Conserving Biodiversity via Adjustable Robust Optimization

Yingxiao Ye
University of Southern California
Los Angeles, California, United States
yingxiao.ye@usc.edu

Christopher Doehring
University of Southern California
Los Angeles, California, United States
doehring@usc.edu

Angelos Georghiou
University of Cyprus
Nicosia, Cyprus
georghiou.angelos@ucy.ac.cy

Hugh Robinson
Panthera
New York, New York, United States
hrobinson@panthera.org

Phebe Vayanos
University of Southern California
Los Angeles, California, United States
phebe.vayanos@usc.edu

ABSTRACT

Human development is a threat to biodiversity and conservation organizations (COs) are purchasing land to protect areas for biodiversity preservation. COs have limited budgets and cannot purchase all the land necessary to perfectly preserve biodiversity, and human activities are uncertain, so exact developments are unpredictable. We propose a multistage, robust optimization problem with a data-driven hierarchical-structured uncertainty set which captures the endogenous nature of the binary (0-1) human land use uncertain parameters to help COs choose land parcels to purchase to minimize the worst-case human impact on biodiversity. In the proposed approach, the problem is formulated as a game between COs, which choose parcels to protect with limited budgets, and the human development, which will maximize the loss to the unprotected parcels. We leverage the cellular automata model to simulate the development based on climate data, land characteristics, and human land use data. We use the simulation to build data-driven uncertainty sets. We demonstrate that an equivalent formulation of the problem can be obtained that presents exogenous uncertainty only and where uncertain parameters only appear in the objective. We leverage this reformulation to propose a conservative K-adaptability reformulation of our problem that can be solved routinely by off-the-shelf solvers like Gurobi or CPLEX. The numerical results based on real data from (Jędrzejewski et al. 2018) show that the proposed method reduces conservation loss by 19.46% on average compared to standard approaches used in practice for biodiversity conservation.

KEYWORDS

Robust Optimization, Endogenous Uncertainty, Multi-Stage Problems, Data-Driven Uncertainty Set

ACM Reference Format:

Yingxiao Ye, Christopher Doehring, Angelos Georghiou, Hugh Robinson, and Phebe Vayanos. 2022. Conserving Biodiversity via Adjustable Robust Optimization. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), Auckland, New Zealand, May 9–13, 2022*, IFAAMAS, 10 pages.

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Auckland, New Zealand.

1 INTRODUCTION

Biodiversity is needed for all life to thrive on Earth. Without it, ecosystems would crumble and humans would not have food to eat or air to breathe. Latin America has many biodiversity-related issues especially because of a dramatic increase in development in the Amazon Rainforest over the last several decades. According to the WWF, twenty percent of the Amazon biome has been lost to human development and it is estimated that 27 percent will be lost by 2030. The Amazon is home to about ten percent of all species on Earth and 3/4 of those are unique to the region. Few are as notable as the jaguar, the only big cat in the Americas and the third largest cat in the world. Jaguars are very reclusive creatures and their population is threatened by human contact and development. Their range can extend over much of Latin America but they especially like heavily canopied areas where they can stay hidden amongst dense rainforest and underbrush. Jaguars, as an umbrella species, help to maintain stability in their ecosystems and are good indicators of overall biodiversity (Jędrzejewski et al. 2018). To preserve biodiversity, in this paper, we propose to protect jaguars. Not-for-profits such as Panthera and other COs attempt to protect land all over the world for the purpose of preserving species and they do work in Latin America where they are trying to preserve jaguar biodiversity.

Unfortunately, given limited resources, it is not possible to protect all important sites. The problem of deciding which parcels of land to protect poses several practical challenges, which we describe below. Typically, conservation plans and objectives set by both governmental and non-governmental organizations extend well into the future while funds dedicated to conservation initiatives are limited. In addition, this limited budget is not made available immediately as a lump sum. Instead, it is received progressively over time as a series of installments (whose value is sometimes a-priori unknown). This implies that conservation plans must take into account both biodiversity and economic considerations from the onset to yield reserves that meet the intended long-term conservation targets (Joseph et al. 2009, Naidoo et al. 2006). Since conservation budgets are made available progressively over time, not all parcels can be protected from the onset and any parcel that has not been protected runs the risk of being lost to human degradation (e.g., transformed to agricultural land, to urban or suburban developments, deforested for use as lumber, or fragmented to build roads). This in turn can interfere with the implementation of conservation plans that do not take into account such contingencies, rendering them suboptimal or even infeasible, see e.g., Meir et al. (2004). From a planning perspective, it raises a significant challenge since conservation prioritization decisions should not only take into account biodiversity (and economic) considerations but also the "chance" that a land parcel will become lost to human land use.

The above highlights the need for decision-support tools to assist in the budget-constrained design of inter-connected networks of habitat reserves. Given the vast number of uncertainties that influence the feasibility and optimality of the conservation actions, the scarcity of data pertinent to inform these uncertainties, and the gravity of the problem of preserving biodiversity, it is imperative to be able to guarantee robustness of the conservation plans even in adverse circumstances. Indeed, biodiversity management policies that fail to acknowledge and account for these uncertainties robustly risk jeopardizing entire ecosystems forever. The uncertain human land use may play adversarially to the conservation plans. Whenever the land parcels is not protected, it runs the risk of getting developed due to the uncertain human land use. To prevent extreme loss in biodiversity due to these uncertainties, the conservation plans should be designed against the uncertain development in the most adverse case. Meanwhile, given the presence of uncertainty in the projected human land use, it is imperative for the network of habitat reserves to be able to adapt to such uncertainties to yield conservation actions that guarantee (with high "chance") the protection of sites that will serve as suitable habitat in the future, the existence of movement corridors, and the preservation of biodiversity. Although data for the biodiversity preservation problem is scarce and noisy, any data available should be leveraged to help build less conservative models of uncertainty.

Biodiversity conservation strategies differ in the extent to which dynamics in conservation targets, environmental conditions and conservation actions are recognized. Most of the literature on biodiversity preservation focuses on deterministic, single-stage models, see e.g. Church et al. (1996), Cocks and Baird (1989), Kiester et al. (1996), Polasky et al. (2001), Saetersdal et al. (1993), Snyder et al. (1999). Only few papers have investigated formulations that capture the multi-stage nature of the problem, see Costello and Polasky (2004), Jafari et al. (2017), Sabbadin et al. (2007). They model the conservation planning problem as a dynamic stochastic integer program. While it can principally be solved by backward induction, the approach is unfortunately plagued by the *curse of dimensionality*, see Bellman (2006). The authors thus end up having to rely on heuristics to be able to solve even moderately sized instances. Realizing that the spatial attributes of the reserve network are critical to species persistence and ecosystem service provision, several authors focused on the spatially explicit reserve design problem, see e.g., Dilkina and Gomes (2010), Dilkina et al. (2011), Le Bras et al. (2013), Onal and Briers (2006), Toth et al. (2009), Williams and Snyder (2005) and the references there-in. Very recently, the dynamic reserve design problem was extended to pay explicit attention to the connectivity of the resulting reserve network. Specifically, Jafari et al. (2017) model connectivity using network flows and, in each period, require that only sites adjacent to already protected ones be used to grow the reserve network.

