Long-Term Autonomy for AUVs Operating Under Uncertainties in Dynamic Marine Environments

Abdullah Al Redwan Newaz[1], Tauhidul Alam[2], Gregory Murad Reis[3], Leonardo Bobadilla[3], and Ryan N. Smith[3]

***Abstract*—There has been significant interest in recent years in the utility and implementation of autonomous underwater and surface vehicles (AUVs and ASVs) for persistent surveillance of the ocean. Example studies include the dynamics of physical phenomena, e.g., ocean fronts, temperature and salinity profiles, and the onset of harmful algae blooms. For these studies, AUVs are presented with a complex planning and navigation problem to achieve autonomy lasting days and weeks under uncertainties while dealing with resource constraints. We address these issues by adopting motion, sensing, and environment uncertainties via a Partially Observable Markov Decision Process (POMDP) framework. We propose a methodology with a novel extension of POMDPs to incorporate spatiotemporally-varying ocean currents as energy and dynamic obstacles as environment uncertainty. Existing POMDP solutions such as the Cost-Constrained Partially Observable Monte-Carlo Planner (POMCP) do not account for energy efficiency. Therefore, we present a scalable Energy Cost-Constrained POMCP algorithm utilizing the predicted ocean dynamics that optimizes energy and environment costs along with goal-driven rewards. A theoretical analysis, along with simulation and real-world experiment results is presented to validate the proposed methodology.**

***Index Terms*—Long-term autonomy, autonomous underwater vehicle navigation, uncertainties, energy constraints.**

I. Introduction

MARINE robotic systems continue to increase their ability to operate independently for progressively longer periods. Existing systems have demonstrated robust, autonomous operations for multiple hours and even days. However, persistent (long-term) navigation capabilities will be critically important for future marine robots as they will be required to operate over periods of days to weeks. While current navigation and mapping algorithms can function over substantial spatial extents, it is currently unclear how to extend these to deal with human-scale spatial and temporal dimensions, as well as deal with the uncertainty of an ever-changing environment. As we look to extend our understanding of the Earth's changing environment, we require these marine robots and robotic systems to comprehend variability across large-scale spatiotemporal dimensions ($>$ 50 km^2 and days to weeks) while reacting to a locally dynamic and uncertain environment.

To enable long-term autonomy for AUVs, we need to handle three primary sources of uncertainties [1, 2]: i) motion uncertainty, which stems from the noise that affects system dynamics; ii) sensing uncertainty, results from noisy sensor measurements, which is also referred to as imperfect state information; and iii) environment uncertainty, caused by uncertain obstacle locations in the dynamic environment as illustrated in Fig 1.

Reducing these uncertainties in the marine environment presents several challenges. First, AUVs have a limited energy budget (limited battery life) to sustain long-term missions and are generally unable to simply stop and recharge. Second, the vehicle motion is significantly impacted by ocean currents, generally on the order of magnitude of the vehicle velocity. Third, the vehicle state can not be accurately estimated due to its error-prone underwater sensor measurements. Fourth, underwater navigation is likely to be affected by many uncertain dynamic obstacles such as ships, boats, etc. Despite these challenges, an AUV needs to successfully navigate the environment while avoiding collisions and maintaining course under the aforementioned sensing, motion, and environment uncertainties and resource constraints.

In this paper, we address the long-term autonomy under uncertainties along with resource constraints through a Partially Observable Markov Decision Process (POMDP) framework. We adopt motion, sensing, and environment uncertainties via the POMDP framework and provide a theoretical analysis. This idea is motivated through recent research efforts in AUV navigation [3, 4] that address motion uncertainty by framing the problem as a Markov Decision Process (MDP). Extending these works to include planning for vehicles under both motion and sensing uncertainties can generally be framed as a POMDP problem. However, very few prior studies (e.g., [5, 6]) utilize a POMDP framework for this application. The referenced research efforts lack rigorous algorithm and theoretic analysis, which we address in this study for applications to the marine environment.

As a novel extension to the existing literature, we introduce spatiotemporally-varying ocean currents as energy; an AUV can save energy by navigating with currents and consume excessive energy by navigating against currents. As such, we account for energy costs for ocean currents and introduce environment costs to avoid collisions with dynamic obstacles

Manuscript received: February 23, 2021; Revised April 25, 2021; Accepted May 29, 2021. This paper was recommended for publication by Editor Pauline Pounds upon evaluation of the Associate Editor and Reviewers' comments. This work is supported in part by the National Science Foundation awards IIS-2034123, IIS-2024733, by the U.S. Department of Homeland Security award 2017-ST-062000002 and by the MRI award 1531322.

[1]A. A. R. Newaz is with the Department of Electrical and Computer Engineering, North Carolina A&T State University, Greensboro, NC, USA (email: aredwannewaz@ncat.edu).

[2]T. Alam is with the Department of Computer Science, Louisiana State University Shreveport, Shreveport, LA, USA (email: talam@lsus.edu).

[3]G. M. Reis and L. Bobadilla are with the School of Computing and Information Sciences, and R. N. Smith is with the Institute of Environment, Florida International University, Miami, FL, USA (email: gregory@cs.fiu.edu; bobadilla@cs.fiu.edu; rysmith@fiu.edu).

Digital Object Identifier (DOI): see top of this page.

Fig. 1: An AUV policy execution conducted at North Beach, Miami, FL, USA. The policy is synthesized while accounting for motion, sensing, and environment uncertainties.

in our POMDP framework for safe, reliable planning under uncertainties. To accurately estimate energy costs, we predict spatiotemporal ocean dynamics using a deep learning model. Since our POMDP framework optimizes multiple objectives and it is not straightforward to utilize the Cost-Constrained POMDP (CC-POMDP) [7], so we extend our POMDP framework to an Energy Cost-Constrained POMDP (ECC-POMDP) framework that optimizes energy and environment costs along with goal-driven rewards.

To address the challenges indicated above, we make the following contributions in this paper: 1) We propose a recurrent neural network (RNN) based learning algorithm for predicting ocean dynamics in a continuous domain from real ocean current data; 2) We present an Energy Cost-Constrained Partially Observable Monte-Carlo Planner (ECC-POMCP) algorithm for solving the ECC-POMDP problem in a continuous state space of a marine environment under motion, sensing, and environment uncertainties that optimizes the trade-off among the rewards, the energy costs, and the collision costs; and 3) We analyze the optimality conditions theoretically for estimating the value function to our presented ECC-POMCP solution. Our theoretical analysis is validated extensively through simulations and physical experiments.

