CRANFIELD UNIVERSITY

SCHOOL OF MANAGEMENT
MSc Logistics and Supply Chain Management

MSc

Academic Year 2009 - 2010

DANIEL FRANCO

ASSESSMENT OF THE CENTRE OF GRAVITY TECHNIQUE FOR THE SOLUTION OF THE FACILITY LOCATION PROBLEM

Supervisor: Dr Andrew Palmer

August 2010

This thesis is submitted in partial fulfilment of the requirements for the degree of MSc
(NB. This section can be removed if the award of the degree is based solely on examination of the thesis)

ABSTRACT

As the importance of strategic supply chain management grows within industries, the decisions that encompass this area have a greater relevance to and higher impact on the competitiveness of the company. When planning the supply chain, the design of the distribution network that will support the company's operations is a critical aspect. It can make the difference between being competitive or not. Therefore, the importance of the decisions that are taken in this process are especially relevant.

A model that is frequently used for solving the Facility Location Problem is the Centre of Gravity model.

The aim of this thesis is to identify the main characteristics, strengths and weaknesses of the Centre of Gravity model for solving the Facility Location Problem. Furthermore, it also looks at how by grouping the customer database using clustering algorithms the solution given by the model be improved.

The results show that the Center of Gravity model can be used for solving the Facility Location Problem, however, it should be used only as a guideline and not as a decision tool since it has some very critical weaknesses due to its simplicity. They also show that the solution given by the model can be sensibly improved using clustering algorithms but there is one specific algorithm that presents the overall best results.

Finally, some suggestions are made for further research on the Center of Gravity model as a tool for solving the Facility Location Problem.

Keywords:

Clustering; Logistics and Supply Chain; Distribution Network Design; Gustafson-Kessel; Fuzzy c-mean

ACKNOWLEDGEMENTS

I would like to express my most sincere gratitude to my thesis supervisor Dr Andrew Palmer for his support, guidance and motivation throughout the research process.

A very special thanks to my parents, Saul Franco and Evelyne Halpert, for believing in me and supporting me every day of my life, without them I would not be where I am today. I would also like to thank my wife Diana Hernandez for the patience and support that she offered me throughout this year. I would also like to mention my brother and his wife, for their support.

To all of my friends from Cranfield, I would like to thank you for making this year one of the most special years of my life.

A final word of gratitude to the SOM faculty and professors for sharing their knowledge and experiences that made this year worth it.

TABLE OF CONTENT

TABLE OF FIGURES

TABLE OF TABLES

1. INTRODUCTION

In a market where globalization is becoming a norm more than a trend the company's supply chains are expected to cover these new necessities and cope with the new demands that clients have. In order to achieve this, the use of decision tools is becoming more important as the number of variables increase exponentially. Clients located around the globe, products manufactured in various countries, numbers of parts, volumes, client's expectations, legislations, resource limitations are just some of the variables that supply chain managers have to take into consideration when planning a new supply chain or remodelling an existing one. With this new complexity, the tools used to help the decision making process should also be reviewed in order to make sure that they have optimum results under these new market conditions.

One of the problems that supply chain managers have to consider is the location of the facilities that will support the distribution of the products. For this, the centre of gravity model has been used as a simple and efficient way of solving the Facility Location Problem (FLP) throughout the years. The model is based on the weighted distances of the different clients that the company wishes to serve. However, as supply chains are becoming more complex than ever with the globalization of the markets, the simplicity of the model raises some questions on its effectiveness and true functionality as a stand-alone decision tool for supply chain managers under the new conditions.

Some evidence has been found showing that the model can have under-optimal solutions when the problem to be solved presents some specific characteristics that are commonly present in today's supply chains. Literature has highlighted that the model can have problems when there is one dominant client that represents more than half of the total demand and also when clients are located a long way from the rest of the demand nodes.

However, there may be ways of overcoming some of these issues, if the centre of gravity were to be used not as a stand-alone decision tool, but rather as part of a more complex FLP solving model. One option of overcoming the issues is to divide the customer database into groups of clients with similar characteristics, also known as clusters. The idea of using clustering algorithms to overcome these issues has been used in the past obtaining better results than by using the centre of gravity (COG) alone. Some literature has been found on similar models that use a hybrid method of Clustering-COG giving the results that some algorithms outperform others. However, the literature has a limited dataset and does not apply in an extensive way the use of the clustering tool for the solution of the FLP. There are some studies but they have limited the research to dividing a database into a maximum of three groups/clusters.

But do these results hold when those same clustering algorithms are used on different datasets of different sizes? Do they give similar results when analysed with a greater number of groups or clusters?

The idea of this research is to see if, by using a hybrid method of clustering and COG, the general result of the COG as a stand-alone decision tool for FLPs within the industries can be improved upon. In order to achieve this, it is necessary to start by defining the FLP within logistics and supply chain. Once this has been worked, it will be necessary to identify the COG as a model for solving the FLP with its strengths and weaknesses. Since the objective is to identify if the COG model can be enhanced by applying clustering algorithms, three different clustering algorithms, including the two algorithms used by Esnaf and Küçükdeniz (2009), will be applied to three different datasets of different sizes and the results analysed to see how the different models perform to identify if there is one algorithm that outperforms the others when combined with the COG.

1.1. Research Question

Can the performance of the COG model for the solution of the FLP be improved by the application of different clustering algorithms?

1.2. Objectives

- Identify the main characteristics that encompass the Facility Location Problem within Logistics and Supply Chain Management.
- Define the Centre of Gravity technique to solve the Facility Location Problem and identify its strengths and weaknesses.
- Apply different Clustering Algorithms to client databases and analyse the solutions given in terms of improvement to distribution costs.

2. LITERATURE REVIEW

The literature review used four main sources, ABI/Inform Global/Trade and Industry (ProQuest) as the main source for information and articles, EBSCO, Science Direct and finally the search engine Google Scholar (scholar.google.com). Through these four sources a wide range of articles for the FLPs were found.

The strategy of research was as shown in Figure 1.

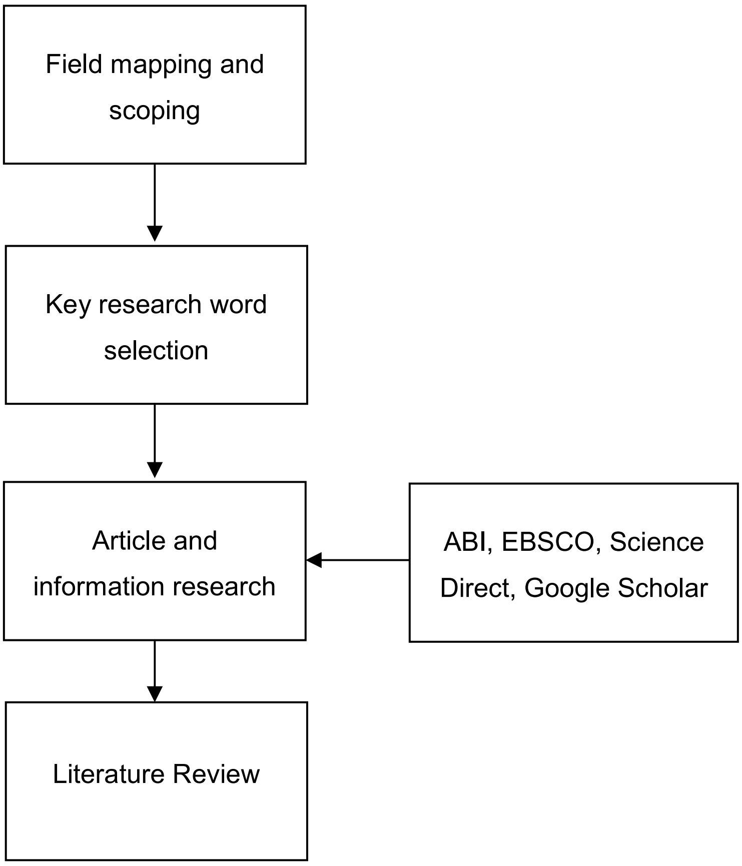

Figure 1 Literature Review Research Strategy (Source: Author)

2.1. Field Mapping and Scoping

The field of FLPs is a wide area of on-going research, therefore it is especially important to map and scope it. Furthermore, the tool of Clustering data must also be put into context during this step. Figure 2 shows the map and scope that was given to the research.

The COG itself is used widely in different areas such as physics and hydrology; it is important to scope the research to its application to solve the FLP. Similarly, FLPs have many variations and there are many forms for solving FLP. Logistics and Supply Chain

Management has also many different areas of research such as demand planning, inventory management and procurement to name a few. Jain et al. (1999) confirm that data clustering has also been used in different areas of research and different contexts. Data mining, data grouping and goodness of fit of the data are some of the examples given.

Since the goal of the research is to analyse how accurate the COG model can be when used to solve the FLP in Logistics and Supply Chain Management and if the results can be improved through the use of data clustering, the area of interest is where these factors overlap in the centre of the graph.

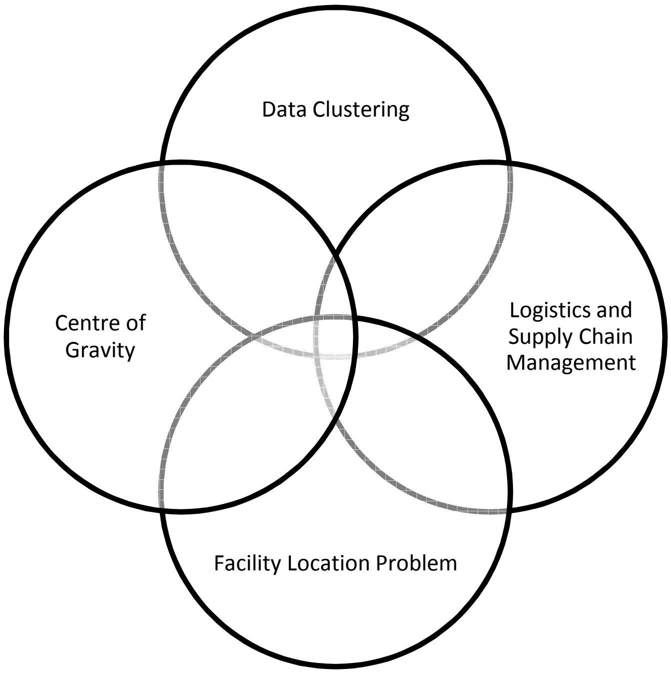

Figure 2 Map and Scope of Research (Source: Author)

2.2. Research Key words selection

All of the fields of interest for the research are widely studied, making it difficult to find articles specifically for the scope of the project. However, some clear key words were chosen for each of the factors that comprise the scope. Table1 lists those keywords divided into each of the factors.

Factor	Key words
Logistics and Supply Chain Management	• Strategic physical network design • Facility location
Clustering	• Clustering algorithms • Data mining with clusters
Facility Location Problem	• Definition of FLP • Classification of FLP • Continuous FLP • Models for solving continuous FLP
Centre of Gravity	• Definition of Centre of Gravity • Solving FLP with COG • Strengths and weaknesses

Table 1 Research Key Words (Source: Author)

2.3. Article and information research

To create a literature review, it is necessary to search in the most significant databases for the related areas of interest as well as other sources such as the internet and theses from previous years. When selecting the databases to be used, the ones selected were ABI (ProQuest), EBSCO and Science Direct. A description of these databases is shown in Table 2.

Databases	Key area	Description
EBSCO-Business Source Complete	Supply chain and logistics management Physical network design	The world's largest scholarly business database provides the greatest collection. Offers more than 2,800 scholarly business journals, including full text for more than 900 peer-reviewed business publications. Coverage includes virtually all subject areas linked to business.

Databases	Key area	Description
ABI/Inform Global/Trade &industry (ProQuest)	Facility location problem and its related issues. Clustering analysis and its related issues	Provides top journals, periodicals in business, management science, computing, transportation from the highest-quality sources of information, and major publishers: working papers, business cases, annual reports, dissertations, etc. nearly 3,000 worldwide business periodicals.
Science direct	Facility location problem Methods for solving FLP Physical network design	More than 2,500 journals and over nine million full-text articles are available.

Table 2 Databases description (Taken from Meeyai (2009), p. 28)

2.4. Review

As the economical environment presents more and more challenges such as the increasing prices of fuel, legislative restrictions transportation within major cities and the increasing responsiveness expectations that clients have, the decisions of supply chain managers are becoming ever more critical and strategic for the success or failure of their organisations.

Christopher (1992) presented the concept that supply chains are competing and not companies. With this in mind, every decision made for the supply chain will have a great impact on the future of the company.

Simchi-Levi et al. (2008) divide supply chain decisions into three distinct categories:

1. Strategic: Deals with long-lasting effect decisions. Within this category lies the definition of the number, capacity and location of the company's facilities.
2. Tactical: Decisions that are usually updated within a year.
3. Operational: Day-to-day operations.

In order to guarantee an efficient supply chain, planning and structuring is an essential and critical part. Figure 3 shows a framework presented by Rushton et al. (2006) for strategically planning and structuring a company's distribution network.

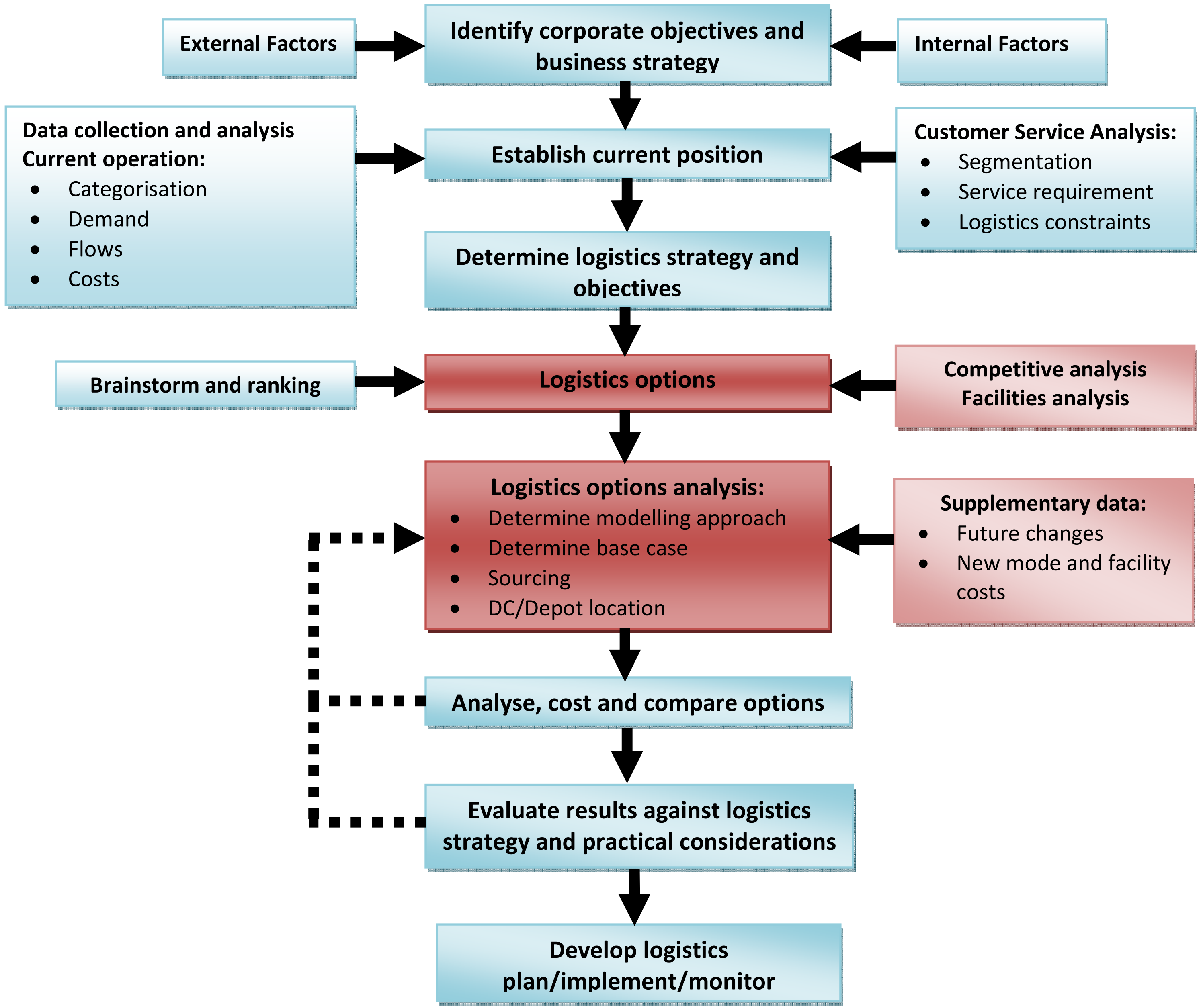

Figure 3 Physical network design framework (Adapted from Rushton et al. (2006))

Following the presented framework, the first two steps require the company to scope into the current situation which involves internal factors such as existing infrastructure, product groups, customer service level, future planned expansions, and external factors such as transport mode availability and trends within the industry. As a result the company may establish its current situation.

Once the company has accomplished this initial analysis, it must go on to define its logistics goals and the strategies that will guide the company into achieving them. When it comes to the next step of defining the logistics options and analysing them, the FLP comes into the scene for defining the optimal location of the necessary facilities in order to reduce distribution costs without affecting the expected service level. This step implies the classification or scoping of the problem, within certain categories that have been defined by different authors, in order to select the optimum approach for solving the problem, taking into consideration that each category has different methods of solution. Revelle (2005) describes the FLP as “Siting facilities in some given space”. This issue of locating the distribution centre/depot is a strategic decision that must be made carefully since it has a critical impact not only in the short term but the long term as well.

This decision usually brings high long term investment. If a wrong decision is made in the facility location part of the process, it will directly affect the costs of the supply chain as well as its efficiency.

These types of problems have been the subject of studies for the past century. From its original form presented by Weber (1909) analysed by Drezner (1995) of locating a warehouse to serve three customers, up to the more complex simulation based models, this is an area of constant development.

Throughout the years, many ways of solving the FLP have emerged, each covering a specific area or variables that make up the problem; however, in order to identify the best way to approach the problem, it is necessary to first evaluate the supply chain landscape and conditions of the organisation.

2.4.1. Classification of the Facility Location Problem

Different classifications for the FLP have been proposed. Esnaf and Küçükdeniz (2009) divide the problem into two main areas which are single facility location problems (SFLPs) and multi facility location problems (MLFPs), both of them being subdivided into capacitated and uncapacitated problems depending on whether the source will have a limited throughput capacity or not. Additionally, ReVelle et al. (2008) divide the models into analytic models, continuous models, network models and discrete location models, depending on the objective of the decision maker. Analytic models can be used to obtain an insight into the situation. They are based on a large number of assumptions that limit the true usefulness of these types of models. Continuous models assume that the facilities can be located in any points of the space. Network models, on the other hand, only consider possible location point within a certain network that is made up from linking the demand nodes. Similarly, the discrete location model assumes that the facilities can be located only in predefined locations.

Even though these classifications are accepted by the literature, for the research done for this project, the classification presented in Eiselt and Sandblom (2004) will be used as the main guideline to define the problem. Figure 4 maps the classification they propose.

According to Eiselt and Sandblom (2004), FLP can be divided into 14 different categories depending on the variables that describe the problem and these that are discussed below. It is worth clarifying that a specific FLP can fall into one or more of the classifications. Each of these factors will directly influence the approach that the decision maker should use in order to solve the FLP.

2.4.1.1. Space Distance

This classification defines the space in which the facility is going to be located. It can either be continuous, where the facility can be located at any point in the space; or discrete, where the facility can only be located at predetermined locations; or network, where the facility can be located within an existing network. Examples of each of these categories are a warehouse location problem where the warehouse could actually be put anywhere within the area of impact (continuous); a location of a supermarket in a city area where the supermarket could only be located in predefined zones defined by the city planning (discrete); the location of a mechanical assistance centre in a highway network that can be located anywhere within that network.

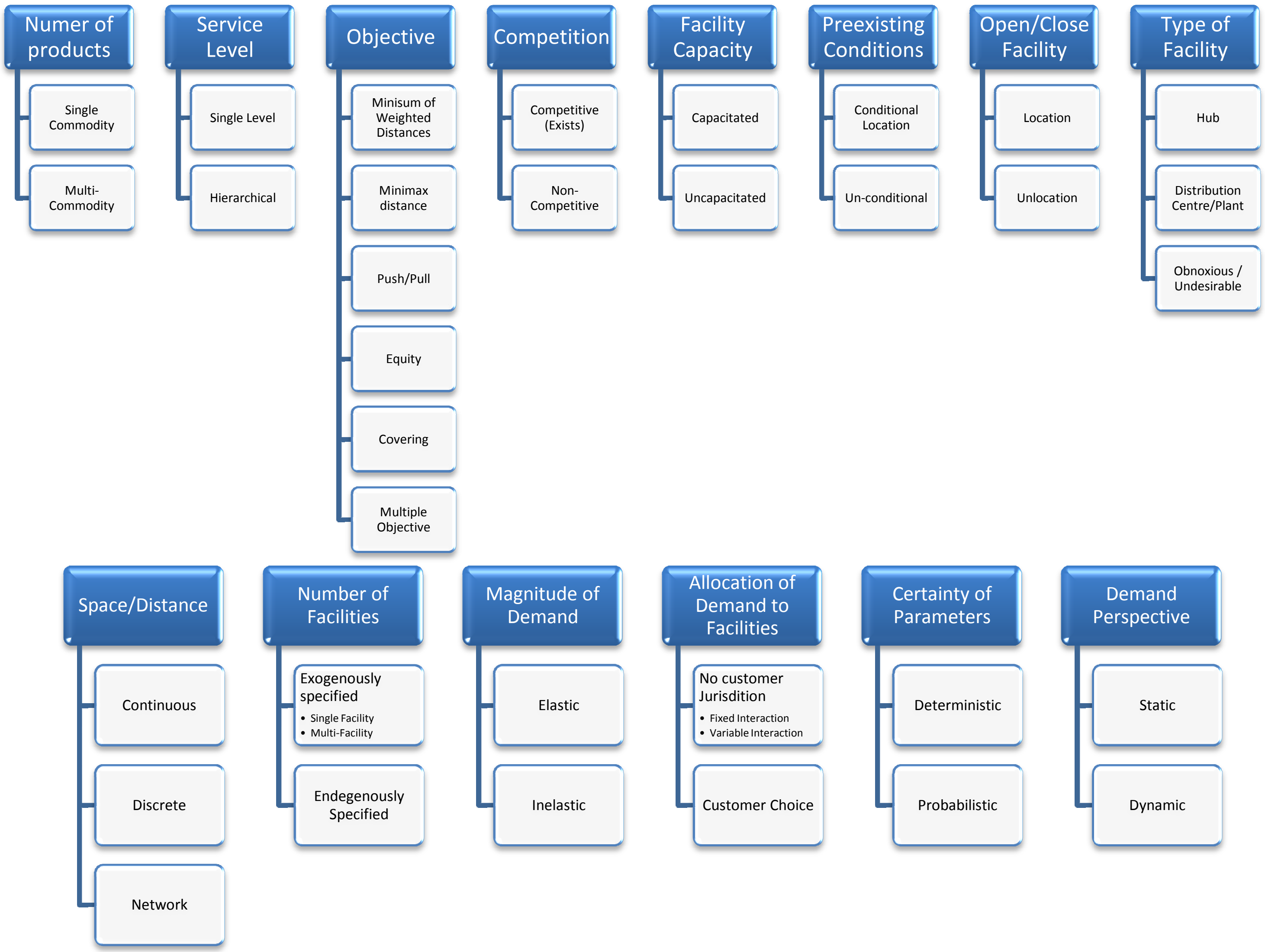

Figure 4 Classification of Location Problems (Adapted from Eiselt and Sandblom (2004))

2.4.1.2. Number of facilities

The number of facilities needed to cover the requirements can be defined by the model that is going to be used to solve the problem, in which case they are considered to be Endegenously specified number of facilities. If the number of facilities are predetermined by the decision maker, and can be considered as an input to the problem, then the problem is considered to be an Exogenously specified model. The latter can be further divided into Single Facility and Multi-facility problems, depending on the numbers that are going to be located. An example of an issue that can define the number of plants that can be opened is a company's budgetary plan.

