Spatially Constrained Geodesign Optimization (GOP) for Improving Agricultural Watershed Sustainability

Yiqun Xie[1], **KwangSoo Yang**[2], **Shashi Shekhar**[1], **Brent Dalzell**[1], **and David Mulla**[1]
[1]University of Minnesota, [2]Florida Atlantic University
xiexx347@umn.edu, yangk@fau.edu, shekhar@umn.edu, bdalzell@umn.edu, mulla003@umn.edu

Abstract

Given an agricultural watershed containing a set of spatial units, and a set of land management practices, the Geodesign Optimization (GOP) aims to find a land management practice for each spatial unit that optimizes overall water quality improvements in the watershed under both budget constraint and spatial constraints (e.g., minimum contiguous area, shape) arising from farm equipment operation practicalities. GOP is important for redesign of agricultural watersheds in Midwestern US to mitigate soil and water quality degradation and loss of habitat. The problem is computationally challenging as a large-scale combinatorial problem (NP-hard) under spatial constraints. Existing optimization techniques do not address spatial constraints, and lead to impractical solutions requiring frequent farm equipment reconfiguration. In this paper, we formalize the spatially-constrained GOP and propose a novel spatial optimizer which explores optimal solution without constraint violations. Our approach is further validated through a Geodesign case study at Seven Mile Creek watershed in Midwestern US.

1 Introduction

Given an agricultural watershed composed of a set of grid cells, and a set of land management practices, Geodesign Optimization problem (GOP) aims to find a choice of land management practice (LMP) for each grid cell that optimizes overall water quality in the watershed, in order to improve sustainability of food and clean water production.

Societal motivation: Rising human population and decreasing poverty require that the world produce enough food to feed another 3 billion people by 2100 (United Nations 2015). Incentives for growing crops have depleted water resources (e.g., Aral Sea, Ogallala aquifer) and affected water quality (e.g., dead zone in Gulf of Mexico) (Eftelioglu et al. 2016). Furthermore, soil erosion which brings sediments and fertilizer residues into rivers also imposes challenges on agriculture, which consumes over 65% of surface fresh water in US (Pimentel et al. 2004). Current agricultural landscape in Midwestern US has led to severe soil and water quality degradation and loss of habitat (The Guardian 2014). Due to the water quality issue, Minnesota has enacted a new state law to enforce vegetative buffers with at least 50-feet width on average along public waters (Laws-Of-Minnesota 2016) (e.g., grass buffer along rivers, lakes). An opportunity exists to better balance food production and environmental benefits by reallocating land management practices. Geodesign Optimization (GOP) aims to address this land allocation problem, and has been recognized by stakeholders and domain scientists as important for improving landscapes in agricultural watersheds (Schively et al. 2016). For the ease of farm equipment operations, the solution needs to avoid spatial fragmentation after reallocation of land management practices. For example, each patch of allocated land management practice should be large enough so that farm equipment can operate smoothly without frequent reconfigurations.

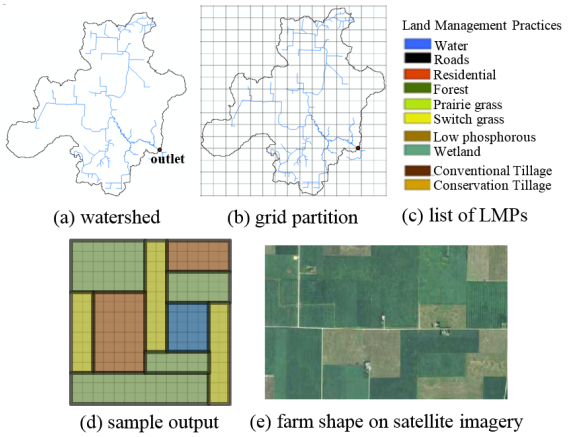
Challenges: GOP is NP-hard (proved in Sec. 3) and is spatially constrained. The problem is computationally challenging due to large number of decision variables needed for real-world watershed.

Related work: In traditional agriculture, land management decisions are mostly made at individual or community level through farming experience. In the worst case scenario, choices of land management practices are made randomly or voluntarily. In case of pollution (e.g., water pollution), changes are heuristically targeted on locations (e.g., steep slope near water bodies) which produce the highest amount of pollution (Galzki, Birr, and Mulla 2011). In addition, existing optimization techniques either cannot address the hard spatial contiguity constraints or yield outcomes far from optimal. Space partitioning techniques (e.g., R-TILE) are also not readily applicable to GOP. Details are discussed in Sec. 4.

Contribution: First, we formulate GOP that honors spatial constraints. Second, we prove GOP is NP-hard. Third, we propose a novel spatial optimizer that heuristically explores optimal solution with no violation of hard spatial constrains.

Validation: A case study at Seven Mile Creek watershed (Minnesota, US) shows our proposed spatial optimizer narrows down the gap to upper-bounding optimal solution with all spatial constraints addressed.

Outline: Sec. 2 defines the Geodesign Optimization problem (GOP). Sec. 3 analyzes the hardness of GOP. Sec. 4 summarizes related work on optimization within a spatial context. Sec. 5 presents a novel spatial optimizer to solve spatially constrained GOP and Sec. 6 evaluates the proposed approach via a detailed case study.



