TRANSPORTED-BELIEF PLANNING FOR MULTI-AGENT SEARCH AND RESCUE UNDER UNCERTAIN DRIFT

Anonymous authors
Paper under double-blind review

ABSTRACT

Maritime search for a drifting survivor requires allocating scarce assets while wind and ocean currents continually reshape where the survivor may be. Each route assignment affects both immediate detection and the team's options for subsequent search. We develop a multi-agent planner that forecasts survivor movement, updates the shared forecast after each non-detection, and coordinates routes by anticipating future search opportunities and teammates' movements. Our distributionally robust formulation derives detection bounds from the mean and covariance of survivor distributions transported by the forecast flow, capturing how routes share uncertainty through survivor drift—Lagrangian transport—rather than through geographic proximity alone. Across 180 simulated search settings based on OSCAR current data from five ocean regions, the planner achieves 50.1% expected mission success, outperforming common benchmark approaches including classical lawnmower search and reinforcement-learning policies. Ablation studies further show that forward planning and teammate modeling both contribute to performance, with gains sustained across the tested fleet sizes, mission scales, and environment conditions. Our work suggests the importance of coordinating current assignments around predicted survivor movement while preserving valuable opportunities for the team's next search.

1 INTRODUCTION

Maritime search and rescue poses a planning problem with a learned model of uncertain drift: how to deploy a scarce fleet when a drifting survivor's location is unknown and the time available for rescue is shrinking. This challenge is further compounded by wind and currents, which reshape the likely search area while assets are en route, so each assignment commits not only current search capacity but also an asset's position for the next decision. A route that appears attractive in isolation may duplicate the survivor movements already covered by teammates, leave other plausible movements unattended, or place an asset far from the team's next valuable search. Moreover, every unsuccessful search changes which survivor movements remain plausible. Effective coordination must therefore balance three linked considerations: immediate detection, differentiated coverage across the fleet, and the future search opportunities preserved by today's assignments.

Geographic dispersion alone does not prevent redundant search: distant routes may encounter the same possible survivor movements, whereas crossing routes may cover different ones (Figure 1). Following those movements—the *Lagrangian* perspective—reveals the additional coverage each assignment contributes. Forecast error creates a second concern: otherwise complementary routes may gain or lose detection value together (Zhang et al., 2017). This resembles portfolio risk, with routes as investments, detection values as returns, and drift scenarios as market conditions. We therefore distinguish complementary coverage within a drift scenario from protection against shared exposure across scenarios.

Operational systems and learned policies already use drifting-target probabilities (Kratzke et al., 2010; Wu et al., 2024); our focus is how successive assignments can anticipate teammates' searches while retaining which survivor paths they already cover. We call the resulting approach *transported-belief planning*, implemented as transported-belief rollout (TBR): it combines shared trajectories with lookahead and teammate coordination. A learned Gaussian-process current forecast (Rasmussen & Williams, 2006) projects possible survivor movements, and unsuccessful searches update their

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

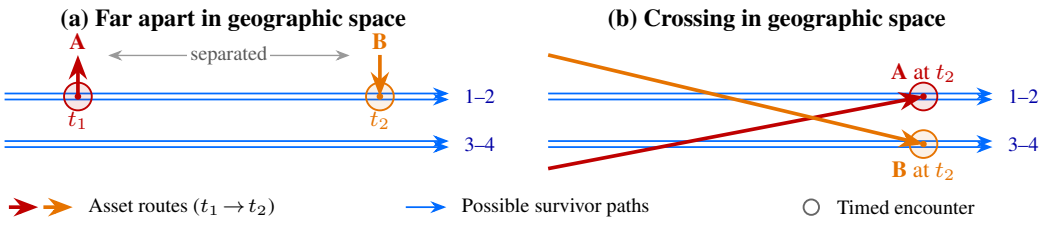


Figure 1: Far apart in space can still mean overlapping search. (a) Separated assets A and B encounter the same survivor paths 1–2 at successive times: high shared-trajectory overlap. (b) Despite crossing routes, A encounters paths 1–2 and B paths 3–4: low overlap and complementary coverage. Rings mark encounters ($t_1 < t_2$). Repeat searches can still help with imperfect sensing.

probabilities. Retaining which trajectories teammates cover lets assignments be valued by the detection they add and the future search options they preserve. A distributionally robust analysis gives a moment-based lower bound on risk-adjusted detection and an additive decomposition of route interactions under stated assumptions.

Across 180 search configurations based on Ocean Surface Current Analysis Real-time (OSCAR) data from five ocean regions, our forward-looking planner achieves 50.1% expected mission success. Given the same trajectory information and action menu, a learned policy reaches 46.5% at the tested training budget. Matched controls show that planning ahead and anticipating teammates' movements both improve success, while discarding which survivor paths teammates cover reduces success by 10.7 percentage points. A mild uncertainty penalty improves success in the worst quarter of wind conditions by 0.82 points at a 0.22-point mean cost. A simpler continuation rule retains success at about a quarter of the planning cost, while tree search on the same belief reaches 51.5%. These alternatives let managers balance complementary coverage, planning effort, and adverse-drift protection.

We make three contributions. We show how a shared transported belief connects immediate detection, complementary team coverage, and future search options; derive moment-based detection lower bounds under stated uncertainty assumptions; and use controlled comparisons to separate the roles of trajectory coverage, lookahead, teammate anticipation, learning, computation, and downside protection. The remainder of the paper reviews the literature (Section 2), develops the model and detection bounds (Section 3), presents the planning framework (Section 4), reports the experiments (Section 5), and concludes (Section 6).

2 Related Work

Search planning under drift. Classical search theory allocates scarce effort according to target-location probabilities and detection value (Koopman, 1957; Stone, 1975). The Search and Rescue Optimal Planning System (SAROPS) transports survivor trajectories, updates them after misses, and optimizes fleet patterns (Kratzke et al., 2010). Probabilistic drift models refine the changing search area (Zhang et al., 2017), while optimization methods address uncertain search paths (Morin et al., 2023) and multi-vehicle coverage (Cho et al., 2022). Our focus is joint team detection on shared paths within multi-step lookahead, tested against map-based coordination using the same transported-belief construction.

Coordinating current and future searches. Assignments must complement teammates' searches and leave assets well placed to continue. Planning under partial observation links actions and evidence to later decisions (Kaelbling et al., 1998); multi-robot search coordinates the team's effort (Chung et al., 2011). Partially Observable Monte Carlo Planning (POMCP) simulates future decisions (Silver & Veness, 2010); multi-agent rollout uses a policy for later actions (Bhattacharya et al., 2021); and Upper Confidence Bounds applied to Trees (UCT) prioritizes promising continuations (Kocsis & Szepesvári, 2006). Shared survivor trajectories make current and follow-up assignments complementary.

Learning to guide search decisions. Learning can provide a drift model or search policy. Maritime applications include weighted coverage (Gao et al., 2019; Ai et al., 2021; Oh et al., 2025), joint route and communication decisions (Lei et al., 2024), drift-informed maps (Wu et al., 2024), and moving-target search (Wang et al., 2022). Other work studies unknown environments (Cao et al., 2024), aerial coordination (Bialas et al., 2026), proximal policy optimization (PPO) and cooperative variants (Schulman et al., 2017; Yu et al., 2022), transformer coordination (Wen et al., 2022), and multi-agent deep deterministic policy gradient (MADDPG) (Lowe et al., 2017). Our reinforcement-learning (RL) comparators include belief-conditioned and local- or joint-observation PPO, MADDPG, and the moving-target search policy MATSMI.

Protection against forecast uncertainty. Routes can fail together under shared forecast error. Distributionally robust optimization controls plausible laws through moments or proximity to a reference (Delage & Ye, 2010; Wiesemann et al., 2014; Esfahani & Kuhn, 2018), while distortion risk emphasizes unfavorable outcomes and admits moment bounds (Shao & Zhang, 2024). For transported survivor beliefs, covariance connects complementarity to the average–adverse-drift tradeoff.

3 Learning a Shared Search Space

At each decision, uncertainty concerns both the drift condition and the survivor path within it. We represent both through drift scenarios containing multiple complete paths. Unlike time-specific probability maps, path identities reveal repeated or complementary coverage within a scenario, while changes in team detection across scenarios reveal shared drift exposure. The same trajectory belief therefore supports fleet coordination and downside analysis.

From current forecasts to survivor trajectories. Assets must search where the survivor may be upon arrival. A two-output Gaussian process (GP) learns both current components jointly, with a squared-exponential spatial kernel and cross-component dependence. Given observed locations and velocities $\mathcal{D}$, let $\mathbf{u}_*$ stack both components at planning-grid locations. The posterior mean vector $\mathbf{m}_{\mathcal{D}}$ gives predicted velocities. The covariance matrix $K_{\mathcal{D}}$ describes their uncertainty and dependence across locations and components. Fitting and conditioning yield the Gaussian posterior

$$\mathbf{u}_* \mid \mathcal{D} \sim \mathcal{N}(\mathbf{m}_{\mathcal{D}}, K_{\mathcal{D}}), \tag{1}$$

Joint field draws preserve this dependence; a spatially uniform random offset adds region-wide drift error. Appendix D gives the GP specification, sampling scheme, and transport recursion.

Figure 2 illustrates survivor transport under sampled currents and wind: particles share the scenario's current field but differ in initial location and unresolved motion. Let Θ denote a generic drift scenario and s index a sampled one. Let Ω be the search domain and T the deadline in ticks. Particle j in scenario s traces positions $x^{s,j} = (X_0^{s,j}, \ldots, X_T^{s,j})$ at ticks $0, \ldots, T$ in the path space $\mathcal{X} = (\Omega \cup \{\dagger\})^{T+1}$; $\dagger$ denotes exit from the domain.

Particle weights $\omega_{s,j}$ start equal and are normalized as $\widetilde{\omega}_{s,j} = \omega_{s,j} / \sum_k \omega_{s,k}$. Let Λ collect the ensemble's distinct paths and $\nu = \sum_{x \in \Lambda} \delta_x$ count them, where δ_x is a unit point mass at x. The scenario belief is the probability-mass function π_s (or π_Θ for a generic scenario):

$$\pi_s(x) = \sum_{j : x^{s,j} = x} \widetilde{\omega}_{s,j}, \qquad x \in \Lambda, \tag{2}$$

For any path-value function f, its scenario average is $\langle f, \pi_s \rangle = \sum_{x \in \Lambda} f(x) \pi_s(x)$. These probabilities guide where further search is worthwhile. A miss shifts path and scenario weights toward movements the fleet was unlikely to detect, redirecting effort without refitting the current model. Exited paths retain probability mass but offer no further search value.

Team coverage and search evidence. Fleet assignments should cover different survivor possibilities while allowing useful repeat searches when sensing is imperfect. Evaluating coverage on shared paths captures both contributions. For N assets, let $\mathcal{F}_t$ contain feasible fleet plans at time t. Plan $p = (p_1, \ldots, p_N) \in \mathcal{F}_t$ specifies asset positions $p_i(\tau)$ at search ticks $\tau \in \mathcal{I}_t(p)$, up to its horizon or deadline T. For path x, let $d_\tau(x, p) = \min_i \|x_\tau - p_i(\tau)\|$ be the distance to the closest

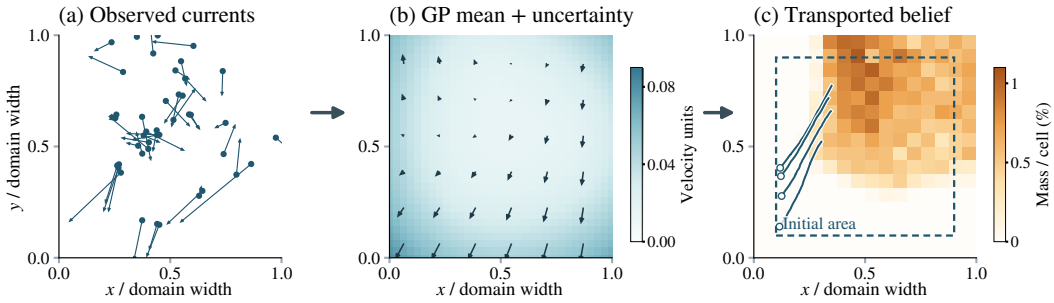


Figure 2: Turning current data into a shared search space. (a) Observed currents inform (b) a GP forecast and its uncertainty. (c) Survivor particles under eight sampled drift scenarios: shading shows their mass at one time, and lines trace selected paths. TBR scores individual paths, not the shaded map; the shading is not a calibrated error bound.

asset, set to infinity after exit. Let $p_{\mathrm{det}}(d)$ be the nonincreasing single-visit detection probability, with $p_{\mathrm{det}}(\infty) = 0$. Conditionally independent sensor outcomes across ticks give

$$\begin{aligned} Y_p(x) &= 1 - \prod_{\tau \in \mathcal{I}_t(p)} \left[1 - p_{\mathrm{det}}(d_\tau(x, p))\right], \\ D_p(\Theta) &= \langle Y_p, \pi_\Theta \rangle, \qquad D_p(s) = \langle Y_p, \pi_s \rangle = \sum_j \widetilde{\omega}_{s,j} Y_p(x^{s,j}). \end{aligned} \tag{3}$$

Here, $Y_p(x)$ is the team's detection probability on path x, and $D_p(s)$ its average under scenario s.

The shaded map in Figure 2 shows where the survivor may be at one time; the paths show how locations connect across time. Managers need these connections to assess complementary coverage and decide where to search after a miss. The next result shows why separate probability maps are insufficient.

Proposition 1 (Path identity affects route value and follow-up decisions)**.** *Two trajectory beliefs may have identical location maps at every search time yet assign different detection values to the same route and produce different future-location beliefs after a miss. Time-indexed probability maps alone therefore do not, in general, determine sequential search value.*

A two-time construction proves the result (Appendix D). Operationally, route scoring and post-miss updating must preserve these temporal connections. TBR does so through shared survivor paths, although a sufficient transition model could retain the same information. For executed fleet assignment a, let $M_a(x) = 1 - Y_a(x)$ be the probability of missing path x and α_s the current scenario probability. Bayes' rule updates path and scenario beliefs:

$$\pi_s^+(x) = \frac{M_a(x)\pi_s(x)}{\langle M_a, \pi_s \rangle}, \qquad \alpha_s^+ = \frac{\alpha_s \langle M_a, \pi_s \rangle}{\sum_r \alpha_r \langle M_a, \pi_r \rangle}. \tag{4}$$

The fleet can then redirect effort toward survivor movements still consistent with the failed search. Scenarios with zero miss probability receive zero weight; if the total miss probability is zero, the miss branch is unreachable.

Protecting plans against shared drift uncertainty. Search managers need to know not only whether a fleet plan performs well on average, but also whether its routes can fail together when realized drift differs from the forecast. For plan p, $D_p(s)$ measures joint detection in drift scenario s. Its mean captures expected success, while its variation captures common downside exposure. We use these two quantities to construct a conservative performance floor.

To state this floor, let $\mathcal{H} = L^2(\mathcal{X}, \nu)$ be the space of square-integrable path-value functions. Fix a nonnegative unit-mass mean belief $\bar{\pi} \in \mathcal{H}$ and a positive, self-adjoint, trace-class covariance budget C. For a probability law Q over beliefs, define

$$\Delta_\Theta = \pi_\Theta - \bar{\pi}, \qquad C_Q = \mathbb{E}_Q[\Delta_\Theta \otimes \Delta_\Theta],$$

where $\otimes$ denotes the outer product. With $\mathbf{1}$ the constant-one function, consider the ambiguity set

$$\mathcal{Q} = \left\{ Q : \begin{array}{ll} Q(\pi_\Theta \geq 0,\ \langle \mathbf{1}, \pi_\Theta \rangle = 1) = 1, & \mathbb{E}_Q \|\pi_\Theta\|_{\mathcal{H}}^2 < \infty, \\ \mathbb{E}_Q[\pi_\Theta] = \bar{\pi}, & C_Q \preceq C \end{array} \right\}. \tag{5}$$

Here $C_Q \preceq C$ means $\langle f, C_Q f \rangle \leq \langle f, Cf \rangle$ for every $f \in \mathcal{H}$. Thus, $\mathcal{Q}$ admits alternative belief laws with the same mean and no more than the specified covariance. This belief covariance differs from GP velocity covariance $K_{\mathcal{D}}$: GP sampling and particle transport produce beliefs π_Θ, whose variation produces C_Q.

Operationally, we evaluate the bound on the retained trajectory ensemble Λ. This makes $\mathcal{H}$ finite-dimensional, with $\nu(\mathcal{X}) = |\Lambda| < \infty$, and allows $\mathcal{Q}$ to reweight the trajectories considered by the planner; $\delta_{\bar{\pi}} \in \mathcal{Q}$ ensures the set is nonempty. The ensemble moments are

$$Q_\alpha = \sum_s \alpha_s \delta_{\pi_s}, \qquad \bar{\pi}_\alpha = \sum_s \alpha_s \pi_s, \qquad C_\alpha = \sum_s \alpha_s (\pi_s - \bar{\pi}_\alpha) \otimes (\pi_s - \bar{\pi}_\alpha).$$

Setting $(\bar{\pi}, C) = (\bar{\pi}_\alpha, C_\alpha)$ gives the empirical score used in Section 4, and carries the same mean and covariance restrictions into every fleet plan's detection value.

Proposition 2 (Moments of belief-dependent coverage)**.** *For any $f \in \mathcal{H}$ and $Q \in \mathcal{Q}$,*

$$\mathbb{E}_Q[\langle f, \pi_\Theta \rangle] = \langle f, \bar{\pi} \rangle, \qquad \mathrm{Var}_Q(\langle f, \pi_\Theta \rangle) = \langle f, C_Q f \rangle \leq \langle f, Cf \rangle. \tag{6}$$

Taking $f = Y_p$ gives $\mathbb{E}_Q[D_p(\Theta)] = \langle Y_p, \bar{\pi} \rangle$ and $\mathrm{Var}_Q(D_p(\Theta)) = \langle Y_p, C_Q Y_p \rangle \leq \langle Y_p, CY_p \rangle$.

Managers can place more weight on unfavorable drift outcomes through a nonnegative, nonincreasing weight $w \in L^2(0,1)$ satisfying $\int_0^1 w(u)\, du = 1$. For a scalar reward Z, define

$$\rho_{w,Q}(Z) = \int_0^1 F_{Z,Q}^{-1}(u) w(u)\, du, \quad \kappa_w = \|w - 1\|_2, \quad \mathcal{V}_w(p) = \inf_{Q \in \mathcal{Q}} \rho_{w,Q}(D_p(\Theta)).$$

Here, $F_{Z,Q}^{-1}$ is the quantile function of Z under Q. The weighted reward $\rho_{w,Q}$ emphasizes low-detection outcomes, and $\mathcal{V}_w$ takes the worst value over the admitted belief laws.

Theorem 1 (Direct detection bound)**.** *Suppose $\mathcal{H}$ is separable, $\nu(\mathcal{X}) < \infty$, $\mathcal{Q}$ is nonempty, and C and w satisfy the conditions above. Then, for every feasible fleet plan,*

$$\mathcal{V}_w(p) \geq [J_{\mathrm{det}}(p)]_+, \qquad J_{\mathrm{det}}(p) = \langle Y_p, \bar{\pi} \rangle - \kappa_w \sqrt{\langle Y_p, CY_p \rangle}, \tag{7}$$

where $[z]_+ = \max\{0, z\}$.

Here, J_{det} balances expected joint detection against sensitivity to shared drift. Setting $\kappa_w = 0$ recovers mean success for a fixed miss-contingent sequence on the empirical ensemble (Proposition 3); positive κ_w favors plans that are more stable across drift scenarios. Proofs appear in Appendix C.

A route decomposition shows when assets share the same exposure. Route p_i alone has path coverage $c_{p_i} = Y_{(p_i)}$ and scenario payoff $R_i(\Theta) = \langle c_{p_i}, \pi_\Theta \rangle$. Their sums are the additive coverage $s_p = \sum_i c_{p_i}$ and payoff $U_p = \sum_i R_i$. For any two routes p, q, define their payoff covariance $k_Q(p,q) = \langle c_p, C_Q c_q \rangle$ and budget interaction $k_{\mathrm{Lag}}(p,q) = \langle c_p, C c_q \rangle$. Then

$$\mathrm{Var}_Q(U_p) = \sum_{i,j} k_Q(p_i, p_j) \leq \sum_{i,j} k_{\mathrm{Lag}}(p_i, p_j) = \langle s_p, C s_p \rangle. \tag{8}$$

Positive $k_Q(p_i, p_j)$ means that the same drift scenarios tend to raise or lower both route values, revealing duplicated exposure even when the routes are geographically separated. When $C = C_Q$, $k_{\mathrm{Lag}} = k_Q$. Because U_p can exceed one, the planner and Theorem 1 instead use nonlinear joint detection D_p; Corollary 1 gives the additive certificate.

In the following section, we use these scenario-dependent detection values to balance immediate coverage, shared drift exposure, and the search opportunities available after a miss.

4 COORDINATING THE SEARCH

Our proposed method, transported-belief rollout (TBR), uses the shared belief to coordinate current and future search through online receding-horizon route selection. A *leg* (one planning step) commits an asset for H ticks, truncated at the deadline; L is the number of lookahead legs. The goal is to make a long mission governable as a sequence of reviewable commitments rather than a fixed route chosen upfront. After each leg, TBR uses the miss-updated belief and remaining time to reconsider the fleet assignment, preserving flexibility as search evidence accumulates.

Balancing immediate detection and follow-up search. At each decision, managers must weigh the best chance of detection now against the fleet positions and search options preserved after a miss. State z summarizes positions, commitments, remaining time, and beliefs, while $\mathcal{A}(z)$ contains the feasible assignments. Assignment a detects now with probability $\bar{d}(z,a) = \sum_s \alpha_s \langle Y_a, \pi_s \rangle$; after a miss, Equation (4) updates the belief and moves the fleet to $\mathcal{T}_{\text{miss}}(z,a)$. The best attainable detection probability before the deadline therefore satisfies

$$W(z) = \sup_{a \in \mathcal{A}(z)} \left\{ \underbrace{\bar{d}(z,a)}_{\text{detect now}} + \underbrace{[1-\bar{d}(z,a)]W(\mathcal{T}_{\text{miss}}(z,a))}_{\text{search option after a miss}} \right\}, \tag{9}$$

with $W = 0$ at the deadline. The best assignment need not maximize immediate detection: a lower current probability can be worthwhile if it preserves stronger follow-up options. For an assignment sequence $\boldsymbol{a}$, let $p(\boldsymbol{a})$ be its fleet track and $V_s(\boldsymbol{a})$ its detection value in scenario s.

Proposition 3 (Miss-contingent sequence representation)**.** *With detection terminal and misses the only nonterminal feedback, a deterministic L-decision policy follows one miss-contingent sequence $\boldsymbol{a} = a_{1:L}$. If $D_{k,s}$ is leg-k detection after earlier misses and the legs partition $\mathcal{I}_t(p(\boldsymbol{a}))$, then*

$$V_s(\boldsymbol{a}) = \sum_{k=1}^{L} \left\{ \prod_{r<k} (1 - D_{r,s}) \right\} D_{k,s} = 1 - \prod_{k=1}^{L} (1 - D_{k,s}) = D_{p(\boldsymbol{a})}(s), \tag{10}$$

Thus success is counted once, and hypothetical miss updating is equivalent to full-track scoring. Averaging over scenarios gives modeled expected mission success, $\hat{\mu}(p(\boldsymbol{a})) = \sum_s \alpha_s V_s(\boldsymbol{a})$. With finite action sets and a horizon to the deadline, Equation (9) maximizes this mean over feasible joint sequences; TBR searches a manageable subset (Appendix D).

Making coordinated assignment tractable. Coordinated search requires each route to be valued against teammate choices: individually attractive routes can duplicate coverage, whereas complementary routes raise fleet detection. The ideal decision therefore selects one feasible L-leg sequence per asset to maximize modeled fleet success. Let $\mathcal{M}_{i,t}$ contain asset i's candidate sequences and binary y_{im} select sequence m; together, y defines fleet track $p(y)$ and mean detection $\hat{\mu}(y)$:

$$\max_{y} \ \hat{\mu}(y) \quad \text{subject to} \quad \sum_{m \in \mathcal{M}_{i,t}} y_{im} = 1 \ (i = 1, \ldots, N), \qquad y_{im} \in \{0,1\}. \tag{11}$$

Exact joint enumeration grows quickly with fleet size and lookahead. To keep replanning responsive, TBR assigns idle assets sequentially, using projected teammate tracks P_{-i} that follow existing commitments and then greedy extensions under the current belief. Each selected first leg becomes a commitment, so later assets adapt to earlier choices. This gains tractability at the cost of possible order dependence. With P_{-i} fixed, let $\mathcal{A}_t$ contain the asset's feasible first legs and $\mathcal{B}_t(a)$ its continuations after leg a. TBR selects a first leg and continuation $b = a_{2:L}$ by

$$a^\star \in \arg\max_{a \in \mathcal{A}_t} \max_{b \in \mathcal{B}_t(a)} \sum_s \alpha_s V_s(a, b; P_{-i}). \tag{12}$$

Here $p(a, b; P_{-i})$ combines the candidate and teammate tracks, and $V_s(a, b; P_{-i}) = D_{p(a,b;P_{-i})}(s)$. Only the first leg is executed before replanning, allowing the next decision to use new evidence. Proposition 3 shows that full-track scoring captures hypothetical miss conditioning. Sampling, candidate menus, limited lookahead, fixed ordering, and teammate projections keep the dynamic problem manageable (Appendix D). Section 5 tests lookahead and teammate projection; assignment order remains untested.

A tractable safeguard for joint assignment. Exact joint detection preserves shared-path identity but makes fleet-wide optimization difficult. For managers requiring explicit coverage, budget, and drift controls, Corollary 1 yields a convex additive assignment safeguard. At a 95% coverage floor, its best of 512 one-leg assignments reaches 35.04% success versus 42.59% for the coordinator, so TBR retains nonlinear joint detection; Appendix G gives the formulation, certificate, and evidence.

Managing downside risk from unfavorable drift. Managers may prefer a plan with nearly the same modeled mean but less exposure to unfavorable drift. For complete candidate track p, let $\hat{\mu} = \langle Y_p, \bar{\pi}_\alpha \rangle$ and $\hat{\sigma}^2 = \langle Y_p, C_\alpha Y_p \rangle$ under the scenario ensemble. The score $\hat{J}_\kappa = \hat{\mu} - \kappa\hat{\sigma}$ makes the tradeoff explicit: $\kappa = 0$ retains expected success, while positive κ favors plans that are more stable across scenarios. At the empirical moments it equals J_{det} with $\kappa_w = \kappa$; Appendix D gives its interpretation and limits. The penalty applies to the complete sequence, not recursively to individual legs or through additive route covariances.

For each first leg, let $b^\star(a) \in \arg\max_{b \in \mathcal{B}_t(a)} \hat{J}_\kappa(a, b)$ and set $m(a) = \hat{\mu}(a, b^\star(a))$. A mean tolerance $\varepsilon \geq 0$ screens these risk-selected continuations before the planner chooses

$$a^\star \in \arg \max_{a:\, m(a) \geq \max_{a'} m(a') - \varepsilon} \hat{J}_\kappa(a, b^\star(a)). \tag{13}$$

This separates two managerial controls: ε limits the mean gap among retained choices, while κ sets the preference for stability. At $\kappa = 0$, the rule recovers Equation (12).

5 Experiments

Search managers need evidence on four decisions: whether better coordination improves rescue probability and speed, which mechanisms create the gain, whether it persists across operating conditions, and whether added planning or downside protection is worth its cost. We study these questions in a common test bed of 180 search configurations spanning five OSCAR regions, three forecast splits, twelve starting conditions, three assets, and 720-tick missions. Targeted tests then vary fleet size, grid regime, current-field mismatch, planning budget, and wind uncertainty to identify when each choice matters.

Mission performance. To judge operational value, Table 1 measures both whether the survivor is found and how quickly. Mission success captures completion; capped detection time, $\mathbb{E}[\min(T_{\mathrm{detect}}, 720)]$, rewards earlier rescue while charging every miss the full 720 ticks. We isolate the sources of value through a planning ladder: the distance-based coordinator on a static belief map, the same rule on the transported belief, and then lookahead and teammate projection. Broader comparisons ask whether learned or classical allocation can deliver the same combination. These include PPO, which learns over TBR's belief and action menu; Belief-MAT, with its own particle belief; joint-observation PPO (JointCPPO), the strongest mean performer among four standard multi-agent RL policies; and a randomized lawnmower sweep.

The headline result is not a tradeoff between speed and success: TBR leads on both. It reaches 50.14% success in 467 capped ticks, adding 7.55 pp and saving 29 ticks versus the transported-belief coordinator, and adding 3.63 pp and saving 34 ticks versus PPO. This is a demanding comparison because the coordinator uses more samples and gains 2.81 pp from teammate-overlap control. For managers, the lower capped time means that higher success does not delay the response: TBR finds more survivors and finds them earlier within the same mission horizon. The advantage is also broad: it is positive in every region and remains 7.55 pp $[6.00, 9.18]$ over the coordinator and 4.04 pp $[3.34, 4.66]$ over PPO in the worst quarter. PPO's validation success rises only 0.45 pp over its final 500 episodes, indicating slowing but incomplete convergence. Together, these results favor plans that preserve valuable follow-up search options rather than exhausting effort on the immediate route.

Table 1: Mission performance across 180 configurations. TBR leads in average and worst-quarter success while reaching detection sooner. Brackets show 95% intervals; worst quarters contain each method's 45 lowest-success or slowest configurations, and bold marks the best estimates.

		Average			Worst quarter	
Planning configuration	Steps ahead	Success (%)	TBR − method (pp) [95% interval]	Capped time	Success (%)	Capped time
Proposed approach						
Transported-belief rollout (TBR)	3	**50.14**	Reference	**467**	**42.96**	**510**
Coordination and lookahead variants						
Static-map coordinator	1	29.68	+20.46 [18.48, 22.54]	563	21.00	618
Transported-belief coordinator	1	42.59	+7.55 [6.88, 8.24]	497	35.40	540
Two-step, teammates fixed	2	44.66	+5.48 [4.50, 6.48]	482	39.41	516
Two-step, teammates projected	2	46.41	+3.73 [3.20, 4.30]	475	40.63	513
Three-step, teammates fixed	3	48.03	+2.11 [1.72, 2.54]	471	41.85	511
Learned RL policies						
PPO	1	46.51	+3.63 [2.93, 4.30]	501	38.91	538
Belief-MAT policy	1	25.99	+24.15 [23.03, 25.30]	585	19.65	620
Multi-agent RL (best of four)	1	22.73	+27.42 [26.54, 28.33]	602	16.18	641
Classical search						
Lawnmower sweep	–	8.58	+41.56 [40.60, 42.59]	668	3.26	700

Why coordination works. Relative to two-step planning with fixed teammates (44.66%), extra lookahead adds 3.36 pp, teammate projection 1.75 pp, and both 5.48 pp (Figure 3). Discarding shared paths costs 10.67 pp; sensor and scenario-averaging changes are small. One-step joint detection reaches 40.77%, below the coordinator's 42.59% (Table 5); direct detection scoring alone does not outperform the coordinator in this benchmark.

