

Seabed Classification using Physics-based Modeling and Machine Learning

Christina Frederick¹, Soledad Villar², and Zoi-Heleni Michalopoulou¹

¹Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102, U.S.A.

²Center for Data Science and Courant Institute of Mathematical Sciences, New York University, New York, NY 10011

July 25, 2020

July 25, 2020

Abstract

In this work, model-based methods are employed along with machine learning techniques to classify sediments in oceanic environments based on the geoacoustic properties of a two-layer seabed. Two different scenarios are investigated. First, a simple low-frequency case is set up, where the acoustic field is modeled with normal modes. Four different hypotheses are made for seafloor sediment possibilities and these are explored using both various machine learning techniques and a simple matched-field approach. For most noise levels, the latter has an inferior performance to the machine learning methods. Second, the high-frequency model of the scattering from a rough, two-layer seafloor is considered. Again, four different sediment possibilities are classified with machine learning. For higher accuracy, 1D Convolutional Neural Networks are employed. In both cases we see that the machine learning methods, both in simple and more complex formulations, lead to effective sediment characterization. Our results assess the robustness to noise and model misspecification of different classifiers.

1 Introduction

Sonar classification of the ocean floor relies on the quality of the acoustic images and the data-processing tools used for interpretation. The pixel resolution in sonar images depends on a number of factors, including the instrument range, geometry, and operating frequency. Due to the multiple sophisticated scattering processes, there is no simple way to interpret each pixel of a sonar image in terms of seafloor parameters such as sediment type, roughness, and layer structure. Well-known approaches such as matched-field processing [1, 2], image-processing based techniques [3, 4, 5], and recently, machine learning algorithms, are successful in known environments for which ground-truth datasets are available [6, 7, 8].

Knowledge of the propagation medium is of paramount importance in source detection, localization, and identification in the ocean. The field of geoacoustic inversion deals with modeling and techniques for obtaining knowledge about the ocean sediments with which the propagating sound

has interacted to facilitate the tasks mentioned above. Extensive research and significant advances have been made through the years in estimating sediment parameters using the full field or select features.

Matched-field inversion (MFI) is a popular approach for inverting for (estimating) geoacoustic properties using measurements of the acoustic field at a number of receiving phones. It stems from matched field processing (MFP) [13], originally designed for source localization, which was extended to invert for environmental parameters with the first effort presented in [14]. Estimates are obtained by comparing measured acoustic fields to replica-predicted fields computed with sound propagation models for numerous values of the unknown parameters. The values that provide the best match between replicas and data according to a preselected criterion are considered to be the desired estimates. MFI initially required significant computation entailing searches in multi-dimensional domains. Global optimization approaches have facilitated the inversion making the process less onerous [15, 16, 17, 18, 19, 20].

Backscattering strength data have also been used for sediment characterization [21, 22, 23, 24, 25, 26] with a variety of methods using backscatter models and statistical techniques. These approaches facilitated inversion both for seafloor roughness and sediment geoacoustic properties.

In the late '90s and early 2000s, feed-forward neural networks gained attention in the field of ocean acoustics because of the ability to overcome issues faced by traditional methods, such as error estimation and time-consuming global or local searches in the parameter space [27, 28, 29]. In these problem settings, neural networks provided efficient, nonlinear approximators of inverse functions.

One drawback of neural networks (and many modern deep learning approaches) is the large amount of training data needed to ensure accurate, reliable performance. Some of the earlier works overcome this using synthetic training data and real test data [30]. Over the last 15 years, feature-based machine learning techniques have also been developed for problems such as target detection and recognition, source localization, and seabed classification [31, 32, 33]. While the corresponding processors can be fully trained with smaller data sets, they require a manual, often tedious, feature extraction process.

Convolutional neural networks (CNNs) offer an alternative approach in which features are instead learned by the network. The neural network can be viewed as a composition of functions, called layers, and the training process consists of determining weights in each layer using batches of labeled data and stochastic optimization. Different layers are used for different purposes; for example, convolutional layers contain filters that are convolved with small patches of the previous layer to capture local features, and max pooling layers downsample the input by taking the maximum over a patch. Layers in a CNN are connected via an entry-wise nonlinear activation function. The most common activation function is $\text{ReLU}(x) = \max(x, 0)$. Modern CNN architectures can contain many layers and connections between them that form acyclic graphs. The complexity and depth of the architecture can capture features and interdependencies of high-dimensional data on a wide range of scales.

The use of CNNs in acoustics is now a very active and rich area of research [34]. Although some studies can be performed using smaller training data sets, [35, 36, 37, 38], most deep learning approaches must develop techniques for accomodating the lack of real sonar data. This can be done by, for example, using a pretrained CNN (trained using optical images) [40, 41], applying image processing techniques for data augmentation [31, 42, 43, 44], or training with synthetic data generated by a physics-based forward solver [30, 45, 46, 39]. Large sonar images can also be

decomposed into many “patches” [37, 47, 48] that form training datasets. Our modeling, described next, is based on a combination of these patch-based techniques for data and physical modeling.

To increase the performance and applicability of current machine learning algorithms, the main challenge is overcoming the lack of large training datasets. In this paper, we demonstrate the potential of physics-based forward modeling to generate training data for seafloor classification problems. The advantages of physics-based modeling for these problems include the ability to accurately label data and fully explore the possible seafloor parameter space. This enhances the training of machine learning models, as well as the evaluation of the reliability and validity of the outcomes.

There are two main contributions of our work aimed at better understanding the limits of modern techniques applied to scenarios with complex physics and limited observational data. The first is physics-based training, in which training datasets are designed using numerical simulations of representative segments of the seafloor domain. At lower frequencies, these template domains are characterized by sound speed, attenuation, and layer thickness. At higher frequencies, finite element modeling (FEM) can incorporate additional small-scale features such as the roughness of the water-sediment interface. The second contribution is a set of machine learning experiments for classifying the top sediment in a two-layer seafloor. In the low-frequency case, these results are compared with known techniques and indicate that machine learning tools provide higher classification accuracy. At high frequencies, deep learning offers a significant improvement in performance over shallow machine learning tools. To our knowledge, this is the first work that attempts material classification from synthetic backscatter generated with FEM-based simulations.

2 Inverse wave problems in ocean acoustics

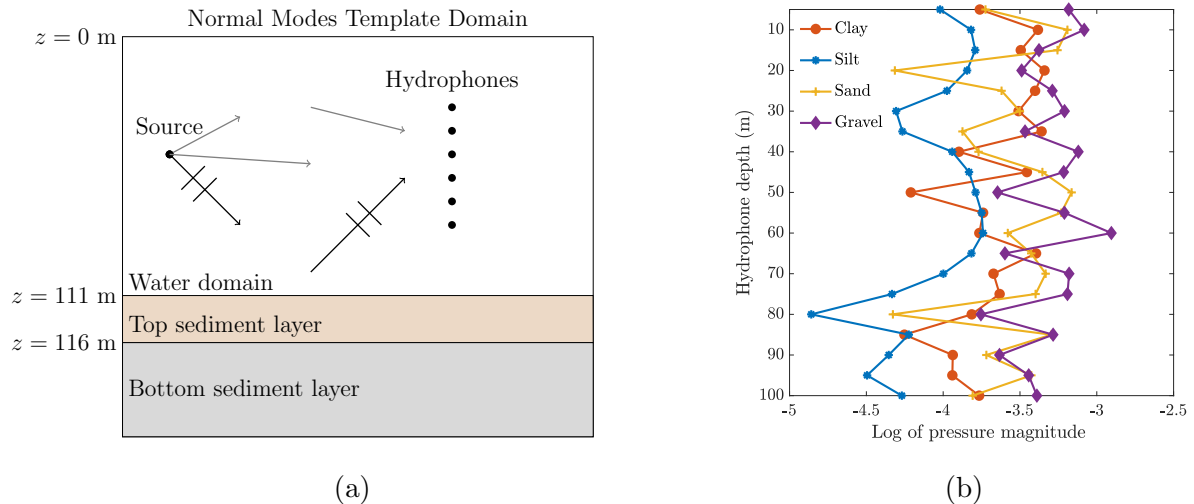


Figure 1: (a) Sound propagation in an oceanic waveguide. (b) Simulated pressure field measurements for four top sediments (pressure is in μPa , simulations are produced with KRAKEN[50]).

For the present work, we model the ocean as a range-independent waveguide with homogeneous water and two sediment layers (see Figures 1 and 3). For ocean bottoms with low shear speeds, a

fluid assumption is generally appropriate [10, 11, 12]. Therefore we treat the sediment layer as a fluid, and model time-harmonic acoustic pressure waves at range x and depth z as solutions $p(x, z)$ to the Helmholtz equation,

$$\rho \nabla \cdot \left(\frac{1}{\rho} \nabla p \right) + \frac{\omega^2}{c^2} p + i\alpha \omega p = s, \quad (1)$$

where ω represents the angular frequency and the spatially varying functions ρ , c , and α , represent the density, sound speed, and attenuation, respectively, of the medium. In seawater, attenuation is negligible at low and intermediate frequencies and it is reasonable to set $\alpha = 0$. The source s and boundary conditions depend on factors, such as incident angle and distance to the seafloor, that are determined by the instrument and environment. Coupling conditions on the water-sediment and sediment-sediment interfaces enforce continuity of the pressure and normal displacement along the fluid-sediment and sediment-sediment interfaces.

