Learning, Optimization, and Planning Under Uncertainty for Wildlife Conservation

Lily Xu

Harvard University, lily_xu@g.harvard.edu

Abstract

Wildlife poaching fuels the multi-billion dollar illegal wildlife trade and pushes countless species to the brink of extinction. To aid rangers in preventing poaching in protected areas around the world, we have developed PAWS, the Protection Assistant for Wildlife Security. We present technical advances in multi-armed bandits and robust sequential decision-making using reinforcement learning, with research questions that emerged from on-the-ground challenges. We also discuss bridging the gap between research and practice, presenting results from field deployment in Cambodia and large-scale deployment through integration with SMART, the leading software system for protected area management used by over 1,000 wildlife parks worldwide.

1 Introduction

The illegal wildlife trade is a multi-billion dollar industry pushing countless species to the brink of extinction [Rosen and Smith, 2010]. Profit-driven poachers will enter protected areas and place snares to entrap animals. To prevent poaching, rangers conduct patrols around these protected areas to detect and confiscate snares. Unfortunately, poachers have the upper hand; wire snares are extremely cheap to make and easy to carry, so poachers can easily blanket the ground with snares. For example, the poachers in Srepok Wildlife Sanctuary are prolific: park managers estimate that there are four snares for every one deer. In contrast, Srepok has only 72 rangers to patrol its 3,750 km^2 — an area roughly the size of

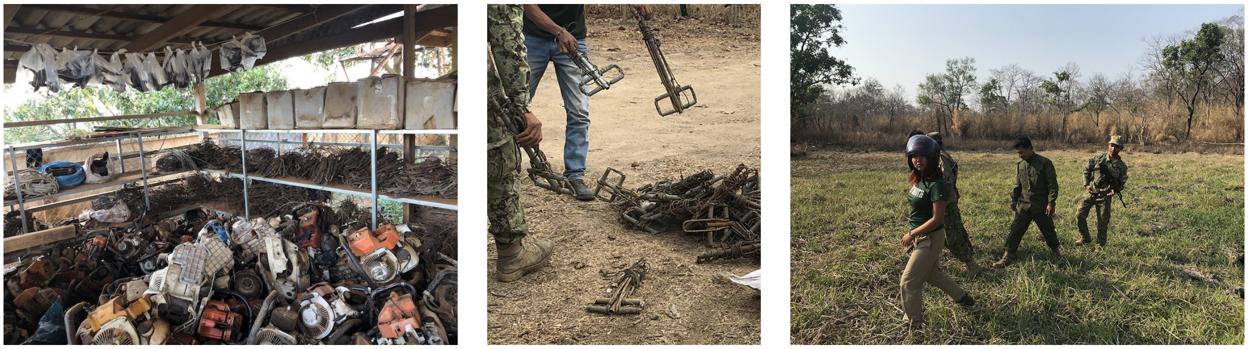


Figure 1: (left) A sampling of the vast number of wire snares and elephant traps confiscated by rangers in Srepok Wildlife Sanctuary, Cambodia. (center) Snares removed during a single patrol. (right) Accompanying rangers on a patrol during a field visit to Srepok in 2019.

Rhode Island. Given this resource imbalance, efficiently planning ranger patrols is critical. Our work aims to help identify the areas with greatest risk of poaching so they can remove as many snares as possible.

Viewed algorithmically, this problem is one of *optimizing* limited resources. The objective is to maximize the expected number of snares that rangers can detect, so that we can remove those snares and prevent wildlife from getting caught. While conducting this optimization, unfortunately, the data we have available are biased and incomplete, necessitating *online learning under uncertainty.* Furthermore, we expect that poachers might eventually learn rangers' behavior and adapt their strategy accordingly, making need for *sequential planning.* These real-world problems, and the computational solutions we have developed, demonstrate that *environmental domains such as wildlife conservation offer a range of fundamental new research challenges related to robust planning and data-driven optimization.*

Our project is called PAWS, the Protection Assistant for Wildlife Security, which has been created in close partnership with a number of conservation organizations, including the World Wide Fund for Nature (WWF) and Wildlife Conservation Society (WCS). We have worked directly with rangers on the frontline, spoken extensively with conservation managers and biodiversity experts, and traveled onsite to Cambodia (Figure 1) to meet with park rangers and experience a wildlife patrol first-hand.

Beyond advancing fundamental research in multi-armed bandits and robust planning, our project also bridges the gap between research and practice, having been tested on-the-ground in Uganda and Cambodia. We are now scaling the system worldwide through integration with SMART, the leading software system for protected area management used by over 1,000 wildlife parks around the world. Although SMART records significant amounts of historical data, its current capabilities are limited to managing data; the missing link is to leverage that data to inform patrol strategy. Our project builds that link, aiding park managers with patrol planning by identifying the most critical areas to patrol, which are either the areas with greatest poaching risk or where we have the greatest uncertainty.

We organize this paper as follows. Beginning with an overview of the domain in Section 2, we underline the urgency of poaching prevention and describe the three parks we work with. In Section 3 summarize the main technical contributions of this project. We then delve into more technical details, beginning with a machine learning approach to predict poaching hotspots in Section 4. In Section 5, we take a deep-dive into algorithmic details of an online learning approach designed for learning and planning in budgeted, combinatorial settings. Our algorithm, LIZARD, offers general theoretical work on multi-armed bandits where we prove that our algorithm improves upon existing regret bounds while also offering a practical approach to patrol planning in data-scarce settings. In Section 6, we consider the robust planning problem, presenting an algorithm to plan sequential patrols under environment uncertainty. Our algorithm, MIRROR, is the first reinforcement learning–based algorithm to calculate minimax regret–optimal policies. From there, we turn to deployment, first describing field tests in Section 7 then scaling up deployment to SMART in Section 8. We reflect on lessons learned from deployment in Section 9 before concluding in Section 10.

2 Wildlife Poaching Crisis

Illegal wildlife poaching is an international problem that threatens biodiversity, ecological balance, and ecotourism [Cooney *et al.*, 2017]. Countless species are being poached to nearextinction: the ivory, horn, and skin of exotic species such as elephants, rhinos, and tigers render them targets for illegal trade of luxury products and medicinal applications [Chase *et al.*, 2016; Spillane, 2015]; other animals like wild pigs and apes are hunted as bushmeat for protein [Warchol, 2004]. Even when their habitats become designated wildlife conservation areas, these animals continue to be at risk due to lack of sufficient resources to protect them from poachers. Timely detection and deterrence of illegal poaching activities in protected areas are critical to combating illegal poaching.

To combat poaching, park rangers conduct patrols through protected areas and use GPS trackers to record their observations. They confiscate animal traps, rescue live animals caught in snares, and monitor wildlife populations [Critchlow *et al.*, 2016]. Their GPS trackers are then synced to the SMART database system [SMART, 2013] to manage their many years of wildlife crime data. However, the data are biased due to the inability of rangers to detect all instances of poaching. There are additional data collection issues due to the nature of these patrols: rangers may have to address an emergency in the field, such as hearing a poacher in the distance, and lose the opportunity to record snares or bullet cartridges they found.

In this paper, we highlight our work with rangers and conservation specialists at Murchison Falls National Park (MFNP) and Queen Elizabeth National Park (QENP) in Uganda, and Srepok Wildlife Sanctuary (SWS) in Cambodia. Combined, these protected areas span over 11,800 sq. km, an area larger than the island country of Jamaica. MFNP and QENP are critically important for ecotourism and conservation in Uganda (Fig. 2), and provide habitat to elephants, giraffes, hippos, and lions [Critchlow *et al.*, 2015]. SWS is the largest protected area in Southeast Asia and is home to elephants, leopards, and banteng. SWS once housed a native population of tigers, but they fell prey to poaching; the last tiger was observed in 2007. In the intervening decade, SWS has been identified as the most promising site in South-

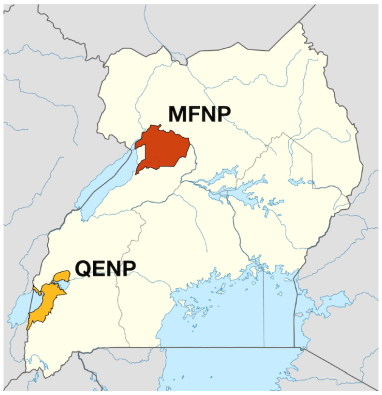


Figure 2: Location of MFNP and QENP in Uganda.

east Asia for tiger reintroduction [Gray *et al.*, 2017]. Effectively managing the landscape by reducing poaching will be critical to successful tiger reintroduction in Srepok.

