Hierarchical Framework for Optimizing Wildfire Surveillance and Suppression Using Human-Autonomous Teaming

Mahdi Al-Husseini,* Kyle H. Wray,† and Mykel J. Kochenderfer‡

Stanford Intelligent Systems Laboratory, Stanford, California 94305

The integration of manned and unmanned aircraft can help improve wildfire response. Wildfire containment failures occur when the resources available to first responders, who execute the initial stages of wildfire management referred to as the initial attack, are ineffective or insufficient. Initial attack surveillance and suppression models have linked action spaces and objectives, making their optimization computationally challenging. The initial attack may be formulated as a multi-agent partially observable Markov decision process (MPOMDP). We divide the initial attack MPOMDP into surveillance and suppression processes with their respective planners operating on different, but constant, time scales. A hierarchical framework iterates between surveillance and suppression planners while also providing collision avoidance. This framework is exemplified by a set of multirotor unmanned aircraft surveying an initial attack fire while a manned helicopter suppresses the fire with a water bucket. Wildfire-specific solver extensions are formulated to reduce the otherwise vast action spaces. The hierarchical framework outperforms firefighting techniques and a myopic baseline by up to 242% for moderate wildfires and 60% for rapid wildfires when simulated in abstracted and actual case studies. We also validate the early dispatching of additional suppression assets using regression models to ensure wildfire containment within thresholds established by wildfire agencies.

Nomenclature

AOA	=	manned aircraft axis of advance
a/A	=	suppression action/set of suppression actions
$\mathcal{B}(x)$	=	whether x is on fire in belief map $\mathcal{B}$
DR	=	drop type
D_m	=	distance between unmanned aircraft and manned aircraft below which P_m is incurred
D_u	=	distance between unmanned aircraft below which P_u is incurred
$\mathcal{D}(x)$	=	instantaneous destruction by wildfire on x in belief map $\mathcal{B}$ with respect to $R(x)$
e/E	=	surveillance action/ set of surveillance actions
F_T, P_T	=	set of x fully suppressed at T and set of x partially suppressed at T
$F(x)$	=	fuel remaining in x
IA_O	=	initial attack fire origin
k	=	function of distance between wildfire and suppression source and manned aircraft cruising speeds
P_i	=	penalty for distance between unmanned aircraft and IA_O
P_M	=	suppression penalty for global resource destruction minimization model
P_m	=	penalty for unmanned aircraft proximity to manned aircraft
P_u	=	penalty for unmanned aircraft proximity to one another
$P(x)$	=	probability that x ignites in t
$p(x, x')$	=	probability that x' ignites x at $t + 1$
$\mathcal{R}$	=	historical wildfire ring array
R_o	=	surveillance reward for belief map updates
R_M	=	suppression reward for localized resource destruction minimization model
$R(x)$	=	resources on x at time t is equal to 0
$\mathcal{S}$	=	surveillance state composed of $(X_{u_1}, Y_{u_1}, Z_{u_1}, X_{u_2}, Y_{u_2}, Z_{u_2})$
T, d_T	=	time step such that T is equal to 1 denotes the first act of wildfire suppression, and duration of T equal to $k * d_t$
t, d_t	=	time step such that t is equal to 0 is the incidence of the wildfire, duration of t
$\mathcal{U}(x)$	=	uncertainty regarding state of x in belief map $\mathcal{B}$
V_e	=	set of x surveyed given action e
$\mathcal{W}(x)$	=	whether x is on fire in the actual wildfire map $\mathcal{W}$
(X_m, Y_m)	=	location in x, y of water drop
$(X_{u_1}, Y_{u_1}, Z_{u_1})$	=	location in x, y, z of first unmanned aircraft U_1
$(X_{u_2}, Y_{u_2}, Z_{u_2})$	=	location in x, y, z of second unmanned aircraft U_2
x	=	wildfire cell
α	=	expended unit of fuel in cell x in t
$\beta(x)$	=	fuel reduced in x due to suppressive activities at T
γ_F, γ_P	=	fuel reduced for a fully suppressed cell and fuel reduced for a partially suppressed cell
$\delta(x)$	=	probability that x is suppressed in T
$\mathcal{H}$	=	suppression state composed of (X_m, Y_m, DR)
$2^{|P_T|}$	=	all outcomes of the partially suppressed set (2^{18})

I. Introduction

EACH year since 2000, an average of 70,600 wildfires have burned through a cumulative 7 million acres, resulting in tens of billions of dollars in damages and thousands of lives lost. Wildfires in 2020 alone were responsible for more than 17,000 destroyed structures and 3500 fatalities [1]. The initial attack occurs when the first set of dispatched assets responds to an incipient wildfire [2]. Initial attack effectiveness significantly influences the outcome of a wildfire's eventual containment, placing a notable burden on the part of first responders [3]. The California Department of Forestry and Fire Protection (CALFIRE) and others define initial attack response success as maintaining 95% of wildfires under 10 acres [4]. The integration of manned and unmanned aircraft as part of an initial attack, as shown in Fig. 1, is introduced to *improve wildfire visibility* and *minimize initial attack fire damage* while further serving to

ensure fire containment to 10 acres through the early requisition of additional suppression assets.

Although there exists ample literature on the optimized maneuver of distributed unmanned aircraft to conduct tactical wildfire management, there is less focus on hierarchical coordination of surveillance and suppression operations and even less research integrating both manned and unmanned aircraft. To our knowledge, this paper is the first to do all three: optimize initial attack surveillance and suppression activities using a hierarchical framework of asynchronous planners to support a human-autonomous aircraft team. Human-autonomous teaming involves manned and unmanned agents collaborating to optimize mission execution [5]. Teaming in aviation applications frees flight crews to focus on complex mission-essential tasks, such as wildfire suppression, while allowing critical but routine secondary tasks, such as wildfire surveillance, to be automated [6]. Despite the opportunity presented by unmanned aircraft, manned aircraft are expected to play a critical role in wildfire response in the foreseeable future. CALFIRE, the premier firefighting aviation program in the world with more firefighting aircraft than any other, continues to rely almost exclusively on manned aircraft for surveillance and suppression [7]. Integrating low- and medium-altitude unmanned aircraft into existing manned fleets provides immediately beneficial and realistic policies that may enhance wildfire control efforts [8]. However, special attention must be paid to industry-standard aviation tactics, techniques, and procedures, and airspace models, to ensure collision avoidance [9].

Multi-agent problems with asynchronous actions and high-dimensional state spaces can be made tractable by using a hierarchical approach to planning [10]. A valuable characteristic of heterogeneous multi-agent models is that action abstraction may be held constant by agent type. In our wildfire scenario, manned aircraft are provided optimal suppression locations but not guided to them. The process by which manned aircraft arrive at those suppression locations, through underlying primitive actions, is based on aerial firefighting standard operating procedures. Unmanned aircraft are explicitly guided to optimal surveillance locations through primitive actions that also satisfy collision avoidance requirements. A single manned aircraft action may occur over the course of multiple unmanned aircraft actions, resulting in an asynchronous action space. All aircraft share a single belief state and possess each other's location. We decompose the initial attack multi-agent partially observable Markov decision process (MPOMDP) into separate surveillance and suppression partially observable Markov decision processes (POMDPs) operating on different time scales with distinct but related reward functions. The POMDPs are simplified into Markov decision processes (MDPs) using shared belief and uncertainty maps. The MDPs are subsequently coupled in a manner exploiting the known structure of an initial attack and respecting collision avoidance requirements. This framework may be adapted to optimize other aerospace teaming applications related to attack, search and rescue, and medical evacuation operations.

This paper's contributions include the following: 1) an adaptable, hierarchical framework for the manned and unmanned multi-agent, multi-objective, partially observable initial attack problem featuring a high-dimensional state and action space; 2) several Monte Carlo tree search (MCTS) extensions informed by wildfire domain knowledge to include three suppression action space restriction mechanisms (ASRs), surveillance and suppression reward models, and an internal wildfire propagation model; 3) simulations demonstrating the efficacy of (1) and (2) given various wind conditions, elevation profiles, and resource topologies, and compared against firefighting techniques and a myopic baseline; and (4) a method for the early dispatching of additional wildfire suppression assets to meet the 10-acre containment standard established by CALFIRE and other wildfire agencies. This framework, solver extensions, and methods validate the use of MDPs to optimize initial attack operations through the provision of collision avoidance, increased flexibility in suppression, and real-time wildfire insights.

II. Related Work

This section reviews related research that applies distributed aircraft coordination frameworks, human-autonomous teaming models, hierarchical structures for multi-agent systems, or some combination of the three, to the wildfire surveillance and/or suppression problem. Further considered is the use of reinforcement learning or probabilistic search techniques to make informed decisions. We especially highlight the technical contributions of Seraj et al., which exist at the confluence of distributed aircraft coordination and hierarchical structures for multi-agent systems, for the joint wildfire surveillance and suppression problem [11]. Similarities and differences with our paper are discussed.

A. Distributed Aircraft Coordination Frameworks

The academic literature provides several examples of distributed unmanned aircraft coordination frameworks that support either wildfire surveillance or suppression (but rarely both). Julian and Kochenderfer demonstrate how deep reinforcement learning may be used to coordinate multiple autonomous fixed-wing aircraft to accurately track wildfire front expansion [12]. Pham et al. similarly introduce a distributed framework for controlling a set of quadcopters to chart a wildfire's progression while avoiding in-flight collisions [13]. Ghamry and Zhang divide the wildfire surveillance problem into three stages: search, confirmation, and observation. In the search stage, the unmanned aircraft team uses a leader-follower approach and moves in geometric formation to the wildfire. On arrival, the unmanned aircraft distribute in accordance with a generated elliptical fire front perimeter [14]. Griffith et al. compare the use of MCTS and mathematical optimization (MO) as applied to the allocation of wildfire suppression teams. MO models the MDP as a mixed-integer linear optimization problem, then applies a commercial solver to determine feasible solutions [15].

B. Human-Autonomous Teaming Models

At its simplest, the relationship between man and machine in a system may be categorized as *human-in-the-loop*, *human-on-the-loop*, or *human-out-of-the-loop*. Human-in-the-loop systems require human approval before action by the unmanned agent. Human-on-the-loop systems involve the human receiving updates from the unmanned agent, and the human in turn providing guidance to the unmanned agent. Human-out-of-the-loop systems have an unmanned agent that acts independently, albeit with initial or occasional guidance from the human [16]. There are surprisingly few research efforts that integrate man and machine in support wildfire management. The academic literature has largely prioritized fully autonomous teams conducting surveillance and suppression operations. There are exceptions. Bjurling et al. consider human-in-the-loop operator control over a swarm of unmanned wildfire surveillance aircraft [17]. Human-autonomous teaming may be defined as manned and unmanned agents working interdependently to accomplish a common goal [5]. Symbiotic autonomy is a form of teaming that accomplishes complex tasks by distributing subtasks and sharing information across multiple agents and agent groups [18]. In this framework, human and autonomous agents may act asynchronously to execute individual subtasks that enhance or inform each other's efforts [19], and may be supported by a "smart environment," which provides shared understanding across the team and thereby improves performance. Seraj and Gombolay
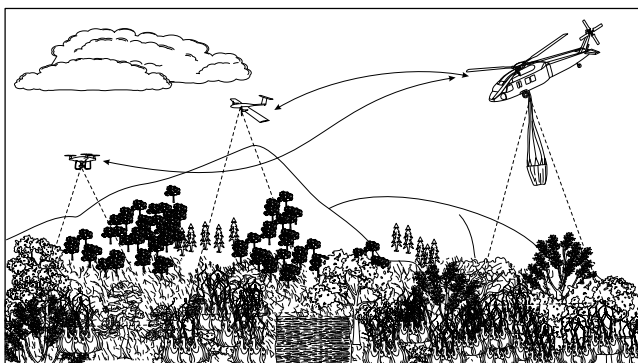

Fig. 1 Manned and unmanned aircraft coordinating to survey and suppress an initial attack fire.

introduce a distributed control framework for a team of unmanned aircraft conducting human-centered surveillance of wildfires and transmitting high-fidelity fire front observations to on-ground firefighters [20].

C.   Hierarchical Structures for Multi-Agent Systems

A hierarchical representation of tasks can enable decision-making across different levels of temporal abstraction for complex, multi-agent systems. Macro-actions are temporally extended actions that can be incorporated into multi-agent decision problems to overcome high-dimension state spaces. The use of macro-actions can result in cooperative agents acting asynchronously. Menda et al. introduce an algorithm that modifies policy gradient estimators for macro-actions, permitting policy optimization in models where agents act asynchronously. They successfully apply this algorithm to the cooperative multi-agent wildfire suppression problem [21]. Hierarchical representation is also a framework on which to scale reinforcement learning to large domains [22]. The category of hierarchical reinforcement learning (HRL) algorithms is large and diverse. With the advent of deep reinforcement learning, HRL has branched into options [23] and subgoal [24] methods. Similarly, there exist several state-of-the-art multi-agent deep reinforcement learning (MARL) algorithms, including MADDPG [25], COMA [26], and QMIX [27]. An in-depth discussion of HRL and MARL algorithms is beyond the scope of this paper. We do, however, highlight the work of Xu et al. in developing the Hierarchical Value Decomposition (HAVEN) framework for solving decentralized partially observable Markov decision process (Dec-POMDP) problems, one of the few algorithms straddling both HRL and MARL [28].

Seraj et al. apply a hierarchical framework of decision-making modules to a perception-action composite team of heterogeneous aircraft, both surveying and suppressing a wildfire [11]. They model the environment as a multi-agent partially observable semi-Markov decision process (MAPOSMDP). A high-level module assigns specialized surveillance tasks to a set of perception UAVs, while a low-level module coordinates a control and planning framework through which all UAVs execute their assigned tasks. A novel reinforcement learning algorithm is proposed that employs a variant of the state-action-reward-state-action (SARSA) algorithm tailored to multi-agent problems. This hierarchical design lends itself to "prolific cooperation between perception and action agents." We instead introduce a hierarchical framework that encourages cooperation between perception agents (unmanned surveillance aircraft), which in turn unilaterally support action agents (manned suppression aircraft). Thus, the perception agents are a complement to the action agents. Action agents are modeled as non-cooperative interacting entities [29], and although provided suppression guidance, are otherwise uncontrolled. Ultimately, we seek to minimize the extent to which unmanned aircraft disrupt manned aircraft operations. We divide the initial attack MPOMDP formulation into surveillance and suppression decision processes operating on different time scales and share information between their planners to ensure collision avoidance and fused information collection. Each subproblem is resolved using MCTS with several domain-specific extensions. We emphasize the integration of manned aircraft conducting suppression operations and demonstrate how certain tactics, techniques, and procedures employed by those aircraft can be supported, or at least avoided, by an unmanned aircraft network.

III.   Problem Statement

We begin by formulating the initial attack problem as an MPOMDP. The MPOMDP is a generalization of the MDP in which multiple agents, each unable to fully observe the underlying world state, collaborate and communicate freely toward one or more shared objectives [30]. The initial attack MPOMDP is represented by $\mathcal{M} = \langle \alpha, \mathcal{S}, \mathcal{A}, P, O, \Omega, \gamma, R \rangle$, where α is the number of agents, $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, P is the transition model, O is the observation function, Ω is the observation space, γ is the discount factor, and R is the reward function [29,31]. Tailoring $\mathcal{M}$ to the initial attack, take the number of agents to equal the number of participating

unmanned and manned aircraft, or $\alpha = \alpha_u + \alpha_m$. Each possible state $s^t \in \mathcal{S}$ at time t includes 1) unmanned aircraft i and manned aircraft j positions and 2) the wildfire state, such that $s^t = [p_i^t, p_j^t, W^t]$. Action space $\mathcal{A}$ is the set of all possible surveillance actions for unmanned aircraft and all possible suppression actions for manned aircraft, giving $\mathcal{A} = (\mathcal{A}_i \times ... \times \mathcal{A}_{\alpha_u}) \times (\mathcal{A}_j \times ... \times \mathcal{A}_{\alpha_m}))$. The transition model P assigns the probability of transitioning from existing state s^{t-1} by action a to state s^t, with $\sum_{s \in \mathcal{S}} P(s^{t-1}, a, s^t) = 1$. The discount factor γ is a hyperparameter that balances short- and long-term rewards. A larger γ prioritizes long-term rewards, whereas a smaller γ prioritizes short-term rewards. Observation function O gives a probability distribution over all possible observations ω after taking action a, resulting in state s. Observation space Ω includes the set of all possible partial wildfire observations by unmanned aircraft i at time t, where $\omega_i^t = [W_i^t]$, $\omega^t = \cup_{i=1}^{\alpha_u} \omega_i^t$, and $\omega^t \in \Omega$. Each unmanned aircraft takes its own surveillance action for which it receives an individual observation. Observations are fused across all unmanned aircraft to attain shared observation ω^t. The shared belief is a probability distribution over $\mathcal{S}$. We update the belief b^t that the current state is s^t, for each s^t, using

$$b^t(s^t) = \beta O(\omega^t, s^t, a^{t-1}) \sum_{s^{t-1} \in \mathcal{S}} b^{t-1}(s^{t-1}) P(s^t, a^t, s^{t-1}) \qquad (1)$$

where b^{t-1} is the initial belief, ω^t is the new shared observation, and β is a normalizing constant [32]. Reward function R returns the reward received when taking action a from state s. Rewards are jointly considered across α aircraft.

The heterogeneity of the agent population suggests that the larger MPOMDP may be decomposed into smaller decision processes tailored to particular subtasks. We partition the initial attack agents into unmanned and manned agent groups, with each group possessing its own action space and a unique level of temporal abstraction. The initial attack itself delineates agent group responsibilities and defines agent group relationships. This permits the overarching MPOMDP to be divided into separable surveillance and suppression POMDPs with differing but linked action spaces and reward models. The number of wildfire states is intractable at $2^{10,000}$, and a proxy in the form of a single modifiable belief map is introduced. The wildfire belief map size matches that of the actual wildfire state and is updated in accordance with unmanned aircraft observations. Surveillance and suppression POMDPs are further simplified into MDPs by assuming the developed wildfire belief map is the actual wildfire state, which we find incurs an acceptable level of error. The resulting surveillance and suppression MDPs are building blocks in a hierarchical framework, and their associated planners repeat at a frequency specific to the wildfire environment. MDPs are P-Complete and finite-horizon POMDP approximations are PSPACE-complete, and the complexity of our approach is similar. This section introduces the stochastic wildfire propagation model, manned and unmanned aircraft dynamics, and state space formulations.