II. Related Work

Planning for AUV autonomy under uncertainties and time-varying ocean currents in marine environments is framed as a decision-theoretic problem in marine robotics. For instance, a graph search-based method [8] is presented to plan time and energy optimal paths in static and time-varying flow fields. This method did not address uncertainties in ocean current forecasting and motion. A non-linear robust model predictive control (NRMPC) [9] method is proposed to compute minimum energy paths subject to time-varying ocean currents and forecast model uncertainty. This method considers a bounded uncertainty and depends on the vehicle kinematic model. Uncertainties in ocean current predictions and navigation are utilized in two stochastic planners that find paths with minimum risk of collision in an ocean environment [10]. However, these planners lacked a proper uncertainty estimation method for ocean currents. This limitation was addressed in their subsequent work [11] by using Gaussian processes augmented with interpolation variance to measure confidence in the uncertainty of noisy predictions. They did not minimize energy consumption in their path planning for AUVs.

The action uncertainty due to ocean currents motivates the use of MDP for marine vehicles planning [12, 3]. One of the methods presented in [4] is built upon an MDP that computes a minimum expected energy policy for marine environments with uncertain and time-varying flow models. This method did not account for collision-free paths. Time-dependent transition probabilities and reward values are calculated on tractable reachable states for a time-varying MDP solution to tackle ocean disturbances that vary with time [13]. These MDP solutions are applicable to a discrete state-space representation, and their accuracy depends on the discretization resolution. To address this issue, an MDP solution in continuous state space is proposed in [14] by expressing the value function as a linear combination of basis functions and approximating the Bellman equation by a partial differential equation. Nonetheless, these MDP based solutions do not address sensing and environment uncertainties in a marine environment.

A sampling-based multi-goal motion planning approach with Monte-Carlo Tree Search (MCTS) computes roadmap tours that enable a robot to reach each goal while reducing the overall distance traveled and the number of times it recharges [15]. This discrete planning method accounts for the energy constraint but does not consider uncertainties in the robot's motion and sensing. On the other hand, an existing POMDP framework in marine robotics [5] considers only motion and sensor uncertainties in AUV navigation. In contrast, our work formulates a variant of the POMDP problem for uncertainties in motion, sensing, and environment along with energy constraints and plans policies for AUVs. To the best of our knowledge, this is one of the few studies that considers dynamic obstacles in marine environments for planning policies.

III. Preliminaries and Problem Formulation

This section begins by defining a decision-theoretic planning problem called POMDP. This definition leads to the formulation of our ECC-POMDP problem that tackles motion, sensing, and environment uncertainties along with energy constraints.

Planning under sensing and motion uncertainties for an AUV in a stochastic, dynamic marine environment is generally framed as a POMDP problem. This POMDP is defined by a tuple $\mathcal{P} = \langle X, U, Y, f, R, h, \gamma \rangle$, where X is a finite set of states x of the vehicle in the environment, U is a finite set of actions u of the vehicle, Y is a finite set of sensor observations y the vehicle may receive, $f(x, u, x') = p(x'|x, u)$ is a probabilistic state transition function, $h(x) = p(y|x, u)$ is a probabilistic observation function, $R(x, u) \in \mathbb{R}$ is an immediate reward for taking an action u in a state x, and $\gamma \in [0, 1)$ is a discount factor. Since the states are not fully observable in POMDPs, the vehicle keeps track of a finite set of belief states $b \in B$, where B is the belief space. A belief state b of the vehicle is defined as a posterior distribution over all possible states given all past actions and sensor observations $b_t(x) = p(x_t|u_0, \ldots, u_{t-1}, y_0, \ldots, y_t)$, which can be updated recursively via the Bayes rule: $b_{t+1}(x') = p(y|x', u) \sum_x p(x'|x, u) b_t(x)$. The POMDP can be considered

as the equivalent belief-state MDP $\langle B, U, T, R, \gamma \rangle$, where B represents the set of reachable beliefs over states in an MDP, $T(b'|b,u) = \sum_{y,x,x'} p(y|x',u)p(x'|x,u)b(x)\delta(b',b)$ is the belief transition function, and $R(b,u) = \sum_x b(x)R(x,u)$ is the immediate reward function on belief states. The solution to a POMDP problem is an optimal policy, π^*, that maximizes expected discounted reward as follows [16]:

$$\max_{\pi} V_R^{\pi}(b) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(b_t, \pi(b_t)|b_0)\right]. \tag{1}$$

Our problem entails to optimize multiple objectives such as goal-driven rewards, costs for dynamic obstacle avoidance, and energy awareness using ocean currents. The ECC-POMDP framework [7] is a generalized variant of POMDP for multi-objective problems. This ECC-POMDP is formally defined by a tuple $\mathcal{C} = \langle X, U, Y, f, R, h, C, \hat{c}, \xi, \hat{e}, \gamma, b_0 \rangle$, where C is a non-negative cost function with an individual threshold $\hat{c}$ and ξ is a non-negative energy function with an individual threshold $\hat{e}$. Similarly, an ECC-POMDP can be converted into an equivalent belief-state CMDP $\langle B, U, T, R, C, \hat{c}, \xi, \hat{e}, \gamma \rangle$, where $C(b,u) = \sum_x b(x)C(x,u)$ and $\xi(b,u) = \sum_x b(x)\xi(x,u)$. Our objective is to compute an optimal policy (solution) to an ECC-POMDP framework that maximizes the expected cumulative reward while bounding the expected cumulative costs and energies:

$$\max_{\pi} V_R^{\pi}(b) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(b_t, \pi(b_t)|b_0)\right] \tag{2}$$

subject to

$$V_C^{\pi}(b) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} C(b_t, \pi(b_t)|b_0)\right] \leq \hat{c},$$

$$V_{\xi}^{\pi}(b) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \xi(b_t, \pi(b_t)|b_0)\right] \leq \hat{e}.$$

To solve the formulated problem above, our methodology first learns to predict ocean dynamics in Section IV and then synthesizes an optimal policy for the ECC-POMDP framework in marine environments under uncertainties in Section V.

