Extension of a-CAMF to Select Groups of Contiguous Cells for Intervention: Computational Cost Vs. Solution Quality

Grethell Castillo Reyes, Dirk Roose, Gerdys Ernesto Jiménez Moya, René Estrella, Floris Abrams and Jos Van Orshoven

October 5, 2023

//UCIENCIA/23
CONVENCIÓN CIENTÍFICA INTERNACIONAL

Extension of A-CAMF to select groups of contiguous cells for intervention: Computational cost vs. solution quality

Grethell Castillo-Reyes[1,2*], Dirk Roose[2], Gerdys Jiménez-Moya[1], René Estrella[3], Floris Abrams[4], Jos Van Orshoven[4]

[1]Data Representation and Analysis Center, University of Informatics Sciences, Havana, Cuba
[2]Department of Computer Science, KU Leuven, Belgium
[3]Research Group Models, Analysis and Simulations, Department of Computer Science, University of Cuenca, Ecuador
[4]Division of Forest Nature and Landscape, Department of Earth and Environmental Sciences, KU Leuven, Belgium
[*]Corresponding author: gcreyes@uci.cu

Abstract
The CAMF method selects in a raster geo-database of a hydrological catchment those cells for which a certain intervention, e.g. afforestation, leads to maximal reduction of sediment export through the outlet of the catchment. This optimization problem is solved by an iterative process in which independent cells with maximal capacity for sediment yield reduction in the catchment are selected. An accelerated version of CAMF (A-CAMF) was implemented with the aim to reduce the computational cost of the iterative process. The cells selected by A-CAMF are not always in the same region and may not be contiguous, which is a disadvantage for some applications, where performing the intervention at several distributed sites requires extra effort and can causes extra costs. In order to provide alternative solutions for these scenarios using A-CAMF, we extend the method to evaluate the selection of sets of contiguous cells in the minimization of sediment yield in a river catchment. We compare the performance of A-CAMF with and without region-growing (in solution quality and execution time) for minimizing sediment outflow using a raster geo-database of the Tabacay catchment in Ecuador.

Keywords: Afforestation, Contiguity, Minimizing flow, Region-growing, Sediment yield reduction

Introduction

Afforestation of certain areas of a river catchment can reduce the outflow of sediment from that catchment (Costin, 1980; Nearing et al., 2005; Heil et al., 2007; Vanwalleghem, 2017). To locate the optimal sites to minimize the sediment outflow, an heuristic method called CAMF was developed. It was first introduced by Vanegas (2010), and extended in several publications, including Vanegas et al. (2012, 2014); Estrella et al. (2014b); Estrella (2015); Castillo-Reyes et al. (2023). The flow minimization with CAMF is based on a) a model for local sediment production, b) a model for sediment transport to neighboring cells, c) a raster geo-database containing elevation data and land use. The method can also be adapted to model the effect of other interventions than afforestation.

A key issue in sediment flow simulation with CAMF is the spatial interaction among cells in the raster data-set. Spatial interaction refers to the fact that changes in the state of a location can have an impact on the state of neighboring or even distant locations (Gersmehl, 1970; Wang, 2017). In the case of CAMF, spatial interaction refers to the phenomenon that afforestation of a cell leads to changes of its characteristics, which in turn affects the amount of sediment flowing from that cell into its downstream neighboring cells and, eventually, into the outlet of the catchment. The total amount of sediment reaching the outlet is called the sediment yield of the catchment.

The simulation of sediment transport in CAMF can be carried out using Single Flow Direction (SFD) or Multiple Flow Direction (MFD) methods (Castillo-Reyes et al., 2023). We analyzed the differences between the two approaches. With the use of MFD methods the spatial interaction increases and the flow simulation with CAMF-MFD, executed in each iteration of the minimization procedure, has a substantially higher computational cost. The total execution time of CAMF can be prohibitively expensive for large geo-databases, since in each iteration only the cell(s) with the maximum sediment outflow reduction are selected.

In order to improve the performance of CAMF, it was parallelized for multi-core processors and also two algorithmic accelerations were introduced. These accelerations substantially reduce the computational cost, with minimal loss of solution quality. The parallelized and accelerated version of the CAMF method is called A-CAMF.

CAMF and A-CAMF iteratively select independent cells for which the effect of a certain intervention on the sediment yield reduction is maximal. However, the highest ranked cells are not necessarily in the same region and may not be contiguous. For some applications, performing the intervention at several distributed sites, e.g. afforestation in mountain regions, requires extra effort and/or causes extra costs.

Several procedures are used in Spatial Decision Support Systems and Geographic Information Systems to solve the contiguous site location problem (Baerwald, 1981; Eastman et al., 1995; Arentze et al., 1996; Cova and Church, 2000a,b; Vanegas et al., 2008, 2009, 2012; Murray et al., 2022; Murray and Church, 2023). Some of them are based on the use of integer programming, requiring many constraints. Therefore, they are computationally intensive for large problems (Xiao, 2006).

Currently, several methods to identify spatially contiguous areas in a large geographical region using a reasonable time frame are available. Several heuristic approaches based on genetic algorithms have been introduced to solve large size problems (Brookes, 1997b,a, 2001; Xiao et al., 2002; Xiao, 2006; Zhang and Armstrong, 2008; Liu et al., 2012; Demetriou et al., 2013; Dai and Ratick, 2014; Li and Parrott, 2016; Liu et al., 2022).

Vanegas et al. (2011, 2014) propose the combination of the region-growing method developed by Church et al. (2003) with CAMF, to locate near-to-optimal contiguous and compact regions in raster maps with on-site suitability, considering weighted multiple criteria. In this case, CAMF is simplified in order to make the method applicable to large sized problems. To achieve this, the sediment flow is computed in a single step procedure instead of the original iterative process in CAMF, neglecting spatial interaction.

With the aim of extending A-CAMF to evaluate different solution alternatives, in this paper we incorporate the analysis of spatial aspects as contiguity by using topological spatial relations. In order to preserve the effect of spatial interaction through the evaluation of off-site characteristics and to be able to select several regions, we follow a similar ranking procedure as in the original CAMF and A-CAMF.

This paper is organized as follows. In a first section we briefly describe the CAMF and A-CAMF methods for selecting optimal sites for an intervention in the presence of spatial interaction. We summarize the main contributions in the literature related to the selection of contiguous regions. Based on the literature review we describe our solution. Using a case study we evaluate our proposal and finally some conclusions are presented.

Selecting sites for intervention with CAMF and A-CAMF

The CAMF method, introduced in Vanegas (2010) and subsequent papers, and the accelerated version A-CAMF use a raster geo-database containing several properties of the catchment:

- a Digital Elevation Model (DEM);
- a land use types map or land cover map;
- the mean annual amount of sediment produced in each cell ($ton\, ha^{-1}\, yr^{-1}$).

