Bayesian Seabed Classification Using Angle-Dependent Backscatter Data From Multibeam Echo Sounders

Knut Landmark, Anne H. Schistad Solberg, Andreas Austeng and Roy Edgar Hansen

Abstract

Acoustical seabed classification is a technology for mapping seabed sediments. Processed multibeam sonar data yield the variation of the seabed scattering strength with incidence angle, and this paper examines the effect of this on classification. A simple Gaussian statistical model is developed for the observed scattering strength, whereby an observation is represented by a piecewise constant function of incidence angle. Provided some data for which the sediment types are known (training data), the statistics for each type can be robustly estimated. Subsequently, standard Bayesian theory is applied to classify new observations. The model was used to compute limits on classification accuracy in terms of the intrinsic scattering strength statistics of the seabed, and to predict whether a logarithmic or linear scale for the data is preferable. Systematic experiments on a North Sea data set with four sediment classes tested how the classification accuracy depends on the piecewise function approximation, incidence angle range, amount of training data, and spatial averaging (combining consecutive pings into one observation). The classifier based on Gaussian statistics performed at least as well as sophisticated algorithms with no assumptions about the data statistics. The best accuracy (95 %) was attained for logarithmic data. The amount of training data needed to achieve this was about 500 pings per class; spatial averaging could be limited to 10–20 pings. Comparable across-track spatial resolution was possible by dividing the full swath into separate independent sectors, but only at reduced accuracy (87 % or less). However, comparable accuracy may be possible by taking into account the spatial relationships of observations.

Index Terms

Bayesian methods, sonar, classification algorithms, seafloor, sediments, remote sensing.

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

I. INTRODUCTION

The purpose of acoustical seabed classification (ASC) and characterization is to map surficial seabed composition with sonars [1]. Multibeam echo sounders (MBES) are particularly useful, as they are now widely used for high-resolution bathymetry. MBES technology has had a significant impact on the marine sciences. Images of seabed morphology produced from multibeam bathymetry tell a detailed story of the physical processes that influence the seabed. ASC adds an extra dimension of information by analyzing and mapping out variations in the seabed acoustical response. This has proven useful, for example, in the field of marine ecology, where ASC is applied to the study of seabed (benthic) habitats [2]–[4]. Seabed ecosystems can be vulnerable to the effects of human activity, from disruptions caused by bottom trawling, to anthropogenic climate change. Between the conflicting interests of resource development and nature conservation, seabed mapping and monitoring will likely only increase in importance.

The solution of a generic classification problem has two main steps [5]: First, to preprocess the data and extract the signal characteristics that are the most effective for discriminating between different classes of signals. Second, to compute a decision rule which assigns a signal to one of several possible classes. The decision rule is often computed based on the statistical distribution of a set of signals for which the class is known, the training data (supervised learning). There are many powerful methods available for computing the decision rule [5], [6], but their relative performance is data dependent, and it is difficult to draw general conclusions about classifier performance based on just a few test data sets. Nevertheless, one should try and answer questions concerning how accurate is the decision rule, how well will the method work on new data sets (generalizability), and when might it not work.

The strategy adopted here is to develop a classification method and consider these questions experimentally as well as with the help of a simple statistical model for the observations. We focus on a property of the seabed acoustical response which can be fairly well modeled and measured. An MBES forms many narrow beams in a wide fan perpendicular to the ship's track, and therefore measures how the seabed scattering strength S varies with the incidence angle θ [7, Ch. 8]. The shape of $S(\theta)$ depends on the seabed reflection coefficient, the roughness spectrum, and the strength of sediment volume scattering [8, Ch. 13–14]. While the received signal statistics can be sensitive to changes in the ensonification pattern on the seabed [9]–[11] (an extraneous factor), properly measured scattering strength is not.

Although seabed characterization based on $S(\theta)$ can be ambiguous due to the complex interplay of scattering mechanisms, several promising results suggest that $S(\theta)$ is a useful feature for seabed classification and characterization. The angular range analysis of [12] is an inverse method for estimating acoustical parameters. It works by fitting simulated and observed features of $S(\theta)$ in several separate angular subsectors. The inversion algorithm is based on the fact that some features correlate with different scattering mechanisms; this fact may also alleviate the problem with ambiguities. Inversion results based on model-data comparison of the $S(\theta)$ function are also reported in [13], [14]. The parameter estimates of [13] are consistent with independent estimates made by inversion of a time-domain (echo shape) model applied to single beam echo sounder data in [15]. The papers [16]–[19] describe statistical or pattern recognition methods applied to the measured $S(\theta)$ function, including Bayesian classifiers, neural networks, and cluster algorithms. Simulation results in [17] showed best performance with a Bayesian classifier, albeit on data with known statistical distributions. A Markov random field based segmentation algorithm for sonar image mosaics, which also

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

takes into account the angular variation, was presented in [20]. Pattern recognition methods are less vulnerable to systematic errors in the data, but are best suited to situations where observations divide naturally into a set of discrete classes, which may not be fully known in advance. This problem is addressed in [21], which represents the backscatter histogram at a single angle with a normal mixture probability density function (PDF), where the number of components is determined from the acoustical data.

This paper describes a statistical supervised classification method, building on preliminary results in [22]. We consider several questions of practical importance concerning ASC applied to $S(\theta)$ measurements: accuracy (discriminating power), spatial resolution, training data requirements, statistics, and PDF estimation. In the statistical approach, an observation of $S(\theta)$ ($\theta \in [\theta_{\min}, \theta_{\max}]$) is identified not with physical seabed parameters, but by its posterior probability $P(\omega|S)$ and likelihood $p(S|\omega)$ with respect to certain classes ω. The classes are defined by ground truth data or other independent knowledge. In theory, the optimum (minimum error) classifier is the Bayes classifier [5], but this presupposes that the PDFs $p(S|\omega)$ are known. To get robust estimates of the PDFs from possibly scarce data, we propose to replace $S(\theta)$ by a vector $\boldsymbol{d} = \begin{bmatrix} d_1 & \dots & d_K \end{bmatrix}^T$, where d_k is the average scattering strength in a segment $\theta_{k-1} \leq \theta < \theta_k$. The number of components in $\boldsymbol{d}$ represents a trade-off between accurate approximation and robustness. If there are too few segments, the classification error rate increases because important information is lost by excessive smoothing. If there are too many segments, good estimation of the PDFs for $\boldsymbol{d}$ requires an unrealistic amount of training data. To achieve a good balance, the segments are determined such that the approximation error is minimized for the data set as a whole (for a fixed number of components K) [23], [24]. The observation vector $\boldsymbol{d}$ retains a simple physical interpretation, and is insensitive to all but very large errors in the estimated incidence angles. Section II-B and Section II-C explain the processing steps for obtaining the observation vectors.

The classification accuracy depends on the amount of averaging, the intrinsic scattering strength S, and the variance of S. Observation vectors result from averaging over time samples, angular segments, and consecutive pings. In Section II-D and appendix A, we compute approximate PDFs for the observation vectors, using a central limit theorem, for the simplest case of uncorrelated components. This is used to compare error bounds for classification with intensity and log-intensity data respectively. It may also be used to generate PDFs from modeled $S(\theta)$ functions.

The method was applied to a North Sea MBES data set. The survey area and supporting data are described in Section II-A. The experimental procedure is explained in Section III-A. The purpose of the experiments was to test how the classification accuracy depends on

- the dimension of the scattering strength vector $\boldsymbol{d}$;
- the amount of training data;
- the amount of spatial averaging (combining data from several pings);
- the extent of the angular sector.

Spatial averaging limits the along-track resolution of ASC mapping. Across-track resolution was also considered by dividing the full sonar swath into separate subsectors. We believe the data set is a useful testing ground for ASC methods. The survey area covers distinct sedimentary environments: a sandy plateau with low present day sedimentation; a deeper trench with thick soft marine depositions; as well as a transition zone between them. A set of sediment samples (gravity cores and grab samples) with laboratory analyses is available, and

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

TABLE I
DEFINITION OF SEDIMENT TYPES

Type	Grain size (mm)	ϕ [a]
Clay	< 0.002	> 9
Silt	0.002-0.063	4-9
Sand	0.063-2.0	(-1)-4
Gravel	> 2.0	< -1

[a] $\phi = -\log_2(D/1\,\text{mm})$, where D is particle diameter.

a previous survey using an alternative ASC method and single beam echo sounder (SBES) provides a reference point [25].

Test results show that good classification accuracy and spatial resolution is possible even with limited training data. Section III-B presents the experimental results, which are discussed in Section IV.

II. METHODS AND DATA

A. Survey area

The survey area stretches from the eastern margin of the North Sea plateau and into the Norwegian Trench (Fig. 1). The North Sea is a shallow continental sea bounded by the British Isles, Norway, and the northern European continent. The Norwegian Trench (or Channel) is a depression in the continental shelf along the coast of southern Norway, with depths of about 300 m in the survey area; its recent geology is influenced by alternating glaciations and inundations during the Quaternary age [26]. The Norwegian Trench is an important trap for fine-grained water-borne sediments in the North Sea [27], and the top seabed layer is soft, silty and rich in organic material [28]. The eastern North Sea plateau is about 100 m deep, and the seabed is covered with sand and gravel [25], [29]. There is almost no present-day accumulation of water-borne, fine-grained sediments in this area. A gradual change in grain size distribution is expected down the western side of the trench. However, this slope is marked by flutes, ditches or elongated pockmarks, which suggests strong action by ocean currents or fluid seepage from below [30], [31].

The sediment sample data set (Fig. 1) confirms this picture. The mass fraction of sand in the samples increases westwards, up to more than 90 % on the plateau. The samples are labeled according to a modified version of Folk's classification system (TABLE I,II) [25], [32]. A previous study covering the area employed single beam echo sounder and the QTC Impact software [33] for seabed classification. Based on the ASC results and the same set of samples, it was concluded that four classes of seabed sediments could be discerned [25]:

1) silt and mud (deep central part of the trench);
2) sand with varying amounts of gravel (shallow plateau);
3) a variety of clay, silt, and sand (lower western slope of the trench);
4) sandy mud with gravel (upper western slope).

There are seven sediment samples inside the studied survey area, classified as either sand (S), sandy silt (sZ) or silt (Z). A fourth class designated U (unknown) was taken to represent a subarea devoid of samples. One of the sand samples (described as fine sand with shell

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

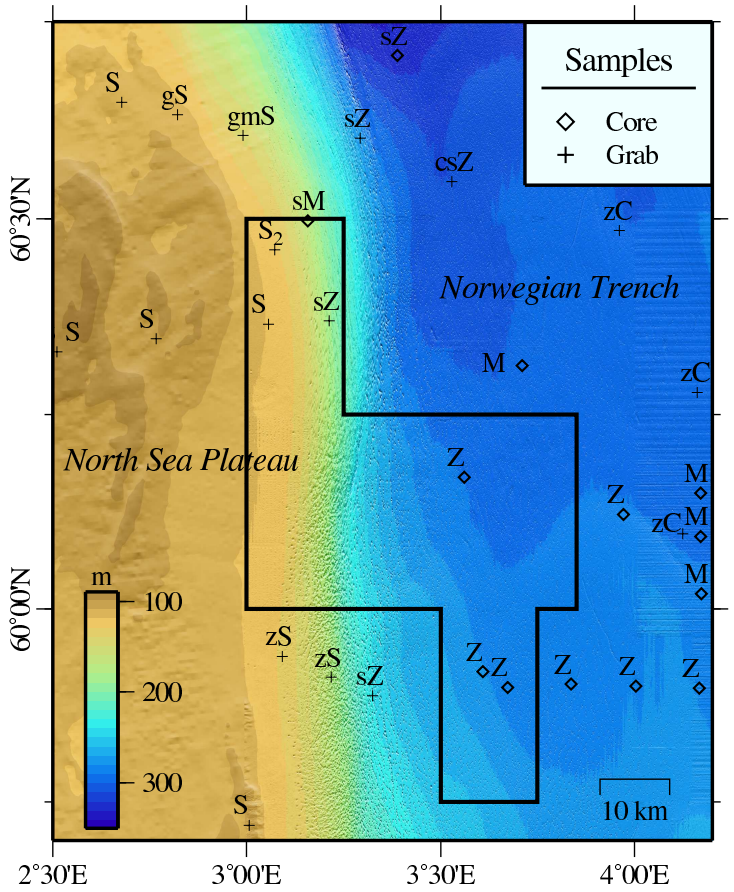


Fig. 1. Bathymetry and sediment samples (gravity cores and grabs) in and around the study area (solid lines) [25]. Letters indicate laboratory classification of samples, e.g. zS means "silty sand" (TABLE I, II).

TABLE II
COMPOSITION OF SEDIMENT SAMPLES

Abbr.	Mass fraction		Description
M	Clay+Silt $> 90\%$	$1:2 <$ Clay:Silt $< 2:1$	Mud
C	Sand < 10 %	Clay:Silt $> 2:1$	Clay
Z	Sand < 10 %	Clay:Silt $< 1:2$	Silt
S	Sand > 90 %	Gravel < 2 %	Sand
sZ	Sand < 50 %	Clay:Silt $< 1:2$	Sandy silt
zS	Sand > 50 %	Clay:Silt $< 1:2$	Silty sand
gS	Gravel 2-30 %	Clay+Silt:Sand $< 1:9$	Gravelly sand
zC	Sand < 10 %	Clay:Silt $> 1:1$	Silty clay
sM	Sand < 50 %	$1:2 <$ Clay:Silt $< 2:1$	Sandy mud
gmS	Gravel 2-30 %	Clay+Silt:Sand $< 1:1$	Gravelly, muddy sand
csZ	Silt > 50 %	$1:2 <$ Clay:Sand $< 2:1$	Clayey, sandy silt

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

fragments) lies in a wedge-shaped area with higher backscatter levels than the surroundings. The sand class was therefore split in two: S and S_2. Hence we defined five classes in our classifier evaluation: S, S_2, Z, sZ and U.

