Peer-Reviewed Technical Communication

Attentive Neural Processes for Fast Trajectory Prediction in Underwater Acoustic Networks

Fernando Fernández Salvador, Juan Parras, and Santiago Zazo

***Abstract*—Accurate underwater localization is crucial for autonomous underwater vehicles, yet it remains challenging due to the complex and dynamic nature of marine environments. Traditional localization methods that rely on anchor nodes and precise synchronization often fail to adapt to real-world underwater acoustic channel conditions characterized by multipath effects, high attenuation, and significant variability. Recent advances in deep learning offer model-free approaches that mitigate some of these issues, but they still struggle with generalization in scenarios with limited data. In this study, we introduce the application of attentive neural processes (ANPs) to underwater localization, enhancing both accuracy and robustness through effective uncertainty modeling and dynamic adaptation based on observed context data. By leveraging the ANP framework, our approach paves the way for extensive offline pretraining on simulated data, followed by rapid few-shot adaptation under real conditions; in this work, we evaluate near-zero-shot generalization. This strategy not only improves predictive performance but also significantly reduces the need for large-scale real-world data, making it highly suitable for dynamic underwater environments. Extensive evaluations demonstrate that ANPs significantly outperform traditional multilayer perceptron approaches, achieving reductions in mean absolute error of up to 25% compared to standard baselines. These results highlight the potential of ANPs to deliver reliable, adaptable, and precise localization solutions in challenging underwater scenarios.**

***Index Terms*—Acoustic telemetry and communication, attentive neural processes (ANPs), modeling and simulation, source localization, trajectory prediction.**

I. Introduction

UNDERWATER localization is a critical enabler for autonomous underwater vehicles (AUVs), sensor networks, and various marine applications ranging from environmental monitoring to military operations. Historically, techniques, such as time-of-arrival (TOA), time-difference-of-arrival (TDOA), angle-of-arrival (AOA), and matched-field processing (MFP) [1], [2] have dominated the field. These model-based approaches rely on precise environmental information (e.g., sound speed profiles and accurate channel models) and are well-grounded in physical principles. However, even slight deviations in these parameters can lead to significant localization errors [3], thereby limiting the applicability of these methods in dynamic underwater environments.

Received 8 June 2025; revised 7 October 2025; accepted 10 November 2025. This work was supported in part by the Spanish Ministry of Science, Innovation, and Universities, in part by the National Research Agency (10.13039/501100011033), and in part by the European Regional Development Fund under Grant PID2023-146540OB-C42 (NEMO4EX).

(Corresponding author: Fernando Fernández Salvador.)

Associate Editor: F. Ferreira.

The authors are with the Information Processing, Telecommunications Center, ETSI Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain (e-mail: fernando.fsalvador@upm.es).

Digital Object Identifier 10.1109/JOE.2025.3647846

In response to these challenges, the past decade has witnessed a surge of data-driven approaches utilising deep learning. Numerous studies have applied convolutional neural networks (CNNs) and recurrent neural networks to learn complex mappings between *raw acoustic sensor samples* and source positions, thus circumventing the need for explicit channel modelling [4]. For example, model-free localization techniques have been shown to improve trajectory prediction by directly inferring source positions from raw data [4]. Other works, such as direct localization approaches using deep CNNs and robust methods for handling model mismatch [5], have demonstrated that neural networks can capture the intricate behaviors of underwater acoustic channels. However, these methods often require large training data sets, lack inherent mechanisms to quantify prediction uncertainty, and can be sensitive to distributional differences between simulated and real data. Although strategies, such as transfer learning and domain randomization have been proposed to mitigate these issues [6], a unified framework that rapidly adapts to new underwater environments while also providing uncertainty-aware predictions remains elusive.

Attentive neural processes (ANPs) offer a promising alternative that addresses many of these challenges. Originally developed to combine the benefits of stochastic processes with the scalability of neural networks [7], ANPs learn distributions over functions from limited data while simultaneously providing uncertainty estimates for their predictions. By embedding an attention mechanism that dynamically weights observed data points, ANPs focus on the most informative features, even when data are sparse or corrupted by noise. Despite their successful application in domains such as robotic grasping and terrestrial navigation [8], to the best of the authors' knowledge, no previous study has applied ANPs to underwater acoustic localization.

The novel contribution of this work is the first systematic application of ANPs for underwater localization. Our approach leverages the adaptability and uncertainty quantification capabilities of ANPs to deliver the following.

1) *Adaptability:* Training on synthetic or simulated data sets under low variability and subsequently fine-tuning in unseen, high-variability conditions.
2) *Uncertainty quantification:* Providing calibrated uncertainty estimates that are critical for risk-sensitive operations and sensor network decision-making.
3) *Data efficiency:* Achieving high performance with relatively few labeled examples by exploiting the probabilistic nature of ANPs.

Our extensive evaluations demonstrate that the proposed ANP-based framework yields up to a 25% reduction in mean absolute error compared to traditional multilayer perceptron (MLP) architectures, underscoring its potential for robust and accurate underwater localization even in the presence of environmental mismatch.

The rest of this article is organized as follows. In Section II, we review the state of the art in underwater localization, contrasting traditional model-based approaches with recent deep-learning methods and introducing the foundations of ANPs; Section III describes our methodology, including the formal problem setup, sensor network and data collection process, and the detailed architectures of both the ANP and the baseline MLP models; Section IV presents the experimental framework and evaluation, reporting quantitative results for low- and high-variance channel scenarios, a qualitative trajectory comparison, and an overall analysis of performance and data-efficiency tradeoffs. Finally, Section V concludes this article with a summary of key findings, discusses the broader implications and limitations of our approach, and outlines directions for future work.

II. Related Work

As mentioned in the Section I, underwater localization has traditionally relied on *model-based* techniques, such as TOA, TDOA, AOA, MFP, extended Kalman filters, and particle filters to estimate target positions from acoustic measurements [3], [9]. While effective under controlled conditions, these methods depend on accurate channel models and precise sensor synchronization. Their performance rapidly degrades under multipath propagation, Doppler shifts, nonline-of-sight scenarios, and sparse sensor deployments [1], [9]. Akyildiz et al. [1] further highlighted the challenges posed by propagation delays, limited bandwidth, and channel variability in underwater networks, motivating the shift toward *model-free* learning approaches.

In this context, deep neural networks (DNNs) have emerged as powerful *model-free* alternatives by learning direct mappings from acoustic features to spatial coordinates. For instance, Parras et al. [4] trained an MLP on power and covariance features, achieving trajectory predictions with mean absolute errors below 50 m in low-variance environments. Building on this, Alonso-Eugenio et al. [10] and Clavijo-Rodríguez et al. [11] validated similar MLP-based pipelines in software-in-the-loop STANAG 5066 simulations and seawater quality monitoring networks, respectively. More recently, Parras and Zazo [12] incorporated reinforcement learning via proximal policy optimization to enhance AUV navigation under unknown disturbances.

However, standard DNNs lack inherent mechanisms to represent uncertainty. Kendall and Gal [13] addressed this by distinguishing between aleatoric and epistemic uncertainty in deep learning, introducing Bayesian techniques to estimate both the prediction mean and variance. Lou et al. [14] extended this idea to mobile robot localization, demonstrating robustness in low-data regimes. Guo et al. [15] exposed miscalibration issues in deep networks and proposed temperature scaling for posthoc confidence correction. More recent efforts have introduced Bayesian calibration frameworks to align predicted variances with empirical errors [16].

To better quantify uncertainty while maintaining flexibility, neural processes (NPs) combine the strengths of neural networks and Gaussian Processes by learning distributions over functions from sparse data [8]. ANPs extend NPs with self- and cross-attention mechanisms to dynamically weight context points, providing both global latent summaries and localized embeddings with improved calibration [7]. Fortuin et al. [17] further integrated Gaussian process priors into the attention structure of ANPs, enhancing sample efficiency and uncertainty quantification. To the best of the authors' knowledge, ANPs have not been applied to the underwater localization problem.

Recently, probabilistic deep models have begun making inroads into underwater applications. Yan et al. [18] employed deep Gaussian processes for AUV trajectory prediction, emphasising predictive uncertainty on real-world marine data. He et al. [19] developed uncertainty-aware learning approaches for acoustic underwater sensor networks, outperforming traditional techniques under noise and sparse sampling conditions. Zhang et al. [20] introduced probabilistic neural networks that adapt effectively to signal delays caused by environmental dynamics, relying on underwater channel models from Stojanovic and Preisig [21].

Despite the progress in model-free and uncertainty-aware techniques, no prior work, to the best of the authors' knowledge, has combined the adaptive context representation and calibrated predictive uncertainty provided by ANPs for underwater localization. Our work addresses this gap by training ANPs on simulated underwater channels, enabling fast, accurate, and uncertainty-aware trajectory predictions that generalize well across both low- and high-variance acoustic conditions.

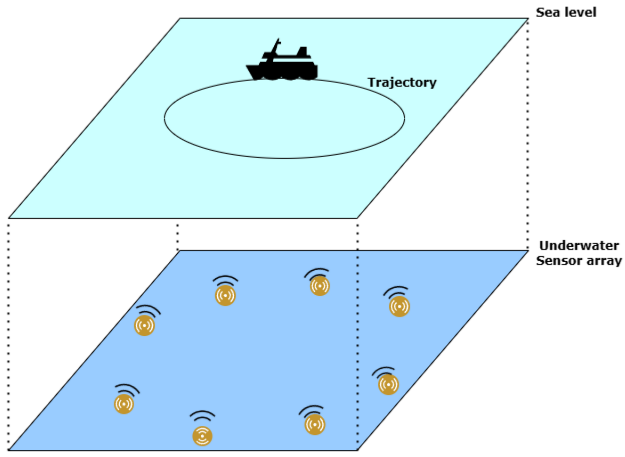

Fig. 1. Illustration of the simulated localization problem, where underwater sensors collect acoustic data from a ship floating at the surface.