3 Overview of Project and Contributions

To help combat poaching, we have developed the Protection Assistant for Wildlife Security (PAWS) as a data-driven approach to identify areas at high risk of poaching throughout protected areas and compute optimal patrol routes. The techniques we have built to help plan patrols for wildlife conservation focus on the central theme of data-driven approaches to optimizing scarce resources in the face of uncertainty.

This PAWS project is a collaboration between computer scientists and conservation practitioners at the World Wide Fund for Nature (WWF), Wildlife Conservation Society (WCS), and Uganda Wildlife Authority (UWA). Joint discussions took the form of countless video calls — monthly leading up to and during field tests then once a week over the course of 6 months during SMART deployment — along with reciprocal site visits. The computer science researchers spent a week in Cambodia to visit Srepok, meet with park managers, and accompany park rangers on a motorbike patrol to better understand on-the-ground constraints. In turn, the conservation practitioners visited Harvard in fall 2019 for a two-day workshop to discuss ongoing challenges in protected area management and ideate potential computational approaches.

Here, we briefly summarize the primary technical thrusts of this project.

Machine learning and risk-averse planning We first focus on data-rich parks that have plentiful historical data that park managers can leverage to understand poacher behavior and plan patrols. The challenge is to make sense of this trove of data to determine how we can optimally allocate our scarce teams of resources. We first leverage this data to build accurate predictive models, focusing on quantifying uncertainty, then leverage that uncertainty to plan risk-averse patrols for rangers [Xu *et al.*, 2020]. Specifically, we leverage uncertainty information from Gaussian processes in an ensemble method designed for class imbalance. We then incorporate the uncertainty of each prediction into a mixed integer linear program (MILP) to determine the optimal patrol strategy, using a scaling parameter to set our threshold for risk. The flexibility of the MILP enables us to accommodate path constraints and other domain-generated constraints, such as starting and ending each patrol by a patrol post or crossing by essential regions for ecological monitoring. We show that planning risk-aware patrols enables us to increase detection of snares by an average of 30%.

Multi-armed bandits for online learning Unfortunately, the majority of parks do not have abundant and accurate historical data that they can leverage. Recognizing that existing computational techniques make unrealistic assumptions on availability of data (e.g., years of patrol data) or time (e.g., infinite time horizons), we focus on patrol planning for these data-scarce settings. We seek to plan *dual-mandate patrols* to simultaneously detect illegal activities and collect valuable data to improve our predictive model and achieve higher long-term reward [Xu *et al.*, 2021a].

We use a multi-armed bandit formulation, where each action represents a patrol strategy, to balance exploration of infrequently visited regions and exploitation of known hotspots. However, traditional bandit approaches compromise short-term performance for long-term optimality, resulting in animals poached and forests destroyed. We develop a novel bandit algorithm, LIZARD, which speeds up performance by leveraging smoothness in the reward function and decomposability of actions. Combining these insights reveals a synergy between Lipschitz-continuity and decomposition as each aids the convergence of the other. *With this approach, we transcend the proven lower regret bound of Lipschitz bandits and generalize combinatorial bandits to continuous spaces.* On top of achieving theoretical no-regret, we also demonstrate that our LIZARD algorithm achieves better short-term performance empirically, increasing the usefulness of this approach in practice — particularly in high-stakes settings where we cannot compromise short-term reward.

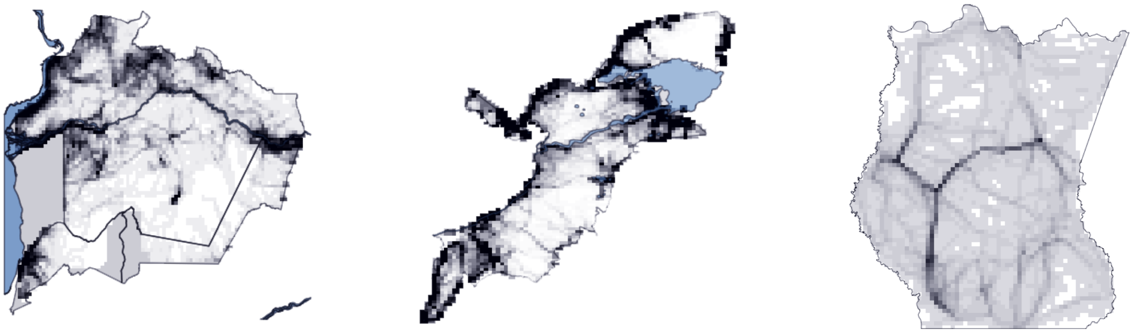
Robust sequential decision-making In parks with extensive history of poaching, we expect the poachers to be sophisticated, with the ability to respond to ranger patrols over time. Conservation biologists understand this behavior to be primarily one of deterrence, where increased ranger patrols deter poachers from returning to one region [Moore *et al.*, 2018]. This deterrence effect of patrols on adversaries' future behavior makes patrol planning a sequential decision-making problem. However, sequential planning techniques such as reinforcement learning (RL) assume an accurate simulator of the environment to enable this planning, but realistically we cannot expect our model to be perfect, given challenges of on-the-ground patrols. Thus, our goal is to plan robust patrols under environment uncertainty.

We focus on robust sequential patrol planning following the minimax regret criterion, formulating the problem as a game between the ranger and nature who controls the parameter values of the poacher behavior [Xu *et al.*, 2021b]. Our solution builds upon the double oracle approach [McMahan *et al.*, 2003], using two reinforcement learning–based oracles and solving a restricted, zero-sum game considering limited defender strategies and parameter values. We propose MIRROR, a framework to calculate minimax regret–optimal policies using RL for the first time. We prove that MIRROR converges to an ε–optimal strategy in a finite number of iterations, overcoming the difficulty of continuous state and action spaces, and empirically evaluate our algorithm on real poaching data. MIRROR improves existing techniques in robust policy planning by enabling the use of minimax regret instead of the standard maximin reward criterion, which tends to be overly conservative.

4 Predicting Poaching Hotspots

Many protected areas have years of historical patrols, which are often recorded on SMART without effective approaches to best leverage this data for future patrol plans. We begin with a supervised learning approach to predict poaching hotspots.

Learning the poachers' behavior is a challenging machine learning problem since the



(a) Murchison Falls National Park, Uganda

(b) Queen Elizabeth National Park, Uganda

(c) Srepok Wildlife Sanctuary, Cambodia

Figure 3: Three protected areas we work with in this paper. Visualized is the relative historical patrol effort for each protected area, calculated as kilometer patrolled per 1×1 km cell. Note that patrol effort is unevenly distributed around the park and many areas have never been patrolled (in white), making clear the need to proactively add data.

wildlife crime datasets are typically extremely imbalanced, with up to 99.6% negative labels; negative labels indicating absence of illegal activity are not reliable due to the difficulty of detecting well-hidden poaching signs in the forest; historical poaching observations are not collected thoroughly and uniformly, resulting in biased datasets; and poaching patterns and landscape features vary from one protected area to another, so a universal predictive model cannot be recommended. A unifying theme is data imbalance and uncertainty.

We note that the ultimate goal is to prescribe effective patrol routes for rangers to maximize the number of snares removed—corresponding to animal lives saved. Thus, rangers are incentivized to conduct patrols with higher certainty of detecting snares. This domain insight inspired us to optimize the patrol plans by measuring the predictive uncertainty in our model. To do so, we use Gaussian processes to quantify uncertainty in predictions of poaching risk and exploit these uncertainty metrics in our optimization problem to increase the robustness of our prescribed patrols.