B. Data preparation

The MBES data and sediment samples were acquired by the Norwegian Defence Research Establishment (FFI) on the research vessel M/S HU Sverdrup II. About 15 GB of data were collected during 600 hours of three separate surveys in the fall of 2005 and late winter of 2006. The three surveys together cover the contiguous area outlined in Fig. 1. The echo sounder was a Kongsberg Maritime EM 1002. The EM 1002 has a center frequency of 95 kHz; the array is semi-cylindrical and emits a fan-shaped beam with 2.0° along-track $3\,\mathrm{dB}$ beamwidth, and it forms 111 beams with 2.0° beamwidth across track. Fig. 2 illustrates the acquisition geometry of an MBES, and Fig. 3 shows a map of the backscatter data. The latter corresponds to Fig. 1, with sounding depth replaced by beam intensity, interpolated onto a regular 10 m^2 grid. The beam intensity is an approximation to the seabed scattering strength [34], [35], and not normalized with respect to angular variations. The angular coverage of the EM 1002 is up to 150° athwartships, but we have only used port and starboard sectors from 0-50°. The present data were acquired with a maximum beam angle (with respect to the vertical) of up to about 65°, but many pings have a narrower angular range. This depends in particular on the water depth, refraction, and possibly also the weather conditions; outer beams may also be missing due to erroneous bottom detection. In addition, other authors [16] have observed that beampattern effects may cause unphysical backscatter levels at high incidence angles. To avoid extrapolating missing data when comparing data from different pings, observations with missing data would have to be discarded. Hence the maximum angle during data processing was set as low as 50° to avoid discarding many observations.

To describe the data processing, we let $y = |\Pi(t)|^2$ denote the squared envelope of the complex pressure time series. The averaged scattering cross section is defined as in [36, Ch. 16]:

$$s(\theta) = \int_\Gamma \sigma(\boldsymbol{r};\theta)\Omega(\boldsymbol{r};\boldsymbol{r}_0)\,\mathrm{d}^2\boldsymbol{r}, \tag{1}$$

where σ is the local scattering cross section. The integration variable $\boldsymbol{r}$ is position on the seafloor; the integral is over the ensonified area Γ of a single beam (at time t); $\Omega(\boldsymbol{r};\boldsymbol{r}_0)$ is the ensonification function, which is the beam pattern expressed as a function of $\boldsymbol{r}$ and $\boldsymbol{r}_0$, the position of the array. It is assumed that Γ is small, so that the incidence angle θ is fixed.

For a long stationary signal, Γ is the beam footprint. The mean square envelope (in $volts^2$), compensated for propagation loss, is then

$$\langle y\rangle = 2s_0^2 s_r^2 s(\theta), \tag{2}$$

where s_0 is the source level and s_r is the receiver sensitivity on the beam's maximum response axis. The factor 2 occurs if s_0 is defined in terms of RMS pressure [8, App. F–G]. For a short MBES pulse of length τ, the across-track width of Γ is limited to $c\tau/2\sin\theta$, and it is not precisely known where Γ lies within the footprint (at particular time t). Nor was $\Omega(\boldsymbol{r},\boldsymbol{r}_0)$ precisely known. Instead (2) is used with $\langle y\rangle$ replaced by the mean intensity, $\bar{y}$, in a window around the echo peak. On average, samples near the echo peak are energy scattered from resolution cells near the maximum of the ensonification function, where its value ($\Omega = 1$) is

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

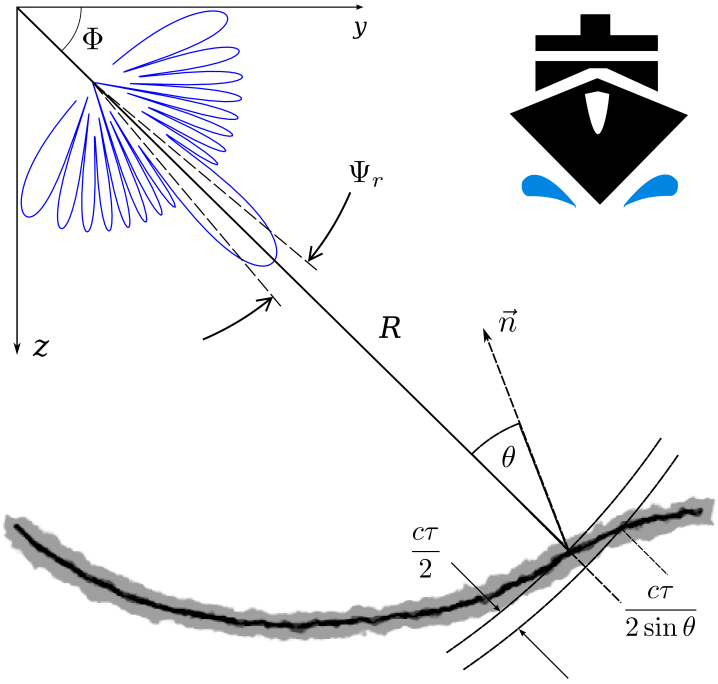


Fig. 2. MBES geometry in the across-track plane. The vessel moves along the x-axis. A single ping with pulse length τ ensonifies a strip of seafloor which is wide in y and narrow in x. The receiver array steers many beams at different depression angles Φ, yielding backscatter response at different incidence angles θ. A resolution cell has a length of $c\tau/2\sin\theta$; the beam footprint length is approximated as $R\Psi_r$. The beampattern shown here has $3\,\mathrm{dB}$ beamwidth $\Psi_r = 10°$ and uniform $-20\,\mathrm{dB}$ sidelobe levels. The actual beamwidth of the EM 1002 is $\Psi_r = 2.0°$, but the precise beampattern is not known.

known by definition of Ω, s_0, and s_r. We understand that the manufacturer accounts for the factors s_0 and s_r, which have been determined experimentally [34], [35]. The corrected beam intensity is

$$y_c = \frac{\bar{y}q^2}{s_0^2 r_0^2 A}, \tag{3}$$

where q is the one-way propagation loss. The factor A is an approximation to

$$\int_\Gamma \Omega \,\mathrm{d}^2\boldsymbol{r} \approx A = R\Psi_t \times \min\left[R\Psi_r, c\tau/2\sin\theta\right],$$

where R is the slant range, and Ψ_t, Ψ_r are the $3\,\mathrm{dB}$ transmit and receive beamwidths. The factor A ensures that $\langle y_c\rangle \approx \langle\sigma\rangle$. The bias caused by the simplification $A \approx \int_\Gamma \Omega\,\mathrm{d}^2\boldsymbol{r}$ is independent of seabed type. In the present work it is only the difference between seabed types which matters. The best way to determine scattering strength would be to measure the total backscattered energy from the beam footprint [37]. However, our recorded data did not include the full time series of the seabed echoes.

We used ray theory [38, Ch. 3] to compute the geometrical spreading loss and the beam direction $\boldsymbol{v}$ at the seafloor. Sound absorption was estimated using conductivity, temperature, and density (CTD) profiles and the empirical formulae of [39]. The seabed shape and surface normal $\boldsymbol{n}$ were calculated with cubic spline interpolation using depth soundings from a set of consecutive pings. (By comparison, the system software assumes that the seabed has a constant slope across the whole swath.) The incidence angle was found from the three-dimensional scalar product $\cos\theta = \boldsymbol{n}\cdot\boldsymbol{v}/\|\boldsymbol{v}\|$.

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

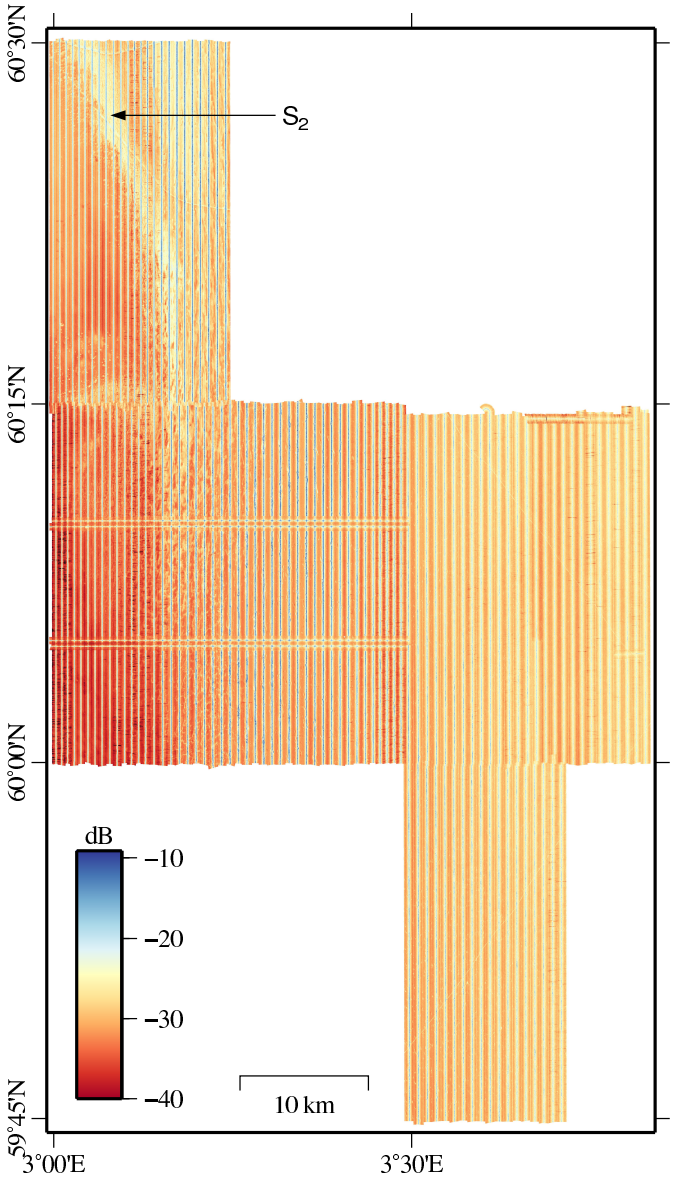


Fig. 3. Scattering strength data (beam intensity) in the survey area. An area of high backscatter levels which contains the S_2 sample (Fig. 1) is also indicated.

C. Observations

Suppose $0 < \theta_1 < \ldots < \theta_{K-1} < \theta_{\max}$ is a subdivision of the incidence angle range of the sonar swath. An observation is a K-dimensional vector $\boldsymbol{d}$, where each component d_k is the average of corrected intensity measurements (3) for incidence angles $\theta \in [\theta_{k-1}, \theta_k)$. As Fig. 4 shows, the vector $\boldsymbol{d}$ is simply a piecewise constant approximation to the full scattering strength observation. A group of measurements can be formed from N_p consecutive pings, but data from port and starboard sides of the swath are kept as separate observations.

For a seabed type ω, we shall estimate a PDF $p(\boldsymbol{d}|\omega)$ for the random vector $\boldsymbol{d}$ with respect to ω. As a rule-of-thumb [40], there should be at least ten times as many training observations for ω as there are components in $\boldsymbol{d}$. To keep K as low as possible without reducing the approximation accuracy too much, the segments $[\theta_{k-1}, \theta_k)$, $k = 1, \ldots, K$, are defined such that the total approximation error for the data set is minimized for fixed K. Hence the segments in Fig. 4 are short where the intensity variation with angle is large, and longer where the

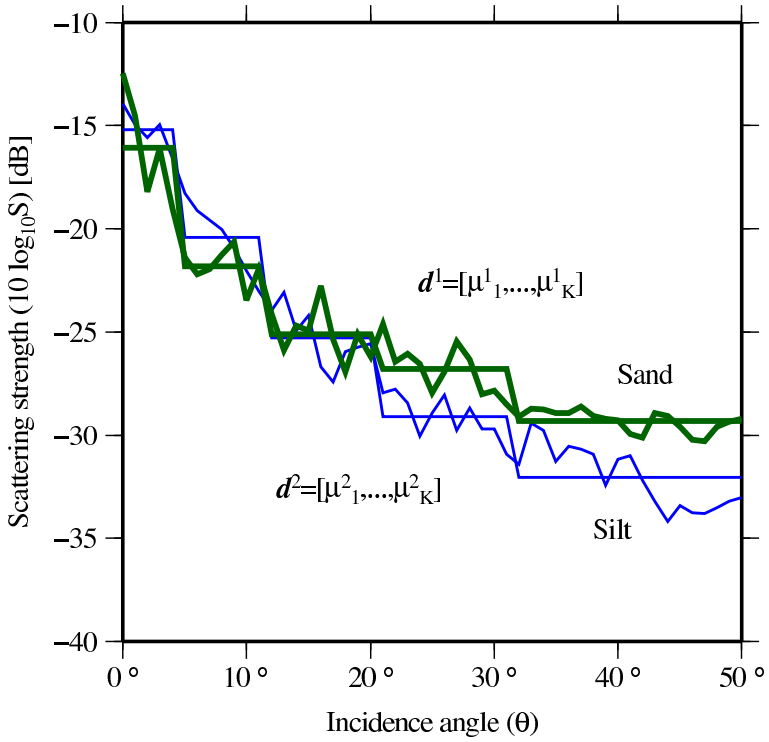


Fig. 4. Two observations of scattering strength versus incidence angle (irregular curves), along with piecewise constant approximations with $K = 5$ segments: sand (thick lines) and silt (thin lines). Each segment is an angular range corresponding to a horizontal line segment in the piecewise constant approximations. The segments are the same for both (all) classes. The K-dimensional observation vectors $\boldsymbol{d}$ comprise the mean values for the K segments. (Data from 10 consecutive pings have been used to form one vector.)

variation is smaller.

