Discover Applied Sciences

RESEARCH Open Access

Active inference motion control for autonomous persistent surveillance

Johan Schubert[1*], Farzad Kamrani[1] and Tove Gustavi[1]

*Correspondence:
Johan Schubert
johan.schubert@foi.se
[1]Swedish Defence Research Agency, Stockholm SE-164 90, Sweden

Abstract

We develop an active inference route-planning method for autonomous control of an intelligent agent. The aim is to reconnoiter a geographical area while balancing search and tracking behaviors to maintain a dynamic situational awareness map. To achieve this, we construct an evidence map that reflects our current understanding of the situation, integrating both positive and "negative" sensor observations of possible target objects collected over time, and diffusing the evidence across the map as time progresses. The generative model of active inference uses Dempster-Shafer theory and a Gaussian sensor model to provide input to the agent. The generative process employs a Bayesian approach to update a posterior probability distribution. We calculate the variational free energy across all locations within the sensor radius by assessing the divergence between a pignistic probability distribution of the evidence map and a posterior probability distribution of the target object, including the level of surprisal associated with receiving new observations. Using the variational free energy, we direct the agent's movement in simulation by taking incremental steps toward a position that minimizes the variational free energy. This approach addresses the challenge of balancing exploration and exploitation, allowing the agent to search large areas while efficiently tracking identified targets. Across 1000 simulation trials, the approach surpasses benchmarks in target observation by an average of 51%–69% across various metrics for a 100-meter sensor radius.

Keywords Active inference, Free energy principle, Dempster-Shafer theory, Autonomous agent, Route-planning, Persistent surveillance

1 Introduction

This paper focuses on active inference [1–3] for the autonomous control of an intelligent agent (e.g., a reconnaissance Unmanned Aerial Vehicle (UAV)) to achieve and maintain accurate situational awareness over time. In scientific literature, this is often referred to as *persistent surveillance* [4]. When the agent's sensors have limited range, it must actively balance exploitation and exploration—tracking familiar objects while searching for unmonitored items of interest to maintain situational awareness. Active inference is a methodology for autonomous decision-making. It is a generic framework based on the concept that a system aims to minimize its surprise when receiving new information. What is unique about active inference is that the method includes two parallel

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

approaches for an agent to seek consistency between reality and the system's description of the environment: either the system's internal representation is updated as new information is received, or actions are taken against the environment to change it so that it is consistent with the agent's perception.

Buckley et al. [3] present the mathematical foundation of the free energy principle, detailing its structure and presenting an agent-based model. The free energy principle integrates action and perception within a unified framework [5]. Its core idea is that all organisms (agents) must counter environmental disturbances and thus minimize surprise. Since surprise is generally unpredictable, agents instead minimize an upper bound called *variational free energy* (VFE). VFE is an information-theoretic concept that quantifies the divergence between two probability distributions and the element of surprise. One distribution reflects our perception of reality, often described as a dynamic situational awareness map (a common operational picture), while the other represents a posterior probability given the current observation. When these two distributions align closely, VFE is low. In active inference, the agent seeks to position itself so that its observations align with its model of reality. Friston [6] shows that this concept broadly applies to biological processes. Furthermore, Friston et al. [7] describe a process theory based on active inference and probabilistic propagation. Levchuk et al. [8] studied self-organizing agent teams that can monitor, decide, and act using active inference. Additionally, free energy enables systems to reason about and explain their decisions to humans. Albarracin et al. [9] emphasize how this helps decision-makers understand AI's reasoning. Da Costa et al. [10] propose that human-inspired decision-making techniques can enhance AI explainability, supported by hierarchical models of human decision processes in active inference. Explainable AI (XAI), driven by a generative model in active inference, offers transparency by allowing systems to share perceptions and updates [11]. Hierarchical generative models enable XAI systems to evaluate themselves by processing outputs from internal components as inputs [12].

Section 2 outlines the problem statement. Section 3 reviews related work. Section 4 introduces the fundamentals of active inference and VFE. Section 5 provides a brief overview of Dempster-Shafer theory, which updates the dynamic situational awareness map. Section 6 details the world model and a simplified sensor model that estimates classification confidence, assuming perfect detection in the simulation. In Sect. 7, we mathematically describe how to control an agent via active inference by minimizing VFE. This section also describes a Dempster-Shafer evidence map that combines positive and "negative" sensor observations and diffuses uncertainty to neighboring cells. We intentionally model active inference in this article using VFE rather than *expected free energy* (EFE) with planning over potential agent trajectories (policies). This approach explores VFE's ability to perform persistent surveillance by employing greedy action selection to minimize a local VFE field within the sensor's current range, thereby lowering computational costs. In the VFE approach, we compute the free energy from current and past observations within the sensor footprint to determine the optimal position, whereas EFE computes the free energy based on potential observations to select an ideal trajectory. In both methods, the agent progresses incrementally toward the optimal position or along the optimal path. Section 8 provides an overview of the implementation details, waypoint selection, and its low computational complexity. Section 9 offers both qualitative and detailed quantitative analysis based on 1000 simulated scenarios, comparing VFE

against its two components and two benchmarks. Section 10 concludes with the main findings of this work.

2 Problem statement and persistent surveillance

2.1 Problem statement

In the field of autonomous systems, the task of continuously monitoring a specific area for an indefinite period is known as persistent surveillance. Real-world applications of persistent surveillance systems include monitoring areas around critical infrastructure and military facilities to detect potential intruders with malicious intent. It can also be used for surveillance at popular beaches to prevent drowning incidents and for monitoring wildlife in sensitive natural areas to support preservation. UAVs are particularly well-suited for this type of surveillance because of their wide-angle views, speed, and the relative absence of obstacles in their operational environments, which facilitates trajectory planning for the autonomous agent.

This article considers a scenario in which autonomous UAVs continuously monitor a designated area to maintain an accurate, up-to-date dynamic situational awareness map. The UAVs are tasked with detecting both stationary and moving targets and, ideally, tracking their approximate locations even when they are out of sensor range. Moving targets must be revisited frequently enough to prevent loss. Moreover, every part of the designated area must be observed regularly, with no region left unmonitored for long periods. Although we model a swarm of UAVs using a physical motion model, one UAV serves as the leader (or agent) responsible for all calculations and waypoint selection to minimize VFE.

In persistent surveillance scenarios, the trajectory planning problem for an autonomous agent is typically defined by a set of specific goals and constraints, each with its own priorities. Various methods can be used to implement the actual trajectory planning or motion control (as discussed in Sect. 7). This paper specifically investigates the feasibility of using active inference with VFE for trajectory planning. The primary research question addressed in this paper can be formulated as follows:

Can active inference with VFE be applied to address the motion control problem for a swarm of UAVs, where a single-sensor leader (the agent) performs multi-objective persistent surveillance over a specified two-dimensional (2D) region?

In experiments, we model a swarm of ten UAVs where only the leader is equipped with a sensor, performs all processing, constructs the situational awareness map, and selects a shared waypoint for the entire swarm.

2.2 An active inference approach to persistent surveillance

To effectively implement the active inference framework for motion control in a persistent surveillance context, several key components must be specified.

A generative model enumerates all possible states and their transitions to address these considerations. Information about these states is derived from observations over time, allowing us to update our understanding of the current state. We model uncertainty using the Dempster-Shafer theory [13, 14]. In this model, states are represented stochastically; each possible transition of the target basic belief from one state (e.g., position) to the next is specified by transition probabilities. The states are dynamically

updated at each time step using a diffusion algorithm. This model indicates the current basic belief that at least one target object exists in each possible state (see Sect. 7.1).

A generative process complements the generative model by handling new observations from the agent's sensor. We use a Bayesian approach in this process. It updates the probabilities for all positions within the agent's sensor radius. This update computes a new probability for each position based on the current sensor model, while accounting for the existing probabilities (see Sect. 7.2).

Because we cannot directly minimize the surprise the agent experiences upon receiving new observations, we instead minimize the VFE, which serves as an upper bound on that surprise. The VFE is defined as the sum of the divergence between two probability distributions and the level of surprisal, and it serves as our objective function to guide the agent's movements. The F_{xy}^t objective function in cell c_{xy} at time t is defined as

$$F_{xy}^t = D_{KL}\left[q_{xy}^t(\vartheta) \parallel p_{xy}^t(\vartheta|\varphi)\right] - ln\left[p_{xy}^t(\varphi)\right] \tag{1}$$

where D_{KL} is the divergence between probability of state $q_{xy}^t(\vartheta)$ and posterior probability of state given observations $p_{xy}^t(\vartheta|\varphi)$. The term $-ln\left[p_{xy}^t(\varphi)\right]$ represents the degree of surprisal from an observation (see Sect. 4).

For each location within the sensor radius, we compare the pignistic probability from the generative model with the probability inferred from observations, aiming to minimize their divergence (see Sect. 7.3).

At each time step, we compute the VFE across all locations within the sensor's radius. The agent then takes a fixed-length step in the direction that minimizes VFE. This approach enables the agent to operate autonomously, striving to maintain an accurate situational awareness map under our model.

2.3 Notations

Table 1 provides notations.

3 Related work

3.1 Persistent surveillance and motion control

Over the years, various methods for addressing the control problem in persistent surveillance (also known as persistent monitoring) have been proposed. This section briefly outlines a few examples.

Hari et al. [15] examine a persistent monitoring mission for a single UAV tasked with repeatedly visiting n targets of equal priority. Because the targets are fixed, the problem reduces to a classic traveling salesman problem. Brown and Anderson [16] consider a more complex maritime surveillance scenario that includes UAV dynamic constraints and formulate a multi-objective optimization problem to maximize information gain and minimize fuel consumption. Feasible and optimal solutions are obtained by combining a trajectory-generation method with a particle swarm optimization algorithm. Similar to the problem discussed here, Hübel et al. [17] introduce a dynamic *information map* subject to information decay and use it to derive a gradient-based control that drives a group of agents to continuously update their situational awareness by surveilling an area. Additionally, the authors propose a time-varying density function that can be integrated into the control algorithm to model moving points of interest. Ramasamy and Ghose [18] propose an approach to increase the probability that autonomous agents

Table 1 Variables and definitions

Variable	Value range	Definition
G	–	grid
c_{xy}	–	cell xy in G
Θ	$\{T, F\}$	frame of discernment
2^{Θ}	$\{\varnothing, \{T\}, \{F\}, \{T, F\}\}$	power set of Θ
T	–	Boolean **true**, indicating target(s) present
F	–	Boolean **false**, indicating no target present
ϑ	$\{T, F\}$	state
φ	$\{T, F\}$	observation
p	[0, 1]	probability of observation
q	[0, 1]	probability of state
t	{0, 0.2, …, 500} s	time step
$q_{xy}^{t}(\vartheta)$	[0, 1]	probability of ϑ in c_{xy} at t
$p_{xy}^{t}(\vartheta)$	[0, 1]	a priori probability of state ϑ in c_{xy} at t
$p_{xy}^{t}(\varphi)$	[0, 1]	probability of obtaining an observation φ in c_{xy} at t
$p_{xy}^{t}(\varphi \mid \vartheta)$	[0, 1]	Likelihood, detection probability of φ given ϑ in c_{xy} at t
$p_{xy}^{t}(\vartheta \mid \varphi)$	[0, 1]	posterior probability of ϑ given φ in c_{xy} at t
$n_{xy}^{t}(\vartheta)$	[0, 1]	basic belief of ϑ in c_{xy} at t
$m_{xy}^{t}(\varphi)$	[0, 1]	basic belief of φ in c_{xy} at t
F_{xy}	$[0, \infty)$	variational free energy in c_{xy}
D_{KL}	$[0, \infty)$	Kullback-Leibler divergence
s_r	{10, 20, …, 100} m	sensor radius
C_{Tmax}	0.7	max basic belief for T positive φ at the sensor position
C_{Fmax}	0.6	max basic belief for F "negative" φ at the sensor position
$C_{T,k}$	0.7	max basic belief for T positive φ at the target position k
σ_T	{7.5, 15, …, 75} m	sigma for T in positive φ from the sensor position
σ_F	{5, 10, …, 50} m	sigma for F in negative φ from the sensor position
σ_{pos}	10 m	sigma for T in positive φ from the target position
d	[0, 100] m	Distance from sensor position or target position
N_{UAV}	10	number of UAVs
$N_{targets}$	7	number of targets
N_{moving}	3	number of moving targets
N_{cells}	220 400	number of cells in G
N_{sim}	1000	number of simulations
$a(agent.c_{xy})$	–	action moving agent at c_{xy}
ϕ	[0, 1]	evaluation score

observe moving targets during surveillance missions. Assuming that the probability of observing a target is non-uniformly distributed across a monitored area, Ramasamy and Ghose assign an *importance* degree to each grid point in a discretized map where the UAV in the scenario has detected a target. The importance of a grid point depends on the number of detections at that point and increases the chance that the UAV will revisit that location. Lastly, additional work documented in the literature employs reinforcement learning for persistent surveillance control. Chen et al. [19] use a multi-agent reinforcement learning approach to learn policies for each agent in a team tasked with continuously monitoring a 2D environment with stationary obstacles. To accomplish this, the problem is modeled such that a penalty is applied at each time step if any point in the environment is left unmonitored. Mishra et al. [4] present another reinforcement-learning-based method for persistent surveillance.

3.2 Active inference for estimation and control

Active inference links perception and action through VFE. Theoretical connections to classical estimation and control show that minimizing VFE yields objectives that combine information-theoretic surprise with control costs, leading to linear-quadratic-Gaussian behavior in linear-Gaussian environments [20]. In practice, active inference has been applied to quadcopter state estimation using dynamic expectation maximization (DEM). DEM is a perception scheme inspired by the brain, based on a data-driven model-learning algorithm [21]. Active inference has also been used for adaptive manipulation control in industrial robots, even when the environment's dynamics are not explicitly modeled [22]. Active inference has been applied to the adaptive control of robot arms using multimodal perception-action and variational autoencoder-based state representations. This approach does not require a dynamic or kinematic model of the robot [23]. Furthermore, active inference has been used to develop a torque controller that integrates raw vision and proprioception in a streamlined design for a 7-degrees-of-freedom Franka Emika Panda robot, capable of online adaptation to changes in dynamics and human interference [24]. It has also been employed for fault-tolerant control in the presence of sensor faults, delivering unbiased state estimation and simplifying action specification [25]. These results support our use of VFE for closed-loop control with a *soft-output sensor* model. Here, soft-output means the sensor model provides a graded confidence value rather than only a hard *Target*/*No Target* label; we define and motivate the model in Sect. 6.2–6.4.