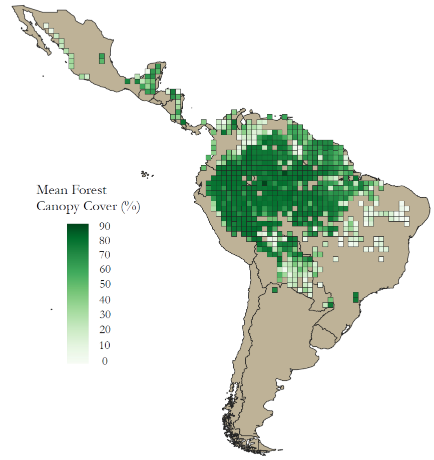
Thus, most approaches fail to capture the dynamic nature of the problem, rely on heuristics to solve it, or ignore the need for reserve connectivity. In addition, it is generally assumed that the distribution of the uncertain parameters in the problem can be perfectly estimated from data. More importantly, and to the best of our knowledge, most existing approaches do not fully utilize the historical data, such as climate and human impacts and animal range shifts data, to understand their relations and provide guidance on the areas of the landscape that are important to protect to help conserve species before they become threatened or endangered. To fill these gaps, in this project, we study data-driven robust approaches for preserving biodiversity. Additionally, motivated by the desiderata of the biodiversity preservation planning above, we propose to model the conservation planning problem as a *multi-stage robust optimization problem* whose objective is to minimize the worst-case level of biodiversity loss to human degradation by the end of the planning horizon. Since its advent over twenty years ago, modern robust optimization (Ben-Tal and Nemirovski 1998, 1999, Kouvelis and Yu 1996) has emerged as a popular alternative to stochastic optimization for optimal decision-making under uncertainty. In contrast to stochastic optimization which models uncertain parameters as random variables (see e.g., Prékopa (1995)), robust optimization takes a deterministic view of uncertainty whereby the decision-maker seeks a solution that is immunized against any realization of the uncertain parameters in a given set, termed *uncertainty set*. The popularity of the robust optimization paradigm has been driven by its modeling power, broad applicability, and computational attractiveness (Bertsimas et al. 2010). We use climate, geographic, and anthropogenic data including precipitation, temperature, canopy, human population density, current jaguar range, protection status, development threat index, latitude, and longitude.

This paper is organized as follows. Section 2 introduces the data involved in this project; Section 3 describes the problem and proposed method to solve it; Section 4 discusses how to construct the uncertainty set; Section 5 provides an equivalent reformulation of the problem which is computationally solvable ; Section 6 provides a computationally tractable approximation approach; Section 7 shows how to implement the proposed method and benchmark it to a Knapsack method, in which we formulate the problem as a knapsack problem and ignore the uncertain development.

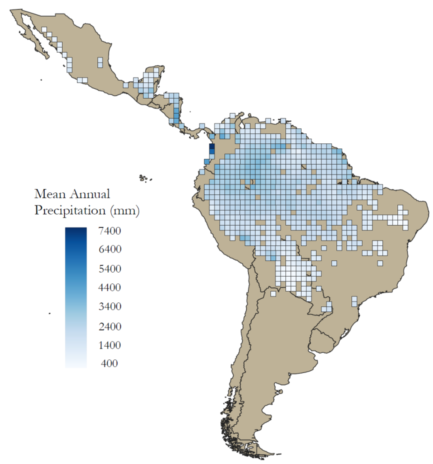
Notation. Throughout this paper, vectors (matrices) are denoted by boldface lowercase (uppercase) letters. The kth element of a vector $\boldsymbol{x} \in \mathbb{R}^n$ ($k \leq n$) is denoted by x_k. Scalars are denoted by lowercase letters, e.g., α or u.

Contributions. In this paper, we propose a multistage robust optimization model for conservation planning which is adaptive to human land use. We leverage the cellular automata simulation to construct a novel data-driven likelihood uncertainty set capturing spatio-temporal dependencies. Though the original formulation is decision dependent and involves objective and constraint uncertainty, which makes the problem hard to solve, we provide an equivalent computationally tractable exogenous reformulation with objective uncertainty only, and we solve it by a proposed approximation approach. We perform numerical experiments based on the real data from (Jędrzejewski et al. 2018) and simulation. The numerical results showed that our method reduces 19.46% conservation loss compared with the general knapsack method.

Figure 1: 1757 Parcels in Latin America. The Area of Each Parcel is 12321 km^2



(a) Canopy Map



(b) Precipitation Map

Figure 2: Geographical Data

2 DATA

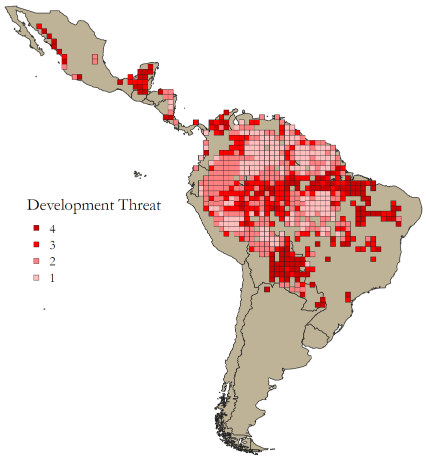
In this paper, we are using different kinds of natural and anthropogenic data. We need data that will help identify good habitat areas for jaguars as well as data to help us model human development. Fortunately, climate data in raster, or image based, form is available online as well as all of the anthropogenic data we need, also in raster form. These rasters allow us to extract data at specific points which is what we do to get data for all of the parcels we are concerned with. The first thing we do is creating a grid of equal-sized parcels in Latin America using a geographical information system software called QGIS. Each parcel is one degree of latitude by one degree of longitude. This totaled to 1757 parcels. To represent each parcel for data extraction, we place a point in the center of each parcel from which we will extract the data we need. Figure 1 shows a representation of these parcels on a map where each parcel has a point in the middle which will be used for extracting data for the parcel.