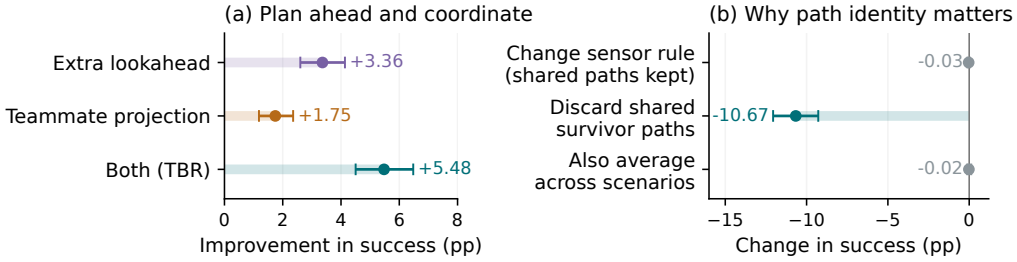


Figure 3: Coordination pays through complementary mechanisms. (a) Lookahead and teammate projection jointly add 5.48 pp. (b) Preserving shared path identity is decisive: separate route averages lose 10.67 pp, while other scoring changes are negligible. Bars show adjusted 95% intervals.

Table 2: Fleet and grid sensitivity (180 configurations per row). TBR remains ahead in every setting; teammate projection becomes more valuable with larger fleets, while deeper planning matters more on larger grids. TBR–coordinator intervals are adjusted 95%.

Assets	Grid	TBR success (%)	Coordinator success (%)	TBR − coordinator (pp) [95% interval]	Third step (pp)	Teammate projection (pp)
2	300	**42.06**	34.70	+7.36 [6.48, 8.27]	+3.75	+1.11
3	300	**50.14**	42.59	+7.55 [6.88, 8.24]	+3.36	+1.75
5	300	**64.18**	55.12	+9.06 [8.05, 10.10]	+2.62	+2.20
3	100	**13.12**	10.27	+2.85 [2.12, 3.63]	+0.02	+0.13
3	500	**59.85**	55.62	+4.23 [3.34, 5.04]	+2.03	+1.30

Sensitivity to fleet size and grid regime. Table 2 suggests potential asset savings: two-asset TBR reaches 42.06%, close to the three-asset coordinator's 42.59%, while three-asset TBR lies near an interpolated four-asset coordinator; these are descriptive comparisons. Teammate-projection gains rise from $+1.11$ to $+1.75$ and $+2.20$ pp for two, three, and five assets (Appendix Table 16). Belief transport adds 11.7–12.9 pp across fleet sizes but only 0.9 pp in 240-tick missions, where another planning step adds 0.02 pp.

Environmental resilience. TBR remains preferred across current and wind conditions. It adds 7.1–8.0 pp under current variation or benign wind and 3.7–4.9 pp under adverse wind in every region. Managers retain a positive margin with imperfect forecasts, although adverse wind narrows it.

Table 3: TBR advantage across current and wind conditions.

What varies	Conditions	Coordinator	TBR	TBR − coordinator
Current pattern	Five ocean regions	39.8–46.3	47.7–53.4	+7.1 to +7.8
True current	Three held-out fields	43.8	51.4–51.5	+7.6 to +7.7
Benign wind	Nominal or gusty	42.8–43.0	50.6–50.8	+7.7 to +8.0
Adverse wind	$\pm 60^\circ$ uncertainty or $\pm 20^\circ$ bias	35.9–42.9	40.2–47.8	+3.7 to +4.9

Managing planning effort and downside risk. Search managers must decide whether deeper route search is worth its delay. Figure 4 compares TBR with greedy continuation and two UCT budgets. Greedy nearly matches TBR at one-quarter of its route-construction time; the broadest five-step UCT adds 1.36 pp in final success but nearly quadruples computation and trails TBR by 1.13 pp at mid-mission. Urgent missions therefore favor TBR or greedy continuation unless marginal final success justifies added search. TBR takes about 10 seconds per mission, or roughly 0.3 seconds per asset-decision on one thread, across fleet sizes and mission regimes. A separate wind-sensitivity test shows that modest protection can improve downside performance: $\kappa = 0.25$ raises lower-quarter success by 0.82 pp $[0.44, 1.22]$ for a 0.22 pp $[0.07, 0.37]$ mean cost.

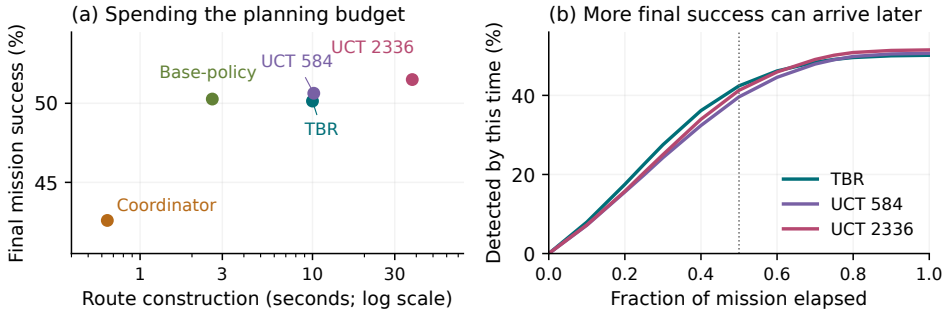


Figure 4: Computation versus rescue urgency. (a) Greedy continuation takes 2.62 versus TBR's 10.00 seconds; broader UCT reaches 51.50% at 37.89 seconds. (b) TBR leads at mid-mission (42.41% versus 41.28%). Together, the panels show that more planning can improve final success while delaying useful action.

6 Conclusion

In this paper, we introduce transported-belief rollout (TBR) to coordinate present and future searches through shared survivor trajectories. Across the tested conditions, it improves success over the coordinator and evaluated learned policies. Controlled comparisons attribute the gain to belief transport, path-specific scoring, lookahead, and teammate projection without implying additive effects. A moment bound makes modest downside protection explicit. The evidence supports matching planning effort to rescue urgency and coordination needs. Future work should test TBR with live forecasts and field trials, online environmental updates, and heterogeneous fleets under sensing and communication constraints.

AI USE STATEMENT

Generative AI tools assisted with manuscript drafting and revision, analysis of reviewer feedback, mathematical derivations and proof checks, experiment design and implementation, result analysis, and consistency checks between equations, code, and reported results. They also assisted with typesetting and cross-reference checks. The authors are responsible for the correctness of the mathematics, experiments, citations, and final submitted content.

ETHICS STATEMENT

The study uses simulated searches and public ocean-current fields, with no human subjects or personally identifying data. Its intended application is search and rescue. These simulations do not establish operational safety: deployment would require validation of local forecasts and sensors, mission constraints, and human supervision. Coordinated-search methods also have potential surveillance applications; responsible use should respect privacy and the authority of search managers. The authors declare no conflicts of interest.

REPRODUCIBILITY STATEMENT

Appendix A maps the main claims to their supporting results. Appendices B and C give the core assumptions and proofs; full theory extensions follow Appendix M. Appendix D specifies the simulator and planner, and Appendix E defines the comparison and statistical protocols. The topical result sections retain learning and simulation budgets, selection history, and additional results. Recorded configurations preserve seeds, budgets, final checkpoint choices, outcomes, and the analysis protocol. The planned supplementary release includes the code, processed fields, frozen configurations, trained checkpoints, and saved analyses needed to reproduce the reported comparisons.

REFERENCES

Bo Ai, Maoxin Jia, Hanwen Xu, Jiangling Xu, Zhen Wen, Benshuai Li, and Dan Zhang. Coverage path planning for maritime search and rescue using reinforcement learning. *Ocean Engineering*, 241:110098, 2021. doi: 10.1016/j.oceaneng.2021.110098.

Sushmita Bhattacharya, Siva Kailas, Sahil Badyal, Stephanie Gil, and Dimitri Bertsekas. Multiagent rollout and policy iteration for POMDP with application to multi-robot repair problems. In *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pp. 1814–1828. PMLR, 2021. URL https://proceedings.mlr.press/v155/bhattacharya21a.html.

Julian Bialas, Mohammad Reza Mohebbi, Michiel J. van Veelen, Abraham Mejia-Aguilar, Robert Kathrein, and Mario Döller. From human teams to autonomous swarms: A reinforcement learning-based benchmarking framework for unmanned aerial vehicle search and rescue missions. *Drones*, 10(2):79, 2026. doi: 10.3390/drones10020079.

Xiao Cao, Mingyang Li, Yuting Tao, and Peng Lu. HMA-SAR: Multi-agent search and rescue for unknown located dynamic targets in completely unknown environments. *IEEE Robotics and Automation Letters*, 9(6):5567–5574, 2024.

Li Chen, Simai He, and Shuzhong Zhang. Tight bounds for some risk measures, with applications to robust portfolio selection. *Operations Research*, 59(4):847–865, 2011. doi: 10.1287/opre.1110.0950.

Sung-Won Cho, Jin-Hyoung Park, Hyun-Ji Park, and Seongmin Kim. Multi-UAV coverage path planning based on hexagonal grid decomposition in maritime search and rescue. *Mathematics*, 10 (1):83, 2022. doi: 10.3390/math10010083.

Timothy H. Chung, Geoffrey A. Hollinger, and Volkan Isler. Search and pursuit-evasion in mobile robotics: A survey. *Autonomous Robots*, 31(4):299–316, 2011.

John H. Cochrane. *Asset Pricing*. Princeton University Press, revised edition, 2005.

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58(3):595–612, 2010. doi: 10.1287/opre.1090.0741.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. *Mathematical Programming*, 171:115–166, 2018.

ESR. OSCAR third degree resolution ocean surface currents. https://doi.org/10.5067/OSCAR-03D01, 2009. Earth & Space Research, accessed 2026.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993. doi: 10.1016/0304-405X(93)90023-5.

Yingying Gao, Guangyin Jin, Yu Guo, Guohai Zhu, Qingqing Yang, and Kewei Yang. Weighted area coverage of maritime joint search and rescue based on multi-agent reinforcement learning. In *2019 IEEE 3rd Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*, pp. 593–597, Chongqing, China, October 2019. IEEE. doi: 10.1109/IMCEC46724.2019.8984116.

Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101(1-2):99–134, 1998.

Levente Kocsis and Csaba Szepesvári. Bandit based monte-carlo planning. In *Machine Learning: ECML 2006*, volume 4212 of *Lecture Notes in Computer Science*, pp. 282–293. Springer, 2006. doi: 10.1007/11871842_29.

Bernard O. Koopman. The theory of search. *Operations Research*, 5(5):613–626, 1957. doi: 10.1287/opre.5.5.613.

Thomas M. Kratzke, Lawrence D. Stone, and John R. Frost. Search and rescue optimal planning system. In *2010 13th International Conference on Information Fusion*, pp. 1–8, 2010. doi: 10.1109/ICIF.2010.5712114. URL https://www.metsci.com/wp-content/uploads/2019/08/Search-and-Rescue-Optimal-Planning-System.pdf.

Chengjia Lei, Shaohua Wu, Yi Yang, Jiayin Xue, and Qinyu Zhang. Joint trajectory and communication optimization for heterogeneous vehicles in maritime SAR: Multi-agent reinforcement learning. *IEEE Transactions on Vehicular Technology*, 73(9):12328–12344, 2024.

Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. In *Advances in Neural Information Processing Systems*, 2017.

Michael Morin, Irène Abi-Zeid, and Claude-Guy Quimper. Ant colony optimization for path planning in search and rescue operations. *European Journal of Operational Research*, 305(1):53–63, 2023. doi: 10.1016/j.ejor.2022.06.019.

Youngon Oh, Jeonghyo Oh, Seunghyun Kim, and Impyeong Lee. Coverage path planning for maritime search and rescue using maskable proximal policy optimization. *Korean Journal of Remote Sensing*, 41(6):1161–1178, 12 2025. doi: 10.7780/kjrs.2025.41.6.17.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

Stephen A. Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976. doi: 10.1016/0022-0531(76)90046-6.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

H. Shao and Z. G. Zhang. Extreme-case distortion risk measures: a unification and generalization of closed-form solutions. *Mathematics of Operations Research*, 49(4):2341–2355, 2024. doi: 10.1287/moor.2022.0156.

David Silver and Joel Veness. Monte-Carlo planning in large POMDPs. In *Advances in Neural Information Processing Systems*, 2010.

Lawrence D. Stone. *Theory of Optimal Search*, volume 118 of *Mathematics in Science and Engineering*. Academic Press, New York, 1975.

Guangcheng Wang, Fenglin Wei, Yu Jiang, Minghao Zhao, Kai Wang, and Hong Qi. A Multi-AUV maritime target search method for moving and invisible objects based on multi-agent deep reinforcement learning. *Sensors*, 22(21):8562, 2022. doi: 10.3390/s22218562. URL https://doi.org/10.3390/s22218562.

Muning Wen, Jakub Grudzien Kuba, Runji Lin, Weinan Zhang, Ying Wen, Jun Wang, and Yaodong Yang. Multi-agent reinforcement learning is a sequence modeling problem. In *Advances in Neural Information Processing Systems*, volume 35, 2022. URL https://papers.nips.cc/paper_files/paper/2022/hash/69413f87e5a34897cd010ca698097d0a-Abstract-Conference.html.

Wolfram Wiesemann, Daniel Kuhn, and Melvyn Sim. Distributionally robust convex optimization. *Operations Research*, 62(6):1358–1376, 2014. doi: 10.1287/opre.2014.1314.

Jie Wu, Liang Cheng, Sensen Chu, and Yanjie Song. An autonomous coverage path planning algorithm for maritime search and rescue of persons-in-water based on deep reinforcement learning. *Ocean Engineering*, 291:116403, January 2024. doi: 10.1016/j.oceaneng.2023.116403.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of PPO in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 2022.

Jinfen Zhang, Ângelo P. Teixeira, C. Guedes Soares, and Xinping Yan. Probabilistic modelling of the drifting trajectory of an object under the effect of wind and current for maritime search and rescue. *Ocean Engineering*, 129:253–264, 2017. doi: 10.1016/j.oceaneng.2016.11.002.

A Appendix guide and evidence map

The supporting material follows the paper's argument: detection and planning foundations, implementation, comparisons, mechanisms, operating conditions, and decision tradeoffs. The final part retains the full theory extensions and their proofs. They analyze an additive covariance selector that TBR does not use.

Table 4: Where to verify the main claims. This is a navigation guide, not an additional analysis. Table and section links lead to the original comparison definitions, interval families, and implementation details.

Main claim or question	Supporting material
Trajectory identity and follow-up decisions	Proposition 1 and Proposition 3; proofs in D
Detection bound and additive assignment	Assumptions in B; proofs in C; optional extensions begin in M
GP, scenario draws, and TBR	D; Algorithm 1
Comparators and statistical units	E: comparison rules and statistical protocol
Table 1: mean, time, and configuration tails	F; Tables 5, 7, 8
Lookahead and teammate projection	Table 9
Shared-path scoring and covariance controls	Tables 11, 12, 10; exact assignment in G
Transport and non-detection updates	Table 13; separate one-leg reference
Table 2: fleet and mission regimes	H; Tables 15, 16
Unseen-current and GP sensitivity	H; Tables 17, 18
Planning effort and rescue urgency	I; Tables 19, 21, 22
Wind downside protection	J; final test in Table 25 and Figure 9
Learned policies and training progress	K; Tables 26, 27, 28
Scope of the evidence	L

Three distinctions used throughout. Shared-path coverage is dependence between routes within a drift scenario; belief covariance describes variation across scenarios. A configuration's expected success integrates survivor trajectories and sensor outcomes; those trajectories are not independent experimental replicates. Finally, Table 1's worst quarter selects difficult starting configurations, whereas the wind tail averages adverse wind offsets within each configuration. These quantities answer different questions.

How to read the retained theory. The directly used results and their proofs appear first. The later extension part covers covariance factorization, menu reduction, assignment, and local replanning. Every theoretical statement and complete proof is retained, including results not needed by the main text. The extension guide explains their scope without presenting them as evaluated components of TBR.

B Belief covariance and transported coordinates

Scope. This section supplies the belief-law assumptions and additive certificate used in Sections 3–4. Proofs of the detection and moment results follow in Appendix C; trajectory and miss-branch proofs accompany the implementation in Appendix D. The optional spectral and assignment extensions begin at Appendix M.

For the moving-target model, $\mathcal{X}$ is the space of target trajectories in Section 3, including the exit state; x denotes a target realization when used as an argument of a coverage function. Spatial waypoint and GP constructions below are identified separately. Integrals on the belief space use the common finite measure ν. Spatial-density formulas are the corresponding special case. The population results require the stated belief law and covariance budget; the finite-scenario rollout is described in Section 4.

Condition on the observations at a planning epoch and fix a common reference flow. A scenario Θ samples environmental uncertainty, while target-location and process uncertainty are integrated into its normalized belief π_Θ. Particle histograms on a finite grid give probability vectors. In continuous space we assume a common finite measure space $(\mathcal{X}, \nu)$ with separable $\mathcal{H} = L^2(\mathcal{X}, \nu)$, square-integrable belief densities, and $\mathbb{E}_Q\|\pi_\Theta\|_2^2 < \infty$. The common mean is $\bar{\pi}$. Define

$$\Delta_\Theta = \pi_\Theta - \bar{\pi}, \qquad C_Q = \mathbb{E}_Q[\Delta_\Theta \otimes \Delta_\Theta].$$

The total target-realization variance obeys

$$\mathrm{Var}(s_p(X)) = \mathbb{E}_Q[\mathrm{Var}(s_p(X) \mid \Theta)] + \mathrm{Var}_Q(U_p).$$

The within-scenario term can depend on the paths and is not asserted to be irreducible or determined by the mean. The objective in (24) specifically evaluates the second level of uncertainty through D_p.

The operator C_Q is positive, self-adjoint, and trace-class. Because $\langle \mathbf{1}, \Delta_\Theta \rangle = 0$, it satisfies $C_Q\mathbf{1} = 0$. We use a fixed positive, self-adjoint trace-class budget C, also with $C\mathbf{1} = 0$, and the plan-independent ambiguity set

$$\begin{aligned}\mathcal{Q} = \{Q : Q(\pi_\Theta \geq 0,\ \langle \mathbf{1}, \pi_\Theta \rangle = 1) = 1,\\ \mathbb{E}_Q\|\pi_\Theta\|_{\mathcal{H}}^2 < \infty, \quad \mathbb{E}_Q \pi_\Theta = \bar{\pi}, \quad C_Q \preceq C\}.\end{aligned} \tag{14}$$

Here $C_Q \preceq C$ means $\langle f, C_Q f \rangle \leq \langle f, Cf \rangle$ for every $f \in \mathcal{H}$. The deterministic belief $\bar{\pi}$ makes this set nonempty. These are model assumptions; fitting a sample covariance alone is not a statistical confidence guarantee that the unknown environment belongs to $\mathcal{Q}$. For a pointwise Mercer expansion we additionally assume a continuous positive-semidefinite kernel on the compact domain; without continuity the spectral identities are read in the L^2 sense, or as ordinary matrix identities on the grid.

Environment covariance is an input, not the planning covariance. The residual-GP construction below is an optional coordinate interpretation, not the fitted current model used by TBR. The implemented two-output GP and scenario draws are specified in Appendix D. Let u be the posterior mean velocity and $\Phi_{\tau,0}$ its invertible backward flow map:

$$\tilde{x} = \Phi_{\tau,0}(x). \tag{15}$$

Observed velocities $\hat{v}_k$ provide residual data

$$y_k = \hat{v}_k - u(X_{\mathrm{obs}}(t_k), t_k), \qquad \tilde{x}_k = \Phi_{t_k,0}(X_{\mathrm{obs}}(t_k)). \tag{16}$$

An environmental GP, componentwise for vector velocity, can model these residuals:

$$\xi_{\mathrm{vel}} \sim \mathcal{GP}(0, K_{\mathrm{vel}}), \qquad K_{\mathrm{vel}} = K_\theta - K_*^\top (K_{\mathrm{train}} + \sigma_{\mathrm{obs}}^2 I)^{-1} K_*'. \tag{17}$$

Sampling drift fields and propagating target particles produces valid beliefs. Only then do we form Δ_Θ and C_Q. In particular, Δ_Θ is not freely sampled as an additive Gaussian density perturbation, and K_{vel} is not identified with belief covariance. Velocity uncertainty can induce belief covariance through propagation; the mass constraint does not rule out such a mapping, but it rules out an unqualified equality of the two kernels.

Coordinate invariance. For the same fixed invertible map T in every scenario (a spatial reference flow, or its trajectory-space extension), set $\tilde{\pi}_\Theta = T_\# \pi_\Theta$ and $\tilde{f} = f \circ T^{-1}$. Then

$$\int \tilde{f}\, d\tilde{\pi}_\Theta = \int f\, d\pi_\Theta.$$

For densities, the pushforward includes the Jacobian of T^{-1}. Transforming both fields and measures consistently preserves means, variances, and the robust certificates. We write K_{Lag} for the kernel of the resulting planning budget in the common material frame and suppress tildes on effort fields. No spectral advantage follows from this identity. A finite-grid invertible change of representation preserves matrix rank; under a measure-preserving unitary transformation the exact covariance spectrum is also preserved. Effects of smoothing, discretization, or fitted-kernel approximation require separate evidence. No Eulerian-versus-material spectral comparison is claimed here.

Proposition 2 gives the moment identity and Equation (8) its route decomposition; its proof is in Appendix C. In all subsequent spectral and menu formulas, $\mu_{U_p} = \mu_p$ and $\sigma_{U_p}^2 = \sigma_p^2$ denote the common scenario mean and its covariance budget, not moments of $s_p(X)$.

$$D_p(\Theta) = \langle Y_p, \pi_\Theta \rangle, \qquad U_p(\Theta) = \langle s_p, \pi_\Theta \rangle. \tag{18}$$

From additive route value to a detection bound. The nonincreasing closest-agent sensor yields

$$Y_p \geq \max_i c_{p_i} \geq s_p/N, \qquad D_p \geq U_p/N. \tag{19}$$

Corollary 1 (Additive detection bound)**.** *Under the assumptions of Theorem 1, let* $\mu_p = \langle s_p, \bar{\pi} \rangle$ *and* $\sigma_p^2 = \langle s_p, C s_p \rangle$*. Then*

$$\inf_{Q \in \mathcal{Q}} \rho_{w,Q}(D_p) \geq \frac{1}{N}[J_{\mathrm{add}}(p)]_+, \qquad J_{\mathrm{add}}(p) = \mu_p - \kappa_w \sigma_p. \tag{20}$$

This corollary supplies a route-additive alternative to the direct fleet bound. Its covariance budget separates as

$$\sigma_p^2 = \sum_{i,j} k_{\mathrm{Lag}}(p_i, p_j), \qquad k_{\mathrm{Lag}}(p, q) = \langle c_p, C c_q \rangle. \tag{21}$$

When $C = C_Q$, the pairwise term is the covariance of individual route detection values. The direct planner retains Y_p and does not replace it by s_p. The complete proof appears in Appendix C.

For densities on a general reference space, use the pushed-forward measure $\tilde{\nu} = T_{\#}\nu$ with $\tilde{f} = f \circ T^{-1}$ and $\tilde{\pi}_\Theta = \pi_\Theta \circ T^{-1}$. This transformation is unitary between the two L^2 spaces. If the reference measure is held fixed instead, the density transformation includes the appropriate Jacobian. A change of coordinates at one planning epoch is distinct from physical propagation of the target belief through time.

Pointwise coverage and the additive certificate

This appendix relates the direct bound in Section 3 to the optional additive and spectral selectors.

Assumption 1 (Lower-tail reward convention)**.** *Fix a measurable, nonnegative, nonincreasing* $w \in L^2(0,1)$ *with* $\int_0^1 w(u)\,du = 1$*. For a square-integrable reward* Z *with law induced by* Q*, define*

$$\rho_{w,Q}(Z) = \int_0^1 F_{Z,Q}^{-1}(u) w(u)\,du, \qquad \kappa_w = \left(\int_0^1 (w(u) - 1)^2\,du \right)^{1/2}. \tag{22}$$

Larger rewards are preferable. The weight emphasizes lower quantiles; the functional is monotone and positively homogeneous. The expectation has $w = 1$ *and* $\kappa_w = 0$*. The worst-*γ *lower-tail average has* $w = \mathbf{1}_{(0,\gamma)}/\gamma$ *and* $\kappa_w = \sqrt{(1-\gamma)/\gamma}$*, for* $0 < \gamma < 1$*;* γ *is distinct from the scenario weights* α_s*.*

This quantile definition replaces any ambiguous survival-versus-CDF distortion convention. When an absolutely continuous cumulative distortion is desired, it is $\phi(u) = \int_0^u w(v)\,dv$; no separate distortion convention is used in the proofs.

Lemma 1 (Pointwise search bounds)**.** *For every feasible plan and target realization under a closest-agent sensor whose single-visit detection probability is nonincreasing in distance,*

$$\max_i c_{p_i}(x) \leq Y_p(x) \leq \min\{s_p(x), 1\}, \qquad Y_p(x) \geq s_p(x)/N.$$

Under the alternative conditionally independent-agent model, write $Y_p^{\mathrm{ind}}(x) = 1 - \prod_i (1 - c_{p_i}(x))$*. Then additionally*

$$1 - e^{-s_p(x)} \leq Y_p^{\mathrm{ind}}(x) \leq \min\{s_p(x), 1\}, \qquad Y_p^{\mathrm{ind}}(x) \geq \max_i c_{p_i}(x) \geq s_p(x)/N.$$

Both detection set functions are monotone submodular for a fixed menu of routes.

The full proof below retains the product-model derivation and supplies the closest-agent coupling argument. The exponential inequality is specific to the independent-agent model: with two colocated agents and one tick of individual detection probability 0.02, the closest-agent probability is $0.02 < 1 - e^{-0.04}$. The coefficient $1/N$ is sharp in both models; when all $c_{p_i} = 1$, team detection is 1 and $s_p = N$.

Lemma 2 (Scenario certificate via a pointwise envelope)**.** *If $f_p \in \mathcal{H}$ and $Y_p(x) \geq f_p(x)$ for every target realization, then $D_p(\Theta) \geq \langle f_p, \pi_\Theta \rangle$ for each valid scenario belief, and*

$$\rho_{w,Q}(D_p) \geq \rho_{w,Q}(\langle f_p, \pi_\Theta \rangle) \quad \textit{for every } Q \in \mathcal{Q}.$$

In particular, $f_p = s_p/N$ gives $\rho_{w,Q}(D_p) \geq \rho_{w,Q}(U_p)/N$.

The direct detection certificate in Theorem 1 uses Y_p itself and needs no envelope. The additive envelope is a tractability step: it permits pairwise path decomposition, without identifying the additive selector with an optimizer of nonlinear detection.

Corollary 1 states the additive certificate in Appendix B. Maximizing J_{add} maximizes that certificate, possibly with extra ties after clipping at zero. It need not maximize J_{det} or the robust detection objective. Taken together, the theorem and corollary retain the combined certificate

$$\inf_{Q \in \mathcal{Q}} \rho_{w,Q}(D_p) \geq \max\{0, J_{\mathrm{det}}(p), J_{\mathrm{add}}(p)/N\}. \tag{23}$$

The relation $D_p \geq U_p/N$ orders their reward-risk values, but does not in general order their mean–standard-deviation bounds: this score is not monotone under pointwise dominance. When $\kappa_w = 0$, expectations preserve the order, so the direct bound already dominates the additive one. The separate statements make each role explicit without needing such an ordering at positive risk weight.

C Proofs for detection and additive coverage

This section proves the moment and detection results used in Sections 3–4, including the pointwise envelope behind the assignment surrogate. Full proofs for the optional extensions are retained in Appendix W.

The belief space is $\mathcal{H} = L^2(\mathcal{X}, \nu)$; $\mu_{U_p} = \mu_p$ and $\sigma_{U_p} = \sigma_p$ denote the scenario mean and covariance budget.

Proof of Theorem 1

The robust value and moment coefficient used in Section 3 are

$$\mathcal{V}_w(p) = \inf_{Q \in \mathcal{Q}} \rho_{w,Q}(D_p), \qquad \kappa_w = \|w - 1\|_2. \tag{24}$$

Proof. For a square-integrable reward Z, write $m = \mathbb{E}_Q Z$ and $q(u) = F^{-1}_{Z,Q}(u)$. Since $\int_0^1 (q - m)\, du = 0$, $\int_0^1 w\, du = 1$, and $\int_0^1 (q - m)^2\, du = \mathrm{Var}_Q\, Z$, Cauchy–Schwarz gives

$$\begin{aligned} \rho_{w,Q}(Z) - m &= \int_0^1 (q(u) - m)(w(u) - 1)\, du \\ &\geq -\|w - 1\|_2 \sqrt{\mathrm{Var}_Q\, Z}. \end{aligned} \tag{25}$$

For $D_p = \langle Y_p, \pi_\Theta \rangle$, the mean is $\langle Y_p, \bar{\pi} \rangle$ and the variance is $\langle Y_p, C_Q Y_p \rangle \leq \langle Y_p, C Y_p \rangle$. Substitution into (25) proves $\rho_{w,Q}(D_p) \geq J_{\mathrm{det}}(p)$ for every $Q \in \mathcal{Q}$.

Finally, $D_p \geq 0$ and $w \geq 0$ imply $\rho_{w,Q}(D_p) \geq 0$. Both lower bounds are independent of Q, so their maximum remains a lower bound after taking the infimum over the same plan-independent $\mathcal{Q}$. □

This proof needs no symmetry assumption or extremal-distribution existence result. Scalar moment-risk bounds and robust portfolio applications are studied by Chen et al. (2011); extreme-case distortion-risk results appear in Shao & Zhang (2024). These results provide related context, but do not imply attainment over valid probability-belief laws with a covariance-operator budget. In particular, $D_p \in [0, 1]$ and $U_p \in [0, N]$, whereas unrestricted scalar moment extremizers may have different support.

PROOF OF PROPOSITION 2

The equivalent centered form writes

$$Z_p := U_p - \mu_p = \langle s_p, \Delta_\Theta \rangle, \qquad \mu_p = \langle s_p, \bar{\pi} \rangle. \tag{26}$$

Then $\mathbb{E}_Q Z_p = 0$ and $\mathrm{Var}_Q(Z_p) = \mathrm{Var}_Q(U_p)$; the variance bound in the proposition applies to both. For individual routes,

$$\mathrm{Cov}_Q(R_i, R_j) = \mathbb{E}_Q\big[\langle c_{p_i}, \Delta_\Theta \rangle \langle c_{p_j}, \Delta_\Theta \rangle\big] = \langle c_{p_i}, C_Q c_{p_j} \rangle = k_Q(p_i, p_j).$$

Thus the portfolio identity concerns variation across drift scenarios after averaging over survivor paths within each scenario. The condition $C_Q \preceq C$ bounds aggregate variance, not each cross-term: it need not imply $k_Q(p_i, p_j) \le k_{\mathrm{Lag}}(p_i, p_j)$ for $i \neq j$. For fixed route variances, lower covariance reduces the variance of additive coverage. The joint detection field Y_p combines encounters nonlinearly; the additive route-pair expansion does not generally apply to it. The centered form above also supports the extensions in Appendices B and N.

