Bayesian Active Learning for Sample Efficient 5G Radio Map Reconstruction

Konstantinos D. Polyzos, Alireza Sadeghi, Wei Ye, Steven Sleder, Kodjo Houssou, Jeff Calder, Zhi-Li Zhang, and Georgios B. Giannakis

***Abstract*—The advent of diverse frequency bands in 5G networks has promoted measurement studies focused on 5G signal propagation, aiming to understand its pathloss, coverage, and channel quality characteristics. Nonetheless, conducting a thorough 5G measurement campaign is markedly laborious given the large number of 5G measurement samples that must be collected. To alleviate this burden, the present contribution leverages principled active learning (AL) methods to prudently select only a *few*, yet most *informative* locations to collect 5G measurements. The core idea is to rely on a Gaussian Process (GP) model to efficiently extrapolate 5G measurements throughout the coverage area. Specifically, an ensemble (E) of GP models is adopted that not only provides a rich learning function space, but also quantifies uncertainty, and can offer accurate predictions. Building on this EGP model, a suite of acquisition functions (AFs) are advocated to query new locations on-the-fly. To account for realistic 5G measurement campaigns, the proposed AFs are augmented with a novel distance-based AL rule that selects informative samples, while penalizing queries at long distances. Numerical tests on 5G data generated by the Sionna simulator and on real urban and suburban datasets, showcase the merits of the novel EGP-AL approaches.**

***Index Terms*—Active Learning, 5G measurement, Radio Map Reconstruction**

I. Introduction

The fifth-generation cellular network, commonly known as 5G, expands its spectrum resources across both the low/mid-frequency bands (<6GHz) and the high-frequency bands (>24GHz) [1]. This brings a remarkable improvement in service quality and outperforms 4G, which mainly operates within frequencies at or below 2GHz. In particular, major 5G operators like ATT and Verizon in the US have been actively deploying C-band (3.7GHz) and mmWave (26GHz) technologies, achieving impressive gigabits-per-second-level throughput and millisecond-level latency [2]–[4].

Nevertheless, applying these frequency bands introduces new characteristics to mobile networks. For instance, mmWave has a much shorter wavelength (about 1∼12mm) [5], making its propagation highly directional, easily blocked and reflected by obstacles. This can lead to complex signal strength behaviors in the real world beyond just distance-based dependency [6], [7].

A plethora of wireless network-related tasks, such as network/radio resource/base station planning, handover optimization, quality of Service (QoS) management, etc., critically rely on a better understanding of radio propagation characteristics in the real world and require quantifying the signal coverage quality of specific base stations. To gain such insight, a large number of in-field measurements need to be conducted.

Obtaining these measurements is a laborious process, since the collection of valid measurements entails a significant amount of effort and the collected data may not even cover all of the locations within the coverage area. The lack of immediate feedback on data quality during in-field measurements can further lead to many redundant and/or useless measuring efforts. To cope with these challenges, existing works [7]–[16] have attempted to reconstruct radio maps only using *limited* measurements (see Sec. II). Albeit interesting, these approaches still fail to fully address the aforementioned practical issues.

The goal of this work is to improve the efficiency of obtaining measurements for reconstructing 5G radio maps by leveraging Bayesian-based active learning (AL) methods. Building on a statistical model to capture the mapping of any location within the coverage area to its corresponding received signal strength (RSS) value, AL leverages a collection of acquisition criteria (i.e., *acquisition functions* (AFs)) to select a *few* highly *informative* locations to obtain RSS measurements. We utilize the Gaussian Process (GP) for the underlying statistical model, which is a well-motivated Bayesian model that is able to effectively extrapolate and interpolate RSS measurements in regions with limited or even no observed data. Along with accurate RSS predictions at unobserved locations, the GP model provides valuable information about the confidence and reliability of predictions by offering a well-quantifiable notion of uncertainty around each predicted RSS value [17]. As GPs are non-parametric models by nature, they do not require explicit assumptions about the functional form or distribution of the data, making them applicable in complex settings.

Nonetheless, their performance hinges on a pre-selected kernel function to evaluate the pairwise similarity of different locations, which subsequently affects the covariance of the corresponding RSS values. Apriori selection of the kernel function is a nontrivial task, that may require domain-knowledge as well. The use of a single pre-selected kernel

Konstantinos D. Polyzos, Alireza Sadeghi and Georgios B. Giannakis are with the Department of Electrical and Computer Engineering at University of Minnesota, Twin Cities (emails: {polyz003, sadeg012, georgios}@umn.edu).

Wei Ye, Steven Sleder and Zhi-Li Zhang are with the Department of Computer Science and Engineering at University of Minnesota, Twin Cities (emails: {ye000094, slede001}@umn.edu and zhzhang@cs.umn.edu).

Kodjo Houssou and Jeff Calder are with the School of Mathematics at University of Minnesota, Twin Cities (emails: {houss001, jwcalder}@umn.edu).

This work was supported by NSF grants 2212318, 2220292, 2126052, 2102312, 2103256, 2312547.

in GP modeling confines the expressiveness of the sought learning function for estimating RSS values. In addition, the complexity of these methods significantly grows as the number of available measurements increases. To cope with expressiveness challenges of single kernel-based GP learning, and to allow for *scalable* and *online* model updates, we will rely on an ensemble of parametric approximants of GP models, similarly as in [18] (see also Sec.II).

Contributions. Our contributions in this work can be summarized in the following aspects:

- To the best of our knowledge, this is the first work that formulates and implements Bayesian-AL for the 5G radio map reconstruction problem using only a small number of measurements.
- We advocate intuitive AL methods that leverage an ensemble of GPs (EGPs) to enhance the model's expressiveness when dealing with multi-modal and complex data distributions. We then leverage parametric function approximants of the GP models in the ensemble using random features (RFs) to offer online and scalable model updates. These are particularly appealing in AL settings where new measurements are processed online.
- To allow for path-efficient acquisition of RSS measurements in new locations, we combine a number of well-motivated AFs with a novel *path efficient* rule that penalizes querying far-distant locations at each iteration of AL.
- We conduct thorough experiments on a total of six simulated datasets with three scenarios and two radio frequencies, and on two real urban and suburban datasets, to corroborate the efficiency and effectiveness of the proposed EGP-based AL in reconstructing 5G radio maps. Our numerical tests show that our EGP-based AL approaches achieve up to 54.8% improvement in terms of normalized mean square error (NMSE) over the EGP model with random sampling. In addition, compared to the EGP-AL methods that do not impose any constraints on the acquisition of new locations to measure, the EGP-based AL approaches that consider the novel path efficient rule, achieve up to 57.1% reduction in terms of required cumulative traveling distance to obtain measurements.

The remainder of the paper is organized as follows. In Sec. II related works are discussed. Active learning formulation for 5G radio map reconstruction is introduced in Sec. III, while the advocated EGP-based AL approach with path-efficient acquisition functions is presented in IV. Upon discussing details about data generation in Sec. V, numerical tests on simulated and real 5G datasets are provided in Sec. VI. Finally, Sec. VII concludes the paper.

Notations. Scalars are denoted by lowercase, column vectors by bold lowercase, matrices by bold uppercase, and sets by calligraphic uppercase fonts. Superscripts $^\top$ and $^{-1}$ denote transpose and inverse respectively; while $\mathbf{1}_N$ stands for the $N \times 1$ all-ones vector; and $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \mathbf{K})$ for the pdf of a Gaussian random vector $\mathbf{x}$ with mean $\boldsymbol{\mu}$, and covariance matrix $\mathbf{K}$. The identity matrix will be represented by $\boldsymbol{I}$, and the all-zeros matrix by $\mathbf{0}$.

II. Related Works

Radio Map Reconstruction. Numerous approaches have been proposed to reconstruct 5G radio maps, comprising both non-parametric methods [8]–[10], [19], [20] and classical parametric ones [11], [12], [21]. More recently, researchers have explored deep learning-based approaches [7], [13]–[16], [22] to enhance the accuracy of radio propagation modeling and better capture its characteristics. However, these methods primarily concentrate on interpolating or extrapolating unmeasured locations based on existing measured points, lacking the ability to guide future sampling —a critical practical challenge. In contrast, the novel active learning method adopted in this paper effectively addresses this challenge by providing a principled method to search for the most informative location to query the next measurement.

Active Learning. AL methods can be broadly categorized into parametric AL [23]–[25] and non-parametric ones [26]–[28]. Although most existing parametric approaches in the literature excel at learning complex representations from data, they are *deterministic* and hence fall short in inherently providing uncertainty quantification. In addition, they are tailored solely for classification tasks, and do not account for regression tasks, such as the RSS prediction problem in this paper; see e.g [29]. On the other hand, Bayesian non-parametric approaches such as GPs, inherently offer uncertainty quantification, and demonstrate superior sample efficiency [17], [27]. Existing GP-based AL approaches typically rely on a single pre-selected kernel, which limits their expressiveness when dealing with data exhibiting multi-modal behavior. To cope with this challenge, existing deep-learning based approaches aim to learn the kernel with neural networks, but they typically require a sufficient number of labels which are not available in several AL settings; see e.g [30]. The proposed AL method in this paper uses a more expressive statistical model that relies on an ensemble of GP models with scalable online model updates as in [31]. Finally, the proposed acquisition strategies in the existing literature do not impose constraints on the data to label, which discourages their application in the 5G map reconstruction problem where the travelling cost of collecting 5G measurements should be minimized. A recent work [32] has considered AL methods that choose query points nearby previously queried ones, though the method is not applicable to 5G signal reconstruction. To cope with this challenge, this work incorporates a novel path efficient acquisition rule into the AL process, where locations at long distances are penalized.

III. Problem Statement

We now formulate the 5G active measurement problem (see Fig.1) and then briefly discuss the limitations of AL when it is directly applied in practice.