4.1 Building a Reliable Predictive Model

To understand poacher behavior, we leverage observations from historical ranger patrols. Patrol observations come from SMART conservation software, which records the GPS loca-

Table 1: About the data available from each park

	MFNP	**QENP**	**SWS**
Number of features	22	19	21
Number of 1×1 km cells	4,613	2,522	3,750
Number of points (6 years)	18,254	19,864	43,269
Percent positive labels	14.3%	4.7%	0.36%
Avg. patrol effort (km/cell)	1.75	2.08	3.96

tion of each observation along with date and time, patrol leader, and method of transport. Rangers enter their observations: animals or humans spotted; signs of illegal activity such as campsites or cut trees; and signs of poaching activity such as firearms, bullet cartridges, snares, or slain animals. We categorize these observations into poaching and non-poaching. Additionally, we rebuild historical patrol effort from these observations by using sequential waypoints to calculate patrol trajectories.

We then incorporate geospatial data about the park. Data specialists at WWF, WCS, and UWA provide us with relevant GIS shapefiles and GeoTIFF files. The features differ between parks, but include terrain features such as rivers, elevation maps, and forest cover; landscape features such as roads, park boundary, local villages, and patrol posts; and ecological features such as animal density and net primary productivity. We use these static features to build data points in our predictive model, either as direct values (such as slope or animal density) or as distance values (such as distance to nearest river).

We build the datasets based on the historical patrol observations. The records are discretized into a set of T time steps and N locations. Each feature vector $\mathbf{x}_{t,n}$ contains multiple time-invariant geospatial features associated with each location (described above) and one time-variant covariate: $c_{t-1,n}$, the amount of patrol coverage in cell n during the previous time step $t-1$, which models the potential deterrence effect of past patrols.

The labels in the predictive classifier are a binary indicator of whether illegal poaching activity was observed in a cell at a given time step. We assign a positive label $y = 1$ to the cell if rangers observed poaching-related activity during that time period and negative label

$y = 0$ if they did not. The data characteristics for each park are described in Table 1

4.2 Predictions with Uncertainty

To account for class imbalance and unreliable negative labels, we use iWare-E, an ensemble method designed for imperfect observations [Gholami *et al.*, 2018]. We enhance iWare-E to explicitly reason about uncertainty of the predictions using Gaussian process (GP) classifiers as the weak learners. GPs are given by the function: $f(\mathbf{x}_i) \sim \mathcal{GP}\left(\mu(\mathbf{X}), \Sigma(\mathbf{X})\right)$, with mean $\mu(\mathbf{X})$ and covariance matrix $\Sigma(\mathbf{X})$. By formally defining the covariance functions, we can use GPs to compute a variance value for each prediction based on confidence from the training data.

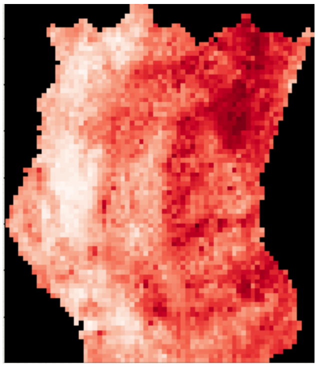


Figure 4: Prediction of poaching risk throughout Srepok.

See Figure 4 for an example of the resulting predictions. In evaluations on the historical park data, we see that our approach consistently improves AUC by 0.100 on average across parks. More importantly, we later describe our on-the-ground evaluations in the form of field tests in Uganda and Cambodia. Additionally, based on this poaching risk and the uncertainty estimates, we can compute an optimal patrol plan by solving a mixed integer linear program to plan patrol routes, incorporating path constraints or other restrictions by parks, such as constraining the patrol to begin and end at a patrol post.

5 Patrol Planning in Data-Scarce Parks

The above results on building a reliable predictive model are promising, but rely on having plentiful historical data from which to train a machine learning model. However, many protected areas lack adequate and unbiased past patrol data, disabling us from learning a reasonable adversary model in the first place [Moreto and Lemieux, 2015]. As one of many examples, Bajo Madidi in the Bolivian Amazon was newly designated as a national park in 2019 [Berton, 2020]. The park is plagued with illegal logging, and patrollers do not have historical data from which to make predictions. Rangers do not want to spend patrol effort solely on information gathering; they must simultaneously maximize detection of attacks. As our work on poaching prediction gets deployed on an ever-larger scale in hundreds of protected areas around the world, addressing this information-gathering challenge is crucial.

Motivated by these practical needs, we focus on conducting *dual-mandate patrols*, with the goal to simultaneously detect illegal activities and collect valuable data to improve our predictive model, achieving higher long-term reward. The key challenge is the exploration–exploitation tradeoff: whether to follow the best patrol strategy indicated by historical data or conduct new patrols to get a better understanding of the attackers. Unfortunately, existing bandit approaches require unrealistically long time horizons to achieve good performance. In the real world, these initial losses are less tolerable and can result in wildlife loss and stakeholders abandoning such patrol-assistance systems. In response, we provide an algorithm with infinite horizon guarantees and also empirically show strong performance in the short term. As we are designing this system for future deployment, it is critical to account for these practical constraints.

We address real-world characteristics of green security domains to design dual-mandate patrols, prioritizing strong performance in the short term as well as long term. Concretely, we introduce LIZARD, a bandit algorithm that accounts for decomposability, Lipschitz-continuity, monotonicity, and historical data present in the poaching prevention domain to produce a more effective online learning algorithm that achieves faster empirical performance and stronger theoretical guarantees.

5.1 Problem Formulation

The protected area for which we must plan patrols is discretized into N targets, each associated with a feature vector $\vec{y}_i \in \mathbb{R}^K$ which is static across the time horizon T. The K features include geospatial characteristics such as distance to river, forest cover, animal density, and slope. In each round, the rangers determine an *effort vector* $\vec{\beta} = (\beta_1, \dots, \beta_N)$ which specifies the amount of effort to spend on each target. The park has a budget B for total effort, i.e., $\sum_i \beta_i \leq B$. In practice, β_i may represent the number of hours spent on foot patrolling in target i. The planned patrols have to be conveyed clearly to the *human* patrollers on the ground to then be executed [Plumptre *et al.*, 2014]. Thus, an arbitrary value of β_i may be impractical. For example, we may ask the patrollers to patrol in an area for 30 minutes, but not 21.3634 minutes. Therefore, we enforce discretized patrol effort levels, requiring $\beta_i \in \Psi = \{\psi_1, \dots, \psi_J\}$ for J levels of effort.

Poachers will place snares in some targets. The reward of a patrol corresponds to the total number of targets where attacks were detected. Let the expected reward of a patrol vector $\vec{\beta}$ be $\mu(\vec{\beta})$. Our objective is to specify an effort vector $\vec{\beta}^{(t)}$ for each timestep t in an online fashion to minimize regret with respect to the optimal effort vector $\vec{\beta}^*$ against a stochastic adversary, where regret is defined as $T\mu(\vec{\beta}^*) - \sum_{tl=1}^{T} \mu(\vec{\beta}^{(t)})$.

In practice, the likelihood of a ranger detecting an attack is dependent on the amount of patrol effort exerted. Thus, spending more time means the human patrollers can check the whole region more thoroughly and are more likely to detect snares. We represent the ranger's expected reward at target i as a function $\mu_i(\beta_i) \in [0, 1]$. We define random variables $X_i^{(t)}$ as the observed reward (attack or no attack) from target i at time t. Then $X_i^{(t)}$ follows a Bernoulli distribution with mean $\mu_i(\beta_i^{(t)})$ with effort $\beta_i^{(t)}$.

5.2 Domain Characteristics

We leverage the following characteristics of poaching domains to direct our approach.

Decomposability The overall expected reward for the ranger is decomposable across targets and additive. For executing a patrol with effort $\vec{\beta}$ across all targets, the expected composite reward is a function $\mu(\vec{\beta}) = \sum_{i=1}^{N} \mu_i(\beta_i)$.

