Supervised Classification of Sound Speed Profiles Via Dictionary Learning

Jhon A. Castro-Correa*

Department of Electrical and Computer Engineering, University of Delaware, Newark, DE USA 19716

Stephanie A. Arnett

Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602 USA

Tracianne B. Neilsen

Department of Physics and Astronomy, Brigham Young University, Provo, UT 84602 USA

Lin Wan

Department of Electrical and Computer Engineering, University of Delaware, Newark, DE USA 19716

Mohsen Badiey

Department of Electrical and Computer Engineering, University of Delaware, Newark, DE USA 19716

*Corresponding author: Jhon A. Castro-Correa, jcastro@udel.edu

Early Online Release: This preliminary version has been accepted for publication in *Journal of Atmospheric and Oceanic Technology* cited, and has been assigned DOI 10.1175/JTECH-D-21-0090.1.The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT

The presence of internal waves (IWs) in the ocean alters the isotropic properties of sound speed profiles (SSPs) in the water column. Changes in the SSPs affect underwater acoustics since most of the energy is dissipated into the seabed due to the downward refraction of sound waves. In consequence, variations in the SSP must be considered when modeling acoustic propagation in the ocean. Regularly, empirical orthogonal functions (EOFs) are employed to model and represent SSPs using a linear combination of basis functions that capture the sound speed variability. A different approach is to use dictionary learning (DL) to obtain a learned dictionary (LD) that generates a non-orthogonal set of basis functions (atoms) that generate a better sparse representation. In this paper, the performance of EOFs and LDs are evaluated for sparse representation of SSPs affected by the passing of IWs. In addition, an LD-based supervised framework is presented for SSP classification and is compared with classical learning models. The algorithms presented in this work are trained and tested on data collected from the shallow water experiment 2006. Results show that LDs yield lower reconstruction error than EOFs when using the same number of basis. In addition, overcomplete LDs demonstrate to be a robust method to classify SSPs during low, medium, and high IW activity, reporting comparable and sometimes higher accuracy than standard supervised classification methods.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

1. Introduction

Internal waves (IWs) are caused by differences in the temperature of water in the ocean. The presence of such anomalies in the seawater creates time-dependent and spatial variations in the sound speed profiles (SSPs) and affects underwater sound propagation due to the downward refraction of sound waves (Katsnelson et al. 2021; Rouseff 2001). Usually, sparse representations of SSPs via empirical orthogonal functions (EOFs) are used to support inversion algorithms for sound speed (North 1984). However, effective sparse representation of SSPs can be compromised due to high perturbations in the water column. Recently, Bianco and Gerstoft (2017) have shown that dictionary learning (DL), an unsupervised machine learning method, is better suited to sparsely represent SSPs. In this paper, DL and EOFs are used to model and classify measured SSPs affected by IWs during the Shallow Water Experiment 2006 (SW06).

Internal waves can be thought of as 4-dimensional phenomena due to their effects in both 3D spatial (x, y, z) and temporal (t) domains. Characterization of the behavior and statistical properties of IWs have been carried out using 3D mapping techniques (Badiey et al. 2013, 2016), resulting in the general understanding of the regimens in the propagation of IWs. Several studies have shown that internal waves create significant variability in the speed of sound, affecting how sound propagates through the ocean, due to the fluctuations in the acoustic modal behavior (Flatté and Tappert 1975; Rouseff 2001; Helfrich and Melville 2006). The variation due to the passing of internal wave packets affects the propagation and reception of acoustic signals underwater because of drastic changes in the acoustic channel (Huang et al. 2008).

EOF analysis has been commonly used to represent SSPs as a linear combination of few orthogonal basis functions that describe the statistics of the sound speed uncertainty (Xu and Schmidt 2006; Abiva et al. 2019). Those resulting sparse representations are employed to aid inversion

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

procedures (Huang et al. 2008), However, strong variations in the SSPs occurring due to the passing of IWs yield a dramatic decay in the reconstruction accuracy of SSPs using EOFs (Sun and Zhao 2020; Roundy 2015). As a result, different approaches such as 3D dimensional modeling (Badiey et al. 2013) or learning methods (Jain and Ali 2006; Bianco and Gerstoft 2017; Sun and Zhao 2020) have been studied for SSPs modeling and reconstruction.

Dictionary learning, an unsupervised learning method, aims to find a set of non-orthogonal basis functions (referred to as atoms) that can sparsely reconstruct signals (Zhang et al. 2015). DL framework has been extensively applied to dimensionality reduction (Tošić and Frossard 2011), pattern recognition (Wright et al. 2010) and sparse representation modeling (Rubinstein et al. 2010). Recent studies have been conducted to sparsely represent SSPs using learned dictionaries (LDs) (Kuo and Kiang 2020). Bianco and Gerstoft (2017) showed LDs are well-suited to generate sparse representations of SSPs using few basis functions. Sun and Zhao (2020) tested the effectiveness of LDs using HYCOM data and supported the results found by Bianco and Gerstoft (2017), concluding that non-orthogonal atoms allow for more flexible dictionaries and produce better sparse representations of SSPs.

Due to the relaxation of the orthogonal requirements and the possibility of generating optimal sparse representations, LDs have also been applied to clustering (Sprechmann and Sapiro 2010) and classification tasks (Tang et al. 2019; Suo et al. 2014). In this approach, specific dictionaries are trained to retain most of the meaningful information of each class. Then, testing data are classified by selecting the dictionary yielding the sparse representation that generates the lowest error (Ramirez et al. 2010; Zhao et al. 2018). Here, classification via dictionary learning is extended to label data containing SSPs affected by low, medium, and high IW activity collected during the SW06 experiment.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

In this paper, there are two main contributions. First, a comparison of the ability to sparsely model SSPs with few basis functions is made between EOFs, complete LDs and overcomplete LDs. Second, a DL-based SSPs classification setting is proposed and assessed via LD. The proposed framework is compared with standard classification algorithms such as support vector machine (SVM) and *k*-nearest neighbors (KNN) classifier. The dictionary atoms are calculated with online dictionary learning, a stochastic gradient-descent approach introduced by Mairal et al. (2009), while sparse coding is performed using the orthogonal matching pursuit (OMP) algorithm and Lasso convex optimization. This paper is structured as follows, in Sec. 2 preliminary concepts and notations are presented. Section 3 introduces the data collected during the SW06 experiment used in this paper. Both EOF analysis and dictionary learning frameworks are introduced in Sec. 4, while results for modeling and classification of SSPs are shown in Sec. 5, followed by the conclusions in Sec. 6.

2. Preliminaries and notation

In this paper matrices are denoted in upper case bold, vectors in lower case bold, and scalars in lower case italic. The ℓ_p-norm of a vector or matrix is represented as $\|\cdot\|_p$, with $1 \leq p < \infty$. For any $\mathbf{X}$, if $p = 2$, $\|\cdot\|_p$ takes the name of Frobenius norm $\|\cdot\|_{\mathcal{F}}$. Notice, these p-norms are entry-wise rather than the induced norms.

Any square diagonalizable matrix $\mathbf{X} \in \mathbb{R}^{n\times n}$ can be factorized into its canonical form via eigen-decomposition satisfying the linear equation $\mathbf{X} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$. Where $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_n] \in \mathbb{R}^{n\times n}$ is the matrix containing eigenvectors and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots \lambda_n) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with the set of eigenvalues. Similarly, any rectangular diagonizable matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ accepts the factorization $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ via singular value decomposition (SVD) (Stewart 1993). Where $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ are the matrices of left and right singular vectors, respectively, and $\mathbf{\Sigma} \in \mathbb{R}^{m\times n}$ contains

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

the singular values. For the cases $\mathbf{X}\mathbf{X}^T$ and $\mathbf{X}^T\mathbf{X}$, the diagonal matrix of the non-zero eigenvalues is related to the singular values of $\mathbf{X}$, such that $\mathbf{\Lambda} = \mathbf{\Sigma}^2 = \mathbf{\Sigma}\mathbf{\Sigma}^T = \mathbf{\Sigma}^T\mathbf{\Sigma}$.

3. Experimental data

The raw data introduced in this work were collected during the Shallow Water acoustic and oceanographic Experiment 2006 (SW06) performed off the coast of New Jersey from mid-July to mid-September in 2006 (Tang et al. 2007; Newhall et al. 2007). During the experiment, 62 acoustic and oceanographic moorings were deployed in a "T" geometry as sketched in Fig. 1. This "T" mooring conformation measured data on an almost constant bathymetry of 80 m in the along-shelf, and a bathymetry across-shelf starting at 600 m going shoreward to 60 m depth. The intersection of the two paths in the "T" was populated with a cluster of 16 moorings to measure the 3D environment. Most of the environmental moorings in the area consisted of temperature, conductivity, and pressure sensors that measured the physical oceanography in the water column.

This paper uses data from mooring SW30, deployed at 39° 01.501' N, 73° 04.007' W, to study the time-evolving SSPs via dictionary learning. The SW30 station was part of the 16 mooring cluster located at the intersection of the "T" geometry deployed in the SW06 experiment. The location of mooring SW30 inside the cluster is marked with a white star in Fig. 1. The SW30 station had 11 unevenly spaced sensors collecting conductivity and temperature profiles from 14 to 83.3 m in a water column with seafloor at 86 m depth. For the present study, temperature, conductivity, and pressure data were extracted from 01 Aug 00:00:00 to 05 Sep 16:00:00 UTC 2006 at SW30 location.

During the SW06 experiment, IW activities were reported initiating at the shelf break and propagating toward the shore after 17 Aug 2006 (Badiey et al. 2013). The transition of internal waves over the area caused highly anisotropic SSPs as the ones depicted in Fig. 2(a), where temperature

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

profiles exhibited significant variations across the water column and caused notable changes in both the acoustic channel and sound transmission.

The study of IWs is a difficult task due to the spatial resolution of the measurements in experiment areas (Badiey et al. 2016). In this paper, only temporal IW anomalies are used for the analyses and applications presented in subsequent sections. As an example, the temporal displacement of an internal wave event spotted from 17 Aug 21:00:00 to 18 Aug 10:00:00 UTC 2006 at SW30 location is presented in Fig. 2. The temperature variability across the water column produced by the passing of IWs provokes changes in the acoustic duct and degrades the acoustic propagation underwater. In Fig. 2(a), the beginning of each stage (approaching, on-set, propagation and tail) of the IW event is marked with a dashed line in Fig. 2(a) and labeled with a geotime t_{g_i} (i.e. t_{g_0}, t_{g_1}, t_{g_2}, and t_{g_3}). These abrupt changes in temperature alter the isotropic properties of the SSPs and produce drastic variations in the SSPs as shown in Figs. 2(c)-(d). Figure 2(c) shows the mean $\mu_{\mathbf{y}_i}$ and standard deviation $\sigma_{\mathbf{y}_i}$ values of the SSPs over the entire time window in panel (a), whereas Fig. 2(d) depicts individual SSPs at geotimes t_{g_i}, with colors matching the vertical lines in part (a).

