STUDIES ON INNOVATION IN SITE SELECTION, PLANNING AND DESIGNING OF PETRO-RETAIL OUTLETS

DISSERTATION

Submitted in partial fulfillment of the

requirements for the award of the degree

of

MASTER OF TECHNOLOGY

(PETRO-INFORMATICS)

Submitted by

N.KARUNAKARAN

(R 060106007)

Under the supervision of
MR. S. PRABAKARAN

COLLEGE OF ENGINEERING STUDIES
UNIVERSITY OF PETROLEUM & ENERGY STUDIES

GURGAON - 122001 (INDIA)
APRIL - 2008

ACKNOWLEDGEMENT

No work of this kind might have seen the light of day unless there must have been some divine inspiration behind it. And I feel myself fortunate enough in this regard to have been blessed with the divine grace of the **Almighty God**

I am grateful to **Mr. S.Prabakaran,** for giving me opportunity for this project, also for his timely help and inspiring guidance without which I would have not successfully completed this project. It is a matter of great pleasure for me to get this opportunity of expressing my deep and sincere of gratitude for his guidance and generous support that gave me sufficient strength to prepare this project.

I am grateful to the **U.P.E.S, Gurgaon Academic Division** for their effort in completion of project.

At last but not the least I would like to thank **my parents** for their everlasting support.

N.Karunakaran

M.Tech(Petro-Informatics)

CERTIFICATE

This is to certify that the dissertation entitled **"STUDIES ON INNOVATION IN SITE SELECTION, PLANNING AND DESIGNING OF PETRO-RETAIL OUTLETS"** which is being submitted in partial fulfillment for the award of Master of Technology in Petro-Informatics by **Mr. N.Karunakaran** is a bonafide work carried out by her under my supervision and guidance. This work has not been published elsewhere or submitted for the award of any other degree.

SP 18/04/08

Mr S.Prabakaran

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION

1.1 Introduction

Petroleum marketers are faced with an ever-changing retail environment. Growth was once primarily a function of new site investments designed to build gasoline sales. Now, growth is much more focused on profit center optimization for existing real estate. Emphasis is being placed on improving profitability by providing consumers with a broader range of convenience offerings.

Driving this change are several factors:

- Prime corner locations simply are not as available as in the past. Growth must come from either investments at "secondary" locations or from existing sites that do not meet their potential [both comprise same-store growth].
- Major gasoline retailers have noted the emergence and success of regional marketers who provide a mix of profit centers that better maximize the value of their real estate. The majors' capital investments are now geared more toward building large convenience store formats at existing sites than on continuing previous habits of prioritizing ground-up investments.
- More and more emphasis is being placed on marketing programs. Credit card rebates and frequent buyer programs permeate the retail environment as everyone continues to compete in a flat-to-moderate demand growth environment.
- Joint ventures and proprietary branded ready-to-eat offerings reflect marketers' attempt to cut into the rising demand in that convenience food product category.
- Many believe that gasoline margins will erode in the long run as competitors seek to drive traffic and extract more value from convenience item margins.

1.1.1 Maximizing the site

Phase 1. Demand is rising correspondingly with new household construction and business start-ups and relocations. The retailer's goal here is to maximize profit center sales with a minimum number of investments while keeping a watchful eye on same-brand cannibalization and competitor cut-off.

Phase 2. Growth is where capital investment decisions focus on existing locations. Key issues evolve around where (at what sites) and how (which profit centers) to prioritize capital. This phase is where most of the investment is being made today.

Phase 3. Growth permeates the entire retail life cycle but often seems to only be a priority for marketers in their latter stages of development. This is where creative marketing programs must be employed to outsmart competitors and target specific customer groups. Ideas on how to revitalize communications and develop customer centric programs will be presented.

1.1.2 Location

The most strategically important decision marketers make in the convenience retail business is where to place their stores. This is true from several perspectives.

First, consider the consumer's perspective. Much research has been done to show that the most important factor determining a consumer's propensity to shop at a given store for gasoline and select merchandise is the speed of the transaction and the "convenience" of the location. The bulk of consumers have secondary preferences for price and service provided that these features are not significantly different from competitor to competitor.

Consumers show a very low propensity to treat your stores as destination shopping locations and mostly relate purchases to an interruption of their daily routines. Of course, there are destination trips for many grocery items but the average store must establish itself as the most convenient location to intercept the bulk of consumer demand.

Second, consider the retailers perspective. Whether you are building a new store or face-lifting an existing one, the incremental business you will generate is, predominately, at the detriment to a competitor or one of your own locations.

In markets where significant demand growth is occurring (Phase 1) there will be less cannibalization, but some will always exist. For that reason it is important to recognize how locations interact with one another in the competitive landscape. Each new or rebuilt location disrupts the network. The retailer's goal must be to understand these interacting relationships and prioritize capital investments to minimize cost and same brand cannibalization while attempting to "cut-off" competitor sales. To achieve this requires a detailed understanding of the nature of supply and demand in the local marketplace.

Site selection decisions should not be left to chance. Because it is so dynamic, the convenience retail business is by far the most difficult retail format in which to select long term successful locations. Roads change, people move, consumer preferences shift and competitors build.

1.2 PRESENT STATE OF KNOWLEDGE, LITERATURE REVIEW & ISSUES & PROBLEMS RELEVANT TO THE STUDY

One of the recent study related to our topics, plans have been developed to build a NASCAR track on valuable wetland habitat on the western shore of Staten Island (Alderson 3/9/05). The NY/NJ Harbor Estuary currently supports many competing uses. "It is trout stream and estuary, water supply and sewer, ship channel and shad river, playground and chamber pot. It is abused, revered, and almost always misunderstood" (Boyle 1969). The Harbor Estuary provides habitat for a number of fish and shellfish species.

And also some researcher worked on this area that the degraded habitat of NY/NJ Harbor is important to ensure that the harbor can continue to be used as an economic, recreational, and ecological resource. The NY/NJ Harbor Estuary Program (HEP) has been working with government agencies, non-governmental organizations (NGOs), and the public since 1996 to protect and restore habitat (Mandarano 2004).

The flowing related research work done on this area.

1. Incorporation (Mandarano 2004) of a GIS into the site selection process would serve to address this issue. Moreover, the exploration of an alternate approach to habitat

restoration site selection provides the opportunity to analyze and assess the current methods.

2. Related study [1] Multi outlets, one of the most important decisions a retailer can make is where to locate a retail outlet. Because convenience is so important to today's consumers a retail store can prosper or fail solely based on its location. Recently a changing retail environment is augmenting the location importance as retail economic groups develop.

3. Another paper uses the literature, when available, to explore how malls are evaluated and perceived by retail managers. Since there are no formal studies that examine this area, this paper concludes by discussing several issues that should be addressed by future research.

4. Another paper they present an integrated distribution center site selection and space requirement problem on a two-stage network in which products are shipped from plants to distribution centers, where they are stored for an arbitrary period of time and then delivered to retailers.

5. In another study [4], GIS is an emerging marketing tool. This study examines the Georgian landscape plant retail market using GIS in conjunction with other traditional market research tools. Spending and logistic regression propensity scores are analyzed for prevailing geographic patterns. This allows retailers to make store location decisions by identifying underserved markets.

6. Another attempt also made by Michael Nwogugu [5] this article critiques existing store-location models and explains the many mathematical and methodological problems inherent in these models. The key issues are that behavioral factors are completely omitted in these models, emphasis on distance in un-warranted and site-specific and retailer specific characteristics are typically omitted in these models. The author introduces new site-selection models. The US retailing sector is used an example.

1.3 Research objectives

The main objectives of this research study are:-

1. To understand the importance of petro-retail outlets.

2. To identify the different site location to install petro-retail outlets by using the GIS Technology for a particular location.

3. To develop the simulation model for designing the selection for Petro-Retail outlets on the basis of exciting model.

4. This simulation model on the basis of to minimize the cost and time schedule by using of the some tools and techniques.(Like PERT/CPM)

5. And to give recommendation and suggestion to install petro-retail outlets in the selected site location.

1.4 Research Methodology

The research techniques used in this research are as follows:

1. Data collection of a large area.

2. Study of various sites of petrol pump station has been already installed.

3. Calculation of distance between the petrol pump station.

4. Collecting information about the locality, lifestyle of that complete area.

5. Analyze the areas where possible installations of petro-retail outlets have been done.

6. Recommend for installation of petro-retail outlets in that area.

1. 5 Organization of this Research Work

Brief outlines of different chapters are given below:-

Chapter 1 contains a brief introduction to the subject to put the problems and investigations in proper perspective. It also provides a brief introduction to the study, motivation for the research, objectives of the research and an outline of organization of this research work with a Chapter wise summary.

Chapter 2 reviews the literature relevant to this research. Literature review is focused on the contemporary work being done towards area of the innovation in site selection, planning, and designing of petro – Designing outlets. Relevant literature on site selection has also been reviewed, in particular, those areas that are relevant to the research being envisaged. Literature on the designing applications of various types of Petro-Retail Outlet Model, Mathematical Formulation of Location Models and graphical representation has also been studied to identify gaps in existing knowledge in the field.

Chapter 3 deals about the location problems, the term facility is used to denote an object whose spatial position we are trying to determine in order to optimize the interaction with other pre-existing objects and also deals about the type of a facility is another important property of the different selection location.

Chapter 4 deals about the construction of a location model, it describes about the decision evels involved in Retail-Outlet network and it explains a possible classification of assessment location explanatory variables and the classical procedures used.

Chapter 5 discuss about the findings and analysis in the graphical representation and the model which is used for the analysis.

Chapter 6 gives about the conclusion and the limitation of the report.

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

In a location-allocation problem (LAP), an optimal number of facilities have to be placed in an area of interest in order to satisfy the customer demand. The optimization objective can be, e.g., to minimize costs or to maximize profit and the situation can be characterized by various constraints. The total location area may also be divided into regions which have each their special characteristics. There are often several plausible scenarios for an LAP representing, for example, different budgets limitations, cost functions, or estimates for market growth. In this Thesis, the emphasis is on retail location-allocation problems and it is illustrated with a case study in retailing.

When dealing with a large number of location-allocation solutions for various scenarios, it is necessary to examine which option is the best one according to specified criteria. The first objective of this Thesis is therefore to define decision criteria for retail applications. The proposed criteria include several monetary, facility related, and customer related attributes.

Besides the facility locations and customer flows, location-allocation solutions are characterized by various derived properties, for example market share, distance to competitors, and the amount of unassigned demand. These derived properties also include the chosen decision criteria. When the total location area is composed of regions, these properties can be expressed for each region and thus have a spatial distribution. In decision support, it is often important to be able to visualize the solutions and their derived properties: finding suitable visualization techniques is also the second main objective of this Thesis. The visualization of geographic properties can be assisted by geographic information systems (GIS).

However, existing tools of GIS software are not sufficient for the visualization of dozens of multi-dimensional location-allocation solutions. This problem can be circumvented either by concentrating on the statistical properties of the solution set instead of examining and comparing the options one by one, or by trying to classify the solutions into groups whose properties can then be more easily compared. The methods of cluster analysis and subgroup discovery could prove useful in grouping the solutions. Also, principal component analysis can be used to reduce the dimensions of the problem and to facilitate visualization.

2.2 Background

In location problems an optimal number of facilities have to be placed in an area of interest in order to satisfy the customer demand. The total area can be subdivided into regions with different characteristics and there may be constraints restricting, e.g., the number of facilities in a region or in a subset of regions. Further constraints may define, for example, the total investment budget, the return of investment per facility, or the capacities of different types of facilities. The constraints and other problem parameters may also be varied from one scenario to another. The location problem can be solved by using, for instance, evolutionary algorithms that provide a set of solutions for each scenario. The key problem is then to determine in what respect these solutions differ and which is the best one for the scenario in question according to some chosen criteria.

Location problems are common optimization problems that occur when choosing the locations of various public, industrial, or commercial service centers in regard to the needs these facilities aim to satisfy. Since location problem are usually complex problems, explicitly formulating the model as well as analyzing the solution alternatives are often carried out by specialized analysts who later communicate the results to the decision maker. The use of experts is also preferred because the implementation of a chosen location option is often costly and, once carried out, cannot be easily altered.

Many practical location problems are multi-criteria in nature and the chosen objective function consists thus of several weighted sub-functions. Besides these optimization targets, there can also be other interesting properties that are only used as constraints or even disregarded in the optimization process. After the optimization, it is necessary to compare the values and spatial distributions of the optimization sub-functions and of other interesting properties in order to find the most suitable location option. Finding a good problem definition can also be crucial in solving location problems: therefore, it is important to compare scenario solution pairs in order to find the problem formulation that produces the best results according to chosen criteria.

