

Informative Path Planning with Guaranteed Estimation Uncertainty

Kalvik Jakkala*, Saurav Agarwal[†], Jason O’Kane* and Srinivas Akella[‡]

*Texas A&M University, [†]Indian Institute of Technology Bombay, [‡]University of North Carolina at Charlotte

Email: {kalvik, jokane}@tamu.edu, saurav.agarwal@iitb.ac.in, sakella@charlotte.edu

Abstract—Environmental monitoring robots often need to estimate data fields (e.g., salinity, temperature, bathymetry) under tight resource constraints. Classical boustrophedon lawnmower surveys provide geometric coverage guarantees but can waste effort by oversampling predictable regions. In contrast, informative path planning (IPP) methods leverage spatial correlations to reduce oversampling, yet typically offer no guarantees on estimation quality. This paper bridges these approaches by addressing IPP with guaranteed estimation uncertainty in complex environments: computing the shortest path whose measurements ensure that the Gaussian process (GP) posterior variance—an intrinsic uncertainty measure that lower-bounds the mean-squared prediction error under the GP model—is upper bounded by a user-specified threshold over the monitoring region.

We propose a three-stage approach for efficient environmental monitoring: (i) learning a GP model from prior information; (ii) transforming the GP kernel into binary coverage maps that identify locations where uncertainty can be reduced below a target threshold; and (iii) planning a near-shortest route to satisfy the global uncertainty constraint. Our approach incorporates non-stationary kernels to capture spatially varying correlations in heterogeneous phenomena and accommodates non-convex environments with obstacles. We provide near-optimal approximation guarantees for both sensing-location selection and the joint selection-and-routing problem under a travel budget. Experiments on real-world topographic data demonstrate that our planners achieve uncertainty targets with fewer sensing locations and shorter travel distances than representative baselines. Furthermore, field experiments with autonomous surface and underwater vehicles validate the real-world feasibility of the approach. Our code is available at: www.sgp-tools.com

I. INTRODUCTION

Monitoring environmental variables such as ocean currents, salinity, and bathymetry is a critical task of broad scientific and societal importance [4]. These quantities are both sensitive indicators of coupled Earth-system dynamics and active drivers of the climate, natural hazards, ecosystems, and resource availability. Accordingly, systematic observation strategies that provide robustness through model-based guarantees, efficiency through resource-aware optimization, and repeatability through automation are essential for effective risk reduction and sustainable management.

Despite its importance, environmental monitoring under finite resource constraints presents significant challenges. A common approach is the boustrophedon *lawnmower* pattern, which provides strong geometric coverage guarantees by sweeping the sensing footprint across the entire region [1, 5, 6]. However, many environmental fields exhibit strong spatial correlations. In these contexts, exhaustive geometric coverage

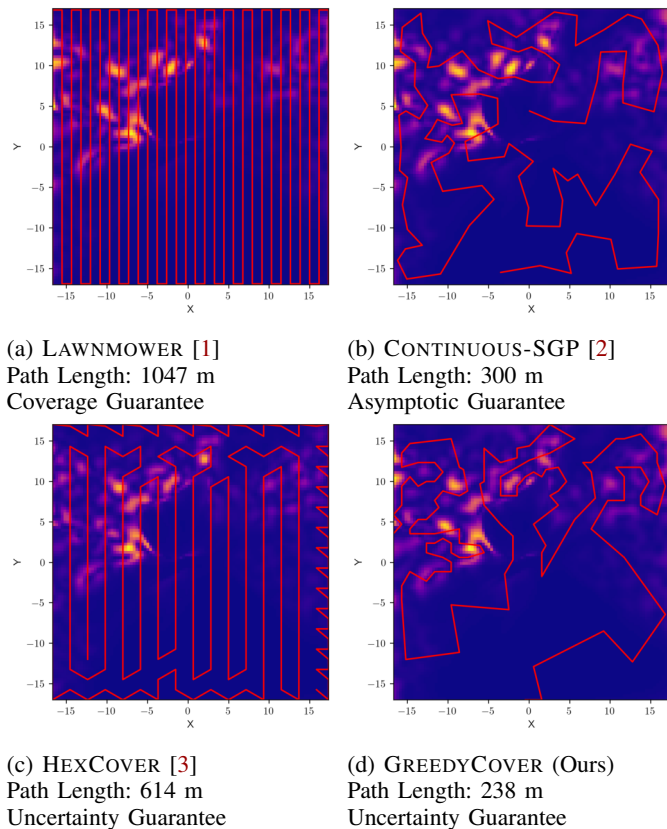


Fig. 1: Comparison of IPP solution paths (red) and the resulting GP predictive uncertainty (heatmaps). While all methods maintain an identical mean predictive uncertainty of 0.02, the proposed GREEDYCOVER approach yields the shortest path while guaranteeing that uncertainty remains below the target threshold across all evaluation locations.

often results in wasted effort, oversampling regions that are easily predictable from nearby observations.

Informative path planning (IPP) leverages spatial correlations—often through Gaussian processes (GPs) [7]—to reduce travel while maintaining reconstruction quality. However, most IPP methods optimize surrogate measures of information gain without certifying the quality of the resulting reconstruction. Moreover, existing approaches typically adopt stopping criteria based on resource budgets, such as mission time, travel distance, or a cardinality constraint on sensing



Fig. 2: An autonomous surface vehicle (ASV) and an autonomous underwater vehicle (AUV) mapping bathymetry during our field trials.

locations. Real-world deployments require a more principled approach: a stopping criterion tied directly to the desired reconstruction accuracy.

Recent research has begun to bridge this gap by introducing *estimation uncertainty guarantees*. These methods provide an upper bound on the GP predictive variance, which in turn lower-bounds the mean-squared error (MSE) [8, 3]. While predictive uncertainty is only a proxy for accuracy, it serves as a rigorous measure of the information gained relative to the prior data, providing a quantitative signal of mission progress. However, existing methods with such guarantees are limited in two critical ways: they generally rely on Radial Basis Function (RBF) kernels—forcing a stationary correlation model—and are derived only for simplified, convex environments. Such constraints are rarely met in the real-world, where environmental fields can be highly non-stationary and operational areas often involve obstacles or complex, non-convex boundaries.

This paper addresses these limitations by developing new methods for IPP with estimation uncertainty guarantees in complex environments. Figure 1 highlights the resulting efficiency gains in a non-stationary bathymetry setting relative to representative baselines. Our main contributions are:

- **IPP with Uncertainty Guarantees:** We present an informative path planning approach that upper-bounds the GP predictive variance, ensuring the maximum predictive uncertainty remains below a user-specified threshold.
- **Non-Stationary and Non-Convex Modeling:** We introduce a method that captures and leverages non-stationary correlations while accommodating non-convex environments containing obstacles and complex boundaries.
- **Near-Optimal Algorithms:** We provide near-optimal approximation algorithms for both sensing-location selection and the joint location-selection and routing problem.
- **Field Validation:** We validate our approach through field trials using autonomous surface and underwater vehicles for bathymetry mapping, demonstrating the practical feasibility of our method in real-world conditions.

II. RELATED WORK

Lawnmower Planners: A foundational approach in environmental monitoring is the boustrophedon (lawnmower) sweep, where a robot follows parallel transects to provide deterministic coverage based on its sensing footprint [1].

This paradigm has been adapted to account for realistic vehicle dynamics, such as Dubins kinematics constraints for autonomous surface vehicles [5], and extended to multi-agent systems where area coverage objectives must be balanced against practical constraints like limited onboard energy or communication range [6].

