# Adaptive Sampling for Online UAV-Aided Radio Map Sensing

Na Zhang‡, Rui Zhou‡, Ming-Yi You⋆, Wenqiang Pu‡
‡Shenzhen Research Institute of Big Data, The Chinese University of Hong Kong, Shenzhen, China
⋆National Key Laboratory of Electromagnetic Space Security, No.36 Research Institute of CETC, China

***Abstract*—Unmanned Aerial Vehicle (UAV)-assisted radio map reconstruction offers a flexible and efficient alternative to labor-intensive process of manual measurements for communication and sensing tasks in next-generation wireless networks. However, the limited flight duration of UAVs necessitates an efficient and adaptive sampling strategy to maximize the value of limited measurements. In this paper, we propose an adaptive sampling scheme that integrates Gaussian Process Upper Confidence Bound (GP-UCB) with cross-entropy (CE) optimization to dynamically balance the exploration-exploitation trade-off during radio map reconstruction and emitter localization. To further enhance trajectory optimization and sampling efficiency, gradient ascent is employed within the CE framework. Unlike conventional methods that focus primarily on uncertainty-based sampling, our approach simultaneously reconstructs the radio map and localizes emitters, ensuring accurate representation of the radio environment. Numerical experiments conducted in simulated radio environments demonstrate that the proposed scheme effectively addresses the dual objectives of emitter localization and radio map exploration, achieving reconstructions with reduced sampling time.**

***Index Terms*—Radio Map, Adaptive sampling, Motion planning, UAV, Cross-Entropy, Gradient ascent.**

## I. Introduction

Radio maps [1]–[3] spatially characterize radio frequency (RF) signal distributions such as received signal strength (RSS) and interference levels, serve as a foundation for wireless network planning, spectrum management, location-based services etc. Radio map reconstruction aims to infer the radio environment of the area of interest from sparse measurements by leveraging data-driven techniques. Traditional radio map construction methods rely on labor-intensive field measurements or deterministic propagation models, which are impractical for large-scale deployments and fail to adapt to dynamic environments. Methods such as matrix completion [4], dictionary learning [5], and deep learning [6] have shown promise in balancing accuracy and measurement efficiency. These methods exploit sparse data and underlying signal structure to generate reliable radio maps with few measurements. Despite these advancements, existing reconstruction works typically depend on real data collected through fixed sensors distributed across the region of interest, which may result in incomplete or biased maps, particularly in scenarios where emitters' positions hold crucial information for accurate reconstruction.

Conversely, the integration of unmanned aerial vehicles (UAVs) into radio environment sensing and monitoring introduces unprecedented flexibility and agility [7]. With the embedding of intelligence, UAV-assisted systems enable on-demand data collection across large or inaccessible areas, while supporting online sampling planning in the face of an unknown environment. Nevertheless, given the limited flying time of UAVs, strategically selecting measurement locations and dynamically optimizing trajectories in response to the reconstruction feedback become crucial for such online settings.

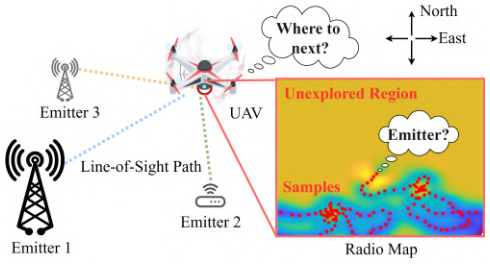



Fig. 1. UAV-assisted radio map sensing.

In this paper, we propose an adaptive sampling scheme for online UAV-assisted radio map sensing, which integrates data-driven reconstruction with intelligent trajectory planning. Unlike prior works [1], [4], [6], [8], [9] that focus solely on radio map reconstruction and rely on uncertainty-based criteria to guide UAV trajectories, our approach tackles a more comprehensive sensing task: simultaneously reconstructing the radio map and estimating emitter locations. This dual objective introduces challenges as conventional uncertainty-based criteria are insufficient. Specifically, uncertainty-based criteria fail to account for regions with high RSS, which require more intensive sampling but are not prioritized under uncertainty-based criteria.