The solutions of location problems often have a geographical interpretation and can be visualized by using geographic information systems (GIS). The visualization is especially interesting in situations where differences in the spatial distributions of solution characteristics can have great consequences; furthermore, this spatial element is rarely accounted for in the optimization process where usually only global values are being compared.

2.3 Introduction to Location Problems

Location problems have been studied ever since Weber posed the first facility location problem at the beginning of the 20th century. This problem has subsequently become known as the Weber problem and it can be formulated as finding the optimal locations of n facilities in Euclidean space so that the transportation costs created in order to satisfy the demand of m customers are minimized. Location analysis did not however become a well established field of study until in the 1960s when the seminal paper by Hakimi established new, important results in location theory and captured the interest of many researchers. Since then, many different types of location models solution methods, and heuristics have been developed, and the field has diversified to such an extent that it has become necessary to propose classification schemes . There is also an extensive bibliography on location analysis provided by Trevor Hale and an online library of location algorithms known as LoLA compiled by the University of Kaiserslautern and Fraunhofer ITWM.

The general objective of location problems is to place an optimal number of service centers, or facilities, in a chosen area. The criteria used to measure the optimality of a set of locations can include for example one or several of the following attributes: investment costs, distance to the customers, estimated profit, or customer coverage. The basic settings of location problems (and also of simple location problems) can be thought to consist of facilities, locations, and customers and of their interrelationships. The definitions and properties of these basic components will be discussed in this Chapter along with some different types of location models. The presentation provided in this Thesis has been influenced by the paper of Scaparra and Scutella which proposes a unified framework for characterizing the different aspects of location problems.

The world is dynamic and there are processes and decisions which evolve in time; in location problems, for example the customer demand can often have temporal variations. These temporal dynamics increase the complexity of the problem and pose additional requirements. To simplify the situation under examination, the location problem is assumed to be static.

2.4 Literature Review

"To those who know it, the Hudson River is the most beautiful, messed up, productive, ignored, and surprising piece of water on the face of the earth" (Boyle 1969). This river, which is part of the equally surprising New York-New Jersey (NY/NJ) Harbor Estuary, was once described as "clear as crystal, and as fresh as milk" (Boyle 1969). The Harbor was teeming with life; it contained an abundance of fish and 350 square miles of oyster beds. Today, the NY/NJ Harbor Estuary is seen by many to be an industrial wasteland, with degraded habitat, contaminated sediments, and polluted water. About 80 percent of the harbor's original benthic habitat and tidal wetlands have been lost. This loss accounts for approximately 300,000 acres, or an area roughly 1.5 times the current area of New York City (Steinberg et al. 2004). Threats to habitat in the NY/NJ Harbor Estuary are a continuing problem.

Recently, plans have been developed to build a NASCAR track on valuable wetland habitat on the western shore of Staten Island (Alderson 3/9/05). While it is true that the NY/NJ Harbor has suffered severe environmental degradation due to industrial pollution, urban development, and harbor dredging, it should not be written off as a lost cause. Despite its history of environmental problems, the Harbor continues to serve as a valuable economic, ecological, and recreational resource for the region. The NY/NJ Harbor Estuary currently supports many competing uses. "It is trout stream and estuary, water supply and sewer, ship channel and shad river, playground and chamber pot. It is abused, revered, and almost always misunderstood" (Boyle 1969). The Harbor Estuary provides habitat for a number of fish and shellfish species. It is located along the Atlantic flyway, providing feeding and resting areas for both migratory and local bird species

(Adams 1998). This diverse ecological habitat also serves as one of the most heavily utilized shipping ports on the east coast of the United States. The NY/NJ Harbor watershed is located in the most densely populated region of the nation, supporting a population of over 20 million people (NY/NJ HEP 1996).

Restoring the degraded habitat of NY/NJ Harbor is important to ensure that the harbor can continue to be used as an economic, recreational, and ecological resource. The NY/NJ Harbor Estuary Program (HEP) has been working with government agencies, non-governmental organizations (NGOs), and the public since 1996 to protect and restore habitat (Mandarano 2004). This paper discusses the measures presently being taken by the New York-New Jersey Harbor Estuary Program to restore habitat in the harbor and explores a method to improve the scientific rigor of current restoration site selection efforts using a geographic information system (GIS).

Concern has been expressed by representatives of agencies that fund and undertake habitat restoration efforts in the harbor, that restoration site selection by the HEP is not scientifically grounded (Mandarano 2004). Incorporation of a GIS into the site selection process would serve to address this issue. Moreover, the exploration of an alternate approach to habitat restoration site selection provides the opportunity to analyze and assess the current methods.

Related study [1] Multi outlets one of the most important decisions a retailer can make is where to locate a retail outlet. Because convenience is so important to today's consumers a retail store can prosper or fail solely based on its location. Recently a changing retail environment is augmenting the location importance as retail economic groups develop.

In the forty-plus years since malls first appeared, these shopping centers have evolved into an integral part of the American culture. However, in recent years the number of malls has increased so dramatically that competition between malls, and even mall closures have become commonplace. Although there is not a large body of work that studies malls from a marketing perspective, the vast majority of what has been published adopts a consumer rather than a managerial perspective. This paper uses the literature, when available, to explore how malls are evaluated and perceived by retail managers. Since there are no formal studies that examine this area, this paper concludes by discussing several issues that should be addressed by future research. [2]

In this paper they presented an integrated distribution center site selection and space requirement problem on a two-stage network in which products are shipped from plants to distribution centers, where they are stored for an arbitrary period of time and then delivered to retailers. The objective of the problem is to minimize total inbound and outbound transportation costs and total distribution center construction cost which includes fixed costs related to their locations and variable costs related to their space requirements for given service levels. Each distribution center is modeled as an *M/G/c* queuing system, in which each server represents a storage slot. We formulate this problem as a nonlinear mixed integer program with a probabilistic constraint. Two cases are considered. For the continuous unbounded size case, we find an approximate formula for the overflow probability and restructure this model into a connection location problem. For the discrete size option case, we reformulate the problem into a capacitated connection location problem with discrete size options. Computational results and a comparison of the two cases are provided. [3].

In another study [4], GIS is an emerging marketing tool. This study examines the Georgian landscape plant retail market using GIS in conjunction with other traditional market research tools. Spending and logistic regression propensity scores are analyzed for prevailing geographic patterns. This allows retailers to make store location decisions by identifying underserved markets.

Another attempt also made by Michael Nwogugu [5], this article critiques existing store-location models and explains the many mathematical and methodological problems inherent in these models. The key issues are that behavioral factors are completely omitted in these models, emphasis on distance in un-warranted and site-specific and retailer specific characteristics are typically omitted in these models. The author introduces new site-selection models. The US retailing sector is used an example.

2.5 The Top 10 Site Selection Techniques

The following 10 items represent a proven framework for assessing site selection opportunities:

1. Collect Demographic Data

One of the driving elements of location research involves the potential number of customers. Understanding where potential customers are concentrated (and whether there are sufficient quantities to support a branch) is like providing the gas to start the engine in identifying possible locations for expansion. The primary pit stop of this information is the U.S. Census Bureau. The 2000 Census provides a number of the lubricants, oils, air and tools for your vehicles; it is a wealth of information on the quantity of persons and households in each market. It is equally important, however, for banks to understand the nature of the potential customers in each market. For example, collecting data on the age, income distribution, educational attainment and occupation type will provide great insight into understanding the market quality of various areas. Visualizing demographic information on maps (see map example) can be a very useful way to understand demographic data.

2. Build an Inventory of Competition

Superior racers know their competition's style, strengths, and weakness, and they utilize this information to their advantage. Financial institutions should do the same. The Federal Deposit Insurance Corporation (FDIC) maintains an on-line database of all bank and savings and loan branch locations including the exact address and deposit levels. The FDIC also maintains a record of branch openings, closings and relocations that is updated every two weeks. It is very easy to access this data through the Internet. The FDIC databases will need to be supplemented by local intelligence on credit union, mortgage and brokerage offices for each target area, as well as an assessment of the quality of each competitor. With an inventory of competition, it will now be possible to identify the competitive intensity of each potential market areas. Some of the more common measures used to evaluate markets include: average deposits per branch, market households per branch, deposit market share, growth/decline of branch deposits, and share of outlets.

3. Generate Market Characteristics

Branch location decisions need to be made with a long-term perspective in mind. market characteristics, such as unemployment levels, planned road improvements and household growth, can provide clues into the overall quality of the market. Generally, local municipal and county government departments are able to provide this type of data along with information on zoning, building permit trends and planned economic development. Again, it is very helpful to display this information on a map to improve your ability to interpret the data.

4. Quantify Market Demand for Financial Products and Services

A race team must answer viability questions before deciding to enter a race; the same is true for financial institutions. Is there sufficient demand for financial products and services in your target markets? Data on average household balances and product usage can be purchased from market research firms by geographic area for both current year volumes and five-year projections. It is also possible to collect information on delivery channel preferences (e.g. traditional branch, in-store branch, ATM, telephone banking or on-line banking) by market area. An understanding of market product usage and balances can provide a basis to assess whether sufficient demand exists to support a new branch office.

5. Understand MCIF Customer Account Information

Your existing customer information provides valuable insight into future expansion opportunities as well. Financial institutions have a real advantage in capturing quality information on current customers as compared to many other industries such as restaurants or retailers. With key customer information including address, balances levels, products, branches visited, and target/profitability segments banks have a wealth of useful knowledge. This information, typically captured within a Marketing Customer Information File (MCIF) database can provide important clues for site selection opportunities. MCIF data can help determine whether your current branches are located close to your high value customers. MCIF data can also be used to evaluate the types of accounts sold through the branch as opposed to alternative banking channels such as telephone banking or PC Banking channels.

6. Recognize Customer Behavior Patterns

Customer develop patterns and preferences in conducting transactions and opening accounts. Developing an understanding of these preferences can allow banks to become more skillful at identifying potential new locations for business expansion. For example, important questions to be asked include:

1. How far are current customers willing to travel to conduct transactions or to open accounts?
2. Do most customers typically conduct transactions at one location or several branches?
3. Are there differences geographically in terms of where new account relationships are opened as opposed to existing relationships?

MCIF systems and transaction files can be used to collect this information. With this data, banks will gain insight into issues such as trade area size (the area where a majority of existing customers are drawn from), how close new locations can be to existing branches, and where the new account growth is occurring that may justify a new branch location.

7. Identify Potential Site Location Opportunities

The information collected on market demographics, competition and existing customers can now be used to identify high quality sites. For example, potential sites can be identified using the following approaches:

1. Household Penetration – Identify high quality markets near existing branch offices where household penetration is low.
2. Core Customer Segments – Identify concentrations of high-value customers without a convenient nearby branch location.
3. High Growth/Limited Competition – Focus on areas with high household growth potential and below average competition to gain a first-mover competitive advantage.

8. Conduct Fieldwork

Field evaluation provides an opportunity to gain firsthand knowledge of the market. This is where the rubber hits the road. Field evaluation tasks can include:

1. Validation of all competitors
2. Evaluation of traffic volumes and shopping patterns
3. Review of specific site locations
4. Identification of retail concentrations and shopping centers and how many financial institutions are located in these areas
5. Analysis of the quality of local neighborhoods to determine whether they are consistent with demographic information
6. Meetings with local agencies such as planning departments, road commissions, and chambers of commerce for updated information on demographic projections, road projects, zoning, and new business development.

9. Identify Site Characteristics

We have already alluded to the impact that good (or bad) site characteristics can have on the sales performance of branch locations. While in the field, an assessment of key site characteristics for any potential locations should involve the following items:

1. Visibility – Will the branch building and signage be clearly visible from the road?
2. Parking – Will a sufficient number of parking spaces be available?
3. Ingress/Egress – Can customer easily enter and exit the site? Are there medians that restrict access to the site?
4. Accessibility – Are there major highways in the area that allow customers convenient access to the area around the potential branch location?
5. Proximity to retail shopping centers – The presence of retail districts will draw customers into an area. Does the site have concentrations of high-quality retail nearby?
6. Safety/Security – Does the site present potential problems for employees or customers?

10. Develop Final Recommendations for Senior Management

Senior management will need to clearly understand the methodology and approach in prioritizing the best locations for branch expansion. Using the data compiled in each of the steps above, a numerical score can be calculated for each site that provides the basis for ranking the desirability of each site location. This ranking should be based on key characteristics such as market size, market growth, competitive intensity, financial product demand, presence of target customer segments, field assessment, and quality of site characteristics. The resulting list of locations should include an explanation of the rationale describing the reasons why each site is recommended or rejected. And finally, using maps and tables to help illustrate why specific sites have been recommended as part of this plan can be extremely beneficial.

