

PAULO CÉSAR RODRIGUES DE LIMA JÚNIOR

Integration of Geographic Information
Systems, Meta-Heuristics and Multi-Criteria
Analysis for Territories Alignment

Dissertação apresentada à Faculdade de Engenharia da Universidade do Porto para a
obtenção do grau de Doutor em Ciências de Engenharia

Orientador: Professor Doutor Jorge Rui Guimarães Freire de Sousa
Co-Orientador: Professor Doutor José António de Sousa Barros Basto

Faculdade de Engenharia da Universidade do Porto

2008

Resumo

Esta tese propõe uma nova abordagem para o problema de balanceamento de territórios usando uma meta-heurística híbrida multi-critério. O problema de balanceamento de territórios é parte de um processo maior, o projecto de territórios, que consiste em atribuir pequenas regiões geográficas às áreas maiores, seguindo os critérios mais relevantes para o planeamento. O principal objectivo do projecto de territórios é a atribuição, às forças de vendas/serviços, de clientes e suas actividades de venda/serviço associadas, procurando obter eficácia e eficiência na fase de operação de um determinado serviço e levando a um aumento da produtividade, das receitas e dos lucros.

O projecto de territórios encontra potencial aplicação num grande número de áreas, por exemplo, a criação de distritos políticos (eleitorais), áreas de escolas, recolha de lixo, serviços sociais (centros de saúde, hospitais, etc), serviços de emergência, venda e distribuição de produtos e equipas de manutenção. O planeamento de territórios das forças de vendas/serviços é altamente complexo, pois procura considerar objectivos que são por vezes antagónicos e deve tentar satisfazer os anseios das principais partes envolvidas: a empresa, os seus clientes e as forças de vendas/serviços.

Com o intuito de procurar integrar estes objectivos conflituosos, foi desenvolvida uma abordagem multi-critério baseada nas meta-heurísticas GRASP e *Tabu Search* para o problema de balanceamento de territórios. O algoritmo foi integrado numa aplicação interactiva e amigável de Sistemas de Informações Geográficas, construindo o procedimento que apelidamos de MultiACE. Este procedimento foi incorporado no software ArcGIS e, deste modo, os planeadores podem, durante a execução do algoritmo, escolher um conjunto diferente de critérios, alterar os parâmetros do algoritmo, gravar soluções intermédias ou introduzir soluções obtidas anteriormente. Nas suas actividades de planeamento, os decisores são capazes de avaliar e analisar um amplo conjunto de soluções alternativas. Além disso, diversos cenários podem ser testados e avaliados num tempo bastante curto.

Esta tese evidencia ainda o potencial da abordagem proposta como uma ferramenta prática e facilmente implementável de apoio à decisão evidenciando o seu uso no contexto de um caso de estudo real que envolveu a equipa de manutenção da sucursal portuguesa de uma empresa multinacional de equipamentos.

Abstract

This thesis proposes a new approach for the territories alignment problem using a multi-criteria hybrid meta-heuristic. The territories alignment problem is part of a bigger procedure, the territories project, which consists of assigning small geographic regions to larger areas following the most relevant criteria for planning. The main objective of the territories project is the assignment of accounts (clients) and their associated selling/service activities to sales/service forces (SSF), looking for obtaining effectiveness and efficiency in the phase of operation of a given service and leading to increased SSF productivity, revenue and profits.

The territories project has potential application in a lot of areas, for instance, the establishment of political districts (electoral), areas of schools, trash collection, social services (health centers, hospitals, etc.), emergency services, sales and distribution of products and maintenance teams. The planning of the SSF territories is highly complex, because it should consider objectives that are sometimes antagonistic and we should try to set out the objectives of the main stakeholders: the company, their customers and the SSF.

In order to tackle these conflicting objectives we developed a multi-criteria approach to the territories alignment problem based in GRASP and Tabu Search meta-heuristics. We also integrated the algorithm for the territories alignment problem in an interactive and user-friendly Geographic Information System application, named MultiACE. This application was embedded in the ArcGIS software and, within it, planners can, during the execution of the algorithm, choose a different set of criteria, change parameters of the algorithm, save intermediate solutions or introduce previously obtained solutions. In their planning activities, planners are able to evaluate and analyze a wide set of alternative solutions. Furthermore, different scenarios can be tested and evaluated in a rather short time.

This thesis also demonstrates the potential of the proposed approach as a practical and readily implementable management decision aid in the context of a current case that involved the team of maintenance workers attached to the Portuguese regional office of a worldwide equipments company.

Résumé

Cette thèse-ci propose une nouvelle approche au problème de l'équilibre entre les territoires, en utilisant une méta-heuristique hybride multicritère. Ce problème fait partie d'un processus plus ample, le projet de territoires, qui consiste à attribuer de petites zones géographiques à des zones plus vastes, selon les critères les plus appropriés au planning. Le principal objectif du projet est l'attribution aux forces de ventes/services des territoires liés à leurs clients, en cherchant plus d'efficience et d'efficacité lors de l'opération d'un service et conduisant, ainsi, à une augmentation de la productivité, des recettes et des profits.

Le projet de territoires peut s'appliquer à une large gamme de domaines, par exemple, la création de districts électoraux, la définition de zones pour des écoles, la collecte des déchets, la localisation des services sociaux (centres de santé, hôpitaux, etc.), des services d'urgence, et encore des services de vente et de distribution de produits, aussi bien que des équipes d'entretien. Le planning de territoires est très complexe, parce qu'il faut considérer plusieurs objectifs, parfois contradictoires, et aussi essayer de satisfaire les aspirations des principaux parts concernés: l'entreprise, les clients et les forces de ventes/services.

Avec le but d'intégrer ces objectives contradictoires, on a développé une approche multicritère fondée sur les méta-heuristiques GRASP et Tabu Search pour le problème de l'équilibre entre les territoires. L'algorithme a également été intégré dans une application de systèmes d'information géographique (SIG), interactive et facile à utiliser, qu'on a appelée MultiACE. Ce développement a été fait sur le software ArcGIS. En l'utilisant, les planificateurs peuvent, au cours de l'exécution de l'algorithme, choisir d'autres critères, changer les paramètres de l'algorithme, forcer des solutions intermédiaires ou d'autres obtenues précédemment. Dans leurs activités de planification, les décideurs peuvent évaluer et examiner un large éventail de solutions alternatives. En outre, plusieurs scénarios peuvent être testés et évalués dans un laps de temps très court.

Dans cette thèse le potentiel de l'approche proposée est encore mis en évidence – tandis qu'outil d'aide à la décision pratique et facile à mettre en œuvre – dans le cadre d'une étude de cas réelle, impliquant les équipes d'entretien du bureau portugais d'une entreprise multinationale d'équipements.

Acknowledgments

This thesis would never have come to fruition without the support I have received from many individuals and organizations. Acknowledgment and thanks must be given to a large number of people, many more than can fit onto a couple of pages. The words and feelings expressed in this Acknowledgment express my sincere appreciation to all of those who have provided invaluable assistance and support during my postgraduate study. The order of appearance of individuals and organizations does not reflect in any way the level of importance of their contribution to the success of the research presented in this Thesis.

First, thanks must go to the Program Alban, for their provision of a scholarship that provided the tuition and living allowance which made it possible for me to undertake research for the PhD degree.

Thanks to the Doctorate in Management and Industrial Engineering for their reception and provision of a very good environment and support for my research during all my PhD time.

My heartfelt thanks are extended to my supervisor, Professor Dr. Jorge Rui Guimarães Freire de Sousa, Professor at the Faculty of Engineering of the University of Porto, who initiated the opportunity for commencing this research. I am most grateful for his guidance, continual encouragement, constructive criticism, and the support he gave me in gaining a broader knowledge of developments in my fields of interest. Without his constant advice, encouragement, and suggestions my research would neither have stayed on the right path nor made good progress.

Many thanks also go to my Co-supervisor, Professor Dr. José António de Sousa Barros Basto, Professor at the Faculty of Engineering of the University of Porto, who gave me far more guidance and assistance than that can be expected from a Co-supervisor. His kind help, guidance and discussion about the structure of the Thesis and several theoretical and methodological aspects of the research, his assistance in pointing me in the right direction regarding the empirical analysis and availability of data, and his support in the use of the CPLEX software package used for this research are greatly appreciated and acknowledged.

Thanks also go to Mrs. Maria Soledade Gouveia Bernardo Medeiros and Mrs. Isabel Cristina Borges Pereira Couto Ribeiro, administrative assistants at Department of Industrial Engineering and Management for their support and kind care when we worked together there.

The knowledge, experience and help kindly provided by many experts over the years are gratefully acknowledged. I particularly thank Professor Francisco Manuel Madureira e Castro Vasques de Carvalho, Associate Professor of University of Porto, who gave useful comments and assistance on PhD making processes and research preferences.

My fraternal thanks to Francisco Gaudêncio Mendonça Freires; without his friendship and help, this PhD would never have been started.

My fraternal thanks to Mauro Augusto da Rosa and Dalva Magro, without their friendship and help, this PhD would never have been completed.

My deepest gratitude goes to my son, Arthur Carvalho de Lima. His warm hugs and tears alleviated my guilty feelings for spending too little time with him. His smile and kisses given to me whenever I returned home late provided me with the energy and gave me the peace of mind needed to finish the Thesis. And his love gave me the greatest courage to overcome all the difficulties I encountered during the period of my PhD study.

Lastly, but certainly not least of all, I am very grateful to my wife Cibeles Costa de Carvalho and to my parents Arislene Rodrigues de Lima and Francisco Clayrton de Lima Franklin (*in memoriam*) for their love and continuous understanding and support.

Contents

Resumo	ii
Abstract.....	iii
Résumé.....	iv
Acknowledgments	v
List of Figures.....	x
List of Tables	xi
List of Acronyms	xii
Chapter 1 Introduction	1
1.1 Context of the research	2
1.2 Scope and relevance of the thesis	3
1.3 Objectives of the thesis.....	5
1.4 Structure of the thesis.....	6
Chapter 2 GIS, MCDA, MH and Territories Project.....	7
2.1 Introduction.....	7
2.2 Geographic Information Systems.....	8
2.3 Multi-Criteria Decision Analysis (MCDA)	10
2.3.1 Introduction to MCDA.....	10
2.3.2 Classification of MCDA Approaches	11
2.4 Heuristics and Meta-Heuristics	13
2.4.1 Main Concepts.....	13
2.4.2 GRASP	16
2.4.3 Tabu Search.....	17
2.5 Territories project.....	19
2.5.1 Territories project objectives and constraints	20
2.5.2 Sales/Services Territories (SST) Project.....	21
2.5.3 When is it Time to Make SST Alignments?	24
2.6 Summary.....	25
Chapter 3 Bibliographical Review.....	26
3.1 Introduction.....	26
3.2 GIS.....	26
3.3 Multi Objective, Multi-Criteria and Meta-Heuristics	29
3.4 Territories, Districting and Partitioning.....	32
3.5 Greedy Randomized Adaptive Search Procedures and Tabu Search..	38
3.6 Summary.....	42

Chapter 4	MultiACE - A Multi-criteria Hybrid Meta-heuristic	43
4.1	Introduction.....	43
4.2	The multi-criteria hybrid meta-heuristic procedure.....	43
4.2.1	Voronoi Diagram Based Solutions	44
4.2.2	Two-Sets Partitioning Strategy to Decrease the Search Space.....	45
4.2.3	GRASP Based Candidates List.....	47
4.2.4	Tabu Search Based Memories.....	49
4.2.5	Multi-criteria Technique	49
4.3	Implementation and application	50
4.4	Data.....	51
4.5	Constraints.....	52
4.6	Objectives.....	53
4.7	Algorithm's Parameters.....	54
4.8	Algorithm Structure	58
4.9	Summary.....	58
Chapter 5	Algorithm Testing	59
5.1	Introduction.....	59
5.2	The Sets of Experiments.....	60
5.3	The Datasets.....	60
5.4	The CPLEX Formulation.....	61
5.5	Computational Results.....	62
5.5.1	Tuning Experiments	62
5.5.2	Exhaustive MultiACE <i>vs.</i> CPLEX Experiments	65
5.5.3	Heuristic MultiACE <i>vs.</i> Exhaustive MultiACE	69
5.6	Summary.....	73
Chapter 6	Solving Real Problems	74
6.1	Introduction.....	74
6.2	Company A Case Background.....	74
6.3	The Territories Assignment Process	76
6.4	The Current Territories Structure.....	77
6.5	New Territories Project.....	84
6.5.1	Achievements and Savings.....	88
6.6	Improving the Solution.....	88
6.7	Summary.....	91
Chapter 7	Conclusions and Future Work	92
7.1	Introduction.....	92
7.2	Main achievements of this thesis.....	93
7.3	Future research.....	94
References		96
Appendix A.	GIS Components and ArcObjects.....	102
A.1	GIS Components	102
A.1.1	Hardware	102

A.1.2	Software.....	102
A.1.3	Data	103
A.1.4	People.....	104
A.1.5	Procedures.....	104
A.2	ArcGIS and ArcObjects	105
Appendix B.	Spatial MCDA Analysis.....	107
B.1	Steps of Spatial MCDA.....	107
B.1.1	Problem Definition.....	108
B.1.2	Evaluation Criteria.....	108
B.1.3	Decision Maker's Preferences.....	108
B.1.4	Decision Rules	109
B.1.5	Sensitivity Analysis.....	112
Appendix C.	A SSF Alignment Procedure.....	113
C.1	A Seven-Step SSF Alignment Procedure	113

List of Figures

Figure 1.1 Territories Project Problem.....	1
Figure 4.1 A Voronoi Diagram	44
Figure 4.2 The Two-sets partitioning strategy applied to a 4 and a 3 territories partitions.....	46
Figure 4.3 The levels of partitions to divide 1000 points into 9 territories.....	46
Figure 4.4 MultiACE – Initial Screen – Parameters Definition	52
Figure 4.5 MultiACE – Initial Screen – Solutions Checking.....	52
Figure 4.6 MultiACE - Number of Territories Selection.....	54
Figure 4.7 MultiACE - Layer, Key Field, Solution Field and Subgroup Selection.....	55
Figure 4.8 MultiACE – Criteria Options	55
Figure 4.9 MultiACE – Stopping Criteria	55
Figure 4.10 MultiACE – Ideal Point Distances Measurements.....	55
Figure 4.11 MultiACE - Physical Distances Measurements.....	56
Figure 4.12 MultiACE – Algorithm’s Memory Parameters	57
Figure 4.13 MultiACE – Solutions Preferences	57
Figure 4.14 Two Different Solutions for the Same Problem.....	57
Figure 4.15 MultiACE – Special Parameter	57
Figure 4.16 MultiACE – Types of Compactness Measurement.....	58
Figure 6.1 Company A – Equipments of the Area Under Study	75
Figure 6.2 Supervisor Territory 01	78
Figure 6.3 Supervisor Territory 02	78
Figure 6.4 Supervisor Territory 03	78
Figure 6.5 Supervisor Territory 04	78
Figure 6.6 Supervisor Territory 05	78
Figure 6.7 Supervisor Territory 06	78
Figure 6.8 Supervisor Territory 07	78
Figure 6.9 Supervisor Territory 09	78
Figure 6.10 Supervisor Territory 11	78
Figure 6.11 Supervisor Territory 12	78
Figure 6.12 Supervisors Territories – Current Balance	80
Figure 6.13 Overlap of the Territories of Supervisors 01, 02, 03 and 12.....	80
Figure 6.14 The 3 Different Areas	82
Figure 6.15 Route in Area 1.....	82
Figure 6.16 Route in Area 2.....	83
Figure 6.17 Route in Area 3.....	83
Figure 6.18 New Supervisors Territories Structure.....	87
Figure 6.19 Supervisor Territory 07 – Maintenance Workers’ Territories	87

List of Tables

Table 5.1 Tuning Experiments - Short-term Memory - Parameters Setting.....	63
Table 5.2 Results for the Short-term Memory Tuning	64
Table 5.3 Short-term Tuning Summary	64
Table 5.4 Tuning Experiments - Time Limit - Parameters Setting	64
Table 5.5 Time Limit Tuning.....	65
Table 5.6 Exhaustive MultiACE vs. CPLEX Experiments – Parameters Setting.....	66
Table 5.7 Number of Minimum and Maximum Territories for The Datasets	66
Table 5.8 Example of The Runs for One Criteria Set.....	67
Table 5.9 Number of Runs Summary	68
Table 5.10 Results Comparison Summary	68
Table 5.11 Heuristic MultiACE vs. Exhaustive MultiACE – Parameters Setting.....	69
Table 5.12 Number of Runs for each Criteria Set and each Bounds Range	70
Table 5.13 Results Comparison Summary	70
Table 5.14 Example of the Runs for One Criteria Set	71
Table 5.15 Example of the Runs for Three Criteria Set	72
Table 6.1 Current Supervisors Territories.....	79
Table 6.2 Supervisors Territories – Current Balance	79
Table 6.3 Technicians Territories – Current Balance – Study Area	81
Table 6.4 Three Areas – A Comparative Analysis	83
Table 6.5 Supervisors Territories – Current Balance	85
Table 6.6 New Territories Structure – Maintenance Workers.....	86
Table 6.7 Improving the Solution – Number of Territories.....	89
Table 6.8 Improved New Territories Structure – Maintenance Workers.....	90

List of Acronyms

AHP	Analytical Hierarchy Process
AMP	Adaptive memory Programming
CCD	Census collection districts
CCP	Capacitated clustering problem
CO	Combinatorial optimization
DM	Decision maker
GA	Genetic algorithms
GIS	Geographic Information Systems
GRASP	Greedy Randomized Adaptive Search Procedure
MADM	Multi-attribute Decision Making
MAUP	Modifiable areal unit problem
MDBP	Maximum distance between points
MDTC	Mean of sum of Euclidean distances to the centroid of the territory
MOCO	Multiple objective combinatorial optimization
MODM	Multi Objective Decision Making
OWA	Ordered Weighted Average
RCL	Restricted Candidate List
SA	Simulated Annealing
SAW	Simple Additive Weighting
SLA	Statistical local areas
SSF	Sales/Services Forces (teams)
SST	Sales/Services Territories
TAP	Territories Alignment Problem
TS	Tabu Search
VBA	Visual Basic for Applications
WGP	Weighted Goal Programming

Chapter 1

Introduction

"We are at the very beginning of time for the human race. It is not unreasonable that we grapple with problems. But there are tens of thousands of years in the future. Our responsibility is to do what we can, learn what we can, improve the solutions, and pass them on."

Richard Feynman (1918 - 1988) US educator & physicist

The research presented in this thesis addresses the application of multi-criteria hybrid meta-heuristics to the territories alignment problem. The territories alignment problem is part of a bigger procedure, the territories project, which consists of assigning small geographical regions to larger areas following the most relevant criteria for planning. The territories project problem is, traditionally, divided in four sub-problems, as presented in Figure 1.1.

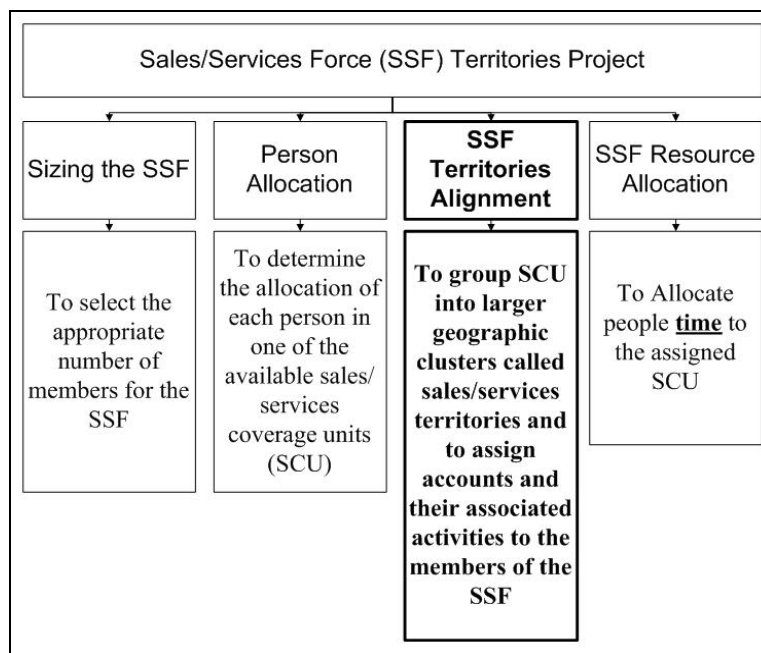


Figure 1.1 Territories Project Problem

The territories project has potential application in a lot of subjects, for instance, the establishment of political districts (electoral), areas of schools, trash collection, social services (health centers, hospitals, etc.), emergency services, sales and distribution of products and maintenance teams. A quite interesting evaluation of the several areas of application of the territories project can be found in (Kalcsics et al. 2005).

In this chapter, Section 1.1 describes the context of this work, while Section 1.2 presents its scope and relevance. The main objectives of this work are outlined in Section 1.3 and the structure of the thesis is presented in Section 1.4.

1.1 Context of the research

The main objective of the territories project is the assignment of accounts (clients) and their associated selling/service activities to SSF, looking for obtaining effectiveness and efficiency in the phase of operation of a given service and leading to increased SSF productivity, revenue and profits. Many sales and services companies use true armies of people to execute several activities related to the marketing area that are the main connection between these companies and their customers.

These SSF should be seen as being crucial for the profitability of the companies, because they are responsible for generating income. At the same time, they generate substantial operating expenses.

The planning of the SSF territories (SST) is highly complex, because it should consider objectives that are sometimes contradictory with each other and the planners should try to set out the objectives of the main stakeholders: the company, their customers and the SSF. The analysis of an immense amount of data, obtained from the most varied sources, is also necessary. Due to this complexity, we need powerful computational tools, allied to methods and heuristics especially developed for the solution of this type of problem.

The shape, alignment and balance of SST largely determine their performance. A study in the USA has shown that 80% of companies are missing 2-7% of unrealized sales because their territories are misaligned (Analytics 2006). According to the same study, poorly aligned and unbalanced territories have negative consequences, including:

- Missed revenue due to not reaching all potential customers and prospects;

- Higher costs due to excessive travel times and travel expenses;
- Inefficiencies due to duplication of efforts among SSF members;
- Low morale and high SSF turnover due to inequitable territories that reward the size of the territory rather than individual effort;
- Poor or excessive customer coverage due to territories too large or too small.

On the contrary, optimally aligned territories that are contiguous and compact in size will help the company to achieve a number of measurable and sustainable benefits including:

- Better customer coverage and customer service leading to increased productivity and revenue;
- Increased sales by prioritizing accounts with the most potential;
- Reduced costs through shorter travel times - often called “windshield time” - and associated travel expenses;
- Improved morale, performance and tenure of sales people due to equitable distribution of accounts and a fair playing field for achieving rewards;
- Competitive advantage through the ability to reach new opportunities faster than their competitors.

There are many ways to assign accounts to SSF members. Since the problem grows exponentially with additional accounts and SSF members, one can imagine the enormity of the task for any reasonably sized SSF. For large, complex SSF, the problem is immense.

The territories optimization software packages use algorithms that evaluate millions of potential alignments to find one that best meets specified criteria regarding territory profitability, workload, sales potential, size, travel, and/or disruption (Zoltners et al. 2004). This kind of algorithm is the core of this research.

1.2 Scope and relevance of the thesis

The territories alignment problem has been largely studied since the 60's and several models and techniques have been proposed to solve it. Most of these approaches are based on set covering or set partitioning formulations, where the objective function consists in the minimization of the sum of the distances. In most cases, set covering and set partitioning models are solved using integer programming techniques often supported by column

generation methods. More recently, several meta-heuristics have been applied to the set covering and the set partitioning problems with very promising results.

In section 2.5 the most important details of this problem are outlined and in section 3.4 a list of references on the territories partitioning problem is presented.

With this work we expect to contribute to the definition of a new perspective in tackling the territories alignment problem, and to provide an innovative multi-criteria approach for its resolution based on meta-heuristics algorithms.

With this purpose, we have designed and implemented a multi-criteria approach to the territories alignment problem using Greedy Randomized Adaptive Search Procedure (GRASP) and Tabu Search (TS) algorithms. GRASP and TS are currently being successfully applied to several combinatorial optimization problems, as they are capable of both generating high quality solutions to many problems within reasonable computation times and being relatively easy to understand and implement. Besides, GRASP and TS do not impose any particular condition to the objective functions or to the constraints (such as linearity or continuity).

In more recent years, there has been a growing interest in multi-criteria meta-heuristics algorithms applied to combinatorial optimization problems (see Chapter 3), but only a very low number of approaches are related to the territories alignment problem (Ehrgott and Gandibleux 2000b).

In this thesis we also propose a new multi-criteria hybrid meta-heuristic algorithm for the territories alignment problem. This algorithm was integrated in an interactive and user-friendly Geographic Information System application, MultiACE, also developed in the scope of this thesis. This application is embedded in the ArcGIS software, although it can also be used as a stand-alone application. Within this application, planners can, during the execution of the algorithms, choose a different set of criteria, change parameters of the algorithms, save intermediate solutions or introduce previously obtained solutions. In their planning activities, planners are able to evaluate and analyze a wide set of alternative solutions. Furthermore, different scenarios can be tested and evaluated in a rather short time.

1.3 Objectives of the thesis

The main goal of this thesis is to develop and implement a new approach to the territories alignment problem that takes into account the multi-criteria nature of the problem, involving a comprehensive set of constraints and objectives, and that can be effectively used to support decision making by planners.

In order to achieve this quite broad goal we have defined the following six more precise objectives:

1. To analyze in detail the structure of the problem in order to identify the components that may influence its complexity and to evaluate the methodologies for the resolution of the SSF territories definition problems in an urban environment.
2. To design a multi-criteria approach for the territories alignment problem. This approach should easily handle the set of constraints and objectives found in the previous stage. It should also quickly provide good feasible solutions.
3. To validate this new approach with real problems. Since our work has been developed in deep collaboration with companies, our approach should also be compared to the solutions currently implemented in these companies.
4. To implement the new approach in a real context, by developing a software application that can be effectively used by operational teams in their planning work.
5. To propose a methodology that contemplates the possibility of using the synergies among the SSF and the logistics forces of the companies.
6. To use GIS and to elaborate indicators for the territories alignment analysis.

This application should be interactive and user-friendly, allowing the planners to select different combinations of objectives and constraints, to control certain parameters of the algorithms, and to quickly evaluate the quality of the solutions provided. The application should also be available to R&D teams for research purposes.

1.4 Structure of the thesis

The remainder of this thesis is organized as follows:

Chapters 2 and 3 outline the main concepts of multi-criteria decision analysis, GIS and meta-heuristics, discuss some of the traditional approaches to the territories alignment problem and present a literature review on the main subjects used to implement the solution.

Chapter 4 introduces our innovative approach to the territories project problem. We then present the multi-criteria hybrid algorithm we have designed.

Chapter 5 presents and discusses the results obtained in some sets of problems created specifically to laboratory tests. Some solutions obtained in laboratory tests are compared to the optimal solutions obtained using the CPLEX optimization software.

Chapter 6 presents and discusses the results obtained in a set of real problems.

Finally, Chapter 7 summarizes the main contributions of this thesis, and presents some topics for future research.

Chapter 2

GIS, MCDA, MH and Territories

Project

"If knowledge can create problems, it is not through ignorance that we can solve them."

Isaac Asimov (1920 - 1992) US science fiction novelist & scholar

2.1 Introduction

Geographic Information Systems, Multi-Criteria Decision Analysis, Meta-Heuristics and Territories Project are the main subjects combined in this thesis to create a new approach for the territories alignment problem. In this work we present a new GRASP-TS hybrid multi-criteria meta-heuristic implemented in a GIS environment to generate feasible solutions for this kind of problem.

In this chapter we present an overview of these four subjects, with a particular emphasis on their application to SST alignment and multi-criteria optimization problems. In Section 2.2 we present the fundamental concepts and describe the main components of the Geographic Information Systems. Section 2.3 outlines the most important information about multi-criteria decision analysis and Section 2.4 presents some meta-heuristics concepts. In Section 2.5 the main concepts of Territories Project are presented.

2.2 Geographic Information Systems

Geographic Information Systems (GIS) are tools that find their application in many areas of knowledge. According to the Environmental Systems Research Institute (ESRI), "A GIS is an organized collection of hardware, software, geographic data and personnel designated to, efficiently, capture, store, update, manipulate, analyze and show all the forms of geographically referenced information." (ESRI 2007).

In a more complete way, GIS can be defined as: "GIS are based on the collection, storage, recovery, analysis and treatment of space, no space and temporary data, aiding the decision making and giving support to the administration, maintenance, operation, analysis and planning activities." (Rodrigues 1990).

GIS technology can be used for scientific researches, resource management, asset management, environmental impact assessment, urban planning, cartography, criminology, history, sales, marketing and logistics. For example, GIS might allow emergency planners to easily calculate emergency response times in the event of a natural disaster, GIS might be used to find wetlands that need protection from pollution, or GIS can be used by a company to site a new business to take advantage of a previously underserved market.

GIS are essentially projected for the manipulation of spatial data, the central component of these systems. All and any data considered as spatial must be mappable, that is, all information has to be linked to a specific object in a map, and the location of the objects in this map should be geographically referenced.

Spatial data are the main subsidy of the spatial analysis, but a GIS is also composed by non-spatial data. The quantitative data analysis can be done as spatially or as non-spatially.

The primordial difference between these two types of analysis is the inclusion or exclusion of spatial factors. Every spatial analysis involves processing about the geography, needing variables with geographic connections. Variables like: weight, height, income and education degree are non-spatial, while location, accessibility to the transport and proximity of service positions are spatial.

The spatial and non-spatial analyses also differ in the type of requested analytical tools. The non-spatial analysis only request a system of database administration and a statistical package, while the spatial ones usually request a complete GIS.

