Selecting sites for afforestation to minimize sediment loss from a river basin: Computational complexity of Single and Multiple Flow Direction Methods in raster databases

Grethell Castillo-Reyes[a,b], René Estrella[c], Karen Gabriels[c], Jos Van Orshoven[c], Floris Abrams[c] and Dirk Roose[a]

[a]Department of Computer Science, KU Leuven, Belgium
[b]Data Representation and Analysis Center, University of Informatic Sciences, Cuba
[c]Division of Forest Nature and Landscape, Department of Earth and Environmental Sciences, KU Leuven, Belgium

ARTICLE INFO

Keywords:
Sediment Flow Simulation
Multiple Flow Direction
Single Flow Direction
Sediment Yield Minimization
Afforestation

ABSTRACT

To minimize the sediment flowing to the outlet of a river catchment with minimal effort or cost, the best areas to perform a certain intervention, e.g., afforestation, must be selected. CAMF is a method that performs this selection process iteratively in a raster geo-database. The original version of CAMF uses a Single Flow Direction (SFD) algorithm to simulate the flow paths. Although SFD is often used in studies related to hydrological applications, it may fail to reflect the nature of flow transport, depending on the topography. This paper describes and analyzes the integration of Multiple Flow Direction (MFD) methods in CAMF, in order to provide more accurate flow simulations in areas with specific topographic characteristics. We compare the computational complexity of CAMF-SFD and CAMF-MFD using two methods: FD8 and D∞, and we discuss the scalability w.r.t. the number of cells involved. We evaluate the behavior of these three variants for sediment yield minimization by afforestation in two catchments with different properties.

CRediT authorship contribution statement

Grethell Castillo-Reyes: Study, Conceptualization, Design, Software development, Analysis and interpretation of results, Writing. **René Estrella:** Generation of data, Analysis and interpretation of results, Review. **Karen Gabriels:** Generation of data. **Jos Van Orshoven:** Conceptualization, Supervision, Analysis and interpretation of results. **Floris Abrams:** Generation of data, Review. **Dirk Roose:** Conceptualization, Design, Methodology, Supervision, Analysis and interpretation of results, Review, Writing.

1 Introduction

Among the main issues analyzed in Spatial Decision Support Systems, land use planning is of vital importance. It comprises the management and modification of the natural environment, focusing on the development of effective strategies for land conservation and urbanization (Randolph, 2012). It allows to select potential sites to carry out a certain intervention, taking into account criteria related to the use of the land, depending on the application.

Sediment delivery to rivers reduces channel and reservoir capacity and leads to water quality problems and biodiversity decline because of suspended mineral and organic substances (Drzewiecki and Mularz, 2008). These issues are particularly undesirable in regions where the river is used for drinking or irrigation water provision or for electricity production.

ORCID(s): 0000-0002-6922-7735 (G. Castillo-Reyes)

Afforestation has proven to be effective to reduce sediment production and delivery and to alleviate the associated water quality problem (Costin, 1980; Nearing et al., 2005; Heil et al., 2007; Vanwalleghem, 2017). It protects the soil surface against the ability of raindrops and runoff to detach and transport soil particles (B.-M. Vought et al., 1995; Piégay et al., 2004). An important question at the beginning of any afforestation project is to identify the most suitable sites to plant the trees. The discrimination between suitable and unsuitable areas for afforestation typically depends on several criteria adopted by forest planners, which can range from on-site and off-site environmental concerns to maximizing financial profits.

This spatial optimization problem can be formulated as a mathematical programming model that can be solved exactly, but typically requires a high amount of computational resources (Fischer and Church, 2003). While this approach has been used for many years in forest planning (Williams and Revelle, 1997), the size of the problems that can be handled remains limited (Vanegas et al., 2011). Therefore, as an alternative, Vanegas (2010); Vanegas et al. (2012, 2014) proposed the CAMF method to locate the sites within a river catchment that should be afforested in order to minimize the amount of sediment that reaches the outlet(s).

Although finding the optimal solution cannot be guaranteed in general, they proved theoretically that CAMF obtains the optimal solution for a small raster under certain conditions. To simulate sediment transport, CAMF uses a simple, computationally efficient and spatially distributed model, that can be used in an optimization context.

A key issue in sediment flow simulation with CAMF is the spatial interaction (SI) among cells in the raster datasets, representing the catchment. SI 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, SI refers to the phenomenon that afforestation of a cell leads to changes of its characteristics, which in turn affect the amount of sediment flowing from that cell into its downstream neighbor cells and, eventually, the sediment yield of the catchment.

The original implementation of CAMF uses a Single Flow Direction (SFD) method, which assumes that flow leaving a cell is delivered entirely to the neighbor at the lowest altitude. Despite the advantage of simplicity of SFD methods, they may fail to reflect the real nature of surface transport processes (Quinn et al., 1991; Wilson and Gallant, 2000; Zhou and Liu, 2002). Therefore Quinn et al. (1991); Costa-Cabral and Burges (1994); Tarboton (1997); Seibert and McGlynn (2007) proposed Multiple Flow Direction (MFD) methods, in which the flow is distributed to one or more of the down-slope neighboring cells, representing sediment flow pathways more appropriately than SFD in some cases, e.g. for the estimation of a specific catchment area, landscape evolution simulations (Gallant and Hutchinson, 2011; Anand et al., 2020) and the computation of topographic indices (Pan et al., 2004; Rampi et al., 2014).

With SFD even a tiny elevation difference between two neighboring cells can have a large effect, since it might determine which cell receives all the outgoing flow. Small elevation differences have a less important effect in a MFD method, since cells with slight differences in elevation receive about the same amount of flow.

In this paper, we develop and analyze MFD variants of CAMF (CAMF-MFD) using two flow direction methods: the Fractional Deterministic Eight-Neighbor (FD8) method (Quinn et al., 1991) and the D∞ method (Tarboton, 1997). Compared to the SFD methods, MFD methods lead to much more complex flow networks, with multiple paths between cells, which not only result in a higher computational cost, but it also makes difficult to determine the order in which the cells network should be traversed to simulate sediment transport. We use a topological sorting algorithm proposed by Kahn (1962) and applied in Anand et al. (2020) to achieve the appropriate ordering of cells in the flow network generated with the MFD methods. We present algorithms for the various steps in the optimization procedure and results for several test cases. We compare the behaviour, the performance and the scalability w.r.t. the problem size of the single and multiple flow variants of CAMF.

This paper is organized as follows. Section 2 provides a general description of CAMF, the calculation of the sediment accumulation raster and the main steps followed by CAMF to select the best areas for afforestation. It also describes particular characteristics of CAMF-SFD and its reformulation to use a MFD method. Section 3 presents the case studies used for the experiments. In Section 4 the results of experiments are discussed and Section 5 presents some conclusions.

2 The CAMF method for minimizing sediment yield

Using a raster geo-database representing a river catchment, the CAMF method (Vanegas, 2010; Vanegas et al., 2012, 2014) selects cells, such that performing a certain intervention in these cells leads to the maximal reduction of sediment yield, i.e. the amount of sediment that reaches the outlet(s), denoted by SY. The cells are selected from a list of candidates, since the intervention can be performed only in cells corresponding to some land use types. Because of SI, cells are selected iteratively. In each iteration, CAMF computes for every candidate cell the sediment yield reduction, denoted by SYR, that will be achieved if the intervention is performed in that cell. The candidate cells are ranked based on their SYR values. The cells with a maximum SYR are selected for the intervention. Although the core structure of CAMF can be applied for different sorts of flow, we applied it for the specific case of sediment yield minimization by afforestation. Therefore, in this paper, we describe CAMF in this context.

The input required by CAMF consists of several characteristics of the study area represented as raster data-sets (Estrella, 2015):

- A Digital Elevation Model (DEM) to compute the gradients and the flow direction matrix to determine the flow paths within a catchment.
- A land use types map or coverage map to identify the cells in which the intervention can take place.
- Mean annual sediment produced locally before and after afforestation ($ton\,ha^{-1}\,yr^{-1}$).

The implemented sediment accumulation procedure requires the following inputs:

- Retention capacity before and after afforestation ($ton\,ha^{-1}\,yr^{-1}$): the amount of sediment that will be retained in a cell.
- Saturation threshold before and after afforestation ($ton\,ha^{-1}\,yr^{-1}$): the amount of sediment exceeding this value will be fully delivered from the cell to its down-slope neighbors.
- Flow factor, computed as a normalized value from the slope (dimensionless). It indicates the fraction of sediment transported from a cell to its down-slope neighbors.

The user also must specify the number of cells to be selected or the required amount of *SYR*.

We now present more details of the main processes in CAMF: (1) the computation of the initial sediment accumulation raster, denoted by *SA*, see sections 2.1-2.3, and (2) the iterative process for the identification of the best cells for the afforestation in order to minimize the flow reaching the outlet(s), see section 2.4.

2.1 Sediment Accumulation

The matrices storing the mean annual sediment produced locally, before and after afforestation, are denoted respectively by α^1 and α^2, while $\alpha_i^k,\ k = 1, 2$ denote the local sediment production in cell i, where $k = 1$ if the cell is not yet afforested and $k = 2$ if the cell is afforested. The Sediment Accumulation matrix *SA* gives for each cell i the total amount of sediment SA_i, i.e., α_i^k plus the sediment received by the cell i from its up-slope neighbours.

The amount of sediment leaving cell i, denoted by D_i, is a function of SA_i. The original version of CAMF uses a piece-wise linear convex function (Vanegas, 2010) defined by three parameters: retention capacity ρ_i^k, saturation threshold σ_i^k and flow factor γ_i^k. If SA_i is less than or equal to the retention capacity ρ_i^k, no sediment leaves the cell. If $SA_i \in (\rho_i^k, \sigma_i^k]$, a fraction γ_i^k of SA_i leaves the cell. If SA_i is larger than the saturation threshold σ_i^k, the difference $SA_i - \sigma_i^k$ is fully delivered from the cell. Hence D_i is modeled by

$$D_i = \begin{cases} 0, & \text{if } SA_i \le \rho_i^k \\ \gamma_i^k(SA_i - \rho_i^k), & \text{if } \rho_i^k < SA_i \le \sigma_i^k \\ \gamma_i^k(\sigma_i^k - \rho_i^k) + (SA_i - \sigma_i^k), & \text{if } SA_i > \sigma_i^k \end{cases} \tag{1}$$

The amount of sediment transported between cells i and j, denoted by $D_{i,j}$, is calculated as

$$D_{i,j} = D_i \times F_{i,j} \tag{2}$$

with D_i the amount of sediment that leaves cell i, calculated using Eq. 1 and $F_{i,j}$ the sediment fraction flowing from

cell i to the down-slope neighbor cell j.

In case of afforestation, the parameters of the piece-wise linear function change when the land use of a cell is converted from its initial state (not afforested) to its new state (afforested), see Fig. 1. Afforestation decreases the local sediment production and the flow factor, i.e., $\alpha_i^2 < \alpha_i^1$ and $\gamma_i^2 < \gamma_i^1$, and increases the retention capacity and the saturation threshold, i.e., $\rho_i^2 > \rho_i^1$ and $\sigma_i^2 > \sigma_i^1$. Hence, the selection of a cell reduces the amount of sediment delivered by this cell and also influences the amount of sediment flowing into the cells lying on the flow path(s) between this cell and the outlet(s).

