An Experimental Benchmark for Geoacoustic Inversion Methods

Julien Bonnel [iD], *Member, IEEE*, Sean P. Pecknold, Paul C. Hines, and N. Ross Chapman, *Fellow, IEEE*

Abstract—Over the past 25 years, there has been significant research activity in development and application of methods for inverting acoustical field data to estimate parameters of geoacoustic models of the ocean bottom. Although the performance of various geoacoustic inversion methods has been benchmarked on simulated data, their performance with experimental data remains an open question. This article constitutes the first attempt of an experimental benchmark of geoacoustic inversion methods. To do so, the article focuses on data from experiments carried out at a common site during the Shallow Water 2006 (SW06) experiment. The contribution of the article is twofold. First, the article provides an overview of experimental inversion methods and results obtained with SW06 data. Second, the article proposes and uses quantitative metrics to assess the experimental performance of inversion methods. From a sonar performance point of view, the benchmark shows that no particular geoacoustic inversion method is definitely better than any other of the ones that were tested. All the inversion methods generated adequate sound-speed profiles, but only a few methods estimated attenuation and density. Also, acoustical field prediction performance drastically reduces with range for all geoacoustic models, and this performance loss dominates over intermodel variability. Overall, the benchmark covers the two main objectives of geoacoustic inversion: obtaining geophysical information about the seabed, and/or predicting acoustic propagation in a given area.

Index Terms—Benchmark, geoacoustic inversion, Shallow Water 2006 (SW06), underwater acoustics.

I. INTRODUCTION

FOR the past three decades there has been renewed focus among researchers in underwater acoustics on sound propagation in shallow water. The traditional wisdom about shallow water acoustics holds that the interaction of sound with the ocean bottom has considerable impact on the acoustical field in the water. Since sound sources in shallow water are relatively close to the seafloor interface, sound reflected from the seafloor and subbottom interfaces or refracted within the bottom and returned

to the water is expected to make a significant contribution. As a consequence, knowledge of the structure and properties of ocean bottom sediment materials is considered essential for making predictions of the sound field in the water. A central issue in research has been development of experimental and theoretical approaches for characterizing the ocean bottom.

Over the same period, geoacoustic inversion of experimental data has become a widely adopted approach. The inversion approach is inherently a remote-sensing approach, based on the observation that sound signals measured at receivers in the water contain information about the ocean medium—including the ocean bottom—through which they have traveled. In a very elementary description, geoacoustic inversion involves estimation of parameters of models that are simplified physical representations of the true ocean bottom. The models generally consist of profiles in depth below the seafloor of the sound speed, attenuation, and density of the ocean bottom sediment materials. In ocean acoustics, the interest is mainly in the first few tens of metres of sediment, but in very low frequency applications to deeper depths of a hundred metres or so are important. The published literature contains a large number of papers reporting successful applications with many different inversion methods that provided geoacoustic models, some with highly detailed structure of the ocean bottom.

The inversion methods reported in the literature are generally model-based techniques that estimate geoacoustic model parameters from acoustical field data (sound pressure) or quantities derived from the field data such as reflection coefficients, travel times of signal arrivals, horizontal wave numbers and group velocities of propagating modes to name just a few. Model-based inversions require some degree of prior knowledge of the ocean bottom at each specific site. This "ground truth" information is an essential component of the inversion that defines, and also constrains, the form of the geoacoustic model that is estimated from the data. Model-based inversions also require calculations of the field quantities, and it is implicitly assumed that the numerical models used in calculating the acoustical fields contain the correct physics. For example, depending on the experimental geometry, it may be necessary to calculate reflection coefficients for spherical waves in doing reflection loss inversions.

The inverse problem itself can be solved by linearized and nonlinear methods, and both approaches have been used to develop different inversion methods. The nonlinear methods were generally cast in terms of matched field inversions based on comparisons between measured data and replicas of the acoustical field that were calculated using candidate geoacoustic

J. Bonnel is with the Applied Ocean Physics and Engineering Department, Woods Hole Oceanographic Institution, Woods Hole, MA 02543 USA (e-mail: jbonnel@whoi.edu).
S. P. Pecknold is with the Defence Research and Development Canada, Dartmouth, NS B2Y 3Z7, Canada (e-mail: sean.pecknold@drdc-rddc.gc.ca).
P. C. Hines is with the Hines Ocean S&T 12 Carlson Court, Dartmouth, NS B2W 5X7, Canada (e-mail: phines50@gmail.com).
N. R. Chapman is with the School of Earth and Ocean Sciences, University of Victoria, Victoria, BC V8W 2Y2, Canada (e-mail: chapman@uvic.ca).

model parameters, and involved exploring multidimensional model parameter spaces. The linearized methods used perturbation theory to determine deviations from an initially assumed geoacoustic model. With the diversity of so many different inversion methods that were in use, the question arose about the relative performance of the methods in estimating accurate and useful geoacoustic models. There was also a fundamental question about the confidence limits of the estimated model parameters.

The first stages in comparing the performance of geoacoustic inversion methods were carried out in benchmarking workshops. The first workshops featured only synthetic data for specific test cases that simulated the ocean bottom in typical shallow water environments. In the first workshop, held in 1997, the ocean bottom environments were range-independent, based on known structures for the geoacoustic models [1]. The models consisted of fluid sediment layers, but included one elastic solid environment and one environment with added noise. In the second workshop, held four years later in 2001, the test cases simulated range-dependent shallow water environments [2]. All the cases were fluid sediment models. The format in each workshop was a blind test in which participants were provided only the calculated acoustical field data, and were not given the input parameters for the geoacoustic models. Calculated fields were provided on grids in depth and range so that participants could choose horizontal or vertical receiver arrays for their inversions. Participants were tasked to invert the geoacoustic model parameters of the test case environments, including sound speed, attenuation, and density of the layered structures, and determine a measure of the uncertainty of the estimates.

The results of both workshops indicated that the inversion methods were capable of generating highly accurate representations of the true geoacoustic models that were used in the test cases. However, questions remained about how well the methods performed with experimental data, whether any one particular approach was superior to others.

A benchmarking exercise involving experimental data presents significantly greater challenges compared to the previous benchmarks with simulated data. For one thing, the acoustical field quantities that were provided as simulated data in the previous workshops must instead be acquired by properly processing experimental data. Another problem is that the true geoacoustic model is not known. As a result, different inversion methods will generate different geoacoustic models from the data used in the inversion: each method is constrained by the quality and resolving power of the experimental data and the assumptions made in the inversion. Consequently, there is a basic question about what metrics are appropriate for assessing the inversion performance. One option is to compare estimated model parameters with ground truth data that are available and reliable. Another is to assess the performance of the estimated models in predicting the acoustical field measured in the experiments. To some extent, the choice of metrics depends on one's objectives in using the estimated geoacoustic model, but the challenge remains in selecting metrics that will provide a comprehensive analysis of strengths and limitations of the various methods. Other questions arise: about the type of data that can be used; the

impact of range dependence at the experimental site that may introduce three-dimensional effects in modeling sound propagation; and knowledge of the variability of the ocean bottom material and oceanographic conditions about the experimental site to name a few. However, it is a very costly venture to conduct an exercise at sea. Instead, it is worth investigating whether data from previous experiments can serve the purpose for benchmarking.

In August–September 2006, the Shallow Water'06 (SW06) experiment, a multipurpose, multi-investigator experiment sponsored by the Office of Naval Research, was carried out close to the shelf break on the New Jersey continental shelf [3]. Although SW06 was not specifically designed as a geoacoustic inversion benchmarking experiment in its overall objectives, experiments carried out in SW06 provided a wealth of high quality data that were used by many different groups for estimating parameters of geoacoustic models to characterize the ocean bottom. The published reports include geoacoustic inversions by researchers who took part in SW06 and others who subsequently used data from the experiments in other inversions, and also direct measurements of sediment geoacoustic properties using specialized techniques. Taken as a whole the results derived from SW06 experimental data provide new information about the dispersion of sound speed and attenuation in the marine sediments at the experimental site over a broad frequency band spanning nearly four decades from ~50 Hz to ~60 kHz.

The hypothesis in this article is that the experiments carried out in SW06 form the basis of an experimental benchmark for geoacoustic inversion methods, and the comparison of the various estimated models using appropriate metrics can be used to address the questions about inversion performance. In this article we develop criteria appropriate for an experimental benchmark of inversion methods, and present a comparison of the performance of the geoacoustic models generated from inversions carried out using data from SW06. As part of the benchmark analysis, the results from the estimated geoacoustic models are set in the context of previous research in marine sediment acoustics.

The remainder of the article is structured as follows. General remarks about the design of an experimental benchmark are introduced in the next section II, and the SW06 experiment is discussed in relation to the benchmark design criteria. The discussion includes a description of the SW06 experimental site and the experiments that were carried out, the ground truth information that was acquired, and general remarks about metrics for assessing the performance of the geoacoustic inversion methods. Technical details about the metrics used in the benchmark analysis are presented in Section III. Section IV presents a brief summary of the measurements made and the methods that were used to invert geoacoustic models for the experimental site in SW06. Section V presents the comparison of inversion performance, separated in two parts. First an analysis is given of midfrequency (2–6 kHz) results that are relevant to the first 1–1.5 m beneath the seafloor, and then the comparison of models estimated from inversions of low frequency data ($<$ 1 kHz) that sampled deeper into the bottom is presented and discussed in terms of inversion performance and critical model parameters.

Section VI introduces and discusses the quantitative benchmark assessment of the estimated geoacoustic models using the metrics. Section VII concludes this article.

II. EXPERIMENTAL BENCHMARKING

A. General Requirements for an Experimental Benchmark

Some basic requirements can be stated first for an experimental benchmark for geoacoustic inversion methods.

1) A common site needs to be established for all experiments that acquire data for use in the inversions.
2) Extensive ground truth needs to be carried out to provide basic prior knowledge of bathymetry and the structure and material properties of the ocean bottom sediment about the experimental site.
3) Continuous sampling of ocean environmental parameters should be carried out during the experimental period. This includes: water sound speed and/or temperature profiles; sea state and wave height; wind speed and direction; and current speed and direction.
4) Continuous monitoring of external sound sources, including shipping and fishing activity, should be maintained.

All these conditions are designed to ensure that the data used in model-based inversions are acquired over the same ocean bottom environment, and oceanographic effects and interference from external sound sources such as passing ships can be taken into account.

It is inherently assumed that the experiments provide:

1) sufficiently high quality data that contain relevant information about the ocean bottom for the inversions;
2) nonacoustic information such as: accurate global positioning system data for positions of sources and receivers; water depths at experimental sites and bathymetry along experimental tracks of sound sources; and sound-speed profiles in the water to name only a few.

Given that these conditions can be met, an experimental benchmark addresses the following questions.

1) Are the inversion methods capable of generating geoacoustic models that are realistic and useful representations of the real ocean bottom?
2) What are the limitations of each inversion method? Which geoacoustic model parameters are well estimated, and which ones are estimated with limited success?
3) Which model parameters have the greatest impact on the acoustical field in the water?

Thus, apart from the comparison of inversion performance, analysis of results of an experimental benchmark also provides other information about the capabilities of present-day geoacoustic inversion methods.

B. SW06 Experimental Site

The experiments in SW06 were carried out from two central sites separated by about 8 km near the shelf break on the New Jersey continental shelf. One site was in a region where the pervasive outer shelf wedge sediment material was within ~1 m of the seafloor, and the other was on a ridge where a
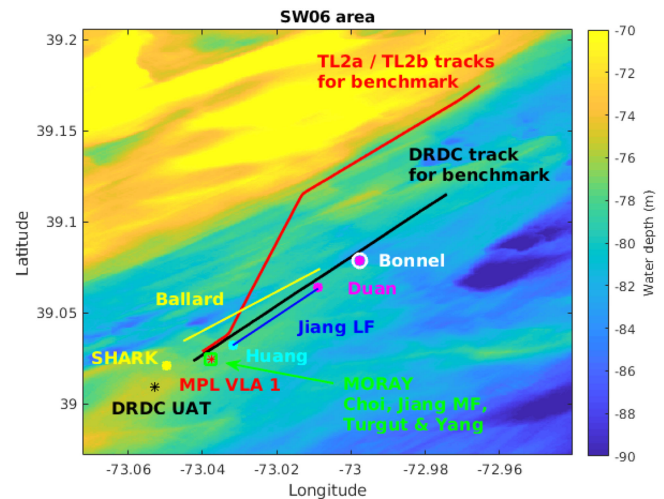

Fig. 1. Chart of experimental site on New Jersey continental shelf. The central site is marked by the red star at MPL VLA 1 and the source trajectory for the source tracks (TL2a/TL2b) used in this article is depicted by the red line. The DRDC vertical array (DRDC UAT) is marked by a black star, and the corresponding DRDC source trajectory is depicted by the black line. The source/receiver configuration of all the inversion methods considered in the benchmark are also reported on the map. All the midfrequency inversions (Choi *et al.* [10], Jiang *et al.* MF [18] and [19], Turgut [23], Yang *et al.* [11]) were performed within 400 m of the MORAY array, depicted by a green square. Most of the low-frequency inversions considered signals recorded on the MPL VLA 1 array (Bonnel *et al.* [15] and [16], Duan *et al.* [17], Huang *et al.* [14], Jiang and Chapman LF [12] and [13]), except Ballard *et al.* [20] and [21] who used the SHARK array shown by the yellow circle. Ballard *et al.* used data from a towed source, whose trajectory is depicted by the yellow line. Jiang and Chapman LF used data from fixed source positions along the blue line. Bonnel *et al.*, Duan *et al.*, and Huang *et al.* used fixed sources, whose positions are depicted by white (Bonnel *et al.*), magenta (Duan *et al.*), and cyan (Huang *et al.*) points.

thin layer of sand (~6 m) covered the underlying outer shelf wedge sediment. Two L-shaped vertical/horizontal arrays of the Applied Research Laboratory at University of Texas, Austin, TX, USA, were deployed at locations about 20 km apart on the sand ridge, and were used to collect both continuous wave (CW) and broadband impulsive data during a three week period. The results of inversions of data acquired over the sand ridge have been reported extensively by Knobles and colleagues [4]–[6], and are summarized very well in a recent paper by Wan *et al.* [7].

In this article, we focus on the former site where measurements of the marine sediment properties were made using *in situ* probes, and geoacoustic models were inverted by several different methods using data recorded at vertical line hydrophone arrays (VLA) and transmitted from sound towed sources along tracks or deployed at fixed stations from the arrays. The experimental benchmark is based on inversions done with data acquired in experiments at the outer shelf wedge site.

The location of the outer shelf wedge site is shown in Fig. 1. Two different vertical arrays were deployed in turn about three weeks apart in August 2006 at the central site, 39°1.4'N 73° 2.7'W. These included the 4-element moored vertical receiving array (MORAY) from the University of Washington Applied Physics Laboratory, Seattle, WA, USA, which was deployed in the first three-week period, and then followed by a 16-element

vertical array from the Marine Physical Laboratory (MPL) of Scripps Institute of Oceanography which was deployed afterward. The water depth at the site was 79 m. Experiments with broadband *in situ* probes measured sound speed and attenuation in the top ∼1.5 m of sediment about the site [8] and [9]. The frequency band of these data spanned 10–60 kHz. Other experiments with lower frequency sound sources including towed projectors on surveyed tracks from the arrays and light bulb implosions provided CW and broadband data over the frequency band from 50 to 3500 Hz that were used in several different inversion methods to estimate geoacoustic profiles to greater depths below the seafloor. The methods included bottom loss inversion [10] and [11], matched field inversion [12]–[14], modal group velocity inversion [15]–[17], and travel time inversion [18] and [19]. Another 32-element vertical (SHARK) array moored about 1 km shoreward of the site was used to record low frequency (50–175 Hz) CW data for modal wave number inversion [20]–[22]. The water depth over the tracks of these experiments increased gradually by only ∼4 m over the ∼8 km towing track. Consequently, most inversions assumed a range independent environment in modeling sound propagation. A total of ten different inversion techniques were applied to the data acquired in the experiments.

