

Source localization in an ocean waveguide using supervised machine learning

Haiqiang Niu, Emma Reeves, and Peter Gerstoft

Citation: [The Journal of the Acoustical Society of America](#) **142**, 1176 (2017); doi: 10.1121/1.5000165

View online: <http://dx.doi.org/10.1121/1.5000165>

View Table of Contents: <http://asa.scitation.org/toc/jas/142/3>

Published by the [Acoustical Society of America](#)

Articles you may be interested in

[Acoustic-intensity striations below the critical depth: Interpretation and modeling](#)

[The Journal of the Acoustical Society of America](#) **142**, EL245 (2017); 10.1121/1.5000325

[Spherical wave scattering from rough surfaces and array processing: Application to sound-speed profile measurement uncertainty analysis](#)

[The Journal of the Acoustical Society of America](#) **142**, 1189 (2017); 10.1121/1.5000234

[Blind deconvolution of shipping sources in an ocean waveguide](#)

[The Journal of the Acoustical Society of America](#) **141**, 797 (2017); 10.1121/1.4976046

[Reciprocity relationships in vector acoustics and their application to vector field calculations](#)

[The Journal of the Acoustical Society of America](#) **142**, 523 (2017); 10.1121/1.4996458

[Application of maximum entropy to statistical inference for inversion of data from a single track segment](#)

[The Journal of the Acoustical Society of America](#) **142**, 737 (2017); 10.1121/1.4996456

[Removal of incoherent noise from an averaged cross-spectral matrix](#)

[The Journal of the Acoustical Society of America](#) **142**, 846 (2017); 10.1121/1.4997923

Source localization in an ocean waveguide using supervised machine learning

Haiqiang Niu,^{a)} Emma Reeves, and Peter Gerstoft

Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0238, USA

(Received 9 December 2016; revised 18 June 2017; accepted 8 August 2017; published online 1 September 2017)

Source localization in ocean acoustics is posed as a machine learning problem in which data-driven methods learn source ranges directly from observed acoustic data. The pressure received by a vertical linear array is preprocessed by constructing a normalized sample covariance matrix and used as the input for three machine learning methods: feed-forward neural networks (FNN), support vector machines (SVM), and random forests (RF). The range estimation problem is solved both as a classification problem and as a regression problem by these three machine learning algorithms. The results of range estimation for the Noise09 experiment are compared for FNN, SVM, RF, and conventional matched-field processing and demonstrate the potential of machine learning for underwater source localization. © 2017 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.5000165>]

[SED]

Pages: 1176–1188

I. INTRODUCTION

Machine learning is a promising method for locating ocean sources because of its ability to learn features from data, without requiring sound propagation modeling. It can be used for unknown environments.

Acoustic source localization in ocean waveguides is often solved with matched-field processing (MFP).^{1–13} Despite the success of MFP, it is limited in some practical applications due to its sensitivity to the mismatch between model-generated replica fields and measurements. MFP gives reasonable predictions only if the ocean environment can be accurately modeled. Unfortunately, this is difficult because the realistic ocean environment is complicated and unstable.

An alternative approach to the source localization problem is to find features directly from data.^{14–17} Interest in machine learning techniques has been revived thanks to increased computational resources as well as their ability to learn nonlinear relationships. A notable recent example in ocean acoustics is the application of nonlinear regression to source localization.¹⁸ Other machine learning methods have obtained remarkable results when applied to areas such as speech recognition,¹⁹ image processing,²⁰ natural language processing,²¹ and seismology.^{22–25} Most underwater acoustics research in machine learning is based on 1990s neural networks. Previous research has applied neural networks to determine the source location in a homogeneous medium,²⁶ simulated range and depth discrimination using artificial neural networks in matched-field processing,²⁷ estimated ocean-sediment properties using radial basis functions in regression and neural networks,^{28,29} applied artificial neural

networks to estimation of geoacoustic model parameters,^{30,31} classification of seafloor³² and whale sounds.³³

This paper explores the use of current machine learning methods for source range localization. The feed-forward neural network (FNN), support vector machine (SVM), and random forest (RF) methods are investigated. There are several main differences between our work and previous studies of source localization and inversion:^{18,26–33}

- (1) Acoustic observations are used to train the machine learning models instead of using model-generated fields.^{26,27,29–31}
- (2) For input data, normalized sample covariance matrices, including amplitude and phase information, are used. Other alternatives include the complex pressure,¹⁸ phase difference,²⁶ eigenvalues,²⁷ amplitude of the pressure field,²⁹ transmission loss,^{30,31} angular dependence of backscatter,³² or features extracted from spectrograms.³³ This preprocessing procedure is known as feature extraction in machine learning.
- (3) Under machine learning framework, source localization can be solved as a classification or a regression problem. This work focuses on classification in addition to the regression approach used in previous studies.^{18,26,28–31}
- (4) Well-developed machine learning libraries are used. Presently, there are numerous efficient open source machine learning libraries available, including TensorFlow,³⁴ Scikit-learn,³⁵ Theano,³⁶ Caffe,³⁷ and Torch,³⁸ all of which solve typical machine learning tasks with comparable efficiency. Here, TensorFlow is used to implement FNN because of its simple architecture and wide user base. Scikit-learn is used to implement SVM and RF as they are not included in the current TensorFlow version. Compared to older neural network implementations, Tensorflow includes improved optimization algorithms³⁹ with better convergence, more robust model with dropout⁴⁰ technique and high computational efficiency.

^{a)}Also at: State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China. Electronic mail: hniu@ucsd.edu

The paper is organized as follows. The input data preprocessing and source range mapping are discussed in Secs. II A and II B. The theoretical basis of FNN, SVM, and RF is given in Secs. II C–II E. Simulations and experimental results in Secs. III and IV demonstrate the performance of FNN, SVM, and RF. In Sec. V, the effect of varying the model parameters is discussed. The conclusion is given in Sec. VI.

II. LOCALIZATION BASED ON MACHINE LEARNING

The dynamics of the ocean and its boundary cause a stochastic relationship between the received pressure phase and amplitude at the array and the source range. After preprocessing we assume a deterministic relationship between ship range and sample covariance matrix. The pressure–range relationship is in general unknown but may be discovered using machine learning methods. The received pressure is preprocessed and used as the input of the machine learning models (Sec. II A). The desired output may be either discrete (classification) or continuous (regression) corresponding to the estimated source range (Sec. II B). The theory of FNN, SVM, and RF are described in Secs. II C–II E.

A. Input data preprocessing

To make the processing independent of the complex source spectra, the received array pressure is transformed to a normalized sample covariance matrix. The complex pressure at frequency f obtained by taking the discrete fourier transform of the input pressure data at L sensors is denoted by $\mathbf{p}(f) = [p_1(f), \dots, p_L(f)]^T$. The sound pressure is modeled as

$$\mathbf{p}(f) = S(f)\mathbf{g}(f, \mathbf{r}) + \epsilon, \quad (1)$$

where ϵ is the noise, $S(f)$ is the source term, and \mathbf{g} is the Green's function. To reduce the effect of the source amplitude $|S(f)|$, this complex pressure is normalized according to

$$\tilde{\mathbf{p}}(f) = \frac{\mathbf{p}(f)}{\sqrt{\sum_{l=1}^L |p_l(f)|^2}} = \frac{\mathbf{p}(f)}{\|\mathbf{p}(f)\|_2}. \quad (2)$$

The normalized sample covariance matrices (SCMs) are averaged over N_s snapshots to form the conjugate symmetric matrix

$$\mathbf{C}(f) = \frac{1}{N_s} \sum_{s=1}^{N_s} \tilde{\mathbf{p}}_s(f) \tilde{\mathbf{p}}_s^H(f), \quad (3)$$

where H denotes conjugate transpose operator and $\tilde{\mathbf{p}}_s$ represents the sound pressure over the s th snapshot. The product $\tilde{\mathbf{p}}_s(f) \tilde{\mathbf{p}}_s^H(f)$ contains an $S(f)S(f)^H$ term, which for large SNR is dominant and thus reduces the effect of the source phase. Preprocessing the data according to Eqs. (2) and (3) ensures that the Green's function is used for localization. Only the real and imaginary parts of the complex valued entries of diagonal and upper triangular matrix in $\mathbf{C}(f)$ are used as input to save memory and improve calculation speed. These

entries are vectorized to form the real-valued input \mathbf{x} of size $L \times (L + 1)$ to the FNN, SVM, and RF.

B. Source range mapping

In the classification problem, a set of source ranges is discretized into K bins, r_1, \dots, r_K , of equal width Δr . Each input vector, \mathbf{x}_n , $n = 1, \dots, N$, is labeled by t_n , where $t_n \in r_k$, $k = 1, \dots, K$; this label represents the true source range class and is the target output for the model. SVM and RF use this classification scheme to train and predict the source range for each sample.

For the FNN, the range class t_n is mapped to a $1 \times K$ binary vector, \mathbf{t}_n , such that

$$t_{nk} = \begin{cases} 1 & \text{if } |t_n - r_k| \leq \frac{\Delta r}{2}, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

$\mathbf{t}_n = t_{n,1}, \dots, t_{n,K}$ therefore represents the expected output probability of the neural network, i.e., the probability that the source is at range r_k for input \mathbf{x}_n . These target vectors are used to train the FNN. The FNN output predictions are given as a softmax distribution with maximum at the predicted range (see Sec. II C).

