Model-Aware UAV Trajectory Planning for Efficient Radio Environment Mapping

Ali Sourchap, Omid Esrafilian, and David Gesbert
Communication Systems Department, EURECOM, Sophia Antipolis, France

***Abstract*—Accurate Radio Environment Mapping (REM) is essential for wireless network optimization and performance analysis. However, conventional data acquisition and REM construction methods face practical challenges, particularly in complex or inaccessible urban areas. Model-driven approaches such as ray tracing (RT) offer a promising alternative but are often limited by unknown electromagnetic (EM) parameters that reduce their accuracy. To overcome this, we propose a model-aware UAV trajectory planning framework for efficient REM generation in challenging environments using limited measurements. The problem is formulated as a joint UAV path planning and REM generation as an inverse modeling problem to collect informative data for inferring unknown EM parameters, thereby improving RT-based REM reconstruction. Simulations in an urban setting demonstrate that the proposed method improves REM estimation accuracy by up to 40% compared with baseline approaches. This work presents the first model-aware UAV trajectory optimization framework for EM parameter learning, enabling accurate and data-efficient REM construction in inaccessible regions.**

I. Introduction

Accurate Radio Environment Maps (REMs) are critical for efficient spectrum utilization, interference management, and wireless network planning. These maps provide a detailed visualization of key spectrum information, such as received signal strength (RSS), channel gain, or Signal-to-Interference-plus-Noise Ratio (SINR), within a specific region. Traditional REM construction relies on distributed ground-level measurements followed by spatial interpolation to estimate radio characteristics at unmeasured locations [1]. However, this approach faces intrinsic limitations: dense sensor deployment is costly, spatial coverage is often sparse, and access to complex urban areas is often restricted, requiring permissions.

In recent years, environment-aware REM estimation has significantly improved prediction accuracy. These methods combine spatial and electromagnetic (EM) information, including 3D urban maps, terrain topology, and transmitter configurations, to build more precise and adaptive REM models that reflect real-world propagation conditions [2]. These data sources enable model-driven methods to reduce dependence on extensive measurements. Standard approaches include empirical path-loss formulations, like 3GPP models [3], and ray-tracing (RT) techniques that simulate wave propagation to predict signal metrics such as RSS and SINR [4]. More recently, hybrid model-based learning frameworks have emerged, combining RT or path-loss models with data-driven learning components. However, these approaches still require training on large measurement datasets. By blending physical modeling and machine learning, these methods improve generalization and robustness under limited measurements, reducing computational cost compared to full RT simulations [5].

When partial environmental data (3D maps and transmitter configurations) are available, RT-based models become particularly appealing for REM generation. They compute metrics like channel gain, RSS, and SINR by tracing propagation paths determined by scene geometry and EM properties [6]. Yet, the accuracy of these predictions depends deeply on the precision of EM parameters. Even with detailed 3D maps, inaccuracies in material characterization or antenna configurations can lead to REM errors, especially in complex urban settings [7].

Recent advances in differentiable RT (Diff-RT) [8] have introduced gradient-based frameworks for learning scene EM characteristics directly from channel measurements. While these methods are theoretically solid and perform well, their accuracy directly depends on the informativeness and spatial diversity of the measurements, which dictates how precisely the underlying EM parameters can be inferred. In practice, achieving adequate spatial resolution is challenging, and this lack of spatial diversity can degrade learning performance.

In recent years, Unmanned Aerial Vehicles (UAVs) have gained attention as flexible and efficient platforms for wireless data acquisition. Unlike traditional methods, UAVs can offer high mobilities that enable channel sensing over wide spatial regions. As a result, UAVs have become an important tool for REM data collection and estimation. Still, their deployment is subject to practical limitations: low-altitude flights in dense urban areas are restricted for safety, and flight time is limited by battery capacity. For these reasons, most existing studies have focused on using UAVs for high-altitude REM interpolation [9] or hybrid airborne–ground data collection for 3D REM construction [10]. In these approaches, UAV measurements are particularly used to estimate the radio environment within the regions directly covered during flight, without addressing REM estimation in unmeasured areas.

This work tackles the challenging problem of constructing accurate REMs using limited UAV measurements collected over hard-to-access regions, such as ground-level urban areas. In this setting, REM construction becomes an inverse modeling task, where unknown EM parameters are inferred from high-altitude measurements and embedded into a ray-tracing model

This work was partially supported by the COBOTPLANET project under the NGI Sargasso funding program, and partially was funded by the HUAWEI France supported Chair on Future Wireless Networks at EURECOM. The work of David Gesbert was partially supported by the 3IA Côte d'Azur Investments (ANR) with reference number ANR-23-IACL-0001.

to reconstruct the REM in inaccessible zones. We address this challenge through a tractable two-stage heuristic that decouples the planning and learning phases. To the best of our knowledge, this is the first model-aware UAV trajectory optimization framework for EM parameter learning, enabling accurate and data-efficient REM generation under UAV power constraints. The main contributions are summarized as:

- A tractable two-stage optimization framework that decouples UAV trajectory planning from EM parameter learning.
- A model-aware trajectory design method using proxy functions to quantify UAV measurement informativeness.
- Model-aided, data-efficient deep learning techniques for REM estimation in urban environments with diverse material types, leveraging Diff-RT enabled backpropagation.

