

UAV-Aided Radio Map Construction Exploiting Environment Semantics

Wenjie Liu^{ID}, *Graduate Student Member, IEEE*, and Junting Chen^{ID}, *Member, IEEE*

Abstract—This paper constructs a full dimensional (6D) radio map to predict the channel gain between any transmitter location and any receiver location based on received signal strength (RSS) measurements between low-altitude aerial nodes and ground nodes. The main challenge is how to describe the signal strength attenuation due to the blockage from the environment. Conventional interpolation-type approaches fail to exploit the close relation between the radio map and the geometry of the environment. This paper proposes to construct radio maps by first estimating and constructing a multi-class 3D virtual obstacle map that describes the geometry of the environment with radio semantics. Mathematically, a least-squares problem is formulated to jointly estimate the virtual obstacle map and the propagation parameters. This problem is found to have a partial quasiconvexity that leads to the development of an efficient parameter estimation and radio map construction algorithm. Numerical experiments confirm that the proposed method substantially reduces the amount of measurement required for achieving the same radio map accuracy. It is also demonstrated that in a unmanned aerial vehicle (UAV)-aided relay communication scenario, a radio map assisted approach for UAV placement can achieve more than 50% capacity gain.

Index Terms—Radio map, environment-aware, radio semantics, unmanned aerial vehicle (UAV).

I. INTRODUCTION

IT BECOMES increasingly important for wireless communication networks to learn about the communication

Manuscript received 25 August 2022; revised 14 December 2022; accepted 28 January 2023. Date of publication 8 February 2023; date of current version 12 September 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1800800, in part by the Basic Research Project through the Hetao Shenzhen-HK S&T Cooperation Zone under Grant HZQB-KCZY-2021067, in part by the National Natural Science Foundation of China under Grant 62171398 and Grant 92067202, in part by the Shenzhen Science and Technology Program under Grant JCYJ20210324134612033 and Grant KQTD20200909114730003, in part by the Guangdong Research under Grant 2019QN01X895 and Grant 2017ZT07X152, and in part by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence under Grant 2022B 1212010001. The associate editor coordinating the review of this article and approving it for publication was J. Hamalainen. (*Corresponding author: Junting Chen.*)

Wenjie Liu is with the School of Science and Engineering and the Future Network of Intelligence Institute (FNii), The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China, and also with the Network Systems Security Group, KTH Royal Institute of Technology, 11428 Stockholm, Sweden (e-mail: wenjieliu@link.cuhk.edu.cn).

Junting Chen is with the School of Science and Engineering and the Future Network of Intelligence Institute (FNii), The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China (e-mail: juntin@link.cuhk.edu.cn).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TWC.2023.3241845>.

Digital Object Identifier 10.1109/TWC.2023.3241845

environment. For example, millimeter-wave, terahertz, and integrated aerial and terrestrial communications favor a line-of-sight (LOS) propagation condition. Thus, knowing the radio environment and even the geometry of the surrounding may help user selection, beamforming, and position optimization for communication nodes [1], [2], [3], [4]. A recent trend is to leverage *radio maps* for optimizing communication networks, where a radio map is a data model that captures the location-dependent wireless channel quality between a transmitter and a receiver [5], [6], [7], [8].

Recent studies have exploited radio maps for unmanned aerial vehicle (UAV) position optimization and trajectory planning for data harvesting, blockage-aware wireless power transfer, and network localization [2], [3], [4], [5], [6], [7], [8]. For example, in a typical scenario of UAV-to-ground communication in dense urban environment, there are buildings and trees that probably appear at arbitrary locations and block the air-to-ground signal. Most earlier works addressed this issue using a *probabilistic* model, which describes the probability of the LOS condition as a function of the elevation angle at the ground node [9], [10]. By contrast, radio map based models can adapt to specific geography environments and determine the actual location-dependent LOS condition and channel gain. Some recent studies on the UAV relay communication further revealed that, when a radio map is partially available, the throughput performance of a position-optimized UAV relay network can be substantially enhanced as compared to the methods based on probabilistic models [2], [7], [8].

However, little is known on how to efficiently construct a full dimensional radio map. The following challenges need to be addressed. First, it usually requires a huge amount of measurement data for radio map construction due to the ample degrees of freedom for a propagation channel. A *full* dimensional radio map for a narrowband single antenna system may still need 6 dimensions to describe the channel quality between any transmitter location and any receiver location in 3D. Second, it is also costly to store, transfer, and share radio maps among communication nodes. Third, it is essential, yet challenging, to embed the environment information to radio maps as wireless channels depend on the geography environment via a complicated mechanism involving signal reflection, diffraction, and scattering. As to be discussed below, pure data-driven environment-blind approaches may result in low efficiency of utilizing the measurement data; classical channel models, such as the simplified probabilistic LOS model, may suffer from poor prediction performance; and ray-

tracing methods suffer from the overwhelming requirement for computing capability and high precision city map information.

This paper attempts to build a full dimensional radio map from received signal strength (RSS) measurements between scattered low altitude aerial nodes and ground nodes. The core idea is to reconstruct the geometry of the propagation environment with *radio semantics* embedded, such that one can infer whether the propagation is under LOS, slightly obstructed, or in deep shadow, and exploit this information to predict the channel gain between any two wireless nodes. Specifically, we build a multi-class 3D environment model, where the environment does *not* necessarily represent the visual appearance of buildings, but it is a model that captures the *semantic* meaning of how the signal strength will be affected by the surrounding. For instance, a pillar made of plastic may impose less attenuation on radio signals than a concrete pillar may do. Mathematically, we construct multi-class 3D virtual obstacles to describe the propagation environment for any pair of wireless nodes in 3D, and based on this, we construct an environment-aware radio map. We show that the proposed approach not only reconstructs the geometry of the propagation environment, but also achieves better accuracy in radio map construction than conventional interpolation-type methods. Moreover, as the information is compressed in the proposed multi-class 3D virtual obstacle model, it becomes easier to convey and share radio maps in the network.

A. Related Work

1) *Data-Driven Approaches*: Radio maps have been studied a lot for indoor localization [11], [12]. In these scenarios, the radio signatures are sampled through RSS measurement over a 2D area, and the focus there was to handle sparsity and interpolate the measurement data. Some of these methods include k -nearest neighbor (KNN) [13], [14], sparse matrix or tensor processing [12], crowdsourcing [15], and Kriging [16], [17], [18]. Based on the recent advance of image processing, deep learning for radio map construction was also investigated in [19], [20], [21], [22], and [23]. Note that these approaches were mainly designed for 2D radio maps, and they may not be easily extended to our scenario of interest.

2) *Model-Based Approaches*: Conventional channel models first label the local area into fine categories, such as urban and sub-urban, and then select a parametric model from the fine category [9], [10]. From the view of signal propagation, the works [24], [25] classify signals into LOS and non-line-of-sight (NLOS) and then adopts path loss model with a city map model. However, the limitation is that the categorization is usually objective, and there are usually a limited number of predefined models and sets of parameters to choose from.

3) *Ray-Tracing*: These methods are based on the 3D model of the environment and compute the physical propagation paths through analyzing any possible reflection, diffraction, and scattering [26], [27]. However, they are not only computationally expensive but also very sensitive to the precision and the accuracy of the information available, including the fine 3D model of the structure and its material.

B. Our Contributions

The paper aims at addressing the following two main issues:

- *How to model the radio map with the geometry of radio environment embedded;*
- *How to develop efficient algorithms to construct the 3D environment and the radio map.*

We develop a radio map model that consists of a multi-class 3D virtual obstacle model to capture the geometry of the radio environment. The key intuition is that if a link is relatively weak considering its propagation distance, then there should be one or more obstacles that block the direct path of the link. With the proposed model, the radio map construction problem is transformed into a joint estimation problem of the propagation parameters and the location and height of the 3D virtual obstacles. A 3D city map is *not* required, although it can help better initialize the algorithm.

Our earlier work [28] clusters the measurement into different propagation conditions, such as LOS and NLOS, and the follow-up work [29], [30] designs a preliminary virtual obstacle model based on the estimated LOS labels. However, the existing approach is an open-loop method, where if the LOS label obtained from [28] were wrong, the error would propagate and be magnified in the subsequent steps in [29]. In this paper, we circumvent this limitation by estimating all *environment semantics* directly from the measurement data.

The novelty and contribution are summarized as follows:

- We develop a novel radio map model that consists of a parametric sub-model that captures the geometry of the propagation environment with radio semantics and a non-parametric sub-model that captures the residual of the shadowing. With such a framework, the model performs well with both small or large amount of training data.
- We formulate a least-squares estimation problem for the radio map construction. While the problem is non-convex with degenerated gradient, we discover and prove the *partial quasiconvexity* of the problem; based on this theoretical result, we develop an efficient algorithm to construct the radio map as well as the geometry of the virtual environment.
- We conduct numerical experiments to verify that the proposed approaches significantly outperform KNN and Kriging for radio map construction using the city data of Shanghai. With the reconstruction of the geometry of the virtual environment, we also demonstrate the performance advantage of applying the proposed radio map model to UAV-assisted wireless communication.

The rest of the paper is organized as follows. In Section II, the multi-degree channel and multi-class virtual obstacle model are established. Section III and Section IV develop the algorithm and establish theoretical results. Applications and their numerical results are demonstrated in Section V. Conclusions are drawn in Section VI.

II. RADIO MAP MODEL

Consider wireless communications between a ground user and a low altitude aerial node over a dense urban environment. Typically, the aerial node can be a relay base station (BS)

carried by a UAV or a sensing device installed on a high tower or on the rooftop. The focus of this paper is to construct a radio map to characterize the channel gain between every terrestrial (ground) user position $\mathbf{p}_u$ and every aerial node (drone) position $\mathbf{p}_d$ pair based on a limited number of measurement samples, where $\mathbf{p}_u, \mathbf{p}_d \in \mathbb{R}^3$.

A. Radio Map With Environment Semantics

Denote the *communication link* $\mathbf{p} = (\mathbf{p}_u, \mathbf{p}_d) \in \mathbb{R}^6$ as the positions of the ground user and aerial node pair. The signal that propagates between $\mathbf{p}_u$ and $\mathbf{p}_d$ consists of multiple paths which experience penetration, reflection, diffraction, and scattering according to the specific environment. We aim at building a radio map $g(\mathbf{p})$ to capture the large-scale effect of the channel gain, including the path loss and the shadowing for each link characterized by the 6-dimensional location $\mathbf{p}$.

Recall a classical channel model $\sum_{k=0}^1 (\beta_k + \alpha_k \log_{10} \|\mathbf{p}_u - \mathbf{p}_d\|_2) \mathbb{I}\{\mathbf{p} \in \mathcal{D}_k\} + \xi$, which is based on the 0-1 categorization of links being in the LOS region $\mathbf{p} \in \mathcal{D}_0$ or in NLOS region $\mathbf{p} \in \mathcal{D}_1$; and ξ is a random variable for the shadowing. We extend such a classical model to a $(K+1)$ -degree model with soft categorization. Specifically, the proposed radio map model $g(\mathbf{p})$ consists of a *deterministic radio map* $\bar{g}(\mathbf{p}; \boldsymbol{\theta}, \mathbf{H})$ and a *residual shadowing map* $\xi(\mathbf{p})$:

$$g(\mathbf{p}) = \bar{g}(\mathbf{p}; \boldsymbol{\theta}, \mathbf{H}) + \xi(\mathbf{p}) \quad (1)$$

where the deterministic radio map

$$\bar{g}(\mathbf{p}; \boldsymbol{\theta}, \mathbf{H}) \triangleq \sum_{k=0}^K (\beta_k + \alpha_k \log_{10} \|\mathbf{p}_u - \mathbf{p}_d\|_2) S_k(\mathbf{p}; \mathbf{H}) \quad (2)$$

is parameterized by $\boldsymbol{\theta} = \{\alpha_k, \beta_k\}_{k=0}^K$ for each path loss sub-model $\beta_k + \alpha_k \log_{10} \|\mathbf{p}_u - \mathbf{p}_d\|_2$ under different degrees of signal obstruction; the term $S_k(\mathbf{p}; \mathbf{H})$ models the likelihood that link $\mathbf{p}$ experiences in the k th degree of signal obstruction with parameter $\mathbf{H}$, to be explained in the next subsection, to capture the semantic information of the propagation environment. The component $\xi(\mathbf{p})$ is a random process assumed with zero mean and bounded variance.

The advantage of the proposed model (1) is as follows: First, as suggested by measurement data that there are rarely sharp edges between LOS and NLOS, $S_k(\mathbf{p}; \mathbf{H})$ approximates the reality better than an indicator function does in a classical model. Second, the geometry of the radio propagation environment will be explicitly built into the model $\bar{g}(\mathbf{p}; \boldsymbol{\theta}, \mathbf{H})$ to assist radio map reconstruction, as to be shown later.

B. Virtual Obstacle Model

We propose to impose a *multi-class virtual obstacle model* for $S_k(\mathbf{p}; \mathbf{H})$, the likelihood of $\mathbf{p}$ being in the k th propagation region $\mathcal{D}_k$ to describe the environment semantics. The general idea is to employ an *equivalent* virtual obstacle at a certain *location*, with appropriate *height* and *type* to intersect with the direct path $\mathbf{p}$ to represent the likelihood $S_k(\mathbf{p}; \mathbf{H})$. Hence, the virtual obstacle may not be mapped to a building in the reality, but serves as a geometry representation of the radio propagation environment.

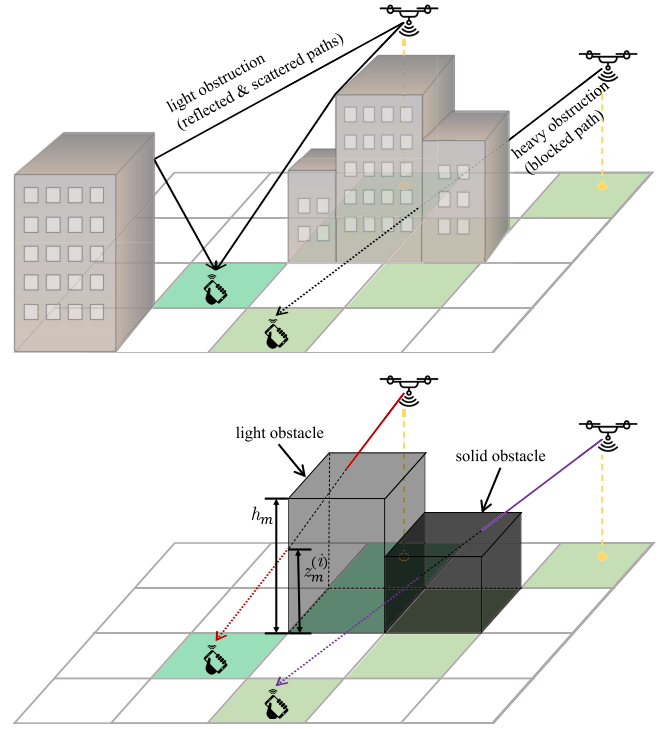


Fig. 1. Left: the multi-path propagation in reality; right: a virtual obstacle structure that captures the semantic information of the propagation environment.

For example, as illustrated in Fig. 1, if a link $\mathbf{p}$ is in deep shadow, then we place a solid virtual obstacle to intersect the direct path of $\mathbf{p}$; on the other hand, if $\mathbf{p}$ is in light shadow, then some light virtual obstacle is in place to intersect the direct path.

