence the reduction of raw data to aggregated values was considered necessary. The data samples were scaled using their accompanying calibration factors provided by CTBTO and an inverse filter of the recording system's frequency response was applied to eliminate the effect of the acquisition chain on the frequency response of the recordings. The broadband signal was then filtered in 5 frequency bands (5-115 Hz, 10-30 Hz, 40-60 Hz, 56-70 Hz, 85-105 Hz) and the squared pressures were averaged over 10 min intervals and transformed into dB re 1 μPa^2. The outcome for the 5-115 Hz band is presented in Figure 1.

Finally, any outliers, i.e. levels greater than 20 dB from the average of the entire time series, were removed to allow better examination of the underlying trend in the data.

Even though this procedure produced much more manageable datasets it was considered that the intended trend analysis would not be heavily influenced by further data reduction as long as the aggregation intervals would not interfere with the seasonal characteristics in the time series. Inspection of the seasonal patterns of the data shown in Figure 1 revealed that further reduction to a daily value would not have a major impact on the trend estimation given the still significant sample size.

Together with the daily averages the 1st, 5th, 50th, 90th, and 99th daily percentiles (P_1, P_5, P_{50}, P_{90} and P_{99}) were also computed based on the previously calculated 10 min averages. These summaries would allow examination of parts of the time series related to ambient (P_1, P_5) and higher (P_{90} and P_{99}) noise levels. At the same time the median would be a more representative parameter than the mean for the middle part of the distribution considering the skewness that can be inferred by examination of Figure 1.

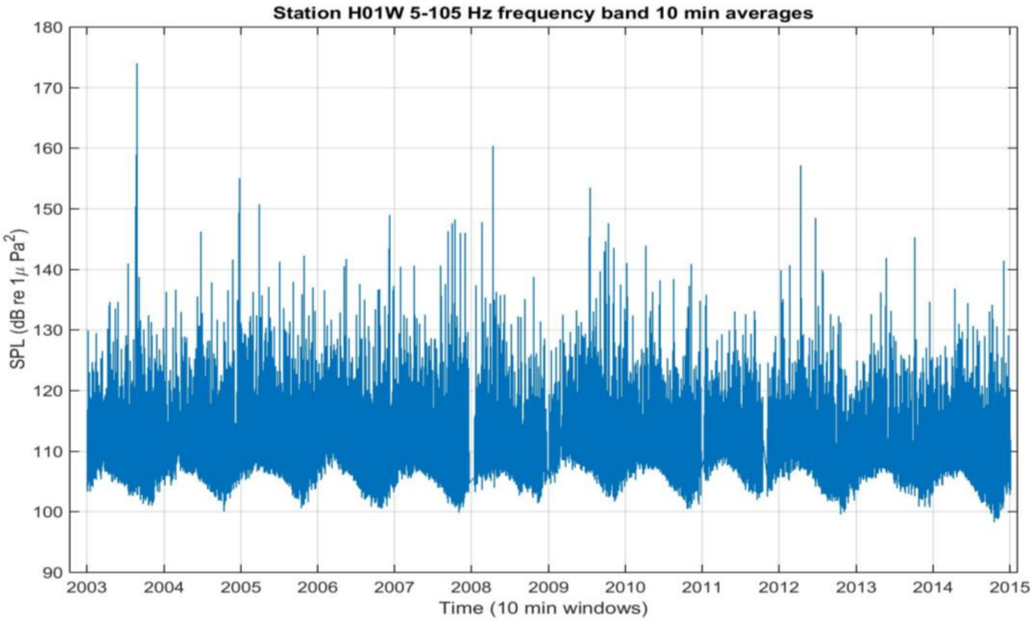

Figure 1, 5-105 Hz band 10 min averages between 01/01/2003 and 01/01/2015.

A very common approach to estimating long term trends is to fit a straight line (Linear Model) and use the gradient as an estimator for the long term trend. Then the associated confidence interval for a stated confidence level (95 % or 99 %) can be evaluated using Gaussian statistics under the assumption that the residual errors associated with the fitted straight line are normally distributed. However, the Linear Model approach is somewhat simplistic for the examined data as it does not take explicit account of seasonal variations. At the same time it is statistically unsound because the data are serially correlated and the residual errors from the model depart appreciably from normality.