Lipschitz-continuity The expected reward at target i is $\mu_i(\beta_i)$, which is dependent on effort β_i. Furthermore, the expected reward depends on the features $\vec{y}_i$ of that target, that is, $\mu_i(\beta_i) = \widetilde{\mu}(\vec{y}_i, \beta_i)$ for all i. As we showed previously, machine learning models (which rely on assumptions of Lipschitz continuity) to predict poaching patterns perform well. Thus, we assume that the reward function $\widetilde{\mu}(\cdot, \cdot)$ is Lipschitz-continuous in feature space as well as across effort levels. That is, two distinct targets in the protected area with identical features will have identical reward functions, and two targets a and b with features $\vec{y}_a$, $\vec{y}_b$ and effort β_a, β_b have rewards that differ by no more than

$$|\widetilde{\mu}(\vec{y}_a, \beta_a) - \widetilde{\mu}(\vec{y}_b, \beta_b)| \leq L \cdot \mathcal{D}((\vec{y}_a, \beta_a), (\vec{y}_b, \beta_b)) \tag{1}$$

for some Lipschitz constant L and distance function $\mathcal{D}$, such as Euclidean distance. Hence, the composite reward $\mu(\vec{\beta})$ is also Lipschitz-continuous.

Monotonicity The more effort spent on a target, the higher the expected reward as our likelihood of finding a snare increases. That is, we assume $\mu(\beta_i)$ is monotonically non-decreasing in β_i. Additionally, we assume that zero effort corresponds with zero reward ($\mu_i(0) = 0$), as rangers cannot prevent attacks on targets they do not visit.

Historical data Finally, many conservation areas have data from past patrols, which we use to warm start the online learning algorithm.

5.3 LIZARD Online Learning Algorithm

Standard bandit algorithms suffer from the curse of dimensionality: the set of arms would be Ψ^N, which has size J^N. Thus, we cast the problem as a combinatorial bandit [Chen *et al.*, 2016]. At each iteration, we choose a patrol strategy $\vec{\beta}$ that satisfies the budget constraint and observe the patrol outcome of each target i under the chosen effort β_i. An *arm* is one effort level β_i on a specific target i; a *super arm* is $\vec{\beta}$, a collection of N arms. By tracking decomposed rewards, we only need to track observations from NJ arms. We now maintain exponentially fewer samples, but the number of arms is still prohibitively

Algorithm 1: LIZARD

```
**Inputs:** Number of targets $N$, time horizon $T$, budget $B$, discretization levels $\Psi$,
  target features $\vec{y}_i$
$n(i, \psi_j) = 0$, $\texttt{reward}(i, \psi_j) = 0 \quad \forall i \in [N], j \in [J]$
**for** $t = 1, 2, \ldots, T$ **do**
    Compute $\text{UCB}_t$ using Eq. 4
    Solve $\mathcal{P}(\text{UCB}_t, B, N, T)$ to select super arm $\vec{\beta}$
    Observe rewards $X_1^{(t)}, X_2^{(t)}, \ldots, X_n^{(t)}$
    **for** $i = 1, 2, \ldots, N$ **do**
        $\texttt{reward}(i, \beta_i) = \texttt{reward}(i, \beta_i) + X_i^{(t)}$
        $n(i, \beta_i) = n(i, \beta_i) + 1$
```

large. WTo address this challenge, we leverage feature similarity between arms to speed up learning, demonstrating the synergy between decomposability and Lipschitz-continuity. We now show how to transfer knowledge between arms with similar effort levels and features.

5.3.1 Upper Confidence Bounds with Similarity

We take an upper confidence bound (UCB) approach where the rewards are tracked separately for different arms. We show that we can incorporate Lipschitz-continuity of the reward functions into the UCB of each arm to achieve tighter confidence bounds.

Let $\bar{\mu}_t(i, j) = \texttt{reward}_t(i, \psi_j)/n_t(i, \psi_j)$ be the average reward of target i at effort ψ_j given cumulative empirical reward $\texttt{reward}_t(i, \psi_j)$ over $n_t(i, \psi_j)$ arm pulls. The *confidence radius* is defined as

$$r_t(i, j) = \sqrt{\frac{3 \log(t)}{2 n_t(i, \psi_j)}} \ . \tag{2}$$

We distinguish between UCB and a term we call $\textsc{self}$UCB. The $\textsc{self}$UCB of an arm (i, j) representing target i with effort level j is the UCB of an arm based only on its own observations, given by

$$\textsc{self}\text{UCB}_t(i, j) = \bar{\mu}_t(i, j) + r_t(i, j) \ . \tag{3}$$

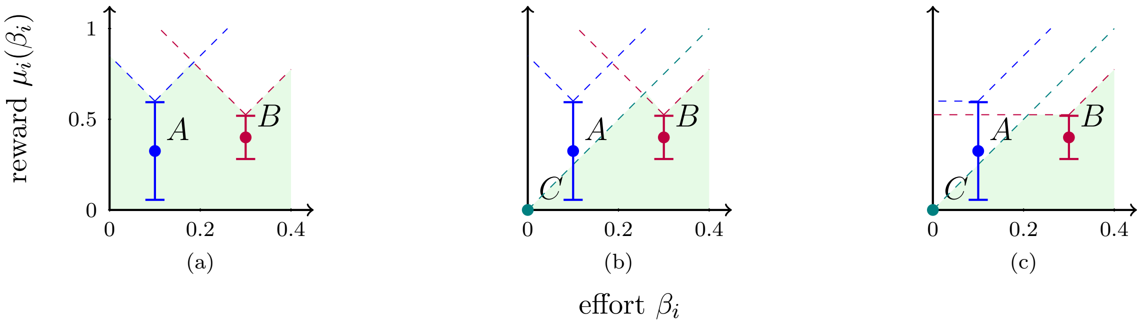


Figure 5: The Lipschitz assumption enables us to prune confidence bounds. We show the impact of each SELFUCBs on the UCBs of other arms in effort space of target i. The solid brackets represent the SELFUCBs. The dashed lines represent the bounds imposed by each arm on the rest of the space. The shaded green region covers the potential value of the reward function at different levels of effort. We visualize the additive effect of (a) Lipschitz-continuity, (b) zero effort yields zero reward, and (c) monotonicity. Note that these plots demonstrate UCBs for one target and that Lipschitz continuity also applies *across* targets based on feature similarity.

This definition of SELFUCB corresponds with the standard interpretation of confidence bounds from UCB1 [Auer *et al.*, 2002]. The UCB of an arm is then the minimum of the bounds of all SELFUCBs as applied to the arm, determined by adding the distance between arm (i, j) and all other arms (u, v) to the SELFUCB:

$$\mathrm{UCB}_t(i,j) = \min_{\substack{u\in[N]\\ v\in[J]}} \{\textsc{self}\mathrm{UCB}_t(u,v) + L \cdot \mathit{dist}\} \tag{4}$$

$$\mathit{dist} = \max\{0, \psi_v - \psi_j\} + \mathcal{D}(\vec{y}_i, \vec{y}_u)$$

which exploits Lipschitz continuity between the arms. See Fig. 5 for a visualization. The distance between two arms depends on the similarity of their features and effort. The first term of *dist* considers similarity of effort level (Fig. 5a); the second considers feature similarity between targets according to distance function $\mathcal{D}$. We define $\mathrm{UCB}_t(i, 0) = 0$ for all $i \in [N]$ due to the assumption that zero effort yields zero reward (Fig. 5b). To address the monotonically non-decreasing reward across effort space, we constrain the first term of *dist* to be nonnegative (Fig. 5c).