IV. Learning Ocean Dynamics

We rely on predicting ocean dynamics for a given time, depth, and geographical location to plan efficiently in marine environments. The existing Regional Ocean Model System (ROMS) [17] provides water current forecasts, which is a sparse prediction over our region of interest [9, 10, 11]. We use a historical dataset here, but this can be applied to recent predictions to enable operations going forward. Effective planning also requires continuous prediction and evaluation of ocean dynamics for which rewards and costs are computed. Therefore, we train a deep neural network to estimate ocean dynamics based on the ROMS dataset and use this model for evaluating the future cost of an action. We frame learning ocean dynamics as a multivariable time series prediction problem in a continuous domain and design a Recurrent Neural Network (RNN) to handle a sequence dependence among the input variables [18].

To learn a multivariable time series ocean dynamics, given an input sequence time series signal $\mu = (\mu_1, \mu_2, \cdots, \mu_T)$ with $\mu_t \in \mathbb{R}^n$, where n is the input variable dimension, our goal is to predict corresponding outputs $\nu = (\nu_1, \nu_2, \cdots, \nu_T)$ at each time with $\nu_t \in \mathbb{R}^m$, where m is the output variable dimension. The learning task is to train a deep RNN to obtain a feed-forward mapping, $\mathcal{F}$, of the prediction sequence from the current state as:

$$(\mu_1, \mu_2, \cdots, \mu_T) = \mathcal{F}(\nu_1, \nu_2, \cdots, \nu_T). \tag{3}$$

To predict time-varying ocean currents, we utilize the Long Short-Term Memory (LSTM) neural network, one of the RNN variants. Particularly, we design an LSTM network with a simple yet effective architecture to forecast ocean currents based on time-varying water current data. Our LSTM network reduces the time consumption overhead.

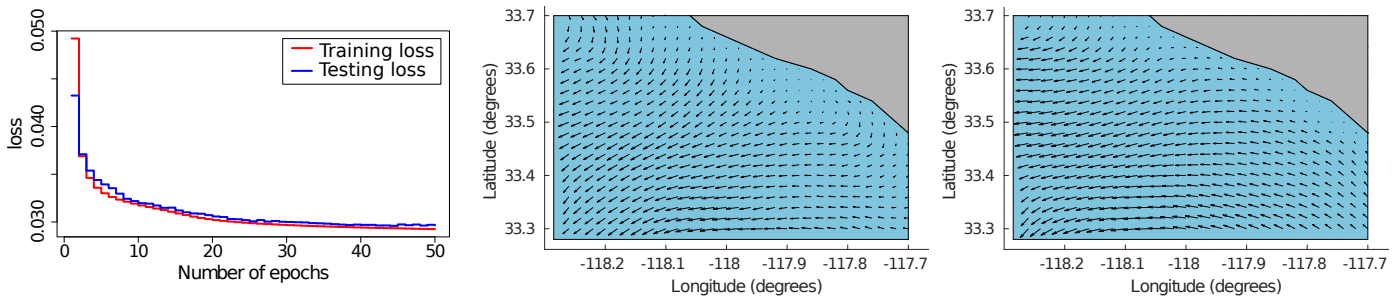
The proposed network structure is composed of LSTM blocks in two layers, followed by a fully connected output layer. Each LSTM block consists of 32 neurons followed by 2 neurons in the output layer. The input shape comprises 1 time step with 3 features such that $\mu_t = \langle \texttt{timestamp}, \texttt{depth}, \texttt{latitude}, \texttt{longitude} \rangle$, and the output is ocean current forecasts in $\mathbb{R}^2$ such that $\nu_t = (\texttt{horizontal\_current}, \texttt{vertical\_current})$. To train the network, we first frame the raw ROMS dataset to be used for a supervised learning problem and then transform this dataset into a trainable dataset by normalizing the input variables. We use the Mean Absolute Error (MAE) loss function and the Adam optimizer for stochastic gradient descent.

The training lasts for 50 epochs with a mini-batch of 72. Fig. 2a shows the training and testing losses on the ROMS dataset. After training, we use the trained model to forecast for the entire test dataset. To understand prediction over the test dataset, we invert the output from the normalization scale back into the original scale. Fig. 2b and Fig. 2c demonstrate the performance of our model. It is obvious from this result that our model can predict as close to ground truth ocean dynamics.

V. Planning in Marine Environments under Uncertainties

Our proposed planning framework consists of two interconnected elements to solve the constrained-POMDP problem with a continuous or large-scale state space. We initially transform this constrained-POMDP problem into an unconstrained problem via the introduction of Linear Programming (LP). Then we approximate the value function of this unconstrained-POMDP by an importance sampling-based algorithm. Once the importance distribution for sampling state transitions is determined, we can find a near-optimal policy for our multi-objective problem.

To find an optimal policy of our constrained-POMDP, we extend the Cost-Constrained POMCP (CC-POMCP) algorithm [7] to an Energy Cost-Constrained POMCP (ECC-POMCP) algorithm that handles both cost and energy in continuous POMDPs. The original CC-POMCP provides a guideline to formulate a constrained-POMDP to an unconstrained belief-state MDP with the scalarized reward function. Therefore, we can leverage a linear program (LP) to obtain an



(a) Learning curves (b) Ground truth ocean dynamics (c) Predicted ocean dynamics

Fig. 2: **Learning ocean dynamics:** (a) Training and testing losses across fifty epochs, (b) Ground truth ocean dynamics for the water surface layer at time 3-hour of the ROMS dataset, (c) Predicted ocean dynamics from a corresponding test input indicates that the model learns to predict ocean dynamics accurately.

optimal stochastic policy. Unfortunately, we cannot directly use the CC-POMCP for our problem since the CC-POMCP action selection rule does not have any notion of energy constraints. In particular, we have three primary factors in action selection: the expected total discounted reward $R(b,u)$, the expected total cost $C(b,u)$ and the expected total energy consumption $\xi(b,u)$. For example, in marine navigation tasks, the distance between the vehicle and the goal location is a useful factor because the vehicle gets a higher discounted reward as it approaches the goal location. The distance between the vehicle and the nearest obstacle is another useful factor, as opposed to rewards, the vehicle gets higher costs when its probability of hitting obstacles is higher. Finally, the vehicle's endurance improvement utilizing ocean currents is another important factor since going against the flow requires more energy consumption than usual. Thus, the derivation of our algorithm starts from optimizing the value function by looking for a policy taking the current belief state as input as follows

$$\min \sum_{b} \delta(b,b_0)V(b) - \lambda_c V_C^{\pi}(b) - \lambda_\xi V_\xi^{\pi}(b) + \lambda_c \hat{c} + \lambda_\xi \hat{e} \quad (4)$$

subject to

$$V(b) \geq R(b,u) - C(b,u)\lambda_c - \xi(b,u)\lambda_\xi + \gamma \sum_{b'} T(b' \mid b,u)V(b'),$$

where $\delta(b,b_0)$ is a Dirac delta function that has the value of 1 if $b = b_0$ and 0 otherwise. By treating λ_c and λ_ξ as constants, we can solve the above problem as an unconstrained belief-state MDP with the reward function $R(b,a) - \lambda_c \hat{c} - \lambda_\xi \hat{e}$. To handle the curse of dimensionality, our proposed policy iteration algorithm utilizes Monte-Carlo Tree Search (MCTS) to find the optimal action selection strategy effectively.

