Robotics: Science and Systems 2020
Corvalis, Oregon, USA, July 12-16, 2020

Agbots 2.0: Weeding Denser Fields with Fewer Robots

Wyatt McAllister*, Joshua Whitman*, Joshua Varghese*, Allan Axelrod*, Adam Davis†, and Girish Chowdhary*

***Abstract*—This work presents a significantly improved strategy for coordinated multi-agent weeding under conditions of partial environmental information. We show that by using Entropic value-at-risk (EVaR) together with the Gittins index, agents can make intelligent decisions about whether to exploit the estimated distribution of weeds in the environment or to explore new areas of the environment. The use of this method improves the performance of agents in comparison to previous methods, resulting in a system which can weed denser fields using fewer robots. Furthermore, we show that for the reward function and environmental dynamics which represent the weeding problem, our system is able to perform comparably to the fully observed case over the real-world range of seed bank densities, while operating under partial observability.**

I. Introduction

The robotic weed management problem can be considered on an abstract level as the challenge of managing a large-scale system with a team of agents, where the objective is to prevent any part of the domain from remaining unattended for too long. Furthermore, the longer any part of the environment is left unattended, the longer it will take to weed. These relationships are not uniform over the entire domain, but depend on a spatially-varying stochastic process. These properties describe many problems beyond weed management; examples include destroying invasive underwater algae blooms, management of disasters such as oil spills or radiation leaks, curing a blight, and fighting wildfires.

Weed management and similar problems are challenging for a number of reasons. First, there is a sharp discontinuity in cost in the event that a weed patch grows too tall. Second, the cost changes dynamically across the field as the weeds grow. Third, agents are only able to exploit a highly rewarding state-action pair periodically, since weeds take time to grow. Planning with a sparse reward/cost function is a difficult problem in robotics that has not been fully solved [1], much less in a dynamic environment such as the one studied in this paper. One effective strategy is to design a less-sparse heuristic reward function [2], but this needs to be done carefully in order to produce performance that is still calibrated to the true objective. Another important challenge is managing the trade-off between exploring unknown parts of the field and exploiting knowledge of the parts that have been observed. This is a fundamental challenge in distributed learning problems, and is especially challenging within dynamic settings [3]. In some sense, the value of information gain [4] must be used, which previous work on the weeding problem failed to take into account [5].

The contribution of this work is that we employ existing bandit theory in an innovative manner to solve the path planning problem in the novel and industry-critical coordinated weeding domain. We extend the reward function from [5], based on weed height and the expected weeding time [6], [7]. Since these quantities grow continuously, the problem of a sparse reward function is eliminated. The overall objective remains the same: preventing any weed from growing too large to be eliminated. We use Entropic value-at-risk (EVaR) to modify the reward function to incorporate our confidence in the predictive model and manage the trade-off between exploration and exploitation [8]. This results in a significant performance improvement in terms of the number of agents required to weed fields with a given seed bank density. To optimize over this cumulative reward function [9], we frame the problem as a cooperative robotics problem involving foraging [10], using techniques from multi-agent task allocation [11].

A. Background: The Herbicide Resistant Weed Problem

Weed management has historically relied on a combination of crop rotation, mechanical weed control, and the use of herbicides [12]. The evolution of herbicide-resistant weeds, coupled with the fact that new herbicide discovery has ceased in the past 30 years, has resulted in a crisis for agricultural weed management [13], [14]. Crop losses due to herbicide resistant weeds are approximately half a billion per year, and may climb to $100 billion per year when chemical control is lost [15]. Evolution of resistance to multiple sites of herbicide action is accelerating, especially in the southern and north-central U.S. grain production regions [16]. Increasingly, farmers are only one site-of-action away from total loss of chemical control. For example, the five-way multiple resistant waterhemp (Amaranthus tuberculatus [Moq.] Sauer) in Illinois is now one gene away from total loss of chemical control [17]. Seeds with bred-in herbicide resistance are exacerbating the herbicide resistance problem in soybean production [18]. An alternative to chemical weeding is mechanical weeding, which most often targets young weeds, including germinating seeds and seedlings that are extremely vulnerable to physical damage.

* Coordinated Science Lab (CSL), University of Illinois, Urbana, IL 61801, USA. {wmcalli2,jewhitm2,joshuav2,allana2,girishc}@illinois.edu
† Dept. of Crop Sciences, University of Illinois. asdavis1@illinois.edu

When weeding mechanically before crop planting, superficial soil disturbance and subsequent soil cultivation can remove germinated weeds. However, hand weeding of young weeds at the two-leaf growth stage is difficult and impractical at scale. Mechanized inter-row cultivation has disadvantages, such as soil compaction due to use of heavy machinery, and an inability to work after the crop canopy closes.

Drones are ineffective for collecting data during much of the crop season, as canopy closure removes aerial visibility of the ground. A team of collaborative low-cost and lightweight mechanical weeding ground robots (termed here as *agbots*, illustrated in Figure 1) may be used to control herbicide-resistant weeds. Such a team of *agbots* can target weeds within and between crop rows, as opposed to tractors, combines, and planters, which cannot be used after the crop canopy closes. The *agbots* are ideal for working in dense fields, since they are small enough to drive over plants without damaging them, and do not compact the soil. This approach necessitates algorithms for managing large fields with the least number of robots.

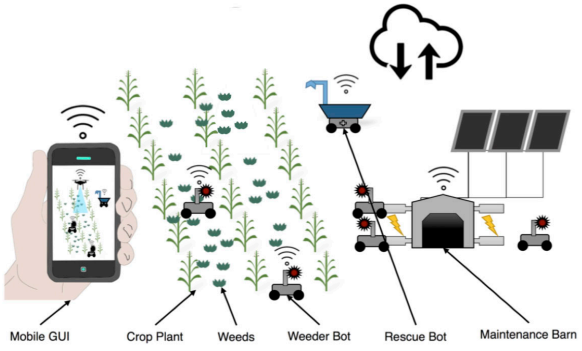

Fig. 1: The *agbot* solution for robotic mechanical weed control is a dynamically configured team of weeder bots, and automated maintenance barns for persistent autonomous weed control, leveraging collaboration.

