Dynamic Coverage Meets Regret: Unifying Two Control Performance Measures for Mobile Agents in Spatiotemporally Varying Environments

Ben Haydon[1], Kirti D. Mishra[1], Patrick Keyantuo[2], Dimitra Panagou[3], Fotini Chow[2], Scott Moura[2], and Chris Vermillion[1]

***Abstract*— Numerous mobile robotic applications require agents to *persistently* explore and exploit spatiotemporally varying, partially observable environments. Ultimately, the mathematical notion of *regret*, which quite simply represents the instantaneous or time-averaged difference between the optimal reward and realized reward, serves as a meaningful measure of how well the agents have exploited the environment. However, while numerous theoretical regret bounds have been derived within the machine learning community, restrictions on the manner in which the environment evolves preclude their application to persistent missions. On the other hand, meaningful theoretical properties can be derived for the related concept of *dynamic coverage*, which serves as an exploration measurement but does not have an immediately intuitive connection with regret. In this paper, we demonstrate a clear correlation between an appropriately defined measure of dynamic coverage and regret, then go on to derive performance bounds on dynamic coverage as a function of the environmental parameters. We evaluate the correlation for several variants of an airborne wind energy system, for which the objective is to adjust the operating altitude in order to maximize power output in a spatiotemporally evolving wind field.**

I. Introduction

Numerous mobile robotic systems are tasked with carrying out missions in or harvesting resources from spatiotemporally varying environments. Examples include wheeled mobile robotic systems [1], [2], unmanned aerial vehicles [3], autonomous sailing drones [4], and tumbleweed rovers [5]. In each of the aforementioned examples, the spatiotemporally varying resource is only measured at a limited set of locations, thereby rendering the environment partially observable and necessitating a balance between exploration (maintaining an accurate map of the environment) and exploitation (carrying out the mission). A final example, which serves as a case study in the present paper, is that of an airborne wind energy (AWE) system. In an AWE system, the conventional tower is replaced with tethers and a lifting body (typically a kite, wing, or aerostat). This allows the system to adjust its altitude in order to seek the highest wind speeds in order to maximize power generation. Since it is impractical to measure the wind at locations far from the AWE system, a balance of exploration and exploitation is once again required (see [6]–[10]). For the present work, we focus our attention on a farm of Buoyant Airborne Turbines (BATs), as depicted in Fig. 1a. Each BAT uses a lighter-than-air shell to lift a horizontal-axis turbine to high altitudes. In our simulation results, we consider topologies with and without the inclusion of supplemental sensing drones (quad rotors), as shown in Fig. 1b, with a sample spatiotemporally varying wind environment (i.e., a set of wind shear profiles) depicted in Fig. 1c. We use this system as a running thread to illustrate key concepts and results throughout the paper.

In all of the aforementioned works, control strategies have been devised that effectively balance exploration and exploitation, as validated through simulation and experimental studies. However, it is desirable to augment these empirical studies with analytic performance guarantees for the purpose of comparing them against each other for a variety of environmental conditions or with extension to other applications. Indeed, numerous publications in the machine learning literature produce analytic bounds on exploration/exploitation performance in terms of a quantity known as *regret*, which represents the instantaneous, time-averaged, or cumulative difference between achieved performance and the performance achieved and optimal performance. While regret serves as an ideal performance measure, each of the aforementioned regret bounds derived in the literature involves some significant restriction on the manner by which the environment evolves, which precludes their extension to the class of systems studied in the present work. In particular, [13], [14] and many others others assume a stationary environment. Among the papers that consider non-stationary environments, which include work on dynamic regret and adaptive regret, the imposition of restrictions on the manner in which the environment is able to evolve temporally are key to deriving regret bounds. These restrictions include abruptly changing environments [15], [16], variation budgets [16], [17], adversarial environments [18], and independence assumptions [19]. In other cases, restrictions such as path length restrictions [20]–[22] and comparator regularity [20] affect the extent to which the regret reflects the true performance of a system, indirectly affecting their application to spatiotemporally varying environments.

Due to the limitations associated with the direct derivation of analytical regret bounds for our class of systems, we turn to an alternative measure known as *dynamic coverage*.