2.2 CAMF-SFD

Originally, the sediment flow simulation in CAMF uses a SFD method, i.e., the sediment leaving a cell flows entirely to the neighbouring cell at lowest altitude. In particular, the Eight-Direction (D8) algorithm (O'Callaghan and Mark, 1984) is used: a cell c delivers all its flow to cell i (one of its eight neighbors), with i determined by

$$arg|min_{i=1,2,...8} Z_i \text{ if } Z_c > Z_i, \tag{3}$$

where Z_c is the elevation for cell c.

The flow direction in CAMF-SFD can be represented as a tree (Fig. 2), where each node corresponds to a raster cell (Fig. 2b), the root node corresponds to the outlet cell and the child-parent edges represent the flow direction following the steepest descent pathway Estrella (2015). To compute the *SA* raster, the tree is used to process the cells in the right order.

As mentioned above and explained in detail in section 2.4, in an optimization iteration a single candidate cell i will be (tentatively) afforested to compute the corresponding *SYR*. In this case, *SA* must not be recomputed completely, but only the SA_i values for all cells on the path between cell i and the root of the tree.

2.3 CAMF-MFD

In a MFD method a cell can receive flow from several cells as well as it can deliver flow to multiple cells, see Fig. 3, which means that the flow network generated with MFD methods is more complex than their SFD counterparts. Therefore, SI and cell connectivity can not be represented as a tree, but as a graph. The main difficulty is to build an appropriate ordering of the cells (nodes of the graph) to calculate *SA* based on the flow network.

To build the graph from a flow network generated with a MFD method, we follow the topological sorting algorithm proposed by (Kahn, 1962). A similar approach is used in Anand et al. (2020) for landscape-evolution simulations, in which the upstream nodes draining to a node i are called "donors" and the nodes to which i delivers flow are called "receivers".

Based on Kahn (1962), Algorithm 1 computes for all the cells an adjacency list with "ancestors" ("donors") and "successors" ("receivers"), following the flow direction matrix F (Figs. 4a, 4b). Cells that deliver sediment to cell i are called its ancestors, and cells that receive sediment from cell i are called its successors. Several methods to compute SA using a graph have been proposed, see e.g. (Qin and Zhan, 2012; Jiang et al., 2013; Barnes, 2019). However, traversing a graph that corresponds to a large data-set is very expensive in terms of memory and execution time. In the minimization process in CAMF, we compute the SA matrix many times. Therefore, in a pre-processing step, Algorithm 2 sorts the cells, so that for each ancestor–successor pair, the ancestor always appears in the sorted list before its successor (Fig. 4c). The SA matrix is then computed cell per cell in the order given by the sorted list, from left to right, see Algorithm 3.

In the case of SFD, SA must not be completely recomputed when a single candidate cell i is (tentatively) afforested to compute the corresponding SYR. However, in the case of MFD, SA_i for all cells on the paths between cell i and the outlet(s) must be recomputed, which corresponds to a substantially larger part of the SA matrix compared to the single path in SFD. Also, the cells i for which SA_i must be recomputed can only be determined by traversing a substantial part of the graph, which is very expensive and requires reading large amounts of data. Therefore, when cell i is (tentatively) afforested, SA_i is recomputed for all cells in the sorted list between cell i and the outlet(s). The latter approach is to be preferred, despite redundant calculations, since it avoids the data accesses required to process the graph. Note that a data access is much more expensive than a floating-point operation on current computers.

Algorithm 1 Create an adjacency list from the flow direction matrix

Input: MFD matrix F

for each cell i in F **do**
  for each neighbor j of i **do**
    if i drains into j **then**
      1. Add j as successor of i in A ▷ A is the adjacency list
      2. Add i as ancestor of j in A
    end if
  end for
end for

Output: Adjacency list A

In the MFD variant of CAMF we use the Fractional Deterministic Eight-Neighbor (FD8) method (Quinn et al., 1991) and the D∞ method (Tarboton, 1997) (see Appendix A for more details). However, the calculation of the SA matrix (Algorithms 1-3) is organised in such a way that several other MFD methods can easily be implemented, e.g. the Digital Elevation MOdel Networks (DEMON) (Costa-Cabral and Burges, 1994), the triangular MFD algorithm (Seibert and McGlynn, 2007) and the MFD algorithm based on Maximum Down-slope Gradient (MFD-md) (Qin et al., 2007) to adaptively model the flow partition based on local topographic attributes.

Algorithm 2 Sort the adjacency list

Input: Adjacency list A

for each cell i in A without ancestors **do**
  1. Store i in E ▷ E stores the cells without incoming ancestors
end for

while E is non-empty **do**
  2. Get cell i from E
  3. Remove cell i from E
  4. Add i to S ▷ S is the sorted list
  for each cell j in A with i as ancestor **do**
    5. Remove i from ancestors of j in A
    if j has no other ancestors **then**
      6. Add j to E
    end if
  end for
end while

Output: Sorted list S

Algorithm 3 Compute Sediment Accumulation SA

Input: Sorted list S, adjacency list A, local sediment production α^k, flow factor γ^k, retention capacity ρ^k, saturation threshold σ^k, with k indicating the values before afforestation ($k = 1$) or after afforestation ($k = 2$)

for each cell i in S **do**
  1. $SA_i \leftarrow \alpha_i^k$ ▷ SA stores the sediment accumulation for every cell
  for each ancestor j of i in A **do**
    if $SA_j \leq \rho_j^k$ **then**
      2. $D_j \leftarrow 0$
    else
      if $SA_j > \rho_j^k$ and $SA_j \leq \sigma_j^k$ **then**
        3. $D_j \leftarrow \gamma_j^k (SA_j - \rho_j^k)$
      else
        if $SA_j > \sigma_j^k$ **then**
          4. $D_j \leftarrow \gamma_j^k (\sigma_j^k - \rho_j^k) + (SA_j - \sigma_j^k)$
        end if
      end if
    end if
    5. $D_{j,i} \leftarrow D_j \times F_{j,i}$ ▷ $F_{j,i}$ is the fraction of sediment delivered from cell j to its neighbor cell i. For SFD $F_{j,i} = 1$, for FD8 and D∞ $F_{j,i}$ is calculated following the formulas in Appendix A.
    6. $SA_j \leftarrow SA_j - D_{j,i}$
    7. $SA_i \leftarrow SA_i + D_{j,i}$
  end for
end for

Output: Sediment accumulation SA

2.4 Minimizing sediment yield at the outlet(s)

Each iteration of the CAMF algorithm consists of the following three steps:

1. Each candidate cell that has not been selected in previous iterations is tentatively selected and the *SA* matrix is recomputed to evaluate the effect of afforesting this cell. In section 2.1 we indicated that when a cell is afforested the amount of sediment delivered to its neighbours decreases. This decrease is propagated to the outlet(s) and results in a decrease in *SY*. Let SY^b be the *SY* in the initial situation; SY_i^k the *SY* if cell i would be afforested in iteration k; SYR_i^k the total *SYR* if cell i would be afforested in iteration k. Then

$$SYR_i^k = SY^b - SY_i^k \tag{4}$$

2. Cells are ranked in descending order based on the SYR_i^k value.
3. The cell(s) with the maximal *SYR* value are added to the set of cells selected for afforestation.

The iterative procedure, see Algorithm 4, ends when the (user-specified) number of selected cells is fulfilled. Note that this stop criterion can easily be replaced by a test on achieving a user-specified *SYR*.

Algorithm 4 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*
  6. Put cell(s) with highest *SYR* in solution set S
end while

Output: Set of selected cells S

Algorithm 3 (computing *SA*) has a linear computational complexity in the number of active cells (with actual values) and the computational cost also depends on the chosen flow direction method, based on the complexity of the graph (e.g. the number of neighboring cells receiving flow from each node of the graph). Hence the computational complexity of Algorithm 4 (determine selected cells) is proportional to the number of active cells and the number of candidate cells (that can be of the order of the number of active cells) and to the number of cells that are selected (since often only one cell is selected per iteration).

3 Case studies

3.1 The Tabacay catchment

The first case study deals with the Tabacay river catchment in Ecuador, with an area of $\approx 6\,639\ ha$ and altitudes ranging from 2 482 m to 3 731 m (above sea level), see Fig. 5. Agriculture and pasture cover 39% of the area. The widespread agricultural land use, even on steep slopes, produces large amounts 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 active cells. Only a part of the cells under agriculture and pasture are considered as candidate cells (27 246 cells).

The initial sediment production map α^1 was computed using the Revised Universal Soil Loss Equation (RUSLE) proposed by Renard et al. (1991), as follows

$$\alpha^1 = R \times K \times LS \times C \times P \tag{5}$$

where α^1: annual soil loss ($ton\,ha^{-1}\,yr^{-1}$); R: rainfall erosivity factor ($MJ\,mm\,ha^{-1}\,h^{-1}\,yr^{-1}$); K: soil erodibility factor ($ton\,h\,MJ^{-1}\,mm^{-1}$); LS: slope length and slope steepness factors; C: cover management factor; P: support practice factor.

We set $P = 1$ and $R = 1\,599\ MJ\,mm\,ha^{-1}\,h^{-1}\,yr^{-1}$ based on Cisneros Espinosa et al. (1999). The K-factor is computed from soil granulometric fractions, see Table 1, using the equation introduced by Wischmeier (1978). The C-factor map is generated by assigning the C-values presented in Table 2, taken from Estrella (2015), to the land cover map (Fig. 6a). The LS-factor map is computed from the DEM using an ArcGIS tool (Conrad, 2003), based on the slope and specific catchment area, following the approach of Desmet and Govers (1996). The LS-factor map and the initial sediment production α^1 are shown in Figs. 6b, 6c.

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 of Eq. 1, i.e., ρ^1, ρ^2, σ^1, σ^2, α^2 and γ^2, listed in Table 3, 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.

Additionally, two smaller regions around the outlet of the catchment were cropped from the original data-set, to evaluate the scalability of CAMF w.r.t. the problem size (number of cells). Figs. 7a, 7b show the corresponding fragments of the initial local sediment production map α^1.

Table 1
K-factor calculated for each soil type in the Tabacay catchment (Estrella, 2015).

Soil Type	Soil Texture	K-factor ($ton\,h\,MJ^{-1}\,mm^{-1}$)
Umbric Leptosol	Sandy clay loam	0.0397
Umbric Andosol	Loam	0.0373
Dystric Cambisol	Loam	0.0376
Ferralic Cambisol	Clay loam	0.0253
Calcaric Regosol	Sandy loam	0.0452
Calcaric Cambisol	Sandy clay loam	0.0290
Eutric Cambisol	Silt loam	0.0518
Eutric Regosol	Sandy loam	0.0521

Table 2
C-factor values for each land use type in the Tabacay catchment (Estrella, 2015).

Land cover type	C-factor
Water	0
Forest	0.001
Bush	0.003
Pasture	0.003
Paramo	0.003
Agriculture	0.2

Table 3
Parameter values used in the experiments with CAMF for the Tabacay case study.