5.3.2 Super Arm Selection

With the computed UCBs, the selection of super arms (patrol strategies) is a knapsack optimization problem. We aim to maximize the value of our sack (sum of the UCBs) subject to a budget constraint (total effort), solved with the following integer linear program $\mathcal{P}$.

$$\begin{aligned} \max_{z} \quad & \sum_{i \in [N]} \sum_{j \in [J]} z_{i,j} \cdot \mathrm{UCB}_t(i,j) && (\mathcal{P}) \\ \text{s.t.} \quad & z_{i,j} \in \{0,1\} && \forall i \in [N], j \in [J] \\ & \sum_{j \in [J]} z_{i,j} = 1 && \forall i \in [N] \\ & \sum_{i \in [N]} \sum\nolimits_{j \in [J]} z_{i,j} \psi_j \leq B \end{aligned}$$

There is one auxiliary variable $z_{i,j}$, constrained to be binary, for each level of patrol effort for each target. Setting $z_{i,j} = 1$ means we exert ψ_j effort on target i. The constraints set β_i by requiring that we pull one arm per target (which may be the zero effort arm $\beta_i = 0$) and mandate that we stay within budget. This integer program has NJ variables and $N+1$ constraints. Pseudocode for the LIZARD algorithm is given in Algorithm 1.

5.4 Regret Analysis

We provide a regret bound for Algorithm 1 with fixed discretization (Sec. 5.4.1), which is useful in practice but cannot achieve theoretical no-regret due to the discretization factor. Thus, we then offer Algorithm 2 with adaptive discretization to achieve no-regret (Sec. 5.4.2), showing that there is no barrier to achieving no regret in practice other than the need for fixed discretization in operationalizing our algorithm in the field. Our regret bound improves upon that of the zooming algorithm of Kleinberg *et al.* [2019] for all problem sizes $N > 1$. In fact, the regret bound for the zooming algorithm is a provable lower bound; we are able to improve this lower bound through decomposition (Sec. 5.4.3). Furthermore, we extend the line of research on combinatorial bandits, generalizing the CUCB algorithm from Chen *et al.* [2016] to continuous spaces.

5.4.1 Fixed Discretization

Theorem 1. Given the minimum discretization gap Δ, the regret bound of Algorithm 1 with SELFUCB is

$$\mathrm{Reg}_{\Delta}(T) \leq O\left(NL\Delta T + \sqrt{N^3\Delta^{-1}T\log T} + N^2L\Delta^{-1}\right) \ . \tag{5}$$

Proof sketch. The regret in Theorem 1 comes from (i) discretization regret in the first term of Eq. 5 and (ii) suboptimal arm selections in the last two terms. The discretization regret is due to inaccurate approximation caused by discretization, where the error can be bounded by rounding the optimal arm selection and fractional effort levels to their closest discretized levels. The suboptimal arm selections are due to insufficient samples across all discretized subarms and have sublinear regret in terms of T. □

5.4.2 Adaptive Discretization

With Theorem 1 we observe that the barrier to achieving no-regret is the discretization error which is linear in all terms, which brings us to adaptive discretization. Adaptive discretization is less practical on the ground, but would be useful in other bandit settings outside ranger patrols where we could more precisely spend our budget, such as facility location. As shown in Algorithm 2, we begin with a coarse patrol strategy, beginning with binary decisions on whether or not to visit each target, then gradually progress to a finer discretization.

Theorem 2. The regret bound of Algorithm 2 with SELFUCB is given by

$$\mathrm{Reg}(T) \leq O\left(L^{\frac{4}{3}}NT^{\frac{2}{3}}(\log T)^{\frac{1}{3}}\right) \ . \tag{6}$$

```
Algorithm 2: Adaptively Discretized LIZARD
Inputs: Number of targets $N$, time horizon $T$, budget $B$, target features $\vec{y}_i$,
  Lipschitz constant $L$
Discretization levels $\Psi = \{0, 1\}$ , gap $\Delta = 1$
$T_k = \frac{N}{L^2 2^{3k}} \log \frac{N}{L^2 2^{3k}} \; \forall k \in \mathbb{N} \cup \{0\}$
$n(i, \psi_j) = 0$, reward$(i, \psi_j) = 0 \; \forall i \in [N], j \in [J]$
for $t = 1, 2, \ldots, T$ do
    if $t > \sum_{j=0}^{k-1} T_j$ then
        Set $\Delta = 2^{-k}$ and $\Psi = \{0, \Delta, ..., 1\}$
    Compute $\mathrm{UCB}_t$ using Eq. 4
    Solve $\mathcal{P}(\mathrm{UCB}_t, B, N, T)$ to select super arm $\vec{\beta}$
    Observe rewards $X_1^{(t)}, X_2^{(t)}, \ldots, X_n^{(t)}$
    for $i = 1, 2, \ldots, N$ do
        reward$(i, \beta_i)$ = reward$(i, \beta_i) + X_i^{(t)}$
        $n(i, \beta_i) = n(i, \beta_i) + 1$
```

Proof sketch. To alleviate the discretization error, we adaptively reduce the discretization gap. We run each discretization gap Δ for $T_\Delta = \frac{N}{L^2 \Delta^3} \log \frac{N}{L^2 \Delta^3}$ time steps to make the discretization error and the selection error of the same order. We then start over with a finer discretization $\Delta/2$ to make the discretization error smaller. After summing the regret from all different phrases of discretization, we achieve sublinear regret as shown in Eq. 6. □

Under reasonable problem settings, T dominates all other variables, so the regret in Theorem 2 is effectively of order $O(T^{\frac{2}{3}} (\log T)^{\frac{1}{3}})$. Our regret bound matches the bound of the zooming algorithm with covering dimension $N = 1$. Our setting is instead N-dimensional, falling into a space with covering dimension $d = N$. The regret bound for a metric space with covering dimension d is $O(T^{\frac{d+1}{d+2}} (\log T)^{\frac{1}{d+2}})$, which approaches $\tilde{O}(T)$ as d approaches infinity [Kleinberg *et al.*, 2008]. Thus, our regret bound improves upon that of the zooming algorithm for any $N > 1$. Theorem 2 signifies that LIZARD can successfully decouple the N-dimensional metric space into individual sub-dimensions while maintaining the smaller regret order, showcasing the power of decomposibility.

5.4.3 Tightening Confidence Bounds

We have so far offered regret bounds to account for decomposability and Lipschitz-continuity across effort space. We now guarantee that the regret bounds continue to hold with Lipschitz-continuity in feature space, monotonicity, and historical information. We first look at how prior knowledge affects the regret bound in the combinatorial bandit setting:

Theorem 3. Consider a combinatorial bandit problem. If the bounded smoothness function given is $f(x) = \gamma x^\omega$ for some $\gamma > 0, \omega \in (0, 1]$ and the Lipschitz upper confidence bound is applied to all m base arms, the cumulative regret at time T is bounded by

$$\text{Reg}(T) \leq \frac{2\gamma}{2-\omega}(6m\log T)^{\frac{\omega}{2}} \cdot T^{1-\frac{\omega}{2}} + \left(\frac{\pi^2}{3}+1\right) mR_{\max}$$

where $R_{\max}$ is the maximum regret achievable.

Theorem 3 matches the regret of the CUCB algorithm [Chen *et al.*, 2016], generalizing combinatorial bandits to continuous spaces. Theorem 3 also allows us to generalize Theorems 1 and 2 to our setting with a tighter UCB from Lipschitz-continuity, which yields Theorem 4:

Theorem 4. The regret bound of Algorithm 1 with UCB is

$$\text{Reg}_\Delta(T) \leq O\left(NL\Delta T + \sqrt{N^3\Delta^{-1}T\log T} + N^2L\Delta^{-1}\right) \tag{7}$$

and the regret bound of Algorithm 2 with UCB is

$$\text{Reg}(T) \leq O\left(L^{\frac{4}{3}}NT^{\frac{2}{3}}(\log T)^{\frac{1}{3}}\right) \ . \tag{8}$$

Finally, when historical data is used, we can treat all the historical arm pulls as previous arm pulls with regret bounded by the maximum regret $R_{\max}$.This yields a regret bound with a time-independent constant and is thus still sublinear, achieving no-regret. Taken together, these capture all the properties of the LIZARD algorithm, so we can state:

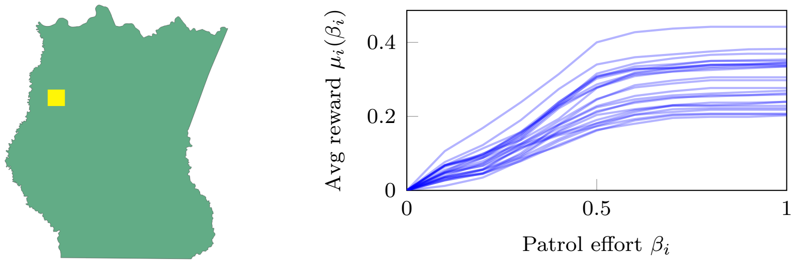


Figure 6: Map of Srepok with a 5 × 5 km region highlighted and the real-world reward functions of the corresponding 25 targets.

