# Model-Based Learning of Underwater Acoustic Communication Performance for Marine Robots

George P. Kontoudis\*, Stephen Krauss, Daniel J. Stilwell

Bradley Dept. of Electrical and Computer Engineering Virginia Tech Blacksburg, VA 24061, USA email: {gpkont,eagle138,stilwell}@vt.edu.

# Abstract

Accurate prediction of acoustic communication performance is an important capability for marine robots. In this paper, we propose a model-based learning methodology for the prediction of underwater acoustic communication performance. The learning algorithm consists of two steps: i) estimation of the covariance matrix by evaluating candidate functions with estimated parameters; and ii) prediction of communication performance. Covariance estimation is addressed with a multi-stage iterative training method that produces unbiased and robust results with nested models. The efficiency of the framework is validated with simulations and experimental data from field trials. The field trials involved a manned surface vehicle and an autonomous underwater vehicle.

*Keywords:* Model-based Learning, Autonomous Underwater Vehicles, Wireless Communications, Spatial Statistics, Kriging

# 1 1. Introduction

Coordination of multiple autonomous underwater agents requires effective
 communication for various cooperative missions [1]. For agents that operate
 underwater, inter-vehicle communication is usually accomplished using wire less underwater acoustic (UWA) signals. In the majority of the literature,

<sup>\*</sup>Corresponding author

wireless communication performance is treated as a deterministic, range-6 dependent function [2, 3, 4, 5, 6, 7, 8, 9]. In the graph theory literature 7 this is also known as r-disk communication graph [10, 11, 12, 13, 14, 15, 16]. 8 Indeed, communication performance is a function of vehicle range, but it is 9 also dependent on many other environmental effects, including multi-path 10 propagation and background noise [17]. In addition to the exchange of data, 11 acoustic communication can also provide vehicle range information to im-12 prove navigation, as global positioning system (GPS) is unavailable in subsea 13 environments [18]. 14

Our goal in this work is to predict UWA communication performance 15 at unvisited locations using a set of communication performance measure-16 ments from nearby locations. We employ a two-step learning methodology 17 that comprises: i) the estimation of covariance parameters and the statistical 18 selection of a covariance function; and ii) the prediction of the communica-19 tion performance and its corresponding variance. Intuitively, the two-step 20 process can be interpreted as first training from data, and then predicting 21 the variable of interest at unvisited locations. The estimation of the covari-22 ance function and of its parameters merits special consideration, because it 23 encodes the assumption on a stationary random field and generalizes the 24 properties of the underlying latent process. Accurate predictions of antici-25 pated communication performance can be exploited to plan better utilization 26 of communication resources. Our general approach may be applicable to ter-27 restrial networks, including aerial and ground communication using radio 28 waves. The main idea is to leverage recent advances in spatial statistics and 20 UWA communication modeling, to provide a realistic statistical prediction 30 of inter-vehicle communication performance for teams of marine robots. 31

In underwater wireless sensor networks, kriging (equivalent to Gaussian 32 processes [19, 20]) has been used to model communication performance in 33 several applications. Horner et al. [21], proposed a methodology based par-34 tially on ordinary kriging for the generation of local and global acoustic 35 communication performance maps to facilitate collaborative navigation. A 36 distributed kriging methodology was used in [22] to estimate coverage holes 37 in large-scale wireless sensor networks. The authors in [23] developed a co-38 operative robust algorithm to compose a spatial map of underwater acoustic 39 communication signals and channel parameters using an  $H_{\infty}$  filter and or-40 dinary kriging. In [24], the acoustic communication performance of micro 41 autonomous underwater vehicles (AUVs) was assessed with field trials. The 42 results of the latter reveal that for non-stationary transmission, i.e. moving 43

vehicle, several factors reduce communication performance, including multi-44 path effect of acoustic transmission and the Doppler effect. In [25], a method-45 ology that combines ordinary kriging and compressive sensing methods, was 46 utilized for prediction of acoustic intensity. Prediction of communication 47 performance has been addressed for radio applications. In [26], the authors 48 employ maximum-likelihood estimation for the parameters of the covariance 49 matrix, logarithmic transformation for the underlying mean towards a model-50 based approach, and compressive sensing for prediction with sparse data. In 51 addition, they show that the location of measurements may improve the pre-52 diction quality. In [27], the authors proposed an ordinary kriging prediction 53 framework with detrended data to build radio environment maps and they 54 also considered positional error of the measurements. Gaussian processes 55 have also been used to build communication maps of known terrestrial envi-56 ronments with multiple agents [28]. Specifically the authors used a Gaussian 57 process with constant mean value [19, (2.38), p.27] (equivalent to ordinary 58 kriging) and squared exponential covariance function. Their methodology 59 uses communication priors based on four communication path-loss models to 60 reduce the uncertainty of the communication maps. In the same spirit, in [29] 61 a Gaussian process with fixed mean function and a squared exponential co-62 variance function is proposed to predict the WiFi channel quality and find the 63 optimal relay position for mobile networks. Ordinary kriging assumes that 64 the underlying process is stationary. In addition, in all of these works it was 65 assumed that the covariance model follows a specific theoretical covariance 66 function. In our work, we formulate the problem as a non-stationary random 67 field with universal kriging, which is equivalent to GPs with model-based 68 fixed basis functions [19, (2.41), p.28]. Moreover, we investigate multiple 69 theoretical models for the statistical selection of the covariance function. 70

Communication performance estimation can be used to estimate the po-71 sition of a vehicle. In [30], the authors employed Gaussian processes to de-72 termine a likelihood model of the received signal strength (RSS) for WiFi to 73 estimate the location of robots. This approach requires to compare a training 74 set of RSS observations to a ground truth map, yet this is a computationally 75 demanding process for large maps. To alleviate the computational burden, 76 the authors in [31] used a Gaussian process latent variable model (GP-LVM) 77 to: i) generate the RSS map, ii) compute the position of the vehicle, and iii) 78 build the seafloor map. In these works, only the RSS measurements were 79 used for the construction of RSS maps. In our work, we also use the distance 80 between communicating vehicles to build basis functions for detrending of 81

<sup>82</sup> non-stationary processes.

In [32], we formulated the UWA communication performance problem using multivariate kriging techniques [33], namely cokriging. More specifically, we compared ordinary kriging with ordinary cokriging for the prediction of UWA communication performance. Cokriging provided better results, but the assumption of stationary mean and the lack of parameter estimation for the covariance matrix revealed deficiencies in both prediction techniques.

*Contributions*: The contribution of this paper is twofold. First, we for-89 mulate the problem as a non-stationary random field and propose basis func-90 tions, inspired by the propagation model. The basis functions are then used 91 to detrend the measurements and allow the implementation of stationary 92 kriging. Second, we introduce an iterative technique to identify theoretical 93 models that describe the unknown underwater acoustic environments. Since 94 the covariance of the UWA propagation model is unknown, we compute the 95 parameters of multiple theoretical covariance functions and based on the 96 Bayesian information criterion we select a theoretical model that fits best 97 to the data. To this end, the iterative technique selects the most suitable 98 theoretical covariance model for each environment. 99

Structure: The remainder of this paper is structured as follows. In Section 2 we formulate the problem, Section 3 discusses the parameter estimation of the covariance matrix, Section 4 focuses on the spatial prediction, Section 6 provides the simulations, the experiments, and the results, and
Section 7 concludes the paper and provides future directions.

#### 105 2. Problem Formulation

In this section we discuss the foundations of random fields, describe the problem, and present the UWA communication performance model. In addition, we formulate the problem as a Gaussian random field.

#### 109 2.1. Foundations

The notation here is standard. The set of real numbers is denoted  $\mathbb{R}$ , the set of all positive real numbers  $\mathbb{R}_{>0}$ , and the set of all non-negative real numbers  $\mathbb{R}_{\geq 0}$ . The transpose and inverse operators are denoted  $(\cdot)^{\intercal}$  and  $(\cdot)^{-1}$ respectively. The expectation, the variance and the covariance operators are represented by  $\mathbb{E}[\cdot]$ ,  $\operatorname{Var}[\cdot]$ , and  $\operatorname{Cov}[\cdot, \cdot]$  respectively. The notation  $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  denotes  $\mathbf{y}$  that is drawn from a Gaussian distribution with a vector of means  $\boldsymbol{\mu}$  and covariance matrix  $\boldsymbol{\Sigma}$ . We denote by  $I_n$  the identity matrix of  $n \times n$  dimension. The vector of n zeros is represented as  $\mathbf{0}_n$  and the matrix of  $n \times m$  zeros as  $\mathbf{0}_{n \times m}$ . The hat  $\hat{y}$  denotes the estimated value of y and the superscript in parenthesis  $\hat{y}^{(n)}$  the *n*-th iteration of an estimation process. The cardinality of the set K is denoted  $\operatorname{card}(K)$ , the absolute values is denoted  $|\cdot|$ , and  $||\cdot||$  denotes the  $L_2$  norm.

Next, we introduce basic notions of random fields. For a more in-depth discussion the reader may refer in [34, 35, 36]. A random field is a stochastic process indexed in the Euclidean space. Let  $Z(\mathbf{x})$  be a random field<sup>1</sup> with covariance  $\operatorname{Cov}[Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})]$  for all  $\mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathbb{R}^m$ , where  $\mathbf{x}$  denotes the spatial coordinates,  $\mathbf{h}$  is the separation vector between two locations, and m is the dimension of the coordinates, e.g. m = 2 for planar coordinates. The variogram is a statistical measure of spatial autocorrelation that is defined by,

$$2\gamma(\mathbf{h}) \coloneqq \mathbf{E}\left[\left(Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})\right)^2\right],\tag{1}$$

where  $\gamma(\mathbf{h}) : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$  is a conditionally negative definite function [33] termed as *semivariogram*. The condition ensures that the variance of the random field  $Z(\mathbf{x})$  is positive.

Lemma 1. A semivariogram function  $\gamma : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$  is a conditionally negative definite function if and only if  $\exp\{-\zeta\gamma\}$  is positive definite for all  $\zeta > 0$ .

## 128 **Proof 1.** The proof follows from [37, p. 74].

A random field is *intrinsically stationary* if both  $E[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 0$ 129 and  $\operatorname{Var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})] = 2\gamma(\mathbf{h})$  for all  $\mathbf{x}, \mathbf{x} + \mathbf{h} \in \mathbf{R}^m$  are satisfied. An 130 intrinsically stationary random field with constant mean  $E[Z(\mathbf{x})] = \mu$  and 131  $\operatorname{Cov}[Z(\mathbf{x}), Z(\mathbf{x} + \mathbf{h})] = C(\mathbf{h})$  is called *second-order stationary*. Note that 132 the covariance function  $C(\cdot)$  is a conditionally positive definite function and 133 stationary—depending only on the separation vector  $\mathbf{h}$  and not on spatial 134 coordinates  $\mathbf{x}$ . Second-order stationarity implies intrinsic stationarity and 135 the Gaussian assumption, yet the converse is not always true. 136

<sup>&</sup>lt;sup>1</sup>Throughout the paper, we use the "random field," "random process," and "random function" interchangeably.

For a second-order stationary random field the correlation function is defined by  $\rho(\mathbf{h}) \coloneqq C(\mathbf{h})/C(\mathbf{0}_m)$ , where  $\rho(\mathbf{h}) \in [-1, 1]$  with  $|C(\mathbf{h})| \leq C(\mathbf{0}_m) =$ Var $[Z(\mathbf{x})]$  and  $C(\mathbf{0}_m) = \sigma^2 + \tau^2$  is the *sill* of the semivariogram with  $\sigma^2$  the *partial sill* and  $\tau^2$  the *nugget effect*. The partial sill  $\sigma^2$  is a semivariogram value where no correlation of data further exists and the nugget  $\tau^2$  represents the variance of the data measurement error at a given location.