Active Learning for 5G Measurements. For a 5G coverage area $\mathcal{X} \subset \mathbb{R}^2$, we are interested in estimating the real-valued received signal strength (RSS) $y(\mathbf{x}) \in \mathbb{R}$ at any desired location (or feature vector) $\mathbf{x} := [\texttt{longitude, latitude}] \in \mathbb{R}^2$; Let $\mathbf{x}_i$ denote the i-th probed location with associated scalar RSS value $y_i \in \mathbb{R}$. We postulate that there exists a

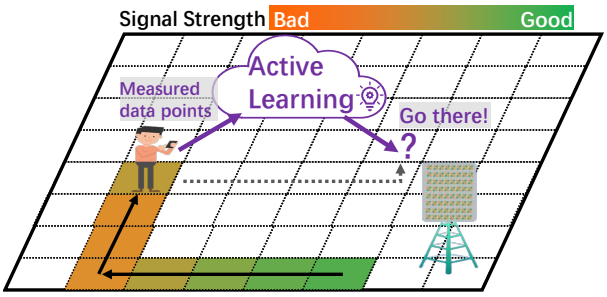


Fig. 1: Accelerating 5G Measurements via Active Learning.

ground truth function $f(\mathbf{x}) : \mathbb{R}^2 \to \mathbb{R}$ which maps a desired location $\mathbf{x}$ to its associated RSS through

$$y = f(\mathbf{x}) + \nu \tag{1}$$

where $\nu \sim \mathcal{N}(\nu; 0, \sigma^2_{\text{noise}})$ represents the observation noise.

To *efficiently* learn an estimate $\widehat{f}(\mathbf{x})$ at any desired $\mathbf{x}$ with only few, but representative, measured RSS data, one can rely on the *active learning* (AL) paradigm; see Fig. 1 for an illustration. AL begins with a relatively small dataset of labeled (measured) data $\mathcal{D}_n := \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ (where n is small). To judiciously update $\mathcal{D}_n$ and build an informative training set, AL leverages an *acquisition function* (AF) $\alpha(\cdot)$ to search for the *most informative* query locations on-the-fly while updating the underlying model $\widehat{f}(\cdot)$. The next query location $\mathbf{x}_{n+1}$ is obtained by solving

$$\mathbf{x}_{n+1} := \arg\min_{\mathbf{x}\in\mathcal{X}} \alpha(\mathbf{x}; \mathcal{D}_n). \tag{2}$$

To obtain the RSS value $y(\mathbf{x}_{n+1}) := y_{n+1}$ corresponding to the queried location $\mathbf{x}_{n+1}$, a field-test[1] can be carried out. Upon obtaining y_{n+1}, AL augments the training dataset $\mathcal{D}_{n+1} := \mathcal{D}_n \cup \{(\mathbf{x}_{n+1}, y_{n+1})\}$ and updates the estimate $\widehat{f}(\cdot)$ on-the-fly. Having an updated $\widehat{f}(\cdot)$, the active learner can predict the RSS at any desired (unmeasured) location $\mathbf{x} \in \mathcal{X}$.

Therefore, to learn $f(\cdot)$ in a *sample-efficient manner* with AL, we aim to properly select a learning model for $f(\cdot)$ and design AFs relying on the chosen model, to effectively guide the acquisition of new query locations to measure in an online manner.

Challenges Conventional AL methods rely on the 'independent data acquisition' assumption, meaning that the active learner can query measurements from any location within the coverage area independently from the previous measurement locations. This limits the applicability of conventional AL in wireless communications, since there exist many practical considerations that one needs to consider when querying a new location, such as traveling cost and measurement runtime complexity. For example, simply selecting a location that may be informative but requires a much longer travel distance is not cost-optimized, as the user may have to travel a prohibitively long distance to take a sufficient number of measurements and may repeatedly traverse the same regions within the measurement area (see Sec. VI). In the ensuing sections, we provide principled methods to facilitate a *path-* and *sample-*efficient data acquisition process.

[1]One can obtain RSS measurements at different locations within $\mathcal{X}$, using some off-the-shelf tools, such as Android-based API [33] running on smartphones or more professional scanners [34].

IV. Active Learning with Ensembles of GPs

To effectively estimate the sought function f with *well-quantifiable uncertainty* in a *sample-efficient fashion*, this work specifically focuses on Gaussian processes (GPs). GPs belong to the family of nonparametric Bayesian models, and will be leveraged in this work to estimate unknown RSS values and also guide the acquisition of new measurement locations, as outlined next.

A. Gaussian Processes

We leverage GPs to probabilistically model $f(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, and subsequently obtain a model over RSS values $\{y(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}\}$. To this aim, let us define a *single* GP-based *prior* distribution over $f(\mathbf{x}) \sim \mathcal{GP}(\mu(\mathbf{x}), \kappa(\mathbf{x}, \mathbf{x}'))$, with mean $\mu(\mathbf{x})$ and a pre-specified kernel function $\kappa(\mathbf{x}, \mathbf{x}')$. Usually, for notational simplicity it is assumed that $\mu(\mathbf{x}) = 0, \forall \mathbf{x} \in \mathcal{X}$. This prior on $f(\mathbf{x})$ implies that for any finite number $n \in \mathbb{N}^+$ of samples, the random vector of function values $\boldsymbol{f}_n := [f(\mathbf{x}_1) \cdots f(\mathbf{x}_n)]$ at locations $\mathbf{X}_n := [\mathbf{x}_1, \ldots \mathbf{x}_n]^\top$ has a *joint Gaussian* distribution; that is $p(\boldsymbol{f}_n|\mathbf{X}_n) = \mathcal{N}(\boldsymbol{f}_n; \mathbf{0}_t, \mathbf{K}_n)$, where $\mathbf{K}_n$ represents the $n \times n$ covariance matrix with (i, j)-th entry $[\mathbf{K}_n]_{i,j} := \text{cov}\left(f(\mathbf{x}_i), f(\mathbf{x}_j)\right) = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ [17].

The random vector $\boldsymbol{f}_n$ is related to the (possibly) noisy observations $\mathbf{y}_n := [y(\mathbf{x}_1), \ldots, y(\mathbf{x}_n)]^\top$ through the batch likelihood $p(\mathbf{y}_n|\boldsymbol{f}_n; \mathbf{X}_n)$ that is assumed to be factored as $p(\mathbf{y}_n|\boldsymbol{f}_n; \mathbf{X}_n) = \prod_{n'=1}^n p(y_{n'}|f(\mathbf{x}_{n'}))$. Capitalizing on the GP prior and the batch likelihood, one can express the joint probability density function (PDF) of the observation vector $\mathbf{y}_n$ and the function value $f(\mathbf{x})$ at any unmeasured location $\mathbf{x}$ as

$$\begin{bmatrix} \mathbf{y}_n \\ f(\mathbf{x}) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}_n + \sigma^2_{\text{noise}}\boldsymbol{I}_n & \mathbf{k}_n^\top(\mathbf{x}) \\ \mathbf{k}_n(\mathbf{x}) & \kappa(\mathbf{x}, \mathbf{x}) \end{bmatrix}\right), \tag{3}$$

where $\mathbf{k}_n(\mathbf{x}) := [\kappa(\mathbf{x}_1, \mathbf{x}), \ldots, \kappa(\mathbf{x}_n, \mathbf{x})]$. Leveraging this joint Gaussian distribution in (3), and *marginalizing* this distribution over $f(\mathbf{x})$, it can be shown that the predictive PDF of $y(\mathbf{x})$ is [17]

$$p(y(\mathbf{x})|\mathcal{D}_n, \mathbf{x}) = \mathcal{N}(y(\mathbf{x}); \bar{y}(\mathbf{x})_{n+1|n}, \sigma^2_{n+1|n}(\mathbf{x})), \tag{4}$$

where

$$\bar{y}(\mathbf{x})_{n+1|\mathbf{n}} = \mathbf{k}_n^\top(\mathbf{x}) \left(\mathbf{K}_n + \sigma^2_{\text{noise}}\boldsymbol{I}_n\right)^{-1}\mathbf{y}_n \tag{5a}$$

$$\sigma^2_{n+1|\mathbf{n}}(\mathbf{x}) = \kappa(\mathbf{x}, \mathbf{x}) - \mathbf{k}_n^\top(\mathbf{x}) \left(\mathbf{K}_n + \sigma^2_{\text{noise}}\boldsymbol{I}_n\right)^{-1}\mathbf{k}_n(\mathbf{x}) + \sigma^2_{\text{noise}}. \tag{5b}$$

with the notation $n+1|\mathbf{n}$ signifying that all n data samples are employed to obtain the predictive PDF at the next $(n+1)$-th location. The mean in (5a) provides a point estimate of the RSS value of location $\mathbf{x}$ and the variance in (5b) quantifies the associated uncertainty.

Although intuitive, the predictive PDF in (5) is subject to certain limitations. Primarily, the predictive mean and variance in equation (5) depend on a *single* pre-selected kernel function which restricts the expressive capacity of the learned function.

In addition, the mean and variance in (5a) and (5b) entail $\mathcal{O}(n^2)$ storage requirements and $\mathcal{O}(n^3)$ computational complexity due to the matrix inversion operation involved, which may become computationally intractable when n is large. In the subsequent sections, we will outline how to circumvent these challenges.