An additional consequence of the IW passing is the instability of vertical displacements that affects acoustic propagation. The periodicity at which a vertically displaced small volume of water oscillates is measured by the Buoyancy frequency N in s^{-1}. The oscillations in the water column are expressed as the squared of the Buoyancy frequency N^2 to obtain real values, as shown in Fig. 2(b). The time interval has been divided into four sections, each categorized by the regimens (1)-(4) describing the IW behavior. These regimens are (1) the approaching, (2) the on-set, (3) the propagation, and (4) the tale of the IW event. The square Buoyancy frequency N^2 is given by $g^2\rho\frac{\beta dS_A - \alpha d\Theta}{dP}$, where g is the gravitational acceleration in $\mathrm{m/s}^2$, ρ is the density in $\mathrm{kg/m}^3$, S_A is the absolute salinity in $\mathrm{g/kg}$, Θ is the conservative temperature in Celsius, P is the pressure in Pascals, and β and α are the saline contraction and thermal expansion coefficients evaluated at the

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

average values of S_A, Θ and P. The mean value of N^2 for each of the four regimens indicated in panel (b) is shown in Fig. 2(e). The behavior of the Buoyancy frequency clearly shows the magnitude of oscillations at different depths during the IW passing.

The sound speed profiles used in this work were derived from environmental measurements at mooring SW30. The temperature, conductivity, and pressure values collected at mooring SW30 were employed to generate water salinity profiles using the equation introduced by Fofonoff and Millard Jr (1983). Then, a discretized version of SSPs in terms of depth is obtained using Eq. 1, which is referred to as the nine-term equation derived by Mackenzie (1981). The nine-term equation utilizes depth, temperature, and salinity profiles to compute discrete samples of SSPs at times t_i, with $i = 0, \ldots, n$ as

$$\begin{aligned} c_z = {} & 1448.96 + 4.591T - 5.304 \times 10^{-2}T^2 + 2.374 \times 10^{-4}T^3 + 1.340(S-35) \\ & + 1.630 \times 10^{-2}Z + 1.675 \times 10^{-7}Z^2 - 1.025 \times 10^{-2}T(S-35) - 7.139 \times 10^{-13}TZ^3, \end{aligned} \tag{1}$$

where c_z is the discretized sound speed profile in m/s, T is the temperature in Celsius, Z is the depth in meters, and S is the salinity ratio PSS-78 (Lewis and Perkin 1981). Even though the salinity ratio is a dimensionless value is commonly reported as ppt (parts per thousand).

The computed SSPs were structured as a group of sample vectors $\mathbf{y}_i$ such that $\mathbf{Y} = [\mathbf{y}_1, \ldots \mathbf{y}_n] \in \mathbb{R}^{m \times n}$, with m discrete depths points (features) and n SSP samples in time. SSP samples were generated every 15 s during the entire period. To obtain a better representation of the sound speed in the water column, shape-preserving piecewise cubic interpolation was used along the depths from 14 m to 40 m with a spacing of 0.5 m, yielding $m = 53$ depth-dependent features. The above resulted in $n = 205,547$ SSP samples for the entire dataset (01 Aug 00:00:00 to 05 Sep 16:00:00 2006).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

As shown in Fig. 2 and 3, internal waves events can be categorized as phenomena with low, medium, and high effects in the water column by considering the amplitude of the internal wave (variation of temperature versus depth). To distinguish between IW activities with low, medium, and high incidence, the extracted SSPs were labeled in four different classes (1)-(4), as detailed in Table 1. These resulting labels serve for the supervised framework presented in the following sections. The classes were inferred via *k*-medoids clustering with Euclidean distance as the measuring metric (Park and Jun 2009). This approach is more robust to noise and outliers as compared to *k*-means because it minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances. The clustering strategy is composed of two steps, the build-step, and the swap-step. In the build-step, each of *k* clusters is associated with a potential medoid, while in swap-step, each point is tested as a potential medoid by checking the sum of within-cluster distances to define a new medoid. At each iteration, every point is then assigned to the cluster with the closest medoid until convergence.

The aforementioned process led to a full-labeled dataset with four distinguishable subsets $\mathbf{Y}_i$, one per class $(i) \in [1, \dots 4]$. Out of the total number of samples in the dataset ($n = 205,547$). Class (1) has $n = 32,992$ SSPs, class (2) $n = 59,630$ samples, class (3) $n = 67,536$ SSPs, whereas class (4) $n = 45,389$ samples. The distribution of classes can be observed in Fig. 4. Basic descriptive statistics were calculated for each subset, as shown in Table 1. The mean $(\mu_{\mathbf{Y}_i})$ and the standard deviation $(\sigma_{\mathbf{Y}_i})$ of all elements in class (i) were computed as $\mu_{\mathbf{Y}_i} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} y_{ij}$ and $\sigma_{\mathbf{Y}_i} = \sqrt{\frac{1}{(n-1)(m-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(y_{ij} - \mu_{\mathbf{Y}_i}\right)^2}$, respectively.

The values of $\mu_{\mathbf{Y}_i}$ and $\sigma_{\mathbf{Y}_i}$ for each class subset provide a general view of the magnitude and variability of the SSPs within class (i). As shown in Table 1 and Fig. 3, those classes with higher $\mu_{\mathbf{Y}_i}$ and $\sigma_{\mathbf{Y}_i}$ present SSPs with the larger fluctuations in depth (classes (1)-(4)), while classes with lower values of $\mu_{\mathbf{Y}_i}$ and $\sigma_{\mathbf{Y}_i}$ exhibit less variable SSPs (classes (1)-(3)).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

The entire labeled dataset presented in Fig. 3(a) was split into training/testing sets to be used for supervised classification of SSPs. Uniform random sampling was used to split the data, where 80% of data were destined for training, while the remaining 20% for testing, and resulted in 164,438 and 41,109 samples, respectively. The training data are meant to train the classification algorithms, while the testing data are used to perform classification and measure the performance of each model. The distribution of training and testing sets for each class is shown in Fig. 4.

4. Sparse representation of SSPs

Sparse representation aims to describe a dataset as a linear combination of few elements (basis) (Rubinstein et al. 2010). These elements capture the relevant statistical information that best describes the data and are combined with a matrix of few non-zero coefficients calculated by imposing an ℓ_p constraint that controls the sparsity level or non-zero elements (Zhang et al. 2015; Wright et al. 2010). In ocean acoustics, SSPs inversions are often regularized by considering a sparse representation of SSPs using EOF analysis (Gerstoft and Gingras 1996; Huang et al. 2008) or DL (Bianco and Gerstoft 2017). This section introduces a method to implement empirical orthogonal function analysis and dictionary learning to represent SSPs as a linear combination of basis functions using measured data from the SW06 experiment [see Sec. 3].

a. Empirical orthogonal functions (EOF)

EOF analysis is employed to reduce the dimensionality and identify meaningful underlying features from a dataset. In statistics, EOF analysis is also known as principal component analysis (PCA) and is described as the way of transforming correlated variables into a smaller number of uncorrelated variables (Abdi and Williams 2010). EOF analysis can simplify a spatial-temporal dataset by transforming it to spatial patterns of variability and temporal projections of these pat-

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

terns (Weare et al. 1976). These spatial patterns are called EOFs and can be thought of as basis functions in terms of variance. The associated temporal projections are the principal component scores (ECs) and are the temporal coefficients of the EOF patterns.

EOFs are computed via an SVD in which a dataset is decomposed in terms of orthogonal basis functions to generate a compressed version of the data. Here, EOF analysis is carried out on a collection of SSPs $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{R}^{m \times n}$ with m features corresponding to depths and n time samples. Initially, the dataset $\mathbf{Y}$ is centered by removing time-mean value across the rows to capture only the variance of each depth feature. Subsequently, an SVD is applied to the zero-mean version of $\mathbf{Y}$.

The SVD of $\mathbf{Y}$ is given by $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Since $\mathbf{Y}$ is a centered matrix with zero mean, the covariance matrix $\mathbf{C_Y}$ can be expressed as $\mathbf{C_Y} = \frac{1}{m-1}\mathbf{Y}\mathbf{Y}^T = \frac{1}{m-1}(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T = \mathbf{U}\frac{\mathbf{\Sigma}^2}{m-1}\mathbf{U}^T$. The last expression is analogous to eigendecomposition, where the terms $\mathbf{U}$ and $\frac{\mathbf{\Sigma}^2}{m-1}$ are the non-zero eigenvectors and eigenvalues of $\mathbf{C_Y}$ respectively. This formulation yields the definition of EOFs. Let $\mathbf{Q}$ be the matrix whose columns contain the EOFs that represent the eigenvectors of the matrix $\mathbf{Y}\mathbf{Y}^T$, such that:

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^T = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \tag{2}$$

where $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_m] \in \mathbb{R}^{m \times m}$ are the eigenvectors (EOFs) of $\mathbf{Y}\mathbf{Y}^T$, and $\mathbf{\Lambda} = \mathbf{\Sigma}^2 \in \mathbb{R}^{m \times m}$ shows the variances of the respective EOF $\mathbf{q}_i$, with $i = 1, \ldots, m$.

Since the EOFs in $\mathbf{Q}$ are ordered from highest to lowest variance, dimension reduction of $\mathbf{Y}$ is addressed using only the first $k \leq m$ leading-order EOFs (denoted by columns of $\mathbf{Q}$). This new low-dimensional space retains meaningful properties from the original data. Typically, $k = 5$ EOFs can explain most of the variance in $\mathbf{Y}$ (Bianco and Gerstoft 2017). In the same way, EOFs are also used for data compression, using the basis functions in the dictionary $\mathbf{Q}$. A sparse

representation of $\mathbf{Y}$ is achieved by expressing the dataset as the product of the EOFs in $\mathbf{Q}$ and a coefficient matrix $\mathbf{C}$ such that

$$\hat{\mathbf{Y}} = \mathbf{QC}, \tag{3}$$

where $\hat{\mathbf{Y}}$ is the sparse representation of the original dataset $\mathbf{Y}$, and $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_n] \in \mathbb{R}^{m \times n}$. A full representation of individual SSPs $\mathbf{y}_i$ can be obtained using $\hat{\mathbf{y}}_i = \mathbf{Qc}_i$, $i = 1, \ldots n$. A sparse representation $\hat{\mathbf{y}}_i$ is computed by using the k leading-order EOFs such that $\hat{\mathbf{y}}_i = \mathbf{Q}_k\mathbf{c}_i$, with $\mathbf{Q}_k \in \mathbb{R}^{m \times k}$ and a coefficient vector $\mathbf{c}_i \in \mathbb{R}^k$. Here, a dictionary is defined as a matrix whose columns comprises of basis functions that retrieve meaningful information from a dataset. In consequence, $\mathbf{Q}_k$ can be thought of as a dictionary containing the first k orthogonal basis functions (EOFs).

b. Sparse coding

Any dataset $\mathbf{Y} \in \mathbb{R}^{m \times n}$ containing m features and n samples can be successfully reconstructed with acceptable error ϵ utilizing a combination of basis vectors in a dictionary $\mathbf{D} \in \mathbb{R}^{m \times k}$, where k represents number of basis functions contained by the dictionary $\mathbf{D}$. If $k < m$ the dictionary is under-complete, if $k = m$ the dictionary is referred to as complete, whereas if $k > m$ the dictionary is considered overcomplete. Data described with sparse coding are assumed to be a linear combination of basis functions $\mathbf{d}_i \in \mathbb{R}^m$ and sparse coding vectors $\mathbf{c}_i \in \mathbb{R}^k$ from the matrix $\mathbf{C} \in \mathbb{R}^{k \times n}$, responsible for defining the synthesis of the data using basis functions.