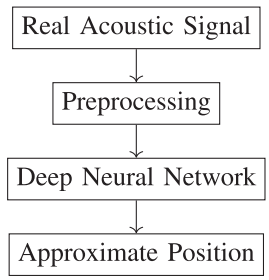

Fig. 2. Flow diagram of the DNN-based trajectory prediction process.

A. Brief Problem Description

Following Parras et al. [4], we address *noncooperative* passive localization: a surface target radiates broadband underwater noise/energy in the 10–20 kHz band, which is measured as *raw acoustic samples* by a static hydrophone network and preprocessed into features. This regime is orthogonal to cooperative self-report systems like automatic identification systems global navigation satellite system (GNSS) [22], which may be unavailable or deliberately disabled in practice.

The underwater localization task typically involves capturing real acoustic signals with an underwater sensor array, as seen in Fig. 1, extracting features via preprocessing, and feeding this data into a deep learning model to estimate the position of the vehicle.

A block diagram can clearly depict this pipeline. As shown in Fig. 2, the process includes the following.

1) *Real acoustic signal:* Raw input from underwater acoustic sensors.

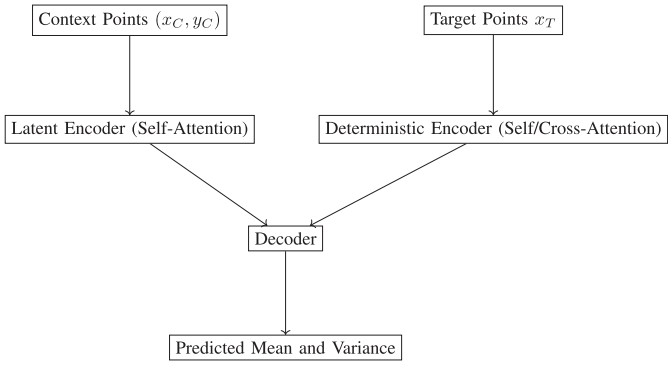

Fig. 3. Schematic of the ANP model. The diagram shows how context points are processed via self-attention (latent encoder) and combined with target-specific features through cross-attention (deterministic encoder) before being decoded into probabilistic predictions.

2) *Preprocessing:* Denoising, normalization, and feature extraction.
3) *DNN:* Learns the mapping from features to spatial coordinates.
4) *Approximate position output:* The model estimates the position of the underwater vehicle.

In this work, Parras et al. [4] stressed the importance of models that generalize beyond controlled environments to highly variable real-world scenarios. This challenge is also echoed in robotics applications with dynamic acoustic conditions [23].

B. Attentive Neural Processes

Traditional models like MLPs, while effective in structured settings, are deterministic and limited in their ability to generalize across unseen conditions [24]. ANPs overcome this by learning distributions over functions and offering both predictions and calibrated uncertainty estimates [7], [25].

Conceptually, an ANP ingests a *context set* of observed pairs (x_c, y_c), where x is the input information and y is the prediction target. A *latent encoder* applies selfattention to aggregate global information into a stochastic latent variable z, which is sampled via the reparameterization trick to capture predictive uncertainty. Simultaneously, a *deterministic encoder* employs crossattention to focus on the most relevant context points for each query x^*, yielding a deterministic embedding r. Here, we denote by x^* a query (or target) input, meaning an unobserved location at which the model must predict the corresponding output y^*. A decoder then combines (z, r, x^*) to output a Gaussian predictive distribution (mean and variance), enabling both point estimates and confidence intervals.

Fig. 3 presents a detailed schematic of the ANP model, illustrating how context points are processed through self-attention in the latent encoder and cross-attention in the deterministic encoder before being merged in the decoder.

This dual-path mechanism is especially well-suited for underwater localization. Training on low-variance simulated channels, followed by rapid fine-tuning, allows the model to adapt to real-world, high-variance conditions. ANPs not only outperform MLPs in generalization, but they also inherently quantify prediction confidence, which is a critical requirement for deployment in unpredictable underwater environments.

Let the context $\mathcal{C} = \{(x_i, y_i)\}_{i=1}^m$ denote the *observed* subset of points from the *same* trajectory being predicted (positions at a fraction of the 50 timestamps) under a fixed sensor topology. Given a query set $\mathcal{T} = \{x_j^*\}_{j=1}^n$, an ANP conditions on $\mathcal{C}$ via the deterministic cross-attention path and the global latent path to produce $p(y^* \mid x^*, \mathcal{C})$ [7]. Crucially, $\mathcal{C}$ is a tunable fraction (2%–80% in our experiments) that trades off sensing cost and accuracy; it does not require training data for *every* possible trajectory.

Although ANPs have been applied in robotics for tasks, such as anomaly detection and sequential data modeling [23], [25], to the bestof the authors' knowledge, this work is the first to explore their potential in underwater localization. We position ANPs as a promising solution for robust, uncertainty-aware trajectory estimation under diverse acoustic channel conditions.

III. Methodology

A. Problem Setup

Underwater localization involves predicting the position of a moving target based on acoustic signals captured by a network of static sensors (see Fig. 1 for an illustrative overview of the problem). In our work, we simulate the underwater environment using the statistical acoustic channel model described in [26].

This model captures the inherent complexities of underwater propagation by incorporating both small-scale phenomena, such as multipath propagation, scattering, and motion-induced Doppler shifts, and large-scale effects, including location uncertainty and environmental variability. The simulation generates acoustic signals by modeling multiple propagation paths. Each path is characterized by distinct gains and delays resulting from a combination of micromultipath components, which are modeled using Gaussian scattering processes. This approach provides a computationally efficient yet realistic representation of underwater acoustic channels, aligning well with experimental observations.

A key feature of the model is the introduction of a control parameter, θ. Rather than treating variability as an implicit byproduct of complex channel dynamics, θ explicitly governs the extent of variability in the simulation. When $\theta = 0$, the model produces a low-variance channel with stable, predictable conditions and minimal interference. Conversely, a value of $\theta = 1$ results in a high-variance channel that exhibits significant multipath effects, Doppler shifts, and other dynamic disturbances. In essence, θ serves as a unified parameter that allows us to seamlessly control and study the impact of varying acoustic conditions on localization performance.

Using this data generation model, we aim to address the following problems in this study.

1) How to accurately predict the trajectories of a moving target in the presence of significant underwater channel variability.
2) How to leverage ANPs to improve prediction accuracy (measured by MAE) and generalization in unseen scenarios compared to traditional DNNs.
3) How to design a flexible and interpretable model that quantifies uncertainty in predictions, which is crucial for applications in dynamic environments.

1) Sensor Network and Data Collection Process: The sensor network is composed of a set of static nodes deployed across a predefined 2-D region. These nodes are responsible for capturing acoustic signals emitted by the moving target. Each sensor records the received signal, extracting its power and covariance features, which reflect the received energy and the phase relationship between different sensors. The propagation of these signals was simulated using the statistical channel model defined by Qarabaqi and Stojanovic [26], which accounts for multipath effects, scattering, and Doppler-induced distortions. This data are then processed to infer the target trajectory. The process of data collection can be described as follows.

a) *Signal emission:* A moving target emits an acoustic signal as it follows its trajectory.

b) *Signal reception:* The static sensors record the received signal, which is affected by channel conditions (parameterized by θ) and noise (fixed SNR of 10 dB). The signal is processed and normalized for consistent feature extraction.

c) *Feature extraction:* In this work, we explored feature extraction methods for power and covariance that were introduced in [4]. These features can help capture both the energy distribution and the intersensor phase relationships of the received acoustic signals.
For power extraction, we compute the power of the received signal as

$$P = \frac{1}{N}\sum_{n=1}^{N} |s[n]|^2 \tag{1}$$

where $s[n]$ is the received signal at time index n and N is the total number of samples.
The covariance features are extracted by computing the covariance matrix C

$$C = \frac{1}{N}\sum_{n=1}^{N} (s[n]-\mu)\,(s[n]-\mu)^* \tag{2}$$

where

$$\mu = \frac{1}{N}\sum_{n=1}^{N} s[n] \tag{3}$$

is the mean of the signal, and the superscript $*$ denotes the Hermitian transpose.
These formulations provide a robust representation of the signal characteristics, and their effectiveness in the context of underwater localization was demonstrated in [4], which motivates their integration and further exploration in our work.

d) *Trajectory reconstruction:* These features are then used to train machine learning models, such as the ANP and MLP, to predict the trajectory of the target.

The generated data channels were divided into two groups based on their θ values: a *low variance group* representing more stable channel conditions and a *high variance group* representing highly dynamic channel conditions. This division reflects real-world scenarios where underwater environments may range from calm and predictable (low θ) to turbulent and highly variable (high θ). Evaluating models on both groups ensures robustness across diverse operational conditions. Each group was processed identically to extract features for training and evaluation.

