

# TOWARDS ROBUST DATA-DRIVEN UNDERWATER ACOUSTIC LOCALIZATION: A DEEP CNN SOLUTION WITH PERFORMANCE GUARANTEES FOR MODEL MISMATCH

*Amir Weiss\**, *Andrew C. Singer*<sup>†</sup>, and *Gregory W. Wornell\**

\*Research Laboratory of Electronics  
Massachusetts Institute of Technology  
{amirwei,gww}@mit.edu

<sup>†</sup>Dept. of Electrical and Computer Engineering  
University of Illinois Urbana-Champaign  
acsinger@illinois.edu

## ABSTRACT

Key challenges in developing underwater acoustic localization methods are related to the combined effects of high reverberation in intricate environments. To address such challenges, recent studies have shown that with a properly designed architecture, neural networks can lead to unprecedented localization capabilities and enhanced accuracy. However, the robustness of such methods to *environmental mismatch* is typically hard to characterize, and is usually assessed only empirically. In this work, we consider the recently proposed data-driven method [18] based on a deep convolutional neural network, and demonstrate that it can learn to localize in complex and *mismatched* environments. To explain this robustness, we provide an upper bound on the localization mean squared error (MSE) in the “true” environment, in terms of the MSE in a “presumed” environment and an additional penalty term related to the environmental discrepancy. Our theoretical results are corroborated via simulation results in a rich, highly reverberant, and mismatch channel.

**Index Terms**— Localization, reverberation, model mismatch.

## 1. INTRODUCTION

Localization in the underwater acoustic (UWA) domain, which is relevant for a variety of important applications [1–4], is challenging due to the environment’s physical properties. In particular, the channel between an acoustic source and a receiver may vary dramatically with their positions. While the related literature is abundant with handcrafted model-based solutions [5], many of these methods rely on isovelocity line-of-sight-based geometry, and are thus sensitive to multipath. Other methods that can theoretically operate in nontrivial environments, such as matched field processing [6], are highly sensitive to model mismatch [7], and are therefore less relevant for some practical purposes.

Due to the success of deep neural networks (DNN) in other domains, data-driven (DD) methods have been recently considered more extensively as potentially viable solutions to the UWA localization problem [8–16]. Specifically, a DD variant of the statistically superior direct localization (DLOC) approach [17, 18] has been proposed in [19]. Yet, while these DD methods are in principle exceptionally powerful, they can be difficult to analyze, and their robustness is generally not guaranteed.

This work is focused on the robustness of a DD-DLOC solution to UWA localization, and its main contribution is an upper bound on the localization mean squared error (MSE) in a mismatched environment. To better contextualize and elaborate on our contributions (listed at the end of Section 2), let us first describe in detail the problem at hand.

## 2. PROBLEM FORMULATION

Consider  $L$  spatially-diverse, time-synchronized receivers at known locations, each consisting of a single omni-directional sensor. Furthermore,

This work was supported, in part, by ONR under Grants No. N00014-19-1-2665 and No. N00014-19-1-2662, and NSF under Grant No. CCF-1816209.

consider the presence of a source, emitting an unknown waveform from an unknown position, denoted by the vector of coordinates  $\mathbf{p} \in \mathcal{V} \subset \mathbb{R}^{3 \times 1}$ , where  $\mathcal{V}$  is a volume of interest. We assume that the source is static during the observation time, and is located sufficiently far from all  $L$  receivers to permit a planar wavefront (far-field) approximation.<sup>1</sup>

In the underwater environment, “reflectors” at unknown positions are potentially present, giving rise to a (possibly rich) multipath channel impulse response (CIR), which can be approximately described using ray propagation (e.g., [20]). For example, the water surface acts as one of the reflectors. In such a model, the acoustic wave that is emitted from the source and measured at the receiver is represented as a sum of (possibly infinitely many) rays, each propagating according to physical properties of the environment (e.g., temperature, bathymetry, etc.). For simplicity, we assume the existence of a set of the environmental parameters, denoted by  $\mathcal{P}_{\text{env}}$ , with which the propagation model is fully characterized.

Formally, the received baseband signal of the  $\ell$ -th receiver is given by

$$x_\ell[n] = \sum_{r=1}^R b_{r\ell} s_{r\ell}[n] + v_\ell[n] \triangleq \mathbf{s}_\ell^T[n] \mathbf{b}_\ell + v_\ell[n] \in \mathbb{C}, \quad (1)$$

$$n = 1, \dots, N, \quad \forall \ell \in \{1, \dots, L\},$$

where we have defined  $\mathbf{s}_\ell[n] = [s_{1\ell}[n] \cdots s_{R\ell}[n]]^T \in \mathbb{C}^{R \times 1}$  and  $\mathbf{b}_\ell = [b_{1\ell} \cdots b_{R\ell}]^T \in \mathbb{C}^{R \times 1}$ , using the notation:

1.  $b_{r\ell} \in \mathbb{C}$  as the unknown attenuation coefficient<sup>2</sup> from the source to the  $\ell$ -th sensor associated with the  $r$ -th signal component (line-of-sight (LOS) or other non-LOS (NLOS) reflections);
2.  $s_{r\ell}[n] \triangleq s(t - \tau_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}}))|_{t=nT_s} \in \mathbb{C}$  as the sampled  $r$ -th component of the unknown signal’s waveform at the  $\ell$ -th sensor, where  $s(t - \tau_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}}))$  is the analog, continuous-time waveform delayed by  $\tau_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})$ , and  $T_s$  is the sampling period; and
3.  $v_\ell[n] \in \mathbb{C}$  as the additive, ambient and internal receiver, noise at the  $\ell$ -th receiver, modeled as a zero-mean random process (not necessarily Gaussian) with an unknown finite variance  $\sigma_{v_\ell}^2$ .

