Compactness and Flow Minimization Requirements in Reforestation Initiatives: An Integer Programming (IP) Formulation

Pablo Vanegas[1,3], Dirk Cattrysse[1], and Jos Van Orshoven[2]

[1] Centre for Industrial Management, Katholieke Universiteit Leuven,
Celestijnenlaan 300A, 3001 Heverlee - Leuven, Belgium
[2] Department of Earth and Environmental Sciences, Katholieke Universiteit Leuven,
Celestijnenlaan 200E, 3001 Heverlee - Leuven, Belgium
[3] Facultad de Ingeniería, Universidad de Cuenca,
Cdla. Universitaria Av. 12 de Abril s/n, Cuenca, Ecuador
Pablo.Vanegas@cib.kuleuven.be, Pablo.Vanegas@ucuenca.edu.ec,
Dirk.Cattrysse@cib.kuleuven.be, Jos.VanOrshoven@biw.kuleuven.be

Abstract. Identifying optimal sites on raster maps is a complex problem when the sites are larger than the cell size. Optimal sites involves a trade-off between the intrinsic characteristics of individual cells and the spatial configuration of the cells. Although there are a number of techniques to solve the site allocation problem, those solutions do not consider spatial interactions between the cells forming the site. This paper presents an Integer Programming Formulation (IP) for allocating a predefined number of cells satisfying the following criteria: 1) minimize flow (water, sediment) reaching the outlet of a watershed, 2) maximize/minimize intrinsic characteristics of the cells, and 3) form a compact patch. Although the core structure of the IP formulation can be applied for different sorts of flow and intrinsic characteristics, it is targeted to a reforestation application. The proposed approach is applied to perform several experiments in two watersheds in South Dakota in the USA for searching a given number of best cells (1) minimizing sediment reaching the watershed outlet,(2) maximizing the environmental criteria, and (3) forming a compact patch. The results obtained with the IP formulation are in agreement with expert assessments of erosion levels, slopes and distances to the riverbeds.

Keywords: Site Location, Site Allocation, Integer Programming, Exact Methods, Flow Minimization.

1 Introduction

Environmental conservation and land use planning often require identifying optimal sites fulfilling particular criteria. That is the case of e.g. locating optimal sites for afforestation [1], for conservation planning [2], or for another particular land use [3]. These examples are mainly dealing with allocation of compact

patches, and do not consider interactions between cells. According to the present knowledge, no contributions have been developed to combine compactness and spatial interactions.

There exist field studies searching for optimal sites where the involved criteria produce interactions between cells. In contrast to simulations, this sort of problems requires including search algorithms for optimality. E.g. in forestation management where decision criteria can be represented by raster maps, the following question regarding sediment flow can arise: *where are the best cells located to be reforested in order to minimize sediment flow reaching the watershed outlet?*. The present paper locates cells to be reforested where *flow production* and *flow delivery* properties are modified for minimizing sediment flow reaching an outlet. Since flow is nonlinear in nature, a piecewise linear convex function is applied to model flow delivery from a cell to one of its neighbors. The flow delivery function needs two breakpoints in order to define three segments. In turn each segment is associated with a specific factor. Breakpoints, segment factors, and flow production in a cell change when it is selected as optimal to be reforested. These changes affect also neighboring cells (spatial interaction). In addition to minimize sediment flow, the selected cells form a compact patch. Compactness is achieved by means of a matrix data structure that allows selecting cells sharing as many borders as possible for minimizing the total perimeter while the area is constant. Moreover, the compact patch also maximizes the environmental criteria expressed through levels of carbon sequestration and nitrate leaching.

This paper formulates an Integer Programming (IP) model for allocating a predefined number of cells (1) minimizing sediment flow reaching the outlet of a watershed, (2) maximizing their intrinsic environmental criteria (maximizing carbon sequestration, minimizing nitrate leaching), and (3) forming a compact patch. The proposed IP is based on the general Network Flow (NF) formulation [4] and on the IP model applied by [1] for allocating multiple criteria compact patches. The next section reviews definitions of spatial relation and interaction, and concepts of Network Flow (NF) models. Section 3 details the proposed IP approach to combine the formulations in [4] and [1]. Finally, section 4 shows and discusses the results obtained for two small watersheds in South Dakota in the USA, and section 5 draws the conclusions.

2   Literature Review

Topology is a mathematical concept that has its origin in the principles of object adjacency and connectedness. It is the way in which spatial objects are linked together, or the set of collective links or spatial relationships between points, lines, and areas [5]. This definition is in agreement with the definition of *Spatial Relation*: it specifies how objects are located in space in relation to some reference objects. Beyond relations, *Spatial Interactions* designate the existence of causal relations in space, the existence of spatial diffusion processes. Topology has been applied to model the *site location/allocation problem* with Integer Programming (IP), nevertheless according to our present knowledge, problems involving spatial interactions have not been tackled with this approach. In this direction, the

proposed IP formulation, based on the general Network Flow (NF) problem, attempts to model spatial interactions between cells in order to minimize flow reaching a predefined outlet. This NF based formulation is combined with a compactness model in order to identify compact sites minimizing flow.

2.1   Network Flow (NF) Problem

According to [4], the NF problem considers a directed network G=(N,A) defined by a set N of n nodes and a set A of m directed arcs. Each arc $(i, j) \in A$ has associated a cost c_{ij} per unit flow, a capacity u_{ij} denoting the maximum flow, and a lower bound l_{ij} for the minimum flow along the arc. A number $b(i)$ is associated to each node; if $b(i) > 0$, i is a supply node, if $b(i) < 0$, i is a demand node, and if $b(i) = 0$, i is a transshipment node.

