Adaptive Monitoring of Stochastic Fire Front Processes via Information-seeking Predictive Control

Savvas Papaioannou, Panayiotis Kolios, Christos G. Panayiotou and Marios M. Polycarpou

***Abstract*— We consider the problem of adaptively monitoring a wildfire front using a mobile agent (e.g., a drone), whose trajectory determines where sensor data is collected and thus influences the accuracy of fire propagation estimation. This is a challenging problem, as the stochastic nature of wildfire evolution requires the seamless integration of sensing, estimation, and control, often treated separately in existing methods. State-of-the-art methods either impose linear-Gaussian assumptions to establish optimality or rely on approximations and heuristics, often without providing explicit performance guarantees. To address these limitations, we formulate the fire front monitoring task as a stochastic optimal control problem that integrates sensing, estimation, and control. We derive an optimal recursive Bayesian estimator for a class of stochastic nonlinear elliptical-growth fire front models. Subsequently, we transform the resulting nonlinear stochastic control problem into a finite-horizon Markov decision process and design an information-seeking predictive control law obtained via a lower confidence bound-based adaptive search algorithm with asymptotic convergence to the optimal policy.**

I. Introduction

The 2025 Southern California wildfires were among the most devastating in the state's history: the Palisades Fire in Los Angeles and the Eaton Fire in Altadena burned nearly 40,000 acres, destroyed over 16,000 structures, and displaced hundreds of thousands of people [1]. Accurate estimation of wildfire propagation is therefore critical for effective disaster response [2]–[4] and informed decision-making [5], [6]. Motivated by this need, this work investigates adaptive monitoring of a stochastic wildfire front using a mobile agent.

The task requires planning the agent's trajectory over a rolling finite horizon to minimize uncertainty in estimating the fire's evolution from sensor data. At each step, the agent re-plans based on the current environment, yielding a complex problem that demands an integrated approach to sensing [7]–[9], estimation [10]–[12] , and control [13]–[15]. Existing methods often address only parts of this problem, either by decoupling sensing, estimation , and control, or by simplifying assumptions [16], [17]. The problem considered in this work relates to *active sensing* [18], *information gathering* [19], and *sensor management* [20]. Sensor management approaches typically address stateless sensors without dynamics, focusing on placement or node selection, and are therefore limited in scenarios where sensor states evolve (e.g., drones equipped with onboard sensors). In such dynamic settings, adaptive control strategies are required. For stochastic dynamic processes, many methods adopt myopic control [21], whereas non-myopic schemes are typically greedy [22] or heuristic with sub-optimality guarantees [23]. Approaches that provide optimality guarantees often assume linear dynamics with Gaussian noise and apply the certainty equivalence principle (CEP), thereby reducing the problem to deterministic optimal control [24]. These assumptions, however, do not hold for the problem addressed in this work.

In summary, this work integrates sensing, estimation, and control within a unified stochastic optimal control (SOC) framework for adaptive wildfire-front monitoring using a mobile agent. We develop a recursive Bayesian estimator for elliptical fire-front dynamics under limited sensing and uncertainty, and reformulate the nonlinear SOC problem as a finite-horizon Markov decision process (MDP). The MDP is solved via a lower-confidence-bound (LCB) guided adaptive search that asymptotically converges to the optimal policy.

II. Problem Formulation

A. Problem Objective

Let a mobile agent be at state y_t at time step t, and let the agent's belief distribution over the fire front state X_t be denoted by $\mathcal{B}_t(X_t|Z_{1:t})$ (abbreviated as $\mathcal{B}_t$) which was computed using measurements $Z_{1:t} = [Z_1, \dots, Z_t]$ up to time step t. The objective is to compute the optimal sequence of control inputs $\{u_{t|t}, \dots, u_{t+T-1|t}\}$ over a finite rolling planning horizon of T time steps that minimizes:

$$\mathbb{E}_{Z_t^{1:T}} \left\{ \sum_{\tau=1}^{T} \mathcal{C}_\tau \left(\mathcal{B}^-_{t+\tau|t}, Z_{t+\tau|t}(u_{t+\tau-1|t}) \right) \right\}. \quad (1)$$

Here, the subscript $t+\tau|t$ denotes predicted quantities at time step $t+\tau$ for $\tau \in \{1, \dots, T\}$ within the planning horizon, based on information available at time step t. The cost function $\mathcal{C}_\tau$ is a bounded, real-valued function that takes as input: a) the predictive density $\mathcal{B}^-_{t+\tau|t}$ i.e., $\mathcal{B}^-_{t+\tau|t}(X_{t+\tau|t}|Z_{1:t+\tau-1|t})$, of the fire front state at time $t+\tau|t$, and b) the future i.e., predicted, measurement $Z_{t+\tau|t}(u_{t+\tau-1|t})$, which will be received given that the agent executes the control input $u_{t+\tau-1|t}$ and moves to the predicted state $y_{t+\tau|t}$. It then returns a value representing the

The authors are with the KIOS Research and Innovation Centre of Excellence (KIOS CoE) and the Department of Electrical and Computer Engineering, University of Cyprus, Nicosia, 1678, Cyprus. E-mail:{papaioannou.savvas, pkolios, christosp, mpolycar}@ucy.ac.cy

This work is supported by the European Union's Horizon Europe program under grant agreement No 101187121 (EUSOME) and the Civil Protection Knowledge for Action in Prevention & Preparedness under grant agreement No. 101193719 (COLLARIS2). It is also supported from the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

uncertainty of the fire front state captured in the resulting (pseudo) posterior belief $\mathcal{B}_{t+\tau|t}$.

An information-rich measurement set is one that reduces the dispersion in the posterior, which is the agent's objective over the planning horizon. The expectation is taken with respect to the future measurement set $Z_t^{1:T} = \{Z_{t+1|t}, \ldots, Z_{t+T|t}\}$. Subsequently, the agent executes the first control input in the sequence, i.e., $u_{t|t}$, transitions to its new state y_{t+1}, receives the real measurement Z_{t+1}, computes the posterior belief $\mathcal{B}_{t+1}(X_{t+1}|Z_{1:t+1})$, and repeats the process described above for time step $t+1$.