Specifically, we first partition the target ground area into M grid cells. Then, a virtual obstacle map can be represented by an $M \times K$ matrix $\mathbf{H}$, where the k th column $\mathbf{h}_k$ of $\mathbf{H}$ represents the height of class- k virtual obstacle, and the m th entry of $\mathbf{h}_k$ represents the height of the virtual obstacle located at the m th grid cell. As such, for each of the m th grid cell, there are K classes of virtual obstacles, each with height $h_{m,k}$, for $k = 1, 2, \dots, K$. With such a notation, the likelihood $S_k(\mathbf{p}; \mathbf{H})$ of link $\mathbf{p}$ belonging to propagation region $\mathcal{D}_k$ can be constructed from the virtual obstacle map $\mathbf{H}$ as follows.

1) *Propagation Regions With Hard Boundary*: It holds that $\sum_k S_k(\mathbf{p}; \mathbf{H}) = 1$ and $S_k(\mathbf{p}; \mathbf{H}) \in \{0, 1\}$, and constructed based on the following rule:

- $\mathbf{p} \in \mathcal{D}_0$, i.e., $S_0(\mathbf{p}; \mathbf{H}) = 1$, if there is no obstacle that intersects with the direct path between $\mathbf{p}_u$ and $\mathbf{p}_d$.
- $\mathbf{p} \in \mathcal{D}_k$, i.e., $S_k(\mathbf{p}; \mathbf{H}) = 1$ for some $k \geq 1$, if a class- k obstacle intersects with the direct path between $\mathbf{p}_u$ and $\mathbf{p}_d$, while no class- l , $l > k$, obstacle intersecting.

Mathematically, denote $\mathcal{B}^{(i)}$ as the set of grid cells that are covered by the direct path between $\mathbf{p}_u^{(i)}$ and $\mathbf{p}_d^{(i)}$. Specifically, if one projects the path $(\mathbf{p}_u^{(i)}, \mathbf{p}_d^{(i)})$ onto the ground, then the projected path passes through the grid cells and only the grid cells in $\mathcal{B}^{(i)}$. For each grid cell $m \in \mathcal{B}^{(i)}$, denote $z_m^{(i)}$ as the altitude when the path $(\mathbf{p}_u^{(i)}, \mathbf{p}_d^{(i)})$ passes over the grid cell.

It follows from the first rule above that $\mathbf{p}^{(i)} \in \mathcal{D}_0$ if $h_{m,k} < z_m^{(i)}$ for all $m \in \mathcal{B}^{(i)}$ and all $0 < k \leq K$, i.e., for all the

relevant grid locations $m \in \mathcal{B}^{(i)}$, all classes of obstacles are below the corresponding critical altitude $z_m^{(i)}$; mathematically, $\mathbb{I}\{\mathbf{p}^{(i)} \in \mathcal{D}_0(\mathbf{H})\} = \mathbb{I}\{h_{m,l} < z_m^{(i)}, \forall m \in \mathcal{B}^{(i)}, \forall l > 0\}$. From the second rule above, if $\mathbf{p}^{(i)} \in \mathcal{D}_k$ for $k > 0$, we must have $h_{m,k} \geq z_m^{(i)}$ for some class- k obstacle at the m th grid cell, mathematically, $\mathbb{I}\{h_{m,k} \geq z_m^{(i)}, \exists m \in \mathcal{B}^{(i)}\} = 1$, and at the same time, we should also have $h_{m,l} < z_m^{(i)}$ for all $m \in \mathcal{B}^{(i)}$ and all $l \geq k+1$, mathematically, $\mathbb{I}\{h_{m,l} < z_m^{(i)}, \forall m \in \mathcal{B}^{(i)}, \forall l > k\} = 1$. To summarize,

$$\mathbb{I}\{\mathbf{p}^{(i)} \in \mathcal{D}_k(\mathbf{H})\} = \begin{cases} \prod_{m \in \mathcal{B}^{(i)}} \prod_{l > 0} \mathbb{I}\{h_{m,l} < z_m^{(i)}\}, & \text{if } k = 0, \\ (1 - \prod_{m \in \mathcal{B}^{(i)}} (1 - \mathbb{I}\{h_{m,k} \geq z_m^{(i)}\})) \times \prod_{m \in \mathcal{B}^{(i)}} \prod_{l > k} \mathbb{I}\{h_{m,l} < z_m^{(i)}\}, & \text{if } k \geq 1. \end{cases} \quad (3)$$

As a useful property, the mathematical model (3) implies that it suffices to focus on the parameter regime $h_{m,1} \geq h_{m,2} \geq \dots \geq h_{m,K}$. This is because, for instance, if a set of parameters $\mathbf{H}'$ differ from $\mathbf{H}$ by $h_{m,k} < h_{m,k+1}$ for one or a few grid cells m and a few virtual obstacle classes k , then the two parameters lead to the same propagation region $\mathcal{D}_k(\mathbf{H}) = \mathcal{D}_k(\mathbf{H}')$ for all k according to (3).

Using the propagation regions defined in (3), the likelihood function $S_k(\mathbf{p}; \mathbf{H})$ in (1) can be chosen as $S_k(\mathbf{p}; \mathbf{H}) \triangleq \mathbb{I}\{\mathbf{p} \in \mathcal{D}_k(\mathbf{H})\}$.

2) *Propagation Regions With Soft Boundary*: The propagation region $\mathcal{D}_k$ modeled in (3) have hard boundaries. In practice, there could be smooth transition from LOS to NLOS regions due to the possible signal diffraction at the edge of the obstacle. To allow the likelihood $S_k(\mathbf{p}; \mathbf{H})$ to take fractional numbers in $[0, 1]$, we extend the hard boundary model by applying a spatial low-pass filter on the propagation regions $\mathcal{D}_k$ defined in (3). Specifically, given a link $\mathbf{p}$, we evaluate a set of neighbor positions with offset ϵ_j from $\mathbf{p}$. By averaging $\mathbb{I}\{\mathbf{p} + \epsilon_j \in \mathcal{D}_k(\mathbf{H})\}$ with weights ω_j , one can obtain the likelihood $S_k(\mathbf{p}; \mathbf{H})$. A common choice of filter coefficients ω_j can be obtained as a function of the distance $\|\epsilon_j\|_2$ from $\mathbf{p}$. In this paper, we consider a spatial filter that consists of a set of J uniform grid points $\epsilon_j, j = 0, 1, \dots, J-1$, in 6D space centered at the origin with ϵ_0 chosen as $\epsilon_0 = \mathbf{0}$, and the weights are chosen as $\omega_j = c \cdot \exp(-\|\epsilon_j\|_2^2 / \sigma_\omega^2)$, where σ_ω is a parameter, and c is a normalization factor such that $\sum_{j=0}^{J-1} \omega_j = 1$.¹ Thus, the likelihood function $S_k(\mathbf{p}; \mathbf{H})$ parameterized by the virtual obstacle map $\mathbf{H}$ is defined as

$$S_k(\mathbf{p}; \mathbf{H}) = \sum_{j=0}^{J-1} \omega_j \mathbb{I}\{\mathbf{p} + \epsilon_j \in \mathcal{D}_k(\mathbf{H})\} \quad (4)$$

which satisfies $\sum_{k=0}^K S_k(\mathbf{p}; \mathbf{H}) = 1$.

¹In general, the weights can be designed using a kernel function, where the smaller $\|\epsilon_j\|_2$, the larger the weight. The choice of the kernel and its parameters can be determined based on signal parameters, such as frequency and bandwidth, and via a cross-validation approach.

In this work, we consider that the smoothing parameters ϵ_j and ω_j are preconfigured and fixed. Therefore, the soft boundary model (3) has the same set of parameters, $\mathbf{H}$, to be estimated as the hard boundary model (3).

III. RADIO MAP CONSTRUCTION VIA ENVIRONMENT MAPPING

In this section, we jointly estimate the propagation parameter θ and the virtual obstacle map $\mathbf{H}$ for constructing the deterministic radio map $\bar{g}(\mathbf{p}; \theta, \mathbf{H})$ in (2).

A. Formulation of the Radio Map Learning Problem

Consider taking measurements at transmit and receive location pairs $\{\mathbf{p}^{(i)}\}$ and recall $\mathbf{p} = (\mathbf{p}_u, \mathbf{p}_d)$. Based on (1), the measured RSS can be written as

$$y^{(i)} = \bar{g}(\mathbf{p}^{(i)}; \theta, \mathbf{H}) + n^{(i)} \quad (5)$$

where $n^{(i)} = \xi(\mathbf{p}^{(i)}) + \tilde{n}^{(i)}$ captures both the random component $\xi(\mathbf{p})$ in (1) for the residual shadowing and the measurement noise $\tilde{n}^{(i)}$ which is assumed as independent and identically distributed with zero mean, variance σ_n^2 , and finite fourth-order moment.

The goal of this section is to estimate parameters θ and $\mathbf{H}$ from the set of noisy measurement data $\{(\mathbf{p}^{(i)}, y^{(i)})\}_{i=1}^N$ obtained from (5). A least-squares problem can be formulated as

$$\underset{\theta, \mathbf{H} \geq \mathbf{0}}{\text{minimize}} \quad f(\theta, \mathbf{H}) \triangleq \frac{1}{N} \sum_{i=1}^N \left[y^{(i)} - \sum_{k=0}^K (\beta_k + \alpha_k d(\mathbf{p}^{(i)})) S_k(\mathbf{p}^{(i)}; \mathbf{H}) \right]^2 \quad (6)$$

where $d(\mathbf{p}^{(i)}) \triangleq \log_{10} \|\mathbf{p}_u^{(i)} - \mathbf{p}_d^{(i)}\|_2$ is the log-distance between $\mathbf{p}_u^{(i)}$ and $\mathbf{p}_d^{(i)}$ for the i th measurement.

Note that the least-squares problem (6) is difficult to solve using a standard solver. This is because the problem is *non-convex* in the joint variable $(\theta, \mathbf{H})$ and the objective function $f(\theta, \mathbf{H})$ is *discontinuous* due to the indicator functions used in (2)–(4). To circumvent these difficulties, we will exploit the property discovered in $f(\theta, \mathbf{H})$.

B. Asymptotic Consistency of Radio Maps

Since the objective function $f(\theta, \mathbf{H})$ in (6) contains randomness due to the measurement noise, we first find a *deterministic proxy* for $f(\theta, \mathbf{H})$ under large N . Denote θ^* and $\mathbf{H}^*$ as the true parameters in (5) and consider the following deterministic proxy function

$$\bar{f}(\theta, \mathbf{H}) \triangleq \frac{1}{N} \sum_{i=1}^N \left[\bar{g}(\mathbf{p}^{(i)}; \theta, \mathbf{H}) - \bar{g}(\mathbf{p}^{(i)}; \theta^*, \mathbf{H}^*) \right]^2 \quad (7)$$

and it is clear that $\bar{f}(\theta^*, \mathbf{H}^*) = 0$.

Lemma 1 (Deterministic Equivalence): Suppose that the random component $\xi(\mathbf{p}^{(i)})$ in the measurement model (5) is weakly dependent, i.e., the covariance satisfies $\text{cov}(\xi(\mathbf{p}^{(i)}), \xi(\mathbf{p}^{(j)})) \rightarrow 0$, as $|i - j| \rightarrow \infty$ and $N \rightarrow \infty$, and moreover, the limit $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \xi(\mathbf{p}^{(i)})^2$ exists and is

finite.² Then, there exists a finite constant $0 < C < \infty$, such that

$$f(\boldsymbol{\theta}, \mathbf{H}) \rightarrow \bar{f}(\boldsymbol{\theta}, \mathbf{H}) + C$$

in probability, for every $(\boldsymbol{\theta}, \mathbf{H})$, as $N \rightarrow \infty$.

Proof: See Appendix A. $\square$

It follows from the above lemma that the parameters $(\boldsymbol{\theta}, \mathbf{H})$ that minimize the deterministic proxy $\bar{f}(\boldsymbol{\theta}, \mathbf{H})$ also minimize the least-squares cost $f(\boldsymbol{\theta}, \mathbf{H})$ asymptotically. As a result, analyzing the property of the deterministic proxy $\bar{f}(\boldsymbol{\theta}, \mathbf{H})$ may inspire efficient algorithms to solve (6) under large N .

The following corollary shows that one may obtain *asymptotically consistent* radio maps even the globally optimal solution $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ differs from the true parameter $(\boldsymbol{\theta}^*, \mathbf{H}^*)$; here, consistency means that radio maps are identical $\bar{g}(\mathbf{p}^{(i)}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}}) = \bar{g}(\mathbf{p}^{(i)}; \boldsymbol{\theta}^*, \mathbf{H}^*)$ at the measurement locations $\{\mathbf{p}^{(i)}\}$.

Corollary 1 (Asymptotic Consistency of Radio Maps): It holds that, as $N \rightarrow \infty$,

$$\bar{f}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{H}}) = \frac{1}{N} \sum_{i=1}^N \left(\bar{g}(\mathbf{p}^{(i)}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}}) - \bar{g}(\mathbf{p}^{(i)}; \boldsymbol{\theta}^*, \mathbf{H}^*) \right)^2 \rightarrow 0$$

where $(\hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ is the globally optimal solution to (6).

Proof: By definition, the minimum value of $\bar{f}(\boldsymbol{\theta}, \mathbf{H})$ is obtained as $\bar{f}(\boldsymbol{\theta}^*, \mathbf{H}^*) = 0$, and therefore, $\bar{f}(\boldsymbol{\theta}, \mathbf{H}) + C$ can be globally minimized to C . Since $f(\boldsymbol{\theta}, \mathbf{H}) \rightarrow \bar{f}(\boldsymbol{\theta}, \mathbf{H}) + C$ from Lemma 1, $f(\boldsymbol{\theta}, \mathbf{H})$ is also asymptotically and globally minimized to $f(\hat{\boldsymbol{\theta}}, \hat{\mathbf{H}}) \rightarrow C$, which implies that $\bar{f}(\hat{\boldsymbol{\theta}}, \hat{\mathbf{H}}) \rightarrow 0$ as $N \rightarrow \infty$. $\square$

C. Solution to the Propagation Parameter $\boldsymbol{\theta}$

It can be easily verified that given the variable $\mathbf{H}$, the problem (6) is convex in $\boldsymbol{\theta}$. To see this, denote $\mathbf{X} \in \mathbb{R}^{N \times (2K+2)}$ as log-distance data matrix, where the even elements in the i th row of $\mathbf{X}$ equal to $d(\mathbf{p}^{(i)})$ and the odd elements in the i th row of $\mathbf{X}$ equal to 1. Arrange the elements in the variable $\boldsymbol{\theta} \in \mathbb{R}^{2K+2}$ as $\boldsymbol{\theta} = [\alpha_0 \ \beta_0 \ \alpha_1 \ \beta_1 \ \cdots \ \alpha_K \ \beta_K]^T$ as the path loss parameter vector for the sub-models. Stack the measurement value $y^{(i)}$ into a vector $\mathbf{y} = [y^{(1)} \ y^{(2)} \ \cdots \ y^{(N)}]^T \in \mathbb{R}^N$. Finally, denote $\mathbf{S} \in \mathbb{R}^{N \times (2K+2)}$ as the likelihood matrix, where $[\mathbf{S}]_{i,2k} = [\mathbf{S}]_{i,2k+1} = S_k(\mathbf{p}^{(i)}; \mathbf{H})$. Then, for a fixed $\mathbf{H}$, problem (6) can be written as

$$\underset{\boldsymbol{\theta}}{\text{minimize}} \ \|(\mathbf{S} \circ \mathbf{X})\boldsymbol{\theta} - \mathbf{y}\|_2^2 \quad (8)$$

where $\circ$ is the Hadamard product, i.e., $[\mathbf{S} \circ \mathbf{X}]_{ij} = [\mathbf{S}]_{ij}[\mathbf{X}]_{ij}$. Problem (8) is unconstrained quadratic programming, and it is convex with respect to (w.r.t.) $\boldsymbol{\theta}$. It can be solved by setting the derivative to zero, and the solution is given by

$$\hat{\boldsymbol{\theta}} = ((\mathbf{S} \circ \mathbf{X})^T (\mathbf{S} \circ \mathbf{X}))^{-1} (\mathbf{S} \circ \mathbf{X})^T \mathbf{y}. \quad (9)$$

Proposition 1: Under $\mathbf{H}^*$, the solution in (9) is an unbiased estimator of $\boldsymbol{\theta}^*$, i.e., $\mathbb{E}\{\hat{\boldsymbol{\theta}}\} = \boldsymbol{\theta}^*$.