The climate and geographical datasets used in this paper are mean temperature, mean precipitation, canopy, latutude, and longitude. All of these datasets came from (Jędrzejewski et al. 2018) Figure 2(a) shows a visualization of the canopy dataset and Figure 2(b) shows the average preciptation. The darker the area in the map, the more canopy exists in that area. This is important because jaguars are more likely to be found in areas with more canopy as they use trees for hunting as well as for shelter (Jędrzejewski et al. 2018). All of these climate-related datasets are important because jaguars have a specific habitat and we use the climate data to make better conservation decisions.

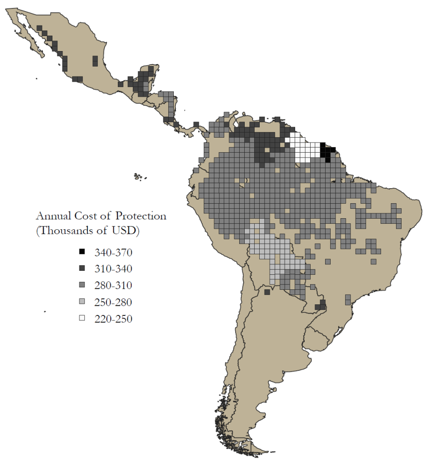
Some of the other data we use is development threat index data from NASA SEDAC (Socioeconomic Data and Applications Center) (Oakleaf et al. 2015). This data is an estimate of the potential for human development in the future. This is the best data available for predicting development patterns so it allows us to figure out what areas to prioritize based on what could be developed sooner. Figure 3(a) shows a visualization of this dataset on a map. The lighter areas represent a lower development threat while the darker areas represent a higher threat.

We also use protection status and current jaguar range from (Jędrzejewski et al. 2018). Protection status allow us to determine areas that are already protected and that will not be developed by humans. We throw out any parcels that had already been protected because they are not of concern for our conservation problem. We also threw out any parcels that do not fall in the jaguar's range as we have no interest in protecting parcels if they do not have jaguars living in. This range is visualized in Figure 4 where any of the areas with color lie within the jaguar's range while the white areas do not.

One important part of the data is to figure out the value of protecting each parcel as well as the cost of protecting each parcel. We calculate the cost in USD for each parcel using land cost estimation techniques used in (Balmford et al. 2003) and (Kark et al. 2009). We gather our data for land area, purchasing parity power, and GNI from the World Bank (https://www.worldbank.org/). This helps us determine whether or not the purchase of the parcel is worth it given the value of the parcel and gives us a way to directly compare



(a) Development Threat Index Map



(b) Cost Map

Figure 3: Threat Index and Cost

the cost of parcels to one another. Figure 3(b) is a visualization of the cost where the darker points represent more expensive parcels and the lighter points represent less expensive parcels.

The best estimation of value is the adjusted jaguar density from (Jędrzejewski et al. 2018). We decide that jaguar density will be the best indicator of value as we want to protect as many jaguars as possible with our efforts. Figure 4 shows a geographic representation of jaguar density.

We choose to use all of this data because they all tie back to the uncertain nature of human activities which we are examining. The data including precipitation, canopy, temperature, and current jaguar range is used because they are some of the best indicators of jaguar presence according to (Jędrzejewski et al. 2018). They are also indicative of areas that would likely be developed by humans for deforestation, as much of the Amazon Rainforest is threatened by human activity, which is why we also use human population density because in places of higher population density, there would be more human activity and development. We also used development threat index because it is our best way of predicting future human development. While the climate variables are helpful in predicting what areas might be developed, we want something that directly shows future development potential and development threat index is the best indicator. Another important part of biodiversity planning is interconnection of the land parcels used to preserve biodiversity. According to (Dilkina et al. 2017), a systems of isolated protected areas would be insufficient to preserve biodiversity and the protected areas instead need to have some connectivity in order to be effective. This is why we use latitude and longitude as part of our data because we want to make sure that geography is a factor as well as all of the other kinds of data.

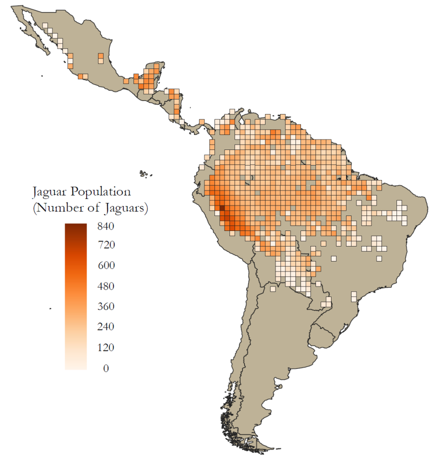


Figure 4: Jaguar Population Map

3 PROBLEM STATEMENT

This problem is multifaceted and involves several challenges. First, there are long planning horizons and very limited budgets available for use, which can vary from year to year. Secondly, there is uncertainty in how humans will use land in the future. Not every parcel can be protected at the same time, so we must find a way to prioritize parcels based on a combination of features, including development potential, so that important parcels can be protected before they are used for other purposes such as logging. If a parcel is not protected, then the potential adversarial human land use may develop it and the jaguars inhabitating there will be lost. Finally, we have access to limited data and must make good use of what we have available to us.

Because of uncertain and complex nature of this problem, we must make our model robust so that it can hold up to such uncertainty. Also, the parcels we choose must be able to sustain biodiversity into the future, even if there are changes in climate factors and in human land use. We must also leverage the data available to us in order to build this model, even though the data is somewhat scarce and noisy. Lastly, our solution must be scalable to large instances. Biodiversity conservation problems involve large amounts of land over massive time horizons. We must take this into account when building our model.

Given all of the aspects of this problem listed above, we are modeling this problem as a multi-stage robust optimization problem whose objective is to minimize the worst case level of biodiversity loss to human degradation by the end of the planning horizon. Robust optimization takes a deterministic view of uncertainty where the decision-maker seeks a solution that is against any realization of the uncertain parameters in a given set, which is called the uncertainty set. Basically, by formulating the conservation problem as a robust optimization, conservation plans and the human development can be regarded as two agents who play against each other. To be specific, human development will play adversarially by maximizing the loss in jaguars in those parcels which are not protected based on the rules determined by the uncertainty set, while the conservation plans need to protect parcels so as to minimize such loss. Next is a description of how we formulated the problem.