Given a covariance function  $C(\mathbf{h})$  the variogram (1) yields,

$$2\gamma(\mathbf{h}) = \operatorname{Var}[Z(\mathbf{x} + \mathbf{h}) - Z(\mathbf{x})]$$
  
=  $\operatorname{Var}[Z(\mathbf{x} + \mathbf{h})] + \operatorname{Var}[Z(\mathbf{x})] - 2\operatorname{Cov}[Z(\mathbf{x} + \mathbf{h}), Z(\mathbf{x})]$   
=  $C(\mathbf{0}_m) + C(\mathbf{0}_m) - 2C(\mathbf{h})$   
=  $2(C(\mathbf{0}_m) - C(\mathbf{h})).$  (2)

We cannot always construct the covariance from the variogram, as the variogram may be unbounded. Thus, let us assume that the random field is *ergodic*. That is as  $\|\mathbf{h}\| \to \infty$  then  $C(\mathbf{h}) \to 0$ . In other words, when the distance between two measurements is very large  $\|\mathbf{h}\| \to \infty$ , there is no spatial correlation  $C(\mathbf{h}) \to 0$ . The limit of (2) as  $\|\mathbf{h}\| \to \infty$  yields,

$$C(\mathbf{h}) = \gamma(\infty) - \gamma(\mathbf{h}), \tag{3}$$

where  $\gamma(\infty) = \sup_{\mathbf{h}} \gamma(\mathbf{h}) < \infty$  is non-negative.

When the variogram depends only on the displacement vector norm, i.e.  $2\gamma(\mathbf{h}) = 2\gamma(||\mathbf{h}||)$ , then the variogram is *isotropic*, otherwise it is *anisotropic*.

## 146 2.2. Problem Formulation

We consider the problem of inter-vehicle UWA communication of two ve-147 hicles. In Fig. 1, we illustrate two cases of UWA communication between 148 two vehicles at range r, with  $\mathbf{x}_{t}$  the position of the transmitting vehicle and 149  $\mathbf{x}_r$  the position of the *receiving* vehicle. The first case is shown in Fig. 1-(a) 150 where the success of the communication event depends solely on a maxi-151 mum communication range Q. This means that if the vehicle range exceeds 152 the communication range r > Q, then the communication cannot be accom-153 plished. In practice, this binary approach is unrealistic, as multiple spatially-154 dependent factors may affect the communication of two vehicles, such as 155 scattering, motion-induced Doppler effect, background noise and change of 156 environmental conditions. To this end, we propose multi-dimensional com-157 munication performance maps for various ranges as illustrated in Fig. 1-(b). 158



Figure 1: Communication scenarios of two autonomous underwater vehicles (AUVs) at range r. The transmitting vehicle is located at position  $\mathbf{x}_t$  and the receiving vehicle at position  $\mathbf{x}_r$ . (a) The communication success relies on a deterministic maximum communication range Q. (b) The communication performance using signal-to-noise ratio (SNR) is predicted for specific vehicle ranges.

More specifically, we assess the communication performance of an UWA network of vehicles for specific ranges by modeling the problem as a spatial Gaussian random field with a spatially varying mean. Note that the Gaussian model is a reasonable assumption, as it has been validated with multiple experimental data [38]. For the evaluation of the communication performance we employ signal-to-noise-ratio (SNR) measurements.

Let the SNR measurements be modeled by,

$$Y(\mathbf{x}; v) = \boldsymbol{\mu}(\mathbf{x}; v) + Z(\mathbf{x}; v) + \boldsymbol{\epsilon}(\mathbf{x}), \tag{4}$$

where  $Y(\mathbf{x}; v) \in \mathbb{R}^n$  is the measurement vector describing a non-stationary 165 random field at spatial coordinates  $\mathbf{x} \in \mathbb{R}^2$ ,  $\boldsymbol{\mu}(\mathbf{x}; v)$  is the deterministic mean 166 (or spatial *trend*),  $Z(\mathbf{x}; v) \sim \mathcal{N}(0, \Sigma(\mathbf{x}; v)) \in \mathbb{R}^n$  is a second-order stationary 167 random field with  $\Sigma(\mathbf{x}; v)$  its covariance matrix, and  $\epsilon \sim \mathcal{N}(0, \tau^2 I_n)$  is an 168 independent and identical distributed (iid) zero-mean Gaussian random field. 169 The mean  $\mu$  is the spatial trend that represents large-scale variability, the 170 second-order stationary random field Z captures medium-scale variability, 171 and the white noise  $\epsilon$  is the small-scale variation of the sensor. The surro-172 gate variable is denoted v and is used to represent model dependence, not 173 explicitly accounted for spatial coordinates x. In the Section 2.3, we identify 174 the surrogate variable by using an UWA propagation channel model. 175

176 Assumption 1. The deterministic mean is decomposed by a linear com-

bination of unknown parameters expressed by  $\boldsymbol{\mu}(\mathbf{x}; v) = \mathbf{X}(\mathbf{x}; v)\boldsymbol{\beta}$ , where  $\mathbf{X}_{(\mathbf{x}; v)} \in \mathbb{R}^{n \times p}$  represents the matrix of known basis functions and  $\boldsymbol{\beta} \in \mathbb{R}^{p}$ the vector of the unknown regressor coefficients.

Since the measurements Y are non-stationary, we detrend the measurements, i.e. remove the mean  $Y-\mu$ , to obtain a stationary random field. Next, with the detrended measurements the covariance matrix  $\Sigma$  is estimated with an iterative scheme. After estimating the covariance matrix  $\Sigma$ , we employ the original measurements Y to perform predictions. A critical component for detrending is the basis functions  $\mathbf{X}$ , thus we are inspired by the propagation model to design  $\mathbf{X}$  and accurately detrend the measurements.

**Remark 1.** The major difference between kriging and Gaussian processes (GPs) is that the former computes the covariance function C through the semivariogram function  $\gamma$  (3). In a second-order spatial random field, this intermediate step provides better estimates for three reasons: i) estimation bias [39, pp. 313-320]; ii) boundedness properties [40, pp. 79–84]; and iii) trend contamination [35, pp. 70–73]. Since this paper regards a second-order spatial random field Z with trend  $\mu$ , we find kriging more suitable over GPs.

## 194 2.3. Communication Performance

For communication performance, we use an UWA propagation channel model and its statistical characterization, described in [41, 17, 38]. The statistical model comprises the physical model of the UWA communication channel and random vehicle perturbations which affect the local SNR. Largescale variability of SNR occurs due to large-scale spatial variations in environmental conditions, evoking local error variations and thus a non-stationary random field.

To approximate the communication performance between two agents we use the SNR. In principle, the higher the SNR, the more likely is to detect the signal. In this work we consider fixed signal power, frequency f, and bandwidth B. Let the power of the transmitted signal be constant, then the SNR yields,

$$SNR = \frac{P_{\rm T}G}{P_{\rm N}},\tag{5}$$

where  $P_{\rm T}$  denotes the power of the transmitted signal, G is the channel gain and  $P_{\rm N}$  is the power of noise. The gain G has been shown to follow a lognormal distribution log  $G \sim \mathcal{N}(\bar{G}, \sigma_{\rm G}^2)$ , where  $\bar{G}$  represents the mean of the log channel gain and  $\sigma_{\rm G}^2$  its variance [42, 38]. On the decibel scale, the source level takes the form of  $S_{\rm I}(f) = 10 \log P_{\rm T}$  and the noise level yields  $NL(f, \omega) =$  $10 \log P_{\rm N}$  [17]. If we neglect variations of water pressure with depth, then the gain on the decibel scale  $g = 10 \log G$  is a Gaussian distribution, expressed as,

$$g(r) = \bar{g}(r) + \nu, \tag{6}$$

where  $\nu \sim \mathcal{N}(0, \sigma_{\nu}^2)$  a zero-mean Gaussian random field. The mean follows,

$$\bar{g}(r) = g_0 - k_0 10 \log \frac{r}{r_{\text{ref}}},$$
(7)

where  $g_0$  is a constant gain,  $r_{\rm ref}$  is reference range (e.g., 1 m in our case), and  $k_0$  is the path loss exponent, provided by taking ensemble averages [43]. Ensemble averages is a method to represent the expected value of a waveform.

Note that (4) has identical structure with the model of the UWA propagation channel model (6). Thus, using (7) we choose v to be the range between transmitting and receiving node, i.e. v = r, and the SNR measurements (4) are expressed,

$$Y(\mathbf{x};r) = \mathbf{X}(\mathbf{x};r)\boldsymbol{\beta} + Z(\mathbf{x};r) + \epsilon(\mathbf{x}).$$
(8)

The specific goal of our UWA performance prediction application is summarized in Problem 1.

**Problem 1.** Predict the communication performance  $\hat{Y}$  and the corresponding variance  $Var[\hat{Y}]$  at unvisited locations  $\mathbf{x}_0$ , provided a set of communication performance measurements Y at locations  $\mathbf{x}$  and the vehicle range r.

#### 210 3. Training of Gaussian Random Field

In this section, we formulate basis functions **X** and use least squares on 211 the training data Y to estimate the unknown regressor coefficients  $\beta$  of the 212 spatial trend  $\mu$ . Then, we remove the trend by subtracting the mean  $\mu$  from 213 the measurements Y to retrieve a stationary random field. The detrended 214 measurements  $Y - \mu$  are used to estimate the parameters of multiple vari-215 ogram functions with a maximum likelihood-based method. Next, we select 216 the most suitable variogram model, based on the Bayesian information crite-217 rion. With the selected variogram model we construct the covariance matrix 218

<sup>219</sup>  $\Sigma$  and use generalized least squares to improve the accuracy of the spatial <sup>220</sup> trend estimator  $\mu$ . The method iterates until the parameters of the vari-<sup>221</sup> ogram function converge.

#### 222 3.1. Spatial Trend Modeling

The random field in (8) is non-stationary due to the spatial trend. Thus, the original measurements cannot be used to estimate the parameters of the variogram. To this end, we seek basis functions **X** to model the spatial trend  $\mu$ , detrend the measurements  $Y - \mu$ , and recover stationarity.

A precise model of the trend is important for spatial extrapolation, ideally arising from the physics of the system [44]. The obvious choice for the elements of the basis function  $\mathbf{X}$  is to employ spatial coordinates as covariates. In spatial statistics, polynomial basis functions of spatial coordinates, e.g.,  $\mathbf{X}(\mathbf{x}) = [1, x, y, xy, x^2, y^2]$ , are often employed [35]. However, polynomial basis functions do not behave well for extrapolation, because they are radially unbounded, i.e. as  $\|\mathbf{x}\| \to \infty$  then  $X(\mathbf{x}) \to \infty$ . To this end, Gaussian radial basis functions (RBF) are widely used in various applications [45], as they provide suitable extrapolation results. In addition, surrogate variables—arising from the physical model of the system—are useful covariates to interpret the behavior of the spatial variation [44]. A Gaussian RBF is described by,

$$X_l(\mathbf{x}; c_l, \sigma_{\mathrm{G},l}^2) = \exp\left(-\frac{(\mathbf{x} - c_l)^2}{2\sigma_{\mathrm{G},l}^2}\right),\tag{9}$$

where  $c_l$  is the center of each measurement, e.g.,  $c_l = 0$  for zero mean measurement error  $\epsilon$  (8). The corresponding variance is denoted  $\sigma_{G,l}^2$ , where in practice is a constant value  $\sigma_{G,l}^2 = \sigma_G^2$  for all l measurements. From (7), it is deduced that the range of the vehicles has a linear-log relationship to the mean. Hence, our proposed hybrid basis function combines Gaussian RBF incorporating spatial coordinates (9) and linear-log range,

$$\mathbf{X}(\mathbf{x};r) = [1, \exp\left(-\frac{(x-c_x)^2}{2\sigma_x^2}\right), \exp\left(-\frac{(y-c_y)^2}{2\sigma_y^2}\right), r, \log r].$$
(10)

For data detrending, since the covariance function is unknown, the generalized least squares (GLS) cannot be used. Thus, we initially estimate the unknown parameters using ordinary least squares (OLS),

$$\widehat{\boldsymbol{\beta}}_{\text{OLS}}^{(1)} = \mathbf{X}(\mathbf{x}; r)^{\dagger} Y(\mathbf{x}; r), \tag{11}$$

where  $\mathbf{X}^{\dagger} = (\mathbf{X}^{\intercal}\mathbf{X})^{-1}\mathbf{X}^{\intercal}, \ \mathbf{X}^{\dagger} \in \mathbb{R}^{p \times n}$  is the Moore-Penrose pseudoinverse of  $\mathbf{X}$ . The estimated unknown parameters  $\widehat{\boldsymbol{\beta}}_{\text{OLS}}^{(1)}$  are not the final estimated unknown regressor values. Instead, we shall employ  $\widehat{\boldsymbol{\beta}}_{\text{OLS}}^{(1)}$  to detrend the measurements and assess their behavior with an iterative technique. The Gaussian residual random field (or detrended data) is expressed,

$$\tilde{Y}(\mathbf{x};r) = Y(\mathbf{x};r) - \mathbf{X}(\mathbf{x};r)\hat{\boldsymbol{\beta}}_{\text{OLS}}^{(1)}.$$
(12)

Assumption 2. The random field of the underlying latent process is secondorder stationary after detrending, i.e.  $\tilde{Y}$  is second-order stationary.

