Computers & Operations Research 153 (2023) 106160

Contents lists available at ScienceDirect

Computers and Operations Research

journal homepage: www.elsevier.com/locate/cor

Spatial optimization of multiple area land acquisition

Alan T. Murray *, Richard L. Church

Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106, USA

ARTICLE INFO

Keywords:
Land allocation
Contiguity
Thresholds
Land use planning
Wildfire

ABSTRACT

Land acquisition was first posed as an integer programming problem by Wright et al. (1983). They defined this problem as the selection of spatial planning units that altogether formed a contiguous area of land of desired size while optimizing one or more objectives. One of the key difficulties in solving the land acquisition as an integer program is ensuring spatial contiguity among the land units that are selected. Also recognized is that some application contexts are also concerned with compactness. Over time, a number of model formulations and solution heuristics have been proposed for the case of delineating one contiguous area, as well as extension to identify multiple independent areas. Unfortunately, existing model formulations are either too computationally complex to solve reasonably sized problems or they rely on one or more simplifying assumptions that limit their utility. We introduce a new model formulation for the multiple land acquisition problem that can be solved by exact methods using commercial mixed-integer programming software, and does so without simplifying assumptions. The key element of this new model formulation is that the contiguity constraints are self-contained and complete and can accommodate any interpretation of land unit neighbor (or adjacency) relationships. Overall, this new model can be applied to problems involving tens of thousands of spatial planning units. Application experience for this general model is detailed, addressing delineated area selection for concentrated fuels removal/reduction efforts in order to reduce wildfire risk for the U.S. Forest Service.

1. Introduction

Land use planning is complicated and complex, requiring details, insights and supporting geographic information about local and regional conditions as well as an understanding of impacts that will result in subsequent land conversion, development and/or change. There are many contexts where an area or areas are necessary for a proposed land use, including parks, recreation facilities, waste processing centers, nature preserves and conservation zones, affordable housing developments, factories, photovoltaic power stations, wind power plants, and many others. Important characteristics of such areas are that they provide the most benefit possible to the intended use or purpose, not exceed a maximum size and are contiguous with respect to the land units/parcels that constitute an individual area.

There has long been interest in the use of geographic information systems and spatial analytics to support land acquisition planning (see Brookes 1997, Church 2002, Ligmann-Zielinska et al. 2008, Xiao and Murray 2019). This paper addresses the land acquisition problem involving the selection of a set of spatial planning units from among a large collection of units to form one or more connected areas, where each planning area is no larger than a specified maximum size and offers the greatest total benefit possible. Wright et al. (1983) was among the first to model land acquisition, identifying one contiguous area adhering to size requirements. This was accomplished using an integer-based mathematical model that sought contiguity in an indirect manner, by minimizing the resulting perimeter or border of the acquired land. Their model was formulated for a landscape defined as a raster (square cells). An extension to account for multiple areas was detailed in Benabdallah and Wright (1992). Fischer and Church (2003) too sought to minimize perimeter in defining clusters of compact reserve areas where the basic spatial unit was a small watershed with the Sierra Nevada mountains of California. Heuristic solution approaches for the single area land acquisition problem have been developed by Brooks (1997) and Church et al. (2003). Spatially explicit approaches to impose land acquisition contiguity have been developed by Williams (2002), Shirabe (2005), and Onal et al. (2016). Specifically, Williams (2002) structured an approach that exploited primal and dual graph complementarity assuming planar neighbor relationships between spatial units. The assumption requires that two units be considered neighbors, or adjacent, only if they share a non-zero length edge. This is often referred to as

* Corresponding author.
E-mail address: amurray@ucsb.edu (A.T. Murray).

https://doi.org/10.1016/j.cor.2023.106160
Received 18 May 2022; Received in revised form 12 January 2023; Accepted 23 January 2023
Available online 27 January 2023

"rook" adjacency in spatial analysis. A more generalized and less restrictive approach was structured in Shirabe (2005), enabling any definition of neighbors to be considered. Contiguity was structured using the sink flow approach of Cova and Church (2000). Both Williams (2002) and Shirabe (2005) were limited to the case of identifying a single area only. Onal et al. (2016) also relied on a principle raised by Cova and Church (2000). Basically, if a unit is selected as the center of a planning area, then any other unit selected must have a neighboring unit pathway leading to the center. Their model was formulated for a raster, like that of Cova and Church (2000), except that the choice of centers was not preselected. To make this computationally tractable, Onal et al. (2016) imposed the condition that that the selection of the neighboring cell be closer to the center, possible only in the context of a raster and a focus on compactness.

The multiple land acquisition problem is related to a number of spatial problems, including districting and region delineation, where spatial units are divided into a set of contiguous areas, each meeting some criteria of size. This includes: the p-regions problem (Duque et al. 2011), the p-compact regions problem (Li et al. 2014) and the max-p-regions (Duque et al. 2012). Each of these region delineation problems involves a spatial partitioning like that of Shirabe (2009). Duque et al (2011) delves into three different approaches to ensure contiguity within each of the resulting regions: flow based, tree based and path-based contiguity. Virtually all of these constraint approaches proved to be computationally challenging to solve, leading to the use of heuristics for real-world applications. Practically speaking, this has been true for the general districting problem as well.

The objective of this paper is to present a new formulation for the single or multiple land acquisition problem that uses a new form of "flow-based constraints." This model is an outgrowth of research supported by the U.S. Forest Service, where the objective is to define multiple contiguous areas for fuels removal or reduction. Such activities have been shown to improve the resilience of forests to wildfire. Here too, work has relied on heuristics to solve the problem. Ager et al. (2013, 2016) reports the use of a heuristic to identify the best wildfire mitigation areas, where each area is limited in size and must be contiguous. Unfortunately, Ager et al. (2013, 2016) were not able to mathematically structure the multiple area land acquisition problem, confounded by the contiguity requirements, offering instead a partial formulation combined with a text based description of the actual problem. Recent research reported in Murray et al. (2022) derived a formal mathematical model of the multiple area land acquisition problem suitable for forest planning involving explicit contiguity requirements.

Model use and application experience was reported in Murray et al. (2022) to support wildfire mitigation efforts, similar to that of Ager et al. (2013, 2016). In contrast to Ager et al. (2013, 2016), or the heuristics of Brookes (1997, 2001) and Nalle et al. (2002), and the approximation approach of Benabdallah and Wright (1992), the mixed-integer program detailed in Murray et al. (2022) is capable of solution by exact methods. However, as problem size grows substantially (e.g., the number of spatial land units and/or the number of areas to identify increase), the associated problem instances are difficult to solve using commercial software. As a result, optimality gaps may be encountered in practice, reflecting that a feasible solution is found, but that it was not possible to close the theoretical bound between the best solution possible established by linear programming relaxation in the branch and bound solution process. Murray et al. (2022) analyzed five regions within the Stanislaus National Forest, ranging in size from 57 up to 1,343 units and involving one to 55 areas to be selected. In total, 23 application instances were solved, with optimality gaps ranging from 0 % to 64.58 % (average of 12.68 %). The largest application instance involved 1,343 land units with 55 areas to be selected, resulting in a mixed-integer program of 223,048 constraints and 582,450 decision variables.