(a) watershed (b) grid partition (c) list of LMPs

(d) sample output (e) farm shape on satellite imagery

Figure 1: Examples of input and output of GOP.

2 Problem statement

2.1 Basic Concepts

Watershed: A watershed (Fig. 1(a)) is a drainage basin, where all surface water converges to a single outlet.

Land management practice (LMP): Land management practice is an agricultural choice on a piece of land. A land management practice has many implications, including cost of investment, change in water quality, etc. Examples of land management practices include conservation tillage, low phosphorous, switch grass, etc (Fig. 1(c)). A land management practice has many implications, including cost of investment, change in water quality, etc.

Soil and Water Assessment Tool (SWAT): SWAT model (Gassman, Reyes, and et al. 2007) operates on a watershed-scale and offers a quantitative approach to simulate the effects of land management practices on environmental quality under a range of conditions (e.g., soil, slope, weather). Inputs of SWAT model include physical information on climate, topography and soil (e.g., raster data layer of soil) as well as land management practices. In this paper, SWAT is used to quantify water quality in the watershed.

InVEST model: InVEST model (Nelson et al. 2009) uses ecological production functions and economic valuation methods to estimate the values of LMPs at different locations. In this paper, InVEST model is used to quantify economic cost of changing LMPs in the watershed.

Grid partition: The study area is discretized into grid-cells. Each cell can only have a single LMP in the output. A grid cell has two coefficients for each LMP choice: 1) cost of investment (cost c), and 2) water quality improvement (benefit b). The coefficients are precomputed for all LMP choices across all grid cells using SWAT model. Each grid cell $g_{ij} = (i, j)$ is identified by its row i and column j.

Tile: A tile is a rectangular region consisting of a group of adjacent grid cells. All grid cells in each tile share the same LMP. Each tile $t = (i_0, j_0, i_1, j_1)$ is identified by its bottom-left grid cell (i_0, j_0) and top-right grid cell (i_1, j_1), where $i_1 > i_0, j_1 > j_0$. Cost c and benefit b of each tile are aggregated from all contained grid cells by summation.

Tiling scheme: A tiling scheme is a tile-partition of the study area. Tiles in a tiling scheme must not have overlaps. Fig. 1(d) shows a sample output of GOP with a tiling scheme and LMP assignments by colors in Fig. 1(c).

2.2 Scope and approximation

Grid Partition. Data (e.g., elevation, soil property) collected from site samples or satellites/UAVs require space discretization. Thus, the study area is approximated by its grid partition and all inputs and outputs are expressed by this discretization. **Spatial constraints.** In a map of LMP decisions, adjacent cells with the same LMP form a patch. In GOP, spatial constraints are imposed on each patch for farm operation practicality. Minimum area constraint is used to avoid spatial fragmentation of LMP. Shape constraint is used to make sure the boundary of each patch is regular in shape. Since farm equipment is normally operated to scan farmland in straight lines and the shape of farmland is often rectangular as shown in Fig. 1(e), the shape constraint used in this paper is set as rectangle as an approximation. This constraint may not be appropriate for farms using central pivot irrigation systems (e.g., circular shaped farms in Kentucky, US), which may be explored in future work. We assume that some land cover types within a watershed are not changeable (e.g., roads, water bodies). The boundaries of these unchangeable places and the boundary of the watershed are often irregular in shape. In this paper, the spatial constraints are relaxed at the boundary of these places. **Rectangular study area.** Finally, to simplify the problem definition and method illustration, the study area used in this paper is rectangular in shape. For watersheds with irregular boundaries, the study area used is either a rectangular bounding box or a rectangular sub-region of the watershed. To deal with empty variables in the rectangular study area (e.g., grid cells outside the watershed or within unchangeable lands), their coefficients are all set to 0 to eliminate their impact on the result. The empty variable effects can also be removed through preprocessing.

2.3 Optimization problem definition

The spatially-constrained GOP is formally formulated as:
Inputs:

- A rectangular study area and its grid partition G, which is a collection of grid cells $\{g_{ij}\}$;
- A list L of choices of land management practices $\{LMP_k\}$;
- A list T of all possible tiles with each choice of LMP: $T = \{t_m = (i_0, j_0, i_1, j_1, k) | \forall (i_0, j_0), (i_1, j_1) \in G, k \in L, (i_0, j_0) \neq (i_1, j_1)\}$. Tiles with the same (i_0, j_0, i_1, j_1) are adjacent in T in ascending order of k;
- A three dimensional matrix $CMAT$ of cost c_{ijk} (cost of investment) for each grid g_{ij} with LMP k; and a matrix $BMAT$ of benefit b_{ijk} (water quality improvement);
- A budget ε, minimum tile area α, minimum tile width β.

Output: A tiling scheme with LMP assignment, denoted by a binary vector S, where each element s_m corresponds to t_m in T. If t_m is chosen, $s_m = 1$; otherwise $s_m = 0$.

