

Comparison of Deep Reinforcement Learning and Model Predictive Control for Real-Time Depth Optimization of a Lifting Surface Controlled Ocean Current Turbine

Arezoo Hasankhani¹, Yufei Tang¹, James VanZwieten², Cornel Sultan³

Abstract—This paper evaluates two strategies, deep reinforcement learning (DRL) and model predictive control (MPC), for maximizing harnessed power from a lifting surface controlled ocean current turbine (OCT) through depth optimization. To address spatiotemporal uncertainties in the ocean current, an online Gaussian Process (GP) is applied, where the prediction error of the ocean current speed is also modeled. We compare the performance of the MPC-based optimization with the DRL-based algorithm (i.e., deep Q-networks (DQN)) using over one week of field collected acoustic doppler current profiler (ADCP) data. The DRL-based algorithm is almost equivalent to the MPC-based algorithm in real-time optimization when the ocean current speed prediction is perfect. However, the performance of the DQN-based algorithm surpasses the MPC-based algorithm when ocean current prediction error is considered. The importance of using the DQN in improving the error-tolerance of the proposed spatiotemporal optimization is verified through the comparative results.

I. INTRODUCTION

To address the increased interest in renewable energy resources due to environmental concerns, marine renewable energy has drawn growing attention. There is high power potential in the Gulf Stream ocean current (i.e., 18.6 GW), with the Florida Current share equal to 5.1 GW [1]. Ocean current turbine (OCT) technologies are being developed to harness the high power potential, where the harnessed power is depth dependent, with the highest time averaged power densities (exceeding 3.0 kW/m^2 in places) located within the top 100 m of the ocean [2]. To locate the OCT near the ocean surface, several technologies have been proposed: variable buoyancy [3]–[5], lifting surface [6], [7], sub-sea winches [8], and surface buoys [9]. In this paper, we focus on lifting surface controlled OCT.

Assuming that the OCT is controlled and navigated through lifting surface technology, spatiotemporal optimization should be developed to determine the optimal ocean depth. Hence, a hierarchical depth optimization and control structure is proposed to maximize the harnessed power from the lifting surface controlled OCT, where the upper level is

defined as the real-time spatiotemporal optimization, and the lower level is determined as the flight controller developed by the lifting surface technology. A similar structure has been proposed for the airborne wind turbine [10], and autonomous underwater vehicle [11]. In this paper, we will focus on the upper level control structure (i.e., spatiotemporal optimization), assuming that the lower level control (i.e., flight control operated by the lifting surface) can navigate to the optimal depth found through the spatiotemporal optimization.

Model predictive control (MPC) is considered in this paper because of its ability to handle highly constrained problems. A large body of literature has investigated the usage of the MPC algorithm for conducting real-time optimization and finding the optimal control solutions [12]. An MPC-based algorithm has been proposed to maximize the harnessed power through finding the optimal flying altitude of the airborne wind turbine [10], as well as the optimal water depth for the autonomous underwater vehicle [11]. The real-time trajectory sequence for autonomous vehicles has been found through the convex quadratic programming-based MPC to avoid collision [13]. The navigation problem of micro aerial vehicles has been addressed by non-linear MPC to avoid obstacles and find the aerial trajectory [14]. To address a helicopter landing and touchdown on ships, two approaches have been proposed based on a shrinking horizon MPC method [15] and an infinite MPC method [16]. Further, the real-time trajectory of a buoyancy controlled OCT has been addressed through a basic MPC algorithm [17]. To maximize the harnessed energy from the wave energy converter, the non-linear MPC algorithm has been used to find the generator's optimal control sequence [18]. It should be noted that MPC-based algorithms are model-based, which guarantee the theoretical assurance of finding the optimal control solution for real-time optimization problems. Still, the MPC-based method is sensitive to the prediction errors in the environment model.

On the other hand, reinforcement learning (RL) can be used to find the optimal control sequence by training using historical data (i.e., data-driven learning), which could be robust to the environment model errors. Among different RL-based approaches, the recently developed deep reinforcement learning (DRL) has gained increased attention due to its superior feature representation capability and significant performance. DRL-based approaches have been applied to address various real-time optimization problems, such as active object detection for the intelligent robotic application

This work was supported by the National Science Foundation under Grant Nos. ECCS-1809164, ECCS-1809404, & OAC-2017597 and the U.S. Department of Energy under Grant No. DE-EE0008955.

¹Department of Computer Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA. {ahasankhani2019, tangy}@fau.edu.

²Department of Civil, Environmental and Geomatics Engineering, Florida Atlantic University, Boca Raton, FL 33431, USA. jvanzwi@fau.edu.

³Department of Aerospace and Ocean Engineering, Virginia Tech, 460 Old Turner St., Blacksburg, VA 24061, USA. csultan@vt.edu.

[19], real-time control of a traffic signal [20], and especially robot navigation [21]. In this paper, we will focus on the real-time depth optimization of a lifting surface controlled OCT, aiming to find the optimal turbine operation depth. In similar applications, autonomous navigation of an unmanned ship has been addressed through DRL in an uncertain ocean environment [22]. Compared to [23], the optimal control sequence for unmanned surface vehicle navigation has been found through a DRL-based algorithm, enhancing protective capability and less sensitivity to the environmental uncertainties [24]. Furthermore, the real-time navigation of an unmanned surface vehicle has been addressed through DRL to avoid the collision under uncertainties arisen by the dynamic obstacles [25].

Lifting surface controlled OCT depth optimization will seek to maximize the harvested electrical power subject to several constraints, such as feasible turbine operation depth range and turbine relocation speed. MPC and DRL are both right-minded candidate methods to address highly constrained optimization problems. The main contribution of this paper is to present quantitative comparisons between the DRL-enabled, learning-based optimization and MPC-enabled, model-based optimization for lifting surface controlled OCT depth optimization. The robustness of these two methodologies is verified in the cases of no error and error existence in the ocean current prediction using over one week of field collected acoustic Doppler current profiler data from the Gulf Stream.