Although flow cost between nodes can vary linearly in the amount of flow, non-linear or piecewise linear functions may be more likely to find in real applications. In a piecewise linear convex model [4] each arc cost $C_{ij}(x_{ij})$ has at most p linear segments: $(0 = d_{ij}^0) < (d_{ij}^1) < (d_{ij}^2) < ...$ denote the breakpoints of the function, where cost varies linearly within the interval $[d_{ij}^{k-1}, d_{ij}^k]$. Therefore, to specify a piecewise linear cost function, it is necessary to specify the breakpoints and the slopes of the linear segments between successive breakpoints. In this sense, a flow x_{ij} can be decomposed into different segments, each one representing flow between two breakpoints d_{ij}^{k-1} to d_{ij}^k. Let y_{ij}^k denote the flow along the kth segment, that is, between d_{ij}^{k-1} and d_{ij}^k. Segment flows y_{ij}^k from the total arc flow x_{ij} can be computed as follows:

$$y_{ij}^k = \begin{cases} 0 & if \ x_{ij} \leq d_{ij}^{k-1} \\ x_{ij} - d_{ij}^{k-1} & if \ d_{ij}^{k-1} \leq x_{ij} \leq d_{ij}^k \\ d_{ij}^k - d_{ij}^{k-1} & if \ x_{ij} \geq d_{ij}^k \end{cases}$$

Since C_{ij} is a convex cost function, $C_{ij}(x_{ij}) = \sum_{k=1}^p c_{ij}^k y_{ij}^k$. Consequently, the following formulation can applied to transform a convex cost flow problem to a minimum cost flow problem:

minimize:

$$\sum_{(i,j) \in A} \sum_{k=1}^p c_{ij}^k \cdot y_{ij}^k \tag{1}$$

s.t:

$$\sum_{\{j:(i,j) \in A\}} \sum_{k=1}^p y_{i,j}^k \quad -$$

$$\sum_{\{j:(j,i) \in A\}} \sum_{k=1}^p y_{j,i}^k = b(i) \qquad \forall i \in N \tag{2}$$

$$0 \leq y_{ij}^k \leq d_{ij}^k - d_{ij}^{k-1} \qquad \forall (i,j) \in A, \forall k = 1, ..., p \tag{3}$$

Constraints in equation 2 are referred as *mass balance constraints*, the first term represents total *outflow* of nodes and the second one total *inflow*. Finally, constraint in equation 3 controls the amount to be delivered from node i to node j along segment k.

2.2 Mathematical Programming Models for Site Location/Allocation

Hof and Bevers [6] formulated four linear programming examples including constraints to avoid adjacency in order to account for biological dispersal. The constraints relate population in a habitat area in a time period (t) to the populations in other areas in a previous time period (t-1), while taking into account the population growth and the immigration dispersion. Church and ReVelle [7] introduce the Maximal Covering Location Model (MCLM), which minimizes the number of facilities to cover each and every demand point on a network. MCLM is modified [8] to determine the optimal deployment of available fire-fighting vehicles. In order to reduce the vulnerability of elements like species, communities, or endemic plants, Church et al. [9] develop a mathematical model for selecting sites for diversity conservation (Biodiversity Management Areas - BMAs). Since the solutions are composed of isolated planning units, to avoid fragmentation, Fischer and Church [10] formulate a mathematical model including the objective of minimizing the outside perimeter of the selected areas. Along these linear approaches, Vanegas et al [1] propose an IP formulation for allocating a set of cells forming a compact patch and maximizing environmental performance. This formulation is applied as an optimality reference to measure the performance of an heuristic solution method for allocating compact patches in raster maps. The idea to represent a mosaic of n cells as a planar graph with vertices and edges is presented by Williams [11], [12]. Each cell is equated with a vertex, and each adjacency relation between a pair of cells is equated with an edge. This idea is also applied by Shirabe [13] to formulate the necessary and sufficient conditions for assembling a connected region with a desired degree of perforation, from no hole to a largest possible number of holes. Although important contributions have been found in mathematical programming for the site allocation problem, none of them are dealing with interactions, e.g. to allocate sites for flow minimization.

3 Materials and Methods

The notation and formulation explained in this section are intended to locate a compact given patch of a given number of cells to be reforested. Reforestation implies modification of *flow production*, *flow delivery factors*, and *environmental criteria* in every selected cell. The first two new conditions affect not only a cell itself but also the state of neighboring cells. For modeling this interaction, a *tree data structure* is constructed from a Single Flow Direction (SFD) map. The compactness criterion requires a *matrix data structure* to allocate cells sharing as many borders as possible in order to minimize the total perimeter of the reforested cells forming the target site. Since *nodes of the tree* and *cells of the matrix*

refer to the same spatial location, both data structures need to be combined to minimize sediment flow and maximize compactness simultaneously.

Whereas subsection 3.1 details the data requirements and their representation, subsection 3.2 explains the IP formulation. Subsection 3.2 includes the notation for minimizing flow at a watershed outlet as well as the notation for allocating a set of cells maximizing their intrinsic environmental performance and forming a compact patch.

## 3.1   Data Description and Representation

Data Description. Twelve raster maps (matrices) are required to identify cells suitable to be reforested to minimize sediment load at a given outlet, and maximize the environmental performance: 1) Flow Direction, 2) Flow Production, 3) Flow Factor, 4) Breakpoint #1, 5) Breakpoint #2, 6) New Flow Production, 7) New Flow Factor, 8) New Breakpoint #1, 9) New Breakpoint #2, 10) Carbon Sequestration, 11) Nitrate Leaching and 12) Location of streams. Cell states change when they are reforested (selected as part of the optimal solution) for minimizing sediment load. In this sense, the first 5 criteria reflect an initial state in every cell and the next 4 represent a new state (after reforestation) affecting also the state of neighboring cells (spatial interaction). Figures 2a and 2b make use of cell j to illustrate the meaning of these criteria.