Although the methods can be used to optimize various types of interventions, we describe them in the context of afforestation. We first briefly review the computation of the sediment flow and sediment accumulation. Afterwards we explain the iterative process to select the cells where afforestation should take place to minimize the sediment exported through the outlet of the catchment.

The sediment accumulated in cell i, denoted by SA_i, is the sum of the sediment produced locally and the sediment flowing into that cell from up-slope cells. The latter depends on the flow direction model and the sediment transport model. As already mentioned, different flow direction models (SFD and MFD) are implemented in CAMF and A-CAMF as well (Castillo-Reyes et al., 2023). The transport model implemented in the current version, determines the amount of sediment leaving cell i, as a convex piece-wise linear function of SA_i, depending on three parameters: retention capacity, flow factor and saturation threshold (Vanegas, 2010). Each of these parameters can have different values for each cell. In our simulations these parameters depend on the initial sediment produced locally, i.e., before afforestation.

If SA_i is below the retention capacity, no sediment flows into neighboring cells. If SA_i is between the retention capacity and the saturation threshold, a fraction of the sediment in the cell, denoted as the flow factor, flows into down-slope cells. If SA_i is above the saturation threshold, the amount of sediment above the threshold is fully delivered to down-slope cells.

When a cell is afforested, the retention capacity and saturation threshold increase, and the local sediment production and flow factor decrease. As a result, afforesting a cell reduces the sediment in that cell and also the sediment that flows into the down-slope cells and, eventually, into the outlet cell(s).

Therefore the methods use the following raster data for the sediment flow simulation in each cell i: sediment produced locally α_i^k, flow factor γ_i^k, retention capacity ρ_i^k and saturation threshold σ_i^k, where k indicates whether the cell is not afforested ($k = 1$) or afforested ($k = 2$). Details of the sediment accumulation computation can be found in Vanegas (2010); Castillo-Reyes et al. (2023).

For a given number of cells to be afforested, CAMF and A-CAMF use an iterative process to select the optimal cells to minimize the sediment reaching the outlet cell(s). Only cells that are under specific land use types, called 'candidate cells', can be selected. In each iteration, each candidate cell that is not yet afforested is *tentatively* afforested, one by one, and the resulting sediment accumulation raster is computed. Hence the sediment accumulated in the outlet cell(s), called the sediment yield SY, and also the sediment yield reduction SYR, due to the afforestation of each individual candidate cell, are known. At the end of the iteration, the cells are ranked in descending order based on their SYR value. CAMF selects in each iteration only the cell(s) with the maximal SYR. However, in order to reduce the number of iterations and the cost of the iterations, A-CAMF computes a complete ranking of all cells only every K iterations. In the intermediate iterations only N cells, that were at the top in the previous complete ranking, are ranked. Further, A-CAMF selects in each iteration several cells for which the intervention reduces the sediment yield at the outlet with nearly the same amount.

The iterative procedure ends when the number of selected cells reaches a user-specified value. Note that this stop criterion can easily be replaced by a test on achieving a user-specified SYR. To check the main steps of both methods, see Algorithm 1.

Algorithm 1 Determine the cells to be selected

Input: Number of cells to be selected n

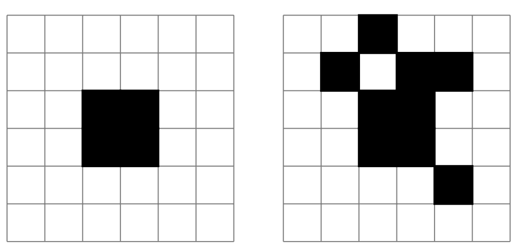
1. $S \leftarrow \emptyset$ ▷ S stores the cells selected
2. $k \leftarrow 0$

while size of $S < n$ **do**
  3. $k \leftarrow k + 1$
  for each candidate cell i that has not been selected **do**
    4. Compute SA matrix and SYR_i^k by tentatively afforesting cell i
  end for
  5. Rank cells according to SYR ▷ With A-CAMF the complete ranking is only computed every K iterations
  6. Put cell(s) with highest SYR in solution set S ▷ With A-CAMF the cells with nearly the same capacity for sediment yield reduction are selected in one iteration
end while

Output: Set of selected cells S

Literature review

In the analysis of spatial relations, a region is considered contiguous (adjacent) if one can move from an identified spatial cell to another cell in the region without leaving the region (Xiao, 2006; Vanegas, 2010). Contiguity does not guarantee compactness (Vanegas, 2010) (see Fig. 1), since a set of contiguous cells can be narrow and elongated or can contain holes that are considered as fragmentation or perforation of the area. In our proposal we only consider the concept of contiguity.



(a) (b)

Figure 1: Representation of spatial relations: (a) contiguous and compact area; (b) contiguous area.

Heuristic methods for implementing contiguity in multiple cell location

As in A-CAMF, a variety of the proposed methods use a graph representation (Xiao et al., 2002; Xiao, 2006) to define the spatial structure in the formulation of the contiguity site search problem. A cell of the raster is equivalent to a vertex in a graph and two adjacent cells are considered to have a direct connection, therefore they are contiguous, Fig. 1b.

As a continuation of the proposal of Eastman et al. (1995) and to manage the shape of a contiguous region, Brookes (1997b) proposes a Parameterized Region-Growing (PRG) method for site location, taking into account particular characteristics on raster suitability maps. To control the shape of the growing region, parameters such as the size, boundary and orientation of an ideal shape are specified. This solution is based on the selection of seeds (initial areas), which is equivalent to identify spatial units for local minima solutions (Vanegas et al., 2010). In a first phase, the cells with suitable characteristics and closest to the seeds are chosen. In a second phase a score is used to indicate the suitability of the cell to be selected as part of a region, evaluating the distance to a seed.

Another heuristic solution, based on the use of seeds, is the Patch Growing Process (PGP) (Church et al., 2003), with application in biological conservation. This method is based on Brookes (1997b,a), but without specific desired shape as a target. A patch begins from an initial seed cell and it is expanded by adding neighbor cells until it reaches a desired total habitat value. The number of edges of the cell shared with a seed or current patch is evaluated in a multi-criteria formulation and a ranking is built according to the suitability of the cells.

The process proposed by Vanegas et al. (2011, 2014) to locate near-to-optimal contiguous and compact regions in raster maps is based on Church et al. (2003) and it consists of three phases: 1) seeds generation 2) region growing and 3) region ranking. An initial map with normalized differences of sediment yield at the outlet(s) is computed as one of the criteria in the seed generation stage. The differences in sediment accumulation are computed in a single step instead of the original iterative process in CAMF, neglecting spatial interaction. The map with normalized sediment flow differences is updated every time with the neighboring cells that are added to the patch. The patches are ranked and the one with the highest score is selected as the best.