C. Ground Truth Data

The real ocean bottom environment is a highly complex system of different sediment materials that is in general variable with depth scales as small as centimeters and range scales of hundreds or even tens of meters. The water depth and depths of interfaces that define different types of sediment material may also vary with range over the same scale. The first and foremost challenge in establishing an experimental benchmark is to characterize the structure and physical properties of the sediment materials about the experimental site as thoroughly as possible. Ground truth information of shallow water sites is usually acquired in two ways: 1) remote sensing sonar and seismic surveys, and 2) physical sampling of the bottom materials by sediment cores and grab samples. Specialized instruments such as cone penetrometers that provide measurements of bearing strength based on the deceleration of the instrument in the sediment are also available. Depending on the water depth, other methods by diver intervention are sometimes used.

A few remarks are necessary to explain what is meant here by ground truth and the extent to which it should be used. The resolution of the ground truth information is a significant factor. For instance, high resolution seismic and sonar surveys can resolve sediment structure to within a meter, and sediment core analysis can resolve much smaller scales. However, geoacoustic inversions generally use acoustical data with much lower resolution, thus constraining the depth scale of the layers that can be resolved in the structure of the estimated geoacoustic model. Also, physical samples such as sediment grabs and cores are point measurements at specific sites, and the information from them is limited to the surficial sediment material by the relatively shallow penetration depths. A fundamental issue in geoacoustic inversion is that the structure

of the estimated geoacoustic model must be consistent with the information about the ocean bottom contained in the data that are used. Otherwise, the estimated model will generally be over parameterized. Although the ground truth information is important knowledge, it should be used only as a guideline for the structure and properties of the estimated model.

There is extensive ground truth information about the region around the experimental site from previous marine geological and geophysical surveys [24], ocean drilling projects [25], and other underwater acoustical experiments [26] and [27]. The following segments of this section describe the results of physical sampling and sonar surveys that characterize the immediate area of the outer shelf wedge experimental site.

D. Chirp Sonar Survey

A high-density grid chirp sonar survey (1–4 kHz) was carried out over the entire region before the experiment in 2000–2001 [24], and additional survey was done in the early stages of SW06. The surveys provided high-resolution bathymetry and subbottom structure over an extensive region of the outer shelf. The most significant features are sub-bottom paleochannels, and the "R"-reflector which is ubiquitous over the experimental area. The sediment above the R-reflector is diffusely reflective clay and sandy-clay. Within this layer is a weakly reflecting boundary about 6–8 m above the R-reflector, caused by a previous erosional episode that separates the diffusely reflective material from a lower layered unit [28] and [29]. The paleochannels consist of sand material, and the profile of sediment material within the channel is usually different from the outer shelf wedge sediment in the surrounding region. The depth of the R-reflector varies over the region from about 10–25 m below the seafloor, and it defines a rough boundary between the clay and sandy-clay outer shelf wedge sediments and older, more consolidated shelly sand sediments with higher sound speed.

E. Physical Samples

An extensive *in situ* probe survey by Goff *et al.* [24] in the general region of the outer shelf wedge site provided sound speed and attenuation measurements at 65 kHz that are characteristic of the first ∼15 cm in the sediment. Grab samples were collected at each probe site to determine type of material and size distribution of sediment grains. Particle size ranged between 1–$2\,\phi$ ($\sim 250 - 500\,\mu$m). Values for sound speed and attenuation were highly variable from site to site in the survey, with values between 1663–1733 m/s for sound speed and ∼38 dB/m (0.99 dB/λ) for attenuation at locations near the outer shelf wedge site.

Sediment cores that provided sound speed, density, and porosity data over the upper portion of outer shelf wedge sediment were also obtained near the experimental site as part of SW06. The cores consisted of primarily stiff clay material with interspersed sandy clay lenses. Analysis of the core material indicated sound speeds of 1600–1650 m/s, relatively high density of 1900–2000 kg/m³, and low porosity between 0.4–0.45 for the upper sediment material within ∼1.5 m of the seafloor. The results suggest that the vicinity of the site is characterized by outer shelf wedge clay and sandy clay sediment that outcrops at

the seafloor at some locations. In some parts of the region, there are thin covering layers of higher sound speed sandy material, ~1730 m/s. These deposits are within the first 10–20 cm of sediments and are not likely to be resolved as distinct layers in inversions based on low frequency sound sources.

F. Experimental Benchmark Metrics

Since the real environment is not exactly known, the ocean bottom is generally assumed to be a system of layers, each of which may be range dependent. However, inversions based on the data from different experiments at the site will estimate different layered approximations to the true environment, depending on the assumptions that are made in the inversion. Design of the metrics to compare the different estimated models presents an additional challenge for establishing the benchmark. To some degree, the metric depends on the objective in the use of the geoacoustic model. For instance, if one wants to learn geophysical information about the seabed, it is important for the estimated models to be as informative as possible. On the other hand, if the overall objective is assessment of sonar performance, one may not care about the physical meaning of the geoacoustic model, but an important metric could be comparison of transmission losses (TL) predicted by the various different models against experimental measurements of the loss.

In this article, we use three different metrics for the comparisons. Technical details about these metrics will be given in the next section III. They are first briefly described here.

The first one is a comparison of estimated geoacoustic model parameters with the ground truth data. This metric is used for assessing the estimated values characteristic of surficial seafloor sediment, within ~1–3 m of the seafloor.

The second metric is a comparison of estimated geoacoustic models from a modal propagation point of view. To do so, the predicted wave numbers are computed, and cross-compared between models. A euclidian distance is defined on the wave number vector, which defines a distance between the estimated geoacoustic models.

A third metric is defined to assess the capability of the geoacoustic models in predicting the acoustical field (i.e., sonar performance). To do so, experimental TL versus range are compared to simulated TL, based on predictions obtained with the estimated geoacoustic models.

G. Experimental Data Used for Evaluating Acoustical Field Prediction

The third metric requires experimental measurements of TL that were made about the site. Three different tracks are considered for the benchmark, they will be referred as TL2a, TL2b, and Defence Research and Development Canada (DRDC). On these three tracks, a towed source emitted continuous tone signals. The source frequencies and towing times are:

1) TL2a (August 25, 2008 from 19 h 20 min to 21 h 15 min): 53, 103, 203, and 253 Hz;
2) TL2b (August 26, 2008 from 06 h 55 min to 08 h 40 min): 302, 402, 502, 702, and 951 Hz;

3) DRDC (August 01, 2008 from 13 h 00 min to 13 h 10 min): 1200 Hz (actually, the source was not a continuous tonal but a series of 0.1 s tonals emitted every 20 s).

For each track, the source trajectories are plotted in Fig. 1. TL2a/TL2b are plotted in red, they are so close to each other that they cannot be distinguished, although TL2b started about 9 h after TL2a ended. The corresponding acoustical field was recorded on the MPL VLA 1, plotted as a red star in Fig. 1. The source trajectory for the DRDC track is plotted as a black line in the figure. The corresponding acoustical field was recorded on an underwater recording station (DRDC UAT), as illustrated by the black star on the same figure. The three tracks (TL2a, TL2b, DRDC) were conducted with increasing range, i.e., from South-West to North-East. No data from these source tows were used in any of the inversions in the benchmark.

III. DESCRIPTION OF THE BENCHMARK METRICS

The first metric is a quantitative analysis of estimated model parameters (Section III-A). The other two metrics (Sections III-B and III-C) involve using the estimated model parameters to predict acoustical quantities, such as wave numbers or TL. Note that this requires using a complete set of geoacoustic parameters, including density and attenuation. Several geoacoustic inversion methods are insensitive to (or do not provide information about) density and/or attenuation. In such case, we have arbitrarily assigned values. This values are identified in Tables I and II with a * superscript. The impact of this choice is also discussed throughout the article.

A. Comparison With Ground Truth

Comparisons with ground truth data can be made with some of the estimated model parameter values. However, care must be taken in applying this metric to ensure that the environmental conditions of the ground truth data and the estimates or measurements from the experiments are reasonably similar. As explained previously, ground truth data are generally obtained in experiments with significantly different resolution and/or in very specific sediment environments.

B. Modal-Based Comparison

As a second metric, the estimated geoacoustic models are used to predict modal wave numbers. Those can easily be plotted and qualitatively compared. Of particular interest here is the definition of a quantitative metric that would allow cross-comparison with different geoacoustic models.

To do so, let us consider two different geoacoustic models, $\mathcal{A}$ and $\mathcal{B}$. They are used to predict two sets of modal wave numbers, $(k_m^{\mathcal{A}})_{m \in [\![1;M_{\mathcal{A}}]\!]}$ and $(k_m^{\mathcal{B}})_{m \in [\![1;M_{\mathcal{B}}]\!]}$, with $M_{\mathcal{A}}$ the number of modes predicted by the model $\mathcal{A}$ and $M_{\mathcal{B}}$ the number of modes predicted by the model $\mathcal{B}$. The distance between the two wave number sets can be easily quantified by computing the euclidian distance

$$d_k(\mathcal{A}, \mathcal{B}) = \sqrt{\sum_{m=1}^{M} |k_m^{\mathcal{A}} - k_m^{\mathcal{B}}|^2} \qquad (1)$$

TABLE I
GEOACOUSTIC MODEL PARAMETERS FOR MIDFREQUENCY INVERSION

Geoacoustic Model	Frequency	Layer	Thickness (m)	Sound speed (m/s)	Attenuation (dB/m/kHz)	Density (kg/m^3)
Yang et al. [11]	3 kHz	1	Half space	1650 $\pm$ 15	0.17	2050 $\pm$100
Jiang et al. [18], [19]	1.5-4.5 kHz	1	21.4	1598 $\pm$ 9	0.095 $\pm$0.01	1800*
		2	Half space	1740*	0.3*	2200*
Choi et al. [10]	2-6 kHz	1	0.2†	1680†	$\diamond$	2100†
		2	21.8	1630 $\pm$ 20	0.05 $\pm$0.01	2000†
		3	Half space	1740†	0.3†	2200†
Turgut [23]	4-10 kHz	1	$\sim$20		0.12 $\pm$0.01	

The superscript † means that a value has been assigned by user, not inverted. The superscript * means that the value has been assigned in this article from previous published data, but was not inverted/defined by user. The symbol $\diamond$ means that Choi *et al.* assigned a nonlinear frequency-dependent attenuation $\alpha = 0.2f^{1.6}$ (f in kHz and α in dB/m).

TABLE II
GEOACOUSTIC MODEL PARAMETERS FOR LOW-FREQUENCY INVERSION

Geoacoustic Model	Frequency (Hz)	Layer	Thickness (m)	Sound speed (m/s)	Attenuation (dB/m/kHz)	Density (kg/m^3)
Ballard et al. [20], [21]	175	1	8†	1670 $\pm$ 12	0.1*	1800*
		2	12†	1585 $\pm$ 19	0.1*	1800*
		3	Half space	1725 $\pm$ 15	0.3*	2200*
Jiang/Chapman [12], [13]	53-703	1	23$\pm$2	1626 $\pm$ 17 / 1591 $\pm$ 15	$\diamond$	1680 $\pm$ 100 / 1820 $\pm$ 200
		2	Half space	1758 $\pm$ 30	0.3*	2340$\pm$600
Huang et al. [14]	53-953	1	24.7$\pm$3	1596 $\pm$ 11 / 1629 $\pm$ 18	0.12 $\pm$ 0.12	1700 $\pm$ 100
		2	Half space	1758 $\pm$ 24	0.12†	2100†
Bonnel et al. [15], [16]	50-200	1	25$\pm$2	1604 $\pm$ 20	0.1*	1800 $\pm$ 200
		2	Half space	2000 $\pm$ 200	0.3*	1480 $\pm$ 600
Duan et al. [17]	50-200	1	20.8$\pm$3	1607 $\pm$ 10	0.05$\pm$0.02	1700 $\pm$ 100
		2	Half space	1900$\pm$100	0.3*	2690 $\pm$ 500

The superscript † means that a value has been assigned by user, not inverted. The superscript * means that the value has been assigned in this article from previous published data, but was not inverted/defined by user. The symbol $\diamond$ means that Jiang and Chapman estimated a nonlinear frequency-dependent attenuation $\alpha = 0.15f^{1.35}$ (f in kHz, and α in dB/m).

with M the smallest number between $M_\mathcal{A}$ and $M_\mathcal{B}$. As a result, if a model predicts more modes, the highest order modes are excluded from the comparison. The underlying assumptions are that the highest order modes have the highest attenuation, and their impact on the predicted acoustical field is minimal.

Note that the distance d_k can be computed for complex wave numbers, and/or for only their real or imaginary part. In the following, d_k will be computed by considering only the wave number real part. This is justified as some of the inversion methods considered in this benchmark did not invert for attenuation. By focusing on the wave number real part, the proposed metric assesses the impact of sound speed and density profiles. It is thus adapted to cross-compare all the inverted models, independently of sediment intrinsic attenuation. In the following, wave numbers are computed using KRAKEN [30].

C. Comparison With Experimental TL

Lastly, the estimated geoacoustic models are used to predict coherent TL versus range, and these simulated TL are compared to experimental TL. All the TL considered in this article are in dB.

1) Definition of Relative TL: Classically, TL are defined as $\mathrm{TL}(r) = \mathrm{RL}(r) - \mathrm{SL}$, with $\mathrm{RL}(r)$ the received level at range r and $\mathrm{SL} = \mathrm{RL}(1\,\mathrm{m})$ the source level at 1 m. The estimation of experimental TL thus requires knowledge of the experimental SL.

The TL2a/TL2b SL have been estimated with an uncertainty of a few decibels [31], which directly translates as the same uncertainty on the TL. A preliminary analysis (not shown here) has demonstrated that this uncertainty is too high for a meaningful quantitative comparison of experimental TL with simulated TL for use in the benchmark.

To circumvent this issue, we define the *relative TL*

$$TL_0(r) = RL(r) - RL_0 \qquad (2)$$

as the received level at range r normalized by a known received level at another range, RL_0. Note that $RL_0 = RL(1 \text{ m}) = SL$ gives the classical definition of TL. However, (2) offers more freedom. Here, we choose RL_0 as the mean RL for ranges between 500 and 1500 m. When necessary and to prevent confusion, the traditional TL normalized to 1 m will be denoted *absolute TL*. The term *relative TL* will be used exclusively to refer to TL_0.

Relative TL can be computed with the acoustical data sets that are used for the benchmark. There is no need for independent information about the SL from external data sets. Here, RL_0 is computed over ranges that are large enough so that near-field effects are minimized. Averaging over the 1 km range aperture helps to mitigate uncertainties associated with source/receiver positions.