Proof. For any $f \in \mathcal{H}$, $\langle f, \pi_\Theta \rangle$ is square integrable because

$$\mathbb{E}_Q |\langle f, \pi_\Theta \rangle|^2 \le \|f\|_2^2 \mathbb{E}_Q \|\pi_\Theta\|_2^2 < \infty.$$

Linearity of expectation gives $\mathbb{E}_Q \langle f, \pi_\Theta \rangle = \langle f, \bar{\pi} \rangle$. The definition of C_Q then gives

$$\mathrm{Var}_Q(\langle f, \pi_\Theta \rangle) = \mathbb{E}_Q \langle f, \Delta_\Theta \rangle^2 = \langle f, C_Q f \rangle \le \langle f, C f \rangle.$$

Take $f = s_p$ or $f = Y_p$. Expanding $s_p = \sum_i c_{p_i}$ in the bilinear form gives the path-pair decomposition. Further,

$$C_Q \mathbf{1} = \mathbb{E}_Q[\Delta_\Theta \langle \Delta_\Theta, \mathbf{1} \rangle] = 0,$$

since each belief and its mean have unit mass. Finite-grid statements follow from the same matrix calculation. No Gaussian assumption on beliefs is required. □

PROOF OF LEMMA 1

Closest-agent sensor.

Proof. Fix a target trajectory and write $q_{i\tau}$ for agent i's detection probability at tick τ. Because single-visit detection is nonincreasing in distance, the closest agent has probability $\max_i q_{i\tau}$. Let U_τ be independent uniform random variables across ticks, shared across agents at each tick. Define

$$A_i = \{U_\tau \le q_{i\tau} \text{ for at least one } \tau\}.$$

Then $\mathbb{P}(A_i) = c_{p_i}$, and

$$\mathbb{P}\left(\bigcup_i A_i\right) = 1 - \prod_\tau \left(1 - \max_i q_{i\tau}\right) = Y_p.$$

Event containment and the union bound give

$$\max_i c_{p_i} \le Y_p \le \min\left\{\sum_i c_{p_i}, 1\right\}, \qquad Y_p \ge \frac{1}{N} \sum_i c_{p_i}.$$

For sets of fixed candidate routes $A \subseteq B$ and $j \notin B$,

$$\mathbb{P}\left(A_j \setminus \bigcup_{i \in A} A_i\right) \ge \mathbb{P}\left(A_j \setminus \bigcup_{i \in B} A_i\right) \ge 0.$$

These are the corresponding marginal detection gains. Thus the set function is monotone and submodular. This fixed-route property makes no claim that adaptive route generation or a mean–deviation objective is submodular. The shared uniforms are a proof coupling, not extra sensor draws. □

Independent-agent product model.

Proof. Fix a feasible fleet plan $p = (p_1, \ldots, p_N)$ and fix an arbitrary target realization x. Define

$$a_i := c_{p_i}(x) \in [0, 1], \qquad i = 1, \ldots, N,$$

and

$$S_p(x) = \sum_{i=1}^{N} a_i.$$

Under the product sensing model, equivalently under conditional independence of the agents' detection events given $X = x$, the team detection probability is

$$Y_p^{\mathrm{ind}}(x) = 1 - \prod_{i=1}^{N}(1 - a_i).$$

We first prove the upper bound. Let $A_i(x)$ denote the event that agent i detects the target conditional on $X = x$, so that

$$\mathbb{P}(A_i(x) \mid X = x) = a_i.$$

Then

$$Y_p^{\mathrm{ind}}(x) = \mathbb{P}\left(\bigcup_{i=1}^{N} A_i(x) \,\middle|\, X = x\right).$$

By the union bound,

$$\begin{aligned} Y_p^{\mathrm{ind}}(x) &\le \sum_{i=1}^{N} \mathbb{P}(A_i(x) \mid X = x) \\ &= \sum_{i=1}^{N} a_i = S_p(x). \end{aligned}$$

Also $Y_p^{\mathrm{ind}}(x) \le 1$ because it is a probability. Therefore

$$Y_p^{\mathrm{ind}}(x) \le \min\{S_p(x), 1\}.$$

This upper bound uses only the union bound; it does not rely on independence.

We next prove the exponential lower bound. For every $u \in [0, 1]$,

$$1 - u \le e^{-u}.$$

Applying this inequality to $u = a_i$ gives

$$1 - a_i \le e^{-a_i}, \qquad i = 1, \ldots, N.$$

Since all factors are nonnegative, multiplying these inequalities preserves the inequality direction:

$$\prod_{i=1}^{N}(1 - a_i) \le \prod_{i=1}^{N} e^{-a_i} = e^{-\sum_{i=1}^{N} a_i} = e^{-S_p(x)}.$$

Subtracting both sides from 1 reverses the inequality:

$$1 - \prod_{i=1}^{N}(1 - a_i) \ge 1 - e^{-S_p(x)}.$$

Using the product-model definition of $Y_p^{\mathrm{ind}}(x)$, we obtain

$$Y_p^{\mathrm{ind}}(x) \ge 1 - e^{-S_p(x)}.$$

Combining the upper and lower bounds gives, for every target realization x,

$$1 - e^{-S_p(x)} \le Y_p^{\mathrm{ind}}(x) \le \min\{S_p(x), 1\}.$$

Equivalently,

$$1 - e^{-S_p(X)} \le Y_p^{\mathrm{ind}}(X) \le \min\{S_p(X), 1\} \qquad \text{almost surely.}$$

Finally, the product-form team detection function is monotone submodular in the selected set of agents or paths. For fixed x, define

$$F_x(A) := 1 - \prod_{i \in A}(1 - a_i), \qquad A \subseteq \{1, \dots, N\}.$$

For $A \subseteq B$ and $j \notin B$, the marginal gain of adding j to A is

$$\begin{aligned} F_x(A \cup \{j\}) - F_x(A) &= \left[1 - (1 - a_j)\prod_{i \in A}(1 - a_i)\right] - \left[1 - \prod_{i \in A}(1 - a_i)\right] \\ &= a_j \prod_{i \in A}(1 - a_i). \end{aligned}$$

Similarly,

$$F_x(B \cup \{j\}) - F_x(B) = a_j \prod_{i \in B}(1 - a_i).$$

Because $A \subseteq B$ and each factor $1 - a_i$ lies in $[0, 1]$,

$$\prod_{i \in A}(1 - a_i) \ge \prod_{i \in B}(1 - a_i).$$

Thus

$$F_x(A \cup \{j\}) - F_x(A) \ge F_x(B \cup \{j\}) - F_x(B),$$

which is the diminishing-returns property. Hence F_x is submodular. Since the marginal gains are nonnegative, F_x is also monotone. □

The additional bound $Y_p^{\mathrm{ind}} \ge s_p/N$ follows from $Y_p^{\mathrm{ind}} \ge c_{p_i}$ for each i, hence $Y_p^{\mathrm{ind}} \ge \max_i c_{p_i} \ge N^{-1} \sum_i c_{p_i}$.

Proof of Lemma 2

Proof. Integrate the pointwise dominance $Y_p \ge f_p$ against each nonnegative unit-mass π_Θ. This gives $D_p \ge \langle f_p, \pi_\Theta \rangle$ for every scenario. Almost-sure dominance implies ordered quantiles, and integration with $w \ge 0$ preserves their order. For $f_p = s_p/N$, positive homogeneity gives the stated factor $1/N$. □

Proof of Corollary 1

Proof. The closest-agent part of Lemma 1 gives $Y_p \ge \max_i c_{p_i} \ge s_p/N$. Integration against each valid belief gives $D_p \ge U_p/N$. Monotonicity and positive homogeneity of $\rho_{w,Q}$, followed by the quantile moment inequality (25), yield

$$\rho_{w,Q}(D_p) \ge \frac{1}{N}\rho_{w,Q}(U_p) \ge \frac{1}{N}\left(\mu_p - \kappa_w \sqrt{\mathrm{Var}_Q\, U_p}\right) \ge \frac{1}{N} J_{\mathrm{add}}(p),$$

where $\mathrm{Var}_Q(U_p) = \langle s_p, C_Q s_p \rangle \le \langle s_p, C s_p \rangle$. Also $\rho_{w,Q}(D_p) \ge 0$. Both bounds hold uniformly over $Q \in \mathcal{Q}$, proving the clipped bound. Maximizing J_{add} therefore maximizes this certificate, possibly with additional ties after clipping. No comparison of its optimizer with the nonlinear detection optimizer is implied. Combining the corollary with Theorem 1 gives Equation (23). □

D Belief model and sensor implementation

This section specifies the learned belief and TBR implementation used in Sections 3–4 and proves the trajectory and continuation results.

The environmental model gives the fleet a common current forecast and a set of plausible departures from it. These fields transport possible survivor locations; detection evidence then changes their probabilities.

Learning the current field. For each region and forecast split, we select 25 training trajectories and sample $n_{\mathrm{tr}} = 500$ rows with the fixed sampling seed 42. A row contains an unscaled location x_n, scaled to $z_n = x_n / L_\Omega \in [0,1]^2$ by domain width L_Ω, and two current components $\mathbf{y}_n \in \mathbb{R}^2$. The two-output GP has constant mean $\mathbf{m} \in \mathbb{R}^2$ and kernel

$$\mathbf{u}(z) \sim \mathcal{GP}(\mathbf{m}, K_\theta),$$
$$K_\theta((z,i),(z',j)) = B_{ij} \exp\left(-\frac{\|z - z'\|^2}{2\ell_{\mathrm{GP}}^2}\right), \qquad B = bb^\top + \mathrm{diag}(v), \quad v \geq 0.$$

Here $i, j \in \{1,2\}$ index velocity components and ℓ_{GP} is the spatial correlation length in normalized coordinates. The task matrix B has a rank-one factor $b \in \mathbb{R}^2$ plus nonnegative diagonal variances $v \in \mathbb{R}^2$; B itself need not have rank one. The symbol θ collects the fitted model parameters. The spatial kernel links nearby observations. Observations obey $\mathbf{y}_n = \mathbf{u}(z_n) + \varepsilon_n$, with independent Gaussian errors across rows. The implemented multitask likelihood uses task-specific noise and a shared noise variance:

$$\Sigma_\varepsilon = \mathrm{diag}(\tau_1^2, \tau_2^2) + \sigma_n^2 I_2, \qquad \varepsilon_n \sim \mathcal{N}(0, \Sigma_\varepsilon).$$

Stack the two components of each location consecutively in $\mathbf{y} \in \mathbb{R}^{2n_{\mathrm{tr}}}$, and let Z collect their normalized locations. Write $\mathbf{m}_Z = \mathbf{1}_{n_{\mathrm{tr}}} \otimes \mathbf{m}$, $K_{ZZ} = K_\theta(Z,Z)$, and $A = K_{ZZ} + I_{n_{\mathrm{tr}}} \otimes \Sigma_\varepsilon$. The log marginal likelihood is

$$\log p(\mathbf{y} \mid \theta) = -\tfrac{1}{2}(\mathbf{y} - \mathbf{m}_Z)^\top A^{-1}(\mathbf{y} - \mathbf{m}_Z) - \tfrac{1}{2}\log\det A - n_{\mathrm{tr}}\log(2\pi).$$

We fit the mean, kernel, and likelihood parameters with 80 Adam steps at learning rate 0.05, starting from $\ell_{\mathrm{GP}} = 0.15$ and initialization seed 20260916 + split. The implementation minimizes the negative log marginal likelihood per scalar observation; the GP fit remains fixed during a mission.

Conditioning and scenario generation. The latent grid-current posterior used in Section 3 is

$$\mathbf{u}_* \mid \mathcal{D} \sim \mathcal{N}(\mathbf{m}_\mathcal{D}, K_\mathcal{D}). \tag{27}$$

On the $G = 32 \times 32$ planning grid Z_*, with $\mathbf{m}_* = \mathbf{1}_G \otimes \mathbf{m}$, GP conditioning gives

$$\mathbf{m}_\mathcal{D} = \mathbf{m}_* + K_{*Z}A^{-1}(\mathbf{y} - \mathbf{m}_Z),$$
$$K_\mathcal{D} = K_{**} - K_{*Z}A^{-1}K_{Z*}.$$

Here $K_\mathcal{D}$ is the covariance of the latent current field. The sampler uses the GP posterior directly, without adding observation noise at the test locations. It forms the dense covariance, symmetrizes it, and adds $10^{-8}I$ for the Cholesky factorization:

$$L_\mathcal{D}L_\mathcal{D}^\top = \tfrac{1}{2}(K_\mathcal{D} + K_\mathcal{D}^\top) + 10^{-8}I_{2G},$$
$$\mathbf{u}_*^{(s)} = \mathbf{m}_\mathcal{D} + L_\mathcal{D}\xi_s + \sigma_r(\mathbf{1}_G \otimes \zeta_s). \tag{28}$$

The vectors $\xi_s \sim \mathcal{N}(0, I_{2G})$ are independent Gaussian innovations; $L_\mathcal{D}\xi_s$ gives jointly correlated field fluctuations. The final term is a spatially uniform random current offset, with $\sigma_r = 0.15$ selected on development data. For the even scenario counts used here, independent $\zeta_s \sim \mathcal{N}(0, I_2)$ in the first half are paired with their negatives in the second half. Thus the offsets average to zero; the finite set of GP draws need not average exactly to $\mathbf{m}_\mathcal{D}$. Bilinear interpolation supplies each particle's local current. Numerical GP solves allow Cholesky jitter (10^{-3} while fitting, 10^{-6} at grid prediction); the explicit sampler jitter above is separate.

Each field, combined with its wind condition, transports initial-location particles into a scenario belief π_s. Scenarios initially have equal weight; non-detection evidence changes both their relative weights and the particle weights within each scenario. The coordinator also receives the transported particle belief. The GP mean guides its candidate construction; its search picture is not restricted to propagation under that mean alone. The resulting belief covariance differs from the velocity covariance $K_\mathcal{D}$. Its radial-coverage diagnostic was 88.4% at nominal 95%. This marginal diagnostic does not establish the mean or covariance conditions needed for a true-environment detection guarantee.

In simulator coordinates, the target update has the form

$$X_{\tau+1} = X_\tau + 0.3\{u_\Theta(X_\tau) + \bar{w}_\Theta(X_\tau, \tau) + g_\tau(1,1)^\top\} + \eta_\tau,$$

where $g_\tau \sim \mathcal{N}(0, 0.8^2)$ is the shared gust in the two velocity components, and $\eta_\tau \sim \mathcal{N}(0, 0.09^2 I)$ at grid size 300. In the nominal regime, the prescribed wind has direction $\pi/4 + 0.4\sin(\tau/80) + 0.15\sin((x+y)/60)$ and magnitude $1.8[1 + 0.2\sin(x/35) - 0.1\cos(y/45)]$. The controlled-wind study adds one mission-persistent angle, uniform on $[-60°, 60°]$. The planning ensemble represents this distribution without observing the evaluation mission's angle. The nominal wind information is supplied by the simulator, rather than obtained from an operational weather-forecast service.

Finite particle representation. For each fixed planning ensemble, let Λ be the union of its distinct scenario trajectories and set $\nu = \sum_{x\in\Lambda} \delta_x$ on $\mathcal{X}$. Each scenario belief is represented by the mass vector

$$\pi_s(x) = \frac{\sum_{j:x^{s,j}=x} \omega_{s,j}}{\sum_j \omega_{s,j}}, \qquad x \in \Lambda.$$

Thus $\langle Y_p, \pi_s \rangle$ is exactly the weighted-particle detection average. The common support is fixed across candidate plans; no continuous density is assumed for these atomic beliefs. A scenario with zero remaining mass has $\alpha_s = 0$ and may be assigned any unit-mass vector without changing any score. The space $L^2(\mathcal{X}, \nu)$ is finite-dimensional, and every probability vector has squared norm at most one. Its empirical moments are $\bar{\pi}_\alpha = \sum_s \alpha_s \pi_s$ and $C_\alpha = \sum_s \alpha_s (\pi_s - \bar{\pi}_\alpha) \otimes (\pi_s - \bar{\pi}_\alpha)$. The empirical law therefore belongs to the class with this mean and covariance bounded by C_α; its covariance is positive and trace-class. This finite representation establishes the theorem's function-space conditions for the empirical model. It does not establish the same mean, common support, or covariance bound for the true environment. The population theorem retains its stated assumptions for a common reference measure ν.

GP illustration. Figure 2 reproduces Gulf Stream split 1 using the 500-row fit and the registered current sampler (ensemble seed 510001). Panel (a) shows every twelfth observation. Panel (b) shows the latent GP mean and the square root of its pointwise covariance trace, excluding the uniform current offset. Panel (c) bins all 8×512 particles after 180 ticks on a 16×16 grid. Each cell reports its count divided by 4096, without renormalizing the 73.95% remaining inside the domain; the rest have exited. The dashed square is the initial uniform support $[30, 270]^2$. Four paths use fixed particle indices $0, 128, 256, 384$ in the first scenario. Initial particles and motion noise are shared across scenarios; field draws and current offsets differ. All panels divide coordinates by domain width. The later distribution includes initial-location, current, wind, and motion uncertainty and is not a visualization of GP variance alone. The example adds no mission-success result; released arrays and scripts reproduce the figure.

Detection and evidence updates. The sensor depends on the closest agent's distance and is nonincreasing:

$$p_{\mathrm{det}}(d) = \begin{cases} 1 - p_{\mathrm{fn}}, & d \le r_{\mathrm{core}}, \\ \beta(1 - p_{\mathrm{fn}}) \exp\left[-\dfrac{(d - r_{\mathrm{core}})^2}{2\ell^2}\right], & d > r_{\mathrm{core}}. \end{cases} \tag{29}$$

Here p_{fn} is the core false-negative probability, $\beta \in [0, 1]$ sets the tail amplitude, and $\ell > 0$ its width; detection at the exit state is zero. Given a survivor trajectory and fleet routes, draws are independent across ticks. After the executed search ticks $\mathcal{T}$, update unnormalized weights by

$$\omega_{s,j} \leftarrow \omega_{s,j} \prod_{\tau\in\mathcal{T}} \left[1 - p_{\mathrm{det}}(d_\tau(x^{s,j}, p))\right], \qquad \alpha_s = \frac{\sum_j \omega_{s,j}}{\sum_{r,j} \omega_{r,j}}. \tag{30}$$

The normalized scenario belief divides its particle weights by their scenario sum. Zero-weight scenarios contribute nothing to subsequent scores. For hypothetical continuation leg k, write $m_k(x) = \prod_{\tau\in\mathcal{I}_k}[1 - p_{\mathrm{det}}(d_\tau(x, p))]$. Then

$$\omega^{(k)}_{s,j} = \omega_{s,j} \prod_{r<k} m_r(x^{s,j}), \qquad D_{k,s} = \frac{\sum_j \omega^{(k)}_{s,j}[1 - m_k(x^{s,j})]}{\sum_j \omega^{(k)}_{s,j}}. \tag{31}$$

If no mass remains, set $D_{k,s} = 0$; its contribution to sequence value is zero. To see why this update gives the first-detection identity, let $W_{k,s} = \sum_j \omega_{s,j} \prod_{r<k} m_r(x^{s,j})$ be the scenario mass remaining before leg k. For $W_{1,s} > 0$, $S_{k-1,s} = W_{k,s}/W_{1,s} = \prod_{r<k}(1 - D_{r,s})$ is the probability

of reaching that leg without detection. Hence $S_{k-1,s}D_{k,s} = (W_{k,s} - W_{k+1,s})/W_{1,s}$ is its first-detection probability. Summing telescopes to $1 - W_{L+1,s}/W_{1,s}$, the fleet-track detection value in Equation (10). Scenario weights α_s at the actual decision boundary average these complete-sequence values; hypothetical misses enter through the within-sequence conditioning. This construction uses the model's detection-only feedback: a successful search terminates the mission, leaving only the history of misses to continue.

Source data and simulator scale. The released current fields use the Ocean Surface Current Analysis Real-time (OSCAR) product at $1/3^\circ$ spatial and five-day temporal resolution (ESR, 2009). Each region averages snapshots dated 5 January, 5 April, 5 July, and 5 October 2018, interpolates them to the simulation domain, and normalizes peak current magnitude to 1.2. The domain $[0, 300]^2$ is an abstract spatial scale. A tick is one simulator update, with no calibrated conversion to seconds or metres. The factor 0.3 in the transport update makes the maximum current-only displacement 0.36 units per tick; nominal wind magnitude 1.8 contributes about 0.54 before spatial modulation, gusts, and process noise. Thus the experiments retain observed current patterns but do not represent a physically calibrated operational deployment.

At grid size 300, agent speed is 1.44 spatial units per tick, the sensor core radius is 9, and its tail width is 3. The initial target location is uniform on $[30, 270]^2$. These are simulator units. The target enters the absorbing state on leaving the domain. Its remaining weight contributes zero to future detection and stays in the denominator of conditional success. With current particle weights, the numerical scenario payoff for a candidate fleet plan is

$$D_p(s) = \frac{\sum_j \omega_{s,j} Y_p(x^{s,j})}{\sum_j \omega_{s,j}}.$$

Candidate scores share the same trajectory samples. The one-, two-, and three-leg methods use the same sensor and non-detection update. Scores stop at the mission deadline. The sensor has $p_{\text{fn}} = 0.12$ and $\beta = 0.03$; the radius and tail width follow the benchmark's grid scaling. The accelerated implementation truncates the tail beyond $r_{\text{core}} + 8\ell$, omitting at most $\beta(1 - p_{\text{fn}})e^{-32} \approx 3.4 \times 10^{-16}$ detection probability per tick.

Dynamic state, miss transition and expected-success recursion. Condition on the current search history and fixed learned environment model. At a decision boundary, write

$$z_t = \big(q_t, \mathcal{C}_t, \{\alpha_{s,t}, \pi_{s,t}\}_{s=1}^{S}, T - t\big),$$

where q_t contains asset positions and $\mathcal{C}_t$ their outstanding route commitments. A feasible joint action $a \in \mathcal{A}(z_t)$ assigns idle assets while honoring existing commitments. It determines fleet motion until the next decision boundary t^+, no later than the deadline. Let $Y_{t,a}(x)$ be detection along this executed fleet segment for survivor path x. Then

$$d_{s,t}(a) = \langle Y_{t,a}, \pi_{s,t} \rangle, \qquad \bar{d}(z_t, a) = \sum_s \alpha_{s,t} d_{s,t}(a).$$

Equation (4) applies with $M_a = 1 - Y_{t,a}$, $d_s(a) = d_{s,t}(a)$ and $\sum_s \alpha_{s,t}[1 - d_{s,t}(a)] = 1 - \bar{d}(z_t, a)$. It updates both within-scenario path laws and scenario probabilities. These formulas apply when the denominators are positive. A scenario with certain detection has zero posterior weight after a miss; its conditional path law can be assigned arbitrarily. If $\bar{d} = 1$, the miss branch is never reached. The transition $\mathcal{T}_{\text{miss}}(z_t, a)$ also advances asset positions, residual commitments and remaining time. A miss therefore favors both paths and drift scenarios the chosen routes were unlikely to detect.

Under this model, search outcomes are the only new information, motion of assets is known, and successful detection terminates the mission. Conditional on a chosen route and a miss, the next belief is deterministic. Conditioning on these two outcomes gives Equation (9): immediate success has value one, while continuation is weighted by its miss probability. Set $W = 0$ at the deadline; when $\bar{d} = 1$, the continuation term is zero. The supremum is over admissible actions; for a finite action set it is a maximum. This is an optimal value under the specified belief model, not a guarantee under the unknown true current law. It is not computed as an experimental optimal benchmark.

Within a finite horizon, the only observation branch requiring further routes is the sequence of misses. Feasible route sequences along this branch thus represent conditional follow-up decisions

without a branching observation tree. For a fixed fleet track, the remaining value from leg k satisfies $V_{k,s} = D_{k,s} + (1 - D_{k,s})V_{k+1,s}$, with $V_{L+1,s} = 0$. Unrolling gives Equation (10), equivalently $V_{1,s} = 1 - \prod_{k=1}^{L}(1 - D_{k,s})$. Stopping after L legs omits value beyond the planning horizon. A finite menu restricts feasible routes as well: with K choices per leg it can contain up to K^L sequences per agent, so direct joint enumeration considers their Cartesian product.

The implemented sequential insertion is distinct from coordinate improvement of a complete baseline fleet plan. It retains first-leg commitments rather than a common full-horizon incumbent for every agent, and unassigned idle agents have no projected tracks. Consequently, no monotonic-improvement claim for a complete fleet objective is made. Assignment-order sensitivity has not been evaluated here.

Diminishing returns for fixed candidate tracks. Lemma 1 already proves the closest-agent detection set function is monotone submodular for each target path. Fix current beliefs, nonnegative scenario weights, common search ticks and complete committed tracks B. For additional complete tracks S, define

$$F_B(S) = \sum_s \alpha_s \langle Y_{B \cup S}, \pi_s \rangle.$$

Nonnegative averaging preserves the pointwise marginal inequalities in the lemma: for $A \subseteq S$ and $q \notin S$,

$$F_B(A \cup \{q\}) - F_B(A) \geq F_B(S \cup \{q\}) - F_B(S) \geq 0.$$

Thus a route adds less when other tracks already cover its detection opportunities. The full closest-agent proof is in Appendix C; it uses a common uniform threshold across agents at each tick to represent detection as a union of events. It does not assume independent sensors across agents. The result holds for fixed complete tracks and current beliefs, not for a risk-penalized score or the changing projected continuations in TBR. In particular, no $1/2$ closed-loop or local fleet-optimality guarantee is inferred here.

Empirical belief moments and the planning score. With teammate tracks fixed, suppress P_{-i} in $V_s(\boldsymbol{a}; P_{-i})$ and $p(\boldsymbol{a}; P_{-i})$ below. On the common finite support of survivor trajectories, let π_s be each scenario's normalized particle law at the decision boundary. Define

$$\bar{\pi}_\alpha = \sum_s \alpha_s \pi_s, \qquad C_\alpha = \sum_s \alpha_s (\pi_s - \bar{\pi}_\alpha) \otimes (\pi_s - \bar{\pi}_\alpha).$$

Equivalently, for a fleet plan selected by y,

$$\hat{\mu}(y) = \sum_s \alpha_s D_{p(y)}(s), \qquad \hat{\sigma}^2(y) = \sum_s \alpha_s [D_{p(y)}(s) - \hat{\mu}(y)]^2. \tag{32}$$

All operators are ordinary matrices on this support. Since $V_s(\boldsymbol{a}) = \langle Y_{p(\boldsymbol{a})}, \pi_s \rangle$, $\hat{\mu} = \langle Y_{p(\boldsymbol{a})}, \bar{\pi}_\alpha \rangle$ and $\hat{\sigma}^2 = \langle Y_{p(\boldsymbol{a})}, C_\alpha Y_{p(\boldsymbol{a})} \rangle$. The empirical score is therefore the untruncated $J_{\det}$ with $\kappa_w = \kappa$, not an assertion that the bound is tight. For the ambiguity set with these empirical moments, Theorem 1 gives the nonnegative certificate $[\hat{J}_\kappa]_+$. A guarantee for an unknown belief law additionally requires that law to belong to the set. For a lower-tail fraction γ, the moment coefficient is $\kappa = \sqrt{(1-\gamma)/\gamma}$. The coefficient $\kappa = 0.25$ matches a lower-tail fraction $16/17$ under the moment formula; it does not directly optimize the measured worst-quarter wind success.

The Bellman equation applies to expected success ($\kappa = 0$). For $\kappa > 0$, the mean–standard-deviation criterion is generally not recursively separable. TBR evaluates the complete finite-horizon scenario payoff vector, then recomputes the score at the next actual decision boundary. This is a receding-horizon risk preference, not a claim of dynamic risk consistency. The score uses joint detection, including saturation of repeated coverage; it does not optimize the additive sum of individual route payoffs.

Mean-value safeguard for the risk variant. For each first leg a, choose a continuation $b^\star(a) \in \arg\max_{b \in \mathcal{B}_t(a)} \hat{J}_\kappa(a, b)$ and set $m(a) = \hat{\mu}(a, b^\star(a))$. The first-leg selection is Equation (13). The tolerance ε compares the estimated means of these risk-selected continuations. It does not compare with the unrestricted best expected-success continuation, and it does not bound realized mission loss. At $\kappa = 0$ the rule recovers Equation (12); frozen settings are reported in Section 5.

Teammate projections and sequential allocation. At each decision boundary, idle agents are assigned in index order before any asset moves. For focal agent i, the projected teammates are exactly those with a nonempty committed plan, including first legs selected for earlier agents at this boundary. Each follows its remaining commitment, then takes greedy legs to the lookahead horizon; detection scores stop at the mission deadline. To specify this continuation, let z be teammate j's projected endpoint and u the future start time. With current belief $b_t = (\alpha_{s,t}, \pi_{s,t})$, the rule is

$$\phi_j(z, u; b_t) \in \arg \max_{c \in \mathcal{A}_j(z,u)} \sum_s \alpha_{s,t} D_c^{j,u}(s; b_t),$$

where $\mathcal{A}_j(z, u)$ is its feasible leg menu and $D_c^{j,u}$ is detection by teammate j alone on leg c. The endpoint and future time advance across projected legs, but b_t remains the belief at the actual decision boundary: these greedy projections do not apply hypothetical non-detection updates. The resulting teammate tracks are fixed across the focal agent's candidate sequences. Their joint detection with each candidate is then evaluated using the conditional updates above. Unassigned idle teammates have no projected track; once assigned, their first legs become commitments for subsequent agents. Thus the policy can depend on assignment order; the evaluated order is fixed by agent index.

Equation (11) defines the ideal joint choice within the finite scenario representation, route menus and planning horizon. It is not a computed optimal benchmark. At $\kappa = 0$, enumeration maximizes the focal agent's estimated sequence detection with teammate tracks fixed. Detection and miss updates use the retained particles, subject to numerical precision and the sensor-tail cutoff described above. This local optimization does not remove the approximations from finite sampling, route restrictions, limited lookahead, projected teammate actions or sequential fleet assignment. Replanning replaces projected continuations with decisions based on actual search history; it does not supply a global optimality guarantee.

Enumeration and computation. For S scenarios, M particles, and at most K candidate legs per branch, the scoring cost is $O(SMHN \sum_{k=1}^{L} K^k)$, plus teammate projection. With eight headings and depth $L = 3$, exhaustive recursion evaluates $8 + 8^2 + 8^3 = 584$ legs for 512 sequences per agent assignment. The reference uses $H = 60$ and $S \times M = 8 \times 512$. The final quiet-host panel records median route construction of 10.00 seconds per mission on one numerical thread (Appendix I). The local mean tolerance is $\varepsilon = 0.005$. These settings specify the evaluated implementation; they are not assumptions of the population detection bound.