## **Assumption 3.** The variogram function is isotropic after detrending.

## 230 3.2. Experimental Semivariogram and Theoretical Models

In this section, we present three commonly used semivariograms and an optimization method to estimate the initial parameters of the semivariogram function. The Matheron empirical semivariogram [46] is used in the majority of the literature for the estimation of the unknown parameters,

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2 \operatorname{card}(N(\mathbf{h}))} \sum_{N(\mathbf{h})} |\tilde{Y}(\mathbf{x} + \mathbf{h}) - \tilde{Y}(\mathbf{h})|^2,$$

where  $N(\mathbf{h}) = \{(o, p) \mid \mathbf{x}_o - \mathbf{x}_p = \mathbf{h}\}$  is the set of measurements at distance  $\mathbf{h}$  and  $\tilde{Y}$  is the vector of the residual measurements (12). The main idea is to compute the experimental semivariogram from the detrended data and then compare it to theoretical semivariogram models. The Matheron empirical semivariogram is unbiased, yet it is highly affected by outliers, due to the squared term. A robust estimator of the experimental semivariogram against outliers is proposed in [47] as,

$$\hat{\gamma}_{\rm CH}(\mathbf{h}) = \frac{\left(\frac{\sum_{N(\mathbf{h})}|\tilde{Y}(\mathbf{x}+\mathbf{h}) - \tilde{Y}(\mathbf{h})|^{1/2}}{\operatorname{card}(N(\mathbf{h}))}\right)^4}{0.914 + \frac{0.988}{2\operatorname{card}(N(\mathbf{h}))} + \frac{0.090}{\operatorname{card}(N(\mathbf{h}))^2}}.$$
(13)

The robustness relies on a transformation which ensures that the fourth root of the transformed distribution produces relatively small skew. Note that we cannot interpolate the experimental semivariogram to obtain a semivariogram, because the conditional negative definiteness property may be violated [40]. Instead, we fit the experimental semivariogram to theoretical models that ensure the desired properties of a semivariogram function.

We consider three potential theoretical semivariogram models which are conditional negative definite. The spherical semivariogram is given by,

$$\gamma_{\rm s}(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} \tau^2 + \sigma^2, & \|\mathbf{h}\| \ge \alpha, \\ \tau^2 + \sigma^2 \left(\frac{3\|\mathbf{h}\|}{2\alpha} - \frac{1}{2} \left(\frac{\|\mathbf{h}\|}{\alpha}\right)^3\right), & \|\mathbf{h}\| \le \alpha, \end{cases}$$
(14)

where the semivariogram parameter vector  $\boldsymbol{\theta} = [\tau^2 \sigma^2 \alpha]^{\mathsf{T}} \in \Theta$  contains the nugget, the partial sill, and the semivariogram range with  $\Theta = \{\boldsymbol{\theta} \in \mathbb{R}^3 \mid \tau^2 \geq 0, \sigma^2 \geq 0, \alpha \geq 0\}$  the parameter space. Second, the exponential semivariogram function,

$$\gamma_{\rm e}(\mathbf{h};\boldsymbol{\theta}) = \tau^2 + \sigma^2 \Big( 1 - \exp\left\{-\frac{\|\mathbf{h}\|}{\alpha}\right\} \Big).$$
(15)

Finally, the Matérn semivariogram function [48],

$$\gamma_{\rm m}(\mathbf{h};\boldsymbol{\theta}) = \tau^2 + \sigma^2 \left( 1 - \frac{(\|\mathbf{h}\|/\alpha)^{\kappa}}{2^{\kappa-1}\Gamma(\kappa)} K_{\kappa} \left(\frac{\|\mathbf{h}\|}{\alpha}\right) \right),$$

where  $\Gamma(\cdot)$  is the gamma function,  $K_{\kappa}$  is the Bessel function of order  $\kappa$ , and  $\kappa$  is the smoothing parameter. The Matérn semivariogram function is a general model, thus we fix the smoothing parameter at  $\kappa = 3/2$  to obtain a mixed polynomial-exponential form,

$$\gamma_{\rm pe}(\mathbf{h};\boldsymbol{\theta}) = \tau^2 + \sigma^2 \left( 1 - \left( 1 + \frac{\sqrt{3} \|\mathbf{h}\|}{\alpha} \right) \exp\left\{ - \frac{\sqrt{3} \|\mathbf{h}\|}{\alpha} \right\} \right). \tag{16}$$

<sup>237</sup> We will employ all semivariogram functions  $C = \{\gamma_s, \gamma_e, \gamma_{pe}\}$  and evaluate <sup>238</sup> their performance.

The next step is to formulate an optimization problem to fit the models C and derive the corresponding parameter vector  $\boldsymbol{\theta}$ . We utilize a weighted

least squares (WLS) approach [49] which yields,

$$\widehat{\boldsymbol{\theta}}_{\text{CWLS}}^{(0)} = \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\Theta} \sum_{g=1}^{N_g} \operatorname{card}(N(\mathbf{h}_g)) \left(\frac{\widehat{\gamma}_{\text{CH}}(\mathbf{h}_g)}{\gamma(\mathbf{h}_g;\boldsymbol{\theta})} - 1\right)^2,\tag{17}$$

where  $N_g$  is the total number of the separation vectors  $\mathbf{h}_g$ . The parameter estimation (17) relies on the residual measurements  $\tilde{Y}$ (12) which incorporate measurement bias. Thus, the estimation is sensitive to the bias of the mean value.

## 243 3.3. Unbiased Semivariogram Model Fitting

In this section, we seek an unbiased estimator for the parameter vector  $\boldsymbol{\theta}$ and a strategy to narrow down the parameter space  $\Theta$ . Maximum likelihood (ML) estimation is used widely in statistics. In spatial statistics, due to high correlation of the observations, ML is known to generate unfavorable outcomes [50]. In addition, when the observations are limited, then the bias of the ML estimation is significant.

An alternative bias-free approach is the restricted maximum likelihood (REML) estimation [51, 52], which makes use of error contrasts to remove the mean dependence from the variance estimates. The main idea is to transform the residual measurements  $\tilde{Y}$  from (12) with a matrix  $\mathbf{A} \in \mathbb{R}^{n \times (n-p)}$  such that,  $\mathbf{A}^{\mathsf{T}} \mathbf{X} = \mathbf{0}_{(n-p) \times p}$  and  $\mathbf{E}[\mathbf{A}^{\mathsf{T}} \tilde{Y}] = \mathbf{0}_{(n-p)}$ , where  $\mathbf{X}$  is the basis function (10). In other words, each column vector of matrix  $\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_{(n-p)}]$  is orthogonal to all columns of  $\mathbf{X}$ . Let us define the error contrast,  $W \coloneqq \mathbf{A}^{\mathsf{T}} \tilde{Y}$  to obtain  $W \sim \mathcal{N}(\mathbf{0}_{(n-p)}, \mathbf{A}^{\mathsf{T}} \mathbf{\Sigma}(\boldsymbol{\theta}) \mathbf{A})$ . Although  $\mathbf{A}$  is not unique, a matrix that satisfies the properties is the orthogonal projection onto the kernel of  $\mathbf{X}$ , that is,  $\mathbf{A} = I_n - \mathbf{X}(\mathbf{X}^{\mathsf{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathsf{T}}$ . We note that  $\mathbf{A}$  does not depend on the estimated mean parameters  $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ . Therefore, the log-restricted likelihood function is defined,

$$L(\boldsymbol{\theta}|W) = -\frac{1}{2} \Big( (n-p)\log(2\pi) + \log|\mathbf{X}^{\mathsf{T}}\mathbf{X}| - \log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \log|\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{X}| - \tilde{Y}^{\mathsf{T}}\boldsymbol{\Pi}(\boldsymbol{\theta})\tilde{Y} \Big),$$
(18)

where  $\Pi(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{X} (\mathbf{X}^{\mathsf{T}} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{X})^{-1} \mathbf{X}^{\mathsf{T}} \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$ , *n* is the measurement vector size, and *p* is the rank of **X**. Next, the log-restricted likelihood (18) is maximized with respect to  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$  to obtain the estimated <sup>253</sup> parameter vector  $\hat{\theta}$ . To reduce the search of the parameter space  $\Theta$ , we use <sup>254</sup> the parameter estimate  $\hat{\theta}_{CWLS}^{(0)}$  (17) as a center value of the initial set of pa-<sup>255</sup> rameters in the optimization scheme. So far we computed three covariance <sup>256</sup> parameter vectors  $\hat{\theta}$  corresponding to three candidate models (14), (15), (16). <sup>257</sup> A benefit of likelihood-based approaches is that they can be combined with <sup>258</sup> statistical model selection tools [53].

## 259 3.4. Statistical Model Selection

The Bayesian information criterion (BIC) is a statistical model selection methodology, introduced by Schwarz in [54]. The BIC is given by,

$$BIC(M_k) = -2\ln \mathcal{L}(\widehat{\theta}_k \mid \tilde{Y}, M_k) + q\ln n, \qquad (19)$$

where  $\mathcal{M} = \{M_k = \Sigma(\widehat{\theta}_k) \mid k = 1, ..., K\}$  is the set of candidate models,  $\widehat{\theta}_k$ 260 denotes the REML estimates of  $\boldsymbol{\theta}_k$ , q = 3 is the dimension of the parameter 261 space  $\Theta$ ,  $\mathcal{L}(\theta_k \mid \tilde{Y}, M_k)$  represents the marginal likelihood corresponding 262 to the density function  $f(\tilde{Y}, M_k \mid \widehat{\theta}_k)$ , and n is the measurement size of 263 the vector  $\tilde{Y}$ . In our case K = 3 corresponds to three different candidate 264 semivariogram functions (14), (15), and (16). In principle, the semivariogram 265 function with the smallest BIC represents the true model, assuming that the 266 real model is listed among the candidate covariance models. One of the major 267 advantages of the BIC is that it satisfies the property of *consistency*. That 268 is even if the true model is not listed among the candidate models, the BIC 269 selects the most parsimonious model closest to the true model, by computing 270 the marginal likelihood with Laplace approximation. 271

Since the BIC (19) is computed in the log-scale, its evaluation may be ambiguous. Thus, we employ the posterior probability of the BIC [55] which is approximated by,

$$P(M_k \mid \tilde{Y}) \approx \frac{\exp\left(-\frac{1}{2}\Delta_k\right)}{\sum_{k=1}^{K} \exp\left(-\frac{1}{2}\Delta_k\right)},\tag{20}$$

where  $\Delta_k = \text{BIC}(M_k) - \text{BIC}^*$  denotes the BIC difference of a candidate model with the minimum BIC candidate model  $\text{BIC}^* = \min_{M_k \in \mathcal{M}} \text{BIC}(M_k)$ . Essentially,  $P(M_k \mid \tilde{Y})$  is a probability mass function, that provides a probability of suitability for each model to the real model.