An optimal sparse solution of the coefficients $\mathbf{c}_i$ such that $i = 1, \ldots, k$ can be obtained by solving an optimization scheme with an ℓ_0-norm constraint used to limit the number of non-zero entries in the vector $\mathbf{c}_i$. The optimization problem with the ℓ_0 penalty is formulated as

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

$$\hat{\mathbf{c}}_i = \underset{\mathbf{c}_i \in \mathbb{R}^k}{\arg\min} \|\mathbf{y}_i - \mathbf{D}\mathbf{c}_i\|_2^2 \quad \text{subject to } \|\mathbf{c}_i\|_0 \leq n_{\text{nz}}, \tag{4}$$

where n_{nz} is the number of non-zero elements in $\mathbf{c}_i$. The optimization problem in Eq. 4 yields an exact solution. However, the ℓ_0-norm constraint brings in a non-convex and NP-hard problem whose solution is practically unreachable (Tošić and Frossard 2011; Ramirez et al. 2010). Approximation algorithms exist to find a suboptimal solution for the NP-hard problem. Greedy algorithms such as matching pursuit (MP) and orthogonal MP (OMP) (Mallat and Zhang 1993) iteratively solve the non-convex problem imposed by the ℓ_0-norm constraint. These algorithms find the suboptimal sparse coefficients $\mathbf{c}_i$ that best approximate the global minimum. Note that the optimization problem presented in Eq. 4 can also be convex-relaxed by changing the ℓ_0-norm term to a ℓ_1 regularization term (Tošić and Frossard 2011); targeting the solution to

$$\hat{\mathbf{c}}_i = \underset{\mathbf{c}_i \in \mathbb{R}^k}{\arg\min} \|\mathbf{y}_i - \mathbf{D}\mathbf{c}_i\|_2^2 + \lambda \|\mathbf{c}_i\|_1, \tag{5}$$

where λ is the regularization parameter. The minimization in Eq. 5 can be solved efficiently as an ℓ_1 least-squared problem using approaches based on convex relaxation such as basis pursuit (BP) denoising (Chen et al. 2001) or Lasso (Hans 2009). In this approach, the exact number of non-zero elements in the vector $\mathbf{c}_i$ is not controlled by the ℓ_1 penalty as in the ℓ_0-norm case.

One good way to compare the performance between EOFs and LDs for sparse representation is by fixing the number of non-zero coefficients n_{nz} in $\mathbf{c}_i$. Since n_{nz} is fixed, the sparse representation of SSPs must be performed with the k-leading basis from EOFs and LDs. This fact permits focus mostly on the ability of DL and EOF to capture relevant features in the data using very few basis functions. As a result, the OMP algorithm with ℓ_0-norm is used to fairly compare the ability of EOFs and DLs to sparsely represent SSPs.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

c. Dictionary learning (DL)

Contrary to EOF analysis in which the basis functions are computed via SVD, DL aims to find a dictionary $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_k] \in \mathbb{R}^{m \times k}$ that is learned directly from a dataset $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_n] \in \mathbb{R}^{m \times n}$. The basis functions contained in the columns of the dictionary $\mathbf{D}$ are called atoms, and are not required to be orthogonal as in EOF analysis. With DL, each signal in $\mathbf{Y}$ can be sparsely represented as a linear combination of few atoms. Sparse coding is used during the process to guarantee the sparse representation of the dataset $\mathbf{Y}$, where an ℓ_p constraint is imposed to the sparse coefficients in $\mathbf{C} = [\mathbf{c}_1, \ldots \mathbf{c}_n] \in \mathbb{R}^{k \times n}$ as shown in Sec. 4(b).

In this method, not only the dictionary atoms $\mathbf{d}_i$ but also the sparse coefficients $\mathbf{c}_i$ have to simultaneously be learned from data. This problem is addressed using iterative algorithms which solve for the coefficient matrix $\mathbf{C}$ and the dictionary $\mathbf{D}$ separately, and alternate the solutions until convergence. Some efficient algorithms to learn dictionaries are the method of optimal directions (MOD) (Engan et al. 1999) and the K-SVD algorithm (Aharon et al. 2006), based on *k*-means clustering. Even tough K-SVD converges faster than MOD, K-SVD is computationally expensive and relies in high memory use when the dataset $\mathbf{Y}$ is large. In this paper, the stochastic online learning algorithm proposed by Mairal et al. (2009) is used to learn dictionaries. This optimization strategy overcomes the computational issues of the K-SVD algorithm and can be applied to large-scale contexts. Here, the learned dictionary $\mathbf{D}$ is chosen from a convex set $\mathcal{C}$ that contains dictionaries whose columns are ℓ_2-normalized

$$\mathcal{C} \coloneqq \left\{ \mathbf{D} \in \mathbb{R}^{m \times k} \text{ s.t. } \forall i = 1, \ldots, k,\ \mathbf{d}_i^T \mathbf{d}_i \leq 1 \right\}. \tag{6}$$

The approach introduced by Mairal et al. (2009), uses stochastic gradient descent to update the dictionary $\mathbf{D}_t$ sequentially by accessing one training sample at a time and using the previous

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

dictionary $\mathbf{D}_{t-1}$ as a warm restart during the iteration t, as shown in Algorithm 1. The optimization problem to solve for $\mathbf{D}$ and $\mathbf{C}$, using the ℓ_1 regularization term is formulated as,

$$\begin{aligned}\left\langle \hat{\mathbf{D}}, \hat{\mathbf{C}} \right\rangle &= \underset{\mathbf{D}\in\mathcal{C},\mathbf{C}\in\mathbb{R}^{k\times n}}{\arg\min} \|\mathbf{Y}-\mathbf{D}\mathbf{C}\|_2^2+\lambda\|\mathbf{C}\|_1 \\ &= \underset{\mathbf{D}\in\mathcal{C},\mathbf{C}\in\mathbb{R}^{k\times n}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n}\left(\|\mathbf{y}_i-\mathbf{D}\mathbf{c}_i\|_2^2+\lambda\|\mathbf{c}_i\|_1\right)\end{aligned} \tag{7}$$

where λ is the regularization parameter applied to the ℓ_1-norm constraint, n is the number of samples in the dataset $\mathbf{Y} \in \mathbb{R}^{m\times n}$, $\mathbf{D}$ is the dictionary in the convex set $\mathcal{C}$, and $\mathbf{C}$ is the matrix with the sparse coefficients. The convex optimization induced by the ℓ_1-norm can be solved efficiently using the least-angle regression Lasso (LARS-Lasso).

d. Dictionary learning for SSP classification

A dictionary $\mathbf{D} \in \mathbb{R}^{m\times k}$ containing a set of k basis functions $\mathbf{d}_i \in \mathbb{R}^m$ can retain much of the statistical properties of a zero-mean dataset $\mathbf{Y} = [\mathbf{y}_1, \dots \mathbf{y}_i] \in \mathbb{R}^{m\times n}\ i \in [1, \dots n]$, composed by m features and n samples. Similarly, a sparse coefficient vector $\mathbf{c}_i \in \mathbb{R}^k$ is employed to reconstruct samples $\mathbf{y}_i \in \mathbf{Y}$, with acceptable error ϵ, using a linear combination of the basis functions $\mathbf{d}_i$ [see Sec. 4(b)-(c)]. Let $\Phi = \{\phi_1, \dots, \phi_l\}_{j=1}^{l}$ denote a set of labeled classes, then dictionary learning can be extended to supervised classification tasks using a measuring metric $\hat{\mathcal{R}}(\mathbf{y}, \hat{\mathbf{y}})$ to classify unlabeled testing data to a specific class ϕ_j. Here, $\hat{\mathcal{R}}(\cdot)$ is a metric to measure the dissimilarity between the sample $\mathbf{y}_i$ and the reconstruction $\hat{\mathbf{y}}_i = \mathbf{D}\mathbf{c}_i$. Usually, when using the OMP algorithm for sparse coding the dissimilarity metric is defined as $\hat{\mathcal{R}}\left(\mathbf{y}, \hat{\mathbf{y}}\right) = \|\mathbf{y} - \mathbf{D}\mathbf{c}\|_2^2$. However, exist cases where the OMP algorithm can be unstable for classification tasks since small variations in the input signal (i.e., sound speed profiles) yield irregular sparse codes that difficult classification (Ramirez et al. 2010).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

Algorithm 1 Online dictionary learning. Adapted from (Mairal et al. 2009)

Require:

$\mathbf{Y} \in \mathbb{R}^{m \times n} \sim \rho(\mathbf{y})$: Training data

T : number of iterations

λ : sparsity level

$\mathbf{D}_0 \in \mathbb{R}^{m \times k}$: Initial dictionary

$\mathbf{A}_0 \leftarrow 0$, $\mathbf{B}_0 \leftarrow 0$ ▷ Initialize $\mathbf{A}_0$ and $\mathbf{B}_0$

for $t = 1$ to T **do**

Draw y_t from $\rho(\mathbf{y})$

Sparse Coding : via OMP or LARS-Lasso algorithm

$$\hat{\mathbf{c}}_t = \underset{\mathbf{c}_t \in \mathbb{R}^k}{\arg\min} \|\mathbf{y}_t - \mathbf{D}_{t-1}\mathbf{c}_t\|_2^2 \quad \text{subject to } \|\mathbf{c}_t\|_0 \leq n_{\text{nz}}$$

or

$$\hat{\mathbf{c}}_t = \underset{\mathbf{c}_t \in \mathbb{R}^k}{\arg\min} \|\mathbf{y}_t - \mathbf{D}_{t-1}\mathbf{c}_t\|_2^2 + \lambda \|\mathbf{c}_i\|_1$$

