Spatio-Temporal Clustering of Polygon Objects and per Object Interventions

Optimizing Remediation of Spatially Dispersed Contaminated Parcels Under an Annual Budget Constraint

Floris Abrams[1,3], Lieve Sweeck[1], Johan Camps[1]
[1]Belgian Nuclear Research Centre (SCK CEN), Mol, Belgium
e-mail: {Floris.Abrams, Lieve.Sweeck, Johan.Camps}@sckcen.be

Dirk Catrysse[2], Jos Van Orshoven[2]
[2]Katholieke Universiteit Leuven (KU Leuven)
Leuven, Belgium
e-mail: {Dirk.Cattrysse, Jos.Vanorshoven}@kuleuven.be

Abstract — **Polygons provide a natural representation for many types of geospatial objects such as agricultural parcels, buildings, and polluted sites. These polygon-based entities form the smallest units used in decision making of real-world problems. Acting on these dispersed entities could result in a heterogeneous and difficult to perform an action plan. Clustering of parcels in larger homogeneous actionable units can improve feasibility and reduce cost. Therefore, a polygon-based clustering can be beneficial for environmental disaster management, where due to the large impacted area or limited availability of labor and financial resources, setting priorities of where, how and when to act are indispensable. This paper presents a spatio-temporal clustering algorithm under a budget constraint to prioritize clusters of parcels for intervention in space and time. The proposed algorithm returns homogeneous actionable clusters in space and time, trading off between effectiveness and feasibility and cost of intervention.**

Keywords-Spatio-temporal clustering; Budget constraint; Disaster management; Multi-Attribute Decision Making; MADM.

I. INTRODUCTION

When dealing with large natural or man-made disasters decision makers are confronted with setting priorities of where, how and when to remediate, because of a limited availability of labor and financial resources. This priority setting is particularly applicable when the impact of actions is costly and has long lasting influences. For spatially distributed sites with variable characteristics, priority setting among the sites and determination of adequate actions are of major importance. The effectiveness of related decisions is typically conditioned by multiple, and often contradicting criteria of economic, social, technical, environmental and human health-related nature. These characteristics of the decision problem make it suitable for the application of a discrete Multi-Attribute Decision Making (MADM) approach. These MADM approaches typically yield a patchwork of sites with different priorities and preferred actions. Because, performing these actions on multiple adjacent sites at the same time, compared to individual sites, is likely to be more feasible and less costly, making clustering of these sites interesting. Although clustering of pixels is quite common for processing and interpreting raster based datasets (e.g., to determine environmental risk areas, like landslides [1]), it is more complicated in polygon-based data datasets. Nonetheless, polygon objects provide a natural representation of real-world geospatial entities. Therefore, it can be more interesting to provide actionable support to decision making based on polygon-based representations of real-world problems. In the field of afforestation multiple MADM approaches where compared to support decision making, when dealing with raster-based datasets for prioritizing afforestation sites [2], [3]. Furthermore, priorities will be conditioned by the resources available at that period, resulting in the need for spreading the actions in time. However, because of this time aspect the initial decision variables, used for prioritization, change due to physical processes in the landscape or due to the decision context. For an afforestation context, the BIOLP model was developed, to determine how a set of land use types should be distributed over space and time in order to optimize the multi-dimensional land performance of a region [4]. However, the used Integer Programming model based on the Balanced Compromise Programming MADM showed the risk of obtaining solutions that are excessively fragmented.

This paper presents a spatio-temporal approach to deal with the clustering of spatially scattered polygon-based parcels, whereby only a set of actions can be performed per year as constrained by annual budgets. The paper explores an innovative extension of the classic region growing principles, adapted to polygon-based data structures and explicitly takes into account the attributes of the polygons to find the optimal compromise solution for the whole cluster. The algorithm is meant to provide actionable and feasible support to decision makers, by proposing a coherent action plan in space and time for the affected region.