A.   Wildfire Propagation

The wildfire propagation problem has been the target of significant research activity, and the resulting models feature varying levels of complexity. We use a stochastic propagation model inspired by that of Bertsimas et al. [33], formulated to simulate the efficacy of the proposed hierarchical surveillance and suppression framework. The propagation model is modified to include the effects of aerial suppression, wind direction and strength, and terrain elevation [34]. Wind and terrain are critical environmental factors affecting not only wildfire spread but unmanned and manned aircraft activity [35]. In keeping with CALFIRE and other wildfire agencies, initial attack fire containment is defined as wildfire control within 2 h or spread not greater than 10 acres. The wildfire model therefore consists of a roughly 10 acre square discretized into a 100×100 grid of 2×2 meter cells. Each cell x features a Boolean value representing whether x is *actually* on fire, $\mathcal{W}(x)$, a Boolean value representing whether x is *believed* to be on fire, $\mathcal{B}(x)$, and an integer value representing the amount of fuel remaining in x, $F(x)$. Fuel is defined as the amount of
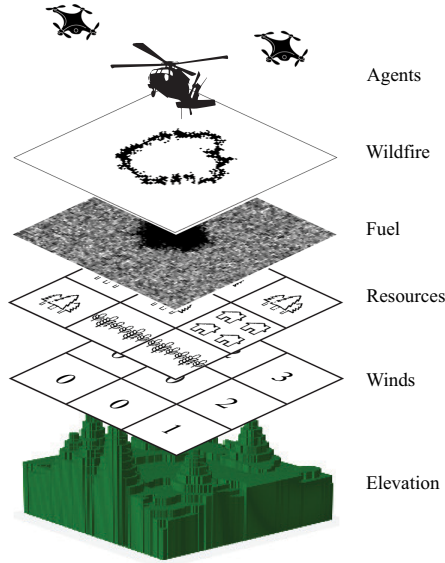
Fig. 2 The initial attack fire propagates based on fuel, winds, and terrain. Included in the initial attack environment are aerial agents surveying and suppressing the wildfire and the resources that are subject to destruction.

combustible material in a cell that contributes to fire behavior and effects [36].

The duration of surveillance time step t is d_t and is held constant at 1 minute. At t, each burning cell can 1) continue burning and expend a unit of fuel α, 2) stop burning by virtue of having expended all available fuel such that $F_t(x) = 0$, or 3) be partially or fully suppressed with probability $\delta(x)$. Similarly, each nonburning cell can ignite with probability $P(x)$, which is a function of the number of neighboring burning cells, remaining fuel, wind strength and direction, and terrain elevation. More specifically, the probability that neighboring cell x' ignites cell x at $t + 1$ is $p(x, x')$. The duration of suppression time step T is d_T, equal to kd_t, where k is a function of both the cruising speed of the suppression asset with and without load and the distance from the initial attack fire to a water replenishing source. There are five suppression actions, each with a different distribution of partially and fully suppressed wildfire grid cells. Notably, suppression affects not only the probability of ignition, but reduces fuel remaining as well [37]. The set of cells fully suppressed at T with zero probability of ignition is F_T, where γ_F is fuel reduced when fully suppressed. The set of cells partially suppressed at T with probability of ignition compounded by P_{partial} is P_T, where γ_P is fuel reduced for a partially suppressed cell. The individual success probabilities of multiple suppressive actions on a single cell x are independent. The resulting wildfire propagation equations are therefore

$$F_{t+1}(x) = \begin{cases} \max(0, F_t(x) - \alpha - \beta(x)) & \text{if } \mathcal{B}_t(x) \\ \max(0, F_t(x) - \beta(x)) & \text{otherwise} \end{cases} \quad (2)$$

$$\beta(x) = \begin{cases} \gamma_F & \text{if } t \bmod k = 0 \text{ and } x \in F_T \\ \gamma_P & \text{if } t \bmod k = 0 \text{ and } x \in P_T \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$P(x) = \begin{cases} 0 & F_t(x) = 0 \\ \delta(x) & \text{if } \mathcal{B}_t(x) \text{ and } F_t(x) > 0 \\ \delta(x)(1 - \prod_{x'}(1 - p(x, x')\mathcal{B}_t(x'))) & \text{if } \neg\mathcal{B}_t(x) \text{ and } F_t(x) > 0 \end{cases}$$
$$(4)$$

$$\delta(x) = \begin{cases} 0 & \text{if } t \bmod k = 0 \text{ and } x \in F_T \\ P_{\text{partial}} & \text{if } t \bmod k = 0 \text{ and } x \in P_T \\ 1 & \text{otherwise} \end{cases} \quad (5)$$

Wind and elevation kernels may be convolved with a grid of $p(x, x')$ to help propagate the wildfire in the direction of the wind or toward up-sloping terrain. A modifiable resource grid is included in the initial attack environment to associate value to areas of importance, such as housing communities in the midst of a forest. Figure 2 depicts inputs to the initial attack fire propagation model, and other features of the wildfire environment. The initial attack model begins with an assignment of fuel quantity and terrain elevation to each grid cell. As shown in Fig. 3, fire is seeded at $t = 0$ min at a handful of cells in the middle of the wildfire grid and allowed to propagate at each consecutive t. The simulation terminates when the initial attack becomes an escaped fire and additional suppression assets are introduced as a matter of policy; this occurs when $t = 120$ min or when the wildfire expands beyond the wildfire grid. The unmanned aircraft arrives 5 minutes after the initial attack fire begins, and the manned aircraft arrives 10 minutes after that. These estimates may be adjusted in the model depending on the proximity of assets to the wildfire.

B. Unmanned Aircraft Dynamics

Two multirotor unmanned aircraft with identical characteristics are introduced to survey the initial attack fire. The unmanned aircraft operate in a $10 \times 10 \times 7$ airspace gridworld consisting of $20 \times 20 \times 20$ meter grid cells and arrive on station at $t = 5$ min. The kinematics of each multirotor unmanned aircraft are abstracted in time, and simple gridworld commands, essentially primitive actions, translate the drone accordingly. More specifically, each drone may be translated up, down, left, right, ascend, descend, or hover in place. Translation may occur once per time step t, and observations are collected en route to the next grid cell. Each drone is limited to the confines of the airspace gridworld, and the action space is pruned at the edges accordingly. Although each unmanned aircraft acts independently, the surveillance MPOMDP is simplified into a POMDP by having the controller take joint actions; by doing so, the surveillance action space increases from 7 to 49. A multirotor design was selected over a fixed-wing design to enable hovering and because multirotor aircraft typically possess the kind of high-quality camera control needed to survey a wildfire.

Each unmanned aircraft must balance coverage with capture. The more elevated the unmanned aircraft, the wider its field of view, but
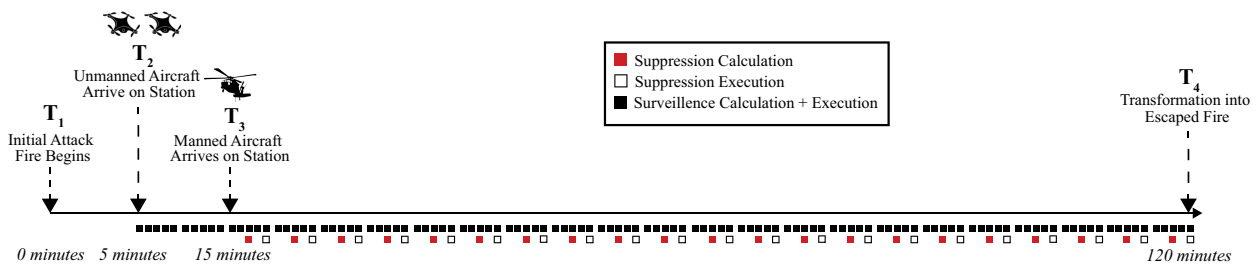

Fig. 3 The initial attack time line depicts the arrival time for unmanned and manned aircraft, the escaped fire transformation, and the calculation and execution of surveillance and suppression actions.

the less resolution it has on the wildfire at a given location. Strategically, this results in a lower altitude selected when the fire is condensed and a higher altitude when the fire is dispersed. The problem becomes more complicated with the addition of a second drone and collision avoidance penalties. As will be demonstrated, the inclusion of multiple unmanned aircraft that can actuate on two or more axes results in unique emergent behaviors, including dispersion, loitering, stacking, and circling. Each unmanned aircraft is penalized by P_u if it is within D_u meters of another unmanned aircraft.

C. Manned Aircraft Dynamics

A rotary-wing manned aircraft, with characteristics similar to those of an S-70 Firehawk helicopter, provides initial attack suppression capabilities through a short-line 660-gallon water bucket. The manned aircraft must fetch water from the nearest adequate water replenishing source before each suppression action. The distance between the water replenishing source and the initial attack fire can vary considerably, but is here assumed to be 10 km. Suppression time step T duration d_T is tied to the distance between the water replenishing source and initial attack fire and cruise speed of the manned aircraft with and without load. For example, an S-70 with an unloaded cruise speed of 140 KIAS and a loaded cruise speed of 80 KIAS, with 10 km between the initial attack fire and water replenishing source, can reasonably expect to perform a drop once every 5 minutes (thus, d_T and $k = 5$, where k is the number of iterations that occur within T). Given an on-station arrival at $t = 15$ min, the manned aircraft can therefore expect to perform 21 drops during the initial attack. These 21 water drops, in addition to efforts of ground-based suppression assets, determine whether an initial attack fire is contained or instead evolves into an escaped wildfire. Ground-based suppression activities are not considered in this paper.

The manned aircraft in an initial attack operates in accordance with best practices, in response to environmental factors, and often, per the guidance of a ground controller in the form of a fire truck, helitack crew, battalion chief, or hand crew [38]. Manned aircraft operations entail not only the placement of the water drop but also the type of drop and aircraft axis of advance in light of winds, smoke, and other aircraft operating in the immediate vicinity [39]. This paper does not attempt to fully capture the decision-making of a pilot in command of a firefighting aircraft but rather seeks to recognize the extent of that decision-making and provide supplemental guidance while ensuring supplemental assets (i.e., unmanned aircraft) do not restrict the manned aircraft's range of movement. Water may be dropped on any cell in the 100×100 wildfire grid. As shown in Fig. 4, there are

five standard drops, each categorized as either a point drop or line drop; this results in a suppression action space of 50,000. The point drop occurs when a suppressing helicopter slows to a near-hover before dumping the contents of its water bucket. The line drop occurs when a suppressing helicopter maintains forward airspeed before dumping the contents of its water bucket. Point drops are condensed in nature and typically used to suppress a condensed fire area, whereas line drops distribute the spread of water or retardant and are ideal for the placement of wet lines. Drop selection, bucket line length, and bucket volume determine the on-ground suppression profile [40]. This paper's model structure may be easily adapted to other aircraft and buckets of varying volumes and line lengths.

Simplifying assumptions include paralleling drop type and axis of advance, early identification of drop placement and type, and generalization of manned aircraft altitude and direction during the drop. The $200 \text{ m} \times 200 \text{ m}$ wildfire grid is small relative to the operational area of a manned aircraft, which may travel more than 10 km to reach a water replenishing source. It becomes reasonable to assume the manned aircraft's axis of advance is linked to its drop selection; a north–south line drop involves the manned aircraft traveling either north to south or south to north through the wildfire grid, such that the axis of advance includes the point of application. The axis of advance of a point drop is less straightforward, given that the manned aircraft slows to a near-hover; however, it is assumed that the axis of advance parallels the vector extending from the water replenishing source to the point of application of the initial attack fire, such that additional maneuvering is not required and time is saved. Figure 4 also depicts the axis of advance associated with each drop type. Each iteration of the surveillance model is held at 1 minute, providing the unmanned aircraft 2 minutes to determine the suppression location and axis of advance. The temporal resolution may be increased to allow the determination of manned aircraft maneuvering at a time nearer to suppression. Finally, for collision avoidance purposes, the unmanned systems are penalized by P_m if they come within D_m meters of the two-dimensional manned aircraft axis of advance at any altitude. This ensures that unmanned aircraft are penalized for proximity to the manned aircraft regardless of bucket-line length, drop altitude, or axis of advance directionality. Both P_m and D_m are larger than P_u and D_u respectively, representing the critical importance of ensuring the unmanned aircraft maintain separation from the manned aircraft and the somewhat lesser importance of ensuring the unmanned aircraft maintain separation from one another.

Given the partial observability of the initial attack fire, a shared belief map B, matching the dimensions of the 100×100 grid wildfire, is
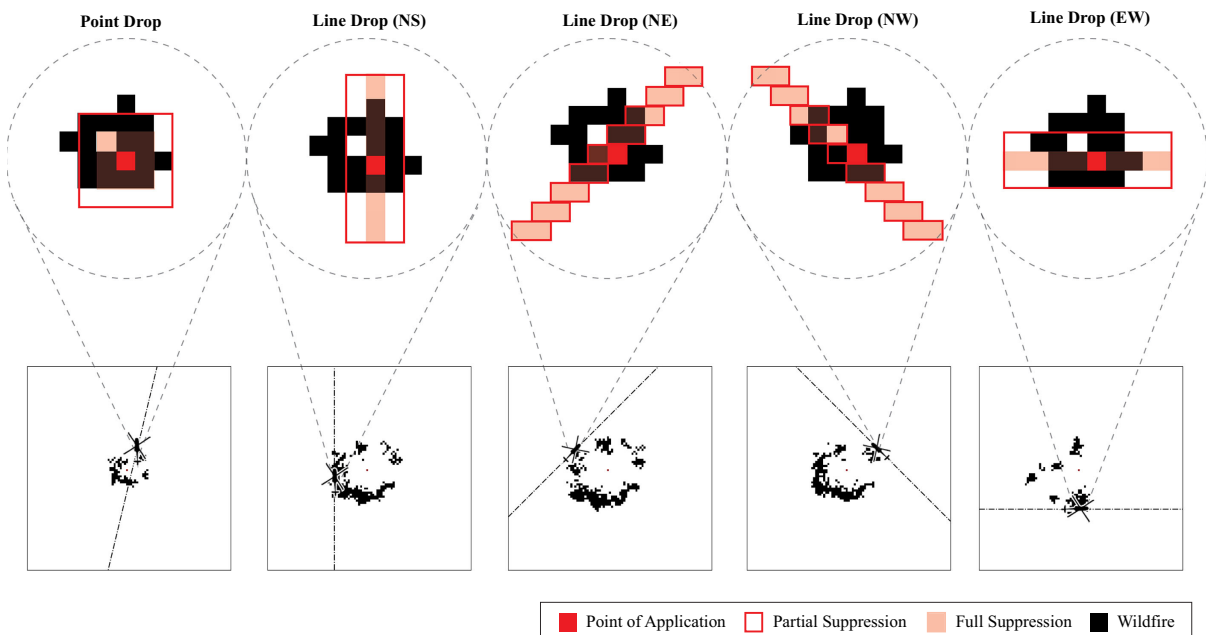

Fig. 4 Five suppression action drop-types and their associated aircraft axis of advance. A line drop and point drop have fundamentally different on-ground suppression characteristics.

introduced and updated at regular intervals. The surveillance planner, executed at each time step t, updates each wildfire cell in B that is observed by either of the two unmanned aircraft to reflect the actual wildfire state. The suppression planner, executed at each time step T, may further update the belief and uncertainty maps at locations where suppression is assured. Figure 5 depicts fuel and belief maps over the course of an initial attack fire. The red and green circular markers represent the unmanned aircraft locations in the two-dimensional representation and the unmanned aircraft trajectory in the last five time steps in the three-dimensional representation. The green square represents a high-value resource area. The ability of the unmanned aircraft to accurately capture the wildfire state is degraded as the wildfire expands.

D. State Space Formulations

The surveillance POMDP is simplified into an MDP by assuming a shared belief map $\mathcal{B}$ to be the actual wildfire state rather than maintaining a probability distribution across $2^{10,000}$ possible wildfire states; this does incur nontrivial error but results in a notable reduction in computational complexity by detaching each instantiation of the evolving wildfire gridworld from the state space. Instead, the surveillance state space, which is explicitly defined, encompasses state variables $(X_{u_1}, Y_{u_1}, Z_{u_1}, X_{u_2}, Y_{u_2}, Z_{u_2})$. This results in a large, but manageable 490,000 states, equivalent to the Cartesian product of sets U_1 and U_2, each constituting all possible locations for an unmanned aircraft in the airspace.

The suppression problem also assumes the shared belief map $\mathcal{B}$ to be the actual wildfire state. Rather than an explicit formulation, the suppression state space is developed by calling a generative model on any number of state–action pairs (s_0, a). The state variables for the suppression MDP are $(X_m, Y_m, DR, 2^{|P_T|})$. This generates a state space with an upper bound of approximately 13 billion states, representing every outcome of each possible suppression action.

IV. Solution

A hierarchical planner derived from wildfire and teaming domain knowledge, shown in Fig. 6 and outlined in Algorithm 1, is used to

split the larger surveillance-suppression MPOMDP into separate surveillance and suppression POMDPs operating on different time scales. A shared belief map is assumed for both models, further simplifying each POMDP into an MDP. The shared belief map is updated using the observed wildfire data and used to generate a wildfire uncertainty map and in the internal wildfire propagation model. The wildfire uncertainty map informs the surveillance planner's reward function. The surveillance planner recommends a surveillance action that updates the observed wildfire data and thus the belief map. The internal wildfire propagation model informs the suppression planner's rewards. The suppression planner recommends a suppression action, which affects the actual wildfire state. We choose to apply MCTS, a simulation-based search algorithm, to both surveillance and suppression planners. The upper confidence bound for trees (UCT) algorithm selects promising actions in the search trees [41]. Several MCTS extensions are introduced to reduce the large surveillance and suppression action spaces. As shown in Algorithm 1, the execution of a planner-recommended action does not always immediately follow the planning process itself; this delay allows heterogeneous planners to adjust the recommendation of future actions based on the expected actions of another (e.g., in support of collision avoidance).