²If the process $\xi(\mathbf{p})$ is segment-wise second-order stationary within each propagation segment, then the limit exists if one samples each propagation segment with a fixed probability, for example, under uniform sampling over the entire area.

Proof: With $\mathbf{H}^*$, the observation model can be written as $\mathbf{y} = (\mathbf{S}^* \circ \mathbf{X})\boldsymbol{\theta} + \mathbf{n}$, where $\mathbf{n}$ is a vector stacking n . Therefore, it is a standard least-squares estimation problem for a linear observation model $\mathbf{y}$ with zero mean noise $\mathbf{n}$. It is well-known that the least-squares estimator (9) in this case is unbiased [31]. $\square$

D. Quasiconvexity in the Environment Parameter $\mathbf{H}$

While problem (6) is convex in $\boldsymbol{\theta}$ by fixing $\mathbf{H}$, it is still non-convex in $\mathbf{H}$ by fixing $\boldsymbol{\theta}$. However, we discover that $\mathbf{H}$ is partially quasiconvex, which can be later exploited for efficient algorithm design.

First, consider the $K = 1$ case, where there are two propagation regions, LOS and NLOS, and the matrix $\mathbf{H}$ degenerates to a column vector $\mathbf{h}$.

Theorem 1 (Quasiconvexity for $K = 1$ Under Soft Boundary): Suppose that the filter coefficient ω_0 in (4) satisfies $\omega_0 \geq \frac{2}{3}$. Given a vector $\mathbf{h}' \succeq \mathbf{h}^*$ and an index m , consider the interval $\mathcal{I}_m \triangleq \{\mathbf{h} \in \mathbb{R}^M : 0 \leq h_m \leq H_{\max}, h_j = h'_j, \forall j \neq m\}$. Then, $\bar{f}(\boldsymbol{\theta}^*, \mathbf{h})$ in (7) is quasiconvex over the interval $\mathcal{I}_m$.

Proof: See Appendix B. $\square$

The above result implies that given a variable $\mathbf{h} \succeq \mathbf{h}^*$, the function $\bar{f}(\boldsymbol{\theta}^*, \mathbf{h})$ is partially quasiconvex w.r.t. to each entry h_m with all the other entries $h_j, j \neq m$, fixed. As a result of the partial quasiconvexity, there exists $\hat{h}_m$, such that, for $h_m < \hat{h}_m$, $\bar{f}$ is non-increasing in h_m , and for $h_m > \hat{h}_m$, $\bar{f}$ is non-decreasing.

Next, consider the case of a general K and the propagation regions being modeled with hard boundaries, i.e., the likelihood function is chosen as $S_k(\mathbf{p}; \mathbf{H}) = \mathbb{I}\{\mathbf{p} \in \mathcal{D}_k(\mathbf{H})\}$ as in (3). The following result shows that the partial quasiconvexity in Theorem 1 also holds.

Theorem 2 (Quasiconvexity Under Hard Boundary): Given a matrix $\mathbf{H}' \succeq \mathbf{H}^*$ and an index (m, k) , define the interval $\mathcal{I}_{m,k} \triangleq \{\mathbf{H} \in \mathbb{R}^{M \times K} : 0 \leq h_{m,k} \leq H_{\max}, h_{j,l} = h'_{j,l}, \forall (j, l) \neq (m, k)\}$. Then, $\bar{f}(\boldsymbol{\theta}^*, \mathbf{H})$ is quasiconvex over the interval $\mathcal{I}_{m,k}$.

Proof: See Appendix C. $\square$

Theorems 1 and 2 imply that if we focus on each individual entry $h_{m,k}$ in the variable $\mathbf{H}$, then $\bar{f}(\boldsymbol{\theta}, \mathbf{H})$ first decreases and then increases. More specifically, according to the fact that $\bar{f}(\boldsymbol{\theta}, \mathbf{H})$ in (7) is constructed from a number of indicator functions in (3) and (4), $\bar{f}(\boldsymbol{\theta}, \mathbf{H})$ appears like a *staircase* function that first steps down along the interval $\mathcal{I}_{m,k}$, reaching the bottom around $h_{m,k}^*$, and then steps up.

To characterize the basin of $\bar{f}$ over $\mathcal{I}_{m,k}$, define a function

$$\bar{f}_{m,k}(h_{m,k}; \boldsymbol{\theta}, \mathbf{H}_{m,k}^-) = \bar{f}(\boldsymbol{\theta}, \mathbf{H}) \quad (10)$$

of the scalar variable $h_{m,k}$ with the other variables $\boldsymbol{\theta}$ and $\mathbf{H}_{m,k}^-$ held fixed, where $\mathbf{H}_{m,k}^- = \{h_{j,l} : \forall (j, l) \neq (m, k)\}$ is a collection of entries from the matrix $\mathbf{H}$ except the (m, k) th one. The basin is defined as the interval $\mathcal{I}_{m,k}(\boldsymbol{\theta}, \mathbf{H}) \triangleq \{z : \bar{f}_{m,k}(z; \boldsymbol{\theta}, \mathbf{H}_{m,k}^-) \leq \bar{f}_{m,k}(h; \boldsymbol{\theta}, \mathbf{H}_{m,k}^-), \forall 0 \leq h \leq H_{\max}\}$. We are interested in the largest value in the basin

$$\hat{h}_{m,k}(\boldsymbol{\theta}, \mathbf{H}) \triangleq \sup\{\mathcal{I}_{m,k}(\boldsymbol{\theta}, \mathbf{H})\} \quad (11)$$

where the algorithm to solve (11) will be developed in Section III-E. With the notation of $\hat{h}_{m,k}(\theta, \mathbf{H})$, the following property can be established.

Theorem 3 (Consistency): Suppose $\mathbf{H} \succeq \mathbf{H}^*$ and consider the interval $\mathcal{I}_{m,k}$ as defined in Theorem 2. Then, $\hat{h}_{m,k}(\theta^*, \mathbf{H}) \geq h_{m,k}^*$. Moreover, given $\mathbf{H}'' \succeq \mathbf{H}' \succeq \mathbf{H}^*$, it holds that $\hat{h}_{m,k}(\theta^*, \mathbf{H}'') \geq \hat{h}_{m,k}(\theta^*, \mathbf{H}') \geq h_{m,k}^*$.

Proof: See Appendix D. $\square$

As inspired from Theorems 1–3, when θ is sufficiently close to θ^* , an efficient algorithm to optimize $\mathbf{H}$ can proceed. First, set the initial value of $\mathbf{H}$ as $h_{m,k} = H_{\max}$ for all m, k . Then, for each element $h_{m,k}$, find $\hat{h}_{m,k}$ that minimizes $\bar{f}(\theta, \mathbf{H})$, and repeat this step until convergence. This approach is summarized in Algorithm 2 and its convergence can be analyzed as follows.

Let $\mathbf{H}(t)$ denotes the variable $\mathbf{H}$ at the t th iteration. Theorem 3 implies the convergence of $h_{m,k}(t)$ for $\theta = \theta^*$. Suppose at $t > 1$, given that $\mathbf{H}(t-1) \succeq \mathbf{H}(t) \succeq \mathbf{H}^*$, Theorem 3 implies that $\hat{h}_{m,k}(t-1) \geq \hat{h}_{m,k}(t) \geq h_{m,k}^*$, $\forall m, k$, under θ^* , and consequently, $\mathbf{H}(t) \succeq \mathbf{H}(t+1) \succeq \mathbf{H}^*$. As $t = 1$ can be checked to satisfy $\hat{h}_{m,k}(1) \geq \hat{h}_{m,k}(2) \geq h_{m,k}^*$ due to the initialization, by induction, $\hat{h}_{m,k}(t) \geq \hat{h}_{m,k}(t+1) \geq h_{m,k}^*$ is satisfied for all t , which means that Algorithm 2 constructs monotonically decreasing and lower bounded sequences $h_{m,k}(t)$.

E. Optimizing $\mathbf{H}$ via Local Polynomial Approximation

The remaining challenge is to compute $\hat{h}_{m,k}(\theta, \mathbf{H})$ in (11). A common approach finding the minimizer of a quasiconvex function over a bounded interval is to perform a bisection search for the critical point. However, as discussed after Theorem 2, the function $\bar{f}(\theta, \mathbf{H})$ appears as a staircase in each variable $h_{m,k}$, where the derivative is zero almost everywhere. Moreover, the staircase function $\bar{f}(\theta, \mathbf{H})$ is not available, but only its noisy counterpart $f(\theta, \mathbf{H})$ in (6) is accessible to the algorithm.

We propose to smooth $f(\theta, \mathbf{H})$ without losing the partial quasiconvex property of $\bar{f}(\theta, \mathbf{H})$. One possibility is to employ local polynomial approximation to estimate $\bar{f}(\theta, \mathbf{H})$ from $f(\theta, \mathbf{H})$. Specifically, we use a polynomial to approximate $\tilde{f}_{m,k}(h; \theta, \mathbf{H}_{m,k}^-)$ in (10) at the neighborhood of $h_{m,k}$:

$$\tilde{f}_{m,k}(h; \mathbf{a}, h_{m,k}) = a_0 + a_1(h - h_{m,k}) + a_2(h - h_{m,k})^2 + \dots$$

where the coefficients $\mathbf{a} = (a_0, a_1, \dots)$ are computed by sampling $f_{m,k}(h; \theta, \mathbf{H}_{m,k}^-)$ over a set $\mathcal{Z}$ of altitudes z in the interval $[0, H_{\max}]$ and minimizing the weighted squared error:

$$\sum_{z \in \mathcal{Z}} \left(f_{m,k}(z; \theta, \mathbf{H}_{m,k}^-) - \tilde{f}_{m,k}(z; \mathbf{a}, h_{m,k}) \right)^2 K_b(z - h_{m,k}). \quad (12)$$

Here, $f_{m,k}(h; \theta, \mathbf{H}_{m,k}^-)$ is a notation defined according to $f(\theta, \mathbf{H})$ in a way similar to the definition of $\tilde{f}_{m,k}(h; \theta, \mathbf{H}_{m,k}^-)$ in (10), and is computed via (6). The term $K_b(u)$ is a kernel function that assigns a high weight if the distance $u = |z - h_{m,k}|$ is small, and a low weight if the distance u is large. It was found that the Epanechnikov kernel $K_b(u) = \frac{3}{4b} \left(1 - \left(\frac{u}{b}\right)^2\right)_+$ minimizes the asymptotic approximation error

Algorithm 1 Construction of the Virtual Environment

Objective: Estimate the virtual environment $\mathbf{H}$ and the propagation parameter θ based on the measurement set $\{(\mathbf{p}^{(i)}, y^{(i)})\}$.

- 1) Initialize $\mathbf{H}(1) = \mathbf{1}H_{\max}$, $\theta(1)$ using expectation-maximization (EM) in [28], and iteration $t = 1$.
- 2) Optimize $\mathbf{H}$: For each (m, k) , update $h_{m,k}(t+1) = \hat{h}_{m,k}$ as defined in (11) based on $\theta(t)$ and $\mathbf{H}_{m,k}^-(t)$. Specifically, the following bisection search is used:
 - a) Initialize $h_{\min} = 0$ and $h_{\max} = H_{\max}$.
 - b) Set $h_{m,k} = \frac{1}{2}(h_{\min} + h_{\max})$ and find the minimizer a_1^* of (12).
 - c) If $a_1^* < 0$, then $h_{\min} \leftarrow h_{m,k}$; if $a_1^* > 0$, then $h_{\max} \leftarrow h_{m,k}$.
 - d) Repeat from Step 2b until $|h_{\max} - h_{\min}| < \epsilon$, for which output $\hat{h}_{m,k} = h_{m,k}$.
- 3) Optimize θ : Update $\theta(t+1) = \hat{\theta}$ based on $\mathbf{H}(t+1)$ according to (9). Set $t \leftarrow t+1$ and repeat from Step 2 until $\frac{1}{MK} \|\mathbf{H}(t+1) - \mathbf{H}(t)\|_F < \epsilon_0$.

of the polynomial $\tilde{f}_{m,k}(h; \mathbf{a}, h_{m,k})$ for a given window size b [32]. Here, we can adapt b according to the volume of the measurement data such that $\tilde{f}_{m,k}$ is smoothed and the gradient is non-degenerated.

It is clear that $a_1^*(h_{m,k})$, from the minimizer of (12), is the approximated (but non-degenerated) gradient of $\bar{f}$ at $h = h_{m,k}$. Note that it suffices to determine the sign of $a_1^*(h_{m,k})$ due to the quasiconvexity in Theorems 1 and 2. Specifically, we can perform a bisection search to seek the minimizer $\hat{h}_{m,k}$ in (11) as follows: (i) Initialize $h_{\min} = 0$ and $h_{\max} = H_{\max}$. (ii) Set $h_{m,k} = \frac{1}{2}(h_{\min} + h_{\max})$ and find the minimizer a_1^* from minimizing (12). (iii) If $a_1^* < 0$, then $h_{\min} \leftarrow h_{m,k}$; if $a_1^* > 0$, then $h_{\max} \leftarrow h_{m,k}$. (iv) Repeat from Step (ii) until $|h_{\max} - h_{\min}| < \epsilon$, and output $\hat{h}_{m,k} = h_{m,k}$.

An alternating optimization approach for a joint estimate of the virtual obstacle height $\mathbf{H}$ and the propagation parameter θ for constructing semantics is summarized in Algorithm 1.

IV. RECONSTRUCTING THE SHADOWING USING KRIGING

The deterministic radio map $\hat{g}(\mathbf{p}) = \bar{g}(\mathbf{p}; \hat{\mathbf{H}}, \hat{\theta})$ constructed from (2) based on the virtual environment $\hat{\mathbf{H}}$ and the propagation parameter $\hat{\theta}$ estimated from Section III only focuses on the degrees of signal blockage of the direct propagation path. To capture the residual shadowing, we propose to use Kriging [16], [17] to construct the residual shadowing map $\xi(\mathbf{p})$ in (1).

For each sample $(\mathbf{p}^{(i)}, y^{(i)})$, denote $\hat{\xi}^{(i)} = y^{(i)} - \hat{g}(\mathbf{p}^{(i)})$ as the residual shadowing at $\mathbf{p}^{(i)}$ over the reconstructed deterministic radio map $\hat{g}(\mathbf{p})$. Thus, a set of data $\mathcal{T} = \{(\mathbf{p}^{(i)}, \hat{\xi}^{(i)})\}$ is formed for constructing $\xi(\mathbf{p})$. Recall the measurement model $y^{(i)}$ in (5) and the noise model $n^{(i)} = \xi(\mathbf{p}^{(i)}) + \tilde{n}^{(i)}$, where $\tilde{n}^{(i)}$ is the measurement noise. We obtain the following model

$$\begin{aligned} \hat{\xi}^{(i)} &= \bar{g}(\mathbf{p}^{(i)}; \theta^*, \mathbf{H}^*) + \xi(\mathbf{p}^{(i)}) + \tilde{n}^{(i)} - \hat{g}(\mathbf{p}^{(i)}) \\ &= \xi(\mathbf{p}^{(i)}) + \tilde{n}^{(i)} + (\bar{g}(\mathbf{p}^{(i)}; \theta^*, \mathbf{H}^*) - \hat{g}(\mathbf{p}^{(i)})) \end{aligned}$$

$$\approx \xi(\mathbf{p}^{(i)}) + \tilde{n}^{(i)} \quad (13)$$

where the approximation is asymptotically accurate because the term $\bar{g}(\mathbf{p}^{(i)}; \boldsymbol{\theta}^*, \mathbf{H}^*) - \hat{g}(\mathbf{p}^{(i)})$ tends to 0 as $N \rightarrow \infty$ according to Corollary 1.