II. Problem Formulation and System Model

As shown in Fig. 1, a UAV collects channel measurements from a base-station Tx over a mission lasting T time steps. The UAV position at step $t \in [0, T]$ is denoted by $\boldsymbol{p}_t = [x_t, y_t, z_0]^\top \in \mathbb{R}^3$, where z_0 is a constant flight altitude. The UAV starts at $\boldsymbol{p}_I$ and finishes at the terminal point $\boldsymbol{p}_F$. Assuming constant velocity and no hovering, we define a discrete action space for the UAV as:

$$\mathcal{A} = \left\{ \overset{\text{north}}{\begin{bmatrix} 0 \\ \Delta d \\ 0 \end{bmatrix}}, \overset{\text{west}}{\begin{bmatrix} -\Delta d \\ 0 \\ 0 \end{bmatrix}}, \overset{\text{south}}{\begin{bmatrix} 0 \\ -\Delta d \\ 0 \end{bmatrix}}, \overset{\text{east}}{\begin{bmatrix} \Delta d \\ 0 \\ 0 \end{bmatrix}} \right\}, \tag{1}$$

where $\Delta d > 0$ is the step size. Given the fixed velocity assumption, we simplify the battery constraint by fixing the total number of planning steps to T. Given action $\boldsymbol{a}_t \in \mathcal{A}$, the next UAV position evolves as $\boldsymbol{p}_{t+1} = \boldsymbol{p}_t + \boldsymbol{a}_t$. By having this action space, the UAV's reachable positions create a discrete set $\mathcal{P}$ at the constant altitude z_0. At each position $\boldsymbol{p}_t \in \mathcal{P}$, the UAV records a channel measurement $\boldsymbol{h}_{\boldsymbol{p}_t}$, from Tx, forming a dataset $\{\boldsymbol{h}_{\boldsymbol{p}_t}\}_{t=1}^{T}$that serves as the basis for REM learning.

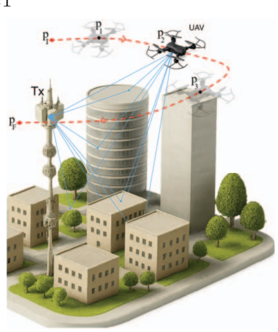


Fig. 1: Efficient radio mapping of inaccessible regions using optimized UAV trajectory measurements

The primary objective is leveraging UAV-collected channel measurements to accurately estimate the REM over a target space (e.g., a ground-level surface). To make this problem computable, we discretized the target space with a per-axis resolution $[\Delta x, \Delta y, \Delta z]$ into a grid $\mathcal{G} \triangleq \{\boldsymbol{p}_v = (x_v, y_v, z_v)^\top \in \mathbb{R}^3 \mid v = [1, V]\}$. We then define the REM as the set of channel impulse responses (CIRs) across this grid, $\Gamma_{\mathcal{G}} \triangleq \{\mathbf{h}_{\boldsymbol{p}_v} \in \mathbb{C}^{N_{\mathrm{S}}} \mid v = [1, V]\}$, where N_{S} is the number of OFDM subcarriers. This general, CIR-based definition allows the later derivation of metrics such as path-gain or RSS.

Assuming that the urban environment consists of M material types [8], for j-th material, $\bar{\boldsymbol{m}}_j = (\varepsilon_r, \sigma, \mathscr{S}, \mathscr{K}, f_s)$ represents relative permittivity, conductivity, scattering coefficient, cross-polarization discrimination, and scattering pattern, respectively. The whole urban environment material catalog is $\mathcal{M} = \{\bar{\boldsymbol{m}}_1, \ldots, \bar{\boldsymbol{m}}_D\}$.

At a measurement point $\boldsymbol{p}_t$, the CIR $\boldsymbol{h}_{\boldsymbol{p}_t}$ can be modeled as the sum of K multi-path components (MPCs) as follows:

$$h(\tau) = \sum_{k=1}^{K} \alpha_k \delta(\tau - \tau_k). \tag{2}$$

In a ray-tracing propagation model, each component (ray) corresponds to a specific propagation path from the transmitter, Tx, to the receiver (i.e. UAV). α_k and τ_k are, respectively, the complex amplitude and path delay of the k-th multipath component. τ_k is proportional to the ray's total length.

If the UAV is equipped with an isotropic antenna, then $\alpha_k \propto \mathbf{T}_k \mathbf{C}^{\mathrm{Tx}}(\theta_k, \varphi_k)$, where, θ_k and φ_k are the zenith and azimuth angles, representing the Angle of Departure (AoD) of the rays. $\mathbf{C}^{\mathrm{Tx}}$ is the transmitter antenna pattern, which determines the gain applied to a transmitted ray as a function of its AoD. $\mathbf{T}_k$ is the transfer matrix corresponding to the k-th ray, and it represents the path's gain and phase changes. It captures all propagation effects, including free-space path loss, signal power degradation due to interaction (e.g., scattering or reflection) with objects along the propagation path, and the incurred phase shifts. At each interaction, the phase change and loss applied to the ray depend on the interaction type, path geometry, and, most importantly, the EM properties $\bar{\boldsymbol{m}}$ of the material it hits (see Appendix A in [6] for details).