A different procedure was applied to the DRDC data. Since ranges less than 1500 m were not available for the DRDC data, the relative TLs could not be normalized at a sufficiently short range to limit the impact of the seabed on the relative TL values. The starting range of the DRDC data is well within the mode stripping zone [32] in which mode amplitudes decrease significantly with range due to interaction with the ocean bottom. On the other hand, accurate SL measurements are available. As a result, absolute TL will be used for the DRDC track.

2) Metric Definition: We define two metrics to quantify the comparison between two TL sets, $TL^{\mathcal{A}}(r)$ and $TL^{\mathcal{B}}(r)$. Note that for the metric definition, it is unimportant to know if the TL have been simulated or experimentally measured. It is also unimportant to know if the TL are absolute or relative.

The aim of the metrics is to quantitatively compare both the overall loss behavior of the TL, and also the detailed interference patterns of the acoustical field. These two features are important in assessing sonar performance [33]: the overall loss behavior is related to a mean detection range, while interference patterns impacts the false alarm probability.

To assess those, we suggest metrics that compare the TL statistical distributions (i.e., histograms) over a given range aperture. Although others have presented direct comparisons between calculated and measured TL over a few kilometers (e.g., [21]), a simple euclidian distance or correlation between measured and calculated TL at every range is not effective for the larger ranges and higher frequencies in our study. A deterministic match between calculated and measured interference patterns is significantly more challenging when more than a few modes are propagating.

Formally, we define $TL^{\mathcal{A}}(r_0)$ and $TL^{\mathcal{B}}(r_0)$ as the TL in dB with r restricted to a given range aperture centered on r_0. Two comparison metrics are then computed. The first one is a simple difference between the TL medians

$$d_{TL}^{med}(\mathcal{A}, \mathcal{B}, r_0) = < TL^{\mathcal{A}}(r_0) > - < TL^{\mathcal{B}}(r_0) > \qquad (3)$$

where the brackets $< \cdots >$ indicates the median. This metric will be called *median distance* from now on. It is adapted to compare the overall loss behavior of the TL.

We further define $\mathcal{H}[TL^{\mathcal{A}}(r_0)]$ and $\mathcal{H}[TL^{\mathcal{B}}(r_0)]$ as the histograms of $TL^{\mathcal{A}}(r_0)$ and $TL^{\mathcal{B}}(r_0)$. The histograms are computed using 20 regular bins, with a minimal (respectively maximal) bin value that corresponds to the smallest minimum (respectively largest maximum) value of $TL^{\mathcal{A}}(r_0)$ and $TL^{\mathcal{B}}(r_0)$; The bin parameters thus depend on the range r_0. Note that the bin size definition has been empirically chosen to provide relatively smooth histograms.

To compare the interference pattern statistics, we compute the *total variation distance* of TL histograms [34]

$$d_{TL}^{TV}(\mathcal{A}, \mathcal{B}, r_0) = \frac{1}{2} \sum_i |\mathcal{H}_i[TL^{\mathcal{A}}(r_0)] - \mathcal{H}_i[TL^{\mathcal{B}}(r_0)]| \qquad (4)$$

where $\mathcal{H}_i$ is the value of the ith bin of the histogram $\mathcal{H}$. An interesting feature of the d_{TL}^{TV} is that its values are between 0 and 1, provided that the histograms are normalized as relative probabilities (i.e., the bin values are the number of elements in the bin divided by the number of elements in the input data). A distance of 0 means that TL distributions are identical (small distance), while 1 means that the TL distributions are completely different (large distance). Indeed, (4) shows that $d_{TL}^{TV} = 0$ if the two histograms are identical, while $d_{TL}^{TV} = 1$ if the histograms are not overlapping. Also, because the comparison is performed on TL histograms, this allows comparison of TL that are computed on different range samples, as long as enough samples are available to estimate the histograms. This will be useful when comparing experimental TL with simulated TL.

As an example, Fig. 2 presents examples of median distance and total variation distance for three absolute TL sets. The TL presented here are actually extracted from the SW06 benchmark, but their specificity is unimportant for now. The left column of Fig. 2 shows two TL sets that are relatively similar, and thus their histograms are similar too. As a result, $d_{TL}^{TV} = 0.22$ is small, and the median distance is $d_{TL}^{med} = 1.03$ dB. In the middle column, the two TL are largely different, with the red one globally smaller than the blue. As a result, the two histograms are centered on different values, which result in relatively large distances: $d_{TL}^{TV} = 0.5$ and $d_{TL}^{med} = 2.15$ dB. Last but not least, the right column presents an intermediate scenario. The 2 TL sets are centered around the same value, which leads to a very small $d_{TL}^{med} = -0.29$ dB. However, the TL distributions are different, which leads to different spreading for the histograms, and thus an intermediate $d_{TL}^{TV} = 0.31$.

For the experimental benchmark, d_{TL}^{TV} and d_{TL}^{med} will be computed using sliding segments of 2 km with 1500 m of overlap (75%). We will now detail how simulated and experimental TL are computed

3) TL Simulation: Range dependent TLs are computed using the parabolic equation (PE) code RAM [35]. The bathymetry of the experimental area is interpolated to define the bathymetry along the acoustical tracks, and subbottom layers are assumed to be parallel to this bathymetry. The seabed layer thicknesses, sound-speed profile (SSP), densities, and attenuation are given by the geoacoustic models under study. Various conductivity temperature depth measurements were recorded during the source tows, and the mean SSP along each track is used as a deterministic input for simulating propagation on each track. The
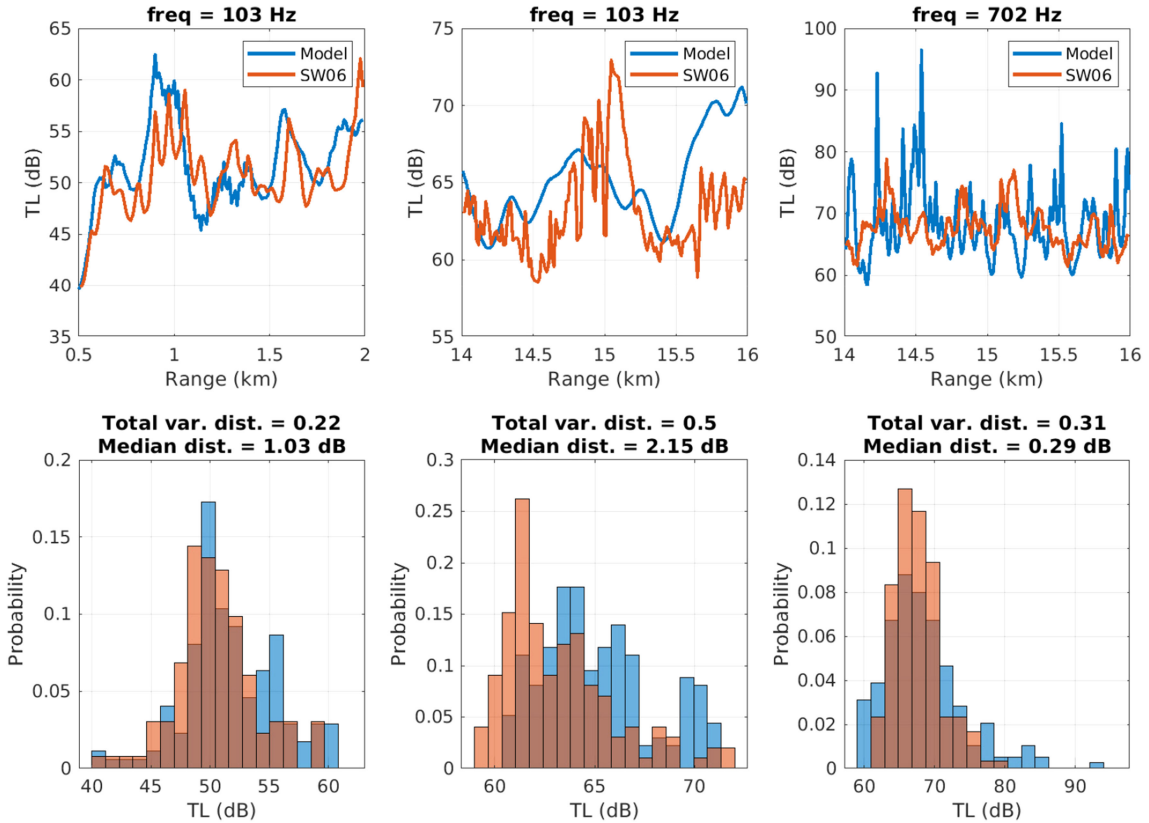
Fig. 2.   Illustration of the total variation distance and median distance to compare different TL sets.

PE code has also been slightly modified to include frequency-dependent volume attenuation in the water column.

It is important to notice that the TL2a/TL2b source trajectory is not aligned with the receiver. The propagation is thus simulated using a N-by-2-D scheme, with a 2-D slice every 10 m along the source trajectory. Also, source depth is adjusted for every slice, based on measurements performed during the experiment. On the other hand, propagation along the DRDC track is simulated using a classical 2-D scheme. The DRDC source depth was also monitored during the experiment, and barely fluctuated during the track.

Other simulation parameters are common for the three tracks. They are as follows: 2-D starter, range step $dr = 1$ m, depth step $dz = 0.1$ m, maximal depth $z_{\max} = 1000$ m, number of Padé coefficients $n_p = 5$, number of stability terms $n_s = 1$, stability constrains used at all ranges, and mean sound speed c_0 chosen as the averaged of the water sound-speed profile. Also, because ship speed is known, TLs are simulated at the doppler shifted frequencies, not at the theoretical source frequencies.

4) Estimation of Experimental TL: The experimental TLs along the TL2a/TL2b tracks are obtained using (2), which requires the estimation of the experimental RL. To do so, the signal is first decimated at $f_s = 2500$ Hz, and divided into snapshots of length L. Power spectra are then computed for each snapshot using fast Fourier transforms. The narrowband $RL(f)$ is estimated at the frequency f by summing B adjacent frequency bins. Chosen parameters for TL2a are $L = 10$ s and $B = 5$ (i.e., 0.4 Hz), and chosen parameters for TL2b are

$L = 5$ s and $B = 5$ (i.e., 0.8 Hz). The reasoning that led to the chosen values for L and B is detailed in Appendix A. Last but not least, the signal-to-noise ratio (SNR) is estimated for each snapshot and each frequency on the TL2a/TL2b track. The SNR is computed as the difference (in dB) of the tone power, minus the average power of the noise in a 10-Hz band encompassing the tone frequency. All the experimental TL with a SNR< 10 dB are removed from the analysis. This effectively removes all the 53 Hz data, as well as several datapoints at other frequencies. The SNR threshold value (10 dB) is a subjective choice, based on empirical observation of the data. Nonetheless, a relatively high SNR has been chosen to focus the benchmark on high quality data, so that performance is mostly driven by geoacoustic properties and not by noise.

The experimental (absolute) TLs along the DRDC track have been provided by DRDC. They were estimated following the method presented in [36]. Although SNR information was not available for this article, the data we use have been previously validated by DRDC, so that SNR is not an issue. All the DRDC data are thus kept for the analysis.

IV. Results From SW06 Experiments

Results from experiments carried out in SW06 for characterizing the ocean bottom sediments fall into two categories: direct measurements of sediment properties made with broadband *in situ* probes at specific locations in the near vicinity of the outer

shelf wedge site, and geoacoustic models estimated by inversion of data acquired in other experiments.

The *in situ* probes penetrated only the upper 1–1.5 m of sediment, and covered the mid- to high-frequency bands from 2–60 kHz. The results of these experiments provide information relevant to the surficial sediment material, and can be compared to the ground truth data from grab samples, cores and other in situ probes.

Geoacoustic inversions from SW06 generally used low frequency data (< 1 kHz), and the estimated geoacoustic models are representative of the upper sediment material to depths of at least the R-reflector ($\sim$20–25 m). The low-frequency inversion results are based on data from relatively long range experiments in which the sound signals interacted with variable sediment types in range along the propagation path, and depth below the seafloor. The geoacoustic models from the inversions are interpreted as models of effective media that are simplified approximations to the real ocean bottom. The estimated values of model parameters such as sound speed and attenuation represent averaged values for the effective sediment medium over the propagation path from the sound source. However, there were also inversions based on the midfrequency data (1–10 kHz) that characterize the surficial sediment material. Comparison with ground truth provides a reasonable metric for assessing the midfrequency inversion results.

Based on the ground truth information from the sonar surveys, three types of geoacoustic profiles were assumed for the ocean bottom in the inversions: a half-space model; a single sediment layer over a half-space; or a two-layer sediment/half-space model. The models were all assumed to be fluid layers described by sound speed, attenuation, and density. The half-space model was appropriate for inversions with higher frequency data (2–3 kHz), while the layered models were used for lower frequency data.

A. In Situ Measurements

Sound speed of surficial sediment material was measured by two different broadband instruments in the near vicinity of the outer shelf wedge site. The Sediment Acoustic-speed Measurement System (SAMS) obtained data from 2 to 20 kHz, sampling to a depth of $\sim$1.5 m [8], and the Geoprobe instrument made measurements from 10 to 60 kHz, sampling sediment within $\sim$0.5 m of the seafloor [9]. The values from SAMS were 1607 ± 11 and 1611 ± 10 m/s for the 2–11- and 10–21-kHz bands, respectively. The Geoprobe measurements were 1640 ± 10, 1650 ± 25, and 1660 ± 10 m/s for bands centred at 15, 35, and 60 kHz, respectively. The results for sound-speed ratio referred to the sound speed at the water bottom are displayed in Fig. 3. It is evident from the figure that there is at best only weak dispersion over the frequency band. For comparison, the averaged sound-speed ratio of point samples from the ground truth survey (see [24], Table 1) that were closest to the outer shelf wedge site is also plotted in the figure (red square). However, the value is characteristic of a surficial covering of coarse grain sandy sediment within 0.15 m of the seafloor. The Geoprobe also measured sound attenuation, with values consistent with
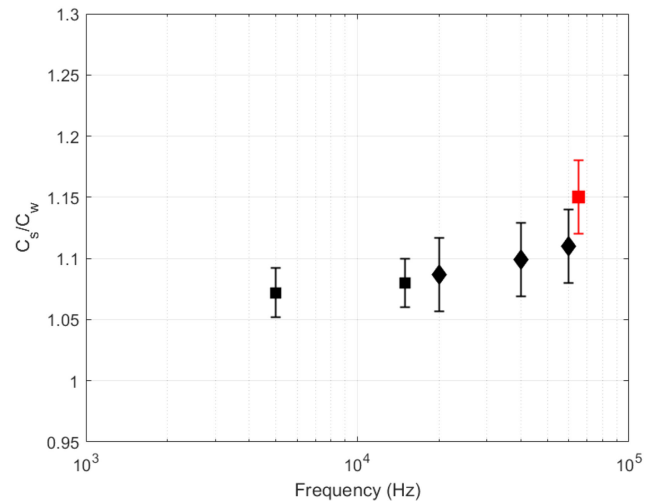

Fig. 3. *In situ* measurements of seafloor sediment sound-speed ratio. Squares: SAMS data [8]; black diamonds: Geoprobe data [9]. The error bars for the Geoprobe data represent the upper and lower bounds of values obtained over each frequency range. The red square shows the result from the 2000–2001 survey [24]. The error bar represents the upper and lower bounds of values from the subset of grab samples taken about the outer shelf wedge site.

0.27 dB/(m $\cdot$ kHz) (0.45 dB/λ) to frequencies over the frequency band $\sim$20–60 kHz [9].

