The Pennsylvania State University

The Graduate School

STOCHASTIC MODELS TO GENERATE SOUND SPEED ENSEMBLES AND CALCULATE TRANSMISSION LOSS ENSEMBLES AT THE NEW ENGLAND SHELF BREAK USING EMPIRICAL ORTHOGONAL FUNCTIONS AND SPARSE DICTIONARY LEARNING

A Thesis in

Acoustics

by

Simeon Benit

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2023

The thesis of Simeon Benit was reviewed and approved by the following:

Andrew Barnard
Professor of Acoustics
Director of the Graduate Program in Acoustics
Thesis Co-Advisor

Ying-Tsong Lin
Associate Scientist with Tenure
Woods Hole Oceanographic Institute
Technical Advisor
Special Signatory

Daniel C. Brown
Assistant Research Professor
Applied Research Lab at Penn State
Thesis Co-Advisor

James Miller
Professor and Department Chair of Ocean Engineering
University of Rhode Island
Special Signatory

ABSTRACT

The New England Shelf Break's (NESB) bathymetry and proximity to the Gulf Stream current make it one of the most complex ocean environments in the world. This makes it a valuable test case for algorithms designed to recognize patterns in large datasets. Empirical Orthogonal Function (EOF) analysis, or Principal Component Analysis (PCA), has historically been the standard statistical method for detecting spatiotemporal patterns in data. These patterns are used to reconstruct datasets accurately and efficiently. Recently, sparse dictionary learning methods like the K-SVD have begun to supplant EOF analysis for this application. Results from these techniques are valuable to underwater acoustic propagation models that require realistic environmental data input to produce useful approximations of the underwater soundscape.

Using World Ocean Database (WOD) *in situ* Conductivity-Temperature-Depth (CTD) measurements taken at the NESB, two stochastic models were developed: one based on the K-SVD, the other based on EOF analysis. Additionally, a new approach to representing sparse coefficients from the K-SVD as a normal random variable was developed. Both models were used to produce sound speed profile (SSP) ensembles. These SSP ensembles were used to calculate Transmission Loss (TL) ensembles from the Range-dependent Acoustic Model (RAM). A probability density function (PDF) of TL ensembles was produced to show the relative likelihood that a TL ensemble would fall within a certain range of TL.

The EOF stochastic model produced SSP ensembles that reliably covered SSP spatial variability at the shelf. While the EOF model did not cover variability at the shelf break, the resulting ensemble envelope was consistent. The K-SVD ensemble envelope was nondeterministic and did not reliably cover SSP spatial variability at the shelf or the shelf break. The EOF stochastic model required that missing values be filled with the depth-wise mean to function, while the K-SVD flexibly handled the imputation of zeros to compensate for missing values. TL ensembles were successfully calculated using K-SVD SSP ensembles. Ultimately, EOF analysis showed more potential as the basis for a stochastic model to inform an underwater acoustic propagation model.

TABLE OF CONTENTS

List of Figures

List of Tables

Acknowledgements

I feel profound gratitude for every person and every opportunity that led me here. My parents, friends, teachers, professors, and colleagues have all contributed in seen and unseen ways. Thank you.

This material is based upon work supported by the Office of Naval Research (ONR) under Award No. [Grant N00014-21-12416].

Disclaimer: Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Office of Naval Research.

Chapter 1

Introduction

The acoustics of the oceans is an extensively studied physical property. The results are applicable to several fields of study including climatology [1], marine biology [2], and oceanography [3]. The oceans' size, variability, and extreme ecosystems are a few of the factors that make it a challenging environment to measure, thus a difficult environment to model. This is especially the case in areas like the New England Shelf Break (NESB), where the combination of sharply varying bathymetry and the Gulf Stream current make it one of the most complex ocean regions [4]–[8]. In this research the choice to separate dictionaries as a function of bathymetry was made to address NESB variation. A cross section of the NESB is shown in Fig. 1.1 The shelf break jet (SBJ) coupled with the change in bathymetry at the shelf break front create a natural separation in the water between the shelf and the shelf break [9]. The differences between the two volumes of water include salinity, temperature, and flow direction. A complete description of the NESB is an ongoing research topic and new discoveries about its complex oceanography are still made as new data acquisition techniques are applied [10].

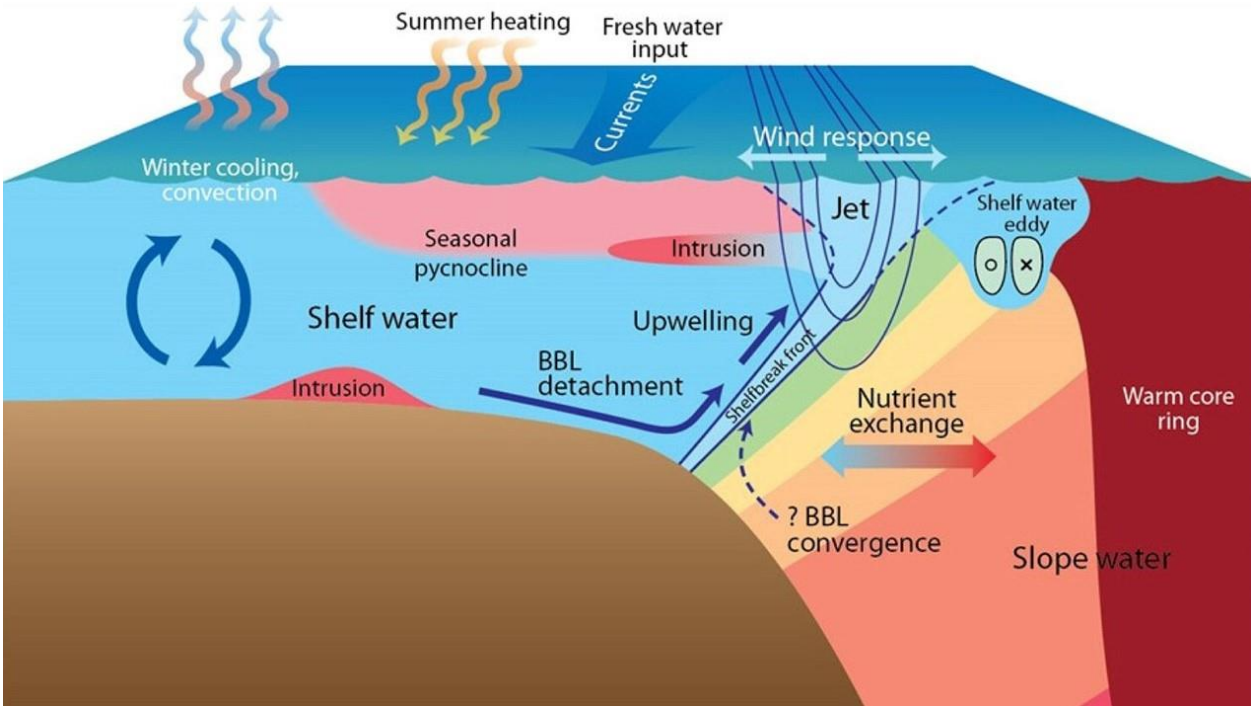


Figure 1.1: Cross section of the ocean at the NESB that depicts the complicated oceanography found in this region [11].

Because of large datasets over time and space, methods to identify patterns in the mass of ocean measurements are essential to developing useful underwater acoustic models. For example, geoacoustic inversion is a computationally intensive inverse problem used to characterize the seabed by analyzing changes in known acoustic signals as they reflect off the ocean floor. The forward models used for this require information about the medium, such as sound speed, in which the signals propagate to accurately predict acoustic behavior [12], [13]. Inverting for sound speed would add another computationally expensive task to the geoacoustic inversion problem. This makes models that can accurately approximate environmental parameters valuable.

Ocean temperature profiles are an invaluable source of information about the oceans. Vast efforts to sample the oceans spatially and temporally are made by organizations and governments worldwide [14], [15]. Devices like the Conductivity-Temperature-Depth (CTD) cast (Fig. 2.1) and Expendable Bathythermographs (XBTs) provide *in situ* temperature profile measurements of the oceans, as well as depth readings that correspond to each sample. CTD

casts include salinity measurements, which makes their data an ideal source for ocean sound speed equations [16]. These measurements are vital for effective underwater acoustic models. Further information on the data sources for this thesis can be found in chapter four.

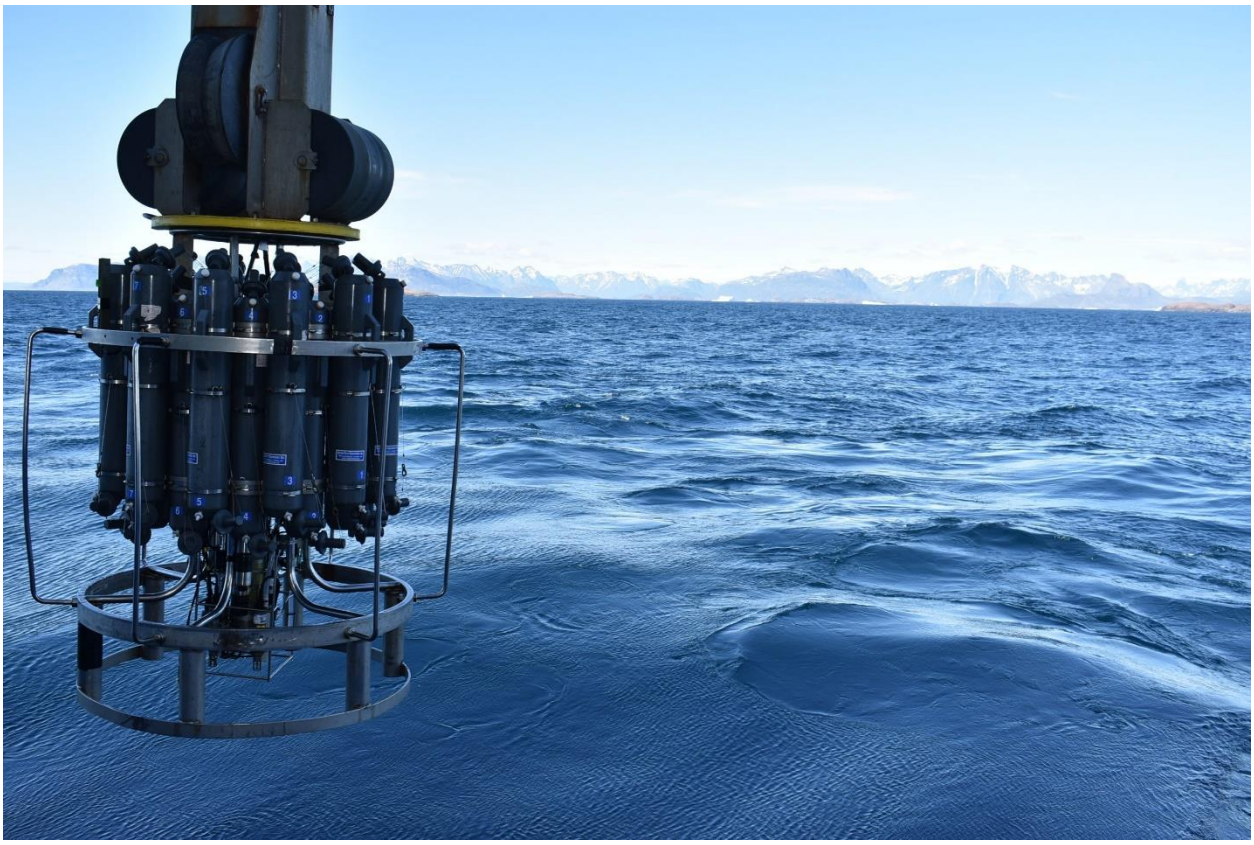


Figure 1.2: A Conductivity-Temperature-Depth (CTD) cast used for *in situ* temperature measurements in the ocean [17].

This research focuses on applying two stochastic models to ocean data from the NESB. The models were independently based on Empirical Orthogonal Functions (EOFs) and the K-SVD, which is an unsupervised dictionary learning (DL) algorithm. The goal of this thesis is to

generate sound speed ensembles[1] informed by the stochastic models and calculate transmission loss (TL) based on the SSP ensembles.

Historically, EOF analysis is a typical second-order statistical method used in oceanography to identify patterns in high-dimensional data. In statistics, EOF analysis is referred to as Principal Component Analysis (PCA) [20], [21]. EOFs are most commonly used in meteorology [22] but are regularly used in other fields including geophysics [23] and climatology [24]. While EOFs are highly effective, machine learning algorithms have recently risen in prominence, primarily due to increases in computational capability. The K-SVD is an iterative method that is a generalization of K-means clustering. It is a two-step process that attempts to greedily find the best representations (coefficients) of an input signal, then update the initial dictionary, as well as the representations established in the first stage. It has been shown that the K-SVD can reconstruct sound speed profiles (SSPs) with greater accuracy, while using less information [25]. Similarly, this thesis investigates the performance of the DL-based stochastic model and the EOF-based stochastic model on data taken at the NESB. Detailed explanations of both these algorithms are given in later chapters.

The forward model used here is the Range-dependent Acoustic Model (RAM) [26]. It is a Parabolic Equation (PE) model that remains an efficient and accurate option for wide-angle underwater acoustic propagation. An in-depth treatment of its theoretical basis can be found in chapter three. Its application in this project is to demonstrate the use of sound speed ensembles created by a stochastic model based on the K-SVD algorithm. This is done by generating TL ensembles as a function of K-SVD produced sound speed ensembles. The TL loss ensembles are also used to construct a TL probability density function that indicates the relative likelihood of a TL ensemble falling within a certain TL interval across the range.

[1] The term “ensembles” was first used in the field of statistical physics by J. Willard Gibbs [18]. Generally, it is defined as statistical microstates of a system [19].

Chapter 2

Empirical Orthogonal Functions and Dictionary Learning

2.1 Empirical Orthogonal Functions

Empirical Orthogonal Function (EOF) analysis is a widely used dimensionality reduction approach. The term "Empirical Orthogonal Functions" was coined in the context of atmospheric science [27], but it is used in various fields of study including underwater acoustics. At its core it is a statistical technique also known as Principal Component Analysis (PCA), where EOFs are the principal components (PCs) [20], [21], [28]. The goal is to compress a dataset while preserving as much variability as possible. EOFs of the dataset are new variables containing elements that represent variance, and thus represent a majority of the statistical information in the dataset. An alternative formulation of its purpose is to find spatiotemporal patterns in large datasets that capture the variance.

In this application, given a dataset $\boldsymbol{Y} \in \mathbb{R}^{m \times t}$, where m is the spatial dimension (depth in this case) and t is the number of discrete observations (SSPs), the depth-wise mean is subtracted from $\boldsymbol{Y}$. The EOFs are the left singular vectors of demeaned $\boldsymbol{Y}$ and are found by applying an eigendecomposition to the covariance matrix $\boldsymbol{YY^T}$

$$\boldsymbol{YY^T} = \boldsymbol{W\Gamma^2 W^T}, \tag{1}$$

where the diagonal elements of $\boldsymbol{\Gamma^2} = [\gamma_1^2, \gamma_2^2, \dots, \gamma_L^2] \in \mathbb{R}^{L \times L}$ are the eigenvalues (variances) of $\boldsymbol{YY^T}$ and $\boldsymbol{W} \in \mathbb{R}^{m \times L}$ contains the eigenvectors of $\boldsymbol{YY^T}$. A basis is formed by $[w_1, w_2, \dots, w_L]$ when $L = m$. This occurs when $t > m$. A special property of the eigendecomposition of a diagonalizable covariance matrix, such as $\boldsymbol{YY^T}$, is that its eigenvectors are the singular vectors of the underlying data set, $\boldsymbol{Y}$. That is, $\boldsymbol{W}$ and $\boldsymbol{W^T}$ are the left and right singular vectors of $\boldsymbol{Y}$,

respectively, and $\boldsymbol{W}$ contains EOFs. An additional special property of a diagonalizable covariance is that the square root of the eigenvalues γ are the singular values of $\boldsymbol{Y}$, which will be important for the stochastic model.

The variance is used to determine the effectiveness of the leading-order EOFs. This measure is important in selecting the number of leading-order EOFs to retain when reconstructing $\boldsymbol{Y}$. The eigenvalues give how much variance is contained in the related eigenvector.

2.2 Dictionary Learning

Sparse Dictionary Learning is a form of unsupervised, representation learning used in many fields of study including compressed sensing ([29], [30]), signal recovery ([31], [32]), and image classification ([33], [34]). Generally, sparse dictionary learning finds an overcomplete set of vectors and their scalar weights to represent an input signal $\boldsymbol{x} \in \mathbb{R}^M$ efficiently. This set of vectors is referred to as a dictionary and comprises columns called atoms $\{\boldsymbol{d}_k\}_{k=1}^K$. The atoms correspond to a vector containing coefficients $\boldsymbol{a} \in \mathbb{R}^K$ that are assumed to be sparse. The dictionary $\boldsymbol{D} \in \mathbb{R}^{M \times K}$ is considered overcomplete when $K > M$, K being the number of atoms. Redundancy in $\boldsymbol{D}$ allows for a more accurate approximation of the input signal relative to an undercomplete ($K < M$) or complete basis ($K = M$); however, overcompleteness implies linear dependence. Given a set of signals $\{\boldsymbol{x}_j\}_{j=1}^N$, where N is the number of signals, dictionary learning algorithms attempt to solve the optimization problem

$$min_{\boldsymbol{D},\boldsymbol{A}} \|\boldsymbol{X} - \boldsymbol{D}\boldsymbol{A}\|_F^2 \; s.t. \; \forall \; 1 \leq j \leq N, \|\boldsymbol{a}_j\|_0 \leq s_0, \tag{3}$$

where $\boldsymbol{X}^{M \times N}$ is a matrix of training data and $\boldsymbol{A} \in \mathbb{R}^{K \times N}$ is a sparse coefficient matrix with $\{\boldsymbol{a}_j\}_{j=1}^N$ column vectors. The maximum number of nonzero coefficients in the representation is s_0. This constraint is expected to be small (i.e., $s_0 \ll M < K$). F denotes the Frobenius norm

[35]. Widely used algorithms for this purpose are Method of Optimal Directions (MOD) ([36], [37]), Online Dictionary Learning ([38], [39]), and the K-SVD [40].

Much of what follows in this section is written to provide the reader with context surrounding the development of the K-SVD. Understanding how the K-SVD fits into the history of machine learning techniques encourages better understanding of its intricacies. Vector Quantization, K-Means, and Method of Optimal Directions are discussed as examples of early dictionary training algorithms that led to the K-SVD. Furthermore, section 2.2.3.1 includes descriptions of three progenitors to the sparse coding algorithm used in this thesis — Orthogonal Matching Pursuit (OMP) [41], Matching Pursuit (MP) [42], and Compressive Sampling Matched Pursuit (CoSAMP) [31]. The sparse coding algorithm used in this thesis is *Coefficient Reuse Orthogonal Matching Pursuit* (CoefROMP) [43].

2.2.1 Vector Quantization and K-Means

Vector Quantization (VQ) uses a codebook (dictionary) with K-codewords (columns) to represent a large set of signals [44]. K-means is the most used algorithm in VQ to solve the clustering problem ([45], [46]) — it is an iterative method that searches for the optimal codebook for VQ. The codebook is trained using K-means, which accomplishes the task by partitioning input signals into K-clusters by nearest neighbor. This is normally done using an l^2-norm as a distance measure to assign signals to their respective cluster.

It includes two steps: 1) sparse coding stage, where, given a codeword, assign signals to their nearest neighbor, and 2) update the new codeword to fit the signal. This structure illustrates its similarity to the K-SVD. How these two relate is explained further in section 2.2.3.

2.2.2 Method of Optimal Directions

Method of optimal directions (MOD) is a pioneering method of dictionary training [36], [37], [47]. Like the K-SVD that came after it, it includes a sparse coding stage and a dictionary update stage. Sparse coding is accomplished iteratively through every signal and can be performed by any typical sparse coding technique. The algorithm is inspired by the Generalized Lloyd Algorithm (GLA) [44]. The major difference between MOD and GLA is that MOD does not require that each new dictionary perform better following an update. It attempts to minimize Eqn. (1) during the dictionary update, but the optimization problem is non-convex, which limits the optimum result to a local minimum. To optimize $\boldsymbol{D}$, the algorithm solves $\boldsymbol{D} = \boldsymbol{Y}\boldsymbol{M}^{+}$, where $\boldsymbol{M}^{+}$is the Moore-Penrose pseudo-inverse [48]. It is important to note that every column of $\boldsymbol{D}$ is required to be normalized.

2.2.3 K-SVD

The K-SVD algorithm is an iterative method that, given an initial ℓ^2 normalized **D**, computes coefficients during a sparse coding stage as part of a sparse coefficient matrix. The atoms in **D** are then modified to best reduce mean squared error by a singular value decomposition. All of this to solve Eqn. (3). The K-SVD is a generalization of the K-means algorithm. K-means can be thought of as approximating a data set using a single atom and a coefficient that must equal one. However, in gain-shape VQ [44], the coefficient is allowed to vary. K-SVD is a generalization in the sense that signals can be approximated using a linear combination of several vectors. Furthermore, K-SVD allows for more than one coefficient, all of which can vary. The sparse coding stage and dictionary update stage in K-means are decoupled. This means that $\boldsymbol{D}$ is fixed during the sparse coding stage, and $\boldsymbol{A}$ is fixed during the dictionary update stage. The generalization of K-means extends to this constraint by allowing $\boldsymbol{A}$ to update during the dictionary update stage. This is done to ensure the next column (atom) update is based on more relevant coefficients.

2.2.3.1 Sparse Coding

Sparse coding is a computational method embedded within dictionary learning algorithms that attempts to represent an input vector $\boldsymbol{x}$ as a linear combination of dictionary atoms $\boldsymbol{d}_k$ and the associated weights $\boldsymbol{a}$:

$$\boldsymbol{x} = \sum_k \boldsymbol{a_k}\, \boldsymbol{d}_k\,. \tag{4}$$

The dictionary $\boldsymbol{D}$ is assumed to be linearly dependent and redundant. The sparse coding problem is often formulated in one of two ways,

$$min_{\boldsymbol{a}_j} \left\|\boldsymbol{x}_j - \boldsymbol{D}\boldsymbol{a}_j\right\|_2^2 \, s.t. \left\|\boldsymbol{a}_j\right\|_0 \leq \; s_0 \tag{5}$$

or

$$min_{\boldsymbol{a}_j} \left\|\boldsymbol{a}_j\right\|_0 \, s.t. \left\|\boldsymbol{x}_j - \boldsymbol{D}\boldsymbol{a}_j\right\|_2^2 \leq \varepsilon, \tag{6}$$

where ε denotes a fixed error threshold. This thesis is focused on Eqn. (5). This task is commonly completed by the pursuit algorithms mentioned in section 2.2. As a reminder, these are Orthogonal Matching Pursuit (OMP) [41], Matching Pursuit (MP) [42], and Compressive Sampling Matched Pursuit (CoSaMP) [31]. These methods are classified as greedy algorithms. Greedy algorithms sequentially select a locally optimal choice with the goal of a global solution.

Matching Pursuit

Matching Pursuit (MP) is a greedy algorithm that iterates through each signal $\boldsymbol{x}$ in search of an estimate $\hat{\boldsymbol{x}}$ [42], [49]. A K atom approximation takes the form of Eqn. (4). Like other pursuit algorithms, the goal is to approximately solve the sparse coding problem shown in Eqn. (5). Assuming $\boldsymbol{D}$ is normalized, a dictionary atom that maximizes the absolute value of the inner product between the signal and each atom is selected. The residual $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{d}\boldsymbol{h}$, where $\boldsymbol{h}$ is the

maximum absolute inner product, is computed, and these steps are repeated until a stopping criterion is met. MP may introduce a component that is non-orthogonal to the span of the first atom selected. This characteristic is removed from the orthogonal matching pursuit. The largest differences between MP and OMP are 1) OMP includes a more complicated approach to finding sparse representations and 2) as more coefficients are selected to represent the signal, the previously identified coefficients for that signal are updated.

Compressive Sampling Matching Pursuit

Compressive sampling matching pursuit (CoSaMP) is based on OMP [31], [41]. Like OMP, it is a greedy pursuit method, which means it builds a solution iteratively by finding a locally optimal choice at each step. The main idea in this method is the selection of the largest components of a signal. Separate from the iteration through signals is an iteration that is governed by some halting criterion. To initialize, the algorithm identifies the largest components of the original signal, solves a least-squares problem that approximates the signal, retains the largest s (sparsity constraint) entries, and finds a residual. After this first iteration, a support (nonzero locations) merging step takes place after the largest signal component identification step. The supports from the previous iteration are merged with the new approximation supports until the halting criterion is met. This support preservation and a pruning step used in CoSaMP are carried over to the pursuit algorithm used in this thesis. They are explained further in the next section.

Orthogonal Matching Pursuit and Coefficient Reuse Orthogonal Matching Pursuit

To understand *Coefficient Reuse Orthogonal Matching Pursuit* (CoefROMP) [43], it is instructive to introduce OMP. OMP is one of the most commonly used pursuit algorithms available to solve the sparse coding problem. Its name is derived from the fact that residual after each iteration is orthogonal to the span of previously selected atoms. This avoids the problem of selecting linear dependent atoms observed in MP. The standard OMP begins by setting the

residual equal to the input vector $\boldsymbol{r} = \boldsymbol{x}$. The maximum absolute value of the inner product between $\boldsymbol{r}$ and $\boldsymbol{D}^T$ is

$$\max_j |\langle \boldsymbol{r}, \boldsymbol{D}^T \rangle|. \tag{7}$$

The atoms in $\boldsymbol{D}$ that correspond to this value are selected. The next step is to solve for coefficients using a least squares solver to approximate $\boldsymbol{a} = \boldsymbol{D}^{\dagger}\boldsymbol{x}$, where $\boldsymbol{D}^{\dagger}$ is the Moore-Penrose pseudoinverse [48]. The residual $\boldsymbol{r} = \boldsymbol{x} - \boldsymbol{D}\boldsymbol{a}$ is calculated and examined by a halting criterion. If the halting criterion is not met, the next iteration begins. Pseudo code in Table 2.1 is a standard OMP example.

OMP($\boldsymbol{D}$, $\boldsymbol{x}$, s)

Input: Dictionary $\boldsymbol{D}$, signal $\boldsymbol{x}$, sparsity s
Output: s-sparse representation $\boldsymbol{a}$ of $\boldsymbol{x}$

Step 1: initialize residual $\boldsymbol{r}_0 = \boldsymbol{x}$ and declare variable to hold indices $I_0 = 0$

for $i = 1, 2, \ldots$ **do**

Step 2: $I_i = I_{i-1} \cup argmax\ \lvert\langle \boldsymbol{D}^T,\ \boldsymbol{r}_{i-1}\rangle\rvert$	{merge indices of max absolute inner product}
Step 3: $\boldsymbol{a}_i = \boldsymbol{D}_{I_i}^{\dagger}\boldsymbol{x}$	{signal estimation by least-squares}
Step 4: $\boldsymbol{r}_i = \boldsymbol{x} - \boldsymbol{D}_{I_i}\boldsymbol{a}_i$	{update residual}
Step 5: if the stopping criterion is met, **return**. If not, **continue**	

end

Table 2.1: OMP pseudo code.

Halting criterion varies. An example is evaluating the sum of residual squared relative to an error threshold ϵ:

$$\sum \boldsymbol{r}^2 < \epsilon. \tag{8}$$

CoefROMP improves the OMP by using an augmentation step and a pruning step from CoSaMP, as well as a "warm-start". Pruning refers to storing the largest s entries from the least-squares solution. This ensures the best s-sparse approximation of $\boldsymbol{a}$**.** The augmentation step adds

coefficients from a previous iteration to a new least squares approximation $\boldsymbol{a}_r = \boldsymbol{D}\boldsymbol{r}$. A "warm-start" considers the largest $s/3$ coefficients from the previous iteration. This is done because the objective function remains at the same height as the previous iteration. The improvements to convergence speed and RMSE brought by a warm-start are not observable without the augmentation and pruning steps [43].

2.2.3.2 Dictionary Update Stage

The dictionary update stage is more complicated than the sparse coding stage. The algorithm as it was originally devised [40] is explained first. For clarity, the following is a list of variables and what their associated indices represent:

1) $\boldsymbol{d}_k \coloneqq$ dictionary atom to update
2) $\boldsymbol{d}_i \coloneqq$ arbitrary dictionary atom
3) $\boldsymbol{a}_k \coloneqq$ coefficient matrix row vector corresponding to $\boldsymbol{d}_k$
4) $\boldsymbol{a}_i \coloneqq$ coefficient row vector corresponding to $\boldsymbol{d}_i$
5) $\boldsymbol{a}_k^S \coloneqq$ nonzero coefficients in row k
6) $\boldsymbol{a}_i^S \coloneqq$ nonzero coefficients in row i

The kth atom is the dictionary atom undergoing an update, and the index i refers to the counter variable. The variant of the algorithm used here is improved [43] from the original, which is also explained. To begin this discussion, it is important to be very clear about how the dimensions of $\boldsymbol{X}$, $\boldsymbol{D}$, and $\boldsymbol{A}$ are related to each other. Columns of dictionary $\boldsymbol{D}$ (atoms) directly correspond to the rows of $\boldsymbol{A}$. When nonzero coefficients are found, they are linked to a dictionary atom $\boldsymbol{d}_k$. The subscript k and superscript S together refer to the nonzero coefficients in the kth row of $\boldsymbol{A}$. The sparse representations $\boldsymbol{a}_j$ in the jth column of $\boldsymbol{A}$ are generated with respect to the signal $\boldsymbol{x}_j$ they represent (hence the term *representations*) from the jth column of $\boldsymbol{X}$. Keeping this in mind is important as one goes through the steps of the dictionary update stage outlined in Table 2.2.

The K-SVD dictionary update stage updates the atoms $\{\boldsymbol{d}_k\}_{k=1}^{K}$ of $\boldsymbol{D}$ sequentially at each i update step, as well as the coefficients $\boldsymbol{a}_j \in \boldsymbol{A}^i$. The sought after improvement is finding an

updated $\boldsymbol{D}$ and $\boldsymbol{A}$ that best reduces mean-squared error. Rewriting the penalty term in the objective function (Eqn. (3)), the result is,

$$\|\boldsymbol{X} - \boldsymbol{DA}\|_F^2 = \left\|\boldsymbol{X} - \sum_{i=1}^{K} \boldsymbol{d}_i \boldsymbol{a}_i\right\|_F^2$$

$$= \|\boldsymbol{E}_k - \boldsymbol{d}_k \boldsymbol{a}_k\|_F^2 , \tag{9}$$

where

$$\boldsymbol{E}_k = \boldsymbol{X} - \sum_{i \neq k} \boldsymbol{d}_i \boldsymbol{a}_i .$$

The summation term is a decomposition of the matrix multiplication $\boldsymbol{DA}$ into a summation of K rank-1 matrices. This is a sum of outer products. $\boldsymbol{E}_k = [\boldsymbol{e}_{i,1}, \dots, \boldsymbol{e}_{i,N}] \in \mathbb{R}^{M \times N}$ is an error matrix that measures error between $\boldsymbol{X}$ and its reconstructions when the kth atom is excluded. Sparsity is not enforced on the rows of $\boldsymbol{X}$. Addressing the many zeros in $\boldsymbol{E}_k$ and $\boldsymbol{a}_k$ is necessary for the singular value decomposition (SVD) to effectively update $\boldsymbol{d}_k$ and $\boldsymbol{a}_k$. This is done by defining $\boldsymbol{E}_k^S$ and $\boldsymbol{a}_k^S$ as nonzero elements of $\boldsymbol{E}_k$ and $\boldsymbol{a}_k$

$$\boldsymbol{E}_k^S = \left\{\boldsymbol{e}_{k,l} \middle| \forall_l, \boldsymbol{a}_k^l \neq 0\right\}, \ \boldsymbol{a}_k^S = \{|\forall_l \, \boldsymbol{a}_k^l \neq 0\}. \tag{10}$$

This manipulation allows for the use of the SVD to decompose $\boldsymbol{E}_k^S$ into $\boldsymbol{E}_k^S = \boldsymbol{U\Sigma V}^T$, where the first column of $\boldsymbol{U}$ is the updated dictionary atom $\widehat{\boldsymbol{d}}_k$ and the product of $\boldsymbol{\Sigma}(1,1)$ and the first column of $\boldsymbol{V}$ is the updated coefficient row vector $\widehat{\boldsymbol{a}}_k^S$.

K-SVD

Goal: Find a dictionary along with sparse representations that best approximate a training data set by solving

$$min_{\boldsymbol{D},\boldsymbol{A}}\|\boldsymbol{X}-\boldsymbol{D}\boldsymbol{A}\|_F^2 \; s.t. \forall \; 1 \; \leq j \; \leq N, \left\|\boldsymbol{a}_j\right\|_0 \; \leq s_0$$

Step 1: initialize $\boldsymbol{D}$ with ℓ^2 normalized columns

for $j = 1, 2, \ldots$ **do** until halting criterion

Step 2: **Sparse Coding Stage** use pursuit algorithm to solve the sparse coding problem for N signals

$$\widehat{\boldsymbol{a}}_j = \; min_{\boldsymbol{a}_j}\left\|\boldsymbol{x}_j-\boldsymbol{D}\boldsymbol{a}_j\right\|_2^2 \; s.t. \left\|\boldsymbol{a}_j\right\|_0 \leq \; s_0 \text{ for } j = 1, 2, \ldots, N$$

Step 3: **Dictionary Update Stage** iterate through K columns of $\boldsymbol{D}$ for $i = 1, 2, \ldots K$ and update it by

- Defining signals that use selected k atom
- Determine error $\boldsymbol{E_k} = \boldsymbol{X} - \sum_{i \neq k} \boldsymbol{d}_i \, \boldsymbol{a}_i^S$
- Choosing columns of $\boldsymbol{E}_k$ that correspond to signals that use selected k atom to get $\boldsymbol{E}_k^R$
- Apply SVD decomposition and get $\boldsymbol{E}_k^R = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$. $\boldsymbol{U}$ is the updated dictionary atom $\widehat{\boldsymbol{d}}_k$ and $\boldsymbol{S}(1,1)\boldsymbol{V}^T$ is the updated sparse representation vector $\widehat{\boldsymbol{a}}_k^R$

end

Table 2.1: Pseudo code for the K-SVD algorithm.

2.2.3.3 Improved Dictionary Update Stage

The improved dictionary update stage found in CoefROMP includes Dictionary Update Cycles (DUCs). The earlier method iterates once through each dictionary to achieve convergence. The dictionary may go through several update cycles to further minimize Eqn. (3). When combined, the DUCs and the adjustments made in CoefROMP relative to OMP have been shown to produce faster convergence to a steady-state solution [43].

Chapter 3

Parabolic Equation

3.1 Historical Background

Application of Parabolic Equation methods (PE) to wave-propagation problems began with electromagnetic waves in the 1940s [50] to calculate diffraction caused by the spherical shape of the earth. Since this original use, the PE has been applied to problems in laser beam propagation [51], plasma physics [52], and seismic wave propagation [53]. It was extended to ocean acoustics in 1973 by Hardin and Tappert [54]. PEs fit into the family of marching partial differential equations, which makes them especially well-suited for range-dependent problems. The ultimate goal of PE approximation methods is to closely approximate the acoustic field during initial analysis to develop approximate equations that are ideal for efficient, high-speed computation. The PE is one of the most efficient and commonly used forward models for range-dependent problems in ocean acoustics.