Parameter	Before afforestation	After afforestation
Sediment production	α^1, calculated by RUSLE, Fig. 6c	$\alpha^2 = 0.83 \times \alpha^1$
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	$\gamma^2 = 0.75 \times \gamma^1$

3.2 The Maarkebeek catchment

The second case study deals with the Maarkebeek river catchment in Belgium, with an area of $\approx 4\,800\ ha$, with altitudes from 14.1 m to 146.9 m, see the DEM in Fig. 8. It is covered mostly by agriculture, with around 10% of its area urbanized, and around 10% afforested (Gabriels et al., 2022). The resolution of the Maarkebeek geo-database is $20m \times 20m$ with 129 097 active cells, of which 53 792 are candidate cells for afforestation (cells under pasture and arable land), see the land cover visualization in Fig. 9. Since the Maarkebeek data-set contains more cells than the Tabacay data-set, the computational cost of executing CAMF on the former is substantially higher.

As in the Tabacay catchment, the initial local sediment production map α^1 (Fig. 10) was generated by means of RUSLE, Eq. 5. The values $P = 1$, $R = 870\ MJ\,mm\,ha^{-1}\,h^{-1}\,yr^{-1}$, the K-factor associated to each soil type (Table 4) and the C-factor assigned to each land use type (Table 5), were obtained from Deproost et al. (2018). The LS-factor map was computed using the same tool as for the Tabacay catchment.

The values for ρ^1, ρ^2, σ^1, σ^2, α^2 and γ^2 in Table 6 were selected such that $\approx 10\%$ of the initial available sediment flows to the outlet.

Table 4
K-factor calculated for each soil type in the Maarkebeek catchment (Deproost et al., 2018).

Soil Texture	K-factor $(ton\,h\,MJ^{-1}\,mm^{-1})$
Silty sand	0.02
Light sandy loam	0.025
Clay	0.042
Sand loam	0.4

Table 5
C-factor values assigned to each land use type in the Maarkebeek catchment (Deproost et al., 2018).

Land cover type	C-factor
Lake and Infrastructure	0
Grass, shrubs (by the side of the road)	0.001
Forest	0.01
Agricultural, pasture and arable land	0.37

Table 6
Parameter values used in the experiments with CAMF for the Maarkebeek case study.

Parameter	Before afforestation	After afforestation
Sediment production	α^1, calculated by RUSLE, Fig. 10	$\alpha^2 = 0.83 \times \alpha^1$
Retention capacity	$\rho^1 = 0.55 \times \alpha^1$	$\rho^2 = 0.73 \times \alpha^1$
Saturation threshold	$\sigma^1 = \alpha^1$	$\sigma^2 = 1.02 \times \alpha^1$
Flow factor	γ^1, normalized slope from DEM	$\gamma^2 = 0.75 \times \gamma^1$

4 Results

All experiments were performed on one core of a Xeon E5-2697 v3 CPU (2.6 GHz) with 128 GB of RAM, with Ubuntu Bionic Linux (18.044.15.0-147-generic x86 64 kernel) as Operating System. Although we performed the experiments on Linux, the implementation of CAMF in C++ can be used under Linux/Unix and Windows.

4.1 Sediment yield reduction with CAMF-SFD and CAMF-MFD

Fig. 12 shows the *SA* rasters of the Tabacay catchment for the initial situation, computed with CAMF-SFD and with the two variants of CAMF-MFD, i.e., CAMF-FD8 and CAMF-D∞. Since ρ^1 and σ^1 are always smaller than α^1, the initial *SY* should be independent of the flow direction method used. For each cell a part of the locally produced sediment and the incoming sediment flows completely to neighbouring (lower) cells. Hence, the total amount of incoming sediment in the outlet cell(s) must be similar regardless of the flow method used. The values reported in Tables 7 and 8 are nearly equal. Note that in our implementation of the FD8 method, the cells at the borders of the raster are treated somewhat differently than in the implementation of the other flow direction methods.

The *SYR* value increases when the number of afforested cells increases, see Table 7, in agreement with the fact that afforestation is an effective means to reduce sediment flow (Costin, 1980; Heil et al., 2007; Nearing et al., 2005). By afforesting a given percentage of candidate cells, the predicted *SYR* value is the highest when CAMF-FD8 is used, and

is the lowest when CAMF-SFD is used.

Table 7
SY and *SYR* obtained with CAMF-SFD, CAMF-FD8 and CAMF-D∞ for the Tabacay catchment. Afforestation of up to 30% of the candidate cells.

Initial *SY* ($ton\ yr^{-1}$)			# Afforested cells	*SYR* ($ton\ yr^{-1}$)			% *SYR*		
SFD	D∞	FD8		SFD	D∞	FD8	SFD	D∞	FD8
37 553	37 558	36 774	5% = 1 362	10 361	10 652	11 244	27.59	28.36	30.58
			10% = 2 724	11 970	12 330	12 945	31.88	32.83	35.20
			20% = 5 448	12 958	13 365	14 034	34.51	35.59	38.16
			30% = 8 172	13 211	13 629	14 310	35.18	36.29	38.91

Fig. 13a shows the *SYR* in function of the number of afforested cells. The extra yield reduction per 1000 selected cells decreases with increasing number of afforested cells, and becomes small after afforesting ≈ 3000 cells, confirming that CAMF first selects the cells leading to the highest *SYR*. Indeed, the cells selected first contribute most to the *SYR*. In this case, by afforesting ≈ 5000 cells (≈ 20% of the candidate cells), the maximum attainable *SYR* is nearly reached, as indicated by the horizontal lines in Fig. 13a.

Fig. 14 shows the cells selected by CAMF-FD8 for afforestation of up to 30% of the candidate cells in the Tabacay catchment. Note that the first selected cells (5% and 10%) are concentrated in areas with high sediment production, as already observed for CAMF-SFD (Vanegas et al., 2012; Estrella et al., 2014b). For example, the 11 candidate cells i with $\alpha_i^2 > 250\, ton\, ha^{-1}\, yr^{-1}$ were selected in the first 25 iterations.

The results for the Maarkebeek catchment are presented in Table 8 and Figs. 15a, 15b. We observe the same behaviour as for the Tabacay catchment. In this case, the difference in results obtained with the D∞ method and the SFD method is insignificant. When only 5% of the cells is afforested ≈ 26% of sediment is reduced (30% in Tabacay with FD8), see Fig. 15b. Also in this case the first selected cells are concentrated in areas with high sediment production.

Table 8
SY and *SYR* obtained with CAMF-SFD, CAMF-FD8 and CAMF-D∞ for the Maarkebeek catchment. Afforestation of up to 5% of the candidate cells.

Initial *SY* ($ton\ yr^{-1}$)			# Afforested cells	*SYR* ($ton\ yr^{-1}$)			% *SYR*		
SFD	D∞	FD8		SFD	D∞	FD8	SFD	D∞	FD8
30 756	30 758	30 739	1% = 538	2 468	2 503	2 617	8.03	8.14	8.51
			2% = 1 076	4 150	4 195	4 415	13.49	13.64	14.36
			3% = 1 614	5 568	5 633	5 918	18.10	18.31	19.25
			4% = 2 152	6 809	6 886	7 240	22.14	22.39	23.55
			5% = 2 687	7 889	7 980	8 403	25.65	25.94	27.34

4.2 Scalability and computational cost of CAMF-MFD

We studied the scalability of the CAMF-SFD and CAMF-MFD algorithms by measuring the execution times for the three data-sets (small, intermediate and original data-set) of both catchments. The characteristics of these data-sets are presented in Tables 9 and 10. The sequential execution times are shown in Table 11.

The SY in the initial situation is computed by calculating the SA matrix. In this case, the number of operations is proportional to the number of active cells. From the small to the intermediate and the original data-set, the number of active cells increases by a factor of 4 and 13 (Tabacay), 4 and 18 (Maarkebeek). However, the execution times grow much slower. On current computers, execution times are mainly determined by the cost of data access, from main memory to caches and to the processor. Note that the calculation of SA uses a very irregular data access pattern. Probably caching works more effectively for larger data-sets. Of course, the execution time of CAMF-FD8 is higher than of CAMF-D∞ and much higher than of CAMF-SFD, due to the SI among cells increases leading to a more complex data structure used to model the flow paths.

To compute the SYR resulting from tentatively afforesting each (not selected) candidate cell i in each iteration of CAMF, the SA matrix is recomputed. As mentioned in section 2.4, this requires to recompute the SA_i values for all cells appearing after cell i in the sorted list S. Thus the number of operations in one iteration is $\approx 0.5n$ larger than the time to compute the SA in the initial situation, with n referring to the number of candidate cells. However, the ratio of the execution times for one iteration and for the computation of the SA in the initial situation is much smaller for the small and intermediate data-sets, while for CAMF-MFD with the original data-set the ratio is $\approx n$. This again shows that the number of operations is a poor indicator for the execution time on current computers.

Table 9
Characteristics of the Tabacay data-set. Small and intermediate data-sets correspond to respectively $\frac{1}{16}$ and $\frac{1}{4}$ of the original data-set; Ratio 1: ratio of intermediate to small data-set; Ratio 2: ratio of original to small data-set.

	Small	Intermediate	Ratio 1	Original	Ratio 2
Dimensions	89 x 87	178 x 173		355 x 346	
# cells	7 743	30 794	4	122 830	16
# active cells	5 475	22 494	4	73 471	13
# candidate cells	2 259	9 859	4	27 246	12

Table 10
Characteristics of the Maarkebeek data-set. Small and intermediate data-sets correspond to respectively $\frac{1}{16}$ and $\frac{1}{4}$ of the original data-set; Ratio 1: ratio of intermediate to small data-set; Ratio 2: ratio of original to small data-set.

	Small	Intermediate	Ratio 1	Original	Ratio 2
Dimensions	116 x 110	232 x 219		464 x 438	
# cells	12 760	50 808	4	203 232	16
# active cells	7 363	32 222	4	129 097	18
# candidate cells	693	11 802	17	73 792	78

In the optimal implementation of CAMF-SFD (Estrella, 2015), the flow connectivity of the cells is represented as

Table 11
CPU times for the calculation of the *SA* matrix in the initial situation and for one iteration, using the Tabacay and Maarkebeek data-sets. Small and intermediate data-sets correspond to respectively $\frac{1}{16}$ and $\frac{1}{4}$ of the original data-set (see Tables 9 - 10).

		CPU times (s)					
	FD method	Tabacay data-set			Maarkebeek data-set		
		Small	Intermediate	Original	Small	Intermediate	Original
Initial *SA*	SFD	0.003	0.006	0.009	0.020	0.027	0.023
	MFD/D∞	0.013	0.018	0.018	0.031	0.038	0.059
	MFD/FD8	0.018	0.023	0.028	0.023	0.036	0.063
1 iteration	SFD	0.476	9.66	110.8	2.572	22.11	867.3
	MFD/D∞	0.870	19.64	316.9	2.750	30.58	1 643.9
	MFD/FD8	1.373	26.67	651.2	6.646	60.77	3 258.4

Table 12
Ratios of CPU times for the calculation of the *SA* matrix in the initial situation and for one iteration, using the Tabacay and Maarkebeek data-sets. Ratio 1: ratio of intermediate to small data-set; Ratio 2: ratio of original to small data-set.