1. *Flow Direction (FD)* is based on elevation maps. FD uses the SFD (Single Flow Direction or D8) model, computing the downslope flow directions by inspecting the 3-by-3 window around the current cell (Moore's Neighborhood). The SFD method assigns to the current cell a unique flow direction, i.e. toward the one with the steepest downslope neighbor.
2. *Flow Production (α)* is a general term associated to levels of any kind of locally produced flow. In the problem at hand, flow production refers to sediment or erosion (Tha_1yr_1).
3. *Flow Factor (γ)* stands for a factor of transporting sediment from a cell to one of its 8 neighbors. In problems where geographic relief plays an important role, as in the problem at hand, slope can act as a multiplier factor for flow delivery.
4. *Breakpoint #1 (σ_1)* is the first breakpoint of a piecewise linear convex function modeling the non-linear nature of sediment flow. Breakpoints are required to define linear segments within the flow delivery functions. In this paper, *breakpoint #1* is considered as the retention capacity: if the total flow in a cell is less than *breakpoint #1*, it will not be delivered to the steepest downslope or any other neighbor.
5. *Breakpoint #2 (σ_2)* is the second breakpoint of the sediment flow delivery function. If total flow in a cell is between *breakpoint #1* and *breakpoint #2*, flow delivered to the dedicated neighbor is equal to *total flow − flow at breakpoint #1* multiplied by the *flow factor* in this cell. On the other hand, if *total flow* is larger than *breakpoint #2*, flow delivered is equal to (*breakpoint #2 − breakpoint #1*) ∗ *flow factor* + (*total flow − breakpoint #2*). This implies that flow exceeding *breakpoint #2* is fully delivered.

6. *New Flow Production (β), New Flow Factor(δ), New Breakpoint #1 (σ₃), and New Breakpoint #2 (σ₄)* are the new values for Flow Production, Flow Factor, Breakpoint #1, and Breakpoint #2 in the cells selected to be reforested; i.e. new parameters of the piecewise linear convex function. The effects of reforestation are assumed to be immediate. This is a gross simplification of reality.

7. *Intrinsic Suitabilities (μ)* are static criteria that do not affect neighboring locations of a cell selected as part of a target patch. In the problem at hand, estimates of *environmental performance attributes* as *carbon sequestration* and *nitrate leaching* obtained in reforested cells stand for intrinsic suitability.

8. *Streams* is a binary grid where cells upholding a value 1 represent streams or riverbeds.

Problem Representation. Locating best cells to be reforested in order to minimize sediment flow requires the specification of an outlet cell where total flow converges. Since the Flow Direction (FD) map assigns a unique path from each cell toward the steepest downslope neighbor, it is feasible to use a tree representation of this map in which the root node equals the outlet cell. Figure 1a shows a schematic representation of the five initial maps applied to minimize flow. The underlying tree structure shown in figure 1b is constructed from the *Flow Direction* map. Each node in the tree corresponds with a cell location, therefore *cell values* in each one of the 12 raster maps (matrices) can be referenced from the
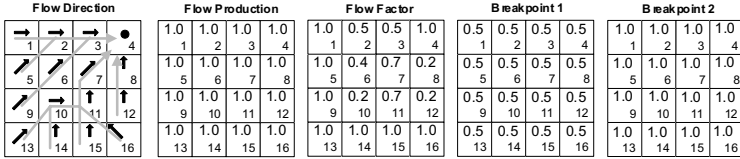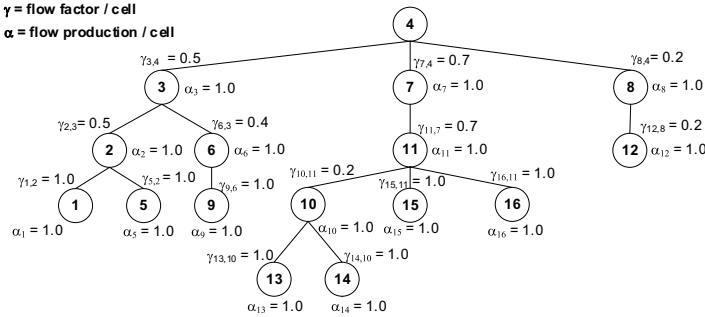

Fig. 1. (a) Data and (b) Problem Representation

corresponding node; e.g. nodes of the tree presented in figure1b are associated with flow factor between node i and node j ($\gamma_{i,j}$), and flow production (α_i) in node i.

Piecewise Linear Convex Function. The proposed IP formulation assigns a piecewise linear convex function (g) to every node in the tree. This function models flow delivery between nodes by decomposing the continuous flow function into finite convex segments, such that g is linear on each segment. Figure 2a shows input and output flow in node j. A function with two breakpoints ($\sigma_{j,1}, \sigma_{j,2}$,) and three segments ($y_{j,k,1}$, $y_{j,k,2}$, $y_{j,k,5}$) is defined to model flow delivery. Accumulated flow (EA_j) corresponds to the sum of flow coming into node j from two children nodes (h and i, fig. 2a) plus the flow produced in the node itself (α_j). Regarding output flow delivered to cell k, it depends on EA_j: 1) when it is less than or equal to the breakpoint #1 ($\sigma_{j,1}$) flow in segment 1 ($y_{j,k,1}$) is completely retained, i.e. no flow is delivered to the parent node (k); 2) if it is less than or equal to breakpoint #2 ($\sigma_{j,2}$) and larger than breakpoint #1 ($\sigma_{j,1}$), a fraction $\gamma_{j,k}$ of the flow in segment 2 ($y_{j,k,2} = EA_j - \sigma_{j,1}$) is delivered; 3) when it is larger than breakpoint #2 ($\sigma_{j,2}$), a fraction $\gamma_{j,k}$ of the flow in segment 2 ($y_{j,k,2} = \sigma_{j,2} - \sigma_{j,1}$) plus the entire flow in segment 3 ($y_{j,k,5} = \sigma_{j,2} - EA_j$) is delivered to the parent node (k).