A. Model-Based REM Generation

Given a high-fidelity 3D map, the precision of REM generation via ray-tracing depends on the accuracy of the radio propagation parameters. Let's denote the complete parameter set for generating the REM for the entire urban area as $\boldsymbol{\Omega}$. For a specific target grid $\mathcal{G}$, the REM $\Gamma_{\mathcal{G}}$ is impacted by a relevant subset of these parameters, $\boldsymbol{\Omega}_{\mathcal{G}} \subseteq \boldsymbol{\Omega}$. This subset includes the transmitter antenna pattern $\mathbf{C}^{\mathrm{Tx}}(\theta, \varphi)$ and the specific set of materials $\mathcal{M}_{\mathcal{G}}$ that interact with propagation paths reaching the grid: $\boldsymbol{\Omega}_{\mathcal{G}} \triangleq \{\mathcal{M}_{\mathcal{G}}, \mathbf{C}^{\mathrm{Tx}}(\theta, \varphi)\}$. By this definition, we assume that other influential parameters, such as antenna orientation and placement, are known and encoded in the 3D map. However, if these parameters are uncertain, they may also be included in $\boldsymbol{\Omega}_{\mathcal{G}}$ and inferred via differentiable ray-tracing techniques. Given the 3D map and $\boldsymbol{\Omega}_{\mathcal{G}}$, the estimated REM, $\widehat{\Gamma}_{\mathcal{G}}$, is produced by ray tracing to compute propagation paths and the corresponding CIRs:

$$\widehat{\Gamma}_{\mathcal{G}}(\boldsymbol{\Omega}_{\mathcal{G}}) = \mathrm{RT}(\text{3D Map}, \boldsymbol{\Omega}_{\mathcal{G}}). \tag{3}$$

In this paper, our goal is to minimize the difference between the estimated and ground-truth REMs. Thus, the REM learning problem can be formulated as:

$$\min_{\boldsymbol{\Omega}_{\mathcal{G}}} \ \frac{1}{V}\sum_{v\in\mathcal{G}} \mathcal{L}\Big(\boldsymbol{h}_{\boldsymbol{p}_v}, \widehat{\boldsymbol{h}}_{\boldsymbol{p}_v}(\boldsymbol{\Omega}_{\mathcal{G}})\Big), \tag{4}$$

where $\mathcal{L}(\boldsymbol{h}, \widehat{\boldsymbol{h}})$ is the per-sample loss function. This function emphasizes two key parameters of the CIR: the total received power, P, and the RMS delay spread, τ_{rms}, (standard definitions in [11]). To quantify the similarity between the actual and simulated CIRs, the loss function $\mathcal{L}$ combines the Symmetric Mean Absolute Percentage Error (SMAPE) of P and τ_{rms}, as:

$$\mathcal{L}(\boldsymbol{h}, \widehat{\boldsymbol{h}}) = \underbrace{\left|\tfrac{\widehat{P}-P}{\widehat{P}+P}\right|}_{\mathcal{L}_{\text{power}}} + \underbrace{\left|\tfrac{\widehat{\tau}_{\text{rms}}-\tau_{\text{rms}}}{\widehat{\tau}_{\text{rms}}+\tau_{\text{rms}}}\right|}_{\mathcal{L}_{\text{delay}}}, \tag{5}$$

The SMAPE terms, $\mathcal{L}_{\text{power}}$ and $\mathcal{L}_{\text{delay}}$, are symmetric, scale-invariant, and bounded in the $[0, 1]$ range. Consequently, the $\mathcal{L}(\boldsymbol{h}, \widehat{\boldsymbol{h}})$ inherits these features: it is bounded in the $[0, 2]$ and provides equal weighting for strong and weak channel gains.

The optimization (4) constitutes an *inverse modeling* problem, where the unknown ray-tracing parameters $\boldsymbol{\Omega}_{\mathcal{G}}$ should be learned from captured channel measurements $\{\boldsymbol{h}_{\boldsymbol{p}_t}\}_{t=1}^{T}$. The studied case in [8] uses measurements collected directly within the target grid $\mathcal{G}$. This contrasts with our problem, which is restricted to UAV-collected measurements along a trajectory that is spatially separate from the target grid (gathering high-altitude measurements to estimate a REM for a grid closer to the ground). Thus, the UAV trajectory must be carefully planned to collect the most informative measurements for learning $\boldsymbol{\Omega}_{\mathcal{G}}$. This challenge makes it necessary to jointly optimize (4) for both path planning and parameter estimation.

B. Joint Optimization Problem

Based on the prior discussions, the optimization in (4) must be reformulated to jointly optimize for both trajectory planning, by taking into account the power constraints, and parameter estimation. This leads to the following joint optimization problem:

$$\begin{aligned} \min_{\{\boldsymbol{a}_t\}_{t=1}^{T}, \boldsymbol{\Omega}_{\mathcal{G}}} \quad & \frac{1}{V}\sum_{v\in\mathcal{G}} \mathcal{L}\Big(\boldsymbol{h}_{\boldsymbol{p}_v}, \widehat{\boldsymbol{h}}_{\boldsymbol{p}_v}(\boldsymbol{\Omega}_{\mathcal{G}})\Big). \\ \text{s.t.} \quad & \boldsymbol{p}_{t+1} = \boldsymbol{p}_t + \boldsymbol{a}_t, \quad t = 1, \ldots, T, \\ & \boldsymbol{p}_1 = \boldsymbol{p}_I, \quad \boldsymbol{p}_T = \boldsymbol{p}_F. \end{aligned} \tag{6}$$

The main challenge in solving (6) is the need to compute or estimate the loss function $\mathcal{L}$ (i.e., finding a lower or upper bound) over the entire grid $\mathcal{G}$. For instance, estimating the loss function using traditional analytical methods (e.g., Fisher information) is intractable, as it would require a closed-form distribution model of the parameters $\boldsymbol{\Omega}_{\mathcal{G}}$, which is not available in our case. On the other hand, numerical approaches, such as Monte-Carlo methods or PCE [12], are also impractical because they would require extensive sampling iterations between measurement collection and ray tracing, resulting in significant computational complexity.