The implications of the above noted details are that multiple area land acquisition is fundamentally important but problem applications remain challenging to solve, and as a result active interest continues with desperate need for continued solution advancements. The ability to identify even slightly improved land acquisition planning alternatives through better modeling and/or solution approaches can translate to a reduction of extreme wildfire events, enhanced sustainability, species persistence, long term viability, economic prosperity, etc. The benefits to the environment, humans and society, as a result, are likely immeasurable. To this end, there are clear advantages to exact approaches capable of establishing solution quality, even when optimality cannot be confirmed (e.g., optimality gap greater than zero). In contrast, heuristic solution approaches cannot generally provide any assurance or measure of solution quality with respect to optimality. Heuristic solutions may be of high quality or of low quality, leaving substantial uncertainty in any identified planning scenario.

The remainder of this paper is organized as follows. The next section offers a new mixed-integer programming formulation of the multiple area land acquisition problem where each area is required to be contiguous and limited in size. Model properties are detailed along with a proof of validity. Planning applications are then examined to demonstrate model performance, focusing on wildfire mitigation efforts by the USDA Forest Service through the Social and Ecological Resilience Across the Landscape project. Seven different regions are examined, considering a number of different areas to be identified, involving up to 10,642 spatial units and 295 areas. Exact solution using commercial optimization software for the 34 different planning application scenarios confirms optimality to within 1.60 % on average, with an observed optimality gap range of 0 % up to a maximum of 4.36 %.

2. Model development

Existing methods that address multiple area land acquisition rely on heuristic methods (e.g., Brookes 1997, 2001, Nalle et al. 2002, Ager et al. 2013, 2016), approximate models using surrogate measures of contiguity (e.g., Benabdallah and Wright 1992), impose restrictive simplifying assumptions (e.g., Williams 2002) and/or are theoretically or computationally limited in some way (e.g., Shirabe 2005, Murray et al. 2022). As a result, theoretical and practical advancements are needed in the optimization of land acquisition. The land acquisition problem addressed in this research involves the selection of spatial units to form a prespecified number of unique areas, where each area is limited in size and contiguous. A mixed-integer formulation of this problem is introduced, proving to have superior solution properties based on the representation of selected areas as sink oriented networks, similar to notions originally conceived of in Cova and Church (2000) and relied upon in Shirabe (2005).

Primary features of problem structure are illustrated for a single selected area in Fig. 1, with the directed arrow width representing smaller or larger flow volumes between node pairs. Each unit in Fig. 1 is 10 acres in size, with selection denoted as a network node. Selection generates/initiates a flow equivalent to the spatial unit size (e.g., 10 acres in this example) destined for the area sink, with flow permitted through selected neighboring land units that are included in the area. Thus, unit 5 initiates 10 acres of flow to unit 4, and similarly units 8 and 9 each initiate 10 acres of flow to neighboring unit 7. The selection of unit 4 initiates 10 acres of flow, combined with the incoming 10 acres of flow from unit 5, destined to neighboring unit 3, giving 20 total acres of flow incoming to unit 3. The selection of unit 7 initiates 10 acres of flow combined with the incoming 20 acres of flow, resulting in 30 acres of flow into neighboring unit 6. This is similarly the case for all other selected land units, with initiated and incoming flow required to terminate at a model determined sink. In this case, the sink is unit 1, with incoming flow of 80 acres and an initiated flow of 10 acres, resulting in 90 acres of flow in total. If the maximum size is 100 acres, then this would be a feasible area since it maintains the size limit and is contiguous, where contiguity is a geographic property that selected units are mutually connected via neighbor conditions through other selected units.

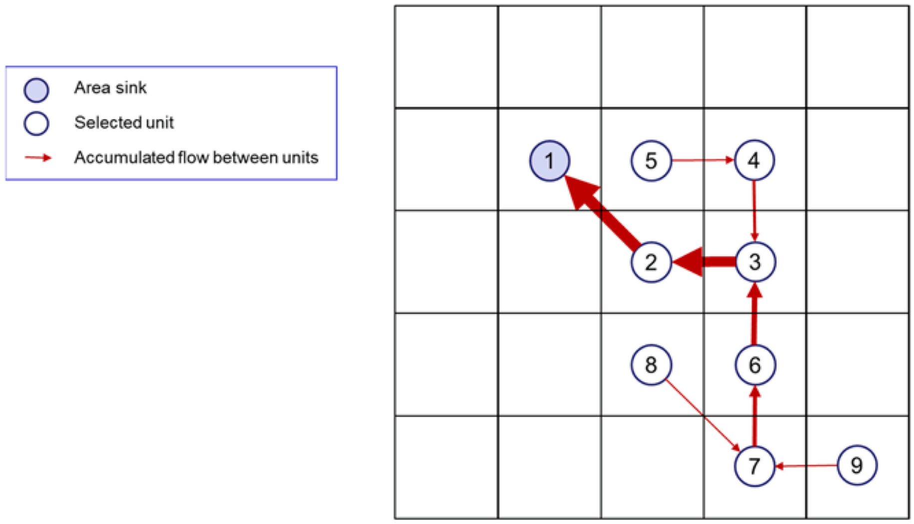

Fig. 1. Spatial land units forming a contiguous area, represented as network flow oriented to indicated sink.

Formalization of multiple area land acquisition as a mixed-integer program is possible. Consider the following notation:

i = index of spatial units (entire set I)
β_i = expected benefit of unit i
α_i = size of unit i
p = number of areas to identify
T = maximum size for an identified area
N_i = set of spatial units that neighbor or are adjacent to unit i

Of major importance are the model decision variables, defined as follows:

$$X_i = \begin{cases} 1 & \text{if unit } i \text{ is selected for usage} \\ 0 & \text{otherwise} \end{cases}$$

$$V_i = \begin{cases} 1 & \text{if unit } i \text{ is selected as an area sink} \\ 0 & \text{otherwise} \end{cases}$$

Y_{ij} = flow from unit i to j destined for sink (initiated and incoming)

$$\tilde{Y}_{ij} = \begin{cases} 1 & \text{if arc between units } i \;\&\; j \text{ is utilized for sink directed flow} \\ 0 & \text{otherwise} \end{cases}$$

The selection decision variable for each unit, X_i, denotes whether a unit is picked for inclusion in an area or not. As noted above, each area must have a sink, from among the units V_i, that serves to spatially organize the area, helping to ensure connectivity between neighboring members of the area by requiring all initiated flow in unit selection to end up at the sink. The decision variables Y_{ij} therefore represent the flow between two neighboring units, both unit selection initiated flow and incoming flow from other selected units. The final set of decision variables is $\tilde{Y}_{ij}$, necessary to ensure that flow from one unit is not split among neighboring units.

A mixed-integer spatial optimization model for multiple area land acquisition follows:

$$\text{Maximize} \sum_i \beta_i X_i \tag{1}$$

$$\text{Subject to} \sum_{j \in N_i} Y_{ij} - \sum_{j \in N_i} Y_{ji} \geq \alpha_i X_i - T V_i \quad \forall i \tag{2}$$

$$Y_{ij} \leq T \;\; \tilde{Y}_{ij} \quad \forall i, j \in N_i \tag{3}$$

$$\sum_{j \in N_i} \tilde{Y}_{ij} \leq 1 \quad \forall i \tag{4}$$

$$T \;\; V_i + \sum_{j \in N_i} Y_{ij} \leq T \quad \forall i \tag{5}$$

$$\sum_i V_i = p \tag{6}$$

$$\sum_{j \in N_i} Y_{ij} \leq T \;\; X_i \quad \forall i \tag{7}$$

$$V_i \leq X_i \quad \forall i \tag{8}$$

$$X_i \in \{0, 1\} \quad \forall i \tag{9}$$

$$V_i \in \{0, 1\} \quad \forall i$$

$$Y_{ij} \geq 0 \quad \forall i, j \in N_i$$

$$\tilde{Y}_{ij} = \{0, 1\} \quad \forall i, j \in N_i$$

The objective, (1), seeks the greatest total benefit possible in the selection of land units to include in formed areas. Constraints (2) establish conservation of flow, plus accounting for initiated flow, and limits total flow into a sink. Constraints (3) require the establishment of an arc flow decision before flow can be allocated. Constraints (4) allow at most one outgoing arc to have flow from a node. Constraints (5) establish terminal conditions for flow at sinks, or rather impose that flow is not allowed to leave a sink. Constraint (6) indicates that exactly p sinks are to be selected, reflecting that p areas are to be geographically delineated. Constraints (7) allow outgoing flow only if the unit has been selected. Constraints (8) permit a sink for an area only from among units selected. Finally, constraints (9) impose binary and non-negativity conditions on decision variables.