Effective Accumulation (EA) in Nodes. The notion of Effective Accumulation (EA) is important at this point to better explain the IP formulation. According to the representation in figure 1, sediment flow which is Effectively Accumulated (EA) in nodes without children (leafs) equals flow produced in the node itself (α_i) without taking into consideration any other contribution or *interaction*. On the other hand, EA for nodes other than leafs is computed level by level, from bottom to top, considering *flow production* (erosion) in node i (α_i) and the flow delivered by its children. Therefore, computation of EA is performed from the leafs until the root is reached. In the piecewise linear convex function in figure 2b, the X axis corresponds with the EA at node j.
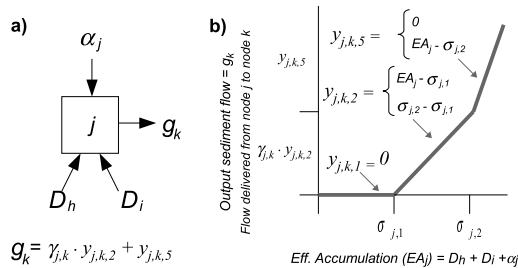

Fig. 2. Piecewise linear convex function describing sediment flow from one cell to another

3.2 Integer Programming Formulation

Network Flow Notation. The IP model, based on the general Network Flow (NF) formulation with piecewise linear convex functions, locates a given number of cells to be reforested in order to minimize sediment flow at a watershed outlet. This formulation requires that each individual link between a pair of nodes (j, k) in the original tree representation (fig 1) is substituted by a set of five links: $y_{j,k,1}$, $y_{j,k,2}$, $y_{j,k,3}$, $y_{j,k,4}$, $y_{j,k,5}$ (fig 3); each link is a pathway for delivering flow from node j to node k. Whereas segments $y_{j,k,1}$ and $y_{j,k,2}$ are used to transport flow to k when node j is not reforested, $y_{j,k,3}$ and $y_{j,k,4}$ are used when node j is reforested. Segment $y_{j,k,5}$ collects any flow exceeding either $\sigma_{j,2}$ when j is not selected or $\sigma_{j,4}$ when j is selected.

The following notation is introduced to facilitate the explanation of the flow minimization section of the IP formulation.

α_i	Flow production in the i_{th} node.
β_i	Flow production in the i_{th} node after it is reforested.
$\gamma_{i,j}$	Flow transport factor from node i to node j (flow factor).
$\delta_{i,j}$	Flow transport factor from node i to node j (flow factor) after node i is selected.
η	Maximum amount of flow reaching the root node, i.e. flow when 0 nodes are reforested.
ζ	Minimum amount of flow reaching the root node, i.e. considering only flow minimization.
$\sigma_{i,1}$	First breakpoint of the piecewise linear convex function in the i_{th} node.
$\sigma_{i,2}$	Second breakpoint of the piecewise linear convex function in the i_{th} node.
$\sigma_{i,3}$	First breakpoint of the piecewise linear convex function in the i_{th} node (i.e. when the node is reforested).
$\sigma_{i,4}$	Second breakpoint of the piecewise linear convex function in the i_{th} node (i.e. when the node is reforested).
$y_{i,j,1}$	Amount of flow transported from node i to node j through segment 1 ($s1$) in the piecewise linear convex function, where $s1$ is in the interval $[0 .. \sigma_{i,1}]$.
$y_{i,j,2}$	Amount of flow transported from node i to node j through segment 2 ($s2$) in the piecewise linear convex function, where $s2$ is in the interval $[\sigma_{i,1} .. \sigma_{i,2}]$.
$y_{i,j,3}$	Amount of flow transported from node i to node j through segment 3 ($s3$) in the piecewise linear convex function, where $s3$ is in the interval $[0 .. \sigma_{i,3}]$.
$y_{i,j,4}$	Amount of flow transported from node i to node j through segment 4 ($s4$) in the piecewise linear convex function, where $s4$ is in the interval $[\sigma_{i,3} .. \sigma_{i,4}]$.
$y_{i,j,5}$	Amount of flow transported from node i to node j through segment 5 ($s5$). Any flow exceeding either $\sigma_{i,2}$ or $\sigma_{i,4}$ will be sent along this segment.
P	Set of cells available to be reforested, i.e. cells that do not belong to riverbeds or background.
φ	Number of target cells.
f_i	Upholds a value 1 when the node i is selected to be reforested.
w_f	Weight assigned to the flow minimization criterion.

In line with the formulation explained in section 2.1, figure 4a shows a generic piecewise linear convex function for nodes not selected as optimal to be reforested in order to minimize flow, and figure 4b shows the function for reforested nodes. Whereas figure 4b represents a modified situation, the former function (figure 4a) models initial conditions in a node. X axes in figures 4a and 4b correspond with sediment flow coming into node j from its children (i_s) plus flow produced in the node j itself (EA_j). When node j is not reforested (fig 4a), flow delivered to its parent (node k) follows the procedure explained in section 3.1. The criteria in this section (3.1) are also applied to determine the amount of flow to be delivered by node j when it is reforested (fig 4b). In the last case, to determine flow along segments $y_{j,k,3}$, $y_{j,k,4}$ and $y_{j,k,5}$, the parameters of the piecewise linear convex function are modified: 1) breakpoints ($\sigma_{j,1}$ by $\sigma_{j,3}$, and $\sigma_{j,2}$ by $\sigma_{j,4}$), 2) segments ($y_{j,k,1}$ by $y_{j,k,3}$ and $y_{j,k,2}$ by $y_{j,k,4}$), 3) flow factors ($\gamma_{j,k}$ by $\delta_{j,k}$), and 4) flow
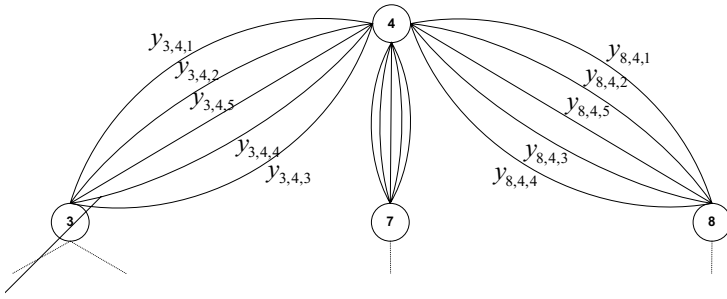
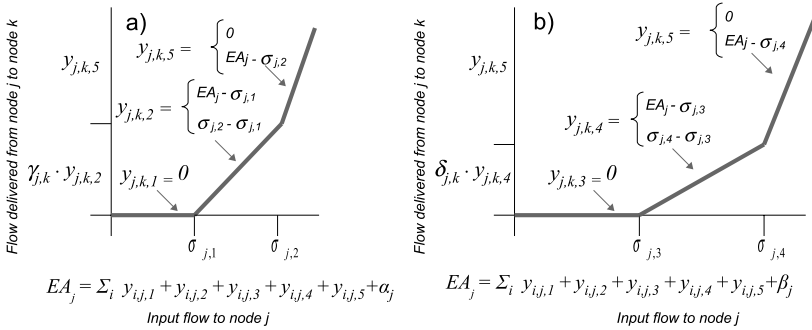**Fig. 3.** Segments for delivering flow from one node to another