Algorithm 1:   Hierarchical planner

Input: Initial wildfire belief $\mathcal{B}_0$, surveillance state $\mathcal{S}_0$, uncertainty map $\mathcal{U}_0$
Output: Updated $\mathcal{B}_k, \mathcal{S}_k, \mathcal{U}_k$ after k iterations, suppression action a
1: **for** $i = 0, 1, \ldots, k-2$ **do**
2:      $e \leftarrow$ Surveillance Planner $(\mathcal{U}_i, \mathcal{S}_i)$
3:      $\mathcal{B}_{i+1}, \mathcal{U}_{i+1}, \mathcal{S}_{i+1} \leftarrow$ survey$(\mathcal{B}_i, \mathcal{U}_i, e)$
4: $a \leftarrow$ Suppression Planner $(\mathcal{B}_{k-1})$
5: $e \leftarrow$ Surveillance Planner $(\mathcal{U}_{k-1}, \mathcal{S}_{k-1}, a)$
6: $\mathcal{B}_k, \mathcal{U}_k, \mathcal{S}_k \leftarrow$ survey$(\mathcal{B}_{k-1}, \mathcal{U}_{k-1}, e)$
7: $\mathcal{B}_k \leftarrow$ suppress$(\mathcal{B}_k, a)$

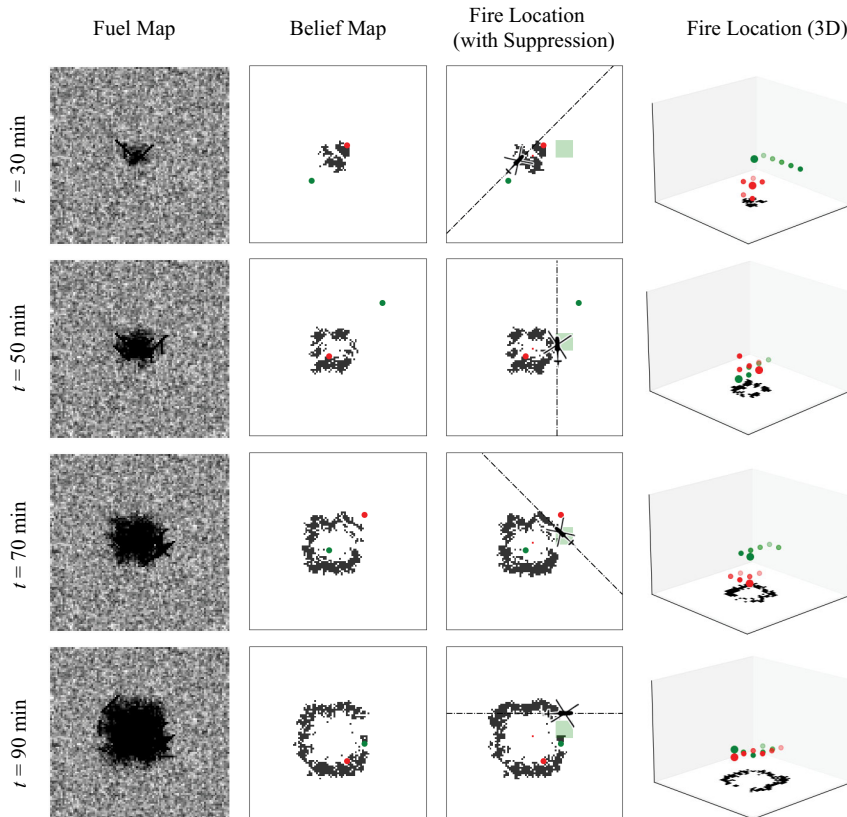| | Fuel Map | Belief Map | Fire Location (with Suppression) | Fire Location (3D) |

Fig. 5   Fuel, belief, and actual wildfire maps in two and three dimensions as the initial attack fire propagates.
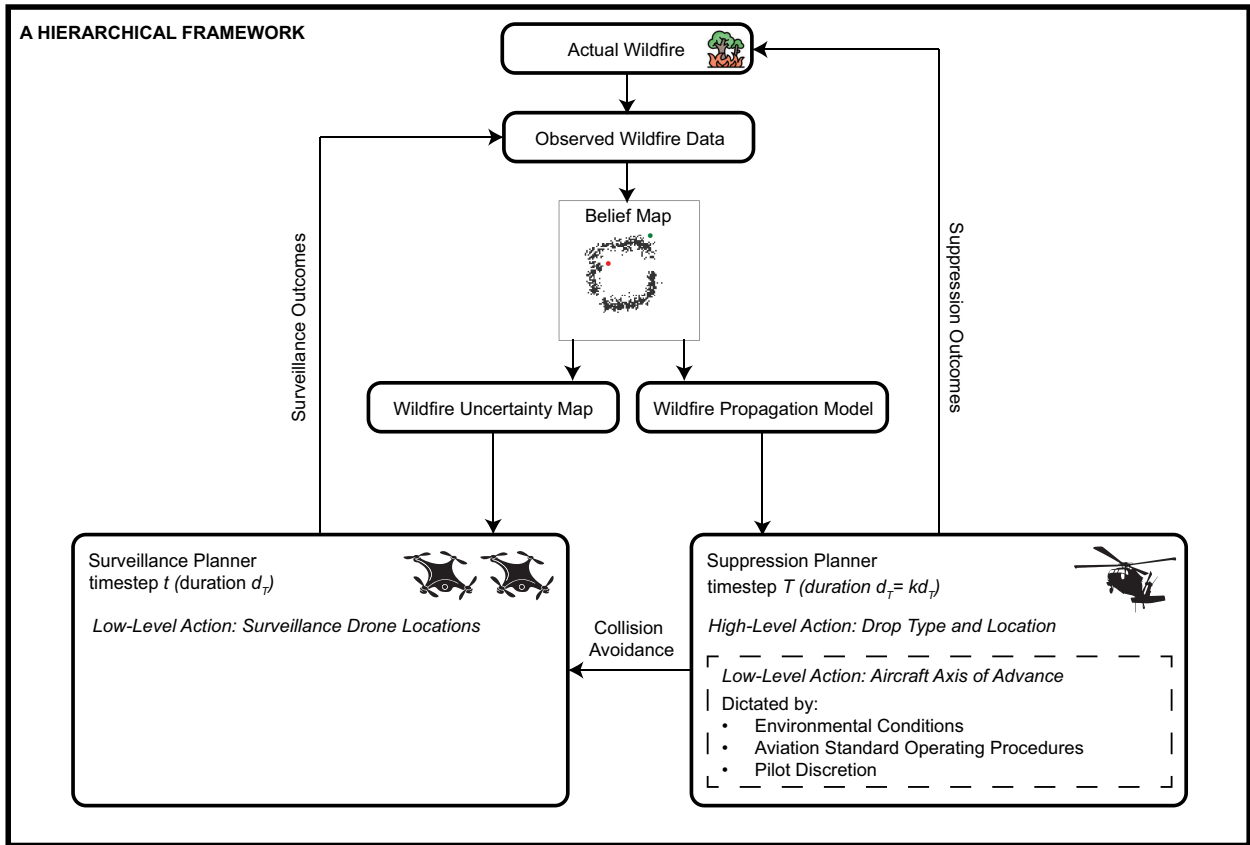
Fig. 6   A hierarchical framework for the initial attack using teaming involving linked surveillance and suppression planners.

A.   MCTS Solver

The introduced hierarchical planner is solver-agnostic. That said, the large state and action spaces coupled with a nonstationary environment in the propagating wildfire encourages the use of an online stochastic planning algorithm like MCTS. MCTS is also an anytime algorithm, which means it can return a valid solution if interrupted before runtime completion [42]. The set of possible actions is represented as edges, and the resulting states are represented as nodes in the search tree. Each MCTS iteration follows four general steps: selection, expansion, simulation, and propagation. MCTS selects nodes and traverses the search tree using a predefined heuristic informed by the problem domain. This heuristic results in an asymmetric search where the most promising actions are prioritized. We choose to apply the standard UCT algorithm to balance exploration versus exploitation in the tree search policy and determine the value of each node n, as follows:

$$\text{UCB}(n) = \bar{u}(n) + c \sqrt{\frac{\log(\text{visits}(n))}{\text{visits}(n')}} \qquad (6)$$

such that $\bar{u}(n)$ is the value of the state at n, c is an exploration constant that adjusts the balance between exploration of new nodes and exploitation of previously visited nodes, $\text{visits}(n)$ is the visitation count for n, and $\text{visits}(n')$ is the visitation count for the parent node of n, n'. The search tree is then expanded by adding a child node to the selected node. State value estimation occurs by simulating a play-out from a node to the end of a predefined planning horizon. A random policy is here applied to estimate the action value during search tree roll-outs. The simulation results are then backpropagated up the search tree, concluding a search iteration. It can be difficult to accurately assess MCTS computational complexity given the many subtasks involved, but in the general case, MCTS runtime and memory complexity scale linearly with the number of search iterations [43,44]. Figure 7 shows how MCTS may be applied to search trees for unmanned aircraft surveying an initial attack wildfire.

Selection is depicted by the bold circles and involves traversing the tree to a select depth using an algorithm like upper confidence bounds applied to trees. The selected node is then expanded by adding a child node, shown in gray. A simulation is performed from the child node to a preordained depth or condition, shown in red. The simulation results are associated with the child node and backpropagated up the search tree. This repeats until a given condition is true, the computation time-limit is met, or a ceiling on search iterations is reached. Given that the state space is evolving in time, we use an internal wildfire model to propagate the wildfire belief map between progressive search tree depths. States that exceed the boundaries of the model, shown crossed out in red, are pruned from the search tree.

B.   Surveillance Planning

There exist 49 possible surveillance actions at time t for a given belief map, equivalent to the Cartesian product of the set of seven actions for two unmanned aircraft. At depth of two, there are 2401 possible actions, and at depth three, there are 1.18×10^5. Given $2^{10,000}$ wildfire states, developing an explicit policy using dynamic programming is infeasible. Instead of an explicit policy formulation, a sampling-based approach with a generative model is used. An additional complication is reward uncertainty, given a stochastic element in the selection process for observations following a selected surveillance action. As each unmanned aircraft increases in altitude, the resolution of the wildfire grid below decreases, and a sampling process is used to balance observation coverage and capture.

Algorithm 2 outlines a process by which a surveillance action with depth one is recommended; as will be later discussed, a probabilistic-search formulation of this process is applied to enable search depths of two and three. Uncertainty map $\mathcal{U}_t$, surveillance state $\mathcal{S}_t$, and, optionally, suppression action a are submitted to the surveillance planner. A surveillance score map $\tilde{\mathcal{E}}$ is initialized. Each surveillance action e in set E is applied to the current surveillance state $\mathcal{S}_t$ to attain the next surveillance state $\mathcal{S}_{t+1}$. The ranging function provides an array of wildfire grid cells $\mathcal{V}_e$, which are observed from the updated
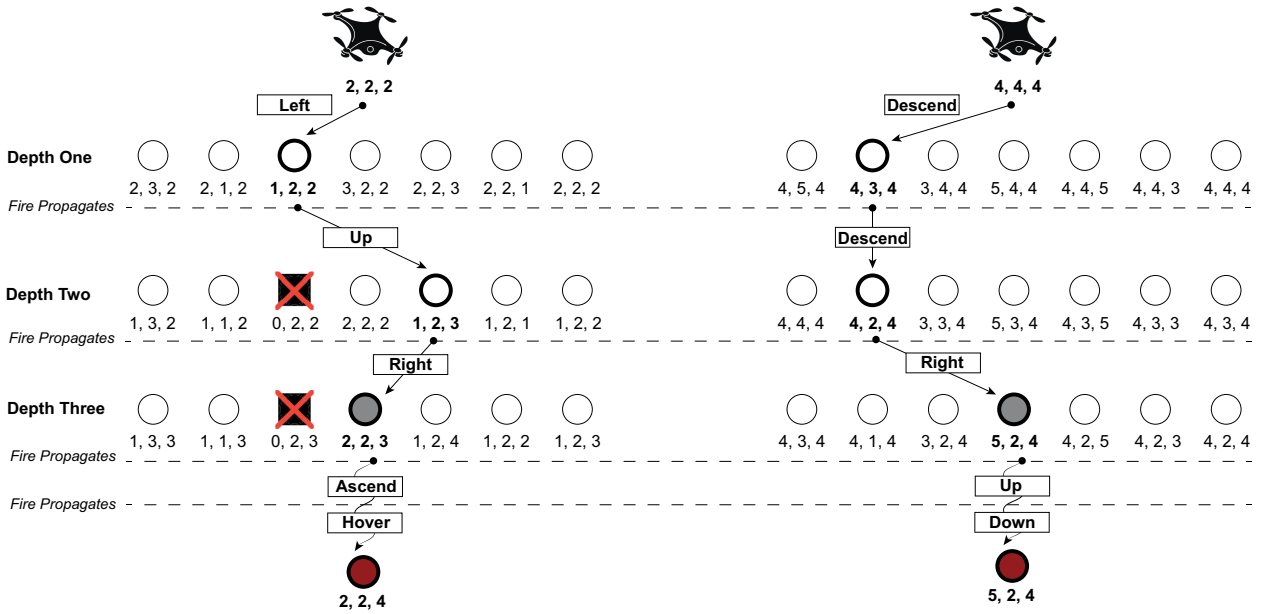
Fig. 7   Monte Carlo search trees depicted for two unmanned aircraft surveying an initial attack wildfire. Each iteration of Monte Carlo tree search has four steps: selection, expansion, simulation, and backpropagation.

Algorithm 2:   [Simplified] Surveillance planner (uncertainty reward model)

Input: Uncertainty map $\mathcal{U}_t$, surveillance state $\mathcal{S}_t$, suppression action a (optional)

Output: Surveillance action e

1: **Initialize** Surveillance action score map $\tilde{\mathcal{E}} \leftarrow \varnothing$
2: **for** action $e_i \in E$ **do**
3:      $S_{t+1} \leftarrow$ apply e_i to S_t
4:      $\mathcal{V}_{e_i} \leftarrow$ ranging(S_{t+1})
5:      $R_{o_{e_i}} \leftarrow \Sigma_{x \in \mathcal{V}_{e_i}} \mathcal{U}_t(x)$
6:      $R_{U_{e_i}} \leftarrow \tau_1 R_{o_{e_i}} + \tau_2 P_u(S_{t+1}) + \tau_3 P_m(S_{t+1}, a) + \tau_4 P_i(S_{t+1})$
7:      $\tilde{\mathcal{E}}[e_i] \leftarrow R_{U_{e_i}}$
8: Recommended surveillance action $e \leftarrow \arg\max_{e_i} \tilde{\mathcal{E}}[e_i]$

surveillance state position as a function of both location and altitude. As mentioned, there is a stochastic element to the ranging function, meaning that observations attained from a particular surveillance position, and thus ensuing rewards, will differ slightly with each simulation. The uncertainty in observation for each wildfire cell x in set $\mathcal{V}_e$, $U(x)$, is then summed to attain R_{o_e}, one of four components of the total surveillance reward R_U for a particular action e. This uncertainty model thus provides a reward corresponding to the extent to which observations reduce overall belief map uncertainty, regardless of the actual state of the wildfire or belief map. As shown in Fig. 8, the uncertainty reward model maintains a 100×100 uncertainty map in which each cell not observed in a particular time step

increases its uncertainty by an amount proportional to its proximity to burning cells (in the belief map), and each cell observed resets its uncertainty to zero. Cells on the outskirts of a wildfire grid with a negligible likelihood of igniting change their uncertainty minimally, while cells at the head of the fire change their uncertainty rapidly. Observations of cells that have not been recently observed reward more than cells that were more recently observed, regardless of whether the cell changes state postobservation.

Reward R_U is further composed of penalties P_u, P_m, and P_i. Unmanned aircraft U_1 and U_2 are penalized for their proximity to one another (P_u) and to the manned aircraft's axis of advance (P_m). The axis of advance and aircraft direction along it may be inferred from environmental data such as winds, aviation and wildfire best practices, and the suppression action a, or alternatively, they may be explicitly passed to the surveillance planner. A small penalty (P_i), proportional to the distance from each unmanned aircraft to the initial attack fire origin IA_O, is added to encourage unmanned aircraft to approach the initial attack fire. Penalty P_i is quickly overtaken by other rewards and penalties. The resulting total reward equation for the unmanned aircraft follows, where τ_1, τ_2, τ_3, and τ_4 are tunable parameters:

$$R_o = \tau_1 \sum_{x \in \mathcal{V}_e} \mathcal{U}_t(x) \qquad (7)$$

$$P_u = \begin{cases} \tau_2 & \text{if } \|U_1 - U_2\| \le D_u \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$
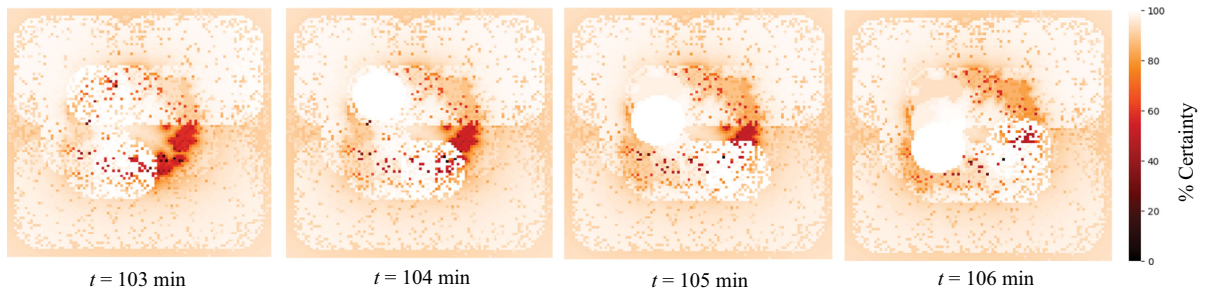

Fig. 8   Wildfire uncertainty map during initial attack fire propagation. Dark red areas suggest nearby fire activity but few recent observations. Grid cells distant from the wildfire, such as at the map periphery, increase their percent uncertainty minimally.

$$P_m = \begin{cases} \tau_3 & \text{if } \|U_1 - AOA\| \le D_m \\ \tau_3 & \text{else if } \|U_2 - AOA\| \le D_m \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$P_i = \tau_4(\|U_1 - IA_O\| + \|U_2 - IA_O\|) \quad (10)$$

$$R_U = R_o - P_u - P_m - P_i \quad (11)$$

The simplified approach in Algorithm 2 becomes computationally intractable when considering depths greater than one, given the large action space and reward uncertainty. The Algorithm 2 reward structure is maintained, and MCTS with UCT is applied to a generative MDP model to search the action space efficiently. The wildfire belief state $\mathcal{B}_t$ is passed to the MDP model to be adjusted and modified, along with the uncertainty map $\mathcal{U}_t$, with a progressive extension of the search-tree's depth. Each MCTS surveillance calculation is capped at 30 s of calculation time to allow each drone to proceed to its new location and attain observations before the next iteration. Any action that would take either a drone beyond gridworld boundaries or any combination of actions that would take both drones into the same grid cell is pruned. The computational complexity of a single MCTS surveillance iteration increases with wildfire growth. For example, given a 30 s restriction and a depth of three, MCTS surveillance conducts roughly 800 iterations at $t = 0$ min, and only 60 iterations at $t = 120$ min. Reducing the MCTS depth as the wildfire progresses ensures a roughly equivalent number of iterations at a depth of one throughout the lifespan of the wildfire. A depth factor of three is reduced to two, and then one, as the wildfire propagates, to maintain a minimum of two runs for each possible action at a depth of one.