While the empirical findings reported later in the paper will highlight that spatial model (1)-(9) does indeed identify contiguous areas of land units that satisfy maximum size limits, as desired, this can be proven as well. A definition of contiguity is now given along with important observations that do not necessarily require formal proof:

Definition 1. Contiguity is a geographic property that selected units are mutually connected via neighbor conditions through only selected units within the identified area.

Observation 1: The selection of units will always approach the maximum limits, given the maximization objective orientation.

Positive benefits β_i associated with land units will always make them desirable for selection, assuming that the size of individual units is less than or equal to the area restriction, e.g., $\alpha_i \leq T$ for all or most units i. Further, it is generally the case that only a subset of units it possible to select, e.g., $\sum_i \alpha_i > T\ p$. A proof of observation 1 is given in Murray et al. (2022), and will not be repeated here.

Observation 2: There will be exactly p areas identified. This is established in constraint (6), where the number of sinks is to be p, $\sum_i V_i = p$.

Observation 3: A sink is only permitted among selected land units. This is established in constraints (8), where the condition $V_i \leq X_i$ prevents a unit from being a sink unless it is selected for area inclusion.

Establishing that the formulated model performs as intended can be approached in different ways. In some cases, application results offer evidence that it does what it claims. Examples also can be used to illustrate what is occurring in the model, providing support for validity. Proofs and theorems too can be undertaken to bolster model validity in various ways. In what follows, we rely on each of these forms to establish model validity.

Theorem 1. Prevention of flow splitting out of a selected node.

Proof: As noted above, the selection of a spatial unit X_i initiates flow equivalent to the size of the unit, α_i, destined to the area sink. This happens in constraint (2). Assume that flow out of a selected unit i is divided, or split, among neighboring units N_i that have also been selected. A split between two (or more) neighboring units $j, j' \in N_i$ implies that $0 < Y_{ij}$ and $0 < Y_{ij'}$ for $j, j' \in N_i$. Constraints (3) require that $\widetilde{Y}_{ij} = 1$ and $\widetilde{Y}_{ij'} = 1$ since $Y_{ij} \leq T\ \widetilde{Y}_{ij}$ and $Y_{ij'} \leq T\ \widetilde{Y}_{ij'}$, respectively. However, this is not possible by constraints (4) since $\widetilde{Y}_{ij} + \widetilde{Y}_{ij'} \leq 1$. This is therefore a contradiction, so flow splitting out of a node is not permissible for a feasible solution. ■.

Theorem 2. Contiguity among units in a selected area.

Proof: Assume that some unit i is selected, e.g., $X_i = 1$, but it is not contiguous to other selected units that comprise the area (Definition 1). Consider two cases, $p = 1$ and $p > 1$.

Case (a): For $p = 1$, two situations can arise. The first is that unit i is the sink, $V_i = 1$. This implies that constraint (2) for unit i would be $\sum_{j\in N_i} Y_{ij} - \sum_{j\in N_i} Y_{ji} \geq \alpha_i\ 1 - T\ 1$, or $\sum_{j\in N_i} Y_{ij} - \sum_{j\in N_i} Y_{ji} \geq \alpha_i - T$. Since unit i is not contiguous with or connected to neighbors, by assumption, then this simplifies to $0 - 0 \geq \alpha_i - T$, or $T \geq \alpha_i$. The assumption of at least one other unit i' being selected implies $\sum_{j\in N_{i'}} Y_{i'j} - \sum_{j\in N_{i'}} Y_{ji'} \geq \alpha_{i'}$ since there is only one sink. This means that $\sum_{j\in N_{i'}} Y_{i'j} - 0 \geq \alpha_{i'}$ because there is no incoming flow (e.g., $\sum_{j\in N_{i'}} Y_{ji'} = 0$). If it is not contiguous by assumption, then $\sum_{j\in N_{i'}} Y_{i'j} = 0$ as there is nowhere for the outgoing flow to go to since $i \notin N_{i'}$. This implies $0 \geq \alpha_{i'}$, which is not possible as it violates constraints (2). The second situation is that unit i is not the sink, $V_i = 0$. This implies that constraint (2) for unit i would be $\sum_{j\in N_i} Y_{ij} - \sum_{j\in N_i} Y_{ji} \geq \alpha_i 1 - T0$, or $\sum_{j\in N_i} Y_{ij} - \sum_{j\in N_i} Y_{ji} \geq \alpha_i$. Since unit i is not contiguous with or connected to neighbors, by assumption, then this simplifies to $0 - 0 \geq \alpha_i$, or $0 \geq \alpha_i$, which is not possible as it violates constraints (2).

Case (b): For $p > 1$, the rationale and proof are similar to Case (a), except now the unit must be part of another selected area, and if not encounters the situations noted for case (a).

Thus, the contradictions of the assumptions hold that, for one or more areas, any selected unit must be contiguous with its associated sink via connected selected units for a feasible solution. ■

Theorem 3. No outflow from a sink unit.

Proof: Assume that there is flow emanating from a selected sink i, $V_i = 1$. This implies that $\sum_{j\in N_i} Y_{ij} > 0$. Observation 2 holds that $X_i = 1$, if it is also a sink. Constraints (5) stipulate that $T + \sum_{j\in N_i} Y_{ij} \leq T$, or $\sum_{j\in N_i} Y_{ij} \leq 0$, which is a contradiction. Thus, no outflow from a sink is possible for a feasible solution.■.

Theorem 4: Maximum size limit among units in a selected area are maintained.

Proof: Suppose the total size of a selected area exceeds the limit T. There are two cases to consider, $p = 1$ and $p > 1$.

Case (a): $p = 1$ implies that $\sum_i \alpha_i X_i > T$. By Theorem 2, all units of a selected area are contiguous. Constraints (7) indicate that any selected unit i, $X_i = 1$, satisfies $\sum_{j\in N_i} Y_{ij} \leq TX_i$, or $\sum_{j\in N_i} Y_{ij} \leq T$. Thus, outgoing flow cannot exceed T. Similarly, from constraints (2) we have $\sum_{j\in N_i} Y_{ij} - \sum_{j\in N_i} Y_{ji} \geq \alpha_i - T\ V_i$, creating two situations (e.g., unit i is a sink, $V_i = 1$, or it is not, $V_i = 0$). When $V_i = 1$, $\sum_{j\in N_i} Y_{ij} - \sum_{j\in N_i} Y_{ji} \geq \alpha_i - T$. By Theorem 3 there is no outgoing flow from a sink, so $\sum_{j\in N_i} Y_{ij} = 0$. This means that $-\sum_{j\in N_i} Y_{ji} \geq \alpha_i - T$, or $T \geq \alpha_i + \sum_{j\in N_i} Y_{ji}$, contradicting the assumption. Alternatively, when $V_i = 0$, then $\sum_{j\in N_i} Y_{ij} - \sum_{j\in N_i} Y_{ji} \geq \alpha_i$, suggesting that the outgoing flow must be at least the incoming and initiated flows, bounded by the limit: $T \geq \sum_{j\in N_i} Y_{ij} \geq \alpha_i + \sum_{j\in N_i} Y_{ji}$. This also contradicts the assumption of an exceeded limit.