An alternative approach more suitable for the examined data involves a class of regression models incorporating seasonal factors (Seasonal Model) used to describe annual periodic trends in addition to an underlying long term trend. This approach significantly improves the regression fit and leads to lower residual errors that appear more normally distributed around zero. Still, however, the conditions for the Central Limit Theorem are not satisfied since the data are serially correlated and so no assumption of normality for the sampling distribution of the estimator can be made. Hence we use a non-parametric approach, specifically bootstrap resampling, to estimate confidence intervals associated with the gradient. The described process involves repeated resampling of the residual errors to generate new datasets that can be fitted with regression models in order to establish the sampling distribution for the estimator. Finally the middle 95 % (or 99 %) of the sampling distribution constitutes the required confidence interval.

Regression with bootstrap resampling of size 10 000 was applied to P_{50} levels for each of the 5 frequency bands using Linear and Seasonal Models in order to examine the differences in the obtained trend estimates and their associated confidence intervals. Also, it was considered interesting to investigate the impact of aggregating over longer time intervals. This was achieved by application of the same procedure to weekly and monthly P_{50} values derived from the initial 10 min windowed dataset.

3. Results and Discussion

The first step of the performed trend analysis was the application of least-squares regression using both Linear and Seasonal Models to the daily, weekly and monthly P_{50} values and examination of the residual errors. The difference between the two model fits is illustrated in Figure 2 where the regressions to the daily P_{50} values for the 10-30 Hz frequency band are shown. Simple inspection of the graph reveals that the Seasonal Model fits the data much better and so it is expected to present significantly lower residual errors.

The residual errors were collected and categorised by year in order to inspect each year's contribution to the overall distribution that was then used as the population for the bootstrap resampling. The upper two plots of Figure 3 present boxplots of the annual residual error distributions while the lower graph shows the resulting overall residual error distributions for the two models. As inferred from Figure 2**,** the Seasonal Model, which incorporates annual seasonalities, appears to yield a much narrower residual error distribution that is also more symmetrical. Both histograms have a high peak at zero, however the Linear Model's distribution is significantly more skewed with a higher spread around the main peak. This implies a poorer fit and its effects are expected to appear in the estimate of the slope as well as the associated confidence interval.

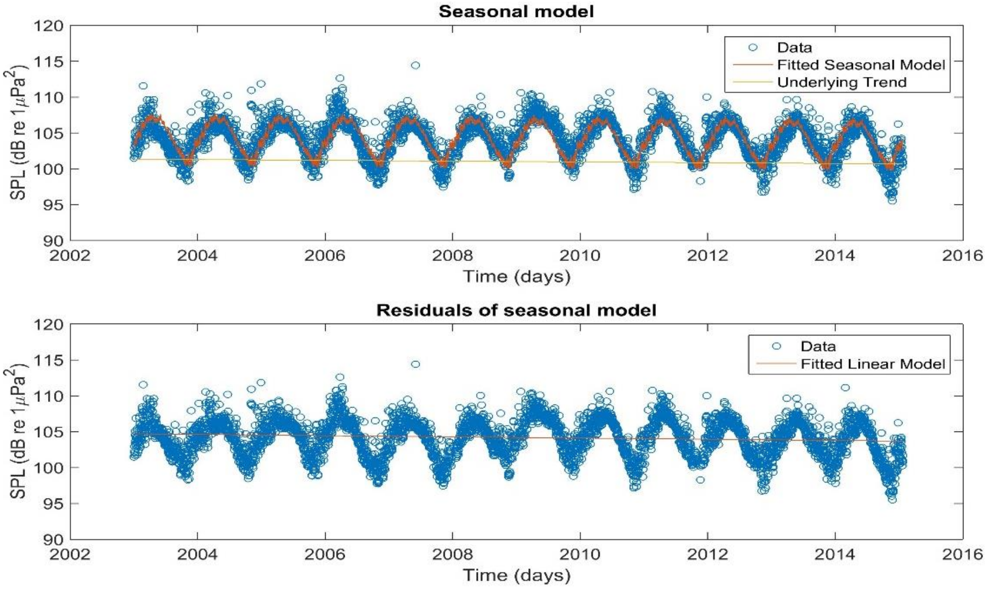

Figure 2, Daily P_{50} 10-30 Hz band data with Seasonal and Linear regression models fitted.