B. Inversion of Geoacoustic Models

Geoacoustic models were inverted from data spanning the low (< 1 kHz) to mid ($\sim$3 kHz) frequency bands. We describe the results of four different inversions of midfrequency data first, and then introduce results of the low frequency inversions. All of the inversions estimated sound speed in the sediment, but it is important to note here that some of the inversions did not invert attenuation and/or density. In some of those cases, values for the unestimated parameters were assigned by the users in their reports. In the remaining cases, appropriate values consistent with values estimated by the other inversions are assigned in this article.

1) Inversions Using Midfrequency Data: Midfrequency bottom loss versus grazing angle data obtained at the MORAY array were inverted by Yang *et al.* assuming a half-space geoacoustic model [11]. This assumption is reasonable for the relatively high frequency data ($\sim$3 kHz), and the estimated model parameters are characteristic of the outer wedge shelf sediment within the first few meters beneath the seafloor. The estimated values for sound speed, attenuation, and density and their uncertainties are listed in Table I. Note that Yang *et al.* actually estimated attenuation at several frequencies (2, 3, 4, and 5 kHz), but did not provide a broadband attenuation model. As a result, only the 3-kHz value is considered in this article, which ensures consistency with other results that are used here [10].

Jiang *et al.* [18] and Jiang and Chapman [19] used travel time difference and spectral ratio data between the signals reflected from the seafloor and R-reflector to infer sound speed and attenuation in the outer shelf wedge sediment above the R-reflector. The data were obtained at the MPL vertical array in a short

range (< 200 m) experiment using linear frequency modulated (LFM) pulses in the band 1.5–4.5 kHz. The estimated values for a single layer geoacoustic model are listed in Table I. The same LFM data were also used by Michalopoulou [37] to invert sound speed and thickness of a single layer model from estimates of arrival times of the sequence of seafloor and subbottom arrivals at the array. Her results for the structure and sound speed were in close agreement with those of Jiang *et al.* [18].

Choi *et al.* [10] constructed a multilayer geoacoustic model including a thin (20 cm) high sound-speed layer at the seafloor, consistent with expectations from ground truth information. Model parameter values for sound speed in the layers were evaluated from analysis of travel time and reflectivity data of midfrequency LFM pulses received at the MORAY array. Other model parameters (e.g., density) were taken from the ground truth data.

These latter two inversions estimated values for only the model parameters of the assumed homogeneous single sediment layer, and did not invert for the parameters of the underlying half space. In all the three inversions, a range-independent environment was assumed. This was reasonable for the short-range (< 300 m) experimental geometry near the site. The estimated values and their uncertainties are listed in Table I.

Turgut [23] used spectral ratio data for the seafloor and R-reflector obtained in a chirp sonar survey to infer sound attenuation in the sediment layer in the band 6–8 kHz. His result is also listed in Table I.

2) Inversions Using Low-Frequency Data: Matched field inversions of low frequency CW tones from 53 to 700 Hz were carried out by Jiang and Chapman [12], [13] and by Huang *et al.* using tones from 53 to 953 Hz [14]. Each group generated estimates of sound speed, attenuation, and density for a single layer/half space geoacoustic model using data recorded on the MPL VLA from the same experiment. Notably, Huang *et al.* [14] assumed a linear increase of sediment sound speed and assumed adiabatic change of water depth, whereas Jiang and Chapman [12], [13], inverted a negative sound-speed gradient assuming a range-independent environment. The latter profile is consistent with the expectation of a low sound-speed layer at depth in the sediment above the "R"-reflector. However, the interval sound speed in the layer is about the same for each geoacoustic model. The model parameters are listed in Table II.

Bonnel *et al.* [15], [16] inverted broadband low frequency (< 200 Hz) modal group velocity data from a light bulb implosion recorded on a single hydrophone to estimate sound speed and density for a single layer/half-space geoacoustic model. The same broadband signal and another implosion at shorter range were used by Duan *et al.* [17]. Using the signal recorded on the MPL VLA (instead of a single hydrophone), Duan *et al.* extended the inversion by including amplitude information of the resolved modes to estimate sound attenuation in the outer wedge sediments. The model parameters of both geoacoustic models are listed in Table II. The light bulb data were also inverted by Taroudakis *et al.* [38] and [39] who applied an approach based on a sparse signal denoising scheme to improve the SNR. They obtained a model similar to the one developed by Bonnel

and Chapman [15] and Bonnel *et al.* [16]. All these inversions assumed a range-independent environment.

Ballard and Becker [20] and Ballard *et al.* [21] adopted a linearized approach that used perturbation theory to relate changes in sound speed to changes in modal wave numbers in inversions of low frequency (< 200 Hz) CW tones. The inversion assumed information about depths of interfaces from the chirp sonar ground truth survey to invert sound speed in a two-layer, range-dependent geoacoustic model. This approach-enabled resolution of the low sound-speed layer above the R-reflector; other inversions of low frequency data were able to resolve only a single layer. The geoacoustic model parameters are listed in Table II.

Rajan and Becker also used a linearized inversion of modal wave numbers and travel times to estimate range-dependent geoacoustic profiles in the experimental area [22]. However, the model relevant to the benchmark track was essentially the same as the one developed by Ballard and Becker [20] and Ballard *et al.* [21].

Similar to the inversions with midfrequency data, the estimates for the sediment model parameters from low frequency data represent effective values assuming homogeneous sediment layers. All the inversions with low frequency data generated estimates for the sound speed of the underlying half-space, but only a small subset inverted for the density and none inverted for the half-space attenuation. This suggests that the experimental data did not carry significant information about these latter two parameters of the half space.

V. EXPERIMENTAL BENCHMARK ANALYSIS: ESTIMATED GEOACOUSTIC MODEL PARAMETERS

This section presents a comparison of estimated values of specific geoacoustic model parameters. All the inversions estimated sediment sound speed, and a few inversions estimated sound attenuation. Comparison of estimated values of these model parameters provides information about their dependence with depth in the sediment, and the dispersion over the frequency band.

The comparison is separated into two parts. Since there is ample evidence from the ground truth information that the sediment above the R-reflector is significantly inhomogeneous with depth, inversions that characterize the surficial sediment structure within 1–3 m beneath the seafloor are considered separately from those other inversions that characterize the sediment structure to depths of the R-reflector. The former includes the inversion of midfrequency data by Yang *et al.* [11]. The latter group includes the inversions by Choi *et al.* [10] and Jiang and Chapman [12], [13] that used midfrequency data, and all the inversions based on low frequency data (see Table II).

A. Estimates of Surficial Sediment Model Parameters

The ground truth metric is appropriate for the inversion by Yang *et al.* [11] based on midfrequency data from which sound speed, density, and attenuation were estimated for the surficial sediment material within 1–3 m of the seafloor. The SAMS [8] and Geoprobe [9] measurements are used as ground truth data for this comparison, since these measurements are relevant to the
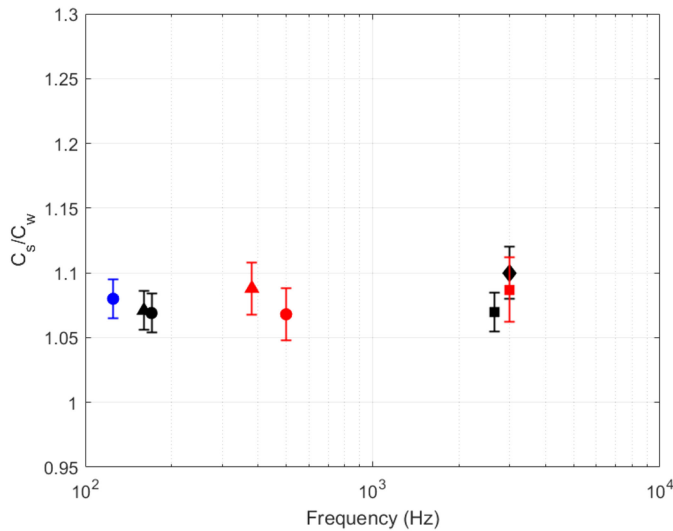
Fig. 4. Inversions of effective sound speed of outer wedge sediment. Results are plotted as ratios of sound speed to bottom water sound speed. Circles: Ballard *et al.* [20], [21] (blue); Bonnel *et al.* [15], [16] (black); Huang *et al.* [14] (red). Triangles: Duan *et al.* [17] (black); Jiang and Chapman LF [12], [13] (red). Squares: Choi *et al.* [10] (red); Jiang *et al.* MF [18], [19] (black) Diamond: Yang *et al.* [11] (black). Note that matched field inversion result are plotted at the center frequency of the band used in the inversion.
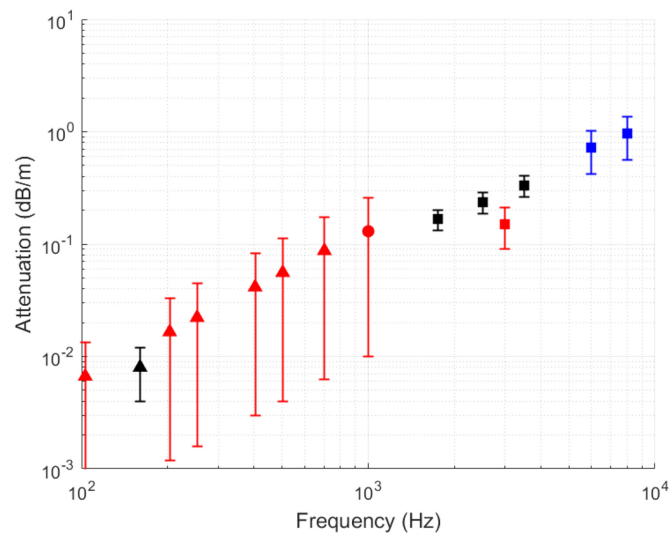
Fig. 5. Inversion of effective attenuation. Circles: Huang *et al.* [14] (red). Triangles: Duan *et al.* [17] (black); Jiang and Chapman LF [12], [13] (red). Squares: Choi *et al.* [10] (red); Jiang *et al.* MF [18], [19] (black); Turgut [23] (blue). Note that matched field inversion results are plotted at the center frequency of the band used in the inversion.

surficial sediment in the immediate vicinity of the experiment. The estimated sound speed of 1650 m/s is consistent with the measured sound speeds, which fall within 1600–1660 m/s. Likewise, the estimated attenuation of 0.33 ± 0.1 dB/(m · kHz) is close to the Geoprobe measurements of 0.27 dB/(m · kHz). The estimated values of sound speed and attenuation are plotted in Figs. 4 and 5, respectively. The estimated density, 2050 kg/m^3 is consistent with the measurements from the sediment core.

The estimates of sound speed and attenuation from this inversion are significantly less than the measured values by Goff *et al.* [24] in the 2000–2001 survey. However, the measurements from the earlier survey sampled only the top ∼10–20 cm of sediment. Since the inversion includes contributions from deeper sediment material, the comparison with the estimates from the inversion indicates that sound speed and attenuation in the upper portion of the sediment decrease substantially in the first 1–3 m beneath the seafloor.

B. Estimates of Effective Sound Speed and Attenuation

This section presents a comparison of estimated values of sound speed and attenuation derived from data that sampled the sediment material to depths of the R-reflector. The estimates represent effective values that characterize the inhomogeneous sediment material to depths of 20–25 m below the seafloor as an assumed single layer of homogeneous material. The result for Ballard and Becker [20] and Ballard *et al.* [21] is an interval sound speed over the two sediment layers (i.e., sound slowness is averaged). The sound-speed estimates are shown in Fig. 4 as relative values to the water sound speed at the bottom versus frequency. Note that the values obtained in the matched field inversions of Jiang and Chapman [12], [13] and Huang *et al.* [14] are plotted at the center frequency of the band used in the inversions.

There are two observations that can be made about these results. First, there is no appreciable trend of the effective sound speed for the sediment layer over the frequency band from 53 to 3000 Hz, with values ranging between 1.06 and 1.09. The inference to draw from this is that there is little evidence of sound-speed dispersion over the frequency band. Second, the estimated values from Yang *et al.* [11] which characterize the upper 1–3 m are slightly greater, >1.1. Other results by Ballard *et al.* [20], [21] (not presented here) that characterize the upper 8 m are also greater than 1.1. When taken as a whole, the results for the effective sound speed are consistent with a general decrease of sound speed in the sediment material to the depth of the R-reflector.

Sound attenuation in the sediment layer was inverted by Choi *et al.* [10], Duan *et al.* [17], Huang *et al.* [14], and Jiang and Chapman [12], [13] who inferred a frequency-dependent nonlinear attenuation: $\alpha(f) = \alpha_0 f^\beta$ (f in kHz) with $\alpha_0 = 0.15 \pm 0.14$ and $\beta = 1.35 \pm 0.2$. These results cover the frequency band from 50 to 3000 Hz and the estimates are shown in Fig. 5, along with the midfrequency measured values by Jiang and Jiang and Chapman [19] and Turgut [23] obtained from spectral ratio data. Note that the results from Jiang and Chapman [12], [13] are plotted at the frequencies used in the matched field inversion to show the nonlinear relation with frequency. The estimate from Huang *et al.* [14] who assumed a linear dependence is shown at 1 kHz for convenience.

The first observation is the near linear frequency dependence ($\beta = 1$) over the midfrequency band (1.75–10 kHz) for the outer shelf wedge sediment, based on the results from the spectral ratio inversions. Although it is possible to extrapolate the linear behavior to lower frequencies, the frequency dependence from

the matched field inversion [12], [13] appears to be different, and the results for the lowest frequencies suggest that there may be different behavior below 250 Hz. With the data sets at hand, a definitive conclusion on the frequency dependence of attenuation is not possible. This reinforces the importance of properly assessing uncertainty associated with geoacoustic inversion results.

The second observation is that the attenuation in the outer shelf wedge sediment is significantly lower than the attenuation expected for sandy sediments based on previous work summarized by Zhou *et al.* [40]. The expected attenuation for sandy sediments, $\alpha(f) = 0.35f^{1.8}$ is greater than the measured values over the entire frequency band shown in Fig. 5. And finally, the estimates from these inversions are all lower than the estimate from the bottom loss inversion of Yang *et al.* [11] which characterized the surficial 1–3 m of sediment. The inference from this is that attenuation may decrease with depth in the sediment.

A brief comment should be made about estimation of density. Only a few of the inversions estimated density in the sediment, with values between 1700 and 1800 kg/m^3 and high uncertainties. These values are much less than the estimate of Yang *et al.* [11], the one closest to the values obtained from the sediment cores. It is not clear why the inversions with lower frequency data obtained values characteristic of "softer," less consolidated material.

VI. EXPERIMENTAL BENCHMARK ANALYSIS: GEOACOUSTIC INVERSION PERFORMANCE

The analysis of the wave number metric and the two TL metrics is presented here. As indicated previously, some inversions did not estimate attenuation and density of either or both of the sediment and the half-space, and so appropriate values were assigned in this article to carry out the calculations associated with the metrics. The metrics for some of the models are therefore dependent to some degree on the values that were assigned for the unestimated parameters. Values for these parameters were assigned from ground truth information and/or results from estimated values from the other experiments. In particular, values are needed for density and attenuation for several geoacoustic models. Arbitrarily round numbers have been chosen for all the models. Assigned values for densities are 1800 kg/m^3 for sediment layer and 2200 kg/m^3 for the basement. Assigned values for attenuation are 0.1 dB/(m · kHz) for sediment layer and 0.3 dB/(m · kHz) for the basement. Lastly, one assigned value for basement sound speed was needed for Jiang *et al.* MF [18] and Jiang and Chapman [19]. An assigned value of 1740 m/s has been chosen to mirror the (assigned) value of Choi *et al.* [10].