Corollary 5. The regret bounds of Algorithm 1 and Algorithm 2 still hold with the inclusion of decomposability, Lipschitz-continuity, monotonicity, and historical data.

Theorem 4 highlights the interplay between Lipschitz-continuity and decomposition. The zooming algorithm achieves the provable lower bound on regret $\tilde{O}(T^{\frac{N+1}{N+2}})$ [Kleinberg, 2004]. *The regret that we achieve improves upon that lower bound for all $N > 1$, which is only possible due to the addition of decomposition.*

5.5 Empirical Evaluation

Conducting experiments using poaching data from Srepok Wildlife Sanctuary, we validate that the addition of decomposition and Lipschitz-continuity not only improves our theoretical guarantees but also leads to stronger empirical performance. We show that LIZARD (Algorithm 1) learns effectively within practical time horizons.

We consider a patrol planning problem with $N = 25$ or 100 targets (each a 1 sq. km grid cell), representing the region reachable from a single patrol post (Figure 6), and time horizon $T = 500$ representing a year and a half of patrols. We use 50 timesteps of historical data, approximately two months of patrol, as we focus on achieving strong performance in parks with limited historical patrols. We compare to three baselines: *CUCB* [Chen *et al.*, 2016], *zooming* [Kleinberg *et al.*, 2019], and *MINION* [Gholami *et al.*, 2018]. Zooming is an online learning algorithm that ignores decomposability, whereas CUCB uses decomposition but ignores similarity between arms. We use *exploit history* as a naive baseline, which greedily

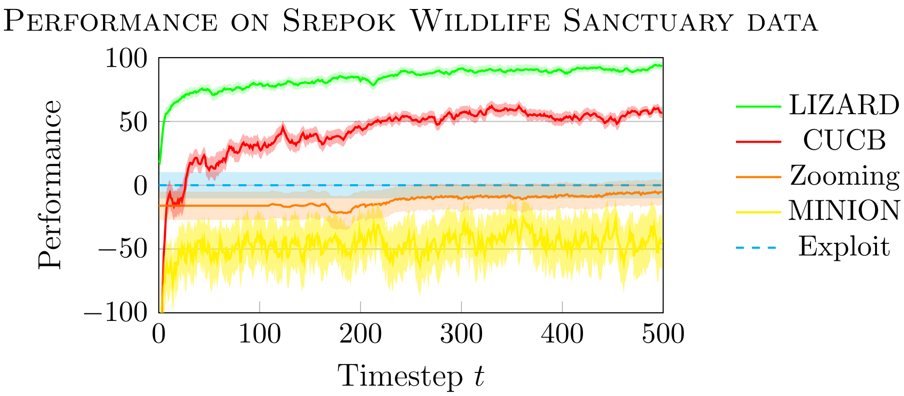


Figure 7: Performance, measured in terms of percentage of reward achieved between $\textsc{optimal} - \textsc{exploit}$, over time. Shaded region shows standard error. Setting shown is $N = 25$, $B = 1$. LIZARD (green) performs best.

exploits historical data with a static strategy. We compute the *optimal* strategy exactly by solving a mixed-integer program over the true piecewise-linear reward functions, subject to the budget constraint.

Fig. 7 shows performance on real-world data from Srepok, evaluated as the reward achieved at timestep t, where the reward of historical exploit is 0 and of optimal is 1. The performance of LIZARD significantly surpasses that of the baselines throughout, with LIZARD providing a particular advantage over CUCB in the early rounds.

6 Robust planning under uncertainty

In parks with extensive history of poaching, we expect the poachers to be sophisticated, with the ability to respond to ranger patrols over time. Conservation biologists understand this behavior to be primarily one of deterrence, where increased ranger patrols deter poachers from returning to one region [Moore *et al.*, 2018]. This deterrence effect of patrols on adversaries' future behavior makes patrol planning a sequential decision-making problem. However, sequential planning techniques such as reinforcement learning (RL) assume an accurate simulator of the environment to enable this planning, but we cannot expect our environmental model to be perfect, given challenges of on-the-ground patrols. Thus, our goal is to plan robust patrols under environment uncertainty.

We focus on robust sequential patrol planning following the minimax regret criterion, formulating the problem as a game between the ranger and nature who controls the parameter values of the poacher behavior [Xu *et al.*, 2021b]. Our solution builds upon the double oracle approach [McMahan *et al.*, 2003] uses two reinforcement learning–based oracles and solves a restricted, zero-sum game considering limited defender strategies and parameter values. We propose MIRROR, a framework to calculate minimax regret–optimal policies using RL for the first time. We prove that MIRROR converges to an ε–optimal strategy in a finite number of iterations, overcoming the difficulty of continuous state and action spaces, and empirically evaluate our algorithm on real poaching data. MIRROR improves existing techniques in robust policy planning by enabling the use of minimax regret instead of the standard maximin reward criterion, which tends to be overly conservative.

7 Deployment: Evaluating PAWS with Field Tests

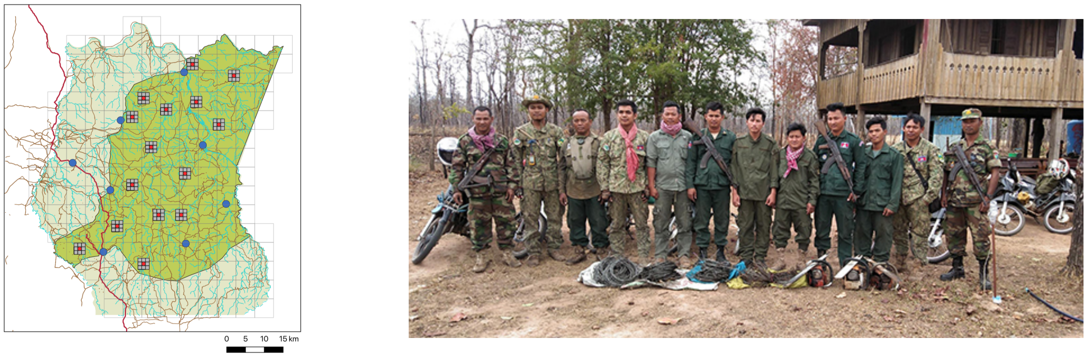


Figure 8: (left) Regions in SWS used for field tests in December 2018. High-, medium-, and low-risk regions are shown in red, yellow, and green, respectively. Blue circles are patrol posts; the rivers and roads are also displayed. (right) Rangers in Srepok Wildlife Sanctuary with snares they removed during our field tests in December 2018. Photo: WWF Cambodia.

We found our algorithms to perform well in experiments on historical park data, but we also want to consider how they fare on the ground to ensure that the results are useful to existing conservation efforts. To do so, we conducted a series of field tests in Uganda and Cambodia, where we demonstrate the ability of our algorithms to uncover previously

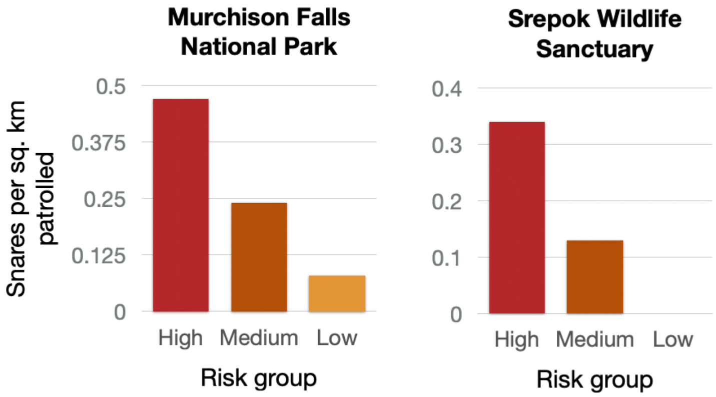


Figure 9: Field test results from deployment in Murchison Falls National Park in Uganda and Srepok Wildlife Sanctuary in Cambodia. The results demonstrate the clear ability of our algorithm to effectively discriminate between relative poaching risk throughout the park, as rangers found many more snares in high-risk regions than in medium- or low-risk ones.

unknown poaching hotspots. The strong performance during these tests has motivated a broader collaboration to deploy our predictive model into widely used conservation software.