Achieving this dual sensing task requires addressing the fundamental exploration-exploitation trade-off: exploration ensures sufficient sampling in areas with high uncertainty, while exploitation focuses on regions critical for emitter localization, as shown in Fig 1. To this end, we leverage Gaussian Processes (GPs) to statistically model the spatial correlation of radio signals. GPs provide both signal predictions and uncertainty metrics across the region of interest, enabling informed decision-

The work of Rui Zhou is supported by the NSFC under Grant 62201362 and Shenzhen Science and Technology Program under Grant No. RCBS20221008093126071.*(Corresponding author: Wenqiang Pu.)*

making during trajectory planning. By strategically combining these two metrics, we design an adaptive sampling scheme that balances exploration and exploitation. This scheme dynamically and iteratively refines UAV trajectories based on reconstruction feedback, ensuring both the efficiency of radio map reconstruction and the accuracy of emitter localization.

## II. System Model

### A. GP Model for Radio Map

Supposing the 2D radio map $f : \mathbb{R}^2 \to \mathbb{R}$ of the area of the interest $\mathcal{X}$ is modeled as a GP [8], [9]:

$$f(\cdot) \sim \mathcal{GP}(\mu(\cdot), \kappa(\cdot,\cdot)),$$

where $\mu(\cdot)$ is the mean function and $\kappa(\cdot,\cdot)$ is the covariance kernel indicating the characteristics of the radio map. For each $\mathbf{x} \in \mathcal{X}$, measurement $y$, i.e., RSS, conducted by the UAV is corrupted with an additive noise,

$$y = f(\mathbf{x}) + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2),$$

where $\sigma_\epsilon^2$ is the noise variance. Denote historical data consisting of $N$ collected measurements as $Y = \{y_0, y_1, \ldots, y_{N-1}\}$, which are sampled at locations $X = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{N-1}\}$. Using GP provides an effective way for predicting map values at any location $\mathbf{x} \in \mathcal{X}$ based on the collected samples. Specifically, GP models $f(\mathbf{x})$ along with $Y$ as a joint multivariate Gaussian distribution given by

$$\begin{bmatrix} f(\mathbf{x}) \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left( \begin{bmatrix} \mu_{\mathbf{x}} \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{x}Y} \\ \Sigma_{Y\mathbf{x}} & \Sigma_{YY} \end{bmatrix} \right),$$

where $\Sigma_{\mathbf{xx}} = \kappa(\mathbf{x},\mathbf{x})$, $\Sigma_{\mathbf{x}Y} = \kappa(\mathbf{x}, X)$, $\Sigma_{Y\mathbf{x}} = \kappa(X,\mathbf{x})$, $\Sigma_{YY} = \kappa(X,X) + \sigma_\epsilon^2 I$. Then, GP regression (GPR) make predictions for location of interest as[1]

$$\begin{aligned} \mu(\mathbf{x}) &= \mu_{\mathbf{x}} + \kappa(\mathbf{x}, X)K^{-1}(Y - \mu_Y), \\ \sigma^2(\mathbf{x}) &= \kappa(\mathbf{x},\mathbf{x}) - \kappa(\mathbf{x}, X)K^{-1}\kappa(X,\mathbf{x}). \end{aligned} \tag{1}$$

where $\mu_{\mathbf{x}}$ and $\mu_Y = [\mu_{\mathbf{x}_0}, \mu_{\mathbf{x}_1}, \ldots, \mu_{\mathbf{x}_{N-1}}]$ are initial value for locations and $K = \kappa(X,X) + \sigma_\epsilon^2 I$. It is important to note that $\mu(\mathbf{x})$ denotes the predicted RSS value at location $\mathbf{x}$, while the posterior variance $\sigma^2(\mathbf{x})$ quantifies the uncertainty associated with this prediction. Locations with higher prediction variance are particularly valuable, as they provide more informative data to reduce the uncertainty of the radio map [10]. Finally, we highlight that $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$ encapsulate critical information regarding the unsampled location $\mathbf{x}$. These metrics are essential for guiding UAV's motion planning, as they help balance between the exploration of more informative areas and the exploitation of regions characterized by higher RSS values. A detailed discussion of this balance is provided in Section III-A.