Objective: Find the binary vector S that maximizes overall benefit:

$$\max_{S} \sum_{m=1, m\in S}^{|S|} b_m \cdot s_m \tag{1}$$

where $b_m = \sum_{i=i_0}^{i_1} \sum_{j=j_0}^{j_1} b_{ijk}$, and i_0, j_0, i_1, j_1 are defined by tile t_m, and $k = mod(m + |L| - 1, |L|) + 1$, which is the LMP index of m^{th} element ("$m + |L| - 1$" is used to cover m that is a multiple of $|L|$). $|L|$ denotes the length of L.

Constraints:

– Binary value constraint on elements in S:

$$s_m \in \{0, 1\} \tag{2}$$

– Total cost is less than ε:

$$\sum_{m=1, m\in S}^{|S|} c_m \cdot s_m \leqslant \varepsilon \tag{3}$$

where c_m is computed in the same way as b_m in (1).

– Each tile t_m has a minimum area α and width β:

$$i_1 - i_0 \geqslant \beta, i_1, i_0 \in t_m \tag{4}$$

$$j_1 - j_0 \geqslant \beta, j_1, j_0 \in t_m \tag{5}$$

$$(i_1 - i_0) \cdot (j_1 - j_0) \geqslant \alpha, i_1, i_0, j_0, j_1 \in t_m \tag{6}$$

– There is no overlap among tiles:

$$\forall t_1, t_2 \in (T \cap S), t_1 \cap t_2 = \phi \tag{7}$$

– The study area is covered by all tiles:

$$\forall g_{ij}, \exists t_m \in (T \cap S), s.t. g_{ij} \in t_m \tag{8}$$

Constraints (4) to (8) are spatial constraints imposed on tiles for practicality concerns (e.g., ease of farm equipment operation and farm scanning). Constraints (4) to (6) are tile-specific constraints, whereby each valid tile must satisfy a minimum area α and width β constraint. For implementation, these constraints can be imposed by preprocessing inputs. For example, before optimization, if a tile t_m in T does not satisfy the area and width constraint, it is removed from list T. Constraints (7) and (8) are spatial constraints imposed on a group of tiles. These constraints are not independent for each tile and there are mutual interactions among tiles. (7) and (8) are expressed at a conceptual level for simplicity. Their exact mathematical formulations are the following.

Constraint (7) requires any pair of tiles t_1 and t_2 in T are disjoint in grid partition G. t_1 and t_2 are disjoint only if no vertex of one tile falls within the other. Thus, for each vertex (i, j) of tile t_1 and the region (i_0, j_0, i_1, j_1) defined by t_2, the following constraints are added:

$$(i - i_0) \cdot (i - i_1) > 0 \tag{9}$$

$$(j - j_0) \cdot (j - j_1) > 0 \tag{10}$$

Constraint (9) shows i is either smaller than i_0 or greater than i_1, and constraint (10) does the same for j. For constraint (8), it is detailed by the following constraint:

$$\sum_{m=1, m\in S}^{|S|} (i_1^m - i_0^m) \cdot (j_1^m - j_0^m) = (i_{max} - i_{min}) \cdot (j_{max} - j_{min}) \tag{11}$$

In (11), $i_{\{0,1\}}^m$, $j_{\{0,1\}}^m$ represent $i_{\{0,1\}}$, $j_{\{0,1\}}$ for tile t_m, and $i_{\{max,min\}}$, $j_{\{max,min\}}$ define the boundary of the study area. A combination of (10) and (11) guarantees the study area is covered by all tiles.

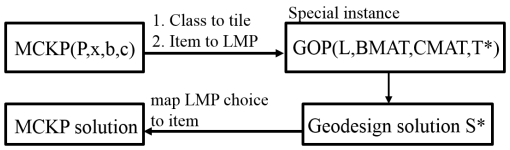


Figure 2: Reduction of MCKP to GOP.

3 Challenges

NP-hardness. The NP-hardness proof follows the well-known result of the multiple choice knapsack problem (MCKP)(Kellerer, Pferschy, and Pisinger 2004).

Definition 1 (Multiple choice knapsack problem)**.** *Given m classes $P_1, P_2, ..., P_m$ of items to pack in a knapsack of capacity ε. Each item $x_{ij} \in P_i$ has two coefficients, a benefit b_{ij} and a cost (e.g., weight) c_{ij}. The MCKP problem is to choose one item from each class so that the overall benefit is maximized without exceeding the cost capacity ε.*

$$\max_{x} \sum_{i=1}^{m} \sum_{j\in P_i} b_{ij} \cdot x_{ij} \tag{12}$$

$$subject\,to, \sum_{i=1}^{m} \sum_{j\in P_i} c_{ij} \cdot x_{ij} \leqslant \varepsilon \tag{13}$$

$$\sum_{j\in P_i} x_{ij} = 1,\ i = 1, 2, ..., m \tag{14}$$

$$x_{ij} \in \{0, 1\}, i = 1, 2, ..., m, j \in P_i \tag{15}$$

Theorem 1. *No polynomial-time algorithm exists for Geodesign Optimization if $P \neq NP$.*