A. Building the Semivariogram

We first build a *semivariogram* from the dataset $\mathcal{T}$ to capture the spatial correlation of $\xi(\mathbf{p})$. Specifically, for a zero mean stationary process $\xi(\mathbf{p})$, the *semivariogram* in terms of an offset $\mathbf{u}$ is defined as $v(\mathbf{u}) = \frac{1}{2} \mathbb{E}\{(\xi(\mathbf{p} + \mathbf{u}) - \xi(\mathbf{p}))^2\}$. It follows that the relation between the semivariogram and the covariance can be characterized as $v(\mathbf{u}) = \mathbb{E}\{\xi(\mathbf{p})^2\} - \mathbb{E}\{\xi(\mathbf{p} + \mathbf{u})\xi(\mathbf{p})\}$. Note that as the residual shadowing process $\xi(\mathbf{p})$ is not available except for its noisy observations $\hat{\xi}^{(i)} = \xi(\mathbf{p}^{(i)}) + \tilde{n}^{(i)}$, the sample semivariogram results in an additive noise offset from the semivariogram $v(\mathbf{u})$ as follows:

$$\begin{aligned} \frac{1}{2} \mathbb{E}\{(\hat{\xi}^{(i)} - \hat{\xi}^{(j)})^2\} &= \frac{1}{2} \mathbb{E}\{(\xi(\mathbf{p}^{(i)}) - \xi(\mathbf{p}^{(j)}))^2\} \\ &\quad + \frac{1}{2} (\mathbb{E}\{(\tilde{n}^{(i)})^2\} + \mathbb{E}\{(\tilde{n}^{(j)})^2\}) \\ &= v(\mathbf{p}^{(i)} - \mathbf{p}^{(j)}) + \sigma_n^2. \end{aligned} \quad (14)$$

A parametric model is usually employed to fit the semivariogram from the data. We compare some common semivariogram models using the datasets described in Section V, with results reported in Fig. 5. It is observed that the exponential model yields the best fit to the data among the spherical model and Gaussian model. In general, the best semivariogram model can be selected from cross-validation.

We adopt the exponential model as an illustrated example:

$$\bar{v}(\mathbf{u}; \boldsymbol{\alpha}) = \alpha_s^2 \left(1 - \exp\left(-\frac{\|\mathbf{u}\|_2}{\alpha_r}\right)\right) \quad (15)$$

where $\boldsymbol{\alpha} = (\alpha_s, \alpha_r)$ is the model parameter to be estimated from the data in $\mathcal{T}$. The parameters of the semivariogram model can be obtained as the solution to the least-squares problem:

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \sum_{i=1}^N \sum_{j=1}^N \left[\bar{v}(\mathbf{p}^{(i)} - \mathbf{p}^{(j)}; \boldsymbol{\alpha}) \right. \\ & \left. + \sigma_n^2 - \frac{1}{2} (\hat{\xi}^{(i)} - \hat{\xi}^{(j)})^2 \right]^2. \end{aligned} \quad (16)$$

B. Constructing the Residual Shadowing

Consider $\mathcal{T}_1 \subseteq \mathcal{T}$ as a subset of data with $N_1 = |\mathcal{T}_1|$ elements that is used in the construction phase. Typically, $N_1 < N = |\mathcal{T}|$ for the consideration of complexity. The residual shadowing $\hat{\xi}(\mathbf{p})$ at each location $\mathbf{p}$ can be constructed from the data from $\mathcal{T}_1 = \{(\mathbf{p}^{(i)}, \hat{\xi}^{(i)})\}$ as

$$\hat{\xi}(\mathbf{p}) = \sum_{i=1}^{N_1} \lambda_i(\mathbf{p}) \hat{\xi}^{(i)} \quad (17)$$

where the set of coefficients $\{\lambda_i(\mathbf{p})\}$ can be determined by minimizing the variance

$$\underset{\boldsymbol{\lambda}}{\text{minimize}} \quad \mathbb{V}\{\hat{\xi}(\mathbf{p}) - \xi(\mathbf{p})\} \quad (18)$$

in which, $\mathbb{V}\{\cdot\}$ is an operator of calculating the variance of the estimation error. The constraint $\mathbf{1}^T \boldsymbol{\lambda} = 1$ is to guarantee that the estimator (17) is unbiased. Substituting (13) and (17) into (18),

$$\begin{aligned} & \mathbb{V}\{\hat{\xi}(\mathbf{p}) - \xi(\mathbf{p})\} \\ &= \mathbb{V}\left\{\sum_{i=1}^{N_1} \lambda_i \hat{\xi}^{(i)} - \xi(\mathbf{p})\right\} \\ &= \mathbb{V}\left\{\sum_{i=1}^{N_1} \lambda_i \xi(\mathbf{p}^{(i)}) + \sum_{i=1}^{N_1} \lambda_i \tilde{n}^{(i)} - \xi(\mathbf{p})\right\} \\ &= \mathbb{E}\left\{\left(\sum_{i=1}^{N_1} \lambda_i \xi(\mathbf{p}^{(i)}) - \xi(\mathbf{p})\right)^2\right\} + \sum_{i=1}^{N_1} \lambda_i^2 \sigma_n^2 \end{aligned}$$

where the third equality is due to the fact that, first, the measurement noise $\tilde{n}^{(i)}$ is with zero mean and independent of the residual shadowing process $\xi(\mathbf{p})$, and second, $\mathbb{E}\left\{\sum_{i=1}^{N_1} \lambda_i \xi(\mathbf{p}^{(i)}) - \xi(\mathbf{p})\right\} = 0$ since the residual shadowing process $\xi(\mathbf{p})$ is assumed to have zero mean.

The Lagrangian method can be employed to solve (18), with the Lagrangian function

$$\begin{aligned} L(\boldsymbol{\lambda}, \mu) &= \mathbb{V}\{\hat{\xi}(\mathbf{p}) - \xi(\mathbf{p})\} + 2\mu \left(\sum_{i=1}^{N_1} \lambda_i - 1\right) \\ &= \mathbb{E}\left\{\left(\sum_{i=1}^{N_1} \lambda_i \xi(\mathbf{p}^{(i)}) - \xi(\mathbf{p})\right)^2\right\} + \sum_{i=1}^{N_1} \lambda_i^2 \sigma_n^2 + 2\mu \left(\sum_{i=1}^{N_1} \lambda_i - 1\right) \\ &= \sum_{i=1}^{N_1} \lambda_i \mathbb{E}\left\{(\xi(\mathbf{p}^{(i)}) - \xi(\mathbf{p}))^2\right\} - \frac{1}{2} \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} \lambda_i \lambda_j \\ &\quad \times \mathbb{E}\left\{(\xi(\mathbf{p}^{(i)}) - \xi(\mathbf{p}^{(j)}))^2\right\} + \sum_{i=1}^{N_1} \lambda_i^2 \sigma_n^2 + 2\mu \left(\sum_{i=1}^{N_1} \lambda_i - 1\right) \\ &= \sum_{i=1}^{N_1} 2\lambda_i v(\mathbf{p}^{(i)} - \mathbf{p}) - \sum_{i=1}^{N_1} \sum_{j=1}^{N_1} \lambda_i \lambda_j v(\mathbf{p}^{(i)} - \mathbf{p}^{(j)}) \\ &\quad + \sum_{i=1}^{N_1} \lambda_i^2 \sigma_n^2 + 2\mu \left(\sum_{i=1}^{N_1} \lambda_i - 1\right) \end{aligned}$$

where μ is the Lagrangian multiplier.

The first-order optimality condition $\nabla_{\boldsymbol{\lambda}} L = \mathbf{0}$ and the constraint $\mathbf{1}^T \boldsymbol{\lambda} = 1$ lead to the following linear equation:

$$\begin{bmatrix} \boldsymbol{\Gamma} - \sigma_n^2 \mathbf{I} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ -\mu \end{bmatrix} = \begin{bmatrix} \boldsymbol{\gamma} \\ 1 \end{bmatrix} \quad (19)$$

where $\boldsymbol{\Gamma}$ is a matrix whose (i, j) th element is given by $\Gamma_{ij} = v(\mathbf{p}^{(i)} - \mathbf{p}^{(j)})$, which can be obtained from the semivariogram $\bar{v}(\mathbf{p}^{(i)} - \mathbf{p}^{(j)})$ constructed in Section IV-A, and $\boldsymbol{\gamma}$ is a vector with the i th element given by $\gamma_i = v(\mathbf{p}^{(i)} - \mathbf{p})$, which can be obtained from $\bar{v}(\mathbf{p}^{(i)} - \mathbf{p})$.

Solving (19) for $\boldsymbol{\lambda}$ leads to the Kriging estimator $\hat{\xi}(\mathbf{p}) = \sum_{i=1}^{N_1} \lambda_i \hat{\xi}^{(i)}$ for the construction of the residual shadowing.

Algorithm 2 Radio Map Construction Based on Virtual Environment With Kriging

Objective: Construct radio maps based on $\mathbf{H}$ and $\boldsymbol{\theta}$, and optionally, a subset of data $\mathcal{T}_1$ for calibration.

Learning Phase:

- 1) Estimate the virtual environment $\hat{\mathbf{H}}$ and the propagation parameter $\hat{\boldsymbol{\theta}}$ via Algorithm 1.
- 2) Construct a dataset from $\mathcal{T} = \{(\mathbf{p}^{(i)}, \hat{\xi}^{(i)})\}$, where $\hat{\xi}^{(i)} = y^{(i)} - \bar{g}(\mathbf{p}^{(i)}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ is the residual shadowing.
- 3) Build a semivariogram $\bar{v}(\mathbf{u})$ according to (16).

Construction Phase:

- 1) Construct a deterministic radio map $\bar{g}(\mathbf{p}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ from the virtual environment using (2).
 - 2) If $\mathcal{T}_1 \subseteq \mathcal{T}$ is available, construct the residual shadowing $\hat{\xi}(\mathbf{p}) = \sum_{i=1}^{N_1} \lambda_i(\mathbf{p}) \hat{\xi}^{(i)}$ with $\lambda_i(\mathbf{p})$ given by the solution of (19); otherwise, $\hat{\xi}(\mathbf{p}) = 0$.
 - 3) The radio map is constructed as $\hat{g}(\mathbf{p}) = \bar{g}(\mathbf{p}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}}) + \hat{\xi}(\mathbf{p})$.
-

C. The Radio Map Construction Algorithm

Based on the virtual environment $\hat{\mathbf{H}}$ and the propagation parameter $\hat{\boldsymbol{\theta}}$ estimated from Section III, a deterministic radio map $\bar{g}(\mathbf{p}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ can be constructed using (2).

In addition, if a subset of data $\mathcal{T}_1$ is available in the construction phase for calibration, then a radio map exploiting virtual environment with Kriging can be constructed. Specifically, based on the offset $\hat{\xi}^{(i)} = y^{(i)} - \bar{g}(\mathbf{p}^{(i)}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ for the various locations in the calibration dataset $\mathcal{T}_1$, the residual shadowing $\hat{\xi}(\mathbf{p})$ can be constructed using Kriging. Finally, the radio map based on the virtual environment with Kriging is given by $\hat{g}(\mathbf{p}) = \bar{g}(\mathbf{p}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}}) + \hat{\xi}(\mathbf{p})$. The overall algorithm is summarized in Algorithm 2.

D. Computational Complexity

The complexity of the proposed radio map construction in Algorithm 2 consists of the part of evaluating the deterministic radio map $\bar{g}(\mathbf{p}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ in (2) and the part of constructing the shadowing component $\hat{\xi}(\mathbf{p})$ in (17). First, the evaluation of $\bar{g}(\mathbf{p}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ consists of computing the $K + 1$ likelihood functions $S_k(\mathbf{p}; \mathbf{H})$, which require to evaluate the blockage relationship involving K classes of virtual obstacles in a number of relevant grid cells as specified in (3). For a target ground area discretized into M grid cells, the number of relevant grid cells for a propagation path $\mathbf{p}$ is $\mathcal{O}(\sqrt{M})$, and hence, the complexity of computing $\bar{g}(\mathbf{p}; \hat{\boldsymbol{\theta}}, \hat{\mathbf{H}})$ is $\mathcal{O}((K + 1)K\sqrt{M})$.

Second, for the construction of the residual shadowing $\hat{\xi}(\mathbf{p})$ using a subset of data $\mathcal{T}_1$ with N_1 samples, it requires a complexity of $\mathcal{O}(N_1^3)$ (19), which dominates the computational complexity. As a result, the total complexity of evaluating $\hat{g}(\mathbf{p})$ for the proposed scheme is $\mathcal{O}(K^2\sqrt{M} + N_1^3)$.

As for benchmarking, the KNN scheme requires to select K_1 nearest points from $\mathbf{p}$ for the construction, and hence, the complexity is $\mathcal{O}(N_1)$. The complexity of the Kriging scheme is $\mathcal{O}(N_1^3)$ as discussed above.

V. NUMERICAL RESULTS

We study a 310 meters by 340 meters area in central Shanghai, as illustrated in Fig. 6. There are dozens of buildings and other objects with heights ranging from 10 to 130 meters. Their shapes include cubes, columns, and some irregular shapes.³ We chose 100 user locations at random on the ground level, and 50,000 UAV locations uniformly at 50 to 120 meters above ground, resulting in a total of 7,188,362 location samples $\{\mathbf{p}^{(i)}\}$ to form our datasets. Thus, if all samples are used for radio map construction, each sample contributes to a radio map granularity of 1,054 m² for a ground node and 150 m³ for an aerial node. Based on the 3D city map and the deployment of the users and UAVs, two radio map datasets are generated:⁴

Dataset A: The radio map is simulated according to the radio map model $g(\mathbf{p})$ in (1) with $K = 2$ and path loss parameters $(\alpha_0, \beta_0) = (-22, -28)$, $(\alpha_1, \beta_1) = (-28, -24)$, and $(\alpha_2, \beta_2) = (-36, -22)$. independent and identically distributed (i.i.d.) Gaussian noise with zero mean and standard deviation $\sigma_n \in \{3, 7\}$ dB is added to model the shadowing combined with measurement noise, and thus, the residual shadowing component will not be constructed for this dataset. In our 3D city map, we treat the objects lower than 15 meters as vegetation and those higher than 15 meters as concrete to model the obstructed line-of-sight (OLOS) and NLOS propagation, respectively.

Dataset B: The radio map $g(\mathbf{p})$ is generated using Remcom Wireless Insite, a commercial 3D ray-tracing software. Up to 6 reflections and 1 diffraction are simulated, and other parameters are set as default. The material of all structures is set as concrete. The waveforms are chosen as narrowband sinusoidal signals at frequency bands 2.5 GHz and 28 GHz, respectively, in different experiments.

The proposed method reconstructs a radio map with a virtual obstacle map of roughly $M = 1,200$ grid cells with 9 meter spacing between grid points. The choice of M is discussed later.