B. Fire Front Propagation Model

In this work, as we discuss next, we employ a stochastic adaptation of the deterministic elliptical fire propagation model proposed in [25]. This model, which is currently utilized in various fire-area simulators [26], describes the spatiotemporal evolution of a fire front using a nonlinear system of first-order differential equations. Specifically, the fire front is represented as an ellipse defined by a series of N vertices that collectively delineate the propagating fire's edge at a specific moment. The spatiotemporal discrete-time dynamics of vertex $i \in \{1, \ldots, N\}$ at time step t are given by:

$$x_t^i = x_{t-1}^i + \Delta t\, \dot{x}_{t-1}^i, \tag{2}$$

where $x_t^i = [\mathrm{x}_t^i, \mathrm{y}_t^i]^\top \in \mathbb{R}^2$ is the state of vertex i composed of 2D Cartesian coordinates, Δt is the sampling interval, and the fire growth velocity at vertex i is given by $\dot{x}_{t-1}^i =$

$$\begin{bmatrix} \dfrac{\alpha_1^2(i)\cos(\theta(i))\,SC(i) - \alpha_2^2(i)\sin(\theta(i))\,CS(i)}{\sqrt{\alpha_2^2(i)CS(i)^2 + \alpha_1^2(i)SC(i)^2}} + C_1(i) \\ \dfrac{-\alpha_1^2(i)\sin(\theta(i))\,SC(i) - \alpha_2^2(i)\cos(\theta(i))\,CS(i)}{\sqrt{\alpha_2^2(i)CS(i)^2 + \alpha_1^2(i)SC(i)^2}} + C_2(i) \end{bmatrix},$$

where:

$$\begin{aligned} SC(i) &= \mathrm{x}_s^i \sin(\theta(i)) + \mathrm{y}_s^i \cos(\theta(i)), \\ CS(i) &= \mathrm{x}_s^i \cos(\theta(i)) - \mathrm{y}_s^i \sin(\theta(i)), \\ C_1(i) &= \alpha_3(i)\, \sin(\theta(i)), \\ C_2(i) &= \alpha_3(i)\, \cos(\theta(i)), \end{aligned} \tag{3}$$

and $[\mathrm{x}_s^i, \mathrm{y}_s^i]^\top$ are the components of the tangent vector at vertex i, providing the local orientation of the fire front at that point.

Environmental conditions, such as fuel type and weather, local to each vertex, affect the forward fire propagation rate and direction. These factors include wind direction and speed, denoted by θ and w_s, respectively, as well as the fire spread rate due to fuel type, denoted by r_f. These parameters are stochastic and can vary throughout the environment. Therefore, each vertex may be affected differently depending on the fire front's extent and the environmental variability.

Specifically, we denote $\theta(i) \in [0, 2\pi]$ as the wind direction affecting vertex i, the wind speed at the location of vertex i as $w_s(i) \in \mathbb{R}^+$, and the fire spread rate as $r_f(i) \in \mathbb{R}^+$. Here, $\theta(i), w_s(i), r_f(i)$, $\forall i$, are random realizations of the wind direction, wind speed, and fire spread rate at the location of vertex i.

The parameters $\alpha_1(i)$, $\alpha_2(i)$, and $\alpha_3(i)$ denote the shape parameters governing the elliptical fire growth from vertex i, representing respectively the lengths of the semi-minor axis, the semi-major axis, and the distance from the ignition point to the center of the ellipse, defined respectively as:

$$\begin{aligned} \alpha_1(i) &= \frac{r_f(i) + \dfrac{r_f(i)}{HB(i)}}{2\,LB(i)}, \\ \alpha_2(i) &= \frac{r_f(i) + \dfrac{r_f(i)}{HB(i)}}{2}, \\ \alpha_3(i) &= \alpha_2(i) - \frac{r_f(i)}{HB(i)}. \end{aligned} \tag{4}$$

where $HB(i)$ is the head-to-back ratio accounting for the difference between the fire's forward (head) and backward (back) spread from the ignition point, while $LB(i)$ is the length-to-breadth ratio which determines the overall elongation of the fire's elliptical shape. These are defined as:

$$\begin{aligned} HB(i) &= \frac{LB(i) + \big(LB(i)^2 - 1\big)^{0.5}}{LB(i) - \big(LB(i)^2 - 1\big)^{0.5}}, \\ LB(i) &= 0.936\, \exp\big(0.2566\, w_s(i)\big) \\ &\quad + 0.461\, \exp\big(-0.1548\, w_s(i)\big) - 0.397. \end{aligned} \tag{5}$$

For a more in-depth description these parameters we refer the reader to [26]. Subsequently, the propagation of the fire front process $X_t = [x_t^1, \ldots, x_t^N]^\top \in \mathcal{X}$ is more compactly expressed as:

$$X_t = \xi(X_{t-1}, E_{t-1}), \tag{6}$$

where $E_t \sim P_E$, with $E_t \in [0, 2\pi]^N \times [0, \infty)^N \times [0, \infty)^N$, denotes a random realization of $\{\theta(i), w_s(i), r_f(i)\}_{i=1}^N$ drawn from the PDF P_E, which captures the stochasticity of the fire front propagation acting as a stationary process noise.

C. Agent Dynamics and Sensing Model

An autonomous mobile agent represented by a point-mass object, evolves inside a bounded planar environment $\mathcal{E} \subset \mathbb{R}^2$ according to discrete-time dynamics of the form [27], [28]:

$$y_t = f_a(y_{t-1}, u_{t-1}) \tag{7}$$

where $y_t \in \mathcal{Y}$ is the state of agent at time t, and $u_t \in \mathcal{U}$ is the control input. In addition, the agent has a finite sensing range for observing its surroundings (i.e., through a camera), which is given by a circular region with radius R_a i.e., $O_t = \{x \in \mathbb{R}^2 \mid \|x - y_t^p\| \leq R_a\}$, where y_t^p is the agent's position at time t.

The agent uses its camera to observe the state of the fire front, i.e., by taking snapshots and determining the location of the fire front from the image snapshots using image processing (i.e., object detection). Due to sensing and image processing imperfections, this process carries a certain degree of inaccuracy, resulting in noisy observations. Specifically,

for fire front vertex i with true state x_t^i, the agent observes the measurement z_t^i inside its sensing range according to:

$$z_t^i = h(x_t^i) + w_t^i, \tag{8}$$

where $h(\cdot)$ is a function that relates the true states to the received measurements, and $w_t^i \sim \mathcal{N}(0, \sigma_z^2 I_{2\times 2})$ represents measurement noise. The noise is independent and identically distributed (i.i.d.) according to a zero-mean Gaussian distribution with variance σ_z^2, where $I_{2\times 2}$ is the 2×2 identity matrix. Additionally, w_t^i is independent of the process noise described in the previous section.

The object detection algorithm often produces multiple fragmented pixel blobs for the same fire-front vertex, so that x_t^i is associated with the measurement vector $[z_t^{i,1}, \ldots, z_t^{i,n_t^i}]$. The number of such blobs depends on the true vertex location, and is modeled as a Poisson random variable with rate $\lambda_t^i(x_t^i)$. Thus, a vertex x_t^i may yield $n_t^i \sim \text{Pois}(\lambda_t^i(x_t^i))$ detections within the sensing range. The resulting measurement set follows a Poisson point process [29] with intensity

$$\gamma_t^i(z|x_t^i, y_t) = \lambda_t^i(x_t^i)\, p(z|x_t^i), \quad x_t^i \in O_t, \tag{9}$$

and $\gamma_t^i(z|x_t^i, y_t) = 0$ otherwise, where $p(z|x_t^i) = \mathcal{N}(z; h(x_t^i), \sigma_z^2 I_{2\times 2})$ is the normalized measurement likelihood restricted to O_t.