In the regression problem, the target output $r_n \in [0, \infty)$ is a continuous range variable for all three models.

C. Feed-forward neural networks

The feed-forward neural network (FNN), also known as multi-layer perceptron, is constructed using a feed-forward directed acyclic architecture, see Fig. 1(a). The outputs are formed through a series of functional transformations of the weighted inputs. In the FNN, the outputs are deterministic functions of the inputs.⁴¹

Here, a three layer model (input layer L_1 , hidden layer L_2 and output layer L_3) is used to construct the FNN. The input layer L_1 is comprised of D input variables $\mathbf{x} = [x_1, \dots, x_D]^T$. The j th linear combination of the input variables is given by

$$a_j = \sum_{i=1}^D w_{ji}^{(1)} x_i + w_{j0}^{(1)}, \quad j = 1, \dots, M, \quad (5)$$

where M is the number of neurons in L_2 and the superscript indicates that the corresponding parameters are in the first layer of the network. The parameters $w_{ji}^{(1)}$ and $w_{j0}^{(1)}$ are called the weights and biases and their linear combinations a_j are called activations. In L_2 , the activations are transformed using an activation function $f(\cdot)$,

$$z_j = f(a_j). \quad (6)$$

The logistic sigmoid was chosen as the intermediate activation function for this study, see Fig. 1(b):

$$f(a) = \sigma(a) = \frac{1}{1 + e^{-a}}. \quad (7)$$

Similarly, for output layer L_3 , the K output unit activations are expressed as linear combinations of z_j

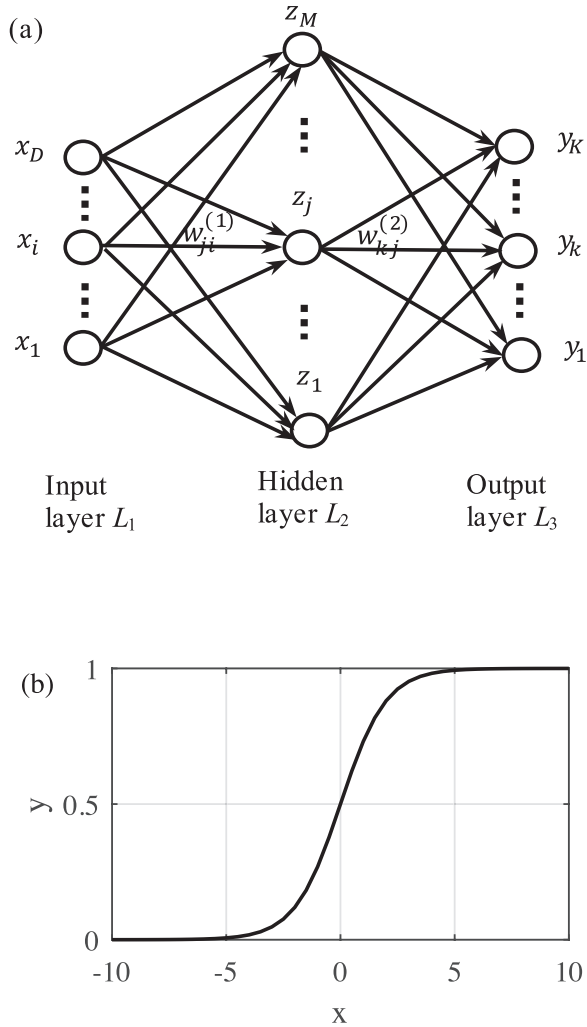


FIG. 1. (a) Diagram of a feed-forward neural network and (b) Sigmoid function.

$$a_k = \sum_{j=1}^M w_{kj}^{(2)} z_j + w_{k0}^{(2)}, \quad k = 1, \dots, K \quad (8)$$

where $w_{kj}^{(2)}$ and $w_{k0}^{(2)}$ represent weights and biases for the second layer.

In the output layer, the softmax function is used as the activation function. The softmax is a common choice for multi-class classification problems.⁴¹ Here, it constrains the output class, $y_k(\mathbf{x}, \mathbf{w})$, to be the probability that the source is at range r_k (Ref. 41):

$$y_k(\mathbf{x}, \mathbf{w}) = \frac{\exp(a_k(\mathbf{x}, \mathbf{w}))}{\sum_{j=1}^K \exp(a_j(\mathbf{x}, \mathbf{w}))}, \quad k = 1, \dots, K \quad (9)$$

where \mathbf{w} is the set of all weight and bias parameters and y_k satisfies $0 \leq y_k \leq 1$ and $\sum_k y_k = 1$.

Before applying the FNN to unlabeled data, the weights and biases \mathbf{w} are determined by training the model on labeled data. Recall that in the FNN case, \mathbf{t}_n is the binary target vector, or true probability distribution (see Sec. II B), and

$y_k(\mathbf{x}_n, \mathbf{w})$ is the estimated probability distribution, for the input \mathbf{x}_n (see Sec. II A).

During training, the Kullback–Leibler (KL) divergence

$$D_{\text{KL}}(\mathbf{t}_n || \mathbf{y}(\mathbf{x}_n, \mathbf{w})) = \sum_k t_{nk} [\ln t_{nk} - \ln y_{nk}], \quad (10)$$

represents the dissimilarity between $y_{nk} = y_k(\mathbf{x}_n, \mathbf{w})$ and t_{nk} , where $\mathbf{t}_n = [t_{n1}, \dots, t_{nK}]$, $k = 1, \dots, K$. Minimizing the KL divergence D_{KL} is equivalent to minimizing the cross entropy function E_n

$$E_n(\mathbf{t}_n, \mathbf{y}(\mathbf{x}_n, \mathbf{w})) = - \sum_k t_{nk} \ln y_{nk}, \quad (11)$$

since the desired output \mathbf{t}_n is constant (independent of \mathbf{w}).

For N observation vectors, the averaged cross entropy and resulting weights and biases are

$$E(\mathbf{w}) = - \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K t_{nk} \ln y_{nk}, \quad (12)$$

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\text{argmin}} \left[- \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K t_{nk} \ln y_{nk} \right]. \quad (13)$$

For the regression problem, there is only one neuron in the output layer representing the continuous range variable. Instead of using Eq. (12), a sum-of-squares error function⁴¹ is minimized

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N |y(\mathbf{x}_n, \mathbf{w}) - r_n|^2, \quad (14)$$

where r_n is the true source range at sample n .

Several optimization methods are provided in the TensorFlow software. In this paper, Adam³⁹ (Adaptive Moment estimation) is used.

D. Support vector machine

Unlike neural networks, support vector machines (SVM) are decision machines that do not provide a posterior probability.⁴¹ Instead, the data are divided into two (or more) classes by defining a hyperplane that maximally separates the classes.

First, for simplicity, assume the input \mathbf{x}_n , $n = 1, \dots, N$ are linearly separable (see Fig. 2) and can be divided into two classes, $s_n \in \{1, -1\}$. The class of each input point \mathbf{x}_n is determined by the form⁴¹

$$y_n = \mathbf{w}^T \mathbf{x}_n + b, \quad (15)$$

where \mathbf{w} and b are the unknown weights and bias. A hyperplane satisfying $\mathbf{w}^T \mathbf{x} + b = 0$ is used to separate the classes. If y_n is above the hyperplane ($y_n > 0$), estimated class label $\hat{s}_n = 1$, whereas if y_n is below ($y_n < 0$), $\hat{s}_n = -1$. The perpendicular distance d of a point \mathbf{x}_n to the hyperplane is the distance between the point \mathbf{x}_n and its projection \mathbf{x}_0 on the hyperplane, satisfying

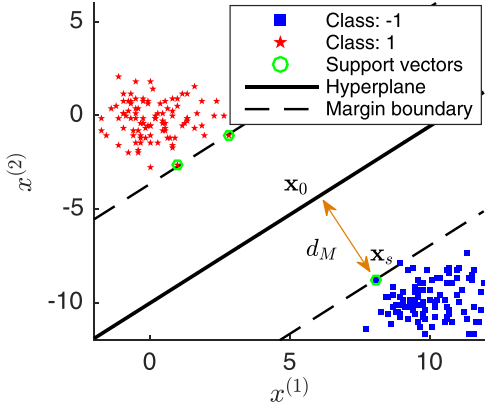


FIG. 2. (Color online) A linear hyperplane learned by training an SVM in two dimensions ($D=2$).

$$\begin{aligned} \mathbf{x}_n &= \mathbf{x}_0 + d \frac{\mathbf{w}}{\|\mathbf{w}\|}, \\ \mathbf{w}^T \mathbf{x}_0 + b &= 0, \end{aligned} \quad (16)$$

where $\|\cdot\|$ is the l_2 norm. From Eq. (16), the distance d is obtained:

$$d(\mathbf{x}_n) = s_n \frac{\mathbf{w}^T \mathbf{x}_n + b}{\|\mathbf{w}\|}, \quad (17)$$

where s_n is added in Eq. (17) to guarantee $d > 0$. The margin d_M is defined as the distance from the hyperplane to the closest points \mathbf{x}_s on the margin boundary (support vectors, see Fig. 2). The optimal \mathbf{w} and b are solved by maximizing the margin d_M :

$$\begin{aligned} \operatorname{argmax}_{\mathbf{w}, b} \quad & d_M, \\ \text{subject to} \quad & \frac{s_n(\mathbf{w}^T \mathbf{x}_n + b)}{\|\mathbf{w}\|} \geq d_M, \quad n = 1, \dots, N. \end{aligned} \quad (18)$$

Equation (18) is equivalent to this optimization problem:⁴¹

$$\begin{aligned} \operatorname{argmin}_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2, \\ \text{subject to} \quad & s_n(\mathbf{w}^T \mathbf{x}_n + b) \geq 1, \quad n = 1, \dots, N. \end{aligned} \quad (19)$$

If the training set is linearly non-separable (class overlapping), slack variables⁴¹ $\xi_n \geq 0$ are introduced to allow some of the training points to be misclassified, corresponding the optimization problem:

$$\begin{aligned} \operatorname{argmin}_{\mathbf{w}, b} \quad & \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{n=1}^N \xi_n, \\ \text{subject to} \quad & s_n y_n \geq 1 - \xi_n, \quad n = 1, \dots, N. \end{aligned} \quad (20)$$

The parameter $C > 0$ controls the trade-off between the slack variable penalty and the margin.