The Standard Parabolic Equation (SPE) is an important steppingstone in the development of PE methods in underwater acoustics. There are different approaches that lead to the SPE; however, the derivation followed later in this document follows Tappert [55]. The SPE is important because it lends itself to efficient numerical solutions like the Split-Step Fourier algorithm (SSF), and thus efficient computation. The SSF algorithm allowed for larger range steps, which lead to less computation relative to commonly used finite-difference solutions. This solution is discussed further in section 3.3 of this chapter. The derivation begins with the frequency domain, elliptic wave equation — the Helmholtz equation — for a constant-density medium in cylindrical coordinates (r, φ, z):

$$\frac{\partial^2 p}{\partial r^2} + \frac{1}{r}\frac{\partial p}{\partial r} + \frac{\partial^2 p}{\partial z^2} + k_0^2 n^2 p = 0. \tag{11}$$

The variable $p(r, z)$ is the acoustic pressure, the reference wavenumber is $k_0 = \frac{\omega}{c_0}$, and the index of refraction is $n(r, z) = \frac{c_0}{c(r,z)}$. The variables r and z represent range and depth, respectively. Assuming azimuthal symmetry (axisymmetric), there is no dependence on φ, which is the coordinate related to circumferential change.

While the SPE is a powerful tool, its accuracy is only reliable for narrowband propagation angles relative to the horizontal. It is sometimes referred to as the standard narrow-angle equation because of its reliance on the narrowband approximation. This assumes an angle of propagation less than 16° with respect to the horizontal. Alternatively, there are wide-angle propagation equations available using the Generalized Parabolic Equation (GPE). The derivation using the operator convention established in [56] provides a basis for many wide-angle approximations. The method used to develop RAM utilizes the GPE operator formalism, allowing it to handle wide-angle propagation. For a detailed review of the development of PEs within underwater acoustics up to 1982, refer to [56].

The forward model used here is the Range-dependent Acoustic Model (RAM) developed by Michael D. Collins [57]. While the model is range-dependent, the medium is split into range independent segments.

3.2 Standard PE and Generalized PE

The development of the SPE rests on the following assumptions outlined in [58]:

1) Constant-density medium
2) Harmonic point source — exp $(-i\omega t)$
3) Azimuthal symmetry — no φ dependence
4) An outgoing cylindrical wave solution that resembles the propagation of a single adiabatic mode — $p(r,z) = \psi(r,z)H_0^{(1)}(k_0 r)$
5) An envelope function that varies slowly with range — $\psi(r,z)$
6) Propagation takes place in the farfield — $k_0 r \ll 1$
7) The small-angle (paraxial) approximation — $\frac{\partial^2 \psi}{\partial r^2} \ll 2ik_0 \frac{\partial \psi}{\partial r}$

After applying the assumptions to Eqn. (11), the result is the SPE:

$$2ik_0 \frac{\partial \psi}{\partial r} + \frac{\partial^2 \psi}{\partial z^2} + k_0^2(n^2 - 1)\psi = 0. \tag{12}$$

These assumptions are carried into the derivation of the GPE except for the paraxial approximation. This is what separates the SPE from the GPE. The GPE is well-suited for wide-angle propagation, in part, because it does not include the paraxial approximation. The derivation underlying RAM is similar to the GPE method referenced in the introduction [56]. To avoid redundancy the derivation of the GPE will be left for the RAM section. An abbreviated example is done here to illustrate the differences in the SPE and GPE. The derivation in [58] begins by defining the depth operator Q and a radial derivative P

$$Q = \sqrt{n^2 + \frac{1}{k_0^2}\frac{\partial^2}{\partial z^2}}, \quad P = \frac{\partial}{\partial r}.$$

Rewriting the elliptic equation using the operator gives

$$[P^2 + 2ik_0P + k_0^2(Q^2 - 1)]\psi = 0. \quad (13)$$

To get to the one-way wave equation, this expression is factored into outgoing and incoming components. This separation allows the selection of only the outgoing wave component. Assuming range-independent media where n does not depend on r, a commuter term that materializes during factorization can be ignored. The expression becomes

$$P\psi = ik_0(Q - 1)\psi. \quad (14)$$

Substituting in the contents of the operators, the equation becomes

$$\frac{\partial\psi}{\partial r} = ik_0\left(\sqrt{n^2 + \frac{1}{k_0^2}\frac{\partial^2}{\partial z^2}} - 1\right)\psi. \quad (15)$$

What makes the GPE and its operator convention important is the many available approximations to the operator Q. Jensen et al. [58] provide an example of how a series expansion of the operator can be used to derive the SPE. Retaining more expansion terms in Q allows for wide-angle propagation.

3.3 Solutions to the PE

3.3.1 Split-Step Fourier Algorithm

The SSF algorithm was first introduced into ocean acoustics by Hardin and Tappert [54] for long-range, narrow-angle propagation. The split-step algorithm is desirable because of its efficiency relative to PE solutions like finite differences (FD). The efficiency of SSF refers to lower computational effort necessary to process large range steps versus the comparably fine range steps FD solutions require. The SSF compromise lies in the physics it must ignore to remain efficient. An example is the necessity to consider bottom interactions negligible. Strong

bottom interactions require the computational grid (Δr, Δz) to be unreasonably fine. This negatively impacts the efficiency of the split-step, which is its primary advantage as a PE solution. The SSF also relies on sound speed varying weakly along the propagation path.

As mentioned in the beginning of this section, SSF is used to solve the SPE, which includes the important small-angle approximation that separates it from the GPE; however, wide-angle PE SSF solutions have been developed beginning with Chapman and Thomson [59]. They apply a square root operator approximation introduced by Feit and Fleck [60] to the split operator form outlined in the GPE derivation in section 3.2.

Wide-angle propagation demands more computation and an algorithm that accurately handles short-range, deep-water problems. Shallow water problems also fall into this category. Finite differences or finite elements (FE) are the PE solvers typically used in these scenarios. While they do not exhibit the SSF's computational efficiency, they can effectively model bottom interactions and wide-angle propagation.

3.3.2 Split-Step Padé

The Split-Step Padé solution is a finite difference solution that allows for wide-angle propagation [26]. When retaining just one series term, the Padé expansion includes the Claerbout approximation that was developed as the standard 40-degree PE [61]. Retaining more terms increases the accuracy relative to a wider propagation angle. The Split-Step Padé allows for larger range steps than other finite difference models, while maintaining acceptable phase errors [58]. A detailed derivation of the solution is done in section 3.4; however, a version of the solution is briefly introduced here to emphasize a point about its efficiency. The summation representation of the Split-Step Padé solution is

$$p(r + \Delta r) = p(r) + \sum_{j=1}^{m} \frac{a_{j,m} X}{1 + b_{j,m} X} p(r), \tag{16}$$

where X is the depth operator and a and b are Padé coefficients. The summation terms can be calculated independently, which enables parallel processing. This is an important factor that contributes to its efficiency.

3.4 RAM

In this section, notation is adjusted to follow Collins' derivation [57]. RAM was selected because of its efficiency and accuracy, as well as its status as a physics-based model. RAM uses the Split-Step Padé solution introduced in section 3.3.2.

The mathematics underlying RAM relies on two assumptions:

1) Outgoing energy dominates (neglect backscattering)
2) Azimuthal symmetry (no ϕ dependence)

While RAM is a range-dependent model, each range step is solved as a range-independent problem. Using cylindrical coordinates, r is the range from a point at source depth z_0. The depth from the ocean surface is z, and ϕ represents azimuth. Each range-independent region begins with the frequency domain elliptical wave equation (note similarities to derivation of GPE). The farfield equation is satisfied by the complex acoustic pressure p

$$\frac{\partial^2 p}{\partial r^2} + \rho \frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial p}{\partial z}\right) + k^2 p = 0, \tag{17}$$

where ρ is medium density, $k = (1 + i\eta\beta)\frac{\omega}{c}$ is the complex wavenumber, ω is the angular frequency, and c is the sound speed in the medium. In the expression for k, $\eta = (40\pi \log_{10} e)^{-1}$ and β is sediment attenuation dB/λ. Considering only outgoing waves and neglecting backscattering after factoring, Eqn. (17) becomes

$$\frac{\partial p}{\partial r} = ik_0(1 + X)^{1/2} p, \tag{18}$$

where $X = k_0^{-2}\left(\rho\frac{\partial}{\partial z}\frac{1}{\rho}\frac{\partial}{\partial z} + k^2 - k_0^2\right)$ is the depth operator, $k_0 = \frac{\omega}{\nu}$ is the representative wavenumber, and ν is the representative phase speed. The solution to Eqn. (18) is

$$p(r + \Delta r, z) = \exp\left(ik_0\Delta r\,(1 + X)^{\frac{1}{2}}\right)p(r, z), \tag{19}$$

where Δr is the range step. At this step the m-term rational, the Padé approximation is applied to get the product representation of the Padé solution:

$$\exp\left(ik_0\Delta r\,(1 + X)^{\frac{1}{2}}\right)p(r, z) \cong 1 + \sum_{j=1}^{m}\frac{a_{j,m}X}{1 + b_{j,m}X}$$

$$= \prod_{j=1}^{m}\frac{1 + \lambda_{j,m}X}{1 + \mu_{j,m}X}$$

$$p(r + \Delta r, z) = \exp(ik_0\Delta r)\prod_{j=1}^{m}\frac{1 + a_{j,m}X}{1 + b_{j,m}X}p(r, z). \tag{20}$$

This version of the solution is designed for single processor runs, as opposed to Eqn. (16). The Padé coefficients $a_{j,m}$ and $b_{j,m}$ [62] are

$$a_{j,m} = \frac{2}{2m + 1}\sin^2\left(\frac{j\pi}{2m + 1}\right) \tag{21}$$

and

$$b_{j,m} = \cos^2\left(\frac{j\pi}{2m + 1}\right), \tag{22}$$

where m is the number of expansion terms.

The fields at the first range step are calculated using a PE self-starter. To avoid the singularity at the point source, the calculation starts at $r = x_0$, where x_0 is on the order of a single wavelength. The expression for the self-starter is

$$p(x_0, z) = \frac{\exp\left(ik_0 x_0 (1+X)^{\frac{1}{2}}\right)}{k_0 (1+X)^{\frac{1}{2}}} \delta(z - z_0). \tag{23}$$

A more detailed treatment can be found in [57] and [63]. A modification to Eqn. (23) is required for a point source

$$p(r_0, z) = \frac{\exp\left(ik_0 r_0 (1+X)^{\frac{1}{2}}\right)}{k_0^{-1/2} (1+X)^{\frac{1}{4}}} \delta(z - z_0). \tag{24}$$

For range-independent problems, Eqn. (20) is marched through each of the range-independent regions. In the case of range-dependent problems, conditions must be set at the vertical interface between each region. To do this, energy conservation is incorporated. To conserve energy a modified depth operator is required

$$\tilde{X} = k_0^{-2} \left(\frac{\rho}{\alpha} \frac{\partial}{\partial z} \frac{1}{\rho} \frac{\partial}{\partial z} \alpha + k^2 - k_0^2 \right), \tag{25}$$

where $\alpha = \left(\frac{\rho}{k}\right)^{1/2}$. Stability constraints that are applied during the coefficient calculation phase resolve Gibbs oscillations due to the different quantities conserved at vertical and horizontal interfaces when sloping interfaces are encountered. The Gibbs oscillations are projected onto the evanescent spectrum. The discretization of the depth operator is done by Galerkin's method detailed in [62]. The specifics of application, and the results are discussed in chapter four.

Chapter 4

Experimental Setup, Results, and Discussion

4.1 Data Acquisition and Curation

The World Ocean Database (WOD) is the world's largest public collection of ocean profile data [15]. Its quality control and uniform formatting made it an obvious choice for this thesis. The data consisted of *in situ* temperature, salinity, and depth measurements from Conductivity-Temperature-Depth (CTD) casts. Sound speed was calculated with temperature, salinity, and depth using the method from [16]. The data was selected using geographic coordinates and date. The date of measurements ranged from 1999 to 2022. To better capture ocean dynamics, the choice was made to segment the data and learn dictionaries as a function of meteorological seasons and bottom depth. This gave a total of eight datasets. The NESB was split into two segments defined as the shelf and shelf break. Table 4.1 shows the depths at which the dictionaries were defined. Depth information was chosen to be standard water depth (Table 4.2) defined by the WOD.

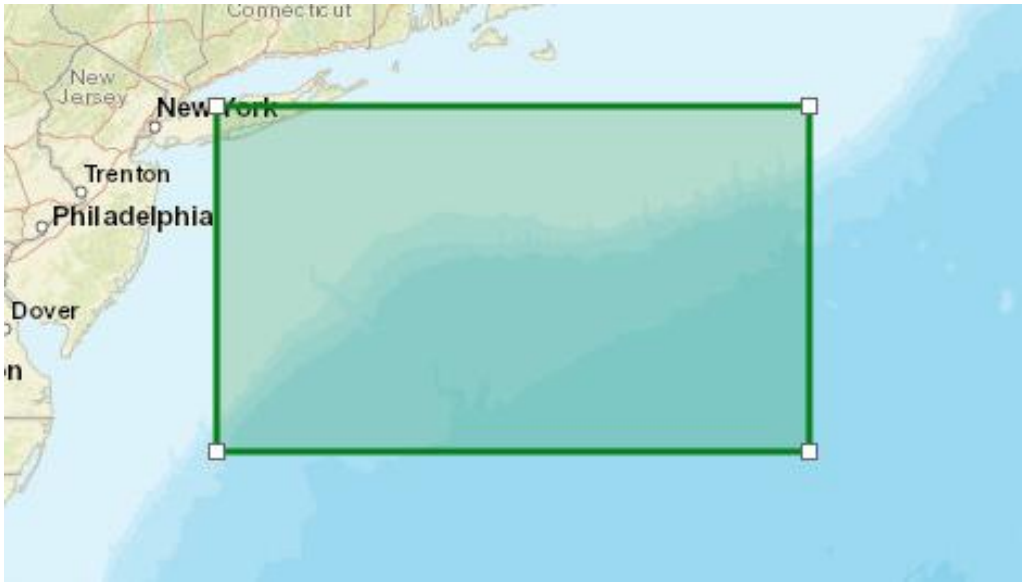


Figure 4.1: All profiles were taken from the area defined by the green box [64].

Region	Depth (m)
Shelf	0 – 50
Shelf Break	0 – 250

Table 4.1: Depth range of each region.

Many profiles did not reach the maximum depth in each region, which resulted in missing values throughout the dataset. This was handled differently for each algorithm. For the K-SVD, missing values (NaNs) were imputed with zeros. For EOFs, NaNs were imputed with the depth-wise mean sound speed. Table 4.3 contains the size of each seasonal dataset. Further information about this choice and its implications were discussed in section 4.2.

Level #	Depth (m)	Level #	Depth (m)	Level #	Depth (m)	Level #	Depth (m)
1	0	36	475	71	2400	106	5900
2	5	37	500	72	2500	107	6000
3	10	38	550	73	2600	108	6100
4	15	39	600	74	2700	109	6200
5	20	40	650	75	2800	110	6300
6	25	41	700	76	2900	111	6400
7	30	42	750	77	3000	112	6500
8	35	43	800	78	3100	113	6600
9	40	44	850	79	3200	114	6700
10	45	45	900	80	3300	115	6800
11	50	46	950	81	3400	116	6900
12	55	47	1000	82	3500	117	7000
13	60	48	1050	83	3600	118	7100
14	65	49	1100	84	3700	119	7200
15	70	50	1150	85	3800	120	7300
16	75	51	1200	86	3900	121	7400
17	80	52	1250	87	4000	122	7500
18	85	53	1300	88	4100	123	7600
19	90	54	1350	89	4200	124	7700
20	95	55	1400	90	4300	125	7800
21	100	56	1450	91	4400	126	7900
22	125	57	1500	92	4500	127	8000
23	150	58	1550	93	4600	128	8100
24	175	59	1600	94	4700	129	8200
25	200	60	1650	95	4800	130	8300
26	225	61	1700	96	4900	131	8400
27	250	62	1750	97	5000	132	8500
28	275	63	1800	98	5100	133	8600
29	300	64	1850	99	5200	134	8700
30	325	65	1900	100	5300	135	8800
31	350	66	1950	101	5400	136	8900
32	375	67	2000	102	5500	137	9000
33	400	68	2100	103	5600		
34	425	69	2200	104	5700		
35	450	70	2300	105	5800		

Table 4.2: Standard Depths Levels defined by the World Ocean Database [65].

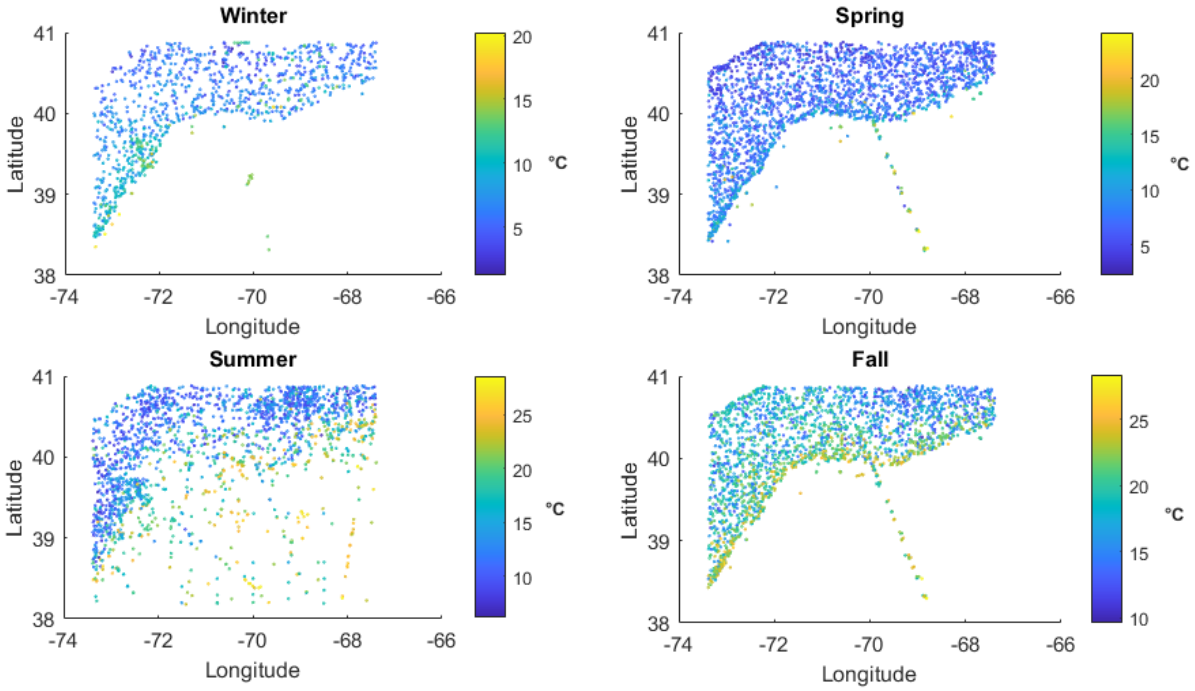


Figure 4.2: Temperature at z = 20 m in each seasonal dataset.

Region	Season	Rows (Depths)	Columns (Profiles)
Shelf	Winter	11	669
	Spring	11	1582
	Summer	11	1462
	Fall	11	1482
Shelf Break	Winter	27	611
	Spring	27	1505
	Summer	27	1165
	Fall	27	1404

Table 4.3: Each region and seasons' data array size.

4.2 Procedure and Analysis

The goal of the procedure was to generate DL informed SSP ensembles as inputs into RAM to calculate TL ensembles. The EOF stochastic model was based on the equation

$$\boldsymbol{y}(z) = \bar{\boldsymbol{y}} + \sum_{i=1}^{e} \gamma_i^2 \boldsymbol{w}_i \,, \tag{26}$$

where $\boldsymbol{y}(z)$ was a single sound speed ensemble, $\bar{\boldsymbol{y}}$ was the mean sound speed profile of the input SSP data, and e was the number of leader-order EOFs. The summation terms γ_i^2 and $\boldsymbol{w}_i$ were the eigenvalues and eigenvectors (EOFs) of an eigendecompositon, respectively. In this research, e = 3. The K-SVD parallel to this model was formulated as

$$\boldsymbol{x}(z) = \bar{\boldsymbol{x}} + \sum_{i=1}^{s} a_i \boldsymbol{d}_i \,, \tag{27}$$

where a_i was a nonzero coefficient from the sparse matrix $\boldsymbol{A}$, $\boldsymbol{d}_i$ was a dictionary atom, and s was the sparsity value.

To compute a matrix of SSP ensembles $\boldsymbol{C}_{EOF}$, a normal random variable $\boldsymbol{P}_{EOF} \in \mathbb{R}^{e \times R}$ was defined, where R was the number of ensembles. This was formulated as

$$\boldsymbol{C}_{EOF} = \bar{\boldsymbol{y}} + \boldsymbol{W}_e(\boldsymbol{P}_{EOF} \odot \boldsymbol{\gamma}) \,, \tag{28}$$

where $\boldsymbol{W}_e$ contained the leading-order EOFs and $\boldsymbol{\gamma}_e$ contained the standard deviations[2] associated with leading-order EOFs. Similarly, the dictionary learning parallel to this equation was

$$\boldsymbol{C}_{DL} = \bar{\boldsymbol{x}} + \boldsymbol{D}_{\boldsymbol{s}}(\boldsymbol{P}_{DL} \odot \sigma)\,, \tag{29}$$

where the relevant dictionary atoms comprised $\boldsymbol{D}_{\boldsymbol{s}}$ and σ was the standard deviation of the relevant coefficients.

A statistical analysis of the K-SVD algorithm's inputs and outputs was completed to inform the development of the stochastic mode. The analysis was restricted to the summer season, but figures for other seasons were included in Appendix D and E. This analysis was done alongside results of the EOF stochastic model for comparison.

4.2.1 K-SVD Analysis

K-SVD parameters were primarily set based on recommendations in a previously referenced paper [43], as well as the statistical analysis in this section. This included the number of iterations through the algorithm, DUCs, and the number of atoms. Sparsity (s) was determined by the statistical analysis explained in the K-SVD statistical analysis section that follows this section. The parameters were made available in Table 4.4, where *nz* was the number of depths.

Ocean Region	Atoms	Sparsity	Iterations	DUCs
Shelf	$nz \times 3$	3	30	2
Shelf Break	$nz \times 3$	4	30	2

Table 4.4: K-SVD parameters for each ocean region[3].

[2] As mentioned in section 2.1, γ^2 is an eigenvalue of the covariance matrix $\boldsymbol{Y}\boldsymbol{Y}^T$. The eigenvalues of covariance matrices are variances. The square root of variance is the standard deviation, and, given special properties of the eigendecomposition of a diagonalizable matrix, these standard deviations are also singular values.

[3] Changing these parameters may drastically change outcomes. This is especially true for sparsity and number of dictionary atoms.

K-SVD Statistical Analysis

To create the DL informed ensembles, statistical analysis of the K-SVD sparsity input, and its outputs was required. Performing this analysis started with posing three questions:

1) What is the optimal sparsity for this dataset?
2) Which atoms are selected most frequently?
3) What distribution do the nonzero coefficients related to these atoms follow?

To answer these questions, the K-SVD was applied to a learning dataset to obtain a learned dictionary (LD). The learning dataset was defined as all profiles not used as elements of the initial dictionary. As a reminder, the dictionary was initialized by randomly sampling $\boldsymbol{X}$. The random samples were removed to create the learning set. Atom selection begins with a single pass through the K-SVD, which outputs $\boldsymbol{D}$ and $\boldsymbol{A}$. The LD exhibits a distribution defined by the number of times each atom was selected to represent a profile. This number was equal to the total number of nonzero coefficients $\left\|\boldsymbol{a}_k^S\right\|_0$ in a given row of the coefficient matrix $\boldsymbol{A}$. Each row in the coefficient matrix **A** corresponds to an atom in **D**. The number of nonzero coefficients related to each individual atom was divided by the total number of nonzero coefficients in $\boldsymbol{A}$ to get the probability that each atom was selected**.** This process determined which atoms were most likely to be selected by the K-SVD. Once these atoms were identified, the s-atoms used the most frequently were stored for SSP ensemble generation. The choice of atoms determined which nonzero coefficients were used by default. The distribution of the nonzero coefficients in the rows of $\boldsymbol{A}$ were assumed to be normal to limit variation between the EOF stochastic model and the K-SVD stochastic model.

Optimal Sparsity

The optimal sparsity value was determined by iterating through the algorithm 1000 times and observing the mean of root mean squared error (RMSE) over those iterations. The representation RMSE for each iteration was computed as

$$\boldsymbol{RMSE}_{pr} = \frac{1}{M_j}\sum_{j=1}^{N}\left(\boldsymbol{x}_j - \widehat{\boldsymbol{x}}_j\right)^2, \tag{30}$$

where $\boldsymbol{RMSE}_{pr}$ is a vector of root mean squared error between a profile $\boldsymbol{x}_j$ and its estimate $\widehat{\boldsymbol{x}}_j$, and

$$RMSE = \frac{1}{N}\sum R\boldsymbol{MSE}_{pr}\,, \tag{31}$$

where $RMSE$ is the mean of the root mean squared error across all profiles. $RMSE$ is the y-axis value in Fig. 4.4. A high-performance computer (HPC) was used to apply the K-SVD 1000 times to each seasonal dataset for every sparsity value s, where $s = 1,2, \ldots 8$. The mean of the 1000 MSE values for each s relative to a regional dictionary is shown in the figure below. Based on RMSE for each season, sparsity for the shelf was chosen to be $s = 3$. Although [43] recommended $s = 1$ based on the size of the depth dimension (Table 4.4), Fig. 4.4 showed that $s = 1$ gave significantly higher representation RMSE and often was not sufficient to capture variability (Fig. 4.3). Also, in Fig. 4.3 the $s = 8$ example was included to demonstrate how ensemble generation behaved at the opposite extreme.

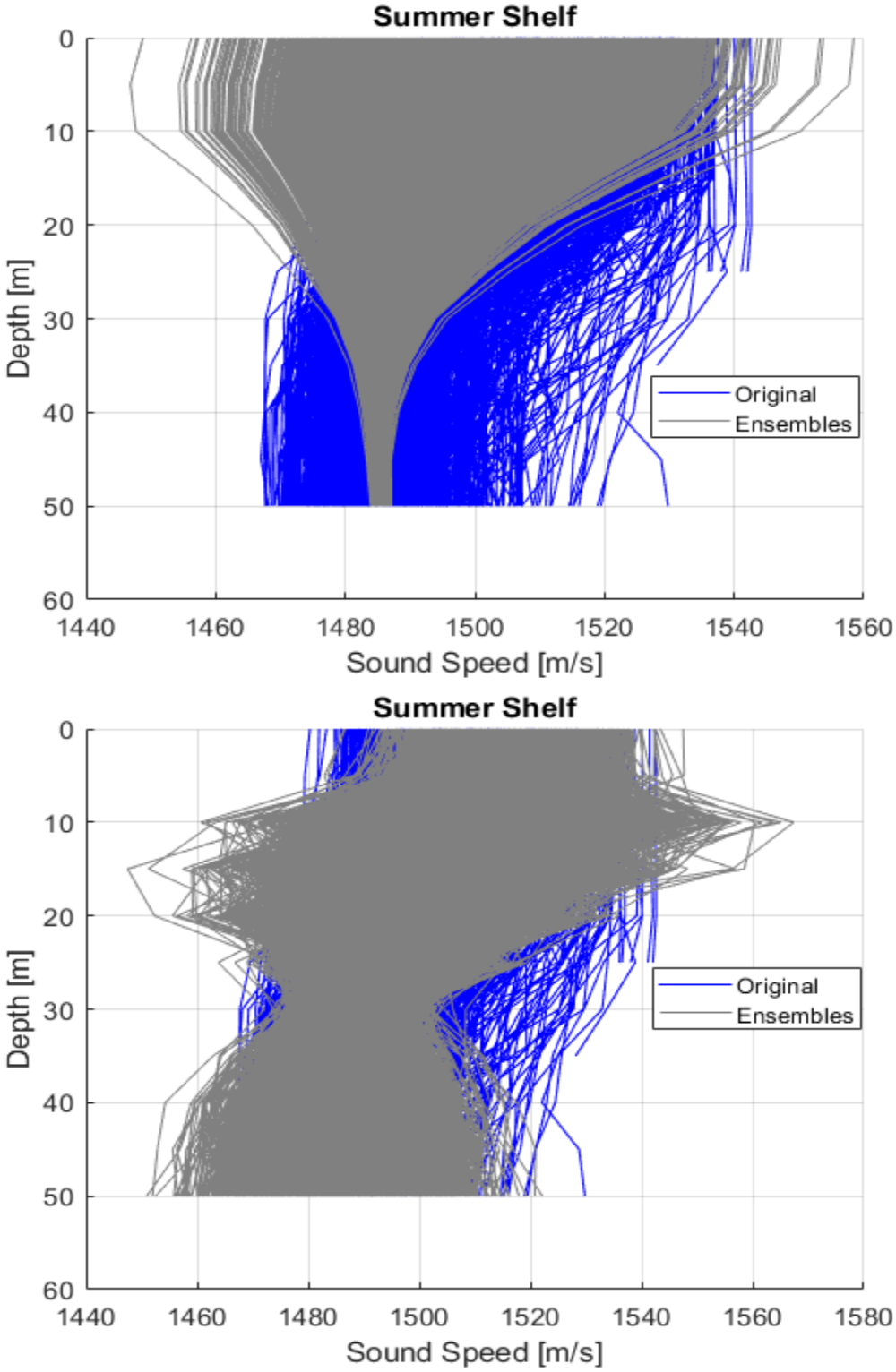


Figure 4.3: Summer shelf K-SVD ensembles when s = 1 (top) and when s = 8 (bottom). The ensembles generated while s = 8 covered SSP variability at lower depths but still show lack of coverage at 30 m.

Even as sparsity was increased, highly unpredictable behavior remained observable in K-SVD ensembles. Over a large sample size, the RMSE in Fig. 4.4 demonstrated that K-SVD results converge to error patterns found in previous work ([25], [43]); however, ensemble behavior was nondeterministic. These issues were addressed in the *K-SVD Ensembles* section of this chapter.

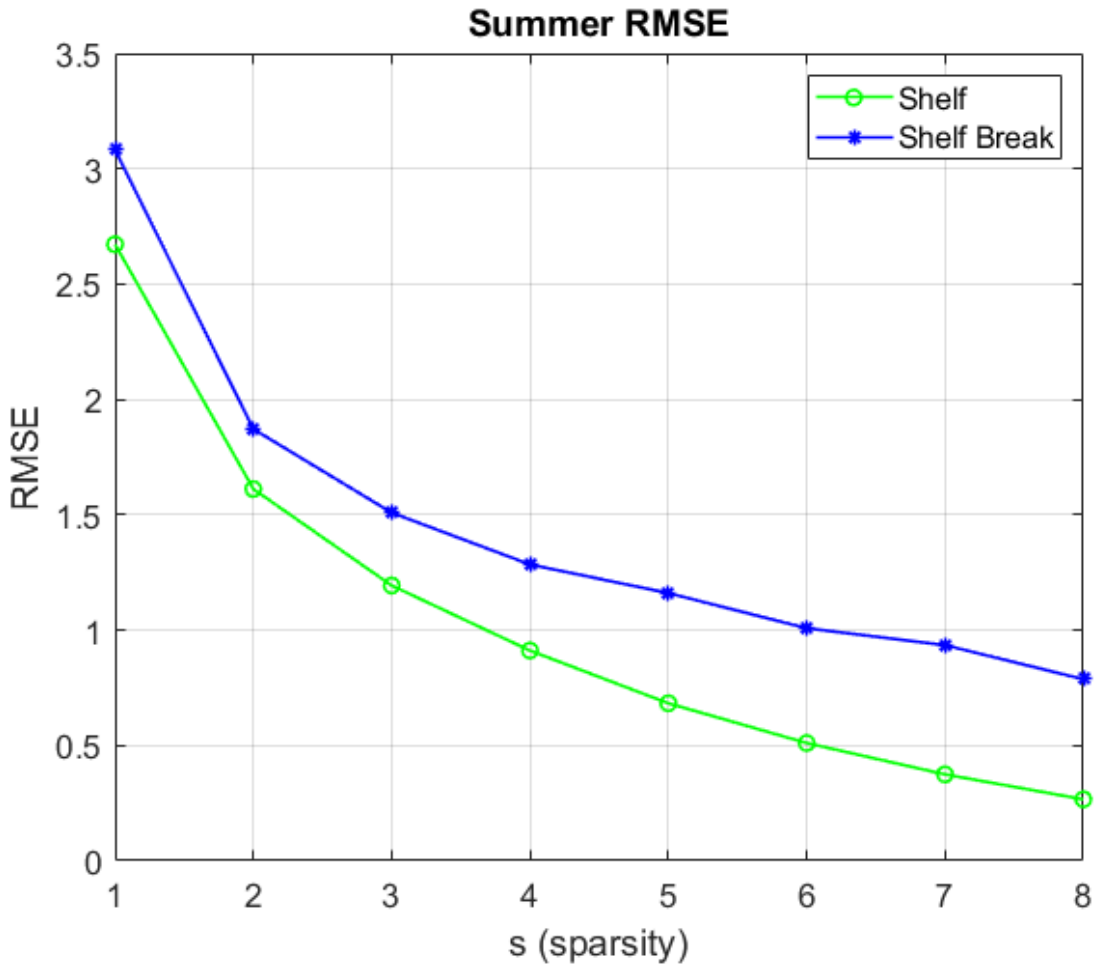


Figure 4.4: RMSE over 1000 iterations at each sparsity value.

Atom Selection

The most frequently selected atoms were determined in the sparse coding stage by CoefROMP. They were identified by summing the number of nonzero coefficients in each row of $\boldsymbol{A}$. The s-atoms that were selected most frequently were used to generate SSP ensembles. The number of times they were used to represent a signal is shown in Fig. 4.5, where the orange highlighted bars indicate the chosen atoms. All 33 atoms in the dictionary are plotted in Fig. 4.6 to show their respective shapes. The green atoms are the atoms most frequently used. The atoms are arranged from highest to lowest usage, and the figure should be read from left to right in a snake pattern.

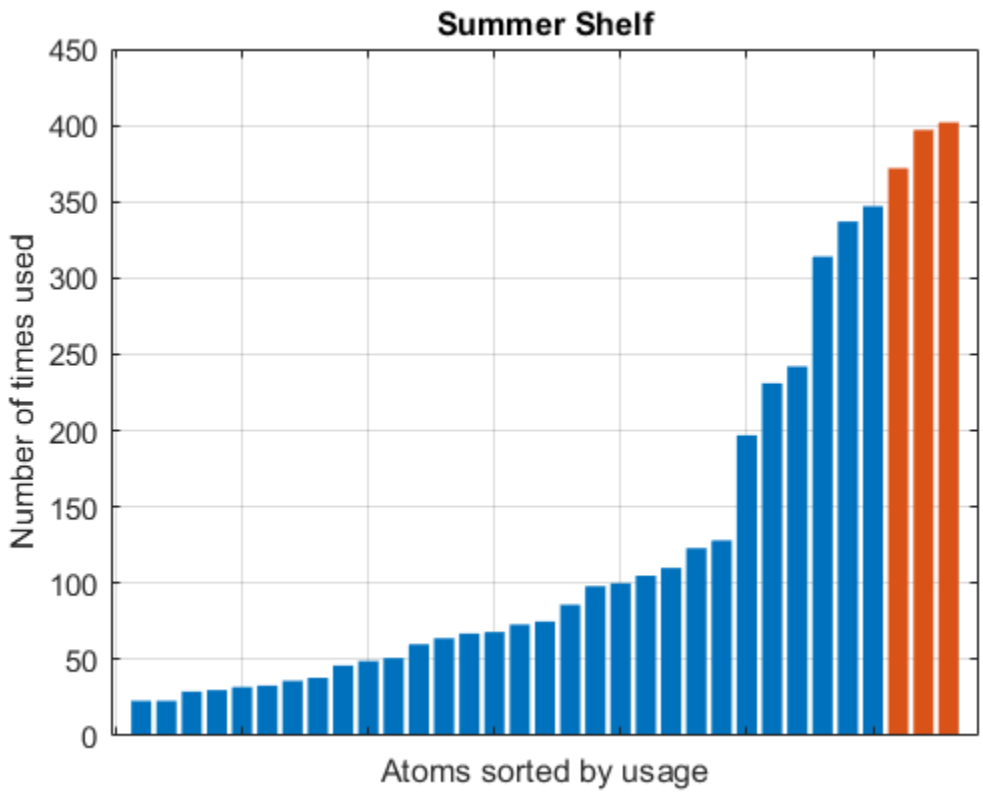


Figure 4.5: Atom usage for summer with atom(s) used for ensemble generation highlighted in orange.