III. Adaptive Monitoring via Information-seeking Predictive Control

A. Fire Front Recursive State Estimation

Bayesian recursive state estimation for systems with fixed-dimensional state and measurement vectors is formulated through the predictor-corrector recursion:

$$\mathcal{B}_t^-(x_t|z_{1:t-1}) = \int p(x_t|x_{t-1})\, \mathcal{B}_{t-1}(x_{t-1}|z_{1:t-1})\, dx_{t-1},$$
$$\mathcal{B}_t(x_t|z_{1:t}) = \frac{p(z_t|x_t)\, \mathcal{B}_t^-(x_t|z_{1:t-1})}{\int p(z_t|x_t)\, \mathcal{B}_t^-(x_t|z_{1:t-1})\, dx_t}, \tag{10}$$

where with slight abuse of notation in Eq. (10) $x_t \in \mathbb{R}^{d_x}$ is the state of the system, $z_t \in \mathbb{R}^{d_z}$ is the received measurement, $p(x_t|x_{t-1})$ is the transitional density governed by the stochastic process dynamics, $p(z_t|x_t)$ is the measurement likelihood function, $\mathcal{B}_t^-(x_t|z_{1:t-1})$ is the predictive belief distribution at time t, and $\mathcal{B}_t(x_t|z_{1:t})$ is the posterior belief of x_t when all measurements $z_{1:t} = [z_1, \ldots, z_t]$ up to time t have been received. Subsequently, given the recursion in Eq. (10) the minimum mean square estimator (MMSE) $\hat{x}_t^{\text{MMSE}}$ is given by:

$$\hat{x}_t^{\text{MMSE}} = \int x_t\, \mathcal{B}_t(x_t|z_{1:t})\, dx_t. \tag{11}$$

However, in our problem, at time t the observation is a point pattern of random cardinality, not a fixed-length vector, hence the standard vector-likelihood underlying Eq. (10) is not directly applicable. We therefore replace the likelihood in Eq. (10) with the appropriate set-likelihood and apply Bayes' rule with that form. To achieve this, first observe that the transitional density $p(x_t|x_{t-1})$ in Eq. (10) becomes:

$$p(X_t|X_{t-1}) = \int \delta(X_t - \xi(X_{t-1}, E_{t-1}))\, P_E(E_{t-1})\, dE_{t-1},$$

as direct consequence of the fire front stochastic dynamics, where $\delta(\cdot)$ is the Dirac delta function. Subsequently, we can compute the predicted belief $\mathcal{B}_t^-(X_t|Z_{1:t-1})$ at time step t, assuming the posterior density at the previous time step, $\mathcal{B}_{t-1}(X_{t-1}|Z_{1:t-1})$, is known. The posterior belief $\mathcal{B}_t(X_t|Z_{1:t})$ can then be obtained by incorporating the measurement set Z_t via the correction step shown in Eq. (10), provided that an expression for the measurement likelihood function $p(Z_t|X_t, y_t)$ is available. This process can then be recursively applied to the next time step.

Proposition: *Let $X_t = [x_t^1, \ldots, x_t^N]^\top$ be the state of the fire front at time t, and let $Z_t = [z_t^1, \ldots, z_t^{m_t}]$ denote the received measurement vector at time step t. The likelihood of Z_t given X_t and y_t is given by: $p(Z_t|X_t, y_t) =$*

$$\frac{1}{m_t!} \exp\left(-\sum_{i=1}^{N} \lambda_i(x_t^i)\right) \prod_{k=1}^{m_t} \left(\sum_{i=1}^{N} \gamma_t^i(z_k|x_t^i, y_t)\right), \tag{12}$$

where $\lambda_t^i(x_t^i) = \int_{O_t} \gamma_t^i(z\,|\,x_t^i, y_t)\, dz$ is the expected number of detections from vertex i, $\gamma_t^i(z\,|\,x_t^i, y_t)$ is the Poisson intensity corresponding to vertex i within the sensing region O_t as defined in Eq. (9), and $m_t!$ denotes the factorial of m_t. Consequently, $p(Z_t|X_t, y_t)$ can be used directly in Eq. (10) enabling the handling of multiple objects and multiple measurements without requiring explicit measurement-to-object association, and allowing the computation of the MMSE as discussed previously.

Proof: Due to the independence of noise realizations in the measurement process, the measurements are conditionally independent given the fire front state X_t. As a result, the Poisson processes defined by Eq. (9) are themselves independent, which implies that the combined set of all measurements generated by all processes forms a superposition of Poisson point processes, with total intensity $\Gamma_t(z|X_t, y_t) = \sum_{i=1}^{N} \gamma_t^i(z|x_t^i, y_t)$, and probability density function (PDF) $p(z|X_t) = \Gamma_t(z|X_t, y_t) \left(\int_{O_t} \Gamma_t(z|X_t, y_t)\, dz\right)^{-1} = \Gamma_t(z|X_t, y_t)\Lambda_t^{-1}$. Subsequently, the joint likelihood $p(Z_t|X_t, y_t)$ of receiving m_t measurements at time t can be decomposed as $p(Z_t|X_t, y_t) = p_m(m_t)\, p(z_t^1, \ldots, z_t^{m_t}|m_t, X_t, y_t)$, where $p_m(m_t|y_t)$ is the probability of receiving m_t observations inside the sensing range, and $p(z_t^1, \ldots, z_t^{m_t}|m_t, X_t, y_t)$ is the conditional joint likelihood function. The term $p_m(m_t, |y_t)$ follows a Poisson distribution with parameter Λ_t, i.e., $p_m(m_t|y_t) = {m_t!}^{-1} \exp(-\Lambda_t)\Lambda_t^{m_t}$. In addition, the joint likelihood decomposes as $p(z_t^1, \ldots, z_t^{m_t}|m_t, X_t, y_t) = \prod_{k=1}^{m_t} p(z_k|X_t, y_t)$, and thus $\prod_{k=1}^{m_t} p(z_k|X_t, y_t) = \prod_{k=1}^{m_t} \Gamma_t(z_k|X_t, y_t)\Lambda_t^{-1}$. Since each vertex only contributes locally to its own measurement process, it follows that $\prod_{k=1}^{m_t} \Gamma_t(z_k|X_t, y_t)\Lambda_t^{-1} = \prod_{k=1}^{m_t} \left(\sum_{i=1}^{N} \gamma_t^i(z_k|x_t^i, y_t)\Lambda_t^{-1}\right)$. Since $\int_{O_t} \Gamma_t(z|X_t, y_t)dz = \sum_{i=1}^{N} \int_{O_t} \gamma_t^i(z|x_t^i, y_t)dz = \sum_{i=1}^{N} \lambda_t^i(x_t^i)$, the result in Eq. (12) follows directly. ∎

B. Information-seeking Predictive Control

The problem in Sec. II-A is addressed via the information-seeking predictive controller shown in Problem (P1), formulated as a receding horizon SOC problem. The goal is to compute control inputs $\{u_{t+\tau-1|t}\}_{\tau=1}^{T}$ that optimize the agent's sensing behavior by minimizing the cumulative uncertainty in the (pseudo) posterior beliefs of the fire front states, as defined in Eq. (13a). At each time step t, only the first control input $u_{t|t}$ is applied, and the process is repeated over a shifted horizon.