7.1 Field Test Setup

Partnering with the Uganda Wildlife Authority and WWF Cambodia to test the predictions made by our machine learning model, we conducted field tests in Murchison Falls National Park and Srepok Wildlife Sanctuary beginning in 2018. Each month, we generated predictions using the predictive learning algorithm described above and classified our recommended areas into three risk groups (low, medium, and high). These areas were all infrequently patrolled in the past, to ensure we test the predictive power of our algorithms rather than relying on past patterns. To reduce bias in data collection, we did not reveal the risk groups to rangers before conducting the experiments.

SWS experiences high seasonality, where many rivers dry up during the dry season, inspiring us to train our model based only on data from dry months (November through April), using a two-month discretization to get three temporal points per year. Using data from January 2015 through April 2018, we made predictions on the November–December 2018 time period for the first set of field tests in December, and repeated for subsequent months. Using these predictions of poaching risk for 1 × 1 km cells, we averaged the risk predictions over the adjacent cells by convolving the risk map to produce 3 × 3 km blocks. We then discarded all blocks with historical patrol effort above the 50th percentile, to ensure we were assessing the ability of our model to make predictions in regions with limited data. From this set of valid blocks, we identified high-, medium-, and low-risk areas by considering blocks with risk predictions within the 80–100, 40–60, and 0–20 percentile. In selecting regions for these field tests we had to also accommodate constraints by the rangers, such as each site region had to be within 5 km from the nearest water source and reasonable accessible from a road. In total, we selected five 3 × 3 km blocks from each of the three risk categories, shown in Fig. 8.

To execute these tests, we gave the park rangers GPS coordinates of the center of each block and asked them to target those regions during their patrols. Again, we kept the study as a blind experiment by not revealing the risk classifications to rangers in advance. Beginning in December 2018, 72 park rangers in teams of eight conducted patrols throughout the park, focusing on our suggested areas.

7.2 Field Test Results

Figure 9 visualizes the key results of our field tests, comparing the number of incidences of poaching activity rangers observed in each region normalized by the number of cells patrolled. Our predictive model effectively evaluates the poaching threat for different regions across the park, with marked success at discriminating between high-risk and low-risk areas. *In Srepok, park rangers found absolutely no poaching activity in low-risk areas, despite exerting a comparable amount of effort in those regions.* This result suggests that revealing our risk predictions to them would grant them valuable insight into their patrol strategy so as to more effectively allocate their limited resources, as they can confidently spend less time patrolling these low-risk regions.

During a single month in Srepok Wildlife Sanctuary in Cambodia patrolling high-risk regions alone, *rangers detected and removed over 1,000 snares — a nearly fivefold increase of an average month.* These tests demonstrate that our algorithms perform well not just in the digital tests that we simulate but also on the ground.

8 Deployment: PAWS Scales to 1000+ Parks

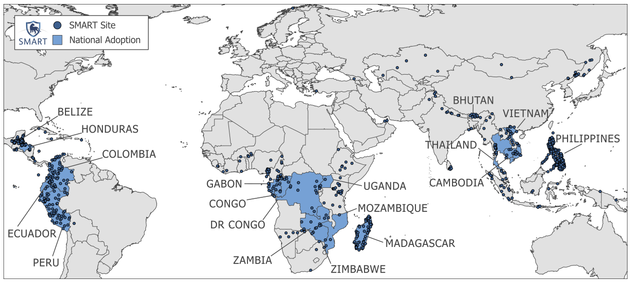


Figure 10: Our PAWS system has been integrated with SMART, used in over 1,000 protected areas worldwide.

Based on the success of our field tests and reciprocal enthusiasm from our collaborators, we are making our predictive algorithms broadly available by integrating our research advances into established software to scale most effectively. We have integrated our PAWS machine learning model into the SMART software, the premier software for wildlife conservation managed by a consortium of nine leading conservation organizations. This integration help democratize AI and enables us to contribute more broadly to the global effort to protect wildlife from poaching, bringing PAWS to protected areas across 60 countries that are using SMART (Figure 10). We are currently in a round of alpha tests to evaluate these predictions in a larger set of parks; conservation managers in Nigeria, Liberia, Zambia, Kenya, and Malaysia have all been running our system and using these predictions to guide their patrol planning.

9 Deployment: Lessons Learned

This research has been conducted in close collaboration with domain experts whose insights have shaped the trajectory of our project. We highlight here a few important lessons learned from this long-term partnership with conservation organizations WWF, WCS, and UWA.

Begin with simple computational approaches. We could not begin by directly implementing a bandits algorithm. Instead, we had to begin with the simplest machine learning strategy, supervised learning, to iterate through the data processing step and understand what would be feasible on the ground. These simple approaches are often easier to evaluate, too. This initial phase is also essential to build relationships and establish a shared language between researchers and practitioners, which for us took the form of weekly or monthly meetings at every stage of the process (project proposal, algorithm design, field testing, deployment).

Incremental deployment before designing ambitious projects. We started out deployments on the ground with data-rich parks and well-resourced parks who had GIS specialists, well-trained rangers, and ranger managers who had a strong understanding of technology and the SMART database. Only after piloting several months of field tests in these settings did we (and our collaborators) feel comfortable taking the leap to integration with SMART.

Integrate domain expertise into algorithm design. The insight that poaching activity is subject to seasonality also resulted from discussions with park rangers, which inspired us to break up the SWS data into rainy vs. dry season and generate those predictive models separately. Our predictive model identified higher poaching risk in the north during dry season and south during rainy season, which garnered immediate positive feedback from the rangers: this aligned with their experiences on the ground and understanding of the accessibility of the terrain.

Consider real-world constraints as research challenges, not limitations. Our initial idea was to plan information-gathering patrols, taking an active learning approach to gather data where the predictive model was most uncertain. However, conservation experts pointed out in our discussions that patrollers could not afford to spend time purely gathering data; they must prioritize preventing illegal poaching, logging, and fishing in the short-term. These priorities inspired dual-mandate patrols and guided our project, particularly our focus on minimizing short-term regret as well as long-term regret. We emphasize that project design and scoping must be made in tandem with domain experts [Bondi *et al.*, 2021].

Evaluate with self-contained experiments. Extended discussions with park rangers made us aware of how much patrols may change over time. First, changes in funding and support may drastically alter patrolling resources; the number of park rangers in SWS in 2018 is over double that of 2016. Second, park rangers must be sensitive to real-time changes in the illegal market. For example, illegal logging rampantly increased in a nearby park in March, so park rangers from SWS were redirected there to provide backup, reducing the sources inside the park. These examples emphasize the importance of conducting self-contained experiments when evaluating model performance.

Real-world deployment is necessary for effective technology transfer. Park rangers are motivated by positive conservation outcomes, not by improvements in AUC. We were thrilled that after two months in SWS, our partners at WWF were enthused by the results and eager to deploy in other parks around the world, which was a critical impetus for the PAWS integration with SMART.

Quality engineering is essential to large-scale deployment. This often cannot be done by academics. A critical component of our integration with SMART has been a partnership with Microsoft AI for Earth, who has provided engineering support along with cloud computing resources. The task of developing production-level software is beyond what academics have capacity (or training) for. Industry partnerships may be helpful here.