Proof. Assume $P \neq NP$, and the number of items is l in each class P_i of MCKP. We construct a mapping from this MCKP problem to a special instance of GOP, in which we assume all input tiles in list T satisfy all the spatial constraints (4) to (8) as inputed (e.g., tiles in T are spatially the same as the optimal tiling scheme but LMP choices remain unknown). Given a $MCKP(P, x, b, c)$, the special instance of GOP can be constructed in polynomial time by assigning all b_{ij}, c_{ij} in MCKP to $BMAT$, $CMAT$ in GOP, so that each $x_{ij} \in P_i$ is translated to a LMP choice on a tile. If GOP can be solved in polynomial time, this special instance can also be solved and we get the optimal vector S. The optimal solution of MCKP can be then constructed by mapping each LMP decision to each item in class P_i, which means MCKP is solved in polynomial time. Since we assume $P \neq NP$, GOP cannot be solved in polynomial time (Fig. 2). □

Spatial constraints. More than MCKP, GOP requires additional combinatorial work on space tiling under the spatial constraints. Constraints (4)-(6) are tile-independent and can be preprocessed to remove tiles in T. Constraints (7) and (8) are global constraints that consider relationships among tiles. The mutual influence each tile has on another poses a challenge to satisfy these constraints. For example, during optimization, the replacement of a tile may be necessary to improve benefit. However, replacing one tile will directly

lead to violations of constraint (7) (or (8)). To resolve the violations, new replacements are necessary but they may lead to further violations. This tight mutual impact among tiles make spatial constraints challenging to address.

Large scale watershed. GOP needs to be performed on large scale watershed (e.g., over millions of grid cells). Construction of T as an input is necessary in order to mathematically formulate spatial constraints. Suppose the grid-partition has a dimension of p by q, and the length of LMP list is l. The length $|T|$ of T is $(p \cdot q \cdot l)^2$. For $(p, q, l) = (1000, 1000, 5)$, $|T| = 2.5 \times 10^{13}$. More challengingly, the number of spatial constraints required can be magnitudes greater than the number of variables.

4 Related work

Fig. 3 shows a classification tree on related optimization work. Two previous efforts on Geodesign have been explored: 1) assisted stakeholder-collaborative geodesign (CG) (Schively et al. 2016) and 2) conventional mixed integer programming (MIP) (Lodi 2010) method without spatial constraints. CG provides a visualization tool that helps multiple stakeholders negotiate the inevitable trade-offs that arise when changes in policy, planning or management are introduced to agricultural watershed. Through the CG visualization tool, stakeholders can set performance goals (e.g., water quality improvement), evaluate several hand-crafted design scenarios, compare tradeoffs and work towards consensus. Fig. 4 shows design maps from CG (F,G,H,I) and MIP (A,B,C,D,E). Solutions from MIP were computed under different budget constraints (e.g., $1 million). In Fig. 4, water quality improvements is measured by sediment reduction. CG provides a software to help stakeholders analyze and create design solutions. The CG solutions yield practical designs for farm equipment operation but water quality achieved are far from optimal. On the contrary, MIP algorithms without spatial constraints, yield optimal water quality but the solution map is spatially fragmented and stakeholders are not willing to implement it.

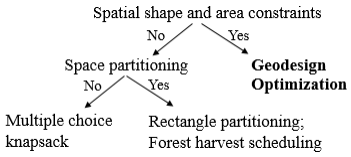


Figure 3: Related work of GOP.

Spatial constraints are studied in forest harvest scheduling problems (FHSP) (McDill, Rebain, and Braze 2002). The spatial constraints imposed in FHSP enforce spatial repulsion among harvest sites (Murray and Weintraub 2002). Differently, GOP enforces spatial contiguity of all types of land management practices. Thus, FHSP techniques currently are not applicable to GOP. It is worth mentioning in FHSP, challenges are also mainly posed by spatial constraints, which tend to be exponential in number of variables.

Rectangular partitioning problems (e.g., R-TILE, R-PACK) have also been studied in optimization (Muthukrishnan, Poosala, and Suel 1999; Berman et al. 2001). These problems and variants only have one value for each grid cell, and all values are included in the solution, which means there is no multiple choice to make on grid cell level. Thus, techniques in R-TILE or R-PACK, which mainly utilize this single-choice characteristic, cannot be directly applied to GOP. In addition, rectangular partitioning problems do not consider minimum area and width constraints.

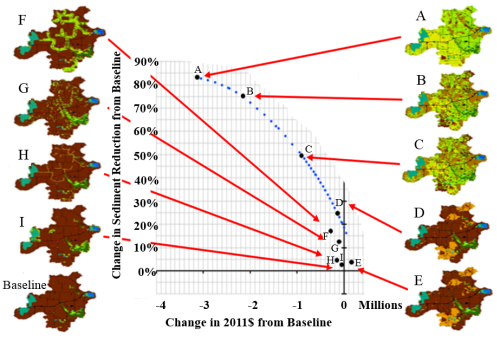


Figure 4: Solutions from CG (left: F,G,H,I) and MIP (right: A,B,C,D,E). LMPs are colored by legend in Fig. 1(c).

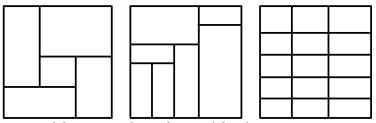


Figure 5: Space tiling frameworks.