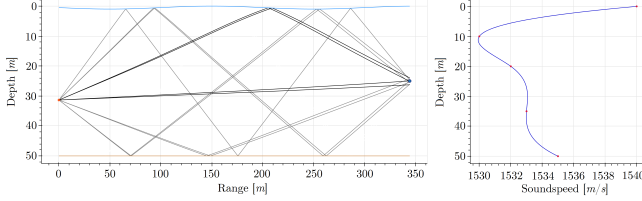
Applying the normalized DFT<sup>3</sup> to (1) yields the equivalent frequency-domain representation for all  $\ell \in \{1, \dots, L\}$ ,

$$\begin{aligned} \bar{x}_\ell[k] &= \sum_{r=1}^R b_{r\ell} \bar{s}[k] e^{-j\omega_k \tau_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})} + \bar{v}_\ell[k] \\ &= \bar{s}[k] \cdot \underbrace{\mathbf{d}_\ell^H[k] \mathbf{b}_\ell}_{\text{a function of } \mathbf{p} \text{ and } \mathcal{P}_{\text{env}}} + \bar{v}_\ell[k] = \bar{s}[k] \cdot \underbrace{\bar{h}_\ell[k]}_{\text{frequency response}} + \bar{v}_\ell[k] \in \mathbb{C}, \quad (2) \end{aligned}$$

<sup>1</sup>This becomes reasonably accurate, e.g., in shallow waters at high frequencies.

<sup>2</sup>In fact, these coefficients are also a function of the source’s position  $\mathbf{p}$  and the environment  $\mathcal{P}_{\text{env}}$ , but we mostly use  $b_{r\ell}$  rather than  $b_{r\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})$  for brevity.

<sup>3</sup> $\bar{\mathbf{z}}$  denotes the normalized discrete Fourier transform (DFT) of  $\mathbf{z}$ .



**Fig. 1.** Illustration of the simulated environment using Bellhop [21]. Left: Ray propagation paths from the source (red) to a receiver (blue). Right: The assumed depth-varying speed of sound profile.

where we have defined  $\bar{h}_\ell[k] \triangleq \mathbf{d}_\ell^H[k] \mathbf{b}_\ell \in \mathbb{C}$ , and

$$\mathbf{d}_\ell[k] \triangleq [e^{-j\omega_k \tau_{1\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})} \dots e^{-j\omega_k \tau_{R\ell}(\mathbf{p}, \mathcal{P}_{\text{env}})}]^H \in \mathbb{C}^{R \times 1},$$

with  $\{\omega_k \triangleq \frac{2\pi(k-1)}{NT_s}\}_{k=1}^N$ . As shorthand notation, we further define

$$\begin{aligned} \bar{\mathbf{x}}_\ell &\triangleq [\bar{x}_\ell[1] \dots \bar{x}_\ell[N]]^T, \quad \bar{\mathbf{s}} \triangleq [\bar{s}[1] \dots \bar{s}[N]]^T, \\ \bar{\mathbf{v}}_\ell &\triangleq [\bar{v}_\ell[1] \dots \bar{v}_\ell[N]]^T, \quad \mathbf{D}_\ell \triangleq [\mathbf{d}_\ell[1] \dots \mathbf{d}_\ell[N]]^T \in \mathbb{C}^{N \times R}, \\ \bar{\mathbf{X}}_\ell &\triangleq \text{Diag}(\bar{\mathbf{x}}_\ell), \quad \bar{\mathbf{S}} \triangleq \text{Diag}(\bar{\mathbf{s}}), \quad \mathbf{H}_\ell(\mathbf{p}, \mathcal{P}_{\text{env}}) \triangleq \text{Diag}(\mathbf{D}_\ell \mathbf{b}_\ell), \end{aligned}$$

where  $\text{Diag}(\cdot)$  forms a diagonal matrix from its vector argument. Note that  $\mathbf{H}_\ell(\mathbf{p}, \mathcal{P}_{\text{env}})$  is generally a nonlinear function of the unknown source position  $\mathbf{p}$  and the environmental parameters  $\mathcal{P}_{\text{env}}$  (e.g., as seen from (2)). With this notation, we may now write (2) compactly as

$$\bar{\mathbf{x}}_\ell = \mathbf{H}_\ell(\mathbf{p}, \mathcal{P}_{\text{env}}) \bar{\mathbf{s}} + \bar{\mathbf{v}}_\ell \in \mathbb{C}^{N \times 1}, \quad \forall \ell \in \{1, \dots, L\}. \quad (3)$$

Given the possibly (highly) intricate relation between  $\mathbf{p}$  and  $\bar{\mathbf{x}}_\ell$ , the complexity of the localization problem, which consists of estimating  $\mathbf{p}$  from the observations  $\{\bar{\mathbf{x}}_\ell\}_{\ell=1}^L$ , is now readily apparent from (3).

To further illustrate the potentially formidable computational task of localizing a source via an exact model-based solution in an environment with nontrivial propagation properties, consider the UWA environment illustrated in Fig. 1. Given a source-receiver pair, the relevant rays propagating from the source to the receiver can be described in a plane perpendicular to an ideal flat bottom, in which the source and the receiver lie. As an example, the complexity of the method [22] in its extended form [19] is  $\Omega(NL^2R^2)$  even before considering that all the  $LR$  time-delays of all the rays to all the receivers must be computed as well.<sup>4</sup>

It is therefore natural to adopt a learning-based solution approach, that will provide a method to localize the source while: (i) judiciously using those environmental characteristics that can be learned during the training phase; (ii) being robust to reasonably moderate deviations from the learned model; and (iii) efficiently approximating a reliable estimator  $\hat{\mathbf{p}} = f(\bar{\mathbf{x}}_1, \dots, \bar{\mathbf{x}}_L)$  that is otherwise computationally infeasible.