Often the relation between y_n and \mathbf{x}_n is nonlinear. Thus Eq. (15) becomes

$$y_n = \mathbf{w}^T \phi(\mathbf{x}_n) + b, \quad (21)$$

where $\phi(\mathbf{x}_n)$ denotes the feature-space transformation. By substituting $\phi(\mathbf{x}_n)$ for \mathbf{x}_n , Eqs. (16)–(20) are unchanged.

The constrained optimization problem can be rewritten in terms of the dual Lagrangian form:⁴¹

$$\begin{aligned} \tilde{L}(\mathbf{a}) &= \sum_{n=1}^N a_n - \frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N a_n a_m s_n s_m k_\phi(\mathbf{x}_n, \mathbf{x}_m), \\ \text{subject to} \quad & 0 \leq a_n \leq C, \\ & \sum_{n=1}^N a_n s_n = 0, \quad n = 1, \dots, N, \end{aligned} \quad (22)$$

where $a_n \geq 0$ are Lagrange multipliers and dual variables, and $k_\phi(\mathbf{x}_n, \mathbf{x}_m) = \phi(\mathbf{x}_n)^T \phi(\mathbf{x}_m)$ is the kernel function. In this study, we use the Gaussian radial basis function (RBF) kernel³⁵

$$k_\phi(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2), \quad (23)$$

where γ is a parameter that controls the kernel shape.

Support vector regression (SVR) is similar to SVM, but it minimizes the ϵ -sensitive error function

$$\mathcal{E}_\epsilon(y_n - r_n) = \begin{cases} 0, & \text{if } |y_n - r_n| < \epsilon, \\ |y_n - r_n| - \epsilon, & \text{otherwise,} \end{cases} \quad (24)$$

where r_n is the true source range at sample n and ϵ defines a region on either side of the hyperplane. In SVR, the support vectors are points outside the ϵ region.

Because the SVM and SVR models are a two-class models, multi-class SVM with K classes is created by training $K(K-1)/2$ models on all possible pairs of classes. The points that are assigned to the same class most frequently are considered to comprise a single class, and so on until all points are assigned a class from 1 to K . This approach is known as the “one-versus-one” scheme,⁴¹ although slight modifications have been introduced to reduced computational complexity.⁴²

E. Random forests

The random forest (RF) (Ref. 43) classifier is a generalization of the decision tree model, which greedily segments the input data into a predefined number of regions. The simple decision tree model is made robust by randomly training subsets of the input data and averaging over multiple models in RF.

Consider a decision tree (see Fig. 3) trained on all the input data. Each input sample, \mathbf{x}_n , $n = 1, \dots, N$, represents a point in D dimensions. The input data can be partitioned into two regions by defining a cutoff along the i th dimension, where i is the same for all input samples \mathbf{x}_n , $n = 1, \dots, N$:

$$\begin{aligned} \mathbf{x}_n &\in \mathbf{x}_{\text{left}} & \text{if } x_{ni} > c, \\ \mathbf{x}_n &\in \mathbf{x}_{\text{right}} & \text{if } x_{ni} \leq c, \end{aligned} \quad (25)$$

where c is the cutoff value, and \mathbf{x}_{left} and $\mathbf{x}_{\text{right}}$ are the left and right regions, respectively. The cost function, G , that is minimized in the decision tree at each branch is³⁵

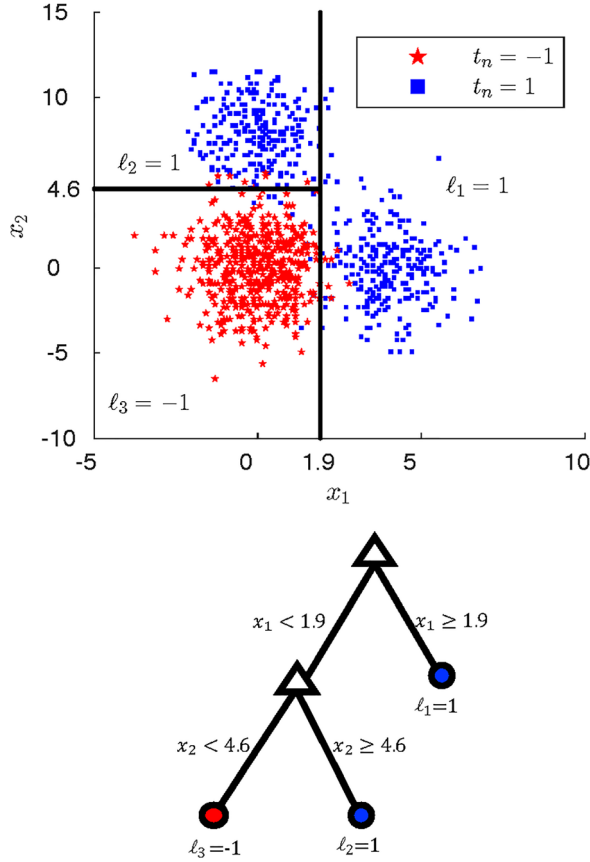


FIG. 3. (Color online) Decision tree classifier and corresponding rectangular regions shown for two-dimensional data with $K=2$ classes ($D=2$, $M=3$) and 1000 training points.

$$c^* = \underset{c}{\operatorname{argmin}} G(c),$$

$$G(c) = \frac{n_{\text{left}}}{N} H(\mathbf{x}_{\text{left}}) + \frac{n_{\text{right}}}{N} H(\mathbf{x}_{\text{right}}), \quad (26)$$

where n_{left} and n_{right} are the numbers of points in the regions \mathbf{x}_{left} and $\mathbf{x}_{\text{right}}$. $H(\cdot)$ is an impurity function chosen based on the problem.

For the classification problem, the Gini index³⁵ is chosen as the impurity function

$$H(\mathbf{x}_m) = \frac{1}{n_m} \sum_{\mathbf{x}_n \in \mathbf{x}_m} I(t_n, \ell_m) \left[1 - \frac{1}{n_m} I(t_n, \ell_m) \right], \quad (27)$$

where n_m is the number of points in region \mathbf{x}_m and ℓ_m represents the assigned label for each region, corresponding to the most common class in the region:³⁵

$$\ell_m = \underset{r_k}{\operatorname{argmax}} \sum_{\mathbf{x}_n \in \mathbf{x}_m} I(t_n, r_k). \quad (28)$$

In Eq. (28), r_k , $k=1, \dots, K$ are the source range classes and t_n is the label of point \mathbf{x}_n in region m , and

$$I(t_n, r_k) = \begin{cases} 1 & \text{if } t_n = r_k, \\ 0 & \text{otherwise.} \end{cases} \quad (29)$$

The remaining regions are partitioned iteratively until regions $\mathbf{x}_1, \dots, \mathbf{x}_M$ are defined. In this paper, the number of

regions, M , is determined by the minimum number of points allowed in a region. A diagram of the decision tree classifier is shown in Fig. 3. The samples are partitioned into $M=3$ regions with the cutoff values 1.9 and 4.6.

For RF regression, there are two differences from classification: the estimated class for each region is defined as the mean of the true class for all points in the region, and the mean squared error is used as the impurity function

$$\ell_m = \frac{1}{n_m} \sum_{\mathbf{x}_n \in \mathbf{x}_m} r_n,$$

$$H(\mathbf{x}_m) = \sum_{\mathbf{x}_n \in \mathbf{x}_m} (\ell_m - r_n)^2, \quad (30)$$

where r_n is source range at sample n .

As the decision tree model may overfit the data, statistical bootstrap and bagging are used to create a more robust model, a random forest.⁴⁴ In a given draw, the input data, \mathbf{x}_i , $i=1, \dots, Q$, is selected uniformly at random from the full training set, where $Q \leq N$. B such draws are conducted with replacement and a new decision tree is fitted to each subset of data. Each point, \mathbf{x}_n , is assigned to its most frequent class among all draws:

$$\hat{f}^{\text{bag}}(\mathbf{x}_n) = \underset{t_n}{\operatorname{argmax}} \sum_{b=1}^B I(\hat{f}^{\text{tree},b}(\mathbf{x}_n), t_n), \quad (31)$$

where $\hat{f}^{\text{tree},b}(\mathbf{x}_i)$ is the class of \mathbf{x}_i for the b th tree.

