Optimizing the consolidation of fragmented small-scale rural holdings

Daniela Ferreira Fernandes

Master's Dissertation

Supervisor: Prof. Pedro Amorim

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia Mecânica

2021-06-28

Abstract

Land fragmentation is a global and timeless problem that compromises the economic viability of land management. In Portugal, it is not uncommon for landowners to have several parcels, sometimes of small sizes, geographically dispersed and with distinct soil occupation. This phenomenon has been linked to rural exodus, fire occurrences and environmental degradation. Addressing this problem is a matter of general interest, however, traditional approaches to land consolidation are time-consuming and costly. Voluntary land-exchange, on the other hand, is a flexible and simpler approach, where the land tenure structure is improved by exchanging parcels amongst the owners. In the present study, this problem and its particularities are described. The main goal is to develop a mathematical model capable of obtaining efficient combinations of land parcel exchanges amongst landowners, reducing distances between parcels belonging to the same holding and increasing contiguous areas.

Different mathematical models were developed and tested using a state-of-the-art solver. The models have in common restrictions at the level of the holdings: of size, must-belong and cannot-belong. For the creation of holdings, both similarity-based and centroid-based criteria were used. A linear model was also proposed. The numerical results evidence that the land-exchange problem is computationally demanding, as the solver failed to achieve optimal solutions for almost all instances and models.

A quadratic minimum sum of pairwise distances and an approximation to the average pairwise distance managed to improve the land tenure structure of a reality-based instance. The quadratic sum is sensitive to the number of parcels and, despite improving the general land tenure, it failed to improve the internal fragmentation of each and every holding. The approximation to the average pairwise distance is more suitable for this purpose, exhibiting a particularity of maintaining the relationship between area and number of plots of each holding, close to the initial.

Both models are similarity-based and present the advantage, over centroid-based models, of being able to include different information, as well as discrete features, in the similarity metric. This characteristic was tested to include contiguity information, which proved effective in promoting the creation of contiguous areas. Furthermore, it improved the algorithm's effectiveness in finding a better solution.

Resumo

A fragmentação da propriedade rústica é um problema global e intemporal que compromete a viabilidade económica da gestão de terrenos rurais. Em Portugal, não é raro os proprietários possuírem várias parcelas, por vezes de pequena dimensão, geograficamente dispersas e com ocupação do solo distinta. Este fenómeno tem sido associado ao êxodo rural, ocorrências de incêndios e degradação ambiental. Abordar este problema é uma questão de interesse público, no entanto, as abordagens tradicionais de emparcelamento rural são morosas e dispendiosas. A troca voluntária de parcelas rurais, por outro lado, é uma abordagem flexível e mais simples, em que a estrutura fundiária é melhorada através da troca de parcelas entre os proprietários. No presente estudo, este problema e respectivas particularidades são descritos. O principal objetivo é desenvolver um modelo matemático capaz de obter combinações eficientes de trocas de parcelas entre proprietários, de forma a reduzir as distâncias entre parcelas pertencentes à mesma propriedade e aumentar as áreas contíguas.

Foram desenvolvidos diferentes modelos matemáticos, e testados recorrendo a um *solver*. Os modelos têm, em comum, restrições ao nível das propriedades, quer da sua dimensão, quer de obrigatoriedade ou impossibilidade de uma determinada parcela pertencer a uma propriedade. Para a formação das propriedades, foram utilizados critérios baseados na semelhança entre parcelas e na distância ao centróide. Um modelo linear foi também proposto. Os resultados numéricos evidenciam que o problema de troca de terrenos é computacionalmente exigente, pois o *solver* foi incapaz de alcançar soluções ótimas para quase todas as instâncias e modelos.

O modelo quadrático da minimização da soma das distâncias par-a-par e uma aproximação da distância média par-a-par conseguiram melhorar a estrutura fundiária de uma instância baseada numa configuração de parcelas real. O modelo quadrático é sensível ao número de parcelas e, apesar de melhorar o panorama geral da estrutura fundiária, não conseguiu melhorar a fragmentação interna de todas as propriedades. A aproximação da distância média par-a-par é mais adequada para este fim, exibindo ainda a particularidade de manter a relação entre a área e número de parcelas de cada propriedade próxima do valor inicial.

Ambos os modelos são baseados na semelhança entre parcelas e, comparativamente a modelos baseados na distância ao centróide, apresentam a vantagem de permitirem a inclusão de variadas informações, bem como características não contínuas, na métrica de semelhança. Este recurso foi testado para incluir informação relativa à contiguidade, o que se revelou eficaz em promover a criação de áreas contíguas e ainda melhorou a eficácia do algoritmo em encontrar uma solução melhor.

Acknowledgements

Firstly, I would like to thank my supervisor Prof. Pedro Amorim and Ricardo Soares, not only for the opportunity to work on this project, but also for the availability and readiness shown during the course of the semester. I would also like to thank the AFBV association for providing the data.

I would like to thank all the friends with whom I have crossed paths. In particular, to Inês, for the longest of friendships and for the endless support. To Ana and Carolina, for these 5 years of companionship, of laughs and shared frustrations, that undoubtedly contributed to making this journey much more memorable.

To Daniel, the purest and most genuine human being I know, for all the tranquility, adventure and perspective you bring into my life.

I would also like to thank my big and loud family, who give real meaning to the word family, in both good and bad times. And to my sister, for all the nonsense and gibberish that never cease to make me laugh. And of course, thank you for the chart, I am forever in your debt. Lastly, I would like to thank my parents to whom I owe everything. Thank you for your patience, support and love, but above all, thank you for everything you taught me, in particular the lessons that cannot be taught by words.

"Remember to look up at the stars and not down at your feet."

Stephen Hawking

Contents

Acronyms and Symbols

01CQP	Zero-One Constrained Quadratic Programming
01UQP	Zero-One Unconstrained Quadratic Programming
APD	Average Pairwise Distance
APDa	Average Pairwise Distance approximation
CAD	Computer Aided Design
FAO	Food and Agriculture Organization of the United Nations
GA	Genetic Algorithm
GIS	Geographic Information System
ILP	Integer Linear Programming
MSD	Mean Squared Deviation
MSSC	Minimum Sum of Squares Clustering
SPD	Sum of Pairwise Distances
SSD	Sum of Squared Distances
WCSS	Within Cluster Sum of Squares
ZIF	Forest Intervention Zone

List of Figures

List of Tables

Chapter 1

Introduction

Land is a finite, non-reproducible consumption resource. In 2020, by 22nd of August, humanity had already depleted earth's one-year capacity to regenerate natural resources, which encompasses demand for food, agricultural and forest products (Lin et al., 2020). With world population still increasing, sustainable land management has been gaining well deserved attention.

One challenge pointed out by the Food and Agriculture Organization of the United Nations (FAO) is land fragmentation. It is defined in the literature as the situation in which a single holding consists of numerous spatially separated parcels which may be small in size and have irregular shapes (Demetriou et al., 2013b).

Land fragmentation is considered to affect primarily the agriculture sector, since parcels are either too small or badly shaped, making it difficult to implement new production standards or to utilize appropriate machinery and technologies (Rembold, 2003) and may require an excessive amount of manual work in the corners and along the boundaries. Additionally, irregular parcel shape prevents proper cultivation of the land, especially for crops which need to be cultivated in rows or series, such is the case of vines and olives. Furthermore, it potentiates neighboring conflicts between landowners (Demetriou et al., 2013b). In short, land fragmentation renders farming economically inefficient, which leads to migration and the abandonment of farmland.

Indisputably, the literature is scarce when it comes to forest land fragmentation. Even though the effects of fragmentation on forest lands are not yet as well understood, small-scale ownership has been explicitly recognized as a major challenge, in terms of cost-effective and cost-competitive management (Hirsch and Schmithüsen, 2010). In turn, high management costs hinder the achievement of sustainable timber production and make the desirable investments in silviculture impossible. It has also been documented that shared ownership generally results in lower felling activity and less qualified silviculture in forestry (Backman and Österberg, 2004).

Land consolidation is a highly effective land management tool that allows for the improvement of the tenure structure of holdings and farms. It increases their economic and social efficiency and could well be the starting point for both food security and sustainable rural livelihoods. Since land consolidation gives mobility to land ownership and other land rights, it may also facilitate the allocation of new areas with specific purposes other than agriculture, such as for public infrastructure or nature protection and restoration, bringing benefits both to right-holders as well as to society in general (Rembold, 2003; Veršinskas et al., 2020).

1.1 Portugal rural depiction

As in most countries, the high fragmentation of rural property in Portugal is a consequence of successional estate inheritances (Bentley, 1990) and it is a recognized problem. The country has one of the highest shares, of 85%, in the number of holdings with size smaller than 5ha recorded in Europe (Demetriou, 2014) and with particular focus in the northern and central regions where the average size of plots is 0,57 hectares. Additionally, it is estimated that more 20% of the territory has no owner or it is unknown (ICNF, 2018).

It is recognized that the existence of a very fragmented land structure in small parcels, together with the rural exodus and the consequent abandonment of land have constituted a strong obstacle to the sustained progress of the rural environment and to socio-economic development. Moreover, the obstacles to economic viability of land management resulting from high fragmentation extend to forestry, where about 92% of forest land is owned by private owners, 6% by local communities and only the remaining 2% by public entities.

In fact, the main land use in the Portuguese territory is forest (30%), and together with shrubland, cover about 70% of the country (ICNF, 2018). The forest has been the basis of a sector of the economy which generates about 100 thousand jobs, that is, about 4% of national employment. The sector also represents around 10% of exports of goods and 2% of Gross Value Added, a value only exceeded in Europe by Finland and Sweden (ENF, 2015). Despite the veracity of these numbers, there are again different realities in the country. In the northern and central rural regions, where small-scale explorations of reduced economic relevance are predominant, the forest dominates the landscape and yet, the socioeconomic indicators do not seem to reveal any resulting benefits from this. On the contrary, the population in the area is increasingly aging, has low income levels and rural abandonment is aggravating. On the other hand, the forest landscape is degraded, misused and unmanaged.

Land fragmentation poses an obstacle to the reinforcement of the competitiveness of the forest sector. Furthermore, it hinders the application and inspection of legally required measures for management and prevention of abiotic risks. An estimated 80% of Portugal's forests are unmanaged. Inadequate silviculture practices predominantly in eucalyptus and pine, the aggressive natural revegetation of abandoned agricultural plots, recently burned areas and forest clearings, and a lack of economic stimulus for promoting opportunities for larger scale biomass removal of understory, noncommercial vegetation and harvest residue exacerbate the growing fuel load problem, which is pointed out as one of the factors that most contribute to the increase in the number of wildfires and burnt area (Beighley and Hyde, 2018; Fernandes et al., 2014; Sequeira et al., 2019). Unsurprisingly, Portugal has one of the highest fire risk rankings in the European continent and is the southern European country most affected by fires (Jesus et al., 2019). These factors, all resulting from lack of management, also make forest more prone to insect and disease problems. The action of fires and harmful biotic agents accelerates ecological degradation and reduces the economic value of forest ecosystems, translating, namely, in the acceleration of soil erosion processes, changes in the water regime and in the reduction of biodiversity (ENF, 2015).

From the several strategies that have been tested, over the years, to overcome the deficiencies of fragmented landownership, the most promising is the Forest Intervention Zone (ZIF), a legal mechanism that facilitates organizing the many small property owners to take collective action. However, while ZIFs have had positive results in landowner identification and participation, and cover a lot of forest area, little improvement to the overall forest situation is being made (Beighley and Hyde, 2018).

1.2 Motivation and objectives

Due to the situation in which the Portuguese rural landscape can be found, namely forest lands, it is believed that the application of land consolidation initiatives would be beneficial.

In the classical form of land consolidation, there is a complete re-structuring of the area. The current plot structure is discarded, a new one is created and the complete infrastructure of the area is often re-planned as well. Such a complex process naturally has the disadvantages of high costs, long implementation times, rigidity and questionable fairness in distribution of land.

Voluntary group consolidation is particularly appealing, due to its flexible nature. Direct state intervention is not mandatory. However, as central government agencies are not involved, additional lots from land banks might not be available and in this case, the number of parcels is not altered. The option of maintaining the current cadastral plots, in spite of the drawbacks to the effectiveness of the land consolidation project since the shape of the plots is an important variable for the cost-efficiency of land management, simplifies and eases the process of land consolidation, given the bureaucratic requirements to alter plot boundaries. Another potential benefit from this approach, especially when considering the economic reality of Portuguese landowners, is the fact that there is no need for additional financial resources from them. There is although, the subsequent necessity to ensure a fair distribution of the land parcels according to the value of the initial holdings, so that no owners become impaired by the exchange.

Considering the above, the problem can be summarized as a reassignment of the parcels among the owners in such a way that fragmentation is minimized, that is, parcels belonging to the same holding are placed as close as possible, while complying with initial values of landowner holdings. The number of possible reassignments is typically very large. For m owners with n parcels, it is m^n. So, even for moderate sizes of 7 owners and 150 parcels the number of possibilities exceeds 10^{126}. Thus, the task of finding an optimal reassignment that simultaneously respects the sizes of the holdings, by trial and error, is hard and time-consuming. Hence, the application of optimization techniques in order to provide results for supporting group voluntary land consolidation processes, gives rise to an optimization problem, of combinatorial and spatial nature, which hereinafter will be referred to as land-exchange.

In this dissertation, the problem of land-exchange is addressed, keeping the focus on the Portuguese reality, namely forest land. The main objective is to model the land-exchange problem, so that it can provide usable results as a voluntary land consolidation tool. In practice, this involves recognizing and gathering criteria and factors important to the problem, and effectively integrate them into the model, as well as obtaining good solutions in reasonable computational times.

1.3 Methodology

In an initial phase, it was sought to understand the current state of the art on land consolidation optimization tools, from the approach and the adopted criteria to the computational results of the algorithm utilized. The problem of land fragmentation was also studied in more detail, seeking to reach a deeper understanding of the criteria that might mitigate the effects of fragmentation on rural property and, particularly, on forests. Other themes were studied, although in less detail, such as spatial optimization, given that the land exchange problem has spatial aspects that must be incorporated into the model, and cluster analysis, since it is a science dedicated to grouping data according to the similarities between them and presents interesting concepts for land-exchange.

Then, the set of criteria that best suits the specificities of the intended exchange model was chosen, and the assumptions were listed, establishing the starting point for the project development. The model was formulated, first, as a clustering problem for which three clustering criteria were proposed. Afterwards, the corresponding formal mathematical models were formulated. In theory, all three criteria are rather similar and were used to intercompare results and performances.

The formulations were implemented in a state-of-art commercial solver and submitted to a set of tests in regular shaped instances. Some necessary modifications were performed in this phase, and the results of the same gathered. The mathematical model results for this set of instances were compared among themselves and with a benchmark.

After validated, the most effective models were chosen and applied in a real parcel layout structure. The results were analysed and final conclusions were drawn.

1.4 Dissertation structure

This dissertation consists of 5 chapters and is organized as follows. The first and current chapter provides an introduction to the fundamental concepts for the scope of this project, in which the expected outcomes and the link between said concepts and the country's situation are included. In light of the above, it proceeds to describe the purpose of the work developed as well as presenting the main objectives and the adopted methodology.

Chapter 2 presents the theoretical concepts of greatest relevance that served as the basis for the preparation of the dissertation, as well as a literature review on land consolidation optimization which provided for decision making throughout the work.

Subsequently, Chapter 3 presents a formal definition of the land-exchange problem by proposing a mathematical programming model. For that purpose, the considered criteria and assumptions for land-exchanges were stated and justified.

Chapter 4 includes a set of computational experiments to explore the features and evaluate the performance of the proposed models. It includes, as well, the results obtained for a real parcel layout.

Finally, Chapter 5 presents the main conclusions of the work developed and possible perspectives for future work, related to the present project.

Chapter 2

Literature review

The land-exchange problem can be seen from several angles. If, on one hand, it is a tool for reducing land fragmentation, and it is therefore important to clarify certain concepts underlying land consolidation, on the other, it is also an optimization problem. Within this last class, since it addresses spatial issues, it can be studied in the scope of spatial optimization. Due to the main objective of the land-exchange problem, of creating compact holdings, it is also possible to study and extract concepts from cluster analysis. Thus, the present chapter, which seeks to introduces theoretical concepts relevant to land-exchange, begins to address land consolidation and fragmentation in more detail. Afterwards, spatial optimization and some of its generalities are tackled, followed by cluster analysis. Finally, the state of the art on land consolidation optimization models and other specific problems that may somehow resemble land-exchange are analyzed in detail.

2.1 Land consolidation

Land consolidation, or sometimes called re-allotment, in a very generalized way, consists of distributing the property to different landowners according to the initial contribution of each one. Land consolidation projects in general aim to improve the land tenure structure by addressing fragmentation, but its purposes can be several, such as village development, nature preservation, outdoor recreation and infrastructure. Nonetheless, rural development is the main driver of this sort of initiative, as it is seen as an instrument for improving production and working conditions in agriculture, forest area and rearrangement of agricultural land in rural areas.

Land consolidation has a close connection with land management and its four main purposes: land tenure, value, use and development. Mentioned for its potential contribution to economic and social reforms (FAO, 2003), many countries have already legislation on land consolidation. According to FAO (2003), there are different approaches to land consolidation. The most complete is comprehensive land consolidation (Akkus et al., 2012; Cay and Iscan, 2011) where, in addition to re-allocating land parcels, rural development actions are taken such as village renewal and erosion control measures including building of natural reserves, construction of rural roads, drainage, environmental protection, social infrastructure and public facilities.

In simplified land consolidation land parcels are re-allocated together with provision of extra land from land banks (Gniadek et al., 2013; Harasimowicz et al., 2017). These projects are often combined with the rehabilitation of infrastructure and sometimes the provision of minor facilities. They do not include the construction of major public works, but they can provide the framework for their construction at a later stage.

The third approach is voluntary land consolidation which involves negotiation within small groups of voluntary landowners (Teijeiro et al., 2020). These projects are based on mutual agreement and are not compulsory. In some countries, voluntary consolidation is the common practice, such is the case of Denmark and the Netherlands where these projects have reached 100 participants.

Individual land consolidation, the last approach, takes place on an informal and sporadic basis. Despite the intervention of state not being expected neither the provision of public facilities, FAO encourages countries to promote consolidation instruments such as joint land use agreements, leasing and retirement schemes.

In the classical form of land consolidation, first and second approaches, there is a complete restructuring of the area, and the current plot structure is discarded and a new one is created. At the same time, in comprehensive land consolidation, the complete infrastructure of the area is replanned. This legal change of ownership is very costly, time-consuming, rigid and its fairness is always questionable. In fact, Borgwardt et al. (2011) states that, in this type of approach, the cost per hectare is about 2000 €, the duration of the procedure often comes close to a decade, and normally each property-owner is forced to participate.