C. Suppression Planning

The suppression model features 50,000 possible actions, encompassing five suppression actions centered on each of the 10,000 wildfire grid cells for a given belief map. The large action space

again prevents an exact policy formulation, and a generative method is instead applied. Reward uncertainty remains a factor, as water buckets fully suppress some wildfire cells but only partially suppress others; the partially suppressed cells are selected via random sampling. An understanding of wildfire dynamics allows for various forms of ASR that minimize the overall action space by 99% or more with minimal consequence. A generated and internally held wildfire propagation mode with limited information, shown in Fig. 6, enables the suppression planner to conduct customized roll-outs to optimize action selection. Two reward models are here introduced: localized and global resource destruction minimization. Global resource destruction minimization calculates the instantaneous destruction of the full 100×100 grid after roll-out, whereas localized destruction minimization calculates the instantaneous destruction of a smaller grid centered on the location of the action taken after roll-out. In both cases, the suppression aircraft is rewarded R_M, or penalized P_M, proportional to the instantaneous destruction caused by the wildfire following one or more roll-out periods in which the internal fire propagation model with limited information propagates postsuppression. Instantaneous destruction is a function of the number of cells burning and the value of the resources contained within those cells. Rewards for minimized instantaneous resource destruction following a roll-out period ensure the long-term impact of water lines can be measured against the short-term impact of more immediate suppressive activities.

Localized resource destruction minimization examines only the area immediately surrounding the suppression activity for reward considerations. Suppressive activities applied to different portions of the wildfire cannot be directly compared due to the differing windows under consideration and must instead be compared against a non-suppressed but equally propagated "reference grid," as illustrated in Fig. 9. Reward is then maximized when the difference between the appropriately localized portions of the reference grid and suppressed grid is greater. Algorithm 3 outlines the process by which a suppression action a is selected using localized rewards. Wildfire belief map $\mathcal{B}_t$ is propagated to roll-out depth RO to attain propagated reference grid $\mathcal{B}_{t+RO}$. An internal wildfire propagation model possessing wind
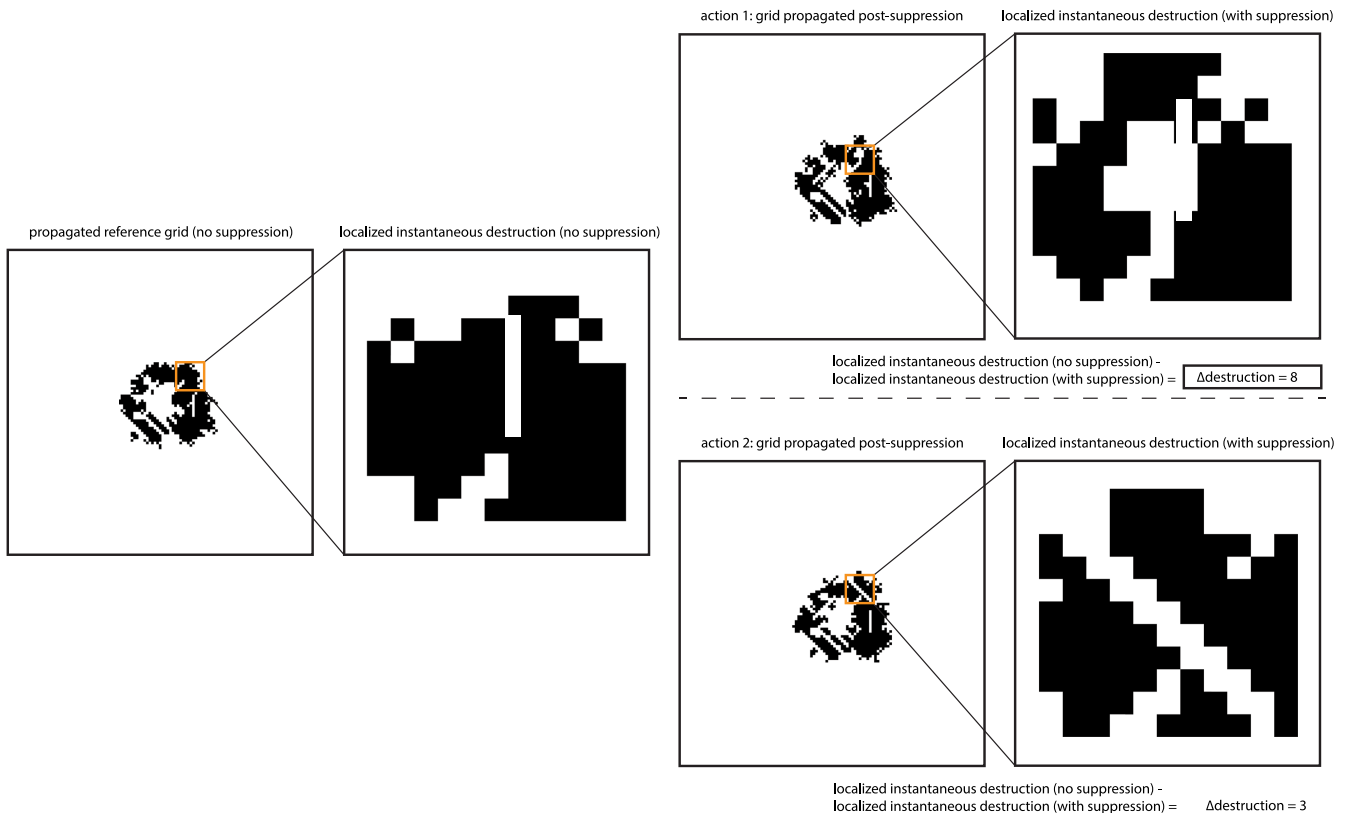

Fig. 9 The localized suppression destruction minimization reward model. A propagated reference grid and postsuppression propagation grid are clipped to attain local subarrays, which are then compared to determine action consequence.

Algorithm 3: [Simplified] Suppression planner (localized destruction minimization reward model)

Input: Wildfire belief $\mathcal{B}_t$, hyperparameter set $\{Q, RO\}$
Output: Suppression action a
1: $\mathcal{B}_{t+RO} \leftarrow \text{propagate}(\mathcal{B}_t, RO)$
2: $A \leftarrow \text{ASR}(\mathcal{B}_t, Q)$
3: **Initialize** Suppression action score map $\tilde{\mathcal{A}} \leftarrow \varnothing$
4: **for** action $a_i \in A$ **do**
5:      where a_i is (X_m, Y_m, DR)
6:      $\mathcal{B}_{st} \leftarrow \text{suppress}(X_m, Y_m, DR, \mathcal{B}_t)$
7:      $\mathcal{B}_{st+RO} \leftarrow \text{propagate}(\mathcal{B}_{st}, RO)$
8:      $\mathcal{L}_s \leftarrow \text{localize}(\mathcal{B}_{st+RO}, X_m, Y_m)$
9:      $\mathcal{L} \leftarrow \text{localize}(\mathcal{B}_{t+RO}, X_m, Y_m)$
10:      $R_{M_i} = \Sigma_{x \in \mathcal{L}} \mathcal{D}(x) - \Sigma_{x \in \mathcal{L}_s} \mathcal{D}(x)$
11:      $\tilde{\mathcal{A}}[a_i] \leftarrow R_{M_i}$
12: Recommended suppression action $a \leftarrow \text{argmax}_{a_i} \tilde{\mathcal{A}}[a_i]$

Algorithm 4: [Simplified] Suppression planner (global destruction minimization reward model)

Input: Wildfire belief $\mathcal{B}_t$, hyperparameter set $\{Q, RO\}$
Output: Suppression action a
1: $A \leftarrow \text{ASR}(\mathcal{B}_t, Q)$
2: **Initialize** Suppression action score map $\tilde{\mathcal{A}} \leftarrow \varnothing$
3: **for** action $a_i \in A$ **do**
4:      where a_i is $\{X_m, Y_m, DR\}$
5:      $\mathcal{B}_{st} \leftarrow \text{suppress}(X_m, Y_m, DR, \mathcal{B}_t)$
6:      $\mathcal{B}_{st+RO} \leftarrow \text{propagate}(\mathcal{B}_{st}, RO)$
7:      $P_{M_i} \leftarrow \Sigma_{x \in \mathcal{B}_{st+RO}} \mathcal{D}(x)$
8:      $\tilde{\mathcal{A}}[a_i] \leftarrow P_{M_i}$
9: Recommended suppression action $a \leftarrow \text{argmin}_{a_i} \tilde{\mathcal{A}}[a_i]$

---

and elevation, but only limited fuel, data is used. While the quality of the internal propagation model affects reward optimization, quality results may be attained for even a low-fidelity model. Belief map $\mathcal{B}_t$ is also submitted to one of three ASR functions with quantile selection Q to attain a smaller action space set A. Each suppression action a_i in set A is then applied to $\mathcal{B}_t$, the results of which are propagated to depth RO to attain $\mathcal{B}_{st+RO}$. A grid with size proportional to RO is centered on action location (X_m, Y_m), and corresponding subarrays $\mathcal{L}$ and $\mathcal{L}_s$ are indexed from $\mathcal{B}_{t+RO}$ and $\mathcal{B}_{st+RO}$ respectively. The summed instantaneous destruction in both subarrays is calculated and subtracted from one another to attain reward R_{M_i}.

Global resource destruction minimization, shown in Fig. 10, allows for the direct comparison of any action taken across the global grid following the appropriate propagation sequence. However, the scope of destruction often causes the impact of the selected suppression activity to get lost amid the extent of the stochastic fire spread. Algorithm 4 outlines the process by which a suppression action a is selected using global rewards. Unlike the localized approach, the global reward model does not initialize a single nonsuppressed, propagated reference grid at the onset. Instead, action roll-out rewards are compared directly. The belief map $\mathcal{B}_t$ is again submitted to one of three ASR functions with quantile selection Q to attain a smaller action space set A. Each suppression action a_i in set A is then applied to $\mathcal{B}_t$, the results of which are propagated to depth RO to attain $\mathcal{B}_{st+RO}$. The summed instantaneous destruction of $\mathcal{B}_{st+RO}$, now global rather than localized, is used to attain penalty P_{M_i}.

Three ASR methods are presented in Fig. 11. ASR method 1 limits suppression actions to cells burning in the belief map; it can be assumed that ideal suppression locations involve at least one burning cell. ASR method 1 reduces the overall action space by 90–95%. ASR method 2 restricts suppression to a predetermined percentage of

cells, here held at 5, 10, or 15%, that are furthest away from the fire origin. ASR method 2 recognizes that suppressing the wildfire exterior and wildfire head results in the greatest likelihood of minimizing overall wildfire propagation. ASR method 2 reduces the overall action space by 99% or more. ASR method 2 may be modified to increase the percentage of considered wildfire cells, the set of which is then randomly selected. A stochastic wildfire does not spread evenly, and consideration of cells to include those not furthest away may better suppress multiple fire regions. ASR method 3 further restricts ASR method 2 to two 60 deg arcs, one centered on the area of highest resource value and the other on the wildfire head. Cells qualified by ASR method 2 that are also within the established arcs qualify for suppression per ASR method 3. These arcs may shift as environmental factors change throughout the simulation.

MCTS with UCT is again applied to a generative MDP model with Algorithm 3 or Algorithm 4 used to determine action rewards. Each MCTS suppression calculation is capped at 120 s of calculation time to allow the suppression calculation to updated belief information while also permitting time to execute the suppression action before the next iteration.

D. Early Dispatch

The development of an intertwined surveillance-suppression model enables the early allocation of additional suppression assets in instances where wildfire growth exceeds even the optimized capabilities of a single aircraft. Challenges include when to consider requesting additional assets, and what thresholds should determine that additional assets are required. The vast majority of initial attack fires do not become escaped fires, and premature dispatching of aircraft is therefore costly and wasteful. At the same time, initial attack fires destined to become escaped fires have limited windows of opportunity in which additional suppression can result in containment; after a certain point, the wildfire spread is too significant to be reasonably contained except by a full-blown suppression operation.

action 1: grid propagated post-suppression   action 2: grid propagated post-suppression

global instantaneous destruction 1   <   global instantaneous destruction 2
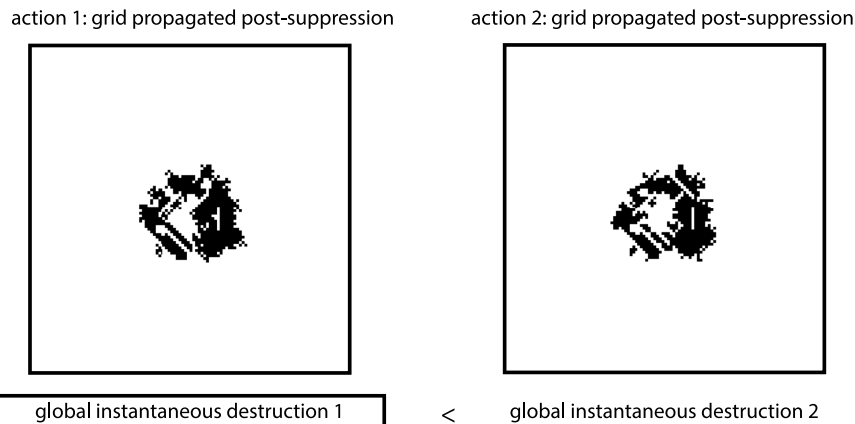
Fig. 10 The global suppression destruction minimization reward model. Unlike in the localized model, there is no propagated reference grid. The wildfire belief map is suppressed, then propagated, and then compared across suppression actions.
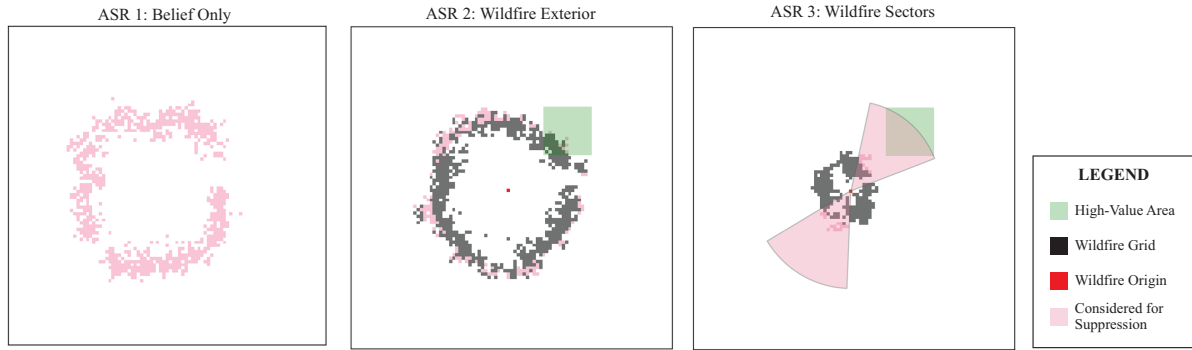
Fig. 11 Three action space restriction schemes: belief only, wildfire exterior, and wildfire sectors based on high-value areas and the wildfire head.

Algorithm 5:   Early dispatch procedures

Input: Wildfire belief $\mathcal{B}_t$, historical wildfire ring array $\mathcal{R}$
Output: Early dispatch recommendation
1: **if** $t \bmod d_T = 0$ **then**
2:      $\mathcal{R}_t \leftarrow \text{radius}(\mathcal{B}_t)$
3:      $\mathcal{R}[t] \leftarrow \mathcal{R}_t$
4:      $\mathcal{R}_{t=120} \leftarrow \text{regression}(\mathcal{R})$
5:      **if** ($t >$ time threshold) and ($\mathcal{R}_{t=120} >$ ring threshold) **then**
6:          Recommend early dispatch
7:      **else**
8:          Recommend against early dispatch

A regression-backed approach is introduced in Algorithm 5, wherein the wildfire ring radius at $t = 120$ min is predicted at each time step T following suppression execution. If the predicted wildfire ring radius exceeds a given time and ring threshold, then a second S-70 is dispatched to the scene to conduct suppression activities.

V.   Experimentation

A.   Environmental Models

1.   Abstracted Case Studies

Three abstracted case studies are presented to simulate the introduced hierarchical framework, surveillance and suppression reward models, and MCTS extensions in various environmental conditions. Cases 1, 2, and 3 are shown in Fig. 12. Case 1 involves flat terrain, variable winds, and a single high-value resource area. Case 2 again features flat terrain, has two high-value resource areas, but also experiences a significant but randomized wind shift at $t = 60$ min. Case 3 has hilly to mountainous elevation, variable winds, and three high-value resource areas. Each progressive case features increased opportunity for destroyed resources.