The mathematical formulation of the inverse problem is to determine a vector of true parameters $m^* \in \mathcal{M}$ using a finite set of observables $\mathbf{Y} \in \mathbb{C}^n$, where $\mathbb{C}$ is the set of complex numbers. Here, $\mathcal{M}$ is a finite-dimensional space of unknown parameters, for example, source location, bathymetry, water column sound speed, and geoacoustic and geometric properties of the sediment layer (density, compressional and shear wave speeds, etc.). In sonar imaging, the observational data is often a set of pressure or backscatter measurements $\mathbf{Y} = (p(x_k, z_k))_{k=1}^n$ recorded at n receiver locations (x_k, z_k) . The forward operator $\mathcal{F} : \mathcal{M} \rightarrow \mathbb{C}^n$ describes the wave propagation process relating the observables to the model parameters.

The inverse problem of determining m^* given $\mathbf{Y} \approx \mathcal{F}(m^*)$ is usually difficult and ill-posed; a number of different seafloor environments can produce similar measurements, and that similarity can be further complicated in the presence of noise. While full waveform inversion (FWI), a standard tool in seismic imaging, achieves success by matching the data with simulations of the forward model, there is no universal technique for solving (1) at frequencies of interest in sonar imaging. Factors such as the source-receiver geometry, frequency and bandwidth, and boundary conditions must be taken into account in formulating a solution technique. Often, a combination of analytic and numerical methods are implemented based on the environmental complexity, desired resolution, and availability of labeled ground truth data. Here we focus on two solution techniques, normal mode propagation in simple geometries and finite element modeling of the scattering process in a more complicated environment having a rough water-sediment interface. In this work, normal mode propagation was used to model the waveguide for acoustic wavelengths on the order of meters, and finite element modeling was used for acoustic wavelengths on the order of centimeters.

3 Seafloor characterization with pressure fields at vertical line arrays

3.1 Problem setup

We consider the scenario of Figure 1, where a source transmits a continuous wave (CW) signal at frequency $\omega = 2\pi f_0$. In this setting ω is the normal modes operating frequency and $f_0 = 400$ Hz. A point source located at range $x_s = 0$ and depth z_s determines the source term $s(x, z) = -\frac{\delta(x)\delta(z-z_s)}{2\pi x}$. The pressure field at a depth z and a distance x from the source can be seen as the solution of the Helmholtz equation (1). Here, we assume that the density is constant

within the water layer and each sediment layer. Also, the environment is assumed to be range independent. The solution operator for the Helmholtz equation corresponds to $\mathcal{F}$ of Section 2.

Assuming that source and receiver location, bottom depth, water column sound speed, and sediment thickness are known, we can invert for sediment sound speed c and attenuation α via MFI (for the experiments that we consider the field is not sensitive to density, for which we cannot reliably invert). An inner product between the normalized pressure field (solution of Eq. 1) computed for different c and α values and the received acoustic field can be calculated. Those values that maximize the inner product are the parameter estimates.

In this work, we are interested in classification rather than parameter estimation. As will be discussed later, MFI can still be employed towards this task, but, here, we are mostly interested in investigating the potential of machine learning in sediment classification. Recently, a sensitivity analysis was performed on parameters in a two-layer seafloor, indicating the promise of these methods [49]. We investigate a low-frequency case with the sound being transmitted by a source and received at a vertical line array with 20 hydrophones (see Figure 1). The phone spacing is 5 m and the source frequency is 400 Hz. The source-array distance is 10 km and the source depth is 50 m. The water column sound speed profile $c(z)$ is a typical shallow water downward refracting profile and the water depth is 111 m. One isovelocity sediment layer is assumed with varying sound speed c and attenuation α , representing different sediment types (clay, silt, sand, gravel) over a chalk halfspace. The sediment thickness is $\tau = 5$ m.

3.2 The data

Using normal modes [50] we synthesize the pressure field at the VLA for the four sediment types; each field has 20 components corresponding to the 20 VLA receivers; these are equispaced with the first one located at a depth of 5 m. Values of 1500, 1575, 1650, and 1800 m/s are the assumed sound speeds for clay, silt, sand, and gravel, respectively [51]. Values of 0.2, 1, 0.8, and 0.6 dB/ λ , where λ is the acoustic wavelength, are the corresponding attenuations. These sediment sound speed-attenuation values are listed in the first column of Table 2. One-thousand noisy field realizations are then synthesized for each sediment type by adding zero-mean complex Gaussian noise. The generated noisy fields are used to train a set of classifiers. The noise-free fields for each class are shown in Figure 1(b). Subsequently, we perturb the sediment sound speed and attenuation values. The new c and α values still represent the same sediment types but now with variation. We create environments for ten sets for each sound speed-attenuation pair (listed in Table 1) and one-thousand noisy realizations for each case for a total of ten-thousand pressure fields. Then, the responses are the predictions $\mathcal{F}(m)$ and observables $\mathbf{Y}$ of Section 2 that form the training and test sets.

Data generation for both training and testing sets was repeated for a number of Signal-to-Noise-Ratios (SNRs) to highlight the effects of noise dependence on the relative classifier performance.

3.3 Sediment Classification

The goal of our experiments is to determine whether sediment class can be determined correctly when the classifiers are tested on data generated for similar, but not identical, environments to those used for the generation of the training data.

Before we implemented machine learning techniques, we constructed a classifier based on MFI; that served well for the performance evaluation of the proposed techniques. The MFI/MFP classifier relied on the computation of a simple inner product between the data to be classified and normalized

Figure 2: **(a)** Classification performance of different trained machine learning classifiers on test data. The y -axis is the test accuracy, the ratio of correct predictions and total predictions. **(b)** and **(c)** Confusion matrices indicate how many instances of the true class (row) were classified as a predicted class (column). The diagonal cells correspond to observations that are correctly classified. The off-diagonal cells correspond to incorrectly classified observations. The bottom rows show the percentages of all the signals predicted to belong to each class that are correctly classified (precision). The rightmost columns in each plot show the percentages of all the signals belonging to each class that are correctly classified (recall). The confusion matrices are given for sediment classification with **(b)** matched field processing, and **(c)** logistic regression, on noisy KRAKEN data (SNR = 18 dB). The results in **(d)**, **(e)** and **(f)**, reflect the performance of the classifiers on a test set with variable top layer thickness.

replica fields calculated for the nominal values of sediment properties. The formula for MFP stems from the calculation of the likelihood function of the unknown parameters including noise variance and source spectrum red [9]. black The replica fields for the four classes are the measurements shown in Figure 1. Given model parameters $m = (c, \alpha)$, noisy pressure measurements at $n = 20$ receiver locations (x_k, z_k) are modeled by $\mathcal{F}(m) = \mathbf{p}(m) = (p(x_k, z_k))_{k=1}^n$, where p is the solution to (1). The MFP classifier predicts the parameters $\hat{m}$ of an observable signal $\mathbf{Y}$ by maximizing the inner product:

$$\hat{m} = \operatorname{argmax}_{m \in \mathcal{M}} \frac{|\langle \mathbf{p}(m), \mathbf{Y} \rangle|^2}{\|\mathbf{p}(m)\|^2}, \quad (2)$$

where the search space consists of material parameters of the four different classes, $\mathcal{M} = \{(c_j, \alpha_j)\}_{j=1}^4$. The MFP classifier required no training.

The following baseline classifiers were implemented in `scikit-learn` [52]: grouping methods, such as k -nearest neighbors (KNN) and Nearest Centroid (NC), Support Vector Classifiers that find the best separating hyperplane (SVC-1), ensemble methods such as Decision Tree Classifiers (DT), feed-forward neural nets, in particular, multi-layer perceptron (MLP), and Logistic Regression (LR). The hyperparameters of each classifier were determined using a randomized grid search over a range of possible parameters. Five-fold cross-validation was used for model selection.

The training data set consisted of the 4,000 labeled noisy pressure field measurements for the nominal values of the four sediment classes. We tested the methods on ten sets of 1,000 realizations for each sediment class, each set for perturbed pairs of sediment parameters. That is, no testing signal was generated for the exact c and α values used to generate the training set vectors. Figure 2(a) shows the overall accuracy of different classifiers. For high noise levels (low SNR), MFP seems to perform well but LR classification exhibits the best performance. While linear regression models fit data with a linear function, logistic regression models instead use the more flexible sigmoid function that is less sensitive to outliers. Unlike linear classifiers, LR models can produce nonlinear decision boundaries, which may improve the ability to discriminate between points that are close together but belong to different classes.