Fig. 4. a) Piecewise linear convex function for cells: a) not selected to be reforested, b) selected to be reforested

production (α_j by β_j). Therefore, when flow is delivered along segments $y_{j,k,2}$ or $y_{j,k,4}$, those flows are multiplied by factors $\gamma_{j,k}$ and $\delta_{j,k}$ respectively. Since these factors are between 0 and 1, flow can be partially delivered. On the contrary, flow along segment $y_{j,k,5}$ is always fully transported from node j to node k.

Compactness and Intrinsic Criteria Notation. In order to fulfill the compactness criterion, a matrix is required for guiding the search of adjacent cells minimizing the perimeter of the target patch. Moreover, the selected cells must also maximize the intrinsic environmental performance criteria (carbon sequestration and nitrate leaching). The following notation is applied in the IP formulation to meet these requirements.

ω_N	Weight assigned to the compactness criterion.
$u_{i,j}$	Upholds a value 1 when the i,j cell and its upper cell are selected.
$d_{i,j}$	Upholds a value 1 when the i,j cell and its down cell are selected.
$l_{i,j}$	Upholds a value 1 when the i,j cell and its left cell are selected.
$r_{i,j}$	Upholds a value 1 when the i,j cell and its right cell are selected.
ρ	Number of intrinsic criteria considered in the decision problem.
ω_k	Weight assigned to the k_{th} intrinsic criterion.
$\mu_{k,i,j}$	Intrinsic suitability value for the k_{th} criterion in the i,j cell.
$c_{i,j}$	Upholds a value 1 when the i,j cell is selected.

IP Formulation for Allocating Cells Fulfilling Compactness and Flow Minimization Criteria. In order to deal with the tree and matrix structures, each part of the combined formulation applies two different sets of variables specifying the nodes or cells selected to be reforested: $F = \{f_1, f_2, ..., f_t\}$ and $C = \{c_{1,1}, ..., c_{1,n}, c_{2,1}, ...c_{2,n}, ..., c_{m,n}\}$. Whereas the set F applies one index to select nodes reducing flow, variables in the set C make use of two indexes to model cells adjacency and to select the ones fulfilling the compactness requirement.

The first part of the objective function (eq 4) in the Integer Programming (IP) formulation minimizes flow reaching the outlet. Flow delivered to the root node (outlet) from its j children is computed by multiplying flow along segments $y_{j,root,2}$ and $y_{j,root,4}$ by their respective flow factors ($\gamma_{j,root}, \delta_{j,root}$). The maximum and minimum possible amount of flow at the root (η, ζ) are used to normalize it. Whereas maximum amount of flow is achieved when no nodes are reforested ($\varphi = 0$), minimum is achieved when flow minimization is the only criterion considered in the objective function ($w_k = 0 \; \forall k, w_N = 0, w_f = 1, \varphi = $ *number of target cells*). The second part of the objective function (eq 4) maximizes the intrinsic criteria of the selected cells, i.e. carbon sequestration and nitrate leaching, where $c_{i,j}$ is a binary variable upholding a value 1 if the cell in row i and column j is part of the target site to be reforested. While ρ is the number of intrinsic criteria considered in the decision problem, ω_k is the importance weight for the k_{th} criterion, and $\mu_{k,i,j}$ upholds a normalized suitability for the k_{th} intrinsic criterion in the cell i, j. The third part of the objective function (eq 4) maximizes the compactness criterion, where ω_N is the weight assigned to this requirement. To construct a compact site, the model considers the Von Neumann neighborhood of a cell (upper, down, left and right cells).

maximize:

$$w_f * (\eta - \sum_j (\gamma_{j,root} \cdot y_{j,root,2} + \delta_{j,root} \cdot y_{j,root,4} + y_{j,root,5}))/(\eta - \zeta) \quad +$$

$$\sum_k^\rho (\omega_k * (\sum_i^m \sum_j^n c_{i,j} * \mu_{k,i,j})/\varphi) \quad + \tag{4}$$

$$\omega_N * (\sum_i^m \sum_j^n 0.25 * (u_{i,j} + d_{i,j} + l_{i,j} + r_{i,j}))/\varphi$$

s.t:

$$EA_j = \sum_i (\gamma_{i,j} \cdot y_{i,j,2} + y_{i,j,5} + \delta_{i,j} \cdot y_{i,j,4}) + \alpha_j \cdot (1 - f_j) + (\beta_j \cdot f_j) \;\; \forall j \tag{5}$$

$$EA_j = y_{j,k,1} + y_{j,k,2} + y_{j,k,3} + y_{j,k,4} + y_{j,k,5} \qquad \forall j, k \tag{6}$$

$$y_{j,k,1} \le \sigma_{1,j} \cdot (1 - f_j) \qquad\qquad\qquad \forall j, k \tag{7}$$

$$y_{j,k,2} \le (\sigma_{2,j} - \sigma_{1,j}) \cdot (1 - f_j) \qquad\qquad \forall j, k \tag{8}$$