F. Performance metric

To quantify the prediction performance of the range estimation methods, the mean absolute percentage error (MAPE) over N samples is defined as

$$E_{\text{MAPE}} = \frac{100}{N} \sum_{i=1}^N \left| \frac{Rp_i - Rg_i}{Rg_i} \right|, \quad (32)$$

where Rp_i and Rg_i are the predicted range and the ground truth range, respectively. MAPE is preferred as an error measure because it accounts for the magnitude of error in faulty range estimates as well as the frequency of correct estimates. MAPE is known to be an asymmetric error measure⁴⁵ but is adequate for the small range of outputs considered.

G. Source localization algorithm

The localization problem solved by machine learning is implemented as follows:

- (1) Data preprocessing. The recorded pressure signals are Fourier transformed and N_s snapshots form the SCM from which the input \mathbf{x} is formed.
- (2) Division of preprocessed data into training and test data sets. For the training data, the labels are prepared based on different machine learning algorithms.
- (3) Training the machine learning models. $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_N]$ are used as the training input and the corresponding labels as the desired output.

- (4) Prediction on unlabeled data. The model parameters trained in step 3 are used to predict the source range for test data. The resulting output is mapped back to range, and the prediction error is reported by the mean absolute percentage error.

III. SIMULATIONS

In this section, the performance of machine learning on simulated data are discussed. For brevity, only the FNN classifier is examined here, although the conclusions apply to SVM and RF. Further discussion of SVM and RF performance is included in Secs. IV and V.

A. Environmental model and source-receiver configuration

Acoustic data are simulated using KRAKEN (Ref. 46) with environmental parameters simulating the Noise09 experiment,⁴⁷ see Fig. 4(a). The source frequency is 300 Hz. The source depth is 5 m in a 152 m waveguide, with a 24 m sediment layer (sound speed 1572–1593 m/s, density 1.76 g/cm³, attenuation coefficient 2.0 dB/λ) and a fluid halfspace bottom (sound speed 5200 m/s, density 1.8 g/cm³, and attenuation coefficient 2.0 dB/λ). The sound speed profile of water column is shown in Fig. 4(b). The vertical array consists of

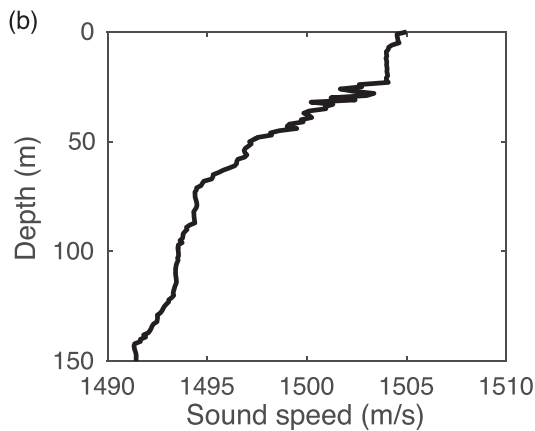
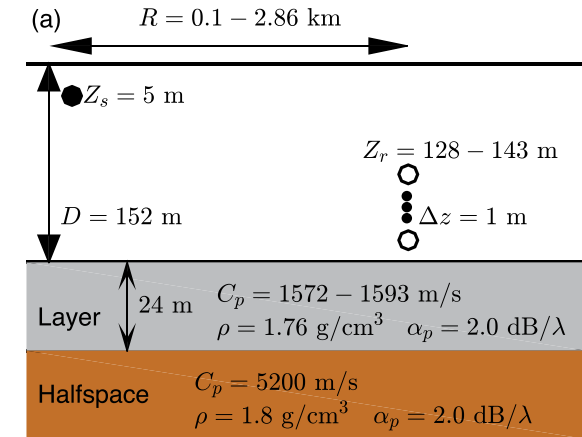


FIG. 4. (Color online) (a) Waveguide parameters and source-receiver configuration. (b) Sound speed profile of water column.

16 receivers spanning 128–143 m depth with inter-sensor spacing 1 m.

A source traveling away from the receiver at 2 m/s is simulated by varying the range from 0.1 to 2.86 km at 2 m intervals. Realizations with different SNRs are generated by adding appropriate complex Gaussian noise to the simulated received complex pressure signals.

Since the source moves in range and the source level is assumed constant, SNR is defined at the most distant range bin

$$\text{SNR} = 10 \log_{10} \frac{\sum_{l=1}^L |\hat{p}_l|^2 / L}{\sigma^2} \text{ (dB)}, \quad (33)$$

where \hat{p}_l is sound pressure signal received by the l th sensor at the longest source-receiver distance and σ^2 represents the noise variance.

B. Input preprocessing and learning parameters

The SCM for a 16–element vertical array is formed at each range point by averaging over $N_s = 10$ successive snapshots (nine snapshots overlapped) according to Eq. (3). The number of neurons in the input layer is therefore $D = 16 \times (16 + 1) = 272$. The range sample interval is 2 m, with 1380 total range samples (1 s duration per snapshot). Thus, a total of $N = 1380$ input matrices constitute the sample set spanning the whole range 0.1–2.86 km.

For each SNR, two realizations of noisy measurements are generated. One realization of size 1380×272 is used for the training set. For the test set, the range sample interval is changed to 20 m, and a realization of size 138×272 is used as input.

In the test set, $K = 138$ output neurons represent ranges from 0.1 to 2.86 km incremented by 20 m. The number of neurons in the hidden layer is $M = 128$. To prevent overfitting, the “keep dropout” technique,⁴⁰ with probability 0.5, is used. The initial learning rate for the Adam optimizer³⁹ is 0.01 and the maximum number of iterations is 1000.

C. Results

The prediction performance is examined for four SNRs (−10, −5, 0, 5 dB). Figure 5 compares range predictions by FNN and the true ranges on test data. For the four SNRs tested, the MAPE for the FNN predictions is 20.6%, 6.5%, 0.2%, and 0.0%, respectively.

As described in Sec. II, the output y_{nk} of FNN represents the probability distribution over a discrete set of possible ranges. To demonstrate the evolution of the probability distribution as the FNN is trained, y_{nk} versus training steps is plotted in Fig. 6 for the signal with SNR 5 dB at range 1.5 km. After 300 training steps, the FNN output probability distribution resembles the target output.

In Fig. 7, the convergence of the FNN algorithm is investigated by plotting the cross entropy Eq. (12) versus the optimization step on training and test data. It shows that the FNN converges after about 300 steps at all SNRs. For low

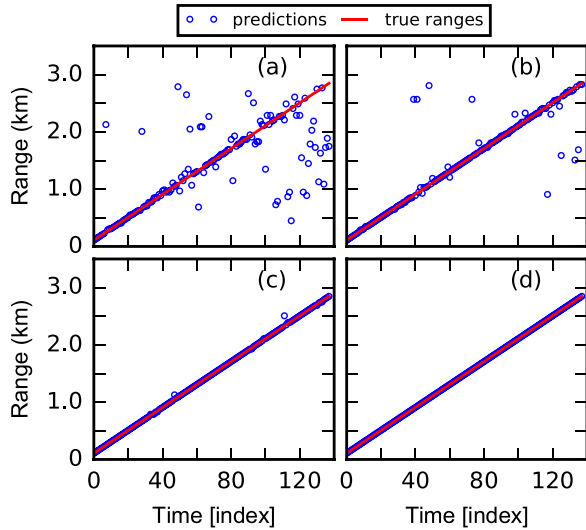


FIG. 5. (Color online) Range predictions by FNN on test data set with SNR of (a) -10 , (b) -5 , (c) 0 , and (d) 5 dB.

SNRs (<0 dB), the FNN classifier generates poor predictions on test data while performing well on training data, which indicates overfitting.

Increasing the training set size can reduce overfitting but additional data may be not available due to experimental or computational constraints. For higher SNRs (e.g., 0 and 5 dB), both test and training errors converge to low cross entropy, indicating good performance. Therefore, best performance of machine learning methods is expected for high SNR.

IV. EXPERIMENTAL RESULTS

Shipping noise data radiated by R/V New Horizon during the Noise09 experiment are used to demonstrate the

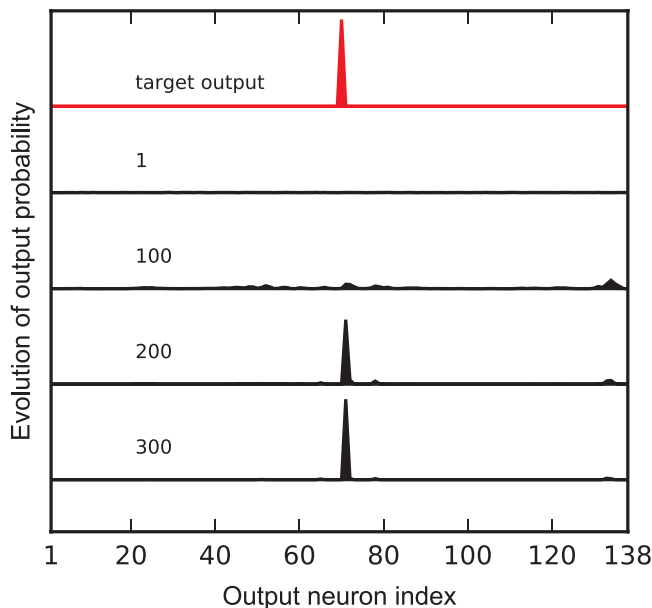


FIG. 6. (Color online) Output probability for range 0.1 – 2.86 km (the true range is 1.5 km) after training steps (1 , 100 , 200 , 300). The top line represents the target output.