B. Ensemble of Gaussian Processes

Instead of relying on a single kernel to estimate $f(\mathbf{x})$, we target a more expressive function space by utilizing a set of kernels $\mathcal{K} := \{\kappa^1, \ldots, \kappa^M\}$ to form a *mixture* of GP models. The set $\mathcal{K}$ constitutes a diverse collection of kernels with different hyper-parameters. Each $\kappa^m \in \mathcal{K}$ induces a unique GP prior m over the function $f(\mathbf{x}) \sim \mathcal{GP}(0, \kappa^m(\mathbf{x}, \mathbf{x}'))$. Constructing a weighted combination of these GP models yields a mixture of GP priors over $f(\mathbf{x})$, as $f(\mathbf{x}) \sim \sum_{m=1}^M w_n^m \mathcal{GP}(0, \kappa(\mathbf{x}, \mathbf{x}'))$, subject to $\sum_{m=1}^M w_n^m = 1$. The per-model weight w_n^m signifies the contribution of m-th kernel in the GP prior over $f(\mathbf{x})$. Using this mixture of GP priors, the posterior distribution of $f(\mathbf{x})$ can be obtained by

$$p\left(f(\mathbf{x})\,|\mathcal{D}_n\right) = \sum_{m=1}^{M} w_n^m p\left(f(\mathbf{x})|\mathcal{D}_n, \kappa = \kappa^m\right), \tag{6}$$

with $w_n^m \propto \Pr(\kappa = \kappa^m)\Pr(\mathcal{D}_n|\kappa = \kappa^m)$, where $\Pr(\mathcal{D}_n|\kappa = \kappa^m)$ represents the marginal likelihood of data $\mathcal{D}_n$ for the m-th GP. For any desired location $\mathbf{x} \in \mathcal{X}$, each GP induces a Gaussian predictive PDF $p(y(\mathbf{x})|\mathcal{D}_n, m, \mathbf{x}) = \mathcal{N}(y(\mathbf{x}); \bar{y}^m_{n+1|n}(\mathbf{x}), \sigma^m_{n+1|n}(\mathbf{x}))$ to estimate the corresponding RSS value. By appropriately updating the mean and variance for each GP (c.f., (5a) and (5b)) and adjusting the weights $\{w_n^m\}_{m=1}^M$ for all GP models, one can obtain [18] the predictive PDF of $y(\mathbf{x})$

$$p(y(\mathbf{x})|\mathcal{D}_n, \mathbf{x}) = \sum_{m=1}^{M} w_n^m \mathcal{N}(y(\mathbf{x}); \bar{y}^m_{n+1|n}(\mathbf{x}), (\sigma^m_{n+1|n}(\mathbf{x}))^2). \tag{7}$$

Although more expressive compared to (4), the predictive PDF in (7) still requires large computational complexity; i.e $\mathcal{O}(Mn^3)$.

C. Random Feature-based EGPs

To bypass this cubic complexity, one can employ parametric approximants to the non-parametric models using the so-called random features (RF) [35]. The main idea of the RF-based approximation is to rely on a shift-invariant kernel $\bar{\kappa}(\mathbf{x}, \mathbf{x}') = \bar{\kappa}(\mathbf{x} - \mathbf{x}')$, normalized as $\bar{\kappa} = \kappa/\sigma_\theta^2$, and express it via the inverse Fourier transformation of its power spectral density $\pi_\kappa(\boldsymbol{\zeta})$ [35]; that is

$$\begin{aligned}\bar{\kappa}(\mathbf{x} - \mathbf{x}') &:= \mathbb{E}_{\pi_\kappa}\left[e^{j\boldsymbol{\zeta}^\top(\mathbf{x}-\mathbf{x}')}d\boldsymbol{\zeta}\right] \\ &= \int \pi_\kappa(\boldsymbol{\zeta}) e^{j\boldsymbol{\zeta}^\top(\mathbf{x}-\mathbf{x}')}d\boldsymbol{\zeta}\end{aligned} \tag{8}$$

where $\int \pi_\kappa(\boldsymbol{\zeta})d\boldsymbol{\zeta} = 1$, so that it can be thought of as a PDF. Since the values of κ are always real, it holds that $\bar{\kappa}(\mathbf{x}-\mathbf{x}') := \mathbb{E}_{\pi_\kappa}\left[\cos(\boldsymbol{\zeta}^\top(\mathbf{x} - \mathbf{x}'))\right]$. Leveraging this new representation, one can approximate the kernel by drawing a sufficiently large number of i.i.d samples $\{\boldsymbol{\zeta}_i\}_{i=1}^D$ from $\pi_\kappa(\boldsymbol{\zeta})$ yielding

$$\hat{\bar{\kappa}}(\mathbf{x} - \mathbf{x}') := \frac{1}{D}\sum_{i=1}^{D} \cos(\boldsymbol{\zeta}_i^\top(\mathbf{x} - \mathbf{x}')) = \boldsymbol{\phi}_\zeta(\mathbf{x})^\top \boldsymbol{\phi}_\zeta(\mathbf{x}'), \tag{9}$$

where the random feature vector $\boldsymbol{\phi}_{\boldsymbol{\zeta}}(\cdot) \in \mathbb{R}^{2D\times 1}$ is defined as

$$\boldsymbol{\phi}_{\boldsymbol{\zeta}}(\mathbf{x}) := \frac{1}{\sqrt{D}}\left[\sin(\boldsymbol{\zeta}_1^\top\mathbf{x}), \cos(\boldsymbol{\zeta}_1^\top\mathbf{x}), \ldots, \cos(\boldsymbol{\zeta}_D^\top\mathbf{x})\right]. \tag{10}$$

Relying on the RF vector $\boldsymbol{\phi}_{\boldsymbol{\zeta}}(\cdot)$, a *linear* and *parametric* approximant of the sought function can be obtained as [35]

$$\check{f}(\mathbf{x}) = \boldsymbol{\theta}^\top \boldsymbol{\phi}_{\boldsymbol{\zeta}}(\mathbf{x}), \quad \boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\theta}; \mathbf{0}_{2D}, \sigma_\theta^2 \boldsymbol{I}_{2D}). \tag{11}$$

This parametric model over $\check{f}(\mathbf{x})$ enables the propagation of the posterior $p(\boldsymbol{\theta}|\mathcal{D}_n) = \mathcal{N}(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_n, \boldsymbol{\Sigma}_n)$ using a recursive Bayesian iteration, which can be updated as new data arrive on-the-fly, as will be shown next.

D. Ensemble of Parametric GPs

To allow for reduced complexity and online model updates that are particularly appealing in the AL setting, we consider an ensemble of parametric GPs. Each GP model m, relies on a shift-invariant and normalized kernel $\bar{\kappa}^m = \kappa^m/\sigma^2_{\theta^m}$ to construct its RF vector $\boldsymbol{\phi}^m_{\boldsymbol{\zeta}}(\cdot)$ by drawing i.i.d random vectors $\{\boldsymbol{\zeta}^m_j\}_{j=1}^D$ from its corresponding power spectral density $\pi_{\bar{\kappa}^m}(\boldsymbol{\zeta})$. Similar to (11), each learner m forms a generative parametric model as

$$p(\boldsymbol{\theta}^m) = \mathcal{N}(\boldsymbol{\theta}^m; \mathbf{0}_{2D}, \sigma^2_{\theta^m}\boldsymbol{I}_{2D}) \tag{12}$$

$$p(f(\mathbf{x})|\kappa = \bar{\kappa}^m, \boldsymbol{\theta}^m) = \delta(f(\mathbf{x}) - \boldsymbol{\phi}^m_{\boldsymbol{\zeta}}(\mathbf{x})^\top\boldsymbol{\theta}^m) \tag{13}$$

$$p(y(\mathbf{x})|\boldsymbol{\theta}^m, \mathbf{x}) = \mathcal{N}(y(\mathbf{x}); \boldsymbol{\phi}^m_{\boldsymbol{\zeta}}(\mathbf{x})^\top\boldsymbol{\theta}^m, \sigma^2_{\text{noise}}). \tag{14}$$

This generative model allows model m to form the posterior PDF $p(\boldsymbol{\theta}^m|\mathcal{D}_n) = \mathcal{N}(\boldsymbol{\theta}^m; \hat{\boldsymbol{\theta}}^m_n, \boldsymbol{\Sigma}^m_n)$ using available data $\mathcal{D}_n$, and the advocated parametric EGP model combines them with the weights $\{w_n^m\}_{m=1}^M$ to form the ensemble predictive PDF over the target $y(\mathbf{x})$, for all $\mathbf{x} \in \mathcal{X}$. This predictive PDF not only offers an estimate of the RSS value at location $\mathbf{x}$, but will further guide the design of a set of acquisition functions to find the next locations to obtain measurements, as will be discussed in the next subsection.

We now show how the EGP model parameters will be updated as a newly acquired measurement at a new location indexed by $n+1$ is obtained. Each model m leverages the learned posterior $p(\boldsymbol{\theta}^m|\mathcal{D}_n)$ to find the predictive PDF of $y(\mathbf{x})$ at any desired target location $\mathbf{x}$ as

$$\begin{aligned}p(y(\mathbf{x})|\kappa =& \bar{\kappa}^m, \mathcal{D}_n, \mathbf{x}) \\ &= \int p(y(\mathbf{x})|\boldsymbol{\theta}^m, \mathbf{x}) p(\boldsymbol{\theta}^m|\mathcal{D}_n) d\boldsymbol{\theta}^m \\ &= \mathcal{N}(y(\mathbf{x}); \hat{\bar{y}}^m_{n+1|n}(\mathbf{x}), (\sigma^m_{n+1|n}(\mathbf{x})))^2,\end{aligned} \tag{15}$$

with mean and variance given by

$$\hat{\bar{y}}^m_{n+1|n}(\mathbf{x}) = \boldsymbol{\phi}^m_{\boldsymbol{\zeta}}(\mathbf{x})^\top \hat{\boldsymbol{\theta}}^m_n \tag{16a}$$

$$(\sigma^m_{n+1|n}(\mathbf{x}))^2 = \boldsymbol{\phi}^m_{\boldsymbol{\zeta}}(\mathbf{x})^\top \boldsymbol{\Sigma}^m_n \boldsymbol{\phi}^m_{\boldsymbol{\zeta}}(\mathbf{x}) + \sigma^2_{\text{noise}}\,. \tag{16b}$$

To estimate the RSS value $y(\mathbf{x})$, the EGP model forms a Gaussian mixture (GM) model by combining all M predictive PDFs to form the ensemble predictive PDF

$$p(y(\mathbf{x})|\mathcal{D}_n,\mathbf{x}) = \sum_{m=1}^{M} w_n^m \mathcal{N}(y(\mathbf{x}); \hat{y}_{n+1|n}^m(\mathbf{x}), (\sigma_{n+1|n}^m)^2(\mathbf{x})). \tag{17}$$

Having this Gaussian mixture model over $y(\mathbf{x})$, one can obtain an estimate of the RSS value along with the corresponding uncertainty, by considering the minimum mean-square error (MMSE) estimator of $y(\mathbf{x})$ along with the variance of the estimator; that is

$$\hat{y}_{n+1|n}(\mathbf{x}) = \sum_{m=1}^{M} w_n^m \hat{y}_{n+1|n}^m(\mathbf{x}) \tag{18a}$$

$$\sigma_{n+1|n}^2(\mathbf{x}) = \sum_{m=1}^{M} w_n^m \Big[(\sigma_{n+1|n}^m(\mathbf{x}))^2 + (\hat{y}_{n+1|n}(\mathbf{x}) - \hat{y}_{n+1|n}^m(\mathbf{x}))^2 \Big]. \tag{18b}$$

where "$n+1|n$" denotes that only the measurement from location n is needed to predict y at the next location $n+1$.