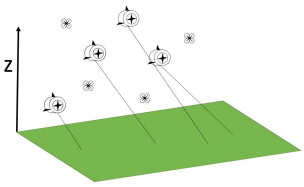
*This work was supported by NSF award numbers 1711579 (Collaborative Research: Multi-Scale, Multi-Rate Spatiotemporal Optimal Control with Application to Airborne Wind Energy Systems) and 2012103 (Persistent Mission Planning and Control for Renewably Powered Robotic Systems).

[1]Ben Haydon, Kirti D. Mishra, and Chris Vermillion are with the Department of Mechanical & Aerospace Engineering, North Carolina State University, Raleigh, NC, USA - 27695 brhaydon@ncsu.edu

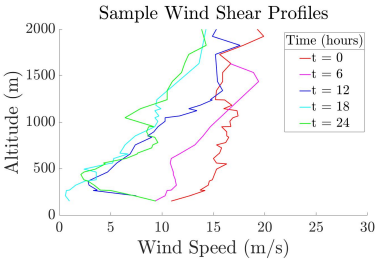
[2] Patrick Keyantuo, Fotini Chow, and Scott Moura are with the Department of Civil & Environmental Engineering, University of California, Berkeley, Berkeley, California, USA - 94720

[3] Dimitra Panagou is with the Department of Aerospace Engineering, University of Michigan, MI, USA - 48109

(a) Altaeros Buoyant Airborne Turbine (BAT) [11]



(b) Four AWE systems as exploitation agents with four quadcopters as exploration agents



(c) Sample wind shear profiles from Cape Henlopen wind profiler [12]

Fig. 1: AWE system considered in this work (left), overall topology for simulation case studies in the paper (middle), and sample wind profile evolution (right).

As described in [3], [23], [24], and [25], dynamic coverage characterizes how well each point within a prescribed domain has been measured (i.e., "covered") at each point in time. The underlying assumption is that the sensors are mobile and, in general, the environment can be time-varying. Dynamic coverage-based control has been applied successfully to intruder detection [23], robustness to sensor death [24], and full-domain exploration using directional sensors [3], [25]. From the standpoint of deriving analytical bounds, coverage carries a distinct advantage over regret in the sense that it is chiefly an exploration-based metric, rendering it largely decoupled from the stochasticity of the environment. However, because it is largely an exploration-based metric, it is not as obviously tied to performance as regret is.

In summary, we have examined two metrics – regret and dynamic coverage. One of these (regret) is precisely the performance metric that we would like to derive bounds for but is fraught with mathematical complications. The other (dynamic coverage) provides a more straightforward mechanism for deriving bounds but is not obviously correlated with true performance. In this paper, we bridge the gap between dynamic coverage and regret by re-tailoring the standard dynamic equations that describe coverage to capture the essential dynamic behavior of regret. This allows us to derive meaningful bounds on an appropriately defined dynamic coverage metric, then demonstrate a clear correlation between that metric and regret. Specifically, we derive hard bounds on average coverage over the domain, then demonstrate that the coverage model does in fact closely correlate to regret, demonstrating the utility of the derived bounds. We illustrate this specifically for the AWE system case study.

II. Problem Topology and Definitions

Each mobile network considered in this work consists of *exploitation agents* and, optionally, *exploration agents*:

- *Exploitation agents* are tasked with either carrying out a prescribed mission or harvesting resources from the environment (e.g., wind energy).
- *Exploration agents* are supplemental agents that are tasked *solely* with sensing; they have no mission to carry out, nor do they harvest the ambient energy.

In the AWE system of Fig. 1b, the exploitation agents are the BATs, which harvest wind energy. The exploration agents are supplemental quadcopters, which serve the sole role of measuring the local wind speed. Here, the exploitation agents are also capable of measuring the wind speed, rendering the supplemental quadcopters optional for the application.

In terms of the underlying control system, the control variable is the vector of spatial location commands to the agents, which are in general subject to mobility constraints. Performance is measured in terms of a *reward function*, $r(z,t)$, which characterizes the value of an exploration agent being located at location z at time t. Given a control policy, which will be denoted by π in the paper, the *realized* reward at time t, denoted as $f_\pi(t)$ is equal to the sum of the individual rewards over all p exploitation agents:

$$f_\pi(t) = \sum_{i=1}^{p} r(z_i, t). \tag{1}$$

In the AWE case study, the individual rewards are taken as the local wind speeds.

