Adaptive Sampling for Fingerprinting Localization

Xinze Li, Hanan Al-Tous, and Olav Tirkkonen
Department of Information and Communications Engineering, Aalto University, Finland
Email: {xinze.li@aalto.fi, hanan.al-tous, olav.tirkkonen}@aalto.fi.

***Abstract*—Accurate fingerprint localization requires extensive numbers of measurements to build a comprehensive data set. We consider adaptive sampling to reduce the cost of measurement campaigns. Existing adaptive sampling algorithms target scalar or low-dimensional features and do not generalize to high-dimensional Channel State Information (CSI) features. To address this challenge, we create a radio environment map using Kriging interpolation. Sample points with high variances are identified in the 2D coverage area. The fingerprints of these points are obtained and used to adaptively update the data set. Simulation results demonstrate the effectiveness of proposed adaptive sampling strategy in improving localization accuracy while significantly reducing data collection effort. Compared to uniformly random sampling, the proposed adaptive sampling method enhances localization accuracy by 35% in root mean squared error when using weighted K nearest neighbor regression. An equivalent localization performance is achieved with 70% fewer training points compared to sampling uniformly at random.**

***Index Terms*—Adaptive sampling, channel state information, fingerprint localization, radio map.**

I. Introduction

Accurate localization has become an essential capability in modern wireless networks, driven by the rapid expansion of location-aware applications, particularly in mission-critical domains such as the Industrial Internet of Things (IIoT). Among various localization techniques, fingerprinting has emerged as a practical and robust solution due to its effectiveness in complex indoor environments and challenging multipath scenarios [1].

Typically, fingerprinting localization involves two phases: an offline training phase, where a model is constructed by collecting fingerprints at predetermined locations, and an online phase, where real-time measurements are matched against the model to estimate locations. Common methods employed for the matching process include Weighted K-Nearest Neighbor (WKNN), Gaussian Process Regression (GPR) and Deep Neural Networks (DNN) [2]. There are a wide options of radio measurements applied for fingerprint localization, such as Received Signal Strength (RSS), Channel Impose Response (CIR) and channel covariance matrices. Among these, channel covariance feature offers superior localization performance while requiring fewer DNN parameters [3], [4].

Radio maps describe how radio waves spread across a geospatial region. A radio map captures spatial distributions of radio measurements such as RSS, Signal-to-Noise Ratio (SNR), or Channel State Information (CSI) across specific geographic areas [5]. The radio map inherently indicates spatial correlations between radio signals and their physical locations, serving critical roles in network planning and localization [6]. However, constructing a comprehensive radio map is resource-intensive and time-consuming, as it requires extensive channel measurements at numerous discrete locations [7].

Identifying the most representative or informative samples can mitigate these challenges. In order to obtain the most representative samples, informative sampling is often driven by Bayesian optimization acquisition functions or by maximum-entropy criteria [8]–[10]. However, standard Bayesian optimization acquisition functions are designed primarily for scalar optimization problems, making them unsuitable for spatially correlated vector-based data, such as CSI-based radio maps. The vector-extension proposed in [10] alleviates this mismatch yet remains computationally prohibitive for high-dimensional inputs. Maximum entropy sampling, a.k.a batch entropy sampling focuses on selecting a subset of a finite universe to maximize information gain. It faces the same issue of computational complexity even for scalar features [8].

Adaptive sampling methods specifically aim at improving fingerprint localization accuracy by intelligently selecting measurement locations. Liu et al. demonstrate that adaptive tubal-sampling on RSS measurements, significantly enhancing localization accuracy [11]. Likewise, Zheng et al. apply an adaptive approach to optimize sampling strategies effectively based on compressive sensing [12]. However, the proposed algorithm is tailored for RSS data and causes high computational complexity when applied to high-dimensional CSI features.

Geostatistical interpolation approaches, such as Kriging methods, offer a more fitting framework for CSI radio maps [9], [13]. Kriging inherently accounts for spatial correlations, utilizing prediction variance as a natural measure of uncertainty. This method has already been applied effectively to construct RSS-based radio maps with limited data [14].

We name a *radio mapping*, a task of a mapping from physical locations to the features. Both fingerprint localization and radio mapping can be constructed from the same data set. The difference is based on the available information in the online phase; for localization the CSI feature is measured, whereas for radio mapping the physical location is provided. We utilize a Bayesian approach to study the relation between adaptive sampling for localization and radio mapping.

In this paper, we propose an adaptive sampling method for CSI-based fingerprint localization with high dimensional features. Our approach explores geostatistical information from Kriging to efficiently identify highly informative sampling points. The primary contributions of this paper are as follows:

- We construct accurate CSI-based radio maps by adopting the channel covariance feature that mitigates the small-

scale fading effect.

- We utilize Kriging interpolation to estimate CSI measurements at unmeasured locations and use the Kriging estimation variance to indicate the quality of estimation.
- We develop a batch adaptive sampling strategy guided by Kriging variance and a clustering algorithm that enhances the informativeness of measurement points and improves the sampling efficiency.
- We follow an iterative approach of enhancing the radio map and use it to find adaptive samples for fingerprint localization.

The remainder of this paper is organized as follows: In Section II, the system model is introduced. In Section III, the adaptive sampling approach is presented. In Section IV, the fingerprint localization and radio map estimation methods are discussed. Simulation results are presented and discussed in Section V. Finally, conclusions are drawn in Section VI.

II. System Model

We consider a communication system with B base stations (BSs), each BS having a uniform linear array (ULA) of M antennas. User Equipment (UE) have one omnidirectional antenna. We assume transmissions based on orthogonal frequency-division multiplexing (OFDM) with N subcarriers, with a cyclic prefix larger than the maximum delay spread of the channels.