2.   Actual Wildfire Case Study

The state of Hawaii is particularly susceptible to wildfires due to an abundance of fire-prone grasses and shrubs and an increasingly warm and dry climate [45,46]. Roughly 0.5% of Hawaii's total landmass burns annually, a proportion greater than or roughly equal to that of

any other American state [47]. We examine an initial attack fire in the Makaha Valley on the island of Oahu on 19 October 2007, which would go on to burn approximately 500 acres [48]. The 2007 Makaha Valley wildfire required 3 days, more than 60 firefighters, and 2 manned helicopters conducting water drops to contain. Although wildfire perimeter data was not available, the initial wildfire location, historical wind data, elevation map, and fuelbed characteristics were aggregated to simulate and recreate the wildfire. Figure 13 shows three photographs of the 19 October 2007 Makaha Valley fire [48], along with a satellite photograph of the 10-acre grid where the wildfire began [49], the associated LANDFIRE fuelbed map [36], and historical wind directionality trends in the Makaha Valley [50]. Surface friction typical of mountainous environments such as the Makaha Valley can result in reduced wind speeds.

B.   Surveillance

1.   Baseline

A myopic baseline is introduced in which unmanned aircraft are jointly rewarded R_o at time step t for the number of cells observed that had a change in state compared to the belief map. Previous literature typically restricts this reward to only newly identified burning cells, as opposed to newly identified extinguished cells, to encourage surveillance along the wildfire head. However, without an accurate identification of the wildfire tail, suppression assets may be guided to previously extinguished cells. Reward R_o may be weighted to emphasize observations of higher-resource cells. Penalties P_u, P_m, and P_i remain in effect.

$$R_o = \tau_1 |\{x | B(x) \neq \mathcal{W}(x)\}| \tag{12}$$

$$R_U = R_o - P_u - P_m - P_i \tag{13}$$

2.   Hyperparameters

A three-dimensional grid search is conducted over a range of discount factors, depths, and exploration constants to identify a high-performing set of three hyperparameters for surveillance MCTS. Table 1 shows all surveillance hyperparameters and considered values. The model performs well when setting the discount factor to 0.95, the depth of search to three, and the exploration constant to 100. The total iteration limit is 1000, and the computation time limit is capped at 30 s.
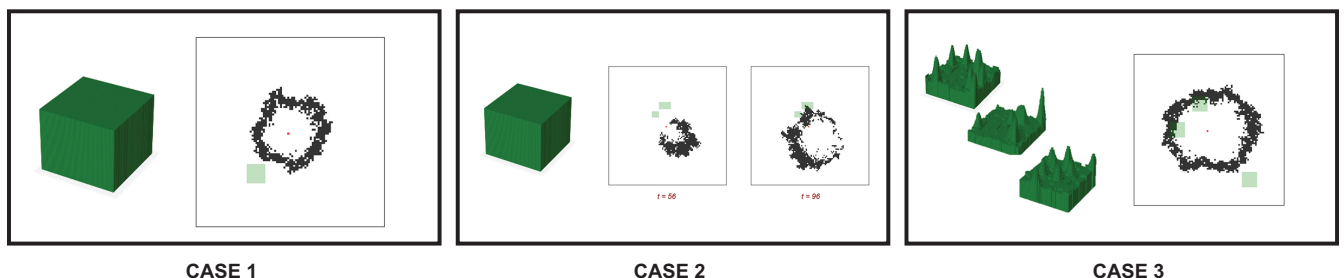

Fig. 12   Three abstracted environmental models featuring different terrain elevation maps, environmental wind profiles, and number and size of high-value areas.
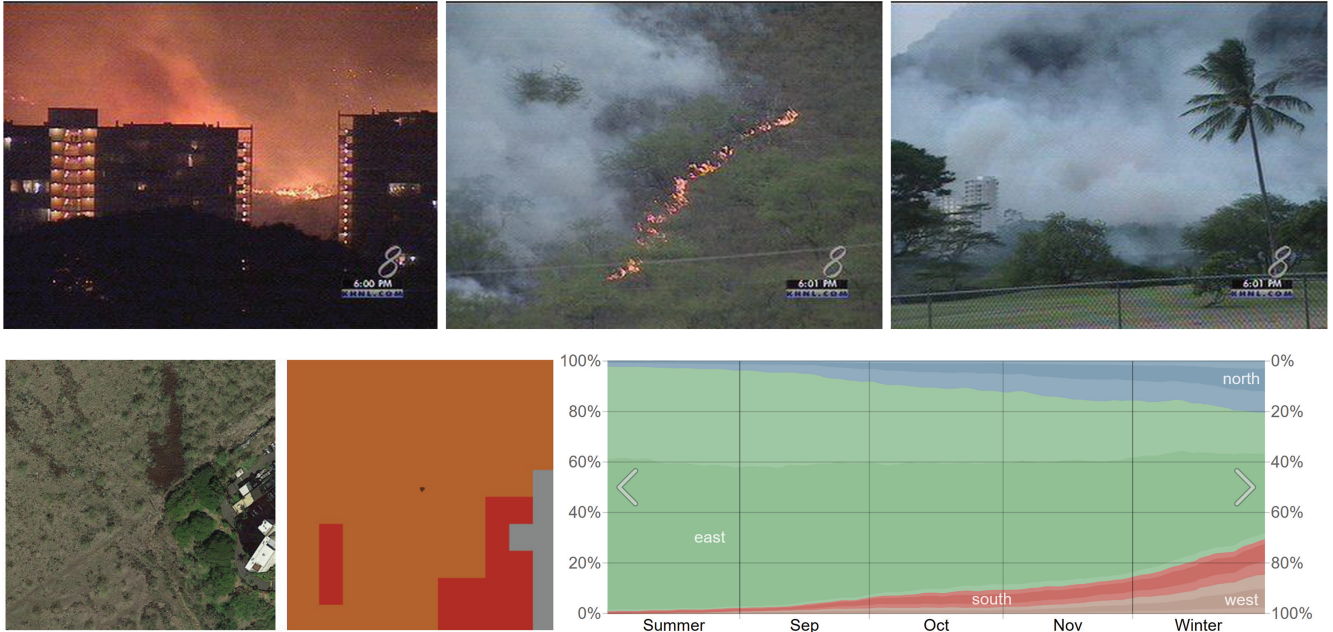
Fig. 13   Top three: photographs of the 19 October 2007 Makaha Valley fire. From bottom left to bottom right: a satellite photograph of the 10-acre grid where the wildfire began, the associated LANDFIRE fuelbed map, and historical wind directionality trends in the Makaha Valley.

Table 1   Surveillance planner hyperparameters and considered values (selected values are in bold)

Hyperparameter	Value, s
Computation time limit	30 s
Iteration limit	1000
Discount factor	{0.80, 0.85, 0.90, **0.95**, 0.99}
Depth	{1, 2, **3**, 4}
Exploration constant	{10, 50, **100**, 200, 1000}

Table 2   Suppression planner hyperparameters and considered values (selected values in bold)

Hyperparameter	Value, s
Computation time limit	120 s
Iteration limit	1000
Discount factor	{0.80, 0.85, 0.90, **0.95**, 0.99}
Depth	{1, **2**, 3}
Exploration constant	{10, 50, **100**, 200, 1000}
ASR selection	{1, **2**, 3}
Quantile	{80, **90**}
Roll-out depth	{5, **10**, 15}

The computational time limit must provide sufficient time for the unmanned aircraft to act on the guidance received before reevaluation.

3. Simulation

Abstracted case study 1 was simulated 20 times with select hyperparameters to compare the myopic baseline against the uncertainty reward model. Suppressive activities were not undertaken to help isolate the effects of both policies without external interference. Two surveillance accuracy metrics were considered: the accuracy of the overall belief map relative to the actual wildfire state and the accuracy of the burning cells in the belief map relative to the actual wildfire state.

C. Suppression

1. Baselines

Firefighting technique and immediate suppression baselines are introduced for comparison against our approach. Firefighting technique reflects traditional firefighting policy and employs a conditions-based, multistep approach to suppression. If resources are unevenly distributed, wet lines are placed around high-value resource areas in an order reflecting their proximity to the wildfire. Thereafter, each drop is aimed at the head of the fire, defined as the distance furthest from the fire origin. The drop selection is made based on the increased proximity of the drop to the fire; for example, a north–south drop when placing a wet line east or west of the fire. Immediate suppression is a myopic policy that rewards actions that maximize the number of wildfire cells extinguished by suppressive activity. MCTS with UCT is applied to optimize immediate suppression policy results.

2. Hyperparameters

Three-dimensional grid searches are conducted to identify a high-performing set of six hyperparameters for suppression MCTS.

We first simulate over discount factor, search depth, and exploration constant, then over ASR selection, quantile, and roll-out depth. Table 2 shows all suppression hyperparameters and considered values. The model performs well when setting search depth to two, exploration constant to 100, ASR selection to two, quantile to 90, and roll-out depth to 10. We note that a search depth of three outperforms a search depth of two for the incipient stages of the wildfire. However, the search depth of three falters precipitously as the action space grows with the wildfire. Similarly, during the later stages of the wildfire's growth, a search depth of one outperforms a search depth of two. This suggests an expanding action space may be supported by progressively restricting the search depth or branching factor over the lifetime of the model through the use of double progressive widening or similar [51]. The total iteration limit is 1000, and the computation time limit is capped at 120 s. The computational time limit must provide sufficient time for the manned aircraft to act on the guidance received before reevaluation.

3. Simulation

Abstracted case study 1 was simulated 20 times with select hyperparameters to compare the immediate suppression baseline and firefighting technique against our localized and global reward models. Surveillance activities were not undertaken to help isolate the effects of all policies without external interference; perfect information was assumed by equating the belief map to the actual wildfire state. Three suppression metrics were considered: total resources destroyed by the wildfire, wildfire flame size (number of burning grid cells), and average wildfire ring radius. Individual wildfires were then categorized as either fully suppressed, contained, or escaped. A fully

suppressed wildfire indicates that suppressive activities were able to extinguish the entirety of the wildfire. A contained wildfire indicates plateauing growth, defined as a wildfire ring size at $t = 120$ min within 10% of the average of the last three (rapid, ultrarapid spread) or four (moderate spread) wildfire ring size states.

D. Joint Surveillance and Suppression

1. Simulation

Abstracted case studies 1, 2, and 3 were simulated 20 times with hyperparameters shown in Tables 1 and 2 to compare the immediate suppression baseline and firefighting technique against our localized and global reward models, using imperfection wildfire information attained using the uncertainty surveillance model. Three suppression metrics were again considered: total resources destroyed by the wildfire, wildfire flame size (number of burning grid cells), and average wildfire ring radius.

E. Early Dispatch Procedures

Early dispatch windows are shown in gray in Fig. 14. Should the predicted wildfire ring size at $t = 120$ min coincide with this window, a secondary aircraft is dispatched to aid in suppressive activity. The predictive accuracy for a case 1 model using linear regression is quite good and stabilizes to within 10% of the final outcome around 20 minutes when there is no suppressive activity and within 40 minutes when there is a single aircraft conducting suppression. In situations involving wind shifts, irregular elevation profiles, or uneven fuel maps, a machine-learning model may be integrated to better predict the extent of the wildfire propagation. Figure 14 also provides insight into the consequences of aircraft suppression across differing wildfire spreads. Unsurprisingly, the less severe the wildfire spread, the greater the impact of added suppressive activity. Perhaps more surprising is the significant impact the addition of a second suppressive aircraft has on all examined wildfires regardless of spread. This suggests a sort of suppression resource "tipping point," after which point wildfire containment is likely—assuming optimal use of resources.

VI. Results

A. Surveillance

1. Emergent Behaviors

Emergence occurs when unique and complex behaviors emerge through the interaction of two or more otherwise simplistic entities. The introduced surveillance planner and associated MCTS solver result in unmanned aircraft exhibiting several emergent behaviors, including dispersion, loitering, circling, and stacking, all shown in Fig. 16. Dispersion occurs when both unmanned aircraft depart from an area of interest to clear room for the manned aircraft, only to return to their original positions immediately following the manned aircraft's departure. Loitering involves one unmanned aircraft avoiding the wildfire entirely while the other maintains full freedom of maneuver around the wildfire. Circling occurs when unmanned aircraft follow one another in a circular pattern. Stacking involves two unmanned aircraft in close lateral proximity with significant altitude between them, such that one unmanned aircraft provides a high-level view of the wildfire while the other provides a condensed low-level view of the wildfire. These emergent behaviors regularly combine with one another (such as circling and stacking), and are typically more prevalent during certain stages of a wildfire's propagation (loitering typically occurs during the initial stages of a wildfire, whereas circling and stacking occur during the latter stages).

2. Simulation Data

As demonstrated in Fig. 15, the uncertainly model increasingly outperforms the belief baseline as the wildfire spread becomes increasingly severe. The belief baseline performs well during the early stages of wildfire response (20–40 min from inception) and when wildfire spread is slow. When the wildfire is small enough to be local to either unmanned aircraft, querying the immediate vicinity is all that is required, especially with a search depth of up to three. As the wildfire expands beyond the reach of the search tree, weighting the recency of past queries on a by-grid cell basis becomes beneficial. The uncertainty model does this by maintaining surveillance recency
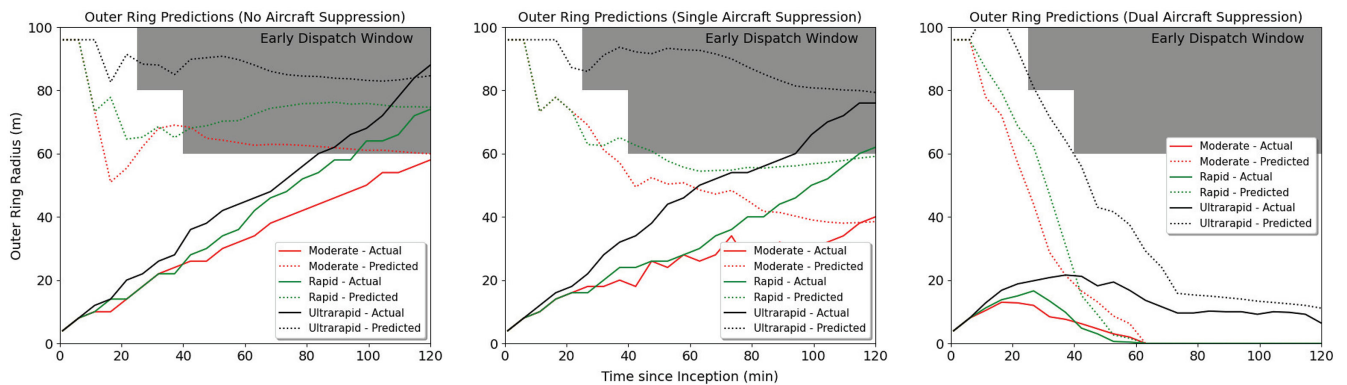

Fig. 14 Linear regression is used to predict wildfire ring size at $t = 120$ min for various case 1 wildfire spread rates with no, single-aircraft, and dual-aircraft suppression applied.
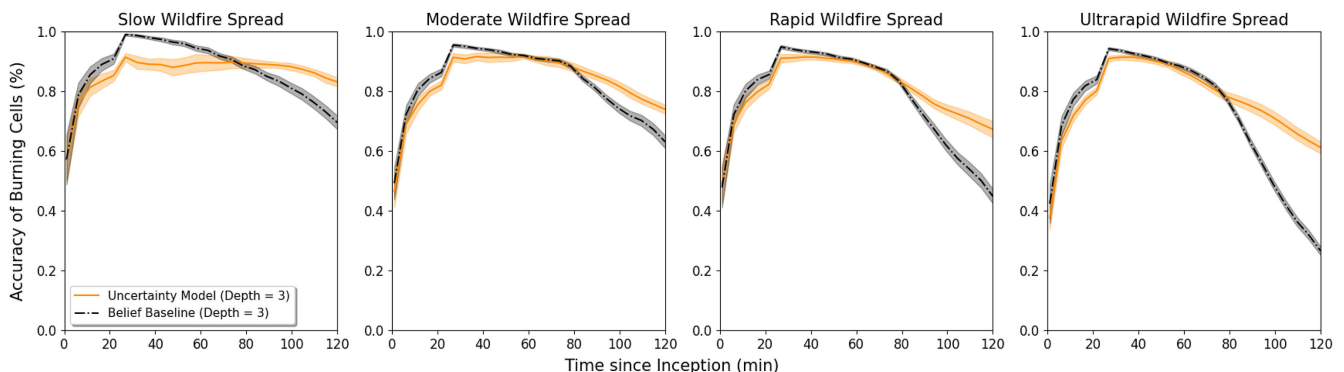

Fig. 15 Surveillance accuracy of burning cells in % with 95% CI ranges for slow, moderate, rapid, and ultrarapid wildfire spread.

data across the entire wildfire in a regularly updated uncertainty map. Given that the uncertainty model rivals or exceeds the performance of the belief baseline in each stage of wildfire propagation and across all spread rates, the uncertainty model will be used exclusively for later joint surveillance and suppression simulations.

B. Suppression

1. Policy Behavior

The three suppression policy categories under consideration each uniquely prioritize different aspects of the wildfire, as evidenced by their behavior in simulation. The comparison of firefighting techniques against optimized destruction minimization policies, shown in Fig. 17, indicates several commonalities, however. The destruction minimization policies also place strategic wet lines, although not with the expediency of the firefighting technique; they prefer to immediately reduce the extent of the initial attack fire rather than preemptively place wet lines alongside high-value areas. This extinguishes the fire in its infancy, when the greatest effect can be had through suppression. Otherwise, the destruction minimization policies also prioritize the head and exterior of the fire. Destruction minimization policies typically select wet lines that simultaneously suppress at least some portion of the wildfire, rather than treating wet lines and suppression as largely exclusive. The immediate suppression baseline is the most myopic of the three categories, as it discards any consideration of future outcomes in favor of extinguishing the greatest amount of wildfire in the present.

2. Simulation Data

As shown in Figs. 18 and 19 and in Table 3, localized and global resource destruction minimization policies outperform the immediate baseline across all wildfire spreads and are likely to outperform firefighting techniques in moderate and rapid wildfire spreads. Both forms of resource destruction minimization appear to perform about evenly. Resource destruction minimization suppression results vary increasingly as the wildfire spread becomes increasingly severe; this may be attributed to the unevenness between wildfires that are fully suppressed and those that escape.