Algorithm 1 Transported-belief assignment at a leg boundary

Require: Shared trajectories and weights, committed routes, menus, L, κ, ε

```
Update shared weights with actual non-detections.
for each idle agent, in index order do
    Project teammates with plans: commitments, then greedy legs on current belief.
    for each feasible first leg a do
        Enumerate continuations, reweighting particles after hypothetical misses.
        Keep b*(a) maximizing Ĵ_κ(a, b); record its mean m(a).
    end for
    Select by Equation (13); commit its first leg before the next agent.
end for
Execute committed routes; repeat using actual evidence at the next boundary.
```

TRAJECTORY INFORMATION AND THE STRUCTURE OF FOLLOW-UP DECISIONS

The following elementary results explain the information retained by the trajectory representation and the continuation structure used by TBR. They do not assert that a particle ensemble is the only adequate representation, or that finite-horizon sequential allocation solves the full fleet problem.

Proof of Proposition 1.

Proof. Consider a two-search-time instance with a fixed feasible route. At each time $\tau \in \{1, 2\}$ choose positions A_τ, B_τ with route detection probabilities $q > r$, respectively, where $0 \le r < q \le 1$.

This requires only a nonconstant distance sensor; in particular, r need not be zero. Define

$$\pi = \tfrac{1}{2}\delta_{(A_1,A_2)} + \tfrac{1}{2}\delta_{(B_1,B_2)}, \qquad \pi' = \tfrac{1}{2}\delta_{(A_1,B_2)} + \tfrac{1}{2}\delta_{(B_1,A_2)}.$$

Both time-τ marginals are $\frac{1}{2}\delta_{A_\tau} + \frac{1}{2}\delta_{B_\tau}$. Conditional independence of sensor draws across times gives

$$\langle Y_p, \pi\rangle = q + r - \tfrac{1}{2}(q^2 + r^2), \qquad \langle Y_p, \pi'\rangle = q + r - qr.$$

Their difference is $(q-r)^2/2 > 0$. After a miss at the first search time, the posterior probability of A_2 is $(1-q)/(2-q-r)$ under π and $(1-r)/(2-q-r)$ under π'. The denominator is positive since $r < q \leq 1$, and these probabilities differ. Thus neither detection nor the future posterior is determined by the time marginals alone. □

The missing information is the coupling of locations across time. A model that retains this coupling through transition laws or a sufficient belief state can also evaluate sequential search; the proposition does not rule out such alternatives. TBR retains it through shared path identities. The example explains why losing that information can matter, but does not predict the effect size of the separate scoring control in Section 5.

Proof of Proposition 3.

Proof. The initial state determines the policy's first action. Detection has no continuation. A miss determines a unique next state and hence the next action. Induction gives the only nonterminal sequence, with arbitrary actions after an unreachable miss history. Conversely, a feasible such sequence specifies the policy on every reachable nonterminal history. By the probability chain rule, survival without detection through leg $k-1$ has probability $\prod_{r<k}(1 - D_{r,s})$. Multiplying by $D_{k,s}$ gives first detection in leg k; these events are disjoint. Their sum is $1 - \prod_{k=1}^{L}(1 - D_{k,s})$, including when an earlier leg detects with probability one. No independence between leg-detection events is required. For the conditional trajectory updates, this value is $D_{p(a_{1:L})}(s)$ by the mass-telescoping identity above: the disjoint segment search ticks together cover $\mathcal{I}_t(p(a_{1:L}))$, so each sensor factor occurs exactly once. □

For completeness, finite action sets also link this identity to the optimal expected-success recursion. Set the continuation value to zero at the deadline. At each preceding decision, randomizing over finitely many feasible actions gives a convex combination of their Bellman values, which cannot exceed the best action. Backward induction therefore supplies a deterministic optimal policy. By Proposition 3, this policy has one continuing sequence of misses. The law of total probability gives its success as $\sum_s \alpha_s V_s = \sum_s \alpha_s D_{p(\boldsymbol{a})}(s)$, with α_s taken at the initial state of this planning problem. Thus, when the decision horizon reaches the deadline, maximizing this mean over all feasible joint miss-contingent sequences yields $W(z)$ for the stated belief model. This argument does not identify the truncated, sequential TBR rule with that full-horizon optimum.

This is a reduction of the observation branches, not of the number of possible action sequences. An unrestricted joint sequence can still require a large search. TBR additionally restricts menus and horizon and fixes projected teammate tracks while assigning each focal agent. New current observations, informative sensor readings beyond detection or a miss, or observed random asset motion would introduce additional branches and fall outside the stated reduction. The mean–standard-deviation preference is still evaluated on the whole sequence, not through a risk-sensitive Bellman recursion.

E Comparison protocols and statistical interpretation

This section supports the method comparisons in Section 5: it defines the information available to each rule, the experimental units, and the scope of the reported intervals.

Evaluation design. We developed the methods on a separate panel and fixed the reported variants before the headline comparison on 180 additional settings: five OSCAR regions, three forecast-training splits, and twelve starting configurations per region and split. Headline missions use three assets on a 300×300 domain for 720 ticks. Each setting integrates expected success over 16,384

simulated survivor trajectories. These trajectories reduce numerical error; starting configurations, rather than individual trajectories, are the experimental units. Confidence intervals resample starts within forecast splits, retain region groups, and adjust within their declared comparison families.

Headline estimates average two independently sampled planning ensembles per setting. Additional frozen studies vary fleet size, domain and mission scale, wind uncertainty, GP construction, and evaluation currents excluded from forecast fitting. Comparisons within each study reuse starts so that method and condition effects are paired.

Method comparisons. All detection planners share the forecast, prior, sensor model, and non-detection updates. TBR and its rollout and tree-search alternatives use 8×512 scenarios/particles, eight headings, and 60-tick legs; the principal coordinator and immediate-detection reference use 32×4096. The coordinator converts the transported belief to a probability map and chooses one leg at a time using coverage and geographic overlap, whereas TBR scores shared survivor trajectories and future search opportunities. Classical and learned comparators retain their specified information and actions; Table 26 records learned-policy inputs and training.

Comparison rules and information

Distance-based coordination. For route waypoint sets $X(p)$ and $X(q)$, define the unnormalized Gaussian overlap

$$k_\sigma(p, q) = \sum_{x \in X(p)} \sum_{y \in X(q)} \exp\left[-\frac{\|x - y\|^2}{2\sigma^2}\right]. \tag{33}$$

The implementation averages sampled waypoint pairs and applies a distance cutoff. Let $\bar{k}_\sigma$ denote that implemented score. For committed teammate routes $\mathcal{C}_t$, define

$$G_t(p) = \sum_{q \in \mathcal{C}_t} \bar{k}_\sigma(p, q). \tag{34}$$

The evolving-belief coordinator minimizes

$$\tfrac{1}{3}\mathcal{N}_t[G_t](p) + \tfrac{1}{3}\mathcal{N}_t[c](p) - \tfrac{1}{3}\mathcal{N}_t[u_t](p), \tag{35}$$

where u_t is posterior route utility, c is route duration, and $\mathcal{N}_t$ maps a term's candidate-menu range to $[0, 1]$, assigning zero when the range vanishes. The weights are fixed. With equal-duration routes, the cost term is constant. Its map is the current weighted histogram of the same transported particles used by the detection planners; cell-decay updates are disabled because the trajectory weights already incorporate search evidence.

Across the configurations, removing the overlap term from the equal-length coordinator changes success by -2.81 pp, adjusted interval $[-3.59, -2.06]$; the earlier component-study estimate is -3.18 pp $[-4.20, -2.13]$. These are ablations of the coordinator's score, not of the rollout. A physical-distance kernel score is distinct from the belief-covariance quantity in Equation 21.

Other comparison methods. The immediate-detection planner maximizes the first-leg scenario-averaged detection probability; its uncertainty-aware variant applies the same mean–standard-deviation rule and tolerance to that one-leg payoff. Synchronized kernel replanning rebuilds joint route menus at common replanning epochs and scores them using physical-distance overlap. Posterior-greedy planning uses the non-transported map to prioritize utility. The unexplored-area heuristic, random walk, and lawnmower sweep use the benchmark's fixed implementations. The main-table lawnmower row is the benchmark's randomized sweep rule, with its recorded randomization and motion-perturbation settings; no best-of-several selection uses these evaluation outcomes.

The belief-conditioned learned comparator is a custom transformer trained with PPO. Its particle-belief representation uses the mean-current forecast, the prescribed wind law, and the initial-location box. It differs from the rollout's correlated-field scenario representation: in particular, it removes escaped particles and renormalizes the remaining mass. The comparison therefore provides common external information through different belief implementations; it does not isolate only learning versus online planning.

The PPO control instead receives TBR's weighted trajectories and scenario weights, fleet state, teammate commitments, search history, and eight-heading leg menu. It uses a learned sequential assignment rule in place of explicit rollout; Appendix K gives its training and evaluation protocol.

Evaluation and statistical protocol

A setting consists of an ocean environment, forecast-training split, and starting configuration. Within a nominal panel, the same 36 starting configurations are evaluated across all five environments, with twelve configurations in each of three splits. For a method difference, we average within each start across its environments, then bootstrap starts within each split and average the split means. The analysis uses 30,000 bootstrap replicates. Bonferroni-adjusted intervals use the number of comparisons in the relevant stage and metric, rather than treating simulated target paths as additional independent experimental units. If m is the recorded comparison count, the percentile endpoints are $0.05/(2m)$ and $1 - 0.05/(2m)$. The original rollout comparison uses $m = 2$, the additional-start and smaller-ensemble confirmations $m = 4$, component removals $m = 12$, and the seed-averaged learned comparison $m = 2$. The PPO policy uses $m = 4$ separately for success, halfway detection, and deadline-assigned detection time. Other family sizes are stated with their results. Adjustments are local to each comparison; they do not give simultaneous coverage over every metric, date, or exploratory stage in the campaign.

Planning trajectories and evaluation trajectories use separate random seeds. Initial target locations use scrambled Sobol samples; Gaussian motion and gust noise generate their subsequent trajectories. The reported number of trajectories is an integration budget, not a count of independent trials. Expected success integrates detection probabilities along the policy's complete route under a history of non-detection. This is valid because a positive detection terminates the mission, and before that event the policy observes the same sequence of negative outcomes. Mean detection time assigns the mission deadline to misses and therefore captures both detection speed and detection probability.

Time conventions. The current headline, horizon, and current-shift summaries align encounter indices with the deadline curves: array index t represents tick $t + 1$. Their capped time equals the recorded miss-inclusive time plus success probability, with the relevant time intervals recomputed after alignment (Appendix F). Breadth panels retain their original index-based time summaries. The convention does not affect success, deadline curves, or computation times.

Interpreting the evidence. Comparison families are local to each study, and descriptive spreads are not confidence intervals. Follow-up status and interval scope are stated with each result; exact sources for Table 1 appear in Appendix F.

F Headline results and configuration-level robustness

This section supports all four outcomes in Table 1. It collects complete method results, regional means, interval provenance, and post hoc configuration-tail analyses.

Complete descriptive comparison

Table 5 reports planning alternatives, controls, classical methods, and learned policies on the 180 configurations. Planning ensembles and training seeds are averaged within each configuration. Capped time assigns 720 ticks to misses; successful-only time conditions on detection.

Table 5: Complete descriptive results across the 180 headline configurations. Success is mean ± SD; capped time assigns 720 ticks to misses; successful-only time excludes misses; the final column averages each method's 45 lowest-success configurations. Learned outcomes average training seeds.

Method	Success (%)	Capped time	Successful-only time	Lowest quarter (%)
Reference and lookahead variants				
TBR	50.14 ± 5.13	467.3	216.8	42.96
Three legs, teammates hold	48.03 ± 4.63	471.5	203.3	41.85
Two legs + teammate model	46.41 ± 4.40	475.4	193.6	40.63
Two-leg rollout	44.66 ± 4.17	481.6	186.9	39.41
Two legs, large ensemble	44.67 ± 4.44	482.2	188.0	38.99
Other planning rules on transported beliefs				
Coordinator	42.59 ± 5.35	496.7	196.8	35.40
Immediate detection	40.77 ± 4.46	498.5	177.2	35.16
Base-policy rollout	50.27 ± 5.62	468.8	220.9	42.57
UCT, 3 legs / 584	49.21 ± 5.04	469.8	212.3	42.21
UCT, 5 legs / 584	50.63 ± 5.76	477.2	240.8	42.50
UCT, 5 legs / 2336	51.50 ± 5.69	470.6	236.4	43.46
Coverage and covariance controls				
TBR, independent coverage	39.44 ± 7.28	527.3	234.6	30.15
TBR, mean-field coverage	39.42 ± 7.25	527.4	234.6	30.18
TBR, particle-product sensor	50.11 ± 5.10	467.5	217.0	42.92
Coordinator, covariance overlap	40.41 ± 5.74	510.8	203.7	32.53
Coordinator, correlation overlap	40.19 ± 5.88	510.4	199.6	32.49
Coordinator, no overlap term	39.78 ± 5.47	513.7	201.4	32.53
Joint covariance assignment, 95% floor	35.04 ± 4.73	537.0	197.7	29.13
Joint additive assignment, 100% floor	29.50 ± 4.47	570.3	212.4	23.86
Classical comparators				
Kernel replanning	22.27 ± 6.72	586.1	119.8	13.63
Posterior-greedy	11.60 ± 4.38	651.5	122.8	6.63
Greedy unexplored	14.69 ± 8.62	636.6	116.9	2.61
Lawnmower	8.58 ± 3.71	668.5	123.7	3.26
Random walk	9.37 ± 8.99	670.9	112.5	0.24
Learned policies				
PPO with TBR information	46.51 ± 5.53	500.8	248.8	38.91
Belief-MAT	25.99 ± 4.71	585.3	202.6	19.65
JointCPPO	22.73 ± 4.77	602.5	204.5	16.18
NeighborPPO	22.72 ± 5.31	599.7	189.4	16.28
MADDPG	20.34 ± 5.17	598.6	124.8	13.32
MATSMI	18.52 ± 5.79	608.0	117.4	10.89

Regional performance and planning-ensemble variation. Table 6 separates regional differences from the pooled headline. TBR exceeds the coordinator in all five regions; five-leg tree search also improves on TBR in all five. Its role as the reference therefore rests on the controlled planning comparisons and detection-time tradeoff, rather than uniformly highest final success among planning alternatives. The two TBR planning ensembles yield 50.26% and 50.02% overall. Their mean absolute difference across the 180 settings is 2.00 pp, despite the small difference in pooled means. These two ensembles quantify the observed sampling variation; they do not establish robustness over arbitrary planning seeds. The regional table averages ensembles within settings and reports descriptive means, without additional regional significance tests.

Average outcomes and difficult configurations

Provenance of the Table 1 intervals. The success column combines recorded comparison families with two marked post hoc contrasts. The static-map coordinator and lawnmower rows retain the baseline-confirmation stage's two families of six success contrasts each: six rules (static-map coordination, posterior-greedy search, unexplored-area search, kernel replanning, random walk, and lawnmower) against two references (TBR and the transported-belief coordinator), with the Bonferroni correction applied separately for each reference. Their saved TBR-minus-method intervals are $[18.48, 22.54]$ and $[40.60, 42.59]$ pp. The transported coordinator and two-step fixed-teammate rows use their recorded three-comparison families; PPO uses four comparisons, Belief-MAT two, and each standard multi-agent policy two per evaluation lattice. The four policies therefore yield

Table 6: Expected mission success by ocean on the headline panel (%). Each row averages 36 settings, with two planning ensembles averaged within each setting. Immediate scores the next leg; base-policy uses greedy continuations. UCT columns give horizon / evaluation threshold. These regional means are descriptive; Table 9 retains the original pooled inference.

Ocean	TBR	Coordinator	Immediate	Base-policy	UCT 3/584	UCT 5/584	UCT 5/2336
Agulhas	49.97	42.19	40.46	49.93	49.31	50.83	51.71
California	53.42	46.29	44.54	53.61	52.45	53.94	54.63
Gulf of Mexico	47.65	39.82	38.45	47.79	46.53	47.92	48.72
Gulf Stream	49.64	42.24	40.10	49.99	48.74	50.01	51.05
Kuroshio	50.03	42.43	40.33	50.02	49.03	50.48	51.39
Mean	50.14	42.59	40.77	50.27	49.21	50.63	51.50

eight contrasts, without a joint eight-way correction. These adjustments are within families, not simultaneous across Table 1. The identical rounded 7.55 pp mean and worst-quarter gains against the coordinator concern different functionals; the MIQP's 7.55 pp deficit is a third, separate comparison.

Configuration-level outcomes. Figure 5 shows the spread of mission success and capped time within each method, complementing the means in Table 1. Each dot averages possible survivor trajectories and sensor outcomes within one configuration; it is not an individual mission. Against the transported-belief coordinator, TBR raises success in 179 of 180 configurations and lowers capped time in 178. Against PPO, the counts are 164 and 180. Both outcomes improve in 178 and 164 configurations, respectively. The regional mean differences favor TBR in all five oceans. These counts describe the saved configurations and are not additional tests.

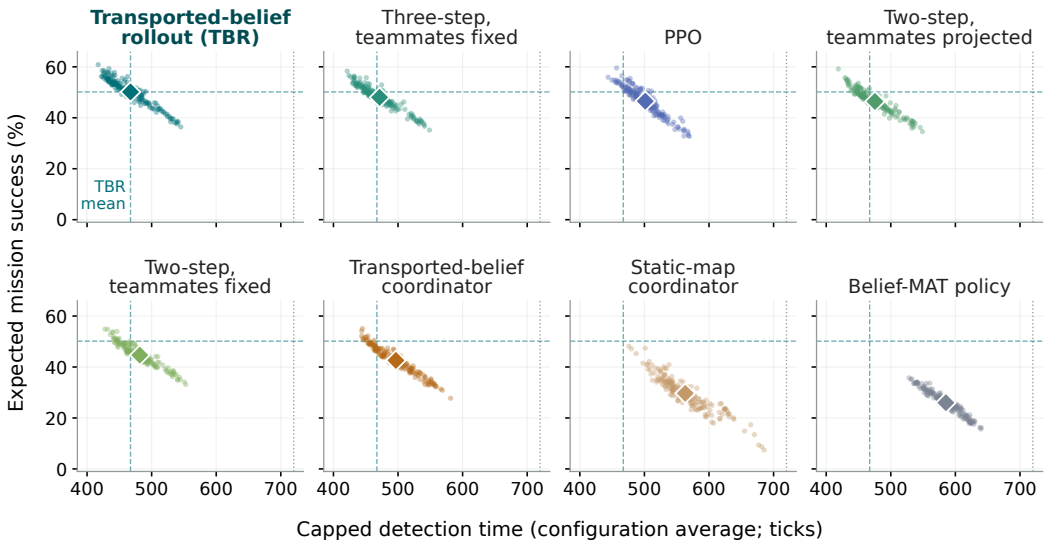


Figure 5: Mission success and capped detection time for the eight methods with the highest average success in Table 1, ranked by average success on identical axes. Each dot averages successful and unsuccessful outcomes within one of the 180 OSCAR configurations; it is not an individual mission outcome. Diamonds give method means; dashed lines mark TBR's mean and the dotted line marks the 720-tick mission deadline. Higher and further left is better. The plot is descriptive; differences and intervals appear in Table 1 and Appendix F.

Difficult searches. A manager also needs to know whether the advantage disappears when rescue is hard. We divide the configurations into four groups of 45, initially using average success of four reference rules—immediate detection, the coordinator, the static-map planner, and kernel replanning—

to rank difficulty. TBR exceeds the coordinator in every group, by 7.86 pp in the hardest and 6.65 pp in the easiest (Figure 6). The direction remains the same under two alternative group definitions. These retrospective summaries describe where the observed gain occurs; their grouping and method means do not constitute additional confirmation tests.

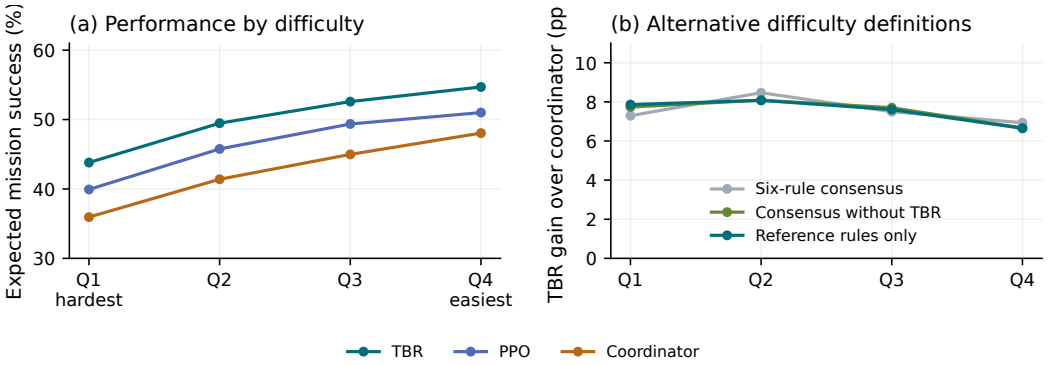


Figure 6: Performance across search difficulty. (a) Settings are ranked by reference-rule success without using TBR or learned policies to form the four groups. (b) The TBR–coordinator gain remains positive when groups instead use a six-rule consensus or that consensus with TBR removed. Each point averages 45 existing configurations; the plots are descriptive.

Table 1 asks two operational questions: how well does a method perform over all deployments, and how well does it perform in the least favorable part of its deployment distribution? All four columns use the same 180 headline configurations. Within a configuration, we first average planning ensembles and, for learned policies, the available training seeds. Thus the tail summaries describe variation across deployments after these averages; they do not measure the worst training seed or planning ensemble.

Success and capped detection time. Success is the probability of detection by tick 720, integrated over the 16,384 truth trajectories. Let $T_{\text{detect}} = \infty$ if the survivor is not detected during the mission. Capped detection time is $\mathbb{E}[\min(T_{\text{detect}}, 720)]$, so an unsuccessful search contributes 720 ticks. This gives credit for both early and eventual detection. If one method's detection probability is at least as large at every deadline, its capped time cannot be higher; a lower capped time alone does not imply dominance at every deadline.

The deadline calculation counts an encounter at array index t at tick $t + 1$. The original time accumulator instead used t. We align the time summaries to the deadline convention by adding each configuration's success probability to its recorded miss-inclusive time, before aggregation or tail selection. Successful-only times increase by one tick. This exact correction uses saved results and requires no new simulation. For the complete discrete deadline curve $F(u) = \Pr(T_{\text{detect}} \le u)$, capped time equals $\sum_{u=0}^{719}[1 - F(u)]$; interpolating the twelve plotted deadlines need not reproduce this sum exactly. The headline descriptive tables, timing figure, and horizon and current-shift time comparisons use the aligned convention. Their original comparison families and bootstrap seeds are retained, with time intervals recomputed from corrected setting-level values. Archived source analyses remain unchanged.

What the worst quarter measures. Worst-quarter success averages a method's 45 lowest-success configurations. Worst-quarter capped time averages its 45 largest capped times. The selected configurations can differ between methods and between the two outcomes. A difference between these tail averages therefore compares two deployment distributions, not performance on a common set of 45 missions. These summaries differ from the wind-risk evaluation, which takes the lowest quarter of wind outcomes *within* a configuration before averaging configurations. The present summaries describe deployment breadth; they are neither worst-case guarantees nor a calibration test of the distributionally robust bound.

Uncertainty in the comparisons. The exploratory analysis uses 30,000 bootstrap replicates. Each replicate resamples the twelve starts within each forecast split, carrying all five ocean regions together and using the same resample for every method. There are 36 starting-configuration groups, not 180 independent observations or $180 \times 16{,}384$ independent trials. We reselect each method's worst quarter in every replicate. The intervals condition on the five evaluated regions, forecast fits, and trained policies.

Table 7 reports unadjusted 95% intervals for the complete set of outcome differences. Table 8 adjusts the four highlighted tail comparisons together: the coordinator and PPO, each on success and capped time. This comparison family was selected after the results were available and remains post hoc. It does not replace the recorded success-comparison families in Table 1, or provide simultaneous coverage across all exploratory comparisons in this appendix.

Table 7: Complete exploratory contrasts for the Table 1 outcomes, including the later planning alternatives. All values are TBR minus the named method: positive success and negative time favor TBR. Success differences are in percentage points and time differences in ticks. Each entry gives the difference and an unadjusted 95% bootstrap interval. These post hoc intervals do not replace Table 1's recorded success-comparison families.

	Average		Worst quarter	
Comparator	Success (pp)	Capped time	Success (pp)	Capped time
Static-map coordinator	+20.46 [+18.94, +22.01]	−95.9 [−103.5, −88.6]	+21.95 [+19.39, +24.80]	−107.3 [−123.4, −93.4]
Transported-belief coordinator	+7.55 [+7.00, +8.11]	−29.4 [−31.8, −27.1]	+7.55 [+6.33, +8.87]	−30.0 [−36.4, −23.5]
Two steps, teammates fixed	+5.48 [+4.68, +6.30]	−14.3 [−18.1, −10.7]	+3.54 [+2.72, +4.74]	−6.0 [−10.0, −2.6]
Two steps, teammates projected	+3.73 [+3.20, +4.29]	−8.1 [−10.7, −5.7]	+2.32 [+1.56, +3.34]	−2.8 [−6.4, +0.2]
Three steps, teammates fixed	+2.11 [+1.71, +2.53]	−4.2 [−6.1, −2.4]	+1.11 [+0.53, +1.89]	−1.0 [−4.4, +2.0]
PPO	+3.63 [+3.10, +4.15]	−34.0 [−37.3, −31.0]	+4.04 [+3.49, +4.53]	−27.4 [−32.3, −22.8]
Belief-MAT	+24.15 [+23.17, +25.15]	−118.0 [−123.3, −112.8]	+23.30 [+21.33, +25.59]	−109.3 [−122.9, −96.9]
JointCPPO	+27.42 [+26.65, +28.21]	−135.2 [−140.0, −130.7]	+26.78 [+25.69, +28.01]	−130.3 [−137.0, −124.5]
Lawnmower sweep	+41.56 [+40.84, +42.32]	−201.2 [−206.3, −196.4]	+39.70 [+38.93, +40.63]	−189.7 [−196.2, −183.6]
Planning alternatives from Table 19				
Base-policy rollout	−0.13 [−0.41, +0.14]	−1.5 [−2.9, −0.05]	+0.39 [−0.02, +0.85]	−2.6 [−4.6, −0.4]
UCT, five steps / 2336	−1.36 [−1.74, −0.97]	−3.3 [−4.9, −1.9]	−0.51 [−1.35, +0.21]	−5.1 [−8.5, −1.4]

Table 8: Family-adjusted 95% intervals. Post hoc tail comparisons highlighted in the main text. Bonferroni adjustment covers these four comparisons together: two comparators and two worst-quarter outcomes. Positive success and negative capped time favor TBR. These are exploratory comparisons, not pre-registered tests.

Comparator	Worst-quarter outcome	Difference [adjusted 95% interval]
Transported-belief coordinator	Success (pp)	+7.55 [+6.00, +9.18]
Transported-belief coordinator	Capped time (ticks)	−30.00 [−37.85, −21.73]
PPO	Success (pp)	+4.04 [+3.34, +4.66]
PPO	Capped time (ticks)	−27.43 [−33.61, −21.57]

Table 9: Family-adjusted 95% intervals. Confirmation success differences in percentage points. Primary comparisons with the smaller two-leg reference use a family of three. Coordinator comparisons are secondary and use the saved three-comparison adjustment; intervals are not simultaneous across all six rows. Every displayed difference is positive in all five regions.

Variant	Reference	Change [adjusted 95% interval]
Three legs + teammate model	Two legs, smaller ensemble	+5.48 [+4.50, +6.48]
Three legs + teammate model	Evolving-belief coordinator	+7.55 [+6.88, +8.24]
Three legs, teammates hold	Two legs, smaller ensemble	+3.36 [+2.60, +4.14]
Three legs, teammates hold	Evolving-belief coordinator	+5.43 [+4.66, +6.21]
Two legs + teammate model	Two legs, smaller ensemble	+1.75 [+1.18, +2.36]
Two legs + teammate model	Evolving-belief coordinator	+3.82 [+3.05, +4.58]

G WHY COORDINATION WORKS

This section supports Figure 3: lookahead and teammate controls are followed by shared-trajectory scoring controls and the separate additive-assignment test. The one-leg removals use a different reference and panel; their effects are not a decomposition of the headline gain.

LOOKAHEAD AND TEAMMATE PROJECTION

Staged design and confirmation. The depth and teammate variants were specified in Addendum 9 of the recorded experiment plan before their original-setting outcomes existed. All use 8 scenarios and 512 particles per scenario. The two factors are depth ($L = 2$ or 3) and teammate continuation (hold position or choose greedy one-leg continuations). The latter choices use the current belief, each teammate's individual detection score, and the eight-heading menu; the projected tracks are then fixed while candidate sequences are assessed. This is a model of later teammate movement, not a joint optimization of future fleet actions. The default three-asset, grid-300, two-leg implementation reproduced the saved reference routes exactly before the breadth runs.

On the original settings, success is 45.86% for the smaller two-leg reference, 49.09% with depth alone, 47.57% with the teammate model alone, and 52.34% with both. Addendum 9b fixed the confirmation comparisons before any variant confirmation outcomes. The 180 confirmation settings use starting seeds beginning at 765000, with planning seeds 510001 and 510002 averaged within each setting. The original variants were selected on starting seeds beginning at 764000. The confirmation set had previously been used for the two-leg family; it is additional-start evidence for the fixed new variants, not an untouched test set for every choice made during the entire campaign. The plan records specify the sequence of choices and evaluations; they are local timestamped experiment specifications, not an external preregistration.

The primary family comprises the three variant-minus-two-leg comparisons. All pass the recorded criterion: an adjusted lower confidence bound above zero and a positive difference in at least four of five oceans. The secondary comparisons against the coordinator and full-ensemble two-leg planner are reported with the saved adjustment for three variants. This adjustment applies within each reference family; it does not provide simultaneous coverage across all families, metrics, or stages. Table 9 gives the principal saved intervals. The expected-success three-leg planner with the teammate model is the resulting lead method; earlier two-leg and risk results retain their original identities.

COVARIANCE OVERLAP AND JOINT DETECTION DIAGNOSTICS

For a route flown alone in drift scenario s, let $d_{p,s}$ be its numerical detection probability and $\bar{d}_p = \sum_s \alpha_s d_{p,s}$. The covariance substitution uses

$$\widehat{C}(p,q) = \sum_s \alpha_s (d_{p,s} - \bar{d}_p)(d_{q,s} - \bar{d}_q)$$

as the coordinator's pairwise overlap; its correlation version divides by both route standard deviations when nonzero. All other normalized utility, cost, and overlap weights are unchanged. The score concerns Lagrangian survivor beliefs but is not the direct score J_{det} or the full additive standard deviation in J_{add}.

Table 10: Descriptive results. Changing only the normalized coordinator's overlap score. Nominal rows use the additional-start settings; wind rows use the separate controlled-wind settings. Tail values refer to wind outcomes, not the lowest-success starting locations.