The rest of this paper is organized as follows. Section II presents the spatiotemporal ocean current modelling. Section III describes our proposed methodologies based on the MPC and DRL algorithms. Section IV presents the simulation results and discussions, and Section V draws the conclusions and future works.

II. MODELING SPATIOTEMPORAL OCEAN CURRENT

To model the spatiotemporal uncertainties in the ocean current, it is important to use real data. The spatial and temporal fluctuations in the current flow are resulted from turbulence, waves, and lower frequency flow structures. Prior to discussing the modeling procedure, we define the notations used to characterize the ocean current. In this regard, the “measured (observed) ocean current” denoted by V should not be confused with the “predicted ocean current” V^* . A “prediction error” e is then added to the predicted ocean current, thereby defining the “estimated ocean current” $\hat{V}$.

In this paper, we use the field measured data by a 75 kHz acoustic Doppler current profiler (ADCP) presented in [2], which were recorded at a latitude of $26.09^\circ N$ and longitude of $-79.80^\circ E$, as shown in Fig. 1. The measurement resolution was 5 m within 400 m water depth, where various current flow data (i.e., current speed, northward current velocity, eastward current velocity, etc.) were recorded. These measured current data were filtered to remove bad data as described in [2], which were mostly measured above a depth of 50 m.

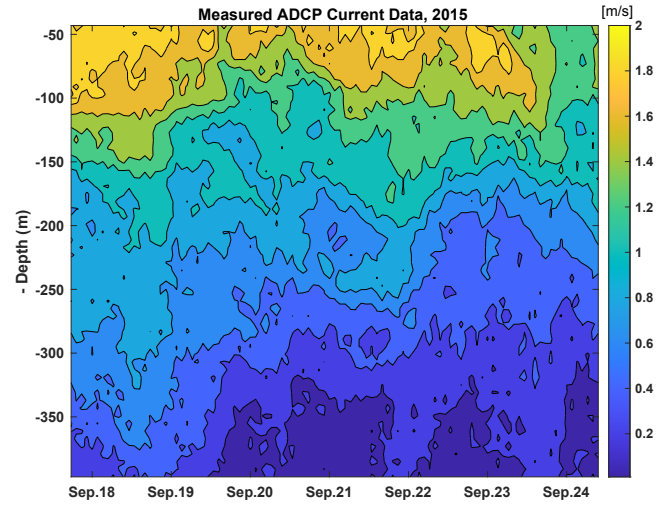


Fig. 1. Time histories of the current flow recorded by a 75 kHz ADCP at a latitude of $26.09^\circ N$ and longitude of $-79.80^\circ E$ [2].

To optimize the trajectory of the turbine and maximize harnessed power, we model spatiotemporal uncertainties and predict future ocean currents. In this paper, we use an online Gaussian process (GP) model to predict the future ocean current, where the GP has been a well suited method to model the ocean current spatiotemporal dynamics [26]. The GP models a noisy process as the water speed v^i by $v^i = f(x^i) + \epsilon$, where x^i is the spatial z^i and temporal t^i values, and ϵ denotes the Gaussian noise. To enable the GP model, a set of the recorded data in the form of $D = [(x^i, v^i) | x \in X, v \in V]$ is given, where x^i denotes a vector of depth z^i and time t^i values for a data point i . To define the GP model, we should find the mean function $m(x)$ and covariance function $k(x, x')$, as follows [27]:

$$f(x) \sim \mathcal{N}(m(x), k(x, x')) \quad (1)$$

$$m(x) = \mathbb{E}[f(x)] \quad (2)$$

$$k(x, x') = \mathbb{E}[(f(x) - m(x))(f(x') - m(x')))] \quad (3)$$

To predict the water speed, the GP predictive function $f(x^*)$ is stated at the future input x^* based on the previous evaluations. Given a set of observed ocean current speed V , the future mean function $m(x^*)$ and covariance function $k(x^*, x')$ are defined in (5) and (6) as follows:

$$f(x^*) \sim \mathcal{N}(m(x^*), k(x^*, x')) \quad (4)$$

$$m(x^*) = k(x^*)^T (K + \sigma^2 I_N)^{-1} V \quad (5)$$

$$k(x^*, x') = k(x^*, x^*) - k(x^*)^T (K + \sigma^2 I_N)^{-1} k_* \quad (6)$$

where σ^2 denotes the Gaussian observation noise hyperparameter, and:

$$k(x^*) = [k(x^*, x^1), k(x^*, x^2), \dots, k(x^*, x^n)] \quad (7)$$

$$K = [k(x^i, x^j)]_{n \times n} \quad (8)$$

To enable the online GP model, a rolling window is applied to predict the future ocean current by (5) and (6). Therefore, a fixed-length rolling window is rolled over the data to avoid the unbounded growth of memory requirements due to increased data, where the old measured velocities are removed with the arrival of a new set of data.

It should be noted that a large number of physical processes occur in the spatiotemporal ocean environment, resulting in large uncertainties in the ocean current flow estimation. Hence, modeling the spatiotemporal uncertainties of the ocean current is intricate, resulting in errors in current prediction and obtained estimation of future ocean current. There exist a multitude of resources for these inaccuracies and errors in the current prediction [28], [29], including sensor error, simplifications of the spatial and temporal model and scale, an imperfection in the ocean current prediction models, etc. The ocean current sensor errors may occur due to data loss and Doppler noise, so the recorded ocean currents are inaccurate. The spatial and temporal model of the ocean current is simplified to reduce the complexity of the calculations, which are another source of error. Furthermore, errors also arise from the inaccuracies and simplifications in the ocean current prediction models.