2.4.1.3. Magnitude of Demand

The demand can be elastic or inelastic, depending on its response to factors such as price. If the product that is being distributed has a low to no response to price, it is regarded as having an inelastic demand. An example of this is a basic commodity such as household salt. On the other hand, cars are an example of a product that has an elastic demand since the price of the product directly affects the volume of the sales.

2.4.1.4. Allocation of Demands to Facilities

This characteristic classifies the FLP into problems where the customers have a part in deciding which facility will serve them or not. If the customer has the opportunity to choose which facility he is going to use, then the problem is classified as a "Customer choice" problem. An example of this type of problem is a supermarket. The customer chooses which supermarket he is going to use. Factors such as closeness can be considered to have a big effect on the decision process; however, other factors such as the distribution of the supermarket, the specific sales that each one has, or simply the ambience of the store can also impact on the decision of the customer. The other class of problem within this classification is known as the one in which the customer has no jurisdiction in terms of the facility that will serve him. An example is the location of a distribution warehouse for a company that delivers the product directly to the customer. In this case, the company is the one that decides which location will serve which customer. Here factors such as closeness to the customer and capacity of the warehouse are some of the predominant issues that decision makers take into account when assigning the demand to a certain distribution centre.

2.4.1.5. Certainty of Parameters

The fifth classification defines the certainty of the parameters of the problem. When the parameters are well known and defined, then the problem is known as a deterministic model; however, in logistics many of the parameters are not known with any certain degree of certainty. In this case the problem is defined as a probabilistic problem. Demand is a typical example of a key parameter that is not known.

2.4.1.6. Demand Perspective

The perspective of the demand can either be static or dynamic. Static demand is considered to be the demand that is regarded as a number that will be unchanged throughout the analysis. Dynamic demand is where the demand volume is considered to change as time passes. If the demand is considered to be dynamic, then the model will most certainly have a probabilistic parameter which would be the demand forecast.

2.4.1.7. Number of Products

The number of products that are going to be distributed will also affect the model. If there is more than one product, then the model is known as a multi-product model; conversely, if there is only one product to be distributed, then the model will be known as a single commodity model. This may affect the optimality of the distribution process due to possible transport inefficiencies or warehousing requirements.

2.4.1.8. Levels of Service

For the FLPs, service levels can be divided into Single or Hierarchical levels of service. The example given by Eiselt and Sandblom (2004) is that of a healthcare facility. Within the healthcare area there are different levels of service such as a doctor’s office, a small clinic or a major hospital. Each one of these facilities is within the healthcare scope but each has different services, usually from a very basic or specialised service to very complete and multi-service facilities. The issue with this type of problem is that it not only requires the decision maker to identify the location of the facilities but also decide on which level of service should be offered at each location to serve the customers’ demands in the most cost effective way. It will clearly be more costly to build a full sized hospital than a smaller doctor’s surgery.

2.4.1.9. Objective

This classification depends exclusively on the decision maker. There are six options defined by Eiselt and Sandblom (2004). These are: Weighted distance, which tries to model the transportation costs that the company would incur for the distribution of products to its customers depending on the distribution of the demand, and distance between the demand and the facility; Minimax distance that tries to minimise the maximum distance between the facility and any given demand node; Push/pull which considers the desirability of the facility for the customer - if the facility is desirable, such as a supermarket, then it is considered to be a pull situation, however if it is a non-desirable facility, such as a landfill, than it would be considered to be a push situation; Equity is the case in which the decision maker wants to make all of the distances between the facility and the demand nodes equal; Covering, which is where the idea is to cover a certain area, a good example of this objective is the location of a fire station where the idea is to have covered every demand node within a certain time or distance; Multiple-objective exists since many of the aforementioned objectives are not mutually excluding.

2.4.1.10. Competition

This considers the existence, or not, of competition within the model. If competition is taken into account, then the model is defined as a competitive model; on the other hand, if the model does not consider the existence of competition then the model is known as a non-competitive model.

2.4.1.11. Facility Capacity

This characteristic is one of the most commonly used to divide the models. If the facility that must be located does not have a maximum throughput capacity, whether it is goods or data, then the problem is known as an uncapacitated problem. If the facility does have a maximum throughput capacity, then it is known as a capacitated problem.

2.4.1.12. Pre-existing Conditions

This classification considers whether there are existing facilities or limitations, or not. If the company already has existing facilities or limitations then the problem is known as a conditional model. If there are no conditions limiting the model then it is known as an unconditional model. Unconditional models usually exist in the first stages of a new logistics network design. When a logistical infrastructure already exists and the objective of the exercise is to re-design, then most probably the model will have existing facilities and conditions that would make it a conditional model.

2.4.1.13. Open/Close Facility

When re-designing a supply chain, one of the objectives may be to close facilities and not necessarily open new ones. In this case, if the problem is trying to close facilities then the model is known as an unlocation model. If, on the other hand, the model is trying to locate new facilities to open, then the model is known as a location model.

2.4.1.14. Type of Facilities

There are different types of facilities that can be located. Examples of some of these types are the location of hubs – these are normally used in the aviation sector and locate airplane hubs to cover the demand in a more efficient manner; Distribution centres (DCs), warehouses or plants; Obnoxious or undesirable facilities, such as landfills or chemical waste processing plants, because customers want to be as far from them as possible.

According to the mix of the characteristics that make up the FLP, different approaches to solving it may appear. Simulation, Tabu Search (Drezner and Drezner, 2007; Mladenovic, 2003; Arostegui et al., 2006), Genetic Algorithms (Alp et al., 2003; Jaramillo et al., 2002; Arostegui et al., 2006), Stochastic Models (Zhou and Liu, 2003), Linear Programming (Erlenkotter, 1978; Lashine et al., 2006), Simulated Annealing (Arostegui et al., 2006), Lagrangian Heuristics (Sridharan, 1993,1991; Chen, 2008) and mixed models (Esnaf and Küçükdeniz, 2009) are some of the most commonly used and most researched methods used for solving the many varieties of FLP.

Cooper (1961) can be said to be the first to work on the fact of multi-facilities to solve the location problem. He describes the general problem as follows:

"Given the location of each of the destinations, the requirements at each of them and a set of shipping costs for the region of interest, it is necessary to determine the number of sources, the location and the capacity for each of them."

This description is further supported by ReVelle and Eiselt (2005). Cooper's model is still based on some of the original assumptions: the fact that there are no restrictions on the permissible capacity of each source and that the shipping cost is independent of the volume shipped. With his method, clients are divided into clusters and served by the different sources in order to minimise the distance between them.

From Cooper's first model, many others have evolved covering some of the other issues from the original model. Wen and Iwamura (2008) build on Zhou and Liu (2003) regarding a fuzzy facility location allocation problem in order to cover the uncertainty of the demand forecast. Their model uses the Hurwicz criterion in order to balance between an optimistic and pessimistic forecast.

Arostegui et al. (2006) highlight the 0-1 MIP, dynamic programming, breakeven analysis, quadratic programming and Fuzzy set theory, as being some of the approaches that have been developed in the past two decades to solve location problems. Mladenovic (2003) uses a Tabu search with variable neighbourhood search to solve the *p*-centre problem, which had been previously used by Ghosh (2002). Alp et al. (2003) apply a Genetic Algorithm to the problem.

However, as can be expected, some models give more optimal results than others. Arostegui et al. (2006) compare the Tabu Search (TS), Simulated Annealing (SA) and Genetic Algorithm (GA) models for solving FLP, coming to the conclusion the TS should be used as the first option due to its overall good results on the tested variations.

Another model that is commonly used to solve the FLP is known as the Centre of Gravity (COG) model.

2.4.2. The Centre of Gravity (COG) Model

The COG is a model that tries to find the point in the plane where the distance to each of the demand points is minimum considering a weighted average. Ballou (1973) highlights the fact that this method is widely used due to its simplicity and ease of use. Ballou (1999) also refers to it as *p*-median, the grid method and the centroid method and defines the COG model with the following equation:

(1) $$MinTC = \sum_i V_i R_i d_i$$

Where:
TC = total transportation cost
V_i= Volume at point i
R_i= Transportation rate to point i
d_i = distance to point i from the facility to be located.

The facility location in the model is calculated by applying the following equations (2) and (3).

$$\bar{X} = \frac{\sum_i V_i R_i X_i / d_i}{\sum_i V_i R_i / d_i} \quad (2)$$

$$\bar{Y} = \frac{\sum_i V_i R_i Y_i / d_i}{\sum_i V_i R_i / d_i} \quad (3)$$

Where $\bar{X}$ and $\bar{Y}$ are the coordinates of the location of the facility and X_i and Y_i are coordinates of the demand points or customers.

The distance between the 2 points can be calculated in different ways depending on the interest of the researcher. However, Love and Morris (1972) have shown that the formula where d_i is calculated by using the following equation (4).

$$d_i = K\sqrt{(X_i - \bar{X})^2 + (Y_i - \bar{Y})^2} \quad (4)$$

where K is a scaling factor to convert the units of coordinate points to a distance measure such as kilometres can be very accurate. This equation represents a straight line between two points and the scaling factor is also used to convert this to an estimated road distance. One of the main problems in a centre of gravity calculation is the issue of distances not allowing for crossing over water or avoiding mountains, for instance. To overcome this, barriers can be introduced to ensure that these obstacles are avoided. The barriers would be in the form of via points meaning that the distance formula would be calculated using the equation 5.

$$d_i = K\sqrt{(X_i - Xj)^2 + (Y_i - Yj)^2} + K\sqrt{\left(X_j - \bar{X}\right)^2 + (Yj - \bar{Y})^2} \quad (5)$$

Where Xj and Yj are the coordinates of the via point.

Figure 5 represents the process for calculating the centre of gravity point according to Ballou (1999).

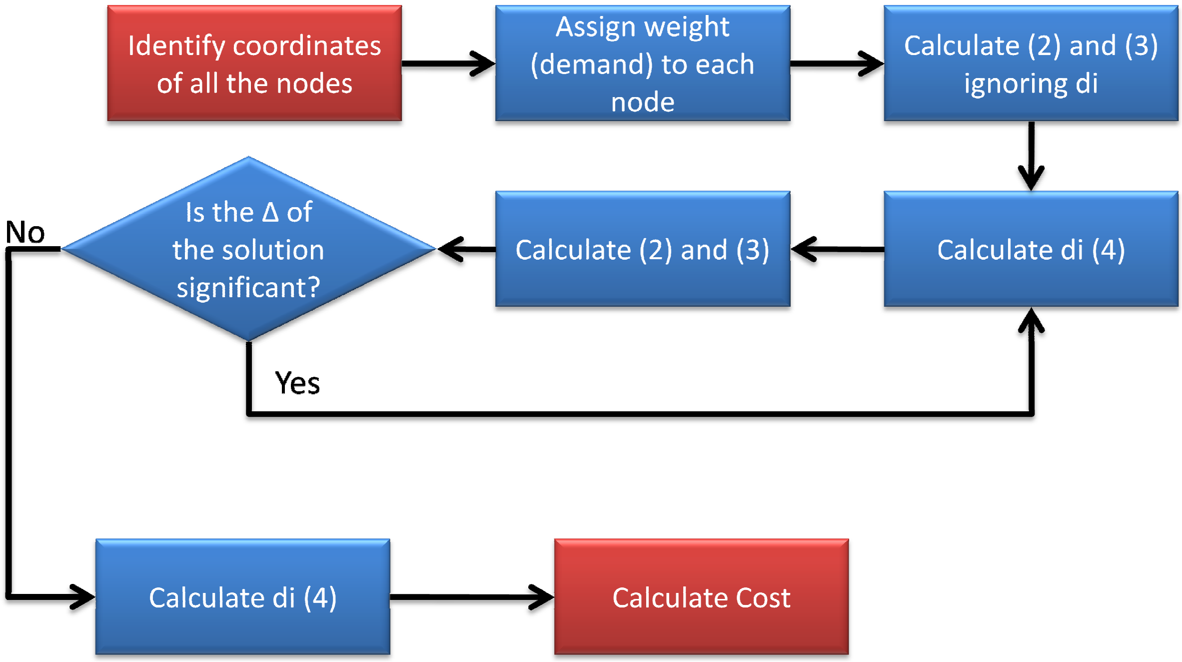

Figure 5 Process for calculating the Centre of gravity (Adapted from Ballou 1999)

The idea of the model is to minimise the total transportation cost of the merchandise from origin to the final destination.

The main principle for this model to work is that it must be used under a Continuous Space problem since the solution given can be found in any place of the space. However, this is not the only characteristic that the problem must have; the problem should also fall into the following set of classifications:

1. Exogenously specified number of facilities since the model does not calculate the optimum number of facilities that must be open, but rather just their location.
2. Static demand. It is important to consider that the model is based on a fixed number that represents the demand/weight of each of the nodes. This number cannot be changed in the model; it can only be changed as a new case for modelling.
3. The customer does not have jurisdiction over the allocation of the demand to each facility since the centre of gravity model requires that the number of customers that must be allocated to a certain facility be defined.
4. Single level service
5. Minisum of weighted distances is the main objective.
6. Non-competitive environment.
7. Uncapacitated facilities due the simplicity of the model; this restriction cannot be considered although it can be modified afterwards as part of the analysis. It can be used only if the decision maker considers it before running the model and allocates

the weights of the nodes to the limit of the facility. However, the model itself will not consider this.

8. It is more commonly used for DC location in the network. Other models exist for solving models that use different types of facilities such as hubs and cross-docking platforms.

Additional to these classifications, another two assumptions are made in order for the model to work. These assumptions are:

1. The cost of setting up a facility is the same at any location of the plane.
2. For simplification reasons, demand is located in one point, known as a node.

However, these assumptions conflict with some of the real-life issues that supply chain managers have to deal with. Some of the issues are:

1. In some cases, there are restrictions on the possibility of locating a facility in a given place. Restrictions can be due to geographical issues such as mountains and unstable ground; they can also be of a legal aspect such as natural reservoirs and city planning; or simply that there is something else already existing at the chosen location.
2. Land costs are not equal in all places. Furthermore, the setting up of a location in the Alps could be assumed to be considerably higher than the setting up of a facility near a city, due to transportation costs.
3. In a large network of clients, one facility may not be enough to cover all of the customers. Throughput capacity and transportation costs can be restricting.
4. Waters (2006) describes demand forecasting as one of the most uncertain areas of logistics. Furthermore, when planning to open new facilities, decisions must be taken based on long-term prediction, adding to the uncertainty of the forecast.
5. Nodes do not consider the number of customers that make up the consolidated demand. From a distribution perspective, serving one customer with a high volume is different from serving many customers with small volumes, even though at the end they add up to the same amount.
6. If one of the high weight clients were to disappear or could no longer be served by the company, the impact of this may be great over the COG, making it very necessary to do a sensitivity analysis of how these big clients impact on the COG. This type of sensitivity analysis is not easy to develop with the COG model since it is a static equation that solves it.

Additionally to these conceptual issues, Ballou (1973) has also found some mathematical issues that may impact on the optimality of the COG model as the main decision making tool for the FLP. There were two main errors identified - both of them regarding the issue that since the COG is based on an average, therefore it has the same flaws that an average has.

Firstly, if within the database there is a demand point with an exceptionally high demand or weight, then the COG will be heavily influenced by this fact, and will tend towards that location.

Secondly, if there is one client that is extremely far away from the rest, this will also affect the accuracy of the COG.

It was also found that as the number of clients on the database increases, the more accurate the solution of the COG will be. This may be explained by the fact that as the database increases, the fewer chances there are that there will be one dominant location over the rest, and the general weights will be better distributed.

For this reason, Ballou's (1973) article concludes that the COG, combined with other tools, can be a better decision making solution than using it by itself.

Following on from this affirmation, Esnaf and Küçükdeniz (2009) used a combination of clustering algorithms and the COG in order to solve an FLP case. However, the study is limited to two algorithms and only analyses up to three clusters. In the research it was concluded that the best solution was given by combining the Gustafson-Kessel clustering algorithm with the COG rather than using the Fuzzy c-mean clustering algorithm, even though in both cases there were great improvements from the initial state of only one cluster.

2.4.2.1. Weaknesses

The main weaknesses that can be defined based on the characterization that has been made are the following:

- In a globalized market where clients can be located all around the globe, the fact that the COG model is susceptible to clients being far apart from the rest can create a lot of problems. In the part, when clients were local, would have been less important as it is today, but now, that clients can be located anywhere in the world, remote locations such as Australia can have a big impact on the COG calculation.

- The solution given by the COG model can be placed anywhere in the global map, making it possible to get a solution in the middle of the ocean or the middle of the Alps. These options would make no sense as possible solutions.
- The COG model does not take into account some costs such as placing a DC, distribution costs, inventory costs, land costs and taxes. These can lead to possible sub-optimal solutions.
- The COG is based on a picture of the demand on a given moment or possibly a forecasted value. These are not dynamic numbers and therefore can be susceptible to market changes or the uncertainty that time brings to any forecast. This can be very dangerous since the investments that are made based on the decisions of a network design conclusion are very high.

2.4.3. Clustering

Jain et al. (1999) define clustering as follows:

"Clustering is the unsupervised classification of patterns into groups"

Through this definition, the authors imply that clustering helps to divide a set of data into groups that have similar characteristics. By doing so, this process can help the user to simplify the analysis of the data as well as find possible patterns that may exist within the dataset. This may help to generalise the studies and study possible unlabelled data that otherwise could not be identified.

Due to its general definition, clustering can, and is, widely used in many areas of study. Health research uses clustering within genetics analysis in order to find similar gene configurations that may simplify the complexity of the genome. Another area where clustering is widely used is in information technology, specifically in network design.

Over time, many clustering techniques have appeared for the different cases that may arise. A generic way of classifying clustering algorithms is presented in Figure 6.

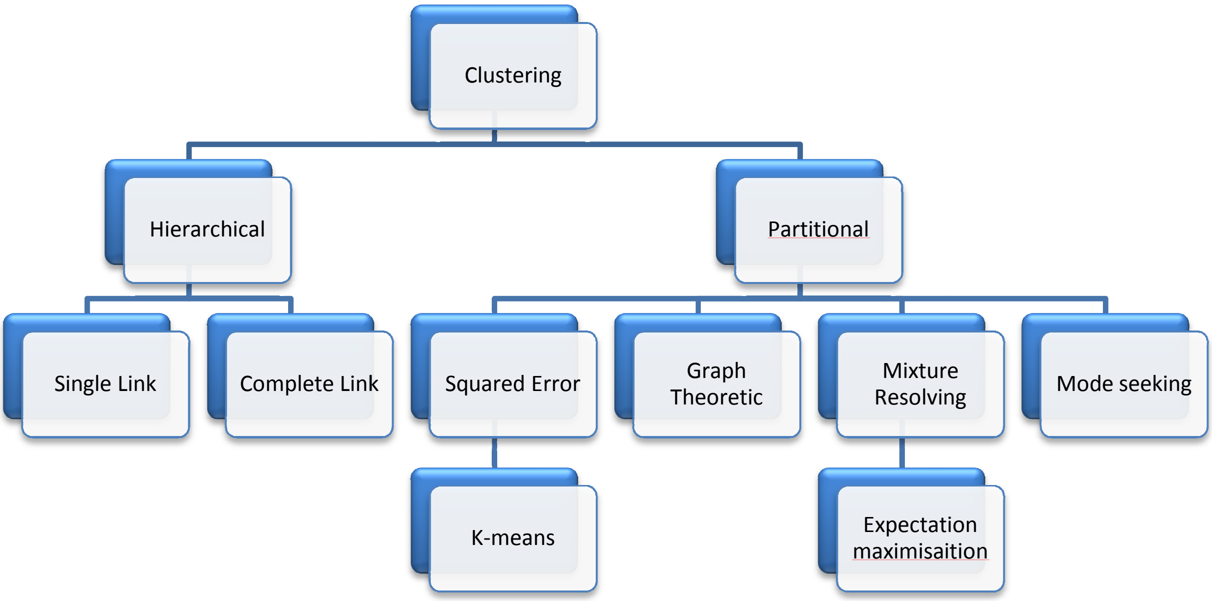

Figure 6 Clustering techniques classification (taken from Jain et al. (1999), p 275)

For a more detailed analysis of each of these categories, the author recommends the reader to review Jain et al. (1999) or Jain and Dubes (1988).

However, for this paper an extra technique will be analysed. This technique can be classified within the Partitional category and it is known as the Nearest Neighbour Clustering Technique (NNCT). This type of technique uses the distance between the nodes in order to classify them within a certain cluster.

In order to run a clustering algorithm it is necessary to have the following information:

- Feature definition: an attribute on which the clustering will be based. For this research it will be the geographical locations.
- Dataset: the database that must be clustered.
- Number of Clusters: the desired number of groups or clusters into which the data will be divided.

These three decisions will determine the outcome of the clustering, therefore it is a time of important decisions where experience and knowledge play a big role. When defining the number of clusters that should be used, it is important to review the size and characteristics of the dataset. If the dataset is relatively small, then a large number of clusters will tend to give very small groups that may not be the most suitable situation. On the other hand, if the dataset is relatively large a small number of clusters may lead to some points not being very well defined to which cluster they best belong to. Both of these situations will be identified in the research that will be developed in this document. In the same way as it will be used throughout this research, the number of clusters can be seen as a variable that will be chosen by the best result that comes out of the analysis.