Materials and methods

In order to expand particular regions, Brookes (1997b); Church et al. (2003); Vanegas (2010) propose to use the optimal suitable cells as seeds. In our research, we opted to follow this approach, considering the first s cells selected by A-CAMF as seeds. These cells have the maximal potential for sediment yield reduction in the catchment outlet cells.

In order to preserve the effect of spatial interaction and to select several clusters (also patches in some literature) we propose a two-stage procedure to select groups of contiguous cells for an intervention: 1) seed generation and 2) region-growing and selection. The ranking procedure is similar to the procedure in CAMF and A-CAMF: the cell(s) with maximal or nearly maximal capacity for sediment yield reduction are selected in each iteration and added to the clusters.

A user-predefined parameter determines the number of seeds (or the number of clusters formed by the seeds, since often neighboring cells are selected in the same or consecutive iterations) that should be selected to apply region-growing.

The two stages are described in detail below, see also Algorithm 2:

1. Seeds generation. Run A-CAMF to select the number of seeds set by the user-defined parameter.
2. Region-growing and selection by an iterative process. In each iteration
   - 2.1 Compute a new list of candidate cells from the original set of candidate cells, but using the constraint that only neighboring cells of already formed clusters are candidate cells and are evaluated to select the best ones, leading to growing clusters of cells, see Fig. 2.
   - 2.2 Execute one iteration of A-CAMF and select from the new list of candidate cells those cell(s) with the highest potential for sediment yield reduction.
   - 2.3 Add the cell(s) selected to the formed clusters in solution set.

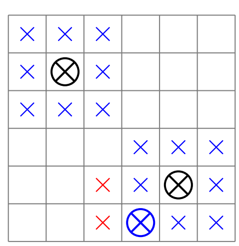


Figure 2: Representation of the selection of sets of contiguous cells in A-CAMF. Assume that all white cells are originally candidate cells. $\otimes$: seeds selected in the first stage; $\times$, $\otimes$: candidate cells computed in the first iteration of the region-growing stage with $\otimes$ the cell selected in the iteration; $\times$: neighbor cells added to the set of candidate cells in the second iteration of the region-growing stage.

The iterative procedure, steps 2.1 to 2.3, ends when the (user-specified) number of selected cells is fulfilled or a relative sediment yield reduction is achieved.

Algorithm 2 Selection of clusters of cells

Input: Number of cells to be selected n; number of seeds s

1. $S \leftarrow \emptyset$ ▷ S stores the selected cells

while size of $S < s$ **do** ▷ Computation of A-CAMF to select first s cells as seeds
  for each candidate cell i that has not been selected **do**
    2. Compute SA matrix and SYR_i^k by tentatively afforesting cell i
  end for
  3. Rank cells according to SYR
  4. Put cell(s) with maximal or nearly maximal SYR in solution set S
end while

while size of $S < n$ **do**
  5. $C \leftarrow \emptyset$ ▷ C stores the new candidate cells computed from the cells stored in S

  for each cell j stored in S **do** ▷ Grow the regions with the candidate neighbors of cells stored in S
    6. Compute candidate neighbor cells of j and store them in C; if not yet in C and not yet in S.
  end for

  for each candidate cell i in C **do**
    7. Compute SA matrix and SYR_i^k by tentatively afforesting cell i
  end for

  8. Rank cells according to SYR
  9. Put cell(s) with maximal SYR in solution set S
end while

Output: Set of selected cells S

Note that this approach does not guarantee the selection of the best possible clusters, but reduces the extra effort and cost of performing the intervention in distant locations. Additionally, it can be considered as another acceleration. Since the set of candidate cells used in each iteration is only a fragment of the original set of candidate cells, the number of simulations (i.e., tentatively afforesting each candidate cell) per iteration is reduced, therefore the computational cost decreases.

Since the list of candidate cells is computed dynamically in each iteration (Fig. 2), it is possible that sub-optimal cells are processed before the optimal ones. Therefore it is not recommended to combine the iterative process in stage 2 (region-growing and selection) with the selection of several cells per iteration and the re-use of a partial ranking in A-CAMF.

Case study

The case study deals with the Tabacay river catchment located in the Republic of Ecuador in South America, with an area of $\approx 6\ 639\ ha$ and altitudes ranging from $2\ 482\ m$ to $3\ 731\ m$ above sea level, see Fig. 3. Agriculture and pasture cover 39% of the area. The widespread agricultural land use (Fig. 4a), even on steep slopes, produces a large amount of sediment, which flows towards the river, resulting in substantial land, river and reservoir degradation.
A raster geo-database of the Tabacay catchment with a resolution of $30m \times 30m$ is used, containing 122 830 cells, of which 73 471 are cells with actual values. Only a part of the cells under agriculture and pasture are considered as candidate cells (27 246 cells).