Contrast to our work: Lawnmower planners guarantee geometric visitation but oversample predictable areas by ignoring data correlations. Our approach instead optimizes for a worst-case posterior uncertainty threshold, enabling non-sweep paths that skip redundant regions while guaranteeing a global uncertainty bound.

Coverage Planners: Our work is also closely related to viewpoint and sensor placement formulations that offer geometric alternatives to exhaustive sweeping. These include randomized art-gallery algorithms designed to select discrete sensor locations that collectively *see* an entire environment [9]. More broadly, the field encompasses decentralized and adaptive strategies targeting geometric visitation or sensing density, particularly in complex or multi-robot settings. For instance, Voronoi-based methods provide coverage guarantees in non-convex environments for networked teams [10], while robust adaptive control laws address uncertainty in the underlying sensor network objectives [11]. In unknown environments, recent methods trade off exploration (mapping free space) with coverage progress to improve operational efficiency [12].

Contrast to our work: Unlike coverage planners that require full-field traversal for geometric visibility, we optimize a *coverage-as-uncertainty* objective. By imposing a global worst-case posterior variance threshold, we certify estimation uncertainty while purposefully skipping regions that do not require direct sensing.

Informative Path Planning Methods: Informative Path Planning (IPP) and Robotic Information Gathering (RIG) focus on planning paths that maximize information gain regarding an unknown field, frequently employing Gaussian Processes (GPs) [7] to model spatial correlations. Submodular formulations have provided a rigorous foundation for this, enabling efficient approximation algorithms for sensor placement [13], including extensions that incorporate travel and routing constraints [14, 15].

In robotic monitoring, IPP has been extensively studied for waypoint optimization in spatiotemporal settings [16], adaptive continuous-space planning for environmental monitoring [17], and Bayesian exploration for mapping [18]. Recent research has expanded into sequential optimization-based approaches [19] and distributed mapping for heterogeneous teams using wavelet-based ergodic exploration [20]. Parallel to these methods, learning-based techniques—such as deep reinforcement learning for adaptive planning [21] and feature tracking [22]—have gained traction. Other efforts focus on model expressiveness and scalability, utilizing multi-scale sampling [23], attentive kernels for information gathering [24, 25], and differentiable optimization via sparse Gaussian Processes [26, 27, 2]. Furthermore, routing-constrained IPP has been applied to high-stakes estimation tasks, such as

localized gas leak rate quantification [28].

Contrast to our work: Unlike most IPP approaches that maximize information gain without providing formal guarantees on estimation quality, we plan with respect to an explicit global uncertainty constraint. By targeting the worst-case posterior variance, our method ensures that the resulting paths satisfy prescribed estimation uncertainty requirements.

IPP Methods with Estimation Uncertainty Guarantees:

The most relevant approaches to our work belong to the specialized class of IPP methods that provide formal *estimation uncertainty guarantees*. Unlike standard information-gathering techniques, these methods certify that a reconstructed field satisfies a user-specified uncertainty threshold. For instance, Suryan and Tokekar investigated minimum-time field learning with multi-robot teams, leveraging GP structures to derive performance guarantees under operational time constraints [8]. Building on this foundation, Dutta *et al.* developed approximation algorithms for the dual sensor placement and shortest-path problems, strengthening theoretical guarantees within convex environments [3]. Collectively, these works demonstrate that rigorous, coverage-style certification can be successfully integrated into the informative sampling paradigm.

Contrast to our work: Since existing uncertainty guarantee-based methods typically assume stationary correlations and convex geometries, they often fail in spatially varying or obstacle-dense environments. We extend these guarantees to expressive, non-stationary models and non-convex domains, providing formal certificates for both near-optimal approximation and estimation uncertainty.

III. PRELIMINARIES: GAUSSIAN PROCESSES

We model the monitored quantity as an unknown function over a domain and use Gaussian processes (GPs) [7] to represent both spatial structure and predictive uncertainty.

GP Prior: A Gaussian process is a stochastic process such that any finite collection of function values is jointly Gaussian. Equivalently, a GP induces a distribution over functions, making it an expressive Bayesian prior for regression. We place a GP prior on the latent data field $f(\cdot)$:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')), \quad (1)$$

where $m(x)$ is the mean function (often taken to be 0 after normalization) and kernel $k(x, x')$ is a positive semi-definite covariance function. The kernel specifies prior correlations:

$$k(x, x') = \text{cov}(f(x), f(x')). \quad (2)$$

Noisy Observations and Posterior Prediction: Given n noisy samples $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$ with $y_i = f(x_i) + \varepsilon_i$ and i.i.d. Gaussian noise $\varepsilon_i \sim \mathcal{N}(0, \sigma_n^2)$, define the stacked training inputs $X = [x_1, \dots, x_n]$ and outputs $y = [y_1, \dots, y_n]^\top$. For a query location x_* , GP regression yields a Gaussian predictive distribution $p(f_* | X, y, x_*) = \mathcal{N}(\mu_*, \sigma_*^2)$, with

$$\mu_* = K_{*n}(K_{nn} + \sigma_n^2 I)^{-1} y, \quad (3)$$

$$\sigma_*^2 = K_{**} - K_{*n}(K_{nn} + \sigma_n^2 I)^{-1} K_{n*}. \quad (4)$$

Here, K_{nn} is the $n \times n$ covariance matrix with $(K_{nn})_{ij} = k(x_i, x_j)$, K_{*n} is the $1 \times n$ vector with $(K_{*n})_i = k(x_*, x_i)$, $K_{n*} = K_{*n}^\top$, and $K_{**} = k(x_*, x_*)$. The posterior variance σ_*^2 provides an intrinsic uncertainty measure that depends only on the input locations (and the GP hyperparameters), which makes it especially useful for planning.

IV. PROBLEM: INFORMATIVE PATH PLANNING WITH GUARANTEED ESTIMATION UNCERTAINTY

System Setup: A robot is tasked with monitoring an environmental field (e.g., salinity, temperature, bathymetry) over a purely spatial or spatio-temporal domain $\mathcal{X} \subset \mathbb{R}^d$. For planning and performance evaluation, we define a finite set of *evaluation points* $\mathcal{V} = \{x_1, \dots, x_N\} \subset \mathcal{X}$ where accurate estimates are required.

Due to accessibility, safety, and operational constraints, the robot can only sample from a restricted set of *candidate sensing locations* $\mathcal{C} = \{c_1, \dots, c_M\} \subseteq \mathcal{X}$. A feasible path is represented by an ordered sequence of waypoints $\mathcal{P} = (p_1, \dots, p_T)$, where each $p_t \in \mathcal{C}$. We denote the travel cost of a path as $\text{cost}(\mathcal{P})$ (e.g., Euclidean distance or boundary-constrained path length).

GP-based Uncertainty Surrogate: Directly optimizing the pointwise Mean Squared Error (MSE) is challenging because it generally depends on the unknown measurement values realized along $\mathcal{P}$. To obtain a tractable objective, we model the unknown field $f : \mathcal{X} \rightarrow \mathbb{R}$ as a Gaussian Process (GP).

The GP provides a posterior variance $\sigma_{\text{post}}^2(x_i | \mathcal{P})$ that depends only on the sensing locations, not the observed values. Since the posterior variance provides a lower bound on the pointwise MSE [29], we define the *worst-case uncertainty* as:

$$J_{\max}(\mathcal{P}) := \max_{x_i \in \mathcal{V}} \sigma_{\text{post}}^2(x_i | \mathcal{P}). \quad (5)$$

The GP is initialized using a representative set of labeled points obtained from pilot missions, historical data, or coarse satellite observations.