(a) TerraSentia

(b) Weeding Apparatus

Fig. 2: Prototypes of TerraSentia Robot and Weeding Apparatus in the field.

Termination of weed seedlings within the critical weed-free period [19], where crops are most vulnerable, is essential to preventing crop yield losses in corn and soybeans [20]. For many crops, weeding may be done under the canopy, and therefore under conditions of partial environmental information.

To address this issue, several companies, such as TerraSentia [21], shown in Figures 2a and 2b, Ecorobotix [22], and Naio-Technologies [23], have developed small agricultural robots for autonomous weeding. For robots like these to be employed at scale, multi-agent planning strategies must solve the problem of coordination in field environments with limited observations.

This work significantly advances existing techniques for coordination of teams of mechanical weeding robots [5], [24]. We use the realistic simulation environment Weed World [24] to test our methods against previous benchmarks, with the TerraSentia robot as the basis for our simulation parameters. Previous work has used the Gittins Index to estimate the reward of weeding a particular row. We show in this work that utilizing the EVaR index can improve the planning performance by effectively managing the explore-exploit trade-off, thereby allowing a fewer number of robots to handle fields with denser populations of weeds.

B. Summary

Section II presents an overview of the Weed World simulation environment used in this work, the underlying generative model for weed growth used by this simulation framework, the planning algorithm and how it was revised, and the experiments used to validate this planning method. Section III explains the insight gained from each of the experiments. Section VI provides a discussion. Finally, conclusions and future work are presented in Section VII.

II. Methods

A. Simulation Environment

The simulation environment, called Weed World (shown in Figure 3), was developed to allow large-scale simulations of coordinated weeding algorithms for multi-robot planning in uncertain environments [24]. This environment incorporates a realistic weed growth model (described in II-B), as well as a framework for multi-agent collaboration, which enables a scalable amount of agents to easily share information. In this environment, the crops are assumed to be arranged in evenly spaced vertical rows.

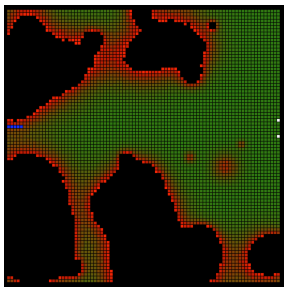

Fig. 3: Simulation environment, Weed World, created in Python. Each cell represents a small 0.8 m square portion of the field. The colors of the squares represent weed seed bank density. Darker colors represent higher density. The agents are shown in solid blue.

B. Weed Generation

The weed growth model is composed of a matrix of cells (each representing 0.8 m^2 for a total of 0.4 hectares) evolving according to a random process, forming a cellular automata model [25]. This study will determine the number of robots needed per acre for effective weed management. Seeds emerge from a finite, fixed seed bank, according to a time-inhomogeneous Poisson process. The parameters used, summarized in Table I, are aligned with the growth model for the common waterhemp weed species specified in [26]. The density of the seed bank in each cell, discretized in the coordinates (x, y, t), is $S(x, y, t)$, which is equal to S_0 (between 600 and 1560 seeds per cell) on average at time $t = 0$. The initial seed bank density in each cell $S_0(x, y)$ is chosen so that the Gini coefficient of concentration (GCC) between all the cells is between 0.31 and 0.35, which ensures that the relative density of weeds aligns with that seen in real experiments [26]. In order to achieve this distribution, we first give each cell a random density between zero and 20 percent of S_0, chosen uniformly at random, and then we create 50 patches of weeds with random centers and random radii up to 20 cells long, and fill those patches with an additional S_0 weeds distributed normally around each center.

The evolution of several fields is shown in Figure 4.

TABLE I: Seed Bank Density Parameters: Consistent with Those Found in [26]

Parameter	GCC	S_0	Np. Patches	Patch Radius
Range	[0.31,0.35]	[600,1560]	50	[0, 20]

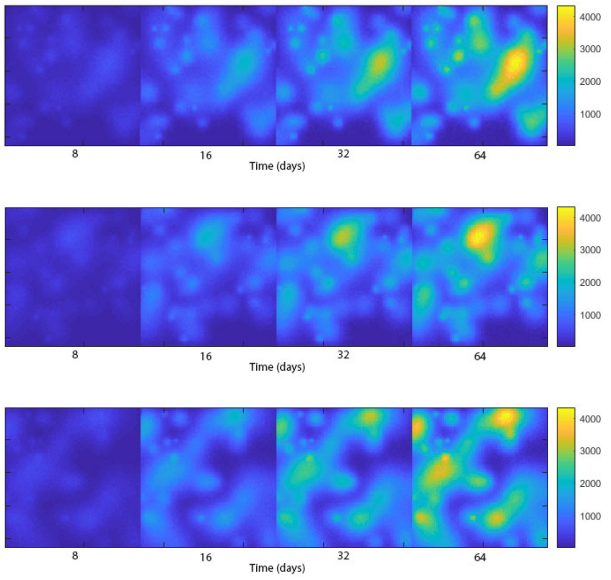

Fig. 4: The evolution of emerged seedling density for different fields over time, as simulated in the Weed World environment. Differing seed bank distributions result in differing evolutions.