The next section provides in depth explanation of the spatio-temporal cluster algorithm. In Section III, the methodology is illustrated for an agricultural region in Belgium, contaminated after a hypothetical accidental release of radionuclides from a nuclear power plant. Finally, Section IV discusses the algorithms and the case study, then Section V draws the conclusions on applicability of the algorithm to help improve decision-making.

 ISBN: 978-1-61208-983-6

II. Spatio-temporal clustering method

In the following section, first the initial priority scores that are the foundation of the algorithm are explained. Second, the temporal aspects of the parcel interventions and third, the different steps in the algorithm are explained.

A. Distance based multi-attribute decision making

An MADM method considering distance in the feature space, named Compromise Programming (CP), was used to rank the set of feasible alternatives. For the CP methodology, a set of independent, operational and non-redundant attributes need to be established as criteria. The performance criteria used will vary significantly between different cases. For each criterion a weight reflecting the importance of the criterion is set by the stakeholders [5]. For this set of criteria, the CP methodology determines the optimal point, a vector of performance attribute values corresponding to an alternative with the best observed performance on each criterion separately. The ideal point is normally unfeasible, because multi-criteria decision problems involve conflicting criteria. Therefore, CP determines a compromise solution by searching for a feasible solution that is closest to the ideal point. The definition of 'closeness' requires the formulation of a distance metric (1), where a larger distance equals a less optimal alternative [6].

$$L = \left[\sum_{i}^{n} w_i^p \left[\frac{f_i^+ - f_i(x)}{f_i^+ - f_i^-} \right]^p \right]^{\frac{1}{p}} \tag{1}$$

- n is the number of criteria under consideration;
- w_i is the relative importance (weight) assigned to performance attribute i;
- p is a parameter that determines the type of distance function, where 2 represents the Euclidian distance;
- f_i^+ is the optimal value for performance criterion i;
- $f_i(x)$ is the value of the i^{th} performance criterion expressed as a function of the decision variables x;
- f_i^- is the anti-ideal corresponding to the i^{th} attribute that is the "worst" value for this attribute.

Distances based on (1), will be standardized within the interval [0-1], where a distance of 0, represents the optimal alternative where no compromise is needed, because it outperforms all other on all criteria. In contrast, a distance of 1 represents an alternative, with the worst score on all criteria simultaneously.

1) Intervention plan of polygon-based parcels

The first question that needs to be answered is: "Where are the sites for which intervention is most urgent situated?". All sites in need for intervention, are identified as feasible alternatives. The CP methodology returns a distance score for each site, representing the priority/urgency of a parcel to be intervened on. From these scores a ranking of the parcels from high priority (small distance) to low priority (large distance) can be made. For the rest of the paper, the scores for the parcels will be referred to as Parcel Priority Scores (PPS).

2) Optimal intervention per parcel

For each parcel, all the feasible intervention actions for that parcel need to be ranked. Once again the ranking of the interventions is based on a distance score, computed by the CP methodology. For the rest of the paper, the scores representing the priority ranking of a remedial action for this specific parcel referred to as the Action Priority Score (APS).

B. Temporal dynamics in MADM

In many decision areas the decision information is collected and evaluated at multiple periods overtime. However, most MADM methods only focus on the decision making problem at a particular period [7]. Therefore, it is necessary to use Dynamic Multi-Attribute Decision Making (dMADM) to tackle the changes of the alternative attributes and weights through time [8]–[10]. For the case of large scale interventions on many different sites, the decision making becomes dynamic due to limits in economic resources or availability of labour forces. Because of these limited resources not all sites can be tackled in one period of time, resulting in postponed interventions. When actions are postponed the initial decision variables may alter and the decision problem needs a multi-period MADM. The amount of parcels that can be acted on each year depends on the annual budget constraint. While performing actions on the most urgent polygons first each of the actions comes at a cost. For each intervention the cost can be calculated based on the cost per unit of area and the size of the parcel. The interventions can be done until the total cost of remediation would exceed the yearly budget, if the budget is reached the remaining parcels are candidates for action in the next period.

Figure 1. The cluster growing procedure applied on 12 polygon-based parcels, with 3 possible actions. Resulting in 2 homogenous clusters.

