A Robust Optimization Approach for Solving Problems in Conservation Planning

Zulqarnain Haider[†], Hadi Charkhgard [*†], and Changhyun Kwon[†]

[†]Department of Industrial and Management Systems Engineering, University of South Florida

Abstract

In conservation planning, the data related to size, growth and diffusion of populations is sparse, hard to collect and unreliable at best. If and when the data is readily available, it is not of sufficient quantity to construct a probability distribution. In such a scenario, applying deterministic or stochastic approaches to the problems in conservation planning either ignores the uncertainty completely or assumes a distribution that does not accurately describe the nature of uncertainty. To overcome these drawbacks, we propose a robust optimization approach to problems in conservation planning that considers the uncertainty in data without making any assumption about its probability distribution. We explore two of the basic formulations in conservation planning related to reserve selection and invasive species control to show the value of the proposed robust optimization. Several novel techniques are developed to compare the results produced by the proposed robust optimization approach and the existing deterministic approach. For the case when the robust optimization approach fails to find a feasible solution, a novel bi-objective optimization technique is developed to handle infeasibility by modifying the level of uncertainty. Some numerical experiments are conducted to demonstrate the efficacy of our proposed approach in finding more applicable conservation planning strategies.

Keywords: conservation planning; robust optimization; invasion control; reserve selection; bi-objective mixed integer linear programming

[*]Corresponding author. E-mail address: *hcharkhgard@usf.edu*

1 Introduction

Conservation Planning concerns itself with the issues related to maintaining and increasing biodiversity. Preserving biodiversity is crucial to human societies and the future of planet Earth. Hence its slow erosion constitutes a threat as consequential as that posed by the climate change (Billionnet, 2013). According the International Union for Conservation of Nature (2017), about 24,000 species out of the 91,000 listed are threatened with extinction. Two of the key issues, among others, resulting in the loss of biodiversity, as identified by the Convention on Biological Diversity (CBD), are land fragmentation and invasive predators. The alteration and loss of the habitats for many species is caused by rampant deforestation, overpopulation, agriculture and other economically beneficial land use alternatives (Polasky et al., 2008).

There is an abundant body of knowledge prescribing the creation of land reserves, geographic regions designated for the preservation of biodiversity, as a way to slow the process of habitat destruction and to protect threatened species from the processes that threaten their existence (Rodrigues et al., 2004). Due to limited monetary and land resources available for conservation and the difficulty of reversing land use decisions in the long term, it is imperative that the reserve selection decision to be based on sound scientific evidence. There is a long history of using optimization methods for reserve selection in assistance to the process of reserve selection (Haight et al., 2000; Polasky et al., 2000; Cabeza and Moilanen, 2001; ReVelle et al., 2002; Arthur et al., 2002; Costello and Polasky, 2004). More recently there has been a growing interest in solving problems of reserve design, i.e., reserve selection with constraints on size, shape, connectivity, compactness and species complementarity (Jafari et al., 2017; Beyer et al., 2016; Haight and Snyder, 2009; Williams et al., 2005; Margules and Pressey, 2000). A brief review of the reserve selection literature and the issues therein is presented in Section 3.

Another major threat to biodiversity and other ecosystem services is the introduction of invasive species (Pejchar and Mooney, 2009). For example, Doherty et al. (2016) estimated that the invasions of mammalian species such as feral cats, rodents and pigs were responsible for massive extinctions (738 vertebrate species) and may have contributed to 58% of the cases of contemporary extinctions of birds, mammals and reptiles. Once established, it is very difficult and costly to fully eliminate an invasive specie. Many mathematical optimization formulations have been presented to manage and control the spread of invasive species . We present a brief review of these formulations in Section 3.

Conservation planning also encompasses other problems besides the two we have mentioned above. Other authors have discussed the use of mathematical optimization to solve a variety of conservation problems (Billionnet, 2013). However, one crucial aspect that has not been sufficiently considered is the issue of noisy information, for example due to imperfect detection of species during surveys(Williams et al., 2005). In their seminal work on systematic conservation planning, Margules and Pressey (2000) point out that conservation planning is riddled with uncertainty. Uncertainty pervades the use of biodiversity surrogates, the setting of conservation targets, decisions about which kinds of land tenure can be expected to contribute to targets and for which features, and decisions about how best to locate, design, implement and manage new conservation areas in the face of limited resources, competition for other uses, and incursions from surrounding areas. New developments in all the planning stages will progressively reduce, but never eliminate, these uncertainties. They recommend that planners, rather than proceeding as if certain, must learn to deal explicitly with uncertainty in ways that minimize the chances of serious mistakes.

Many problems in conservation planning require information about state variables (e.g., species abundance, occupancy), rates that pertain to the dynamic of ecological systems (e.g., population growth rate, movement rate), or conservation value of land parcels among other variables (Williams et al., 2005). Ignoring these potential sources of uncertainties may lead to bad decisions. Many studies have addressed these uncertainties with probabilistic and stochastic approaches. These approaches, although a big step up on the deterministic models, do not handle the uncertainty sufficiently. This is due to the fact that there are always certain inhibiting assumptions regarding the nature of the uncertainty in these methods. More precisely, due to sparsity of the data available, it is overly optimistic to try and over fit this data into certain probability distributions.

To deal with the issue of uncertainty and the lack of sufficient probabilistic information, there has long been a discussion of using robust optimization (see, for instance, Beyer et al. 2016). But we were not able to find any study that exploits this technique. In this paper, we propose to use robust optimization for conservation planning and optimal control of invasive species.

Since robust optimization by Bertsimas and Sim (2004) accounts for the wort case scenarios, it ensures that the problem is tractable and near optimal in the face of large uncertainty. When using the robust approach, the decision maker will know the quantum of parametric uncertainty they are protected against when they deploy the decisions and policies recommended by the robust counterpart of a formulation. In this paper, we also show another crucial value of the robust optimization. For some conservation problems, if the uncertainty is very large it may be infeasible to find a solution that meet a budget constraint. We have developed a bi-objective optimization approach that addresses this problem. Our approach gives managers the possibility to visualize how much uncertainty can be handled for a given budget. As we show in subsequent sections, this knowledge can have profound policy implications. We come up with a novel bi-objective optimization formulation to model this approach and develope it further.

This paper is organized as follows: In Section 2, we describe the robust optimization approach that we have used. In Section 3, we review existing basic optimization formulations developed for two fundamental problems in conservation planning. In Sections 4 and 5, we introduce a robust optimization approach for the invasive control problem and the reserve selection problem, respectively, and present some numerical experiments. Finally, in Section 6, we state our concluding remarks.

2 Preliminaries

2.1 Robust optimization

Robust optimization is a principal method to address data uncertainty in mathematical programming formulations. This method has been successfully applied to solve many problems (under uncertainty) when the exact distribution for the data is unknown or difficult to determine or otherwise using stochastic optimization techniques is computationally impractical. In general, robust optimization is a conservative approach that seeks to protect the decision maker against the worst realizations of outcomes. The focus of this study is the robust optimization technique developed by

Bertsimas and Sim (2004) since it allows for controlling the degree of conservatism of the solution.

Let $\boldsymbol{c}$ be an n-vector, $\boldsymbol{A}$ be an $m \times n$ matrix, and $\boldsymbol{b}$ be an m-vector. The deterministic optimization formulations in this study are in the form of mixed integer linear programs, i.e.,

$$\begin{aligned} &\min \ \boldsymbol{cx} \\ &\text{s.t.} \ \ \boldsymbol{Ax} \leq \boldsymbol{b} \\ &\qquad \boldsymbol{x} \geq \boldsymbol{0} \\ &\qquad x_i \in \mathbb{Z} \qquad\qquad\qquad \text{for } i = 1, \dots, n_1, \end{aligned}$$

where $\boldsymbol{x}$ is the vector of variables containing n_1 number of integer variables, and n_2 number of continuous variables (note that $n = n_1 + n_2$). Also, all coefficients are rational, i.e., $\boldsymbol{A} \in \mathbb{Q}^{m \times n}$, $\boldsymbol{b} \in \mathbb{Q}^m$, and $\boldsymbol{c} \in \mathbb{Q}^n$. In all proposed formulations in this study, the data uncertainty effects only the elements of the matrix $\boldsymbol{A}$. To avoid any unnecessary confusion, we next explain a customized version of the robust optimization technique developed by Bertsimas and Sim (2004) that works on this specific class of optimization problems.

We do not make any assumption about the exact distribution of each entry a_{ij} of the matrix $\boldsymbol{A}$. However, it is assumed that reasonable estimates for the mean value of the coefficient $\overline{a}_{ij}$ and its range $\widehat{a}_{ij}$ are available. In other words, we assume that each entry a_{ij} takes value in $[\overline{a}_{ij} - \widehat{a}_{ij}, \overline{a}_{ij} + \widehat{a}_{ij}]$. Note that $\widehat{a}_{ij}$ can be equal to 0.