Condition	Overlap score	Mean (%)	Wind lower quarter (%)
Nominal	Geometric	42.59	–
Nominal	Covariance	40.41	–
Nominal	Correlation	40.19	–
Uncertain wind	Geometric	37.75	31.37
Uncertain wind	Covariance	34.75	29.90
Uncertain wind	Correlation	35.82	30.62

Table 11: Descriptive results. Coordination controls. Nominal and wind studies each have 180 settings; nominal results average two planning ensembles. The wind column uses a separate starting panel. Tail success is the average over the worst quarter of wind offsets within each setting.

Scoring rule	Nominal mean (%)	Wind mean (%)	Wind tail (%)
Joint detection	50.14	41.42	33.17
Shared trajectories, separate sensors	50.11	41.35	33.03
Separate average route coverage	39.44	30.76	21.97
Also average drift scenarios	39.42	30.55	21.80

Covariance penalties address a different interaction. The tested pairwise covariance substitution lowers coordinator success from 42.59% to 40.41%, a loss of 2.19 pp $[-2.78, -1.60]$ relative to geometric overlap. Under wind it loses 2.99 pp $[-3.66, -2.35]$ in mean and 1.48 pp $[-2.49, -0.54]$ in lower-quarter success. Correlation overlap also loses: 1.93 pp $[-2.50, -1.39]$ in wind mean and 0.75 pp $[-1.46, -0.003]$ in the wind lower quarter. A favorable single setting was inspected before specifying the original covariance study; the aggregate and additional-start outcomes do not support its superiority. These outcomes delimit this heuristic, not the validity of the population detection bound.

Convex joint-assignment formulation. The convex MIQP chooses one route per asset while minimizing covariance exposure subject to coverage and budget constraints. Table 5 reports the exact $8^3 = 512$-assignment check: a 95% coverage floor reaches 35.04% success versus 29.50% at 100%, but remains 7.55 pp below the coordinator. The gap shows the practical cost of replacing nonlinear joint trajectory coverage with additive route values. Corollary 1 provides the certificate, and Appendix O gives the low-rank conditions.

Tables 11 and 12 give the absolute outcomes and successive contrasts underlying Figure 3(b).

Separating sensor combination from shared trajectories. The reference evaluates the closest-agent sensor along each common target trajectory. A new particle-product control instead combines own-route and teammate detection on each of those same trajectories, before averaging. The independent alternative averages own-route and teammate detection separately, then combines their misses. A final alternative also averages teammate detection across drift scenarios. All arms retain the original closest-agent sensor for executed and hypothetical non-detection updates and for mission evaluation; only their candidate-scoring rule changes.

For a fixed candidate and drift scenario s, let a_j and b_j be own-route and teammate detection probabilities on the same survivor trajectory j. Write $\mathbb{E}_{j|s}$ for the average using its current normalized particle weights. The diagnostic scores obey

$$\begin{aligned} D_{\mathrm{pp},s} &= \mathbb{E}_{j|s}[1 - (1 - a_j)(1 - b_j)], \\ D_{\mathrm{ind},s} &= 1 - (1 - \mathbb{E}_{j|s} a_j)(1 - \mathbb{E}_{j|s} b_j), \\ D_{\mathrm{pp},s} - D_{\mathrm{ind},s} &= -\operatorname{Cov}_{j|s}(a_j, b_j). \end{aligned} \tag{36}$$

Table 12: Family-adjusted 95% intervals. Success changes (pp) from successive changes to fleet scoring. Each interval uses the recorded within-stage comparison adjustment. The sensor control retains shared trajectory identities; separate averages discard them; the last change additionally averages teammate coverage across drift scenarios.

Change	Outcome	Effect [adjusted 95% interval]
Sensor combination	Nominal	-0.03 [-0.11, +0.04]
Sensor combination	Wind mean	-0.07 [-0.17, +0.02]
Sensor combination	Wind tail	-0.14 [-0.32, +0.03]
Separate coverage averages	Nominal	-10.67 [-12.06, -9.28]
Separate coverage averages	Wind mean	-10.60 [-11.93, -9.33]
Separate coverage averages	Wind tail	-11.06 [-12.49, -9.65]
Also average scenarios	Nominal	-0.02 [-0.12, +0.12]
Also average scenarios	Wind mean	-0.21 [-0.43, -0.02]
Also average scenarios	Wind tail	-0.17 [-0.36, +0.01]

This identity concerns variation over survivor trajectories *within* one drift scenario. The theorem's C_Q concerns variation of the beliefs π_Θ *across* drift scenarios. The diagnostic therefore does not redefine C_Q or measure the isolated benefit of its risk penalty.

Changing the sensor combination while preserving trajectory identity changes nominal success by -0.032 pp $[-0.112, 0.042]$ and wind success by -0.070 pp $[-0.166, 0.017]$. The wind-tail change is -0.137 pp $[-0.322, 0.027]$. These intervals lie within the stated ± 1 pp tolerance. Separately averaging own/team coverage then loses 10.666 pp $[9.282, 12.058]$ nominally, 10.595 pp $[9.330, 11.926]$ under wind, and 11.061 pp $[9.653, 12.490]$ in wind-tail success. All five regions show a loss. In this diagnostic the large effect is therefore primarily associated with shared trajectory information, rather than the tested sensor combination. Route scores must retain which possible survivor movements teammates are likely to cover, so that another agent can cover what they miss.

Further averaging teammate coverage across scenarios changes nominal success by -0.019 pp $[-0.123, 0.124]$, wind success by -0.206 pp $[-0.428, -0.021]$, and the wind tail by -0.169 pp $[-0.361, 0.009]$. Thus the additional across-scenario averaging has a much smaller effect here. These successive interventions have conditional effects and do not allocate the entire headline gain among independent components.

The nominal control uses the $765000+$ headline starts with two planning ensembles; the wind control uses the separate $763000+$ panel. Each contains 180 settings. A small preliminary example was inspected before the local control plan was recorded, following the earlier joint-versus-independent diagnostic. We retain that staged design history and use the complete panels, rather than the preliminary example, for the claims above. Figure 1 illustrates shared trajectory coverage within one drift scenario. Across-scenario belief covariance enters the population detection bound separately; the schematic does not depict an empirical covariance-overlap gain.

Belief components in the one-leg reference

Tables 13 and 14 retain the separate one-leg controls behind the belief-transport and update claims. The reference plans only the next assignment. Removing transport costs 22.29 pp $[20.86, 23.74]$ and replacing non-detection conditioning with fixed cell decay costs 8.00 pp $[6.37, 9.64]$. These adjusted effects belong to this one-leg panel, not to the shared-trajectory rollout scoring control above.

Table 13: Family-adjusted 95% intervals. Component changes to the immediate-detection planner on a separate set of 180 settings, with reference success 41.77%. Each change holds the remaining configuration fixed; intervals are Bonferroni-adjusted. Disabling exact non-detection updates substitutes fixed cell decay.

Removed component	Success (%)	Change (pp)	95% interval
Forward propagation	19.48	-22.29	[-23.74, -20.86]
Wind information	28.90	-12.88	[-14.37, -11.20]
Exact non-detection update	33.78	-8.00	[-9.64, -6.37]
Initial-location prior	34.21	-7.56	[-8.69, -6.41]

Table 14: Additional one-component changes to the immediate-detection reference used in Table 13. These are configuration checks, with adjusted intervals, rather than an ablation of the rollout.

Configuration change	Success (%)	Change (pp)	95% interval
Four headings instead of eight	39.19	-2.58	[-4.12, -0.96]
No coherent residual scenario mode	41.70	-0.08	[-0.68, +0.51]
Unnormalized GP coordinates	41.82	+0.04	[-0.50, +0.54]
120-tick legs instead of 60	41.94	+0.17	[-1.27, +1.53]
16×1024 instead of 32×4096	42.18	+0.40	[-0.09, +0.86]

H FLEET, MISSION, AND CURRENT-FIELD SENSITIVITY

This section supports Table 2 and the current-field robustness discussion. It separates fleet and mission changes from frozen-forecast mismatch and GP-training sensitivity; these panels do not form one common factorial experiment.

FLEET SIZES AND MISSION REGIMES

Fleet and grid protocols. The fleet study uses two and five assets at grid 300 with starting seeds beginning at 766000. The base locations are shared across the two fleet sizes, but changing the fleet changes the search problem. Each panel has five oceans, three splits, and twelve starts per split. The grid study uses three assets, starting seeds beginning at 767000, and grid sizes 100 and 500. The forecast model is fitted to the training trajectories for each grid. The simulator scales the horizon to 240 and 1200 ticks, respectively, along with speed, sensor footprint, and process noise. Its wind law is unchanged. These are operating-regime comparisons; neither success differences between rows nor their timing ratios isolate grid resolution.

The initial grid-500 study completed all 15 environment/split tasks.

After the two-leg grid-500 result was inspected, Addendum 9d specified two follow-ups on the same settings: a larger 16×1024 two-leg ensemble and the three-leg planner with teammate modeling at 8×512. The larger ensemble changes success by -0.23 pp $[-1.09, 0.58]$ relative to the smaller two-leg reference and by -0.52 pp $[-1.74, 0.74]$ relative to the coordinator. The combined planner improves on those references by 4.52 and 4.23 pp. Addendum 9e evaluated the combined planner at both fleet sizes, grid 100, and under wind uncertainty. Addendum 9f subsequently completed the depth-by- teammate comparison on each breadth panel. At grid 500, depth alone adds 2.03 pp $[1.24, 2.81]$ and the teammate model alone 1.30 pp $[0.50, 2.03]$. At grid 100, depth alone is unresolved, while the teammate change is small and positive. Table 16 gives all panels. The wind row compares the smaller new variants to a full-ensemble two-leg reference, so its component differences do not hold the simulation budget fixed. These follow-ups use already-examined settings and are sensitivity studies.

Operating breadth in mission outcomes. TBR's success increases from 42.06% with two assets to 64.18% with five; its gains over the coordinator are 7.36 and 9.06 pp, compared with 7.55 pp for the three-asset headline. This does not establish computational scaling to larger fleets. Across mission regimes, TBR reaches 13.12% versus 10.27% in the smallest and 59.85% versus 55.62% in the largest. Belief transport adds 11.68, 12.91, and 12.90 pp for two, three, and five assets, respectively,

Table 15: Descriptive results. Mission success (%) across fleet and operating-regime studies; 180 settings per row. The three-asset, grid-300 row uses the configurations and averages two planning ensembles for the belief-based rules. Other rows use separate breadth starts and one planning ensemble. Grid changes also scale horizon, speed, sensing, and motion noise; the wind law is unchanged.

Setting	TBR	Two legs	Transported coordinator	Static-map coordinator	One leg
2 assets, grid 300	**42.06**	36.87	34.70	23.03	32.83
3 assets, grid 300	**50.14**	44.66	42.59	29.68	40.77
5 assets, grid 300	**64.18**	57.82	55.12	42.22	53.56
3 assets, grid 100	**13.12**	12.91	10.27	9.35	11.47
3 assets, grid 500	**59.85**	55.33	55.62	46.88	54.14

Table 16: Family-adjusted 95% intervals. Success changes (pp) from adding depth, the teammate model, or both, relative to each panel's two-leg reference. Intervals adjust within the three-comparison family. The wind reference uses a larger ensemble, so that row is not a fixed-budget factorial.

Panel	Third leg	Teammate model	Both
2 assets	+3.75 [+3.08, +4.42]	+1.11 [+0.44, +1.75]	+5.19 [+4.34, +6.00]
3 assets	+3.36 [+2.60, +4.14]	+1.75 [+1.18, +2.36]	+5.48 [+4.50, +6.48]
5 assets	+2.62 [+1.70, +3.62]	+2.20 [+1.43, +3.03]	+6.36 [+5.45, +7.27]
Grid 100	+0.02 [-0.13, +0.18]	+0.13 [+0.02, +0.24]	+0.21 [+0.05, +0.40]
Grid 500	+2.03 [+1.24, +2.81]	+1.30 [+0.50, +2.03]	+4.52 [+3.71, +5.27]
Uncertain wind	+0.94 [+0.48, +1.41]	+0.33 [-0.05, +0.72]	+1.48 [+0.85, +2.11]

but only 0.92 pp in the shortest missions. There, adding a third leg with teammates held fixed yields 0.02 pp $[-0.13, 0.18]$. Figure 7 and Table 15 provide the absolute outcomes; the following table reports the component controls.

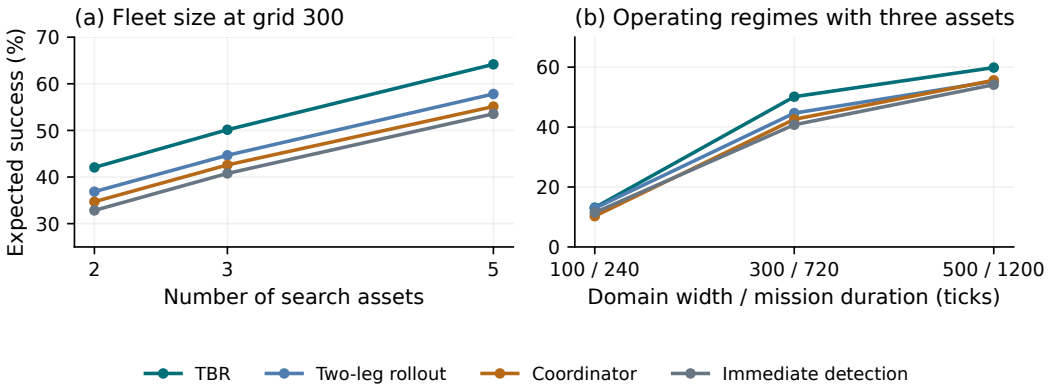


Figure 7: Fleet and mission breadth. (a) Two, three, and five assets on the 300×300 domain. (b) Three assets across domains and mission durations. Each configuration has 180 settings. The three-asset, grid-300 result uses two planning ensembles on the headline starts; other configurations use separate breadth starts and one ensemble. Domain changes also alter speed, sensing, and motion noise. "Two-leg rollout" holds teammates in place beyond their current commitments.

Detection-time and computation details for these regimes appear in Table 23.

Current fields excluded from fitting and selection

The current-transfer study completed all 45 environment–split–date tasks. It reuses the fitted 2018 current models and fixes the planning ensembles, residual-offset recipe, menu, sensor, and planner

settings. The truth fields are independently retrieved OSCAR snapshots from 2014. The requests and cached filenames use 5 January, 5 May, and 5 September; the server composite timestamps stored inside the files are 6 January, 7 May, and 6 September. The manuscript reports these actual composite dates. The fields use the public spatial interpolation and the 2018-mean scale factor, without rescaling each new field to the old peak speed.

Each date uses five regions, three forecast splits, twelve starts, two planning ensembles, and $16{,}384$ truth trajectories per setting. The starts begin at seed 770000 and were not used for the headline panel. The same starts recur across dates, giving 540 date–setting evaluations per arm and 4320 route evaluations overall. Every route is unchanged across dates because the forecast and the no-detection execution path remain fixed; only the evaluation survivor dynamics change.

The three pooled success contrasts form the recorded primary family. Their adjusted intervals compare TBR with the coordinator ($+7.61$ pp, $[6.87, 8.41]$), base-policy rollout ($+0.29$ pp, $[-0.16, 0.72]$), and five-leg UCT at threshold 584 (-0.32 pp, $[-0.81, 0.18]$). The latter two intervals lie inside the specified ± 1 pp equivalence range. Resampling retains repeated date evaluations within starting-setting groups. The corresponding detection-time contrasts form a separate family; all favor TBR (Table 17).

Table 17: Pooled current-transfer contrasts, TBR minus the alternative, with adjusted 95% intervals. Success and detection time use separate three-comparison families. Capped time uses the deadline curve's $t + 1$ convention with misses at the deadline; negative values favor TBR. Time intervals are recomputed after alignment, retaining the original families and resampling seed. The same 180 starting configurations recur over three dates.

Alternative	Success difference (pp)	Capped-time difference (ticks)
Coordinator	$+7.61$ $[6.87, 8.41]$	-29.41 $[-33.17, -25.93]$
Base-policy rollout	$+0.29$ $[-0.16, 0.72]$	-3.91 $[-5.94, -1.94]$
UCT, five legs / 584	-0.32 $[-0.81, 0.18]$	-9.53 $[-11.23, -7.77]$

Pooled success is 51.42% for TBR and 43.81% for the coordinator, with positive gains on each date (Figure 8). The 540 date–setting evaluations reuse 180 starts and are not independent new starting configurations. The forecasts and routes remain frozen across dates, so this is a current-field mismatch study rather than adaptation.

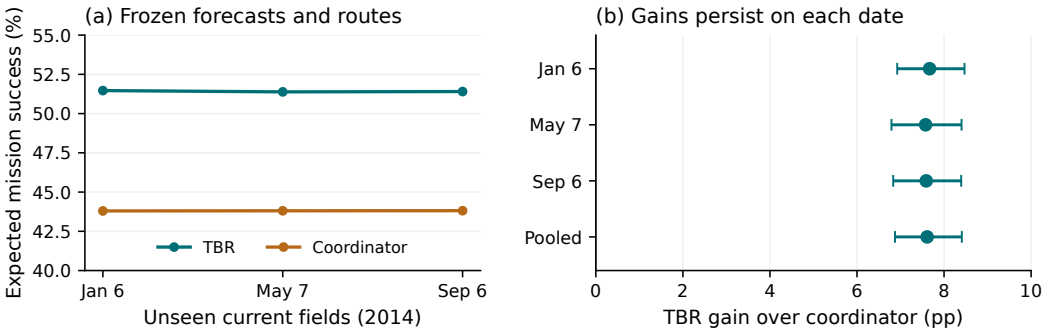


Figure 8: Robustness to current-field shifts with frozen forecasts and routes. (a) Mean mission success on three held-out OSCAR composite dates in 2014. (b) TBR minus coordinator, with the original adjusted 95% intervals. The same 180 new starts recur across dates; pooled inference retains that repetition. Other planning alternatives are reported in Appendix H.

A secondary bootstrap resamples dates and then starts, giving TBR-minus- coordinator $+7.61$ pp $[7.25, 7.99]$. With only three dates, it is a descriptive check and does not justify inference over arbitrary dates. The saved primary adjusted intervals remain the reported evidence. Current-field errors range from 0.16 to 0.38 in simulator velocity units, and success within a setting changes by 1.8 pp on average between dates. The experiment therefore establishes robustness under these stationary

current-field shifts, not calibrated operational forecasts, time-varying currents, or arbitrary forecast horizons.

GP TRAINING AND FORECAST SKILL

Sensitivity to learning and scenario construction. The registered study uses 180 new settings (five regions, three training splits, twelve starts), two planning ensembles, and 16,384 truth trajectories per setting. All 75 tasks completed. Training sets are nested subsets of 250, 500, or 1000 rows; the optimizer and all other settings stay fixed. Separate variants remove or double the spatially uniform scenario offset. Table 18 reports these one-factor changes. Its primary family tests changes in the TBR-minus-coordinator advantage, adjusting over four comparisons; the reference advantage is 7.67 pp, adjusted interval $[6.81, 8.58]$ in the five-recipe advantage family. These are comparisons on a new panel, not additional rows for the headline table.

Table 18: Sensitivity to the learned-current specification on 180 new settings, with two planning ensembles averaged within each setting. The last column is the change in TBR's advantage over the coordinator relative to the reference; intervals adjust over the four registered changes. All lie within the declared ± 1.5 pp range. Successes are percentages, advantages are percentage points.

Recipe	GP rows	Offset SD	TBR	Coord.	Gap	Gap change [95% interval]
Reference	500	0.15	49.96	42.28	7.67	–
Fewer observations	250	0.15	49.85	42.16	7.69	+0.02 [-0.46, +0.50]
More observations	1000	0.15	49.94	42.23	7.71	+0.04 [-0.48, +0.54]
No current offset	500	0.00	49.37	41.05	8.31	+0.64 [-0.10, +1.37]
Larger current offset	500	0.30	49.60	42.03	7.56	−0.11 [-0.78, +0.56]

The four gap-change intervals lie wholly within the predeclared ± 1.5 pp range. The advantage is positive in all five regions for every recipe. Removing the offset lowers TBR success by 0.59 pp, $[-0.96, -0.22]$, and coordinator success by 1.23 pp, $[-1.81, -0.62]$: a larger relative advantage does not mean a better forecast recipe. The GP's mean field RMSE is 0.179, 0.179, and 0.181 velocity units with 250, 500, and 1000 observations. This flat response shows no benefit from more rows under the fixed fitting recipe; it does not isolate model misspecification from optimization or data coverage. Across the 15 reference fits, spatial correlations with the field range from 0.00 to 0.21, and RMSE is 0.99–1.07 times that of a regional-mean predictor. The study supports sensitivity conclusions within these ranges, not calibrated uncertainty or transfer to unseen dates. Its variable runtime is excluded from the quiet-host timing comparisons.

A separate headline-setting diagnostic re-centers each planning ensemble on the registered true current while retaining its residual draws, wind law, menus, and seeds. TBR rises from 50.14% to 51.49%, a gain of 1.35 pp $[0.98, 1.70]$; the coordinator rises by 0.79 pp $[0.34, 1.23]$. These are references for the value of improved current information under this protocol, not upper bounds on achievable search performance.

I PLANNING EFFORT, DETECTION DEADLINES, AND RUNTIME

This section supports Figure 4 and the planning-effort discussion. Final success, early detection, and route-construction cost remain separate outcomes, with each timing panel identified.

PLANNING ALTERNATIVES ON THE TRANSPORTED BELIEF

A shared forecast can support several ways of spending the planning budget. The transported-belief rollout (TBR) enumerates three-leg sequences and projects greedy teammate continuations. The alternatives hold the forecast, 8×512 scenario/particle budget, eight-heading menu, 60-tick legs, closest-agent sensor, and non-detection updates fixed. The base-policy rollout enumerates the first leg and uses greedy immediate-detection continuations for the agent and projected teammates, following the rollout principle (Bhattacharya et al., 2021). The implemented continuation scores use the same root-scenario weights and within-scenario updates as the shared planning code. It is a finite-horizon implementation, not a claim of the general base-policy improvement guarantee.

UCT uses bandit selection (Kocsis & Szepesvári, 2006), a greedy base rollout, and exploration constant 0.05. Its nominal payoff-evaluation thresholds at depth five are 584 and 2336; the smaller threshold equals the number of nodes in a complete three-leg, eight-heading tree, $8 + 8^2 + 8^3$. A completed rollout or initial coverage of all first-leg options can exceed a threshold. Consequently the budget specifies a stopping rule, not an exact wall-clock or universally exact evaluation count. The wall-clock UCT variant and a five-leg base-policy variant were tested on the initial settings only. The additional-start confirmation contains the three-leg base-policy rollout and the two evaluation-threshold UCT variants, each with two planning ensembles.

The confirmation differences from the three-leg reference are $+0.13$ pp $[-0.20, 0.47]$ for the base-policy rollout, $+0.49$ pp $[0.03, 0.99]$ for smaller UCT, and $+1.36$ pp $[0.89, 1.82]$ for larger UCT. Adjustment uses three comparisons within each reference family. The base-policy interval falls within the specified ± 1 pp tolerance. Larger UCT improves expected success, but its capped detection time is 470.6 ticks versus 467.3 for the reference; the smaller UCT value is 477.2 ticks. Expected final success and earlier detection are different objectives. The final timing study records median route times of 2.62, 10.17, and 37.89 seconds for these alternatives, versus 10.00 for the reference. These are complete route-construction times from a separate 15-mission panel, rather than timings over all 180 success-evaluation settings (Appendix I).

Enlarging the two-leg ensemble changes success from 44.66% to 44.67%, with its difference interval inside ± 1 pp (Table 5). Exploring continuations is more useful here than increasing that particle budget.

Table 19 consolidates the final success, halfway detection, and dedicated timing study; a missing time denotes an unmeasured configuration, not zero computation.

Table 19: Descriptive results. Sensitivity to lookahead depth, search strategy, and planning effort. All configurations share the forecast, trajectory ensemble, action menu, sensor, and non-detection updates. Success and halfway detection average the 180 configurations and two planning ensembles. Route times are medians from the separate 15-mission timing study and exclude forecast/setup; a dash denotes an unmeasured configuration in that timing study. Tree-search thresholds limit search effort but are not exact evaluation counts.

Planning configuration	Lookahead segments	Final success (%)	Detected by halfway (%)	Route time (s/mission)
TBR (reference)	3	50.14	42.41	10.00
Greedy-continuation rollout	3	50.27	41.99	2.62
Tree search, threshold 584	3	49.21	42.12	–
Tree search, threshold 584	5	50.63	39.62	10.17
Broader tree search, threshold 2336	5	51.50	41.28	37.89

Separating horizon from search strategy. The follow-up fixes the same inputs and compares exhaustive TBR with UCT at depths three and five, both at threshold 584. It reuses the 180 headline settings and averages planning ensembles 510001 and 510002 within each setting. Truth integration uses $16{,}384$ trajectories. The exploration constant remains 0.05; the two contrasts were declared before execution. This is a targeted follow-up on examined settings, not a new-start test of the method family. Bonferroni adjustment covers the two contrasts within each metric, with the same start resampling and five-region groupings.

Three-leg UCT reaches 49.21%, 0.93 pp $[0.49, 1.37]$ below TBR, with lower success in all five regions. The difference interval is not contained in the declared ± 1 pp equivalence range. Extending UCT to five legs raises success by 1.42 pp $[0.93, 1.91]$, in all five regions. These comparisons support enumeration at three legs and a benefit from deeper search within this UCT implementation. They do not establish a general ranking of search algorithms across horizons or tuning choices.

Final success and earlier rescue. The three-leg UCT control increases capped detection time by 2.48 ticks $[0.56, 4.43]$ relative to TBR; extending UCT to five legs adds 7.41 ticks $[5.52, 9.18]$. At half the mission, mean detection is 42.41% for TBR, 42.12% for three-leg UCT, and 39.62% and 41.28% for five-leg UCT at thresholds 584 and 2336. Higher final success can therefore come with fewer early detections, a distinction that matters for time-sensitive rescue.

Table 20: Descriptive horizon and search-strategy results. UCT evaluation counts are observed means and maxima per decision; 584 is TBR's full-tree count. Capped time assigns the 720-tick deadline to misses and uses the deadline curve's $t + 1$ convention. Route times in this shared-host stage are reported separately from the dedicated timing study in Table 22.

Method	Evaluations (mean / max)	Success (%)	Capped time (ticks)	Route (s)
TBR, 3 legs	584 / 584	50.14	467.3	11.65
UCT, 3 legs	587 / 592	49.21	469.8	12.71
UCT, 5 legs	589 / 598	50.63	477.2	11.69

Table 21: Post hoc bootstrap timing comparisons from saved headline routes, not part of the frozen analysis. Differences are TBR minus the alternative: positive halfway detection and negative capped time favor TBR. Time uses the deadline curve's $t + 1$ convention, with failures assigned tick 720. All eight intervals adjust together; resampling retains region groups and averages the two planning ensembles within each setting. This retrospective analysis does not change the original final-success endpoint or its intervals.

Alternative	Halfway detection (pp)	Capped time (ticks)
Base-policy rollout	$+0.42$ $[+0.02, +0.80]$	-1.47 $[-3.44, +0.52]$
UCT, 3 legs / 584	$+0.29$ $[-0.31, +0.88]$	-2.48 $[-4.94, -0.16]$
UCT, 5 legs / 584	$+2.78$ $[+2.23, +3.31]$	-9.89 $[-11.87, -7.87]$
UCT, 5 legs / 2336	$+1.13$ $[+0.63, +1.62]$	-3.34 $[-5.56, -1.35]$

A retrospective analysis of the saved routes compares TBR with base-policy rollout and all three UCT alternatives at the halfway deadline and on deadline-assigned detection time (Table 21). The eight contrasts form one Bonferroni-adjusted family, with 30,000 resamples of starting-setting groups within splits and two ensembles averaged within each setting. TBR's halfway advantage over five-leg UCT at threshold 584 is 2.78 pp $[2.23, 3.31]$; the advantage over three-leg UCT is 0.29 pp $[-0.31, 0.88]$. The halfway comparison was motivated by the observed curves and review, so these intervals are exploratory, not independent confirmation. Original final-success comparisons retain their families and intervals. Time comparisons use the aligned $t + 1$ convention, recomputed with their original families and resampling seeds (Appendix F). No cutoff was selected by a new sweep.

Implementation check. Four depth-three tasks initially stalled because search could revisit fully expanded subtrees without making progress toward its evaluation threshold. The repair tracks closed subtrees and selects open children. All three arms were rerun; six pre-repair task files are archived and excluded. The depth-five results at threshold 584 reproduce the earlier rows exactly. A further audit at threshold 2336 likewise reproduces all 360 rows (180 settings, two ensembles), including success, detection time, route hashes, and deadline curves. Its measured count averages 2343 evaluations per decision, with maximum 2360. We retain the original adjusted success intervals for both depth-five comparisons; reproducing the same rows is an implementation audit, not additional independent evidence.

Dedicated runtime measurements

Table 19 combines headline success with the dedicated route timings; Table 22 gives the preparation, setup, and route costs separately.

The final timing study uses five oceans and three starts per ocean with one numerical thread. It directly times complete route construction. The 15 missions differ from the 180 headline success settings, so the main frontier combines two studies. Median route times are 0.65 s for the coordinator, 1.46 s for two legs with teammate modeling, 2.62 s for three-leg base-policy rollout, 10.00 s for the three-leg reference, and 10.17/37.89 s for the two UCT thresholds. Observed maxima and setup costs are in Table 22.

The shared forecast-preparation measurement is 3.52 s and bundles the requested target variants; scenario setup is measured separately.

Table 22: Final timing study across five oceans and three starts per ocean. Values are medians except for maximum route time. The GP column is shared forecast preparation; route construction excludes preparation and setup.

Planner	Shared GP (s)	Setup (s)	Route median (s)	Route max (s)
Coordinator	3.52	2.13	0.65	1.84
Immediate detection	3.52	2.06	4.24	10.39
Two legs + team model	3.52	0.51	1.46	5.22
Base-policy rollout	3.52	0.51	2.62	2.94
Three-leg reference	3.52	0.50	10.00	19.80
Tree search, 584	3.52	0.50	10.17	10.37
Tree search, 2336	3.52	0.50	37.89	70.35

Table 23: Detection time, with misses assigned the local deadline, and median route-construction time. Horizons are 240, 720, and 1200 ticks at grids 100, 300, and 500; rows represent different mission regimes.

Setting	Method	Time (ticks)	Route (s)
2 assets, grid 300	Ours	516.5	7.05
2 assets, grid 300	Two legs	530.2	1.02
2 assets, grid 300	Coordinator	544.6	0.97
3 assets, grid 300	Ours	466.8	11.44
3 assets, grid 300	Two legs	481.2	1.85
3 assets, grid 300	Coordinator	496.3	0.94
5 assets, grid 300	Ours	383.0	20.26
5 assets, grid 300	Two legs	396.4	2.81
5 assets, grid 300	Coordinator	416.5	1.23
3 assets, grid 100	Ours	216.5	2.53
3 assets, grid 100	Two legs	216.2	0.40
3 assets, grid 100	Coordinator	219.7	0.26
3 assets, grid 500	Ours	641.1	18.87
3 assets, grid 500	Two legs	670.4	2.63
3 assets, grid 500	Coordinator	679.7	1.69

We omit cross-method per-decision counters because their timed regions differ. For the transported-belief coordinator, the stored internal counter ends after histogram construction, before route selection; treating it as full decision latency would understate cost. The outer per-mission route timer includes the complete construction for every policy and is the comparison used here.