After initializing the simulation, a certain number of days, d_0, are allowed to elapse before the agbots begin weeding. The number of emerging weeds in each cell, N_{emerge}, is a randomly generated Poisson variable with mean, $\lambda\,(x, y, t)$, such that 90 percent of the seed bank, $S\,(x, y, t)$, emerges in T_{total}, which is two months. This emergence rate is aligned with past work [27]–[31], all of which present measurements of the seed bank densities for various species of weeds, and provide an analysis of weed growth models for these species.

$$\lambda_t\,(x, y, t) = \frac{0.9 \cdot \Delta t \cdot S\,(x, y, t)}{T_{\text{total}}}, \quad \lambda_0 = \frac{0.9 \cdot d_0 \cdot S_0}{T_{\text{total}}} \tag{1}$$

$$N_{\text{emerge}}\,(x, y, t) = \text{Poi}\,(\lambda_t\,(x, y, t)) \tag{2}$$

$$S\,(x, y, t) = \max\left\{0, S_0 - \sum_{t'=t_0}^{t} N_{\text{emerge}}\,(x, y, t')\right\} \tag{3}$$

The weed density in each cell, $\zeta\,(x, y, t)$, grows as seeds emerge from the seed bank.

$$\zeta\,(x, y, t) = \sum_{t'=t_{\text{last weeded}}}^{t} N_{\text{emerge}}\,(x, y, t') \tag{4}$$

The maximum weed height in each cell, $\delta(x, y, t)$, increases at a fixed, upper-bounded, rate of Γ inches per day. This assumption is valid up until the point at which weeds grow explosively, and mechanical weeding becomes impossible.

$$\delta\,(x, y, t) = \left(\frac{t_{\text{curr.}} - t_{\text{last weeded}}}{60 \cdot 60 \cdot 24}\right)\Gamma \tag{5}$$

C. State, Action, and Reward Model

Here, $N_{\text{dim}} = 85$ is the number of rows, N_{agents} is the number of agents, $Y_{\text{len.}} = 64$ m is the length of each row, and $R_W\,(x, y, t)$ is the reward per cell (x, y) at time t.

The environmental state, S, depends on the x and y positions of each agent in I. The action, $a_i\,(t)$, is defined to be the target row chosen by each agent.

$$S \equiv \{1, ..., N_{\text{dim}}\} \tag{6}$$

$$I \equiv \{1, ..., N_{\text{agents}}\} \tag{7}$$

$$x_i\,(t) \in S, \quad y_i\,(t) \in S \quad \forall i \in I \tag{8}$$

$$a_i\,(t) \in A \equiv S \tag{9}$$

Since we require agents to finish the rows they begin, only the x location is relevant for the state.

$$x_i\,(t) \in S \quad \forall i \in I \tag{10}$$

We choose the reward associated with a cell to be the maximum height in that cell, prioritizing the regions with the tallest weeds, which prevents weeds from exceeding the maximum height our system can weed and thus causing major yield loss.

The reward for each row is the sum of the reward for each cell in the row, $R_W(x, y, t)$.

$$R_W(x, y, t) = \delta(x, y, t) \quad \forall x \in S, y \in S, \quad \forall i \in I \tag{11}$$

The agents keep track of the estimated density and maximum height for each observed cell, using this to estimate a total scalar reward for each row. This is the only required information for the reward.

$$R_i(a_i(t)) = \sum_{y=1}^{N_{\text{dim}}} R_W(a_i(t), y, t) \quad a_i(t) \in A \tag{12}$$

D. Previous Planning Algorithm

In [24], a time delayed reward was used, where each agent receives its reward for a row after completing it.

The planned operation time for a given row is the sum of the time it takes to move to the proposed row, $T_{\text{to row}}$, the time it takes to move down it, $T_{\text{down row}}$, and the time it takes to weed all the cells in the row, $T_{\text{weed row}}$, which depends on the weed density of each cell but is at most 2 minutes per cell.

$$T_i(x_i(t), a_i(t)) = T_{\text{to row}} + T_{\text{down row}} + T_{\text{weed row}} \tag{13}$$

$$T_{\text{to row}} = \frac{(a_i(t) - x_i(t))}{v} \tag{14}$$

$$T_{\text{down row}} = \frac{Y_{\text{len.}}}{v} \tag{15}$$

$$T_{\text{weed row}} = \min\left\{2\min \cdot N_{\text{dim}}, T_{\text{kill}} \sum_{y(t)=1}^{N_{\text{dim}}} \zeta(x, y, t)\right\} \tag{16}$$

The Gittins Index, $G(X_i)$, is known to be an optimal metric for planning on tasks with an uncertain termination time and *known* statistics [6]. Here, x is the state of the bandit, τ is the stopping time, r is the reward, and γ is the discount factor.

$$G(x(t)) = \sup_{\tau}\left\{\frac{\mathbb{E}\left[\sum_{t=0}^{\tau}\gamma^t r(x(t)) | x(0) = x\right]}{\mathbb{E}\left[\sum_{t=0}^{\tau}\gamma^t | x(0) = x\right]}\right\} \tag{17}$$

For our domain, the termination time, τ, is not a planning parameter, but is equal to $T_i(x_i(t), a_i(t))$, and the reward is delayed until that time. This resulted in a heuristic index based on Gitten's Index, $\bar{G}_i(a, x)$, as follows:

$$\bar{G}_i(a, x) = \frac{\gamma^{T_i(x_i(t), a_i(t))} R_i(a_i(t))}{\sum_{t=0}^{T_i(x_i(t), a_i(t))}\gamma^t} \tag{18}$$

E. Revised Planning Algorithm

Our goal is to improve the planning algorithm from [24], which placed full confidence in the estimated reward for each row, to more effectively address the uncertainty of our model of this dynamic field environment. We want to find an optimization index which yields improved explore-exploit performance.

The difference between this work and [24] is that for unobserved rows, we now have a prediction for the reward, which varies across the field, and may grow more accurate as we explore the space further. We need to account for the value of information gained for a candidate row via predictive inference.