Limited data inspire research directions to close the gap. These continued partnerships uncover new research directions as well. One challenge for under-resourced parks is that PAWS requires geospatial data about the land, such as rivers, roads, land cover, and animal density. During initial testing of PAWS with SMART, some park managers were getting nonsensical predictions — but it turns out they were trying to make predictions with just a single feature: park boundary. Parks without GIS specialists may not have additional predictive features. To remedy this data deficiency, we leverage publicly available remote sensing data, which provide global imagery across time [Guo *et al.*, 2020]. In this case, our close partnership with conservation NGOs and active role in the deployment phase enabled us to identify a key gap in making this work scalable to resource-constrained parks.

10 Conclusion

Overall, our PAWS project demonstrates an end-to-end pipeline to develop effective algorithms for wildlife conservation by assisting rangers with patrol planning. We describe three main technical thrusts, each of increasing complexity (supervised learning, online learning, robust reinforcement learning), which are each grounded in urgent real-world challenges in conservation. We present results from field tests that demonstrate the effectiveness of PAWS on the ground, leading to large-scale deployment via integration with established conservation software SMART. In our lessons learned, we share insights from years of close collaborations with conservation practitioners.

Along the way, we demonstrate that real-world impact does not preclude technical novelty. Among our key contributions is LIZARD, an integrated algorithm for online learning in green security domains: on top of achieving theoretical no-regret, we also demonstrate improved short-term performance empirically, increasing the usefulness of this approach in practice—particularly in high-stakes environments where we cannot compromise short-term reward. These results validate our approach of treating real-world conditions not as constraints but rather as useful features that lead to faster convergence.

We hope that this project serves as an example that interdisciplinary collaborations, particularly between researchers and practitioners, offer a trove of inspiration for novel research questions. The process is often laborious requiring additional challenges with communication, implementation, and understanding existing problems in other fields, but the benefits and potential for impact are innumerable.

We wish to close by stating that our work is intended to serve as an assistive technology, helping rangers identify potentially snare-laden regions they otherwise might have missed, given that these parks can be up to thousands of square kilometers and rangers can only patrol a small number of regions at each timestep. We do not intend this work to replace the critical role that rangers play in conservation management; no AI/OR tool could substitute for the complex skills and domain insights that rangers and park managers have to plan and conduct patrols.

References

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine Learning*, 47(2-3):235–256, 2002.

Eduardo Franco Berton. Rare trees are disappearing as 'wood pirates' log Bolivian national parks. `https://news.mongabay.com/2020/01/rare-trees-are-disappearing-as-wood-pirates-log-bolivian-national-parks/`, 2020.

Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A. Killian. Envisioning communities: A participatory approach towards AI for social good. In *Proc. 4th AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES-21)*, 2021.

Michael J Chase, Scott Schlossberg, Curtice R Griffin, Philippe JC Bouché, Sintayehu W Djene, Paul W Elkan, Sam Ferreira, Falk Grossman, Edward Mtarima Kohi, Kelly Landen, et al. Continent-wide survey reveals massive decline in african savannah elephants. *PeerJ*, 4:e2354, 2016.

Wei Chen, Yajun Wang, Yang Yuan, and Qinshi Wang. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *Journal of Machine Learning Research*, 17(1):1746–1778, 2016.

Rosie Cooney, Dilys Roe, Holly Dublin, Jacob Phelps, David Wilkie, Aidan Keane, Henry Travers, Diane Skinner, Daniel WS Challender, James R Allan, et al. From poachers to protectors: engaging local communities in solutions to illegal wildlife trade. *Conservation Letters*, 10(3):367–374, 2017.

R Critchlow, AJ Plumptre, M Driciru, A Rwetsiba, EJ Stokes, C Tumwesigye, F Wanyama, and CM Beale. Spatiotemporal trends of illegal activities from ranger-collected data in a ugandan national park. *Conservation Biology*, 29(5):1458–1470, 2015.

Rob Critchlow, Andrew J Plumptre, Bazil Alidria, Mustapha Nsubuga, Margaret Driciru, Aggrey Rwetsiba, F Wanyama, and Colin M Beale. Improving law-enforcement effectiveness and efficiency in protected areas using ranger-collected monitoring data. *Conservation Letters*, 2016.

Shahrzad Gholami, Sara Mc Carthy, Bistra Dilkina, Andrew Plumptre, Milind Tambe, Margaret Driciru, Fred Wanyama, Aggrey Rwetsiba, Mustapha Nsubaga, Joshua Mabonga, Tom Okello, and Eric Enyel. Adversary models account for imperfect crime data: Forecasting and planning against real-world poachers. *AAMAS '18*, 2018.

Thomas Gray, Rachel Crouthers, K Ramesh, J Vattakaven, Jimmy Borah, Mks Pasha, Thona Lim, Phan Channa, R Singh, Barney Long, S Chapman, O Keo, and M Baltzer. A framework for assessing readiness for tiger panthera tigris reintroduction: a case study from eastern cambodia. *Biodiversity and Conservation*, 05 2017.

Rachel Guo, Lily Xu, Andrew Plumptre, Drew Cronin, Francis Okeke, and Milind Tambe. Enhancing poaching predictions for under-resourced wildlife conservation parks using remote sensing imagery. In *NeurIPS Workshop on Machine Learning for the Developing World*, 2020.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Multi-armed bandits in metric spaces. In *Proceedings of the 40th ACM Symposium on Theory of Computing (STOC 2008)*, pages 681–690. ACM, 2008.

Robert Kleinberg, Aleksandrs Slivkins, and Eli Upfal. Bandits and experts in metric spaces. *Journal of the ACM (JACM)*, 2019.

Robert D Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. In *Advances in Neural Information Processing Systems 17 (NeurIPS-04)*, pages 697–704, 2004.

H Brendan McMahan, Geoffrey J Gordon, and Avrim Blum. Planning in the presence of cost functions controlled by an adversary. pages 536–543, 2003.

Jennifer F Moore, Felix Mulindahabi, Michel K Masozera, James D Nichols, James E Hines, Ezechiel Turikunkiko, and Madan K Oli. Are ranger patrols effective in reducing poaching-related threats within protected areas? *Journal of Applied Ecology*, 55(1):99–107, 2018.

William D Moreto and Andrew M Lemieux. Poaching in uganda: Perspectives of law enforcement rangers. *Deviant Behavior*, 36(11):853–873, 2015.

Andrew J Plumptre, Richard A Fuller, Aggrey Rwetsiba, Fredrick Wanyama, Deo Kujirakwinja, Margaret Driciru, Grace Nangendo, James EM Watson, and Hugh P Possingham. Efficiently targeting resources to deter illegal activities in protected areas. *Journal of Applied Ecology*, 51(3):714–725, 2014.

Gail Emilia Rosen and Katherine F Smith. Summarizing the evidence on the international trade in illegal wildlife. *EcoHealth*, 7(1):24–32, 2010.

SMART. Spatial monitoring and reporting tool. `http://smartconservationtools.org/`, 2013.

James J. Spillane. Africa's elephant population: Permanently declining or sustainable? *The Eastern African Journal of Hospitality, Leisure & Tourism*, 3(1):1–19, 2015.

Greg L Warchol. The transnational illegal wildlife trade. *Criminal Justice Studies*, 17(1):57–73, 2004.

Lily Xu, Shahrzad Gholami, Sara Mc Carthy, Bistra Dilkina, Andrew Plumptre, Milind Tambe, Rohit Singh, Mustapha Nsubuga, Joshua Mabonga, Margaret Driciru, et al. Stay ahead of poachers: Illegal wildlife poaching prediction and patrol planning under uncertainty with field test evaluations. In *Proc. IEEE 36th International Conference on Data Engineering (ICDE-20)*, 2020.

Lily Xu, Elizabeth Bondi, Fei Fang, Andrew Perrault, Kai Wang, and Milind Tambe. Dual-mandate patrols: Multi-armed bandits for green security. In *Proc. 35th AAAI Conference on Artificial Intelligence (AAAI-21)*, 2021.

Lily Xu, Andrew Perrault, Fei Fang, Haipeng Chen, and Milind Tambe. Robust reinforcement learning under minimax regret for green security. In *Proc. 37th Conference on Uncertainty in Artifical Intelligence (UAI-21)*, 2021.