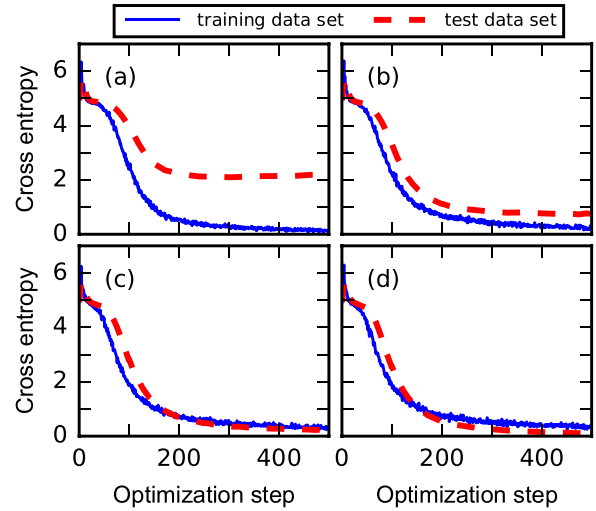


FIG. 7. (Color online) Cross entropy Eq. (11) versus optimization steps on training (solid) and test (dashed) data with SNR of (a) -10 , (b) -5 , (c) 0 , and (d) 5 dB.

performance of the FNN, SVM and RF localization. The experiment geometry is shown in Fig. 8, with bottom-moored vertical linear arrays (VLAs) indicated by triangles and the three ship tracks used for range estimation. The hydrophone sampling rate was 25 kHz.

The data from VLA2, consisting of 16 hydrophones at 1 m spacing, are used for range estimation. The frequency spectra of shipping noise recorded on the top hydrophone during the three periods are shown in Fig. 9. The striations indicate that the source was moving. The SNR decreases with increasing source-receiver distance.

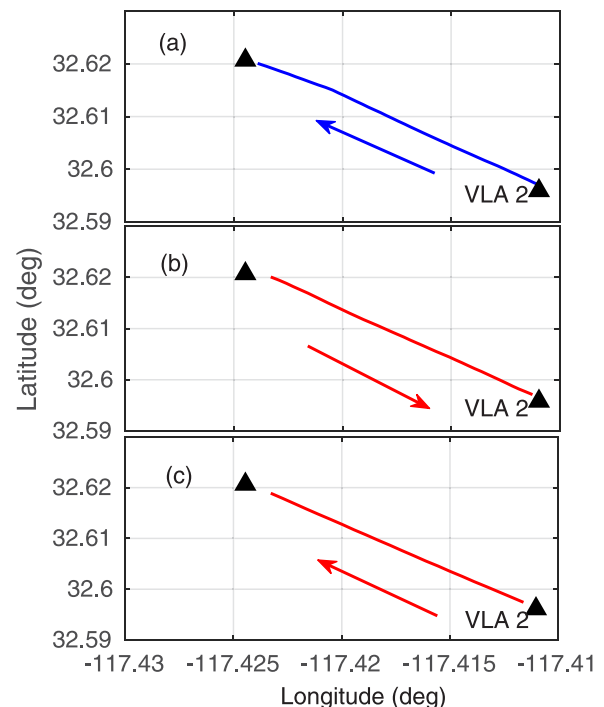


FIG. 8. (Color online) Ship tracks for Noise09 experiment during the periods (a) $01/2031/2009$, $01:43$ – $02:05$ (training data, ship speed 2 m/s), (b) $01/31/2009$, $01:05$ – $01:24$ (Test-Data-1, ship speed -2 m/s), and (c) $02/04/2009$, $13:41$ – $13:51$ (Test-Data-2, ship speed 4 m/s).

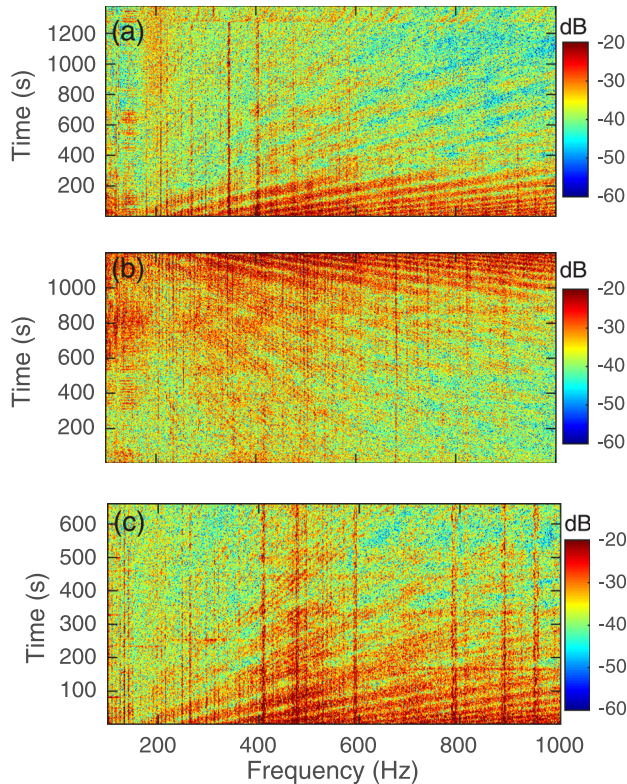


FIG. 9. (Color online) Spectra of shipping noise during periods (a) 01/31/2009, 01:43–02:05, (b) 01/31/2009, 01:05–01:24, and (c) 02/04/2009, 13:41–13:51.

Data from period 01:43–02:05 on January 31, 2009 are used as the training set and 01:05–01:24 on January 31 and 13:41–13:51 on February 4 are used as the test sets (Test-Data-1 and Test-Data-2).

The GPS antenna on the New Horizon is separated from the noise-generating propeller by a distance L_d . To account for this difference we use the range between the propeller and VLA2 as the ground truth range R_g :

$$R_g = \begin{cases} R_{GPS} - L_d & \text{for training data and Test-Data-2,} \\ R_{GPS} + L_d & \text{for Test-Data-1,} \end{cases} \quad (34)$$

where R_{GPS} represents the range between the GPS antenna and VLA2. According to the R/V New Horizon handbook, $L_d = 24.5$ m. In the following, the ranges have been corrected by Eq. (34).

A. Input preprocessing and learning parameters

For the training set and both test sets, the 16×16 SCM at each range and frequency, averaged over 10 successive 1-s snapshots, is used as input. There are 1380 samples in the training data set and 120 samples in each of the test data sets (samples are drawn every 10 s for Test-Data-1 and 5 s for Test-Data-2). The source-receiver range 0.1–3 km is divided into $K = 138$ discrete range points.

As in the simulations in Sec. III B, the keep probability for training dropout of the FNN is 0.5, the initial learning rate is 0.01 and the maximum iteration step is 1000. The number

of neurons in the hidden layer is chosen as $M = 128$ for 1 frequency and $M = 1024$ for 66 frequencies (see Sec. IV B).

For the SVM classifier, Gaussian radial basis function kernel is used. The parameters γ [Eq. (23)] and C [Eq. (20)] were tested over $[10^{-3} \ 10^{-1}]$ and $[10 \ 10^3]$, respectively. Values of $\gamma = 10^{-2}$ and $C = 10$ are found to be optimal.

For the RF method, the number of trees bagged is 500, with a minimum of 50 samples required for each leaf.

The performance of all test cases for the FNN, SVM, RF, and conventional MFP is summarized in Tables I and II.

B. SCM inputs

Because the shipping noise has a wide frequency band as seen from Fig. 9, the performance of the machine learning with the single and multi-frequency inputs is investigated. The FNN classifier is again used an example to illustrate the benefit of using multiple frequencies.

Input SCMs are formed at 550, 950, and 300–950 Hz with 10 Hz increments (66 frequencies). For the multi-frequency input, the SCMs are formed by concatenating multiple single-frequency SCM input vectors. For example, the dimension of a single frequency input sample is 272, whereas the multi-frequency input has a dimension $272 \times N_f$ for N_f frequencies. The FNN is trained separately for each case and the source-receiver range is then predicted at the selected frequencies.

The prediction results on the two test data sets are shown in Figs. 10(a)–10(f) along with R_g . For single frequency inputs, the minimum error is 12% [Fig. 10(d)] at 550 Hz and the largest error is 18% at 950 Hz [Fig. 10(e)], both on Test-Data-2. For multi-frequency inputs, the prediction error is 8% on Test-Data-1 and 6% on Test-Data-2, indicating a performance improvement by using multiple frequencies. In general, the FNN predictions are better at close ranges due to higher SNR, as expected from the simulation results. However, the FNN with multi-frequency inputs performs well regardless of source range.