In this work, we introduce the ECC-POMCP, an approximate policy iteration algorithm, summarized in Algorithm 1. This ECC-POMCP utilizes MCTS in the belief space and the particle representation of belief states. In MCTS, each belief node contains a set of particles that resembles the corresponding approximated belief. MCTS is governed by two policies: a tree policy and a rollout policy. In the policy evaluation step, we apply the rollout policy to compute the intermediate belief-action function, or Q-value by simulating a look-ahead search on the POMDP model. The rollout policy guides the Monte

Algorithm 1 ECC-POMCP: Policy Iteration with Ocean Dynamics and Dynamic Obstacles

Input: Belief-state CMDP, $\langle B, U, T, R, C, \hat{c}, \xi, \hat{e}, \gamma \rangle$ and learned ocean model $\mathcal{F}$
Output: Optimal Policy, π_λ^*
1: Initialize λ_c and λ_ξ
2: **repeat**
3: Policy Evaluation for $\pi(u \mid b)$
4: For a given belief b and an action u, compute $Q_R(b,u), Q_C(b,u)$, and $Q_\xi(b,u;\mathcal{F})$
5: Obtain the joint belief-action value $Q_\lambda^{\oplus}(b,u)$
6: Policy Improvement on $\langle B, U, T, R - \lambda_c \hat{c} - \lambda_\xi \hat{e}, \gamma \rangle$
7: Update the policy based on the joint action-value as:

$$\pi_\lambda^* = \arg\max_{u \in U(h)} Q_\lambda^{\oplus}(b,u)$$

8: **until** Converge()

Carlo simulation toward a promising subspace by utilizing domain knowledge. After each simulation, the belief-action pairs for reward, cost, and energy are updated as follows:

$$\begin{aligned} Q_R(b,u) &= Q_R(b,u) + \frac{R - Q_R(b,u)}{N(b,u)} \\ Q_C(b,u) &= Q_C(b,u) + \frac{C - Q_C(b,u)}{N(b,u)} \\ Q_\xi(b,u;\mathcal{F}) &= Q_\xi(b,u;\mathcal{F}) + \frac{\xi - Q_\xi(b,u;\mathcal{F})}{N(b,u)}, \end{aligned} \quad (5)$$

where $N(b)$ is the number of simulations performed through b, $N(b,u)$ is the number of times action u is selected in b, In contrast, in the policy evaluation step, the tree policy selects an action utilizing the Partially Observable Upper Confidence Bounds for Trees (PO-UCT) algorithm [16] as follows:

$$\begin{aligned} \arg\max_{u \in U(b)} Q_\lambda^{\oplus}(b,u) &= Q_R(b,u) - \lambda_c Q_C(b,u) \\ &\quad - \lambda_\xi Q_\xi(b,u;\mathcal{F}) + \epsilon \sqrt{\frac{\log N(b)}{N(b,u)}}, \end{aligned} \quad (6)$$

where ϵ is a constant that balances exploration and exploitation for a search algorithm.

It is important to note that our algorithm is scalable since it inherits the scalability of the CC-POMCP algorithm. Like the

CC-POMCP, we generate samples for rewards and collision costs from the POMDP model. However, unlike the original CC-POMCP, we generate samples for energy costs from our learned ocean dynamics model $\mathcal{F}$.

VI. Theoretical Analysis

The computational load of Algorithm 1 can be divided into three parts. First, during policy evaluation, a POMCP tree of depth D contains only $O(|\overline{U}|^D|\overline{Y}|^D)$ nodes, where $|\overline{U}|$ and $|\overline{Y}|$ are the sizes of the action subset and the observation subset such that $\overline{U} \subset U$ and $\overline{Y} \subset Y$, respectively. Second, after learning the ocean dynamics, the prediction over ocean dynamics can be obtained with a constant $O(1)$ time complexity. Finally, the selection of joint action-value can be obtained using linear programming which has $O(|\overline{U}|)$ time complexity [7].

Next, we will analyze the optimality conditions for estimating the value function using Linear Programming and the approximation errors of our algorithm to provide a bound on its performance.

Lemma 1. *Let $\mathcal{M}_1 = \langle B, U, T, R_1, \gamma\rangle$, $\mathcal{M}_2 = \langle B, U, T, R_2, \gamma\rangle$, and $\mathcal{M}_3 = \langle B, U, T, R_3, \gamma\rangle$ be three (belief-state) MDPs differing only in the reward function, and V_1^π, V_2^π and V_3^π be their corresponding value functions under a fixed policy π, respectively. Then, the value function of the joint MDP $\mathcal{M} = \langle B, U, T, qR_1 + \hat{q}R_2 + \bar{q}R_3, \gamma\rangle$, with the policy π is $V^\pi(b) = qV_1^\pi(b) + \hat{q}V_2^\pi(b) + \bar{q}V_3^\pi(b)$ for all $b \in B$.*

Proof. We use the mathematical induction to prove this Lemma motivated by the idea in [7].
Base step for all b

$$\begin{aligned} V^0(b) &= \sum_u \pi(a \mid b)[qR_1(b,u) + \hat{q}R_2(b,u) + \bar{q}R_3(b,u)] \\ &= q\sum_u \pi(u \mid b)R_1(b,u) + \hat{q}\sum_u \pi(u \mid b)R_2(b,u) \\ &+ \bar{q}\sum_u \pi(u \mid b)R_3(b,u) \\ &= qV_1^0(b) + \hat{q}V_2^0(b) + \bar{q}V_3^0(b) \end{aligned}$$