Leveraging the learned PDF in (17) along with the mean and variance in (18), one can find the next query location $\mathbf{x}_{n+1}$ to obtain the RSS measurement y_{n+1} by minimizing the AFs as will be elaborated in the next subsection IV-E. Upon obtaining $\mathbf{x}_{n+1}$ and y_{n+1}, one can update the EGP model parameters. Specifically, each model m uses Bayes' rule along with (16) to update its weight w_n^m as

$$w_{n+1}^m = \frac{w_n^m \mathcal{N}(y_{n+1}; \hat{y}_{n+1|n}^m, (\sigma_{n+1|n}^m)^2)}{\sum_{m'=1}^{M} w_n^{m'} \mathcal{N}(y_{n+1}; \hat{y}_{n+1|n}^{m'}, (\sigma_{n+1|n}^{m'})^2)}, \tag{19}$$

and its posterior PDF as

$$p(\boldsymbol{\theta}^m|\mathcal{D}_{n+1}) = \frac{p(\boldsymbol{\theta}^m|\mathcal{D}_n)p(y_{n+1}|\boldsymbol{\theta}^m,\mathbf{x}_{n+1})}{p(y_{n+1}|\boldsymbol{\theta}^m,\mathcal{D}_n,\mathbf{x}_{n+1})} = \mathcal{N}(\boldsymbol{\theta}^m; \hat{\boldsymbol{\theta}}_{n+1}^m, \boldsymbol{\Sigma}_{n+1}^m), \tag{20}$$

where

$$\hat{\boldsymbol{\theta}}_{n+1}^m = \hat{\boldsymbol{\theta}}_n^m + \frac{\boldsymbol{\Sigma}_n^m \boldsymbol{\phi}_{\boldsymbol{\zeta}}^m(\mathbf{x}_{n+1})(y_{n+1} - \hat{\bar{y}}_{n+1|n}^m)}{(\sigma_{n+1|n}^m)^2} \tag{21a}$$

$$\boldsymbol{\Sigma}_{n+1}^m = \boldsymbol{\Sigma}_n^m - \frac{\boldsymbol{\Sigma}_n^m \boldsymbol{\phi}_{\boldsymbol{\zeta}}^m(\mathbf{x}_{n+1})\boldsymbol{\phi}_{\boldsymbol{\zeta}}^m(\mathbf{x}_{n+1})^\top \boldsymbol{\Sigma}_n^m}{(\sigma_{n+1|n}^m)^2}. \tag{21b}$$

The required complexity of updating the model parameters in (21) is $\mathcal{O}(M((2D)^2 + 2Dn)))$, which is smaller than the $\mathcal{O}(Mn^3)$ complexity of the original EGP-based model.

E. Acquisition functions for EGP-based active learning.

This section introduces a number of EGP-based AFs, which depend on the learned predictive PDF $p(y(\mathbf{x})|\mathcal{D}_n)$ in (17), to guide the acquisition process; that is to find the next most informative location within the coverage area, i.e., $\mathbf{x}_{n+1} \in \mathcal{X}$, and obtain its associated RSS measurement y_{n+1}.

1) Weighted variance: To exploit the uncertainty of each GP model, we leverage the variance of all GP models' posterior PDFs and combine them to form the weighted variance AF as

$$\alpha^{\mathrm{wVar}}(\mathbf{x};\mathcal{D}_n) := \sum_{m=1}^{M} w_n^m (\sigma_{n+1|n}^m(\mathbf{x}))^2. \tag{22}$$

The intuition is that locations with high uncertainty can be informative for the training dataset. Although $\alpha^{\mathrm{wVar}}(\cdot)$ well captures the uncertainty of each GP model, it ignores the information offered by the predicted means (c.f. (16a)), motivating the next AF.

2) Query-by-committee: Considering M GP models as members of a committee, the Query-by-committee (QBC) AF chooses the next location to be measured where the committee members exhibit the most disagreement; that is

$$\alpha^{\mathrm{QBC}}(\mathbf{x};\mathcal{D}_n) := \sum_{m=1}^{M} w_n^m \left(\hat{\bar{y}}_{n+1|n}^m(\mathbf{x}) - \bar{y}_n(\mathbf{x}) \right), \tag{23}$$

where $\hat{\bar{y}}_{n+1|n}^m(\mathbf{x})$ represents the estimated mean of RSS values at location $\mathbf{x}$ by learner m (c.f., (16a)), and $\bar{y}_n(\mathbf{x}) := \sum_{m=1}^{M} w_n^m \hat{\bar{y}}_{n+1|n}^m(\mathbf{x})$ denotes the weighted mean of the committee. Note that only the predictive means are used in $\alpha^{\mathrm{QBC}}(\cdot)$ and the predictive variances are not considered.

3) Variance of GP mixture: To simultaneously account for the *means* and *variances* associated with each GP model's predictive PDF, the variance of GP mixture AF combines the the per-learner uncertainty in (22) and committee disagreement in ((23) (c.f., (18b)) as

$$\alpha^{\mathrm{GPM-var}}(\mathbf{x},\mathcal{D}_n) := \sigma_{n+1|n}^2(\mathbf{x}). \tag{24}$$

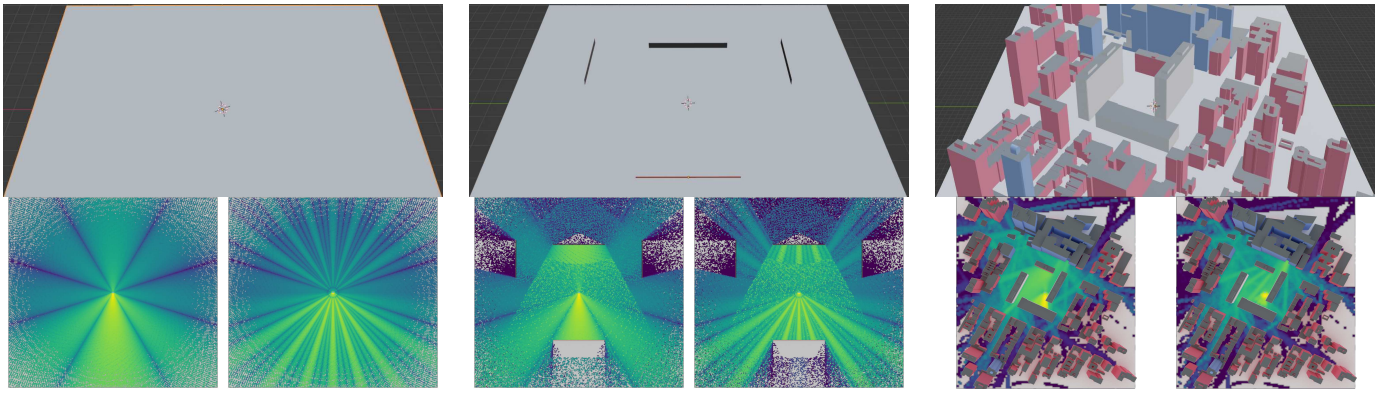
4) Weighted entropy: Relying on the entropy as an alternative measure of the associated uncertainty, one can form an AF using a weighted combination of the entropies associated with the predictive PDFs of all learners as

$$\alpha^{\mathrm{wEnt}}(\mathbf{x};\mathcal{D}_n) := \sum_{m=1}^{M} w_n^m \log\left((\sigma_{n+1|n}^m(\mathbf{x}))^2\right). \tag{25}$$

5) Entropy of GP mixtures: Instead of employing the entropies of individual GP models independently as in (25), we can leverage the entropy of the learned GP-mixture. While an analytical closed-form expression for this entropy does not exist, we leverage a closed-form lower bound; see e.g [31]. To this aim, let us first define the entropy of the learned function $\hat{f}(\mathbf{x})$ at any location $\mathbf{x}$

$$\mathcal{H}(\hat{f}(\mathbf{x})|\mathcal{D}_n) := -\int_{-\infty}^{\infty} \sum_{m=1}^{M} w_n^m \mathcal{N}(\hat{f}(\mathbf{x})|\hat{\bar{y}}_{n+1|n}^m(\mathbf{x}), (\sigma_{n+1|n}^m(\mathbf{x}))^2) \times \log(p(\hat{f}(\mathbf{x})|\mathcal{D}_n))\, d\hat{f}(\mathbf{x}) \tag{26}$$

where the integration here is taken over all the *random function values* that $\hat{f}(\mathbf{x})$ can take at any desired, but fixed $\mathbf{x} \in \mathcal{X}$.



(a) Scenario-1: Open Space. (b) Scenario-2: Reflection & Blockage. (c) Scenario-3: Urban Canyon.

Fig. 2: Radio coverage maps of diverse scenarios rendered with Blender (top) and Sionna ray-tracing simulator at 3.7GHz (bottom left) and 26GHz (bottom right).