#### 276 3.5. Nested Semivariogram Model

So far we assumed that the variation of the underlying process is purely represented by either a spherical (14), or an exponential (15), or a polynomialexponential (16) variogram model. However, in many cases, the spatial variability is more complex, and thus a combination of semivariogram models interprets the latent process more precisely. The nested [40] (or compositional [56]) semivariogram function is defined by,

$$\gamma_{\text{nest}}(\mathbf{h};\widehat{\boldsymbol{\theta}}_{\text{s},k},\widehat{\boldsymbol{\theta}}_{\text{e},k},\widehat{\boldsymbol{\theta}}_{\text{pe},k}) \coloneqq \xi_1 \gamma_{\text{s}}(\mathbf{h};\widehat{\boldsymbol{\theta}}_{\text{s},k}) + \xi_2 \gamma_{\text{e}}(\mathbf{h};\widehat{\boldsymbol{\theta}}_{\text{e},k}) + \xi_3 \gamma_{\text{pe}}(\mathbf{h};\widehat{\boldsymbol{\theta}}_{\text{pe},k}),$$
(21)

277 where  $\xi_k \in (0, 1)$ , and  $\sum_{k=1}^{K} \xi_k = 1$ .

Proposition 1. Any convex combination of semivariograms is a semivari ogram.

**Proof 2.** Let  $\gamma_k$  be a semivariogram and  $\gamma_{-k} = {\gamma_l}_{l \neq k}$  a vector of semivariograms other than  $\gamma_k$ . Since  $\gamma : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$  and  $\xi_k \in (0,1)$ , then  $\gamma_{\text{nest}} = \sum_{k=1}^{K} \xi_k \gamma_k > 0$  for  $||\mathbf{h}|| \neq 0$ . Moreover,  $\exp\{-\zeta \gamma_{nest}\}$  is positive definite for all  $\zeta > 0$ . Hence, from Lemma 1 any convex combination of variograms  $\gamma_{nest}$  is a variogram.

The nested semivariogram is similar in spirit to [56], yet the authors used directly the BIC. Since the BIC is in the log-scale (19), it does not scale well with the nested semivariogram. Alternatively, we employ the posterior probabilities of BIC  $\xi_k = P(M_k \mid \tilde{Y})$  that satisfy  $\xi_k \in (0, 1)$  and  $\sum_{k=1}^{K} \xi_k = 1$ .

## 289 3.6. Iterative Parameter Training

For the iterative parameter training we utilize the estimated covariance matrix  $\Sigma(\hat{\theta}^{(1)})$ . The covariance matrix allows the implementation of the generalized least squares (GLS) to improve the estimation of the mean. The GLS mean estimate is described by,

$$\widehat{\boldsymbol{\beta}}_{\text{GLS}}^{(2)} = \left( \mathbf{X}^{\mathsf{T}} \boldsymbol{\Sigma} \left( \widehat{\boldsymbol{\theta}}^{(1)} \right)^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^{\mathsf{T}} \boldsymbol{\Sigma} \left( \widehat{\boldsymbol{\theta}}^{(1)} \right)^{-1} Y.$$
(22)

Sequentially, the residual measurements (12) yield,

$$\tilde{Y}(\mathbf{x};r) = Y(\mathbf{x};r) - \mathbf{X}(\mathbf{x};r)\hat{\boldsymbol{\beta}}_{\text{GLS}}^{(2)}.$$
(23)

In addition, the GLS mean estimation facilitates a more accurate determination of the covariance function. To this end, we employ the detrended measurements (23) and iterate the covariance training. The training is terminated when,

$$\|\widehat{\boldsymbol{\theta}}^{(s)} - \widehat{\boldsymbol{\theta}}^{(s-1)}\| \le \eta \tag{24}$$

where  $\eta \in \mathbb{R}_{>0}$  is a small error threshold. At every iteration we expect lower BIC values (19). Essentially, after the second iteration, the change on the mean and covariance estimate is insignificant [40, pp. 196–200], [57, 58], and usually the training is terminated.

## **4.** Spatial Prediction

In this section, we describe universal kriging [35, 59, 33], a spatial prediction technique that predicts values at locations of interest, based on measurements from other locations Y and the estimated covariance matrix  $\Sigma$ . The main difference from the ordinary kriging lies in the mean value of the random field, which is not assumed to be constant. More specifically, provided measurements Y at locations  $\mathbf{x} \in \mathbb{R}^2$  the random field is described by (8). We use a linear unbiased estimator,

$$\hat{Y}(\mathbf{x}_0; r) = \sum_{i=1}^n \omega_i Y(\mathbf{x}_i; r) = \boldsymbol{\omega}^{\mathsf{T}} Y(\mathbf{x}; r),$$
(25)

where  $\mathbf{x}_0 \in \mathbb{R}^2$  is the location of interest,  $\boldsymbol{\omega} = [\omega_1 \dots \omega_n]^{\mathsf{T}} \in \mathbb{R}^n$  are the weights we seek to obtain, and  $Y(\mathbf{x}; r)$  are the raw measurements, i.e. not the residuals. The unbiasedness of the predictor is ensured by  $\mathrm{E}[\hat{Y}(\mathbf{x}_0; r) - Y(\mathbf{x}_0; r)] = 0$ , that yields a system of equations known as universality conditions,  $\boldsymbol{\omega}^{\mathsf{T}}\mathbf{X} = X_0^{\mathsf{T}}$ , where  $X_0 \in \mathbb{R}^p$  is the vector of known basis functions at the location of interest. Next, we formulate the unconstrained minimization problem of the prediction variance with multiple Lagrange multipliers  $\boldsymbol{\lambda} \in \mathbb{R}^p$  to include the universality conditions. The solution is,  $\boldsymbol{\omega}_{\mathrm{UK}} = \Gamma_{\mathrm{UK}}^{-1}\boldsymbol{\gamma}_{\mathrm{UK}}$ , where  $\boldsymbol{\omega}_{\mathrm{UK}} = [\boldsymbol{\omega}^{\mathsf{T}} \boldsymbol{\lambda}_{\mathrm{UK}}^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^{n+p}$  is a stacked vector that contains the weights  $\boldsymbol{\omega}$  and the Lagrange multipliers  $\boldsymbol{\lambda}_{\mathrm{UK}}$  to minimize the mean square prediction error. The non-singular matrix  $\Gamma_{\mathrm{UK}} \in \mathbb{R}^{(n+p) \times (n+p)}$  captures the *redundancy* 

of measurements and is given by,

$$\boldsymbol{\Gamma}_{\text{UK}} = \begin{bmatrix} \gamma(\mathbf{x}_{1}, \mathbf{x}_{1}) & \dots & \gamma(\mathbf{x}_{1}, \mathbf{x}_{n}) & 1 & X_{2}(\mathbf{x}_{1}) & \dots & X_{p}(\mathbf{x}_{1}) \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma(\mathbf{x}_{n}, \mathbf{x}_{1}) & \dots & \gamma(\mathbf{x}_{n}, \mathbf{x}_{n}) & 1 & X_{2}(\mathbf{x}_{n}) & \dots & X_{p}(\mathbf{x}_{n}) \\ 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ X_{2}(\mathbf{x}_{1}) & \dots & X_{2}(\mathbf{x}_{n}) & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{p}(\mathbf{x}_{1}) & \dots & X_{p}(\mathbf{x}_{n}) & 0 & 0 & \dots & 0 \end{bmatrix} \\ \coloneqq \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{X} \\ \mathbf{X}^{\mathsf{T}} & \mathbf{0}_{p \times p} \end{bmatrix},$$

The semivariogram vector  $\boldsymbol{\gamma}_{\mathrm{UK}} \in \mathbb{R}^{(n+p)}$  considers the *closeness* of the measurements to the location of interest  $\mathbf{x}_{0}$ ,

$$\boldsymbol{\gamma}_{\mathrm{UK}} = \begin{bmatrix} \gamma(\mathbf{x}_0, \mathbf{x}_1) & \dots & \gamma(\mathbf{x}_0, \mathbf{x}_n) & 1 & X_2(\mathbf{x}_0) & \dots & X_p(\mathbf{x}_0) \end{bmatrix}^{\mathsf{T}} \coloneqq \begin{bmatrix} \boldsymbol{\gamma}_0 \\ X_0 \end{bmatrix}.$$

The decoupled coefficients in terms of the covariance matrix yield,

$$\boldsymbol{\omega}^{\mathsf{T}} = \left(\mathbf{c}_0 + \mathbf{X}(\mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}(X_0 - \mathbf{X}^{\mathsf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{c}_0)\right)^{\mathsf{T}}\boldsymbol{\Sigma}^{-1},\tag{26}$$

with Lagrange multipliers,

$$\boldsymbol{\lambda}_{\mathrm{UK}}^{\mathsf{T}} = -(X_0 - \mathbf{X}^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \mathbf{c}_0)^{\mathsf{T}} (\mathbf{X}^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}.$$
(27)

Hence, the predictive distribution of UK with a covariance matrix is,

$$\hat{Y} \mid Y, \mathbf{x}, r \sim \mathcal{N} \Big( [\mathbf{c}_0 \Sigma^{-1} + (X_0 - \mathbf{c}_0 \Sigma^{-1} \mathbf{X}) (\mathbf{X}^{\mathsf{T}} \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^{\mathsf{T}} \Sigma^{-1}] Y,$$

$$C(\mathbf{0}_m) - \mathbf{c}_0 \Sigma^{-1} \mathbf{c}_0^{\mathsf{T}} + (X_0 - \mathbf{c}_0 \Sigma^{-1} \mathbf{X}) (\mathbf{X}^{\mathsf{T}} \Sigma^{-1} \mathbf{X})^{-1} (X_0 - \mathbf{c}_0 \Sigma^{-1} \mathbf{X})^{\mathsf{T}} \Big)$$
(28)

## <sup>295</sup> 5. Model-Based Learning Framework

In this section, we discuss the structure and the algorithm of the communication performance prediction technique.



Figure 2: The two-step learning process. The first step is the training of the Gaussian random field that yields a covariance matrix and the second step the spatial prediction of the communication performance.

## <sup>298</sup> 5.1. Learning Structure

The two-step process is depicted in Fig. 2. We start by collecting 299 measurements of communication performance (SNR) along with the vehi-300 cle range. Given those measurements we seek to predict the communication 301 performance at unvisited locations. The first step is the training of the Gaus-302 sian random field to obtain a covariance matrix, while the second is spatial 303 prediction at unvisited locations with universal kriging. The objective of 304 the first step is to determine the most suitable covariance function and its 305 parameters characterizing the underlying latent process. The block of the 306 covariance matrix is depicted in light red. The goal of the second step is to 307 predict the SNR at unvisited locations and its corresponding variance, where 308 their blocks are depicted in light red accordingly. 300

The training step comprises three modules: i) the data detrending; ii) the parameter estimation; and iii) the iterative training. The data detrending includes the hybrid basis function formulation (10) and the OLS computation (11). Next, the detrended measurements are used to compute the candidate semivariogram functions (14), (15), (16). The semivariograms are provided

to the estimation module which is also a multistage process. The estima-315 tion module first computes the covariance parameters to be used as initial 316 conditions, by employing the Cressie and Hawkins robust experimental semi-317 variogram (13) and a weighted least squares estimation with Cressie weights 318 (17). The next stage is the REML estimation that optimizes the objective 319 likelihood function (18) and results in three bias-free covariance parameter 320 vectors. The last stage of the estimation module considers the selection of 321 the most suitable covariance model among the three candidates with the 322 posterior BIC (20). Whenever the posterior probability of BIC indicates 323 suitability of less than a probability threshold, we compute a nested semi-324 variogram. The last module describes an iterative training for the selection 325 of the covariance matrix. Since we have obtained a covariance matrix, the 326 mean estimates can be improved by computing the GLS (22). Subsequently, 327 we recompute the residual random function and run the estimation module 328 to obtain a new covariance matrix. The training iterates until the parame-329 ters of the covariance matrix converge (24). For the numerical experiments 330 reported herein, convergence requires no more than two iterations. 331

The second step is the spatial prediction. Given the measurements, the model-based basis functions, and the covariance matrix from the previous step we use the location of interest to solve the universal kriging and obtain the kriging weights (26),(27). Finally, we predict the SNR at the location of interest and corresponding SNR variance (28).