Problem Statement: The *informative path planning problem with guaranteed estimation uncertainty* aims to find the minimum-cost path that ensures the posterior variance at every evaluation point remains below a user-specified threshold σ_{tar}^2 :

$$\begin{aligned} \mathcal{P}^* \in \arg \min_{\mathcal{P} \in \mathcal{C}^*} \quad & \text{cost}(\mathcal{P}) \\ \text{s.t.} \quad & J_{\max}(\mathcal{P}) \leq \sigma_{\text{tar}}^2, \\ & \text{cost}(\mathcal{P}) \leq D \quad (\text{optional}), \end{aligned} \quad (6)$$

where $\mathcal{C}^* := \bigcup_{T \geq 1} \mathcal{C}^T$ denotes the set of all finite waypoint sequences over $\mathcal{C}$, and D is an optional budget constraining the total path cost.

V. NEAR-OPTIMAL INFORMATIVE PATH PLANNING

This section presents our approach to solving the informative path planning problem with guaranteed estimation uncertainty. We first describe how we leverage available prior information to compute coverage maps. We then use these maps in two algorithms—GREEDYCOVER and GCBCOVER—which address sensing-location selection and the joint problem of sensing-location selection and route planning, respectively.

A. Coverage Map Construction Theory

We begin by fitting a Gaussian process (GP) to the available prior information. Depending on the data source, we use either a non-stationary kernel [24] or a physics-informed kernel [30]. Using the learned GP, we construct *coverage maps* that specify which evaluation points $v \in \mathcal{V}$ can be driven below the user-specified target posterior variance σ_{tar}^2 by taking a measurement at a candidate sensing location $c \in \mathcal{C}$.

Ideally, coverage would be evaluated jointly over *sets* of sensing locations, since the posterior variance at an evaluation point depends on all measurements collected from a path. However, informative path planning is NP-hard [16]; exact evaluation would require computing posterior variances for many subsets of the power set of $\mathcal{C}$, which is computationally infeasible. We therefore adopt a scalable alternative: for each candidate location $c \in \mathcal{C}$, we compute the set of evaluation points that would satisfy the variance threshold if we were to observe *only* at c . These per-location coverage maps can then be combined efficiently during planning.

We justify this construction with the following results. The first provides an explicit prior covariance threshold under which a single measurement at a sensing location can reduce the posterior variance at a given evaluation point below σ_{tar}^2 . The second formalizes the monotonicity property that conditioning on additional measurements can only further reduce posterior variance; hence, the single-location condition yields a conservative coverage region for multi-measurement paths.

Theorem 1 (Minimum Required Prior Covariance).

Let $f \sim \mathcal{GP}(0, k)$ with independent Gaussian observation noise variance σ_n^2 , and consider a single noisy observation at candidate location $c \in \mathcal{C}$. Fix an evaluation location $v \in \mathcal{V}$. For any target posterior variance $\sigma_{\text{tar}}^2 \in (0, k(v, v))$, the condition

$$\sigma_{\text{post}}^2(v | c) \leq \sigma_{\text{tar}}^2$$

holds if and only if

$$|k(c, v)| \geq \sqrt{(k(v, v) - \sigma_{\text{tar}}^2)(k(c, c) + \sigma_n^2)}. \quad (7)$$

Thus, achieving a posterior variance of at most σ_{tar}^2 at v requires that the prior covariance between c and v exceed the threshold in Equation 7.

Please refer to the Appendix for the proof.

Theorem 2 (Variance Reduction with Multiple Locations).

Under the setup of Theorem 1, consider n sensing locations $c_1, \dots, c_n \in \mathcal{C}$. Fix an evaluation location $v \in \mathcal{V}$ and a target variance $\sigma_{\text{tar}}^2 \in (0, k(v, v))$. Suppose there exists $j \in \{1, \dots, n\}$ such that

$$|k(c_j, v)| \geq \sqrt{(k(v, v) - \sigma_{\text{tar}}^2)(k(c_j, c_j) + \sigma_n^2)}.$$

Then the posterior variance at v conditioned on all observations satisfies

$$\sigma_{\text{post}}^2(v | c_{1:n}) \leq \sigma_{\text{tar}}^2.$$

Please refer to the Appendix for the proof.

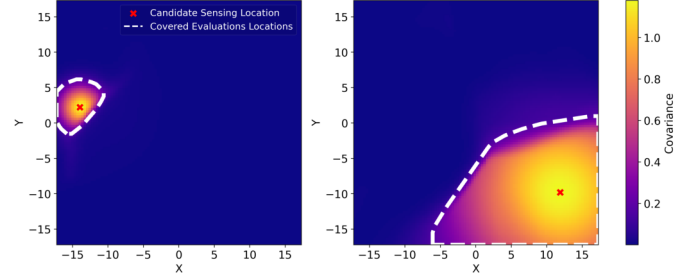


Fig. 3: Coverage maps for two candidate sensing locations under a non-stationary kernel. Each dotted white blob indicates the set of evaluation locations for which that candidate location can guarantee the target posterior-variance threshold.

B. Binary Coverage Map Construction

Together, Theorems 1 and 2 motivate a per-location notion of coverage: a sensing location c_j covers an evaluation point v_i if the prior covariance $k(c_j, v_i)$ exceeds the threshold in Equation 7. We encode this relationship in a binary matrix $B \in \{0, 1\}^{M \times N}$. For each candidate sensing location $c_j \in \mathcal{C}$ and evaluation point $v_i \in \mathcal{V}$, we set $B_{ji} = 1$ (i.e., sensing at c_j covers v_i) if

$$\mathbb{I}\left\{|k(c_j, v_i)| \geq \sqrt{(k(v_i, v_i) - \sigma_{\text{tar}}^2)(k(c_j, c_j) + \sigma_n^2)}\right\},$$

and $B_{ji} = 0$ otherwise. Figure 3 shows example coverages.

This binary representation provides two practical benefits. First, it reduces memory compared to storing real-valued covariances for all pairs in $\mathcal{C} \times \mathcal{V}$. Second, it enables fast planning-time updates via set operations: the evaluation points covered by a path can be computed by unions over the rows of B corresponding to the visited candidate locations, without repeatedly evaluating GP posterior variances for different candidate subsets.

C. GREEDYCOVER: IPP via Greedy Selection and TSP

Given the binary coverage matrix B from Section V-B, our first planning approach proceeds in two stages: (i) greedily select a subset of candidate sensing locations that maximizes the number of covered evaluation points, and (ii) construct a path that visits all selected locations using a TSP solver [31].

1) *Greedy Sensor Selection*: Starting from an empty selected set $S_0 = \emptyset$ and an empty covered set $U_0 = \emptyset$, the greedy procedure iteratively adds the candidate location that covers the largest number of *new* (previously uncovered) evaluation points. Let $S_t \subseteq \{1, \dots, M\}$ denote the indices of selected candidates after iteration t , and let $U_t \subseteq \mathcal{V}$ denote the set of evaluation points covered by S_t . The marginal gain of adding candidate $j \notin S_t$ is

$$\Delta(j | S_t) := |\{v_i \in \mathcal{V} : B_{ji} = 1, v_i \notin U_t\}|.$$

At each iteration, we select

$$j^* \in \arg \max_{j \notin S_t} \Delta(j | S_t),$$

update $S_{t+1} = S_t \cup \{j^*\}$, and expand the covered set

$$U_{t+1} = U_t \cup \{v_i \in \mathcal{V} : B_{ji} = 1\}.$$

The loop terminates once all evaluation points are covered (i.e., $|U_t| = N$) or once no remaining candidate yields additional coverage (i.e., $\max_{j \notin S_t} \Delta(j | S_t) = 0$). Algorithm 1 in the Appendix summarizes this procedure. The worst-case running time of the algorithm is $\mathcal{O}(M^2N)$, when $M \leq N$.