The remainder of this paper is devoted to the study of the desired merits of a learning-based solution as mentioned above. Specifically,

- We demonstrate that a DD approach can yield a computationally efficient DLOC solution that is “stable”—in the sense of its MSE performance—with respect to model mismatch deviations.
- We provide analytical guarantees on the MSE localization performance of an estimator designed for one environment, when applied to data from a another environment (i.e., model mismatch).

### 3. LEARNING TO LOCALIZE

Let us assume that we do not have precise knowledge of the environment  $\mathcal{P}_{\text{env}}$  or the propagation model, based on which  $\mathbf{H}_\ell(\mathbf{p}, \mathcal{P}_{\text{env}})$  is computed for any  $\mathbf{p}$ . Rather, assume that a dataset of received signals with labeled source positions  $\mathcal{D}_{\text{train}}^{(J)} \triangleq \{(\bar{\mathbf{x}}_1^{(i)}, \dots, \bar{\mathbf{x}}_L^{(i)}, \mathbf{p}^{(i)})\}_{i=1}^J$  of size  $J$  is available. Such a dataset can be obtained in several ways (see [19] for details).

<sup>4</sup>For environments such as the one in Fig. 1, this may already be infeasible.

### 3.1. A Data-Driven Direct Localization Method

In a recent work [19], an architecture based on a deep convolutional neural network (CNN) has been proposed as a DD-DLOC solution for the UWA localization problem. This architecture is trained in two phases. In the first phase, three “sub-models” are trained to estimate range, azimuth and inclination. In the second phase, these three models are fused into a “global model”, that provides an (enhanced) estimate of the source position. The successful operation of this solution has been demonstrated for a 3-ray isovelocity propagation model, which theoretically allows for accurate DLOC even in the absence of LOS (see [22] for simulation and experimental results). While this is already a powerful capability, in this work we demonstrate that this solution generalizes to even richer environments, which are “closer”—in some well-defined sense—to the real physical medium in which UWA localization systems operate.

To this end, in this work we consider a simulated UWA environment as our virtual testbed, and adopt the architecture proposed in [19], to which we introduce a two main modifications. First, we now directly concatenate the outputs of the three sub-models, such that the modified architecture already provides a working solution at the end of the first phase of training (mentioned above). Second, during training the CNN-based solution is further matched to the attenuation law of the channel, by normalizing the signal power with the average attenuation magnitude in the volume of interest. For more implementation details of this architecture, see the supplementary material for this paper, [23].

To go beyond the 3-ray model [22], we generate the CIRs using the Bellhop simulator [21]. This way, the propagation model can be made exceptionally rich, taking into account nontrivial effects such as “bending” rays due to depth-varying speed of sound, different bathymetry and surface geometries, etc. Thus, the simulated received signals are more similar to true recorded signals in a real environment. For more details, see [23].

As we demonstrate in Section 4.2, our architecture is able to capture the environment’s characteristics relevant for inference regarding the source position. While this is a significant part of the contribution of this work, we nonetheless choose to elaborate on yet another important aspect—robustness to model mismatch. Although a carefully designed simulator (such as Bellhop) can provide good approximations of the real CIRs, it is still only an *approximation* that naturally suffers (to some extent) from modeling error relative to the respective true CIRs as well as unmodelled effects. For example, a simulator may provide a CIR with a finite number of rays, when in practice there may be an infinite number (with finite energy). Therefore, and especially given that analytical characterization of the output of DNNs is generally a hard task, we next turn to the main discussion of this paper—an upper bound on the MSE of the proposed CNN-based solution under model mismatch.

### 4. PERFORMANCE GUARANTEES UNDER A MISMATCHED PROPAGATION MODEL

For simplicity of the exposition, we assume hereafter that  $\sigma_{v_\ell}^2 = \sigma_v^2$  for all  $\ell \in \{1, \dots, L\}$ . However, this assumption can be relaxed, and with the necessary slight modifications, the results below are still valid.

The signal model (1) more faithfully characterizes the true underlying channel, e.g., when  $R \rightarrow \infty$ , in which case  $\sum_{r=1}^{\infty} |b_{r\ell}|^2 < \infty$  for all  $\ell \in \{1, \dots, L\}$ , when the power of the received signal is finite. However, practical localization solutions would take into account only a finite number of NLOS components (if, at all), and would therefore suffer from model mismatch. Nevertheless, if  $R$  is chosen to be sufficiently large, such that most of the energy of the received signal is captured by the model, it is reasonable to expect that the devised localization method would still perform properly, with some slight performance degradation relative to the expected performance in the absence of model mismatch. Below we provide a rigorous proof that this intuition is indeed correct.

Moreover, we quantify “which portions” of the channel must be modeled in order to still maintain MSE performance guarantees.