III. Regret and Coverage Formulations

A. Regret

In this work, we define regret as the difference between achieved performance and the performance of a comparator that performs optimally given omniscient information, i.e.:

$$R_{\text{inst}}(\pi, t) = f^*(t) - f_\pi(t). \tag{2}$$

Here, $R_{\text{inst}}(\pi, t)$ is the *instantaneous* regret associated with control strategy π operating at time t, $f^*(t)$ is the reward at time t associated with an optimal omniscient strategy, and $f_\pi(t)$ is the reward achieved by strategy π at time t. We also define *cumulative* regret as:

$$R(\pi, t_k) = \sum_{i=1}^{k} \Big(f^*(t_i) - f_\pi(t_i) \Big) \tag{3}$$

and *average* regret as:

$$R_{\text{avg}}(\pi, t_k) = \frac{1}{k} \sum_{i=1}^{k} \Big(f^*(t_i) - f_\pi(t_i) \Big) \tag{4}$$

B. Coverage

Dynamic coverage is a quantity representing both the *amount* of measurements taken near a point and the *quality* of those measurements, where sensing quality is quantified by a *sensing function*, denoted S_i. One candidate sensing function, which we use in the AWE example, is:

$$S_i(z, z_i(t)) = Ae^{-(\frac{\|z - z_i(t)\|}{r})^2} \tag{5}$$

where $S_i(z, z_i(t))$ is the value of the sensing function of agent i at location z and time t, $z_i(t)$ is the location of sensing agent i at time t, and A and r are parameters related to magnitude and width of the sensing function, respectively. The sensing function specifically characterizes the ability of a sensor to measure the environment at location z, based on how far that sensor is from z.

Based on our sensing functions, we now define our model of coverage dynamics as follows:

$$\dot{q}(z,t) = (1 - q(z,t)) \sum_{i=1}^{n} S_i(z, z_i(t)) - \alpha q(z,t) \tag{6}$$

where $q(z,t) \in [0,1]$ is coverage as a function of spatial parameter, n is the total number of sensing agents, and α is a parameter defining the "coverage loss rate." The parameter α can be tuned to match the time scale of the environment, reflecting the fact that old measurements lose value over time and allowing coverage to both increase and decrease. Coverage at any spatial location is continuously defined on the interval $[0,1]$ (noting that if $q(z,0) \in [0,1]$, the first term of $\dot{q}(z,t)$ will result in an increase in $q(z,t)$ bounded above by 1, and the second term will result in a decrease in $q(z,t)$ bounded below by 0). These features differentiate the proposed formulation from non-decreasing and boolean coverage formulations that can be found elsewhere in the literature. In this work, we will also examine average coverage, denoted by q_{avg} and obtained by integrating coverage over the spatial domain:

$$q_{\text{avg}}(t) \triangleq \frac{1}{z_m - z_0} \int_{z_0}^{z_m} q(z,t)\, dz \tag{7}$$

With constant sensor placements (and therefore constant $S_i(z, z_i(t))$), coverage will converge to an equilibrium value that depends on the relative values of A and α. However, with effective control, the sensing will be time-varying due to mobile sensors, so coverage will only converge to within some set, exhibiting limit cycle type behavior. For the AWE case study with one BAT and three auxiliary drones, an example of the spatiotemporal behavior of coverage can be seen in Fig. 2.

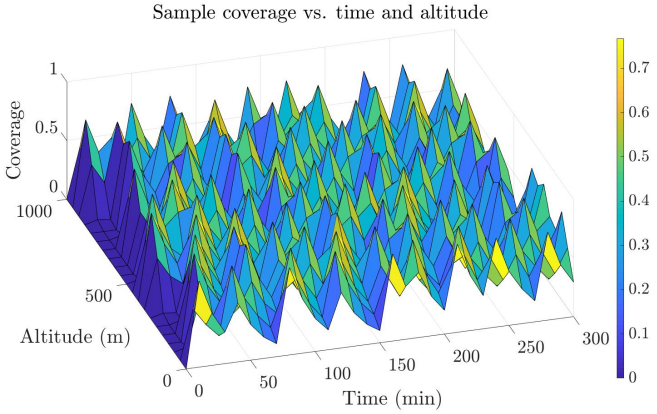


Fig. 2: Sample coverage surface plot. Limit cycle type behavior can be seen in the repetitive cycles experienced at each spatial location.