As the SNR increases, there was a marked difference between MFP and other classifiers, which all outperform MFP. Figure 2(b) and (c) show the confusion matrices for MFP and LR for an SNR of 18 dB. We observed that MFP mislabeled signals from the silt class with gravel, whereas LR incorrectly classified about half of the signals from the clay class as silt, but had superior performance for the other materials. An interesting inference can be made from the LR classification results. Signals classified as clay, sand, or gravel were likely to be correctly characterized. In the

case of silt, however, there was a possibility of misclassification. Identifying a feature that further distinguishes clay from silt measurements could potentially resolve the ambiguity.

Following this experiment, we tested the classifiers on more realistic and challenging data. Although we kept the training dataset the same, the test dataset was generated using perturbations of thickness in addition to sound speed and attenuation; this is something faced in practice where prior but not exact knowledge on the layer thickness is given. Again, we generated 40,000 test patterns corresponding to ten perturbations of the sound speed, attenuation, and thickness. The thicknesses were drawn from a normal distribution with a mean of 5 m and a standard deviation of 2. The sound speed and attenuation values are those in Table 2. The accuracy of the different classifiers vs. SNR is shown in Fig. 2(d) and confusion matrices for MFP and LR are illustrated in (e) and (f). The overall behavior was similar to the experiment with fixed thickness, for all classifiers, with slightly lower accuracy.

4 Backscatter

If the spatial scale of the seafloor roughness is much smaller than the acoustic wavelength, a source ping smoothly reflects from the seafloor, and analytical methods described in the previous section are preferable because of the accuracy and low computational cost. On the other hand, when the spatial scale of seafloor heterogeneities is similar to the acoustic wavelength, these methods fail to approximate the measured backscatter, and more accurate modeling is needed.

Finding exact solutions to (1) can be very difficult, even for homogeneous environments with simple geometries, and at higher frequencies, the problem becomes even more complex, with features such as roughness complicating the task of modeling and classification. In the case of rough, layered sediments, numerical methods such as the finite element method (FEM) provide an attractive option. The FEM discretization of the governing acoustic wave equations in underwater domains enables accurate modeling of the seafloor environment [53, 54, 55, 56, 57, 58, 59, 60]. Not limited by constraints on the degree of surface roughness or range dependence, FEM offers flexibility and can incorporate physics generated from realistic scenarios. In computational ocean acoustics applications, FEM is usually used for providing benchmark solutions, short-range scattering, or as a part of a hybrid approach [61, 62, 63, 64] in which the scattering is approximated using a FEM discretization of a small region near the seafloor and the reverberant field is modeled using traditional, more efficient techniques.

The primary disadvantage of FEM and other numerical methods for solving high-frequency Helmholtz equations is the high computational cost of approximating $O(\lambda^{-1})$ unknowns in each spatial dimension, where λ is the acoustic wavelength. Computing solutions for large spatial domains in the high-frequency regime, in which each forward solve may need to resolve up to billions of unknowns, is a notoriously hard problem and a topic of intense interest in the scientific computing community [65]. In practice, the balance between accuracy and computational cost often results in model simplifications that limit the recovery of small-scale features in a range of complex underwater environments. A main challenge in ocean acoustics is to find ways to improve existing models in a way that is both physically justified and computationally efficient.

To address the scalability of the high-frequency models to large seafloor domains, we divide the region near the seafloor into segments and perform local simulations on these subdomains. The resulting simulations produce acoustic templates from small, representative seafloor environments that can be assumed to have homogeneous properties. Our rationale for this model reduction is

based on the framework of multiscale modeling for partial differential equations, in which a low-cost *large-scale* solver of an effective model is coupled with a *fine-scale* solver that resolves detail. The range of validity for these methods is problem-dependent and our justification in the present work is based on physical assumptions of scale-separation and spatial stationarity in the problem.

The forward process can be summarized in terms of three main stages: (I) the incoming wave from the source to the seafloor; (II) scattering and reflection from the seafloor; and (III) traveling waves recorded by receivers (See Figure 3). Stages (I) and (III) are relatively easy to model for a wide range of source frequencies. It should be noted that variable wave propagation velocity in the water can easily be handled by existing tools, e.g. geometrical optics. The critical interactions with the seafloor that occur in stage (II) deserve accurate modeling at high frequencies to resolve significant contributions from scattering effects near the ocean floor. In this case, fine-scale features such as seafloor roughness must be incorporated in modeling.

Given the availability of tools for modeling wave propagation at lower frequencies, we limit the present discussion to modeling the high-frequency scattering that occurs in stage (II) and testing the performance of these models in sediment classification. Though it is out of the scope of the present work, it is an important problem to further analyze multiscale couplings between the different stages in the forward process and find ways to systematically build hybrid algorithms that adapt to more complicated physical processes occurring in the ocean.

Below we will discuss how we approach the backscatter modeling for high-frequency scattering and sediment characterization.

4.1 Problem setup

We consider models for acoustic backscatter on template domains shown in Figure 3(a). When the acoustic wave generated by a source located near the surface reaches this template domain, it can be modeled as an incident plane wave traveling in the direction given by the vector $k_\theta = \frac{\omega}{c_w}(\sin \theta, -\cos \theta)$, where θ is the incident angle with respect to the z axis,

$$p_{in}(x, z) = e^{-ik_\theta \cdot (x, z)}. \quad (3)$$

The interaction of these waves with the seafloor produces scattered waves in multiple directions due to objects on the seafloor, the roughness of the water-sediment interface caused by sand ripples, as well as heterogeneities in the sediment, including interactions between multiple sediment layers. The total pressure field p solves the homogeneous equation (1) with $s = 0$ and can be expressed in the water domain as the sum of the incident field and the scattered wavefield, $p = p_{in} + p_{scatt}$.

Sufficiently far from caustics, p_{scatt} can be approximated by the superposition

$$p_{scatt}(x, z) \simeq \sum_{\beta} P_{\beta}(x, z) e^{ik_{\beta} \cdot (x, z)}, \quad (4)$$

where $P_{\beta}(x, z)$ is the complex amplitude of the outgoing wave traveling in the direction β and is assumed to be locally smooth and independent of wavelength. For co-located sources and receivers, the backscattered direction is $\tilde{\theta} = \pi/2 + \theta$, and we model the received signal as backscatter measurements,

$$\mathbf{Y} = (|P_{\tilde{\theta}}(x_k, z_*)|)_{k=1}^n. \quad (5)$$

where (x_k, z_0) is a set of observation points located at a fixed depth $z \equiv z_0$ above the seafloor.

For unlayered seabeds, there has been significant progress in quantifying the dominant contributions to backscatter at high frequencies [66, 67, 68, 69, 70]. Many approaches rely on cancellation effects from neighboring waves to simplify the model, and the situation is even more complicated when there are multiple sediment layers within the acoustic penetration range. Numerical approaches for approximating far-field scattering from targets using FEM include boundary-element methods [56, 71] and techniques for post-processing solutions using the Helmholtz-Kirchhoff (HK) boundary integral formulation of the scattering problem [72]. Another technique called numerical microlocal analysis (NMLA) [73, 74] aims to isolate direction angles and amplitudes of waves crossing an observation location (x_0, z_0) without relying on knowledge of the solution on the entire boundary of a scatterer. The main idea is that if the acoustic wavelength $\lambda = 2\pi/k$ contains variations on a scale sufficiently small with respect to the seafloor geometry, the solution $p(x, z)$ to (1) behaves like a finite superposition of plane waves (4). Then, discrete measurements of the solution on a observation circle, $p(x_0 + r \cos(t_j), z_0 + r \sin(t_j))$, can be expressed as a 2D Jacobi-Anger expansion that contains the amplitudes and angles of interest. By filtering the vector of observations using the fast Fourier transform, the method recovers the dominant directions and amplitudes of scattered waves. Recently, research and implementations for high-frequency Helmholtz equations further improve the accuracy and stability of the method [75, 76].

The plots in Figure 3(b) show the acoustic response approximated by FEM and NMLA at $z = \bar{f} - 1.5$ m above the seafloor (where $\bar{f}$ is the average of $f(x)$) from a rough, two-layer seafloor model. Figure 3(a) shows approximate backscatter signals $|P_{\hat{\theta}}(x, z_0)|$ in (5) modeled with FEM and NMLA on the domain $0 < x < 2, -2 < z < 2$ for environments with varying environmental parameters $m = (c, \rho, \tau)$ and a fixed realization of the water-sediment interface. The signals corresponding to the environment having an upper sediment layer with the largest thickness, $\tau = 1$ m, are visually well-separated owing to the high attenuation of the signal in the sediment. For $\tau = .25$ m, however, the plots of the signals reflect the effects of the complex interactions that occur when the acoustic energy penetrates to the bottom sediment layer. These plots indicate the difficulty and ill-posedness of the inverse problem in the multi-layer setting, and why machine learning, described below, can play a valuable role in discriminating these signals based on material type or other classes.