Voluntary group exchange, on the other hand, in spite of usually presenting less effective results, is a more flexible, cheaper and a voluntary alternative. These projects involve negotiations with landowners and require their approval. Consequently, voluntary projects tend to be small, and best suited to address small and localized problems. In some countries, voluntary projects usually have fewer than ten participants. However, in Denmark, almost all land consolidation projects are carried out in a completely voluntary process and are typically based on negotiations with about 50 land owners, although some projects have involved about 100 participants.

To employ land fragmentation reducing measures, it is necessary to understand it first. Although many investigations have been made to comprehend the outcomes of fragmentation, surprisingly, there is no universal definition of fragmentation. In fact, the term land fragmentation has been used for very different problems. Van Dijk (2004) drew attention to this problem and discerned four different types of fragmentation. First, ownership fragmentation, a situation in which there is a high number of landowners, resulting in very small holdings. A high number of land users, the second referred type, results in very small land management units or parcels. Internal fragmentation consists in a large number of parcels exploited per owner. A fourth and last type, is the discrepancy between ownership and use, which when present in a large extent, hinders land use. Generally, all types can be found, in different measures, in a fragmented landscape. In the scheme in Figure 2.1, the first three mentioned types of fragmentation are visible, given that, in a small area with a high concentration of landowners and plots, there are fragmented holdings, some of which are small, and whose generally small and irregularly shaped parcels are dispersed.

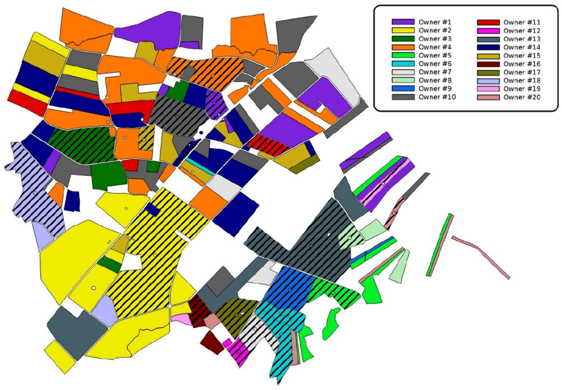


Figure 2.1: Schematic representation of real a case of fragmentation. Source: Teijeiro et al. (2020)

Just as there is no universal definition, there is no standard measure or metric for assessing and comparing land fragmentation. Most commonly used parameters are the average holding size, the number of parcels in each holding, the size of each parcel, the shape of each parcel, the spatial distribution of parcels and the size distribution of parcels. Most authors who tried to measure fragmentation have either used one of the above parameters or integrated more than one, which also rises the problem of defining the weights of each used parameter (Demetriou, 2014). Additionally, very few have took ownership type factors in consideration, with the exception of Demetriou et al. (2013b), who developed a methodology which encompasses, beside dispersion, size and shape of parcels, dual and shared ownership, and also accessibility of parcels.

Western land consolidation experts who tried to demonstrate the importance of land consolidation have mostly focused on the second and third mentioned types of fragmentation. Several models, seeking to understand the effects of fragmentation on agriculture, proved that decreasing the distance of parcels to the farm and the distance between parcels saves time, a better parcel shape raises yields and increased parcel size both saves time and raises yields (Latruffe and Piet, 2014; Lu et al., 2018).

The forest-related literature is more ambiguous in its conclusions. Given the long cycle of timber management, few studies have undertaken an input-output analysis with cross-sectional data or data obtained from monitoring over several years. In an attempt to study the effects of small forestland in forestry, Zhu et al. (2020) concluded that fragmentation results in lower volumes and profits from timber harvests. Furthermore, the additional costs for fuel and labor resulting from unproductive travel time between scattered lots can extend to forestry.

Evaluating the effects from an environmental point of view, in Portugal, fragmentation contributes to the occurrence of large fires. The existence of a landscape mosaic with different land uses and covers disrupts the fire spread, thus helping to prevent the occurrence of large fires (Sil et al., 2019). However, extreme fragmentation has an opposite effect of the heterogeinization of the landscape. Instead it contributes to homogenization, both directly, as the plots are too small, and indirectly, as the economic inefficiency resulting from fragmentation leads to inadequate forestry practices, such as the plantation in large scale of highly flammable species, for example the *Eucalyptus*, due to the perception that there is a lower economic risk associated to this species production due to its high productivity, short harvesting cycles and the guarantee for wood destination to the national pulp and paper industry (Alegria et al., 2019).

2.2 Spatial optimization

Spatial optimization is the use of mathematical and computational methods to identify the best solution(s) to geographic decision problems, which usually involves maximizing or minimizing one or more objectives under strictly defined conditions. Spatial optimization techniques have been applied in various geographic fields of study including land-use planning, natural resource management, political geography, school districting, medical geography, location-allocation selection, routing, retailing, and urban design.

The objective(s) and constraints of a generic model are generally specified using explicit mathematical notation as in Equation 2.1 and 2.2.

$$\text{Minimize} \quad g(x) \tag{2.1}$$

$$\text{Subject to} \quad f_i(x) \geq b_i \quad \forall i \tag{2.2}$$

What distinguishes spatial optimization from other forms of optimization is the explicit representation of space in the decision variables, coefficients, functions and/or constraining conditions used to solve the problem. As geography or space becomes part of the model structure by design, the resulting optimization problem is unique in the sense that geographically based variables, coefficients, functions and constraining conditions have spatially interdependent relationships and properties that are often challenging to abstract and model.

A common approach is to abstract geographical units as points, such as centroids for census tracts, blocks, cities, traffic analysis zones, etc. Compared to complex spatial objects, a point representation, or node in graph theory, makes data handing and problem solving easier, given that evaluation of spatial relationships based on points is more straightforward. On the other hand, with point simplification, loss of geographical detail introduces uncertainties and even errors, such is the case with proximity when areal units are replaced by points (Church, 1999).

The most common of spatial relationships and properties are distance, adjacency, connectivity, contiguity, containment, intersection, shape, districts and pattern. The first four will be addressed with further detail as they are the most relevant to the scope of the project.

2.2.1 Distance

Perhaps the most typical concern in geographic problems is proximity between places, either in terms of physical distance or travel time. Distance is the displacement from a geographic location to another location, and can be measured in many ways, depending on the intended study purpose. In a two-dimensional space, Euclidean, Manhattan, or network distances are all possible and have been used in spatial optimization application. It is possible for it to be incorporated as a function and/or constraint when locations are not known in advance, as is the case for facility location in the Weber problem using the Euclidean metric. However, it is also possible to utilize distance as model coefficients, where they are derived in advance.

2.2.2 Adjacency and contiguity

The concepts of contiguity and adjacency are closely related. The relationship of two geographic features next to each other, or neighboring, is known as adjacency. As for contiguity, an usual description in the literature related to land management is that a group of spatial objects is contiguous if it is possible to travel between any two points inside the set of objects without leaving it (Cova and Church, 2000). Formally, two polygons are said to be contiguous when they share a common edge or arc. This is the reason behind Geographic Information Systems (GIS) that describe contiguity as the topological concept that allows the data model to determine adjacency between two spatial objects represented by polygons. In Figure 2.2, on the left, plots are abstracted by polygons, and plots that share an arc are contiguous. The information can be stored and used in different formats according to the spatial problem, either in adjacency matrices or lists.

2.2.3 Connectivity

The ability to get from one location to another or moving unimpeded from one location to another is known as connectivity. It is defined through arc-node topology and is the basis for many network tracing and path-finding operations. Connected arcs are determined by searching through the list for common nodes. Connectivity and contiguity are related in the sense that assessing contiguity often involves transforming space into a corresponding network, where nodes represent the spatial objects and arcs between nodes are defined if the objects are contiguous. In Figure 2.2 such relationship is demonstrated as polygons, on the left, which share an arc and are thus contiguous, are abstracted as points, on the right, and edges connect contiguous plots.

2.2.4 Spatial optimization challenges

Spatial optimization challenges include, besides mathematical abstraction of spatial problems, the integration of the algorithm with the whole package of technical solutions used in the process of optimization, striving towards the building of integrated systems supporting this process. Most of such systems rely on GIS techniques (Demetriou et al., 2013a; Tourino et al., 2003) and a lot fewer on CAD (Computer Aided Design) platforms. Additionally, there is the need for better data models.

The ongoing research to develop computationally efficient solution algorithms to spatial optimization represents the last challenge. Algorithms for solving spatial optimization problems can be either exact or heuristic. Exact algorithms are guaranteed to find an optimal solution by exhausting all possibilities or exploiting problem properties. There are many possible exact approaches, including derivative based techniques, enumeration, linear programming, integer programming

Figure 2.2: Plot layout. On the left: abstraction of lots as polygons; On the right: abstraction of lots as nodes

with branch and bound, etc. as well as more problem specific methods like the Hungarian algorithm, transportation simplex, out-of-kilter algorithm, Dijkstra's algorithm, among others.

Alternatively, heuristics are specific strategies, finding characteristically good solutions, but generally lacking any capacity to verify or validate the quality of the solution. The idea is to explore the solution space, in a strategical way, so that it is possible to identify feasible solutions in short times. Some heuristics start with one or multiple feasible solutions. According to the specific search rule of the heuristic, a new solution(s) is identified, and the current solution(s) is updated.

It may happen that no exact method is able to solve a problem, or that the computational effort exceeds allotted time, among other reasons that render exact methods unfeasible or inadequate. In fact, spatial problems, which are often NP-hard and have large data size, such as political, school and healthcare districting problems, resort frequently to simulated annealing, tabu search, genetic algorithms, and many other heuristic methods.

2.3 Clustering

Clustering or cluster analysis is the task of partitioning a data set into subsets, called clusters, in such a way that data points in the same cluster are as similar as possible, and data points in different clusters are as dissimilar as possible. According to Ge (2008), clustering methods can be classified into different classes, according to different perspectives, view Figure 2.3.

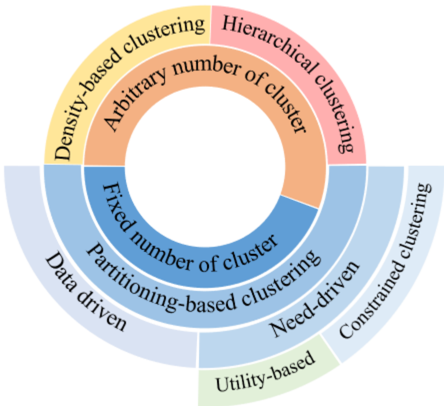


Figure 2.3: Classification of clustering methods. Adapted from: Ge (2008)

Generally speaking, clustering methods can have a specified number of clusters in advance or it can be decided by the clustering algorithm. The clustering methods can be classified into three major categories, based on the properties of the generated cluster, namely, partitioning, hierarchical and density based clustering. The first, partitioning-based methods, tend to find round-shaped clusters. Hierarchical methods can discover, as the name implies, a hierarchy of clusters from which clusters at different resolutions can be extracted. Density-based methods are good for finding arbitrarily shaped clusters.

The interest in clustering for the scope of the land-exchange problem resides in the first mentioned class. Partitional clustering is often modeled as an optimization problem and the quality of a cluster is measured by an objective function, which typically measures the compactness of the generated clusters. Depending on the purpose, partitioning-based methods can be either data-driven or need-driven. Data-driven methods seek to discover the true structure of the underlying data distribution and most traditional partitioning-based methods such as *k-Center* and *k-Means* can be viewed as data-driven methods.

As for need-driven methods, these achieve the same goal as that of data-driven methods while complying with application needs or background knowledge, as is the case of the need to form clusters of predetermined sizes in the land-exchange problem. Need-driven methods can be subdivided in utility-based or constrained clustering. For simple application needs, the objective function is defined by an utility function which instead of measuring the compactness of the generated clusters, measures the utility of a clustering in decision making. Therefore, this sort of clustering can yield more useful and feasible clusters than the ones generated by data-driven clustering.

However, it may fail to discover desired clusters for applications with complex application needs or background knowledge. In some problems, in order to make similarities inside clusters more significant, it is preferable to address other needs, for example the size of the groups, by imposing constraints, thus resorting to constrained clustering.

The constraints found in constrained clustering problems can fall into three categories: instance-level, cluster-level and model-level constraints. While instance-level impose specific requirements on pairs such as must-link and cannot-link, model-level constraints are enforced on a clustering. But it is the cluster-level constraints that arouse the most interest and pose new challenges.

In the most popular partitional algorithm, *k-means*, each data object is assigned to the cluster whose center (centroid) is the nearest. This is a squared error algorithm, meaning that it minimizes the within-cluster sum of squares (WCSS), and it converges to a optimum but does not ensure the global minimum. There are alternatives that aim to alleviate this problem, such as the *global k-means* but it is computationally expensive for large datasets. Additionally, because of its nature, the *k-means* algorithm can be used only if all attributes are continuous, or treated as such. The same does not happen with *k-medoids*, a similar algorithm with the difference that the representative of each cluster is a data object, that can deal with all types of attributes and all dissimilarity measures. However, the major obstacle to applying this type of algorithms to constrained clustering, is the fact that assigning a data point to the nearest cluster representative may not always be satisfied due to possible conflicts with constraints. As a result, they can fail to find a solution that satisfies all the constraints even when such a solution exists.

On the other hand, general and declarative frameworks using generic optimization tools offer the flexibility of handling a wide variety of user constraints, and finding an exact solution of the problem whenever one exists. A framework based on Integer Linear Programming (ILP) and column generation has been proposed by Aloise et al. (2012), where user constraints are handled within the branch-and-bound search, used for generating new columns. Using Constraint Programming, a generic framework has been developed in Dao et al. (2017), with a global constraint to compute and prune the search space for the WCSS criterion. Guns et al. (2016) suggested a repetitive branch-and-bound algorithm is combined with Constraint Programming used to compute tight lower and upper bounds.

Overall, Minimum Sum-of-Squares Clustering (MSSC), Min-Sum k-Clustering and other closely related clustering problems are NP-hard, so it is not simple to find the optimal solution in a short time. Most exact approaches to these clustering problems, resort to algorithms such as branch-and-bound and branch-and-cut. The hardest task while devising exact algorithms is to compute good lower bounds in a reasonable amount of time, in order to speed up the search process. Hence, although finding an exact solution may be of high importance on valuable datasets, it is only a reality for small datasets.

2.4 Optimization models in land consolidation

Methods of mathematical optimization for land consolidation have a long history and are constantly being developed, which is a result of the timeless and worldwide problems of fragmentation of land in rural areas, development of algorithms and problem-solving techniques and the improvement of computational performance. The existing optimization methods range from very general to problem specific formulations and from the strict mathematical to more intuitive formulations (Cay and Iscan, 2006). Between the reviewed models, two categories can be distinguished, whether they rearrange the parcel structure through the re-allocation of the owners' lands into blocks, or whether they re-assign current parcels to different landowners. It is worth mentioning that, even though the aim of this project falls in the second group, both are analyzed in this section, since the literature on the latter is scarce, and the former is useful for understanding the different frameworks of land consolidation optimization models.

In general, criteria in mathematical models are given as follows:

- Attending to landowner's wishes;
- Keeping constant foundations;
- Reducing parcel number;
- Minimizing parcel distances.

A number of different approaches and algorithms to obtain the final solution can also be recognized in the existing optimization models. One of the suggested ways is the use of heuristic methods, proposed by Tourino et al. (2003). In this approach, simulated annealing method was proposed for searching and narrowing the set of feasible parcel layouts. This model was embedded in a GIS environment and furthermore considers landowners' petitions and soil classes.

Another group of heuristic solutions are the methods based on genetic algorithms, which are an effective tool for solving difficult computational optimizing processes in the field of land consolidation (Akkus et al., 2012; Demetriou et al., 2013a). Genetic algorithm (GA) is an optimization algorithm based on genetic and natural selection mechanisms. It randomly tries to achieve a better solution in each iteration, by submitting existing solutions to operators (crossover, mutation) and ensuring the survival of the fittest at the end of each iteration. This is also the algorithm used by Teijeiro et al. (2020) in a land exchange approach which considers land values and provides the option of securing lots to owners if they so desire. Also, multiple fitness functions, such as distance between parcels, distance to reference point and number of parcels, are tested in this paper.

The application of genetic algorithms requires some parameters such as population size, mutation and crossover probabilities. These are very important for GA performance and it is known that the interaction between them is quite complex. Its adjustments are done by the user, usually before the evolution. Therefore, trial and error method is frequently used which takes a considerable amount of time. In addition, stagnation of the population may occur, and the algorithm converges to a local optimal solution. For these reasons, Ertunç et al. (2018) created a hybrid system, using fuzzy logic to adjust the GA parameters and adapt the genetic operators, enhancing in this way GA's performance. A further use of fuzzy logic methods obtained results on the example of the village Agalar that were largely in line with the 'interview-based land reallocation model', what indicates the possibility of using this method in the actual consolidation work (Cay and Iscan, 2011).

Another group constitutes methods based on linear programming, which due to the low speed of computers, for a long time, resulted only in the optimal allocation of the land of individual owners in a small number of blocks (complexes) of land. One of the approaches of these methods is the one proposed by Cay and Iscan (2006) which, after dividing the land into blocks by planning an optimal network for roads and channels, solves the problem of determining how much land and from which block should be given to a farm. Other suggested models resort to simplified cost function to express to some extent the preferences of landowners and the traveling time based on straight-line distances (Ayranci, 2007); minimize average real distance between farm and parcels (Harasimowicz et al., 2017); and use multi-criteria analysis, such as Analytic Hierarchy Process method to support the process of obtaining the optimized land configuration due to the preferences of the landowners (Cay and Uyan, 2013).

Other optimization methods, used in land consolidation, differ from the classical ones. This group includes the Minimum k-Star Clustering method, an integer binary program, used to solve the model of land consolidation assuming the change of the landowners while maintaining the existing arrangement of plots (Borgwardt et al., 2011). Another method proposed by Brieden and Gritzmann (2004) is to use a quadratic optimization model. Designed as a tool which can support the implementation of lend-lease agreements, it maximizes the distance between the centers of mass of holdings.

The three reviewed models that opt for maintaining the current plot (Borgwardt et al., 2011; Brieden and Gritzmann, 2004; Teijeiro et al., 2020) use different algorithms and methods, but all with good computational results. Both the Quadratic Optimization Model and Minimum k-Star Clustering resort to approximation algorithms to linearize the problem, and the latter then applies a small heuristic to solve integrality problems that result from the approximation. Both run in under a minute and are satisfied with near-optimal solutions. The Genetic Algorithm for Parcel Exchange resorts to parallelization techniques to improve its performances from more than 5 hours to under 10 minutes. In each of these models, the objective is fulfilled, through Euclidean distances between parcels, with the assumption that their reduction implies small driving distances, but also rewards adjacent lots to be assigned to the same holding. While land values considerations for fair distribution and the possibility to secure plots are integrated in all three models, other type of preferences, for parcel characteristics for example, are not.