Using Jensen's inequality yields a lower bound for the entropy $\mathcal{H}(\hat{f}(\mathbf{x})|\mathcal{D}_n)$ at any location $\mathbf{x} \in \mathcal{X}$ as follows

$$\mathcal{H}(\hat{f}(\mathbf{x})|\mathcal{D}_n) \geq - \sum_{m=1}^{M} w_n^m \log \left(\int_{-\infty}^{\infty} \mathcal{N}(\hat{f}(\mathbf{x})|\hat{\bar{y}}_{n+1|n}^m(\mathbf{x}), (\sigma_{n+1|n}^m(\mathbf{x}))^2) \times p(\hat{f}(\mathbf{x})|\mathcal{D}_n)\, d\hat{f}(\mathbf{x}) \right). \quad (27)$$

The term inside the logarithm in (27) admits a closed-form expression and hence the lower-bound can be expressed analytically as

$$\alpha^{\mathrm{GPM-Ent}}(\mathbf{x}; \mathcal{D}_n) := \sum_{m=1}^{M} \log \left(\sum_{m'=1}^{M} w_n^{m'} \psi_n^{m,m'} \right), \quad (28)$$

where $\psi_n^{m,m'}$ captures the intra-GP interactions and is given by

$$\psi_n^{m,m'} := \int_{-\infty}^{\infty} \mathcal{N}(\hat{f}(\mathbf{x})|\hat{\bar{y}}_{n+1|n}^m(\mathbf{x}), (\sigma_{n+1|n}^m(\mathbf{x}))^2) \times \mathcal{N}(\hat{f}(\mathbf{x})|\hat{\bar{y}}_{n+1|n}^{m'}(\mathbf{x}), (\hat{\sigma}_{n+1|n}^{m'}(\mathbf{x}))^2) d\hat{f}(\mathbf{x}) = \mathcal{N}(\hat{\bar{y}}_{n+1|n}^m; \hat{\bar{y}}_{n+1|n}^{m'}, (\sigma_{n+1|n}^m)^2 + (\sigma_{n+1|n}^{m'})^2). \quad (29)$$

While these AFs cover a diverse range of important ideas within AL, they impose no constraints on the spatial locations of the next query point, which can lead to prohibitively long travelling distances required to collect the necessary measurements dictated by the AFs. To address this limitation, we will couple these EGP-based AFs with a minimum distance-based criterion, described in the next subsection.

F. Path efficient EGP-based AFs

To allow for efficient and cost-effective data acquisition strategies, thereby enhancing the practical feasibility and economic viability of acquiring measurements and reconstructing 5G radio maps, we consider AFs with the following form

$$\alpha(\mathbf{x}, \mathcal{D}_n) := \begin{cases} \alpha^{\mathrm{EGP}}(\mathbf{x}, \mathcal{D}_n) & \text{w.p.} \quad \varepsilon \\ \alpha^{\mathrm{EGP}}(\mathbf{x}, \mathcal{D}_n)\, c(\mathbf{x}, \mathbf{x}_{n-1}) & \text{w.p.} \quad 1-\varepsilon \end{cases} \quad (30)$$

TABLE I: Simulation configurations.

	Parameters	Descriptions		
Data	**Scenario**	OpenSpace	Reflection	UrbanCanyon
	3.7GHz Samples	780.5K	841.7K	47.9K
	26GHz Samples	779.8K	744.5K	41.3K
Antenna Configs	**Rx Pattern**	Isotropic		
	Tx Pattern	TR 38.901		
	Tx Power	24.0 dB		
	Ray Interactions	9		
	Clipping Range	(0.0, 160.0] dB		
	Cell Size	5m x 5m		

(a) Urban Canyon. (b) Suburban.

Fig. 3: Radio coverage maps for real RSS measurements. Green color represents strong RSS, while red poor RSS.

where $c(\mathbf{x}, \mathbf{x}_{n-1}) := \frac{1}{\lambda + \mathrm{dist}(\mathbf{x}, \mathbf{x}_{n-1})}$ is a distance-penalizing function with λ being a hyperparameter to be tuned, $\alpha^{\mathrm{EGP}}(\cdot)$ is one of the above EGP-based AFs, and $\varepsilon > 0$ is a small hyperparameter.[2] With a small probability $0 < \varepsilon < 1$ our novel AF in (30) allows exploring $\mathcal{X}$ while adhering to the EGP-based AF $\alpha^{\mathrm{EGP}}(\mathbf{x}, \mathcal{D}_n)$. With a much larger probability $1 - \varepsilon$ our AF is strongly encouraged to select points *nearby* the most recently obtained measurement $\mathbf{x}_n$ by penalizing far distant points via the path-cost function $c(\mathbf{x}, \mathbf{x}_{n-1})$. This allows the AF to explicitly acknowledge that obtaining measurements from diverse locations in 5G scenarios can be costly and resource-intensive, while still allowing a user-controlled degree of exploration through the parameter ε. This is crucial in several AL applications including the 5G radiomap reconstruction that this paper focuses on.

[2]The abbreviation w.p. stands for *with probability*.

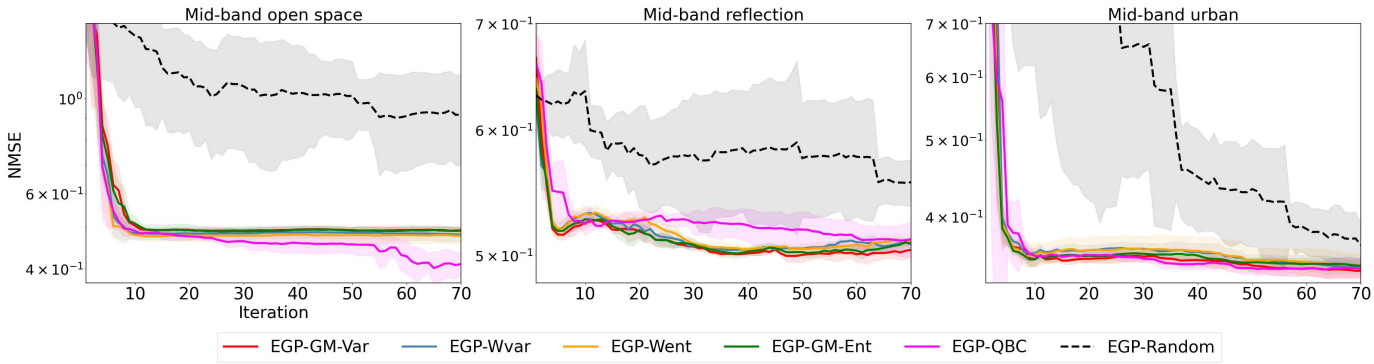


Fig. 4: NMSE vs iteration for 3.7GHz mid-band data under (a) open space; (b) reflect scenarios; and (c) urban canyon.

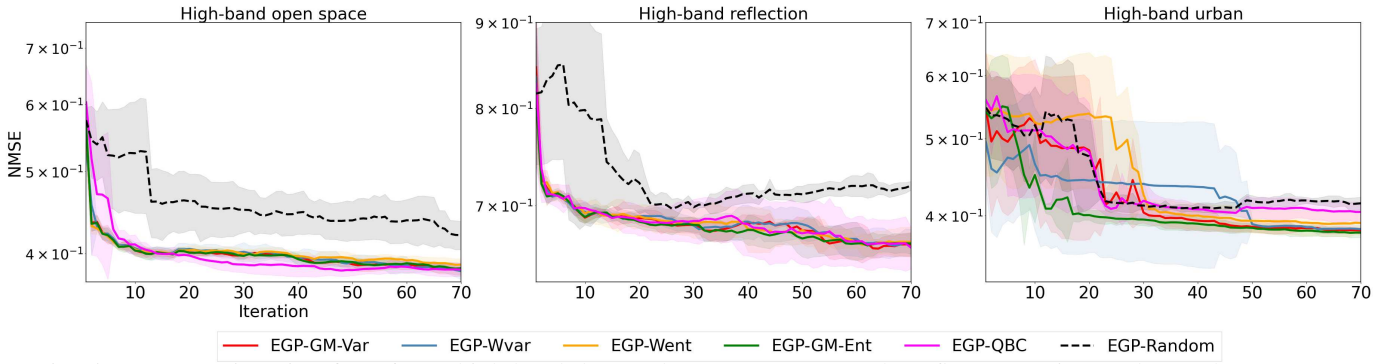


Fig. 5: NMSE vs iteration for 26GHz high-band data under (a) open space; (b) reflect scenarios; and (c) urban canyon.

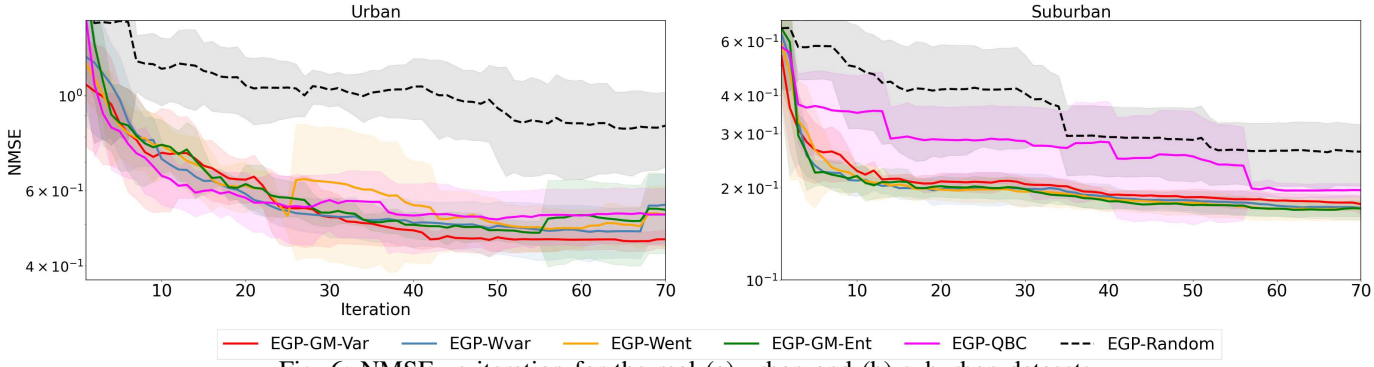


Fig. 6: NMSE vs iteration for the real (a) urban and (b) suburban datasets.

V. Dataset Generation

We now introduce the dataset utilized for evaluation, which comprises simulated and real measurements. The simulated dataset replicates three distinct scenes: open space, reflection & blockage, and urban canyon. Meanwhile, the real measurements consist of two scenes: suburban and urban canyons.

