UNIVERSITY
INSTITUTE
OF LISBON

A Decision Support System for Land Consolidation: Graph Modeling and Multi-Objective Optimization

Miguel de Carvalho Valentim

Master in Integrated Business Intelligence Systems

Supervisor:
PhD Vítor Manuel Basto Fernandes, Associated Professor with Aggregation,

ISCTE - Instituto Universitário de Lisboa

September, 2025

TECHNOLOGY
AND ARCHITECTURE

Departamento de Ciências e Tecnologias da Informação

A Decision Support System for Land Consolidation: Graph Modeling and Multi-Objective Optimization

Miguel de Carvalho Valentim

Master in Integrated Business Intelligence Systems

Supervisor:
PhD Vítor Manuel Basto Fernandes, Associated Professor with Aggregation,

ISCTE - Instituto Universitário de Lisboa

September, 2025

To everyone who asked, 'How's the thesis going?' – this is for you. It is over.

Acknowledgments

These past two years have been long and demanding, and I am certain that it would not have been possible to write this document and complete this master's degree without the support and contributions of several people who accompanied me along the way.

To my supervisor, Professor Vítor Manuel Basto Fernandes, I express my deepest gratitude for his guidance, for his availability even beyond the original deadlines, and for the valuable suggestions that proved essential to the completion of this work.

To my colleagues and friends, Catarina Pontes and João Mota, who not only marked my undergraduate years but also made this master's journey lighter and more rewarding. Without your presence and support, this path would have been much harder.

To my family, my sincere thanks for the unwavering encouragement and for instilling in me the resilience that was decisive in times of difficulty.

To my father and brother, thank you for your unconditional support and for always being by my side. Knowing that I could count on you made everything easier.

Finally, to the three most important women in my life, my mother, Diana, and my grandmother, I express my heartfelt gratitude for the strength, determination, and perseverance you have always transmitted to me. Thank you for reminding me that every path must be followed to the end and, above all, for never letting me forget that I was not raised to give up.

Resumo

A fragmentação fundiária constitui um desafio persistente em Portugal, onde muitas propriedades rurais são compostas por parcelas pequenas e dispersas. Esta dissertação desenvolve um sistema de apoio à decisão para orientar processos voluntários de emparcelamento e apoiar iniciativas públicas em curso. O estudo segue uma abordagem estruturada de ciência de dados, utilizando informação geoespacial da Madeira para simular diferentes cenários de propriedade. As parcelas foram agrupadas em blocos maiores, e foi concebido um algoritmo para sugerir trocas que aumentam a contiguidade e reduzem a fragmentação, mantendo as perdas globais de área em níveis residuais.

Os resultados de três cenários mostraram melhorias consistentes: a dimensão média das explorações aumentou cerca de 11%, o número de parcelas descontínuas por proprietário foi praticamente reduzido para metade e os custos do emparcelamento foram distribuídos de forma equilibrada entre os participantes. Estes resultados indicam que métodos baseados em dados podem gerar benefícios concretos, mesmo em contextos rurais complexos.

Para além do contributo técnico, o sistema destaca-se pela transparência e interpretabilidade: cada passo do processo de consolidação pode ser explicado e verificado, o que o torna adequado para contextos de política pública. As conclusões oferecem apoio direto a programas como o *Emparcelar para Ordenar*, sugerindo que ferramentas algorítmicas podem reforçar a equidade, a eficiência e a confiança nos processos de consolidação fundiária.

Palavras-chave: fragmentação fundiária; emparcelamento; apoio à decisão; desenvolvimento rural; Portugal.

Abstract

Land fragmentation is a persistent challenge in Portugal, where many rural properties consist of small and scattered parcels. This thesis develops a decision support system to guide voluntary land consolidation and support ongoing public initiatives. The study follows a structured data science process that applies Madeira geospatial data to simulate different ownership scenarios. The land parcels were grouped into larger blocks and an algorithm was designed to suggest exchanges that increase contiguity and reduce fragmentation while keeping the overall loss of land minimal.

Experiments with three scenarios showed consistent improvements: the average landholding size increased by about 11%, the number of disconnected parcels per owner was nearly halved and the consolidation costs were fairly distributed among participants. These results indicate that data-driven methods can produce tangible benefits, even in complex rural landscapes.

Beyond the technical contribution, the system emphasizes transparency and interpretability: Each consolidation step can be explained and verified, making the approach suitable for policy contexts. The findings provide concrete support for programs such as *Emparcelar para Ordenar*, suggesting that algorithmic tools can enhance fairness, efficiency, and trust in land consolidation policies.

Keywords: land fragmentation; land consolidation; decision support; rural development; Portugal.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

This first chapter provides an overview of the study, highlighting the objectives, the need, and the methodology adopted for the investigation.

1.1. Contextualization

Territorial management has become increasingly central to policy and planning efforts, particularly in countries facing demographic, economic, and environmental challenges. In Portugal, territorial organization has evolved in response to the need for better regional balance, efficient resource use, and sustainable rural development. This evolution was marked by the consolidation of national planning frameworks, such as the *Programa Nacional da Política de Ordenamento do Território* (PNPOT), which laid the foundations for modern spatial planning [20].

In recent decades, the growing complexity of land-related challenges such as depopulation of the interior, abandonment of agricultural land, and uncoordinated urban expansion has highlighted structural problems such as land fragmentation and the absence of a complete cadastral system. These issues have hindered investment, reduced agricultural productivity, and made governance more difficult.

To respond to these constraints, geospatial technologies have gained a prominent role in the governance of land and territory. Geographic Information Systems (GIS), which emerged internationally in the 1960s and later expanded in application across Europe, have become indispensable tools in land management. These systems allow for the integration and analysis of spatial data, enabling planners and policy makers to simulate development scenarios, identify areas of under-utilization, and plan more rational land allocations [24].

In the Portuguese context, the adoption of GIS and other spatial decision-support technologies has become especially relevant given the widespread lack of reliable cadastral records and the pressing need for more efficient land use. These tools, coupled with recent national initiatives such as BUPi and the *Emparcelar para Ordenar* program, provide an opportunity to address historical inefficiencies and implement intelligent data-driven territorial planning strategies.

1.2. Motivation

The motivation behind the execution of this project arises from both urgent territorial challenges and new technological and institutional opportunities.

(1) Overcoming Fragmentation and Under-Utilization: Two-thirds of the land in Portugal is wild or underutilized rural land, generating significant social, economic, and environmental challenges. In particular, the extreme fragmentation of rural land ownership, where landowners often manage many small, scattered parcels, creates structural barriers to productivity and hinders sustainable development [10]. Reducing the number of disconnected land `clusters` per owner and promoting contiguous holdings is critical for improving land use efficiency.

(2) Capitalizing on Policy Momentum: Recent public policies have shown a strong commitment to tackling land-related issues. Initiatives such as the cadastral mapping of rural properties (BUPi – Balcão Único do Prédio) and Decree-Law No. 29/2020, which established the *Emparcelar para Ordenar* program, provide a legal and institutional framework to consolidate fragmented plots, increase the size and continuity of landholdings, and enhance their economic sustainability. These measures aim to build more resilient rural territories through ecological and economic revitalization [13].

(3) Leveraging Technological Innovation: Advances in Geographic Information Systems (GIS), artificial intelligence, and optimization algorithms have created new avenues for modeling and addressing land fragmentation. By treating consolidation as a computational problem, it becomes possible to simulate property exchanges, optimize spatial configurations, and generate actionable strategies for decision makers. This project positions itself at the intersection of technology and policy, reflecting global trends toward smart land management systems [5].

(4) Encouraging Sustainable and Inclusive Growth: The project is guided by broader goals of promoting sustainability and socioeconomic equity in rural regions. By algorithmically increasing the average land area per owner, reducing land fragmentation, and minimizing net area losses during exchanges, the solution enables more viable and efficient agricultural practices. This technical approach directly supports the revitalization of rural territories and aligns with the principles of territorial cohesion and sustainable development of the European Union [23].

This project integrates computational methodologies with policy analysis to create a robust decision support system. The ultimate goal is to provide actionable solutions that address land fragmentation through optimized, fair ,and sustainable land consolidation strategies, contributing to a more intelligent and resilient territorial management model in Portugal.

1.3. Objectives

Smart land management is a crucial response to the interconnected challenges facing societies around the world today, from urban pressures and climate change to environmental degradation and rural depopulation. One of the most critical components of this effort is the ability to reorganize property boundaries in a rational and optimized way. Rather than arbitrarily reallocating land, this project uses computational tools to facilitate voluntary land consolidation, guided by spatial proximity, parcel area, and potential gains in productive efficiency.

This research investigates the opportunities and trade-offs of applying optimization techniques to support smart territorial governance. Particular attention is given to the use of heterogeneous data sources and the implications for data privacy, especially in light of GDPR compliance. The ultimate goal is to provide applied knowledge and technological frameworks that can inform land use policies and decision-making.

The proposed system is built using data from the public platform *Sistema de Identificação Parcelar* (iSIP), made available by the *Instituto de Financiamento da Agricultura e Pescas* (IFAP), which serves as a central authority for the management of geospatial and cadastral data in Portugal.

1.3.1. Specific Objectives

(1) **Evaluate Land Consolidation Potential:** Quantify the potential to increase the average land area per owner and reduce land fragmentation (number of `clusters` per owner) through optimized property exchanges in a target region.
(2) **Develop Optimization Algorithms:** Design and implement a computational model based on geospatial data and parcel adjacency graphs, capable of performing iterative, multi-objective land consolidation that maximizes spatial efficiency and minimizes owner loss.
(3) **Case Study and Territorial Focus:** Apply the proposed methodology to a specific territorial unit (a parish in Madeira) to demonstrate feasibility and discuss how similar models could be applied to prioritize other regions.
(4) **Validation and Metrics:** Define and apply performance metrics to evaluate the effectiveness of land consolidation outcomes, including the average land area per owner, the net variation in area per owner and the number of contiguous land `clusters` per owner.

1.4. Research Questions

This project addresses the following key research questions, structured around the central goal of developing a data-driven, optimization-based decision support system for land consolidation:

(1) **What is the potential to increase the average land area per owner through optimized land exchanges in a fragmented territory?**
(2) **To what extent can an algorithmic approach reduce land fragmentation—measured by the number of disconnected** `clusters` **per owner, while preserving productive capacity?**
(3) **How can a tailored graph-based optimization algorithm be used to balance competing objectives such as increasing average area, reducing fragmentation, and minimizing net area loss?**
(4) **What metrics best capture the effectiveness and fairness of land consolidation results, and how can they support multi-criteria evaluation of different solutions?**

1.5. Research Methodology

The project followed the CRISP-DM framework, from business understanding to modeling and evaluation, ensuring a structured and iterative development of the land consolidation tool. This approach enabled the transformation of geospatial data into actionable insights through the implementation of a custom optimization algorithm. The methodology involved analyzing spatial parcel data, constructing adjacency relationships, simulating block-to-block exchanges, and evaluating results across multiple metrics such as average land area per owner, number of `clusters`, and net area variation. The full methodological details are presented in Chapter 3.

1.6. Structure of the Project

The project is divided into five main chapters that allow for a structured and coherent evolution of the study.

The first chapter presents the research problem along with the motivation and objectives that guide this work. It introduces the challenges of territorial governance in Portugal and explains how land fragmentation, underutilization, and policy momentum form the basis for the proposed optimization approach.

Chapter 2 discusses the theoretical foundations of the study through a literature review. It explores the evolution of land use planning, cadastral systems, and the emergence of smart land management practices. The chapter also reviews relevant technologies such as GIS and optimization algorithms, as well as national and European policy frameworks that support land consolidation initiatives.

Chapter 3 outlines the adopted methodology and the data sources. Describes the rationale behind the CRISP-DM framework and details each of its six phases, from business understanding to deployment. This chapter also describes the geospatial data used in the study (from iSIP), the construction of adjacency graphs, the assignment of synthetic landowners, and the implementation of the custom algorithm for land consolidation.

Chapter 4 presents the empirical analysis and results of the study. It evaluates the performance of the proposed algorithm in reducing fragmentation, increasing the average land area per owner, and minimizing net area loss. Different execution runs are analyzed to explore tradeoffs between these competing objectives. The results are supported by performance metrics and visualizations, including the construction of a Pareto front that highlights non-dominated solutions.

The final chapter summarizes the key findings, discusses their implications for sustainable land governance in Portugal, and reflects on the contributions of the study to both policy and practice. It also outlines limitations of the current model and offers directions for future work, such as testing in other regions, expanding the optimization criteria, or integrating socioenvironmental variables into the decision support system.

CHAPTER 2

Literature Review

2.1. Territorial Management Overview

2.1.1. Historical Context

Portugal's evolution of territorial management reflects the nation's socioeconomic and environmental priorities. Territorial governance has served as a central pillar of spatial organization, gradually transitioning from localized agrarian land use practices to more structured and centralized planning frameworks shaped by industrialization and urban growth.

As governance matured, public authorities began to assume a greater role in regulating land use, balancing regional disparities, and promoting sustainable development. A landmark in this evolution was the creation of the *Programa Nacional da Política de Ordenamento do Território* (PNPOT), which established strategic guidelines for territorial cohesion and planning at national, regional and local levels [20].

Despite these advances, Portugal's territorial governance has long been constrained by the lack of reliable, complete, and updated cadastral information. The lack of tools to reform land ownership, especially in rural areas, allowed fragmentation to persist unchecked for decades. In many regions, small, scattered land holdings reduced the efficiency of agricultural operations and created barriers to investment, modernization, and coordinated development.

The emergence of initiatives such as BUPi and the growing availability of geospatial technologies have opened new paths to address these legacy issues. For the first time, it has become possible to simulate and evaluate consolidation strategies using accurate spatial data and algorithmic logic. This research builds directly on that transformation, proposing a computational framework that operationalizes land consolidation, offering practical means to confront a structural inefficiency that has long shaped the Portuguese rural landscape.

2.1.2. Present-day Issues in Management of Territory

2.1.2.1. *Uncontrolled Urbanization and Coastal Concentration* This phenomenon, known as coastal urbanization, has characterized urban development in the last century in Portugal, where the population and infrastructure have become intensely concentrated along the western and southern coastal zones. This trend has exacerbated regional disparities, leaving vast inland territories underdeveloped, sparsely populated, and increasingly disconnected from national investment priorities.

Studies, including those by [23], highlight the challenges faced in coastal areas, such as overburdened infrastructure, environmental degradation, housing pressure, and unsustainable land use patterns. At the same time, abandonment of rural areas has led to socioeconomic decline, inefficient land occupation, and greater exposure to ecological risks such as fire and erosion.