Entropic value-at-risk (EVaR) [8] is a principled way to optimize with regard to the reward and information gain. The parameter $\alpha \in (0, 1)$ is our confidence in our reward estimate, and X is the reward distribution.

$$\text{EVaR}[X; 1-\alpha] := \inf_{\eta > 0}\left\{\frac{1}{\eta}\ln\left(\mathbb{E}_P\left[e^{\eta X}\right]/\alpha\right)\right\} \tag{19}$$

EVaR is an index based on the Chernoff bound.

$$\mathbb{P}(X \geq \text{EVaR}[X; 1-\alpha]) \leq \alpha \tag{20}$$

As in [32], we set our confidence α as follows:

$$\alpha = e^{-D_{KL}(Q||P)} \tag{21}$$

Here, Q and P are the distributions for the emergence time before and after a given measurement respectively, and D_{KL} is the Kullback-Leibler divergence, which uniquely quantifies the information gain of Q relative to P [33]. We obtain a modified version of the equation for EVaR:

$$\text{EVaR}[X; 1-\alpha] := \inf_{\eta > 0}\frac{1}{\eta}\left(\ln\left(\mathbb{E}_P\left[e^{\eta X}\right]\right) + D_{KL}(Q||P)\right) \tag{22}$$

Equation 22 highlights that EVaR is a probabilistically meaningful optimization equation that includes an "exploration bonus" based on the information gain, represented by the D_{KL} term. In this way we have a probabilistically meaningful and inquisitive planning algorithm. In addition, EVaR is linear just like the expected value, meaning that linear transformations of the problem space will not result in different solutions. This is not the case for other exploration bonus methods [34]–[37]. Finally, EVaR exhibits the properties of strong monotonicity [38] and stop-loss ordering [39], which both speak to the increased ability to discern optimality amongst similarly valued random variables, as compared to the expected value, average value-at-risk, and value-at-risk.

While most work in financial mathematics assumes a constant confidence parameter, α, in our case, this parameter changes with each measurement made. However, the partial stochastic ordering [40] guarantees of strong monotonicity and stop-loss ordering hold at each time step, and since the value of α is constant throughout the field at each time step, these ordering guarantees hold at each time step. Finally, as sampling goes to infinity, the information gain converges to zero; therefore, our confidence function α converges to one, and our EVaR based index converges to the Gittins index.

In order to compute the KL divergence, we create a distribution for the emergence time for each cell, T_{emerge}.

We keep track of the time since we have weeded a cell using the current weeding time.

$$T_{\text{since weed}}(x, y, t) = t - T_{\text{last weed}}(x, y) \tag{23}$$

During an observation, we calculate the time since emergence and the emergence time as below, based on the observed maximum height.

$$T_{\text{since emerge}}(x, y, t) = \frac{\hat{\delta}(x, y, t)}{\Gamma} \tag{24}$$

$$\begin{aligned} T_{\text{emerge}}(x, y, n) = T_{\text{since weed}}(x, y, t) \\ - T_{\text{since emerge}}(x, y, t) \end{aligned} \tag{25}$$

We then keep an average of the emergence time for each visit to a cell.

$$\bar{T}_{\text{emerge}}(x, y) = \frac{1}{N_{\text{visits}}} \sum_{n=1}^{N_{\text{visits}}} T_{\text{emerge}}(x, y, n) \tag{26}$$

At each time step in the simulation, we make a histogram of $\bar{T}_{\text{emerge}}(x, y)$ at times t and $t-1$ and construct probability mass functions P, and Q. These distributions are used to compute the KL divergence, which then goes into the information gain term within EVaR, where $|P|$ is the size of P.

$$D_{KL}(P||Q) = \sum_{i=0}^{|P|} P(i) \left(\log(P(i)) - \log(Q(i))\right) \tag{27}$$

Supposing that the empirical estimates are unbiased, both P and Q will converge to the same true pmf in the limit of observations. Therefore, $D_{KL}(P||Q)$ would asymptotically converge to zero, causing EVaR to converge to the Gittins index [41] over time.

Finally, in order to plan while incorporating our confidence in the environmental model, we utilize a similar algorithm to that used previously in [24], but with the new planning index which leverages EVaR. The average reward, $\bar{R}_i(a_i(t))$, is computed as the sum of rewards for all agents for each row weeded, divided by the total number of rows weeded, $N_{\text{rows weeded}}$.

$$\bar{R}_i(a_i(t)) = \frac{\sum_{t=0}^{t_{\text{curr.}}} \sum_{i=0}^{N_{\text{agents}}} R_i(a_i(t))}{N_{\text{rows weeded}}} \tag{28}$$

The information index of a row, $I(a_i(t))$, is the number of rows which would be newly observed by going to that row.

$$I(a_i(t)) = \sum_{i=-r_{obs}}^{r_{obs}} \mathbb{I}_{\{\text{is observed}(x=a_i(t)+i)\}} \tag{29}$$

We then compute a new Gittins index using EVaR and pick the maximum.

$$\hat{G}_i(a, x) = \frac{\gamma^{T_i(x_i(t), a_i(t))} \text{EVaR}\left[\bar{R}_i(a_i(t) I(a_i(t))); 1-\alpha\right]}{\sum_{t=0}^{T_i(x_i(t), a_i(t))} \gamma^t} \tag{30}$$

We are then able to choose the action which maximizes the new Gittins index for every agent.

$$a_i(t) = \arg\max_{a_i(t)} \hat{G}_i(a, x) \quad a_i(t) \in A \tag{31}$$

F. Overall Architecture

We assume an architecture in which agents are assigned rows from a centralized planner which has access to a global environmental model. This model aggregates the observations from all the agents at every step in the simulation. The overall process is as follows.

1) When agent i completes a row, it becomes idle and waits for a new row to be assigned according to Equation 31.
2) It proceeds to the new row and weeds it.
3) At every time step t during weeding, agent i sends its observations on visited rows to the centralized planner.

Before any row assignment, predictive inference is used to compute the EVaR index, which then goes into the bandit planner described in Equation 31, as summarized in Figure 5. We assume a discretized and fixed time step, in which observations are aggregated, processed, and sent to the centralized planner. Communication delays may change this time step slightly in a real-world scenario. However, as long as the time step is much smaller than the time taken to weed a row, the time step will not affect decision making.

In practice, communication delays can be minimized by assigning a team of agents to each region of a larger field, and using our method to coordinate their actions. This work presents a study on one-acre fields, in which the assumption of reliable communication is reasonable.