For these reasons, we propose breaking down (6) into a two-stage optimization approach. The first stage involves proposing a model-aware offline trajectory planning, and the second stage executes parameter learning for $\boldsymbol{\Omega}_{\mathcal{G}}$ via Diff-RT. This allows for a simpler optimization while preserving the core goal of minimizing the discrepancy between $\Gamma_{\mathcal{G}}$ and its estimate $\widehat{\Gamma}_{\mathcal{G}}$.

III. Two-Stage Heuristic Optimization

Our proposed Heuristic Optimization consists of two stages: **Stage 1: Model-Aware Trajectory Planning** and **Stage 2: Parameter Learning via Differentiable Ray-Tracing**. In Stage 1, we propose an offline trajectory planning method based on model-aware proxy functions. These functions are introduced to quantify the expected information of potential measurements, enabling the planner to design a trajectory that gathers high-quality and sufficient data for the learning process. Afterward, the collected measurements are used in Stage 2 by employing the Diff-RT to learn the parameters, and then to estimate the REM.

A. Model-Aware Trajectory Planning

In this section, we propose an offline trajectory design aimed at capturing the most informative measurements for learning the unknown REM parameters. The main challenge lies in quantifying the informativeness of each measurement. However, deriving a closed-form analytical metric would require prior knowledge of the parameter distribution, which is unavailable and therefore infeasible. To address this, we introduce a proxy reward function that estimates measurement informativeness, reformulating the problem as a heuristic reward maximization task. We refer to this as a model-aware trajectory since it leverages the RT-based propagation model to evaluate measurement quality. This approach uses a prior parameter set, $\boldsymbol{\Omega}_0 \triangleq \{\mathcal{M}_0, \mathbf{C}_0^{\text{Tx}}(\theta, \varphi)\}$, to perform initial RT simulations for candidate measurements $\boldsymbol{p}_t \in \mathcal{P}$, providing a baseline for trajectory evaluation.

First, let's define the total reward, comprising two terms: i) the Accumulated Material Reward (AMR), and ii) the Accumulated Antenna Reward (AAR). The general form of the total reward function is:

$$R_{1:T} = \lambda_1 R_{1:T}^{\mathcal{M}} + \lambda_2 R_{1:T}^{\text{A}}, \tag{7}$$

where $R_{1:T}^{\mathcal{M}}$ is AMR, proxying how informative the collected measurements are for learning $\mathcal{M}_{\mathcal{G}}$. Similarly, $R_{1:T}^{\text{A}}$ is AAR, proxying how well the trajectory supports learning $\mathbf{C}^{\text{Tx}}(\theta, \varphi)$. The weights $\lambda_1, \lambda_2 > 0$ control the tradeoff between these objectives. The reward terms are detailed below.

1) Accumulated Material Reward: For the material set $\mathcal{M}_{\mathcal{G}}$ (containing $M_{\mathcal{G}}$ material types) the collected data should be sufficiently diverse to provide information about ideally all materials. At the same time, the quality of the information for each material must be high enough to ensure that each parameter set $\bar{\boldsymbol{m}}_j$ can be reliably learned during Stage 2.

To assess the quality of the measurements for learning the materials, we introduce the Material Contribution Ratio (MCR) coefficient, allowing us to quantify the impact of each material in the collected measurements. For each measurement collected at $\boldsymbol{p}_t$ and material $j \in [1, M_{\mathcal{G}}]$, let $\mathcal{J}_j$ denote the set of propagation rays that interact at least once with the j-th material. Assuming that after performing RT, the CIR obtained at position $\boldsymbol{p}_t$ consists of K rays, the MCR is then defined as:

$$\rho_j(\boldsymbol{p}_t) \triangleq \frac{\sum_{k\in\mathcal{J}_j}|\alpha_k|}{\sum_{k=1}^{K}|\alpha_k|}, \tag{8}$$

where α_k are the complex coefficients defined in (2). Note, a higher $\rho_j(\boldsymbol{p}_t)$ means a higher contribution of j-th material. The Accumulated MCR up to time τ for material j is:

$$\mathcal{S}^{\mathcal{M}}_{j,1:\tau} = \sum_{t=1}^{\tau}\rho_j(\boldsymbol{p}_t). \tag{9}$$

For the whole trajectory (at $\tau = T$), the total accumulated MCR for material j is denoted by $\mathcal{S}^{\mathcal{M}}_{j,1:T}$.

From an information-theoretic perspective, $\mathcal{S}^{\mathcal{M}}_{j,1:\tau}$ serves as a proxy vector for information gain, meaning that a higher $\mathcal{S}^{\mathcal{M}}_{j,1:\tau}$ leads to a higher mutual information between the measurements and the j-th material parameter, $I(\bar{\boldsymbol{m}}_j;\{\boldsymbol{h}_t\}_{t=1}^{T})$ [13]. Therefore, we define the AMR by summing over the MCRs for all materials over the trajectory as follows:

$$R^{\mathcal{M}}_{1:T} = \sum_{j=1}^{M_{\mathcal{G}}}\mathcal{S}^{\mathcal{M}}_{j,1:T}. \tag{10}$$

Maximizing $R^{\mathcal{M}}_{1:T}$ encourages trajectories to collect the most overall information, serving as a proxy to maximize the total gathered mutual information.