CHAPTER 3: STUDY OF LOCATION PROBLEMS

3.1 Basic Components of Location Problems

3.1.1 Facilities

In location problems, the term facility is used to denote an object whose spatial position we are trying to determine in order to optimize the interaction with other pre-existing objects. The classical examples for facilities include objects such as retail outlets, schools, hospitals, warehouses, and plants or in general any other industrial, commercial, or public structures. Some less conventional uses include, for example, trees in landscape conservation, territories in territory design, and different types of water sources in water source management. The facilities are usually characterized by, among other things, their number, type, and costs. Other facility-connected properties can include for instance profit, capacity, attraction range (within which the customers are drawn to the facility), and the type of service provided.

One of the properties characterizing a location problem is the number of new facilities to be located in the given area. The simplest case is the single-facility problem in which only one new facility is to be established. The more general case is called the multi-facility problem in which the aim is to locate simultaneously more than one facility. The number of facilities can be either pre-specified or determined during the optimization process. In the latter case, there might be limitations to the number of facilities to be established, for example a minimum or a maximum or both. The situation also often involves a trade-off between, e.g., improved customer coverage and increased costs for operating a greater number of facilities. The optimization problem can also include the possibility to close existing facilities or to change their type.

The type of a facility is another important property: among other things, it can determine the kind of service provided the production capacity of the facility, its profit, and its costs. In the simplest case, all the facilities are supposed to be identical with respect to their size and the kind of service they offer. However, it is often necessary to locate facilities that differ from one another, for instance hospitals and smaller health care units. Location models can also be differentiated into single-service and multi-service types, based on whether the facilities can provide only one or many services.

It can also be taken into consideration whether the facilities can supply an infinite demand or whether their production and supply capacity is limited. In this respect, the problems are often classified into uncapacitated and capacitated. The production capacity of a facility might also depend on the site where it is to be established. Finally, the representations of the facilities can also have different configurations: facilities can be thought to occupy point or area locations, or even to take a certain specialized shape, for example a graph or a tree. Path-shaped location models are often used in transportation planning problems that concern for example locating new roads and highways, railroad lines, or airlines routes. Path-shaped facilities also arise in telecommunications systems, for example in designing the locations of fiber optic cables.

The total cost concerning the facilities can be divided into fixed expenses (incurring upon the opening, the closure, or the change of type of a facility) and variable costs. These fixed costs can be connected to the specific location or to the facility type and they remain the same no matter how many customers frequent the facility or how far away they might be. Variable charges can be for example storage costs which depend on the number of unsold items or service delivery costs which are usually a function of the distance between users and services.

3.1.2 Locations

The second essential element of location problems is the physical space, or locations, where the facilities can be positioned. The set of eligible locations has three possible representations: discrete, continuous, and network. Continuous space models are sometimes referred to as site-generation models since the generation of appropriate sites is left to the model at hand. Conversely, discrete space models are sometimes referred to as site-selection models since we have a priori knowledge of the site candidates. In discrete models, the decision maker has to determine a limited number of plausible sites for facility locations. This is the most natural way to deal with problems where e.g. the land availability is limited and new facilities can only be opened at some fixed points in the area under consideration. For example, this approach is adopted by Leaoetal. in studying the allocation of waste landfills in growing urban areas. Discrete models can also be used when abstracting from a site location to a region.

On the other hand, in many location problems the space for facility sites is continuous, i.e. all the possible site locations are determined by one or more continuously varying coordinates. These continuous location problems are usually considered in Euclidean space R^2 or, more generally, in an n-dimensional space R^n. For evident geographical reasons, two-dimensional problems are the most popular, but more abstract settings may require more than two dimensions. For example, product positioning in marketing can be represented as a location problem in an n-dimensional attribute space.

The network-based model is the third type of location models that can be distinguished with respect to the locations. Problems defined on networks can be either continuous or discrete depending on whether the links of the network are considered as a continuous set of candidate locations or whether only the nodes are eligible for the placement of new facilities. The underlying graph can also have different structures. Tree-like structures are particularly popular: for example, distribution and telecommunications systems can often be modeled with star-like trees. Examples of the use of networks in location problems also include the placement of health care units in rural areas in order to guarantee good health service coverage and the modeling of public school districts.

The eligibility of location points can also be defined by many other features: the available space may be restricted by forbidden zones where the facilities may not be located or there may be conditions forcing to choose at least (or, conversely, at most) one site in a given region. There might also be additional compatibility constraints that depend on existing infrastructure: for instance locating a health care center in an area where it is difficult to access because of poor transportation connections would not be ideal.

3.1.3 Customers

Traditionally, the generic term customer is used in location problems to denote a person requiring accessibility to a service or to the supply of a good. The word can also be used more abstractly to indicate any object which must interact with a new facility: examples include communities in rural areas requiring emergency services, urban areas demanding waste disposal sites, or remote terminals in communications networks. Facility problems arise from the need to satisfy the demand of a set of customers by the optimal location of facilities and it is therefore crucial to know the distribution, demand, and behaviour of these customers.

With regard to the distribution, customers can by assumed to be either uniformly distributed over a given area or they can be located at specific points in space or at vertices in a network

Demand is expressed by assigning each customer a weight which expresses the amount of service required. The customer can be a single user or a representation of an area for which the service is destined, such as a community, a district, or a region. In the latter case, the weight is used to account for the total demand arising from the area in question and it might be, for example, a function of the population size. Both for single users and larger areas, demand might not always be deterministic. For example, for facilities that provide non-essential services (e.g. restaurants, retail retail-outlets, or ATM machines), customer demand is often a function of the cost of the service and of the distance to the facility. For example, so called gravity models can be used in modeling nondeterministic demand. On the other hand, if the facilities provide essential services like water, electricity, or health care, customer demand is often assumed to be deterministic and known a priori. Moreover, customer demand might not always be static: there can be, for example, temporal variations. In retailing, demand and supply are often expressed in discrete units (either natural or artificial) but they might also be continuous. As facilities may provide different types of services, customers may also demand multiple services or goods. Problems concerning demand for several kinds of services are sometimes referred to as multi-commodity problems

Customers can be characterized according to their behaviour. In some cases, customers are free to choose the facility from which they will be served; in these applications, they might always prefer the closest facility or there might be some other criterion to reflect their preferences.

3.1.4 Interrelationships

Facility-facility relationships can be used to model synergistic or competitive interactions among new and existing centers. This can be done, for example, in the two following ways: first, by allowing a flow of services or goods from one facility to another in the model and, second, by assuming that the facilities operate in a competitive environment and that their relative positions affect the performance of the whole system. The latter aspect is often typical of economic location problems which deal with commercial activities such as retail outlets.

Customer-facility relationships describe the way the users' demand is satisfied by the supply centers. One key element is the number of facilities that customers use to satisfy their demand: in some applications, customers might be allowed to use only a single facility whereas in others, each customer's needs might be served by several centers. Customers and facilities are also linked through the customers' preferences for a supply center location: customers can find the closeness of a facility desirable and would like to have it located as close as possible, they might be indifferent to the facility location, or the facility could be considered undesirable as is the case of obnoxious, polluting, and dangerous activities. There are several ways to determine which facility a customer will patronize when situated within the sphere of attraction of two or more facilities. The customer's demand might be, for example, equally divided among the facilities or it might all go to the facility that is judged to be the most attractive.

Facility-location relationships usually concern investment costs and capacity restrictions related to particular sites. They might be due to specific geographic features, to existing infrastructure, or to territorial and administrative regulations which might cause the construction of a facility to be more costly than at other locations or, for example, set limits to the production capacity of the facility.

Customer-location relationships are also an essential part of location problems it is necessary to measure the interaction among the locations where the facilities are situated and the customers' access to the provided services in order to guide the optimization process towards a satisfactory result. In general, the quality of interactions is thought to be directly related to the relative spatial positions of the customers and the facility and it is usually expressed by some notion of distance. There are several different distance measures, the most familiar being the Euclidean distance. This distance is a special case of L_p-distances that can be mathematically expressed for two points x and y with coordinates (x_1, x_2) and (y_1, y_2)as:

$$L_p(x, y) = \left(|x_1 - y_1|^p + |x_2 - y_2|^p\right)^{1/p} \tag{3.1}$$

For Euclidean distance, p = 2. Other often used distance measures include, for example, road distance and travel time. The distance measures can also help to specify further efficiency criteria, for example by converting the estimated distances into representative costs.

Customer-customer relationships are especially relevant in models including market externalities such as queuing delays or congestion. In these situations, the utility of one customer is negatively or positively affected by the actions of other customers. Examples of negative congestion externalities include traffic jams in transportation systems, noise pollution in residential areas, or waiting time in service centers. It is supposed that when deciding which facility to patronize, customers not only consider the distance they have to travel but also the cost associated with the congestion externalities.

3.2 Combination of the Basic Components into a Model

The basic elements of facility location problems can be combined into a model by defining four factors: (i) actors, (ii) objective(s), (iii) available instruments, and (iv) environmental constraints. The model formulated a location problem in retailing responds to the following questions: who are the decision-making agents?; what are the goals of these agents?; what instruments may the agents use to achieve their objective(s)?; and what constraints prevent the agents from achieving unlimited amounts of their objective(s)? Constraints include, for example, parameters defining the environment and the nature of competition or market structure. The final aspect of a model is choosing the relevant time horizon.

In this Thesis, it is assumed that the situation is characterized by perfect competition so that a single retailer has no influence on the prices. Also, the availability of inputs is not taken into consideration since the problem concerns a product market. The time horizon is assumed to be static. The retail example considered in this work can be classified by using the definitions given in Section 2.1 as a capacitated multi-facility, multi-service multi-commodity, and multi-source problem with discrete location points and Euclidean distances. The chosen decision criteria are also later defined for this general category of location problems.

There are many different ways to combine all the aspects relating facilities, locations, and customers into a mathematical location model. The way to build these models becomes clearer in the following Section where some traditional types of location problems are presented.

3.2.1 Mathematical Formulation of Location Models

The most extensively studied and classical location problems are perhaps the p-median, p-center, warehouse, and covering problems. The p-median problem is to place p new facilities, called medians, on a network in order to minimize the weighted sum of distances from each demand node to the nearest facility. The formulation of the p-center problem is almost identical, except that the objective is this time to minimize the maximum distance between a demand node and the closest facility. This approach is useful when it is important to make the longest customer-facility distance as short as possible, as is the case, e.g., in placing emergency facilities. In warehouse problems, the aim is to find the best sites for intermediate stocking points, i.e. warehouses, while planning distribution systems. In covering problems, the goal is to "cover" as many customers as possible and to capture their demand by placing facilities at locations that provide potential users with optimal access to service facilities in terms of distance or travel time. Certain covering models can be adapted to the settings of retail location problems; these models will be formulated shortly in this Section in order to give the reader an impression of what the optimization problem will actually look like.

In order to represent the models mathematically, the following notation will be used:

I : set of facility location candidates

J : set of facility customers

x_j : demand of customer j

d_{ij} : distance between customer j and the candidate location i

y_i : binary decision variable which takes the value 1 or 0 according to whether a facility is established at a location $i \in \hat{I}$ or not.

Facility location candidates are denoted with $\hat{I}$ since I is later used to represent the set of chosen locations. The listed variables form in fact the basic elements of most location problems although the way to combine them together varies in different types of location models.

3.2.2 On the Solution of Location Problems

In general, facility location problems belong to the class of NP-hard problems. As a consequence, exact solution methods are restricted to small problems and many heuristics have been proposed to find quickly approximate solutions for realize problems. These heuristics and meta-heuristics include for example tabu search procedures, simulated annealing, dual-based heuristics, scatter search, and evolutionary algorithms. Each location problem type has naturally its own set of most popular heuristics.

The location optimizer developed at Fraunhofer AIS makes use of evolutionary algorithms that can be used to cover large search areas. Moreover, the solution efficiency can be easily improved as the knowledge of the problem domain increases. The optimizer can be used to provide solutions to different scenarios; in these scenarios, for example the budget, the assumptions concerning market development, or the type of the objective function can be varied. The solutions provided by the optimizer consist of the number, types, and locations of facilities and of the corresponding customer flows. These attributes will be denoted in this Thesis as the basic solution properties. In addition, there are several so called derived properties concerning for example costs, profit, the amount of satisfied demand, and customer coverage.