Case (b): $p > 1$ follows along similar lines. Since an area is contiguous by Theorem 2 and no outflow from a sink is possible by Theorem 3, this case collapses to that of $p = 1$, case (a).

Therefore, these cases violate the assumption, so a feasible solution will not exceed the maximum size of an area. ■

The above observations and theorems establish the validity of the multiple area land acquisition formulation, (1)-(9). This will be further supported in the application results that follow in the next section. Also important, however, is expected problem size in application in terms of the number of decision variables and constraints. This is generally an indicator of the expected success of mixed-integer program solution through branch and bound approaches. Define $\overline{N} = \sum_i N_i / |I|$, the average number of observed spatial neighbors, where indicates the number of elements in the set. The number of decision variables is: $|I| + |I| + |I|\overline{N} + |I|\overline{N} = 2|I|(1 + \overline{N}) \approx O(|I|\overline{N})$. The notation $O()$ indicates on the order of, effectively a coefficient multiplier that is substantially smaller than the indicated set size or number. The number of constraints is: $|I| + |I|\overline{N} + |I| + |I| + 1 + |I| + |I| = |I|(5 + \overline{N}) + 1 \approx O(|I|\overline{N})$. Notice that neither the number of decision variables nor the number of constraints is dependent on the number of areas to be selected, p. An important property of this formulation is that model size (e.g., number of decision variables and number of constraints) is invariant to the number of areas selected, a point that will be further highlighted later in the paper.

3. Application results

The developed model, (1)-(9), is applied in support of USDA Forest Service wildfire mitigation efforts. Considered here is the Social and Ecological Resilience Across the Landscape project region within the Stanislaus National Forest in California, where wildfire risk was considered in terms of suppression difficulty. Spatial units, I, were defined by the USDA Forest Service to have homogeneous vegetation and topography, with benefit derived as a measure of departure from wildfire resilience. Thus, benefit β_i for spatial unit i was based on forest structure, including mean clump size, trees per acre and proportion of gaps, and represents a measure of forest resilience. The goal is therefore to select a prespecified number of areas that are most beneficial, maintaining contiguity within an area that is limited to at most 100 acres in size (e.g., $T = 100$). The intent is to transition vegetation conditions through mitigation to within historical ranges in order to decrease extreme event wildfire susceptibility.

Seven different planning regions are considered, varying in number of management units, from $|I| = 57$ to $|I| = 10{,}642$. Summary characteristics can be found in Table 1 in terms of region size as well as minimum, mean and maximum spatial unit size α_i and benefit β_i. The range

Table 1
Application details and attribute [minimum, mean, maximum] summary.

$\|I\|$	Region (acres)	α_i	β_i
57	747.69	[2, 13.12, 55.15]	[0, 16.59, 167.75]
130	1,752.47	[2, 13.48, 101.85]	[0, 7.78, 46.71]
315	3,279.88	[0.22, 10.41, 77.84]	[0, 12.34, 94.70]
608	5,546.97	[0.22, 9.12, 70.28]	[0, 8.83, 119.73]
1,343	15,778.25	[0.89, 11.75, 217.06]	[0, 13.69, 198.80]
4,696	48,322.17	[0.22, 10.29, 179.47]	[0, 10.31, 258.08]
10,642	117,917.75	[0.22, 11.08, 217.06]	[0, 12.41, 270.63]

of areas selected effectively spans up to 25 % of each region, a desired USDA Forest Service goal in many cases depending on resource availability.

A desktop personal computer (Intel Xeon E5 CPU, 2.30 GHz with 96 GB RAM) running Windows 10 was used for the analysis. The multiple area land acquisition model, (1)-(9), was implemented in Python and solved using GUROBI (version 9.5). Default solution settings were applied.

The seven different regions combined with desired outcomes gives 34 different planning application scenarios. Exact solution details are offered in Table 2, summarized by region (e.g., $|I|$) and number of areas to be identified (e.g., p). For each scenario, column headings represent the following. "Rows" indicates the number of model constraints and "Columns" indicates the number of model decision variables, in accordance with expected theoretical problem size characteristics outlined previously. The "Objective" is that specified in (1). "Optimality gap" is the theoretical bound remaining if the branch and bound process is terminated prior to verifying optimality or unable to verify optimality, as a percentage. Finally, "Time" is the solution time to solve the problem instance, in seconds. In this case, a maximum time limit of one hour (e.g., 3,600 s) was imposed on the solution process. Any remaining gap is unlikely to be resolved with further computational processing, even after 24 + hours. Examining Table 2, the scenario involving $|I| = 57$ units with 1 area to be identified has 582 constraints and 706 decision variables. The optimal solution was confirmed, reflected in an optimality gap of 0.00 %, with an objective value of 312.8903. Total solution time in this case was 0.1 s. Looking further down Table 2, the scenario of $|I| = 10{,}642$ with 295 areas to be identified had 117,923 constraints and 150,708 decision variables. The best feasible solution found was 57,977.1320, with a terminating optimality gap of 3.85 % after 3600.0 s.

Table 2 indicates that 10 of the 34 problems were solved to optimality, with no remaining gap. The average optimality gap across the 34 problem scenarios was 1.60 %, with an observed maximum of 4.36 %. The summary for each scenario given in Table 2 reflects a spatial configuration of units selected as part of individual areas to undergo coordinated wildfire mitigation efforts. As an example, the optimal mitigation plan for scenario $|I| = 608$ with one area to be identified is shown in Fig. 2. This area is contiguous and maintains the 100 acre limit. As the optimality gap is 0.0 %, there is no other feasible solution that has a better total benefit than 211.8159. Other scenario solutions can be visualized as well. Fig. 3 shows the best five areas ($p = 5$) found for $|I| = 1{,}343$. In this case, the optimality gap is 0.88 % with the objective value of 1,242.9900. Fig. 4 depicts the scenario for $|I| = 4{,}696$ of $p = 50$. The total benefit for this configuration of units is 10,935.9646, with an optimality gap of 2.37 %. Finally, Fig. 5 illustrates the 100 areas ($p = 100$) for $|I| = 10{,}642$, giving a total benefit of 22,970.1493 with a terminating optimality gap of 3.17 % after 3600.0 s.

4. Discussion

There are many items worth further discussion associated with contiguity, planning application and observed computational performance. To begin, it is important to note that the reported findings

Table 2
Solution summary of different scenarios for wildfire mitigation by application region.