In this study, we use the GP model to estimate the future ocean current speed according to the ocean currents recorded by a 75 kHz ADCP. Although the measured ocean currents are filtered to remove bad data, Doppler noise, spatial averaging, and other errors still distort the measurements used as the inputs to our prediction model. As explained, the sensor errors may mislead the prediction model, where the observed ocean currents differ from the real ocean current. To simplify our model, we only consider the depth coordinate at a specific latitude and longitude to interpret the spatial model of the ocean environment. Finally, to avoid the curse of dimensionality due to the substantial increase in the recorded ocean current, we apply a fixed-size rolling window. In summary, the error can occur due to the above-mentioned factors, resulting in the deviation of the current estimations from the true values. To account for the error in ocean current prediction, we introduce an error value e to the predicted ocean current value by (4) and rewrite the predicted ocean current V^* as the estimated ocean current $\hat{V} = V^* + e$.

III. PROPOSED OPTIMIZATION METHODOLOGY

A. Optimization Problem Formulation

This paper targets solving real-time spatiotemporal optimization to maximize the harnessed power from the lifting surface controlled OCT. More specifically, the OCT can be treated as an “autonomous underwater vehicle” but with the primary role of energy generation. Given this primary role, it is critical that the OCT maintains an accurate spatiotemporal estimate of the Gulf Stream current profile and navigate itself at or near the depth with the most intensive ocean flow.

The lifting surface controlled OCT is based on an 8 degree of freedom OCT design presented in [30], which includes twin counter rotating 20 m diameter variable pitch rotor blades. A schematic diagram of the lifting surface controlled

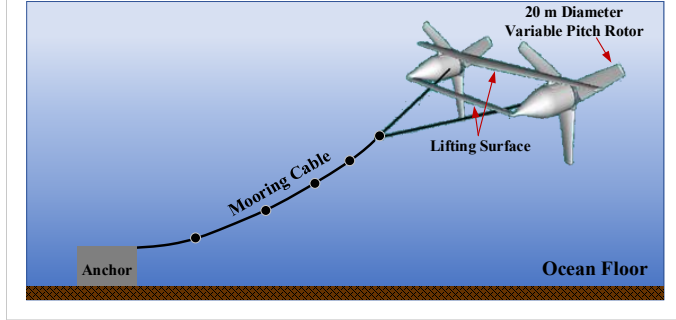


Fig. 2. Schematic diagram of the lifting surface controlled OCT.

OCT is shown in Fig. 2. The generated power of the twin rotor OCT P is formulated as follows:

$$P_t = 2 \times \frac{1}{2} \rho A C_p v_t^3(z_t) - P_{\Delta z} \quad (9)$$

$$= 1024 \times 100\pi \times 0.415 \times v_t^3(z_t) - P_{\Delta z}$$

where ρ is the ocean water density, A is the swept area of the OCT rotor, C_p denotes the average power coefficient, and $P_{\Delta z}$ is the power consumed to change the operating depth, which is negligible compared to the first term due to the lifting surface controlled design of the OCT. In the following, we design two different approaches based on MPC and DRL to find the optimal sequence of depths. Note that these optimal depths are calculated to maximize the harnessed power from the OCT in a given time horizon, where the effect of error in ocean current prediction is also evaluated.

B. Model Predictive Control

The MPC-based optimization algorithm uses the model of the system to determine the optimal control solutions. The optimization problem is formulated at each time step according to the OCT model to minimize a defined cost function in a prediction horizon N due to the problem constraints. It should be noted that the real-time spatiotemporal optimization is solved in a prediction horizon, where only the first value of the obtained control sequence is defined as the optimal depth in our problem. The MPC-based algorithm is employed with the outlook of the prediction horizon (i.e., length of the sliding window) of N , where the sliding window will be rolled in the next time step, as shown in Fig. 3. Hence, an optimal depth sequence of N (Eq. (12)) is calculated through solving an objective function, where its elements are determined based on the first element of control sequence found at each time step (Eq. (11)).

Fig. 4 shows a schematic of the proposed MPC-based optimization algorithm for our problem. The objective function is defined as Eq. (10) subject to Eq. (11)- Eq. (13):

$$J^* = \min_Z \sum_{i=k}^{k+N-1} [\omega_1 l_1(z_{i|k}, v_{i|k}) + \omega_2 l_2(z_{i|k}, v_{i|k})] \quad (10)$$

s.t.

$$z(i+1|k) = z_{0|k}^* \quad (11)$$

$$Z = [z_{0|k}, \dots, z_{N|k}] \quad (12)$$

$$\hat{V} = V^* + e \quad (13)$$

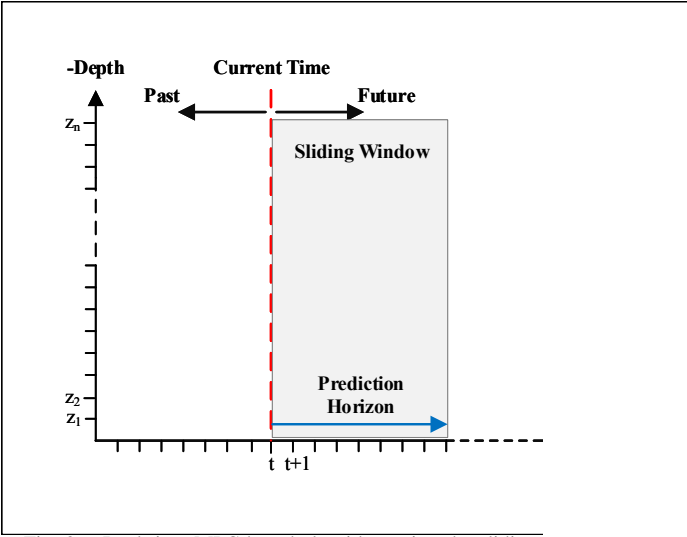


Fig. 3. Real-time MPC-based algorithm using the sliding the optimal solution of the spatiotemporal optimization.