#### 4.1. Upper Bounds on the MSE for Mismatched Environment

To facilitate the derivations below, we introduce a compact representation of (3). Let  $\bar{\mathbf{x}} \triangleq [\bar{\mathbf{x}}_1^T \dots \bar{\mathbf{x}}_L^T]^T \in \mathbb{C}^{NL \times 1}$ , with which (3) reads

$$\bar{\mathbf{x}} = \mathbf{H}(\mathbf{p}, \mathcal{P}_{\text{env}})\bar{\mathbf{s}} + \bar{\mathbf{v}} \in \mathbb{C}^{NL \times 1}, \quad (4)$$

where  $\mathbf{H}(\mathbf{p}, \mathcal{P}_{\text{env}}) \triangleq [\mathbf{H}_1^T(\mathbf{p}, \mathcal{P}_{\text{env}}) \dots \mathbf{H}_L^T(\mathbf{p}, \mathcal{P}_{\text{env}})]^T \in \mathbb{C}^{NL \times N}$  and  $\bar{\mathbf{v}} \triangleq [\bar{\mathbf{v}}_1^T \dots \bar{\mathbf{v}}_L^T]^T$ . Further, let  $Q \triangleq \mathcal{CN}(\mathbf{0}, \Sigma_Q)$ ,  $P \triangleq \mathcal{CN}(\mathbf{0}, \Sigma_P)$  be the distributions of the presumed and the true<sup>5</sup> models for (4) with environmental parameters  $\mathcal{Q}_{\text{env}}, \mathcal{P}_{\text{env}}$ , respectively, where

$$\begin{aligned} \Sigma_Q &\triangleq \mathbb{E}_Q [\bar{\mathbf{x}}\bar{\mathbf{x}}^H] = \sigma_s^2 \mathbf{H}(\mathbf{p}, \mathcal{Q}_{\text{env}})\mathbf{H}^H(\mathbf{p}, \mathcal{Q}_{\text{env}}) + \sigma_v^2 \mathbf{I}_{NL}, \\ \Sigma_P &\triangleq \mathbb{E}_P [\bar{\mathbf{x}}\bar{\mathbf{x}}^H] = \sigma_s^2 \mathbf{H}(\mathbf{p}, \mathcal{P}_{\text{env}})\mathbf{H}^H(\mathbf{p}, \mathcal{P}_{\text{env}}) + \sigma_v^2 \mathbf{I}_{NL}. \end{aligned} \quad (5)$$

Let  $\hat{\mathbf{p}}: \bar{\mathbf{x}} \rightarrow \mathbb{R}^{K_P \times 1}$  be a location estimator of  $\mathbf{p}$ , and define  $\boldsymbol{\varepsilon}_{\mathbf{p}} \triangleq \hat{\mathbf{p}} - \mathbf{p}$ , its respective estimation error vector. We have the following result:

**Proposition 1.** Denote  $Q_\varepsilon$  and  $P_\varepsilon$  as the distributions of  $\boldsymbol{\varepsilon}_{\mathbf{p}}$  when  $\bar{\mathbf{x}}$  is distributed according to  $Q$  and  $P$ , respectively. Then the MSE of  $\hat{\mathbf{p}}$  evaluated for the true model  $P$  is upper bounded by

$$\mathbb{E}_P [\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2] \leq \mathbb{E}_Q [\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2] + \delta(P_\varepsilon, Q_\varepsilon), \quad (6)$$

where  $\delta(P_\varepsilon, Q_\varepsilon) \triangleq \sqrt{\text{Var}_Q (\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2) \chi^2(P_\varepsilon \| Q_\varepsilon)}$ , and  $\chi^2(P_\varepsilon \| Q_\varepsilon)$  denotes the chi-square divergence (CSD) of  $P_\varepsilon$  from  $Q_\varepsilon$ .

*Proof.* The bound (6) is obtained by directly applying [24, Th. 1]. ■

**Remark 1.** The bound (6) is agnostic to the type/form of model mismatch.

The meaning of Prop. 1 is the following. If for the presumed environment  $\mathcal{Q}_{\text{env}}$ , which is used for training of the neural network, the localization error is  $Q_\varepsilon$ -distributed, and this distribution is “not too far”—in the CSD sense—from the true localization error distribution  $P_\varepsilon$ , then the MSE for the real data  $P$  is also “not too far”—in the sense of (6)—from the MSE for the presumed data  $Q$ .

While (6) is the strong version of the type of bound [24], it is instructive to analyze a looser version thereof (stated below), which nevertheless provides better insights as to the type of more “tolerable” forms of mismatch. To this end, we will make use of the following technical lemmas.

**Lemma 1.** Let  $\bar{\mathbf{h}}_Q^{(k)}$ , shorthand for  $\bar{\mathbf{h}}_Q^{(k)}(\mathbf{p}, \mathcal{Q}_{\text{env}})$ , be defined as<sup>6</sup>

$$\bar{\mathbf{h}}_Q^{(k)}(\mathbf{p}, \mathcal{Q}_{\text{env}}) \triangleq [\bar{h}_1[k] \dots \bar{h}_L[k]]^T \in \mathbb{C}^{L \times 1}, \quad (7)$$

where  $\{\bar{h}_\ell[k] = \mathbf{d}_\ell^H[k] \mathbf{b}_\ell\}$  in (7) are for  $\mathcal{Q}_{\text{env}}$ . Then the eigenvalues of

$$\widetilde{\mathbf{H}}_Q^{(k)} \triangleq \sigma_s^2 \bar{\mathbf{h}}_Q^{(k)} \bar{\mathbf{h}}_Q^{(k)H} + \sigma_v^2 \mathbf{I}_L \quad (8)$$

are given by

$$\boldsymbol{\lambda}_k^{(Q)} = [\sigma_s^2 \|\bar{\mathbf{h}}_Q^{(k)}\|^2 + \sigma_v^2 \underbrace{\sigma_v^2 \dots \sigma_v^2}_{L-1 \text{ terms}}]^T \in \mathbb{R}_+^{L \times 1}. \quad (9)$$

*Proof.* The proof follows by identifying (8) as a special rank-one “updated matrix”<sup>7</sup> of  $\sigma_v^2 \mathbf{I}_L$ , observing that  $\sigma_s \bar{\mathbf{h}}_Q^{(k)}$  is an eigenvector of  $\sigma_v^2 \mathbf{I}_L$  with an eigenvalue  $\sigma_v^2$ , and applying [25, Th. 2.1.] to get (9). ■

<sup>5</sup>Possibly with  $R \rightarrow \infty$ , or a different type of model mismatch.