For each row $i \in \{1, \dots, m\}$ of the matrix $\boldsymbol{A}$, we introduce a number Γ_i (defined by users) to adjust the the required level of conservatism in the proposed robust optimization formulation. This number simply imposes an upper bound on the number of entries of row i of the matrix $\boldsymbol{A}$ that can reach their worst-case values. Given that $\boldsymbol{Ax} \leq \boldsymbol{b}$ and all variables are non-negative, the worst-case value for the entry a_{ij} of the matrix $\boldsymbol{A}$ is $\overline{a}_{ij} + \widehat{a}_{ij}$. So, higher the value of Γ_i, higher the degree of conservatism. Γ_i can only take values in the interval $[0, |J_i|]$ where $J_i = \{j : \widehat{a}_{ij} > 0\}$. We assume that if $\Gamma_i \notin \mathbb{Z}$ then at most $\lfloor \Gamma_i \rfloor$ number of entries of row i of the matrix $\boldsymbol{A}$ can reach their worst-case values, i.e., $\overline{a}_{ij} + \widehat{a}_{ij}$. One other entry r_i can reach the value of $\overline{a}_{ij} + (\Gamma_i - \lfloor \Gamma_i \rfloor)\widehat{a}_{ij}$. The robust optimization formulation that seeks to conduct the optimization against the worst-case scenario under these stated assumptions can be presented as follows:

$$\begin{aligned}
&\min \boldsymbol{cx} \\
&\text{s.t. } \sum_{j=1}^{n} \overline{a}_{ij} x_j + \max_{\{S_i \cup \{r_i\}:\ S_i \subseteq J_i,\ |S_i| \leq \lfloor \Gamma_i \rfloor,\ r_i \in J_i \setminus S_i\}} \Big\{ \sum_{j \in S_i} \widehat{a}_{ij} x_j + (\Gamma_i - \lfloor \Gamma_i \rfloor) \widehat{a}_{ir_i} x_{r_i} \Big\} \leq b_i \quad \text{for } i = 1, \ldots, m \\
&\quad \boldsymbol{x} \geq \boldsymbol{0} \\
&\quad x_i \in \mathbb{Z} \qquad \text{for } i = 1, \ldots, n_1.
\end{aligned}$$

It can be shown that this formulation has the following equivalent mixed integer linear programming formulation (Bertsimas and Sim, 2004):

$$\begin{aligned}
&\min \boldsymbol{cx} \\
&\text{s.t. } \sum_{j=1}^{n} \overline{a}_{ij} x_j + z_i \Gamma_i + \sum_{j \in J_i} p_{ij} \leq b_i && \text{for } i = 1, \ldots, m \\
&\quad z_i + p_{ij} \geq \widehat{a}_{ij} x_j && \text{for } i = 1, \ldots, m \text{ and } j \in J_i \\
&\quad p_{ij} \geq 0 && \text{for } i = 1, \ldots, m \text{ and } j \in J_i \\
&\quad z_i \geq 0 && \text{for } i = 1, \ldots, m \\
&\quad \boldsymbol{x} \geq \boldsymbol{0} \\
&\quad x_i \in \mathbb{Z} && \text{for } i = 1, \ldots, n_1.
\end{aligned}$$

In this study, we call this formulation the *robust counterpart formulation*.

2.2 Bi-objective mixed integer linear programming

Many real-world problems involve multiple objectives. In many practical cases, objectives are conflicting and hence finding a feasible solution that simultaneously optimizes all of them is usually impossible. Consequently, in practice, decision makers want to understand the trade off between these objectives before choosing a suitable solution. Here, we introduce a specific class of multi-objective optimization problems that we used in this study.

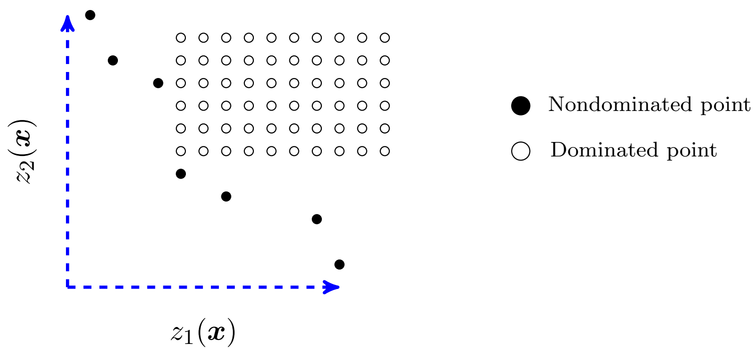

Figure 1: The nondominated frontier.

Let $\boldsymbol{c}^1$ and $\boldsymbol{c}^2$ be n-vectors. A *Bi-Objective Mixed Integer Linear Program* (BOMILP) can be stated as follows:

$$\begin{aligned} &\min \; \{z_1(\boldsymbol{x}) := \boldsymbol{c}^1\boldsymbol{x}, z_2(\boldsymbol{x}) := \boldsymbol{c}^2\boldsymbol{x}\} \\ &\text{s.t. } \; \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b} \\ &\qquad \boldsymbol{x} \geq \boldsymbol{0} \\ &\qquad x_i \in \mathbb{Z} \qquad\qquad \text{for } i = 1, \dots, n_1, \end{aligned}$$

where $\mathcal{X} := \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{b},\ \boldsymbol{x} \geq \boldsymbol{0},\ x_i \in \mathbb{Z} \text{ for } i = 1, \dots, n_1\}$ represents the *feasible set in the decision space*, and the image $\mathcal{Y}$ of $\mathcal{X}$ under vector-valued function $\boldsymbol{z} = (z_1, z_2)$ represents the *feasible set in the objective/criterion space*, i.e., $\mathcal{Y} := \{\boldsymbol{y} \in \mathbb{R}^2 : \boldsymbol{y} = \boldsymbol{z}(\boldsymbol{x}_1, \boldsymbol{x}_2) \text{ for some } (\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{X}\}$. It is assumed that $\mathcal{X}$ is *bounded*, and all coefficients are rational.

We now introduce a fundamental concept. A feasible solution $\boldsymbol{x} \in \mathcal{X}$ is called *efficient* or *Pareto optimal*, if there is no other feasible solution $\boldsymbol{x}' \in \mathcal{X}$ such that $z_1(\boldsymbol{x}') \leq z_1(\boldsymbol{x})$ and $z_2(\boldsymbol{x}') < z_2(\boldsymbol{x})$ or $z_1(\boldsymbol{x}') < z_1(\boldsymbol{x})$ and $z_2(\boldsymbol{x}') \leq z_2(\boldsymbol{x})$. If $\boldsymbol{x}$ is efficient, then image of this solution in the criterion space, i.e., $\boldsymbol{z}(\boldsymbol{x})$, is called a *nondominated point*.

Generating the set of all nondominated points, the so-called *nondominated frontier* is the primary goal in bi-objective optimization. An illustration of a typical nondominated frontier can be found in Figure 1.

It is well known that if at most one of the objective functions contains continuous decision variables, then the set of nondominated points of a BOMILP is finite (Stidsen et al., 2014). In such a case, well-known algorithms like the *ϵ-constraint method* can be used for computing the nondominated frontier of the problem (Chankong and Haimes, 1983; Boland et al., 2015). The bi-objective optimization problem developed in this study has the mentioned property, and hence whenever appropriate, we employ the *ϵ-constraint method* to solve it.

3 Optimization Models in Conservation Planning

As mentioned in the introduction, optimization methods have been used in conservation problems including reserve selection, reserve design, landscape fragmentation, forest management, control of invasive species, protection of genetic diversity and wildfire control (Billionnet, 2013). We illustrate the benefits of robust optimization methods for conservation with the two types of problems that we introduced earlier: reserve selection and control of invasive species.

3.1 Control of Invasive Species

Spatial and Temporal control of invasive species is an important problem in conservation planning. The simplest deterministic formulation for containing the spread of an alien invader was presented by Hof (1998). They divide the land under consideration into M identical square parcels. The invading species grows by a constant rate g every time period. Once a control action is implemented in a parcel i, the invader is supposed to be completely eliminated. There is also some diffusion or spread of the species to parcel i from parcel j. Table 1 shows the mathematical notation used in this basic deterministic formulation.