This asymmetrical development pattern has also accelerated land fragmentation in rural areas, as demographic shrinkage and inheritance practices divide land into increasingly smaller disconnected plots. The result is a growing divide between hyperconsolidated coastal zones and structurally fragmented interior regions.

This territorial imbalance threatens the goal of national cohesion and calls for differentiated policy tools that address the specific needs of low-density rural regions. In this context, intelligent land management solutions, particularly those that promote land consolidation and efficient use of agricultural space, can serve as targeted interventions to revitalize neglected areas. By optimizing property structures and reducing land fragmentation in the interior, such strategies contribute to rebalancing development dynamics and enabling more equitable use of the national territory.

2.1.2.2. *Territorial Fragmentation and Rural Abandonment* The fragmentation of land, reflected in the division of property into multiple small and spatially dispersed plots, is one of the most pressing structural problems in rural Portugal. This pattern, especially prevalent in agricultural zones, leads to inefficiencies in land use, hinders the adoption of modern production techniques, and prevents economies of scale. Owners often manage multiple noncontiguous parcels, increasing operational costs and reducing productive capacity.

This fragmentation not only affects individual landowners, but also undermines regional planning, environmental sustainability, and agricultural competitiveness. The cumulative impact is a landscape that is difficult to manage, inefficient to exploit, and vulnerable to degradation. In particular, fragmented parcels are often abandoned or underused, contributing to the spread of wild vegetation, fire risk, and the disruption of ecological continuity.

These problems are further exacerbated by demographic pressures, such as aging of the population and sustained emigration from rural areas, which weaken the economic base and social fabric of interior regions [10]. The result is a vicious cycle of abandonment and disinvestment.

From a territorial governance perspective, solving fragmentation requires more than voluntary land exchanges or individual initiatives. It demands structured and scalable strategies capable of reorganizing the land in a way that increases the average area per owner, improves contiguity, and minimizes the loss of productive area. Computational approaches, such as the one proposed in this research, offer the ability to model and optimize such land consolidations, even in complex or partially documented cadastral environments.

2.1.2.3. *Consequences of Not Having a Modern Cadastral System* One of the main structural limitations of territorial governance in Portugal is the absence of a complete and standardized cadastral system. In many rural regions, land boundaries remain imprecise, ownership records are fragmented or outdated, and formal documentation is often nonexistent. These shortcomings introduce legal uncertainty, hinder land transactions, and complicate efforts to implement coordinated spatial planning.

The lack of a reliable register also undermines the delivery of agricultural subsidies, the protection of property rights, and the enforcement of environmental regulations, all of which depend on accurate parcel identification. Furthermore, it weakens the institutional foundation needed for land reorganization initiatives, which require verified ownership structures and precise spatial data.

Although platforms like BUPi are gradually addressing these deficiencies, large parts of the territory, especially in low-density rural areas, still lack adequate infrastructure, digitized records, and citizen participation. As a result, top-down territorial policies often remain disconnected from operational realities on the ground.

However, the growing availability of partial cadastral and geospatial data opens the door to new types of solutions. Computational models, such as the one developed in this research, can simulate land consolidation scenarios using spatial information alone, relying on parcel geometry and adjacency rather than full legal ownership datasets. These tools offer a complementary path forward: helping to identify and prioritize consolidation opportunities even before the cadastral system is fully operational.

2.2. Smart Land Management and Geospatial Technologies

2.2.1. Definition and Scope

Smart land management is a modern integrated approach to land use and governance that uses digital innovation to address long-standing territorial challenges. It is a diverse and interdisciplinary field that combines elements of spatial planning, environmental monitoring, data science, and computational modeling. Its goal is to maximize the utility and sustainability of land resources by enabling more informed, transparent, and data-driven decisions in both urban and rural contexts.

As part of this paradigm, spatial information plays a central role. Geographic Information Systems (GIS), remote sensing technologies, and spatial data infrastructures form the technical foundation of smart land management. These tools allow governments and planners to visualize, analyze, and monitor territorial patterns in real time and at scale. High-performance computing and advanced algorithms further extend these capabilities by enabling the simulation of future scenarios, the detection of inefficiencies, and the optimization of spatial configurations.

A key application of smart land management lies in the addressing the problem of land fragmentation, particularly in rural areas where property division and abandonment are widespread. In such contexts, intelligent tools can be deployed to consolidate fragmented plots, improve the contiguity of landholdings, and enhance productive efficiency. This requires the integration of spatial datasets with algorithmic approaches capable of proposing viable land exchange configurations.

The current research situates itself within this framework by developing a decision support system (DSS) that operates on cadastral and geospatial data to propose optimized land consolidation strategies. It exemplifies how smart land management principles can be operationalized to support rural development, improve territorial governance, and address the critical issue of land fragmentation in Portugal [24].

2.2.2. The Role of Geographic Information Systems (GIS)

Smart land management is largely enabled by Geographic Information Systems (GIS), which provide a digital framework for capturing, storing, analyzing, and displaying spatial data. These systems allow public authorities, researchers, and planners to model and manage the territory with a level of precision and scalability that would otherwise be unfeasible.

Beyond administrative tasks, GIS plays a critical role in land analysis, spatial optimization, and planning. It supports a wide range of advanced geospatial operations, such as:

- Calculating parcel areas directly from geometry.
- Detecting spatial adjacency between parcels using geometric relationships.
- Building spatial networks (graphs) to simulate block-level exchanges between owners.
- Dissolving contiguous plots by owner to compute updated cluster structures and consolidation outcomes.

In the context of this project, GIS was indispensable for transforming raw cadastral geometries into structured inputs for the consolidation algorithm. Geospatial data enabled the creation of adjacency graphs, the identification of contiguous ownership blocks, and the evaluation of consolidation gains in terms of area, connectivity, and efficiency. GIS thus provided the technical infrastructure required to formulate and solve the consolidation problem as a spatial optimization task.

Rather than serving only as a mapping tool, GIS operates here as a core component of the decision support system, integrating geometry, topology, and optimization logic to inform territorial governance in a scalable and data-driven manner [11].

2.2.3. Heterogeneous Data Integration

Data integration across heterogeneous sources is a fundamental requirement for smart land management. It involves combining datasets with varying structures and content, such as cadastral maps, geospatial imagery, property registration records, and environmental data, into a unified and interoperable system. This integration enables more accurate land governance, comprehensive spatial analysis, and effective decision-making.

However, the process presents significant challenges, particularly in rural territories where cadastral records are often incomplete or outdated. Variations in data formats, inconsistencies in spatial resolution, and missing ownership information complicate efforts to create coherent land databases. Platforms like iSIP and BUPi have been instrumental in mitigating these issues in Portugal by applying state-of-the-art geospatial processing techniques and enforcing interoperability standards.

Key steps in this integration process include:

- **Georeferencing:** Aligning cadastral data with high-resolution imagery to accurately define property boundaries.
- **Data Standardization:** Converting heterogeneous formats into common schemas for processing and analysis.

- **Remote Sensing:** Leveraging satellite imagery and aerial photography to fill in missing or outdated spatial data.
- **Dynamic Updates:** Incorporating continuous updates from local actors to reflect changes in land use and ownership in real time.

Support for these processes comes from technologies such as high-definition imaging and machine learning algorithms that enhance spatial precision and processing efficiency. The resulting territorial databases provide reliable input for land policy and planning, supporting initiatives such as the Common Agricultural Policy (CAP) while aligning with broader EU directives [13].

In the context of this research, the project demonstrates that even when ownership data is unavailable or incomplete, effective modeling is still feasible. Using only geospatial parcel boundaries and generating simulated ownership scenarios, the algorithm constructs a functional representation of a suitable land structure for optimization. This reinforces the idea that smart land management systems can operate on partial or evolving datasets, as long as spatial integrity is preserved. Heterogeneous data integration, therefore, is not only a technical goal but also an enabler of resilient and scalable computational solutions for territorial consolidation.

2.2.4. Geospatial Technologies: Their Influence on Territorial Governance

The integration of geospatial technologies into territorial governance has reshaped the way land-related decisions are formulated, executed, and evaluated. By providing spatially explicit information and supporting transparent decision-making, these technologies empower local and national authorities to address complex and urgent land management issues. Key areas of impact include:

- **Territorial Fragmentation:** Identifying patterns of scattered landownership and proposing strategies for consolidating fragmented plots.
- **Urban Sprawl Tracking:** Monitoring the expansion of built-up areas and ensuring sustainable infrastructure development.
- **Rural Revitalization:** Supporting the reactivation of neglected regions through spatially informed interventions.
- **Environmental Sustainability:** Evaluating land-use impacts on ecological systems and designing mitigation measures.

Geospatial technologies enable not only the layering and integration of diverse spatial datasets, but also advanced spatial analysis, scenario simulation, and stakeholder coordination. In this way, they promote more intelligent and adaptive governance frameworks, particularly in contexts characterized by fragmentation, abandonment, or outdated cadastral infrastructure.

This research builds on these capabilities directly by implementing a spatial optimization model that simulates land consolidation through a graph-based algorithm. By leveraging parcel geometry and adjacency data, the model demonstrates how geospatial technologies can go beyond analysis to actively guide territorial restructuring. The result is a prototype of a decision support system (DSS) that uses spatial computation to inform rural policy, improve land use efficiency, and promote a more equitable and sustainable land distribution. In this sense, geospatial technologies act not only as diagnostic tools but also as operational engines of territorial transformation.

2.3. Legislative and Policy Framework

2.3.1. National Policies and Initiatives

Territorial management in Portugal has undergone a profound evolution over the years, driven by pressing socio-economic and environmental needs. One of the main instruments in this evolution is the *Programa Nacional da Política de Ordenamento do Território* (PNPOT), which defines the strategic framework for land use and spatial planning. This program addresses a wide range of challenges—including urban sprawl, regional disparities, and environmental degradation—by establishing guiding principles and development models that promote coherent and sustainable territorial organization [20]. PNPOT also serves as a reference for aligning regional and municipal plans with national policy goals.

A central component of territorial management is the cadastral system (*cadastro predial*), which provides the legal and spatial basis for identifying ownership, delineating parcel boundaries and supporting the enforcement of land policy. The persistent absence of a fully operational and comprehensive cadastral system has historically limited the ability of authorities to effectively plan, regulate, and optimize land use.

To counteract land fragmentation in rural areas and promote agricultural revitalization, the Portuguese government launched the *Emparcelar para Ordenar* initiative. This program encourages voluntary consolidation of dispersed parcels to increase the size of landholdings, improve productive efficiency, and ensure long-term sustainability of rural territories [10]. The initiative represents a policy commitment to reverse decades of structural disaggregation in the rural landscape.

However, while these national programs articulate important strategic goals, they do not specify the technical means by which consolidation should be implemented or evaluated. This gap opens space for innovative approaches, such as algorithmic models, that can simulate and optimize land reallocation scenarios based on geospatial data and equity considerations.

The decision support system developed in this research aligns and operationalizes the objectives of *Emparcelar para Ordenar*, offering a scalable and data-driven method to evaluate consolidation opportunities, simulate property exchanges and generate fair, efficient, and territorially coherent solutions. As such, it provides a valuable tool for supporting the effective implementation of national territorial management policies.

2.3.2. Rules of the European Union

A significant portion of Portugal's territorial management policies is shaped by directives from the European Union (EU), particularly those focused on agriculture, sustainability, and rural development. Among these, the Common Agricultural Policy (CAP) stands out as a key influence. CAP provides financial support and strategic guidance to promote sustainable land use, ensuring that agricultural production is maximized while environmental protection is maintained.

Through its implementation, Portugal has been able to align its rural development strategies with broader European objectives, strengthening economic resilience and environmental responsibility in rural areas [13].

Also central to this framework is the regulatory standard [3], which requires the incorporation of geospatial data into cadastral information systems. In October 2023, this regulation reinforces the obligation to maintain updated spatially referenced land records that support effective monitoring, planning, and compliance with EU agricultural policies. These standards are intended to improve the transparency, efficiency and territorial cohesion of rural and agricultural governance.

However, meeting these European objectives requires more than policy alignment: it demands concrete technical solutions that can translate regulatory principles into practice. The algorithmic approach developed in this research responds directly to these challenges. By optimizing land consolidation through geospatial data and owner-level simulations, the model supports sustainable and efficient land use, aligns with CAP's goals of productivity and environmental balance, and promotes equitable territorial development.

In addition, the system facilitates compliance with EU directives by providing a mechanism to identify and evaluate consolidation scenarios that meet both national needs and European standards. It exemplifies how advanced spatial technologies and decision support systems can bridge the gap between legislative vision and implementation on the ground.

In summary, the interaction between EU policy, national initiatives, and geospatial innovation forms the foundation for a new generation of smart land management strategies. The present research contributes to this paradigm by offering a technically robust and policy-aligned tool to tackle fragmentation, improve land governance, and advance sustainable territorial development in Portugal.

2.4. AI and the Future of the Land Management System

2.4.1. Framing AI within Territorial Governance

Artificial intelligence (AI) is revolutionizing land management by shifting from static intuition-based planning to data-driven decision-making processes. Broadly defined as the replication of human cognitive functions by machines, AI encompasses a wide range of techniques, from machine learning and neural networks to symbolic reasoning and heuristic optimization [18]. These technologies support territorial governance by enabling advanced analysis, predictive modeling, and spatial optimization of land use, resource distribution, and environmental impact.

In the context of territorial management, AI holds particular promise in addressing complex spatially embedded problems such as land fragmentation, abandonment, and inefficient land use. By transforming geospatial data into actionable insights, AI enables planners and policy makers to explore consolidation scenarios, evaluate spatial trade-offs, and simulate the effects of land use interventions.

This research aligns with the emerging trend of vertical AI: domain-specific intelligent systems tailored to solve complex problems in narrowly defined fields. Rather than employing generic AI methods such as genetic algorithms or simulated annealing (covered in Section 2.5), the proposed model is grounded in spatial graph analysis and rule-based multi-objective heuristics. Using structured geospatial data, it simulates block-level land exchanges and optimizes key performance indicators, such as area efficiency and contiguity.

Importantly, this approach demonstrates that AI in land governance does not require opaque or black-box models. In contrast, it thrives on interpretability: the transformation logic is explicit (e.g., adjacency detection, area gain computation, cluster reduction), the optimization goals are transparent, and the outcomes are fully traceable. As such, this type of AI offers not only analytical power, but also trust and accountability, key qualities for public-sector applications involving land and property.