5 Geodesign Optimization (GOP)

We introduce two novel approaches for GOP: 1) dynamic growth tiling framework (DGTF) and 2) spatial optimizer.

5.1 Dynamic growth tiling framework (DGTF)

There are three classes of rectangular space tiling frameworks in general (Fig. 5) (Muthukrishnan, Poosala, and Suel 1999), namely 1) arbitrary; 2) hierarchical; and 3) $p \times q$. For GOP, the optimal tiling scheme can be arbitrary. Although the optimal solution is not expected to be found within tractable time, it is still ideal to guarantee it is within the potential search space of the optimizer.

For $p \times q$ tiling framework, the output has p rows and q columns. In this case, it does not cover tiling schemes that have more than p rows or q columns at places, as well as those with a prime total number of tiles. For hierarchical tiling framework, straight lines are used in each iteration to bi-partition a rectangular region. Suppose the study area has height h and width w. At the root level, a straight line is used to partition the entire study area into two halves. The length of this split line is at least $min(h, w)$. More generally, at level lv (root level $lv_{root} = 1$), the length of the split line is at least $\frac{min(h,w)}{2^{lv-1}}$. In an arbitrary tiling scheme, there may not exist a straight line (formed by adjacent tile sides) with such long lengths. As the minimum area and

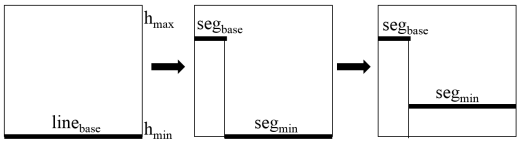


Figure 6: Example steps of DGTF.

width constraints get smaller, it may get harder to form these long straight lines in a tiling scheme.

Arbitrary tiling framework is ideal for this case. To the best of our knowledge, there is yet no systematic approach that is able to reproduce any arbitrary tiling scheme that satisfies spatial constraints. Thus, we propose a dynamic growth tiling framework (DGTF), which gives a procedural way that can produce any arbitrary tiling schemes.

DGTF contains two phases: 1) initialization and 2) growth. **In initialization phase**, one side of the rectangular study area is set as the global baseline $line_{base}$ for growth. The direction of $line_{base}$ is considered as horizontal and height of $line_{base}$ is minimal ($h_{min} = 0$). The height at the opposite side of the study area is maximal (h_{max}). **In growth phase** (Fig. 6), individual constrained tiles are iteratively growed to fill up the study area in a sequential manner from "low to high". A new term, base-segment seg_{base}, denotes a horizontal line segment formed by adjacent tile boundary parts (including $line_{base}$ part) that is not yet covered by any tiles. Minimum base-segment seg_{min} denotes the base-segment with the lowest height (h^{seg}_{min}). The growth phase works as follows. **Step 1:** Find the current seg_{min}. If $d_l = length(seg_{min}) \geq \beta$, $d_h = (h_{max} - h^{seg}_{min}) \geq \beta$ and $d_l \cdot d_h \geq \alpha$, go to Step 2-1; otherwise go to Step 2-2; **Step 2-1:** Grow a constrained tile at the left vertex of seg_{min}. The tile is constrained means it satisfies the minimum area α and width β constraints. **Step 2-2:** Expand an adjacent constrained tile of seg_{min} to fill up seg_{min} or the vertical space left, depending on which part violates α or β. **Step 3:** Update the set of seg_{base} to reflect the new growth, and then terminate if $h^{seg}_{min} = h_{max}$ after update, or otherwise continue with Step 1. Fig. 6 provides a toy example illustrating the steps of DGTF.

Theorem 2 shows DGTF can be used to generate any arbitrary tiling schemes. This does not mean DGTF will form the optimal arbitrary tiling scheme but it guarantees this optimal tiling scheme is contained in its potential search space. In other words, following the sequential procedure of DGTF, any arbitrary tiling scheme can be reproduced, which cannot be done with hierarchical and $p \times q$ tiling frameworks. In real implementation, how each tile is growed in Step 2-1 is determined by a specific user-specified guidance function.

Theorem 2. *All arbitrary tiling schemes are contained by the potential search space of DGTF.*

Proof. Define a stage variable v (integer) of DGTF, which increases by 1 each time a new tile is growed. Given a spatially constrained tiling scheme, DGTF reproduces it by sequentially adding tile t on seg_{min} at each stage v. If a tiling scheme cannot be reproduced, then there must be a stage v_u, such that the next tile t_{u+1} cannot be put onto current seg_{min}. If v_u exists, then any tile $t_{u' \leq u}$ that contains seg_{min} (or part of seg_{min}) as a side must have no adjacent tile sharing seg_{min} in the given tiling scheme. Since the study area is entirely covered by tiles, seg_{min} must have $h^{seg}_{max} = h_{max}$. Thus, the next tile t_{u+1} must not exist which contradicts the original set up. Thus, v_u does not exist. □

5.2 Spatial optimizer

The spatial optimizer is a heuristic optimizer built on top of DGTF, which tackles the large combinatorial problem by heuristically generating constrained tiles to form output vector S. The spatial optimizer has three major components: 1) linear programming (LP) relaxation; 2) DGTF-based space tiling; 3) land management practice (LMP) rearrangement.