3.3 Active inference for navigation, exploration, and bandit problems

For mobile agents, active inference within hierarchical generative models enables goal-oriented navigation with topologically consistent maps and practical robot deployments [26]. Modular active inference systems support flexible goal-driven navigation, obstacle avoidance, and high-confidence path selection, with strong zero-shot generalization to new settings [27]. Curiosity-driven learning for robotic tasks, grounded in the free energy principle, employs Bayesian neural networks to represent epistemic uncertainty and model complex behaviors [28]. Moreover, retrospective (residual) surprise has been introduced as a computational element in active inference, serving as a lower bound on the EFE [29]. In decision-making, contextual multi-armed bandits (CMABs) extend the classical bandit problem by conditioning action selection on observed contextual information. Recent active inference models of CMABs [30, 31] use EFE to guide context-based action selection, balancing exploration and exploitation under uncertainty—a behavior we aim to reproduce in our approach by minimizing a spatial free-energy field over the evidence map.

3.4 Multi-agent active inference, organizational adaptation, and complex tasks

In multi-agent settings, active inference is used to design adaptive organizations in which roles and structures change by minimizing team-level free energy [32–34]. For complex robotic tasks, active inference can be combined with behavior trees to enable continuous planning and robust execution with fewer nodes [35]. Meanwhile, non-modular, cognitively inspired active inference architectures are explored to improve robustness to unknown inputs [36]. Recent work extends active inference to autonomous driving by incorporating action-oriented priors that link perception and control, leading

to more human-like, collision-avoidant behavior [37]. In socially interactive and multi-agent scenarios, active inference applies to human-robot kinesthetic interaction, where meta-priors adjust compliance and counter-forces during physical contact [38], and to empathic, socially compliant agents that adapt their movements to surrounding individuals and contextual norms [39], demonstrating the flexibility of the free-energy framework for coordination and interaction tasks. Active inference also applies to hierarchical, embodied perception-action loops, where multiple sensorimotor pathways associated with different body parts are dynamically combined, enabling the robot to reconfigure control and activate only the necessary joints for a given task [40]. Our method differs in that we operationalize VFE over a spatial Dempster-Shafer evidence map and a deterministic soft-output sensor model, then guide the agent to move in the direction that minimizes this per-cell objective at each step.

3.5 Our contribution

Although active inference is a relatively new concept, it has been tested in various applications of autonomous agent planning and control, as shown in Sect. 3.2–3.4. However, to our knowledge, the literature has not explored or described persistent surveillance using VFE.

Action selection in active inference is generally regarded as planning over policies, such as waypoint trajectories, by minimizing the EFE. Our approach instead offers a computationally efficient minimization over a VFE field. At each time step, the agent advances a fixed step toward the point on the map with the lowest VFE and repeats this process. This greedy approach provides a fast, reactive controller well-suited for persistent surveillance.

In addition, we separate observation and state evidence by introducing a Dempster-Shafer-based evidence map and a deterministic soft-output sensor model to generate cell-by-cell evidential scores. We steer the swarm of UAVs by minimizing a grid-based free-energy objective that employs the Kullback-Leibler divergence [41] to compare a long-term basic belief, expressed as a pignistic distribution [42], with an instantaneous posterior probability derived from the latest observation—including observation surprisal—rather than solely maximizing map intensity or minimizing revisit time.

4 Active inference and variational free energy

The generative model and the generative process are stochastic. Let the frame of discernment be the set of all possible states, denoted by $\Theta = \{T, F\}$, where T denotes the presence of at least one target object, and F represents the absence of a target object. In this section, we use the Bayesian probabilities p (for observation ϕ) and q (for state ϑ).

4.1 The generative model

The generative model enumerates all possible states and transitions [1–3]. In this context, states are represented by grid cells, each corresponding to a possible position on a map. The agent and target objects can transition between states, moving from any grid cell to its nearest neighboring cells.

The probability in the generative model is denoted by $q^t_{xy}(\vartheta)$, where ϑ is the state and q^t_{xy} is the probability for a grid cell c_{xy} at time t. In each grid cell, the probability of at least one target (T) or no targets (F) sums to 1, thus we have

$$q_{xy}^t(\vartheta = T) + q_{xy}^t(\vartheta = F) = 1 \tag{2}$$

We determine the probabilities q_{xy}^t for all grid cells c_{xy} in our stochastic situational awareness map based on all observations p_{xy}^t gathered throughout the scenario. These observations form our understanding of the situation. Section 7 describes how p_{xy}^t and q_{xy}^t are calculated when scouting with an agent.

4.2 The generative process

In the generative process, we calculate the posterior probability of target presence given observations. At each time step, we have an incoming a priori probability $p_{xy}^t(\vartheta)$. This probability is equal to the posterior probability at the previous time step.

We set

$$p_{xy}^t(\vartheta) = p_{xy}^{t-1}(\vartheta \mid \varphi) \tag{3}$$

In addition, we initialize the a priori probability of at least one target (T) in c_{xy} for time 0 according to

$$p_{xy}^0(\vartheta = T) = \epsilon \tag{4}$$

Using Bayes' theorem, we can calculate the posterior probability $p_{xy}^t(\vartheta \mid \varphi)$ in the grid cell c_{xy} for the current time step t based on the a priori probability.

We have

$$p_{xy}^t(\vartheta \mid \varphi) = \frac{p_{xy}^t(\varphi \mid \vartheta)}{p_{xy}^t(\varphi)} p_{xy}^t(\vartheta), \tag{5}$$

where $p_{xy}^t(\varphi \mid \vartheta)$ represents the likelihood, which indicates the detection probability for an observation $\phi = T$, as determined by the sensor model, when we have a target object $\vartheta = T$ in the grid cell c_{xy} at time t. In this context, $p_{xy}^t(\varphi)$ represents the probability of obtaining an observation. The probability can be calculated according to

$$p_{xy}^t(\varphi) = p_{xy}^t(\varphi \mid \vartheta = T) \cdot p_{xy}^t(\vartheta = T) + p_{xy}^t(\varphi \mid \vartheta = F) \cdot p_{xy}^t(\vartheta = F), \tag{6}$$

where $p_{xy}^t(\vartheta = T)$ denotes the a priori probability discussed earlier, and

$$p_{xy}^t(\vartheta = F) = 1 - p_{xy}^t(\vartheta = T) \tag{7}$$

Additionally, $p_{xy}^t(\varphi \mid \vartheta = T)$ represents the likelihood, which is the detection probability of the sensor model, with

$$p_{xy}^t(\varphi \mid \vartheta = F) = 1 - p_{xy}^t(\varphi \mid \vartheta = T) \tag{8}$$

4.3 The variational free energy

Based on the generative model's $q_{xy}^t(\vartheta)$ and the generative process's $p_{xy}^t(\vartheta \mid \varphi)$, we can now calculate the VFE F_{xy}^t for all grid cells c_{xy} at time t. Here, F_{xy}^t is our objective function (1), which we aim to minimize over all c_{xy} within sensor range.

We have

$$F_{xy}^{t} = D_{KL}\left[q_{xy}^{t}(\vartheta) \parallel p_{xy}^{t}(\vartheta|\varphi)\right] - ln\left[p_{xy}^{t}(\varphi)\right] = \left[\sum_{\vartheta \in \{T,F\}} q_{xy}^{t}(\vartheta) \cdot ln\left(\frac{q_{xy}^{t}(\vartheta)}{p_{xy}^{t}(\vartheta|\varphi)}\right)\right] - ln\left[p_{xy}^{t}(\varphi)\right], \tag{9}$$

where D_{KL} is the Kullback-Leibler divergence [41]. The term $-ln\left[p_{xy}^{t}(\varphi)\right]$ represents the degree of surprisal from an observation, $q_{xy}^{t}(\vartheta)$ is the probability according to the generative model, and $p_{xy}^{t}(\vartheta|\varphi)$ is the posterior probability according to the generative process, calculated at each time point based on Bayes' theorem. Finally, $p_{xy}^{t}(\varphi)$ is the probability of an observation as derived in Eq. (6).

In active inference, the action that minimizes F_{xy}^{t} is chosen. This action may involve moving to c_{xy} or in the direction of c_{xy}.

5 Dempster-shafer theory

Calculating a basic belief field requires managing uncertain data that originate from observations—such as questions of general credibility and specifics about time and location—as well as from sensors' detection capabilities.

Our reasoning is grounded in Dempster-Shafer theory [13, 14], which is well suited to this type of problem. Its main advantage is handling and merging uncertain data without relying on unrealistic assumptions. The key is to describe the uncertain data accurately and realistically, ensuring the analysis rests on the most informative foundation.

Dempster-Shafer theory relies on observations supporting a specific hypothesis (positive observations) alongside evidence against it ("negative" observations). Unlike classical probability theory, a basic belief value below 1 does not imply disbelief equal to 1 minus the positive basic belief. We lack knowledge of how to allocate the total basic belief of 1, leaving room for uncertainty. This approach distinguishes between evidence contradicting a hypothesis and a lack of knowledge. Conversely, if a scan of an area yields no indication, we model it as a "negative" observation with a basic belief value less than 1—the remaining basic belief represents uncertainty rather than basic belief in the presence of a target. This way of representing observations—either as a positive observation supporting the presence of a target and leaving the remaining basic belief as uncertainty, or as a "negative" observation contradicting the presence of a target and leaving the rest of the basic belief as uncertainty—is a physically and conceptually sound method for expressing uncertainty in surveillance, where we do not simultaneously obtain basic belief both for and against the presence of a target at the same location based on a single observation. It aligns with our interpretation of uncertain observations from soft-output sensors.

In Dempster-Shafer theory, basic belief is assigned to a proposition through a basic belief assignment, denoted by m. The proposition is represented by a subset A of an exhaustive set of mutually exclusive possibilities, referred to as a frame of discernment, denoted by Θ.

The basic belief function m is defined as a function from the power set of Θ to the interval [0, 1],

$$m : 2^{\Theta} \rightarrow [0,1], \tag{10}$$

where $m(\emptyset) = 0$ and all assigned basic beliefs sum to 1,

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{11}$$

Here, $m(A)$ represents the basic belief assigned to A.

If we receive additional information about the same hypothesis from another source, we combine the two basic belief functions to form a more comprehensive understanding. This is achieved by computing the orthogonal combination using Dempster's rule. Let B be a subset of m_1 and C be a subset of m_2. The combination of m_1 and m_2 results in a new basic belief function $m_1 \oplus m_2$, where

$$m_1 \oplus m_2(A) = K \cdot \sum_{B \bigcap C = A} m_1(B) \cdot m_2(C), \tag{12}$$

where K is a normalization constant

$$K = \left(1 - \sum_{B \bigcap C = \emptyset} m_1(B) \cdot m_2(C)\right)^{-1} \tag{13}$$

This normalization is required because, by definition, no basic belief can be assigned to $\emptyset$. To combine multiple basic belief functions, first combine the initial two, then combine the resulting basic belief with the third, and continue in this manner.

6 World model and sensor model

In this section, and in the next, we use basic beliefs m (for observation) and n (for state).

6.1 World model

The world model is represented as a 2D map of a designated surveillance area. This map is divided into a grid of cells, indexed by a coordinate system. Using a 2D representation is a deliberate simplification, justified by the fact that the scenario under examination is much larger in the x- and y-dimensions than in the z-dimension. The positions of both the agent and any target objects are indicated using cell indices in the format c_{xy}, where x and y are integers. Typically, each cell is large enough to contain the agent or a target object within the surveillance area. However, if a target is larger than a cell, it is assigned to the cell containing its center of gravity.

Fig. 1 shows how the agent's location in the real world relates to its sensor radius and how basic beliefs are derived from sensor data.

The 2D map serves as the foundation for the environment's generative model. The world model includes both stationary and moving targets. Two predefined movement patterns are available for targets: stationary (no movement) or straight-line movement at a constant speed.

We model uncertainty about target presence using Dempster-Shafer theory [13, 14]. Each cell is associated with two evidence-based values [43]. The first is a basic belief, denoted as $n^t_{xy}(\vartheta = T)$, stored in grid cell c_{xy}, which ranges from 0 to 1. It represents the basic belief at time t that at least one target object is in c_{xy}. The second is a basic belief, also between 0 and 1, denoted as $n^t_{xy}(\vartheta = F)$. This indicates the basic belief that, at time t, a target is absent. Note that each basic belief in a cell c_{xy} is an independent distribution with its own frame of discernment. This design choice of using a grid with one

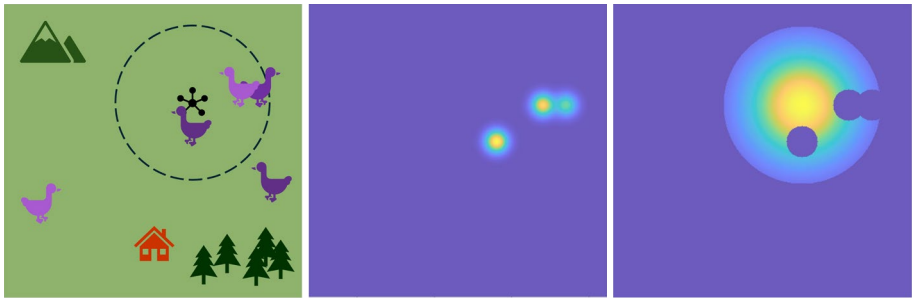

Fig. 1 The left image shows a schematic of the surveillance area in the real world, featuring the UAV agent represented as a star formation and targets shown as purple ducks. The dashed circle represents the sensor range, covering three ducks in the left image. The middle image highlights three positive target detections, with basic belief shown around each. The right image demonstrates a "negative" observation within the sensor radius, excluding the area of positive basic belief around the targets

independent frame of discernment per grid cell, rather than a more complex frame of discernment that incorporates all grid cells in the entire scenario, is to avoid combinatorial explosion. The formulas for diffusing basic belief to neighboring cells in the generative model are designed to be independent of the grid's topology. For example, moving from the square grid topology used in this experiment to a hexagonal topology requires only changing the parameter N, which represents the number of neighboring cells in Eqs. (18) and (19) (see Sect. 7.1). This change reduces N from 8 to 6.

6.2 Sensor characteristics

The agent's sensor is assumed to be an intelligent image sensor, designed for ground scouting. When the agent is active, the sensor takes aerial-view images at predetermined time intervals and uses real-time image-analysis software to process them. The image analysis process is assumed to consist of three main steps:

1. Localization of possible target objects in the images.
2. Classification of the target candidates, where the image processing algorithm returns (for each target candidate): (i) a classification as *Target* (T) or *No Target* (F), and (ii) a value in the [0, 1] interval denoting the classification confidence.
3. (Only for target candidates classified as *Target*.) Estimation of the target candidates' positions. For simplicity, we assume that the image sensor has direct access to the agent's GPS coordinates and its orientation, making it possible for the sensor to compute position estimates in the global coordinate system directly.