IV. Bounding Dynamic Coverage

Due to a lack of direct dependence on the stochastic resource, performance guarantees for dynamic coverage are more easily derived than are (meaningful) performance guarantees for regret. To this end, we use this section to specify methods to determine upper and lower bounds for the maximum achievable average coverage (MAAC), the largest *reachable* average value of the state vector $\underline{q}$, where $\underline{q} = [\underline{q}(z_1), \underline{q}(z_2), ..., \underline{q}(z_m)]^T$, for some unspecified time. This definition depends on the definition of reachability:

Definition 1: A coverage state $\underline{q}$ is defined to be reachable if there exists some τ and some sensor trajectory vector $\underline{z}(t) = [z_1(t), z_2(t), ..., z_n(t)]^T$ defined for $t \in [0, \tau]$ such that for all spatial locations i, subject to $q(z_i, 0) = 0$ and the dynamic model in (6), $q(z_i, \tau) = \underline{q}(z_i)$. The set of all reachable states $\underline{q}$ is denoted by Q.

The utility of this definition is that a coverage state is deemed unreachable if it is impossible to create some control trajectory that is capable of reaching that state. Unreachable states represent coverage levels unobtainable without adding more sensing capability to the system. We now also define maximum achievable average coverage (MAAC) as:

$$q_{\text{MAAC}} = \max_{\underline{q} \in Q} \frac{1}{m} \sum_{i=1}^{m} q(z_i) \tag{8}$$

$$\underline{q}_{\text{MAAC}} = \arg\max_{\underline{q} \in Q} \frac{1}{m} \sum_{i=1}^{m} q(z_i) \tag{9}$$

where q_{MAAC} is the maximum achievable value of coverage averaged across the domain discretized into m distinct locations, $\underline{q}_{\text{MAAC}}$ is the state vector corresponding to MAAC, and Q is the set of all *reachable* values of the coverage state vector. Noting that the set Q is closed and bounded (i.e., compact), this maximum is guaranteed to exist. Note that the difficulty in calculating q_{MAAC} lies entirely in identifying a set of constraints that dictate the set Q of reachable states. However, in order to bound q_{MAAC}, we do not need to create a set solely consisting of reachable states; we merely need one that is guaranteed to contain q_{MAAC}. To this end, we introduce the following lemma, which states that it must be possible to achieve an instantaneously constant (zero time derivative) average coverage at the MAAC state:

Lemma 1: For any MAAC state, $\underline{q}_{\text{MAAC}}$ (9), there exists $\underline{z}$ for which $\frac{1}{m} \sum_{i=1}^{m} \dot{q}(z_i) \triangleq \dot{q}_{\text{avg}} = 0$ where $\dot{q}(z_i)$ is the time derivative of coverage as defined in (6).

Proof: We will show that it is possible to choose a sensor trajectory vector, $\underline{z}(t)$, such that the left and right limits of $\dot{q}_{\text{avg}}$ both converge to the same nonnegative value for any MAAC state. We begin with the left limit. By continuity of $q(z,t)$ and the fact that $\underline{q}_{\text{MAAC}}$ is a local maximizer, it follows that:

$$\lim_{t \to t_{\max}^-} \dot{q}_{\text{avg}} \geq 0 \tag{10}$$

for some time $t_{\max}$ corresponding to a local maximal value of average coverage. Thus, the value of the derivative of coverage using a left hand limit is nonnegative.

To address the right limit, we need to show that for some $\underline{z}$, $\dot{q}$ will be continuous (and $\dot{q}$ will be equal to its left hand limit). To do so, we start by examining the dynamic model (6), noting that if $\dot{q}(z_i, t_{\max})$ is continuous for all z_i at $t_{\max}$, average coverage will also be continuous as it is a sum of a finite number of continuous functions. Next, as the right hand side of (6) is comprised solely of bounded terms with no division, $\dot{q}(z_i,t)$ is finite and therefore $q(z_i,t)$ is continuous. Finally, note that it is possible to choose instantaneously stationary sensor locations (constant $\underline{z}$) such that $\sum_{i=1}^{n} S_i(z, z_i(t))$ is instantaneously continuous. As is is possible to represent $\dot{q}(z_i,t)$ as the finite sum and product of continuous functions, it is possible to choose $\underline{z}$ such that $\dot{q}_{\text{avg}}$ will be continuous. This continuity combined with (10) guarantees that the $\dot{q}_{\text{avg}}$ exists for some $\underline{z}$ and is nonnegative. Given that $\dot{q}_{\text{avg}}$ exists for some maximum, $\dot{q}_{\text{avg}} = 0$ is guaranteed, proving the lemma. ■