We consider the problem faced by a conservation planner, such as Panthera, charged with protecting a site that consists of $|\mathcal{I}|$ land parcels indexed in the set $\mathcal{I}$ over the finite planning horizon $\mathcal{T} := \{1, \ldots, T\}$. At the beginning of each period (usually each year) $t \in \mathcal{T}$, the conservation planner is allocated a conservation budget $b_t \in \mathbb{R}_+$ which can be used to protect land parcels from the site (e.g., by purchasing them). We let $c_i \in \mathbb{R}_+$ denote the cost of introducing parcel i in the reserve (usually corresponding to the purchasing cost). Note that if a parcel is purchased at time t it will remain part of the reserve in the future. We assume that the sequence of funds $\{b_t\}_{t\in\mathcal{T}}$ that will be made available to the conservation planner throughout the planning horizon along with the corresponding sequence of parcel protection costs $\{c_i\}_{i\in\mathcal{I}}$ are both perfectly known at the beginning of the planning horizon. Associated with each parcel i is its conservation value $v_i \in \mathbb{R}_+$, which encodes the richness of its genetic and ecological biodiversity which is assumed known and fixed. Between consecutive periods, each land parcel risks becoming developed (i.e., being lost to human degradation), in which case its conservation value vanishes, representing the fact that all genetic and ecological biodiversity present in the parcel is lost. We let $\xi_{it} \in \{0, 1\}$ represent the status of land parcel i at time t. Thus, $\xi_{it} = 1$ if and only if land parcel i has been developed on or before time t. For notational convenience, we let $\boldsymbol{\xi}_t := \{\xi_{it}\}_{i\in\mathcal{I}}$ collect the (uncertain) statuses of all parcels at time t. In the spirit of robust optimization, we propose to model the uncertain parameters in our problem as deterministic variables that are constrained to lie in an uncertainty set, which we denote by $\mathcal{U} \subseteq \{0, 1\}^{|\mathcal{I}|\times|\mathcal{T}|}$. This set reflects fundamental known properties that the original random quantities would satisfy with high probability in the absence of the conservation planner's intervention. As an illustrative example, $\mathcal{U}$ will capture the requirement that if a parcel i is developed at time $t \in \mathcal{T}$, then it remains developed at time $t + 1$ by means of the constraint $\xi_{i,t+1} \geq \xi_{it}$. We will introduce our uncertainty set $\mathcal{U}$ in the next section. Throughout this paper, we will make the assumption that $\mathcal{U}$ is polyhedral. We argue that polyhedral sets are already rich enough to capture many of the dynamics of interest.

The decision variable, which is denoted by $x_{it} \in \{0, 1\}$, is 1 if and only if the land parcel i has been protected on or before time t. It is taken after the history of parcel development statuses $\boldsymbol{\xi}^{t-1} := \{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{t-1}\}$ has been revealed but before future developments $\{\boldsymbol{\xi}_\tau\}_{\tau\geq t}$ are observed. This motivates us to model the decision x_{it} to protect parcel i at time t as a function that maps the history of observations $\boldsymbol{\xi}^{t-1}$ to conservation actions, i.e., $x_{it} : \{0, 1\}^{|\mathcal{I}|} \to \{0, 1\}$. Thus, x_{it} is modeled as a binary adaptive decision variable or decision rule. The requirement that x_{it} only depends on the history of observations reflects the causal nature of decisions is referred to as non-anticipativity in the stochastic-programming literature. For convenience, we let $\mathcal{N}$ denote the space of all non-anticipative decision rules. Explicitly,

$$\mathcal{N} := \left\{ \begin{array}{ll} x_{it}(\cdot): & x_{it} \in \{0,1\}^{|\mathcal{I}|} \to \{0,1\} \\ & \text{and } x \text{ is adaptive to } \boldsymbol{\xi}^{t-1} \ \ \forall i \in \mathcal{I}, t \in \mathcal{T} \end{array} \right\}.$$

We always assume that at the beginning, all the parcels have not been developed or protected, which means $\xi_{i0} = 0, \ \ \forall i \in \mathcal{I}$ and $x_{i0}(\boldsymbol{\xi}^0) = 0 \ \ \forall \boldsymbol{\xi} \in \mathcal{U}, \ \forall i \in \mathcal{I}$. Also, we hope that the planner can make decision at first time stage before the nature develops parcels, so we set $\xi_{i0} = 0 \ \ \forall i \in \mathcal{I}$. Crucially, the only motivation for the conservation planner to protect a land parcel is that it will preclude it from becoming developed in the future. Thus, by suitably deciding which parcels to protect, the conservation planner is able to strategically influence or modify the uncertainty set $\mathcal{U}$. He is then faced with a new policy-dependent uncertainty set $\Xi : \mathcal{N} \to \{0, 1\}^{|\mathcal{I}|\times|\mathcal{T}|}$, expressible as:

$$\Xi(\boldsymbol{x}) = \{\boldsymbol{\xi} \in \mathcal{U} : \xi_{it} \leq 1 - x_{it}(\boldsymbol{\xi}^{t-1}), \forall i \in \mathcal{I}, t \in \mathcal{T}\}. \quad (1)$$

Relative to $\mathcal{U}$, $\Xi(\boldsymbol{x})$ involves one additional constraint for each parcel-time-period pair. This constraints stipulates that if a parcel is protected at time $t - 1$, then it can no longer be lost to human degradation at time t. In particular $\Xi(\boldsymbol{x}) \subseteq \mathcal{U}$ for all $\boldsymbol{x}$. This model affords an intuitive interpretation: by making strategic conservation plans, the decision-maker (conservation planner) can reduce uncertainty by precluding undesirable scenarios from materializing. The objective of the conservation planner is to minimize the worst-case conservation value lost to human land-use across the entire site. Mathematically, the robust biodiversity conservation problem affected by human land use uncertainty is expressible as

$$\begin{array}{ll} \text{minimize} & \sup\limits_{\boldsymbol{\xi}\in\Xi(\boldsymbol{x})} \sum\limits_{i\in\mathcal{I}} v_i \xi_{iT} \\ \text{subject to} & \boldsymbol{x} \in \mathcal{N} \\ & \left. \begin{array}{l} \sum\limits_{i\in\mathcal{I}} c_i [x_{it}(\boldsymbol{\xi}^{t-1}) - \boldsymbol{x}_{i,t-1}(\boldsymbol{\xi}^{t-2})] \leq b_t \\ \left. \begin{array}{l} x_{it}(\boldsymbol{\xi}^{t-1}) - x_{i,t-1}(\boldsymbol{\xi}^{t-2}) \leq 1 - \xi_{i,t-1} \\ x_{it}(\boldsymbol{\xi}^{t-1}) \geq x_{i,t-1}(\boldsymbol{\xi}^{t-2}) \\ x_{it}(\boldsymbol{\xi}^{t-1}) \in \{0,1\} \end{array} \right\} \forall i \in \mathcal{I} \end{array} \right\} \begin{array}{l} \forall t \in \mathcal{T}, \\ \boldsymbol{\xi} \in \Xi(\boldsymbol{x}). \end{array} \end{array} \quad (2)$$

The objective function of the problem corresponds to the worst-case loss in conservation value achieved by an adaptive conservation strategy $\boldsymbol{x}$ when "nature" can select parcels to develop from the policy-dependent uncertainty set $\Xi(\boldsymbol{x})$, i.e., from those parcels that are not yet protected. The first constraint corresponds to a budget constraint. The second constraint stipulates that if a site has already been lost to human degradation, it can no longer be introduced in the reserve. The third constraint captures the requirement that if a parcel i is protected at time $t \in \mathcal{T}$, then it remains protected at time $t + 1$. All constraints are enforced robustly, i.e, for all possible realizations of $\boldsymbol{\xi}$ in the policy dependent uncertainty set $\Xi(\boldsymbol{x})$.