NNCTs follow a similar iterative process. This process can be defined as follows:

1. The dataset that will be analysed must be clearly defined and selected.
2. A feature must be chosen. This feature represents the factor that will define the similarity of the data. In the case of nearest neighbour the feature is distance by default.
3. There are many different algorithms that can be used on the dataset. This step is especially critical since the choice of a given algorithm will have big impact on the outcome. Each algorithm will most likely divide the same dataset in different clusters, even when using the same number of clusters. In other words, a specific data point can be assigned to different clusters depending on the algorithm that is used. To determine the best algorithm, research and experience are two main attributes that the decision maker should have.
4. Once the algorithm is chose, the number of clusters must be determined. Again, the number of clusters depends on the size of the dataset, the characteristics of the dataset and the experience of the decision maker. However, it can be seen as a variable that will be part of the outcomes of the model. Such is the case of the research of this document.
5. The chosen algorithm is applied to the dataset and the outcome of this is the classification of one of the units into the cluster where it best fits according to the chosen feature.
6. Point 5 is repeated until all data is assigned to a cluster. When the clustering algorithm is Fuzzy, the result is not a specific cluster but rather a number that determines how well the data fits into each of the clusters. If the algorithm is hard, the result is that each data is assigned to only one cluster. Once all data is assigned, the clustering process finishes.

To help understand how clustering is used Figure 7 presents a basic flowchart on the steps that must be taken in order to cluster the data.

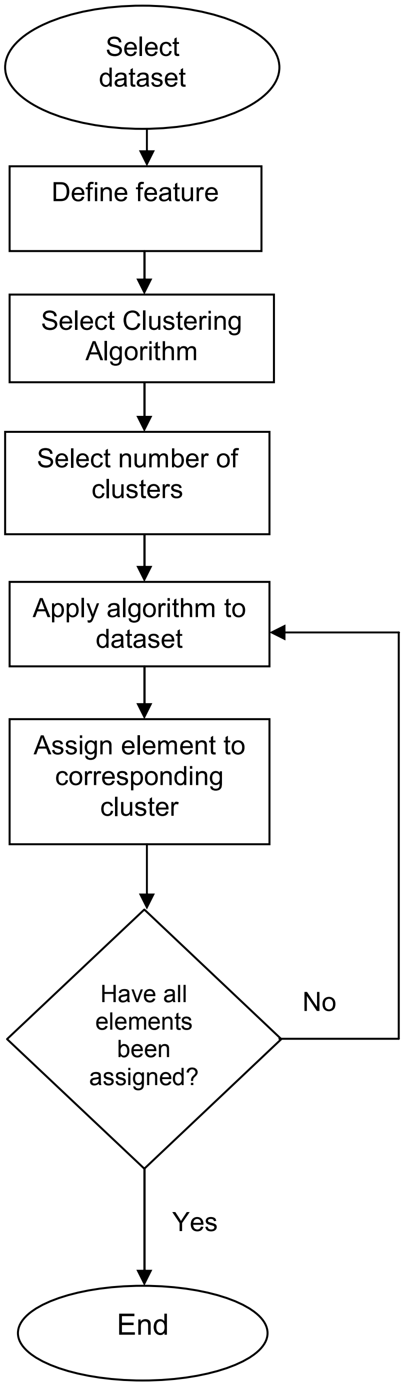

Figure 7 Application of clustering algorithms to a database flowchart (Adapted from Jain et al., 1999)

All of the aforementioned clustering techniques can present different characteristics within each cluster. The techniques can be Deterministic or Stochastic, Incremental or Non-incremental, Hard or Fuzzy, Agglomerative or Divisive, and Monothetic or Polythetic. For a specific explanation of all of these mentioned characteristics, again it is suggested that the reader reviews Jain et al. (1999). For this research, however, some of these characteristics will be reviewed due to their impact on the methodology and data analysis.

Hard clustering techniques are those techniques that assign each of the nodes to one, and only one, cluster. It does not allow for fuzzy results where a node can be partially allocated to one node and not the other. On the other hand, fuzzy techniques allocate each node to each cluster in a partial way depending on how well that specific node fits into the determined cluster. The model assigns a value between 0 and 1, 0 being if the node has no relation to that cluster and 1 being a perfect fit. This is important because the analysis that will be used in this research will all be based on fuzzy clustering techniques. It is important to consider that a node may be in the midpoint of two clusters, in which case it would be the decision maker's responsibility to assign that node to one of the clusters. This decision can influence the result of the analysis especially if the node in question has a high weight/demand assigned to it.

Jain et al. (1999) and Bose and Chen (2009) agree with the fact that the most common fuzzy clustering technique is known as the c-means (FCM). This model was initially proposed by Bezdek (1981). Since then, some modifications have appeared where the Gustafson-Kessel (GK) algorithm is highlighted by Bose and Chen (2009) as one of the most commonly used. Since these two algorithms are the most common for fuzzy techniques they will be used for this research. Additionally, since Esnaf and Küçükdeniz (2009) use those same two algorithms, they will be used in order to compare the results in terms of which algorithm gives the better results. However, to extend this research a third algorithm will be used.

The Gath-Geva (GG) model will also be analysed. This model was initially presented by Gath and Geva (1989) and has also been widely used for research in different areas.

3. METHODOLOGY

The research and analysis followed the process shown in Figure 8.

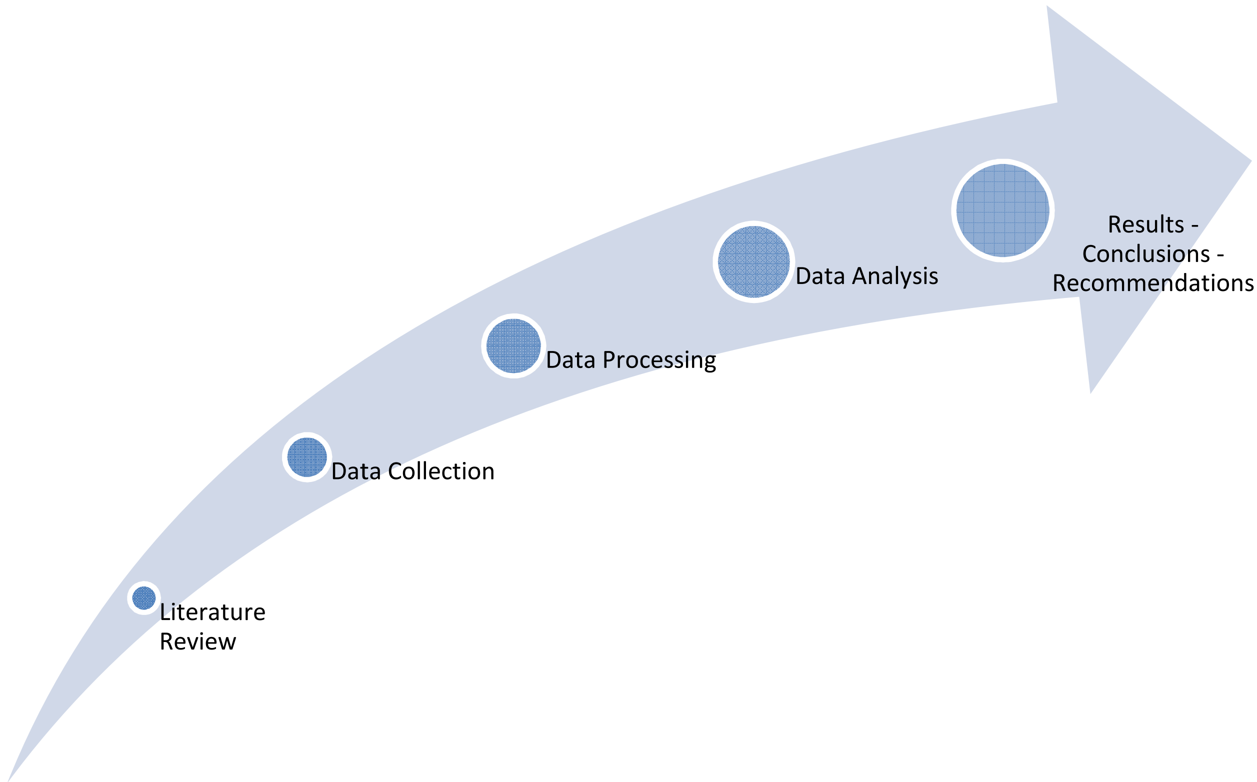

Figure 8 Process followed throughout the research for the thesis. (Source: Author)

The first step of the process is based on a literature review in order to identify the research that has already been done on the different subjects. The second step is to identify and collect the data that will be used in order to evaluate the model that is proposed. The third step is to process the data collected in the previous step. The fourth step is to analyse the results shown by the data processing step. Lastly, results and conclusions are made based on the findings given by the model.

3.1. Data Collection

In order to evaluate the algorithm models in different situations, three different databases of clients, their location and demands were chosen. The databases were the following:

1. Portuguese Paper Producer (PPP): The database comes from a case study of the company PPP. It is comprised of 54 clients, located throughout Europe. For this

database, the supplier plant is located in Lisbon (Lat: 38.7 North, Long: 9.1 West).The total demand for the clients is 2,783 pallets divided between the clients.

2. Database provided by Dr Andrew Palmer: It is comprised of 292 clients throughout Europe. For this database the supplier was assumed as located in Adro (Lat: 45.57 North; Long: 9.97 East). The total demand for the clients is 53,191 pallets of product divided between the clients.
3. Database provided by Dr Andrew Palmer: It is comprised of 521 clients located throughout Europe. For this database the supplier was assumed as located in Adro (Lat: 45.57 North; Long: 9.97 East). The total demand served by the DC is of 39.471,488 pallets.

Appendix A shows a list of clients, their location and the demand for each of the 3 databases.

3.2. Data Processing

Data processing was done in two steps. The first step was to apply the chosen clustering algorithms to each of the datasets. The second step was to apply the clustering solutions to an Excel model designed and programmed by the author, in order to calculate the total transportation, inventory holding and DC throughput costs.

3.2.1. Database clustering

This process was done using a mathematically specific software called MATLAB. This software is a licensed software with copyright protection held by The MathWorks, inc. The software uses a specific programming language that allows the user to create complex mathematical functions. For ease of use, users who have already created the code for common functions share these codes by means of toolboxes. These toolboxes are mostly freeware and can be found on the internet. Therefore, apart from a researcher’s own code, the program is made up of modules (toolboxes) with different functions. There are toolboxes for financial mathematics, specific engineering mathematics and so on. For this research MATLAB’s Clustering toolbox was used as the main tool for analysis.

MATLAB’s results are given in a numerical and graphical form. The numerical form is given as a matrix of NxM, where N is the number of clients and M is the number of clusters evaluated. Each client is assigned a number for each cluster that represents how well that client fits into that cluster. The sum of these numbers must equal to 1. For this exercise the client was allocated to the cluster with the highest fit, since it would not be logistically rational to serve the same client from different DCs. Figure 9 shows how MATLAB provides the numerical results. Column A is the name or code of the clients; columns B to H represent the grade of fitness that the client has to each of the clusters; column I was inserted manually in order to simplify the analysis and it shows the highest number of fitnesses; column J has also been inserted manually and it shows the cluster that has the highest value, therefore it is the cluster to which that specific client is assigned.

	A	B	C	D	E	F	G	H	I	J
1	Client	1	2	3	4	5	6	7	Maximum	Cluster
2	3997015	0.03372464	0.68957575	0.03811391	0.06537971	0.04128222	0.01467682	0.11724695	0.68957575	2
3	3997100	0.0329893	0.69498966	0.03733413	0.06418852	0.04044103	0.01429864	0.11575872	0.69498966	2
4	3997013	0.03299461	0.69466531	0.03724537	0.06448602	0.0402373	0.01413907	0.11623233	0.69466531	2
5	3993012	0.03287578	0.69423768	0.03684526	0.06526758	0.03947854	0.01358052	0.11771464	0.69423768	2
6	3992021	0.03262225	0.69222539	0.03620646	0.06675688	0.03828179	0.01265965	0.12124759	0.69222539	2
7	Catania	0.02548721	0.75439004	0.02980367	0.04993039	0.03293892	0.01146646	0.09598331	0.75439004	2
8	3989100	0.01878014	0.81145756	0.02281552	0.03673293	0.02588516	0.00894631	0.07538238	0.81145756	2
9	Palermo	0.02638728	0.73612706	0.02984771	0.05626508	0.0316108	0.00986904	0.10989302	0.73612706	2

Figure 9 MATLAB numerical solution for seven clusters using FCM algorithm on database 3 (Source: Author)

MATLAB also gives a graphical solution. Figure 10 shows an example of this graphical solution. As can be seen, each cluster is represented by layers of circles. These layers show how well each client fits into that given cluster. Each client is represented as a point in the graph and the centre of each of the seven clusters is shown as the smallest circle on the graph. It is important to highlight that the centre shown on this graph does not represent the COG for each of the cluster datasets because up to this point the demand or weight for each of the clients has not been taken into account.

Each database was evaluated using three different clustering algorithms, and being divided into 2, 3, 4, 5, 6, 7 and 8 clusters. The three algorithms used are Fuzzy c-mean (FCM), Gustafson-Kessel (GK) and Gath-Geva (GG). The first cluster case is considered to be equal for all three clustering algorithms since it represents the case where all the clients are serviced from the same location. This process returns a total of 22 different options of clustering for each dataset.

In order to take into account the weights and calculate the centre of gravity, an Excel model was designed and programmed. This model was used for the second step of the data processing.

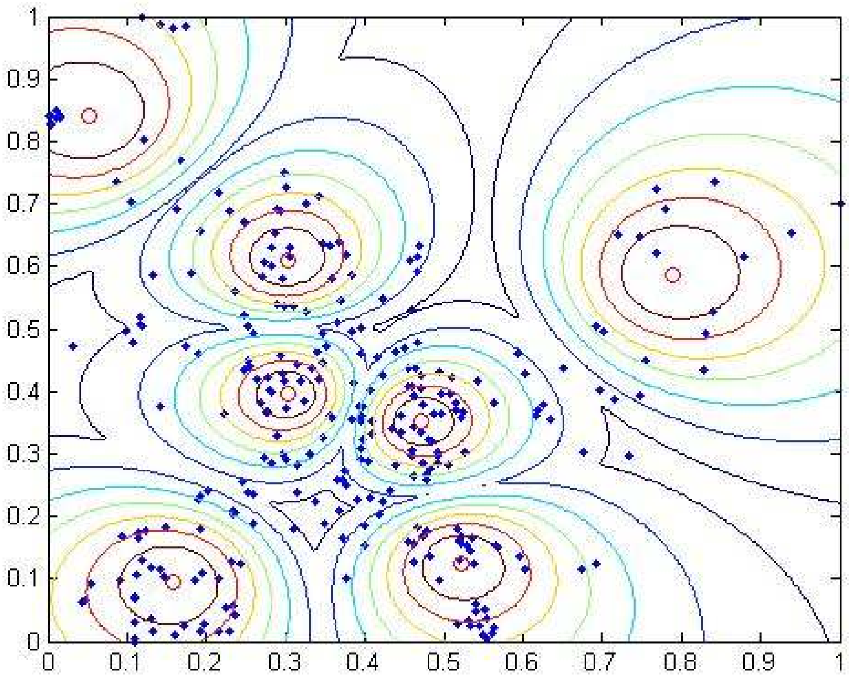

Figure 10 Graphical solution for seven clusters using FCM algorithm on database 3 (Source: Author)

3.2.2. Excel model: Centre of Gravity and costs per cluster

The Excel model was designed and programmed by the author especially for this research work. The objective of the model is to calculate the COG of each of the clusters that were obtained from the previous step and the costs that the company would incur if that strategy were to be used.

The model assumes that each cluster will have a DC that will serve all of the demand of the customers assigned to that specific cluster. Furthermore, that DC will be located at the COG of that cluster.

For calculating the COG, the formulae described in chapter 2.4.2 are used. After reviewing some examples, it was defined that after the fifth repetition the changes in the coordinates were sufficiently small in order to get an idea of where the COG should be located.

As mentioned in chapter 2.4.2, the COG is used in continuous spaces, therefore the DC can be placed at any point of the plane and all of the assumptions defined in the aforementioned chapter will apply. Additionally, the model does not consider any ‘over cost’ the company may have to pay because of multimodal transportation, specifically when sending product from mainland Europe to the United Kingdom and vice versa.

Table 3 lists and describes the input data that must be entered into the model before running it.

Fixed Local delivery cost	Cost per pallet for distributing from the depot to local clients
Variable Local delivery cost	Cost per pallet per kilometre for distribution from the depot to local clients
Wiggle Factor	Factor used in order to add an amount of Kms to consider the curves and extra kilometres that a truck must travel
Depot Cost	Cost to maintain a Depot in operation
Variable Depot Cost	Cost per pallet of throughput in a given depot
Safety stock	Stock quantity that must be stored in order to support the operations during replenishment. The number is given in weeks of inventory
Product Value	The value of a pallet of product
Holding Cost	The cost that the company incurs when holding a pallet in stock
Fixed Trunking Cost	Cost of replenishing a pallet from the plant to the depots
Variable Trunking Cost	Cost per pallet per km for replenishing from the main plant to the depots
Plant location	The coordinates of the plant from where the product originates and is distributed to each of the depots
Cluster	This number represents the number of the cluster to which the location belongs. If there are no Clusters but just one data set, then this number should be 1 A maximum of 10 clusters can be analysed at a time
Customer Location	The country of the location. This data is optional
Location	City or name of the location
Demand/Weight	The demand or the weight corresponding to the location
Latitude	Latitude coordinates of the corresponding location. If the coordinate is SOUTH it should be entered as a negative number
Longitude	Longitude coordinates of the corresponding location. If the coordinate is WEST, it should be entered as a negative number

Table 3 List of input data for the model (Source: Author)

The outputs of the model are listed in Table 4.

Base iteration	The initial iteration of the COG
Iteration 1	Each iteration afterwards will tend to close into a number.
Iteration 2	
Iteration 3	Usually after this iteration the number tends to close into the final. Afterwards the changes are minimal
Iteration 4	
Iteration 5	Suggested location to be used.
X_i	Latitude of the centre of gravity corresponding to the iteration i
Y_i	Longitude of the centre of gravity corresponding to iteration i
Customers served	The total number of clients assigned to that cluster
Volume served	Volume of cases or pallets assigned to the customers from the selected cluster
Costs	
Fixed Depot	Fixed cost of maintaining a DC regardless of the use it has
Var Depot	Cost that is based on the throughput of product that goes through a given cluster
Fixed Trunking	Cost per pallet for transporting it from the plant to the DC
Var Trunking	Cost per pallet per kilometre for transporting it from the plant to the DC
Fixed Local Delivery	Cost per pallet for transporting it from the local DC (COG of the cluster) to the client
Variable Local Delivery	Cost per pallet per kilometre for transporting it from the local DC (COG of the cluster) to the client
Holding	Cost of opportunity that the company assumes because of held inventory
Total Cost	The sum of all of the previous costs

Table 4 Outputs from the Excel model (Source: Author)

Additionally to these datasets, the model also represents the data in a graphical way in order to help the user to analyse the results given.

The first graph is a bar graph that shows a discrimination of each of the costs for each cluster. This helps to firstly compare how each cluster impacts on the total costs; secondly it allows the user to analyse how each cost affects the total cost of each cluster. Figure 11 shows an example of this graph.

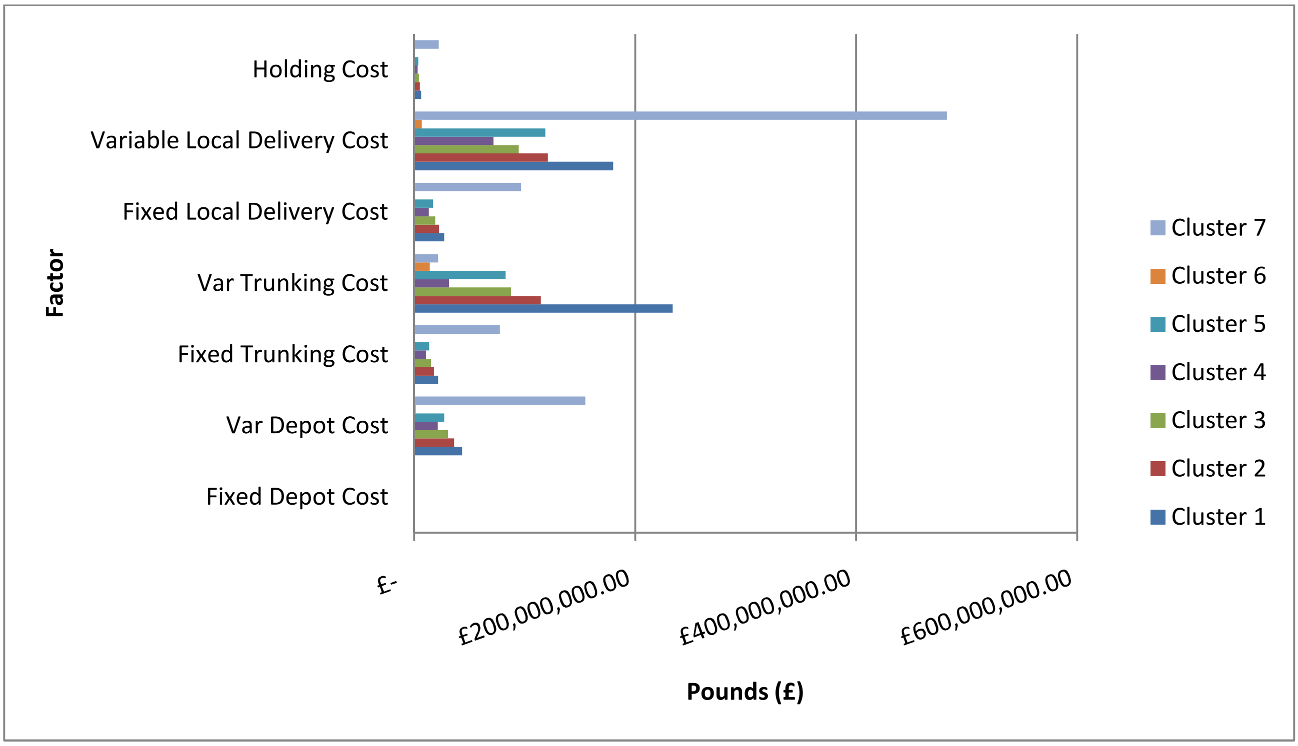

Figure 11 *Bar graph for seven clusters using FCM algorithm on database 3 (Source: Author)*

The second graph that can be found is a pie chart. This chart represents the percentage of participation of the total cost that each cluster represents. This chart is useful when identifying the most expensive clusters in order to concentrate any further analysis on them. Figure 12 shows an example of this chart.