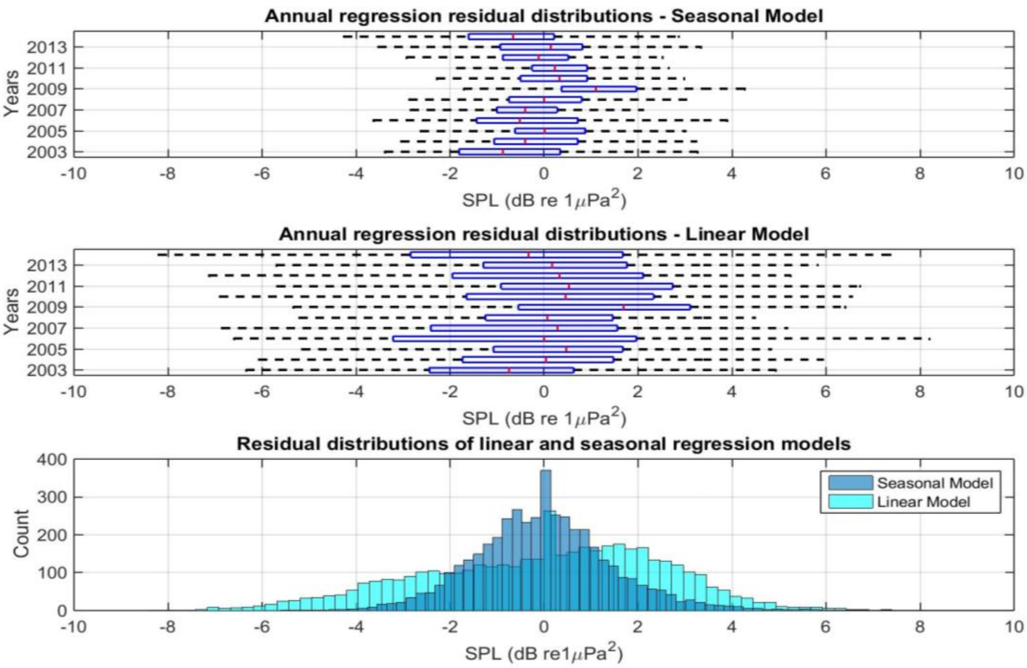

Figure 3, Annual and overall residual error distributions from applied regressions on the 10 – 30 Hz frequency band.

The distributions of Figure 3 were used as populations from which samples were drawn with replacement in order to form new datasets on which regression was applied. This procedure was repeated 10 000 times in order to allow a healthy sample size for the sampling distribution of the estimated trend. The resulting values of the rate of change in dB/year are presented in Figure 4.

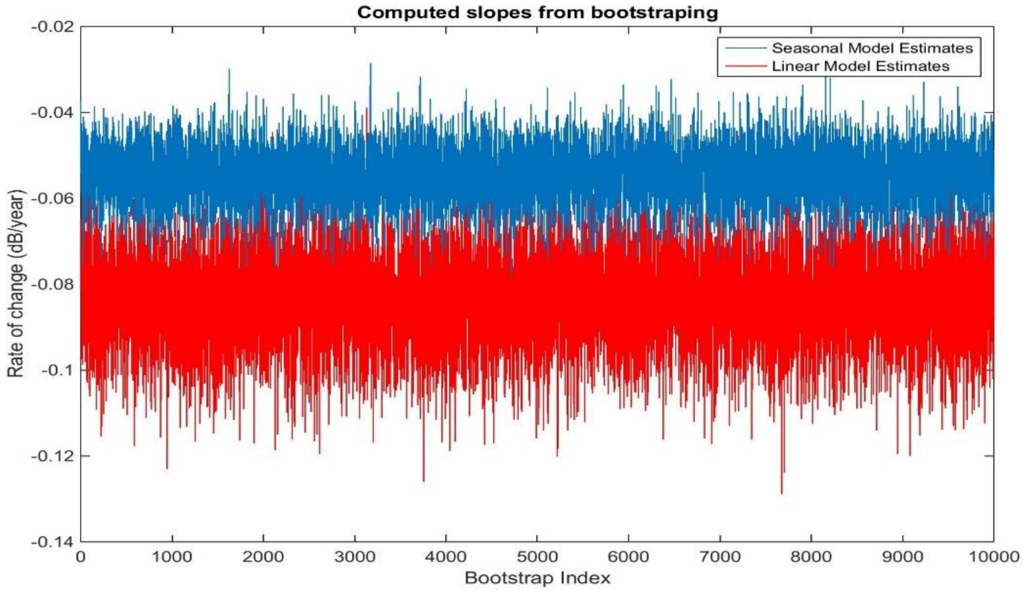

Figure 4, Computed slopes for the 10-30 Hz frequency band from bootstrap resampling.