[1]In GPR, the kernel hyperparameters are recursively updated by maximizing the likelihood as new samples are collected. Since this is a standard procedure in GPR, the detailed steps are omitted for brevity [8], [9].

### B. Dynamic Model of UAV

At time $t$, the UAV is modeled as a continuous-time nonlinear dynamical system governed by the ordinary differential equation (ODE):

$$\dot{\mathbf{s}}_t = g(\mathbf{s}_t, \mathbf{u}_t),$$

where $\mathbf{s}_t \in \mathcal{X}$ denotes the system state and $\mathbf{u}_t \in \mathcal{U}$ represents the control input. The state space $\mathcal{X}$ encompasses all kinematically admissible configurations of the vehicle, actuated by inputs drawn from the action space $\mathcal{U}$. Concretely, the system state vector $\mathbf{s_t} = [r_t^x, r_t^y, \theta_t]^\top$ integrates the planar position $\mathbf{x}_t = [r_t^x, r_t^{y]}]^\top$ and the heading orientation $\theta_t$. The control input $\mathbf{u}_t = [v_t, w_t]^\top$ comprises the linear velocity $v_t$ and the rotation rate $w_t$.

However, frequent real-time state updates impose computationally intractable overhead for feedback-based motion planning systems. To address this, the dynamic motion model of the system over a discretization interval $\Delta_t = t' - t$ is then formulated as [11]:

$$\begin{aligned} r_{t'}^x &= \begin{cases} r_t^x + \frac{v_t}{w_t}(\sin(\theta_t + w_t \cdot \Delta_t) - \sin\theta_t), & w_t \neq 0 \\ r_t^x + v_t \sin\theta_t \cdot \Delta_t, & \text{otherwise} \end{cases}, \\ r_{t'}^y &= \begin{cases} r_t^y - \frac{v_t}{w_t}(\cos(\theta_t + w_t \cdot \Delta_t) - \cos\theta_t), & w_t \neq 0 \\ r_t^y + v_t \cos\theta_t \cdot \Delta_t, & \text{otherwise} \end{cases}, \\ \theta_{t'} &= \theta_t + w_t \cdot \Delta_t. \end{aligned} \tag{2}$$

Besides, we apply first-order linearization to the dynamic model, which yields the discrete-time linear state transition function as follows:

$$\mathbf{s}_{t+1} = \mathcal{L}(\mathbf{s}_t, \mathbf{u}_t) = A_t \mathbf{s}_t + B_t \mathbf{u}_t, \tag{3}$$

where $A_t = \frac{\partial \mathcal{L}}{\partial \mathbf{s}_t}$ represents the Jacobian matrix of current state to the previous state and $B_t = \frac{\partial \mathcal{L}}{\partial \mathbf{u}_t}$ the Jacobian matrix of current state to the control parameters, derived from (2).

### C. Motion Planning Formulation

The UAV actively senses the radio environment and evaluates surrounding regions with a reward function $R_t(\mu_t(\cdot), \sigma_t^2(\cdot))$ derived from GP. To enhance foresight and assimilate richer environmental information, receding horizon planning (RHP) method is integrated into motion planning.

The RHP methodology [12] operates as follows: At each decision epoch $t$, the UAV plans a trajectory comprising $H$ future waypoints. Denoting the current state as $\mathbf{s}_t^0$, the UAV optimizes its path over a discrete-time horizon $\{t + h\Delta_t | h = 1, \ldots, H\}$. For notational compactness, we define the predicted state sequence $\{\mathbf{s}_{t+\Delta_t}, \mathbf{s}_{t+2\Delta_t}, \ldots, \mathbf{s}_{t+H\Delta_t}\} \equiv \{\mathbf{s}_t^1, \mathbf{s}_t^2, \ldots, \mathbf{s}_t^H\}$. A candidate trajectory $\pi = \{\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^H\}$ thus accumulates a total reward

$$J_t = \sum_{h=1}^{H} R_t(\mathbf{x}_t^h),$$

where $\mathbf{x}_t^h = \lfloor \mathbf{s}_t^h \rfloor$ denotes the planar position (obtained via projection $\lfloor \cdot \rfloor$) of the state $\mathbf{s}_t^h$.