Evolutionary algorithms normally provide a set of solutions ranked by their fitness. Combined with different scenarios, dozens, even hundreds of location solutions can thus be easily obtained. It is important to be able to compare these options for several reasons. The first reason is that the solutions can have very different properties although they all had been generated by using the same scenario. Usually, the objective function consists of several weighted sub-functions and the values of these sub-functions can therefore be different in each solution. Moreover, the retailer might also have a preference concerning some attributes that are not included in the constraints or in the objective function. On the other hand, finding a good problem definition (i.e. a scenario) can prove to be important in guiding the optimization process towards a good result; it is thus useful to compare the qualities of solutions created by using different scenarios. It might also be uncertain what kind of parameter is the most realistic one (there can be, e.g., different estimates for market growth) and it would be useful to compare to what extent the optimal solutions differ for the alternative scenarios.

3.3 Defining Decision Criteria

In this Chapter, several decision criteria are proposed for retail location problems. The attributes are first formulated globally, that is for the total area under consideration. This formulation is subsequently developed further in order to express the attributes for regions that constitute the total area. Other possible decision criteria that are not included are considered shortly in Section . The Chapter begins with the definition of available input data and output obtained from the optimizer. These data are the main source exploited in defining the decision criteria that are explored after defining the retailers' major objectives.

3.3.1 Input and Output Data

No matter how the actual location optimization problem is formulated, there are almost always certain common input variables. Respectively, there is also a certain number of attributes that are generally obtained as an output from the optimizer. Some input and output variables that are frequently encountered in retail and geo-marketing applications are represented. These variables are also relevant in our case study. The usual characteristics of the variables (i.e. whether they are region. service. or facility type dependent) in the case study are also listed as an example. Naturally, the input and output are dependent on the chosen location model; besides the variables represented, there could also be, for example, data on the transportation costs. On the other hand, not all the attributes defined might be relevant in every situation. In addition, the form and the coordinates of the total area and of its sub-regions must also be known.

3.3.2 Defining the Objectives

According to a research most location problems can be classified into four general categories on the basis of their objective function:

(1) Cost minimization
(2) Demand oriented
(3) Profit maximization
(4) Environmental concern

These categories are not completely separate; for example, cost minimization and demand satisfaction can be viewed as surrogates for profit maximization. Cost minimization is the most traditional of these approaches, and it is often formulated as the minimization of some type of distance which represents the transportation costs. In demand oriented approaches, the aim is to maximize the amount of satisfied demand and the objectives are typically based on some measure of proximity or accessibility to a facility in the form of, for example, coverage, demand assignment, or other demand measurement. Although profit would seem to be the most natural objective at least for private sector enterprises, it is not very often used as an objective. According to Current et al., this may be because demand satisfaction and cost minimization are viewed as strategies to identify profit maximizing locations. In trying to maximize profit, there could, however, be a trade-off between cost and demand: the more we minimize the costs, the less we succeed in satisfying the demand and vice versa.

In retail applications, the three first objectives (cost, demand, and profit) can perhaps be assumed to be more preoccupying than environmental considerations. Besides these most common criteria, for example market share and visibility could be included in the goals of the optimization. Further attributes can then be developed to measure each of these objectives when necessary in the form of value tree analysis often used in multiple criteria decision analysis (MCDA).

Profitability is a natural objective at least for all private enterprises. Profit can usually be calculated directly when the amount of satisfied demand is known. Maximizing the demand satisfied by the retailer's services is another natural objective. The potential demand could also be considered as a decision criterion: the greater the demand potential within an area, the more interesting it is to establish facilities there. However, looking only at the retail potential leaves something unconsidered: the established facilities might not have enough capacity to respond to the entire demand. In the case of capacitated location problems, it seems therefore more suitable to use a criterion such as unassigned or satisfied demand.

Minimizing the costs of establishing and of running facilities is another obvious objective. The costs can usually be divided into variable and fixed costs or, in some cases, into one-time and recurring costs. Maximizing the visibility of the retailer (or of the brand) in the area could also be considered relevant. Although visibility is usually managed with different marketing strategies, it could also be important that the facilities are well placed and seen by as many potential users as possible so that they "remind" the customers of their existence. Measuring visibility is not quite straightforward since it includes factors concerning infrastructure and demography; as a simplified measure, the number of facilities or their density could perhaps be used. The number of customers could perhaps also be included as an attribute for this objective in order to include aspects such as word-of-mouth advertising.

Increasing the market share of the retailer or of the brand is in many cases a major objective. When the current market situation is known, market share can be calculated directly, or, when there is a lack of information, the number of new purchases could perhaps be used as a surrogate. The number of purchases can in general be deduced from the amount of satisfied demand. There is also another possible objective that has been left out from Figure 1: the aim of the retailer could also be to beat the competitor(s). The objective could be considered achieved when, for example, a market share larger than that of the competitor's is obtained, or, in the extreme, when the competitor is forced out of the market. This competitive objective could also be considered to be encompassed by market share maximization.

There are also other possible objectives, for instance the accessibility of the facilities to the customers. This could however be considered as a sub-objective in trying to maximize demand. This and other possible decision criteria are developed in more detail in the following Section.

3.3.3 Global Decision Criteria

The decision attributes developed here can be divided into the three following categories: monetary, facility related, and customer related attributes. This classification is used since it is somewhat less ambiguous than a classification under the overlapping main objectives (profit, demand, cost, visibility, and market share) would be. It is assumed that there is more than one service and the decision criteria are thus expressed specifically for each service type when possible. There are also m different types of costs (e.g. opening, change of facility type, and closure costs).

The following notation will be used to describe the input variables.

L : set of services

T : set of facility types

x_j^l : demand of customer j for service l

s_i^l : supply capacity of a facility located at site i for service l

c_i^k : cost of type k for a facility at location i, $k \in K = \{1,...m\}$

g_i^l : financial gain obtained from satisfying a unit of demand for service l by a facility located at i

The output variables obtained from the optimization are defined as:

$I \subseteq I$: set of chosen facility locations

w_{ij}^l : fraction of customer j's demand for service l provided by facility i, $0 \leq w_{ij}^l \leq 1$

w_{0j}^l : fraction of customer j's demand for service l not provided by any facility, $0 \leq w_{0j}^l \leq 1$

It must naturally hold that

$$w_{0j}^l + \sum_{i \in I} w_{ij}^l = 1 \quad \forall j \in J \quad \forall l \in L \qquad (3.2)$$

It is implicitly assumed that there can be different types of facilities and that su pply capacity, gain, and costs can depend on them; for example, $g_i^l = g_i^l(t_i)$ where t_i is the type of the facility located at site i. In this formulation, there are only fixed costs and e.g. no transportation costs to be accounted for. The facility locations candidates are discrete as well as the actual facility locations $i \in I$ i chosen in the optimization. In this formulation, there can only be one facility at a given site; the facility located at site i can therefore be simply denoted as facility i. In Section 3.4, where regional decision criteria are formulated, it is naturally assumed that more than one facility can be placed in a given region.

3.3.4 Regional Decision Criteria

If it is assumed that the total area is subdivided into regions, the previously defined criteria can be calculated separately for each of them; there is also a new decision criterion describing customer movement from one region to another. The global decision criteria defined for the total area are assumed to be a sum over the corresponding regional properties. We define

R : set of regions

$F(r) \subset I$: set of facility locations in region $r \in R$.

$C(r) \subset J$: set of customers located in region $r \in R$.

The regions are assumed to be disjoint:

$$r_i \cap r_j = \phi \quad r_i, r_j \in R \quad \forall i, j, i \neq j \tag{3.3}$$

The formulas presented in this Section are to be interpreted from a facility-centered point of view: the facilities are always supposed to be located in the region under consideration whereas the customers can also come from neighboring regions. The situation could naturally also be turned around and investigated from the customers' viewpoint. The facility-centered approach has been chosen in order to be able to inspect, e.g the profit and cost per region.

The regional cost can be expressed as

$$\sum_{i \in F(r)} \sum_{k \in K} c_i^k \tag{3.4}$$

and the regional gain for service l is respectively

$$\sum_{i \in F(r)} \sum_{j \in J} g_i^l w_{ij}^l x_i^l \tag{3.5}$$

Now the profit for region r is obtained as

$$\sum_{i \in F(r)} \left(\sum_{j \in J} \sum_{l \in L} g_i^l w_{ij}^l d_i^l - \sum_{k \in K} c_i^k \right) \tag{3.6}$$

The budget can naturally be expressed separately for each region only if it has been defined as a regional variable in the input data.

When the total area consists of regions, customer movement can be used to describe the amount of customer demand that is satisfied outside the customers' home regions. When the directions of this movement are known, it can be determined which facilities or regions are particularly attractive to customers. The amount of "import," that is the demand coming from other regions to region $r \in R$, can be expressed mathematically as

$$\sum_{i \in F(r)} \sum_{j \notin C(r)} w_{ij}^{l} x_{i}^{l} \tag{3.7}$$

Respectively, the "export" from region r can be expressed as

$$\sum_{i \notin F(r)} \sum_{j \in C(r)} w_{ij}^{l} x_{i}^{l} \tag{3.8}$$

The amount of demand satisfied by the facilities in region r is

$$\sum_{i \in F(r)} \sum_{j \in C(r)} w_{ij}^{l} x_{i}^{l} \tag{3.9}$$

and unassigned demand (potential demand – satisfied demand) can be formulated as

$$\sum_{j \in C(r)} w_{0j}^{l} x_{i}^{l} \tag{3.10}$$

Customer movement could be included as a decision criterion if the aim were, for example, to increase the attractiveness of a specific region. The distance customers-facilities can also be calculated region specifically by looking at the distance between the facilities in a certain region and the customers using these facilities (the customers might naturally come from other regions as well).

The capacity surplus/deficit can be formulated for a region $r \in R$ as

$$\sum_{i \in F(r)} \left(s_{i}^{l} - \sum_{j \in J} w_{ij}^{l} x_{i}^{l} \right) - \sum_{j \in C(r)} w_{0j}^{l} x_{i}^{l} \tag{3.11}$$

The amount of facility dispersion can also be calculated separately for each region using the chosen dispersion measure. Calculating the competition per region could however be more complicated since the situation is also influenced by competition in neighboring regions.

3.4 Further Suggestions for Decision Criteria

There are also many other factors that may be of interest in evaluating the quality of a location solution: these factors could describe for instance the compatibility with demographic structure and existing infrastructure or characterize the suitability of the solution considering the growth rate of competitors, market saturation, and so forth.

Such factors have not been included in the defined decision criteria for various reasons. The first reason is that many aspects can be taken into consideration in estimating the potential demand: for example regional demand values are influenced by the regions' demographic structure as well as by market saturation. Many of the chosen decision criteria are in turn dependent on the potential demand. Taking the infrastructure into consideration might be very useful but its influence is difficult to formulate if there is no specific application: some businesses might benefit from being situated in pedestrian zones whereas others would profit from locating in big shopping malls etc. On the other hand, the effects of infrastructure might also be taken into account in determining the demand function or the potential locations. The competitors' growth rate might also be of interest but it can be difficult to estimate, especially when the current competition situation changes as a retailer establishes new facilities.

3.5 Making Multi criteria Decisions with Spatial Data

In addition to pure visualization methods, various decision support tools can be of great help when comparing the properties of several location options. Most importantly, multiple criteria methods allow to weigh several conflicting decision criteria simultaneously. In general, MCDA methods calculate an evaluation score in order to obtain a ranking of the options. Let us define

x : an option

n : number of decision criteria $\rightarrow n$ attributes for each option $x = (x_1, ... x_n)$

ω_k : weight assigned to attribute k

$v_k(x_k)$: value (or score) of option x in regard to attribute k

Decision support methods calculate a score $V(x)$ for each option x usually by using an additive model where the total score is a weighted sum of attribute specific scores:

$$V(x) = \sum_{k=1}^{n} \omega_k v_k(x_k) \tag{3.12}$$

In dealing with regional data, it would be interesting to find the best solution for each region according to some decision criteria; this requires including the regional dimension of the attributes in the scoring process. If the weights of the attributes are assumed to be equal in all the regions, i.e. $\omega_{kr} = \omega_k \forall r \in R$, the evaluation score $V_r(x)$ for region r can be calculated as

$$V_r(x) = \sum_{k=1}^{n} \omega_k v_{kr}(x_{kr}) \tag{3.13}$$

where $v_{kr}(x_{kr})$ denotes the value of attribute k in region r for option x. The regional approach is also important since using the global solution attributes as decision criteria would mean losing the spatial dimension of data.