The channel vector between UE u and BS b over subcarrier n at time-sample s is

$$\mathbf{h}_{u,b,n,s} \in \mathbb{C}^{M\times 1}\,.$$

The channel coefficients model both path loss as well as large scale and small scale multipath fading effects.

A. Covariance Matrix

The covariance matrix of the channel between UE u and BS b is

$$\mathbf{R}_{u,b} = \frac{1}{S\,N}\sum_{s=0}^{S-1}\sum_{n=0}^{N-1}\mathbf{h}_{u,b,n,s}\,\mathbf{h}_{u,b,n,s}^{\mathrm{H}}\,, \tag{1}$$

where $\mathbf{R}_u \in \mathbb{C}^{M\times M}$, and S is the number of time samples. In $\mathbf{R}_{u,b}$, the effect of small scale fading is averaged which endows covariance-based features with robustness to small scaling fading. The channel covariance captures power and angle of arrival information. Note that the channel covariance feature $\mathbf{R}_{u,b}$ is smaller size as compared to the CSI feature$\{\mathbf{h}_{u,b,n,s}\}$ where all frequency and time samples of UE u the channel are stacked. The covariance matrix is used as a feature for localization in [3], [4]. For a fixed transmission power, the received power $p_{u,b}$ is proportional to the average channel gain, $p_{u,b} \sim \mathrm{Tr}\;\mathbf{R}_{u,b}$.

B. Logarithm Covariance Feature

For a positive definite Hermitian matrix $\mathbf{R}$, the matrix logarithm can be calculated through eigenvalue decomposition:

$$\mathbf{R} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}},$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix. Then

$$\log(\mathbf{R}) = \mathbf{U}\,\mathrm{diag}\left([\log\lambda_1,\ldots,\log\lambda_M]\right)\mathbf{U}^{\mathrm{H}},$$

where $\lambda_m > 0$ for $m = 1,\ldots,M$. Studies have shown that the log-covariance feature outperforms the standard covariance feature for localization and channel charting applications [4], [15].

We create the corresponding one BS feature $\mathbf{z}^{(b)} \in \mathbb{R}^{M^2\times 1}$ by extracting the real diagonal elements, followed by the real and imaginary parts of the sub-diagonal elements from the Hermitian matrix. Then the *log-covariance* feature is constructed from all BSs, i.e., $\mathbf{z} = [\mathbf{z}^{(1)\mathrm{T}},\ldots,\mathbf{z}^{(B)\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{F\times 1}$, where $F = BM^2$ is the feature length.

C. Data Set Construction

For fingerprint localization and radio mapping, we consider a data set comprising measured CSI features and their corresponding locations, represented as $\mathcal{D} \in \{(\mathbf{z}_u,\mathbf{x}_u)\}_{u=1}^{U}$, with $\mathbf{x}_u \in \mathbb{R}^{2\times 1}$ and $\mathbf{z}_u \in \mathbb{R}^{F\times 1}$.

In this paper, our analysis targets limited data set sizes. Notably, DNN-based fingerprint localization approaches typically require substantial training data, thus limiting their efficacy in scenarios with small size data sets. Consequently, this study emphasizes the evaluation and utilization of GPR and WKNN localization, which perform well with small size data sets [16], [17]

III. Problem Formulation and Solution

Since constructing measurement-based radio maps demands considerable resources due to extensive data acquisition requirements, (i.e., time-consuming and labor-intensive) our investigation focus on identifying optimal strategies for selecting the most representative samples. We aim to efficiently establish accurate and reliable fingerprint databases while minimizing data collection efforts.

A. Localization versus Radio Mapping

Fingerprint localization can be considered as a mapping from CSI feature to location coordinates and their associated uncertainty, expressed as

$$\Gamma_{\mathrm{L}} : \mathbf{z}_j \mapsto \mathbf{x}_j, \boldsymbol{\sigma}^2_{\mathrm{x},j}\,, \tag{2}$$

where $\boldsymbol{\sigma}^2_{\mathrm{x},j} \in \mathbb{R}^{2\times 1}_{+}$ denotes the variance of the location estimate [3]. In this formulation, large output variances indicate that the data set / model is not able to accurately predict the location. This indicates that the corresponding pair $(\mathbf{z}_j,\mathbf{x}_j)$ should be added to the fingerprint database to improve accuracy. Because the localization model Γ_{L} relies on acquired measurements to evaluate uncertainty, it raises the question: *Can similar uncertainty analysis be conducted prior to data acquisition?*

To address this, we consider the inverse mapping approach, i.e., radio mapping:

$$\Gamma_{\mathrm{RM}} : \mathbf{x}_j \mapsto \mathbf{z}_j, \boldsymbol{\sigma}^2_{\mathrm{z},j}\,, \tag{3}$$

where $\boldsymbol{\sigma}^2_{\mathrm{z},j} \in \mathbb{R}_+^{F\times 1}$ denotes the uncertainty of the predicted CSI feature. We consider radio map estimations that provide both the feature estimate and its variance. This estimator is trained on a landmark data set $\mathcal{D}$. For any location $\mathbf{x}_j \notin \mathcal{D}$, we determine corresponding uncertainty indicator $\boldsymbol{\sigma}^2_{\mathrm{z},j}$. A high variance implies that the existing dataset inadequately represents the CSI features at $\mathbf{x}_j$, and direct measurement at $\mathbf{x}_j$ is required to improve the model Γ_{RM}.