Empirically, we observe that when $\mathcal{S}^{\mathcal{M}}_{j,1:\tau}$ exceeds a threshold ($\mathcal{S}^{\mathcal{M}}_{j,1:\tau} > \Lambda_M$), Diff-RT achieves reliable learning performance for the corresponding material parameters, $\bar{\boldsymbol{m}}_j$. We implement this concept by defining a capped information vector, $\hat{\mathcal{S}}^{\mathcal{M}}_{j,1:\tau}$, where each element is defined as: $\hat{\mathcal{S}}^{\mathcal{M}}_{j,1:\tau} = \min\left(\mathcal{S}^{\mathcal{M}}_{j,1:\tau},\Lambda_M\right)$. This approach indicates that once a material has surpassed the information threshold, it has likely provided the required information for parameter learning $\bar{\boldsymbol{m}}_j$, encouraging the planner to seek information from other materials that have not yet met the threshold. We can then rewrite (10) as:

$$R^{\mathcal{M}}_{1:T} = \sum_{j=1}^{M_{\mathcal{G}}}\hat{\mathcal{S}}^{\mathcal{M}}_{j,1:T}. \tag{11}$$

However, simply maximizing (11) may lead to unbalanced exploration; we employ an entropy term to encourage balanced measurement collection. This term serves as a proxy for maximizing the minimum mutual information achieved for any single material. We denote $H^{\mathcal{M}}_{1:T}$ as the entropy of the capped accumulated information:

$$H^{\mathcal{M}}_{1:T} = -\sum_{j=1}^{M_{\mathcal{G}}}\tilde{s}_j\log\tilde{s}_j \quad \text{,where} \quad \tilde{s}_j = \frac{\hat{\mathcal{S}}^{\mathcal{M}}_{j,1:T}}{\sum_{k=1}^{M_{\mathcal{G}}}\hat{\mathcal{S}}^{\mathcal{M}}_{k,1:T}}. \tag{12}$$

The Entropy reaches its maximum when the accumulated MCRs in (9) is uniformly distributed, encouraging trajectories that provide balanced coverage of all materials. By including the entropy term, the final form of AMR is:

$$R^{\mathcal{M}}_{1:T} = \lambda_{11}\sum_{j=1}^{M_{\mathcal{G}}}\hat{\mathcal{S}}^{\mathcal{M}}_{j,1:T} + \lambda_{12}H^{\mathcal{M}}_{1:T}, \tag{13}$$

where the weights $\lambda_{11},\lambda_{12} > 0$ regulate the trade-off between measurement quality and diversity, providing a tractable heuristic for learning the parameters of $\mathcal{M}_{\mathcal{G}}$.

2) Accumulated Antenna Reward: Akin to the AMR, we define the Accumulated Antenna Reward (AAR), $R^{\mathrm{A}}_{1:T}$, for learning the antenna pattern $\mathbf{C}^{\mathrm{Tx}}(\theta,\varphi)$. The objective is to plan a trajectory that collects measurements containing rays with a diverse range of AoD, providing sufficient samples from different antenna directions to ensure accurate pattern learning. In our setting, where the UAV operates at a fixed altitude, controlling the zenith angle (θ) variation is challenging. Thus, we focus on designing trajectories that capture rays transmitted across a broad range of azimuth angles (φ), while maintaining quality rays, i.e. with power levels distinguishable from noise.

To evaluate the quality of measurements for antenna pattern learning, similar to MCR, we define the Azimuth Contribution Ratio (ACR) which quantifies the contribution of each ray within the collected measurements. To maintain tractability, the azimuth domain is discretized into N_φ intervals. Consider a measurement collected at $\boldsymbol{p}_t$ with K total rays. Let $\mathcal{Z}_i$ denote the set of rays whose departure azimuths fall within interval $i\in[1,N_\varphi]$. The contribution of interval i to the measurement at $\boldsymbol{p}_t$ is then given by the ACR as follows:

$$\upsilon_i(\boldsymbol{p}_t) \triangleq \frac{\sum_{k\in\mathcal{Z}_i\setminus\mathcal{Z}_{\mathrm{LoS}}}|\alpha_k|}{\sum_{k=1}^{K}|\alpha_k|}, \tag{14}$$

where α_k is the complex coefficients in (2), and $\mathcal{Z}_{\mathrm{LoS}}$ denotes the line-of-sight (LoS) paths, which are excluded from $\mathcal{Z}_i$ to prevent the strong LoS dominance. Since the prior antenna pattern, $\mathbf{C}^{\mathrm{Tx}}_0(\theta,\varphi)$, is isotropic, line-of-sight rays yield high channel gains that overshadow the non-LoS components. Excluding LoS ray highlights the reflected and scattered rays, providing richer angular diversity for antenna pattern learning. This exclusion would be less critical if the prior pattern were more directional. Consequently, the cumulative ACR up to time τ for interval i is given by $\mathcal{S}^{\mathrm{A}}_{i,1:\tau} = \sum_{t=1}^{\tau}\upsilon_i(\boldsymbol{p}_t)$.

Similar to MCR, the ACR will help to gather high-quality measurements. We also introduce an entropy terms (analogous to (12)), promoting diverse ray sampling from the antenna pattern. Finally, the AAR is defined as follows:

$$R^{\mathrm{A}}_{1:T} = \lambda_{21}\sum_{i=1}^{N_\varphi}\hat{\mathcal{S}}^{\mathrm{A}}_{i,1:T} + \lambda_{22}H^{\mathrm{A}}_{1:T}, \tag{15}$$

where $\hat{\mathcal{S}}^{\mathrm{A}}_{i,1:T}$ is defined similarly to $\hat{\mathcal{S}}^{\mathcal{M}}_{i,1:T}$ by applying a threshold Λ_A to $\mathcal{S}^{\mathrm{A}}_{i,1:T}$, and $H^{\mathrm{A}}_{1:T}$ is the entropy term, defined as in (12) over azimuth intervals $i\in[1,N_\varphi]$. The weights $\lambda_{21},\lambda_{22} > 0$ balance measurement quality and diversity.