Were we dealing with a single site selection problem, the approach would be relatively straightforward: a scoring method could be used to rank the candidate locations, in this case regions, by assigning weights to their decision attributes and dynamic displays would allow to visualize how changes in the weights affect the outcome. In the previous formulation, the location candidates would represent the options x. When comparing different location solutions obtained from an optimization process, the situation is however much more complicated: the options are no more single regions but sets of regions including the number and the types of facilities located in them. The score of a solution sol in region r is now obtained as

$$V_r(sol) = \sum_{k=1}^{n} \omega_k v_{kr}(sol_{kr}) \tag{3.14}$$

The scoring is a mapping of type $R \rightarrow IR$ and regional scores obtained for different solutions can thus be visualized by using methods. It could also be possible to visualize dynamically the effects of changing the weights of attributes by showing the change in the solution scores in each region. Besides only visualizing the regional scores, it is be possible to combine them into a global one: when using an additive model, this can be done by calculating the weighted sum of regional scores. The global score for a solution sol would thus be

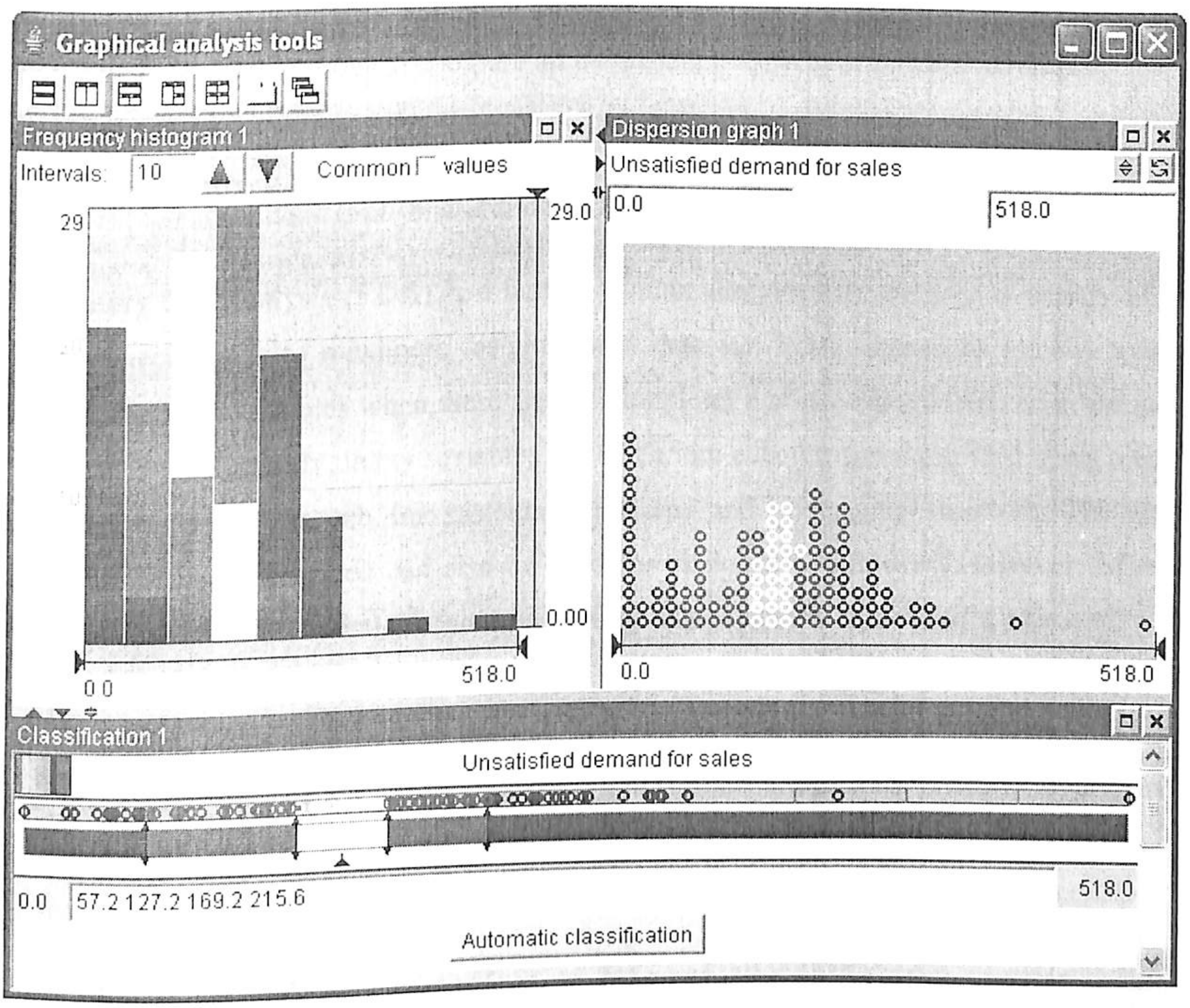
$$V(sol) = \sum_{r \in R} \omega_r V_r(sol) \qquad (3.15)$$

3.6 Multivariate Data Analysis Techniques

When analyzing large quantities of multi-dimensional data, techniques developed in data mining can often prove useful. For example, cluster analysis can be used to classify the data into disjoint, homogeneous groups and subgroup discovery can be used to search for statistically unusual datasets. On the other hand, highly dimensional data are often transformed into lower dimensional data where coherent patterns can be detected more clearly by applying a statistical method called principal component analysis (PCA). This Section provides a closer look at various data analysis methods, beginning with a short introduction to classical descriptive statistics. The presented methods can be used to support decision making when studying multi-relational data such as a large number of location options. The latter source also offers information on basic statistics and various data mining techniques.

3.6.1 Descriptive Statistics

Descriptive statistics can be used to characterize a large number of location solutions and their attributes. There are three major qualities that are usually examined: the distribution, the central tendency, and the dispersion of variables. To describe central tendency, mean, median, or mode can be calculated. Dispersion can be measured with standard deviation and variance, and the frequency distribution can be visualized with the help of histograms or bar charts. In addition, the quartiles, maximum and minimum values, skewness and kurtosis, etc. might also be calculated. In visualizing the data, various tools such as scatter plots, box-and-whiskers plots, stem-and-leaf plots, parallel co-ordinates plots, and correlation diagrams can be used. A few examples of graphical analysis tools from Common GIS can be seen in Figures. Some a bit more exotic representations include for example Andrews' plots and representing multivariate data as Chernoff faces.



The available methods depend on the type of variables investigated and the selection is naturally more limited for nominal attributes than for ordinal, interval, or ratio variables. The interactions of variables can also be examined for instance by calculating their correlations. There are also several methods of outlier detection that can help to test the uniformity of a set of solutions and also the quality of the optimizer. In dealing with a large number of solutions, it could also be beneficial to use data tables: after placing the solutions as rows and their characteristics as columns (this could also include an indicator for the scenario), a sorting method could be used to view the best (or worst) solutions according to some criterion. More information and examples on visualizing multivariate data can be found.

One setback of using non-cartographical displays is that they do not necessarily offer a very intuitive representation of the situation. Another problem is the large number of attributes that have to be examined in order to compare all the different aspects of location solutions.

3.6.2 Exploratory Data Analysis

Exploratory data analysis (EDA) is a branch of data analysis that employs a variety of mostly graphical techniques to maximize insight into a data set. EDA is used to identify systematic relations between variables when there are no (complete) a priori expectations as to the nature of those relations. Typically, many variables are taken into account and compared using a selection of techniques in the search for systematic patterns and underlying structure. The graphical statistical methods described can also be used for exploratory purposes. However, EDA is not merely a collection of statistical and data mining techniques; it is more of a philosophy to data analysis that postpones the usual assumptions about what kind of model the data will follow and allows the data itself to reveal its underlying structure. The seminal work in the field is John Tukey's Exploratory Data Analysis; among more recent publications, Keim's paper provides a good summary of different visual techniques.

Computational exploratory data analysis methods include both simple basic statistics and more advanced exploratory techniques designed to identify patterns in multivariate data sets. The basic statistical exploration methods include, for instance, examining the distributions of variables, reviewing large correlation matrices for coefficients that meet certain thresholds, or examining multi-way frequency tables. Multivariate exploratory techniques designed to identify patterns in multivariate data sets comprehend for example cluster analysis, factor analysis, discriminate analysis, multidimensional scaling, linear and nonlinear regression, correspondence analysis, time series analysis, and classification trees.

A large selection of powerful exploratory data analytic techniques is also offered by graphical data visualization methods that can help to identify relations, trends, and outliers hidden in unstructured data sets. Perhaps the most common and historically first widely used technique is called brushing which is an interactive method allowing to select specific data points or subsets of data and to identify their characteristics or to examine their effects on relations between relevant variables. Other graphical exploratory techniques include, e.g., function fitting and plotting, data smoothing, overlaying and merging of multiple displays, categorizing data, splitting or merging subsets of data in graphs, aggregating data in graphs, identifying and marking subsets of data that meet specific conditions, shading, and plotting confidence intervals and confidence areas.

3.6.3 Cluster Analysis

Cluster analysis is a procedure for identifying groups of instances that are as homogeneous as possible within the groups and as different as possible from other groups. This methods is used in order to obtain a relatively small number of (in the ideal case) homogeneous groups instead of a large number of separate instances. Clusters are not usually allowed to overlap and each instance in the instance space has to be assigned to a cluster, that is the clusters have to be mutually exclusive and exhaustive. The set of clusters represents a partitioning of the instance space on the basis of two or more classification variables. The similarity or distance function measuring the likeness of the instances plays therefore an important part in the analysis.

Cluster analysis has been formulated mathematically as follows. We define an instance space X and a set of instances $S \subseteq X$. Let q be a quality function:

$$q : X \rightarrow IR\,()$$

Given S and q, the task of cluster analysis is to find a clustering

$$C = \{C_1, \ldots C_k\} \quad \text{where } C_i \subseteq S \quad \forall i = 1, \ldots k \; ()$$

so that $q(C)$ is maximized and (optionally)

$$C_i \cap C_j = \phi \qquad \forall i, j \in \{1, \ldots, k\}, i \neq j \text{ and } \bigcup_{i=1}^{k} C_i = S \; ()$$

The quality function q is often defined on the basis of some kind of distance function:

$$\text{dist}: X \times X \rightarrow IR^{+}$$

There are two main types of clustering algorithms, namely hierarchical and partitional clustering. Hierarchical clustering techniques can be divided further into agglomerative and divisive clustering methods. In some algorithms, the number of clusters has to be fixed beforehand, in others it gets determined by the algorithm during the clustering process. More information on different clustering methods can be found for instance in.

The definition of a distance function dist is essential in clustering in order to measure the dissimilarity between patterns. Examples of different types of distance measures for measuring the distance between two instances include for example the Minkowski (or Lp) distances, the statistical distance, and the Mahlanobis distance. The most popular metric for continuous space remains the Euclidean distance which is a special case of the Minkowski distances. Once a distance measure between individual instances has been chosen, one must also define a way to measure the distance between clusters or groups of data. Methods used in hierarchical clustering include for example the centroid method, nearest-neighbor method, furthest-neighbors method, average linkage method, and Ward's method. As opposed to a distance measure, a similarity function could also be defined. Similarity measures are usually viewed as being conceptually equivalent to distance measures; however, it is more common to consider the dissimilarity between two patterns by using a distance measure. Distance functions can also combine qualitative and quantitative differences.

Cluster analysis is often used for example in marketing to partition consumers into market segments on the basis of customer surveys. Cluster analysis is also used as a tool in scientific data exploration, information retrieval and text mining, image segmentation, computational biology, object recognition, and in many others fields. Ideally, the clustering procedure should not only provide us with a list of the subgroups and of their instances but also a compact description that characterizes the objects in each cluster.

3.6.4 Discriminate Analysis

Discriminate analysis (DA) can be used to determine which variables discriminate between two or more mutually exclusive groups. The discriminating variables are determined by testing whether the means for a variable are significantly different in different groups as is also done in using one-way or multivariate analysis of variance (ANOVA/MANOVA). Many different variables are usually included in the study in order to see which ones contribute the most to the discrimination between groups. After finalizing the model and finding the discriminating variables, DA can be further used to predict group membership and to classify new cases. In addition, discriminate analysis can be useful in evaluating the quality of clustering solutions.

In DA, the aim is to form a set of discriminate functions from variables describing the data. This is done by inspecting the matrix of total variances and co-variances and the matrix of pooled within-group variances and co-variances. The two matrices are compared via multivariate F tests to determine whether there are any significant differences between groups in regard to all the variables. If there are statistically significant differences between the matrices, we can proceed to see which of the variables have significantly different means across the groups. DA determines an optimal combination of variables so that the first discriminate function provides the most overall discrimination between groups, the second provides the second most, and so on. Moreover, the functions will be independent or orthogonal, or in other words, their contributions to the discrimination between groups will not overlap.