Runtime across fleet and mission regimes

Table 23 supports the per-asset computation discussion in Section 5. It uses the earlier breadth timing panel and the recorded index-based detection times, not the dedicated runtime panel above or the aligned capped-time convention in Table 1.

J Downside protection under wind uncertainty

This section supports the mean–tail tradeoff in Section 5 with the final frozen $\kappa = 0.25$ evaluation.

Frozen penalty and forecast-mismatch evaluation

Choosing protection against adverse wind requires specifying an acceptable mean-success cost before evaluating a penalty. We implemented Equation (32) for both two and three legs with projected teammate continuations. Before running new starts, the selection rule considered $\kappa \in \{0.25, 0.5, 1\}$. A value qualified when its estimated mean cost was at most 0.5 pp in both nominal and controlled-wind development panels; among qualifying values, the largest wind lower-quarter gain selected $\kappa = 0.25$. Those development panels had already been used in earlier studies. The selected value,

Table 24: Frozen choices on 180 new starting settings, reused across five conditions. Two ensembles are averaged per setting. The last column is the penalty effect on mean success (pp), with its within-condition adjusted 95% interval. Wind bias and increased gust variability are not represented by the planner.

Condition	Coordinator	$\kappa = 0$	$\kappa = 0.25$	Penalty effect [95% interval]
Nominal	42.82	50.78	50.97	+0.19 [+0.03, +0.36]
Uncertain wind	38.00	41.65	41.44	-0.22 [-0.37, -0.07]
Wind bias $+20^\circ$	42.88	47.76	47.65	-0.11 [-0.43, +0.18]
Wind bias -20°	35.93	40.16	40.04	-0.13 [-0.37, +0.11]
Stronger gusts	42.96	50.63	50.86	+0.22 [+0.03, +0.40]

Table 25: Descriptive results. Wind uncertainty on the final new-start evaluation. Lower-quarter success averages the worst quarter of 64 wind offsets within each setting, then averages settings. Both lookahead depths use the teammate model and 8×512 scenarios/particles.

Method	Mean (%)	Lower quarter (%)
Three legs, $\kappa = 0$	41.65	32.41
Three legs, $\kappa = 0.25$	41.44	33.23
Two legs, $\kappa = 0$	40.39	32.84
Two legs, $\kappa = 0.25$	40.33	33.41
Coordinator, evolving belief	38.00	30.77
Immediate detection	38.69	32.52

comparators, mean-loss tolerance, and evaluation conditions were frozen before the final new-start results.

The final panels use starting seeds beginning at 768000, twelve per split and five regions, with two independent planning ensembles averaged per setting. The three-leg penalty comparison holds depth, teammate model, route menu, and 8×512 belief budget fixed. Nominal evaluation uses the usual prescribed wind. The represented-uncertainty condition draws one persistent wind offset from $[-60^\circ, 60^\circ]$ and gives the planner that distribution, not the realized offset. Two bias conditions rotate the true wind by $+20^\circ$ or -20° while the planner assumes nominal wind. The gust condition increases true gust standard deviation to 1.2 while planning keeps 0.8. Bias/gust outcomes integrate eight blocks of 4096 truth trajectories per setting. The same starts are reused across conditions.

Four primary mean comparisons are adjusted within each condition: penalized versus unpenalized three legs, three versus two legs, unpenalized three legs versus the coordinator, and penalized three legs versus the coordinator. A separate four-comparison family applies to the wind lower-quarter outcome in the represented-uncertainty condition. These intervals do not supply campaign-wide simultaneous coverage. For the bias and gust conditions, the stored lower-quarter summary coincides with the mean because there is no mixture of persistent wind offsets; it is not used as evidence of wind-tail protection. The lowest quarter of starting-setting successes is a different descriptive metric.

With uncertain wind, mean success changes by -0.217 pp $[-0.367, -0.072]$, while lower-quarter success improves by 0.820 pp $[0.437, 1.217]$. The lower-quarter change is positive in all five regions, ranging from $+0.63$ to $+1.26$ pp. The mean interval lies above the predeclared -0.5 pp acceptable loss. The nominal mean change is $+0.193$ pp $[0.030, 0.358]$, and the stronger-gust change is $+0.221$ pp $[0.034, 0.401]$. Both bias-effect intervals cross zero. These results support the selected risk tradeoff. They establish neither a universal best coefficient nor a finite-sample certificate. The two-leg penalized planner reaches 33.41% wind lower-quarter success, descriptively above the three-leg penalized value; that difference is not a primary superiority test.

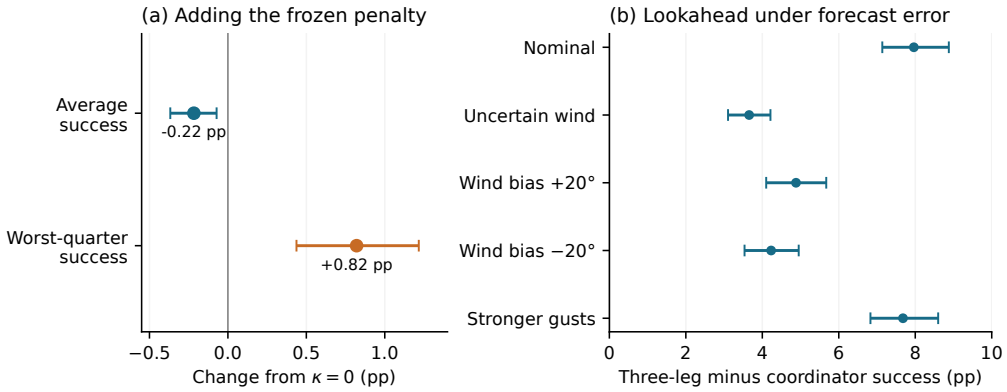


Figure 9: Average performance and adverse wind. (a) TBR with the preselected penalty $\kappa = 0.25$ minus $\kappa = 0$ under represented wind uncertainty. (b) Expected-success TBR minus the coordinator across wind conditions. Each condition uses 180 new starting settings. Intervals adjust within condition and metric families, not simultaneously over the whole figure.

K Learned policies, training, and seed variation

This section supports the learned-policy comparisons in Table 1. It separates validation progress, training returns, and final evaluation outcomes.

Headline learned-policy comparison

The classical and learned comparators were evaluated on the headline starting settings. No sweep variant was selected using those outcomes. Table 5 gives the classical panel with aligned capped times; Table 27 and Figure 11 show the completed five-policy, three-seed results in each ocean region. Overall learned means weight regions equally after averaging training seeds.

The four newly trained families contain 60 completed models: five regions, three seeds, and 2000 training episodes per model. The Belief-MAT panel adds 15 previously trained 500-episode models, for 75 models in total. NeighborPPO and JointCPPO use the published discrete-action implementations with sixteen headings. The published MADDPG and MATSMI implementations use continuous actions; the latter adds the map representation specified in Table 26. They use the current three-asset/grid-300 environment settings and their published training defaults. Belief-MAT has the particle representation described in Appendix E. These observations, representations, and action spaces differ, so this panel compares the evaluated systems rather than changing only the learning algorithm.

A closer learning control gives PPO access to TBR's weighted survivor trajectories and scenario weights, fleet state, teammate commitments, and eight-heading, 60-tick action menu. Three policies are trained in each region for 5,000 episodes over all three forecast splits, using starts and planning and truth seeds disjoint from evaluation. Held-out validation success increases throughout training, with markedly smaller gains near the final checkpoint (Figure 10). Averaged over seeds, the policy reaches 46.51% success, 3.63 pp $[2.93, 4.30]$ below TBR and 3.92 pp $[3.03, 4.81]$ above the coordinator, with the same sign in all five regions. It detects 7.16 pp $[6.32, 8.02]$ fewer survivors by mid-mission than TBR. This comparison supports explicit rollout at the tested training budget; it does not bound what a larger or differently trained policy could achieve. Its roughly 8 ms network decision time excludes shared belief preparation and is therefore not inserted into the separate route-construction timing panel.

Validation progress with the shared belief. Average validation success increases from 14.5% at episode 500 to 48.0% at episode 5000; the final 500 episodes add 0.45 pp. Every run uses its final checkpoint. The validation and headline evaluation starts differ, so the 48.0% and 46.51% means are not the same outcome. The curves show slowing improvement, not proof that further training has no value.

Table 26: Protocol summary; no confidence intervals. Inputs and actions for the evaluated learned policies. Every policy has three training seeds per region. Budgets are episodes per model; the implementations retain their published training defaults and use the current environment settings. Different representations and action spaces are retained.

Policy	Observation / representation	Action	Episodes
PPO	Weighted trajectories and scenario weights; fleet state and teammate commitments	Same eight headings and 60-tick legs	5000
Belief-MAT	512-particle belief; 16×16 map; mean-current forecast	16 options, 30 ticks	500
NeighborPPO	Local patch (radius 2); two neighbor slots	16 headings	2000
JointCPPO	Joint observations with local patches (radius 2)	16 headings	2000
MADDPG	Normalized coordinates of all assets	Continuous 2D action	2000
MATSMI	All-asset coordinates; local exploration map compressed into eight directions	Continuous 2D action	2000

Table 27: Descriptive results. Complete learned-policy results across the configurations: three independently trained models per policy and region, 75 models in total. Ranges show training-seed variation. Belief-MAT uses 500 training episodes per model and the four other policies 2000.

Policy	Region	Mean (%)	Seed range (%)
Belief-MAT	Agulhas	25.24	[18.94, 32.32]
Belief-MAT	California	25.33	[19.36, 29.76]
Belief-MAT	Gulf of Mexico	21.31	[19.13, 23.64]
Belief-MAT	Gulf Stream	28.86	[22.01, 34.34]
Belief-MAT	Kuroshio	29.23	[24.16, 32.25]
JointCPPO	Agulhas	21.68	[20.56, 23.61]
JointCPPO	California	25.48	[24.56, 26.83]
JointCPPO	Gulf of Mexico	23.29	[18.92, 30.37]
JointCPPO	Gulf Stream	20.00	[16.21, 22.29]
JointCPPO	Kuroshio	23.19	[20.32, 25.94]
NeighborPPO	Agulhas	26.45	[22.61, 29.02]
NeighborPPO	California	19.19	[8.79, 34.22]
NeighborPPO	Gulf of Mexico	23.33	[18.61, 27.41]
NeighborPPO	Gulf Stream	22.68	[12.70, 28.73]
NeighborPPO	Kuroshio	21.96	[15.55, 26.13]
MADDPG	Agulhas	20.45	[12.92, 25.77]
MADDPG	California	18.72	[16.57, 19.85]
MADDPG	Gulf of Mexico	21.36	[19.65, 24.10]
MADDPG	Gulf Stream	20.52	[17.60, 23.63]
MADDPG	Kuroshio	20.65	[19.32, 21.69]
MATSMI	Agulhas	16.30	[12.79, 18.29]
MATSMI	California	19.39	[17.56, 22.00]
MATSMI	Gulf of Mexico	19.45	[15.71, 24.08]
MATSMI	Gulf Stream	20.91	[18.68, 22.32]
MATSMI	Kuroshio	16.52	[14.99, 18.46]

The seed-averaged means are 25.99% for Belief-MAT, 22.73% for JointCPPO, 22.72% for NeighborPPO, 20.34% for MADDPG, and 18.52% for MATSMI. Each policy's regional mean lies below the three-leg reference in all five oceans. MADDPG and MATSMI differ from the reference by -29.80 pp $[-30.48, -29.14]$ and -31.63 pp $[-32.36, -30.89]$, respectively. Belief-MAT, JointCPPO, and NeighborPPO differences are -24.15 pp $[-25.30, -23.03]$, -27.42 pp $[-28.33, -26.54]$, and -27.42 pp $[-28.17, -26.67]$. The recorded adjustment applies within each declared comparison family. Intervals resample starts after averaging trained seeds; they condition on these models. The regional seed ranges expose training variation beyond the seed-averaged comparison.

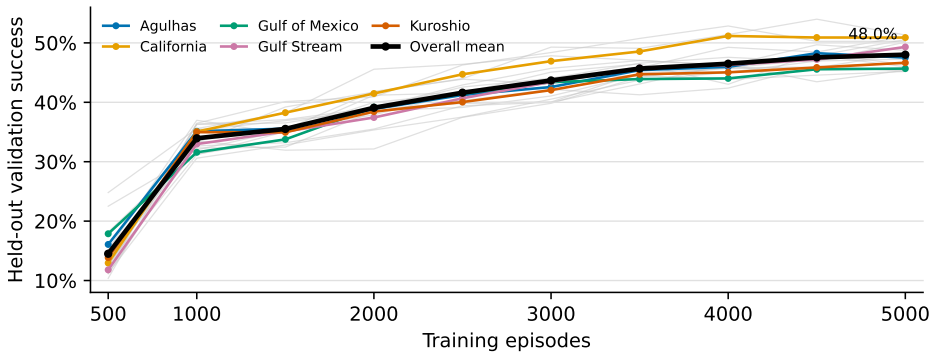


Figure 10: Learning with the planner's information. Thin gray lines show held-out validation success for the 15 region–seed runs; colored lines average three seeds within each region, and black averages all runs. The final checkpoint is used for headline evaluation. These curves measure validation progress, not performance on the evaluation configurations.

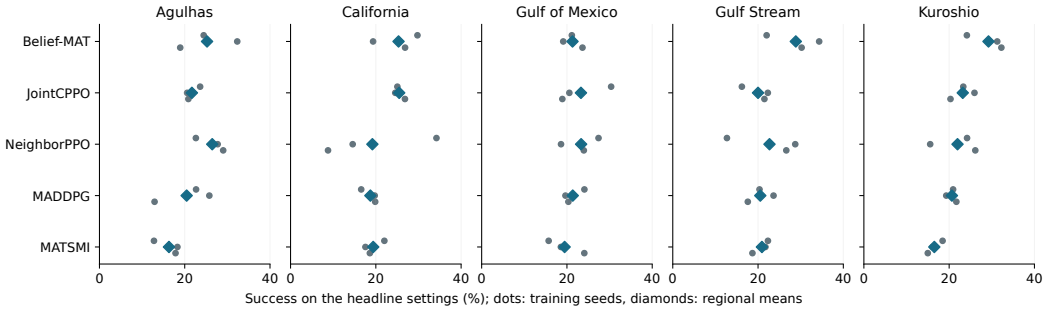


Figure 11: Variation among the other learned search systems. Dots represent three trained models per policy and region; diamonds mark regional means on the configurations. Belief-MAT uses 500 training episodes and the other four policies 2000. Inputs and action spaces differ across these systems; the PPO control is reported separately above.

Training histories and late-budget behavior. Figure 12 uses the saved returns of the same fifteen Belief-MAT, fifteen MADDPG, and fifteen MATSMI models evaluated above. Table 28 compares equal-length windows at the end of each method's budget. The window definitions are descriptive, chosen after training; no convergence test or checkpoint-selection rule is inferred. The native PPO wrappers retained final actors and evaluation outcomes but did not persist their training-return histories. Accordingly no learning curve or convergence claim is supplied for NeighborPPO or JointCPPO. Their final outcomes and seed variation remain in Table 27 and Figure 11. PPO retained all held-out validation histories (Figure 10).

The saved MADDPG and MATSMI returns improve rapidly early and then level off on average; Belief-MAT remains noisy without a sustained late increase. All fifteen runs per method are shown. These are training-return patterns, not convergence claims about held-out mission success.

Table 28: Descriptive late-training changes for the saved baseline histories. Each row averages five regions and three seeds. Return columns compare the penultimate and final quarter of the training budget; reward scales differ by method. Episode success is measured during training, not by headline route integration. Dashes denote unavailable histories, not zero success.

Method	Episodes	Training return		Training success (%)	
		Previous quarter	Final quarter	Previous quarter	Final quarter
Belief-MAT	500	6.18	6.08	–	–
MADDPG	2000	-32.10	-30.05	36.00	35.97
MATSMI	2000	-0.41	-1.55	47.31	45.59

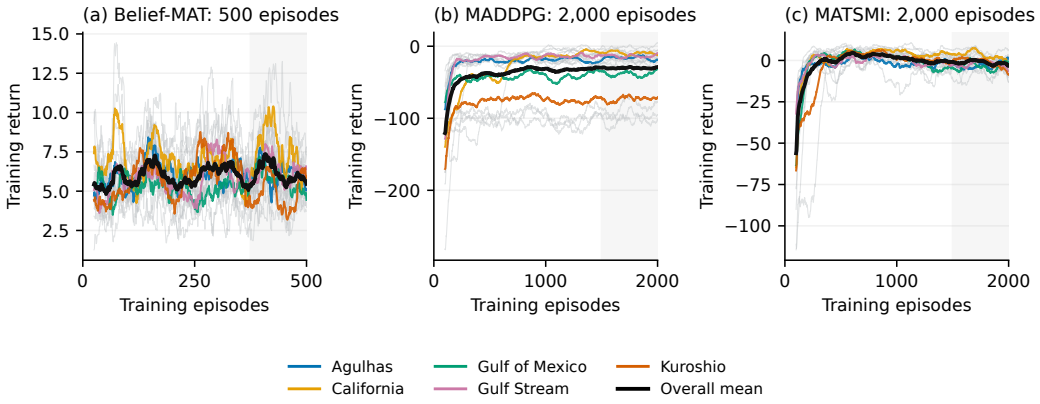


Figure 12: Saved training-return curves for the evaluated Belief-MAT, MADDPG, and MATSMI models. Gray lines are the fifteen runs per method, colored lines are three-seed regional means, and black is the overall mean. Trailing averages use 25 episodes for Belief-MAT and 100 for the other methods; shading marks the final quarter of training. Returns have different reward scales across methods and are distinct from validation or test success.

L Scope and Limitations

This section states the limits of the evidence in Section 5 and of the theoretical guarantees.

The experiments identify three useful planning ingredients: shared survivor trajectories, lookahead, and projected teammate searches. A separate frozen penalty study measures a wind-tail benefit at a small mean cost. The main comparison covers five ocean regions, three assets, and grid 300; additional studies use two and five assets, grids 100 and 500, and controlled wind uncertainty. These selected panels establish breadth across the tested conditions. They do not cover every fleet–grid combination, and changing grid size also changes mission duration, kinematics, sensing, and relative wind influence.

Operational use requires matching the model to the forecasts and search information available to a rescue organization. The planner currently receives a prescribed wind model, a known initial-location prior, and a spatial GP current forecast fitted before the mission. Search outcomes update survivor beliefs; the experiments do not refit that GP from new environmental readings. In the fitted benchmark models, spatial correlation with the training-era fields is only 0.00–0.21 and field RMSE is comparable to a regional-mean predictor. Re-centering the planning ensemble on the registered true current improves TBR by 1.35 pp under the same wind law, so the experiments provide limited evidence for learning fine current structure. New-start bias and gust tests introduce controlled forecast mismatch. The frozen planners also retain their advantage over the coordinator on three 2014 current fields excluded from fitting and selection (Appendix H). The transfer study changes stationary currents while retaining the wind law. Operational wind forecasts and lead-time-dependent prediction still require validation. Ocean reanalysis and synthetic wind also cover only part of the sensor failures, asset constraints, and drift uncertainty encountered at sea.

The detection bound provides a conservative planning criterion when its belief-moment assumptions hold. The GP and transport model supply the implementation's estimated moments. The finite ensemble has not been shown to upper-bound true belief covariance, and its radial-coverage diagnostic falls below the nominal level. Planning further approximates the fleet problem through finite menus, limited lookahead, and sequential assignments with projected teammate behavior. Global fleet optimality is therefore outside the established results.

The experiments clarify which uses of dependence help. Preserving shared trajectories matters strongly in the sensor-control diagnostic; substituting pairwise covariance overlap in the coordinator loses mean and tail success. The joint covariance assignment improves on strict additive coverage after allowing a 5% coverage tolerance, but remains below the coordinator and TBR. The first concerns trajectory dependence within a drift scenario, while the theorem concerns belief covariance across scenarios. The selected mild penalty improves the specified wind-tail outcome on new starts, with a measured mean cost. These results support the tested risk preference without establishing universal tail dominance.

Computation and learning comparisons concern specified systems. Base-policy rollout matches TBR with less computation, whereas deeper UCT improves success. PPO, trained for $5{,}000$ episodes with three seeds per region, remains 3.63 pp below TBR; its held-out gains slow near the final checkpoint. Other learned systems differ in observations, actions, budgets, and representations, so the comparisons characterize these implementations, not a general limit on learning. Final route timing excludes forecast preparation and setup and comes from 15 missions; it characterizes the tested implementation rather than deployment latency.

M THEORY EXTENSIONS: COVARIANCE-AWARE ALLOCATION

These extensions ask how an additive coverage certificate can support structured fleet assignment when larger route menus make direct enumeration costly. They develop spectral coordinates, truncation control, constrained assignment, and replanning constructions, with complete proofs in Appendix W. Their guarantees concern the stated additive models and assumptions; TBR instead scores joint detection on shared trajectories. The static-map scenario-menu construction is also distinct from the time-aligned one-leg assignment control evaluated in Appendix G.

Extension question	Location
How can covariance be represented and truncated?	N, O
How can menus support constrained fleet assignment?	P–R
What changes under alternative learning and local models?	S–U
Where are the full derivations?	W

N MERCER FACTORIZATION DETAILS

This extension factorizes the covariance budget defined in Section 3. The first statement diagonalizes the covariance operator, providing coordinates for the additive assignment model.

For the positive trace-class budget, let $\lambda_1 \geq \lambda_2 \geq \cdots > 0$ index the nonzero modes; if there are finitely many, pad the spectral sums with zeros. The nullspace contributes zero to covariance. In L^2,

$$(Cf)(\tilde{x}) = \int_{\mathcal{X}} K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') f(\tilde{x}')\, d\nu(\tilde{x}'), \tag{37}$$

$$K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') = \sum_{r \geq 1} \lambda_r \psi_r(\tilde{x}) \psi_r(\tilde{x}'). \tag{38}$$

For the pointwise Mercer argument, additionally assume a continuous positive semidefinite kernel on a compact metric realization space with a finite Borel measure of full support. Without those additional conditions, the operator identities below hold in L^2, or as matrix identities on a grid.

Proposition 4 (Mercer eigenfactorization of the Lagrangian covariance operator)**.** *Let C be a positive, self-adjoint trace-class operator on the separable space $\mathcal{H}$. Then C is Hilbert–Schmidt and compact, and its nonzero modes admit orthonormal eigenfunctions ψ_r with nonnegative, summable eigenvalues λ_r. The kernel expansion in* (38) *holds in $L^2(\mathcal{X} \times \mathcal{X})$. If C has a continuous positive-semidefinite kernel on compact metric $\mathcal{X}$ and ν is a finite Borel measure of full support, the expansion also converges uniformly. In either case,*

$$C\psi_r = \lambda_r \psi_r, \qquad r \geq 1. \tag{39}$$

Theorem 2 (Spectral representation of the covariance budget)**.** *Using the covariance eigenfactorization,*

$$\sigma_p^2 = \sum_{r \geq 1} \lambda_r V_r(p)^2, \qquad V_r(p) = \int_{\mathcal{X}} s_p \psi_r \, d\nu. \tag{40}$$

The full proof is in Appendix W.

To make the transport-aware interaction structure explicit at the path level, define the exact pairwise overlap decomposition

$$\sigma_{U_p}^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} k_{\mathrm{Lag}}(p_i, p_j). \tag{41}$$

and the path-level mode loadings

$$v_{p_i}^{(r)} := \int_{\mathcal{X}} c_{p_i}(\tilde{x})\, \psi_r(\tilde{x})\, d\nu(\tilde{x}), \tag{42}$$

together with the fleet-level overlap loading in mode r

$$V_r(p) := \sum_{i=1}^{N} v_{p_i}^{(r)} = \int_{\mathcal{X}} s_p(\tilde{x})\, \psi_r(\tilde{x})\, d\nu(\tilde{x}). \tag{43}$$

For rank R, the corresponding low-rank pairwise overlap is

$$k_{\mathrm{Lag},R}(p_i, p_j) := \sum_{r=1}^{R} \lambda_r v_{p_i}^{(r)} v_{p_j}^{(r)} = \left\langle \tilde{v}_{p_i}, \tilde{v}_{p_j} \right\rangle, \tag{44}$$

so Mercer coordinates factorize the same covariance-induced overlap object rather than replace it.

This interpretation parallels factor-risk decompositions in asset pricing (Ross, 1976; Fama & French, 1993; Cochrane, 2005). The dominant Mercer modes play the part of latent systematic factors, $V_r(p)$ are analogous to portfolio factor loadings, and λ_r are risk weights that penalize concentration in common modes. In SAR, these factors summarize belief variation induced by uncertain transport; they are not the eigenmodes of velocity covariance by definition.

O Low-rank truncation and risk-scaled coordinates

This appendix develops the optional spectral and menu reductions introduced in Appendix N. The first step is the high-mean menu reduction. The reference set and nonempty retained menu are defined in Appendix T.

Proposition 5 (Approximate mean equalization on a high-mean menu)**.** *For every $p \in \mathcal{M}_\varepsilon$, there exists $\delta_p \in [-\varepsilon, 0]$ such that*

$$\mu_{U_p} = \mu^* + \delta_p. \tag{45}$$

Hence, on $\mathcal{M}_\varepsilon$,

$$\mu_{U_p} - \kappa_w \sigma_{U_p} = \mu^* - \kappa_w \sigma_{U_p} + \delta_p, \qquad \delta_p \in [-\varepsilon, 0]. \tag{46}$$

In particular, for any $p, q \in \mathcal{M}_\varepsilon$,

$$\left| \left(\mu_{U_p} - \kappa_w \sigma_{U_p}\right) - \left(\mu_{U_q} - \kappa_w \sigma_{U_q}\right) + \kappa_w (\sigma_{U_p} - \sigma_{U_q}) \right| \le \varepsilon. \tag{47}$$

Thus, within $\mathcal{M}_\varepsilon$, the robust ranking is an ε-perturbation of variance ranking.

Since

$$\sigma_{U_p}^2 - \sigma_{U_p,R}^2 = \sum_{r>R} \lambda_r V_r(p)^2,$$

the bound in (54) directly certifies the omitted variance. By monotonicity of the square-root map on $\mathbb{R}_+$, it also implies

$$0 \le \sigma_{U_p} - \sigma_{U_p,R} \le \sqrt{\lambda_{R+1}}\, \|s_p\|_{\mathcal{H}}, \tag{48}$$

which is the quantity that directly controls the robust-objective gap later in (53).

To pass from fleet-level modal loadings to a path-selection problem, define

$$\tilde{v}_{p_i} := \left(\sqrt{\lambda_1} v_{p_i}^{(1)}, \ldots, \sqrt{\lambda_R} v_{p_i}^{(R)}\right) \in \mathbb{R}^R. \tag{49}$$

Then the rank-R covariance-overlap penalty (equivalently, truncated variance) can be written as

$$\sigma_{U_p,R}^2 = \sum_{r=1}^{R} \lambda_r V_r(p)^2 = \left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2. \tag{50}$$

Thus the truncated robust variance depends only on the aggregate risk-scaled overlap vector in $\mathbb{R}^R$, not on the original infinite-dimensional effort field.

Expanding (50) gives

$$\left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2 = \sum_{i=1}^{N} \|\tilde{v}_{p_i}\|_2^2 + 2 \sum_{1 \le i < j \le N} \langle \tilde{v}_{p_i}, \tilde{v}_{p_j} \rangle. \tag{51}$$

Equation (51) provides the pairwise interpretation of the reduced model: robustness is governed by *risk-weighted covariance-induced overlap*, not by raw geometric overlap alone.

Proposition 6 (Exact robust lattice principle). *Let $\mathcal{M}_{\mathrm{eq}} \subseteq \mathcal{F}_N$ be any retained menu of team plans $p = (p_1, \ldots, p_N)$ such that $\mu_{U_p} = \bar{\mu}$ for all $p \in \mathcal{M}_{\mathrm{eq}}$. For $\kappa_w > 0$, maximizing the rank-R truncated additive surrogate*

$$\bar{\mu} - \kappa_w \sigma_{U_p,R}$$

over $\mathcal{M}_{\mathrm{eq}}$ is equivalent to solving

$$\min_{p \in \mathcal{M}_{\mathrm{eq}}} \left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2. \tag{52}$$

For $\kappa_w = 0$, all equal-mean plans tie; variance minimization is only a tie-breaking rule.

Since the menu $\mathcal{M}_\varepsilon$ from (76) is only approximately equalized, we introduce the following notation for the approximation discussion:

$$B_\varepsilon := \sup_{p \in \mathcal{M}_\varepsilon} \|s_p\|_{\mathcal{H}}, \qquad \mathcal{J}(p) := \mu_{U_p} - \kappa_w \sigma_{U_p}, \qquad \mathcal{J}_R(p) := \mu^* - \kappa_w \sigma_{U_p,R}.$$

The next corollary states the uniform approximation gap for the additive surrogate. A truncated variance alone underestimates the full budget; a reported detection certificate must retain the tail bound (and the factor $1/N$).

Corollary 2 (Approximate robust lattice principle on a high-mean menu). *For every $p \in \mathcal{M}_\varepsilon$,*

$$0 \le \mathcal{J}_R(p) - \mathcal{J}(p) \le \varepsilon + \kappa_w \sqrt{\lambda_{R+1}}\, B_\varepsilon. \tag{53}$$

Corollary 2 marks the transition from functional analysis to combinatorial optimization. Once each path is mapped to risk-scaled Mercer coordinates, the planner no longer optimizes directly over nonlinear team-detection expressions under infinite-dimensional uncertainty; it minimizes a low-rank factorized representation of the same covariance-induced overlap objective. On $\mathcal{M}_\varepsilon$, the truncated lattice selector

$$\min_{p \in \mathcal{M}_\varepsilon} \left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2$$

is therefore a uniform additive surrogate for maximizing $\mathcal{J}(p)$, with explicit gap control from (53). In particular, if $p, q \in \mathcal{M}_\varepsilon$ satisfy

$$\left|\mathcal{J}_R(p) - \mathcal{J}_R(q)\right| > 2\Big(\varepsilon + \kappa_w \sqrt{\lambda_{R+1}}\, B_\varepsilon\Big),$$

then their ordering under $\mathcal{J}_R$ is preserved under $\mathcal{J}$. Appendix P translates this design principle into a convex MIQP over the retained menu, together with exact and scalable algorithms tailored to menu size.

THE OMITTED SPECTRAL TAIL

Corollary 3 (Truncation error bound). *For every feasible plan p,*

$$0 \le \sigma_{U_p}^2 - \sigma_{U_p,R}^2 \le \lambda_{R+1} \|s_p\|_{\mathcal{H}}^2. \tag{54}$$

Corollary 3 makes R an explicit accuracy–complexity knob. Appendix T records the corresponding high-mean menu reduction and risk-scaled coordinate algebra.