		Ratios			
	FD method	Tabacay data-set		Maarkebeek data-set	
		Ratio 1	Ratio 2	Ratio 1	Ratio 2
Initial *SA*	SFD	2.19	3.45	2.46	1.07
	MFD/D∞	2.38	1.38	0.20	2.35
	MFD/FD8	1.28	0.03	0.21	1.06
1 iteration	SFD	20.28	232.72	8.60	337.18
	MFD/D∞	22.41	361.39	11.13	598.15
	MFD/FD8	19.43	474.33	9.14	490.27

a tree where the root corresponds to the outlet of the catchment and child nodes deliver sediment to their parents. Note that, in this case, only the cells on the path between cell i and the outlet (one branch of the tree) must be recomputed, but this optimization is not implemented in our implementation of CAMF.

5 Conclusions

The raster-based CAMF method selects sites where an intervention has maximal impact, e.g., to minimize sediment delivery at the outlet(s) of a river catchment due to afforestation. The original CAMF used a Single Flow Direction method for flow simulation. We described and analyzed the integration in CAMF of two variants of Multiple Flow Direction methods, i.e., FD8 and D∞, in which the sediment leaving a cell flows into more than one down-slope neighbor, providing more accurate flow simulations in some topographies, e.g., to model divergent flow. Since the calculation of sediment accumulation in CAMF only depends on the ordering of the nodes of the graph representing the flow direction network in the raster cells, the flow direction method can be selected depending on the application and the characteristics of the terrain. Furthermore, the software design allows to easily implement other multiple flow direction methods that uses a 3×3 neighborhood. Also MFD methods that use an adaptive approach for computing certain parameters, such as the flow partitioning exponent, based on a local topographic attribute, as in Qin et al.

(2007), can be implemented, since such attributes are available or can easily be computed from the DEM.

The various algorithmic steps are presented in detail, highlighting the higher complexity of the sediment transport simulation in the Multiple Flow Direction variants. This paper reports on the application of CAMF for afforestation planning in two river catchments: the Tabacay catchment in Ecuador and the Maarkebeek catchment in Belgium. The amount of sediment leaving a cell is modeled by a piece-wise linear function, parameterized by retention capacity, saturation threshold and a flow factor.

The results show that CAMF iteratively selects, from a large set of candidate cells and in the presence of spatial interaction, those cells for which the marginal contribution to the sediment yield reduction by afforestation is highest. A major characteristic of the selected cells is that they lie in regions with a high local sediment production. For the retention capacities and saturation thresholds used in this paper, the three variants, i.e., CAMF-SFD, CAMF-FD8 and CAMF-D∞, produce nearly the same sediment yield at the outlet(s) in the initial situation. For a given number of afforested cells, CAMF-FD8 predicts a slightly higher sediment yield reduction than CAMF-D∞ and CAMF-SFD.

The two CAMF-MFD variants have a substantially larger computational cost per iteration than CAMF-SFD, due to the higher cost to update the sediment accumulation matrix after afforesting a single cell. The computational cost of the FD8 variant of CAMF-MFD is much larger than of the variant that uses D∞, due to the fact that the number of cells involved in FD8 is higher than in D∞. Hence there is a clear need to exploit parallelization and to reduce the cost of each iteration and also the number of iterations. This topic is addressed in a forthcoming paper.

6 Acknowledgments

The authors would like to acknowledge the VLIR-UOS projects “Networks 2019 Phase 2 Cuba ICT” (grant for PhD research of G. Castillo Reyes) and “Global Minds KU Leuven” (Short term Research Stay of R. Estrella).

Code availability section

Name of the code/library: CAMF-MFD 1.0

Open source license: GPL 3.0

Developer: Grethell Castillo Reyes

Contact: Department of Computer Science, Celestijnenlaan 200A box 2402, 3001 Leuven, Belgium; Data Representation and Analysis Center, San Antonio de los Baños Km 2½, University of Informatic Sciences, Cuba.

Email: grethell.castilloreyes@kuleuven.be, gcreyes@uci.cu. Tel.: +32496998509

Year First Available: 2022

Hardware requirements: Code was tested on one core of a Xeon E5-2697 v3 CPU (2.6 GHz) with 28 cores and 128 GB of RAM

Program language: C++

Software required: Geographic Data Abstraction Library (GDAL). Source code available at the link: `https://gdal.org/download.html`. For Unix systems it is also available at the repositories with the name libgdal-dev.

Program size: 136 KB

The source codes are available for downloading at the link: `https://gitlab.kuleuven.be/u0123674/camf-mfd`

Appendix A: MFD methods implemented in CAMF

In the FD8 model (Quinn et al., 1991), the proportion of sediment flowing to each down-slope neighbor cell is given by

$$F_{i,j} = \frac{\tan G_{i,j} \times L_{i,j}}{\sum_{m=1}^{n} \tan G_{j,m} \times L_{j,m}} \tag{A.1}$$

with $F_{i,j}$ the fraction of sediment delivered from cell i to its neighbor cell j; n the number of down-slope neighboring cells; $G_{i,j}$ the gradient between cell i and neighbor cell j; $L_{i,j}$ contour length: $0.5 \times$ cell size for a cardinal neighbour and $0.354 \times$ cell size for a diagonal neighbour.

The D∞ method (Tarboton, 1997), allows to minimize the dispersion present with FD8. Using this approach the flow is distributed to maximum two neighbors, representing the flow direction as a single angle in radians counter-clockwise from east. The flow is divided between the two down-slope cells in accordance with the angle of the steepest descent of the eight neighbor cells.

When the angle represents one of the cardinal neighbors: 0; $\pi/2$; π; $3\pi/2$ or one of the diagonal neighbors: $\pi/4$; $3\pi/4$; $5\pi/4$; $7\pi/4$, the flow only occurs in one of these directions. In other cases, the proportion of flow distributed between two neighbors is calculated as indicated in Fig.17.

References

Anand, S.K., Hooshyar, M., Porporato, A., 2020. Linear layout of multiple flow-direction networks for landscape-evolution simulations. Environmental Modelling & Software 133, 104804. URL: https://www.sciencedirect.com/science/article/pii/S1364815220305934, doi:https://doi.org/10.1016/j.envsoft.2020.104804.

B.-M. Vought, L., Pinay, G., Fuglsang, A., Ruffinoni, C., 1995. Structure and function of buffer strips from a water quality perspective in agricultural landscapes. Landscape and Urban Planning 31, 323–331. URL: https://www.sciencedirect.com/science/article/pii/016920469401057F, doi:https://doi.org/10.1016/0169-2046(94)01057-F.

Barnes, R., 2019. Accelerating a fluvial incision and landscape evolution model with parallelism. Geomorphology 330, 28–39. URL: https://www.sciencedirect.com/science/article/pii/S0169555X19300029, doi:https://doi.org/10.1016/j.geomorph.2019.01.002.

Cisneros Espinosa, P.J., Gabriëls, D., Van Meirvenne, M., 1999. Mapping of soil erosion risk zones in a part of the watershed of rio Paute (Ecuador). URL: http://lib.ugent.be/catalog/rug01:000500575.

Conrad, O., 2003. Tool LS Factor. URL: https://saga-gis.sourceforge.io/saga_tool_doc/8.0.1/ta_hydrology_22.html. Accessed: 2022-01-11.

Costa-Cabral, M.C., Burges, S.J., 1994. Digital Elevation Model Networks (DEMON): A model of flow over hillslopes for computation of contributing and dispersal areas. Water Resources Research 30, 1681–1692. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/93WR03512, doi:https://doi.org/10.1029/93WR03512.

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.

Deproost, P., Renders, D., Wauw, J., Ransbeeck, N., Verstraeten, G., 2018. Herkalibratie van WaTEM/SEDEM met het DHMV-II als hoogte- model omgevingvlaanderen.be Eindrapport. Technical Report. KU Leuven. Belgium.

Desmet, P., Govers, G., 1996. A GIS procedure for automatically calculating the USLE LS Factor on topographically complex landscape units. Journal of Soil and Water Conservation 51, 427–433.

Drzewiecki, W., Mularz, S.C., 2008. Simulation of water soil erosion effects on sediment delivery to dobczyce reservoir.

Estrella, R., 2015. Where to afforest? Single and multiple criteria evaluation methods for spatio-temporal decision support, with application to afforestation. Ph.D. thesis. Faculty of Bioscience Engineering, KU Leuven. URL: https://lirias.kuleuven.be/retrieve/345463.

Estrella, R., Cattrysse, D., Van Orshoven, J., 2014a. Comparison of three ideal point-based multi-criteria decision methods for afforestation planning. Forests 5, 3222–3240. URL: https://www.mdpi.com/1999-4907/5/12/3222, doi:10.3390/f5123222.

Estrella, R., Vanegas, P., Cattrysse, D., Van Orshoven, J., 2014b. Trading off accuracy and computational efficiency of an afforestation site location method for minimizing sediment yield in a river catchment, Rückemann, Claus-Peter. International Academy, Research, and Industry Association (IARIA). pp. 94–100. URL: https://lirias.kuleuven.be/retrieve/267799.

Fischer, D.T., Church, R.L., 2003. Clustering and Compactness in Reserve Site Selection: An Extension of the Biodiversity Management Area Selection Model. Forest Science 49, 555–565. URL: https://doi.org/10.1093/forestscience/49.4.555, doi:10.1093/forestscience/49.4.555.

Gabriels, K., Willems, P., Van orshoven, J., 2022. An iterative runoff propagation approach to identify priority locations for land cover change minimizing downstream river flood hazard. Landscape and Urban Planning 218, 104262. URL: https://www.sciencedirect.com/science/article/pii/S0169204621002255, doi:https://doi.org/10.1016/j.landurbplan.2021.104262.

Gallant, J.C., Hutchinson, M.F., 2011. A differential equation for specific catchment area. Water Resources Research 47. URL: https://agupubs.