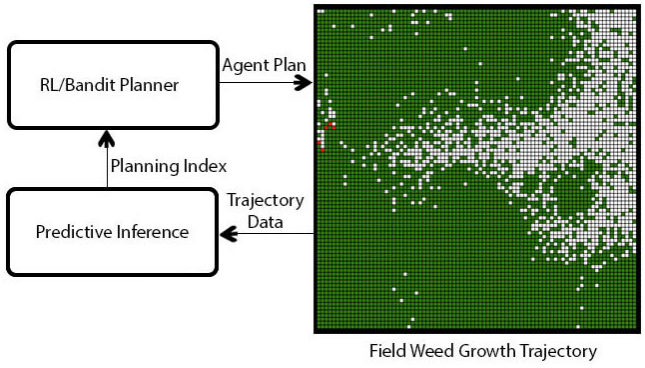

Fig. 5: **Overall Architecture:** The red squares within the portion of the figure showing the field represent the robotic agents.

G. Experiment Plan

This section presents an outline for the experiments to test for improvement over the planning algorithm in [5] (Section II-D), detailed in Section II-E. In [5], it was determined that wider ranges of observation for each agent improved performance in every case, by comparing the case where observation radius $r_{\text{obs}} = 1$ to the case where $r_{\text{obs}} = 0$, which had worse performance. It was also observed that beyond a critical point of two days for the number of days of weed growth allowed before the simulation starts (denoted d_0), the system could not succeed regardless of the number of agents used. Therefore, for these experiments, we set $d_0 = 1$ and set the observation radius $r_{\text{obs}} = 1$, which is the largest observation distance reasonable for robots operating under an occluded crop canopy. These experiments compare the performance of the prediction scheme from [5] with Gitten's Index used for the planning index, with the new prediction scheme using EVaR in the planning index. We do not use the fully-observed scenario, $r_{\text{obs}} = \infty$ with EVaR, as this would make exploration irrelevant since there is no uncertainty in the environmental model. We did not compare to the case of $r_{\text{obs}} = 0$, because the current technology for the robot prototypes used to determined the simulation parameters in this work allow for $r_{\text{obs}} = 1$ and the restriction on the amount of information would flatten the differences between the old method and EVaR.

We conduct eight experiments, each with 1000 trials with varying initial parameters shown in Table II, in a simulated field of 0.4 hectares, gridded in 0.8 m^2 cells. Each trial is run for 2 days of simulated time, as this was found to be the critical weeding period for the environment. The algorithm is run with the original prediction scheme, and the planner using the Gittins index (Old), where the planner has knowledge of each cell adjacent to one of the agents, in order to establish a benchmark of previous performance using the new Python Weed World simulation. The algorithm is then run with the EVaR planning index (New), with the old prediction scheme using the average reward in previously weeded rows, in order to observe how the revision to the planner affects performance. Next, the algorithm is run with full observability of the environment provided to the planner (Obs.), to see how performance improves. Last, we compare to the baseline case of a lawn mower pattern (Mow.), which we expect all other algorithms to outperform.

In Table II, an X denotes a parameter for a Monte Carlo run over the ranges shown in Table III. For S_0, the real-world range is between 600 and 1560 for the weed species of interest [26]. We chose to double this range for our simulations in order to find an upper bound on the seed bank density that can be weeded for each algorithm with the given number of agents. This better shows the feasibility of each method with respect to changes in these parameters, and allows us to more accurately determine the sensitivity of each algorithm to the weed density.

TABLE II: Initial Experimental Parameters: Here, N_{agent} is the number of agents, S_0 is the initial seed bank density. An X denotes a parameter for a Monte Carlo run over the ranges in Table III. We use a value of 15 for N_{agent} to ensure there are enough agents to weed the field. Here, 1080 is the median of the real-world range for S_0, which is determined to be between 600 and 1560, for the weed species of interest [26]. The original planning and prediction methods are from [5].

Exp.	1	2	3	4	5	6	7	8
Plan	Old	New	Old	Mow	Old	New	Old	Mow
Pred	Old	Old	Obs	NA	Old	Old	Obs	NA
N_{agent}	15	15	15	15	X	X	X	X
S_0	1080	1080	1080	1080	X	X	X	X

For Experiments 1 - 4, we used the values 15 for N_{agent}, and 1080 for S_0. These values are chosen to ensure that there are enough agents to weed the field, and that the robots have enough time to do so before the weeds grow too tall. The mean and standard deviation of the average reward for the environment (total reward over all the cells divided by the number of agents) are plotted at each time step. This provides a clear comparison between the different cases (the original prediction scheme versus fully observed case, the Gittins index versus the EVaR index, and the lawn mower patter), and showcases the effect of each method used on weeding performance. The general trend in the performance shows the stages of weeding for each algorithm tested, and the maximum possible reward each can achieve.

In Experiment 5-8, Monte Carlo runs are performed over a range of values for the parameters N_{agent}, and S_0, for the lawnmower pattern. A heat map of the terminal reward (total reward over all the cells at the end of the simulation divided by the number of agents) is plotted with respect to initial seed bank density S_0 and number of agents N_{agents}.

Within every Monte Carlo run, we consider a success to be a case in which an algorithm is able to keep the total maximum heights from each cell in the field, per agent, under the value 1000. This is an adjustable threshold, which physically represents a system which drives the weed height to an average of just over one-tenth of one inch per agent for each cell.

III. Results

IV. Experiments 1 - 4

In Figure 6, the mean and standard deviation of the average reward for the environment (total reward over all the cells divided by the number of agents) are plotted at each time step for each algorithm in Experiments 1-4. Though we see slight statistical differences in numerical performance, the general trend in the reward is the same for all of these algorithms, which all succeed in fields with these nominal parameters. The lawn mower pattern does worse than any of the others, as we expect.

TABLE III: Ranges for Experimental Parameters in Monte Carlo Runs: d_0 is then number of days of weed growth before the start of weeding, and r_{obs} is the observation radius. The real-world range for S_0 is between 600 and 1560, for the weed species of interest [26], but we chose to double this range in order to test the upper limits of our algorithms.