The additive optimization objective is

$$\max_{p \in \mathcal{F}_N} J_{\mathrm{add}}(p), \qquad J_{\mathrm{add}}(p) = \mu_p - \kappa_w \sigma_p. \tag{55}$$

Using only rank-R coordinates requires the omitted standard-deviation allowance in a population detection certificate:

$$\inf_{Q \in \mathcal{Q}} \rho_{w,Q}(D_p) \ge \frac{1}{N} \max\left\{0, \mu_p - \kappa_w \left(\sigma_{p,R} + \sqrt{\lambda_{R+1}} \|s_p\|_{\mathcal{H}}\right)\right\}.$$

P SCENARIO MENU CONSTRUCTION AND CONSTRAINED MIQP DETAILS

This appendix retains a scenario-menu construction for the additive covariance selector. Its static-map and footprint features differ from the time-aligned detection score used by the rollout. References below to the implementation describe the earlier scenario-menu machinery; the current experiment implementation is specified in Appendix D. The static percentage-mass maps $U^{(s)}$ below are not TBR's trajectory beliefs; this construction does not describe the evaluated 512-assignment control in Appendix G.

Environment learning and path-centric candidate generation. Based on the statistical layer in Appendix S, Stage I begins by learning environment uncertainty and then propagating target uncertainty under sampled flows. For each flow design, we fit a posterior model of the ambient drift field from trajectory data and write

$$\widehat{u}(\cdot) \sim \Pi_u(\cdot \mid \mathcal{D}_t^{\mathrm{env}}),$$

where Π_u denotes the environment-posterior law. We then draw S posterior realizations

$$\widehat{u}^{(1)}, \ldots, \widehat{u}^{(S)} \sim \Pi_u(\cdot \mid \mathcal{D}_t^{\mathrm{env}}),$$

and convert each realization into a scenario-specific utility map.

On the target side, we use the posterior density π_t from Appendix S, which is approximated by a particle prior propagated under each sampled flow realization and process noise, then mapped to a grid utility field $U^{(s)} \in \mathbb{R}^G$ (with G grid cells). The implementation stores percentage mass: $U_g^{(s)} = 100\pi_g^{(s)}$, up to floating-point rounding, with nonnegative entries and total mass 100. Collecting over $s = 1, \ldots, S$ gives

$$\mathcal{U} = \{U^{(1)}, \ldots, U^{(S)}\}.$$

This shared ensemble is used consistently for candidate generation, cross-scenario evaluation, and empirical Mercer embedding, so the exploitation and robustness surrogates in Stage I are built from the same uncertainty sample set. Given a scenario map $U^{(s)}$, one candidate path per agent is generated by a sequential greedy planner on a time-indexed motion graph. Let $p = (x_0, x_1, \ldots, x_T)$ with $x_t \in \Omega$. At each step, the planner considers eight compass moves and a stay-put action. If the current state is x_t and a candidate successor is x', the local score is

$$\mathrm{Score}(x_t, x') = \frac{R(x'; U^{(s)})}{c_{\mathrm{step}}(x_t, x')} + \xi, \tag{56}$$

where $R(x'; U^{(s)})$ is utility in the footprint centered at x', c_{step} is the one-step motion cost, and ξ is a small tie-breaking perturbation. The step cost is

$$c_{\mathrm{step}}(x_t, x') = c_0 + \lambda_{\mathrm{turn}}\, \mathbf{1}\{\Delta\theta(x_t, x') > \pi/2\}, \tag{57}$$

with base cost c_0 and heading-change penalty term $\Delta\theta(x_t, x')$. The next state is thus chosen greedily as the maximizer of (56), repeated until the planning budget is exhausted.

Within each scenario, the N agent paths are generated sequentially on a shared working map. After each selected step, utility in the visited detection footprint is harvested and then removed from the working map. This discourages immediate revisits by later steps and later agents, providing a lightweight within-scenario anti-duplication mechanism without solving a full joint dynamic program. The global robust diversification decision is deferred to Stage II. For each scenario s and agent i, the planner returns a feasible path $p_i^{(s)}$, its realized utility, travel cost, and visited grid indices. Hence each agent has a raw scenario menu

$$\widehat{\mathcal{M}}_i = \{p_i^{(1)}, \ldots, p_i^{(S)}\}.$$

The intent is not to treat each scenario-generated path as globally robust, but to populate a rich candidate set spanning distinct plausible environment realizations.

Cross-scenario evaluation and empirical Mercer coordinates. Consistent with the path-centric construction above, each raw candidate is evaluated on *all* scenario maps, not only on the map that generated it. Let $\mathcal{I}(p)$ be the union of visited square-footprint grid cells of candidate path p. Its utility under scenario s is

$$u_p^{(s)} = \sum_{g \in \mathcal{I}(p)} U_g^{(s)}. \tag{58}$$

The corresponding Monte Carlo mean utility is

$$\widehat{\mu}_p^{\mathrm{fp}} = \frac{1}{S} \sum_{s=1}^{S} u_p^{(s)}. \tag{59}$$

This decouples discovery from assessment: a path may be found in one realization, but it is kept only if it performs well across the full sampled uncertainty set.

The same ensemble is then used to summarize how each candidate responds to scenario variation. Define

$$\bar{U} = \frac{1}{S} \sum_{s=1}^{S} U^{(s)}, \qquad H^{(s)} = U^{(s)} - \bar{U}.$$

Using these centered maps, we form the scenario Gram matrix

$$K_{ss'} = \frac{1}{S-1} \langle H^{(s)}, H^{(s')} \rangle, \qquad s, s' = 1, \dots, S, \tag{60}$$

and, for each candidate path p, define centered scenario utilities

$$h_p = \left(u_p^{(1)} - \widehat{\mu}_p^{\mathrm{fp}}, \dots, u_p^{(S)} - \widehat{\mu}_p^{\mathrm{fp}} \right)^\top.$$

If $A \in \mathbb{R}^{S \times R}$ collects orthonormal retained eigenvectors of the scenario Gram matrix, the risk-scaled Mercer coordinate is

$$\widehat{v}_p^{\mathrm{fp}} = \frac{1}{\sqrt{S-1}} A^\top h_p \in \mathbb{R}^R. \tag{61}$$

Thus, each candidate carries both a mean-utility score $\widehat{\mu}_p^{\mathrm{fp}}$ and a compact uncertainty signature $\widehat{v}_p^{\mathrm{fp}}$, both derived from the same scenario ensemble.

Alignment with the coverage model. Let $b_p(g) = \mathbf{1}\{g \in \mathcal{I}(p)\}$ be the binary union-of-footprints field. The code's score obeys the exact identity

$$u_p^{(s)}/100 = \langle b_p, \pi^{(s)} \rangle, \tag{62}$$

not, in general, $\langle c_p, \pi^{(s)} \rangle$ for the soft detector in (29). Repeated visits, false negatives, the soft tail, and the square-footprint approximation distinguish the two. The empirical spectral construction is therefore an exact factorization of sample *footprint-utility* covariance (up to truncation and numerical error). It is not by itself a certificate for soft-detector mission success. The coverage-weighted construction would use $\sum_g c_p(g) \pi_g^{(s)}$ in place of $u_p^{(s)}/100$; this revision does not substitute new scores into the reported experiments.

The percentage convention scales means and coordinates by 100, and quadratic covariance terms by 10^4. Probability-scale certificates require the inverse scaling. With $S > 1$ equally weighted scenarios, define

$$\bar{\pi}_S = S^{-1} \sum_s \pi^{(s)}, \qquad C_{\mathrm{emp}} = S^{-1} \sum_s (\pi^{(s)} - \bar{\pi}_S)(\pi^{(s)} - \bar{\pi}_S)^\top.$$

This is the exact covariance of the uniform finite-scenario law. The code's $1/(S-1)$ sample convention corresponds on probability scale to $\widehat{C} = S C_{\mathrm{emp}}/(S-1)$, which upper-bounds C_{emp} and also satisfies $\widehat{C}\mathbf{1} = 0$. It does not automatically upper-bound covariance outside this empirical law. Empirical means, covariance budgets, and Monte Carlo errors must be distinguished when making a population guarantee.

The implemented selector also has separate Gaussian-overlap and spectral families. The OFFLINE KERNEL and synchronized kernel replanner select with averaged physical-distance Gaussian path

pairs and utility/cost terms; the retained scenario coordinates support the spectral ablation. Thus availability of covariance features in a menu does not imply that every selector uses them. The asynchronous policy likewise uses physical-distance overlap. The displayed MIQP describes the ideal spectral objective; its detection interpretation requires the corresponding coverage-weighted features and stated belief model.

High-mean frontier screening. For each agent, Stage I screens the raw menu to a near-best mean-utility frontier. Let

$$\widehat{\mu}_i^{\star,\mathrm{fp}} = \max_{p \in \widehat{\mathcal{M}}_i} \widehat{\mu}_p^{\mathrm{fp}}$$

be the best Monte Carlo mean for agent i. A candidate is retained when

$$\widehat{\mu}_p^{\mathrm{fp}} \geq \alpha \widehat{\mu}_i^{\star,\mathrm{fp}}, \tag{63}$$

with default retention factor $\alpha = 0.95$. Within this near-best subset, dominated mean–cost options are removed, and the menu is trimmed if needed while preserving representative high-utility and low-cost candidates.

Stage I therefore outputs agent-indexed screened menus

$$\mathcal{M}_i^{\varepsilon} = \{p_{i1}, \ldots, p_{iK_i}\},$$

where each retained path is feasible, near-best in mean utility, and annotated with

$$(\mu_{im}, c_{im}, \tilde{v}_{im}).$$

Constrained MIQP extension. Proposition 7 gives a model for the truncated additive covariance surrogate that can be solved with modern exact optimization technology. To incorporate nominal performance and resource usage, let μ_{im} denote the estimated mean additive coverage utility of path p_{im} and let c_{im} denote its travel/resource cost.

To retain the high-mean logic in realistic finite samples setting, we can further impose a fleet-level utility floor and a resource budget cap. Let

$$U_{\mathrm{ref}} := \sum_{i=1}^{N} \max_{1 \leq m \leq K_i} \mu_{im} \tag{64}$$

be the nominal reference value of the screened menus, and let $\alpha \in (0, 1]$ be a retention fraction. The utility floor is then

$$\sum_{i=1}^{N} \sum_{m=1}^{K_i} y_{im} \mu_{im} \;\geq\; \alpha U_{\mathrm{ref}}. \tag{65}$$

Given a fleet resource budget B is enforced, we also enforce

$$\sum_{i=1}^{N} \sum_{m=1}^{K_i} y_{im} c_{im} \;\leq\; B. \tag{66}$$

The resulting operational model is

$$\begin{aligned}
\min_{y} \quad & \left\| \sum_{i=1}^{N} \sum_{m=1}^{K_i} y_{im} \tilde{v}_{im} \right\|_2^2 \\
\text{s.t.} \quad & \sum_{m=1}^{K_i} y_{im} = 1, && i = 1, \ldots, N, \\
& \sum_{i=1}^{N} \sum_{m=1}^{K_i} y_{im} \mu_{im} \geq \alpha U_{\mathrm{ref}}, \\
& \sum_{i=1}^{N} \sum_{m=1}^{K_i} y_{im} c_{im} \leq B, \\
& y_{im} \in \{0, 1\}, && i = 1, \ldots, N,\ m = 1, \ldots, K_i.
\end{aligned} \tag{67}$$

In matrix form, stacking the variables y_{im} into a vector $y \in \{0,1\}^K$ with $K = \sum_i K_i$, one obtains

$$\begin{aligned} \min_{y} \quad & y^\top Q y \\ \text{s.t.} \quad & Ay = \mathbf{1}, \\ & \mu^\top y \geq \alpha U_{\text{ref}}, \\ & c^\top y \leq B, \\ & y \in \{0,1\}^K, \end{aligned} \tag{68}$$

where A is the block assignment matrix, μ and c collect the path-level mean utilities and costs, and

$$Q_{(i,m),(j,n)} = \langle \tilde{v}_{im}, \tilde{v}_{jn} \rangle.$$

The per-agent screening guarantee is relative to the sampled footprint means. To obtain the population fleet condition used by the high-mean propositions, mean-estimation error and coverage approximation must be controlled separately. In an uncoupled product menu, per-agent additive allowances ε_i imply the fleet allowance $\varepsilon = \sum_i \varepsilon_i$; using the same ε for every agent permits a fleet loss of $N\varepsilon$. A fleet budget can make the sum of independent best means unattainable, so feasibility of the displayed utility floor must also be checked.

When the utility floor and budget constraints are inactive, this operational model reduces to the covariance-overlap core discussed in Appendix O, with only the one-path-per-agent assignment constraints retained. In a pooled common-menu view, that same reduction corresponds to replacing agent-indexed assignment constraints by a single cardinality condition.

Proposition 7 (Convex MIQP)**.** *Let $W \in \mathbb{R}^{R \times K}$ collect the agent-indexed vectors $\tilde{v}_{im}$, and let $Q = W^\top W$. Then $Q \succeq 0$, and the matrix formulation is a convex mixed-integer quadratic program with linear side constraints. Its only nonconvex restrictions are the binary assignments.*

The full proof is in Appendix W.

Q BEAM-SEARCH BACKEND

This appendix retains the proposed large-menu spectral backend, including its optional local-improvement step. It specifies an algorithmic construction, not a claim that the rollout or historical Gaussian-rule experiments executed this exact backend or the one-agent swaps.

Suppose a partial solution has already selected a set of paths with aggregate Mercer exposure

$$g = \sum_{(j,n) \in \mathcal{A}} \tilde{v}_{jn} \in \mathbb{R}^R.$$

If candidate (i,m) is added next, the increase in the robust term is

$$\Delta_{im}(g) := \|g + \tilde{v}_{im}\|_2^2 - \|g\|_2^2 = 2\langle g, \tilde{v}_{im} \rangle + \|\tilde{v}_{im}\|_2^2. \tag{69}$$

Thus, once the retained Mercer coordinates are available, evaluating a marginal move requires only an inner product in $\mathbb{R}^R$. This is exactly where low-rank reduction delivers computational leverage: search is executed in an R-dimensional uncertainty space rather than on the underlying spatial grid.

The proposed scalable algorithm processes agent-indexed menus sequentially. Consider a *state* variable at depth i that consists of

$$(g, U, C, \mathcal{A}),$$

where g is the current factorized overlap exposure, U is accumulated mean utility, C is accumulated cost, and $\mathcal{A}$ is the selected-candidate set. Starting from the empty state, the algorithm expands each state by one candidate from the next agent-indexed menu. After expansion, it retains only the best B_{beam} states according to a score that favors low exposure while pruning states that cannot satisfy the utility floor or budget under optimistic completion bounds.

A possible ranking score for a partial state is

$$\text{Score}(g, U, C) = \|g\|_2^2 + \rho_U \left[\alpha U_{\text{ref}} - U - U_{\text{rem}}^{\max}\right]_+ + \rho_B \left[C - B\right]_+, \tag{70}$$

Algorithm 2 Beam-search backend with local improvement

Require: Agent-indexed screened menus $\{\mathcal{M}_i^{\varepsilon}\}_{i=1}^{N}$; path statistics $(\mu_{im}, c_{im}, \tilde{v}_{im})$; utility floor αU_{ref}; budget B; beam width B_{beam}
Ensure: Feasible fleet plan $\hat{p}$, or report that no feasible assignment was found
1: Initialize the beam with the empty state $(0, 0, 0, \varnothing)$
2: **for** $i = 1, \ldots, N$ **do**
3: Expand each beam state by all candidates in $\mathcal{M}_i^{\varepsilon}$
4: Update each child state's factorized overlap exposure using (69)
5: Discard states that are provably unable to meet the utility floor or budget
6: Keep the B_{beam} states with lowest heuristic score $\mathrm{Score}(g, U, C)$ from (70)
7: **end for**
8: If the beam has no complete feasible state, report failure; otherwise select its best feasible state
9: Perform 1-swap local improvement over the agent-indexed menus
10: Return the resulting fleet plan $\hat{p}$

where $U_{\mathrm{rem}}^{\max}$ is an upper bound on the utility that can still be collected from the remaining agents and $[a]_+ = \max\{a, 0\}$. The constants ρ_U and ρ_B are large feasibility penalties used only in heuristic ranking. Importantly, the underlying design objective is unchanged: the search still targets low-exposure fleets subject to nominal-performance safeguards.

After constructing a complete assignment, the proposed backend applies a 1-swap local-improvement phase. For each agent i, the incumbent choice is tentatively replaced by alternative candidates from $\mathcal{M}_i^{\varepsilon}$, and the swap is accepted if it improves the true objective in (67) while preserving feasibility. Because the robust term updates via (69), each swap can be evaluated in $O(R)$ time once the current aggregate exposure is known.

Algorithm 2 summarizes the scalable structured-solver backend.

Algorithm 2 is the large-scale backend for the same structured covariance-overlap selector, extending the formulation to sizes where exact global optimization is impractical while preserving the same objective and constraints.

Feasibility pruning and heuristic beam ranking are separate operations. Later coordinates can offset the current aggregate vector, so $\|g\|^2$ alone is not a lower bound for a completed assignment. For an exact branch-and-bound alternative, the attainable sum of remaining coordinates lies inside its coordinatewise bounding box. The squared distance of $-g$ to that box is a valid objective lower bound. Maximum remaining utility and minimum remaining cost provide feasibility bounds. Pruning only with valid bounds preserves exactness; fixed-width beam pruning generally does not.

R SYNCHRONIZED COVARIANCE-SELECTOR CONSTRUCTION

The spectral pseudocode describes the ideal covariance-feature selector. The evaluated kernel-family variants use physical-distance Gaussian overlap instead; Appendix P gives the implementation correspondence. Optimization exactness refers to the selected algebraic objective, not the full robust detection problem.

The algorithm below retains the ideal spectral version of synchronized replanning.

The ideal covariance-feature replanner is the structured receding-horizon counterpart of the offline selector. At each replanning epoch, it constructs a local menu from the current drone states and updated posterior map, then resolves the structured utility-floor selector on that local menu and commits one leg per agent. See Algorithm 3 for this construction; the implemented Gaussian variant is distinguished above.

The key distinction from the adaptive heuristic in Appendix E is that ONLINE KERNEL REPLAN operates at synchronized replanning events: it rebuilds the local menus at $t = 0$, every `replan_every` steps, or whenever any current plan is exhausted. Between those events, the policy simply executes the currently committed legs.

Algorithm 3 Synchronized covariance-feature replanner (ideal construction)

Require: Current agent states; newly arrived observations; local scenario bundle; retained rank R; replanning period τ_{replan}
Ensure: Next committed leg for each agent
1: **if** $t = 0$ **or** t hits the fixed replanning period τ_{replan} **or** any current plan is exhausted **then**
2: Assimilate newly arrived observations and update the posterior / probability map
3: Rebuild a local candidate menu for each agent from its current state
4: Evaluate local candidates under the same scenario-bundle and retained-rank machinery to obtain $(\mu_{im,t}, c_{im,t}, \tilde{v}_{im,t})$
5: Solve the structured one-path-per-agent selector on the local menu using the exact or beam-search backend
6: Commit the selected leg for each agent until the next replanning event
7: **end if**

S General learning-layer construction and sensing models

The following GP construction is a possible statistical layer for the theory. It retains the separate target and environment models. The current experiments instead propagate a particle belief with the learned current model and prescribed wind law in Appendix D; they do not fit the target-trajectory GP below or use operational buoy and weather feeds.

Learning layer. At each planning epoch the decision maker conditions on two data streams:

$$\mathcal{D}_t^{\mathrm{tar}} = \{\text{target observations up to time } t\},$$
$$\mathcal{D}_t^{\mathrm{env}} = \{\text{environmental observations up to time } t\}.$$

We model these two objects separately because they serve different purposes in the optimization.

Target GP and posterior belief. Let $X(\tau) \in \mathbb{R}^2$ denote the target location at time τ. We treat the target trajectory as a latent process with a GP-based predictive law,

$$X(\tau) \sim \mathcal{GP}\big(m_X(\tau), K_X(\tau, \tau')\big),$$

whose prior mean and covariance may incorporate the last known position, mission-specific motion assumptions, and drift information. For linear-Gaussian observations and a compatible Gaussian conditioning model, conditioning on $\mathcal{D}_t^{\mathrm{tar}}$ and the available drift information yields a predictive distribution for the target state at the planning epoch,

$$X_t \mid \mathcal{D}_t^{\mathrm{tar}}, \mathcal{D}_t^{\mathrm{env}} \sim \mathcal{N}\big(\mu_t^X, \Sigma_t^X\big), \tag{71}$$

with associated spatial density

$$\pi_t(x) := \mathbb{P}\big(X_t \in dx \mid \mathcal{D}_t^{\mathrm{tar}}, \mathcal{D}_t^{\mathrm{env}}\big)/dx. \tag{72}$$

Unsuccessful searches generally require a non-Gaussian likelihood update and need not preserve this Gaussian form. The density π_t summarizes the nominal predictive target state. The certificate uses the scenario mean $\bar{\pi}$ obtained from valid propagated beliefs; these objects agree only when the predictive model is chosen consistently with that ensemble.

Environment GP and transport uncertainty. Let $V(x, \tau) \in \mathbb{R}^2$ denote the ambient drift field at location x and time τ. We model the environment field with a GP prior,

$$V(x, \tau) \sim \mathcal{GP}\big(m_V(x, \tau), K_V((x, \tau), (x', \tau'))\big),$$

updated with buoys, drifters, weather products, and other environmental observations. Standard GP conditioning yields a posterior mean field and covariance,

$$u(x, \tau) := \mathbb{E}\big[V(x, \tau) \mid \mathcal{D}_t^{\mathrm{env}}\big], \qquad K_{\mathrm{env}}\big((x, \tau), (x', \tau')\big) := \mathrm{Cov}\big(V(x, \tau), V(x', \tau') \mid \mathcal{D}_t^{\mathrm{env}}\big). \tag{73}$$

The posterior mean field u defines a common reference flow. The environmental posterior generates drift scenarios, and target propagation under each scenario induces a normalized belief. The

covariance of these beliefs, rather than the velocity covariance itself, determines how expected path coverage co-varies. The nominal target predictive model supplies the initial state distribution; the scenario mean $\bar{\pi}$ is the mean belief used by the certificate.

The optimization theory requires valid scenario beliefs, their specified common mean, and a covariance budget. The environmental model is one way to construct those beliefs; Appendix P states the empirical implementation and its coverage approximation.

Detection-model details. For a single visit at y and a target at x, let $d = \|x - y\|$. The two-radius family used in the earlier derivation has the form

$$p_{\mathrm{det}}^{\mathrm{old}}(d) = (1 - p_{\mathrm{fn}})\mathbf{1}\{d \le r_{\mathrm{core}}\} + \beta e^{-d^2/(2\ell^2)}\mathbf{1}\{d > r_{\mathrm{core}}\}. \tag{74}$$

Its tail limit at the core boundary is $\beta e^{-r_{\mathrm{core}}^2/(2\ell^2)}$. The evaluated simulator uses Equation 29, with radius-shifted tail $\beta(1 - p_{\mathrm{fn}})e^{-(d-r_{\mathrm{core}})^2/(2\ell^2)}$ and boundary limit $\beta(1 - p_{\mathrm{fn}})$. Both are bounded single-visit probability models for valid parameters. The closest-agent envelope additionally requires nonincreasing detection with distance. Equation (29) satisfies this for $\beta \le 1$; the old family does so only if $\beta e^{-r_{\mathrm{core}}^2/(2\ell^2)} \le 1 - p_{\mathrm{fn}}$. Continuity at the core boundary is unnecessary. The fleet sensing convention also matters: the main paper uses the closest agent at each tick; the product model is stated separately in Lemma 1 and its proof.

T High-mean menu reduction and risk-scaled coordinates

This appendix records the high-mean screening and low-rank algebra. Let $\mathcal{F}_N^{\mathrm{menu}} \subseteq \mathcal{F}_N$ be the feasible fleet combinations from a specified raw candidate menu. All approximation comparisons below are relative to this reference class; they do not bound candidate-generation loss. Define

$$\mu^* := \sup_{p \in \mathcal{F}_N^{\mathrm{menu}}} \mu_{U_p}, \tag{75}$$

and define the high-mean menu

$$\mathcal{M}_\varepsilon := \left\{ p \in \mathcal{F}_N^{\mathrm{menu}} : \ \mu_{U_p} \ge \mu^* - \varepsilon \right\}. \tag{76}$$

Assume $\mathcal{M}_\varepsilon$ is nonempty; this holds for $\varepsilon > 0$, or for $\varepsilon = 0$ when the supremum is attained. For a fixed rank R, define the truncated variance

$$\sigma_{U_p,R}^2 := \sum_{r=1}^{R} \lambda_r V_r(p)^2. \tag{77}$$

Proposition 5 gives the approximation argument on $\mathcal{M}_\varepsilon$, while Corollary 2 quantifies the resulting robust-objective gap.

With the risk-scaled path coordinates

$$\tilde{v}_{p_i} := \left(\sqrt{\lambda_1} v_{p_i}^{(1)}, \ldots, \sqrt{\lambda_R} v_{p_i}^{(R)}\right) \in \mathbb{R}^R,$$

the rank-R covariance-overlap penalty becomes

$$\sigma_{U_p,R}^2 = \left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2,$$

so the retained Mercer modes define a finite-dimensional design problem with explicit error control through Corollary 3.

U Incremental replanning surrogate

This appendix records incremental formulas for the ideal belief-covariance selector and a conditional local-kernel comparison. For candidate paths p and q, the exact covariance-overlap kernel is

$$k_{\mathrm{Lag}}(p, q) := \int_{\mathcal{X}} \int_{\mathcal{X}} c_p(\tilde{x}) K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') c_q(\tilde{x}') \, d\nu(\tilde{x}) \, d\nu(\tilde{x}'). \tag{78}$$

For any fleet plan $P = (p_1, \dots, p_m)$, the covariance-overlap penalty is

$$R(P) := \sum_{i=1}^{m} \sum_{j=1}^{m} k_{\mathrm{Lag}}(p_i, p_j).$$

For this additive variance objective, the increment contains both self variance and covariance with committed routes: if $\mathcal{C}_t$ is the committed set at epoch t, then

$$\Delta R(p \mid \mathcal{C}_t) = R(\mathcal{C}_t \cup \{p\}) - R(\mathcal{C}_t) = k_{\mathrm{Lag}}(p, p) + 2 \sum_{q \in \mathcal{C}_t} k_{\mathrm{Lag}}(p, q).$$

The following statement concerns a separate spatial waypoint patch, and assumes a local kernel shape; it does not derive that shape from a velocity GP or from normalization of belief maps. A globally positive Gaussian kernel need not satisfy the belief mass constraint $C\mathbf{1} = 0$. Local approximation, quadrature, pair averaging, truncation, and candidate-dependent self-overlap must all be accounted for before transferring a detection certificate to a Gaussian online score.

Proposition 8 (Short-horizon consistency of the waypoint local Gaussian surrogate)**.** *Suppose, on the waypoint patch under consideration, K_{Lag} admits the local radial representation*

$$K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') = \kappa(\|\tilde{x} - \tilde{x}'\|), \qquad \kappa(r) = \kappa(0) \exp\left(-\frac{r^2}{2\ell^2}\right) + E(r),$$

with Taylor remainder

$$|E(r)| \le C_2' \frac{r^4}{\ell^4}, \qquad 0 \le r \le 3\ell.$$

Suppose the backward flow map satisfies

$$\|\Phi_{\tau,0}(x) - x\| \le \delta$$

for every waypoint under consideration. For two candidate paths p, q with waypoint sets $X(p)$ and $X(q)$, define the waypoint-discretized Lagrangian path-pair kernel

$$k_{\mathrm{Lag}}^{\mathrm{wp}}(p, q) := \sum_{x \in X(p)} \sum_{y \in X(q)} K_{\mathrm{Lag}}\big(\Phi_{\tau,0}(x), \Phi_{\tau,0}(y)\big),$$

and the local Gaussian surrogate with bandwidth $\sigma = \ell$ by

$$k_\sigma(p, q) := \sum_{x \in X(p)} \sum_{y \in X(q)} \exp\left(-\frac{\|x - y\|^2}{2\ell^2}\right).$$

Let

$$M_p := |X(p)|, \qquad M_q := |X(q)|, \qquad \rho_{pq} := \max_{x \in X(p),\, y \in X(q)} \|x - y\|.$$

If $\rho_{pq} \le \ell$ and $\delta \le \ell$, then

$$\left| k_{\mathrm{Lag}}^{\mathrm{wp}}(p, q) - \kappa(0) k_\sigma(p, q) \right| \le M_p M_q \left(C_1 \frac{\delta}{\ell} + C_2 \frac{\rho_{pq}^4}{\ell^4} \right),$$

where one may take

$$C_1 = 4|\kappa(0)| + 128 C_2', \qquad C_2 = 8 C_2'.$$

Thus each waypoint-discretized Lagrangian cross-overlap term is tracked by the local Gaussian kernel up to $O(\delta/\ell + \rho_{pq}^4/\ell^4)$ in the short-horizon regime.

The proof is provided in Appendix W.

These conditional identities compare an ideal belief kernel on waypoint pairs with an unnormalized Gaussian score. Connecting them to the implemented online score also requires averaging and cutoff corrections, coverage quadrature, and the full covariance increment, including self-overlap.

For time-stamped waypoints, each occurrence of $\Phi_{\tau,0}$ is evaluated at that waypoint's timestamp. The proof uses the same displacement bound for both maps and therefore also covers distinct waypoint times. The exact inequality

$$\big| \|\Phi_{\tau_x,0}(x) - \Phi_{\tau_y,0}(y)\| - \|x - y\| \big| \le \|(\Phi_{\tau_x,0}(x) - x) - (\Phi_{\tau_y,0}(y) - y)\|$$

shows that relative displacement can sharpen the pairwise perturbation bound. It does not establish a deformation bound for the evaluated environments.

V EXACT COVARIANCE-OVERLAP SELECTOR PSEUDOCODE

The spectral pseudocode describes the ideal covariance-feature selector. The evaluated kernel-family variants use physical-distance Gaussian overlap instead; Appendix P gives the implementation correspondence. Optimization exactness refers to the selected algebraic objective, not the full robust detection problem.

Algorithm 4 Exact covariance-overlap selector

Require: Screened menus $\{\mathcal{M}_i^{\varepsilon}\}_{i=1}^{N}$; path statistics $(\mu_{im}, c_{im}, \tilde{v}_{im})$; optional utility retention fraction α; optional budget B
Ensure: Selected fleet plan $p^\star = (p_1^\star, \dots, p_N^\star)$
1: If side constraints are active, compute U_{ref} from (64)
2: Stack all path vectors into $W = [\tilde{v}_{11} \ \cdots \ \tilde{v}_{NK_N}]$ and form $Q \leftarrow W^\top W$
3: Build the assignment matrix A and, when needed, the side-constraint vectors μ, c
4: Solve (68), dropping any inactive utility-floor or budget constraints
5: If infeasible, report infeasibility; otherwise, for each agent i, extract the unique candidate with $y_{im}^\star = 1$
6: Return the corresponding fleet plan $p^\star$

W PROOFS FOR THE THEORY EXTENSIONS

This section retains the complete spectral, menu-reduction, assignment, and local-surrogate proofs. These results extend the additive covariance model; they are not additional performance guarantees for TBR.