A. Analysis of Wave Number Metric

The wave number metric enables analysis of similarities between estimated geoacoustic models. Here, the wave number metrics are computed at the frequencies of the sources that have been used on the TL2a, TL2b, and DRDC tracks (from 53 to 1200 Hz). As a result, wave number and TL metrics can be computed on the same set of frequencies.

As indicated previously, the metric considers only the real parts of the wave numbers, and will thus be sensitive primarily to the sound-speed structure of the model. Information related to attenuation in the sediment is derived from the imaginary parts, and a qualitative sense of attenuation is shown in Fig. 6 where the real and imaginary parts are plotted for the different frequency bands. It is evident from inspection of the values of the imaginary parts in the figure that the inversion by Huang *et al.* [14] generated the highest attenuation, and the one by Duan *et al.* [17] the least. A quantitative assessment of the effect of estimated attenuation is reserved for the analysis of the TL metrics. It is also evident from the figure that different models may generate different numbers of modes. However, the difference lies in high order modes. Because those modes have a high imaginary part, they are highly attenuated. This effect should thus have a minimal impact on the acoustical field at long range.

The result of the wave number metric is displayed in the matrix in Fig. 7. As explained in Section III-B, the entries indicate the calculated distance d_k (see (1)) between wave numbers generated by the different models, with the colors signifying the degree of similarity: white is closest and most similar, black is farthest and least similar. Similarities that are evident between several different models are assessed in terms of salient features of the sound-speed profiles estimated by the different geoacoustic models.

There is strong similarity between Bonnel and Chapman [15] and Bonnel *et al.* [16] and Huang *et al.* [14] at all frequencies. These models feature thick sediment layers (∼25 m) and have similar average sound speed in the layer (∼1605 m/s). Although the similarity between Bonnel and Chapman [15] and Bonnel *et al.* [16] and Huang *et al.* [14] is the strongest, there is also a strong similarity between Huang *et al.* [14], Bonnel and Chapman [15] and Bonnel *et al.* [16], Duan *et al.* [17] and Jiang *et al.* [18], [19]. All these models feature roughly the same average sound speed in the sediment layer. Their differences in layer thickness and/or sound-speed profile appear to be less important.

There is a medium strength similarity between Ballard and Becker [20] and Ballard *et al.* [21] and Jiang and Chapman ("Jiang LF" on Fig. 7) [12], [13] at all frequencies. Although the similarity between the two is not that high, a striking feature is that they are clearly different from all the other models. Interestingly, these two models are the only ones that feature decreasing or low sound speed with depth in the sediment.

At low frequencies (<200 Hz), there is strong similarity between Duan *et al.* [17] and Huang *et al.* [14]; these models are similar in average sound speed in the layer (∼1605 m/s). However, the similarity decreases as frequency increases, likely due to the relatively low sound speed at the seafloor interface of Huang *et al.* [14]. The models of Duan *et al.* [17] and Choi *et al.* [10] are relatively similar over all frequencies. These models feature thin sediment layers (∼21 m). At high frequencies, (>500 Hz) there is strong similarity between Choi *et al.* [10], and Jiang and Chapman ("Jiang LF" in Fig. 7) [12], [13]. These models feature the largest values of sound
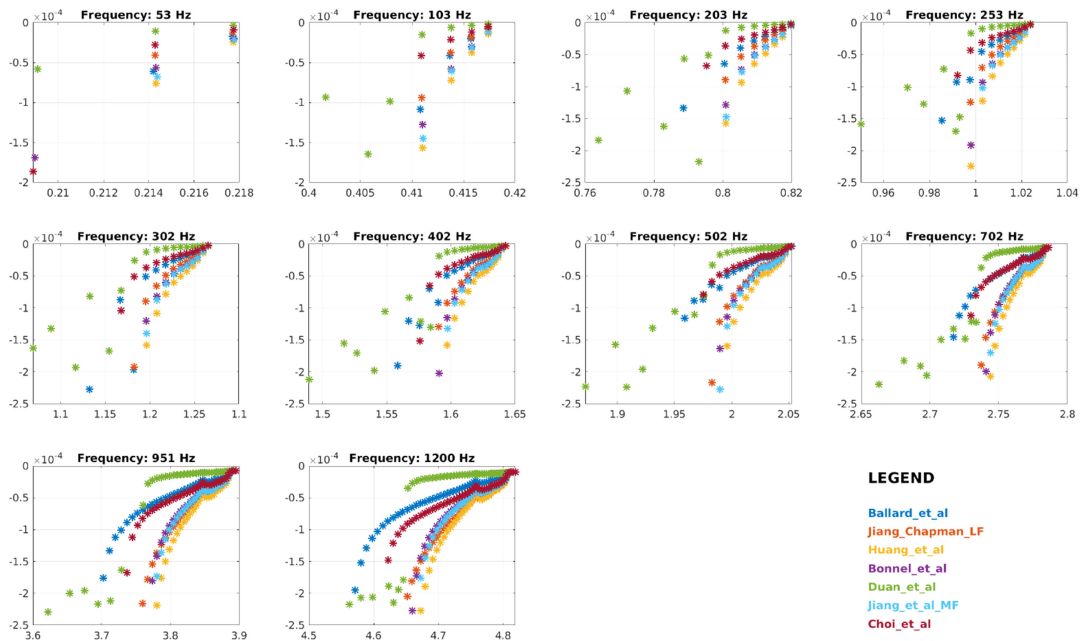
Fig. 6. Wave numbers predicted by all the geoacoustic models over the frequency band 53–1200 Hz. On each panel, the horizontal axis is the real part of the wave number, and the vertical axis is the imaginary part of the wave number. The color code is detailed on the bottom right of the figure, it will be used consistently for Figs. 8–12.
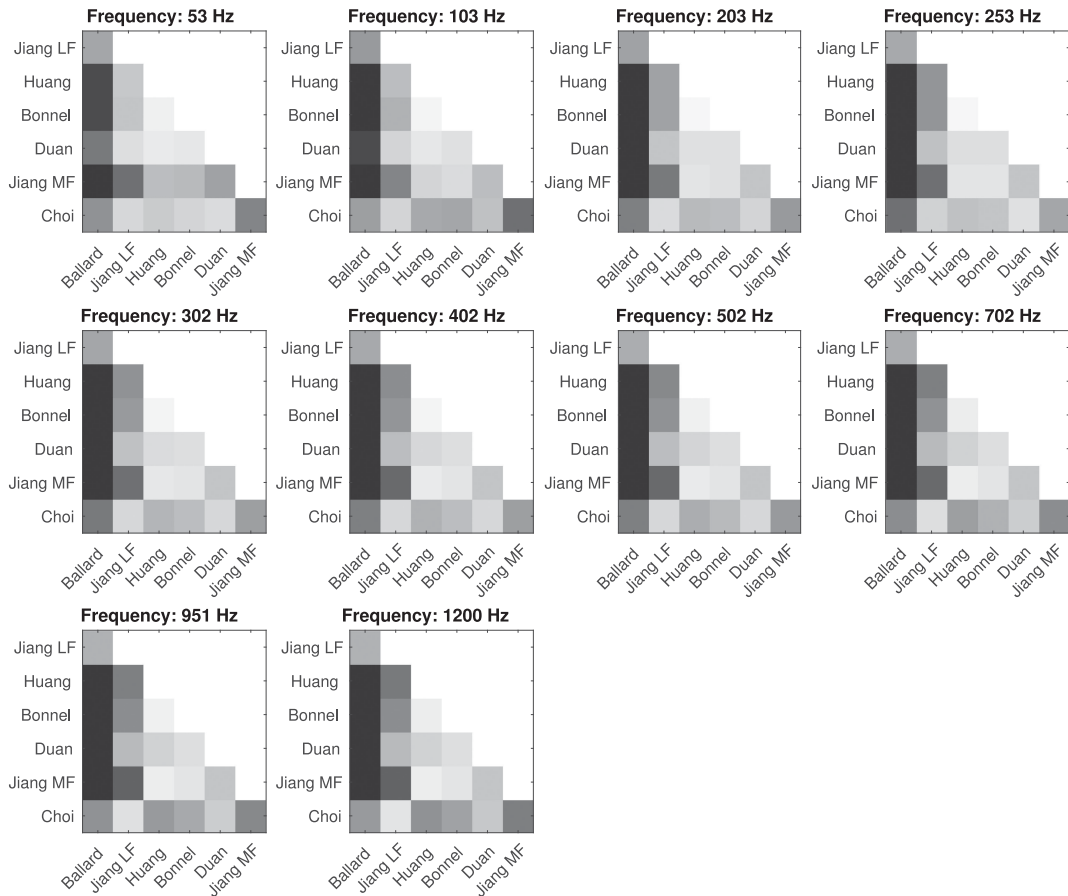

Fig. 7. Matrices of wave number distance between all the geoacoustic models. The panels show results for frequencies from 53 to 1200 Hz. On each panel, the colorscale linearly goes from white (minimal distance of zero between a model and itself) to black (maximal distance). Note that distance matrices are symmetric, but the upper-right half of each panel is not plotted to prevent redundancy.

speed at the seafloor (>1630 m/s). Medium strength similarity is evident between Duan *et al.* [17] and Bonnel and Chapman [15] and Bonnel *et al.* [16] at all frequencies. These models are close in average sound speed in the layer (∼1605 m/s) and feature high values of half-space sound speed (∼2000 m/s).

The analysis of the wave number metric shows that features of the sound-speed profile, such as sound speed at the seafloor, average sound speed in the sediment, and variation of sound speed with depth, can lead to similarities in the estimated models. The extent to which the features affect performance of the models in making realistic predictions of the sound field is investigated with the TL metrics.

B. Analysis of TL Metrics

The wave number metric allowed an intercomparison between the different geoacoustic models. The objective of the TL metrics is to enable a comparison of each geoacoustic model with experimental data. To do so, the TL metrics are computed with (3)–(4), with $\mathcal{A}$ the TL simulated using a geoacoustic model, and $\mathcal{B}$ the experimental TL. As explained in Section III-C, the metrics for tracks TL2a/TL2b are computed for relative TL only, and for experimental data with SNR >10. For the DRDC track, the metrics are computed for absolute TL over the whole track.

By nature, the TL metrics are range dependent. Because the considered data have been collected on VLAs, several depths are also available. The resulting metrics are thus range and depth dependent. To facilitate understanding, the metrics are first averaged over depth. Corresponding curves are given in Appendix B. Their most interesting feature is that the total variation distance globally increases with range. On the other hand, the spread between models of the median distance also tends to increase with range. These results illustrate that our ability to predict the acoustical field with even well-estimated geoacoustic models decreases with range. Environmental mismatch is always present, and its impact increases with range.

Next, the two TL metrics are range-averaged to provide single distance values per frequency for each geoacoustic model. Fig. 8 shows the median distance results, while Fig. 9 presents the total variation distance results. Such figures enable a synoptic and quantified evaluation of the performance of each geoacoustic model in predicting the acoustical field.

The median distance is analyzed first. A negative distance means that a geoacoustic model under-estimates the TL, while a positive distance means that the TL are over-estimated. In other words, if the median distance is positive, there is too much effective attenuation in the model; if the median distance is negative then the model lacks attenuation.

The median distance for the model by Duan *et al.* [17] is always negative, except at 103 Hz. This is consistent with the small attenuation value of this model and with the small imaginary parts predicted for the wave numbers (see Fig. 6). The result suggests that the assumption of linear frequency dependence associated with the estimate fails to model the observed TL as frequency increases. Nonetheless, the model performs well at 103 Hz, which is the only frequency included in the band

that was used for inversion. The low-frequency inversion by Jiang and Chapman [12], [13] is the only one providing a nonlinear estimate of the attenuation. Interestingly, the model performs well at all frequencies. On the other hand, the model by Huang *et al.* [14] and Bonnel and Chapman [15] and Bonnel *et al.* [16] have a similar behavior at all frequencies. These two models over-estimate TL at every frequency except at 702 Hz, where their median distance is virtually 0. This similarity is consistent with their small wave number distance. Two low frequency models (Ballard *et al.* [20], [21] and Bonnel and Chapman [15], [16]) did not invert attenuation and were assigned canonical values assuming a linear variation with frequency. It is nonetheless interesting to see that they present different features. The model by Ballard and Becker [20] and Ballard *et al.* [21] performs very well at most frequencies, notably below 402 Hz. Above 502 Hz, it tends to slightly underestimate TL. On the other hand, the model by Bonnel and Chapman [15] and Bonnel *et al.* [16] clearly over-estimates TL for frequencies up to 502 Hz, but is doing better for frequencies above 702 Hz. This demonstrates that attenuation is not the only factor driving the median distance metrics, and suggests that the sound-speed profile by Ballard and Becker [20] and Ballard *et al.* [21], a layered sediment with a slow speed at depth, is more appropriate than the isovelocity sediment by Bonnel and Chapman [15] and Bonnel *et al.* [16].

The two midfrequency models by Jiang *et al.* [18] and Jiang and Chapman [19] and Choi *et al.* [10] over-estimate the TL (positive median distance) for frequencies below 402 Hz, and have different behavior for higher frequencies. This behavior is unexpected, as the two inversions were performed using data with frequency above 1.5 kHz, and one would expect them to present similar performances at the highest frequencies. Nonetheless, their median distances are of the same order of magnitude as those obtained with the low-frequency models, demonstrating their ability to predict low-frequency acoustical fields reasonably well.

Overall, the total variation distance is less intuitive to understand than the median distance, because it takes into account the overall loss of the TL (as quantified by the median distance) and also the detailed interference pattern. As a result, a model that performs poorly in terms of median distance will also perform poorly in terms of total variation distance. As an example, this is the case for the models by Huang *et al.* [14], Bonnel and Chapman [15] and Bonnel *et al.* [16] and Jiang *et al.* (MF) [18] and Jiang and Chapman [19] at 402 Hz. On the other hand, one model may be better than another for the median distance, with opposite performance for the total variation distance. As an example, this is the case for the models by Ballard and Becker [20] and Ballard *et al.* [21] and Jiang and Chapman (LF) [12] and [13] at 502 Hz. Although the absolute value of the median distance is smaller for Jiang and Chapman (LF) [12], [13], the total variation distance is smaller for Ballard *et al.* [20] and [21]. This is due to the spread of the TL distribution caused by the interference patterns, which in this case is better predicted by the Ballard *et al.* model.