2) Diagram of Sensor Network and Data Collection: Fig. 4 illustrates the sensor network setup and the data collection process. The sensors are uniformly distributed across the environment, ensuring comprehensive spatial coverage. The moving target's trajectory represents a typical path in underwater environments, with signal emissions modeled as periodic pulses. This plot also illustrates the influence of channel conditions (θ) and noise on the received signals. These sensors capture signal features, including power and covariance, and process them to infer the trajectory of the target.

3) System Assumptions and Data Acquisition: We assume centralized offline training with raw multisensor waveforms available from the hydrophone network (wired where needed due to bandwidth), while at runtime, only features or final position estimates may be transmitted to a gateway over low-rate acoustic links. This reflects practical UASN constraints on bandwidth/latency [1], [27] and isolates our contribution to the learning model (ANP versus MLP) rather than online networking. All axes in our topology/trajectory figures are labeled in meters, and default layouts place sensors within approximately $\pm$1–2 km in x and $\pm$1 km in y.

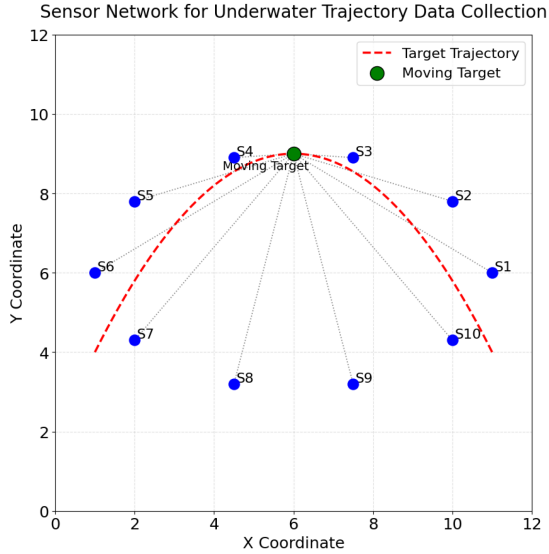

Fig. 4. Example illustration of the sensor network for underwater trajectory data collection. The moving target emits acoustic signals that are received by sensors, which extract the features for trajectory prediction. The uniform sensor placement ensures complete spatial coverage, while the effects of channel variability (θ) are simulated using a statistical channel model.

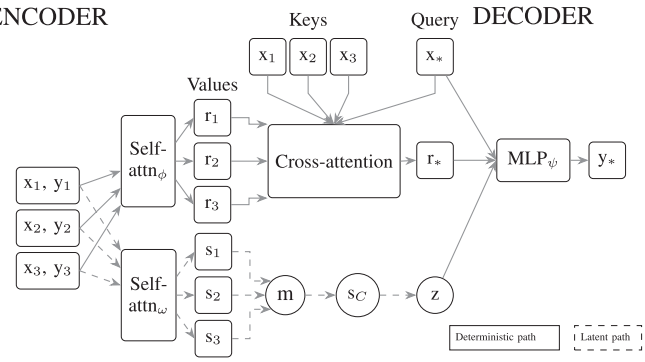

Fig. 5. Architecture of the ANP. Latent encoder uses self-attention to generate a global latent vector, while the deterministic encoder applies cross-attention to produce query-specific context representations. Adapted from [7].

B. ANP Architecture

ANPs enhance NPs by integrating attention mechanisms that separately capture global trends and query-specific details [7], [28]. As shown in Fig. 5, the model has two encoder branches: latent and deterministic, and a decoder. We let: $\omega : \mathbb{R}^{d_x+d_y} \to \mathbb{R}^d$ and $\phi : \mathbb{R}^{d_x} \to \mathbb{R}^d$ denote both learnable MLPs that embed context pairs and queries, respectively. Here, d_x and d_y denotes the dimension of input X and target y, and d is the dimension of the latent space.

a) Latent encoder with self-attention: Given a context set $C = \{(x_i, y_i)\}_{i=1}^{|C|}$, we first embed each pair into representations $h_i = \omega(x_i, y_i)$ and stack them into $H = [h_1, \ldots, h_{|C|}] \in \mathbb{R}^{|C|\times d}$. We then apply scaled dot-product self-attention [29]

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{\text{QK}^\top}{\sqrt{\text{d}_\text{k}}}\right)\text{V} \tag{4}$$

with

$$Q = HW_Q,\ K = HW_K,\ V = HW_V \tag{5}$$

where $W_Q, W_K : \mathbb{R}^d \times \mathbb{R}^{d_k}$ and $W_V : \mathbb{R}^d \times \mathbb{R}^{d_v}$ are linear projections. The dot products $QK^\top$ measure pairwise similarity, and scaling by $\sqrt{d_k}$

stabilizes gradients. The softmax converts scores into attention weights that aggregate the rows of V. From a signal-processing standpoint, the attention output can be regarded as a (nonparametric) estimator: for a given query q, it forms a data-adaptive, similarity-weighted average of the value vectors $\sum_i \alpha_i(q, k_i)\, v_i$ with $\alpha_i \propto \exp(q^\top k_i/\sqrt{d_k})$. This "attention-as-estimator" view has been exploited in spectrum cartography: Spatial TransfOrmers for Radio Map Estimation cast attention as a gridless, permutation- and Euclidean-equivariant estimator for spatial interpolation, achieving state-of-the-art radio-map reconstruction [30].

This produces a *global summary* $s_C \in \mathbb{R}^{d_v}$ that encodes the overall context structure. We write per-point self-attention outputs as s_i, and define

$$s_C \;=\; \frac{1}{|C|}\sum_{i=1}^{|C|} s_i. \tag{6}$$

A Gaussian latent variable $z \sim \mathcal{N}\big(\mu_\omega(s_C), \sigma^2_\omega(s_C)\big)$ is then sampled via the reparameterization trick [31], where $\mu_\omega, \sigma_\omega$ are learnt MLPs. The stochastic z captures epistemic uncertainty about the underlying function [8].

b) Deterministic encoder with self and cross-attention: The deterministic encoder refines predictions for each target x^*. First, context embeddings H undergo a self-attention layer (as above), producing per-point vectors r_i (stacked into $\widetilde{H} = [r_1, \ldots, r_{|C|}]$), which preserve permutation equivariance [29]. Next, for each query embedding $h^* = \phi(x^*)$, we perform *cross attention* against $\widetilde{H}$

$$r_* = \mathrm{softmax}\left(\frac{\mathrm{h^*W_Q}\,(\widetilde{\mathrm{H}}\mathrm{W_K})^\top}{\sqrt{\mathrm{d_k}}}\right)\left(\widetilde{\mathrm{H}}\mathrm{W_V}\right). \tag{7}$$

Here, h^*W_Q serves as the query, $\widetilde{H}W_K$ as keys, and $\widetilde{H}W_V$ as values. Cross-attention weights each context point by its relevance to x^*, yielding a *query-specific summary* r_*[25], [32]. This pathway is deterministic, sharpening local detail. In Fig. 5, the "Keys" block represents $\widetilde{H}W_K$ and the "Values" block represents $\widetilde{H}W_V$, even though the raw inputs x_i and the self-attention outputs r_i are shown as their respective data icons.

c) Decoder: It concatenates the latent representation, deterministic embedding, and target query $(z,\ r_*,\ x^*)$ into a unified representation and feeds them through a small MLP, with parameters ψ, to output Gaussian parameters

$$p\,(y^* \mid x^*, C) = \mathcal{N}\left(y^*;\ \mu_\psi(x^*),\ \sigma^2_\psi(x^*)\right). \tag{8}$$

Thus, providing both a point estimate and a calibrated uncertainty interval [23].

d) Roles of the latent and deterministic encoders: The latent and deterministic encoders serve complementary purposes, enabling the ANP to balance flexibility, precision, and uncertainty modeling.

1) The *latent encoder* provides a *global probabilistic foundation* for the model. By capturing overarching patterns in the data, it ensures that the ANP accounts for high-level trends and relationships across trajectories. The stochastic latent variable z allows the model to generate diverse predictions and quantify uncertainty, making it robust to noisy and variable environments.
2) The *deterministic encoder* refines predictions by focussing on *local, query-specific details*. Its cross-attention mechanism ensures that each target point is informed by the most relevant context points, enabling dynamic adaptation to varying data distributions. This is crucial for trajectory prediction in scenarios with large variability.

e) Training Objective: We assume that the target outputs Y given inputs X follow a Gaussian distribution, which allows the model to learn both the mean and variance of predictions, thus quantifying uncertainty, a process known as *calibration* [16], [33]. The loss maximizes the evidence lower bound, combining a negative log-likelihood (NLL) term

$$\begin{aligned}\mathcal{L}_{\text{NLL}} \;&=\; -\ln p(\mathbf{y} \mid \mathbf{x}^*, \mathbf{r}_*, \mathbf{z}) \\ &= \frac{1}{2}\ln\left((2\pi)^k\,|\Sigma(x^*)|\right) + \frac{1}{2}\,(\mathbf{y} - \mu_\psi(\mathbf{x}^*, \mathbf{r}_*, \mathbf{z}))^T \\ &\quad \Sigma^{-1}\ (\mathbf{y} - \mu_\psi(\mathbf{x}^*, \mathbf{r}_*, \mathbf{z}))\end{aligned} \tag{9}$$

where $\Sigma(x^*) = \mathrm{diag}(\sigma^2_\psi(\mathrm{x}^*))$ is the diagonal covariance matrix of our predictive Gaussian. The $\mathcal{L}_{\text{NLL}}$ measures the likelihood of the target data under the predicted distribution, with a kullback-leibler (KL) divergence regularizer

$$D_{\text{KL}}\left(\mathcal{N}(\mu_1, \Sigma_1) \parallel \mathcal{N}(\mu_2, \Sigma_2)\right) = \frac{1}{2}t \tag{10}$$

where t is

$$t = \left(\ln\frac{|\Sigma_2|}{|\Sigma_1|} - k + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_2 - \mu_1)^T\Sigma_2^{-1}(\mu_2 - \mu_1)\right). \tag{11}$$

The KL regularizer encourages the approximate posterior $q(z \mid C)$ to remain close to the prior $p(z)$ and prevents overconfident uncertainty estimates [15]. Proper calibration of the predictive variance has been shown to be critical in decision-making applications, as it ensures that reported confidence levels faithfully reflect true error rates [16], [33].