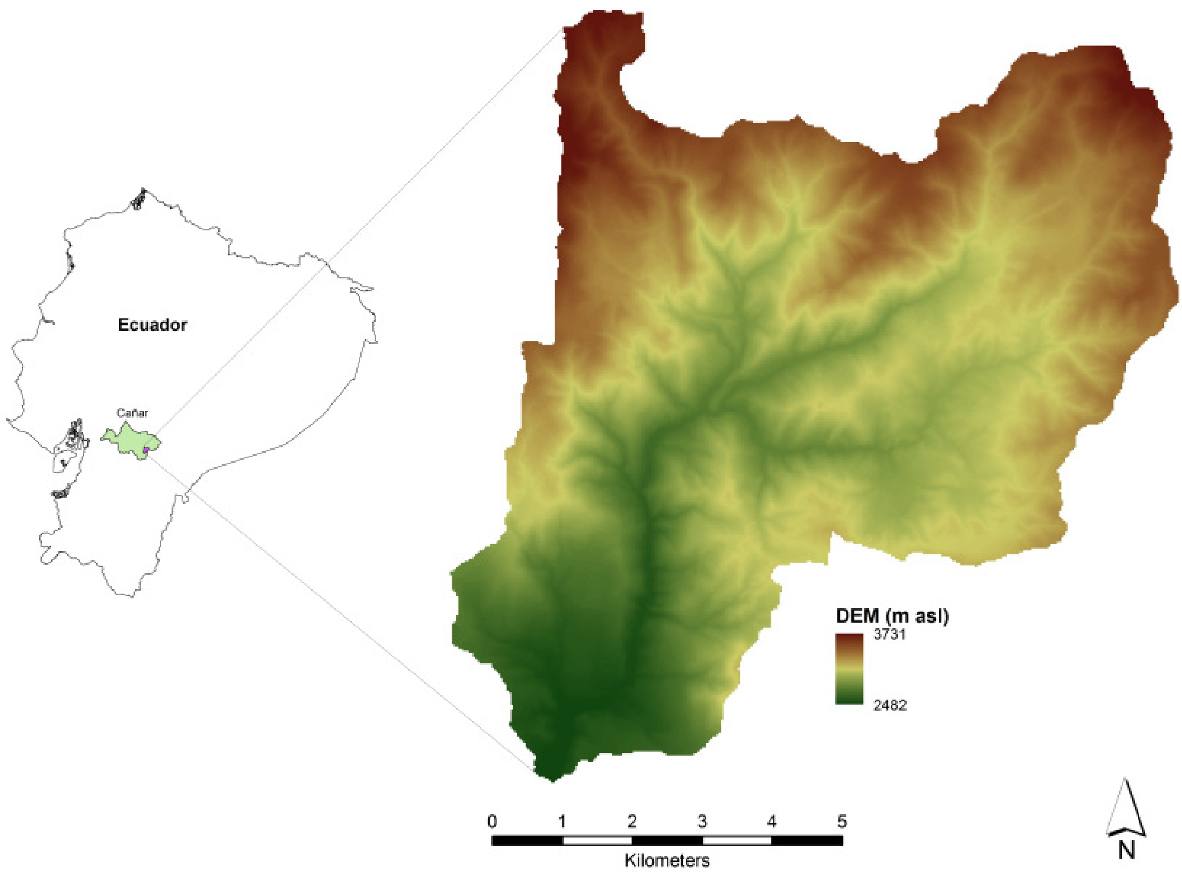
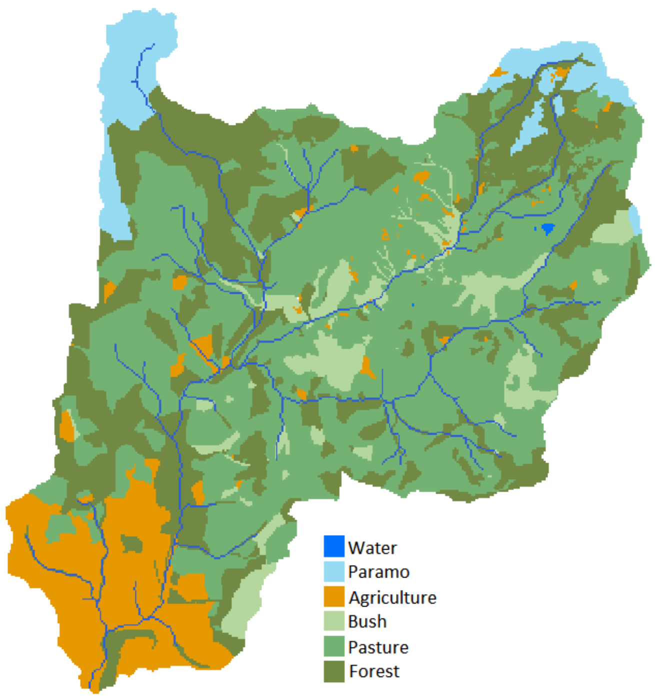
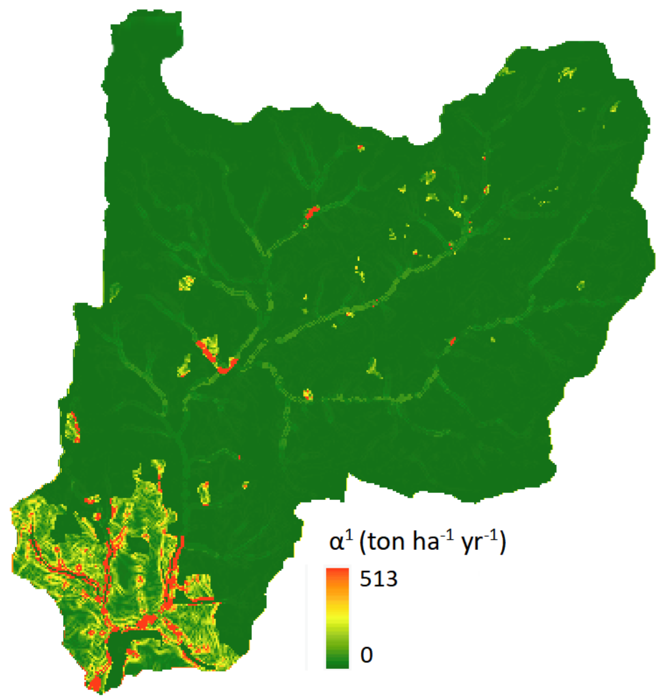


Figure 3: Digital Elevation Model of the Tabacay catchment and its location in Ecuador (Estrella et al. (2014a)).

The initial sediment production map α^1 was computed using RUSLE (Renard et al., 1991). We refer to Castillo-Reyes et al. (2023) for details about the computation of α^1 and the parameters values used in RUSLE. The land use map and the initial sediment production α^1 are shown in Figs. 4a, 4b. The initial flow factor map γ^1 was computed by a linear transformation of the original slope map using min-max normalization (Han et al., 2012). The values of the other parameters to calculate SA (Castillo-Reyes et al., 2023), i.e., ρ^1, ρ^2, σ^1, σ^2, α^2 and γ^2, listed in Table 1, are taken from Estrella (2015). Note that Estrella (2015) used a slightly different initial local sediment production map α^1, since the LS-factor was computed differently.



(a)



(b)

Figure 4: (a) Land cover map of the Tabacay river catchment (Estrella, 2015). (b) Initial local sediment production map α^1 ($ton\, ha^{-1}\, yr^{-1}$) of the Tabacay river catchment, calculated by means of RUSLE. $\min_{\alpha^1} = 0\, ton\, ha^{-1}\, yr^{-1}$, $\max_{\alpha^1} = 513\, ton\, ha^{-1}\, yr^{-1}$, $\overline{\alpha^1} = 2.54\, ton\, ha^{-1}\, yr^{-1}$.

Table 1: Parameter values used in the experiments with A-CAMF for the Tabacay case study.

Parameter	**Before afforestation**	**After afforestation**
Sediment production	α^1, calculated by RUSLE, Fig. 4b	$\alpha^2 = 0.83 \times \alpha^1$
Retention capacity	$\rho^1 = 0.37 \times \alpha^1$	$\rho^2 = 0.61 \times \alpha^1$
Saturation threshold	$\sigma^1 = 0.96 \times \alpha^1$	$\sigma^2 = 0.98 \times \alpha^1$
Flow factor	γ^1, normalized slope from DEM	$\gamma^2 = 0.75 \times \gamma^1$

Results and Discussion

All experiments have been performed on a Xeon E5-2697 v3 CPU (2.6 GHz) with 28 cores and 128 GB of RAM, with Ubuntu Bionic Linux as Operating System. A-CAMF is implemented in C++ and it can also be executed under Windows. We use 28 threads to take advantage of the available 28 cores.