Accordingly, the control sequence over the horizon is $\mathbf{U}_t = \{\mathbf{u}_t^0, \mathbf{u}_t^1, ...\mathbf{u}_t^{H-1}\}$ with $\mathbf{u}_t^i = [v_t^i, w_t^i]^\top$. This yields the receding-horizon optimization problem:

$$\begin{aligned} \max_{\mathbf{U}_t} \ & J_t = \sum_{i=0}^{H-1} R_t(\lfloor g(\mathbf{s}_t^i, v_t^i, w_t^i) \rfloor) \\ \text{s.t. } & \mathbf{s}_t^{i+1} = A_t \mathbf{s}_t^i + B_t \mathbf{u}_t^i, \ \forall i, \\ & \mathbf{x}_t^i \in \mathcal{X}, \ \mathbf{u}_t^i \in \mathcal{U}, \ \forall i, \\ & v_t^i \in [v_{lb}, v_{ub}], w_t^i \in [w_{lb}, w_{ub}], \ \forall i, \end{aligned} \tag{4}$$

where $v_{lb}$, $v_{ub}$, $w_{lb}$, $w_{ub}$ represent the physical actuation limits of this vehicle.

## III. Adaptive Sampling Strategy

Our adaptive sampling procedure, as depicted in Algorithm 1, consists of two parts: data fitting and decision making. In data fitting part, GP regression introduced in Section II-A is adopted to fit the sampled history data and then predict the reward derived from GP for new locations. In the decision-making part, the UAV will determine its next movement through motion planning.

**Algorithm 1** Online UAV-assisted Adaptive Sampling

1: Initialization: Starting state $\mathbf{x}_0$, Sampling dataset $\mathcal{D} = \{(\mathbf{x}_0, y_0)\}$, Metamodel $f \leftarrow \text{GPR}(\mathcal{D})$, $t = 0$
2: **repeat**
3: Update reward function $R_t$ based on $f$
4: Trajectory $\pi_t = \{\mathbf{x}_t^1, \mathbf{x}_t^2, ..., \mathbf{x}_t^H\}$ = Planning$(R_t, \mathbf{x}_t)$
5: Fly along $\pi_t$ and obtain new samples $\mathcal{D} \cup \{(\mathbf{x}_t^1, y_t^1), ...\}$
6: Update Metamodel $f \leftarrow \text{GPR}(\mathcal{D})$
7: $t \leftarrow t + 1$
8: **until** $t$ achieves the maximum mission time $T_m$

### A. Exploration-Exploitation Trade-off Reward

The selection of the reward function $R_t(\mathbf{x}_t)$ in problem (4) is crucial for effectively guiding the UAV's motion. A straightforward approach is to set $R_t(\mathbf{x}_t) = \mu(\mathbf{x}_t)$, which focuses on sampling points with high RSS values. However, this method may lead the algorithm to converge to a local optimum. Conversely, a strategy that prioritizes high uncertainty, i.e., $R_t(\mathbf{x}_t) = \sigma(\mathbf{x}_t)$, could overlook valuable information surrounding the signal source. To address this trade-off between exploration and exploitation, the GP-UCB (Upper Confidence Bound) [13] framework is adopted, which provides a balanced approach that combines both strategies. Inspired by the understanding that uncharted areas require increased exploration as time progresses, our reward function is defined as follows

$$R_t(\mathbf{x}_t) = \mu(\mathbf{x}_t) + \beta_t \sigma(\mathbf{x}_t), \beta_t = \sqrt{2 \log(\frac{|D| t^a \pi^2}{6\delta})},$$

