A statistical model for classifying ambient noise in the sea*

OCEANOLOGIA, 39 (3), 1997.
pp. 227–235.

KEYWORDS
Ambient noise
Sea-state classification

Wiesław Kiciński
The Polish Naval Academy,
Gdynia

Manuscript received January 14, 1997, in final form June 18, 1997.

Abstract

The article discusses the results of a multidimensional approach to the investigation of ambient noise. The methodology for the characterisation of ambient noise is discussed and the results of its application in the classification of sea-states are considered.

The main problem in classifying sea-state using multidimensional statistical methods is to determine the distinctive features of ambient noise. The data were processed in the time and frequency domains. The autoregressive model of ambient noise was applied in the time domain, and its coefficients were used as discriminatory terms for classifying and identifying sea-states. The third octave spectrum was used to extract the distinctive features of ambient noise in the frequency domain. The data were analysed in the 2–128 Hz frequency band.

The data sets were processed using Fisher's linear functions. The order of the autoregressive model and the third octave frequencies were found to classify the sea state. The distinctive features of ambient noise were determined with respect to time and frequency. Finally, the parameters of two statistical models were used to classify sea states.

1. Introduction

There are some fields where sea surface sounds play a significant role and a knowledge of ambient noise characteristics is often required. Here, underwater acoustics or hydrometrology can be mentioned as examples.

* This work was supported by grant No. 8T11B02710 from the Polish State Committee for Scientific Research. Parts of this paper were presented at the conference on Sea Surface Sound, Southampton, U. K., 1997.

Various schemes for underwater noise have been developed, usually based on the descriptions and analyses of statistical descriptions of the signal and noise processes. One-dimensional, parametric or non-parametric statistical models are in general use. Tests for independence, homogeneity and normality are applied in the first stage of underwater noise analysis. The second stage – statistical analyses of sea noises – includes the determination of cumulative probabilities, standard deviation, skew, kurtosis, decorrelation time, *etc.*

Ambient noise properties, such as spectrum level and statistical characteristics, have been examined experimentally by a number of investigators. Dyer (1973), and Dyer and Mikhalevskiy (1978) investigated deep-ocean ambient noise in the *ca* 20–200 Hz frequency range. Their statistical descriptions of ambient noise were based on the probability density of the short-term average mean square pressure in the underwater environment. Jobst and Adams (1977) analysed low-frequency ambient noise using statistical tests to establish the independence, time stationarity, Gaussian distributions of samples and statistical homogeneity of underwater noise. Wilson (1985) applied hypothesis testing techniques and statistical tests for randomness, normality and homogeneity in the analysis of surface reverberation generated by projecting sound at the wind-roughened surface of a freshwater lake. Near-surface ocean ambient noise measurements at very low frequencies were investigated by Cotaras *et al.* (1988). They analysed the data statistically by determining the cumulative probabilities, standard deviation, skew, kurtosis and mean spectral level. Bouvet and Schwartz (1988) reported that knowledge of the ambient noise probability density function is the central problem in underwater signal detection. He analysed some models of probability density functions for ambient noises. Klusek (1990) used the classical statistical analysis of data for investigating the contribution of underwater noise sources to the acoustic field of a shallow sea.

The purpose of this paper is to present statistical models for measuring underwater noise and to apply them to a classification of sea-state. The main objective is to show that the methods of time series analysis and discriminant analysis are very useful and powerful tools in the measurement of sea noise. The data used in this paper were collected off the southern Baltic coast during a period of one year.

2. Measuring and data preprocessing

The methodology of sea-state classification considered here is supported by the classification of ambient noise characteristics (Kiciński and Kozaczka, 1996).

Two approaches for investigating ambient noise characteristics are presented:

- in the time domain,
- in the frequency domain.

Ambient noise measurements were carried out in coastal waters of the southern Baltic Sea in 1995. Noise was measured under a variety of hydrological conditions, given in Tab. 1.