#### 337 5.2. Algorithm

The main routine of the communication performance predictor is pre-338 sented in Algorithm 1. The initialConditions module assigns initial val-339 ues to the semivariogram parameter vector  $\widehat{\boldsymbol{\theta}}^{(0)}$ . More specifically, the partial 340 sill  $\sigma^2$  is assumed to be the variance of the residual measurements (12), the 341 nugget effect  $\tau^2$  and the semivariogram range  $\alpha$  are selected according to 342 the sensor sensitivity and characteristics respectively. The initial covariance 343 matrix estimate  $\Sigma(\widehat{\theta}^{(0)})$  is set equal to the identity matrix. Next, the algo-344 rithm proceeds to the iterative parameter estimation process. We consider 345 three semivariogram functions (14), (15), (16) at each iteration. The **basis** 346 function computes  $\mathbf{X}$  according to (10). The GLS function implements the 347 GLS (22) to estimate the mean regressor parameters  $\widehat{\boldsymbol{\beta}}^{(s)}$ . Note that in the 348 first iteration the initial covariance matrix is the identity matrix, and hence 349 the algorithm implements an OLS regression (11). The function detrend is 350

**Algorithm 1** Learning of UWA Communication Performance

**Input:** Y, x, r,  $x_0$ , n, p, q,  $\gamma$ ,  $\varphi$   $\eta$ **Output:**  $\widehat{Y}$ ,  $Var[\widehat{Y}]$  $1: \ \widehat{\boldsymbol{\theta}}^{(0)} \gets \texttt{initialConditions}(Y)$ 2:  $\Sigma(\widehat{\boldsymbol{\theta}}^{(0)}) \leftarrow I_n; k \leftarrow 0;$ 3:  $\mathbf{X} \leftarrow \texttt{basis}(\mathbf{x}; r);$ 4: for s = 1 to S do ▷ Start training  $\widehat{\boldsymbol{\beta}}^{(s)} \leftarrow \mathtt{GLS}(Y, \mathbf{X}, \boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}^{(s-1)}));$ 5:  $\tilde{Y}^{(s)} \gets \texttt{detrend}(Y, \mathbf{X}, \widehat{\boldsymbol{\beta}}^{(s)});$ ▷ Non-stationarity 6: 
$$\begin{split} & \text{for each } \boldsymbol{\gamma} \in \mathcal{C} \text{ do} \\ & \widehat{\boldsymbol{\theta}}_k^{(s-1)} \leftarrow \text{CWLS}(\tilde{Y}, \boldsymbol{\gamma}, \widehat{\boldsymbol{\theta}}^{(s-1)}); \\ & \widehat{\boldsymbol{\theta}}_k^{(s)} \leftarrow \text{REML}(\tilde{Y}, \mathbf{X}, n, p, \boldsymbol{\gamma}, \widehat{\boldsymbol{\theta}}_k^{(s-1)}); \end{split}$$
7: 8:  $\triangleright$  Robustness 9: ▷ Unbiasedness  $M_k \leftarrow \mathbf{\Sigma}(\widehat{\boldsymbol{\theta}}_k^{(s)});$ 10:  $\operatorname{BIC}_k \leftarrow \operatorname{BIC}(\tilde{Y}, n, q, \widehat{\boldsymbol{\theta}}_k^{(s)}, M_k);$ 11:  $k \leftarrow k + 1;$ 12:end for 13: $\operatorname{BIC}^{\star} \leftarrow \min_{M_k \in \mathcal{M}} \{\operatorname{BIC}(M_k)\};$ 14:for k = 1 to K do 15: $\Delta_k \leftarrow \texttt{diffBIC}(BIC_k, BIC^*);$ 16:end for 17:for k = 1 to K do  $\triangleright$  Model selection 18: $P(M_k \mid \tilde{Y}) \leftarrow \texttt{postBIC}(\Delta_k);$ 19: end for 20:if  $\max_{M_k \in \mathcal{M}} \{ P(M_k \mid \tilde{Y}) \} < \varphi$  then  $\Sigma(\hat{\theta}^{(s)}) \leftarrow \operatorname{nested}(P(M_k \mid \tilde{Y}), \hat{\theta}_k^{(s)});$ 21: ▷ Covariance 22: else 23: $\boldsymbol{\Sigma}(\widehat{\boldsymbol{\theta}}^{(s)}) \leftarrow \max_{M_k \in \mathcal{M}} \{ P(M_k \mid \widetilde{Y}) \};$ 24: $\begin{array}{l} \mathbf{end} \ \mathbf{if} \\ \mathbf{if} \ \| \widehat{\boldsymbol{\theta}}^{(s)} - \widehat{\boldsymbol{\theta}}^{(s-1)} \| \leq \eta \ \mathbf{then} \end{array}$ 25:26: $\triangleright$  Iteration criterion 27:break; 28:end if 29: end for  $\triangleright$  End training 30:  $\widehat{Y}$ ,  $\operatorname{Var}[\widehat{Y}] \leftarrow \operatorname{UK}(Y, \mathbf{x}, r, \mathbf{X}, \mathbf{x}_0, \Sigma(\widehat{\boldsymbol{\theta}}^{(s)}));$  $\triangleright$  Prediction

employed to compute the residual measurements (or detrended data)  $\hat{Y}^{(s)}$  by subtracting the estimated spatial trend from the measurements (12). With the detrended data, the function CWLS computes initial values for the estimation of the semivariogram parameter vector  $\hat{\theta}_{k}^{(s-1)}$  by solving a WLS minimization problem (17). Next, the REML module implements the REML

(18) to estimate the semivariogram parameter vector  $\hat{\theta}_k^{(s)}$ . The BIC function 356 calculates the BIC (19) and the diffBIC computes the difference of each 357 candidate with the lowest BIC<sup>\*</sup>. Then, the postBIC calculates the posterior 358 BIC (20) that assign probabilities of suitability for each candidate model with 359 the underlying latent process. When the highest probability of the posterior 360 BIC falls below a threshold  $\varphi$ , the **nested** function computes the covariance 361 matrix with a nested semivariogram (21). The iterative training procedure 362 is terminated when the semivariogram parameter estimation converges to 363 an  $\eta$ -neighborhood (24). Finally, we utilize the estimated covariance matrix 364  $\mathbf{\Sigma}(\widehat{\boldsymbol{ heta}}^{(s)})$  and the measurements to solve the universal kriging and obtain SNR 365 prediction  $\hat{Y}$  at the unvisited locations of interest  $\mathbf{x}_0$  and its corresponding 366 variance  $\operatorname{Var}[\hat{Y}]$  (28). 367

## 368 5.3. Computational Complexity

The time complexity of the training is  $\mathcal{O}(n^3)$  for computing the inverse 369 and determinant of the covariance matrix  $\Sigma$ . These computations are per-370 formed repeatedly in (18) to find the hyperparameters  $\boldsymbol{\theta}$  that maximize the 371 log-restricted likelihood. Next, we store the inverse covariance  $\Sigma^{-1}$  and n 372 measurements, which result in  $\mathcal{O}(n^2+mn)$  space complexity. For small robots 373 with limited RAM memory capacity, the space complexity may be more re-374 strictive than the time complexity. The prediction mean and variance (28)375 require  $\mathcal{O}(n)$  and  $\mathcal{O}(n^2)$  computations respectively. 376

#### 377 6. Simulations and Experiments

In this section, we provide simulations and experiments to demonstrate the efficacy of the proposed methodology.

#### 380 6.1. Simulation Environment

The simulation environment is developed with a well-established, statistical UWA channel model that incorporates 34 parameters and interprets multipath formation, motion-induced Doppler, surface scattering, and largescale variability of the channel geometry [38]. This channel model has been exhaustively compared to experimental data from multiple underwater missions, which varied in location, season, time duration, weather conditions, static nodes, and moving AUVs.

The SNR measurements consist of three components as described in (5). 388 The channel gain (6) is computed for signal frequency f = 25 kHz, band-389 width B = 5 kHz, surface height 100 m, and vehicle depths  $z_1 = 80$  m 390 and  $z_2 = 50$  m. The navigation depth corresponds to shallow water, where 391 the speed of sound can be considered constant [17]. We set the source level 392  $S_1 = 180 \text{ dB}$  which is a realistic value for UWA acoustic modems operat-393 ing in such signal frequencies. The large-scale parameters, i.e. path gain 394 and propagation delay, are computed using the BELLHOP model [60]. The 395 Doppler parameters were generated using first-order dynamics. Since the 396 vehicles maintain constant velocity, the drifting parameters were neglected. 397 Each vehicle depth remained constant during the simulation, yet the depth 398 of each vehicle is different. 399

In addition, we impose local ambient noise to the synthetic data (denom-400 inator of (5)). The local ambient noise is captured with: i) uniform noise; 401 ii) linear noise; iii) single non-zero Gaussian distribution; and iv) two non-402 zero Gaussian distributions. We evaluate the ambient noise over a grid of 403 points in the space  $\mathcal{S} \coloneqq \mathbb{X} \times \mathbb{Y}$ , where  $\mathbb{X} = [-2000, -1990, \dots, 3000]$  and 404  $\mathbb{Y} = [0, 10..., 5000]$  in meters. The ambient noise for the space of interest 405 outputs values  $NL(\mathbf{x}) \in [7.75, 50]$  in dB, resulting in both mild and extreme 406 environments. 407

The evaluation of the predictions is accomplished with two metrics. The 408 first metric is the mean square error (MSE), MSE =  $1/n_u \sum_{u=1}^{n_u} (\hat{Y}(\mathbf{x}_{0,u}; r_u) -$ 409  $Y(\mathbf{x}_{0,u}; r_u))^2$ , where  $n_u$  is the number of unknown responses at locations 410 of interest. Next, the negative log predictive density (NLPD) [61] follows, 411 NLPD =  $-1/n_u \sum_{u=1}^{n_u} \log p(y_u \mid \mathbf{x}_{0,u}; r_u)$ , where the distribution is provided 412 by  $p(y_u \mid \mathbf{x}_{0,u}; r_u) \sim \mathcal{N}(\hat{Y}(\mathbf{x}_{0,u}; r_u), \sigma_{\mathrm{UK}}^2(\mathbf{x}_{0,u}; r_u))$  (28). The NLPD loss char-413 acterizes not only the error of the mean value, but more importantly the 414 uncertainty bound. More specifically, both under- and over-confident predic-415 tions are penalized. 416

#### 417 6.2. Simulation Results

We compare five prediction techniques: i) the ordinary kriging (OK) with exponential semivariogram (15); ii) the OK with Matérn semivariogram (16); iii) the universal kriging (UK) with linear trend and exponential semivariogram (15); iv) the UK with linear trend and Matérn semivariogram (16); and v) the proposed model-based learning method with hybrid basis function and semivariogram model selected by the posterior BIC or formed as a nested

Cases			Exponential Semivariogram Parameters		
Training Set	Validation Set	Bias	$\begin{matrix} \text{OK} \\ \sigma^2,  \alpha,  \tau^2 \end{matrix}$	$\begin{array}{c} \text{UK} \\ \sigma^2,  \alpha,  \tau^2 \end{array}$	
150 Long-distant	519	$-10 \\ 0 \\ +10$	$\begin{array}{c} 93.27,17822,0\\ 92.83,17738,0\\ 92.74,17722,0\end{array}$	60.59, 12381, 0 61.00, 12378, 0 60.79, 12382, 0	
500 Short-distant	169	$-10 \\ 0 \\ +10$	80.91, 16888, 0 80.86, 16877, 0 80.31, 16763, 0	70.66, 14792, 0 71.51, 14969, 0 71.15, 14895, 0	

Table 1: Training with Exponential Semivariogram

OK–Ordinary kriging; UK–Universal kriging.