4.2 The data

Backscatter signals were generated for a fixed source frequency $\omega = 15$ kHz, incident angle $\theta = \pi/12$ radians, and sound attenuation $\alpha = 0.5$ dB per 1 meter per 1 kHz. In the current study we chose not to distinguish attenuation based on material type. For our current purposes, we chose to isolate the small-scale effects of the seafloor by assuming that $f = \bar{f} + \tilde{f}$, where $\tilde{f}$ is a mean-zero function with prescribed root mean square (RMS) height and correlation length. A random selection of the available environmental parameters, summarized in Table 2, was used to generate the balanced datasets of backscatter signals (5). We set the sound speed and density of water to be $c_w = 1500$ m/s and $\rho_w = 1000$ kg/m³ respectively.

Full wave solutions were generated using the COMSOL Multiphysics® Acoustics Module [77] for the solution of the Helmholtz equation on a 2 m $\times$ 2 m square domain using a P^2 - finite element method discretization, which translates to about 10^5 degrees of freedom per template. For more details on sonar forward modeling using COMSOL Multiphysics® see [58, 53]. Then, NMLA implemented in MATLAB, is applied to approximate backscatter at $n = 1024$ locations (x_k, z_0) , representing a uniformly spaced horizontal discretization of the water domain and a depth 1 meter

above the seafloor $z_0 = \bar{f} - 1$.

4.3 Sediment classification

The top part of Table 3 shows the accuracy of different machine learning classifiers implemented in `scikit-learn` as described in 3.3, a Support Vector Classifier with radial basis function kernel and a Random Forest Classifier.

These baseline classifiers were trained on $N = 20,000$ signals taken from the training data set, 20% of which are set aside for validation. The performance of the classifiers was evaluated on 5,000 signals from each of the four different test environments with different realizations of the water-sediment interface f . Results from these experiments, shown in Table 3 indicate that the baseline machine learning classifiers did not generalize to different test sets generated from environments with slightly perturbed parameters. Based on our findings, described next, the deeper neural networks exhibited a better classification performance on the validation data and test environments.

4.3.1 Implementation of CNNs

Initially, we constructed two smaller convolutional neural networks, ‘CNN-3’ and ‘CNN-4’, having three and four convolutional layers, respectively. In both cases, each convolutional layer is followed by a batch normalization layer, ReLU activation functions, and a max pooling layer, and the filter sizes vary from 16×1 to 256×1 . We also considered the following well-known deeper CNNs by adapting them to the 1D setting: AlexNet [78] (8 convolutional layers), GoogleNet [79] (22 convolutional layers), ResNet50 [80] (50 convolutional layers), and VGG-19 [81] (19 convolutional layers). MATLAB code and the trained models are publicly available¹.

We use the Deep Learning toolbox in MATLAB to train CNNs using $N = 80,000$ signals taken from the training data set, 20% of which are set aside for hold-out validation. The performance is evaluated using 5,000 signals from each of the four different test sets. The labels of the test signals correspond to the closest match, in terms of absolute distance, between the sound speed of the top layer and the four sound speed classes used in the training dataset.

Learnable parameters in each CNN are determined using the ADAM method for stochastic optimization [83]. Training options for the ADAM optimizer were chosen based on trial and error. We chose a mini-batch size of 500; the learning rate schedule varied between $1e-4$ and $1e-3$, with a 50% drop in the learning rate scheduled periodically (every 10-50 epochs) during the training; the training data are shuffled before every epoch; the validation data are shuffled before each network validation; the gradient threshold was set to 1; and the L_2 regularization factor was set to $1e-4$. The termination criteria were determined to ensure a decreasing validation loss. For the deeper CNNs, the training was performed on one NVIDIA Tesla P100 16GB GPU or two GPUs in parallel, as needed. Figure 4 shows plots of the cross-entropy training loss and validation loss for six different CNN architectures. The cross-entropy loss function is defined as $\ell(P, Q) = -\sum_{t=1}^M \sum_{x \in X} P_t(x) \log Q_t(x)$ where M is the number of classes (4 in our case), X is the set of observations (either the training or validation set), $P_t(x)$ indicates whether the true label of x is t , and $Q_t(x)$ is the predicted probability of x belonging to class t according to the classifier.

The training times and test accuracy (discussed next) for each classifier are listed in Table 3.

¹<https://github.com/cf87/Seabed-Classification-2020>

4.3.2 Results

Figures 5(a) and (b) contain plots of the mean test accuracy of different CNNs used to classify data with varying SNR. The mean is taken over four realizations of noise added to each of the 5,000 signals, resulting in 20,000 predicted labels. Out of the classifiers that were tested, the CNNs that showed the most robustness to noise were AlexNet-1D, GoogleNet-1D, and to a lesser extent, CNN-3. This performance was also seen for the other test environments not shown.

One criticism that deep learning techniques receive is the lack of interpretability, especially when compared with classifiers like SVM and KNN. In physical scenarios, such as seabed classification, predictive models that rely on the governing equations offer the ability to test many different environments. This advantage can lead to gained insight on the capability to generalize. Classification results for the four test environments in our experiments shed light on the sensitivity of the machine learning outcomes to physical parameters. Figure 6(a) and (b) contains confusion matrices for the best classifier, AlexNet-1D, corresponding to noisy test data with an SNR of 20. There was significant misclassification among all classifiers of signals in Test 3 corresponding to responses from environments with clay and silt top layer. This indicates that the layer thickness can have a significant influence on misclassification, which is consistent with physical intuition about the interaction between thin sediment layers and indicates the need to resolve the layer thickness with high accuracy to obtain better classification results. The classifier performance in Test 4 shows improved accuracy.

Although GoogleNet-1D demonstrated similar accuracy and robustness to noisy data, the training times, listed in the bottom part of Table 3, indicate the superiority of AlexNet-1D (under 11 minutes) over GoogleNet-1D (over 7 hours). Although the smaller, custom CNNs have low training times and performed well on test sets with low SNR, they did not show robustness to noise. Even at low SNR, AlexNet-1D outperformed the best baseline machine learning classifiers, however, in the case of noisy data, the baseline classifiers performed much better overall than the custom CNNs, ResNet50-1D, and VGG19-1D. Interestingly, SVC performed the best on the Test 3 dataset corresponding to an environment with perturbed layer thickness, which was the most challenging test case.

An interesting outcome of these numerical experiments is that some of the deep learning classifiers that perform well on the validation set are vulnerable to noise and small perturbations in the geoacoustic properties of the test environments. Both AlexNet-1D and GoogleNet-1D demonstrated good generalization behavior whereas ResNet-1D, VGG19-1D, and CNN-4 did not generalize at all. The poorest performing classifier makes predictions that have about the same accuracy as a random guess. Understanding in what context deep learning classifiers are good at generalizing is currently an open question within machine learning (see for instance [84]).

4.4 Further Discussion

One way to visualize the behavior of deep neural networks is by plotting the activation patterns of different layers for a given input. Figure 7(a) contains plots of the activations of the first convolutional layer in the trained AlexNet-1D for four different input test signals. The pixel intensities signify which channels are activated. These plots give a sense of which filters in each layer are activated by different inputs, for example, the filters in the 50-100 range are activated by input signals from the gravel class much more than the other three materials. Furthermore, these plots highlight the ease of separating signals with strong activations in this layer (sand and gravel environments)

from signals with weaker activations (silt and clay environments).

Viewing activations in this way across many signals can be challenging since the set of activations in each layer corresponds to a high-dimensional vector. One way to approach the task is the t-Distributed Stochastic Neighbor Embedding (t-SNE), [85], which constructs a probability distribution defined on pairs of activation vectors, where the probability of sampling two vectors increases as the distance between the vectors decreases. Typically, this distribution is modeled with a Gaussian kernel. Then, gradient descent is used to find a map between the high-dimensional distribution and a Student's t -distribution, defined on a low-dimensional space, that minimizes the Kullback–Leibler divergence. The Student's t -distribution arises from estimating the mean of a normally distributed population; it is symmetric and bell-shaped but it has heavier tails than a Gaussian distribution. We refer the reader to the t-SNE paper [85] for a formal definition and a justification of the visualization technique. The t-SNE plots in Figure 7(b) correspond to the layer activations for all of the signals in the Test 4 dataset corresponding to the first max pooling layer, final convolutional layer, and the softmax layer that produces a probability distribution on the set of possible classes. This visualization suggests that in the initial layer and fifth convolutional layer, the activations corresponding to different class boundaries overlap, however, in the final fully connected layer, the class boundaries are more separated. This indicates that the full architecture of AlexNet-1D is needed for increased accuracy.