LP relaxation. Due to the hardness of GOP, it is not expected to find the optimal solution sol^*. However, an upper bound $\widehat{sol^*}$ on the optimal solution can be efficiently computed (e.g., interior point method (Karmarkar 1984)) through a relaxation on spatial and binary integer constraints ($\{0, 1\}$ set to $[0, 1]$ interval). To grow a tile in DGTF, a guidance function $f(i_1, j_1, k)$ (k is a LMP choice) is necessary to determine which (i_1, j_1, k) yields the locally optimal tile for the current growth step anchoring at (i_0, j_0). $\widehat{sol^*}$ is used in $f(i_1, j_1, k)$ to give global-based hints to DGTF.

DGTF-based space tiling. The growth of each tile t starts at a fixed vertex (i_0, j_0). A set of candidate tiles defined by (i_1, j_1)s need to be enumerated to determine which maximizes $f(i_1, j_1, k)$:

$$f(i_1, j_1, k) = \frac{B(i_1, j_1, k)/B^*(i_1, j_1)}{C(i_1, j_1, k)/C^*(i_1, j_1)} \tag{16}$$

where $B(i_1, j_1, k) = \sum_{i=i_0}^{i_1} \sum_{j=j_0}^{j_1} b(i, j, k)$ and $C(i_1, j_1, k) = \sum_{i=i_0}^{i_1} \sum_{j=j_0}^{j_1} c(i, j, k)$ are the sums of benefit and cost for the candidate tile, and $B^*(i_1, j_1) = \sum_{i=i_0}^{i_1} \sum_{j=j_0}^{j_1} b^*(i, j)$ and $C^*(i_1, j_1) = \sum_{i=i_0}^{i_1} \sum_{j=j_0}^{j_1} c^*(i, j)$ are sums of benefit and cost from $\widehat{sol^*}$ for the same candidate. $b^*(i, j), c^*(i, j)$ are the benefit and cost achieved at (i, j) in $\widehat{sol^*}$.

The guidance function $f(i_1, j_1, k)$ is a normalized ratio of $\frac{benefit}{cost} = \frac{B(i_1, j_1, k)}{C(i_1, j_1, k)}$ of a candidate tile using hints from $\widehat{sol^*}$. For each candidate corner grid cell (i_1, j_1), $f(i_1, j_1, k)$ increases monotonically as $B(i_1, j_1, k)$ increases and decreases as $C(i_1, j_1, k)$ increases. The normalization with $B^*(i_1, j_1)$ and $C^*(i_1, j_1)$ is used to avoid $B(i_1, j_1, k)$ and $C(i_1, j_1, k)$ derailing far from that of $\widehat{sol^*}$. The local optimal (i_1, j_1, k) selected is:

$$(i_1, j_1, k) = \begin{cases} arg \max\limits_{i_1, j_1, k} f(i_1, j_1, k), & \forall \frac{C(i_1, j_1, k)}{C^*(i_1, j_1)} \leq 1 \\ arg \min\limits_{i_1, j_1, k} C(i_1, j_1, k), & \text{otherwise} \end{cases} \tag{17}$$

In Eq. (17), the condition is used to control the overflow on cost. Since $\widehat{sol^*}$ is obtained using LP relaxation, its cost

is expected to be tightly under budget given solution is feasible. Thus, confining $C(i_1, j_1, k)$ by bound $C^*(i_1, j_1)$ helps keep overall investment within budget during optimization. If there exists no (i_1, j_1, k) satisfying this local cost constraint, the one with minimum cost is selected.

To efficiently compute the summations in $f(i_1, j_1, k)$, integral images (Viola and Jones 2004) are precomputed in linear time to evaluate each summation (e.g., $B(i_1, j_1, k)$) in $O(1)$ time.

LMP rearrangement. In this final 2-phase step, the output sol_{DG} of DGTF-based tiling scheme generation is used as an initial seed. **Phase 1 - Constraint violation elimination:** In this phase, budget violation, if any, is eliminated by tile-level LMP adjustment. Denote $cost_{DG}$, $bene_{DG}$ as the total cost and benefit achieved by sol_{DG}, and denote $cost(t, k)$, $bene(t, k)$ as the cost and benefit of a single tile t in sol_{DG} with LMP choice k. If k is the same as the LMP choice in sol_{DG}, it is denoted as k_{DG}. In each iteration a single tile (t, k) is selected: $(t, k) = arg\max_{t,k}\ bene_{DG} + bene(t, k) - bene(t, k_{DG})$, where $(t, k) \in set_{tk} = \{(t, k)|cost(t, k)/cost(t, k_{DG}) < 1\}$. The selected tile t and alternative LMP choice k on t guarantees the adjusted benefit is maximized given that total cost is reduced by this tile-level adjustment. The adjustment terminates as soon as the budget constraint is satisfied. If $set_{tk} = \phi$ and budget constraint ε is still violated, the solution is not feasible given sol_{DG}. **Phase 2 - Pairwise trading:** In this phase, the spatial optimizer explores pairs of tiles (t, t') in the current solution, and potential change of LMP assignments on t and t' in order to further improve the objective function without violation of budget constraint ε. For a pair of tiles, reassigning their LMPs create a pairwise trading of cost and benefit. In each iteration, a pair of tiles (t, t') and their LMP choices (k, k') are selected. Denote $set_{tt'kk'}$ as a set of (t, t', k, k')s that can improve the total benefit while keeping the total cost under budget ε. The (t, t', k, k') is selected as: $(t, t', k, k') = arg\max_{t,t',k,k'}\ [bene(t, k) - bene(t, k_{DG})] + [bene(t', k') - bene(t', k'_{DG})]$, where $(t, t', k, k') \in set_{tt'kk'}$. Pairwise trading terminates when $set_{tt'kk'} = \phi$.