Problem (P1): *Information-seeking Predictive Control*

$$\min_{\{u_{t+\tau-1|t}\}_{\tau=1}^{T}} \mathbb{E}_{Z_t^{1:T}} \left\{ \sum_{\tau=1}^{T} \nu^{\tau} \mathcal{C}_{\tau}\big(\mathcal{B}_{t+\tau|t}(\cdot|Z_{1:t+\tau|t})\big) \right\} \tag{13a}$$

subject to:

$$\mathcal{B}^{-}_{t+\tau|t} = \int p(X_{\tau}|X_{\tau-1})\, \mathcal{B}_{t+\tau-1|t}(X_{\tau-1}|\cdot)\, dX_{\tau-1}, \tag{13b}$$

$$\mathcal{B}_{t|t} = \mathcal{B}_{t|t-1}, \tag{13c}$$

$$y_{t+\tau|t} = f_a(y_{t+\tau-1|t}, u_{t+\tau-1|t}), \tag{13d}$$

$$y_{t|t} = y_{t|t-1}, \tag{13e}$$

$$\mathcal{B}_{t+\tau|t}(\cdot|Z_{1:t+\tau|t}) \propto p\left(Z_{t+\tau|t}|X_{\tau}, y_{t+\tau|t}\right) \mathcal{B}^{-}_{t+\tau|t}, \tag{13f}$$

$$y_t \in \mathcal{Y}, u_t \in \mathcal{U}, X_t \in \mathcal{X}, Z_t \in \mathcal{Z}, \tag{13g}$$

$$\mathcal{C}_{\tau} \in [0,1], \nu \in (0,1], \tau = \{1, \ldots, T\}. \tag{13h}$$

For each prediction time step $\tau \in \{1, \ldots, T\}$ in the horizon, the agent predicts the fire front state X_{τ} forward in time using the Bayesian prediction step, based on the transition density $p(X_{\tau}|X_{\tau-1})$ and the (pseudo) posterior belief from the previous time step, $\mathcal{B}_{t+\tau-1|t}(X_{\tau-1}|Z_{1:t+\tau-1|t})$, as shown in Eqs. (13b)-(13c). The constraints in Eqs. (13d)-(13e) arise from the agent dynamics, which are assumed to be deterministic in this work. At the predicted time step τ, given the agent's predicted state $y_{t+\tau|t}$ and sensing range $O_{t+\tau|t}$, the agent receives the predicted measurement set $Z_{t+\tau|t}$. This set is then used to compute the posterior belief $\mathcal{B}_{t+\tau|t}(X_{\tau}|Z_{1:t+\tau|t})$ via the Bayesian correction step, as shown in Eq. (13f), using the joint likelihood function $p(Z_{t+\tau|t}|X_{\tau}, y_{t+\tau|t})$ and the predicted density. This posterior distribution subsequently becomes the prior for the next prediction step, continuing the recursive process. The predicted measurements $Z_{t+\tau|t}$ represent hypothetical observations based on the planned control inputs and the anticipated fire front state. Since actual measurements are only available after executing the control actions, the objective in Eq. (13a) requires taking an expectation over all possible future measurement sequences. This enables informed decision-making by accounting for potential outcomes without executing the corresponding trajectories.

Equations (13b)-(13f) admit no closed form: the model is nonlinear and non-Gaussian with set-valued (multi-object, multi-measurement) observations, so Kalman-type filters are inapplicable. Therefore, the recursion is implemented using Sequential Importance Resampling (SIR), i.e., particle filtering. Specifically, the belief $\mathcal{B}_{\tau}$ is represented by a set of weighted particles $\mathcal{B}_{\tau} = \{w_{\tau}^{(i)}, X_{\tau}^{(i)}\}_{i=1}^{N_s}$, where $X_{\tau}^{(i)} = [x_{\tau}^{1}, \ldots, x_{\tau}^{N}]^{\top}$. These particles are propagated to the next time step according to the process dynamics and reweighted using the likelihood function to compute the posterior. For notational convenience, we will often write $t+\tau|t$ as τ when no ambiguity arises. The functional $\mathcal{C}_{t+\tau|t} : \mathcal{B}_{t+\tau|t}(X_{t+\tau|t}|Z_{1:t+\tau|t}) \rightarrow [0,1]$ in Eq. (13a) quantifies the uncertainty of the fire-front state $X_{t+\tau|t}$ encoded in the posterior distribution at time $t+\tau|t$, conditioned on all hypothetical measurements $Z_{1:t+\tau|t}$. This uncertainty is measured by the *Risk-Weighted Dispersion* (RWD) defined as:

$$\mathcal{C}_{t+\tau|t}(\mathcal{B}_{t+\tau|t}) = \frac{1}{\omega} \sum_{\varepsilon \in \tilde{\mathcal{E}}} \mathcal{R}(\varepsilon) \det(\Sigma^{\varepsilon}_{t+\tau|t}), \tag{14}$$

where the environment $\mathcal{E}$ is discretized in space to form a 2D grid $\tilde{\mathcal{E}}$, composed of a finite number of non-overlapping, equally sized cells $\tilde{\mathcal{E}} = \{\varepsilon_1, \ldots, \varepsilon_{|\tilde{\mathcal{E}}|}\}$, such that $\bigcup_{i=1}^{|\tilde{\mathcal{E}}|} \varepsilon_i = \mathcal{E}$. The term $\mathcal{R}(\varepsilon) \in [0,1]$ denotes the risk value associated with cell ε, reflecting the severity of fire presence in that region. The random quantity $\det(\Sigma^{\varepsilon}_{t+\tau|t})$ is the determinant of the sample covariance matrix $\Sigma^{\varepsilon}_{t+\tau|t}$ computed from all particle points residing in cell ε at time step $t+\tau|t$, and ω is a scaling factor ensuring that $\mathcal{C}_{t+\tau|t}(\mathcal{B}_{t+\tau|t}) \in [0,1]$. Finally, the parameter $\nu \in (0,1]$ in Eq. (13a) is a discount factor that controls the relative importance of future decisions.

C. Adaptive LCB-guided Policy Search

Problem (P1) is a stochastic, multi-dimensional, nonlinear, and non-convex optimization problem that cannot be directly solved in its original form. However, we can address an equivalent version by reformulating (P1) as a Markov Decision Process (MDP) [30]. To achieve this, we assume that the agent's control inputs $u_t \in \mathcal{U}$ can be reduced to a finite set $\mathbb{U}$, consisting of $|\mathbb{U}|$ discrete control vectors $\hat{u}_t \in \mathbb{U}$. This discretization, in turn, leads to a finite set of possible agent states $\hat{y}_t \in \mathbb{Y}_t$. Consequently, (P1) can be reformulated as an MDP $\langle \mathcal{S}, \mathbb{U}, \mathcal{T}, \mathcal{C} \rangle$, where $\mathcal{S}$ is the state space of the system, and an individual state $s \in \mathcal{S}$ is represented as the tuple $s_t = (\mathcal{B}_t, \hat{y}_t)$. Note that the fire front process evolves independently of the agent's actions. The transition function $\mathcal{T} : \mathcal{S} \times \mathbb{U} \rightarrow \mathcal{S}$ describes the evolution of the system in response to agent actions. Although the agent's actions are deterministic in our setting, the transition function $\mathcal{T}$ remains stochastic due to the randomness in the agent's observations upon executing an action. Specifically, we have: $\mathcal{T} : p\left(s'_t = (\mathcal{B}', \hat{y}') \mid s_{t-1} = (\mathcal{B}, \hat{y}), \hat{u}_{t-1} = u\right)$. The cost function $\mathcal{C}$ assigns a cost to a specific state s' resulting from applying action $\hat{u}$ at state s. With slight abuse of notation, we denote it as $\mathcal{C}(s' = (\mathcal{B}, \hat{y}))$, which effectively operates on the posterior belief $\mathcal{B}$ in state s', derived from the agent state $\hat{y}$. The cost is defined according to Eq. (14).