PROOF OF THEOREM 2

Proof. Here $\sigma_p^2 = \langle s_p, C s_p \rangle$ is the scenario-coverage variance budget from Corollary 1. Proposition 2 gives

$$\mathrm{Var}_Q(U_p) = \langle s_p, C_Q s_p \rangle \le \sigma_p^2,$$

with equality for $C = C_Q$. The following spectral identity factors this same budget; it does not redefine the random variable as a velocity exposure.

For a fleet plan $p = (p_1, \dots, p_N)$,

$$s_p(\tilde{x}) = \sum_{i=1}^{N} c_{p_i}(\tilde{x}).$$

Since each $c_{p_i} \in [0,1]$, we have $0 \le s_p \le N$. Because $\mathcal{X}$ has finite measure,

$$\|s_p\|_{\mathcal{H}}^2 \le N^2 \nu(\mathcal{X}) < \infty,$$

so $s_p \in \mathcal{H}$.

Let $s_{p,0}$ be the projection onto $\ker C$. Expand the remaining component in the nonzero eigenmodes from Proposition 4:

$$s_p(\tilde{x}) = s_{p,0}(\tilde{x}) + \sum_{r \ge 1} V_r(p)\psi_r(\tilde{x}), \qquad V_r(p) = \langle s_p, \psi_r \rangle_{\mathcal{H}} = \int_{\mathcal{X}} s_p(\tilde{x})\psi_r(\tilde{x})\, d\nu(\tilde{x}).$$

Let

$$s_{p,m} := s_{p,0} + \sum_{r=1}^{m} V_r(p)\psi_r.$$

Then $s_{p,m} \to s_p$ in $\mathcal{H}$. Since C is bounded by Proposition 4, we also have $C s_{p,m} \to C s_p$ in $\mathcal{H}$. Moreover, using $C s_{p,0} = 0$ and $C\psi_r = \lambda_r \psi_r$,

$$C s_{p,m} = \sum_{r=1}^{m} \lambda_r V_r(p)\psi_r.$$

Therefore, by continuity of the inner product and orthogonality of the nullspace to the nonzero modes,

$$\begin{aligned}
\sigma_p^2 &= \langle s_p, C s_p \rangle_{\mathcal{H}} \\
&= \lim_{m\to\infty} \langle s_{p,m}, C s_{p,m} \rangle_{\mathcal{H}} \\
&= \lim_{m\to\infty} \left\langle \sum_{r=1}^{m} V_r(p)\psi_r, \sum_{s=1}^{m} \lambda_s V_s(p)\psi_s \right\rangle_{\mathcal{H}} \\
&= \lim_{m\to\infty} \sum_{r=1}^{m}\sum_{s=1}^{m} V_r(p)\lambda_s V_s(p)\langle \psi_r, \psi_s \rangle_{\mathcal{H}} \\
&= \lim_{m\to\infty} \sum_{r=1}^{m} \lambda_r V_r(p)^2. \\
&= \sum_{r\geq 1} \lambda_r V_r(p)^2.
\end{aligned}$$

This proves the Mercer-mode variance identity.

It remains to identify the fleet loading with the sum of path-level loadings. Since $s_p = \sum_{i=1}^{N} c_{p_i}$, linearity of the inner product gives

$$\begin{aligned}
V_r(p) &= \int_{\mathcal{X}} s_p(\tilde{x})\psi_r(\tilde{x})\, d\nu(\tilde{x}) \\
&= \sum_{i=1}^{N} \int_{\mathcal{X}} c_{p_i}(\tilde{x})\psi_r(\tilde{x})\, d\nu(\tilde{x}) \\
&= \sum_{i=1}^{N} v_{p_i}^{(r)}.
\end{aligned}$$

For a rank-R truncation, define

$$\tilde{v}_{p_i} = \left(\sqrt{\lambda_1} v_{p_i}^{(1)}, \ldots, \sqrt{\lambda_R} v_{p_i}^{(R)} \right) \in \mathbb{R}^R.$$

Then

$$\begin{aligned}
\sigma_{p,R}^2 &= \sum_{r=1}^{R} \lambda_r V_r(p)^2 \\
&= \sum_{r=1}^{R} \lambda_r \left(\sum_{i=1}^{N} v_{p_i}^{(r)} \right)^2 \\
&= \sum_{r=1}^{R} \left(\sum_{i=1}^{N} \sqrt{\lambda_r} v_{p_i}^{(r)} \right)^2 \\
&= \left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2 .
\end{aligned}$$

Thus the rank-R covariance-overlap penalty is the squared Euclidean norm of the summed risk-scaled path coordinates. Zero-eigenvalue coordinates, if present, contribute zero to the variance. □

PROOF OF PROPOSITION 7

Proof. Let

$$Q = W^\top W, \qquad W \in \mathbb{R}^{R\times K}.$$

For any $z \in \mathbb{R}^K$,

$$z^\top Q z = z^\top W^\top W z = \|Wz\|_2^2 \geq 0.$$

Hence

$$Q \succeq 0.$$

Equivalently, the Hessian of the quadratic objective

$$y \mapsto y^\top Q y$$

is

$$2Q \succeq 0,$$

so the objective is convex on the continuous relaxation $y \in \mathbb{R}^K$.

The assignment constraints

$$Ay = \mathbf{1}$$

are linear. The optional utility-retention and budget constraints,

$$\mu^\top y \geq \alpha U_{\mathrm{ref}}, \qquad c^\top y \leq B,$$

are also linear. Finally, the discrete restrictions are binary:

$$y \in \{0,1\}^K.$$

Thus the continuous relaxation has a convex quadratic objective and linear constraints, and the only nonconvexity comes from the binary restrictions. Therefore the screened-menu selector is a convex mixed-integer quadratic program with linear side constraints. □

PROOF OF PROPOSITION 4

The following detailed integral proof uses the additional pointwise Mercer conditions. For a general positive trace-class budget, the compact self-adjoint spectral theorem yields the same L^2 identities: its nonnegative eigenvalues are summable, hence square-summable, and $\|C\|_{\mathrm{HS}}^2 = \sum_r \lambda_r^2 < \infty$.

Proof. Let

$$(Cf)(\tilde{x}) := \int_{\mathcal{X}} K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') f(\tilde{x}')\, d\nu(\tilde{x}').$$

The planning budget C is assumed positive and self-adjoint. In the continuous-kernel case, K_{Lag} is symmetric and positive semidefinite; these are properties of the belief covariance budget, not consequences of a velocity-GP posterior formula.

Since K_{Lag} is continuous on compact $\mathcal{X} \times \mathcal{X}$, it is square-integrable. Hence the associated integral operator is Hilbert–Schmidt:

$$\|C\|_{\mathrm{HS}}^2 = \int_{\mathcal{X}} \int_{\mathcal{X}} |K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}')|^2\, d\nu(\tilde{x})\, d\nu(\tilde{x}') < \infty.$$

Therefore C is bounded and compact on $\mathcal{H}$.

For $f, g \in \mathcal{H}$, boundedness of K_{Lag} and finite measure of $\mathcal{X}$ imply

$$\begin{aligned}\int_{\mathcal{X}} \int_{\mathcal{X}} |f(\tilde{x}) K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') g(\tilde{x}')|\, d\nu(\tilde{x}')\, d\nu(\tilde{x}) &\leq \|K_{\mathrm{Lag}}\|_\infty \|f\|_{L^1(\mathcal{X},\nu)} \|g\|_{L^1(\mathcal{X},\nu)} \\ &\leq \|K_{\mathrm{Lag}}\|_\infty \nu(\mathcal{X}) \|f\|_{\mathcal{H}} \|g\|_{\mathcal{H}} < \infty.\end{aligned}$$

Thus Fubini's theorem applies. Using symmetry of K_{Lag},

$$\begin{aligned}\langle f, Cg \rangle_{\mathcal{H}} &= \int_{\mathcal{X}} \int_{\mathcal{X}} f(\tilde{x}) K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') g(\tilde{x}')\, d\nu(\tilde{x}')\, d\nu(\tilde{x}) \\ &= \int_{\mathcal{X}} \int_{\mathcal{X}} g(\tilde{x}') K_{\mathrm{Lag}}(\tilde{x}', \tilde{x}) f(\tilde{x})\, d\nu(\tilde{x})\, d\nu(\tilde{x}') \\ &= \langle Cf, g \rangle_{\mathcal{H}}.\end{aligned}$$

Hence C is self-adjoint. Since the planning kernel is positive semidefinite,

$$\langle f, Cf \rangle_{\mathcal{H}} \geq 0 \qquad \text{for all } f \in \mathcal{H},$$

so C is positive.

By the spectral theorem for compact self-adjoint positive operators, C admits nonnegative eigenvalues $\{\lambda_r\}_{r\geq 1}$ and orthonormal eigenfunctions $\{\psi_r\}_{r\geq 1}$. If the positive-eigenvalue eigenfunctions do not span all of $\mathcal{H}$, we adjoin an orthonormal basis of $\ker(C)$ and assign eigenvalue zero to those additional basis functions. Together these give a complete orthonormal eigenbasis of $\mathcal{H}$. In the ordered spectral sums we index only the nonzero modes; the separate nullspace projection contributes zero. If the rank is finite, zero modes may be appended to those finite sums.

Since K_{Lag} is continuous, symmetric, and positive semidefinite on compact $\mathcal{X}$, Mercer's theorem gives the uniformly convergent expansion

$$K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') = \sum_{s\geq 1} \lambda_s \psi_s(\tilde{x})\psi_s(\tilde{x}'), \qquad \lambda_s \geq 0.$$

Zero-eigenvalue basis elements, if adjoined, contribute zero terms to this expansion.

For completeness, we verify directly that the Mercer functions diagonalize the integral operator. Let r correspond to a positive-eigenvalue Mercer function. Then ψ_r is continuous on compact $\mathcal{X}$, hence bounded. Therefore the uniform convergence of the Mercer expansion implies uniform convergence, in $\tilde{x}'$, of

$$\sum_{s\geq 1} \lambda_s \psi_s(\tilde{x})\psi_s(\tilde{x}')\psi_r(\tilde{x}'),$$

so termwise integration is justified. Hence

$$\begin{aligned}(C\psi_r)(\tilde{x}) &= \int_{\mathcal{X}} \sum_{s\geq 1} \lambda_s \psi_s(\tilde{x})\psi_s(\tilde{x}')\psi_r(\tilde{x}')\, d\nu(\tilde{x}') \\ &= \sum_{s\geq 1} \lambda_s \psi_s(\tilde{x}) \int_{\mathcal{X}} \psi_s(\tilde{x}')\psi_r(\tilde{x}')\, d\nu(\tilde{x}') \\ &= \sum_{s\geq 1} \lambda_s \psi_s(\tilde{x}) \langle \psi_s, \psi_r\rangle_{\mathcal{H}} = \lambda_r \psi_r(\tilde{x}).\end{aligned}$$

For an adjoined zero-eigenvalue basis function, the same identity holds by construction:

$$C\psi_r = 0 = \lambda_r \psi_r.$$

Therefore $C\psi_r = \lambda_r \psi_r$ for every basis function ψ_r. These are standard Mercer covariance-operator facts; see, e.g., Rasmussen & Williams (2006). □

PROOF OF PROPOSITION 5

Proof. Assume $\mathcal{M}_\varepsilon$ is nonempty and fix $p \in \mathcal{M}_\varepsilon$. Since

$$0 \leq s_q(X) \leq N$$

for every feasible plan q, each posterior mean μ_{U_q} is finite and

$$0 \leq \mu_{U_q} \leq N.$$

Hence

$$\mu^* := \sup_{q \in \mathcal{F}_N^{\mathrm{menu}}} \mu_{U_q}$$

is finite.

Define

$$\delta_p := \mu_{U_p} - \mu^*.$$

By definition of μ^* as a supremum over the feasible reference menu,

$$\mu_{U_p} \leq \mu^*.$$

Since $p \in \mathcal{M}_\varepsilon$, we also have

$$\mu_{U_p} \geq \mu^* - \varepsilon.$$

Combining these inequalities gives

$$-\varepsilon \leq \delta_p \leq 0.$$

Thus

$$\mu_{U_p} = \mu^* + \delta_p, \qquad \delta_p \in [-\varepsilon, 0],$$

which proves (45).

Substituting into (46) gives, for every $p \in \mathcal{M}_\varepsilon$,

$$\mu_{U_p} - \kappa_w \sigma_{U_p} = \mu^* - \kappa_w \sigma_{U_p} + \delta_p, \qquad \delta_p \in [-\varepsilon, 0].$$

Thus, on $\mathcal{M}_\varepsilon$, the full additive surrogate differs from the constant-mean variance-control proxy

$$\mu^* - \kappa_w \sigma_{U_p}$$

by an additive perturbation δ_p whose magnitude is at most ε.

Now take any $p, q \in \mathcal{M}_\varepsilon$. Then

$$\begin{aligned}\big(\mu_{U_p} - \kappa_w \sigma_{U_p}\big) - \big(\mu_{U_q} - \kappa_w \sigma_{U_q}\big) &= \big(\mu^* - \kappa_w \sigma_{U_p} + \delta_p\big) - \big(\mu^* - \kappa_w \sigma_{U_q} + \delta_q\big) \\ &= -\kappa_w(\sigma_{U_p} - \sigma_{U_q}) + (\delta_p - \delta_q).\end{aligned}$$

Equivalently,

$$\big(\mu_{U_p} - \kappa_w \sigma_{U_p}\big) - \big(\mu_{U_q} - \kappa_w \sigma_{U_q}\big) + \kappa_w(\sigma_{U_p} - \sigma_{U_q}) = \delta_p - \delta_q.$$

Because

$$\delta_p, \delta_q \in [-\varepsilon, 0],$$

the diameter of this interval is ε, and therefore

$$|\delta_p - \delta_q| \le \varepsilon.$$

Consequently,

$$\Big|\big(\mu_{U_p} - \kappa_w \sigma_{U_p}\big) - \big(\mu_{U_q} - \kappa_w \sigma_{U_q}\big) + \kappa_w(\sigma_{U_p} - \sigma_{U_q})\Big| \le \varepsilon.$$

This proves (47).

In particular, when $\varepsilon = 0$, all plans in $\mathcal{M}_0$ have exactly the same mean μ^*. Hence, if $\kappa_w > 0$, maximizing $\mu_{U_p} - \kappa_w \sigma_{U_p}$ over $\mathcal{M}_0$ is exactly equivalent to minimizing σ_{U_p} over $\mathcal{M}_0$. If $\kappa_w = 0$, the objective is constant over the equal-mean menu, so variance minimization is a tie-breaker rather than a strict equivalent objective. □

PROOF OF PROPOSITION 6

Proof. By Theorem 2, for any feasible plan $p = (p_1, \dots, p_N)$, the rank-R truncated variance satisfies

$$\sigma^2_{U_p,R} = \sum_{r=1}^{R} \lambda_r V_r(p)^2, \qquad V_r(p) = \sum_{i=1}^{N} v^{(r)}_{p_i}.$$

Therefore

$$\sigma^2_{U_p,R} = \sum_{r=1}^{R} \lambda_r \left(\sum_{i=1}^{N} v^{(r)}_{p_i} \right)^2.$$

Define the risk-scaled path coordinates by

$$\tilde{v}^{(r)}_{p_i} := \sqrt{\lambda_r}\, v^{(r)}_{p_i}, \qquad \tilde{v}_{p_i} := \left(\tilde{v}^{(1)}_{p_i}, \dots, \tilde{v}^{(R)}_{p_i} \right) \in \mathbb{R}^R.$$

Then

$$\begin{aligned}\sigma^2_{U_p,R} &= \sum_{r=1}^{R} \left(\sum_{i=1}^{N} \tilde{v}^{(r)}_{p_i} \right)^2 \\ &= \left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2.\end{aligned}$$

On $\mathcal{M}_{\mathrm{eq}}$, the mean term is constant:

$$\mu_{U_p} = \bar{\mu} \qquad \text{for all } p \in \mathcal{M}_{\mathrm{eq}}.$$

If $\kappa_w > 0$, maximizing the rank-R truncated robust objective

$$\bar{\mu} - \kappa_w \sigma_{U_p,R}$$

over $\mathcal{M}_{\text{eq}}$ is equivalent to minimizing $\sigma_{U_p,R}$ over $\mathcal{M}_{\text{eq}}$. Since $\sigma_{U_p,R} \geq 0$ and $t \mapsto t^2$ is increasing on $\mathbb{R}_+$, this is equivalent to minimizing $\sigma^2_{U_p,R}$. Using the identity above, the problem becomes

$$\min_{p \in \mathcal{M}_{\text{eq}}} \left\| \sum_{i=1}^{N} \tilde{v}_{p_i} \right\|_2^2 .$$

This is exactly (52).

If $\kappa_w = 0$, then the truncated robust objective is constant on $\mathcal{M}_{\text{eq}}$, so every plan in $\mathcal{M}_{\text{eq}}$ is optimal. In this degenerate risk-neutral case, the squared-norm objective should be interpreted as a covariance-overlap tie-breaker rather than as a strict equivalent objective. □

PROOF OF COROLLARY 2

Proof. Fix $p \in \mathcal{M}_\varepsilon$. Recall that

$$\mathcal{J}(p) = \mu_{U_p} - \kappa_w \sigma_{U_p}, \qquad \mathcal{J}_R(p) = \mu^* - \kappa_w \sigma_{U_p,R}.$$

The proxy $\mathcal{J}_R$ replaces the path-dependent mean μ_{U_p} with the screening-set supremum μ^*; this is why the screening error $\mu^* - \mu_{U_p}$ enters the gap below. We have

$$\begin{aligned} \mathcal{J}_R(p) - \mathcal{J}(p) &= \big(\mu^* - \kappa_w \sigma_{U_p,R}\big) - \big(\mu_{U_p} - \kappa_w \sigma_{U_p}\big) \\ &= (\mu^* - \mu_{U_p}) + \kappa_w (\sigma_{U_p} - \sigma_{U_p,R}). \end{aligned}$$

Because $p \in \mathcal{M}_\varepsilon$, we have

$$\mu^* - \varepsilon \leq \mu_{U_p} \leq \mu^*,$$

and hence

$$0 \leq \mu^* - \mu_{U_p} \leq \varepsilon.$$

By Corollary 3,

$$0 \leq \sigma_{U_p} - \sigma_{U_p,R} \leq \sqrt{\lambda_{R+1}}\, \|s_p\|_{\mathcal{H}}.$$

Moreover, in the setting of the moment-robust bound, $\kappa_w \geq 0$. Hence the truncation contribution is nonnegative and satisfies

$$0 \leq \kappa_w (\sigma_{U_p} - \sigma_{U_p,R}) \leq \kappa_w \sqrt{\lambda_{R+1}}\, \|s_p\|_{\mathcal{H}}.$$

Let

$$B_\varepsilon := \sup_{q \in \mathcal{M}_\varepsilon} \|s_q\|_{\mathcal{H}}.$$

This constant is finite. Indeed, since $0 \leq c_{q_i} \leq 1$ for each agent,

$$0 \leq s_q(\tilde{x}) \leq N,$$

and because $\mathcal{X}$ has finite measure,

$$\|s_q\|_{\mathcal{H}} \leq N \sqrt{\nu(\mathcal{X})}.$$

Thus

$$B_\varepsilon \leq N \sqrt{\nu(\mathcal{X})} < \infty.$$

Consequently, for every $p \in \mathcal{M}_\varepsilon$,

$$0 \leq \sigma_{U_p} - \sigma_{U_p,R} \leq \sqrt{\lambda_{R+1}}\, B_\varepsilon.$$

Combining the mean-screening error and the truncation error gives

$$0 \leq \mathcal{J}_R(p) - \mathcal{J}(p) \leq \varepsilon + \kappa_w \sqrt{\lambda_{R+1}}\, B_\varepsilon.$$

This proves (53). □

PROOF OF COROLLARY 3

Proof. Fix a feasible fleet plan p. By Proposition 4, the eigenfunctions $\{\psi_r\}_{r\geq 1}$ are the nonzero orthonormal modes of C; let $s_{p,0}$ denote the nullspace component. Hence

$$s_p(\tilde{x}) = s_{p,0}(\tilde{x}) + \sum_{r\geq 1} V_r(p)\psi_r(\tilde{x}), \qquad V_r(p) = \langle s_p, \psi_r\rangle_{\mathcal{H}}.$$

Define the rank-R projection and tail by

$$s_{p,R} := \sum_{r=1}^{R} V_r(p)\psi_r, \qquad s_{p,>R} := \sum_{r>R} V_r(p)\psi_r.$$

By Theorem 2,

$$\sigma^2_{U_p} = \sum_{r\geq 1} \lambda_r V_r(p)^2, \qquad \sigma^2_{U_p,R} = \sum_{r=1}^{R} \lambda_r V_r(p)^2.$$

Therefore

$$\sigma^2_{U_p} - \sigma^2_{U_p,R} = \sum_{r>R} \lambda_r V_r(p)^2.$$

Since the eigenvalues are nonincreasing,

$$\sum_{r>R} \lambda_r V_r(p)^2 \leq \lambda_{R+1} \sum_{r>R} V_r(p)^2.$$

By orthogonality of the retained, tail and nullspace components,

$$\sum_{r>R} V_r(p)^2 = \|s_{p,>R}\|^2_{\mathcal{H}} \leq \|s_p\|^2_{\mathcal{H}}.$$

Thus

$$0 \leq \sigma^2_{U_p} - \sigma^2_{U_p,R} \leq \lambda_{R+1}\|s_p\|^2_{\mathcal{H}}.$$

This proves (54).

We also record the corresponding standard-deviation bound. Let

$$\Delta_R := \sigma^2_{U_p} - \sigma^2_{U_p,R}.$$

Since $\sigma_{U_p} \geq \sigma_{U_p,R} \geq 0$, we may apply the elementary inequality

$$\sqrt{a} - \sqrt{b} \leq \sqrt{a-b}, \qquad a \geq b \geq 0.$$

Indeed,

$$(\sqrt{a-b} + \sqrt{b})^2 = a + 2\sqrt{(a-b)b} \geq a.$$

Taking $a = \sigma^2_{U_p}$ and $b = \sigma^2_{U_p,R}$ gives

$$0 \leq \sigma_{U_p} - \sigma_{U_p,R} \leq \sqrt{\Delta_R}.$$

Combining this with the variance truncation bound yields

$$0 \leq \sigma_{U_p} - \sigma_{U_p,R} \leq \sqrt{\lambda_{R+1}}\,\|s_p\|_{\mathcal{H}}.$$

This proves (48). □

PROOF OF PROPOSITION 8

Proof. Fix two candidate paths p, q with waypoint sets $X(p)$ and $X(q)$. Recall the waypoint-discretized Lagrangian path-pair kernel

$$k^{\mathrm{wp}}_{\mathrm{Lag}}(p,q) = \sum_{x\in X(p)} \sum_{y\in X(q)} K_{\mathrm{Lag}}\big(\Phi_{\tau,0}(x), \Phi_{\tau,0}(y)\big),$$

and the local Gaussian surrogate

$$k_\sigma(p,q) = \sum_{x\in X(p)} \sum_{y\in X(q)} \exp\left(-\frac{\|x-y\|^2}{2\ell^2}\right), \qquad \sigma = \ell.$$

Fix one waypoint pair $(x,y) \in X(p) \times X(q)$. Define

$$\tilde{x} := \Phi_{\tau,0}(x), \qquad \tilde{y} := \Phi_{\tau,0}(y),$$

and set

$$r := \|x-y\|, \qquad \tilde{r} := \|\tilde{x}-\tilde{y}\|.$$

By the short-horizon flow assumption,

$$\|\tilde{x}-x\| \le \delta, \qquad \|\tilde{y}-y\| \le \delta.$$

Using the reverse triangle inequality followed by the triangle inequality,

$$\begin{aligned} |\tilde{r}-r| &= |\|\tilde{x}-\tilde{y}\| - \|x-y\|| \\ &\le \|(\tilde{x}-\tilde{y})-(x-y)\| \\ &\le \|\tilde{x}-x\| + \|\tilde{y}-y\| \\ &\le 2\delta. \end{aligned}$$

Since $r \le \rho_{pq} \le \ell$ and $\delta \le \ell$,

$$\tilde{r} \le r + 2\delta \le \rho_{pq} + 2\delta \le 3\ell.$$

Thus the Taylor-remainder hypothesis applies to $\tilde{r}$. Moreover,

$$\tilde{r} + r \le (r+2\delta) + r = 2r + 2\delta \le 2\rho_{pq} + 2\delta \le 4\ell.$$

Therefore,

$$|\tilde{r}^2 - r^2| = |\tilde{r}-r|(\tilde{r}+r) \le 2\delta \cdot 4\ell = 8\delta\ell.$$

For $a, b \ge 0$, the map $u \mapsto e^{-u}$ is 1-Lipschitz on $\mathbb{R}_+$, so

$$|e^{-a} - e^{-b}| \le |a-b|.$$

Applying this with

$$a = \frac{\tilde{r}^2}{2\ell^2}, \qquad b = \frac{r^2}{2\ell^2},$$

gives

$$\left|\exp\left(-\frac{\tilde{r}^2}{2\ell^2}\right) - \exp\left(-\frac{r^2}{2\ell^2}\right)\right| \le \frac{|\tilde{r}^2 - r^2|}{2\ell^2} \le 4\frac{\delta}{\ell}.$$

By the assumed local radial representation,

$$K_{\mathrm{Lag}}(\tilde{x},\tilde{y}) = \kappa(\tilde{r}) = \kappa(0)\exp\left(-\frac{\tilde{r}^2}{2\ell^2}\right) + E(\tilde{r}).$$

Since $\tilde{r} \le \rho_{pq} + 2\delta \le 3\ell$, the Taylor-remainder bound gives

$$|E(\tilde{r})| \le C_2' \frac{(\rho_{pq}+2\delta)^4}{\ell^4}.$$

Using

$$(a+b)^4 \le 8(a^4+b^4), \qquad a, b \ge 0,$$

with $a = \rho_{pq}$ and $b = 2\delta$, we obtain

$$(\rho_{pq}+2\delta)^4 \le 8\rho_{pq}^4 + 128\delta^4.$$

Since $\delta \le \ell$,

$$\frac{\delta^4}{\ell^4} = \left(\frac{\delta}{\ell}\right)^4 \le \frac{\delta}{\ell}.$$

Hence

$$|E(\tilde{r})| \le 8C_2' \frac{\rho_{pq}^4}{\ell^4} + 128C_2' \frac{\delta}{\ell}.$$

Combining the Gaussian perturbation and Taylor-remainder bounds, we get the pointwise waypoint-pair estimate

$$\begin{aligned}
&\left| K_{\mathrm{Lag}}(\tilde{x}, \tilde{y}) - \kappa(0) \exp\left(-\frac{\|x-y\|^2}{2\ell^2}\right) \right| \\
&\quad = \left| \kappa(0) \left[\exp\left(-\frac{\tilde{r}^2}{2\ell^2}\right) - \exp\left(-\frac{r^2}{2\ell^2}\right) \right] + E(\tilde{r}) \right| \\
&\quad \le 4|\kappa(0)| \frac{\delta}{\ell} + 8C_2' \frac{\rho_{pq}^4}{\ell^4} + 128C_2' \frac{\delta}{\ell} \\
&\quad = \left(4|\kappa(0)| + 128C_2'\right) \frac{\delta}{\ell} + 8C_2' \frac{\rho_{pq}^4}{\ell^4}.
\end{aligned}$$

Thus, defining

$$C_1 := 4|\kappa(0)| + 128C_2', \qquad C_2 := 8C_2',$$

we have

$$\left| K_{\mathrm{Lag}}\big(\Phi_{\tau,0}(x), \Phi_{\tau,0}(y)\big) - \kappa(0) \exp\left(-\frac{\|x-y\|^2}{2\ell^2}\right) \right| \le C_1 \frac{\delta}{\ell} + C_2 \frac{\rho_{pq}^4}{\ell^4}.$$

Summing over all waypoint pairs and applying the triangle inequality,

$$\begin{aligned}
&\left| k_{\mathrm{Lag}}^{\mathrm{wp}}(p,q) - \kappa(0) k_\sigma(p,q) \right| \\
&\quad = \left| \sum_{x \in X(p)} \sum_{y \in X(q)} \left[K_{\mathrm{Lag}}\big(\Phi_{\tau,0}(x), \Phi_{\tau,0}(y)\big) - \kappa(0) \exp\left(-\frac{\|x-y\|^2}{2\ell^2}\right) \right] \right| \\
&\quad \le \sum_{x \in X(p)} \sum_{y \in X(q)} \left| K_{\mathrm{Lag}}\big(\Phi_{\tau,0}(x), \Phi_{\tau,0}(y)\big) - \kappa(0) \exp\left(-\frac{\|x-y\|^2}{2\ell^2}\right) \right| \\
&\quad \le M_p M_q \left(C_1 \frac{\delta}{\ell} + C_2 \frac{\rho_{pq}^4}{\ell^4} \right).
\end{aligned}$$

This proves the claimed short-horizon consistency of the waypoint local Gaussian surrogate.

For a committed set $\mathcal{C}_t$, let $H_{\mathrm{wp}}(p) = \sum_{q \in \mathcal{C}_t} k_{\mathrm{Lag}}^{\mathrm{wp}}(p,q)$ and $G_{\mathrm{wp}}(p) = \sum_{q \in \mathcal{C}_t} k_\sigma(p,q)$. Summing the proved pair bounds yields

$$|H_{\mathrm{wp}}(p) - \kappa(0) G_{\mathrm{wp}}(p)| \le \sum_{q \in \mathcal{C}_t} M_p M_q \left(C_1 \frac{\delta}{\ell} + C_2 \frac{\rho_{pq}^4}{\ell^4} \right).$$

This compares unnormalized cross terms only. G_{wp} is not the averaged, cutoff score G_t in the implementation, and H_{wp} omits the self-term and factor of two in the full increment ΔR. Thus the inequality alone is not an online detection certificate.

For the spatial-patch special case, if the exact continuous path-pair covariance overlap

$$k_{\mathrm{Lag}}^{\mathrm{cont}}(p,q) = \int_\Omega \int_\Omega c_p(\tilde{x}) K_{\mathrm{Lag}}(\tilde{x}, \tilde{x}') c_q(\tilde{x}') \, d\tilde{x} \, d\tilde{x}'$$

is used instead, then a separate quadrature approximation assumption is required. For example, if

$$\varepsilon_{\mathrm{quad}}(p,q) := \left| k_{\mathrm{Lag}}^{\mathrm{cont}}(p,q) - k_{\mathrm{Lag}}^{\mathrm{wp}}(p,q) \right|,$$

then the triangle inequality yields

$$\left| k_{\mathrm{Lag}}^{\mathrm{cont}}(p,q) - \kappa(0) k_\sigma(p,q) \right| \le \varepsilon_{\mathrm{quad}}(p,q) + M_p M_q \left(C_1 \frac{\delta}{\ell} + C_2 \frac{\rho_{pq}^4}{\ell^4} \right).$$

□