Inductive step for all b

$$\begin{aligned} V^{k+1}(b) &= \sum_u \pi(u \mid b)[qR_1(b,u) + \hat{q}R_2(b,u) + \bar{q}R_3(b,u) \\ &+ \gamma T\sum_{b'}(b'|b,u)V^k(b')] \\ &= q\sum_u \pi(u \mid b)\left[R_1(b,u) + \gamma T\sum_{b'}(b'|b,u)V^k(b')\right] \\ &+ \hat{q}\sum_u \pi(u \mid b)\left[R_2(b,u) + \gamma T\sum_{b'}(b'|b,u)V^k(b')\right] \\ &+ \bar{q}\sum_u \pi(u \mid b)\left[R_3(b,u) + \gamma T\sum_{b'}(b'|b,u)V^k(b')\right] \\ &= qV_1^{k+1}(b) + \hat{q}V_2^{k+1}(b) + \bar{q}V_3^{k+1}(b) \end{aligned}$$

Thus,

$$\begin{aligned} V^\pi(b) &= \lim_{k\to\infty} V^k(b) \\ &= \lim_{k\to\infty}\left(V_1^k(b) + V_2^k(b) + V_3^k(b)\right) \\ &= qV_1^\pi(b) + \hat{q}V_2^\pi(b) + \bar{q}V_3^\pi(b) \end{aligned}$$

□

Theorem 1. *For a given belief set B and a piecewise-linear and convex (PWLC) value function $V_\lambda^\pi(b)$ of the belief b such that for all $b \in B$, the optimal policy π_λ^* can be obtained using Linear Programming (LP).*

Sketch of the proof. An important property of PWLC functions is that sum and max operators also preserve the convexity property. The value function $V_\lambda^\pi(b)$ can be decomposed into three parts: an approximated piecewise linear reward function, an approximated piecewise linear cost function and an approximated piecewise linear energy function. Since the value function $V_\lambda^\pi(b)$ is computed using sum and max operations over these PLWC functions, the value function $V_\lambda^\pi(b)$ also preserves the convexity property. Thus, given the PWLC representations of reward, cost and energy functions, for any λ, we can obtain a corresponding unique π_λ^* by maximizing $V_\lambda^\pi(b)$ using LP [7, 19]. □

We approximate our value function $V_\lambda^\pi(b)$ utilizing a weighted particle filter. To formally prove the error bound of the approximated value function, we modify a generic convergence result for particle filtering introduced in [20, 21].

Theorem 2. *Let $\varepsilon_{t|t}(\epsilon)$ be the random perturbation function with a constant error ϵ. Let V_λ^* be the value function of the optimal policy π^*. The error introduced by the proposed algorithm is bounded as follows*

$$\mathbb{E}\left[\left(\tilde{V}_\lambda^\pi(b) - V_\lambda^\pi(b)\right)^2\right]^{\frac{1}{2}} \leq \frac{\sqrt{\eta} + \sqrt{\varepsilon_{t|t}(\epsilon)}}{\sqrt{N}}\left(V_\lambda^*\right),$$

where η is a constant, $\tilde{V}_\lambda^\pi(b)$ is the approximate value function, and N is the sample size.

Proof. One has

$$\tilde{V}_\lambda^\pi(b) - V_\lambda^\pi(b) = \tilde{V}_\lambda^\pi(b) - \hat{V}_\lambda^\pi(b) + \hat{V}_\lambda^\pi(b) - V_\lambda^\pi(b). \quad (7)$$

Then the Minkowski's inequality gives

$$\begin{aligned} &\mathbb{E}\left[\left(\tilde{V}_\lambda^\pi(b) - V_\lambda^\pi(b)\right)^2\right]^{\frac{1}{2}} \\ &\leq \mathbb{E}\left[\left(\tilde{V}_\lambda^\pi(b) - \hat{V}_\lambda^\pi(b)\right)^2\right]^{\frac{1}{2}} + \mathbb{E}\left[\left(\hat{V}_\lambda^\pi(b) - V_\lambda^\pi(b)\right)^2\right]^{\frac{1}{2}} \end{aligned} \quad (8)$$

Let G_t be the $\sigma-$field generated by $\{x_{t|t-1}^{(i)}\}_{i=1}^N$ particles, then using multinomial resampling we can obtain

$$\mathbb{E}\left[\left(\tilde{V}_\lambda^\pi(b)\right)\middle| G_t\right] = \hat{V}_\lambda^\pi(b) \quad (9)$$

and

$$\begin{aligned} &\mathbb{E}\left[\left(\tilde{V}_\lambda^\pi(b) - \mathbb{E}\left[\tilde{V}_\lambda^\pi(b)\middle| G_t\right]\right)^2\middle| G_t\right] \\ &= \mathbb{E}\left[\left(\tilde{V}_\lambda^\pi(b) - \hat{V}_\lambda^\pi(b)\right)^2\middle| G_t\right] \\ &= \frac{1}{N}\left(\tilde{V}_\lambda^\pi(b)^2 - \left(\tilde{V}_\lambda^\pi(b)\right)^2\right) \\ &\leq \frac{\eta}{N}\left(V_\lambda^*\right)^2 \end{aligned} \quad (10)$$