C. Spatio-temporal clustering algorithm with budget constraint

To determine spatial coherent clusters of parcels with the same intervention action, a spatially explicit clustering algorithm is used. The algorithm operates in a similar fashion as a region growing algorithm, where it consecutively checks if it could add one of the neighboring parcels to the cluster, depending on the similarity between their PPS and RPS. The clustering algorithm is iterative and consists of three phases: The cluster initialisation, followed by the cluster growing procedure and lastly end of growth phase.

1) Cluster initialisation

To optimally allocate resources, the most urgent sites should be treated first. Therefore, the seed parcel is the one with the lowest PPS (smallest distance to the optimal point) and will be selected as the first parcel in the cluster.

2) Cluster growing procedure

After the seed parcel has been determined, the cluster growing procedure attempts to find neighbouring parcels, which can be added to the seed parcel and later the growing cluster. Parcels in a cluster have the same intervention action and are acted on simultaneously. Adding more parcels to the cluster enlarges the cluster, therefore, creating larger actionable units which are preferred from the perspective of reducing the complexity and operational cost of the intervention. But the addition of candidates with a higher PPS or a different optimal APS to the cluster, results in lower performance of the cluster, compared to the set of individual parcels. It is therefore important to find a compromise remediation action between all parcels on the cluster level. The procedure is shown graphically in Figure 1 and pseudo code in Figure 3.

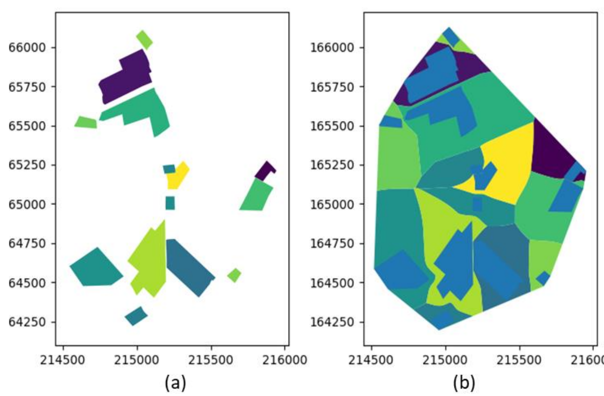

Figure 2. Initial set of distributed parcels (a) and VP computed by the EMT, resulting in a partitioned space (b).

a) Determination of the parcel neighbors

Compared to a raster dataset, where pixels are spatially arranged in a systematic way and neighbours are easy to define, in a vector data set of spatially distributed polygons determining the neighbours is more challenging. To define neighbouring polygons, which are not necessarily sharing a border but rather separated by boundaries such as roads or a small stream, a technique called morphologic tessellation (MT) is used. At the core of MT lies the Voronoi tessellation (VT), a method of geometric partitioning of the 2D space, where a planar set of 'seed points' generate a series of polygons, known as Voronoi polygons (VP). Each VP encloses the portion of the plane that is closer to its seed than to any other polygon [11]. From the partitioned space, the neighbours can be determined by the respected VPs sharing borders. An example of the portioning by VPs is given in Figure 2. Our clustering algorithm makes use of an enclosed tessellation based on the enhanced morphological tessellation algorithm (EMT). EMT allows to set limits to the expansion of the MT, limiting the allowed distance between parcels that can be considered to be neighbours. Further, it allows to set break lines (e.g., larger rivers, administrative boundaries), which the VPs are not permitted to trespass. The VP constructed by the EMT algorithm capture the spatial configuration of all parcels, from which the neighbouring parcels of each parcel can be determined. The EMT algorithm is accessible from an open-source python package (http://docs.momepy.org). Fleischmann (2019, 2020) provide more information regarding the EMT methodology.