Mathematically, getting $\boldsymbol{d}$ from measurements can be viewed as a two-step binning process, defined by two non-decreasing onto functions b and c,

$$[0, \theta_{\text{max}}] \xrightarrow{b} \{1, 2, \ldots N\} \xrightarrow{c} \{1, 2, \ldots K\}\,, \;\; K < N.$$

We let $\langle A \rangle$ denote the arithmetic mean of the elements of a discrete set A. A full observation (irregular curve in Fig. 4) is thus a vector $\boldsymbol{\varsigma} = \begin{bmatrix} \varsigma_1 & \ldots & \varsigma_N \end{bmatrix}$, where the component ς_i is the average of all measurements in the ith bin for a group of pings,

$$\varsigma_i = \langle y_c(\theta) \,|\, b(\theta) = i \rangle\,. \tag{4}$$

The number of bins N depends on the angular coverage and resolution. In this paper $\theta_{\text{max}} = 50°$ and $N = 50$, so the bin length is $1°$ (about the ratio of angular coverage to number of beams). A data set of M observations $\boldsymbol{\varsigma}^1, \ldots, \boldsymbol{\varsigma}^M$ is represented by an $M \times N$ matrix

$$\boldsymbol{D} = \begin{bmatrix} \xrightarrow{N \text{ bins}} & & \\ \varsigma_1^1 & \cdots & \varsigma_N^1 \\ \vdots & \ddots & \vdots \\ \varsigma_1^M & \cdots & \varsigma_N^M \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \downarrow M \text{ observations} \end{matrix}$$

The segments correspond to the horizontal lines in Fig. 4 and are defined by c. Hence $\boldsymbol{\varsigma}$ is replaced by the observation vector $\boldsymbol{d} = \begin{bmatrix} \mu_1 & \ldots & \mu_K \end{bmatrix}$, where $\mu_k = \langle \varsigma_i \,|\, c(i) = k \rangle$ is the mean of all components in the kth segment. For a given number of segments K, the approximation

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

error depends on c, which is determined adaptively from $\boldsymbol{D}$. We use a method from [23], [24], which minimizes the total squared approximation error for the data set $\boldsymbol{D}$. The optimization problem is

$$\min_{c} \sum_{k=1}^{K} \sum_{j=1}^{M} \sum_{i \in I_k} \left(\varsigma_i^j - \mu_k^j\right)^2, \tag{5}$$

where $I_k = \{\, n \mid c(n) = k \,\}$. A fast and straightforward algorithm for solving (5) can be found in [23]. The solution for c can be implemented as a linear transform in the following way. Let n_k be the number of elements in I_k, and define $N_k = \sum_{i=1}^{k-1} n_k,\ N_1 = 0$. Let $\boldsymbol{w}_k$ be the $N \times 1$ vector with $1/n_k$ in rows $N_k + 1 \ldots N_{k+1}$ and 0 elsewhere. Then

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{w}_1 & \ldots & \boldsymbol{w}_K \end{bmatrix} \tag{6}$$

is an $N \times K$ matrix such that $\boldsymbol{d} = \boldsymbol{\varsigma W}$ is the feature vector of observation $\boldsymbol{\varsigma}$.

The optimization (5) is performed on the whole data set (all observations and seabed types), or a representative subset, not just the training samples. The segments are the same for all seabed types. The main point is to reduce the dimension of $\boldsymbol{\varsigma}$ (i.e. $K < N$); the PDF $p(\boldsymbol{d}|\omega)$ can be estimated more accurately when K is small. This is relevant for ASC when training observations must be picked near a few ground truth sites and therefore are limited in number. Failing to ensure a high ratio between training set size and sample space dimension may impair the classification accuracy, as is commonly seen in pattern recognition problems [5].

D. Bayesian classification

Let $p(\boldsymbol{d}|\omega_i)$ denote the PDF for feature vector $\boldsymbol{d}$ with respect to seabed class ω_i. The Bayes decision rule [5, Ch. 2] states that a new observation $\boldsymbol{x}$ should be assigned to the class with the highest posterior probability. By Bayes' theorem, this means selecting class ω_i if

$$p(\boldsymbol{x}|\omega_i)P(\omega_i) > p(\boldsymbol{x}|\omega_j)P(\omega_j) \text{ for all } j \neq i, \tag{7}$$

where $P(\omega_i)$ is the prior probability assigned to class ω_i. The PDFs may be estimated parametrically or non-parametrically. We used a multinormal model and equal prior probabilities. Thus,

$$p(\boldsymbol{d}|\omega_i) = \frac{(2\pi)^{-K/2}}{|\hat{\boldsymbol{\Sigma}}_i|^{1/2}} \exp\left[-\frac{1}{2}(\boldsymbol{d} - \hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}_i^{-1} (\boldsymbol{d} - \hat{\boldsymbol{\mu}}_i)\right], \tag{8}$$

where $\hat{\boldsymbol{\mu}}_i = (1/N) \sum_{k=1}^{N} \boldsymbol{d}_k^i$ is the mean vector from N training samples in class ω_i, and

$$\hat{\boldsymbol{\Sigma}}_i = \frac{1}{N} \sum_{k=1}^{N} (\boldsymbol{d}_k^i - \hat{\boldsymbol{\mu}}_i)(\boldsymbol{d}_k^i - \hat{\boldsymbol{\mu}}_i)^T$$

is the sample covariance matrix for observations in class ω_i.

Classification was implemented with the Matlab toolbox PRTools4 from TU Delft [41]. We have tested the three common variants of the Bayesian normal classifier: Quadratic discriminant classifier (QDC), linear discriminant (LDC), and uncorrelated (UDC). UDC is the assumption that the covariance matrices are diagonal; LDC is the assumption that the covariance matrices are equal ($\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$ for all classes ω_i); QDC is general (no assumption on covariance structure). LDC or UDC may be preferred over QDC when $\boldsymbol{\Sigma}$ is near-singular (typically, with few and closely correlated training samples). For QDC, the decision

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

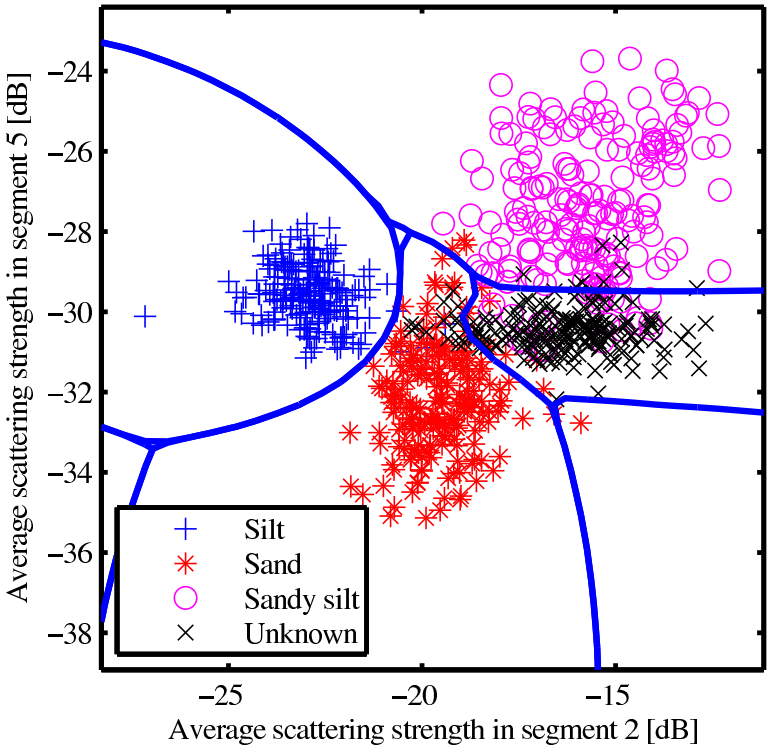


Fig. 5. Scatterplot and decision boundaries for the QDC in two dimensions. Segment 2 is the angular range $[5°, 12°\rangle$, and segment 5 is the range $[32°, 50°\rangle$. We used the forward floating search method [42] to pick the two most informative features from a five-dimensional approximation of scattering strength. The plot contains 800 observations (200 per class), which were randomly selected from large training areas as explained in Section III-A and Fig. 7.

boundaries (points of equality in (7)) are $(K-1)$-dimensional quadric hypersurfaces, as the two-dimensional example in Fig. 5 shows.

The critical step is PDF estimation, not the assumption of normality in (8). However, using a normal model reduces the risk of overfitting the model to the data, and simplifies the PDF estimation task to computing $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}^{-1}$. By definition, a feature vector $\boldsymbol{d}$ results from a triple averaging over time samples, over pings, and over incidence angle segments. It is therefore a sum of many random vectors, each component of which is a sum of intensities from separate sonar resolution cells. With the common assumption of statistical independence, it follows below that the PDFs for the observation vectors should be asymptotically normal.

Suppose the PDF $f_{Y|\theta}(y)$ describes the distribution of beam intensity measurements at angle θ, for a particular seabed type. The seabed type is characterized by a distribution $f_{S|\Theta}(s|\theta)$ of the average scattering strength (1). To be more specific, we consider the compound model in [36, Ch. 16.3], where it is assumed that intensity measurements collected from regions with a fixed value of s are exponentially distributed (i.e. Gaussian complex pressure). The compound PDF is

$$f_{Y|\Theta}(y|\theta) = \int_0^\infty \frac{1}{s} e^{-y/s} f_{S|\Theta}(s|\theta)\,\mathrm{d}s. \tag{9}$$

Measurements in the kth segment are for random angles in $[\theta_{k-1}, \theta_k)$, so we regard Θ as a

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

continuous random variable with uniform PDF

$$f_\Theta(\theta;k)=\begin{cases}1/\left(\theta_k-\theta_{k-1}\right) & \text{if } \theta\in[\theta_{k-1},\theta_k)\\ 0 & \text{otherwise.}\end{cases} \tag{10}$$

We write the expected value in segment k with respect to Θ as

$$\langle A\rangle_k=\frac{1}{\theta_k-\theta_{k-1}}\int_{\theta_{k-1}}^{\theta_k}A(\theta)\,\mathrm{d}\theta. \tag{11}$$

The marginal distribution

$$\begin{aligned}f_Y(y;k)&=\int_{\theta_{k-1}}^{\theta_k}f_{Y|\Theta}(y|\theta)f_\Theta(\theta;k)\,\mathrm{d}\theta\\&=\int_0^\infty\frac{1}{s}e^{-y/s}\left\langle f_{s|\Theta}(s|\theta)\right\rangle_k\,\mathrm{d}s\end{aligned} \tag{12}$$

is a model for the normalized histogram of intensity measurements in the kth segment. For log-intensity measurements, (12) must be replaced with

$$\begin{aligned}f_W(u;k)&=\int_0^\infty f_{W|S}(u)\left\langle f_{S|\Theta}(s|\theta)\right\rangle_k\,\mathrm{d}s,\\W&=\log Y,\\f_{W|S}(u)&=\frac{e^u}{s}e^{-e^u/s}.\end{aligned} \tag{13}$$

Now form the vector $\boldsymbol{Z}=\begin{bmatrix}Y_1 & \dots & Y_K\end{bmatrix}^T$, where Y_k is a random variable with PDF as in (12) or (13). If $\boldsymbol{Z}^1,\dots,\boldsymbol{Z}^M$ are M independent random vectors distributed as $\boldsymbol{Z}$, with mean $\boldsymbol{\mu}$ and finite covariance matrix $\boldsymbol{\Gamma}$, then

$$\frac{\boldsymbol{Z}^1+\ldots+\boldsymbol{Z}^M}{\sqrt{M}}\xrightarrow{\mathrm{D}}N(\sqrt{M}\boldsymbol{\mu},\boldsymbol{\Gamma}), \tag{14}$$

where $\xrightarrow{\mathrm{D}}$ means convergence in distribution [43, Ch. 11]. For sufficiently large M the PDF for the average

$$\boldsymbol{m}=\frac{\boldsymbol{Z}^1+\ldots+\boldsymbol{Z}^M}{M} \tag{15}$$

approaches the normal distribution $N(\boldsymbol{\mu},\boldsymbol{\Gamma}/M)$. The average $\boldsymbol{m}$ (15) is a model for the observation vector $\boldsymbol{d}$, although in practice we allow for different number of measurements in different segments. The number of independent measurements increases with N_p (number of pings) and decreases with K (number of segments). With the notation

$$\varsigma(\theta)=\mathrm{E}[S|\Theta=\theta] \tag{15a}$$

$$\nu^2(\theta)=\mathrm{Var}[S|\Theta=\theta] \tag{15b}$$

$$\varsigma_L(\theta)=\mathrm{E}[\log S|\Theta=\theta] \tag{15c}$$

$$\nu_L^2(\theta)=\mathrm{Var}[\log S|\Theta=\theta], \tag{15d}$$

then from the appendix (28)–(29),

$$\mu_k\equiv\mathrm{E}[d_k]=\mathrm{E}[Y]=\langle\varsigma\rangle_k \tag{16}$$

$$M\Sigma_{kk}\approx\mathrm{Var}[Y]=2\left\langle\nu^2\right\rangle_k+\left\langle\varsigma^2\right\rangle_k+\mathrm{Var}_\Theta\left[\varsigma\right] \tag{17}$$

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

for linear intensity. For log-intensity, the result is (30)–(31)

$$\mu_k \equiv \mathrm{E}[d_k] = \mathrm{E}[\log Y] = \langle \varsigma_L \rangle_k - \gamma \tag{18}$$

$$M\Sigma_{kk} \approx \mathrm{Var}[\log Y] = \left\langle \nu_L^2 \right\rangle_k + \frac{\pi^2}{6} + \mathrm{Var}_\Theta \left[\varsigma_L\right]. \tag{19}$$

Here γ is the Euler-Mascheroni constant.