The difference between Γ_{RM} and Γ_{L} lies in their searching spaces. The radio mapping evaluates CSI uncertainty over the two-dimensional physical space, whereas the localization model operates in a high-dimensional CSI feature space, becoming computationally prohibitive to search over F as it grows. Furthermore, the physical domain is known as a prior, i.e., the boundary of the localization area, whereas the CSI feature domain is not known a prior. It is worth to mention that the localization framework itself can be used to check if a measured feature is informative or not.

B. Motivation Example

Wireless signals exhibit spatial smoothness in channel covariance, enabling interpolation techniques like GPR and Kriging for estimating unmeasured locations or features[1]. As GPR is hard to implemented with high dimensional outputs, we use Kriging for radio mapping framework Γ_{RM}, while GPR is employed for localization model Γ_{L}. Both models provide an estimation and its associated uncertainty. We train both models on a data set $\mathcal{D}$ and construct a test data set $\mathcal{D}_{\mathrm{T}} \in \{(\mathbf{z}_j, \mathbf{x}_j)\}$. During testing, we identify the top J points with the highest total variance from each model and assign then to the sets $\mathcal{D}_{\mathrm{RM}}$ and $\mathcal{D}_{\mathrm{L}}$. The total variance is computed as $\sigma^2_{\mathrm{q},j} = \mathbf{1}^{\mathrm{T}}\boldsymbol{\sigma}^2_{\mathrm{q},j}$, where $\mathrm{q} \in \{\mathrm{x},\mathrm{z}\}$ and $\mathbf{1}$ is an all-ones vector of appropriate dimension. We define the intersection of the two sets

$$\mathcal{D}_{\mathrm{in}} = \mathcal{D}_{\mathrm{RM}} \cap \mathcal{D}_{\mathrm{L}}\,.$$

The overlap ratio $|\mathcal{D}_{\mathrm{in}}|/J$ quantifies the relation between the two models identifying high-uncertainty points, and $|\mathcal{J}|$ is the cardinality of set $\mathcal{J}$.

We consider 200 training samples to learn Γ_{L} and Γ_{RM} and 1800 test samples. The CSI features are created by the Quadriga simulator as detailed in Section V. Figure 1, illustrates the idea of the motivation example. The intersection set size $|\mathcal{D}_{\mathrm{in}}|$ and the overlap ratio are reported in Table I. The single most uncertain point is the same for both localization and radio mapping models. As J increases, the overlap ratio decreases. The radio mapping can be used to detect localization uncertainty regions by searching over the two-dimensional physical space, motivating an adaptive sampling strategy guided by estimation variance of Γ_{RM}.

C. Adaptive Sampling for Localization based on Radio Map

The motivation example above shows that single-point selection scheme works well for both radio-map and localization models. However, choosing a batch of $J > 1$ points is discussed

[1]GPR and Kriging are discussed in Section IV.

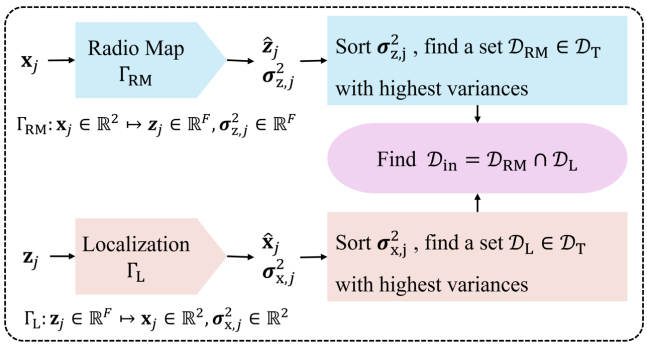

Fig. 1. An illustrative example is presented to find the common set of uncertainty estimations, where Γ_{RM} and Γ_{L} are mapping and its inverse.

TABLE I
THE PERCENTAGE OF THE SAME UNCERTAINTY ESTIMATIONS FOR MODELS Γ_{RM} AND Γ_{L} WITH DIFFERENT J

J	1	5	9	13	25
$\lvert\mathcal{D}_{\mathrm{in}}\rvert/J$ %	100	60	44	46	48

in the literature under maximum Shannon entropy sampling problem, which is NP-hard problem. Several approaches to obtain an efficient solution are discussed in [8].

Can we add multiple samples each time to improve the efficiency without the need to solve a complex optimization problem?

Let us investigate the physical locations with high variances from the motivation example. The total variance heat-map and the top $J = 25$ highest variance points from radio mapping framework are shown in Figure 2. The figure shows that the high variance points are clustered, shown with red cross. It is expected that if a point or few points in each cluster are used to update the radio map, then the performance of radio map will improve. Adding all data points with high variances is not an efficient strategy.

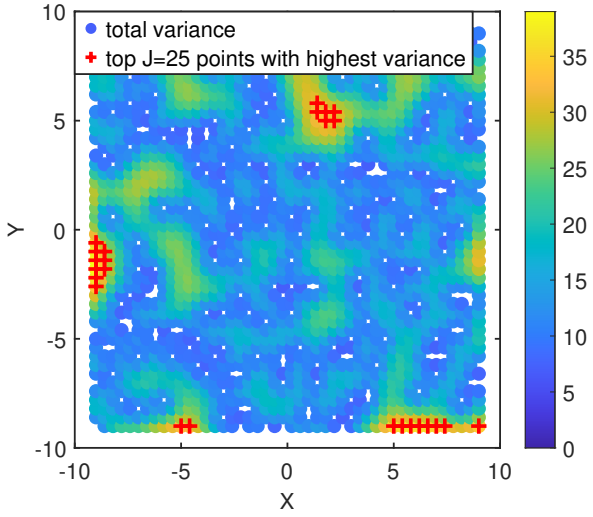

Fig. 2. The heatmap of estimation variance. Landmarks are white colored, and high variance points are red colored.