Consequently, the output from the intelligent image sensor consists of:

- Estimated positions for all objects classified as targets (if any).
- For each object classified as target (if any), the basic belief that the object in question actually corresponds to a real-world target (i.e., the classification confidence). In this article, we refer to this abstraction as a soft-output sensor model: the intelligent image sensor associates each target detection with a graded confidence value in the interval [0, 1], rather than providing only a hard *Target*/*No Target* decision. This matches modern object detectors, which return localized detections along with class-specific scores or probabilities.

In addition, the sensor is assumed to have the following characteristics:

- The sensor has a limited field of view within which it can reliably detect and classify targets. On ground level, the size of the circular 2D area that the sensor can observe depends both on the sensor's characteristics and on the agent's flight altitude. For fixed flight altitudes, the radius of the observed area can be assumed fixed and is denoted s_r.
- There is a non-negligible risk that the sensor may miss detecting targets present in the observed area.
- The target position estimates, computed by the intelligent image sensor, are somewhat imprecise. They are, however, known to be symmetrically distributed around the true target positions with variance, σ_{pos}^2.

Section 6.3 explains how the sensor model is implemented in the simulation environment. Section 6.4 demonstrates how sensor output is transformed into the input format required by the algorithms that calculate the VFE across the surveilled area.

6.3 Implementation of the sensor model

A simplified mock-up of the intelligent image sensor described in Sect. 6.2 was implemented in the simulation environment and used in the experiments. The mock-up sensor generates output directly from its knowledge of the true target positions in the simulation environment. Its behavior is based on the premises that the real-world image sensor is trained for object classification in aerial images, and that a real sensor's ability to locate and classify potential target objects in an image primarily depends on the distance between the sensor and the object. The most notable simplifications in the sensor implementation are:

- All targets that come within sensor range are detected with probability 1.
- The sensor does not generate any false target detections.
- The target position estimates, generated by the implemented sensor model, are free from noise.

The output from the mock-up sensor is generated as follows:

- Estimated positions of detected targets are replaced by true target positions.
- Instead of classification confidences generated in the target classification process, the mock-up sensor returns *simulated classification confidences* for each detected target $k = 1, \ldots, K$. A Gaussian function is used to model the decline in classification confidence, $C_{T,k}$, as a function of the distance, d_k, between the sensor and a detected target, as shown in Eq. (14). We have.

$$C_{T,k}(d_k) = C_{Tmax} \cdot e^{-\frac{d_k^2}{2\sigma_T^2}}, \tag{14}$$

where σ_T determines the rate of decline in classification confidence (see Fig. 2), while $C_{Tmax} \leq 1$ is a scale factor that sets the maximum achievable value of $C_{T,k}$.

Note that the soft-output sensor avoids Monte Carlo sampling through deterministic soft observations.

In reality, weather conditions and terrain can influence the probability of object detection, classification precision, and classification confidence. However, the dependence of $C_{T,k}$ on these external factors is not modeled in the simulated sensor.

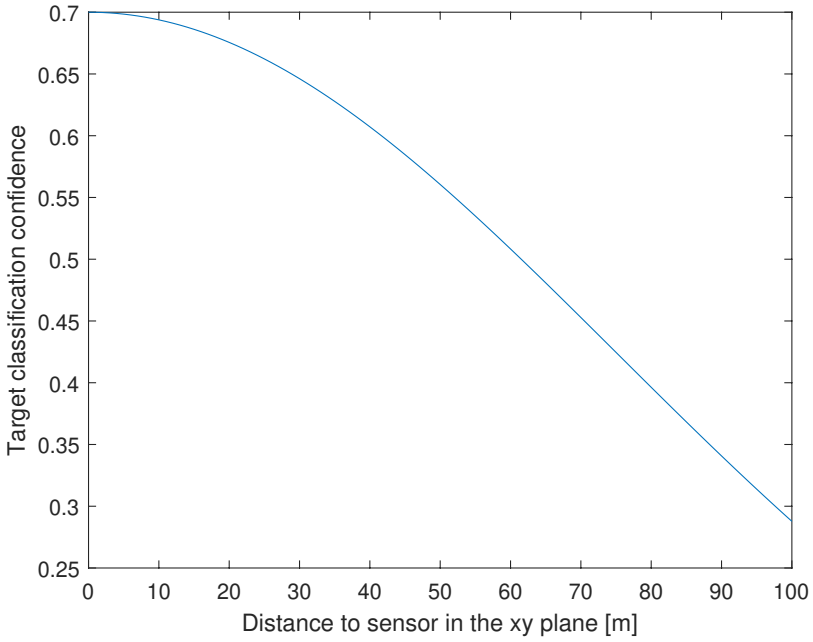

Fig. 2 A Gaussian function models the decline in classification confidence as the distance from the sensor increases. The curve has $C_{Tmax} = 0.7$, $\sigma_T = \frac{3}{4}s_r$ (with $s_r = 100$ m) and corresponds to positive target classification

6.4 Assignment of basic belief from sensor output

In this section, we describe how evidence – such as sensor output – combined with reasonable assumptions is used to assign basic belief values for *Target* (T) and *No Target* (F) to individual grid cells in the simulation area. For this purpose, we introduce the notations $m^t_{xy}(\varphi = T)$ and $m^t_{xy}(\varphi = F)$ to represent the basic beliefs of *Target* and *No Target* (respectively), stored in cell c_{xy} at time t (as in 6.1). Once converted to this representation, the information can be used to update the world model's state (Sect. 7).

To start with, consider the special case where exactly one target ($k = 1$) has been detected by the sensor at time t. The available evidence for *Target* in the simulation area consists of sensor output: the estimated target position coordinates. ($x_k = 1$, $y_k = 1$) and the associated classification confidence $C_{T,\,k\,=\,1}$. Thus, in cell c_{x_1,y_1} we assign $m^t_{x_1,y_1}(\varphi = T) = C_{T,\,k\,=\,1}$. However, as we know that the position estimate is imprecise, we also choose to assign positive (≥ 0) basic belief values for *Target* in the cells surrounding c_{x_1,y_1}. We refer to these basic beliefs, one per cell, as observations (to be consistent with standard free energy nomenclature [3]). The basic belief values in the cells around cell c_{x_1,y_1} are set to match a 2D Gaussian function centered in c_{x_1,y_1} with standard deviation σ_{pos} (given as information associated with the sensor, see Sect. 6.2) and maximum value $C_{T,\,1}$ at c_{x_1,y_1}. The assignment of basic belief values associated with the detection is limited to the cells at a distance $\leq 2\sigma_{pos}$ from c_{x_1,y_1}. All cells outside the $2\sigma_{pos}$ radius, as well as all cells outside the sensor's range, are assigned a basic belief value $m^t_{x_1,y_1}(\varphi = T) = 0$. If there are multiple target detections at the same time, separate basic belief functions $m^t_{xy,k}(\varphi = T)$ are defined for $k = 1, \ldots, K$, equivalently to the case of $k = 1$ described above. With $d_{xy,k}$ denoting the distance between cell c_{xy} and the detected target k, and $d_{xy,s}$ denoting the distance between cell c_{xy} and the sensor, we get basic belief of *Target* k at cell c_{xy}:

$$m^t_{xy,k}(\varphi = T) = \begin{cases} C_{T,k} \cdot e^{-\frac{d^2_{xy,k}}{2\sigma^2_{pos}}}, & (xy) | \, d_{xy,s} \leq s_r, \;\; d_{xy,k} \leq 2\sigma_{pos} \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

The basic belief functions $m^t_{xy,k}(\varphi = T)$ are combined by Dempster's rule [13]. We have,

$$m^t_{xy}(\varphi = T) = \bigoplus_{k=1}^{K} m^t_{xy,k}(\varphi = T) \tag{16}$$

where $\oplus$ is Dempster's rule.

The current sensor model does not provide explicit evidence for *No Target*, but the lack of target observations in a cell within the sensor's range can be regarded as indirect evidence that the cell is empty of targets. We model this by assigning positive (≥ 0) basic belief values for *No Target*, $m^t_{xy}(\varphi = F)$, to all cells located within sensor range *but not inside the* $2\sigma_{pos}$ *radius of any positive target detection*. These basic belief values are modeled by a Gaussian function centered at the sensor's position, with an appropriately chosen standard deviation σ_F and a maximum value, C_{Fmax} (see Fig. 3).

With $d_{xy,k}$ and $d_{xy,s}$ defined as before, we get basic belief on *No Target* at cell c_{xy}:

$$m^t_{xy}(\varphi = F) = \begin{cases} C_{Fmax} \cdot e^{-\frac{d^2_{xy,k}}{2\sigma^2_{pos}}}, & (xy) | \, d_{xy,s} \leq s_r, \;\; d_{xy,k} \geq 2\sigma_{pos} \; \forall \, k \\ 0, & \text{otherwise} \end{cases} \tag{17}$$

7 Controlling an intelligent agent by minimizing VFE

This Section describes our implementation of VFE, in which the generative model is represented as a Dempster-Shafer evidence map with basic beliefs, and its pignistic transform produces the probability distribution employed in VFE minimization.

In Sect. 7.1, we explain the generative model, including its representation and all formulas necessary to update the situational awareness map. This involves fusing sensor observations into the map and diffusing the map's uncertainty to neighboring cells over time. In Sect. 7.2, we outline the generative process and explain how to compute the posterior distribution using Bayes' theorem from a priori, likelihood, and observation distributions. Section 7.3 then integrates these components, illustrating how to combine the pignistic and posterior distributions when calculating VFE. Finally, Sect. 7.4 describes the action rule that the agent uses to assign each UAV swarm member a new common waypoint.

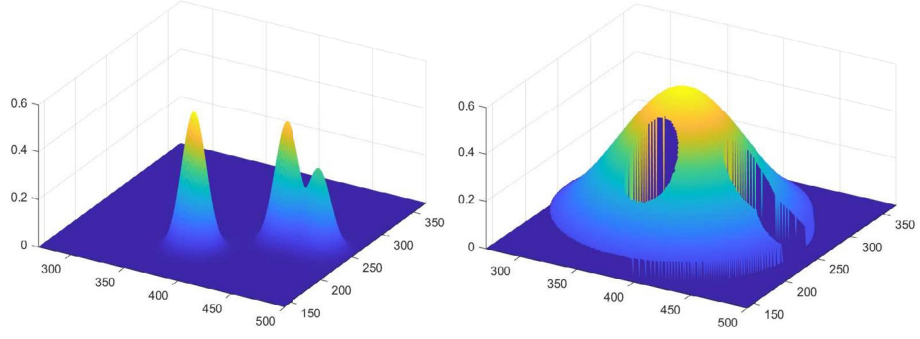

Fig. 3 Combined basic belief assignments $m^t_{xy,k}(\varphi = T)$ for *Target* associated with three observations (on the left), and basic belief assignments $m^t_{xy}(\varphi = F)$ for *No Target* (on the right). For positions within the sensor radius, $m^t_{xy}(\varphi = F) > 0$, other than those near the three targets

7.1 The generative model

This work presents a situational awareness map that uses an evidence map, denoted G, as the world model. The evidence map is structured as a grid of *cells*. Each cell corresponds to a specific segment of the monitored geographical area, as a grid is overlaid on the map. Thus, each cell represents a distinct portion of the surveillance zone. In the generative model, we represent uncertainty using basic beliefs, denoted by m (for observations) and n (for states), as defined in Sect. 6.

Each cell c_{xy} in the evidence map is associated with two basic beliefs: $n^t_{xy}(\vartheta = T)$ and $n^t_{xy}(\vartheta = F)$. Here, $n^t_{xy}(\vartheta = T)$ indicates the basic belief that the cell at time t contains at least one object of interest, while $n^t_{xy}(\vartheta = F)$ represents the basic belief that the cell is empty of targets. The following conditions apply to the basic beliefs: $n^t_{xy}(\vartheta = T), n^t_{xy}(\vartheta = F) \geq 0$, and $n^t_{xy}(\vartheta = T) + n^t_{xy}(\vartheta = F) \leq 1$ for all cells c_{xy} at all times t. The pair $\left(n^t_{xy}(\vartheta = T), n^t_{xy}(\vartheta = F)\right)$ represents the basic belief for cell c_{xy} at time t, with uncertainty defined as $1 - n^t_{xy}(\vartheta = T) - n^t_{xy}(\vartheta = F)$.

In Fig. 4, the green area represents $n^t_{xy}(\vartheta = T)$, the red area represents $n^t_{xy}(\vartheta = F)$, and the white area indicates the uncertainty.

The uncertainty $1 - n^t_{xy}(\vartheta = T) - n^t_{xy}(\vartheta = F)$ can be considered a genuine uncertainty in cell c_{xy}, which the agent should strive to minimize across all cells c_{xy} at all times to maintain a current situational awareness map.

Instead of assigning a single piece of evidence for a positive detection to a single *set* of grid cells within the assumed sensor detection radius, we assign evidence to each grid cell within that radius. This approach offers a more detailed representation, as each grid cell c_{xy} is assigned a value, while significantly reducing the computational complexity of updating these values in response to new sensor measurements.

Our experiment uses a Gaussian distribution with a standard deviation equal to the sensor radius, ensuring that the center cell receives the highest value (Fig. 2). This approach yields a set of distinct basic belief distributions – one for each grid cell. The result provides satisfactory resolution and a relevant spread in the evidence-based situational awareness map. The diffusion model outlined in Eqs. (18) and (19) represents target-motion uncertainty and produces a dynamic evidence-based situational awareness map.

Because the observed environment is dynamic, the value of older information diminishes over time. In the evidence map, this drop in certainty about the locations of already detected objects is shown by the spread of basic belief to adjacent cells. Equation (18) describes this outward diffusion of presence evidence as a random-walk process. Conversely, Eq. (19) constrains the inward propagation of the lack of evidence, signifying that an area is considered empty only if its surrounding cells are also empty.