$\mathbf{A}_t \leftarrow \mathbf{A}_{t-1} + \hat{\mathbf{c}}_t \hat{\mathbf{c}}_t^T$ ▷ Update $\mathbf{A}_t$ using $\hat{\mathbf{c}}_t$

$\mathbf{B}_i \leftarrow \mathbf{B}_{t-1} + \mathbf{y}_t \hat{\mathbf{c}}_t^T$ ▷ Update $\mathbf{B}_t$ using $\hat{\mathbf{c}}_t$ and $\mathbf{y}_i$

Dictionary Update : compute $\mathbf{D}_t$ with $\mathbf{D}_{t-1}$ as initialization

$$\begin{aligned} \mathbf{D}_t &= \underset{\mathbf{D} \in \mathcal{C}}{\arg\min} \frac{1}{t} \sum_{i=1}^{t} \left(\|\mathbf{y}_i - \mathbf{D}\mathbf{c}_i\|_2^2 + \lambda \|\mathbf{c}_i\|_1 \right) \\ &= \underset{\mathbf{D} \in \mathcal{C}}{\arg\min} \frac{1}{t} \left(\frac{1}{2} \mathrm{Tr}\left(\mathbf{D}^T \mathbf{D} \mathbf{A}_t\right) - \mathrm{Tr}\left(\mathbf{D}^T \mathbf{B}_t\right) \right) \end{aligned}$$

end for

Return $\mathbf{D}_T$ (learned dictionary)

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

In consequence, for classification contexts, the ℓ_0-norm constraint is replaced by an ℓ_1-norm regularization term to balance the trade-off between reconstruction error and sparsity, mitigating in this way undesirable outcomes resulting from OMP. The use of ℓ_1-norm constraint leads to the updated dissimilarity metric

$$\hat{\mathcal{R}}\left(\mathbf{y}, \hat{\mathbf{y}}\right) = \|\mathbf{y} - \mathbf{D}\mathbf{c}\|_2^2 + \lambda\|\mathbf{c}\|_1, \tag{8}$$

which is in fact equal to Eq. 5. The terms in Eq. 8 takes into account the reconstruction error ($\|\mathbf{y} - \mathbf{D}\mathbf{c}\|_2^2$) and the complexity of the sparse decomposition ($\lambda\|\mathbf{c}\|_1$). In other words, the reconstruction error measures the quality of the approximation while the complexity is measured by the ℓ_1-norm of the optimal $\mathbf{c}$.

Additionally, the minimization problem to learn dictionaries [refer to Eq. 7] is modified to train the class-specific dictionaries, following the scheme introduced by Ramirez et al. (2010)

$$\begin{aligned}\min_{\{\mathbf{D}_j, \mathbf{C}_j\}_{j=1,\ldots,l}} \sum_{j=1}^{l} \left\{ \|\mathbf{Y}_j - \mathbf{D}_j\mathbf{C}_j\|_2^2 + \lambda \sum_{j'=1}^{k} \|\mathbf{c}_{j'}^j\|_1 \right\} + \\ \eta \sum_{j' \neq j} \|\mathbf{D}_{j'}^T \mathbf{D}_j^T\|_{\mathcal{F}}^2,\end{aligned} \tag{9}$$

where j and j' are sub-indices to denote elements from different classes $\phi_{j'}, \phi_j \in \Phi$, η is a regularization parameter that penalizes the coherence between inter-class dictionaries $\mathbf{D}_j$ and $\mathbf{D}_{j'}$. The last term in Eq. 9 promotes incoherence between the different class dictionaries by weakening the ability of a dictionary $\mathbf{D}_j$ to classify correctly for other classes $(j') \neq (j)$. In other words, the class-specific dictionary $\mathbf{D}_j$ for class (j) is trained using Eq. 7 while considering the coherence with dictionaries from other classes at the same time (i.e. dictionaries $\mathbf{D}_{j'}$ s.t. $j' \neq j$). This optimization process yields class-specific dictionaries than are very different from each other and

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

contains the most relevant information from the data. For further details, refer to Ramirez et al. (2010).

The classification setting presented here is comprised by five steps, (1) **dictionary training:** class-specific dictionaries $\mathbf{D}_j$ are trained using labeled data in each class ϕ_j, these class-specific learned dictionaries are meant to capture meaningful information about data in the class ϕ_j. This step is performed by combining Algorithm. 1 with Eq. 9. (2) **Sparse coding on testing data:** sparse coding via LARS-Lasso algorithm with ℓ_1 is applied to each class-specific dictionary $\mathbf{D}_j$ and testing data $\mathbf{y}_i^{\text{test}}$ to compute the sparse coefficients $\mathbf{c}_i^{\text{test}}$. (3) **Reconstruction:** the j^{th} class dictionary $\mathbf{D}_j$ and set of sparse coefficients $\mathbf{c}_i^{\text{test}}$ are used to compute the i^{th} reconstruction $\hat{\mathbf{y}}_i^j = \mathbf{D}_j\mathbf{c}_i^{\text{test}}$. As a result of this, a sparse representation $\hat{\mathbf{y}}_i^j$ is generated for each class $\phi_j \in \Phi$, $j = 1, \ldots, l$. (4) **Error calculation:** dissimilarity between the each reconstruction $\hat{\mathbf{y}}_i^j$ and testing data $\mathbf{y}_i^{\text{test}}$ is computed via $\hat{\mathcal{R}}(\mathbf{y}^{\text{test}}, \hat{\mathbf{y}}_i^{\text{j}})$. (5) **Labeling:** the testing dataset $\mathbf{y}_i^{\text{test}}$ is assigned to the class $\hat{j}$ that generates the lowest error $\hat{\mathcal{R}}(\cdot)$ as follows

$$\hat{j} = \underset{j=1,\ldots,l}{\arg\min}\, \hat{\mathcal{R}}\left(\mathbf{y}_i^{\text{test}}, \hat{\mathbf{y}}_i^{\text{j}}\right), \tag{10}$$

where $\hat{j}$ is the class assigned to the unlabeled testing data sample $\mathbf{y}_i^{\text{test}}$.

5. Results and Discussion

To illustrate the practicability of the methods presented in Sec. 4, both DL and EOF analysis are applied to data collected at SW30 station during the SW06 experiment [Sec. 2]. In this section, DL and EOFs are implemented to sparsely represent SSPs altered by the passing of IWs using OMP algorithm for sparse coding. In addition, dictionary learning is later employed for SSP classification following the scheme described in Sec. 4(d), using the LARS-Lasso algorithm for sparse coding.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

The matrix $\mathbf{Q} \in \mathbb{R}^{m \times k}$ containing the EOFs is calculated via SVD [Eq. 2], and the learned dictionary $\mathbf{D} \in \mathbb{R}^{m \times k}$ is obtained using the convex optimization shown in Eq. 7, via online dictionary learning algorithm. Equation 7 is solved iteratively using the LARS algorithm with a tolerance parameter $\delta = 1e^{-8}$, number of iterations $T_{\text{max}} = 2000$, and regularization parameter $\lambda = 1$ (Pedregosa et al. 2011). Furthermore, sparse coding via OMP algorithm with $T_{\text{max}} = 2000$ and is used to compute the coefficient matrix $\mathbf{C} \in \mathbb{R}^{k \times n}$ to compare performance of EOFs and DL [Sec. 5(a)]. The number of non-zero elements in $\mathbf{c}_i$, computed via OMP algorithm, are set to exactly $n_{\text{nz}} = 3$ non-zero coefficients. For classification, only DL is used and sparse coding is done via LARS-Lasso algorithm with $\lambda = 1$, $\eta = 1$, $T_{\text{max}} = 2000$, and $\delta = 1e^{-8}$ [see Sec. 5(b)].

a. Dictionary learning and EOF analysis for SSPs sparse representation

Due to the reward for sparsity and the absence of an orthogonality requirement, learned dictionaries (LD) can provide an alternative for a sparse representation that might yield more accurate reconstructions than traditional methods, such as empirical orthogonal functions. Bianco and Gerstoft (2017) showed that LDs are suitable for representing sound speed profiles in underwater environments, even outperforming EOFs with few basis functions.

In order to test the versatility of dictionary atoms in highly fluctuating environments such as those with internal waves, both EOFs and DL are compared using a portion of data collected from 17 Aug 00:00:00 to 18 Aug 23:59:59 UTC 2006 at SW30 station during high internal wave activity. The data used to compared both methods are shown between the two black lines drawn in Fig. 4(a), with $n = 11520$ SSPs sampled every 15 s measured at $m = 53$ different depths. EOFs are calculated for $k = 53$ basis functions. Similarly, a complete dictionary with $k = 53$ and an overcomplete dictionary with $k = 100$ are also computed. Sparse coding based on both EOFs and LDs is performed using $n_{\text{nz}} = 3$ non-zero coefficients via OMP algorithm. The resulting EOFs and

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

atoms from the LD are shown in Fig. 5. Each EOF/atom contains meaningful information about the sound speed variability in terms of depth. Notice only the leading-order EOFs in Fig. 5(a) capture the variability of SSPs, whereas the SSP variance is distributed on all atoms of both complete [Fig. 5(b)] and overcomplete dictionary [Fig. 5(c)].

The explained variance ratio per EOF/atom is shown in Fig. 6 and complements the findings presented in Fig. 5. Most of the variance of the SSPs is mainly concentrated in the first $k = 5$ leading-order EOFs [Fig. 6(a)], while in the case of LDs, the variance is shared among most of the basis functions [Figs. 6(a)-(b)]. Given these findings, SSPs can be reconstructed effectively using only the first leading-order EOFs, while in the case of LDs, the SSPs can be sparsely represented by a sparse combination of almost any atom in the dictionary.

As stated in Sec. 4, the EOFs in the matrix $\mathbf{Q}$ are computed via SVD and have the property of being normalized orthogonal basis functions. Conversely, atoms in a dictionary $\mathbf{D}$ are not required to be orthogonal and are restricted to have unit ℓ_2-norm as stated in Eq. 6. Coherence among column entries in $\mathbf{Q}$ and $\mathbf{D}$ can be calculated by computing the Gram matrix $\mathbf{G_Q} = |\mathbf{Q}^T\mathbf{Q}|$ for EOFs, and $\mathbf{G_D} = |\mathbf{D}^T\mathbf{D}|$ for LDs. For the EOF analysis, $\mathbf{G_Q} = \mathbf{I}$, as shown in Fig. 7(a), because of the orthogonal properties of EOFs. In contrast, atoms in the complete and overcomplete LDs are not necessarily orthogonal, as displayed in Fig. 7(b)-(c).