We define a finite-horizon open-loop policy over T steps as the control input sequence $\pi = \{\hat{u}_{t+\tau-1|t}\}_{\tau=1}^{T} \in \mathbb{U}^T$. Let $\Pi_T = \mathbb{U}^T$ denote the set of all such admissible policies. Each policy is a predetermined sequence of actions evaluated via

simulation of the MDP's stochastic state transitions. Given that the agent starts from an initial state $s_t = (\mathcal{B}_t, \hat{y}_t)$ at time step t, a policy $\pi \in \Pi_T$ can then be simulated under the MDP to obtain:

$$V_t^{\pi}(s_t) = \mathbb{E}\left\{\sum_{\tau=1}^{T} \nu^{\tau} \mathcal{C}_{\tau}\left(s_{t+\tau|t}^{\pi}\right)\right\}, \tag{15}$$

where $s_{t+\tau|t}^{\pi}$ is the state encountered at level τ in the horizon under policy π when starting from state s_t. Subsequently, the optimal policy that minimizes the expected cumulative cost over the horizon is given by

$$\pi^{\star} = \arg\min_{\pi \in \Pi_T} \; V_t^{\pi}(s_t) \tag{16}$$

Observe that the optimization problem in Eq. (16) is equivalent to Problem (P1) under the assumption of a finite number of admissible control inputs. Solving this equivalent MDP formulation however, introduces its own set of challenges. In particular, the continuous belief-space represented by particles and the stochastic nature of the measurements result in an effectively infinite state space. This makes classical dynamic programming (DP) methods, such as value iteration and backward induction, infeasible. These approaches rely on a finite state space to compute value functions backward from the planning horizon and therefore do not apply here. Furthermore, the transition dynamics are not explicitly known in closed form, nor can they be represented in a tabular format.

To solve this problem we build upon the Upper Confidence Bound 1 (UCB1) framework [31] and we treat each sequence $\pi = \{\hat{u}_{t+\tau-1|t}\}_{\tau=1}^{T} \in \mathbb{U}^T$ as an *arm* in a multi-armed bandit. We use rollouts to simulate policy outcomes and employ a UCB1-like adaptive selection strategy to efficiently balance exploration and exploitation. In particular, the UCB1 algorithm iteratively computes an upper confidence bound score on the expected reward for each arm (where arms are considered the actions), by adding the sample mean reward of the arm to an exploration bonus that depends on both the total number of arm pulls and the number of pulls for that specific arm. This formulation is grounded in the "optimism under uncertainty" principle and leverages concentration inequalities (i.e., Chernoff-Hoeffding bounds) to guarantee near-optimal regret (i.e., the expected loss between the optimal policy and selected policies). Based on these scores, the policy always selects the arm with the highest current upper confidence bound, which enables asymptotic convergence to the optimal policy [31]. The proposed adaptive-search algorithm iteratively computes a Lower Confidence Bound (LCB) on the expected total cost associated with each T-finite control sequence, which is then used to adaptively select the next policy to simulate. Specifically, for an agent at state s_t, we define the LCB score of a policy $\pi \in \Pi_T$ at iteration n as:

$$\mathrm{LCB}^n(\pi, s_t) = \begin{cases} Q^n(\pi, s_t) - \sqrt{\frac{2\ln n}{I^n(\pi)}}, & \text{if } I^n(\pi) \neq 0, \\ \mathrm{LCB}_{\min}^n, & \text{otherwise,} \end{cases} \tag{17}$$

where $I^n(\pi)$ denotes the number of times policy π has been simulated up to iteration n, and $Q^n(\pi, s_t) = \frac{1}{I^n(\pi)} \sum_{i=1}^{n} \delta(\pi^i - \pi) L(\pi, s_t)$ is the sample mean of the cumulative cost incurred by policy π. Here, $L(\pi, s_t) = \sum_{\tau=1}^{T} \nu^{\tau} \mathcal{C}_{\tau}\left(s_{t+\tau|t}^{\pi}\right)$ is the total discounted cost of executing policy π starting from state s_t, and $\delta(\cdot)$ is the discrete Dirac delta function. The constant $\mathrm{LCB}_{\min}^n < -\sqrt{2\ln n}$ ensures that any policy not yet selected by iteration n will have an LCB value lower than that of any previously selected policy, thereby guaranteeing it will be explored with higher probability in subsequent iterations. Finally, the term $\sqrt{\frac{2\ln n}{I^n(\pi)}}$ serves as an exploration bonus, encouraging the selection of under-explored policies. Under the standard UCB1 assumptions, the expected fraction of iterations on which LCB selects a suboptimal policy is $O\big(\frac{\ln n}{n}\big)$; hence the optimal policy is selected with asymptotic frequency 1 in expectation as $n \to \infty$.

Algorithm 1 Adaptive LCB-guided Search

Input: Policy set Π_T, Initial state $s_t = (\mathcal{B}_t, \hat{y}_t)$
Initialize $I^0(\pi) = 0, \forall \pi \in \Pi_T$
for $n = 1, \ldots, (n_{\max} \geq |\Pi_T|)$ **do**
  Sample policy: $\hat{\pi} = \arg\min \mathrm{LCB}^n(\pi, s_t)$ (Eq. (17))
  for $\tau = 1, \ldots, T$ **do**
    Compute predictive density: $\mathcal{B}_{t+\tau|t}^{-}$ via Eq. (13b)
    Move to new state: $\hat{y}_{t+\tau|t} = f_a\left(\hat{y}_{t+\tau-1|t}, \hat{\pi}_{\tau}\right)$
    Sample meas.: $Z_{t+\tau|t} \sim p(Z|X_{t+\tau|t}, \hat{y}_{t+\tau|t})$
    Compute posterior: $\mathcal{B}_{t+\tau|t}$ via Eq. (13f)
    Compute stage cost: $c_{\tau} = c_{\tau-1} + \nu^{\tau}\mathcal{C}_{\tau}(\mathcal{B}_{t+\tau|t})$
  end for
  Update LCB: $Q^n(\hat{\pi}, s_t) = \mathrm{mean}\left(Q^{n-1}(\hat{\pi}, s_t), c_T\right)$
    $I^n(\hat{\pi}) = I^n(\hat{\pi}) + 1$
end for
Output: Optimal policy $\pi^{\star} = \arg\min \; \mathrm{LCB}^{n_{\max}}(\pi, s_t)$

The complete LCB-guided adaptive search algorithm is shown in Algorithm 1. At each time step t, the algorithm identifies the optimal policy $\pi^{\star}$ over the horizon $\{t+\tau|t\}_{\tau=1}^{T}$. The agent then executes the first control input of $\pi^{\star}$, transitions to a new state, receives the "real" measurement, computes the posterior belief $\mathcal{B}_{t+1}$, and the algorithm is re-applied at the next time step $t+1$. We should note that intelligent pruning techniques, such as ϵ-suboptimal reductions [19], can be designed and integrated into the proposed approach to focus on the most promising set of control inputs at each time step thereby reducing the runtime complexity.