C. Joint Surveillance and Suppression

1. Simulation Data

Localized and global resource destruction minimization policies outperformed the immediate baseline for moderate wildfire spreads and are likely to outperform the immediate baseline for rapid and ultrarapid wildfire spreads. Localized and global resource destruction minimization policies are also likely to outperform firefighting techniques in moderate wildfire spreads and are on par for rapid and ultrarapid wildfire spreads. Both forms of resource destruction minimization continue to perform about evenly. Joint surveillance and suppression results are depicted in the Appendix, in Table A1 and Figs. A1 and A2. Notably, imperfect as opposed to perfect surveillance information has a significantly more detrimental effect on the optimized policies as opposed to firefighting techniques. Accurate wildfire data is required to fight the wildfire in a manner
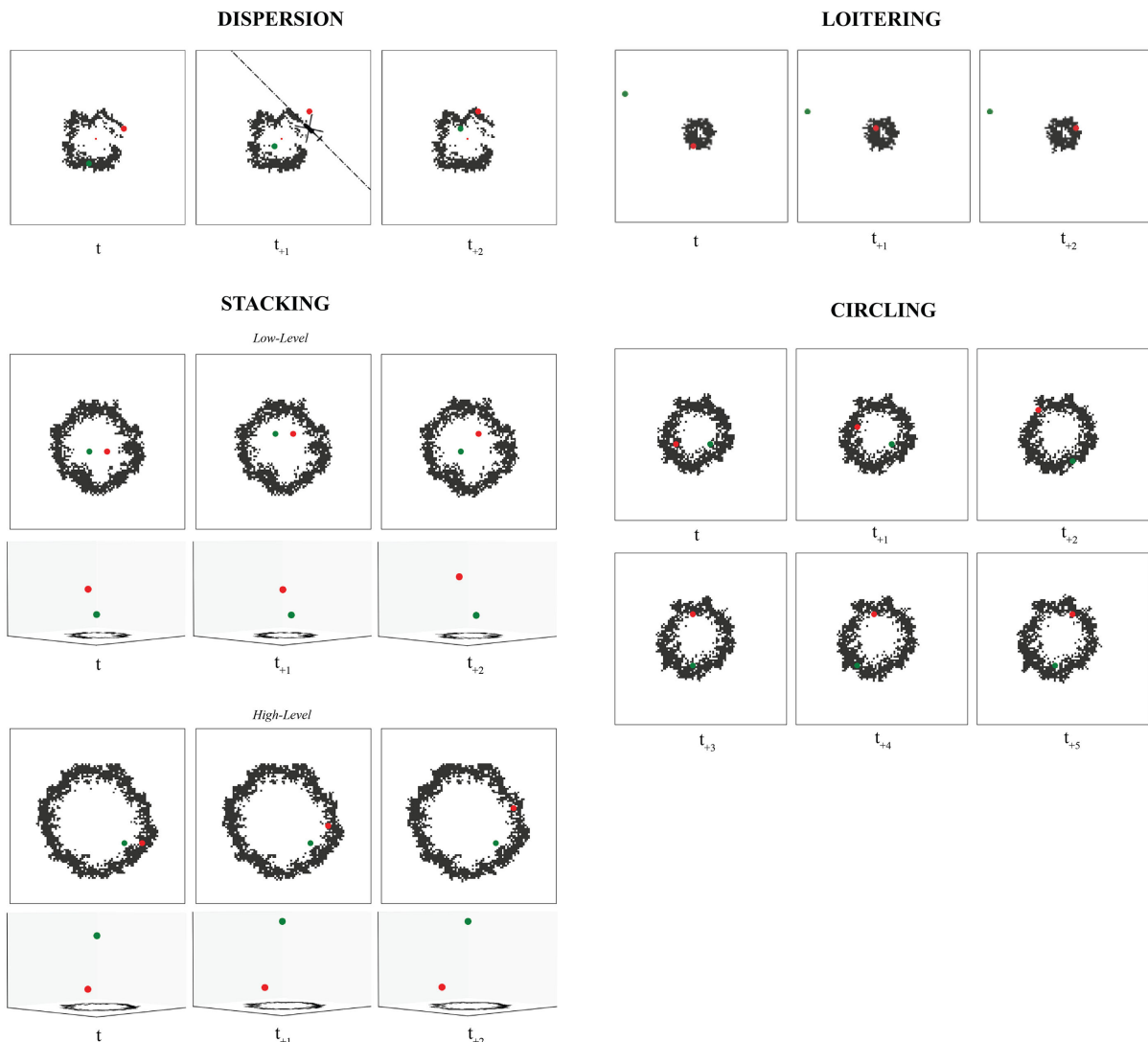

Fig. 16 Unmanned aircraft exhibiting various forms of emergent behavior to include dispersion, loitering, stacking, and circling.

Firefighting Technique (Baseline)

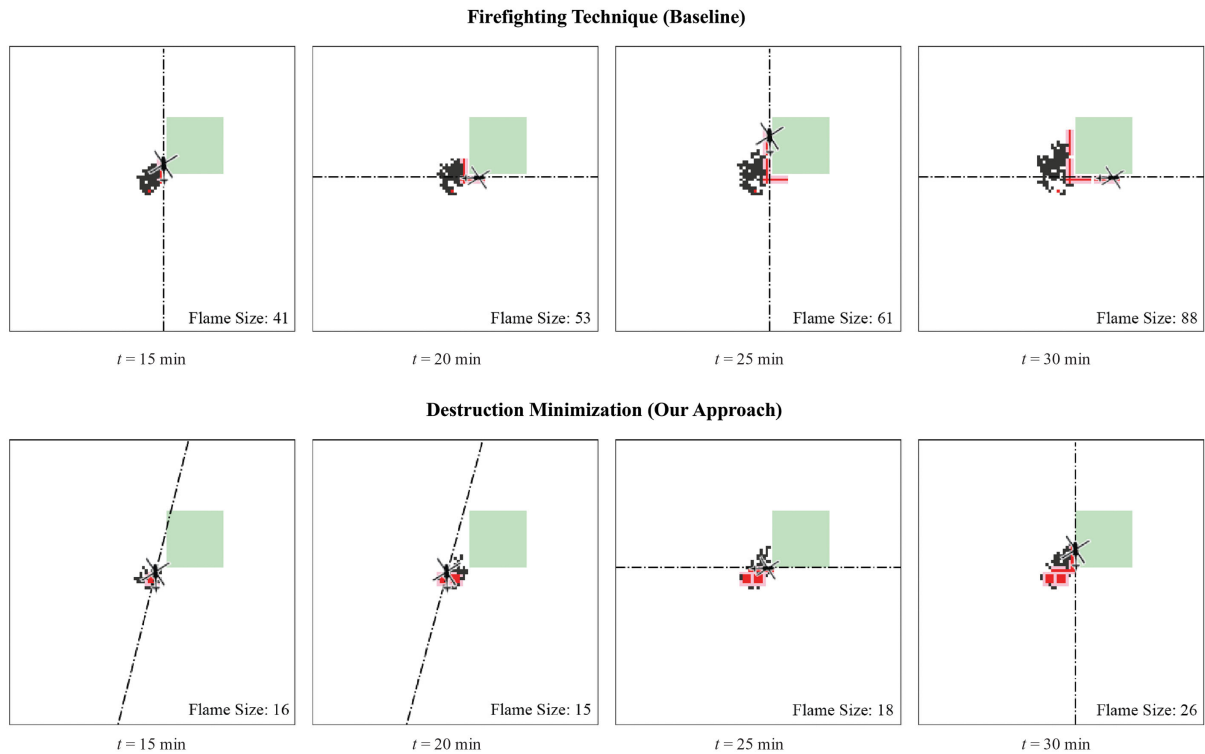**Destruction Minimization (Our Approach)**

Fig. 17   Limitations of an overly proactive approach in firefighting technique compared to the strategically proactive approach in our destruction minimization models.
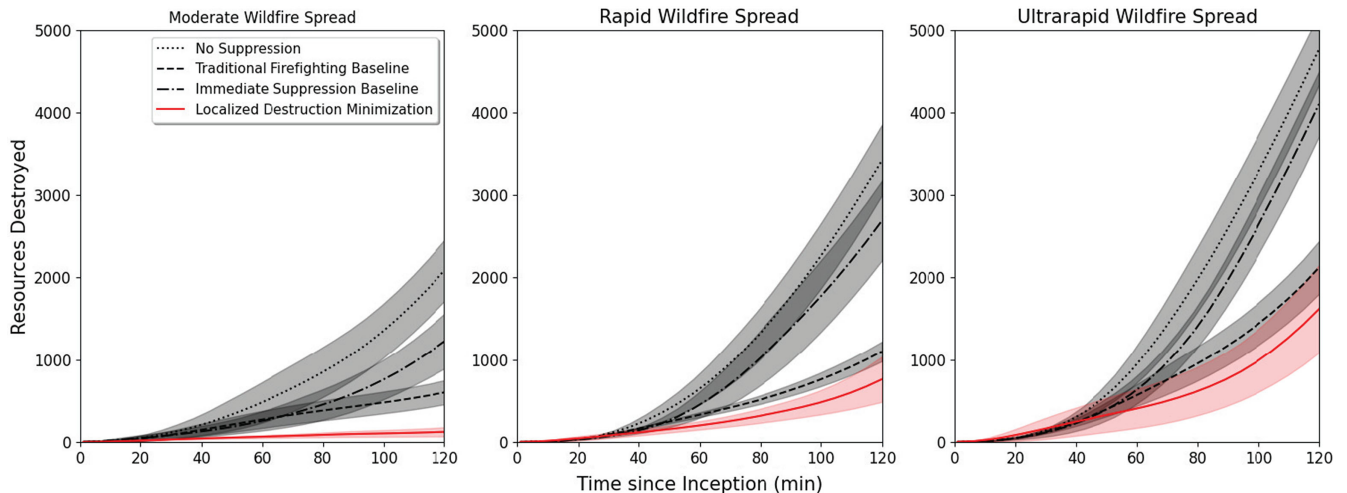

Fig. 18   A 20-run average comparison of suppression policy selection on resources destroyed over time with 95% CI ranges for moderate, rapid, and ultrarapid wildfire spreads in case 1 given a perfect surveillance information assumption.

that results in full suppression, whereas placing wet lines along boundaries and high-value areas is more forgiving. These results hold for cases 2, 3, and 4, despite the increase in environmental complexity. It may be inferred that improving surveillance accuracy is among the most effective ways to optimize suppression results.

Figure 20 analyzes the end-state of the propagated wildfires postsuppression by bucketing them into one of three categories: fully suppressed, contained, or escaped. Escaped fires represent between 1 and 17% of all wildfires in the United States but result in 97% of the overall burned landmass [52,53]. Escaped fire containment requires a significant expansion in suppression capability by firefighting agencies through the employment of multiple ground and air assets. As shown, the percentage of escaped wildfires in moderate, rapid, and ultrarapid wildfire spreads all greatly exceed 1–17% when firefighting techniques are applied. The firefighting technique resulted in containment in

5–15% of wildfire simulations across all spreads, outperforming the immediate suppression baseline and no suppression whatsoever. This indicates our focus on initial attack fires with the potential to become escaped fires. For reference, the joint surveillance and suppression framework with an optimized resource destruction minimization model applied results in a 100% full suppression rate when the wildfire spread corresponds to an 83% full suppression rate using firefighting techniques. Figure 20 also demonstrates the effectiveness of resource destruction minimization models, with and without early dispatch, relative to the immediate suppression baseline and firefighting techniques. The only instances of fully suppressed wildfires across all three wildfire spreads were when resource destruction minimization models were applied. The moderate wildfire spread chart does not feature an early dispatch variation of the destruction minimization model due to the slow propagation sequence not triggering the selected early dispatch window.
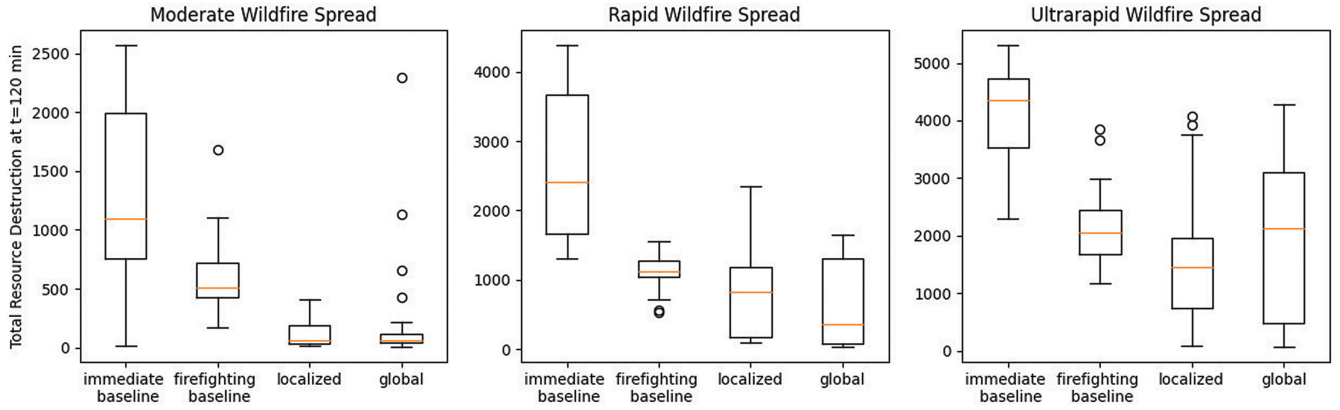
Fig. 19   A 20-run comparison of suppression policy selection on resources destroyed at $t = 120$ min for moderate, rapid, and ultrarapid wildfire spreads in case 1 with a perfect surveillance information assumption.

Table 3   Suppression performance (final destruction), case 1, 20 run average, 95% CI

Suppression method (perfect information assumption)	Moderate fire spread	Rapid fire spread	Ultrarapid fire spread
	Baselines		
Firefighting technique	600.40 ± 146.52	1100.18 ± 121.02	2120.49 ± 320.49
Immediate suppression	1223.14 ± 334.92	2687.47 ± 483.72	4107.61 ± 393.40
	Our methods		
Localized destruction minimization	$\mathbf{122.66 \pm 55.57}$	$\mathbf{760.37 \pm 275.75}$	1617.31 ± 532.71
Global destruction minimization	$\mathbf{272.49 \pm 241.16}$	$\mathbf{669.38 \pm 274.02}$	1910.78 ± 628.51

Bold indicates methods that outperform both baselines by a statistically significant margin ($\alpha = 0.05$).
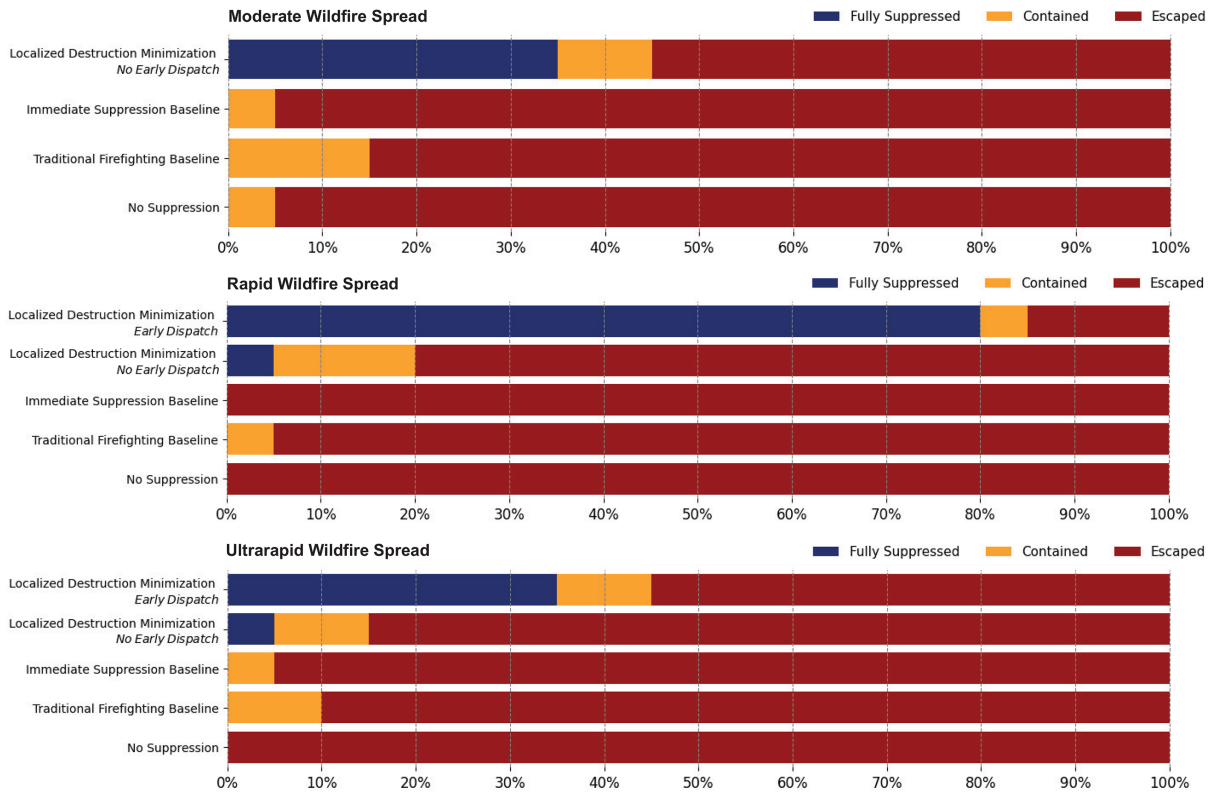

Fig. 20   A 20-run comparison of suppression policy selection to include early dispatching on wildfire status category at $t = 120$ min for moderate, rapid, and ultrarapid wildfire spreads in case 1 with imperfect surveillance information.

VII.   Discussion

A.   System Capability

The introduced hierarchical framework is designed to integrate unmanned surveillance aircraft into existing initial attack operations with minimal disruption to participating manned aircraft.

It is therefore important to consider the assumptions that enable this capability to function as intended. Those assumptions are related to network architecture, autonomy modes in the case of network degradation, free and stable communication, and enabling hardware.