Having defined both AMR and AAR, we can now form our model-aware trajectory optimization problem as follows:

$$\begin{aligned} \max_{\{\boldsymbol{a}_t\}_{t=1}^{T}} \quad & R_{1:T} = \lambda_1 R^{\mathcal{M}}_{1:T} + \lambda_2 R^{\mathrm{A}}_{1:T} \\ \text{s.t.} \quad & \boldsymbol{p}_{t+1} = \boldsymbol{p}_t + \boldsymbol{a}_t, \quad t = 1,\ldots,T, \\ & \boldsymbol{p}_1 = \boldsymbol{p}_I, \quad \boldsymbol{p}_T = \boldsymbol{p}_F. \end{aligned} \tag{16}$$

Since we assume prior parameters $\boldsymbol{\Omega}_0$, the reward term $R_{1:T}$ can be computed offline, allowing problem (16) to also be solved offline. To address problem (16), which represents a discrete sequential decision-making task, we employ an N-greedy search strategy to efficiently explore the action space.[1] At each decision step t, the UAV evaluates every action $\boldsymbol{a}_t \in \mathcal{A}$ and greedily selects the action maximizing the cumulative reward over the next N steps. The optimal action $\boldsymbol{a}_t^*$ is:

$$\boldsymbol{a}_t^* = \arg\max_{\boldsymbol{a}_t \in \mathcal{A}} \max_{\substack{\{\boldsymbol{a}_{t+1},\ldots,\boldsymbol{a}_{t+N-1}\} \\ \subseteq \mathcal{A}}} \sum_{n=0}^{N-1} r_{t+n}(\boldsymbol{a}_{t+n}), \quad r_{t+n} = R_{1:t+n} - R_{1:t+n-1} \tag{17}$$

Here, r_{t+n} denotes the immediate reward at step $t+n$, representing the gain in cumulative reward. The parameter N governs the trade-off between optimality and computational cost: larger N offers greater foresight but increases complexity. Additionally, a *Safe-Return* mechanism activates when the remaining steps are just enough to reach the final point $\boldsymbol{p}_F$. Once triggered, the UAV selects the return path that maximizes cumulative reward while ensuring arrival at $\boldsymbol{p}_F$ by $t \leq T$.

B. Parameter Learning via Differentiable Ray-Tracing

After trajectory execution, the collected measurements $\{\boldsymbol{h}_{\boldsymbol{p}_t}\}_{t=1}^T$ are used to learn the radio parameters $\boldsymbol{\Omega}_{\mathcal{G}}$. We use differentiable ray-tracing Framework [8] to model and learn the channel. The transmitter antenna pattern is represented as a mixture of χ spherical Gaussians:

$$\mathbf{C}^{\text{Tx}}(\theta, \varphi; \boldsymbol{\Theta}_A) = \eta_{\text{rad}} \sum_{i=1}^{\chi} \frac{\omega_i}{\beta_i} \exp\big(\lambda_i \boldsymbol{\mu}_i^\top \hat{\boldsymbol{r}}(\theta, \varphi)\big), \tag{18}$$

where $\beta_i = \frac{2\pi}{\lambda_i}(e^{\lambda_i} - e^{-\lambda_i})$ and $\hat{\boldsymbol{r}}(\theta, \varphi)$ is the unit direction vector in spherical coordinate. The Antenna learnable parameter set is denoted as $\boldsymbol{\Theta}_A = \{\eta_{\text{rad}}, \{\omega_i, \lambda_i, \boldsymbol{\mu}_i\}_{i=1}^{\chi}\}$, allowing us to have a differentiable model of the antenna patterns.

To simplify the learning process, we assume all materials have the same scattering pattern (f_s) [8], which is determined a priori [7]. The remaining EM parameters for each material $\bar{\boldsymbol{m}}_j$ to learn are $\{\varepsilon_{r_j}, \sigma_j, \mathscr{S}_j, \mathscr{K}_j\}$. The Over-Parameterization (OP) method from [8] learns material variables from a moderately small shared latent space. Although this OP works for a small set of materials, the shared coupling is not well-suited to urban environments with many diverse materials, where it tends to limit flexibility and slow the convergence of per-material variables. To address this, our approach uses an MLP network, g_{mat} (weights $\boldsymbol{\Theta}_M$), that takes a one-hot vector for material index j as input to produce its EM parameter:

$$g_{\text{mat}}(j; \boldsymbol{\Theta}_M) = \{\varepsilon_{r_j}, \sigma_j, \mathscr{S}_j, \mathscr{K}_j\}. \tag{19}$$

This approach decouples materials, increases their expressivity, and yields faster and more reliable parameter convergence.