The relaxation of the orthogonality requirement for LD atoms leads to more flexible dictionaries and richer data representation. As a result of this, better compression of data is achieved using dictionary learning. For both LDs and EOFs, increasing the number of basis functions k and non-zero coefficients n_{nz} yields lower representation errors. To study sparse representation of SSPs, the OMP algorithm is used to compute sparse codes with $1 \leq n_{\text{nz}} \leq 12$ for EOFs and LDs. The root-mean-squared error (RMSE) between the sparse representation $\hat{\mathbf{Y}}_i$ and the original data $\mathbf{Y}$ for is measured as

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\hat{y}_{ij}\right)^2}{mn}}, \tag{11}$$

where m are depths, and n is the number of SSPs in $\mathbf{Y}$. The RMSE values as a function of n_{nz}, shown in Fig. 8, provides a method for comparing the different sparse representations. It is clear that for $n_{\text{nz}} \geq 8$, the EOFs, complete LD, and overcomplete LD give the almost the same RMSE. For $n_{\text{nz}} < 7$, however, the RMSE using EOFs is larger than for LDs, with the overcomplete LD yielding the lowest RMSE for small values of n_{nz}.

As an attempt to demonstrate the efficacy of both EOFs and LDs to sparsely represent SSPs, six individual random samples $\mathbf{y}_i$ are chosen from the dataset $\mathbf{Y}$ and are depicted with solid green lines in Fig. 9. The sparse representation of SSPs $\hat{\mathbf{y}}_i$ of the six samples using the EOFs and the complete LD, both with $k = 53$ and $n_{\text{nz}} = 3$ are shown in Fig, 9. The dashed orange line presents the time-mean value from the entire dataset, dotted blue line shows the SSP reconstructed using EOFs whereas dash-dotted red line the reconstruction using LDs. In addition, the absolute error between each real sample $\mathbf{y}_i$ and the reconstruction $\hat{\mathbf{y}}_i$ in terms of depth was computed as $|\mathbf{y}_i - \hat{\mathbf{y}}_i|$, and is shown with color bars (to the right of each line plot) for both LD and EOFs. Dark colors in the bars correspond to small errors, whereas light colors represent high absolute errors. For each case, only with $n_{\text{nz}} = 3$ basis functions both EOFs complete LD yield a good representation of SSPs. Even with internal wave events passing by, complete LD provides a nearly perfect representation of the six SSPs and outperforms conventional EOF analysis in the sparse reconstruction of sound speed profiles.

b. Classification of SSPs via dictionary learning

In previous sections is mentioned that learned dictionaries $\mathbf{D}$ can retain meaningful information from a dataset $\mathbf{Y} \in \mathbb{R}^{m\times n}$ that can or not be labeled in a class ϕ_j, $j = 1, \ldots, l$. However, limitations

for data classification arise when few n_{nz} atoms are used due to the impossibility of representing the variability in the data. Overcomplete dictionaries, due to their redundant nature, are suitable for classification since they can perform data analysis with functions that are likely to match the characteristics in different classes of data.

In this section, the dictionary learning framework is used to label sound speed profiles following the classification setting introduced in Sec. 4(d). Complete and overcomplete LDs are employed to classify data with low, medium, and high occurrence of internal wave events in the ocean. For the best analysis, the data introduced in sec. 2 and Table 1 are used for to test the algorithms in this section.

As the extracted SSPs from the experiment are not initially labeled, each SSP sample is labeled into four classes (1)-(4) depending on the internal wave level using *k*-medoids algorithm. Once all data samples are labeled, the effectiveness of LD atoms to infer a class-type for SSPs is studied using the training/testing sets described in Sec. 3. Here, class-specific dictionaries are learned using training data, to then, classify unlabeled SSPs samples in the testing set utilizing a dissimilarity metric $\hat{\mathcal{R}}(\mathbf{y}, \hat{\mathbf{y}})$, [Eq. 10]. The scheme introduced in Sec. 4(d) is implemented for complete ($k = 53$) and overcomplete ($k = 100$ and $k = 300$) dictionaries. In addition, dictionary learning is compared with support vector machine (SVM) and *k*-nearest neighbor (KNN) algorithms, both commonly used for supervised classification tasks. KNN is implemented using $k = 100$ neighbors and Euclidean distance as metric, whereas ℓ_2 regularized SVM with Gaussian kernel is employed with a one-vs-the-rest configuration for multi-class classification.

The LDs models, KNN and SVM are trained using the training set with $n = 164,438$ samples and four different classes and then tested on the remaining $n = 41,109$ unseen samples, whose labels are inferred by the classification model. That is, the trained classifier $h \in \mathcal{H}$ aims to assign a label to unlabeled data $\mathbf{y}_i^{\text{test}}$, such that $\hat{j}_i = h(\mathbf{y}_i^{\text{test}})$, $\hat{j} = 1, \ldots, l$. Here, the accuracy is reported

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

for each model, and accounts for the number of correct classifications over the total number of samples in the testing set. This metric is calculated as

$$\text{accuracy} = \frac{\sum_{i=1}^{n} \mathbb{1}(\hat{j}_i = j_i)}{n} \times 100\% \tag{12}$$

where $\mathbb{1}(\cdot)$ is the indicator function, n is the total number of samples in the dataset, j_i is label or ground truth for the i^{th} sample, and $\hat{j}_i$ is the inferred class for the i^{th} SSP, such that $\hat{j}_i = h(\mathbf{y}_i^{\text{test}})$.

The classification results for KNN, SVM and LDs are shown in Table 2, where accuracy is reported for all the models using the entire testing data along with the performance for individual subsets corresponding to each independent class (j). As demonstrated in Sec. 5(a), the overcomplete LD yields lower reconstruction errors than complete LDs. This fact is also corroborated by results in Table 2, where the overcomplete LD with $k = 300$ reaches higher accuracy than LDs with $k = 53$ and $k = 100$ for all the four classes.

It is notable that results of the overcomplete LD with $k = 300$ are comparable to SVM and KNN and are even better at differentiating between classes with high internal wave activity (classes (2) and (4)). Notice the misclassification of SSPs by LDs is due to the possible lack of sufficient information about the variability retrieved by the k atoms. It is clear to see that the more atoms the dictionary uses, the more the accuracy will be. Therefore, the classification results reported for LDs in Table 2 can be improved by increasing the number of atoms used. This study shows that classification of SSPs via dictionary learning is feasible and can be extended to large-context scenarios as long as exists labeled data to compute specific dictionaries for each class.

From results in Table 2, it is possible to conclude overcomplete dictionary learning offers a good alternative for classifying SSPs with high internal wave activity if using sufficient k atoms. The relaxation in the orthogonal constraint in the basis functions allows DLs to capture the most representative information from data. Each class-specific dictionary $\mathbf{D}_j$ tends to have different

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

patterns than others from different classes ($\mathbf{D}_{j'}$). The coherence $\mathbf{G}_{\mathbf{D}_j}$ for class-specific learned dictionaries with $k = 53$, $k = 100$, and $k = 300$ atoms are shown in Fig. 10. Notice that for different values of k and class, each class-specific dictionary presents different and distinctive patterns than are being learned from data.

It is important to remark that if the number of k atoms increases, the dictionary will lead to higher accuracy with a cost of an increment in the complexity of the convex optimization. Therefore, it is important to consider the trade-off between accuracy and complexity when training LDs, as is done with any learning model, such as neural networks. In cases where there are not sufficient labeled training data, it is possible to apply data processing techniques such as data augmentation to increase the variability and the number of data samples within a class (Castro-Correa et al. 2021).

Similar to previous research (Bianco and Gerstoft 2017; Sun and Zhao 2020), overcomplete LDs perform well as a sparse representation algorithm for SSPs. Dictionary learning outperforms EOFs because it tends to distribute the SSPs energy among all the atoms and its ability to generate non-orthogonal functions. Furthermore, as few as n_{nz} non-zero coefficients are needed to consistently provide a complete enough representation to achieve very low training error even when dictionaries are trained on SSPs in the presence of internal waves. Results demonstrate competitive performance of LDs with respect to standard classification models. The amount of data, the sparsity level provided by n_{nz}, and the number of atoms k are crucial factors to obtain optimal sparse representations of SSPs.

6. Conclusions

Both dictionary learning and empirical orthogonal functions (EOFs) were implemented to sparsely represent SSPs disturbed by the passing of internal waves in the SW06 experiment. Due

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

to their redundant nature and their ability to generate non-orthogonal basis functions, overcomplete learned dictionaries (LD) showed better performance for reconstructing SSPs than the EOFs and the complete LD when using the best $n_{\text{nz}} = 3$ combination of basis functions.

The presence of internal waves in the water column causes highly anisotropic SSPs. The variability in the SSPs induced by IWs was reflected in higher errors when both EOF and dictionary learning frameworks were applied. Under those circumstances, overcomplete dictionaries with $k > m$ atoms were shown to achieve even better compression of SSPs than EOFs and the complete LD $k = m$. The improvement in the reconstruction of SSPs was produced due to the relaxation of the orthogonal requirements, and the number of atoms used in the dictionary.

In this paper, the classification of SSPs via dictionary learning was introduced. Here, specific dictionaries were built for each class of internal wave activity, and testing datasets were classified by finding the dictionary corresponding to the most accurate sparse representation. Results demonstrated that overcomplete learned dictionaries trained on labeled data are suitable to classify SSPs successfully. When the training data are representative and are labeled by a class, the resulting overcomplete dictionary can be effectively applied to other datasets to classify SSPs.

This work provides insights into the application of learned dictionaries for the representation and classification of SSPs. Further analyses are required to find the optimal number of non-zero coefficients n_{nz} and k atoms used for an optimal sparse representation of SSPs. Future research needs to be conducted to find a better-suited dissimilarity cost $\hat{\mathcal{R}}(\cdot)$ that can yield higher classification performance, while futher studies are required to determine if undercomplete LDs $(k < m)$ are suitable for sparse representation and dimensionality reduction of SSPs.

Sparse representation of sound speed profiles using dictionary learning promotes and expedites research into internal waves. A representative dictionary of basis functions is an efficient way to store and generate thousands of unique sound speed profiles, given that the dictionary succinctly

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

and effectively represents the small-scale variability to model. Using learned dictionaries to generate realistic and variable training datasets may further the progress of machine learning in many undersea applications.

Acknowledgments. This research was supported by Office of Naval Research grants No. N00014-21-1-2424 and N00014-21-1-2760.

Data availability statement. Data used in this work were collected during the Shallow Water Experiment 2006 (SW06). Due to privacy and ethical concerns, neither the data nor the source of the data can be made available.

References

Abdi, H., and L. J. Williams, 2010: Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, **2 (4)**, 433–459.