Cases			Matérn Semivariogram Parameters			
Training Set	Validation Set	Bias	$\begin{matrix} \text{OK} \\ \sigma^2,  \alpha,  \tau^2 \end{matrix}$	$\bigcup_{\sigma^2,\ \alpha,\ \tau^2}^{\rm UK}$	Model-based UK $\sigma^2,  \alpha,  \tau^2$	
150 Long-distant	519	$-10 \\ 0 \\ +10$	368.73, 2994, 0.20	364.42, 2994, 0.20	14.42, 711, 0.19	
500 Short-distant	169	$-10 \\ 0 \\ +10$	$14.13,  548,  0.23 \\111.17,  1611,  0.26$	13.78, 548, 0.23	82.04, 1495, 0.26	

Table 2: Training with Matérn Semivariogram

OK-Ordinary kriging; UK-Universal kriging.

structure. The OK formulation is discussed in [32]. In the first four prediction techniques, we select the exponential and the Matérn semivariogram functions, as they are widely used in the literature. Each agent collects measurements of communication performance (SNR) and vehicle range r from visited locations. Since the global paths are known, the agents are aware of their range at the unvisited locations. The GEOR package [62] is used to implement the geostatistical methodologies.

## 431 6.2.1. Training

For the evaluation of the robustness in training, we perform 30 simulations with added bias on the measurements. We consider one noise profile scenario of two non-zero Gaussian distributions. The trajectories of the mobile robots as well as the ambient noise distribution are illustrated in the top right image of Figure 3. The black solid and dotted line represent the lawnmower paths of agent 1 and 2 respectively. We consider two cases: i) the long-distant prediction; and ii) the short-distant prediction. In the long-distant prediction

case, each agent collects 75 measurements while in the short-distant case 250 439 measurements of SNR and range. We seek to predict the communication per-440 formance in the long-distant case of 260 and 259 and in the short-distant-case 441 of 85 and 84 unvisited locations for agent 1 and 2 respectively. The effect 442 of the bias to the semivariogram estimation, i.e. robustness, is investigated 443 by adding a systematic error to the measurements. The added biases are: 444 i) +10; ii) -10; and iii) no bias. We observe in Tables 1 and 2 that the added 445 bias does not affect the training of the proposed technique, resulting in the 446 same semivariogram function and semivariogram parameters. In both OK 447 and UK methods with exponential semivariogram, the estimated parameters 448 are clearly affected by the added bias. In the OK prediction method with 449 Matérn semivariogram, the added bias affects only the long-distant case of 450 +10 added bias, yet the difference is significant. The UK prediction method 451 with Matérn semivariogram is not affected by the added bias. Note that 452 the the posterior BIC selected the Matérn semivariogram as the true model. 453 Evidently, when a statistical model selection methodology is not employed, 454 yet the true semivariogram model is spontaneously selected, then the param-455 eter estimation appears less variation with added bias. However, we can-456 not always rely on heuristic assumptions, ignoring statistical model selection 457 methods. In addition, in many cases a single semivariogram function may 458 not be adequate to fully describe the underlying latent process. After us-459 ing the posterior BIC to select the true semivariogram function, the REML 460 successfully removes the bias from the parameter estimation, regardless of 461 the systematic error direction, i.e. sign of the bias. Therefore, the proposed 462 methodology constitutes a robust and bias-free alternative of the maximum 463 likelihood estimator. 464

## 465 6.2.2. Prediction

For the evaluation of the prediction we perform 180 simulations, compris-466 ing 9 training datasets at 4 ambient noise profile scenarios and 5 prediction 467 techniques. The size of the training dataset varied proportionally from 10%468 up to 90% of the data. The remainder data act as the validation dataset of 469 the learning process. The distant horizon of the extrapolation is associated 470 to the proportion of the training data, e.g., 10% of training data correspond 471 to the longest distant prediction and 90% to the shortest distant prediction. 472 The spatial environmental conditions and the global path of the vehicles are 473 shown in the top row of Figure 3. The MSE and NLPD are presented in the 474 middle and bottom row of Figure 3 respectively. 475



Figure 3: The color map on the top row depicts the ambient noise distribution that deteriorates the UWA communication performance. The solid black and dotted black lines correspond to the lawnmower paths of agent 1 and agent 2 respectively. In all cases we use 9 proportions of the training data to make predictions. (a) MSE and NLPD values for the uniform noise distribution case. (b) MSE and NLPD values for the linear noise distribution case. (c) MSE and NLPD values for the one source of non-zero Gaussian noise distribution case. (d) MSE and NLPD values for the two non-zero Gaussian noise distribution case.

In the first noise distribution scenario, i.e. uniform noise, randomness 476 arises mostly from the statistical characterization of the UWA channel model 477 (see Figure 3-(a)). That is mild ambient noise conditions, which often appear 478 in deep ocean. In shallow water environments, uniform ambient noise occur 479 when vehicles navigate in areas with no nearby shipping and mild weather 480 conditions. Clearly, the proposed method outperforms the rest techniques 481 both in terms of prediction accuracy and uncertainty quantification. Es-482 pecially, for long-distant prediction the difference is significant, making our 483 model-based approach three orders of magnitude more accurate in terms of 484 MSE and the uncertainty bounds almost one order of magnitude more re-485 alistic according to NLPD. As more data are incorporated in the training 486 dataset, the rest methods improve their accuracy and uncertainty quantifi-487

cation metrics. However, only in the shortest distant prediction case, i.e. 488 90% training dataset, the rest methods are comparable with our technique. 489 The results advocate that for mild ambient noise conditions the proposed 490 model-based learning technique vastly outperforms the compared methods 491 and can be even used for long-distant extrapolation. Next, we impose lin-492 ear ambient noise distribution to the UWA channel model, as presented in 493 Figure 3-(b). Linear ambient noise corresponds to a spatially large source of 494 noise that almost equally and progressively deteriorates the communication 495 performance of the vehicles. Similarly to the uniform noise case, the results 496 show better predictions from all other methods, where after the 40% train-497 ing dataset the predicted values become accurate with almost zero error. 498 Yet, the uncertainty of the proposed technique is overconfident, reporting 499 similar NLPD values with the rest methods. The results reveal that for 500 the linear ambient noise distribution scenario, our methodology outperforms 501 the rest techniques and produces accurate predictions for 40% and larger 502 training datasets. However, the uncertainty quantification is overconfident 503 in all cases. 504

A single spatially small and intense source of noise is presented in Fig-505 ure 3-(c). Such noise sources often appear in Nature and they consider to be 506 the main reason of conservativeness in long-distant extrapolated predictions. 507 Apparently, the spatially small source of noise obscure the UK methods and 508 slightly favors the OK techniques. However, the proposed model-based UK 509 methodology outperforms the rest techniques by one order of magnitude 510 on the mean predictions and quantifies the uncertainty better according to 511 NLPD. Thus, our learning method advocates to higher level of robustness for 512 unexpected spatially small and intense source of noise. We extend the previ-513 ous case using two spatially small and intense sources of noise with different 514 magnitude, as illustrated in Figure 3-(d). The proposed method outperforms 515 the rest techniques for the majority of the training dataset cases in terms of 516 MSE. The biggest competitor is the most parsimonious form of prediction 517 the OK, yet in only three out of nine training datasets the OK produces lower 518 communication performance error values. The uncertainty quantification is 519 reasonable in all techniques except for long distant predictions of UK with 520 linear trend and exponential semivariogram. Although in unexpected noisy 521 environments the model-based techniques are expected to be inefficient, our 522 method outperforms the other techniques in the vast majority of the cases 523 in terms of prediction and quantifies reasonably well the uncertainty. 524

In addition to the evaluation of prediction metrics, the effectiveness of

Table 3: Posterior BIC-based Selection of Semivariogram Function

	Semivariogram-posterior BIC [%]				
% of Data	Uniform Noise	Linear Noise	1 Gaussian Source Noise	2 Gaussian Source Noise	
10	S-33; E-33; M-34	S	М	М	
20	S-31; E-31; M-38	S-51; E-49	Μ	Μ	
30	S-32; E-32; M-36	$\mathbf{S}$	Μ	Μ	
40	S-33; E-32; M-35	S	Μ	Μ	
50	S-32; E-32; M-36	S	Μ	Μ	
60	S-26; E-32; M-42	$\mathbf{S}$	Μ	Μ	
70	S-9; E-87; M-4	$\mathbf{S}$	Μ	Μ	
80	S-33; E-33; M-34	Μ	Μ	Μ	
90	S-16; E-67; M-17	Μ	Μ	Μ	

S-Spherical; E-Exponential; M-Matérn.



Figure 4: Comparison of nested semivariogram with the three candidate semivariogram functions for the uniformly distributed noise scenario.

nested semivariogram is illustrated. In Table 3, we list the semivariograms 526 as selected by the posterior BIC for all 9 training datasets and 4 ambient 527 noise profile scenarios. Interestingly, in the linear noise distribution sce-528 nario the posterior BIC changes the semivariogram function from spherical 529 to Matérn at the 80% and 90% training data. This means that even if we 530 select one semivariogram model for a specific case, there are no guarantees 531 that the same semivariogram will describe the latent process with updated 532 training datasets. Moreover, we observe in Table 3 that all semivariograms 533 are nested for the uniform noise distribution, thus we focus our attention on 534 this scenario. In Figure 4, we compare the MSE and NLPD of the nested 535 semivariogram with the three candidate semivariograms. Notably, the mean 536 predictions are identical for single and nested semivariograms. Yet, the un-537 certainty quantification for nested functions is consistently better with all 538 training datasets. This advocates that the proposed technique with nested 539



Figure 5: The top row depicts the trajectories of the SV and the 690-AUV. The light gray line corresponds to the SV trajectory during the day, the blue line depicts the trajectory of the SV for the current mission, and the maroon colored line represents the path of the 690-AUV. The bottom row shows the vehicle range and output SNR of the corresponding mission.



Figure 6: The VIRGINIA TECH 690-AUV used in the field trials.

semivariogram quantifies more realistically the uncertainty, without compro mising accuracy.