Regardless of whether the agent observed the state of cell c_{xy} between times t and $t+1$, the basic beliefs $n^t_{xy}(\vartheta = T)$ and $n^t_{xy}(\vartheta = F)$ in the cell are updated according to

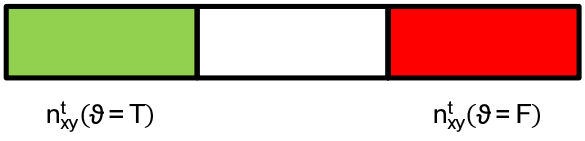

Fig. 4 Illustration of evidence and uncertainty in cell c_{xy}. Green indicates $n^t_{xy}(\vartheta = T)$, red indicates $n^t_{xy}(\vartheta = F)$, and white shows residual uncertainty $1 - n^t_{xy}(\vartheta = T) - n^t_{xy}(\vartheta = F)$; this per-cell Dempster–Shafer pair forms the basis of the evidence map G

Eqs. (18) and (19). This process is separate from the fusion with new observations, which is handled in Eqs. (20) and (21). The update for the basic belief of *Target* is given by:

$$n_{xy}^{t+1}(\vartheta = T) = 1 - \left(1 - \frac{n_{xy}^{t}(\vartheta = T)}{N+1}\right) \cdot \prod_{(ij) \in \{neighbors\ of\ c_{xy}\}} \left(1 - \frac{n_{ij}^{t}(\vartheta = T)}{N+1}\right) \tag{18}$$

where N is the number of neighbors of the cell c_{xy}. This means that the basic belief $n_{xy}^{t}(\vartheta = T)$ of cell c_{xy} is shared among its neighbors and itself (i.e., among nine grid cells in the case of a square grid) and then combined using Dempster's rule. This causes a diffusion of basic belief $n_{xy}^{t}(\vartheta = T)$, over an increasingly larger area over time (in a manner similar to an explosion).

Suppose $n_{xy}^{t}(\vartheta = T) > 0$ in cell c_{xy}. Then some of the basic belief $n_{xy}^{t}(\vartheta = T)$ will be propagated to its N neighbors. Conversely, if $n_{ij}^{t}(\vartheta = T) > 0$ in adjacent cells, some basic belief will be propagated into cell c_{xy}.

Furthermore, for *No Target*, we have basic belief

$$n_{xy}^{t+1}(\vartheta = F) = n_{xy}^{t}{}^{\left(\frac{1}{N+1}\right)}(\vartheta = F) \cdot \prod_{(ij) \in \{neighbors\ of\ c_{xy}\}} n_{ij}^{t}{}^{\left(\frac{1}{N+1}\right)}(\vartheta = F) \tag{19}$$

where $n_{xy}^{t+1}(\vartheta = F)$ indicates that c_{xy} is empty of targets at time $t+1$. This suggests that if cell c_{xy} will be empty at the next time step (i.e., $n_{xy}^{t+1}(\vartheta = F) = 1$), then both its neighboring cells and the cell itself must currently be empty at time t, meaning $n_{xy}^{t}(\vartheta = F) = 1$ for all c_{xy} and all neighbors of c_{xy}. Equation (19) contracts basic belief $n_{xy}^{t}(\vartheta = F)$ inward (in a manner similar to an implosion). Our empirical findings suggest that this leads to stable uncertainty dynamics.

Sensor measurements continuously update the evidence map according to Eqs. (20) and (21) below.

If we obtain positive observations, the basic belief $m_{xy}^{t+1}(\varphi = T)$ assigned by (16) is combined with the existing basic belief pair $\left(n_{xy}^{t+1}(\vartheta = T), n_{xy}^{t+1}(\vartheta = F)\right)$, derived by Eqs. (18) and (19) using Dempster's rule [13]. Note that $m_{xy}^{t+1}(\varphi = T) = 0$ for c_{xy} outside $2\sigma_{pos}$ of all target positions.

We have basic belief of *Target* and *No Target*, respectively,

$$n_{xy}^{t+1^*}(\vartheta = T) = \frac{n_{xy}^{t+1}(\vartheta = T) + m_{xy}^{t+1}(\varphi = T) \cdot \left(1 - n_{xy}^{t+1}(\vartheta = T) - n_{xy}^{t+1}(\vartheta = F)\right)}{1 - m_{xy}^{t+1}(\varphi = T) \cdot n_{xy}^{t+1}(\vartheta = F)} \tag{20}$$

and

$$n_{xy}^{t+1^*}(\vartheta = F) = \frac{\left(1 - m_{xy}^{t+1}(\varphi = T)\right) \cdot n_{xy}^{t+1}(\vartheta = F)}{1 - m_{xy}^{t+1}(\varphi = T) \cdot n_{xy}^{t+1}(\vartheta = F)} \tag{21}$$

It is assumed that Eqs. (20) and (21) are always applied after Eqs. (18) and (19), at time step $t+1$. The basic beliefs $n_{xy}^{t+1}(\vartheta = T)$ and $n_{xy}^{t+1}(\vartheta = F)$ on the right-hand side of Eqs. (20) and (21) represent the results from the diffusion update at time step $t+1$. In

contrast, the values $n_{xy}^{t+1*}(\vartheta = T)$ and $n_{xy}^{t+1*}(\vartheta = F)$ on the left-hand side, indicated with an asterisk $(_*)$, are the combined results from the two updates at the same time step $t+1$.

We receive "negative" observations in grid cells c_{xy} within the sensor radius but outside $2\sigma_{pos}$ of all target positions, and assign basic belief $m_{xy}^{t+1}(\varphi = F)$ that is combined with $\left(n_{xy}^{t+1}(\vartheta = T), n_{xy}^{t+1}(\vartheta = F)\right)$ using Dempster's rule.

We have basic belief of *Target* and *No Target*, respectively,

$$n_{xy}^{t+1^*}(\vartheta = T) = \frac{n_{xy}^{t+1}(\vartheta = T) \cdot \left(1 - m_{xy}^{t+1}(\varphi = F)\right)}{1 - n_{xy}^{t+1}(\vartheta = T) \cdot m_{xy}^{t+1}(\varphi = F)} \tag{22}$$

and

$$n_{xy}^{t+1^*}(\vartheta = F) = \frac{n_{xy}^{t+1}(\vartheta = F) + m_{xy}^{t+1}(\varphi = F) \cdot \left(1 - n_{xy}^{t+1}(\vartheta = T) - n_{xy}^{t+1}(\vartheta = F)\right)}{1 - n_{xy}^{t+1}(\vartheta = T) \cdot m_{xy}^{t+1}(\varphi = F)} \tag{23}$$

Refer to Fig. 5 for snapshots of the evidence distributions $n_{xy}^{t+1*}(\vartheta = T)$ and $n_{xy}^{t+1*}(\vartheta = F)$ after calculating Eqs. (20)–(23).

7.2 The generative process

In the generative process, we use a Bayesian approach with probabilities denoted by p. When we receive a new observation, our goal is to calculate the posterior distribution, $p_{xy}^{t+1}(\vartheta|\varphi)$. We must determine this probability distribution across all grid cells, c_{xy}, to achieve this. This requires a sensor model in the form of a likelihood distribution $p_{xy}^{t+1}(\varphi|\vartheta)$ and an observation model $p_{xy}^{t+1}(\varphi)$, as well as an a priori distribution $p_{xy}^{t+1}(\vartheta)$.

We begin with the a priori distribution. It is important to note that the a priori distribution at each time step equals the posterior distribution from the previous time point.

We have a priori probability

$$p_{xy}^{t+1}(\vartheta) = p_{xy}^{t}(\vartheta|\varphi) \tag{24}$$

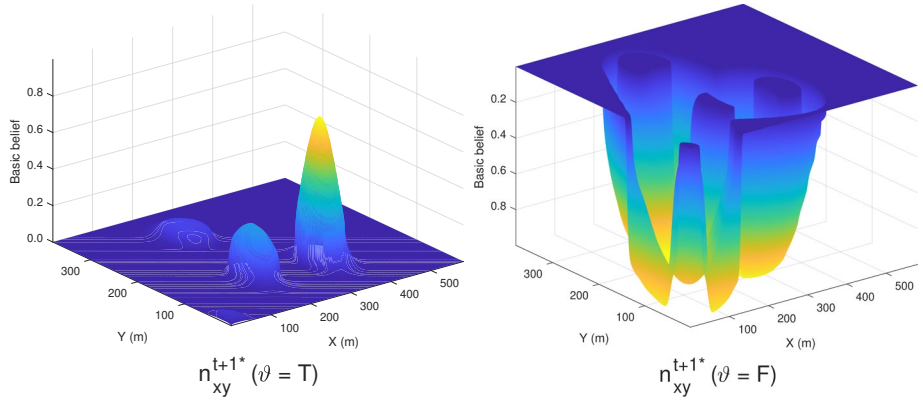

Fig. 5 The figure illustrates the evidence distribution $n_{xy}^{t+1^*}(\vartheta = T)$ in each cell c_{xy} over the surface immediately after detecting a fourth target object in a simple ground scenario (on the left). At time step $t+1$, the evidence distribution $n_{xy}^{t+1^*}(\vartheta = F)$ is shown (on the right). For illustration, the distribution $n_{xy}^{t+1^*}(\vartheta = F)$ is shown upside down in the figure

Initially, for $t = 0$, we have set

$$p_{xy}^{0}(\vartheta = T) = \epsilon = \frac{N_{targets}}{N_{cells}} \tag{25}$$

where $N_{targets}$ are the assumed number of targets in the area under investigation, and N_{cells} is the number of cells in the area.

Using a likelihood distribution and our sensor model, we assign probabilities to each grid cell c_{xy} within the sensor's radius, based on two Gaussian distributions.

To calculate the probability of an observation, denoted as $p_{xy}^{t}(\varphi)$, we can express it as follows:

$$p_{xy}^{t+1}(\varphi) = p_{xy}^{t+1}(\varphi|\vartheta = T) \cdot p_{xy}^{t+1}(\vartheta = T) + p_{xy}^{t+1}(\varphi|\vartheta = F) \cdot p_{xy}^{t+1}(\vartheta = F), \tag{26}$$

where $p_{xy}^{t+1}(\vartheta = T)$ is the a priori probability derived in (24) and

$$p_{xy}^{t+1}(\vartheta = F) = 1 - p_{xy}^{t+1}(\vartheta = T) \tag{27}$$

Additionally, the likelihood distribution $p_{xy}^{t}(\varphi|\vartheta = T)$ is calculated using a Gaussian distribution, similarly to the method described in Sect. 6 and Eq. (14).

Since the sensor returns a deterministic score, we define the likelihood directly from it. Hence, the identification in (28) is merely shorthand rather than an inference about the state,

$$\begin{aligned} p_{xy}^{t+1}(\varphi|\vartheta = T) &\triangleq n_{xy}^{t+1}(\vartheta = T), \\ p_{xy}^{t+1}(\varphi|\vartheta = F) &= 1 - p_{xy}^{t+1}(\varphi|\vartheta = T) \end{aligned} \tag{28}$$

Using the three probability distributions (a priori $p_{xy}^{t+1}(\vartheta)$, likelihood $p_{xy}^{t+1}(\varphi|\vartheta)$, and observation $p_{xy}^{t+1}(\varphi)$), we can calculate the desired posterior distribution using Bayes' theorem.

We have posterior probability

$$p_{xy}^{t+1}(\vartheta|\varphi) = \frac{p_{xy}^{t+1}(\varphi|\vartheta)}{p_{xy}^{t+1}(\varphi)} p_{xy}^{t+1}(\vartheta) \tag{29}$$

Figure 6 displays the posterior distribution alongside Fig. 5.

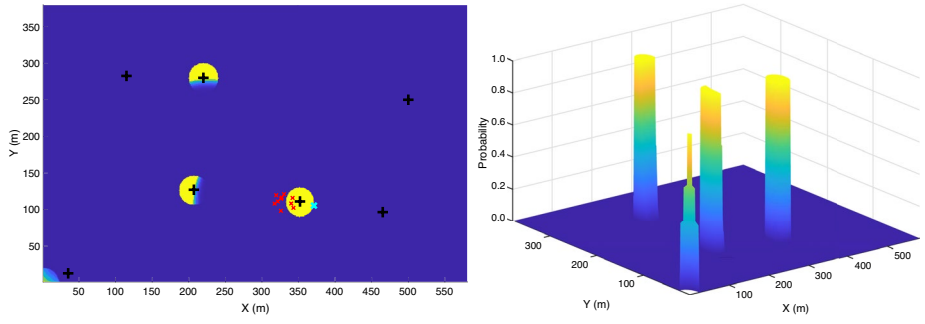

Fig. 6 The posterior distribution $p_{xy}^{t+1}(\vartheta|\varphi)$, calculated after the most recent observation, using Bayes' rule, Eq. (29), the prior, Eqs. (24)–(27), and the Gaussian sensor likelihood, Eq. (28), at the same time step as Fig. 5, for comparison with the fused basic belief shown there

7.3 The variational free energy

The basic belief distribution $n_{xy}^{t+1*}(\vartheta = T)$ represents the distribution from the generative model that we aim to compare with the posterior probability distribution derived from the current observation $p_{xy}^{t+1}(\vartheta|\varphi)$. Notably, $n_{xy}^{t+1*}(\vartheta = T)$ encompasses values over all subsets of the frame of discernment Θ, including Θ itself (where $|2^{\Theta}| = 3$), while $p_{xy}^{t}(\vartheta|\varphi)$ is limited to values associated with the elements of Θ (where $|\Theta| = 2$). Therefore, we must convert the basic belief distribution into a probability distribution to enable comparison between the two distributions. This conversion is achieved through a pignistic transformation [42], defined as follows:

$$BetP(\omega) = \sum_{\omega \in X} \frac{m(X)}{|X|} \tag{30}$$

where $\omega \in \Theta$ and $X \subseteq \Theta$.

We have pignistic probabilities

$$\begin{aligned} q_{xy}^{t+1^*}(\vartheta = T) &= \frac{n_{xy}^{t+1^*}(\vartheta = T)}{|\{T\}|} + \frac{n_{xy}^{t+1^*}(\vartheta = \Theta)}{|\Theta|} \\ &= n_{xy}^{t+1^*}(\vartheta = T) + \frac{1 - n_{xy}^{t+1^*}(\vartheta = T) - n_{xy}^{t+1^*}(\vartheta = F)}{2} \\ &= \frac{1}{2}\left[1 + n_{xy}^{t+1^*}(\vartheta = T) - n_{xy}^{t+1^*}(\vartheta = F)\right] \end{aligned} \tag{31}$$

and

$$\begin{aligned} q_{xy}^{t+1^*}(\vartheta = F) &= \frac{n_{xy}^{t+1^*}(\vartheta = F)}{|\{F\}|} + \frac{n_{xy}^{t+1^*}(\vartheta = \Theta)}{|\Theta|} \\ &= n_{xy}^{t+1^*}(\vartheta = F) + \frac{1 - n_{xy}^{t+1^*}(\vartheta = T) - n_{xy}^{t+1^*}(\vartheta = F)}{2} \\ &= \frac{1}{2}\left[1 + n_{xy}^{t+1^*}(\vartheta = F) - n_{xy}^{t+1^*}(\vartheta = T)\right], \end{aligned} \tag{32}$$

where $q_{xy}^{t+1^*}(\vartheta)$ represents a pignistic probability derived from the pignistic transformation of the basic belief distribution $n_{xy}^{t+1*}(\vartheta)$. In Eqs. (31) and (32), we have $|\{T\}| = |\{F\}| = 1$ and $|\Theta| = 2$. Figure 7 illustrates the transformed pignistic probability distribution of the evidence distribution shown in Fig. 5.