The formulation presented by Hof (1998) can be represented as follows:

$$\text{(D1)} \qquad \min \sum_{i=1}^{M}\sum_{t=1}^{T} v_{it} \tag{1}$$

$$\sum_{i=1}^{M} x_{it} \leq U \qquad \text{for } i = 1, \ldots, T \tag{2}$$

$$v_{i0} = a_i \qquad \text{for } i = 1, \ldots, M \tag{3}$$

Table 1: Mathematical notation used in the basic formulation for invasive species control

Variables	
v_{it}	A non-negative variable that captures the population size of invasive species in parcel i at the beginning of time period t
x_{it}	A binary variable that is equal to 1 if parcel i is treated in time period t
Parameters	
T	The number of time periods in the planning horizon
M	The number of parcels in the land under consideration
p_{ji}	The proportion of population of parcel j that diffuses into parcel i between periods t and t+1
b_{it}	A sufficiently large value that provides an upper bound for the population of the invasive species in parcel i
g	The growth rate of the invasive species at any time period
U	The maximum number of parcels that can be treated at any time period
a_i	The initial population of invasive species in parcel i at the beginning of time period 0

$$v_{it} + b_{it} \sum_{t'=1}^{t} x_{ik} \geq \sum_{j=1}^{M} p_{ji}(1+g)v_{jt-1} \qquad \text{for } i = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{4}$$

$$v_{it} \geq 0 \qquad \text{for } i = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{5}$$

$$x_{it} \geq \{0, 1\} \qquad \text{for } i = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{6}$$

The objective function minimizes the population size of the invasive species on all parcels during a planning horizon with T time steps. Constraint (2) guarantees that the number of parcels treated in each time period is not greater than the imposed upper bound, i.e., U. Finally, Constraints (3) and (4) simply capture the population size of invasive species in each parcel at the beginning of each time period. We assume that b_{it} is sufficiently large for $i = 1, \ldots, M$ and $t = 1, \ldots, T$, i.e., regardless of the value of v_{jt-1} for $j = 1, \ldots, M$ we must have that $b_{it} \geq \sum_{j=1}^{M} p_{ji}(1+g)v_{jt-1}$. Therefore, the value of b_{it} can be computed recursively by using $b_{it} = \sum_{j=1}^{M} p_{ji}(1+g)b_{jt-1}$ and $b_{i0} = a_i$. In light of this observation, the objective function and constraint (4) together imply that $v_{it} = 0$ if parcel i is treated in or before time period t and $v_{it} = \sum_{j=1}^{M} p_{ji}(1+g)v_{jt-1}$ otherwise. Hence, implicitly, treating a parcel more than once is unnecessary in this formulation. In practice it is computationally advantageous to add the following valid inequalities to the formulation,

$$\sum_{t=1}^{T} x_{it} \leq 1 \qquad \text{for } i = 1, \ldots, M \tag{7}$$

Note that these valid inequalities are not part of the original formulation introduced by Hof (1998). However, we decide to use them since in practice we have observed that they can reduce the solution time by a factor of around 2. Next, we briefly review some of the more advanced formulations.

Büyüktahtakın et al. (2011) present a dynamic binary nonlinear deterministic formulation for invasion control. Their objective function incorporates a damage function that minimizes the total damage to some resources in all parcels of land in all time periods. They also use a weight parameter to represent the relative importance of a resource. In their case they consider a logistic growth for the population of invasives. Their formulation also considers different kill rates depending on the critical population density in a parcel. The budget and labor constraints in every time period are also considered. According to the authors, the large size of the dynamic problem and the nonlinearity make the application of direct optimization methods impossible. Instead, they analyze and compare the most frequently suggested strategies and their consequences.

Epanchin-Niell and Wilen (2012) present a two dimensional, temporally and spatially explicit formulation of the spread of invasive species. They then present a spatially explicit MIP formulation for optimal control of the invasion. They minimize the present value of the sum of control costs and invasion damages across space and time. They incorporate two types of invasion control: clearing of invaded patches and preventing spread from invaded to non-invaded patches. They do not consider any growth formulation for within patch growth.

Büyüktahtakın et al. (2014) consider a multiobjective optimization approach for invasion control. Their formulation is an extension of Büyüktahtakın et al. (2011) in which they consider separate objective functions for each resource under threat due to invasion. They find out a vector of the best possible objective function values and try to minimize the "distance" of a Pareto optimal vector of objective functions from the above vector.

Büyüktahtakın et al. (2015) consider an age structured bio-economic formulation of invasive species control. They explicitly take into account the biological dynamics of a population and try to minimize the damage done by invasive species throughout the time horizon.

Another simple linear programming based approach to invasion control is proposed by Hastings et al. (2006). They assume that at initial stages, the growth is density independent and hence linear and that physical removal rather than a biological action that stunts the growth of the species be used as a control strategy. They consider a linear age or stage structured population of the invader. At different stages (age), the invader has different growth, survival rates, and fecundity.

For an exhaustive review of the work that considers uncertainty in invasion control we refer the readers to the review paper by Epanchin-Niell and Hastings (2010). We will mainly concern ourselves with how the current body of work in invasion control tries to handle uncertainty. In summary, the formulations with uncertainty fall into two categories: Models that consider stochasticity in some parameters and formulations that consider lack of information about some parameters. The point of introduction, growth, diffusion, response to control and the harmful impacts of an invasion can all be stochastic processes. The data on these process for a new invader can be sparse and hence it is often difficult to construct representative probability distributions. The optimization formulations also contain parameters whose values for a particular application are simply not known and their estimation remains difficult especially before or early in an invasion. Even during the invasion, accurate detection and measurement of invasions is difficult especially for animal species like the Burmese Pythons in South Florida (Bonneau et al., 2016). These multiple uncertainties are an impediment to devising optimal control policies or evaluating their impacts. The approaches that have been tried so far include information gap theory, learning models, partially observable Markov decision processes and Bayesian analysis (For detailed references see Epanchin-Niell and Hastings, 2010).

More recently, Haight and Polasky (2010) used partially observable Markov decision process for controlling an invasive species with imperfect information about the level of infestation. The formulation they describe considers multiple, mutually exclusive, management actions and multiple levels of infestation. Different states represent different levels of infestation. In the absence of perfect information, the decision maker only has a vector of probabilities about the level of infestation in a time period. At the beginning of each period, the manager's problem is to determine the best action given the set of beliefs about the infestation state.

Kotani et al. (2011) also consider a stochastic dynamic programming formulation where they consider growth and measurement uncertainties. They describe these uncertainties as random variables with a certain distribution. The random variable associated with growth uncertainty in a period represents uncontrollable stochasticity while the one associated with measurement uncertainty reflects potentially controllable uncertainty. The authors show that these uncertainties could significantly alter not only qualitative features of optimal policies, but also their value functions.

3.2 Reserve Selection Problem

The reserve selection problem is another well studied problem in the conservation planning literature. The problem consists in selecting a proportion of a given geographic area for the purposes of conserving a given species or a set of species. It is oftentimes prohibitive to earmark a very large geographic area for species conservation because of opportunity costs associated with alternative, high economic value land use. Multiple variations on the basic reserve selection problem have been presented over the years. For a more exhaustive review of the science of reserve selection, readers can refer to Williams et al. (2005), Haight and Snyder (2009), Billionnet (2013), and Beyer et al. (2016).

In our paper, we use the basic reserve selection formulation presented in Beyer et al. (2016). There is a cost associated with the selection of each reserve, and each reserve contributes to the conservation of species of interest. The optimization procedure selects parcels of land so as to minimize the cost while achieving some explicit target conservation values for each species.

Table 2 details the mathematical notation used in the basic deterministic reserve selection formulation. We divided the land under consideration into M parcels. If the conservation targets for species $k \in \{1, \ldots, K\}$ were achieved, the parcels were assumed to be conserved.

The formulation presented by Beyer et al. (2016) can be represented as follows:

Table 2: Mathematical notation used in the basic formulation for reserve selection

Variables	
x_i	A binary variable that is equal to 1 if parcel i is selected as reserve for species conservation
Parameters	
K	The number of species under consideration
M	The number of parcels in the land under consideration
c_i	The cost of selecting parcel i as a reserve
w_{ik}	The conservation value of parcel i for species k
W_k	The target conservation value that must be achieved for species k

$$\text{(D2)} \qquad \min \sum_{i=1}^{M} c_i x_i \tag{8}$$

$$\sum_{i=1}^{M} w_{ik} x_i \geq W_k \qquad \text{for } k = 1, \ldots, K \tag{9}$$

$$x_i \in \{0, 1\} \qquad \text{for } i = 1, \ldots, M \tag{10}$$

The objective function minimizes the total cost of conservation. Constraint (9) ensures that the target conservation value for each specie is achieved. Next, we explain some fundamental features of the reserve selection problem and the existing formulations for this problem in the literature.

- In all existing formulations, the area under consideration is divided into parcels of land. One or more species can be considered for conservation. The decision of reserve selection can be single period or a multi-period dynamic decision (Costello and Polasky, 2004; Snyder et al., 2005; Tóth et al., 2011; Strange et al., 2006).

- In the basic deterministic version of the problem, there is a cost of selecting a parcel of land. More elaborate approaches to determine cost consider sophisticated economic models to get a more complete picture of these costs (Polasky et al., 2008; Tóth et al., 2011). Besides the cost or the number of reserves, other objectives like species or genetic diversity can also be considered (Cabeza and Moilanen, 2001). Some researchers have used multiobjective optimization to handle more than one objective functions together (Memtsas, 2003; Snyder et al., 2004).