Abiva, J., T. Fabbri, and R. Vicen-Bueno, 2019: Automatic classification of sound speed profiles using pca and self-organizing map techniques. *OCEANS 2019-Marseille*, IEEE, 1–10.

Aharon, M., M. Elad, and A. Bruckstein, 2006: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, **54 (11)**, 4311–4322.

Badiey, M., L. Wan, and J. F. Lynch, 2013: Statistics of internal waves measured during the shallow water 2006 experiment. *The Journal of the Acoustical Society of America*, **134 (5)**, 4035–4035.

Badiey, M., L. Wan, and J. F. Lynch, 2016: Statistics of nonlinear internal waves during the shallow water 2006 experiment. *Journal of Atmospheric and Oceanic Technology*, **33 (4)**, 839–846.

Bianco, M., and P. Gerstoft, 2017: Dictionary learning of sound speed profiles. *The Journal of the Acoustical Society of America*, **141 (3)**, 1749–1758.

Castro-Correa, J. A., M. Badiey, T. B. Neilsen, D. P. Knobles, and W. S. Hodgkiss, 2021: Impact of data augmentation on supervised learning for a moving mid-frequency source. *The Journal of the Acoustical Society of America*, **150 (5)**, 3914–3928, doi:10.1121/10.0007284, URL https://doi.org/10.1121/10.0007284, https://doi.org/10.1121/10.0007284.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001: Atomic decomposition by basis pursuit. *SIAM review*, **43 (1)**, 129–159.

Engan, K., S. O. Aase, and J. H. Husoy, 1999: Frame based signal compression using method of optimal directions (mod). *1999 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, Vol. 4, 1–4.

Flatté, S. M., and F. D. Tappert, 1975: Calculation of the effect of internal waves on oceanic sound transmission. *The Journal of the Acoustical Society of America*, **58 (6)**, 1151–1159.

Fofonoff, N. P., and R. Millard Jr, 1983: Algorithms for the computation of fundamental properties of seawater. *UNESCO Technical Papers in Marine Sciences*.

Gerstoft, P., and D. F. Gingras, 1996: Parameter estimation using multifrequency range-dependent acoustic data in shallow water. *The Journal of the Acoustical Society of America*, **99 (5)**, 2839–2850.

Hans, C., 2009: Bayesian lasso regression. *Biometrika*, **96 (4)**, 835–845.

Helfrich, K. R., and W. K. Melville, 2006: Long nonlinear internal waves. *Annu. Rev. Fluid Mech.*, **38**, 395–425.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

Huang, C.-F., P. Gerstoft, and W. S. Hodgkiss, 2008: Effect of ocean sound speed uncertainty on matched-field geoacoustic inversion. *The Journal of the Acoustical Society of America*, **123 (6)**, EL162–EL168.

Jain, S., and M. Ali, 2006: Estimation of sound speed profiles using artificial neural networks. *IEEE Geoscience and Remote Sensing Letters*, **3 (4)**, 467–470.

Katsnelson, B. G., O. A. Godin, and Q. Zhang, 2021: Observations of acoustic noise bursts accompanying nonlinear internal gravity waves on the continental shelf off new jersey. *The Journal of the Acoustical Society of America*, **149 (3)**, 1609–1622.

Kuo, Y.-H., and J.-F. Kiang, 2020: Estimation of range-dependent sound-speed profile with dictionary learning method. *IET Radar, Sonar & Navigation*, **14 (2)**, 194–199.

Lewis, E., and R. Perkin, 1981: The practical salinity scale 1978: conversion of existing data. *Deep Sea Research Part A. Oceanographic Research Papers*, **28 (4)**, 307–328.

Mackenzie, K. V., 1981: Nine-term equation for sound speed in the oceans. *The Journal of the Acoustical Society of America*, **70 (3)**, 807–812.

Mairal, J., F. Bach, J. Ponce, and G. Sapiro, 2009: Online dictionary learning for sparse coding. *Proceedings of the 26th annual international conference on machine learning*, 689–696.

Mallat, S. G., and Z. Zhang, 1993: Matching pursuits with time-frequency dictionaries. *IEEE Transactions on signal processing*, **41 (12)**, 3397–3415.

Newhall, A. E., and Coauthors, 2007: Acoustic and oceanographic observations and configuration information for the whoi moorings from the sw06 experiment. Tech. rep., Woods Hole Oceanographic Institution MA.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

North, G. R., 1984: Empirical orthogonal functions and normal modes. *Journal of Atmospheric Sciences*, **41 (5)**, 879–887.

Park, H.-S., and C.-H. Jun, 2009: A simple and fast algorithm for k-medoids clustering. *Expert Systems with Applications*, **36 (2, Part 2)**, 3336–3341, doi:https://doi.org/10.1016/j.eswa.2008.01.039, URL https://www.sciencedirect.com/science/article/pii/S095741740800081X.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, **12**, 2825–2830.

Ramirez, I., P. Sprechmann, and G. Sapiro, 2010: Classification and clustering via dictionary learning with structured incoherence and shared features. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE, 3501–3508.

Roundy, P. E., 2015: On the interpretation of eof analysis of enso, atmospheric kelvin waves, and the mjo. *Journal of Climate*, **28 (3)**, 1148–1165.

Rouseff, D., 2001: Effect of shallow water internal waves on ocean acoustic striation patterns. *Waves in Random Media*, **11**, 377–393.

Rubinstein, R., A. M. Bruckstein, and M. Elad, 2010: Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, **98 (6)**, 1045–1057.

Sprechmann, P., and G. Sapiro, 2010: Dictionary learning and sparse coding for unsupervised clustering. *2010 IEEE international conference on acoustics, speech and signal processing*, IEEE, 2042–2045.

Stewart, G. W., 1993: On the early history of the singular value decomposition. *SIAM review*, **35 (4)**, 551–566.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

Sun, S., and H. Zhao, 2020: Sparse representation of sound speed profiles based on dictionary learning. *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, IEEE, 484–488.

Suo, Y., M. Dao, U. Srinivas, V. Monga, and T. D. Tran, 2014: Structured dictionary learning for classification. *arXiv preprint arXiv:1406.1943*.

Tang, D., and Coauthors, 2007: Shallow water'06: A joint acoustic propagation/nonlinear internal wave physics experiment. *Oceanography*, **20 (4)**, 156–167.

Tang, W., A. Panahi, H. Krim, and L. Dai, 2019: Analysis dictionary learning based classification: Structure for robustness. *IEEE Transactions on Image Processing*, **28 (12)**, 6035–6046.

Tošić, I., and P. Frossard, 2011: Dictionary learning. *IEEE Signal Processing Magazine*, **28 (2)**, 27–38.

Weare, B. C., A. R. Navato, and R. E. Newell, 1976: Empirical orthogonal analysis of pacific sea surface temperatures. *Journal of Physical Oceanography*, **6 (5)**, 671–678.

Wright, J., Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, 2010: Sparse representation for computer vision and pattern recognition. *Proceedings of the IEEE*, **98 (6)**, 1031–1044.

Xu, W., and H. Schmidt, 2006: System-orthogonal functions for sound speed profile perturbation. *IEEE Journal of Oceanic Engineering*, **31 (1)**, 156–169.

Zhang, Z., Y. Xu, J. Yang, X. Li, and D. Zhang, 2015: A survey of sparse representation: algorithms and applications. *IEEE access*, **3**, 490–530.

Zhao, C., Z. Feng, X. Wei, and Y. Qin, 2018: Sparse classification based on dictionary learning for planet bearing fault identification. *Expert Systems with Applications*, **108**, 233–245.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

LIST OF TABLES

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

TABLE 1: Four resulting classes after applying k-medoids to SSP data extracted from SW30 (01 Aug 00:00:00 to 05 Sep 16:00:00 UTC 2006). The columns in the table show the number of SSP samples per class, the mean $(\mu_{\mathbf{Y}})$, and standard deviation $(\sigma_{\mathbf{Y}})$ for the SSPs in each class.

Class	Number of SSPs	$\mu_{\mathbf{Y}}$ (m/s)	$\sigma_{\mathbf{Y}}$ (m/s)
(1)	33,992	1513.46	12.93
(2)	59,630	1497.51	11.34
(3)	67,536	1490.81	7.61
(4)	45,389	1503.64	14.95
Total	**205,547**	**1499.49**	**13.91**

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

TABLE 2: Classification results for KNN, SVM, complete LD, and overcomplete LDs for testing data. Accuracy in (%) is reported for each classification model. The first four rows in the table correspond to independent classes in the testing set, while the last row presents the overall accuracy for each model in the complete testing set.

Dataset	KNN	SVM	Complete LD ($k = 53$)	Overcomplete LD ($k = 100$)	Overcomplete LD ($k = 300$)
Testing - only class (1)	97.91	99.38	95.61	98.51	**99.71**
Testing - only class (2)	**98.68**	89.39	83.65	90.00	95.46
Testing - only class (3)	98.30	**99.72**	89.16	92.51	95.42
Testing - only class (4)	97.18	72.50	93.63	95.65	**98.79**
Testing - complete	**98.10**	90.64	89.59	93.44	96.87

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

LIST OF FIGURES

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

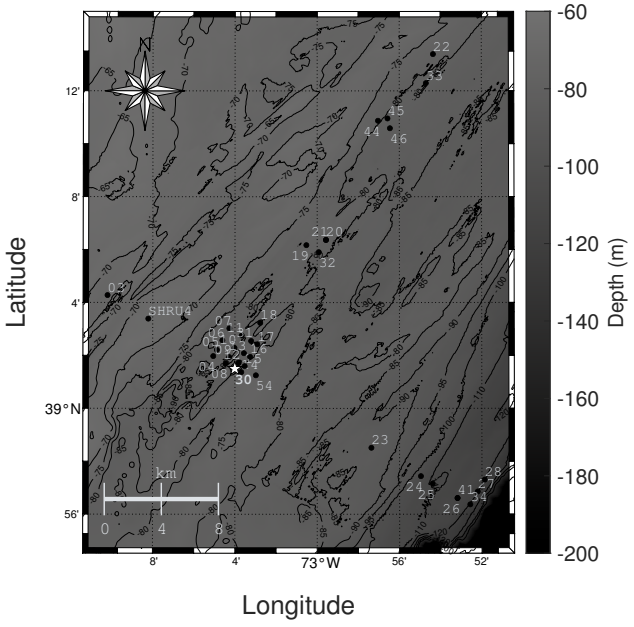

FIG. 1: Some of the SW06 moorings along-shelf and across-shelf conforming a "T" geometry. SSPs used in this work are derived from data collected at SW30 station, which is marked with a white star in the figure.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