A GIS is most often associated with maps. A map, however, is only one of the ways you can work with geographic data in a GIS, and only one type of product generated by a GIS. This is important, because it means that a GIS can provide a great deal more of problem-solving capabilities than using a simple mapping program or adding data to an on-line mapping tool. A GIS can be viewed in three ways (ESRI 2007):

- The Database View: A GIS is a unique kind of database of the world - a geographic database (geo-database). It is an "Information System for Geography". Fundamentally, a GIS is based on a structured database that describes the world in geographic terms;
- The Map View: A GIS is a set of intelligent maps and other views that show features and feature relationships on the earth's surface. Maps of the underlying geographic information can be constructed and used as "windows into the database" to support queries, analysis, and editing of the information. This is called geo-visualization;
- The Model View: A GIS is a set of information transformation tools that derive new geographic datasets from existing datasets. These geo-processing functions take information from existing datasets, apply analytic functions, and write results into new derived datasets.

Together, these three views are critical parts of an intelligent GIS and are used at varying levels in all GIS applications.

According to (Guo et al. 2001), the role of a GIS in a framework, for instance a zone system, consists of three aspects:

- Firstly, it acts as a good source of input data since a wide variety of topological and demographic information is available within most common GIS products;
- Secondly, when provided with the zoning plan, a GIS serves to generate and to store the resulting zoning system. A GIS also facilitates the archiving and distribution of aggregated data;
- Thirdly, the graphical functionality in most GIS provides the user a readily available medium to visualize the resulting zone design and hence speed up the feedback loop in the design process. Furthermore, a GIS provides a mechanism for the user to make manual edition to the zoning system, a task that would be difficult without a GIS.

Finally, these capabilities are the reason why we have chosen to integrate our model into a GIS. Our objective is to provide a very powerful and flexible tool for the decision makers (DM) to solve territories alignment problems. Appendix A presents a more complete definition of the GIS components and some interesting information about ArcGIS and ArcObjects, the commercial GIS software that we decided to use.

2.3 Multi-Criteria Decision Analysis (MCDA)

2.3.1 Introduction to MCDA

According to (Matos 2005), when we have to make a decision, there are some procedures that we follow in a more or less systematic way:

- We identify alternatives, options or hypotheses of decision. For instance, a group of cars from among which we intend to choose one, a list of candidates that we intend to order, possible itineraries for a trip, a group of proposals in a public contest, etc.
- We verify the viability of the alternatives (eliminating the ones that don't satisfy certain eliminatory criteria, the ones that are technically called constraints). For instance, we discard the cars whose cost is bigger than our financial availability, candidates without certain minimum qualifications for the position, etc.
- We define the evaluation criteria, relevant aspects or points of view to have into account when we compare two alternatives in terms of preference. For instance, the cost (we prefer an alternative with smaller cost than the other), the professional experience (we usually prefer a candidate with more experience), the technical quality (we prefer a project with better technical quality).

In many circumstances of our lives, these steps are interconnected and mixed, and the decision seems to appear in a rather informal way. However, in more complex situations or when the decisions are made in an organizational environment, where they have to be justified or, at least, explained, we need to follow a formal procedure that guarantees the consistence of the global process.

Given that the territories alignment problem is complex and aims to assist, at the same time, different objectives of the several stakeholders, we concluded that multi-criteria techniques might be a good contribution to find solutions to the territories alignment problem.

2.3.2 Classification of MCDA Approaches

Decision Analysis is a set of systematic procedures for analyzing complex decision problems. These procedures include dividing the decision problems into smaller, more understandable ones; analyzing each sub-problem; and integrating the sub-problems in a logical manner in order to produce a meaningful solution (Malczewski 1997).

In general, MCDA problems involve six components (Keeney and Raiffa 1976):

- A goal or a set of goals the decision maker wants to achieve;
- The decision maker or a group of decision makers involved in the decision making process with their preferences with respect to the evaluation criteria;
- A set of evaluation criteria (objectives and/or physical attributes);
- The set of decision alternatives;
- The set of uncontrollable (independent) variables or states of nature (decision environment);
- The set of outcomes or consequences associated with each alternative attribute pair.

MCDA techniques can be used to identify a single most preferred option, to rank options, to list a limited number of options, for subsequent detailed evaluation, or to distinguish acceptable from unacceptable possibilities.

There are many MCDA approaches which differ in how they combine and utilize the data. MCDA approaches can be classified on the basis of the major components of multi-criteria decision analysis. Three different classifications can be made:

- Multi objective decision making (MODM) versus multi-attribute decision making (MADM);
- Individual versus group decision maker problems;
- Decisions under certainty versus decisions under uncertainty.

The distinction between MADM and MODM is based on the evaluation criteria which are the standards of judgments or rules on which the alternatives are ranked according to their desirability. Criterion is a general term and includes both the concepts of attributes and objectives.

An attribute is a measurable quantity whose value reflects the degree to which a particular objective is achieved. An objective is a statement about the desired state of the system under consideration. It indicates the directions of improvement of one or more attributes. Objectives are functionally related to, or derived from, a set of attributes (Malczewski 1999).

There might be formal relationships between objectives and attributes, but usually the relationship is informal. To assign an attribute to a given objective, two properties which are comprehensiveness and measurability should be satisfied. An attribute is comprehensive if its value sufficiently indicates the degree to which the objective is met. And it is measurable if it is reasonably practical to assign a value in a relevant measurement scale. The ratio, interval, ordinal and binary scales are suitable for measurement of attributes, whereas nominal scale is not, since it does not allow an ordering of the alternatives (Janssen 1992).

MADM problems require that choices are made among alternatives described by their attributes. The set of attributes is given explicitly and multi-attribute problems have a finite set of feasible alternatives. Unlike MADM, MODM problems require that means-ends relationships be specified, since they deal explicitly with the relationship of attributes of alternatives to higher level objectives. MODM involves designing the alternatives and searching for the best decisions among an infinite or very large set of feasible alternatives. Each alternative is defined implicitly in terms of the decision variables and evaluated by means of objective functions.

Both MADM and MODM problems can be further classified as individual and group decision making depending on the goal-preference structure. If there is a single goal preference, the problems is considered as individual decision making regardless of the number of decision makers involved in the process. However, if the individual or interest groups are characterized by different goal preferences, the problem becomes the group decision making one.

The other classification depends on the certainty of the decision. If the decision maker has perfect knowledge of the decision environment and the amount of knowledge available is enough, then the decision is considered as decision under certainty. However, most of the real world decisions involve some aspects that are unknown and difficult to predict. This type of decisions is referred as decisions under uncertainty. The decisions under uncertainty can be further subdivided into fuzzy and probabilistic decision making. The probabilistic decisions are handled by probability theory and statistics. And the outcome of a stochastic event is either true or false. However, if the situation is ambiguous, the problem is structured

as the degree of how much an event belongs to a class. This type of problems is handled by fuzzy sets theory.

Appendix B presents the steps of a spatial MCDA analysis and the major elements involved in a spatial decision making process, including the MCDA methods that can be used to solve multi-criteria problems in a GIS based tool.

In this thesis, we develop a new approach for the territories alignment problem. This approach can be classified as:

- **Multi-criteria (multi-attribute):** because we defined only one objective function and the alignment objectives are represented as numeric attributes for the territories;
- **Individual decision making**, if there is only one attribute for the territories, or **group decision making**, if there are more than one objective and these objectives represent different interests of the main stakeholders;
- **Decision under certainty:** because every attribute is represented numerically and the decision maker must have perfect knowledge of the decision environment.

In Chapter 4 we outline the MCDA decision rule method used in the MultiACE algorithm to help finding feasible solutions for the territories alignment multi-criteria problem.

2.4 Heuristics and Meta-Heuristics

2.4.1 Main Concepts

A good definition of heuristics and meta-heuristics can be found at (Ehrgott and Gandibleux 2004).

"Heuristics. A heuristic is defined as a technique which seeks good (i.e. near-optimal) solutions at a reasonable computational cost without being able to guarantee optimality, to state how close to optimality a particular feasible solution is or, in some cases, even to guarantee feasibility. Often heuristics are problem-specific, so that a method which works for one problem cannot be used to solve a different one.

Meta-heuristics¹. In contrast, meta-heuristics are powerful techniques applicable generally to a large number of problems. A meta-heuristic refers to an iterative master strategy that guides and modifies the operations of subordinate heuristics by combining intelligently different concepts for exploring and exploiting the search space. A meta-heuristic is a solution concept. The adaptation to a specific problem uses heuristics as solution methods."

Combinatorial optimization is a branch of optimization. Its domain is optimization problems where the set of feasible solutions is discrete or can be reduced to a discrete one, and the goal is to find the best possible solution. It can be also faced as a branch of applied mathematics and computer science, related to operations research, algorithm theory and computational complexity theory that sits at the intersection of several fields, including artificial intelligence, mathematics and software engineering.

In the context of meta-heuristics, the main goal of combinatorial optimization is to find a discrete mathematical object (such as a bit string or permutation) that maximizes (or minimizes) an arbitrary function specified by the user. These objects are generically called states, and the set of all candidate states is the search space. The nature of the states and the search space are usually problem-specific.

The function to be optimized is called the goal function, or objective function, and is usually provided by the user as a black-box procedure that evaluates the function on a given state. Depending on the meta-heuristic, the user may have to provide other black-box procedures that, say, produce a new random state, produce variants of a given state, pick one state among several or provide upper or lower bounds for the goal function over a set of states.

Some meta-heuristics maintain at any instant a single current state, and replace that state by a new one. This basic step is sometimes called a state transition or move. The move is uphill or downhill depending on whether the goal function value increases or decreases. The new state may be constructed from scratch by a user-given generator procedure. Alternatively, the new state may be derived from the current state by a user given mutator procedure; in this case the new state is called a neighbor of the current one. Producers and mutators are often probabilistic procedures. The set of new states that can be produced by the mutator is the neighborhood of the current state.

More sophisticated meta-heuristics maintain, instead of a single current state, a current pool with several candidate states. Then, the basic step may add or delete states from

¹ The name combines the Greek prefix "meta" ("beyond", here in the sense of "higher level") and "heuristic" (heuriskein, "to find").

this pool. User-given procedures may be called to select the states to be discarded, and to generate the new ones to be added. The latter may be generated by combination or crossover of two or more states from the pool.

A meta-heuristic may also keep track of the current optimum, the optimum state among those already evaluated so far.

Since the set of candidates is usually very large, meta-heuristics are typically implemented so that they can be interrupted after a client-specified time budget. If not interrupted, some exact meta-heuristics will eventually check all candidates, and use heuristic methods only to choose the order of enumeration; therefore, they will always find the true optimum, if their time budget is large enough. Other meta-heuristics give only a weaker probabilistic guarantee, namely that, as the time budget goes to infinity, the probability of checking every candidate tends to 1.

Some well-known meta-heuristics are:

- Random optimization;
- Local search;
- Greedy algorithm and hill-climbing;
- Random-restart hill climbing;
- Best-first search;
- Genetic algorithms;
- Simulated annealing (SA);
- Tabu search;
- Ant colony optimization;
- GRASP;
- Stochastic Diffusion Search;
- Generalized extreme optimization;
- Harmony search.

Innumerable variants and hybrids of these techniques have been proposed, and many more applications of meta-heuristics to specific problems have been reported (See Chapter 3). This is an active field of research, with a considerable literature, a large community of researchers and users, and a wide range of applications.

Since our work is based in a GRASP - TS hybrid meta-heuristic, in the next Sections, we outline the main concepts regarding these two techniques.

2.4.2 GRASP

According to (Blum and Roli 2003), GRASP is a simple meta-heuristic that combines constructive heuristics and local search. GRASP is an iterative procedure, composed of two phases: solution construction and solution improvement. The best found solution is returned upon termination of the search process.

The solution construction mechanism is characterized by two main ingredients: a dynamic constructive heuristic and randomization. Assuming that a solution “ s ” consists of a subset of a set of elements (solution components), the solution is constructed step-by-step by adding one new element at a time. The choice of the next element is done by picking it uniformly at random from a candidate list. The elements are ranked by means of a heuristic criterion that gives them a score as a function of the (myopic) benefit if inserted in the current partial solution. The candidate list, called *restricted candidate list* (RCL), is composed of the best elements. The heuristic values are updated at each step, thus the score of each element changes during the construction phase, depending on the possible choices. This constructive heuristic is called dynamic, in contrast to the static one which assigns a score to each element only before starting the construction.

The length of the RCL determines the strength of the heuristic bias. In the extreme case (the length of the RCL is 1) the best element would be added, thus the construction would be equivalent to a deterministic Greedy Heuristic. On the opposite, if the length of the RCL is the number of elements in the set, the construction would be completely random (indeed, the choice of an element from the candidate list is done at random). Therefore, the length of the RCL is a critical parameter which influences the sampling of the search space.

The solution improvement phase of the algorithm is a local search process, which may be a basic local search algorithm such as iterative improvement, or a more advanced technique such as SA or TS. GRASP can be effective if two conditions are satisfied:

- The solution construction mechanism samples the most promising regions of the search space;
- The solutions built up by the constructive heuristic belong to basins of attraction of different local minima solutions.

The first condition can be met by the choice of an effective constructive heuristic and an appropriate length of the candidate list, whereas the second condition can be met by choosing the constructive heuristic and the local search in a way such that they fit well. The description of GRASP as given above indicates that a basic GRASP does not use the history of the search process. The only memory requirement is storing the problem instance and

keeping the best so-far solution. This is one of the reasons why GRASP is often outperformed by other meta-heuristics. However, due to its simplicity, it is generally very fast and it is able to produce quite good solutions in a very short amount of computation time. Furthermore, it can be successfully integrated into other search techniques.

2.4.3 Tabu Search

According to (Blum and Roli 2003), TS is among the most cited and used meta-heuristics for combinatorial optimization problems. TS explicitly uses the history of the search, both to escape from local minima and to implement an explorative strategy.

The simplest version of the TS algorithm applies a best improvement local search as basic ingredient and uses a *short term memory* to escape from local minima and to avoid cycles. The short term memory is implemented as a *tabu list* that keeps track of the most recently visited solutions and forbids moves toward them. The neighborhood of the current solution is thus restricted to the solutions that do not belong to the tabu list. In the following we will refer to this set as *allowed set*. At each iteration the best solution from the allowed set is chosen as the new current solution. Additionally, this solution is added to the tabu list and one of the solutions that were already in the tabu list is removed (usually in a FIFO order). Due to this dynamic restriction of allowed solutions in a neighborhood, TS can be considered as a dynamic neighborhood search technique. The algorithm stops when a termination condition is met. It might also terminate if the allowed set is empty, that is, if all the solutions in the search space are forbidden by the tabu list.

The use of a tabu list prevents from returning to recently visited solutions. Therefore, it prevents from endless cycling and forces the search to accept even uphill moves. The length of the tabu list (i.e., the tabu tenure) controls the memory of the search process. With small tabu tenures the search will concentrate on small areas of the search space. On the opposite, a large tabu tenure forces the search process to explore larger regions, because it forbids revisiting a higher number of solutions. The tabu tenure can vary during the search, leading to more robust algorithms.

However, the implementation of short term memory as a list that contains complete solutions is not practical, because managing a list of solutions is highly inefficient. Therefore, instead of the solutions themselves, solution attributes are stored. Attributes are usually components of solutions, moves, or differences between two solutions. Since more than one attribute can be considered, a tabu list is introduced for each of them. The set of attributes and the corresponding tabu lists define the tabu conditions which are used to filter the

neighborhood of a solution and generate the allowed set. Storing attributes instead of complete solutions is much more efficient, but it introduces a loss of information, as forbidding an attribute means assigning the tabu status to probably more than one solution. Thus, it is possible that unvisited solutions of good quality are excluded from the allowed set. To overcome this problem, aspiration criteria are defined which enable the inclusion of a solution in the allowed set even if it is forbidden by tabu conditions. Aspiration criteria define the aspiration conditions that are used to construct the allowed set. The most commonly used aspiration criterion selects solutions which are better than the current best one.

Tabu lists are only one of the possible ways of taking advantage of the history of the search. They are commonly identified with the usage of short term memory. Information collected during the whole search process can also be very useful, especially for a strategic guidance of the algorithm. This kind of long-term memory is usually added to TS by referring to four principles: *recency*, *frequency*, *quality* and *influence*. Recency-based memory records for each solution (or attribute) the most recent iteration it was involved in. Frequency-based memory keeps track of how many times each solution (attribute) has been visited. This information identifies the regions (or the subsets) of the solution space where the search was confined, or where it stayed for a high number of iterations. This kind of information about the past is usually exploited to diversify the search. The third principle (i.e., quality) refers to the accumulation and extraction of information from the search history in order to identify good solution components. This information can be usefully integrated in the solution construction. Other meta-heuristics (e.g., ant colony optimization) explicitly use this principle to learn about good combinations of solution components. Finally, influence is a property regarding choices made during the search and can be used to indicate which choices have shown to be the most critical. In general, the TS field is a rich source of ideas. Many of these ideas and strategies have been and are currently adopted by other meta-heuristics.

Chapter 3 presents some works that can clarify the definitions of GRASP and TS and a list of interesting papers using these techniques in several areas.

In this thesis we present a hybrid meta-heuristic that uses some concepts found both in GRASP and TS. These concepts are:

- From GRASP, the RCL and the randomization for candidates list;
- From TS, the short-term and long term memory structures.

In Chapter 4 we explain how we implemented these procedures and how we fitted them together in our meta-heuristic.

2.5 Territories project

The territories project problem, that has been referred to as the automatic zoning design problem, the land allocation problem, the redistricting problem, the region partitioning problem and the districting problem, is an important geographic problem that is present in a great number of geographic tasks. Zone design algorithms have been used in school districting, in the design of zones with appropriate characteristics for posterior socio-economic and epidemiological analysis, in the design of SST, in the electoral districting and in the design of census output geography.

According to (Guo et al. 2001), the distinct characteristic of spatial analysis and modeling is in their use of geo-referenced data, which are often zonal in nature (e.g. census collection districts). In order to uncover a pattern within the data and to make the problem size manageable, it is common in most spatial studies to further aggregate the data into larger areal units or zones. In aggregating adjacent basic areal units to form a larger zone, the original individual attribute values of the basic units are replaced by a single value. The uniqueness of each areal unit and the variation for the whole area is often lost. Depending on the aggregation scale and the spatial configuration, the construction of new zones can yield different representations of the same data set. Any analytical or statistical results derived from the data may vary depending on the particular zoning system chosen. The significance of this so-called modifiable areal unit problem (MAUP) has been long recognized. It has been suggested that one way of approaching the MAUP is to tailor the zoning system for the particular study at hand. The zone design issue was often ignored in the past due to lack of awareness of MAUP and the absence of appropriate technology. Instead, researchers and practitioners typically used official sets of zones (e.g. statistical local areas, or postcodes). Only recently, with the emergence of GIS technology, the MAUP has regained interest. The availability of digitized spatial data and the proliferation of GIS products enable the user to aggregate areal units to produce alternative zoning systems. However, the problem of insuring that the resulting zoning system is meaningful remains a major impediment.

2.5.1 Territories project objectives and constraints

The criteria on what constitutes a meaningful zoning system lie in the purpose of the studies and depend on the experience of the zone designer to define. The zoning criteria are often guided by the problem specifications or restricted by the available data. In the problem of electoral districting, aside from its obviously political aspect, the process must be evaluated against specified redistricting criteria. For instance, some of the criteria are constitutionally required while other geographical and political concerns may be advocated. The most commonly used set of criteria includes contiguity, compactness and population equality.

Another application area is the transport planning process. In the conventional aggregate transportation modeling, the study region is divided into areal units known as transport analysis zones (TAZ). Aggregation is required to reduce the level of detail of the region under study.

Researchers and practitioners have put forward diverse opinions on what design criteria to consider when devising TAZ. These include:

- Minimum variation in spatial characteristics of zones (i.e. area, size, shape);
- Zonal homogeneity in socio-economic characteristics (i.e. number of households, income level);
- Maximum spatial interaction (i.e. number of trips);
- Contiguity and compactness of zones;
- Compatibility with natural and physical barriers to travel;
- Equal trip generation/attraction;
- Traffic load points as centroids;
- Compatibility with other existing spatial units (i.e. administrative divisions, previous zoning systems, census tract boundaries, political boundaries);
- Decision maker's preference in number of zones.

It is observed that these zoning objectives revolve around three fundamental elements: definition of connectedness, measure of the individual areal units, and measure of inter-unit relationships.

According to (Bacao et al. 2005a), the constraints of the zone design problem are similar to the ones that characterize the clustering problem. Let the set of initial areal units be $X = \{x_1, x_2, \dots, x_n\}$, where x_i is the i -th areal unit. Let the number of zones be K . Let Z_i be the set of all the areal units that belong to zone Z_i . Then:

$$\begin{aligned}
Z_i &\neq \emptyset, \text{ for } i=1, \dots, K, \\
Z_i \cap Z_j &= \emptyset, \text{ for } i \neq j, \\
\bigcup_{i=1}^K Z_i &= X
\end{aligned} \tag{2.1}$$

These constitute the set of constraints that can be applied equally in clustering and in zone design. Nevertheless, in zone design an additional constraint has to be included, which accounts for contiguity and creates a more complex problem. This constraint limits the set of acceptable solutions to the problem and consists in assuring contiguity between all the areal units that build up a zone. Contiguity means that each areal unit in a zone is connected to every other areal unit via areal units that are also in the zone.

One of the reasons for the zone design problem being especially difficult is due to the size of the solution space. The dimension of a usual real world problem makes unfeasible any attempt to explicitly enumerate all the possible solutions. The calculation of the total number of possible solutions for a zone design problem is similar to the clustering problem and is given by the Stirling number of the second kind (Altman 1998). If we have “ n ” unit blocks and want to create “ k ” districts, the number S of possible districts is:

$$S(n, k) = \frac{1}{k!} \sum_{i=0}^k (-1)^i \left(\frac{k!}{(k-i)!i!} \right) (k-i)^n \tag{2.2}$$

Additionally, in terms of computational complexity, the zone design problem has been shown to be NP-Complete (Crescenzi and Kann 1998). Thus, heuristic techniques seem to be the best road available to produce solutions to the problem in reasonable computational time. This is certainly a compromise but guaranteed optimality seems at this stage simply too difficult.

2.5.2 Sales/Services Territories (SST) Project

For large field organizations, the most critical resource is the aggregate face-to-face time that the SSF has available to call on customers and prospects. With the rising cost of an industrial field call going from \$30 in 1965 to \$60 in 1975 to over \$200 in 2000, we are seeing a doubling and then a tripling in the cost of a call. The cumulative effect of poor SSF deployment (i.e. calls not made on accounts that increase sales or profits the most) drives up the cost of sales and increases lost revenue. On the positive side, by using the management principle of leverage, a very small improvement in SSF deployment can produce enormous improvements in sales and profits (Pinals 2001).

The SST definition can be seen as the division of the company's customers group in smaller subsets, following criteria or objectives defined during the strategic planning. To

understand the problem linked to the SST project, we previously need to know why the subject is so important for companies in general.

According to (Zoltners et al. 2004), the changes that happen in the SSF are important because of the SSF costs and because of their direct connection to sales and their impacts in the profitability of the companies. In general, the SSF have a high cost, but they are also the responsible for the sales, they are the ones that generate incomes. Besides that, they are the “means of living” of the companies; for many companies' customers, the sales people “are the company”. The best SSF are dynamic, because they suffer changes and they adapt as the market develops. That constant changing is, at the same time, a problem and a source of opportunities for the companies.

In a general way, the reasons that lead SSF to suffer changes are:

- External factors (customers, competition, and environment);
- Internal factors (corporate strategies);
- Performance factors (performance challenges).

We can generalize the SST project problem for any organization that distributes goods or services and that needs to optimize the use of its resources (people, vehicles, time, money) invested in a certain geographical area. During the last four decades, several researchers tried, in an objective way, to move forward with the theory of the SST project for the most varied types of companies, customers and products.

SST alignment is not an exact problem and there are no rights or wrong alignments. An alignment that may be perfect for one business may be a disaster for another, even if they sell the same products. Typically, a SST alignment is a balancing act both in concept and in procedure. A set of territories that have the absolute minimum travel time may not be practical because the market potential is so wildly skewed. Similarly, a territory alignment containing territories that are all balanced with respect to sales ($\pm 1\%$) may be unworkable due to the huge variation in drive times from territory to territory. As a result, the best alignments are carefully tailored to suit the needs and objectives of a particular business and will almost always be a compromise of several factors.

In an alignment, there is typically a set of alignment criteria or objectives that are considered in the process. One of the most common objectives is the balancing of workload and equalizing of sales potential. Balancing of workload or potential is important for two reasons. A SST with too many customers or prospects means that sales opportunities are left on the table because there is too much work to do. A territory with too few customers or prospects means that a SSF member's time is being wasted and potential is not being

exploited. The combination of wasted selling time and missed sales opportunities restrains revenue growth and increases expenses.

Minimizing travel and reducing or eliminating territory overlap are frequently considered as additional objectives in SST alignment. Poorly defined territories result in increased travel or “windshield” time and reduced selling time. Proper representative placement, good travel access and minimal territory overlap are achieved in an efficient territory alignment.

Because the objectives of SSF redeployment differ from case to case, it is always a good idea to determine the companies’ specific objectives before the alignment process begins. By doing this, we can then be certain when our newly aligned territories meet our goals.

Benefits of territory alignment include:

- Reduced travel time due to having geographically manageable territories;
- Improved customer and prospect coverage due to creating territories that allow dedicated representatives to better manage the number of opportunities as opposed to having a too large coverage area;
- Increased sales due to optimized territories that allow underutilized representatives to cover more opportunities;
- A better alignment of pay-for-performance objectives that ensure that the representative is rewarded as opposed to the opportunity existing in his territory (this also improves morale).

SST alignment also improves the ability of executives to analyze territory effectiveness and conduct predictive modeling to experiment with realignment theories.

Since a territory alignment (or realignment) is typically event driven, there is usually a single objective that is the foremost in the minds of those orchestrating the change. For example, there may be a need to increase or decrease the number of field personnel, a merger may imply the combining of two or more sales organizations, or a new product may require that a dedicate SSF be spun off from the main SSF. In most cases, there are goals and objectives associated to these events. You may or may not have input into or control over these events but you should make an effort to understand what needs to be accomplished and be in a position to suggest alternatives or techniques to make the task easier.

Questions that one may want to ask prior to embarking on a territory optimization or alignment project include:

- How many field SST, districts and regions are needed?

- Are there components to the existing sales hierarchy that cannot change? If so, what are they?
- What data variables need to be considered in the alignment?
- What is the primary variable to be used for balancing territories?
- What percent deviation from this goal is acceptable?
- Do existing SST need to be considered?

2.5.3 When is it Time to Make SST Alignments?

Most companies face a territory alignment question at least once a year when forecasting revenue growth, market expansion and new hires for the following year. It also happens that unforeseen territory issues show up throughout the year, at any time.

Often an “event” takes place in a company that causes territory disruption and creates a sense of urgency around territory planning. The event is often a typical business situation. It might be:

- Merger or acquisition;
- New sales channel;
- Major product launches (or retirements);
- Hiring of sales people;
- Change in “go-to-market” strategy;
- Emergence of a new competitor;
- Downsizing.

Any one of these events will cause territory disruption and likely requires:

- A re-assignment of sales/services resources;
- A re-alignment of SST.

And herein lies the pitfall for many companies: in the rush to respond to the event and meet business mandates, territories are quickly cobbled together. New territories get appended to old, adjacent territories are merged, new territory maps are drawn quickly using basic mapping software or other primitive tools, or territory re-alignment is never addressed - and the result is a bunch of territories that over-promise and under-deliver, eventually collapsing like a house of cards.

To avoid the “house of cards” pitfall, sales/services executives and managers need an intelligent, straightforward and proven process to re-align and optimize territories.

Appendix C presents a seven-step process that will help to move towards the desired outcomes. At the completion of these seven steps, an organization will have an in-depth

understanding and visualization of how, why and where to deploy their resources to convert their current state into a much more efficient state.

2.6 Summary

In this chapter we have presented the main concepts in Geographic Information Systems, Multi-Criteria Decision Analysis, Meta-Heuristics and Territories Project. We also outlined the basic concepts of GRASP and TS algorithms and we have briefly described some of the most important criteria used in territories alignment problems. This brief introduction sets up the ground for the choice of the detailed approaches to be used along this work, in particular in Chapter 4.

Chapter 3

Bibliographical Review

"You will find it a very good practice always to verify your references sir."

Martin Routh (1755 - 1854) English classical scholar

3.1 Introduction

This chapter is dedicated to works on the subjects outlined in Chapter 2 and the main objective is not so much to provide an extensive bibliography, but rather to highlight some important papers that served as a basis to create the new approach presented in this thesis. In the sequel, Section 3.2 deals with the GIS concepts and applications. Section 3.3 presents some papers on the Multi Objective Combinatorial Optimization and Meta-heuristics. Section 3.4 presents a few works on the territories districting and partitioning problem. Finally, Section 3.5 outlines some works on GRASP and TS techniques.

3.2 GIS

(Murphy 1995) develops a work that looks for better defining the concept of GIS and verifying if they can be used as decision support systems. Special problems and analysis associated to the spatial data are also discussed. The conclusions that GIS can aid the decision making are complemented with the implications of their use in the companies and in the administration activities.

(Mennecke and Crossland 1996) introduce GIS to the researchers of the information systems area at the same time they suggest research opportunities in the areas related to

administration, organizational impacts, and evaluation of the effectiveness of the decision making and social impacts in the developed and in development countries.

(Keenan 1998) suggests the use of GIS in the decision making area related to the vehicle routing. Besides the routes optimization algorithms introduction, the author tries to point out the importance of the junction of the capacities of geographical data manipulation with the science of administration, turning the decision making process more complete, fast and effective.