In this dissertation, a land-exchange model, oriented towards forest consolidation, is proposed. It includes features to ensure the most common preferences of owners and implements a new functionality of guaranteeing compatibility between parcels and owners. The problem is defined mathematically, using insights and concepts of clustering and spatial optimization. An exact solution method is applied, by resorting to a state-of-the-art solver.

Chapter 3

Problem statement

Land consolidation projects through the exchange of parcels raise a problem that is not often studied: the land-exchange problem. In this chapter, the problem is presented. First, it is provided an overview of the factors to be taken into account, as well as some assumptions and explanatory notes. Then, it is shown that the optimization of land-exchange can be perceived as a clustering problem, and three criteria commonly used in cluster analysis are adapted to land-exchange. The required input data and their formats are listed, and how they are processed in a phase prior to optimization is further described. Subsequently, the problem is formally defined as a zero-one constrained quadratic program. Other variants, resulting from other clustering criteria, as well as linearization are presented.

3.1 Land-exchange problem framing

The land-exchange problem is a land consolidation optimization tool to assist in the task of reallocating lots among landowners in such a way that, as it is with all land consolidation projects, land tenure fragmentation is minimized. In addition, it is designed for voluntary use and to be performed without the need for direct state intervention. For this reason, current land parcels are to be maintained and therefore, while their location and shape remain intact, ownership is exchanged between owners. Land banks, which are used for provision of additional lands, are assumed to be unavailable in this sort of approach. Hence, parcels are treated as discrete, unshareable, and finite objects, to be reassigned among the participating owners, and the problem is of a combinatorial nature. Moreover, as parcels are not altered, distances and other spatial relationships between them may be derived in advance.

With regard to minimizing land fragmentation, as seen in Chapter 2, it is a rather comprehensive concept, to which many factors contribute. Given the internal fragmented situation in which one can expect to find Portuguese forestland and landowners, great improvements in the land tenure can be achieved by placing the parcels that belong to the same landowners as close as possible. Other factors, such as road access and parcel shape will not be taken into account in fragmentation measurement for the reason that they are not altered by the project. Furthermore, it is assumed that owners have full ownership of their parcels and of all that is contained in them. Consequently, improvements in the fragmented situation will be achieved only through spatial criteria.

The fact that the optimization model is designed for a voluntary basis gives special emphasis to right holders' satisfaction with the proposed solution. In one hand, an unbiased distribution must be ensured, which is explored in further detail in the next section, and on the other hand, each owner's personal views should be attended whenever possible. One type of criteria that many models neglect is landowners' wishes and preferences. Given the subjective nature and diversity of this sort of criteria attending to owners' preferences can be an exhaustive task. Section 3.1.2 details the approach chosen for this challenge.

3.1.1 "At least as well off" principle

A mentioned advantage of Land-Exchange is the fact that the owners do not need additional financial resources. However, this implies the necessity to ensure a fair distribution of the land parcels so that no owners become impaired by the exchange. The "at least as well off" principle, described as the key principle of land consolidation by FAO, is used to further this concept. The principle is stated as follows: *"(...) the situation of legitimate landowners and other right holders cannot be made worse by the implementation of a land consolidation project" (Veršinskas et al., 2020)*

It can be challenging to apply this principle. Presumably, participant owners have land parcels that differ in many ways, besides the most obvious feature - its area. In practice, this means a reasonable evaluation method is necessary to access each parcel's value. By doing so it is then possible to compare each owner's initial and final holding values. Ideally, both values should be identical, but this would turn the task of finding a feasible combination of holdings that simultaneously constitutes an improvement in the fragmented tenure pattern, very difficult, if not impossible. Thus, some slack between the initial and proposed holding value is allowed.

Nonetheless, the "at least as well off" principle should not be understood as referring exclusively to owner's land or property values before and after the project (Veršinskas et al., 2020) which means it can be accomplished, provided that the owner personal gains from consolidation outweigh the eventual losses of property value allowed by the slack.

3.1.2 Preferences and petitions

The opinions of the landowners, sometimes neglected, are of utmost importance even more because, since the exchange is voluntary, they can determine the acceptance of the proposed solution. Overall, attending to petitions and preferences can be an exhaustive task, as these are subjective in nature, may refer to many different criteria, e.g., location, type of soil, proximity to water sources, etc., and are not easy to assess and quantify. Even so, many models attempt to integrate this functionality. Through the land consolidation models included in the bibliographic review and other projects, it was possible to verify that the owners' requests were mostly about preferred soil types, location, and plots they intended to keep, often because of existing infrastructures.

As it is designed to be used on forest lands, unlike the agricultural framework models where farms play an important role as they are not exchangeable and distance from farm to farmland is very important, major infrastructures are not expected. Nevertheless, the option for landowners not to exchange certain parcels is incorporated into the model. It has been observed that fixed lots affect the quality of the solution negatively if these are present in a large number or are close to each other (Borgwardt et al., 2011; Teijeiro et al., 2020). Therefore, this functionality should be discouraged, except for parcels that have infrastructure, long-term investments such as orchards or any other feature, that makes the parcel indispensable, for the acceptance of the proposal by the owners.

Regarding other expressed interests, namely those concerning the characteristics of the plots, it is reasonable to assume that these are born from the preference of landowners for profitable land. Naturally, whether a parcel is profitable or not depends on the intended use for it. Since the Portuguese forest and agroforest landscape can be heterogeneous, especially in the hereabouts of agricultural and urban areas, the land-uses that can be found can be diverse. On the other hand, other factors, mainly legal, can restrict certain land-uses. For instance, recent legislation on population control of the species Eucalyptus, used on a large scale to supply the cellulose industry, forbids its plantation in areas occupied by other species and because of fuel management measures, in certain areas the minimum distance between fast-growing trees is 10 meters, which can compromise the economic viability of production of these species.

The proposed solution to deal with the aforementioned is to impose compatibility between landowner and parcel. Rather than meeting the specific wishes of landowners, this method outlines the profile of landowners in terms of the use they give to land. Also, it determines the suitability of each plot for the different land-uses in question, which can be done by applying land use suitability models to the study area and cross-checking this information with the legal framework that may exist for the zone. Having gathered this information about the owners and parcels, it is then possible to cross-check them to ensure that no parcel is assigned to an owner who has no use for it.

Compatibility is a concept usually implemented in land-use optimization models and can be defined as a spatial property where neighboring parcels have land-uses that do not conflict (Yao et al., 2018). In this model, compatibility is established between users and plots, but based on land-use, and ensuring it, is an additional constraint, which can compromise the main purpose if it is too restrictive. However, based on the premise that nearby parcels are likely to have similar characteristics and be regulated by the same rules, it is believed that, as long as the owners' preferences are not too restraining, it is possible to obtain a consolidated solution that appeals to all owners. In this way, compatibility might not only increase the probability of acceptance of the proposal but also improve the use of parcels' biophysical characteristics and enhance compliance with the territory planning rules.

3.1.3 Spatial criteria

In general, the most relevant impediments to the economic viability of land management, resulting from land fragmentation, are considered to be small plots and long driving distances between plots, the first because it makes economies of scale impossible and the second since transport costs increase.

Although the structure of the plots is not changed by the project, the intended effect of increasing the size of the plots can be achieved by placing adjacent plots in the same holding. The adjacency between parcels is not difficult to determine. GIS programs have the ability to determine contiguity between polygons and export this information under pairs of adjacent plots.

In an ideal scenario the holdings would be contiguous, however, considering the possible initial situations of the fragmented landscape, this cannot be the only criterion for creating holdings. Unlike the continuous blocks of land in comprehensive consolidation projects, the fact that the plots do not form a contiguous area is to be expected, not only due to the existence of natural boundaries, such as rivers, roads and steep slopes, but also because, given its voluntary character, holdings are most likely to be scattered and mixed with parcels not available for exchange. Hence, a measurement based on distance is also necessary to deal with scattered parcels.

Given the impact of driving distances on increasing transport costs and unproductive times, a good practice when assessing land fragmentation in holdings is to consider the total distance traveled by an owner to visit all his parcels in one round-trip. However, seeking to minimize this measure would constitute an extension of the NP-hard traveling salesman problem, in which, for each combination of holdings, the permutations that result in the shortest round-trips would have to be determined, and therefore, the problem would be formulated in a more complex manner than it needs to be. Bearing in mind that the potential to improve fragmented land tenure increases with the number of parcels and landowners involved in the exchange, this method was discarded as it would not be feasible for larger problems.

Therefore, measures based on dispersion using Euclidean distances and the centers of the parcels are considered sufficiently accurate to reduce the driving distance between plots. Different dispersion indicators can be obtained, either through the distance of all points to a reference point, or by calculating the distances between pairs of points. When the reference point is the holding's centroid, this measure and the sum of pairwise distances are equivalent, but the pairwise distances are established in advance and, as they represent a relationship between two parcels, they can be combined with other non-continuous attributes, such as contiguity.

In fact, if the holdings are created from the pairwise relationships between parcels then it is possible to combine several factors (*e.g.* distance, contiguity, current soil occupation similarity, *etc.*) in one single similarity metric, with the intended weights for each factor, even before the optimization.

3.2 Optimization model

Let the area in which the optimization will take place consist of a set of n parcels, $P = \{p_1, ..., p_n\}$, represented by their centers in a *2 - dimensional Euclidean space*, and a set $O = \{o_1, ..., o_m\}$ be the m participating owners, where usually $n >> m$. Figure 3.1 shows a weighted graph $G = (V, E, w_V, w_E)$ embedded in a 2 - dimensional space where each vertex $v_i \in V$ corresponds to the center of parcel i for $i = 1, ..., n$. A weight $w_V(v)$ returns the value of the parcel associated with the vertex v. As the graph is completed and undirected, there are $\binom{n}{2}$ edges. For each edge there is a weight w_E which represents the distance metric between the parcels associated with the endpoints of said edge.

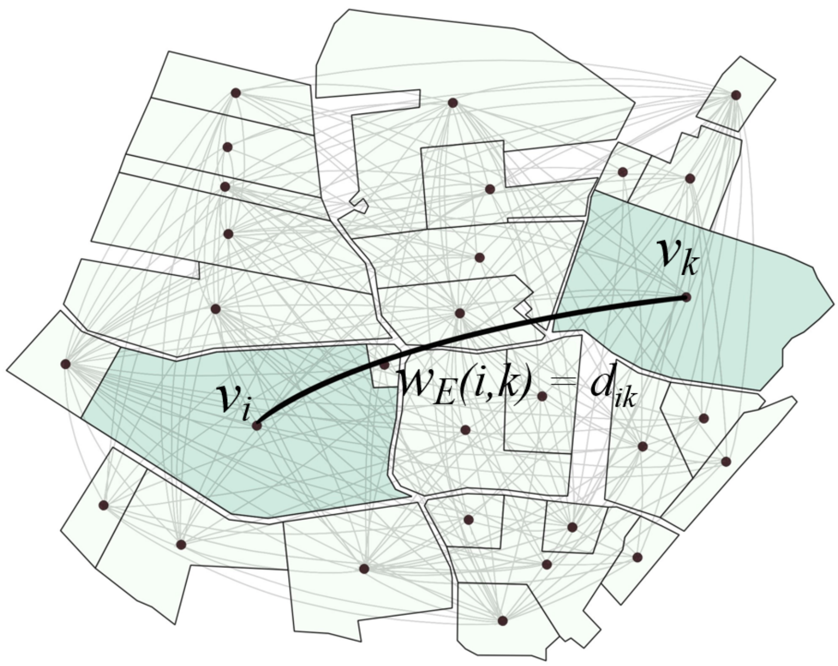


Figure 3.1: Graph representation of the land-exchange problem

The land-exchange problem is therefore to partition the vertices V into m disjoint clusters $\{H_1, ..., H_m\}$, according to their degree of proximity. A clustering criterion, to be later optimised, is needed as it is part of the problem definition. In the next sections the following three criteria will be addressed:

- minimum sum of pairwise distances, which consists in minimizing the total weight of the edges that have both endpoints in the same cluster;
- minimum average pairwise distance, which minimizes the sum of the average weight of the edges of each holding. Similar to the first one, but regards the number of edges of each holding, thus returning the sum of the average distance between two parcels for each holding;

- minimum sum of squares, which is to minimize the sum of squared distances of all parcel centers to their “cluster center”, which can be any point in space.

Another important aspect of the problem is that, as the H_j subset, which is the holding assigned to the j-th owner, must satisfy certain restriction on cluster weights related to the values of the holding, the problem is cluster-level constrained. Thus, the land-exchange problem can be perceived as a need-driven clustering problem, more specifically a cluster-level constrained clustering problem.

3.2.1 Inputs

Throughout this dissertation, it has been mentioned the fact that a land-exchange project is a engaging process, in the sense that, by design, it involves all stakeholders and requires communication with them. To apply the fundamental principles and ideas described in the framework section it is therefore necessary to gather the necessary information. This encompasses the preferences of landowners, which is one of the reasons why communication is indispensable, as well as all the characteristics of their property, whether they are of geometrical character or related to the uses and values of parcels. In the absence of data from which to infer the best method for inserting it in the model, the input parameters presumed to be necessary and sufficient are described below:

- The set of n land parcels to be reassigned;
- The coordinates (p_i^x, p_i^y) of the center of each parcel i;
- The value of each parcel i, s_i, given by area a_i multiplied by an evaluation index e_i;
- The number of l land-uses in the study area (examples);
- For each parcel, a vector $\mathbf{v}_i \in \{0,1\}^l$, describes the suitability or unsuitability of parcel i for each land use;
- The set of m landowners;
- The initial partition of the parcels; $H = \{H_1, ..., H_m\}$, being H_j the initial holding of owner j;
- A vector $\mathbf{w}_j \in \{0,1\}^l$ for each owner which, similarly to $\mathbf{v}_i$, records the acceptability of land uses by owner j;
- The contiguity relationships between each pair of parcels (i,k), for $i < k < n$. $\gamma_{i,k} = 1$ states the existence of contiguity, $\gamma_{i,k} = 0$ otherwise;
- The tolerance allowed, θ, between the initial and final values of holdings.

3.2.2 Pre-processing

After inserting the model inputs, there are certain steps that can be executed at a stage prior to the model itself. Such is the case of calculating the values of initial holdings, necessary to ensure that the holding value proposed by the model does not deviate excessively. The initial holding evaluation is stored in a parameter, s_j^{init} for $j \in O$, and its calculation is done as indicated in Expression 3.1.

$$s_j^{init} = \sum_{i \in H_j} s_i \tag{3.1}$$

The value of each parcel, s_i, is given by the multiplication of its area a_i and an evaluation index e_i. In the simplest case, $e_i = 1$, and the evaluation of initial and proposed holdings is based on the area of the same. This parameter may also be based on the price of the square meter, or a combination of other factors decided with the consent of the group of landowners.

Another contribution resulting from this phase is the establishment of compatibility between parcels and owners. Recalling that compatibility refers to land-uses, if there are l land-uses in the area to be optimized, then, for each parcel, there is a vector $\mathbf{v}_i$ of dimension l, which expresses the suitability for each land-use as a matter of yes or no, through binary components. For example, given the set of land-uses $L = \{$ *Fruit orchard, Nut orchard, Eucalyptus plantation, Marine Pine plantation* $\}$, the vector $\mathbf{v}_i = \{1,1,0,0\}$, tells us that the biophysical characteristics and legal framework in which parcel i is included, make it suitable for a fruit or nut orchard and unsuitable for Eucalyptus and Marine Pine plantations. Similarly, the vector $\mathbf{w}_j$ registers whether each land-use pleases the owner j. The objective of ensuring that no parcel is assigned to an owner who has no use for it, is then achieved by respecting the mathematical inequality 3.2.

$$\mathbf{v_i} \cdot \mathbf{w_j} \geq 1 \quad , \forall i \in P, \forall j \in O \tag{3.2}$$

As it will be seen in more detail, there is a variable to describe whether parcel i is assigned to owner j. Therefore, in this phase prior to optimization, pairs (i, j) are defined when condition 3.2 is verified and stored in a set, such that: $I = \{(i, j) \in P \times O : \mathbf{v_i} \cdot \mathbf{w_j} \geq 1\}$. These pairs will serve as indexes to the decision variables. Compatibility could also be imposed by model constraints or penalization points, but in this way, not only does it avoid increasing the complexity of the problem by adding constraints or multi-objective objective function, but it is also possible to reduce the number of variables in case of existence of incompatibilities.

The last procedure of pre-processing is the calculation of the similarity metric between 2 parcels, necessary for the similarity-based models. This metric can be Euclidean pairwise distances between coordinates of parcel centers given by Expression 4.2.

$$d_{ik} = ||p_i - p_k|| \quad \text{for} \;\; i < k \; < n \tag{3.3}$$

Or, it can be a combination of spatial relationship factors between parcels. Given the spatial errors induced by point abstraction, Euclidean distances can be adapted to better represent other relationships between parcels, such is the case of contiguity. In Expression 4.3, Euclidean distances and contiguity are combined in order to result a null distance if parcels i and k are contiguous.

$$d_{ik} = (1 - \gamma_{ik}) \, ||p_i - p_k|| \quad \text{for} \;\; i < k \; < n \tag{3.4}$$

In this way, in a minimization problem, a pair of contiguous parcels is preferred over a pair of non-contiguous even if the pairwise Euclidean distances are equal, which encourages the creation of contiguous holdings.

3.2.3 Minimum sum of pairwise distances

The minimum sum of pairwise distances land-exchange problem, consists on assigning the n parcels to m disjoint holdings, in such a way that the sum of the distances between every two parcels belonging to the same holding is minimized, and the holdings respect the value limits determined for each owner.

3.2.3.1 Zero-one quadratic minimum sum of pairwise distances

The minimum sum of pairwise distances land-exchange problem can be formulated as a Constrained Zero-One Quadratic Program. Zero-one quadratic programming seeks to minimize a quadratic objective function, along with the condition that each variable is restricted to take on a value of either zero or one. These problems can be divided into constrained (01CQP), where the land-exchange problem belongs, and unconstrained (01UQP), but both are combinatorial problems known to be NP-hard.