Discriminate functions can be interpreted by using standardized coefficients which are given for each variable in each discriminate function. The larger the standardized coefficient, the greater is the contribution of the respective variable to the discrimination between groups. Another way to determine which variables define a particular discriminate function is to look at the factor structure matrix which contains the correlations between the variables and the discriminate functions. The standardized coefficients do not tell between which groups a discriminate function is the most of the least discriminating; for this purpose, the means of the discriminate functions across groups have to be examined. In a spatial interpretation, means are centroids and differences in the locations of centroids show the dimensions along which the groups differ.

Multiple discriminate analyses is computationally very similar to MANOVA, and all assumptions of MANOVA (e.g. normal distribution of variables and homogeneity of variances between groups) apply; however, the violation of these assumptions does not always seem to be dangerous. Logistic regression answers basically the same questions as discriminate analysis and it is sometimes preferred to discriminate analysis since it is more flexible in its assumptions and in the types of data that can be analyzed: it can handle both categorical and continuous variables and the predictors do not have to be normally distributed or of equal variance within each group. A comparison of different statistical methods can be found.

3.7 New Retail outlet-Location Models

The following is a retail-outlet location model. Using the relevant literature on Fuzzy systems. Much of the data used can be collected using questionnaires, publicly available information and interviews. The Primary Trade Area is defined as an x-y mile radius around the selected site; and the Secondary Trade Area is defined as an a-b mile radius around the selected site (these demarcations can be changed depending on the product/service).

Let,

- P = No of individuals having vehicle.

- O_i = monthly operating costs at location i.
- S_i = minimum monthly net-sales at location i.
- CF_i = minimum monthly net-cash-flow to be earned at outlet at location i.
- R_i = minimum monthly rent at location i.
- T_c = no of vehicles that pass the location each day.
- ρ_p = average probability of product returns.
- B_c = percent of vehicle traffic that will purchase items from outlet.
- E_c = percent of vehicle traffic that will enter the outlet.
- S_a = average purchase per customer.
- F_c = total square footage of all direct-competitor retail outlet in the trade area.
- F_a = square footage of proposed outlet at location i.
- $F_{average}$ = average square footage of all competing outlets in the trade area.
- S_{fb} = average sales for existing competing retail-outlets before proposed retail-outlet is built.
- S_{fa} = average sales for existing competing retail-outlets after proposed retail-outlet is built.
- S_{fr} = average sales for existing outlets in the region/state.
- S_c = total actual sales for all direct-competitor retail outlet in the Primary Trade Area.
- S_e = total estimated sales for all direct-competitor retail outlet during the next three years.

MU = Migrative-Use factor, which measures the extent to which people aged 18 years or more, in communities in the Primary and Secondary Trade Areas migrate out of their own communities and the Trade Areas, between 6am-10pm, in order to shop, spend leisure time or work in other areas. It also measures the characteristics and buying propensity of the population in the Trade area. $0 < MU < 1$. $MU \rightarrow 0$, as a higher percentage of percentage of the population in the Primary and Secondary Trade Areas aged 18 years or more, a) work outside the Primary Trade Area, b) spend leisure time outside the Primary Trade Area, c) shop outside the Primary Trade Area.

TS = Time factor, which measures the impact of availability of time (or lack thereof) on shopping habits of people aged 18-65 years in the Primary Trade Area and the Secondary Trade Area. Time also has an important effect on propensity to shop in retail locations. $0 < TS < 1. TS \rightarrow 0$, as 1) a higher percentage of the population in the Primary and Secondary Trade Areas aged 18 years or more, have more than X hours available for shopping and leisure each week.

X = pricing of products in the retail-outlet. $0 < X < 1. X \rightarrow 0$, as the retailer's prices become less than prices of competing retail retail-outlets in the Primary Trade Area and The Secondary Trade Area.

D = new development activity within two miles of the site, which will generate additional traffic to the immediate neigborhood. $0 < D < 1. D \rightarrow 0$, as probability of new development within two miles, increases.

V = variety in the retailer's retail-outlet. $0 < V < 1$. This factor measures the depth of variety in the retailer's retail-outlet. Some retailers sell a wider variety of products, and or have ancillary items such as ATM machines, restaurants, pharmacies, etc. Variety is also measured relative to competing retail-outlet, and also with respect to ambience of the retail-outlets. $V \rightarrow 0$, as variety in the retailer's retail-outlets increases.

A = attractiveness of retailer's retail-outlet. $0 < A < 1$. This measures relative attractiveness – retailer's brand, retail-outlet layout, quality of products, etc.. $A \rightarrow 0$, as attractiveness of retail-outlet increases.

M = proximity to major 'attractors' that exist or will exist within the next three years (such as the Central Business District, convention centers, universities, government institutions, shopping mall that sells other types of goods, etc.). $0 < M < 1. M \to 0$, as proximity to major attractors is greater.

T = measures attractiveness of traffic conditions to and from the location – measures traffic congestion, landscaping, presence/absence of offensive sites/images along the route, visibility of site from 500-900 meters in each direction etc.. $0 < T < 1$. T does not measure distance. $T \to 0$, as attractiveness of traffic conditions increase.

IC = the retailer's incremental costs of building/opening the physical retail-outlet. $0 < IC < 1. IC \to 0$, as the retailer's incremental costs of building the physical retail-outlet becomes greater.

U = community resistance to selection of site. This index measures resistance from municipal authorities and residents to selection of the site to for use as a retail outlet location. Such resistance results in additional costs that make the site more expensive and may negatively impact post-occupancy sales – the costs include litigation costs, public hearings, environmental impact studies, community meeting, impact fees, etc. $0 < U < 1. U \to 0$, as the costs of resistance increases.

- Brand Factor (B):

$$B = e^{-(\alpha T + \beta A + \lambda V + \mu M + \pi X + \theta D + (\delta * TS) + (\varsigma * MU))} \quad (3.15)$$

$$0 < B < 1.$$

$$\alpha, \beta, \lambda, \mu, \pi, \theta, \delta, \varsigma \in (0,1)$$

$\alpha, \beta, \lambda, \mu, \pi, \theta, \delta, \varsigma$ are weights attached to T, A, V, M, X, D, TS, MU respectively to reflect their relative importance with respect to each other. The natural log scale is used because it reflects the change in scale as various factors changes. Factors such as attractiveness, Variety, pricing refer to qualities of the retailer and its brand relative to competing retail outlet in the trade area.

At each given potential site, the retailer has at least three alternatives, each of which results in drastically different capital commitments for the retailer:

Alternative 1:-

Purchase land and then build a retail-outlet – the retailer's capital commitment will be the sum of the following:

R_b = monthly interest for amount borrowed

E = equity invested

T_t = monthly amortization of transaction costs (amortized over time t)

$P_{c1} = (R_b + E + T_t)$

Alternative 2:-

Purchase land and building – the retailer's capital commitment will be the sum of the following:

R_b = monthly interest for amount borrowed

E = equity invested

T_t = monthly amortization of transaction costs (amortized over time t)

$P_{c2} = (R_b + E + T_t)$

Alternative 3:-

Have a developer purchase/lease land and build a retail-outlet, and then lease the property from the developer – the retailer's capital commitment will be the monthly lease payments over time t plus the amortized transaction costs (where $t \in T$, and t is the average number of months that the retailer stays at each site, and T is the total lease term (typically 20-30 years), the sum of which is denoted L_d.

Alternative 4:-

Lease an existing retail-outlet - the retailer's capital commitment will be the monthly lease payments over time t, plus the amortized transaction costs (where $t \in T$, and t is the average number of months that the retailer stays at each site, and T is the total lease term (typically 20-30 years), the sum of which is denoted L_d.

Its assumed that taxes, utilities and insurance costs are the same under property rights. Then,typically:

$P_{c1} > P_{c2} > L_1 > L_d$

$L_1 \approx L_0$

$L_d > L_0$

The component of the site selection decision that concerns property rights, results in an incremental cost

$\psi = \{Max(P_{c1}, P_{c2}, L_1, L_d) - L_0\}$

where L_0 is the normal lease payment which is already reflected in O_i, the operating expenses.

The retailer's multi-criteria objective function in site-selection will be:

$$Max\left[S_a \Omega (1-\rho_p)(T_c B_c E_c)^* 30^* e^{-\{((S_c+S_e)/S_{ta})^* MU\}}\right]$$

$$Min[Min(\{O_i + OV_i + CF_i + \psi\},0)]$$

Hence, the site-selection decision, is also calculated as whether the site meets certain conditions:

1. $\left[S_a \Omega (1-\rho_p)(T_c B_c E_c)^* 30^* e^{-\{((S_c+S_e)/S_{ta})^* MU\}}\right] > Max[S_i > \{O_i + OV_i + CF_i + \psi\},0]$
2. $[S_a \Omega (1-\rho_p)(T_c B_c E_c)^* 30^* \{L_n ((S_{ta}/(S_c+S_e))^* MU)\} * B * U] > Max[S_i > \{O_i + OV_i + CF_i + \psi\},0]$
3. $Max[\{O_i + OV_i + R_i + \psi Y\},0] < S_i$
4. $B > 0.5$

5. $(\partial S_c / \partial S_{ta}) > 1$

6. $(\partial S_c / \partial P) > 1$

7. $(\partial S_{ta} / \partial F_c) > 1$

8. $(\partial S_{fh} / \partial S_{fr}) > 1$

9. $(\partial S_{fh} / \partial S_{fr}) > Max[(\partial S_c / S_{ta}), 1]$

CHAPTER 4 CONSTRUCTION A LOCATION MODEL

4.1 INTRODUCTION

In spite of the heterogeneity observed in different European countries generally the retail sector are going in restructuring phase for such factors as increasing consumer mobility increasing electronic commerce. changing house hold size, concentration of market power, and changes in planning legislation to justify new trends in retailing. The change in the consumer's behavior and the fact that consumers are more demanding force the retail groups to invest strongly at stores, betting in the proximity and quality of goods and services strategy.

Fig-1 represents a possible schema for different levels of retail network location decisions. We considered a phase decision methodology, although the interaction among the different phase is possible. especially between the services offered and store location. Studies based on enquiries, indicate that the decision maker mind separates decision about choosing the geographical regions and choosing store locations within the region.

In first phase, based on expansion strategic policy for the network is selected a zone or region where new stores will be installed, the number of units to be built, along with implementation timings. After the selection of some alternative locations is made, usually with the help of real estate agencies it is intended in a second phase to accomplish the final store site choice. For this decision level the number of commercial plus academically published models suggests the researchers feel a need or rational and formula use of information.

The third decision level concerning store and services design is the level most linked to the concept of services quality and customer satisfaction as mentioned. The physical design of the services facility has an important role to play.

Retailers have always understood location as paramount but understanding all aspects of store performance site potential in addition to consumer behavior demands a great amount of information. including geographical, demographic, socioeconomic, and competition data.