In the appendix we have used these formulae to calculate error bounds for discrimination between pairs of classes, parameterized by M and the standard deviation of the intrinsic scattering strength. This may in turn be used to predict whether it is better to apply the classification method to intensity or log-intensity data in general. Since we do not know the underlying distributions $f_{S|\Theta}(s|\theta)$, the maximum entropy principle [44] is invoked to make the comparisons. From (43)–(44) in the appendix,

$$\nu^2(\theta) = \left(e^{\nu_L^2} - 1\right) \varsigma^2(\theta),$$

so ν is proportional to ς, and $\nu = 0$ for $\nu_L = 0$. If we assume for simplicity that the proportionality factor (hence ν_L) is angle-independent and class-independent, then the error bound (36) can be plotted as a function of the single parameter ν_L. This gives a family of error bound curves indexed by the number M. Fig. 6 shows one example for the classes S and Z. To make this plot we have used the class means of the observed scattering strength $\varsigma_L(\theta)$ (full observations with $N = 50$ bins). These error bounds are for classification in $K = 5$ dimensions. Fig. 6 suggests that log-intensity is the better choice. The picture is similar for e.g. the pair sZ and U.

III. Experiments and results

A. Experimental setup

The purpose of the experiments was to see how the classification accuracy depended on

- the dimension of the scattering strength vector $\boldsymbol{d}$;
- the amount of training data;
- the amount of spatial averaging;
- the extent of the angular sector.

These factors determine the reliability, spatial resolution, and practical applicability of the method. We evaluated the classifier in two alternative ways. Both involved the following recipe:

1) Select random, independent training and test data for each known class (seabed type).
2) Estimate the PDFs (8) using the training set.
3) Count the number of misclasssified test samples in repeated experiments to obtain an average error rate, the probability of error.

The first way was with small training areas centered on the sediment sample sites. This would be the natural approach to mapping new areas where little is known about the seabed geology in advance. There are seven samples inside the survey area, representing four classes (Section II-A and Fig. 1), and each defined the center of a 3 km^2 block of training data. Two additional blocks of data were taken from an area far from any samples. These data represented the class U (unknown, not sampled). We defined test blocks similar in number and size in the vicinity of the sample sites; they did not overlap with the training areas, but the seabed type could nevertheless be asserted with reasonable confidence based on prior

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

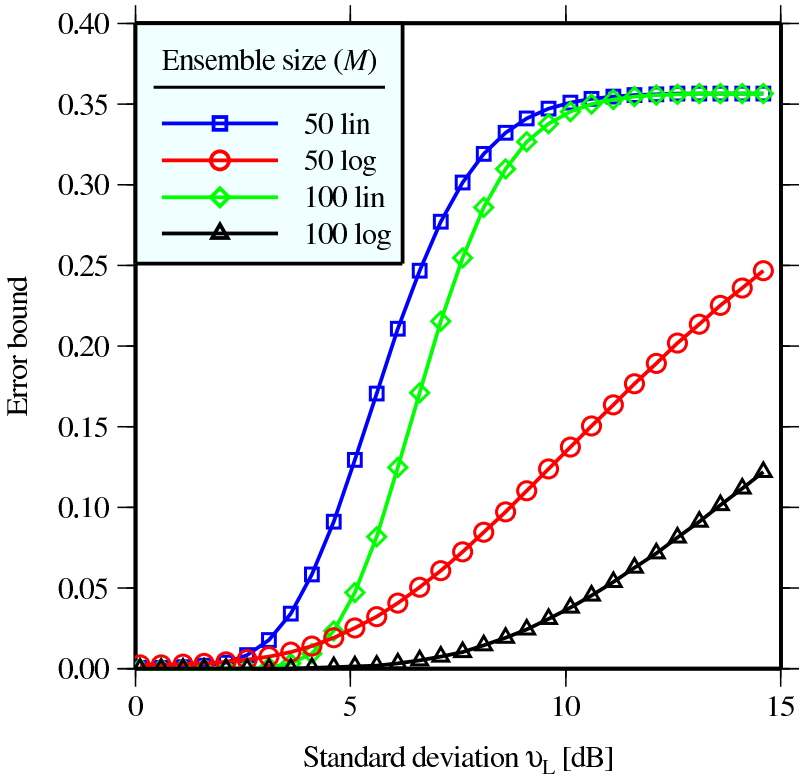


Fig. 6. Error bounds for discrimination between the S and Z classes ($K = 5$ segments). It is assumed that the variance is the same for both classes. With five segments, the values $M = 50$ and $M = 100$ correspond roughly to 10–20 pings spatial averaging. Error bounds for both intensity and log-intensity data are shown for comparison.

knowledge. Randomly drawn observations from the training areas were used to estimate the PDFs.

The second way was with more data and more careful randomization. This was an attempt to make the results and conclusions more robust. For each class we defined a large rectangular training area (rectangles with thick outlines in Fig. 7). Each training area was divided into smaller cells by a 20×20 grid subdivision (for better legibility, Fig. 7 illustrates this for 4×4 grids only). In the trials, half of the cells were assigned to training and the rest to testing. This was done in a random manner, with new random selection for each run. In each run, a specified number of training observations, ranging from 5 to 500 per class, were randomly selected from the combined set of 200 training cells and used to estimate the PDFs.

In its basic form, the present method yields two observations per group of pings, one on either side of the vessel. This limits the lateral (across-track) spatial resolution of the final class map, i.e. the number of independently classified seabed patches. To see how the method may handle environments where the seabed type varies across the swath, we extracted data from smaller, complementary angular sectors. We ran the same set of tests as above on each subsector independently of the others. The sectors were specified in two alternative ways. One approach was to define equiangular sectors, for example 0–25° and 25°–50°. The segments in each sector were computed from (5) applied to measurements in each sector separately. The number of segments K was set equal for all sectors. The other approach was to apply (5) to all data in the full angular range as before, and assign an equal number of segments to each sector. For example, with $K = 8$ (we let K be a power of two), we defined two sectors

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

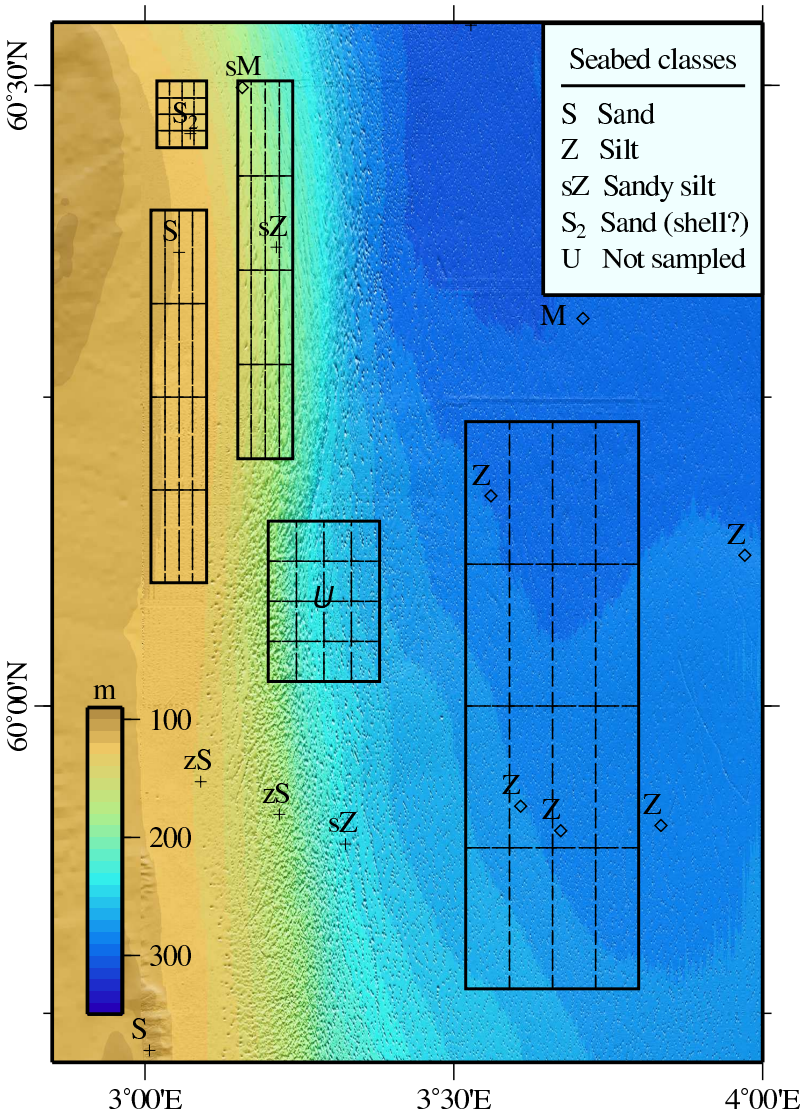


Fig. 7. Training areas (large rectangles with thick lines) with 4×4 subdivisions for five seabed classes in the survey area. In practice 20×20 subdivisions were used. For each test run, cells were randomly assigned for training and testing in a fixed proportion, and a specified number of observations were selected from the set of training cells and used to estimate the PDFs. See also TABLE I, II and Fig. 1.

with segments 1–4 and 5–8 respectively, and four sectors with segments 1–2, 3–4, 5–6, and 7–8 respectively. In the latter approach the PDF for a sector can be obtained from the full PDF by marginalizing out the segments outside the sector, which for normal distributions just means discarding the corresponding components.

B. Results

1) Accuracy: We ran the tests on $y_c(\theta)$ and $10 \log_{10} y_c(\theta)$ data, where y_c was the corrected, dimensionless beam intensity (Section II-B). The results presented here are for the classes S, Z, sZ and U. The S_2 class was left out because its small extent limits the amount of independent training and test data, and because we are less confident about the true class and nature of the S_2 data. However, when all five classes were taken into account, using small training areas around the samples, the results were again very similar to those of Fig. 8–

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

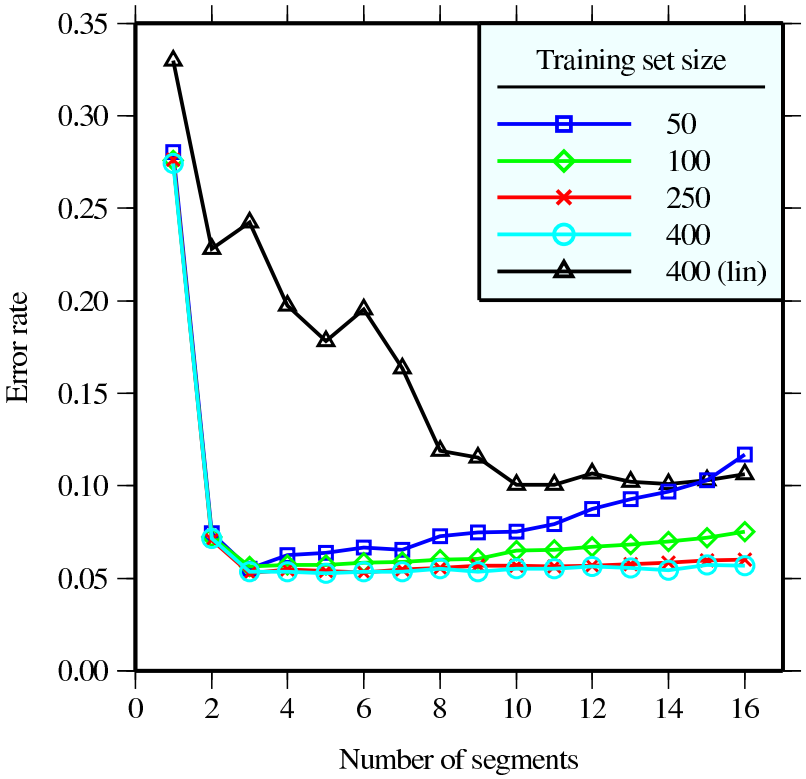


Fig. 8. Error rate vs. observation vector dimension (number of segments). The topmost curve is for intensity data, the rest are for log-intensity data (QDC, $N_p = 20$ pings averaging.)

Fig. 10. A single example (five classes, log-intensity, LDC, 50 pings averaging) can be found in [22].

As Fig. 8 shows, the best results were obtained with log-intensity (decibel) data, about 5 % less errors. Therefore, the results in Fig. 8–Fig. 11 are for log-intensity data, except for the topmost curve in Fig. 8. The estimated error rates were similar for the two approaches described in Section III-A, showing the same trends and minimum error rate. The main difference was a better error rate in very low (1–2) dimensions when using the large training areas with randomized subdivision. The results in this section have all been obtained with large training areas. The incidence angle domain was 0–50°.

The variable parameters are the number of training observations per class, the number of segments (observation vector dimension K), the number of pings (N_p) in the spatial averaging, and the classifier itself (QDC, LDC, or UDC). As the minimum error rate was attained with about five segments, we set $K = 5$ in the plots where observation vector dimension is not the independent variable. For each set of parameters, the estimated error rate is the average fraction of misclassifications from five random experiments.