Thus, the decision variables of this model are binary variables that aim to describe to which landowners each parcel is assigned.

$$x_{ij} = \begin{cases} 1, & \text{if parcel } i \text{ is assigned to owner } j \\ 0, & \text{otherwise} \end{cases} \qquad \forall (i,j) \in I$$

Expressions 3.5 to 3.9 represent the formulation for the minimum sum of pairwise distances land-exchange. The objective function given by Expression 3.5 consists in the minimization of the sum of the distances between each pairs of parcels that belong to the same holding.

$$\text{Minimize} \quad \sum_{j=1}^{m}\sum_{i=1}^{n-1}\sum_{k=i+1}^{n} d_{ik}x_{ij}x_{kj} \tag{3.5}$$

Subject to:

$$\sum_{j=1}^{m} x_{ij} = 1 \quad ,\forall i \in P \tag{3.6}$$

$$\sum_{i=1}^{n} s_i x_{ij} \geq s_j(1-\theta) \quad ,\forall j \in O \tag{3.7}$$

$$\sum_{i=1}^{n} s_i x_{ij} \leq s_j(1+\theta) \quad ,\forall j \in O \tag{3.8}$$

$$x_{ij} \in \{0,1\} \tag{3.9}$$

Equation 3.6 ensures that each parcel is only attributed to one owner, as parcels are considered indivisible and non shareable items. Constraints 3.7 and 3.8 specify the lower and upper limits, respectively, on holdings' values. These limits result from each owner's initial holding evaluation, s_j, with a decrease or increase of $\theta\%$, and ensure fair distribution. They can, however, be disabled together or separately, if desired. Lastly, all decision variables are declared as binaries (Equation 3.9).

On a final note, when a landowner j wishes to keep a particular parcel i but still include it in the model, this can be ensured by adding a constraint as Equation 3.10.

$$x_{ij} = 1 \tag{3.10}$$

3.2.3.2 Linear minimum sum of pairwise distances

Since linearizations of quadratic and, more generally, polynomial programming problems, enable the application of well-studied mixed-integer linear programming techniques, the quadratic minimum sum of pairwise distances land-exchange problem is reformulated as a linear programming.

This linearization is achieved by replacing each product of variables in the objective function of the quadratic problem, $x_{ij}x_{kj}$, with a new variable $y_{ikj} \in [0,1]$. This variable describes if parcel i and k are both assigned to holding j. Since the variables x_{ij} are kept in the linear constraints, it is necessary to establish the relationship between them and the linearization variables. The inequalities 3.11 to 3.13 can be used, as these ensure that $y_{ikj} = 1$ if and only if $x_{ij} = 1$ and $x_{kj} = 1$.

$$y_{ikj} \leq x_{ij} \tag{3.11}$$

$$y_{ikj} \leq x_{kj} \tag{3.12}$$

$$y_{ikj} \geq x_{ij} + x_{kj} - 1 \tag{3.13}$$

Since this is a minimization problem and all variables x_{ij} are present in the objective function, only Equation 3.13 is needed.

Therefore, the linear formulation of the minimum pairwise distance is given by Expressions 3.14 - 3.19.

$$\text{Minimize} \quad \sum_{j=1}^{m}\sum_{i=1}^{n-1}\sum_{k=i+1}^{n} d_{ik}y_{ikj} \tag{3.14}$$

Subject to:

$$\sum_{j=1}^{m} x_{ij} = 1 \quad ,\forall i \in P \tag{3.15}$$

$$\sum_{i=1}^{n} s_i x_{ij} \geq s_j\,(1-\theta) \quad ,\forall j \in O \tag{3.16}$$

$$\sum_{i=1}^{n} s_i x_{ij} \leq s_j\,(1+\theta) \quad ,\forall j \in O \tag{3.17}$$

$$y_{ikj} \geq x_{ij} + x_{kj} - 1 \quad ,\forall (i,k,j) \in I' \tag{3.18}$$

$$x_{ij} \in \{0,1\},\ \ y_{ikj} \in [0,1] \tag{3.19}$$

Note that for the linearized formulation, I' is defined in the pre-processing phase, in a similar way as $I : I' = \{(i,k,j) : \mathbf{v_i} \cdot \mathbf{w_j} \geq 1 \wedge \mathbf{v_k} \cdot \mathbf{w_j} \geq 1,\ \ i,k \in P,\ \ j \in O\}$

3.2.4 Minimum average pairwise distances

The fact that the minimum sum of pairwise distances sums all the distances between each pair of parcels within the holdings without taking into account their size, makes holdings with a greater number of parcels, exhibit a significantly higher value for this metric, without this meaning that their parcels are more dispersed. Bearing this in mind, the expression suggested in this section, the minimum average pairwise distance, sums the average pairwise distance of each holding.

If n_j is the number of parcels in holding H_j, then there are $\binom{n_j}{2} = \frac{n_j(n_j-1)}{2}$ pairwise distances to be summed for that same holding. In order to obtain the average pairwise distance, the sum of the pairwise distances of each holding must be divided by the number of pairwise combinations of parcels which belong to the holding. Let n_j^e define the number of pairwise combinations for holding j, the new objective function is then expressed by expression 3.20.

$$\text{Minimize} \quad \sum_{j=1}^{m} \frac{\sum_{i=1}^{n-1}\sum_{k=i+1}^{n} d_{ik}\,x_{ij}x_{kj}}{n_j^e} \tag{3.20}$$

The need for the number of pairwise combinations of each holding introduces a new set of variables, n_j^e, and Constraints 3.21, which define the new variables as the number of pairwise combinations.

$$n_j^e = \frac{\sum_{i=1}^{n} x_{ij}\,(\sum_{i=1}^{n} x_{ij} - 1)}{2} \quad ,\forall j \in O \tag{3.21}$$

The problem is further subject to Constraints 3.6 to 3.10, that constrain the minimum sum of pairwise distances.

Approximation

When compared to the minimum sum pairwise distance, which might appear simple but is already difficult to solve, the introduction of a variable in the denominator, in the latter formulation, adds even more complexity to the problem.

The method found to simplify the average pairwise distance formulation is to use the initial holdings, since, due to Constraints 3.7 and 3.8, their size should not change significantly. Then, n_j^e is calculated based on the number of parcels of the initial holdings, and thus, it is a parameter which can be calculated beforehand and Constraints 3.21 are not necessary.

3.2.5 Minimum sum of squared distances

The within cluster sum of squares, which consists in summing the distances of each entity to the centroid of the cluster to which it belongs, is one of the most used criteria in clustering problems. Following this approach, the minimum sum of squared distances land-exchange problem, focuses on minimizing the sum of squared Euclidean distances between each parcel center and the centroid of the holding to which it belongs. This is expressed mathematically by expression 3.22.

$$\text{Minimize} \quad \sum_{j=1}^{m}\sum_{i=1}^{n} x_{ij}||p_i - c_j||^2 \tag{3.22}$$

The additional variable c_j, for $1 \leq j \leq m$, represents the centroid of holding j and can be obtained by the arithmetic mean of the centers of the parcels belonging to the cluster, as expressed in Equation 3.23.

$$c_j = \frac{\sum_{i=1}^{n} x_{ij} p_i}{\sum_{i=1}^{n} x_{ij}} \tag{3.23}$$

However, because $\sum_{i=1}^{n} x_{ij}||p_i - q||^2$ attains its global minimum for $q = c_j$ (Peng and Xia, 2005), there is no need to enforce Equation 3.23 and doing so, would detract from problem solving efficiency. Thus, the problem is subject to the same restrictions as the minimum sum of pairwise distances problem, which are Equation 3.6 through 3.10.

Chapter 4

Results

The present chapter begins with a description of the instances used for testing. It includes the computational results for the minimum sum of pairwise distances, minimum average pairwise distance and minimum sum of squares formulations as well as for the existing linearizations and approximations. The formulations are analyzed both for efficiency and instance improvements, and compared for the benchmark instance. After, they are tested for reality-based instances.

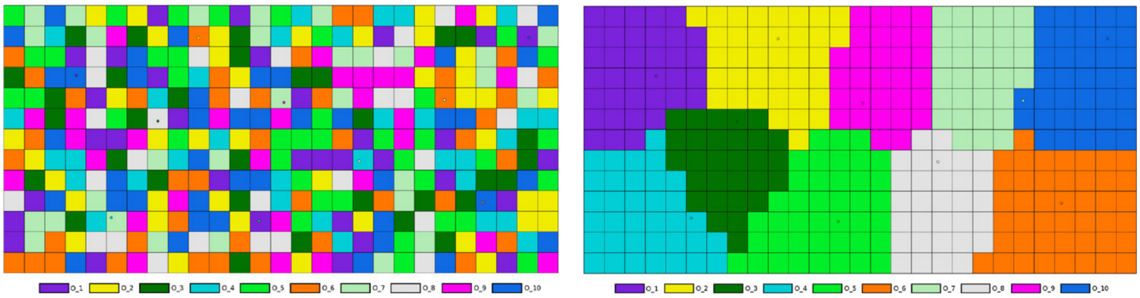
All tests were done using the state-of-the-art solver Gurobi Optimizer 9.1.1. The model was developed in the Python programming language, version 3.8, by resorting to the Python-Gurobi API. A standard laptop computer was used, with an Intel(R) Core(TM) i7-6500U CPU @ 2.50GHz, with 2 cores and 4 threads, and 8 GB of installed RAM.

4.1 Test instances

In a first phase, in order to test the feasibility of the model(s), synthetic instances were used with land parcels represented by equal adjacent squares, whose centers are one unit apart from each other. The initial ownership assignment was randomized. For future reference, the dimensions of these instances are expressed as A - B, with A being the number of parcels, and B the number of owners. Because the instances are uniform, the shortest distance between two parcels corresponds to a pair of adjacent parcels, and so, Euclidean distances between parcels' centers were used throughout this phase. Compatibility was generalized for all parcels and owners and therefore had no influence in the outcome.

4.1.1 Benchmark instance

One of the uniform instances used to test the model is similar to the one produced by Teijeiro et al. (2020). In this article, an instance of 13 by 27 plots and 10 owners is used to validate the results of the genetic algorithm used. The original and the optimized distribution for the average parcel distance fitness function are given and can be seen in Figure 4.1a and 4.1b, respectively.



(a) Original synthetic parcel distribution with random owners

(b) Algorithm results using average distance between parcels fitness function

Figure 4.1: Benchmark instance. Source: Teijeiro et al. (2020).

Since the objective function values are not available, the results were replicated and the average and total pairwise distances were calculated assuming a unitary distance between adjacent plots. Table 4.1 presents these results, as well as the registered maximum decrease and increase in holdings' values.

Table 4.1: Benchmark results

Total parcel distance	Average parcel distance	Max decrease / increase in value of holdings
19849	31.994	- 9,76% / 14,81%

4.1.2 Ribeira de Fráguas

The models were further tested in an instance based on a real plot structure located in the parish of Ribeira de Fráguas, Albergaria-a-Velha. The data, provided by the Baixo Vouga Forestry Association, includes the shape and location of the plots, and the owner to which each plot belongs. It was not possible to obtain information related to the preferences of the owners, either in terms of land-uses or of the parcels they wish to keep. Likewise, the biophysical characteristics of the plots and their suitability for land-uses were not taken into account. Consequently, the tests carried out in these instances only consider spatial factors and, in the analysis of the results, the adequacy of the solution to each owners' preferences and acceptance of the solution are not evaluated.

The information was visualized and manipulated using the QGIS 3.18 software. The integration of this tool with the Python-Gurobi API was not considered in the scope of the dissertation and therefore, the data was transferred manually between softwares.

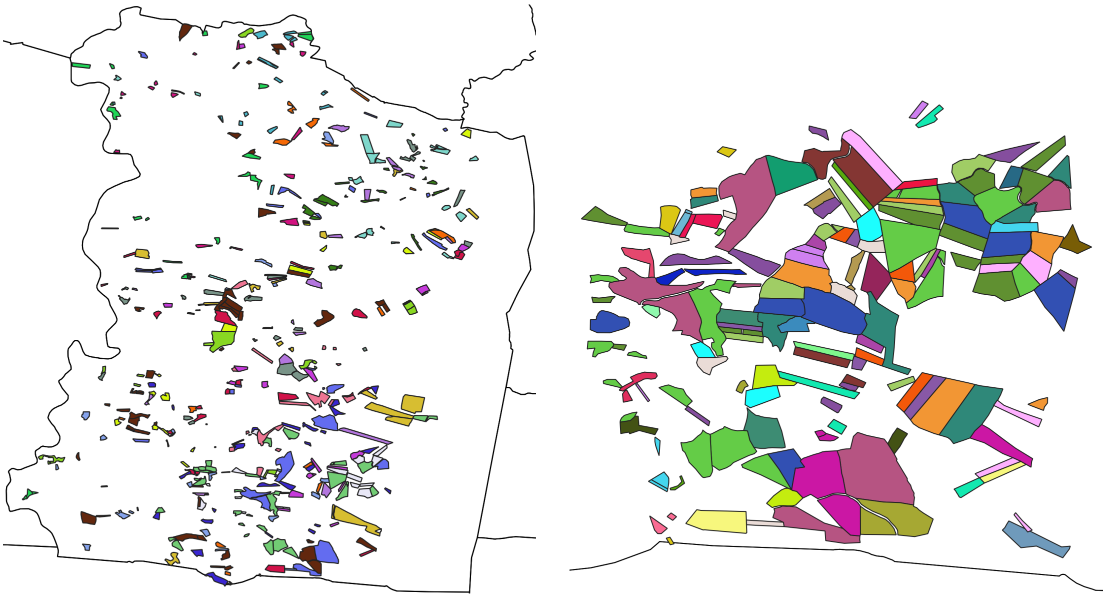
The original data file contained 1085 plots, represented by polygons. Among these, 28 polygons with excessively small areas were identified and considered as inaccuracies and errors in the creation of the polygons, and therefore, they were eliminated. The remaining 1057 plots range from 0.031 to 7.1 hectares. The average size of the plots is 0.53 hectares and only 12.5% of the plots are larger than 1 hectare. The parcels are spread over the 2,675 hectares of the parish which results in a low land coverage, particularly, on the northern part of the parish.

With regard to the quality of information, some inconsistencies were found. Firstly, some plots overlapped in certain areas, which should not happen. The opposite situation was also verified, in which there is a gap between the boundaries of two parcels, despite these being expected to be contiguous. What leads to this assumption is the fact that the gap is too small, making it impossible for a third parcel to exist and, it does not appear to exist discontinuities such as roads and water courses, which was observed through the overlay with satellite maps. The fact that there are gaps between the boundaries of parcels that should be contiguous makes it impossible to quickly extract information on the adjacency relationships. Also observed using satellite maps, while some parcels crossed by roads encompass this information, by being constituted by 2 polygons that contain only productive forest or agricultural area, other parcels, in the same situation, are constituted by a single polygon that includes the area crossed by the road. This situation is not ideal, since in the second case, the road area contributes to the evaluation of the holding to which it belongs.

The total of the parcels is held by 145 landowners. Naturally, the holdings are heterogeneous and vary in terms of number of parcels, area and dispersion. The average number of parcels per holding is 7.3, yet there are holdings consisting of only one parcel as well as holdings with up to 49 parcels. Figure A.1 in Appendix A, contains the visualization of the structure of the parcels, coded by colors, although, sometimes, the distinction is not noticeable due to the high number of landowners.

In fact, with so many different owners, it is very difficult in practice, to enter into negotiations with all interested parts. On the other hand, due to the combinatorial nature of the problem, 1057 parcels and 145 owners result in an excessive problem size, which compromises the effectiveness in finding good solutions that improve everyone's situation. Even after reducing the size of the problem for 944 parcels and 85, by excluding uninteresting holding for the land-exchange problem, such as holdings already consolidated or with few isolated parcels, the problem exhausted the available memory, in preliminary tests. It was therefore decided to significantly reduce the scale of the problem. A new decision problem arises of selecting m owners with whom to initiate the land-exchange. This strategic decision is important, in the sense that it will affect the efficiency of the land-exchange and therefore, it is desirable to identify owners who provide sufficient and preferably the best opportunities for improvement of the tenure structure in the region.

A good practice would be to choose the m owners in order to maximize the length of boundaries shared by parcels of different owners. This would increase the opportunities for creating contiguous areas. It is also the reasoning behind the work of Borgwardt and Schmiedl (2014), where this problem is studied in greater detail and solved. However, as this approach runs into an NP-hard problem, and the purpose of using this instance was to test the efficiency of the land-exchange models, a more practical approach was chosen. The 20 owners with the most parcels were selected. This results in a new instance whose initial distribution can be seen in Figure 4.2a. It is composed by 20 landowners and 426 dispersed parcels. The opportunities to form contiguous areas are underutilized, particularly in the lower right corner.



(a) 426-20 instance (b) 159-42 instance

Figure 4.2: Ribeira de Fráguas instances

Since the plots are dispersed, a second instance was created to study the behavior of the models, together with different similarity metrics, when the land coverage is high. The lower right corner was selected for this purpose. The parcel boundaries were also corrected as deemed appropriate in order to be able to withdraw the information, concerning contiguities. These corrections, of very small dimension, did not result in the alteration of the centroid coordinates nor the areas of the plots. The instance, which can be observed in Figure 4.2b, consists of 159 parcels and 42 landowners.

4.2 Computational results

In this section, the numerical results obtained by solving the uniform instances with the minimum sum of pairwise distances (SPD), minimum average pairwise distances (APD) and minimum sum of squared distances (SSD) formulations, with a time limit of 3 hours are exposed, and an analysis of the characteristics of each formulation is performed. Additionally, the performance of the linearized SPD and the approximation of the APD are assessed, by comparing their solutions against those of the respective original formulation. For all the formulations, the corresponding sum of pairwise distance and average pairwise distance are calculated for comparison.

After evaluating the performance of the models for the uniform instances, the models whose results were positive, are applied to the instances built from the tenure structure of the parish of Ribeira de Fráguas. A small experiment is also carried out involving the modification of the similarity metric capturing contiguity relationships, in order to study its impact on the solution proposed by the solver.

4.2.1 Minimum sum of pairwise distances

The results of the commercial solver's performance in dealing with the minimum sum of pairwise distances land-exchange problem and the impact of optimization on the holdings are included in Table 4.2. It includes the best objective, which corresponds to the SPD, and also the calculated APD for the obtained holdings, for future comparison purposes. Regarding performance assessment, the table indicates the average run-time in minutes, if optimality was reached and the optimality percentage gap. In terms of the impact on holdings, the maximum percentage increase and decrease in value are given ($\pm\Delta s_j$), as well as the mean squared deviation from initial holding values (MSD). The last two columns refer to the percentage improvements registered in SPD and APD values in relation to the initial situation. As an example, a SPD improvement of 70.0% corresponds to a reduction of the SPD value to 30% of the initial value. The results were obtained for a tolerance of 10% both for maximum allowed increase and decrease in holding values.