The total loss is then computed as

$$\mathcal{L} = \mathcal{L}_{\text{NLL}} + D_{\text{KL}}. \tag{12}$$

Intuitively, the ANP works by first extracting critical information from the observed context. The latent encoder captures overall trends and global patterns, while the deterministic encoder refines this by focussing on the most relevant features for each target prediction. The decoder then uses this combined representation to generate a probability distribution over the target outputs, effectively predicting the position while also quantifying uncertainty.

C. MLP Architecture

For comparison, we implement a baseline MLP that flattens the preprocessed input features into a single vector and passes them through three hidden layers of neurones, with rectified linear unit (ReLU) activations and dropout regularization to mitigate overfitting [34]. The output layer produces the estimated trajectory, and the model is trained by minimising the mean squared error (MSE). A basic representation of this architechture can be seen in Fig. 6.

This architecture is adapted from the DNN approach described in [4] and is widely used in many applications, ensuring a fair baseline for evaluating the probabilistic performance of our ANP model. Detailed implementation aspects, including input reshaping, training procedures, and hyperparameter choices for the MLP, can be found in the Appendix A.

IV. Experimental Setup, Evaluation, and Analysis

In this section, we describe the experimental framework and present a detailed analysis of our results. We first introduce the common setup, including data set details, preprocessing, and training parameters. Then, we present three experiments that evaluate the performance of the ANP and the baseline MLP models under different underwater channel

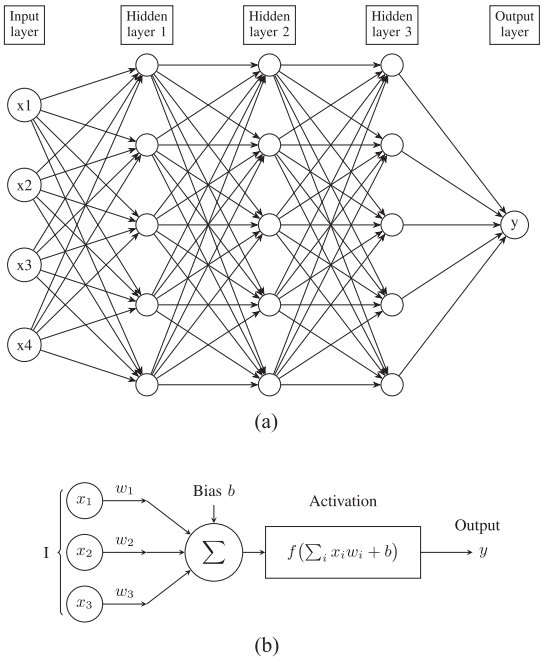

Fig. 6. Standard representation of a MLP architecture with four inputs features, followed by three fully connected hidden layers of five neurons each, producing a single output. (b) Representation of a single-neuron operation, inputs are weighted by and summed, a bias is added, and an activation function $f(\cdot)$ (in our case a **Re**ctified **L**inear **U**nit) is applied to yield the neuron output.

conditions. Each experiment includes plots that illustrate the results, along with annotations for clarity.

Our code is publicly available to foster reproducibility and future research in: Github.

A. Common Experimental Setup and Parameters

Data generation and processing:

1) *Acoustic band and grid:* We simulate a 10–20-kHz band with a frequency resolution of $\Delta f = 25$ Hz (401 frequency points), using a stochastic shallow-water channel with small- and large-scale variability [26]. The IFFT-domain impulse responses thus have ≈ 0.1 ms resolution over a ≈ 40 ms window.
2) *Array and SNR:* Unless stated otherwise, we use ten hydrophones with typical spacing within a $\sim$ 1-km scale and an SNR of $= 10$ dB.
3) *Raw tensors:* The raw filtered data are arranged as $(401, 50, 100, 10)$ (frequency/time taps, points per trajectory, trajectories, sensors); features (power/covariance) follow Parras [4]. This means that we generate 100 trajectories for each channel θ, and each trajectory has 50 points. More details about the data preparation can be found in the Appendix A.
4) *Channel variability:* The data set is divided into two groups based on the channel variability parameter θ. The low-variance group covers $\theta = 0.0$ to 0.5, stratified into six subgroups. The high-variance group covers $\theta = 0.6$ to 1.0, stratified into five subgroups.
5) *Train-validation-test split:* A 70/20/10 stratified split (based on θ) is used to ensure balanced representation across different channel conditions.

B. Training Procedures and Hyperparameters

a) For the ANP: We employ a two-phase training regimen. In the first phase, the model is trained exclusively on low-variance channel data to learn the underlying trajectory dynamics under relatively stable conditions. In the second phase, we evaluate its few-shot performance on high-variance data, testing the ability of the ANP to generalize without additional fine-tuning. To study the impact of the available context, we vary the fraction of observed trajectory points between 2% and 80%, measuring prediction accuracy as a function of context size.

b) For the MLP: We trained the model both on individual θ-stratified subgroups. Its architecture consists of an input layer, three hidden layers of 128 neurones each with ReLU activations, and dropout layers with a rate of 0.1 to mitigate overfitting [34]. A final output layer produces the predicted trajectory coordinates.

In addition to training MLP models on individual θ-stratified subgroups, to test cross-domain robustness without attention or latent conditioning, we train a single MLP on the aggregate of low-variance channels (θ=0.0, . . . , 0.5), following the domain randomization idea used for sim-to-real robustness [6]. This model, which we denote *DR-MLP*, shares architecture and hyperparameters with the specialist MLPs; only the training set differs.

Evaluation metric: MAE is used to quantify the average prediction error

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{13}$$

MAE is especially useful in our study since we are dealing with real distances in underwater localization. This metric provides an intuitive measure of error in the same unit as the target variable (e.g., meters), making it easier to interpret in practical scenarios. In addition, MAE is less sensitive to large outliers compared to metrics like mse, which is important in our context, where small positional deviations can be critical.

C. Experiment I: Low-Variance Channel Evaluation

a) Objective: The goal of this experiment is to assess how well the ANP model, trained on low-variance channel data ($\theta \in [0.0, 0.5]$), predicts trajectories compared to both MLP subgroup models, each trained on a single θ value, and a domain random MLP trained on the entire low-variance range.

b) Methodology: We train the ANP exclusively on low-variance simulations, varying the fraction of observed trajectory points (*context size*) between 2% and 80%. Validation is performed on six stratified θ-stratified subgroups, measuring the MAE for each.

c) Statistical Significance via Adjusted p-Values: To verify whether the observed MAE differences in Table I are merely noise or truly meaningful, we performed a paired, nonparametric Wilcoxon signed-rank test for each θ-stratified subgroup. Concretely, for each column (θ), we have the following.

1) Identified the "reference" model as the one attaining the lowest mean MAE for that subgroup.
2) Collected per-trajectory absolute errors of the reference model versus each other model (all MLPs and the ANP at each context size).
3) Ran a Wilcoxon signed-rank test between the two error distributions to produce an unadjusted p-value.
4) Adjusted these p-values across all pairwise comparisons in that column using the Holm–Bonferroni procedure to control the family-wise error rate at $\alpha = 0.05$.

TABLE I
MAE IN METERS FOR LOW-VARIANCE DATA ($\theta = 0.0 - 0.5$): ANP VERSUS MLP SUBGROUPS AND DR-MLP

Model / Context	0.0	0.1	0.2	0.3	0.4	0.5
MLP(0.0)	**19.1**⋆	40.6	60.3	133.0	271.1	405.0
MLP(0.1)	48.0	**19.5**⋆	68.1	141.8	278.9	435.7
MLP(0.2)	60.1	56.3	**26.4**⋆	111.8	234.0	363.9
MLP(0.3)	86.6	88.3	87.3	41.9⋆	187.2	315.5
MLP(0.4)	186.0	187.2	184.1	145.1	104.8	288.8
MLP(0.5)	397.6	394.5	400.6	320.7	260.8	144.6⋆
DR-MLP	38.3	37.9	41.0	68.6	126.3	187.5
ANP(2%)	35.2	37.7	45.3	68.4	128.0	192.2
ANP(4%)	31.5	29.8	39.6	53.9	112.4	174.1
ANP(6%)	27.8	25.9⋆	34.5⋆	44.9	100.9	161.1
ANP(8%)	25.0⋆	23.7⋆	31.9⋆	39.6	93.5	153.4
ANP(10%)	23.2⋆	22.8⋆	30.2⋆	38.0	90.0	149.7
ANP(15%)	22.3⋆	21.8⋆	29.0⋆	36.7⋆	88.2⋆	146.7⋆
ANP(25%)	21.0⋆	20.8⋆	28.9⋆	36.5	85.3⋆	144.4⋆
ANP(40%)	19.6⋆	19.7⋆	27.5⋆	35.5⋆	82.3⋆	141.2⋆
ANP(60%)	20.0⋆	20.4⋆	28.7⋆	35.4⋆	**81.5**⋆	**139.1**⋆
ANP(80%)	19.8⋆	21.3⋆	29.5⋆	**35.2**⋆	81.8⋆	139.2⋆

Bold letters indicates best performing model per channel, and the Star denotes no significant difference with the best performance using Wilcoxon test with significance level 0.05. ANP is unmatched for low θs and from 15% context.