(Afonso et al. 1998) present a model and make the specification of a prototype for a decision making for marketing problems that incorporates fuzzy theory to GIS. The architecture of the proposed system is based on three main components: a decision making system that is leaning in a GIS and in an evaluation system that uses fuzzy logic.

(Forster 2000) executes, through the analysis of interviews accomplished in several logistic services area companies, a survey of the use of GIS. The author concludes that, although it already exists a group of systems that present characteristics existent in GIS, the same is confused with mapping programs and its use is in the operational sphere, not being used during the planning process or during the decision making.

(Mohraz 2000) presents the application of a GIS together with a database. The main characteristic of this system is enabling the visualization of the information present in the company's database directly in thematic maps, facilitating the analysis and definition of patterns that can help during the decision making and planning processes.

(Zhao 2000) makes the integration of statistical classification techniques and GIS in order to obtain a marketing analysis. Through that integration, customers' data and information coming from censuses supply the base for the establishment of customers' profiles and for mapping the areas of larger interest for companies, facilitating the allocation of resources and the decision making process.

(Verter and Kara 2001) present a methodology for assessment of the hazardous materials transport risk in a multi-commodity, multiple origin-destination setting. The proposed risk assessment methodology was integrated in a GIS. A GIS-based model of the truck shipments of dangerous goods was developed and a government agency could use it to estimate the impact of alternative policies that could alter the carriers' route choices.

(Monteiro et al. 2001) have the location of wind energy distributed generation structures as their area of application of GIS. The work shows how GIS can be used as spatial decision making systems.

(Vlachopoulou et al. 2001) develop a process and make its application in the area of site selection for depots implantation. According to the authors, the process allows, with the aid of GIS and some specifically developed routines, the definition of places to assist location needs, in agreement with the corporate objectives.

(Macmillian 2001) makes the junction of a simulated annealing heuristic and a GIS to get the partition of a set of areas in smaller subsets. The algorithm tries to obtain contiguous subsets and to minimize the population difference among them. Besides, the author presents a new procedure for proximity evaluation of the subsets and compares it with a previously developed one.

(Zhao and Garner 2001) integrate, through Dynamic Data Exchange programming, two systems, a GIS and a Multi-Criteria Decision Making (MCDM) system for the analysis and selection of bank agencies that should be closed. The work searches to use the multi-criteria analysis and the spatial analysis to improve the evaluation process and to facilitate the process of decision making.

(Church 2002) approaches the use of GIS in the evaluation and site selection area. It is then made an inclusive bibliographical review of the main works accomplished in this area, showing that the application of GIS in the location science grows as being a multidisciplinary area. Also, some research areas involving GIS and the location science are discussed.

(Thomas 2002) demonstrates the modeling and the use of GIS in the urban areas revitalization area. The analysis and selection of areas for revitalization are made with the aid of a decision system based on a GIS and the author points out the importance of the system for the decision making of government agencies, which decide on the resource allocation for the whole population.

In (Zhao 2002) and (Zhao 2004), the author makes a complementary study of the integration of multi-criteria decision systems and GIS and, after defining the whole process for that integration, accomplishes the implementation of the system proposed in a real case study. It is then shown the importance of the research and of the integration of the two study fields.

(Nasirin 2003), based on the capacity of GIS for great databases manipulation and using the “grounded theory” approach, explores cases of study with the objective of noticing and presenting the success factors that justify the use of GIS in projects of retail companies.

(Hess et al. 2004) specify the use of GIS as a decision making tool for the marketing area. The area of marketing information systems is then approximated to the one of spatial

analysis. An integration of the existent information is proposed and considerations for future researches seeking to approach this theme are made.

(Marinoni 2004) focuses the development of an ArcGIS VBA macro which enables the user to derive criteria weights with the analytical hierarchy process and which allows a mapping of the land use assessment results by a weighted summation of GIS raster data sets. He also provides a dynamic link library for the calculation of the eigen-values and eigen-vectors of a square matrix.

(Bacao et al. 2005b) briefly review different initialization procedures for the k-means algorithm, one of the most widely used clustering techniques used in geographical information science problems, and propose Kohonen's Self-Organizing Maps as the most convenient method, given the proper training parameters. They show that, in the final stages of its training procedure, the Self-Organizing Map algorithm is exactly the same as the k-means algorithm. Thus, they propose the use of Self-Organizing Maps as possible substitutes for the more classical k-means clustering algorithms.

(Zhao 2005) presents a study using GIS together with databases censuses for the evaluation of markets for retail companies. Besides the functionalities found in used GIS, the author also points out the importance of the visualization of the information that can be obtained with GIS and the improvement of the decisions that can be made with the use of the tool.

3.3 Multi Objective, Multi-Criteria and Meta-Heuristics

According to (Zitzler 1999), many real-world problems involve two types of problem difficulties: i) multiple, conflicting objectives and ii) a highly complex search space. On the one hand, instead of a single optimal solution, competing goals give rise to a set of compromise solutions generally denoted as Pareto-optimal. In the absence of preference information, none of the corresponding trade-offs can be said to be better than the others. On the other hand, the search space can be too large and too complex to be solved by exact methods. Thus, efficient optimization strategies are required to be able of dealing with both difficulties.

In this work, (Zitzler 1999) makes the comparison and the improvement of existing multi-objective evolutionary algorithms and their application to system design problems in

computer engineering. In detail, the major contributions are: development of an experimental methodology to compare multi-objective optimizers; execution of an extensive comparison of numerous evolutionary techniques in which further aspects, such as the influence of elitism and the population size, are also investigated; proposal of a novel approach to multi-objective optimization, the strength Pareto evolutionary algorithm; and addressing of two complex multi-criteria applications using evolutionary algorithms.

(Ehrgott and Gandibleux 2000a) provide a survey of the research in an annotated bibliography of multiple objective combinatorial optimization (MOCO). They present a general formulation of MOCO problems, describe the main characteristics of MOCO problems, and review the main properties and theoretical results for these problems. The main parts of the paper are a section on the review of the available solution methodology, both exact and heuristic, and a section, organized problem by problem, on the annotation of the existing literature in the field. They conclude the paper by stating open questions and areas of future research.

Still in the field of the references, (Ehrgott and Gandibleux 2000b) brings a list with two hundred and thirty four works that form a bibliography related to the multi-objective combinatorial optimization. A formulation is presented for this problem type, describing its main properties and its theoretical results. Open subjects and areas for future investigations are also presented.

According to (Jaszkiewicz 2001):

“MOCO problems arise in many areas: production scheduling, project scheduling, staff scheduling, time tabling, production facilities design, vehicle routing, telecommunication routing, investment planning and location. The number of successful real-life applications of MOCO reported in the literature is, however, relatively low in comparison with other areas of multiple objective optimization or single objective combinatorial optimization. Clearly, MOCO problems constitute one of the most difficult fields of optimization. Following the success of meta-heuristics in single objective optimization many researchers proposed the use of meta-heuristics in multiple objective cases. Since the proposition of the Vector Evaluated Genetic Algorithm in 1985 a significant number of different multiple objective meta-heuristics have been proposed. The goal of the methods is to generate a set of approximately Pareto-optimal solutions being an approximation of the whole or a part of the non-dominated set. At present, multiple objective meta-heuristics constitute one of the most active fields of research in multiple objective optimization. The use of multiple objective meta-heuristics is one of the most promising approaches to MOCO. The goal of this work is to advance the methodology of multiple objective meta-heuristics for MOCO problems.”

(Jaszkiewicz 2001) does a review of the methodology of multiple objective meta-heuristics. The review focuses on both characterization of the main algorithms and the issue of evaluation of multiple objective meta-heuristics. In this work a new approach that evaluates approximations of the non-dominated set on the basis of some assumptions about

the decision maker's preferences is presented. Then, two new multiple objective meta-heuristics developed by the author are presented. They are based on two well-known single objective methods: simulated annealing and genetic local search. This work describes also a two-phase approach to interactive analysis of MOCO problems.

(Ehrgott and Gandibleux 2004) present a review of heuristics and meta-heuristics elaborated for the solution of problems of multi-objective combinatorial optimization.

As (Bong et al. 2004) say, redistricting is a spatial multi-criteria decision problem but there is a limited research attempt in terms of decision-making. The limitations in existing solutions create an opportunity for designing a more structured and systematic method. On the other hand, redistricting based on multi-objective definition helps to consider relationship among objectives and it is a more realistic solution to the real-world problem. Therefore, there is a need to employ a multi-objective meta-heuristic technique as its solution and a GIS as its tool. The authors present a state-of-the-art multi-objective meta-heuristic for a redistricting process to draw territory lines for geographical or spatial zones for the purpose of space control. Their paper discusses the current applications and solutions by providing an in-depth description of the existing research study especially in the aspects of decision-making and optimization search method. It also gives a discussion on the role of GIS. Lastly, the paper presents the necessity of further research in multi-objective meta-heuristic to clarify the possible extension and enhancement in the technical and theoretical aspects.

(Pereira et al. 2004) present a work that deals with the problem of dividing a territory in "homogeneous" areas. Each area is composed by a group of elementary territory units. When multiple criteria are considered, the problem of enumerating all the possible solutions is known as being "NP-hard". It is then presented a method for solution based on an evolutionary algorithm with local search. This algorithm is then applied in a case study of the public transportation system of Paris, namely in the areas that compose the tariff system.

(Wei and Chai 2004) present a multi-objective hybrid meta-heuristic approach for an intelligent territory zoning model. The model uses a GIS and uses techniques of multi-objective combinatorial optimization as its components. The hybrid meta-heuristic proposed consists of the symbiosis between TS and scatter search and it is used to generate non-dominated solutions.

(Zitzler et al. 2004) give an overview of evolutionary multi-objective optimization with the focus on methods and theory. In the first part of the paper, basic principles of multi-objective optimization and evolutionary algorithms are presented, and various algorithmic

concepts such as fitness assignment, diversity preservation, and elitism are discussed. In the second part, they provide a tutorial that includes some recent theoretical results on the performance of multi-objective evolutionary algorithms and addresses the question of how to simplify the exchange of methods and applications by means of a standardized interface.

(Hashimoto 2004) aims at contributing for the estimation of electrical performance in the distribution of electrical energy. The electrical performance estimation is formulated according to an optimization problem where the objective functions correspond to an evaluation of occurrence probability, and correspond to a proximity evaluation of calculated parameters with values obtained by measurement as well. Specific heuristics are also proposed to get solutions with load diversity and unbalanced loads. In order to adequately apply these heuristics, a meta-heuristic evolutionary method to build feasible solutions is proposed and applied, and ranked according to Pareto's concept.

In (Coello 2006) can be found several works that show the various forms of applying multi-objective techniques and evolutionary algorithms in the resolution of the most varied types of combinatorial optimization problems.

3.4 Territories, Districting and Partitioning

(Hess 1971) approaches the theme of sales districts based on his experiences and researches. The author makes the proposal of a model and executes its implementation in the sales forces of a pharmaceutical company and of a computers company. The developed model, based on the analogy with the problem of defining electoral districts, searches to solve the problem of the territorial division of salespersons' areas and tries to guarantee that the obtained territories are appropriately balanced by the total effort of the respective customers. This effort is related with criteria defined by the managers of the company. Some examples: the amount of visits drafted for each customer, the maximum number of clients visited for each salesperson, the potential of market of the areas, the total of sales in each territory or even a value considered among two or more of the criteria previously mentioned. One of the pre-requisites of the model is the compactness of the territories obtained, reducing the time of trip and getting larger available time for the sales activities.

(Zoltners 1976) presents two integer linear programming models for the sales territories alignment where the main objective, based on the premises of each model, is maximizing the profit obtained by the sales forces.

(Parasuraman 1977) developed, together with sales executives of a great company, a sales territories analysis model. The model treats the sales operations in an intuitive way for the sales managers, using quantitative methods and computational simulation. It is a model guided to assist the sales forces managers, facilitating the analysis of the profitability of the territories and aiding in the sales resources allocation decision making.

(Zoltners 1980) shows the modeling of a practical conceptual structure for the sales resources allocation. Integer programming models are evaluated and a general model for sales resources allocation is developed, trying to treat the great variety of information used by the sales companies.

To deal with the problems of sales forces sizes and products and markets allocation, (Lodish 1980) develops a model that classifies as “easy to understand” and “easy to implement”. The model has basically two components. The first is a procedure that tries to foresee the sales and the effects in the profit of the allocation of a product to the sales forces dimensioning policies. The second is a procedure that tries to isolate the most profitable policies considering the constraints imposed by the firm.

The sales territories alignment problem is, once again, treated by (Zoltners 1983), and he develops an algorithm of alignment optimization that can, according to the author's data, establish contiguous territories, balanced by one more alignment criteria, and that it is also compatible with geographical considerations.

(Lodish et al. 1988) report the work executed in partnership with a pharmaceutical company for the resolution of the problem of defining the total amount of resources that should be invested in its sales force. The work shows how the modeling process was executed and developed and the positive results that were achieved with the proposed methodology.

(Howick and Pidd 1990) review models that try to solve some questions related with the sales forces: size of the sales forces, sales people's time allocation and sales territories alignment. They also supply, for each category of models, the reasons of its application and a discussion of its technical content. Ideas for futures developments and investigations are also supplied.

(Drexl and Haase 1999) approach the problem of the distribution of sales forces as an inter-relationship of four sub-problems: sales force dimensioning, sales people allocation,

sales territories alignment and sales resources allocation. It is then proposed a model that tries to solve the four sub-problems simultaneously. For the solution of the model, a new method is presented, which intends to solve instances of big scale present in practical cases. According to the authors, the developed method is effective and efficient, being useful for the marketing administration.

The problem of the alignment of territories is treated by (Zoltners and Lorimer 2000) as being an important productivity tool that is neglected by the sales companies. Some important points and benefits that can bring a good alignment are highlighted and it is also developed a process that looks for executing a sales force alignment in a correct, objective and reliable way.

(Kalcsics et al. 2001) treat the problem of sales territories project not as being a problem of grouping small units into larger units, but as a discrete problem of m -medians. In this manner, the case is seen as a “NP-hard” problem and a heuristic is developed for its solution. This algorithm is then implemented in a software package that the authors call BusinessManager.

(Sinha and Zoltners 2001), based on their experiences along 25 years in the modeling and project of sales territories, define the main points, the obstacles and the ways of overcoming them, the activities and the decisions of a sales forces changes implementation process. In a didactic and direct way, interesting insights are supplied for all the process phases and the most important doubts related to them are clarified.

(Barker 2001) investigates the impact of the characteristics of salespeople, activities of sales managers and territory design as antecedents of sales organization performance among Canadian sales forces. According to the author, and based on his multivariate analysis of variance of the responses from 102 field sales managers, more effective sales organizations have salespeople who are more motivated, employ managers who are better leaders and are more satisfied with the design of their territories. The author says that sales managers must adopt better people and communication skills to train and lead their salespeople who are expected to provide better service and establish long-term relations with their customers.

(Guo et al. 2001) describe the development and applications of a computational tool, referred to as the Multi-Objective Zoning and AggRegation Tool (MOZART), for devising optimal zoning systems. The tool aims to account for a diverse range of criteria that may arise in customizing zones to suit different contexts. This flexibility is achieved by treating the problem of zone design as a problem of partitioning a mathematical graph to meet a set of generalized objectives.

(Scaparra and Scutella 2001) try to fit the large variety of location models within a general unified framework, which arises from the description of the three buildings blocks of location problems, namely: facilities, customers, and locations. They provide evidences of how a particular problem specification can be stated mathematically as an optimization problem by adequately combining into a compact and workable model the main features that characterize and relate these three elements.

(Bergey et al. 2003) refer to the problem of partitioning a physical power grid into economically viable districts (distribution companies) as the electrical power districting problem (EPDP). Their research identifies the fundamental characteristics required to appropriately model and solve an EPDP and proposes a solution methodology implemented as a decision support system (DSS) featuring a visualization tool allowing decision makers (DM) to explore what they refer to as a “soft efficient frontier”. This DSS was found to effectively support DM at The World Bank in solving an EPDP in the context of a case study for the Republic of Ghana.

(Zoltners and Sinha 2005), once again, based on their experiences along 30 years in the modeling and implementation of sales territories, approach the problem of the alignment of territories. They supply information since the evolution of this area until the process that was developed for the realignment of sales forces, also going through the analysis of real projects executed by the authors.

(Kirkizoglu 2005) develops a genetic algorithm (GA) to solve the sales territories alignment problem. The criteria that the author uses for the balance of the territories are: necessary effort to assist the customers, proximity of the territories and low rupture customer-salesperson. The algorithm was tested in small populations and the results are supplied in the work, however no comparison with other methods was accomplished.

(Kalcsics et al. 2005) present a new heuristic for the territories definition. The authors make an introduction of the problem of territories definition, approach the main criteria used for each application area, and make the modeling of the problem and its implementation, integrating it into a GIS.

(Bacao et al. 2005a) made a work where the GA were used to deal with the complex problem of zone design, that comprises a large number of geographical tasks, electoral districting being probably the most well known. They describe and formalize mathematically the electoral districting problem, present different problem encodings, suited to GA optimization, together with different objective functions. They also give a practical real world example and perform tests in order to evaluate the effectiveness of the GA approach.

(Shirabe 2005) considers a problem of allocating spatial units to particular uses in order to form “regions” according to specified criteria, which is here called “spatial unit allocation”. This paper proposes a new exact formulation of contiguity that can be incorporated into any mixed integer programming model for spatial units allocation. The resulting model guarantees the reinforcement of contiguity regardless of other included criteria such as compactness. The author also proposes a classification scheme for spatial properties that helps to address a variety of spatial properties in establishing spatial allocation criteria.

(Negreiros and Palhano 2006) define the capacitated centered clustering problem (CCCP) as the problem of defining a set of clusters with limited capacity and maximum similarity per cluster. Each cluster is composed of individuals from whom we can compute a centre value and, hence, determine a similarity measure. The clusters must cover the demands of their individuals. In this work, they present two variations of this problem (p-CCCP and Generic CCCP) and their mathematical programming formulations. They also propose a two-phase polynomial heuristic algorithm to solve it.

Among the algorithms already developed for the solution of the set partitioning problem, some authors treat it as a “P-median Problem” and, due to the applicability of these algorithms, we accomplished a deeper investigation of the existent literature.

(Narula and Ogbu 1976) present a branch-and-bound algorithm for the solution of the p-median problem where the limits are obtained through the resolution of the lagrangean relaxation of the problem of p-median using the method of sub-gradient optimization. The implementation of the algorithm and the application in small instances of the problem were both accomplished.

(Rolland et al. 1996) develop an efficient procedure using TS heuristic for the p-median problems. A comparison of this algorithm with two other heuristics is accomplished in instances up to 500 points, being shown that the results obtained by the proposed procedure are better.

(Bozkaya et al. 1997) try to develop an effective heuristic using the GA principles. The authors make a comparison of their method with other algorithms and obtain, for some situations, better results, mainly because it better avoids the problem of local optima.

(Pirkul et al. 1999) developed an application allowing to find and to visualize solutions for the p-median problem. The system is called VisOpt and presents a series of configuration options for the problems. The authors performed tests and comparisons with other resolution methods.

(Lorena et al. 2000) make the integration of a p-median model to GIS ArcView and SPRING. The code that was integrated to these GIS implements a recent lagrangean/surrogate heuristic approach. In this approach, the viability of the dual solution is made through an alternate location-allocation heuristic. The work presents some computational tests using data of the municipal district of São José dos Campos, Brazil, topping at the maximum of 3282 points and 1141 medians.

(Corrêa et al. 2002) propose a GA for the solution of capacitated p-median problems. They also present a new heuristic for the mutation operator and compare their results with a TS algorithm.

(Taillard 2003) presents new heuristic methods for solving a class of hard centroid clustering problems including the p-median, the sum-of-squares clustering and the multi-source Weber problems. The first method proposed is a candidate list search that produces good solutions in a short amount of time if the number of centers in the problem is not too large. The second method is a general local optimization approach that finds very good solutions. The third method is designed for problems with a large number of centers; it decomposes the problem into sub problems that are solved independently. Numerical results show that these methods are efficient - dozens of best solutions known to problem instances of the literature have been improved - and fast, handling problem instances with more than 85,000 entities and 15,000 centers - much larger than those solved in the literature. The expected complexity of these new procedures is discussed and shown to be comparable to that of an existing method which is known to be very fast.

(Reese 2005) accomplishes an investigation work that summarizes the literature on the solution methods for the p-median problem, both capacitated and non-capacitated, formulated on a graph or a net.

(Klose and Drex1 2005) review some of the contributions to the problem of locating facilities and allocating customers. In particular, continuous location models, network location models, mixed-integer programming models, and respective applications are summarized.

(Jackson et al. 2007) treat the p-median problem under a new distance measure, the directional rectilinear distance, which requires the assigned supply point for a given demand point to lie above and to the right of it. In a previous work, they showed that the directional p-median problem is polynomially solvable in one dimension; they give here an improved solution by reformulating the problem as a special case of the constrained shortest path problem and they also present an efficient heuristic to solve it. According to the authors,

compared to the robust Teitz and Bart heuristic, their heuristic enjoys substantial speedup with short loss in terms of solution quality, making it an ideal choice for real-world applications with thousands of demand points.

3.5 Greedy Randomized Adaptive Search Procedures and Tabu Search

(Pitsoulis and Resende 2001) define GRASP as a meta-heuristic for combinatorial optimization usually implemented as a multi-start procedure. Each iteration is made up of a construction phase, where a randomized greedy solution is built up, and a local search phase, which starts at that solution and applies iterative improvement until a local optimum is found. They give an overview of GRASP, describe the basic building blocks of a GRASP and cover enhancements to the basic procedure, including reactive GRASP, hybrid GRASP, and intensification strategies.

(Glover and Laguna 1993) define the TS technique and say that, more particularly, TS is based on the premise that problem solving, to be appropriately qualified as intelligent, must incorporate adaptive memory and responsive exploration. The adaptive memory feature of TS allows the implementation of procedures that are capable of searching the solution space economically and effectively. Since local choices are guided by information collected during the search, TS contrasts with memoryless designs that heavily rely on semi-random processes implemented as a form of sampling. Examples of memoryless methods include semi-greedy heuristics and the prominent “genetic” and “annealing” approaches inspired by metaphors of physics and biology. Adaptive memory also contrasts with rigid memory designs typical of branch and bound strategies.

One can find many introductory works, tutorials and reference papers about TS and GRASP. For instance, about TS: (Glover 1989), (Glover 1990b), (Glover 1990a), (Glover and Laguna 1993), (Glover et al. 1993), (Hertz et al. 1995), (Laguna and Glover 1996), (Hertz et al. 1997), (Glover and Laguna 1998); and about GRASP: (Resende 2001), (Resende and Ribeiro 2003), (Festa and Resende 2004).

Other interesting papers that are related to this proposal are listed and annotated as follows.

(Battiti and Tecchiolli 1994) propose an algorithm for combinatorial optimization where an explicit check for the repetition of configurations is added to the basic scheme of TS. In their Tabu scheme, the appropriate size of the list is learned in an automated way by reacting to the occurrence of cycles. In addition, if the search appears to be repeating an excessive number of solutions too often, the search is then diversified by making a number of random moves proportional to a moving average of the cycle length. The reactive scheme is compared to a “strict” Tabu scheme that forbids the repetition of configurations and also to schemes with a fixed or randomly varying list size.

(Trejos et al. 1999) present an original method for partitioning by automatic classification, using the optimization technique of TS. The method uses a classical TS scheme based on transfers for the minimization of the within variance; it introduces in the tabu list the indicator of the transferred object. This method is compared with two stochastic optimization-based methods proposed by the authors (one based on simulated annealing and the other on a GA), and with the classical k-means and Ward methods. Results of the TS are significantly better than the classical and genetic methods, and slightly better than their simulated annealing method.

(Delmaire et al. 1999) treat the single source capacitated plant location problem as a discrete location problem that takes into account capacities in the plants to be opened and imposes that clients are served from a single open plant. The authors propose a hybrid heuristic that embeds reactive GRASP in a TS algorithm as a diversification method that provides elite candidate sets. Intensification is also done.

(Cano et al. 2000) present a new proposal for Cluster Analysis based on a GRASP, with the objective of overcoming the convergence to a local solution. It uses a probabilistic greedy Kaufman initialization method for getting initial solutions and the K-Means algorithm as a local search algorithm. The new proposal will become a new initialization approach to K-Means. Hence, they have compared some initialization methods for the K-Means algorithm: Random, Forgy, Macqueen and Kaufman, with the GRASP. Their results suggest that the Kaufman initialization method performs better than the other three methods and the hybrid GRASP - K-Means with Kaufman initialization improves these results even more. The proposed method obtains high quality solutions for the benchmark problems considered.

(Claro and Sousa 2001) present an object-oriented framework for multi-objective local search, integrating several generic strategies. The fundamental concepts of MOCO are introduced, followed by an overview of Multi-objective Meta-heuristics. A short review of

object-oriented approaches to meta-heuristics precedes the presentation of the framework, in which the general architecture and a template for multi-objective local search are highlighted. The paper ends with a short set of concluding remarks and lines for future developments.

(Resende 2001) and (Resende and Ribeiro 2003) explain the basic GRASP in detail and enhancements to the basic procedure are described. Several applications of GRASP are reported, showing how this method can find good approximate solutions to operations research problems and industrial applications.

(Ho and Haugland 2002) consider the Vehicle Routing Problem with Time Windows and Split Deliveries and present a solution method based on TS and four different neighborhood structures. This problem is an extension of the Vehicle Routing Problem, where the start of service at each customer must be within a time window, and a customer may be serviced by more than one vehicle. The heuristic gives promising results when applied to realistic instances of the problem. It is also adapted to the problem where splitting is not an option, and experimental results have shown that the heuristic has improved 5 of the 56 best published solutions to the Solomon benchmarks.

(Blum and Roli 2003) say that the field of meta-heuristics for the application to combinatorial optimization problems is a rapidly growing field of research. This is due to the importance of combinatorial optimization problems for both the scientific and the industrial world. They give a survey of the nowadays most important meta-heuristics from a conceptual point of view and outline the different components and concepts that are used in the different meta-heuristics in order to analyze their similarities and differences. They also introduce a framework in order to put different intensification and diversification components in relation to each other.

(Festa and Resende 2004) present an annotated bibliography of GRASP. The bibliography contains: background material; tutorials and surveys; enhancements to the basic method; hybrid methods; software; parallel GRASP; graph theory; quadratic and other assignment problems; location, layout, and cutting; covering, clustering, packing, and partitioning; routing; sequencing and scheduling; logic; manufacturing; transportation; telecommunications; electrical power systems; biology; VLSI design; drawing; and miscellaneous topics. The collection includes papers published in journals, books, M.Sc. and Ph.D. dissertations, and unpublished technical reports.

(Ahmadi and Osman 2005) propose a merger GRASP and Adaptive Memory Programming (AMP) into a new GRAMPS framework for the capacitated clustering problem

(CCP). A learning process is kept in charge of tracking information on the best components in an elite set of GRAMPS solutions. The information is strategically combined with problem-domain data to restart the construction search phase. At an early stage of the constructions, priorities are given to problem-domain data and progressively shifted toward generated information as the learning increases. GRAMPS is implemented with an efficient local search descent based on a restricted k -interchange neighborhood. Extensive experiments are reported on a standard set of benchmarks from the literature and on a new set of large instances. The results show that GRAMPS has an efficient learning mechanism and is competitive with the existing methods in the literature.

(Batun 2005) brings a study that aims to solve the Sales Territory Alignment Problem (STAP) by using a TS algorithm. This algorithm tries to construct the working areas of sales representatives of a firm considering the sales potential, number of customers and customer profile at each demand point. The author shows only preliminary results to small instances of the problem and says that the main difficulty to use a linear mathematical model is ensuring travel efficiency by using adjacency requirement.

(Haugland et al. 2005) consider the problem of designing districts for vehicle routing problems with stochastic demands. In particular, demands are assumed to be uncertain at the time the districts are being made, and they are only revealed after the districting decisions are determined. TS and multi-start heuristics for this stochastic districting problem are developed and compared. Some computational results show that TS is better than multi-start.

(Resende and Werneck 2006) present a multi-start heuristic for the uncapacitated facility location problem, based on a very successful method they originally developed for the p -median problem. They show extensive empirical evidence to the effectiveness of their algorithm in practice. For most benchmarks instances in the literature, they obtain solutions that are either optimal or a fraction of a percentage point away from it. Even for pathological instances, created with the sole purpose of being hard to tackle, their algorithm can get very close to optimality if enough time is given. It consistently outperforms other heuristics in the literature.

3.6 Summary

In this chapter we have presented a bibliographical review of the main building block subjects of this thesis. We also outlined the papers related to our research which served as a basis to the development of the new algorithm presented in the next chapter.

Chapter 4

MultiACE - A Multi-criteria Hybrid

Meta-heuristic

"The most exciting phrase to hear in science, the one that heralds new discoveries, is not "Eureka!" (I found it!) but "That's funny ..."

Isaac Asimov (1920 - 1992) US science fiction novelist & scholar

4.1 Introduction

This chapter presents a multi criteria hybrid meta-heuristic approach for the territories alignment problem (TAP). The model employs a Geographic Information System (GIS) and uses multi-criteria combinatorial optimization techniques as its components. The proposed hybrid meta-heuristic consists of the symbiosis between TS and GRASP and it is used to generate non-dominated alternatives. The general procedure and its algorithms are presented as well as its implementation in the GIS environment.

4.2 The multi-criteria hybrid meta-heuristic procedure

The hybrid meta-heuristic developed in this study uses TS and GRASP concepts in order to improve the search process. The symbiosis of these well-known meta-heuristic methods helps to support spatial data and process, introducing diversification and an

efficient memory structure to identify a wider range of alternatives and to analyze the trade-offs between alternatives.

The new algebraic approach presented in this thesis is based on a combination of:

- a Voronoi Diagram based strategy, to create solutions;
- a Two-sets partitioning strategy, to decrease the search space;
- a GRASP based candidates list;
- Tabu Search based short and long term memories;
- a multi-criteria technique to define the best solutions.