where $|D|$ denotes the dimension of the input space, $a \geq 2$ and $\delta \in (0, 1)$ are pre-determined parameters. In this formulation, the term $\beta_t$ effectively scales the uncertainty, enabling our strategy to dynamically adjust its focus on exploration as time progresses. Notably, when $a = 2$, this configuration reduces to the typical GP-UCB. While our approach places a significant emphasis on exploring regions with high uncertainty. Numerical simulations presented in Section IV demonstrate that this setting fosters a more comprehensive understanding of the environment over time. As a result, it enhances the decision-making capabilities of the UAV's sampling strategy, allowing for more informed and effective exploration of the environment.

### B. Gradient Enhanced Cross-Entropy Optimization

The cross-entropy (CE) method [14] is a powerful technique for addressing the optimization problem outlined in (4). This approach treats the optimization task as an estimation problem for rare events, employing importance sampling to efficiently find solutions. Let us denote the optimal value of the objective function for each planning step as:

$$\gamma_t^* = \max_{\mathbf{U}_t} J_t,$$

which can be considered as a rare event in this context. The process of identifying the optimal control parameters from a sampling distribution involves iteratively simulating these rare events until the distribution converges towards a delta function. Consequently, the reward $\gamma_t^k$ in iteration $k$ approaches the optimal value $\gamma_t^*$. By focusing on elite samples, the CE method efficiently narrows down the search space, leading to improved convergence and performance in solving the optimization problem.

However, the random sampling procedure inherent in the CE method does not fully leverage the gradient information provided by the reward function, which can result in inefficient sampling. To address this, we propose a gradient enhanced CE method that facilitates guided sampling. Denote the control parameters as

$$z = \{v_0, w_0, v_1, w_1, \ldots, v_{H-1}, w_{H-1}\},$$

where $z$ follows a Gaussian mixture distribution, expressed as $z \sim \mathcal{N}(\mu_z, \Sigma_z)$. Our proposed gradient enhanced CE method involves updating the mean $\mu_z$ and the covariance $\Sigma_z$ based on the samples, integrating gradient information to refine the sampling process. The detailed steps for implementing this approach are outlined below.

**Initialization.** Set $\rho_0 = \{\mu_0^z, \Sigma_0^z\}$ as the initial parameters of the distribution for $z$, where $z \sim \mathcal{N}(\mu_0^z, \Sigma_0^z)$.

**Sampling.** For each CE iteration $k \in \{1, 2, \ldots, K\}$, draw $G$ samples $\{z_1, z_2, \ldots, z_G\}$ from $\mathcal{N}(\mu_k^z, \Sigma_k^z)$, while ensuring that the control samples and the generated trajectories remain within the specified operational boundaries.

**Optimization.** For each sampled trajectory, use the gradient information from the reward model to optimize the control parameters. This is achieved by updating the parameters as follows:

$$z_g^{m+1} \leftarrow z_g^m + \eta \nabla_z J_t^m,$$

where $m \in \{1, 2, \ldots, M\}$ denotes the optimization step, and $\eta$ represents the step size.

TABLE I
SIMULATED RADIO MAP SETTING.

| **Emitter** | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| **Power** [W] | 1.2 | 0.2 | 0.8 | 1.5 | 0.5 |
| **Height** [m] | 5 | 3 | 2 | 4 | 1 |
| **Longitude** | -71.3561 | -71.3573 | -71.3635 | -71.3624 | -71.3661 |
| **Latitude** | 42.3346 | 42.3412 | 42.3403 | 42.3369 | 42.3342 |

**Parameter Update.** Let the reward $\gamma_t^k$ be the $\alpha$ percentile of $J_t$ under $\rho_k$, choose the elite set $\epsilon_k$ of the samples evaluated with $J_t(z_g) > \gamma_t^k$. Update the parameter estimation $\mu_{k+1}^z$, $\Sigma_{k+1}^z$ for Gaussian distribution based on expectation-minimization method [14].