Simulated Dataset. Following [36], we employ Blender [37] and Sionna [38] to produce radio maps and use them as ground truth for the evaluation. Blender is an open-source software that can create different 3D environments. We utilize its Open Street Map (OSM) and Mitsuba plugins to load the real-world map and render the digital scenes. Subsequently, these scenes are imported into Sionna, an open-source link-level ray-tracing simulator that is compatible with the 3GPP channel models, to generate radio coverage maps that we sample from.

The upper panel of Fig. 2 depicts the loaded digital scenes, illustrating three distinct scenarios: (1) a large open space with the line-of-sight propagation of radio signals; (2) a more complex wall scenario featuring multiple walls to demonstrate reflection, diffraction, and signal blockage; and (3) a highly complex downtown urban canyon, where the signals are reflected and obstructed by the surrounding buildings. The lower panel of Fig. 2 shows the simulated radio coverage maps. We also investigate the propagation characteristics of different frequency radio waves across 3.7GHz (mid-band) and 26GHz (high-band) in light of commercial 5G networks. For each scenario-frequency pair, we place the transmitter (Tx) at the center location with a directional antenna oriented at the negative y-axis. The map is divided into sets of cells, each containing a receiver (Rx) with an omnidirectional antenna, generating a total of N unique samples of the form $[x, y, RSS]$. To determine the RSS value, rays are fired from a Tx at the origin and are allowed to interact with the environment, i.e., reflect, diffract, etc., up to a bounded number of times. The RSS is aggregated from per intersecting rays [38], clipped to a range. For specific configuration parameters of our dataset, we refer the reader to Table I.

Real Measurement Dataset. In addition to simulated datasets,

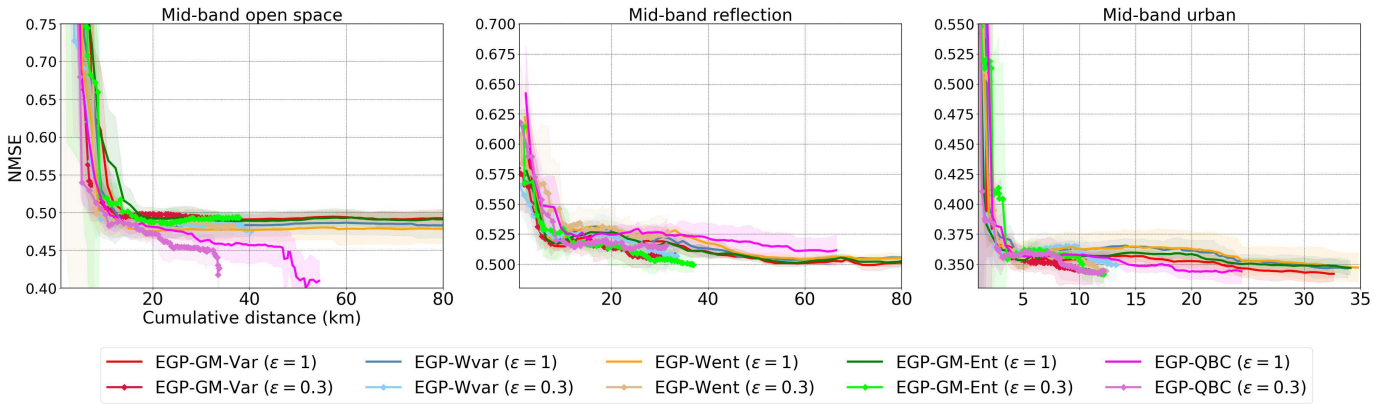


Fig. 7: NMSE vs cumulative distance for 3.7GHz mid-band data under (a) open space; (b) urban canyon; and (c) reflect datasets.

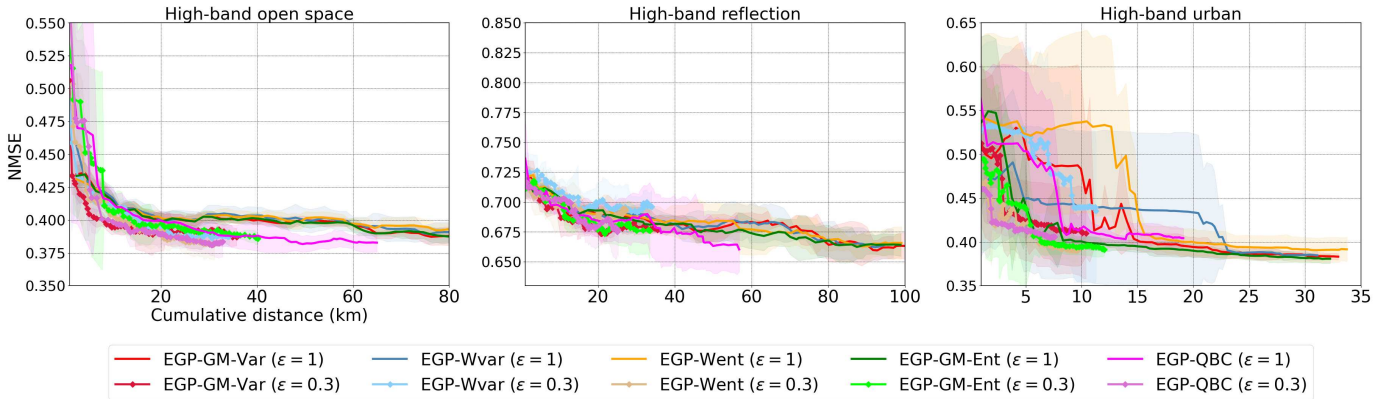


Fig. 8: NMSE vs cumulative distance 26GHz high-band data under (a) high-band open space; (b) high-band urban canyon; and (c) high-band reflect datasets.

we have gathered real RSS radio measurements from urban and suburban environments. Specifically, we used a smartphone as an Rx, which was connected to a commercial 5G network by running the "iperf3" application to maintain active connection. The phone is also connected to a laptop equipped with the professional software XCAL [34] to read RSS measurements from the phone's chipset-level diagnostics. This enabled us to collect a total of 27k data samples from an urban area of size $0.8\ km^2$, and 18k data samples from a suburban area of size $1.0\ km^2$. Similar to the simulated data, each real data sample is represented by $[x, y, RSS]$. The visualization of the collected real radio RSS map is depicted in Fig. 3.

VI. Evaluation

In this section, we evaluate the performance of our EGP-based AL methods on the simulated 5G mid-band, high-band and real urban and suburban datasets described above. We will first quantify the advantages of the advocated EGP-based AL approaches over random sampling with EGPs and then highlight the benefits of our novel path-aware method. We will also conduct an in-depth study with visualized results to demonstrate the efficacy of those AL algorithms in the 5G measurement problem.

A. Experiment Settings

We denote the advocated EGP-based AL methods that employ the AFs in (22)-(28) as 'EGP-WVar', 'EGP-QBC', 'EGP-GM-Var', 'EGP-Went', and 'EGP-GM-Ent' respectively. We also implement the 'EGP-Random' baseline for comparison, which relies on the EGP model as well but randomly selects new locations to query at each time slot. For all competing approaches, 100 initial labeled (measured) data samples are considered for training, and at each iteration of the AL process, one sample is queried from an unlabeled set consisting of 700 available unmeasured locations. Their performance is evaluated on a test set $\mathcal{T} := \{(\mathbf{x}_{n'}^{\text{test}}, y_{n'}^{\text{test}})\}_{n'=1}^{N^{\text{test}}}$ consisting of $N^{\text{test}} = 5000$ test locations. As a figure of merit the normalized mean squared error (NMSE) is used, which is expressed as $\text{NMSE}_n := \frac{1}{N^{\text{test}}} \sum_{n'=1}^{N^{\text{test}}} (\hat{y}_{n'|n}^{\text{test}} - y_{n'}^{\text{test}})^2 / \sigma_y^2$, where $\hat{y}_{n'|n}^{\text{test}}$ represents the point estimate of the RSS value of test location n' upon having processed location n, and $\sigma_y^2 := \mathbb{E}\|\mathbf{y}_{N^{\text{test}}}^{\text{test}} - \mathbb{E}\{\mathbf{y}_{N^{\text{test}}}^{\text{test}}\}\|^2$, where $\mathbf{y}_{N^{\text{test}}}^{\text{test}} := [y_1^{\text{test}} \dots y_{N^{\text{test}}}^{\text{test}}]^\top$.

For each method, the EGP model consists of 11 GP learners, each capitalizing on a distinct radial basis function (RBF) kernel with characteristic length scale chosen from the set $\{10^c\}_{c=-4}^{6}$. The kernel hyperparameters of each GP learner are obtained by maximizing the marginal log-likelihood utilizing the *sklearn* Python package. To allow for scalability and online model updates, $D = 50$ RFs are employed to yield the RF-