Combining the results of Lemma 1, which demonstrate that the MAAC state is one where the time derivative of average coverage is equal to zero, with additional constraints that are imposed by the number of agents in the system, an upper bound on MAAC can be computed as the solution to a constrained optimization problem, as specified in the following theorem:

Theorem 1: Given a spatial domain discretized into m locations, n sensing agents, coverage parameters A and α, and the coverage dynamics (6), the solution to the following optimization problem will be an upper bound on MAAC:

$$\max_{\underline{q}, \underline{z}} \quad \frac{1}{m} \sum_{i=1}^{m} q(z_i) \tag{11}$$

$$\text{subject to :} \quad q(z_i) \leq \frac{nA}{nA+\alpha} \; \forall \; i \tag{12}$$

$$\sum_{j=1}^{m} \left((1 - q(z_j,t)) \sum_{i=1}^{n} S_i(z_j, z_i(t)) - \alpha q(z_j,t) \right) \geq 0 \tag{13}$$

Proof: First, we will show that this optimization problem has a solution. To do this, we note that the objective function is linear in the decision variables therefore it is continuous. The domain of the optimization problem is bounded by:

$$\begin{aligned} 0 \leq q(z_i) \leq 1 \; \forall \; i = 1...m \\ z_1 \leq z_j \leq z_m \; \forall \; j = 1...n \end{aligned} \tag{14}$$

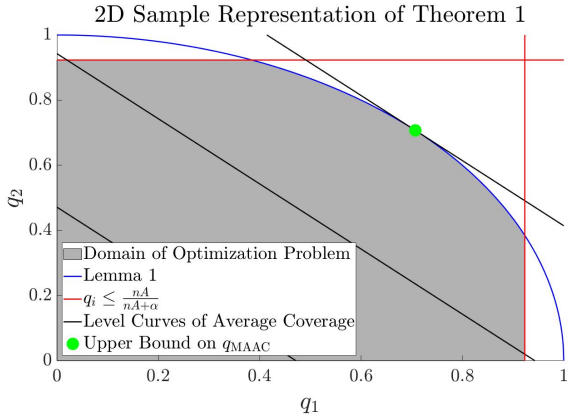


Fig. 3: Sample 2D problem constraints. Constraints are upper bounds on q_1 and q_2 separately, as well as a bound corresponding to Lemma 1. The gray region is a set containing the set Q of all reachable states. The maximum value of $q_{\text{avg}} = \frac{1}{2}(q_1+q_2)$ is the point where the gray region intersects the greatest level curve of q_{avg}.

Because the domain is closed and bounded, a maximum is guaranteed to exist. In the multidimensional case, the second set of inequalities will hold individually for each dimension, defining the boundaries of the domain.

To prove that the optimization problem yields an upper bound on maximum achievable coverage, we must show that the maximum achievable coverage is contained in the domain of the optimization problem. We do this by evaluating each constraint to show that it does not eliminate the MAAC state.

The first set of constraints is an element-by-element consideration of maximum achievable coverage. According to the dynamic model (6), $q(z_i,t)$ will be maximized when $\sum_{i=1}^{n} S_i(z,t)$ maintains its maximal value nA and the system is allowed to converge to an equilibrium. In this case:

$$(1 - q_{\max}(z_i))nA - \alpha q_{\max}(z_i) = 0 \tag{15}$$

$$q_{\max}(z_i) = \frac{nA}{nA+\alpha} \tag{16}$$

As this is an upper bound on the coverage that can be attained at any single point, the first set of constraints does not eliminate any reachable states.