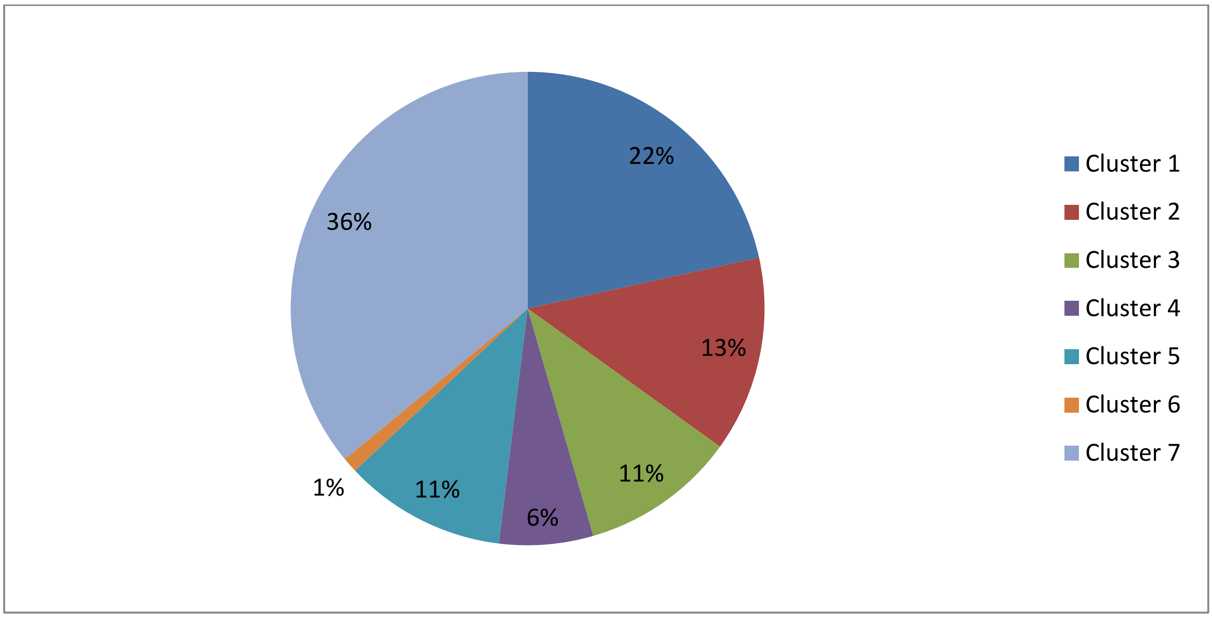

Figure 12 *Pie chart of the participation of each cluster on the total company distribution costs for seven cluster analysis using FCM algorithm on database 3 (Source: Author)*

The third graph is still for cost analysis. It is another pie chart representing how much each of the analysed costs actually affects the total cost. This chart does not analyse each cluster separately, but rather analyses the distribution network as a whole. It is especially useful when analysing the main areas of opportunities that the company has in terms of which costs are making the most impact. Figure 13 shows an example of this chart for the analysis of seven clusters using FCM algorithm on database 3.

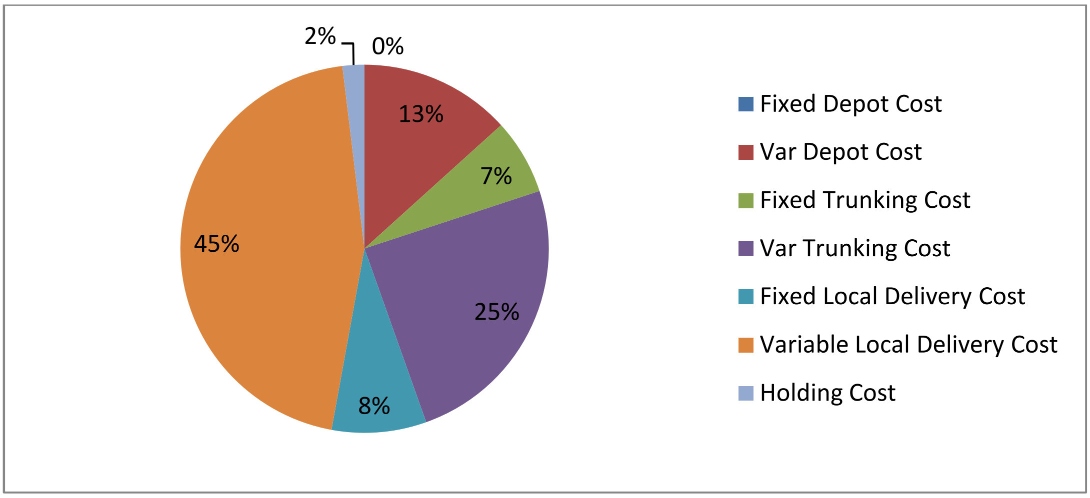

Figure 13 Pie chart of total costs categories over the total distribution network cost for seven cluster analysis using FCM algorithm on database 3. (Source: Author)

The fourth chart is also a pie chart that represents the total number of clients served discriminated by clusters. This analysis helps to analyse the actual size of each of the clusters allowing the decision maker to see if a certain cluster is really necessary, due to the low number of clusters. Figure 14 shows an example of the chart for a seven cluster analysis using the FCM algorithm on database 3.

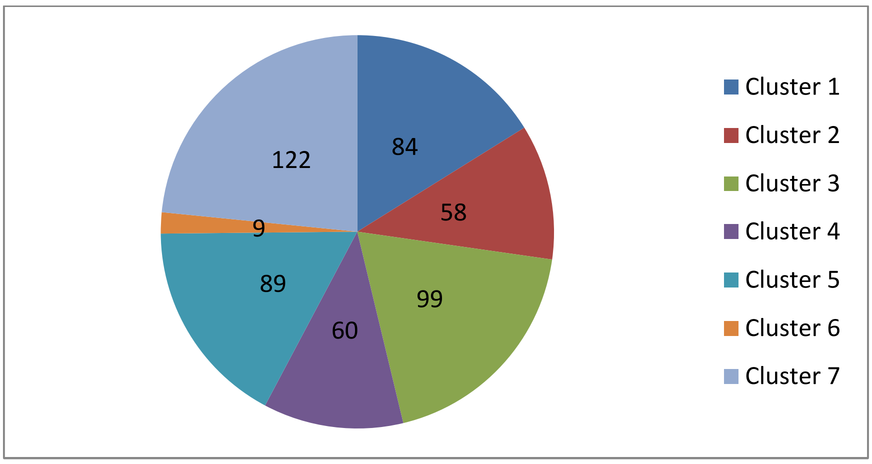

Figure 14 Pie chart of number of clients served per cluster for seven cluster analysis using FCM algorithm on database 3. (Source: Author)

The fifth chart is also a pie chart and it represents the volume of product that is served by each of the clusters. This can also help the company analyse where the greatest demand is concentrated and where it may not be useful to actually open a DC due to low volume. Figure 15 shows an example of this chart. Figure 15 is an example of this chart.

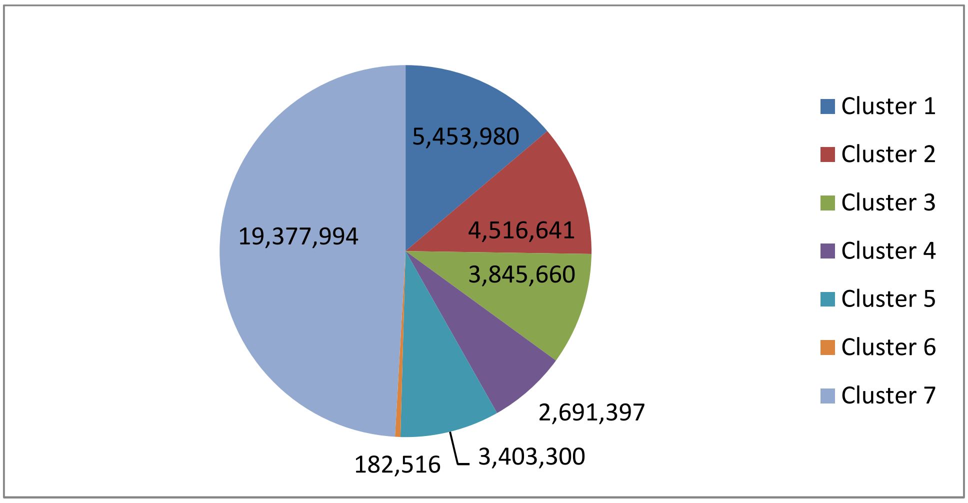

Figure 15 Volume of product assigned to each of the clusters for seven cluster analysis using FCM algorithm on database 3. (Source: Author)

The last graph is a dispersion graph that helps to identify if there are any customers that have high demands that could affect the locations given by the COG model. Figure 16 shows an example of the output for database 3.

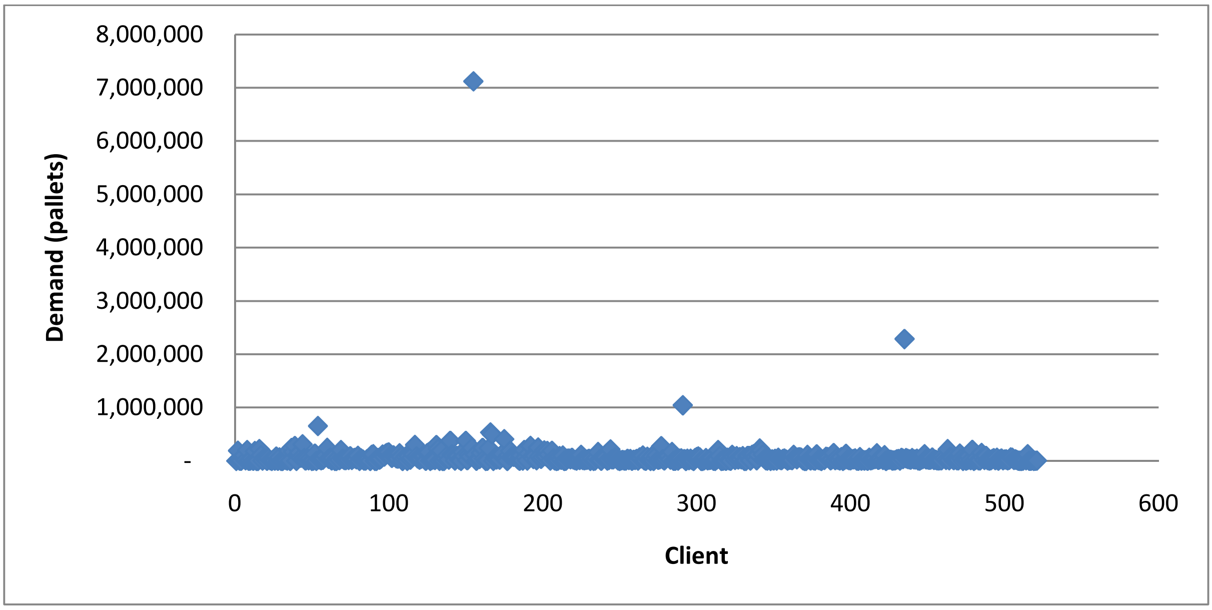

Figure 16 Dispersion analysis of the demand for database 3. (Source: Author)

4. EXECUTION/ANALYSIS

As it has been defined, the model requires a set of inputs in order to provide a solution. For this analysis the input data was taken from the PPP case study. Table 5 shows the input data that were used.

Fixed Local delivery cost	£5.00 per pallet
Variable Local delivery cost	£0.10 per pallet per kilometre
Wiggle Factor	20%
Depot Cost	£2,000 per depot
Variable Depot Cost	£8.00 per pallet of throughput
Safety stock	0.50 weeks of stock
Product Value	£600.00 per pallet
Holding Cost	20%
Fixed Trunking Cost	£4.00 per pallet
Variable Trunking Cost	£0.03 per pallet per kilometre

Table 5 Input data used for the model. (Source: Author)

With this input data and the aforementioned databases, the model was run and the following data were given. The results will initially be given by database in order to discriminate each one separately for later, to do an overall analysis of them.

4.1. Data base 1

Description:

Name of Company: Portuguese Paper Producer (PPP).
Source: Case study of the PPP company.
Number of clients: 54 clients.
Location of the clients: Throughout Europe.
Location of supplying plant: Lisbon, Portugal (Lat: 38.7 North, Long: 9.1 West).
Overall demand: 2,783 pallets.

Results:

The demand of this database has a special feature represented, i.e. there is one dominant demand node located in London. As Figure 17 shows, London stands out as a dominant demand point that could alter the reliability of the COG model. It may be expected that according to Ballou (1973), the COG of the cluster that has London as one of its cities is likely to be towards this location.

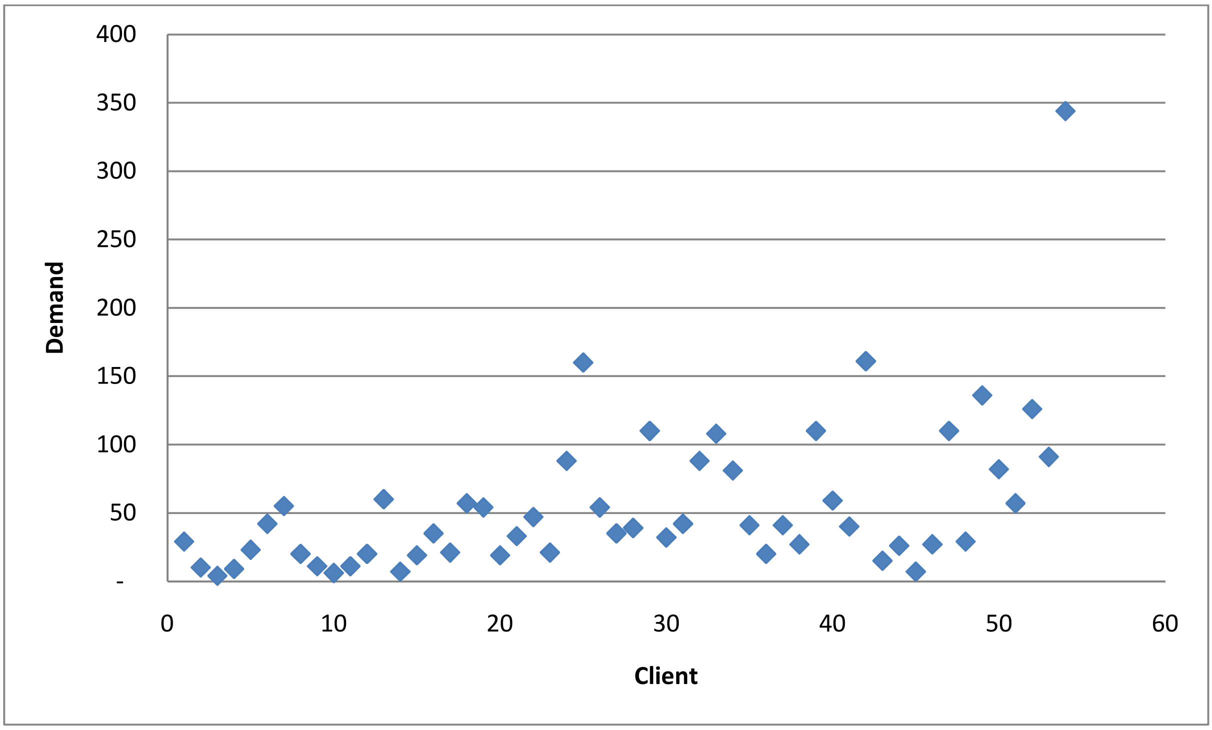

Figure 17 Distribution of the demand points (clients) from database 1. (Source: Author)

When reviewing the data given, it can be seen that the GK algorithm gives a fast improvement from one cluster to two clusters. However, after that point the benefits of this clustering algorithm are flattened. The overall best result is given by the Fuzzy c-mean algorithm. This data initially supports what was found by Esnaf and Küçükdeniz (2009), where with a limited number of clusters the best solution is given by the GG algorithm. However, with more than three clusters, the image changes towards giving a better solution by using the FCM algorithm. Figure 18 shows the comparative evolution of the overall distribution costs by algorithm and number of clusters. It can further be analysed that after the five cluster analysis, the costs tend to stay within a small range. The most savings are made between one and three clusters, even though the best result would be from using six clusters with a Fuzzy c-means algorithm for the clustering of clients.

Figure 18 *Overall distribution costs for PPP database by Clustering algorithm between one and eight clusters. (Source: Author)*

Before analysing the distribution costs resulting from this analysis, it is important to first look at the evolution of the COG of the clients as the clusters are created. Figure 19 shows the evolution of the COG of each cluster when using the FCM algorithm.

The figure shows how the centres of gravity change of location as the number of clusters increase. This change can be explained by the fact that as the number of clusters vary the demand nodes may change from one cluster to another. This variation changes the demand structure of each cluster impacting on the location of the COG.

It can be highlighted that the issue of the COG model not taking into account any geographical restrictions can be evidenced in this figure. In four cases, including the economically best scenario of six clusters, the COG of a cluster falls on a location in the Alps. This limits the possibility of this scenario actually being the most logistically efficient because the costs of building a DC in that precise location can be excessive up to the point of restricting the possibility of applying the result. Apart from this cost, transportation costs me be also restrictive due to the possible lack of transportation routes that can reach the specified location.

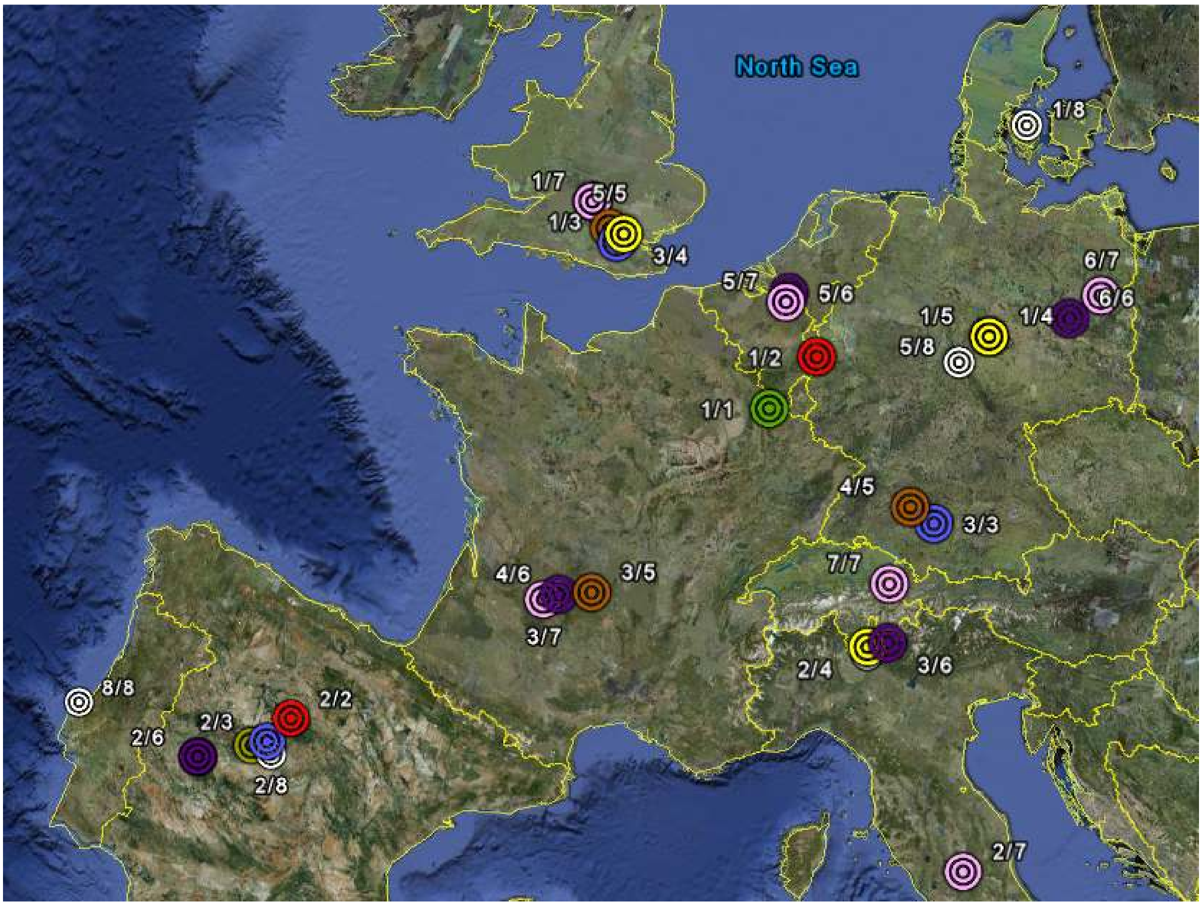

Figure 19 Geographical evolution of the centre of gravity while increasing the number of clusters using FCM algorithm on database 1. (Source: Author; Google Maps)

The colour code used in the graph is the following:

- Green: 1 Cluster Analysis
- Red: 2 Cluster Analysis
- Blue: 3 Cluster Analysis
- Yellow: 4 Cluster analysis
- Brown: 5 Cluster analysis
- Purple: 6 Clusters
- Pink: 7 Cluster analysis
- White: 8 Cluster analysis

Each mark represents where the COG would fall within a certain cluster. It can be seen that as the number of clusters increases, the more the clusters concentrate around similar areas. The main areas of concentration can be identified as London, Madrid, Mid-west France and mid-Germany. However, it is clear that the point where there is a consensus of locations would be the London area. This is an issue that the COG presents, and supports the findings of Ballou (1973).

The most cost efficient is achieved by utilising six clusters calculated by the FCM algorithm. Figure 20 shows the graphical results from applying the clustering algorithm to the database.

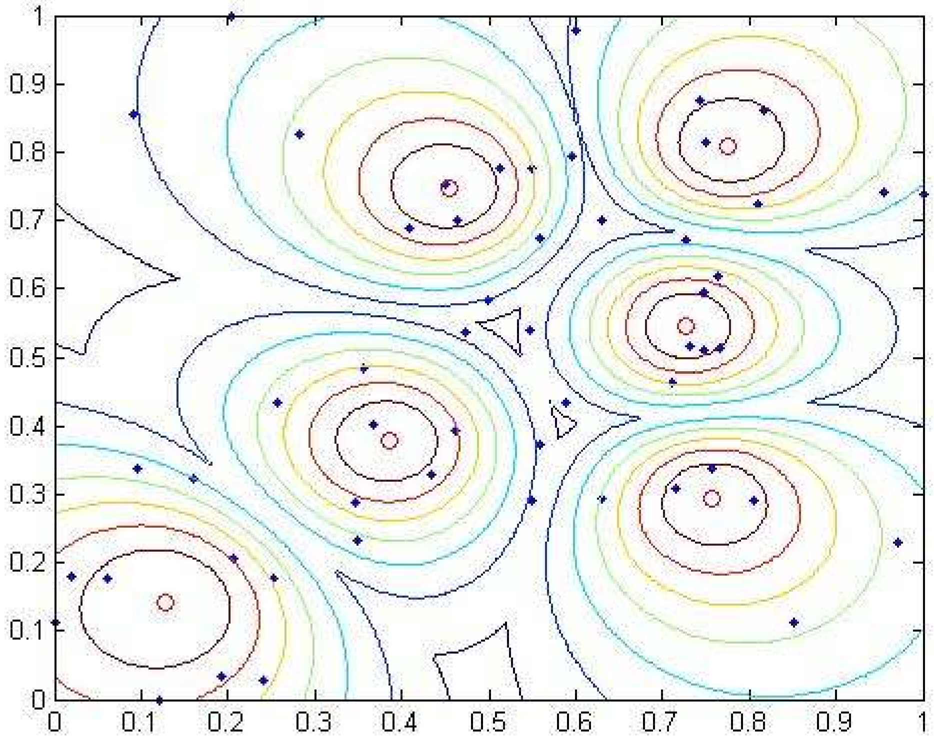

Figure 20 Graphical representation of six cluster analysis using FCM algorithm on database 1. (Source: Author)

It helps to graphically identify the concentration of clients per cluster. Additionally, it helps to see how in general each of the demand points actually fits into these clusters. For example, it can be seen that some of the demand points are so close to the limit of one of the clusters and the adjacent one, that they could, at some point, be reconsidered with regard of the clusters they are assigned and if that could bring some logistical benefits.