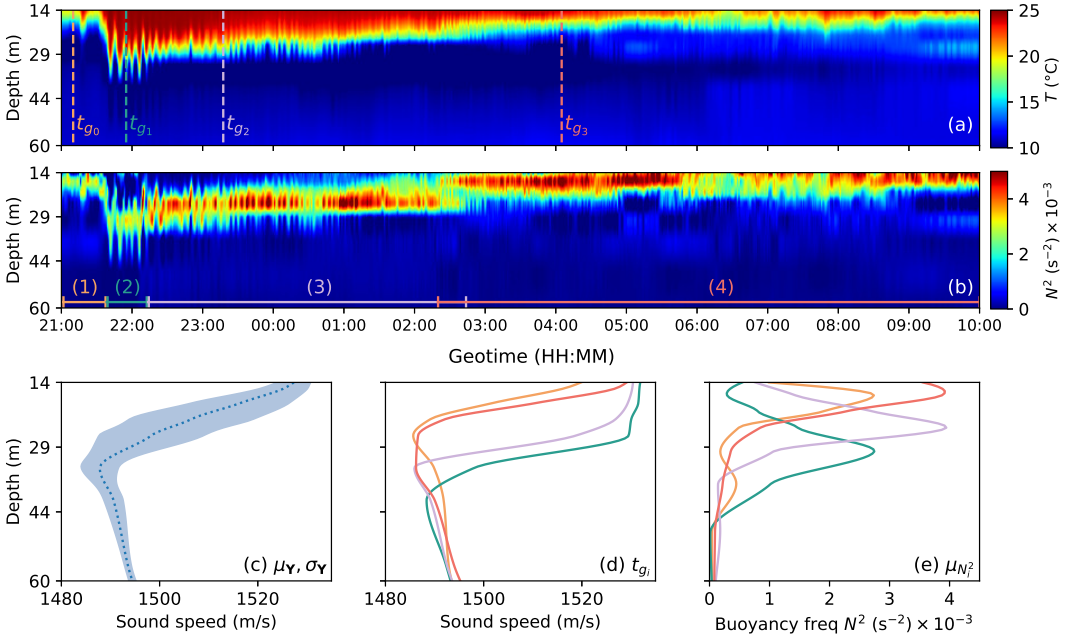

FIG. 2: IW event spotted from 17 Aug 21:00:00 to 18 Aug 10:00:00 UTC 2006. (a) Temporal evolution of temperature profiles at mooring SW30. (b) The square of the Buoyancy frequency N^2, in terms of depth, divided into four sections corresponding to different regimens. (c) Standard and mean of the SSPs presented in part (a). (d) Individual SSP samples at each geotimes t_{g_i} with colors matching the vertical dashed lines in part (a). (e) Buoyancy frequency N^2 for each of the four regimens shown in part (b).

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

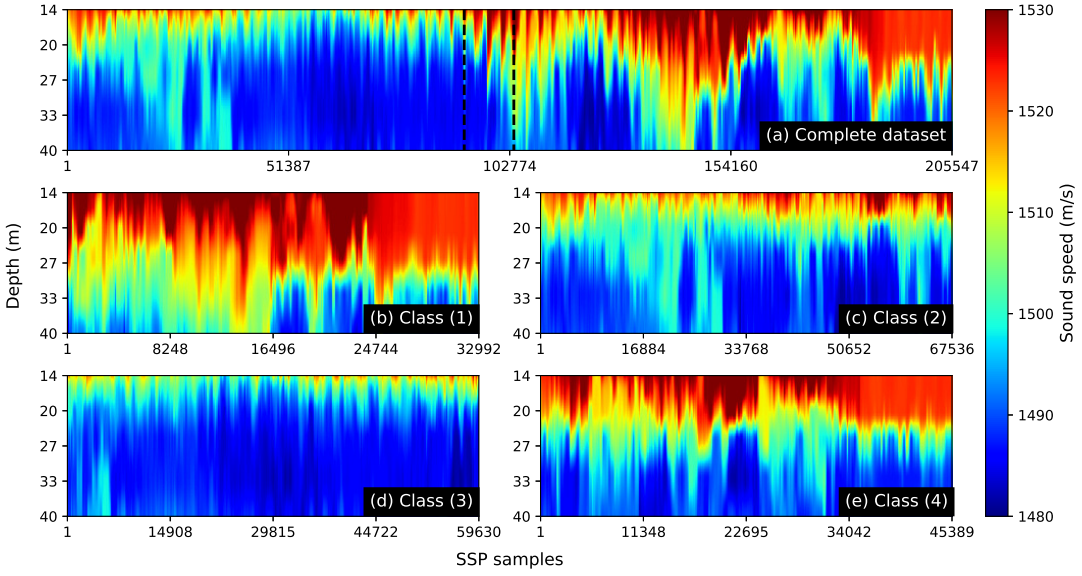

FIG. 3: Sound speed profiles from SW30 station calculated using the nine-term equation (Mackenzie 1981). The SSPs are depth-dependent as indicated with the y-axis. The x-axis shows the number of SSPs in each panel. These SSPs are extracted every 15 seconds from 01 Aug 00:00:00 to 05 Sep 16:00:00 UTC 2006. In (a), a total $n = 205,547$ of SSPs are extracted during the period described. In this work, k-medoids clustering is used to label the data. The resulting classes from (1) to (4) that are shown in panels (b), (c), (d) and (e), respectively. Notice that the SSPs on each class are not ordered in time.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

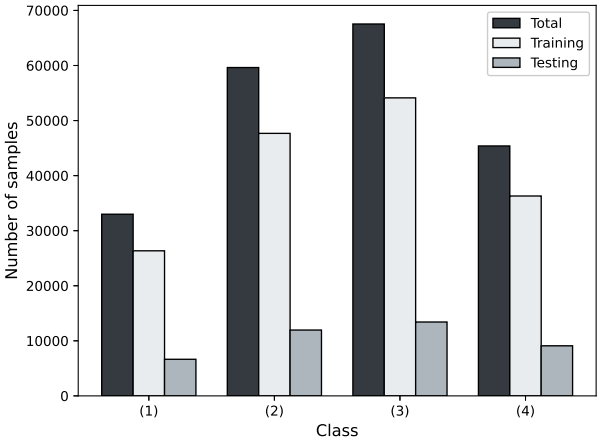

FIG. 4: Distribution of data into the four classes calculated using *k*-medoids algorithm. The original data are split into training and testing sets, whose number of samples per class are also depicted in the plot.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

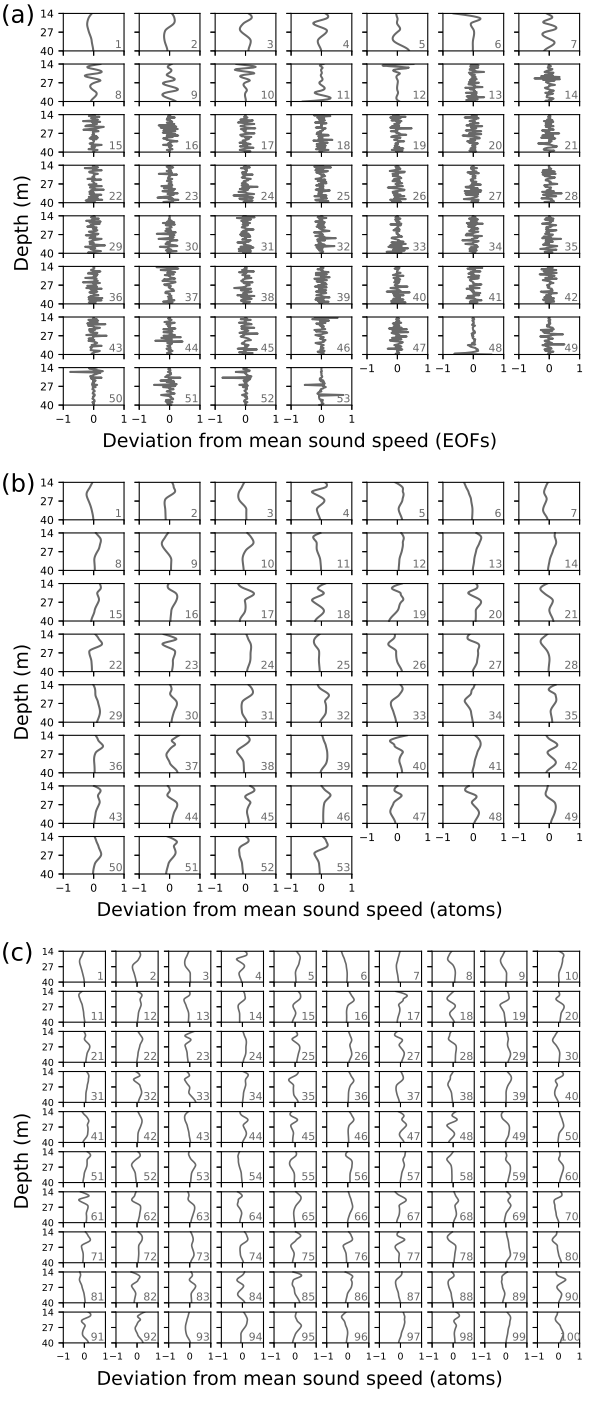

FIG. 5: Basis functions computed via (a) EOF analysis, (b) complete DL, and (c) overcomplete DL. Only the first leading-order EOFs describe variability in the SSPs, whereas variance is distributed along all atoms in the LDs.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

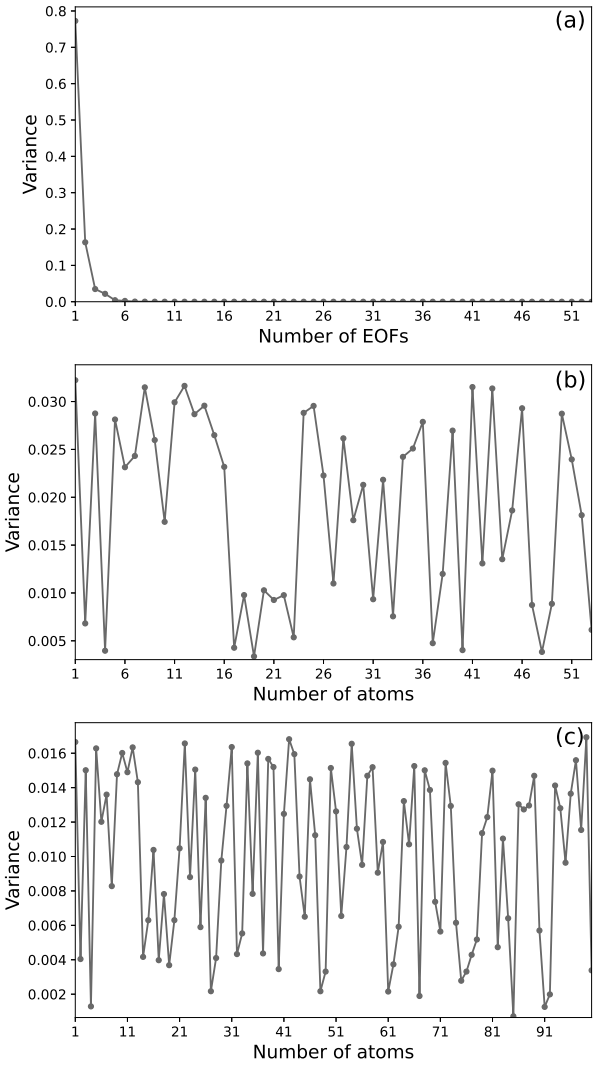

FIG. 6: SSP explained variance ratio for (a) EOFs and (b) complete LD entries with $k = 53$, and (c) overcomplete LD with $k = 100$ atoms using $n_{nz} = 3$ non-zero coefficients.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

