

A Genetic Algorithm For Fair Land Allocation

Alex Gliesch¹, Marcus Ritt¹ (Advisor)

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul

Abstract. *The goal of agrarian reform projects is the redistribution of farmland from large latifundia to smaller, often family farmers. One of the main problems the Brazilian National Institute of Colonization and Agrarian Reform (INCRA) has to solve is to subdivide a large parcel of land into smaller lots that are balanced with respect to certain attributes. This problem is difficult since it considers several constraints originating from legislation as well as ethical considerations. Current solutions are computer-assisted, but manual, time-consuming and error-prone, leading to rectangular lots of similar areas which are uneven with respect to soil aptitude and access to hydric resources. In this thesis, we propose a genetic algorithm to produce fair land subdivisions automatically. We present a greedy randomized constructive heuristic based on location-allocation to generate initial solutions, as well as mutation and recombination operators that consider characteristics of the problem. Experiments on real-world and artificial instances confirm the effectiveness of the different components of our method, and show that it leads to more fair solutions than those currently applied in practice.*

1. Motivation

The Territorial Organization in Agrarian Reform Projects and Environmental Planning Problem (PROTERRA, acronym for the name in Portuguese: *Problema de Organização Territorial em Projetos de Reforma Agrária e Planejamento Ambiental*) consists of subdividing a large parcel of agrarian land with a single owner into several smaller lots which are to be designated to family farmers, under a set of constraints. PROTERRA is directly associated with the agrarian reform process, which is one of the main factors in the sustainable development of countries. Successful implementations of agricultural settlements have shown to have direct impacts on land use, food production, job creation and eradication of hunger. In Brazil, the technical regulations for the design of agrarian settlements are controlled by the Brazilian National Institute of Colonization and Agrarian Reform (INCRA). However, the current land allocation process is mainly manual, time-consuming and often uneven, usually leading to rectangular lots of similar areas but which fail to consider intrinsic attributes such as soil aptitude or access to hydric resources, and have difficulties handling regions with irregularly-shaped boundaries.

Problem instances are defined as land areas representing farmable soil and which must be allocated to lots. Instances may also define water areas representing hydric resources such as rivers or lakes, and natural preservation areas representing regions of environmental importance and which must not be part of the allocation, so as to avoid undue interventions that may affect the feasibility of settlement. Instance topologies are provided to us by experts as ESRI shapefiles. We adopt a discrete approach and represent these areas as two-dimensional rectangular grids, which translate directly to planar

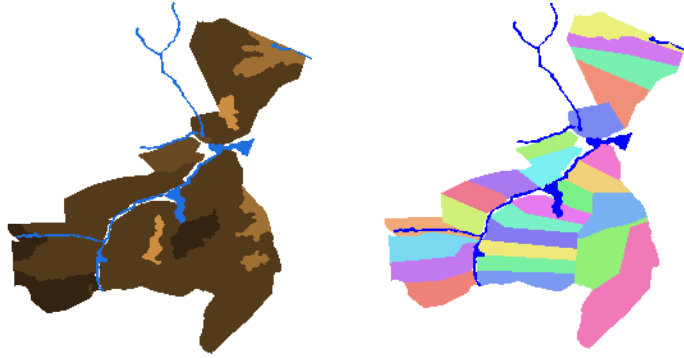


Figura 1. Left: the topology of the real-world instance Veredas, in the Brazilian state of Minas Gerais. White areas represent outer boundaries or nature preservation areas, blue represents water areas, and the different tints of brown represent the different soil aptitudes, darker tints being higher aptitudes. Right: the current manual subdivision for this instance. Each color corresponds to a lot.

graphs. This allows us to use a number of existing techniques useful in related problems, in particular districting. Each discretized cell represents areas of 300m^2 up to 1000m^2 .

Besides its size and topology, an instance is characterized by its soil: the usable land is divided into regions with different classes of aptitude with respect to their potential to grow crops. According to [Ferreira 2015], these aptitude values are obtained by an analysis of physical, agronomic and social factors. Typically areas nearest to water resources have an easier access to irrigation, and have the highest soil aptitude. In the discrete representation, each node u of the input graph is associated with a real value q_u representing the average soil aptitude in that area, using the class-to-value conversion proposed by [Ferreira 2015].