C. Source localization as a classification problem

Source localization is first solved as a classification problem. Only the best MAPE obtained by FNN (Sec. II C), SVM (Sec. II D) and RF (Sec. II E) is shown here (Fig. 11). These results are summarized in Table I.

The lowest MAPE is achieved by the SVM, with 2% on both data sets. RF also reaches 2% MAPE for Test-Data-2

TABLE I. Best MAPE rate of FNN, SVM, RF, and MFP predictions.

Model	MAPE	
	Test-Data-1 (%)	Test-Data-2 (%)
FNN classifier	3	3
SVM classifier	2	2
RF classifier	3	2
FNN regressor	10	5
SVM regressor	42	59
RF regressor	55	48
MFP	55	36
MFP with measured replica	19	30

TABLE II. Parameter sensitivity of FNN, SVM, and RF classifiers.

(a): FNN classifier					
No. of hidden layers	No. of hidden neurons	No. of classes	No. of snapshots	MAPE	
				Test-Data-1 (%)	Test-Data-2 (%)
1	1024	1380	10	7	5
1	1024	690	10	3	6
1	1024	276	10	6	8
1	1024	138	10	8	6
1	1024	56	10	7	4
1	1024	28	10	10	4
1	1024	14	10	16	7
1	1024	138	1	10	5
1	1024	138	5	6	3
1	1024	138	20	8	3
1	64	138	10	9	9
1	128	138	10	7	7
1	256	138	10	8	6
1	512	138	10	8	4
1	2048	138	10	7	5
2	128	138	10	9	8
2	256	138	10	9	9
2	512	138	10	6	8

(b): SVM classifier

γ	C	No. of classes	No. of snapshots	MAPE	
				Test-Data-1 (%)	Test-Data-2 (%)
10^{-2}	10	1380	10	2	3
		690	10	2	3
		276	10	4	3
		138	10	2	2
		56	10	3	3
		28	10	5	3
		138	1	17	5
		138	5	2	3
		138	20	3	2

(c): RF classifier

No. of trees	No. of samples per leaf	No. of classes	No. of snapshots	MAPE	
				Test-Data-1 (%)	Test-Data-2 (%)
500	50	1380	10	4	10
		690	10	3	4
		276	10	3	3
		138	10	3	2
		56	10	9	5
		28	10	13	9
		138	1	20	15
138	5	6	5		
138	20	3	2		

and 3% for Test-Data-1. FNN has 3% MAPE for both test sets. The performance of these three machine learning algorithms is comparable when solving range estimation as a classification problem.

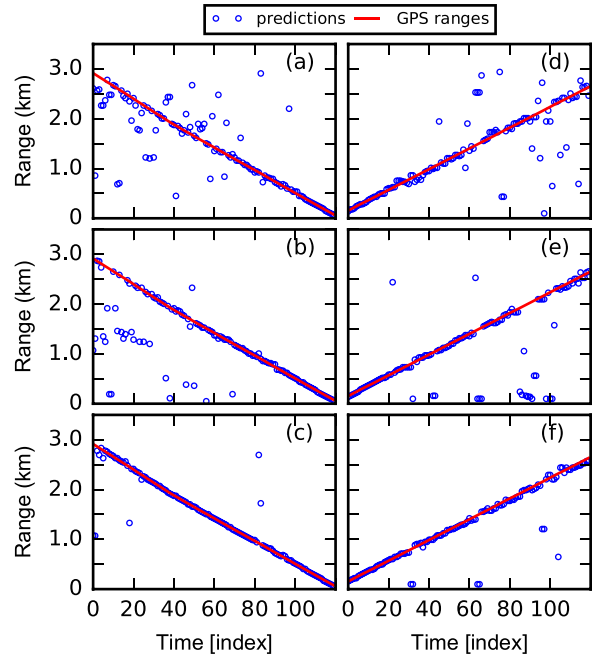


FIG. 10. (Color online) Range predictions on Test-Data-1 (a, b, c) and Test-Data-2 (d, e, f) by FNN. (a), (d) 550 Hz, (b), (e) 950 Hz, (c), (f) 300–950 Hz with 10 Hz increment, i.e., 66 frequencies. The time index increment is 10 s for Test-Data-1, and 5 s for Test-Data-2.

The performance of these machine learning algorithms with various parameters (e.g., number of classes, number of snapshots and model hyper-parameters) is examined in Sec. V.

D. Source localization as a regression problem

Source localization can be solved as a regression problem. For this problem, the output represents the continuous range

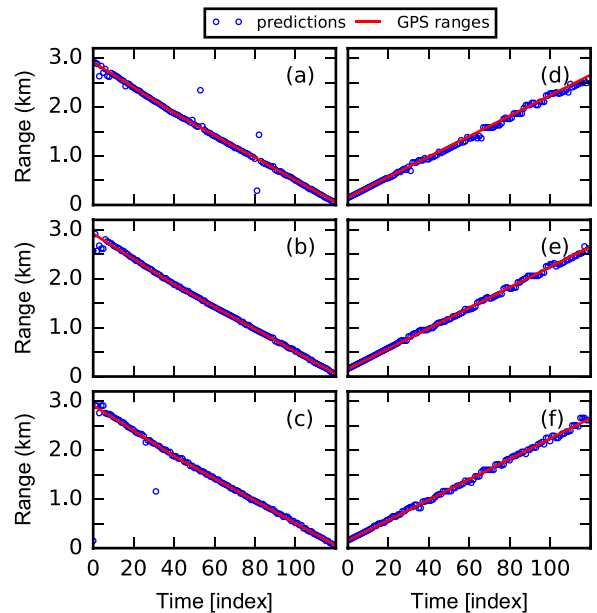


FIG. 11. (Color online) Source localization as a classification problem. Range predictions on Test-Data-1 (a, b, c) and Test-Data-2 (d, e, f) by FNN, SVM and RF for 300–950 Hz with 10 Hz increment, i.e., 66 frequencies. (a),(d) FNN classifier, (b),(e) SVM classifier, (c),(f) RF classifier.

and the input remains the vectorized covariance matrix Eq. (3). In this case, the highly nonlinear relationship between the covariance matrix elements and the range, caused primarily by modal interference, is difficult to fit a nonlinear model. In the training process, the input data remain the same, the labels are direct GPS ranges, and the weights and biases are trained using least-squares objective functions.

The range predictions by FNN with different number of hidden layers along with the GPS ranges are given in Fig. 12. Increasing the number of hidden layers increases the nonlinearity of the model and allows a larger number of parameters to be learned, thus significantly reducing the error for FNN regression. Figure 13 shows the results of SVM [Figs. 13(a) and 13(c)] and RF regressors [Figs. 13(b) and 13(d)] on two data sets. For these methods, since additional layers cannot be added to increase the nonlinearity, the performance lags FNN. The best MAPE values for each regressor are shown in Table I. Compared with classifiers, the FNN, SVM and RF degrade significantly for solving regression tasks.

E. Conventional matched-field processing

The Bartlett MFP¹ is applied to Noise09 data for comparison. Two kinds of replica fields are used in the Bartlett processor. The first is generated by KRAKEN using the Noise09 environment in Fig. 4, with the corresponding ambiguity surfaces and maximum peaks shown in Fig. 14. We use the measured data (i.e., training data, 01/31/2009, 01:43–02:05) as the second group of replica fields as proposed in Ref. 14. The results are shown in Fig. 15. For each case, both single frequency (550 Hz) and broadband (300–950 Hz) are considered.

From Figs. 14 and 15, the Bartlett MFP fails to determine source positions using a single frequency, while the FNN still

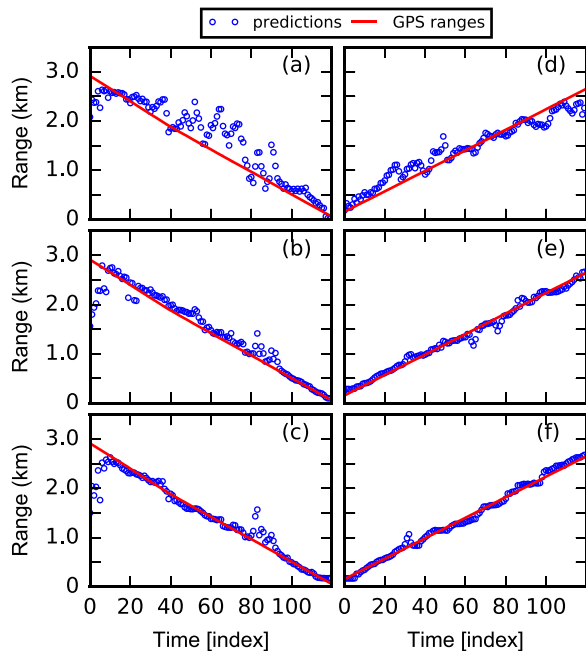


FIG. 12. (Color online) Source localization as a regression problem. Range predictions on Test-Data-1 (a, b, c) and Test-Data-2 (d, e, f) by FNN for 300–950 Hz with 10 Hz increment, i.e., 66 frequencies. (a),(d) 1 hidden layer, (b),(e) 2 hidden layers, (c),(f) 3 hidden layers. Each hidden layer consists of 512 neurons.