The total variation distance shows two interesting results. First, there is relatively little difference in the values for all the models at each frequency. The second is the behavior of
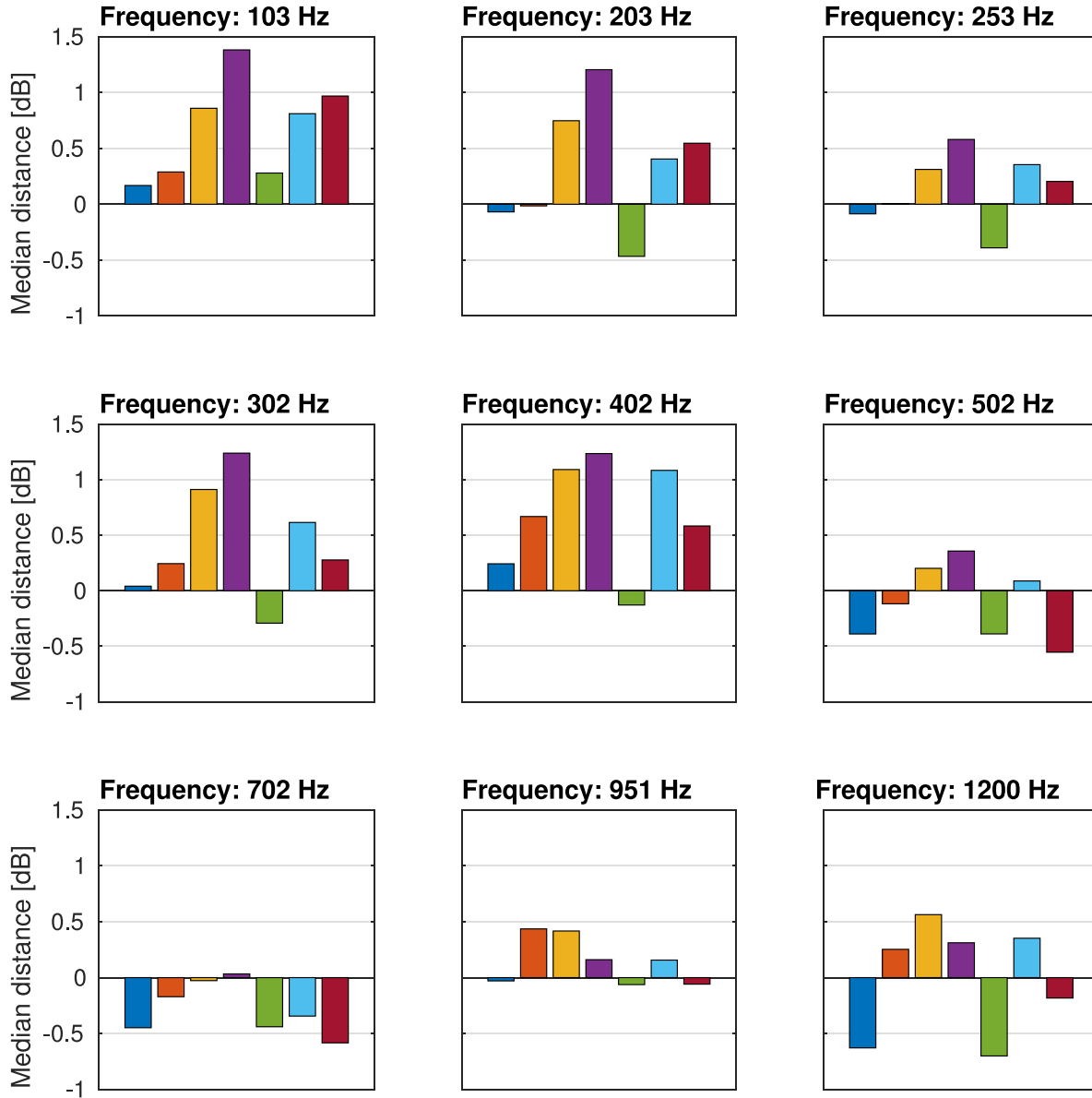
Fig. 8. Range averaged median distance (d_{TL}^{med}) between experimental and simulated TL. The vertical scale is the same on all the panels, and the color code is the same than on Fig. 6 (dark blue: Ballard *et al.*, orange: Jiang and Chapman LF, yellow: Huang *et al.*, purple: Bonnel *et al.*, green: Duan *et al.*, light blue: Jiang *et al.* MF, red: Choi *et al.*).

the global pattern over frequencies for all the models. This distance is very high for low frequency (103 and 203 Hz), decreases to reach a minimum at 502 Hz, and then increases gradually for higher frequencies. We believe this behavior is due to two different phenomena. At low frequencies, only a few modes are propagating, and thus interfering together. The resulting interference patterns are fully deterministic. They oscillate slowly, so that only a few oscillations are available to estimate the TL histograms in the range window. As a result, any environmental/position mismatch causes mismatch with interference patterns in the measured data; this can drastically impact the TL histogram, and thus increase the total variation distance. As frequency increases, the interference structure is more complicated because the number of modes increases, and the impact of this type of mismatch decreases. On the other

hand, at sufficiently high frequency, the acoustical field becomes sensitive to small environmental details which are not modeled here, such as range-dependent effects in the water columns or sediment inhomogeneities. This constrains our ability to predict TL accurately, and thus increases the total variation distance.

At first look, the total variation distance for the DRDC data (1200 Hz) seems to follow the previous pattern, with distances larger than those at 953 Hz. However, one has to remember that the DRDC TL were computed on a different track, with ranges smaller than on the TL2a/TL2b track. As a result, a direct comparison of the (range-averaged) DRDC results with the (range-averaged) TL2a/TL2b results must be done cautiously. Since the TL metrics tend to increase with range, it is likely that the TL metrics for the DRDC data would have been even higher if
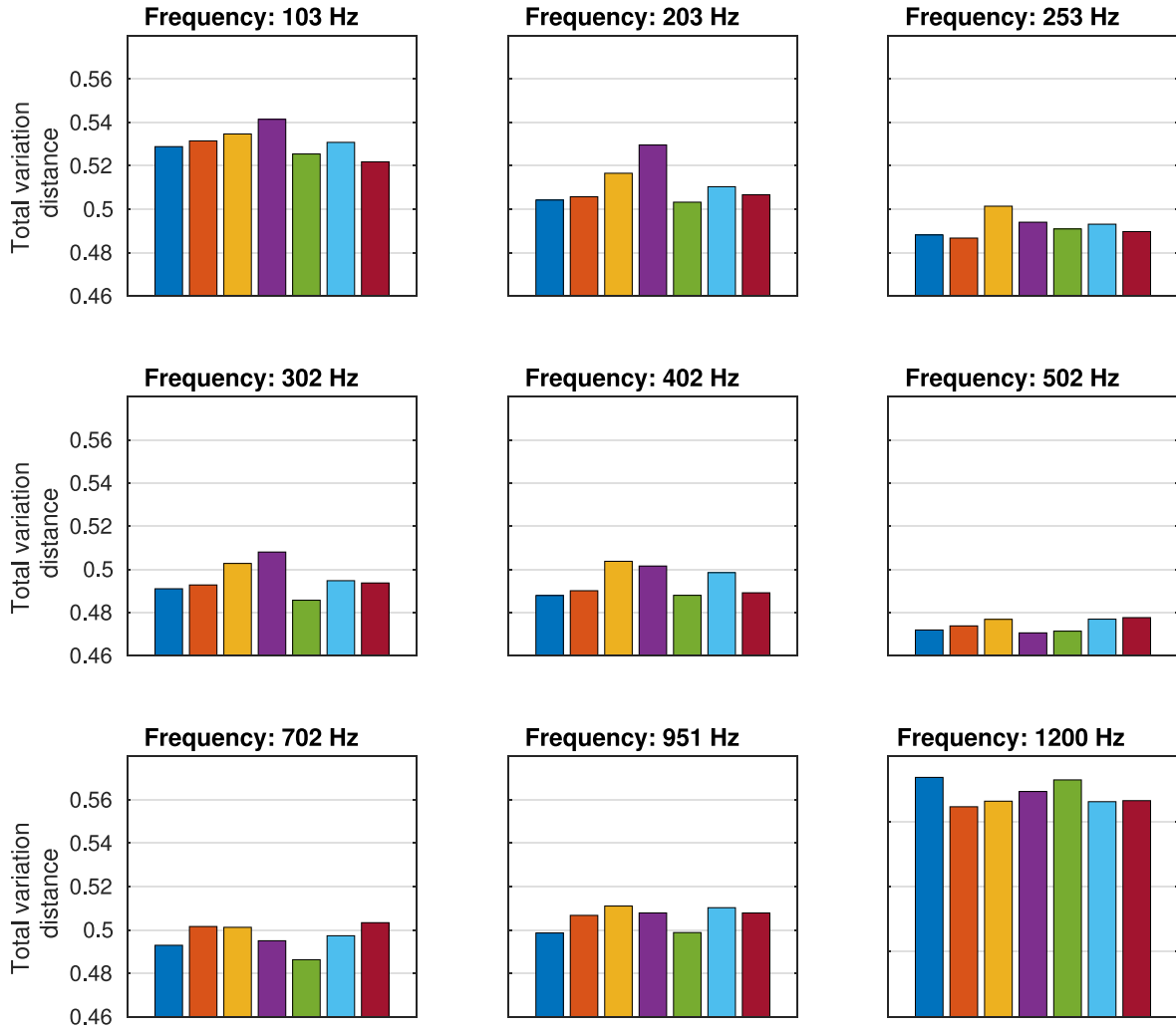
Fig. 9.   Range averaged total variation distance ($d_{\text{TL}}^{\text{TV}}$) between experimental and simulated TL. The vertical scale is the same on all the panels, and the color code is the same than on Fig. 6 (dark blue: Ballard *et al.*, orange: Jiang and Chapman LF, yellow: Huang *et al.*, purple: Bonnel *et al.*, green: Duan *et al.*, light blue: Jiang *et al.* MF, red: Choi *et al.*).

it could have been computed on the same range than the one used for the TL2a/TL2b tracks. As a reminder, range-dependent TL metrics are available in Appendix B. In particular, Fig. 12 shows that the total variation distance at 1200 Hz is notably higher than for other frequencies at similar range. Since the median distance at 1200 Hz is relatively similar to what is obtained at lower frequency, it means that the high total variation distance is mostly driven by the detailed interference pattern (i.e., the TL spread around their median value). This may be due to our inability to correctly predict the field at such frequency, or to the experimental estimate of the TL which is different for this specific frequency (DRDC track). Since the DRDC raw data is not available, one cannot provide a definitive conclusion.

The absolute value of the median distance and the total variation distance are also averaged over frequency, for frequencies between 103 and 953 Hz (TL2a/TL2b tracks). The result is presented in Fig. 10. One can see that the two best models from the perspective of median distance are the ones by
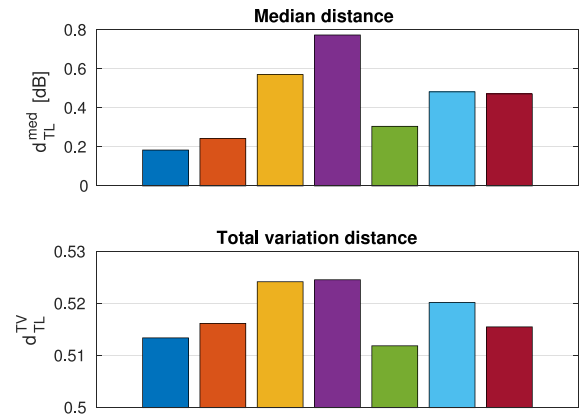

Fig. 10.   Frequency averaged median distance ($d_{\text{TL}}^{\text{med}}$) and total variation distance ($d_{\text{TL}}^{\text{med}}$) for the TL2a and TL2b tracks ($103 < f < 953$ Hz). The color code is the same than on Fig. 6 (dark blue: Ballard *et al.*, orange: Jiang and Chapman LF, yellow: Huang *et al.*, purple: Bonnel *et al.*, green: Duan *et al.*, light blue: Jiang *et al.* MF, red: Choi *et al.*).

Ballard and Becker [20] and Ballard *et al.* [21] and Jiang and Chapman (LF) [12], [13]. Interestingly, they are the only ones that predict a decrease of sound speed within the sediment layer. This is a very important point, illustrating the importance of detailed seabed sound-speed profile to predict low and midfrequency acoustical field in shallow water. On the other hand, one can see that Bonnel's model [15], [16] has the worst median distance, which may be linked to the fact that the inversion method by Bonnel *et al.* used a single sensor, a relatively narrow frequency band, and did not estimate attenuation. On the other hand, Duan *et al.*'s model [17] is always better than Bonnel *et al.*'s. This is an interesting result, as Duan *et al.*'s method is an improvement of Bonnel *et al.*'s method: Duan *et al.* used a wider frequency band (obtained by source deconvolution), a full VLA (instead of a single sensor) and inverted for attenuation. Here, inversion performance is clearly related to the inversion method complexity. Last but not least, it is noteworthy that Huang *et al.*'s model [14] is the only one with a sediment sound speed that increases within the first 20 m (i.e., above the R-reflector). This feature impacts the sound field prediction at all frequencies, and is reflected here by a large median distance and the largest total variation distance.

However, the intermodel variability of the total variation distance after averaging over range and frequency is very small. The results lie between 0.50 and 0.53. This must be compared with the range variability of the metric at a given frequency (see Fig. 12), with results between 0.40 and 0.65. Overall, the performance degradation with range seems much greater than the performance variation between models.

A final detail of interest is the fact that the models by Ballard and Becker [20] and Ballard *et al.* [21] and Jiang and Chapman [12], [13] are the ones with the best performance from the median distance point of view, while the model by Duan *et al.* [17] outperforms them from the total variation distance point of view (although the difference is very small). On the one hand, it is comforting to note that Ballard *et al.*'s and Jiang and Chapman's models have a similar behavior, since they are the only ones to predict a low sound speed at depth, within the sediment layer. On the other hand, it is interesting to note that no model is significantly better than all the others. The performance of geoacoustic models depends on the metrics that is used, and one must define/use metrics adapted to one's goals.

VII. CONCLUSION

This article describes the first experimental benchmark of geoacoustic inversion methods. The article demonstrates that the SW06 data set is appropriate for benchmarking inversion methods, because extensive knowledge of the ground truth is available, and also because the extensive acoustical data acquired in the experiment enables analysis of the benchmark from a sonar performance point of view. As stated in the study, the performance of geoacoustic models depends on the metrics that are used, and one must define/use metrics adapted to the goals. Here, we have introduced several metrics to assess the performance. First, a wave number metric is defined to quantify differences by estimating a distance between models. Second, two metrics based on TL statistics are defined to compare

predicted TL with experimental TL from a sonar performance point of view. Overall, the benchmark analysis shows that no geoacoustic inversion method is definitely better than any other one.

All the inversion methods under study provided an estimated sound-speed profile in the seabed, which is obviously the most important seabed parameter. Although all the estimated sound-speed profiles are different, they are all relatively effective in predicting the acoustical field at long range. The benchmark illustrated that the average sound speed in the sediment layer is an important parameter, and most models in the study have a relatively similar average sound speed to depths of 20–25 m below the seafloor. However, the benchmark also showed that details like depth variation are also desirable. As an example, the models by Ballard *et al.* and Jiang and Chapman are the only ones with a slow sound speed at the base of the sediment layer: they are also close in terms of wave number distance, and outperform all the other models in terms of median distance.

The benchmark illustrated the importance of attenuation to predict the acoustical field. This is not a surprise, but it remains a critical point to make in a geoacoustic inversion context. Only a few of the inversion methods estimated attenuation in the sediment. Indeed, many methods (particularly those based on modal/ray travel times, or real parts of horizontal wave numbers) are not sensitive to attenuation, and thus do not provide estimates for attenuation. If models from those methods are to be used to predict the acoustical field, an additional attenuation estimation must be performed (or an appropriate value must be picked from the literature). Preliminary analysis not presented here showed that the model performance from the median distance perspective seriously degrades depending on the value of attenuation in the sediment. As an example, if $0.2 \text{ dB}/(\text{m} \cdot \text{kHz})$ is used in the sediment layer (which is within the uncertainty of several inversions), then the TL metrics show significant increase and the considered model becomes clearly worse than the others.