**Termination.** Terminate until the preset iterations end or KL-divergence between $\mathcal{N}(\mu_k^z, \Sigma_k^z)$ and $\mathcal{N}(\mu_{k+1}^z, \Sigma_{k+1}^z)$ smaller than a fixed threshold. Then choose $z_k$ which has the highest reward as the final solution $z^*$.

## IV. SIMULATION RESULTS

In this section, we compare the sampling trajectory, reconstruction error, and computation time of the proposed algorithm against the uncertainty-based baseline [8], [9]. All experiments are run on server with 2 CPUs of which the type is Intel(R) Xeon(R) Gold 6230R CPU @ 2.10GHz.

### A. Experiment Setup

**Radio Map Setting:** To evaluate the algorithms, a radio map is generated using MATLAB, based on a coverage map that utilizes a free space propagation model without any building blocks. In the defined region $\mathcal{X}$, five emitters operating on the same frequency are deployed, with the coverage map extending to a maximum range of 600 m. Detailed specifications of the emitters can be found in Table I.

**UAV Setting:** The UAV equipped with a sensor capable of collecting signal strength data flying at a constant altitude of 10 m is considered. Its velocity and rotation rate constraints are defined as follows: the lower and upper bounds of the UAV's velocity are $v_{lb} = 0.5$ m/s and $v_{ub} = 8$ m/s, respectively, while the rotation rate bounds are $w_{lb} = -1.2$ rad/s and $w_{ub} = 1.2$ rad/s. The sampling interval is $\Delta_t = 1$ s, and the receding horizon length is $H = 5$. Ultimately, it executes the two steps that yield the highest expected reward. This approach allows for efficient exploration of the environment while optimizing data collection.

**Algorithm Setting:** To examine the influence of the exploration-exploitation trade-off on the adaptive sampling algorithm, we set $\delta = 0.9$ and $|D| = 2$ (Lat.,Lon.) for $\beta_t$ and compare the reconstruction results across various settings of $a = 2, 4, 6, 8$. For the gradient enhanced cross-entropy optimization, we establish the total number of iterations $K = 2$, the optimization steps $M = 5$, and the number of sampling trajectories $G = 20$. Additionally, the step size $\eta$ is fine-tuned separately for both velocity and rotation rate. For the initialization of control parameters, $\mu_0^z$ is typically set around the midpoint of the admissible range with $\Sigma_0^z$ set to unity.

TABLE II
COMPARISON OF POSITIONING ERROR FOR DIFFERENT SETTINGS.

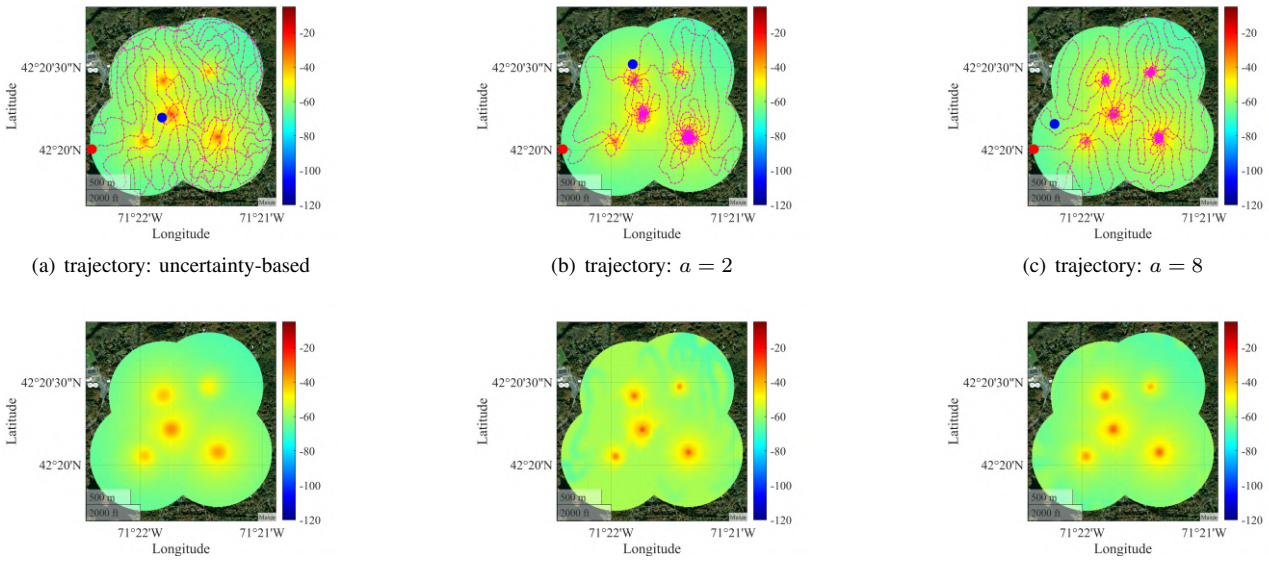
| **Setting** | Uncertainty | $a = 2$ | $a = 8$ |
|---|---|---|---|
| **Lon. Err.** | $2.46e-5$ | $1.43e-5$ | $1.51e-5$ |
| **Lat. Err.** | $4.84e-5$ | $1.84e-5$ | $0.78e-5$ |