2.5. Optimization in Territorial Management

Optimization has become a key approach to solving spatial inefficiencies in territorial management. In rural Portugal, the persistence of fragmented landholdings creates operational barriers to agricultural modernization, infrastructure development, and efficient land use. Addressing this problem requires computational strategies capable of reorganizing land ownership into more contiguous and productive configurations.

Optimization models provide a structured means of evaluating alternative land arrangements based on spatial, economic, and social criteria. These models aim to consolidate scattered plots into fewer, larger, and better-connected parcels while minimizing negative impacts, such as loss of productive area or inequality between owners.

2.5.1. Using Multilayer Networks to Optimize Properties

A powerful modeling approach involves the use of multilayer networks. In this structure, one layer represents land parcels and their spatial properties, while another layer encodes ownership or usage rights. Interlayer connections reflect ownership relations, while intralayer edges define spatial proximity or adjacency between parcels.

This research incorporates the logic of multilayer networks through a simplified graph-based structure. Each node represents a block of contiguous parcels owned by an individual, and edges reflect the adjacency to blocks owned by different individuals. This structure allows for the identification of high-potential land exchanges that would lead to better parcel aggregation. By simulating transfers between owners based on proximity and area balance, the algorithm generates improved configurations without requiring a complete cadastral reorganization.

2.5.2. Algorithmic Approaches

Several algorithmic strategies have been applied to territorial optimization problems, particularly those involving land fragmentation:

- **Genetic Algorithms (GA):** Inspired by biological evolution, GAs are often used for multi-objective optimization. In land management, they can balance competing goals such as agricultural efficiency and equitable land distribution [8].

- **Simulated Annealing (SA):** A probabilistic technique that searches for global optima by exploring a wide range of solutions while gradually reducing randomness. It has been used to optimize spatial patterns and land use configurations [23].
- **Very Fast Simulated Annealers (VFSA):** An efficient variation of SA that accelerates convergence, making it suitable for large-scale spatial datasets such as rural parcel maps [7].

Although these approaches offer flexibility, they typically involve unclear internal dynamics and substantial hyperparameter tuning. In contrast, the algorithm developed in this research follows a domain-specific rule-based heuristic with explicit decision logic. It runs on a spatial graph of ownership blocks, enumerates adjacent candidates for exchange, and scores each potential absorption using an interpretable edge weight:

$$w_{A,B} \;=\; \mathrm{benefit}_{A,B} \;-\; \alpha \cdot \frac{|\,\mathrm{area}_A - \mathrm{area}_B\,|}{\mathrm{area}_A + \mathrm{area}_B},$$

where $\mathrm{benefit}_{A,B}$ captures the contiguity gain from merging blocks and α penalizes asymmetric swaps. This makes every choice auditable: stakeholders can observe the adjacency that motivated the edge, the computed benefit, the applied penalty, and the thresholding that led to selection.

Rather than defining three independent objectives that risk redundancy, the model is guided by a single organizing principle ,improving the structural coherence of landholdings, which is operationalized through three tightly related indicators: (i) mean land area per owner (viability), (ii) number of contiguous ownership clusters (fragmentation), and (iii) net area variation (fairness of redistribution). The algorithm continues iteratively (match $\rightarrow$ absorb $\rightarrow$ dissolve $\rightarrow$ rebuild graph) until no edge with positive weight remains or progress stalls. Interpretability is preserved because each step corresponds to observable spatial operations, each exchange is justified by a transparent score, and all intermediate states (blocks, adjacencies, weights, selections) can be logged for ex-post verification. This yields a pragmatic, policy-ready alternative to general-purpose optimizers: it achieves measurable consolidation gains while keeping the decision pathway traceable and explainable.

2.6. Sources and Availability of Data

2.6.1. Role of iSIP (Sistema de Identificação Parcelar)

Sistema de Identificação Parcelar (iSIP) serves as a central repository of cadastral and geospatial data in Portugal. Operated by the Instituto de Financiamento da Agricultura e Pescas (IFAP), the platform offers detailed parcel-level information such as boundaries, land use classifications, and legal constraints. It plays a pivotal role in the management of the territory, the validation of agricultural subsidies, and the support of geospatial analysis for land-related decision-making. [10].

2.6.2. Institutional Integration and Future Interoperability

Recent efforts have aimed to integrate institutional datasets to improve the transparency and efficiency of land governance. A key development is the protocol signed between eBUPi and

IFAP, which enables the sharing of cadastral data between systems. IFAP provides information on the iSIP package, while BUPi provides legal and ownership data, creating a two-way data flow that enriches both platforms [12].

Future integration efforts will further consolidate these systems into a unified cadastral infrastructure that combines legal ownership records with georeferenced spatial data [3]. This will allow for dynamic updates on parcel characteristics and property rights, facilitating compliance with EU regulations, and enabling more coherent, data-driven territorial policies. For algorithmic tools such as the one developed in this research, such interoperability will enhance scalability and ensure that spatial optimization models remain accurate and legally valid over time.

2.6.3. Data Accessibility and Recent Improvements

Although Portugal has made notable progress in digitizing cadastral information, several challenges remain, especially in rural areas. Lack of standardization, fragmented formats, and incomplete coverage continue to hinder seamless integration and legal validation of geospatial models.

However, the situation is improving. The online version of iSIP now provides public access to parcel boundaries, land classifications, orthophotos, and associated metadata. This transparency supports broader civic participation and offers a foundation for building reliable spatial decision support systems. Although ownership data remain restricted, geometric and topological information is sufficient to simulate consolidation strategies, as demonstrated in this investigation.

2.6.4. Effects of Data Integration in Agriculture

Centralized access to land data has already produced tangible benefits for farmers. By consolidating cadastral, regulatory, and environmental information in accessible platforms, stakeholders can better plan resource allocation, comply with policy requirements, and adopt modern precision farming techniques [10].

This improved data environment also strengthens participation in agricultural support programs, both national and EU-funded, where accurate geospatial information is increasingly required for eligibility. The land consolidation algorithm proposed in this thesis complements these trends by offering a tool that improves land structure and promotes scalable agricultural practices. In this way, algorithmic land management becomes a key enabler of more sustainable, productive, and equitable rural development.

In summary, the reviewed literature and institutional context reveal a persistent challenge of land fragmentation in rural Portugal, exacerbated by cadastral incompleteness and unequal development patterns. Geospatial technologies, optimization algorithms, and emerging AI methodologies have shown potential in addressing these issues through more intelligent and data-driven land management. Building on these foundations, the following chapter presents the methodological framework adopted in this research, describing the structure, logic, and implementation of a computational system designed to support and simulate land consolidation strategies.

CHAPTER 3

Methodology and Data

This chapter presents the methodological approach adopted to develop a decision support system for land consolidation in Portugal. Building upon the contextual and technological foundations discussed in the previous chapters, the methodology integrates geospatial data processing, ownership simulation, and graph-based optimization to address land fragmentation in a structured and scalable manner.

To ensure analytical rigor and adaptability, the project follows the Cross-Industry Standard Process for Data Mining (CRISP-DM), a widely used framework for guiding data-driven projects. The chapter begins by outlining this framework (Section 3.1), then details the data sources and pre-processing steps (Section 3.2), followed by a comprehensive explanation of the algorithm design, execution and evaluation (Sections 3.3 and 3.4). It concludes with a reflection on methodological limitations and a synthesis of key points (Sections 3.5 and 3.6).

3.0.1. CRISP-DM Framework

The development of this land consolidation system was guided by the Cross-Industry Standard Process for Data Mining (CRISP-DM), a widely adopted framework to structure data-driven projects in all domains [26]. CRISP-DM is composed of six iterative phases that ensure both analytical depth and alignment with real-world challenges.

- **Business Understanding:** Define strategic goals related to land fragmentation, such as increasing the average parcel size, reducing the fragmentation of `owners` and allowing a more equitable land distribution.
- **Data Understanding:** Explore geospatial and cadastral datasets (e.g., iSIP, BUPi) to assess quality, completeness, and suitability to model spatial relationships and ownership structures.
- **Data Preparation:** Structure the data to construct a multilayer graph network, including cleaning geometries, simulating `owners`, identifying adjacent `parcels` and filtering eligible participants.
- **Modeling:** Develop a graph-based heuristic algorithm capable of iteratively identifying mutually beneficial land exchanges that reduce fragmentation and promote contiguity.
- **Evaluation:** Use interpretable metrics (e.g., area per owner, number of `clusters`, net area change) to evaluate solution quality and identify optimal trade-off configurations.
- **Deployment:** Store results as reusable CSV files for further analysis and visualization. Although not implemented in a production system, the prototype demonstrates practical applicability for real-world use cases.

The iterative nature of CRISP-DM allows the process to loop back, particularly between modeling and evaluation, when refinements or insights emerge. This flexibility supports a gradual improvement of the algorithm while maintaining alignment with the real-world complexity of rural land governance.

Each phase of CRISP-DM is addressed in this chapter: the **Business Understanding** is informed by the contextual and policy insights developed in Chapters 1 and 2; the **Data Understanding and Preparation** are covered in Section 3.2; the core **Modeling** logic is detailed in Section 3.3; the **Evaluation** metrics are presented in Section 3.4; and broader considerations regarding limitations and scalability are addressed in Sections 3.5 and 3.6.

3.1. Data Collection and Preparation

3.1.1. Sources and Structure of the Dataset

The dataset used in this study comes from the public version of the *Sistema de Identificação Parcelar* (iSIP), a national platform managed by the *Instituto de Financiamento da Agricultura e Pescas* (IFAP). iSIP provides geospatial data on land `parcels` throughout Portugal, including geometry, land use, and cadastral information. However, due to access restrictions and encoding limitations in mainland datasets where parcel geometries are often obfuscated or encoded, the present study focuses on the Autonomous Region of Madeira, where parcel geometries are publicly available in WKT format and compatible with standard geoprocessing workflows.

The selected dataset comprises 35,046 land `parcels`. Each parcel is represented as a polygon geometry (typically of type `MULTIPOLYGON`) and includes an attribute for area calculation (`Shape_Area`). The geometries were converted into a GeoDataFrame using Python's `geopandas` and `shapely` libraries, enabling operations such as adjacency detection, merging of contiguous `parcels`, and centroid computation.

All geometries are defined in the EPSG:3763 – PT-TM06/ETRS89 coordinate system, which ensures accurate area computation in square meters. The `Shape_Area` attribute, derived from this projection, plays a central role in ownership aggregation and optimization evaluation throughout the methodology.

Since real ownership data are not available, a synthetic ownership structure was created through a custom simulation procedure. This simulation is implemented via the `LandDistributor` class, which receives two input parameters: the desired number of synthetic `owners` and the maximum allowed variation in the number of `parcels` assigned per owner. The simulation process follows four main steps:

(1) Compute the average number of `parcels` per owner based on the total number of `parcels`;
(2) Generate a list of target parcel counts for each owner, sampled from a uniform distribution centered on the average and bounded by the variation parameter;
(3) Randomly assign `parcels` to `owners` while respecting the target distribution;
(4) Assign a unique `owner_id` to each parcel.

This process is deliberately spatially agnostic: `parcels` are assigned independently of location or proximity. The result is a fragmented initial ownership structure that reflects the complexity and disorder commonly observed in rural landholding systems. This synthetic distribution creates the conditions necessary to test the consolidation logic under realistic fragmentation.

Following the initial assignment, a filtering step is applied using the `PlotFilter` class. Owners with only a single parcel are excluded from the dataset, as they cannot participate in exchanges, an essential mechanism of the optimization algorithm. This ensures that the simulation retains only actors with consolidation potential, improving computational efficiency and analytical relevance.

At the end of the preparation process, each parcel in the dataset contains the following attributes:

- `parcel_id`: A unique identifier for each land unit;
- `geometry`: A valid polygon object representing the shape and location of the parcel;
- `Shape_Area`: The area of the parcel in square meters;
- `owner_id`: A synthetic owner assigned via the controlled distribution algorithm.

This structured and filtered dataset serves as the foundation for all subsequent modeling steps. In Chapter 4, three different ownership distribution scenarios are tested, each varying the number of `owners` and parcel variability, to assess their impact on the effectiveness of land consolidation strategies.

3.1.2. Pre-Processing Pipeline

After simulating ownership and structuring the base dataset, a sequence of pre-processing operations is performed to transform individual `parcels` into block-level entities. This pipeline enables the construction of a suitable spatial graph for optimization and ensures that only relevant `owners` participate in the consolidation process.

Step 1 – Adjacency Detection using STRtree The first step identifies spatial adjacencies between `parcels` using an efficient spatial index. Specifically, the algorithm employs an `STRtree`—a spatial index structure from the `shapely` library designed for fast geometric queries. All parcel geometries are indexed, and for each parcel, the tree is queried to detect intersecting or touching neighbors. The output is a list of parcel-to-parcel adjacency pairs, representing direct spatial connections between land units.

Step 2 – Block Formation by Owner Once parcel-level adjacencies are identified, the next step aggregates them by ownership. Parcels that belong to the same owner and are spatially adjacent are grouped into contiguous components, hereafter referred to as *blocks*. Each block corresponds to a connected cluster of `parcels` under the same owner, ensuring geographic contiguity and single agent control. These blocks become the basic units of consolidation in the subsequent optimization process.

Step 3 – Filtering of Single-Block Owners Not all simulated `owners` are relevant for consolidation. Owners whose `parcels` form only one block are excluded from the dataset at this stage.

Since they lack internal fragmentation, these owners contribute neither to exchange dynamics nor benefit from consolidation. Removing them reduces computational overhead and ensures the model focuses on meaningful cases where structural improvement is possible.

Step 4 – Block Dissolution and Geometry Update Within each owner, the geometry of each block is computed by dissolving the corresponding parcels using geopandas.dissolve(). This operation merges adjacent polygons into a single geometry per block. The area and centroid of each resulting block are recalculated, and adjacency links between blocks owned by different individuals are updated. This ensures that the spatial graph reflects the current state of the system after every structural change.

pre-processing Output Upon completion of the pipeline, the dataset is transformed into a clean and compact structure ready for optimization. It includes:

- A list of blocks, each with a unique geometry, area, centroid, and owner ID;
- A reduced set of owners, each controlling at least two blocks;
- A spatial adjacency graph connecting blocks owned by different agents.

These outputs serve as input for the graph construction and matching logic described in Section 3.2.1. An overview of this pre-processing logic is illustrated in Figure 3.1.