Table 1. Hydrological conditions

Sea-state [B]	Wind force [B]	Water temperature [°C]	Air temperature [°C]	Atmospheric pressure [hPa]
1–6	1–6	–5–22	–10–24	1010–1025

Mounted on a stationary measurement system, the hydrophones were submerged to 20 metres below the sea surface. The ambient noise output of the hydrophones was recorded in analogue form on magnetic tape. Each sample was 60 s long. All tonal measurements of ambient noise were preselected by perceptive methods. In this way the outlier measurements were excluded from data set and only 'quiet' measurements were included in the later analysis. Finally, two sets of data were formed. The first one – consisting of 50 samples – was processed as a training set to categorise sea-state classes. Analysing the parameters of hydrological conditions, 3 sea-state classes were determined *a priori*:

- class 1 – sea-state 1–2 B,
- class 2 – sea-state 3–4 B,
- class 3 – sea-state 5–6 B.

All training data were divided into 3 groups. Next, they were indexed according to the *a priori* determined sea-state classes.

The second set of data – consisting of 20 samples – was employed as a testing set in the sea-state identification procedure.

The ambient noise signal and statistical computations were analysed using a digital signal analyser and a computer system with the STATISTICA and QUATTRO PRO programs.

3. Models of sea-state classification

The methodology of sea-state investigation is based on a multidimensional statistical description of data and consists of three stages:

- class determination – grouping objects into classes according to their 'similarity',
- allocation – devising a classification rule from a training set of already classified objects,
- identification – recognising objects as belonging to classes determined *a priori*.

3.1. Sea-state classification in the time domain

In the time domain, ambient noise in the sea was treated as a time series (Box and Jenkins, 1970) at the output of a root mean square (RMS) detector. The root mean square level of underwater noise was determined for each record of an ambient noise signal. The average root mean square levels of ambient noise signal are shown in Fig. 1.

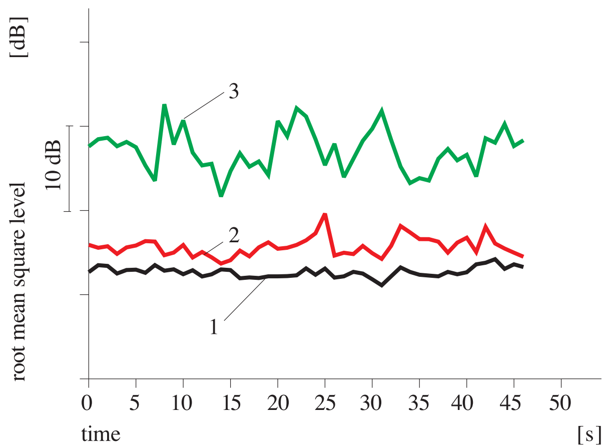

Fig. 1. The average root mean square levels of ambient noise: 1 – sea-state 1–2 B, 2 – sea-state 3–4 B, 3 – sea-state 5–6 B

The autoregessive model (AR) of the signal at the detector output can be expressed as follows:

$$y(k) + a_1(1)y(k-1) + ... + a_n(n)y(k-n) = e_n(k), \tag{1}$$

where
$e_n(k)$ – denotes the model of residual.

The coefficients of the AR model of ambient noise were applied as discriminatory terms of the sea-state. For all training data sets, 3rd order coefficients of the AR model plus constant were estimated. After this, discriminant analysis was used to classify the measurement data (Fisher's linear functions were applied here; Hand, 1981).

The results of determining discriminant variables are presented in Tab. 2. The standard Wilk's λ statistic is used to denote the statistical significance of the discriminatory power of the current model; this can range from 1.0 (no discriminatory power) to 0.0 (perfect discriminatory power). The statistics for overall discrimination are computed as the ratio of the determinant of within-groups covariance matrix to the determinant of the total covariance matrix. The partial lambda (λ_p) is used to determine the unique contribution of the respective variable to the discrimination between groups. It is computed as the multiplicative increment in lambda that results from adding the respective variable. The standard F statistic is used as a variable to indicate its statistical significance in the discrimination between groups, that is, it is a measure of the extent to which a variable makes a unique contribution to the prediction of group membership. Each significant level p is associated with its respective F.

Table 2. Summary of discriminant function analysis

$N = 50$	Wilk's λ	λ_p	F	p
a_0	0.5099	0.7243	8.37	0.0008
a_1	0.4909	0.7524	7.23	0.0019
a_2	0.5035	0.7336	7.98	0.0011
a_3	0.4722	0.7822	6.12	0.0045

It can be seen (Tab. 2) that the variable a_0 contributes the most, the variable a_2 the second most, the variable a_1 the third most and the variable a_3 the least to the overall discrimination. Generally speaking, the smaller the value of Wilk's partial lambda, the greater the contribution to the overall discrimination.