Table 4.2: Results of the SPD land-exchange using the commercial solver.

n	m	**SPD**	APD	time	opt.	gap	$\pm\Delta s_j$	MSD	$\downarrow$ SPD	$\downarrow$ APD
25	4	104.95	5.69	7.8 (s)	y	0.0	-0.0 / +0.0	0.0	44.8	46.4
49	4	534.29	7.73	180	n	11.7	-7.1 / +9.1	1.0	48.5	47.4
64	4	1039.7	8.62	180	n	52.3	-5.6 / +7.1	0.5	49.3	49.3
81	5	1391.1	11.24	180	n	100	-5.9 / +6.7	0.4	52.3	52.6
100	5	2419.6	12.33	180	n	100	-7.4 / +6.3	1.6	53.1	53.0
150	6	4837.7	16.01	180	n	97.7	-10.0 / +10.0	3.0	59.7	60.0
200	7	8022.2	20.35	180	n	99.7	-9.4 / +7.7	7.4	64.7	64.3
250	8	11738	24.62	180	n	100	-8.8 / +7.7	4.5	67.4	67.8
300	9	15292	27.94	180	n	100	-10 / +9.7	6.22	68.7	68.8
351	10	20250	33.3	180	n	100	-9.8 / +10.0	7.4	69.6	69.3

The results for the minimum sum of pairwise distances land-exchange emphasize the computational complexity of solving combinatorial problem with exact methods. For small-sized instances, with parcels up to 25 and 4 owners the solver handles the problem effectively, as the optimal solution was reached within 10 seconds. However, as the number of parcels and landowners increases, Gurobi fails to compute the optimal solution within the time limit. For 81 parcels and 5 landowners, the performance is further degraded as the gap after 3 hours of optimization is consistently close to 100%.

Yet, the improvement both for total and average pairwise distances is not overly degraded. Although, increasing the size of the instance, implies the remaining gap is close and even 100%, with no guarantees of optimality, the improvements registered also increase. This is expected given that there are more opportunities for improvement in larger instances, but these improvements are also reflected in the holdings as the solver is generally able to retrieve contiguous and compact holdings. For visual results, see Figure B.1 in Appendix B.

A new set of tests was run with 0% tolerance, to understand the relationship between the allowed variations in holding values and the SPD. In table 4.3 the SPD, time in minutes, if optimality was reached and the gap are presented, both for 0% and 10% tolerance.

Table 4.3: Comparison of results of the SPD land-exchange using the commercial solver for $\theta =$ 0% and 10%.

		θ=0%				θ=10%			
n	m	SPD	time	opt.	gap	SPD	time	opt.	gap
25	4	104.95	4.8(s)	y	0.0	104.95	7.8(s)	y	0.0
49	4	542.92	28.8	y	0.0	534.29	180	n	11.7
64	4	1054.4	180	n	49.7	1039.7	180	n	52.3
81	5	1397.4	180	n	100	1391.1	180	n	100
100	5	2493.4	180	n	100	2419.6	180	n	100
150	6	4967.5	180	n	95.9	4837.7	180	n	97.7
200	7	8332.9	180	n	99.6	8022.2	180	n	99.7
250	8	11859	180	n	99.2	11738	180	n	100
300	9	16030	180	n	99.5	15292	180	n	100
351	10	21301	180	n	99.7	20254	180	n	100

The first observation that can be made is that, for $\theta = 0\%$, the optimality was reached for instances up to 49 plots and in shorter times, when compared with the results of 10% tolerance. However, with the exception of instance 25 - 4, the value of SPD is lower when the tolerance is set for 10%. This is a result of two factors. First, looking at the visual results, B.2 in Appendix B, for some larger instances, the solver was more effective in finding contiguous and compact holdings with tolerance set to 10%. On the other hand, tolerance makes it possible to reach optimal solutions with a lower SPD value. Although it was only observed for the instance of 49 parcels, where the optimal solution for θ=0% has a higher SPD than the non-optimized solution for θ=10%, this behavior is intrinsic to SPD land-exchange formulation. For each holding, the model sums $\binom{n_j}{2}$ distances, being n_j the number of parcels in holding j, and, for this reason, SPD is very sensitive to the number of observations in each cluster. In the case of uniform instances, the sum of the distances are minimal for holdings with a similar number of parcels which results in the homogenization of the size of the holdings. In fact, it was observed that the holdings with the initial highest scores, had the highest losses and, to the smaller holdings, more parcels were assigned, corresponding to the largest positive variations verified, thus homogenizing the size of the holdings. Hence, this metric accentuates the confrontation between minimizing the sum of pairwise distances and minimizing variations in the value of holdings of landowners.

4.2.2 Comparison between linear and quadratic SPD

The quadratic and linear formulations of the SPD begin to diverge in the size of the problem. The number of variables and constrains, ignoring any possible incompatibilities between owners and parcels which could result in fewer variables, both for the quadratic and the linear formulation are presented in Table 4.4.

Table 4.4: Number of variables and constraints of quadratic and linear SPD.

Formulation	Number of variables	Number of constraints
Quadratic	nm	$n+2m$
Linear	$\frac{n(n+1)}{2}m$	$n+2m+\frac{n(n-1)}{2}m$

By replacing each product $x_{ij}x_{kj}$ by a new variable y_{ikj} and adding a set of linear constraints that forces y_{ikj} to be equal to $x_{ij}x_{kj}$, the size of the problem increases by $n(n-1)m/2$ variables and constraints. The performances of both formulations for the set of uniform instances were compared, and the results are exposed side by side in Table 4.5. The results include the best SPD obtained, its improvement over the initial situation, the run-time in minutes, whether the optimal solution was reached and the gap. The instances were solved with a 0% tolerance since the interest resided in the performance. The results show two distinct behaviors. In a first phase, for instances

Table 4.5: Comparison of results of the linear and quadratic SPD land-exchange using the commercial solver, for θ=0%.

		Linear					Quadratic				
n	m	SPD	time	opt.	gap	↓ SPD	SPD	time	opt.	gap	↓ SPD
25	4	104.95	2.9(s)	y	0.0	44.8	104.95	4.8(s)	y	0.0	44.8
49	4	542.92	62.7(s)	y	0.0	47.7	542.92	28.8	y	0.0	47.6
64	4	1054.4	3.65	y	0.0	48.6	1054.4	180	n	49.7	48.6
81	5	1394.6	180	n	0.79	52.1	1397.4	180	n	100	52.1
100	5	2883.5	180	n	14.3	44.1	2493.4	180	n	100	51.6
150	6	6849.4	180	n	36.3	43.0	4967.5	180	n	95.9	58.6
200	7	16641	180	n	77.6	26.7	8332.9	180	n	99.6	63.3
250	8	25069	180	n	85.2	30.5	11859	180	n	99.2	67.1
300	9	31482	180	n	91.2	35.6	16030	180	n	99.5	67.2
351	10	46213	180	n	100	30.6	21301	180	n	99.7	68.0

up to 64 parcels and 4 owners, the linear model obtained optimal solutions for all three instances and in less than 5 minutes, whereas, for the quadratic one, this was only possible for the smallest instance, and, for 49 parcels and 4 owner, it took the solver almost 30 minutes to reach optimality. For 81 parcels, the linear model still achieved a SPD value slightly lower than that achieved by the quadratic formulation, and the gap was very close to 0%.

However, from this instance onwards, the performances of the models are reversed. None is able to reach the optimality, but, when assessing the obtained holdings, it is clear that their quality is degraded for the linear formulation. While the quadratic model, although unable to produce good bounds, returns somewhat contiguous and compact holdings, the linear model produces dispersed holdings, view B.3 in Appendix B. This degradation is visible in the SPD improvement values as they decrease with larger instances, for the linear formulation.

The origin of this divergence could be in the size of the problem. For smaller instances, the linearization of the objective function eliminates the quadratic terms that are more complex to deal with, and allows taking advantage of the specialization of Gurobi in solving MILPs. On the other hand, the linear formulation introduces more variables and constraints, and for larger instances the increment is significant. For example, for 200 parcels and 7 owners, there are a maximum of 1400 variables for the quadratic formulation, while for the linear model there could be as much as 140,700 variables. Thus, it may happen that, for instances above a certain size, the gains from reducing complexity by eliminating quadratic terms, are overcome by the added difficulty of the increased size of the problem.

4.2.3 Minimum average pairwise distance

In Table 4.6, the same parameters used for the SPD describe the results of the average distance between pairs. The results reveal the difficulty that the average pairwise distance model imposes. Even for the smallest instance, 25 - 4, which the minimum sum of pairwise distances model solves in under 10 seconds, the average pairwise distance function cannot prove the optimality of the solution within 3 hours of optimization. As for the immediately following instance, 49 - 4, the model violates the constraints, and for instances with 64 parcels and upwards, Gurobi fails to produce incumbents within the established time limit, and therefore, such results are not included.

Table 4.6: Results of the APD land-exchange using the commercial solver.

n	m	**APD**	SPD	time	opt.	gap	$\pm\Delta s_j$	MSD	↓ SPD	↓ APD
25	4	5.69	104.95	180	n	32.5	0.0 / 0.0	0.0	44.8	46.4

For the approximation model both the objective function and the real average pairwise distance are given in Table 4.7. Besides these parameters, the corresponding SPD, the average runtime in minutes and the percentage gap of the solver are included, as well as if optimality was reached. Information on variation of holdings is presented and instances were solved with a 10% tolerance both for the lower and upper limits on holdings' value.

Table 4.7: Results of the APDa land-exchange using the commercial solver.

n	m	**APDa**	APD	SPD	time	opt.	gap	$\pm\Delta s_j$	MSD	↓ APD
25	4	5.69	5.69	104.95	6.9(s)	y	0.0	0.0 / 0.0	0.0	46.4
49	4	7.67	7.67	553.57	180	n	10.5	-9.1 / 7.1	0.5	47.8
64	4	8.64	8.65	1073.1	180	n	100	-7.1 /- + 5.6	0.5	49.1
81	5	11.2	11.1	1406.9	180	n	100	-6.7 / +5.9	0.8	53.0
100	5	12.16	12.21	2581.8	180	n	100	-6.3 / 7.4	1.6	53.5
150	6	16.04	16.05	5058.7	180	n	98.6	-10 / +4.0	1.0	59.9
200	7	20.05	20.03	8292.1	180	n	100	-7.7 / 9.4	2.3	64.8
250	8	24.09	24.10	11834	180	n	100	-3.8 / +5.3	1.3	68.3
300	9	27.76	27.82	15835	180	n	100	-8.3 / +3.2	1.1	69.0
351	10	31.54	31.63	20411	180	n	100	-7.4 / 9.3	3.8	70.8

The results of the approximation are more satisfactory than those of the APD. The solver is only able to solve the instance of 25 parcels to optimality, which is done in under 10 seconds. Larger instances are not solved within the established time limit of 3 hours. Similarly to the SPD, this does not manifest itself in the quality of the solutions found. The obtained holdings are compact and continuous, refer to Figure B.4 in Appendix B, and the improvements in APD and SPD are consistently increasing. In fact, the behavior of this formulation, with respect to solver efficiency, is very similar to that of SPD, which is expected since the formulations are very similar.

Regarding the proximity of the objective function to the average pairwise distance, the difference, for this set of tests, does not exceed 1% of the real value. Although promising, it does not allow to conclude anything about the usefulness of this function, since the instances which are being dealt with are uniform, and so the initial number of parcels, which defines the parameter n_j^e for each holding, is proportional to the size of the holdings, constrained to be close to the initial value. In real instances, with parcels of different areas and values, this proportionality relationship may or may not be verified, so further tests must be carried out, to draw conclusions.

4.2.4 Minimum sum of squared distances

Firstly, the minimum sum of squared distances land-exchange formulation was adapted since the solver only accepts quadratic terms and, therefore, does not accept the objective function $\sum_{j=1}^{m}\sum_{i=1}^{n} x_{ij}||p_i - c_j||^2$. An auxiliary variable was added, d_{ij}, to replace $||p_i - c_j||^2$. Thus, the Constraint 4.1 was added and the problem becomes quadratically constrained.

$$d_{ij} = ||p_i - c_j||^2 \quad ,\forall (i,j) \in I \tag{4.1}$$

This formulation was then tested for the same set of uniform instances and with $\theta = 10\%$. In Table 4.8, the results for this formulation are presented, and include the final sum of squared distances, the corresponding total and average pairwise distance, in addition to all the parameters that have been described for previous results.

Table 4.8: Results of the SSD land-exchange using the commercial solver.

n	m	**SSD**	SPD	APD	time	opt.	gap	$\pm\Delta s_j$	MSD	$\downarrow$ SPD	$\downarrow$ APD
25	4	25.54	104.95	5.69	105	y	0.0	0.0 /0.0	0.0	44.8	46.4
49	4	100.7	548.7	7.77	180	n	88.1	0.0 /0.0	0.0	47.1	47.1
64	4	163.2	1039.7	8.62	180	n	91.4	-5.6 / +7.1	0.5	49.3	49.3
81	5	241.8	1419.4	11.5	180	n	100	-5.9 / +6.7	0.8	51.3	51.5
100	5	373.6	2517.3	12.6	180	n	100	-7.4 / +9.5	2.0	51.2	52.0
150	6	629.8	4978.2	16.2	180	n	100	-8.0 / +7.7	3.6	58.6	59.7
200	7	980.5	8017.4	20.3	180	n	100	-9.4 / +8.3	4.6	64.6	64.4
250	8	1335	11588	24.0	180	n	100	-9.7 / +7.7	3.3	67.9	68.4
300	9	1747	15601	28.1	180	n	100	-10 / +9.1	6.2	68.1	68.7
351	10	2196	19775	32.4	180	n	100	-9.8 / +9.1	9.6	70.3	70.1

The first observation is that, again, it was only possible for the solver to reach the optimal solution for the smallest of instances, and it took 105 minutes, a time considerably higher than the 10 seconds achieved by the SPD and the APDa formulations. Even so, and analogously to its counterparts, the holdings obtained present quality, Figure B.5 in Appendix B, which is reflected in the percentage of improvement verified in relation to the initial situation.

4.2.5 Comparison of formulations and benchmark

Comparing all formulations, it is possible to start by discarding the average pairwise distance, given its ineffective performance. Regarding the linearization of the SPD, despite positive results for small instances, the scope of optimization requires a good performance for larger instances, which renders this formulation inadequate.

This restricts the range of potential formulations to SPD, APDa and SSD. Although none of them can reach optimality for instances greater than 25 parcels, and for a tolerance of 10%, the solutions obtained are satisfactory, as the holdings, with few exceptions, are compact and contiguous. Furthermore, the SPD and APDa formulations have very close computational performances, which is to be expected since they are very similar. Regarding the improvements observed in the initial situation of each instance, the three formulations present similar values. The SPD formulation exhibits the best values for the sum of pairwise distances and the APDa for the average distance, both due to slight differences that do not exceed 2%.

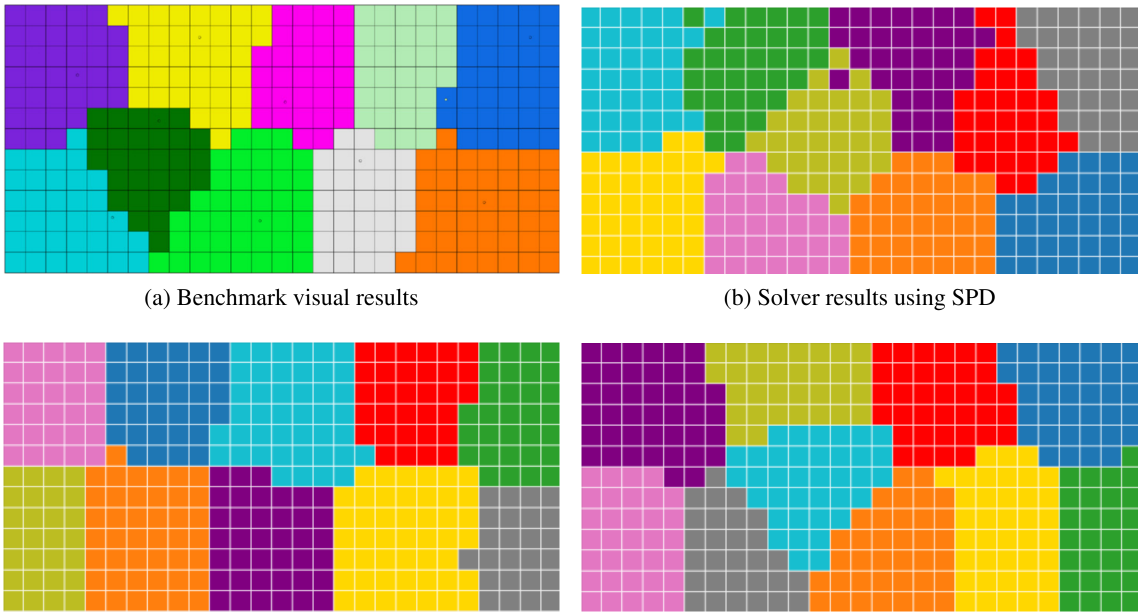
Concerning the performance of these three formulations for the benchmark instance, Table 4.9 summarizes the results for the 3 hours of optimization. The upper limit of 10% has been removed to better simulate the benchmark results.

Table 4.9: Comparison between SPD, APDa and SSD, for benchmark instance, 3h run-time.

Formulation	SPD	ADP	$\pm\Delta s_j$	↓ SPD	↓ APD
GA	19849	31.99	- 9.8 / 14.8	70.2	70.5
SPD	19608	32.52	-9.8 / 22.2	70.5	70.0
APDa	20411	31.54	-7.4 / 9.3	69.3	70.8
SSD	19775	32.35	-9.8 / 9.1	70.3	70.1

The values of total and average pairwise distances and improvements verified, both for the benchmark and the proposed formulations, present similar values. Both the SPD and SSD formulations obtain slightly lower values for the sum of the pairwise distances while the APDa succeeds in finding an average pairwise distance value lower than that found in the benchmark. These are small differences that do not exceed 1% in the registered improvements. To better assess the obtained holding the visual results are provided in Figure 4.3.

Observing the visual results, it is noticeable, in Figure 4.3b, that the SPD function does not produce completely contiguous and compact holdings, since there are parcels isolated from the holdings they belong to. However, of all the results, it is the one that presents a smaller value for the sum of pairwise distances. The fact that a solution that is visibly not the best presents the best values for the minimum sum of pairwise distances metric, raises the doubt if this is the



(a) Benchmark visual results

(b) Solver results using SPD

(c) Solver results using APDa

(d) Solver results using SSD

Figure 4.3: Results of benchmark instance, 3h run-time.

most appropriate metric to guarantee the reduction of fragmentation. As for the APDa and SSD formulations, in Figures 4.3c and 4.3d respectively, both produce contiguous holdings, with a tendency of SSD to form more roundish shapes.

Since the results were obtained for 3 hours of run-time and in the article from which the benchmark was taken it is stated that the results were obtained in about 10 minutes, the tests were redone to study the performance under shorter times.