where J^* is the MPC cost function, Z is the cont defined as the depth sequence in our problem and $l_2(z_{i|k}, v_{i|k})$ represent the cost function terms denote the weights of cost function terms. $z_{0|k}^*$ element of the optimal control sequence Z^* , as

$$Z^* = [z_{0|k}^*, z_{1|k}^*, \dots, z_{N|k}^*]$$

The first objective function term $l_1(z(i|k), v(i|k))$ is defined to maximize the harnessed power from the OCT through minimizing the expected power of the $E(P(z(i|p)))$ due to the power relation defined in Eq. (9). To mitigate the uncertainties in the current flow, we define the second term as suggested in [31]. The first term and the second term of the cost function are defined as follows:

$$l_1(z_{i|k}, v_{i|k}) = -E(P(z_{i|k}, v_{i|k})) \quad (15)$$

$$l_2(z_{i|k}, v_{i|k}) = \sum_{z'=z_l}^{z'=z_u} \sigma^c(v(t(i), z')|V, \hat{V}) \quad (16)$$

where $E(P(z(i|k), v(i|k)))$ shows the expected power, z_l and z_u are the lower bound and upper bound of the ocean depths. $\sigma^c(v(t(i), z')|V, \hat{V})$ denotes the conditional standard deviation of the future estimated ocean current from the measured ocean current.

C. Deep Reinforcement Learning

Reinforcement learning is a data-driven, learning-based technique to find the optimal control solutions (i.e., optimal depth in our problem). It should be noted that RL is formulated as a Markov decision process, where the OCT as an agent observes the current state $s_t \in S$, takes action $a_t \in A$ according to the learned policy π , resulting in a reward $r_t \in R$ and state change. The RL-based optimization learns to map each state to the optimal action to maximize the cumulative discounted reward $R_t = \sum_{i=0}^{\tau} \gamma^i r_{t+i}$. The proposed design for the real-time spatiotemporal optimization based on DRL is shown in Fig. 5.

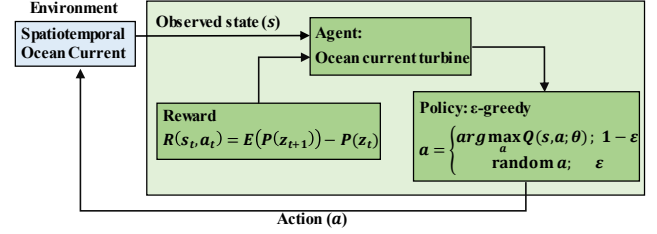


Fig. 5. Proposed architecture for the DRL-based real-time spatiotemporal optimization.

The state, action, reward, and policy notations should be described according to the problem, which are defined in our real-time spatiotemporal optimization. The state is defined as the position of the OCT and the current flow speeds at a different depth, formulated as follows:

$$S = \{z_t, v_t^1, v_t^2, \dots, v_t^n\} \quad (17)$$

The action space is determined as the potential optimal depth due to the current state (i.e., the position of the OCT and the current flow speeds), as follows:

$$A = \{z^1, z^2, \dots, z^n\} \quad (18)$$

In our proposed spatiotemporal optimization algorithm, we define the reward $R(s_t, a_t)$ of an OCT taking action a_t at state s_t . The reward is defined as the power change that occurred due to the depth change, as follows:

$$R(s_t, a_t) = \begin{cases} +1, & E(P(z_{t+1})) - P(z_t) > \delta \\ 0, & \text{else} \end{cases} \quad (19)$$

The policy $\pi : S \rightarrow R$ guides the agent (i.e., OCT in our problem) to choose the optimal action a_t according to the current state s_t . The ϵ -greedy policy can be applied to balance the exploitation along with the exploitation in the process of choosing optimal actions. However, a simple ϵ -greedy algorithm with a constant ϵ (i.e., a constant rate of exploration vs. exploitation during the training process) does not justify an ensured optimal policy. Therefore, we use an adaptive ϵ -greedy policy [32], where the value of ϵ decreases over the training procedure. Eventually, the policy $\pi(s_t, a_t)$ determines the probability of choosing a set of feasible actions, where the adaptive ϵ -greedy policy is

applied to calculate the probability of each action, as follows:

$$a_t = \pi(s_t, a_t) = \begin{cases} a_t \leftarrow \operatorname{argmax}_{a_t} Q(s_t, a_t), & 1 - \varepsilon \\ a_t \leftarrow \text{random } a \in A, & \varepsilon \end{cases} \quad (20)$$

$$\varepsilon = \varepsilon_{\min} + (\varepsilon_{\max} - \varepsilon_{\min})e^{-d \times n_e} \quad (21)$$

where d and n_e are decay factor and episode number, respectively. $Q(s_t, a_t)$ is the Q-value determined by Bellman equation, as follows:

$$Q(s_t, a_t) = r_t + \gamma \max_{a_t} Q(s_{t+1}, a_t) \quad (22)$$

The future state s_{t+1} is determined by the current state s_t and action policy π , and our method aims to find the optimal policy π^* to maximize the cumulative reward. We use the DRL approach to find the optimal policy. Among different DRL algorithms, including deep Q-networks (DQN), deep deterministic policy gradient methods, trust region policy optimization, etc. [33], we choose the DQN, described in the following section.