- All existing formulations contain parameters related to the target value of conservation to be achieved for each species under consideration and the contribution of each parcel to the the species' conservation. In the more basic formulations, contribution was modeled through binary parameters representing presence/absence of the species for each parcel and the target was to make sure that each species is represented in the optimal choice of parcels (ReVelle et al., 2002). Other more advanced formulations use some geographical, ecological or biological surrogates and/or some survey or sightings data alongside statistical modeling (Margules and Pressey, 2000; Austin, 2002) to come up with estimates of spatial species distribution.

- Almost all formulations consider constraints to ensure some sort of spatial arrangement of parcels. This is done to achieve connectivity, contiguity, compactness, shape, size or certain boundary or buffer zone requirements for the reserve (Williams et al., 2005; Westphal et al., 2007; Jafari and Hearne, 2013; Wang and Önal, 2015; Beyer et al., 2016).

- Many existing formulations consider uncertainty related to one or more of the parameters described above. Probabilistic reserve selection formulations assign probabilities to species presence rather than using a binary variable (for presence (1) and absence (0)). These formulations either maximize the expected coverage (expected coverage approach) (Polasky et al., 2000) or the number of species covered, where a species is considered covered if its cumulative presence probability exceeds a predetermined threshold (threshold approach). There are no formulations, however, that consider the parametric uncertainty related to value of each parcel to each specie or the cost of acquiring a parcel of land in the formulation we have presented (Haight et al., 2000; Arthur et al., 2002).

Some recent review papers, for instance Billionnet (2013) and Beyer et al. (2016), emphasize the necessity of dealing with uncertainty related to these parameters, they recommend robust optimization approaches to account for this parametric uncertainty. Despite its great potential, robust optimization has been under utilized to address problems of conservation and natural resource management.

4 A Robust Optimization Approach for the Invasion Control Problem

In this section, we consider the basic deterministic formulation by Hof (1998), i.e., (D1), and incorporate uncertainty in some parameters through the robust optimization approach presented in Section 2.1. More precisely, we assume that all parameters of (D1) are known with certainty except the diffusion rate p_{ji} for $i = 1, \ldots, M$ and $j = 1, \ldots, M$. We assume that $p_{ji} \in [\overline{p}_{ji} - \widehat{p}_{ji}, \overline{p}_{ji} + \widehat{p}_{ji}]$ and $\widehat{p}_{ji} > 0$ for all $i = 1, \ldots, M$ and $j = 1, \ldots, M$.

We denote the robust counterpart formulation of (D1) with (R1). This formulation can be easily constructed using the techniques developed in Section 2.1. Interested readers can find (R1) in Appendix A.

4.1 Numerical Experiments

To compare the performance of (D1) and (R1), a random instance is generated by setting $M = 40$, $T = 5$, $U = 2$ and $r = 0.05$ in this section. We randomly chose a geographical region in central Florida and divide it into M equal parcels. The value of a_i for each $i \in \{1, \ldots, M\}$ is generated randomly from $(-4, 4)$. We set negative values to zero. The value of diffusion rate p_{ji} is considered to be inversely proportional to the square of Euclidean distance between the parcels for each $j \in \{1, \ldots, M\}$ and $i \in \{1, \ldots, M\}$, i.e., $\overline{p}_{ji} = 1/(d_{ji})^2$. Let $\alpha, \beta \geq 0$ be two user-defined parameters, we assume that $\widehat{p}_{ji} = \beta\overline{p}_{ji}$ and $\Gamma_i^t = \alpha M$ for $i = 1, \ldots, M$, $j = 1, \ldots, M$, and $t = 1, \ldots, T$. It is worth mentioning that to compare (R1) and (D1), we assume that $p_{ji} = \overline{p}_{ji}$ in (D1) for $i = 1, \ldots, M$ and $j = 1, \ldots, M$. Next, we conduct a set of experiments on the generated instance by choosing different values for α and β.

We first note that solving (D1) and (R1) usually results in completely different solutions and objective values. The objective function for both formulations is the total presence of the invasive

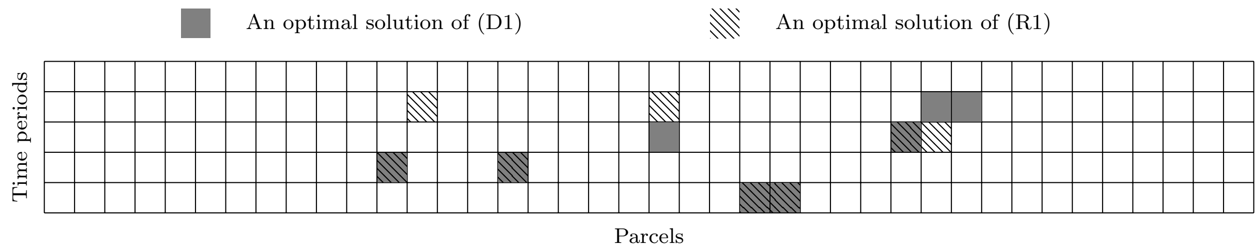

Figure 2: Optimal values of x_{it} for $i = 1, \ldots, M$ and $t = 1, \ldots, T$ produced by (D1) and (R1) for an instance of invasion control problem.

species over all parcels of land and all time periods. However, the robust counterpart formulation (R1), cannot possibly achieve an objective value less than that of the deterministic formulation (D1) due to increased diffusion rates caused by considering uncertainty. Observe that the main decision variables of the problem are x_{it} for $i = 1, \ldots, M$ and $t = 1, \ldots, T$ and they indicate whether a parcel i is selected for treatment in a time period t or not. We can easily plot the optimal values of these variables produced by (D1) and (R1). An illustration of this observation when $\alpha = 0.3$ and $\beta = 0.05$ can be found in Figure 2 in which each cell represents x_{it}, and it is hatched/filled only if $x_{it} = 1$. As expected, the recommended values generated by (D1) and (R1) are different.

Since optimal solutions of (R1) and (D1) may differ significantly, and since their respective objective function values cannot be fairly compared due to increased diffusion rates in (R1), we introduce a few techniques to be able to compare the solutions and show the value of robust optimization. Since the main decision variables of the problem are x_{it} for $i = 1, \ldots, M$ and $t = 1, \ldots, T$, let $\boldsymbol{x}^d$ and $\boldsymbol{x}^r$ be the optimal values of these decision variables produced by solving (D1) and (R1), respectively. Furthermore, for a given $\overline{\boldsymbol{x}}$, we define $D(\overline{\boldsymbol{x}})$ as the optimal objective value of (D1) when we set $x_{it} = \overline{x}_{it}$ for $i = 1, \ldots, M$ and $t = 1, \ldots, T$ in (D1). Similarly, for a given $\overline{\boldsymbol{x}}$, we define $R(\overline{\boldsymbol{x}})$ as the optimal objective value of (R1) when we set $x_{it} = \overline{x}_{it}$ for $i = 1, \ldots, M$ and $t = 1, \ldots, T$ in (R1). Using these definitions, four values can be computed:

- $D(\boldsymbol{x}^d)$: This is the the objective value that is reached when we choose to implement the deterministic decision $\boldsymbol{x}^d$ made on assumptions of zero uncertainty. Luckily, in reality, there is no uncertainty and hence the worst case scenario (as defined by the robust optimization formulation) does not arise.

- $D(\boldsymbol{x}^r)$: This is the objective value that is reached if we use $\boldsymbol{x}^r$ i.e., when the original decision is made on assumptions of worst case scenario. And luckily, it later turns out that the worst case scenario (as defined by the robust optimization formulation) does not arise.

- $R(\boldsymbol{x}^r)$: This is the objective value we get if we choose to implement $\boldsymbol{x}^r$, the decision made to protect against the worst case. And it later turns out that, true to our assumptions, uncertainty was in fact present and hence the worst case scenario (as defined by the robust optimization formulation) does come about.

- $R(\boldsymbol{x}^d)$: This is the objective value we achieve if we use $\boldsymbol{x}^d$ i.e., the decision made on assumptions of zero uncertainty. However, unlucky for us, the worst case scenario (as defined by the robust optimization formulation) does arise in reality.