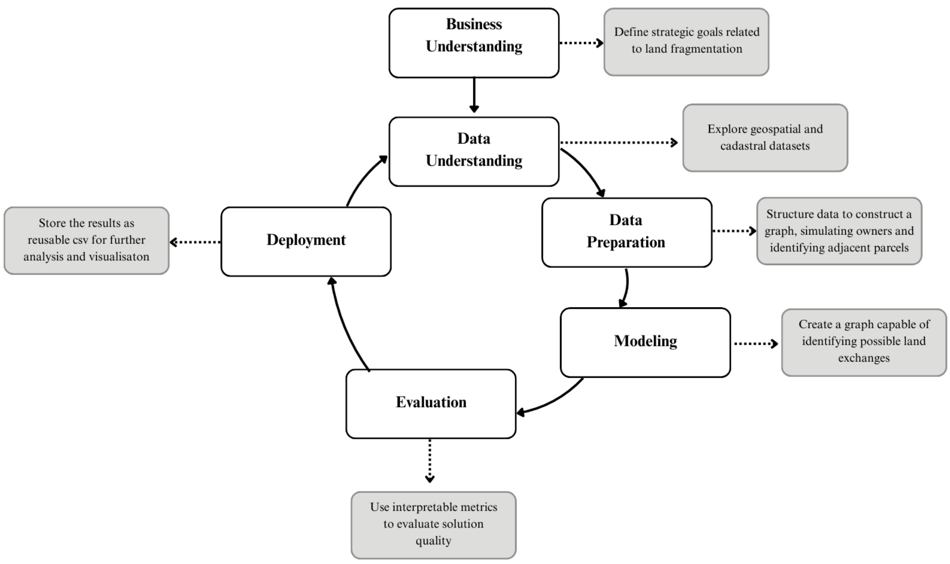

Figure 3.1. *Overview of the pre-processing pipeline from* parcels *to blocks and graph structure*

3.2. Algorithm Design and Execution

3.2.1. Graph-Based Model

At the heart of the land consolidation algorithm lies a graph-based model that encodes the spatial and ownership structure of land parcels as a network. This abstraction enables the simulation of land exchanges through computational operations on nodes and edges, facilitating the identification of mutually beneficial swaps between landowners.

In this model, each **node** represents a contiguous *block* of land—i.e., a group of spatially connected `parcels` owned by a single individual, as defined during the pre-processing stage. Each node contains the following attributes:

- `geometry`: The unified polygon resulting from dissolving the block's constituent `parcels`;
- `area`: The total area of the block;
- `owner_id`: Identifier of the block's owner;
- `centroid`: The geometric center used in spatial computations.

An **edge** connects two `nodes` if their corresponding blocks are adjacent in space, either sharing a common boundary or being within a small distance threshold. Each edge denotes a potential exchange between two `owners`, where one block could be transferred in a way that improves the spatial coherence of their holdings.

To quantify the quality of each potential exchange, a numerical **weight** is assigned to each edge. This weight reflects the net benefit of the swap, balancing two components:

(1) **Contiguity Gain (benefit$_{AB}$)**: Measures whether the new block would increase the geographic contiguity of the recipient's holdings (e.g., by merging two previously disconnected blocks).

(2) **Area Fairness Penalty**: Penalizes imbalanced swaps where the exchanged blocks differ significantly in area.

The total weight of an edge connecting blocks A and B is computed as:

$$w_{AB} = \text{benefit}_{AB} - \alpha \cdot \left| \frac{\text{area}_A - \text{area}_B}{\text{area}_A + \text{area}_B} \right|$$

where:

- benefit$_{AB}$ is the structural gain in contiguity for both `owners` if the exchange occurs;
- α is a tunable parameter that controls the penalty imposed by asymmetric area swaps.

The resulting graph is a weighted, undirected network where `edges` encode both spatial feasibility and strategic value. This graph is used by the optimization engine (detailed in Section 3.2.2) to identify a maximum weight matching: a set of non overlapping block exchanges that maximize the overall benefit.

After each iteration of the matching algorithm, selected exchanges are performed, affected blocks are dissolved, and the graph is rebuilt to reflect the updated ownership and adjacency structure. This iterative process continues until the convergence criteria are met.

Figure 3.2 illustrates the conceptual structure of this graph model, with `nodes` representing blocks and `edges` representing consolidation opportunities between neighboring `owners`.

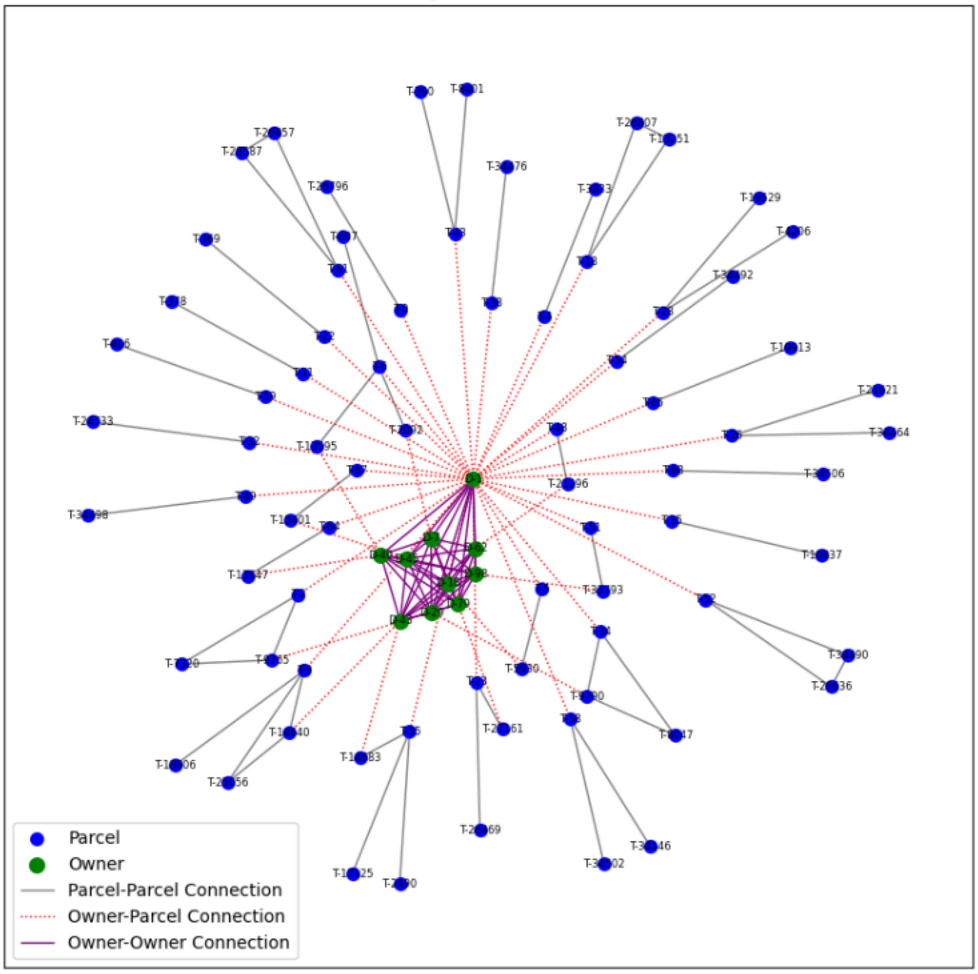

Figure 3.2. *Example of a block-level graph where* `nodes` *represent contiguous land blocks and* `edges` *represent consolidation opportunities*

3.2.2. Matching and Exchange Logic

After constructing the graph of spatial blocks and their potential consolidation benefits, the algorithm proceeds to identify and execute land exchanges through an iterative matching process. The goal is to maximize spatial efficiency by merging fragmented ownership structures without imposing rigid constraints on area equality.

The matching step is performed using the Edmonds algorithm for the maximum weight matching in undirected graphs, implemented via the `networkx` Python library. This classical algorithm identifies a set of nonoverlapping `edges` (block pairs) whose total edge weight is maximized, meaning it prioritizes the most strategically beneficial exchanges in each round.

Absorption Rule. For each pair of adjacent blocks A and B, the algorithm applies an asymmetric absorption rule: the smaller block is absorbed into the larger one. This choice reflects real-world logic, where expanding contiguous territory is preferable to fragmenting it. As a result:

- The owner of the larger block gains the area and geometry of the smaller block.
- The previous owner relinquishes the absorbed block and sees their total area reduced.
- The system records this exchange as a block absorption event.

Example. Suppose block A (owned by owner 1) has an area of 6,000 m^2, and block B (owned by owner 2) has 3,500 m^2. If the two blocks are adjacent and the edge between them has a high weight, indicating a potential contiguity gain for owner 1, the algorithm selects this pair for matching. Since B is smaller, it is absorbed by A, and owner 1's holdings expand, potentially reducing the number of `clusters` in their ownership. This change modifies the overall structure of the ownership network and adjacency relationships.

Dissolution and Graph Update. After all exchanges in the current round are completed, the algorithm executes a `dissolve` operation to merge any newly contiguous blocks belonging to the same owner. This has three main effects:

- Reduces internal fragmentation by unifying connected components.
- Updates the geometries and areas of the new blocks.
- Alters spatial adjacencies between blocks, requiring the graph to be rebuilt.

The graph is reconstructed after each round:

- **Nodes** are redefined to reflect new block geometries and owner assignments.
- **Edges** are recalculated by checking the adjacency between the updated blocks.
- **Weights** are recomputed using the same formula introduced in Section 3.2.1, incorporating updated area ratios and contiguity benefits.

Termination. The iterative process continues until one of the following conditions is met:

- No more beneficial exchanges are identified (i.e., all edge weights are non positive).
- A predefined runtime threshold is reached.

Through this cycle of matching, absorption, dissolution, and graph rebuilding, the model progressively enhances the spatial structure of land ownership. The logic ensures that only high-impact exchanges are performed, while maintaining flexibility and fairness throughout the optimization process.

3.2.3. Objective Function and Optimization Criteria

The land consolidation algorithm is guided by a multi-objective optimization framework that evaluates several criteria after each round of exchanges. Rather than relying on a single composite score, the model assesses the following three goals independently, enabling a more nuanced and interpretable optimization process:

- **Maximize average area per owner** - indicator: increase in mean landholding area;
- **Minimize net area loss** - indicator: total net difference in area exchanged;
- **Reduce the number of ownership `clusters` (blocks)** - indicator: decrease in the number of disconnected contiguous units per owner.

These criteria reflect the core goals of rural territorial planning: reducing fragmentation, increasing landholding viability, and maintaining equity in area distribution. After each iteration, the model computes these indicators to assess whether the consolidation process produces desirable outcomes.

No Fixed Weights. Unlike traditional multi-objective models that collapse all goals into a weighted sum, this algorithm deliberately avoids assigning fixed importance to any criterion. Instead, it evaluates progress across all dimensions and tracks improvements independently. This design supports Pareto analysis, allowing stakeholders to later explore trade-offs between area gain and area fairness without being constrained by predefined priorities. Such flexibility is essential in public-sector contexts where decision-making must accommodate competing interests and adapt to evolving policy goals.

Termination Criteria The optimization loop continues until one or more of the following stopping conditions is met:

- **Runtime limit**: A predefined maximum execution time ensures that the algorithm remains computationally feasible even on large datasets.
- **Stagnation in metrics**: If key metrics, such as the number of absorbed blocks or the average area per owner, stop improving across consecutive iterations, the algorithm assumes it has reached a local optimum and halts.
- **Empty matching set**: When no beneficial exchanges remain in the graph (that is, all edge weights are zero or negative), the matching algorithm yields no valid swaps, signaling the exhaustion of optimization opportunities.

Together, this multi-objective strategy and adaptive termination logic enable the algorithm to explore a diverse solution space without sacrificing transparency or computational control. The resulting land configurations form a Pareto frontier of non-dominated alternatives, each representing a distinct trade-off among consolidation, fairness, and structural coherence.

3.3. Metrics and Evaluation

To assess the performance and impact of each land consolidation run, the algorithm records a comprehensive set of metrics that evaluate spatial structure, area fairness, and algorithmic efficiency. These indicators serve three main purposes: (i) evaluating solution quality, (ii) enabling comparison across runs and parameterizations, and (iii) informing decision-making by revealing trade-offs between competing objectives.

All results are saved in a structured output file named `consolidacao_runs.CSV`, where each row corresponds to one complete execution of the algorithm. The CSV includes both input parameters (e.g., `alpha`, `base_benefit`, `min_block`) and outcome metrics, ensuring full traceability. These data can be reused for diagnostic analysis, benchmarking, or the construction of Pareto fronts - allowing stakeholders to compare alternative configurations with different trade-off balances. Such visualizations are explored in the results chapter (Chapter 4).

The main evaluation metrics are:

- **Average Area per Owner (Initial and Final):** Mean land area held by each owner before and after consolidation. An increase in this metric reflects improved territorial viability and successful land aggregation.

- **Number of Ownership Clusters (Initial and Final):** Total number of disconnected contiguous blocks per owner, used as a direct indicator of spatial fragmentation. For example, in one test run, the average cluster count dropped from 3.2 to 1.7 per owner.
- **Net Area Exchanged (`diff_area_final`):** Net sum of gains and losses across all `owners`. Ideally, this should approach zero, indicating balanced exchanges with minimal total loss.
- **Area Exchanged:** Cumulative amount of land area transferred during consolidation. This contextualizes the structural change in terms of the required land movement.
- **Number of Rounds Executed:** Total iterations of matching, swapping, and graph rebuilding until convergence or timeout, indicative of how long the system required to reach equilibrium.
- **Total Number of Blocks Absorbed:** Reflects how many blocks were eliminated by consolidation - serving as a proxy for structural simplification.
- **Number of Parcels (Initial and Final):** Indicates changes in the number of individual geometries after dissolution and merges. The lower post-exchange parcel count reflects successful integration.
- **Number of Active Owners:** Number of `owners` with at least two `parcels` who were eligible and included in the consolidation process after filtering.
- **Runtime (seconds):** Total processing time for the run. Useful for assessing computational cost, especially across larger datasets.
- **Key Parameters Used:** Inputs such as `alpha`, `base_benefit`, `min_block`, and number of `owners`-captured to support reproducibility.

All metrics are tracked dynamically during execution, allowing the system to:

- Monitor the evolution of key indicators across consolidation rounds;
- Detect diminishing returns or algorithmic stagnation;
- Generate Pareto front visualizations that expose trade-offs between objectives (e.g., maximizing area vs. minimizing net loss).