Theorem 1 (Completeness): *Let Π_T be the set of all candidate policies in the LCB framework. Then for every optimal policy $\pi^{\star} \in \Pi_T$, there exists iteration n such that $\pi^{\star}$ is selected at least once by the algorithm.*

Proof: By construction of the LCB algorithm, each policy $\pi \in \Pi_T$ is initialized in a manner that guarantees it is tried at least once. Thus, any policy that has never been tried up to a certain point will have a higher LCB score than those policies with a finite sample count, ensuring it is selected at least once. Hence $\pi^{\star}$ is included among the solutions eventually tried. ■

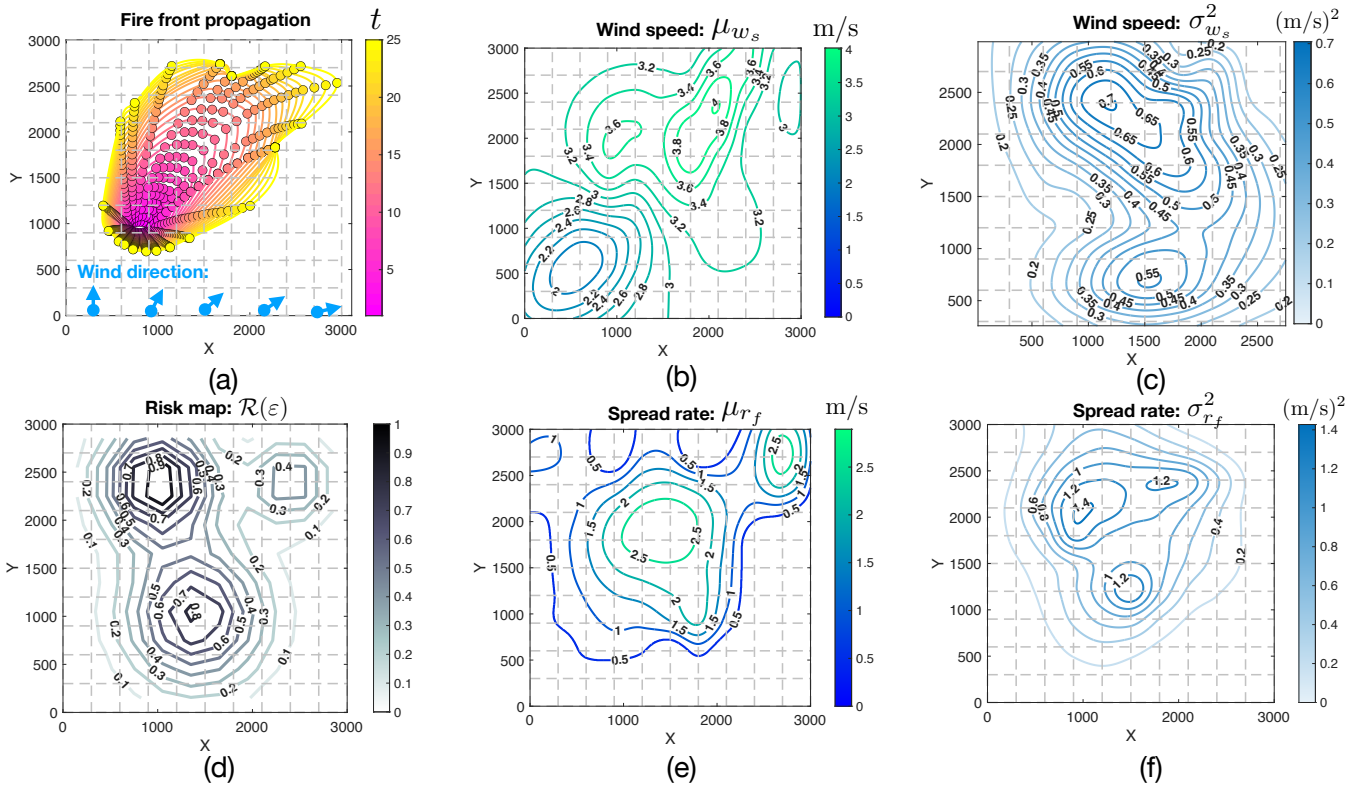

Fig. 1. Simulation Setup: (a) Fire front true evolution, (b)(c) Wind speed parameters, (d) Risk map, and (e)(f) Fire spread rate parameters.

Theorem 2 (Asymptotic Convergence): *Suppose there is a unique optimal policy $\pi^\star$ whose expected cost is strictly smaller than that of any other $\pi \in \Pi_T$. Under LCB, the fraction of times $\pi^\star$ is selected converges to 1 as $n \to \infty$, the per-iteration regret decays on the order of $O\big(\frac{\ln(n)}{n}\big)$, and LCB converges to $\pi^\star$ at that rate.*

Proof: By Theorem 1, every policy in Π_T is selected at least once under LCB. Moreover, LCB follows the same selection rule as UCB1 for minimization problems, adapted to entire control sequences, effectively treating each policy as a bandit arm with a stationary cost distribution and stage costs bounded in the range $[0, 1]$. Consequently, the classical result on UCB1 [31] applies directly: the total number of times any suboptimal policy is selected is $O(\ln(n))$. Hence the fraction of times a suboptimal policy is chosen decreases as $O(\frac{\ln(n)}{n})$, which implies the expected regret shrinks at that same rate. Thus, as $n \to \infty$, $\frac{\ln(n)}{n} \to 0$ and the fraction of times the optimal policy $\pi^\star$ is selected converges to 1. ■