If these values exist then it is easy to show that $D(\boldsymbol{x}^d) \leq D(\boldsymbol{x}^r) \leq R(\boldsymbol{x}^r) \leq R(\boldsymbol{x}^d)$. In other words, using $\boldsymbol{x}^r$ results in less fluctuations in the objective value if realizations of data different than those anticipated arise in practice. To illustrate this observation, we assume that $\beta \in \{0.5, 1, 2\}$ and $\alpha \in \{0.25, 0.5, 0.75\}$, and run 9 experiments by applying different combinations for values of α and β. Experiments $1, 2, \ldots, 9$ are defined to be precisely $(\beta = 0.5, \alpha = 0.25), (\beta = 0.5, \alpha = 0.5), \ldots, (\beta = 2, \alpha = 0.75)$. The scaled results are reported in Figure 3 in which the vertical axis shows the ratio, i.e., $\frac{D(\boldsymbol{x}^d)}{R(\boldsymbol{x}^d)}$, $\frac{D(\boldsymbol{x}^r)}{R(\boldsymbol{x}^d)}$, $\frac{R(\boldsymbol{x}^r)}{R(\boldsymbol{x}^d)}$, and $\frac{R(\boldsymbol{x}^d)}{R(\boldsymbol{x}^d)}$, and the horizontal axis shows the experiment number. Observe that $\frac{D(\boldsymbol{x}^d)}{R(\boldsymbol{x}^d)} \approx \frac{D(\boldsymbol{x}^r)}{R(\boldsymbol{x}^d)}$ (in fact $\frac{D(\boldsymbol{x}^d)}{R(\boldsymbol{x}^d)}$ is slightly better/smaller than $\frac{D(\boldsymbol{x}^r)}{R(\boldsymbol{x}^d)}$ in all experiments). This implies that $\boldsymbol{x}^r$ is almost optimal for (D1). So, if we use $\boldsymbol{x}^r$, and we are lucky in a sense that the worse scenario (as defined by the robust optimization formulation) does not arise in reality, we have almost lost nothing. Also observe that $\frac{R(\boldsymbol{x}^r)}{R(\boldsymbol{x}^d)}$ is up to 14% better/smaller than $\frac{R(\boldsymbol{x}^d)}{R(\boldsymbol{x}^d)}$. This implies that if the worse case scenario (as defined by the robust optimization formulation) arises then we are up to 14% better off by using $\boldsymbol{x}^r$. These two observations clearly illustrate the value of the proposed robust optimization, and the fact that $\boldsymbol{x}^r$ is a better choice in practice. Of course, as enunciated earlier, the assumption of uncertainty and the robust formulation used to handle it, also impacts the gist of managerial decision making by prescribing to treat different parcels of land in different time periods as compared with those suggested by the deterministic

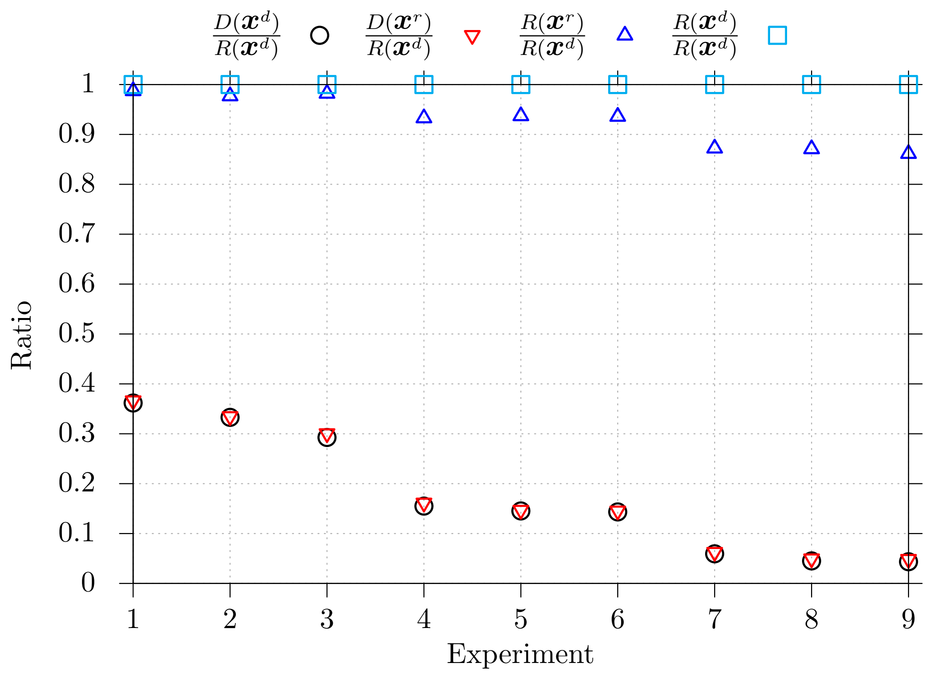

Figure 3: Comparing the optimal solution generated by (D1) and (R1) in 9 different experiments.

formulation.

5 A Robust Optimization Approach to the Reserve Selection Problem

In this section, the robust counterpart formulation of the basic reserve selection problem (Beyer et al., 2016), explained in Section 5, is developed. We assume that the cost coefficients for parcels, i.e., c_i for $i = 1, \dots, M$, and the target values for species, W_k for $k = 1, \dots, K$, are known with certainty. Note that this assumption may often be reasonable because the cost can be approximated using a range of values from alternative economic models, and the species conservation target can be elicited from stakeholders. w_{ik}, represent the contribution of parcel i to species k and is considered to be uncertain for $i = 1, \dots, M$ and $k = 1, \dots, K$. These values often depend on a number of biological factors and therefore need thorough examination of the ecological characteristics of a certain parcel of land. The estimates for these biological factors are prone to estimation errors and changes over time.

More specifically, we assume that $w_{ik} \in [\overline{w}_{ik} - \widehat{w}_{ik}, \overline{w}_{ik} + \widehat{w}_{ik}]$ and $\widehat{w}_{ik} > 0$ for all $i = 1, \ldots, M$ and $k = 1, \ldots, K$. We denote the robust counterpart formulation of (D2) with (R2). This formulation can be easily constructed using the techniques developed in Section 2.1. Interested readers can find (R2) in Appendix B.

5.1 Numerical Experiments

The goal of this section is to compare the performance of (D2) and (R2). We evaluated the performance of (D2) and (R2) with simulated data. The simulated data are generated by setting $M = 40$ and $K = 5$. Also, c_i for $i = 1, \ldots, M$ are randomly generated by using the discrete uniform distribution in the interval $[100, 1000]$. Furthermore, $\overline{w}_{ik}$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$ are randomly drawn from the normal distribution with mean of 0 and standard deviation of 5. Values less than zero are truncated to zero. This implies that on average 50% of the parcels do not contribute to the conservation of a particular species. We also set $W_k = 0.5 \sum_{i=1}^{M} \overline{w}_{ik}$ for $k = 1, \ldots, K$. Let $\alpha, \beta \geq 0$ be two user-defined parameters, we assume that $\widehat{w}_{ik} = \beta \overline{w}_{ik}$ and $\Gamma_k = \alpha M$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$. It is worth mentioning that to compare (R2) and (D2), we assume that $w_{ik} = \overline{w}_{ik}$ in (D2) for $i = 1, \ldots, M$, and $k = 1, \ldots, K$. Next, we conduct some experiments on the simulated data by choosing different values for α and β.

In rest of this section, we assume that (D2) is always feasible, but we make no such assumption about (R2). We first note that to compare (D2) and (R2), we cannot use the same technique developed in Section 4.1 because this problem has only one type of decision variables, and so the solution corresponding to (D2) is unlikely to be feasible for (R2). This implies that using the solution of the deterministic formulation would be a poor choice because under the worst case scenario (as defined by the robust optimization formulation), the deterministic solution fails to satisfy the target values for species of interest.

Based on this observation, one may be tempted to make the determination to always use the robust optimization formulation (R2), because it is always feasible. However, as we subsequently explain, that is not necessarily the case. In general, the structure of (R2) is such that it is quite possible to increase the degree of uncertainty in (R2) to such an extent so as to render the problem

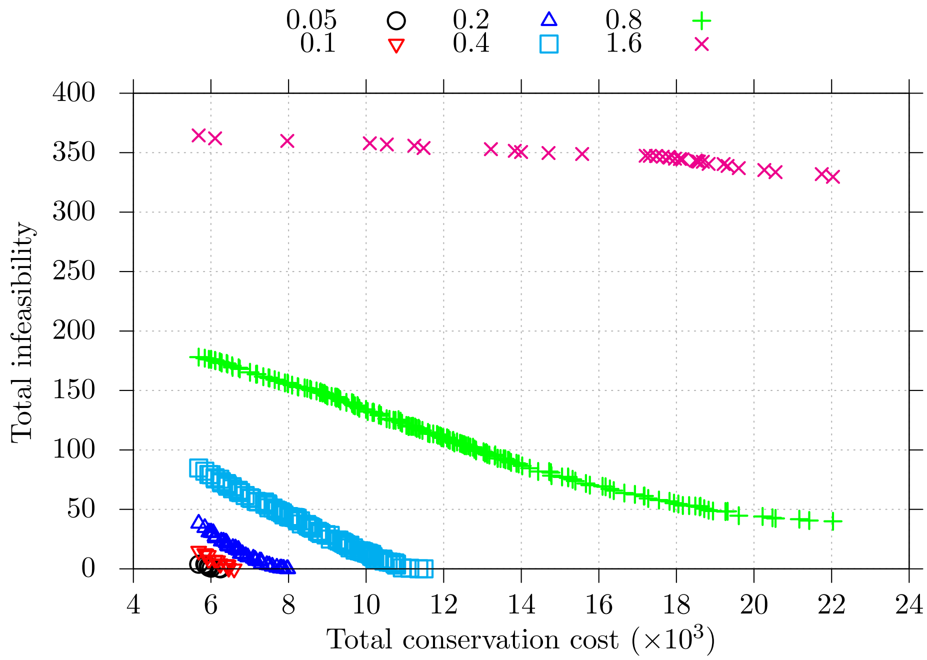

Figure 4: The nondominated frontier of (R3) for different values of β when $\alpha = 0.25$. Each curve represents the non dominated frontier for a different value of β. The axes represent the values of the respective objective functions.

infeasible.