Figure 1 shows the real-world instance Veredas, located in the Brazilian state of Minas Gerais, on the left, and the current manual allocation used by INCRA on the right. One can observe that the manual allocation is uneven in area, appears not to consider soil aptitude at all, and does not always respect river boundaries.

Given an input instance defined by a planar graph $G = (V, E)$ and the desired number of lots k , we define a solution as a partition $V = \bigcup_{i \in [k]} C_i$ of the basic areas into k lots. Lots are required to be contiguous regions of farmable land, i.e. each C_i induces a connected subgraph of G . Water and natural reserves act as obstacles and may not divide a lot. For simplicity, man-made entities such as existing architecture or road network are represented in the instances as natural reserves, since they also serve the purpose of obstacles.

In addition to lot contiguity, a fair allocation must also satisfy geographical, political and fairness constraints. In this work we have considered three main requirements:

1. Accessibility: lots are not allowed to completely enclose one another.
2. Balance: lots with access to a water resource must not be larger in area than lots without direct access to water.
3. Equality: the largest lot may not be more than 3 times larger in area than the smallest lot.

4. Aptitude deviation: the standard deviation in absolute soil quality among lots must be as low as possible.

Other variant criteria, which are not considered in this thesis, include satisfying a minimum lot area while maximizing the number of lots allocated, balancing soil capacity and access to road networks, or generating lots that meet some geometric compactness measure.

2. Proposed method

Our main motive in this thesis is to develop an efficient, automatic method for generating land subdivisions that satisfy the specified requirements. We propose a genetic algorithm (GA) to achieve this. Our algorithm iteratively evolves a population of solutions by generating offspring solutions from a set of individuals, which are selected from the current population based on their fitness. The offspring solutions are obtained through recombination and mutation operators inspired by the natural evolution process. Individuals in the initial population are generated by a randomized greedy heuristic, and each next generation is composed of some of the best individuals from the previous generation, some offspring solutions, and some newly-generated random solutions.

Algorithm 1 A genetic algorithm for PROTERRA

```

1:  $P \leftarrow \{\text{greedyConstructiveHeuristic}() \mid i \in [p]\}$ 
2:  $b \leftarrow$  best solution in  $P$ 
3: repeat
4:    $P' \leftarrow$  best  $e$  solutions in  $P$ 
5:   for  $i \in [o]$  do
6:      $(S_1, S_2) \leftarrow \text{selectParents}(P)$ 
7:      $S_3 \leftarrow \text{crossover}(S_1, S_2)$ 
8:      $S_3 \leftarrow \text{mutate}(S_3)$ 
9:      $P' \leftarrow P' \cup \{S_3\}$ 
10:   $P \leftarrow P' \cup \{\text{greedyConstructiveHeuristic}() \mid i \in [r]\}$ 
11:   $b \leftarrow$  best solution in  $P$ 
12: until stopping criterion satisfied
13: return  $b$ 

```

Algorithm 1 outlines the main method. The population P is initially obtained by generating p solutions using a greedy constructive algorithm (line 1). The main loop of the algorithm (lines 3–12) iteratively updates the current population. The new population P' consists of the best e solutions of the current population (line 4), o solutions obtained by the recombination process (lines 5–9), and r solutions generated using the same randomized greedy process used for the initial population (line 10), where $e + o + r = p$. By adding r randomized solutions at each iteration, we ensure that the population stays varied and avoid that all individuals become related or very similar. The parent solutions of each new offspring solution are uniquely selected by a 3-tournament in P (line 6). The crossover and mutation operators (lines 7 and 8) are stochastic, and consider specifics of the problem. At each iteration we update the best overall solution b and return it at the end (lines 2, 11 and 13). The algorithm ends when a stopping criterion (typically a time limit) is reached.

2.1. Fitness

Our proposed algorithms satisfy the connectivity and accessibility requirements by design, but not the others. We use a fitness function $\varphi = (A, \lambda, \sigma)$ which considers three criteria: balance (A), equality (λ) and aptitude deviation (σ) in lexicographical order. We have further proposed methods to update φ efficiently when solutions are subject to small changes, such as the reassignment of a node. These can be found in more detail in Section 2.2.2 of the attached thesis.