In this paper, we consider a simple approach for adaptive batch sampling that combines variance ranking with clustering in the physical domain. We first partition the search area into clusters to promote spatial diversity, then select the point with the highest variance within each cluster.

The proposed adaptive sampling approach is described in Algorithm 1. It is initialized with clustered areas of physical locations and randomly selected U_{init} points. In each iteration, the radio map model Γ_{RM} is updated and then used to interpolate the CSI feature, outputting variances for sampling points. The α locations with highest total variance are found from each clustered area. Then considering the measurement at β top-variance sampling points to update the data set. It is worth to mention that the adaptive sampling Algorithm 1 can be used with any localization framework, for examples KWNN and GRP.

Algorithm 1 Adaptive Sampling for Radio Map

Require: Number of initial samples U_{init}, maximum number of samples $U_{\max}$, number of clusters α, number of added points at each iteration β, with $\beta \leq \alpha$.

1: **Start:** Set the number of iterations $i = 1$. Generate random U_{init} sampled locations and obtain corresponding CSI features. Set current data set

$$\mathcal{D} = \{(\mathbf{x}^{(i,1)}, \mathbf{z}^{(i,1)}), \ldots, (\mathbf{x}^{(i,U_{\text{init}})}, \mathbf{z}^{(i,U_{\text{init}})})\}.$$

2: **Cluster locations:** Create a 2D grid of locations and create α clusters of physical locations.
3: **while** $|\mathcal{D}| < U_{\max}$ **do**
4: **Kriging estimation** using the current data set $\mathcal{D}$.
5: **Evaluate the features and variances** for the 2D grid of physical locations.
6: **Find the location with highest variance from each clustered area:** $[\mathbf{x}_\star^{(i,1)}, \ldots, \mathbf{x}_\star^{(i,\alpha)}]$ and total variances $\mathbf{v}_\star = [v_\star^{(i,1)}, \ldots, v_\star^{(i,\alpha)}]$.
7: **Find the next β sampling points:** sort $\mathbf{v}_\star$ and sample at top β corresponding locations.
8: **Obtain the CSI features** $\{\mathbf{z}^{i,l}\}$ at the β locations.
9: **Update the data set:**

$$\mathcal{D} = \mathcal{D} \cup \{(\mathbf{x}^{(i,1)}, \mathbf{z}^{(i,1)}), \ldots, (\mathbf{x}^{(i,\beta)}, \mathbf{z}^{(i,\beta)})\}.$$

10: **Update number of iterations** $i = i + 1$.
11: **end while**

IV. Localization and Radio Map Algorithms

A. Localization

The localization is solved in a supervised way, employing GPR and WKNN methods due to their suitability for small data sets [16], [17].

1) Gaussian Process Regression: Fingerprint-based localization can be formulated as a multi-task GPR problem. Consider a spatial location $\mathbf{x}_u \in \mathbb{R}^2$ and its associated noisy fingerprint observation $\mathbf{z}_u \in \mathbb{R}^F$ [18]:

$$\mathbf{x}_u = f(\mathbf{z}_u) + \boldsymbol{\varepsilon}_u, \quad \boldsymbol{\varepsilon}_u \sim \mathcal{N}(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}), \quad u = 1, \ldots, U, \quad (4)$$

where U is the number of observations, $\mathbf{I}$ is the identity matrix of size 2×2 and σ_ε^2 is the noise variance. Considering an intrinsic coregionalisation GP prior, we can write the Gaussian process

$$f(\mathbf{z}) \sim \mathcal{GP}\Big(\mathbf{0},\, k(\mathbf{z}, \mathbf{z}')\, \mathbf{B}\Big), \quad (5)$$

where $k(\ ,\ )$ is a squared-exponential kernel and $\mathbf{B} \in \mathbb{R}^{2\times 2}$ is a positive-semidefinite coregionalisation matrix that captures inter-feature correlations. Define the block kernel matrices $\mathbf{K}_{UU} = \big[k(\mathbf{z}_i, \mathbf{z}_j)\mathbf{B}\big]_{i,j=1}^{U}$, $\mathbf{K}_{U*} = [k(\mathbf{z}_i, \mathbf{z}_*)\mathbf{B}]_{i=1}^{U}$, and $\mathbf{K}_{**} = k(\mathbf{z}_*, \mathbf{z}_*)\mathbf{B}$.

Then the posterior predictive distribution at a test point $\mathbf{x}_*$ is Gaussian with [19]

$$\mathbb{E}[\mathbf{x}_* \mid \mathcal{D}] = \mathbf{K}_{U*}^{\mathrm{T}}\big(\mathbf{K}_{UU} + \sigma_\varepsilon^2 \mathbf{I}\big)^{-1}\mathbf{X}, \quad (6)$$

$$\mathrm{Cov}[\mathbf{x}_* \mid \mathcal{D}] = \mathbf{K}_{**} - \mathbf{K}_{U*}^{\mathrm{T}}\big(\mathbf{K}_{UU} + \sigma_\varepsilon^2 \mathbf{I}\big)^{-1}\mathbf{K}_{U*}, \quad (7)$$

where $\mathbf{X} = [\mathbf{x}_1^{\mathrm{T}}, \ldots, \mathbf{x}_U^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{2U\times 1}$.

The scalar kernel hyper-parameters, the entries of $\mathbf{B}$, and the noise variance σ_ε^2 are estimated by maximizing the log marginal likelihood approach as in [18].