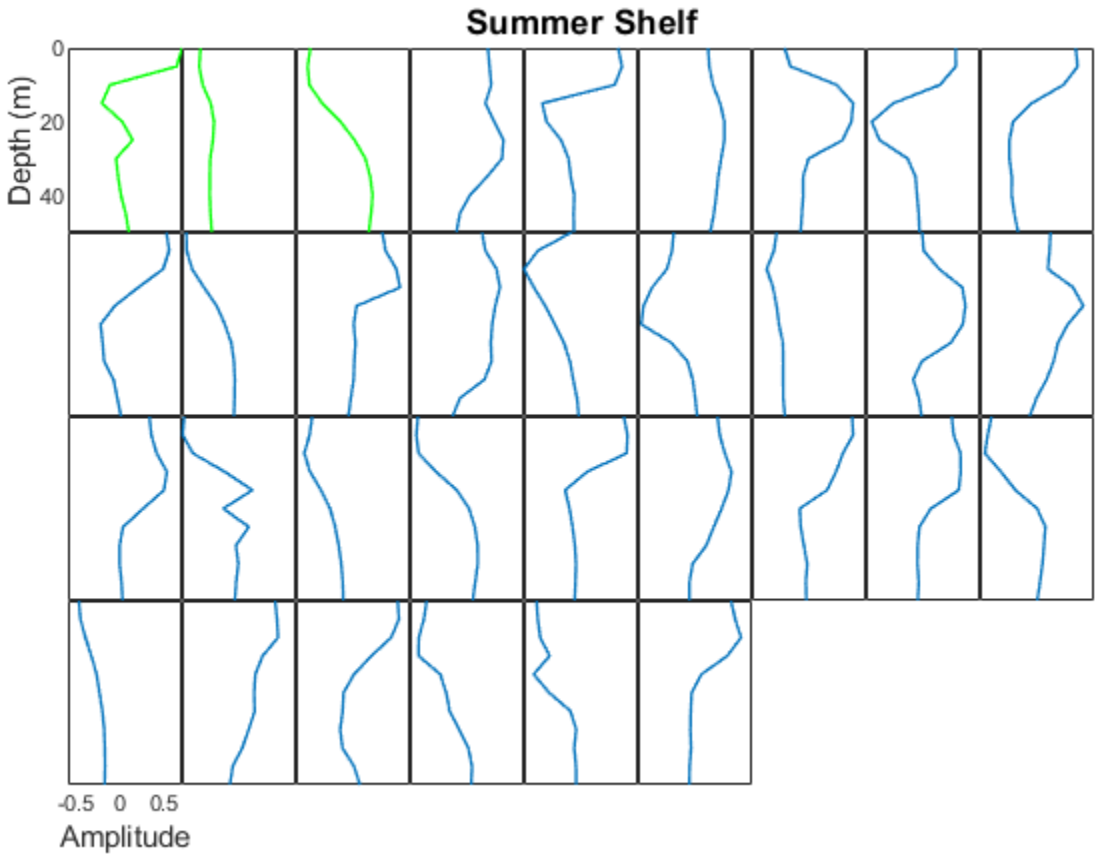


Figure 4.6: Dictionary atoms sorted according to most frequent usage. Green atoms were used most. Ordering moves from left to right in a snake pattern.

Coefficient Distribution

Coefficient values varied widely within each row of the coefficient matrix. This variation presented a challenge when trying to determine an acceptable distribution. Ultimately, to be consistent with the EOF stochastic model, the coefficients were assumed to show a normal distribution. Two methods to create the distribution were tried:

Method 1: Fit a distribution to each individual row of nonzero coefficients and determine the standard deviation for each.

Method 2: Fit a distribution to all nonzero coefficients from each selected row in $\boldsymbol{A}$ as a single vector and determine the standard deviation.

The green points in Fig. 4.7 and Fig. 4.8 represent the combined normal coefficient distributions of each individual row of $\boldsymbol{A}$ used to generate ensembles with Method 1 and Method 2, respectively. These separate distributions each had their own mean and standard deviation related to the nonzero coefficients in the relevant rows of $\boldsymbol{A}$. The purple points were the nonzero coefficients from the relevant rows of $\boldsymbol{A}$. There were three observations made from this depiction: 1) the visible gaps in the purple points at zero were a result of sparsity and only selecting nonzero coefficients, 2) the normally distributed coefficients (green, Fig. 4.7) from Method 1 overrepresented high amplitude coefficients from $\boldsymbol{A}$ (purple), and 3) the normally distributed coefficients (green, Fig. 4.8) did not capture some high amplitude nonzero coefficients from $\boldsymbol{A}$. Method 2 gave a more centered, or narrowed, distribution of coefficients. While some of the outlier coefficients were not adequately represented, Method 2 gave a better representation of where the majority of nonzero coefficients (purple) lied. The Gaussian fit in Fig. 4.9 shows that the distribution lacked density around the mean. A distribution with greater kurtosis would improve coefficient coverage. For completeness, separate scatter plots of the coefficients from Method 2 were presented in Fig. 4.10.

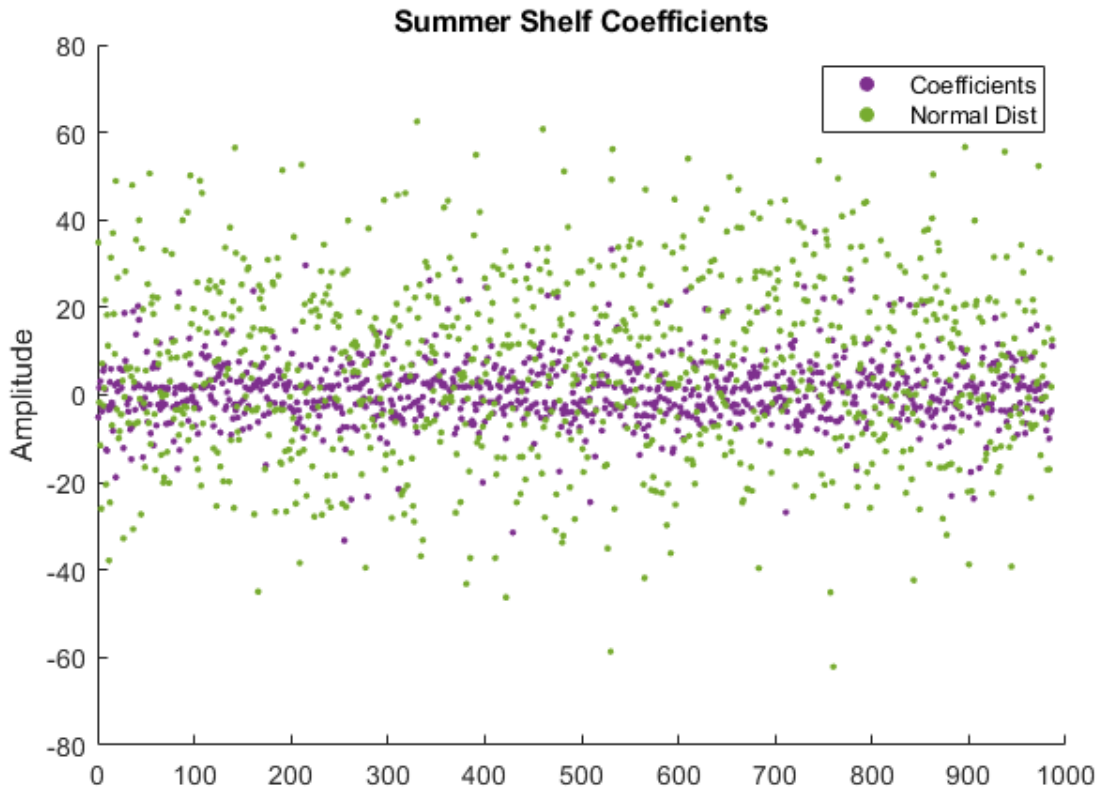


Figure 4.7: Nonzero coefficients (purple) from $\boldsymbol{A}$ that correspond to the most frequently selected atoms. The green dots were generated by a normal distribution with Method 1.

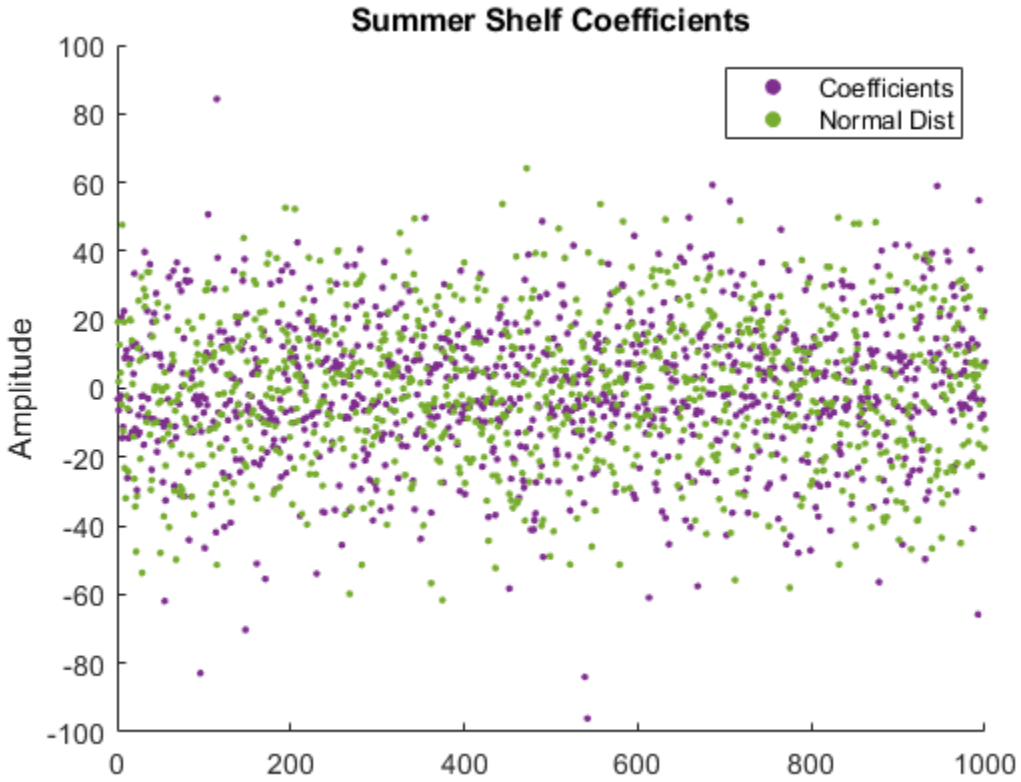


Figure 4.8: Nonzero coefficients from $\boldsymbol{A}$ (purple) and coefficients from the normal distribution (green) generated with Method 2.

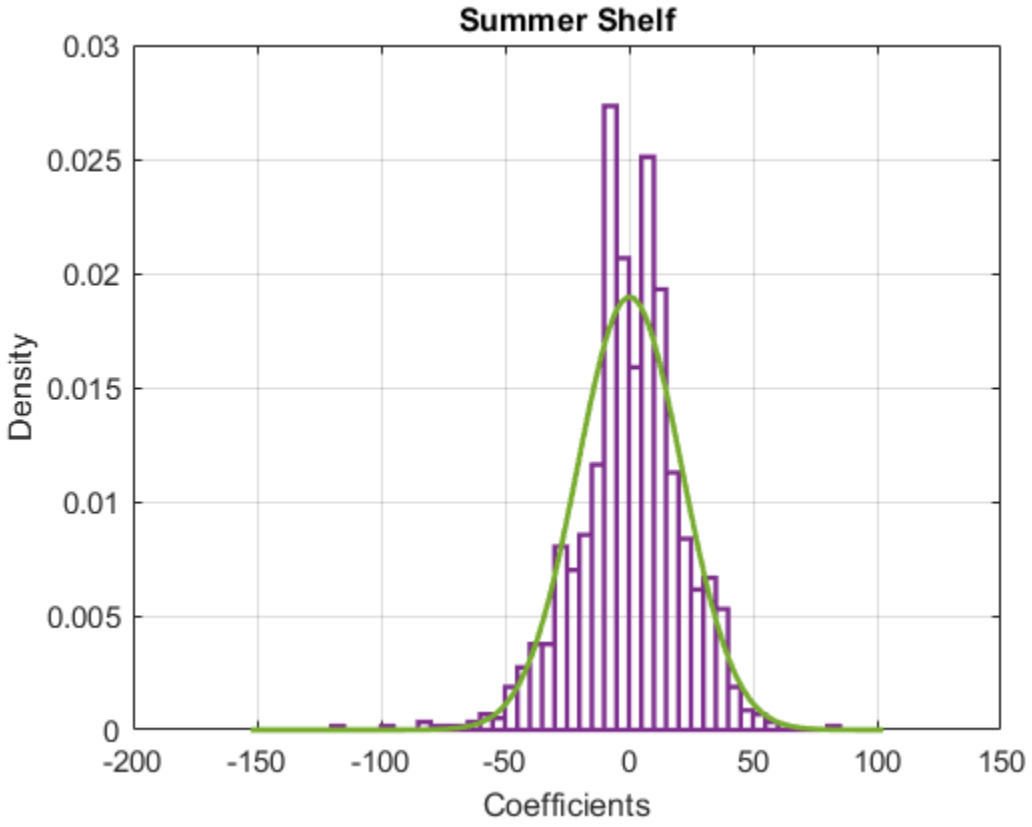


Figure 4.9: Method 2 gaussian fit to nonzero coefficients from summer shelf dataset.

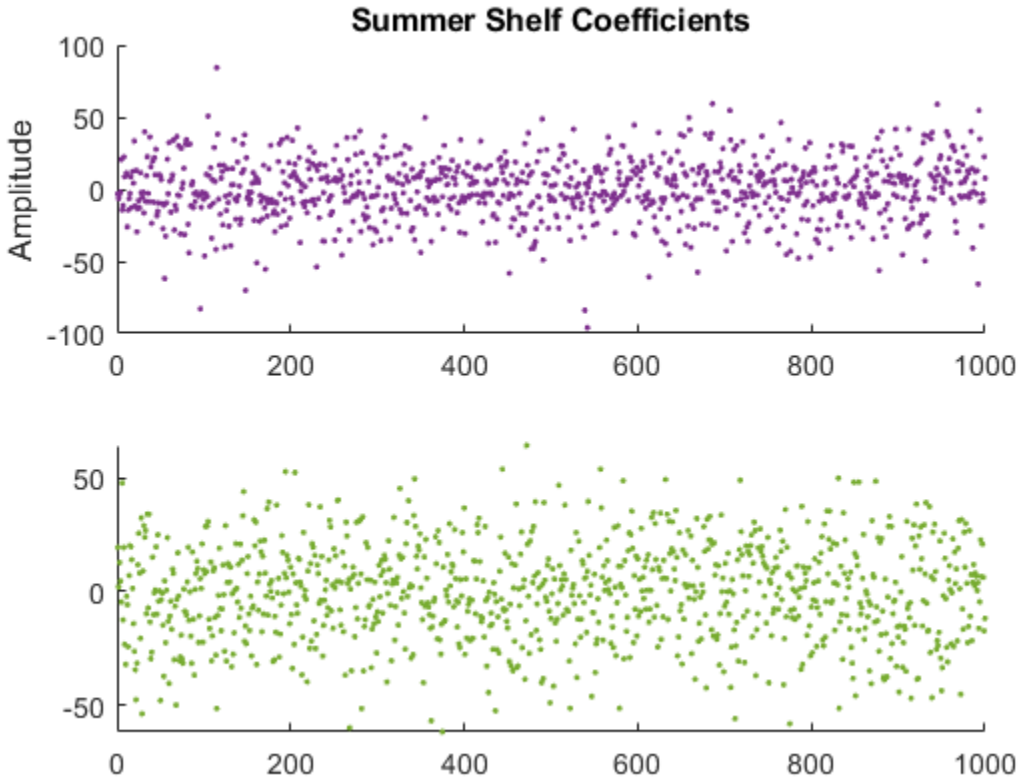


Figure 4.10: Coefficients from the K-SVD (top), and coefficients generated with Method 2 (bottom).

K-SVD Ensembles

K-SVD ensembles shown in Fig. 4.11 and Fig. 4.12 were the result of ensembles generated during computation that also produced the TL ensembles found . This run will be referred to as TLRUN1 for the rest of the thesis. The ensembles depicted in Fig. 4.11 poorly represented variability in SSP measurements. The nondeterministic nature of the K-SVD ensemble envelope was an important focus. Fig. 4.12 pointed to the selected dictionary atoms as the driver of the ensemble envelope. While a more accurate distribution where outliers were represented fully would likely create a distribution with greater variability, the dictionary atoms were primarily responsible for the shape of the ensembles. Given the shape of the mean profile shown in Fig. 4.13, the expectation was that the algorithm would most frequently select atoms resembling that shape. Fig. 4.6 indicated the algorithm chose unexpected atoms. This also called

into question the atom selection method for the stochastic model, thus how atoms are selected to represent profiles during the sparse coding stage. In CoefROMP, like OMP, atoms are selected to approximate an input vector based on the maximum absolute inner product between the residual and the dictionary (Eqn. (7)). method of determining similarity may predispose the K-SVD to poor spatial variability representation.

A fundamental difference between the K-SVD and EOF analysis is their respective goals. The goal of EOF analysis is to capture spatial variability by decomposing a covariance matrix, while the goal of the K-SVD can be summarized as order reduction by solving a minimization problem (Eqn. (3)). While EOF analysis is also a powerful tool for dimensionality reduction, the K-SVD has been proven to be superior in this respect; however, when EOF analysis is applied to a problem that suits its strength (spatial variability capture), it outperforms the K-SVD as applied in this thesis.

For a more granular view of ensemble shape, fifty random samples of the ensembles and the dataset were selected to be shown in Fig. 4.14. The ensembles cluster toward the center of input data. Fig. 4.15 was an arbitrary run added to indicate how drastically the K-SVD ensembles can vary from run-to-run on the same dataset. Notice that the mean profile varied between these two examples, although they were from the same dataset. This was the consequence of removing profiles from the input dataset that were randomly selected during dictionary initialization. It is plausible to assume that if profile removal affected the mean profile, this could account for the unpredictability of the ensembles, but that was not the case. Nondeterministic behavior continued whether the profiles used for dictionary initialization were removed or not. The mean ensemble profile and the mean profile for the arbitrary run are shown in Fig. 4.16.

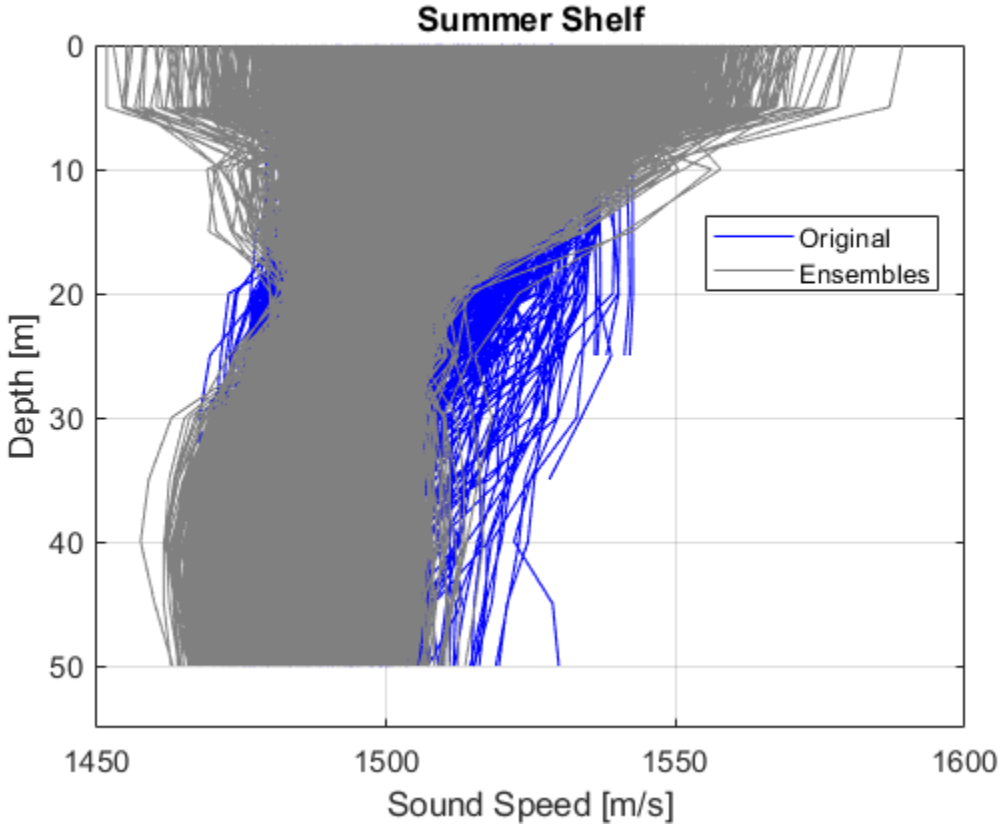


Figure 4.11: K-SVD ensembles from TLRUN1 with s = 3 overlayed onto original sound speed profiles.

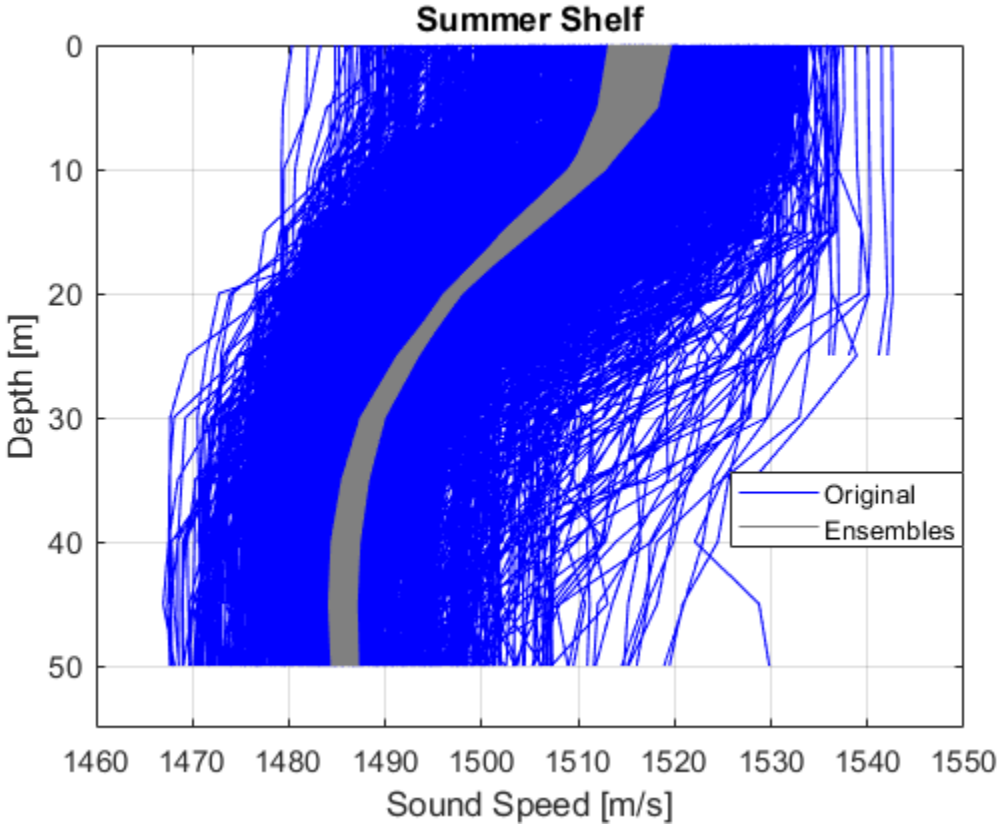


Figure 4.12: These ensembles are the product of the selected dictionary atoms and a normal distribution with zero mean and standard deviation of 1. The genesis of the ensemble envelope in Fig. 4.11 was visible in this depiction. This illustrates that the shape of the ensembles was primarily influenced by the dictionary atoms.

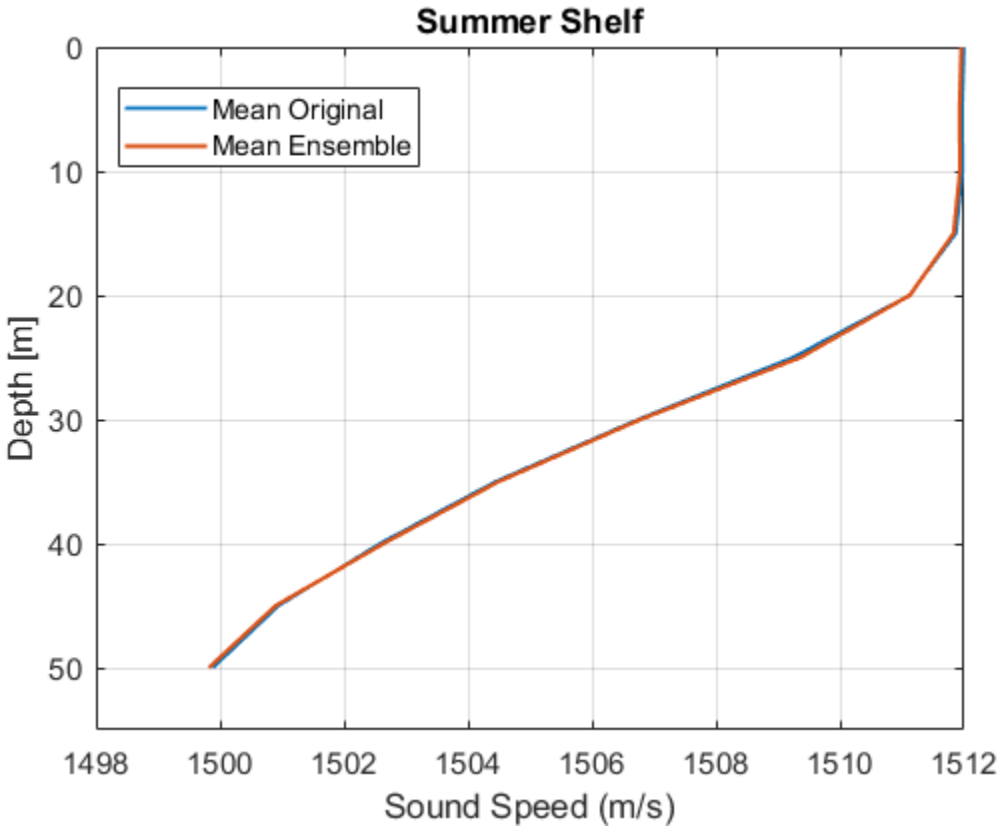


Figure 4.13: Mean SSP ensemble and mean SSP profile from TLRUN1 (Fig. 4.5).

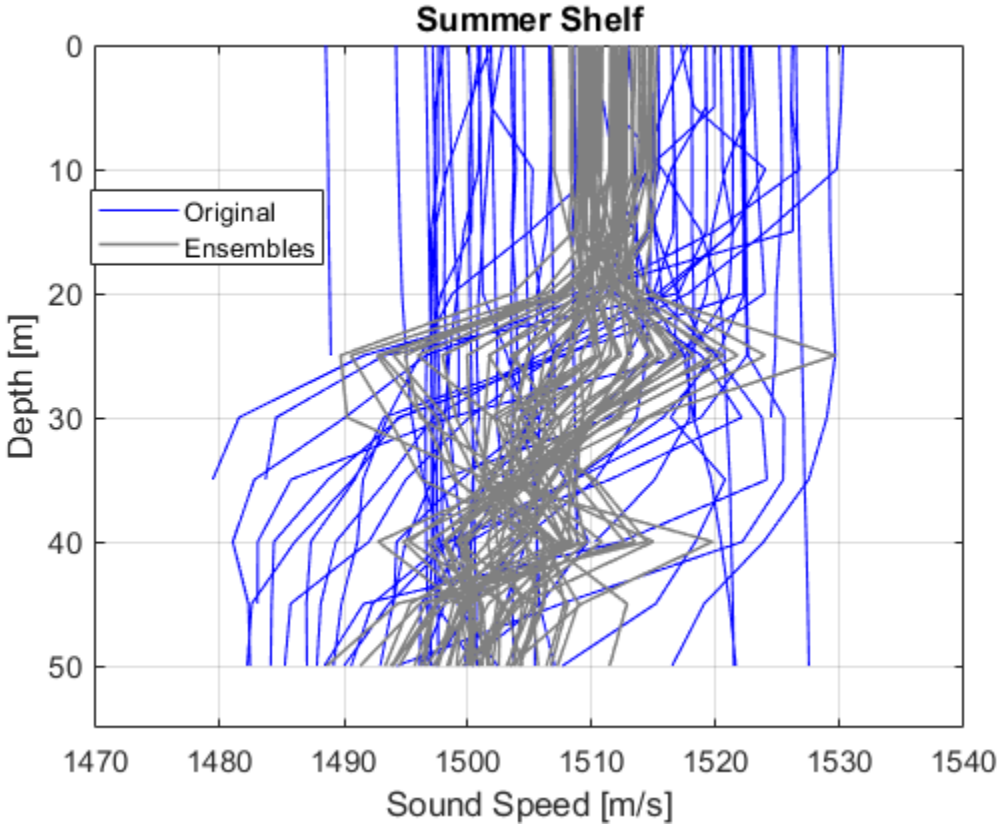


Figure 4.14: Fifty Random samples from TLRUN1. The K-SVD ensembles cluster toward the center of the input data.

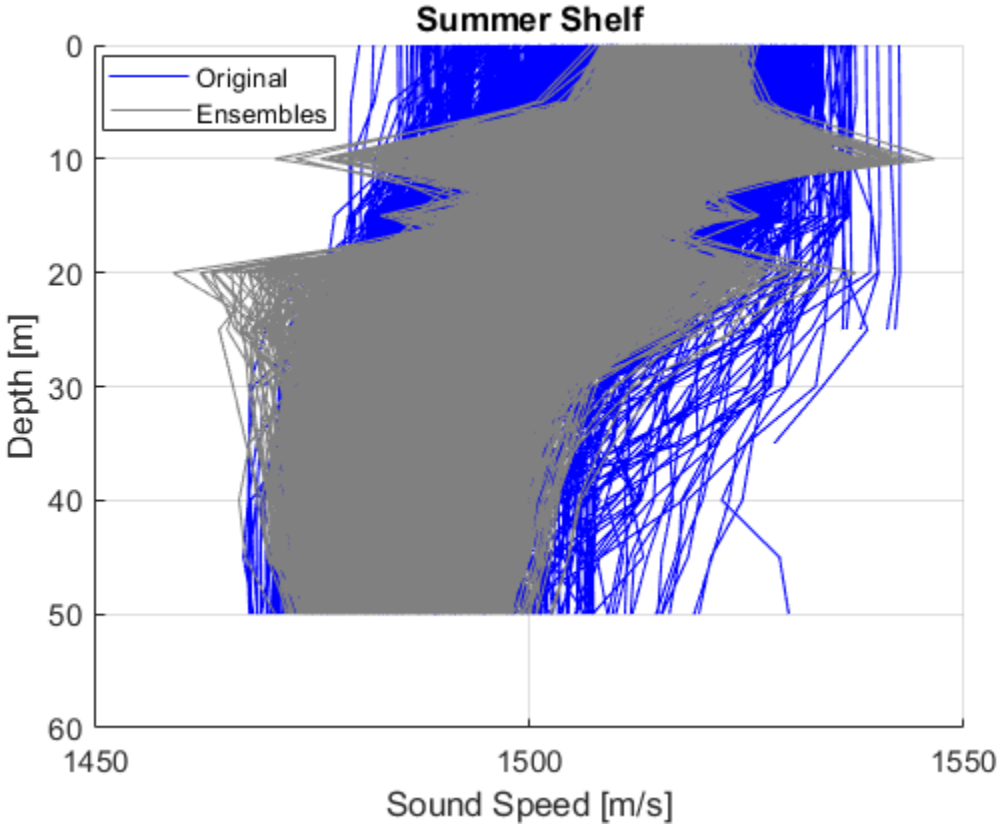


Figure 4.15: K-SVD generated SSP ensembles (N = 5000) over the original summer shelf sound speed data. This was an arbitrary run of the K-SVD algorithm to demonstrate the nondeterministic behavior of the ensembles.

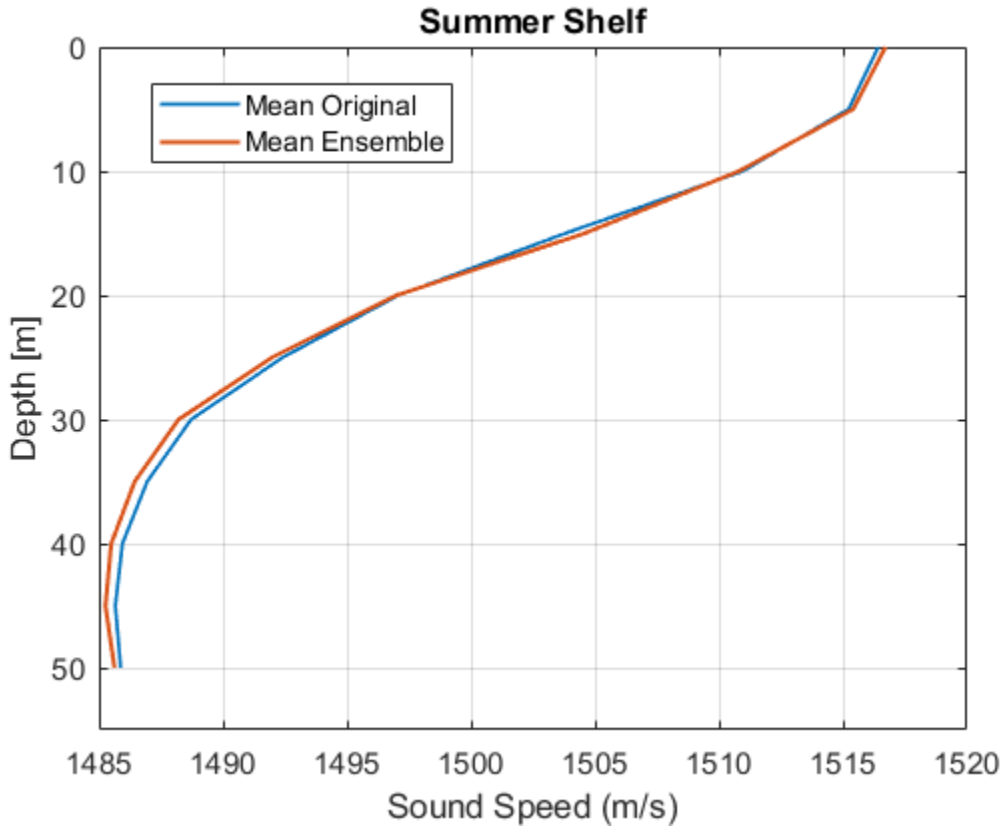


Figure 4.16: Mean SSP ensemble and mean SSP profile from arbitrary run in Fig. 4.8.

Similar to the shelf region, the K-SVD ensembles were unpredictable and did not cover variability. At deeper depths, SSP data was less dense, which contributed to the already sometimes narrow SSP variability coverage given by K-SVD ensembles (Fig. 4.17).