2.2. Constructive heuristic

We use a greedy algorithm based on the location-allocation method [Hess et al. 1965] to generate new solutions. It consists of two phases. In the location phase k seed nodes, each corresponding to a lot, are first selected randomly from the usable land. We then expand these nodes into weighted Voronoi regions, where the weight of a node u is q_u , and reassign the seeds as the geometric centroids of each corresponding region. Next, in the allocation phase we expand these seeds into lots, now accounting for the fitness function. We maintain a list of valid candidate assignments (i.e., which result in a connected and accessible solution) of a node to a lot, and iteratively apply the best b candidates w.r.t. φ , where b is a parameter. We stop once all nodes have been assigned to some lot.

2.3. Recombination

We combine parent solutions by maintaining contiguous regions that are assigned to the same lot in both parents, while reassigning the remaining areas. Since there is no direct correspondence between lots in different solutions, we first establish one by solving a maximum weight perfect matching problem where weights correspond to the sizes of lot intersections. Next, we mark as unassigned all regions not in the intersection of the corresponding lots. Finally, we apply the allocation algorithm of Section 2.2 to re-expand the lots and obtain a complete solution.

2.4. Mutation

The mutation operator first marks as unassigned all nodes within a graph distance of 2 nodes from the borders between lots. Then, the unassigned parts are re-expanded into lots using the allocation algorithm of Section 2.2.

3. Results

We have compared the solutions produced by the proposed GA to the existing manual allocations by the responsible government entity INCRA on 4 real-world instances located in the Brazilian state of Minas Gerais: “Veredas”, “Olhos D’Água”, “Iucatã” and “Belo Vale”. Table 1 shows the results. We have executed the GA for 30 minutes and report averages of 5 replications. Figure 2 visually compares the best solution encountered by the GA and the official allocation.

In all cases, the GA produces solutions that are better with respect to all three components of the objective function (in each component A , λ and σ , lower is better). The manual allocations are highly imbalanced with respect to river access (A) in instances Belo Vale, Iucatã and Olhos D’Água with respect to lot sizes (λ) in instances Iucatã, Olhos D’Água and Veredas, whereas the GA found significantly more fair solutions for all four

Tabela 1. Comparison of our genetic algorithm to manual allocation made by INCRA.

Instance	A (m)		λ		σ	
	Manual	GA	Manual	GA	Manual	GA
Belo Vale	19.7	0.0	2.77	1.83	5,378.5	1,888.6
Iucatã	735.9	0.0	22.38	2.83	18,444.9	2,806.5
Olhos D'Água	830.5	0.0	28.81	1.93	19,972.8	11,716.1
Veredas	0.0	0.0	5.78	1.70	3,964.4	1,161.5