2) WKNN: WKNN is a straightforward fingerprint localization technique, whose accuracy heavily depends on reference point density, chosen distance metric, weight function, and neighbor count. For UE $*$ to-be localized, the K nearest neighbors in the database are selected using a feature distance, typically Euclidean distance. Let $\mathcal{A}_*$ be the neighbor set of UE $*$, $d_{*,i} = \|\mathbf{z}_* - \mathbf{z}_i\|_2$ is the feature distance between the new CSI feature $\mathbf{z}_*$ and CSI feature $\mathbf{z}_i$. The weight for neighbor i is

$$\omega_{*,i} = \frac{g\,(d_{*,i})}{\sum_{i\in\mathcal{A}_*} g\,(d_{*,i})},$$

where $g(d)$ a weight function and $g(d) = \exp(-\mu\, d)$ is the commonly used exponential weight function with tunable parameter μ [16].

The estimated position is a linear combination of neighbors,

$$\hat{\mathbf{x}}_* = \sum_{i\in\mathcal{A}_*} \omega_{*,i}\, \mathbf{x}_i\,,$$

where $\mathbf{x}_i$ denotes the known position of neighbor i.

B. Radio Map Interpolation

Since radio mapping inverts the localization model, one could apply GPR for feature estimation by swapping inputs and outputs of the localization GPR discussed in Section IV-A1. However, GPR becomes computationally prohibitive for high-dimensional CSI, e.g. the kernel matrix size $UF \times UF$ [17]. Instead, we use Kriging interpolation, which efficiently estimates CSI features by modeling their spatial correlation through a semivariogram, which describes the degree of spatial dependence among CSI features.

1) Semivariogram Modeling: The theoretical semivariogram for a stationary vector field $\mathbf{z}(\mathbf{x})$ is [14]

$$\gamma(\mathbf{h}) = \frac{1}{2}\,\mathbb{E}\left[\|\mathbf{z}(\mathbf{x}) - \mathbf{z}(\mathbf{x}+\mathbf{h})\|_2^2\right]. \tag{8}$$

Given samples $\{\mathbf{z}(\mathbf{x}_u)\}_{u=1}^{U}$, the empirical semivariogram is

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|\mathcal{P}(\mathbf{h})|} \sum_{(i,j)\in\mathcal{P}(\mathbf{h})} \|\mathbf{z}(\mathbf{x}_i) - \mathbf{z}(\mathbf{x}_j)\|_2^2, \tag{9}$$

where $\mathcal{P}(\mathbf{h}) = \{(i,j) : \|\mathbf{x}_i - \mathbf{x}_j\|_2 \approx \|\mathbf{h}\|_2\}$ is the set of all index pairs whose locations are separated by a distance approximately equal to $\|\mathbf{h}\|_2$. We ignore the direction of vector $\mathbf{h}$. Generally, the semivariogram values are function of separation distance h_l, for for multiple steps $l = 1, \ldots, L$. In the case of empirical semivariogram, separation distance interval $h_l \pm \delta$ is used rather than exact distances, and usually isotropic conditions are assumed. One then fits a model (e.g., linear, exponential, Gaussian, etc.) to $\hat{\gamma}(\mathbf{h})$ and then uses it in the Kriging Interpolation.

2) Kriging Interpolation: We wish to estimate the vector-valued random field $\mathbf{z}(\mathbf{x}_*)$ at an unmeasured location $\mathbf{x}_* \in \mathbb{R}^2$, given U noisy observations [20]

$$\mathbf{y}_u = \mathbf{z}(\mathbf{x}_u) + \boldsymbol{\varepsilon}_u, \quad u = 1, \ldots, U, \tag{10}$$

where measurement errors $\{\boldsymbol{\varepsilon}_u\}$ are zero-mean, i.i.d. with covariance $\sigma_\varepsilon^2\mathbf{I}$, and $\mathbf{z}(\mathbf{x})$ is second-order stationary with mean function $\mathbf{m}(\mathbf{x})$ and covariance $C(\mathbf{x}, \mathbf{x}')$.

In Kriging interpolation method, the prediction at the unknown value $\mathbf{z}_*$ at a new location $\mathbf{x}_*$ is from the weighted known values $\mathbf{y}_u$ in the radio map database [14]. We seek the Best Linear Unbiased Estimator (BLUE) [20], [21]

$$\hat{\mathbf{z}}_* = \sum_{u=1}^{U} \lambda_u\, \mathbf{y}_u + \boldsymbol{\lambda}_0, \tag{11}$$

with weights $\boldsymbol{\lambda} = [\lambda, \ldots, \lambda_n]^{\mathrm{T}}$ and $\boldsymbol{\lambda}_0$ chosen to minimize the unbiased mean squared error:

$$\min_{\boldsymbol{\lambda},\boldsymbol{\lambda}_0} \mathbb{E}\left[\|\hat{\mathbf{z}}_* - \mathbf{z}(\mathbf{x}_*)\|_2^2\right], \quad \text{s.t.} \quad \mathbb{E}\left[\hat{\mathbf{z}}_*\right] = \mathbf{m}(\mathbf{x}_*). \tag{12}$$