In the following sections, we discuss and present these components and a step-by-step description of the meta-heuristic solution procedure, its basic structure, initialization, subset generation and combination, dominance comparison, quality measurement and adaptive memory structure used. We also present the overall algorithm for the meta-heuristic designed for the multi-objective TAP.

4.2.1 Voronoi Diagram Based Solutions

A Voronoi Diagram, named after Georgy Voronoi, also called a Voronoi tessellation, a Voronoi decomposition, or a Dirichlet tessellation (after Lejeune Dirichlet), is a special kind of decomposition of a metric space determined by distances to a specified discrete set of objects in the space, e.g., determined by a discrete set of points (Fortune 1992).

Let S be a set of n sites in Euclidean space of dimension d . For each site c (the centers of the territories) of S , the Voronoi cell $V(c)$ (the territory) of c is the set of points that are closer to c than to other sites of S . The Voronoi diagram $V(S)$ is the space partition induced by Voronoi cells (Figure 4.1).

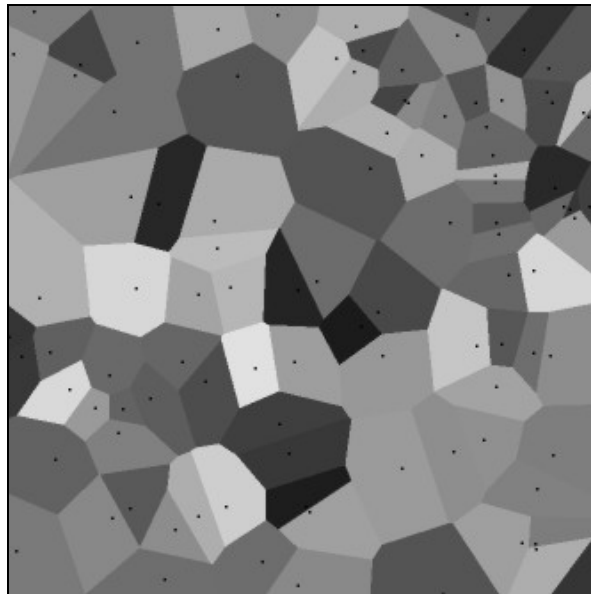


Figure 4.1 A Voronoi Diagram

The most important reason for choosing this strategy is its ability to create contiguous solutions for our problem, since the Voronoi Diagram always creates subsets that are contiguous around the centers that we select.

According to this, a solution for our problem can be created following these two simple steps:

- Selection of a number “ n ” of points as our territories “centers” (we use “centers” to indicate the points we use to generate the territories, not in the geometric sense);
- Allocation of each of all the other points to its closest center.

The other reason to use this strategy is that to store, to create or to recreate a solution, we just need the “centers”. This advantage is very useful when we are dealing with big sets of points and we need to make a lot of calculations based on the “centers” positions.

This Voronoi based strategy has an inconvenient characteristic for the alignment of territories: it doesn’t work very well when applied in a density heterogeneous area. In Chapter 6 we use our meta-heuristic in a real case study when we present a way to overcome this problem.

4.2.2 Two-Sets Partitioning Strategy to Decrease the Search Space

The strategy adopted to decrease the solutions search space divides the problem into some minor problems that can be solved more efficiently. It basically subdivides the original problem into new, smaller and less complex problems, starting with the original set of points and dividing it into two parts as follows:

Proportional to $\frac{n-1}{2}$ and $\frac{n+1}{2}$, if n is odd,

both proportional to $\frac{n}{2}$, if n is even.

where “ n ” is the number of territories. Then, this procedure is repeated until it generates n territories. This procedure is represented in the Algorithm 1.

The number of sub-divisions is always $n-1$ and, in the end, we obtain all subparts proportional to each other. Figure 4.2 shows an example of the strategy applied to divide 120 points into 4 and 3 territories divisions and Figure 4.3 shows an example of all the levels of partitions to divide 1000 points into 9 territories.

```

1. Initialize the number of territories
2. Calculate the NumberOfPartitions and the NumberOfPartitionLevels
3. for Level=1 to NumberOfPartitionLevels
4.   if Level = 1 then
5.     NumberToDivide = NumberOfTerritories
6.     if NumberToDivide is even then
7.       PartitionsMatrix(1, 1) = NumberToDivide / 2
8.       PartitionsMatrix(2, 1) = NumberToDivide / 2
9.     else
10.      PartitionsMatrix(1, 1) = (NumberToDivide - 1) / 2
11.      PartitionsMatrix(2, 1) = (NumberToDivide + 1) / 2
12.    end if
13.  else
14.    for j=1 to 2(Level-1)
15.      NumberToDivide = PartitionsMatrix(j, Level - 1)
16.      if NumberToDivide = 1 then
17.        PartitionsMatrix(2 * j - 1, Level) = 0
18.        PartitionsMatrix(2 * j, Level) = 0
19.      end if
20.      if NumberToDivide > 1 And NumberToDivide is even then
21.        PartitionsMatrix(2 * j - 1, Level) = NumberToDivide / 2
22.        PartitionsMatrix(2 * j, Level) = NumberToDivide / 2
23.      end if
24.      if NumberToDivide > 1 And NumberToDivide is odd then
25.        PartitionsMatrix(2 * j - 1, Level) = (NumberToDivide - 1) / 2
26.        PartitionsMatrix(2 * j, Level) = (NumberToDivide + 1) / 2
27.      end if
28.    end for
29.  end if
30. end for

```

Algorithm 1 The Partition Scheme Algorithm Structure

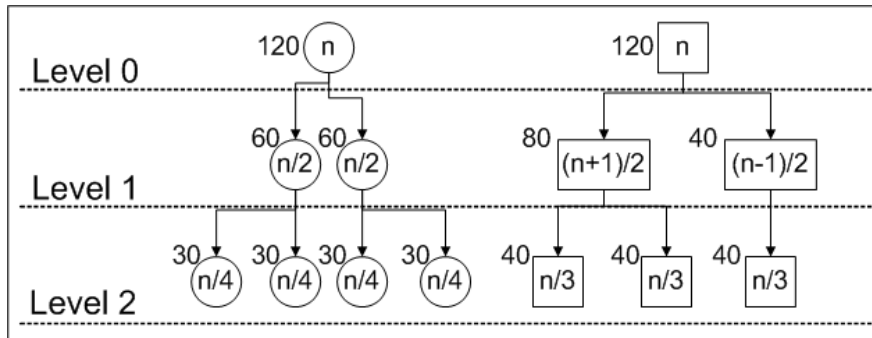


Figure 4.2 The Two-sets partitioning strategy applied to a 4 and a 3 territories partitions

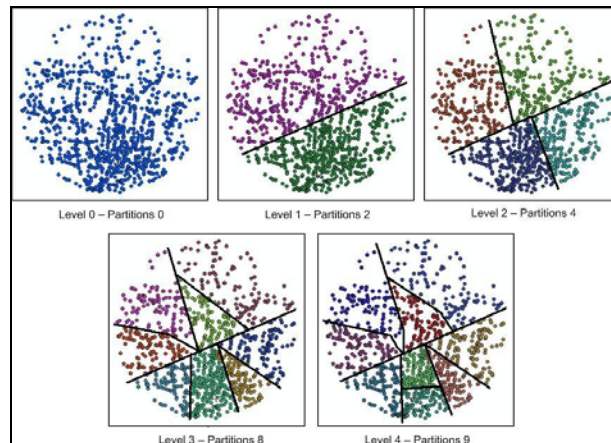


Figure 4.3 The levels of partitions to divide 1000 points into 9 territories

For each one of the levels of partitions defined with Algorithm 1, we also define the bounds for the summation of the selected criteria for the partitions. The higher level (for instance, the Level 2 in Figure 4.2) receives the bounds defined by the DM and the other levels receive bounds according to the number of the level, as follows:

$$UpperBound_i = \frac{UpperBound_{i+1}}{2}$$

$$LowerBound_i = \frac{LowerBound_{i+1}}{2}$$

where:

- i is the number of the level of partition;
- $UpperBound_i$ is the upper bound for the summation of the selected criteria for the partitions level of partition i ;
- $UpperBound_{i+1}$ is the upper bound for the summation of the selected criteria for the partitions level of partition $i+1$;
- $LowerBound_i$ is the lower bound for the summation of the selected criteria for the partitions level of partition i ;
- $LowerBound_{i+1}$ is the lower bound for the summation of the selected criteria for the partitions level of partition $i+1$.

This strategy reduces the search space because it decreases the number of possible candidates to the centers selection (Voronoi Diagram). Since we always divide each set into two, and only two, subsets, we always select two points as centers and allocate all the others to one, and only one, of these centers. This is a very fast way to generate and test solutions, because we don't need to improve the created solution in any way, we just create it and test.

In every division the main objective is to create two sub-sets that are proportionally divided. The algorithm tries to keep all the subsets balanced, when considering the attributes selected by the DM.

4.2.3 GRASP Based Candidates List

When we want to divide a set of points into two subsets, the first step, according to our algorithm, is to select the centers. We use a greedy strategy to select these centers, based in a candidates list sorted according to the Ideal Point Method. In this method the alternatives are ranked according to their distance from an ideal point. The ideal point is defined as the most desirable, weighted, hypothetical alternative (decision outcome). The closest alternative to the ideal point is the best alternative. The separation is measured in terms of metric distance (Janssen 1992).

This list is obtained after testing all the possible centers. This testing procedure is done allocating a “ k ” fixed number of points to each one of these centers. This “ k ” number of points can be, for instance, the mean of the number of points. The steps for the centers testing are:

- Selection of a center;
- Allocation of the nearest k points to this center to create a territory;
- Calculation of the distance of this territory to the ideal territory.

The distance to the Ideal Point (ideal territory) can be calculated using one of the three following measurements:

$$d_E(S, Ideal) = \sqrt{\sum_{k=1}^m w_k (|S_k - Ideal_k|)^2}$$

$$d_M(S, Ideal) = \sum_{k=1}^m w_k |S_k - Ideal_k|$$

$$d_C(S, Ideal) = \max_k \{w_k |S_k - Ideal_k|\}$$

where:

- d_E is the Euclidean Distance;
- d_M is the Manhattan Distance;
- d_C is the Chebyshev Distance;
- S is the territory that we want to compare to the ideal territory;
- S_k is the attribute k of S ;
- m is the number of attributes;
- $Ideal$ is the Ideal Point;
- $Ideal_k$ is the attribute k of the Ideal Point;
- w_k is the scale factor for the attribute k .

Each one of these three measurements has a different characteristic and a particular influence on the algorithm’s behavior. These differences can be explained by the “indifference curves” of the measurements. For instance, the Euclidean and Manhattan measurements try to minimize all the attributes at the same time, while the Chebyshev measurement tries to minimize the maximum value of the attributes.

These steps are executed with all the possible candidates and they are sorted according to the calculated distances. Since a wide exploration of the solution space is important to effectively make the algorithm navigate into various regions of the search domain, one important step in the proposed approach is to create solutions with a

diversification purpose to encourage the search process to examine random regions of the solution space.

We generate a solution using the candidates list, but we include randomness in this process. The centers ranking defines a probability ranking as well. This probability ranking is defined as follows:

$$\Pr(c) = \frac{i_c - n}{1 - n}(\delta) + \varphi$$

where, $\Pr(c)$ is the probability of the point c to be selected as a center, i_c is the position of the point in the candidates list, n is the number of points, φ is the lowest probability value and $\delta + \varphi$ is the highest probability value. If we choose $\delta=0.8$ and $\varphi=0.1$, the best center will have the probability of 90% and the worst center the probability of 10%. The range 90%-10% is used by the algorithm to select the centers. These values permit to choose the best centers while including a randomness that can help to search a wider space and to avoid local optima at the same time.

4.2.4 Tabu Search Based Memories

The MultiACE algorithm uses two memory structures: a short-term memory and a long-term memory. The long-term memory is simply a list of the solutions that have been tested. The short-term memory is more interesting and affects the way the algorithm explores the search space.

The type of TS list implemented in this algorithm is a list containing an attribute. When a center is selected and its territory is not a feasible one, this center receives a number equal to the number of times that it won't participate in the solutions creation process. It is a kind of quarantine. After this quarantine, the point can be chosen one more time as a center. The quarantine's size can be defined as a fixed one, but in our work it is dependent on the size of the points set, because during the whole territories process this number can vary from division to division.

In this TS list, a solution is represented by a pair of centers and if one of these territories is not feasible the other will probably be neither. So, we cannot apply the idea that a high quality solution can be found using part of a low quality solution, because we cannot use a part of a non-feasible solution.

4.2.5 Multi-criteria Technique

The multi-criteria technique used in the MultiACE algorithm is the Distance to the Ideal Point Technique. This technique was chosen because of the following reasons:

- It is numerically easy to understand and to implement in an algorithm;
- It is easy to insert the DM's preferences;
- It is very efficient (fast), and effective, considering the computational efforts.

A feasible territory is a territory whose attributes are inside the limits defined by the DM (considering the Ideal Point and the error bounds). For each one of the divisions, an Ideal Point is calculated. Then, the territories of the solutions are compared to the attributes bounds, and if both the territories are feasible, the solution is feasible.

If a solution is feasible, it is then compared to the best found solution. We use again the same concept of Ideal Point. But this time, the Ideal Point is “perfect balanced”, that is, the smaller the difference between the attributes of the solution's territories, the better the solution.

Each solution receives a grade, calculated according to its distance to the Ideal Point and the DM's weights. The smaller is this grade, the better will be the solution. So, the algorithm tries to minimize it. This grade is calculated as follows:

$$SG = \sum_j w_j \left(\sum_{k=1}^K a_{jk} - \sum_{l=1}^L a_{jl} \right)$$

where:

- SG is the solution grade;
- j is the number of criteria selected by the DM;
- w_j is the weight of the criteria j defined by the DM;
- K is the number of points in the first subset;
- L is the number of points in the second subset;
- a_{jk} is the value of the attribute j to the k point of the first subset;
- a_{jl} is the value of the attribute j to the l point of the second subset.

4.3 Implementation and application

The framework just described was implemented in ArcGIS version 9.2, a GIS software product of the Environmental Software Research Institute (ESRI). Being the world's leading GIS software supplier, ESRI provides GIS software products and related services to clients worldwide. ArcGIS, a high-end GIS software, is ESRI's flagship product, which has

the capabilities of automation, modification, management, analysis and display of geographical information.

The implementation coding was conducted with Visual Basic Application (VBA) embedded in ArcGIS and run on a Pentium Core 2 Duo 2.13 GHz PC with 2 Gb RAM. The input data to the multi-objective TAP was stored in the form of map layers called features layers of a personal geo-database format, which were handled and visualized using the ArcGIS. Each map layer contains a set of objects that were considered as elements of alternative solutions. As it was not possible to use the GIS standard operations alone to generate the solutions, the proposed multi-objective hybrid meta-heuristic method was specifically designed, coded and aggregated in the VBA code to tackle multi-objective TAP. The algorithm just described was coded and tested on some laboratory sets (randomly generated).

Instead of stating one complete mathematical model we preferred to build up more building blocks of broader applicability, since there are a lot of applications, each being slightly different from the other, requiring a different model. The same thought was applied to the objective(s). Since the TAP are of multi-objective nature, thus a single objective function is not easy to formulate, but it depends on the application, whose criteria are viewed as (hard) constraints and which should be optimized.

The methods we presented try to achieve the different goals in a heuristic way. Clearly we do not compute Pareto solutions or the Pareto frontier in a strict mathematical definition. In this section, we will first discuss the data requirements. Then, we will explain the setting of the constraints, objectives and algorithm options. The experiments will be reported in Chapter 5.

4.4 Data

To run the algorithm, the user (DM) needs:

- A personal geo-database layer with points representing the units to aggregate into territories. We are aware that working with a point-representation of the basic units could lead to problems with the compactness of territories, but the geographic obstacles, on the other hand, can be easily handled in our

approach by delivering to the algorithm the network distances between points;

- A database containing all the attributes of each unit (point).

The finding of solutions is done by ArcGIS and the MultiACE algorithm. The user must supply some parameters and choose some options (Figure 4.4).

MultiACE - Territories Balancer

Number of Territories: 4

Buttons: Run Balancing, Stop Searching, Hide Details

Tabs: Balancing | Solutions Checking | Special

Points Selection: Choose a points layer: Equipavalent, Choose the key field: CODGPS, Choose the solution field: TRIESTE, Subgroup selection: %CODE, 10

Balancing Criteria:

- First criterion: Conta, (%X-), (%X+), Weight: 4, 7, 1
- Second criterion: HORA, 5, 8, 2
- Third criterion: 5, 5, 1

Solutions Preferences:

- Compact territories as well as possible
- Consider decision maker's weight preferences

Stop Criteria:

- Processing time (min): 1
- Number of solutions to test: 100000
- Stop when find the first feasible solution
- Show each improved solution on the map
- Show every partition

Physical Distances:

- Euclidean Distances
- Manhattan Distances
- Chebyshev Distances
- Network Distances

Streets type:

- Consider One-Way Streets
- Consider Two-Way Streets

Short term memory: 40, Long term memory: 50000

Try all the possible solutions when the set is smaller or equal to: 450

Increase bounds if it is necessary?

Ideal Point Distances:

- Euclidean Distances
- Manhattan Distances
- Chebyshev Distances

Level: 1 of 2 - Partition: 1 of 3 - #1.S.: 945 - #1.S.: 7 - #1.L.S.: 6 - S.G.: 0,002 - Time: 60
 Level: 2 of 2 - Partition: 2 of 3 - #1.S.: 1540 - #1.S.: 44 - #1.L.S.: 41 - S.G.: 0,144 - Time: 60
 Level: 2 of 2 - Partition: 3 of 3 - #1.S.: 1005 - #1.S.: 42 - #1.L.S.: 40 - S.G.: 0,051 - Time: 60

Figure 4.4 MultiACE – Initial Screen – Parameters Definition

After the solutions generation, we can check the proposed solution and its compactness measurements (Figure 4.5).

MultiACE - Territories Balancer

Number of Territories: 4

Buttons: Run Balancing, Stop Searching, Hide Details

Tabs: Balancing | Solutions Checking | Special

Points selection: Choose a points layer: Equipavalent, Choose the key field: CODGPS, Choose the solution field: TRIESTE

Check: The solution!!!

Criteria:

- First criterion: Conta
- Second criterion: HORA
- Third criterion:

Type of Compact Measure:

- Mean of distances
- Sum of distances
- Maximum ray
- Maximum distance between points

Code	Crt. 01 Summ	Crt. 02 Summ	Crt. 03 Summ	Compact Measure
2504	211	278,9967	0	188,1815
2600	213	260,73806899	0	132,323196411133
3000	206	274,7643	0	126,1017
2607	212	276,438812255	0	86,325996388255

Mean: 210,500, 272,734, 528,932

Minimum: 206,000, 260,730, 80,326

Maximum: 213,000, 278,997, 188,182

% Minimum: 2,138, 4,399, 180

% Maximum: 1,188, 2,296,

Figure 4.5 MultiACE – Initial Screen – Solutions Checking

4.5 Constraints

Constraints in the multi-objective spatial territories balancing problem can be divided into hard and soft constraints. Soft constraints will consider some criteria that are treated as soft requirements in an objective function, while hard constraints are the unchangeable factors such as rivers and one-way streets, for instance.

In order to avoid splitting a territory with respect to natural boundaries such as major rivers or bodies of water, the DM must choose the network distances option. The algorithm will read a text file containing all the distances between each pair of points and will use these distances during the calculation process. This file of distances can be obtained using the Network Analyst embedded in ArcGIS 9.2.

The soft constraints include:

- Each basic unit is assigned to one district;
- The number of districts;
- No basic unit can be assigned to an unselected center;
- The limits to the territories sizes (balance) are taken into account.

4.6 Objectives

Two main objectives are considered in this model:

- The first objective is that the territories must be contiguous and as compact as possible. There are many ways to measure compactness and the MultiACE uses the summation of the Euclidean distances of all points to the geometric center of the territories. This objective is achieved using the Voronoi Diagram based strategy. Depending on the preferences of the DM, the algorithm seeks for the most compact solution or for the most preferred solution, based on the DM preference weights.
- The second objective is that the territories must be balanced by one or more criteria. This means that the summation of the attributes for every territory must lie within the defined interval. This objective (or constraint) can be formulated in the following way:

$$(1 - \beta_j)(Ideal_{j1}) \leq \left(\sum_{k=1}^K a_{jk} \right) \leq (1 + \varphi_j)(Ideal_{j1})$$

and

$$(1 - \beta_j)(Ideal_{j2}) \leq \left(\sum_{l=1}^L a_{jl} \right) \leq (1 + \varphi_j)(Ideal_{j2})$$

where:

- j is the number of criteria selected by the DM;
- K is the number of points in the first subset;

- L is the number of points in the second subset;
- $Ideal_{j1}$ is the value of the attribute j to the Ideal Point 1; (Each subset is compared to a different Ideal Point, since the subsets can be of different sizes. If the subsets are not of different sizes, the Ideal Point 1 is equal to the Ideal Point 2.);
- $Ideal_{j2}$ is the value of the attribute j to the Ideal Point 2;
- a_{jk} is the value of the attribute j to the k point of the first subset;
- a_{jl} is the value of the attribute j to the l point of the second subset;
- β_j is the lower bound for the attribute j of the subsets;
- φ_j is the upper bound for the attribute j of the subsets.

4.7 Algorithm's Parameters

The algorithm's parameters are:

- Number of territories to generate;
- Layer of points to process;
- Criteria to use in the multi-criteria analysis;
- Weights of each criterion used in the analysis and the solutions preferences;
- Stopping criteria;
- The type of measurement for the distance between points.

The number of territories to generate is the first and the easiest parameter to choose and understand. According to this number, the algorithm will elaborate the scheme of partitions to follow until the final solution (Figure 4.6).

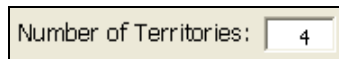


Figure 4.6 MultiACE - Number of Territories Selection

The layer of points to process is the layer of information that will be used in the analysis. The user has to choose not only the layer, but the key field, the field where the solution will be stored, and, if it is necessary, a field that indicates a subgroup selection. The subgroup selection is used to generate territories using only a selected set of units that are in the selected layer (Figure 4.7).

Figure 4.7 MultiACE - Layer, Key Field, Solution Field and Subgroup Selection

The criteria (up to three) are the fields where the values of the attributes for each unit are stored. The DM first selects how many criteria he will use then he selects each field separately. The last definition about criteria is the weight definition. These weights will guide the algorithm through the solutions search (Figure 4.8).

Figure 4.8 MultiACE - Criteria Options

The stopping criteria are the criteria the algorithm will use to stop the solutions search. The options are: a time limit, a limit to the number of solutions tested, finding of the first feasible solution or just the DM will (Figure 4.9).

Figure 4.9 MultiACE - Stopping Criteria

To measure the distance to the “Ideal Point”, we can use three options: Euclidean distances, Manhattan distances or Chebyshev distances (Figure 4.10) and to measure the physical distance between points, we can use four options: the three just mentioned or Network distances, and in this one there are still two options, considering two-way or one-way streets (Figure 4.11).

Figure 4.10 MultiACE - Ideal Point Distances Measurements

Figure 4.11 MultiACE - Physical Distances Measurements

The algorithm reads the Network distances from a file and the Euclidean, Manhattan and Chebyshev distances are calculated as follows:

$$d_E(a,b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2}$$

$$d_M(a,b) = |x_a - x_b| + |y_a - y_b|$$

$$d_C(a,b) = \max(|x_a - x_b|, |y_a - y_b|)$$

where:

- d_E is the Euclidean Distance;
- d_M is the Manhattan Distance;
- d_C is the Chebyshev Distance;
- x_a, y_a are the coordinates for the point A;
- x_b, y_b are the coordinates for the point B.

There are three special and important parameters that influence the way the algorithm works. We can choose if the algorithm will look for solutions in a heuristic way or if it will make an exhaustive search. The exhaustive search is indicated when we have small sets of points, because the algorithm can be more efficient. We have to define the size of the sets that will make the algorithm to use the exhaustive search procedure. We also define the maximum number of analyzed solutions to keep in memory (long-term memory) and the percentage of the number of points that will define the short-term memory (Figure 4.12).

The next decision to make is related to the objective function (Figure 4.13). The DM must decide if the algorithm will look for most compact solutions or if it will look for solutions that are most preferred by the DM. This is an important question, because it can influence the shape of the territories and the final results. Figure 4.14 shows two different territories solutions for the same problem. The first one shows a set of territories optimized for the DM preferences (weights), while the second shows a set of territories that are more compact than the first solution.

☒ Try all the possible solutions when the set is smaller or equal to
 Short term memory: Long term memory:

Figure 4.12 MultiACE – Algorithm's Memory Parameters

Solutions Preferences
☒ Compact territories as well as possible
☐ Consider decision maker's weight preferences

Figure 4.13 MultiACE – Solutions Preferences

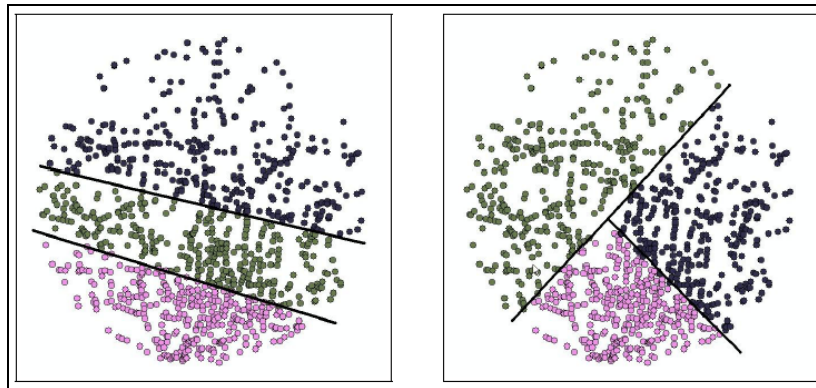


Figure 4.14 Two Different Solutions for the Same Problem

The next parameter to define is an optional parameter and is related to the bounds defined for each one of the partitions carried out by the algorithm. The DM can allow the algorithm to increase the bounds considered by the algorithm for the partition until it finds a feasible solution for that partition, if it is necessary. This parameter is useful to help finding solutions when the DM defines very restricted bounds for the territories criteria balancing (Figure 4.15).

☐ Increase bounds if it is necessary?

Figure 4.15 MultiACE – Special Parameter

There is still a special parameter to define, but only when we are checking the proposed solutions (Figure 4.16). It is the territories Compactness Measurements. We can calculate it in four ways:

- Mean of the distances from the points to the centroid of the territory;
- Summation of the distances from the points to the centroid of the territory;
- Maximum radius of the territory. The distance from the centroid of the territory to the most distant point;
- Maximum distance between two points of a territory.

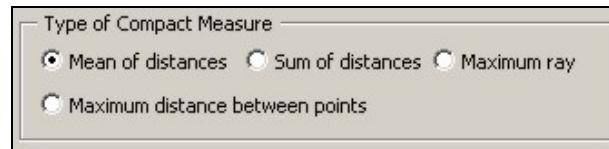


Figure 4.16 MultiACE – Types of Compactness Measurement

4.8 Algorithm Structure

The MultiACE algorithm can be summarized in the structure listed in the Algorithm

2.

1. *Initialize the number of territories*
2. *Calculate the Number of Partition Levels (NumLev)*
3. *Calculate the Number of Partitions in each Level (NumPartOfLev)*
4. *Generate the Partitions Scheme*
5. *for Level=1 to NumLev do*
6. *for Partition=1 to NumPartOfLev*
7. *Calculate the Proportion Between the Territories*
8. *Read the points attributes*
9. *Calculate the Ideal Point*
10. *Read or Calculate the distances between points*
11. *Evaluate each point as a possible center*
12. *Generate the candidates list*
13. *while (Keep calculating solutions = TRUE)*
14. *Define the centers to generate a solution*
15. *Generate the solution*
16. *Calculate the statistics of the solution*
17. *Evaluate the solution*
18. *Test the Stopping Criteria*
19. *end while*
20. *end for*
21. *end for*

Algorithm 2 MultiACE Algorithm Structure

4.9 Summary

In this chapter we have presented the main concepts used in the MultiACE algorithm. We outlined the multi-criteria hybrid meta-heuristic elaborated and implemented in a GIS environment and we have described some of the most important options that the algorithm offers to the DM to use in territories alignment problems. In the next Chapters we will present the tests and the application of the algorithm in a real case study.

Chapter 5

Algorithm Testing

"We have to learn again that science without contact with experiments is an enterprise which is likely to go completely astray into imaginary conjecture."

Hannes Alfvén (1908 – 1995) Swedish physicist

5.1 Introduction

This chapter presents the tests performed to evaluate the effectiveness of the algorithm proposed in this thesis in the context of some synthetic (laboratory) datasets².

We decided to use this kind of datasets due to the lack of specific datasets and solved instances available in repositories. All these cases were randomly generated and we analyzed three sets of experiments. Each one of these sets of experiments was carried out to evaluate a specific feature of the algorithm. The random points in all datasets were generated with the random points' generator available in ArcGIS and the values to the criteria were generated using the random procedure available in VBA.

In next sections, we outline the sets of experiments, we explain the procedure used to generate the datasets, we present the results obtained in each one of the trials and we highlight the main characteristics of the solutions obtained with the algorithm proposed in this thesis.

² All datasets are included on the CD-ROM attached to this thesis and will be send upon request by the author to interested readers.

5.2 The Sets of Experiments

We performed a series of carefully prepared tests on synthetic data generated with random procedures. These tests were grouped into three main types of experiments:

- **Tuning Experiments:** where we could verify the best combination of two parameters: the maximum amount of time to run the algorithm and the size of the short-term memory;
- **Exhaustive³ MultiACE vs. CPLEX Experiments:** where we could compare the solutions generated using the Exhaustive MultiACE to the optimal solutions generated using the CPLEX solver;
- **Heuristic⁴ MultiACE vs. Exhaustive MultiACE:** where we could compare the solutions generated using the Exhaustive MultiACE to the solutions generated using the Heuristic MultiACE procedure.