We evaluate the behaviour of the region-growing approach for the Tabacay river catchment and we compare the results with our previous results for this catchment, using A-CAMF for selecting independent cells (see Tables 2 - 3). For the simulation of flow paths we configured A-CAMF to use a MFD method, specifically the Fractional Deterministic Eight-Neighbor (FD8) variant.

We first use A-CAMF to afforest a given number of cells (5% to 30% of the number of candidate cells). The results show that indeed the region-growing approach leads to a sub-optimal solution with respect to the sediment yield reduction. Indeed, RD, the relative difference in sediment yield reduction between A-CAMF without region-growing and A-CAMF with region-growing, is between 11.6% and 14.2 % when 100 seeds are selected and between 8.5% and 11.3% for 200 seeds selected. However, since only the neighbors of the seeds or already formed clusters are allowed as candidate cells, the number of cells to be evaluated in the simulations is smaller and the execution time is reduced by a factor of 4 - 5, i.e., using region-growing leads to a speedup of 4 - 5. The locations of the cells selected by both approaches are shown in Figs. 5 - 8.

Table 2: Performance of A-CAMF for selecting independent cells (IC) and clusters of cells (RG) with 100 seeds. RD: relative difference between SYR when the original selection (independent cells) is used and SYR after selecting clusters of cells; T_{28}: CPU time (s) on 28 cores; AS_{28}: Algorithmic Speedup on 28 cores; # afforested cells: 5%, 10%, 20% and 30% of the candidate cells.

# Afforested cells	SYR ($ton\,yr^{-1}$)		RD (%)	T_{28}		AS_{28}
	IC	RG		IC	RG	
5% - 1 362	11 243.77	9 939.35	11.60	40 996	10 753	3.81
10% - 2 724	12 944.78	11 109.88	14.17	82 319	17 006	4.84
20% - 5 448	14 034.18	12 192.21	13.12	171 612	34 788	4.93
30% - 8 172	14 310.08	12 625.25	11.77	245 954	61 720	3.98

Table 3: Performance of A-CAMF for selecting independent cells (IC) and clusters of cells (RG) with 200 seeds. RD: relative difference between SYR when the original selection (independent cells) is used and SYR after selecting clusters of cells; T_{28}: CPU time (s) on 28 cores; AS_{28}: Algorithmic Speedup on 28 cores; # afforested cells: 5%, 10%, 20% and 30% of the candidate cells.

# Afforested cells	SYR ($ton\,yr^{-1}$)		RD (%)	T_{28}		AS_{28}
	IC	RG		IC	RG	
5% - 1 362	11 243.77	10 289.15	8.49	40 996	14 081	2.91
10% - 2 724	12 944.78	11 482.97	11.30	82 319	20 397	4.03
20% - 5 448	14 034.18	12 574.25	10.40	171 612	36 115	4.75
30% - 8 172	14 310.08	13 008.26	9.09	245 954	59 612	4.13

Note that when 200 seeds are used, in some cases the speedup decreases in comparison with the speedup when 100 seeds are selected. By selecting more seeds, the number of candidate cells, evaluated in each iteration, increases which affects the total execution time. However, since the seeds are the cells with the maximal capacity for sediment yield reduction, RD decreases when the number of seeds increases.

We also use A-CAMF to achieve a user-specified value for the relative amount of sediment reduction in the outlet cells. In this case, using region-growing affects the number of cells that must be afforested. Table 4 shows the results after running A-CAMF to reduce 35% of the total initial SY, denoted by SY^0. Selecting independent cells only 9.66% of the candidate cells must be selected for afforestation, while using the region-growing approach 38% of the cells must be afforested to reach the required reduction (see also maps in Fig. 9).

Table 4: Performance of A-CAMF for selecting independent cells (IC) and clusters of cells (RG) with 100 seeds. Stopping criteria: reduction of 20% and 35% of $SY^0 = 36\,774\,ton\,yr^{-1}$. T_{28}: CPU time (s) on 28 cores; % of cells selected: % of cells selected from the total set of candidate cells (27 246).

Seeds	% SYR from SY^0	Selection method	# cells selected	% of cells selected	T_{28}
100	20	IC	319	1.17	14 870
		RG	437	1.60	5 871
	35	IC	2 633	9.66	162 987
		RG	10 414	38.22	115 333

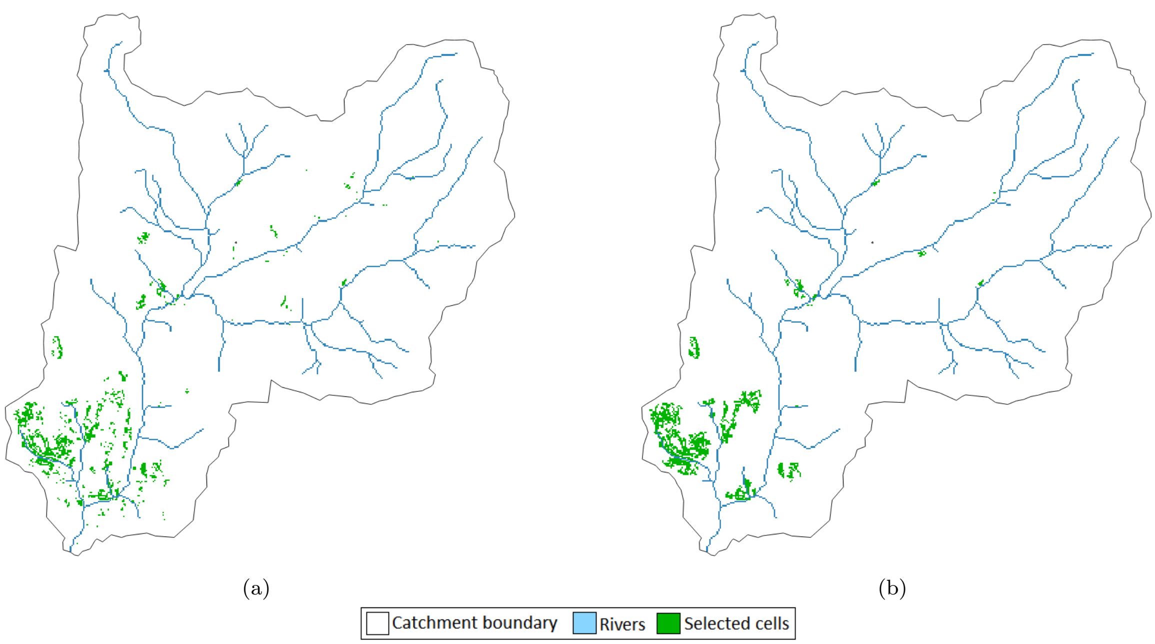


Figure 5: Areas selected by A-CAMF for afforestation in the Tabacay river catchment for (a): independent cell selection; (b): region-growing with 100 seeds. Selection of 5% of the candidate cells.