IV. Evaluation

1) Simulation Setup: To evaluate the proposed approach, we used the following setup: the environment $\mathcal{E} \subset \mathbb{R}^2$ is square, with side length 3×10^3 m in each dimension, whereas its discrete representation $\tilde{\mathcal{E}}$ consists of a 10×10 grid with equally sized cells, as shown in Fig. 1(a) with gray dotted lines. The fire front state X_t comprises $N = 20$ fire front vertices (shown in Fig. 1(a)), initially forming an ellipse centered at $(x, y) = (800, 900)$, i.e., the ignition point, with semi-major and semi-minor axes of lengths 120 m and 60 m, respectively. Each of these vertices evolves according to Eq. (2) with $\Delta t = 60$s. The environmental conditions are as follows: wind direction θ, wind speed w_s, and fire spread rate r_f due to fuel are defined for each cell $\varepsilon \in \tilde{\mathcal{E}}$. Specifically, the mean wind direction (with North aligned with the y-axis) varies uniformly across the x-axis from North to North-East to East, as illustrated by the blue arrows in Fig. 1(a), following a von Mises distribution with concentration parameter $\kappa = 500$ in every cell i.e., $\theta(\varepsilon) \sim \mathcal{VM}(\mu_{\theta(\varepsilon)}, \kappa_{\theta(\varepsilon)})$. Subsequently, the wind speed and fire spread rate are modeled as rectified Gaussian distributions, i.e., $w_s(\varepsilon) \sim \mathcal{N}_R(\mu_{w_s(\varepsilon)}, \sigma^2_{w_s(\varepsilon)})$ and $r_f(\varepsilon) \sim \mathcal{N}_R(\mu_{r_f(\varepsilon)}, \sigma^2_{r_f(\varepsilon)})$. The superposition of these random variables over the grid for each $\varepsilon \in \tilde{\mathcal{E}}$ is shown in Fig. 1(b) and Fig. 1(c) for the wind speed, and in Fig. 1(e) and Fig. 1(f) for the fire spread rate, respectively. Finally, the risk $\mathcal{R}(\varepsilon)$ associated with each cell, indicating the severity of fire in that region is shown in Fig. 1(d). Consequently, the fire front state X_t evolves in continuous space and the environmental parameters θ, w_s, and r_f influencing each fire front vertex are obtained by associating it with the nearest cell in the discretized environment. The mobile agent (i.e., a drone) evolves according to $\hat{y}_t = \hat{y}_{t-1} + d_R\big[\cos_d(\vartheta), \sin_d(\vartheta)\big]^\top$, and is controlled via the input $\hat{u}_t = [d_R \Delta t,\ \vartheta]$, where $d_R \in \{3, 6\}$ m/s and $\vartheta \in \{0, 90, 180, 270\}$ deg. We assume that the drone operates at a fixed altitude of 250 m and is equipped with a wide-angle field-of-view camera with a viewing angle of 120 deg, resulting in a circular sensing

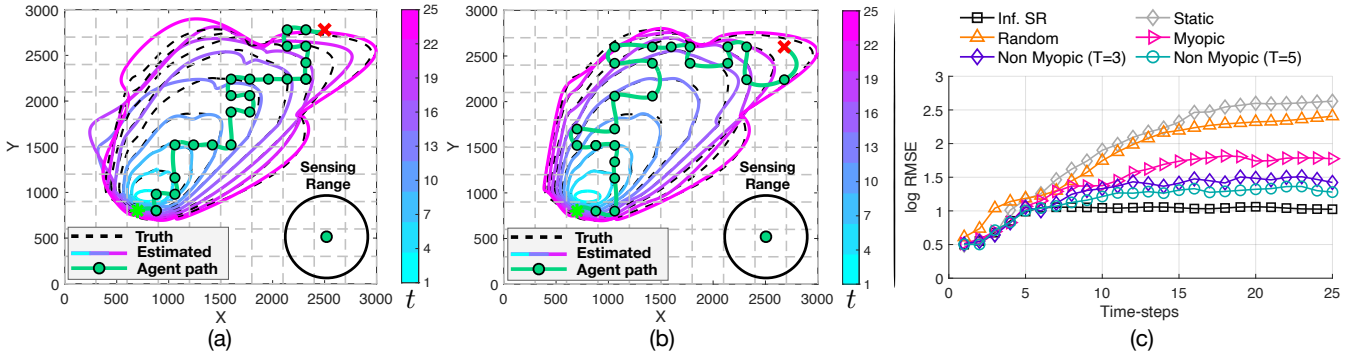

Fig. 2. Performance Evaluation: (a) Agent trajectory ($\star$ and $\times$ indicate start and stop states respectively), and fire front state (i.e., estimated with solid line, and true state with dotted line) obtained with Alg. 1, over 25-steps experiment, with myopic settings (i.e., horizon $T = 1$), (b) Same result with non-myopic settings (i.e. horizon $T = 3$), (c) Estimation error for different configurations of the proposed approach and comparisons with baselines.

range with radius $R_a = 250 \tan_d(120/2) \approx 430$ m. The measurement noise is set to $\sigma_z = 3.5$ m, and we assume a fixed intensity $\lambda_t^i(x_t^i) = 5$, for all $i \in \{1, \dots, N\}$ and all t. Finally, $\nu = 0.99$, $n_{\max}$ varies depending on the horizon, and the Bayes recursion in Eq. (10) is implemented as a SIR particle filter with $N_s = 2000$ particles.

2) Results: Figure 1(a) illustrates the true evolution of the fire front over a simulation period of $T_s = 25$ time steps, influenced by the environmental conditions described above. The fire-spread physics in this model support propagation in all directions from the ignition point. As a result, even under strong wind and fuel conditions, the upwind (back) perimeter of the fire continues to advance, albeit at a slower rate, as depicted. Furthermore, Fig. 1(a)-(c) illustrates how variations in wind direction, wind speed, and spread rate influence the propagation of the fire front, either by accelerating or decelerating its advancement.

The output of Alg. 1 for this setup is shown in Fig. 2(a) and Fig. 2(b), corresponding to horizon lengths of $T = 1$ (myopic) and $T = 3$ (non-myopic), respectively. The agent is initialized at position $(x, y) = (700, 800)$, as indicated by the green asterisk, running Alg. 1 in a rolling horizon fashion, as discussed in Sec.III-C. The green line in the figures is the agent's final trajectory over 25 time steps, resulting from the minimization of the expected cumulative RWD over the planning horizon at each time step. The dotted black lines show the true evolution of the fire front, while the time color-coded lines represent the estimated front; ideally, these two should align closely. As shown, the non-myopic approach plans multiple steps ahead and achieves improved performance compared to the myopic strategy, which plans greedily. Specifically, the myopic behavior of the agent in Fig. 2(a) prevents it from targeting the high-risk region in the top-left corner of Fig. 1(d), which in this scenario is also associated with high uncertainty, as illustrated in Figs. 1(c) and 1(f). This limitation results in significant estimation errors, as shown.

Subsequently, Fig. 2(c) illustrates the performance of the proposed approach in terms of root mean square error (RMSE) i.e., $\sqrt{\mathbb{E}\big((\hat{X}_t - X_t)^2\big)}$, between the estimated fire front state $\hat{X}_t$ and the true state X_t over 25 time steps, using a uniform risk map. Specifically, we conducted 50 Monte Carlo trials, randomly initializing both the fire front and the agent's position within the simulation environment described earlier. The figure presents the average $\log_{10}$ RMSE per time step per vertex over the 25-step simulation, comparing six different approaches. The baseline, denoted as *Inf. SR*, corresponds to an agent with infinite sensing range that remains stationary and simply runs the particle filter. In this case, the RMSE arises solely from measurement noise and multiple detections, with no influence from control actions-representing the best achievable performance under the given settings. The *Static* approach involves a stationary agent with a finite sensing range, while the *Random* approach uses an agent also with finite sensing range that selects a random control input at each time step. As expected, both approaches result in significant errors. The figure also includes the proposed method evaluated with different planning horizon lengths, i.e., $T = 1$ (myopic), and non-myopic settings with $T = 3$ and $T = 5$, and clearly demonstrates improved performance as the planning horizon increases. Note that the baseline is unattainable in this setting due to the agent's limited sensing capabilities.