5 Conclusions

We employed a model-based approach for designing training and test datasets of acoustic templates for capturing the relevant physics of representative patches of the seafloor. At low frequencies, this is accomplished with normal mode propagation, and at higher frequencies, local modeling on smaller computational domains enables fast, parallelizable simulations. An underlying assumption is the spatial stationarity of the seafloor, which is reasonable in situations where roughness statistics are similar over an area larger than the ensonified area. In this study we performed sediment classification in a two-layer seafloor, varying both geoacoustic parameters (sound speed, density) and geometric parameters (roughness, thickness) in the training and test data.

In low-frequency models, standard machine learning classifiers, such as logistic regression models and support vector machines, outperformed traditional matched-field processing techniques, especially when the test data had a low signal-to-noise ratio. Confusion matrices summarizing the classification results indicate that signals from certain classes have a higher likelihood of misclassification, namely silt and clay. On the other hand, predictions of signals from gravel and sand classes are more likely to be correct. For backscatter data, the standard machine learning classifiers demonstrated poor accuracy and did not generalize well to test environments with added noise and perturbed parameters. Some of the deep learning classifiers, namely AlexNet and GoogleNet, adapted to 1D signals, were more costly to train but demonstrated higher accuracy and improved generalization. These results also indicate that the layer thickness can have a significant influence on misclassification rates. Further investigation must place an emphasis on resolving the layer thickness with high accuracy.

Our results indicate the promise of machine learning and deep learning for the difficult problem of geoacoustic classification. The results from our simulations highlight the need to test models for a broad spectrum of environments to ensure generalization. Producing a well-performing deep learning model requires thorough experimental design. Several new directions can be explored with

this framework, for example, finding elastic properties of sediments and incorporating the influence of the material type on the statistical properties of the seafloor roughness.

Acknowledgements

The authors are grateful to Kyunghyun Cho (CDS, NYU) for valuable discussions. The research of CF is supported by NSF DMS 1720306. The research of ZHM is supported by ONR N000142012029 and N000141812125. SV is partly supported by NSF DMS 1913134, EOARD FA9550-18-1-7007 and the Simons Algorithms and Geometry (A&G) Think Tank. The project started while SV was with the NYU Center for Data Science and Courant Institute of Mathematical Sciences.

References

- [1] A. B. Baggeroer and W. A. Kuperman, “Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem,” *J. Acoust. Soc. Am.* **83**(2), 571–587 (1988).
- [2] A. Tolstoy, “Applications of matched-field processing to inverse problems in underwater acoustics,” *Inverse Problems* **16**(6), 1655–1666 (2000).
- [3] Y. Rzhanov, L. Fonseca, and L. Mayer, “Construction of seafloor thematic maps from multi-beam acoustic backscatter angular response data,” *Computers and Geosciences* **41**, 181–187 (2012).
- [4] S. Reed, I. Tena Ruiz, C. Capus, and Y. Petillot, “The fusion of large scale classified side-scan sonar image mosaics,” *IEEE Transactions on Image Processing* **15**(7), 2049–2060 (2006).
- [5] P. Cervenka and C. De Moustier, “Sidescan Sonar Image Processing Techniques,” *IEEE Journal of Oceanic Engineering* **18**(2), 108–122 (1993).
- [6] R. P. Gorman and T. J. Sejnowski, “Analysis of hidden units in a layered network trained to classify sonar targets,” *Neural Networks* **1**(1), 75–89 (1988).
- [7] R. P. Gorman and T. J. Sejnowski, “Learned Classification of Sonar Targets Using a Massively Parallel Network,” *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**(7), 1135–1140 (1988).
- [8] S. Perry and L. Guan, “Detection of small man-made objects in multiple range sector scan imagery using neural networks,” *Proceedings of the CAD/CAC conference* **4**, 2108–2114 (2001).
- [9] P. Gerstoft and C. Mecklenbrauker, “Ocean acoustic inversion with estimation of a posteriori probability distributions,” *The Journal of the Acoustical Society of America* **104**, 808–819 (1998).
- [10] D. M. F. Chapman, “What are we inverting for?,” in *Inverse Problems in Underwater Acoustics* (Springer New York, 2001), 1–14.
- [11] D. R. Jackson and M. D. Richardson, *High-Frequency Seafloor Acoustics* (Springer, New York, New York, USA, 2007), 616.

- [12] D. R. Jackson, D. Winebrenner, and A. Ishimaru, “Application of the composite roughness model to high-frequency bottom backscattering,” *The Journal of the Acoustical Society of America* **79**(May), 1410 (1986).
- [13] A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993), 1–212.
- [14] E. Livingston and O. Diachok, “Estimation of average under-ice reflection amplitudes and phases using matched-field processing,” *Journal of the Acoustical Society of America* **86**, 1909–1919 (1989).
- [15] M. D. Collins, W. A. Kuperman, and H. Schmidt, “Nonlinear inversion for ocean-bottom properties,” *J. Acoust. Soc. America* **92**(5), 2770–2783 (1992).
- [16] P. Gerstoft, “Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions,” *The Journal of the Acoustical Society of America* **95**(2), 770–782 (1994).
- [17] J. Shorey, L. Nolte, and J. Krolik, “Computationally efficient Monte Carlo estimation algorithms for matched field processing in uncertain ocean environments,” *Journal of Computational Acoustics* **2**(3), 285–314 (1994).
- [18] S. E. Dosso, “Quantifying uncertainty in geoacoustic inversion. I. a fast Gibbs sampler approach,” *The Journal of the Acoustical Society of America* **111**(1), 129–142 (2002).
- [19] D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, “Broadband sound propagation in shallow water and geoacoustic inversion,” *The Journal of the Acoustical Society of America* **113**(1), 205–222 (2003).
- [20] Z.-H. Michalopoulou and U. Ghosh-Dastidar, “Tabu for matched-field source localization and geoacoustic inversion,” *Journal of the Acoustical Society of America* **115**(1), 135–145 (2004).
- [21] Z.-H. Michalopoulou, D. Alexandrou, and C. de Moustier, “Application of a maximum likelihood processor to acoustic backscatter for the estimation of seafloor roughness parameters,” *The Journal of the Acoustical Society of America* **95**(5), 2467–2477 (1994).
- [22] Z. Michalopoulou and D. Alexandrou, “Bayesian modeling of acoustic signals for seafloor identification,” *The Journal of the Acoustical Society of America* **99**(1), 223–233 (1996).
- [23] A. Caiti and S. M. Jesus, “Acoustic estimation of seafloor parameters: A radial basis functions approach,” *The Journal of the Acoustical Society of America* **100**(3), 1473–1481 (1996).
- [24] G. Steininger, C. W. Holland, S. E. Dosso, and J. Dettmer, “Seabed roughness parameters from joint backscatter and reflection inversion at the Malta Plateau,” *Journal of the Acoustical Society of America* **134**(3), 1833–1842 (2013).
- [25] B. Engquist, C. Frederick, Q. Huynh, and H. Zhou, “Seafloor identification in sonar imagery via simulations of helmholtz equations and discrete optimization,” *Journal of Computational Physics* **338**, 477–492 (2017).
- [26] B. Zou, J. Zhai, Z. Qi, and Z. Li, “A comparison of three sediment acoustic models using bayesian inversion and model selection techniques,” *Remote Sensing* **11**, 1–30 (2019).

- [27] Y. Stephan, X. Demoulin, and O. Sarzeaud, “Neural direct approaches for geoacoustic inversion,” *Journal of Computational Acoustics* **6**(01n02), 151–166 (1998).
- [28] B. Chakraborty, V. Kodagali, and J. Baracho, “Sea-floor classification using multibeam echosounding angular backscatter data: a real-time approach employing hybrid neural network architecture,” *IEEE Journal of Oceanic Engineering* **28**(1), 121–128 (2003).
- [29] Z.-H. Michalopoulou, D. Alexandrou, and C. de Moustier, “Application of neural and statistical classifiers to the problem of seafloor characterization,” *IEEE Journal of Oceanic Engineering* **20**(3), 190–197 (1995).
- [30] J. Benson, N. R. Chapman, and A. Antoniou, “Geoacoustic model inversion using artificial neural networks,” *Inverse Problems* **16**(6), 1627 (2000).
- [31] E. Coiras, P. Y. Mignotte, Y. R. Petillot, J. Bell, and K. Lebart, “Supervised target detection and classification by training on augmented reality data,” *IET Radar, Sonar and Navigation* **1**(1), 83–90 (2007).
- [32] S. Reed, Y. Petillot, and J. Bell, “An automatic approach to the detection and extraction of mine features in sidescan sonar,” *IEEE Journal of Oceanic Engineering* **28**(1), 90–105 (2003).
- [33] Q. Kong, D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft, “Machine learning in seismology: Turning data into insights,” *Seismological Research Letters* **90**(1), 3–14 (2019).
- [34] M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle, “Machine learning in acoustics: Theory and applications,” *The Journal of the Acoustical Society of America* **146**(5), 3590–3628 (2019).
- [35] D. P. Williams, “On the use of tiny convolutional neural networks for human-expert-level classification performance in sonar imagery,” *IEEE Journal of Oceanic Engineering* 1–25 (2020).
- [36] X. Ye, C. Li, S. Zhang, P. Yang, and X. Li, “Research on side-scan sonar image target classification method based on transfer learning,” *OCEANS 2018 MTS/IEEE Charleston, OCEAN 2018* 1–6 (2019).
- [37] T. Berthold, A. Leichter, B. Rosenhahn, V. Berkhahn, and J. Valerius, “Seabed sediment classification of side-scan sonar data using convolutional neural networks,” *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017 - Proceedings* **2018-January**, 1–8 (2018).
- [38] D. Van Komen, T. B. Neilsen, D. P. Knobles, and M. Badiey, “A convolutional neural network for source range and ocean seabed classification using pressure time-series,” *Proceedings of Meetings on Acoustics* **36**(1), 070004 (2019).
- [39] D. Van Komen, T. B. Neilsen, K. Howarth D. P. Knobles, and P. H. Dahl, “Seabed and range estimation of impulsive time series using a convolutional neural network,” *The Journal of the Acoustical Society of America* **147**(5), EL403-EL408 (2020).
- [40] I. Kvasic, N. Miskovic, and Z. Vukic, “Convolutional Neural Network Architectures for Sonar-Based Diver Detection and Tracking,” *OCEANS 2019 - Marseille* 1–6 (2019).