367 onlinelibrary.wiley.com/doi/abs/10.1029/2009WR008540, doi:https://doi.org/10.1029/2009WR008540.
368 Gersmehl, P., 1970. Spatial interaction. Journal of Geography 69, 522–530. URL: https://doi.org/10.1080/00221347008981861, doi:10.
369 1080/00221347008981861.
370 Han, J., Kamber, M., Pei, J., 2012. Data Preprocessing, in: Han, J., Kamber, M., Pei, J. (Eds.), Data Mining. Third Edition ed.. Morgan Kaufmann,
371 Boston. The Morgan Kaufmann Series in Data Management Systems, pp. 83–124. URL: https://www.sciencedirect.com/science/
372 article/pii/B9780123814791000034, doi:https://doi.org/10.1016/B978-0-12-381479-1.00003-4.
373 Heil, G., Muys, B., Hansen, K., 2007. Environmental Effects of Afforestation in North-Western Europe – From Field Observations to Decision
374 Support. doi:10.1007/1-4020-4568-9.
375 Jiang, L., Tang, G., Liu, X., Song, X., Yang, J., Liu, K., 2013. Parallel contributing area calculation with granularity control on massive grid terrain
376 datasets. Computers & Geosciences 60, 70–80. URL: https://www.sciencedirect.com/science/article/pii/S0098300413001921,
377 doi:https://doi.org/10.1016/j.cageo.2013.07.003.
378 Kahn, A.B., 1962. Topological sorting of large networks. Commun. ACM 5, 558–562. URL: http://doi.acm.org/10.1145/368996.369025,
379 doi:10.1145/368996.369025.
380 Nearing, M., Jetten, V., Baffaut, C., Cerdan, O., Couturier, A., Hernandez, M., Le Bissonnais, Y., Nichols, M., Nunes, J., Renschler, C.,
381 Souchère, V., van Oost, K., 2005. Modeling response of soil erosion and runoff to changes in precipitation and cover. CATENA 61,
382 131–154. URL: https://www.sciencedirect.com/science/article/pii/S0341816205000512, doi:https://doi.org/10.1016/
383 j.catena.2005.03.007. Soil Erosion under Climate Change: Rates, Implications and Feedbacks.
384 O'Callaghan, J.F., Mark, D.M., 1984. The extraction of drainage networks from Digital Elevation Data. Computer Vision, Graphics, and Image
385 Processing 28, 323 – 344. URL: http://www.sciencedirect.com/science/article/pii/S0734189X84800110, doi:https://doi.
386 org/10.1016/S0734-189X(84)80011-0.
387 Pan, F., Peters-Lidard, C.D., Sale, M.J., King, A.W., 2004. A comparison of geographical information systems–based algorithms for computing the
388 topmodel topographic index. Water Resources Research 40. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/
389 2004WR003069, doi:https://doi.org/10.1029/2004WR003069.
390 Piégay, H., Walling, D.E., Landon, N., He, Q., Liébault, F., Petiot, R., 2004. Contemporary changes in sediment yield in an alpine mountain basin due
391 to afforestation (the upper drôme in France). CATENA 55, 183–212. URL: https://www.sciencedirect.com/science/article/pii/
392 S0341816203001188, doi:https://doi.org/10.1016/S0341-8162(03)00118-8. Geomorphic Impacts of Rapid Environmental Change.
393 Qin, C.Z., Zhan, L., 2012. Parallelizing flow-accumulation calculations on graphics processing units—from iterative dem preprocessing algorithm
394 to recursive multiple-flow-direction algorithm. Computers & Geosciences 43, 7–16. URL: https://www.sciencedirect.com/science/
395 article/pii/S0098300412000787, doi:https://doi.org/10.1016/j.cageo.2012.02.022.
396 Qin, C.Z., Zhu, A.X., Pei, T., Li, B., Zhou, C., Yang, L., 2007. An adaptive approach to selecting flow partition exponent for multiple flow direction
397 algorithm. International Journal of Geographical Information Science 21, 443–458. doi:10.1080/13658810601073240.
398 Quinn, P., Beven, K., Chevallier, P., Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modelling using digital
399 terrain models. Hydrological Processes 5, 59–79. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.3360050106,
400 doi:10.1002/hyp.3360050106.
401 Rampi, L.P., Knight, J.F., Lenhart, C.F., 2014. Comparison of flow direction algorithms in the application of the cti for mapping wetlands in
402 minnesota. Wetlands 34, 513–525. URL: https://doi.org/10.1007/s13157-014-0517-2, doi:10.1007/s13157-014-0517-2.
403 Randolph, J., 2012. Environmental Land Use Planning and Management. Island Press. URL: https://books.google.be/books?id=
404 Kf83YgEACAAJ.

Renard, K.G., Foster, G.R., Weesies, G.A., Porter, J.P., 1991. RUSLE: Revised Universal Soil Loss Equation. Journal of Soil and Water Conservation 46, 30–33. URL: https://www.jswconline.org/content/46/1/30.

Seibert, J., McGlynn, B.L., 2007. A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models. Water Resources Research 43. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006WR005128, doi:https://doi.org/10.1029/2006WR005128.

Tarboton, D.G., 1997. A new method for the determination of flow directions and upslope areas in grid Digital Elevation Models. Water Resources Research 33, 309–319. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/96WR03137, doi:https://doi.org/10.1029/96WR03137.

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

Vanegas, P., Cattrysse, D., 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., Van Orshoven, J., 2011. A multiple criteria heuristic solution method for locating near to optimal contiguous and compact sites in raster maps, in: Geocomputation, sustainablity and environmental planning. 1st ed.. Springer. volume 348, pp. 35–56.

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

Vanwalleghem, T., 2017. Soil Erosion and Conservation. John Wiley & Sons, Ltd. pp. 1–10. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118786352.wbieg0381, doi:https://doi.org/10.1002/9781118786352.wbieg0381.

Wang, J., 2017. Economic Geography: Spatial Interaction. John Wiley & Sons, Ltd. pp. 1–4. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118786352.wbieg0641, doi:https://doi.org/10.1002/9781118786352.wbieg0641.

Williams, J.C., Revelle, C.S., 1997. Applying mathematical programming to reserve selection. Environmental Modeling & Assessment 2, 167–175.

Wilson, J., Gallant, J., 2000. Terrain Analysis: Principles and Applications. Earth sciences: Geography, Wiley. URL: https://books.google.be/books?id=1311_4-zvy4C.

Wischmeier, W.H., 1978. Predicting rainfall erosion losses : a guide to conservation planning , 58 p. : ill., maps ; 26 cm. —USDAURL: https://handle.nal.usda.gov/10113/CAT79706928.

Yang, T.H., Chen, Y.C., Chang, Y., Yang, S.C., Ho, J.Y., 2015. Comparison of different grid cell ordering approaches in a simplified inundation model. Water 7, 438–454. doi:10.3390/w7020438.

Zhou, Q., Liu, X., 2002. Error assessment of grid-based flow routing algorithms used in hydrological models. International Journal of Geographical Information Science 16, 819–842. URL: https://doi.org/10.1080/13658810210149425, doi:10.1080/13658810210149425.

List of Figures

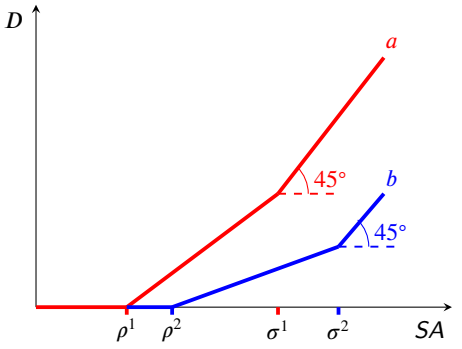


Figure 1: Piece-wise linear convex functions to compute the amount of sediment that leaves a cell. *SA*: sediment accumulation ($ton\,ha^{-1}\,yr^{-1}$); D: amount of sediment leaving the cell ($ton\,ha^{-1}\,yr^{-1}$). Case a (red curve): the cell is not afforested and b (blue curve): the cell is afforested; ρ^1, ρ^2: retention capacity before and after afforestation ($ton\,ha^{-1}\,yr^{-1}$); σ^1, σ^2: saturation threshold before and after afforestation ($ton\,ha^{-1}\,yr^{-1}$).

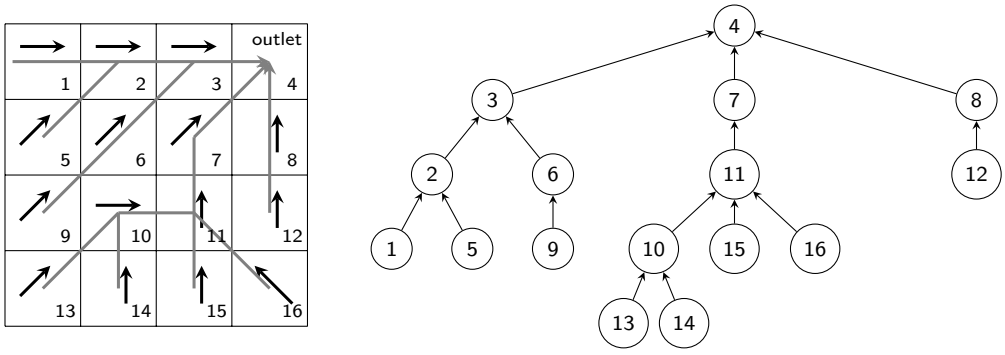
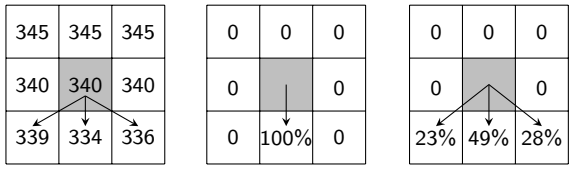


Figure 2: (a): Fragment of a flow direction matrix, where cell 4 represents the outlet, black arrows represent flow direction and gray arrows denote the branches of the corresponding tree representation; (b): Tree representation of matrix in (a).



(a) Elevations (b) SFD (c) MFD

Figure 3: Difference in flow simulation in CAMF-SFD and CAMF-MFD.

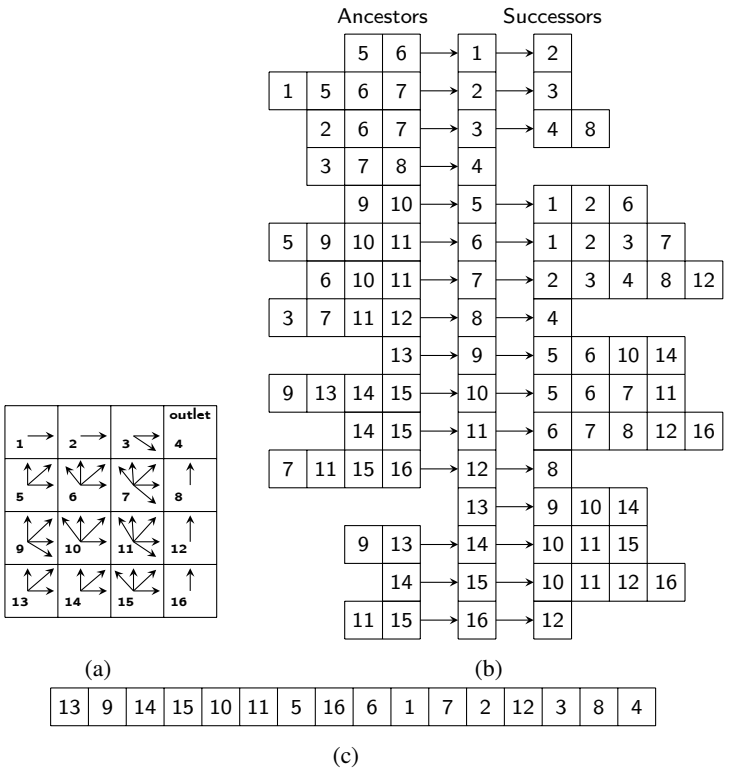


Figure 4: Representation of flow direction in CAMF-MFD. (a) Fragment of a flow direction matrix, where cell 4 represents the outlet; (b) Adjacency list representing the spatial interaction graph; (c) Topologically sorted list of cells.