Finally, the last constraint does not eliminate the MAAC state by Lemma 1. Since the optimization problem has a solution and the MAAC state is in the domain of the problem, the solution is guaranteed to be at least as large as MAAC and is therefore an upper bound. ■

This constrained optimization is illustrated in Fig. 3 for a sample problem with only two spatial locations (and therefore two coverage state variables, q_1 and q_2). The red upper bounds correspond to constraints imposed by the number of sensors, given mathematically in (12), and the blue curve corresponds to the bound from Lemma 1. Now that we have derived upper bounds on MAAC, we will derive lower bounds:

Theorem 2: A lower bound on MAAC can be obtained as the solution to the following optimization problem, given an

TABLE I: Statistical Model Hyperparameters

σ	l_z	l_t
5.1 m/s	270 m/s	22 min

initial state $\underline{q}(z_i(0), 0)\ \forall\ i$ and the dynamic model (6):

$$\begin{aligned} \max_{\underline{z}(t)} \quad & \frac{1}{m}\sum_{i=1}^{m} q(z_i, t_f) \\ \text{subject to :} \quad & \dot{q}(z,t) = (1 - q(z,t))\sum_{i=1}^{n} S_i(z, z_i(t)) \\ & \qquad - \alpha q(z,t) \end{aligned} \tag{17}$$

Proof: By the definition of MAAC, any reachable state will correspond to a lower bound, so a search for a lower bound can be done by simply beginning at the initial state of zero coverage and searching for the control input sequence that maximizes average coverage. Since any state found this way will have a corresponding control sequence, it must be reachable and will therefore be a lower bound on MAAC. ■

V. Predictive Modeling and Simulation

Using the AWE system as a case study, we now demonstrate the correlation between dynamic coverage and regret through simulation. We begin by reviewing statistical modeling tools from [26] and [27] that are used to (i) generate a synthetic data set for simulations and (ii) generate real-time characterizations of regret statistics during the execution of candidate control strategies.

We begin by defining a mean and covariance function for our reward. Specifically, we consider a mean function of zero in this work, along with a squared exponential covariance kernel, given by:

$$k(z,t,z',t') = \sigma^2 e^{-\frac{(z-z')^2}{2l_z^2}} e^{-\frac{(t-t')^2}{2l_t^2}}. \tag{18}$$

where (z,t) and (z',t') are the two points for which the covariance is calculated. The parameters σ, l_z, and l_t, which are given for the AWE study in Table I, are commonly referred to as hyperparameters and can be adjusted in order to reflect the desired statistical properties of the model.

To generate synthetic data, we generate a white noise signal and use the covariance kernel as a coloring filter to transform the signal into one with the desired hyperparameters. Full details are given in [27].

To maintain running estimates of conditional reward statistics to be used for control, Gaussian Process (GP) modeling is used. GP modeling provides a framework for maintaining calculations of prediction mean and prediction variance, which represent the conditional mean and variance of the prediction error, conditioned upon data collected up until that point in time, and are given by [28]:

$$\begin{aligned} \mu_* &= K(z_*, z)K(z,z)^{-1} y \\ \sigma_*^2 &= K(z_*, z_*) - K(z_*, z)K(z,z)^{-1}K(z, z_*) \end{aligned} \tag{19}$$

Here, μ_* and σ_*^2 are the prediction mean and prediction variance respectively of the set of test points z_* based on historical data y at points z. The function $K(\cdot,\cdot)$ is the matrix form of the covariance kernel whose dimensions are given by the lengths of the two input vectors.

A. Upper Confidence Bound Control Strategy

In this work, we take the time step to be 10 minutes such that the system can reasonably traverse the domain within a single time step, reducing the decision variable to the next operating location. The decision of which spatial location(s) to visit at each time step (altitudes, in the case of the AWE system) is made in this work through an upper confidence bound (UCB) [29] control strategy. This strategy explicitly trades off exploration and exploitation through the maximization of the following acquisition function:

$$\alpha_t(\mu_t(z), \sigma_t(z)) = \mu_t(z) + \sqrt{\beta_t}\sigma_t(z), \tag{20}$$

where β_t is a parameter that defines the relative weighting of $\mu_t(z)$ and $\sigma_t(z)$ in the acquisition function. The control decision associated with this acquisition function is:

$$z_{\text{next}} = \arg\max_z \alpha_t(\mu_t(z), \sigma_t(z)) \tag{21}$$

Each time the acquisition function is evaluated, points with high variance and points with high expected value are both valued. Larger β_t values correspond to greater emphasis on exploration over exploitation.