Finally, the trainable expression for the j-th material's parameters is given by $\bar{\boldsymbol{m}}_j(\boldsymbol{\Theta}_M) = \{g_{\text{mat}}(j; \boldsymbol{\Theta}_M), f_s\}$. The parameters $\boldsymbol{\Theta}_A$ and $\boldsymbol{\Theta}_M$ are found by optimizing the following problem over the collected measurements:

$$\boldsymbol{\Theta}_A^*, \boldsymbol{\Theta}_M^* : \arg\min_{\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_M} \sum_{t=1}^{T} \mathcal{L}\Big(\boldsymbol{h}_{\boldsymbol{p}_t}, \widehat{\boldsymbol{h}}_{\boldsymbol{p}_t}(\boldsymbol{\Theta}_A, \boldsymbol{\Theta}_M)\Big). \tag{20}$$

Upon convergence, the learned parameters $\boldsymbol{\Omega}_{\mathcal{G}}^* = \{\mathcal{M}_{\mathcal{G}}^*, \mathbf{C}^{\text{Tx}}(\theta, \varphi; \boldsymbol{\Theta}_A^*)\}$ are composed, using to generate the final REM over the target grid: $\widehat{\Gamma}_{\mathcal{G}} = \text{RT}(\text{3D Map}, \boldsymbol{\Omega}_{\mathcal{G}}^*)$, where $\mathcal{M}_{\mathcal{G}}^* = \{\bar{\boldsymbol{m}}_1(\boldsymbol{\Theta}_M^*), \cdots, \bar{\boldsymbol{m}}_{M_{\mathcal{G}}}(\boldsymbol{\Theta}_M^*)\}$.

IV. Experiments

We evaluate our method using Sionna v0.19.2 [14] with a high-fidelity 3D urban map of Birmingham city from AccuCities [15]. The map contains 14 material types (walls, rooftops, terrain, asphalt, concrete, etc.). Ground truth EM parameters are assigned based on established ranges in [16]. The target REM grid $\mathcal{G}$ is located 1.5m above ground, discretized with granularity $[\Delta x, \Delta y] = [0.45, 0.45]$ m. The transmitter uses a 2×2 antenna array with 0.5λ spacing and a vertical 3GPP TR 38.901 pattern, operating at 3.8 GHz with $N_S = 1024$ subcarriers. It is located at an altitude of 15 m, with its position shown in Figs. 2 and 3(a) . All materials share the same scattering pattern f_s from [17] with parameters $\alpha_r = 5$, $\alpha_i = 8$, $\lambda = 0.8$. Fig. 2 shows the urban scene structure and the corresponding true path gain radio map. The UAV flies at an altitude of $z_0 = 22$ m, with candidate locations $\mathcal{P}$ forming a grid with $\Delta d = 10$ m spacing. The initial location $\boldsymbol{p}_I$ is a random point in front of the antenna's main lobe. The trajectory is a loop ($\boldsymbol{p}_I = \boldsymbol{p}_F$) limited to $T = 40$ steps.

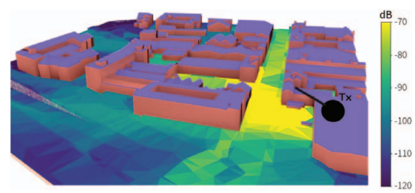
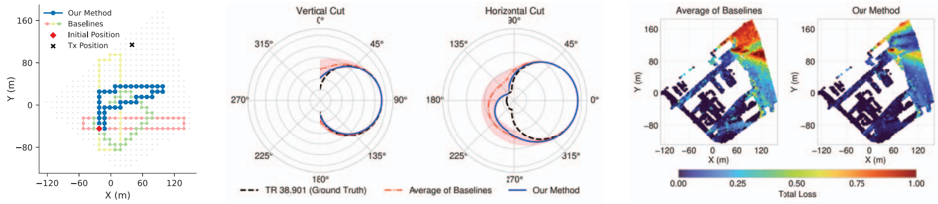


Fig. 2: 3D urban and path gain maps (dB) at 1.5 m above ground, as well as the antenna and its direction.

The prior $\boldsymbol{\Omega}_0$ assumes a simple isotropic pattern for $\mathbf{C}_0^{\text{Tx}}(\theta, \varphi)$ and sets all material parameters to $\{\varepsilon_r = 3, \sigma = 0.026, \mathscr{S} = 0.3, \mathscr{K} = 0.3\}$. We set the horizon of greedy planner $N = 6$. The main reward weights are $\lambda_1 = 0.25$ and $\lambda_2 = 0.75$, prioritizing antenna pattern (C^{Tx}) learning, with uniform gain-entropy weights ($\lambda_{ij} = 0.5$). For the learning stage, the antenna pattern is modeled with $\chi = 3$ spherical Gaussians. The material parameters are learned by an MLP with $[192 \times 96]$ hidden layers and Swish activation.

We compare our model-based method against six arbitrary baselines, which include fixed rectangular shapes (vertical, horizontal, and diagonal) and region-specific paths. Fig. 3(a) illustrates the optimized trajectory alongside three examples of the baseline trajectories. The optimized trajectory is planned to travel to different parts of the map, while gathering measurements from a wide range of ray azimuths. After the stage 2 learning process, using $T = 40$ measurements ($\{\boldsymbol{h}_{\boldsymbol{p}_t}\}_{t=1}^T$)

[1] Alternatively, a dynamic programming approach is possible but intractable, suffering the curse of dimensionality from the exponential-size state space.



(a) (b) (c)

Fig. 3: Trajectory and Learning results. From left to right: (a) Optimized trajectory vs. baselines. (b) Learned antenna pattern (vertical/horizontal cuts). (c) Per-point REM loss $\mathcal{L}(\cdot)$ over the target grid $\mathcal{G}$.

collected along each trajectory, we compared our method against the average performance of baselines. For $\mathcal{M}_{\mathcal{G}}^*$ parameter learning, our obtained trajectory achieved a relative error of 182.2%, significantly outperforming the baseline average of 368.3%. [2] In addition, we applied the OP method from [8] to the measurements collected along all trajectories. The resulting relative error was 3961.78%, indicating this over-parametrization is unsuitable for learning $\mathcal{M}_{\mathcal{G}}^*$ in our system. Fig. 3(b) shows our method estimated more accurately the transmitter antenna pattern $\mathbf{C}^{\mathrm{Tx}}(\theta, \varphi; \mathbf{\Theta}_A^*)$. The improvement is most significant for the antenna's rear lobes, which were poorly exposed to the UAV's starting position.