TABLE II
MAE IN METERS FOR HIGH-VARIANCE DATA ($\theta = 0.6 - 1.0$): ANP VERSUS MLP SUBGROUPS AND DR-MLP

Model / Context	0.6	0.7	0.8	0.9	1.0
MLP(0.6)	**182.0**⋆	326.9	311.0	351.8	388.2
MLP(0.7)	326.8	**207.6**⋆	305.2	341.5	377.6
MLP(0.8)	354.5	364.7	**225.7**⋆	338.3	385.0
MLP(0.9)	375.9	349.3	314.4	293.1⋆	384.9
DR-MLP	226.5⋆	238.7⋆	242.3⋆	272.9⋆	349.1
ANP(2%)	231.4⋆	288.0	254.3⋆	296.1	312.5⋆
ANP(4%)	213.9⋆	272.5	245.7⋆	288.6⋆	307.9⋆
ANP(6%)	199.0⋆	261.1	239.9⋆	282.1⋆	303.4⋆
ANP(8%)	189.2⋆	253.2⋆	234.0⋆	278.7⋆	299.9⋆
ANP(10%)	186.6⋆	250.3⋆	230.3⋆	276.9⋆	299.0⋆
ANP(15%)	185.9⋆	247.8⋆	229.0⋆	275.1⋆	296.5⋆
ANP(25%)	187.6⋆	246.5⋆	228.7⋆	272.4⋆	295.6⋆
ANP(40%)	187.5⋆	242.6⋆	227.0⋆	267.9⋆	288.9⋆
ANP(60%)	189.6⋆	241.5⋆	226.0⋆	**266.9**⋆	282.8⋆
ANP(80%)	189.3⋆	239.5⋆	225.6⋆	267.9⋆	**280.4**⋆

Bold letters indicates the best performing model per channel, and the Star denotes no significant difference with the best performance using Wilcoxon test with significance level 0.05. The transferred ANP model generalizes good enough to match the best MLPs in each channel, despite these being trained on that specific data.

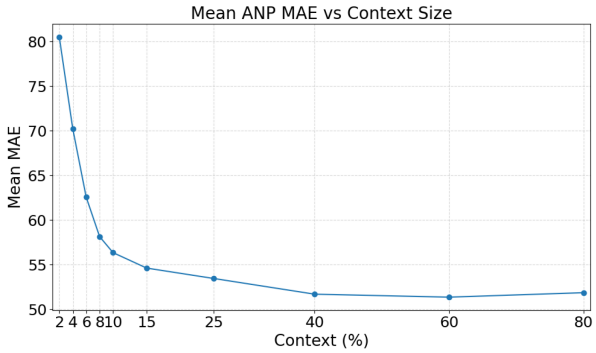

Fig. 7. Mean MAE versus context size for the ANP on low-variance data. Improves sharply from 2% to 10% and it plateaus at 40%.

An adjusted p-value > 0.05 tells us that *we cannot reject* the null hypothesis of equal performance at the 5% level; those models are statistically indistinguishable from the best performer in the subgroup.

d) Graphical Analysis: Table I shows the evolution of MAE across $\theta \in \{0.0, 0.1, \ldots, 0.5\}$ and various context sizes. The top six rows correspond to the MLP subgroup models; each model achieves a low error on its own θ, but it fails to generalize even to neighbouring channel conditions (e.g., for $\theta = 0.0$, MAE jumps from 17 to over 120 m at $\theta = 0.3$). The DR-MLP model (seventh row) smooths this variability; its MAE ranges from around 47 to 170 m, but it remains markedly higher than the best ANP. The subsequent rows trace the MAE of the ANP: with only 2% context, the ANP already beats the DR-MLP in some θ -stratified subgroups, with errors ranging from 34 to 182 m. As context grows to 25%, the error of the ANP falls below 35 m across the first four θ-stratified subgroups and remains around 80 m even at $\theta = 0.4$. The only subgroup that still poses some difficulties is $\theta = 0.5$, but even for this group, the ANP outperforms all models at 25% context.

Fig. 7 plots the average MAE of the ANP across all six θ values as a function of context size. The curve drops steeply from 86 m at 2% context to 56 m at 10%; then, from there, it declines more gradually to a plateau near 50 m beyond 40%. The sweet spot of around 15% to 25% demonstrates once again how good the ANP performance is, with relatively low costs in terms of data.

e) Results and Discussion: These results demonstrate three key benefits of the ANP in low-variance conditions. First, from just 4% context onward, the ANP matches and then surpasses the DR-MLP baseline across most θ-stratified subgroups, highlighting its ability to interpolate information across trajectories. Second, its MAE decreases sharply with small increases in observed points; most of the significant improvements occur between 2% and 10% context, indicating efficient use of limited data. Third, beyond 25% context, the gains diminish, suggesting a practical tradeoff point where prediction accuracy and context-collection cost balance. Finally, this context-size parameter offers a direct tradeoff between precision and speed: larger contexts yield more precise results but require proportionally more observed samples (and thus more processing time). In contrast, MLP subgroup models, lacking probabilistic attention, fail to generalize beyond their training θ, and the DR-MLP, though more robust, cannot match the combined precision and data efficiency of the ANP.

D. Experiment II: High-Variance Channel Generalization

a) Objective: In this experiment, we test the ability of the ANP to generalize in a nearzeroshot setting to highvariance channel conditions ($\theta \in [0.6, 1.0]$) despite having been trained only on low-variance data. We compare its performance against the MLP trained on each high-variance subgroup and a DR-MLP trained on the entire high-variance range.

b) Methodology: The ANP, with fixed weights from low-variance training, is evaluated on five stratified high-variance subgroups ($\theta = 0.6, 0.7, 0.8, 0.9, 1.0$). For each subgroup, we measure the MAE as we vary the context size from 2% to 80% of the trajectory points. In parallel, we train the MLP on individual θ-stratified subgroups and a DR-MLP on the full high-variance set, recording their MAEs under the same context-size protocol.

c) Statistical Significance: As before, markers "⋆" in Table II indicate where the difference in MAE versus the best model of the column is not statistically significant at the 5% level (paired Wilcoxon test with Holm–Bonferroni correction).

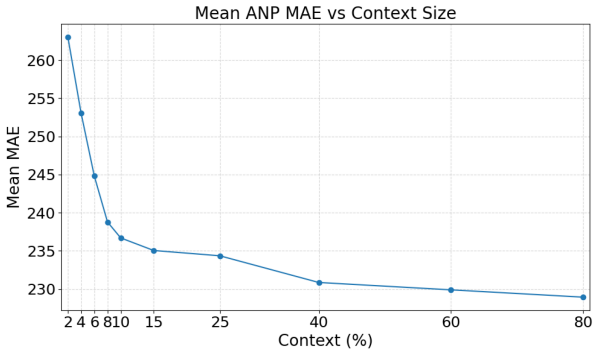

Fig. 8. Mean MAE versus context size for the ANP on high-variance data. The model is more sensitive than at low variance, showing constant improvement.

d) Graphical Analysis: Table II presents the evolution of MAE across the five high-variance θ values. The top five rows show MLP subgroup performance: each model performs well on its own channel but degrades sharply elsewhere (e.g., the $\theta = 0.6$ MLP jumps from 173 m at its trained condition to over 360 m by $\theta = 1.0$). The DR-MLP smooths this variability again, but its average MAE remains higher than before (240–330 m). In contrast, the ANP (subsequent rows) maintains a MAE between 200 and 290 m even with only 4% context. As context increases, the error of the ANP steadily declines to roughly 180–280 m at 25% context. Notably, beyond 40% context, the gains taper off, indicating a saturation point in high-variance settings.

Fig. 8 plots the mean MAE of the ANP across all five high-variance subgroups versus context size. As expected from previous experiments, the curve shows a steeper decline at small context sizes, from around 260 m at 2% to 235 m at 15%, and then it continues to improve noticeably until it flattens around 230 m beyond 60% context, highlighting diminishing returns on additional observations under highly variable channels.

e) Results and Discussion: Despite lacking direct exposure to high-variance training data, the ANP consistently outperforms both subgroup-specific models and DR-MLP across all context sizes. Its probabilistic latent representation and attention-based conditioning enable robust adaptation to unseen channel disturbances, whereas deterministic MLPs either overfit their narrow training θ or cannot balance the heterogeneous distribution. The performance curve of the ANP further reveals that, although high-variance noise inflates absolute errors (mean MAE $\approx$ 230 m versus $\approx$ 50 m in low-variance), context-driven uncertainty reduction remains effective, plateauing once sufficient information has been observed. These findings confirm the superior generalization and uncertainty modelling capabilities of the ANP in challenging underwater environments. They also show that, as expected, as uncertainty increases, larger samples of context are needed to reduce the error.