$$y_{j,k,3} \le \sigma_{3,j} \cdot f_j \qquad\qquad\qquad\qquad \forall j, k \tag{9}$$

$$y_{j,k,4} \le (\sigma_{4,j} - \sigma_{3,j}) \cdot f_j \qquad\qquad\quad \forall j, k \tag{10}$$

$$c_{i,j} - x_{i,j} \geq 0 \qquad \forall i,j; \quad \forall x_{i,j}, \ where \ x_{i,j} \in \{u_{i,j}, d_{i,j}, l_{i,j}, r_{i,j}\} \qquad (11)$$

$$c_{i-1,j} - u_{i,j} \geq 0 \qquad \forall i,j \qquad (12)$$

$$c_{i+1,j} - d_{i,j} \geq 0 \qquad \forall i,j \qquad (13)$$

$$c_{i,j-1} - l_{i,j} \geq 0 \qquad \forall i,j \qquad (14)$$

$$c_{i,j+1} - r_{i,j} \geq 0 \qquad \forall i,j \qquad (15)$$

$$c_{i,j} = f_k \qquad \forall k; \quad \forall i,j \qquad (16)$$

$$f_k = 0 \qquad \forall k \notin P \qquad (17)$$

$$\sum_i^m \sum_j^n c_{i,j} = \varphi \qquad (18)$$

$$c_{i,j} \in \{0,1\} \qquad \forall i,j \qquad (19)$$

$$u_{i,j} \in \{0,1\} \qquad \forall i,j \qquad (20)$$

$$d_{i,j} \in \{0,1\} \qquad \forall i,j \qquad (21)$$

$$l_{i,j} \in \{0,1\} \qquad \forall i,j \qquad (22)$$

$$r_{i,j} \in \{0,1\} \qquad \forall i,j \qquad (23)$$

$$f_k \in \{0,1\} \qquad \forall k \qquad (24)$$

The binary variable $u_{i,j}$, will uphold a value 1 if the cell i,j is selected as part of the target patch (reforested, $c_{i,j} = 1$), and at the same time its upper neighbor cell is *also* selected. The same judgment is applied for assigning values to the binary variables $d_{i,j}$, $l_{i,j}$, $r_{i,j}$, which are associated to the *down*, *left*, and *right* neighbor cell respectively. Since the objective function is dealing with normalized values, and applies the Von Neumann neighborhood in the compactness criterion, the sum of $u_{i,j}$, $d_{i,j}$, and $r_{i,j}$ is multiplied by a 0.25 factor. In this manner, while the maximum level regarding the compactness contribution of the cell i,j is 1 (four neighbors of $c_{i,j}$ are *also* selected), the minimum level is 0 (no neighbors of $c_{i,j}$ are *also* selected to be reforested).

Equations 5 and 6 balance input and output flows at node j. In equation 5, flow in node j equals flow coming from its children nodes (i_s) plus flow produced in the node j itself. In fact equations 5 and 6 stand for the Effective Accumulation (EA), and unlike explanation in section 3.1, these equations consider that a node can be either reforested ($f_j = 1, \delta_{i,j}, \beta_j$) or not ($1 - f_j = 1, \gamma_{i,j}, \alpha_j$). Amount produced in the *jth* node is given either by β_j when the node is reforested, or by α_j when it is not. Although flow larger than 0 can also assigned to segments $y_{i,j,1}$ and $y_{i,j,3}$, they are not considered in equation 5 in order to model the problem in such a way that flow in node i is retained when it is lower than $\sigma_{i,1}$ or $\sigma_{i,3}$ (fig 4). Equation 6 computes flow delivered from node j to k, where k is the parent node. Even though in this equation all segments are summed, only segments $y_{j,k,1}$, $y_{j,k,2}$ and $y_{j,k,5}$ are used when the *jth* node is not selected to be reforested, in this case segments $y_{j,k,3}$ and $y_{j,k,4}$ uphold a value 0 (not used). On the contrary, the formulation makes use of segments $y_{j,k,3}$, $y_{j,k,4}$ and $y_{j,k,5}$ when the *jth* node is reforested. When flow leaving node j exceeds either $\sigma_{j,2}$ (node j is not reforested) or $\sigma_{j,4}$ (node j is reforested), this flow is delivered

through segment $y_{j,k,5}$. In any case, sediment flow along this segment is always fully delivered to the parent node (k). The amount of flow delivered along each segment is constrained by equations 7 and 8 when the jth node is not reforested $(1-f_j = 1)$, and by equations 9 and 10 in case the jth node is reforested $(f_j = 1)$.

Regarding compactness, the set of constraints in equation 11 guarantees that $u_{i,j}$, $d_{i,j}$, $l_{i,j}$, or $r_{i,j}$ can have a value 1 if and only if the cell i,j is selected $(c_{i,j} = 1)$ to be reforested. Constraint in equation 12 assures that if $u_{i,j}$ is equal to 1, the matching cell in the set of binary variables C, will be necessarily equal to 1 $(u_{i,j} = c_{i-1,j})$. Constraints from equation 13 to equation 15 articulate the previous statement for variables $d_{i,j}$, $l_{i,j}$, and $r_{i,j}$. These integrality constraints (equations 12 to 15), tend to effectively reduce the perimeter in order to form a compact and contiguous area.

Although C refers to cells, and F to nodes, both sets refer to the same spatial location, so that every variable in C has a corresponding variable in F. Whereas constraints in equation 16 keep the integrality between these two sets of variables, equation 17 assures that the solution is composed by nodes belonging to the set P of available ones. Finally, equation 18 restricts the number of selected cells (or nodes) to be equal to φ, and equations 19 to 24 designate variables $c,u,d,l,r,$and f as binary.

4   Results

Data are taken from the demonstration set accompanying Grass 6.3.0 windows software. Two watersheds within South Dakota in the USA (fig 5c), made up of 299 and 536 cells, are used to test the IP formulation. Raster maps in these regions are composed of cells of $30m$ x $30m$.