Fig. 8 shows the error rate as a function of observation vector dimension. The minimum error rate is about 5 %, it is achieved with log-intensity data, and is quite stable up to 16 dimensions. However, the importance of dimensional reduction is apparent when the number of training observations is small. Even with much training data, the error rate increases in more than 16 dimensions, sometimes sharply. Fig. 9 shows the effects of training set size and spatial averaging. These results suggest that adequate estimation of the PDFs was achieved with about 50–100 observations per class, which is about 10–20 times the observation vector dimension (for $K = 5$). Increasing the number of pings in a group of measurements from 10 to

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

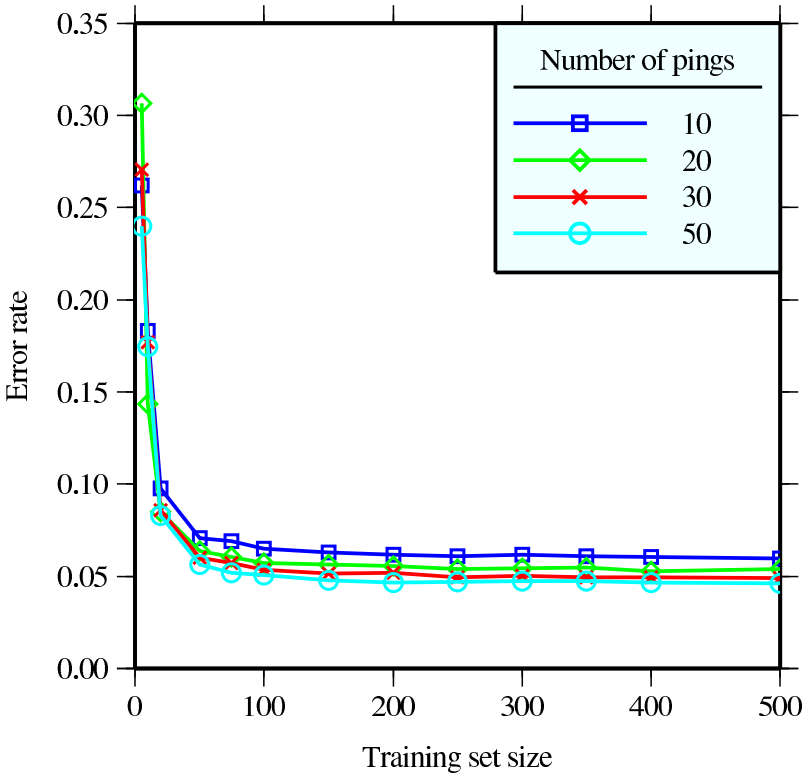


Fig. 9. Error rate vs. training set size (per class). (QDC, $K = 5$ segments.)

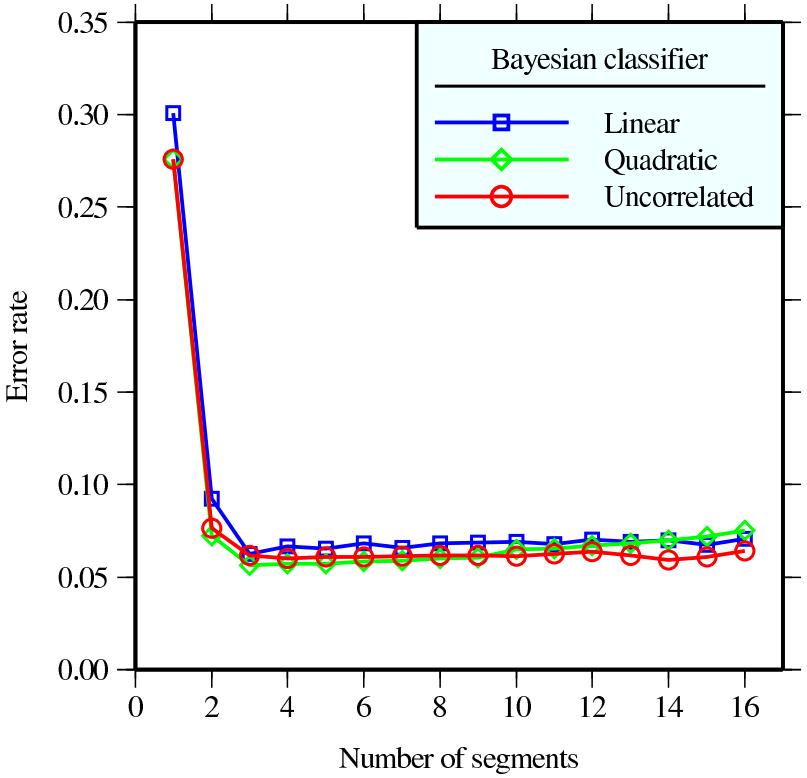
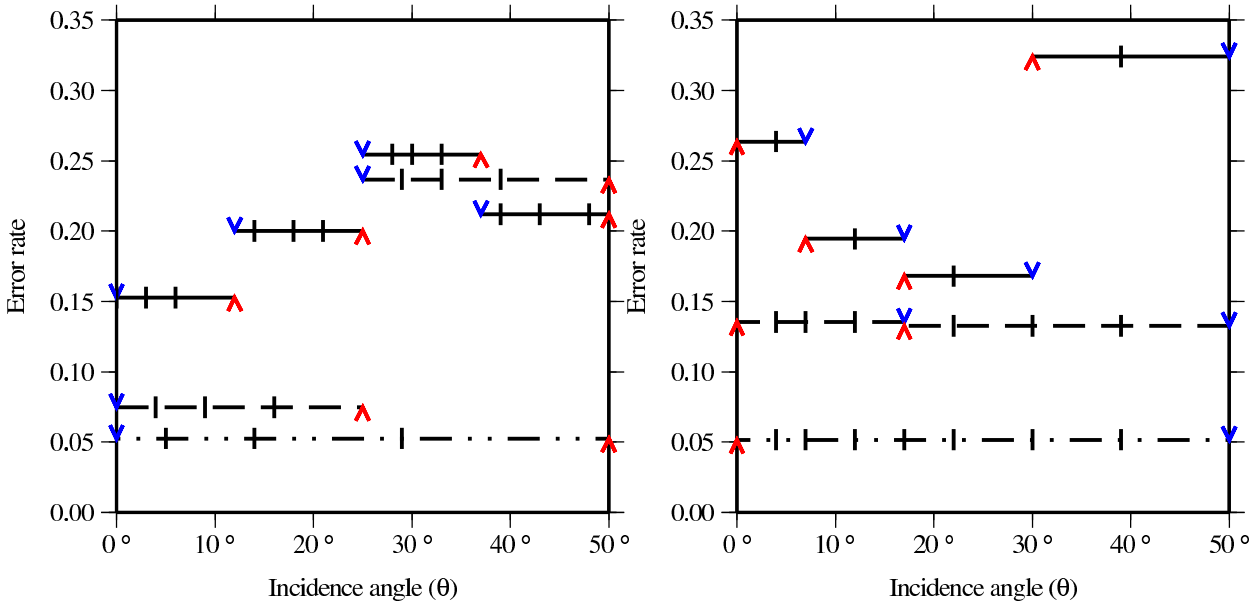


Fig. 10. Error rate vs. feature vector dimension (all classifiers). (100 training observations per class.)

50 reduces the variance, and hence the error rate, as Fig. 6 also implies. Fig. 10 compares the classifiers QDC, LDC and UDC. QDC performs best, but LDC and UDC are not much worse.

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133



(a) Equiangular sectors with four segments each (b) Eight globally determined segments and subsets

Fig. 11. Error rates for observations restricted to angular subsectors (QDC, 20 pings averaging, 300 training observations per class). Each horizontal line between $\wedge$ and $\vee$ marks the width of a sector; the vertical bars mark the limits of segments. In (11b) there are 8 segments determined for the full angular range, with subsectors comprising two or four segments. In (11a), the sectors have been obtained by dividing the the full range into two or four equiangular parts, with four separately determined segments per sector.

In high dimensions, the reduced ratio between training set size and dimension causes more unstable estimates of the inverse covariance matrix, which affects QDC the worst. However, in 16 dimensions or less, the performance degradation of QDC is not significant.

2) Lateral resolution: For equiangular subsectors (Section III-A), the lateral (across-track) resolution is, roughly, the swath width on the seafloor divided by the number of sectors. Recall that we also tested an alternative subdivision method: determine the segments using all data (full swath), and assign an equal number of segments to each subsector. Fig. 11a and Fig. 11b show the error rates for the two approaches. The error rate for each subsector has been plotted as a horizontal line, whose extent indicates the opening angle of the sector. The vertical bars on each sector line show the breakpoints of the segments (Section II-C). In Fig. 11a, each subsector has been obtained by dividing the full swath into one, two, or four subsectors; in Fig. 11b, there are eight segments in total, and sectors with two, four, or eight segments.

For any partition into complementary sectors, the overall accuracy is limited by the accuracy of the worst sector. In these examples (Fig. 11), the best partition consist of the two sectors at 13 % error rate in Fig. 11b. This results in an overall error rate of 13 %. By comparison, the equiangular partitions in Fig. 11a do not yield acceptable error rates at moderate-to-high incidence angles. For almost every sector, the segments are longer at high angles, because the data vary less with angle in this regime. This is as predicted by scattering physics [36, Ch. 13–14]. The plots also show that using the full sector (0-50°) gives the lowest error rate.

3) Sediment map: The classifier (decision rule) was applied to the whole data set to produce a seabed class map of the survey area (Fig. 12). To make this map, the training data were restricted to 2.8 km^2 blocks containing the sediment samples shown in Fig. 1 (inside the

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

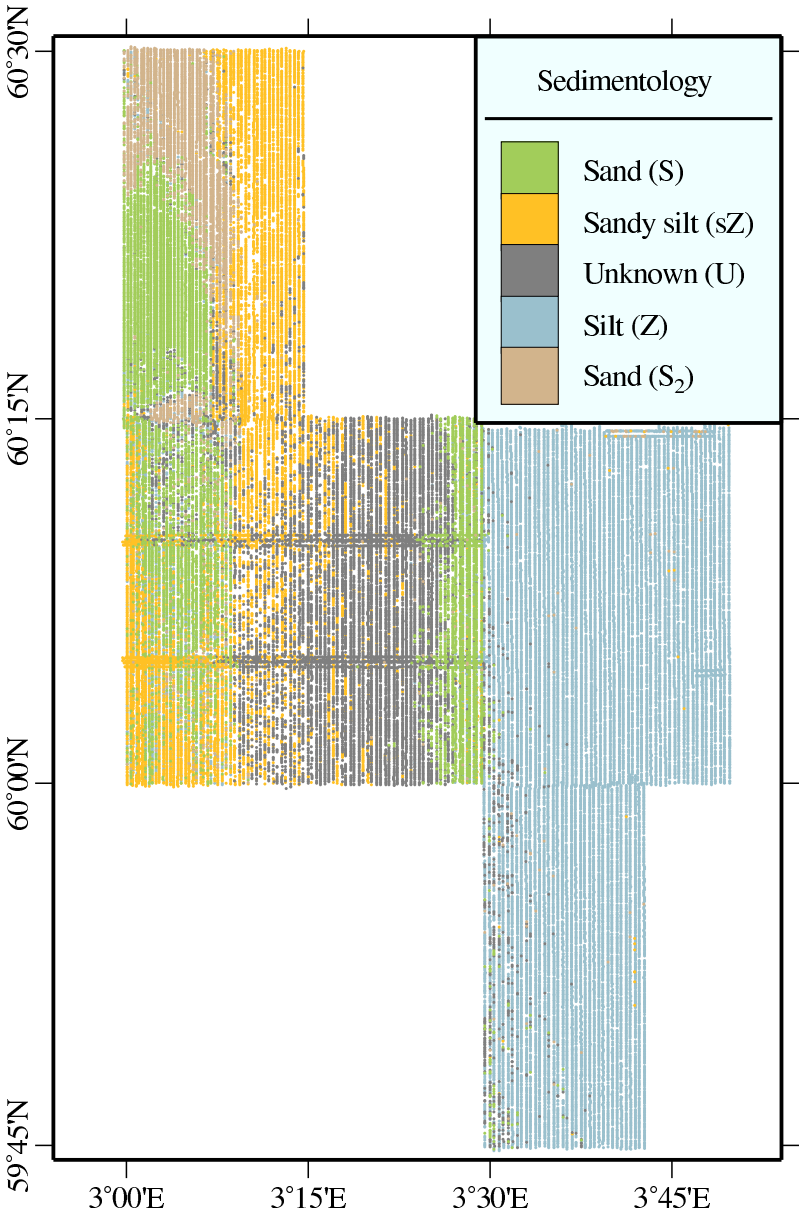


Fig. 12. Classification result for the survey area (five classes).

survey area). We included the S_2 sample, so the map has five sediment classes. The scattering strength observations were represented by a piecewise constant function approximation with $K = 5$ segments. The spatial averaging was $N_p = 20$ pings. The PDFs $p(\boldsymbol{x}|\omega_i)$ were computed using (8) with no assumptions about $\hat{\boldsymbol{\Sigma}}_i$ (i.e. QDC). Each observation $\boldsymbol{d}$ was then classified according to the decision rule (7) with equal prior probabilities. Each observation was assigned the geographical coordinates of the midpoint of its angular sector (on the seafloor). The classified observations were then preprocessed with a 100 m^2 spatial block mode filter before each point was plotted with color codes to get Fig. 12. The map shows that the boundary between the sZ and U classes in particular is not crisp; sZ and U tend to blend into each other. We did anticipate that sZ and U would be similar. In fact, most of the misclassifications found during error rate estimation can be attributed to confusion between the sZ and U classes [22].

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

C. Statistics

In Section II-D it was argued that the PDF $p(\boldsymbol{d}|\omega_i)$ for observations within a class ω_i could be approximated by a multivariate normal model (8). This argument was based on the central limit theorem, so one should only expect it to hold if there are many statistically independent measurements for each observation vector. Moreover, the rate of convergence will be affected by the level of variance in the intrinsic scattering strength within each class. We have also assumed, based on available background information, that the training areas (Fig. 7) are homogeneous, i.e. cover a single sediment class. Since ground truth is limited, it is possible that the empirical distributions are contaminated by samples from other sediment types. Other types of outliers include those caused by acoustical shadows, and those which occur when the sonar bottom tracking algorithm fails for several consecutive pings.