<sup>6</sup>Recall that  $\bar{h}_\ell[k]$ , defined in (2), depends on the environment  $\mathcal{Q}_{\text{env}}$  or  $\mathcal{P}_{\text{env}}$ .

<sup>7</sup>Using the same terminology as the one used [25].

**Lemma 2.** [26, Ch. 7] Assume  $\det(\Sigma_P) \neq 0$ . Let  $\boldsymbol{\Omega} \triangleq \Sigma_Q \Sigma_P^{-1}$ , and denote  $\boldsymbol{\Omega} \mathbf{U} = \mathbf{U} \text{Diag}(\boldsymbol{\psi})$  as the eigendecomposition of  $\boldsymbol{\Omega}$ . Then, if

$$\psi_m \neq \psi_n^*, \quad \forall m \neq n \in \{1, \dots, NL\},$$

$\mathbf{U}$  is a joint diagonalizer (JD) of  $(\Sigma_Q, \Sigma_P)$ . In particular,

$$\Sigma_Q = \mathbf{U} \text{Diag}(\boldsymbol{\lambda}^{(Q)}) \mathbf{U}^H, \quad \Sigma_P = \mathbf{U} \text{Diag}(\boldsymbol{\lambda}^{(P)}) \mathbf{U}^H, \quad (10)$$

for  $\boldsymbol{\lambda}^{(Q)} \triangleq [\boldsymbol{\lambda}_1^{(Q)} \dots \boldsymbol{\lambda}_N^{(Q)}]^T, \boldsymbol{\lambda}^{(P)} \triangleq [\boldsymbol{\lambda}_1^{(P)} \dots \boldsymbol{\lambda}_N^{(P)}]^T \in \mathbb{R}_+^{NL \times 1}$ , where  $\{\boldsymbol{\lambda}_k^{(Q)} \in \mathbb{R}_+^{L \times 1}\}_{k=1}^N$  are defined in (9), and  $\{\boldsymbol{\lambda}_k^{(P)} \in \mathbb{R}_+^{L \times 1}\}_{k=1}^N$  are defined similarly but for the environment  $\mathcal{P}_{\text{env}}$ .

We now provide a generally looser, yet more insightful version of Prop. 1.

**Theorem 1.** Assume the conditions of Lemma 2 hold, and let  $\mathbf{U}$  be the JD of  $(\Sigma_Q, \Sigma_P)$  with  $\boldsymbol{\lambda}^{(Q)}, \boldsymbol{\lambda}^{(P)}$  as defined in (10). Then, if

$$2 \|\bar{\mathbf{h}}_Q^{(k)}\|^2 + \text{SNR}^{-1} > \|\bar{\mathbf{h}}_P^{(k)}\|^2, \quad \forall k \in \{1, \dots, N\}, \quad (11)$$

where  $\text{SNR} \triangleq \frac{\sigma_s^2}{\sigma_v^2}$ , the MSE of  $\hat{\mathbf{p}}$  for the true model  $P$  is upper bounded by

$$\mathbb{E}_P [\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2] \leq \mathbb{E}_Q [\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2] + \delta(P, Q), \quad (12)$$

where  $\delta(P, Q) \triangleq \sqrt{\text{Var}_Q (\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2) \Delta^2}$  with

$$\Delta^2 + 1 \triangleq \prod_{k=1}^N \frac{(\text{SNR} \|\bar{\mathbf{h}}_Q^{(k)}\|^2 + 1)^2}{(\text{SNR} \|\bar{\mathbf{h}}_P^{(k)}\|^2 + 1) (\text{SNR} (2 \|\bar{\mathbf{h}}_Q^{(k)}\|^2 - \|\bar{\mathbf{h}}_P^{(k)}\|^2) + 1)}. \quad (13)$$

*Proof.* By applying [24, Cor. 1], we obtain

$$\mathbb{E}_P [\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2] \leq \mathbb{E}_Q [\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2] + \sqrt{\text{Var}_Q (\|\boldsymbol{\varepsilon}_{\mathbf{p}}\|^2) \chi^2(P \| Q)}. \quad (14)$$

Since in this case  $P = \mathcal{CN}(\mathbf{0}, \Sigma_P), Q = \mathcal{CN}(\mathbf{0}, \Sigma_Q)$ , computation of  $\chi^2(P \| Q)$  by evaluating the integral implied by [24, Eq. 4] yields

$$\chi^2(P \| Q) = \frac{\det(\Sigma_Q)}{\det(\Sigma_P) \det(2\mathbf{I}_{NL} - \Sigma_P \Sigma_Q^{-1})} - 1. \quad (15)$$

By assumption, the conditions of Lemma 2 hold. Therefore, substituting (10) into (15) readily gives

$$\Delta^2 \triangleq \prod_{k=1}^N \prod_{\ell=1}^L \left( \frac{\gamma^2[k, \ell]}{2\gamma[k, \ell] - 1} \right) - 1, \quad (16)$$

under the condition

$$\gamma[k, \ell] \triangleq \frac{\lambda_k^{(Q)}[\ell]}{\lambda_k^{(P)}[\ell]} > \frac{1}{2}, \quad \forall k \in \{1, \dots, N\}, \quad \forall \ell \in \{1, \dots, L\}. \quad (17)$$

This arises because, after joint diagonalization, the entries of the transformed complex-normal (CN) vectors become uncorrelated, and due to their circular symmetry, they are statistically independent. Therefore, by virtue of the tensorization property of the CSD, each term of the product (16) corresponds to a CSD (plus one) of two *univariate* zero-mean CN random variables, which must be non-negative.