By substituting the right hand side of Eqn. (8), we can get

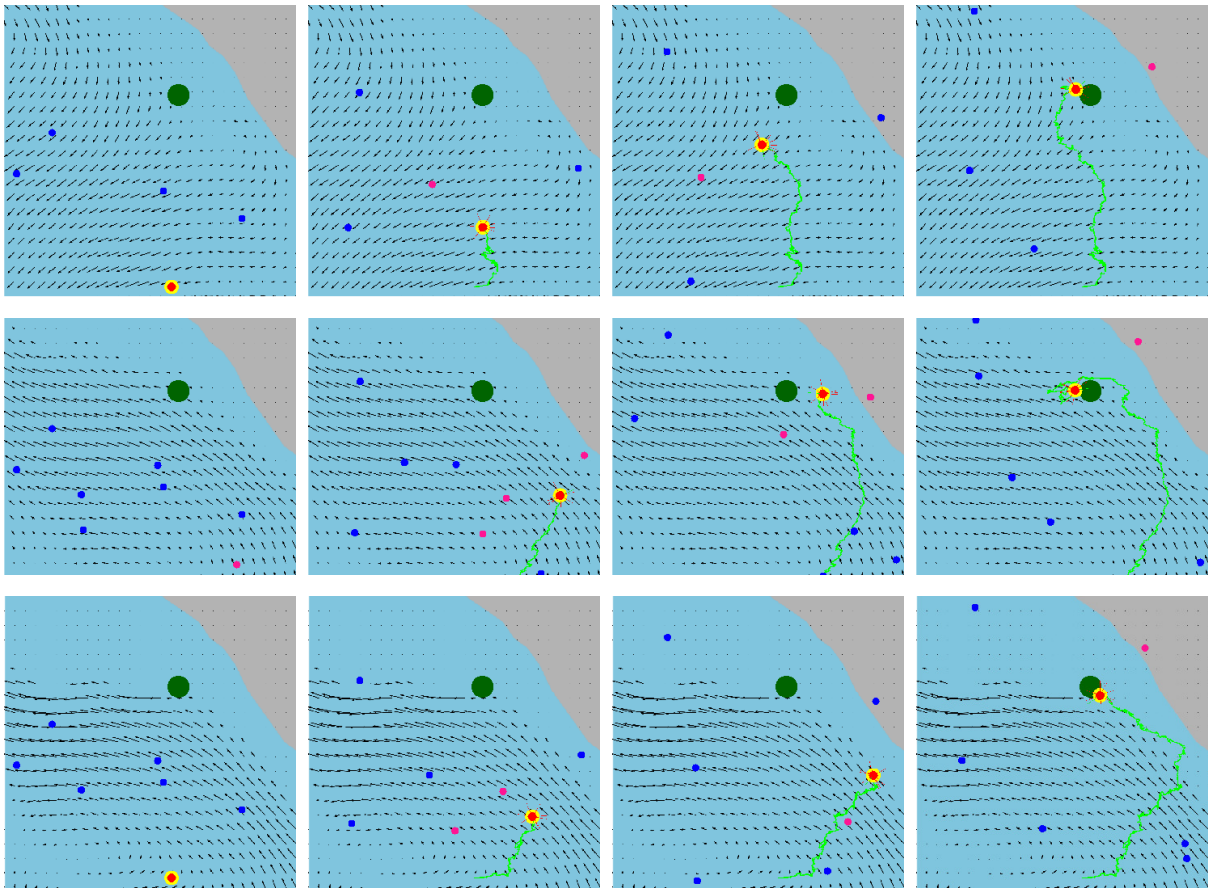


Fig. 3: **Simulation of our ECC-POMCP algorithm:** Executed trajectories delineated with the green lines of the vehicle (red circle) from its initial location to the goal location (green circle) applying the synthesized policies for the first water current layer (0 m depth) in the top row and for the sixth water current layer (30 m depth) in the middle row and for the eleventh water current layer (100 m depth) in the bottom row. The blue circles correspond with dynamic obstacles and magenta circles are potential collidable obstacles. The red lines around the vehicle represent a set of preferred actions of a belief state.

$$\mathbb{E}\left[\left(\tilde{V}_{\lambda}^{\pi}(b)-V_{\lambda}^{\pi}(b)\right)^{2}\right]^{\frac{1}{2}} \leq \frac{\sqrt{\eta}}{\sqrt{N}}\left(V^{*}\right)+\frac{\sqrt{\eta}}{\sqrt{N}}\left(V_{\lambda}^{*}\right) \leq \frac{\sqrt{\eta}+\sqrt{\varepsilon_{t|t}(\epsilon)}}{\sqrt{N}}\left(V_{\lambda}^{*}\right) \tag{11}$$

□

VII. Experiments

To validate the effectiveness of our methodology this section presents experiments in both simulated and real-world marine environments under uncertainties. We conduct our simulations using a general-purpose laptop with an Intel Core i7 (Eight Generation) with 16GB DDR4 RAM, running the Ubuntu 20.04 LTS operating system.

A. Simulation Results

We use the ROMS [17] ocean current data observed in the Southern California Bight (SCB) region to evaluate our method. We learn our predicted ocean model at continuous locations from the ROMS data that provides ocean currents at discrete locations. The 3-D ocean environment is taken into account as a simulated environment for the vehicle movements having 2-D ocean surfaces at different water current layers or depths (e.g., 0 m, 5 m, 10 m, 15 m, 20 m, 30 m, 40 m, and so on). Each 2-D ocean current layer is tessellated into a grid map. Each tessellated water current layer is a 21×29 grid map with a horizontal spatial resolution of 1 km$\times$1 km. We assume that the AUV is equipped with ultrasonic range sensors and capable of detecting near obstacles. Thus, the AUV detects obstacles based on the sensor measurements, where the Gaussian white noise is incorporated to encapsulate sensing uncertainty.

We implement the particle filter for planning under uncertainties and constraints for many water current layers from our ROMS ocean current prediction data. The particle filter utilizes a dead-reckoning method in the absence of sensor measurements. However, we consider that the AUV periodically visits the water surface to keep the uncertainty tractable. From our algorithm implementation, we obtain a set of policies as output from the layer-wise policy synthesis. Fig. 3 illustrates the executed trajectories of the vehicle applying the synthesized policies that avoid dynamic obstacles for the same pair of given initial and goal locations at different water current layers.

We examine the trajectories of an AUV under different ocean current layer, time, and strength. In Fig. 3, the red circle, the green circle, and the green lines represent the AUV, the

goal location, and the executed trajectories, respectively. We overlay flow fields for demonstrating ocean currents, where the sky-blue region represents the navigable water area, and the gray region represents the land area. The blue dots represent unknown obstacles, navigating randomly in the area. The pink dots represent the states when obstacles are detected by the AUV with a higher confidence. As we increase the number of obstacles in the scene, the AUV spends more time to navigate to the goal location on the same flow field.

Our method attempts to balance the travel time, avoiding possible collisions and energy consumptions under uncertainties. It prioritizes safety by avoiding collisions and then leverages the ocean current to advance to the goal location. As we can also observe from Fig. 3, the ECC-POMCP policies usually take a longer path instead of taking a straight direct path toward the goal location. This is because it is preferable to utilize the direction of ocean currents to minimize energy consumption. Furthermore, the ECC-POMCP policies can leverage the fact that following in the same direction of the ocean current allows the AUV to obtain a faster net speed, resulting in a shorter time to reach the goal location.