E. Experiment III: Trajectory Prediction Comparison (Qualitative Analysis)

a) Objective: This experiment offers a qualitative side-by-side comparison of the ANP versus the DR-MLP on two representative trajectories selected from one of the low-variance test sets. Beyond aggregate MAE scores, we inspect how each model recovers the overall shape, smoothness, and, uniquely for the ANP, the uncertainty bands of the true path.

b) Methodology: We randomly select two trajectories and withhold 25% of each as context. The ANP uses those context points to predict the full 2-D trace and to output a point-wise standard deviation,

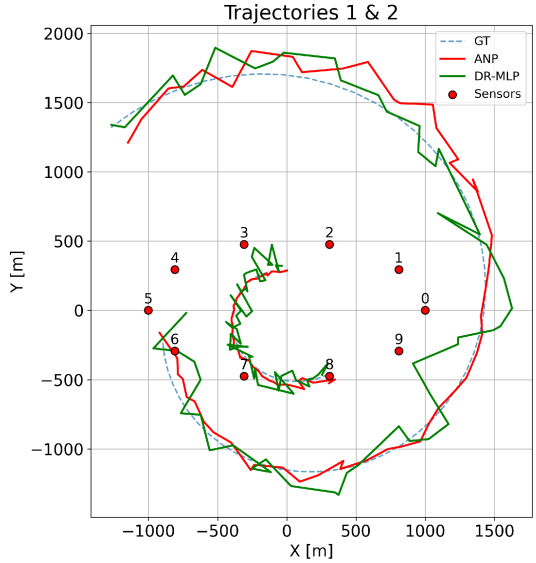

Fig. 9. 2-D trajectory overlays: ground truth (dashed blue), ANP (red), DR-MLP (green). Scale is in meters.

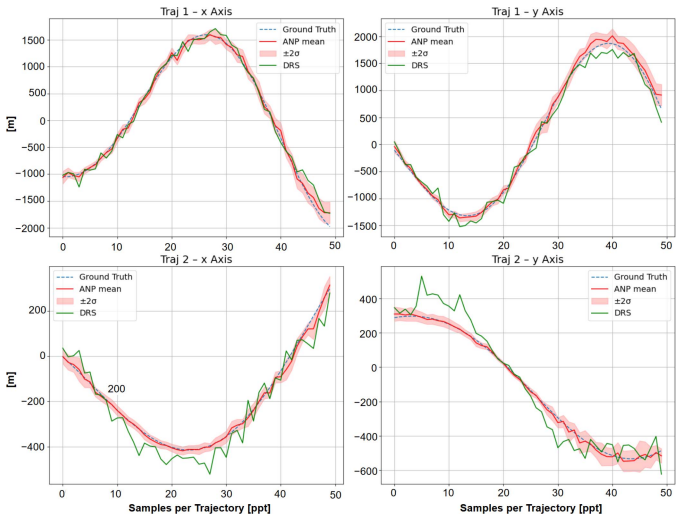

Fig. 10. Axis-wise predictions with ANP $\pm 2\sigma$ (shaded red) versus DR-MLP (green). The x-axis corresponds to the total number of points per trajectory and the scale of y-axis is in meters.

whereas the DR-MLP simply regresses the 2-D coordinates. We then visualize

1) the full 2-D reconstruction of each trajectory (ground truth versus ANP versus DR-MLP), and
2) the separate x- and y-axes predictions for each trajectory, overlaid with the $\pm 2\sigma$ confidence intervals of the ANP and the point estimates of the MLP.

c) Graphical Analysis: Fig. 9 shows the true paths (dashed blue) alongside the red prediction of the ANP and the green prediction of the DR-MLP. In both examples, the ANP more closely follows the curvature and continuity of the ground truth, while the DR-MLP exhibits sharper deviations and jagged corrections.

Fig. 10 breaks each trajectory into its x- and y-components (top: Trajectory 1; bottom: Trajectory 2). The solid red line represents the ANP mean prediction; shaded in light red is the $\pm 2\sigma$ band, and the green line indicates the DR-MLP output. The uncertainty envelope of

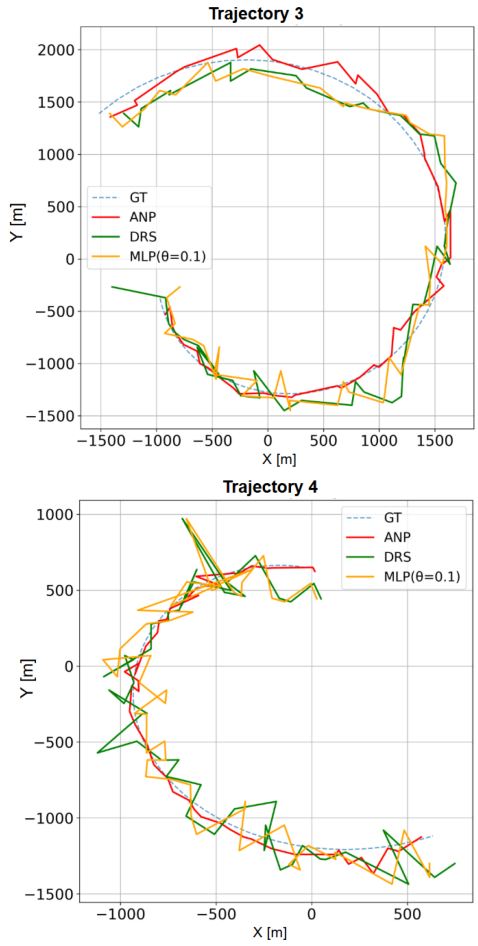

Fig. 11. Qualitative comparison with all baselines: ground truth (dashed blue), specialist MLP (orange), DR-MLP (green), and ANP (red).

the ANP widens in regions of higher curvature or sparse context, plausibly reflecting genuine epistemic uncertainty; the DR-MLP, lacking a probabilistic output, cannot convey such confidence.

We have also extended the overlays to include the prediction of the specialist MLP trained on the same θ in two new trajectories from the $\theta(=0.1)$-stratified subgroup. For brevity, we present representative examples as a two-panel figure in Fig. 11, where we can see how, compared to the DR-MLP, the specialist MLP reduces some local deviations but remains less stable than the ANP.

d) Results and Discussion: Visually, the ANP delivers smoother and more coherent reconstructions: its red traces hug the ground truth continuously, whereas the DR-MLP (green) often overshoots or oscillates (see Fig. 9). The axis-wise plots further reveal that the $\pm 2\sigma$ bands of the ANP reliably expand in segments of high change, offering calibrated uncertainty while remaining tight in well observed regions (see Fig. 10). Quantitatively, over these two examples, the ANP reduces MAE by roughly 25% relative to the DR-MLP, underscoring the benefit of combining a global latent representation with local attention for reliable, uncertainty-aware trajectory estimation.

F. Experiment IV: Impact of Sensor Topology

a) Motivation: Prior sections used the standard ellipsoidal layout inherited from [4]. To assess robustness to the geometry of the array, we added two new topologies to our study: (1) *Random* scatter within the operational area and (2) *Alignment* along the x-axis (poor cross-range diversity). For each topology, we regenerated data with identical trajectories and retrained all models. Fig. 12 shows one validation trajectory and the ten hydrophones for each topology. Axes are in meters; sensors span roughly $[-2, 2]$ km in x and $[-1, 1]$ km in y. We report the MAE (meters) across $\theta \in \{0.0, 0.1, \ldots, 0.5\}$.

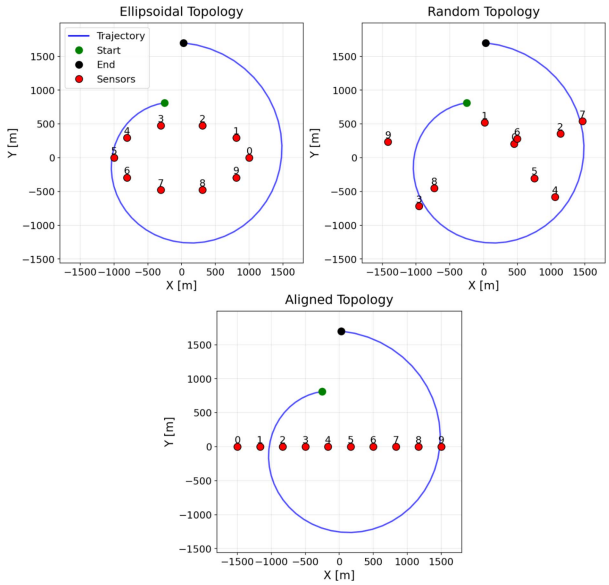

Fig. 12. Same trajectory with the three sensor topologies: Ellipsoidal (left), Random (mid), and Aligned (right). Axes in meters.