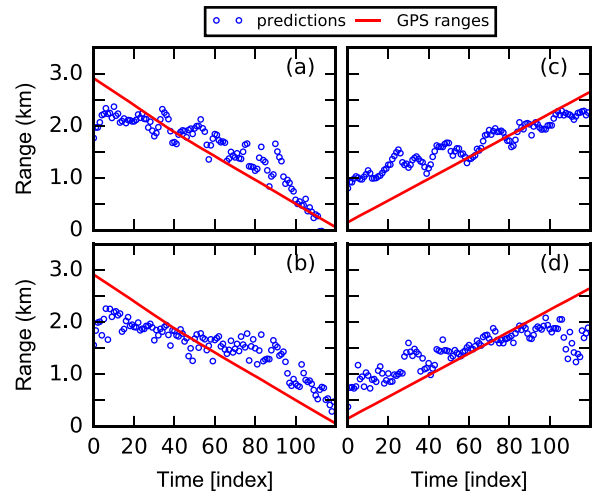


FIG. 13. (Color online) Source localization as a regression problem. Range predictions on Test-Data-1 (a, b) and Test-Data-2 (c, d) by SVM and RF for 300–950 Hz with 10 Hz increment, i.e., 66 frequencies. (a), (c) SVM for regression, (b),(d) RF for regression.

generates a number of reasonable predictions [see Fig. 10(a)]. Despite improved performance using broadband MFP, there are some errors due to sidelobes. The MAPE of MFP predictions is shown in Table I. The minimum MAPE of Bartlett MFP is 19% on Test-Data-1 and 30% on Test-Data-2, which is much larger than the machine learning classifiers.

V. DISCUSSIONS

A. Range resolution

The number of classes, corresponding to the resolution of range steps, was varied to determine its effect on range estimation results. Previously (see Sec. IV) $K = 138$ classes were used, corresponding to a range resolution of 20 m. The MAPE for predictions with different numbers of output classes by FNN, SVM, and RF classifiers is given in Table II with 10 snapshots averaged for each sample. These three classifiers perform well for all tested range resolutions.

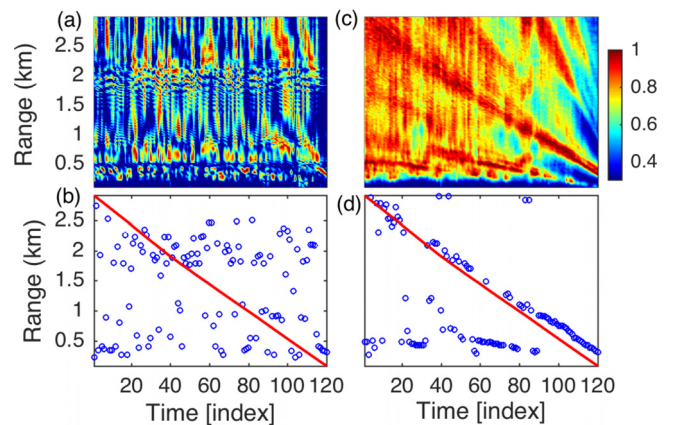


FIG. 14. (Color online) Localization using Bartlett matched-field processing based on synthetic replica fields. (a) Ambiguity surface and (b) maximum peaks for 550 Hz, (c) ambiguity surface and (d) maximum peaks for 300–950 Hz with 10 Hz increment. Circles and solid lines denote predictions and GPS ranges, respectively.

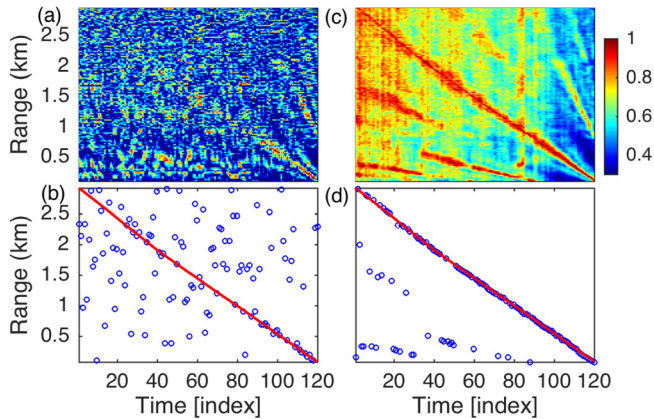


FIG. 15. (Color online) Localization using Bartlett matched-field processing (training data as replica fields). (a) Ambiguity surface and (b) maximum peaks for 550 Hz, (c) ambiguity surface and (d) maximum peaks for 300–950 Hz with 10 Hz increment. Circles and solid lines denote predictions and GPS ranges, respectively.

B. Snapshots

The number of snapshots averaged to create the SCMs may also affect the performance. Increasing the number of snapshots makes the input more robust to noise, but could introduce mismatch if the source is moving or the environment is evolving across the averaging period. The range estimation methods are tested using 1, 5, and 20 snapshots and the corresponding MAPE is shown in Table II. All of the three models degrade with 1 snapshot due to low SNR and become robust with more snapshots.

C. Number of hidden neurons and layers for FNN

The MAPE of FNN with different numbers of hidden neurons and layers is given in Table II. Increasing the number of hidden neurons increases the number of parameters to fit in the FNN model. As a result, more of the variance in the data is captured. FNN has the minimum error when the number of hidden neurons is chosen as 128 or 2048 for Test-Data-1 (7%) and 512 for Test-Data-2 (4%). In the case of Test-Data-1, 256, 512, and 1024 have a similar result (8%, 8%, 8%). As shown in Table II, the FNN with two hidden layers did not improve the prediction performance for classification.

D. Kernel and regularization parameters for SVM

When using a Gaussian radial basis function kernel, the parameters γ in Eq. (23) and the regularization parameter C in Eq. (20) determine the best separation of the data by SVM. The MAPE versus these two parameters on two data sets is shown in Fig. 16. As seen from the result, there exists an optimal interval for these two parameters (i.e., $10 < C < 10^3$ and $10^{-3} < \gamma < 10^{-1}$). The SVM fails when γ and C are out of this interval, but is robust when γ and C are within the appropriate range.

E. Number of trees and minimum samples per leaf for RF

The number of decision trees and the minimum samples per leaf³⁵ are the most sensitive parameters for the RF.

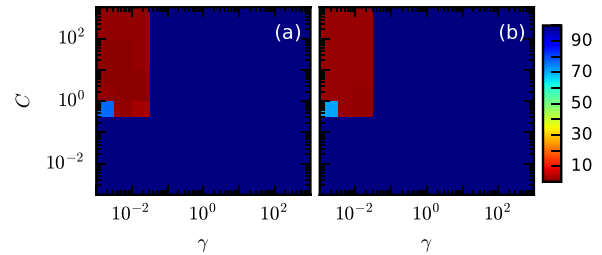


FIG. 16. (Color online) MAPE of the SVM classifier on (a) Test-Data-1 and (b) Test-Data-2. γ is the kernel parameter and C is the regularization parameter. There are 138 output classes and 10 snapshots averaged for each sample.

Figure 17 shows the MAPE versus these two parameters. The RF parameters have a smaller range of possible values than SVM, but the RF classifier will not fail for any of these choices. The RF classifier has the best performance for more than 500 trees and 20 to 50 minimum samples per leaf.

F. CPU time

The three machine learning models are efficient in CPU time (on a 2015–iMac). For the FNN with 1024 hidden neurons and 1000 training steps, it took 140 s for training and predicting the results. The SVM took 110 s, and the RF (500 trees and 50 minimum samples per leaf) was the fastest at 52 s.

G. Multiple sources and deep learning

In our study, only one source is considered. The simultaneous multiple source localization problem is more challenging, especially for sources close to each other. Solving this problem with FNN is a multiple binary classification problem and will require additional training data.

Although the FNN with one hidden layer works well for the data sets in this paper, more complicated machine learning algorithms, e.g., deep learning, may be necessary for more complicated experimental geometries or ocean environments.

VI. CONCLUSION

This paper presents an approach for source localization in ocean waveguides within a machine learning framework. The localization is posed as a supervised learning problem and solved by the feed-forward neural networks, support vector machines and random forests separately. Taking advantage of the modern machine learning library such as

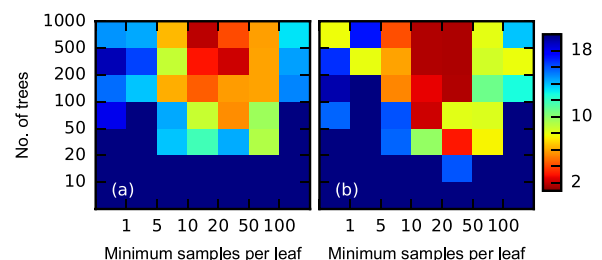


FIG. 17. (Color online) MAPE of the RF classifier versus the number of trees and the minimum samples per leaf on (a) Test-Data-1 and (b) Test-Data-2. There are 138 output classes and 10 snapshots averaged for each sample.