A MPOMDP, as opposed to a Dec-POMDP, has no impediment to communication [30]—and yet no network is infallible. In our framework, communication exists between unmanned aircraft, between manned aircraft, and between unmanned and manned aircraft. Although there are several feasible approaches to aircraft network architecture, we propose a dual-layer system. The low-level layer is a decentralized mesh network with self-healing properties where aircraft communicate their location in real-time to other networked aircraft. This robust layer aids in collision avoidance. The high-level layer is hierarchical and mirrors the introduced framework. One unmanned aircraft is assigned to be the unmanned network lead. This assignment may rotate such that the unmanned aircraft most central to the mesh, in terms of relative position or strength of network connection, becomes the lead. The unmanned network lead is responsible for 1) receiving and fusing observation data across unmanned aircraft, 2) updating the wildfire belief map, 3) resolving the surveillance planner using the updated wildfire belief map, 4) disseminating surveillance guidance to unmanned aircraft, and 5) disseminating the updated wildfire belief map to all aircraft. A manned network lead may be assigned if there are multiple participating manned aircraft. Alternatively, manned aircraft may independently resolve their own suppression planners. This suits when one manned aircraft is actively suppressing the fire while the other is replenishing water off-station.

Human-autonomous frameworks typically traverse an autonomy hierarchy in response to changes in the network and operating environment. This has significant repercussions for system scalability and robustness. While the ideal case assumes a shared wildfire belief map and centralized decision-making within homogeneous agent groups, this need not be true in our framework. Individual decision processes can be resolved independently if the high-level network layer is degraded. This highlights one advantage of dividing the larger MPOMDP into smaller MDPs. Given lapses in communication with the respective network lead, all aircraft can proceed, if suboptimally, to the best of their independent ability and per their individual MDPs and belief maps. If the network lead is damaged, another aircraft assumes the network lead role. If a nonlead aircraft is damaged, the framework continues to operate unabated, albeit with fewer observations or suppression actions. Separately, it is critical that the low-level network layer remain intact. An unmanned aircraft that disconnects from the decentralized mesh network and is therefore no longer receiving location information must depart the wildfire vicinity immediately or risk collision. Given the inherent risks involved with manned aircraft integration, we would expect additional fail-safe measures to mitigate collision avoidance concerns during communication lapses.

The initial attack is limited to 10 acres, or a roughly 200-m-by-200-m grid. This is small enough to ensure reliable communication between aircraft using any one of a number of wireless protocols, including WiFi or Bluetooth 5.0. The only exception occurs when the suppression aircraft departs the wildfire to replenish its water bucket. The nearest water replenishing source may be upward of 10 km away, in which case there is limited communication between manned aircraft off-station and the unmanned network lead on-station. This results in the manned aircraft possessing a stale wildfire belief map. Suppression planner performance using the stale wildfire belief map is determined by comparing the frequency of suppression to how quickly the wildfire expands. In slowly and moderately propagating fires, optimizing suppression on a belief map a few minutes old is still likely to provide useful results. Alternatively, the suppression planner can resolve on the condition that the manned aircraft receives an updated wildfire belief map from the unmanned network lead while en-route back to the fire. This ensures a more accurate wildfire belief at the time of suppression planner execution but reduces the amount of time available for computation. Satellite-based communication or the presence of a nearby ground station can mitigate network reliability concerns.

The initial attack wildfire is constrained in time and is either suppressed, contained, or becomes an escaped fire at the 120-min mark. The vast majority of manned suppression aircraft can remain airborne for at least 2 h. For example, the S-70 has 150 min of flight time when filled with fuel. More limiting is the battery life for unmanned aircraft, especially those with multirotors. There exist a handful of higher-end electric and hybrid surveillance quadcopters

that have between 55 and 120 minutes of flight time, depending on the payload. This flight time is expected to incrementally improve with advancements in lithium batteries and composite materials. The presence of recharge stations or the dispatch of a second fleet of unmanned aircraft, can help maintain continuous surveillance operations over the course of an expanding, incipient wildfire.

B. Limitations and Future Work

The hierarchical framework introduced, while promising, is subject to certain limitations. These can generally be categorized as model-specific, solver-specific, or domain-specific. From those limitations, we identify opportunities for further analysis.

1. Model-Specific

MDPs are excellent tools for capturing the dynamics underpinning and uncertainties affecting complex systems, though they are not without limitations. MDPs are subject to the "curse of dimensionality," meaning that they grow exponentially as the number of states and actions increases [10]. Through a combination of wildfire-specific constraints, probabilistic search algorithms, and the decomposition of the MPOMDP into hierarchically arranged subproblems, we overcame the high-dimensional state and action spaces associated with the initial attack and attained significant and meaningful results. As the initial attack grows beyond the confines of its regulated boundaries and becomes an escaped fire, we would expect the state and suppression action spaces to swell. Additionally, if more aircraft were introduced, surveillance and suppression action spaces would exponentially enlarge. In both cases, the introduced constructs would become less effective. To maintain a similar resolution for surveillance and suppression operations for an escaped fire with an increase in the number of participating aircraft, we suggest an added hierarchical layer, now within homogeneous groupings in addition to just between them. This transforms our hierarchy from an iterating series of planners into an iterating series of subhierarchies. We leave a more extensive review of the nested MDP hierarchy design and associated computational complexity for the escaped wildfire problem to future research. A second model-specific limitation is surveillance planner frequency and the merging of collision avoidance penalties and wildfire surveillance rewards into one objective function. While the duration between surveillance decisions can be easily modified, there needs to be sufficient time between decisions to 1) apply the solver and get results and 2) actually execute surveillance operations and attain observations. This time scale may differ from that required for robust collision avoidance, which is like to operate at a much higher frequency. In application, the frequency of location sharing between aircraft would be greater than that of observation sharing. We therefore suggest a low-level "detect and avoid" system for each unmanned aircraft. The surveillance planner introduced then effectively keeps unmanned aircraft away from the expected manned aircraft axis of advance, while the low-level detect and avoid system informed by location sharing data adjusts for unexpected changes in manned aircraft trajectory.

2. Solver-Specific

MCTS is an online planning algorithm and therefore an effective solver choice for the initial attack problem. MCTS can account for nonstationary behavior in the initial attack wildfire. This includes instantaneous changes in wildfire propagation direction and intensity, as demonstrated with the randomized wind shift simulated in abstracted case study two. Despite its upside, MCTS has certain limitations that must be addressed. MCTS suffers from sample inefficiency in large search spaces, is prone to high variance and faces domain-specific challenges related to its internal model. We address state aggregation via nested hierarchical design in the model-specific limitation section above, which can minimize an otherwise expansive search tree and enable MCTS to obtain accurate statistics and make informed decisions. MCTS roll-outs can vary significantly due to the random nature of simulations. This may result in inconsistent action value estimations and induce problematic noise. Parallelization, variance reduction, and progressive widening techniques can mitigate these concerns. MCTS uses an internal model to conduct simulations,

and solver performance is therefore a consequence of how well that internal model reflects reality. This is not an issue when MCTS is used for deterministic games, but becomes concerning when applied to environments rife with uncertainty. The disparity between reality and the internal model is exacerbated with depth. As such, a typical disadvantage of model-based methods is that compounding errors make long-horizon roll-outs unreliable. The depth considered during surveillance planning is shallow enough to avoid these compounding errors while still remaining useful due to the frequency of reevaluation.

3. Domain-Specific

While we have sought to develop a high-fidelity initial attack model, we have not so far addressed fire intensity and flame height, smoke plumes, mixed fleet operations, and the ember attack. Unmanned aircraft attain higher resolution observation data by descending, but doing so may put them at risk of catching fire. Descending below an established safety distance from the ground may result in a penalty applied to the violating aircraft's objective function. A more tailored approach involves modifying the safety distance as a function of flame intensity or height for a given cell column. We have assumed a binary mapping of the wildfire state, but we can reasonably include an assessment of both flame intensity, given the appropriate sensors, and flame height using three-dimensional wildfire propagation and observation models. Pham et al. apply a safety distance to unmanned aircraft tracking a wildfire and also introduce a flame intensity model as an artificial potential field [13]. Smoke plumes obscure surveillance aircraft's vision-based observations, induce significant errors in sensor measurements [54], and modify manned suppression aircraft's axis of advance. Smoke sensors or low-cost cameras outfitted for unmanned aircraft may be used to develop three-dimensional keep-out geofences [55] for smoke plumes. We address agent heterogeneity by separating available assets into homogeneous manned and unmanned aircraft groups. The suppression patterns and partial suppression probabilities in this paper are tailored to a 660-gallon water bucket on a short line hauled by an S-70 Firehawk helicopter. These patterns and probabilities may be adjusted for any

manned aircraft and water suppression platform. We would expect fundamentally different results for a CH-47 Chinook hauling a 2600-gallon water bucket on a long line. As the initial attack wildfire escapes, a variety of new manned aircraft arrive on station, including fixed-wing tankers, helicopters carrying various sizes of water buckets, and smoke-jumper aircraft. The set of joint actions across the diversity of manned aircraft may give rise to interesting emergent behaviors but must also comply with Fire Traffic Area altitudes, orbits, and routing structures. The set of considered neighboring cells during wildfire propagation may be expanded to incorporate the spread of embers across sizable distances in the case of an ember attack. An ember attack occurs when wildfires in strong wind conditions carry embers beyond the fire front. We focus on the general form of the initial attack and therefore consider only adjacent neighbors during propagation. MDP models exist that have incorporated airborne ember spread for larger wildfires [12].

VIII.   Conclusions

The coordination of wildfire surveillance and suppression activities using manned and unmanned aircraft in tandem is an effective means of reducing wildfire propagation severity and minimizing wildfire destruction. A hierarchical framework involving iterating surveillance and suppression planners is introduced to divide an otherwise intractable MPOMDP into optimizable subproblems acting on asynchronous but otherwise consistent time scales. Surveillance, suppression, and joint surveillance and suppression models with unique MCTS extensions applied are compared in simulation across abstracted and actual case studies. We demonstrate how a hierarchical approach to Markov decision processes may be used to ensure collision avoidance between unmanned and manned aircraft operating in close proximity and how teaming aerospace operations may extend into wildfire management. We further find that our hierarchical framework with a resource destruction minimization reward model significantly outperforms firefighting techniques and a myopic baseline in preventing initial attack fires from developing into escaped fires.

Appendix: Joint Wildfire Surveillance-Suppression Performance Across Policies

Table A1   Joint surveillance-suppression performance (final destruction), 20 run average, 95% CI

Case	Suppression method	Moderate	Rapid	Ultrarapid
	Uncertainty surveillance model	Fire spread	Fire spread	Fire spread
	Baselines			
	Firefighting technique	663.16 ± 153.86	1416.30 ± 270.56	2504.41 ± 259.58
1	Immediate suppression	1247.83 ± 351.86	2523.71 ± 442.56	3909.05 ± 373.87
	Our methods			
	Localized destruction minimization	444.48 ± 151.70	1455.15 ± 244.14	2727.90 ± 480.10
	Global destruction minimization	$\mathbf{348.12 \pm 133.91}$	1602.76 ± 390.03	2581.25 ± 305.61
	Baselines			
	Firefighting technique	1050.80 ± 134.23	1586.44 ± 130.94	2515.49 ± 227.91
2	Immediate suppression	1552.63 ± 298.54	2653.19 ± 364.15	3354.86 ± 364.23
	Our methods			
	Localized destruction minimization	$\mathbf{417.43 \pm 209.84}$	1210.73 ± 489.17	2494.31 ± 610.34
	Global destruction minimization	$\mathbf{445.21 \pm 233.31}$	1549.57 ± 535.60	2581.25 ± 305.61
	Baselines			
	Firefighting technique	1792.59 ± 382.81	2396.35 ± 465.55	3728.30 ± 441.38
3	Immediate suppression	1532.93 ± 592.88	4173.03 ± 677.81	6483.51 ± 665.34
	Our methods			
	Localized destruction minimization	$\mathbf{447.74 \pm 185.17}$	$\mathbf{1493.11 \pm 532.93}$	4778.15 ± 892.61
	Global destruction minimization	$\mathbf{934.72 \pm 406.79}$	1753.15 ± 551.94	4145.07 ± 651.05
	Baselines			
	Firefighting technique	1061.27 ± 90.40	2020.31 ± 131.77	3160.39 ± 165.56
4	Immediate suppression	690.67 ± 160.69	2010.83 ± 262.97	3690.32 ± 372.43
	Our methods			
	Localized destruction minimization	$\mathbf{288.61 \pm 113.47}$	$\mathbf{1329.51 \pm 226.86}$	3290.26 ± 407.79
	Global destruction minimization	$\mathbf{434.65 \pm 143.71}$	1477.92 ± 321.53	2800.06 ± 332.47

Bold indicates methods that outperform both baselines by a statistically significant margin ($\alpha = 0.05$).
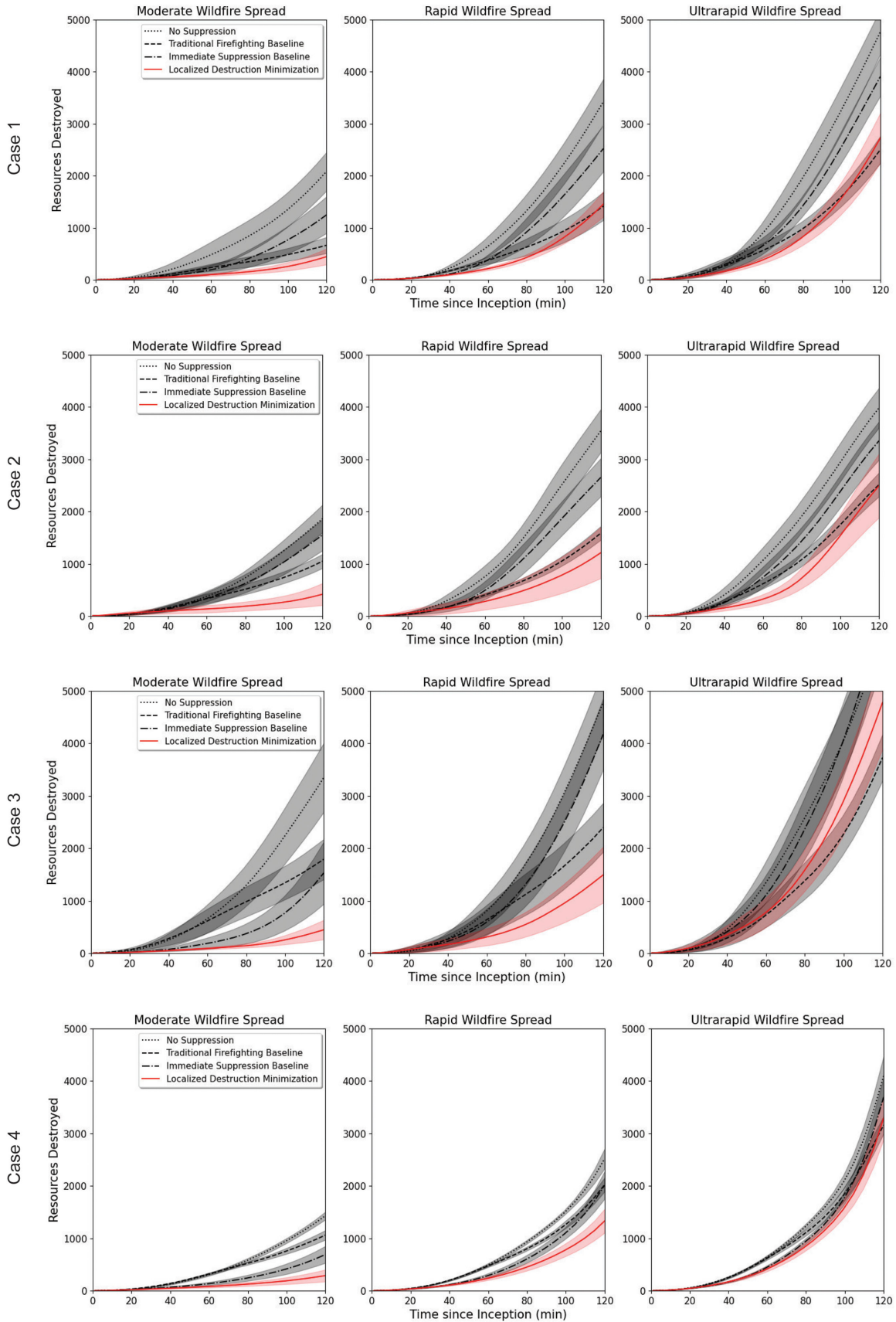
Fig. A1 **A 20-run average comparison of suppression policy selection on resources destroyed over time with 95% CI ranges for moderate, rapid, and ultrarapid wildfire spreads in all four cases with imperfect surveillance information.**
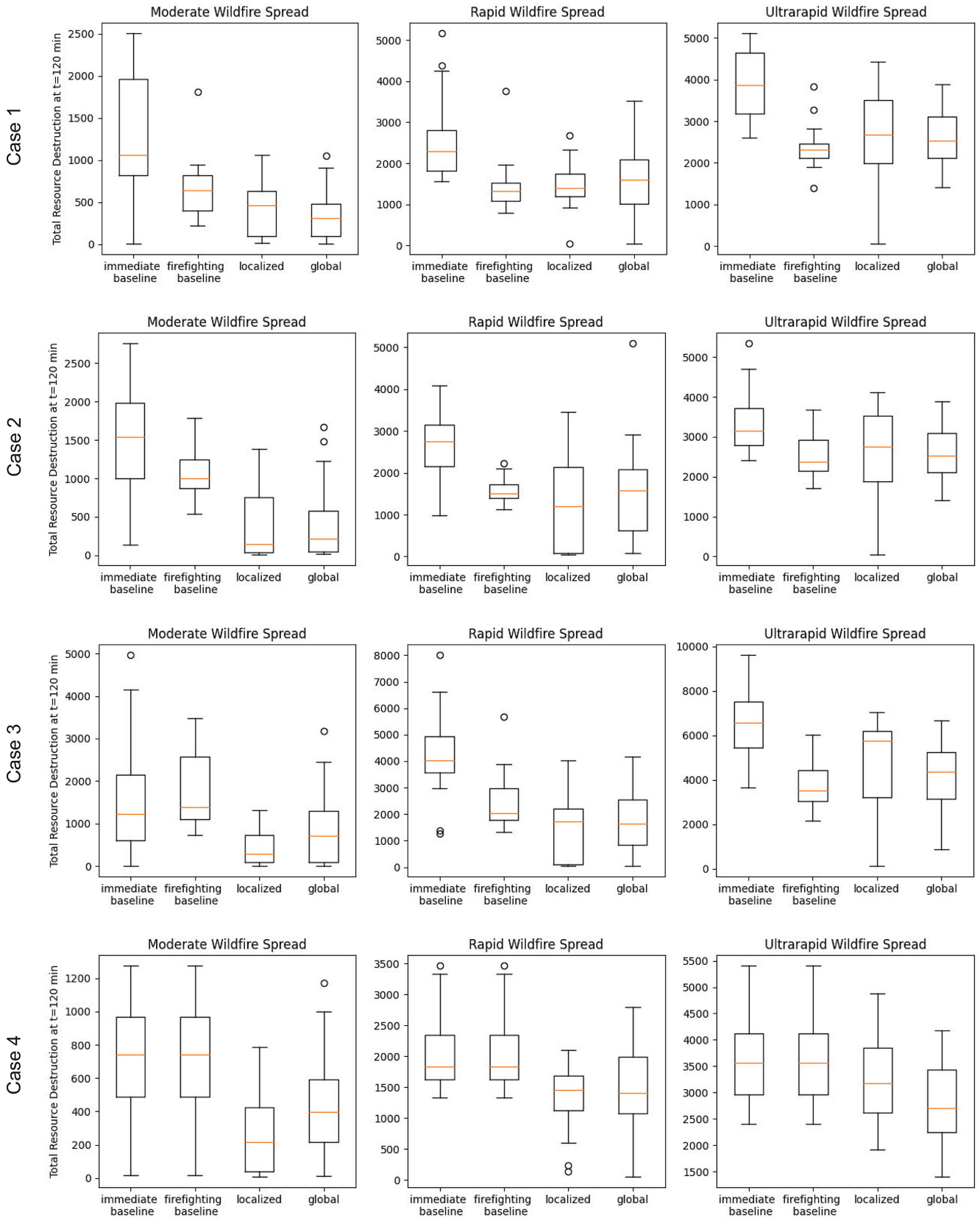
Fig. A2   A 20-run comparison of suppression policy selection on resources destroyed $t = 120$ min for moderate, rapid, and ultrarapid wildfire spreads in all four cases with imperfect surveillance information.