- [41] P. Zhu, J. Isaacs, B. Fu, and S. Ferrari, “Deep learning feature extraction for target recognition and classification in underwater sonar images,” in *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, IEEE (2017), 2724–2731.
- [42] Y. Xu, Y. Zhang, H. Wang, and X. Liu, “Underwater image classification using deep convolutional neural networks and data augmentation,” in *2017 IEEE International Conference on Signal Processing, Communications and Computing, ICSPCC 2017* (2017), Vol. 2017-January, 1–5.
- [43] J. Ding, B. Chen, H. Liu, and M. Huang, “Convolutional Neural Network with Data Augmentation for SAR Target Recognition,” *IEEE Geoscience and Remote Sensing Letters* **13**(3), 364–368 (2016).
- [44] H. T. Nguyen, E. H. Lee, and S. Lee, “Study on the classification performance of underwater sonar image classification based on convolutional neural networks for detecting a submerged human body,” *Sensors (Switzerland)* **20**(1) (2020).
- [45] H. Niu, E. Reeves, and P. Gerstoft, “Source localization in an ocean waveguide using supervised machine learning,” *The Journal of the Acoustical Society of America* **142**(3), 1176–1188 (2017).
- [46] Y. Ren and Y. Qi, “Waveguide invariant and range estimation based on phase-shift-compensation of underwater acoustic spectrograms,” in *AIP Conference Proceedings*, AIP (2012), Vol. 1495, 627–633.
- [47] M. S. Scheide, “Using Deep Learning for Automatic Classification of Marine Habitats in HiSAS Imagery,” (July), 155 (2018).
- [48] J. Chen and J. E. Summers, “Deep convolutional neural networks for semi-supervised learning from synthetic aperture sonar (SAS) images,” 173rd Meeting of Acoustical Society of America and 8th Forum Acusticum **30**(2017), 055018 (2019).
- [49] T. B. Neilsen, M. K. Transtrum, D. F. Van Komen, and D. P. Knobles, “Optimal experimental design for machine learning using the Fisher information matrix,” *The Journal of the Acoustical Society of America* **144**(3), 1730–1730 (2018).
- [50] M. Porter, “The Kraken Normal Mode Program,” (2017).
- [51] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer New York, New York, NY, 2011), 1–500.
- [52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
- [53] M. J. Isakson, “High Frequency Acoustic Reflection and Transmission in Ocean Sediments,” *Development* (512), 1–10 (2014).
- [54] G. Bao, “Finite element approximation of time harmonic waves in periodic structures,” *SIAM journal on numerical analysis* **32**(4), 1155–1169 (1995).

- [55] O. P. Bruno and F. Reitich, “Solution of a boundary value problem for the helmholtz equation via variation of the boundary into the complex domain,” (1992).
- [56] J. J. Shirron and T. E. Giddings, “A finite element model for acoustic scattering from objects near a fluid-fluid interface,” *Computer Methods in Applied Mechanics and Engineering* **196**(1-3), 279–288 (2006).
- [57] C. P. Vendhan, G. C. Diwan, and S. K. Bhattacharyya, “Finite-element modeling of depth and range dependent acoustic propagation in oceanic waveguides,” *The Journal of the Acoustical Society of America* **127**(6), 3319–3326 (2010) 10.1121/1.3392440.
- [58] M. J. Isakson, R. a. Yarbrough, and N. P. Chotiros, “A finite element model for seafloor roughness scattering,” *Proceedings of the International Symposium on Underwater Reverberation and Clutter* 173–180 (2008).
- [59] F. B. Jensen and M. Zampolli, “Finite-element modeling in ocean acoustics: Where are we heading?,” *AIP Conference Proceedings* **1272**, 11–22 (2010).
- [60] Q. Huynh, K. Ito, and J. Toivanen, “A Fast Helmholtz Solver for Scattering by a Sound-soft Target in Sediment,” in *Domain Decomposition Methods in Science and Engineering XVI* (Springer Berlin Heidelberg, 2007), 595–602.
- [61] S. G. Kargl, A. L. Espana, K. L. Williams, J. L. Kennedy, and J. L. Lopes, “Scattering From Objects at a Water-Sediment Interface: Experiment, High-Speed and High-Fidelity Models, and Physical Insight,” *IEEE Journal of Oceanic Engineering* **40**(3), 632–642 (2014).
- [62] J. E. Murphy and S. A. Chin-Bing, “A Seismo-Acoustic Finite Element Model for Underwater Acoustic Propagation,” in *Shear Waves in Marine Sediments* (Springer Netherlands, Dordrecht, 2011), 463–470.
- [63] A. T. Abawi, W. A. Kuperman, and M. D. Collins, “The coupled mode parabolic equation,” *The Journal of the Acoustical Society of America* **102**(1), 233–238 (1997).
- [64] M. Y. Trofimov, S. B. Kozitskiy, A. D. Zakharenko, and P. S. Petrov, “A ray mode parabolic equation, and examples of its application in shallow water acoustics propagation problems,” *Proceedings of Meetings on Acoustics* **24**(1), 070010 (2015).
- [65] B. Engquist and H. Zhao, “Approximate separability of the Green’s function of the Helmholtz equation in the high frequency limit,” *Communications on Pure and Applied Mathematics* **71**(11), 2220–2274 (2018).
- [66] S. L. Broschat and E. I. Thorsos, “An investigation of the small slope approximation for scattering from rough surfaces. Part II. Numerical studies,” *The Journal of the Acoustical Society of America* **101**(5), 2615–2625 (2002).
- [67] A. G. Voronovich, “Wave diffraction by rough interfaces in an arbitrary plane-layered medium,” *Waves Random Media* **10**(2), 253–272 (2000).
- [68] V. Jaud, C. Gervaise, Y. Stephan, and A. Khenchaf, “Modeling of high-frequency roughness scattering from various rough surfaces through the small slope approximation of first order,” *Open Journal of Acoustics* **2**(1), 1–11 (2012).

- [69] W. A. Kuperman and H. Schmidt, “Selfconsistent perturbation approach to rough surface scattering in stratified elastic media,” *The Journal of the Acoustical Society of America* **86**(4), 1511–1522 (2005).
- [70] D. R. Jackson and A. N. Ivakin, “Modeling issues in scattering by layered seabeds,” in *Proceedings of the 11th European Conference on Underwater Acoustics*, 5 (2016), 1180–1187.
- [71] P. Gerstoft and H. Schmidt, “A boundary element approach to ocean seismoacoustic facet reverberation,” *The Journal of the Acoustical Society of America* **89**(4), 1629–1642 (2005).
- [72] M. Zampolli, A. Tesei, G. Canepa, and O. A. Godin, “Computing the far field scattered or radiated by objects inside layered fluid media using approximate Green’s functions,” *The Journal of the Acoustical Society of America* **123**(6), 4051–4058 (2008).
- [73] J. D. Benamou, F. Collino, and O. Runborg, “Numerical microlocal analysis of harmonic wavefields,” *Journal of Computational Physics* **199**(2), 717–741 (2004).
- [74] Y. Landa, N. M. Tanushev, and R. Tsai, “Discovery of point sources in the Helmholtz equation posed in unknown domains with obstacles,” *Communications in Mathematical Sciences* **9**(3), 903–928 (2011).
- [75] J. Benamou, F. Collino, and S. Marmorat, “Numerical microlocal analysis revisited,” INRIA (2011).
- [76] J. Fang, J. Qian, L. Zepeda-Núñez, and H. Zhao, “A hybrid approach to solve the high-frequency Helmholtz equation with source singularity in smooth heterogeneous media,” *Journal of Computational Physics* **371**, 261–279 (2018).
- [77] C. AB, “COMSOL Multiphysics® v. 5.2. www.comsol.com.” .
- [78] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems* (2012), 1097–1105.
- [79] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), 1–9.
- [80] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), 770–778.
- [81] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556* (2014).
- [82] <https://github.com/cf87/Seabed-Classification-2020>.
- [83] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *International Conference of Learning Representations (ICLR)* (2015).
- [84] B. Recht, R. Roelofs, L. Schmidt, and V. Shankar, “Do cifar-10 classifiers generalize to cifar-10?,” *arXiv preprint arXiv:1806.00451* (2018).