Furthermore, Figure 21 shows the geographical location of the centres of gravity for the six clusters with the most cost efficient network given by the analysis. When analysing this option in more detail it is found that the number of clients served per DC may be too low to actually consider this option. Moreover, in the clusters where there is a clear dominant location, the COG tends to be towards that point. Cluster 1 tends to be in London, cluster 3 in Milan and cluster 6 between Brussels and Dusseldorf. The other centres of gravity are located in the middle point of the demand points, because there is no one dominant demand point.

Figure 22 shows the client distribution per cluster. These figures may not be logistically viable even though, from a costs analysis, they may give the best option. To open a DC to serve a maximum of 11 clients, can be complicated and might become inefficient. The decision maker can consider opening small hubs or cross-docking platforms for the locations that have low volume. However, this would imply the COG model to be used for a different purpose different from the one that is Nevertheless it is important to also review the volume of product that would go through each of the DCs. Figure 23 shows this distribution. From this chart it can also be seen that there are two DCs that might not be efficient. Those of clusters 2 and 4 would be DCs with very low volume of throughput - again questioning the logistical relevance of this scenario.

Figure 21 COG locations for six cluster analysis of database 3 using FCM algorithm. (Source: Author; Google Maps)

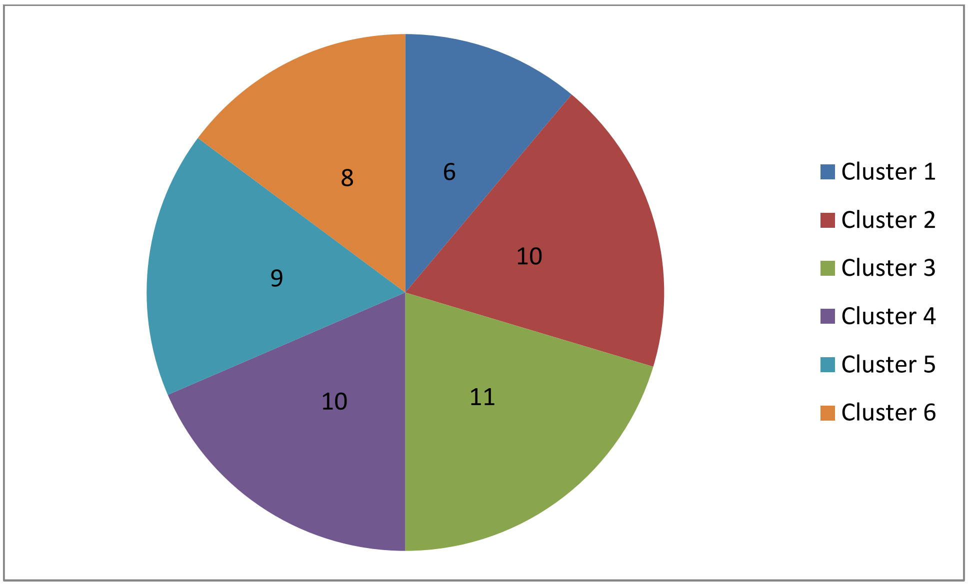

Figure 22 Client distribution per cluster using six clusters with FCM algorithm on database 1. (Source: Author)

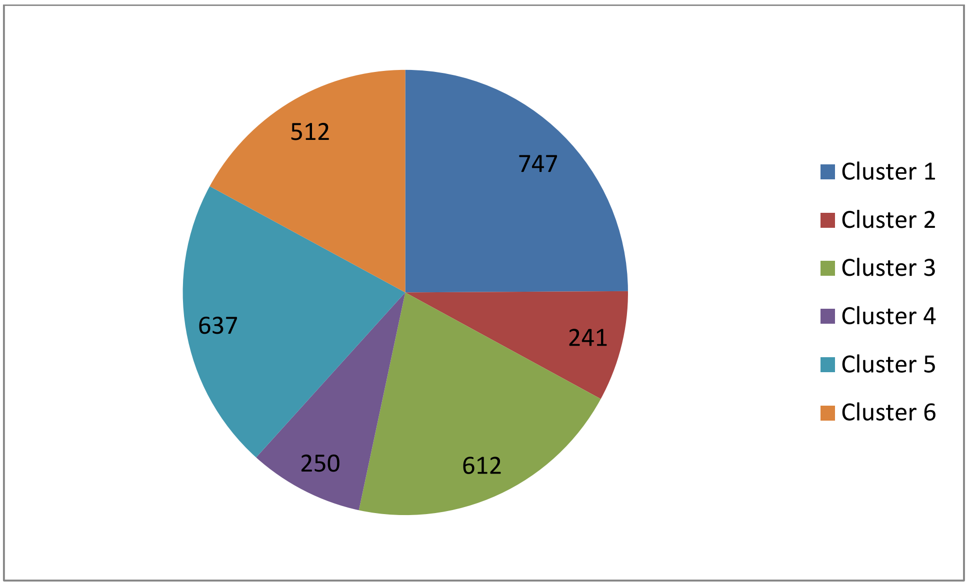

Figure 23 Volume distribution per cluster using six clusters with FCM algorithm on database 1. (Source: Author)

Analysing this scenario from a costing point of view, the cluster that represents the most costs is cluster 5. As Figure 24 shows, clusters 1 and 5 are the two most expensive clusters of the 6, accounting for more than 50% of the total distribution costs for the company. Even though there is a close relation between the throughput and the total costs, the cluster with the highest throughput does not represent the highest distribution costs. The dominating cost for this scenario, as Figure 25 shows, is the variable trunking costs for clusters 1 and 5. This can be explained by the great distances that the product must travel from the plant to each of the DCs.

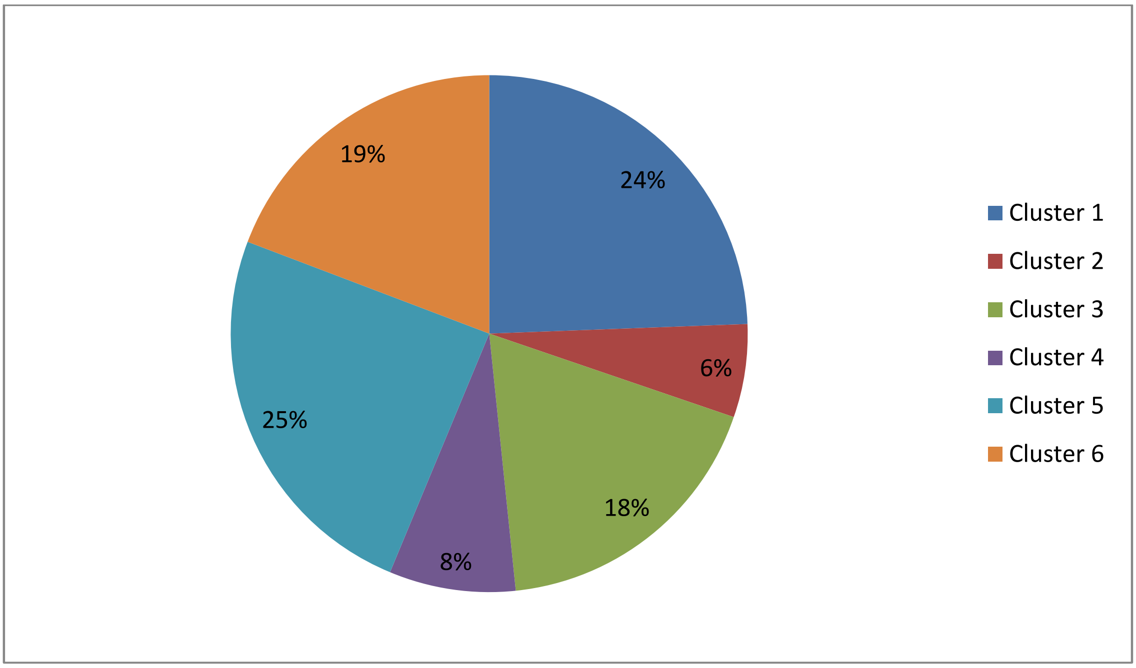

Figure 24 Distribution costs discriminated per cluster of a six cluster analysis of database 1 using FCM algorithm. (Source: Author)

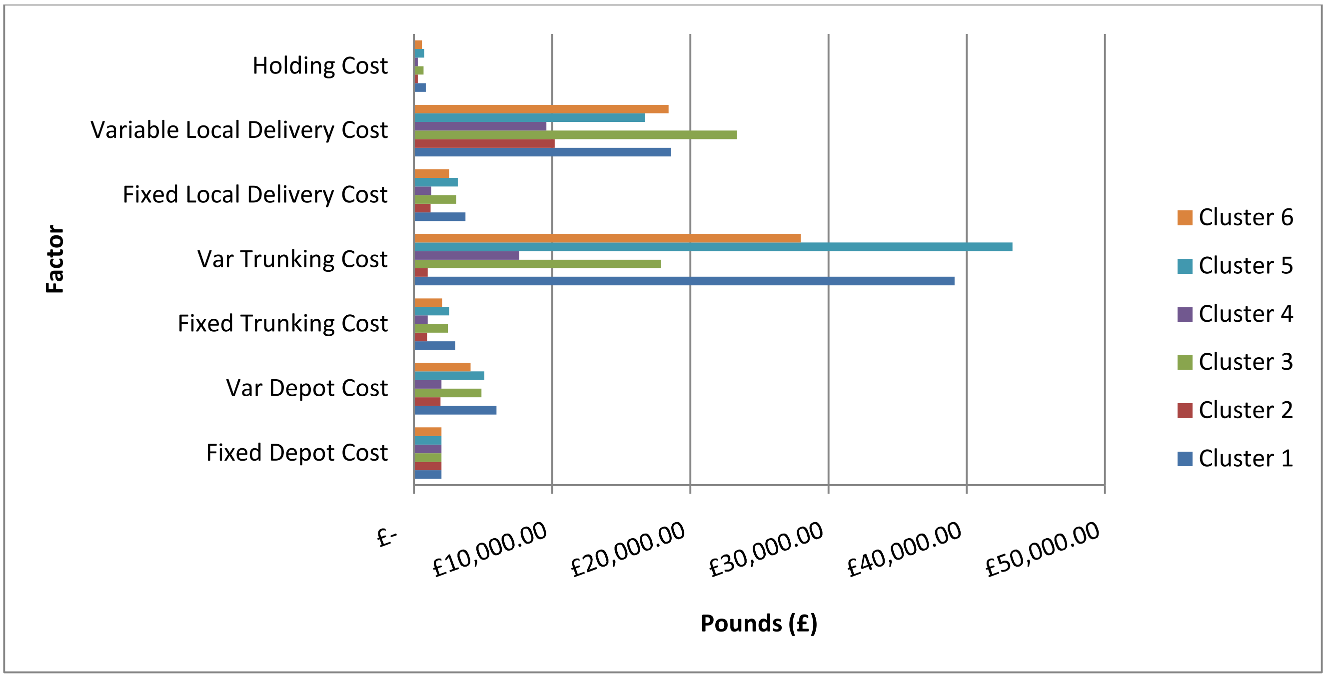

Figure 25 Distribution of costs per factor per cluster on a six cluster analysis of database 1 using FCM algorithm. (Source: Author)

From a more global perspective, if each of the factors is analysed separately, it can be seen that in this case the overall variable trunking costs represent almost 50% of the total cost. The second highest are the variable local delivery costs that account for 32% of the total. Figure 26 shows a clearer picture of this data.

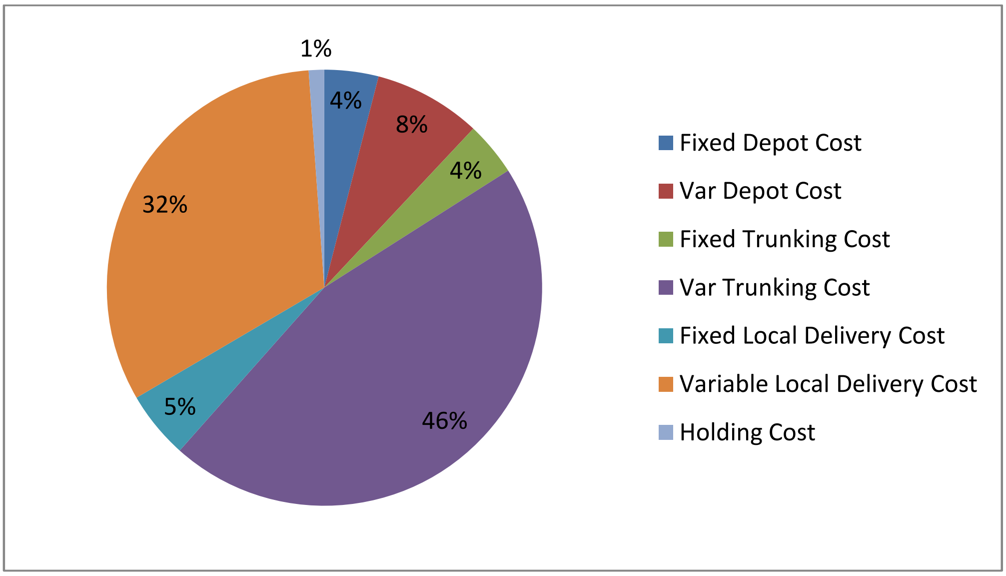

Figure 26 Participation per factor over overall distribution costs for a six cluster analysis using FCM algorithm on database 1. (Source: Author)

A more logistically optimum solution would look for a well distributed throughput for each of the clusters in order to have reasonable parity in terms of efficiency and workload. For this case, possibly the balance between cost efficiency and logistical efficiency can be found using the GK algorithm with a three cluster analysis. This solution presents a great reduction in distribution costs and it makes better use of the general logistical capacities that the company may implement.

Figure 27 shows the location of the centres of gravity for each of the clusters. It can be seen again that there is one location based in the London area for cluster 1. However, the other two COGs do not tend to be towards any single location. This is because within the clients that are attended by these other two DCs, there is no one dominant demand point. Additionally, it can be seen that the viability of implementing this scenario is much higher, given that all three locations are geographically possible. When analysing the distribution of clients, there is a more evenly distributed database than the one from the previously analysed scenario. However, when analysing the demand distribution, Cluster 2 has a considerably lower throughput than that of the other two clusters. In this case, due to the geographical proximity between the location of the DC for cluster 2 and the main plant, it could be considered to simply attend to those clients from this location and just maintain the other two DCs. Figure 28 and Figure 29 show the distribution of clients and demand respectively.

As it was expected, the cluster that includes London tends to be near that area due to the high volume that this specific node has when compared to the rest of the database.

Figure 27 Location of centre of gravity for a three cluster analysis using GK algorithm on database 1. (Source: Author; Google Maps)

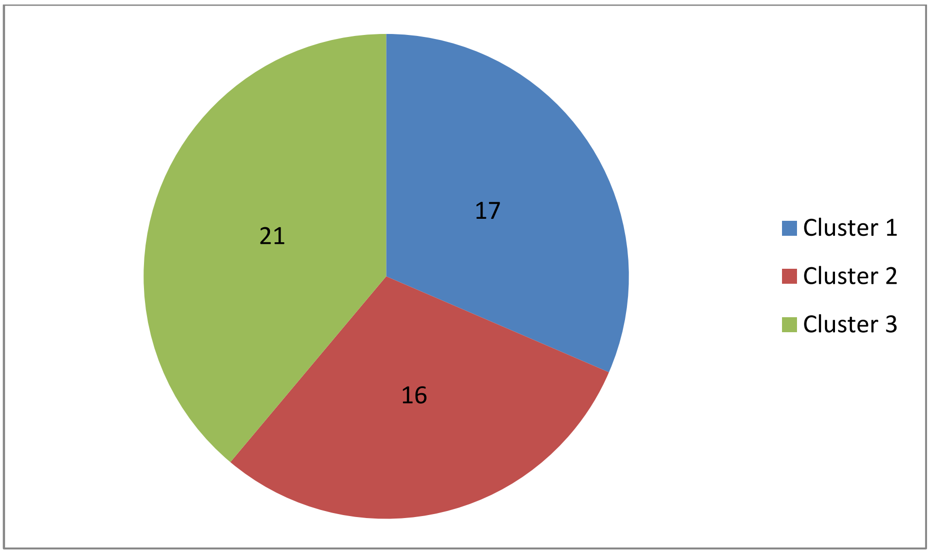

Figure 28 Distribution of clients per cluster on a three cluster analysis using GK algorithm on database 1. (Source: Author)

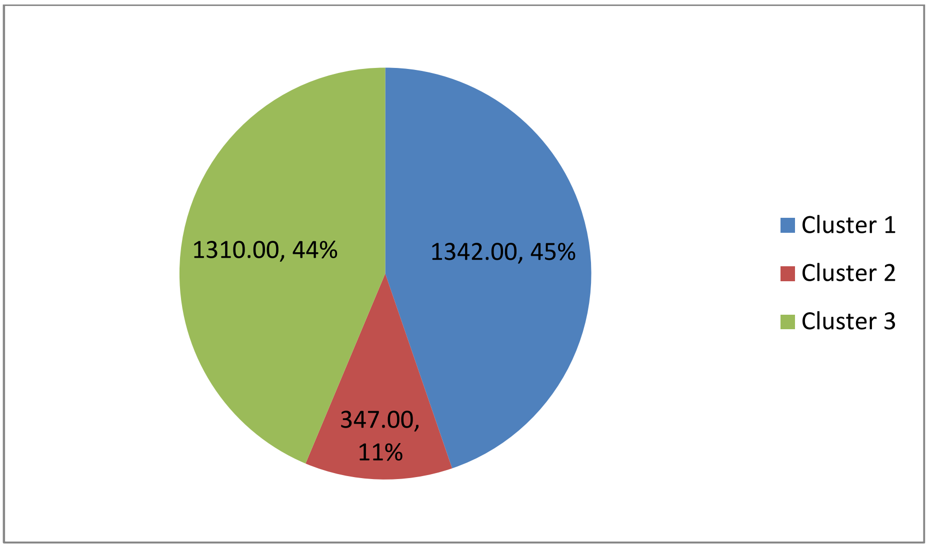

Figure 29 Distribution of demand per cluster on a three cluster analysis using GK algorithm on database 1. (Source: Author)

From a cost perspective, the main costs are represented by variable local and variable trunking costs to and from clusters 1 and 3. In general, cluster 2 is not a very big factor economically since it only represents 10% of the overall network costs.

A big difference in terms of costing between this scenario and that previously analysed is that on an overall basis, variable local delivery costs become the most dominant factor. This can be explained by considering that since there are fewer DCs, the distances from the DC to each client it serves increases. Figure 30 shows the cost distribution for this scenario.

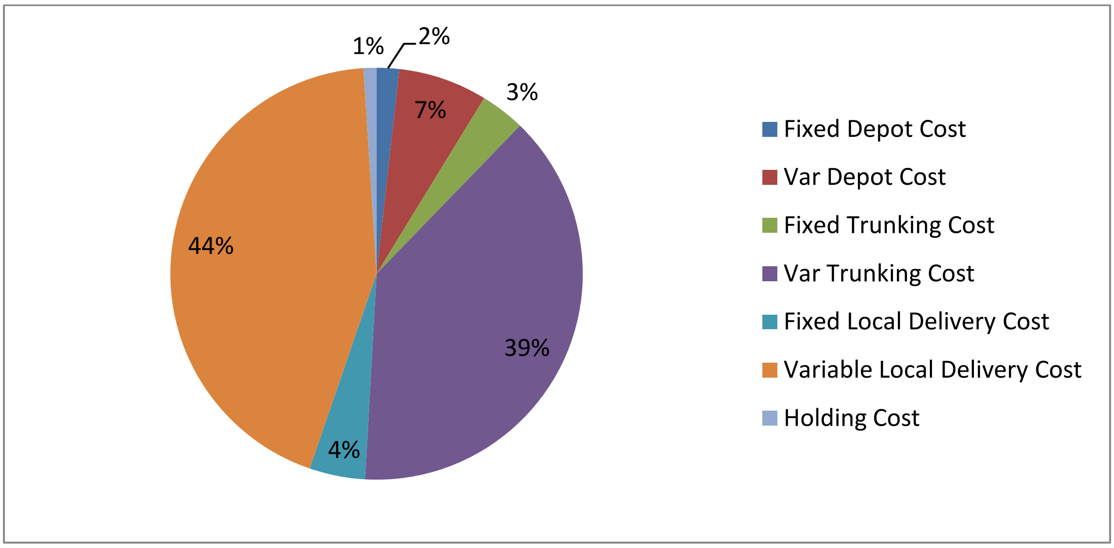

Figure 30 Participation per factor over total distribution costs for a three cluster analysis using GK algorithm on database 1. (Source: Author)

Variable trunking costs become the second highest value. A further impact of this distribution is that if the recommendation is to close the DC located in the COG of cluster 2, then the trunking costs would decrease and the variable local delivery costs would increase.

4.2. Database 2

Description:

Name of company: Unknown
Source: Data base supplied by Dr Andrew Palmer.
Number of clients: 292.
Location of the clients: Throughout Europe
Location of supplying plant: Adro, Italy (Lat: 45.57, Long: 9.97)
Overall demand: 53,191 pallets

Results:

Before starting to analyse the results of the model, it is important to review the composition of the demand nodes. Figure 31 shows the dispersion of the demand. It can be seen that there is no single dominant node that represents a high percentage of the total demand. With this in mind it could be expected that the GOG of the clusters would not be highly influenced by this factor, therefore allowing a higher degree of confidence over the results that the model may show. There are some nodes that may have more impact than others; however, it would be highly improbable that in a given cluster one of the nodes would represent more than half of the total demand of the cluster. This is an issue that has been referenced before, i.e. that was found by Ballou (1973).

When reviewing the costing data shown in Figure 32, it can be seen that the distribution costs for this database maintain a strong decreasing trend as the number of clusters increase. The best results throughout the analysis are given by using the FCM algorithm. In only one case, with four clusters, does the GK algorithm outperform the FCM algorithm.