## 542 6.3. Field Trials

The experimental data were collected from field trials conducted at Clay-543 tor lake near Dublin, VA in December 2019. A manned surface vehicle (SV) 544 and the VIRGINIA TECH 690-AUV [63] were used in the field trials. The 545 SV is equipped with an omnidirectional acoustic transducer and a Woods 546 Hole Oceanographic Institute (WHOI) Micromodem-2 [64]. The AUV (pic-547 tured in Figure 6), can operate at a depth of 500 m for up to 24 hours. It is 548 equipped with a suite of navigational sensors, sidescan sonar, and the WHOI 549 Micromodem-2. 550

The SV transmitted acoustic packets to the AUV every 10 seconds. The acoustic transmission each lasted 3.5 seconds and had a carrier frequency of

Mission	Duration [s]	Transmitted Signals	SNR Occurrence	SNR Success	Outliers
1	475.44	48	54	48	6
2	498.76	50	52	48	4
3	523.64	53	65	48	17
4	538.36	54	67	58	9

 Table 4:
 Output SNR Values for Four-Waypoint Experiments

f = 25 kHz and bandwidth of B = 4 kHz. The transponder mounted to 553 the SV was submerged at a depth  $z_1 = 1$  m while the AUV traveled at a 554 depth of  $z_2 = 3.35$  m. That is clearly shallow water navigation which makes 555 the acoustic communication even more challenging. The maximum depth 556 of Claytor lake is 35 m. We conducted four missions whose trajectories are 557 illustrated in the top row of Figure 5. The SV trajectory throughout the day 558 is shown in light gray, the SV trajectory during each mission is highlighted in 559 blue, and the AUV trajectory is demonstrated in maroon. In all missions, the 560 AUV traversed identical waypoint paths (waypoints shown in black circles). 561 The speed of the AUV was constant at 1.6 m/s. The SV was manned-driven 562 with different path for each mission. The missions were conducted in mild 563 weather conditions with no nearby shipping. Note that the field tests were 564 conducted in December, when no extramural activities take place at the 565 lake. Thus, the only obvious source of ambient noise was from the SV. The 566 noise arising from the SV was time-varying, as it was driven at different low 567 speeds. In missions 1 and 2 the SV used the propulsion system to navigate 568 and traversed longer paths. That is to intentionally create ambient noise. In 569 missions 3 and 4 the propulsion system of the SV was not used, i.e. the SV 570 was floating, which resulted in lower ambient noise. We used GPS for the 571 SV position, while the AUV position was estimated by the AUV's unscented 572 Kalman filter (UKF). 573

The SNR measurements were collected by the WHOI Micromodem-2 at 574 the output of the equalizer. This SNR metric is used in the literature for the 575 evaluation of communication performance [65, 66, 67]. The disadvantages of 576 the output SNR are: i) averages the SNR for a communication event; and ii) 577 provides positive rounded numbers—compromising the SNR measurement 578 accuracy. In the bottom row of Figure 5, we present the SNR in blue solid 579 line, the corresponding vehicle range in green solid line, and the outliers in 580 red squares. Clearly, there exists a coupling of the range and the SNR. More 581 specifically, as the range increases the SNR decreases. Note that the default 582



Figure 7: The eMSE and eNLPD metrics for all five prediction methods in four missions.

value of the WHOI modem to report output SNR outliers is -9.99 dB [64], yet we plot them at -1.00 dB for scaling purposes. In Table 4, we list the statistics of communication events.

#### 586 6.4. Experimental Results

Similarly to Section 6.2, we compare five prediction techniques. However, 587 the MSE and NLPD cannot be used, as the true value of the communication 588 performance at the location of the measurements is unknown during field 589 trials; measurements are corrupted by multiple sources of error. Hence, to 590 proceed with our analysis we refer to the metrics as empirical MSE (eMSE) 591 and empirical NLPD (eNLPD) accordingly. The eMSE and eNLPD values 592 for various proportions of data are presented in Figure 7. The 40% pro-593 portion of data includes at least 20 measurements for the longest distant 594 prediction, while the 80% proportion of data corresponds to the shortest dis-595 tant prediction case. In some cases the predicted values report almost zero 596 uncertainty, making the eNLPD values very high. At these cases, we con-597 sider that the corresponding method has failed, as uncertainty quantification 598 is a key element in communication performance prediction. Since we are 590 interested in evaluating low scaled eNLPD values, we set its upper bound 600 to be 100. In Figure 8 the prediction mean and standard deviation of three 601 techniques: i) OK with exponential semivariogram; ii) UK with linear trend 602 and exponential semivariogram; and iii) our method are presented. We select 603



Figure 8: The prediction mean and standard deviation for three methods and three proportions of data in four missions.

the exponential semivariogram for both OK and UK, because they provide better predictions in terms of eMSE and eNLPD (see Figure 7). The top row of Figure 8 corresponds to 40% proportion of data, the middle row to 59%, and the bottom row to 80%.

In mission 1 the high ambient noise affects the performance of the UK 608 techniques. Our method performs similarly to the OK methods for long-609 distant predictions and profoundly better for short-distant predictions. The 610 eNLPD values are acceptable in all predictions techniques and cases, except 611 one case of UK prediction with Matérn semivariogram. In mission 2 the high 612 ambient noise affects the performance of all prediction techniques. Only 613 the proposed method quantifies the uncertainty, while all other methods fail 614 as shown in Figure 8-(b). Both parsimonious methods of OK report lower 615 error values, yet with zero variance. Although the error metrics of the pro-616 posed technique are not satisfactory for this mission, the uncertainty of the 617 proposed method is quantified, as indicated by the eNLPD in Figure 7-(b) 618 and the prediction plots in Figure 8-(b). Paradoxically, all methods produce 619 higher error values as more measurements are collected. In mission 3 and 4, 620

	Semivariogram-posterior BIC [%]				
% of Data	Mission 1	Mission 2	Mission 3	Mission 4	
40	- C	C 99. E 99. M 94		C 24. E 20. M 24	
40 51	S	S-33; E-33; M-34 S-33·F-33·M-34	E	S-34; E-32; M-34	
59	S	S-32; E-33; M-35	E	S-33; E-33; M-34	
67	S-29; E-37; M-34	S-33; E-32; M-35	$\mathbf{E}$	S-33; E-33 ;M-34	
80	S-38; E-29; M-33	S-36; E-31; M-33	$\mathbf{E}$	S-33; E-33; M-34	

Table 5: Posterior BIC-based Selection of Semivariogram Function

S-Spherical; E-Exponential; M-Matérn.

the low ambient noise results in better predictions for our method than the 621 rest techniques. Particularly, in mission 3 the error and uncertainty metrics 622 of our technique are significantly better from the rest methods as illustrated 623 in Figure 8-(c). In mission 4, all techniques provide acceptable results in 624 terms of eMSE, yet the proposed method is the most accurate and the only 625 one that quantifies the uncertainty. All other methods fail. In addition, even 626 though the other techniques provide low eMSE values, their uncertainty is 627 overconfident as indicated by the eNLPD values in Figure 7-(d). Realistic 628 uncertainty bounds are reported only from the proposed methodology, while 629 all other techniques provide predictions with zero variance as presented in 630 Figure 8-(d). Note that for higher vehicle depth the ambient noise deterio-631 rates, favoring the proposed methodology, as noise scenarios are similar to 632 missions 3 and 4. 633

In Table 5, we list the selected semivariograms based on the posterior BIC. It is evident that there is no dominant semivariogram function and that the nested semivariogram was employed in many cases. This emphasizes the importance of nested models and the necessity of statistical model selection techniques for field trials in complex environments.

#### <sup>639</sup> 7. Conclusion and Future Work

This paper proposes a model-based, data-driven learning technique for prediction of underwater acoustic communication performance in AUVs beyond the observation area. In both traditional ordinary kriging (OK) and universal kriging (UK) methods the estimated parameters are affected by the artificially added bias, leading to different parameter values. We show that the proposed model-based learning yields accurate predictions, outperforming up to three orders of magnitude other kriging methods in simulations.

More specifically, for all ambient noise profile scenarios both OK and UK pre-647 diction methods that used in the comparison produce high error values and 648 quantify the uncertainty poorly. Moreover, the nested semivariogram func-649 tion improves drastically the uncertainty quantification. In addition, experi-650 mental results reveal significantly better predictions with our method for low 651 and high ambient noise environments. The proposed technique reports real-652 istic uncertainty bounds in all missions, which other mission methods often 653 fail to generate. In unpredictable and high ambient noise environments, our 654 method outperforms in prediction accuracy the techniques assessed herein. 655

A disadvantage of the proposed technique arises from the computational 656 requirements of the iterative training. More specifically, the training step 657 entails the computation of three candidate covariance functions with corre-658 sponding parameters at every iteration. We found in practice that the recur-659 sive method usually terminates after two iterations, for which the execution 660 of the training step requires six times more computations than the traditional 661 OK and UK methods with fixed semivariogram functions. Another drawback 662 that is subject to all techniques stems from the communication. In particu-663 lar, all agents must communicate their measurements to every other agent in 664 multi-robot missions. To this end, our focus in ongoing work is on decentral-665 ized approximate methods to implement kriging in multi-robot systems with 666 reduced computational complexity and limited inter-vehicle communication 667 [68]. This will allow even large networks with big data to use the proposed 668 technique in real-time. 669

## 670 Acknowledgements

This work was supported by the Office of Naval Research via grants N00014-18-1-2627 and N00014-19-1-2194.

#### 673 References

- [1] I. F. Akyildiz, D. Pompili, T. Melodia, Underwater acoustic sensor
  networks: Research challenges, Ad Hoc Networks 3 (2005) 257–279.
  doi:10.1016/j.adhoc.2005.01.004.
- [2] W. Burgard, M. Moors, C. Stachniss, F. E. Schneider, Coordinated
  multi-robot exploration, IEEE Transactions on Robotics 21 (2005) 376–
  386. doi:10.1109/TRD.2004.839232.

- [3] J. Fink, A. Ribeiro, V. Kumar, Robust control of mobility and communications in autonomous robot teams, IEEE Access 1 (2013) 290–309. doi:10.1109/access.2013.2262013.
- [4] L. V. Nguyen, S. Kodagoda, R. Ranasinghe, G. Dissanayake,
   Information-driven adaptive sampling strategy for mobile robotic wire less sensor network, IEEE Transactions on Control Systems Technology
   24 (2015) 372–379. doi:10.1109/tcst.2015.2435657.
- [5] X. Yu, M. A. Hsieh, Synthesis of a time-varying communication net work by robot teams with information propagation guarantees, IEEE
   Robotics and Automation Letters 5 (2020) 1413–1420. doi:10.1109/
   lra.2020.2967704.
- [6] H. Yu, C.-C. Lim, R. Hunjet, P. Shi, Flocking and topology manipulation based on space partitioning, Robotics and Autonomous Systems
   124 (2020) 103328. doi:10.1016/j.robot.2019.103328.
- [7] Y. Sung, A. K. Budhiraja, R. K. Williams, P. Tokekar, Distributed assignment with limited communication for multi-robot multi-target tracking, Autonomous Robots 44 (2020) 57–73. doi:10.1007/s10514-019-09856-1.
- [8] M. Minelli, J. Panerati, M. Kaufmann, C. Ghedini, G. Beltrame,
  L. Sabattini, Self-optimization of resilient topologies for fallible multirobots, Robotics and Autonomous Systems 124 (2020) 103384. doi:10.
  1016/j.robot.2019.103384.
- [9] E. Olcay, F. Schuhmann, B. Lohmann, Collective navigation of a multirobot system in an unknown environment, Robotics and Autonomous
  Systems 132 (2020) 103604. doi:10.1016/j.robot.2020.103604.
- [10] F. Bullo, J. Cortes, S. Martinez, Distributed control of robotic networks: A mathematical approach to motion coordination algorithms,
  volume 27, Princeton University Press, 2009.
- [11] J. Cortés, Distributed Kriged Kalman filter for spatial estimation, IEEE
   Transactions on Automatic Control 54 (2009) 2816–2827. doi:10.1109/
   tac.2009.2034192.