The normal PDFs are on the form $p(\boldsymbol{d}|\omega_i) \propto \exp(-r^2/2)$, where

$$r^2 = (\boldsymbol{d} - \hat{\boldsymbol{\mu}}_i)^T \hat{\boldsymbol{\Sigma}}_i^{-1} (\boldsymbol{d} - \hat{\boldsymbol{\mu}}_i) \tag{20}$$

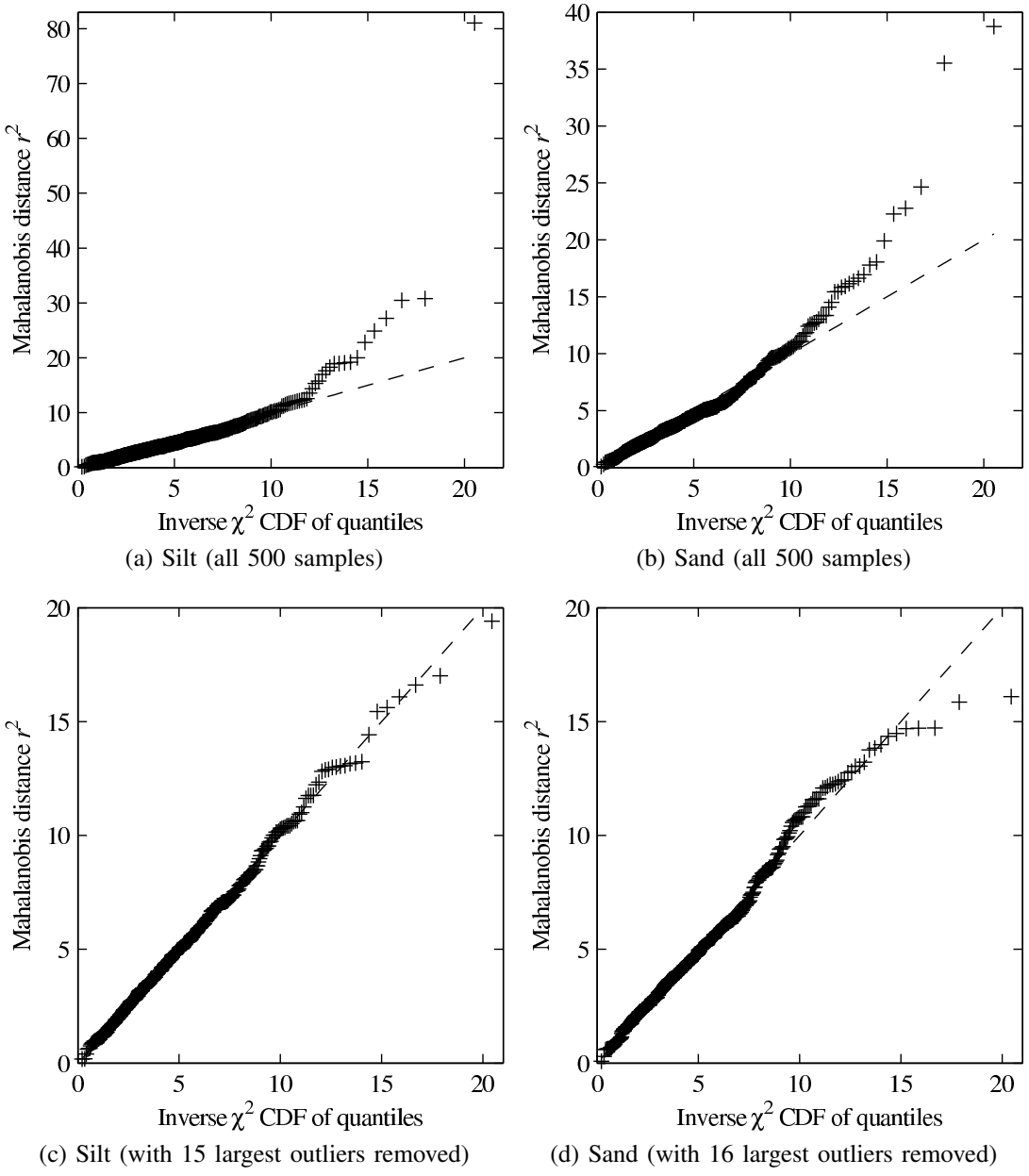
is the (squared) Mahalanobis distance. The classification accuracy is therefore directly influenced by the statistical distribution of r^2. For a true multinormal distribution, the standardized residuals $\boldsymbol{\epsilon} = \boldsymbol{\Sigma}_i^{-1/2}(\boldsymbol{d} - \boldsymbol{\mu}_i)$, where $\boldsymbol{\Sigma}_i^{-1/2}$ is the Cholesky decomposition of $\boldsymbol{\Sigma}_i^{-1}$, form an uncorrelated Gaussian random sequence with zero mean and unit standard deviation. Therefore, $r^2 = \boldsymbol{\epsilon}^T\boldsymbol{\epsilon}$ is the sum of the squares of K standard independent normal random variables, and follows a $\chi^2(K)$-distribution with K degrees of freedom (where K is the dimension of the observation vectors). Thus one way to assess the multinormal assumption is to consider a quantile-quantile plot of the empirical distribution of r^2 versus the $\chi^2(K)$-distribution [45, Ch. 4.6]. This is shown in Fig. 13 for the silt (Z) and sand (S) classes. To make these plots, $n = 500$ random observations from each class (with $K = 5$ dimensions) were drawn from the large training areas in Fig. 7. The spatial averaging was $N_p = 30$ pings. The observed values of r^2 were sorted in ascending order and plotted against $F^{-1}(p_i, K)$, where $p_i = (i-0.5)/n$, $i = 1, \ldots n$ are n equally spaced quantiles, and F^{-1} is the inverse cumulative distribution function (CDF) for $\chi^2(K)$. For a multinormal distribution, the observations are expected to lie on the line with unit slope (dashed line).

Fig. 13a and Fig. 13b reveal some large outliers. To reduce the influence of outliers on $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$, hence make the outliers stand out more, $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\Sigma}}$ were robustly estimated using the minimum covariance determinant estimator [46], [47]. Fig. 13c and Fig. 13d show the same data sets after first removing the 15 (for silt) and 16 (for sand) most extreme observations. Note that the tail of the distribution in both classes is sensitive to how many observations are considered as outliers. If too many or too few observations are excluded, the remaining extreme observations will deviate more from the unit slope curve. However, for r^2 less than about 10 or 12, the observations follow the unit slope curve very closely.

IV. Discussion

The purpose of this work has been to test a method for acoustical mapping of seabed composition using measured scattering strength versus incidence angle. The questions that arose concerned how effective the $S(\theta)$ function is for seabed type discrimination; what is a suitable statistical model for this function; what spatial resolution can be achieved; and how much training data or prior knowledge about backscatter statistics is required. The main findings for the North Sea data set were:

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133



(a) Silt (all 500 samples)

(b) Sand (all 500 samples)

(c) Silt (with 15 largest outliers removed)

(d) Sand (with 16 largest outliers removed)

Fig. 13. Quantile-quantile plots for the distribution of the robustly determined Mahalanobis distance r^2 versus the $\chi^2(K)$-distribution with K degrees of freedom; here $K = 5$ (the dimension of the observation vectors). Fig. (a) and (b) show the result for 500 random observations from the silt (Z) and sand (S) classes respectively, drawn from the large training areas shown in Fig. 7. Fig. (c) and (d) show the result after first eliminating the 15 largest outliers for silt and the 16 largest outliers for sand. For a multinormal distribution, the observations are expected to lie on the line with unit slope (dashed line).

- The minimum error rate was about 5 %; most errors occured because we have split a large area with sparse ground truth into two classes which are physically similar.
- When the spatial averaging is decreased from 50 pings to 10 pings, the error rate increases from about 5 % to about 7 %. As this increase is not very large in absolute terms, it seems acceptable to use 10–20 pings averaging, which corresponds to a distance of about 25–50 m (5 knots, 1 Hz ping rate).
- The minimum error rate was attained with 3–5 segments in the piecewise constant approximation of $S(\theta)$. Using a single segment (overall average scattering strength) yields

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

a comparatively poor result, about 30 % errors. The less training data available, the worse performance in higher dimensions.

- Getting a low error rate required about 50–100 training observations per class. This agrees with the long-established rule-of-thumb that the ratio between training set size and dimension should be at least 10 (since here, $K = 3$–5) [40]. For complex classifiers, the ratio should be higher, which is a reason to choose the simple Bayesian normal classifier. With training data restricted to small areas near ground truth sites, the results were almost as good as when data were picked from large training areas in a random manner.
- The lowest error rate was achieved with QDC, the performance of LDC and UDC was marginally worse.
- For data restricted to smaller incidence angle sectors, the error rate was frequently above 15 % (Fig. 11). However, for the sectors 0–17°and 17°–50°(either side of the ship), the overall error rate was 13 % (Fig. 11b). By overall error rate we mean the highest error rate observed in any sector. If accurate classification can be done in several independent, complementary sectors, there will be more observations across the swath, hence better across-track resolution in the class map. We think that the 13 % error rate can be improved by taking into account contextual information, by weighting the prior probabilities according to the neighboring observations. Across-track resolution comparable to the along-track resolution may therefore be within reach for the present data set, but has not been examined further in this paper.
- A key question with respect to classification accuracy is how well the multinormal PDF approximates the true density. The analysis of Section III-C suggests that, excluding some outliers, the observed data are consistent with the assumptions. Each seabed type forms a cluster which is well represented by a normal PDF in the feature space, even when observations are picked at random across large areas. Because the feature vectors are formed by taking averages, the PDFs should theoretically approach normal distributions when the number of independent measurements is large. Averaging also reduces the variance of each distribution, which improves the classification accuracy.

The less training data, the more important feature reduction becomes. The present approach, piecewise constant approximation, is not the only possible solution. Alternatives include the principal components transform (PCA), polynomial approximation, and wavelet transforms. We tried to use the PCA transform on the scattering strength observations (4), and the results were similar to those of Fig. 8–Fig. 10. The piecewise constant approximation has three main advantages. First, each component of a feature vector has a physical meaning, and each component measures the same quantity, the average scattering strength. Second, taking averages over angular sectors which are typically wider than the uncertainty of the incidence angle estimates, reduces the significance of this uncertainty. Third, each feature vector component corresponds to an angular sector which ensonifies a separate patch of seafloor. Therefore there is a natural way to increase the across-track resolution of the present method, as discussed in connection with Fig. 11.

In practice, across-track resolution hinges not on the number of independently classified observations, but on the ability to detect class transitions within the swath. Therefore, one way to proceed would be to construct an algorithm which detects discontinuities in $S(\theta)$, and suggests a partition of the incidence angle domain. Then each partition (sector) would

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

be classified independently. The experiments of this paper (Fig. 11) merely suggest how successful classification using smaller sectors would be.

With sufficient training data, the error rate was stable in dimensions less than $K = 16$. More sophisticated classification methods like support vector machines [6, Ch. 12] did not improve results, nor did non-parametric PDF estimation techniques. This indicates that the 5 % limit on accuracy is inherent in the data, particularly in the unclear distinction between the U and sZ classes, and also due to insufficient ground truth in these two areas. The scattering strength in the sZ area is surprisingly high, but consistently so in data from three different cruises in the area. The western boundary of the sZ region also closely follows a transition in seabed morphology (Fig. 12). On a large scale, the seabed is much more rough in the sZ and U regions, with a very irregular pattern of depressions. This area corresponds roughly to the western slope of the Norwegian Trench. The depressions may be a result of erosion by a combination of water currents and fluid seepage up through the shallow geological layers, which causes seabed sediments to be lifted into suspension. This theory was put forward by Hovland [30], who observed very shallow gas blankets in seismic boomer data from a part of the western slope, with apparent vertical escape routes matching the occurrence of elongated depressions in the seabed. A similar phenomenon has been observed further south in the Norwegian Trench (the Skagerrak) [31]. It is not possible based on two grab samples to say whether the seabed has large roughness also at the scale of ultrasonic scattering, but the presence of vertically migrating shallow gas may be a partial explanation for the observed high scattering strength.

The fact that the statistical signal model (9) is formulated in terms of the spatially averaged scattering strength s (1) (instead of σ) reflects a subtle point in ASC. Namely, the signal statistics depend on the ensonification function $\Omega(\boldsymbol{r}, \boldsymbol{r}_0)$ in a way which is not simple to compensate for. This can be a problem because Ω is not related to sediment properties, but rather to beampattern, incidence angle, pulse length and water depth. However, the present method operates only on the mean, corrected intensity $\langle y_c \rangle \approx \langle \sigma \rangle$ (3), and is therefore not sensitive to Ω.