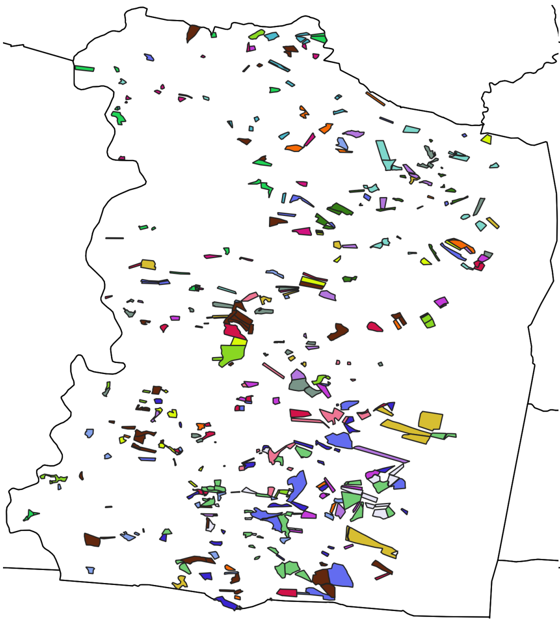
A time limit of 15 minutes was imposed. Both APDa and SSD succeed in forming contiguous holdings, and again, the results, in Table 4.10, are very close to those obtained by the benchmark. The SSD converged under 10 minutes to a reduction of 70.1% in the sum of pairwise distances while the APDa manages to surpass the benchmark with an average pairwise distance of 31.74 reached at 610 seconds. Regarding SPD, which has already exhibited difficulties for 3 hours of optimization, in 15 minutes it could not find a competitive solution. The configurations can be observed in Figure B.6 of Appendix B.

Table 4.10: Comparison between SPD, APDa and SSD, for benchmark instance, 15 minutes run-time.

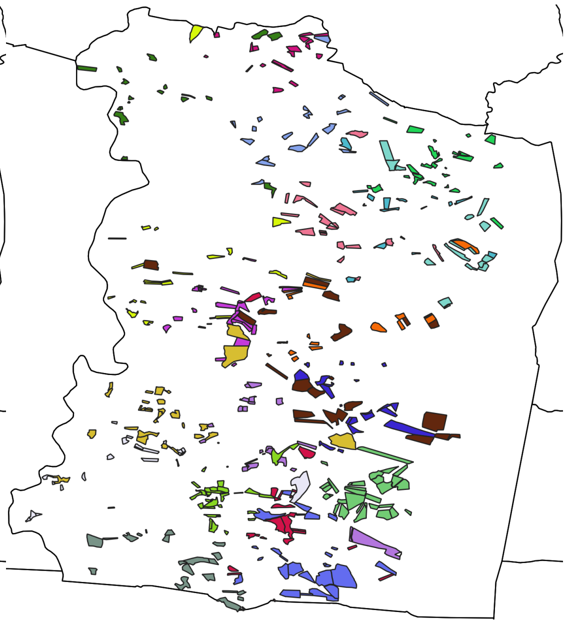
Formulation	SPD	ADP	↓ SPD	↓ APD
SPD	20729	34.4	68.9	68.2
APDa	20461	31.7	69.3	70.7
SSD	19866	32.6	70.2	69.9

4.2.6 Results for Ribeira de Fráguas instance

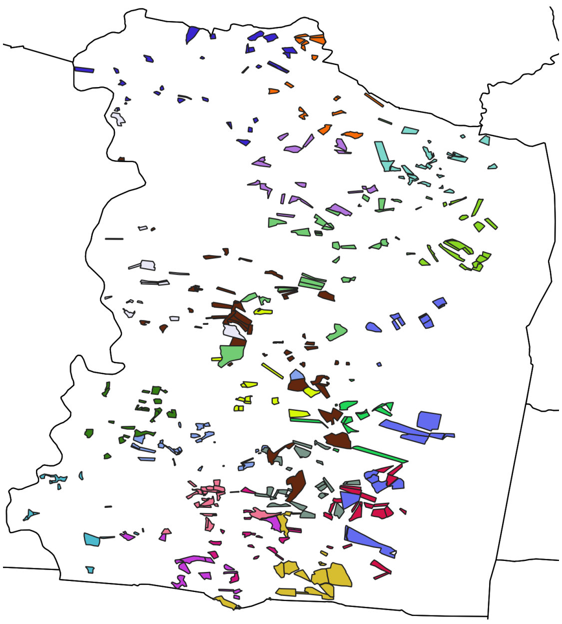
Despite the SPD formulation having presented less satisfactory results for the regular instances of higher dimensions, it was decided to nonetheless, test it for the reality-based instance. Thus, the SPD, APDa and SSD formulations were applied to the instance built from the fragmented structure of Ribeira de Fráguas, by the 20 holdings with more parcels. All tests were performed within 3 hours and with a tolerance of 10%. The resulting tenure structure can be observed in Figure 4.4.



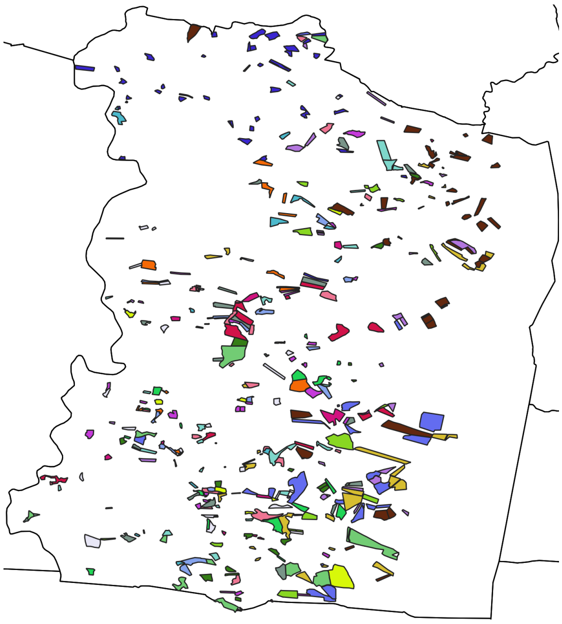
(a) Current holdings



(b) Solver results for SPD



(c) Solver results for APDa



(d) Solver results for SSD

Figure 4.4: Results of 426-20 Ribeira de Fráguas instance, 3h run-time.

For some holdings, the improvements are notorious since parcels are concentrated and separated from other holdings, but this is not always the case. The improvements were therefore translated into values which are described in Table 4.11. The average distance between two parcels belonging to the same holding was used to compare results since this measure is not susceptible to the number of parcels of the holding. For each formulation and each holding, the reduction (or increment) relative to the current value of this measure is presented. The number of parcels in each holding is also presented.

Table 4.11: Results of SPD, APDa and SSD, for 426-20 Ribeira de Fráguas instance, 3h run-time.

	Current		SPD		APDa		SSD	
Holding	APD	n	% APD	n	% APD	n	% APD	n
12	2071.6	15	-61.61	28	-78.12	15	11.64	30
20	1752.3	23	-61.38	21	-74.62	24	28.88	21
64	1929.1	21	-56.90	22	-73.78	20	-10.53	15
80	1235.3	27	-57.36	22	-31.26	30	-7.27	39
89	1457.1	16	-48.87	17	-71.02	22	54.25	25
91	2358.7	49	-49.51	19	-57.68	53	-16.74	36
197	2322.8	17	-78.27	22	-83.14	14	-35.67	19
200	2130.9	16	-70.41	23	-71.87	16	13.14	28
252	1000.0	32	-60.34	28	-15.76	34	143.35	11
299	1507.9	19	-70.06	18	-71.69	17	39.39	12
338	2084.2	17	-52.45	19	-76.42	13	-17.36	10
353	1241.2	18	-53.22	20	-77.88	23	77.53	24
537	2483.6	27	-76.25	24	-52.10	19	-27.43	24
541	2778.6	16	-71.11	17	-78.14	16	-40.57	14
705	2098.0	17	-77.18	24	-73.07	15	-3.82	31
784	889.5	17	-22.97	16	-28.35	13	181.01	12
963	894.2	18	-12.43	21	-14.75	15	114.40	29
1461	1928.3	25	-64.07	19	-71.34	27	-8.88	14
9001	2197.1	16	-84.16	29	-81.51	18	16.97	14
9022	769.0	20	9.06	17	-48.62	22	161.70	18
Total	35129	426	-61.46	426	-66.37	426	14.57	426

Beginning with the results for the SSD solution, both the values of the average distance between parcels and Figure 4.5d, reveal that this model, although effective in previous uniformed tests, failed to provide a good solution for this reality-based instance as the average distance between parcels belonging to the same holding is 14.57% higher than the current.

As for the SPD and APDa formulations, the overall average pairwise distance is reduced for both models, by 61.46% and 66.37%, respectively. However, for the SPD, the situation of holding number 9022 worsened, with the average distance increasing by 9.06%, which is not desirable as all holdings should be improved to promote acceptance of the proposed solution. Regarding the number of parcels attributed to each holding, it was found that for the APDa formulation, this number remains close to the initial one. This behaviour is assumed to be a result of using the initial number of parcels to calculate the average pairwise distance. In the case of dispersed instances, with few contiguous parcels, this characteristic can represent an advantage, since it maintains the relation of area-number of parcels of the holding. However, for instances with high land coverage, in which the relation between area and number of parcels of the holding is of little importance because contiguous areas can be formed by a high number of parcels, this characteristic may prevent reaching a better solution.

Lastly, the objective function value registered for the APDa formulation, of 11,477, compared to the real APD value obtained of 11,815, represents a difference of less than 3% of the real value. Thus, despite the high heterogeneity of the parcel areas, the APDa formulation provided a good approximation of the APD measure.

4.2.7 Comparison of Euclidean and modified Euclidean similarity metrics

The test instance built from the lower right region of the parish of Ribeira de Fráguas, is designed to study the performances of different criteria and similarity metrics in the allocation of parcels in an instance with high land coverage. The analysis by holding, as performed for the previous section, will not be carried out, since 10 out of the 42 owners only own one parcel, and so it is expected that their situation will not improve and may even worsen, if they are assigned more than one parcel or one of a smaller area. The analysis of the results focus on the assignment of adjacent parcels to the same holding, according to the similarity metric. For this purpose, two different similarity metrics were compared. The first (Equation 4.2), the Euclidean distance between the centroids of two plots, which has been used in all tests up until now. The second, a modified Euclidean distance in order to return a null value if the pair of parcels are contiguous, which can be defined by Equation 4.3. The reasoning behind this metric is that, being a minimization problem, the assignment of a null distance between each pair of adjacent parcels prioritizes the formation of contiguous holdings.

$$d_{ik} = ||p_i - p_k|| \quad \text{for } \; i < k \; < n \tag{4.2}$$

$$d_{ik} = (1 - \gamma_{ik})\,||p_i - p_k|| \quad \text{for } \; i < k \; < n \tag{4.3}$$

Both metrics were combined with the SPD and APDa formulations, and tested in the instance for a time limit of three hours. The resulting reassignments can be visualized in Figure 4.5. The results were also translated to values to better assess them. Table 4.12 presents the best objective of each solution, the sum and average pairwise distances calculated for each configuration, as well as the number of pairs of contiguous parcels included in the same holding.

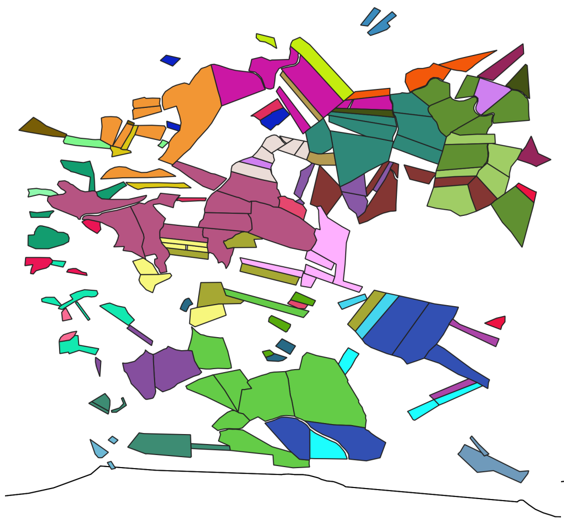
Evidently, the objective function of the SPD - Euclidean formulation corresponds to the measured SPD. As for the APDa formulation, the O.F. value of 5435 is very close to the real value of

Table 4.12: Comparison between Euclidean and modified metrics, combined with SPD and APDa, for 3h, for 159-42 Ribeira de Fráguas instance.

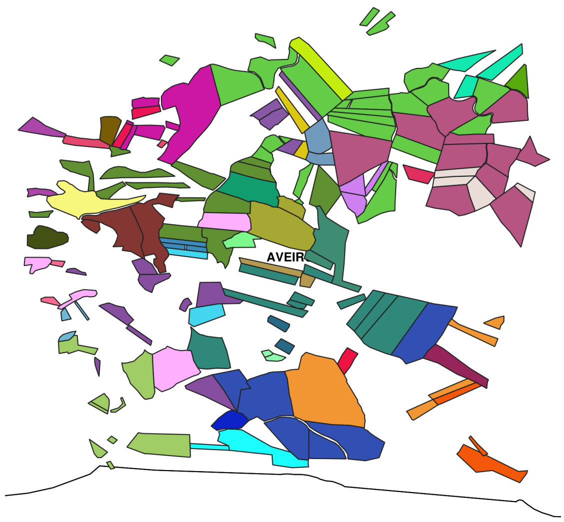
Formulation	O.F.	SDP	ADP	NC
Current	-	310211	18528	9
SPD - Euclidean	65139	65139	9507	49
APDa - Euclidean	5435	184710	5540	44
SPD - Modified	50175	56588	7856	67
APDa - Modified	3527	164075	4816	66

the average pairwise distance, of 5540. This difference is approximately 1.9% of the real value, once again validating the approximation.

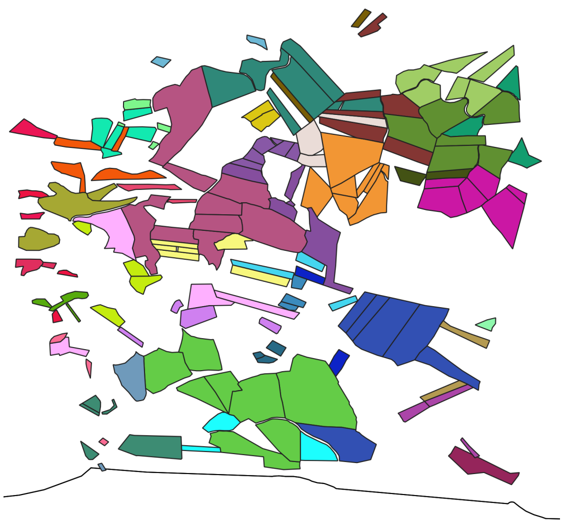
Regarding contiguity, the modified metric to represent these relationships proved to be effective in promoting the inclusion of pairs of adjacent parcels in the same holdings. For both the SPD and APDa criteria, the number of pairs of adjacent parcels belonging to the same holding increased by 18 and 22, respectively. Analyzing the SPD and APD values, which were calculated equally for all results, using the Euclidean distances between the centroids of the plots, it was verified that the solutions of the modified similarity metric had lower values compared to their Euclidean metric counterparts. Thus, the inclusion of information regarding the adjacency between parcels in the similarity metric not only promoted the creation of holdings with more contiguous areas, but also improved the algorithm's efficiency, reaching better solutions.



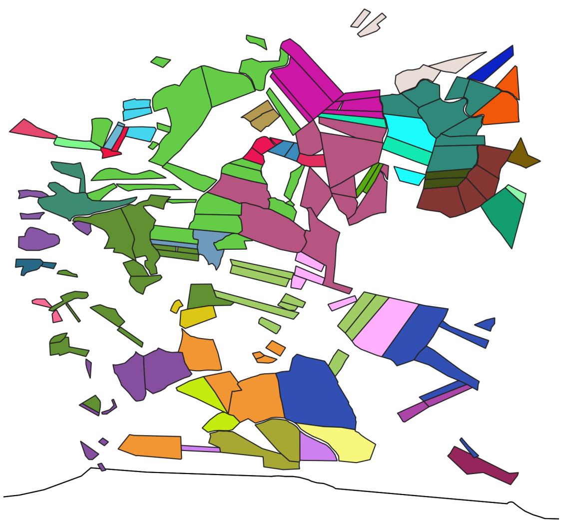
(a) Solver results for SPD and Euclidean metric



(b) Solver results for APDa and Euclidean metric



(c) Solver results for SPD and modified metric



(d) Solver results for APDa and modified metric

Figure 4.5: Results of 159-42 Ribeira de Fráguas instance, for SPD and APDa land-exchange, for different similarity metrics, 3h run-time.

Chapter 5

Conclusion and future works

This dissertation presented the problem of land-exchange, a voluntary form of land consolidation, that aims to address the issues resulting from land fragmentation in rural environments. Special attention was given to the particularities of fragmented Portuguese forests, namely legal and environmental matters that restrict the uses of the parcels. The main objective was to minimize the distances between parcels belonging to the same holding, preferably forming contiguous areas, by reassigning the parcels, while meeting landowners' preferences and the specific range of values for each holding, according to their initial evaluation. Three different criteria were suggested for minimizing the distances between parcels, and they were the sum of the pairwise distances (SPD), the sum of the average pairwise distance of each holding (APD), and the sum of the squared distances to the center of each holding (SSD). Each problem was formally defined, the SPD as a Zero-One Quadratic Program, and both APD and SSD as Nonlinear Programming. A linearization of the SPD was also formulated as well as a quadratic approximation of the APD (APDa). All models were implemented and studied using a state-of-the-art solver.

In a first phase, the computational experiments were performed by solving uniform instances, with square, equal and evenly spaced parcels. The performance of the solver revealed the computational complexity of the land-exchange problem. The linear SPD was the only formulation capable of achieving optimal results for instances up to 64 parcels and 4 owner and a gap very close to 0% for the 81-5 instance. However, for larger instances, the performance is very degraded, and the solver is unable to present consolidated holdings. This behavior is assumed as an outcome of the increased size of the problem resulting from the addition of linearization variables and constraints. Since the land-exchange problem presents more opportunities for improvement for larger instances, the proposed linearization is inadequate. For the quadratic SPD, APDa and SSD criteria, it was only possible to reach the optimal solution for the smallest instance with 25 parcels and 4 owners, which, for the first two referred criteria, was achieved within 10 seconds and, for the latter, in about 25 minutes. For the APD and for the same instance, Gurobi failed to determine the optimal solution within the 3 hour time limit, which reveals its increased complexity. Furthermore, for larger instances, the solver failed to produce incumbents, and for this reason this formulation was discarded. As for the results of SPD, APDa and SSD for larger instances, although Gurobi was not capable of reaching optimality, the solutions obtained present quality, as holdings, with few exceptions, are compact and contiguous. The metrics used to assess the quality of the solution, the reduction of the sum and the average pairwise distances over the initial distribution, verify the improvements resulting from the application of the three models.