A. Radio Map Reconstruction

We first evaluate the performance of radio map reconstruction from Dataset A. The performance is evaluated in mean absolute error (MAE) $e = \mathbb{E}\{|\hat{g}(\mathbf{p}) - g(\mathbf{p})|\}$ for the reconstructed radio map. To make a fair comparison, the following baseline schemes are evaluated:

- 1) KNN [13], [14]: To construct the channel quality at each 6D location $\mathbf{p}$, the algorithm first selects 5 measurement samples that are closest to $\mathbf{p}$ from the training set $\{\mathbf{p}^{(i)}\}$ and form the neighbor set as $\mathcal{N}(\mathbf{p})$; then, the channel quality at $\mathbf{p}$ is computed as $\hat{g}(\mathbf{p}) = \mu^{-1} \sum_{i \in \mathcal{N}(\mathbf{p})} w(\mathbf{p}, \mathbf{p}^{(i)}) y^{(i)}$, where $w(\mathbf{p}, \mathbf{p}^{(i)}) = \exp[-\|\mathbf{p} - \mathbf{p}^{(i)}\|_2^2 / (2s^2)]$ with a properly chosen parameter $s = 55$ meters and $\mu = \sum_{i \in \mathcal{N}(\mathbf{p})} w(\mathbf{p}, \mathbf{p}^{(i)})$ is a normalizing factor.

³The 3D city map is available at <https://www.openstreetmap.org>

⁴The code, datasets, settings and user placement map are available at <https://github.com/6wj/radiomap-uav>

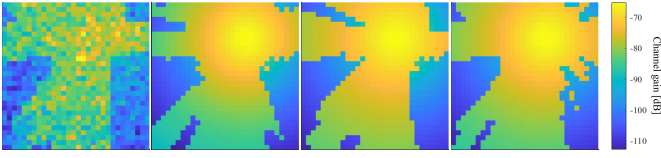


Fig. 2. 2D slices of radio maps from Dataset A at a fixed user position and fixed UAV height, where each pixel represents the link quality between the UAV at the corresponding (x, y) position and the ground user at a fixed position. From left to right: the true radio map, and the deterministic radio maps $\hat{g}(\mathbf{p}; \hat{\theta}, \hat{\mathbf{H}})$ reconstructed from 400, 900, 2500 measurement samples, respectively.

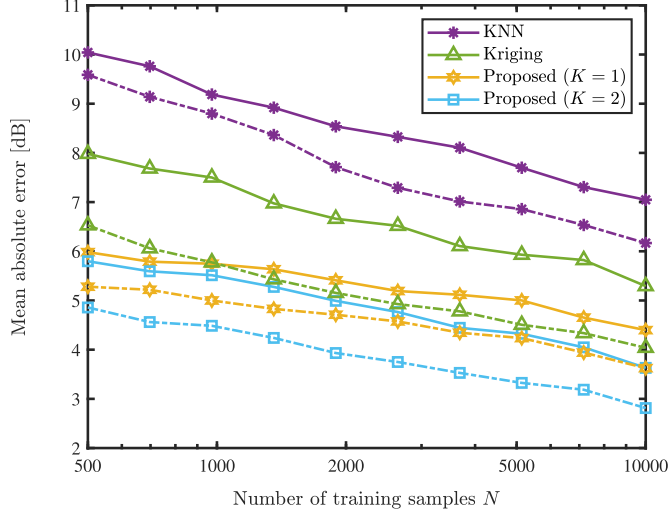


Fig. 3. Reconstruction error versus the number of training samples N for Dataset A, where solid lines are for $\sigma_n = 7$ dB noise and dashed lines for $\sigma_n = 3$ dB noise.

- 2) Kriging [17]: The radio map $\hat{g}(\mathbf{p})$ is constructed based on all the measurement samples $\{(\mathbf{p}^{(i)}, y^{(i)})\}$ using a similar model as in (17) and the model parameters are computed using the method in Section IV.

Several 2D slices of reconstructed radio maps from Dataset A are demonstrated in Fig. 2, where the geometry structure that captures the radio semantics can be observed.

Fig. 3 shows the MAE for the radio map reconstruction in terms of the number of training samples for Dataset A. The proposed schemes follow Algorithm 2 under $N_1 = 0$. It is observed that the proposed radio map reconstruction method can reduce the MAE by 2–4 dB, corresponding to an order of reduction in the sampling complexity, *e.g.*, the proposed method requires only 500 samples to achieve a similar or lower MAE that is achieved by Kriging or KNN using 5,000 samples. Table I summarizes the reconstruction performance in MAE over the parameter K in the proposed schemes using Dataset A with $\sigma_n = 3$ dB noise. We observe that increasing K from 0 to 2 indeed substantially enhances the reconstruction performance; but further increasing K to 3 causes slight performance degradation due to the mismatch with the model that generates the synthetic data in Dataset A.

Fig. 4 shows the MAE versus the number of training samples from Dataset B. The proposed schemes follow Algorithm 2 under $N_1 = N$. We obtain four observations:

- **Environment-Awareness:** The proposed method, which estimates both the environment-aware deterministic radio map $\hat{g}(\mathbf{p})$ based on the constructed virtual environment

TABLE I
RECONSTRUCTION MAE VERSUS K

K	0	1	2	3
$N = 500$	8.63	5.29	4.77	4.89
$N = 2500$	8.68	4.64	3.82	3.88

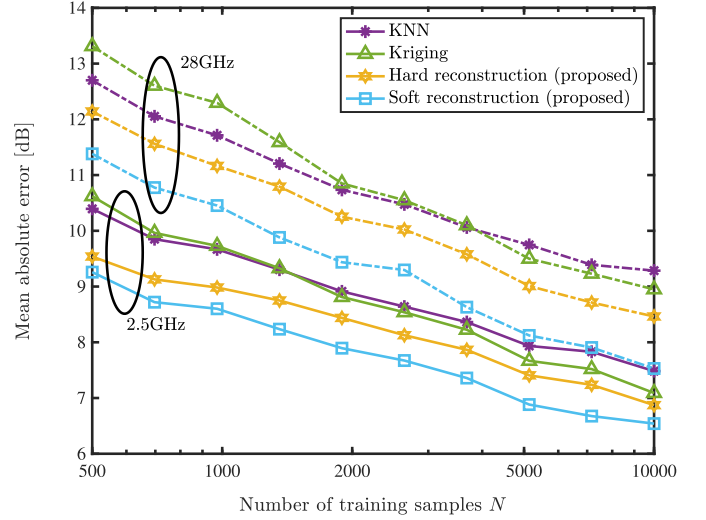


Fig. 4. Reconstruction error versus the number of training samples N for Dataset B, where solid lines are from data of the frequency at 2.5 GHz and dashed lines are from 28 GHz data.

$\mathbf{H}$ and the residual shadowing $\xi(\mathbf{p})$, can bring down the MAE by 1–2 dB, corresponding to more than 50% reduction in measurement samples. The constructed virtual environment $\mathbf{H}$ yields to an improved estimate on the type of propagation $\mathbf{p} \in \mathcal{D}_k$ via the likelihood function $S_k(\mathbf{p}; \mathbf{H})$, resulting in a better estimation on the channel quality compared to the baseline schemes. This confirms that recovering the virtual geometry of propagation environment with radio semantics does help radio map reconstruction.

- **Impact of the Semivariogram Model:** Fig. 5 shows the fitting results for several common semivariogram models used for Kriging. It is observed that the fitting performance is not sensitive to the choice of the semivariogram model. Note that the best choice of the semivariogram model may depend on the propagation environment. Therefore, in practice, one way wish to employ cross-validation for the choice an appropriate semivariogram model.

- **Performance Comparison between KNN and Kriging.** (i) Comparison over simulated data: Dataset A is generated using (1)–(3), where the shadowing component $\xi(\mathbf{p})$ is generated by a stationary Gaussian process within each propagation region. Note that Kriging is known to be the best unbiased minimum variance estimator under the stationary assumption. Although the whole process $\xi(\mathbf{p})$ is non-stationary across propagation regions, it is stationary within each propagation region. These properties may contribute to the good performance of Kriging under simulated data, and this may explain the substantial performance gain of Kriging over KNN. (ii) Comparison over ray-tracing data: Dataset B is generated by

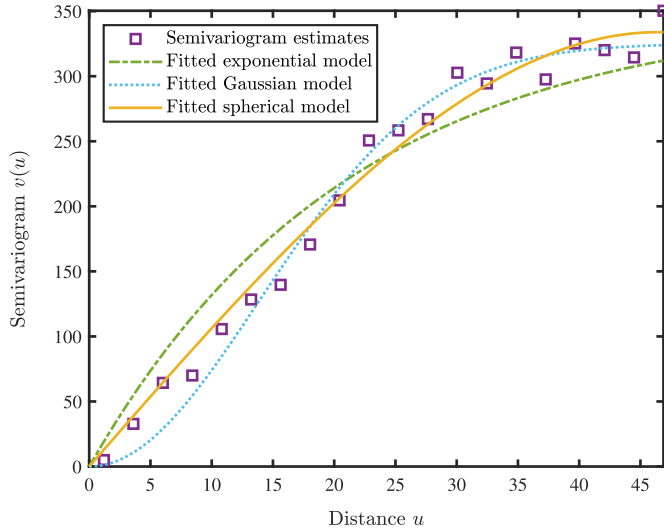


Fig. 5. Semivariogram fitting under Dataset B.

3D ray-tracing, and the stationary assumption on $\xi(\mathbf{p})$ no longer holds even within a propagation region, substantially violating the stationary assumption for the (ordinary) Kriging we tested. This may explain the smaller performance gap between Kriging and KNN.

- **Performance Comparison between Simulated and Ray-Tracing Data:** The model in (1)–(3) that generates the simulated data in Dataset A is relatively simple, where only different levels of signal obstruction are considered. The mechanism of producing the ray-tracing data is very complicated, apart from signal obstruction, signal reflection, diffraction, and scattering are considered and simulated in Dataset B. Since the proposed method is designed based on the model in Equations (1)–(3), it shows a larger performance gain in simulated data than in ray-tracing data.

B. Reconstructing the Geometry of the Radio Environment

We demonstrate the recovered geometry of the surroundings that represent the propagation environment with radio semantics.

Fig. 6 demonstrates the virtual environment reconstruction. For visualization purpose, a simplified Dataset A under $K = 1$ is used, where all the objects in the city map are treated as concrete, and thus, there are only LOS and NLOS propagation regions. Measurement noise of $\sigma_n = 7$ dB is added to the data. It is observed that the geometry of the reconstructed radio environment is roughly consistent with the city map under such a simplified configuration.

Note that the two maps have different physical meanings. The city map represents objects seen by visual light via reflection and scattering, whereas, the virtual obstacle map represents objects “seen” by radio signals via penetration, diffraction, reflection, and scattering, etc. The virtual obstacle map serves as a low dimensional (2D) geometry interpretation of the radio environment.

Furthermore, we obtain some insights from our experiments for the choice of the parameters K and M in constructing the virtual obstacle map.

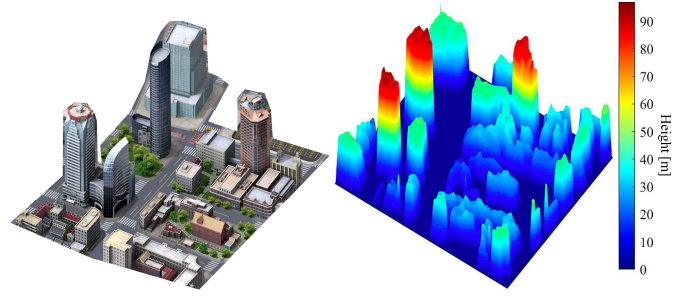


Fig. 6. The left figure is the 3D city map used for generating Dataset A. The right figure demonstrates the reconstructed virtual environment $\hat{\mathbf{H}}$ from Algorithm 1, where the color represents the height of the virtual obstacle in meters.

- **Performance-Complexity Tradeoff:** The number of virtual obstacle types K and the number of grid cells M affect the model complexity as seen in Section II-B. Specifically, the number of parameters to be estimated (such as $h_{m,k}$ in (3)) scales as $\mathcal{O}(MK)$. In an ideal case, a larger model may provide a better approximation to reality. However, it also costs a higher computational burden to estimate the model parameters.
- **Spatial Resolution and Measurement Data Requirement Tradeoff:** The parameters K and M also affect the spatial resolution of the reconstructed radio map. A larger M corresponds to a finer spacing of grid cells for estimating the virtual obstacles, and a larger K corresponds to more types of virtual obstacles per grid cell. Therefore, it is expected that a model with larger K and M may represent finer details in a radio map, and likewise, require more measurement data and more computational resources.

Our experiments suggest that the best practice for choosing M and K is such that there are on average 5–20 measurement links passing over a grid cell for each type of virtual obstacle. Moreover, our earlier work also studied a method to dynamically adjust the resolution M locally according to the amount of measurement data [30].

C. Application in UAV-Aided Relay Communication

Consider a scenario of placing a UAV relay in 3D to establish a relay link for two ground users probably in deep shadow in a dense urban environment. For demonstration purpose, suppose that a half-duplex decode-and-forward relay strategy is used for a narrowband single antenna system. We evaluate the end-to-end capacity from one user to the other via the UAV relay. The capacity is clearly a function of the UAV position $\mathbf{p}_d$ and the radio environment: $C_{\text{DF}}(\mathbf{p}_d) = \frac{1}{2}W \min \{\log_2(1 + \kappa P_b g_{d,b}(\mathbf{p}_d)), \log_2(1 + \kappa P_r g_{u,d}(\mathbf{p}_d))\}$, where $W = 100$ MHz is the bandwidth, $\kappa = 0.5$ is a discount factor to capture the modulation and coding loss, $P_b, P_r = 104$ dB is the ratio of the transmission power 20 dBm over the received noise power $N_0 W$ with $N_0 = -164$ dBm/Hz, and $g_{d,b}(\mathbf{p}_d), g_{u,d}(\mathbf{p}_d)$ are channel gain depending on the UAV position.

We propose to maximize the relay channel capacity by optimizing the UAV position $\mathbf{p}_d$ using the radio map $\tilde{g}(\mathbf{p}) = \hat{g}(\mathbf{p}) + \xi(\mathbf{p})$ constructed in this paper. To benchmark the

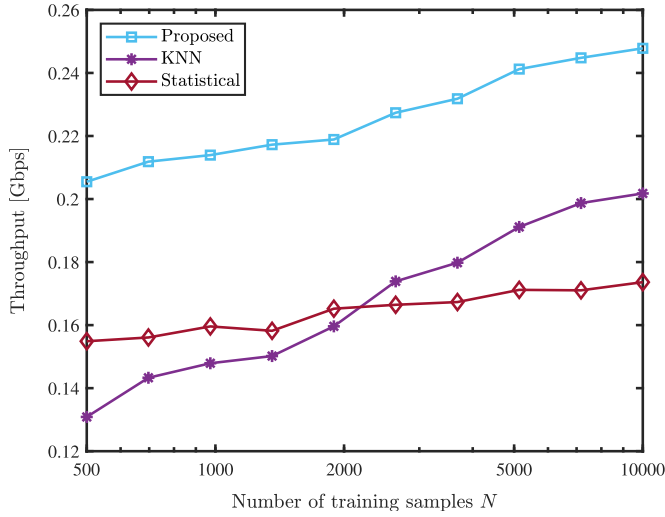


Fig. 7. Average UAV-aided relay channel capacity evaluated over 1,225 user positions from Dataset B in a dense urban environment.

performance, we also evaluate baselines that optimize the UAV position based on the radio map, or propagation model, constructed by the following schemes found in the recent literature:

- 1) Statistical Map [9], [10]: Define $p(\phi(\mathbf{p}))$ as the LOS probability of the user-UAV position pair $\mathbf{p} = (\mathbf{p}_u, \mathbf{p}_d)$ given the elevation angle $\phi(\mathbf{p})$ from the user to the UAV. Then, the channel gain $g_{d,b}(\mathbf{p}_d)$ or $g_{u,d}(\mathbf{p}_d)$ is estimated as $p(\phi(\mathbf{p}))G_0(\mathbf{p}) + (1 - p(\phi(\mathbf{p})))G_1(\mathbf{p})$, where $G_k(\mathbf{p}) = a_k + b_k \log_{10} \tilde{d}(\mathbf{p})$ with $\tilde{d}(\mathbf{p})$ being the propagation distance. The empirical distribution $p(\phi)$ is obtained via offline training based on data $\{(\mathbf{p}^{(i)}, y^{(i)})\}_{i=1}^N$, and the parameters a_k, b_k are empirically fitted from data under LOS (for $k = 0$) or under NLOS (for $k = 1$).
- 2) KNN-based Radio Map: The radio map is constructed using the KNN approach described in Section V-A.