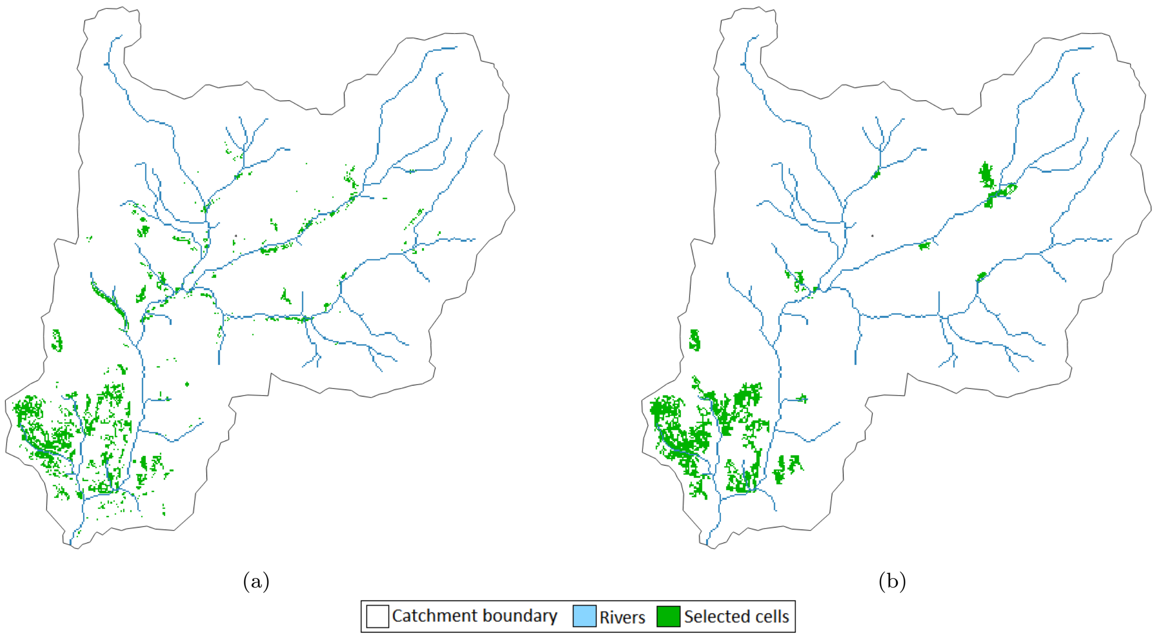


Figure 6: Areas selected by A-CAMF for afforestation in the Tabacay river catchment for (a): independent cell selection; (b): region-growing with 100 seeds. Selection of 10% of the candidate cells.

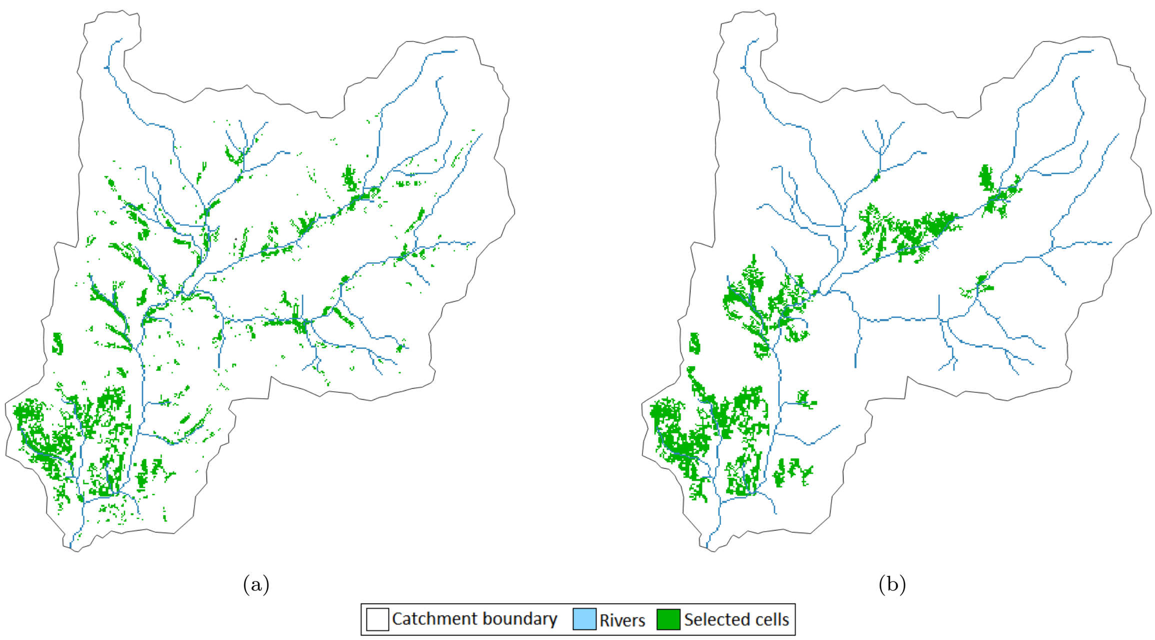


Figure 7: Areas selected by A-CAMF for afforestation in the Tabacay river catchment for (a): independent cell selection; (b): region-growing with 100 seeds. Selection of 20% of the candidate cells.

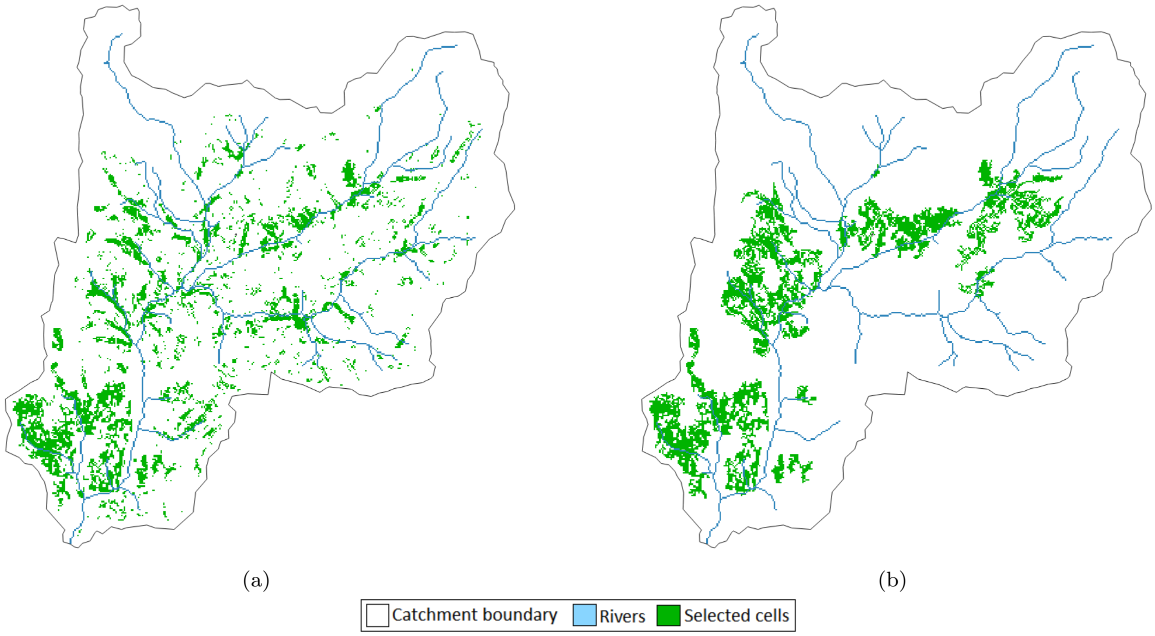


Figure 8: Areas selected by A-CAMF for afforestation in the Tabacay river catchment for (a): independent cell selection; (b): region-growing with 100 seeds. Selection of 30% of the candidate cells.