Besides the problem of lateral resolution, the present method has two main limitations. The first is the possibility that different seabed types may have similar scattering strength functions. The second is the problem of determining how many seabed types are recorded in the data, especially when there is little available ground truth, or when there is a gradual change in sediment composition, hence no clear boundary between distinct seabed types. Our data set illustrates these dilemmas. In advance we supposed that the sediment grain size distribution would change gradually from shallow to deep water. Thus the sand region between the U and Z areas in Fig. 12 (at about $03\,^\circ 30\,'\mathrm{E}$) was an unexpected classification result. There are three alternative explanations. There may actually be sand in the area, because sand particles that are moved by currents from the plateau possibly accumulate at the bottom of the slope. A core sample at $03\,^\circ 23\,'\mathrm{E},\ 59\,^\circ 40\,'\mathrm{N}$ contained a large sand fraction (28 %); likewise a core sample at $03\,^\circ 23\,'\mathrm{E},\ 60,^\circ\ 43\,'\mathrm{N}$ had a 50 cm sandy surficial layer [25]. Alternatively, there may be a silty sediment with scattering strength similar to that of sand, due to a special combination of surface roughness and scattering from suspended fluid inclusions or coarse material. The third possibility is that the likelihood $p(\boldsymbol{d}|\omega)$ of the data is small with respect to all the predefined classes $\omega \in \{\mathrm{S},\ \mathrm{S}_2,\ \mathrm{Z},\ \mathrm{sZ},\ \mathrm{U}\}$. This implies that a new class should be introduced, or else a seabed type (here: sand) with low likelihood may get a high posterior probability due to the low evidence factor $p(\boldsymbol{d})$ in Bayes' theorem. The unsupervised Bayesian approaches of [33],

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

[48] and [21] determine the number of classes based on a Gaussian mixture model for the complete data set. Similar solutions may extend the present method.

The amount of training data needed to achieve a 5 % error rate is about 100 observations per class using 30 pings averaging. If a 7 % error rate is acceptable, 50 observations using 10 pings averaging are sufficient. Then, in a survey with 400 m line spacing and a 1 Hz ping rate at 5 kn speed, sufficient training data could be picked from 0.25 km^2 areas surrounding the ground truth sites. Because the classification is based on a physical quantity (mean scattering strength) which is instrument independent, properly processed data from any calibrated MBES survey could in principle be used for training. Moreover, with modeled scattering strength functions $\varsigma(\theta)$ and a given level of intrinsic variance $\nu^2(\theta)$, the equations from this paper can be used to estimate theoretical PDFs, and the likelihood of observations with respect to the theoretical PDFs can be studied. From this point of view, our method is therefore related to the inverse theory approach of e.g. [12], [13].

The present results are compatible with works like [12]–[14], [16]–[19], which all obtained good results by applying classification or inverse methods to the $S(\theta)$ function. There are of course successful alternative strategies which analyze fluctuations in the backscatter levels for discriminating between sediment types. An important example is texture analysis by means of second-order statistics (gray level co-occurrence matrices) applied to corrected sidescan mosaics [48]–[51]. In this approach, the pixel values in the mosaic are quantized to, say, L values (gray levels). Then, for any rectangular neighborhood in the mosaic, one may form an $L \times L$ matrix $\boldsymbol{G}(\Delta x, \Delta y)$, such that G_{ij} is the relative frequency of two pixels with gray levels i and j respectively being separated by pixel distance $(\Delta x, \Delta y)$ [52]. From $\boldsymbol{G}(\Delta x, \Delta y)$ one may derive several attributes which represent the character of the texture [52], [53].

This kind of information is discarded when using only scattering strength versus angle. One solution, based on the angular range analysis (inversion) of [12], is to visually inspect backscatter mosaics to ensure that the inversion is applied to data from acoustically similar areas [54]. For classification too, it seems that a hybrid method is possible, where both the angular variation and the backscatter scintillations are taken into account. For example, consider again as in Section II-C a subdivision of the incidence angle range, $0 < \theta_1 < \ldots < \theta_{K-1} < \theta_{\max}$, where for a given K the subdivision is such that the variation with angle is effectively minimized within each segment. Assume that the beamformed data have been corrected for propagation loss and beampattern variations, and mapped to a gray level range. Then within each segment, for a sequence of N_p consecutive pings, one may form an analogous co-occurrence matrix $\boldsymbol{G}(\Delta x, \Delta\theta)$, where now Δx is along-track separation between pairs of samples, and $\Delta\theta$ is the difference in estimated incidence angle. A special case would be to set $\Delta x \approx 0$ and consider only separation in angle. This results in additional features for each segment besides the mean scattering strength, e.g. the entropy and homogeneity properties used in [50]. As shown in Section III-B2 and Fig. 11, it may be possible to achieve a reasonable classification accuracy also after dividing the sonar swath into smaller separate sectors, and consequently get a better lateral resolution. We hypothesize that this accuracy can be improved by augmenting the observation vector $\boldsymbol{d}$ (Section II-C) with such texture attributes.

V. Conclusion

We have presented a statistical seabed classification method for multibeam echo sounders, and discussed its practical applicability. The method uses the observed scattering strength

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

dependence on incidence angle $S(\theta)$. To obtain a robust statistical description from scarce training data, $S(\theta)$ is represented by a piecewise constant function approximation. Theoretical error bounds and experimental results on a North Sea data set showed that good classification accuracy and spatial resolution is possible even with limited training data.

APPENDIX A

We want to compute asymptotic error bounds for discrimination between two seabed types, ω_1 and ω_2, in terms of the mean and variance of the scattering strength. It is useful to reinterpret (9) as a multiplicative noise model similar to the product model for clutter statistics in radar images [55]. The PDF for the product of two random variables, $Y = SX$, is

$$p_Y(y) = \int_{-\infty}^{\infty} \frac{1}{|s|} p_{S,X}(s, \frac{y}{s})\, \mathrm{d}s. \tag{21}$$

If $S \geq 0$, S and X are independent, and

$$p_X(x) = e^{-x}, \tag{22}$$

then (21) reduces to (9) [with $p_S(s) = f_{s|\theta}(s)$]. X represents noise (speckle) which is exponentially distributed with unit variance.

The variance decomposition formula applied to Y and S is

$$\mathrm{Var}[Y] = \mathrm{Var}[\mathrm{E}[Y|S]] + \mathrm{E}[\mathrm{Var}[Y|S]]. \tag{23}$$

By assumption (9), $\mathrm{E}[Y|S] = S$ and $\mathrm{Var}[Y|S] = S^2$, so

$$\begin{aligned} \mathrm{Var}[Y|\Theta] &= \mathrm{Var}[S|\Theta] + \mathrm{E}[S^2|\Theta] \\ &= 2\mathrm{Var}[S|\Theta] + \mathrm{E}[S|\Theta]^2. \end{aligned} \tag{24}$$

This result can also be found directly from the relation

$$\mathrm{E}[Y^n|\Theta] = n!\, \mathrm{E}[S^n|\Theta], \tag{25}$$

which follows from (9), but (23) is valid for other noise models as well. The PDF for $\log X$ is

$$f_{\log X}(u) = e^u e^{-e^u},$$

with $\mathrm{E}[\log X] = -\gamma$, the Euler-Mascheroni constant, and $\mathrm{Var}[\log X] = \pi^2/6$. Thus, since S and X are independent,

$$\begin{aligned} \mathrm{Var}[\log Y|\Theta] &= \mathrm{Var}[\log S + \log X] \\ &= \mathrm{Var}[\log S] + \frac{\pi^2}{6}, \end{aligned} \tag{26}$$

$$\mathrm{E}[\log Y|\Theta] = \mathrm{E}[\log S] - \gamma. \tag{27}$$

With notation as in (11) and (16), we find using (24)-(27) that

$$\mu_k \equiv \mathrm{E}[d_k] = \mathrm{E}[Y] = \langle \varsigma \rangle_k \tag{28}$$

$$M\Sigma_{kk} = \mathrm{Var}[Y] = 2\left\langle \nu^2 \right\rangle_k + \left\langle \varsigma^2 \right\rangle_k + \mathrm{Var}_\Theta\left[\varsigma\right] \tag{29}$$

for linear intensity, and

$$\mu_k \equiv \mathrm{E}[d_k] = \mathrm{E}[\log Y] = \langle \varsigma_L \rangle_k - \gamma \tag{30}$$

$$M\Sigma_{kk} = \mathrm{Var}[\log Y] = \left\langle \nu_L^2 \right\rangle_k + \pi^2/6 + \mathrm{Var}_\Theta\left[\varsigma_L\right] \tag{31}$$

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

for log-intensity. The first equality in (29) and (31) is only strictly true in the asymptotic (large M) limit. By construction (5), the breakpoints $\theta_0, \ldots, \theta_K$ are such that the variation of mean intensity with angle is minimal within a segment (for a fixed number of segments K). Therefore, the third term in (29) and (31) is small compared to the first term for most segments.

For modeled (or observed) scattering strength with associated variance at N angles, these formulae can be implemented using (6). If n_k is the number of bins in the kth segment, and $\boldsymbol{n} = \begin{bmatrix} n_1 & \ldots & n_K \end{bmatrix}^T$, then

$$\boldsymbol{\mu} = \boldsymbol{\varsigma}\boldsymbol{W} \tag{32}$$

$$M\mathrm{diag}(\boldsymbol{\Sigma}) = 2\left(\boldsymbol{\nu} \star \boldsymbol{\nu}\right)\boldsymbol{W} + \left(\boldsymbol{\varsigma} \star \boldsymbol{\varsigma}\right)\boldsymbol{W} + \\ + \left(\boldsymbol{\varsigma} - \boldsymbol{m}\right) \star \left(\boldsymbol{\varsigma} - \boldsymbol{m}\right)\boldsymbol{W} \tag{33}$$

$$\boldsymbol{m} = \boldsymbol{\mu}\boldsymbol{W}^T \star \boldsymbol{n},$$

where $\boldsymbol{\varsigma} = \begin{bmatrix} \varsigma_1 & \ldots & \varsigma_N \end{bmatrix}$, $\boldsymbol{\nu} = \begin{bmatrix} \nu_1 & \ldots & \nu_N \end{bmatrix}$, and $\star$ denotes element-wise multiplication. For log-intensity, the formulae are

$$\boldsymbol{\mu} = \boldsymbol{\varsigma}_L\boldsymbol{W} \tag{34}$$

$$M\mathrm{diag}(\boldsymbol{\Sigma}) = \left(\boldsymbol{\nu}_L \star \boldsymbol{\nu}_L\right)\boldsymbol{W} + \frac{\pi^2}{6} \\ + \left(\boldsymbol{\varsigma}_L - \boldsymbol{m}\right) \star \left(\boldsymbol{\varsigma}_L - \boldsymbol{m}\right)\boldsymbol{W} \tag{35}$$

$$\boldsymbol{m} = \boldsymbol{\mu}\boldsymbol{W}^T \star \boldsymbol{n}.$$

In the asymptotic limit when $p(\boldsymbol{d}|\omega_1)$ and $p(\boldsymbol{d}|\omega_2)$ are normal distributions, the Bhattacharyya bound for the probability of error is [5, Ch. 2.8]

$$P(\text{error}) \leq \sqrt{P(\omega_1)P(\omega_2)}e^{-\beta}, \tag{36}$$

where

$$\beta = \frac{1}{8}\left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right)^T \left[\frac{\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2}{2}\right]^{-1} \left(\boldsymbol{\mu}_2 - \boldsymbol{\mu}_1\right) \\ + \frac{1}{2}\log\frac{\det\frac{1}{2}\left[\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2\right]}{\sqrt{\det\boldsymbol{\Sigma}_1 \det\boldsymbol{\Sigma}_2}}. \tag{37}$$

The angular segments correspond to separate patches of seabed. We therefore consider an error bound for data with independent features (the UDC assumption holds true). Consequently the covariance matrix is diagonal, with entries given by (29) or (31). Then

$$\beta = \frac{1}{2}\sum_{k=1}^{K}\left[\frac{\left(\langle\varsigma_2\rangle_k - \langle\varsigma_1\rangle_k\right)^2 M/2}{\mathrm{Var}[Y;k]_1 + \mathrm{Var}[Y;k]_2} + \log f\left(\sqrt{\rho_k}\right)\right], \tag{38}$$

with

$$\rho_k = \frac{\mathrm{Var}[Y;k]_1}{\mathrm{Var}[Y;k]_2}, \tag{39}$$

$$f(x) = \frac{1}{2}\left(x + \frac{1}{x}\right). \tag{40}$$

If $x \geq 0$, then $f(x) \geq 1$, with minimum for $x = 1$. Hence both terms inside the brackets are always greater than zero. The second term vanishes when the LDC assumption is correct.

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

When the variance becomes large, the second term dominates, and $P(\text{error})$ tends to a limit determined by the variance ratio ρ_k. The absolute upper limit (with equal prior probabilities) is $P(\text{error}) = 0.5$ (pure guessing). For log-intensity we must replace $\langle \varsigma \rangle_k$ with $\langle \varsigma_L \rangle_k$, and $\mathrm{Var}[Y;k]$ (29) with $\mathrm{Var}[\log Y, k]$ (31).

To compare classifiers based on intensity and log-intensity, we need to compare the means and variances for S and $\log S$ without knowledge of the underlying scattering cross section PDF $f_{S|\Theta}(s|\theta)$. For a given set of constraints, the least biased estimate for $f_{S|\Theta}(s|\theta)$ is the distribution $p(s)$ which maximizes the entropy [44], which for continuous distributions is [5, App. A.7]

$$H = -\int_{-\infty}^{\infty} p(s) \log p(s)\, \mathrm{d}s. \tag{41}$$

If we constrain the mean and variance of $\log S$,

$$\mathrm{E}[\log S|\Theta] = \int \log s f_{S|\Theta}(s|\theta)\, \mathrm{d}s = \varsigma_L(\theta),$$
$$\mathrm{Var}[\log S|\Theta] = \int (\log s - \varsigma_L(\theta))^2 f_{S|\Theta}(s|\theta)\, \mathrm{d}s = \nu_L^2(\theta),$$

then the maximum entropy PDF for S is log-normal. Thus [56],

$$f_{S|\Theta}(s|\theta) = \frac{1}{s\nu_L\sqrt{2\pi}} \exp\left[-\frac{(\log s - \varsigma_L)^2}{2\nu_L^2}\right], \tag{42}$$

for which

$$\varsigma = e^{\varsigma_L + \nu_L^2/2}, \tag{43}$$
$$\nu^2 = \left(e^{\nu_L^2} - 1\right) e^{2\varsigma_L + \nu_L^2}. \tag{44}$$

Hence, given the scattering strength $\varsigma_L(\theta)$ from a physics model or observation, we have the necessary quantities to compare the Bhatacharyya bound (36) for Y and $\log Y$, parameterized by the variance $\nu_L^2(\theta)$ and the number M. These quantities are given by (28)–(31), (38)–(40), and (43)–(44).