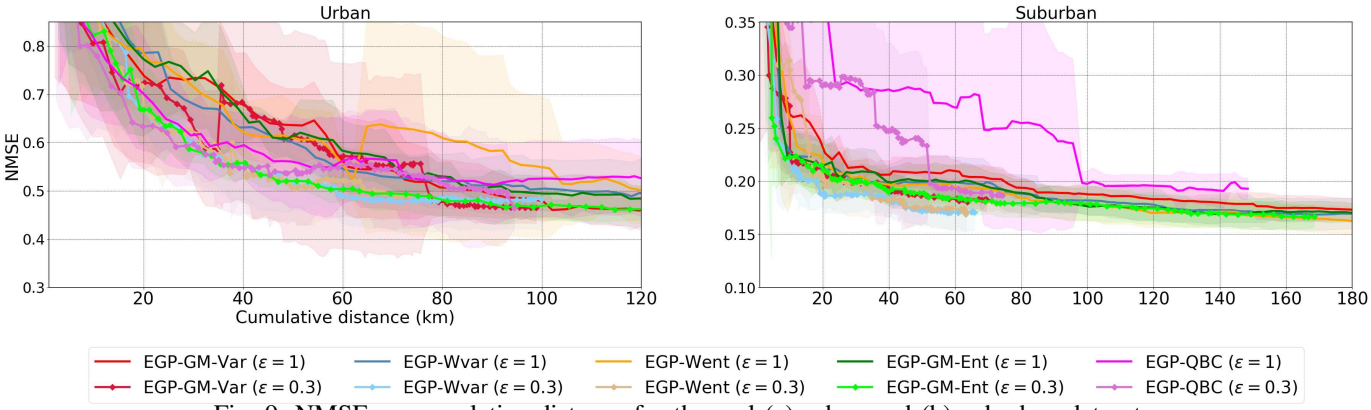
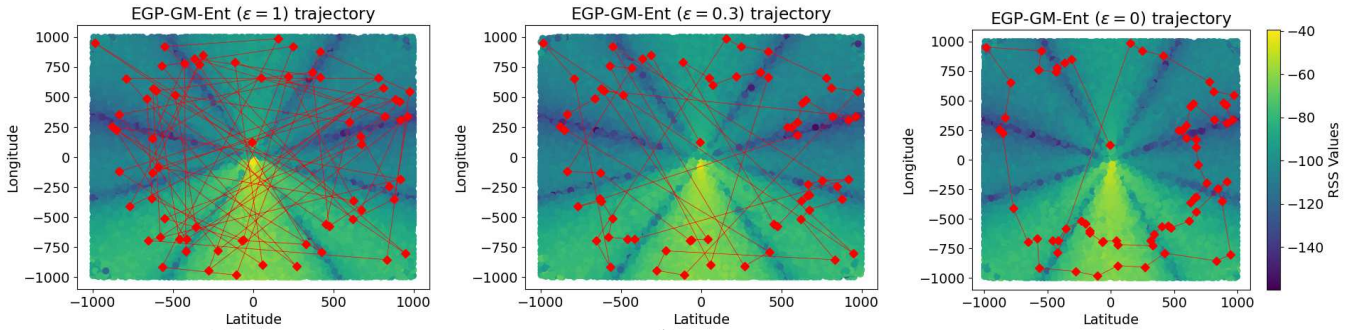


Fig. 9: NMSE vs cumulative distance for the real (a) urban and (b) suburban datasets.



(a) (b) (c)

Fig. 10: Trajectory of the queried locations from the EGP-GM-Ent method for (a) $\varepsilon = 1$; (b) $\varepsilon = 0.3$; and $\varepsilon = 0$ in the mid-band open space dataset.

based parametric and approximate GP models in the ensemble.

B. Performance of AL in the 5G Measurement Problem

Effectiveness of EGP-based AFs. To demonstrate the significance of the EGP-based AFs, we first compare them with the 'EGP-Random' baseline. To this end, the path-cost function $c(\mathbf{x}, \mathbf{x}_{n-1})$ is not taken into account; that is $\varepsilon = 1$ in (30).

Figs. 4, 5 and 6 illustrate the average NMSE performance of all competing approaches over 10 independent runs along with the corresponding highlighted standard deviation, on the mid-band, high-band and real urban and suburban datasets, respectively, for 70 iterations of the AL process; that is 70 queried locations that are measured. It is evident that in all mid-band, in almost all high-band and in all real datasets, the advocated EGP-based AL methods markedly outperform the 'EGP-Random' baseline that does not rely on AL, achieving up to 54.8% improvement. This showcases the significance of judiciously selecting the few, but most informative, locations to measure so as to accurately predict the RSS values in unknown locations. Although effective, these AFs by themselves do not impose constraints on the traveling distance needed to collect the necessary measurements, which may limit their application in settings with restricted travel budgets.

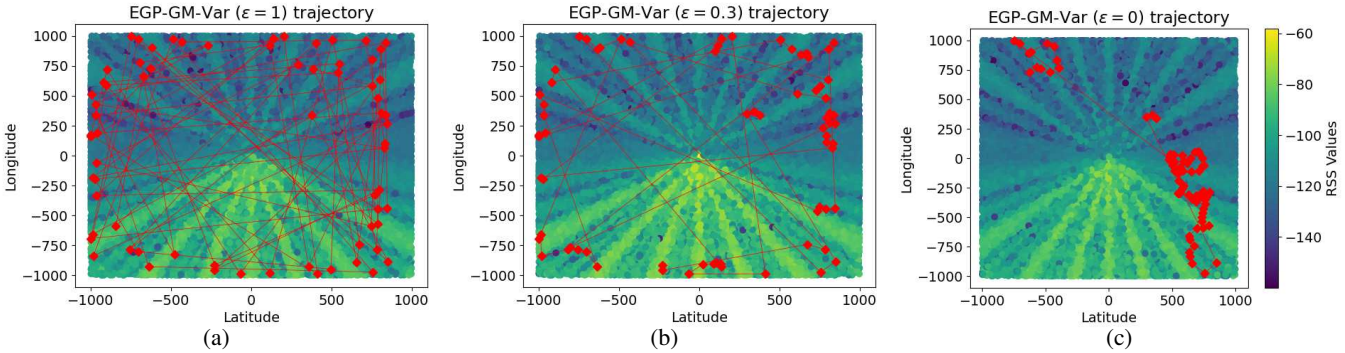
Path-aware AF. We now show the importance of adopting the novel path-cost related AF in (30) that allows for accurate RSS estimates with less required traveling distance. Regarding the exploration-exploitation parameter ε in (30), we set $\varepsilon = 0.3$, meaning that our method selects a location to query not far away with probability 0.7 or selects any location without any path constraint with probability 0.3, allowing for further exploration.

In Figs. 7, 8 and 9, the average NMSE performance along with the corresponding standard deviation are reported, with respect to the cumulative traveling distance for 70 iterations of the AL process. It can be clearly seen that in both simulated mid-band and high-band, and real urban and suburban datasets, all EGP-AL methods that use (30) with $\varepsilon = 0.3$ have comparable or even lower NMSE compared to the EGP-AL counterparts that do not consider any path constraint ($\varepsilon = 1$), while at the same time requiring significantly less traveling distance (up to 57.1% reduction in terms of required cumulative traveling distance).

C. In-depth Results Study

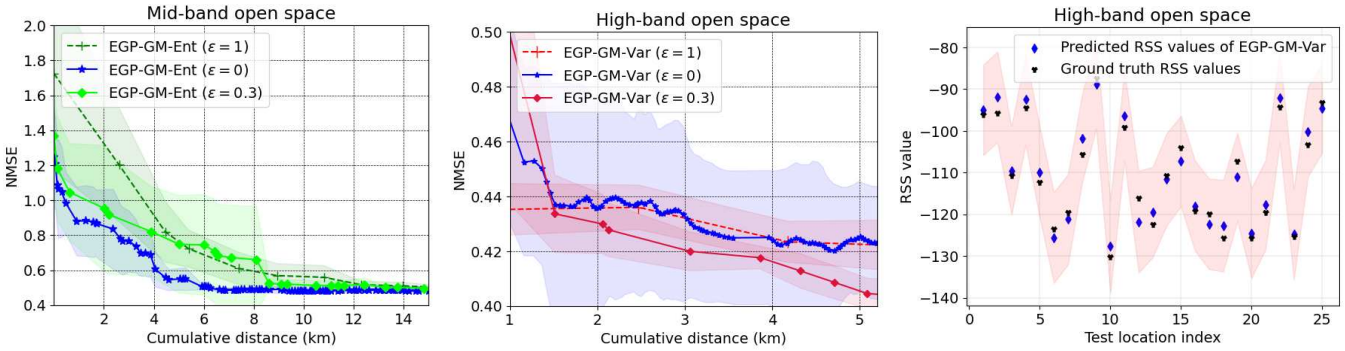
Sampling Trajectory. To further demonstrate the efficiency of the advocated method, Figs. 10 and 11 depict the trajectories of the queried locations obtained by the 'EGP-GM-Ent' and 'EGP-GM-Var' AFs for the mid-band and high-band open space datasets, respectively, for $\varepsilon \in \{0, 0.3, 1\}$[3]. As expected, in both datasets the corresponding required traveling distance significantly reduces as the value of ε decreases. Interestingly, in the mid-band open space dataset, the EGP-GM-Ent approaches with $\varepsilon = 1$ and $\varepsilon = 0.3$ query very similar measurement locations; however, the latter entails much less traveling cost, as shown in Fig. 10.

[3]Trajectories of other AL methods and datasets are omitted due to space limitations.



(a) (b) (c)

Fig. 11: Trajectory of the queried locations from the EGP-GM-Var method for (a) $\varepsilon = 1$; (b) $\varepsilon = 0.3$; and $\varepsilon = 0$ in the high-band open space dataset.



(a) (b) (c)

Fig. 12: NMSE vs cumulative distance of the EGP-GM-Ent method for different values of ε for (a) mid-band open space dataset; (b) high-band open space dataset; and (c) uncertainty quantification performance of EGP-GM-Var on the High-band open space dataset (RSS values are in dB).

Hyper-Parameter Sensitivity. Fig. 12a illustrates the NMSE performance of the 'EGP-GM-Ent' methods with $\varepsilon = 1, 0.3, 0$ respectively for a fixed cumulative distance budget, where it can be seen that the latter two can query more locations and thus achieve superior NMSE performance compared to the $\varepsilon = 1$ case. Similarly, Fig. 12b depicts the NMSE performance of the EGP-GM-Var approach for different values of ε and a given distance budget for the high-band OpenSpace dataset. It is evident that the case of $\varepsilon = 0.3$ enjoys substantially lower NMSE compared to the $\varepsilon = 0$ and $\varepsilon = 1$ cases whereas the AL methods with $\varepsilon = 0$ and $\varepsilon = 1$ have similar performance. This emanates from the fact that although the approach with $\varepsilon = 0$ queries many more locations compared to $\varepsilon = 1$, for a distance budget of 5.2km, these locations are only from two clusters and are not from different regions of the radio map as shown in Fig. 11c. This showcases the importance of small but non-zero values of ε that do not so *heavily* penalize distant locations but instead allow for further exploration that yields improved prediction performance without the need for large traveling costs. Finally, to evaluate the well-quantifiable uncertainty offered by the advocated EGP-AL methods, Fig. 12c shows the predicted RSS values of the EGP-GM-Var ($\varepsilon = 0.3$) method on some test locations of the high-band OpenSpace dataset along with σ-confidence intervals, where it is intuitive that the ground truth RSS values fall inside these intervals.