We can now calculate the VFE of each grid cell c_{xy} within the sensor radius by measuring the divergence between $q_{xy}^{t+1^*}(\vartheta)$ and $p_{xy}^{t+1}(\vartheta|\varphi)$, along with the degree of surprisal represented by $p_{xy}^{t+1}(\varphi)$.

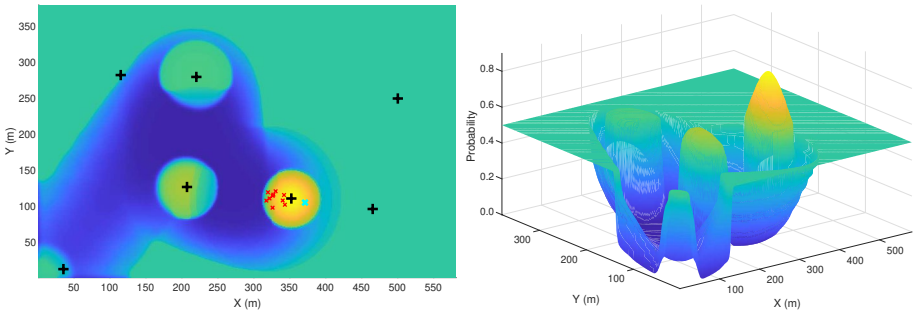

Fig. 7 A pignistic probability distribution $q_{xy}^{t+1^*}(\vartheta)$

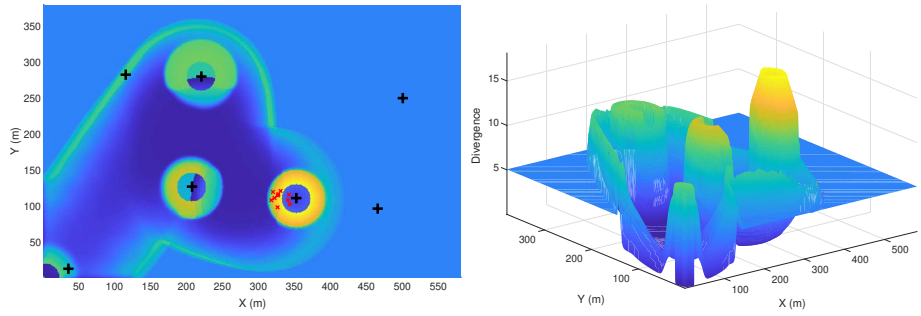

Fig. 8 The divergence between the model and the actual observation. Cell-wise $D_{KL}\left[q_{xy}^{t+1^*}(\vartheta) \parallel p_{xy}^{t+1}(\vartheta|\varphi)\right]$, which is the first term of Eq. (33) within the sensor footprint ($r = 100$ m), illustrating where the model's basic belief and the posterior disagree

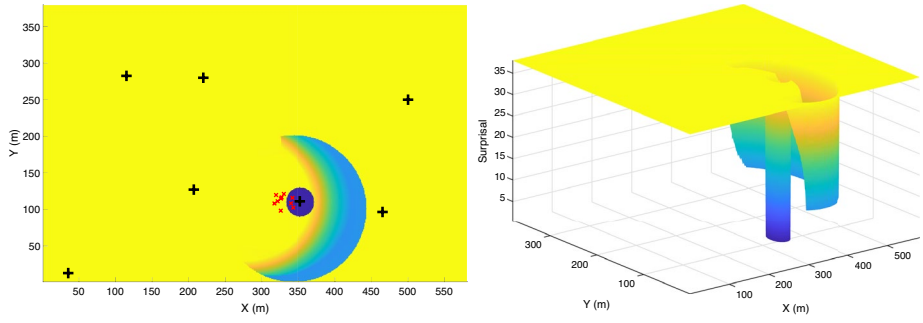

Fig. 9 The level of surprise. The surprisal $-ln\left[p_{xy}^{t+1}(\varphi)\right]$, which represents the second term of Eq. (33), is shown inside the sensor footprint. Outside the footprint, where $p(\varphi) = 0$, surprisal is infinite

Figure 8 illustrates the divergence between the pignistic probability distribution $q_{xy}^{t+1^*}(\vartheta)$ from the generative model and the posterior distribution $p_{xy}^{t+1}(\vartheta|\varphi)$ from the generative process.

Figure 9 displays the degree of surprisal, indicated as $-ln\left[p_{xy}^{t+1}(\varphi)\right]$.

We have VFE in cell c_{xy}

$$\begin{aligned} F_{xy} &= D_{KL}\left[q_{xy}^{t+1^*}(\vartheta) \parallel p_{xy}^{t+1}(\vartheta|\varphi)\right] - ln\left[p_{xy}^{t+1}(\varphi)\right] \\ &= \left[\sum_{\vartheta \in \{T,F\}} q_{xy}^{t+1^*}(\vartheta) \cdot ln\left(\frac{q_{xy}^{t+1^*}(\vartheta)}{p_{xy}^{t+1}(\vartheta|\varphi)}\right)\right] - ln\left[p_{xy}^{t+1}(\varphi)\right], \end{aligned} \tag{33}$$

where D_{KL} is the Kullback-Leibler divergence [41].

The VFE F_{xy} is calculated for all grid cells within the sensor radius, then minimized across them. Finally, the VFE is shown in Fig. 10.

7.4 Action selection

The agent selects an action a upon arriving at its current waypoint at location $c_{x_1y_1}$. It then advances toward a new waypoint at grid cell $c_{x_2y_2}$, which is placed in the direction of $c_{x_0y_0}$ that minimizes F_{xy} at a fixed distance equal to half the sensor radius. If $c_{x_0y_0} = c_{x_1y_1}$, the direction is undefined and the waypoint remains at the current position, i.e., $c_{x_2y_2} = c_{x_1y_1}$. We have action for an agent in cell $c_{x_1y_1}$

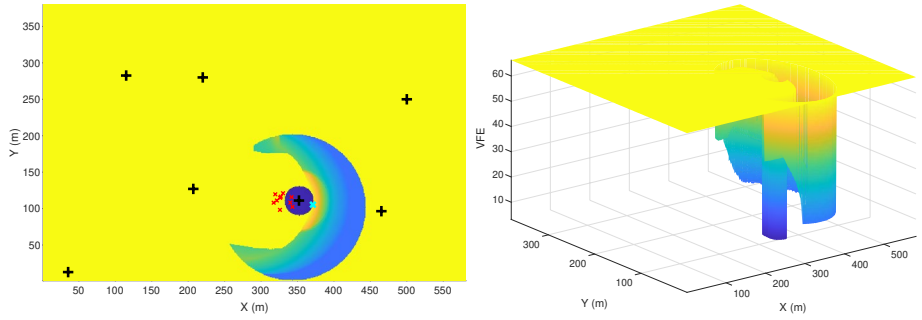

Fig. 10 VFE, denoted as F_{xy}, for all grid cells c_{xy}. F_{xy} is calculated using Eq. (33)

$$a\left(agent.c_{x_1y_1}\middle|\begin{array}{l}(x_0,y_0)=argmin_{xy}\,F_{xy}\\(x_2,y_2)=(x_1,y_1)+(x_0-x_1,y_0-y_1)\cdot\frac{s_r}{2\sqrt{(x_0-x_1)^2+(y_0-y_1)^2}}\end{array}\right)\rightarrow agent.c_{x_2y_2}, \tag{34}$$

where $agent.c_{x_1y_1}$ is the current position of the agent and $agent.c_{x_2y_2}$ is the updated position. The leader assigns this waypoint to the swarm, and the low-level controller moves the UAVs incrementally toward it. Consequently, employing active inference, we gradually move the agent toward grid cell $c_{x_2y_2}$, thereby moving in the direction that minimizes F_{xy}.

8 Implementation for an intelligent agent

Implementing agent control using the free energy principle and active inference requires performing calculations, as outlined in Sect. 7. We conduct UAV control experiments using active inference in MATLAB. However, several essential implementation details and design choices must be considered.

The generative model and process are evaluated at each simulation time step. Importantly, the VFE is calculated only for positions within the sensor radius; in this simulation, we evaluate sensor radii from 10 to 100 m. This limitation is not merely a design choice but a fundamental aspect of the model. According to Eq. (33), the VFE is defined as the sum of the Kullback-Leibler divergence between two probability distributions, $D_{KL}\left[q_{xy}^{t+1^*}(\vartheta)\parallel p_{xy}^{t+1}(\vartheta|\varphi)\right]$, and the surprisal, which is calculated as $-ln\left[p_{xy}^{t+1}(\varphi)\right]$. Here, $p_{xy}^{t+1}(\varphi)$ represents the probability of the observation. Outside the sensor radius, $p_{xy}^{t+1}(\varphi)$ equals zero, causing $-ln\left[p_{xy}^{t+1}(\varphi)\right]$ to approach infinity. This observation highlights that surprisal becomes infinite when faced with the unobserved. Therefore, we restrict the VFE assessment to within the sensor's radius. Outside this boundary, the likelihood is undefined in the absence of observations, so we do not compute VFE in those areas.

Another design choice concerns the distance to the next waypoint. Once the position with the lowest VFE is identified (Fig. 11), we calculate the direction from the agent's current position to that position and place the next waypoint $0.5s_r$ along that direction (i.e., 50 m when s_r = 100 m). The value $0.5s_r$ was chosen as a heuristic compromise: it keeps the waypoint within the currently evaluated sensor footprint, ensures sufficient overlap between consecutive footprints, and still allows meaningful spatial progress before replanning. The agent then takes incremental motion-control steps toward the waypoint. At each simulation step, the agent updates its observations and the situational awareness map. The agent designates a new waypoint once it reaches the current one.

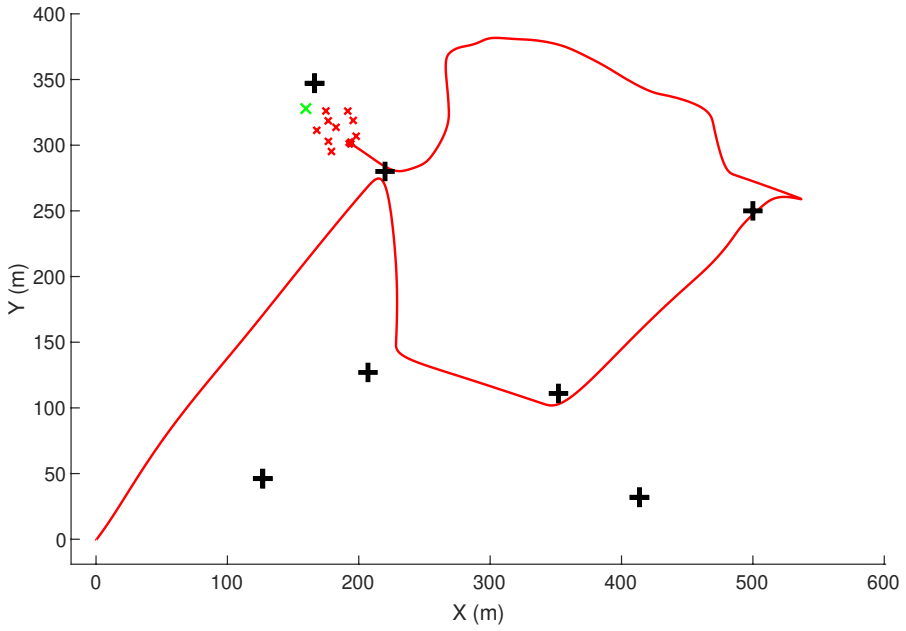

Fig. 11 The trajectory, waypoints, and the position of the minimum VFE. Red crosses represent the UAVs. The leader (i.e., the agent with the red trajectory) selects a new waypoint along a vector pointing toward the minimum VFE (indicated by the green cross). The black crosses indicate the exact locations of the targets (viewed from a global perspective), which the agent does not know directly except through sensor observations; the green "×" marks *argmin* F_{xy}, which the agent uses to set the next waypoint

The factor 0.5 functions as a controller parameter rather than a result of the VFE formulation and was not optimized in this study. Therefore, the results reported are specific to this setting. Smaller fractions would lead to more frequent, localized waypoint updates, while larger fractions would reduce footprint overlap and responsiveness to changes in the VFE field.

The agent collects sensor observations and updates the probabilistic map G every 0.2 s. When the VFE-based swarm-control strategy takes effect, the minimum free-energy point is determined. A waypoint is then set at 50% of the sensor radius from the agent's position, oriented toward the minimum free-energy point. Upon reaching the waypoint, the process repeats: a new minimum free-energy point is identified, and the cycle continues until the scenario concludes. The same step length applies to the VFE, divergence-only, and surprisal-only controllers. The random and fixed-track strategies follow their respective predefined motion rules. The waypoint, set by the agent, is shared by all UAVs.

The physical control of ten UAVs is simulated using an in-house Multi-Agent Dynamic Simulator (MADS). However, one of the UAVs (the agent) serves as the leader and performs all planning. During flight, MADS manages the low-level motion control of the individual UAVs using a spring-damper system to spread them out and avoid collisions.

The computational complexity per step of our planner is

$$O\left(N_{UAV}^2 + N_{targets}N_{cells} + s_r^2\right) \tag{35}$$

where N_{UAV} is the number of UAVs in the swarm, $N_{targets}$ is the number of targets, N_{cells} is the number of grid cells, and s_r is the sensing radius expressed in grid cells. The term $O\left(N_{UAV}^2\right)$ arises primarily from pairwise calculations in the low-level swarm simulation.

The term $N_{targets}$ is a conservative upper bound on the number of detected targets by the sensor. The term $O\left(N_{targets}N_{cells}\right)$ represents the grid evidence and posterior updates, together with full map operations performed independently of the number of targets. Given the current evidence and posterior maps, the VFE values are calculated for the candidate cells within the sensor footprint and the minimum is selected. Because the number of candidate cells grows with the area of the sensor footprint, the complexity of the underlying VFE evaluation and minimization is therefore $O\left(s_r^2\right)$.

Not all operations occur at every simulation tick. Diffusion and fusion of new observations occur at each tick, while VFE minimization occurs when a new waypoint is selected. Equation (35) should therefore be interpreted as a conservative upper bound for a simulation tick in which these updates are performed.