### B. Results and Analysis

Fig. 2 illustrates a comparison of sampling trajectories and reconstructed radio maps under different exploration-exploitation trade-off settings. As anticipated, sampling based solely on uncertainty drives the UAV toward unexplored areas. While this approach ensures that sampling points are distributed across the entire map, it neglects critical information about the source location. In contrast, subfigures (b) and (c) demonstrate the effectiveness of balancing exploration and exploitation, where the UAV focuses more on regions with higher RSS. Consequently, the sampling point density near the source is sufficiently high to accurately indicate the source's location. A detailed comparison of positioning error is provided in Table II. Furthermore, as the parameter $a$ increases, the proportion of uncertainty in the reward function grows more rapidly over time, resulting in improved exploration of the radio map.

Fig. 3 presents the reconstruction mean squared error (MSE) for different values of parameter $a$. The numerical experiments were repeated three times, and the resulting curves were smoothed using a moving average with a window size of 15 for clarity. The results indicate that larger value of $a$ reduces the MSE more rapidly during the early stages of the sampling process, but the differences in MSE gradually diminish, with only a small gap remaining by the end of the process. This demonstrates that while increased exploration accelerates initial reconstruction, the trade-off has less impact as the sampling process progresses and sufficient measurements are obtained. Table II compares the localization error (in degree) averaged across five emitters, where the locations of emitters are coarsely estimated by the coordination of reconstructed map's peaks. It is evident that the uncertainty-based method results in larger localization errors despite achieving the smallest MSE in radio map reconstruction. In contrast, the proposed sampling method with $a = 8$ strikes a better balance between accurate localization and effective radio map reconstruction.

Fig. 4 compares the computation time for the standard CE method with $G = 50$ and $K = 10$, and the gradient enhanced CE method with $G = 20$, $K = 2$, and $M = 5$. While both configurations achieve similar MSEs in radio map reconstruction, their computation times differ. It can be observed that the standard CE method is more time-consuming due to its reliance on larger sample sizes, which requires additional computational effort. Note that the computation time per step increases as more data is incorporated into the GPR, leading to longer prediction times for the reward function and fluctuations in the computation time are caused by resampling events when the sampling process encounters




(a) trajectory: uncertainty-based (b) trajectory: $a = 2$ (c) trajectory: $a = 8$

(d) reconstructed radio map: uncertainty-based (e) reconstructed radio map: $a = 2$ (f) reconstructed radio map: $a = 8$

Fig. 2. Adaptive sampling results under different settings. The simulated radio map is demonstrated above the terrain map, and the signal strength is denoted from high to low as from red to blue. In (a)-(c), the red point and the blue point represent the starting point and end point of the UAV, which conducts 500 planning sessions and collects 1000 samples in one sampling task. The pink points and trajectories are the sampling way points and sampling trajectories. Reconstructed radio map at the end of the adaptive sampling task are depicted in (d)-(f).