Problem (2) is a multi-stage adaptive optimization problem involving high-dimensional binary (0-1) randomness and an endogenous uncertainty set affected by the conservation policy of the planner.

4 DESCRIPTION OF THE UNCERTAINTY SET

In this section, we construct the the uncertainty set $\mathcal{U}$ in our robust model in Section 3, based on the data that we have at our disposal. In real world applications, the neighboring interaction plays an important role in parcel development. This means that rather than independent random development among parcels, a parcel is more likely to be developed if more of its neighboring parcels have been developed. This motivates the use of the cellular automata, a commonly-used model in land use change literature. In the cellular automata model, each parcel can be classified as developed or undeveloped. In each iteration, the probability of getting developed of each parcel is calculated based on both the threat index and the status of the neighboring parcels, i.e.

$$prob_i^t = \frac{TI_i}{10} \cdot \frac{1 + \sum_{j \in \Omega_i} \mathbb{1}_{\{j \text{ is developed}\}}}{1 + |\Omega_i|}, \tag{3}$$

where TI_i is the threat index of parcel i and Ω_i is the set of the neighbors of parcels i. To emphasize the importance of connectivity of protected parcels as discussed in Section 2 and motivated by the fact that parcels with similar geographical characteristics tend to develop (or not develop) together, we divide the whole area into 9 clusters based on the threat index and geographical features, such as location and average temperature. Then, we define the neighbors of a parcel as the adjacent parcels which are in the same cluster. After the probability of getting developed is calculated, the development of the parcel i is considered follows $Bernoulli(prob_i)$. Thus, we use the cellular automata model to simulate the development of each parcel for different time stages, and the development risk of each parcel is calculated as the proportion of development in the simulation. The development risk p_i^t is regarded as the probability of getting developed for parcel i at time stage t. The uncertainty set is constructed to only allow the realizations of development whose likelihood is beyond an user-input parameter $\boldsymbol{\lambda}$, which controls the robustness of the model. The uncertainty set is as follows.

$$\mathcal{U} = \left\{ \begin{array}{ll} \xi \in \{0,1\}^{|\mathcal{I}| \times |\mathcal{T}|} : & \\ \xi_{it} \geq \xi_{i,t-1} & \forall i \in \mathcal{I}, t \in \mathcal{T} \\ \prod_{i \in \mathcal{I}} (p_i^t)^{\xi_{it}} \cdot (1 - p_i^t)^{(1-\xi_{it})} \geq \lambda^t & \forall t \in \mathcal{T} \end{array} \right\}.$$

5 EQUIVALENT REFORMULATION

Problem (2) is a multistage endogenous robust optimization problem. Due to the large size of parcels and the rolling horizons, it is impractical to solve Problem (2) directly. We need an equivalent reformulation which is easier to handle. The following theorems provide an equivalent formulation in which we get rid of the constraints uncertainty and the uncertainty set is independent to the solution $\boldsymbol{x}$.

Equivalent Reformulation with Deterministic Constraints. We introduce a variant of Problem (2) from which we eliminate the second constraint to obtain

$$\begin{array}{ll} \text{minimize} & \sup\limits_{\xi \in \Xi(\boldsymbol{x})} \sum\limits_{i \in \mathcal{I}} v_i \sum\limits_{t \in \mathcal{T}} [\xi_{it} - \xi_{i,t-1}] \left[1 - x_{it}(\xi^{t-1})\right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{N} \\ & \left. \begin{array}{l} \sum\limits_{i \in \mathcal{I}} c_{it} [x_{it}(\xi^{t-1}) - x_{i,t-1}(\xi^{t-2})] \leq b_t \\ \left. \begin{array}{l} x_{it}(\xi^{t-1}) \geq x_{i,t-1}(\xi^{t-2}) \\ x_{it}(\xi^{t-1}) \in \{0,1\} \end{array} \right\} \forall i \in \mathcal{I} \end{array} \right\} \begin{array}{l} \forall t \in \mathcal{T}, \\ \xi \in \Xi(\boldsymbol{x}). \end{array} \end{array} \tag{4}$$

Theorem 5.1. *The robust biodiversity conservation Problem* (2) *is equivalent to Problem* (4).

Equivalent Reformulation with Exogenous Uncertainty Set. We introduce the problem

$$\begin{array}{ll} \text{minimize} & \sup\limits_{\xi \in \mathcal{U}} \sum\limits_{i \in \mathcal{I}} v_i \sum\limits_{t \in \mathcal{T}} [\xi_{it} - \xi_{i,t-1}] \left[1 - x_{it}(\xi^{t-1})\right] \\ \text{subject to} & \boldsymbol{x} \in \mathcal{N} \\ & \left. \begin{array}{l} \sum\limits_{i \in \mathcal{I}} c_{it} [x_{it}(\xi^{t-1}) - x_{i,t-1}(\xi^{t-2})] \leq b_t \\ \left. \begin{array}{l} x_{it}(\xi^{t-1}) \geq x_{i,t-1}(\xi^{t-1}) \\ x_{it}(\xi^{t-1}) \in \{0,1\} \end{array} \right\} \forall i \in \mathcal{I} \end{array} \right\} \begin{array}{l} \forall t \in \mathcal{T}, \\ \xi \in \mathcal{U}. \end{array} \end{array} \tag{5}$$

Theorem 5.2. *Problem* (5) *is a conservative approximation of the robust biodiversity conservation Problem* (2)*. If the constraints in the uncertainty set* $\mathcal{U}$ *except the non-decreasing constraints can be written in the form* $f^T \xi_t \leq g$*, where all the components of* f *are non-negative, then the robust biodiversity conservation Problem* (2) *is equivalent to Problem* (5).