The various experiments performed compare the effect of algorithm parameters variation and datasets variation, and, in all the tests we performed, we compared the effects in distortion and time to generate the solutions. The distortion considered is the summation of the distances to the centroid of the territories. These tests have been designed in order to measure the effect of the variation of some parameters: the number of points in the dataset, the number of criteria, the number of territories generated and the bounds defined for the criteria. The mathematical programming optimization experiments were coded with AMPL and solved using the commercial optimizer CPLEX (version 10.0) running on Windows XP.

5.3 The Datasets

We generated seven random datasets with 100, 200, 300, 400, 500, 600 and 700 points using the random points' generator available in ArcGIS. For each one of these points, we generated three criteria values, using when necessary the random procedure available in VBA.

- Criterion 1: 1, for all points;

³ Exhaustive – All the possible solutions for each partition are tested.

⁴ Heuristic – Using the GRASP based candidates list and the TS based memories to select the solutions for each partition.

- Criterion 2: a random value in the range of 1000 to 10000;
- Criterion 3: a random value in the range of 0 to 2.

The first criterion was intentionally defined as 1 to every point because, when we use this criterion, we will be creating territories balanced according to the number of points. For the other two criteria we defined very different ranges to show that the algorithm can deal with different scales criteria.

5.4 The CPLEX Formulation

Since we were using 7 synthetic datasets, we needed to know the optimal solutions for our instances. To obtain these optimal solutions, we developed a formulation for our problem and used the CPLEX solver to find the optimal solutions and to compare them to the solutions obtained with the MultiACE algorithm.

We developed a p-median based binary formulation (Lorena and Senne 2004), adding the necessary constraints to create what we defined as “multi-criteria balanced p-median formulation” (a similar formulation can be found in (Won and Chang Lee 2004) and the author called it “modified p-median formulation”), as follows:

Decision variables:

$$x_{ij} = \begin{cases} 1, & \text{if point } i \text{ is allocated to point } j \\ 0, & \text{case else} \end{cases} \quad (1)$$

$$x_{jj} \in \{0, 1\} \quad (2)$$

Objective function:

$$\min \sum_{j \in I} \sum_{i \in I} d_{ij} x_{ij} \quad (3)$$

Subject to:

$$\sum_{j \in I} x_{ij} = 1, \forall i \in I \quad (4)$$

$$\sum_{i \in I} x_{ii} = m \quad (5)$$

$$x_{ij} \leq x_{jj}, \forall i, \forall j \in I \quad (6)$$

$$x_{jj} (1 - \alpha_{ka}) \left(\frac{\sum_{i \in I} c_{ki}}{m} \right) \leq \sum_{i \in I} c_{ki} x_{ij}, \forall k, \forall j \in I \quad (7)$$

$$x_{jj}(1 + \alpha_{kb}) \left(\frac{\sum_{i \in I} c_{ki}}{m} \right) \geq \sum_{i \in I} c_{ki} x_{ij}, \forall k, \forall j \in I \quad (8)$$

where:

- I is the set of distinct points;
- d_{ij} is the Euclidean distance between point i and point j ;
- m is the number of territories;
- K is the set of criteria defined for the problem;
- c_{ki} is the value of the criterion k of the point i ;
- a_{ka} is the lower bound percentage defined for the criterion k ;
- a_{kb} is the upper bound percentage defined for the criterion k .

The objective of formulation (1)-(8) is to minimize the summation of the Euclidean distances from each point to the center it is allocated to.

Constraints (4)-(6) define the regular p-median problem. Constraint (4) defines that each point can be allocated to only one center. Constraint (5) defines the number m of territories to be created and constraint (6) defines that a point cannot be allocated to a point that is not a center of a territory.

Constraints (7) (lower bounds) and (8) (upper bounds) describe the feasible region of the “multi-criteria balanced p-median formulation”. The addition of these constraints leads to the creation of balanced solutions according to the defined criteria.

5.5 Computational Results

5.5.1 Tuning Experiments

The first set of experiments was carried out to tune two important parameters of the algorithm: the time limit for each partition and the size of the short-term memory. These parameters are very important because they strongly influence the algorithm’s behavior.

The first parameter tuned was the size of the short-term memory. We carried out the experiments varying the parameters of our laboratory instances as shown in Table 5.1.

Table 5.1 Tuning Experiments - Short-term Memory - Parameters Setting

Parameter	Value(s)
Number of Points (Datasets)	100, 200, 300, 400, 500, 600 and 700
Number of Territories	2
Criteria Sets	7 possible combinations of 3 criteria
Criteria Bounds	We varied the bounds of our criteria, but used the same range for the lower and upper bounds of every criterion of the instances. We used only 2 ranges (-/+5% or -/+10%) because we believe that these bounds can represent real cases bounds. Bounds smaller than -/+5% are too much restrictive and greater than -/+10% are almost unbalanced
Time Limit	We defined a limit of 1 minute to each partition. We decided to give a small amount of time, allowing that the effect of the amount of short-term memory could be exposed
Short-Term Memory	We varied the amount of the short-term memory, from 10% to 100% and carried out 10 runs for each dataset. <u>Table 5.2</u> shows an example of the results obtained for 1 set of runs
Physical Distance Between Points	Euclidean
Preferences on Solutions	Territories as compact as possible
Ideal Point Measurement	Euclidean
Criteria Weights	1 to every criterion

We ran the algorithm 9800 times (7 criteria sets $\times$ 10 memory sizes $\times$ 10 runs $\times$ 7 datasets $\times$ 2 bound ranges). An example of summary of these runs can be found in Table 5.2. For each group of 10 short-term memory tests, we collected the amount of memory that generated the maximum number of tested and feasible solutions. The summary of the occurrences of these maxima is presented in Table 5.3.

According to this summary, we can verify that to these datasets the best range of memory is between 40% and 50% of the number of points of the instances. The reason why this occurs is that the short-term memory works as quarantine and if we give too much of this kind of memory, the algorithm will put a lot of possible centers in this quarantine, and this will reduce the size of the search space it can explore.

On the other hand, these values for the best range were, in some way, expected because the greater the number of solutions tested, the greater the probability of finding the best solutions.

Table 5.2 Results for the Short-term Memory Tuning

Memory (%)	Bound: +/- 5%		Bounds: +/- 10%	
	Number of Solutions Tested	Number of Feasible Solutions	Number of Solutions Tested	Number of Feasible Solutions
10	175	25	107	47
20	623	66	331	99
30	1107	137	781	213
40	1673	237	944	285
50	1006	138	1701	480
60	54	4	264	81
70	61	11	83	24
80	55	6	67	17
90	55	5	74	24
100	61	11	75	25

Table 5.3 Short-term Tuning Summary

Memory (%)	Occurrences of The Maximum of the Number of Tested and Feasible Solutions		
	Bound: +/- 5%	Bound: +/- 10%	Total
10	0	0	0
20	2	5	7
30	39	8	47
40	97	52	149
50	352	352	704
60	0	73	73
70	0	0	0
80	0	0	0
90	0	0	0
100	0	0	0
Total	490	490	980

The second parameter we tuned was the time limit for the partitions. We carried out the experiments varying the parameters of our laboratory instances as shown in Table 5.1, but we varied the parameters “Time Limit” and “Short-Term Memory” according to Table 5.4.

Table 5.4 Tuning Experiments - Time Limit - Parameters Setting

Parameter	Value(s)
Time Limit	We varied the time limits from 30 to 600 seconds, with increments of 30 seconds, to each partition
Short-Term Memory	We used a fixed value of 40%. We have chosen this value because it is in the range of the best solutions obtained in the short-term tuning

We ran the algorithm 1960 times (7 criteria sets x 20 runs x 7 datasets x 2 bound ranges). For each group of 20 runs (Table 5.5 shows an example), we can notice that the number of tested and feasible solutions does not increase a lot if we give more time to the algorithm. For these given datasets, the ranges of 120 to 180 seconds are very efficient amounts of time, since the number of tested and feasible solutions is near to the maximum that the algorithm can explore.

Table 5.5 Time Limit Tuning

Dataset	Time (s)	Bound: +/- 5%		Bounds: +/- 10%	
		Tested Solutions	Feasible Solutions	Tested Solutions	Feasible Solutions
100	30	544	87	432	124
	60	1878	243	294	83
	90	2150	265	1394	406
	120	2075	249	1361	362
	150	2068	277	1464	415
	180	2185	267	1416	417
	210	2250	269	1443	408
	240	2173	259	1275	397
	270	2331	254	1046	339
	300	2432	276	1564	452
	330	2293	294	1182	341
	360	2136	275	1463	415
	390	2232	294	1520	428
	420	2244	270	1502	450
	450	2336	284	1198	374
	480	2204	316	1472	419
	510	2284	310	1575	422
	540	2293	278	1556	448
	570	2452	305	1562	429
	600	2391	274	1479	406

5.5.2 Exhaustive MultiACE *vs.* CPLEX Experiments

In this set of experiments we decided to use the Exhaustive MultiACE to show that the strategy to create solutions (Voronoi based strategy) and the strategy to decrease the search space (Two-set partitioning strategy) work very well and can generate very high-quality solutions. So, we did not use the GRASP candidates list and the TS based memories in this set of experiments. We carried out the experiments varying the parameters of our problems as shown in Table 5.6.

We ran the algorithm 238 times, but we just could compare 229 results to the ones from the CPLEX runs. Nine of these were aborted because they would take more than 43200 seconds (12 hours) to find a solution. Table 5.8 shows an example⁵ of the runs for one criteria set and Table 5.9 shows the number of runs for each criteria set and each bound range.

⁵ All summaries of the runs are included on the CD-ROM attached to this thesis.

Table 5.6 Exhaustive MultiACE vs. CPLEX Experiments – Parameters Setting

Parameter	Value(s)
Number of Points (Datasets)	100, 200, 300, 400 and 500
Number of Territories	We did, for each one of the datasets, a set of splittings, shown in Table 5.7. We defined a maximum number of territories to each one of the datasets based on: $\frac{\text{Maximum Number of Territories}}{\text{Number of Points in the Dataset}} \leq 1,5\%$ The exception is the 100 points dataset, because we cannot divide it in less than 2 territories. We adopted these limits because, for the real cases we know, we rarely find territories smaller than the ones we obtain with these limits
Criteria Sets	7 possible combinations of 3 criteria
Criteria Bounds	We varied the bounds of our criteria, but used the same range for the lower and upper bounds of every criterion of the instances. We used only 2 ranges (-/+5% or -/+10%) because we believe that these bounds can represent real cases bounds. Bounds smaller than -/+5% are too much restrictive and greater than -/+10% are almost unbalanced
Time Limit	120 seconds
Short-Term Memory	We used a fixed value of 40%. We have chosen this value because it is in the range of the best solutions obtained in the short-term tuning
Physical Distance Between Points	Euclidean
Preferences on Solutions	Territories as compact as possible
Ideal Point Measurement	Euclidean
Criteria Weights	1 to every criterion

Table 5.7 Number of Minimum and Maximum Territories for The Datasets

Dataset	Min Territories	Max Territories
100	2	2
200	2	3
300	2	4
400	2	6
500	2	7
600	2	9
700	2	10

Table 5.8 Example of The Runs for One Criteria Set

Dataset	Number of territories	Bound: +/- 5%							Bound: +/- 10%						
		CPLEX TIME (s)	MultiACE TIME (s)	CPLEX DIST. SUMM. (m)	CPLEX DIST. SUMM. (m)*	CPLEX DIFF. %	MultiACE DIST. SUMM. (m)	DIST. SUMM. DIFF %	CPLEX TIME(s)	MultiACE TIME(s)	CPLEX DIST. SUMM. (m)	CPLEX DIST. SUMM. (m)*	CPLEX DIFF. %	MultiACE DIST. SUMM. (m)	DIST. SUMM. DIFF %
100	2	8	2	7312	7306	-0,08%	7297	-0,11%	6	3	7307	7299	-0,11%	7297	-0,02%
200	2	108	20	14315	14333	0,12%	14335	0,01%	180	21	14304	14311	0,05%	14313	0,02%
	3	109	26	11570	11575	0,04%	11697	1,06%	102	26	11508	11530	0,20%	11718	1,63%
300	2	874	64	21380	21389	0,04%	21387	-0,01%	1613	64	21379	21391	0,05%	21387	-0,02%
	3	1352	83	16977	16977	0,00%	17416	2,59%	1021	83	16955	16845	-0,65%	17389	3,23%
	4	535	81	14127	14128	0,01%	14140	0,08%	496	81	14110	14107	-0,02%	14110	0,02%
400	2	3569	150	29952	29848	-0,35%	29781	-0,23%	3270	148	29943	29929	-0,04%	29781	-0,50%
	3	2571	195	23875	23868	-0,03%	24297	1,80%	5041	191	23831	23819	-0,05%	24282	1,95%
	4	3084	191	19557	19535	-0,11%	19534	-0,01%	1722	189	19528	19526	-0,01%	19510	-0,08%
	5	1798	205	17672	17659	-0,07%	17901	1,37%	4830	202	17649	17681	0,18%	17924	1,37%
	6	1415	212	15795	15761	-0,22%	16373	3,89%	1798	200	15731	15670	-0,38%	16348	4,32%
500	2	8326	287	35658	35664	0,02%	35664	0,00%	12530	286	35658	35664	0,02%	35664	0,00%
	3	5720	372	28576	28583	0,03%	29040	1,60%	6916	369	28575	28579	0,02%	29037	1,60%
	4	4495	364	24033	24023	-0,04%	24161	0,57%	5371	364	24004	23990	-0,06%	24179	0,78%
	5	4975	390	21960	21963	0,01%	22418	2,07%	4163	390	21914	21905	-0,04%	22264	1,64%
	6	12137	388	19878	19852	-0,13%	20549	3,51%	3693	388	19818	19807	-0,05%	20563	3,82%
	7	3589	392	18248	18223	-0,14%	18636	2,27%	3257	386	18227	18216	-0,06%	18585	2,03%

Remark: The CPLEX calculates the total summation of distances between each point and its territory center (CPLEX DIST. SUMM.). The MultiACE calculates the summation of distances between each point and the centroid of the territory (MultiACE DIST. SUMM.). After obtaining the CPLEX solutions, we imported them to the ArcGIS and calculated the summation of distances between each point and the centroid of the territory (CPLEX DIST. SUMM.*). So, we could compare the results obtained with MultiACE to the ones obtained with CPLEX.

Table 5.9 Number of Runs Summary

Criteria	Number of Runs		
	Bound: +/- 5%	Bound: +/- 10%	Total
1	17	17	34
2	17	17	34
3	17	17	34
1 and 2	17	17	34
1 and 3	17	17	34
2 and 3	13	17	30
1, 2 and 3	13	16	29
Total	111	118	229

After this, we compared the 229 runs according to their summation of distances and the results are shown in Table 5.10. The results show that the proposed hybrid meta-heuristic gives high-quality solutions. In our opinion, the results obtained with CPLEX allow a further appreciation of the quality of the results obtained with the heuristics.

We can notice that in 94,76% of the runs, the difference between MultiACE results and CPLEX results is not greater than 5% and the amount of time to find the solutions, for these datasets, is always greater when we use the CPLEX formulation. The mean of the differences is 1,3% and the greatest value is 19,9%, but it occurred just once. The greatest differences appeared when we set the values for the number of territories near to the maximum number for each dataset and when we tried to balance three criteria at the same time, increasing the complexity of the partition problem.

For all runs, CPLEX is the most time-consuming method. For this set of experiments, the required computation times for the smallest datasets (100 and 200 points) are not significantly larger than the ones of MultiACE but as the size of problems grows the required computation times become much larger, compared to MultiACE.

Table 5.10 Results Comparison Summary

Experiment Summary			
Summ. Of Distances Differences	Occurrences	Occurrences (%)	Occurrences Cumulative (%)
$\leq 1\%$	107	46,72%	46,72%
$>1\%$ and $\leq 2\%$	43	18,78%	65,50%
$>2\%$ and $\leq 3\%$	37	16,16%	81,66%
$>3\%$ and $\leq 4\%$	22	9,61%	91,27%
$>4\%$ and $\leq 5\%$	8	3,49%	94,76%
$>5\%$	12	5,24%	100,00%
Total	229	100,00%	

5.5.3 Heuristic MultiACE vs. Exhaustive MultiACE

Once we have shown that the Exhaustive MultiACE can generate high-quality solutions, we decided to find out if our heuristic method using the strategy to decrease the search space can perform well when compared to the exhaustive method. Then, we carried out the experiments varying the parameters of our problems, as shown in Table 5.6, but we varied the parameters “Time Limit” and “Short-Term Memory” according to Table 5.11.

Table 5.11 Heuristic MultiACE vs. Exhaustive MultiACE – Parameters Setting

Parameter	Value(s)
Time Limit	For the heuristic experiments, 120 seconds for each partition. For the exhaustive experiments, the time required to finish the partitions.
Short-Term Memory	We used a fixed value of 40%. We have chosen this value because it is in the range of the best solutions obtained in the short-term tuning

We ran the algorithm 952 times, 476 heuristically and 476 exhaustively. Table 5.12 shows the number of runs for each criteria set and each bound range and we can notice that the algorithm solved 99,26% (945 out of 952) of the problems. After this, we could make 470 comparisons between the methods.

We compared the 470 pairs of results according to their summation of distances and the results are shown in Table 5.13. We can notice that in 84,68% of the runs, the difference between the heuristic and the exhaustive strategies is not greater than 5%. The mean of the differences is 2,0% and the greatest value is 36,8% (only one occurrence). Once again, the greatest differences appeared when we set the values for the number of territories near to the maximum number for each dataset and when we tried to balance three criteria at the same time, increasing the complexity of the partition problem (see Table 5.15).

Finally, Table 5.14 shows an example of the runs for one criteria set and Table 5.15 shows an example of the runs for three criteria set.

Table 5.12 Number of Runs for each Criteria Set and each Bounds Range

Criteria	Bounds	Exhaustive		Heuristic	
		Solved	Failed	Solved	Failed
1	-5% / +5%	34	0	34	0
	-10% / +10%	34	0	34	0
2	-5% / +5%	34	0	34	0
	-10% / +10%	34	0	34	0
3	-5% / +5%	34	0	34	0
	-10% / +10%	34	0	34	0
1 and 2	-5% / +5%	34	0	34	0
	-10% / +10%	34	0	34	0
1 and 3	-5% / +5%	34	0	33	1
	-10% / +10%	34	0	34	0
2 and 3	-5% / +5%	34	0	33	1
	-10% / +10%	34	0	34	0
1, 2 and 3	-5% / +5%	33	1	30	4
	-10% / +10%	34	0	34	0
Subtotal		475	1	470	6
Total		952			

Table 5.13 Results Comparison Summary

Experiment Summary			
Summ. Of Distances Differences	Occurrences	Occur. (%)	Cumulative (%)
≤1%	298	63,40%	63,40%
>1% and ≤2%	41	8,72%	72,13%
>2% and ≤3%	30	6,38%	78,51%
>3% and ≤4%	16	3,40%	81,91%
>4% and ≤5%	13	2,77%	84,68%
>5%	72	15,32%	100,00%
Total	470	100,00%	100,00%

Table 5.14 Example of the Runs for One Criteria Set

Dataset	Number of Territories	Bound: +/- 5%						Bound: +/- 10%					
		Exhaustive		Heuristic		Dif %		Exhaustive		Heuristic		Dif %	
		Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)
100	2	4	7.297,457	120	7.297,457	1400,00%	0,00%	5	7.297,457	120	7.297,457	1100,00%	0,00%
200	2	20	14.334,621	120	14.342,918	200,00%	0,06%	22	14.313,470	120	14.315,598	172,73%	0,01%
	3	30	11.697,073	240	11.908,535	300,00%	1,81%	27	11.629,328	240	11.721,894	344,44%	0,80%
300	2	76	21.386,645	120	21.387,421	-21,05%	0,00%	68	21.386,645	120	21.388,389	-11,76%	0,01%
	3	86	17.372,357	240	17.461,084	39,53%	0,51%	82	17.388,591	240	17.448,777	46,34%	0,35%
	4	100	14.139,942	360	14.149,095	80,00%	0,06%	81	14.109,954	360	14.133,439	122,22%	0,17%
400	2	148	29.780,633	120	29.781,552	-59,46%	0,00%	148	29.780,633	120	29.781,429	-59,46%	0,00%
	3	192	24.297,142	240	24.531,381	-37,50%	0,96%	191	24.282,013	240	24.451,511	-37,17%	0,70%
	4	188	19.534,198	360	19.534,198	-4,26%	0,00%	188	19.509,526	360	19.513,319	-4,26%	0,02%
	5	201	17.901,255	480	17.907,962	19,40%	0,04%	201	17.906,513	480	17.862,777	19,40%	-0,24%
	6	200	16.373,360	600	16.377,761	50,00%	0,03%	200	16.347,610	600	16.435,825	50,00%	0,54%
500	2	286	35.663,715	120	35.666,821	-79,02%	0,01%	286	35.663,715	120	35.671,350	-79,02%	0,02%
	3	371	29.039,619	240	29.118,833	-67,65%	0,27%	367	29.007,848	240	29.019,597	-67,30%	0,04%
	4	362	24.160,613	360	24.254,588	-50,28%	0,39%	362	24.178,502	360	24.184,966	-50,28%	0,03%
	5	387	22.404,298	480	22.455,167	-37,98%	0,23%	386	22.263,615	480	22.513,236	-37,82%	1,12%
	6	384	20.549,112	600	20.382,656	-21,88%	-0,81%	382	20.544,960	600	20.533,504	-21,47%	-0,06%
	7	390	18.636,322	720	18.815,003	-7,69%	0,96%	387	18.585,176	720	18.724,373	-6,98%	0,75%
600	2	490	43.524,514	120	43.530,949	-87,76%	0,01%	490	43.524,514	120	43.524,514	-87,76%	0,00%
	3	634	35.566,862	240	35.687,382	-81,07%	0,34%	628	35.554,890	240	35.533,272	-80,89%	-0,06%
	4	619	28.462,574	360	28.468,070	-70,92%	0,02%	618	28.476,045	360	28.437,930	-70,87%	-0,13%
	5	664	26.555,849	480	26.601,397	-63,86%	0,17%	660	26.493,661	480	26.778,141	-63,64%	1,07%
	6	656	24.113,213	600	24.195,283	-54,27%	0,34%	656	24.077,363	600	24.148,510	-54,27%	0,30%
	7	664	22.091,927	720	22.216,831	-45,78%	0,57%	662	22.090,235	720	22.083,139	-45,62%	-0,03%
	8	656	21.496,236	840	21.404,781	-35,98%	-0,43%	654	21.429,959	840	21.315,494	-35,78%	-0,53%
	9	667	20.154,326	960	19.480,805	-28,04%	-3,34%	667	20.221,293	960	20.217,462	-28,04%	-0,02%
700	2	775	51.080,497	120	51.091,629	-92,26%	0,02%	782	51.055,193	120	51.055,950	-92,33%	0,00%
	3	1001	41.275,562	240	41.382,989	-88,01%	0,26%	995	41.417,284	240	41.353,997	-87,94%	-0,15%
	4	978	34.844,060	360	34.786,965	-81,60%	-0,16%	1014	34.867,447	360	35.106,728	-82,25%	0,69%
	5	1033	31.818,225	480	32.055,889	-76,77%	0,75%	1058	32.131,012	480	32.532,739	-77,32%	1,25%
	6	1025	28.651,071	600	28.718,235	-70,73%	0,23%	1045	28.546,087	600	28.610,768	-71,29%	0,23%
	7	1031	26.713,651	720	26.640,836	-65,08%	-0,27%	1064	26.698,198	720	26.932,143	-66,17%	0,88%
	8	1018	24.627,520	840	24.621,365	-58,74%	-0,02%	1055	24.614,982	840	24.658,280	-60,19%	0,18%
	9	1039	23.152,553	960	23.163,615	-53,80%	0,05%	1072	23.027,346	960	23.040,742	-55,22%	0,06%
	10	1038	22.339,648	1080	22.206,463	-47,98%	-0,60%	1078	22.256,751	1080	22.900,682	-49,91%	2,89%

Table 5.15 Example of the Runs for Three Criteria Set

Dataset	Number of Territories	Bound: +/- 5%						Bound: +/- 10%					
		Exhaustive		Heuristic		Dif %		Exhaustive		Heuristic		Dif %	
		Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)	Time (s)	Dist. Sum.(m)
100	2	5	7.297,457	120	7.297,457	2300,00%	0,00%	5	7.297,457	120	7.297,457	2300,00%	0,00%
200	2	23	14.456,871	120	15.562,356	421,74%	7,65%	26	14.361,610	120	14.844,418	361,54%	3,36%
	3	FAILED	FAILED	FAILED	FAILED			35	11.870,895	240	13.470,387	585,71%	13,47%
300	2	70	21.388,217	120	21.390,604	71,43%	0,01%	69	21.386,645	120	21.455,393	73,91%	0,32%
	3	90	17.431,313	240	17.609,997	166,67%	1,03%	91	17.388,591	240	17.468,223	163,74%	0,46%
	4	87	14.411,734	360	18.879,758	313,79%	31,00%	89	14.134,567	360	15.088,803	304,49%	6,75%
400	2	157	29.780,633	120	29.852,559	-23,57%	0,24%	161	29.780,633	120	29.822,178	-25,47%	0,14%
	3	206	24.475,202	240	24.947,800	16,50%	1,93%	208	24.462,041	240	24.948,623	15,38%	1,99%
	4	201	21.526,054	360	23.193,984	79,10%	7,75%	205	19.515,657	360	20.594,854	75,61%	5,53%
	5	217	17.925,928	480	21.112,084	121,20%	17,77%	221	17.933,853	480	20.929,307	117,19%	16,70%
	6	214	20.061,341	FAILED	FAILED			219	18.155,925	600	17.564,335	173,97%	-3,26%
500	2	309	35.673,460	120	35.683,633	-61,17%	0,03%	311	35.663,715	120	35.668,275	-61,41%	0,01%
	3	397	29.097,029	240	29.113,980	-39,55%	0,06%	402	28.969,643	240	29.169,275	-40,30%	0,69%
	4	388	24.511,733	360	26.668,693	-7,22%	8,80%	394	24.241,451	360	26.902,608	-8,63%	10,98%
	5	416	24.624,386	480	23.128,522	15,38%	-6,07%	420	26.316,278	480	24.919,300	14,29%	-5,31%
	6	410	20.367,739	600	20.541,933	46,34%	0,86%	419	20.566,749	600	20.875,084	43,20%	1,50%
	7	416	20.842,982	720	20.939,857	73,08%	0,46%	421	18.980,996	720	18.949,306	71,02%	-0,17%
600	2	529	43.524,514	120	43.524,514	-77,32%	0,00%	533	43.524,514	120	43.524,514	-77,49%	0,00%
	3	672	35.642,975	240	35.694,331	-64,29%	0,14%	686	35.553,837	240	35.539,693	-65,01%	-0,04%
	4	660	31.928,827	360	33.252,454	-45,45%	4,15%	675	29.860,596	360	32.870,623	-46,67%	10,08%
	5	715	28.630,752	480	30.112,976	-32,87%	5,18%	721	26.527,114	480	30.088,601	-33,43%	13,43%
	6	698	28.599,547	600	32.766,139	-14,04%	14,57%	725	24.798,003	600	26.434,537	-17,24%	6,60%
	7	696	26.262,065	720	24.071,311	3,45%	-8,34%	721	22.946,845	720	24.775,678	-0,14%	7,97%
	8	686	25.882,412	840	26.707,072	22,45%	3,19%	709	22.802,858	840	24.073,024	18,48%	5,57%
	9	701	23.612,680	FAILED	FAILED			727	21.094,389	960	26.339,440	32,05%	24,86%
700	2	812	51.156,056	120	51.975,821	-85,22%	1,60%	842	51.082,610	120	51.287,264	-85,75%	0,40%
	3	1057	41.751,217	240	41.529,791	-77,29%	-0,53%	1080	41.380,848	240	41.653,051	-77,78%	0,66%
	4	1023	35.256,480	360	37.131,446	-64,81%	5,32%	1063	34.703,967	360	35.450,523	-66,13%	2,15%
	5	1108	31.400,397	480	34.614,582	-56,68%	10,24%	1135	31.343,863	480	31.569,365	-57,71%	0,72%
	6	1085	29.130,336	600	32.458,323	-44,70%	11,42%	1125	28.689,044	600	29.661,645	-46,67%	3,39%
	7	1094	28.116,511	720	29.844,197	-34,19%	6,14%	1127	26.889,510	720	27.494,251	-36,11%	2,25%
	8	1082	26.042,015	840	27.205,189	-22,37%	4,47%	1113	24.523,490	840	26.073,699	-24,53%	6,32%
	9	1103	26.387,023	960	29.412,102	-12,96%	11,46%	1135	24.510,159	960	26.163,163	-15,42%	6,74%
	10	1100	23.806,836	FAILED	FAILED			1135	22.363,821	1080	24.644,866	-4,85%	10,20%

5.6 Summary

In this chapter we have presented the tests performed to evaluate the effectiveness of the algorithm proposed in this thesis in the context of some synthetic (laboratory) datasets.

We carried out three different sets of experiments: a Tuning Experiments set; an Exhaustive MultiACE vs. CPLEX Experiments set; and a Heuristic MultiACE vs. Exhaustive MultiACE Experiments set.

Computational experiments have been performed over these three different sets of test problems. The results obtained with the different approaches have been analyzed and commented.

We have shown that the solutions generated using the Voronoi based strategy and the Two-sets partitioning strategy are quite good and obtained in a rather shorter time, when compared to the optimal solutions obtained with CPLEX.