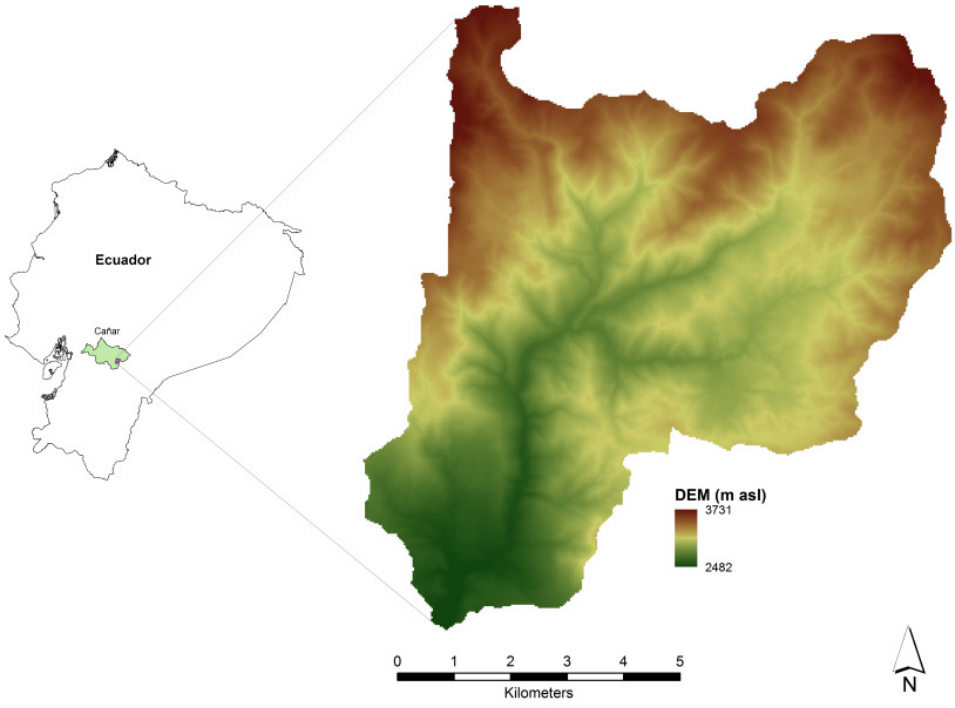


Figure 5: Digital Elevation Model of the Tabacay catchment and its location in Ecuador (Estrella et al., 2014a).

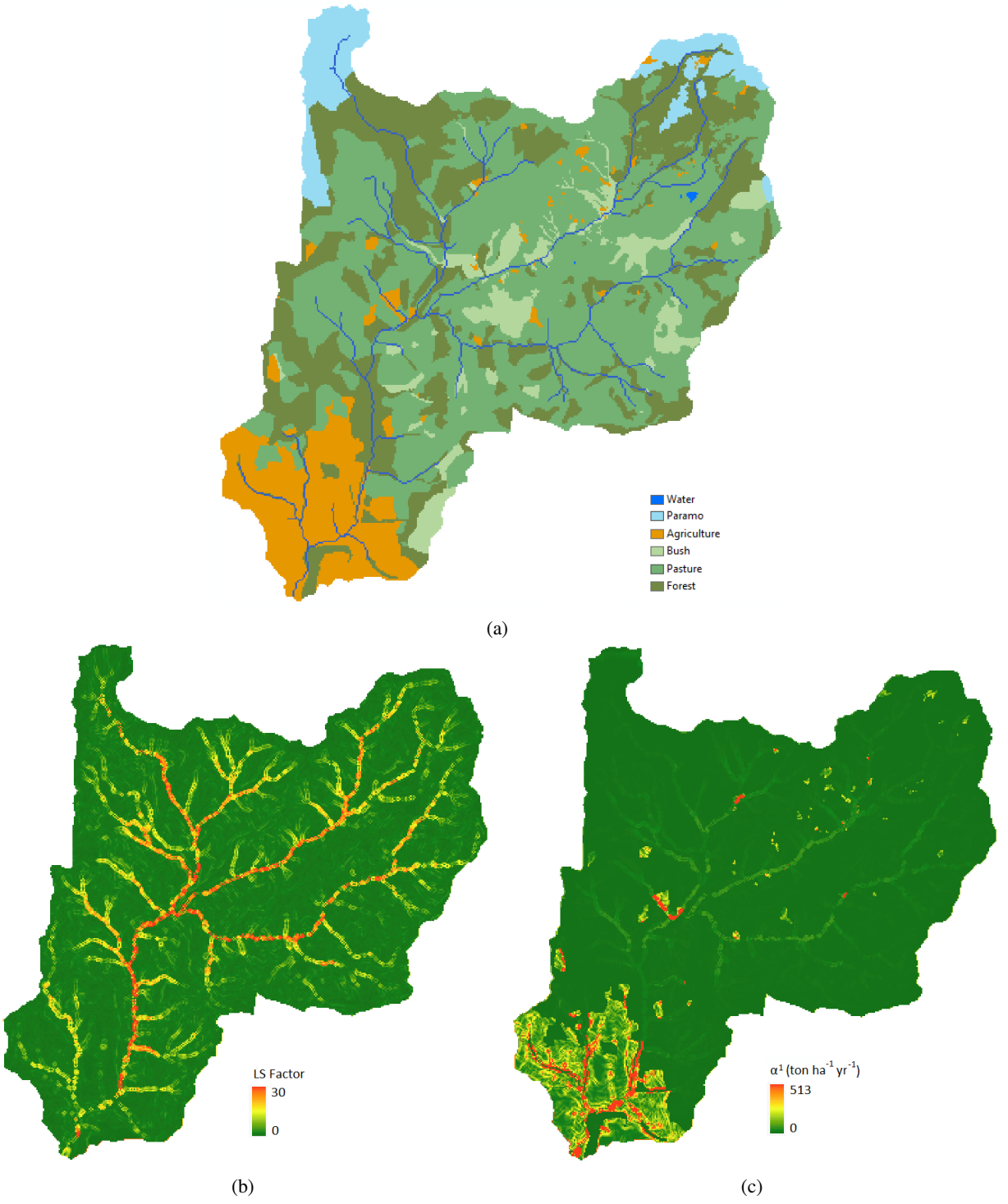
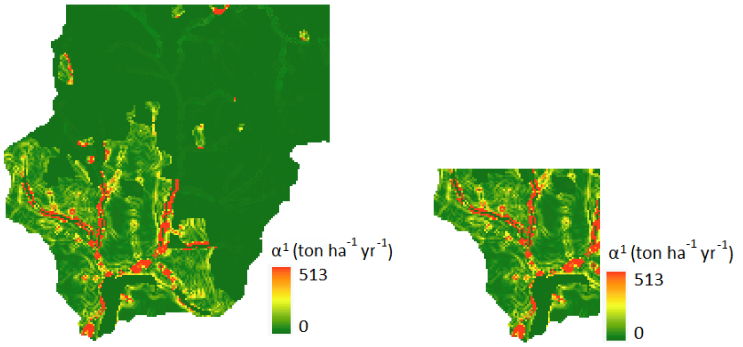


Figure 6: (a): Land cover map of the Tabacay catchment (Estrella, 2015); (b): LS-factor map computed with the ArcGIS tool; (c): Initial local sediment production map α^1 $(ton\, ha^{-1}\, yr^{-1})$ calculated by means of RUSLE.



(a) (b)

Figure 7: Maps cropped from the initial local sediment production map α^1 $(ton\, ha^{-1}\, yr^{-1})$ in the Tabacay catchment. (a): $\frac{1}{4}$ of the original raster data-set; (b): $\frac{1}{16}$ of the original raster data-set.

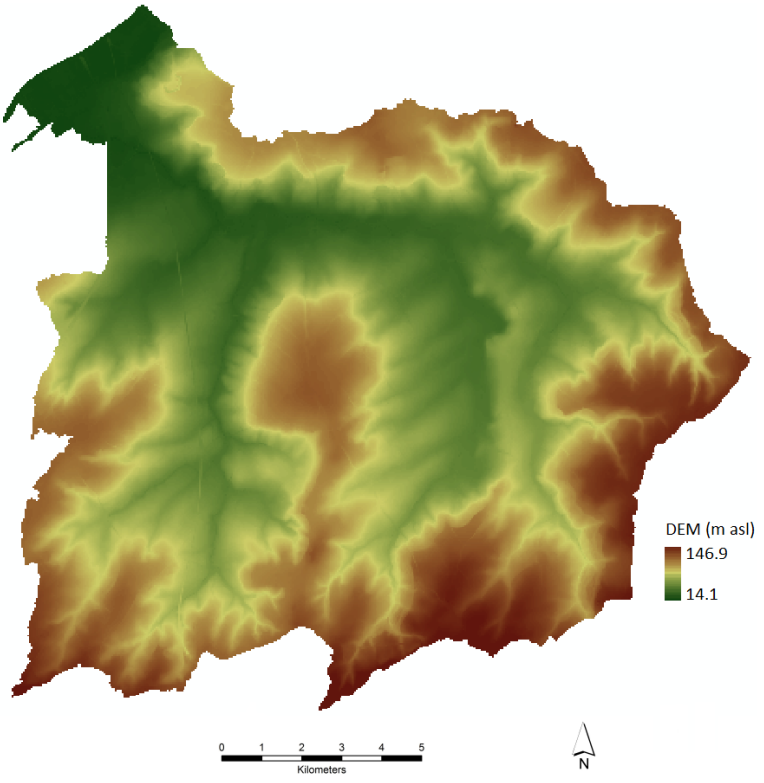


Figure 8: Digital Elevation Model of the Maarkebeek catchment in Belgium (Gabriels et al., 2022).

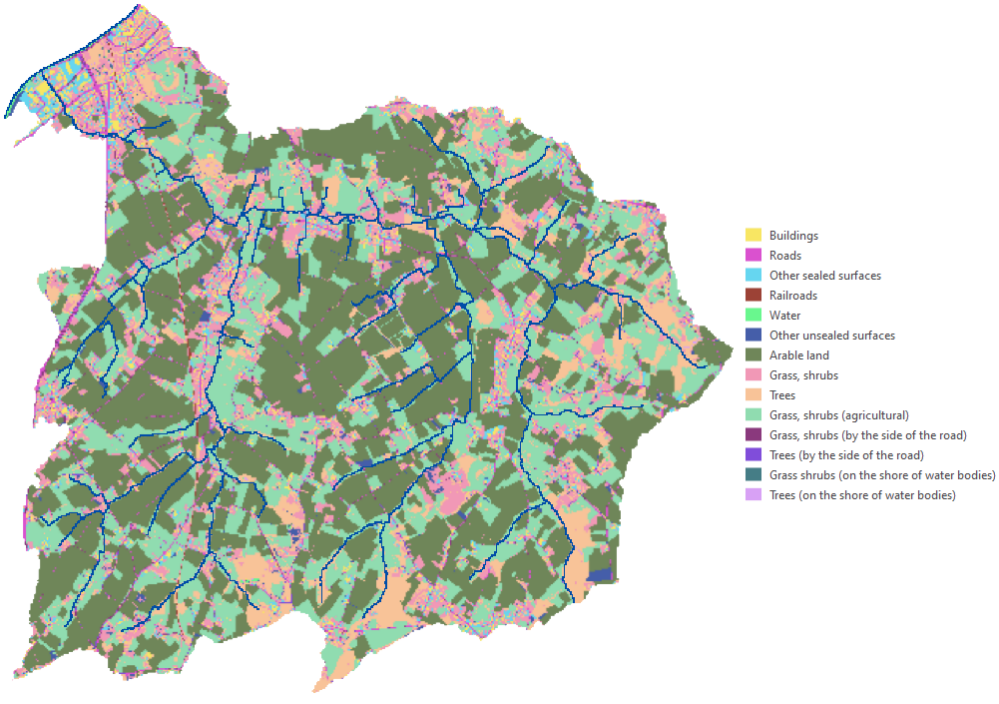


Figure 9: Land cover map of the Maarkebeek catchment (Gabriels et al., 2022).