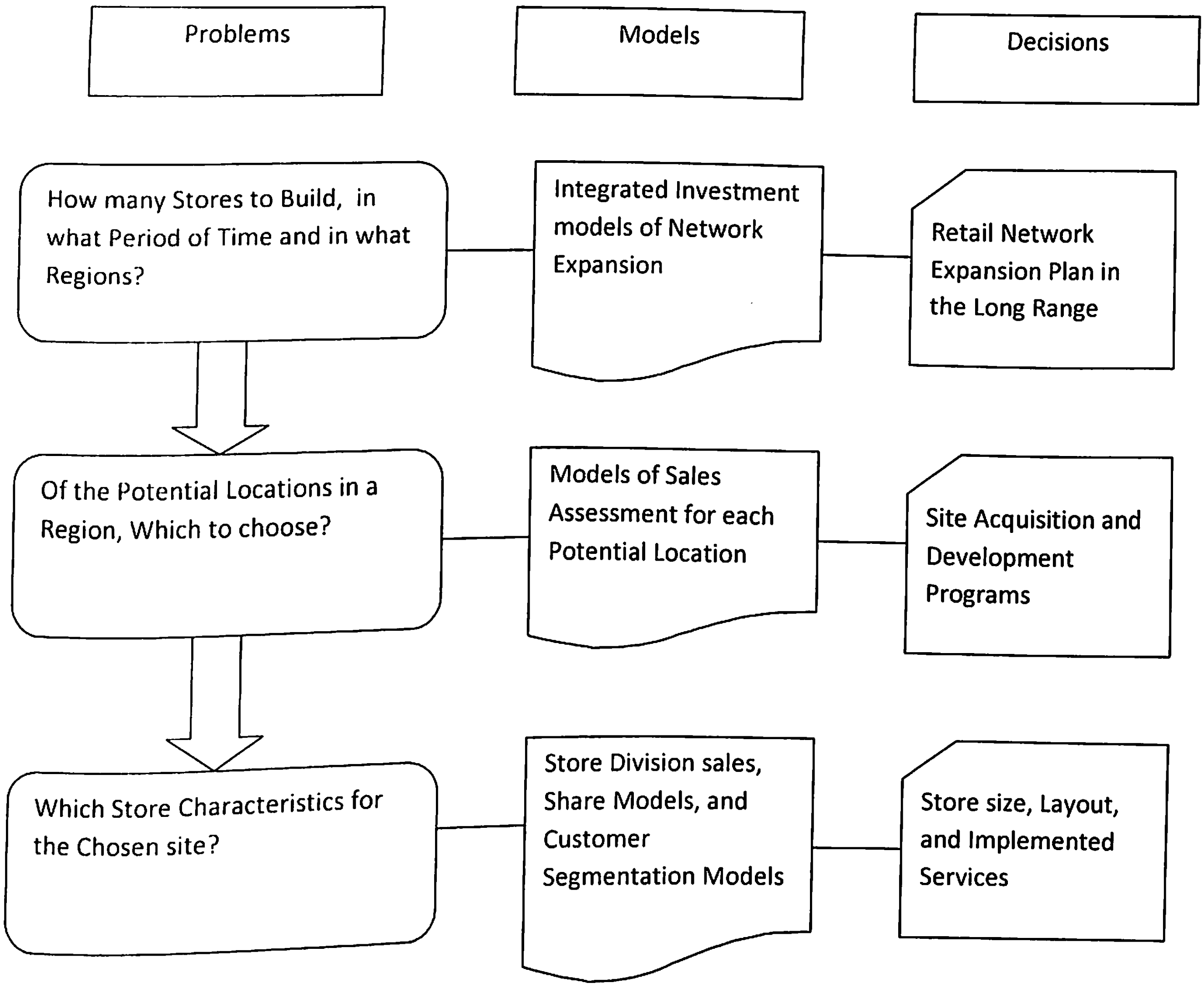


Figure 1Decision levels involved in Retail-Outlet Network

Figure-2 suggests a possible classification of assessment location explanatory variables. In opposition to trade area evaluation, the site and store variables intend to evaluate internal factors or the store offer. Store size is traditionally singled out as the most important outlet characteristics, as an independent branch of the schema outlines.

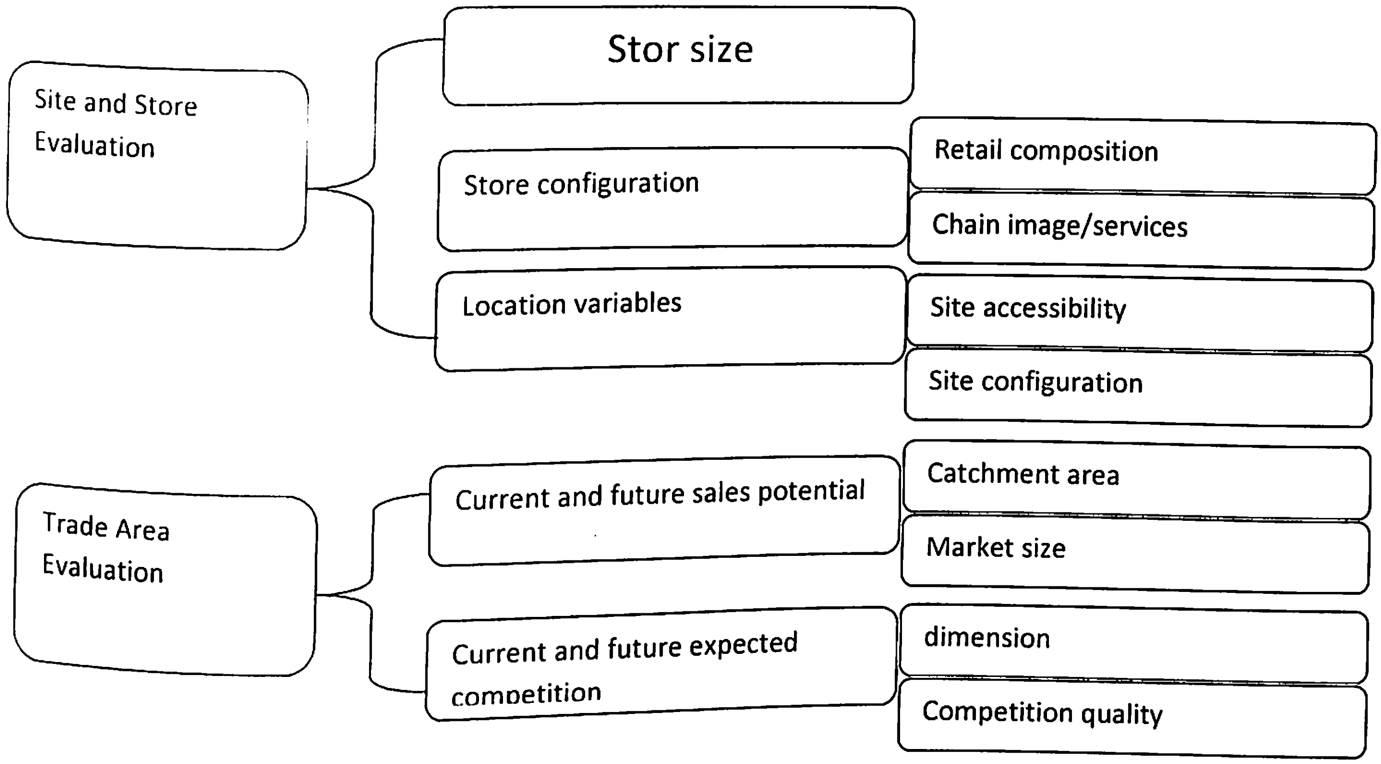


Figure2 Classification of Assessment Location Explanatory Variables

4.2 CLASSICAL PROCEDURES:

Development of accurate sales forecast is central to successful retail site selection. However, empirical methods of retail forecasting suffered from an excessive subjectivity of analysis, as well as an inability to consider simultaneously the impacts of multiple variables. Progressively statistical sales forecasting methods contributed to define the relationship between store sales and influential variables, such as site, demographics along with competitive characteristics. Pastor also emphasizes the importance of new techniques focusing both on finding optimal sites for new outlets and forecasting sales, based o objective criteria.

In the middle 80's a medium UK supermarket chain, named Tesco PLC decisively invested in a remarkably successful project aimed at forecasting store turnover with only 10% mean error. The need of such an expensive and long study is justified in the words of Atrtwell by the fundamentals need to make better site acquisition decisions, thus reducing the risk associated with multi-million-pound investments. They began with simple regression models, evolving later to gravitational models. The value of store location analysis is now well established in Tesco, and is considered a commercial advantage in the United Kingdom with 15.6% of the food market.

The techniques and models like analogue techniques and gravitational models which are heavily used and studied now-days in spite of been known for long ago. Te last three sections concentrate in newer promising approaches and trend on the problem of locating retail stores, on the trade area delimitation and on using GIS for location analysis.

4.3 ANALOGUE-BASED PROCEDURES:

The simplest site location technique involves rules of thumb used by location analysts that apply a combination of experience, empirical observation along with trial and error to isolate the key factors, which appear to affect directly the performance of a store location. One example is the checklist method referred into several attributes. A manager checks each factor separately for the proposed location and analogue stores used for constructing a profile of the strengths and weakness of the sites. Those profiles are then compared and a rating is produced based on rapid solution for a location decision. However, these simple rule based methods may be over simplistic and subjective.

The analogue-based forecast models are a natural outcome of the former attempt to overcome their lack of objectivity. Developed by Applebaum since the 30's analogue methods in their simpler form do not require many data are reasonably objective and allow the inclusion of the analysts business experience and intuition. These models are still used by companies when the chain's number of stores does not justify the development of more advanced models.

The methodology starts by classifying several analogue commercial spaces according to a group of store attributes as empirical benchmarks. Each store in the analogue group is evaluated, and the location-related factors of the new site are measured against a pre-defined scale. Using weights r other simple techniques, a store sorting list is ten produced which rates the new store relatively to the remaining ones. A .forecast interval for the sales of the new store is thus defined. This interval is as narrow as the annual sales difference between the stores rated immediately before and after the new store.

A large portugueses retail group uses this method . the procedure starts by identifying key variables that characterize the trade area. Those variables are then used in the identification of analogue areas. By considering the number of stores or the total floor space already installed in composition with the sales potential of the analogue areas it is possible to conclude it it has potential for the installation of new stores. This method is put in practice with the aid of GIS packages in addition to extensive database including stores of several chains controlled by the same and completing retail groups.

Recent projects explore a combination of intuitive insight with normative approaches to develop analogue decisions support systems. a recent example is the work presented by Clarke. These authors describe how the analogies identified by a qualitative system can be visualized effectively for use in location analysis. The use cognitive mapping and other visualization techniques for exploratory data analysis and decision support.

Analogue-based forecast models have been appearing in recent years, associated with statistical techniques. Example are the analogues based on regression referred by Simkin in his survey conducted among U.K multiple retailers, which at the time were the primary location procedure used and discriminate analysis. Another example is cluster analysis in analogue store group formation and market segmentation.

4.4 GRAVITATIONAL MODELS

Gravitational models are derived from the laws of Newtonian physics based on the balance between the store attractively and the distance to the potential customers. In the initial work of Reilly the Law of retail gravitation related the share f customers that an outlet attracts a being inversely proportional to distance they must travel and directly related to the store dimension. A similar formulation was pioneered by Huff to calculate a probability tat a customer patronizes a facility.

Gravitational models procedure start by the definition of the trade area of store location. This area is the divided in smaller zones of homogeneous demographic environments in addition to competitive characteristics respecting geographical obstacles like rivers, motorway lines, railways, etc. Each zone is then analyzed for the calculation of potential sales originating in that zone. The distribution of this potential is made by using equation for several sale points according to a distance function where the parameter actually determined by regression for analogue stores, reflects the sensitivity of customers to distance .

This mode evolved from store dimension to include a variety of other factors in the attraction concept as competition or demographic variables which could extend to the multiple retailer context. An example is the work presented by Staney and Sewall where multiple site or store variables characterizing the different stores of a network are aggregated in an attractively measure. These procedures are used to evaluate total network configurations in a market. It has also the advantage of explicitly incorporating distance among sales points as well as considering items as population and competition which in highly competitive markets can be decisive. They can also provide estimates of sales impacts on the construction layout modifications or amplification of sale pints on competition as cannibalism in sister stores. Therefore they seem uniquely suited for location strategies simulation in the long run.

4.5 ANALOGUE BASED REGRESSION MODELS:-

This is one of the most commonly applied store turnover forecasting model and Hermandez and Bennison. It was first applied to retail location analysis in the 1960's. it is particularly appropriate for retailers with highly segmented market appeals for example clothing retailers, restaurants, bookshop, and jewelers, but is widely used and misused in all areas of retail.

The variables are measured for existing stores, which are reasonably analogous to new locations in study. Those data are then used to calibrate a linear statistical equation. It's important to note that the equation includes just those variables, termed independent or predictor variables, which are found to be significantly correlated with store turnover. Their uses simply involve entering the measurements for a new site on the relevant predictor variables and compute the required estimate of store turnover.

An example of a dependence model with some complexity is the S.L.A.M model, Store Location Assessment Model, described by Simkin which uses addictive or multiplicative expression and many variables. In these models, demographic as well as competitive variables can be defined in relation to one or more zones of the new store retail trade area and different models can evaluate each zone.

In 1998 Quintino and Leitao present an example for the Portuguese retail gasoline market. Those authors obtained a total of 7 linear and multiplicative regression models one for a general case named basic model and six for the same number of service station segments. The estimates produced were more accurate than those made by previous models and are currently being used to support investment decisions of largest company. They also introduced the concept of anchor variables as a model selection criterion. For identical quality of fit models with better stability of the parameters of the anchor variables where favored in order to obtain marketing consistency and interpretability.

Regression techniques are also used in attraction modes parameter calibration that may be applied to multi outlet store assessment as the work presented by Achabal. The development of MULTIple LOCation (MULTILOC) model extends the Multivariate Compretitive Interactive (MCI) model to the multi store location problem utilizing a random search procedure combined with an interchange heuristic to identify optimal or near-optimal sets of location.