Although, using the optimal solution obtained by (D2) does not seem to be a good choice. At the same time, however, employing (R2) as currently defined may not be a good option either as the issue of possible infeasibility persists. Consequently, we need to revise (R2) in order to ensure that the revised formulation is always feasible (if (D2) is feasible).

We revised (R2) by adding a secondary objective to the problem (in addition to the total conservation cost) to measure the total infeasibility of the robust formulation. It is worth mentioning that the total infeasibility can be interpreted as the amount of decrease in the uncertainty ranges in order to make the robust counterpart formulation feasible. We denote the revised formulation by (R3). It is worth mentioning that (R3) can be easily written as a BOMILP (with a finite number of nondominated points) by using a few linearization techniques. Interested readers can find (R3) and its linearization process in Appendix C.

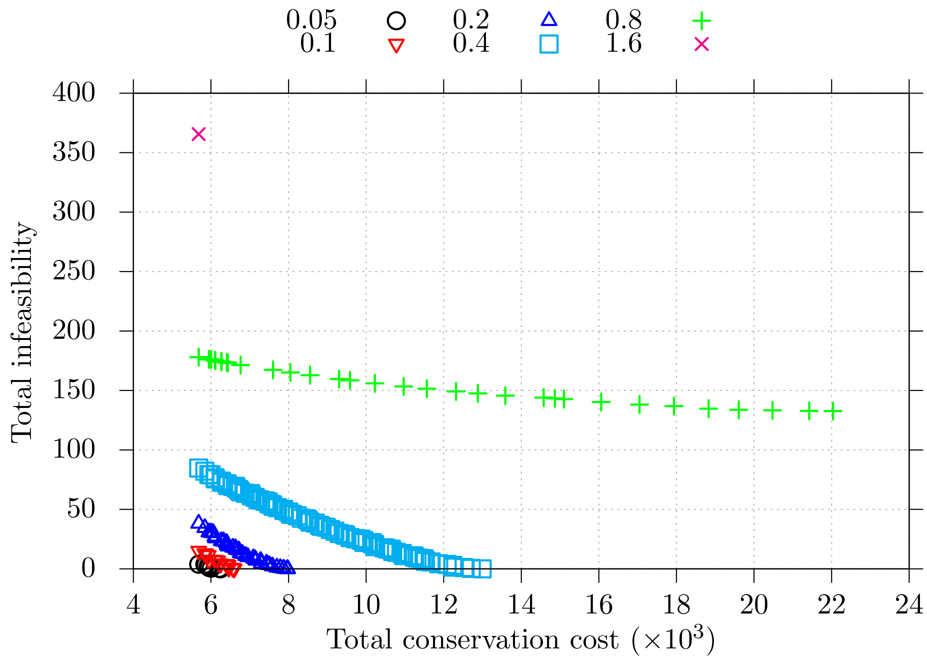

Figure 5: The nondominated frontier of (R3) for different values of β when $\alpha = 0.5$. Each curve represents the non dominated frontier for a different value of β. The axes represent the values of the respective objective functions.

We use the ϵ-constraint method to solve the BOMILP corresponding to (R3) for

$$\beta = \{0.05, 0.1, 0.2, 0.4, 0.8, 1.6\}$$

and $\alpha \in \{0.25, 0.5\}$. Figures 4 and 5 show the (exact) nondominated frontier of the problem for different values of β when $\alpha = 0.25$ and $\alpha = 0.5$, respectively. We next make a few observations about these figures.

- Intuitively, it can be seen that the upper left and lower right endpoints of each nondominated frontier provide information about (D2) and (R2), respectively. More specifically, the upper left point shows the infeasibility of the optimal solution obtained with (D2) under the worst case scenario. It also shows the optimal cost of conservation by allowing (R3) to handle this amount of infeasibility when the worst case scenario arises. Similarly, the lower right endpoint shows the total infeasibility of (R2), and the optimal cost of conservation by allowing (R3) to handle this amount of infeasibility.

- For a fixed α and β, the decision maker(s) can visualize the nondominated frontier and choose a desirable point. Obviously, choosing a point closer to the lower right corner indicates that the decision makers adopt a more conservative or risk averse approach since the lower right point represents the status of (R2).

- For a fixed α, the nondominated frontier shifts upward, i.e., the value of infeasibility increases, as β increases, and it converges towards a plateaue. This is not surprising because as β increases, the uncertainty goes up. Indeed, increasing β adds more uncertainty to the problem. We observe that it is possible that the nondominated frontier becomes so flat that eventually reduces to a single point, see for instance $\beta = 1.6$ and $\alpha = 0.5$. Note that this is also true for the other extreme case representing very small or no uncertainty. If the uncertainty is very small then it is expected that the nondominated frontier would become so steep that it would eventually reduce to a single point.

These observations imply that (for the formulation of the reserve selection problem explored in this study), the decision makers would not gain much by trying to handle a large amount of uncertainty using the proposed robust optimization technique. At higher values of assumed uncertainty, the nondominated frontier is probably so flat that to improve the infeasibility a little bit, we need to increase the conservation cost significantly. Similarly, if the assumed uncertainty is very small the robust optimization technique is almost equivalent to the deterministic formulation. Even in this extreme case, using the robust optimization approach is not particularly helpful. Nevertheless, for a reasonable amount of uncertainty using the proposed robust optimization approach seems to be quite helpful. For instance, the nondominated frontiers of the problem when $\beta = 0.2$, $\beta = 0.4$ or $\beta = 0.8$ and when $\alpha = 0.25$ or $\alpha = 0.5$ in Figures 4 and 5 seem very promising since the lower right endpoint of these nondominated frontiers has the total infeasibility value of zero, but the upper left point has a significant total infeasibility value. Also observe that since these curves are generally steeper i.e., we have significant gains on infeasibility without losing much on the conservation cost, the decision maker can genuinely achieve a trade-off between his competing objectives by choosing a suitable point on these nondominated frontiers.

6 Conclusion

Many conservation problems involve a lot of uncertainty, which may not always be captured with probability distributions. In this study, we explored the idea of applying robust optimization techniques for solving conservation problems while accounting for high levels of uncertainty. To the best of our knowledge this is the first study in applying a robust optimization approach in conservation planning problems. We illustrated our proposed approach with two types of problems: the invasion control problem and the reserve selection problem. More importantly, we developed novel techniques to compare the results obtained by the proposed robust optimization approach and the corresponding deterministic formulation. We hope that the applicability, versatility, and performance of our approach encourages practitioners and researchers to implement it to address important issues in natural resource management and conservation.

Acknowledgement

The authors are grateful to Nahid Jafari for stimulating conversations about the use of robust optimization for conservation planning. The authors are also thankful to Julien Martin from U.S. Geological Survey for helpful discussions. The views and opinions expressed in this manuscript, however, are solely those of the authors.

References

Arthur, J. L., R. G. Haight, C. A. Montgomery, S. Polasky. 2002. Analysis of the threshold and expected coverage approaches to the probabilistic reserve site selection problem. *Environmental Modeling and Assessment* **7**(2) 81–89.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling* **157**(2) 101–118.

Bertsimas, D., M. Sim. 2004. The price of robustness. *Operations Research* **52**(1) 35–53.

Beyer, H. L., Y. Dujardin, M. E. Watts, H. P. Possingham. 2016. Solving conservation planning problems with integer linear programming. *Ecological Modelling* **328** 14–22.

Billionnet, A. 2013. Mathematical optimization ideas for biodiversity conservation. *European Journal of Operational Research* **231**(3) 514–534.

Boland, N., H. Charkhgard, M. Savelsbergh. 2015. A criterion space search algorithm for biobjective integer programming: The balanced box method. *INFORMS Journal on Computing* **27**(4) 735–754.

Bonneau, M., F. A. Johnson, C. M. Romagosa. 2016. Spatially explicit control of invasive species using a reaction–diffusion model. *Ecological Modelling* **337** 15–24.