2) *Submodularity and Approximation Guarantees*: The objective implicitly optimized by the above greedy method is a classic monotone submodular set function [32, 33, 13]. Define the *coverage function* with $S \subseteq \{1, \dots, M\}$:

$$F(S) := \left| \bigcup_{j \in S} \{v_i \in \mathcal{V} : B_{ji} = 1\} \right|, \quad (8)$$

which counts how many evaluation points are covered by the selected candidates. This function is:

- *Monotonically Increasing*: If $Y \subseteq Z$, then $F(Y) \leq F(Z)$, since adding sensing locations can only enlarge the union of covered points.
- *Submodular*: It satisfies diminishing returns. For all $Y \subseteq Z$ and any location $l \notin Z$,

$$F(Y \cup \{l\}) - F(Y) \geq F(Z \cup \{l\}) - F(Z),$$

since the set of newly covered points contributed by l can only shrink as the baseline set grows.

The GREEDYCOVER procedure addresses the *submodular set cover* problem: minimizing the number of sensing locations required to reach coverage $F(\mathcal{V})$. Given that $F(S)$ is an integer-valued monotone submodular function, the result by Wolsey [33] ensures that the greedy solution S satisfies

$$F(S) \leq \text{OPT} \cdot \left(1 + \ln \max_{j \in \{1, \dots, M\}} F(\{j\})\right),$$

where OPT denotes the minimum number of locations required to cover all evaluation points $v_i \in \mathcal{V}$. Thus, the greedy method provides a provably near-optimal selection for minimizing the sensing resources necessary to satisfy the coverage objective.

3) *Path Construction via TSP*: Given the selected index set S , we construct a low-cost route that visits all sensing locations $\{c_j : j \in S\}$. We treat these locations as nodes in a complete Euclidean graph, with edge weights given by pairwise distances between locations. We then invoke a TSP solver (e.g., the Christofides algorithm [31]) to compute a path through the nodes, which we use as the execution path.

While our formulation constrains sensing locations to the valid environment—including non-convex domains—the straight-line edges of a Euclidean TSP path may still traverse infeasible regions, such as obstacles or areas outside the boundary. To ensure feasibility, we post-process the path using standard motion-planning techniques (e.g., grid-based or sampling-based planners [34]) to replace straight-line segments with collision-free, admissible paths.

4) *Limitations of the Decoupled Approach*: This two-stage pipeline—first selecting sensing locations to satisfy the coverage requirement, then routing them with a TSP solver—has an important limitation: it provides *no joint approximation guarantee* for the combined information–travel objective. The submodular approximation guarantees provided by Wolsey [33] apply only to the selection of sensing locations and do not account for path cost. In practice, greedy selection may choose sensing locations that are geographically dispersed, leading to unnecessarily long routes even though other, slightly different selections could satisfy the estimation uncertainty requirements at a lower travel cost. This limitation motivates our next method, which more tightly couples sensing-location selection and routing.

D. GCBCOVER: IPP via Generalized Cost–Benefit Algorithm

Our second planning method couples sensing-location selection and routing via the generalized cost–benefit (GCB) algorithm [35, 36]. Unlike the decoupled approach, which first maximizes coverage and then computes a path, GCB explicitly trades off *marginal coverage gain* against the *marginal increase in path cost* at each iteration.

1) *Generalized Cost–Benefit Selection with Coverage Maps*: We reuse the binary coverage matrix $B \in \{0, 1\}^{M \times N}$ from Section V-B and the coverage function in Equation 8. Let $\text{cost}(S)$ denote the routing cost of a selected index set $S \subseteq \{1, \dots, M\}$, defined as the length of a (near-)shortest route that visits the sensing locations $\{c_j : j \in S\}$.

At iteration t , given the current sensing set S_t with coverage $F(S_t)$ and routing cost $\text{cost}(S_t)$, each remaining candidate $j \notin S_t$ is scored by a cost–benefit ratio constructed from:

(i) the *marginal coverage gain*:

$$\Delta F(j | S_t) := F(S_t \cup \{j\}) - F(S_t),$$

and (ii) the *marginal routing-cost increment*:

$$\Delta \text{cost}(j | S_t) := \text{cost}(S_t \cup \{j\}) - \text{cost}(S_t).$$

The candidate with the largest ratio:

$$\Delta F(j | S_t) / \Delta \text{cost}(j | S_t)$$

is selected, provided it satisfies the routing budget D , effectively prioritizing candidates that yield substantial new coverage for a minimal increase in routing cost.

Because a purely ratio-greedy strategy can be suboptimal under a strict budget, GCB also computes a standard greedy coverage solution and truncates it to satisfy the budget. The algorithm then returns whichever of the budget-feasible solutions attains higher coverage. This comparison step is also necessary for the approximation guarantee stated below. Algorithm 2 in the Appendix summarizes the GCB algorithm.

The worst-case running time is dominated by repeated routing cost evaluations. Let $T(n)$ denote the time required to compute the routing cost for n sensing locations (e.g., via a TSP solver). Using the Christofides algorithm [31], where

$T(n) = \mathcal{O}(n^3)$, the resulting worst-case complexity of GCB is $\mathcal{O}(M^2(N + M^3))$ for $M \leq N$.

In our experiments, we use a faster heuristic TSP routine. Given the current solution path, we estimate the routing cost increase upon adding a candidate sensing location using a fast heuristic (see Algorithm 3 in the Appendix). Once a location is selected via the GCB ratio, we invoke a full TSP solver to refine the path. This hybrid strategy significantly reduces runtime while maintaining high solution quality.

2) *Approximation Guarantees Under Routing Constraints:* As established in Section V-C, the coverage objective F is monotone submodular. Similarly, the routing cost $\text{cost}(\cdot)$ is a monotone set function defined by the length of a TSP path through the selected sensing locations [35]. Consequently, the problem can be framed as a submodular maximization task subject to a routing constraint:

$$\max_{S \subseteq \{1, \dots, M\}} F(S) \quad \text{s.t.} \quad \text{cost}(S) \leq D.$$

The generalized cost–benefit (GCB) algorithm provides a constant-factor approximation guarantee for this class of problems. In particular, under mild conditions on the routing cost function—namely, approximate submodularity with submodularity ratio α_c , bounded curvature κ_c , and a $\psi(n)$ -approximation factor for the TSP solver—the GCB solution satisfies the following bi-criterion guarantee [35]:

$$F(S) \geq \frac{1}{2} \left(1 - \frac{1}{e}\right) \text{OPT},$$

where OPT denotes the optimal coverage achievable under a slightly tighter effective budget (with the tightening depending on α_c , κ_c , $\psi(n)$, and the largest feasible solution size). Equivalently, GCB achieves at least a $\frac{1}{2}(1 - 1/e) \approx 31.6\%$ fraction of an *approximate* optimum, where the approximation reflects the quality of the TSP solver. We refer to Zhang and Vorobeychik [35] for a more detailed explanation of the approximation guarantee.