- [12] R. K. Williams, G. S. Sukhatme, Constrained interaction and coordination in proximity-limited multiagent systems, IEEE Transactions on Robotics 29 (2013) 930–944. doi:10.1109/tro.2013.2257578.
- [13] L. Sabattini, N. Chopra, C. Secchi, Decentralized connectivity maintenance for cooperative control of mobile robotic systems, The International Journal of Robotics Research 32 (2013) 1411–1423. doi:10.1177/ 0278364913499085.
- [14] Y. Kantaros, M. M. Zavlanos, Distributed intermittent connectivity
  control of mobile robot networks, IEEE Transactions on Automatic
  Control 62 (2016) 3109–3121. doi:10.1109/tac.2016.2626400.
- [15] R. Khodayi-mehr, Y. Kantaros, M. M. Zavlanos, Distributed state estimation using intermittently connected robot networks, IEEE Transactions on Robotics 35 (2019) 709–724. doi:10.1109/tro.2019.2897865.
- [16] K. Khateri, M. Pourgholi, M. Montazeri, L. Sabattini, A connectivity preserving node permutation local method in limited range robotic networks, Robotics and Autonomous Systems (2020) 103540. doi:10.1016/j.robot.2020.103540.
- [17] M. Stojanovic, J. Preisig, Underwater acoustic communication channels:
   Propagation models and statistical characterization, IEEE Communications Magazine 47 (2009) 84–89. doi:10.1109/mcom.2009.4752682.
- [18] H. P. Tan, R. Diamant, W. K. Seah, M. Waldmeyer, A survey of techniques and challenges in underwater localization, Ocean Engineering 38 (2011) 1663–1676. doi:10.1016/j.oceaneng.2011.07.017.
- [19] C. E. Rasmussen, C. K. Williams, Gaussian Processes for Machine
  Learning, 2 ed., Cambridge, MA, USA: MIT Press, 2006.
- [20] R. B. Gramacy, Surrogates: Gaussian Process Modeling, Design and Op timization for the Applied Sciences, Chapman Hall/CRC, Boca Raton,
   Florida, 2020. http://bobby.gramacy.com/surrogates/.
- [21] D. Horner, G. Xie, Data-driven acoustic communication modeling for
  undersea collaborative navigation, in: IEEE OCEANS-San Diego, 2013,
  pp. 1–10. doi:10.23919/0CEANS.2013.6741261.

- [22] M. Umer, L. Kulik, E. Tanin, Spatial interpolation in wireless sensor networks: Localized algorithms for variogram modeling and kriging,
  Geoinformatica 14 (2010) 101. doi:10.1007/s10707-009-0078-3.
- [23] W. Wu, A. Song, P. Varnell, F. Zhang, Cooperatively mapping of the underwater acoustic channel by robot swarms, in: ACM International Conference on Underwater Networks & Systems, 2014, p. 20. doi:10. 1145/2671490.2674572.
- [24] Q. Tao, Y. Zhou, F. Tong, A. Song, F. Zhang, Evaluating acoustic communication performance of micro autonomous underwater vehicles in confined spaces, Frontiers of Information Technology and Electronic Engineering 19 (2018) 1013–1023. doi:10.1631/fitee.1700841.
- [25] J. Sun, S. Liu, F. Zhang, A. Song, J. Yu, A. Zhang, A kriged compressive sensing approach to reconstruct acoustic fields from measurements collected by underwater vehicles, IEEE Journal of Oceanic Engineering (2020). doi:10.1109/J0E.2020.2974270.
- [26] M. Malmirchegini, Y. Mostofi, On the spatial predictability of communication channels, IEEE Transactions on Wireless Communications 11 (2012) 964–978. doi:10.1109/twc.2012.012712.101835.
- R. Augusto, C. Panazio, On geostatistical methods for radio environ ment maps generation under location uncertainty, Journal of Communi cation and Information Systems 33 (2018). doi:10.14209/jcis.2018.8.
- [28] A. Q. Li, P. K. Penumarthi, J. Banfi, N. Basilico, J. M. O'Kane, I. Rekleitis, S. Nelakuditi, F. Amigoni, Multi-robot online sensing strategies for the construction of communication maps, Autonomous Robots 44 (2019) 299–319. doi:10.1007/s10514-019-09862-3.
- [29] J. Kim, P. Ladosz, H. Oh, Optimal communication relay positioning
   in mobile multi-node networks, Robotics and Autonomous Systems 129
   (2020) 103517. doi:10.1016/j.robot.2020.103517.
- [30] B. F. D. Hähnel, D. Fox, Gaussian processes for signal strength-based
   location estimation, in: Robotics: Science and Systems, 2006.

- [31] B. Ferris, D. Fox, N. D. Lawrence, WiFi-SLAM using Gaussian process
  latent variable models, in: International Joint Conferences on Artificial
  Intelligence, volume 7, 2007, pp. 2480–2485.
- [32] G. P. Kontoudis, D. J. Stilwell, A comparison of kriging and cokriging for estimation of underwater acoustic communication performance, in: ACM International Conference on Underwater Networks & Systems, 2019. doi:10.1145/3366486.3366515.
- [33] H. Wackernagel, Multivariate geostatistics: An introduction with applications, 3 ed., New York, NY, USA: Springer-Verlag, 2003.
- [34] G. Matheron, Principles of geostatistics, Economic Geology 58 (1963)
  1246–1266.
- <sup>783</sup> [35] N. Cressie, Statistics for spatial data, 2 ed., Wiley, 1993.
- [36] R. J. Adler, J. E. Taylor, Random fields and geometry, 1 ed., New York,
   NY, USA: Springer, 2007.
- [37] C. Berg, J. P. R. Christensen, P. Ressel, Harmonic analysis on semigroups: Theory of positive definite and related functions, volume 100, Springer, 1984.
- [38] P. Qarabaqi, M. Stojanovic, Statistical characterization and computationally efficient modeling of a class of underwater acoustic communication channels, IEEE Journal of Oceanic Engineering 38 (2013) 701–717. doi:10.1109/joe.2013.2278787.
- [39] W. A. Fuller, Introduction to statistical time series, 2 ed., John Wiley
   & Sons, 1996.
- [40] R. Webster, M. A. Oliver, Geostatistics for environmental scientists, 1
   ed., New York, NY, USA: Wiley, 2007.
- [41] M. Stojanovic, On the relationship between capacity and distance in an underwater acoustic communication channel, ACM SIGMOBILE Mobile Comput. and Communic. Review 11 (2007) 34–43. doi:10.1145/ 1347364.1347373.

- [42] P. Cardieri, T. S. Rappaport, Statistics of the sum of lognormal variables
   in wireless communications, in: IEEE Vehicular Technology Conference,
   2000, pp. 1823–1827. doi:10.1109/vetecs.2000.851587.
- [43] P. Qarabaqi, M. Stojanovic, Modeling the large scale transmission
  loss in underwater acoustic channels, in: IEEE Allerton Conference on Communication, Control, and Computing, 2011, pp. 445–452.
  doi:10.1109/allerton.2011.6120201.
- [44] P. J. Diggle, P. J. Ribeiro Jr, Model-based geostatistics, 1 ed., New York,
   NY, USA: Springer, 2007.
- [45] H. Takeda, S. Farsiu, P. Milanfar, Kernel regression for image processing
  and reconstruction, IEEE Transactions on Image Processing 16 (2007)
  349–366. doi:10.1109/tip.2006.888330.
- <sup>813</sup> [46] G. Matheron, Traité de géostatistique appliquée, volume 1, Editions
   <sup>814</sup> Technip, 1962.
- [47] N. Cressie, D. M. Hawkins, Robust estimation of the variogram: I,
  Journal of the International Association for Mathematical Geology 12
  (1980) 115–125.
- [48] B. Matérn, Spatial variation, 2 ed., Berlin, Germany: Springer-Verlag,
   1960; reprinted 1986.
- [49] N. Cressie, Fitting variogram models by weighted least squares, Journal
  of the International Association for Mathematical Geology 17 (1985)
  563-586. doi:10.1007/bf01032109.
- [50] P. K. Kitanidis, Parametric estimation of covariances of regionalized
  variables, Journal of the American Water Resources Association 23
  (1987) 557-567. doi:10.1111/j.1752-1688.1987.tb00832.x.
- [51] D. A. Harville, Bayesian inference for variance components using only
  error contrasts, Biometrika 61 (1974) 383-385. doi:10.1093/biomet/
  61.2.383.
- [52] D. L. Zimmerman, Computationally efficient restricted maximum like lihood estimation of generalized covariance functions, Mathematical
   Geology 21 (1989) 655–672. doi:10.1007/bf00893314.

- I. J. Myung, Tutorial on maximum likelihood estimation, Journal of Mathematical Psychology 47 (2003) 90–100. doi:10.1016/
  s0022-2496(02)00028-7.
- <sup>835</sup> [54] G. Schwarz, et al., Estimating the dimension of a model, The Annals of Statistics 6 (1978) 461–464. doi:10.1214/aos/1176344136.
- [55] A. A. Neath, J. E. Cavanaugh, The Bayesian information criterion: Background, derivation, and applications, Wiley Interdisciplinary Reviews: Computational Statistics 4 (2012) 199–203. doi:10.1002/wics.
  199.
- <sup>841</sup> [56] D. Duvenaud, J. R. Lloyd, R. Grosse, J. B. Tenenbaum, Z. Ghahramani,
  <sup>842</sup> Structure discovery in nonparametric regression through compositional
  <sup>843</sup> kernel search, in: International Conference on Machine Learning, 2013.
- <sup>844</sup> [57] P. K. Kitanidis, Generalized covariance functions in estimation, Mathematical Geology 25 (1993) 525–540. doi:10.1007/bf00890244.
- <sup>846</sup> [58] T. Hengl, G. B. Heuvelink, A. Stein, A generic framework for spatial
  <sup>847</sup> prediction of soil variables based on regression-kriging, Geoderma 120
  <sup>848</sup> (2004) 75–93. doi:10.1016/j.geoderma.2003.08.018.
- [59] A. Lichtenstern, Kriging methods in spatial statistics, Thesis, Technical
   University of Munich, 2013.
- <sup>851</sup> [60] M. B. Porter, The BELLHOP manual and user's guide: Preliminary <sup>852</sup> draft, 2011.
- <sup>853</sup> [61] J. Quinonero-Candela, C. E. Rasmussen, F. Sinz, O. Bousquet,
  <sup>854</sup> B. Schölkopf, Evaluating predictive uncertainty challenge, in: Machine
  <sup>855</sup> Learning Challenges Workshop, 2005, pp. 1–27.
- <sup>856</sup> [62] P. J. Ribeiro Jr, P. J. Diggle, The geoR package, https://cran.
   r-project.org/web/packages/geoR/index.html, 2007. Version 1.7 5.2.1.
- <sup>859</sup> [63] B. McCarter, S. Portner, W. L. Neu, D. J. Stilwell, D. Malley, J. Minis, Design elements of a small AUV for bathymetric surveys, in: <sup>861</sup> IEEE/OES Autonomous Underwater Vehicles, 2014, pp. 1–5. doi:10. <sup>862</sup> 1109/auv.2014.7054428.

- [64] E. Gallimore, J. Partan, I. Vaughn, S. Singh, J. Shusta, L. Freitag, The
   WHOI micromodem-2: A scalable system for acoustic communications
   and networking, in: MTS/IEEE OCEANS, 2010, pp. 1–7. doi:10.1109/
   oceans.2010.5664354.
- L. Freitag, S. Singh, Performance of micro-modem PSK signaling under
   variable conditions during the 2008 RACE and SPACE experiments, in: IEEE OCEANS, 2009, pp. 1–8. doi:10.23919/oceans.2009.5422310.
- <sup>870</sup> [66] B. Tomasi, J. Preisig, M. Zorzi, On the predictability of underwater
  <sup>871</sup> acoustic communications performance: The KAM11 data set as a case
  <sup>872</sup> study, in: Intern. Workshop on Underwater Networks, 2011, pp. 1–7.
  <sup>873</sup> doi:10.1145/2076569.2076570.
- <sup>874</sup> [67] L. Freitag, S. Singh, Performance of micro-modem PSK signaling with
  <sup>875</sup> a mobile transmitter during the 2010 MACE experiment, in: IEEE
  <sup>876</sup> OCEANS-Genova, 2015, pp. 1–7. doi:10.1109/oceans-genova.2015.
  <sup>877</sup> 7271734.
- <sup>878</sup> [68] G. P. Kontoudis, D. J. Stilwell, Decentralized nested Gaussian processes
  <sup>879</sup> for multi-robot systems, in: IEEE International Conference on Robotics
  <sup>880</sup> and Automation, 2021.