Two important features are obvious in the above graph. The first is a difference in the mean values around which each dataset is spread. Specifically the Linear Model appears to systematically give lower estimates of the trend than the Seasonal Model. This happens because the Linear Model does not describe adequately the deterministic part of the data and exhibits a much worse fit than the Seasonal Model, as the inspection of the residual errors also showed. The second characteristic is a difference in the spread with the Linear Model's estimates presenting higher variance. This is inherited from the higher spread in the initial regression's residual errors.

Using the central 95 % of the estimated slope distributions the 95 % confidence interval for each model's slope estimate was constructed. Then the magnitude of change was computed by multiplying by the number of years over which the regression was applied. The same procedure was performed for weekly and monthly P_{50} values and the resulting estimates for the 10-30 Hz frequency band are shown in Figure 5.

Figure 5 shows that as the aggregating interval increases (from daily to monthly) the less precise the estimates become. In time series analysis it is common practice to use averaging or similar aggregation techniques as a way of removing noise and smoothing data. However such techniques must be used carefully as smoothing can hide important characteristics of the data and in combination with the reduction of the sample size can eventually impact heavily the uncertainties for the estimated parameter. This effect is exaggerated when models that do not describe the data accurately are used as we see by comparing the estimates based on the monthly P_{50} values. Moreover, the discussed offset between the estimates achieved by the two models appears for all different aggregation intervals while the precision of the Seasonal Model is always higher.

In order to better understand the importance of choosing an appropriate model and aggregation interval for the estimation of long term trends one should compare the results obtained by the Linear Model using monthly P_{50} values to those provided by the Seasonal Model based on daily values. It is clear that based on the same dataset the conclusions drawn by the two approaches can be quite different since for the Linear Model the null hypothesis that there is no significant trend in the data cannot be rejected while the opposite happens when the Seasonal Model is considered. That is because zero is included in the 95 % confidence interval for the monthly based Linear Model while this is not the case when a daily based Seasonal Model is considered.