Deep Q-networks: A DQN is a neural network-based RL, which maps the relation between states and actions, where the DQN as the RL-based algorithm should interact with an environment by observing a sequence of the current state, action, reward, and future state (s_t, a_t, r_t, s_{t+1}) . The DQN uses the neural network as a nonlinear estimation to find the optimal action estimation Q^* , where all weights in the neural network are denoted by θ , which are calculated through minimizing the Mean Square Error (MSE), as follows:

$$L(\theta) = [(r_t + \gamma \max_{a_t} Q(s_{t+1}, a_t)) - Q(s_t, a_t; \theta)]^2 \quad (23)$$

where $r_t + \gamma \max_{a_t} Q(s_{t+1}, a_t)$ is denoted as target Q-value found through the target network, which is utilized to provide a stable update on the neural network and defined by similar architecture with the main network but a different set of weight parameters θ^- .

After the training procedure of the DQN is completed according to Eq. (23), the optimal weights of the neural network θ^* are obtained. Hence, the optimal Q-value is determined by $Q(s_t, a_t; \theta^*)$. In our problem, the DQN is trained offline using the previously measured ocean current V , where the constructed DQN is then used to find the optimal depth in a real-time application, feeding with the estimated ocean currents $\hat{V}$.

IV. RESULTS AND DISCUSSIONS

To quantify the performance of the MPC-based and DRL-based spatiotemporal optimization, the testing results of the proposed methodology on a sample 1.4 MW lifting surface controlled OCT over one week (from September 18, 2015 to September 26, 2015) are presented in this section. To enable the MPC-based algorithm, the utilized time step is 1 hour, prediction horizon is $N = 10$, assuming $z_l = 40$ m and $z_u = 150$ m since the maximum ocean currents occur in the top 100 m depths. For the DQN-based algorithm, we select a network with two hidden layers and use a buffer size of $5e5$ (including 94 MB data), a batch size of 64,

$\gamma = 0.5$, $\varepsilon_{\min} = 0.01$, $\varepsilon_{\max} = 1$, $d = 0.01$, and $n_e = 3000$, where the target network is updated every 4 steps. To train the DQN network, we use a four-week measured ocean current data set (from August 20 to September 17, 2015), where the constructed network is then tested on a one-week period that indicated above. It should be noted that the DQN algorithm is sensitive to the network parameters, and fine-tuning through trial and error are used to determine the parameters. The results are shown in two cases of with and without ocean current prediction error.

A. Results

To highlight the differences between the proposed MPC- and DRL-based algorithms, two cases of testing “no prediction error” and “current speed prediction error” are considered. In the first case, the current prediction error e is assumed zero. Hence, the estimated ocean current speed $\hat{V}$ is equal to the predicted ocean current V^* . To include the current prediction error, we assume that $\hat{V}$ is deviated from the predicted ocean current V^* by e , as dictated by the high spatiotemporal uncertainties in the ocean current.

Fig. 6 shows the simulation results over one week, where the initial depth of OCT is set to 50 m. The optimal trajectories determined by the MPC-based algorithm and the DRL-based algorithm are demonstrated in Fig. 6 (a), where the obtained optimal depths look similar. However, two differences in choosing the optimal depth (i.e., September 19 and September 23), indicating the difference between the selection strategy of the MPC and DRL algorithms. As shown in this figure, both algorithms tend to navigate the OCT through the higher ocean current speeds, verifying the performance of the proposed spatiotemporal optimization methodology. The harnessed power from the lifting surface controlled OCT using two spatiotemporal algorithms is shown in Fig. 6 (b).

To understand the effect of “current speed prediction error”, the simulation results are obtained considering the difference between estimated ocean current and predicted ocean current. To evaluate the results with the first case, a similar initial depth of 50 m and the same simulation period are set. Since the prediction error has a noisy nature, the optimal depth sequence is different each time. Hence, the tests are performed multiple times, and the obtained results are shown in Table I. We performed 4 different tests, and the results are shown in terms of cumulative energy and average power for MPC and DRL algorithms. As shown in this table, a significant decrease is observed compared to the baseline (i.e., the case without prediction error), which is more observable in the MPC algorithm.

We show the optimal depth and the harnessed power of Test 1 in Fig. 7, where the estimated ocean current is significantly different from the true ocean current. Note that the prediction error is defined as 20% noise disturbances [34]. As shown in Fig. 7 (a), the obtained trajectory by the DRL follows the higher current flow, resulting in a larger harvested power despite the more depth fluctuations at several time steps; still, the MPC algorithm is misled and

TABLE I

COMPARISON OF DRL AND MPC WITH OCEAN CURRENT PREDICTION ERROR IN FOUR SAMPLE TESTS. THE OBTAINED SOLUTIONS ARE COMPARED WITH THE BASELINE (I.E., THE CASE WITHOUT PREDICTION ERROR).

Test No.	MPC			DRL		
	Cumulative Energy [MWh]	Average Power [kW]	Decrease to Baseline [%]	Cumulative Energy [MWh]	Average Power [kW]	Decrease to Baseline [%]
Test 1	105.5	628.2	9.2	108.5	645.7	6.2
Test 2	104.4	621.1	10.2	108.1	643.4	6.5
Test 3	104.5	622.3	10.1	108.0	642.8	6.6
Test 4	105.7	629.2	9.1	108.2	644.1	6.4