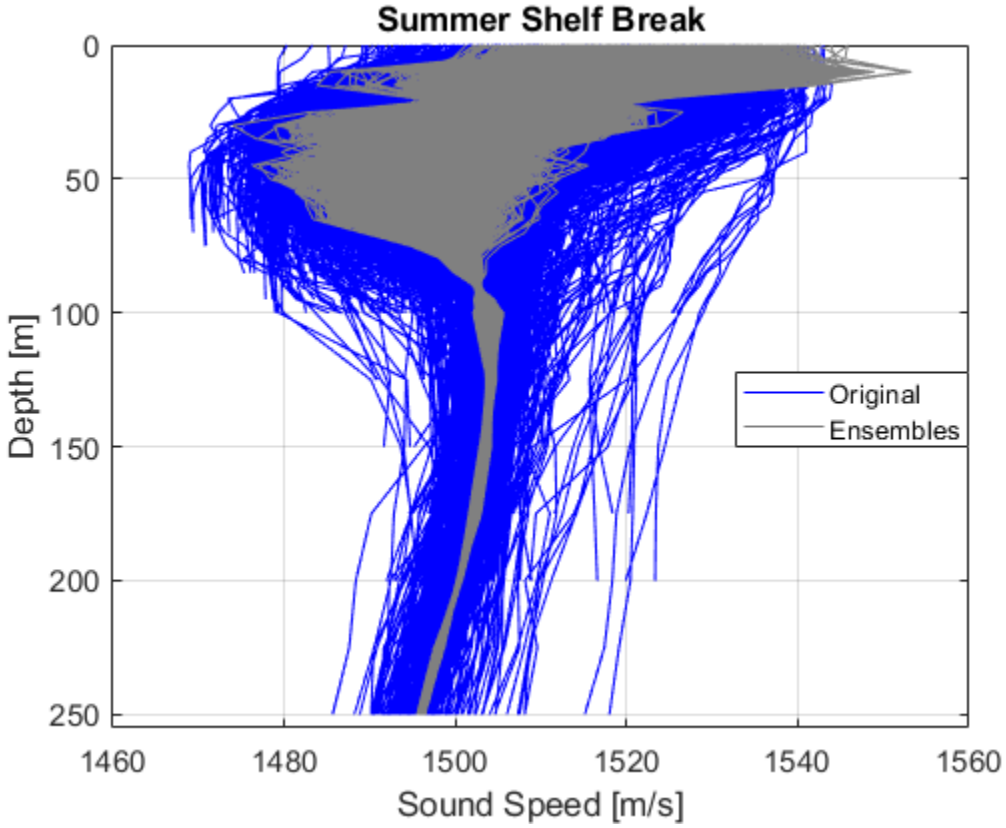


Figure 4.17: K-SVD ensembles with s = 4 at the shelf break. As SSP profiles get less dense with an increase in depth, ensembles represent less of the variability of SSP profiles.

4.2.2 EOF Analysis

EOF ensembles were generally more consistent than K-SVD ensembles in capturing SSP profile variability. For the case shown in Fig. 4.18, three leading-order EOFs were used. The EOF mean ensemble profile shown in Fig. 4.20 more consistently followed the input SSP measurement profile over the different seasons and regions, although this does come with a qualification. The EOF algorithm required that all missing values in the SSP measurement input be filled with its own depth-wise mean. This likely contributes to a bias with respect to mean ensemble profile fit to the mean profile; however, the algorithm simply does not work without this step. The difference between K-SVD (Fig. 4.16) and EOFs (Fig. 4.20) in mean profile replication was a typical result.

To further illustrate the disparity in variability coverage between EOF and K-SVD ensembles, the mean absolute deviation (MAD)[4] was computed as

$$\boldsymbol{MAD} = \frac{1}{R}\sum_{i=1}^{R} \boldsymbol{c}_i - \bar{\boldsymbol{c}}, \tag{32}$$

where R is the number of ensembles, $\boldsymbol{c}_i$ is an ensemble, and $\bar{\boldsymbol{c}}$ is the mean ensemble. $\boldsymbol{MAD}$ for ensembles and original SSP data at the summer shelf are shown in Fig. 21. Accurate spatial variability coverage results in a $\boldsymbol{MAD}$ curve that matches the $\boldsymbol{MAD}$ of the SSP data. The EOF maximum absolute difference between the EOF ensemble $\boldsymbol{MAD}$ and the SSP data $\boldsymbol{MAD}$ was 1.2 m/s, while the maximum absolute difference between the K-SVD ensemble $\boldsymbol{MAD}$ and the SSP data $\boldsymbol{MAD}$ was 5.5 m/s. This result was visible in Fig. 11, where the K-SVD ensembles overshot spatial variability at depths from 0 m to 10 m and did not cover variability at 20 m.

[4] Mean absolute deviation for the original SSP data was calculated using the same form.

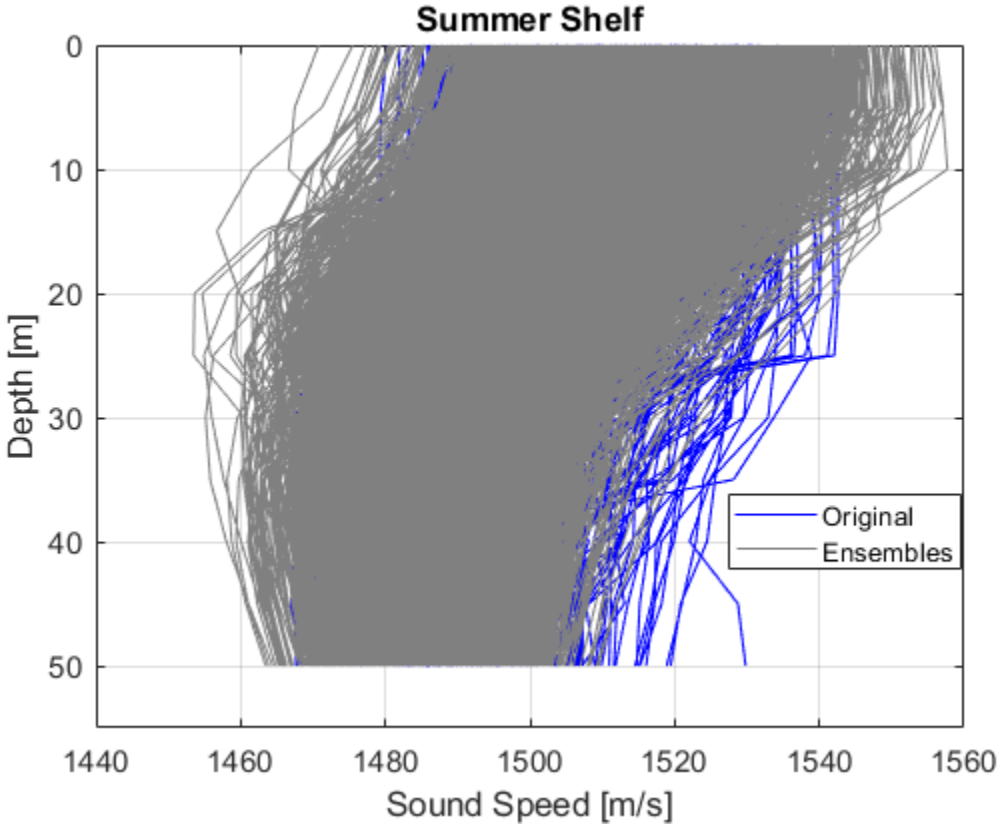


Figure 4.18: EOF ensembles overlayed onto original SSP profiles. Three EOFs were used to generate ensembles.

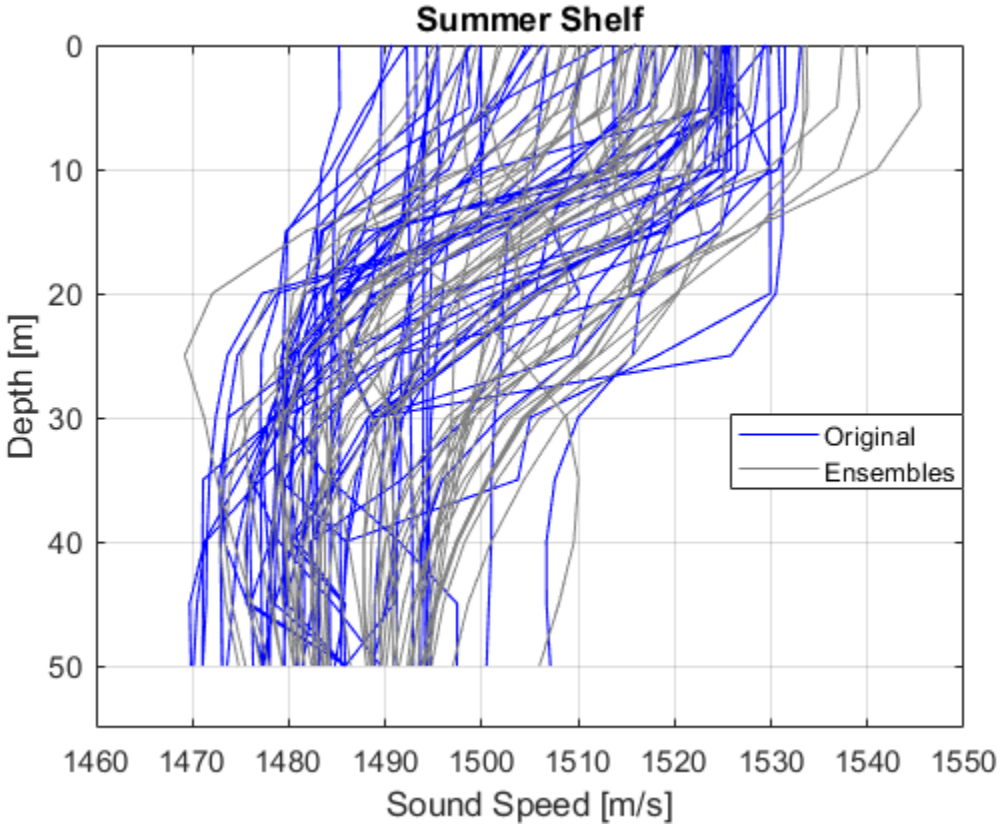


Figure 4.19: Random sampling of fifty EOF ensembles and profiles. The individual ensembles do not converge to the cener like the K-SVD ensembles.

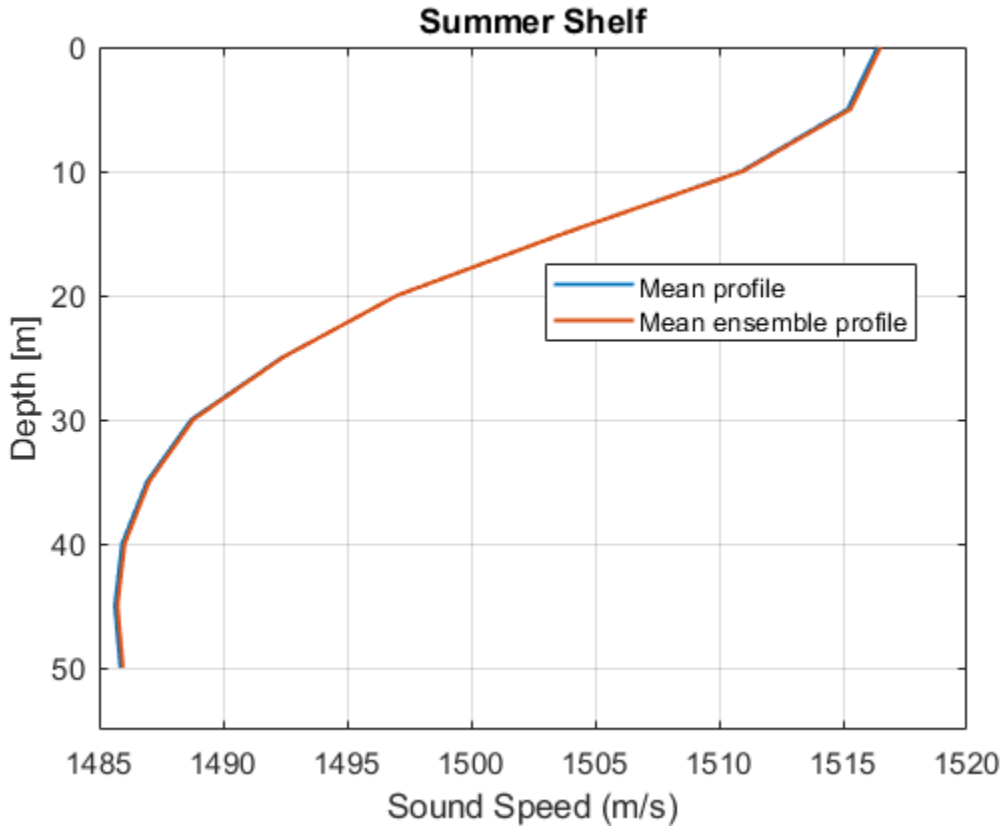


Figure 4.20: EOF mean SSP ensemble profile and mean SSP profile for summer shelf dataset.

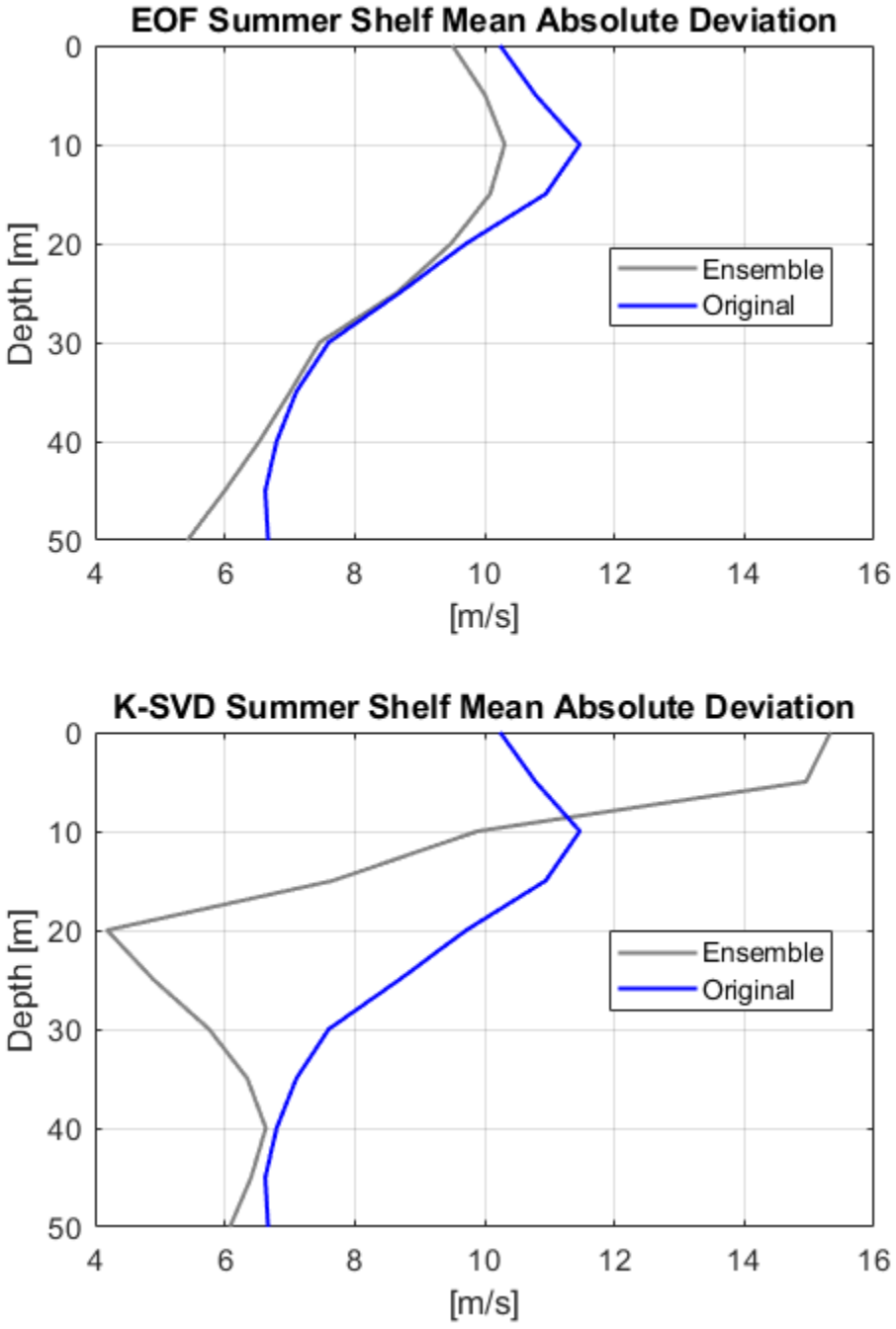


Figure 4.21: Summer shelf mean absolute deviation (MAD) for EOFs (top) and the K-SVD (bottom). Degree of similarity between the lines indicates how well each algorithm covers spatial variability. EOF ensemble MAD closely matches SSP MAD, while K-SVD MAD shows a significant overrepresentation at some depths and underrepresentation at others.

The three leading-order EOFs used for ensemble generation were identified in green in Fig. 4.22. The EOFs were ordered by greatest variance. Variability was largely captured in the first four EOFs, which was consistent with prior work on this subject [25]. The K-SVD performed as expected with variability distributed across atoms (Fig. 4.6). An observation of note was the atoms selected most frequently to represent data in the K-SVD. There was an expectation that the K-SVD might most frequently use dictionary atoms that resembled the leading order EOFs. This was not the case and likely contributed to the EOFs ensemble generation advantage.

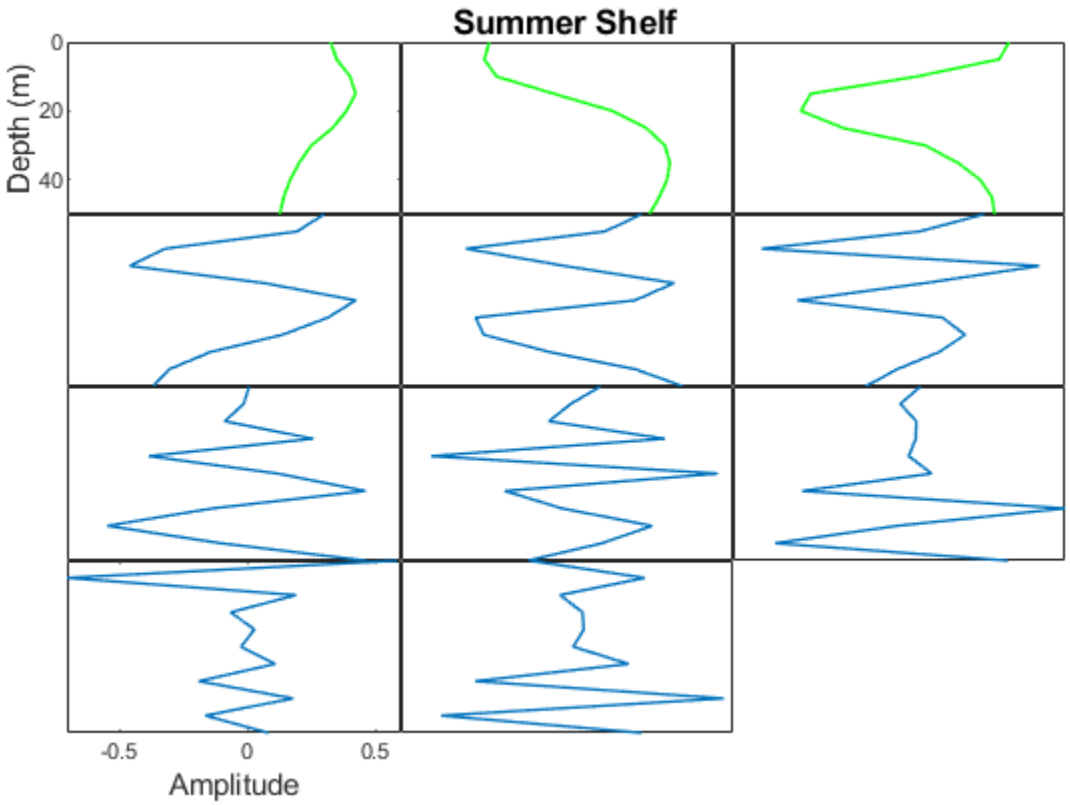


Figure 4.22: Summer shelf EOFs ordered by variance from left to right in a snake pattern. The three leading-order EOFs used for ensemble generation are shown in green.

At the shelf break, four leading-order EOFs were used to represent the data. The EOF ensembles in Fig. 4.23 did not cover variability at depths beyond approximately 70 meters. The lack of variability coverage was due to a decrease in valid input data at deeper depths. As the

density of *in situ* measurements decreased with depth, more data points were necessarily filled with the depth-wise mean. The ensemble convergence toward the center of the original data at deeper depths was the result of this imputation choice.

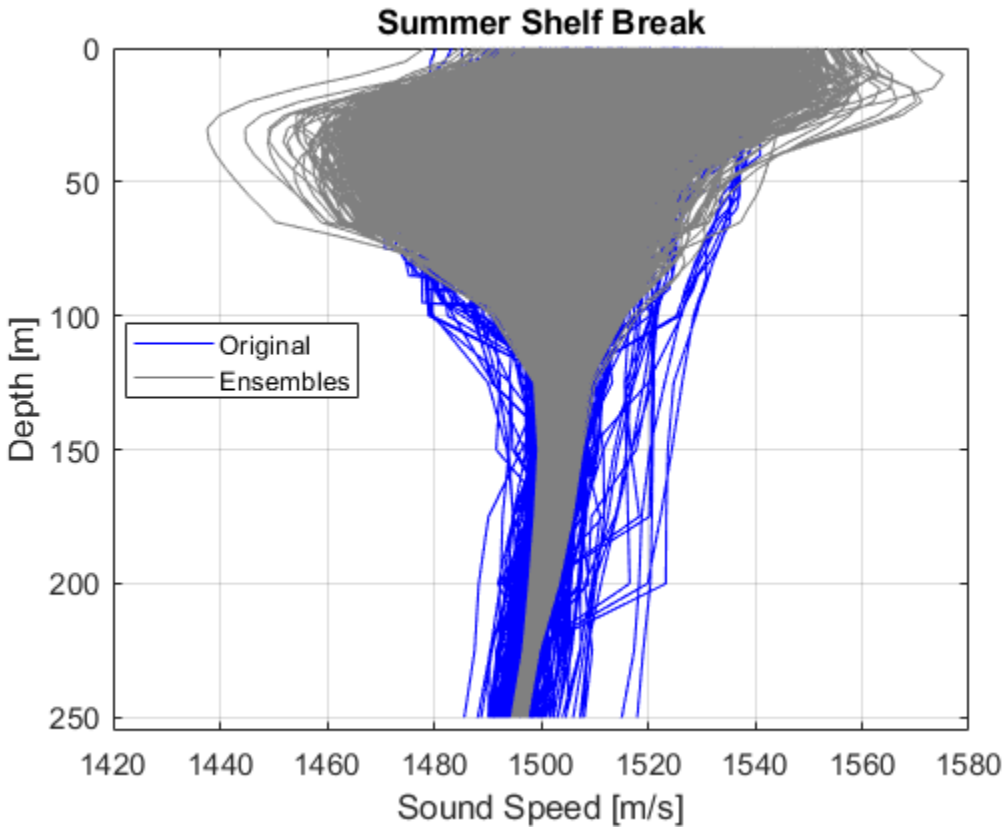


Figure 4.23: EOF ensembles at the shelf break. Four leading-order EOFs were used.

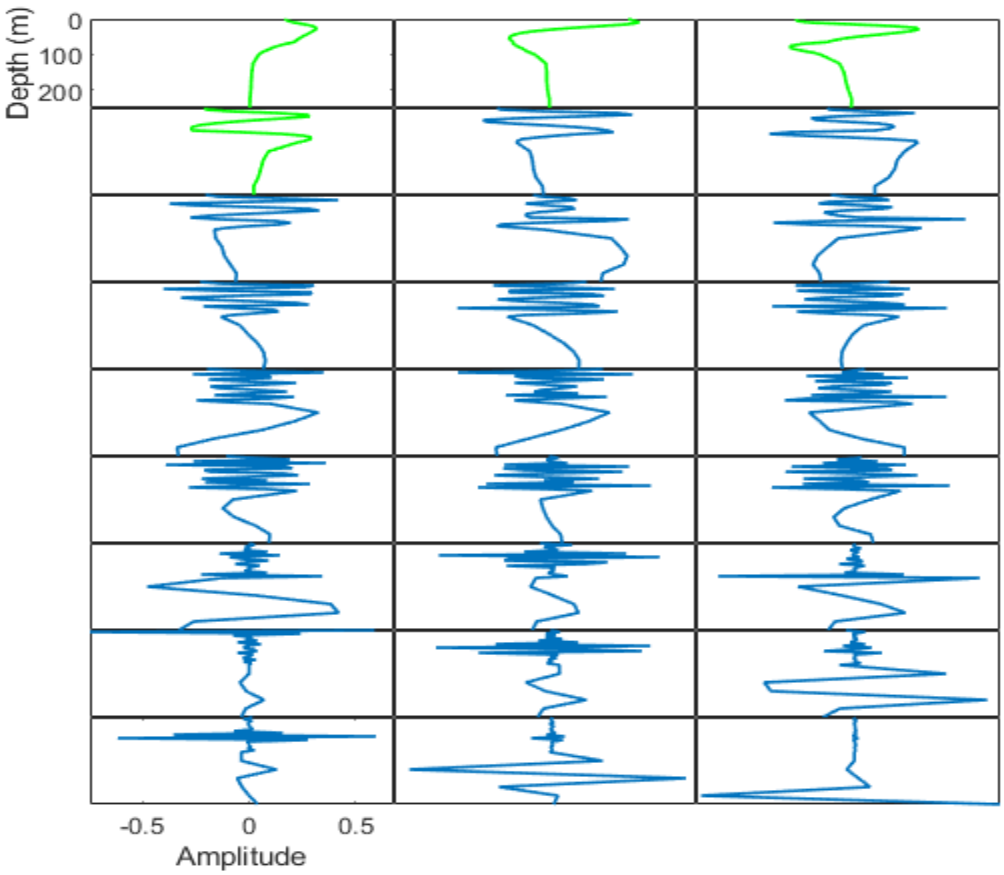


Figure 4.24: Summer shelf break EOFs sorted in order of descending variance. Green atoms represent the four leading-order EOFs used for ensemble generation.

4.2.3 Transmission Loss Ensembles and Distributions

RAM accepts several inputs that include frequency, source depth, range step, and an individual sound speed profile. For the summer shelf dataset, the source frequency was f_s = 250 Hz, and the environment was range-independent with a bottom boundary z_b = 50 m and bottom density ρ_b = 1800 kg/m^3. Source depth z_s and receiver depth z_r were both 35.71 m. There are several more inputs into RAM that have been placed in Appendix C. The K-SVD SSP ensembles from the summer shelf dataset were used as inputs into RAM to calculate the TL ensembles in Fig. 4.25. SSP and TL ensembles were created for each meteorological season at both the shelf

and shelf break, but the focus on the summer shelf dataset continued here. The TL ensembles were used to compute the probability density function as a function of TL in Fig. 4.26. This was a measure of the relative likelihood that TL ensembles would fall within the defined area.

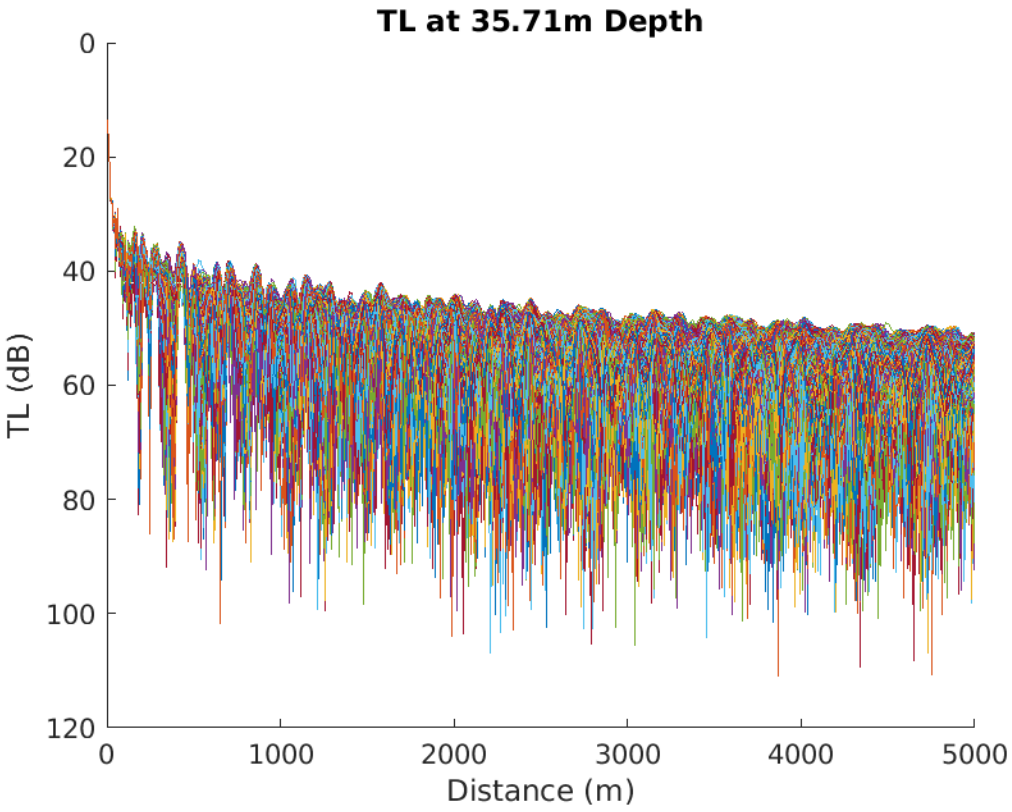


Figure 4.25: Five thousand TL ensembles from summer shelf dataset with a source depth z_s = 35.71 m and receiver depth z_r = 35.71 m.

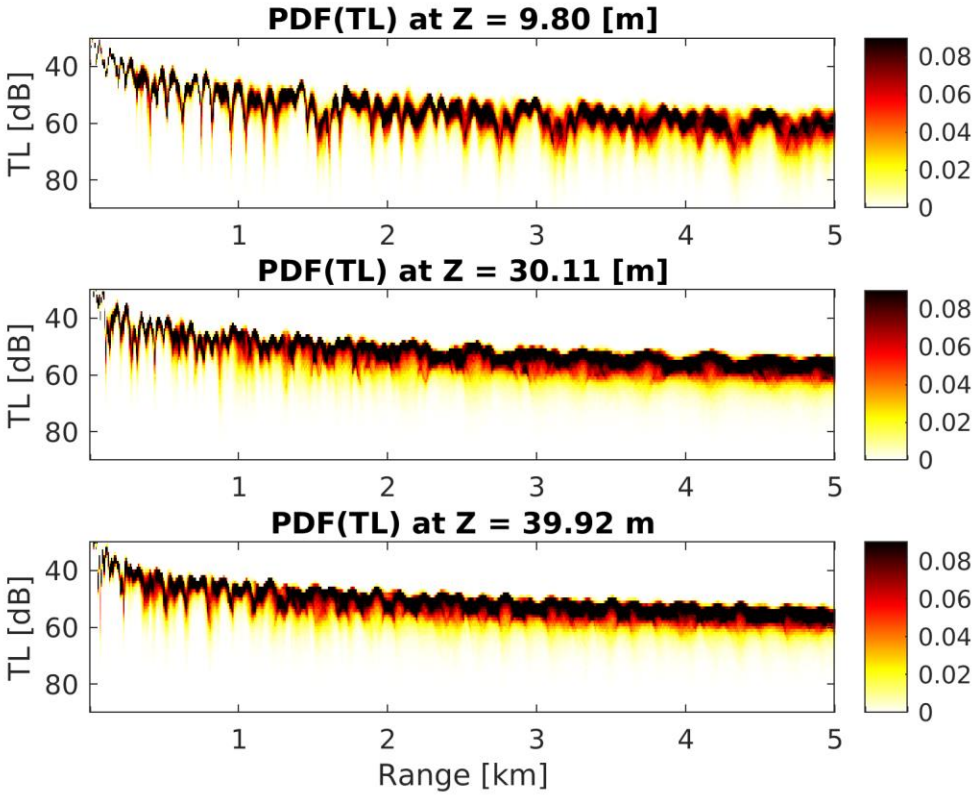


Figure 4.26: Transmission Loss probability density function for summer shelf, where Z was receiver depth. Source depth was 36 m. PDF(TL) was a measure of the relative likelihood a TL ensemble would fall within the observed TL boundaries.

Chapter 5

Conclusion

The main goal of this thesis was to generate SSP ensembles from a stochastic model and calculate TL ensembles. The TL ensembles were computed as a function of SSP ensembles generated from a K-SVD informed stochastic model. This stochastic model was inspired by an EOF stochastic model. A statistical analysis completed on K-SVD outputs determined the stochastic model inputs. The statistical analysis was motivated by answering three questions related to the K-SVD dictionary learning algorithm:

1) What is the optimal sparsity for this dataset?
2) Which atoms are selected most frequently?
3) What distribution do the nonzero coefficients related to these atoms follow?

The optimal sparsity was determined to be $s = 3$ for data defined at the shelf, and $s = 4$ for the shelf break. Representation RMSE (Eqn. (31)) for several values of s, ensemble coverage of original data at given sparsity, and recommendations from [43] were taken into account for this decision.

Question two was answered by identifying which s rows of $\boldsymbol{A}$ contained the most nonzero coefficients. Because of the inherent relationship between the rows of $\boldsymbol{A}$ and the columns (atoms) of $\boldsymbol{D}$, the identified s rows corresponded to the s atoms most frequently used to represent input signals. This was not a deterministic property in the sense that which atoms were chosen most frequently changed with every completed K-SVD run.

The coefficient distribution was assumed to be normal with standard deviation equal to the standard deviation of the coefficients in relevant rows. This was chosen to agree with the EOF stochastic model that assumes normally distributed singular vectors with a standard deviation equal to the square root of singular values to create ensembles. The normal distribution

applied to the K-SVD coefficients was not an ideal fit. A distribution that better captures outliers could help in stabilizing ensemble production.

The EOF stochastic model proved to cover SSP spatial variability reliably and with greater accuracy than the K-SVD stochastic model. Accuracy was quantified by computing the $\boldsymbol{MAD}$ of both algorithm's ensembles and the SSP data (Fig. 4.21). The maximum absolute differences between the ensemble $\boldsymbol{MAD}$ and SSP $\boldsymbol{MAD}$ for the EOF and K-SVD were 1.2 m/s and 5.5 m/s, respectively. The K-SVD ensemble envelope exhibited nondeterministic behavior that resulted in erratic spatial variability coverage. Fig. 4.12 gave reason to believe dictionary atoms were responsible for ensemble envelope shape. The atom shapes did not appear representative of the mean SSP profile in the datasets examined. The leading-order EOF shapes were indicators of their success in generating consistent ensembles as they more closely resembled the mean SSP profile. A definitive explanation for the K-SVD behavior was not identified; however, key insights were made that may lead to a solution in the future. The most plausible reason that EOF analysis showed superior spatial variability coverage is that its explicit goal is to capture spatial variability. The K-SVD's goal is to solve an optimization problem by minimizing Eqn. (3), which results in order reduction. The unexpected shapes of the most frequently selected dictionary atoms were a result of this goal. Furthermore, the stochastic model developed in this thesis inherently favors EOF analysis. The stochastic models had access only to the e leading-order EOFs and the analogous s most frequently selected K-SVD atoms. This benefited EOF analysis because data variance was concentrated in the lead-order EOFs (Fig. 4.22), while data variance was distributed throughout the LD from the K-SVD (Fig. 4.6).

Despite these drawbacks, the K-SVD did produce accurate mean ensemble profiles relative to the mean profile. The K-SVD also allowed for the imputation of zeros for missing values instead of filling missing data with the depth-wise mean like EOF analysis required.

Both the K-SVD and the EOF model poorly represented SSP variability at deeper depths in the shelf break region. This was primarily due to a decrease in SSP data density as depth increased.