Table 3. The results of classifying the training data set

Classification matrix of training data Rows: observed classification Columns: predicted classification				
Classes	Percent correct	ST1_2 $p = 0.56$	ST3_4 $p = 0.28$	ST5_6 $p = 0.16$
ST1_2	89.28	25	3	0
ST3_4	64.28	5	9	0
ST5_6	62.50	2	1	5
Total	78.00	32	13	5

The discriminant analysis procedure contains two steps. In the first one, the coefficients of classification functions are determined and in the second one, the data is classified. The results of classifying the training data set are shown in Tab. 3.

3.2. Sea-state classification in the frequency domain

The power spectrum level characterises ambient noise in the frequency domain. Each recorded ambient noise signal was processed by the digital signal analyser.

Let

$$X_N = (x_1, x_2, x_3, ..., x_N), \tag{2}$$

where

x_i – i-th component of the power spectrum level,

N – dimensional vector of the measurement space of sea ambient noise in the frequency domain.

The measurement vectors were preprocessed for dimensionality reduction. To this end, the third octave spectrum was computed. Taking eq. (2) into account, the feature vector of sea ambient noise can be expressed as follows:

$$Y_p = (y_1, y_2, y_3, ..., y_p) = f(x_1, x_2, x_3, ..., x_N), \tag{3}$$

where

y_i – i-th component of the third octave spectrum level.

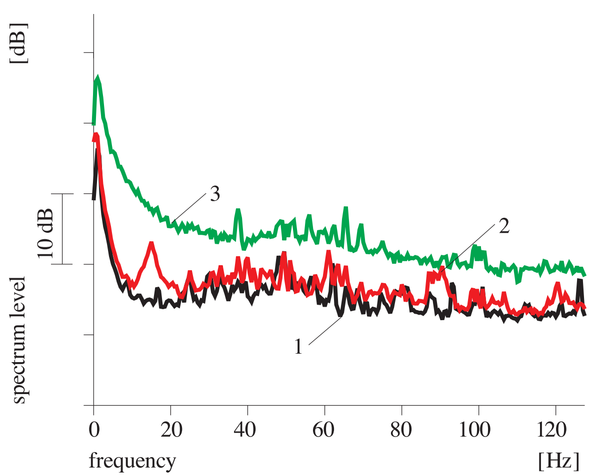

Fig. 2. Average spectrum level of ambient noise: 1 – sea-state 1–2 B, 2 – sea-state 3–4 B, 3 – sea-state 5–6 B

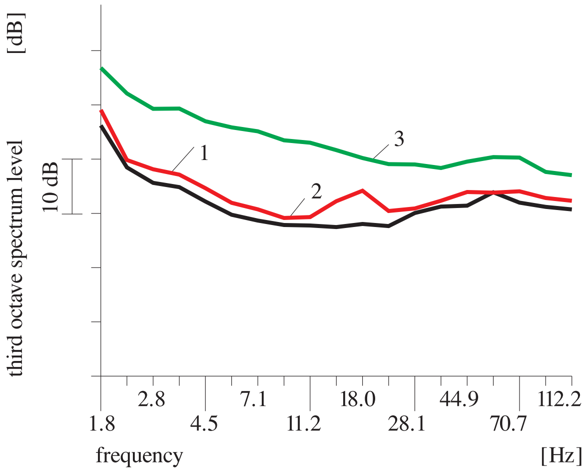

Fig. 3. Average third octave spectrum level of ambient noise: 1 – sea-state 1–2 B, 2 – sea-state 3–4 B, 3 – sea-state 5–6 B

Records of digital data were analysed in the 2–128 Hz frequency band. All data sets were divided into three subsets (classes) of sea-states determined *a priori*: 1–2 B, 3–4 B and 5–6 B. The average spectrum level and the third octave spectrum of ambient noise were computed for each class (Figs. 2 and 3). The records of the third octave spectrum of ambient noise data were used for classifying the data sets. The results of determining the discriminant variables are presented in Tab. 4. The variables V1...V24 represent the centroid frequency of the third octave spectrum bands covering 2–128 Hz.