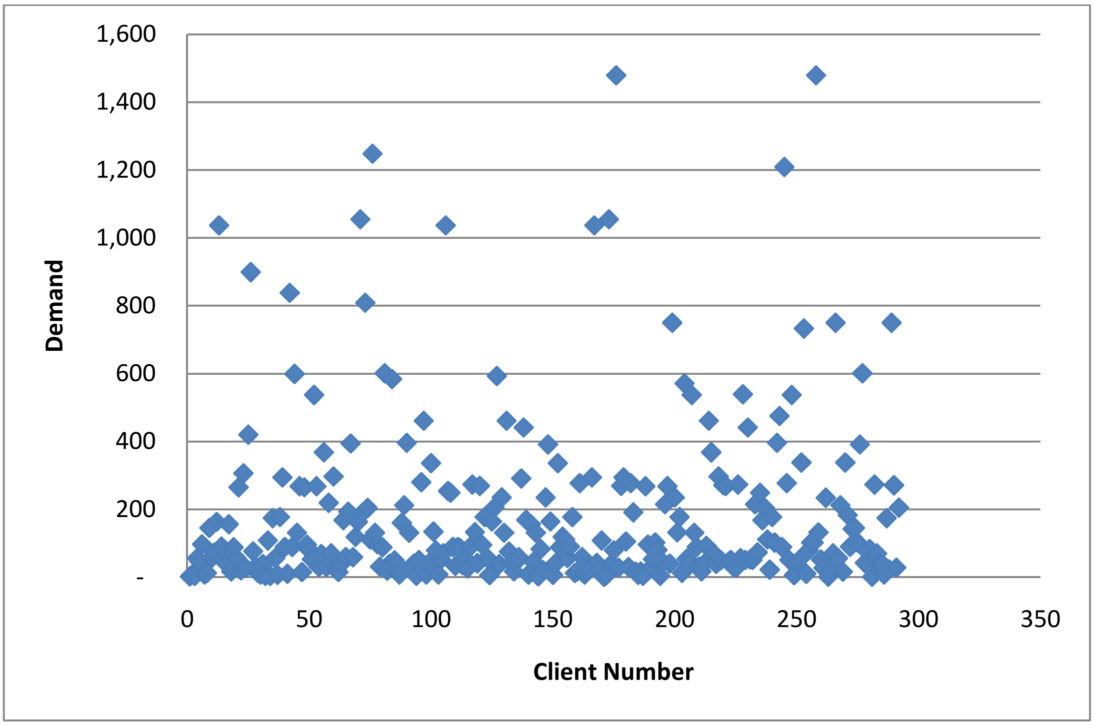

Figure 31 *Distribution of the demand per client of database 2. (Source: Author)*

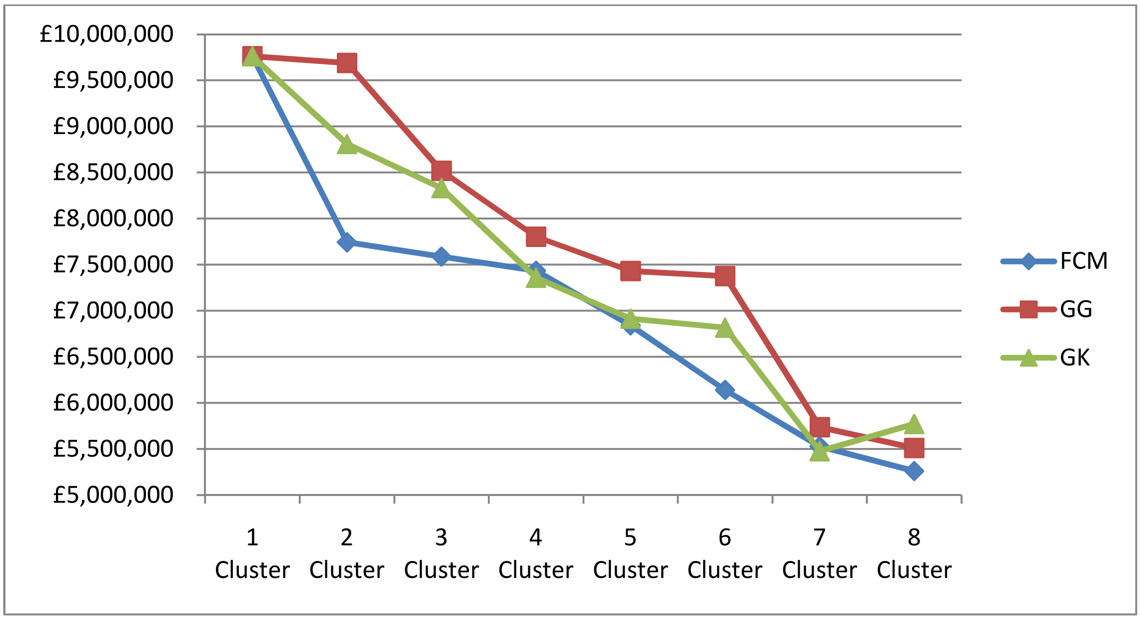

Figure 32 *Total distribution costs for database 2 by clustering algorithm between one and eight cluster analyses. (Source: Author)*

By the time an eight cluster analysis is done, there have been savings for the company of approximately 50% compared to the base case, but the savings are not as great from one scenario to the other. The savings that the company would achieve by only applying the hybrid algorithm with two clusters, compared to the COG model by itself, are approximately

£2,000,000. From there, the savings maintain a low trend but are less steep as the number of clusters grows.

Table 6 shows the difference in cost between one scenario and the next one as a percentage, using the overall best algorithm which is the FCM algorithm. It clearly shows that the savings from going from one cluster to two clusters is 20% - a big impact. From there on, the impact of difference is not as big.

From 1 to 2 clusters	20.68%
From 2 to 3 clusters	2.00%
From 3 to 4 clusters	2.02%
From 4 to 5 clusters	7.96%
From 5 to 6 clusters	10.26%
From 6 to 7 clusters	9.99%
From 7 to 8 clusters	4.81%

Table 6 Variation of costs from one clustering scenario to the next using FCM algorithm on database 2. (Source: Author)

For further analysis, Figure 33 shows the geographical evolution of the centres of gravity of the model, using the FCM algorithm as the number of clusters change. The colour code used on the figure is the same as before, as follows:

- Green: 1 Cluster Analysis
- Red: 2 Cluster Analysis
- Blue: 3 Cluster Analysis
- Yellow: 4 Cluster analysis
- Brown: 5 Cluster analysis
- Purple: 6 Clusters
- Pink: 7 Cluster analysis
- White: 8 Cluster analysis

The COG of the dataset is located in Germany near the border with Switzerland. However, this location is only present in this initial scenario; it never repeats itself throughout the analysis. As the number of clusters increases, there are a few critical nodes that are present in many of the scenarios. These locations are in Spain, Romania, United Kingdom, Sweden and Northern Germany. The other locations vary a great deal between one scenario and the next.

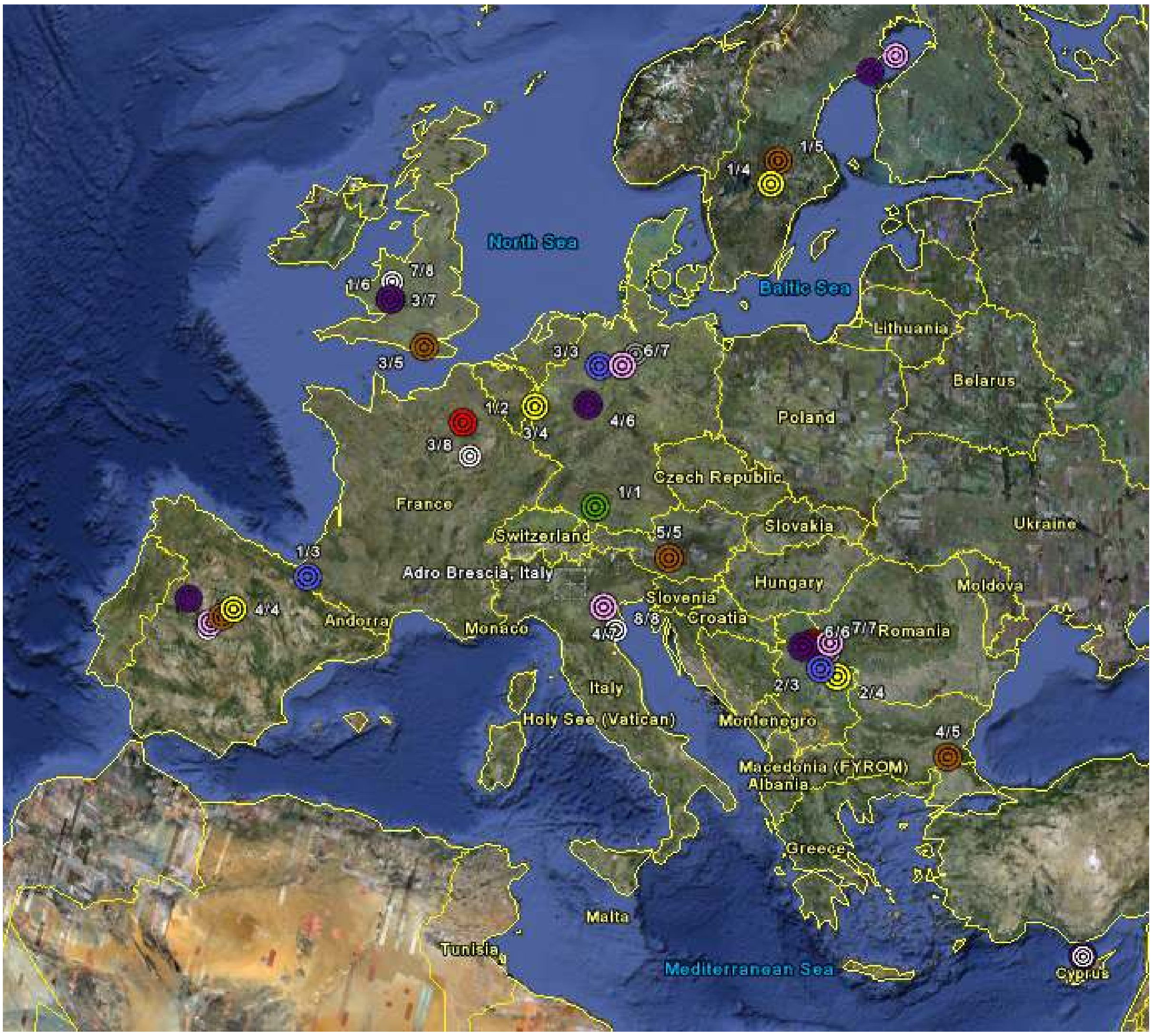

Figure 33 Geographical evolution of the centres of gravity while increasing the number of clusters using FCM algorithm on database 2. (Source: Author; Google Maps)

Since the most cost efficient solution is given by using the FCM algorithm and eight clusters, this will be the scenario of interest for the analysis. Figure 34 shows the distribution of the centres of gravity for each of the clusters. This scenario has a DC inn all of the common locations plus others in Cyprus, eastern France and northern Italy.

However, one of the main flaws that the COG has is evidenced in this scenario. Two of the proposed locations, the one located near Sweden and the one near Cyprus, are given in the Gulf of Bothnia and the Mediterranean sea respectively, making them impossible to apply. If the decision maker were to use the COG location as the exact point where the DC was to be located, than this scenario would be unfeasible. Therefore a different scenario that is not as cost efficient would have to be considered.

Figure 34 Geographical location of the centres of gravity for eight cluster analysis of database 2 using FCM algorithm. (Source: Author; Google Maps)

By analysing Figure 33 again, some other scenarios can be discarded for this same reason. The use of both seven and six clusters has the same issues. It is important to highlight that as can be seen from Figure 32, four other scenarios using different algorithms have a better result than the five cluster FCM scenario. When locating the centres of gravity in all of those options, it is found that in all of them there is a similar issue of having at least one unfeasible location. Figure 35 shows the geographical location of the centres of gravity for each of the four scenarios. Having discarded all of the other options, the most cost effective and geographically feasible is the five cluster analysis using FCM algorithm

a. Seven cluster analysis using GK algorithm with database 2.

b. Eight cluster analysis using GG algorithm with database 2

c. *Seven cluster analysis using GG algorithm in database 2*

d. *Eight cluster analysis using GK algorithm in database 2*

Figure 35 Geographical location of four scenarios using GG and GK algorithms on database 2. (Source: Author; Google Maps)

Figure 36 shows the geographical location of the centres of gravity for the five different clusters using the FCM algorithm.

According to this scenario the DCs would be located in the United Kingdom, Sweden, Spain, Bulgaria and Austria.

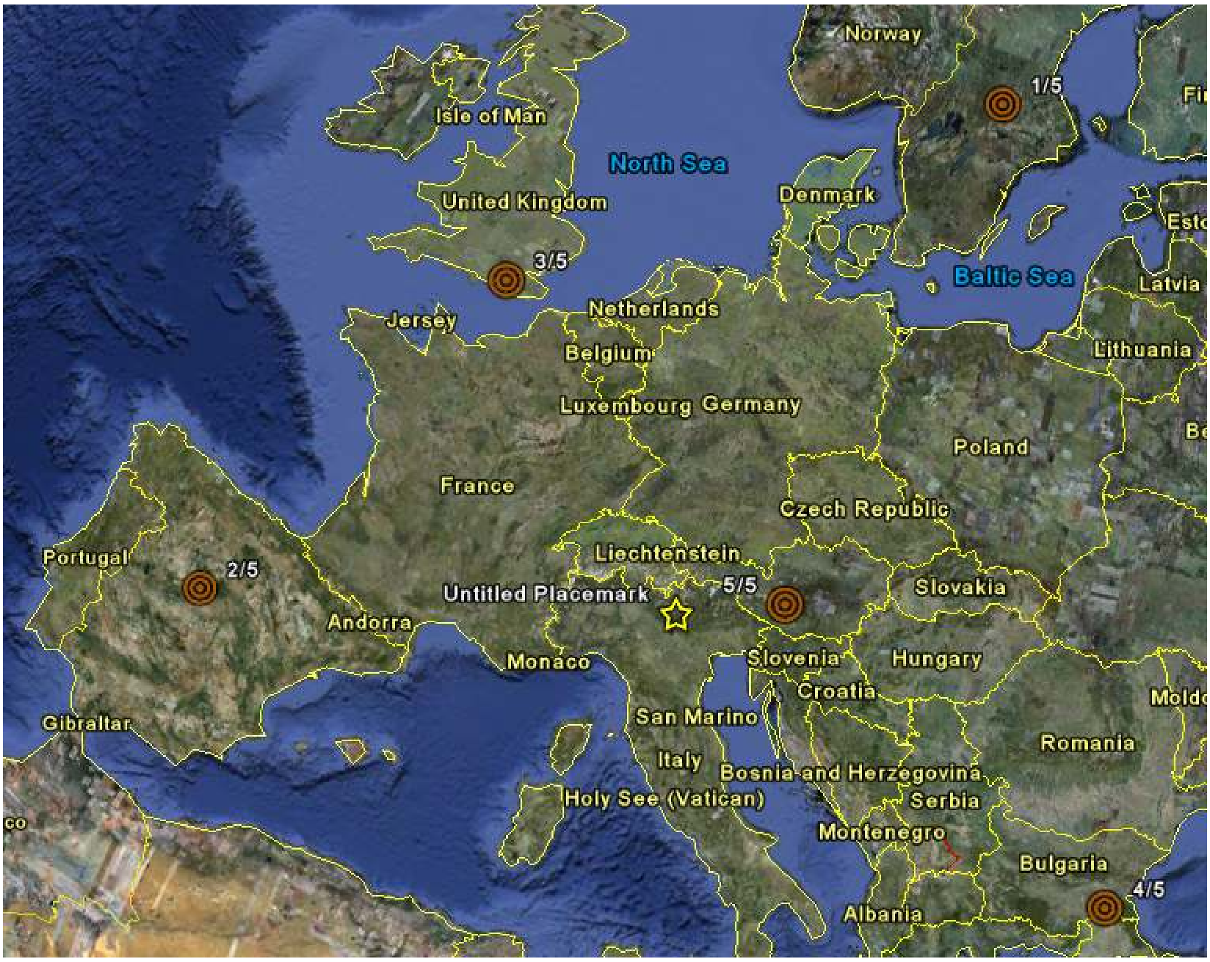

Figure 36 Geographical location of the centres of gravity for a five cluster analysis on database 2 using FCM algorithm. (Source: Author; Google Maps)

The distribution of the demand for each of the clusters is represented in Figure 37. Cluster 5 covers 36% of the total demand. Additionally, it can be expected that the COG for each of the models is relevant since there is no single node in each of the clusters that represents a high percentage over the total amount the cluster covers, as Figure 38 shows. Each of the points in the chart represents a demand node or client.

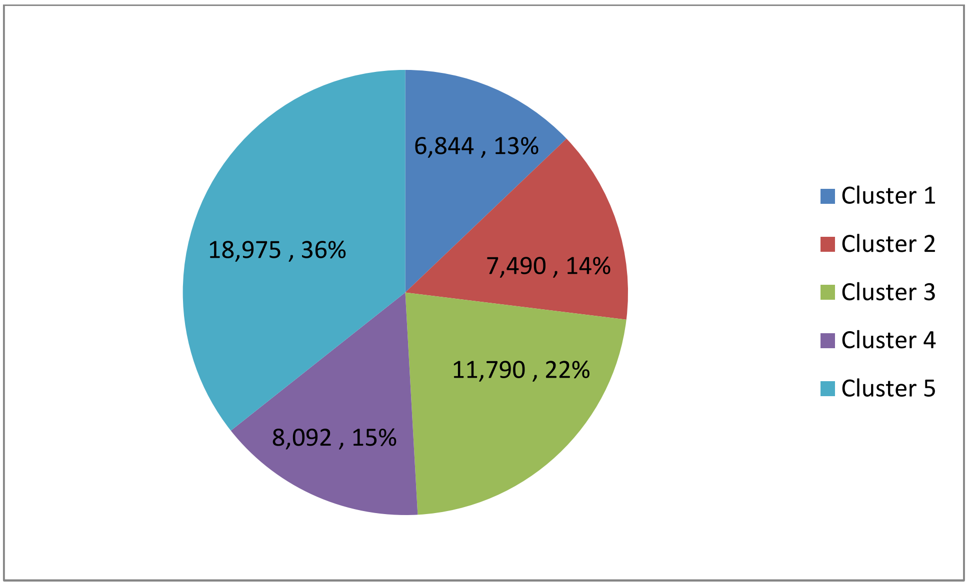

Figure 37 Distribution of the demand by cluster in a five cluster analysis using FCM algorithm on database 2. (Source: Author)

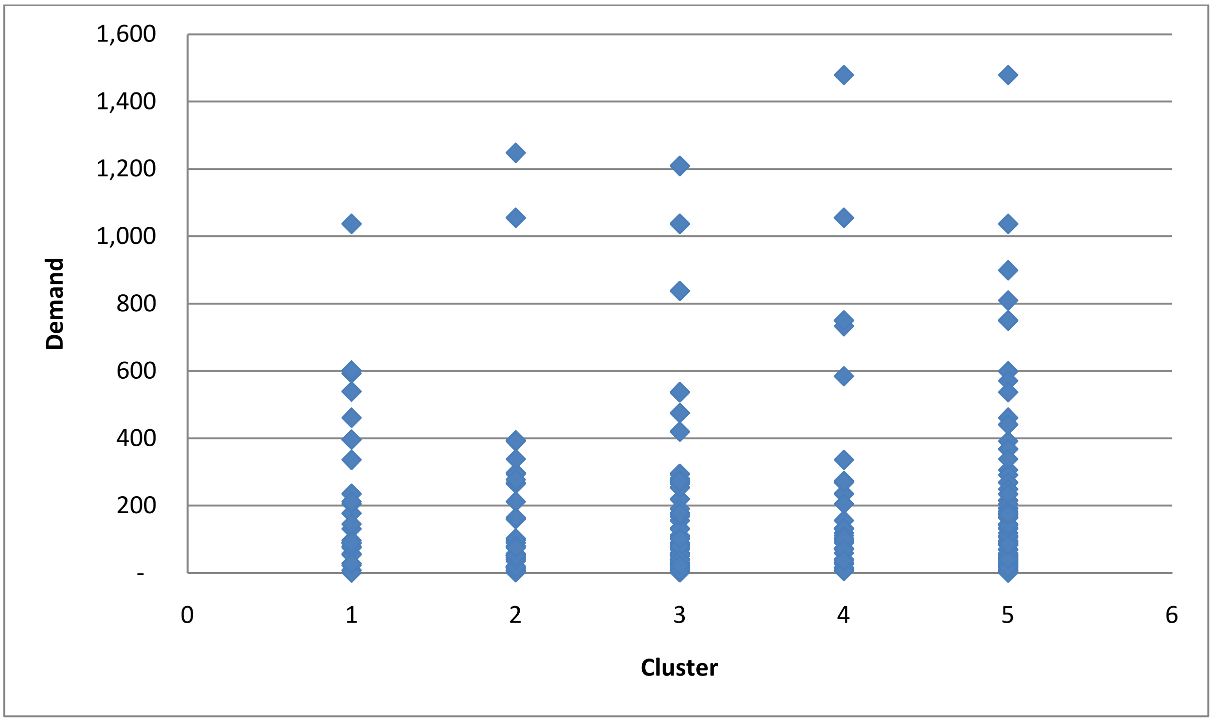

Figure 38 Distribution of the demand per client for each node in a five cluster analysis using FCM algorithm on database 2. (Source: Author)

When analysing the consolidated costs of this distribution network it can be seen from Figure 39 that the dominant cost is the variable local delivery cost, representing 58%. The second highest is the variable trunking costs that represent 28%.

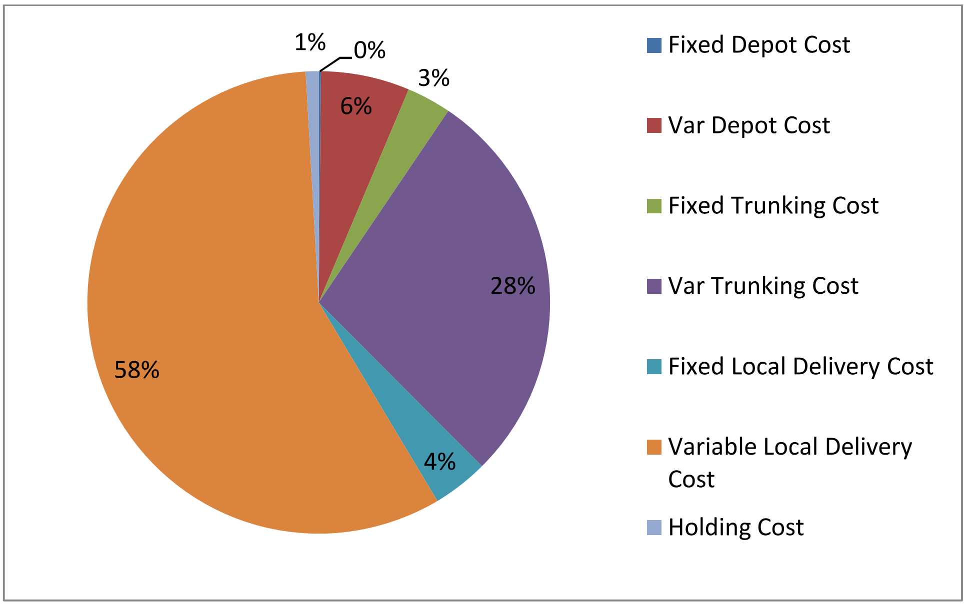

Figure 39 *Consolidated network distribution costs by factor for a five cluster analysis of database 2 using FCM clustering algorithm. (Source: Author)*

When reviewing the specific costs per cluster, it is clearly identified that cluster 5 has an extremely high variable local delivery cost. Figure 40 helps to visualise this issue. The variable local delivery cost is the highest cost for all of the five clusters. However, even though cluster 5 has the highest local delivery cost, it does not represent the highest trunking cost.