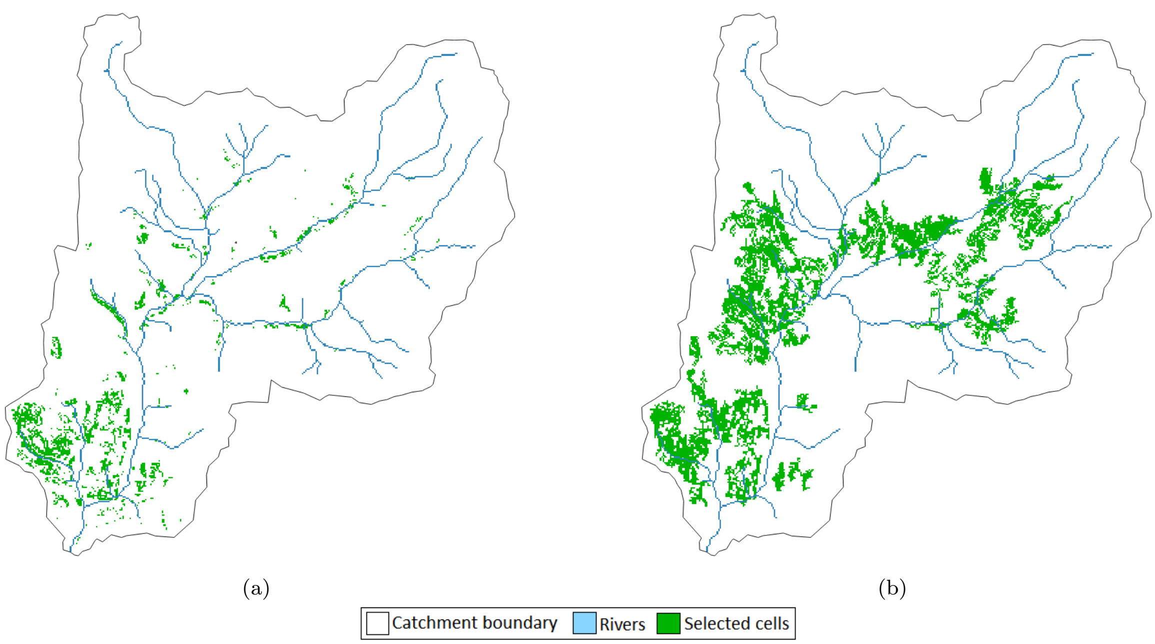


Figure 9: Areas selected by A-CAMF for afforestation in the Tabacay river catchment to reduce 35% of the sediment yield in the outlet, when (a): independent cell selection is used; (b): region-growing is used with 100 seeds.

Conclusions

The contiguity concept in spatial relationships is an important constraint in several land use applications, e.g. the afforestation of mountain regions. Several techniques, comprising integer programming and heuristic approaches, have been developed to consider contiguity in the selection of optimal sites for an intervention. The generation of seeds as starting points for the contiguous regions is a technique proposed in a variety of the heuristic solutions.

In this paper, we presented an extension of A-CAMF to select several contiguity regions for afforestation while still preserving the effect of spatial interaction. As seeds we select the firs s cells with the maximal capacity for sediment yield reduction selected by A-CAMF. The number of seeds is a user-defined parameter.

We compared the performance of A-CAMF using region-growing and A-CAMF without region-growing, in solution quality and computational cost, using a river catchment located in Ecuador. Using region-growing reduces the computational cost, since the set of candidate cells used in each iteration is smaller than the original set of candidate cells. However, since the candidate cells are limited to the neighbors of the seeds or already formed clusters, the optimal cells are often not evaluated and selected.

For the current test case, to reduce the 35% of sediment load in the outlet of the river catchment, $\approx 38\%$ of the candidate cells must be afforested if the contiguity constraint is evaluated, while in the other case only a small part of the candidate cells are selected. This clearly reflects that not only the sediment reduction should be evaluated, but also the cost of the intervention, since often the budget is important in real applications. The cost of afforestation depends, for example, on the price of the land, on the actions needed to prepare the land, to plant the trees and to maintain the plantation, which is a function of accessibility in which the region-growing plays an important role. Therefore, we consider the solution as a supporting point for a decision maker, who will focus in the advantages and disadvantages depending on the field of application, the intervention to be applied and the budget required.

The literature reflects that also other constraints, as compactness, are significant in this field. In order to also consider compactness besides contiguity in A-CAMF, other constrains should be added to select the new set of candidate cells per iteration, limiting the cells to the neighbor cells that meet certain characteristics. Additionally, the inclusion of several constraints in the selection of the optimal sites for sediment yield reduction in a hydrological catchment, shows a clear need to consider a multi-criteria evaluation in A-CAMF, which is an interesting topic for forthcoming research.

Acknowledgments

The authors acknowledge the support of the VLIR-UOS projects "Afforestation and avoidance of deforestation to protect mountain cities from floods and mud flows" and "Networks 2019 Phase 2 Cuba ICT" (PhD research grant of G. Castillo-Reyes).

References

Arentze, T. A., Borgers, A. W. J., and Timmermans, H. J. P. (1996). An efficient search strategy for site-selection decisions in an expert system. *Geographical Analysis*, 28(2):126–146.

Baerwald, T. J. (1981). The site selection process of suburban residential builders. *Urban geography.*, 22(4).

Brookes, C. J. (1997a). A genetic algorithm for locating optimal sites on raster suitability maps. *Transactions in GIS*, 2(3):201–212.

Brookes, C. J. (1997b). A parameterized region-growing programme for site allocation on raster suitability maps. *International Journal of Geographical Information Science*, 11(4):375–396.

Brookes, C. J. (2001). A genetic algorithm for designing optimal patch configurations in GIS. *International Journal of Geographical Information Science*, 15(6):539–559.