In our setting, F is exactly monotone submodular; thus, all approximation effects arise from the routing component (i.e., the submodularity ratio/curvature of $\text{cost}(\cdot)$ and the approximation quality of the TSP solver). Consequently, our method inherits the same worst-case approximation factor as the underlying GCB framework while directly optimizing a coverage objective tailored to informative sensing.

VI. EXPERIMENTS

This section evaluates the proposed method through two experimental studies. First, we benchmark our methods against representative baseline approaches using real-world topographic datasets. Second, we validate the practical feasibility of our approach through field deployments.

A. Benchmark Experiments

We benchmarked the proposed methods on the Shuttle Radar Topography Mission (SRTM) dataset [37], which provides high-resolution digital elevation data worldwide. We focused on the region near (45°N, 123°W), chosen for its

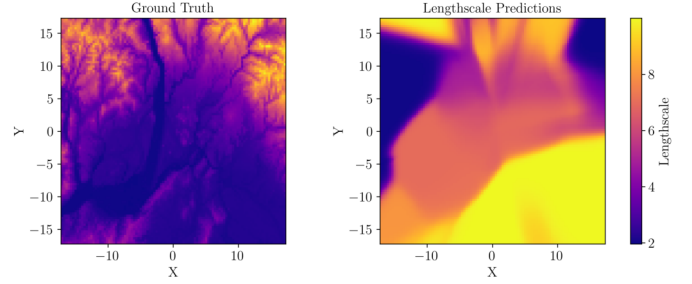


Fig. 4: Left: SRTM ground truth data (45°N, 123°W). Right: Lengthscales from a non-stationary kernel.

complex, non-stationary spatial structure; results for three additional locations are reported in the Appendix. All benchmarks were run on a computer with an Intel i9-14900K CPU and 64 GB of RAM.

To model the environment, we employ a non-stationary Attentive kernel [24], which utilizes a mixture of Radial Basis Function (RBF) kernels to capture spatially varying correlations. We optimize the kernel hyperparameters by fitting a Gaussian process (GP) [7] to data collected along an initial pilot path. To generate this path, we uniformly sample locations in the domain $\mathcal{X}$, cluster them using k -means [38], and select 10 resulting centroids as waypoints. A Traveling Salesperson Problem (TSP) solver [31] is then used to compute a route through them. After collecting 350 measurements along this pilot trajectory to initialize the GP, the learned kernel function and likelihood noise variance are provided to each planner to generate solution paths. Figure 4 shows the SRTM ground truth for (45°N, 123°W), along with the spatially varying lengthscales learned by the Attentive kernel.

We evaluated our proposed methods, GREEDYCOVER (Section V-C) and GCBCOVER (Section V-D), considering GCB-COVER in both unconstrained and budgeted-distance settings. We compared these methods against two established baselines. The first, HEXCOVER [3], generates IPP solutions with estimation uncertainty guarantees. Since HEXCOVER requires a stationary RBF kernel, we set its lengthscale to the minimum value inferred by the trained Attentive kernel—a conservative choice that usually preserves the uncertainty guarantee. The second, CONTINUOUS-SGP [2], is a high-performance IPP method known to outperform mutual-information and variance-based approaches, though it lacks formal uncertainty guarantees.

None of the planning methods explicitly account for measurements collected from the pilot path; during planning, they use only the learned kernel and noise variance. GREEDYCOVER, GCBCOVER, and CONTINUOUS-SGP can be warm-started by seeding the selected sensing locations with already visited points and then planning the remaining route, potentially reducing the number of additional sensing locations (and thus total travel distance). However, HEXCOVER does not support such initialization. For a consistent comparison, we therefore incorporate the pilot measurements only during

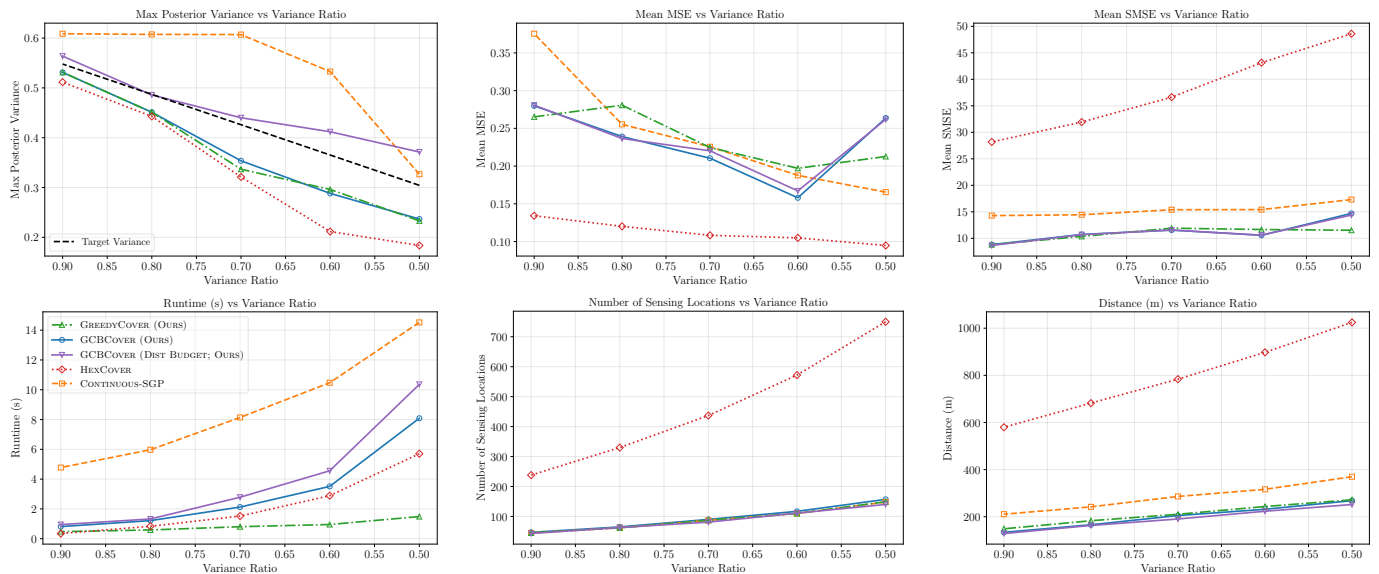


Fig. 5: Benchmark results on the SRTM dataset (45°N , 123°W). Top row: Max posterior variance, MSE, and SMSE. Bottom row: Runtime, sensing locations, and path length. GREEDYCOVER and GCBCOVER satisfy uncertainty targets with fewer sensing locations and shorter paths than HEXCOVER and CONTINUOUS-SGP. In budgeted scenarios, GCBCOVER prioritizes strict adherence to travel limits over further variance reduction. Lower values represent better performance across all metrics.

evaluation, not during planning.

We utilized a GP trained on pilot data to determine the maximum posterior variance across a discrete evaluation grid. We define the target variance threshold as a fixed fraction of this maximum, varying the fraction from 0.9 to 0.5. For the distance-budgeted configuration of GCBCOVER, the budget is defined as the length of its unconstrained solution minus 20 m. This design choice anchors the budget to the physical requirements of the uncertainty constraint while imposing a deliberate shortfall to challenge the planner’s efficiency. Additionally, as CONTINUOUS-SGP requires the number of sensing locations to be specified a priori, we set this value to match the number of locations generated by the GREEDYCOVER solution.