Acknowledgments

References

[1] Hoover, K., and Hanson, L. A., "Wildfire Statistics," 2023, https://crsreports.congress.gov/product/pdf/IF/IF10244.

[2] Hirsch, K. G., Corey, P. N., and Martell, D. L., "Using Expert Judgment to Model Initial Attack Fire Crew Effectiveness," *Forest Science*, Vol. 44, No. 4, 1998, pp. 539–549.
https://doi.org/10.1093/forestscience/44.4.539

[3] Rahn, M., *Initial Attack Effectiveness: Wildfire Staffing Study*, Montezuma Publ., San Diego State Univ., San Diego, CA, 2010, pp. 1–13.

[4] "2022 Strategic Fire Plan Amador El Dorado Unit," California Department of Forestry and Fire Protection, 2022, https://osfm.fire.ca.gov/media/irefprho/2022-amador-el-dorado-alpine-sacramento-unit-fire-plan.pdf.

[5] O'Neill, T., McNeese, N., Barron, A., and Schelble, B., "Human–Autonomy Teaming: A Review and Analysis of the Empirical Literature," *Human Factors*, Vol. 64, No. 5, 2022, pp. 904–938.
https://doi.org/10.1177/0018720820960865

[6] Gaydos, S. J., and Curry, I. P., "Manned-Unmanned Teaming: Expanding the Envelope of UAS Operational Employment," *Aviation, Space, and Environmental Medicine*, Vol. 85, No. 12, 2014, pp. 1231–1232.
https://doi.org/10.3357/ASEM.4164.2014

[7] "Firefighting Aircraft Recognition Guide," California Dept. of Forestry and Fire Protection, 2019, https://www.fire.ca.gov/media/4950/aviation-guide-2019-access.pd.

[8] Ambrosia, V. G., Wegener, S., Zajkowski, T., Sullivan, D., Buechel, S., Enomoto, F., Lobitz, B., Johan, S., Brass, J., and Hinkley, E., "The Ikhana Unmanned Airborne System (UAS) Western States Fire Imaging Missions: From Concept to Reality (2006–2010)," *Geocarto International*, Vol. 26, No. 2, 2011, pp. 85–101.
https://doi.org/10.1080/10106049.2010.539302

[9] Albaker, B., and Rahim, N., "A Survey of Collision Avoidance Approaches for Unmanned Aerial Vehicles," *2009 International Conference for Technical Postgraduates (TECHPOS)*, Inst. of Electrical and Electronics Engineers, New York, 2009, pp. 1–7.
https://doi.org/10.1109/TECHPOS.2009.5412074

[10] Ghavamzadeh, M., Mahadevan, S., and Makar, R., "Hierarchical Multi-Agent Reinforcement Learning," *Autonomous Agents and Multi-Agent Systems*, Vol. 13, No. 2, 2006, pp. 197–229.
https://doi.org/10.1007/s10458-006-7035-4

[11] Seraj, E., Chen, L., and Gombolay, M. C., "A Hierarchical Coordination Framework for Joint Perception-Action Tasks in Composite Robot Teams," *IEEE Transactions on Robotics*, Vol. 38, No. 1, 2021, pp. 139–158.
https://doi.org/10.1109/TRO.2021.3096069

[12] Julian, K. D., and Kochenderfer, M. J., "Distributed Wildfire Surveillance with Autonomous Aircraft Using Deep Reinforcement Learning," *Journal of Guidance, Control, and Dynamics*, Vol. 42, No. 8, 2019, pp. 1768–1778.
https://doi.org/10.2514/1.G004106

[13] Pham, H. X., La, H. M., Feil-Seifer, D., and Deans, M. C., "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, 2018, pp. 1537–1548.
https://doi.org/10.1109/TSMC.6221021

[14] Ghamry, K. A., and Zhang, Y., "Cooperative Control of Multiple UAVs for Forest Fire Monitoring and Detection," *2016 12th IEEE/ASME International Conference on Mechatronic and Embedded Systems and Applications (MESA)*, Inst. of Electrical and Electronics Engineers, New York, 2016, pp. 1–6.
https://doi.org/10.1109/MESA.2016.7587184

[15] Griffith, J. D., Kochenderfer, M. J., Moss, R. J., Mišic, V. V., Gupta, V., and Bertsimas, D., "Automated Dynamic Resource Allocation for Wildfire Suppression," *Lincoln Laboratory Journal*, Vol. 22, No. 2, 2017, pp. 38–59.

[16] Hobbs, K. L., and Li, B., "Safety, Trust, and Ethics Considerations for Human-AI Teaming in Aerospace Control," *AIAA Scitech 2024 Forum*, AIAA Paper 2024-2583, 2024.
https://doi.org/10.2514/6.2024-2583

[17] Bjurling, O., Granlund, R., Alfredson, J., Arvola, M., and Ziemke, T., "Drone Swarms in Forest Firefighting: A Local Development Case Study of Multi-Level Human-Swarm Interaction," *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, Assoc. for Computing Machinery, New York, 2020, pp. 1–7.
https://doi.org/10.1145/3419249.3421239

[18] Basich, C., Svegliato, J., Wray, K. H., Witwicki, S., Biswas, J., and Zilberstein, S., "Competence-Aware Systems," *Artificial Intelligence*, Vol. 316, March 2023, Paper 103844.
https://doi.org/10.1016/j.artint.2022.103844

[19] Veloso, M., Biswas, J., Coltin, B., and Rosenthal, S., "Cobots: Robust Symbiotic Autonomous Mobile Service Robots," *24th International Joint Conference on Artificial Intelligence*, Citeseer, AAAI Press, Washington, D.C., 2015, pp. 4423–4429.
https://doi.org/10.5555/2832747.2832901

[20] Seraj, E., and Gombolay, M., "Coordinated Control of UAVs for Human-Centered Active Sensing of Wildfires," *2020 American Control Conference (ACC)*, Inst. of Electrical and Electronics Engineers, New York, 2020, pp. 1845–1852.
https://doi.org/10.23919/ACC45564.2020

[21] Menda, K., Chen, Y.-C., Grana, J., Bono, J. W., Tracey, B. D., Kochenderfer, M. J., and Wolpert, D., "Deep Reinforcement Learning for Event-Driven Multi-Agent Decision Processes," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 4, 2018, pp. 1259–1268.
https://doi.org/10.1109/TITS.6979

[22] Barto, A. G., and Mahadevan, S., "Recent Advances in Hierarchical Reinforcement Learning," *Discrete Event Dynamic Systems*, Vol. 13, Nos. 1–2, 2003, pp. 41–77.
https://doi.org/10.1023/A:1022140919877

[23] Bacon, P.-L., Harb, J., and Precup, D., "The Option-Critic Architecture," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31, No. 1, Assoc. for the Advancement of Artificial Intelligence (AAAI) Press, Washington, D.C., 2017, pp. 1726–1734.
https://doi.org/10.1609/aaai.v31i1.10916

[24] Vezhnevets, A. S., Osindero, S., Schaul, T., Heess, N., Jaderberg, M., Silver, D., and Kavukcuoglu, K., "Feudal Networks for Hierarchical Reinforcement Learning," *International Conference on Machine Learning*, PMLR/JMLR, Cambridge, MA, 2017, pp. 3540–3549.
https://doi.org/10.48550/arXiv.1703.01161

[25] Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., and Mordatch, I., "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," *Advances in Neural Information Processing Systems*, Vol. 30, Curran Assoc. Inc., Red Hook, NY, 2017, pp. 6382–6393.
https://doi.org/10.48550/arXiv.1706.02275

[26] Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., and Whiteson, S., "Counterfactual Multi-Agent Policy Gradients," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32, No. 1, Assoc. for the Advancement of Artificial Intelligence (AAAI) Press, Washington, D.C., 2018, pp. 2974–2982.
https://doi.org/10.1609/aaai.v32i1.11794

[27] Rashid, T., Samvelyan, M., De Witt, C. S., Farquhar, G., Foerster, J., and Whiteson, S., "Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning," *Journal of Machine Learning Research*, Vol. 21, No. 1, 2020, pp. 7234–7284.
https://doi.org/10.48550/arXiv.2003.08839

[28] Xu, Z., Bai, Y., Zhang, B., Li, D., and Fan, G., "Haven: Hierarchical Cooperative Multi-Agent Reinforcement Learning with Dual Coordination Mechanism," *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, No. 10, Assoc. for the Advancement of Artificial Intelligence (AAAI) Press, Washington, D.C., 2023, pp. 11,735–11,743.
https://doi.org/10.1609/aaai.v37i10.26386

[29] Pierre, J.-E., Sun, X., and Fierro, R., "Multi-Agent Partial Observable Safe Reinforcement Learning for Counter Uncrewed Aerial Systems," *IEEE Access*, Vol. 11, July 2023, Paper 78192.
https://doi.org/10.1109/ACCESS.2023.3298601

[30] Kochenderfer, M. J., *Decision Making Under Uncertainty: Theory and Application*, MIT Press, Cambridge, MA, 2015, p. 180.

[31] Gmytrasiewicz, P. J., and Doshi, P., "A Framework for Sequential Planning in Multi-Agent Settings," *Journal of Artificial Intelligence Research*, Vol. 24, July 2005, pp. 49–79.
https://doi.org/10.1613/jair.1579

[32] Kaelbling, L. P., Littman, M. L., and Cassandra, A. R., "Planning and Acting in Partially Observable Stochastic Domains," *Artificial Intelligence*, Vol. 101, Nos. 1–2, 1998, pp. 99–134.
https://doi.org/10.1016/S0004-3702(98)00023-X

[33] Bertsimas, D., Griffith, J. D., Gupta, V., Kochenderfer, M. J., and Mišic, V. V., "A Comparison of Monte Carlo Tree Search and Rolling Horizon Optimization for Large-Scale Dynamic Resource Allocation Problems,"

European Journal of Operational Research, Vol. 263, No. 2, 2017, pp. 664–678.
https://doi.org/10.1016/j.ejor.2017.05.032

[34] Papadopoulos, G. D., and Pavlidou, F.-N., "A Comparative Review on Wildfire Simulators," *IEEE Systems Journal*, Vol. 5, No. 2, 2011, pp. 233–243.
https://doi.org/10.1109/JSYST.2011.2125230

[35] Saikin, D. A., Baca, T., Gurtner, M., and Saska, M., "Wildfire Fighting by Unmanned Aerial System Exploiting Its Time-Varying Mass," *IEEE Robotics and Automation Letters*, Vol. 5, No. 2, 2020, pp. 2674–2681.
https://doi.org/10.1109/LSP.2016.

[36] "Hawaii LANDFIRE—LF 2020 Fuel Characteristic Classification System Fuelbeds," Landfire, 2020, https://www.landfire.gov/viewer/ [retrieved 19 Aug. 2023].

[37] Hansen, R., "Corrigendum to: Estimating the Amount of Water Required to Extinguish Wildfires Under Different Conditions and in Various Fuel Types," *International Journal of Wildland Fire*, Vol. 21, No. 6, 2012, pp. 778–778.
https://doi.org/10.1071/WF11022_CO

[38] "CALFIRE Sonoma-Lake-Napa Unit 2023 Strategic Fire Plan," California Department of Forestry and Fire Protection, 2023, https://osfm.fire.ca.gov/media/e33a3ior/2023-sonoma-lake-napa-unit-fire-plan.pdf.

[39] Kal'avský, P., Petríček, P., Kelemen, M., Rozenberg, R., Jevčák, J., Tomaško, R., and Mikula, B., "The Efficiency of Aerial Firefighting in Varying Flying Conditions," *International Conference on Military Technologies (ICMT)*, Inst. of Electrical and Electronics Engineers, New York, 2019, pp. 1–5.
https://doi.org/10.1109/ICMT47403.2019

[40] Ault, R., Mooney, C., and Thomasson, J., "Drop Patterns for the Bambi-Max Bucket and the Bell 212 Helitanker," Aug. 2012, https://library.fpinnovations.ca/en/viewer?file=%2fmedia%2fFOP%2f9331.pdf#search=Drop%20patterns%20for%20the%20Bambi-Max%20bucket%20and%20the%20Bell%20212%20helitanker&phrase=false.

[41] Kocsis, L., and Szepesvári, C., "Bandit Based Monte-Carlo Planning," *European Conference on Machine Learning*, Springer–Verlag, Berlin, 2006, pp. 282–293.
https://doi.org/10.1007/11871842

[42] Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., and Colton, S., "A Survey of Monte Carlo Tree Search Methods," *IEEE Transactions on Computational Intelligence and AI in Games*, Vol. 4, No. 1, 2012, pp. 1–43.
https://doi.org/10.1109/TCIAIG.2012.2186810

[43] Li, W., Liu, Y., Ma, Y., Xu, K., Qiu, J., and Gan, Z., "A Self-Learning Monte Carlo Tree Search Algorithm for Robot Path Planning," *Frontiers in Neurorobotics*, Vol. 17, July 2023, Paper 1039644.
https://doi.org/10.3389/fnbot.2023.1039644

[44] Powley, E., Cowling, P., and Whitehouse, D., "Memory Bounded Monte Carlo Tree Search," *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, Vol. 13, No. 1, Assoc. for the Advancement of Artificial Intelligence (AAAI) Press, Washington, D.C., 2017, pp. 94–100.
https://doi.org/10.1609/aiide.v13i1.12932

[45] Ellsworth, L. M., Litton, C. M., Dale, A. P., and Miura, T., "Invasive Grasses Change Landscape Structure and Fire Behaviour in Hawaii," *Applied Vegetation Science*, Vol. 17, No. 4, 2014, pp. 680–689.
https://doi.org/10.1111/avsc.12110

[46] Zhu, T. R., "Modeling Fuels and Wildfire Behavior in Hawaiian Ecosystems," Ph.D. Dissertation, Univ. of Hawai'i at Manoa, Honolulu, HI, 2019.

[47] "State of Hawai'i 2018 Hazard Mitigation Plan," Tetra Tech, 2018, https://dod.hawaii.gov/hiema/files/2020/06/2018-State-HI-HMP-Update-100218.pdf.

[48] Kim, L., "Makaha Valley Residents Return Home as Fire Nears Containment," Hawaii News Now, Oct. 2007, https://www.hawaiinewsnow.com/story/7243086/makaha-valley-residents-return-home-as-fire-nears-containment/.

[49] "Makaha Valley 1: 1200," Google Maps, 2023, https://goo.gl/maps/4KqzyzjqPYCz3Soy5 [retrieved 19 Aug. 2023].

[50] "Wind Direction in the Fall in Makaha Valley," Weather Spark, 2023, https://weatherspark.com/s/96/2/Average-Fall-Weather-in-M%C4%81kaha-Valley-Hawaii-United-States#Figures-WindDirection [retrieved 19 Aug. 2023].

[51] Sunberg, Z., and Kochenderfer, M., "Online Algorithms for POMDPs with Continuous State, Action, and Observation Spaces," *Proceedings of the International Conference on Automated Planning and Scheduling*, Vol. 28, Assoc. for the Advancement of Artificial Intelligence (AAAI) Press, Washington, D.C., 2018, pp. 259–263.
https://doi.org/10.1609/icaps.v28i1.13882

[52] Reimer, J., Thompson, D. K., and Povak, N., "Measuring Initial Attack Suppression Effectiveness Through Burn Probability," *Fire*, Vol. 2, No. 4, 2019, p. 60.
https://doi.org/10.3390/fire2040060

[53] Calkin, D. E., Gebert, K. M., Jones, J. G., and Neilson, R. P., "Forest Service Large Fire Area Burned and Suppression Expenditure Trends, 1970-2002," *Journal of Forestry*, Vol. 103, No. 4, 2005, pp. 179–183.
https://doi.org/10.1093/jof/103.4.179

[54] Yuan, C., Zhang, Y., and Liu, Z., "A Survey on Technologies for Automatic Forest Fire Monitoring, Detection, and Fighting Using Unmanned Aerial Vehicles and Remote Sensing Techniques," *Canadian Journal of Forest Research*, Vol. 45, No. 7, 2015, pp. 783–792.
https://doi.org/10.1139/cjfr-2014-0347

[55] Kim, J., and Atkins, E., "Airspace Geofencing and Flight Planning for Low-Altitude, Urban, Small Unmanned Aircraft Systems," *Applied Sciences*, Vol. 12, No. 2, 2022, p. 576.
https://doi.org/10.3390/app12020576

D. Casbeer
Associate Editor