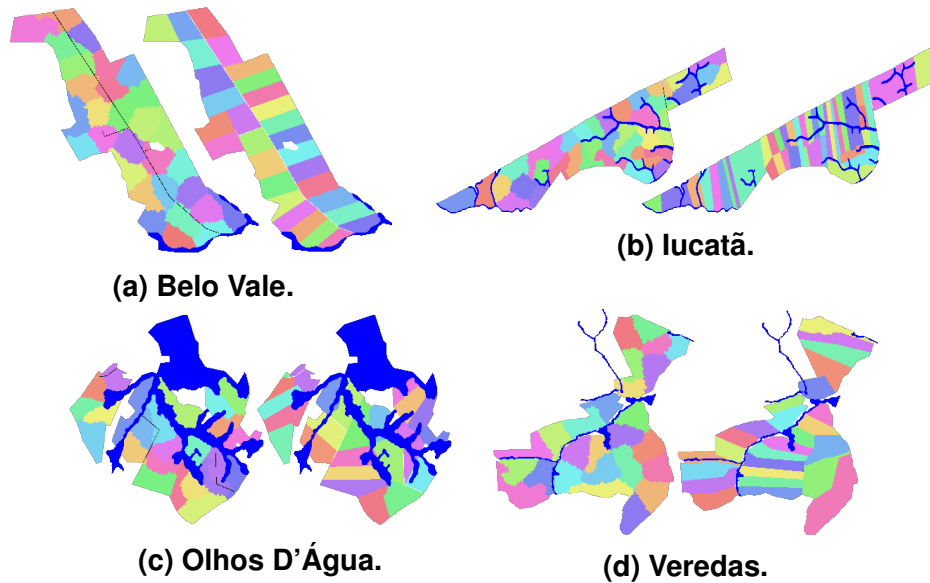


Figura 2. Comparison between the allocation produced by the GA (left) and the manual allocation done by INCRA (right) for instances Belo Vale 2a, Iucatã 2b, Olhos D'Água 2c and Veredas 2d.

instances. The standard deviation in soil aptitude (σ) among lots found by the GA is a factor of 1.7 to 7 times lower, on average.

Additionally, we have conducted extensive experiments to calibrate the different parameters of the algorithm (i.e. the batch size b) and assess the effectiveness of the different components of the method. These experiments can be seen in more detail in Sections 2.6.3–2.6.6 of the attached thesis.

4. Impact

The subject of this thesis was the main topic of the “I Workshop de Problemas de Parcelamento Territorial”, hosted by our research group at the Federal University of Lavras. In this workshop we presented public talks on our published and current research.

Additionally, our research has been awarded the Google Latin America Research Awards in both 2017 and 2018 (grants 25111 and 26568). As a result, we have presented our works to the community in two events hosted by Google in São Paulo (2017) and Belo Horizonte (2018).

4.1. Publications

The main results of this thesis have been published as a full paper at the Genetic and Evolutionary Computation Conference (Qualis A1) [Gliesch et al. 2017] in July 2017.

Moreover, the problem we considered in this thesis can be seen as a districting problem. Districting aims at grouping basic geographical units into groups subject to specific requirements, and has a wide range of applications, in particular electoral and commercial territory design. Since the publication of this thesis, we have adapted some of our methods and proposed a tabu search heuristic for a problem in commercial districting that significantly outperformed existing methods in the literature. This work was a best paper candidate at the European Conference on Evolutionary Computation in Combinatorial Optimisation (Qualis B1) [Gliesch et al. 2018].

Additionally, a technical paper in which we consider a generic approach to a set of districting models has been recently submitted to a top conference in the field. Further, a fourth paper in which a graphical user interface implementing the algorithms developed in this thesis is tested against a relevant user base is being prepared for publication, and is expected to be submitted to a high-quality journal in the upcoming months.

4.2. Current and future work

The research conducted during the development of this thesis is ongoing, and will be the main subject of Alex Gliesch's doctorate. Currently, we address two main lines of work. The first is on fair land allocation, and its main goal is to improve the current problem model, such that an optimal solution could be applied to practice with little or no modification. Further, we aim to adapt current methods to the new requirements and to instances with finer discretizations. The second, more general line of work aims at studying solutions to a number of related problems in districting, with the ultimate objective of developing a general heuristic method that can handle the most common constraints and objective functions found in the literature. More details about our research group and be found in sites.google.com/site/fairlandallocation.

Referências

- Ferreira, F. M. (2015). Aptidão agrícola das terras como função de otimização para o ordenamento territorial e planejamento ambiental: uma análise do SOTER-PA. Master's thesis, Programa de Pós-Graduação em Extensão Rural, Universidade Federal de Viçosa, Brazil.
- Gliesch, A., Ritt, M., and Moreira, M. C. O. (2017). A genetic algorithm for fair land allocation. In *Genetic and Evolutionary Computation Conference - GECCO '17*, pages 793–800, New York, New York, USA. ACM Press.
- Gliesch, A., Ritt, M., and Moreira, M. C. O. (2018). A Multistart Alternating Tabu Search for Commercial Districting. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 10782 LNCS, pages 158–173.
- Hess, S. W., Weaver, J. B., Siegfeldt, H. J., Whelan, J. N., and Zitlau, P. A. (1965). Nonpartisan Political Redistricting by Computer. *Operations Research*, 13(6):998–1006.