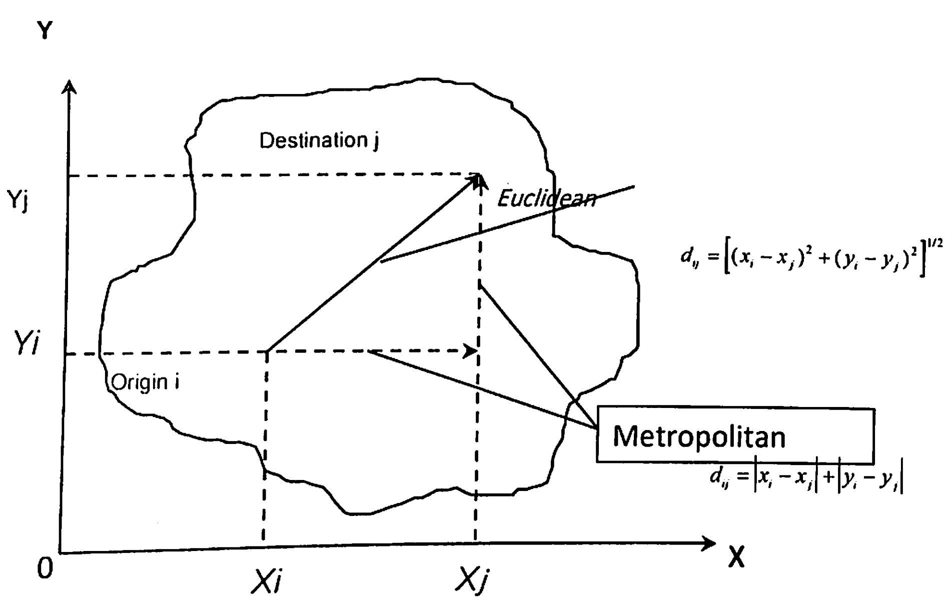
The main obstacle to the use of this method is the high number of necessary observations as much as more explanatory variables are included in the model and more segments are to be considered. The secret that surrounds the competition information makes it difficult to obtain the necessary number of observations. This explains why mainly companies with high number of own sale points have used these models. Regression techniques do not consider as an entire network but rather evaluate sites in isolation. They treat all customers within the trade the same way, regardless f their actual distance from the site and they evaluates specific sites but do not search for optimal ones. The difficulties in evaluating some factors in numerical scales is another weakness in applying regression or gravity models which can be overcome by applying discriminate procedures.

CHAPTER 5 FINDINGS AND ANALYSIS

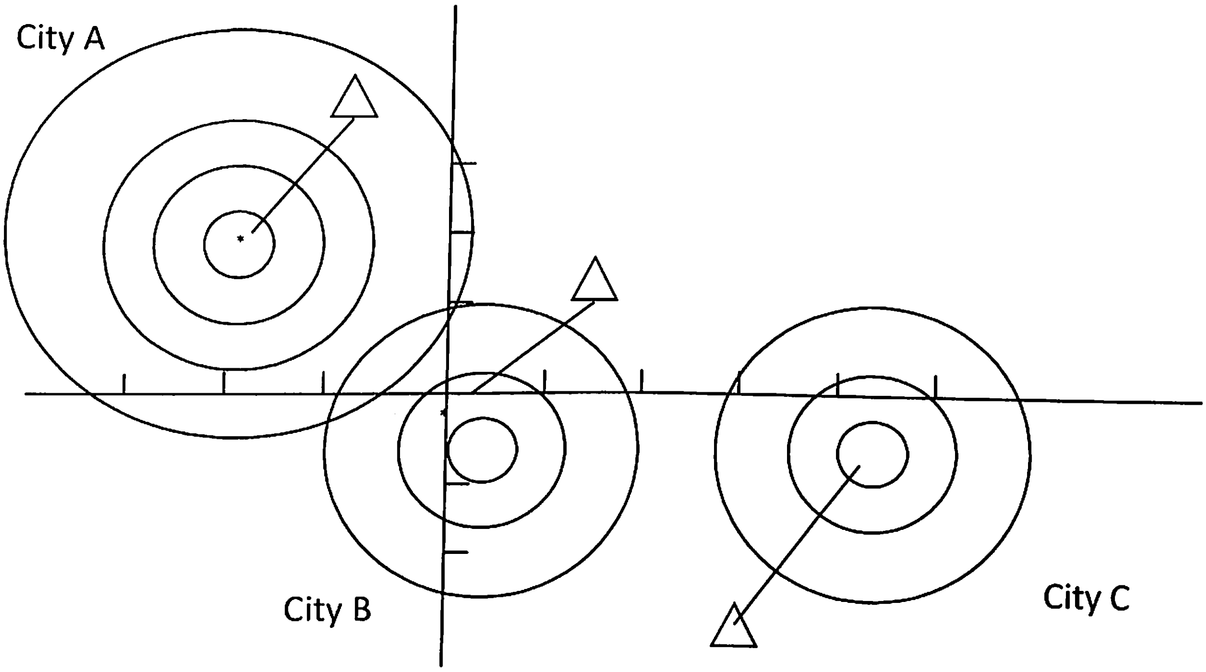
From the above discussion we obtained the flowing findings and analysis in the graphical representation.

Geographic Representation

Location on a Plane



Effect of Optimization Criteria



1. Maximize Utilization

(City C: elderly find distance a barrier)

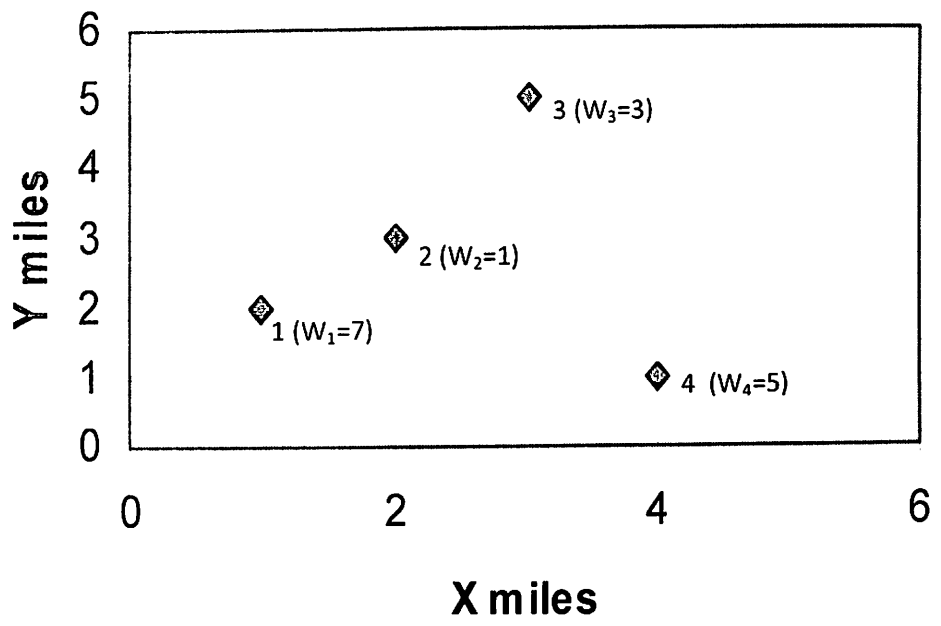
2. Minimize Distance per Capita

(City B: centrally located)

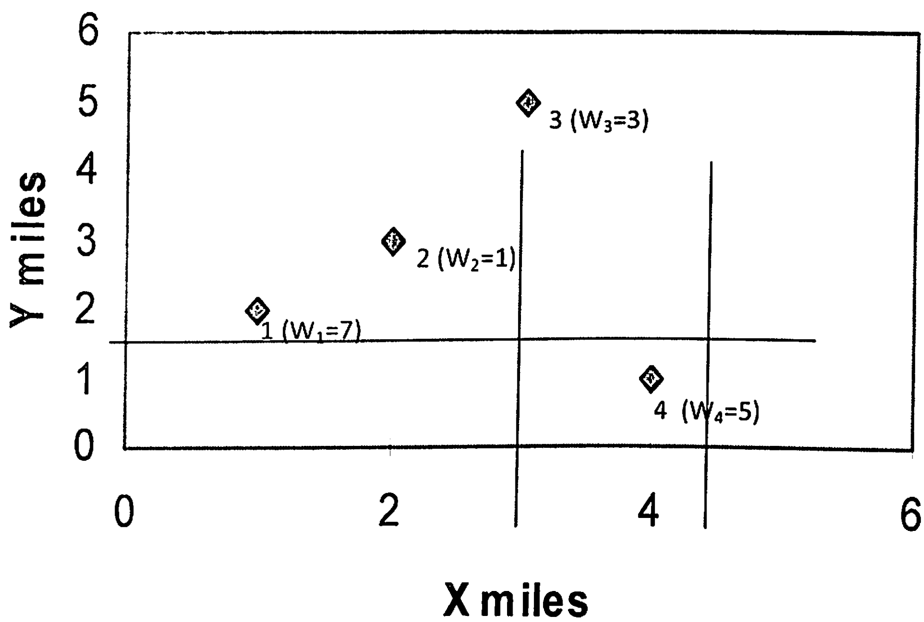
3. Minimize Distance per Visit

(City A: many frequent users)

Single Facility Location Using Cross Median Approach



Single Facility Location Using Cross Median Approach



Huff Retail Location Model

First. a gravity analogy is used to estimate attractiveness of store j for customers in area i.

Aij= Attraction to store j for customers in area i

Sj = Size of the store (e.g. square feet)

Tij= Travel time from area i to store j

lambda = Parameter reflecting propensity to travel

$$A_{ij} = \frac{S_j}{T_{ij}^{\lambda}}$$

Second, to account for competitors we calculate the probability that customers from area i will visit a particular store j.

$$P_{ij} = \frac{A_{ij}}{\sum_{j=1}^{n} A_{ij}}$$

Third, annual customer expenditures for item k at store j can now be calculated.

Pij = Probability customers from area i travel to store j

Ci = Number of customers in area i (e.g. census track)

Bik = Annual budget for product k for customers in area i

m = Number of customer areas in the market region

$$E_{jk} = \sum_{j=1}^{m} \left(P_{ij} C_i B_{ik} \right)$$

Fourth, market share of product k purchased at store j can now be calculated.

$$M_{jk} = \frac{E_{jk}}{\sum_{i=1}^{m} (C_i B_{ik})}$$

CHAPTER 6 CONCLUSION

1. Implementation of a GIS-based approach to habitat restoration site selection that works in conjunction with the current consensus decision method would require the commitment of the HEP to collect additional spatial data.
2. The store-location decision remains critical to the survival of retailing companies between 2002-2005. approximately 80% of all new retail chain stores were single-tenant properties. The Retailing industry is experiencing substantial and fundamental long-term changes that continue to affect the real estate industry. Most existing store location models are inaccurate and grossly mis-specified.
3. The criteria included notably costs, revenue, profit, and budget as monetary criteria; the number of facilities, capacity surplus or deficit, facility dispersion, and competition as facility related criteria; and market share, satisfied demand, distance from customers to facilities, and customer movement as customer related criteria. However, not all of these criteria are relevant in every situation and in certain cases there may also be other criteria that should be included. The applicability of the criteria depends ultimately on the features of the location problem in question.
4. The purpose of the study is to propose, create, and evaluate ways to visualize and to analyze location-allocation solutions and their different aspects. Most of the common visualization methods were tried and illustrated with the help of CommonGIS. These techniques are sufficient when only a few solutions and a few variables need to be visualized at a time.
5. There are still many remaining practical problems that have to be solved before the proposed integrated approach can be put into practice. For example, the proposed multivariate data analysis techniques (cluster analysis, subgroup discovery, PCA, and discriminate analysis) might prove too time-consuming to compute.

REFERENCES

1. Armando B.Mendes & Isabel Hall Themido "Multi Outlet Retail Site Location Assesment: A state of the Art"

2. Beverly K. Brockman, Gary L. Benton, L. W. Turley, "Retail Site Selection in a Mall Context"

3. Simin Huang, Rajan Batta, Rakesh Nagi "An Integrated model for space determination and site selection of distribution centers" November 2003.

4. Chandler McClellan, Steven Turner, Lewell Gunter, Forrest Stegelin "A Location Analysis of Landscape Plant Retail Outlet Markets in Georgia" February 18, 2004.

5. Michael Nwogugu, "SITE SELECTION IN THE US RETAILING INDUSTRY".