Algorithm: Spatio-temporal cluster approach

```
Input: collection of polygon-based parcels, yearly budget (budget), similarity threshold (T)

Create data structure R to store parcels
Add parcels in need for remediation to R

Create data structure S to store remediated parcels
Create data structure RC to store remediation clusters
Create data structure CC to store cluster candidates

Set t to 1
Set BT to budget
Compute PPS for each parcel

While size of R > 0 do

    Select parcel with lowest PPS from R as Seed Parcel (SP)
    Compute APS values for each remedial action
    Select all feasible remediation actions for period t
    Determine the optimal action for seed parcel as action_SP
    Determine neighboring parcels of SP as candidates
    Add candidates to CC

    IF BT - remediation cost action_SP > 0 do
        Set BT to BT - remediation cost action_SP
        Add period to SP
        Add SP to RC
        While CC > 0 do
            Compute composite score (PPS + APS) for all candidate-action combinations
            Select candidate-action_opt with lowest composite score for the whole cluster as CP
            If composite score_cp for action_opt – composite score_SP for action_SP < T do
                If BT - remediation cost CP > 0 do
                    Add period to CP
                    Add CP to RC
                    Determine neighboring parcels of CP as new_candidates
                    Add new_candidates to C
                Else do
                    Set t to t +1
                    Set BT to budget + BT
                    End while
                Endif
            Else do
                End while
            Endif

        End while
        add RC to S
        Remove RC from initial set R

    Else
        Set t to t +1
        Set BT to budget + BT
    Endif
End while

Output: solution set (S)
```

Figure 3. Pseudo code of the spatio-temporal cluster approach, determining the optimal year and remedial clusters.

a) Determining the optimal neighbour

To determine the candidate parcel for growing the cluster, it is necessary to find the parcel and action combination to add to the cluster, with the lowest increase in composite score. The composite score of the cluster is the sum of all PPS scores of the included parcels and their APS scores for the optimal action of the cluster. From this follows, that adding a parcel to the cluster could change the remediation of the whole cluster, resulting in a different compromise solution within the cluster.

This compromise solution is not necessarily the optimal solution for all parcels individually, but from the perspective of the collective composite score of the cluster. When the best parcel is found, it should be checked if it is still similar enough to the seed pixel to be added and that the budget constraint is not exceeded. If for the candidate parcel none of the previous stated thresholds are exceeded, the parcel is added to the cluster and the composite score of the cluster is adapted to the new situation. The growth will continue by the adding the new neighbors and determining the optimal one. This procedure is repeated until the end of growth phase is reached as described in the next paragraph.

3) End of growth

a) Similarity threshold

To determine whether a neighboring parcel can still be added to the cluster, the similarity between the cluster seed and the candidate parcel their composite scores are checked. The difference between both scores cannot exceed the predefined similarity threshold. The threshold is defined by the user, according to its preference to optimality or ease of implementation of the remediation strategy. The reasoning behind the threshold is that when the difference between seed and candidate is large, resources will be used on less urgent parcels or sub optimal interventions will be performed.

b) Budget constraint

The budget constraint limits the amount of resources that can be allocated to intervention in each period. The implementation of a budget constraint in the spatial clustering algorithm, ensures that the cluster cannot exceed the budget for the given period and the clustering is therefore halted once the available budget is reached. Once the budget will be exceeded, the cluster growing is stopped and the polygon's attributes for the new period are determined, then the cluster initialization phase can be started for the new period.

III. CASE STUDY

The results shown in this section are based on a hypothetical accidental radioactive release, affecting 157 polygon-based agricultural parcels in Flanders, Belgium. For a budget of 400.000 euro per year a remediation plan can be designed, that ensures that all parcels are remediated so that after remediation food can be produced in line with the legally set contamination limits.

A. Environmental remediation characterisitcs

A parcel is defined by its on-site characteristics or attributes such as: geographic location, environmental characteristics, agricultural practices. These attributes form the basis for the decision criteria. For the intervention 5 remedial actions are determined, all with a different local and environmental impact and remediation efficiency. The feasibility of the intervention depends on the contamination level and the crop type because some remedial actions are unsuitable for specific agricultural crops or inadequate to reduce the contamination levels below the allowed levels. For example, ploughing actions cannot be performed on parcels with perennial crops. The criteria to assess remedial actions, can vary largely based on the geographical region, contamination type, included stakeholders and data availability.