Comparing the SPD, APDa and SSD formulations, it was found that, expectably, the SPD presents the best results for the metric of the sum of the pairwise distances. Yet, the visual results indicate that the holdings obtained by this criterion are not the best, as holdings are not entirely contiguous, which suggests that this metric is not the best to promote nor to assess the reduction of fragmentation. The performance of the solver for these three models was further compared with the genetic algorithm developed by Teijeiro et al. (2020), and the resolution of the APDa and SSD models through the solver presented competitive performances.

Regarding the tests performed on instances of real proportions, only spatial issues were addressed and analyzed, as other information on owners' preferences was lacking. Gurobi was incapable of handling an instance formed by the remaining 944 parcel and 85 owners after eliminating uninteresting holdings for the exchange, since it exhausted the available memory. A smaller instance, formed by the 20 holdings with the highest number of parcels, was created and was solved with the SPD, APDa and SSD models. The SPD and APDa formulations presented improvements over the current land tenure structure. A more detailed analysis of the changes of each holding revealed that for the SPD, one of the holdings suffered an increase in the average distance between parcels, which could compromise the acceptance of the solution. The APDa, on the other hand, managed to provide improvements in the average distances of all holdings, whilst keeping the numbers of parcels attributed to each holding close to current values. This feature, although it seems to be advantageous for the instance in question, should be analyzed in more detail and for several parcel layouts with different characteristics. Regarding the SSD formulation applied to the reality-based instance, the solver solution, after 3 hours of optimization, did not produce any improvement. On the opposite, it caused an increase in the average distance between plots of about 15%. Thus, despite its effectiveness in regular instances, the SSD model is not suitable for reality-based instances, and therefore, neither for the land-exchange problem.

A last computational experiment was carried out to verify the contribution of the similarity metric in the results of the similarity-based models. For this purpose, a region with high land coverage and contiguous areas was chosen, which happened to have a high number of owners and holdings composed of a single parcel, which is not ideal for land-exchange. The Euclidean distance metric was compared with a distance modified to be null for adjacent plots, and the use of the latter not only increased the number of pairs of contiguous plots included in the same holding but also surpassed the performance of the models with the Euclidean metric, since it reached solutions with lower values for this measure.

It is acknowledged that the tests performed on instances with real features are insufficient to draw definitive conclusions. Even so, the APDa model, followed by the SPD, present themselves as promising tools for small problems of voluntary land consolidation. However, they are not exempt from a more in-depth study of the impact of each criterion at the level of holdings. The features of the models related to the owners' preferences, with regard to maintaining parcels and ensuring compatibility, were only tested to ensure their functioning, however their impact on the quality of the solution was not evaluated, as it should be. Including, studying and validating these factors implies communication with the landowners.

On the other hand, both models are similarity-based and allow the integration of other information beyond spatial data, such as soil characteristics and use, which can contribute to the formation of contiguous areas of the same land-use, and create economies of scale. In this way, the land-exchange problem might benefit from exploring this similarity metric.

A better integration with a GIS, a recognized difficulty common to spatial optimization problems, is also needed to visualize and manipulate data more easily. Likewise, uncertainties and errors in data models, which are to be expected, should be anticipated and taken care of.

Furthermore, although the models developed are able to provide solutions that represent an improvement in the general land tenure structure for small problems, the extreme fragmentation found in the Portuguese landscape demands more from these models. Thus, either an effective complementary mechanism is developed for choosing the owners whose participation in the land exchange enhances opportunities for improvements in the land tenure structure, or the algorithm's capacity is increased. For the latter, in order to be able to solve larger instances within a reasonable time, perhaps it would be beneficial to resort to heuristic or even hybrid approaches.

Bibliography

Akkus, M. A., Karagoz, O., and Dulger, O. (2012). Automated land reallotment using genetic algorithm. *INISTA 2012 - International Symposium on INnovations in Intelligent SysTems and Applications*, (July 2012).

Alegria, C., Pedro, N., do Carmo Horta, M., Roque, N., and Fernandez, P. (2019). Ecological envelope maps and stand production of eucalyptus plantations and naturally regenerated maritime pine stands in the central inland of portugal. *Forest Ecology and Management*, 432:327–344.

Aloise, D., Hansen, P., and Liberti, L. (2012). An improved column generation algorithm for minimum sum-of-squares clustering. *Mathematical Programming*, 131(1-2):195–220.

Ayranci, Y. (2007). Re-allocation aspects in land consolidation: A new model and its application.

Backman, M. and Österberg, T. (2004). Land Consolidation in Sweden. pages 1–13.

Beighley, M. and Hyde, A. C. (2018). Portugal Wildfire Management in a New Era Assessing Fire Risks , Resources and Reforms Report Basis : Why Listen to Us ? Because You ' re Listening to Them. (February).

Bentley, J. W. (1990). Wouldn't you like to have all of your land in one place? land fragmentation in Northwest Portugal. *Human Ecology*, 18(1):51–79.

Borgwardt, S., Brieden, A., and Gritzmann, P. (2011). Constrained Minimum-k-Star Clustering and its application to the consolidation of farmland. *Operational Research*, 11(1):1–17.

Borgwardt, S. and Schmiedl, F. (2014). Threshold-based preprocessing for approximating the weighted dense k-subgraph problem. *European Journal of Operational Research*, 234(3):631–640.

Brieden, A. and Gritzmann, P. (2004). A Quadratic Optimization Model for the Consolidation of Farmland by Means of Lend-Lease Agreements. *Operations Research Proceedings*, pages 324–331.

Cay, T. and Iscan, F. (2006). Optimization in Land Consolidation. pages 1–11.

Cay, T. and Iscan, F. (2011). Fuzzy expert system for land reallocation in land consolidation. *Expert Systems with Applications*, 38(9):11055–11071.

Cay, T. and Uyan, M. (2013). Evaluation of reallocation criteria in land consolidation studies using the analytic hierarchy process (ahp). *Land Use Policy*, 30(1):541–548.

Church, R. L. (1999). Location modelling and GIS. *Geographical Information Systems 2nd Edition*, pages 293–303.

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

Dao, T. B. H., Duong, K. C., and Vrain, C. (2017). Constrained clustering by constraint programming. *Artificial Intelligence*, 244:70–94.

Demetriou, D. (2014). *The Development of an Integrated Planning and Decision Support System (IPDSS) for Land Consolidation*.

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

Demetriou, D., Stillwell, J., and See, L. (2013b). A new methodology for measuring land fragmentation. *Computers, Environment and Urban Systems*, 39:71–80.

ENF (2015). Resolution of the Council of Ministers no. 6-B/2015.

Ertunç, E., Çay, T., and Haklı, H. (2018). Modeling of reallocation in land consolidation with a hybrid method. *Land Use Policy*, 76(July 2017):754–761.

FAO (2003). *The design of land consolidation pilot projects in Central and Eastern Europe*.

Fernandes, P. M., Loureiro, C., Guiomar, N., Pezzatti, G. B., Manso, F. T., and Lopes, L. (2014). The dynamics and drivers of fuel and fire in the Portuguese public forest. *Journal of Environmental Management*, 146:373–382.

Ge, R. (2008). *Clustering with cluster-level constraints*. PhD thesis.

Gniadek, J., Harasimowicz, S., Janus, J., and Pijanowski, J. M. (2013). Optimization of the parcel layout in relation to their average distance from farming settlements in the example of Mściwojów village, Poland. *Geomatics, Landmanagement and Landscape*, 2(February 2016):25–35.

Guns, T., Dao, T. B. H., Vrain, C., and Duong, K. C. (2016). Repetitive Branch-and-Bound using constraint programming for constrained minimum sum-of-squares clustering. *Frontiers in Artificial Intelligence and Applications*, 285:462–470.

Harasimowicz, S., Janus, J., Bacior, S., and Gniadek, J. (2017). Shape and size of parcels and transport costs as a mixed integer programming problem in optimization of land consolidation. *Computers and Electronics in Agriculture*, 140:113–122.

Hirsch, F. and Schmithüsen, F. J. (2010). Private forest ownership in Europe. *ETH Zurich Research Collection*.

ICNF (2018). Perfil Florestal - Portugal. pages 1–4.

Jesus, S.-M.-A., Tracy, D., Roberto, B., Giorgio, L., Alfredo, B., Daniele, D. R., Davide, F., Pieralberto, M., Tomas, A. V., Hugo, C., and Fabio, L. (2019). *Advance EFFIS report on Forest Fires in Europe, Middle East and North Africa 2019*. Number March.

Latruffe, L. and Piet, L. (2014). Does land fragmentation affect farm performance? A case study from Brittany, France. *Agricultural Systems*, 129:68–80.

Lin, D., Wambersie, L., Wackernagel, M., and Hanscom, P. (2020). Calculating Earth Overshoot Day 2020: Estimates Point to August 22nd. Technical report.

Lu, H., Xie, H., He, Y., Wu, Z., and Zhang, X. (2018). Assessing the impacts of land fragmentation and plot size on yields and costs: A translog production model and cost function approach. *Agricultural Systems*, 161(January):81–88.

Peng, J. and Xia, Y. (2005). A cutting algorithm for the minimum sum-of-squared error clustering. In *Proceedings of the 2005 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics.

Rembold, F. (2003). Land fragmentation and its impact in Central and Eastern European countries and the Commonwealth of Independent States.

Sequeira, C. R., Montiel-Molina, C., and Rego, F. C. (2019). Historical fire records at the two ends of Iberian Central Mountain System: Estrela massif and Ayllón massif. *Investigaciones Geograficas*, (72):31–52.

Sil, A., Fernandes, P., Rodrigues, A. P., Alonso, J., Honrado, J., Perera, A., and Azevedo, J. (2019). Farmland abandonment decreases the fire regulation capacity and the fire protection ecosystem service in mountain landscapes. *Ecosystem Services*, 36:100908.

Teijeiro, D., Corbelle Rico, E., Porta, J., Parapar, J., and Doallo, R. (2020). Optimizing parcel exchange among landowners: A soft alternative to land consolidation. *Computers, Environment and Urban Systems*, 79(September 2019):101422.

Tourino, J., Parapar, J., Doallo, R., Boullón, M., Rivera, F. F., Bruguera, J. D., González, X. P., Crecente, R., and Álvarez, C. (2003). A GIS-embedded system to support land consolidation plans in Galicia. *International Journal of Geographical Information Science*, 17(4):377–396.

Van Dijk, T. (2004). Land consolidation as Central Europe ' s panacea reassessed. pages 1–21.

Veršinskas, T., Vidar, M., Hartvigsen, M., Arsova, K. M., van Holst, F., and Gorgan, M. (2020). *Legal guide on land consolidation*.

Zhu, Z., Xu, Z., Shen, Y., and Huang, C. (2020). How forestland size affects household profits from timber harvests: A case-study in China's Southern collective forest area. *Land Use Policy*, 97(May 2018):103380.

Appendix A

Ribeira de Fráguas land tenure

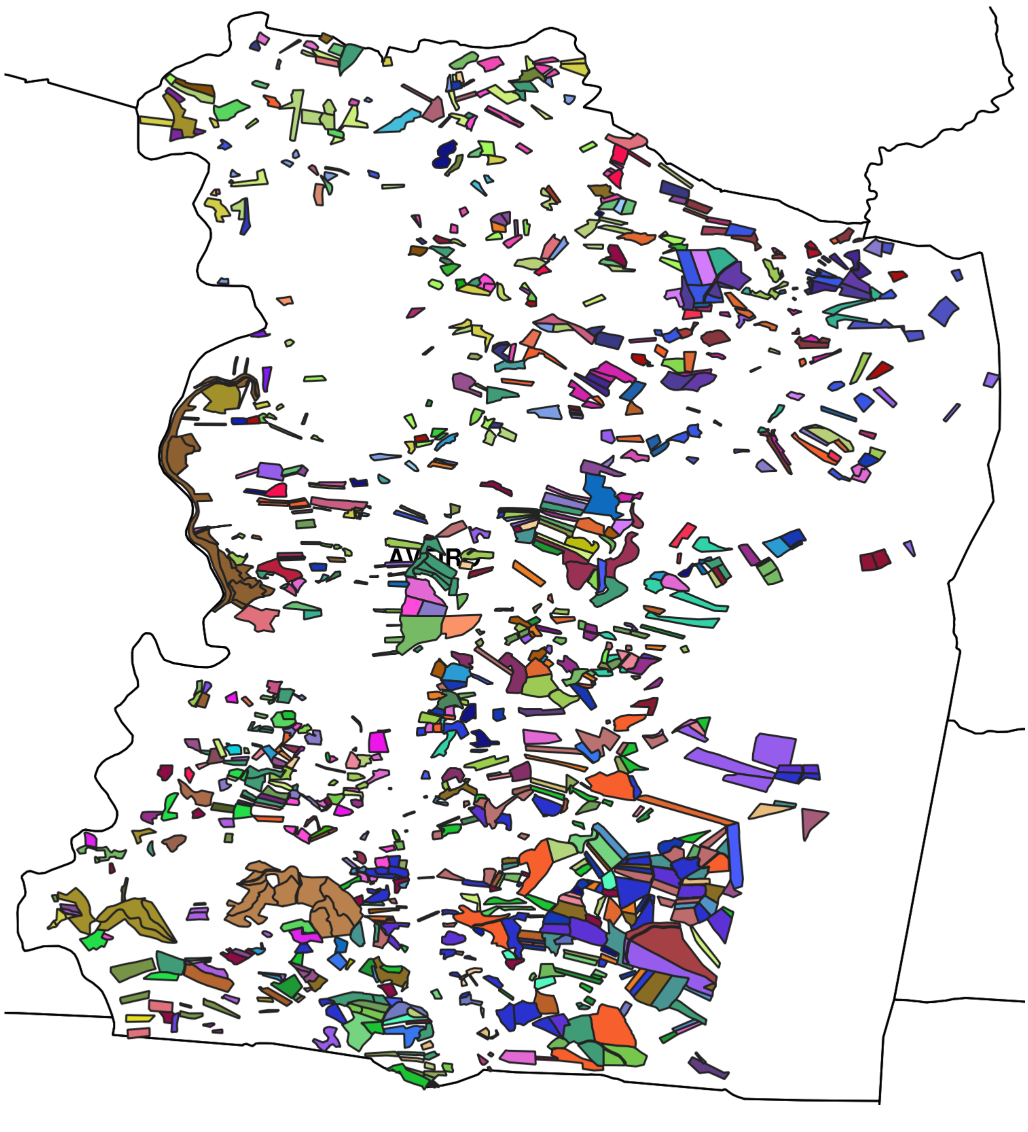
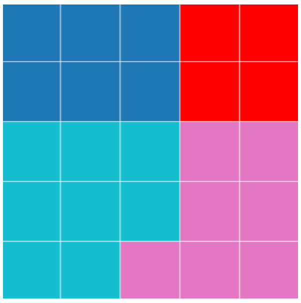


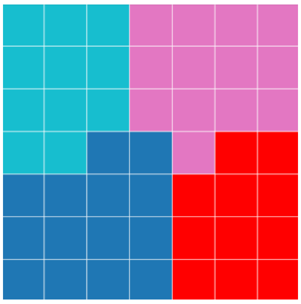
Figure A.1: Incomplete land tenure structure of the parish of Ribeira de Fráguas with 1057 parcels and 145 owners.

Appendix B

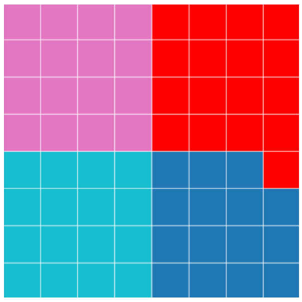
Visual results of solver solutions



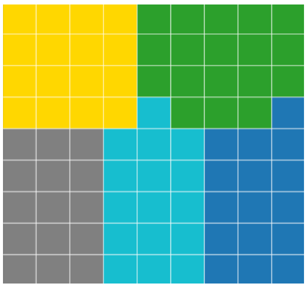
(a) 25 - 4



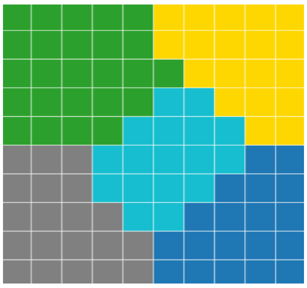
(b) 49 - 4



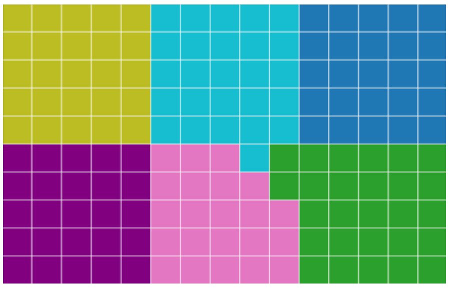
(c) 64 - 4



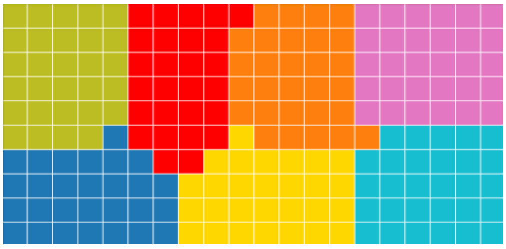
(d) 81 - 5



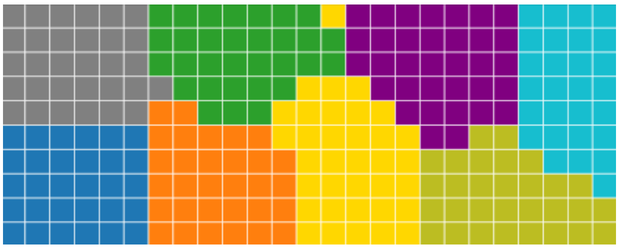
(e) 100 - 5



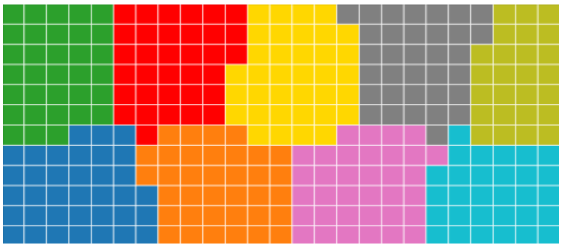
(f) 150 - 6



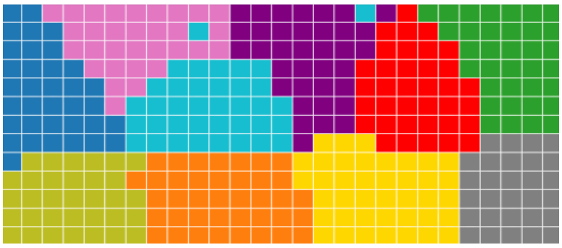
(g) 200 - 7



(h) 250 -8

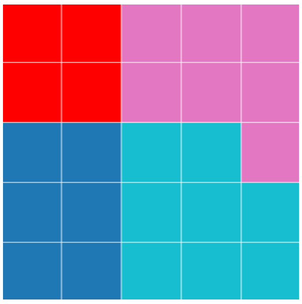


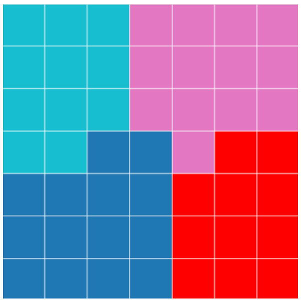
(i) 300 - 9

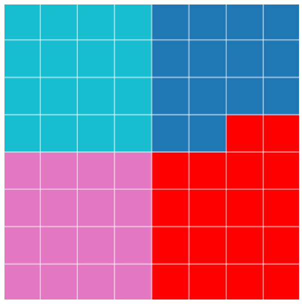


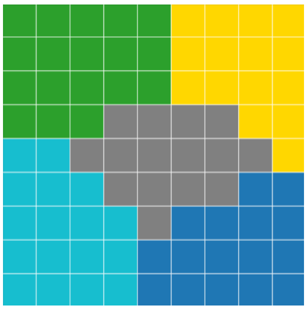
(j) 351 - 10

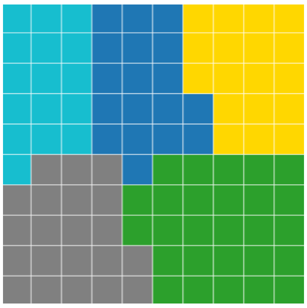
Figure B.1: Results for SPD, $\theta = 10\%$, 3h run-time.