Fig. 7 demonstrates the average capacity versus the number of training samples N from Dataset B for the 2.5 GHz case. Statistical method requires the least number of training samples before its performance saturates, because the model has the least parameters and a poor capability to describe the radio environment. The capacity of both the KNN-based approach and the proposed approach increases as the number of training samples N gets larger. The significant gain of the proposed scheme over the KNN-based approach can be understood from the fact that the proposed method reconstructs a radio map much more accurate than the KNN approach does. Specifically, the proposed scheme achieves more than 50% capacity gain over the KNN-based approach in the small sample regime (around 500 training samples); in addition, it also outperforms the statistical scheme by more than 30% capacity gain and the performance gain further increases when more training samples are used. Thus, we may conclude that an accurate radio map may substantially enhance the performance in a wireless communication system.

VI. CONCLUSION

This paper developed a radio map model and estimation algorithms for constructing 6D radio maps. The core idea

is to construct a multi-class obstacle model from the RSS measurement data to capture the geometry of the environment with radio semantics. A joint estimation problem of the propagation parameters and the 3D virtual environment map is formulated. While the estimation problem is non-convex with degenerated gradient, we discover the partially quasiconvex property, which leads to the development of an efficient parameter estimation radio map construction algorithm. Numerical results demonstrated that by exploiting radio semantics, the required measurement data for achieving a same radio map construction accuracy can be reduced by more than a half. It is also found that when the proposed radio map is applied to UAV placement for UAV-aided relay communication, more than 50% relay capacity gain can be achieved.

APPENDIX A PROOF OF LEMMA 1

Under $y^{(i)} = \bar{g}(\mathbf{p}^{(i)}; \boldsymbol{\theta}^*, \mathbf{H}^*) + n^{(i)}$, we obtain

$$\begin{aligned} f(\boldsymbol{\theta}, \mathbf{H}) &= \frac{1}{N} \sum_{i=1}^N \left[y^{(i)} - \sum_{k=0}^K (\beta_k + \alpha_k d(\mathbf{p}^{(i)})) S_k(\mathbf{p}^{(i)}; \mathbf{H}) \right]^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left[\bar{g}(\mathbf{p}^{(i)}; \boldsymbol{\theta}^*, \mathbf{H}^*) + n^{(i)} \right. \\ &\quad \left. - \sum_{k=0}^K (\beta_k + \alpha_k d(\mathbf{p}^{(i)})) S_k(\mathbf{p}^{(i)}; \mathbf{H}) \right]^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left[(\Delta^{(i)})^2 + 2\Delta^{(i)}n^{(i)} + (n^{(i)})^2 \right] \end{aligned} \quad (20)$$

$$= \bar{f}(\boldsymbol{\theta}, \mathbf{H}) + \frac{1}{N} \sum_{i=1}^N [2\Delta^{(i)}n^{(i)} + (n^{(i)})^2] \quad (21)$$

where, in (20), we denote

$$\Delta^{(i)} = \bar{g}(\mathbf{p}^{(i)}; \boldsymbol{\theta}^*, \mathbf{H}^*) - \sum_{k=0}^K (\beta_k + \alpha_k d(\mathbf{p}^{(i)})) S_k(\mathbf{p}^{(i)}; \mathbf{H})$$

and thus, $\bar{f}(\boldsymbol{\theta}, \mathbf{H}) = \frac{1}{N} \sum_{i=1}^N \Delta^{(i)}$ in (21) as from (7).

It remains to show that the second term in (21) converges to C in probability.

We first show the convergence of $\frac{1}{N} \sum_{i=1}^N 2\Delta^{(i)}n^{(i)}$. For brevity, we denote $X_i = 2\Delta^{(i)}n^{(i)} = 2\Delta^{(i)}(\xi^{(i)} + \tilde{n}^{(i)})$ from the observation model, where we have further denoted $\xi^{(i)} \triangleq \xi(\mathbf{p}^{(i)})$. It follows that $\mathbb{E}\{X_i\} = 0$ since both $\xi^{(i)}$ and $\tilde{n}^{(i)}$ have zero mean. As a result, $\text{cov}(X_i, X_j) = \mathbb{E}\{X_i X_j\} = 4\Delta^{(i)}\Delta^{(j)}\text{cov}(\xi^{(i)}, \xi^{(j)})$ for $i \neq j$, since $\xi^{(i)}$ and $\tilde{n}^{(i)}$ are independent with zero mean.

Note that $\Delta^{(i)}$ is a deterministic function with bounded values due to its definition and the fact that the regions of $\boldsymbol{\theta}$, $\mathbf{H}$, and $\mathbf{p}^{(i)}$ are bounded, *i.e.*, there exists C_0 , such that $4|\Delta^{(i)}\Delta^{(j)}| < C_0$ for all i, j . Therefore, we must have $|\text{cov}(X_i, X_j)| \leq C_0|\text{cov}(\xi^{(i)}, \xi^{(j)})|$ for $i \neq j$. Since $|\text{cov}(\xi^{(i)}, \xi^{(j)})| \rightarrow 0$ as $|i - j| \rightarrow \infty$ and $N \rightarrow \infty$, it follows that $|\text{cov}(X_i, X_j)| \rightarrow 0$.

We are interested in the variance $\mathbb{V}\{\frac{1}{N} \sum_{i=1}^N X_i\}$, which equals to

$$\frac{1}{N^2} \sum_{i=1}^N \mathbb{V}\{X_i\} + \frac{2}{N^2} \sum_{i=1}^N \sum_{j=i+1}^N \text{cov}(X_i, X_j) \quad (22)$$

where, in the first term, $\mathbb{V}\{X_i\} \leq C_1$ since $\xi^{(i)}$ and $\tilde{n}^{(i)}$ have bounded variance and $\Delta^{(i)}$ is bounded. In the second term, the condition $|\text{cov}(X_i, X_j)| \rightarrow 0$ implies that for any $\epsilon_1 > 0$, there exists a finite N_1 such that $\text{cov}(X_i, X_j) < \epsilon_1$ for all $j - i > N_1$. So, it follows that

$$\begin{aligned} & \sum_{j=i+1}^N \text{cov}(X_i, X_j) \\ &= \sum_{j=i+1}^{i+N_1} \text{cov}(X_i, X_j) + \sum_{j=i+N_1+1}^N \text{cov}(X_i, X_j) \\ &\leq \sum_{j=i+1}^{i+N_1} \sqrt{\mathbb{V}\{X_i\}\mathbb{V}\{X_j\}} + \sum_{j=i+N_1+1}^N \epsilon_1 \\ &\leq N_1 C_1 + N \epsilon_1 \end{aligned}$$

where the first term in the second line is due to the Cauchy-Schwarz inequality.

As a result, $\mathbb{V}\{\frac{1}{N} \sum_{i=1}^N X_i\}$ from (22) can be upper bounded as

$$\begin{aligned} \mathbb{V}\left\{\frac{1}{N} \sum_{i=1}^N X_i\right\} &\leq \frac{1}{N^2} \sum_{i=1}^N C_1 + \frac{2}{N^2} \sum_{i=1}^N (N_1 C_1 + N \epsilon_1) \\ &= \frac{(1 + 2N_1)C_1}{N} + 2\epsilon_1. \end{aligned}$$

Therefore, for any $\epsilon > 0$, one can choose $\epsilon_1 = \epsilon/3$ and $N_2 = 3(1 + 2N_1)C_1/\epsilon$, such that $\mathbb{V}\{\frac{1}{N} \sum_{i=1}^N X_i\} < \epsilon$ for all $N > N_2$. This establishes that $\mathbb{V}\{\frac{1}{N} \sum_{i=1}^N X_i\} \rightarrow 0$ as $N \rightarrow \infty$.

As a result, by Chebyshev's inequality, we have

$$\mathbb{P}\left\{\left|\frac{1}{N} \sum_{i=1}^N X_i - \mathbb{E}\{X_1\}\right| > \epsilon_0\right\} \leq \frac{\mathbb{V}\{\frac{1}{N} \sum_{i=1}^N X_i\}}{\epsilon_0^2} \rightarrow 0$$

as $N \rightarrow \infty$. This shows that $\frac{1}{N} \sum_{i=1}^N X_i \rightarrow 0$ in probability.

Finally, to show the convergence of $\frac{1}{N} \sum_{i=1}^N (n^{(i)})^2$ in (21), we have

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N (n^{(i)})^2 &= \frac{1}{N} \sum_{i=1}^N (\tilde{n}^{(i)})^2 + 2 \frac{1}{N} \sum_{i=1}^N \tilde{n}^{(i)} \xi^{(i)} \\ &\quad + \frac{1}{N} \sum_{i=1}^N (\xi^{(i)})^2 \end{aligned}$$

where the first term and the second term respectively converges to $\mathbb{E}\{(\tilde{n}^{(i)})^2\} = \sigma_n^2$ and $\mathbb{E}\{\tilde{n}^{(i)} \xi^{(i)}\} = 0$ in probability, as $N \rightarrow \infty$, due to the weak law of large number and the fact that $\tilde{n}^{(i)}$ and $\xi^{(i)}$ are independent;⁵ the third term converges to a finite value C_2 from the assumption of the lemma.

Therefore, we have shown that $f(\boldsymbol{\theta}, \mathbf{H}) \rightarrow \bar{f}(\boldsymbol{\theta}, \mathbf{H}) + C$ for $C = \sigma_n^2 + C_2$ in probability as $N \rightarrow \infty$ for every $(\boldsymbol{\theta}, \mathbf{H})$.

⁵More rigorously, the convergence of the second term can be proven by Chebyshev's inequality following a similar procedure as proving the convergence of $\frac{1}{N} \sum_{i=1}^N X_i$.

APPENDIX B PROOF OF THEOREM 1

To simplify the notations of the proof, we denote $\gamma_k^{(i)} = \beta_k^* + \alpha_k^* d(\mathbf{p}^{(i)})$ as the path loss for $d(\mathbf{p}^{(i)})$ and the k th degree of signal obstruction, $\mathbb{I}_k^{(i,j)}(\mathbf{h}) = \mathbb{I}\{\mathbf{p}^{(i)} + \boldsymbol{\epsilon}_j \in \mathcal{D}_k(\mathbf{h})\}$, $\forall j = 1, \dots, J-1$, $\mathbb{I}_k^{(i)}(\mathbf{h}) = \mathbb{I}\{\mathbf{p}^{(i)} \in \mathcal{D}_k(\mathbf{h})\}$. Accordingly, $\bar{g}^{(i)} \triangleq \bar{g}(\mathbf{p}^{(i)}; \boldsymbol{\theta}^*, \mathbf{h}^*) = \sum_{k=0}^1 \gamma_k^{(i)} S_k(\mathbf{p}^{(i)}; \mathbf{h}^*)$ from (2), and $S_k(\mathbf{p}^{(i)}; \mathbf{h}^*) = \sum_{j=0}^{J-1} \omega_j \mathbb{I}_k^{(i,j)}(\mathbf{h}^*)$ from (4).

Define variable $\mathbf{h}_{-m}$ as a vector from $\mathbf{h}$ but with the m th element h_m removed. Thus, by restricting $\bar{f}(\boldsymbol{\theta}, \mathbf{h})$ in (7) to take value in the interval $\mathcal{I}_m$, we have

$$\begin{aligned} \bar{f}_m(h_m; \boldsymbol{\theta}^*, \mathbf{h}_{-m}) &= \frac{1}{N} \sum_{i=1}^N \left[\bar{g}^{(i)} - \sum_{k=0}^1 \gamma_k^{(i)} \sum_{j=0}^{J-1} \omega_j \mathbb{I}_k^{(i,j)}(\mathbf{h}) \right]^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left[\bar{g}^{(i)} - \sum_{k=0}^1 \gamma_k^{(i)} \omega_0 \mathbb{I}_k^{(i)}(\mathbf{h}) \right. \\ &\quad \left. - \sum_{k=0}^1 \gamma_k^{(i)} \sum_{j=1}^{J-1} \omega_j \mathbb{I}_k^{(i,j)}(\mathbf{h}) \right]^2. \end{aligned} \quad (23)$$

Substituting $\bar{g}^{(i)}$ into (23), we obtain

$$\begin{aligned} \bar{f}_m(h_m; \boldsymbol{\theta}^*, \mathbf{h}_{-m}) &= \frac{1}{N} \sum_{i=1}^N \left[\sum_{k=0}^1 \gamma_k^{(i)} \omega_0 \mathbb{I}_k^{(i)}(\mathbf{h}^*) - \sum_{k=0}^1 \gamma_k^{(i)} \omega_0 \mathbb{I}_k^{(i)}(\mathbf{h}) \right. \\ &\quad \left. + \sum_{k=0}^1 \gamma_k^{(i)} \sum_{j=1}^{J-1} \omega_j \mathbb{I}_k^{(i,j)}(\mathbf{h}^*) - \sum_{k=0}^1 \gamma_k^{(i)} \sum_{j=1}^{J-1} \omega_j \mathbb{I}_k^{(i,j)}(\mathbf{h}) \right]^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left[\underbrace{\omega_0 \sum_{k=0}^1 \gamma_k^{(i)} (\mathbb{I}_k^{(i)}(\mathbf{h}^*) - \mathbb{I}_k^{(i)}(\mathbf{h}))}_{(a)} \right. \\ &\quad \left. + \underbrace{\sum_{j=1}^{J-1} \omega_j \sum_{k=0}^1 \gamma_k^{(i)} (\mathbb{I}_k^{(i,j)}(\mathbf{h}^*) - \mathbb{I}_k^{(i,j)}(\mathbf{h}))}_{(b)} \right]^2. \end{aligned} \quad (24)$$

We have three cases for the value of $\sum_{k=0}^1 \gamma_k^{(i)} (\mathbb{I}_k^{(i)}(\mathbf{h}^*) - \mathbb{I}_k^{(i)}(\mathbf{h}))$ or $\sum_{k=0}^1 \gamma_k^{(i)} (\mathbb{I}_k^{(i,j)}(\mathbf{h}^*) - \mathbb{I}_k^{(i,j)}(\mathbf{h}))$, i.e., 0, $\gamma_0^{(i)} - \gamma_1^{(i)}$, and $\gamma_1^{(i)} - \gamma_0^{(i)}$. Notice that, $\gamma_0^{(i)} > \gamma_1^{(i)}$ from our model (1) and smaller k means less signal obstruction. $\mathbb{I}_k^{(i,j)}(\mathbf{h}^*) - \mathbb{I}_k^{(i,j)}(\mathbf{h}) \in \{-1, 0, 1\}$.