After applying the Lagrange multipliers theorem, we have [21]

$$\hat{\mathbf{z}}(\mathbf{x}_*) = \mathbf{m}(\mathbf{x}_*) + \mathbf{c}_*^{\mathrm{T}}(\mathbf{C} + \sigma_\varepsilon^2\mathbf{I})^{-1}(\mathbf{Y} - \mathbf{M}), \tag{13}$$

where $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_U]^{\mathrm{T}}$, $\mathbf{M} = [\mathbf{m}(\mathbf{x}_1), \ldots, \mathbf{m}(\mathbf{x}_U)]^{\mathrm{T}}$, the covariance vector and matrices $\mathbf{c}_* = [C(\mathbf{x}_*, \mathbf{x}_1), \ldots, C(\mathbf{x}_*, \mathbf{x}_U)]^{\mathrm{T}}$ and $\mathbf{C} = [C(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^{U}$. The Kriging variance is

$$\sigma_{\mathrm{K}}^2(\mathbf{x}_*) = \mathbf{c}_*^{\mathrm{T}}(\mathbf{C} + \sigma_\varepsilon^2\mathbf{I})^{-1}\mathbf{c}_*. \tag{14}$$

The random field is assumed second order stationary, then the mean function is constant. We assume $\mathbf{m}(\mathbf{x}) = \mathbf{0}$. The covariance and semivariogram are linked by using a lag $\mathbf{h}$ over physical spaces,

$$\gamma(\mathbf{h}) = \frac{1}{2}\,\mathbb{E}\left[\|\mathbf{z}(\mathbf{x}) - \mathbf{z}(\mathbf{x}+\mathbf{h})\|_2^2\right] = C(0) - C(\mathbf{h}). \tag{15}$$

Thus, one can model the semivariogram $\gamma(\mathbf{h})$ and recover $C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$ when assembling $\mathbf{C}$ and $\mathbf{k}_0$.

TABLE II
SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
Center Freq.	3.5 GHz	Subcarrier Spa.	30 kHz
Scenario	InF-SL	Bandwidth	10 MHz
BS Tx Power	20 dBm	Noise Power	-174 dBm
BS Height	1.5 m	UE Height	1 m
BS Array, M	8 ULA	UE Array	1
BSs, B	4		

V. SIMULATION

A. Simulation Setup

We consider Non-Line-of-Sight (NLoS) environment to evaluate performances, specifically an Indoor Factory Sparse Low (InF-SL) scenario defined in [22]. The simulation parameters are summarized in Table II. The factory layout consists of 4 BSs located at xy-coordinates $[-10,\ 10]$ m, $[-10,\ -10]$ m, $[10,\ 10]$ m and $[10,\ -10]$ m, where 2000 UEs are on a grid with 0.4 m spacing. The basis of evaluation is synthetic channel data generated with the QuaDRiGa simulator, considering large-scale and small-scale effects including multi-path fading [23]. Parameters such as delay spread, angle-of-arrival, and angle-of-departure distributions are based on the InF-SL scenario from [22]. We compute the log-covariance feature from $B = 4$ BSs. The UE is set to move in a random direction with a speed of 5 km/h, during the time in which the CSI samples are recorded, ending at the labeled position. We consider $S = 50$ time samples, one taken every 0.5 ms.

B. Kriging and Adaptive Sampling

We set the number of clusters $\alpha = 20$, the number of samples added at each iteration $\beta = 15$, the number of initial samples $U_{\mathrm{init}} = 50$, maximum number of samples $U_{\max} = 300$ and $U_{\max} = 1000$.

Before the Kriging interpolation in each iteration, the parameters of semivariogram are estimated. As the linear model gives the best fits for the empirical semivariogram, we consider a bounded linear model given as

$$\gamma(h) = \begin{cases} \frac{c\,h}{a}, & \text{for } 0 \le h \le a, \\ c, & \text{for } h > a, \end{cases} \tag{16}$$

where c represents sill and a is range and the lag distance is $h = \|\mathbf{x}_i - \mathbf{x}_j\|_2$. The fitted semivariogram is shown in Figure 3.

We apply the purposed batch adaptive sampling to increase the number of training points (landmark points) and investigate the behavior of the updated semivariogram with current training samples. The results are illustrated in Figure 4. The changes in the semivariogram gradually stabilize with more measured samples. After around 300 samples collected, the slope of semivariogram is stabilized.

For Kriging interpolation, we plot the interpolation error

$$e_j = \sqrt{\frac{\|\mathbf{z}_j - \hat{\mathbf{z}}_j\|_2^2}{F}},$$

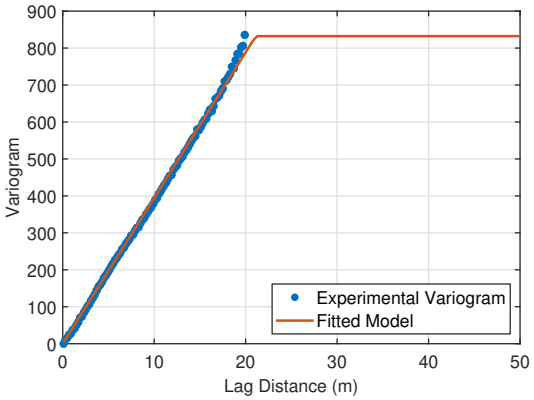

Fig. 3. The semivariogram is fitted with linear model (16), which is created with 1000 points found by adaptive sampling.

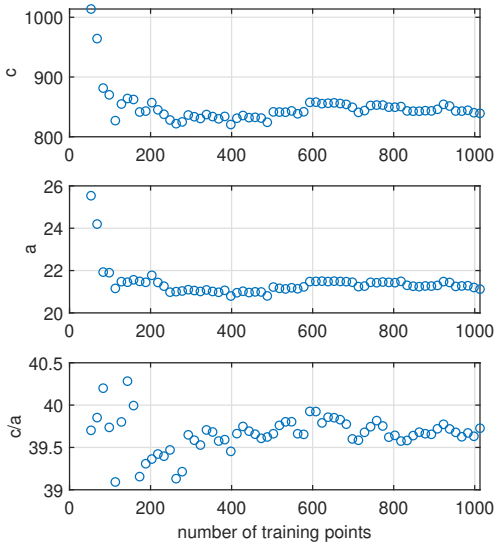

Fig. 4. Semivariogram model parameters, i.e., c, a and c/a change with the number of training points (x-axis).

and estimated total variance, with the number of training points, 50, 300, respectively, shown in Figures 6. From the heatmap, the points near the measured locations have smaller error and variance values. Both error and variance of radio map estimation decreases with more adaptive sampling points added. By expand the size of measurements, the distribution of variance tends to stable during the adaptive sampling as shown in Figure 5.