This complexity reflects the current MATLAB/MADS implementation rather than an inherent property of VFE or the proposed sensor model. In particular, the dependence on N_{UAV} stems from the low-level swarm simulation. In a real system, computational cost would depend on the image-processing pipeline, the sensor-output representation, and the map-update implementation.

In practice, for a 580×380 grid with ten UAVs, seven targets, and a sensing radius of 10–100 cells (i.e., 10–100 m), each simulation step takes on average 0.12 s, which amounts to 300 s computation time for a 500 s scenario with five simulation ticks per second (i.e., faster than real time) on a standard laptop (MATLAB; Intel i7; 32 GB RAM).

Since the leader handles sensing and waypoint selection, adding follower UAVs does not provide independent observations within the current architecture. It mainly influences the low-level swarm interaction terms (N_{UAV}). Additionally, larger maps and denser target fields increase the time complexity of the current implementation by up to $O\left(N_{targets}N_{cells}\right)$.

The measured average of 0.12 s per step applies only to the tested configuration of ten UAVs, seven targets, a 580×380 grid, and sensor radii of 10–100 cells. We have not evaluated larger swarms, denser target fields, or larger maps; therefore, we should not extrapolate the runtime or surveillance performance beyond this configuration.

A thorough evaluation of the method requires identifying the most important criteria for the task. Based on these criteria, we must determine whether the approach adequately meets the established goals. This involves generating various random scenarios, analyzing the method's performance, and comparing it with other methods. This process allows us to draw statistically reliable conclusions about the method's effectiveness within a given world model.

9 Results

This section examines the characteristics and performance of the implemented VFE-based motion control strategy for autonomous persistent surveillance. Section 9.1 describes the experimental setup. Section 9.2 provides a qualitative assessment of the control strategy's performance using simulation data. Section 9.3 delivers a quantitative comparison between the VFE-based method and four other control strategies, highlighting key aspects of VFE-based motion control. The control strategies include the two VFE components, divergence and surprisal, analyzed separately to understand their individual impact on control performance. Additionally, two reference strategies are included for benchmarking: a random-walk method and a fixed-track approach, in which the

agent loops a predetermined path designed for efficient area coverage. The five strategies are assessed based on criteria such as target detection, sustained observation, and accurate target localization. Results show that VFE-based motion control surpasses the benchmarks in tracking over time, while the fixed-track method is most effective for target discovery. Of the two VFE components, divergence has a stronger influence on performance than surprisal.

9.1 Experimental setup

In these performance evaluations, we use a 580 m by 380 m scenario map with 1 m by 1 m grid cells for a 2D environment. We provide a thorough quantitative analysis of five control strategies using 1000 randomly generated 500-second scenarios, each updated every 0.2 s with new sensor observations, yielding 2.5 million observations across all simulations. Each scenario includes four fixed targets randomly placed across the map and three moving targets starting at random positions along the edges. These moving targets begin at time 0 and move in straight lines toward randomly assigned final locations on the opposite side of the map. Their speeds are 1, 1.5, and 2 m/s, respectively. The swarm of ten UAVs starts at the lower-left corner of the scenario map at time 0. Initially, it traverses a path passing the four fixed targets in a random order, collecting initial sensor observations for the probabilistic (common operational picture) map G using a simplified sensor with perfect detection, no false alarms, and fixed classification reliability. Once the fourth fixed target is reached, the agent switches to one of the evaluated control strategies, which is then applied for the remainder of the scenario.

The UAVs' average speed across all scenarios is 4.23 m/s. The targets' speeds are intentionally chosen so that some are likely to exit the simulation area before the 500-second scenario concludes. This sometimes causes the loss of observed targets, prompting the swarm to search for new ones. On rare occasions, all three targets might leave the simulation area before the scenario finishes.

9.2 Qualitative analysis

The findings indicate that the technique performs well. Once it reaches the first four targets along the preset waypoints, the agent maintains a balanced approach, alternating between search and track behaviors as it continues to explore the area.

The agent's trajectory, shown as a red line over a 500-second scenario (Fig. 12), illustrates a dynamic balance between pursuing targets marked with black "+" symbols and exploring new areas. Black trajectories represent the three moving targets, while the four stationary targets have no trajectories. The green "×" indicates the minimum VFE at the current time.

We examine how the swarm's behavior shifts over time in a specific scenario. In Fig. 12, upon reaching the fourth fixed target at (420, 180), control strategies are activated for the remainder of the scenario. The UAV swarm detects a northbound target (dashed line) moving at 1.5 m/s and follows it northward. En route, it encounters a slower southbound target (dotted line) moving at 1 m/s and turns south to pursue it, approaching the nearby fourth fixed target for closer inspection. Subsequently, it detects a fast eastbound target (solid line) moving at 2 m/s and follows it eastward across the region. The swarm then reencounters the slow-moving target and accompanies it southward toward the endpoint at (540, 0). When the slow target departs the area, the swarm heads westward

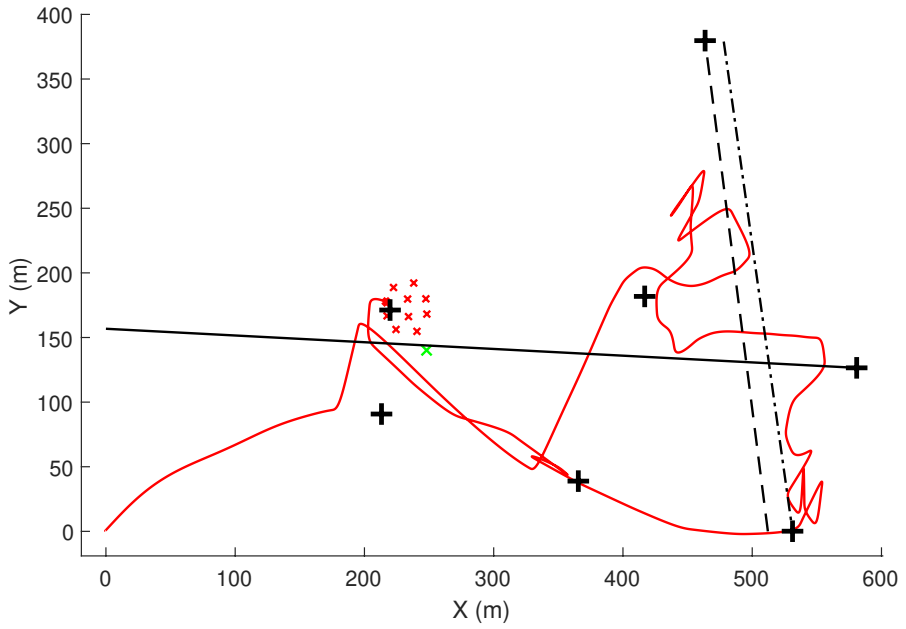

Fig. 12 The agent's path combines surveillance with target tracking. The trajectory, waypoints, and the position corresponding to the minimum VFE are shown. The UAVs, marked with red crosses, allow their leader (the agent) to select a new waypoint along a vector pointing toward the minimum free-energy point, shown as a green cross. Black crosses indicate the targets' true locations from a global perspective, which the agent knows only through sensor observations. The black "+" at the end of the trajectories indicates the target's final position as it leaves the map and simulation

and commences searching, first rediscovering the third fixed target, then the second at (220, 170). This sequence illustrates how the swarm can switch targets based on VFE and smoothly transitions from tracking to searching.

A qualitative analysis of the images using a stochastic basic belief model, along with probabilistic observations and their divergences, provides insights into the agent's level of consistency between its model and observations. This understanding, combined with the concept of VFE, clarifies the reasons and timing behind the agent's decision to scout or pursue a target. When analyzing the sequence of images throughout a scenario, a behavior emerges that is both intuitively understandable and analytically coherent.

9.3 Quantitative analysis

This section employs a simulation-based approach to examine whether active inference can address the motion-control challenge for a UAV swarm, in which a single-sensor leader (the agent) performs multi-objective persistent surveillance (Sect. 2.1).

As mentioned earlier, persistent surveillance can be seen as a multi-objective control problem. In the scenario under consideration, the UAV swarm aims to track all targets within the designated area. Due to sensor limitations, the swarm must alternate between searching for new targets and revisiting known ones to update their locations. To enhance situational awareness, additional measures—such as observing target activity, which often requires longer tracking periods—may be incorporated. Whether employing hybrid control or other motion-control methods, the system should produce appropriate behaviors and maintain a balanced, effective approach.

To evaluate the performance of a persistent surveillance motion control algorithm, simulations can measure how much the swarm's world model diverges from the actual

target locations. However, this score offers limited insight into UAV behavior under the control algorithm. Since this work is exploratory, we have opted not to calculate a final evaluation score. Instead, we focus on experiments that analyze the characteristics of VFE-based control by collecting relevant metrics and simulation statistics, and comparing VFE to benchmarks.

Behaviors that are particularly important to analyze in experiments include:

- Basic belief in target presence at the true target positions.
- Reliable target detection measured by the percentage of scenarios where different numbers of moving targets are detected.
- Even distribution of observation time across moving targets.
- Tracking continuity over a limited consecutive period.

Section 9.3.1 and 9.3.2 describe the experiments conducted to analyze behavior under different control strategies and present their findings. These experiments used the five control algorithms mentioned earlier. Controls based on divergence and surprisal are included to explore how these components influence the overall characteristics of VFE-based control. Additionally, the random walk and fixed-track controls are included for comparison.

9.3.1 Evaluation of basic belief density relative to true target positions

The primary goal of the agent leading the UAV swarm is to keep current estimates of the targets' locations throughout each simulation. Although this might seem like a simple success measure, a closer look reveals that evaluating goal achievement is more complex. Relying on a basic belief map to represent the world has limitations. For instance, it does not distinguish among targets; all observations are treated equally, regardless of which target is observed. As a result, this representation makes it difficult to analyze position estimation errors for individual targets. Moreover, determining the exact number of targets within the surveyed area on the map is challenging with this approach. It is also worth considering whether measuring a target's position estimation error remains meaningful if it exceeds a threshold, after which the estimate is *deemed unreliable*. In practice, it may be more relevant to monitor the extent to which a target position estimate is *sufficiently close* to the actual target location.

Considering the above points, a simplified metric evaluates goal achievement by measuring support for true target positions over time. Instead of computing position estimates and their errors for each target, this metric sums the basic beliefs assigned to map cells within a 10-meter radius of the true position. This approach provides a more robust measure than relying solely on the basic belief at the exact target location. For simplicity, we refer to the area within 10 m of target j at time k as A_j^k. Therefore, the evaluation score is defined as follows:

$$\begin{aligned}\phi \; = \frac{1}{N_{sim}} \sum_{i=1}^{N_{sim}} \frac{1}{N_{moving}} \sum_{j=1}^{N_{moving}} \frac{1}{t_{stop}\,(i,j) - t_{start}\,(i)} \\ \sum_{k=t_{start}(i)}^{t_{stop}(i,j)} \frac{1}{\left|A_j^k \bigcap G\right|} \sum_{(x,y)|c_{xy} \in A_j^k \bigcap G} n_{xy}^k \,(\vartheta \; = T)\,,\end{aligned} \tag{36}$$

where N_{sim} denotes the number of simulations, N_{moving} is the number of moving targets, $t_{start}(i)$ is the time in scenario number i when control strategies take effect for the remainder of the scenario (after having reached the fourth fixed target), $t_{stop}(i, j)$ is the end-time in scenario i when target j leaves the area (or it is 500 if it remains within the simulation area at the end of simulation), A_j^k is the area with a 10 m radius around target j at time k, and $A_j^k \bigcap G$ is the part of A_j^k that is within G. Note that $\left|A_j^k \bigcap G\right| < \left|A_j^k\right|$ when the target is within 10 m of the border of the simulation area.

The proposed metric does not penalize false indications of target presence at locations without targets. This simplification is based on the understanding that the soft-output sensor used in the simulations does not produce false observations. A more comprehensive version of this metric could be developed by setting:

- $\phi^*_{target} = \phi$
- $\phi^*_{\neg\, target} = \phi$, with the last summation in Eq. (36) evaluated over $(x, y)|\, c_{xy} \notin A_j^k \bigcap G$ instead of $(x, y)|\, c_{xy} \in A_j^k \bigcap G$,

and define the evaluation score as

$$\phi^* = \frac{\phi^*_{target}}{\phi^*_{\neg\, target}} \tag{37}$$

A high evaluation score, ϕ^*, suggests that the positive basic belief in the world model centers on the true target locations rather than being evenly distributed across the area. However, this score alone does not tell the whole story. There may still be false target detections that, if given enough basic belief, could cause the agent to take suboptimal actions.