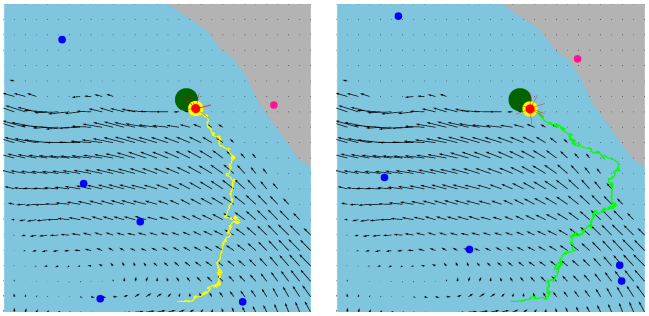
B. Performance Analysis

We compare our ECC-POMCP algorithm with the baseline CC-POMCP algorithm [7]. Table I demonstrates the efficacy of the proposed ECC-POMCP algorithm. To understand how the ECC-POMCP can help energy efficiency via minimizing the mission time, we perform three experiments for different water current layers or depths (0 m, 30 m, and 100 m) in the absence of obstacles. We use a constant acceleration controller for this purpose. Therefore, the AUV navigates with a lower average speed when it moves against the water current and vice-versa. We observe that the trajectories generated by the ECC-POMCP avoid navigating against the strong current and follow longer but energy-efficient paths based on the Eqn. (6). Hence, the AUV is capable of navigating with higher velocity while utilizing the water currents to reach the goal location, resulting in faster mission completion times. Fig. 4 demonstrates the trajectories of the CC-POMCP and ECC-POMCP algorithms under varying water current layers and strengths. As we can see in Fig 4, the CC-POMCP algorithm generates a shorter trajectory while the ECC-POMCP algorithm generates a longer but energy-efficient trajectory utilizing water currents.

Exp.	Traj. Length (m)		Avg. Velocity (m/s)		Mission Time (min)	
	ECC	CC	**ECC**	CC	**ECC**	CC
1	469.75	456.76	0.547	0.452	14.30	16.82
2	685.38	456.76	0.624	0.375	18.28	20.28
3	523.60	456.76	0.559	0.440	15.59	17.29

TABLE I: Trajectory length, average velocity, and mission completion time for different water current layers (depths) in absence of obstacles. In contrast to the CC-POMCP, the AUV can navigate with higher velocity while following the ECC-POMCP policy. Hence, even if the trajectory is longer, the overall mission completion time is faster.

Table II presents the performance comparison between CC-POMCP and ECC-POMCP algorithms in terms of trajectory lengths, policy synthesis times, and average step rewards.



(a) CC-POMCP trajectory (b) ECC-POMCP trajectory

Fig. 4: Yellow and green lines represent trajectories of CC-POMCP and ECC-POMCP policies, respectively. A CC-POMCP policy generates a shorter trajectory while an ECC-POMCP policy generates an energy efficient trajectory utilizing water current.

Here we quantitatively evaluate the average step rewards over trajectories, trajectory lengths, and policy synthesis times of two algorithms subject to different water current strengths and different numbers of dynamic obstacles. The resultant policies of the ECC-POMCP algorithm achieve higher average rewards with the cost of spending more synthesis time to utilize water currents to reach the goal location. Conversely, the resultant policies of the CC-POMCP algorithm take a shorter path with lower average rewards due to the lack of energy efficiency.

C. From Simulation to Physical Implementation

We extend our marine navigation simulations in a real-world scenario using the YSI Ecomapper [22], an AUV that can navigate up to 7.408 km/h in speed and up to 100 meter depth. The AUV performs its trajectory at the bay surface in January 2021 starting from the initial location (lat. 25.9128625°, long. −80.1378406666667°) to the goal location (lat. 25.9115051396367°, long. −80.1371944635927°). The experiment is conducted off the coast of North Beach, Miami Beach, FL, USA, in a region surrounded by mangroves with shallow and clear water. The environment consists of navigable water areas and virtual obstacles, representing dynamic obstacles such as boats, shipwrecks, tree debris, and so on. The AUV also conducts its mission at 3.704 km/h, for around 30 minutes. We use a DJI Mavic Pro drone to track the trajectory of the AUV autonomously. As the AUV navigates, it localizes with a noisy GPS sensor and its belief states are estimated using a particle filter. We compute an offline policy in the presence of dynamic obstacles and water current. Fig. 5 demonstrates our policy on the water surface layer and overlays the AUV trajectory applying the policy on a GeoTIFF image. In Fig. 5, the top row represents a simulated trajectory from our computed offline policy, and the bottom row represents the real-world execution of the trajectory through an AUV mission.

VIII. Conclusion

We present a methodology with a novel extension of the POMDP framework for AUVs with resource constraints operating under motion, sensing, environment uncertainties

Water Current Layer or Depth (m)	Number of Obstacles	Average Step Reward		Trajectory Lengths (km)		Policy Synthesis Time (s)	
		ECC	CC	**ECC**	CC	**ECC**	CC
1 (0)	3	0.695391	0.626621	70.701	53.047	11.8763	8.6283
1 (0)	4	0.641426	0.626165	64.123	51.011	13.2603	9.0808
6 (30)	4	0.646165	0.490734	86.306	51.012	14.5801	9.058
6 (30)	8	0.618052	0.445625	79.030	83.022	14.3976	15.3926
11 (100)	4	0.646165	0.586158	55.253	51.012	9.7763	8.9387

TABLE II: Trajectory lengths, average step rewards, and policy synthesis times for different numbers of obstacles and ocean current variability at different depths. Even though synthesizing optimal policies using the CC-POMCP algorithm takes less times and trajectory lengths, the policies generated by our ECC-POMCP algorithm outperform the CC-POMCP's policies in terms of average step rewards. This is because the CC-POMCP algorithm ignores energy costs while synthesizing policies.