- For term (a): We have, for both $k = 0$ and 1 , term (a) $\in \{0, \omega_0(\gamma_0^{(i)} - \gamma_1^{(i)}), \omega_0(\gamma_1^{(i)} - \gamma_0^{(i)})\}$.
- For term (b): We obtain a lower bound and an upper bound as

$$\sum_{j=1}^{J-1} \omega_j (\gamma_1^{(i)} - \gamma_0^{(i)}) \leq (b) \leq \sum_{j=1}^{J-1} \omega_j (\gamma_0^{(i)} - \gamma_1^{(i)}). \quad (25)$$

Set $\omega_0 \geq \frac{2}{3}$, it follows that $0 \leq \sum_{j=1}^{J-1} \omega_j \leq \frac{1}{3}$ since $\sum_j \omega_j = 1$, and hence,

$$[\pm \omega_0 + \sum_{j=1}^{J-1} \omega_j]^2 \geq [\sum_{j=1}^{J-1} \omega_j]^2. \quad (26)$$

By (25) and (26), for any i , w.r.t. the value of term (a),

- Substituting the upper bound in (25) into $[(a) + (b)]^2$ in (24), we obtain

$$\begin{aligned} & [\pm\omega_0(\gamma_0^{(i)} - \gamma_1^{(i)}) + \sum_{j=1}^{J-1} \omega_j(\gamma_0^{(i)} - \gamma_1^{(i)})]^2 \\ & \geq [0 + \sum_{j=1}^{J-1} \omega_j(\gamma_0^{(i)} - \gamma_1^{(i)})]^2. \end{aligned}$$

- Substituting the lower bound in (25) into $[(a) + (b)]^2$ in (24), we obtain

$$\begin{aligned} & [\pm\omega_0(\gamma_0^{(i)} - \gamma_1^{(i)}) + \sum_{j=1}^{J-1} \omega_j(\gamma_1^{(i)} - \gamma_0^{(i)})]^2 \\ & \geq [0 + \sum_{j=1}^{J-1} \omega_j(\gamma_1^{(i)} - \gamma_0^{(i)})]^2. \end{aligned}$$

Hence,

$$\begin{aligned} & [\pm\omega_0(\gamma_0^{(i)} - \gamma_1^{(i)}) + \sum_{j=1}^{J-1} \omega_j \sum_{k=0}^1 \gamma_k^{(i)} (\mathbb{I}_k^{(i,j)}(\mathbf{h}^*) - \mathbb{I}_k^{(i,j)}(\mathbf{h}))]^2 \\ & \geq [0 + \sum_{j=1}^{J-1} \omega_j \sum_{k=0}^1 \gamma_k^{(i)} (\mathbb{I}_k^{(i,j)}(\mathbf{h}^*) - \mathbb{I}_k^{(i,j)}(\mathbf{h}))]^2 \quad (27) \end{aligned}$$

where $\pm\omega_0(\gamma_0^{(i)} - \gamma_1^{(i)})$ is for $\mathbb{I}_k^{(i)}(\mathbf{h}^*) \neq \mathbb{I}_k^{(i)}(\mathbf{h})$ and 0 is for $\mathbb{I}_k^{(i)}(\mathbf{h}^*) = \mathbb{I}_k^{(i)}(\mathbf{h})$ in term (a).

Denote the right hand side in (27) as $A^{(i)}$, the left hand side as $B^{(i)}$, and two sets $\mathcal{I}_A(\mathbf{h}) = \{i \mid \mathbb{I}_k^{(i)}(\mathbf{h}) = \mathbb{I}_k^{(i)}(\mathbf{h}^*)\}$, $\mathcal{I}_B(\mathbf{h}) = \{i \mid \mathbb{I}_k^{(i)}(\mathbf{h}) \neq \mathbb{I}_k^{(i)}(\mathbf{h}^*)\} = \{1, 2, \dots, N\} \setminus \mathcal{I}_A(\mathbf{h})$.

Recall the obstacle indicator function in (3)

$$\begin{aligned} & \mathbb{I}\{\mathbf{p}^{(i)} \in \mathcal{D}_k(\mathbf{H})\} \\ & = (1 - \prod_{j \in \mathcal{B}^{(i)}} (1 - \mathbb{I}\{h_{j,k} \geq z_j^{(i)}\})) \prod_{j \in \mathcal{B}^{(i)} l > k} \mathbb{I}\{h_{j,l} < z_j^{(i)}\} \end{aligned}$$

and we denote a set

$$\mathcal{L}_m(\mathbf{h}) = \{i \mid h_j < z_j^{(i)}, \forall j \in \mathcal{B}^{(i)} \setminus \{m\}\}$$

as measurement samples that can only be blocked by m th grid. Then, w.r.t. h_m , values of $\mathbb{I}_k^{(i)}(\mathbf{h})$ have two types: (i) dependent of h_m , i.e., index $i \in \mathcal{L}_m(\mathbf{h})$, and by the condition $h_j^* < h_j$ and $h_j < z_j^{(i)}, \forall j \in \mathcal{B}^{(i)} \setminus \{m\}$,

$$\begin{aligned} & \mathbb{I}_k^{(i)}(h_m; \mathbf{h}_{-m}) \\ & = (1 - (1 - \mathbb{I}\{h_{m,k} \geq z_{m,k}^{(i)}\})) \prod_{l > k} \mathbb{I}\{h_{m,l} < z_m^{(i)}\} \end{aligned}$$

so $\mathbb{I}_0^{(i)}(h_m; \mathbf{h}_{-m}) = \mathbb{I}\{h_m < z_m^{(i)}\} = \mathbb{I}_0^{(i)}(h_m; \mathbf{h}_{-m}^*)$ and $\mathbb{I}_1^{(i)}(h_m; \mathbf{h}_{-m}) = \mathbb{I}\{h_m \geq z_m^{(i)}\} = \mathbb{I}_1^{(i)}(h_m; \mathbf{h}_{-m}^*)$; (ii) independent of h_m , i.e., index $i \in \{1, 2, \dots, N\} \setminus \mathcal{L}_m(\mathbf{h})$, and then $\mathbb{I}_k^{(i)}(h_m; \mathbf{h}_{-m})$ will be constant. Hence, we only need to consider the type (i) samples. For index $i \in \mathcal{L}_m(\mathbf{h})$,

- LOS measurement samples satisfy $h_m^* < z_m^{(i)}$. As shown in Fig. 8, when $h_m < z_m^{(i)}$, we obtain $\mathbb{I}_k^{(i)}(\mathbf{h}) = \mathbb{I}_k^{(i)}(\mathbf{h}^*)$ and $[(a) + (b)]^2$ in (24) taking value $A^{(i)}$. When

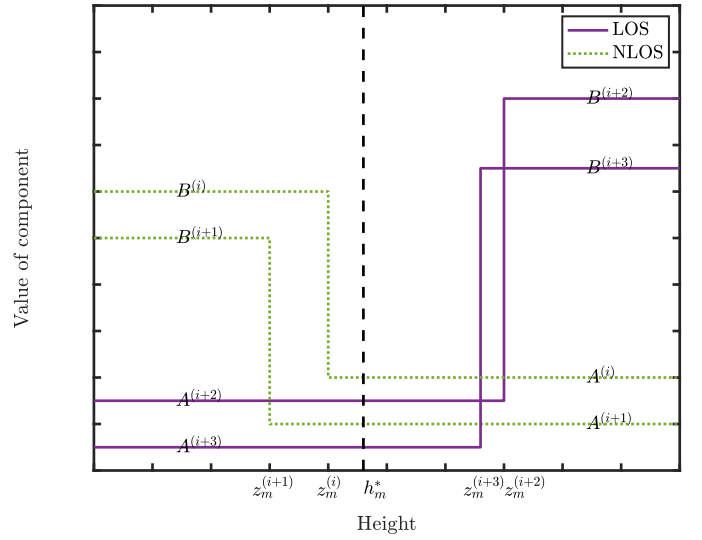


Fig. 8. An example for the value of component $[\bar{g}^{(i)} - \sum_{k=0}^1 \gamma_k^{(i)} \sum_{j=0}^{J-1} \omega_j \mathbb{I}_k^{(i,j)}(\mathbf{h})]^2$ vs. h_m .

$h_m \geq z_m^{(i)}$, we obtain $\mathbb{I}_k^{(i)}(\mathbf{h}) \neq \mathbb{I}_k^{(i)}(\mathbf{h}^*)$ and $[(a) + (b)]^2$ taking value $B^{(i)}$.

- NLOS measurement samples satisfy $h_m^* \geq z_m^{(i)}$. As shown in Fig. 8, when $h_m \geq z_m^{(i)}$, we obtain $\mathbb{I}_k^{(i)}(\mathbf{h}) = \mathbb{I}_k^{(i)}(\mathbf{h}^*)$ and $[(a) + (b)]^2$ in (24) taking value $A^{(i)}$. When $h_m < z_m^{(i)}$, we obtain $\mathbb{I}_k^{(i)}(\mathbf{h}) \neq \mathbb{I}_k^{(i)}(\mathbf{h}^*)$ and $[(a) + (b)]^2$ taking value $B^{(i)}$.

Therefore, the smaller the distance between h_m and h_m^* is, the larger the set $\mathcal{I}_A(h_m; \mathbf{h}_{-m})$ will be, i.e., the more $\mathbb{I}_k^{(i)}(h_m; \mathbf{h}_{-m}) = \mathbb{I}_k^{(i)}(h_m^*; \mathbf{h}_{-m}^*)$. $\forall \epsilon > 0$, when $h_m < h_m^*$, $\mathcal{I}_A(h_m - \epsilon; \mathbf{h}_{-m}) \subseteq \mathcal{I}_A(h_m; \mathbf{h}_{-m})$, and when $h_m > h_m^*$, $\mathcal{I}_A(h_m + \epsilon; \mathbf{h}_{-m}) \subseteq \mathcal{I}_A(h_m; \mathbf{h}_{-m})$. Intuitively, for index $i \in \mathcal{L}_m(\mathbf{h})$, the value of $[(a) + (b)]^2$ in (24) is shown in Fig. 8, and for index $i \in \{1, 2, \dots, N\} \setminus \mathcal{L}_m(\mathbf{h})$, the value is constant in h_m . We can see the summation of all LOS is increasing in h_m and is constant from zero to a height which is greater than h_m^* . Similarly, the summation of all NLOS is decreasing in h_m and is constant from zero to a height which is less than h_m^* , so the summation of all LOS and NLOS is quasiconvex. Then, we will show the relation between the sets and the function value mathematically.

$$\bar{f}_m(h_m; \boldsymbol{\theta}^*, \mathbf{h}_{-m}) = \sum_{i \in \mathcal{I}_A(\mathbf{h})} A^{(i)} + \sum_{i \in \mathcal{I}_B(\mathbf{h})} B^{(i)}$$

where $0 < A^{(i)} \leq B^{(i)}$. $\forall \epsilon > 0$, when $\forall h_m < h_m^*$, by using $\mathcal{I}_A(h_m - \epsilon; \mathbf{h}_{-m}) \subseteq \mathcal{I}_A(h_m; \mathbf{h}_{-m})$ and $\mathcal{I}_A(h_m; \mathbf{h}_{-m}) = \{1, 2, \dots, N\} \setminus \mathcal{I}_A(h_m; \mathbf{h}_{-m})$, we finally have $\bar{f}_m(h_m - \epsilon; \boldsymbol{\theta}^*, \mathbf{h}_{-m}) \geq \bar{f}_m(h_m; \boldsymbol{\theta}^*, \mathbf{h}_{-m})$. Similarly, $\forall \epsilon > 0$, when $\forall h_m > h_m^*$, $\bar{f}_m(h_m + \epsilon; \boldsymbol{\theta}^*, \mathbf{h}_{-m}) \geq \bar{f}_m(h_m; \boldsymbol{\theta}^*, \mathbf{h}_{-m})$. To sum up, $\bar{f}_m(h_m; \boldsymbol{\theta}^*; \mathbf{h}_{-m})$ is quasiconvex in h_m .

APPENDIX C PROOF OF THEOREM 2

By Theorem 1, we already have $\bar{f}_{m,1}(h_{m,1}; \boldsymbol{\theta}^*, \mathbf{H}_{m,1}^-)$ is quasiconvex under the given condition, where $\mathbf{H}_{m,k}^-$ is a matrix representing $\mathbf{H}$ removing $h_{m,k}$. Then, we generalize

the proof to $K > 1$, when $S_k(\mathbf{p}^{(i)}; \mathbf{H})$ is chosen as the indicator function, *i.e.*, $\omega_0 = 1$. As mentioned previously, $\forall m$, given $\mathbf{H}_{m,1}^- \succeq \mathbf{H}_{m,1}^-$ and obstacle indicator function (3), values of $\mathbb{I}_k^{(i)}(\mathbf{H})$ have two types: (i) $\mathbb{I}_0^{(i)}(\mathbf{H}) = \mathbb{I}\{h_{m,1} < z_m^{(i)}\}$, $\mathbb{I}_1^{(i)}(\mathbf{H}) = \mathbb{I}\{h_{m,1} \geq z_m^{(i)}\}$; (ii) independent of $h_{m,1}$. The case (i) is also known as measurement samples which can only be blocked by m th grid and denote it as

$$\mathcal{L}_{m,K}(\mathbf{H}) = \left\{ i \mid h_{j,K} < z_{j,K}^{(i)}, \forall j \in \mathcal{B}^{(i)} \setminus \{m\} \right\} \quad (28)$$

where $K = 1$. In a general $K > 1$ case, we denote a set $\mathcal{L}_{m,k}(\mathbf{H})$ for samples only blocked by m th grid and no more obstructed than k th class obstacle

$$\mathcal{L}_{m,k}(\mathbf{H}) = \left\{ i \mid h_{j,l} < z_j^{(i)}, \forall l \geq k, \forall j \in \mathcal{B}^{(i)} \setminus \{m\} \right\}.$$

For quasiconvex of NLOS obstacle, *i.e.*, obstacle of the most obscured propagation, we can similarly follow the pipeline of previous $K = 1$ situation, function (7) becomes

$$\bar{f}_{m,K}(h_{m,K}; \boldsymbol{\theta}^*, \mathbf{H}_{m,k}^-) = \frac{1}{N} \sum_{i=1}^N [\bar{g}^{(i)} - \sum_{k=0}^K \gamma_k^{(i)} \mathbb{I}_k^{(i)}(\mathbf{H})]^2. \quad (29)$$

Recall that $\bar{g}^{(i)} = \sum_{k=0}^K \gamma_k^{(i)} \mathbb{I}_k^{(i)}(\mathbf{H}^*)$ is from (2), and we denote $\gamma_k^{(i)} = \beta_k^* + \alpha_k^* d(\mathbf{p}^{(i)})$ as the path loss for $d(\mathbf{p}^{(i)})$ and the k th degree of signal obstruction, $\mathbb{I}_k^{(i)}(\mathbf{H}) = \mathbb{I}\{\mathbf{p}^{(i)} \in \mathcal{D}_k(\mathbf{H})\}$. Substitute $\bar{g}^{(i)}$ into (29). We obtain

$$\begin{aligned} \bar{f}(h_{m,K}; \boldsymbol{\theta}^*, \mathbf{H}_{m,k}^-) &= \frac{1}{N} \sum_{i=1}^N \left[\sum_{k=0}^K \gamma_k^{(i)} \mathbb{I}_k^{(i)}(\mathbf{H}^*) - \sum_{k=0}^K \gamma_k^{(i)} \mathbb{I}_k^{(i)}(\mathbf{H}) \right]^2 \\ &= \frac{1}{N} \sum_{i=1}^N \left[\sum_{k=0}^K \gamma_k^{(i)} (\mathbb{I}_k^{(i)}(\mathbf{H}^*) - \mathbb{I}_k^{(i)}(\mathbf{H})) \right]^2. \end{aligned}$$