C. Localization with Adaptive Sampling

To evaluate localization performance, the localization error distribution is shown by Cumulative Distribution Function (CDF). We focus on numerical estimates of the Root Mean Squared Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{U}\sum_{u=1}^{U} \|\mathbf{x}_u - \hat{\mathbf{x}}_u\|_2^2}\,, \tag{17}$$

as well as the 80th and 90th percentile distance error.

We compare the performance of adaptive sampling with benchmark sampling methods as follows:

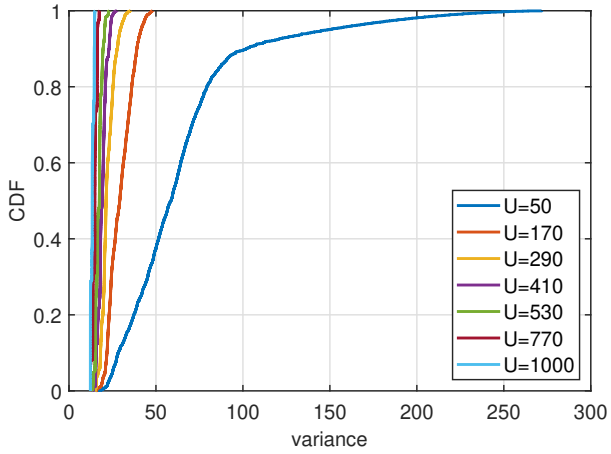

Fig. 5. The distribution of total variance estimation for adaptive sampling.

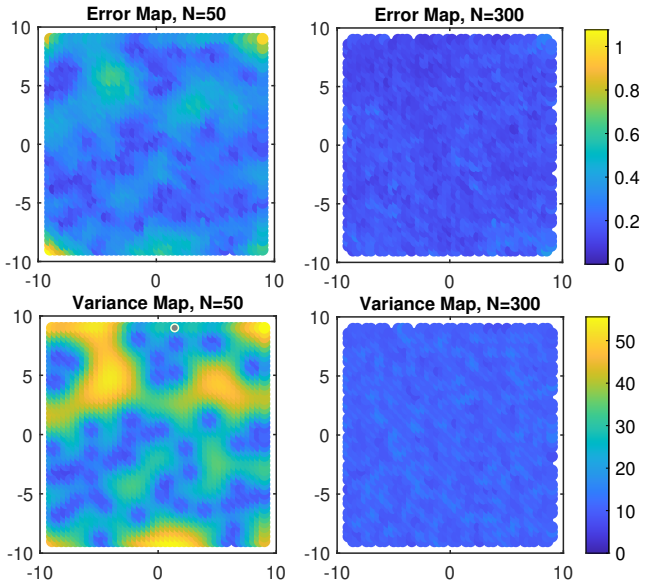

Fig. 6. Heatmap for Kriging interpolation estimation error and variance with 50 random points and 300 points found by adaptive sampling.

- Uniformly random training points
- Single-adaptive sampling: One point is added per iteration.

We consider the following settings:

- $U_{\max} = 300$, corresponding random sampling density = 0.75 sample/m^2.
- $U_{\max} = 1000$, corresponding random sampling density = 2.5 sample/m^2.

The localization error distribution is shown in Figure 7, using WKNN and GPR localization methods. For WKNN, the number of neighbors $K = 5$ and an exponential weighting function is used. The numerical results are summarized in Table III.

Compared to random sampling, the proposed batch-adaptive sampling based on radio maps consistently reduces localization error. With approximately 300 training points, batch sampling lowers the GPR RMSE by 0.12 m and the WKNN RMSE by 0.14 m. Moreover, batch sampling is faster than single-point adaptive sampling while delivering comparable accuracy.

TABLE III
LOCALIZATION ERROR WITH DIFFERENT NUMBER OF TRAINING POINTS.

Method	No. Training Points	Sampling	RMSE	80%	90%
GPR	300	Batch Adaptive	**0.36**	**0.64**	**0.79**
GPR	300	Random	0.48	0.68	0.85
WKNN	300	Batch Adaptive	0.26	0.45	0.53
WKNN	300	Single Adaptive	**0.25**	**0.42**	**0.50**
WKNN	300	Random	0.40	0.66	0.85
WKNN	1000	Batch Adaptive	**0.18**	**0.32**	**0.37**
WKNN	1000	Random	0.24	0.42	0.49

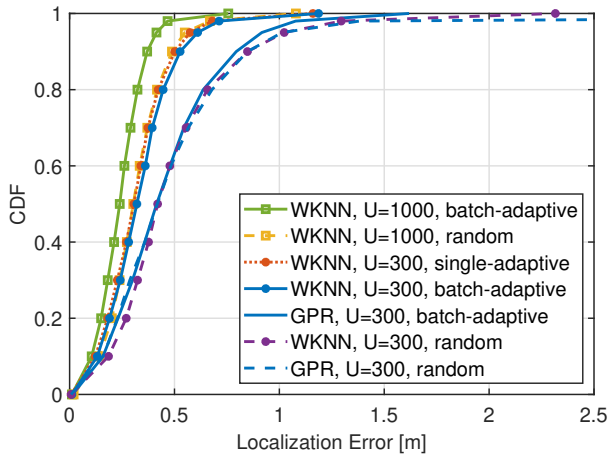

Fig. 7. Localization error distribution with 300 and 1000 training points based on GPR and WKNN. Sampling methods are batch adaptive sampling, single adaptive sampling and randomly selected.