6 Validation

We evaluate the performance of the proposed spatial optimizer through a detailed case study at Seven Mile Creek watershed in Midwestern, US. The watershed drains directly to the Minnesota River, which is a tributary of the Mississippi River. Although the watershed has an average slope of less than 2%, its flat upland transitions quickly into a ravine-zone before draining into the Minnesota River. The current crop landscape in the Seven Mile Creek watershed is mostly corn and soy bean, similar to landscape across Midwestern US agricultural areas. Conventional LMP choices on this landscape does not produce a well-balanced design between productivity and environment quality, and has led to soil and water quality degradation, as well as loss of habitat.

Experiment setup. The study area is a rectangular 5760-acre subset of the Seven Mile Creek watershed, and is partitioned into 30×30 square-foot grid cells. The dimension of the grid is 680×410, which has a quarter million grid cells. Five land management practices (LMP) are considered: conservation tillage, low phosphorous, prairie grass, switch grass and stover (all selected by local agricultural scientists). Water quality improvement (WQI) is measured by percentage of sediment reduction in water. WQI (benefit b_{ijk}) and cost of investment (cost c_{ijk}) are calculated from detailed Soil and Water Assessment Tool (SWAT) modeling.

The minimum area α and width β constraints are varied in the case study to show the trend of water quality improvement as α and β increase. Unit of α and β is in grid-cell. Seven sets of (α,β) are tested (Fig. 7). For better visualization, adjacent tiles sharing a common boundary are merged. The budget for all experiments is $100,000. Fig. 7(a) shows the result of LP relaxation, in which the LMP choice colored for each grid cell is the one with the largest value. We can see the tiling schemes generated by DGTF attempt to approximate the original distribution of LMP patches in $\widehat{sol^*}$ with all spatial constraints satisfied. When (α, β) is set to $(8, 2)$, which is about $(800m^2, 20m)$, WQI shows only a 1.99% gap to $\widehat{sol^*}$ (63.61% to 65.60%), which is an acceptable difference considering 65.60% is a relaxed upper bound. The reduction of water quality improvement is gradual as (α, β) increases. As α and β increase, more LMP choices are forced to be made at non-optimal locations (expensive in this case study) due to the extended spatial contiguity constraints. There is a big decrease from Fig. 7(g) to 7(h), which makes the water quality improvement of 7(h) below half of $\widehat{sol^*}$. One explanation is the difficulty to include more constrained tiles which can improve the water quality without exceeding the budget limit due to large (α, β) values.

Runtime is measured on a 64-bit Window 8 laptop with CORE i7 (Fig. 7). The longest runtime is 201.3 seconds for $(\alpha = 8, \beta = 2)$. Runtime is mainly affected by total number of growth in DGTF, and the size of search space of each growth step. The search space of each growth is limited by the length of seg_{min} (Fig. 6). As α increases, total number of growth and number of tiles in LMP rearrangement is reduced. As β increases, seg_{min} tends to increase which enlarges the search space. Since both α and β are increasing, the runtime shows multiple peaks that reflects the interplay of the two variables.

7 Conclusion and Future Work

In this paper, we formulate the spatially constrained GOP as a combinatorial optimization problem. The problem is important for landscape redesign towards improving the sustainability of food and clean water production in agricultural watersheds, which has triggered broad interests among real stakeholders and domain scientists. GOP is computationally challenging as a NP-hard problem with a huge volume of decision variables. We propose a spatial optimizer which approaches the upper-bounding optimal solution without violation of hard spatial constraints. In a case study, the spatial optimizer is able to narrow the gap to the upper bound of the optimal solution. The future work includes: 1) explore tighter upper bound and lower bound on solution quality to evaluate the performance of spatial optimizer more accurately; 2) explore acceleration of exact algorithms to solve