Consequently, the superior $\mathbf{\Omega}_{\mathcal{G}}^*$ estimation yields a more accurate REM. As illustrated in Fig. 3(c), our method achieves a lower per-point loss, $\mathcal{L}(\boldsymbol{h}_{\boldsymbol{p}_v}, \widehat{\boldsymbol{h}}_{\boldsymbol{p}_v})$, across the target grid $\mathcal{G}$, performing particularly well behind the antenna's main lobe. Quantitatively, our average grid loss (4) is 0.12 ± 0.18 versus the baseline's 0.26 ± 0.31, more than twofold enhancement. The average absolute logarithmic error, $\left|\bar{P}_{db} - \widehat{\bar{P}}_{db}\right|$, is also reduced to 0.69 ± 1.08 dB from the baseline's 1.72 ± 2.08 dB.

To generalize our findings, we conducted a Monte-Carlo simulation over random loop trajectories with random initial points. Fig. 4 plots the average CDF of the per-point loss in the estimated REM, $\mathcal{L}\left(\boldsymbol{h}_{\boldsymbol{p}_v}, \widehat{\boldsymbol{h}}_{\boldsymbol{p}_v}\right)$, illustrating the distribution of estimation errors across grid points. Fig. 4 confirms our method's clear out-performance, with its CDF curve consistently above the baseline's. This is quantified by the 95th-percentile grid loss $\mathcal{L}$ of 0.67 (vs. 1.18 for the baseline, $\sim 40\%$ reduction). Similarly, the improvement holds for the channel-gain average absolute logarithmic error, with our 95th percentile at 4.34 dB versus the baseline's 9.22 dB ($\sim$ 5 dB reduction). Moreover, the plot shows that using OP-material training [8] on the optimized trajectory still performs poorly, with a 95th percentile similar to the baseline.

V. Conclusion

In this paper, we proposed a model-aware UAV trajectory planning framework for efficient REM construction in hard-to-reach regions. The proposed two-stage approach combines an offline trajectory planner with Diff-RT to infer key EM parameters from UAV measurements. Simulation results show notable estimation accuracy gains over baselines. This work provides a data-efficient framework leveraging UAVs and RT-propagation models to map inaccessible radio environments.

[2]The high relative errors are expected, as conductivity σ, one of the per-material EM parameters, spans a wide theoretical range of $[10^{-6}, 10^{+6}]$ S/m.

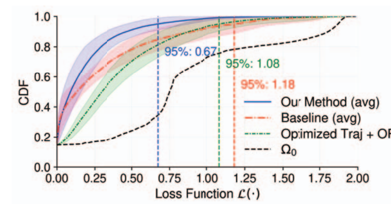


Fig. 4: Average CDF of the loss function. The $\mathbf{\Omega}_0$ curve represents the loss obtained using only prior parameters.

References

[1] M. Pesko *et al.*, "Radio environment maps: The survey of construction methods," *KSII Transactions on Internet and Information Systems*, 2014.
[2] B. Feng *et al.*, "A recent survey on radio map estimation methods for wireless networks," *Electronics*, 2025.
[3] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz," Valbonne, France, Tech. Rep. TR 38.901, 2024.
[4] S. Bakirtzis *et al.*, "Radio propagation modelling: To differentiate or to deep learn, that is the question," 2025.
[5] V. V. Ratnam *et al.*, "Fadenet: Deep learning-based mm-wave large-scale channel fading prediction and its applications," *IEEE Access*, 2021.
[6] F. A. Aoudia *et al.*, "Sionna rt: Technical report," 2025.
[7] A. W. Mbugua *et al.*, "Review on ray tracing channel simulation accuracy in sub-6 GHz outdoor deployment scenarios," *IEEE Open Journal of Antennas and Propagation*, 2021.
[8] J. Hoydis *et al.*, "Learning radio environments by differentiable ray tracing," 2023.
[9] A. Ivanov *et al.*, "Limited sampling spatial interpolation evaluation for 3D radio environment mapping," *Sensors*, 2023.
[10] B. S. Shawel *et al.*, "A deep-learning approach to a volumetric radio environment map construction for uav-assisted networks," *International Journal of Antennas and Propagation*, 2024.
[11] M. Varela and M. Sanchez, "Rms delay and coherence bandwidth measurements in indoor radio channels in the UHF band," *IEEE Transactions on Vehicular Technology*, 2001.
[12] A. C. M. Austin *et al.*, "Application of polynomial chaos to quantify uncertainty in deterministic channel models," *IEEE Transactions on Antennas and Propagation*, 2013.
[13] Y. Satsangi *et al.*, "Maximizing information gain in partially observable environments via prediction reward," 2020.
[14] J. Hoydis *et al.*, "Sionna," 2022.
[15] Accucities, "3D model of Birmingham," 2024. [Online]. Available: https://accucities.com/3d-model-of-birmingham/
[16] S. J. P. Ng, "Investigating the EM performance of building materials for mid-band 5G in urban built environment: An exploratory study," Ph.D. dissertation, Nanyang Technological University, 2023.
[17] V. Degli-Esposti *et al.*, "Measurement and modelling of scattering from buildings," *IEEE Transactions on Antennas and Propagation*, 2007.