TensorFlow and Scikit-learn, the machine learning models are trained efficiently. Normalized sample covariance matrices are fed as input to the models. Simulations show that FNN achieves a good prediction performance for signals with SNR above 0 dB even with deficient training samples. Noise09 experimental data further demonstrates the validity of the machine learning algorithms.

Our results show that classification methods perform better than regression and MFP methods. The three classification methods tested (FNN, SVM, RF) all performed well with the best MAPE 2%–3%. The SVM performs well if the model parameters are chosen appropriately, otherwise it degrades significantly. By contrast, the FNN and RF are more robust on the choices of parameters.

The experimental results show that multi-frequency input generates more accurate predictions than single frequency (based on FNN). In the current study, the training and test data were from the same ship. In a realistic application, data from multiple ships of opportunity can be used as training data by taking advantage of the Automatic Identification System (AIS), a GPS system required on all cargo carriers. The tracks were quite similar and it would be interesting to compare the performance of the three methods as tracks deviate.

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research under Grant No. N00014–1110439 and the China Scholarship Council.

¹F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, 2nd ed. (Springer Science & Business Media, New York, 2011), Chap. 10.

²H. P. Buckner, “Use of calculated sound fields and matched field detection to locate sound source in shallow water,” *J. Acoust. Soc. Am.* **59**, 368–373 (1976).

³H. Schmidt, A. B. Baggeroer, W. A. Kuperman, and E. K. Scheer, “Environmentally tolerant beamforming for high-resolution matched field processing: Deterministic mismatch,” *J. Acoust. Soc. Am.* **88**, 1851–1862 (1990).

⁴M. D. Collins and W. A. Kuperman, “Focalization: Environmental focusing and source localization,” *J. Acoust. Soc. Am.* **90**, 1410–1422 (1991).

⁵E. K. Westwood, “Broadband matched-field source localization,” *J. Acoust. Soc. Am.* **91**, 2777–2789 (1992).

⁶A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, “An overview of matched field methods in ocean acoustics,” *IEEE J. Ocean. Eng.* **18**, 401–424 (1993).

⁷D. F. Gingras and P. Gerstoft, “Inversion for geometric and geoacoustic parameters in shallow water: Experimental results,” *J. Acoust. Soc. Am.* **97**, 3589–3598 (1995).

⁸Z. H. Michalopoulou and M. B. Porter, “Matched-field processing for broad-band source localization,” *IEEE J. Ocean. Eng.* **21**, 384–392 (1996).

⁹C. F. Mecklenbraüker and P. Gerstoft, “Objective functions for ocean acoustic inversion derived by likelihood methods,” *J. Comput. Acoust.* **8**, 259–270 (2000).

¹⁰C. Debever and W. A. Kuperman, “Robust matched-field processing using a coherent broadband white noise constraint processor,” *J. Acoust. Soc. Am.* **122**, 1979–1986 (2007).

¹¹S. E. Dosso and M. J. Wilmut, “Bayesian multiple source localization in an uncertain environment,” *J. Acoust. Soc. Am.* **129**, 3577–3589 (2011).

¹²S. E. Dosso and M. J. Wilmut, “Maximum-likelihood and other processors for incoherent and coherent matched-field localization,” *J. Acoust. Soc. Am.* **132**, 2273–2285 (2012).

¹³S. E. Dosso and M. J. Wilmut, “Bayesian tracking of multiple acoustic sources in an uncertain ocean environment,” *J. Acoust. Soc. Am.* **133**, EL274–EL280 (2013).

¹⁴L. T. Fialkowski, M. D. Collins, W. A. Kuperman, J. S. Perkins, L. J. Kelly, A. Larsson, J. A. Fawcett, and L. H. Hall, “Matched-field processing using measured replica fields,” *J. Acoust. Soc. Am.* **107**, 739–746 (2000).

¹⁵C. M. A. Verlinden, J. Sarkar, W. S. Hodgkiss, W. A. Kuperman, and K. G. Sabra, “Passive acoustic source localization using sources of opportunity,” *J. Acoust. Soc. Am.* **138**, EL54–EL59 (2015).

¹⁶Z. Lei, K. Yang, and Y. Ma, “Passive localization in the deep ocean based on cross-correlation function matching,” *J. Acoust. Soc. Am.* **139**, EL196–EL201 (2016).

¹⁷H. C. Song and Chomgun Cho, “Array invariant-based source localization in shallow water using a sparse vertical array,” *J. Acoust. Soc. Am.* **141**, 183–188 (2017).

¹⁸R. Lefort, G. Real, and A. Drémeau, “Direct regressions for underwater acoustic source localization in fluctuating oceans,” *Appl. Acoust.* **116**, 303–310 (2017).

¹⁹G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, “Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,” *IEEE Signal Proc. Mag.* **29**, 82–97 (2012).

²⁰A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.* 1097–1105 (2012).

²¹R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Mach. Learn. Res.* **12**, 2493–2537 (2011).

²²M. L. Sharma and M. K. Arora, “Prediction of seismicity cycles in the himalayas using artificial neural networks,” *Acta Geophys. Polonica* **53**, 299–309 (2005).

²³S. R. García, M. P. Romo, and J. M. Mayoral, “Estimation of peak ground accelerations for Mexican subduction zone earthquakes using neural networks,” *Geofis. Int.* **46**, 51–63 (2006).

²⁴A. Köhler, M. Ohrnberger, and F. Scherbaum, “Unsupervised pattern recognition in continuous seismic wavefield records using Self-Organizing Maps,” *Geophys. J. Int.* **182**, 1619–1630 (2010).

²⁵N. Riahi and P. Gerstoft, “Using graph clustering to locate sources within a dense sensor array,” *Signal Process.* **132**, 110–120 (2017).

²⁶B. Z. Steinberg, M. J. Beran, S. H. Chin, and J. H. Howard, “A neural network approach to source localization,” *J. Acoust. Soc. Am.* **90**, 2081–2090 (1991).

²⁷J. M. Ozard, P. Zakarauskas, and P. Ko, “An artificial neural network for range and depth discrimination in matched field processing,” *J. Acoust. Soc. Am.* **90**, 2658–2663 (1991).

²⁸A. Caiti and T. Parisini, “Mapping ocean sediments by RBF networks,” *IEEE J. Ocean. Eng.* **19**, 577–582 (1994).

²⁹A. Caiti and S. Jesus, “Acoustic estimation of seafloor parameters: A radial basis functions approach,” *J. Acoust. Soc. Am.* **100**, 1473–1481 (1996).

³⁰Y. Stephan, X. Demoulin, and O. Sarzeaud, “Neural direct approaches for geoacoustic inversion,” *J. Comput. Acoust.* **6**, 151–166 (1998).

³¹J. Benson, N. R. Chapman, and A. Antoniou, “Geoacoustic model inversion using artificial neural networks,” *Inverse Problems* **16**, 1627–1639 (2000).

³²Z. H. Michalopoulou, D. Alexandrou, and C. Moustier, “Application of neural and statistical classifiers to the problem of seafloor characterization,” *IEEE J. Ocean. Eng.* **20**, 190–197 (1995).

³³A. M. Thode, K. H. Kim, S. B. Blackwell, C. R. Greene, C. S. Nations, T. L. McDonald, and A. M. Macrander, “Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys,” *J. Acoust. Soc. Am.* **131**, 3726–3747 (2012).

³⁴M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, “TensorFlow: Large-scale machine learning on heterogeneous distributed systems,” Software available from [tensorflow.org](https://www.tensorflow.org) (2015).

³⁵F. 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 É. Duchesnay, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

- ³⁶J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, D. Warde-Farley, and Y. Bengio, "Theano: A CPU and GPU math compiler in Python," Proc. 9th Python in Science Conf. (2010), pp. 1–7.
- ³⁷Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of 22nd ACM international conference on Multimedia* (2014), pp. 675–678.
- ³⁸R. Collobert, S. Bengio, and J. Mariétoz, "Torch: A modular machine learning software library," No. EPFL-REPORT-82802, Idiap (2002).
- ³⁹D. Kingma and B. Jimmy, "Adam: A method for stochastic optimization," arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014).
- ⁴⁰N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
- ⁴¹C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006), Chaps. 4, 5, and 7.
- ⁴²R. Fan, P. Chen, and C. Lin, "Working set selection using second order information for training support vector machines," *J. Mach. Learn. Res.* **6**, 1889–1918 (2005).
- ⁴³L. Breiman, "Random forests," *Mach. Learn.* **45**, 5–32 (2001).
- ⁴⁴L. Breiman, "Bagging predictors," *Mach. Learn.* **24**, 123–140 (1996).
- ⁴⁵P. Goodwin and R. Lawton, "On the asymmetry of the symmetric MAPE," *Int. J. Forecasting* **15**, 405–408 (1999).
- ⁴⁶M. B. Porter, The KRAKEN Normal Mode Program, <http://oalib.hlsresearch.com/Modes/AcousticsToolbox/manualtml/kraken.html> (Last viewed 11/1/2009).
- ⁴⁷S. H. Byun, C. M. A. Verlinden, and K. G. Sabra, "Blind deconvolution of shipping sources in an ocean waveguide," *J. Acoust. Soc. Am.* **141**, 797–807 (2017).