Finally, the production of TL ensembles based on K-SVD SSP ensembles was accomplished. These TL ensembles were also used to construct a TL probability density function that has implications outside the scope of this thesis.

Future Work

Identifying or developing a distribution that more accurately captures the behavior of coefficient selection is an intriguing direction for research. Weights that better represent the distribution of nonzero coefficients may contribute to more reliable K-SVD ensemble shapes.

When missing values were present in the dataset, K-SVD representation RMSE showed anomalies of up to 3.19 m/s when $s = 3$, while the mean over the entire sample size was 1.19 m/s. The percentage of missing values relative to valid values has an effect on reconstruction. Research to determine a critical valid value percentage that a K-SVD input matrix is required to have for these anomalies to decrease in frequency is promising.

Further study into the fundamental difference between EOF analysis and the K-SVD is another promising direction for research. The K-SVD is highly amenable to changes in parameters, as well as alternative sparse coding methods. A deeper understanding of the theoretical differences may lead to choices that strengthen the K-SVD's ability to cover the spatial variability of input data.

Using this stochastic model (K-SVD or EOF) has potential to contribute to estimating the probability of detecting marine mammal noise through passive acoustic monitoring using TL PDF information shown in Fig. 4.26.

Bibliography

[1] K. F. Woolfe, S. Lani, K. G. Sabra, and W. A. Kuperman, “Monitoring deep-ocean temperatures using acoustic ambient noise,” *Geophysical Research Letters*, vol. 42, no. 8, pp. 2878–2884, 2015, doi: 10.1002/2015GL063438.

[2] J. G. LeBien and J. W. Ioup, “Species-level classification of beaked whale echolocation signals detected in the northern Gulf of Mexico,” *The Journal of the Acoustical Society of America*, vol. 144, no. 1, pp. 387–396, Jul. 2018, doi: 10.1121/1.5047435.

[3] J. Lynch, G. Gawarkiewicz, Y.-T. Lin, T. Duda, and A. Newhall, “Impacts of Ocean Warming on Acoustic Propagation Over Continental Shelf and Slope Regions,” *Oceanog*, vol. 31, no. 2, Jun. 2018, doi: 10.5670/oceanog.2018.219.

[4] R. C. Beardsley, W. C. Boicourt, and D. V. Hansen, “Physical Oceanography of the Middle Atlantic Bight,” *Am Soc of Limnology and Oceanography Conf on Middle Atlantic Continental Shelf and the Ne*, p. 20, 1975.

[5] F. Aikman III and E. S. Posmentier, “Stratification and shelf-slope interaction in the Middle Atlantic Bight: A numerical study,” *Journal of Geophysical Research: Oceans*, vol. 90, no. C3, pp. 4895–4905, 1985, doi: 10.1029/JC090iC03p04895.

[6] G. Gawarkiewicz, T. G. Ferdelman, T. M. Church, and G. W. Luther, “Shelfbreak frontal structure on the continental shelf north of Cape Hatteras,” *Continental Shelf Research*, vol. 16, no. 14, pp. 1751–1773, Dec. 1996, doi: 10.1016/0278-4343(96)00014-3.

[7] J. A. Benthuysen, “Linear and Nonlinear Stratified Spindown over Sloping Topography,” Ph.D., Massachusetts Institute of Technology, United States -- Massachusetts, 2010. [Online]. Available: https://ezaccess.libraries.psu.edu/login?qurl=https%3A%2F%2Fwww.proquest.com%2Fdissertations-theses%2Flinear-nonlinear-stratified-spindown-over-sloping%2Fdocview%2F820633134%2Fse-2%3Faccountid%3D13158

[8] G. Gawarkiewicz *et al.*, “The Changing Nature of Shelf-Break Exchange Revealed by the OOI Pioneer Array,” *Oceanography*, vol. 31, no. 1, pp. 60–70, 2018.

[9] P. S. Fratantoni, R. S. Pickart, D. J. Torres, and A. Scotti, “Mean Structure and Dynamics of the Shelfbreak Jet in the Middle Atlantic Bight during Fall and Winter,” *Journal of Physical Oceanography*, vol. 31, no. 8, pp. 2135–2156, Aug. 2001, doi: 10.1175/1520-0485(2001)031<2135:MSADOT>2.0.CO;2.

[10] W. G. Zhang and G. G. Gawarkiewicz, “Dynamics of the direct intrusion of Gulf Stream ring water onto the Mid-Atlantic Bight shelf,” *Geophysical Research Letters*, vol. 42, no. 18, pp. 7687–7695, 2015, doi: 10.1002/2015GL065530.

[11] J. Cook, “Pioneer Array Design: Building on the known to explore the unknown,” *Ocean Observatories Initiative*. https://oceanobservatories.org/wp-content/uploads/2019/12/tjoo_a_1679609_f0003_oc-1280.jpg (accessed Feb. 22, 2023).

[12] P. L. Nielsen, M. Siderius, and P. Gerstoft, “Range-dependent geoacoustic inversion: results from the Inversion Techniques Workshop,” *IEEE Journal of Oceanic Engineering*, vol. 28, no. 3, pp. 414–423, Jul. 2003, doi: 10.1109/JOE.2003.816677.

[13] J. Bonnel *et al.*, “Geoacoustic inversion on the New England Mud Patch using warping and dispersion curves of high-order modes,” *The Journal of the Acoustical Society of America*, vol. 143, no. 5, pp. EL405–EL411, May 2018, doi: 10.1121/1.5039769.

[14] Argo, “Argo float data and metadata from Global Data Assembly Centre (Argo GDAC).” SEANOE, 2023. doi: 10.17882/42182.

[15] T. P. Boyer *et al.*, “World Ocean Database 2018.” NOAA National Centers for Environmental Information, 2018. Accessed: Feb. 11, 2023. [Online]. Available: https://www.ncei.noaa.gov/products/world-ocean-database

[16] K. V. Mackenzie, “Nine-term equation for sound speed in the oceans,” *The Journal of the Acoustical Society of America*, vol. 70, no. 3, pp. 807–812, Sep. 1981, doi: 10.1121/1.386920.

[17] I. Le Bras, “HAI_0031-scaled.jpg (2560×1711).” https://oceanobservatories.org/wp-content/uploads/2021/08/HAI_0031-scaled.jpg (accessed Feb. 13, 2023).

[18] J. W. Gibbs, *Elementary principles in statistical mechanics: developed with especial reference to the rational foundation of thermodynamics*. in Yale bicentennial publications. New York: Scribner’s sons, 1902.

[19] M. Schwarts, “Lecture 7: Ensembles.” 2019. Accessed: Mar. 28, 2023. [Online]. Available: https://scholar.harvard.edu/files/schwartz/files/7-ensembles.pdf

[20] H. Hotelling, “Analysis of a complex of statistical variables into principal components,” *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, Sep. 1933, doi: 10.1037/h0071325.

[21] I. T. Jolliffe and J. Cadima, “Principal component analysis: a review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.

[22] A. Hannachi, I. T. Jolliffe, and D. B. Stephenson, “Empirical orthogonal functions and related techniques in atmospheric science: A review,” *International Journal of Climatology*, vol. 27, no. 9, pp. 1119–1152, 2007, doi: 10.1002/joc.1499.

[23] R. Chi-Durán, M. S. Avery, N. Knezek, and B. A. Buffett, “Decomposition of Geomagnetic Secular Acceleration Into Traveling Waves Using Complex Empirical Orthogonal Functions,” *Geophysical Research Letters*, vol. 47, no. 17, p. e2020GL087940, 2020, doi: 10.1029/2020GL087940.

[24] A. Navarra and V. Simoncini, *A Guide to Empirical Orthogonal Functions for Climate Data Analysis*. Dordrecht: Springer Netherlands, 2010. doi: 10.1007/978-90-481-3702-2.

[25] M. Bianco and P. Gerstoft, “Dictionary learning of sound speed profiles,” *J. Acoust. Soc. Am.*, vol. 141, no. 3, pp. 1749–1758, Mar. 2017, doi: 10.1121/1.4977926.

[26] M. D. Collins, “A split-step Padé solution for the parabolic equation method,” *The Journal of the Acoustical Society of America*, vol. 93, no. 4, pp. 1736–1742, 1993, doi: https://doi-org.ezaccess.libraries.psu.edu/10.1121/1.406739.

[27] E. N. Lorenz and Statistical Forecasting Project (Massachusetts Institute of Technology), *Empirical Orthogonal Functions and Statistical Weather Prediction*. in Scientific report. Massachusetts Institute of Technology, Department of Meteorology, 1956. [Online]. Available: https://books.google.com/books?id=q7sJAQAAIAAJ

[28] K. Pearson, “LIII. On lines and planes of closest fit to systems of points in space,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, Nov. 1901, doi: 10.1080/14786440109462720.

[29] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006, doi: 10.1109/TIT.2006.871582.

[30] R. G. Baraniuk, “Compressive Sensing [Lecture Notes],” *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, Jul. 2007, doi: 10.1109/MSP.2007.4286571.

[31] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, May 2009, doi: 10.1016/j.acha.2008.07.002.

[32] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006, doi: 10.1002/cpa.20124.

[33] A. Olaode, G. Naghdy, and C. Todd, “Unsupervised Classification of Images: A Review,” *International Journal of Image Processing*, vol. 8, pp. 2014–325, Sep. 2014.

[34] K. Xu, W. Yang, G. Liu, and H. Sun, “Unsupervised Satellite Image Classification Using Markov Field Topic Model,” *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 1, pp. 130–134, Jan. 2013, doi: 10.1109/LGRS.2012.2194770.

[35] G. H. Golub and C. F. Van Loan, *Matrix computations*, Fourth edition. in Johns Hopkins studies in the mathematical sciences. Baltimore: The Johns Hopkins University Press, 2013.

[36] K. Engan, S. O. Aase, and J. Hakon Husoy, “Method of optimal directions for frame design,” in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, Mar. 1999, pp. 2443–2446 vol.5. doi: 10.1109/ICASSP.1999.760624.

[37] K. Engan, “Frame Design Using FOCUSS With Method of Optimal Directions”.

[38] L. Bottou, “On-line learning and stochastic approximations,” in *On-line learning in neural networks*, USA: Cambridge University Press, 1999, pp. 9–42.

[39] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, “Online dictionary learning for sparse coding,” in *Proceedings of the 26th Annual International Conference on Machine Learning*, in ICML ’09. New York, NY, USA: Association for Computing Machinery, Jun. 2009, pp. 689–696. doi: 10.1145/1553374.1553463.

[40] M. Aharon, M. Elad, and A. Bruckstein, “K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006, doi: 10.1109/TSP.2006.881199.

[41] S. CHEN, S. A. BILLINGS, and W. LUO, “Orthogonal least squares methods and their application to non-linear system identification,” *International Journal of Control*, vol. 50, no. 5, pp. 1873–1896, Nov. 1989, doi: 10.1080/00207178908953472.

[42] S. G. Mallat and Z. Zhang, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993, doi: 10.1109/78.258082.

[43] L. N. Smith and M. Elad, “Improving Dictionary Learning: Multiple Dictionary Updates and Coefficient Reuse,” *IEEE Signal Process. Lett.*, vol. 20, no. 1, pp. 79–82, Jan. 2013, doi: 10.1109/LSP.2012.2229976.

[44] A. Gersho and R. M. Gray, *Vector quantization and signal compression*. in The Kluwer international series in engineering and computer science, no. SECS 159. Boston: Kluwer Academic Publishers, 1992.

[45] J. Macqueen, “Some Methods for Classification and Analysis of Multivariate Observations,” in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 1967, pp. 281–297.

[46] J. A. Hartigan and M. A. Wong, “Algorithm AS 136: A K-Means Clustering Algorithm,” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979, doi: 10.2307/2346830.

[47] K. Engan, S. O. Aase, and J. H. Husøy, “Multi-frame compression: theory and design,” *Signal Processing*, vol. 80, no. 10, pp. 2121–2140, Oct. 2000, doi: 10.1016/S0165-1684(00)00072-4.

[48] A. Ben-Israel, “The Moore of the Moore-Penrose inverse,” *ELA*, vol. 9, Jan. 2002, doi: 10.13001/1081-3810.1083.

[49] “Matching Pursuit Algorithms - MATLAB & Simulink.” https://www.mathworks.com/help/wavelet/ug/matching-pursuit-algorithms.html (accessed Mar. 27, 2023).

[50] M. A. Leontovich and V. A. Fock, “Solution of the Problem of Propagation of Electromagnetic Waves along the Earth’s Surface by Method of Parabolic Equations,” *Journal of Physics—USSR*, vol. 10, pp. 13–24, 1946.

[51] P. L. Kelley, “Self-Focusing of Optical Beams,” *Phys. Rev. Lett.*, vol. 15, no. 26, pp. 1005–1008, Dec. 1965, doi: 10.1103/PhysRevLett.15.1005.

[52] V. I. Karpman, *Nonlinear Waves in Dispersive Media*. Pergamon Press, 1975.

[53] J. F. Claerbout, “Coarse grid calculations of waves in inhomogeneous media with application to delineation of complicated seismic structure,” *Geophysics*, vol. 35, no. 3, pp. 407–418, 1970.

[54] F. D. Tappert and R. H. Hardin, “Applications of the split-step fourier method to the numerical solution of nonlinear and variable coefficient wave equations,” presented at the SIAM 1972 Fall Meeting, Austin, Texas: SIAM Review, 1973, p. 423.

[55] F. D. Tappert, “The parabolic approximation method,” in *Wave Propagation in Underwater Acoustics*, New York: Springer, Berlin, Heidelberg, 1977, pp. 224–287.

[56] J. A. Davis, D. White, and R. C. Cavanagh, “NORDA Parabolic Equation Workshop,” Naval Ocean Research and Development Activity, Stennis Space Center, MS, TN-143, 1982.

[57] M. D. Collins, “User’s Guide for RAM versions 1.0 and 1.0p,” Naval Research Laboratory, Washington DC, USA, 1995.

[58] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, “Chapter 6: Parabolic Equations,” in *Computational Ocean Acoustics: Second Edition*, 2nd ed.in Modern Acoustics and Signal Processing. New York: Springer, 2011, pp. 457–529.

[59] D. J. Thomson and N. R. Chapman, “A wide-angle split-step algorithm for the parabolic equation,” *The Journal of the Acoustical Society of America*, vol. 74, no. 6, pp. 1848–1854, Dec. 1983, doi: 10.1121/1.390272.

[60] M. D. Feit and J. A. Fleck, “Light propagation in graded-index optical fibers,” *Appl. Opt., AO*, vol. 17, no. 24, pp. 3990–3998, Dec. 1978, doi: 10.1364/AO.17.003990.

[61] J. F. Claerbout, *Fundamentals of geophysical data processing: with applications to petroleum prospecting*. Palo Alto, CA: Blackwell Scientific Publications, 1985.

[62] M. D. Collins, “Applications and time-domain solution of higher-order parabolic equations in underwater acoustics,” *The Journal of the Acoustical Society of America*, vol. 86, no. 3, pp. 1097–1102, Sep. 1989, doi: 10.1121/1.398101.

[63] M. D. Collins, “A self-starter for the parabolic equation method,” *The Journal of the Acoustical Society of America*, vol. 92, no. 4, pp. 2069–2074, Oct. 1992, doi: 10.1121/1.405258.

[64] “Dataset Search,” *National Centers for Environmental Information*. https://www.ncei.noaa.gov/access/search/dataset-search?startDate=2015-01-05T00:00:00&endDate=2015-01-05T23:59:59 (accessed Feb. 21, 2023).

[65] “World Ocean Database Select.” https://www.ncei.noaa.gov/access/world-ocean-database-select/depth_definition.html (accessed Jan. 05, 2023).

[66] A. J. Jerri, “The Shannon sampling theorem—Its various extensions and applications: A tutorial review,” *Proceedings of the IEEE*, vol. 65, no. 11, pp. 1565–1596, Nov. 1977, doi: 10.1109/PROC.1977.10771.

[67] C. L. Pekeris, “Theory of Propagation of Explosive Sound in Shallow Water,” in *Geological Society of America Memoirs*, Geological Society of America, 1948, pp. 1–116. doi: 10.1130/MEM27-2-p1.

[68] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, “Chapter 2: Layered Media and Waveguides,” in *Computational Ocean Acoustics: Second Edition*, 2nd ed.in Modern Acoustics and Signal Processing. New York: Springer, 2011, pp. 65–153.

Appendices

Appendix A

K-SVD Example

To demonstrate the K-SVD's versatility, it was applied as a means to compress and reconstruct ocean temperature profiles at the NESB. The dataset consisted of water temperature profiles retrieved by an Expendable BathyThermograph (XBT) at the NESB (same geographical location as the CTD data) shown in Fig. A.1. This is a compressive sensing application based on the principle that a signal can be reconstructed from much fewer samples than prescribed by the Nyquist-Shannon Sampling Theorem [29], [30], [32], [66]. All K-SVD parameters were set in accordance with [43]. Fig. A.1 shows an example of an XBT device. Fig. A.2 shows the original and reconstructed data at a depth of 20 m, as well as the residuals between the two. The original dataset was stored in a *.mat* binary file with a size of 417 kilobytes. The dictionary, sparse coefficient matrix, and mean temperature profile required to reconstruct the data was 234 kilobytes. This was a 43% decrease in file size.

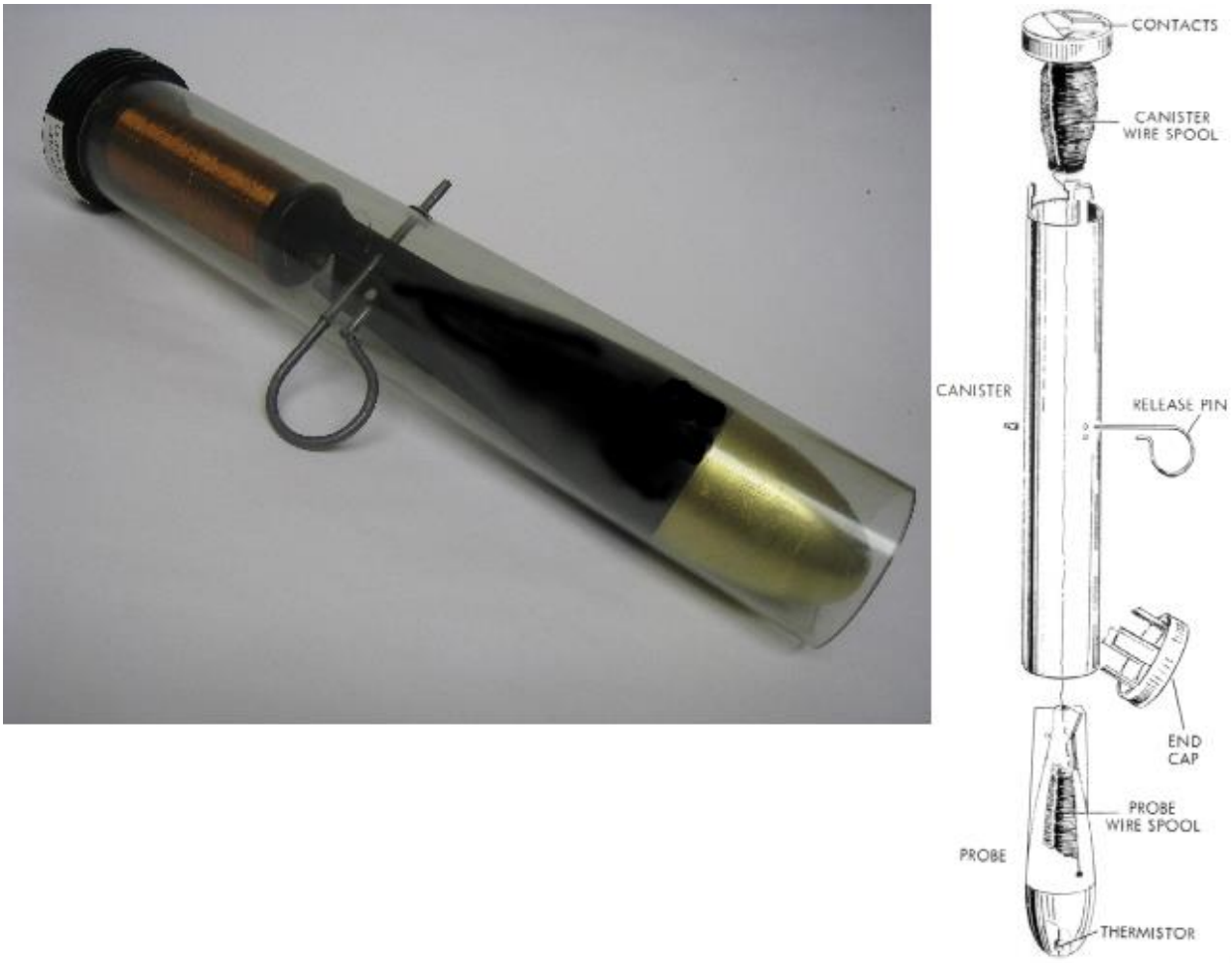


Figure A.1: Expendable BathyThermograph (XBT) device used to measure water temperature (above). Inner-workings of the XBT device (right).

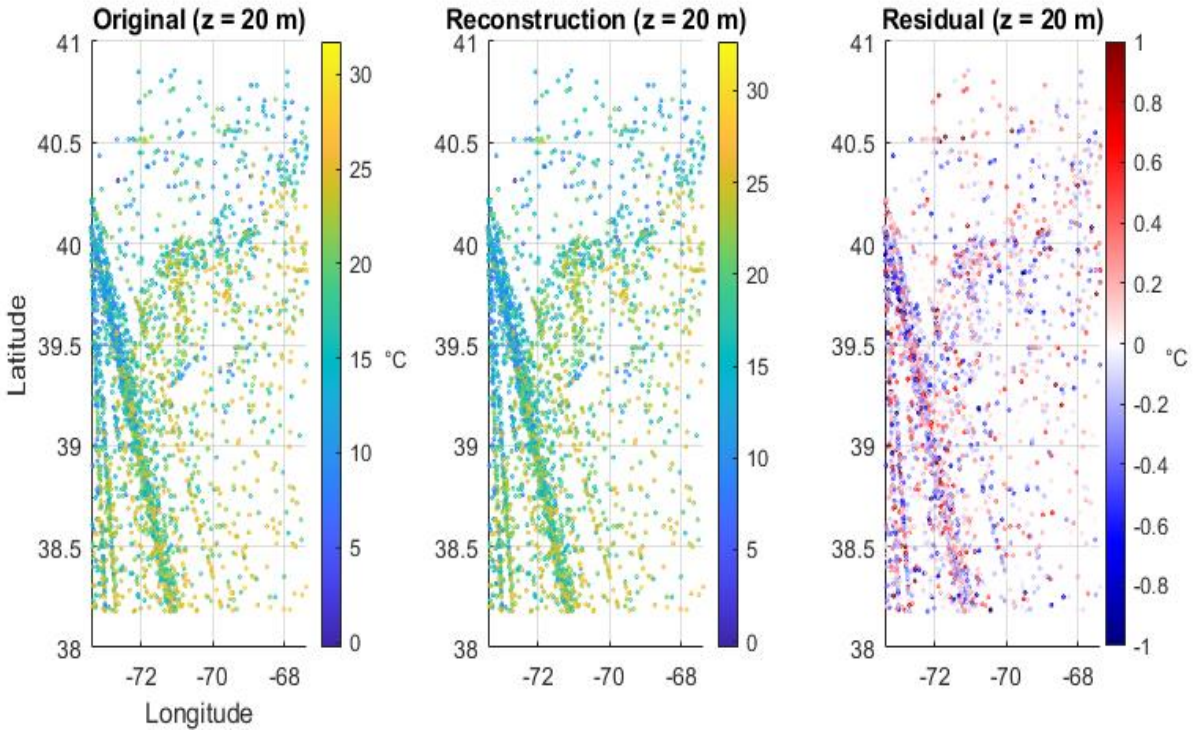


Figure A.2: Sample of ocean temperature profiles at the NESB for z = 20 m. Original data (left), K-SVD reconstruction (middle), and residual between the two (right).

Appendix B

RAM Example

The range-independent Pekeris waveguide solution was chosen as an ideal case to demonstrate the output of RAM. The Pekeris waveguide shown in Fig. B.1 is a horizontally stratified environment with an isovelocity water layer above an isovelocity fluid halfspace that simulates a simple bottom layer [67], [68]. By modeling the bottom layer as an infinite fluid halfspace, how energy leaks out of the waveguide becomes apparent. The modes that escape the waveguide surpass a critical angle of propagation relative to the horizontal and leak energy into the fluid halfspace. These are known as leaky modes.

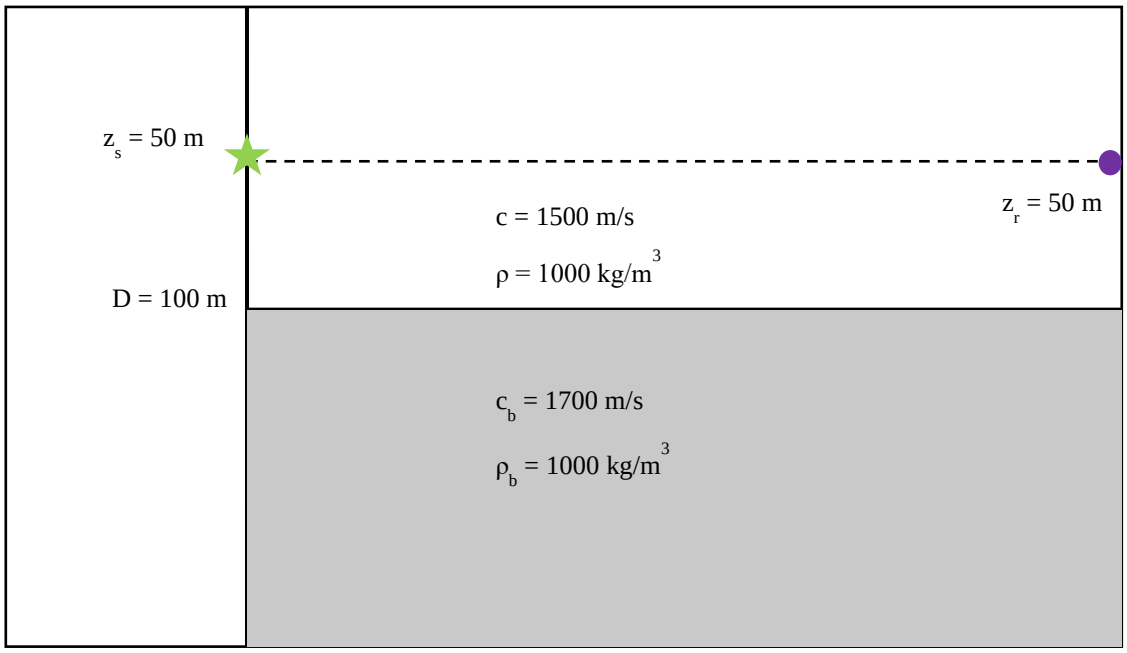


Figure B.1: Pekeris waveguide environment - isovelocity water layer above an isovelocity half space with a different sound speed.

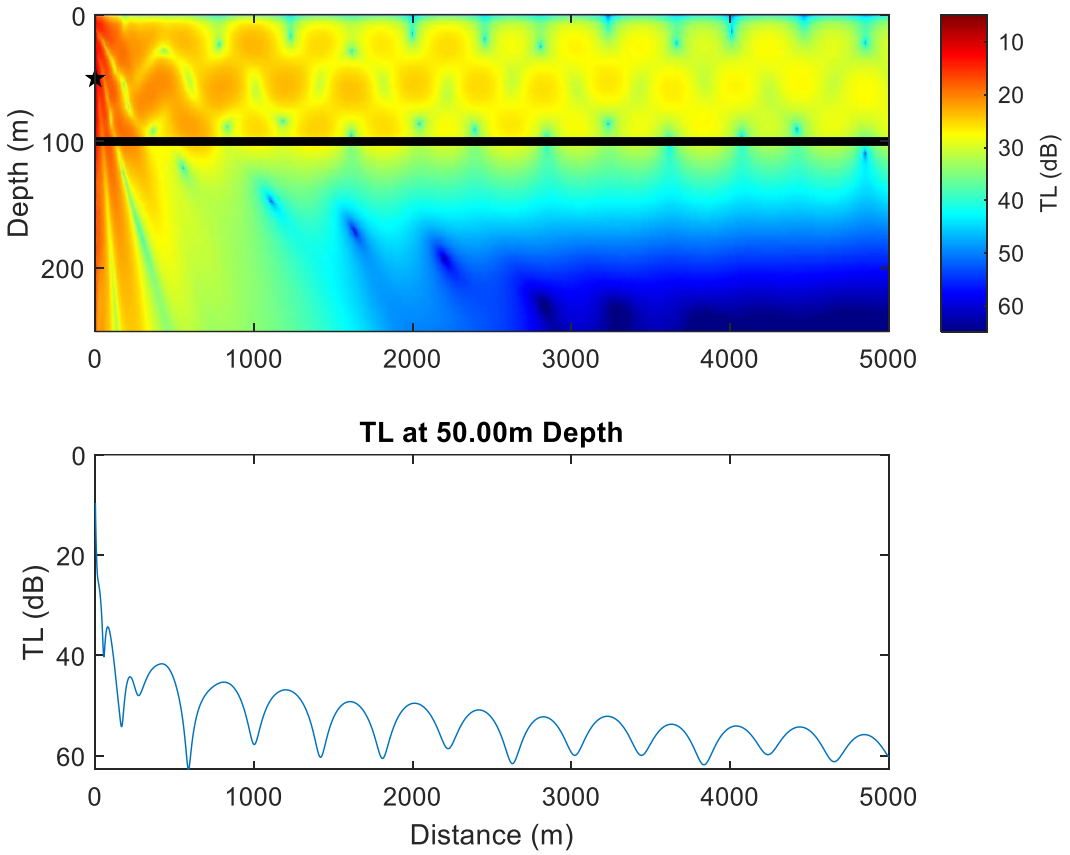


Figure B.2: Colormap showing TL in the Pekeris waveguide with a source frequency of 50 Hz, and source and receiver depth of 50 m. The leaky modes are the three finger-like beams within 500 m (in range) of the source (above). TL at the receiver is shown at 50 m.

They are visible near the source, penetrating the bottom layer in Fig. B.2. The source frequency is 50 Hz, and the source depth is 50 m. The receiver depth is also at 50 m. RAM models the Transmission Loss as a function of depth and range.

As a benchmark for performance, a Pekeris waveguide example from [68] shown in Fig. B.3 with source frequency f_s = 50 Hz, source depth (z_s) and receiver depth (z_r) equal to 36 m, sound speed in water c_w = 1500 m/s, sound speed in the bottom layer c_b = 1800 m/s, density of water ρ_w = 1000 kg/m^3, and density in bottom layer ρ_b = 1800 kg/m^3 was modeled in RAM (Fig. B.4). A visual inspection showed that they agree.

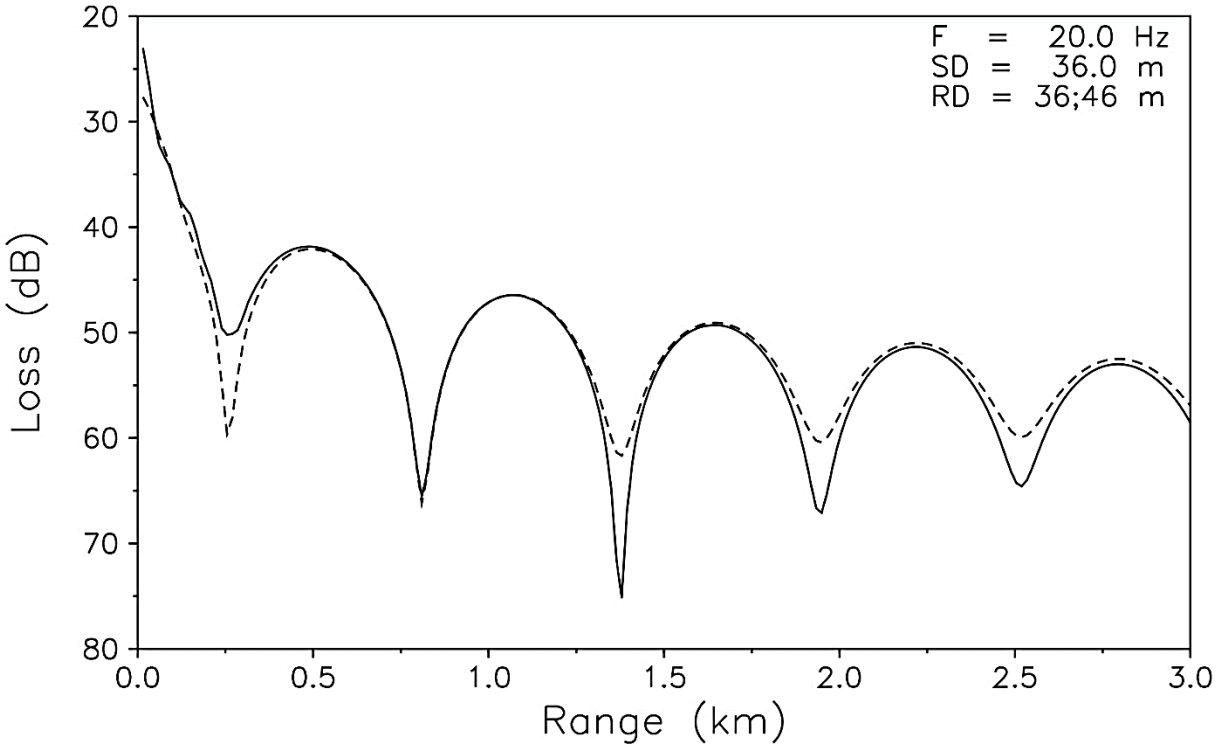


Figure B.3: Transmission Loss in a Pekeris waveguide of source with frequency 20 Hz and waveguide depth of 100 meters. The source depth and receiver depth are 36 meters (modified) [68]. Solid curve is TL at receiver depth of 36 m and dashed curve is TL at receiver depth of 46 m.

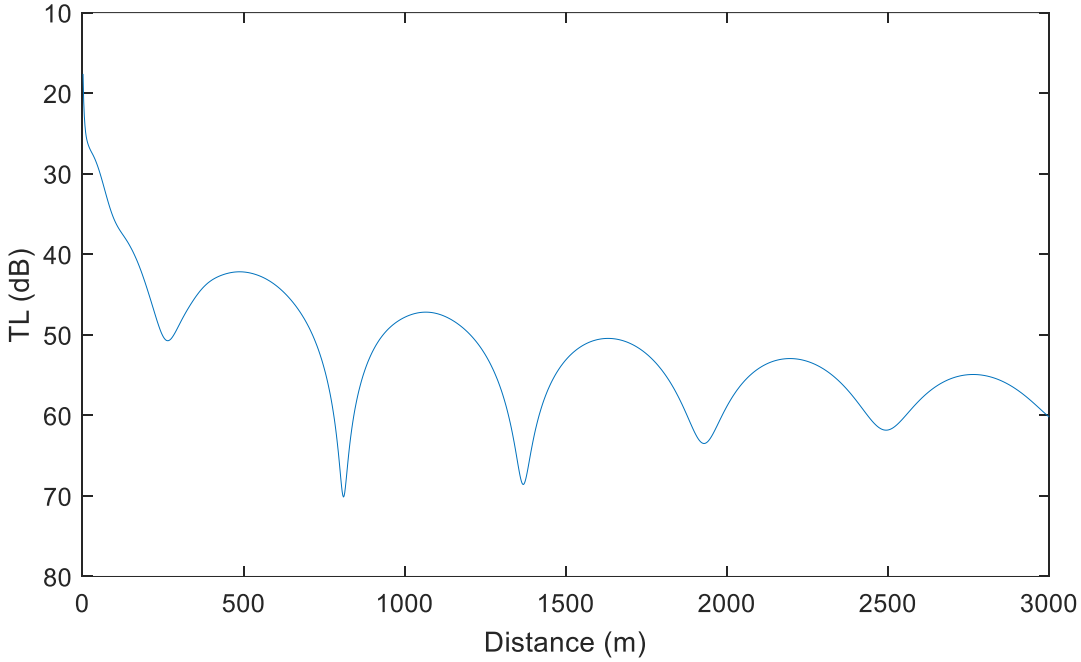


Figure B.4: RAM solution for Transmission Loss with identical parameters of solid curve in Fig. 5.5.