Büyüktahtakın, İ. E., Z. Feng, G. Frisvold, F. Szidarovszky, A. Olsson. 2011. A dynamic model of controlling invasive species. *Computers & Mathematics with Applications* **62**(9) 3326–3333.

Büyüktahtakın, İ. E., Z. Feng, F. Szidarovszky. 2014. A multi-objective optimization approach for invasive species control. *Journal of the Operational Research Society* **65**(11) 1625–1635.

Büyüktahtakın, İ. E., E. Y. Kıbış, H. I. Cobuloglu, G. R. Houseman, J. T. Lampe. 2015. An age-structured bio-economic model of invasive species management: insights and strategies for optimal control. *Biological Invasions* **17**(9) 2545–2563.

Cabeza, M., A. Moilanen. 2001. Design of reserve networks and the persistence of biodiversity. *Trends in Ecology & Evolution* **16**(5) 242–248.

Chankong, V., Y. Y. Haimes. 1983. *Multiobjective Decision Making: Theory and Methodology*. Elsevier Science, New York.

Costello, C., S. Polasky. 2004. Dynamic reserve site selection. *Resource and Energy Economics* **26**(2) 157–174.

Doherty, T. S., A. S. Glen, D. G. Nimmo, E. G. Ritchie, C. R. Dickman. 2016. Invasive predators and global biodiversity loss. *Proceedings of the National Academy of Sciences* **113**(40) 11261–11265.

Epanchin-Niell, R. S., A. Hastings. 2010. Controlling established invaders: integrating economics and spread dynamics to determine optimal management. *Ecology Letters* **13**(4) 528–541.

Epanchin-Niell, R. S., J. E. Wilen. 2012. Optimal spatial control of biological invasions. *Journal of Environmental Economics and Management* **63**(2) 260–270.

Haight, R. G., S. Polasky. 2010. Optimal control of an invasive species with imperfect information about the level of infestation. *Resource and Energy Economics* **32**(4) 519–533.

Haight, R. G., C. S. Revelle, S. A. Snyder. 2000. An integer optimization approach to a probabilistic reserve site selection problem. *Operations Research* **48**(5) 697–708.

Haight, R. G., S. A. Snyder. 2009. Integer programming methods for reserve selection and design. *Spatial conservation prioritization: quantitative methods and computational tools.(A Moilanen, KA Wilson, HP Possingham, eds.). Oxford University Press, New York, NY* 43–57.

Hastings, A., R. J. Hall, C. M. Taylor. 2006. A simple approach to optimal control of invasive species. *Theoretical Population Biology* **70**(4) 431–435.

Hof, J. 1998. Optimizing spatial and dynamic population-based control strategies for invading forest pests. *Natural Resource Modeling* **11**(3) 197–216.

International Union for Conservation of Nature. 2017. The IUCN Red List of Threatened Species. URL `http://www.iucnredlist.org/about/summary-statistics`.

Jafari, N., J. Hearne. 2013. A new method to solve the fully connected reserve network design problem. *European Journal of Operational Research* **231**(1) 202–209.

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

Kotani, K., M. Kakinaka, H. Matsuda. 2011. Optimal invasive species management under multiple uncertainties. *Mathematical Biosciences* **233**(1) 32–46.

Margules, C. R., R. L. Pressey. 2000. Systematic conservation planning. *Nature* **405**(6783) 243–253.

Memtsas, D. P. 2003. Multiobjective programming methods in the reserve selection problem. *European Journal of Operational Research* **150**(3) 640–652.

Pejchar, L., H. A. Mooney. 2009. Invasive species, ecosystem services and human well-being. *Trends in Ecology & Evolution* **24**(9) 497–504.

Polasky, S., J. D. Camm, A. R. Solow, B. Csuti, D. White, R. Ding. 2000. Choosing reserve networks with incomplete species information. *Biological Conservation* **94**(1) 1–10.

Polasky, S., E. Nelson, J. Camm, B. Csuti, P. Fackler, E. Lonsdorf, C. Montgomery, D. White, J. Arthur, B. Garber-Yonts, et al. 2008. Where to put things? spatial land management to sustain biodiversity and economic returns. *Biological Conservation* **141**(6) 1505–1524.

ReVelle, C. S., J. C. Williams, J. J. Boland. 2002. Counterpart models in facility location science and reserve selection science. *Environmental Modeling and Assessment* **7**(2) 71–80.

Rodrigues, A. S., S. J. Andelman, M. I. Bakarr, L. Boitani, T. M. Brooks, R. M. Cowling, L. D. Fishpool, G. A. Da Fonseca, K. J. Gaston, M. Hoffmann, et al. 2004. Effectiveness of the global protected area network in representing species diversity. *Nature* **428**(6983) 640–643.

Snyder, S., C. ReVelle, R. Haight. 2004. One-and two-objective approaches to an area-constrained habitat reserve site selection problem. *Biological Conservation* **119**(4) 565–574.

Snyder, S. A., R. G. Haight, C. S. ReVelle. 2005. A scenario optimization model for dynamic reserve site selection. *Environmental Modeling and Assessment* **9**(3) 179–187.

Stidsen, T., K. A. Andersen, B. Dammann. 2014. A branch and bound algorithm for a class of biobjective mixed integer programs. *Management Science* **60**(4) 1009–1032.

Strange, N., B. J. Thorsen, J. Bladt. 2006. Optimal reserve selection in a dynamic world. *Biological Conservation* **131**(1) 33–41.

Tóth, S. F., R. G. Haight, L. W. Rogers. 2011. Dynamic reserve selection: Optimal land retention with land-price feedbacks. *Operations Research* **59**(5) 1059–1078.

Wang, Y., H. Önal. 2015. Optimal design of compact and connected nature reserves for multiple species. *Conservation Biology* .

Westphal, M. I., S. A. Field, H. P. Possingham. 2007. Optimizing landscape configuration: a case study of woodland birds in the mount lofty ranges, south australia. *Landscape and Urban Planning* **81**(1) 56–66.

Williams, J. C., C. S. ReVelle, S. A. Levin. 2005. Spatial attributes and reserve design models: a review. *Environmental Modeling and Assessment* **10**(3) 163–181.

Appendices

A Robust Formulation of (D1)

Based on our discussion in Section 2.1, it is easy to show that the robust counterpart formulation of (D1) can be stated as:

$$\text{(R1)} \quad \min \quad \sum_{i=1}^{M}\sum_{t=1}^{T} v_{it} \tag{11}$$

$$\sum_{i=1}^{M} x_{it} \leq U \qquad \text{for } i = 1, \ldots, T \tag{12}$$

$$v_{i0} = a_i \qquad \text{for } i = 1, \ldots, M \tag{13}$$

$$v_{it} + b_{it}\sum_{t'=1}^{t} x_{ik} \geq \sum_{j=1}^{M} \overline{p}_{ji}(1+g)v_{jt-1} + z_{it}\Gamma_i^t + \sum_{j=1}^{M} q_{jit} \qquad \text{for } i = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{14}$$

$$z_{it} + q_{jit} \geq \widehat{p}_{ji}(1+g)v_{jt-1} \qquad \text{for } i = 1, \ldots, M \text{ and } j = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{15}$$

$$q_{jit} \geq 0 \qquad \text{for } i = 1, \ldots, M \text{ and } j = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{16}$$

$$z_{it} \geq 0 \qquad \text{for } i = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{17}$$

$$v_{it} \geq 0 \qquad \text{for } i = 1, \ldots, M \text{ and } t = 1, \ldots, T \tag{18}$$

$$x_{it} \geq \{0, 1\} \qquad \text{for } i = 1, \ldots, M \text{ and } t = 1, \ldots, T, \tag{19}$$

where Γ_i^t for $i = 1, \ldots, M$ and $t = 1, \ldots, T$ is a user-defined parameter showing the level of conservatism in constraint (14).

Note that we assume that b_{it} is sufficiently large, i.e., regardless of the value of v_{jt-1}, z_{it}, and q_{ijt} for $j = 1, \ldots, M$, $j = 1, \ldots, M$ and $t = 1, \ldots, T$, we must have that

$$b_{it} \geq \sum_{j=1}^{M} \overline{p}_{ji}(1+g)v_{jt-1} + z_{it}\Gamma_i^t + \sum_{j=1}^{M} q_{jit}.$$

Therefore, the value of b_{it} should be computed differently in (R1). This can be done using the following proposition.