By structuring output in a reusable tabular format, this system bridges computational modeling with real-world application, allowing technical output to directly support territorial planning decisions. The evaluation strategy complements the algorithmic logic, ensuring coherence between the model design and the performance analysis.

3.4. Limitations

Although this research advances the state-of-the-art in spatially optimized land consolidation, several limitations were encountered during its development. These can be grouped into four main categories: data availability, model simplifications, scalability constraints, and practical relevance.

Data Availability and Geographic Scope

The analysis was restricted to the Autonomous Region of Madeira due to data accessibility constraints. Although the national iSIP platform provides parcel-level information for the entire country, the geometries for mainland Portugal are encoded in a proprietary format, rendering them incompatible with standard spatial analysis tools. Only Madeira provided geometry in WKT format, allowing direct integration into the GIS pipeline. As such, the study lacks national coverage and does not evaluate the regional variation in land structure.

Model Simplifications

Several simplifications were required during algorithm design. In particular, the model does not account for contextual or economic land attributes, such as proximity to main roads or urban centers, soil quality, or infrastructure access. Although datasets were identified for features such as road hierarchies and urban polarities (e.g., for Lisbon), equivalent data was unavailable for Madeira. Therefore, these features were excluded even though the model architecture could technically accommodate them. As a result, all optimization was based strictly on geometric criteria: area, adjacency, and contiguity.

Scalability and Computational Constraints

The iterative consolidation algorithm, based on graph matching and repeated geometry updates, is increasingly demanding as the number of `parcels` and `owners` increases. Although the model performed efficiently in the Madeira case (35,000 `parcels`), applying it to full national datasets would likely require further optimization or parallelization. The runtime was limited (for example, 15 minutes per run), but the complexity of the algorithm increases non-linearly with the size of the graph, which may limit its applicability in larger-scale deployments.

Practical Relevance and Realism

Due to the absence of public ownership records, the land`owners` in the dataset were synthetically simulated using a randomized algorithm. Although this allows for controlled experimentation, it introduces a degree of artificiality and does not reflect true ownership patterns, landholding histories, or social dynamics. The results must therefore be interpreted as exploratory simulations rather than precise forecasts. Furthermore, the algorithm assumes that block-to-block exchanges are feasible whenever they are spatially advantageous, removing legal, institutional, or behavioral barriers to land reallocation.

These limitations do not undermine the conceptual validity of the model, but highlight opportunities for future refinement. Chapter 4 discusses how some of these challenges shaped the empirical results, while Chapter 5 outlines possible directions for extending the framework toward broader applicability and realism.

3.5. Summary

This chapter presented the methodological framework used to design, implement and evaluate a land consolidation system for Portugal. Based on CRISP-DM, the process began with geospatial data from the Autonomous Region of Madeira, where parcel geometries were accessible and ownership was simulated under controlled conditions. The data was structured into connected blocks and modeled as a graph, with `nodes` representing ownership units and `edges` representing potential exchanges. A maximum weight matching algorithm guided iterative optimization, balancing three core objectives: maximizing the average area per owner, minimizing net area loss, and reducing fragmentation. The metrics recorded in each run provide a transparent basis for performance evaluation. In the next chapter, this framework is applied to the test dataset, and the resulting trade-offs in land consolidation outcomes are analyzed.

CHAPTER 4

Results and Discussion

4.1. Experimental Scenarios

This section outlines three distinct scenarios designed to test the algorithm under varying levels of land fragmentation.

To assess the performance and scalability of the proposed land consolidation algorithm, three experimental scenarios were developed. Each scenario is based on the same dataset - 35,046 land `parcels` from the autonomous region of Madeira, but differs in how ownership is synthetically distributed. The goal is to simulate varying degrees of land fragmentation and ownership concentration, allowing systematic observation of the algorithm's behavior under different territorial configurations.

All scenarios apply the same optimization logic, as defined in Chapter 3, but differ in their structural complexity. This design ensures that variations in results are attributable to the initial ownership pattern rather than to algorithmic parameters.

4.1.1. Scenario 1: Highly Fragmented Ownership

This scenario simulates an extreme fragmentation context, where a small number of `owners` hold many dispersed `parcels`. Specifically, the 35,046 `parcels` are assigned to only 3,000 synthetic `owners`, resulting in a high average of 11.68 `parcels` per owner (ranging from 2 to 24). This exaggerates the baseline complexity and allows the algorithm to be tested in a worst-case configuration for consolidation.

4.1.2. Scenario 2: Realistic Distribution Based on National Averages

This scenario aims to reflect a more realistic ownership structure, based on Portuguese land statistics and broader southern European trends. According to FAO, DGADR and INE reports, most smallholders in Portugal own between 3 and 6 `parcels`, with fragmentation particularly severe in the North and Center [16, 14, 17].

To approximate these conditions, `parcels` were assigned to 6,000 synthetic `owners`, resulting in an average of 5.84 `parcels` per owner and a range between 2 and 9. This configuration serves as a policy-relevant baseline, allowing the evaluation of the algorithm's behavior under empirically grounded assumptions.

4.1.3. Scenario 3: Concentrated Ownership

This scenario represents a context with minimal fragmentation, a proxy for a near-optimal land structure. Here, the same number of `parcels` is distributed among 10,000 `owners`, yielding an average of 3.5 `parcels` per owner (with minimal variation between 2 and 5). This structure is used to evaluate the performance of the algorithm in cases where the potential for improvement is less and the ownership is already relatively aggregated.

4.1.4. Scenario Summary

Table 4.1 provides a consolidated view of the three scenarios, including the number of `owners`, the characteristics of the parcel distribution, and the average level of fragmentation. These configurations define the baseline conditions for the simulations in Chapter 4.

Table 4.1. *Overview of the three experimental scenarios*

Scenario	Owners	Min Parcels	Max Parcels	Avg Parcels / Owner
Highly Fragmented	3,000	2	24	11.68
Realistic (Portugal Avg)	6,000	2	9	5.84
Concentrated	10,000	2	5	3.50

These scenarios define a structured testbed for evaluating algorithmic performance across a spectrum of real and hypothetical consolidation challenges. Although ownership assignments are synthetic, they are grounded in relevant statistical distributions and serve to model representative planning contexts.

4.2. Metrics Used for Evaluation

To evaluate the behavior of the algorithm in the three experimental scenarios, a set of quantitative metrics was defined. These indicators are extracted from the structured output file `consolidacao_runs.CSV`, which stores the results of each execution in reproducible format. Together, they allow for the analysis of spatial fragmentation, equity of redistribution, algorithmic efficiency, and potential trade-offs introduced by the optimization process.

The metrics are grouped into three main categories: **fragmentation and spatial structure**, **equity and redistribution**, and **computational efficiency**. Their relevance for both technical assessment and real-world applicability is summarized in Table 4.2.

4.2.1. Fragmentation and Spatial Structure

These metrics capture how effectively the algorithm reduces land fragmentation and improves parcel contiguity:

- **Number of Clusters (`clusters`):** Represents the total number of contiguous ownership units after consolidation. A lower value indicates stronger spatial aggregation.
- **Mean Area per Owner (Start and Final):** Provided by `mean_area_start` and `mean_area_final`, these values track changes in landholding size before and after the process.
- **Delta and Percentage Gain (`delta_mean`, `pct_mean`):** Quantify absolute and relative improvements in average area, standardizing comparisons between scenarios of different sizes.

4.2.2. Equity and Redistribution Effects

This group of metrics evaluates the fairness of land exchanges and the distribution of losses:

- **Net Area Difference (`diff_area`):** Measures the net imbalance in the area exchanged. Ideally close to zero, this reflects the transaction symmetry.
- **Total, Mean and Max Area Lost (`area_perdida_total`, `media`, `max`):** Indicate the global and individual magnitude of land losses, offering insight into the cost of reallocation.
- **Share of Owners with Losses (`donos_com_perda`, `percent`):** Reveal how widely sacrifices are distributed among participants - a key measure of perceived fairness.

4.2.3. Computational Efficiency and Reproducibility

The following metrics are related to the execution cost and help contextualize the performance under different scenarios.

- **Runtime (`runtime_sec`):** Total time required per execution, reflecting algorithmic scalability.
- **Algorithm Parameters:** Each execution logs key hyperparameters used during graph construction and optimization:
  - `alpha`: Penalization for area differences in `edges`.
  - `base_benefit`: Minimum utility threshold for exchanges.
  - `min_block_1` and `min_block_2`: Constraints on the minimum block count per owner.

  These parameters are critical for reproducibility and interpretability.

Table 4.2. *Summary of evaluation metrics and their relevance*

Metric	**Definition**	**Interpretation and Policy Relevance**
`clusters`	Total number of contiguous blocks owned	Lower values indicate improved spatial cohesion and reduced fragmentation.
`mean_area_initial / mean_area_final`	Average area per owner before/after	Key indicator of consolidation success.
`delta_mean_area`	Absolute change in average area	Captures net gain in landholding size per owner.
`pct_mean_area`	Relative gain (%)	Enables comparison across scenarios of different scale.
`net_area_difference`	Net imbalance in area exchanged	Reflects fairness and symmetry of transactions.
`total_area_lost`	Total area lost by all owners	Measures global sacrifice / efficiency cost.
`mean_area_lost / max_area_lost`	Mean and max area lost per affected owner	Highlights individual burden, informing equity evaluation.
`owners_with_loss / pct_owners_with_loss`	Count and percentage of owners with losses	Indicates how widely losses are distributed in the population.
`runtime_sec`	Execution time	Proxy for computational scalability.
`alpha, base_benefit, min_block_x`	Algorithm parameters	Support interpretability and diagnostics; scenario-agnostic analysis.

4.2.4. Use of Metrics in Analysis

These metrics serve as the basis for all subsequent evaluations in Chapter 4. They are used to:

- Compare algorithm performance across the three scenarios;
- Visualize Pareto trade-offs between efficiency and equity;

- Interpret statistical patterns in fragmentation reduction and area distribution;
- Diagnose algorithmic behavior under varying structural conditions.

4.3. Scenario Analysis and Comparative Evaluation

4.3.1. Scenario 1: 3000 Owners

This scenario represents a consolidation effort over a moderately sized region with approximately 3000 `landowners`. The algorithm performed consistently in all executions, exhibiting both stability and efficiency. The **average area per owner** increased from approximately 2130 to 2350 units, yielding a mean gain of 10.9%. This improvement was robust in all runs, with gains ranging between 8% and 16%, demonstrating convergence to near-optimal configurations.

The total area **removed** from the system averaged 3.7×10^6 units, which, although the smallest in absolute terms among all scenarios, represents a proportionally significant share due to the smaller overall size of the region. In particular, more than **52% of the** `owners` **retained or expanded** their initial landholdings, while the remaining 48% contributed land. Among contributors, the average loss was around 2540 units, indicating moderately distributed sacrifices.

Figure 4.1 shows the **correlation matrix**, which reveals a near-zero correlation between `delta_mean_area` and both `total_lost_area` and `pct_owners_with_loss`. This suggests that larger losses or more contributors do not necessarily lead to significantly higher gains, pointing to an **equilibrium point** in the optimization space.

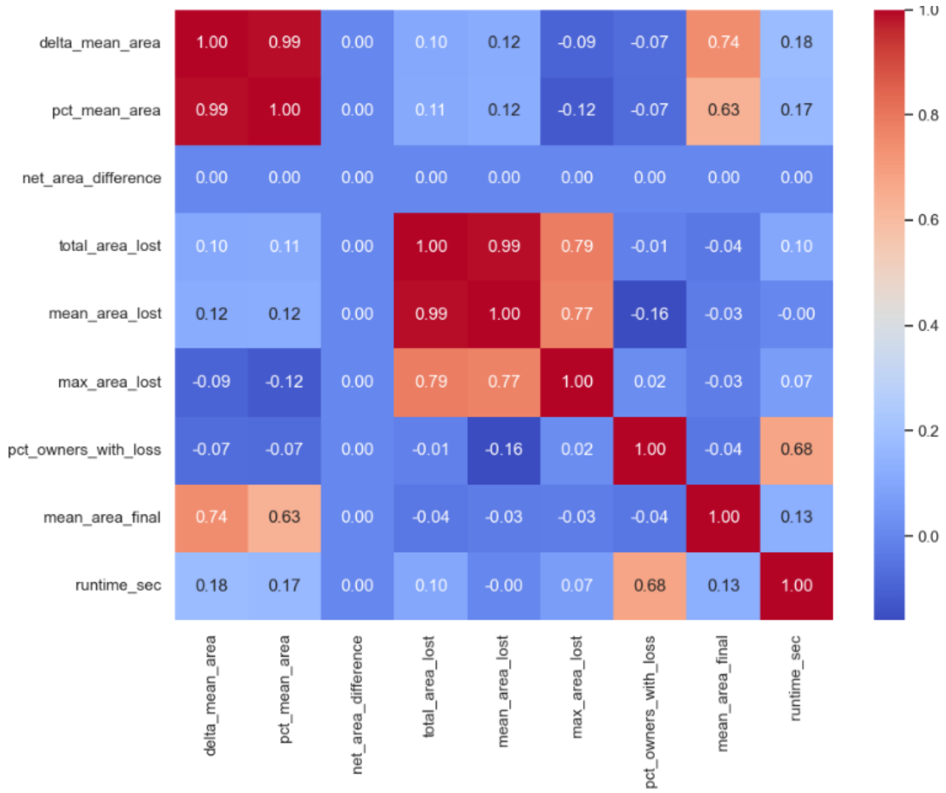

Figure 4.1. *Scenario 1 (3000 Owners) - Correlation matrix. Low correlation suggests independence between global gains and losses.*

Figure 4.2 illustrates the distribution of results across different simulations, showing the relationship between total area lost and the resulting mean area. The red line connects the best-performing configurations, outlining an empirical trade-off curve. The nearly flat shape suggests diminishing returns: additional land losses lead to only marginal increases in the mean consolidated area.