In the simulations, the evaluation score ϕ was calculated for five control strategies: VFE, divergence, surprisal, random walk, and fixed-track. The results are summarized in Table 2; Fig. 13. Clearly, none of the methods performs well with small sensor radii. However, the fixed-track approach, which is designed to scan the entire area for targets, has a small but consistent advantage. Using the fixed-track approach as a benchmark, we observe that for sensor radii of 40 m or more, the VFE-based control produces higher basic belief near true target positions. For smaller sensor radii, the fixed-track method has a slight edge. As the sensor radius increases, VFE's benefit over the benchmark

Table 2 Evaluation score ϕ is the sum of the basic beliefs across all grid cells within 10 m of *any* of the three moving targets

Sensor radius	VFE	Divergence	Surprisal	Random	Fixed track	VFE/FT
10 m	0.0007	0.0007	0.0006	0.0008	0.0015	0.4667
20 m	0.0040	0.0043	0.0044	0.0029	0.0078	0.5128
30 m	0.0186	0.0158	0.0170	0.0066	0.0208	0.8942
40 m	0.0418	0.0381	0.0315	0.0112	0.0376	1.1117
50 m	0.0678	0.0657	0.0399	0.0161	0.0531	1.2768
60 m	0.0984	0.0951	0.0481	0.0213	0.0695	1.4158
70 m	0.1224	0.1259	0.0552	0.0271	0.0859	1.4249
80 m	0.1491	0.1610	0.0626	0.0336	0.0907	1.6439
90 m	0.1765	0.1830	0.0690	0.0404	0.1050	1.6810
100 m	0.1995	0.1827	0.0745	0.0469	0.1179	1.6921

It is calculated for the five control functions and different sensor radii. We observe that VFE outperforms the fixed-track benchmark for sensor radii of 40 m or more

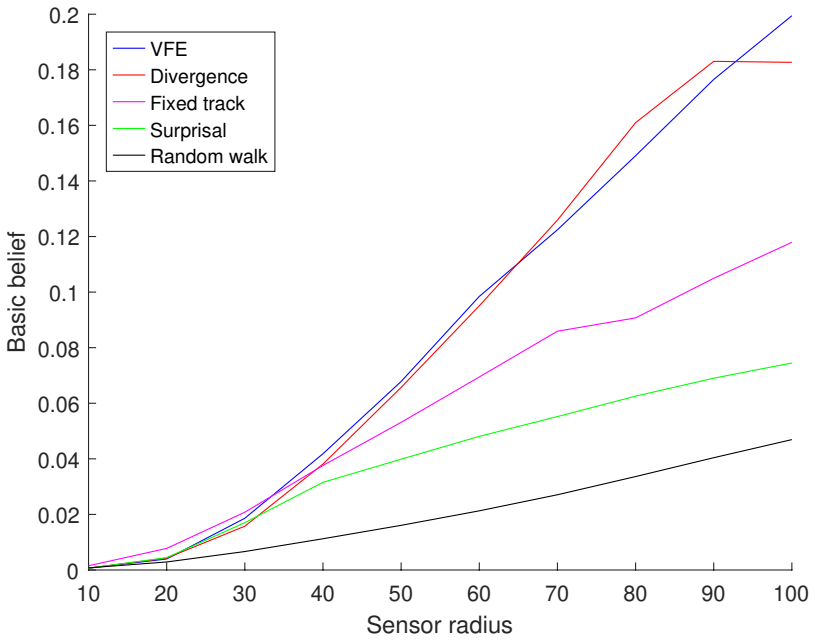

Fig. 13 Evaluation score ϕ, is the sum of the basic beliefs over each second for three moving-target trajectories. The scores are averaged across 1000 scenarios and sensor radii ranging from 10 m to 100 m. The figure compares the VFE control function (blue) with divergence (red), surprisal (green), random (black), and fixed-track (magenta)

grows. This is explained by VFE's greater reliance on longer sensor radii when maneuvering to detect new targets, compared with benchmarks that do not actively maneuver based on sensor input. Specifically, at a 100-meter sensor radius, VFE surpasses the benchmark by 69.2%. The table shows that the divergence component mainly influences VFE's performance, but the gap between VFE and divergence at the 100-meter radius also reflects some contribution from surprisal. Control strategies that rely solely on surprisal or randomness are ineffective.

The scenarios in the simulations vary. Therefore, our evaluation reflects both a comparison of these control strategies and, to some extent, the potential results achievable in such scenarios.

The data in Table 2; Fig. 13 represent averages from 1000 randomly generated scenarios. To demonstrate the variability in outcomes across these scenarios, we use boxplots (Fig. 14) that show the median and the 2nd and 3rd quartiles. The boxes show the interquartile range; the whiskers show the remaining lower and upper ranges according to the boxplot convention used. Notably, high medians for VFE and divergence suggest strong performance, but their large boxes and long upper whiskers also indicate high variance. In contrast, the surprisal component and the random walk demonstrate both low performance and high variance. The fixed-track control has a relatively good median compared to other control strategies, but it is notable for its significantly low variance.

Additionally, based on the basic belief about target presence, we independently evaluate three moving targets. Since these targets only differ in speed and are otherwise unrelated, we assign numbers to them based on the agent's ability to observe each one. In Fig. 15, the target with the highest basic belief of presence in each scenario—defined as the one with the top basic belief across all grid cells within 10 m of its position—is

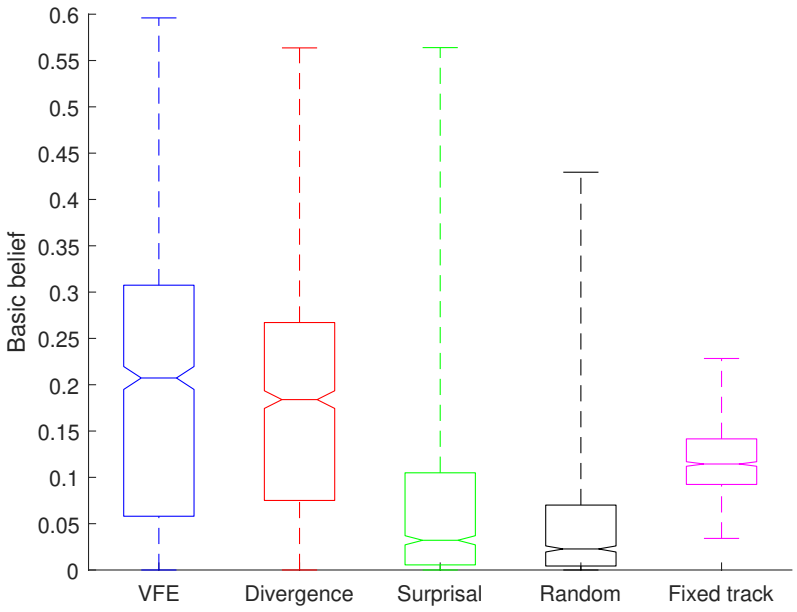

Fig. 14 Boxplot showing the evaluation score ϕ. Results come from 1000 scenarios per control function, with a sensor radius of 100 m. The boxes represent the 2nd and 3rd quartiles across all scenarios

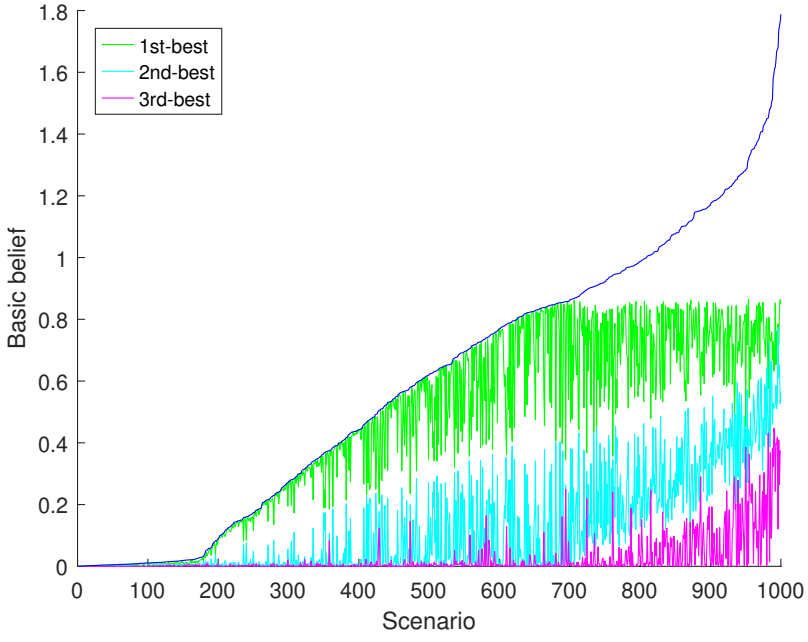

Fig. 15 Basic belief per second for three moving-target trajectories across 1000 scenarios with a sensor radius of 100 m, using the VFE control function. The blue line shows the sum of the basic beliefs (across the three basic belief distributions) for all targets, while the green, cyan, and magenta lines depict the 1st-, 2nd-, and 3rd-best observed moving targets, respectively. These four curves are ordered from lowest to highest based on the overall sum of basic beliefs across all targets (blue). The *x*-axis index corresponds to the same scenario across all four curves

Table 3 The proportion of scenarios in which different numbers of moving targets are detected across 1000 scenarios, using a 100-meter sensor radius, for the five control functions

Number of targets	VFE	Divergence	Surprisal	Random	Fixed track
0	0.017	0.017	0.036	0.051	0
1	0.136	0.131	0.205	0.205	0
2	0.383	0.374	0.404	0.409	0.011
3	0.464	0.478	0.355	0.335	0.989

Table 4 Average observation time in seconds for at least one moving target within sensor range, and for each of the three moving targets across 1000 scenarios, with a sensor radius of 100 m, for the five control functions

Observation in seconds	VFE	Divergence	Surprisal	Random	Fixed track
Any moving target	219.8	199.3	92.1	54.6	145.5
1st best	195.2	170.2	78.5	46.1	75.8
2nd best	48.5	56.1	19.6	13.3	56.0
3rd best	9.4	12.6	3.8	2.4	35.7

These figures may include multiple observation segments and do *not* necessarily indicate continuous observation

shown in green. The second-most-observed target is shown in cyan, and the third in magenta. The 1000 scenarios are arranged by the overall basic belief that at least one target is observed. From this, we see that just over half of VFE's performance comes from detecting the most visible moving target, while less than half results from tracking the second- and third-best observed targets. This illustrates how VFE balances tracking targets already in view with searching for new ones.

9.3.2 Discovery and observation times

We next analyze discovery rates and observation times for different control strategies, specifically for the case with a 100 m sensor radius. We begin by comparing discovery rates across the five strategies.

Table 3 shows that VFE detects all three targets at least once in 46.4% of scenarios. It also detects two targets with a 38.3% chance. A target is considered detected if the sensor is within 100 m of the target's location for at least 1 simulation tick. The table suggests that strategies without active target tracking are more likely to uncover new targets, which explains the high discovery rate observed in the fixed-track benchmark. Moreover, VFE and divergence exhibit a noticeably high rate of detecting two or more targets.

Examining the observation times in Table 4, VFE consistently achieves the longest average observation times across 1000 scenarios, outperforming other control strategies. VFE observes at least one moving target within sensor range for an average of 219.8 s, surpassing alternatives. Across all scenarios, at least one moving target remains in the area for an average of 374.0 s after control activation. Consequently, 58.8% of available time is spent tracking targets, with the remainder allocated to exploring for new ones. Furthermore, VFE is particularly effective at tracking its top target, whereas divergence performs better when tracking its second-best target.

Figure 16 shows the distribution of observation times across five control strategies for any moving target. The wide variance reflects the range of difficulty across scenarios. Boxplots indicate that VFE generally outperforms the other strategies, though it also exhibits the greatest variance. The boxes show the median and interquartile range for

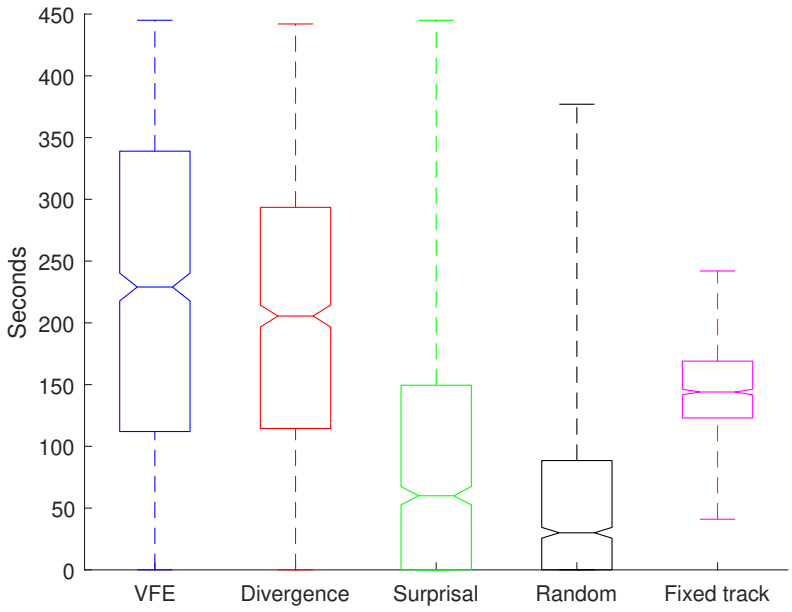

Fig. 16 Boxplot of observation times in seconds for any of the three moving targets across 1000 scenarios, with a 100-meter sensor radius, and across the five control functions. These values may include multiple observation segments and do *not* necessarily indicate continuous observation. The boxes show the 2nd and 3rd quartiles across all scenarios

each control strategy across all scenarios. As noted earlier, VFE outperforms divergence due to its surprisal component. VFE also outperforms the other strategies, primarily due to its divergence component. The random strategy performs poorly roughly 75% of the time, while the fixed-track benchmark's performance mostly falls within the first or second quartiles of VFE.

Finally, we analyze the observation times for the three targets using VFE across all scenarios. In Fig. 17, scenarios are arranged by the total observation time for the three targets. Each scenario highlights the target with the most observations in green with a ragged line, the second-most-observed target in cyan, and the third in magenta. The blue line indicates the total observation time for all three targets. The figure clearly shows that the target with the most observations in each scenario dominates, while the second-most-observed target also plays a significant role. As shown in Table 4, across all scenarios, the most frequently observed target accounts for 77.1% of observations, and the second most frequently observed target accounts for 19.2%. These findings from Fig. 17 align with those from Fig. 15, which analyzed the basic belief of observing each target. In that figure, the most observed target received 78.1% basic belief of target presence, and the second-most-observed target received 18.1%.

The target observation time indicates how well different control strategies perform in tracking, but it may comprise multiple observation segments that add up to the total observation time and may not fully reflect tracking consistency. Examining the average length of the longest continuous observation segment for each of the three targets shows variations in tracking performance among the most, second-most, and least observed targets. As shown in Table 5, the VFE control strategy maintains continuous observation of its top target for an average of 130.7 s, surpassing the other four strategies. Divergence

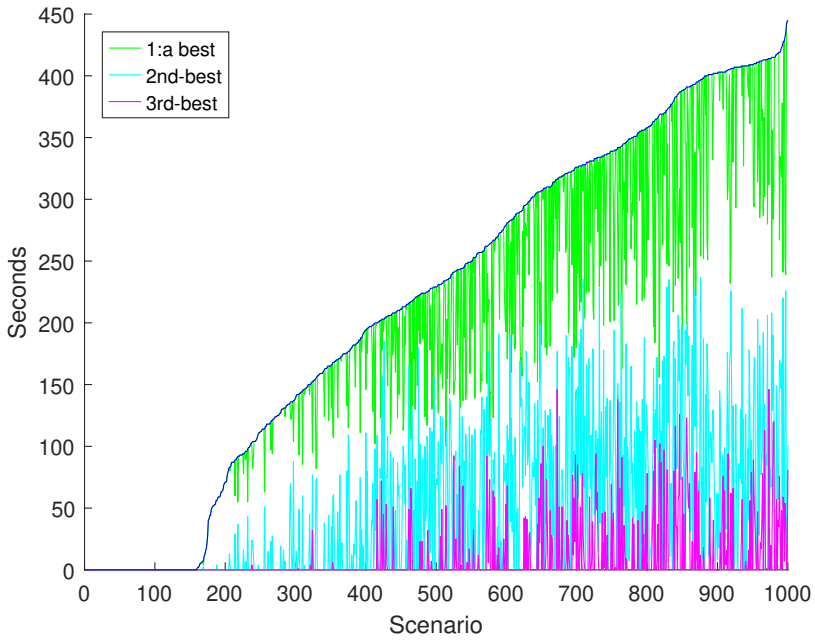

Fig. 17 Observation time in seconds for each of three moving targets across 1000 scenarios, with a sensor radius of 100 m. It compares the VFE control function for all targets (blue) with the top three moving targets ranked by observation time: 1st (green), 2nd (cyan), and 3rd (magenta). All four are sorted from lowest to highest by total seconds observed across all targets (blue). The *x*-axis index matches across all four curves. The numbers may represent multiple observation segments and do *not* necessarily reflect continuous observation

Table 5 Average maximum continuous observation time in seconds for three moving targets across 1000 scenarios, with a sensor radius of 100 m, for the five control functions

Observation in seconds	VFE	Divergence	Surprisal	Random	Fixed track
1st best	130.7	124.0	64.3	37.7	36.6
2nd best	31.1	43.9	15.2	10.8	26.5
3rd best	4.3	9.6	2.6	1.9	17.6

The numbers represent the average duration of the longest *continuous* observation segment across all scenarios

achieves the longest average duration on its second-best target, while the fixed-track approach is most effective at observing its least-tracked target. Comparing Table 5 with Table 4, the longest continuous observation time for the best-tracked target, as a percentage of total observation time, is 67.0% for VFE, 72.9% for divergence, 81.9% for surprisal, 69.0% for random walk, and 48.2% for the fixed-track benchmark across all scenarios. This illustrates how frequently each control strategy revisits each target, with surprisal focused on tracking, fixed-track on searching, and VFE balancing both.