Proposition 1. *Let* $\boldsymbol{u_{i,t-1}} := (u^1_{i,t-1}, u^2_{i,t-1}, \ldots, u^M_{i,t-1})$ *such that* $u^j_{i,t-1} := \widehat{p}_{ji}(1+g)b_{j,t-1}$ *for* $i = 1, \ldots, M$ *and* $t = 1, \ldots, T$*. Also, let* $u^{(j)}_{i,t-1}$ *be the* j*-th largest component of* $\boldsymbol{u_{i,t-1}}$*. For each,* $i \in \{1, \ldots, M\}$ *and* $t \in \{1, \ldots, T\}$*,* b_{it} *can be computed recursively by using*

$$b_{it} = \sum_{j=1}^{M} \overline{p}_{ji}(1+g)b_{jt-1} + \sum_{j=1}^{\lfloor \Gamma_i^t \rfloor} u^{(j)}_{it-1} + (\Gamma_i^t - \lfloor \Gamma_i^t \rfloor) u^{(\lceil \Gamma_i^t \rceil)}_{i,t-1},$$

and $b_{i0} = a_i$*.*

Proof. We first note that based on the discussion given in Section 2.1, (R1) is equivalent to

$$\min \quad \sum_{i=1}^{M} \sum_{t=1}^{T} v_{it} \tag{20}$$

$$\sum_{i=1}^{M} x_{it} \leq U \qquad \text{for } i = 1, \ldots, T \tag{21}$$

$$v_{i0} = a_i \qquad \text{for } i = 1, \ldots, M \tag{22}$$

$$v_{it} + b_{it} \sum_{t'=1}^{t} x_{ik} \geq \sum_{j=1}^{M} \overline{p}_{ji}(1+g)v_{jt-1} +$$

$$\max_{\{S_i^t \cup \{r_{it}\}:\ S_i^t \subseteq J_i^t,\ |S_i^t| \le \lfloor \Gamma_i^t \rfloor,\ r_{it} \in J_i^t \setminus S_i^t\}} \left\{ \sum_{j \in S_i^t} \widehat{p}_{ji}(1+g)v_{jt-1} + (\Gamma_i^t - \lfloor \Gamma_i^t \rfloor)\widehat{p}_{r_{it},i}(1+g)v_{r_{it},t-1} \right\}$$

$$\text{for } i = 1,\dots,M \text{ and } t = 1,\dots,T \tag{23}$$

$$v_{it} \ge 0 \qquad \text{for } i = 1,\dots,M \text{ and } t = 1,\dots,T \tag{24}$$

$$x_{it} \ge \{0,1\} \qquad \text{for } i = 1,\dots,M \text{ and } t = 1,\dots,T. \tag{25}$$

Based on Constraints (23), b_{it} where $i \in \{1,\dots,M\}$ and $t \in \{1,\dots,T\}$ is sufficiently large if regardless of the value of v_{jt-1} for $j = 1,\dots,M$, we have that

$$b_{it} \ge \sum_{j=1}^{M} \overline{p}_{ji}(1+g)v_{jt-1} + \max_{\{S_i^t \cup \{r_{it}\}:\ S_i^t \subseteq J_i^t,\ |S_i^t| \le \lfloor \Gamma_i^t \rfloor,\ r_{it} \in J_i^t \setminus S_i^t\}} \left\{ \sum_{j \in S_i^t} \widehat{p}_{ji}(1+g)v_{jt-1} + (\Gamma_i^t - \lfloor \Gamma_i^t \rfloor)\widehat{p}_{r_{it},i}(1+g)v_{r_{it},t-1} \right\}.$$

It is evident that

$$\max_{\{S_i^t \cup \{r_{it}\}:\ S_i^t \subseteq J_i^t,\ |S_i^t| \le \lfloor \Gamma_i^t \rfloor,\ r_{it} \in J_i^t \setminus S_i^t\}} \left\{ \sum_{j \in S_i^t} \widehat{p}_{ji}(1+g)v_{jt-1} + (\Gamma_i^t - \lfloor \Gamma_i^t \rfloor)\widehat{p}_{r_{it},i}(1+g)v_{r_{it},t-1} \right\} \le \sum_{j=1}^{\lfloor \Gamma_i^t \rfloor} u_{it-1}^{(j)} + (\Gamma_i^t - \lfloor \Gamma_i^t \rfloor) u_{i,t-1}^{(\lceil \Gamma_i^t \rceil)},$$

for $i = 1,\dots,M$ and $t = 1,\dots,T$. So, the result follows. □

B Robust Formulation of (D2)

Based on our discussion in Section 2.1, it is easy to show that the robust counterpart formulation of (D2) can be stated as:

$$\text{(R2)} \qquad \min \sum_{i=1}^{M} c_i x_i \tag{26}$$

$$\sum_{i=1}^{M} \overline{w}_{ik} x_i - z_k \Gamma_k - \sum_{i=1}^{M} q_{ik} \ge W_k \qquad \text{for } k = 1,\dots,K \tag{27}$$

$$z_k + q_{ik} \geq \widehat{w}_{ik} x_i \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{28}$$

$$q_{ik} \geq 0 \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{29}$$

$$z_k \geq 0 \qquad \text{for } k = 1, \ldots, K \tag{30}$$

$$x_i \in \{0, 1\} \qquad \text{for } i = 1, \ldots, M, \tag{31}$$

where Γ_k for $k = 1, \ldots, K$ is a user-defined parameter showing the level of conservatism in Constraint (27).

C Revised Robust Formulation of (D2)

It is evident from Constraint (27) that the term $-z_k\Gamma_k - \sum_{i=1}^{M} q_{ik}$ cannot take a positive value and so it can force more x_i to take the value of 1 (in comparison to (D2)). However, this in itself can force the value of $-z_k\Gamma_k - \sum_{i=1}^{M} q_{ik}$ to become even more negative due to Constraint (28). Thus, it is possible for (R2) to be infeasible. The higher the degree of uncertainty in (R2), i.e., larger the values of $\widehat{w}_{ik}$ and Γ_k for $i = 1, \ldots, M$ and $k = 1, \ldots, K$, larger the probability of this outcome arising.

So, to deal with infeasibility of (R2), we propose a revised formulation as follows:

$$\text{(R3)} \qquad \min \sum_{i=1}^{M} c_i x_i \tag{32}$$

$$\min \sum_{i=1}^{M} \sum_{k=1}^{K} \epsilon_{ik} \tag{33}$$

$$\sum_{i=1}^{M} \overline{w}_{ik} x_i - z_k\Gamma_k - \sum_{i=1}^{M} q_{ik} \geq W_k \qquad \text{for } k = 1, \ldots, K \tag{34}$$

$$z_k + q_{ik} \geq (\widehat{w}_{ik} - \epsilon_{ik}) x_i \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{35}$$

$$0 \leq \epsilon_{ik} \leq \widehat{w}_{ik} \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{36}$$

$$q_{ik} \geq 0 \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{37}$$

$$z_k \geq 0 \qquad \text{for } k = 1, \ldots, K \tag{38}$$

$$x_i \in \{0, 1\} \qquad \text{for } i = 1, \ldots, M, \tag{39}$$

where ϵ_{ik} is a new continuous variable that is introduced in order to revise/reduce the value of $\widehat{w}_{ik}$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$. This can be observed from Constraints (35) and (36). In consequence, the new objective function, $\sum_{i=1}^{M} \sum_{k=1}^{K} \epsilon_{ik}$, simply measures the infeasibility of (R2) with respect to the value of $\widehat{w}_{ik}$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$. To understand the formulation better, we now explore two extreme cases. In the first case we suppose that $\epsilon_{ik} = 0$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$. In such a scenario, (R3) is precisely equivalent to (R2). Now, in the second case, let us suppose that $\epsilon_{ik} = w_{ik}$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$. In this case, the optimal solution of (D2) is also optimal for (R3) because we now have the option to set $z_k = 0$ and $q_{ik} = 0$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$. So, this formulation captures the essence of both (D2) and (R2), and it is guaranteed to be feasible (since we assume that (D2) is feasible).

Note that solving this bi-objective optimization problem returns the trade-off between the total cost of conservation, i.e., the first objective function, and the total infeasibility, i.e., the second objective function. However, the proposed formulation is not linear. In order to linearize it, a new non-negative variable $\widehat{\epsilon}_{ik}$ can be introduced to capture the value of the bilinear term $\epsilon_{ik} x_j$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$, and then Constraint (35) can be replaced by the following constraints:

$$z_k + q_{ik} \geq \widehat{w}_{ik} x_i - \widehat{\epsilon}_{ik} \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{40}$$

$$\widehat{\epsilon}_{ik} \leq \epsilon_{ik} \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{41}$$

$$\widehat{\epsilon}_{ik} \leq \widehat{w}_{ik} x_i \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{42}$$

$$\widehat{\epsilon}_{ik} \geq \epsilon_{ik} - \widehat{w}_{ik} x_i - \widehat{w}_{ik} \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K \tag{43}$$

$$\widehat{\epsilon}_{ik} \geq 0 \qquad \text{for } i = 1, \ldots, M \text{ and } k = 1, \ldots, K. \tag{44}$$

This linearization is valid since if $x_i = 1$ then $\widehat{\epsilon}_{ik} = \epsilon_{ik}$ and if $x_i = 0$ then $\widehat{\epsilon}_{ik} = 0$ for $i = 1, \ldots, M$ and $k = 1, \ldots, K$.