The benchmark also showed that sediment density was not well estimated in the inversions. Some inversion methods (notably those based on ray travel times) are insensitive to density, and most inversion methods are only weakly sensitive to density. As a result, density is a parameter that is usually overlooked in inversion study. As an example of the impact, we arbitrarily assigned density values for two models considered in this benchmark, and some of the metrics (notably the wave number distance) were sensitive to the chosen values. We believe it is justified to arbitrarily assign values for density as part of an inversion study if the parameter is not estimated by the inversion method. Nonetheless, the chosen value should be judiciously chosen and specified, because it may impact the estimation of other geoacoustic parameters (e.g., sound speed).

Last but not least, all the inversion methods considered here assume that the seabed is a fluid medium. The effect of shear wave propagation in the sediment is thus ignored. However, nothing prevents using the benchmark method and the proposed metrics for future inversion studies that would include shear speed estimates.

Another important message in the article is the importance of correctly processing the experimental TL as part of the benchmark. This includes two difficulties. The first difficulty

is that the traditional TL definition depends on the SL, which may be unknown within a few dBs. In this article, this issue was circumvented by using the relative TL, which can be estimated from the data without knowledge of the SL. The second difficulty is the proper estimation of the received level, which is needed to compute (absolute and/or relative) TL. Although details about received level computations are omitted in most ocean acoustics papers, the choice of signal processing parameters that are used to do so may impact the result. A rationale to find good parameters is presented in Appendix A. Also, the benchmark illustrated the importance of assessing the experimental SNR to reject data with weak SNR: all data at 53 Hz for track TL2a were rejected for this study.

Current inversion methods also focus on estimating uncertainties of parameter values. Parameter uncertainty was generally ignored in this benchmark, because the SW06 inversions did not provide consistent uncertainty estimates. However, it will be particularly important to properly account for these uncertainties for future benchmark studies. We believe an uncertainty-aware (Bayesian?) benchmark is the next step. Doing so would require to revisit the SW06 data, re-run all the inversion methods and assess uncertainties. Another option would be to consider another data set, such has the Seabed Characterization Experiment (SBCEX [41]). Although it is too early for a benchmark using SBCEX, as most inversion studies are still ongoing, it is interesting to note that uncertainty estimate is a primary objective of SBCEX. Whether working on SW06, SBCEX, or any other data set, it will be of paramount importance that all the inversion methods that are benchmarked provide a consistent uncertainty proxy. As an example, a standard deviation may not be appropriate to describe a posterior probability density (as obtained with Bayesian inversion) if it is not Gaussian. In any case, the benchmark method has been designed so that uncertainty can be taken into account without further modifications. Indeed, if posterior probability distributions (or any other uncertainty proxy) of geoacoustic parameters are available, one can predict probability distributions of TL, and use the proposed TL metrics.

Overall, the benchmark analysis of geoacoustic inversion methods used in SW06 proved to be a difficult job, because the exercise mixes different experimental designs (e.g., array or single receiver, fixed or towed source, true or synthetic aperture, ...), different data processing (e.g., using raw time series, estimating cross-spectral density matrices, modal filtering, etc.) with different inversion methods (linear versus nonlinear inversion, frequentist versus Bayesian inference, exhaustive grid search versus probabilistic sampling, etc.). Another simpler option, from the benchmark perspective, would be to use a single inversion method and to benchmark the performance of various experimental designs and associated data processing. This could already be done for subsets of inversion methods from SBCEX that use the same inverse algorithm, e.g., trans-dimensional inversion [42]–[44]. There is a clear advantage of this approach since it focuses the benchmark on experimental design and data processing, without dealing with optimization/inversion issues. However, this narrows the benchmark to a subset of inversion studies; a global meta-study, as performed in this article, would be impossible.

Another important point is that inversion methods rely on a variety of experimental designs and requirements. Some inversion methods are very simple from an operational point of view (e.g., a single hydrophone and a source of opportunity) while others are much more complicated (e.g., dedicated source and receiver arrays). Also, different inversion methods require different prior knowledge to be effective, such as sound-speed profile in the water column, seabed layering, etc. For some applications (e.g., generating a high-resolution model of the seabed), performance may be of paramount importance even if the experimental cost is high. On the other hand, for other applications (e.g., low-resolution source localization), it may be better to have average performance, but simpler experimental design. In any case, informed choices must be made, and that requires a proper characterization of inversion experimental performances.

Last but not least, from the sonar performance point of view, a thorough (and uncertainty-aware) benchmark of geoacoustic inversion methods will need to take into account both the spatial and temporal variability of the acoustical field. With today's experimental designs, it seems impossible to untangle the spatial and temporal variability. Doing so will require a dedicated at-sea experiment that includes traditional source/receiver arrays, as well as acoustical coring devices to collect reliable ground truth (e.g., [45] and [46]). For low-frequencies, the best experimental design is probably a fixed source and a distributed/sparse array of receivers. The receiving array, spread over an area of interest, will allow the estimation of the spatial variability, while the temporal variability could be ruled out using the repeatability of sound transmissions provided by the fixed source. Such a design is now practical, as each element of the receiving array could be a single hydrophone autonomous unit, and inversion could be performed using single hydrophone methods which do not require the array to be synchronized.

APPENDIX A
ESTIMATION OF EXPERIMENTAL TL

As explained in the article, TL are estimated using (2), which requires the estimation of the experimental RL. To do so, the signal is divided into snapshots of length L, power spectra are then computed for each snapshot, and the narrowband $RL(f)$ is estimated at the frequency f by summing B adjacent frequency bins.

Actually, the estimated values of $RL(f)$ depend on L and B. A double tradeoff needs to be found here. Indeed, if L is too small, then there may not be enough SNR and $RL(f)$ will be overestimated. On the other hand, if L is too large, then the source significantly changes position within the snapshot, and $RL(f)$ is not representative of propagation at a given range. Moreover, for a given L, B must be large enough to cover the tone (otherwise $RL(f)$ is underestimated), and small enough not to encompass too much noise (otherwise $RL(f)$ is overestimated).

A convenient way to correctly choose is to plot $RL(f)$ as a function of B and L, and to identify the behavior previously explained. Here, this has been done for all the considered frequencies, and for at least three positions within the track (short
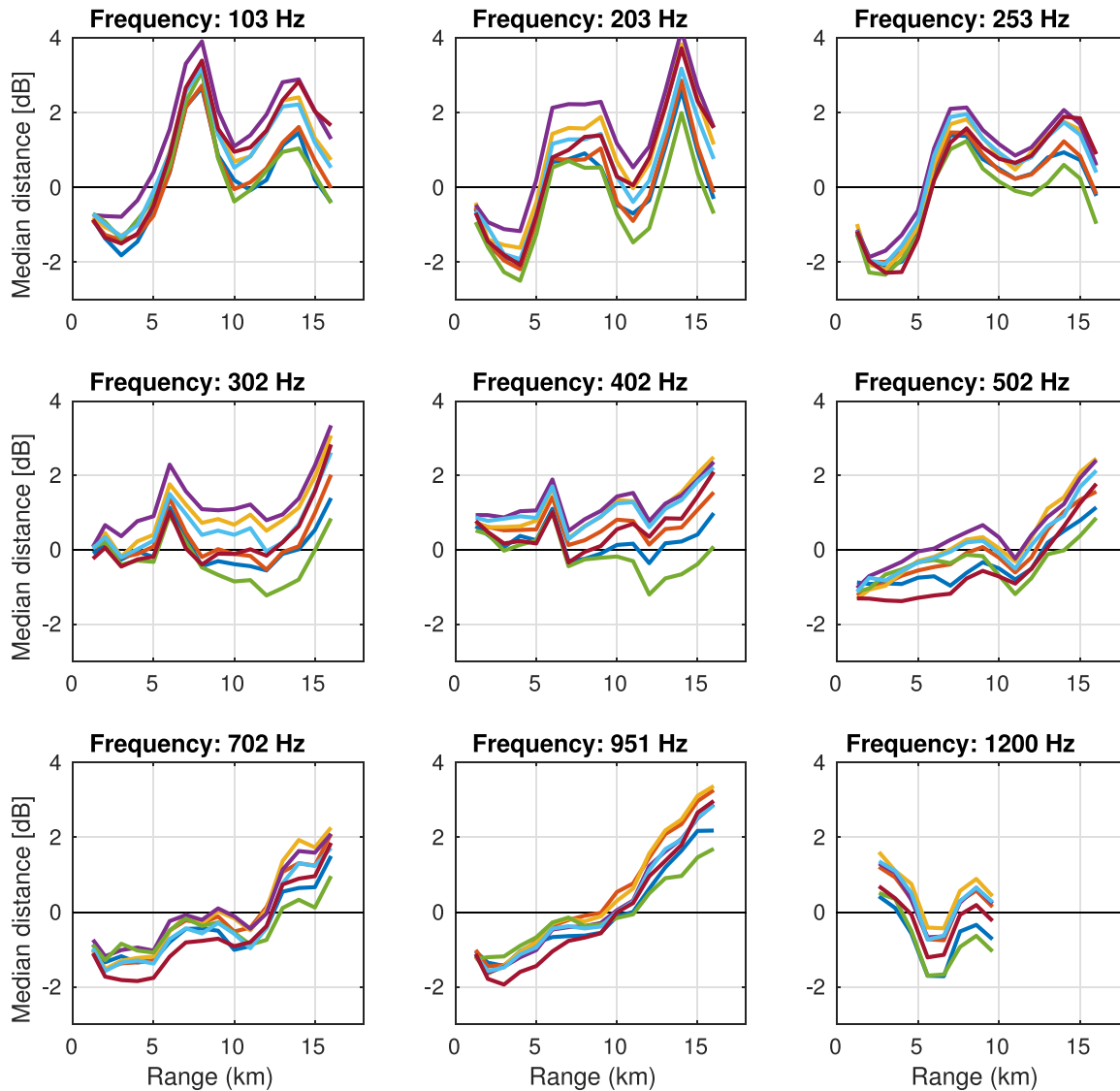
Fig. 11. Median distance $d_{\mathrm{TL}}^{\mathrm{med}}$ between experimental and simulated TL. The vertical and horizontal scales are the same on all the panels, and the color code is the same than on Fig. 6 (dark blue: Ballard *et al.*, orange: Jiang and Chapman LF, yellow: Huang *et al.*, purple: Bonnel *et al.*, green: Duan *et al.*, light blue: Jiang *et al.* MF, red: Choi *et al.*).

range, intermediate range, and large range) to cover various SNR conditions. Chosen parameters for TL2a are $L = 10$ s and $B = 5$ (i.e., 0.4 Hz), and chosen parameters for TL2b are $L = 5$ s and $B = 5$ (i.e., 0.8 Hz). Note that L is larger for TL2a to compensate for higher ambient noise at low-frequencies. Note also that there is a 50% overlap between snapshots for TL2a, and no overlap for TL2b, so that the snapshot density along the track is relatively equivalent for TL2a and TL2b.

APPENDIX B
RANGE-DEPENDENT TL METRICS

This appendix presents range-dependent TL metrics: Fig. 11 shows the median distance and Fig. 12 shows the total variation distance. Overall, the absolute values of the two distances tend to increase with range. This is a logical result, which shows

that the importance of correctly modeling the seabed increases with range.

Another point of interest is that the median distance may switch from positive to negative values (and vice versa) for a given model at a given frequency. As a result, the range averaged results, presented in Fig. 8, may be relatively small with respect to the range-dependent results (Fig. 11), because positive and negative values compensate for each other in the averaging process. Nonetheless, this averaging process allows for negative and positive values of the median distance. This in turn enables an assessment of the global attenuating behavior of each models, which is discussed in details in Section VI-B.
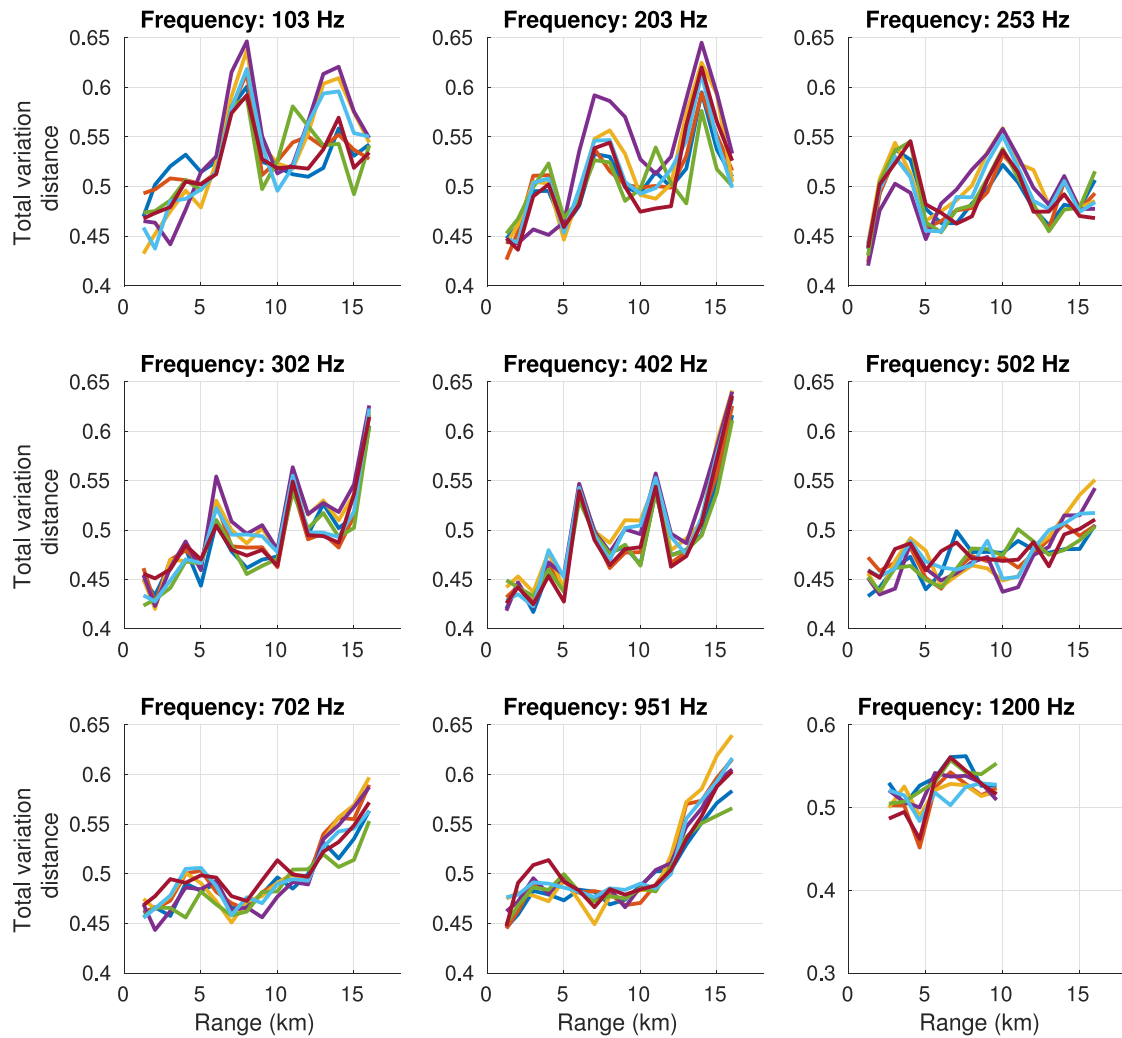
Fig. 12. Total variation distance $d_{\mathrm{TL}}^{\mathrm{TV}}$ between experimental and simulated TL. The vertical and horizontal scales are the same on all the panels, and the color code is the same than on Fig. 6 (dark blue: Ballard *et al.*, orange: Jiang and Chapman LF, yellow: Huang *et al.*, purple: Bonnel *et al.*, green: Duan *et al.*, light blue: Jiang *et al.* MF, red: Choi *et al.*)

REFERENCES

[1] A. Tolstoy, N. Chapman, and G. Brooke, "Workshop'97: Benchmarking for geoacoustic inversion in shallow water," *J. Comput. Acoust.*, vol. 6, pp. 1–28, 1998.