Appendix C

RAM Input Definitions

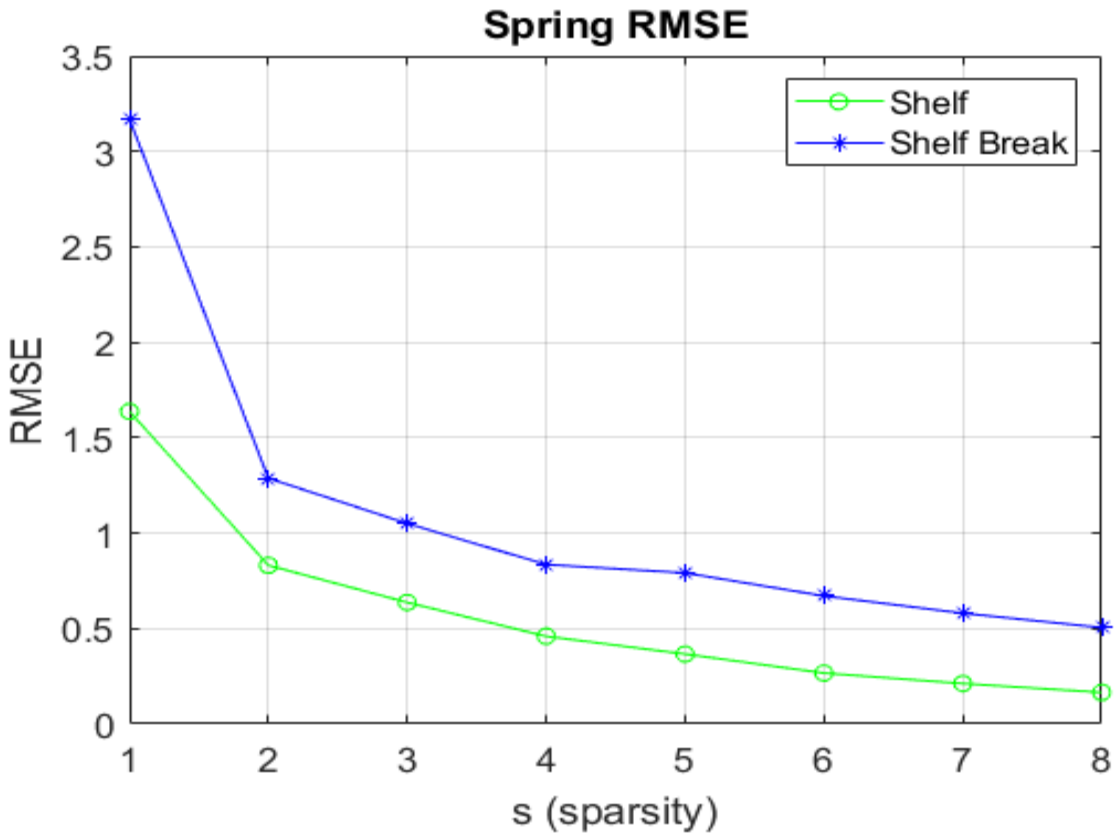
Code snippet from *ram.m* function with input definitions.

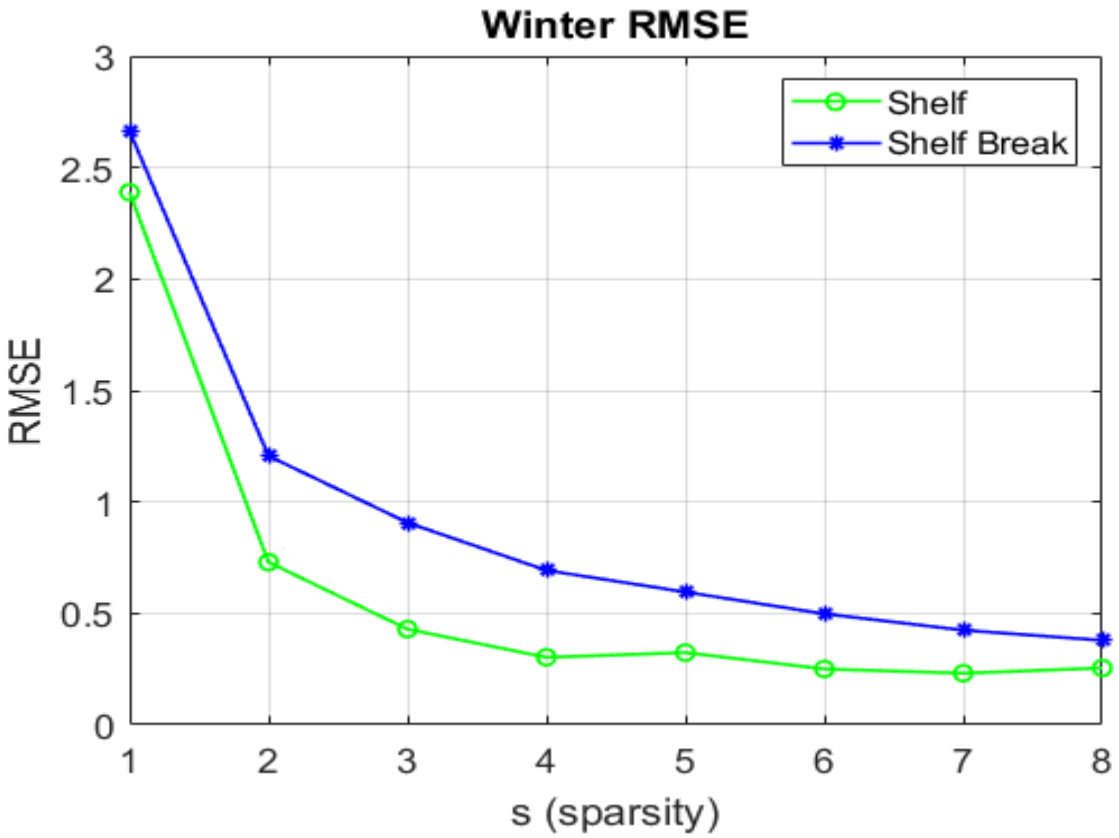
```
function [fld,zg,rout]=ram(frq,zsrc,dim,rg,dr,zmax,dz,dzm,zmplt,c0,np,ns,rs,...
                           rb,zb,rp,zw,cw,zs,cs,zr,rho,za,attn)

% frq       frequency
% zsrc      source depth
% dim       2-d starter (dim=2) or 3-d starter (dim=3)
% rg        vector of output ranges, the first range must be the PE starter position
% dr        range step
% zmax      max depth
% dz        depth grid increment
% dzm       decimate output depth grid (dzm=1, no decimation)
% xmplt     maximum depth of output
% c0        "mean" sound speed
% np        # of pade coefficients
% ns        # of stability terms
% rs        stability range
% rb        bathymetry range
% zb        bathymetry
% rp        profile ranges(nr)
% zw        sound speed grid depth(nzw)
% cw        sound speed(nr,nzw)
% zs        sediment speed grid depth(nzs)
% cs        sediment speed(nr,nzs)
% zr        density depth grid(nzr)
% rho       density(nr,nzr)
% za        attenuation depth grid(nza)
% attn      attenuation(nr,nza)
%
```

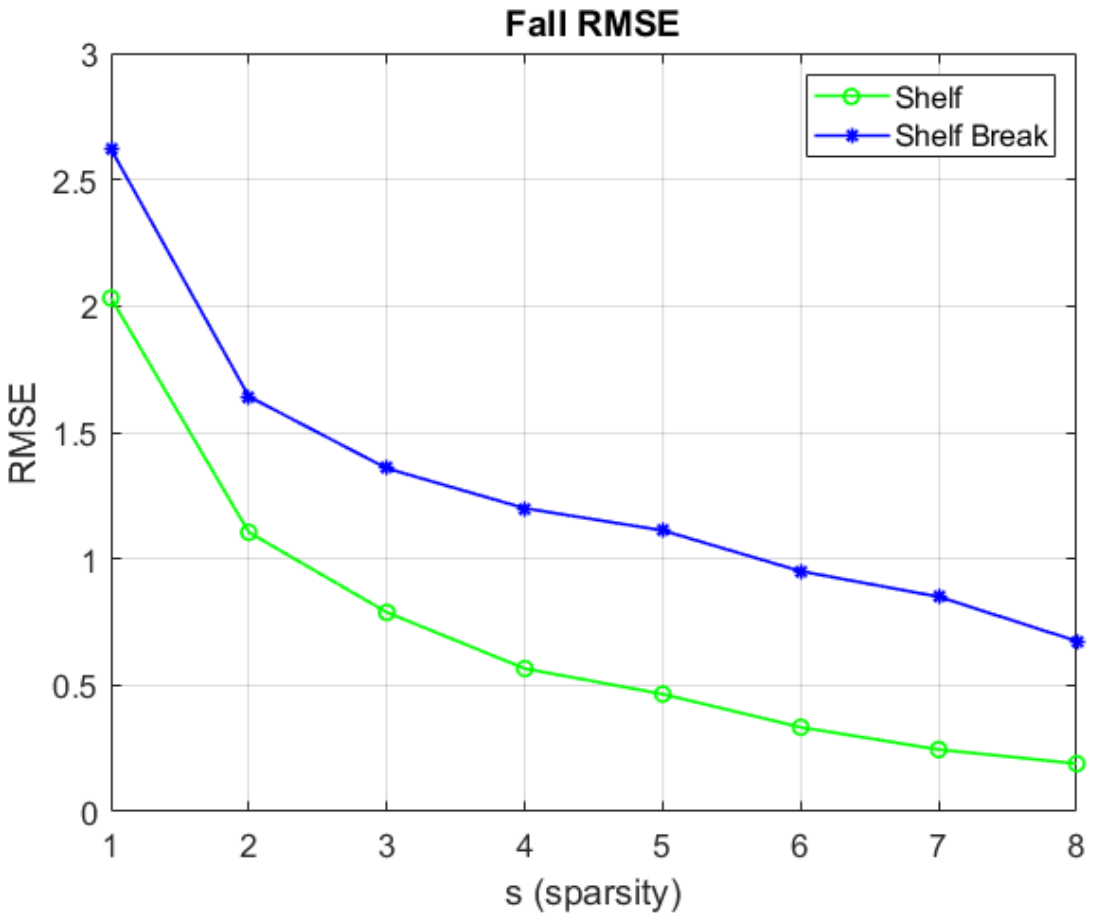
Appendix D

K-SVD Related Figures

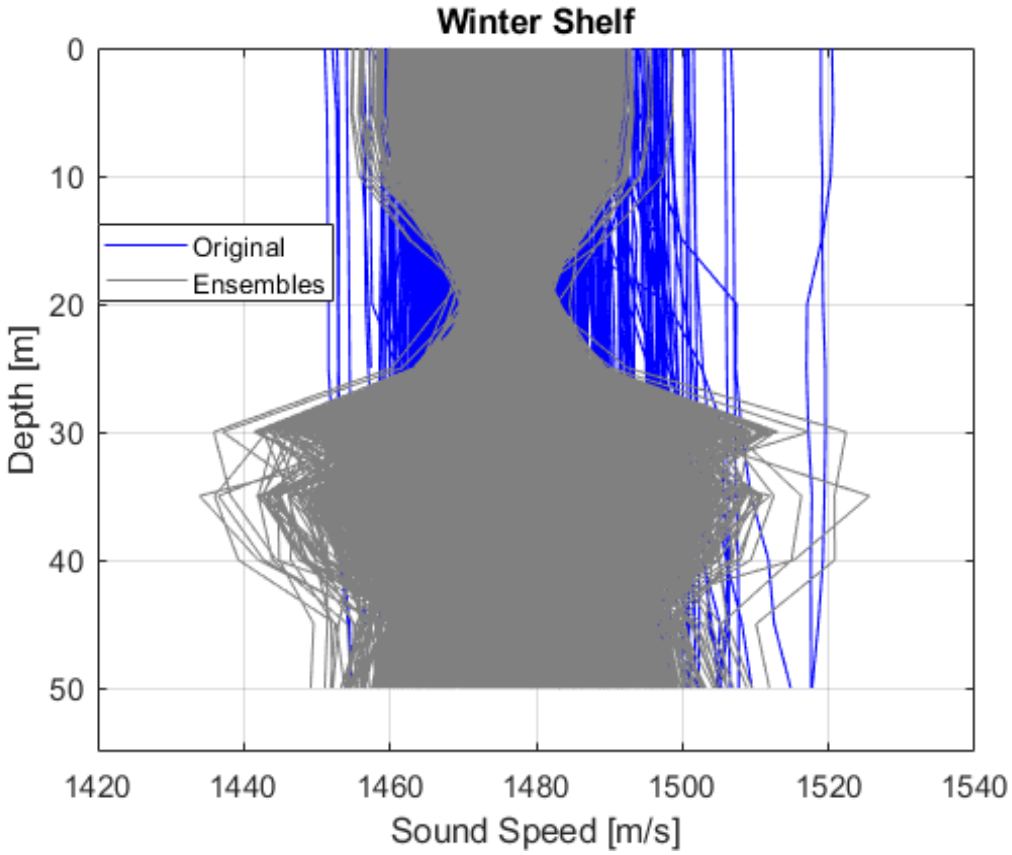
K-SVD representation error figures and ensemble figures for each season and region.

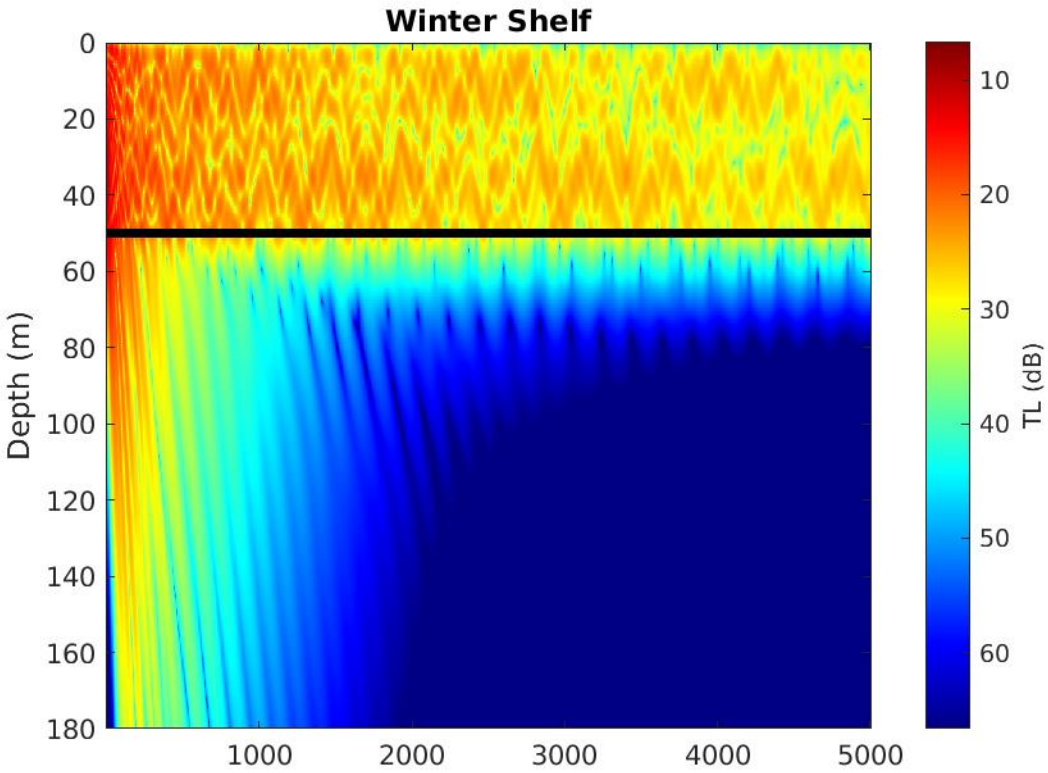

Spring RMSE
Shelf
Shelf Break
RMSE
s (sparsity)
0
0.5
1
1.5
2
2.5
3
3.5
1
2
3
4
5
6
7
8

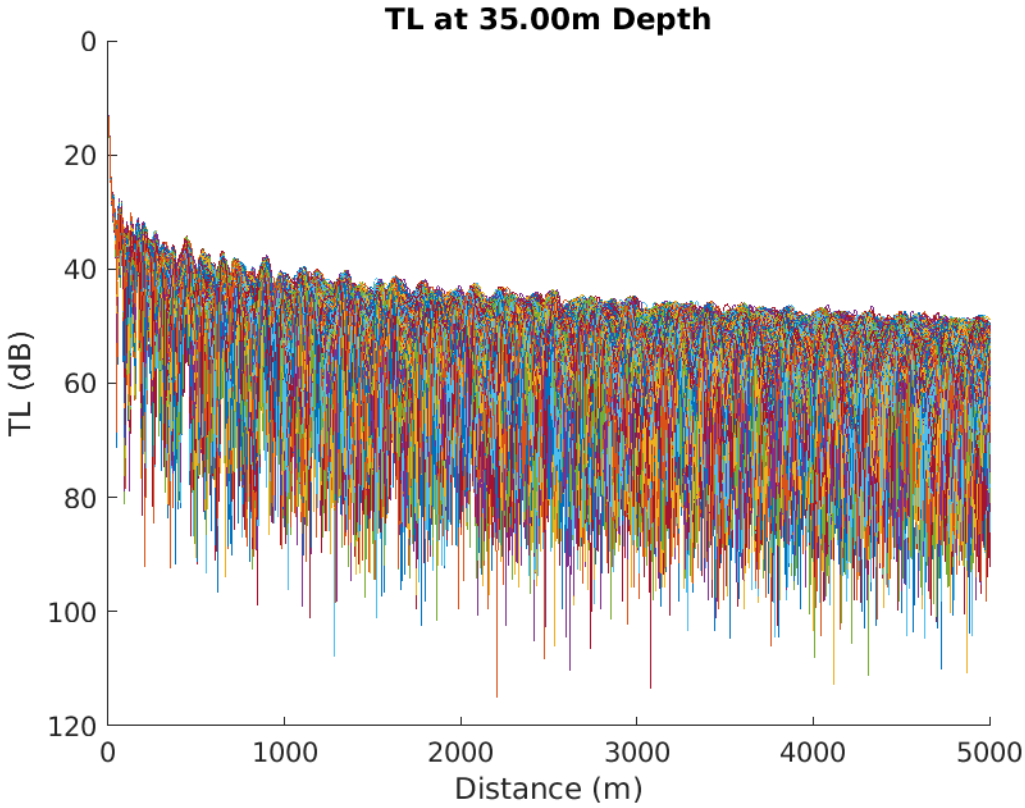

Winter RMSE
Shelf
Shelf Break
RMSE
s (sparsity)
0
0.5
1
1.5
2
2.5
3
1
2
3
4
5
6
7
8

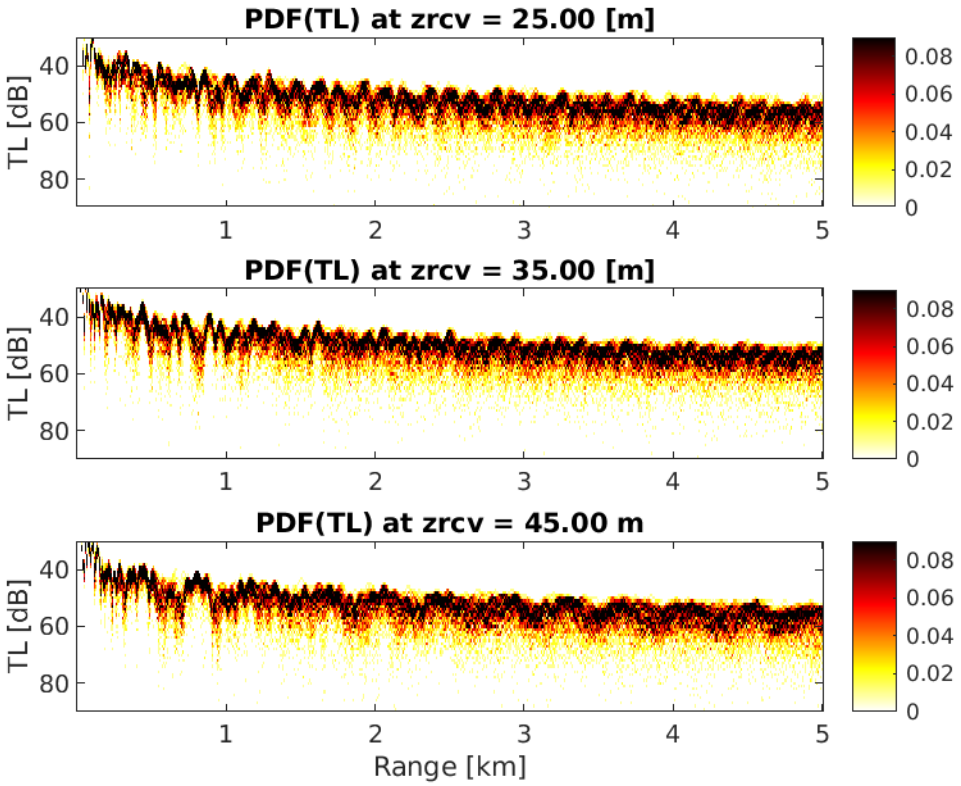

Fall RMSE
Shelf
Shelf Break
RMSE
s (sparsity)
0
0.5
1
1.5
2
2.5
3
1
2
3
4
5
6
7
8

Winter Shelf

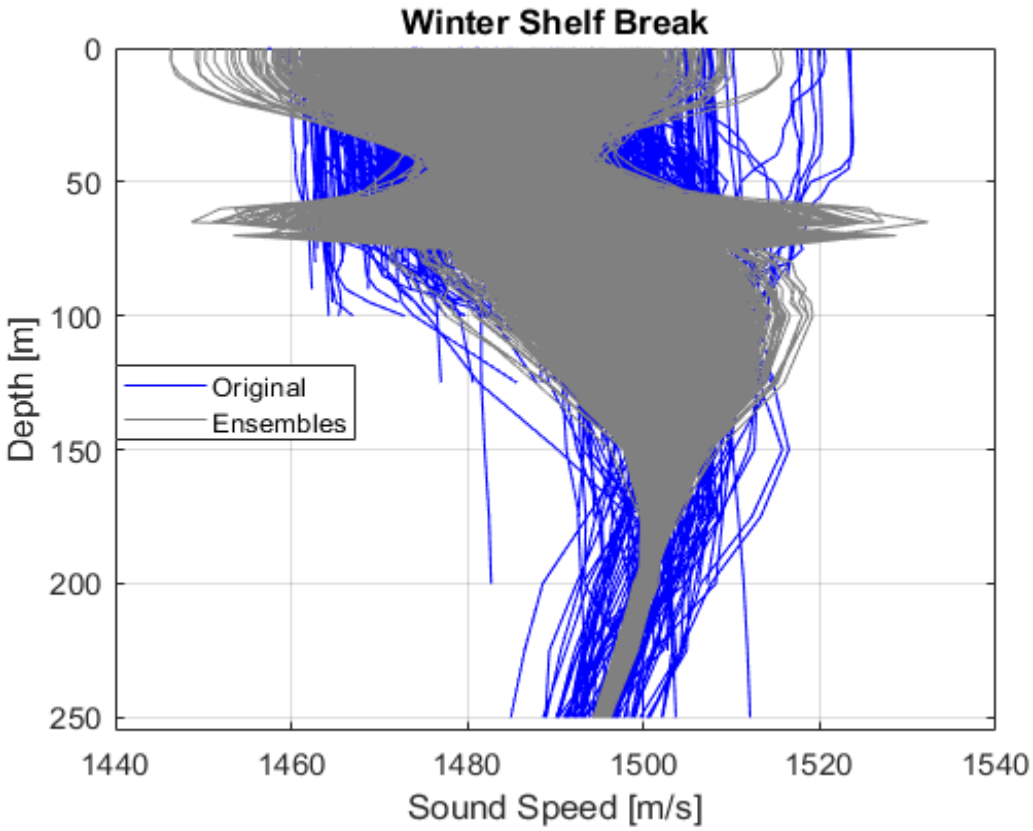


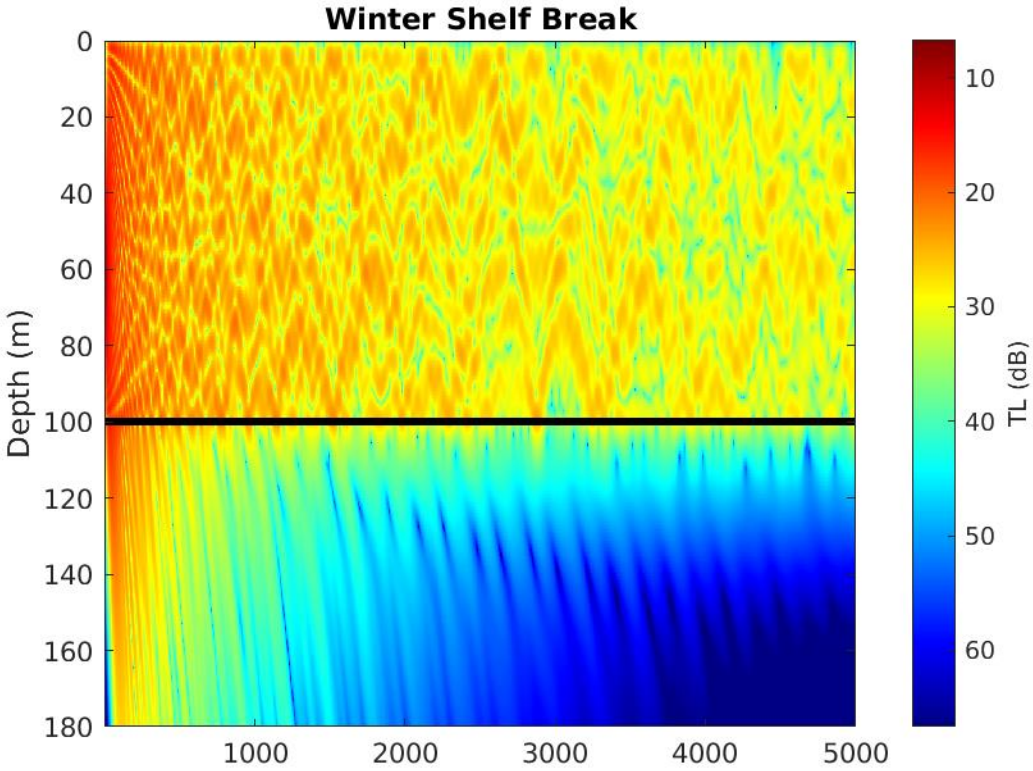

Winter Shelf
Depth (m)
0
20
40
60
80
100
120
140
160
180
1000
2000
3000
4000
5000
TL (dB)
10
20
30
40
50
60

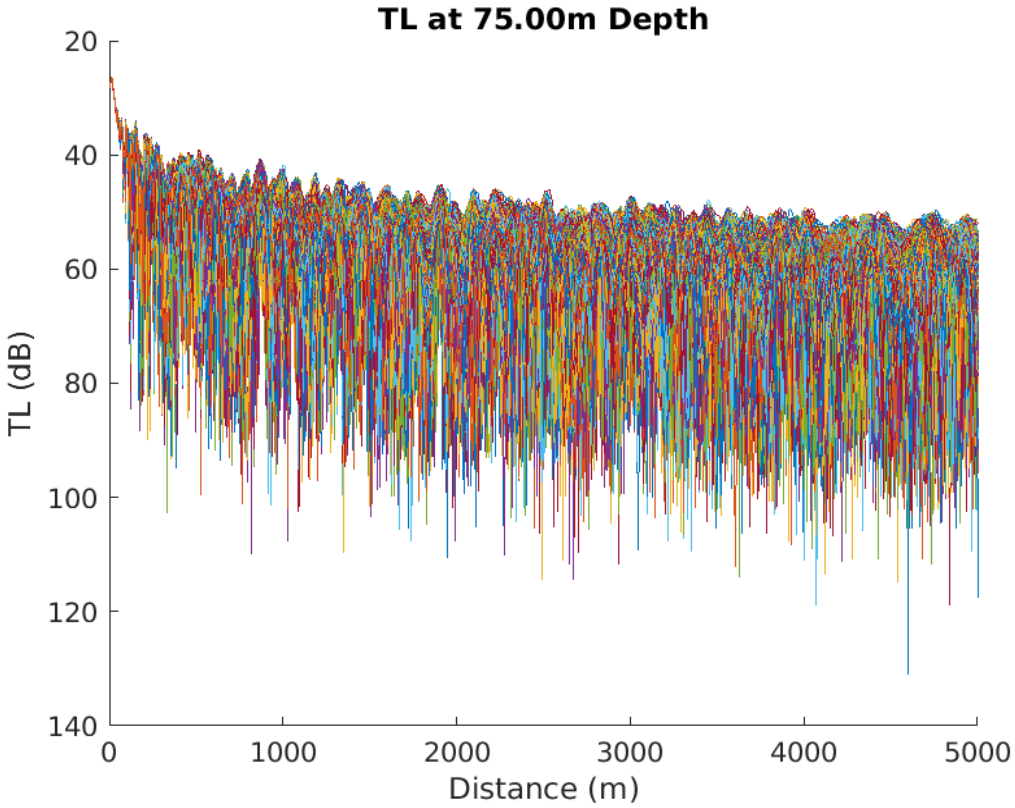

TL at 35.00m Depth
TL (dB)
0
20
40
60
80
100
120
0
1000
2000
3000
4000
5000
Distance (m)

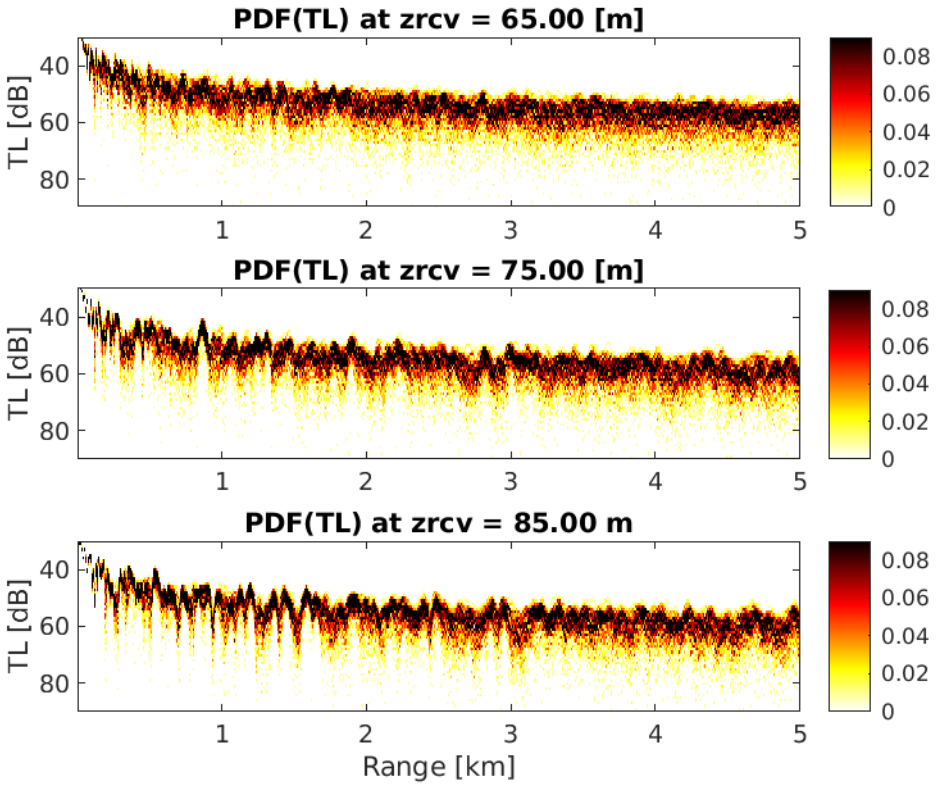

PDF(TL) at zrcv = 25.00 [m]
PDF(TL) at zrcv = 35.00 [m]
PDF(TL) at zrcv = 45.00 m
TL [dB]
40
60
80
1
2
3
4
5
Range [km]
0.08
0.06
0.04
0.02
0

Winter Shelf Break

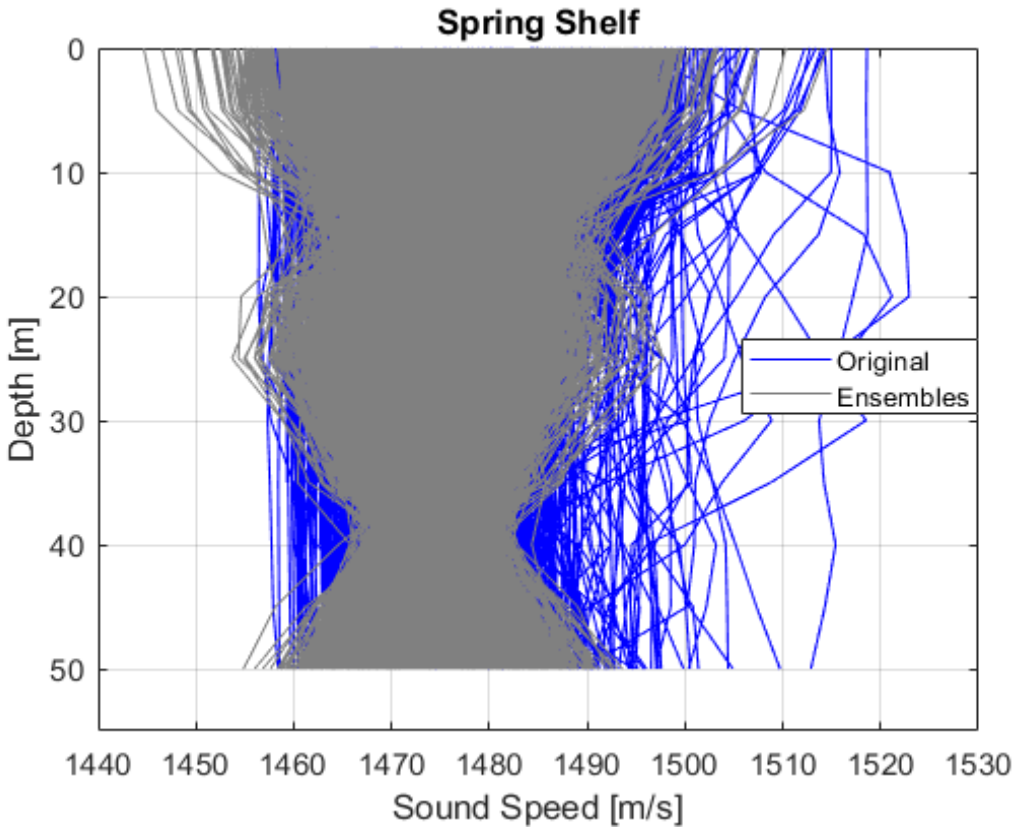


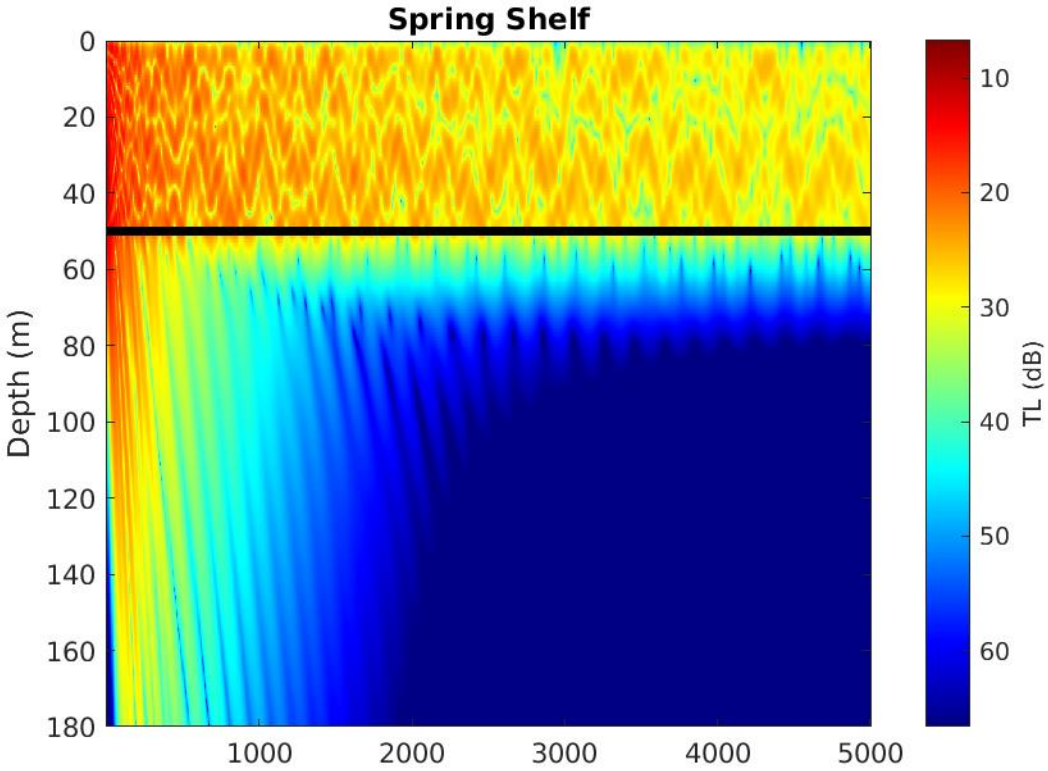

Winter Shelf Break
Depth (m)
0
20
40
60
80
100
120
140
160
180
1000
2000
3000
4000
5000
10
20
30
40
50
60
TL (dB)