B. Simulation Results

In this section, we present simulations of the AWE system with several auxiliary sensing agents, an example of which is shown in Fig. 1b. Here, we consider one BAT with various numbers of sensing agents operating in a 1-dimensional domain where wind speed varies with time and altitude. We control the BAT with the upper confidence bound (UCB) algorithm with $\sqrt{\beta_t} = 5$ and the sensors with a variety of control algorithms based on current coverage level. For each combination of sensor control algorithm and number of sensors, we ran a suite of simulations and plotted the average regret vs. average coverage averaged over the set of simulations. Varying the number of sensors and associated control algorithms allows for direct variation of the system's exploration capabilities, affecting the average coverage, which in turn indirectly affects the average regret. As can be seen in Fig. 4, this results in a strong correlation (Spearman correlation coefficient = -0.912) between regret and average coverage.

We further analyze the bounds on maximum achievable average coverage by taking results from simulations and comparing them to the calculable bounds. We obtained the results in Table II using a sensor control strategy that optimizes coverage gain over the current time step irrespective of future time steps. The deviation between the actual maximum and the lower bound on MAAC is due to suboptimality of the sensor control algorithm.

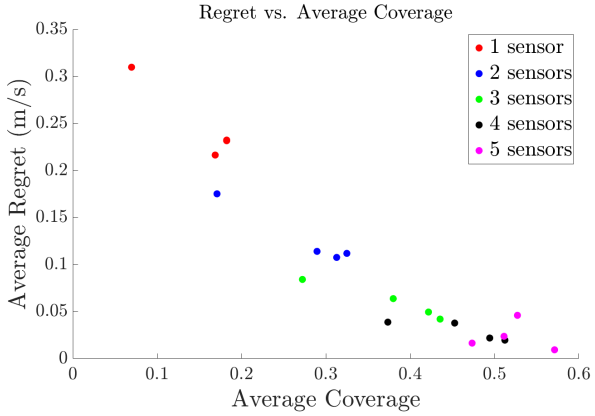


Fig. 4: Correlation between coverage and regret

TABLE II: Maximum Achievable Average Coverage (MAAC) Bounds vs. Actual

Upper Bound	Lower Bound	Actual Maximum
0.5384	0.4598	0.451

VI. Conclusion

This work presented a novel dynamic coverage formulation with properties that allowed for a link between dynamic coverage and regret for systems operating in spatiotemporally varying environments. Specifically, we presented upper and lower bounds on maximum achievable average coverage (MAAC) and also showed that a simple algorithm could come close to attaining the lower bound. We then showed a link between average regret and average coverage.

References

[1] A. Solanas and M. Garcia, "Coordinated multi-robot exploration through unsupervised clustering of unknown space," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2004.

[2] M. Quann, L. Ojeda, W. Smith, D. Rizzo, M. Castanier, and K. Barton, "An Energy-Efficient Method for Multi-Robot Reconnaissance in an Unknown Environment," *American Control Conference*, 2017.

[3] W. Bentz, T. Hoang, E. Bayasgalan, and D. Panagou, "Complete 3-D dynamic coverage in energy-constrained multi-UAV sensor networks," *Autonomous Robots*, vol. 42, pp. 825–851, 2018.

[4] C. Meinig, N. Lawrence-Slavas, R. Jenkins, and H. Tabisola, "The use of Saildrones to examine spring conditions in the Bering Sea: Vehicle specification and mission performance," 10 2015, pp. 1–6.

[5] A. E. Hartl and A. P. Mazzoleni, "Dynamic modeling of a wind-driven tumbleweed rover including atmospheric effects," *Journal of Spacecraft and Rockets*, vol. 47, no. 3, pp. 493–502, 2010.

[6] A. Bafandeh and C. Vermillion, "Real-time altitude optimization of airborne wind energy systems using Lyapunov-based switched extremum seeking control," *American Control Conference*, pp. 4990–4995, 07 2016.

[7] A. Baheri, S. Bin-Karim, A. Bafandeh, and C. Vermillion, "Real-time control using Bayesian optimization: A case study in airborne wind energy systems," *Control Engineering Practice*, vol. 69, pp. 131–140, 12 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967066117302101?via%3Dihub

[8] A. Bafandeh, S. Bin-Karim, A. Baheri, and C. Vermillion, "A Comparative Assessment of Hierarchical Control Structures for Spatiotemporally-Varying Systems, with Application to Airborne Wind Energy," *Control Engineering Practice*, vol. 74, pp. 71–83, 12 2018.

[9] S. Bin-Karim, A. Bafandeh, A. Baheri, and C. Vermillion, "Spatiotemporal Optimization Through Gaussian Process Based Model Predictive Control: Case Study in Airborne Wind Energy," *IEEE Transactions on Control Systems Technology*, vol. 27, pp. 798–805, 3 2017.