The formulation developed in the course of this paper is a useful tool to make decisions regarding the location of the best sites to be reforested. The proposed formulation searches for a set of cells minimizing sediment flow reaching the watershed outlet when their transport function changes, i.e. breakpoint #1, breakpoint #2, and flow factor are modified. Moreover this set of cells must form a compact patch optimizing two intrinsic characteristics: maximize carbon sequestration (μ_1) and minimize nitrate leaching (μ_2). While the left side of table 1a shows the parameters used for the current and new *flow production, flow factor, breakpoint #1 and breakpoint #2*, the right side defines 9 test cases assigning different weights to *compactness, flow minimization, carbon sequestration and nitrate leaching*. Since compactness and flow minimization are the most interesting criteria, these are prioritized in the test cases.

The following maps are prepared by means of Grass 6.3.0: 1) *Flow Direction* (FD), 2) Erosion representing *Flow Production* (α), 3) a normalized slope map for *Flow Factor* (γ), 4) riverbeds for *Streams*, 5) land cover to estimate Carbon Sequestration (CS) (μ_1), and 6) soils to estimate Nitrate Leaching (μ_2). 0.5 and 1.0 are the original values placed in every cell of maps representing *Breakpoint #1* (σ_1) and *Breakpoint #2* (σ_2) respectively, *New Breakpoint #1* (σ_3) and *New Breakpoint #2* (σ_4) are assigned values of 1.0 and 2.0 (table 1a). Finally, a

Table 1. Results obtained when identifying optimal cells for maximizing the intrinsic performance, maximizing compactness, and minimizing flow reaching the outlet in basins 1 and 2

a)

Parameter	Initial Value	Parameter	Test Values	Parameter	T1	T2	T3	T4	T5	T6	T7	T8	T9
								(Weights)					
Flow Yield (α_i) Erosion		New Flow Yield (β_i)	Erosion	Compactness (ω_N)	0,25	0,4	0,52	0,61	0,7	0,2	0,16	0,13	0,1
Flow Factor (γ_i) Slope		New Flow Factor (δ_i)	0,2	Flow reduction (ω_f)	0,25	0,2	0,16	0,13	0,1	0,4	0,52	0,61	0,7
Breakpoint 1 $(\sigma_{i,1})$ 0,5		New Breakpoint 1 $(\sigma_{i,3})$	1,0	Carbon Sequestration $(\omega_{k=1})$	0,25	0,2	0,16	0,13	0,1	0,2	0,16	0,13	0,1
Breakpoint 2 $(\sigma_{i,2})$ 1,0		New Breakpoint 2 $(\sigma_{i,4})$	2,0	Nitrate Leaching $(\omega_{k=2})$	0,25	0,2	0,16	0,13	0,1	0,2	0,16	0,13	0,1

b)

Test Cases	Basin 1 Size = 299 cells				Basin 2 Size = 536 cells			
	Objective Value	Time [sec]	Compact. Index	Flow at root	Objective Value	Time [sec]	Compactness	Flow at root
T1	0,73	23	0,54	755,21	0,76	67	0,61	1540,21
T2	0,72	37	0,55	755,98	0,77	80	0,88	1539,75
T3	0,73	42	0,88	758,69	0,78	298	0,88	1539,75
T4	0,75	84	0,88	758,69	0,79	387	0,88	1539,75
T5	0,76	87	0,88	758,69	0,79	9087	0,88	1539,75
T6	0,76	24	0,40	751,38	0,78	308	0,81	1539,71
T7	0,79	16	0,37	750,34	0,80	74	0,65	1539,63
T8	0,82	21	0,33	749,92	0,83	66	0,51	1537,84
T9	0,86	20	0,29	748,79	0,86	64	0,47	1536,8

fraction 0.2 is assigned to every cell in the *New Flow Factor* (δ) map. *New Flow Production* (β) keeps the same values as the initial *flow production* (α). This last condition is not as in reality but the absence of erosion levels after reforestation forces to it, nevertheless it does not affect the generality of the model.

The test cases of the IP model are implemented by means of the Lingo language v7 to select 30 objective cells. Table 1b shows the results obtained with the IP model: objective values, computation times, compactness index (given by equation 25) and flow reaching the root. Objective values increase with increasing the weights assigned to both criteria: compactness and flow minimization. Nevertheless, according to the test cases T6 to T9 in table 1b, objective values for basins 1 and 2 are larger when larger weights are assigned to flow minimization, i.e. this criterion has more impact in the objective function. This behavior is most likely produced because the cells selected to be reforested were initially contributing with high amounts of sediment flow to the outlet. On the other hand, computation times needed by the model increase with the weight assigned to compactness. Compact areas seem to be more hardly achieved. In agreement with the logic of the IP formulation, while the compactness increases when larger weights are given to this criterion (test cases T2 to T5), flow reaching the root decreases with larger weights to flow minimization (test cases T6 to T9).