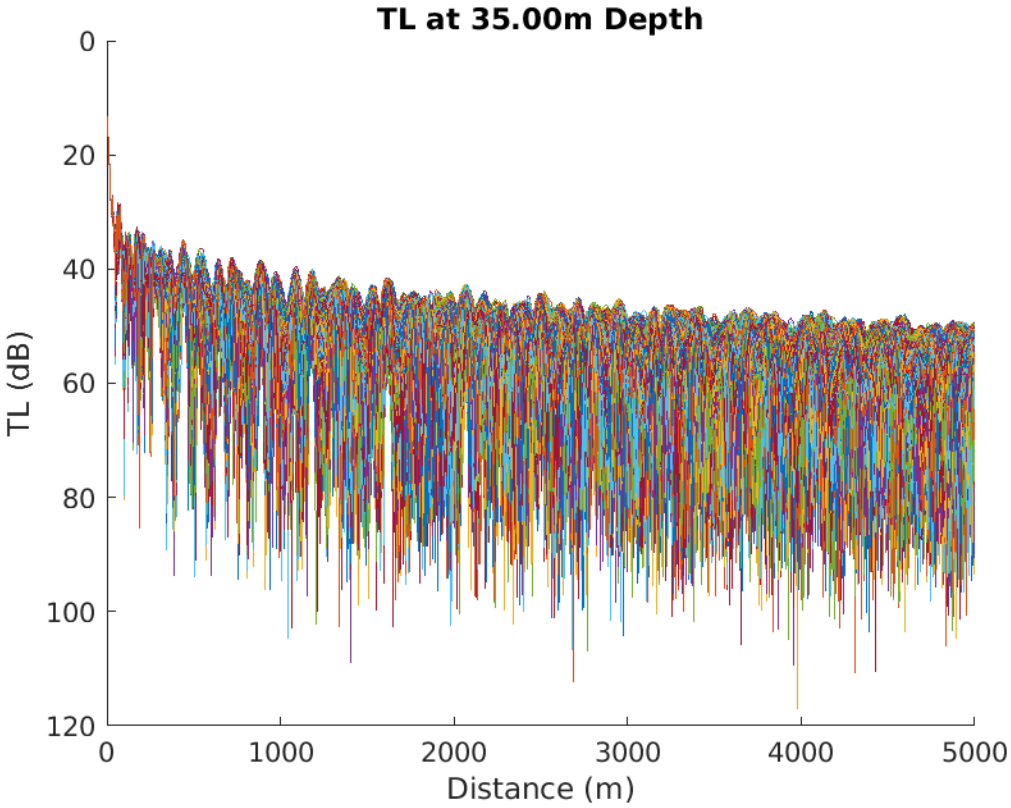

TL at 75.00m Depth
TL (dB)
20
40
60
80
100
120
140
0
1000
2000
3000
4000
5000
Distance (m)

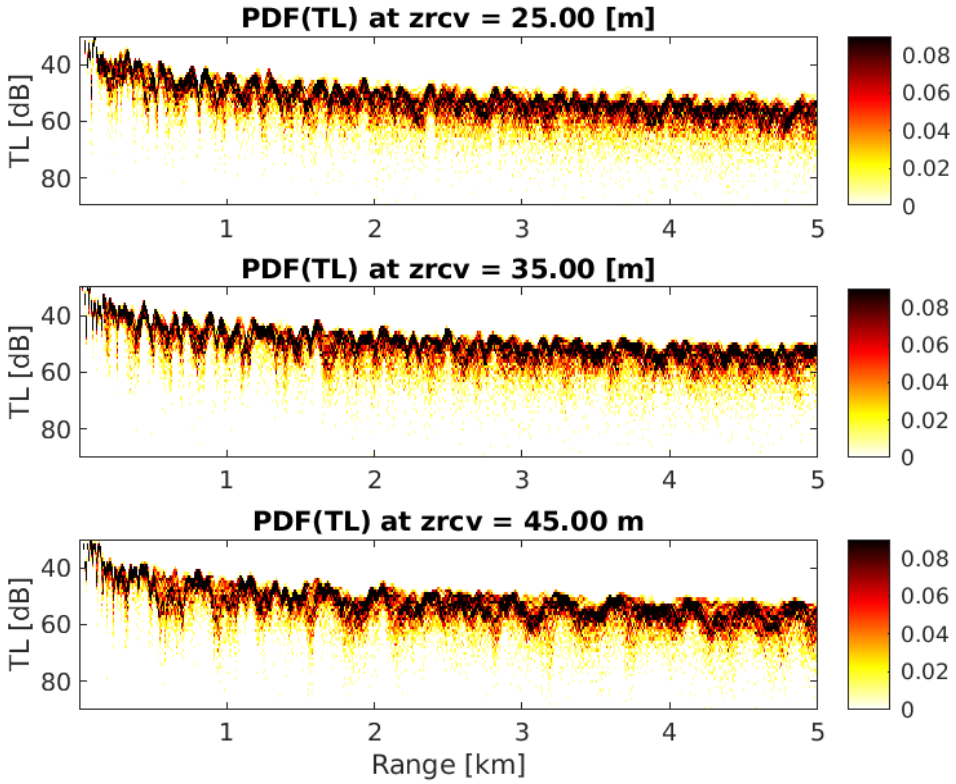

PDF(TL) at zrcv = 65.00 [m]
PDF(TL) at zrcv = 75.00 [m]
PDF(TL) at zrcv = 85.00 m
TL [dB]
40
60
80
1
2
3
4
5
Range [km]
0.08
0.06
0.04
0.02
0

Spring Shelf

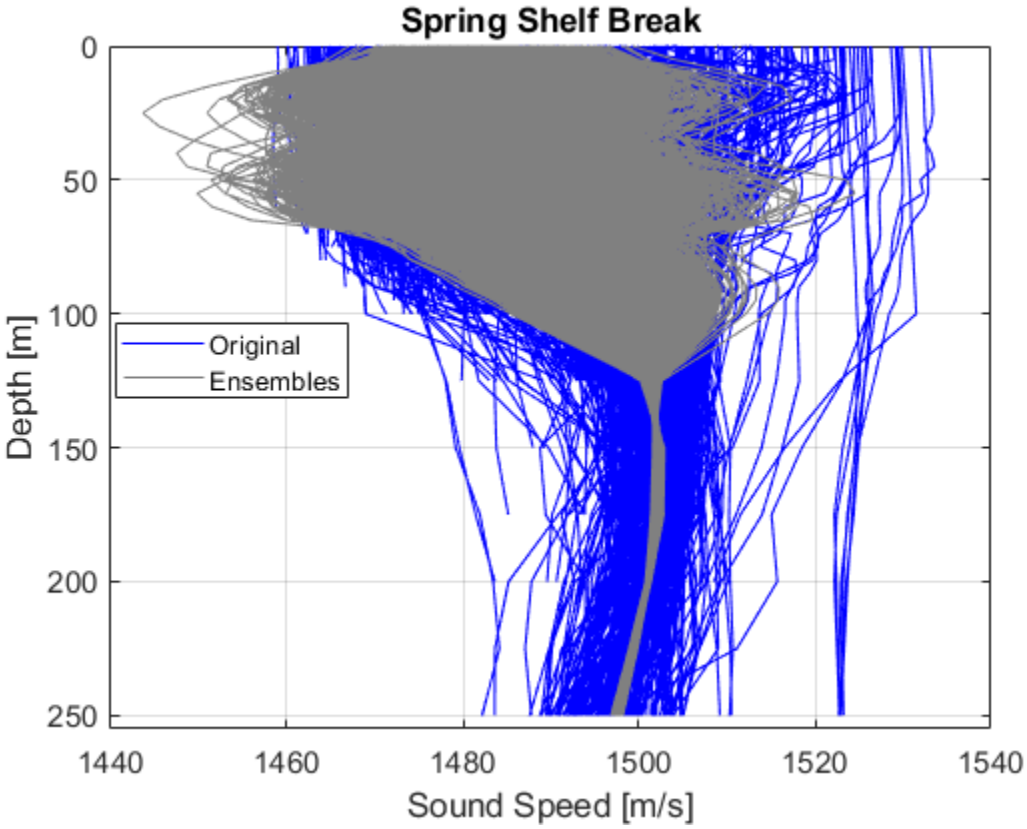


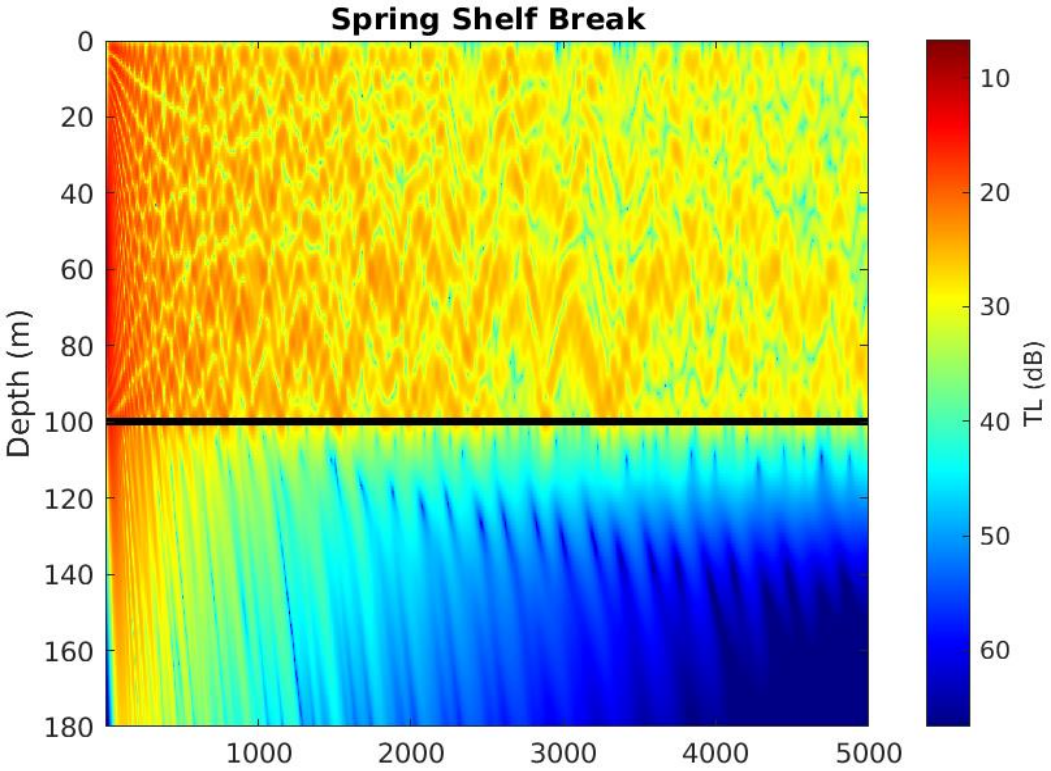

Spring Shelf
Depth (m)
0
20
40
60
80
100
120
140
160
180
1000
2000
3000
4000
5000
TL (dB)
10
20
30
40
50
60

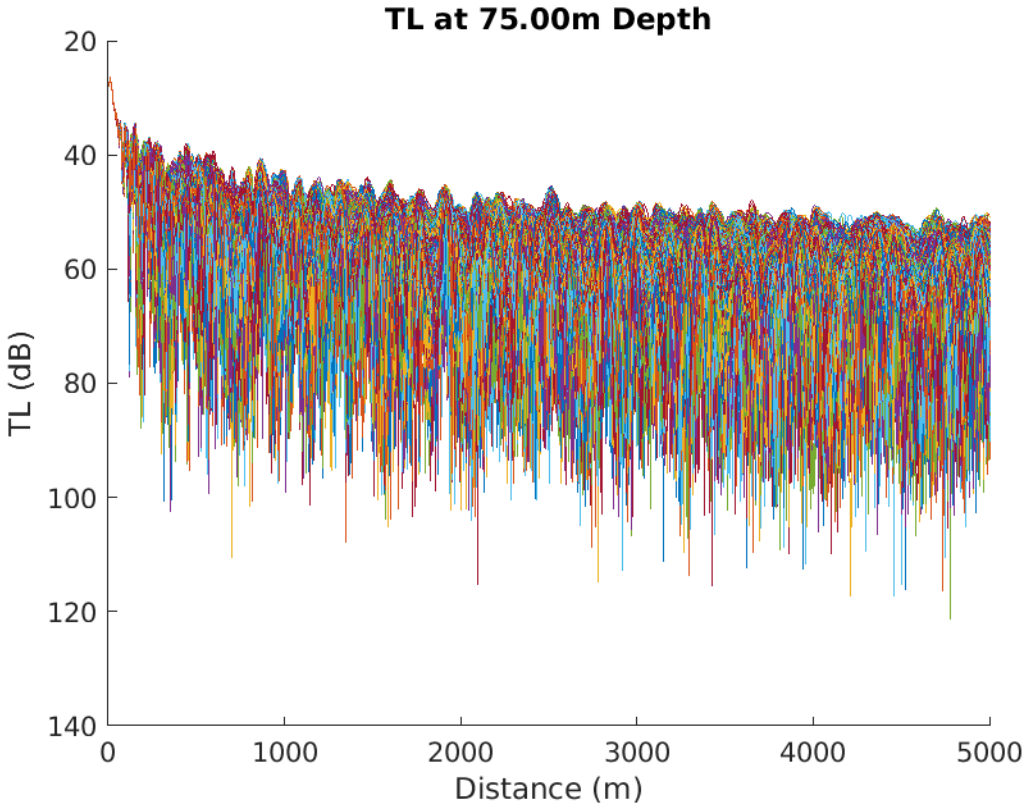

TL at 35.00m Depth
TL (dB)
0
20
40
60
80
100
120
0
1000
2000
3000
4000
5000
Distance (m)

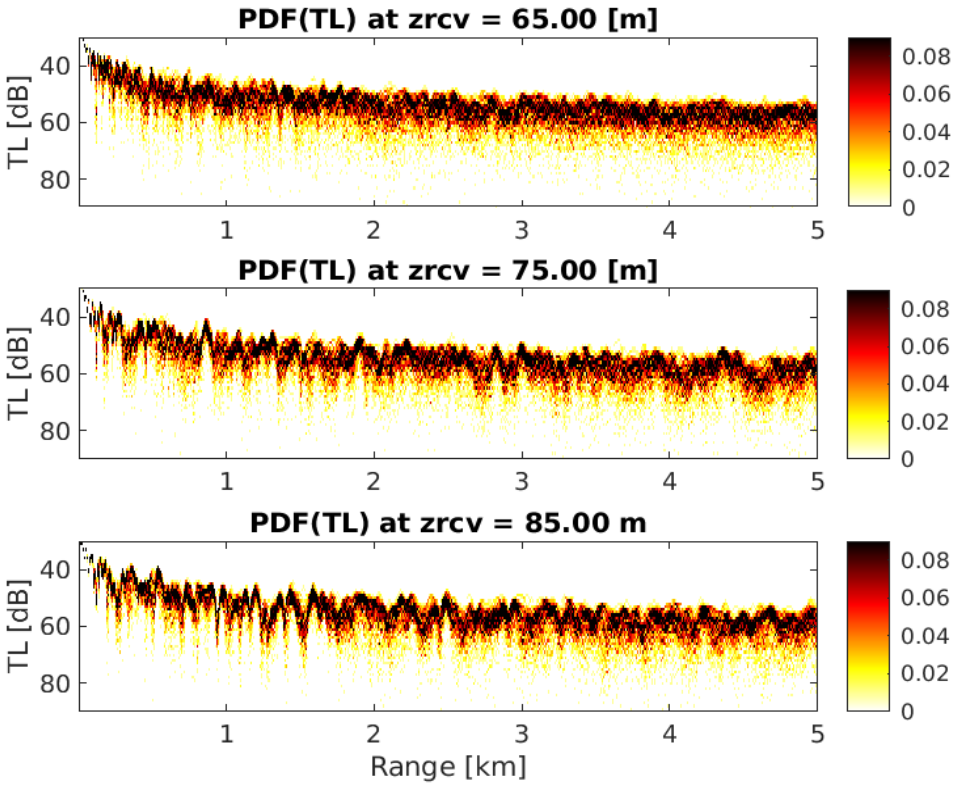

PDF(TL) at zrcv = 25.00 [m]
PDF(TL) at zrcv = 35.00 [m]
PDF(TL) at zrcv = 45.00 m
TL [dB]
40
60
80
1
2
3
4
5
Range [km]
0.08
0.06
0.04
0.02
0

Spring Shelf Break

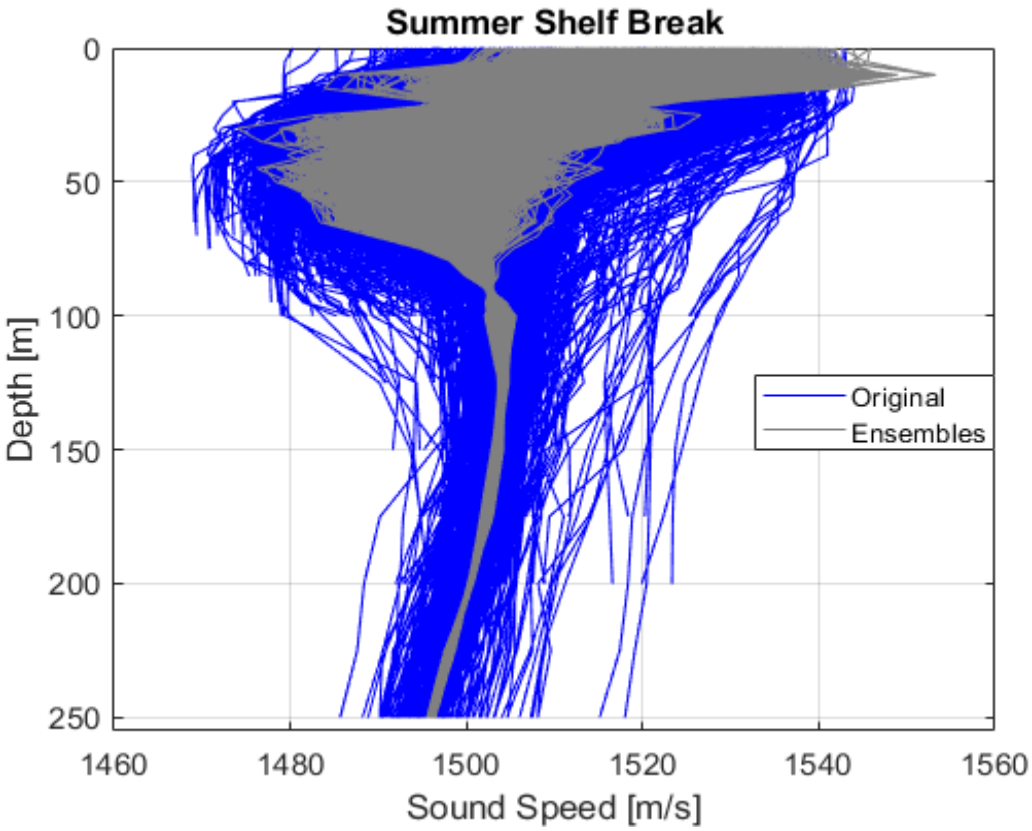


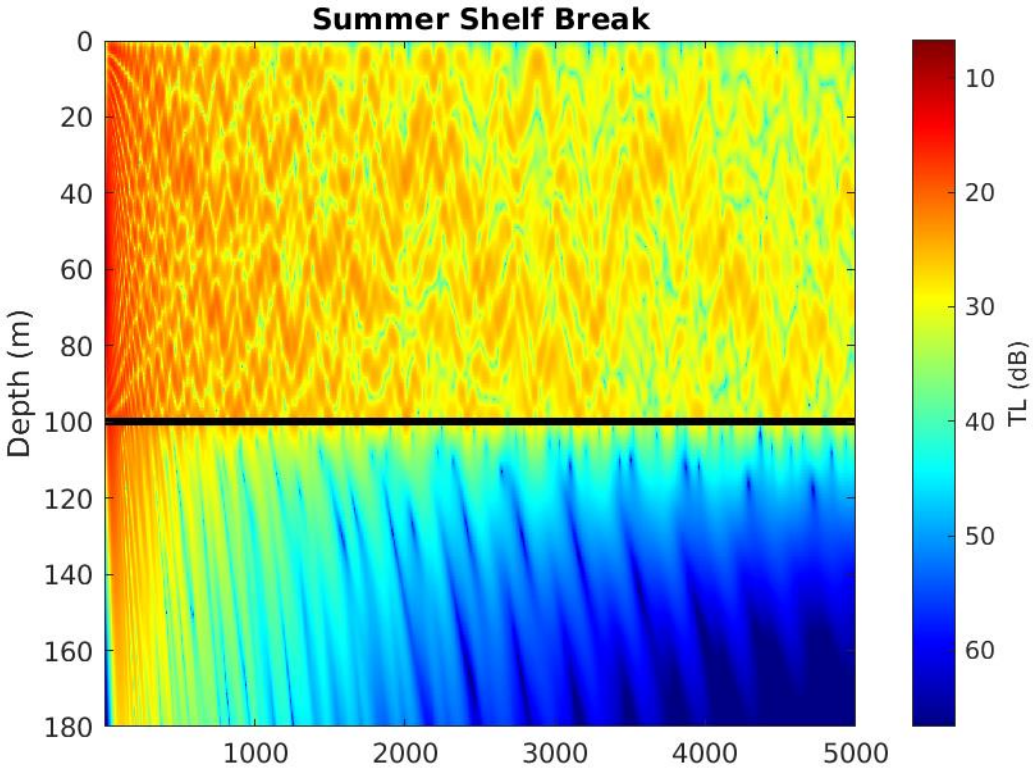

Spring Shelf Break
Depth (m)
0
20
40
60
80
100
120
140
160
180
1000
2000
3000
4000
5000
TL (dB)
10
20
30
40
50
60

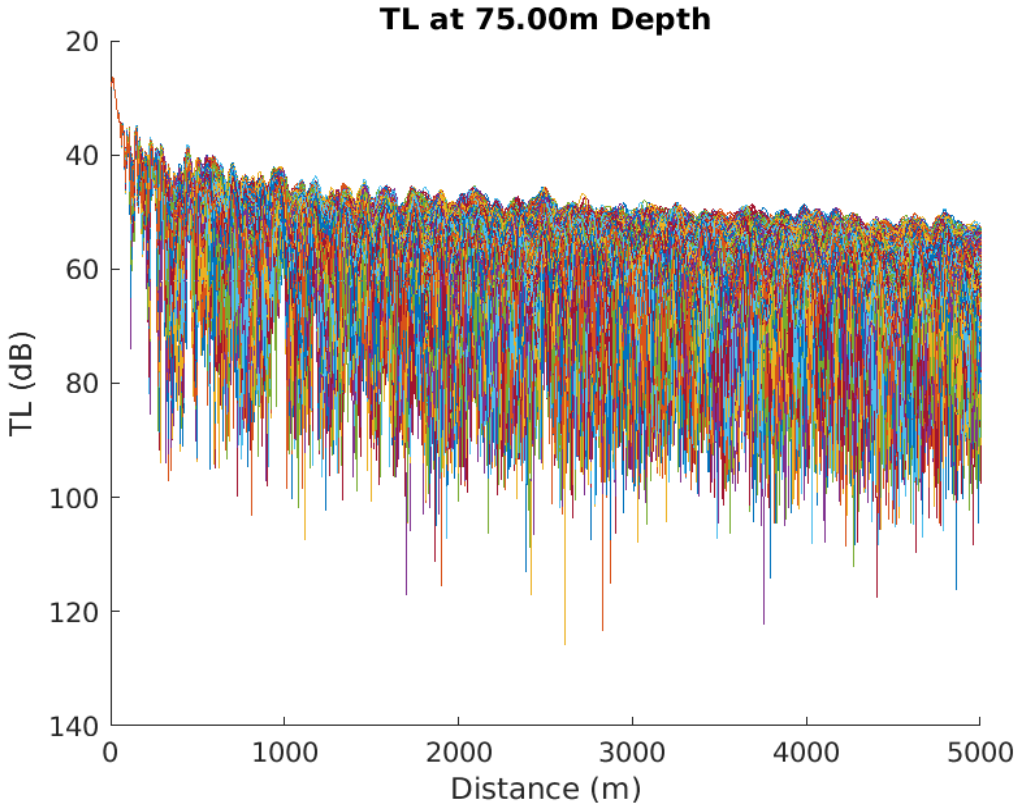

TL at 75.00m Depth
TL (dB)
20
40
60
80
100
120
140
0
1000
2000
3000
4000
5000
Distance (m)

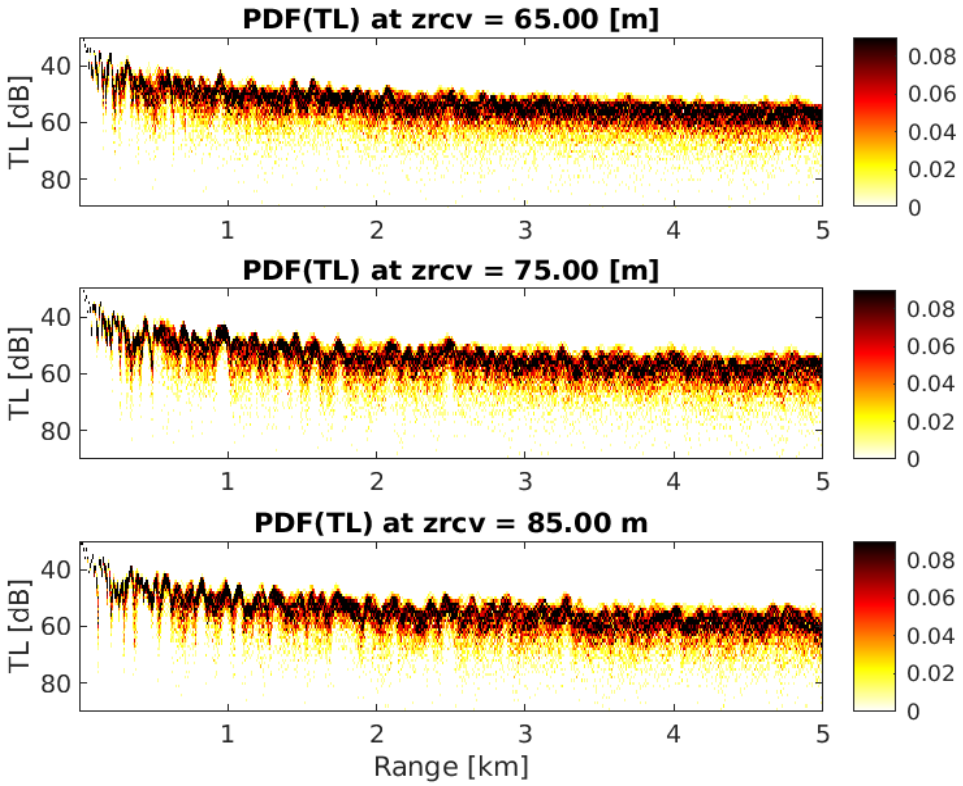

PDF(TL) at zrcv = 65.00 [m]
PDF(TL) at zrcv = 75.00 [m]
PDF(TL) at zrcv = 85.00 m
TL [dB]
40
60
80
1
2
3
4
5
Range [km]
0.08
0.06
0.04
0.02
0

Summer Shelf Break

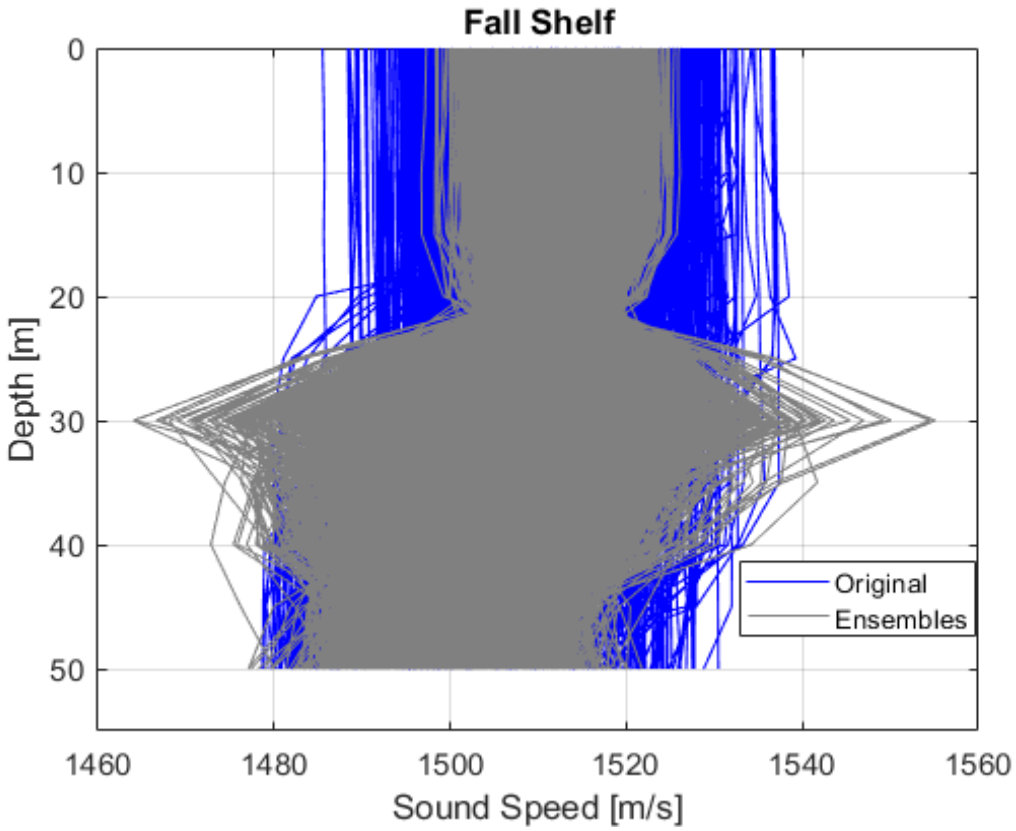


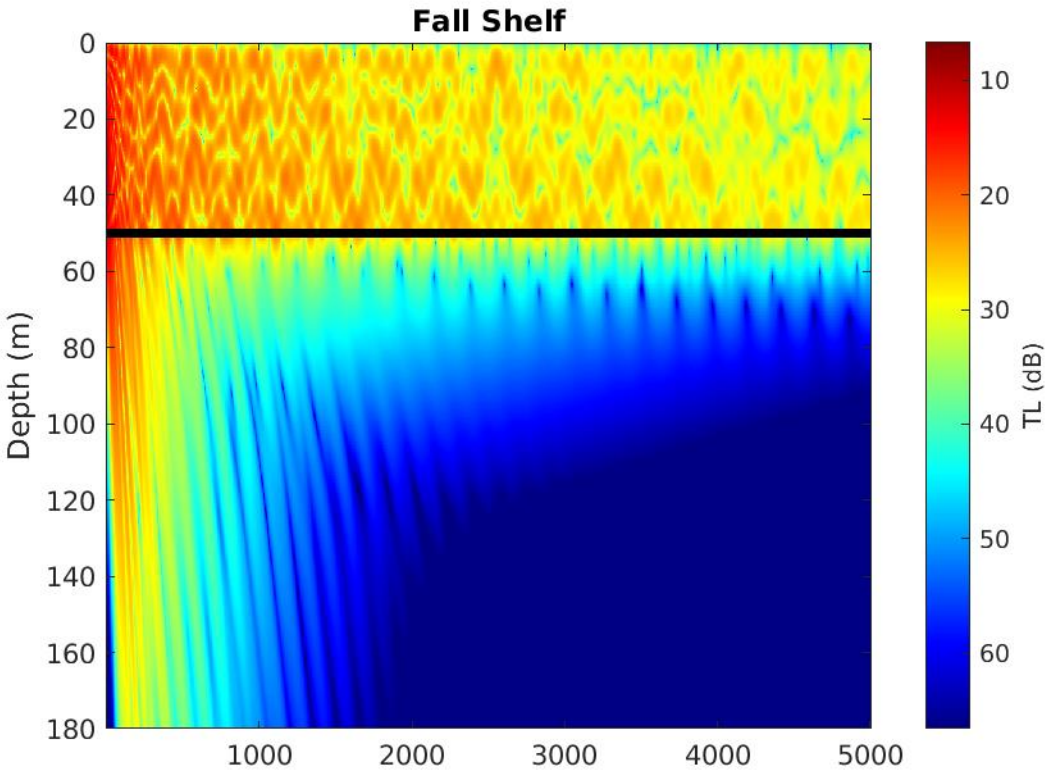

Summer Shelf Break
Depth (m)
0
20
40
60
80
100
120
140
160
180
1000
2000
3000
4000
5000
TL (dB)
10
20
30
40
50
60