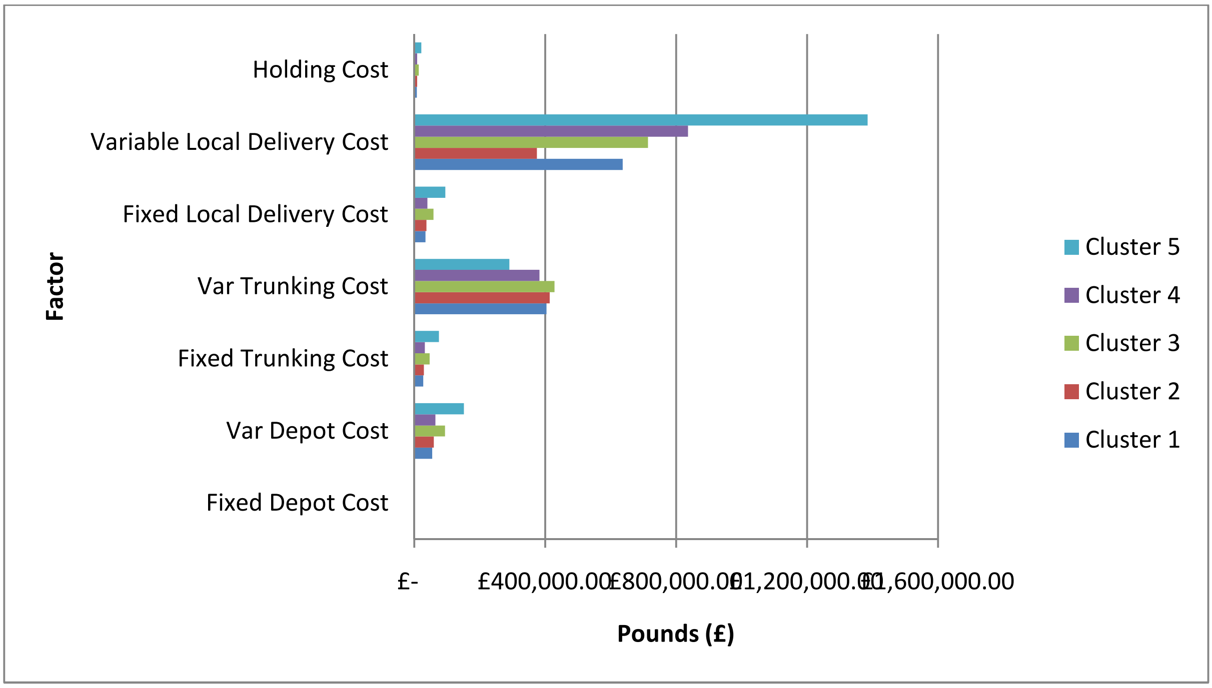

Figure 40 *Distribution of the costs per cluster per factor for 5 cluster analysis of database 2 using FCM clustering algorithm. (Source: Author)*

This contrast can be explained by viewing the location of the plant, represented by the yellow star on Figure 36, and the dispersion of the demand points assigned to cluster 5 by the clustering algorithm that can be seen in Figure 41 (identifiable by the number in the centre circle of each cluster in the figure). The high number of nodes assigned to cluster 5 and the fact that they are widely dispersed within the cluster, indicates that the variable local delivery cost of that node is high. However, the proximity of the COG to the plant makes the variable trunking costs very low, even though the volume is high.

Not surprisingly, the cluster with the highest overall cost is cluster 5 as Figure 42 shows. It can also be seen that even though cluster 1 attends only 10% of the clients and 13% of the demand, it is not the cheapest of the clusters since it is the furthest away from the plant and with a highly dispersed client configuration.

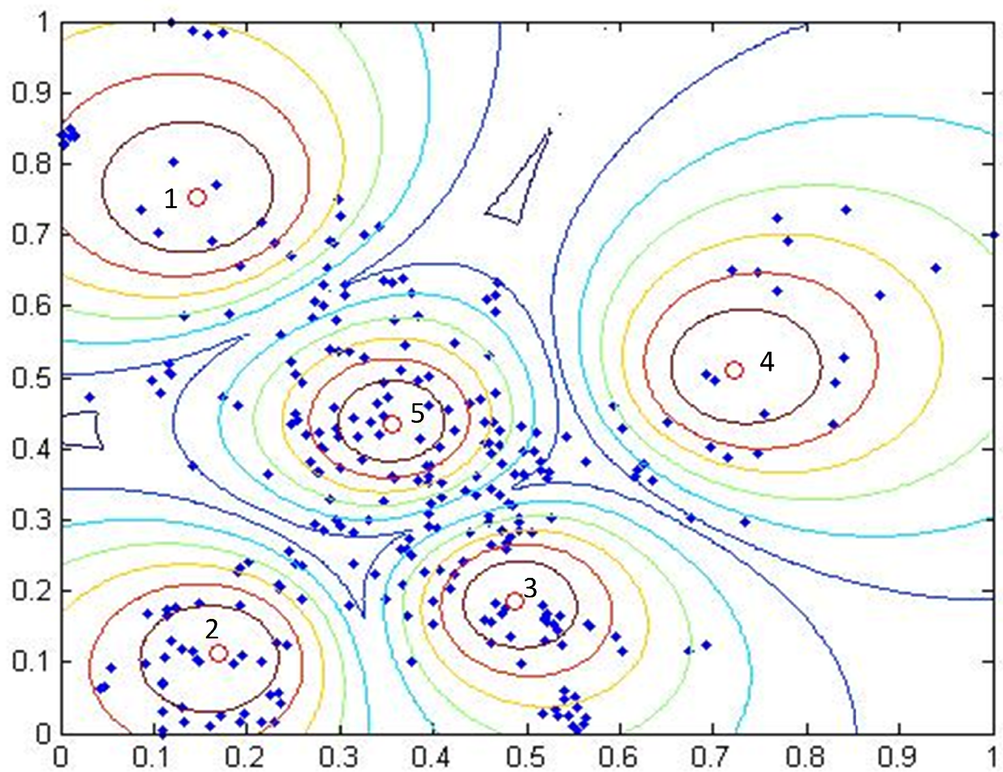

Figure 41 Graphical representation of the five clustering analysis of database 2 using FCM algorithm. (Source: Author)

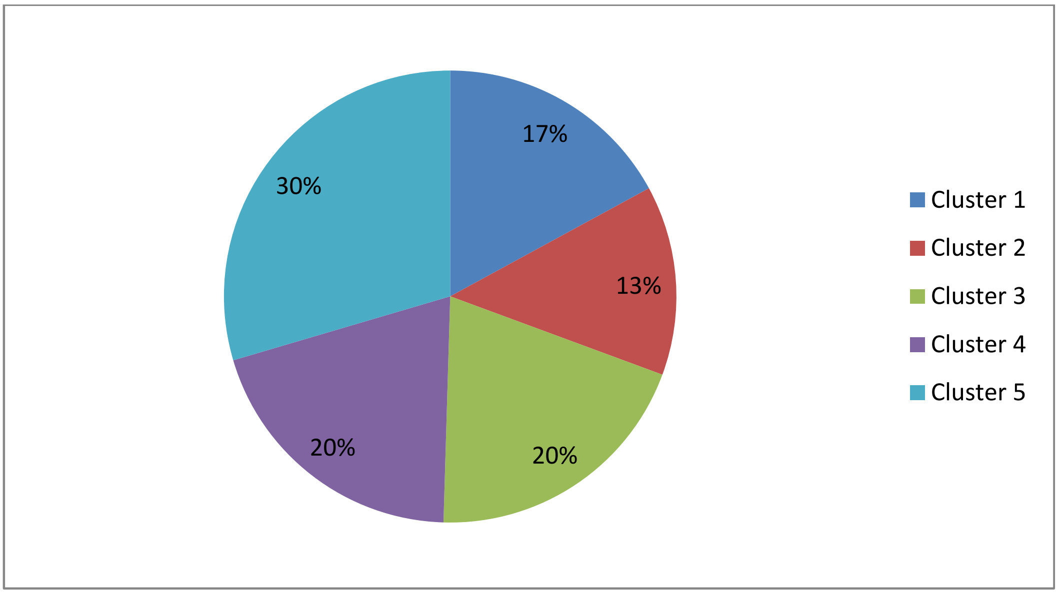

Figure 42 Participation per cluster over the overall distribution cost for a five cluster analysis of database 2 using FCM algorithm. (Source: Author)

4.3. Database 3

Description:

Name of company: Unknown

Source: Data base supplied by Dr Andrew Palmer.

Number of clients: 521.

Location of the clients: Throughout Europe

Location of supplying plant: Adro, Italy (Lat: 45.57, Long: 9.97)

Overall demand: 39,471,488 pallets

Results:

The analysis of the demand per client for this database shows a very dominant client. Figure 43 shows the distribution of the dataset clients where it can be seen that the demand volume per client is very stable under 1,000,000 pallets except for two points, but one in particular. This exceptionally high volume node represents

It can be expected that the cluster that has one of these two nodes will have its COG near or on those coordinates. However, this will be further analysed later in this section. These nodes are located in the proximities of Turin, Italy and in the proximities of Watford, Herts, UK.

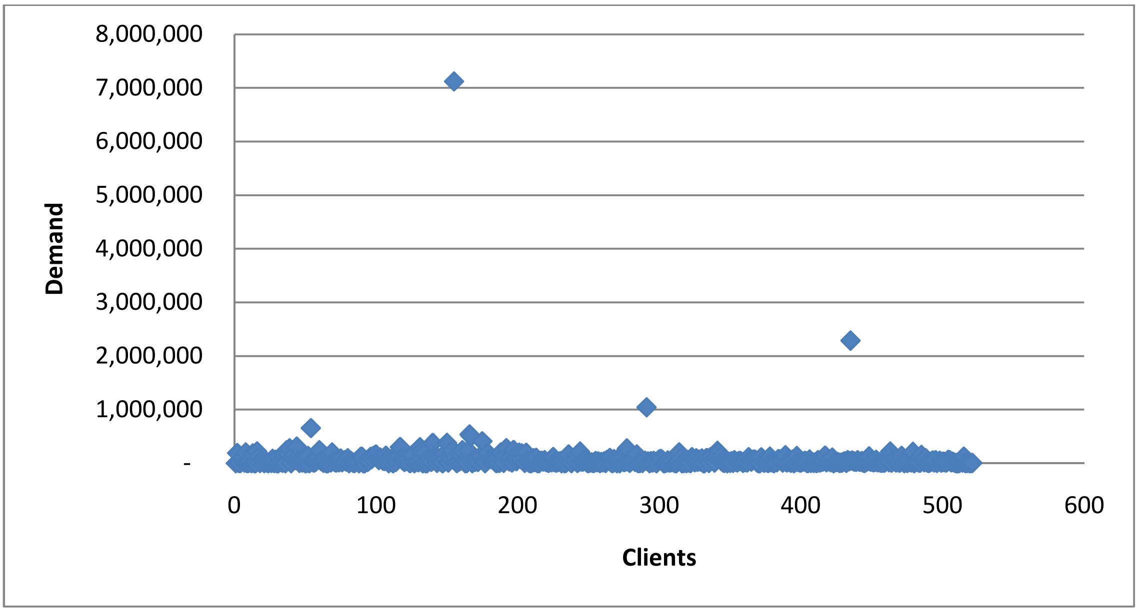

Figure 43 Demand distribution per client for database 3. (Source: Author)

The results from this database are somewhat different from those found on the other cases. First of all, when analysing the evolution of the total costs from one cluster up to eight clusters in both of the previous cases, there was a great saving generated in at least one of the three algorithms when going from the base case (1 cluster) to 2 clusters. On database 1 it was using the GK algorithm, with database 2 it was present using the FCM algorithm. However, as can be seen in Figure 44, with database 3 by changing from one cluster to two there is no real change in either of the algorithms.

Figure 44 Total distribution costs for database 3 by clustering algorithms between one and eight cluster analyses. (Source: Author)

A second difference is that, even though the FCM algorithm has an overall best result, it does not give the optimal solution. With this database the optimum solution is found using the GK algorithm with a seven cluster analysis.

Another clear characteristic is that the costs tend to be towards a value of between £2,400,000,000 and £2,200,000,000. The FCM algorithm is the fastest to get to this value reaching it with four clusters. The other two algorithms require more clusters to reach that same range. This is where decision makers must analyse the information. The complexity of managing four DCs could be expected to be less intensive than that found when working with seven DCs.

However, for this analysis, decisions will be based only on the information given by the model, therefore the best solution is found using seven clusters with GK algorithm.

Figure 45 shows the geographical location of the centres of gravity for the selected scenario.

Figure 45 Geographical location of centres of gravity for a seven cluster analysis using GK algorithm with database 3. (Source: Author; Google Maps)

The yellow star once again represents the location of the main distribution plant from where all of the products are dispatched to the DCs. If the distribution of the demand per cluster, as shown in Figure 46, is analysed, it can be seen that cluster 4 attends the Turin location.

When reviewing the location of the COG of this cluster in Figure 45, cluster 4 is located in exactly the same coordinates as the Turin node.

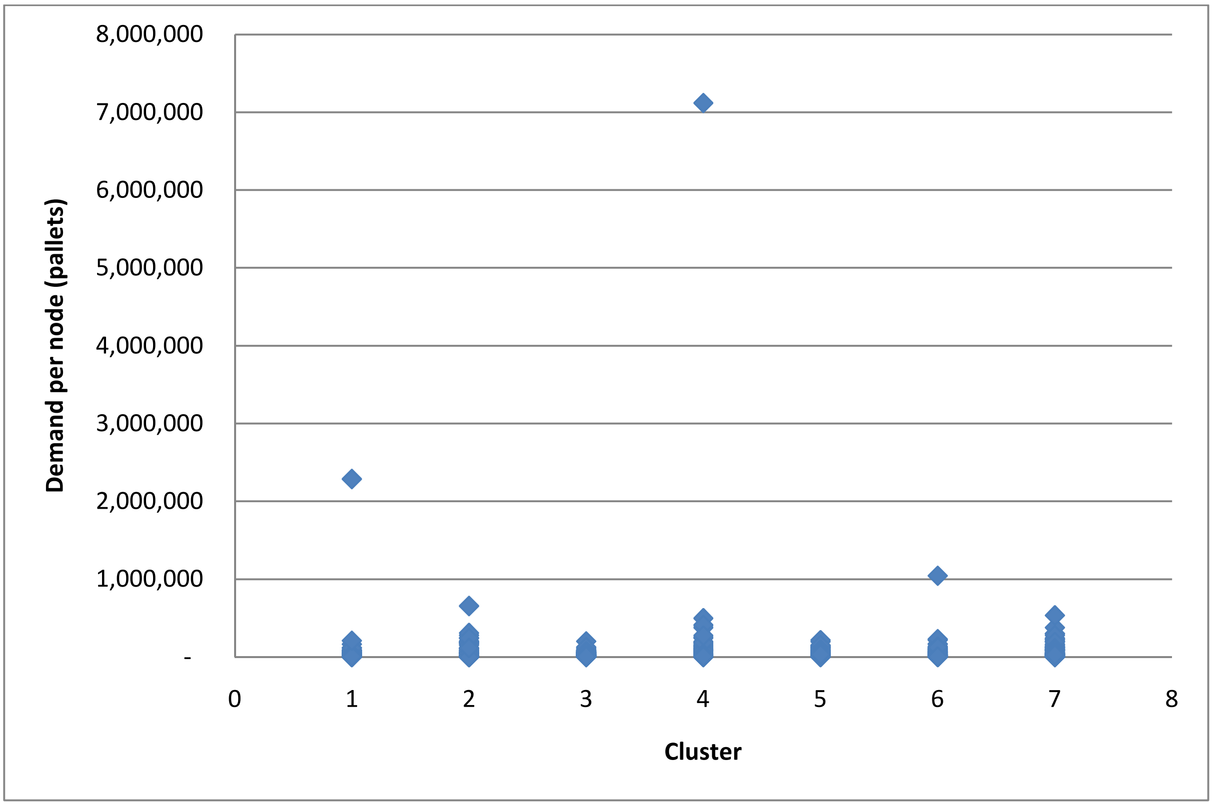

Figure 46 Demand distribution per node in a seven cluster analysis of database 3 using GK algorithm. (Source: Author)

The same issue can be found with cluster 1, where the second highest node is attended from, and its COG, as seen in Figure 45, is in that same location.

To analyse the impact that these two nodes have over their respective clusters, Figure 47 shows with a yellow pin facsimile the geographical location of the centres of gravity of clusters 1 and 4 with the assumption that neither the Turin nor the Watford Herts nodes exist. The location of cluster 1 is dramatically impacted by the fact of the existence of that one particular high volume node. It goes from being located in southern UK to northern France.

The location of cluster 4 goes from being near Turin to being closer to Milan. Even though it has an impact, this is not as dramatic as the change seen with cluster 1. This may be explained by the distribution of the clients. Most of the clients attended by cluster 4 are located in the vicinities of Turin. On the other hand, the locations of the clients attended by cluster 1 are in the majority within in continental Europe.

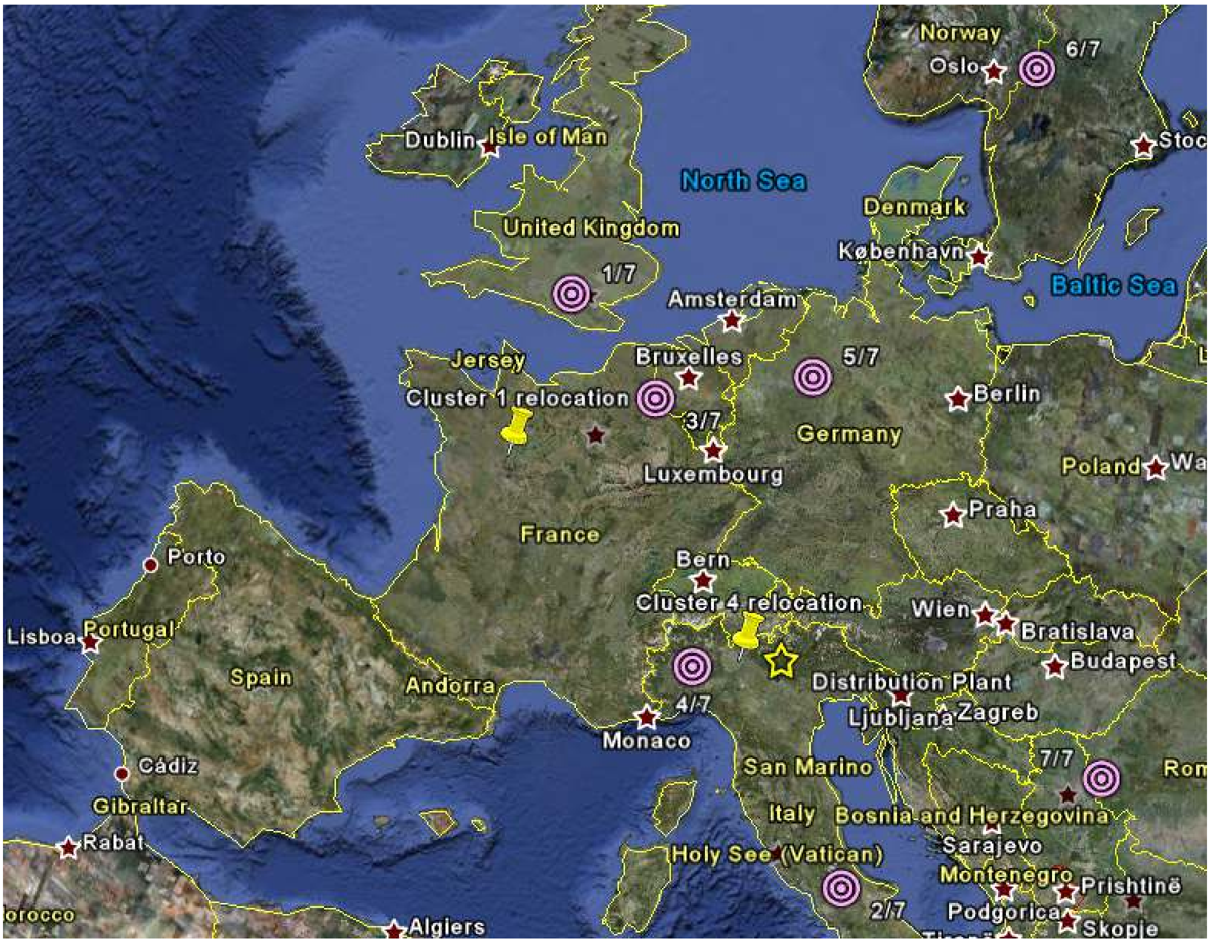

Figure 47 Location of the centres of gravity for a seven cluster analysis of database 3 using GK algorithm with the assumption of two nodes not existing. (Source: Author; Google Maps)

The total volume assigned to cluster 4 comes to a total of 13,378,096 pallets, which means that the Turin location represents more than 50% of it, supporting the theory presented by Ballou (1973).

In this scenario the number of clients is similarly distributed between all of the clusters; however, the volume is not. Figure 48 and Figure 49 show how these two traits behave in the scenario.