VII. Conclusions

This work contributes novel Bayesian AL approaches for sample- and path-efficient 5G radio measurement problems. The advocated AL approaches judiciously select only a few representative locations to collect RSS measurements. To effectively estimate the sought function that predicts RSS values across the coverage area, an ensemble of GPs is utilized. This EGP model not only provides accurate RSS predictions with quantifiable uncertainty, but also offers a suite of pertinent AFs to guide the acquisition of new locations to measure. To further accommodate real-world 5G measurement campaigns, where traveling distances to collect RSS measurements might be limited, these AFs are coupled with a novel path-efficient rule that penalizes queries at long distances. Numerical tests on three simulated scenarios with two different radio frequencies, showcase the significance of the proposed AL methods in the 5G radio map reconstruction task.

References

[1] F. C. Commission, "America's 5G future," https://www.fcc.gov/5G, 2023, accessed: 2023-07-29.

[2] R. A. Fezeu, E. Ramadan, W. Ye, B. Minneci, J. Xie, A. Narayanan, A. Hassan, F. Qian, Z.-L. Zhang, J. Chandrashekar *et al.*, "An in-depth measurement analysis of 5G mmWave PHY latency and its impact on end-to-end delay," in *Intl. Conf. on Passive and Active Netw. Meas.* Springer, 2023, pp. 284–312.

[3] K. Shaw, "What are the benefits of 5G mmWave?" https://www.verizon.com/business/resources/articles/s/what-are-the-benefits-of-5g-mmwave/, 2023, accessed: 2023-07-29.

[4] R. de Fremery, "What is C-Band, and what does it mean for 5G?" https://www.verizon.com/business/resources/articles/s/what-is-c-band-and-what-does-it-mean-for-5g/, 2023, accessed: 2023-07-29.

[5] S. A. Kali Pickens, "What is mmWave technology?" https://www.samsung.com/global/business/networks/insights/podcasts/0110-what-is-mmwave-technology/, 2023, accessed: 2023-07-29.

[6] A. Narayanan, E. Ramadan, R. Mehta, X. Hu, Q. Liu, R. A. Fezeu, U. K. Dayalan, S. Verma, P. Ji, T. Li *et al.*, "Lumos5G: Mapping and predicting commercial mmWave 5G throughput," in *Proc. of ACM Intl. Meas. Conf.*, New York, NY, October 27 - 29, 2020, pp. 176–193.

[7] W. Ye, X. Hu, T. Liu, R. Sun, Y. Li, and Z.-L. Zhang, "5GNN: extrapolating 5G measurements through GNNs," in *Proc. of the 1st Intl. Workshop on Graph Neural Networking*, ser. GNNet '22. New York, NY, USA: Association for Comput. Machinery, 2022, p. 36–41.

[8] Y. Zhang and S. Wang, "K-nearest neighbors Gaussian process regression for urban radio map reconstruction," *IEEE Commun. Lett.*, vol. 26, no. 12, pp. 3049–3053, 2022.

[9] M. Hamid and B. Beferull-Lozano, "Non-parametric spectrum cartography using adaptive radial basis functions," in *Intl. Conf. on Acoust., Speech, and Signal Proc.,*, New Orleans, LA, March 5-9 2017, pp. 3599–3603.

[10] B. Khalfi, B. Hamdaoui, and M. Guizani, "Airmap: scalable spectrum occupancy recovery using local low-rank matrixapproximation," in *IEEE Global Commun. Conf.*, Abu Dhabi, UAE, Dec. 9-13, 2018, pp. 206–212.

[11] F. Shen, Z. Wang, G. Ding, K. Li, and Q. Wu, "3D compressed spectrum mapping with sampling locations optimization in spectrum-heterogeneous environment," *IEEE Trans. Wireless Commun.*, vol. 21, no. 1, pp. 326–338, 2021.

[12] G. Zhang, X. Fu, J. Wang, X.-L. Zhao, and M. Hong, "Spectrum cartography via coupled block-term tensor decomposition," *IEEE Trans. Signal Process.*, vol. 68, pp. 3660–3675, 2020.

[13] G. Chen, Y. Liu, T. Zhang, J. Zhang, X. Guo, and J. Yang, "A graph neural network based radio map construction method for urban environment," *IEEE Commun. Lett.*, vol. 27, no. 5, pp. 1327–1331, 2023.

[14] J.-H. Seong and D.-H. Seo, "Selective unsupervised learning-based Wi-Fi fingerprint system using autoencoder and GAN," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 1898–1909, 2019.

[15] Q. Niu, Y. Nie, S. He, N. Liu, and X. Luo, "RecNet: a convolutional network for efficient radiomap reconstruction," in *IEEE Intl. Conf. on Commun.*, Kansas City, MO, May 20-24, 2018, pp. 1–7.

[16] H. Zou, C.-L. Chen, M. Li, J. Yang, Y. Zhou, L. Xie, and C. J. Spanos, "Adversarial learning-enabled automatic wifi indoor radio map construction and adaptation with mobile robot," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 6946–6954, 2020.

[17] C. E. Rasmussen and C. K. Williams, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006.

[18] Q. Lu, G. Karanikolas, Y. Shen, and G. B. Giannakis, "Ensemble gaussian processes with spectral features for online interactive learning with scalability," in *Proc. of AISTATS*, Palermo, Italy, June 3-5, 2020, pp. 1910–1920.

[19] D. Romero, D. Lee, and G. B. Giannakis, "Blind radio tomography," *IEEE Trans. Sig. Process.*, vol. 66, no. 8, pp. 2055–2069, 2018.

[20] W. Liu and J. Chen, "Geography-aware radio map reconstruction for uav-aided communications and localization," in *-IEEE Intl. Conf. Communications*. IEEE, 2021.

[21] V.-P. Chowdappa, C. Botella, J. J. Samper-Zapater, and R. J. Martinez, "Distributed radio map reconstruction for 5g automotive," *IEEE Intel. Transport. Sys. Magazine*, vol. 10, no. 2, pp. 36–49, 2018.

[22] S. Roger, M. Brambilla, B. C. Tedeschini, C. Botella-Mascarell, M. Cobos, and M. Nicoli, "Deep-learning-based radio map reconstruction for v2x communications," *IEEE Trans. Vehicular Tech.*, 2023.

[23] O. Sener and S. Savarese, "Active learning for convolutional neural networks: A core-set approach," *Intl. Conf. on Learn. Rep.*, Vancouver, BC, April 30-May 3, 2018.

[24] J. T. Ash, C. Zhang, A. Krishnamurthy, J. Langford, and A. Agarwal, "Deep batch active learning by diverse, uncertain gradient lower bounds," *Intl. Conf. on Learn. Rep.*, New Orleans, LA, May 6-9, 2019.

[25] K. Jedoui, R. Krishna, M. Bernstein, and L. Fei-Fei, "Deep bayesian active learning for multiple correct outputs," *arXiv:1912.01119*, 2019.

[26] A. Locatelli, A. Carpentier, and S. Kpotufe, "Adaptivity to noise parameters in nonparametric active learning," in *Proc. of Conf. on Learn. Theory*, Amsterdam, Netherlands, July 7-10, 2017.

[27] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Active learning with Gaussian processes for object categorization," in *IEEE Intl. Conf. on Computer Vision*, Rio de Janeiro, Brazil, Oct 14-21, 2007, pp. 1–8.

[28] A. Krause and C. Guestrin, "Nonmyopic active learning of Gaussian processes: an exploration-exploitation approach," in *Proc. of Intl. Conf. on Machine Learn.*, Corvalis Oregon, June 20-24, 2007, pp. 449–456.

[29] K. Miller and J. Calder, "Poisson reweighted laplacian uncertainty sampling for graph-based active learning," *SIAM Journal on Mathematics of Data Science*, vol. 5, no. 4, pp. 1160–1190, 2023.

[30] J. Huang, Z. Gojcic, M. Atzmon, O. Litany, S. Fidler, and F. Williams, "Neural kernel surface reconstruction," in *Proc. IEEE/CVF Conf. Comp. Vision Pat. Recog.*, 2023, pp. 4369–4379.

[31] K. D. Polyzos, Q. Lu, and G. B. Giannakis, "Weighted ensembles for active learning with adaptivity," *arXiv:2206.05009*, 2022.

[32] J. Chapman, B. Chen, Z. Tan, J. Calder, K. Miller, and A. L. Bertozzi, "Novel batch active learning approach and its application on the synthetic aperture radar datasets," in *Algorithms for Synthetic Aperture Radar Imagery XXX*, vol. 12520. SPIE, 2023, pp. 96–111.

[33] G. A. D. Communities, "Cellsignalstrengthnr," https://developer.android.com/reference/android/telephony/CellSignalStrengthNr, 2023, accessed: 2023-07-29.

[34] ACCUVER, "XCAL: PC based advanced 5G network optimization solution," accessed: 2023-07-29. [Online]. Available: https://accuver.com/sub/products/view.php?idx=6

[35] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," *Proc. of Neural Info. Process. (NeurIPS)*, pp. 1177–1184, Vancouver, B.C., Dec. 8-11, 2008.

[36] J. Hoydis, "Sionna RT: scene creation with Blender using OpenStreetMap," https://www.youtube.com/watch?v=7xHLDxUaQ7c, 2023, accessed: 2023-07-10.

[37] blender.org, "Blender," https://www.blender.org/, 2023, accessed: 2023-07-10.

[38] J. Hoydis, S. Cammerer, F. Ait Aoudia, A. Vem, N. Binder, G. Marcus, and A. Keller, "Sionna: An open-source library for next-generation physical layer research," *arXiv:2203.11854*, Mar. 2022.