$$SI = \frac{\sqrt{area}}{0.282 \cdot perimeter} \tag{25}$$

Figure 5a shows erosion, slope, streams, carbon sequestration and nitrate leaching maps used in the test cases T1 to T9 in watershed 2. Figure 5b shows the location of the 30 optimal cells selected to be reforested. The resulting maps in figure 5b show that the test case with the highest weight assigned to compactness (T5) achieves a compact patch near the riverbed. On the other hand, the test case with the highest weight on flow minimization (T9) result in a fragmented pattern. Moreover the low weights assigned to carbon sequestration and nitrate leaching are also reflected in the results.
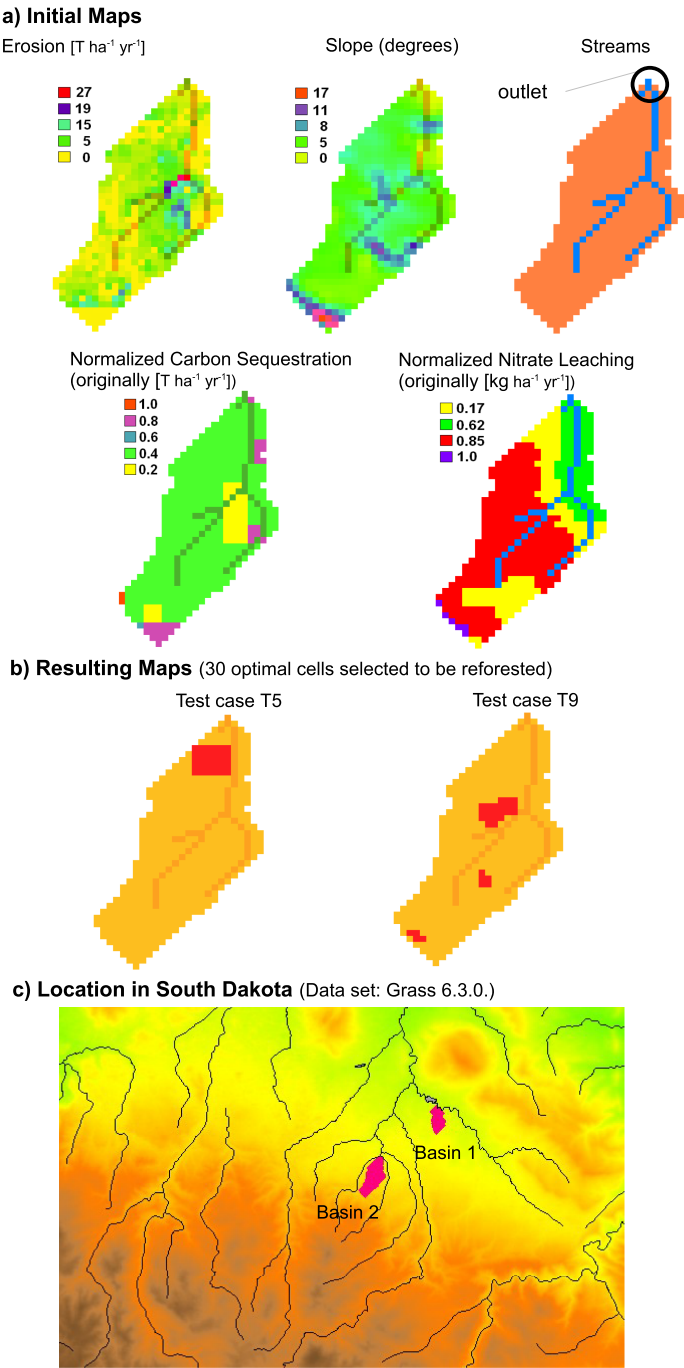
Fig. 5. Allocation of 30 compact cells minimizing flow at the outlet in basin 2

5   Conclusions and Future Work

An Integer Programming (IP) formulation has been proposed and tested in order to identify sets of cells for reforestation (1) minimizing sediment flow reaching a watershed outlet, (2) maximizing intrinsic performance, and (3) maximizing compactness. The formulation is based on two data structures: a tree for minimizing flow by means of the general Network Flow (NF) model, and a matrix for allocating compact sites.

Several test cases were configured and applied to two watersheds in order to analyze the IP formulation. The results show that the IP model is appropriate to combine the requirements of flow minimization, compactness and intrinsic suitability maximization. According to the resulting maps, an adequate configuration of the weights can give priority either to compactness or flow. Regarding computation times, the compactness requirement is harder to achieve than flow minimization.

The results obtained are in agreement with expert assessments of erosion levels, slopes and distances to the riverbeds; however the IP model requires a high number of variables, mainly produced because of the five links needed in the network flow part, and the four extra variables (u,d,l,r) for modeling compactness. This memory requirement will restrict the size of the problems. Finally, further evaluation of validity is required by comparisons with process models.

References

1. Vanegas, P., Cattrysse, D., Van Orshoven, J.: 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., Gavrilova, M.L. (eds.) ICCSA 2008, Part I. LNCS, vol. 5072, pp. 389–404. Springer, Heidelberg (2008)
2. Church, R., Gerrard, R., Gilpin, M., Stine, P.: Constructing cell-based habitat patches useful in conservation planning. Annals of the Association of American Geographers 93, 814–827 (2003)
3. Gilbert, K., Holmes, D., Rosenthal, R.: A multiobjective discrete optimization model for land allocation. Management Science 31, 1509–1522 (1985)
4. Ahuja, R., Magnanti, T., Orlin, J.: Network Flows: Theory algorithms, and applications. Prentice-Hall, Englewood Cliffs (1993)
5. Van-Orshoven, J.: Introduction to spatial data modelling and functionality of geospatial technology. Department of Earth and Environmental Sciences. K.U.Leuven (2007)
6. Hof, J., Bevers, M.: Direct spatial optimization in natural resource management: Four linear programming examples. Annals of Operations Research 95, 67–91 (2000)
7. Church, R., ReVelle, C.: The maximal covering location model. Regional Science Association 32, 101–118 (1974)
8. Dimopoulou, M., Giannoikos, I.: Spatial optimization of resources deployment for forest-fire management. International Transactions in Operational Research 8, 523–534 (2001)

9. Church, R., Stoms, D., Davis, F., Okin, B.: Planning management activities to protect biodiversity with a gis and an integrated optimization model. In: Proceedings of the Third international conference/workshop on Integrating GIS and environmental modeling (1996)
10. Fischer, D., Church, R.: Clustering and compactness in reserve site selection: An extension of the biodiversity management area selection model. Forest Science 49, 555–565 (2003)
11. Williams, J.: A linear-size zero-one programming model for the minimum spanning tree problem in planar graphs. Networks 39, 53–60 (2001)
12. Williams, J.: A zero-one programming model for contiguous land acquisition. Geographical Analysis 34, 330–349 (2002)
13. Shirabe, T.: Modeling topological properties of a raster region for spatial optimization. In: Proceedings of the 11th International Symposium on Spatial Data Handling (2004)