[10] L. Dunn, C. Vermillion, F. K. Chow, and S. Moura, "On Wind Speed Sensor Configurations and Altitude Control in Airborne Wind Energy Systems," *American Control Conference*, 2019.

[11] O. Yallop, "Inflatable turbines: the windfarms of the future?" 2014, accessed: 2019-03-29.

[12] C. Archer, "Wind Profiler at Cape Henlopen," 2014. [Online]. Available: https://www.ceoe.udel.edu/our-people/profiles/carcher/fsmw

[13] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Gaussian Process Bandits without Regret: An Experimental Design Approach," *CoRR*, 2009. [Online]. Available: http://arxiv.org/abs/0912.3995

[14] A. Krause and C. S. Ong, "Contextual Gaussian Process Bandit Optimization," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2011, pp. 2447–2455. [Online]. Available: http://papers.nips.cc/paper/4487-contextual-gaussian-process-bandit-optimization.pdf

[15] A. Garivier and E. Moulines, "On Upper-Confidence Bound Policies for Switching Bandit Problems," in *Algorithmic Learning Theory*, J. Kivinen, C. Szepesvári, E. Ukkonen, and T. Zeugmann, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 174–188.

[16] L. Wei and V. Srivatsva, "On abruptly-changing and slowly-varying multiarmed bandit problems," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 6291–6296.

[17] O. Besbes, Y. Gur, and A. Zeevi, "Stochastic Multi-Armed-Bandit Problem with Non-stationary Rewards," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 199–207.

[18] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The non-stochastic multiarmed bandit problem," *SIAM journal on computing*, vol. 32, no. 1, pp. 48–77, 2002.

[19] A. Slivkins and E. Upfal, "Adapting to a Changing Environment: the Brownian Restless Bandits," in *21st Conference on Learning Theory (COLT)*, July 2008, pp. 343–354. [Online]. Available: https://www.microsoft.com/en-us/research/publication/adapting-to-a-changing-environment-the-brownian-restless-bandits/

[20] A. Jadbabaie, A. Rakhlin, S. Shahrampour, and K. Sridharan, "Online optimization: Competing with dynamic comparators," in *Artificial Intelligence and Statistics*. PMLR, 2015, pp. 398–406.

[21] S. Bubeck, Y. Li, H. Luo, and C.-Y. Wei, "Improved path-length regret bounds for bandits," in *Conference On Learning Theory*. PMLR, 2019, pp. 508–528.

[22] L. Zhang, S. Lu, and T. Yang, "Minimizing dynamic regret and adaptive regret simultaneously," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 309–319.

[23] B. Liu, O. Dousse, P. Nain, and D. Towsley, "Dynamic coverage of mobile sensor networks," *IEEE Transactions on Parallel and Distributed systems*, vol. 24, no. 2, pp. 301–311, 2012.

[24] A. Sekhar, B. Manoj, C. Siva, and R. Murthy, "Dynamic coverage maintenance algorithms for sensor networks with limited mobility," in *Third IEEE International Conference on Pervasive Computing and Communications*. IEEE, 2005, pp. 51–60.

[25] D. Panagou, D. M. Stipanović, and P. G. Voulgaris, "Distributed dynamic coverage and avoidance control under anisotropic sensing," *IEEE Transactions on Control of Network Systems*, vol. 4, no. 4, pp. 850–862, 2016.

[26] B. Haydon, J. Cole, L. Dunn, P. Keyantuo, T. Chow, S. Moura, and C. Vermillion, "Empirical Regret Bounds for Control in Spatiotemporally Varying Environments: A Case Study in Airborne Wind Energy," *ASME Dynamic Systems and Control Conference*, 10 2019.

[27] ——, "Empirical Regret Bounds for Control in Spatiotemporally Varying Environments: A Case Study in Airborne Wind Energy," *Journal of Dynamic Systems, Measurement and Control (under review)*.

[28] C. E. Rasmussen and H. Nickisch, "GPML Matlab Code version 4.2," 2018. [Online]. Available: http://www.gaussianprocess.org/gpml/code/matlab/doc/index.html

[29] D. Cox and S. John, "SDO: A statistical method for global optimization," *Multidisciplinary Design Optimization: State of the Art*, pp. 1241 – 1246 vol.2, 11 1992.