9.3.3 Discussion of quantitative results

The quantitative analysis shows that VFE generally outperforms the fixed-track benchmark at sensor radii of 40 m or greater. Specifically, at a 100-meter radius, VFE exceeds the benchmark by 69.2% in the basic belief of target presence (Table 2) and by 51.1% in observation duration (Table 4). It also excels at observing the top two targets, with continuous observation durations of 161.8 (= 130.7 + 31.1) and 63.1 (= 36.6 + 26.5) seconds for VFE vs. fixed-track, respectively (Table 5). Nevertheless, this high level of continuous tracking entails a trade-off: VFE detects at least two targets in 84.7% of scenarios,

compared with the 100% detection rate of the fixed-track benchmark. The results demonstrate that VFE tracks targets for longer durations than the benchmarks, yet identifies fewer targets in comparison to the fixed-track. A significant portion of VFE's effectiveness is attributed to the divergence component, with surprisal also contributing, as illustrated by the boxplots in Figs. 14 and 16.

10 Conclusions

A simulation of an intelligent agent demonstrates its capacity to autonomously decide when to engage in scouting activities and when to track specific targets. This decision-making process depends entirely on data acquired through the agent's sensor. As the agent operates, it navigates from its current location toward the designated grid cell that minimizes VFE. VFE quantifies the degree of alignment between the agent's perception of reality and its current observations. Minimizing VFE enhances the agent's comprehension and adaptability to its environment. Moreover, the planner operates efficiently: in our tested configuration, each step takes approximately 0.12 s on average, resulting in a total of 300 s of computation time for a 500-second scenario on standard hardware, demonstrating its potential for near-real-time operation at the evaluated scale, though scalability to larger swarms, denser target fields, and larger maps has not been established. A qualitative assessment of the agent's behavior throughout the simulation reveals promising results. Quantitative data collected during the simulations indicate that the agent surpasses benchmarks in target observation. VFE significantly exceeds the fixed-track benchmark, demonstrating a 69.2% increase in the basic belief of target presence and a 51.1% increase in target observation time for a 100-meter sensor radius. The agent's capacity to balance search and tracking behaviors signifies substantial operational autonomy and efficiency. In future work we will extend the approach to EFE by planning over possible agent trajectories (policies) and comparing their performance against computational costs.

Author contributions
Johan Schubert, Farzad Kamrani, and Tove Gustavi contributed equally.

Funding
Open access funding provided by Swedish Defence Research Agency. This work was funded by the Swedish Ministry of Defence and the Swedish Defence Materiel Administration.

Data availability
Data sharing is not applicable to this article, as no datasets were generated or analyzed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 8 June 2026 / Accepted: 1 October 2026

References

1. Parr T, Pezzulo G, Friston KJ. Active inference: The free energy principle in mind, brain, and behavior. Cambridge, MA: The MIT Press; 2022. https://doi.org/10.7551/mitpress/12441.001.0001.
2. Namjoshi SV. Fundamentals of active inference. Cambridge, MA: The MIT Press; 2026.

3. Buckley CL, Kim CS, McGregor S, Seth AK. The free energy principle for action and perception: A mathematical review. J Math Psychol. 2017;81:55–79. https://doi.org/10.1016/j.jmp.2017.09.004.
4. Mishra M, Poddar P, Agrawal R, Chen J, Tokekar P, Sujit PB. Multi-agent deep reinforcement learning for persistent monitoring with sensing, communication, and localization constraints. IEEE Trans Autom Sci Eng. 2025;22:2831–43. https://doi.org/10.1109/TASE.2024.3385412.
5. Friston K. The free-energy principle: a unified brain theory? Nat Rev Neurosci. 2010;11:127–38. https://doi.org/10.1038/nrn2787.
6. Friston K. Life as we know it. J R Soc Interface. 2013;10(86):20130475. https://doi.org/10.1098/rsif.2013.0475.
7. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active inference: A process theory. Neural Comput. 2017;29(1):1–49. https://doi.org/10.1162/NECO_a_00912.
8. Levchuk G, Pattipati K, Fouse A, Serfaty D. (2017). Application of free energy minimization to the design of adaptive multi-agent teams. In Proceedings of SPIE 10206, Disruptive Technologies in Sensors and Sensor Systems, 102060E. https://doi.org/10.1117/12.2263542
9. Albarracin M, Hipólito I, Essafi Tremblay S, Fox JG, René G, Friston K, et al. Designing explainable artificial intelligence with active inference: A framework for transparent introspection and decision-making. In: Buckley CL, editor. Active Inference, IWAI 2023, vol. 1915. Cham: Springer; 2023. p. 123–44. https://doi.org/10.1007/978-3-031-47958-8_9
10. Da Costa L, Lanillos P, Sajid N, Friston K, Khan S. How active inference could help revolutionise robotics. Entropy (Basel). 2022;24(3):361. https://doi.org/10.3390/e24030361.
11. Parr T, Pezzulo G. Understanding, explanation, and active inference. Front Syst Neurosci. 2021;15:772641. https://doi.org/10.3389/fnsys.2021.772641.
12. Sandved-Smith L, Hesp C, Mattout J, Friston K, Lutz A, Ramstead MJD. Towards a computational phenomenology of mental action: modelling meta-awareness and attentional control with deep parametric active inference. Neurosci Conscious. 2021;2021(1):niab018. https://doi.org/10.1093/nc/niab018.
13. Dempster AP. A generalization of Bayesian inference. J Royal Stat Soc Ser B. 1968;30(2):205–47. https://doi.org/10.1111/j.2517-6161.1968.tb00722.x.
14. Shafer G. A mathematical theory of evidence. Princeton, NJ: Princeton University Press; 1976. https://doi.org/10.1515/9780691214696.
15. Hari SKK, Rathinam S, Darbha S, Kalyanam K, Manyam SG, Casbeer D. Optimal UAV route planning for persistent monitoring missions. IEEE Trans Robot. 2020;37(2):550–66. https://doi.org/10.1109/TRO.2020.3032171.
16. Brown A, Anderson D. Trajectory optimization for high-altitude long-endurance UAV maritime radar surveillance. IEEE Trans Aerosp Electron Syst. 2020;56(3):2406–21. https://doi.org/10.1109/TAES.2019.2949384.
17. Hübel N, Hirche S, Gusrialdi A, Hatanaka T, Fujita M, Sawodny O. Coverage control with information decay in dynamic environments. IFAC Proc. 2008;41(2):4180–5. https://doi.org/10.3182/20080706-5-KR-1001.00703.
18. Ramasamy M, Ghose D. (2016). Learning-based preferential surveillance algorithm for persistent surveillance by unmanned aerial vehicles. In Proceedings of the 2016 International Conference on Unmanned Aircraft Systems, pp. 1032–1040. IEEE, Piscataway, NJ. https://doi.org/10.1109/ICUAS.2016.7502678
19. Chen J, Baskaran A, Zhang Z, Tokekar P. (2021). Multi-agent reinforcement learning for visibility-based persistent monitoring. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2563–2570. IEEE, Piscataway, NJ. https://doi.org/10.1109/IROS51168.2021.9635898
20. van de Laar T, Özçelikkale A, Wymeersch H. Application of the free energy principle to estimation and control. IEEE Trans Signal Process. 2021;69:4234–44. https://doi.org/10.1109/TSP.2021.3095711.
21. Bos F, Meera AA, Benders D, Wisse M. (2022). Free energy principle for state and input estimation of a quadcopter flying in wind. In Proceedings of the 2022 International Conference on Robotics and Automation, pp. 5389–5395. IEEE, Piscataway, NJ. https://doi.org/10.1109/ICRA46639.2022.9812415
22. Pezzato C, Ferrari R, Hernández Corbato C. A novel adaptive controller for robot manipulators based on active inference. IEEE Robot Autom Lett. 2020;5(2):2973–80. https://doi.org/10.1109/LRA.2020.2974451.
23. Meo C, Lanillos P. (2021). Multimodal VAE active inference controller. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2693–2699. IEEE, Piscataway, NJ. https://doi.org/10.1109/IROS51168.2021.9636394
24. Meo C, Franzese G, Pezzato C, Spahn M, Lanillos P. Adaptation through prediction: multisensory active inference torque control. IEEE Trans Cogn Dev Syst. 2023;15(1):32–41. https://doi.org/10.1109/TCDS.2022.3156664.
25. Baioumy M, Pezzato C, Ferrari R, Corbato CH, Hawes N. (2021). Fault-tolerant control of robot manipulators with sensory faults using unbiased active inference. In Proceedings of the 2021 European Control Conference, pp. 1119–1125. IEEE, Piscataway, NJ. https://doi.org/10.23919/ECC54610.2021.9654913
26. Çatal O, Verbelen T, Van de Maele T, Dhoedt B, Safron A. Robot navigation as hierarchical active inference. Neural Netw. 2021;142:192–204. https://doi.org/10.1016/j.neunet.2021.05.010.
27. Scholz F, Gumbsch C, Otte S, Butz MV. Inference of affordances and active motor control in simulated agents. Front Neurorobot. 2022;16:881673. https://doi.org/10.3389/fnbot.2022.881673.
28. Kawahara D, Ozeki S, Mizuuchi I. (2022). A curiosity algorithm for robots based on the free energy principle. In Proceedings of the 2022 IEEE/SICE International Symposium on System Integration, pp. 53–59. IEEE, Piscataway, NJ. https://doi.org/10.1109/SII52469.2022.9708819
29. Katahira K, Kunisato Y, Okimura T, Yamashita Y. Retrospective surprise: a computational component for active inference. J Math Psychol. 2020;96:102347. https://doi.org/10.1016/j.jmp.2020.102347.
30. Wakayama S, Ahmed N. (2023). Active Inference for autonomous decision-making with contextual multi-armed bandits. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation, pp. 7916–7922. https://doi.org/10.1109/ICRA48891.2023.10160593
31. Wakayama S, Candela A, Hayne P, Ahmed N. Active inference for bandit-based autonomous robotic exploration with dynamic preferences. IEEE Trans Robot. 2025;41:3841–51. https://doi.org/10.1109/TRO.2025.3577041.
32. Levchuk G, Pattipati K, Fouse A, Serfaty D. (2017). Application of free energy minimization to the design of adaptive multi-agent teams. In Proceedings SPIE 10206, Disruptive Technologies in Sensors and Sensor Systems, 102060E. https://doi.org/10.1117/12.2263542

33. Levchuk G, Fouse A, Pattipati K, Serfaty D, McCormack R. (2018). Active learning and structure adaptation in teams of heterogeneous agents. In Proceedings SPIE 10653, Next-Generation Analyst VI, 1065305. https://doi.org/10.1117/12.2305875
34. Levchuk G, Pattipati K, Serfaty D, Fouse A, McCormack R. Active inference in multiagent systems: Context-driven collaboration and decentralized purpose-driven team adaptation. In: Artificial Intelligence for the Internet of Everything. Cambridge, MA: Academic Press; 2019. p. 67–85. https://doi.org/10.1016/B978-0-12-817636-8.00004-1.
35. Pezzato C, Corbato CH, Bonhof S, Wisse M. Active inference and behavior trees for reactive action planning and execution in robotics. IEEE Trans Robot. 2023;39(2):1050–69. https://doi.org/10.1109/TRO.2022.3226144.
36. Baltieri M, Buckley CL. (2019). Nonmodular architectures of cognitive systems based on active inference. In Proceedings of the 2019 International Joint Conference on Neural Networks, pp. 1–8. https://doi.org/10.1109/IJCNN.2019.8852048
37. Nozari S, Krayani A, Marin P, Marcenaro L, Gomez DM, Regazzoni C. Exploring action-oriented models via active inference for autonomous vehicles. EURASIP J Adv Signal Process. 2024;92. https://doi.org/10.1186/s13634-024-01173-9.
38. Sawada H, Ohata W, Tani J. Human-robot kinaesthetic interactions based on the free-energy principle. IEEE Trans Syst Man Cybern Syst. 2025;55(1):366–79. https://doi.org/10.1109/TSMC.2024.3481251.
39. Matsumura T, Esaki K, Yang S, Yoshimura C, Mizuno H. Active inference with empathy mechanism for socially behaved artificial agents in diverse situations. Artif Life. 2024;30(2):277–97. https://doi.org/10.1162/artl_a_00416.
40. Esaki K, Matsumura T, Minusa S, Shao Y, Yoshimura C, Mizuno H. Dynamical perception-action loop formation with developmental embodiment for hierarchical active inference. In: Buckley CL, editor. Active Inference. IWAI 2023, vol. 1915. Cham: Springer; 2023. p. 14–28. https://doi.org/10.1007/978-3-031-47958-8_2
41. Kullback S, Leibler RA. On information and sufficiency. Ann Math Stat. 1951;22(1):79–86. https://doi.org/10.1214/aoms/1177729694.
42. Smets P, Kennes R. The transferable belief model. Artif Intell. 1994;66(2):191–234. https://doi.org/10.1016/0004-3702(94)90026-4.
43. Morbée M, Tessens L, Aghajan H, Philips W. Dempster-Shafer based multi-view occupancy maps. Electron Lett. 2010;46(5):341–3. https://doi.org/10.1049/el.2010.0105.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.