The proof now continues with two phases. In *Phase 1*, we obtain closed-form expressions for the eigenvalues  $\boldsymbol{\lambda}^{(Q)}, \boldsymbol{\lambda}^{(P)}$ . In *Phase 2*, we express  $\{\gamma[k, \ell]\}$  in terms of these eigenvalues, to obtain (11) and (13).

*Phase 1:* The elements  $\{\lambda_k^{(Q)}[\ell], \lambda_k^{(P)}[\ell]\}$  of  $\{\boldsymbol{\lambda}_k^{(Q)}, \boldsymbol{\lambda}_k^{(P)}\}$  can be efficiently computed, as follows. Recall that  $\{\mathbf{H}_\ell\}$  are all diagonal. Hence, the covariance matrices  $\boldsymbol{\Sigma}_Q$  and  $\boldsymbol{\Sigma}_P$  (5) have a block structure, where each  $N \times N$  block is a diagonal matrix. Thus, by applying the appropriate permutation matrix  $\mathbf{\Pi}$  and its inverse  $\mathbf{\Pi}^T$  on both sides, we have (focusing on  $\boldsymbol{\Sigma}_Q$ , since the logical argument is identical for  $\boldsymbol{\Sigma}_P$ )

$$\boldsymbol{\Sigma}_Q = \mathbf{\Pi}_{\text{row}}^T \left( \mathbf{\Pi}_{\text{row}} \boldsymbol{\Sigma}_Q \mathbf{\Pi}_{\text{col}} \right) \mathbf{\Pi}_{\text{col}}^T \triangleq \mathbf{\Pi}_{\text{row}}^T \widetilde{\boldsymbol{\Sigma}}_Q \mathbf{\Pi}_{\text{col}}^T,$$

where  $\mathbf{\Pi}_{\text{row}}, \mathbf{\Pi}_{\text{col}}$  permute the rows and columns, respectively, and where

$$\widetilde{\boldsymbol{\Sigma}}_Q = \text{Blkdiag} \left( \widetilde{\mathbf{H}}_Q^{(1)}, \dots, \widetilde{\mathbf{H}}_Q^{(N)} \right) \in \mathbb{C}^{NL \times NL},$$

and  $\text{Blkdiag}(\cdot, \dots, \cdot)$  forms a block diagonal matrix from its (square) matrix arguments. Since any permutation matrix is a unitary matrix,  $\det(\boldsymbol{\Sigma}_Q) = \det(\widetilde{\boldsymbol{\Sigma}}_Q)$ . Moreover, since  $\widetilde{\boldsymbol{\Sigma}}_Q$  is a block diagonal matrix, it follows that, for every  $k$ ,  $\lambda_k^{(Q)}$  are the eigenvalues of  $\widetilde{\mathbf{H}}_Q^{(k)}$  (8). Using the same arguments,  $\lambda_k^{(P)}$  are the eigenvalues of  $\widetilde{\mathbf{H}}_P^{(k)}$ . We conclude that  $\boldsymbol{\lambda}^{(Q)}, \boldsymbol{\lambda}^{(P)}$  can be computed via  $N$  eigendecomposition of  $L \times L$  matrices, rather than a single (naïve) eigendecomposition of an  $NL \times NL$  matrix. To complete phase 1, we have from Lemma 1 that

$$\lambda_k^{(D)}[\ell] = \begin{cases} \sigma_s^2 \|\bar{\mathbf{h}}_D^{(k)}\|^2 + \sigma_v^2, & \ell = 1, \\ \sigma_s^2, & 2 \leq \ell \leq L, \end{cases} \quad D \in \{Q, P\}. \quad (18)$$

*Phase 2:* Using the closed-form expressions (18), we now have

$$\gamma[k, \ell] = \frac{\lambda_k^{(Q)}[\ell]}{\lambda_k^{(P)}[\ell]} = \begin{cases} \frac{\sigma_s^2 \|\bar{\mathbf{h}}_Q^{(k)}\|^2 + \sigma_v^2}{\sigma_s^2 \|\bar{\mathbf{h}}_P^{(k)}\|^2 + \sigma_v^2}, & \ell = 1, \\ 1, & 2 \leq \ell \leq L. \end{cases} \quad (19)$$

Therefore, substituting (19) into (16), the latter is further simplified into

$$\begin{aligned} \Delta^2 + 1 &= \prod_{k=1}^N \prod_{\ell=1}^L \left( \frac{\gamma^2[k, \ell]}{2\gamma[k, \ell] - 1} \right) = \prod_{k=1}^N \left( \frac{\gamma^2[k, 1]}{2\gamma[k, 1] - 1} \right) \\ &= \prod_{k=1}^N \frac{(\sigma_s^2 \|\bar{\mathbf{h}}_Q^{(k)}\|^2 + \sigma_v^2)^2}{(\sigma_s^2 \|\bar{\mathbf{h}}_P^{(k)}\|^2 + \sigma_v^2) (\sigma_s^2 (2\|\bar{\mathbf{h}}_Q^{(k)}\|^2 - \|\bar{\mathbf{h}}_P^{(k)}\|^2) + \sigma_v^2)}. \end{aligned}$$

Using  $\text{SNR} = \frac{\sigma_s^2}{\sigma_v^2}$ , we obtain (13), and with (19), (17) simplifies to (11). ■

It is instructive to evaluate (13) in the following signal-to-noise ratio (SNR) limiting cases.