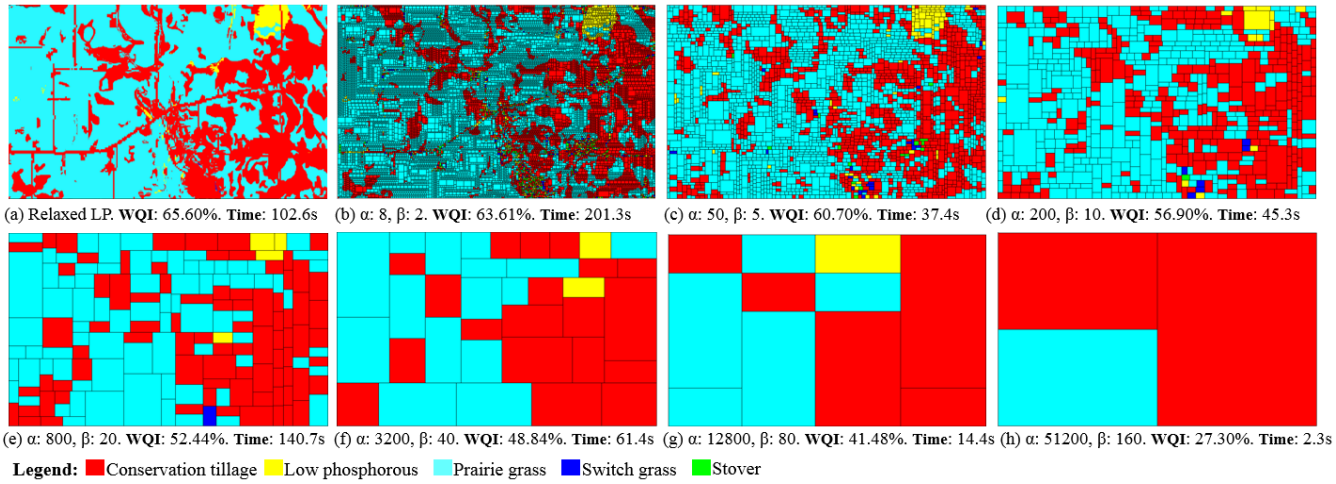


Figure 7: Results on water quality improvements (WQI) and runtime of the spatial optimizer with varying spatial constraints.

special cases of GOP in feasible time and reach global optimality; and 3) explore acceleration of the proposed spatial optimizer to solve GOP on a larger scale (e.g., big data).

8 Acknowledgments

This material is based upon work supported by the National Science Foundation under Grants No. 1541876, 1029711, IIS-1320580, 0940818 and IIS-1218168, the USDOD under Grants No. HM1582-08-1-0017 and HM0210-13-1-0005, the OVPR U-Spatial and Minnesota Supercomputing Institute (MSI) at the University of Minnesota.

References

Berman, P.; DasGupta, B.; Muthukrishnan, S.; and Ramaswami, S. 2001. Improved approximation algorithms for rectangle tiling and packing. In *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '01, 427–436.

Eftelioglu, E.; Jiang, Z.; Ali, R.; and Shekhar, S. 2016. Spatial computing perspective on food energy and water nexus. *J. of Environmental Studies and Sciences* 6(1):62–76.

Galzki, J.; Birr, A. S.; and Mulla, D. J. 2011. Identifying critical agricultural areas with 3-meter lidar elevation data for precision conservation. *J. Soil Water Conserve* 66(6):423–430.

Gassman, P. W.; Reyes, M. R.; and et al. 2007. The soil and water assessment tool - historical development applications, and future research directions. *Tran. of the American Soc. of Ag. and Bio. Engineers* 50:1211–1240.

Karmarkar, N. 1984. A new polynomial-time algorithm for linear programming. *Combinatorica* 4(4):373–395.

Kellerer, H.; Pferschy, U.; and Pisinger, D. 2004. *The Multiple-Choice Knapsack Problem*. 317–347.

Laws-Of-Minnesota. 2016. Buffer and soil loss statutes. http://www.bwsr.state.mn.us/buffers.

Lodi, A. 2010. *50 Years of Integer Programming 1958-2008: From the Early Years to the State-of-the-Art*. Springer Berlin Heidelberg.

McDill, M.; Rebain, S.; and Braze, J. 2002. Harvest scheduling with area-based adjacency constraints. *Forest Science* 48(4):631–642.

Murray, A. T., and Weintraub, A. 2002. Scale and unit specification influences in harvest scheduling with maximum area restrictions. *Forest Science* 48(4):779–789.

Muthukrishnan, S.; Poosala, V.; and Suel, T. 1999. *On Rectangular Partitionings in Two Dimensions: Algorithms, Complexity and Applications*. 236–256.

Nelson, E.; Mendoza, G.; Regetz; J., P.; and et. al. 2009. Modeling multiple ecosystem services, biodiversity conservation, commodity production, and tradeoffs at landscape scales. *Frontier in Eco. and the Envir.* 7(1):4–11.

Pimentel, D.; Berger, B.; Filiberto, D.; Newton, M.; Wolfe, B.; Karabinakis, E.; Clark, S.; Poon, E.; and Abbett, E. 2004. Water resources: Agricultural and environmental issues. *BioScience* 54(10):909–918.

Schively, C.; Runck, B.; Pitt, D.; and et al. 2016. Collaborative geodesign to advance multifunctional landscapes. *J. of Landscape and Urban Planning* Accepted.

The Guardian. 2014. Corn farming in the midwest heavily taxes water resources and supply. `www.theguardian.com/sustainable-business/2014/jun/23/corn-farming-midwest-water-pollution-nitrates-scarcity`.

United Nations. 2015. World population prospects. The 2015 revision. `https://esa.un.org/unpd/wpp/publications/files/key_findings_wpp_2015.pdf`.

Viola, P., and Jones, M. J. 2004. Robust real-time face detection. *International J. of Computer Vision* 57(2):137–154.