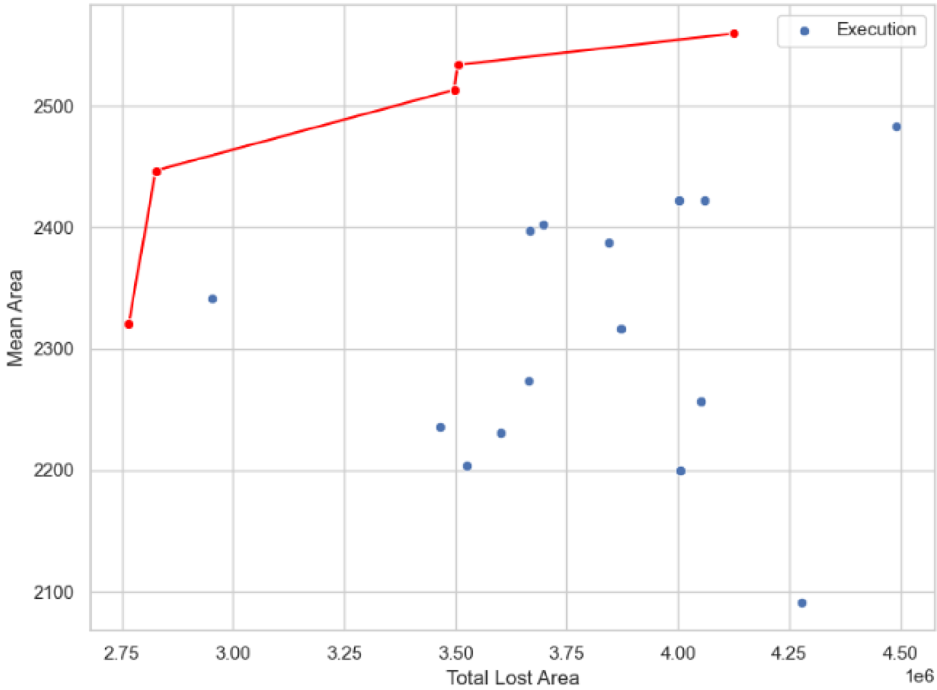

Figure 4.2. *Scenario 1 (3000 Owners) - Distribution of simulation results. The red line connects the best-performing simulations.*

4.3.2. Scenario 2: 6000 Owners (Realistic Case)

This scenario simulates a land structure aligned with empirical patterns in Portugal and Southern Europe. The 6,000-owner setup yielded average area gains of 12.0%, with execution outcomes ranging from 8% to 20%. The algorithm adapted well to this more realistic configuration, confirming its **scalability and resilience**.

The **total area removed** averaged 4.3×10^6 units. Approximately 50% of the `owners` experienced no losses, while those who did contributed about 1500 units on average. This relatively low loss per owner is illustrated in Figure 4.3, where the distribution is centered and well behaved, with no outliers.

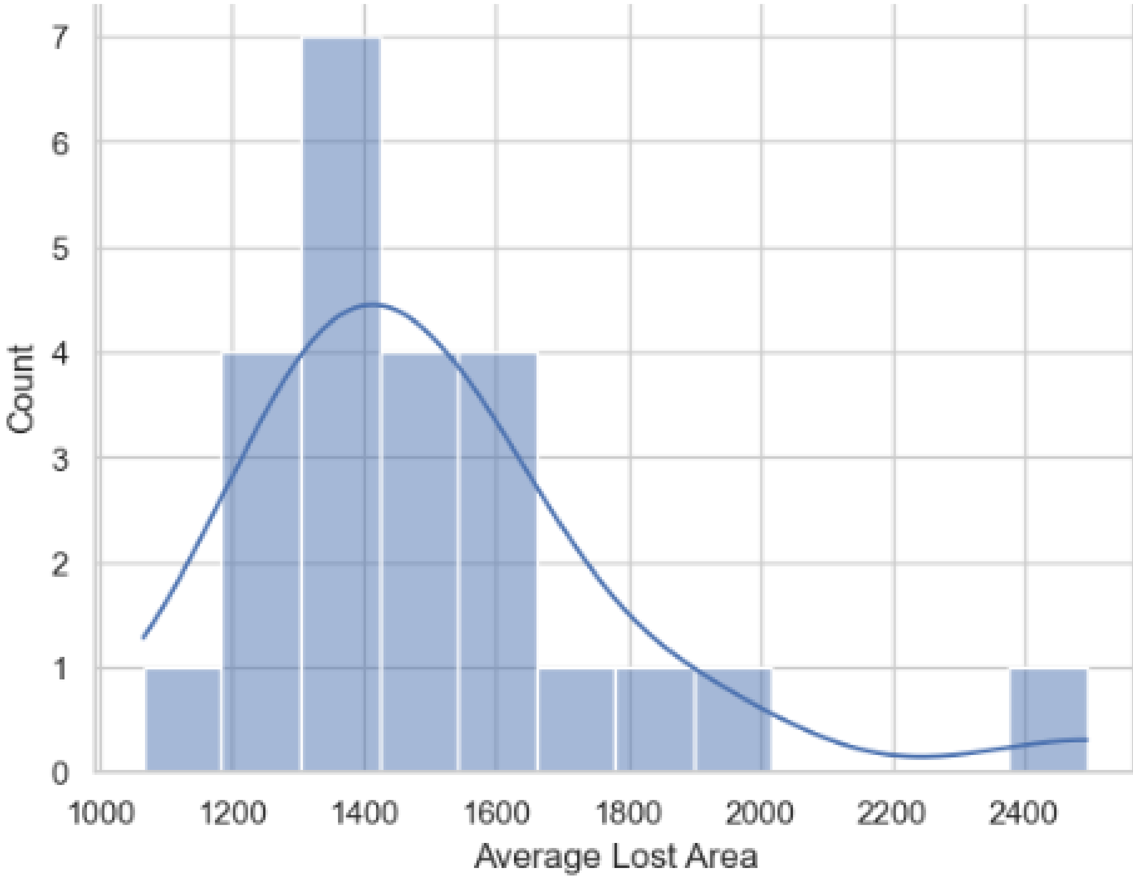

Figure 4.3. *Scenario 2 (6000 Owners) - Area loss distribution. Sacrifices are moderate and equitably spread.*

Figure 4.4 presents the **correlation matrix**. Here, `delta_mean_area` correlates moderately with `total_area_lost` ($r \approx 0.44$) and more strongly with `percentage_donos_com_perda` ($r \approx 0.78$). This suggests that wider participation in contributions tends to improve global outcomes, an insight that is especially relevant for policy decisions regarding incentivization and fairness.

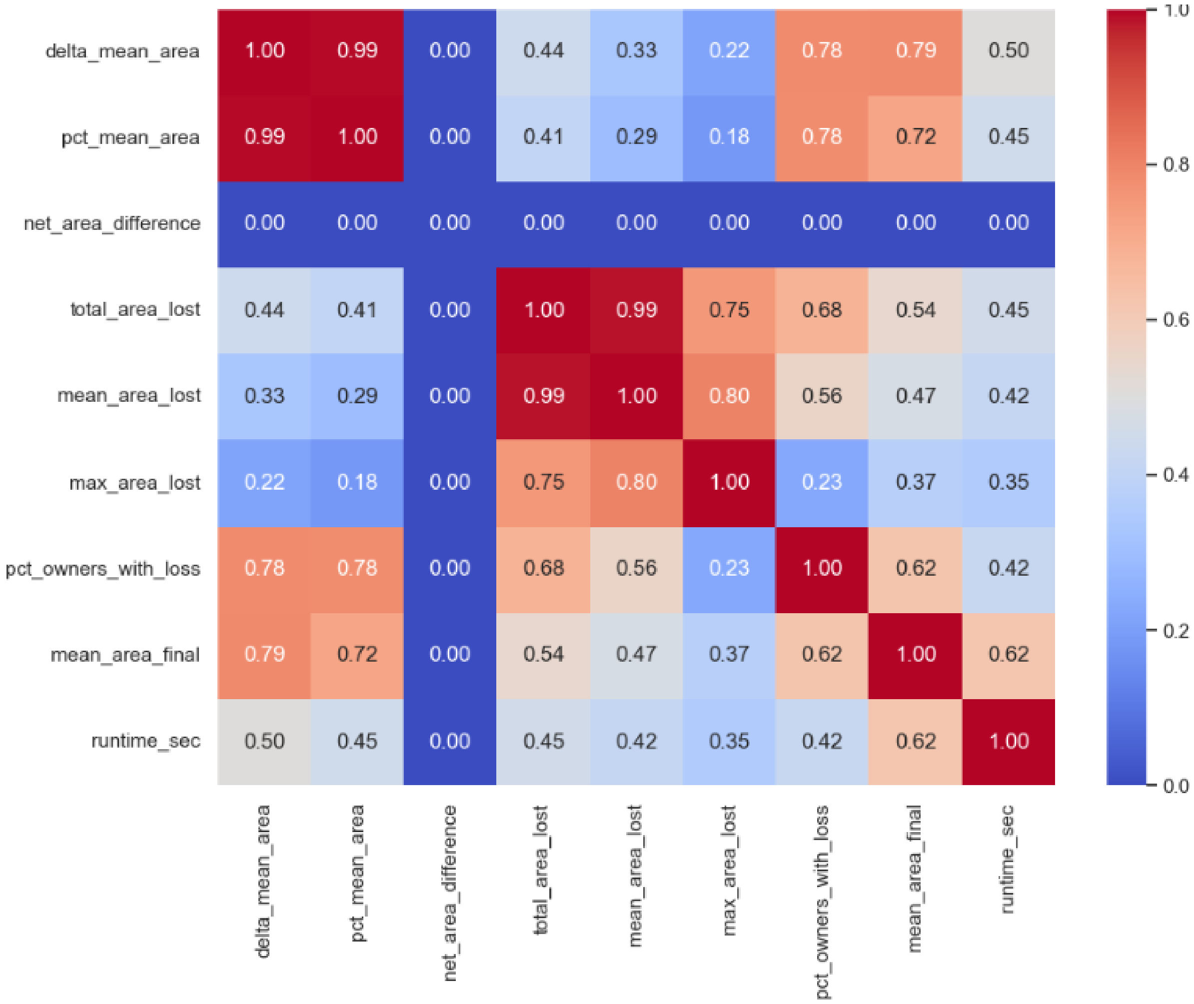

Figure 4.4. *Scenario 2 (6000 Owners) - Correlation matrix. Increased participation is associated with greater consolidation gains.*

As shown in Figure 4.5, the results illustrate the distribution of the simulated solutions, highlighting the empirical relationship between the total lost area and the mean consolidated area. Compared to Scenario 1, the trade-off curve appears more pronounced, indicating a sharper balance between land losses and consolidation gains.

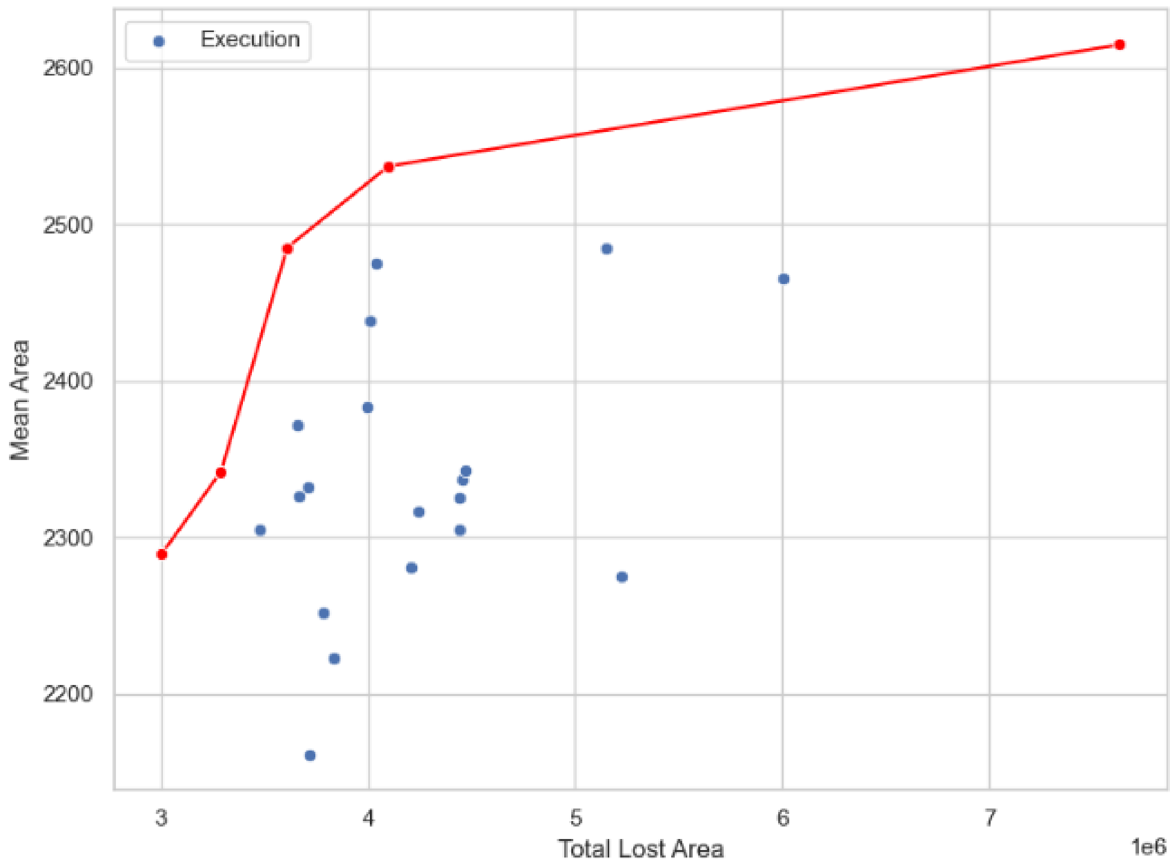

Figure 4.5. *Scenario 2 (6000 Owners) - Distribution of simulation results. The red line connects the best-performing simulations.*

4.3.3. Scenario 3: 10000 Owners

The final scenario simulates a highly distributed ownership pattern with 10,000 `owners`. Despite the larger population, the algorithm produced stable gains (mean increase of 11.0%) with lower average sacrifices, highlighting its **robustness in large-scale settings**.

The average area lost per contributing owner was around 950 units - the lowest in all scenarios ,while more than 55% of the `owners` retained or expanded their holdings. Figure 4.6 shows that losses are small and well distributed, confirming the equitable design **of the system**.