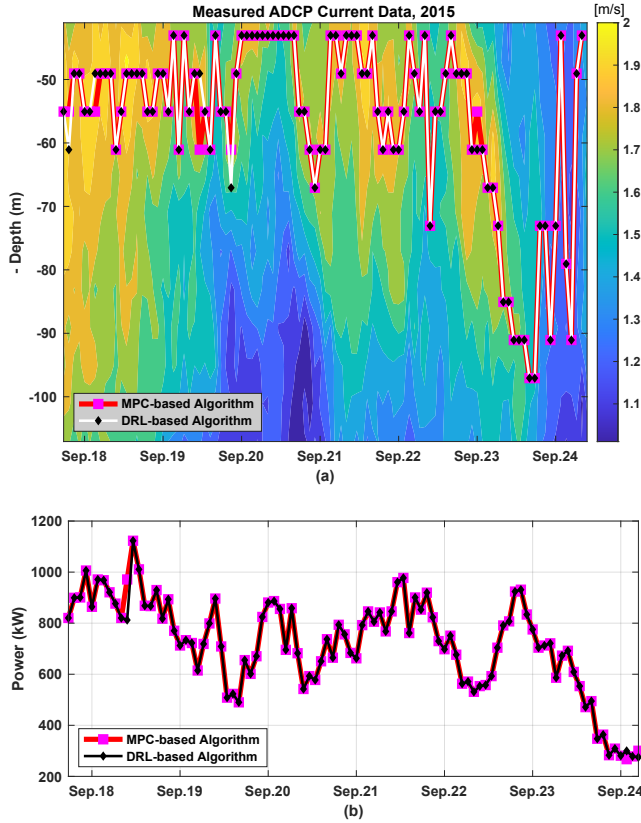


Fig. 6. Comparison of an optimal sequence determined by (i) MPC-based algorithm and (ii) DRL-based algorithm in case of “no prediction error”. (a) Optimal position; and (b) Optimal power.

deviated significantly from the optimal depths. Therefore, a multitude of the determined depths by the MPC-based algorithm, especially on September 22 and September 23, are deviated from the optimal depth and current, as shown in Fig. 6 (a). The DRL algorithm outperforms the MPC algorithm since it learns the optimal policy through offline training of the DQN network with the previously measured ocean current; hence, the DRL is less affected by the ocean current prediction error.

B. Discussions

Through a careful evaluation of the results under the “no error” case, it is verified that the model-based algorithm (i.e., MPC) and the learning-based algorithm (i.e., DRL)

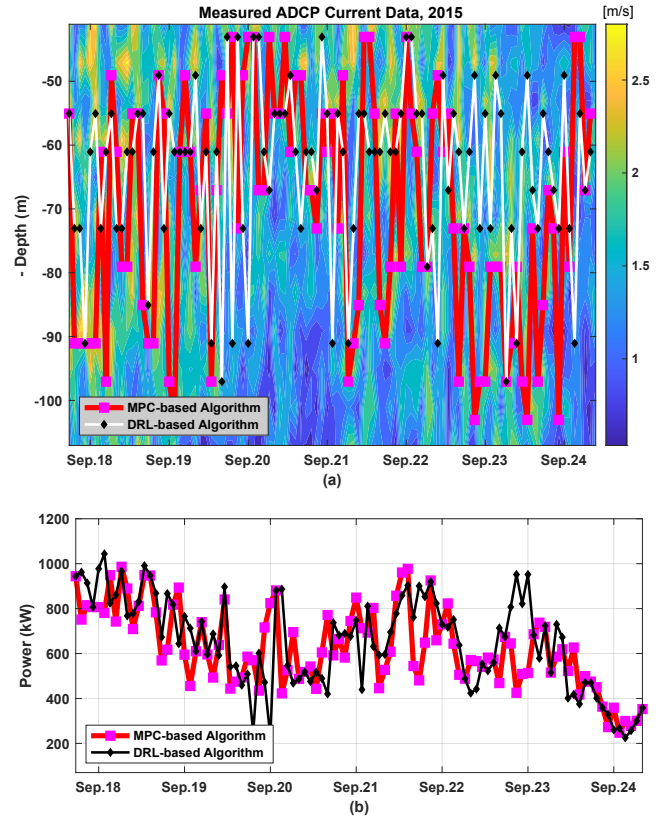


Fig. 7. Comparison of an optimal sequence determined by (i) MPC-based algorithm and (ii) DRL-based algorithm in case of “ocean current prediction error”. (a) Optimal position; and (b) Optimal power.

can optimally find the trajectory as the control sequence in the real-time spatiotemporal optimization problem. The final result of MPC and DRL algorithms in terms of cumulative energy and average power under the “no error” case and “current speed prediction error” are presented in Table II. It should be noted that the results of the “current speed prediction error” case are reported as an average of the results obtained by 100 tests. To understand the effect of the current prediction error, the decrease compared to the “no error” case is also included in the table. The obtained cumulative energy over one week is 116.3 MWh and 115.6 MWh for MPC and DRL, respectively. Further, the average harnessed power over a period of one week is 692.1 kW for MPC and 688.1 kW for DRL.

TABLE II

COMPARISON OF CUMULATIVE ENERGY AND AVERAGE POWER WITH AND WITHOUT OCEAN CURRENT PREDICTION ERROR USING MPC AND DRL. THE CASE WITHOUT PREDICTION ERROR IS A BASELINE.

Parameter	[Unit]	MPC	DRL
Without Prediction Error			
Cumulative Energy	[MWh]	116.3	115.6
Average Power	[kW]	692.1	688.1
With Prediction Error			
Cumulative Energy	[MWh]	105.2	108.3
Average Power	[kW]	626.2	644.6
Decrease to Baseline	[%]	9.5	6.3

It is worthwhile to mention that the comparison study is not trivial. It is tough to set an even ground and evaluate the two optimization algorithms (MPC along with DRL) adequately. To realize a fair comparison, we make an effort to define a similar condition for both approaches. Due to the high error in the estimation of ocean current, an input disturbance rejection can be applied for improving DRL algorithm [35] and MPC algorithm [36], which is beyond the scope of our study.