V. Conclusion

This paper considers the problem of fire front monitoring under uncertainty by formulating it as a stochastic optimal control problem that integrates sensing, estimation, and control. A recursive Bayesian estimator was developed for elliptical-growth fire front processes, and the control problem was formulated as a finite-horizon Markov Decision Process (MDP). An information-seeking control law was then derived using a lower confidence bound (LCB)-based adaptive search, enabling optimal risk-aware planning.

References

[1] E. Kajita, "Notes from the field: Emergency department use during the los angeles county wildfires, january 2025," *MMWR. Morbidity and Mortality Weekly Report*, vol. 74, 2025.

[2] S. Papaioannou, P. Kolios, T. Theocharides, C. G. Panayiotou, and M. M. Polycarpou, "Towards automated 3d search planning for emergency response missions," *Journal of Intelligent & Robotic Systems*, vol. 103, no. 1, p. 2, 2021.

[3] ——, "A Cooperative Multiagent Probabilistic Framework for Search and Track Missions," *IEEE Transactions on Control of Network Systems*, vol. 8, no. 2, pp. 847–858, 2020.

[4] S. Papaioannou, S. Kim, C. Laoudias, P. Kolios, S. Kim, T. Theocharides, C. Panayiotou, and M. Polycarpou, "Coordinated crlb-based control for tracking multiple first responders in 3d environments," in *2020 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2020, pp. 1475–1484.

[5] S. Papaioannou, P. Kolios, T. Theocharides, C. G. Panayiotou, and M. M. Polycarpou, "3d trajectory planning for uav-based search missions: An integrated assessment and search planning approach," in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2021, pp. 517–526.

[6] S. Papaioannou, P. Kolios, C. G. Panayiotou, and M. M. Polycarpou, "Synergising human-like responses and machine intelligence for planning in disaster response," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–8.

[7] S. Papaioannou, P. Kolios, T. Theocharides, C. G. Panayiotou, and M. M. Polycarpou, "Jointly-optimized searching and tracking with random finite sets," *IEEE Transactions on Mobile Computing*, vol. 19, no. 10, pp. 2374–2391, 2019.

[8] ——, "Cooperative receding horizon 3d coverage control with a team of networked aerial agents," in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 4399–4404.

[9] ——, "Integrated ray-tracing and coverage planning control using reinforcement learning," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 7200–7207.

[10] ——, "Decentralized search and track with multiple autonomous agents," in *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE, 2019, pp. 909–915.

[11] S. Papaioannou, C. Laoudias, P. Kolios, T. Theocharides, and C. G. Panayiotou, "Joint estimation and control for multi-target passive monitoring with an autonomous uav agent," in *2023 31st Mediterranean Conference on Control and Automation (MED)*. IEEE, 2023, pp. 176–181.

[12] S. Papaioannou, P. Kolios, and G. Ellinas, "Distributed estimation and control for jamming an aerial target with multiple agents," *IEEE Transactions on Mobile Computing*, vol. 22, no. 12, pp. 7203–7217, 2022.

[13] S. Papaioannou, P. Kolios, T. Theocharides, C. G. Panayiotou, and M. M. Polycarpou, "Rolling horizon coverage control with collaborative autonomous agents," *Philosophical Transactions A*, vol. 383, no. 2289, p. 20240146, 2025.

[14] ——, "Jointly-optimized trajectory generation and camera control for 3d coverage planning," *IEEE Transactions on Mobile Computing*, 2025.

[15] S. Papaioannou, P. Kolios, C. G. Panayiotou, and M. M. Polycarpou, "Data-driven predictive planning and control for aerial 3d inspection with back-face elimination," in *2025 European Control Conference (ECC)*, 2025, pp. 2160–2166.

[16] H. X. Pham, H. M. La, D. Feil-Seifer, and M. C. Deans, "A distributed control framework of multiple unmanned aerial vehicles for dynamic wildfire tracking," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 4, pp. 1537–1548, 2020.

[17] P. Sujit, D. Kingston, and R. Beard, "Cooperative forest fire monitoring using multiple UAVs," in *2007 46th IEEE conference on decision and control*. IEEE, 2007, pp. 4875–4880.

[18] M. Lauri and R. Ritala, "Stochastic control for maximizing mutual information in active sensing," in *IEEE International Conference on Robotics and Automation*, 2014, pp. 1–6.

[19] N. Atanasov, J. Le Ny, K. Daniilidis, and G. J. Pappas, "Information acquisition with sensing robots: Algorithms and error bounds," in *2014 IEEE International conference on robotics and automation (ICRA)*. IEEE, 2014, pp. 6447–6454.

[20] A. O. Hero and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064–3075, 2011.

[21] P. Dames, M. Schwager, V. Kumar, and D. Rus, "A decentralized control policy for adaptive information gathering in hazardous environments," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 2807–2813.

[22] A. Singh, A. Krause, C. Guestrin, and W. J. Kaiser, "Efficient informative sensing using multiple robots," *Journal of Artificial Intelligence Research*, vol. 34, pp. 707–755, 2009.

[23] G. A. Hollinger and G. S. Sukhatme, "Sampling-based robotic information gathering algorithms," *The International Journal of Robotics Research*, vol. 33, no. 9, pp. 1271–1287, 2014.

[24] Y. Kantaros, B. Schlotfeldt, N. Atanasov, and G. J. Pappas, "Sampling-based planning for non-myopic multi-robot information gathering," *Autonomous Robots*, vol. 45, no. 7, pp. 1029–1046, 2021.

[25] G. D. Richards, "An elliptical growth model of forest fire fronts and its numerical solution," *International Journal for Numerical Methods in Engineering*, vol. 30, no. 6, pp. 1163–1179, 1990.

[26] M. A. Finney, *FARSITE, Fire Area Simulator–model development and evaluation*. US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 1998, no. 4.

[27] S. Papaioannou, P. Kolios, T. Theocharides, C. G. Panayiotou, and M. M. Polycarpou, "UAV-based receding horizon control for 3D inspection planning," in *2022 International Conference on Unmanned Aircraft Systems (ICUAS)*, 2022, pp. 1121–1130.

[28] ——, "Unscented optimal control for 3d coverage planning with an autonomous uav agent," in *2023 International Conference on Unmanned Aircraft Systems (ICUAS)*. IEEE, 2023, pp. 703–712.

[29] R. L. Streit and R. L. Streit, *The Poisson point process*. Springer, 2010.

[30] R. Sutton and A. Barto, "Reinforcement learning: An introduction," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 1054–1054, 1998.

[31] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Machine Learning*, vol. 47, pp. 235–256, 2002.