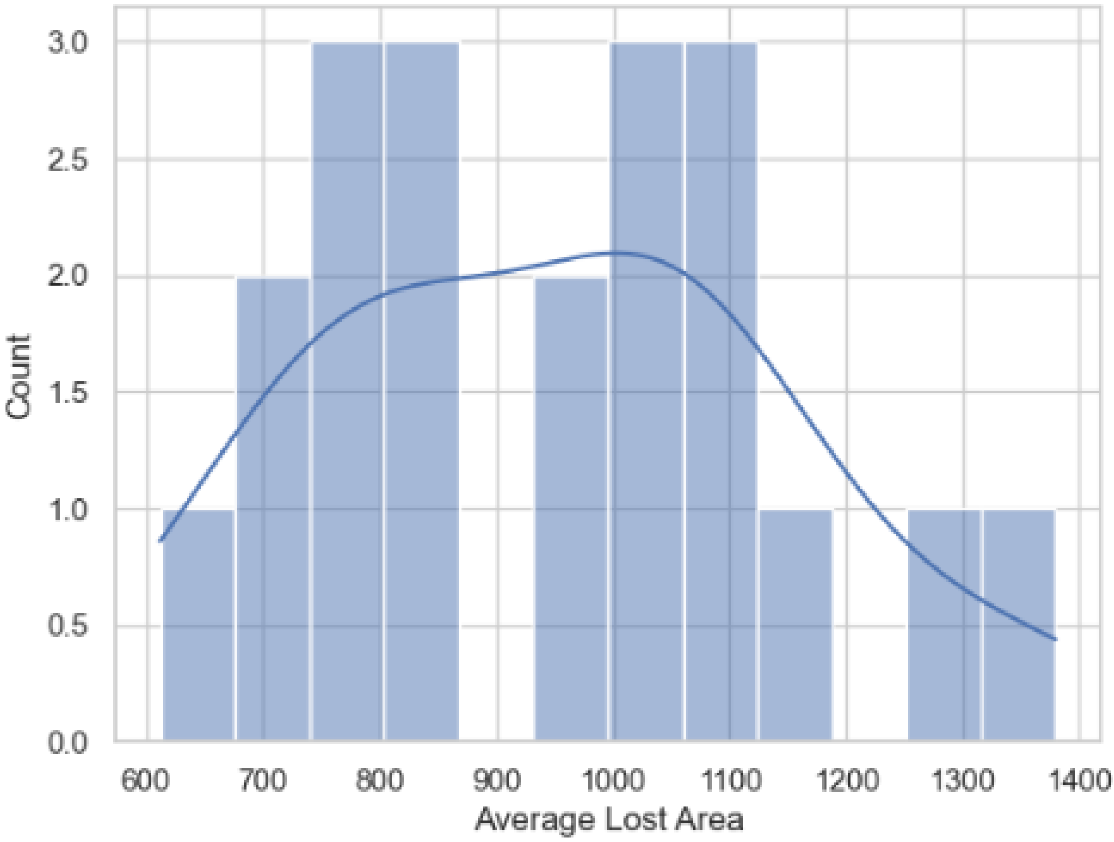

Figure 4.6. *Scenario 3 (10000 Owners) - Area loss distribution. Minimal per-owner sacrifice, widely distributed.*

Figure 4.7 presents weak correlations between `delta_mean` and the loss-related indicators, suggesting that additional sacrifices bring diminishing marginal returns. The system self-regulates toward balanced outcomes without requiring extensive intervention.

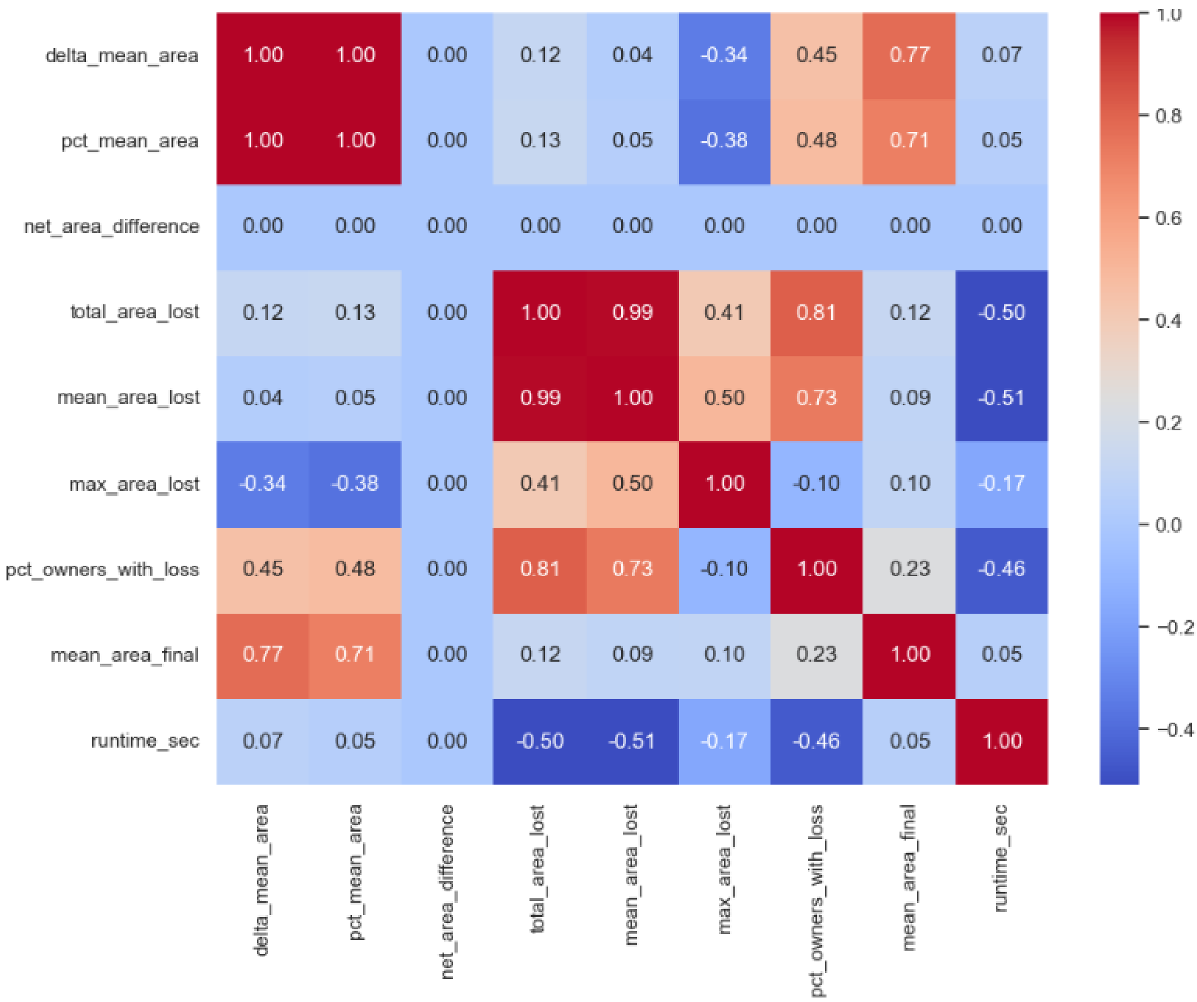

Figure 4.7. *Scenario 3 (10000 Owners) - Correlation matrix. Weak associations imply an emergent balance.*

Finally, Figure 4.8 illustrates the distribution of simulation results, strengthening the presence of diminishing returns. Although higher total losses tend to yield larger mean consolidated areas, the upper envelope flattens, indicating that excessive land removal offers limited additional benefits.

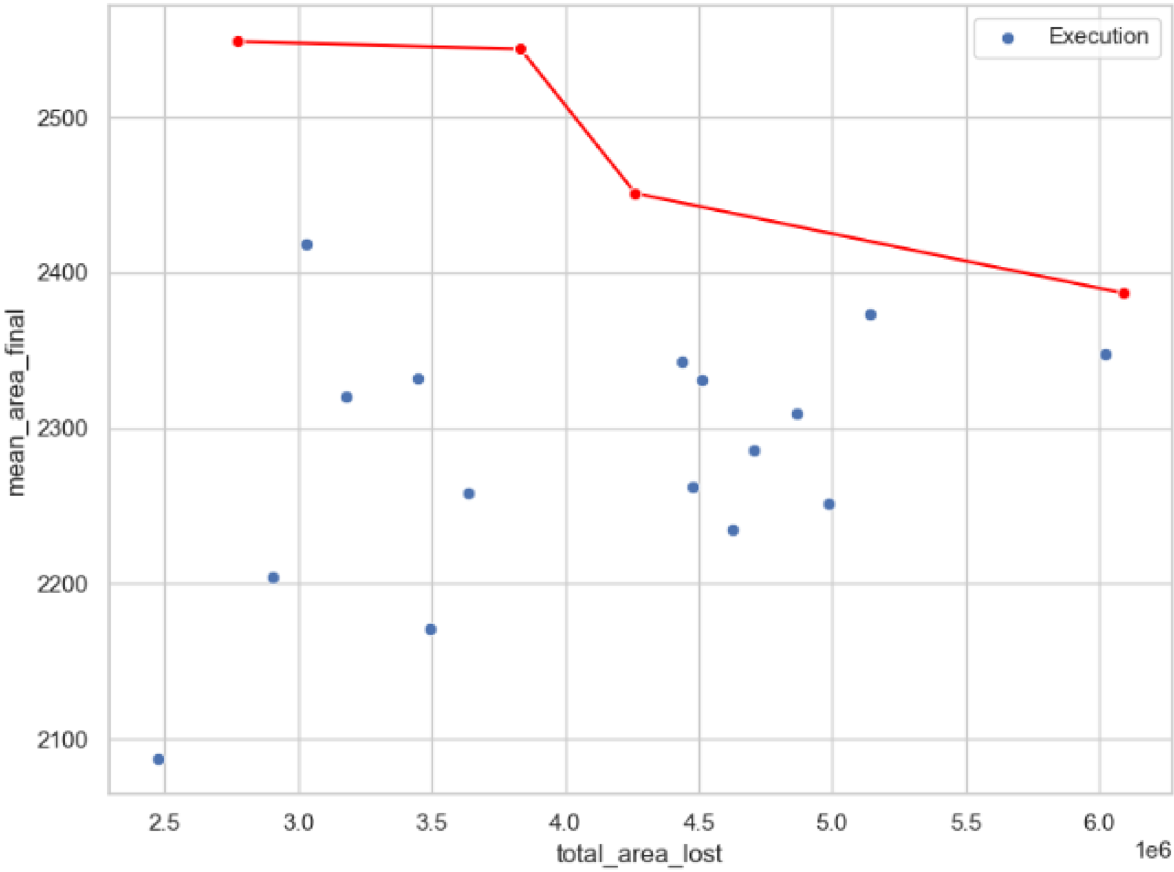

Figure 4.8. *Scenario 3 (10000 Owners) - Distribution of simulation results. The red line connects the best-performing simulations.*

4.3.4. Comparative Analysis Across Scenarios

The comparative evaluation of the three experimental scenarios reveals critical insights on the consistency, fairness, and capacity to scale of the algorithm. Table 4.3 synthesizes the most relevant outcome indicators in all settings.

Table 4.3. *Average metrics across scenarios*

Metric	3000 Owners	6000 Owners (Real)	10000 Owners
Mean Gain (Δ_{mean}) [units]	231.5	253.3	230.8
Relative Mean Gain (%)	10.9%	12.0%	11.0%
Total Area Lost ($\times 10^6$ units)	3.69	4.26	4.15
Mean Area Lost per Contributor	2540	1502	951
% of Owners with Loss	48.5%	47.0%	43.3%
Final Mean Area per Owner	2352	2361	2323

Key Comparative Findings:

- **Stable Efficiency Gains:** The algorithm consistently delivers area gains between 10. 9% and 12. 0%, regardless of the scale of the scenario. This reflects a high degree of algorithmic robustness and reproducibility across heterogeneous ownership structures.

- **Improved Equity with Scale:** Although the total amount of land removed from the system increases slightly in absolute terms, the burden per contributing owner decreases significantly. This trend, from 2540 units (3,000 `owners`) to only 951 units (10,000 `owners`), suggests that larger populations enable **softer individual sacrifices** through distributed contribution.
- **Higher Inclusiveness in Larger Scenarios:** The percentage of `owners` who avoid losses increases from 51. 5% to 56. 7% in all scenarios. This pattern points to increasing inclusion and a more socially acceptable redistribution process as the scale increases.
- **Pareto-Convergent Solutions:** In all three scenarios, most of the solutions fall close to the Pareto frontier. The 6,000-owner (realistic) case displays the most balanced behavior, with visible trade-offs and moderate variability, suggesting an ideal compromise between consolidation benefit and land sacrifice.
- **Fragmentation Remains Controlled:** Fragmentation indicators in Table 4.4 reveal that the average cluster counts remain highly stable in all settings. Although variability increases slightly, the mean remains constant, indicating that the algorithm preserves spatial cohesion even under complex territorial conditions.

Table 4.4. *Fragmentation indicators per scenario*

Indicator	**3000 Owners**	**6000 Owners (Real)**	**10000 Owners**
Average Clusters	31 680	30 991	31 270
Median	31 748	31 336	30 917
Standard Deviation	170	773	897

Scenario-Level Summary:

- *3000 Owners:* Delivers strong reproducibility and consistent optimization, but at the cost of higher individual sacrifices.
- *6000 Owners:* Achieves the most favorable balance between gains and losses, making it the most **efficient and equitable** configuration overall.
- *10000 Owners:* Minimizes average losses and maximizes inclusiveness, though with slightly more dispersion in results, potentially due to lower consolidation potential at the outset.

Critical Perspective:

These results collectively affirm the algorithm's **scalability, fairness, and adaptability**. As the number of `owners` increases, the model maintains performance while lowering individual costs, a particularly desirable feature for public policy implementation. However, the slightly increased variance in the largest scenario suggests that future research might explore more adaptive parameter tuning to ensure consistency in very large-scale consolidations.

In summary, this comparative evaluation confirms that the proposed method is not only technically effective but also socially responsive, offering robust support for consolidation initiatives in diverse real-world contexts.

4.4. Implications for Real-World Policy

Although the simulations presented in this chapter are based on synthetic ownership structures, their results provide valuable guidance for the real-world implementation of algorithmic land consolidation. Consistent area gains, equitable sacrifice patterns, and stable behavior in diverse scenarios underscore the practical feasibility of this method under varying territorial conditions.

Scalability and Adaptability: As demonstrated in Section 4.3.4, the algorithm maintains its performance as the number of `owners` increases, without requiring structural changes to the optimization logic. This robustness across scales suggests potential for deployment in both small community-driven initiatives and large-scale national restructuring programs. The limited increase in per-owner sacrifice with rising population, observed in all scenarios, further reinforces its capacity to operate under socially acceptable thresholds.