$\|I\|$	p	Rows	Columns	Objective, (1)	Optimality gap (%)	Time (seconds)
57	1	582	706	312.8903	0.00	0.1
57	2	582	706	506.2098	0.00	1.3
57	3	582	706	667.0241	0.00	13.7
130	1	1,365	1,688	131.2376	0.00	6.9
130	3	1,365	1,688	357.4882	0.00	32.1
130	6	1,365	1,688	610.4840	0.36	3600.0
315	1	3,338	4,154	206.5037	0.00	5.2
315	5	3,338	4,154	988.2890	0.47	3600.0
315	11	3,338	4,154	2,028.0640	1.30	3600.0
608	1	6,601	8,336	211.8159	0.00	8.0
608	5	6,601	8,336	1,007.7249	0.34	3600.0
608	10	6,601	8,336	1,930.7770	0.85	3600.0
608	15	6,601	8,336	2,761.2166	1.12	3600.0
608	19	6,601	8,336	3,323.6051	1.78	3600.0
1,343	1	14,620	18,494	284.6195	0.00	3.0
1,343	5	14,620	18,494	1,242.9900	0.88	3600.0
1,343	10	14,620	18,494	2,343.5952	1.57	3600.0
1,343	15	14,620	18,494	3,369.5102	1.75	3600.0
1,343	20	14,620	18,494	4,346.2283	2.28	3600.0
1,343	30	14,620	18,494	6,250.9360	2.19	3600.0
1,343	40	14,620	18,494	7,999.2666	2.07	3600.0
1,343	55	14,620	18,494	10,356.7595	2.49	3600.0
4,696	1	51,371	65,172	293.3076	0.00	23.8
4,696	5	51,371	65,172	1,308.3556	2.95	3600.0
4,696	10	51,371	65,172	2,495.4732	4.04	3600.0
4,696	50	51,371	65,172	10,935.9646	2.37	3600.0
4,696	80	51,371	65,172	16,245.7193	3.17	3600.0
4,696	120	51,371	65,172	22,486.9881	3.58	3600.0
10,642	1	117,923	150,708	321.5257	0.00	228.8
10,642	50	117,923	150,708	12,322.0298	4.36	3600.0
10,642	100	117,923	150,708	22,970.1493	3.17	3600.0
10,642	150	117,923	150,708	32,590.6136	3.75	3600.0
10,642	200	117,923	150,708	42,032.0435	3.24	3600.0
10,642	295	117,923	150,708	57,977.1320	3.85	3600.0

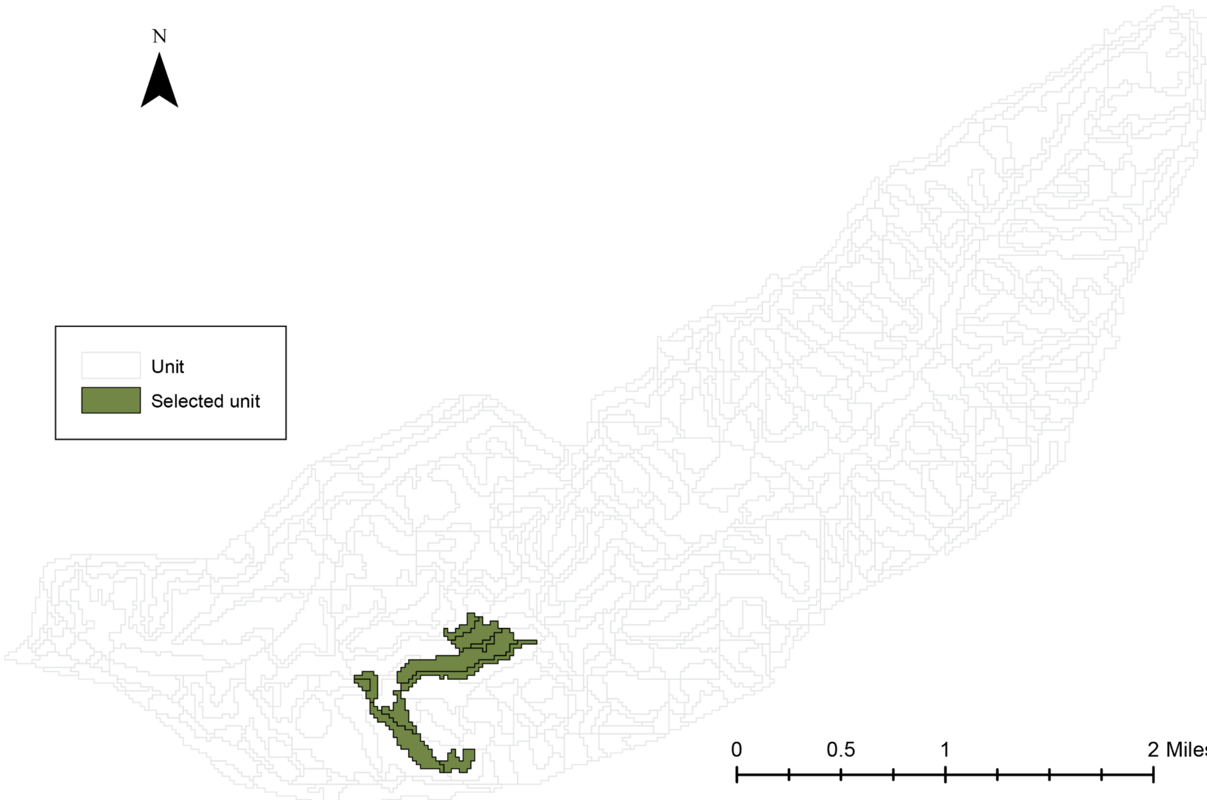

Fig. 2. Selected areas for $|I| = 608$ ($p = 1$, $T = 100$).

Fig. 3. Selected areas for $|I| = 1{,}343$ ($p = 5$, $T = 100$).

summarized in Table 2 reflect the largest sized problems addressed using an exact solution approach, whether in the context of single area or multiple area land acquisition. Shirabe (2005) dealt with 179 units, seeking only a single area. Murray et al. (2022) considered up to 1,343 units in multiple area selection. Restrictive assumptions or related contiguity variants addressed problem sizes of 144 units in Williams (2002), 20 units in Magnanti and Raghavan (2005), 115 units in Onal and Briers (2006), 1,363 units in Carvajal et al. (2013), 1,000 units in Haouari et al. (2013), 1,600 units in Onal et al. (2016) and 2,000 units in Wang et al. (2018). Given this, the reported findings in Table 2 highlight the advancement in exact solution capabilities offered by the new formulation of multiple area land acquisition.

While not the focus of this research, the observed results in Table 2 may be more precisely contrasted with the approach detailed in Murray et al. (2022). Again, they derived a spatial optimization model that represents an extension of Shirabe (2005) to address multiple area land acquisition. Formulation of this model requires the following additional notation:

k = index of areas to be selected

M = large value (total number of spatial units in this case)

$$X_{ik} = \begin{cases} 1 & \text{if unit } i \text{ is selected for treatment as a member of area } k \\ 0 & \text{otherwise} \end{cases}$$

$$V_{ik} = \begin{cases} 1 & \text{if unit } i \text{ is selected as area } k \text{ sink} \\ 0 & \text{otherwise} \end{cases}$$