For term $\sum_{k=0}^K \gamma_k^{(i)} (\mathbb{I}_k^{(i)}(\mathbf{H}^*) - \mathbb{I}_k^{(i)}(\mathbf{H}))$ in it: We have

$$\sum_{k=0}^K \gamma_k^{(i)} (\mathbb{I}_k^{(i)}(\mathbf{H}^*) - \mathbb{I}_k^{(i)}(\mathbf{H})) = \begin{cases} A^{(i)}, & \text{if } \mathbb{I}_k^{(i)}(\mathbf{H}^*) = \mathbb{I}_k^{(i)}(\mathbf{H}) \\ B^{(i)}, & \text{if } \mathbb{I}_k^{(i)}(\mathbf{H}^*) \neq \mathbb{I}_k^{(i)}(\mathbf{H}) \end{cases}$$

where $A^{(i)} = 0$ and $B^{(i)} \neq 0$. Denote two sets $\mathcal{I}_A(\mathbf{H}) = \{i \mid \mathbb{I}_k^{(i)}(\mathbf{H}) = \mathbb{I}_k^{(i)}(\mathbf{H}^*)\}$, $\mathcal{I}_B(\mathbf{H}) = \{i \mid \mathbb{I}_k^{(i)}(\mathbf{H}) \neq \mathbb{I}_k^{(i)}(\mathbf{H}^*)\} = \{1, 2, \dots, N\} \setminus \mathcal{I}_A(\mathbf{H})$ and similar to the proof of Theorem 1, the smaller the distance between $h_{m,K}$ and $h_{m,K}^*$ is, the larger the set $\mathcal{I}_A(h_{m,K}; \mathbf{H}_{m,k}^-)$ will be, *i.e.*, the more $\mathbb{I}_k^{(i)}(h_{m,K}; \mathbf{H}_{m,k}^-) = \mathbb{I}_k^{(i)}(h_{m,K}^*; \mathbf{H}_{m,k}^-)$. Therefore, $\bar{f}_{m,K}(h_{m,K}; \boldsymbol{\theta}^*, \mathbf{H}_{m,k}^-)$ is quasiconvex in $h_{m,K}$. We have the sets $\mathcal{L}_{m,K}(\mathbf{H})$ and class- K obstacle heights $h_{m,K}$, where $m = 1, 2, \dots, M$. Followed by NLOS obstacle, the obstacle of less obscured propagation is class- $K - 1$ and the set is

$$\mathcal{L}_{m,K-1}(\mathbf{H}) = \left\{ i \mid h_{j,l} < z_j^{(i)}, \forall l \geq K-1, \forall j \in \mathcal{B}^{(i)} \setminus \{m\} \right\}$$

in which $l = K - 1, K$. On the condition of the sets $\mathcal{L}_{m,K}(\mathbf{H})$ and class- K obstacle heights $h_{m,K}$, $\mathcal{L}_{m,K-1}(\mathbf{H})$ can be also written as

$$\begin{aligned} \mathcal{L}_{m,K-1}(\mathbf{H}) &= \left\{ i \mid h_{j,K-1} < z_j^{(i)}, \forall j \in \mathcal{B}^{(i)} \setminus \{m\} \right\} \cap \mathcal{L}_{m,K}(\mathbf{H}) \end{aligned}$$

which is now also similar to the previous situation (28). In other words, when we have the most obscured propagation situation, the less obscured propagation situation can be recursively get. As a result, for all m, k , quasiconvexity of $\bar{f}_{m,k}(h_{m,k}; \boldsymbol{\theta}^*, \mathbf{H}_{m,k}^-)$ is proved and then $\bar{f}(\boldsymbol{\theta}^*, \mathbf{H})$ is element-wise quasiconvex for each individual element $h_{m,k}$.

APPENDIX D

PROOF OF THEOREM 3

First, the optimal solution $\mathcal{H}^*$ to minimize function (7) are intervals, because in the objective function, variable $\mathbf{H}$ is only in an indicator function taking 1 by comparing the range of those values. Denote $\mathcal{M}_m$ as the index set of measurements $(\mathbf{p}^{(i)}, y^{(i)})$ that the link $(\mathbf{p}_u^{(i)}, \mathbf{p}_d^{(i)})$ passes over the m th grid cell. Then, by the proof of Theorem 2, similar to $\mathcal{L}_{m,k}(\mathbf{H})$, denote a set $\mathcal{Q}_{m,k}(\mathbf{H})$ for samples passing m th grid and no more obstructed than k th class obstacle, *i.e.*,

$$\mathcal{Q}_{m,k}(\mathbf{H}) = \left\{ i \mid h_{j,l} < z_j^{(i)}, \forall l \geq k, \forall j \in \mathcal{B}^{(i)}, \forall i \in \mathcal{M}_m \right\} \quad (30)$$

and the optimal set $\mathcal{Q}_{m,k}(\mathbf{H}^*)$

$$\mathcal{Q}_{m,k}(\mathbf{H}^*) = \left\{ i \mid h_{j,l}^* < z_j^{(i)}, \forall l \geq k, \forall j \in \mathcal{B}^{(i)}, \forall i \in \mathcal{M}_m \right\}$$

So it follows that $\mathcal{Q}_{m,1}(\mathbf{H}) \subseteq \mathcal{Q}_{m,2}(\mathbf{H}) \subseteq \dots \subseteq \mathcal{Q}_{m,K}(\mathbf{H})$, $\forall m, K > 1$. Recall the obstacle indicator function in (3), then it is obvious that $h_{m,1} \geq h_{m,2} \geq \dots \geq h_{m,K}$, $\forall m, K > 1$. Given the condition that $\mathbf{H}' \succeq \mathbf{H}^*$, we thus have

$$\mathcal{Q}_{m,k}(\mathbf{H}') \subseteq \mathcal{Q}_{m,k}(\mathbf{H}^*), \quad \forall m, k. \quad (31)$$

From the proof of Theorem 2 and Fig. 8, denote the k th class obstacle height estimation of m th grid

$$\begin{aligned} \hat{h}_{m,k} &= \sup \arg \min_{h_{m,k}} \bar{f}_{m,k}(h_{m,k}; \boldsymbol{\theta}^*, \mathbf{H}_{m,k}^-) \\ &= \min(z_m^{(i)}), \quad i \in \mathcal{Q}_{m,k}(\mathbf{H}'). \end{aligned} \quad (32)$$

Therefore, by (31)–(32),

$$\begin{aligned} \hat{h}_{m,k} &= \min(z_m^{(i)}), \quad i \in \mathcal{Q}_{m,k}(\mathbf{H}') \\ &\geq \min(z_m^{(i)}), \quad i \in \mathcal{Q}_{m,k}(\mathbf{H}^*) \\ &= h_{m,k}^*. \end{aligned}$$

Similarly, for $k = K, K - 1, \dots, 1$ and for all m , we have the same results, so $\hat{\mathbf{H}} \succeq \mathbf{H}^*$.

As for the $\hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}'') \geq \hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}')$ under $\mathbf{H}'' \succeq \mathbf{H}' \succeq \mathbf{H}^*$: When $\mathbf{H}'$ taking value of $1H_{\max}$, $\mathbf{H}''$ can only equal to $1H_{\max}$ and

$$\begin{aligned} \hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}') &= \min(z_m^{(i)}), \quad i \in \mathcal{Q}_{m,k}(1H_{\max}) \\ &= \hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}'') \end{aligned}$$

so $\hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}'') \geq \hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}')$ and $H_{\max} \geq \hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}')$. For arbitrary $\mathbf{H}'$, when $\mathbf{H}'' \succeq \mathbf{H}' \succeq \mathbf{H}^*$ is satisfied, by (30) $\mathcal{Q}_{m,k}(\mathbf{H}'') \subseteq \mathcal{Q}_{m,k}(\mathbf{H}')$ and similar to (32)

$$\begin{aligned} \hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}'') &= \min(z_m^{(i)}), \quad i \in \mathcal{Q}_{m,k}(\mathbf{H}'') \\ &\geq \min(z_m^{(i)}), \quad i \in \mathcal{Q}_{m,k}(\mathbf{H}') \\ &= \hat{h}_{m,k}(\boldsymbol{\theta}^*, \mathbf{H}') \end{aligned}$$

thus it always holds.

REFERENCES

- [1] Y.-G. Lim, Y. J. Cho, M. S. Sim, Y. Kim, C.-B. Chae, and R. A. Valenzuela, "Map-based millimeter-wave channel models: An overview, data for B5G evaluation and machine learning," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 54–62, Aug. 2020.
- [2] X. Mo, Y. Huang, and J. Xu, "Radio-map-based robust positioning optimization for UAV-enabled wireless power transfer," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 179–183, Feb. 2019.
- [3] S. Zhang and R. Zhang, "Radio map-based 3D path planning for cellular-connected UAV," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1975–1989, Mar. 2021.
- [4] O. Esrafilian, R. Gangula, and D. Gesbert, "Three-dimensional-map-based trajectory design in UAV-aided wireless localization systems," *IEEE Internet Things J.*, vol. 8, no. 12, pp. 9894–9904, Jun. 2020.
- [5] Y. Zeng and X. Xu, "Toward environment-aware 6G communications via channel knowledge map," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 84–91, Jun. 2021.
- [6] X. Xia, Y. Wang, K. Xu, and Y. Xu, "Toward digitalizing the wireless environment: A unified A2G information and energy delivery framework based on binary channel feature map," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6448–6463, Aug. 2022.
- [7] Y. Zeng, X. Xu, S. Jin, and R. Zhang, "Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4205–4220, Jul. 2021.
- [8] Q. Hu, Y. Cai, A. Liu, G. Yu, and G. Y. Li, "Low-complexity joint resource allocation and trajectory design for UAV-aided relay networks with the segmented ray-tracing channel model," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 6179–6195, Sep. 2020.
- [9] A. Al-Hourani, S. Kandeepan, and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 2014, pp. 2898–2904.
- [10] C. You and R. Zhang, "Hybrid offline-online design for UAV-enabled data harvesting in probabilistic LoS channel," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3753–3768, Jun. 2020.
- [11] Q. Jiang, Y. Ma, K. Liu, and Z. Dou, "A probabilistic radio map construction scheme for crowdsourcing-based fingerprinting localization," *IEEE Sensors J.*, vol. 16, no. 10, pp. 3764–3774, May 2016.
- [12] Y. Zhang and L. Ma, "Radio map crowdsourcing update method using sparse representation and low rank matrix recovery for WLAN indoor positioning system," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1188–1191, Jun. 2021.
- [13] K. S. Ni and T. Q. Nguyen, "Adaptable K-nearest neighbor for image interpolation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Las Vegas, NV, USA, Mar. 2008, pp. 1297–1300.
- [14] R. Deng, Z. Jiang, S. Zhou, S. Cui, and Z. Niu, "A two-step learning and interpolation method for location-based channel database construction," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Abu Dhabi, United Arab, Dec. 2018, pp. 1–6.
- [15] Z. Han, J. Liao, Q. Qi, H. Sun, and J. Wang, "Radio environment map construction by Kriging algorithm based on mobile crowd sensing," *Wireless Commun. Mobile Comput.*, vol. 2019, pp. 1–12, Feb. 2019.
- [16] H. Braham, S. B. Jemaa, G. Fort, E. Moulines, and B. Sayrac, "Fixed rank Kriging for cellular coverage analysis," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4212–4222, Aug. 2017.
- [17] K. Sato and T. Fujii, "Kriging-based interference power constraint: Integrated design of the radio environment map and transmission power," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 1, pp. 13–25, Mar. 2017.
- [18] V.-P. Chowdappa, C. Botella, J. J. Samper-Zapater, and R. J. Martinez, "Distributed radio map reconstruction for 5G automotive," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 2, pp. 36–49, Apr. 2018.
- [19] U. Masood, H. Farooq, and A. Imran, "A machine learning based 3D propagation model for intelligent future cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.
- [20] A. Massa, D. Marcantonio, X. Chen, M. Li, and M. Salucci, "DNNs as applied to electromagnetics, antennas, and propagation—A review," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, no. 11, pp. 2225–2229, Nov. 2019.
- [21] Y. Teganya and D. Romero, "Deep completion autoencoders for radio map estimation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1710–1724, Mar. 2021.
- [22] R. Levie, C. Yapar, G. Kutyniok, and G. Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 4001–4015, Jun. 2021.
- [23] S. Shrestha, X. Fu, and M. Hong, "Deep spectrum cartography: Completing radio map tensors using learned neural models," *IEEE Trans. Signal Process.*, vol. 70, pp. 1170–1184, 2022.
- [24] C. Fan, X. Zhong, and J. Wei, "BS-to-ground channel reconstruction with 3D obstacle map based on RSS measurements," *IEEE Access*, vol. 7, pp. 99633–99641, 2019.
- [25] O. Esrafilian, R. Gangula, and D. Gesbert, "Map reconstruction in UAV networks via fusion of radio and depth measurements," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2021, pp. 1–6.
- [26] A. Eleryan, M. Elsabagh, and M. Youssef, "Synthetic generation of radio maps for device-free passive localization," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Houston, TX, USA, Dec. 2011, pp. 1–5.
- [27] N. Suga, R. Sasaki, M. Osawa, and T. Furukawa, "Ray tracing acceleration using total variation norm minimization for radio map simulation," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 522–526, Mar. 2020.
- [28] J. Chen, U. Yatnalli, and D. Gesbert, "Learning radio maps for UAV-aided wireless networks: A segmented regression approach," in *Proc. IEEE Int. Conf. Commun.*, Paris, France, May 2017, pp. 1–6.
- [29] J. Chen, O. Esrafilian, D. Gesbert, and U. Mitra, "Efficient algorithms for air-to-ground channel reconstruction in UAV-aided communications," in *Proc. IEEE Global Telecommun. Conf.*, Singapore, Dec. 2017, pp. 1–6.
- [30] B. Zhang and J. Chen, "Constructing radio maps for UAV communications via dynamic resolution virtual obstacle maps," in *Proc. Int. Workshop Signal Process. Adv. Wireless Commun.*, Atlanta, GA, USA, May 2020, pp. 1–5.
- [31] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
- [32] J. Fan, *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability*, vol. 66. Evanston, IL, USA: Routledge, 1996.



Wenjie Liu (Graduate Student Member, IEEE) received the B.Eng. degree in electronic information engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 2019, and the M.Phil. degree in computer and information engineering from The Chinese University of Hong Kong, Shenzhen, in 2021. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology. He was a Visiting Student/an Exchange Student at Peking University and the Harbin Institute of Technology and interned at the Intel Asia-Pacific Research and Development Center. His research interests mainly include radio map, security, and privacy in wireless communication systems.



Juntong Chen (Member, IEEE) received the B.Sc. degree in electronic engineering from Nanjing University, Nanjing, China, in 2009, and the Ph.D. degree in electronic and computer engineering from the Hong Kong University of Science and Technology (HKUST), Hong Kong, SAR, China, in 2015.

From 2014 to 2015, he was a Visiting Student with the Wireless Information and Network Sciences Laboratory, MIT, Cambridge, MA, USA. He is currently an Assistant Professor with the School of Science and Engineering and the Future Network of Intelligence Institute (FNii), The Chinese University of Hong Kong, Shenzhen (CUHK-Shenzhen), Guangdong, China. Prior to joining CUHK-Shenzhen, he was a Post-Doctoral Research Associate with the Ming Hsieh Department of Electrical Engineering, University of Southern California (USC), Los Angeles, CA, USA, from 2016 to 2018, and the Communication Systems Department of EURECOM, Sophia-Antipolis, France, from 2015 to 2016. His research interests include channel estimation, MIMO beamforming, machine learning, and optimization for wireless communications and localization. His current research focuses on radio map sensing, construction, and application for wireless communications. He was a recipient of the HKITIT Post-Graduate Excellence Scholarships in 2012. He was nominated as the Exemplary Reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS in 2018. His paper received the Charles Kao Best Paper Award from WOCC 2022.