We also demonstrated that the algorithm can be very effective when we use the GRASP candidates list and the TS based memories (Heuristic MultiACE), since they produce a considerable improvement in the performance of the meta-heuristic, as compared to the one of Exhaustive MultiACE.

Chapter 6

Solving Real Problems

"The most beautiful thing we can experience is the mysterious. It is the source of all true art and science."

Albert Einstein (1879 - 1955) US (German-born) physicist

6.1 Introduction

The purpose of this chapter is to demonstrate the potential of the proposed approach as a practical and readily implementable management decision aid in the context of an current case involving the maintenance team of a Portuguese regional office of a worldwide equipments company (called "Company A" for confidentiality reasons). The regional managers at Company A were interested in learning how our approach could be used to evaluate and possibly improve their current territories structure. The following sections include a description of the procedures for data collection, parameters estimation employed and solution obtained using the GIS software tools and the algorithm proposed in this thesis.

6.2 Company A Case Background

The company sells new equipments and provides maintenance to new and old equipments. There is a specific division responsible for the maintenance tasks, and, in Portugal, the operation is divided in sub-areas. The area in which this study was accomplished is a sub-area in the North of Portugal and is shown in Figure 6.1. The area

under study is almost restricted to the district of Porto, with just a few equipments out of this district.

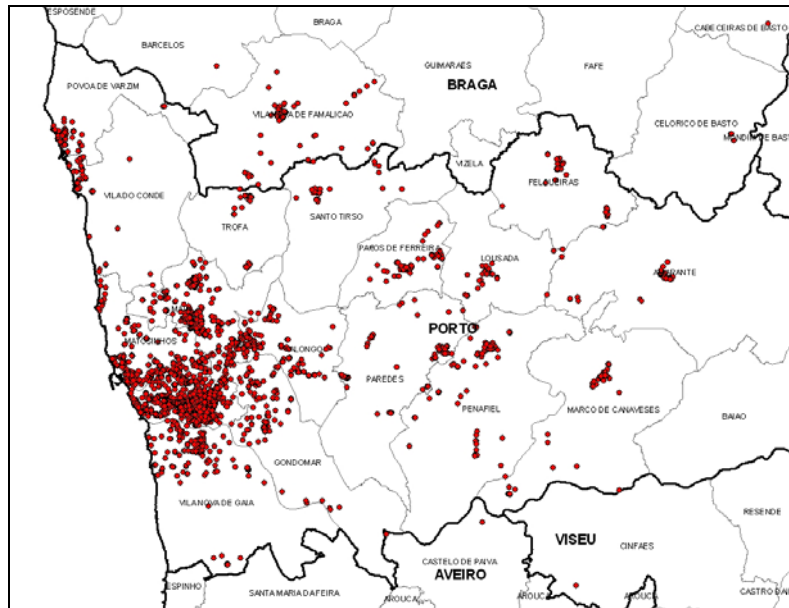


Figure 6.1 Company A – Equipments of the Area Under Study

Company A designs, manufactures, installs, maintains, and modernizes equipments for almost every type of building requirement worldwide. The company is dedicated to latest-technology engineering, and mechanical and microprocessor-technology products designed and rigorously tested for comfort, efficiency, and reliability.

In the area under analysis, the company maintains approximately six thousands equipments supported by a team of technical experts and solution providers. In the office, ten supervisors coordinate and lead various crews that maintain and repair customers' equipments.

There are two special supervisors: one that takes care of a special group of equipments and other one that takes care of a special type of client's equipments. The other eight supervisors take care of the remaining equipments, approximately five thousands and two hundreds.

Each technician has a list of equipments to provide maintenance. This list is what we define as the technician's territory. The supervisor is responsible for all equipments of the technicians of his group and he manages all the back-office work related to the equipments.

The technicians are supposed to visit each one of the equipments of his list once a month to perform both preventive maintenance (adjusting, examining and repairing or replacing worn components, lubricating and cleaning parts) and services for customer trouble calls (callbacks). Customer callbacks are usually assigned to the responsible maintenance technician, if available.

6.3 The Territories Assignment Process

We implemented the territories assignment process according to (Zoltner et al. 2004): a central alignment that acts as a benchmark, with local adjustments. The central benchmark alignment parameters (criteria) were defined by the Company A managers and were based on objective business criteria. Figure C.1 (Appendix C) shows the process steps followed to accomplish the study.

In Step 1, the managers determined the main objectives and the alignment criteria⁶ as:

- To distribute equipments equitably. First, define the supervisors territories and then subdivide these territories into technicians' territories. The number of supervisors territories should change to 7 territories, instead of 8, and the technicians' territories should be in the range of 108 to 132 equipments (120 equipments $\pm 10\%$);
- To distribute workload equitably and, since there is a calculated maintenance time⁷ for each equipment, the total sum of a territory's maintenance time should never be more than 136 maintenance hours per month;
- To merge the Company A territories and the territories of the small companies subcontracted by Company A;
- To decrease the number of regular supervisors from 8 to 7;
- To create contiguous territories;
- To create territories as compact as possible.

In Step 2, a database was developed. This database included all the equipments locations and alignment attributes (workload). The first step to develop this database was to locate all equipments using a GPS device and this data collection was accomplished with the help of the team of maintenance workers. We visited all the equipments of their territories and marked points to everyone⁸ of them using a GPS device. This process step took a little bit longer than planned because it was accomplished in July, August and September, and some of the technicians were in vacation. The current hierarchical organization information (supervisor, technician, region, districts, etc.) was inserted in the database too.

⁶ These criteria were defined after some brainstorming sessions and after some scenarios proposed using the MultiACE algorithm. Here, we just listed the final group of criteria.

⁷ Based on characteristics of the equipments such as: age, size, technology, etc.

⁸ Each equipment received latitude and longitude coordinates.

In Step 3, we proposed a new territories structure based on the predetermined objectives and constraints. This analysis was carried out using the MultiACE algorithm. We did not make a preliminary personnel assignment at this time. The shape and balance of the proposed territories were based on the combination of the criteria. The result of this step was an optimized territory alignment.

In Step 4, we tried to take into account the real-world factors and organizational uniqueness that shapes the maintenance workers environment. The management team audited and adjusted the territories. The audit group included the regional and district managers.

In Step 5, we developed optimal territory alignments. After this, a personnel assignment was proposed and adjusted, since the specific geography and account assignments for each territory were known.

In Step 6, the alignments and personnel assignments were audited and finalized with the help of the management team.

In the following sections we will detail all the steps and all the decisions we made to achieve the final result. The problems and special solutions are detailed as well.

6.4 The Current Territories Structure

After developing the database, we inserted all the information in a personal geodatabase⁹ and started to use the GIS environment. Our first task was to analyze the current situation of the territories. This evaluation is very important to define the regions or subgroups of equipments that need more attention or are less balanced.

The current supervisor areas are shown from Figure 6.2 to Figure 6.11 and just with these maps we can notice some problems and unbalanced situations. As the figures show, the supervisor territories have different shapes and geographical distributions. For instance, the territory of the supervisor 06 is the smallest while the territory of the supervisor 12 is the biggest. Table 6.1 shows, for each territory, its number of equipments, its workload¹⁰, its maximum distance between points (MDBP), and its mean sum of Euclidean distances to the centroid of the territory (MDTC).

⁹ ArcGIS file format.

¹⁰ Summation of the theoretical work mean-time of each equipment in the territory.



Figure 6.2 Supervisor Territory 01

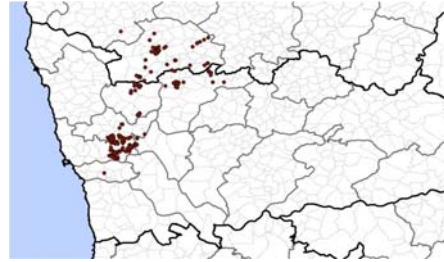


Figure 6.3 Supervisor Territory 02

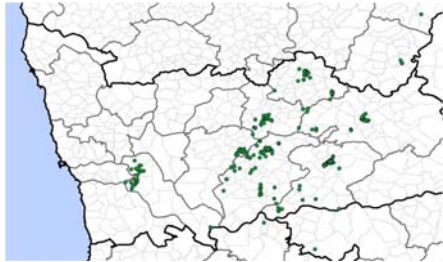


Figure 6.4 Supervisor Territory 03

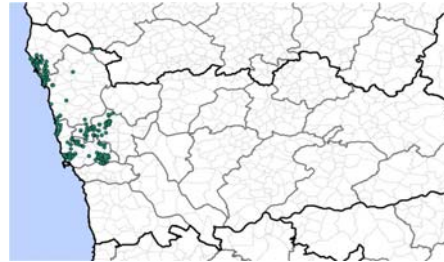


Figure 6.5 Supervisor Territory 04



Figure 6.6 Supervisor Territory 05



Figure 6.7 Supervisor Territory 06



Figure 6.8 Supervisor Territory 07

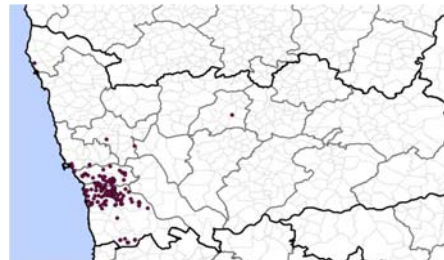


Figure 6.9 Supervisor Territory 09



Figure 6.10 Supervisor Territory 11

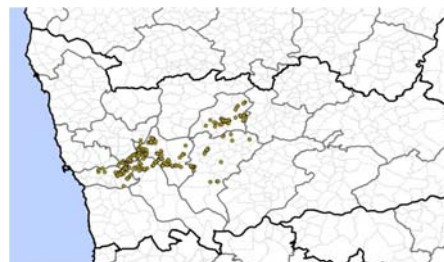


Figure 6.11 Supervisor Territory 12

Table 6.1 Current Supervisors Territories

Supervisor	Equipments		Workload (hours)		Compact Measurement (m)	
	Quantity	% of Mean	Quantity	% of Mean	MDBP	MDTC
01	581	96,87%	651,59	97,82%	26667,29	5493,97
02	639	106,54%	685,71	102,95%	35869,67	9251,38
03	612	102,03%	652,05	97,89%	64062,85	10865,15
04	736	122,71%	782,99	117,55%	27446	8365,32
05	712	118,71%	779,06	116,96%	7683,58	1985,25
06	545	90,86%	556,36	83,53%	3769,54	767,11
07	524	87,36%	588,47	88,35%	5279,69	1355,53
09	804	134,04%	943,66	141,67%	44128,78	4191,92
11	162	27,01%	290,87	43,67%	11219,11	1391,1
12	683	113,87%	730,12	109,61%	29949,24	7458,22
Minimum	162		290,87			
Maximum	804		943,66			
Mean	599,8		666,09			
Standard Deviation	167,99		163,46			
Total	5998		6660,89			

Table 6.2 and Figure 6.12 show the current supervisors' territories balance situation for the area under study.

Table 6.2 Supervisors Territories - Current Balance

Supervisor	Equipments		Workload (hours)	
	Quantity	% of Mean	Quantity	% of Mean
01	578	101,05%	645,67	105,23%
02	633	110,66%	673,09	109,69%
03	375	65,56%	403,55	65,77%
04	736	128,67%	782,99	127,60%
05	691	120,80%	747,39	121,80%
06	545	95,28%	541,92	88,32%
07	507	88,64%	562,79	91,72%
09	400	69,93%	436,1	71,07%
12	683	119,41%	729,01	118,81%
Minimum	375		403,55	
Maximum	736		782,99	
Mean	572		613,61	
Standard Deviation	120,40		128,16	
Total	5148		5522,51	

For the final analysis, we excluded, from the 5998 equipments set, the supervisor 11 territory (special client equipments), some equipment transferred to other regional office and some equipments integrated to the special equipments supervisor. In Table 6.2 we can see the distance between the minimum and maximum number of equipments in the territories and we can notice the unbalanced current state.

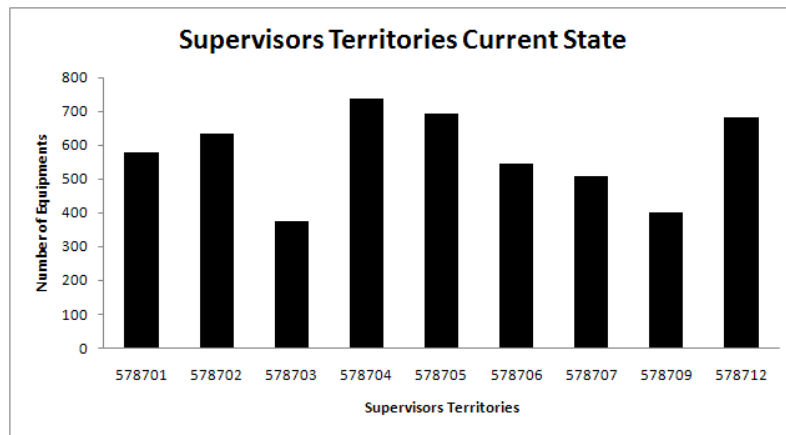


Figure 6.12 Supervisors Territories – Current Balance

Besides the different sizes and shapes, the supervisors' territories have a contiguity problem. There are overlaps in some areas, as shown in Figure 6.13. The same problem occurs to a lot of territories of the maintenance workers' and it is an important factor that decreases the productivity of the teams and increases the company costs.

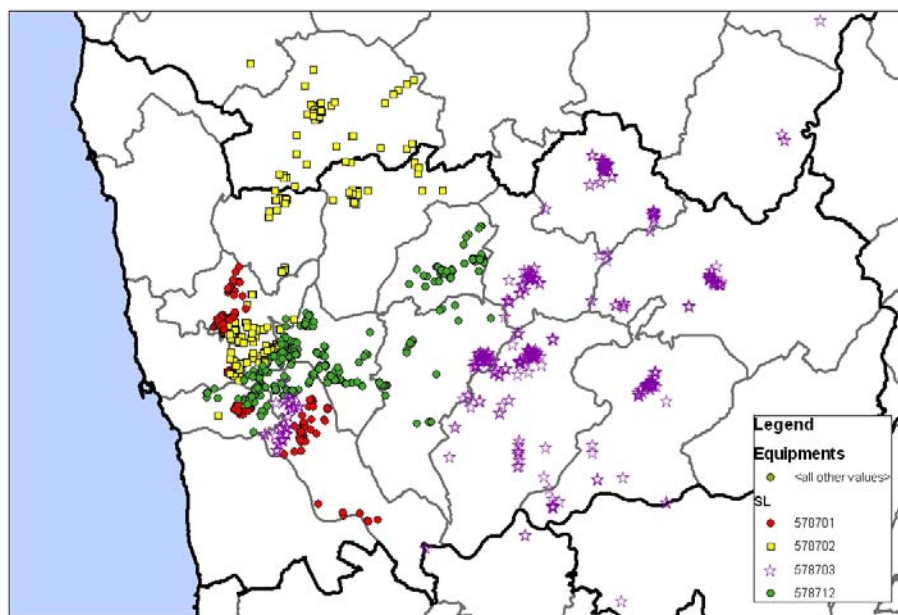


Figure 6.13 Overlap of the Territories of Supervisors 01, 02, 03 and 12

Table 6.3 shows the current maintenance teams territories balance situation for the study area. One can notice in Table 6.3 the distance between the minimum and maximum number of equipments in the workers' territories and we can notice the unbalanced current state. We can also notice the subcontracted small companies' territories: 0301, 0504, 0907 and 1201 (in bold).

Table 6.3 Technicians Territories – Current Balance – Study Area

Supervisor	Equipments		Workload (hours)	
	Quantity	% of Mean	Quantity	% of Mean
0101	126	102,80%	141,93	107,94%
0102	102	83,22%	109,36	83,17%
0103	124	101,17%	146,16	111,16%
0104	104	84,85%	121,13	92,12%
0105	122	99,53%	127,09	96,66%
0201	125	101,98%	134,37	102,19%
0202	136	110,96%	140,35	106,74%
0203	121	98,72%	128,3	97,58%
0204	124	101,17%	134,65	102,40%
0205	127	103,61%	135,42	102,99%
0301	55	44,87%	55,88	42,50%
0302	127	103,61%	140,83	107,10%
0303	57	46,50%	59,12	44,96%
0304	136	110,96%	147,73	112,35%
0401	125	101,98%	132,07	100,44%
0402	131	106,88%	139,43	106,04%
0403	112	91,38%	120,98	92,01%
0404	130	106,06%	138,2	105,11%
0405	122	99,53%	131,52	100,03%
0406	116	94,64%	120,79	91,87%
0501	132	107,69%	155,12	117,97%
0502	132	107,69%	139,58	106,15%
0503	109	88,93%	110,23	83,84%
0504	318	259,44%	342,46	260,45%
0601	106	86,48%	99,09	75,36%
0602	100	81,59%	100,26	76,25%
0603	113	92,19%	110,73	84,21%
0604	108	88,11%	112,89	85,86%
0605	118	96,27%	118,96	90,47%
0701	82	66,90%	87,19	66,31%
0702	90	73,43%	108,64	82,62%
0703	98	79,95%	108,66	82,64%
0704	116	94,64%	131,64	100,11%
0705	121	98,72%	126,66	96,33%
<i>0902</i>	<i>7</i>	<i>5,71%</i>	<i>7,44</i>	<i>5,66%</i>
<i>0903</i>	<i>3</i>	<i>2,45%</i>	<i>2,86</i>	<i>2,17%</i>
<i>0904</i>	<i>8</i>	<i>6,53%</i>	<i>9,7</i>	<i>7,38%</i>
<i>0905</i>	<i>14</i>	<i>11,42%</i>	<i>17,49</i>	<i>13,30%</i>
0907	368	300,23%	398,61	303,15%
1201	433	353,26%	451,64	343,48%
1202	134	109,32%	151,6	115,29%
1203	116	94,64%	125,77	95,65%
Minimum	3		2,86	
Maximum	433		451,64	
Mean	122,58		131,49	
Standard Deviation	79,08		84,27	
Total	5148		5522,51	

The territories 0902, 0903, 0904 and 0905 (in italic) lost almost all the equipments and we should use these workers to create new territories.

After the analysis of the current supervisors and maintenance workers territories, we noticed that the area under analysis needed to be subdivided before we start to generate the

new territories structure. The reason for this procedure is that the study area is very large and has a heterogeneous distribution of the equipments. So, the managers decided to divide the equipments into 3 areas (see Figure 6.14): the first one, a very compact and central area, the second one, bigger than the first area and around it, and the third one, a very big and dispersed area in the outskirts.

To divide these areas we used only the ArcGIS tools (selection and database tools), according to the managers will. We carried out this partition trying to use the geographical characteristics of the areas, either visually or using the database tools.

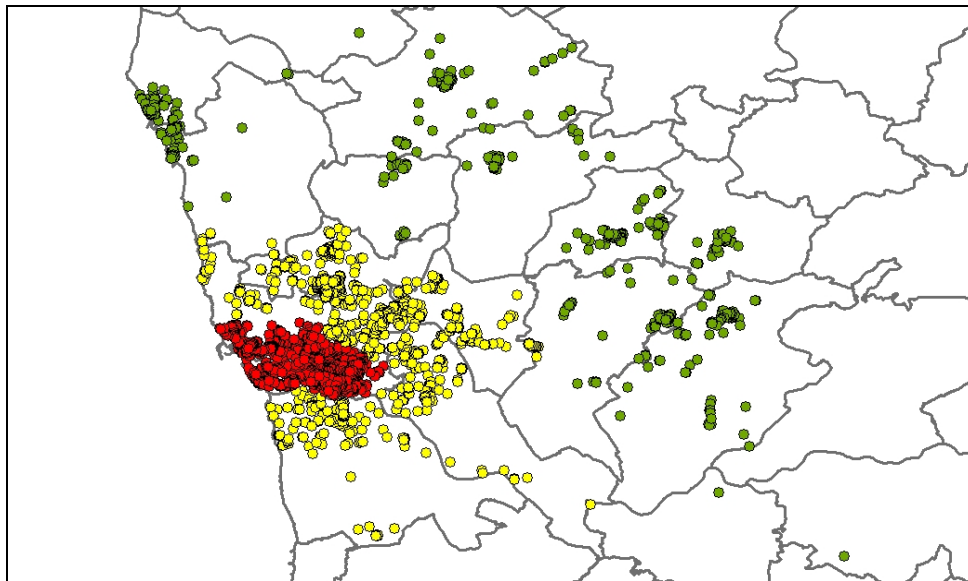


Figure 6.14 The 3 Different Areas

To confirm the differences between these areas we used the Network Analyst to create minimum distances routes passing in every point of the areas (see Figure 6.15, Figure 6.16 and Figure 6.17) and we used the total distance of the routes to calculate the mean distance between equipments (MDBE) and the mean distance between stops (MDBS), shown in Table 6.4.

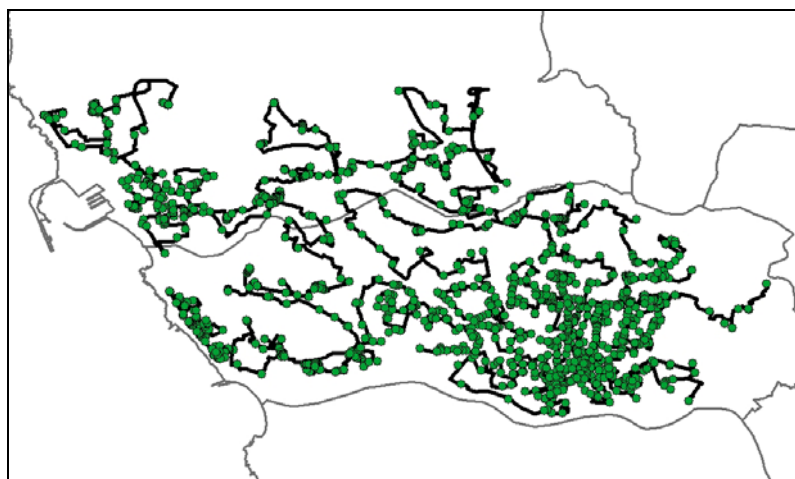


Figure 6.15 Route in Area 1

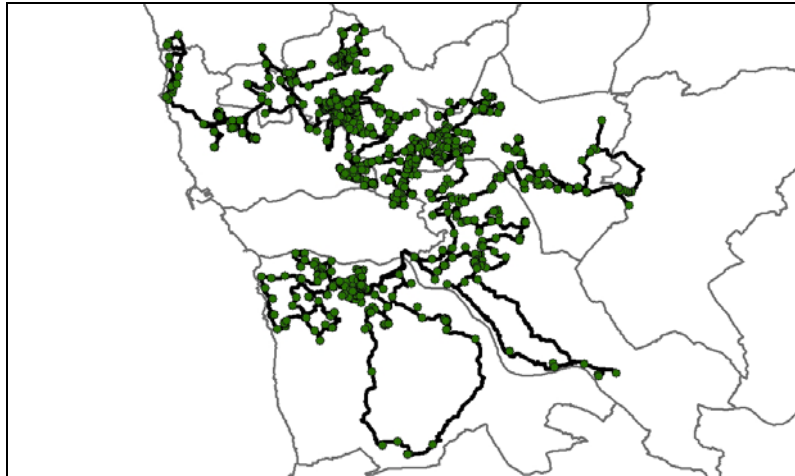


Figure 6.16 Route in Area 2

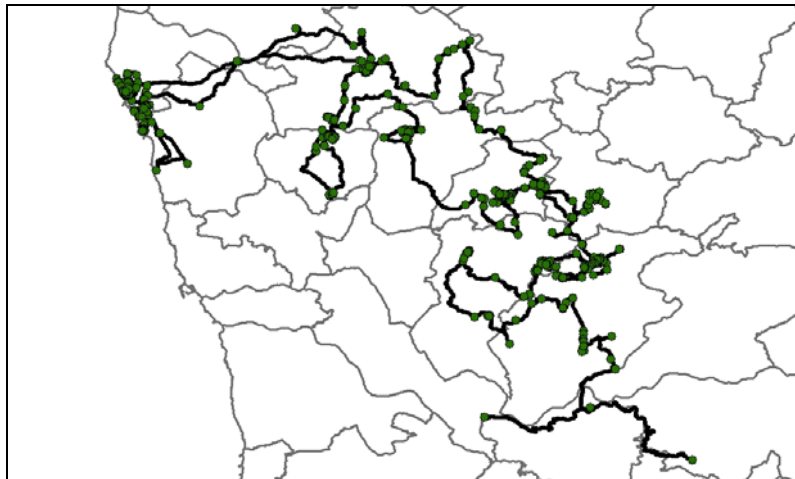


Figure 6.17 Route in Area 3

Table 6.4 Three Areas – A Comparative Analysis

Area	Equipments	Stops	Total Distance (m)	MDBE (m)	MDBS (m)
1	2160	1248	211919,4	98,11	169,80
2	2007	1015	474507,8	236,42	467,49
3	981	559	455099,4	463,91	814,13

According to the comparative analysis of the 3 areas, the managers decided to change the bounds of the number of equipments for the areas 2 and 3. They did not set new values for the bounds but decided that areas 2 and 3 could have territories 5-10% smaller than the territories in area 1. The reason for this changing is that the mean distances between equipments in these areas are very large, increasing the travel time between equipments.

6.5 New Territories Project

After the evaluation of the current situation, the next step was to project the new territories for the supervisors. We divided the 5148 equipments¹¹ into 7 territories using some different strategies to create them, but, to run the MultiACE algorithm we used the following options:

- Number of supervisors territories: 3;
- First criterion: Number of equipments;
- First criterion bounds: +10% and -10%;
- First criterion weight: 1;
- Second criterion: Workload;
- Second criterion bounds: +10% and -10%;
- Second criterion weight: 1;
- Processing time limit: 10 minutes;
- Ideal point distance measurement: Euclidean distances;
- Physical distance measurement: Network distances;
- Streets type: two-way distances;
- Short term memory: 30% of the number of points;
- Long term memory: 200000 solutions.

The first supervisor territory was obtained manually using the ArcGIS selection tools and the MultiACE to generate a summary of the criteria values of this territory. We used this approach to create the first territory because it covers a very sparse area and the algorithm aims to solve problems in homogeneous areas. After this, we selected part of the Area 3 to create the supervisor territory 01.

The remaining points of Area 3 points and all the points of Area 2 were used to generate 3 territories and, finally, Area 1 was used to generate the last 3 territories, with the help of the MultiACE algorithm.

After the generation of the solutions, we analyzed them and did some small manual adjustments. The proposed solutions were very good and we just changed some points from one territory to another to adjust the shape of the territories. The last adjustment in the

¹¹ The equipments were aggregated into 2819 different points, and we used theses points to create the territories.

supervisors' territories was carried out only after we calculated the Technicians territories, but, at this time, we did just the whole technician territories changes, moving from one Supervisor territory to another. The final supervisors' territories are shown in Table 6.5.

Table 6.5 Supervisors Territories – Current Balance

Supervisor	Equipments		Workload (hours)		Compactness Measurement (m)	
	Quantity	% of Mean	Quantity	% of Mean	MDBP	MDTC
01 ¹²	762	103,61%	807,32	102,33%	60354,13	10765,61
02	794	107,96%	869,02	110,15%	24870,64	7910,59
03	752	102,25%	817,01	103,56%	23291,99	5681,51
04	798	108,51%	851,98	107,99%	11704,06	3226,56
05	697	94,77%	762,16	96,61%	7015,9	1947,96
06	723	98,31%	785,58	99,58%	6128,06	1539,69
07	622	84,58%	629,43	79,78%	5730,88	951,03
Minimum	622		629,43			
Maximum	798		869,02			
Mean	735,43		788,93			
Standard Deviation	57,10		73,37			
Total	5148		5522,51			

The next step was to subdivide each one of the supervisors' territories into the new maintenance workers' territories. At this time, we used the MultiACE algorithm to create all the territories. For each one of the Supervisors' territories we applied the MultiACE algorithm and, since we did not have very large sets of points, we used the exhaustive method to create better solutions. The MultiACE parameters used are:

- Number of territories: 5 or 6;
- First criterion: Number of equipments;
- First criterion bounds: +10% and -10%;
- First criterion weight: 1;
- Second criterion: Workload;
- Second criterion bounds: +10% and -10%;
- Second criterion weight: 1;
- Exhaustive method;
- Ideal point distance measurement: Euclidean distances;
- Physical distance measurement: Network distances;
- Streets type: two-way distances.

The final technicians territories is shown in Table 6.6, that also shows, for each territory, its number of equipments, its workload, its MDBP, its MDTC and its Maximum Radius from Territory Centroid (MRTC). The territory with the minimum number of

¹² This area was obtained manually.

equipments and workload is out of the bound determined by the managers because it is located in the most dispersed area (see its compactness measurements). Figure 6.18 and Figure 6.19 show the supervisors territories and, for each one of them, its technicians territories.