Remark 5.1. *Note that when the uncertainty set* $\mathcal{U}$ *does not satisfy the condition in Theorem 5.2, we cannot claim that Problem* (2) *is not equivalent to Problem* (5)*. Here is an counterexample: consider a single stage problem with two parcels, i.e.* $|\mathcal{I}| = 2, |\mathcal{T}| = 2$*. When* $v_1 = v_2 = 1, c_{11} = 10, c_{21} = 1, b_1 = 2$ *and the uncertainty set* $\mathcal{U} = \left\{\xi \in \{0,1\}^2 : \xi_{11} - \xi_{21} \leq 0\right\}$*, the optimal objective value to Problem* (2) *is 0, while the optimal objective value to Problem* (5) *is 1.*

Remark 5.2. *Problem* (5) *has exogenous uncertainty only, binary adaptive variables, and uncertainty in the objective only and thus can be solved with off the shelf techniques from the literature (Bertsimas and Georghiou 2015, 2017, Hanasusanto et al. 2016, Subramanyam et al. 2019, Vayanos et al. 2011). In section 7, we show the performance of a static (i.e. conservative) approximation to the Problem* (5)*, which is sufficient to illustrate the benefits of incorporating uncertainty directly in the model.*

6 APPROXIMATION APPROACH

The scenario-based extensive form of Problem (5) is as follows

$$
\begin{aligned}
\text{minimize} \quad & \tau \\
\text{subject to} \quad & \tau \in \mathbb{R} \\
& \left.\begin{aligned}
& \tau \geq \sum_{i \in \mathcal{I}} v_i \sum_{t \in \mathcal{T}} (\xi_{it} - \xi_{i,t-1})(1 - x^{\xi^t}_{i,t-1}) \\
& \left.\begin{aligned}
& \sum_{i \in \mathcal{I}} c_{it} (x^{\xi^t}_{it} - \boldsymbol{x}^{\xi^t}_{i,t-1}) \leq b_t \\
& \left.\begin{aligned}
& x^{\xi^t}_{it} \geq x^{\xi^t}_{i,t-1} \\
& x^{\xi^t}_{it} \in \{0, 1\}
\end{aligned}\right\} \forall i \in \mathcal{I}
\end{aligned}\right\} \forall t \in \mathcal{T}
\end{aligned}\right\} \forall \xi \in \mathcal{U}
\end{aligned}
\tag{6}
$$

Given that the huge number of scenarios in the uncertainty set $\mathcal{U}$ makes the problem computationally intractable, we approximate the optimal solution of Problem (6) by solving the following static problem

$$
\begin{aligned}
\text{minimize} \quad & \tau \\
\text{subject to} \quad & \tau \in \mathbb{R} \\
& \tau \geq \sum_{i \in \mathcal{I}} v_i \sum_{t \in \mathcal{T}} (\xi_{it} - \xi_{i,t-1})(1 - x_{i,t-1}) \quad \forall \xi \in \mathcal{U} \\
& \left.\begin{aligned}
& \sum_{i \in \mathcal{I}} c_{it} (x_{it} - \boldsymbol{x}_{i,t-1}) \leq b_t \\
& \left.\begin{aligned}
& x_{it} \geq x_{i,t-1} \\
& x_{it} \in \{0, 1\}
\end{aligned}\right\} \forall i \in \mathcal{I}
\end{aligned}\right\} \forall t \in \mathcal{T}.
\end{aligned}
\tag{7}
$$

In Problem (7), all decision variables are here and now. In other words, it is a one-stage optimization problem and we call it StaticApprox. This problem then can be solved by constraint generation algorithm, in which the problem is initialized with an empty uncertainty set and scenarios in the uncertainty set are added progressively when the solution is not optimal. In Section 7, we compare the performance of StaticApprox with the following knapsack problem.

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i \in \mathcal{I}} v_i x_i \\
\text{subject to} \quad & \sum_{i \in \mathcal{I}} c_i x_i \leq b_1 \\
& x_i \in \{0, 1\} \quad \forall i \in \mathcal{I}.
\end{aligned}
\tag{8}
$$

We can think of the reserve selection problem in a knapsack problem framework, which is described as follows. At each time stage the conservation planner is given a certain amount of budget which the planner can use to purchase some parcels to protect. The objective is to maximize the value of parcels protected in the final time stage. This problem can be modeled as the following integer linear programming problem.

$$
\begin{aligned}
\text{maximize} \quad & \sum_{i \in \mathcal{I}} v_i x_{iT} \\
\text{subject to} \quad & \sum_{i \in \mathcal{I}} c_i [x_{it} - x_{i,t-1}] \leq b_t \quad \forall t \in \mathcal{T} \\
& \left.\begin{aligned}
& x_{it} \geq x_{i,t-1} \\
& x_{it} \in \{0, 1\}
\end{aligned}\right\} \forall i \in \mathcal{I},\ \forall t \in \mathcal{T}.
\end{aligned}
\tag{9}
$$

In Section 7, we will compare the solution given by the StaticApprox to Problem (5) with the solution given by the knapsack problem (denoted as Knapsack).

Note that when the robustness level $\frac{1}{\lambda}$ is larger than a certain threshold, if we keep increasing Γ, the performance of StaticApprox will be getting similar to the knapsack problem. Intuitively, when the uncertainty set is large enough, all the parcels which have not been protected will be developed immediately. In this case, minimizing the loss is equivalent to maximizing the protected value, which is what the knapsack problem optimizes for. We summarize this in the following proposition.

PROPOSITION 6.1. *For sufficient small $\boldsymbol{\lambda}$, particularly, the uncertainty set $\mathcal{U}$ contains non-decreasing constraints only, the following two statements are true.*

(a) Given any optimal solution $\bar{\boldsymbol{x}}$ to Problem (2)*, there is an optimal solution $\tilde{\boldsymbol{x}}$ to Problem* (8) s. t. $\{\tilde{x}_i\}_{i \in \mathcal{I}} = \{\bar{x}_{i1}\}_{i \in \mathcal{I}}$.

(b) Given any optimal solution $\tilde{\boldsymbol{x}}$ to Problem (8)*, there is an optimal solution $\bar{\boldsymbol{x}}$ to Problem* (2) s. t. $\{\bar{x}_{i1}\}_{i \in \mathcal{I}} = \{\tilde{x}_i\}_{i \in \mathcal{I}}$.

7 NUMERICAL EXPERIMENTS

Data Preprocessing. We cleaned up the data by erasing any points with missing values and were left with 1678 data points to use. After that, we removed parcels which are already protected according to (Jędrzejewski et al. 2018). From there, we were left with 1310 available points. We narrowed our data down further by excluding any points outside of the jaguar's current natural range, which was also from (Jędrzejewski et al. 2018), since there would be benefit in protecting an area where we know jaguars do not live. This left us with 692 parcels. Figure 5 shows the protection status data and the parcels left after data cleaning.