The ocean environment is noisy, highly affected by the ocean current's spatiotemporal uncertainties, and it is highly probable that we fail to precisely model and predict the future ocean current. Hence, the prediction error should be added to the ocean current prediction model, resulting in the difference between the estimated and predicted ocean currents. After adding the "current prediction error" to our GP model, we investigate the high deviation from the optimal trajectory found in the "no error" case, highlighting the importance of the error modeling for practical applications. Both algorithms are deviated from the optimal trajectory, while the learning-based algorithm is significantly better than the model-based algorithm to find the optimal solutions in a noisy environment. The average harnessed power from the lifting surface controlled OCT is reduced by 9.5% to 626.2 kW for MPC, while the power decreasing is 6.3% for DRL, verifying the error-tolerance of the learning-based algorithm compared to the model-based algorithm. The obtained cumulative energy over one week is significantly decreased compared to the "no error" case, that is 105.2 MWh and 108.3 MWh for the MPC algorithm and DRL algorithm, respectively.

Different policies provoke a difference between the net energy in choosing the optimal depth, where the MPC algorithm follows the system and environment model; still, the DRL algorithm learns the policy from its interactions with the environment. Hence, the MPC algorithm can guarantee the ensured performance in the non-noisy environment and outperform the DRL, while the DRL can find semi-optimal solutions after fine-tuning its parameters through offline training of the network. Therefore, improving the prediction of ocean current boosts the results obtained from the MPC algorithm. Meanwhile, the DRL extracts the robust and efficient high-level features from the noisy and uncertain ocean current historical data to beat the MPC algorithm.

V. CONCLUSIONS

Real-time spatiotemporal optimization for a case study of a lifting surface controlled OCT was presented. The optimization target was to find a sequence of optimal turbine operation depths to maximize the harnessed power. DRL and MPC strategies were applied to solve the optimization problem and find the optimal control sequence. Simulations were run for the two cases of "no prediction error" and "ocean current prediction error" to verify the effectiveness. We showed that the DRL and MPC algorithms operated similarly when the estimated ocean current was assumed similar to the predicted current. However, the DRL outperformed the MPC in case of high prediction error due to the spatiotemporal uncertainties in the ocean current.

Future work will focus on testing the performance and robustness of the proposed methodology over various modeling errors. Also, one of the major drawbacks of the DRL and neural network-based algorithms is the generalization issue, which will be carefully tested in future works and will be justified with a stochastic MPC [37] and a robust MPC [38]. The robust MPC will be explored for rejection of disturbance and uncertainties of the model, in addition to improving the prediction of the ocean current. Meanwhile, the robust DRL [39] will also need to be tested to fairly compare the robust model-based and learning-based approaches. It is of interest to extend the proposed methodology to optimize an OCT array and design a real-time spatiotemporal optimization framework for the OCT array, assuming wake effects between multiple OCTs, optimal navigation, and collision avoidance.

REFERENCES

- [1] X. Yang, K. Haas, and H. Fritz, "Evaluating the potential for energy extraction from turbines in the gulf stream system," *Renewable energy*, vol. 72, pp. 12–21, 2014.
- [2] M. C. P. M. Machado, J. H. VanZwieten, and I. Pinos., "A measurement based analyses of the hydrokinetic energy in the gulf stream," *Journal of Ocean and Wind Energy*, vol. 3, no. 1, pp. 25–30, 2016.
- [3] T. Ueno, S. Nagaya, M. Shimizu, H. Saito, and N. Handa, "Development and demonstration test for floating type ocean current turbine system conducted in kuroshio current," in *2018 OCEANS-MTS/IEEE Kobe Techno-Oceans (OTO)*. IEEE, 2018, pp. 1–6.
- [4] A. Hasankhani, J. VanZwieten, Y. Tang, B. Dunalp, A. D. Luera, C. Sultan, and N. Xiros, "Modeling and numerical simulation of a buoyancy controlled ocean current turbine," *International Marine Energy Journal*, Submitted.
- [5] Y. Tang, Y. Zhang, A. Hasankhani, and J. VanZwieten, "Adaptive super-twisting sliding mode control for ocean current turbine-driven permanent magnet synchronous generator," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 211–217.
- [6] Minesto – power from tidal and ocean currents. [Online]. Available: <http://www.minesto.com>
- [7] A. Siddiqui, K. Naik, M. Cobb, K. Granlund, and C. Vermillion, "Lab-scale, closed-loop experimental characterization, model refinement, and validation of a hydrokinetic energy-harvesting ocean kite," *Journal of Dynamic Systems, Measurement, and Control*, vol. 142, no. 11, 2020.
- [8] A. Baheri, P. Ramaprabhu, and C. Vermillion, "Iterative 3d layout optimization and parametric trade study for a reconfigurable ocean current turbine array using bayesian optimization," *Renewable energy*, vol. 127, pp. 1052–1063, 2018.
- [9] K. Shirasawa, K. Tokunaga, H. Iwashita, and T. Shintake, "Experimental verification of a floating ocean-current turbine with a single rotor for use in kuroshio currents," *Renewable Energy*, vol. 91, pp. 189–195, 2016.