Table 6.6 New Territories Structure – Maintenance Workers

Technician	Equipments		Workload (hours)		Compact Measurement (m)		
	Quantity	% of Mean	Quantity	% of Mean	MDBP	MDTC	MRTC
0101	104	90,91%	114,81	93,55%	10219,57	2544,76	7849,6
0102	109	95,28%	116,3	94,76%	7902,42	2743,36	4892,22
0103	89	77,80%	92,54	75,41%	24399,07	4349,11	20970,43
0104	108	94,41%	114,7	93,46%	8790,52	2091,87	6991,3
0105	119	104,02%	117,73	95,93%	17552,45	6427,68	10454,12
0106	117	102,27%	125,43	102,21%	11016,57	3490,84	5986,19
0107	116	101,40%	125,82	102,52%	15135,96	1874,98	8638,7
0201	108	94,41%	116,51	94,94%	3536,88	564,85	2008,93
0202	120	104,90%	128,32	104,56%	8177,33	2009,47	5532,8
0203	105	91,78%	110,67	90,18%	3144,76	643,63	1661,57
0204	118	103,15%	127,06	103,53%	4515,56	1125,45	2522,92
0205	114	99,65%	136,46	111,19%	1621,88	385,2	920,67
0206	116	101,40%	129,3	105,36%	4313,65	1021,64	3071,89
0207	113	98,78%	120,7	98,35%	14376,46	1951,14	9771,11
0301	99	86,54%	104,23	84,93%	6228,19	1987,39	4403,48
0302	116	101,40%	125,54	102,30%	5462,32	1765,84	3101,91
0303	108	94,41%	115,88	94,42%	11960,67	3389,45	8531,43
0304	110	96,15%	122,49	99,81%	3143,68	617,72	1942,48
0305	101	88,29%	113,46	92,45%	4063,09	1181,83	2292,3
0306	116	101,40%	128,19	104,46%	4211,32	1272,75	2856,5
0307	102	89,16%	107,22	87,37%	18199,87	3331,65	14286,3
0401	118	103,15%	120,71	98,36%	2948,57	635,69	1765,94
0402	110	96,15%	109,57	89,28%	6540,05	2106,91	3609,37
0403	117	102,27%	125,74	102,46%	2967,22	772,99	1823,83
0404	111	97,03%	115,91	94,45%	4299,33	1410,58	2780,2
0405	114	99,65%	124,45	101,41%	3583,46	1024,15	2299,7
0406	120	104,90%	134,74	109,79%	4022,1	1143,64	2024,33
0407	108	94,41%	120,86	98,49%	2429,9	499,39	1631,53
0501	121	105,77%	127,12	103,58%	2755,04	1046,33	1683,41
0502	110	96,15%	129,97	105,91%	1504,55	399,3	845,86
0503	119	104,02%	129,2	105,28%	2278,04	662,59	1355,31
0504	114	99,65%	126,17	102,81%	2637,41	764,04	1776,66
0505	120	104,90%	125,09	101,93%	2259,94	643,73	1521,71
0506	113	98,78%	124,61	101,54%	2766,64	944,56	1552,41
0601	118	103,15%	137,18	111,78%	2107,93	633,08	1170,32
0602	117	102,27%	132,27	107,78%	2807,9	760,63	1980
0603	115	100,52%	120,88	98,50%	2955,79	562,8	1488,59
0604	122	106,64%	136,24	111,01%	1677,83	426,2	1055,98
0605	121	105,77%	126,47	103,05%	1502,11	447,23	940,12
0606	130	113,64%	132,54	108,00%	1360,1	399,97	736,31
0701	126	110,14%	133,86	109,07%	1982,33	502,43	1139,13
0702	116	101,40%	120,71	98,36%	1435,69	401,55	944,01
0703	130	113,64%	131,94	107,51%	3321,81	652,68	2537,08
0704	123	107,52%	122,01	99,42%	1345,12	332,91	841,27
0705	127	111,01%	120,92	98,53%	1320,79	316,95	746,7
Minimum	89		92,54				
Maximum	130		137,18				
Mean	114,4		122,72				
Standard Deviation	8,00		9,03				
Total	5148		5522,51				

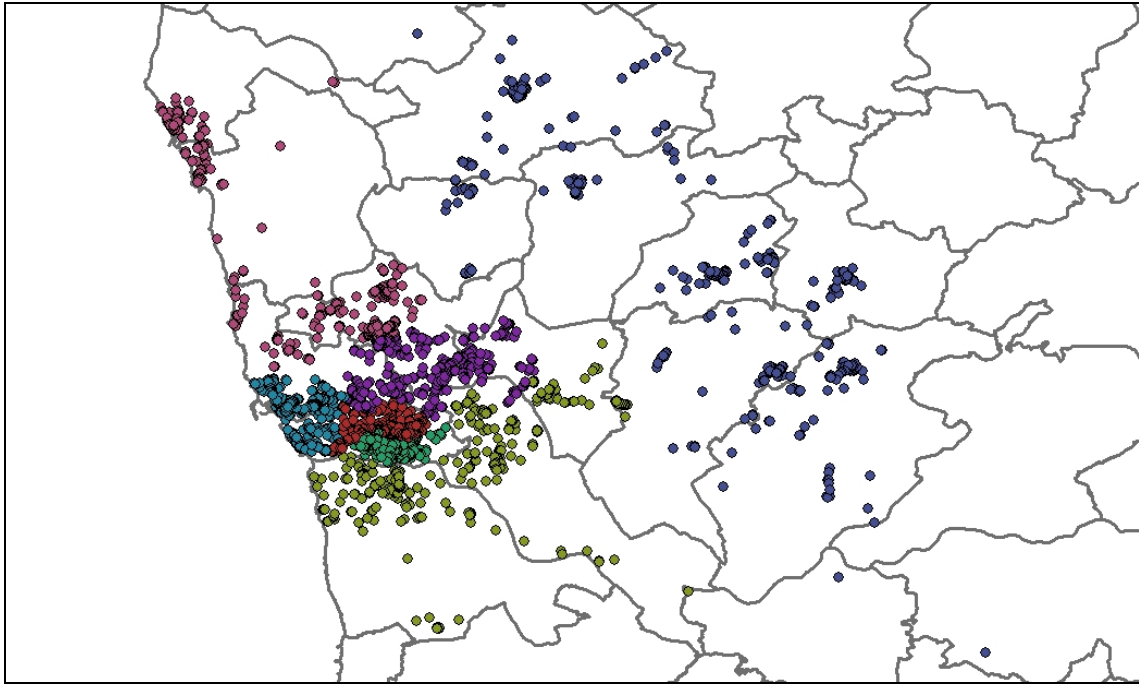


Figure 6.18 New Supervisors Territories Structure

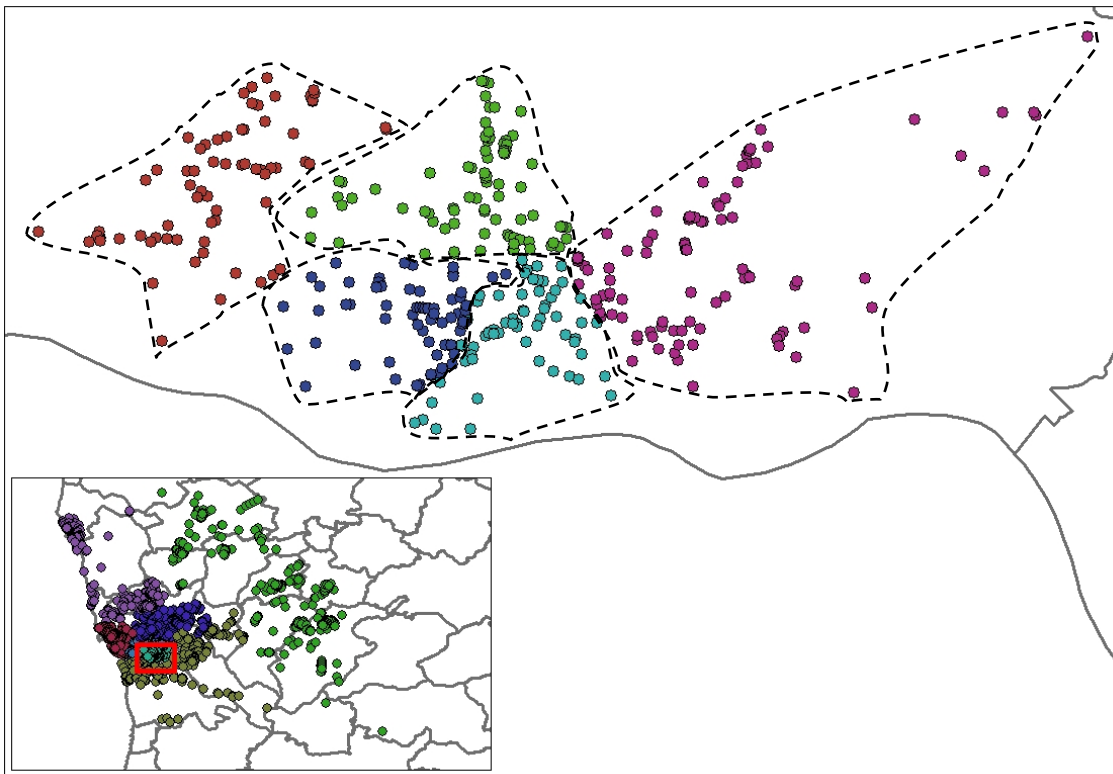


Figure 6.19 Supervisor Territory 07 - Maintenance Workers' Territories

6.5.1 Achievements and Savings

The new territories structure shown in Table 6.6 was implemented and the main achievements and savings obtained with it are:

- The managers were able to quickly get an overview of where all the territories were and where there was some overlap in the structure. There are no more territories overlaps and the territories are now better geographically defined. It is easy to decide what maintenance worker will take care of new equipments. They have been able to better define areas of responsibility and loading;
- The maintenance workers' territories are more compact, so the total travel time decreased, improving the productivity of the maintenance team;
- According to the MRTC compactness measurement, the managers decided to decrease the number of cars available to the maintenance team. They decided that the territories with MRTC smaller than 1700 meters (bold territories in Table 6.6) would not have a car. The result was a decreasing of eight cars from the maintenance fleet, with direct and indirect costs reduction of approximately eighty thousands Euros per year.

6.6 Improving the Solution

The solution represented by the new structure is one of the six scenarios studied during the whole study. The managers could explain their ideas, the constraints we should respect and their subjective objectives for the new structure.

Some of these constraints and subjective goals influenced the final result and we believed that we could improve this solution. So, we decided to create a different structure to compare to the selected one and show some improvements that could be made.

The changes we decided to make are:

- **Number of supervisors:** instead of seven supervisors, we decided to use eight;
- **Number of maintenance workers:** instead of forty five, we decided to use forty six;

One can think that these changes would increase the direct and indirect costs, but we intended to show that the selection of a structure (scenario) in a territories project is a trade-off among the several objectives the managers have to achieve.

To create the territories for the supervisors, we used the three Areas defined previously, but this time we divided the Area 3 into two territories, Area 2 into three territories and Area 1 into three territories.

The next step was to divide the supervisors' territories into maintenance workers' territories. We used the same criteria and MultiACE parameters defined previously, but this time we changed the number of territories according to Table 6.7. The improved new territories structure is shown in Table 6.8.

Table 6.7 Improving the Solution – Number of Territories

Supervisor	Maintenance Territories
11	6
12	6
13	6
21	6
22	6
23	6
31	5
32	5
Total	46

The main remark we can make from the data shown in Table 6.8 is that the number of territories with MRTC smaller than 1700 meters increased from 15 to 20, decreasing thirteen cars from the maintenance fleet, with direct and indirect costs reduction of approximately ninety six thousands Euros per year, including the costs of one more supervisor, one more maintenance worker and one more vehicle for the supervisor. This kind of trade-off between the size of the maintenance team and costs reduction can be made easily and many scenarios can be analyzed in a very short time.

Table 6.8 Improved New Territories Structure – Maintenance Workers

Territory	Equipments		Workload (hours)		MRTC
	Quantity	% of Mean	Quantity	% of Mean	
111	114	99,65%	126,17	110,29%	1686,66
112	121	105,77%	128,04	111,93%	1683,41
113	113	98,78%	124,61	108,93%	1552,41
114	120	104,90%	125,09	109,34%	1521,71
115	119	104,02%	129,20	112,94%	1355,31
116	110	96,15%	129,97	113,61%	845,86
121	117	102,27%	132,27	115,62%	1980,00
122	118	103,15%	120,71	105,51%	1765,94
123	115	100,52%	120,88	105,66%	1488,59
124	118	103,15%	137,18	119,92%	1170,32
125	126	110,14%	134,78	117,82%	1139,13
126	122	106,64%	136,24	119,09%	1055,98
131	130	113,64%	132,87	116,14%	2537,08
132	118	103,15%	123,72	108,15%	950,21
133	121	105,77%	126,47	110,55%	940,12
134	123	107,52%	122,93	107,46%	841,27
135	127	111,01%	122,77	107,32%	746,70
136	128	111,89%	131,39	114,85%	740,78
211	120	104,90%	128,32	112,17%	5532,80
212	116	101,40%	129,30	113,03%	3071,89
213	118	103,15%	127,06	111,07%	2522,92
214	108	94,41%	116,51	101,84%	2008,93
215	108	94,41%	120,86	105,65%	1631,53
216	114	99,65%	136,46	119,28%	920,67
221	99	86,54%	104,23	91,11%	4403,48
222	110	96,15%	109,57	95,78%	3609,37
223	111	97,03%	115,91	101,32%	2780,20
224	114	99,65%	124,45	108,79%	2299,70
225	120	104,90%	134,74	117,78%	2024,33
226	117	102,27%	126,66	110,72%	1673,83
231	101	88,29%	106,24	92,87%	9427,35
232	108	94,41%	115,88	101,29%	8531,43
233	116	101,40%	125,54	109,74%	3101,91
234	116	101,40%	130,05	113,68%	2856,50
235	101	88,29%	113,46	99,18%	2292,30
236	110	96,15%	122,49	107,07%	1942,48
311	102	89,16%	113,99	99,64%	10719,57
312	88	76,92%	84,22	73,62%	10638,55
313	110	96,15%	117,38	102,61%	6240,20
314	86	75,17%	93,87	82,06%	4217,79
315	104	90,91%	109,53	95,74%	1665,10
321	88	76,92%	91,84	80,28%	22203,75
322	100	87,41%	102,51	89,61%	11274,06
323	108	94,41%	119,20	104,19%	4670,80
324	93	81,29%	98,58	86,17%	4331,88
325	102	89,16%	110,41	96,51%	1304,93
Minimum	86		84,22		
Maximum	130		137,18		
Mean	111,91		120,31		
Standard Deviation	10,40		12,26		
Total	5148		5534,56		

6.7 Summary

In this chapter we have described the application of GIS tools and the algorithm proposed in this thesis in a real case study. The Company A needed to restructure its supervisors and maintenance workers' territories, while merging territories of other subcontracted companies and splitting part of the old territories to other regional office.

We have presented the utilization of GIS tools to support and help the decisions makers to perform some geographic based analysis and we also applied the MultiACE algorithm to propose solutions for the territories definition and alignment problem.

Chapter 7

Conclusions and Future Work

"The outcome of any serious research can only be to make two questions grow where only one grew before."

Thorstein Veblen (1857 - 1929) US economist & social philosopher

7.1 Introduction

This thesis has addressed the territories alignment problem as a critical component of the operational planning process in sales/services companies. A particular emphasis was given to the case of a Portuguese company and the tool developed and used, has been described in more detail.

We have proposed a new perspective on approaching TAP that is closer to the real problems and to the planners work and we have identified and discussed several different objectives and constraints used by companies in their alignment process. In order to tackle these simultaneous and conflicting objectives, a multi-objective approach based on GRASP and TS has been developed.

This new hybrid meta-heuristic is based on a combination of a Voronoi Diagram based strategy, to create solutions; a two-sets partitioning strategy, to decrease the search space; a GRASP based candidates list; Tabu Search based short and long term memories; and a multi-criteria technique to define the best solutions. We presented and discussed these components and a step-by-step description of the meta-heuristic solution procedure, its basic structure, initialization, subset generation and combination, dominance comparison, quality measurement and adaptive memory structure used.

The algorithm was tested using a series of carefully prepared tests on synthetic data generated with random procedures and the various experiments performed compared the

effect of algorithm parameters variation and datasets variation. We also compared the algorithm's results to the ones obtained using the CPLEX solver.

Finally, we also demonstrated the potential of the proposed approach as a practical and readily implementable management decision aid in the context of a current case that involved the team of maintenance workers attached to the Portuguese regional office of a worldwide equipments company.

7.2 Main achievements of this thesis

Territories alignment is an important process in the management of sales/services, usually involving the work of several planners. As a result of this process, important savings can be made with no deterioration of the overall quality and cost. Any advances in methods that support the decision making process of the planners can therefore have a significant impact. Thus, in this context we see the major contributions of our work as follows:

- The research has provided a new and valuable insight into the complex nature of the territories alignment problem. The use of a combination of: a Voronoi Diagram based strategy; a Two-sets partitioning strategy; a GRASP based candidates list; a Tabu Search based short and long term memories; and a multi-criteria technique, obtained very good results, when we analyzed and compared them to CPLEX results;
- A novel approach to the TAP has been developed and tested. The MultiACE algorithm has been tested on synthetic problems and has produced satisfactory results. The tests performed on real problems have shown that the algorithm can be used as a tool for the planning process. We may conclude that the application represents a high valuable contribution to support the territories alignment process;
- The MultiACE algorithm was integrated in an advanced interactive tool which can be used in different contexts. At the operational level, it represents a valuable tool to quickly produce alternative solutions. At the tactical level it can be used to simulate different operating scenarios and evaluate the impact of changes in territories. Furthermore, it can be also applied by planners to

support the testing of new prerequisites and new criteria can be easily incorporated;

- The MultiACE algorithm is a user-friendly tool that can be used by planners and decision makers without any special skill in programming or optimization. The person only needs to receive some basic instructions on the way the algorithm works and on the effect of some parameters on the final results, on top of having basic knowledge on GIS systems, familiarity with the specific GIS software and careful data acquisition;
- Our general multi-objective approach can be extended to other territories problems involving several different objectives. In fact, territories partitioning problems are natural extensions of our problem, but we believe that the approach could be also applied to other partitioning problems, such as data clustering. Naturally, different components should be devised and a different set of criteria should be found, but in principle the general structure of the approach could be easily adapted.

7.3 Future research

As it is natural, a research work such as the one described in this thesis is never fully closed. Several issues do clearly deserve further research. In the following points we briefly present some of these issues:

- The heuristic could have been tested handling several other activity measures. In this case while examining a partition, one could take different activity measures into account when determining the best territories structure;
- The algorithm could have been prepared to consider prescribed and forbidden territory centers. This means that we could have imposed some fixed territory centers at the beginning, which had to be taken into account, or, in the other way around, some basic areas that could not be allowed to be selected as district centers;
- In the case in which some districts are already given at the beginning of the planning process the algorithm could have been adapted to take the already existing territories into account and possibly add additional clients to them;

- The heuristic could have been incorporated into a larger framework in order to apply it to different practical planning problems. For example scenarios where a limit on the maximal allowed geographic extent of the territories has to be taken into account;
- We could have avoided the number of territories to be fixed in advance. Instead, the algorithm would choose the appropriate number of districts in such a way that the planning criteria would be best fulfilled. For example, the partition of the basic areas in the region under consideration into as few as possible territories such that the size of all territories should be below a certain maximal bound;
- The heuristic could have been incorporated into an open source framework in order to allow it to be used and tested by several planners, decision makers and developers around the world. This could be a good way to spread the utility of the GIS in the field of territories planning. This action could also encourage other developers to solve different kinds of alignment problems, with different kinds of criteria and in different situations. This open source implementation could be the cornerstone of a promising field of applied research.

References

- Afonso, P. H., Gomes, M. R., and Abrantes, M. G. "A Geomarketing Decision Support System Based on Fuzzy Set Theory." *3rd International Conference on GeoComputation*, University of Bristol, United Kingdom.
- Ahmadi, S., and Osman, I. H. (2005). "Greedy random adaptive memory programming search for the capacitated clustering problem." *European Journal of Operational Research*, 162(1), 30-44.
- Altman, M. (1998). "Districting Principles and Democratic Representation," California Institute of Technology, Pasadena, California.
- Analytics, M. (2006). "Three Steps to More Productive Sales Territories - A Mapping Analytics Executive Brief."
- Bacao, F., Lobo, V., and Painho, M. (2005a). "Applying genetic algorithms to zone design." *Soft Comput.*, 9(5), 341-348.
- Bacao, F., Lobo, V., and Painho, M. (2005b). "The self-organizing map, the Geo-SOM, and relevant variants for geosciences." *Computers & Geosciences*, 31(2), 155-163.
- Barker, A. T. (2001). "Salespeople characteristics, sales managers' activities and territory design as antecedents of sales organization performance." *Marketing Intelligence & Planning*, 19(1), 21 - 28.
- Battiti, R., and Tecchiolli, G. (1994). "The Reactive Tabu Search." *ORSA Journal on Computing*, 6(2), 126 - 140.
- Batun, S. "A Tabu Search Algorithm for Sales Territory Alignment Problem." *35th International Conference on Computers and Industrial Engineering*, Istanbul, Turkey, 251 - 256.
- Bergey, P., Ragsdale, C., and Hoskote, M. (2003). "A Simulated Annealing Genetic Algorithm for the Electrical Power Districting Problem." *Annals of Operations Research*, 121(1-4), 33-55.
- Blum, C., and Roli, A. (2003). "Metaheuristics in combinatorial optimization: Overview and conceptual comparison." *ACM computing surveys*, 35(3), 268-308.
- Bong, C. W., Chai, W. Y., and Wong, C. W. "State-of-the-art multiobjective metaheuristic for redistricting." 763-769 Vol.2.
- Bozkaya, B., Zhang, J., and Erkut, E. (1997). "An Effective Genetic Algorithm for The P-Median Problem." 97-2, University of Alberta - Faculty of Business - Department of Finance and Management Science, Edmonton, Alberta.
- Cano, J. R., Cordon, O., Herrera, F., and Sanchez, L. (2000). "A Greedy Randomized Adaptive Search Procedure to the Clustering Problem." *DECSAI-000113*, ETS de Ingenieria Informatica. University of Granada., Granada, Spain.
- Charnes, A., and Cooper, W. (1961). *Management models and industrial applications of linear programming*, Wiley, New York.
- Church, R. L. (2002). "Geographical information systems and location science." *Computers & Operations Research*(29), 541 - 562.

- Claro, J., and Sousa, J. P. "An Object-Oriented Framework for Multiobjective Local Search." *4th Metaheuristics International Conference*, Porto, Portugal.
- Coello, C. A. C. (2006). "References on Evolutionary Multiobjective Optimization." <http://delta.cs.cinvestav.mx/~ccoello/EMOO/EMOObib.html>.
- Corrêa, E. S., Steiner, M. T. A., Freitas, A. A., and Carnieri, C. (2002). "A Genetic Algorithm for the P-Median Problem."
- Crescenzi, P., and Kann, V. (1998). "A compendium of NP optimization problems." *Tech. Rep. SI/RR-95/02*, Dipartimento di Scienze della Informazione, Università di Roma La Sapienza.
- Delmaire, H., Diaz, J. A., Fernandez, E., and Ortega, M. (1999). "Reactive GRASP and tabu search based heuristics for the single source capacitated plant location problem." *INFOR*, 37, 194-225.
- Drexler, A., and Haase, K. (1999). "Fast Approximation Methods for Sales Force Deployment." *Management science*, 45(10), 1307 - 1323.
- Ehrgott, M., and Gandibleux, X. (2000a). "An Annotated Bibliography of Multi-objective Combinatorial Optimization." 62/2000, Fachbereich Mathematik - Universität Kaiserslautern, Kaiserslautern, Germany.
- Ehrgott, M., and Gandibleux, X. (2000b). "A survey and annotated bibliography of multiobjective combinatorial optimization." *OR Spectrum*, 22(4), 425-460.
- Ehrgott, M., and Gandibleux, X. (2004). "Approximative solution methods for multiobjective combinatorial optimization." *TOP (Spanish journal of operations research)*, 12(1), 1 - 89.
- ESRI. (2007). *What is GIS?*, <http://www.gis.com/whatisgis/index.html>.
- Festa, P., and Resende, M. G. C. (2004). "An Annotated Bibliography of GRASP." AT&T Labs Research Technical Report.
- Forster, M. (2000). "Review of the use of Geographical Information Systems in the Marketing and Planning of Logistics Services." School of Management and Languages at Heriot-Watt University - The Logistics Research Centre (LRC).
- Fortune, S. (1992). "Voronoi Diagrams and Delaunay Triangulations." *Computing in Euclidean Geometry*, D.-Z. D. a. F. Hwang, ed., World Scientific, 193--223.
- Glover, F. (1989). "Tabu Search - Part I." *ORSA Journal on Computing*, 1(3), 190-206.
- Glover, F. (1990a). "Tabu Search - Part II." *ORSA Journal on Computing*, 2(1), 4-32.
- Glover, F. (1990b). "Tabu search: A tutorial." *Interfaces*, 20(4), 74.
- Glover, F., and Laguna, M. "Tabu Search." *Modern Heuristic Techniques for Combinatorial Problems*, Oxford, England.
- Glover, F., and Laguna, M. (1998). *Tabu Search*, Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Glover, F., Taillard, E., and Werra, D. d. (1993). "A user's guide to tabu search." *Annals of Operations Research*, 41, 3-28.
- Guo, J., Trinidad, G., and Smith, N. (2001). "MOZART: A Multi-Objective Zoning and Aggregation Tool."
- Guptill, S. C., and Morrison, J. L. (1995). "Elements of Spatial Data Quality." International Cartographic Association, I. C. Association, ed., Elsevier Science, 153 - 166.

- Hashimoto, K. (2004). "Técnicas de Otimização Combinatória Multiobjetivo Aplicadas na Estimativa do Desempenho Elétrico de Redes de Distribuição," Escola Politécnica da Universidade de São Paulo, São Paulo.
- Haugland, D., Ho, S. C., and Laporte, G. (2005). "Designing delivery districts for the vehicle routing problem with stochastic demands." *To appear in European Journal of Operational Research*.
- Hertz, A., Taillard, E., and Werra, D. d. "A Tutorial on Tabu Search." *Proc. of Giornate di Lavoro AIRO'95 (Enterprise Systems: Management of Technological and Organizational Changes)*, Italy, 13-24.
- Hertz, A., Taillard, E., and Werra, D. d. (1997). "Tabu Search." *Local Search in Combinatorial Optimization*, J. K. Lenstra, ed., Wiley-Interscience, Chichester, England, 121-136.
- Hess, R. L., Rubin, R. S., and Jr., L. A. W. (2004). "Geographic information systems as a marketing information system technology." *Decision Support Systems*(38), 197 - 212.
- Hess, S. W. (1971). "Experiences with a Sales Districting Model: Criteria and Implementation." *Management science*, 18(4), 41.
- Ho, S. C., and Haugland, D. (2002). "A Tabu Search Heuristic for the Vehicle Routing Problem with Time Windows and Split Deliveries." *Computers & Operations Research*, 31(12), 1947 - 1964.
- Howick, R. S., and Pidd, M. (1990). "Sales force deployment models." *European Journal of Operational Research*, 48(2), 295 - 310.
- Jackson, L. E., Rouskas, G. N., and Stallmann, M. F. M. (2007). "The directional p-median problem: Definition, complexity, and algorithms." *European Journal of Operational Research*, 179(3), 1097-1108.
- Janssen, R. (1992). *Multiobjective Decision Support for Environmental Management*, Kluwer Academic, Dordrecht.
- Jaszkiewicz, A. (2001). *Multiple Objective Metaheuristic Algorithms for Combinatorial Optimization*, Politechnika Poznańska, Poznań.
- Kalcsics, J., Melo, T., Nickel, S., and Gundra, H. "Planning Sales Territories - A Facility Location Approach." *Operations Research Proceedings 2001*, Springer Verlag Berlin, 141 - 148.
- Kalcsics, J., Nickel, S., and Schroder, M. (2005). "Towards a Unified Territory Design Approach. Applications, Algorithms and GIS Integration." ITWM.
- Keenan, P. B. (1998). "Spatial Decision Support Systems for Vehicle Routing." *Decision Support Systems*, 22, 65-71.
- Keeney, R. L., and Raiffa, H. (1976). *Decisions with multiple objectives: preferences and value tradeoffs*, Wiley, New York.
- Kirkizoglu, Z. "A Genetic Algorithm Approach For Sales Territory Alignment Problem." *35th International Conference on Computers and Industrial Engineering*, Istanbul, 1185 - 1190.
- Kirkwood, C. W. (1997). *Strategic Decision Making: Multiobjective Decision Analysis with spreadsheets*, Duxbury Press, Belmont.
- Klose, A., and Drexl, A. (2005). "Facility location models for distribution system design." *European Journal of Operational Research*, 162(1), 4-29.

- Laguna, M., and Glover, F. (1996). "What is Tabu Search?" *Colorado Business Review*, LXI(5).
- Lodish, L. M. (1980). "A User-Oriented Model for Sales Force Size, Product, and Market Allocation Decisions." *The Journal of marketing*, 44(3), 70.
- Lodish, L. M., Curtis, E., Ness, M., and Simpson, M. K. (1988). "Sales Force Sizing and Deployment Using a Decision Calculus Model at Syntex Laboratories." *Interfaces*(18), 5 - 20.
- Lorena, L. A. N., and Senne, E. L. F. (2004). "A column generation approach to capacitated p-median problems." *Computers & Operations Research*, 31(6), 863-876.
- Lorena, L. A. N., Senne, E. L. F., Paiva, J. A. d. C., and Marcondes, S. P. B. (2000). "Integracao de Um Modelo de P-Mediana a Sistemas de Informacoes Geográficas."
- Macmillan, W. (2001). "Redistricting in a GIS Environment: An Optimisation Algorithm Using Switching-Points." *Journal of Geographical Systems*(3), 167 - 180.
- Malczewski, J. (1997). "Propagation of errors in multicriteria location analysis: a case study." *Multiple Criteria Decision Making*, G. Fandel and T. Gal, eds., Springer - Verlag, Berlin, 154 - 155.
- Malczewski, J. (1999). *GIS and Multicriteria Decision Analysis*, John Wiley & Sons.
- Marinoni, O. (2004). "Implementation of the analytical hierarchy process with VBA in ArcGIS." *Computers & Geosciences*, 30(6), 637-646.
- Matos, M. (2005). "Notas para as disciplinas de SAD (LCI) e SAD (MGI) - FEUP." Porto, 23.
- Mennecke, B. E., and Crossland, M. D. "Geographic Information Systems: Applications and Research Opportunities for Information Systems Researchers." *29th Annual Hawaii International Conference on System Sciences*, Hawaii, 537 - 546.
- Mohraz, K. "Geographical Analysis in SAP Business Information Warehouse." *8th ACM Symposium on GIS*, Washington, 191 - 193.
- Monteiro, C., Miranda, V., Ramírez-Rosado, I. J., Morais, C., García-Garrido, E., Mendoza-Villena, M., Fernández-Jiménez, L. A., and Martínez-Fernández, A. "Spatial Decision Support System for Site Selection Permitting of Distributed Generation Facilities." *2001 IEEE Porto Power Tech Conference*, Porto, Portugal.
- Murphy, L. D. "Geographic Information Systems: Are They Decision Support Systems?" *28th Annual Hawaii International Conference on System Sciences*, Hawaii, 131 - 140.
- Narula, S. C., and Ogbu, U. I. (1976). "An Algorithm for The P-Median Problem." 709 - 713.
- Nasirin, S. (2003). "DSS implementation in the UK retail organisations: a GIS perspective." *Information & management*, 40(4), 325.
- Negreiros, M., and Palhano, A. (2006). "The capacitated centred clustering problem." *Computers & Operations Research*, 33(6), 1639-1663.
- Parasuraman, A. (1977). "A Management-Oriented Model for Allocating Sales Effort." *Journal of marketing research*, 14(1), 22.