The reason for including the temporal dynamics in this case study is the altering of certain decision variables in environmental contamination problems. A natural phenomenon, called natural attenuation, causes the mass, toxicity, volume or concentration of contaminants in the soil or groundwater to reduce over time. This implies that the contamination decreases overtime without interference of specific remedial actions. For radioactive contaminations in particular, the reduction of the contaminant is even more strongly determined by radioactive decay, its half-live. For a remedial action to be considered feasible, it should be able to reduce the contamination levels below the legally allowed limits. From the dynamic character of the contamination follows, that after a certain period of time other remedial options can become available, which outperform the previously selected option. Consequently, the remedial actions for each parcel should be revised on a regular basis to ensure they are still optimal for this time period.

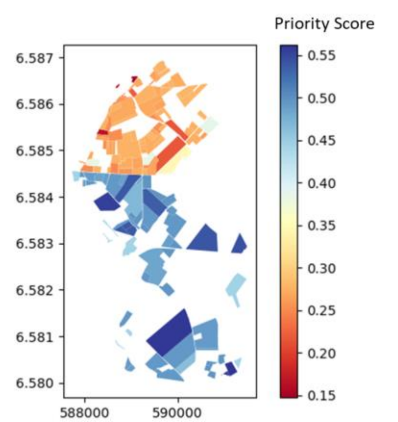

Figure 4. Parcel priority score (PPS) for the affected agricultural parcels, the smaller the more urgent the remediation.

B. Compromise solution on a per parcel basis

The PPS of each parcel is shown in Figure 4, parcels with a low PPS are identified as urgent to remediate. Further, for each specific parcel, the APS for all feasible remediation actions is determined. A specific example for a pasture parcel is shown in figure 5. It is important to acknowledge, that for each parcel, all feasible remedial

Copyright (c) IARIA, 2022. ISBN: 978-1-61208-983-6

actions, possess a RPS score, where a lower score represents a more optimal remedial action for that specific parcel.

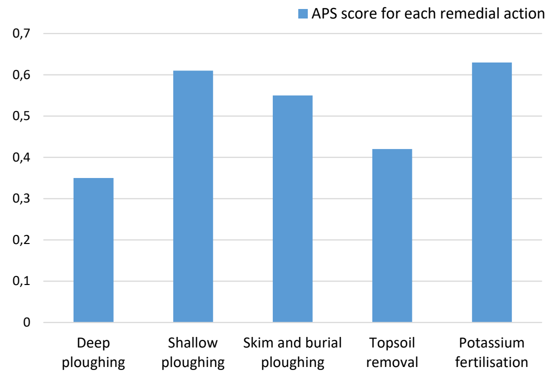

Figure 5. Action Priority Score (APS) for the different candidate remedial actions for an agricultural parcel with cereal cultivation.

C. *Spatio-temporal cluster for the affected region*

A comparison between the action sites per year with similarity threshold = 0 and similarity threshold = 0.25 is shown in Figure 6. A similarity threshold of 0, corresponds to a situation where no candidate parcel will be good enough to add to the cluster, resulting in a remediation plan without clusters. For different values of the similarity threshold a comparison between the remediation plans can be made based on for example: the cost, waste production and time needed for remediation. The comparison between different similarity thresholds is out of the scope of this paper as it relies heavily on characteristics of the contamination and remediation actions.

TABLE 1. THE GROWING PROCEDURE OF A CLUSTER FOR 5 ITERATIONS.