To evaluate each method, we combined measurements from the pilot path with those collected at the solution waypoints and computed six metrics: (i) maximum posterior variance, (ii) mean squared error (MSE) [7], (iii) standardized mean squared error (SMSE) [7], (iv) algorithm runtime, (v) number of selected sensing locations, and (vi) total solution path length. Figure 5 presents the benchmark results for (45°N , 123°W), while Figure 6 illustrates representative solution paths. Additional results are available in the Appendix.

Across the full range of threshold ratios, all methods successfully satisfy the target variance requirements, with two notable exceptions: distance-budgeted GCBCOVER, which prioritizes the travel distance limit over variance reduction, and CONTINUOUS-SGP, which offers no formal guarantees.

Computational efficiency varies significantly across the evaluated methods. GREEDYCOVER is consistently the fastest, outperforming both HEXCOVER and CONTINUOUS-SGP. For HEXCOVER, while sensing-location selection requires less

than one second, the overall runtime is dominated by the TSP solver. This overhead is driven by HEXCOVER selecting a larger number of sensing locations than GREEDYCOVER, which increases the complexity of the subsequent routing. In contrast, GCBCOVER is a little slower due to the tight coupling between its selection and routing phases. However, this integrated approach allows GCBCOVER to enforce strict distance constraints while maintaining near-optimal approximation guarantees for the joint selection and routing problem.

From a deployment perspective, the critical metrics are the number of sensing locations and the total path length. GREEDYCOVER and GCBCOVER consistently satisfy the target variance threshold using fewer sensing locations and shorter travel distances. Although HEXCOVER yields the lowest maximum posterior variance and a corresponding improvement in MSE, this advantage stems primarily from collecting a larger volume of data due to its inability to leverage non-stationary correlations. Conversely, GREEDYCOVER and GCBCOVER utilize the Attentive kernel to model non-stationarity in the underlying field, effectively capturing spatially varying correlations to prioritize the most informative sensing locations.

B. Autonomous Surface Vehicle Field Trial

We validated the proposed method in a real-world deployment using a custom-built autonomous surface vehicle (ASV) [39] to map lake bathymetry. The monitoring region was restricted to a non-convex subset of the lake. The ASV platform (Figure 2) was equipped with GPS, a single-beam sonar, and a Raspberry Pi 5 for onboard computation.

We conducted an IPP trial using GREEDYCOVER. We first generated a pilot path with ten waypoints via k -means and

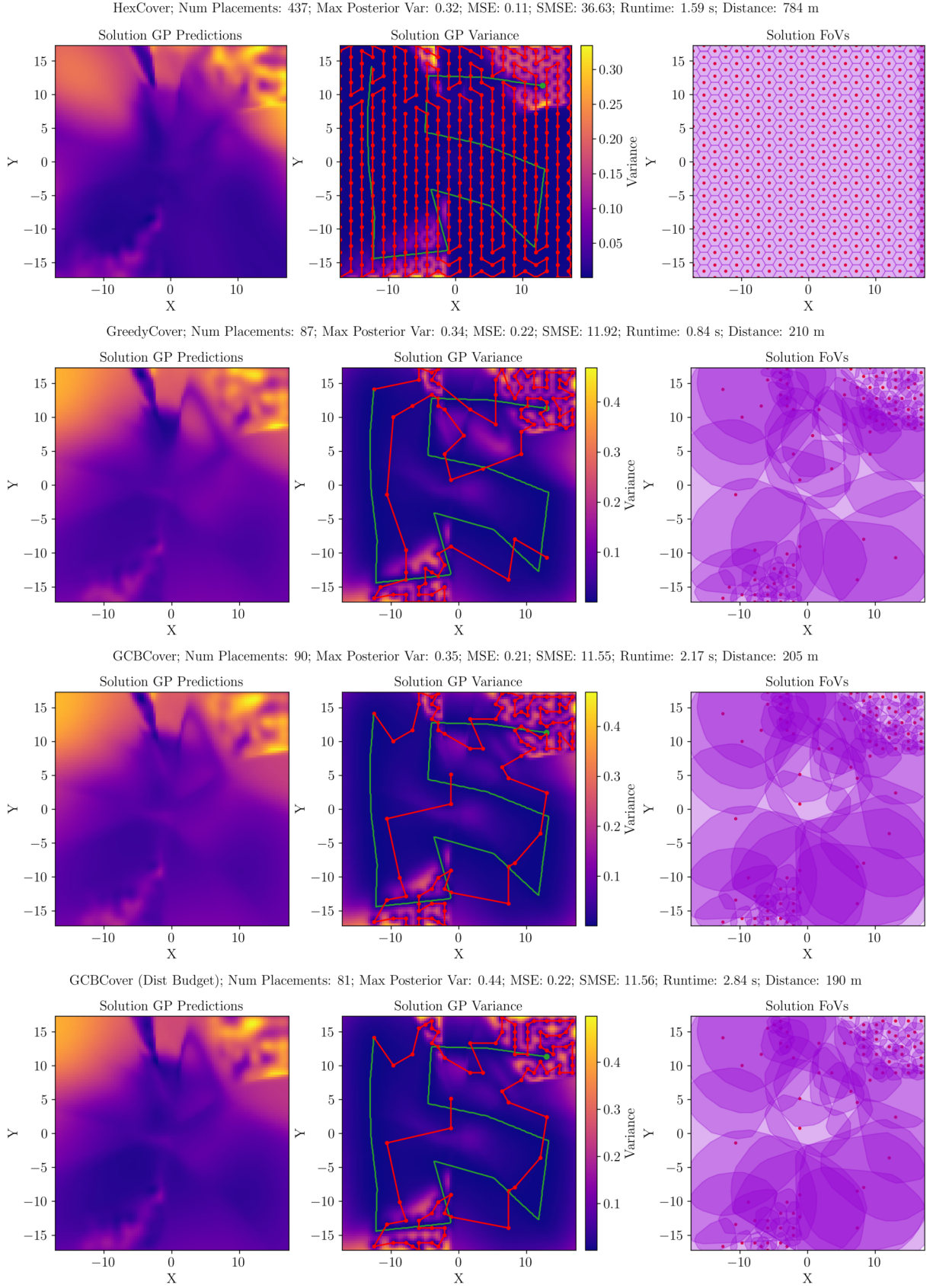


Fig. 6: Comparison of solutions on the SRTM dataset (45°N , 123°W) for a 0.7 variance ratio (target variance threshold: 0.43). Red: informative paths and sensing locations; Green: initial pilot path; Violet: approximate sensor fields-of-view.