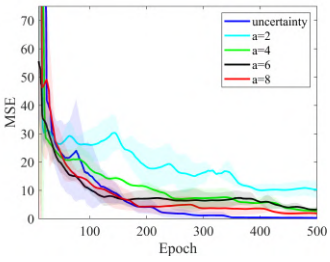



Fig. 3. Reconstruction MSE.

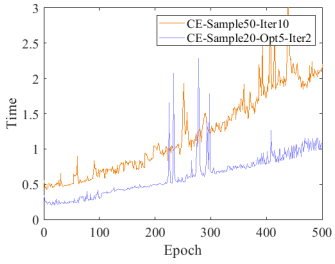



Fig. 4. Computation time comparison.

boundary constraints.

## V. Conclusion

In this work, we formulate adaptive sampling for online UAV-assisted radio map sensing as an online sequential optimization problem, comprising two key components: model fitting and motion planning. We utilize GPR to construct a black-box metamodel of the radio map, and design a reward function inspired by GP-UCB to guide the sampling process. The UAV's motion planning is driven by a specifically designed reward function that balances exploration and exploitation and a gradient enhanced cross-entropy optimization method is developed to generate the trajectory for the UAV, ensuring efficient and effective data collection throughout the sampling process.

## References

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

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

[3] L. Xu, L. Cheng, J. Chen, W. Pu, and X. Fu, "Radio map estimation via latent domain plug-and-play denoising," 2025. [Online]. Available:https://arxiv.org/abs/2501.13472

[4] H. Dong, W. Pu, R. Zhou, X. Fu, and F. Yin, "Integrated interpolation and matrix completion for radio map estimation: A convex optimization approach," International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025, pp. 1–5.

[5] S. Zhang, T. Yu, B. Choi, F. Ouyang, and Z. Ding, "Radiomap inpainting for restricted areas based on propagation priority and depth map," IEEE Transactions on Wireless Communications, vol. 23, no. 8, pp. 9330–9344, 2024.

[6] R. Levie, Ç. Yapar, G. Kutyniok, and G. Caire, "RadioUNET: Fast radio map estimation with convolutional neural networks," IEEE Transactions on Wireless Communications, vol. 20, no. 6, pp. 4001–4015, 2021.

[7] D. Gesbert, O. Esrafilian, J. Chen, R. Gangula, and U. Mitra, "UAV-aided RF mapping for sensing and connectivity in wireless networks," IEEE Wireless Communications, vol. 30, no. 4, pp. 116–122, 2023.

[8] X. Chen, X. Zhong, Z. Zhang, L. Dai, and S. Zhou, "High-efficiency urban 3D radio map estimation based on sparse measurements," 2024. [Online]. Available: https://arxiv.org/abs/2408.04205

[9] R. Shrestha, D. Romero, and S. P. Chepuri, "Spectrum surveying: Active radio map estimation with autonomous UAVs," IEEE Transactions on Wireless Communications, vol. 22, no. 1, pp. 627-641, 2023.

[10] G. Morales and J. Sheppard, "Adaptive sampling to reduce epistemic uncertainty using prediction interval-generation neural networks," 2024. [Online]. Available: https://arxiv.org/abs/2412.10570

[11] Y. T. Tan, A. Kunapareddy, and M. Kobilarov, "Gaussian process adaptive sampling using the cross-entropy method for environmental sensing and monitoring," in 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018, pp.6220–6227.

[12] K. Bergman, O. Ljungqvist, T. Glad, and D. Axehill,"An optimization-based receding horizon trajectory planning algorithm," IFAC-PapersOnLine, vol. 53, no. 2, pp. 15 550–15 557, 2020.

[13] N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, "Gaussian process optimization in the bandit setting: No regret and experimental design,"arXiv preprint arXiv:0912.3995, 2009.

[14] M. Kobilarov, "Cross-entropy motion planning," The International Journal of Robotics Research, vol. 31, no. 7, pp. 855–871, 2012.