Iterations \ Parcels	A PPS = 0.15	B PPS= 0.17	C PPS= 0.28	D PPS= 0.35	E PPS= 0.41
Iteration I (Seed parcel)	**APS_1: 0.17** APS_2: 0.22 APS_3: 0.33	**APS_1: 0.18** APS_2: 0.23 APS_3: 0.20	APS_1: 0.30 APS_2: 0.24 **APS_3: 0.11**	APS_1: 0.20 **APS_2: 0.11** APS_3: 0.17	**APS_1: 0.15** APS_2: 0.22 APS_3: 0.26
Iteration II (A+B)	**APS_1: 0.35 PPS: 0.32 Composite score$_1$: 0.67**				
Iteration III (A+B+C)	**APS_3: 0.64 PPS: 0.60 Composite score$_3$: 1.24**				
Iteration IV (A+B+C+D)	**APS_2: 0.80 APS: 0.95 Composite score$_2$: 1.75**				
Iteration V (A+B+C+D+E)	**APS_1: 1 PPS: 1.36 Composite score$_1$: 2.36**				

D. *Compromise solutions in the clusters*

To highlight the process of finding a compromise between all parcels on the cluster level, five iterations of the growing procedure are shown in Table 1. The growing procedure determines the optimal remediation action for the 5 parcels based on a similarity threshold of 0.25. The yellow cells show the current parcels in the cluster and the remedial action of the cluster is shown based on the subscript number. The APS values in bold show the optimal action per parcel. From Table 1, it becomes clear that while a cluster grows, the optimal remediation for all parcels within the cluster changes based on the clusters compromise solution. In iteration III, the optimal remediation on the cluster level, is the worst performing action for the seed parcel (A), and the second best action for B. Nevertheless, from the perspective of the cluster action 3 is the best compromise solution.

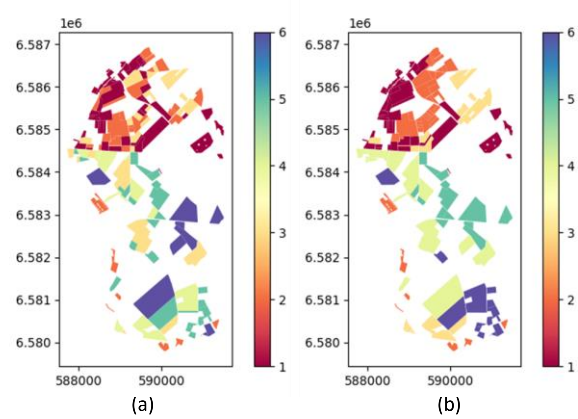

Figure 6. Spatial distribution of the remediated parcels per year after spatio-temporal clustering with a similarity threshold of 0 (a) and a similarity threshold of 0.25 (b). The colors depict the year of remediation per parcel or cluster.

IV. DISCUSSION

Determining the PPS and APS of a parcel, should be done with great care, because it can influence the rest of the process. For the purpose of this research, the CP methodology was used but other distanced based methodologies can replace it. Because of the use of two distance-based metrics, PPS and RPS, the composite distance score still has a physical meaning (distance to the optimal). When other MADM procedures ELECTRE or PROMETHEE are used, this should be done carefully to make sure both scores are still compatible to sum up. Figure 6 shows the clusters that are formed based on the budget constraint and similarity threshold. Reducing the budget, would spread the remediation plan over more years and would interact more with the cluster growing mechanisms, because of more early stops. If the similarity threshold would become bigger, less optimal clusters are allowed and the deviation from the optimal situation of the clusters would grow. The similarity threshold influences the cluster heterogeneity and therefore the compromise solution per cluster becomes more important. Because both values for PPS and APS range between 0 -1, similarity thresholds can range from 0.1 to 0.5. From Table 1, it is clear that more heterogeneous parcels in the cluster result in more changes

 ISBN: 978-1-61208-983-6

in the compromise remedial action throughout the growing of the cluster.

When working with polygon-based datasets, topological errors, such as gaps, overlap and sliver polygons occur. Relying solely on these topological relationships, can have major impacts on determining the neighbors. Our approach is not impacted by these errors. The utility of the algorithm was shown based on an environmental remediation case study, where clusters of remediated parcels would reduce cost and effectivity, compared to non-clustered approaches. Nevertheless, other use cases could benefit from a similar approach. For example, when afforesting a large region, not all sites can be afforested at the same time. Further, every plot has a certain suitability and urgency to be afforested. In addition, afforesting connected parcels, with a similar tree composition, would severely reduce the cost of planting and also improve the ecological connectivity of the landscape. Therefore, finding optimal clusters of areas to afforest with similar tree compositions could be facilitated with our proposed algorithm, for raster datasets this was already done [13].