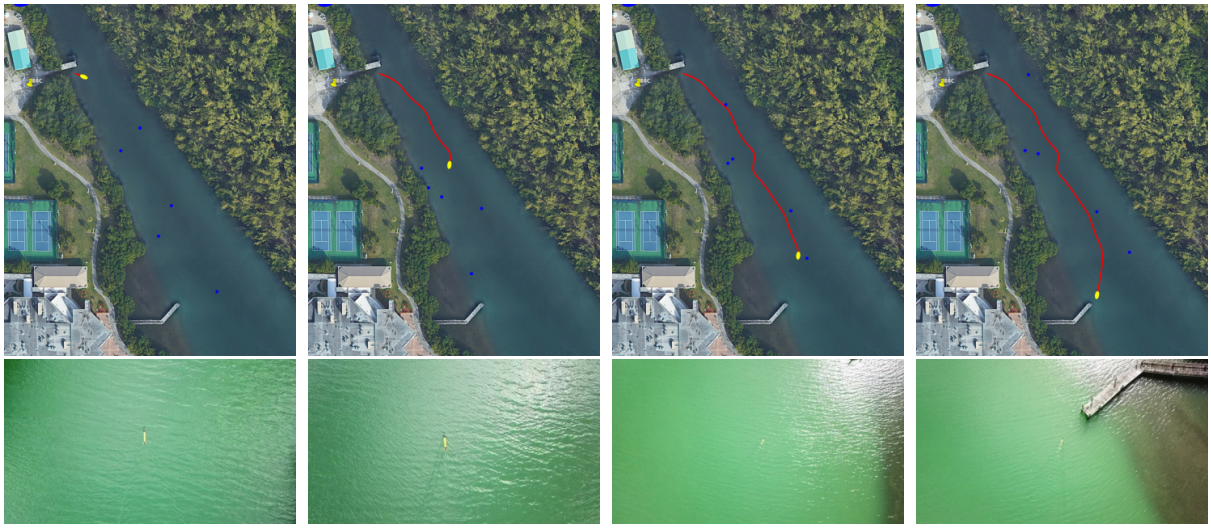


Fig. 5: (top) A simulated trajectory (red line) from a synthesized offline policy on the water surface layer with virtual dynamic obstacles (blue circles). (bottom) The trajectory execution of the offline policy with an AUV mission off the coast of Miami Beach, FL, USA.

in dynamic marine environments. First, a recurrent neural network based learning algorithm is proposed to predict ocean dynamics in a continuous domain. Second, an ECC-POMCP algorithm in a continuous state space utilizing the predicted ocean dynamics is presented to synthesize the optimal policy as a solution to the ECC-POMDP problem. Third, we provide a complexity analysis of the value function to guarantee its optimality along with an approximation error bound for the ECC-POMCP algorithm. Finally, we validate the effectiveness of our methodology through simulations and experiments.

REFERENCES

[1] A.-A. Agha-Mohammadi, S. Chakravorty, and N. M. Amato, "FIRM: Sampling-based feedback motion-planning under motion uncertainty and imperfect measurements," *Int. J. Robot. Res.*, vol. 33, no. 2, pp. 268–304, 2014.

[2] Y. Wang, A. A. R. Newaz, J. D. Hernández, S. Chaudhuri, and L. E. Kavraki, "Online partial conditional plan synthesis for POMDPs with safe-reachability objectives: Methods and experiments," *IEEE Trans. Autom. Sci. Eng.*, 2021.

[3] L. Liu and G. S. Sukhatme, "A solution to time-varying Markov decision processes," *IEEE Robot. Autom. Lett.*, vol. 3, no. 3, pp. 1631–1638, 2018.

[4] D. Kularatne, H. Hajieghrary, and M. A. Hsieh, "Optimal path planning in time-varying flows with forecasting uncertainties," in *Proc. IEEE Int. Conf. Robot. Autom.*, pp. 4857–4864, 2018.

[5] H. Kurniawati and V. Yadav, "An online POMDP solver for uncertainty planning in dynamic environment," in *Proc. Int. Symp. Robot. Res.*, pp. 611–629, 2016.

[6] T. Alam, A. Al Redwan Newaz, L. Bobadilla, W. H. Alsabban, R. N. Smith, and A. Karimoddini, "Towards energy-aware feedback planning for long-range autonomous underwater vehicles," *Frontiers Robot. AI*, vol. 8, p. 7, 2021.

[7] J. Lee, G.-H. Kim, P. Poupart, and K.-E. Kim, "Monte-Carlo tree search for constrained POMDPs," in *Proc. Advances Neural Inf. Process. Syst.*, pp. 7923–7932, 2018.

[8] D. Kularatne, S. Bhattacharya, and M. A. Hsieh, "Going with the flow: a graph based approach to optimal path planning in general flows," *Auto. Robots*, vol. 42, no. 7, pp. 1369–1387, 2018.

[9] V. T. Huynh, M. Dunbabin, and R. N. Smith, "Predictive motion planning for AUVs subject to strong time-varying currents and forecasting uncertainties," in *Proc. IEEE Int. Conf. Robot. Autom.*, pp. 1144–1151, 2015.

[10] A. A. Pereira, J. Binney, G. A. Hollinger, and G. S. Sukhatme, "Risk-aware path planning for autonomous underwater vehicles using predictive ocean models," *J. Field Robot.*, vol. 30, no. 5, pp. 741–762, 2013.

[11] G. A. Hollinger, A. A. Pereira, J. Binney, T. Somers, and G. S. Sukhatme, "Learning uncertainty in ocean current predictions for safe and reliable navigation of underwater vehicles," *J. Field Robot.*, vol. 33, no. 1, pp. 47–66, 2016.

[12] T. Alam, G. M. Reis, L. Bobadilla, and R. N. Smith, "A data-driven deployment and planning approach for underactuated vehicles in marine environments," *IEEE J. Ocean. Eng.*, vol. 46, no. 2, pp. 372–388, 2021.

[13] J. Xu, K. Yin, and L. Liu, "Reachable space characterization of Markov decision processes with time variability," in *Proc. Robot.: Sci. Syst.*, 2019.

[14] J. Xu, K. Yin, and L. Liu, "State-continuity approximation of Markov decision processes via finite element methods for autonomous system planning," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 5589–5596, 2020.

[15] Y. Warsame, S. Edelkamp, and E. Plaku, "Energy-aware multi-goal motion planning guided by Monte Carlo search," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, pp. 335–342, 2020.

[16] D. Silver and J. Veness, "Monte-Carlo planning in large POMDPs," in *Proc. Advances Neural Inf. Process. Syst.*, pp. 2164–2172, 2010.

[17] A. F. Shchepetkin and J. C. McWilliams, "The Regional Oceanic Modeling System (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model," *Ocean Model.*, vol. 9, no. 4, pp. 347–404, 2005.

[18] J. A. Caley and G. A. Hollinger, "Environment prediction from sparse samples for robotic information gathering," in *Proc. IEEE Int. Conf. Robot. Autom.*, pp. 10577–10583, 2020.

[19] M. Araya-López, O. Buffet, V. Thomas, and F. Charpillet, "A POMDP extension with belief-dependent rewards," in *Proc. Advances Neural Inf. Process. Syst.*, pp. 64–72, 2010.

[20] X.-L. Hu, T. B. Schon, and L. Ljung, "A basic convergence result for particle filtering," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1337–1348, 2008.

[21] D. Crisan and A. Doucet, "A survey of convergence results on particle filtering methods for practitioners," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 736–746, 2002.

[22] YSI Incorporated, "YSI EcoMapper." https://www.ysi.com/ecomapper, 2021.