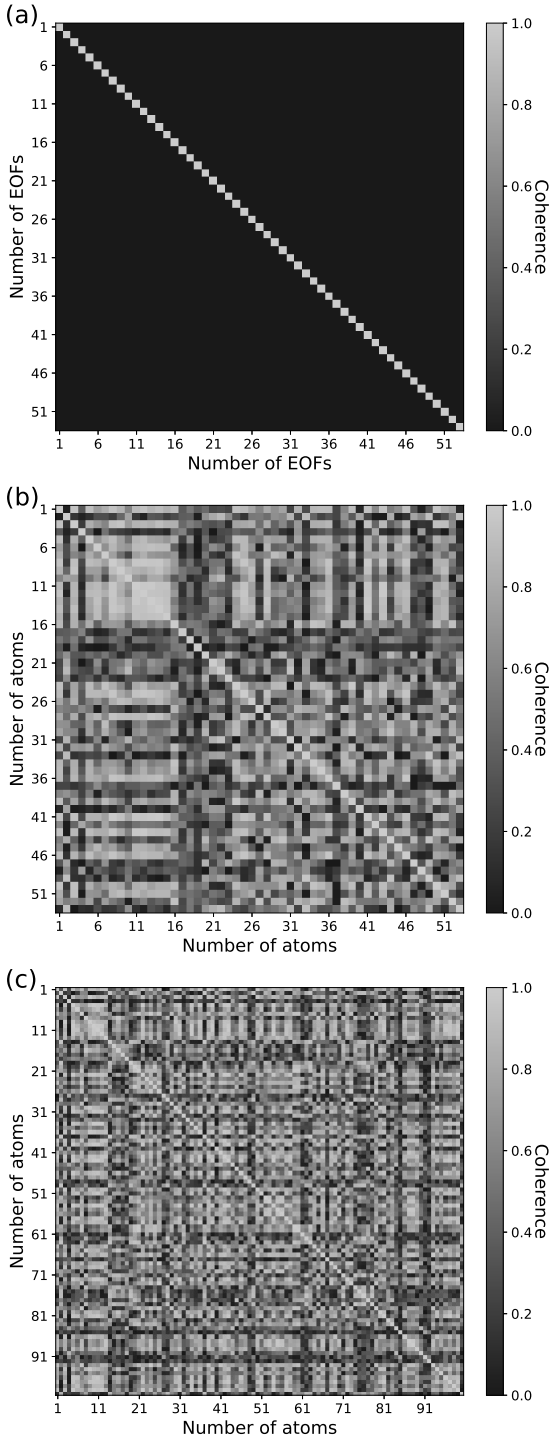

FIG. 7: Coherence matrices for (a) EOFs, (b) complete LD entries, and (c) overcomplete LD [see Fig. 5] using $n_{\mathrm{nz}} = 3$ non-zero coefficients. Eigenvectors in the EOF dictionary are orthogonal, whereas those of the LDs are not.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

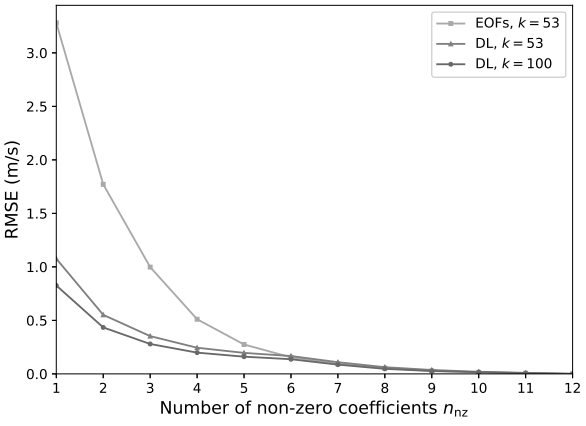

FIG. 8: Root mean square error (RMSE) of EOFs and complete LD, both with $k = 53$, and overcomplete LD with $k = 100$ when different numbers of non-zeros coefficients n_{nz} used for sparse coding.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

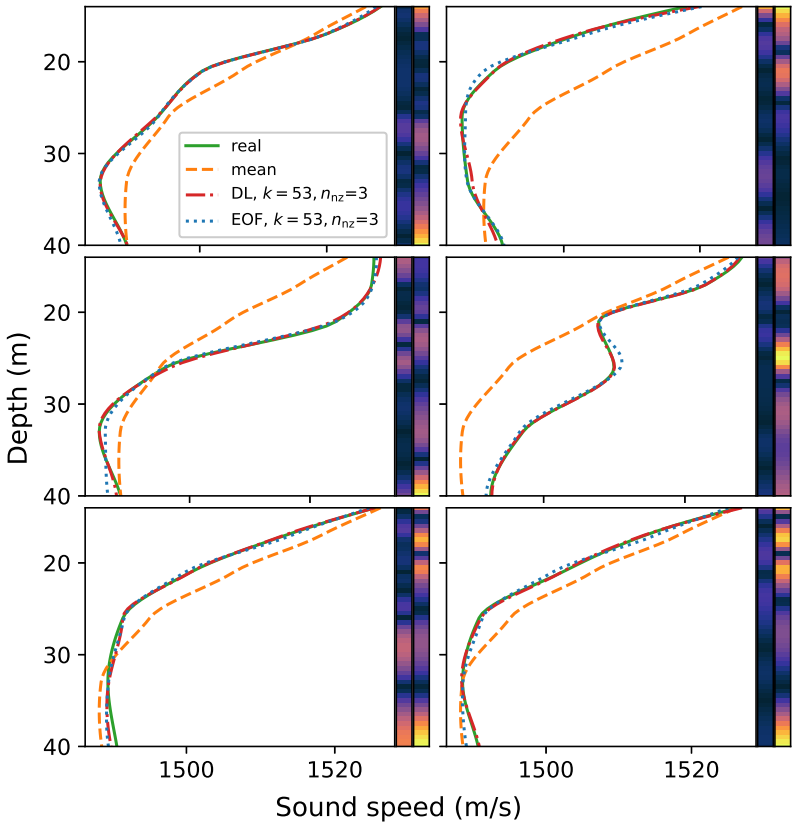

FIG. 9: Sparse representation of six SSP samples using complete DL and EOF with $n_{\text{nz}} = 3$ non-zero coefficients and $k = 53$ atoms and EOFs. For each sample, color bars present the absolute error between the sample and the reconstruction in terms of depth for LD and EOFs, respectively.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.

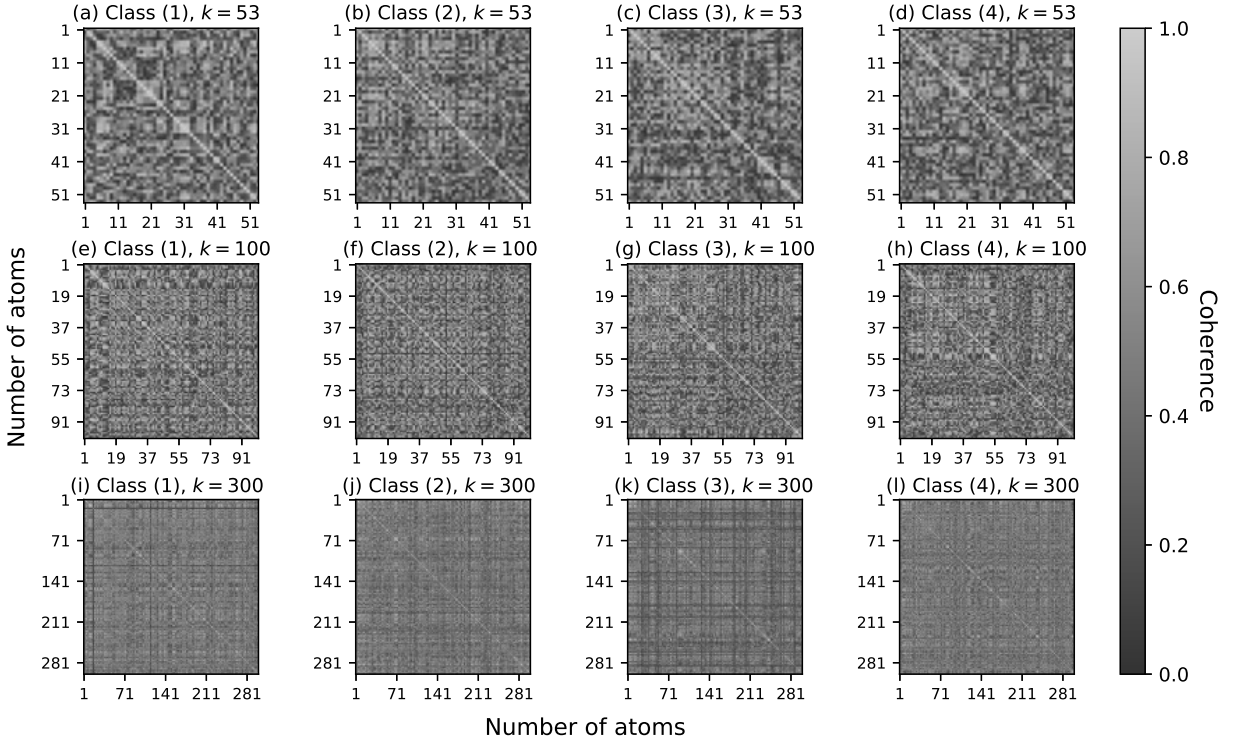

FIG. 10: Coherence of individual class-specific learned dictionaries (LD) for classification of SSPs. (a), (b), (c) and (d) are coherence for the complete LD ($k = 53$) for each class (1)-(4). (e), (f), (g) and (h) are coherence of overcomplete LDs with $k = 100$ atoms for all the classes. (i), (j), (k) and (l) are coherence for LDs with $k = 300$ for the four classes.

Accepted for publication in *Journal of Atmospheric and Oceanic Technology*. DOI 10.1175/JTECH-D-21-0090.1.