[2] N. Chapman, S. Chin-Bing, D. King, and R. Evans, "Benchmarking geoacoustic inversion methods for range-dependent waveguides," *IEEE J. Ocean. Eng.*, vol. 28, no. 3, pp. 320–330, Jul. 2003.

[3] D. Tang *et al.*, "Shallow Water '06: A joint acoustic propagation/nonlinear internal wave physics experiment," *Oceanography*, vol. 20, no. 4, pp. 156–167, 2007.

[4] D. Knobles, P. Wilson, J. Goff, and S. Cho, "Seabed acoustics of a sand ridge on the New Jersey continental shelf," *J. Acoust. Soc. Amer.*, vol. 124, no. 3, pp. EL151–EL156, 2008.

[5] D. P. Knobles, J. A. Goff, R. A. Koch, P. S. Wilson, and J. A. Shooter, "Effect of inhomogeneous sub-bottom layering on broadband acoustic propagation," *IEEE J. Ocean. Eng.*, vol. 35, no. 4, pp. 732–743, Oct. 2010.

[6] S. Stotts, R. Koch, S. Joshi, V. Nguyen, V. Ferreri, and D. Knobles, "Geoacoustic inversions of horizontal and vertical line array acoustic data from a surface ship source of opportunity," *IEEE J. Ocean. Eng.*, vol. 35, no. 1, pp. 79–102, Jan. 2010.

[7] L. Wan, M. Badiey, and D. P. Knobles, "Geoacoustic inversion using low frequency broadband acoustic measurements from L-shaped arrays in the Shallow Water 2006 experiment," *J. Acoust. Soc. Amer.*, vol. 140, no. 4, pp. 2358–2373, 2016.

[8] J. Yang, D. Tang, and K. Williams, "Direct measurement of sediment sound speed in Shallow Water'06," *J. Acoust. Soc. Amer.*, vol. 124, no. 3, pp. EL116–EL121, 2008.

[9] A. Turgut and T. Yamamoto, "In situ measurements of velocity dispersion and attenuation in New Jersey shelf sediments," *J. Acoust. Soc. Amer.*, vol. 124, no. 3, pp. EL122–EL127, 2008.

[10] J. W. Choi, P. H. Dahl, and J. A. Goff, "Observations of the r reflector and sediment interface reflection at the Shallow Water '06 central site," *J. Acoust. Soc. Amer.*, vol. 124, no. 3, pp. EL128–EL134, 2008.

[11] J. Yang, D. R. Jackson, and D. Tang, "Mid-frequency geoacoustic inversion using bottom loss data from the Shallow Water 2006 experiment," *J. Acoust. Soc. Amer.*, vol. 131, no. 2, pp. 1711–1721, 2012.

[12] Y.-M. Jiang and N. R. Chapman, "Bayesian geoacoustic inversion in a dynamic shallow water environment," *J. Acoust. Soc. Amer.*, vol. 123, no. 6, pp. EL155–EL161, 2008.

[13] Y.-M. Jiang and N. R. Chapman, "The impact of ocean sound speed variability on the uncertainty of geoacoustic parameter estimates," *J. Acoust. Soc. Amer.*, vol. 125, pp. 2881–2895, 2009.

[14] C. Huang, P. Gerstoft, and W. Hodgkiss, "Effect of ocean sound speed uncertainty on matched-field geoacoustic inversion," *J. Acoust. Soc. Amer.*, vol. 123, no. 6, pp. EL162–EL168, 2008.

[15] J. Bonnel and N. Chapman, "Geoacoustic inversion in a dispersive waveguide using warping operators," *J. Acoust. Soc. Amer.*, vol. 130, pp. EL101–EL107, 2011.

[16] J. Bonnel, S. Dosso, and R. Chapman, "Bayesian geoacoustic inversion of single hydrophone light bulb data using warping dispersion analysis," *J. Acoust. Soc. Amer.*, vol. 134, pp. 120–130, 2013.

[17] R. Duan, N. R. Chapman, K. Yang, and Y. Ma, "Sequential inversion of modal data for sound attenuation in sediment at the New Jersey shelf," *J. Acoust. Soc. Amer.*, vol. 139, no. 1, pp. 70–84, 2016.

[18] Y.-M. Jiang, N. R. Chapman, and P. Gerstoft, "Short range travel time geoacoustic inversion with vertical line array," *J. Acoust. Soc. Amer.*, vol. 124, no. 3, pp. EL135–EL140, 2008.

[19] Y.-M. Jiang and N. R. Chapman, "Measurement of low-frequency sound attenuation in marine sediment," *IEEE J. Ocean. Eng.*, vol. 35, no. 1, pp. 70–78, Jan. 2010.

[20] M. S. Ballard and K. M. Becker, "Geoacoustic inversion on the new jersey margin: Along and across the shelf," *J. Acoust. Soc. Amer.*, vol. 124, no. 3, pp. EL141–EL145, 2008.

[21] M. Ballard, K. Becker, and J. Goff, "Geoacoustic inversion for the new jersey shelf: 3-D sediment model," *IEEE J. Ocean. Eng.*, vol. 35, no. 1, pp. 28–42, Jan. 2010.

[22] S. Rajan and K. Becker, "Inversion for range-dependent sediment compressional-wave-speed profiles from modal dispersion data," *IEEE J. Ocean. Eng.*, vol. 35, no. 1, pp. 43–58, Jan. 2010.

[23] A. Turgut, "Wideband attenuation measurements in New Jersey shelf sediments," *J. Acoust. Soc. Amer.*, vol. 124, no. 4, pp. 2468–2468, 2008.

[24] J. A. Goff *et al.*, "Seabed characterization on the New Jersey middle and outer shelf: Correlatability and spatial variability of seafloor sediment properties," *Mar. Geol.*, vol. 209, no. 1, pp. 147–172, 2004.

[25] J. C. Hathaway *et al.*, "Preliminary summary of the 1976 Atlantic Margin Coring project of the US geological survey," U.S. Geological Surv., Reston, VA, USA, Open-File Rep. 76-844, 1976.

[26] R. Cederberg, W. Siegmann, and W. Carey, "Influence of geoacoustic modeling on predictability of low-frequency propagation in range-dependent, shallow-water environments," *J. Acoust. Soc. Amer.*, vol. 97, no. 5, pp. 2754–2766, 1995.

[27] R. J. Cederberg, W. M. Carey, and W. L. Siegmann, "Modal analysis of geoacoustic influences on shallow-water propagation," *IEEE J. Ocean. Eng.*, vol. 22, no. 2, pp. 237–244, Apr. 1997.

[28] S. P. Gulick, J. A. Goff, J. A. Austin Jr., C. R. Alexander Jr., S. Nordfjord, and C. S. Fulthorpe, "Basal inflection-controlled shelf-edge wedges off New Jersey track sea-level fall," *Geology*, vol. 33, no. 5, pp. 429–432, 2005.

[29] T. A. Davies, J. A. Austin Jr., M. B. Lagoe, and J. D. Milliman, "Late quaternary sedimentation off New Jersey: New results using 3-D seismic profiles and cores," *Mar. Geol.*, vol. 108, no. 3/4, pp. 323–343, 1992.

[30] M. B. Porter, "The KRAKEN normal mode program," Naval Res. Lab., Washington, DC, USA, DTIC Document, Tech.Rep. NRL/MR/5120-92-6920, 1992.

[31] J. Sagers and D. Knobles, "SW06 source level estimation for three specific tow tracks: TL1a, TL1b, ITC," Appl. Res. Lab., Univ. Texas at Austin, Austin, TX, USA, ARL TL-EV-14-69, 2014.

[32] D. Weston, "Intensity-range relations in oceanographic acoustics," *J. Sound Vib.*, vol. 18, no. 2, pp. 271–287, 1971.

[33] P. Abbot and I. Dyer, "Sonar performance predictions incorporating environmental variability," in *Impact of Littoral Environmental Variability of Acoustic Predictions and Sonar Performance*. Springer, 2002, pp. 611–618.

[34] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *Int. Statist. Rev.*, vol. 70, no. 3, pp. 419–435, 2002.

[35] M. D. Collins, "Users guide for RAM versions 1.0 and 1.0 p," Naval Res. Lab, Washington, DC, USA, vol. 20375, 1995.

[36] S. P. Pecknold, K. W. Masui, and P. C. Hines, "Transmission loss measurements and geoacoustic sensitivity modeling at 1.2 khz," *J. Acoust. Soc. Amer.*, vol. 124, no. 3, pp. EL110–EL115, 2008.

[37] Z. Michalopoulou, "Stochastic sediment property inversion in Shallow Water 06," *J. Acoust. Soc. Amer.*, vol. 142, no. 5, pp. EL495–EL499, 2017.

[38] M. Taroudakis, C. Smaragdakis, and N. R. Chapman, "Inversion of acoustical data from the Shallow Water 06 experiment by statistical signal characterization," *J. Acoust. Soc. Amer.*, vol. 136, no. 4, pp. EL336–EL342, 2014.

[39] M. Taroudakis, C. Smaragdakis, and N. R. Chapman, "Denoising underwater acoustic signals for applications in acoustical oceanography," *J. Comput. Acoust.*, vol. 25, no. 2, pp. 1750015-1–1750015-23, 2017, doi: 10.1142/S0218396X17500151.

[40] J.-X. Zhou, X.-Z. Zhang, and D. Knobles, "Low-frequency geoacoustic model for the effective properties of sandy seabottoms," *J. Acoust. Soc. Amer.*, vol. 125, no. 5, pp. 2847–2866, 2009.

[41] D. Knobles and P. Wilson, "SBCEX: Seabed characterization experiment," 2017. [Online]. Available: https://wwwext.arlut.utexas.edu/sbcex/. Accessed: Jan. 2020.

[42] J. Bonnel, S. Dosso, D. Eleftherakis, and R. Chapman, "Trans-dimensional inversion of modal dispersion data on the New England mud patch," *IEEE J. Ocean. Eng.*, to be published, doi: 10.1109/JOE.2019.2896389.

[43] J. Belcourt, C. W. Holland, S. E. Dosso, J. Dettmer, and J. A. Goff, "Depth-dependent geoacoustic inferences with dispersion at the new england mud patch via reflection coefficient inversion," *IEEE J. Ocean. Eng.*, to be published, doi: 10.1109/JOE.2019.2900115.

[44] D. Tollefsen, S. E. Dosso, and D. Knobles, "Ship-of-opportunity noise inversions for geoacoustic profiles of a layered mud-sand seabed," *IEEE J. Ocean. Eng.*, to be published, doi: 10.1109/JOE.2019.2896389.

[45] J. Yang and D. Tang, "Direct measurements of sediment sound speed and attenuation in the frequency band of 2–8 kHz at the target and reverberation experiment site," *IEEE J. Ocean. Eng.*, vol. 42, no. 4, pp. 1102–1109, Oct. 2017.

[46] M. S. Ballard *et al.*, "In situ measurements of compressional wave speed during gravity coring operations in the new england mud patch," *IEEE J. Ocean. Eng.*, to be published, doi: 10.1109/JOE.2019.2924560.

Julien Bonnel (S'08–M'11) received the Ph.D. degree in signal processing from Grenoble Institut National Polytechnique, Grenoble, France, in 2010.

From 2010 to 2017, he was an Assistant/Associate Professor with Laboratoire des Sciences et Technologies de l'Information, de la Communication et de la Connaissance (Lab-STICC, UMR CNRS 6285), Ecole Nationale Suprieure de Techniques Avances de Bretagne, ENSTA Bretagne, Brest, France. Since September 2017, he has been an Associate Scientist with the Woods Hole Ocenographic Institution, Woods Hole, MA, USA. His research in signal processing and underwater acoustics includes time-frequency analysis, source detection/localization, geoacoustic inversion, acoustical tomography, passive acoustic monitoring, and bioacoustics.

Dr. Bonnel was the recipient of the 2019 A.B. Wood Medal for his distinguished contributions to the application of underwater acoustics.

Sean P. Pecknold received the Ph.D. degree in physics from McGill University, Montral, QC, Canada, in 2000.

He is currently a Defence Scientist with the Underwater Sensing Section, Defence Research and Development Canada's Atlantic Research Centre, Dartmouth, NS, Canada, where his research interests include acoustical propagation uncertainty and geoacoustic parameter estimation.

Dr. Pecknold is a member of the Canadian Acoustical Association and the Acoustical Society of America.

Paul C. Hines was born in Glace Bay, NS, Canada. He received the B.Sc. (honors) in engineering-physics from Dalhousie University, Halifax, NS, Canada, in 1981, and the Ph.D. degree in physics from University of Bath, Bath, U.K, in 1988.

In 1981, he joined Defence Research and Development Canada (DRDC), Dartmouth, NS, Canada. From 1985 to 1988, he was with the University of Bath. From his return to DRDC in 1988 until his departure in March 2014, he led several groups and managed a variety of acoustical research projects for both DRDC and the U.S. Office of Naval Research. He is a Seasoned Experimentalist and has been the Chief Scientist for several collaborative international research trials. He is currently the President of Hines Ocean S&T, Inc., Dartmouth, NS, Canada, in addition to conducting research at Dalhousie University in the Department of Electrical and Computer Engineering and Department of Oceanography. His research interests include Anti-Submarine Warfare, mine and torpedo countermeasures, rapid environmental assessment, acoustical scattering, sound-speed dispersion, vector sensor processing, sonar classification and tracking, continuous active sonar, and the application of aural perception in humans, target classification in sonar.

Dr. Hines is a Fellow of the Acoustical Society of America, and a Distinguished Lecturer of the IEEE Ocean Engineering Society. He received the Chesterman Medal from the University for Outstanding Research in Physics for his research on acoustic scattering from ocean boundaries.
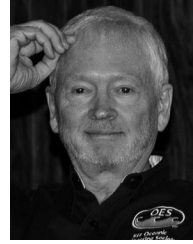
N. Ross Chapman (M'97–SM'03–F'10) received the Ph.D. degree in physics from the University of British Columbia in 1975, and started work in Defence Research in Canada as Leader of the Ocean Acoustics Group (1976–1995), before moving on to the University of Victoria as a Research Chair in Ocean Acoustics, and Director of the Centre for Earth and Ocean Sciences at the University of Victoria (2000–2006). He recently retired from the University of Victoria following a career of over 35 years in underwater acoustics, signal processing, and acoustical oceanography. He has authored or coauthored over 130 papers and book chapters, and continues to maintain his research interests in ocean acoustical propagation and modeling, advanced signal processing, ocean ambient noise measurement and modeling, and inversion of ocean seabed properties from acoustical field data.

Dr. Chapman is an elected fellow of the Acoustical Society of America, and served as the Editor-in-Chief for the IEEE JOURNAL OF OCEANIC ENGINEERING from 2012 to 2017.