TABLE III
MAE IN METERS FOR LOW-VARIANCE DATA ($\theta = 0.0$-0.5): MODELS VERSUS ANP(25%)—**RANDOM** TOPOLOGY

Model	0.0	0.1	0.2	0.3	0.4	0.5
MLP($\theta = 0.0$)	**24.4**	23.8	39.6	108.3	225.0	287.0
MLP($\theta = 0.1$)	26.1	**25.6**	40.5	112.0	232.9	295.3
MLP($\theta = 0.2$)	37.9	35.7	30.9	92.6	222.1	298.5
MLP($\theta = 0.3$)	82.7	94.6	73.1	71.8	180.3	252.1
MLP($\theta = 0.4$)	189.2	220.0	194.3	160.4	156.7	220.3
MLP($\theta = 0.5$)	345.1	357.7	350.4	326.8	276.4	171.9
DR-MLP	33.1	32.8	37.2	65.0	142.5	170.3
ANP(25%)	27.6	26.4	**28.7**	**48.1**	**86.5**	**121.7**

Bold indicates the best value within each row.

b) Discussion: ANP(25%) achieves the lowest mean MAE across all topologies. Aligned layouts are clearly the most challenging due to weak cross-range diversity (geometric dilution); yet, ANP still halves the DR-MLP error on average. *Moreover, the impact of topology is strongest for low-variance channels and diminishes as θ increases.* Inspecting Tables III and IV, for $\theta \in \{0.0, 0.2\}$, the ANP error with an Aligned layout is roughly 3.5 and 2.7 times larger, respectively, than with a Random layout; whereas by θ=0.5, the gap shrinks to about 1.3 times. This trend shows that the position-error lower bound scales with the measurement-noise variance and geometry-dependent factors; as noise/variability rises, the absolute error grows for *all* topologies, and the relative differences attributable to geometry become less pronounced. Consequently, under harsher channels, performance is increasingly dominated by the stochastic disturbance rather than by array geometry, and ANP's advantage stems primarily from its probabilistic conditioning and attention rather than from topology alone [35], [36]. These results confirm that the proposed probabilistic attention framework confers robustness across various deployment geometries.

TABLE IV
MAE IN METERS FOR LOW-VARIANCE DATA ($\theta = 0.0$-0.5): MODELS VERSUS ANP(25%)—**ALLIGNED** TOPOLOGY

Model	0.0	0.1	0.2	0.3	0.4	0.5
MLP($\theta = 0.0$)	202.6	228.3	233.0	257.5	291.9	337.7
MLP($\theta = 0.1$)	195.7	218.0	224.6	253.6	295.7	344.5
MLP($\theta = 0.2$)	202.0	217.8	217.2	245.8	288.6	337.9
MLP($\theta = 0.3$)	219.3	234.9	227.5	230.4	268.5	325.7
MLP($\theta = 0.4$)	279.0	283.6	278.1	265.3	269.4	302.3
MLP($\theta = 0.5$)	312.3	327.1	314.3	299.0	268.9	277.6
DR-MLP	196.0	217.8	219.6	231.5	254.4	281.6
ANP(25%)	**97.0**	**79.0**	**77.7**	**87.8**	**119.1**	**154.7**

Bold indicates the best value within each row.

G. Overall Discussion and Analysis

Across all experiments, the ANP demonstrated clear and consistent advantages over traditional MLP-based methods. In the low-variance setting, the ANP reduced MAE from roughly 65–50 m, a 23% improvement, while in the high-variance scenario, it maintained performance within 10%–15% of its low-variance results despite never having trained on those channel conditions. These quantitative gains are complemented by qualitative improvements: visual inspections of trajectory plots reveal that ANP predictions adhere more closely to the ground truth, producing smoother, more continuous paths compared to the jagged outputs of MLP models.

A key driver of the success of the ANP is its ability to leverage varying amounts of contextual information. When only 4% of trajectory points are observed, the model already matches or outperforms a DR-MLP; as context grows to 20%, the MAE of the ANP declines sharply, after which further increases yield diminishing returns beyond 40%. This context sensitivity highlights an attractive tradeoff: practitioners can balance the cost of data collection (i.e., the number of acoustic observations) against the desired accuracy, stopping once the MAE curve begins to flatten.

The results in Tables I and II vividly illustrate the robustness of the ANP across both low and highvariance θ-stratified subgroups. Unlike MLPs trained on single θ values, which suffer dramatic error spikes when evaluated under mismatched channel conditions, the ANP maintains a relatively flat error surface. Likewise, the context versus MAE plots (see Figs. 7 and 8) confirm that most of the performance benefit is captured with modest context sizes. Finally, Figs. 9 and 10 provide visual evidence of the superior qualitative behaviour of the ANP: its attention-based conditioning unifies global latent structure and local query relevance to generate prediction curves that closely follow the true trajectories.

These findings carry important implications for underwater localization. First, the probabilistic framework of the ANP affords reliable uncertainty quantification and rapid adaptation to new environments, making it particularly suited to real-world operations where channel characteristics can change unpredictably. Second, its data efficiency, achieving high accuracy with limited context, reduces the need for extensive measurement campaigns. Together, the numerical and visual advantages of the ANP suggest it as a compelling alternative to conventional MLP approaches for robust, adaptable, and precise underwater trajectory prediction.

V. CONCLUSION

In this work, we have shown that ANPs represent a powerful and data-efficient alternative to traditional MLP baselines for underwater trajectory prediction. Through extensive experiments on both low- and high-variance acoustic channels, the ANP consistently delivered lower MAEs and smoother trajectory estimates. Even with 4% of a trajectory, the ANP approaches the performance of the DR-MLP; at 8% context, it matches the DR-MLP baseline across most θ -stratified subgroups, and by 15% context, its MAE drops below 35 m for low-variance channels. At 25% context, errors stabilize around 33–34 m for $\theta \leq 0.3$ and remain around 80 m even at $\theta = 0.4$. In high-variance settings, despite never having seen such data during training, the ANP outperforms both subgroup-specific models and DR-MLP by over 15%, exhibiting a flatter, lower error surface as context grows.

These results underscore two fundamental strengths of ANPs. First, their *data efficiency*: a handful of context points delivers dramatic accuracy gains, enabling reliable predictions with minimal observations. Second, their *robust generalization*: the combination of a global latent encoder and a cross-attention deterministic path empowers near zero-shot adaptation to unseen channel conditions. Crucially, the user can tune the *context size* to navigate a direct tradeoff between precision and speed; more samples yield finer predictions but require additional measurements and computation, allowing deployment to be tailored to resource and latency constraints.

Beyond accuracy, ANPs inherently produce calibrated uncertainty estimates, an essential feature for risk-sensitive marine operations, such as search-and-rescue, environmental monitoring, and autonomous navigation. In contrast, deterministic MLPs lack built-in uncertainty modeling and suffer dramatic performance degradation when evaluated outside their narrow training θ-stratified subgroups.

Looking forward, several avenues can further amplify ANPs' impact in underwater localization.

1) *Adaptive context selection:* Dynamically choosing the most informative context points could boost efficiency without sacrificing accuracy.
2) *Enhanced uncertainty modeling:* Integrating more sophisticated probabilistic priors or posthoc calibration methods may tighten confidence intervals under extreme variability.
3) *Real-world deployment:* Validating on field-collected underwater data sets will expose additional challenges, including complex sensor noise, multipath interference, and nonstationary environments. These factors will definitively help guide model refinements.
4) *Cross-domain extensions:* Applying ANPs to 3-D underwater positioning or terrestrial localization tasks can test their generalizability and drive broader adoption.
5) *Front-end detection:* In this project, we assumed that the acoustic window of interest is available to isolate model comparisons. A practical system will integrate a front-end detector/gating stage before feature extraction and ANP inference, and evaluate end-to-end performance in sea trials.

In summary, ANPs deliver a unified, uncertainty-aware framework for fast, accurate, and adaptable trajectory prediction in dynamic underwater environments, laying a strong foundation for both future research and real-world deployment.

APPENDIX

A. Data Preparation

The data preparation process involved several steps to reshape, split, and format the raw acoustic and trajectory data for use with the ANP and MLP models. These steps ensured compatibility with each model's requirements and maintained consistency for fair evaluation.

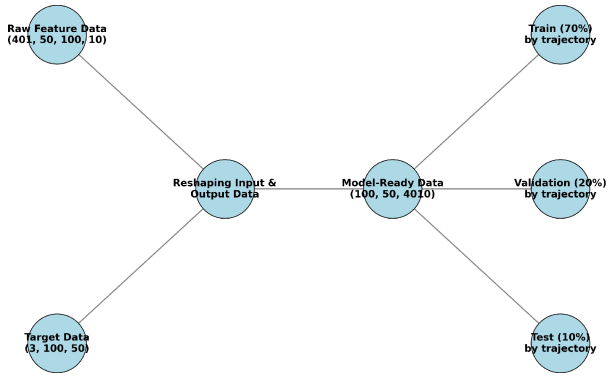

Fig. 13. Data preparation process for ANP and MLP Models. Raw data is reshaped, split into train, validation and test sets, and formatted. Stratified splitting for balanced representation of channel variability (θ).

1) *Raw data reshaping:* Input data, representing acoustic features captured by the sensors, was reshaped from its original dimensions of $(401, 50, 100, 10)$ (time points, points per trajectory, number of trajectories, number of sensors) to $(100, 50, 4010)$ for ANP and MLP use. This format preserves trajectory structure for the ANP while enabling compatibility with regression-based models. The output trajectory data, originally shaped as $(3, 100, 50)$ (coordinates, trajectories, points per trajectory), was reshaped to $(100, 50, 3)$, preserving the trajectory point structure.
2) *Feature processing for MLP:* The reshaped input and output data for the MLP were flattened into input–output pairs, with each trajectory represented as a single feature vector. The final MLP data set dimensions were (5000, 4010) for the input and (5000, 3) for the target labels.
3) *Feature processing for ANP:* For the ANP, the data were prepared in a context-target split format. Context points, randomly selected from the input–output pairs, provided the observed data for training. This probabilistic setup allows the ANP to generalise predictions to unobserved target points and adapt to varying levels of available contextual information.
4) *Data splitting:* The data sets for both ANP and MLP were stratified based on the θ parameter, ensuring a balanced representation of channel variability in all the splits: training (70%), validation (20%), and test (10%). Shuffling was applied to reduce bias and ensure representative training data.
5) *Saving processed data:* Processed data was saved in serialized. pkl files, along with the corresponding θ values. This organization facilitates consistent evaluation across models and channels.