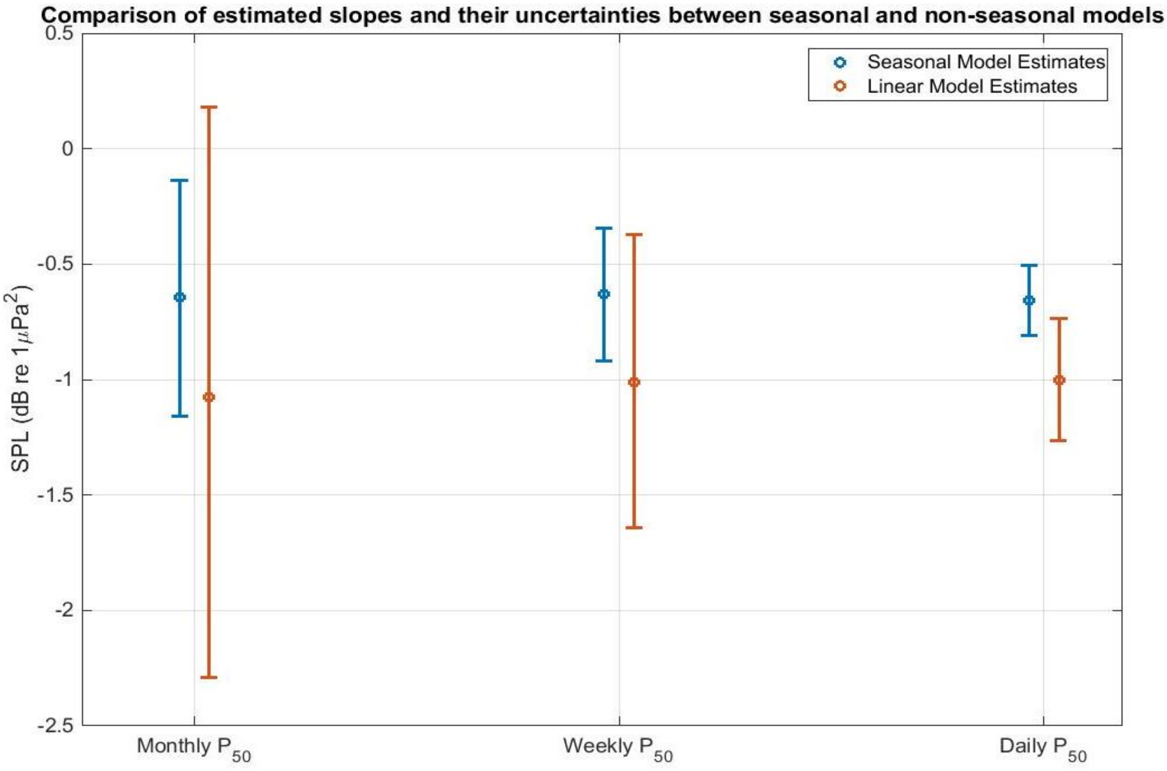

Figure 5, Confidence intervals for the estimated slopes for the 10-30 Hz frequency band using the two regression models for three different aggregation times.

Finally, application of the described trend analysis to band-limited frequency intervals allowed the examination of possible changes in the frequency content of the noise. The results are summarised in Table 1.

Table 1, Estimated magnitude of change in dB re 1μPa^2 between 2003 and 2015 at 5 frequency bands using both regression models. All presented estimates are based on daily P_{50} levels.

Frequency Band (Hz)	Linear Model			Seasonal Model		
	Lower Limit	Slope Estimate	Upper Limit	Lower Limit	Slope Estimate	Upper Limit
5 - 105	-1.49	-1.26	-1.03	-1.13	-0.97	-0.82
10 - 30	-1.26	-1.00	-0.73	-0.80	-0.65	-0.49
40 - 60	-1.76	-1.56	-1.36	-1.49	-1.35	-1.21
56 - 70	-1.01	-0.80	-0.60	-0.76	-0.61	-0.45
85 -105	-0.09	0.08	0.26	0.04	0.20	0.36

Table 1 shows a common decrease in all frequency bands apart from the 85 -105 Hz band where a marginal increase is seen. The most significant level decrease appears in the 40 - 60 Hz band which is commonly associated with shipping noise. An interesting observation is that the Linear Model's systematic underestimation of the slope failed to identify the increase in the 85 -105 Hz frequency band as statistically significant yielding a Type II error while the increase was successfully detected by the Seasonal Model.

4. Conclusions

Trend analysis of underwater low frequency noise recorded by CTBTO station H01W at Cape Leeuwin was performed using a Seasonal Model incorporating annual seasonal trends and a less complex Linear Model. Then the uncertainties associated with each estimate were computed using bootstrap resampling of the residual errors from the regressions. The results were compared in order to explore the reliability of the estimates that can be achieved. Moreover the impact of aggregating data over longer time intervals on the precision of the estimated trends was discussed and possible changes in the frequency content of noise between 2003 and 2015 were explored.

It was shown that estimates of long term trend can be more reliable when seasonal characteristics of the data are incorporated in the regression model. Precision decreases as the data sample size decreases something that gets even more exaggerated by the utilisation of the simple Linear Model. On the other hand, use of more appropriate models with temporal resolution that maintains the original data's seasonal characteristics provides substantially improved estimates. From the described analysis it was concluded that daily P50 noise levels recorded at Cape Leeuwin decreased between 2003 and 2015 at all but one of the examined frequency bands with the 40-60 Hz band presenting the highest change.

Computation of the uncertainties associated with estimates is a fundamental requirement for statistical inference as it provides a degree of confidence for the precision of that estimate. Neglecting this step leads to estimates with no indication of their quality which can potentially lead to wrong conclusions. This can have serious implications if these conclusions are then used as the basis for further actions including policy making decisions.

5. Acknowledgements

This project has been funded by Department for Business, Energy and Industrial Strategy of the United Kingdom.

6. References

[1] Asariotis, R., Benamara, H., Finkenbrink, H., Hoffmann, J., Lavelle, J., Misovicova, M.,, “Review of maritime transport 2011. Technical Report E.11.II.D.4.,” United Nations conference on trade and development., 2011.

[2] McDonald M.A., Hildebrand J.A., Wiggins S.M., “Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California.,” *J. Acoust. Soc. Am.,* vol. 120, no. 2, pp. 711-8, 2006.