- [85] L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” *Journal of machine learning research* **9**(Nov), 2579–2605 (2008).
- [86] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision* (2017), 618–626.
- [87] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” *International Conference of Learning Representations (ICLR)* (2015).

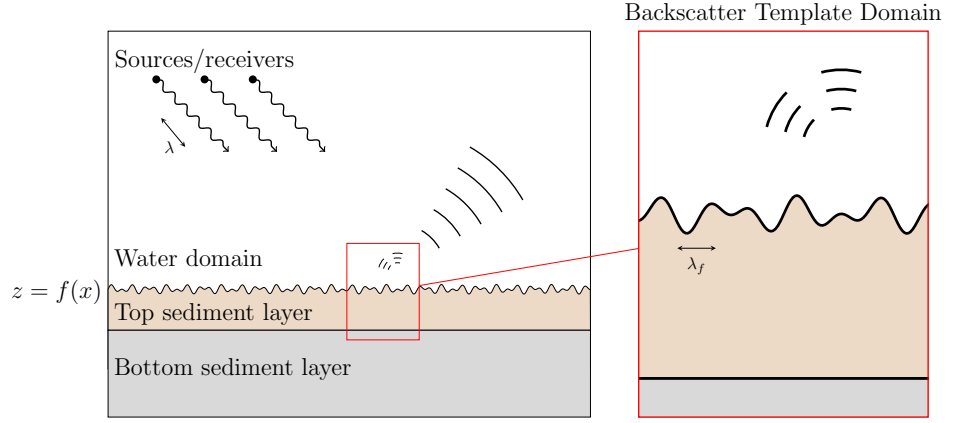
6 Tables

Parameter	Training	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9	Test 10
c (m/s)											
Clay	1500	1517	1521	1517	1546	1517	1529	1526	1518	1527	1518
Silt	1575	1577	1592	1582	1574	1591	1586	1596	1581	1584	1580
Sand	1650	1658	1632	1663	1647	1652	1644	1642	1672	1648	1655
Gravel	1800	1794	1795	1799	1796	1792	1802	1809	1791	1801	1784
α (dB/ λ)											
Clay	0.2	0.271	0.077	0.296	0.175	0.224	0.160	0.066	0.256	0.215	0.166
Silt	1.0	1.050	1.064	0.954	0.892	0.982	1.039	1.188	1.046	0.922	1.189
Sand	0.8	1.042	0.843	0.914	0.708	0.839	0.681	0.970	0.881	0.595	0.797
Gravel	0.6	0.495	0.683	0.547	0.458	0.666	0.616	0.740	0.651	0.680	0.775
τ (m)											
All	5	*	*	*	*	*	*	*	*	*	*

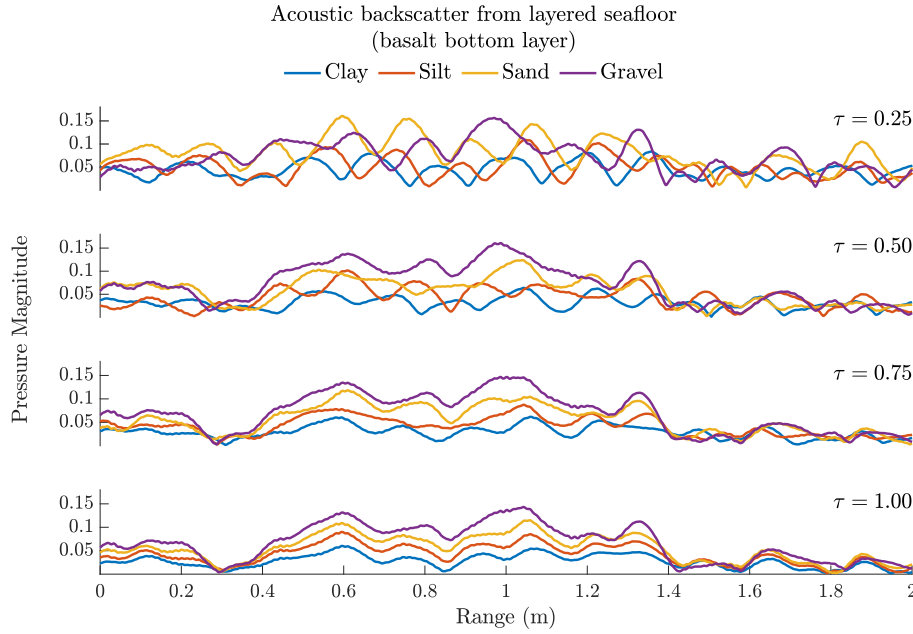
Table 1: Environmental parameters varied to generate synthetic data with KRAKEN. An ‘*’ indicates the parameters used are the same as those used in the training data.

Parameter	Units	Training	Test 1	Test 2	Test 3	Test 4
Roughness						
RMS height	(cm)	.5	.46	.46	.46	.46
RMS corr. len.	(cm)	2	1.5	1.5	1.5	1.5
Sound Speed (c_{top}, c_{bottom})						
Clay	(m/s)	(1500, 5250)	*	(1501.57, *)	*	(1501.57, 5254.58)
Silt	(m/s)	(1575, 5250)	*	(1577.03, *)	*	(1577.03, 5254.65)
Sand	(m/s)	(1650, 5250)	*	(1648.13, *)	*	(1648.13, 5246.58)
Gravel	(m/s)	(1800, 5250)	*	(1802.07, *)	*	(1802.07, 5254.71)
Density ($\rho_{top}, \rho_{bottom}$)						
Clay	(kg/m ³)	(1500, 2700)	*	(1500.66, *)	*	(1500.66, 2704.57)
Silt	(kg/m ³)	(1700, 2700)	*	(1697.99, *)	*	(1697.99, 2699.85)
Sand	(kg/m ³)	(1900, 2700)	*	(1898.89, *)	*	(1898.89, 2703.00)
Gravel	(kg/m ³)	(2000, 2700)	*	(1802.07, *)	*	(1802.07, 2696.42)
Thickness						
τ	(m)	{.25, .5, .75, 1}	*	*	{0.253, 0.504, 0.725, 1.004}	*

Table 2: Environmental parameters varied to generate synthetic backscatter data. An ‘*’ indicates the parameters used are the same as those used in the training data. Among the parameters that are fixed throughout all of the data are the acoustic frequency $\omega = 15$ kHz, incident angle $\theta = \pi/12$ radians, and domain size of 2×2 meters.



(a)



(b)

Figure 3: (a) High-frequency scattering from a two-layer seafloor with rough water-sediment interface given by $z = f(x)$. The acoustic wavelength λ of the incoming source ping is comparable to the spatial wavelength λ_f of the water-sediment interface. (b) Approximate backscatter signals corresponding to environments with varying sediments and top layer thicknesses (τ).

Fig4.pdf

Figure 4: Cross entropy loss as a function of training epoch for six CNN classifiers.