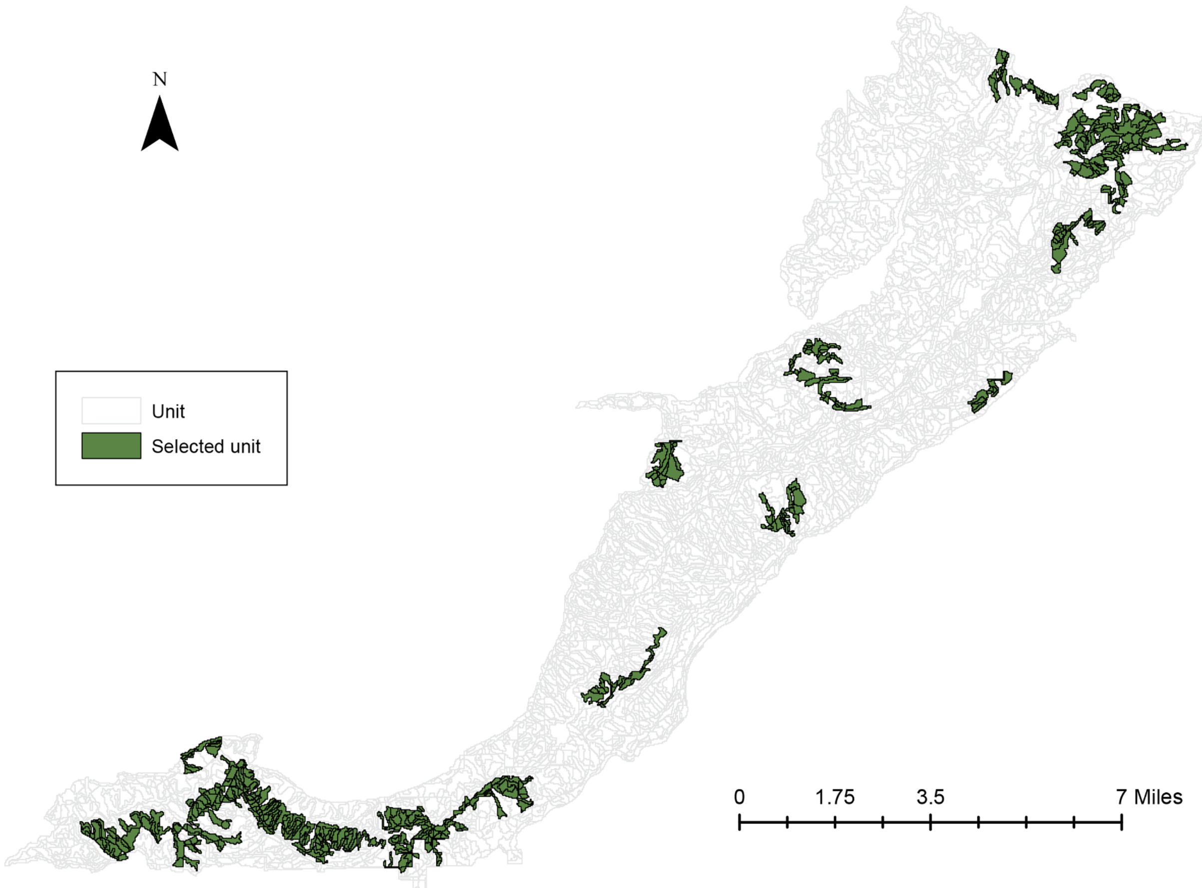

Fig. 4. Selected areas for $|I| = 4{,}696$ ($p = 50$, $T = 100$).

Y_{ijk} = accumulated flow from unit i to j destined for area k sink.

The model offered in Murray et al. (2022) is as follows:

$$\text{Maximize} \sum_i \sum_{k=1}^{p} \beta_i X_{ik} \tag{10}$$

$$\text{Subject to} \sum_i \alpha_i X_{ik} \leq T \quad \forall k \tag{11}$$

$$\sum_{j \in N_i} Y_{ijk} - \sum_{j \in N_i} Y_{jik} \geq X_{ik} - M \ \ V_{ik} \quad \forall i, k \tag{12}$$

$$\sum_i V_{ik} = 1 \quad \forall k \tag{13}$$

$$\sum_{j \in N_i} Y_{ijk} \leq (M-1) X_{ik} \quad \forall i, k \tag{14}$$

$$V_{ik} \leq X_{ik} \quad \forall i, k \tag{15}$$

$$\sum_{k=1}^{p} X_{ik} \leq 1 \quad \forall i \tag{16}$$

$$X_{ik} \in \{0, 1\} \quad \forall i, k \tag{17}$$

$$V_{ik} \in \{0, 1\} \quad \forall i, k$$

$$Y_{ijk} \geq 0 \quad \forall i, j, k$$

The objective, (10), is equivalent to (1), seeking the greatest total benefit. Constraints (11) impose size limits on each identified area. Constraints (12) ensure conservation of flow through a unit and account for initiated flow, similar in nature to constraints (2). Constraints (13) require exactly-one sink for each area. Constraints (14) allow outcoming flow only if the unit has been selected. Constraints (15) permit a sink for an area only from among units selected. Constraints (16) limit a unit to being part of at most one area. Finally, constraints (17) impose binary and non-negativity conditions on decision variables.

A major distinction of (1)-(9) compared with (10)-(17) is associated problem size, with (10)-(17) requiring decision variables and constraints that are a function of p, the number of areas to be identified. This formulation has $2|I|p + 3p + |I|$ constraints and $|I|p(2 + \overline{N})$ decision variables. There are some noteworthy structural differences as well, such as knapsack constraints (11) that impose area size restrictions and the fact that flow is not area based. Rather, the selection of a unit in (10)-(17) initiates a single unit of flow in constraints (12), whereas the flow initiated in constraints (2) reflects the total area of a unit.

While Murray et al. (2022) used Xpress to solve a subset of scenarios reported in Table 2, with an average optimality gap of 12.68 %, solution using Gurobi offers more meaningful comparison. Table 3 reports Gurobi solution details using model (10)-(17) for each problem scenario detailed in Table 2. In general, the larger planning instances (e.g., $p > 1$ for $|I| = 4{,}696$ and $|I| = 10{,}642$) with millions of constraints and decision variables having double digit or worse optimality gaps are clearly beyond capabilities to identify high quality solutions, at least based on commercial solver used, default parameters and/or current computing configuration.

An interesting and noteworthy characteristic of the developed model, (1)-(9), is that problem size does not change when the number of areas to be identified changes. Comparatively, application problem sizes are generally substantially larger for their approach, effectively limiting problems that can be addressed. Table 3 summarizes problem size and computational limitations in solution. Note in particular that $|I| =$