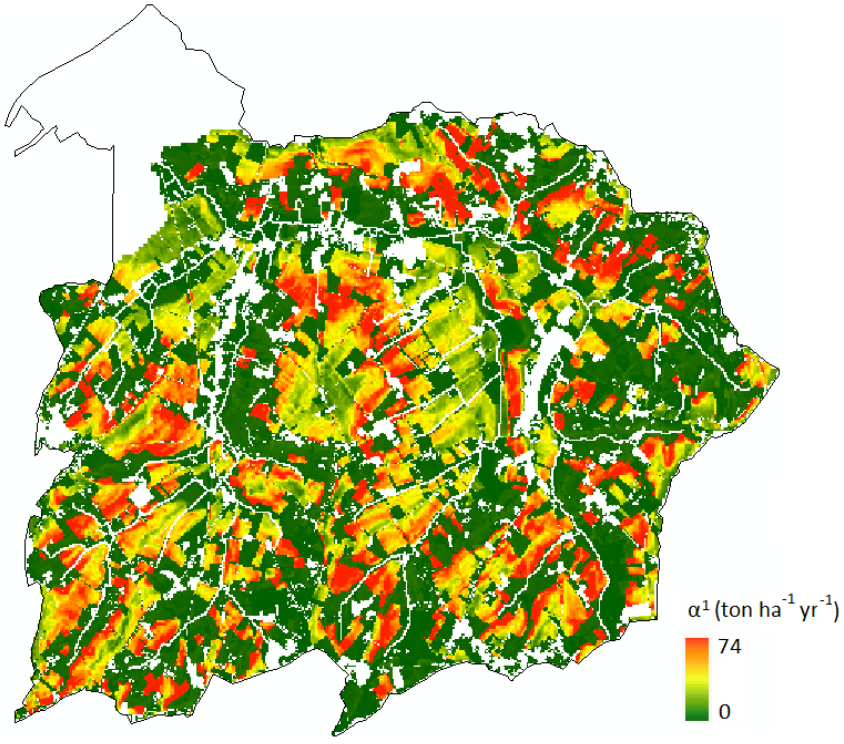
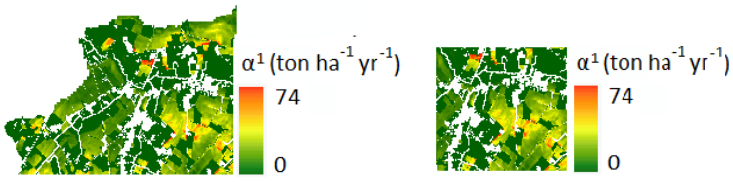


Figure 10: Initial local sediment production map α^1 $(ton\, ha^{-1}\, yr^{-1})$ of the Maarkebeek catchment calculated by means of RUSLE.



(a) (b)

Figure 11: Maps cropped from the initial local sediment production map α^1 $(ton\, ha^{-1}\, yr^{-1})$ in the Maarkebeek catchment. (a): $\frac{1}{4}$ of the original raster data-set; (b): $\frac{1}{16}$ of the original raster data-set.

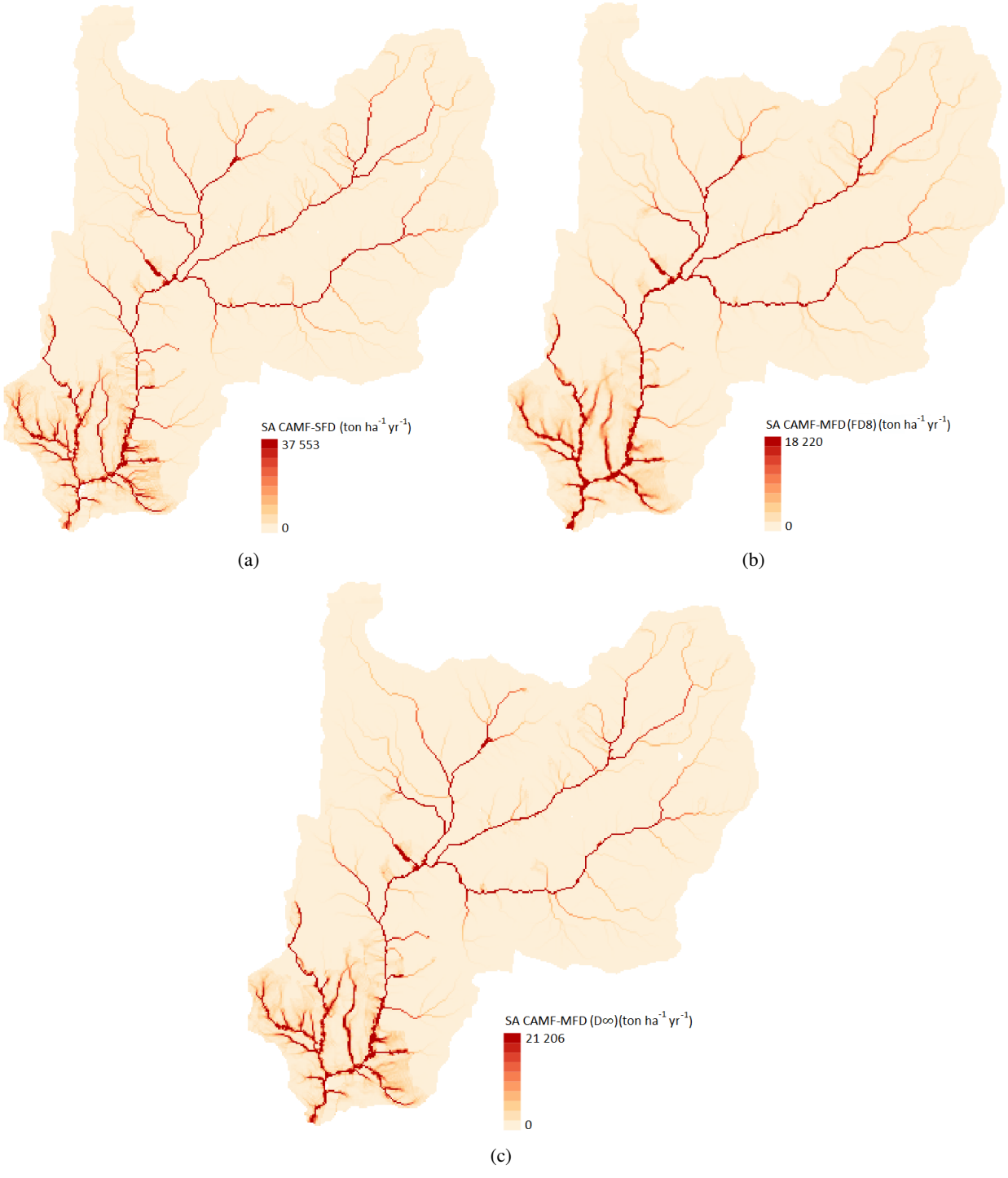
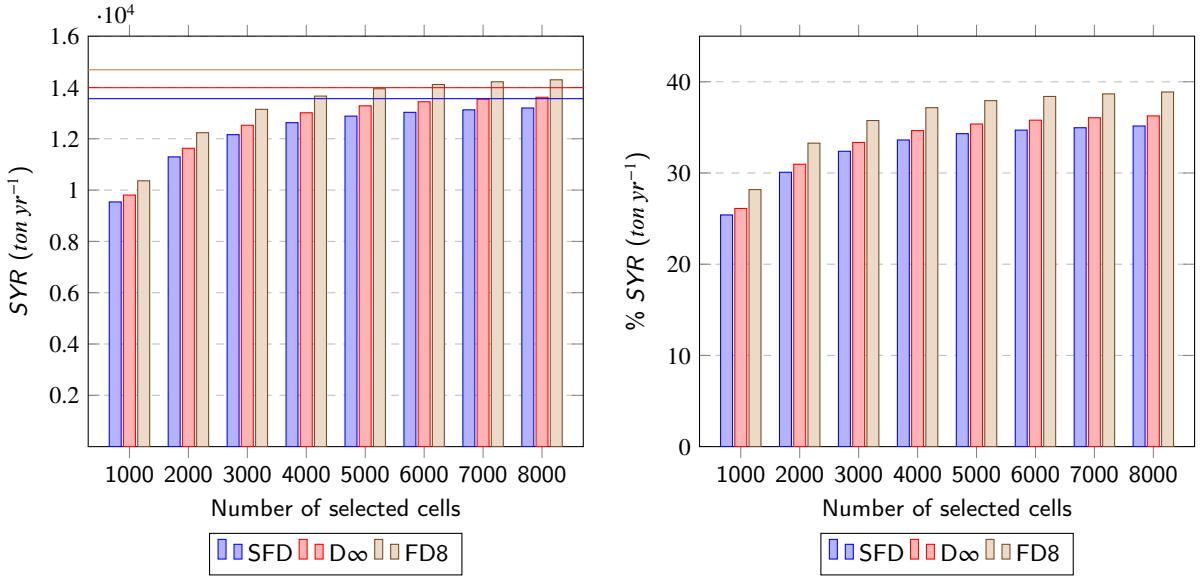


Figure 12: Sediment accumulation (SA) ($ton\,ha^{-1}\,yr^{-1}$) for the initial situation in the Tabacay catchment computed by (a): CAMF-SFD; (b): CAMF-FD8; (c) CAMF-D∞. For CAMF-FD8 and CAMF-D∞ the outlet consists of several (contiguous) cells. In this cases SY is defined as the sum of the SA values in these outlet cells.



(a) *SYR* ($ton\ yr^{-1}$) for Tabacay

(b) % *SYR* ($ton\ yr^{-1}$) for Tabacay

Figure 13: Evolution of *SYR* ($ton\ yr^{-1}$) for the Tabacay catchment using CAMF-SFD, CAMF-FD8 and CAMF-D∞. (a): Absolute *SYR*. Horizontal lines: maximum attainable *SYR* by afforesting all candidate cells; (b): Relative *SYR*.