**High SNR Regime:** To better understand what governs the gap  $\delta(P, Q)$  at high SNR, define  $\rho_k \triangleq \|\bar{\mathbf{h}}_P^{(k)}\|^2 / \|\bar{\mathbf{h}}_Q^{(k)}\|^2$ , and observe that

$$\lim_{\text{SNR} \rightarrow \infty} \Delta^2 = \prod_{k=1}^N \frac{1}{\rho_k (2 - \rho_k)} - 1.$$

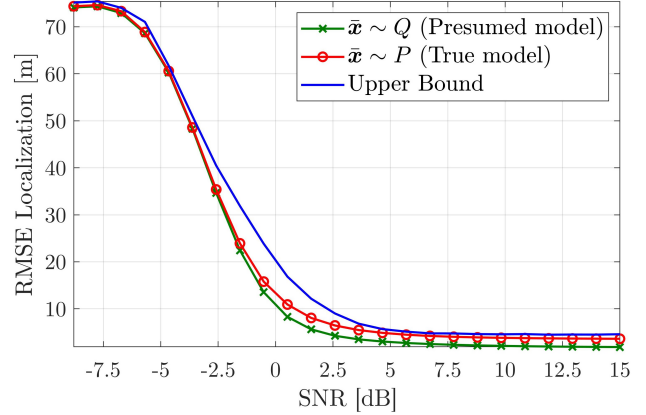
Thus, the gap is governed by product of functions of all the *per-frequency* energy deviations of the true frequency response from the presumed one.

**Low SNR Regime:** In this case, for any set of finite  $\{\rho^{(k)}\}$  per-frequency energy deviations, at low SNR, it follows from (19) that

$$\forall k \in \{1, \dots, N\} : \lim_{\text{SNR} \rightarrow 0^+} \gamma[k, 1] = 1 \implies \lim_{\text{SNR} \rightarrow 0^+} \Delta^2 = 0.$$

Hence, in the low SNR limit, the model mismatch is essentially irrelevant (in the sense of the attainable MSE).

A “conceptual interpolation” between these two extreme cases suggests that as the SNR increases, the model mismatch has an increasingly stronger effect on the gap  $\Delta^2$ , and consequently on the gap  $\delta(P, Q)$ .



**Fig. 2.** Localization RMSE vs. SNR for the presumed and true models,  $Q$  and  $P$ , respectively, and the upper bound (6).

#### 4.2. Simulation Results: DLOC in a Mismatched Environment

We consider the same scenario described in [19] with  $L = 4$  receivers, and a single source whose position was drawn (once, and then fixed) from a 3-dimensional uniform distribution supported on the volume of interest. However, instead of the 3-ray propagation model as in [19], we now use the Bellhop simulator to simulate a substantially richer environment, and generate the resulting CIR of each source-receiver pair corresponding to a pre-specified set of environmental parameters. For the training dataset  $\mathcal{D}_{\text{train}}^{(J)}$ , we use the environment  $\mathcal{Q}_{\text{env}}$  (similar to the one illustrated in Fig. 1), and for the test dataset we use  $\mathcal{P}_{\text{env}}$ , in which we use a *different* depth-varying speed of sound profile. The technical details pertaining to these environments are deferred to the supplementary material [23].

Fig. 1 shows the root MSE (RMSE) vs. the SNR<sup>8</sup> obtained by our CNN-based solution for: (i) a testset that is  $Q$ -distributed, the same as the training set, from the environment  $\mathcal{Q}_{\text{env}}$ , and; (ii) a testset that is  $P$ -distributed, stemming from the mismatched environment  $\mathcal{P}_{\text{env}}$ . We also plot the upper bound (6), which was evaluated numerically.<sup>9</sup> The results are based on averaging  $10^6$  independent trials per SNR point.

First, the robustness of the proposed DD-DLOC solution is evident, as reflected from the graceful performance degradation of the red curve relative to the green. Further, the upper bound (6) is not only fairly tight at the low and high SNR regimes, but is also informative about the SNR region in which the threshold phenomenon takes place. These results demonstrate the successful operation of our DD-DLOC solution in a rich and highly reverberant UWA environment, and at the same time corroborate our theoretical result in Prop. 1.

## 5. CONCLUDING REMARKS

The theoretical results of this work have an operational interpretation and an important implication to the DD-DLOC problem. They suggest that a system with a DNN that was trained on a dataset representing one environment ( $\mathcal{Q}_{\text{env}}$ ), which gives rise to a collection of possible CIRs, may still perform well (in the sense of the upper bounds (6) or (14)) if it is deployed in a different environment ( $\mathcal{P}_{\text{env}}$ ) that gives rise to a *different* collection of possible CIRs, as long as the overall mismatch is not too large (e.g., (11)). This has been demonstrated when using the deep CNN discussed in Section 3 for a simulated reverberant UWA environment.

An important aspect to be studied through future lines of this work is the domain-specific, informed incorporation of randomness to the training dataset. Such randomness should be selected so as to balance between localization accuracy and the MSE-robustness to CIR deviations.

<sup>8</sup>More accurately, a normalized SNR, relative to the average CIR attenuation.

<sup>9</sup>The CSD was estimated using the method in [27, 28], see also [29].