To calibrate the uncertainty set, we used the threat index and K-Means clustering to create a hierarchy of four groups. The highest level group is all of Latin America and includes all of our points. To get the next level, we use the development threat data to divide all parcels into four clusters. We did this because we want to prioritize development threat as a factor in our decisions to protect parcels. The development threat data we used came from NASA SEDAC (Oakleaf et al. 2015) put every area of land on a 1-4 scale of threat with 1 being the least threatened and 4 being the most threatened. The third level was to create the subclusters within each initial cluster. We used K-Means clustering on that group and got nine clusters. We decided on the number of clusters to use in each clustering based on an R function called NbClust (Charrad et al. 2014). The parameters we used to generate these initial clusters were latitude, longitude, precipitation, mean temperature, canopy, and human population density. The reasons for using these

Figure 5: Protection Status

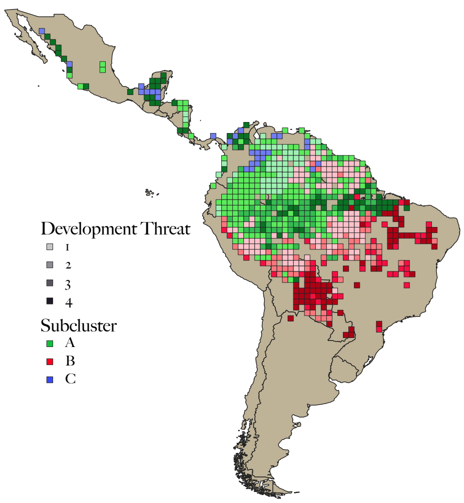
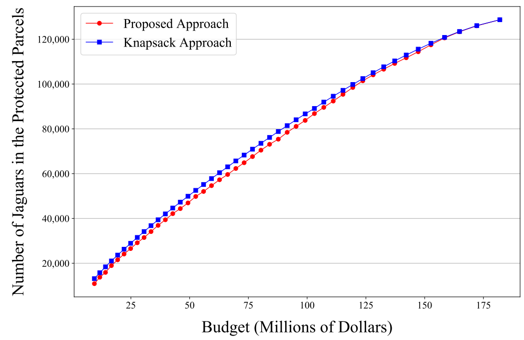
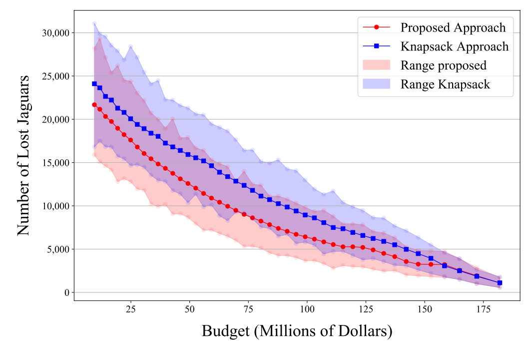


Figure 6: Nine Initial Clusters



(a) Number of Jaguars in Protected Parcels



(b) Number of Jaguars Lost to Development

Figure 7: Performance of proposed approach and of knapsack method in terms of Jaguars preserved/in conserved areas

specific datasets is discussed in Section 2. All of these values were normalized. Figure 6 shows what these nine clusters look like.

Results. In the numerical experiment, we consider a single-stage problem, which can be solved to optimality within 10 minutes. We benchmark our proposed approach against the knapsack approach. Figure 7 shows the performance of the proposed approach and knapsack approach in terms of the value of protected parcels and the average value of lost parcels for different budgets. The red curve is the proposed approach and the blue one is the knapsack approach. The shaded area is the range between the maximal lost and minimal lost in the 1000 simulated samples. Although in Figure 7(a) on the left, knapsack approach has a better performance in terms of the value of protected parcels, in real world applications, however, what we truly want is to decrease the loss due to development. So in terms of the number of lost jaguars, which is shown in Figure 7(b), the proposed approach can prevent more jaguar loss. Actually, the proposed approach can decrease the loss by 19.46 percentage in average.

8 CONCLUSIONS AND FUTURE RESEARCH

This paper formulated the biodiversity conservation problem as a game between the conservation organizations and human development. It proposed a multistage robust optimization model to minimize the value of loss due to the development. We constructed a likelihood uncertainty set based on the data from (Jędrzejewski et al. 2018), and reformulate the robust optimization problem with endogenous constraint uncertainty into a robust optimization problem with exogenous objective uncertainty, which is computationally tractable. We used a proposed StaticApprox method to approximately solve the problem. The numerical results suggest that the StaticApprox method outperforms the general knapsack method.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1763108. P. Vayanos, Y. Ye and C. Dohering gratefully acknowledge support from the National Science Foundation through NSF Research Experiences for Undergraduates program.

REFERENCES

A. Balmford, K. J. Gaston, S. Blyth, A. James, and V. Kapos. 2003. Global variation in terrestrial conservation costs, conservation benefits, and unmet conservation needs. *Proceedings of the National Academy of Sciences of the United States of America* 100, 3 (2 2003), 1046–50. https://doi.org/10.1073/pnas.0236945100

R. Bellman. 2006. *Dynamic Programming*. Courier Dover Publications.

A. Ben-Tal and A. Nemirovski. 1998. Robust convex optimization. *Mathematics of Operations Research* 23, 4 (1998), 769–805.

A. Ben-Tal and A. Nemirovski. 1999. Robust Solutions of uncertain linear programs. *Operations Research Letters* 25 (1999), 1–13.

D. Bertsimas, D. B. Brown, and C. Caramanis. 2010. Theory and applications of robust optimization. *SIAM Rev.* 53, 3 (2010), 464–501.

D. Bertsimas and A. Georghiou. 2015. Design of Near Optimal Decision Rules in Multistage Adaptive Mixed-Integer Optimization. *Operations Research* 63, 3 (2015), 610–627. https://doi.org/10.1287/opre.2015.1365 arXiv:https://doi.org/10.1287/opre.2015.1365

D. Bertsimas and A. Georghiou. 2017. Binary decision rules for multistage adaptive mixed-integer optimization. *Mathematical Programming* (24 Mar 2017). https://doi.org/10.1007/s10107-017-1135-6

M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs. 2014. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *Journal of Statistical Software* 61, 6 (2014). https://doi.org/10.18637/jss.v061.i06

R. L. Church, D. M. Stoms, and F. W. Davis. 1996. Reserve selection as a maximal covering location problem. *Biological Conservation* 76, 2 (1996), 105 – 112. https://doi.org/10.1016/0006-3207(95)00102-6

K. D. Cocks and I. A. Baird. 1989. Using mathematical programming to address the multiple reserve selection problem: An example from the Eyre Peninsula, South Australia. *Biological Conservation* 49, 2 (1989), 113 – 130. https://doi.org/10.1016/0006-3207(89)90083-9

C. Costello and S. Polasky. 2004. Dynamic reserve site selection. *Resources and Energy Economics* 26, 2 (6 2004), 157–174. https://doi.org/10.1016/j.reseneeco.2003.11.005

B. Dilkina and C. P. Gomes. 2010. Solving connected subgraph problems in wildlife conservation. In *International Conference on Integration of Artificial Intelligence (AI) and Operations Research (OR) Techniques in Constraint Programming*. Springer Berlin Heidelberg, 102–116.