Parameter	Range
r_{obs}	1
d_0	1
N_{agent}	[5,20]
S_0	[600,3120]

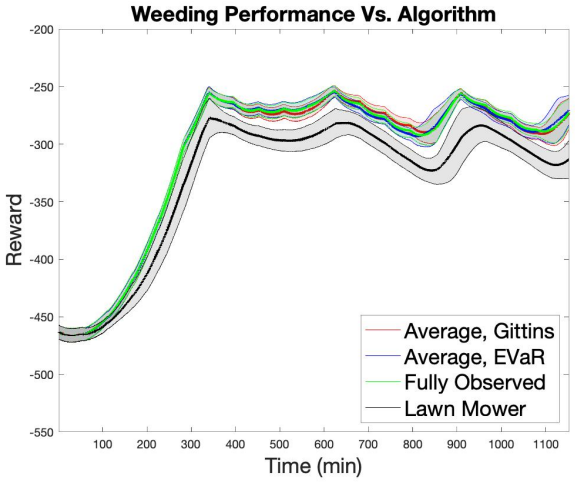

Fig. 6: **Weeding Performance vs. Algorithm**

V. Experiments 5 - 8

For each experiment, the contour map of the terminal reward for 1000 trials with number of agents and seed bank density is shown. The red end of the spectrum represents a terminal reward of zero, meaning the field has been weeded completely, and the blue end represents a high nonzero terminal reward, a strong failure case. Each blue circle represents a simulated trial, and the black dashed line represents the maximum seed bank density the system can weed for each number of agents.

In Figure 7, we examine the case when the Gittins index is used for the planner. *The maximum density that can be handled with eight agents is significantly less than the maximum of the real-world seed bank range.* When more agents are used, the system succeeds over the entire range of seed bank densities.

In Figure 8, we examine the case when EVaR is used for the planner. The system succeeds with only eight agents for the entire real-world seed bank densities range, though not for the entire worst-case range. When more than eight agents are used, the system succeeds over the entire range of seed bank densities. *Our experiment shows that in most real-world scenarios, EVaR will yield better performance at a lower cost.*

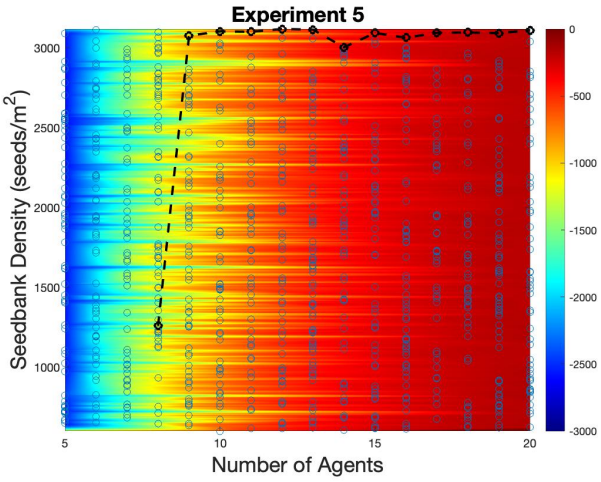

Fig. 7: **Number of Agents vs. Seed Bank Density, Gittins Index, Partially Observed**

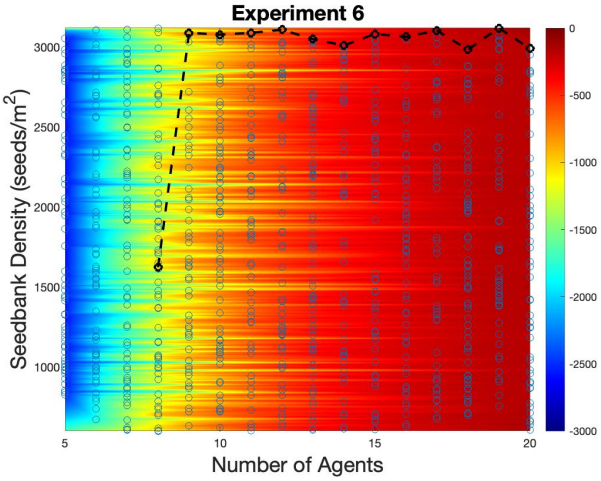

Fig. 8: **Number of Agents vs. Seed Bank Density, EVaR Index, Partially Observed**

In Figure 9, we examine the case with full environmental information. Full environmental information is unrealistic, but serves as a useful comparison point. We find that the system can handle extremely high seed bank densities with only eight agents, though quite the entire tested range. Seven agents are still insufficient to succeed at any realistic seed bank density even with full observability. This suggests that *eight agents is a critical number for one-acre fields*.

In Figure 10, we examine the case of the baseline planner using a lawn mower pattern. *From this figure, we see that with eight agents, the system fails for every seed bank density tested.* Nine agents are needed for success. This demonstrates that the lawn mower pattern does much worse than any algorithm tested in the main body of the text.

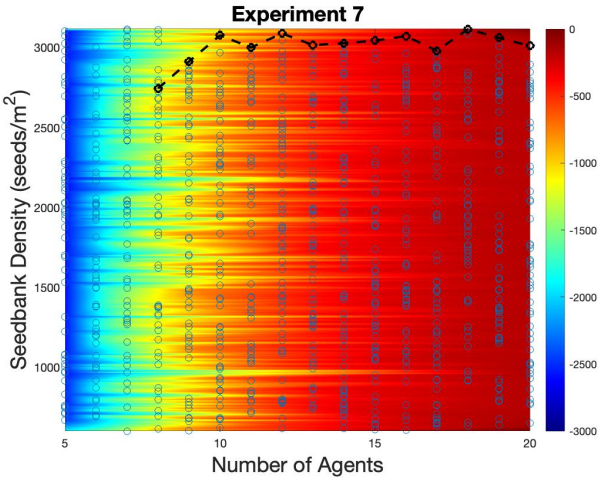

Fig. 9: **Number of Agents vs. Seed Bank Density, Full Environmental Information**