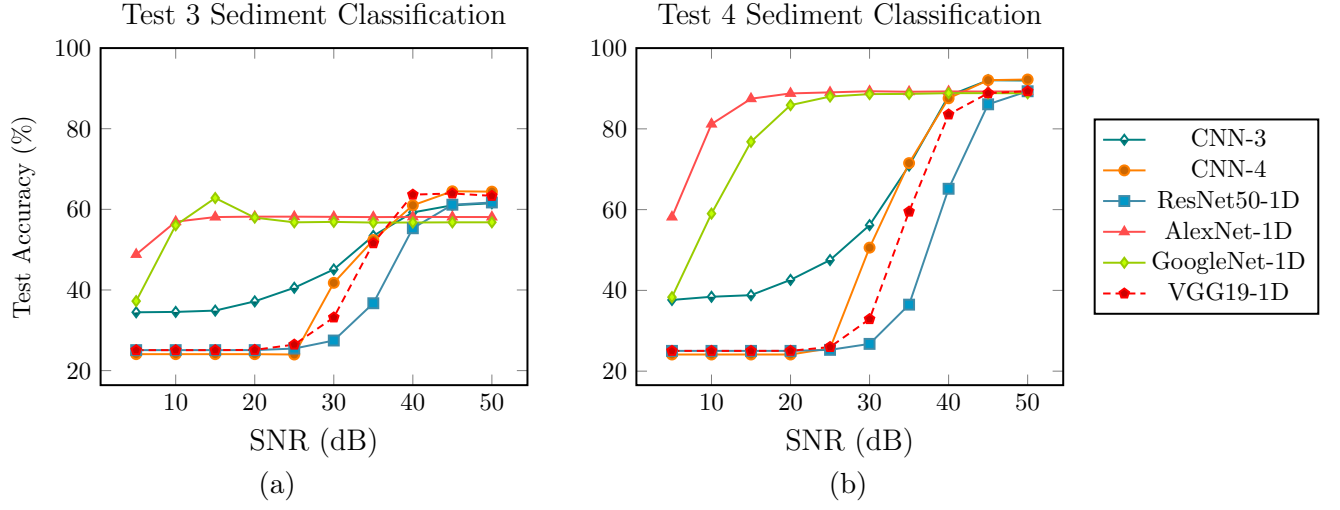


Figure 5: Classification performance of different trained CNNs on the (a) Test 3 and (b) Test 4 environments.

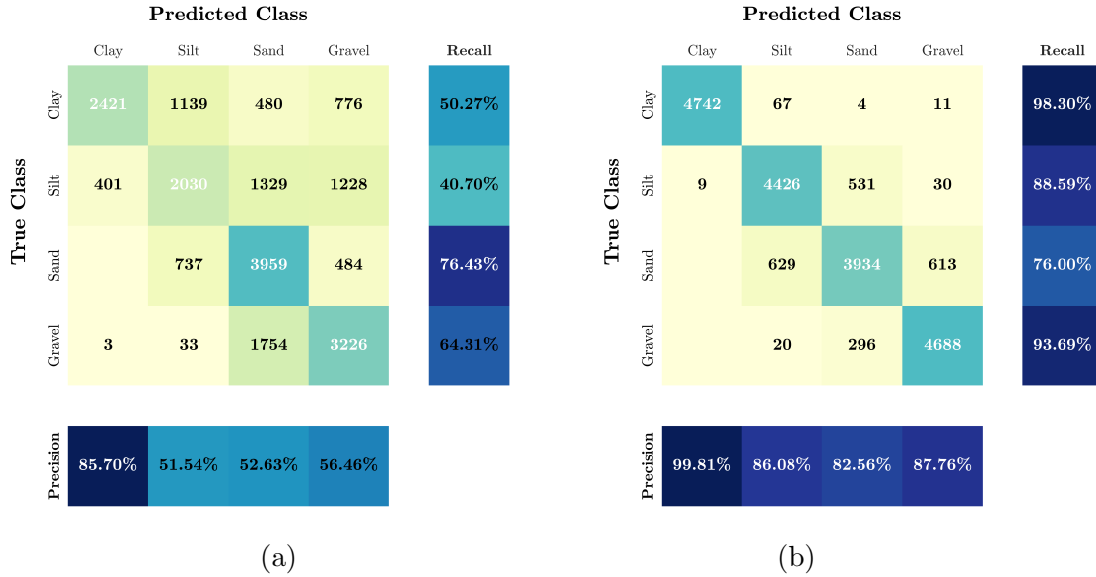
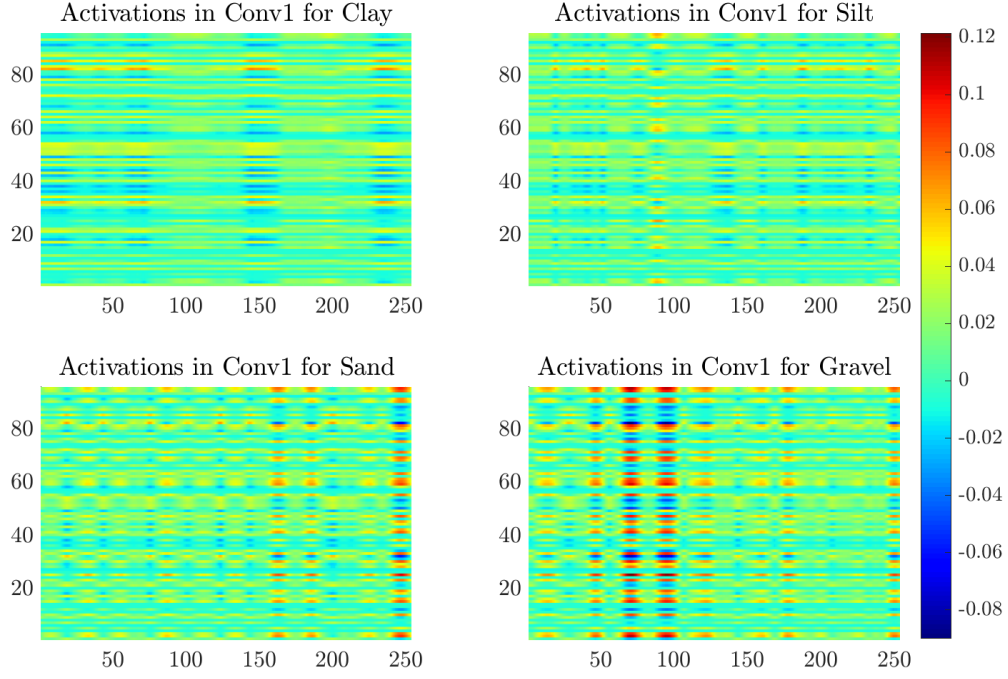
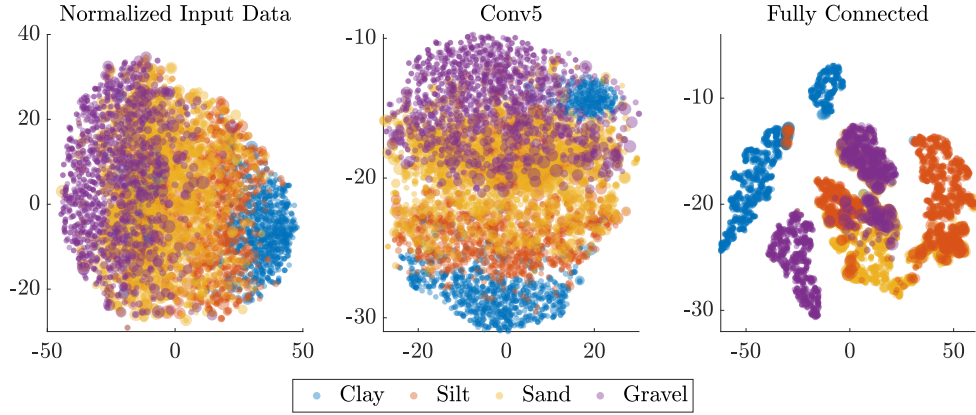


Figure 6: Confusion matrices for sediment classification using AlexNet-1D on noisy test data (SNR = 20 dB) on the (a) Test 3 and (b) Test 4 environments.



(a)



(b)

Figure 7: (a) Activations in the 'Conv1' layer of the trained AlexNet-1D classifier given input test signals from environments with different sediment types. (b) t-SNE visualization of the activations in three different layers of the trained AlexNet-1D classifier applied to test data from the Test 4 environment.

Classifier	Training Time	Classification Accuracy (%)				
		Val.	Test 1	Test 2	Test 3	Test 4
	(CPU)					
Nearest Centroid	0:00:03	63.3	66.63	64.54	53.37	64.67
Random Forest Classifier	0:00:07	71.5	71.74	70.15	53.86	71.44
Decision Tree Classifier	0:00:10	55.1	55.95	54.36	43.17	53.94
K Neighbors Classifier	0:00:34	55.7	59.41	57.08	46.29	57.36
SVC (radial basis kernel)	0:02:05	77.3	76.62	75.76	56.63	76.19
SVC (linear kernel)	0:02:18	67.4	59.44	58.14	42.98	58.36
Multilayer Perceptron	0:05:38	73.6	73.04	72.43	51.47	73.43
Logistic Regression	0:45:52	67.7	57.38	55.74	41.54	56.36
	(GPU)					
AlexNet-1D	00:05:30	88.95	89.72	88.46	58.19	88.78
CNN-3	00:08:28	91.61	43.45	42.49	37.18	42.60
CNN-4	00:10:47	92.41	24.34	24.10	24.08	24.12
ResNet50-1D	01:00:20	89.62	26.16	25.08	25.09	25.02
GoogleNet-1D	07:30:01	86.03	86.14	86.61	57.95	85.88
VGG19-1D	31:37:46	90.31	26.19	25.12	25.15	25.04

Table 3: Training time and accuracy of different shallow machine learning tools (top) and CNNs (bottom) applied to a validation set (no added noise) and four test sets (noise added with SNR = 20 dB).