## 6. REFERENCES

- [1] J. Waterston, J. Rhea, S. Peterson, L. Bolick, J. Ayers, and J. Ellen, "Ocean of things: Affordable maritime sensors with scalable analysis," in *Proc. OCEANS Conf.*, 2019, pp. 1–6.
- [2] S. Sun, S. Qin, Y. Hao, G. Zhang, and C. Zhao, "Underwater acoustic localization of the black box based on generalized second-order time difference of arrival (GSTDOA)," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7245–7255, 2020.
- [3] G. Tuna and V. C. Gungor, "A survey on deployment techniques, localization algorithms, and research challenges for underwater acoustic sensor networks," *Int. J. Commun. Syst.*, vol. 30, no. 17, pp. e3350, 2017.
- [4] M. Erol-Kantarci, H. T. Mouftah, and S. Oktug, "Localization techniques for underwater acoustic sensor networks," *IEEE Commun. Mag.*, vol. 48, no. 12, pp. 152–158, 2010.
- [5] H.-P. Tan, R. Diamant, W. K. Seah, and M. Waldmeyer, "A survey of techniques and challenges in underwater localization," *Ocean Engineering*, vol. 38, no. 14–15, pp. 1663–1676, 2011.
- [6] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics," *IEEE J. Ocean. Eng.*, vol. 18, no. 4, pp. 401–424, 1993.
- [7] M. D. Collins and W. A. Kuperman, "Focalization: Environmental focusing and source localization," *J. Acoust. Soc. Am.*, vol. 90, no. 3, pp. 1410–1422, 1991.
- [8] M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle, "Machine learning in acoustics: Theory and applications," *J. Acoust. Soc. Am.*, vol. 146, no. 5, pp. 3590–3628, 2019.
- [9] H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li, "Deep-learning source localization using multi-frequency magnitude-only data," *J. Acoust. Soc. Am.*, vol. 146, no. 1, pp. 211–222, 2019.
- [10] A. Testolin and R. Diamant, "Underwater acoustic detection and localization with a convolutional denoising autoencoder," in *Proc. of CAMSAP*, 2019, pp. 281–285.
- [11] Z. Gong, C. Li, and F. Jiang, "A machine learning-based approach for auto-detection and localization of targets in underwater acoustic array networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15857–15866, 2020.
- [12] R. Chen and H. Schmidt, "Model-based convolutional neural network approach to underwater source-range estimation," *J. Acoust. Soc. Am.*, vol. 149, no. 1, pp. 405–420, 2021.
- [13] R. Lefort, G. Real, and A. Drémeau, "Direct regressions for underwater acoustic source localization in fluctuating oceans," *Applied Acoustics*, vol. 116, pp. 303–310, 2017.
- [14] Y. Wang and H. Peng, "Underwater acoustic source localization using generalized regression neural network," *J. Acoust. Soc. Am.*, vol. 143, no. 4, pp. 2321–2331, 2018.
- [15] D. Qin, J. Tang, and Z. Yan, "Underwater acoustic source localization using LSTM neural network," in *2020 39th Chinese Control Conference (CCC)*, 2020, pp. 7452–7457.
- [16] L. Houégnigan, P. Safari, C. Nadeu, M. André, and M. van der Schaar, "Machine and deep learning approaches to localization and range estimation of underwater acoustic sources," in *2017 IEEE/OES Acoustics in Underwater Geosciences Symposium (RIO Acoustics)*, 2017, pp. 1–6.
- [17] A. J. Weiss, "Direct position determination of narrowband radio frequency transmitters," *IEEE Signal Process. Lett.*, vol. 11, no. 5, pp. 513–516, 2004.
- [18] L. Wang, Y. Yang, and X. Liu, "A direct position determination approach for underwater acoustic sensor networks," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13033–13044, 2020.
- [19] A. Weiss, T. Arikan, and G. W. Wornell, "Direct localization in underwater acoustics via convolutional neural networks: A data-driven approach," in *Proc. 2022 IEEE 32nd Int. Workshop Mach. Learn. Signal Process.*, 2022, pp. 1–6.
- [20] P. C. Etter, *Underwater acoustic modeling and simulation*, CRC Press, 2018.
- [21] M. B. Porter, "The BELLHOP manual and user's guide: Preliminary draft," *Heat, Light, and Sound Research, Inc., La Jolla, CA, USA, Tech. Rep.*, vol. 260, 2011.
- [22] A. Weiss, T. Arikan, H. Vishnu, G. B. Deane, A. C. Singer, and G. W. Wornell, "A semi-blind method for localization of underwater acoustic sources," *IEEE Trans. Signal Process.*, vol. 70, pp. 3090–3106, 2022.
- [23] A. Weiss, A. C. Singer, and G. W. Wornell, "Towards robust data-driven underwater acoustic localization: A deep CNN solution with performance guarantees for model mismatch, supplementary materials," <https://www.weissamir.com/project/DLOC>, 2022.
- [24] A. Weiss, A. Lancho, Y. Bu, and G. W. Wornell, "A bilateral bound on the mean squared error for estimation in model mismatch," *Submitted to IEEE Int. Symp. Inf. Theory*, 2023.
- [25] J. Ding and A. Zhou, "Eigenvalues of rank-one updated matrices with some applications," *Applied Mathematics Letters*, vol. 20, no. 12, pp. 1223–1226, 2007.
- [26] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.
- [27] B. Póczos and J. Schneider, "On the estimation of alpha-divergences," in *Proc. of Int. Conf. Artif. Int. Statist.*, 2011, pp. 609–617.
- [28] B. Póczos, L. Xiong, D. J. Sutherland, and J. Schneider, "Non-parametric kernel estimators for image classification," in *Proc. of IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2989–2996.
- [29] J. J. Ryu, S. Ganguly, Y.-H. Kim, Y.-K. Noh, and D. D. Lee, "Nearest neighbor density functional estimation from inverse laplace transform," *IEEE Trans. Inf. Theory*, 2022.