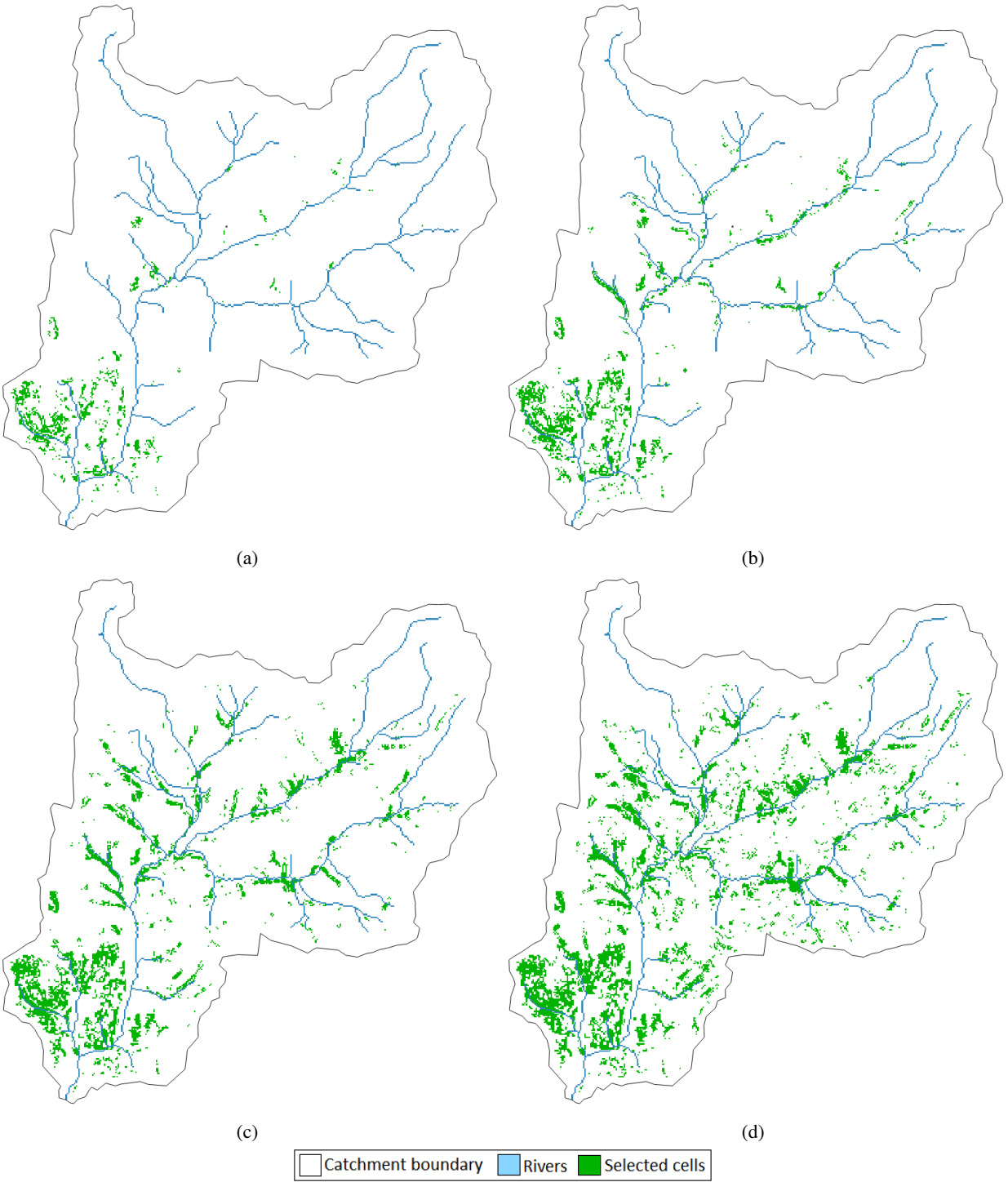
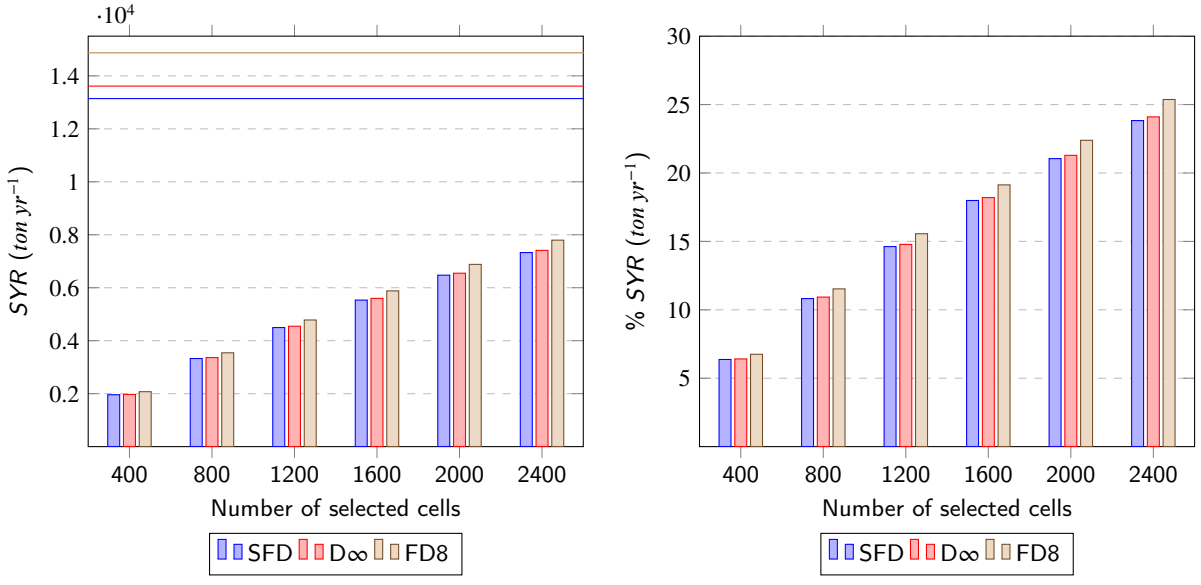


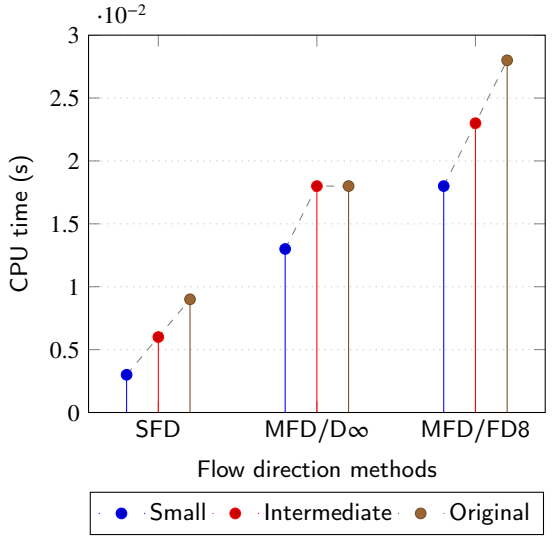
Figure 14: Areas selected by CAMF-FD8 for afforestation in the Tabacay catchment. (a): 5% of candidate cells; (b): 10% of candidate cells; (c): 20% of candidate cells; (d): 30% of candidate cells.



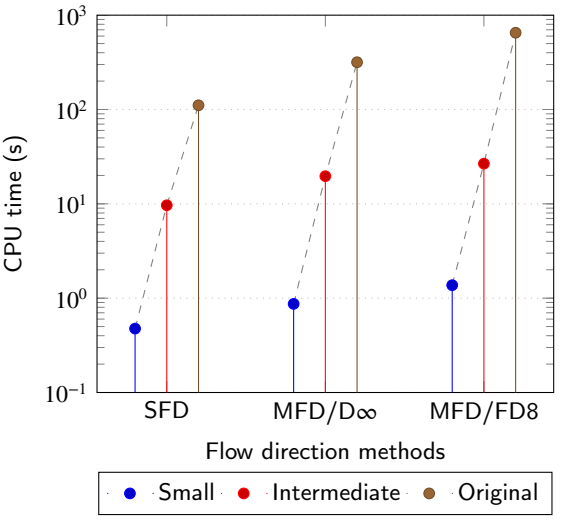
(a) *SYR* (*ton* yr^{-1}) for Maarkebeek

(b) % *SYR* (*ton* yr^{-1}) for Maarkebeek

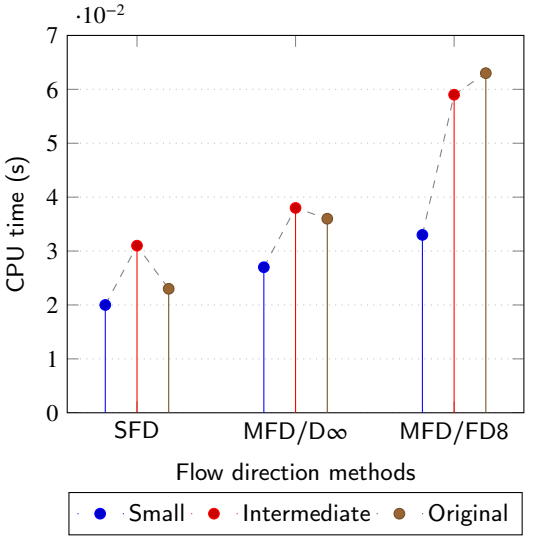
Figure 15: Evolution of *SYR* (*ton* yr^{-1}) for the Maarkebeek catchment using CAMF-SFD, CAMF-FD8 and CAMF-D∞. (a): Absolute *SYR*. Horizontal lines: maximum attainable *SYR* by afforesting all candidate cells; (b): Relative *SYR*.



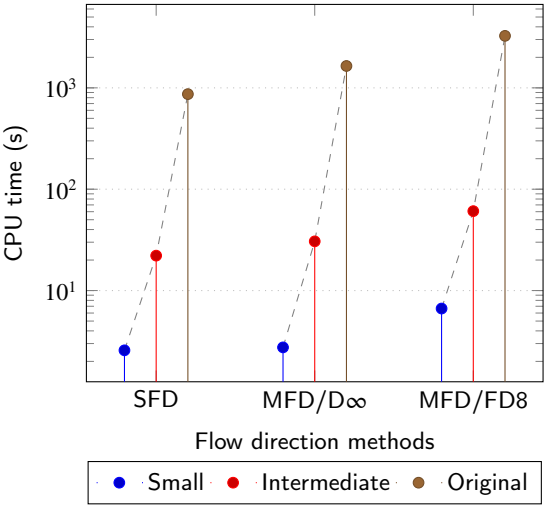
(a) Calculation of initial *SA* - Tabacay data-set



(b) One iteration - Tabacay data-set



(c) Calculation of initial *SA* - Maarkebeek data-set



(d) One iteration - Maarkebeek data-set

Figure 16: Timings for data-sets of increasing size. Small and intermediate data-sets correspond to respectively $\frac{1}{16}$ and $\frac{1}{4}$ of the original data-set (see Tables 11-12). (a)-(c): Timings for calculating the initial *SA* matrix (linear scale); (b)-(d): Timings for executing one iteration of the optimization loop (log scale).

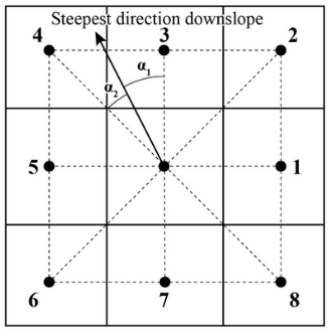


Figure 17: Flow direction computed with the D∞ method, taken from Yang et al. (2015), based on Figure 2 in (Tarboton, 1997). $\alpha_1 + \alpha_2 = 45°$. Proportion flowing to neighbor cell 4 $= \alpha_1/(\alpha_1 + \alpha_2)$; Proportion flowing to neighbor cell 3 $= \alpha_2/(\alpha_1 + \alpha_2)$.