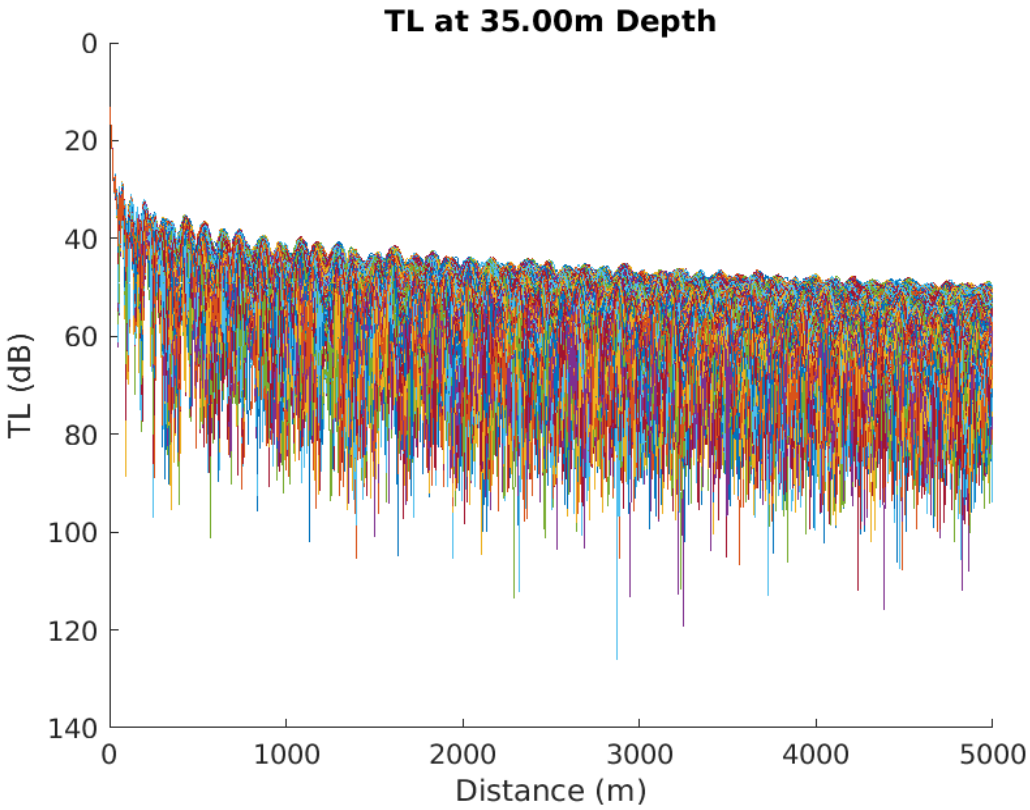

TL at 75.00m Depth
TL (dB)
20
40
60
80
100
120
140
0
1000
2000
3000
4000
5000
Distance (m)

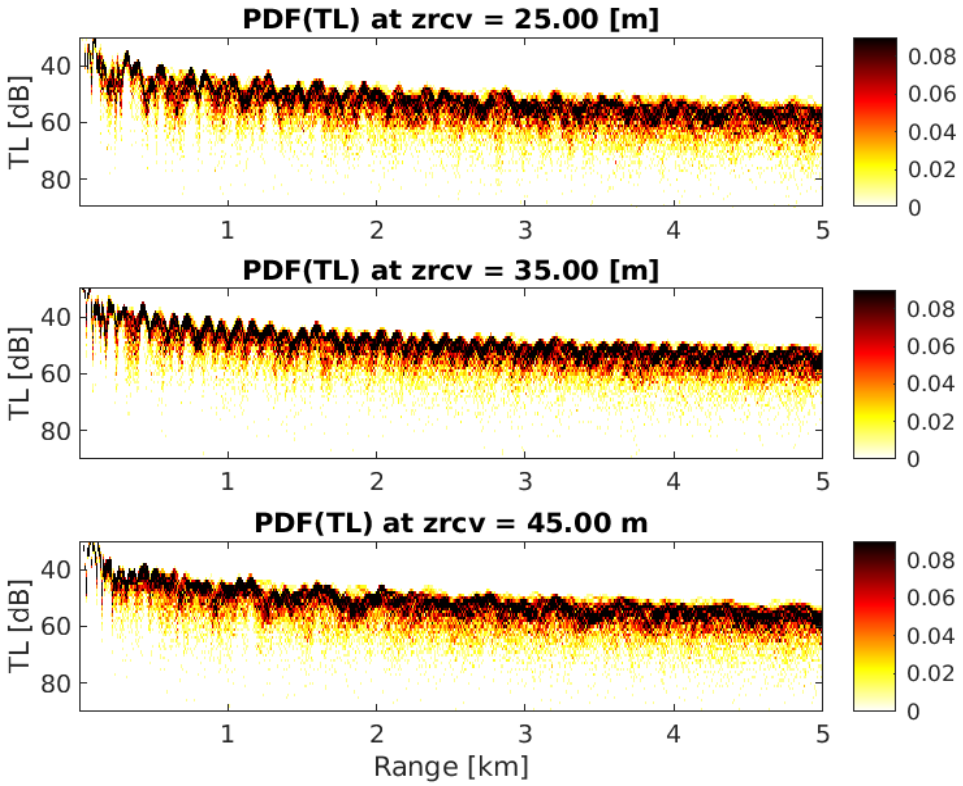

PDF(TL) at zrcv = 65.00 [m]
PDF(TL) at zrcv = 75.00 [m]
PDF(TL) at zrcv = 85.00 m
TL [dB]
40
60
80
1
2
3
4
5
Range [km]
0.08
0.06
0.04
0.02
0

Fall Shelf

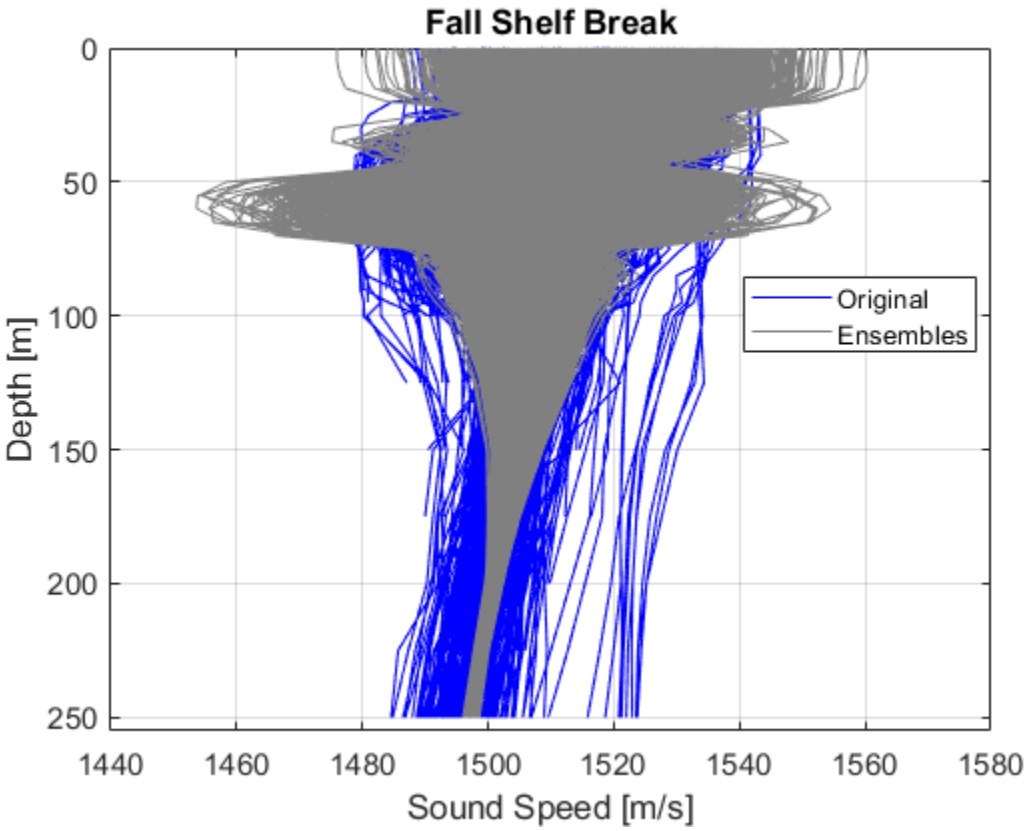

Fall Shelf
Depth [m]
0
10
20
30
40
50
1460
1480
1500
1520
1540
1560
Sound Speed [m/s]
Original
Ensembles

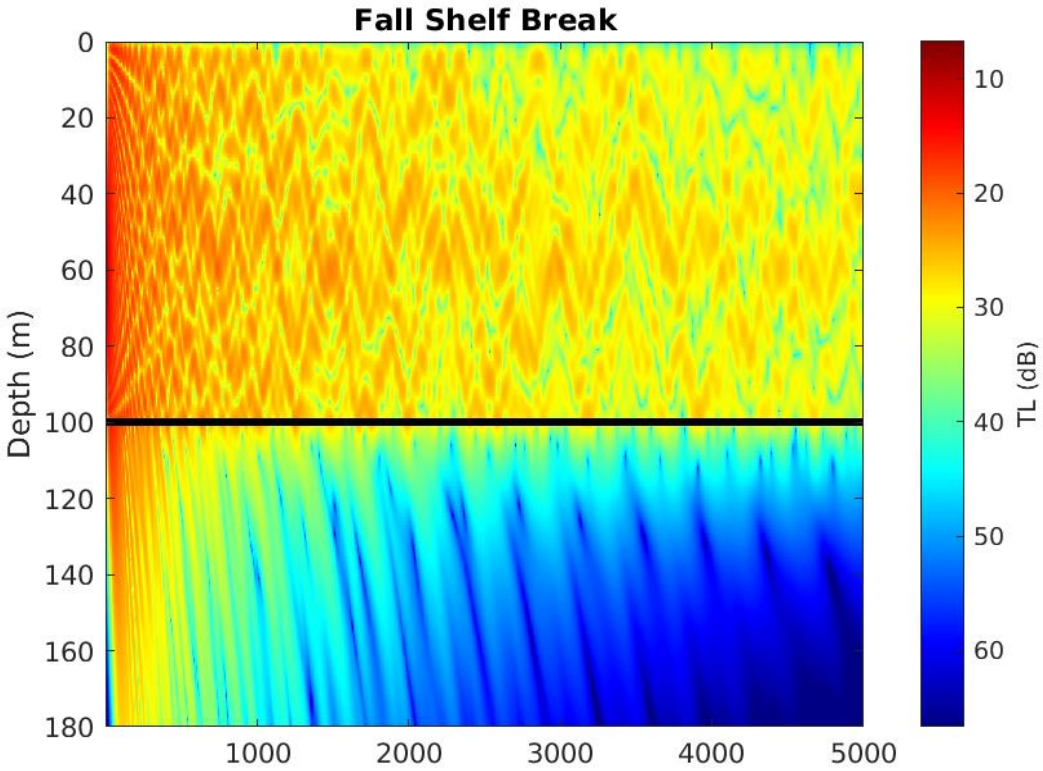

Fall Shelf
Depth (m)
0
20
40
60
80
100
120
140
160
180
1000
2000
3000
4000
5000
TL (dB)
10
20
30
40
50
60

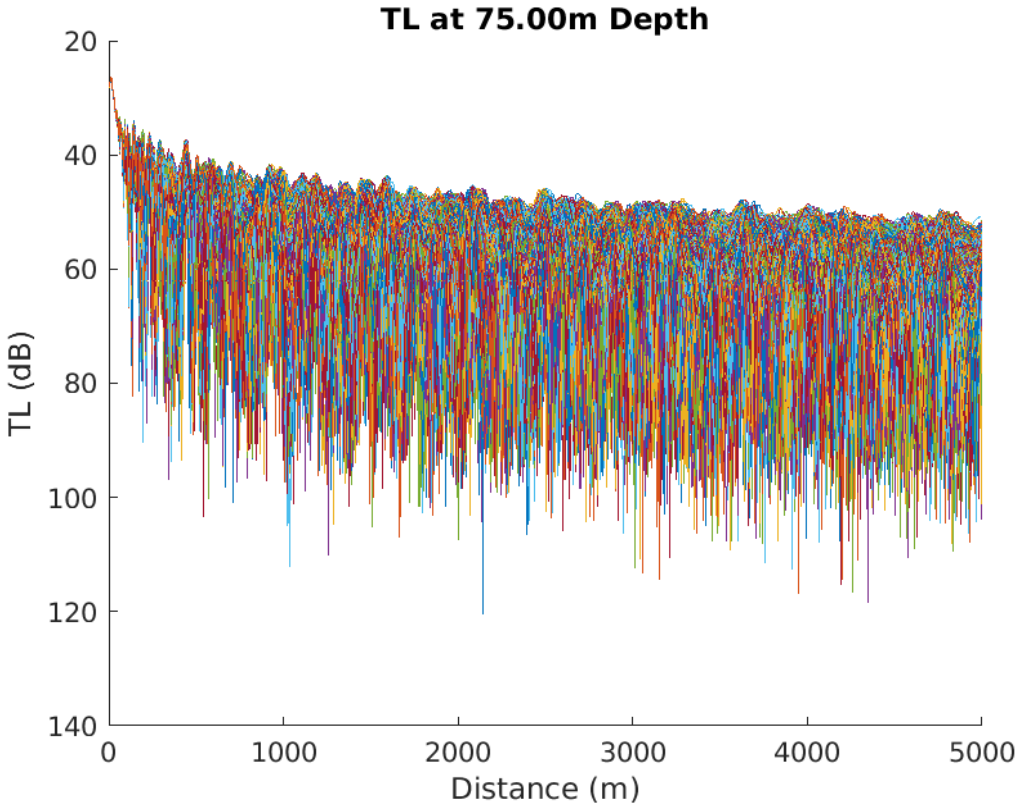

TL at 35.00m Depth
TL (dB)
0
20
40
60
80
100
120
140
0
1000
2000
3000
4000
5000
Distance (m)

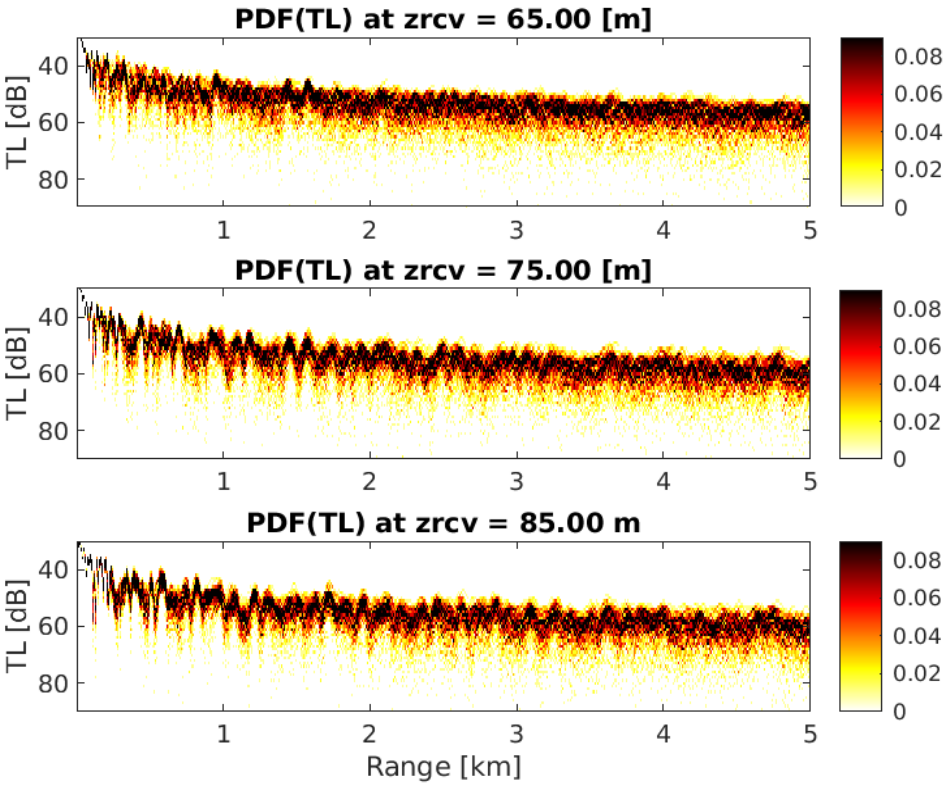

PDF(TL) at zrcv = 25.00 [m]
PDF(TL) at zrcv = 35.00 [m]
PDF(TL) at zrcv = 45.00 m
TL [dB]
40
60
80
1
2
3
4
5
Range [km]
0.08
0.06
0.04
0.02
0

Fall Shelf Break




Fall Shelf Break
Depth (m)
0
20
40
60
80
100
120
140
160
180
1000
2000
3000
4000
5000
TL (dB)
10
20
30
40
50
60

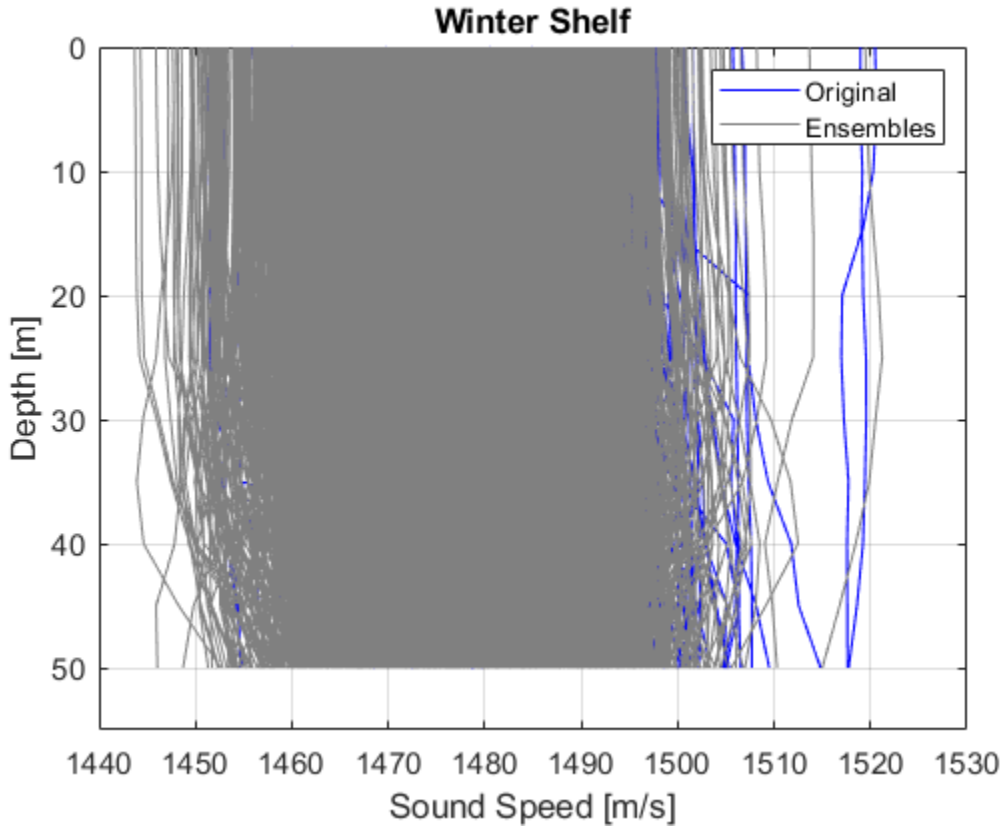

TL at 75.00m Depth
TL (dB)
20
40
60
80
100
120
140
0
1000
2000
3000
4000
5000
Distance (m)

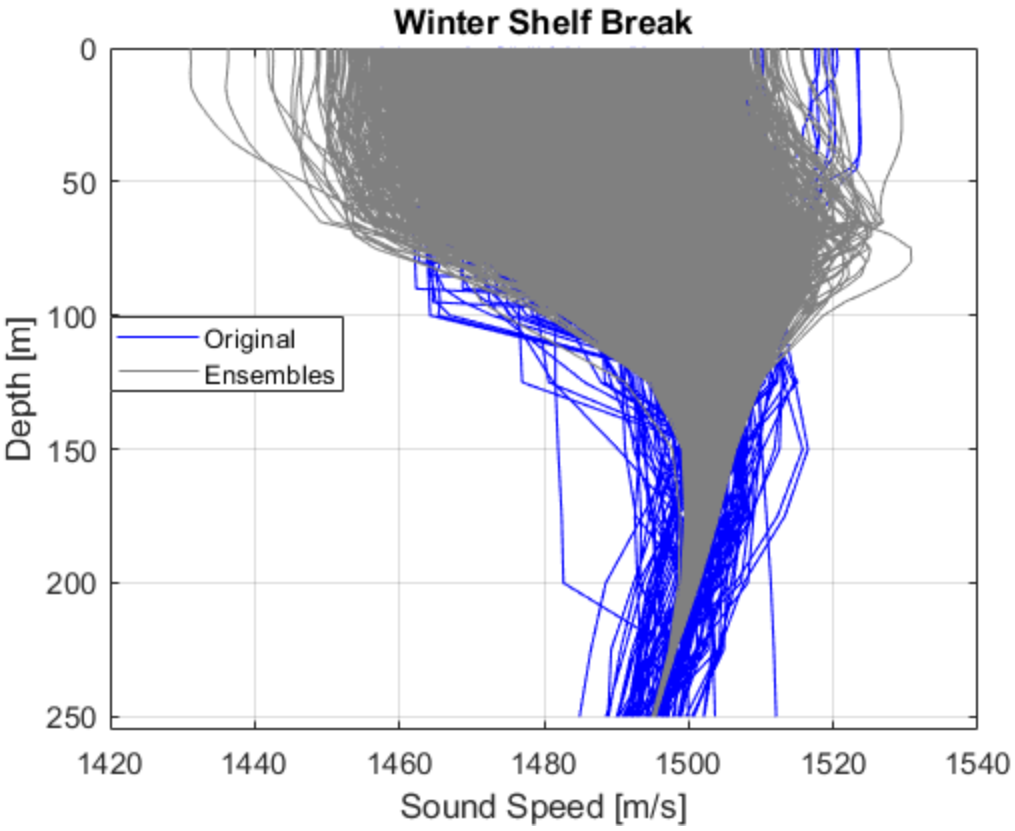

PDF(TL) at zrcv = 65.00 [m]
PDF(TL) at zrcv = 75.00 [m]
PDF(TL) at zrcv = 85.00 m
TL [dB]
40
60
80
1
2
3
4
5
Range [km]
0.08
0.06
0.04
0.02
0

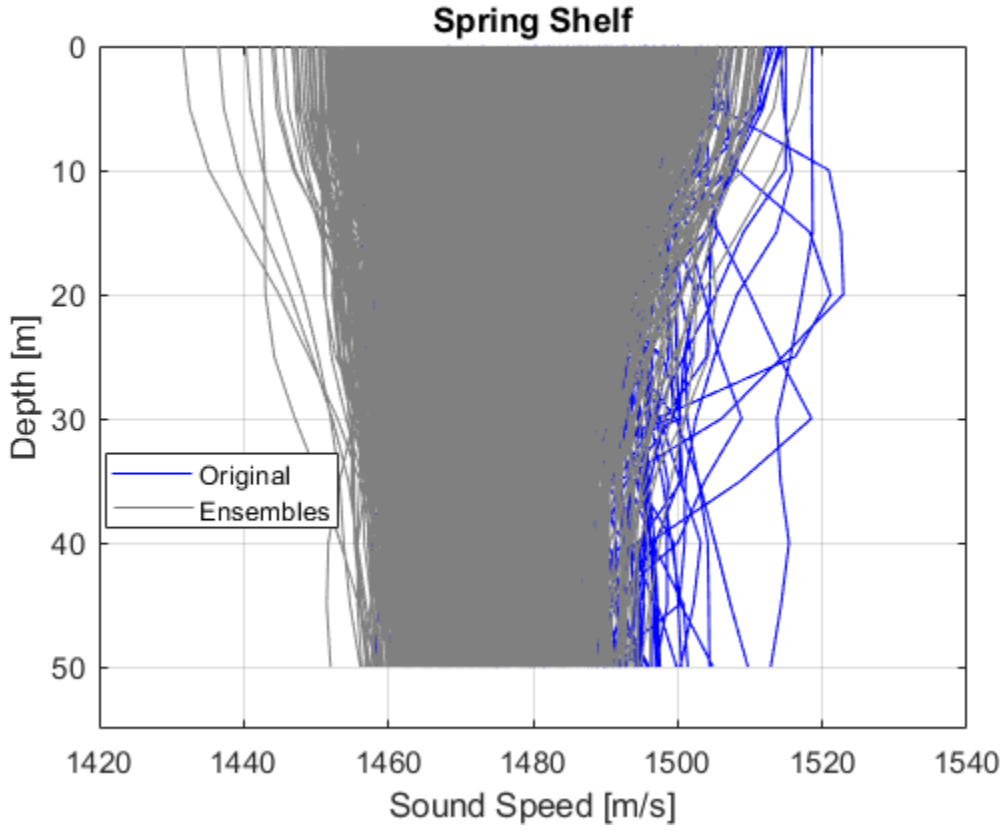
Appendix E

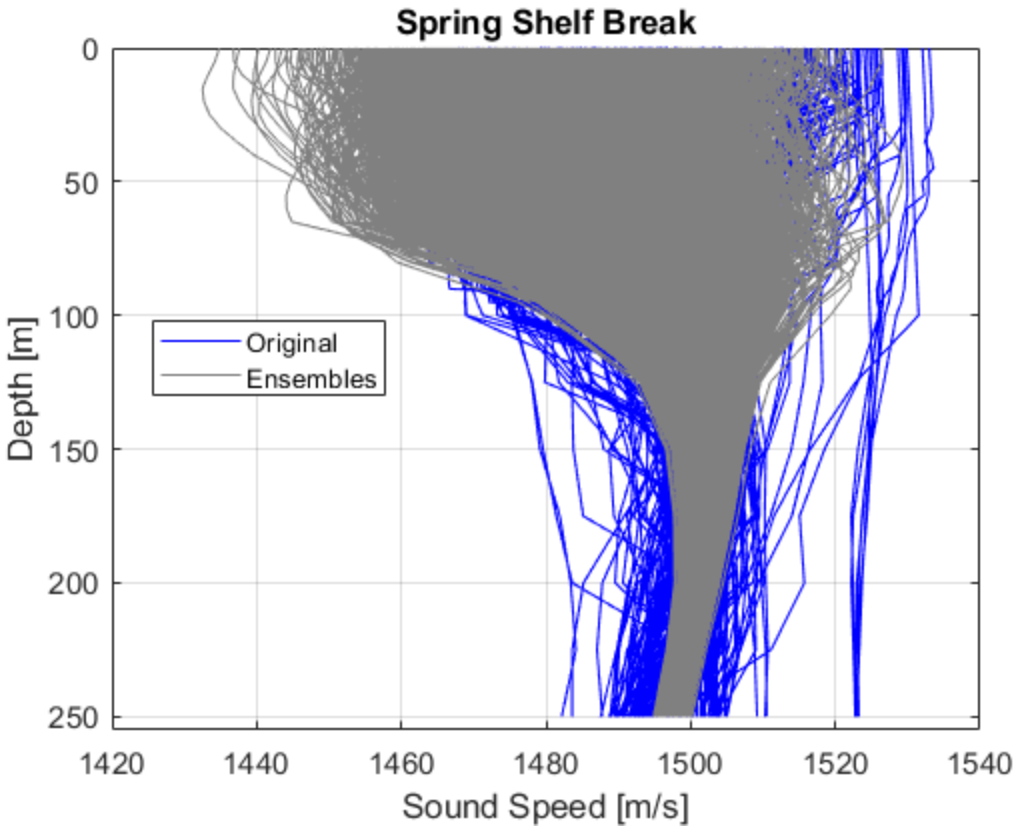
Empirical Orthogonal Function Related Figures

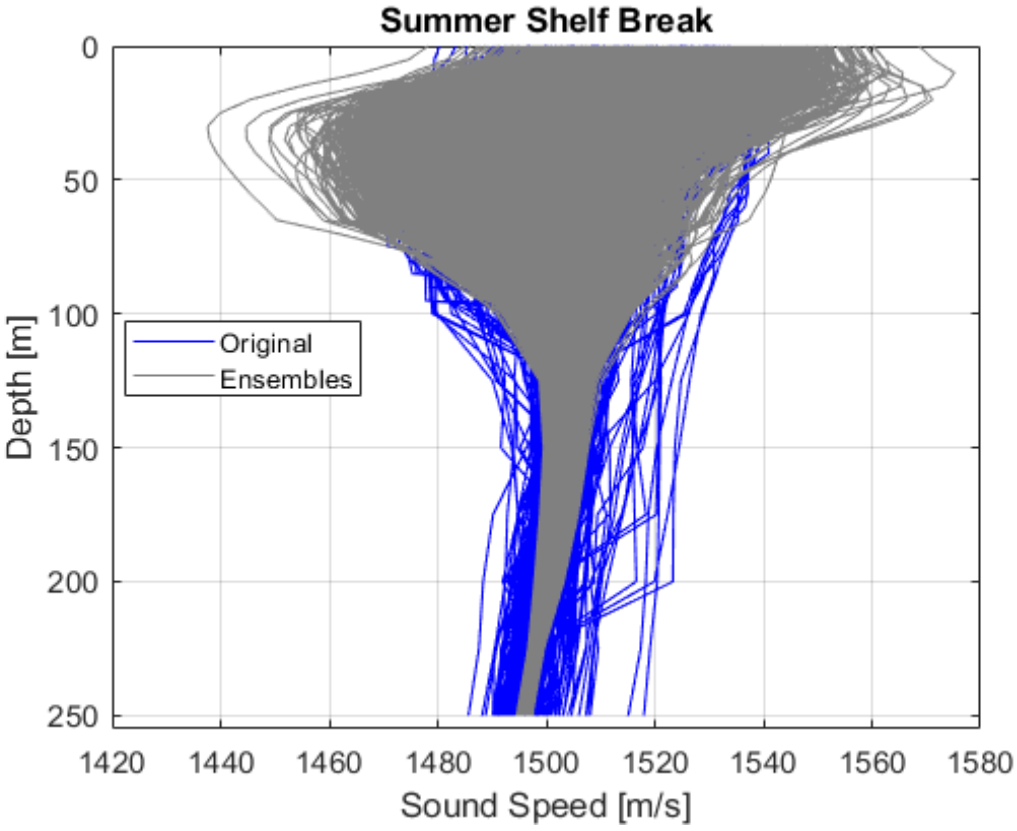
EOF ensemble figures for each season and region.

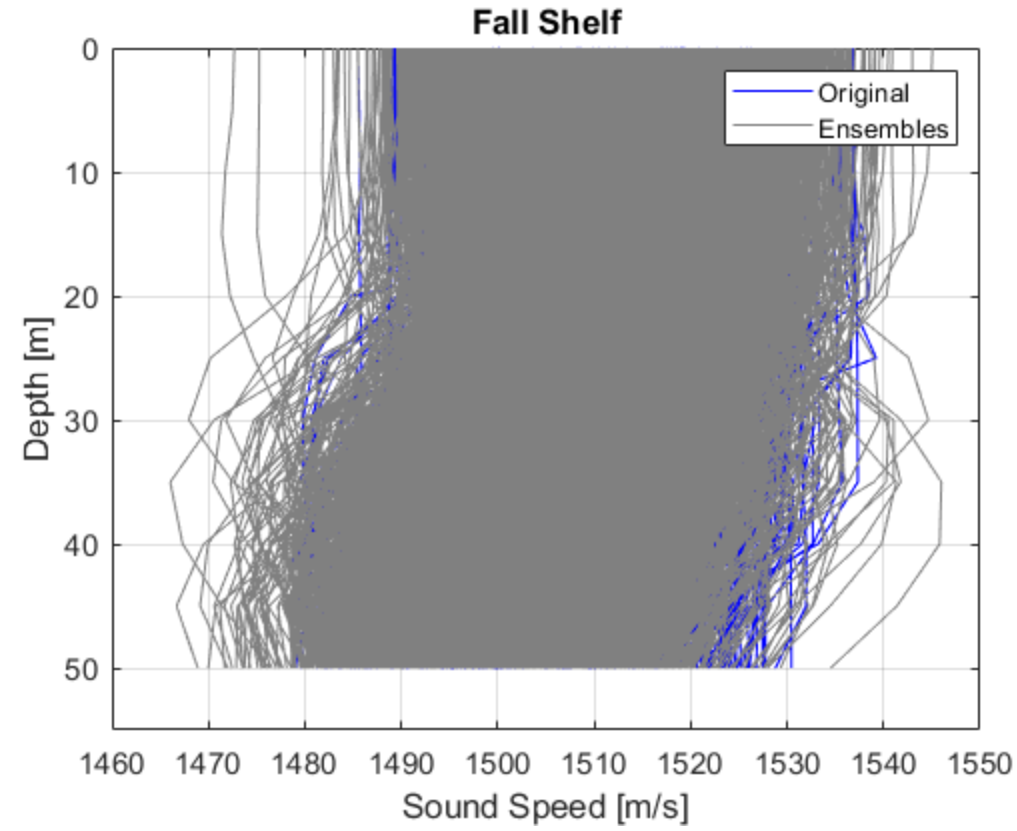


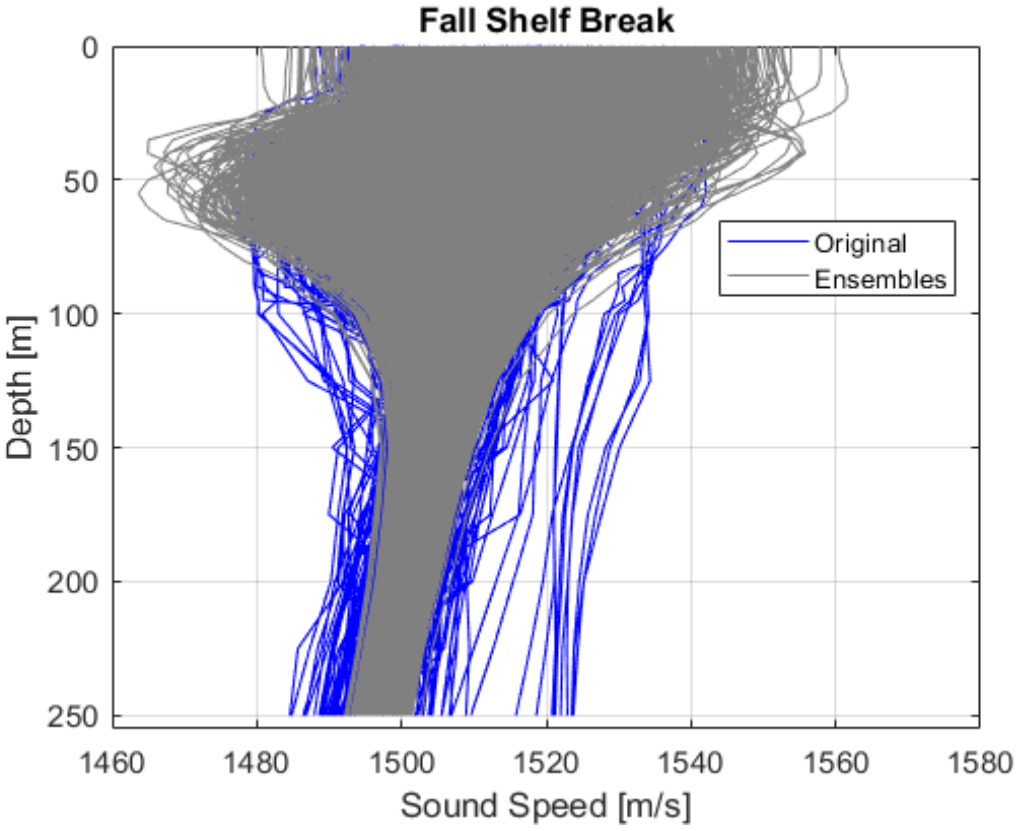

Winter Shelf Break
Depth [m]
Sound Speed [m/s]
Original
Ensembles
0
50
100
150
200
250
1420
1440
1460
1480
1500
1520
1540


Spring Shelf
Depth [m]
Sound Speed [m/s]
Original
Ensembles
0
10
20
30
40
50
1420
1440
1460
1480
1500
1520
1540


Spring Shelf Break
Depth [m]
Sound Speed [m/s]
Original
Ensembles
0
50
100
150
200
250
1420
1440
1460
1480
1500
1520
1540


Summer Shelf Break
Depth [m]
Sound Speed [m/s]
Original
Ensembles
0
50
100
150
200
250
1420
1440
1460
1480
1500
1520
1540
1560
1580


Fall Shelf
Original
Ensembles
Depth [m]
Sound Speed [m/s]
0
10
20
30
40
50
1460
1470
1480
1490
1500
1510
1520
1530
1540
1550


Fall Shelf Break
Original
Ensembles
Depth [m]
Sound Speed [m/s]
0
50
100
150
200
250
1460
1480
1500
1520
1540
1560
1580