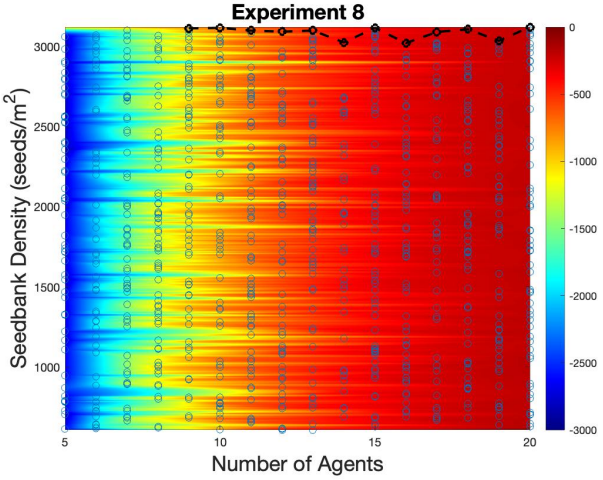

Fig. 10: **Number of Agents vs. Seed Bank Density, Gittins Index, Partially Observed**

VI. Discussion

These results show that compared to previous algorithms, the EVaR-based index enables teams of robots to weed fields with higher average background seed bank densities with the same number of agents. We see that EVaR even exhibits comparable performance to the fully observed scenario, over the real-world range of seed bank densities from 600 to 1560 for this weed species. Since our reward function is only concerned with the height of weeds, and since weeds emerge quickly after being destroyed and begin growing at a consistent rate, models of the seed bank density and growth patterns are not necessary for high performance in most realistic fields. Only when the field becomes unrealistically dense does the fully observed case start to outperform the case of EVaR with partial information.

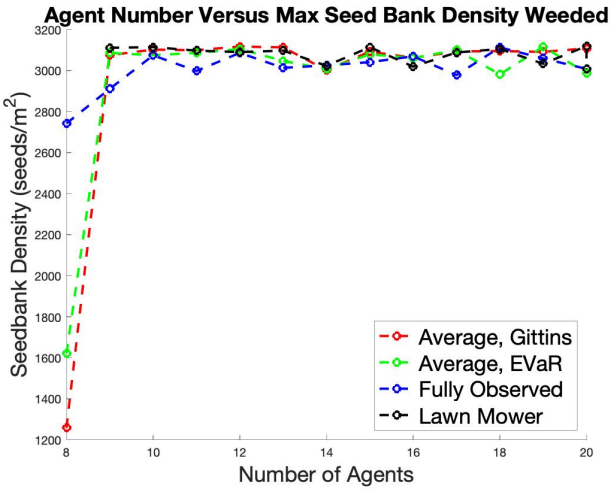

Fig. 11: **Number of Agents vs. Seed Bank Density, Summary:** We compare the maximum seed bank density weeded for a given number of agents for each algorithm.

The Weed World simulation environment establishes a worst-case scenario for weeding performance. We assume weeds grow aggressively, without growth being curbed by competition from crops or bad weather. Robotic agents use conservative estimates for speed and weeding time, which make effective use of each agent critical. In order to establish a baseline, the weeding time is much slower than that of the TerraSentia robot. The goal of the system is to ensure weeds never grow taller than the mechanical system can eliminate, entering a regime of explosive growth. For common waterhemp, which grows to have a large and strong stem, this can mean full yield loss, due to inability to harvest the crop with combine harvesters. Though, as shown in Figure 11 the use of EVaR only saves one robot per acre, and only does so for fields with high density, it guarantees that any field in this real-world range of seed bank densities can be managed without the risk of explosive weed growth.

VII. Conclusion

The use of the revised planning algorithm leveraging Entropic value-at-risk allows the use fewer robots to achieve the goal of preventing weeds from stifling crop yield. Our simulations show that fields with significantly greater and more challenging hidden seed banks of weeds can be managed effectively using this new approach. We observe that the ability to predict weed density evolution is surprisingly non-critical to the success of robots in this domain. Further work will be done in bringing machine learning and robotics expertise to this challenging and pressing problem in agriculture.

VIII. Acknowledgements

Support provided for this work by the joint USDA National Institute of Food and Agriculture and the National Science Foundation Cyber Physical Systems program (USDA NIFA 2018-67007-28379, NSF#1739874).

REFERENCES

[1] W. D. Smart and L. P. Kaelbling, "Effective reinforcement learning for mobile robots," in *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No. 02CH37292)*, vol. 4, pp. 3404–3410, IEEE, 2002.

[2] G. Flaspohler, V. Preston, A. P. Michel, Y. Girdhar, and N. Roy, "Information-guided robotic maximum seek-and-sample in partially observable continuous environments," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 3782–3789, 2019.

[3] A. Axelrod and G. Chowdhary, *The Explore-Exploit Dilemma in Non-stationary Decision Making under Uncertainty*, ch. The Explore-Exploit Dilemma in Nonstationary Decision Making under Uncertainty. 2198-4182, Springer international publishing, 1 ed., 2015.

[4] J. T. Kent, "Information gain and a general measure of correlation," *Biometrika*, vol. 70, no. 1, pp. 163–173, 1983.

[5] W. McAllister, D. Osipychev, G. Chowdhary, and A. Davis, "Agbots: Weeding a field with a team of autonomous robots," *Computers and Electronics in Agriculture*, vol. 163, p. 104827, 2019.

[6] R. Weber, "On the gittins index for multiarmed bandits," *The Annals of Applied Probability*, pp. 1024–1033, 1992.

[7] M. J. Matarić, "Reinforcement learning in the multi-robot domain," *Autonomous Robots*, vol. 4, no. 1, pp. 73–83, 1997.

[8] A. Ahmadi-Javid, "Entropic value-at-risk: A new coherent risk measure," *Journal of Optimization Theory and Applications*, vol. 155, no. 3, pp. 1105–1123, 2012.

[9] C. Amato, G. Chowdhary, A. Geramifard, N. K. Ure, and M. J. Kochenderfer, "Decentralized control of partially observable markov decision processes," in *The 52nd IEEE Conference on Decision and Control (CDC)*, 2013.

[10] Y. U. Cao, A. S. Fukunaga, and A. Kahng, "Cooperative mobile robotics: Antecedents and directions," *Autonomous Robots*, vol. 4, no. 1, pp. 7–27, 1997.