- Pereira, F. T., Figueira, J., Mousseaux, V., and Roy, B. (2004). "Multiple Criteria Districting Problems, Models, Algorithms, and Applications: The Public Transportation Paris Region Pricing System." 53.
- Pinals, D. (2001). "Effective Sales Force Deployment - Use or Misuse of Your Most Critical Resource."
- Pirkul, H., Gupta, R., and Rolland, E. (1999). "VisOpt: a visual interactive optimization tool for P-median problems." *Decision Support Systems*(26), 209 - 223.
- Pitsoulis, L. S., and Resende, M. G. C. (2001). "Greedy Randomized Adaptive Search Procedures." AT&T Labs Research Technical Report.
- Reese, J. (2005). "Methods for Solving the p-Median Problem: An Annotated Bibliography." Trinity University, San Antonio.
- Resende, M. G. C. (2001). "Greedy randomized adaptive search procedures (GRASP)." *Encyclopedia of Optimization*, P. M. Pardalos, ed., Kluwer Academic Publishers, 373-382.
- Resende, M. G. C., and Ribeiro, C. C. (2003). "Greedy randomized adaptive search procedures." *Handbook of Metaheuristics*, G. Kochenberger, ed., Kluwer Academic Publishers, 219-249.
- Resende, M. G. C., and Werneck, R. F. (2006). "A hybrid multistart heuristic for the uncapacitated facility location problem." *European Journal of Operational Research*, 174(1), 54-68.
- Rodrigues, M. "Introducao ao Geoprocessamento." *Simposio Brasileiro de Geoprocessamento*, Sao Paulo, 1 - 26.
- Rolland, E., Schilling, D. A., and Current, J. R. (1996). "An efficient tabu search procedure for the p-Median Problem." *European Journal of Operational Research*(96), 329 - 342.
- Saaty, T. L. (1980). *The Analytic Hierarchy Process*, McGraw-Hill.
- Scaparra, M. P., and Scutella, M. G. (2001). "Facilities, Locations, Customers: Building Blocks of Location Models. A Survey." *TR-01-08*, Dipartimento di Informatica.
- Shirabe, T. (2005). "A Model of Contiguity for Spatial Unit Allocation." *Geographical Analysis*, 37(1), 2-16.
- SIC. (2004). *Components of a GIS*, <http://maic.jmu.edu/sic/gis/components.htm>.
- Sinha, P., and Zoltners, A. A. (2001). "Sales-Force Decision Models: Insights from 25 Years of Implementation." *Interfaces*, 3(31), S8 - S44.
- Taillard, É. D. (2003). "Heuristic Methods for Large Centroid Clustering Problems." *Journal of Heuristics*, V9(1), 51-73.
- Thomas, M. R. (2002). "A GIS-based decision support system for brownfield redevelopment." *Landscape and urban planning*, 58(1), 7.
- Trejos, J., Piza, E., and Murillo, A. (1999). "A Tabu Search Algorithm for Partitioning." Preprint 1-1999 CIMPA, University of Costa Rica, San Jose, Costa Rica.
- Verter, V. V., and Kara, B. B. Y. (2001). "A GIS-based framework for hazardous materials transport risk assessment." *Risk analysis*, 21(6), 1109-20.
- Vlachopoulou, M., Silleos, G., and Manthou, V. (2001). "Geographic Information Systems in Warehouse Site Selection Decisions." *International Journal of Production Economics*(71), 205-212.

- Wei, B. C., and Chai, W. Y. (2004). "A Multiobjective Hybrid Metaheuristic Approach for GIS-based Spatial Zoning Model." *Journal of Mathematical Modelling and Algorithms*(3), 245 - 261.
- Won, Y., and Chang Lee, K. (2004). "Modified p-median approach for efficient GT cell formation." *Computers & Industrial Engineering*, 46(3), 495-510.
- Yager, R. R. (1988). "On ordered weighted averaging aggregation operators in multi criteria decision making." *IEEE Transactions on Systems, Man and Cybernetics*, 1(18), 183 - 190.
- Zeleny, M. (1982). *Multiple criteria decision making*, McGraw Hill, New York.
- Zhao, L. "Integrating rank correlation techniques with GIS for marketing analysis." *5th International Conference on GeoComputation*, University of Greenwich, United Kingdom.
- Zhao, L. (2002). "The Integration of Geographical Information Systems and Multicriteria Decision Making Models for the Analysis of Branch Bank Closures," University of New South Wales, Sydney.
- Zhao, L. "Branch Bank Closures in Sydney - A Geographical Perspective and Analysis." *Geoinformatics 2004 - Proc. 12th Int. Conf. on Geoinformatics - Geospatial Information Research - Bridging the Pacific and Atlantic*, University of Gavle, Sweden.
- Zhao, L. "Census Mapping for Market Research and Products Advertisement." *XXII International Cartographic Conference*, La Coruna, Spain.
- Zhao, L., and Garner, B. J. (2001). "Developing GIS Tools to Integrate MCDM Models for the Analysis of Bank Branch Closures."
- Zitzler, E. (1999). "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach." *IEEE transactions on evolutionary computation*, 3(4), 257.
- Zitzler, E., Laumanns, M., and Bleuler, S. "A Tutorial on Evolutionary Multiobjective Optimization." *Metaheuristics for Multiobjective Optimisation*, Berlin, 3-37.
- Zoltners, A. A. (1976). "Integer Programming Models for Sales Territory Alignment to Maximize Profit." *Journal of marketing research*, 13(4), 426.
- Zoltners, A. A. (1980). "Integer Programming Models for Sales Resource Allocation." *Management science*, 26(3), 242.
- Zoltners, A. A. (1983). "Sales Territory Alignment: A Review and Model." *Management science*, 29(11), 1237.
- Zoltners, A. A., and Lorimer, S. E. (2000). "Sales Territory: An Overlooked Productivity Tool." *Journal of Personal Selling & Sales Management*, XX(3), 139 - 150.
- Zoltners, A. A., and Sinha, P. (2005). "Sales Territory Design: Thirty Years of Modeling and Implementation." *Marketing Science*, 24(3), 313 - 331.
- Zoltners, A. A., Sinha, P., and Lorimer, S. E. (2004). *Sales Force Design for Strategic Advance*, Palgrave Macmillan, New York.

Appendix A.

GIS Components and ArcObjects

A.1 GIS Components

A GIS can be divided into five components: Hardware, Software, Data, People, and Procedures. All these components need to be in balance for the system to be successful. No single part can run without the other (SIC 2004).

A.1.1 Hardware

Hardware consists of the technical equipment needed to run a GIS including a computer system with enough power to run the software, enough memory to store large amounts of data, and input and output devices such as scanners, digitizers, GPS data loggers, media disks and printers.

A.1.2 Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are:

- Tools for the input and manipulation of geographic information;
- A database management system (DBMS);
- Tools that support geographic query, analysis, and visualization;
- A graphical user interface (GUI) for easy access to tools.

There are many different GIS software packages available today. All packages must be capable of data input, storage, management, transformation, analysis and output, but the appearance, methods, resources, and ease of use of the various systems may be very different. Today's software packages are capable of allowing both graphical and descriptive data to be stored in a single database, known as the object-relational model. Before this

innovation, the geo-relational model was used. In this model, graphical and descriptive data sets were handled separately. The modern packages usually come with a set of tools that can be customized to the user's needs.

A.1.3 Data

Possibly the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house or purchased from a commercial data provider. A GIS will integrate spatial data with other data resources and can even use a DBMS, used by most institutions to organize and maintain their data, to manage spatial data.

Perhaps the most time consuming and costly aspect of initiating a GIS is creating a database. There are several things to consider before acquiring geographic data. It is crucial to check the quality of the data before obtaining it. Errors in the data set can add many unpleasant and costly hours to implementing a GIS and the results and conclusions of the GIS analysis most likely will be wrong. Several guidelines to look at include (Guptill and Morrison 1995):

- **Lineage:** This is a description of the source material from which the data were derived and the methods of derivation, including all transformations involved in producing the final digital files. This should include all dates of the source material and updates and changes made to it;
- **Positional Accuracy:** This is the closeness of an entity in an appropriate coordinate system to that entity's true position in the system. The positional accuracy includes measures of the horizontal and vertical accuracy of the features in the data set;
- **Attribute Accuracy:** An attribute is a fact about some location, set of locations, or features on the surface of the earth. This information often includes measurements of some sort, such as temperature or elevation or a label of a place name. The source of error usually lies within the collection of these facts. It is vital to the analysis aspects of a GIS that this information is accurate;
- **Logical Consistency:** Deals with the logical rules of structure and attribute rules for spatial data and describes the compatibility of a datum with other data in a data set. There are several different mathematical theories and models used to test logical consistency such as metric and incidence tests, topological and order related tests. These consistency checks should be run at different stages in the handling of spatial data;

- **Completeness:** This is a check to see if relevant data is missing with regards to the features and the attributes. This could deal with either omission errors or spatial rules such as minimum width or area that may limit the information.

A.1.4 People

The people are the component who actually makes the GIS work. They include a plethora of positions including GIS managers, database administrators, application specialists, systems analysts and programmers. They are responsible for maintenance of the geographic database and provide technical support. People also need to be educated to make decisions on what type of system to use. People associated with a GIS can be categorized into: viewers, general users, and GIS experts.

- **Viewers** are the public at large whose only need is to browse a geographic database for referential material. These constitute the largest class of users;
- **General Users** are people who use GIS to conducting business, performing professional services and making decisions. They include facility managers, resource managers, planners, scientists, engineers, lawyers, business entrepreneurs, etc;
- **GIS experts** are the people who make the GIS work. They include GIS managers, database administrators, application experts, systems analysts and programmers. They are responsible for the maintenance of the geographic database and for the provision of technical support to the other two classes of users.

A.1.5 Procedures

Procedures include how the data will be retrieved, input into the system, stored, managed, transformed, analyzed and finally presented in a final output. The procedures are the steps taken to answer the question arising from our problem. The ability of a GIS to perform spatial analysis and answer these questions is what differentiates this type of system from any other information systems.

The transformation processes include such tasks as adjusting the coordinate system, setting a projection, correcting any digitized errors in a data set, and converting data from vector to raster or raster to vector.

A.2 ArcGIS and ArcObjects

The GIS software we have chosen to implement our algorithm is the ArcGIS 9.2. This is one of the most used GIS software in Portugal and our school has a license to use it. Other important reason for choosing this software was the possibility to easily implement the algorithm using VBA and ArcObjects. In this section we present the main characteristics of ArcGIS and ArcObjects.

ArcGIS is the name of a group of geographic information system software product lines produced by ESRI. At the desktop GIS level, ArcGIS can include:

- ArcReader, which allows one to view and query maps created with the other Arc products;
- ArcView, which allows one to view spatial data, create maps and perform basic spatial analysis;
- ArcEditor which includes all the functionality of ArcView and more advanced tools for manipulation of shapefiles and geodatabases;
- ArcInfo is the most advanced version of ArcGIS, which includes added capabilities for data manipulation, editing and analysis.

There are also server-based ArcGIS products, as well as ArcGIS products for PDAs. Extensions can be purchased separately to increase the functionality of ArcGIS.

The 9.2 release of the ArcGIS suite constituted a major change in ESRI's software offerings, aligning all their client and server products under one software architecture known as ArcGIS, developed using Microsoft Windows COM standards and designed to store data in a new proprietary RDBMS-hosted format: what ESRI termed a geodatabase.

One major difference is the programming (scripting) languages available to customize or extend the software to suit particular user needs. The ArcGIS 9.2 release includes a geoprocessing environment that allows execution of traditional GIS processing tools (such as clipping, overlay, and spatial analysis) interactively or from any scripting language that supports COM standards. Although the most popular of these is Python, others have been used, especially perl and VBScript.

ArcGIS 9.2 also includes a visual programming environment called ModelBuilder that allows users to graphically link geoprocessing tools into new tools called models. These models can be executed directly or exported to scripting languages which can then execute in

batch mode (launched from a command line), or they can undergo further editing to add branching or looping.

ArcObjects are platform independent software components, written in C++, that provide services to support GIS applications, either on the desktop in the form of thick and thin clients or on a server for Web and traditional client/server deployments. Because this architecture supports a number of unique ArcGIS products with specialized requirements, all ArcObjects are designed and built to support a multi-use scenario.

ArcObjects comprises a technology framework, an object-oriented geographic data model, an integrated library of software components and a rich collection of developer oriented resources. The ArcObjects software components are delivered as an organized collection of object-components.

The key to the success of components is that they implement in a very practical way many of the object-oriented principles now commonly accepted in software engineering. Components encapsulate behind a well-defined interface the manipulation methods and the data that characterize each instantiated object. This promotes structured and safe system development. Components facilitate software reuse because they are self-contained building blocks that can easily be assembled into larger systems. They also support inheritance¹³ and polymorphism¹⁴.

The object-component approach to software development provides a framework for anyone to extend a data model or customize a geographic information system. In the basic object model approach, only the original GIS vendor has complete customization capabilities and is free from performance and functionality bottlenecks.

Also, because of the closed nature of object implementations, users are bound to proprietary macro languages when they want to undertake customization. With the ArcObjects object-component model, users extend the data model with exactly the same technology as ESRI used to build ArcInfo. As a consequence, users have more options and their objects will perform just as well. To the user, there is absolutely no difference between the ESRI base object-components and third party developer custom components.

¹³ Inheritance is the ability to reuse functionality in other components by including reference to another object's state and behavior.

¹⁴ Polymorphism describes the process whereby each object has its specific implementation for operations such as draw, create, and delete.

Appendix B.

Spatial MCDA Analysis

B.1 Steps of Spatial MCDA

Any spatial decision problem can be structured into three major phases (Malczewski 1999):

- Intelligence phase, which examines the existence of a problem or the opportunity for change;
- Design phase, which determines the alternatives;
- Choice phase, which decides the best alternative.

The major elements involved in spatial decision making process are presented in Figure B.1 and the five most important of them are discussed below.

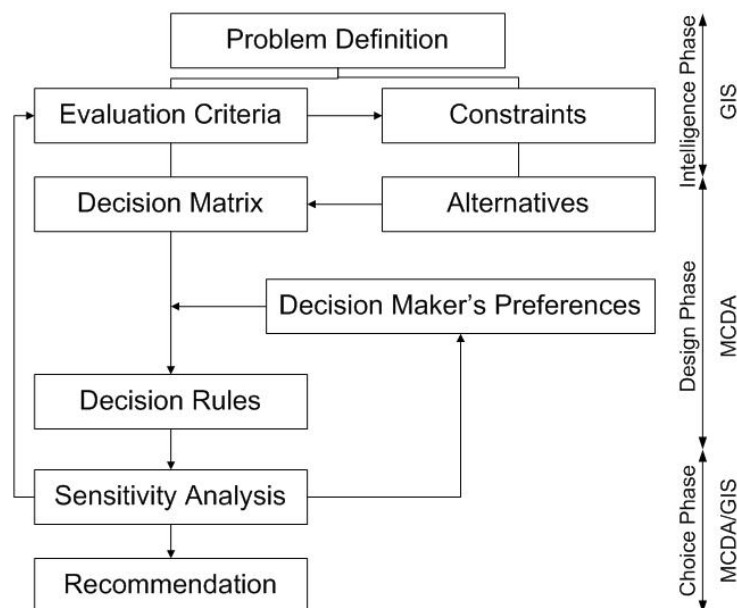


Figure B.1 - Framework for spatial multi criteria decision analysis (Malczewski 1999).

B.1.1 Problem Definition

A decision problem is the difference between the desired and existing state of the real world. It is a gap which is recognized by a decision maker. Any decision making process begins with the recognition and the definition of the problem. This stage is in the intelligence phase of decision making and it involves searching the decision environment for conditions, obtaining, processing and examining the raw data to identify the problems.

The GIS capabilities for storage, management, manipulation and analysis are used in this stage which provides major support.

B.1.2 Evaluation Criteria

After the determination of the problem, the set of evaluation criteria which includes attributes and objectives should be designated (Keeney and Raiffa 1976). This stage involves specifying a comprehensive set of objectives that reflects all relevant concerns to the decision problem and measures the achievement of those objectives which are defined as attributes. Because the evaluation criteria are related to geographical entities and the relationships between them, they can be represented in the form of maps which are referred to as attribute maps. GIS data handling and analyzing capabilities are used to generate inputs to spatial decision making analysis.

B.1.3 Decision Maker's Preferences

According to (Kirkwood 1997), a weight can be defined as a value assigned to an evaluation criterion which indicates its importance relative to other criteria under consideration.

Assigning weights of importance to evaluation criteria accounts for:

- The changes in the range of variation for each evaluation criterion;
- The different degrees of importance being attached to these ranges of variation.

There are four different techniques when assigning the weights: Ranking, Rating, Pairwise Comparison and Trade-off Analysis Methods.

- **Ranking Method:** This is the simplest method for evaluating the importance of weights which includes that every criterion under consideration is ranked in the order of the decision maker's preferences. Due to its simplicity, the method is very attractive. However, the larger the number of criteria used, the less appropriate is the method.

- **Rating Method:** The method requires the decision maker to estimate weights on the basis of a predetermined scale. One of the simplest rating methods is the point allocation approach. It is based on allocating points ranging from 0 to 100, where 0 indicates that the criterion can be ignored and 100 represents the situation where only one criterion needs to be considered. Another method is ratio estimation procedure which is a modification of the point allocation method. A score of 100 is assigned to the most important criterion and proportionally smaller weights are given to criteria lower in the order. The score assigned for the least important attribute is used to calculate the ratios.
- **Pairwise Comparison Method:** The method involves pairwise comparisons to create a ratio matrix. It takes pairwise comparisons as input and produces relative weights as output.
- **Trade-off Analysis Method:** In this method, a decision maker is required to compare two alternatives with respect to two criteria at a time and to assess which alternative is preferred. Trade-offs define a unique set of weights that will allow all of the equally preferred alternatives in the trade-offs to get the same overall value/utility. There is an assumption in this method that the trade-offs the decision maker is willing to make between any two criteria do not depend on the levels of other criteria.

B.1.4 Decision Rules

The criterion map layers and weightings must be integrated to provide an overall assessment. This is accomplished by an appropriate decision rule or aggregation function. Since a decision rule provides an ordering of all alternatives according to their performance with respect to the set of evaluation criteria, the decision problem depends on the selection of the best outcome. The most often used decision rules are discussed below.

- **Simple Additive Weighting:** It is a simple and the most often used multi-attribute decision technique (Malczewski 1997). The method is based on the weighted average. An evaluation score is calculated for each alternative by multiplying the scaled value given to the alternative of that attribute by the weights of relative importance directly assigned by the decision maker followed by summing the products for all the criteria;

- **Analytical Hierarchy Process (AHP):** The AHP developed by (Saaty 1980) is a technique for analyzing and supporting decisions in which multiple and competing objectives are involved and multiple alternatives are available. The method is based on three principles: decomposition, comparative judgment and synthesis of priorities. In the AHP, the first step is that a complex decision problem is decomposed into simpler decision problems to form a decision hierarchy. When developing a hierarchy, the top level is the ultimate goal of the decision. The hierarchy decreases from the general to the more specific until a level of attributes are reached. Each level must be linked to the next higher level. Once the decomposition is completed, cardinal rankings for objectives and alternatives are required. This is done by using pairwise comparisons which reduce the complexity of decision making since two components are considered at a time. The final step is to combine the relative weights of the levels obtained in the previous step to produce composite weights. This is done by means of a sequence of multiplications of the matrices of relative weights at each level of the hierarchy;
- **The value/utility function method:** The method is based on multi-attribute utility theory. The value function approach is applicable in the decision situation under certainty (deterministic approach) which assumes that the attributes are known with certainty whereas the utility function approach is convenient for the uncertainty conditions (probabilistic in nature). The value/utility function involves two elements: the single attribute value/utility function to transform attribute levels into an interval value/utility scale; and the trade-off analysis for defining the weights. By multiplying the utilities by the weights, the trade-offs among the attribute utilities are taken into account in the multi-attribute utility function. The overall utility or value for any alternative is a weighted average of the single attribute utilities;
- **The Ideal Point Method:** In the ideal point method the alternatives are ranked according to their separation from an ideal point. The ideal point is defined as the most desirable, weighted, hypothetical alternative (decision outcome). The alternative, closest to the ideal point, is the best alternative. The separation is measured in terms of metric distance (Janssen 1992);
- **Outranking Methods.** These methods which are also known as concordance methods are based on a pairwise comparison of alternatives. They provide an

ordinal ranking and sometimes only a partial ordering of the alternatives which means that it can only express which alternative is preferred but cannot indicate how much. The best known outranking method is the Elimination and Choice Translating Reality (ELECTRE I) and several modifications of this method have been suggested (ELECTRE II, III, IV, PROMETHEE I and II). The basic elements of this method are concordance measures which are the set of all criteria for which alternative i is not worse than the competing alternative i' and discordance measures which are the set of all criteria for which alternative i is worse than the competing alternative i' . These indicators are calculated for all pairs of alternatives and then the alternatives with the highest concordance value and with the lowest discordance value are found.

- **Ordered Weighted Average (OWA):** The OWA method has been developed in the context of fuzzy set theory (Yager 1988). There are three basic types of aggregation operators on fuzzy sets: operators for the intersection of fuzzy sets (the MIN operations); operators for the union of fuzzy sets (the MAX operations); averaging operators. (Yager 1988) introduced an aggregation technique based on the OWA operator, which is a generalization of the three basic aggregation functions. OWA is a weighted sum with ordered evaluation criteria. Thus, in addition to the criterion weights which are assigned to each evaluation criteria to indicate their relative importance, order weights are used. The order weights are associated with the criterion values on the location by location basis. Each of them is assigned to a given location's attribute value, put in decreasing order, without considering from which attribute map the value comes.
- **Goal Programming:** The goal programming method was proposed by (Charnes and Cooper 1961). It is a form of linear programming for multiple goals (evaluation criteria). Linear programming identifies from the set of feasible solutions the point which optimizes a single objective, whereas goal programming determines the point that best satisfies the set of goals in the decision problem which aims to minimize the deviations from the goals.
- **Compromise Programming:** It is a MODM method based on the displaced ideal concept which assumes that the choice among alternatives depends on the point that is used as a reference (Zeleny 1982). This point is the ideal point which defines the optimal value for each objective, considered separately.

Compromise programming attempts to minimize the distance from the ideal solution.

B.1.5 Sensitivity Analysis

In many decision rules it is assumed that complete information is available so that the criterion outcomes of each alternative are precisely known. However, in real world situations, this is not the case and analysis should be made to investigate whether the preliminary conclusions are robust or not. Sensitivity analysis aims to identify the effects of changes in the inputs which are geographical data and the decision maker's preferences on the outputs, in other words, on the ranking of alternatives. If the changes do not significantly affect the outputs, then the ranking is assumed as robust and satisfactory. If the result is unsatisfactory, it should be returned to the problem formulation step.

Appendix C.

A SSF Alignment Procedure

C.1 A Seven-Step SSF Alignment Procedure

According to (Zoltners et al. 2004), the best way to accomplish successful SST project is to have a central alignment that acts as a benchmark, with local adjustments. The central benchmark alignment should be created by an objective party, and should be based on objective business criteria. It should use consistent logic for determining staffing needs across the area. It provides quantifiable criteria against which all alignments can be judged. Local adjustments to this central benchmark can then ensure that local conditions are taken into account.

The process of incorporating local input also facilitates the acceptance of the realignment by the entire sales/services organization. Figure C.1 illustrates a seven-step process for realignment that integrates central benchmarking and local adjustment.

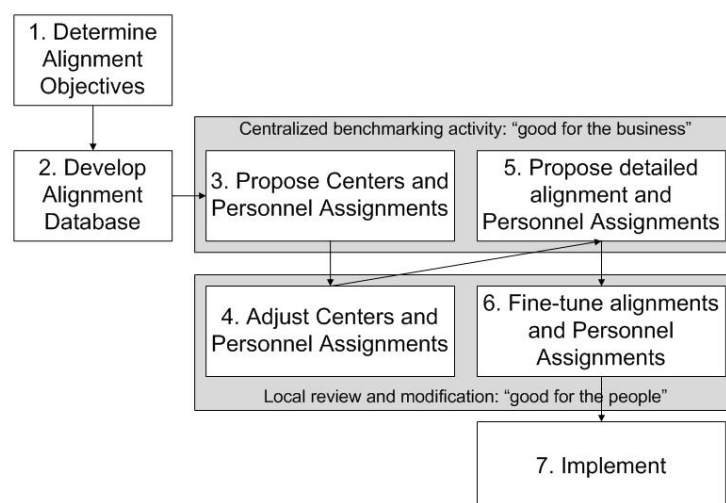


Figure C.1: A process for realignment - central benchmark with local review

In Step 1, alignment criteria such as “maximize profits”, “distribute workload equitably” and “minimize disruption” are determined. This step clarifies management's objectives and current strategy. It also defines constraints and available resources and identifies internal and external data sources, which can be tapped for the steps that follow. This step is a comprehensive review of the organization's goals and objectives, current and future strategy, constraints, resources and data availability.

In Step 2, a database is developed. The database usually includes customer and prospect locations, travel time data, and alignment attributes such as market potential, sales, and workload. The first step to more productive territories is to gather existing customer, prospect and territory data. This data may reside in a spreadsheet or centralized database. Or it may need to be collected from various sources. Internal data may be supplemented with external market or demographic data for each territory to better understand market penetration or potential. The more data we have to start with, the more informed decisions we can make later in the process.

If there is a hierarchical SSF organization with territories rolling up to districts or to regions, this information is considered and overlaid on maps and aggregated in charts and tables. Revenue, leads, forecasts or other attribute data associated with each existing territory or account also is included in this initial analysis phase. Missed opportunities or customers might be outlined. Oversaturated or underserved areas stand out in a glance. The product of this analysis is a series of maps, charts and reports that provide the current “lay of the land”. While the main purpose of this step is to gain an unbiased understanding of the existing state of the current territories, often a number of unpredicted situations occur when the data is presented. Inequitable, illogical or unbalanced territories may become immediately and visually apparent.

In Step 3, proposed SST are determined centrally, based on business needs. This analysis defines where each member of the SSF should be located in order to cover the market effectively. Preliminary personnel assignments are made at this time. This gives management a preview of where to start hiring salespeople and managers (if an expansion is planned), who stays with the sales force (if a downsizing is anticipated), or who needs to relocate (if several sales forces merge and are integrated).

The shape and balance of proposed territories can be based on any number and combination of criteria. Therefore, it is essential to understand which criteria are important to the company's business. Otherwise, it will likely end up with the wrong territory alignment.

If the decision makers are unsure about which criteria to use for balancing territories, a series of rapid alignment scenarios will provide insight. Territory alignment and optimization software play a key role in delivering quick, accurate results. Criteria to test may include:

- Revenue potential for each territory;
- Number of accounts or workload in each territory;
- Number of territories required/desired;
- Location of sales people in relation to customers and prospects in each territory;
- Lead distribution within each territory;
- Proximity to company's stores/offices/service centers;
- A modeled attribute, for example, a score based on a customer/prospect profile.

The result of this step is an optimized territory alignment, a sort of sales territory utopia. But the process is not complete. The decision makers still must take into account the real-world factors and organizational uniqueness which all SSF operate under.

In Step 4, the management team audits and adjusts the SST and personnel assignments. In small SSF, the leader of the SSF performs this task, while in larger SSF an audit group is in charge of the task. The audit group should include multiple management levels, such as regional or divisional directors and district managers.

Invariably a number of real-world conditions infringe upon the ideal alignment. The manager's goal, therefore, should be to create an optimized territory alignment within the bounds of real-world conditions. In this step, the knowledge of sales executives and regional/local sales managers is crucial. The decision makers may know, for example, that certain accounts must remain with specific sales people, or that some sales people cannot be moved, or that the company has given control of some areas to distributors or partners. This phase of territory alignment is typically interactive, with discussions going back and forth among various sales managers as to which conditions must be met. The optimized pro forma alignment is manually tweaked as needed.

In Step 5, optimal territory alignments are developed centrally. Territories are grouped into regions or districts to create the management structure if required. Also, the proposed personnel assignments are adjusted once again, since the specific geography and account assignments for each territory are now known.

In Step 6, the alignments and personnel assignments are audited and finalized once again with the help of the management team. The input of first-line managers is critical at this point in the process. Since these managers oversee the SSF members directly, they are intimately familiar with the needs of local markets and SSF personnel. Their input facilitates a smooth implementation of the new alignment in Step 7 of the process.

Once a final territory alignment is agreed upon, every first-line manager should sign off on it and be prepared to “sell it” to their SSF. Since the alignment will be balanced, fair and customized to your specific situation, selling it to SSF should be relatively easy.