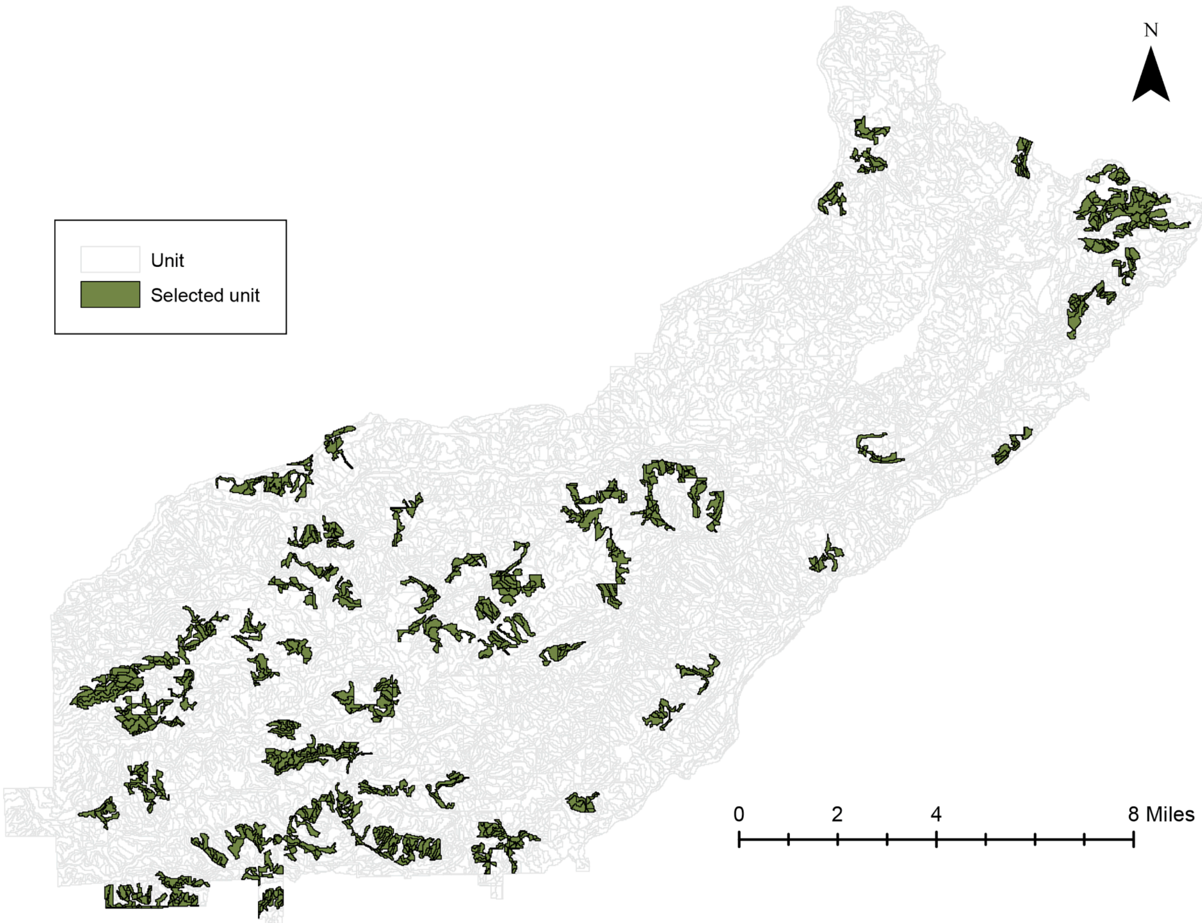

Fig. 5. Selected areas for $|I| = 10{,}642$ ($p = 100$, $T = 100$).

10,642 for $p = 295$ has 9,429,402 constraints and 25,368,820 decision variables. This can be contrasted with 117,923 constraints and 150,708 decision variables observed in Table 2. The result is that Table 3 shows an extremely poor solution after one hour (e.g., a total benefit of 3,558.9994, with an optimality gap of 1,065,191.33 %). Even when computational time is extended to 24 h, the objective improves to only 39,378.1295 with an optimality gap of 54.44 %. This is but a fraction of the best solution found using (1)-(9) of 57,977.1320. Similar comparative findings can be noted for all of the 25 scenarios in Table 3 that could not find and/or verify optimality.

Noted previously was that Ager et al. (2013, 2016) detail a heuristic solution approach. Murray et al. (2022) indicate that the ForSys heuristic of Ager et al. (2013, 2021) was not able to identify solutions as good as those reported in Table 3, though limited in considering only those scenarios up to $|I| = 1{,}343$. The finding was that the heuristic solutions were even worse than the observed solution quality in Table 3. The reality, however, is that heuristics do not give any sort of solution quality bound, so they may be good or not so good, as discussed previously. In this case, the empirical evidence is that the identified solutions using the heuristic of Ager et al. (2013, 2021) are well outside the quality bounds observed in Table 2 using the exact approach reflected in (1)-(9).

The average optimality gap of 1.60 % in Table 2 is rather remarkable, particularly considering results reported in Murray et al. (2022) for smaller size problems (see Table 3). Nevertheless, any observed application scenario optimality gaps suggest opportunity for advancing solution capabilities, either reducing observed gaps or enabling larger more complicated problem instances to be addressed. The literature focused on imposing contiguity in spatial configurations may offer alternative model formulation approaches or perhaps ways to strengthen (1)-(9). Though the work is couched in terms of various assumptions and/or focused on a different underlying problem, opportunity for advancement could include the approaches of Williams (2002), Magnanti and Raghavan (2005), Onal and Briers (2006), Toth et al. (2009), Duque et al. (2011), Carvajal et al. (2013), Haouari et al. (2013), Onal et al. (2016) and/or Pais et al. (2021). Additionally, work in districting may well offer potential as well (e.g., Openshaw 1977, Shirabe 2009, Rios-Mercado 2020). Of course, these are open research questions with no obvious answers at this point, but there are important reasons to explore these as additional considerations may warrant new approaches, such as area proximity to roads, spatial dispersion of identified areas, and others.

5. Conclusions

This paper has highlighted that land use planning is complicated and complex. The use of spatial analytics, and in particular spatial optimization, is critical for accounting for local and regional conditions as well as developing an understanding of decision making impacts associated with land use plans. The focus in this work was multiple area land acquisition, where a desired number of areas offering the greatest benefit are sought that conform to maximum size limitations and where each area is contiguous with respect to the spatial units that it is comprised of. The paper highlights capabilities to address the selection of multiple areas, contiguity constraints that are self-contained and complete, any interpretation of geographic neighbors, and the ability to solve planning applications involving tens of thousands of units.

A new mixed-integer programming formulation of the multiple area

Table 3
Comparative wildfire mitigation results, identified using (10)-(17) and solved using Gurobi.

$\|I\|$	p	Rows	Columns	Objective, (10)	Optimality gap (%)	Time (seconds)
57	1	230	410	312.8903	0.00	0.1
57	2	403	820	506.2098	0.00	6.3
57	3	576	1,230	667.0241	0.00	8.0
130	1	522	974	131.2376	0.00	9.4
130	3	1,306	2,922	357.4882	1.82	3600.0
130	6	2,482	5,844	607.7779	2.49	3600.0
315	1	1,262	2,392	206.5037	0.00	10.3
315	5	5,050	11,960	982.0606	3.37	3600.0
315	11	10,732	26,312	2,031.2943	2.73	3600.0
608	1	2,434	4,776	211.8159	0.00	17.6
608	5	9,738	23,880	1,007.6154	2.44	3600.0
608	10	18,868	47,760	1,878.8248	5.11	3600.0
608	15	27,988	71,640	2,742.9004	2.70	3600.0
608	19	35,302	90,744	3,293.6027	3.59	3600.0
1,343	1	5,374	10,590	284.6195	0.00	3.0
1,343	5	21,498	52,950	1,242.9900	4.13	3600.0
1,343	10	41,653	105,900	2,321.9458	6.22	3600.0
1,343	15	61,808	158,850	3,340.2570	6.00	3600.0
1,343	20	81,963	211,800	4,309.7218	5.69	3600.0
1,343	30	122,273	317,700	6,054.3838	7.11	3600.0
1,343	40	162,583	423,600	7,622.6004	8.42	3600.0
1,343	55	202,893	529,500	9,178.3723	8.02	3600.0
4,696	1	223,048	582,450	9,460.2042	13.10	3600.0
4,696	5	18,786	37,282	293.3076	0.00	68.9
4,696	10	75,146	186,410	1,283.6892	10.83	3600.0
4,696	50	145,596	372,820	2,275.8786	23.12	3600.0
4,696	80	709,196	1,864,100	8,276.8398	38.68	3600.0
4,696	120	1,131,896	2,982,560	12,789.5899	33.44	3600.0
10,642	1	1,695,496	4,473,840	17,033.6758	38.61	3600.0
10,642	50	42,570	85,996	321.5257	0.00	501.8
10,642	100	1,607,042	4,299,800	9,117.5226	47.13	3600.0
10,642	150	3,203,442	8,599,600	15,132.4418	60.68	3600.0
10,642	200	4,799,842	12,899,400	1,838.2025	1,772.94	3600.0
10,642	295	6,396,242	17,199,200	2,664.8241	964,474.41	3600.0

land acquisition problem was detailed, ensuring that each area is contiguous and limited in size. Theoretical properties were derived in order to prove validity of the model. Extensive computational experience was offered through planning application scenario evaluation of USDA Forest Service wildfire mitigation. In total, 34 different planning application scenarios involving problem instances of up to 10,642 spatial units and 295 areas were solved by an exact approach using commercial optimization software. Optimal solutions could be confirmed for 10 scenarios, and where overall solution quality was within 1.60 % of optimum on average for all 34 planning scenarios. These problems represent the largest problem size characteristics addressed to date using an exact approach in land acquisition. The work is an important advance in many ways, including the derived model, problem size invariance associated with areas to be identified, solution capabilities and the overall confirmed quality of solutions found in practice.