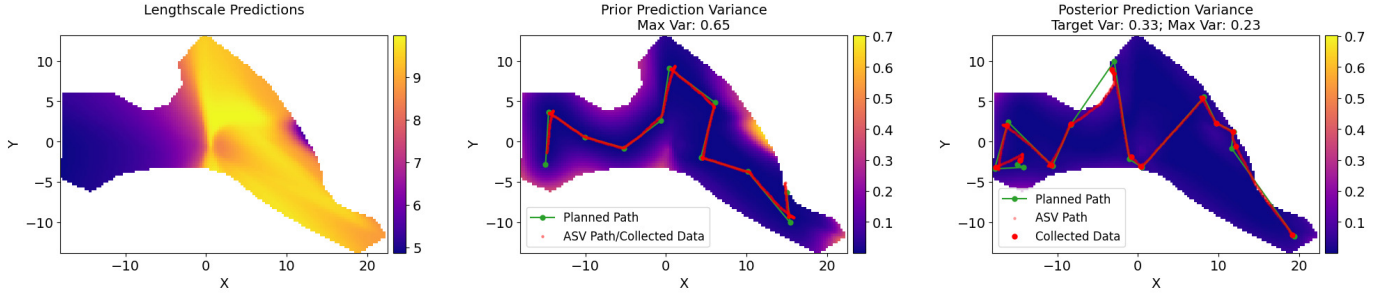


Fig. 7: Autonomous surface vehicle (ASV) field-trial results using GREEDYCOVER. Left: learned lengthscales from the non-stationary kernel. Middle: GP prior prediction variance conditioned on the initial pilot data. Right: GP posterior prediction variance after conditioning on the IPP solution data.

collected 544 depth measurements along this path. Using this data, we fit a GP with an Attentive kernel [24]. We set the target maximum posterior variance to (50%) of the maximum variance obtained after executing the pilot path, and then computed the IPP solution. To accelerate convergence and reduce path length, we warm-started GREEDYCOVER with the pilot data. Consequently, the planned IPP solution avoided revisiting locations near the pilot path and prioritized regions whose uncertainty remained above the target threshold. After executing the IPP solution, we updated the GP using measurements collected *at the visited waypoints* and evaluated the resulting maximum posterior variance. Additional implementation details are provided in the code Appendix. The learned lengthscales, together with the planned and executed pilot and informative paths, are shown in Figure 7.

Despite deviations from the planned path, the method successfully reduced uncertainty such that the maximum-variance threshold was satisfied over the monitoring region. We also observed that the IPP sensing locations remained within the region boundary, and the low-level trajectory planner [40] adjusted the executed path to remain inside the non-convex domain, consistent with the non-convex domain constraints. Details of the ocean field trial with an Aqua2 AUV [41, 42] are provided in the Appendix.

VII. LIMITATIONS

A central limitation is the reliance on a well-specified GP kernel: if the kernel is misspecified, the bound on predictive variance may fail to be a strict lower bound on the true mean-squared prediction error. An important extension is *online* IPP that adapts a non-stationary kernel from streaming data (e.g., POAM [25]) and provides guarantees that are robust to hyperparameter uncertainty. Alternatively, correlations could be inferred from physics-based priors or historical data, reducing—or potentially eliminating—the need for pilot data collection and onboard kernel learning.

Second, our guarantees are enforced on a finite evaluation set $\mathcal{V}$ and therefore may not hold uniformly over the continuous domain. Future work could exploit regularity of the GP variance (e.g., continuity or Lipschitz properties) to guide grid-resolution selection and lift guarantees from $\mathcal{V}$ to the

continuous domain.

Finally, the field trials highlight execution uncertainty (e.g., localization error and imperfect tracking). Incorporating these effects via chance-constrained formulations would further improve real-world reliability.

VIII. CONCLUSION

We presented two planners for *informative path planning (IPP) with guaranteed estimation uncertainty* that compute near-shortest (or budget-feasible) sensing paths whose measurements ensure that the Gaussian process (GP) posterior variance at every evaluation point falls below a user-specified threshold. This provides coverage-style guarantees while leveraging correlation-aware sampling.

To make the problem tractable, we fit a GP to available prior information and construct *binary coverage maps* by thresholding kernel-induced covariance, reducing the global variance requirement to a discrete coverage problem over evaluation points. This reduction enables efficient planning with guarantees: GREEDYCOVER yields a near-optimal approximation for selecting sensing locations, while GCBCOVER balances marginal coverage gain against marginal path length to obtain near-optimal solutions under a strict distance budget. Unlike geometric sweep patterns, both methods adapt sampling to the learned correlation structure.

Our approach leverages *non-stationary* kernels to adaptively sample complex fields, concentrating efforts in high-variation regions while avoiding redundancy in predictable areas. The approach naturally extends to *non-convex* domains containing obstacles. Benchmarks on SRTM topography confirm that GREEDYCOVER and GCBCOVER outperform existing IPP baselines—including those with and without formal uncertainty guarantees—by satisfying uncertainty constraints while minimizing both number of sensing locations and travel distance. Finally, field trials using autonomous surface and underwater vehicles for bathymetry mapping validate the real-world feasibility and robustness of our methods.

Future work includes IPP with guaranteed estimation uncertainty that accommodates vehicle dynamics, motion uncertainty, continuous sensing, and decentralized multi-robot deployments that maintain robustness in large-scale, communication-limited settings.

ACKNOWLEDGEMENTS

We thank Pratap Tokekar and Shamak Dutta for sharing code for the HexCover method, and Weizhe Chen for assistance with the attentive kernel code. We are grateful to Artur Wolek for providing access to the ASV platform, and we thank Ninh Nguyen for support during the ASV field trial. We also thank Bennett Carley and Jeremy Mallette for assistance with Aqua2 setup, and Elias Sokolova, Jalil Francisco Chavez Galaviz, and Chelsey Edge for their help deploying the Aqua2 AUV in the ocean. Finally, we thank Edwin Meriaux and Shuo Wen for capturing aerial footage of the AUV experiment.