V. CONCLUSIONS AND FUTURE WORK

With the proposed algorithm, dispersed polygon-based parcels can be clustered in space and time for given intervention under an annual budget constraint. In addition, the utility of the algorithm shows promise for many other fields of application. The extension of the region growing principles from a raster data set to polygons is a useful approach for dealing with real-world problems. Further, explicitly taking into account the attributes of all parcels in the cluster, during the cluster growing procedure gives rise to interesting compromise solutions from a cluster perspective. More research on the impact of the similarity threshold is needed and future work should also attempt at defining the similarity threshold in a way, that can more easily be understood by decision makers.

ACKNOWLEDGMENT

This work was supported by a PhD grant for Floris Abrams from the Belgian Nuclear Research Centre (SCK CEN).

REFERENCES

[1] S. K. Nath, A. Sengupta, and A. Srivastava, “Remote sensing GIS-based landslide susceptibility & risk modeling in Darjeeling–Sikkim Himalaya together with FEM-based slope stability analysis of the terrain,” in *Natural Hazards*, vol. 108, no. 3, Springer Netherlands, 2021, pp. 3271–3304.

[2] R. Estrella, W. Delabastita, A. Wijfels, D. Catrice, and J. Van Orshoven, “Comparison of multicriteria decision making methods for selection of afforestation sites,” *Rev. Int. géomatique*, vol. 24, no. 2, pp. 143–157, 2014.

[3] R. Estrella, W. Delabastita, A. Wijfels, D. Catrice, and J. Van Orshoven, “Comparison of multicriteria decision making methods for selection of afforestation sites,” *Rev. Int. géomatique*, vol. 24, no. 2, pp. 143–157, 2012.

[4] R. Estrella, D. Cattrysse, and J. Van Orshoven, “An integer programming model to determine land use trajectories for optimizing regionally integrated ecosystem services delivery,” *Forests*, vol. 7, no. 2, pp. 1–27, 2016.

[5] V. Belton and T. J. Stewart, *Multiple criteria decision analysis: An integrated approach*. 2002.

[6] J. Malczewski and C. Rinner, *Multicriteria Decision Analysis in Geographic Information Science*, no. Massam 1993. 2015.

[7] Z. Xu, “On multi-period multi-attribute decision making,” *Knowledge-Based Syst.*, vol. 21, no. 2, pp. 164–171, 2008.

[8] Y. Chen and B. Li, “Dynamic multi-attribute decision making model based on triangular intuitionistic fuzzy numbers,” *Sci. Iran.*, vol. 18, no. 2 B, pp. 268–274, 2011.

[9] M. Karatas, “Multiattribute Decision Making Using Multiperiod Probabilistic Weighted Fuzzy Axiomatic Design,” *Syst. Eng.*, vol. 20, no. 4, pp. 318–334, 2017.

[10] Q. Dong and Y. Guo, “Multiperiod multiattribute decision-making method based on trend incentive coefficient,” *Int. Trans. Oper. Res.*, vol. 20, no. 1, pp. 141–152, 2013.

[11] M. Fleischmann, A. Feliciotti, O. Romice, and S. Porta, “Morphological tessellation as a way of partitioning space: Improving consistency in urban morphology at the plot scale,” *Comput. Environ. Urban Syst.*, vol. 80, no. May 2019, p. 101441, 2020.

[12] M. Fleischmann, “momepy: Urban Morphology Measuring Toolkit,” *J. Open Source Softw.*, vol. 4, no. 43, p. 1807, 2019.

[13] P. Vanegas, D. Cattrysse, A. Wijffels, and J. Van Orshoven, “Finding sites meeting compactness and on- and off-site suitability criteria in raster maps,” in *2nd International Conference on Advanced Geographic Information Systems, Applications, and Services, GEOProcessing 2010*, 2010, pp. 15–20.