[11] B. P. Gerkey and M. J. Matarić, "A formal analysis and taxonomy of task allocation in multi-robot systems," *International Journal of Robotics Research*, vol. 23, no. 9, pp. 939–954, 2004.

[12] D. L. Shaner, "Lessons learned from the history of herbicide resistance," *Weed Science*, vol. 62, no. 2, pp. 427–431, 2014.

[13] I. Heap, "The international survey of herbicide resistant weeds," *Heap, I*, 2017.

[14] B. D. Maxwell, M. L. Roush, and S. R. Radosevich, "Predicting the evolution and dynamics of herbicide resistance in weed populations," *Weed Technology*, vol. 4, no. 1, pp. 2–13, 1990.

[15] M. Livingston, J. Fernandez-Cornejo, and G. B. Frisvold, "Economic returns to herbicide resistance management in the short and long run: The role of neighbor effects," *Weed Science*, vol. 64, no. sp1, pp. 595–608, 2016.

[16] M. V. Bagavathiannan and J. K. Norsworthy, "Multiple-herbicide resistance is widespread in roadside palmer amaranth populations," *PloS One*, vol. 11, no. 4, p. e0148748, 2016.

[17] C. M. Evans, *Characterization of a novel five-way-resistant population of waterhemp (Amaranthus tuberculatus)*. PhD thesis, University of Illinois at Urbana-Champaign, 2016.

[18] J. Gressel, A. J. Gassmann, and M. D. Owen, "How well will stacked transgenic pest/herbicide resistances delay pests from evolving resistance?," *Pest Management Science*, vol. 73, no. 1, pp. 22–34, 2017.

[19] B. D. Booth, S. D. Murphy, and C. J. Swanton, *Weed ecology in natural and agricultural systems*. CABI Pub., 2003.

[20] E. R. Page, D. Cerrudo, P. Westra, M. Loux, K. Smith, C. Foresman, H. Wright, and C. J. Swanton, "Why early season weed control is important in maize," *Weed Science*, vol. 60, no. 3, pp. 423–430, 2012.

[21] TerraSentia, "Terrasentia robot - earthsense, inc.," *TerraSentia*, 2017.

[22] ecorobotix, "Technology for the environment," *ecorobotix*, 12 2018.

[23] N. Technologies, "Autonomous weeding for agricultural robots - naio technologies," *Naio Technologies*, 2017.

[24] W. McAllister, D. Osipychev, G. Chowdhary, and A. Davis, "Multi-agent planning for coordinated robotic weed killing," in *Intelligent Robots and Systems (IROS), 2018 IEEE/RSJ International Conference on*, p. 1, IEEE, 2018.

[25] B. Chopard and M. Droz, *Cellular automata*. Springer, 1998.

[26] D. Mulugeta and D. E. Stoltenberg, "Seed bank characterization and emergence of a weed community in a moldboard plow system," *Weed Science*, pp. 54–60, 1997.

[27] D. Nordby, R. Hartzler, and K. Bradley, "Biology and management of waterhemp," *Glyphosate, Weeds, and Crop Sciences, Purdue University Extension, publication GWC-13.12*, 2007.

[28] B. J. Schutte and A. S. Davis, "Do common waterhemp (amaranthus rudis) seedling emergence patterns meet criteria for herbicide resistance simulation modeling?," *Weed Technology*, vol. 28, no. 2, pp. 408–417, 2014.

[29] R. Werle, L. D. Sandell, D. D. Buhler, R. G. Hartzler, and J. L. Lindquist, "Predicting emergence of 23 summer annual weed species," *Weed Science*, 2014.

[30] B. A. Sellers, R. J. Smeda, W. G. Johnson, J. A. Kendig, and M. R. Ellersieck, "Comparative growth of six amaranthus species in Missouri," *Weed Science*, vol. 51, no. 3, pp. 329–333, 2003.

[31] M. J. Horak and T. M. Loughin, "Growth analysis of four amaranthus species," *Weed Science*, vol. 48, no. 3, pp. 347–355, 2000.

[32] A. Axelrod, L. Carlone, G. Chowdhary, and S. Karaman, "Data-driven prediction of evar with confidence in time-varying datasets," in *2016 IEEE 55th Conference on Decision and Control (CDC)*, pp. 5833–5838, IEEE, 2016.

[33] A. Hobson, "A new theorem of information theory," *Journal of Statistical Physics*, vol. 1, no. 3, pp. 383–391, 1969.

[34] R. Houthooft, X. Chen, Y. Duan, J. Schulman, F. De Turck, and P. Abbeel, "Vime: Variational information maximizing exploration," in *Advances in Neural Information Processing Systems*, pp. 1109–1117, 2016.

[35] R. Allamaraju, H. Kingravi, A. Axelrod, G. Chowdhary, R. Grande, J. P. How, C. Crick, and W. Sheng, "Human aware uas path planning in urban environments using nonstationary mdps," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1161–1167, IEEE, 2014.

[36] D. Pathak, P. Agrawal, A. A. Efros, and T. Darrell, "Curiosity-driven exploration by self-supervised prediction," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 16–17, 2017.

[37] J. Achiam and S. Sastry, "Surprise-based intrinsic motivation for deep reinforcement learning," *arXiv preprint arXiv:1703.01732*, 2017.

[38] A. Ahmadi-Javid and M. Fallah-Tafti, "Portfolio optimization with entropic value-at-risk," *European Journal of Operational Research*, 2019.

[39] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas, *Actuarial theory for dependent risks: measures, orders and models*, ch. 1, p. 149. John Wiley & Sons, 2006. Citing stop-loss ordering.

[40] M. Denuit, J. Dhaene, M. Goovaerts, and R. Kaas, *Actuarial theory for dependent risks: measures, orders and models*, ch. 1, pp. 105–106. John Wiley & Sons, 2006. Citing stochastic ordering.

[41] A. Axelrod and G. Chowdhary, "A dynamic risk form of entropic value at risk," in *AIAA Scitech 2019 Forum*, p. 0392, 2019.