The reshaping and splitting ensured that both models received appropriately formatted data. The stratified splitting based on θ values maintained fairness in model evaluation, highlighting their respective strengths in handling channel variability.

B. Diagram of Data Preparation Workflow

Fig. 13 illustrates the data preparation workflow, showing how raw acoustic data and trajectory information were processed into formats suitable for the ANP and MLP models.

The workflow begins with raw acoustic data and trajectory information, originally structured as $401 \times 50 \times 100 \times 10$ for the input and $3 \times 100 \times 50$ for the output. The data are reshaped to meet the input requirements of the models, ensuring compatibility with their architectures. For the ANP model, the structure preserves the temporal and spatial relationships necessary for context-target splits ($100 \times 50 \times 4010$). For the MLP model, the data are flattened to create feature-target pairs for regression tasks (5000×4010 for inputs and 5000×3 for outputs). The final step involves splitting the processed data into training (70%), validation (20%), and test (10%) subsets, with stratified sampling ensuring that all levels of channel variability (θ) are represented in both subsets.

REFERENCES

[1] I. F. Akyildiz, D. Pompili, and T. Melodia, "Underwater sensor networks: Applications, advances, and challenges," *Philos. Trans. Roy. Soc. A: Mathematical, Phys. Eng. Sci.*, vol. 370, no. 1958, pp. 158–175, 2005.
[2] L. Cheng, S. Zhou, J.-H. Cui, and A. C. Bagtzoglou, "Underwater localization in sparse 3D acoustic sensor networks," in *Proc. IEEE Infocom 27th Conf. Comput. Commun.*, 2008, pp. 798–806.
[3] M. Stojanovic and J. Preisig, "Acoustic underwater communications: The state of the art," *IEEE J. Ocean. Eng.*, vol. 33, no. 4, pp. 112–119, Apr. 2007.
[4] J. Parras, S. Zazo, I. A. Pérez-Álvarez, and J. L. Sanz González, "Model free localization with deep neural architectures by means of an underwater WSN," *Sensors*, vol. 19, no. 16, 2019, Art. no. 3530. [Online]. Available: https://www.mdpi.com/1424-8220/19/16/3530
[5] 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," in *Proc. ICASSP 2023-2023 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2023, pp. 1–5.
[6] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *Proc. 2017 IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)s*, 2017, pp. 23–30.
[7] H. Kim et al., "Attentive neural processes," in *Proc. Int. Conf. Learn. Representations*, 2019.
[8] M. Garnelo et al., "Neural processes," *ICML Workshop Theor. Found. Appl. Deep Generative Models*, 2018.
[9] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
[10] V. Alonso-Eugenio, V. Guerra, S. Zazo, and I. Pérez-Álvarez, "Software-in-loop simulation environment for electromagnetic underwater wireless sensor networks over STANAG 5066 protocol," *Electronics*, vol. 9, no. 10, 2020, Art. no. 1611.
[11] A. Clavijo–Rodrıguez, V. Alonso–Eugenio, S. Zazo, and I. A. Pérez–Álvarez, "Softwareinloop simulation of an underwater wireless sensor network for monitoring seawater quality: Parameter selection and performance validation," *Sensors*, vol. 21, no. 3, 2021, Art. no. 966.
[12] J. Parras and S. Zazo, "Robust deep reinforcement learning for underwater navigation with unknown disturbances," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2021, pp. 3440–3444.
[13] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
[14] Y. Lou, D. Hong, Y. Pan, H.-Y. Hsieh, and K. Bekris, "Learning uncertainty for data-efficient localization with deep neural networks," in *Proc. 2021 IEEE Int. Conf. Robot. Automat. (ICRA)*, 2021, pp. 9467–9473.
[15] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Int. Conf. Mach. Learn. (ICML)*, vol. 70, 2017, pp. 1321–1330. [Online]. Available: https://arxiv.org/abs/1706.04599
[16] J. Huang, S. Park, and O. Simeone, "Calibration-aware Bayesian learning," *IEEE 33rd Int. Workshop Mach. Learn. Signal Process.*, Rome, Italy, pp. 1–6, 2023, doi: 10.1109/MLSP55844.2023.10285894.
[17] V. Fortuin and T. Rochussen, "Sparse gaussian neural processes," in *Proc. 7th Symp. Adv. Approx. Bayesian Inference*, Singapore, Apr. 2025, vol. 289, pp. 194–219.
[18] S. Yan, C. Luo, H. Wang, J. Zhao, and C. Wang, "Learning-based autonomous underwater vehicle localization using deep Gaussian processes," in *Proc. 2021 IEEE Int. Conf. Robot. Automat. (ICRA)*, 2021, pp. 10239–10245.
[19] R. He, W. Zhang, Y. Shen, and Q. Zhao, "Uncertainty-aware learning for underwater acoustic sensor networks," in *Proc. 2021 IEEE Int. Conf. Commun. (ICC)*, 2021, pp. 1–6.

[20] W. Zhang, X. Fang, M. Li, and F. Zuo, "AUV trajectory estimation via probabilistic neural networks," *IEEE Access*, vol. 8, pp. 151154–151164, 2020.

[21] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE Commun. Mag.*, vol. 47, no. 1, pp. 84–89, 2009.

[22] "Automatic identification system (AIS) overview," USCG Navigation Center, Accessed 2025-10. [Online]. Available: https://www.navcen.uscg.gov/automatic-identification-system-overview

[23] Z. Zhang et al., "Multi-domain neural process model based on source attention for anomaly detection of industrial robots," *Control Eng. Pract.*, vol. 135, 2023, Art. no. 105394.

[24] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.

[25] Y. Qin, J. Zhu, J. Qin, W. Wang, and D. Zhao, "Recurrent attentive neural process for sequential data," *NeurIPS Workshop*, 2019, *arXiv:1910.09323*.

[26] P. Qarabaqi and M. Stojanovic, "Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels," *IEEE J. Ocean. Eng.*, vol. 38, no. 4, pp. 701–717, Apr. 2013.

[27] I. F. Akyildiz et al., "State-of-the-art in protocol research for underwater sensor networks," in *Proc. WUWNet, 2006, Surv. UASN Architectures Constraints*.

[28] A. Suresh and S. Srinivasan, "Improved attentive neural processes," *Bayesian Deep Learn. NeurIPS 2019 Workshop*, Vancouver, BC, Canada, pp. 1–10, 2019. [Online]. Available: https://bayesiandeeplearning.org/2019/papers/28.pdf

[29] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, 2017, pp. 5998–6008.

[30] P. Q. Viet and D. Romero, "Spatial transformers for radio map estimation," in *Proc. IEEE Int. Conf. Commun.*, 2025, pp. 6155–6160, doi: 10.1109/ICC52391.2025.11160709.

[31] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," in *Proc. Second Int. Conf. Learn. Representations*, 2014.

[32] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. Int. Conf. Learn. Representations*, 2015.

[33] M. Zecchin, S. Park, O. Simeone, M. Kountouris, and D. Gesbert, "Robust Bayesian learning for reliable wireless ai: Framework and applications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 9, no. 4, pp. 897–912, Aug. 2023, doi: 10.1109/TCCN.2023.3261300.

[34] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, 2014, pp. 1929–1958.

[35] B. Jin, X. Xu, and T. Zhang, "Robust time-difference-of-arrival (TDOA) localization using weighted least squares with cone tangent plane constraint," *Sensors*, vol. 18, no. 3, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/3/778

[36] Y. Zhao, C. Wu, P. Wu, K. Zhu, and X. Deng, "A microwave radiometer residual inversion neural network based on a deadband conditioning model," *J. Mar. Sci. Eng.*, vol. 11, no. 10, 2023, Art. no. 1887. [Online]. Available: https://www.mdpi.com/2077-1312/11/10/1887

Fernando Fernández Salvador received the B.S. degree in technomathematics from the Technische Universität Hamburg, Hamburg, Germany, in 2021, the M.Sc. degree in signal processing and machine learning for Big Data from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2023, and the M.Sc. degree in research in artificial intelligence from the Universidad Nacional de Educación a Distancia, Madrid, in 2025. He is currently working toward the Ph.D. degree in machine learning with UPM.

His research interests include probabilistic deep learning and neural processes.

Juan Parras received the B.S. degree in telecommunications engineering from the Universidad de Jaén, Jaén, Spain, in 2014, and the M.Sc. and Ph.D. degrees in telecommunications engineering from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2016 and 2020, respectively.

He is currently an Assistant Professor with UPM. His research interests include deep generative models, deep reinforcement learning, game theory, and optimization with health and communications applications.

Santiago Zazo is currently a Dr. Engineer with the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1995. In 1998, he joined UPM, where he is currently a Full Professor in signal theory and communications. More recently, he has been mostly focused on deep learning with medical applications, distributed optimization, game theory, and reinforcement learning. He is the author/coauthor of more than 40 journal articles and about 200 conference papers. His main research focuses on signal processing.