B. Dilkina, R. Houtman, C. P. Gomes, C. A. Montgomery, K. S. McKelvey, K. Kendall, T. A. Graves, R. Bernstein, and M. K. Schwartz. 2017. Trade-offs and efficiencies in optimal budget-constrained multispecies corridor networks. *Conservation Biology* 31, 1 (2 2017), 192–202. https://doi.org/10.1111/cobi.12814

B. Dilkina, K. J. Lai, and C. P. Gomes. 2011. Upgrading shortest paths in networks. In *International Conference on AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*. Springer Berlin Heidelberg, 76–91.

G. A. Hanasusanto, D. Kuhn, and W. Wiesemann. 2016. K-adaptability in two-stage distributionally robust binary programming. *Operations Research Letters* 44, 1 (2016), 6 – 11. https://doi.org/10.1016/j.orl.2015.10.006

N. Jafari, B. L. Nuse, C. T. Moore, B. Dilkina, and J. Hepinstall-Cymerman. 2017. Achieving full connectivity of sites in the multiperiod reserve network design problem. *Computers & Operations Research* 81 (2017), 119–127.

W. Jędrzejewski, H. S. Robinson, M. Abarca, K. A. Zeller, G. Velasquez, E. A. D. Paemelaere, J. F. Goldberg, E. Payan, R. Hoogesteijn, E. O. Boede, K. Schmidt, M. Lampo, Á. L. Viloria, R. Carreño, N. Robinson, P. M. Lukacs, J. J. Nowak, R. Salom-Pérez, F. Castañeda, V. Boron, and H. Quigley. 2018. Estimating large carnivore populations at global scale based on spatial predictions of density and distribution – Application to the jaguar (Panthera onca). *PLOS ONE* 13, 3 (3 2018), e0194719. https://doi.org/10.1371/journal.pone.0194719

L. N. Joseph, R. F. Maloney, and H. P. Possingham. 2009. Optimal allocation of resources among threatened species: a project prioritization protocol. *Conservation biology* 23, 2 (2009), 328–38.

S. Kark, N. Levin, H. S. Grantham, and H. P. Possingham. 2009. *Between-country collaboration and consideration of costs increase conservation planning efficiency in the Mediterranean Basin*. Technical Report. www.pnas.org/cgi/content/full/

A. R. Kiester, J. M. Scott, B. Csuti, R. F. Noss, B. Butterfield, K. Sahr, and D. White. 1996. Conservation Prioritization Using GAP Data. *Conservation Biology* 10, 5 (1996), 1332–1342. https://doi.org/10.1046/j.1523-1739.1996.10051332.x

P. Kouvelis and G. Yu. 1996. *Robust Discrete Optimization and Its Applications (Nonconvex Optimization and Its Applications (closed))* (1st ed.). Springer. http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0792342917

R. Le Bras, B. Dilkina, Y. Xue, C. P. Gomes, K. S. McKelvey, M. K. Schwartz, and C. A. Montgomery. 2013. Robust network design for multispecies conservation. In *AAAI*.

E. Meir, S. Andelman, and H. P. Possingham. 2004. Does conservation planning matter in a dynamic and uncertain world? *Ecology Letters* 7, 8 (2004), 615–622. https://doi.org/10.1111/j.1461-0248.2004.00624.x

Robin Naidoo, Andrew Balmford, Paul J. Ferraro, Stephen Polasky, Taylor H. Ricketts, and Mathieu Rouget. 2006. Integrating Economic Costs into Conservation Planning. *Trends in Ecology & Evolution* 21, 12 (2006), 681–687.

J. R. Oakleaf, C. M. Kennedy, S. Baruch-Mordo, P. C. West, J. S. Gerber, L. Jarvis, and J. Kiesecker. 2015. A World at Risk: Aggregating Development Trends to Forecast Global Habitat Conversion. *PLOS ONE* 10, 10 (10 2015), e0138334. https://doi.org/10.1371/journal.pone.0138334

H. Onal and R. A. Briers. 2006. Optimal Selection of a Connected Reserve Network. *Oper. Res.* 54, 2 (2006), 379–388. arXiv:http://or.journal.informs.org/cgi/reprint/54/2/379.pdf

S. Polasky, J. D. Camm, and B. Garber-Yonts. 2001. Selecting Biological Reserves Cost-Effectively: An Application to Terrestrial Vertebrate Conservation in Oregon. *Land Economics* 77, 1 (2001), 68–78.

A. Prékopa. 1995. *Stochastic Programming*. Kluwer Academic Publishers.

R. Sabbadin, D. Spring, and C. Rabier. 2007. Dynamic reserve site selection under contagion risk of deforestation. *Ecological Modelling* 201, 1 (2007), 75 – 81. https://doi.org/10.1016/j.ecolmodel.2006.07.036 Management, Control and Decision Making for Ecological Systems.

M. Saetersdal, J. M. Line, and H.J.B. Birks. 1993. How to maximize biological diversity in nature reserve selection: Vascular plants and breeding birds in deciduous woodlands, western Norway. *Biological Conservation* 66, 2 (1993), 131 – 138. https://doi.org/10.1016/0006-3207(93)90144-P

S. Snyder, L. Tyrell, and R. Haight. 1999. An optimization approach to selecting re- search natural areas in national forests. *Forest Science* (1999).

A. Subramanyam, C.E. Gounaris, and W. Wiesemann. 2019. K-adaptability in two-stage mixed-integer robust optimization. *Mathematical Programming* (2019).

S. F. Toth, R. G. Haight, S. A. Snyder, S. George, J. R. Miller, M. S. Gregory, and A. M. Skibbe. 2009. Reserve selection with minimum contiguous area restrictions: An application to open space protection planning in suburban Chicago. *Biological Conservation* 142, 8 (2009), 1617 – 1627. https://doi.org/10.1016/j.biocon.2009.02.037

P. Vayanos, D. Kuhn, and B. Rustem. 2011. Decision rules for information discovery in multi-stage stochastic programming. In *Proceedings of the 50th IEEE Conference on Decision and Control*. 7368–7373.

J. C. Williams and S. A. Snyder. 2005. Restoring Habitat Corridors in Fragmented Landscapes using Optimization and Percolation Models. *Environmental Modeling and Assessment* 10, 3 (2005), 239–250.