ACKNOWLEDGMENT

The authors thank the Norwegian Defence Research Establishment (FFI) for permission to use the data, especially A. Karlsen and his group, who collected and made the data available. Dr. A. Ommundsen of FFI gave valuable comments on the manuscript. The authors also thank Dr. A. C. Jensen, who provided Matlab code for piecewise constant function approximation, and R. Eckhoff of Kongsberg Maritime, who helped them understand the signal processing of the EM echo sounders. Finally, two anonymous reviewers provided valuable comments and suggestions that helped improve the manuscript.

REFERENCES

[1] J. T. Anderson, D. Van Holliday, R. Kloser, D. G. Reid, and Y. Simard, "Acoustic seabed classification: current practice and future directions," *ICES Journal of Marine Science*, vol. 65, no. 6, pp. 1004–1011, Sep. 2008.

[2] C. J. Brown and P. Blondel, "Developments in the application of multibeam sonar backscatter for seafloor habitat mapping," *Applied Acoustics*, vol. 70, no. 10, pp. 1242–1247, Oct. 2009.

[3] R. J. Kloser, J. D. Penrose, and A. J. Butler, "Multi-beam backscatter measurements used to infer seabed habitats," *Continental Shelf Research*, vol. 30, no. 16, pp. 1772–1782, Sep. 2010.

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

[4] A. Rattray, D. Ierodiaconou, L. Laurenson, S. Burq, and M. Reston, "Hydro-acoustic remote sensing of benthic biological communities on the shallow South East Australian continental shelf," *Estuarine Coastal and Shelf Science*, vol. 84, no. 2, pp. 237–245, Sep. 2009.

[5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY: Wiley-Interscience, 2001.

[6] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning. Data Mining, Inference and Prediction*, 2nd ed., ser. Springer Series in Statistics. New York, NY: Springer, 2009.

[7] X. Lurton, *An Introduction to Underwater Acoustics: Principles and Applications*, 2nd ed. Berlin/DE: Springer, 2010.

[8] D. R. Jackson and M. D. Richardson, *High-Frequency Seafloor Acoustics*. New York, NY: Springer, 2007.

[9] S. T. McDaniel, "Fluctuations of the acoustic field Bragg-backscattered from rough surfaces," *Waves in Random Media*, vol. 6, no. 3, pp. 229–249, 1996.

[10] G. Le Chenadec, J. Boucher, and X. Lurton, "Angular dependence of *K*-distributed sonar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1224–1235, May 2007.

[11] A. P. Lyons, S. F. Johnson, D. A. Abraham, and E. Pouliquen, "High-frequency scattered envelope statistics of patchy seafloors," *IEEE J. Ocean. Eng.*, vol. 34, no. 4, pp. 451–458, Oct. 2009.

[12] L. Fonseca and L. Mayer, "Remote estimation of surficial seafloor properties through the application angular range analysis to multibeam sonar data," *Mar. Geophys. Res.*, vol. 28, no. 2, pp. 119–126, Jul. 2007.

[13] K. Haris, B. Chakraborty, C. De, R. G. Prabhudesai, and W. Fernandes, "Model-based seafloor characterization employing multi-beam angular backscatter data-a comparative study with dual-frequency single beam," *J. Acoust. Soc. Am.*, vol. 130, no. 6, pp. 3623–3632, Dec. 2011.

[14] B. Chakraborty, H. W. Schenke, V. Kodagali, and R. Hagen, "Seabottom characterization using multibeam echosounder angular backscatter: An application of the composite roughness theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 5, pp. 2419–2422, Sep. 2000.

[15] C. De and B. Chakraborty, "Model-based acoustic remote sensing of seafloor characteristics," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3868–3877, Oct. 2011.

[16] G. Canepa and N. G. Pace, "Seafloor segmentation from multibeam bathymetric sonar," in *Proceedings of the Fifth European Conference on Underwater Acoustics, ECUA 2000*, M. E. Zakharia, P. Chevret, and P. Dubail, Eds., vol. 1. Lyon, France: European Commision, Jul. 2000, pp. 361–367.

[17] Z.-H. Michalopoulou, D. Alexandrou, and C. de Moustier, "Application of neural and statistical classifiers to the problem of seafloor characterization," *IEEE J. Ocean. Eng.*, vol. 20, no. 3, pp. 190–197, Jul. 1995.

[18] L. J. Hamilton and I. Parnum, "Acoustic seabed segmentation from direct statistical clustering of entire multibeam sonar backscatter curves," *Continental Shelf Research*, vol. 31, no. 2, pp. 138–148, Feb. 2011.

[19] B. Chakraborty, V. Kodagali, and J. Baracho, "Sea-floor classification using multibeam echo-sounding angular backscatter data: A real-time approach employing hybrid neural network architecture," *IEEE J. Ocean. Eng.*, vol. 28, no. 1, pp. 121–128, Jan. 2003.

[20] S. Dugelay, "Deep seafloor characterization with multibeam echosounders using image segmentation and angular acoustic variations," in *OCEANS '96. MTS/IEEE. Prospects for the 21st Century. Conference Proceedings*, vol. 3. Ft Lauderdale, FL: IEEE, Sep. 1996, pp. 1545–1550.

[21] D. G. Simons and M. Snellen, "A Bayesian approach to seafloor classification using multi-beam echosounder backscatter data," *Applied Acoustics*, vol. 70, no. 10, pp. 1258–1268, Oct. 2009.

[22] K. Landmark, A. H. S. Solberg, C.-I. C. Nilsen, A. Austeng, and R. E. Hansen, "Bayesian seabed classification using angle-dependent backscatter data from multibeam sonar," in *11th European Conference on Underwater Acoustics 2012 (ECUA 2012)*, ser. Proceedings of the Institute of Acoustics Volume 34 Pt.3. Institute of Acoustics (IOA), Oct. 2012, pp. 417–424.

[23] A. C. Jensen and A. S. Solberg, "Fast hyperspectral feature reduction using piecewise constant function approximations," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 547–551, Oct. 2007.

[24] G. Winkler and V. Liebscher, "Smoothers for discontinuous signals," *J. Nonparametric Statist.*, vol. 14, no. 1/2, pp. 203–222, 2002.

[25] E. J. Eidem and K. Landmark, "Acoustic seabed classification using QTC IMPACT on single-beam echo sounder data from the Norwegian Channel, northern North Sea," *Continental Shelf Research*, vol. 68, pp. 1–14, Oct. 2013.

[26] H. P. Seirup *et al.*, "Quaternary of the Norwegian Channel: glaciation history and paleooceanography," *Norsk geologisk tidsskrift*, vol. 75, no. 2-3, pp. 65–87, 1995.

[27] H. de Haas, E. Okkels, and T. C. E. van Weering, "Recent sedimentation and organic carbon burial in a shelf area: the North Sea," *Norges Geologiske Undersøkelse Bulletin*, vol. 430, pp. 57–65, 1996.

[28] E. S. Andersen, S. R. Østmo, C. F. Forsberg, and S. J. Lehman, "Late- and post-glacial depositional environments in the Norwegian Trench, northern North Sea," *Boreas*, vol. 24, pp. 48–63, 1995.

[29] L. Rise and K. Rokoengen, "Surficial sediments in the Norwegian sector of the North Sea between $60°30'$ and $62°$N," *Marine Geology*, vol. 58, pp. 287–317, 1984.

[30] M. Hovland, "Elongated depressions associated with pockmarks in the western slope of the Norwegian Trench," *Marine Geology*, vol. 51, no. 1-2, pp. 35–46, Feb. 1983.

[31] R. Bøe, L. Rise, and D. Ottesen, "Elongate depressions on the southern slope of the Norwegian Trench (Skagerrak): morphology and evolution," *Marine Geology*, vol. 146, no. 1-4, pp. 191–203, Apr. 1998.

[32] R. L. Folk, "The distinction between grain size and mineral composition in sedimentary-rock nomenclature," *J. Geol.*, vol. 62, no. 4, pp. 344–359, 1954.

[33] J. M. Preston, A. C. Christney, L. S. Beran, and W. T. Collins, "Statistical seabed segmentation—from images and echoes to objective clustering," in *Proc. Seventh European Conf. Underwater Acoustics ECUA 2004*. Delft, The Netherlands: Delft University of Technology, Jul. 2004, pp. 813–818.

[34] E. Hammerstad, "Backscattering and seabed image reflectivity," Kongsberg Maritime, EM technical note, Jan. 2000. [Online]. Available: http://www.km.kongsberg.com/ks/web/nokbg0397.nsf/AllWeb/C2AE0703809C1FA5C1257B580044DD83/$file/EM_technical_note_web_BackscatteringSeabedImageReflectivity.pdf?OpenElement

[35] R. Eckhoff, "Back scatter corrections," by personal communication, Kongsberg Maritime, Aug. 2009.

[36] D. R. Jackson and M. D. Richardson, *High-Frequency Seafloor Acoustics*. New York, NY: Springer, 2007.

[37] A. N. Gavrilov and I. M. Parnum, "Fluctuations of seafloor backscatter data from multibeam sonar systems," *IEEE J. Ocean. Eng.*, vol. 35, no. 2, pp. 209–219, Apr. 2010.

[38] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, 2nd ed., ser. Modern Acoustics & Signal Processing. New York, NY/US: Springer, 2011.

[39] R. E. Francois and G. R. Garrison, "Sound absorption based on ocean measurements. Part 2: Boric acid contribution and equation for total absorption," *J. Acoust. Soc. Amer.*, vol. 72, no. 6, pp. 1879–1890, 1982.

[40] A. K. Jain and B. Chandrasekaran, "Dimensionality and sample size considerations in pattern recognition practice," in *Handbook of statistics*, P. R. Krishnaiah and L. N. Kanal, Eds. Amsterdam: North-Holland, 1982, vol. 2, pp. 835–855.

[41] R. P. W. Duin *et al.* (2007) PRTools4 A Matlab Toolbox for Pattern Recognition. [Online]. Available: http://www.prtools.org/

[42] P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recogn. Lett.*, vol. 15, no. 11, pp. 1119–1125, Nov. 1994.

[43] L. Breiman, *Probability*, ser. Classics in applied mathematics, vol. 7. Philadelphia, PA: SIAM, 1992.

[44] E. T. Jaynes, "Information theory and statistical mechanics," *Phys. Rev.*, vol. 106, no. 4, pp. 620–630, May 1957.

[45] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1988.

[46] P. J. Rousseeuw and K. V. Driessen, "A fast algorithm for the minimum covariance determinant estimator," *Technometrics*, vol. 41, no. 3, pp. 212–223, Aug. 1999.

[47] S. Verboven and M. Hubert, "LIBRA: a MATLAB library for robust analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 75, no. 2, pp. 127–136, Feb. 2005.

[48] J. Preston, "Automated acoustic seabed classification of multibeam images of Stanton Banks," *Applied Acoustics*, vol. 70, no. 10, pp. 1277–1287, Oct. 2009.

[49] T. B. Reed and D. Hussong, "Digital image processing techniques for enhancement and classification of SeaMARC II side scan sonar imagery," *J. Geophys. Res.*, vol. 94, no. B6, pp. 7469–7490, Jun. 1989.

[50] P. Blondel and O. Gómez Sichi, "Textural analyses of multibeam sonar imagery from Stanton Banks, Northern Ireland continental shelf," *Applied Acoustics*, vol. 70, no. 10, pp. 1288–1297, Oct. 2009.

[51] V. A. I. Huvenne, P. Blondel, and J.-P. Henriet, "Textural analyses of sidescan sonar imagery from two mound provinces in the Porcupine Seabight," *Marine Geology*, vol. 189, no. 3-4, pp. 323–341, Sep. 2002.

[52] F. Albregtsen. (2008, Nov.) Statistical Texture Measures Computed from Gray Level Cooccurence Matrices. Dept. Informatics, Univ. Oslo. [Online]. Available: http://www.uio.no/studier/emner/matnat/ifi/INF4300/h08/undervisningsmateriale/glcm.pdf

[53] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 4, pp. 610–621, Nov. 1973.

Dette er en postprint-versjon / This is a postprint version.
DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133

[54] L. Fonseca, C. Brown, B. Calder, L. Mayer, and Y. Rzhanov, "Angular range analysis of acoustic themes from Stanton Banks Ireland: A link between visual interpretation and multibeam echosounder angular signatures," *Applied Acoustics*, vol. 70, no. 10, pp. 1298–1304, Oct. 2009.

[55] K. D. Ward, "Compound representation of high resolution sea clutter," *Electron. Lett.*, vol. 17, no. 16, pp. 561–563, Aug. 1981.

[56] NIST/SEMATECH e-Handbook of Statistical Methods. [Last accessed: Oct. 15 2012]. National Institute of Standards and Technology. [Online]. Available: http://www.itl.nist.gov/div898/handbook/

[57] P. Wessel and W. H. F. Smith, "New, improved version of Generic Mapping Tools released," *EOS Trans. Amer. Geophys. U.*, vol. 79, no. 47, p. 579, 1998.

DOI til publisert versjon / DOI to published version: 10.1109/JOE.2013.2281133