Castillo-Reyes, G., Estrella, R., Gabriels, K., Van Orshoven, J., Abrams, F., and Roose, D. (2023). Selecting sites for afforestation to minimize sediment loss from a river basin: Computational complexity of Single and Multiple Flow Direction Methods in raster databases. *Computers & Geosciences*, 171:105269.

Church, R. L., Gerrard, R. A., Gilpin, M., and Stine, P. (2003). Constructing cell-based habitat patches useful in conservation planning. *Annals of the Association of American Geographers*, 93(4):814–827.

Costin, A. (1980). Runoff and soil and nutrient losses from an improved pasture at ginninderra, southern tablelands, new south wales. *Crop & Pasture Science*, 31:533–546.

Cova, T. J. and Church, R. L. (2000a). Contiguity constraints for single-region site search problems. *Geographical Analysis*, 32(4):306–329.

Cova, T. J. and Church, R. L. (2000b). Exploratory spatial optimization in site search: a neighborhood operator approach. *Computers, Environment and Urban Systems*, 24(5):401–419.

Dai, W. and Ratick, S. J. (2014). Integrating a raster geographical information system with multi-objective land allocation optimization for conservation reserve design. *Transactions in GIS*, 18(6):936–949.

Demetriou, D., See, L., and Stillwell, J. (2013). A spatial genetic algorithm for automating land partitioning. *International Journal of Geographical Information Science*, 27(12):2391–2409.

Eastman, J., Jin, W., Kyem, P., and Toledano, J. (1995). Raste procedure for multi-criteria/multi-objective decisions. *Photogrammetric Engineering & Remote Sensing*, 61:539–547.

Estrella, R. (2015). *Where to afforest? Single and multiple criteria evaluation methods for spatio-temporal decision support, with application to afforestation*. PhD thesis, KU Leuven.

Estrella, R., Cattrysse, D., and Van Orshoven, J. (2014a). Comparison of three ideal point-based multi-criteria decision methods for afforestation planning. *Forests*, 5(12):3222–3240.

Estrella, R., Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2014b). Trading off accuracy and computational efficiency of an afforestation site location method for minimizing sediment yield in a river catchment. pages 94–100. Rueckemann, Claus-Peter, International Academy, Research, and Industry Association (IARIA).

Gersmehl, P. (1970). Spatial interaction. *Journal of Geography*, 69(9):522–530.

Han, J., Kamber, M., and Pei, J. (2012). Data Preprocessing. In Han, J., Kamber, M., and Pei, J., editors, *Data Mining*, The Morgan Kaufmann Series in Data Management Systems, pages 83–124. Morgan Kaufmann, Boston, Third edition.

Heil, G., Muys, B., and Hansen, K. (2007). *Environmental Effects of Afforestation in North-Western Europe â€“ From Field Observations to Decision Support*. Springer.

Li, X. and Parrott, L. (2016). An improved genetic algorithm for spatial optimization of multi-objective and multi-site land use allocation. *Computers, Environment and Urban Systems*, 59:184–194.

Liu, H., Yan, F., and Tian, H. (2022). Towards low-carbon cities: Patch-based multi-objective optimization of land use allocation using an improved non-dominated sorting genetic algorithm-II. *Ecological Indicators*, 134:108455.

Liu, X., Li, X., Shi, X., Huang, K., and Liu, Y. (2012). A multi-type ant colony optimization (MACO) method for optimal land use allocation in large areas. *International Journal of Geographical Information Science*, 26(7):1325–1343.

Murray, A. T. and Church, R. L. (2023). Spatial optimization of multiple area land acquisition. *Computers & Operations Research*, 153:106160.

Murray, A. T., Church, R. L., Pludow, B. A., and Stine, P. (2022). Advancing contiguous environmental land allocation analysis, planning and modeling. *Journal of Land Use Science*, 17(1):572–590.

Nearing, M., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., Hernandez, M., Le Bissonnais, Y., Nichols, M., Nunes, J., Renschler, C., Souchére, V., and van Oost, K. (2005). Modeling response of soil erosion and runoff to changes in precipitation and cover. *CATENA*, 61(2):131–154.

Renard, K. G., Foster, G. R., Weesies, G. A., and Porter, J. P. (1991). RUSLE: Revised Universal Soil Loss Equation. *Journal of Soil and Water Conservation*, 46(1):30–33.

Vanegas, P. (2010). *A spatially explicit approach to the site location problem in raster maps with application to afforestation*. PhD thesis, KU Leuven, KU Leuven, Belgium.

Vanegas, P., Cattrysse, D., and Orshoven, J. (2012). Allocating reforestation areas for sediment flow minimization: an integer programming formulation and a heuristic solution method. *Optimization and Engineering*, 13:247–269.

Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2008). Comparing exact and heuristic methods for site location based on multiple attributes: An afforestation application. In Gervasi, O., Murgante, B., Laganà, A., Taniar, D., Mun, Y., and Gavrilova, M. L., editors, *Computational Science and Its Applications – ICCSA 2008*, pages 389–404. Springer Berlin Heidelberg.

Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2009). Compactness and flow minimization requirements in reforestation initiatives: An integer programming (IP) formulation. In Gervasi, O., Taniar, D., Murgante, B., Laganà, A., Mun, Y., and Gavrilova, M. L., editors, *Computational Science and Its Applications – ICCSA 2009*, pages 132–147. Springer Berlin Heidelberg.

Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2010). Compactness in spatial decision support: A literature review. In Taniar, D., Gervasi, O., Murgante, B., Pardede, E., and Apduhan, B. O., editors, *Computational Science and Its Applications – ICCSA 2010*, pages 414–429. Springer Berlin Heidelberg.

Vanegas, P., Cattrysse, D., and Van Orshoven, J. (2011). A multiple criteria heuristic solution method for locating near to optimal contiguous and compact sites in raster maps. In *Geocomputation, sustainability and environmental planning*, volume 348, pages 35–56. Springer, First edition.

Vanegas, P., Cattrysse, D., Wijffels, A., and Van Orshoven, J. (2014). Compactness and flow minimization requirements in reforestation initiatives: A heuristic solution method. *Annals of Operations Research*, 219(1):433–456.

Vanwalleghem, T. (2017). *Soil Erosion and Conservation*, pages 1–10. John Wiley & Sons, Ltd.

Wang, J. (2017). *Economic Geography: Spatial Interaction*, pages 1–4. John Wiley & Sons, Ltd.

Xiao, N. (2006). An evolutionary algorithm for site search problems. *Geographical Analysis*, 38(3):227–247.

Xiao, N., Bennett, D. A., and Armstrong, M. P. (2002). Using evolutionary algorithms to generate alternatives for multiobjective site-search problems. *Environment and Planning A: Economy and Space*, 34(4):639–656.

Zhang, X. and Armstrong, M. P. (2008). Genetic algorithms and the corridor location problem: Multiple objectives and alternative solutions. *Environment and Planning B: Planning and Design*, 35(1):148–168.