The research does suggest a number of avenues for future research. First, there remain challenges for model solution using commercial solvers. One is solution time and the other is closing optimality gaps. While the new model formulation is a major advance, enabling larger problem instances to be addressed and faster solution times, such problems remain a research challenge in terms of effective model structure, branching strategies, etc. as optimality gaps often are encountered in practice. A second area of research is broad assessment and comparison of alternative approaches for imposing contiguity. Alternative approaches do exist, and it may be that combining them in creative ways could prove to offer enhanced model structure for better, faster solution. A third area of research would be exploration of extended modeling considerations, such as the inclusion of a compactness measure. It is unclear whether such extension degrades or improves model solution. A final avenue for future research would be understanding the impacts of network density, and how solvability changes under different application conditions.

CRediT authorship contribution statement

Alan T. Murray: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft. **Richard L. Church:** Conceptualization, Writing – review & editing, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We wish to acknowledge staff of the Stanislaus National Forest, the Social and Ecological Resilience Across the Landscape (SERAL) project team (especially Pete Stine and Carlos Ramirez), and Dr. Patricia Manley of the USFS Pacific Southwest Research Station. We also acknowledge the data processing and analysis efforts of Amelia Pludow at UCSB. Funding for this research through the Pacific Southwest Research Station of the US Forest Service, Department of Agriculture, is gratefully acknowledged (#20-CR-11272139-061).

References

Ager, A.A., Vaillant, N.M., McMahan, A., 2013. Restoration of fire in managed forests: a model to prioritize landscapes and analyze tradeoffs. Ecosphere 4 (2), 1–19.

Ager, A.A., Day, M.A., Vogler, K., 2016. Production possibility frontiers and socioecological tradeoffs for restoration of fire adapted forests. J. Environ. Manage. 176, 157–168.

Benabdallah, S., Wright, J.R., 1992. Multiple subregion allocation models. J. Urban Plann. Dev. 118 (1), 24–40.

Brookes, C.J., 1997. A parameterized region-growing programme for site allocation on raster suitability maps. Int. J. Geogr. Inf. Sci. 11 (4), 375–396.

Brookes, C.J., 2001. A genetic algorithm for designing optimal patch configurations in GIS. Int. J. Geogr. Inf. Sci. 15 (6), 539–559.

Carvajal, R., Constantino, M., Goycoolea, M., Vielma, J.P., Weintraub, A., 2013. Imposing connectivity constraints in forest planning models. Oper. Res. 61 (4), 824–836.

Church, R.L., 2002. Geographical information systems and location science. Comput. Oper. Res. 29 (6), 541–562.

Church, R.L., Gerrard, R.A., Gilpin, M., Stine, P., 2003. Constructing cell-based habitat patches useful in conservation planning. Ann. Assoc. Am. Geogr. 93 (4), 814–827.

Cova, T.J., Church, R.L., 2000. Contiguity constraints for single-region site search problems. Geogr. Anal. 32 (4), 306–329.

Duque, J.C., Church, R.L., Middleton, R.S., 2011. The p-regions problem. Geogr. Anal. 43 (1), 104–126.

Duque, J.C., Anselin, L., Rey, S.J., 2012. The max-p-regions problem. J. Reg. Sci. 52 (3), 397–419.

Fischer, D.T., Church, R.L., 2003. Clustering and compactness in reserve site selection: an extension of the biodiversity management area selection model. For. Sci. 49 (4), 555–565.

Haouari, M., Maculan, N., Mrad, M., 2013. Enhanced compact models for the connected subgraph problem and for the shortest path problem in digraphs with negative cycles. Comput. Oper. Res. 40 (10), 2485–2492.

Li, W., Church, R.L., Goodchild, M.F., 2014. The p-compact-regions problem. Geogr. Anal. 46 (3), 250–273.

Ligmann-Zielinska, A., Church, R.L., Jankowski, P., 2008. Spatial optimization as a generative technique for sustainable multiobjective land-use allocation. Int. J. Geogr. Inf. Sci. 22 (6), 601–622.

Magnanti, T.L., Raghavan, S., 2005. Strong formulations for network design problems with connectivity requirements. Networks 45 (2), 61–79.

Murray, A.T., Church, R.L., Pludow, B.A., Stine, P., 2022. Advancing contiguous environmental land allocation analysis, planning and modeling. J. Land Use Sci. 17 (1), 572-590.

Nalle, D.J., Arthur, J.L., Sessions, J., 2002. Designing compact and contiguous reserve networks with a hybrid heuristic algorithm. For. Sci. 48 (1), 59–68.

Onal, H., Briers, R.A., 2006. Optimal selection of a connected reserve network. Oper. Res. 54 (2), 379–388.

Onal, H., Wang, Y., Dissanayake, S.T., Westervelt, J.D., 2016. Optimal design of compact and functionally contiguous conservation management areas. Eur. J. Oper. Res. 251 (3), 957–968.

Openshaw, S., 1977. Optimal zoning systems for spatial interaction models. Environ Plan A 9 (2), 169–184.

Pais, C., Carrasco, J., Elimbi Moudio, P., Shen, Z.-J., 2021. Downstream Protection Value: Detecting critical zones for effective fuel-treatment under wildfire risk. Comput. Oper. Res. 131, 105252.

Rios-Mercado, R.Z. (Ed.), 2020. Optimal Districting and Territory Design. Springer Nature.

Shirabe, T., 2005. A model of contiguity for spatial unit allocation. Geogr. Anal. 37 (1), 2–16.

Shirabe, T., 2009. Districting modeling with exact contiguity constraints. Environ. Plann. B. Plann. Des. 36 (6), 1053–1066.

Tóth, S.F., Haight, R.G., Snyder, S.A., George, S., Miller, J.R., Gregory, M.S., Skibbe, A. M., 2009. Reserve selection with minimum contiguous area restrictions: An application to open space protection planning in suburban Chicago. Biol. Conserv. 142 (8), 1617–1627.

Wang, Y., Onal, H., Fang, Q., 2018. How large spatially-explicit optimal reserve design models can we solve now? An exploration of current models' computational efficiency. Nature Conservation 27, 17.

Williams, J.C., 2002. A zero-one programming model for contiguous land acquisition. Geogr. Anal. 34 (4), 330–349.

Wright, J., ReVelle, C., Cohon, J., 1983. A multiobjective integer programming model for the land acquisition problem. Reg. Sci. Urban Econ. 13 (1), 31–53.

Xiao, N., Murray, A.T., 2019. Spatial optimization for land acquisition problems: A review of models, solution methods, and GIS support. Trans. GIS 23 (4), 645–671.