REFERENCES

- [1] Howie Choset and Philippe Pignon. [Coverage Path Planning: The Boustrophedon Cellular Decomposition](#). In *Field and Service Robotics*, pages 203–209, London, 1998. Springer London.
- [2] Kalvik Jakkala and Srinivas Akella. [Fully differentiable sensor placement and informative path planning](#). *The International Journal of Robotics Research*, 2025.
- [3] Shamak Dutta, Nils Wilde, Pratap Tokekar, and Stephen L. Smith. [Approximation Algorithms for Robot Tours in Random Fields with Guaranteed Estimation Accuracy](#). In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 7830–7836, 2023.
- [4] Gary M. Lovett, Douglas A. Burns, Charles T. Driscoll, Jennifer C. Jenkins, Myron J. Mitchell, Lindsey Rustad, James B. Shanley, Gene E. Likens, and Richard Haeuber. [Who needs environmental monitoring?](#) *Frontiers in Ecology and the Environment*, 5(5):253–260, 2007.
- [5] Nare Karapetyan, Jason Moulton, Jeremy S. Lewis, Alberto Quattrini Li, Jason M. O’Kane, and Ioannis M. Rekleitis. [Multi-robot Dubins Coverage with Autonomous Surface Vehicles](#). *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2373–2379, 2018.
- [6] Saurav Agarwal and Srinivas Akella. [Area Coverage With Multiple Capacity-Constrained Robots](#). *IEEE Robotics and Automation Letters*, 7(2):3734–3741, 2022.
- [7] Carl Edward Rasmussen and Christopher K. I. Williams. [Gaussian Processes for Machine Learning](#). MIT Press, Cambridge, USA, 2005.
- [8] Varun Suryan and Pratap Tokekar. [Learning a Spatial Field in Minimum Time With a Team of Robots](#). *IEEE Transactions on Robotics*, 36(5):1562–1576, 2020.
- [9] H. González-Banos and Jean-Claude Latombe. [A Randomized Art-Gallery Algorithm for Sensor Placement](#). In *Proceedings of the Seventeenth Annual Symposium on Computational Geometry*, page 232–240, New York, NY, USA, 2001.
- [10] Andreas Breitenmoser, Mac Schwager, Jean-Claude Metzger, Roland Siegwart, and Daniela Rus. [Voronoi coverage of non-convex environments with a group of networked robots](#). In *2010 IEEE International Conference on Robotics and Automation*, pages 4982–4989, 2010.
- [11] Mac Schwager, Michael P. Vitus, Samantha Powers, Daniela Rus, and Claire J. Tomlin. [Robust Adaptive Coverage Control for Robotic Sensor Networks](#). *IEEE Transactions on Control of Network Systems*, 4(3), 2017.
- [12] Amanda Bouman, Joshua Ott, Sung-Kyun Kim, Kenny Chen, Mykel J. Kochenderfer, Brett Lopez, Ali-akbar Agha-mohammadi, and Joel Burdick. [Adaptive Coverage Path Planning for Efficient Exploration of Unknown Environments](#). In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 11916–11923, 2022.
- [13] Andreas Krause, Ajit Singh, and Carlos Guestrin. [Near-Optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies](#). *Journal of Machine Learning Research*, 9(8):235–284, 2008.
- [14] Chandra Chekuri and M. Pal. [A recursive greedy algorithm for walks in directed graphs](#). In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS’05)*, pages 245–253, 2005.
- [15] Lorenzo Bottarelli, Manuele Bicego, Jason Blum, and Alessandro Farinelli. [Orienteering-based informative path planning for environmental monitoring](#). *Engineering Applications of Artificial Intelligence*, 77:46 – 58, 2019.
- [16] Jonathan Binney, Andreas Krause, and Gaurav S. Sukhatme. [Optimizing waypoints for monitoring spatiotemporal phenomena](#). *The International Journal of Robotics Research*, 32(8):873–888, 2013.
- [17] Gregory Hitz, Enric Galceran, Marie-Eve Garneau, François Pomerleau, and Roland Siegwart. [Adaptive continuous-space informative path planning for online environmental monitoring](#). *Journal of Field Robotics*, 34(8):1427–1449, 2017.
- [18] Gilad Francis, Lionel Ott, Roman Marchant, and Fabio Ramos. [Occupancy map building through Bayesian exploration](#). *The International Journal of Robotics Research*, 38(7):769–792, 2019.
- [19] Joshua Ott, Mykel J. Kochenderfer, and Stephen Boyd. [Approximate sequential optimization for informative path planning](#). *Robotics and Autonomous Systems*, 182: 104814, 2024.
- [20] Ananya Rao, Howie Choset, and David Wettergreen. [Wavelet-Based Distributed Coverage for Heterogeneous Agents](#). In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6630–6637, 2025.
- [21] Julius Rücker, Liren Jin, and Marija Popović. [Adaptive Informative Path Planning Using Deep Reinforcement Learning for UAV-based Active Sensing](#). In *2022 International Conference on Robotics and Automation (ICRA)*, pages 4473–4479, 2022.
- [22] Jose Fuentes, Paulo Padrão, Abdullah Al Redwan Newaz, and Leonardo Bobadilla. [Learning-Based Adaptive Navigation for Scalar Field Mapping and Feature Tracking](#). In *2025 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1–7, 2025.
- [23] Sandeep Manjanna and Gregory Dudek. [Data-driven](#)

- selective sampling for marine vehicles using multi-scale paths. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017.
- [24] Weizhe Chen, Roni Khardon, and Lantao Liu. [AK: Attentive Kernel for Information Gathering](#). In *Proceedings of Robotics: Science and Systems*, New York City, NY, USA, June 2022.
- [25] Weizhe Chen, Lantao Liu, and Roni Khardon. [POAM: Probabilistic Online Attentive Mapping for Efficient Robotic Information Gathering](#). In *Robotics: Science and Systems*, 2024.
- [26] Kalvik Jakkala and Srinivas Akella. [Multi-Robot Informative Path Planning from Regression with Sparse Gaussian Processes](#). In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024.
- [27] Kalvik Jakkala and Srinivas Akella. [Fully Differentiable Adaptive Informative Path Planning](#). In *2025 IEEE International Conference on Robotics and Automation*, pages 5431–5437, Atlanta, GA, May 2025.
- [28] Kalvik Jakkala and Srinivas Akella. [Probabilistic Gas Leak Rate Estimation Using Submodular Function Maximization With Routing Constraints](#). *IEEE Robotics and Automation Letters*, 7(2):5230–5237, 2022.
- [29] Johan Wagberg, Dave Zachariah, Thomas Schon, and Petre Stoica. [Prediction Performance After Learning in Gaussian Process Regression](#). In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1264–1272. PMLR, 20–22 Apr 2017.
- [30] Marc Härkönen, Markus Lange-Hegermann, and Bogdan Raită. [Gaussian Process Priors for Systems of Linear Partial Differential Equations with Constant Coefficients](#). In *Proceedings of the 40th International Conference on Machine Learning, ICML*. JMLR, 2023.
- [31] Nicos Christofides. [Worst-Case Analysis of a New Heuristic for the Travelling Salesman Problem](#). *Operations Research Forum*, 3(1):20, Mar 2022.
- [32] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. [An analysis of approximations for maximizing submodular set functions-I](#). *Mathematical Programming*, 14(1):265–294, Dec 1978.
- [33] Laurence A. Wolsey. [An analysis of the greedy algorithm for the submodular set covering problem](#). *Combinatorica*, 2:385–393, 1982.
- [34] K.M. Lynch and F.C. Park. *Modern Robotics*. Cambridge University Press, 2017.
- [35] Haifeng Zhang and Yevgeniy Vorobeychik. [Submodular Optimization with Routing Constraints](#). In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, page 819–825, 2016.
- [36] Pao-Te Lin and Kuo-Shih Tseng. [Improvement of Submodular Maximization Problems With Routing Constraints via Submodularity and Fourier Sparsity](#). *IEEE Robotics and Automation Letters*, 8(4):1927–1934, 2023.
- [37] Tom G Farr, Paul A Rosen, Edward Caro, Robert Crippen, Riley Duren, Scott Hensley, Michael Kobrick, Mimi Paller, Ernesto Rodriguez, Ladislav Roth, et al. [The Shuttle Radar Topography Mission](#). *Reviews of geophysics*, 45(2), 2007.
- [38] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer New York, 2006.
- [39] Michael Brancato and Artur Wolek. [Adaptive Sampling of a Stationary Gaussian Spatial Process by a Team of Robots With Heterogeneous Dynamics and Measurement Noise Variance](#). *IEEE Access*, 12:94407–94423, 2024.
- [40] Anders Strand-Holm Vinther and Anders Strand-Holm Magnus. [Pathfinding in Two-dimensional Worlds](#). Master’s thesis, Aarhus University, 2015.
- [41] Gregory Dudek, Philippe Giguere, Chris Prahacs, Shane Saunderson, Junaed Sattar, Luz-abril Torres-Mendez, Michael Jenkin, Andrew German, Andrew Hogue, Arlene Ripsman, Jim Zacher, Evangelos Milios, Hui Liu, Pifu Zhang, Marti Buehler, and Christina Georgiades. [AQUA: An Amphibious Autonomous Robot](#). *Computer*, 40(1): 46–53, 2007.
- [42] Edwin Meriaux, Shuo Wen, David Widhalm, Zhizun Wang, Junming Shi, Mariana Sosa Guzm’an, Kalvik Jakkala, Bennett A. Carley, Elias Sokolova, Yogesh Girdhar, Monika Roznere, Jason O’Kane, Junaed Sattar, and Gregory Dudek. [AMP2026: A Multi-Platform Marine Robotics Dataset for Tracking and Mapping](#). In *23rd Conference on Robots and Vision*, 2026.
- [43] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications and Signal Processing. Wiley, 1991.