Notably, 300 points by batch-sampled achieve the same WKNN localization performance as 1000 by randomly chosen points.

VI. CONCLUSIONS

In this paper, we have considered a batch adaptive sampling method for constructing CSI-based fingerprint radio maps. By exploiting Kriging-based geostatistical model, our approach efficiently handles high-dimensional CSI by identifying informative sampling points via spatial variance over the 2D domain. Simulations have showed that equivalent localization performance is achieved with 70% fewer training points compared to uniformly random sampling approach, and that batch sampling maintains performance while improving efficiency over single-sampling scheme. Overall, our method enhances fingerprint localization accuracy while minimizing measurement effort and cost. For future works, the number of clusters and batch size can be optimized and we can investigate DNN localization with adaptive sampling.

ACKNOWLEDGMENT

REFERENCES

[1] P. Castro, P. Chiu, T. Kremenek, and R. Muntz, "A probabilistic room location service for wireless networked environments," in *Int. Conf. Ubiquitous Comput.* Springer, 2001, pp. 18–34.

[2] X. Zhu, W. Qu, T. Qiu, L. Zhao, M. Atiquzzaman, and D. O. Wu, "Indoor intelligent fingerprint-based localization: Principles, approaches and challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2634–2657, 2020.

[3] G. Tian, I. Yaman, M. Sandra, X. Cai, L. Liu, and F. Tufvesson, "Deep-learning-based high-precision localization with massive MIMO," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 2, pp. 19–33, 2023.

[4] X. Li, H. Al-Tous, S. E. Hajri, and O. Tirkkonen, "Channel covariance based fingerprint localization," in *IEEE VTC2024-Fall*, 2024, pp. 1–7.

[5] S. Bi, J. Lyu, Z. Ding, and R. Zhang, "Engineering radio maps for wireless resource management," *IEEE Wirel. Commun.*, vol. 26, no. 2, pp. 133–141, 2019.

[6] D. Romero and S.-J. Kim, "Radio map estimation: A data-driven approach to spectrum cartography," *IEEE Signal Process. Mag.*, vol. 39, no. 6, pp. 53–72, 2022.

[7] A. E. C. Redondi, "Radio map interpolation using graph signal processing," *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 153–156, 2017.

[8] C.-W. Ko, J. Lee, and M. Queyranne, "An exact algorithm for maximum entropy sampling," *Operations Research*, vol. 43, no. 4, pp. 684–691, 1995.

[9] J. N. Fuhg, A. Fau, and U. Nackenhorst, "State-of-the-art and comparative review of adaptive sampling methods for kriging," *Archives of Computational Methods in Engineering*, vol. 28, pp. 2689–2747, 2021.

[10] S. Ament, S. Daulton, D. Eriksson, M. Balandat, and E. Bakshy, "Unexpected improvements to expected improvement for bayesian optimization," *Adv. Neural Inf. Process. Syst.*, vol. 36, pp. 20 577–20 612, 2023.

[11] X.-Y. Liu, S. Aeron, V. Aggarwal, X. Wang, and M.-Y. Wu, "Adaptive sampling of RF fingerprints for fine-grained indoor localization," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2411–2423, 2015.

[12] H. Zheng, M. Gao, Z. Chen, X.-Y. Liu, and X. Feng, "An adaptive sampling scheme via approximate volume sampling for fingerprint-based indoor localization," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2338–2353, 2019.

[13] D. G. Krige, "A statistical approach to some basic mine valuation problems on the witwatersrand," *Journal of the Southern African Institute of Mining and Metallurgy*, vol. 52, no. 6, pp. 119–139, 1951.

[14] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *IEEE International Symposium on Dynamic Spectrum Access Networks*, 2012, pp. 422–433.

[15] P. Kazemi, H. Al-Tous, T. Ponnada, C. Studer, and O. Tirkkonen, "Beam SNR prediction using channel charting," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13 130–13 145, 2023.

[16] X. Li, H. Al-Tous, S. E. Hajri, and O. Tirkkonen, "Enhanced weighted k-nearest neighbor positioning," in *IEEE VTC2024-Spring*, 2024, pp. 1–6.

[17] K. Vuckovic, S. Hosseini, F. Hejazi, and N. Rahnavard, "A CSI-based data-driven localization framework using small-scale training datasets in single-site MIMO systems," *IEEE Trans. Wireless Commun.*, 2024.

[18] E. V. Bonilla, K. Chai, and C. Williams, "Multi-task gaussian process prediction," *Adv. Neural Inf. Process. Syst.*, vol. 20, 2007.

[19] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[20] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Statistical Science*, vol. 4, no. 4,409-435, 1989.

[21] M. Marinescu, "Explaining and connecting kriging with gaussian process regression," *arXiv preprint arXiv:2408.02331*, 2024.

[22] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," TS 38.901, Jan. 2018, version 14.3.0.

[23] S. Jaeckel, L. Raschkowski, K. Borner, and L. Thiele, "QuaDRiGa: A 3-D multi-cell channel model with time evolution for enabling virtual field trials," *IEEE Trans. Antennas Propag.*, vol. 62, no. 6, pp. 3242–3256, Jun. 2014.