Table 4. Summary of discriminant function analysis

$N = 50$	Wilk's λ	λ_p	F	p
V6	0.1895	0.8453	3.38	0.0447
V9	0.2294	0.6984	7.98	0.0013
V10	0.1764	0.9081	1.87	0.1682
V11	0.2413	0.6640	9.35	0.0005
V12	0.2265	0.7071	7.66	0.0016
V13	0.2722	0.5885	12.93	0.0001
V14	0.2044	0.7837	5.10	0.0110
V15	0.2204	0.7269	6.94	0.0027
V16	0.1788	0.8959	2.14	0.1310
V18	0.1732	0.9250	1.49	0.2364
V19	0.1734	0.9237	1.52	0.2305

The Wilk's partial λ statistic is used, as before, to denote the statistical significance of the discriminatory power of the current model. One can see (Tab. 4) that the variable V13 contributes the most, the variable V11 the second most, the variable V9 the third most and the variable V18 the least to the overall discrimination. It may be concluded at this point that the frequencies represented by V13, V11, V12, V15 and V9 are the major variables allowing to discrimination between the measurement groups. The results of classifying the training data set are shown in Tab. 5.

Table 5. The results of classifying the training data set

Classification matrix of training data Rows: observed classification Columns: predicted classification				
Classes	Percent correct	ST1_2 $p = 0.56$	ST3_4 $p = 0.28$	ST5_6 $p = 0.16$
ST1_2	100.00	28	0	0
ST3_4	64.28	4	9	1
ST5_6	75.00	1	1	6
Total	86.00	33	10	7

3.3. Verification of models

The multidimensional models of ambient noise described above were employed in sea-state identification. Thus, a data set consisting of 20 samples was tested. The results are presented in Tab. 6.

Table 6. Results of testing measurement models of ambient noise

	Number of measurement data	Measurement model of ambient noise	
		in time domain	in frequency domain
TS1_2	10	7	9
TS3_4	8	3	4
TS5_6	2	1	2
Total	20	11	15
Percent correct	100	55	75

4. Conclusions

Two models of sea-state classification were considered and tested. The main aim was to determine the distinctive parameters of a hydroacoustical signal for ambient noise classification. The sea state was classified using a distinct multidimensional feature of ambient noise in the time and frequency domains. As was pointed out above, better results in sea-state classification can be obtained in the frequency domain by applying the third octave spectrum as a parametrised signal of ambient noise. The methods of ambient noise analysis in the time domain can be extended to the ARMA time series model.

Discriminant analysis of ambient noise measurements were carried out in the 2–128 Hz frequency band, as this is the most useful band for the recognition of underwater noise distortions.

References

Bouvet M., Schwartz S. C., 1988, *Underwater noises: statistical modeling detection, and normalization*, J. Acoust. Soc. Am., 83, 1023–1033.

Box G. E. P., Jenkins G. M., 1970, *Time series analysis*, Holden Day, San Francisco, 574 pp.

Cotaras F. D., Fraser I. A., Merklinger H. M., 1988, *Near-surface ocean ambient noise measurement at very low frequencies*, J. Acoust. Soc. Am., 83, 1345–1359.

Dyer I., 1973, *Statistics of distant shipping noise*, J. Acoust. Soc. Am., 53, 564–570.

Dyer I., Mikhalevskiy P. N., 1978, *Approximations to distant shipping noise statistics*, J. Acoust. Soc. Am., 63, 732–738.

Hand D. J., 1981, *Discrimination and classification*, John Wiley and Sons, New York, 218 pp.

Jobst W. J., Adams S. L., 1977, *Statistical analysis of ambient noise*, J. Acoust. Soc. Am., 62, 63–71.

Kiciński W., Kozaczka E., 1996, *Classification of underwater noise by linear discriminant functions*, [in:] *3rd European Conference on Underwater Acoustics*, J. S. Papadakis (ed.), FORTH–IACM Heraklion, Crete, 1135–1139.

Klusek Z., 1990, *The conditions for sound propagation in the southern Baltic*, Rozpr. i monogr., 1, Inst. Oceanol. PAN, Sopot, 269 pp., (in Polish).

Wilson G. R., 1985, *A statistical analysis of surface reverberation*, J. Acoust. Soc. Am., 74, 249–255.