[3] Ross, D., “Ship sources of ambient noise,” *IEEE J. Ocean. Eng.,* vol. 30, pp. 257-261, 2005.

[4] Hildebrand, J., “Anthropogenic and natural sources of ambient noise in the ocean.,” *Mar. Ecol. Prog. Ser.,* vol. 395, pp. 5-20, 2009.

[5] Nieukirk, S., Mellinger, D., Moore, S., Klinck, K., Dziak, R., Goslin, J, “Sounds from airguns and fin whales recorded in the mid-Atlantic ocean 1999-2009,” *J. Acoust.Soc. Am.,* vol. 131, pp. 1102-1112, 2012.

[6] Potter, J., Thillet, M., Douglas, C., Chitre, M., Doborzynski, Z., Seekings, P., “Visual and passive acoustic marine mammal observations and high-frequency seismic source characteristics recorded during a seismic survey,” *IEEE J. Ocean. Eng,* vol. 32, pp. 469-483, 2007.

[7] Madsen, P.,Wahlberg, M., Tougaard, J., Lucke, K., Tyack, P., “Wind turbine underwater noise and marine mammals: implications of current knowledge and data needs,” *Mar. Ecol. Prog. Ser.,* vol. 309, pp. 279-295, 2006.

[8] Southall, B., Bowles, A., Ellison, W., Finneran, J., Gentry, R., Greene, C., Kastak, D., Ketten, D.,, "Marine mammal noise exposure criteria: initial scientific recommendations.," *Aquat. Mamm.,* vol. 33, pp. 411-521, 2007.

[9] Holt, M. M., Noren, D. P., Veirs, V., Emmons, C. K., and Veirs, S., "Speaking up: Killer whales (Orcinus orca) increase their call amplitude in response to vessel noise," *J. Acoust. Soc. Am.,* vol. 125, p. EL27–EL32, 2009.

[10] Charif, R., Mellinger, D., Dunsmore, K., Firstup, K., Clark, C., "Estimated source levels of fin whale (Balaenoptera physalus) vocalizations: adjustments for surface interference," *Mar. Mamm. Sci.,* vol. 18, pp. 81-98, 2002.

[11] Mooney, T., Nachtigall, P., Breese, M., Vlachos, S., Au, W., "Predicting temporary threshold shifts in bottlenose dolphin (Tursiops truncatus): the effects of noise level and duration," *J. Acoust. Soc. Am.,* vol. 125, pp. 1816-1826, 2009a.

[12] European Parliament and the Council of the European Union, "Directive 2008/56/ec of the European parliament and of the council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive)," *Off. J. Eur. Union,* vol. L164, pp. 19-40, 2008.

[13] European Commission, 2010, "Commission Decision of 1 September 2010 on Criteria and Methodological Standards on Good Environmental Status of Marine Waters. (2010/477/eu)," Available from: http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:232:0014:0024:EN:PDF.

[14] deGroot-Hedlin, C., and Orcutt, J. (Eds), "Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Hydroacoustics," *Pure and Applied Geophysics,* vol. 158, no. 3, pp. 421-626, 2001.

[15] Miksis-Olds, J., L., Bradley,D., L., Niu, X., M., "Decadal trends in Indian Ocean ambient sound," *J.Acoust.Soc.Am,* vol. 134, no. 5, pp. 3464-3475, 2013.

[16] Prior,M.,K., Brown, D., J., Haralabus, G., Stanley, J., "Long-term monitoring of ambient noise at CTBTO hydrophone stations," in *Proceedings of the 11th European Conference on Underwater Acoustics*, Edinburgh, 2012.

[17] Sabra,K., G., Fried, S., Kuperman, W., A., Prior, M., K., "On the coherent components of low-frequency ambient noise in the Indian Ocean," *J. Acoust. Soc. Am.,* vol. 133, no. 1, p. EL25, 2012.

[18] van der Schaar, M., Ainslie,M., A., Robinson, S., P., Prior, M., K., André, M., "Changes in 63 Hz third-octave band sound levels over 42 months recorded at four deep-ocean observatories," *Journal of Marine Systems,* vol. 130, pp. 4-11, 2014.

[19] Efron, B., Tibshirani,R., An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability 57), vol. 57, Chapman and Hall, 1993.

[20] Efron, B., The Jacknife, the Bootstrap and Other Resampling Plans, Philadelphia: Society for Industrial and Applied Mathematics, 1982.