Feasibility with Real Data: The data requirements of the method, as discussed in Chapter 3, are intentionally minimal: basic parcel geometries and owner identifiers are sufficient to simulate reallocation processes. In many countries, such datasets already exist through cadastral registers, municipal databases, or satellite-assisted parcel mapping. Where formal data are incomplete or outdated, participatory mapping initiatives and crowd-sourced surveys have proven effective in reconstructing usable baselines.

Simulation as a Planning Tool: Simulations play a critical role in testing policy options without real-world consequences. They allow decision-makers to explore trade-offs, such as consolidation depth versus owner equity, and assess the sensitivity of outcomes to legal or social constraints. In addition, as shown in this study, metrics such as Pareto frontiers and loss distributions can support transparent stakeholder engagement by anticipating the range of individual impacts.

Digitalization Gaps and Informal Systems: Even in settings with limited digital infrastructure or informal land tenure systems, the proposed approach can offer value. Regions with low cadastral coverage, such as parts of Sub-Saharan Africa, South Asia, or post-conflict zones, often face challenges in deploying traditional consolidation programs. In these contexts, synthetic simulations can serve as planning guides for gradual formalization, helping prioritize interventions and structure incentives. The low dependence of the algorithm on high-resolution data makes it particularly suitable for such transitional environments.

Institutional and Legal Considerations: Despite its technical feasibility, adoption in the real-world must be accompanied by legal and institutional safeguards. Issues such as voluntary participation, rights restitution, and dispute resolution should be included in the policy framework. In addition, community acceptance may depend on prior consultations, transparency of reallocation logic, and the provision of compensatory mechanisms for loss-bearing participants. These aspects fall outside the algorithmic scope, but can be directly informed by its outputs.

Supporting Policy Co-Design: By quantifying expected distributions of benefit and sacrifice, this method can support the co-design of land policy instruments. For example, equity indicators can guide eligibility criteria for subsidies; Pareto curves can help define efficiency thresholds; and simulation outputs can inform community engagement materials. As argued in Chapter 2, data-driven tools are not substitutes for participatory governance, but they can significantly enhance clarity, predictability, and accountability in decision-making.

Conclusion: In sum, the consolidation algorithm presented in this thesis demonstrates not only computational soundness but also promising alignment with policy realities. Its ability to generate fair, scalable, and interpretable results, even in imperfect data environments, position it as a viable complement to traditional land management strategies. Although legal adaptation and social integration remain critical prerequisites, the findings in this chapter reaffirm the transformative potential of simulation-based approaches for land governance.

4.5. Summary

This chapter has demonstrated that scalable land consolidation is not only technically feasible, but also socially balanced. Through structured simulations in three ownership scenarios, the proposed algorithm consistently achieved robust, equitable, and interpretable results.

In the 3,000-owner scenario, gains were high and stable, though the burden of loss was more concentrated among fewer contributors. The case of 6,000 `owners` - representing a realistic Portuguese distribution - offered the most balanced result in terms of efficiency, fairness, and fragmentation control. Finally, the 10,000-owner scenario confirmed the method's scalability, showing that as the system scales, individual sacrifices become smaller and more evenly distributed.

One critical insight emerging from this analysis is that **equity improves with scale**. As the number of participants increases, both the percentage of affected `owners` and the average loss per contributor decrease, a dynamic that improves social acceptability and reinforces the potential of the algorithm for real-world application.

These findings set the foundation for the final chapter, which synthesizes key conclusions, reflects on practical deployment challenges, and outlines directions for future work in data-driven land governance.

CHAPTER 5

Discussion and Policy Implications

5.1. Critical Analysis of Results in Light of the Research Questions

This research was set up to investigate whether a data-driven, algorithmic approach could contribute meaningfully to the land consolidation process in rural contexts, particularly in Portugal. The investigation was guided by three main research questions, each of which is critically addressed in the following, in light of the findings presented in Chapter 4.

Research Question 1: Can a decision support system based on graph modeling effectively assist in generating land consolidation scenarios?

The findings provide strong evidence that the proposed DSS can automate the generation of land exchange scenarios that improve structural indicators of landholding. The performance of the system, tested in a real-world case study, consistently increased the average area per owner and reduced the number of disconnected land blocks. For example, optimization led to gains of approximately 10–11% in the average parcel size and a reduction of fragmentation by nearly half. These quantitative results demonstrate that the graph-based approach is not only technically feasible but also capable of delivering meaningful improvements in spatial structure.

However, the effectiveness of the system is highly dependent on the quality of the input data and the assumptions embedded in the model. For instance, no real cadastral inaccuracies or legal constraints were considered, which may limit its direct applicability in complex real-life scenarios. Despite this, the results are promising indications of the potential of this method.

Research Question 2: Is it possible to balance multiple, potentially conflicting objectives (e.g., increasing area, reducing fragmentation, minimizing land loss) within the same decision model?

The multi-objective optimization strategy adopted in the system successfully generated solutions that reflect different trade-offs among competing goals. Rather than imposing fixed priorities, the model explored a solution space that allowed decision makers to choose scenarios based on preferred balance points. This flexibility is particularly important in land consolidation, where stakeholder preferences and policy constraints can vary.

In practice, the results showed that it is possible to achieve improvements in all objectives simultaneously, though not always to the same degree. For example, reducing fragmentation often implied a modest increase in area loss. The capacity to generate and compare such trade-offs transparently is a central strength of the proposed system. However, the system currently does not model qualitative objectives such as social equity or ecological preservation, which limits its scope.

Research Question 3: Can the decision support system generate interpretable and transparent outputs suitable for use in policy and planning contexts?

Transparency and interpretability were embedded in the architecture of the system from the beginning. Each stage of the algorithm, from initial graph modeling to the generation of optimized exchange plans, is traceable and explainable. This makes the system compatible with governance processes that demand accountability and fairness.

In addition, the output includes metrics and comparative visualizations that help stakeholders understand the implications of each scenario. This characteristic aligns with the goals of Portugal's *Emparcelar para Ordenar* initiative, which emphasizes voluntary participation and trust in the consolidation process.

Nonetheless, the full potential of the system in real policy contexts remains to be tested. The case study confirmed its technical viability, but a broader stakeholder engagement and institutional integration would be necessary to validate its practical value.

In general, the research questions were answered affirmatively, with empirical support and clear evidence of impact. At the same time, the analysis acknowledges that real-world implementation would require adaptations to account for legal, social, and institutional complexities that were beyond the scope of this study.

5.2. Evaluation of Objectives and Limitations

The primary objective of this dissertation was to design, implement, and evaluate a decision support system capable of assisting land consolidation processes through the use of graph-based modeling and multi-objective optimization. This section critically assesses the extent to which the defined objectives were achieved while also reflecting on the main limitations that emerged during the development and application of the system.

First, the development of a prototype decision support system was completed successfully. The system integrates geospatial land data with graph-based structures to model the relationships between parcels and owners. A multi-objective optimization algorithm was implemented to generate exchange scenarios that consider different goals simultaneously, such as maximizing parcel size, minimizing fragmentation, and reducing land loss. This implementation represents a concrete and operational contribution, confirming that it is possible to create a practical tool to assist in complex land planning challenges.

Second, the system was applied to a real-world case study, validating its functionality under realistic conditions. The empirical results confirmed that the system could generate feasible and beneficial consolidation solutions within a defined geographic area. This demonstrated the potential practical value of the approach and fulfilled the goal of evaluating the system beyond theoretical scenarios.

Third, the dissertation aimed to provide transparency and traceability throughout the decision-making process, a requirement for real-world policy integration. This objective was also met:

the modular structure of the system and the clarity of the optimization outputs enable users and decision-makers to understand the steps involved and the reasoning behind each scenario.

Despite these achievements, several limitations emerged that restrict the scope and generalizability of the study. The system was tested in a single geographic area, which, while suitable for initial validation, does not allow for broad conclusions about its performance in diverse land contexts. Future applications in regions with different ownership patterns, legal frameworks, or landscape characteristics may require adaptations.

Furthermore, the model made several simplifying assumptions, such as perfect data quality, full voluntary participation by landowners, and the exclusion of socioeconomic or environmental constraints. These assumptions were necessary to ensure tractability, but may reduce the realism of the scenarios produced. The algorithm also did not incorporate factors such as land value, access, or ecological preservation, which are often relevant in actual land consolidation processes.

In summary, while the core objectives of the study were successfully met, the limitations highlighted here define clear boundaries for the conclusions that can be drawn. They also point to important areas for future enhancement and broader testing, which are discussed in the next section.

5.3. Future Work and Final Remarks

Although the developed decision support system has demonstrated its potential to improve land consolidation processes, the research also opens several promising directions for future work. These directions aim to strengthen the robustness, realism, and applicability of the proposed approach, as well as to broaden its contribution to both research and practice in land management.

First, future research should focus on expanding the application of the system to different geographic and socioeconomic contexts. The current study relied on a single case study to validate the methodology, which was appropriate to test feasibility but limits generalization. Applying the model to other regions, particularly those with diverse land tenure systems or varying levels of cadastral completeness, would help assess its adaptability and scalability. Such comparative analyses could reveal how ownership structures, parcel patterns, or policy frameworks influence the effectiveness of optimization-based land consolidation.

Second, the model could be improved by incorporating additional dimensions beyond spatial structure. Future versions of the decision support system could integrate variables such as land value, soil quality, accessibility, or environmental constraints. Including these aspects would allow for a more holistic evaluation of consolidation outcomes, balancing not only efficiency and fairness, but also economic and ecological sustainability. Inclusion of owner preferences or behavioral components could also make the system more realistic, as land consolidation often depends on voluntary participation and perceived fairness.

Third, further work should aim to improve the optimization algorithm itself. Although the current heuristic approach proved effective and interpretable, there is potential for integrating advanced metaheuristic or hybrid methods to explore the solution space more efficiently, especially in larger datasets. Incorporating adaptive parameters or learning-based techniques could enhance performance without compromising transparency. Similarly, developing a user-friendly interface with visual analytics capabilities would facilitate the adoption by policy makers and planners, bridging the gap between computational modeling and decision making.

From a policy perspective, the system could be piloted within real institutional frameworks, such as Portugal's *Emparcelar para Ordenar* initiative. Testing the tool in collaboration with government agencies or local communities would provide valuable feedback on usability, trust, and the social dynamics of implementation. This would represent an essential step toward transforming the system from an academic prototype into a functional policy instrument.

In conclusion, this dissertation has shown that a data-driven, graph-based approach can make a meaningful contribution to addressing rural land fragmentation. The developed system achieved tangible improvements in land structure while maintaining fairness and interpretability, qualities essential for real-world policy use. Although further refinements are needed, the research provides a solid foundation for future innovation at the intersection of geospatial analytics, optimization, and land governance.

Bibliography

[1] (2007). Decreto-lei nº 87/2007.

[2] (2012). Decreto-lei nº 195/2012.

[3] (2022). Regulamento delegado (ue) 2022/1172 da comissão.

[4] Agência para a Modernização Administrativa (2025). Zonas de intervenção florestal - dados.gov.pt. Acedido em agosto de 2025.

[5] Aleta, A. and Moreno, Y. (2019). Multilayer networks in a nutshell. *Annual Review of Condensed Matter Physics*, 10:45–62.

[6] Alijo, M., Abdoun, O., Bachran, M., and Bergam, A. (2020). Optimization by hybridization of a genetic algorithm with the promothee method: Management of multicriteria localization. *arXiv preprint arXiv:2002.04068*.

[7] Almeida, F. and Santos, J. (2023). Governing with intelligence: The impact of artificial intelligence on public policy. *Information*, 15(9):556.

[8] Alqahtani, G. D. et al. (2014). A computational comparison between optimization techniques for wells placement problem: Mathematical formulations, genetic algorithms and very fast simulated annealing. *Journal of Materials Science and Chemical Engineering*, 2(10):59.

[9] Correia, M. M. M. (2010). *Avaliação Ambiental Estratégica: Aplicação aos Planos Municipais de Ordenamento do Território*. PhD thesis, Faculdade de Ciências e Tecnologia.

[10] de Agricultura e Desenvolvimento Rural, D.-G. (2020). Emparcelar para ordenar. Accessed: 2025-01-06.

[11] de Beires, R. S. (2013). *O Cadastro e a Propriedade Rústica em Portugal*. Fundação Francisco Manuel dos Santos.

[12] de Financiamento da Agricultura e Pescas (IFAP), I. (2025a). Quem somos. Accessed: 2025-01-06.

[13] de Financiamento da Agricultura e Pescas (IFAP), I. (2025b). Sip - informação básica. Accessed: 2025-01-06.

[14] Direção-Geral de Agricultura e Desenvolvimento Rural (2022). Estudo de avaliação da fragmentação da estrutura fundiária. Accessed: 2025-08-03.

[15] Direção-Geral do Território (2025). Programa emparcelar para ordenar. Acedido em agosto de 2025.

[16] FAO (2020). Land fragmentation: a barrier to sustainable rural development. Accessed: 2025-08-03.

[17] Instituto Nacional de Estatística (2019). Censos 2019 - agricultura: Caracterização da estrutura fundiária. Accessed: 2025-08-03.

[18] Laudon, K. C. and Laudon, J. P. (2003). *Essentials of Management Information Systems: Managing the Digital Firm*.

[19] Mafra, F. and Silva, J. A. (2004). *Planeamento e Gestão do Território*. Sociedade Portuguesa de Inovação.

[20] Oliveira, F. P. (2019). *Direito do Urbanismo: Do Planeamento à Gestão*. Braga Aedrl, 3 edition. Carta Europeia de Ordenamento do Território, aprovada pelo Conselho da Europa em 1983.

[21] OpenAI (2025). Chatgpt (versão gpt-5). Ferramenta de inteligência artificial generativa. Disponível em `https://chat.openai.com`. Acedido em setembro de 2025.

[22] Papudo, R. M. (2007). *Indicadores de Ordenamento do Território: Uma Proposta Metodológica*. PhD thesis, PhD Thesis.

[23] Pereira, M. (2009). Desafios contemporâneos do ordenamento do território: para uma governabilidade inteligente do(s) território(s). *Prospectiva e Planeamento*, 16:77–102.

[24] Portugal, E. (2022). O que são os sig? Accessed: 2025-01-06.

[25] Qian, K., Mao, L., Liang, X., Ding, Y., and Gao, J. (2020). Ai agent as urban planner: Steering stakeholder dynamics in urban planning via consensus-based multi-agent reinforcement learning. *arXiv preprint arXiv:2002.04068*.

[26] Shearer, C. (2000). The crisp-dm model: The new blueprint for data mining. SPSS White Paper.