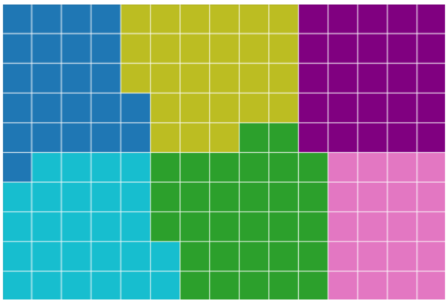

(a) 25 - 4

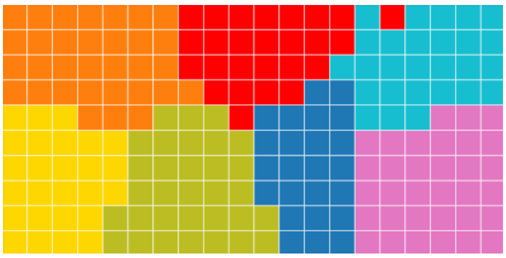

(b) 49 - 4

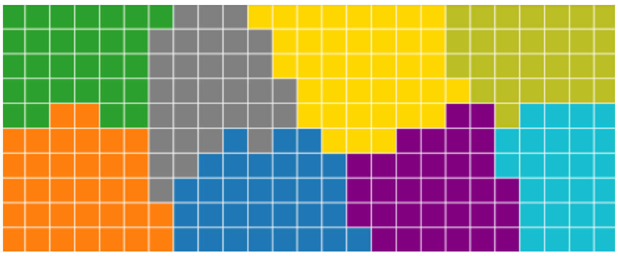

(c) 64 - 4

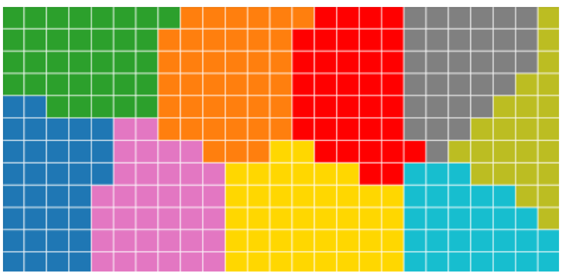

(d) 81 - 5


(e) 100 - 5


(f) 150 - 6


(g) 200 - 7


(h) 250 -8


(i) 300 - 9

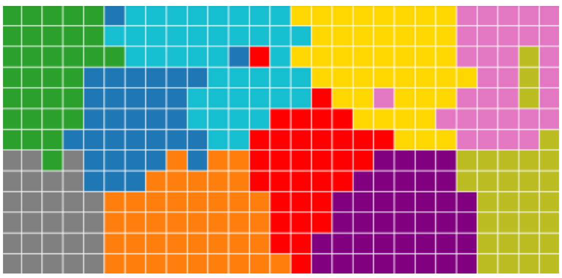
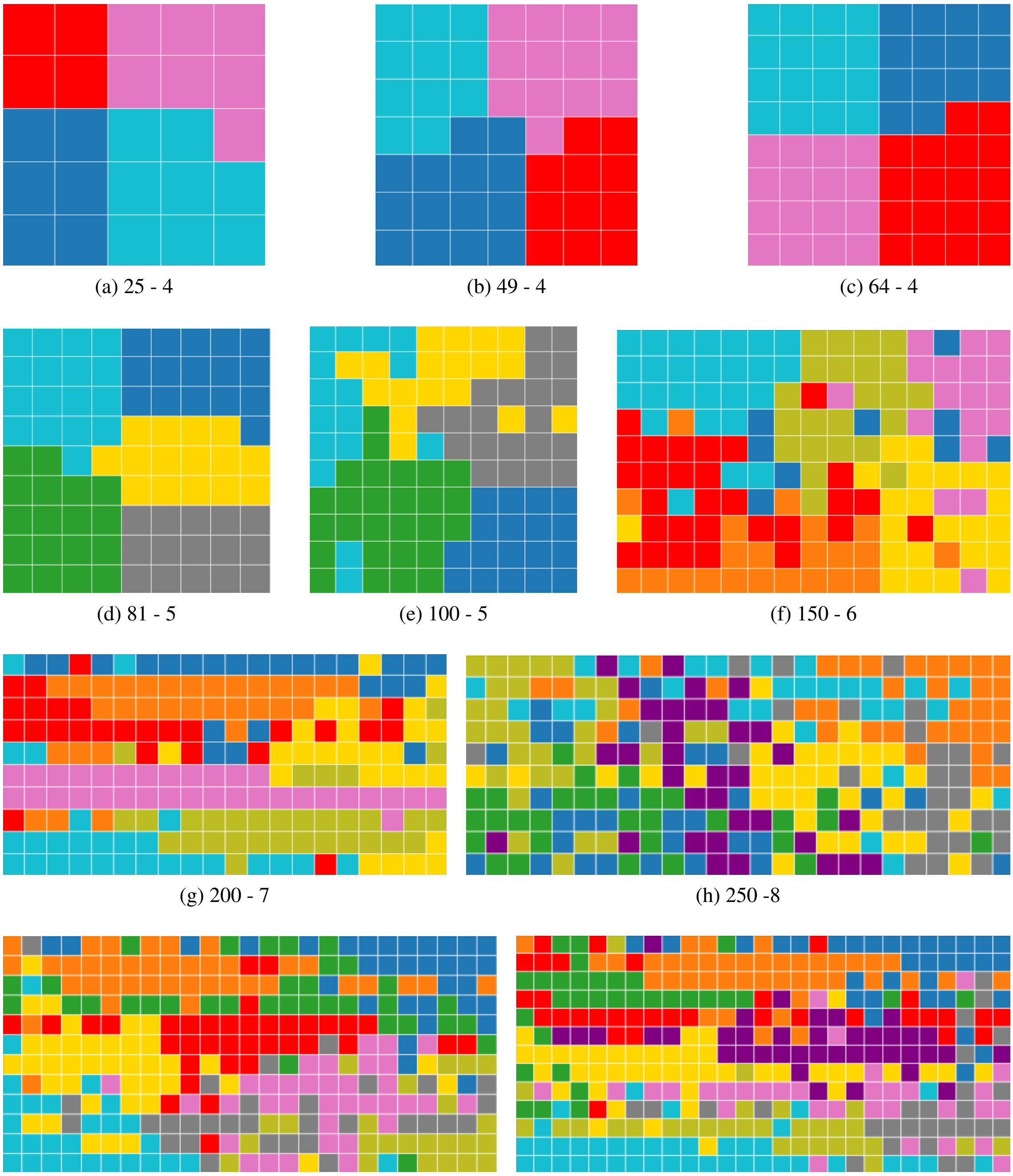

(j) 351 - 10

Figure B.2: Results for SPD, $\theta = 0\%$, 3h run-time.



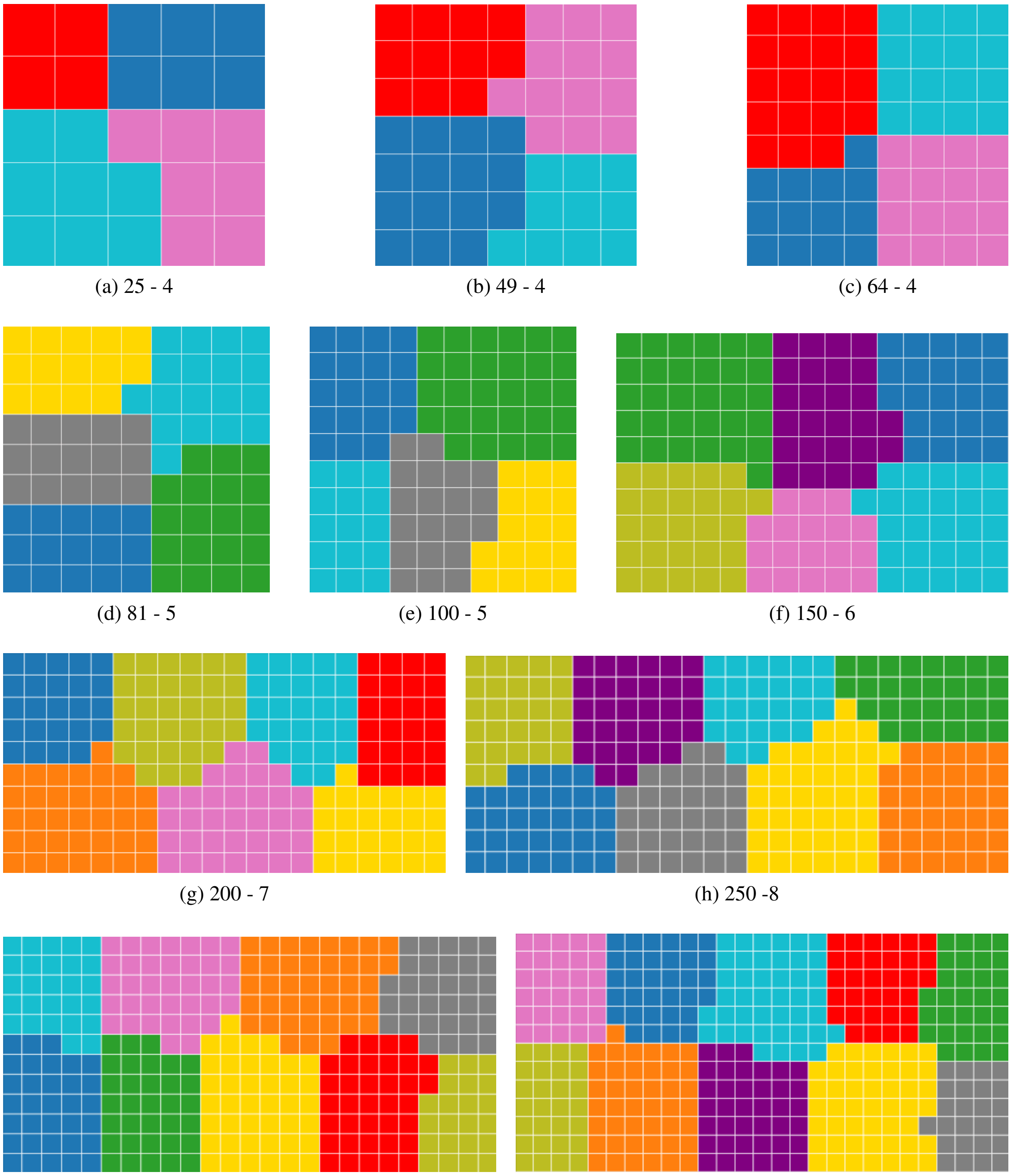
(a) 25 - 4 (b) 49 - 4 (c) 64 - 4

(d) 81 - 5 (e) 100 - 5 (f) 150 - 6

(g) 200 - 7 (h) 250 -8

(i) 300 - 9 (j) 351 - 10

Figure B.3: Results for linear SPD, $\theta = 0\%$, 3h run-time.



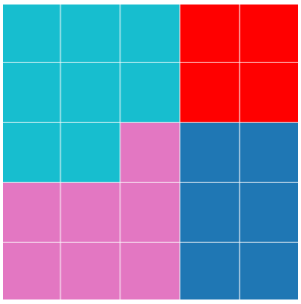
(a) 25 - 4 (b) 49 - 4 (c) 64 - 4

(d) 81 - 5 (e) 100 - 5 (f) 150 - 6

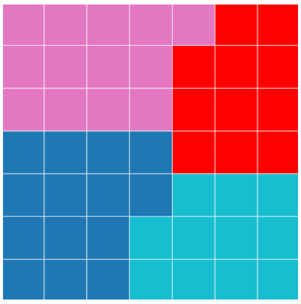
(g) 200 - 7 (h) 250 -8

(i) 300 - 9 (j) 351 - 10

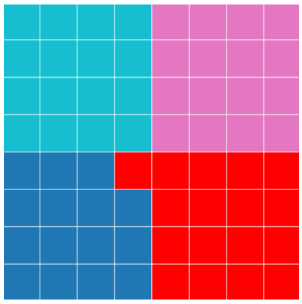
Figure B.4: Results for APDa, $\theta = 10\%$, 3h run-time.



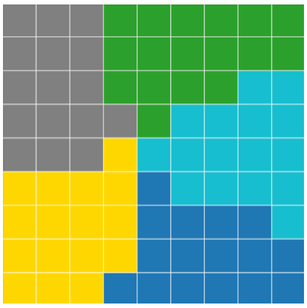
(a) 25 - 4



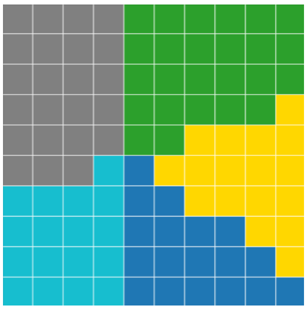
(b) 49 - 4



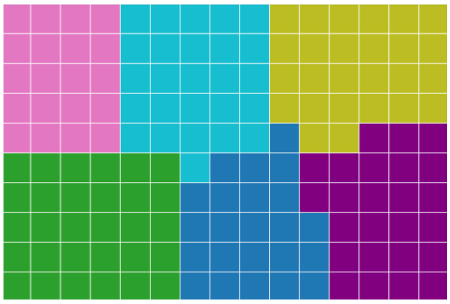
(c) 64 - 4



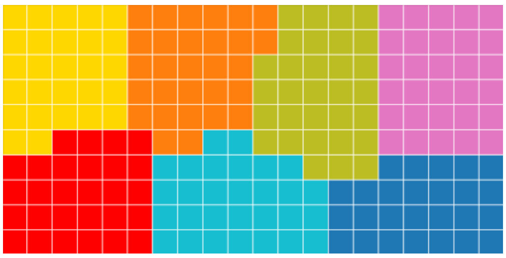
(d) 81 - 5



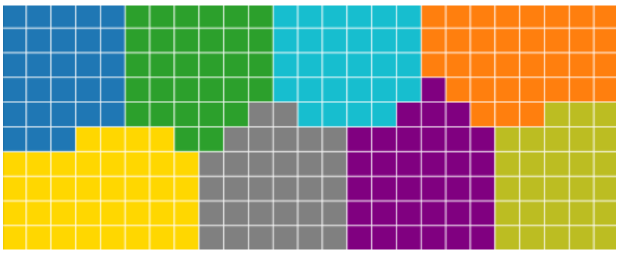
(e) 100 - 5



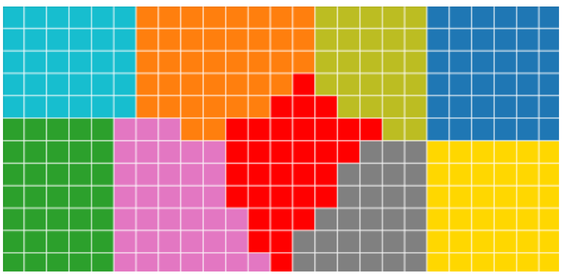
(f) 150 - 6



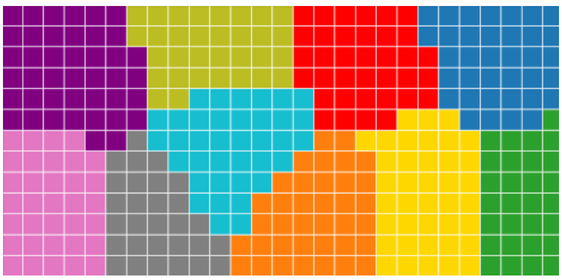
(g) 200 - 7



(h) 250 -8

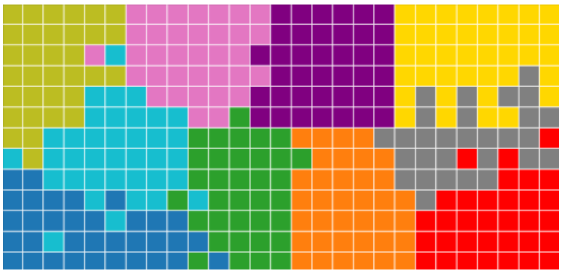


(i) 300 - 9

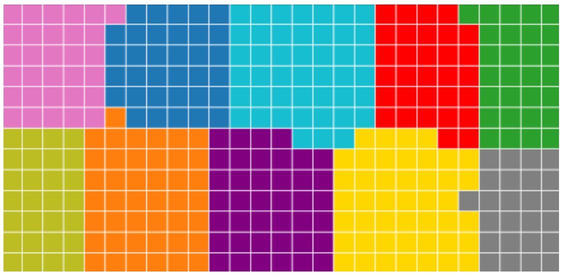


(j) 351 - 10

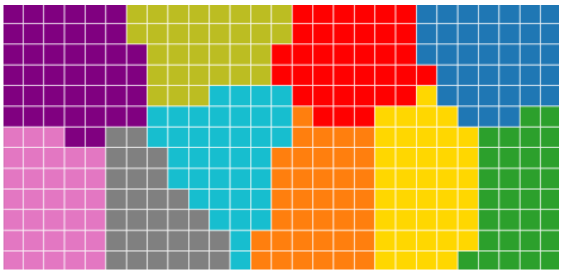
Figure B.5: Results for SSD, $\theta = 10\%$, 3h run-time.



(a) SPD



(b) APDa



(c) SSD

Figure B.6: Benchmark instance results for 15 minutes run-time.