- [10] J. Deese and C. Vermillion, "Real-time experimental optimization of closed-loop crosswind flight of airborne wind energy systems via recursive gaussian process-based adaptive control," in *2020 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2020, pp. 516–521.
- [11] S. Bin-Karim, M. Muglia, and C. Vermillion, "Centralized position optimization of multiple agents in spatiotemporally-varying environment: a case study with relocatable energy-harvesting autonomous underwater vehicles in the gulf stream," in *2019 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2019, pp. 264–269.
- [12] E. F. Camacho and C. B. Alba, *Model predictive control*. Springer Science & Business Media, 2013.
- [13] Z. Wang, G. Li, H. Jiang, Q. Chen, and H. Zhang, "Collision-free navigation of autonomous vehicles using convex quadratic programming-based model predictive control," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 3, pp. 1103–1113, 2018.
- [14] E. Small, P. Sopasakis, E. Fresk, P. Patrinos, and G. Nikolakopoulos, "Aerial navigation in obstructed environments with embedded nonlinear model predictive control," in *2019 18th European Control Conference (ECC)*. IEEE, 2019, pp. 3556–3563.
- [15] W. B. Greer and C. Sultan, "Shrinking horizon model predictive control method for helicopter–ship touchdown," *Journal of Guidance, Control, and Dynamics*, vol. 43, no. 5, pp. 884–900, 2020.
- [16] W. B. Greer and C. Sultan, "Infinite horizon model predictive control tracking application to helicopters," *Aerospace Science and Technology*, vol. 98, p. 105675, 2020.
- [17] T. D. Ngo, C. Sultan, J. H. VanZwieten, and N. I. Xiros, "Model predictive control for moored ocean current turbines," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 875–880.
- [18] D. Son and R. W. Yeung, "Optimizing ocean-wave energy extraction of a dual coaxial-cylinder wec using nonlinear model predictive control," *Applied energy*, vol. 187, pp. 746–757, 2017.
- [19] X. Han, H. Liu, F. Sun, and X. Zhang, "Active object detection with multistep action prediction using deep q-network," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3723–3731, 2019.
- [20] T. Tan, T. Chu, and J. Wang, "Multi-agent bootstrapped deep q-network for large-scale traffic signal control," in *2020 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2020, pp. 358–365.
- [21] E. Marchesini and A. Farinelli, "Discrete deep reinforcement learning for mapless navigation," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 10 688–10 694.
- [22] S. Guo, X. Zhang, Y. Zheng, and Y. Du, "An autonomous path planning model for unmanned ships based on deep reinforcement learning," *Sensors*, vol. 20, no. 2, p. 426, 2020.
- [23] M. Abdelaal, M. Fränzle, and A. Hahn, "Nonlinear model predictive control for trajectory tracking and collision avoidance of underactuated vessels with disturbances," *Ocean Engineering*, vol. 160, pp. 168–180, 2018.
- [24] X. Wu, H. Chen, C. Chen, M. Zhong, S. Xie, Y. Guo, and H. Fujita, "The autonomous navigation and obstacle avoidance for usvs with anoa deep reinforcement learning method," *Knowledge-Based Systems*, p. 105201, 2020.
- [25] X. Xu, Y. Lu, X. Liu, and W. Zhang, "Intelligent collision avoidance algorithms for usvs via deep reinforcement learning under colregs," *Ocean Engineering*, vol. 217, p. 107704, 2020.
- [26] 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," *Journal of Field Robotics*, vol. 33, no. 1, pp. 47–66, 2016.
- [27] K.-C. Ma, L. Liu, and G. S. Sukhatme, "Informative planning and online learning with sparse gaussian processes," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2017, pp. 4292–4298.
- [28] P. F. Lermusiaux, C.-S. Chiu, G. G. Gawarkiewicz, P. Abbot, A. R. Robinson, R. N. Miller, P. J. Haley, W. G. Leslie, S. J. Majumdar, A. Pang *et al.*, "Quantifying uncertainties in ocean predictions," HARVARD UNIV CAMBRIDGE MA, Tech. Rep., 2006.
- [29] Y. Tang, J. VanZwieten, B. Dunlap, D. Wilson, C. Sultan, and N. Xiros, "In-stream hydrokinetic turbine fault detection and fault tolerant control-a benchmark model," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 4442–4447.
- [30] J. VanZwieten, F. Driscoll, A. Leonessa, and G. Deane, "Design of a prototype ocean current turbine—part ii: flight control system," *Ocean engineering*, vol. 33, no. 11-12, pp. 1522–1551, 2006.
- [31] S. Bin-Karim, A. Bafandeh, A. Baheri, and C. Vermillion, "Spatiotemporal optimization through gaussian process-based model predictive control: A case study in airborne wind energy," *IEEE Transactions on Control Systems Technology*, vol. 27, no. 2, pp. 798–805, 2017.
- [32] X. Liang, X. Du, G. Wang, and Z. Han, "A deep reinforcement learning network for traffic light cycle control," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 2, pp. 1243–1253, 2019.
- [33] Y. Lin, J. McPhee, and N. L. Azad, "Comparison of deep reinforcement learning and model predictive control for adaptive cruise control," *IEEE Transactions on Intelligent Vehicles*, 2020.
- [34] T. D. Ngo, C. Sultan, J. H. VanZwieten, and N. I. Xiros, "Constrained control of moored ocean current turbines with cyclic blade pitch variations," *IEEE Journal of Oceanic Engineering*, 2020.
- [35] Z.-q. Su, M. Zhou, F.-f. Han, Y.-w. Zhu, D.-l. Song, and T.-t. Guo, "Attitude control of underwater glider combined reinforcement learning with active disturbance rejection control," *Journal of Marine Science and Technology*, vol. 24, no. 3, pp. 686–704, 2019.
- [36] Z. Sun, Y. Xia, L. Dai, K. Liu, and D. Ma, "Disturbance rejection mpc for tracking of wheeled mobile robot," *IEEE/ASME Transactions On Mechatronics*, vol. 22, no. 6, pp. 2576–2587, 2017.
- [37] S. Di Cairano, D. Bernardini, A. Bemporad, and I. V. Kolmanovsky, "Stochastic mpc with learning for driver-predictive vehicle control and its application to hev energy management," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 3, pp. 1018–1031, 2013.
- [38] D. Q. Mayne, M. M. Seron, and S. Raković, "Robust model predictive control of constrained linear systems with bounded disturbances," *Automatica*, vol. 41, no. 2, pp. 219–224, 2005.
- [39] H. Zhang, H. Chen, C. Xiao, B. Li, D. Boning, and C.-J. Hsieh, "Robust deep reinforcement learning against adversarial perturbations on observations," *arXiv preprint arXiv:2003.08938*, 2020.