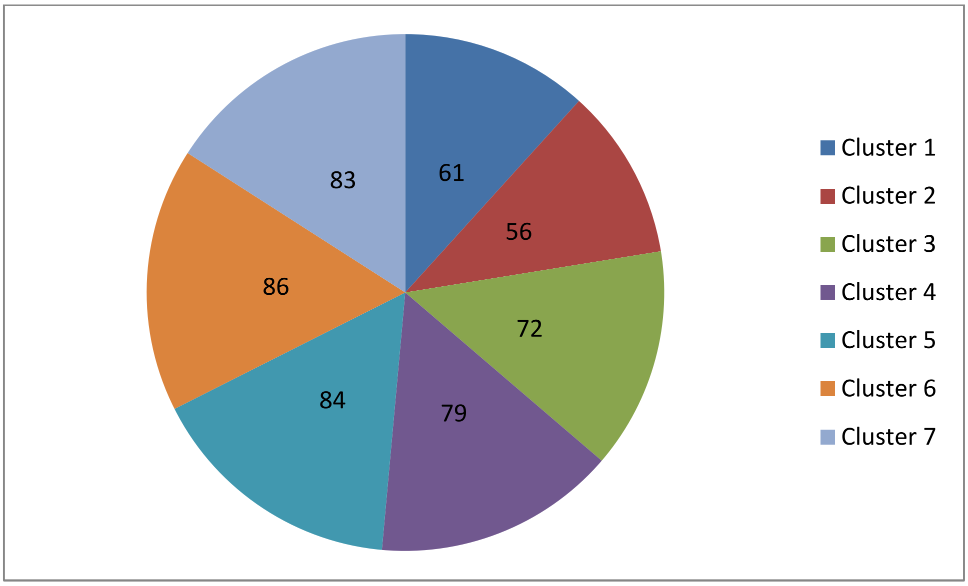

Figure 48 Number of clients per cluster in a seven cluster analysis of database 3 using GK algorithm. (Source: Author)

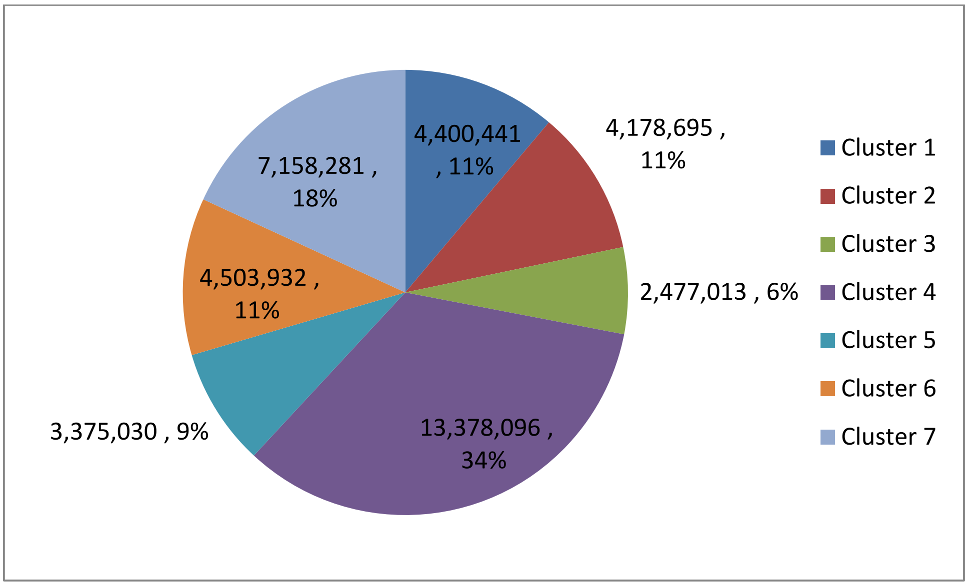

Figure 49 Demand distribution by cluster for a seven cluster analysis of database 3 using GK algorithm. (Source: Author)

When starting to analyse the costs of the scenario, it can be found that the local delivery cost of the whole distribution network is the dominant cost as can be seen in Figure 50. Figure 51 additionally shows that the variable depot cost, a factor that had not been represented in the previous databases, is now important, especially in cluster 4.

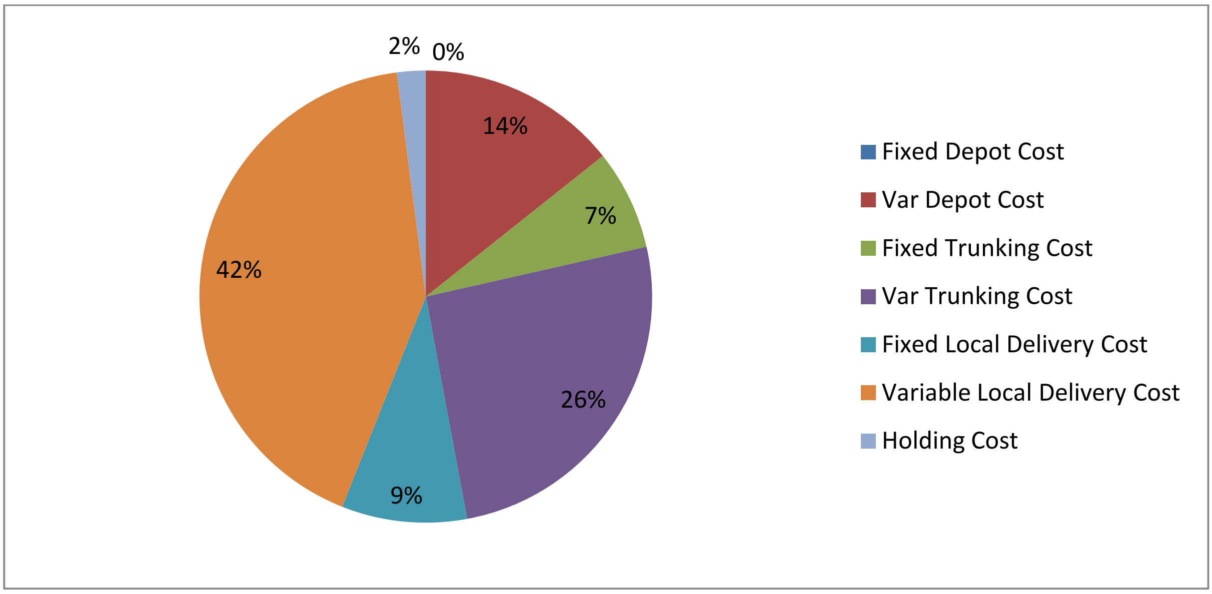

Figure 50 Impact of consolidated cost factors over the overall distribution costs for a seven cluster analysis of database 3 using GK algorithm. (Source: Author)

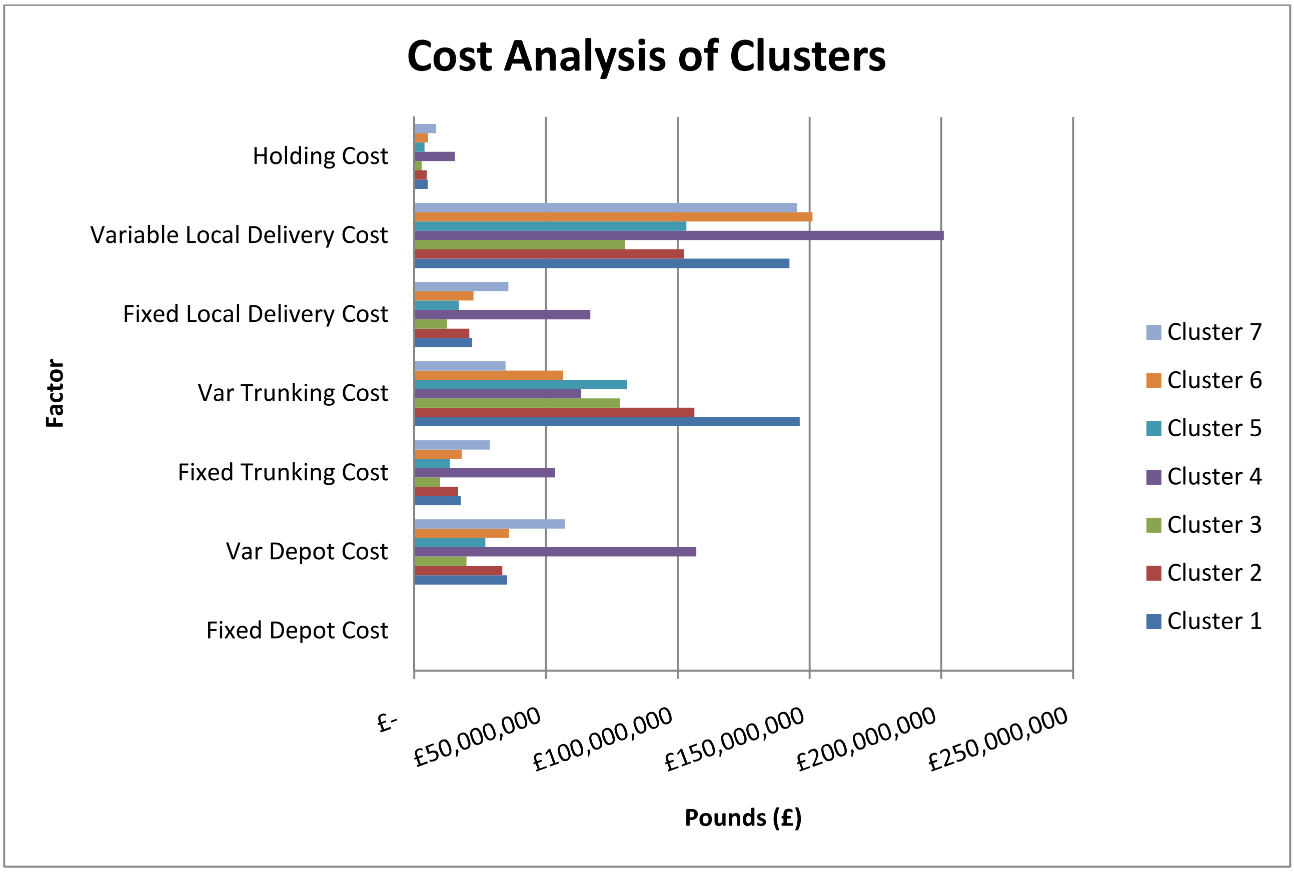

Figure 51 Distribution of cost factors per cluster on a seven cluster analysis of database 3 using GK algorithm. (Source: Author)

5. DISCUSSION OF RESULTS

In order to achieve the objectives set for this research it was necessary to begin by understanding the relevance of the Facility Location Problem (FLP) within the supply chain. As the importance of strategic supply chain management grows within industries, the decisions that encompass this area have a greater relevance to and higher impact on the competitiveness of the company. When planning the supply chain, the design of the distribution network that will support the company’s operations is a critical aspect. It can make the difference between being competitive or not. Therefore, the importance of the decisions that are taken in this process are especially relevant.

Part of distribution network strategic planning is the location of the DCs and other facilities that will allow the company to serve its customers to their expected service level in the most cost efficient way. Here is where the FLP falls into the picture of relevance in the supply chain. The FLP is a wide ranging problem that tries to model the reality of the distribution necessities for each industry.

The FLP has been widely reviewed and researched in past years. Starting from seminal works by Weber (1909) and Cooper (1961) many other works have been created since then. However, there has yet to appear a model that combines a significant number of the multiple variables that the FLP has. As a consequence, research has been created with a view to attacking one or two specific issues at a time. The FLP is divided into different categories depending on the variables that impact the supply chain under study.

One model for solving the FLP that is frequently used is the Centre of Gravity (COG) model. The definition of this model and its strengths and weaknesses are the main outcome of the second objective. The COG model applied to the FLP is a simple model that identifies the point where the average distance between that point and the weighted demands of each demand point of the system is at a minimum. The model can be used in various situations, but has some fundamental characteristics that must be taken into account. One of them is the fact that the problem that is to be solved has to fall within the following categories from the classification of FLP problems described by Eiselt and explained in section 2.4.1:

- Continuous space: the location of the plant can be at any point in the space regardless of any limitations.
- Exogenously specified: The number of DCs that are to be opened is predefined.
- Deterministic: All of the clients’ locations are known as well as their demands.

- The customer does not choose which location he will use to purchase his product.

This model is easy to use and it brings to the decision maker a tool for solving the FLP.

However, due to the same simplicity of the model, it has some identified flaws that can impact on the optimality of the solution. The literature identifies two specific weaknesses:

- It can be highly vulnerable to the existence of one dominant client in terms of volume within the database.
- It can be highly vulnerable to the existence of a client that is significantly geographically far from the rest of the clients of the database.

These two represent a fundamental problem to the COG; however, other weaknesses can also be found in the assumptions of the model. Assuming that a DC can be located anywhere is something that is not true; having to input the expected demand in the long term brings high uncertainty to the analysis.

One way of overcoming the issue of having clients geographically apart is to divide the clients into groups and plan on attending them from different locations. This is something that has been done in the past using what is known as clustering algorithms. Clustering algorithms are a mathematical way of dividing sets of data into subsets that have some characteristics in common. This technique is used in many fields including medicine, biology and computer networking.

In this case, the clustering algorithms look to divide the databases into groups of nearby clients and by doing so the distance issue is expected to be overcome. In order to review this, and solve the third objective of this work, three different databases of different sizes were analysed using three different clustering algorithms. Additionally, each clustering algorithm was calculated using from one up to eight clusters.

The idea of having three databases is to analyse the impact on distribution networks of different sizes, and see if the results from the COG model, in terms of distribution costs, can be improved by using the clustering of the clients.

The results show that all of the distribution network costs are decreased by using the clustering algorithms to accompany the COG model. Furthermore, the selection of the clustering algorithm also has an impact on the optimality of the final result. On all of the databases the Fuzzy c-means (FCM) clustering algorithm gave a general best result, even though in some specific cases it was outperformed by the Gustafson - Kessel (GK) algorithm. The Gath – Geva (GG) algorithm presented the worst overall results and therefore should not be considered as an option.

In all of the cases, the most dominant cost is the transportation costs, being local delivery cost the overall highest. The second highest was the trunking cost. As the number of clusters increased it was found that the impact on this cost structure changed as the trunking costs increased and local delivery costs decreased. This is explained by the fact that by having more DCs in the network the distances that have to be travelled for local delivery decrease since the client is closer to a certain DC.

The research proved that some of the aforementioned weaknesses do affect the COG model in terms of its realistic use for decision making. In database 2 the fact that the model considers the possibility of geographically locating the DC at any point in the plane denies the decision maker the use of the most optimum solution since one of the DCs is located in a non-admissible location in the sea. In a not so extreme measure this same issue is found in database 1 where the location of one of the DCs in the economically optimum solution is given in a location that can be considered as to having difficult access which is the Alps near the frontier between Switzerland and Italy.

It is also shown how a customer with an extreme high demand can greatly affect the location of the centre of gravity. This is clear in database 1 where London is a location with an extremely high demand. The location of the centre of gravity of the cluster that includes London tended to be located in the vicinities of this city. But an even clearer case is identifiable in database 3. There is a clear dominant node in Watford Herts, UK. This node that accounts for nearly one-fifth of the total demand has a notable impact on the location of the centre of gravity of the cluster where it is located. When analysing this impact it was found that the location of the centre of gravity of the cluster changed from the UK when the node was considered, to north-east France when the assumption of this node not existing was taking into consideration. This dramatic deviation of location can make the distribution network become highly inefficient if the company were to lose that one customer.

All of these issues show that the COG model can have serious flaws as a tool to solving the FLP. Its simplicity brings many weaknesses. Nevertheless, it can be used as a guidance tool when analysing the information. The results of the COG model can clearly be improved by using the various clustering algorithms, however, the algorithm that gave an overall best result is the FCM algorithm. This contradicts the conclusions of Esnaf and Küçükdeniz (2009) who conclude that the GK model gives a better solution. The solution given by this research does not however deny that the GK model is adequate since in some cases it did give an optimum solution, specifically in database 3. The GG model is a model that cannot be recommended given the results of the research since at no point did it give an optimum solution.

This research has helped to identify that by dividing the clients into geographical clusters one of the fundamental weaknesses that the COG model has, can be overcome. Additionally, the study suggests that by combining the COG model with the FCM clustering algorithm the overall results are better than the other two clustering algorithms that were analysed.

6. CONCLUSIONS AND CRITIQUE

In concluding this thesis, the objectives of this research were:

- Identify the main characteristics that encompass the Facility Location Problem within Logistics and Supply Chain Management.

This has been done following a literature review on the FLP and how it impacts on the supply chain. The findings are that the FLP is a key part of the designing phase of the supply chain and it has a great impact on the overall performance of the future distribution network. The FLP impacts on decisions that are usually considered for long term and require high investments such as the opening or closing of a distribution centre or the reassignment of clients to existing DCs. The FLP was identified to be categorized depending on the characteristics of the problem and the variables that should be taken into account. However, due to the complexity of the problems, it is currently unclear on how to have models that can help solve multi-variable FLP.

- Define the Centre of Gravity technique to solve the Facility Location Problem and identify its strengths and weaknesses.

This has been done by an exhaustive literature review coming from seminal works such as Weber (1909) and Cooper (1961) to more recent works that utilize the COG as a hybrid method with other modelling techniques such as clustering techniques like that presented by Esnaf and Küçükdeniz (2009). The COG is identified as a linear equation defined in equation 1.

The main strength that was found is the fact that the COG is a simple model that can help identify possible locations where to place the DCs in order to minimize the weighted distance to the customers. However, some weaknesses were also found that put to doubt the applicability of the COG as a standalone decision tool. The fact that the COG can fall anywhere in the plane irrespective of geographical limitations is an example of such weaknesses.

- Apply different Clustering Algorithms to client databases and analyse the solutions given in terms of improvement to distribution costs.

After going through some bibliography and identifying the characteristics of clustering algorithms, how they work and the different classifications, three different algorithms were chosen for applying onto the databases. Additionally, in order to review the impact that the hybrid model has on databases of different sizes, three databases of varying sizes were chosen.

From the research done on clustering algorithms, it was identified that one of the main inputs for the algorithms is the number of clusters into which the dataset is to be divided. Therefore, in order to quantify the impact that this decision has on the final result, each dataset was analysed using each of the three algorithms with different number of clusters from 1 cluster up to 8 clusters.

Each of the results of the algorithm was then passed through a model developed by the author in order to quantify the distribution costs that each scenario has. The costs were divided into Truncking (fixed and variable), local delivery (fixed and variable), depot (fixed and variable)and holding costs.

The overall research question was “Can the performance of the COG model for the solution of the FLP be improved by the application of different clustering algorithms? In answer to this it was shown that the results given by the COG model can improve if a clustering algorithm is used in order to divide the clients into groups, of course this bringing the necessity to open one DC for each cluster. From the research it was shown that the COG-FCM clustering algorithm hybrid would bring the overall best result.

The research shows that the COG model can be used to solve the FLP given a certain framework and conditions. However, it would not be recommended to use this method as a stand-alone decision making tool.

As an outcome of this research, some topics for further research have been identified in order to develop the clustering-COG hybrid model proposed by the author. The research that has been done lacks of the possibility to analyse the true impact that the distance between nodes has in terms of the location of the centre of gravity. This is an identified weakness of the COG model that should be analysed. Another topic for further research is a more in-depth analysis of the impact of high volume nodes. Although this issue was touched in the present research, further sensitivity analysis can help to quantify the impact that losing a key client could bring to a distribution network that was planned using the COG model.

The data that was used as input for the model was taken from the PPP study case. This dataset was not changed throughout the research. A possible sensitivity analysis of how the variation of certain factors such as the fixed warehouse costs, product value or safety stock policies can impact on the decision of opening more than one DC.

BIBLIOGRAPHY

Alp, O., Erkut, E. and Drezner, Z. (2003), "An Efficient Genetic Algorithm for the p-Median Problem", *Annals of Operations Research,* vol. 122, no. 1/4, p. 21.

Arostegui, M. A.,Jr, Kadipasaoglu, S. N. and Khumawala, B. M. (2006), "An empirical comparison of Tabu Search, Simulated Annealing, and Genetic Algorithms for facilities location problems", *International Journal of Production Economics,* vol. 103, no. 2, p. 742.

Ballou, R. H. (1973), "Potential Error in the Center of Gravity Approach to Facility Location", *Transportation Journal,* vol. 13, no. 2, pp. 44-50.

Ballou, R.H. (1999), “Business Logistics Management: Planning, Organization & Controlling the Supply Chain”, *Modern Purchasing,* vol. 42, no. 11, p. 14, Toronto

Bezdek, J.C., (1981), “Pattern Recognition with Fuzzy Objective Function Algorithms”, New York: Plenum Press.

Bose, I. and Chen, X. (2009), “A Method for Extension of Generative Topographic Mapping for Fuzzy Clustering”, *Journal of the American Society for Information Science and Technology,* vol. 60, no. 2, pp. 363-371.

Chen, C. and Ting, C. (2008), "Combining Lagrangian heuristic and Ant Colony System to solve the Single Source Capacitated Facility Location Problem", *Transportation Research.Part E, Logistics & Transportation Review,* vol. 44, no. 6, p. 1099.

Chistopher, M. (1992), “Logistics and Supply Chain Management”, Pitmans, London, UK.

Cooper, L. (1961), “Location-Allocation Problems”, *Operations Research,* vol. 11, 331-343.

Drezner, Z. (1995), “Facility Location: A Survey of Applications and Methods”, Springer Verlag.

Drezner, T. and Drezner, Z. (2007), "The gravity p-median model", *European Journal of Operational Research,* vol. 179, no. 3, p. 1239.

Eiselt, H.A., and Sandblom, C.-L. (2004), “Decision Analysis, Location Models, and Scheduling Problems”, Springer-Verlag.

Erlenkotter, D. (1978), "A Dual-Based Procedure for Uncapacitated Facility Location", *Operations Research,* vol. 26, no. 6, p. 992.

Esnaf, S. and Küçükdeniz, T. (2009), "A fuzzy clustering-based hybrid method for a multi-facility location problem", *Journal of Intelligent Manufacturing,* vol. 20, no. 2, p. 259.

Gath, I., and Geva, B. (1989) “Unsupervised Optimal Fuzzy Clustering”, *IEEE Transactions on Patterns Analysis and Machine Intelligence*, Vol. 11, no.7, pp 773-781.

Ghosh, D. (2003), "Neighborhood search heuristics for the uncapacitated facility location problem", *European Journal of Operational Research,* vol. 150, no. 1, p. 150.

Jain, A.K. and Dubes R.C. (1988), “Algorithms for Clustering data”, Prentice-Hall advanced reference series, Prentice-Hall, inc.

Jain, A.K., Murty, M.N. and Flynn, P.J., (1999), “Data Clustering: A Review”, *ACM Computing Surveys*, vol. 31, no. 3, pp. 264-323.

Jaramillo, J. H., Bhadury, J. and Batta, R. (2002), "On the use of genetic algorithms to solve location problems", *Computers & Operations Research,* vol. 29, no. 6, pp. 761-779.

Lashine, S. H., Fattouh, M. and Issa, A. (2006), "Location/allocation and routing decisions in supply chain network design", *Journal of Modeling in Management,* vol. 1, no. 2, p. 173.

Love, R. F. and Morris, J.G., (1972), “Modeling Inter-City Road Distances by Mathematical Functions”, *Operational Res. Quart.,* vol. 23, no. 1, p. 61.

Meeyai, S., (2009), “A Hybrid Approach for the Design of Facility Location and Supply Chain Network Under Supply and Demand Uncertainty: A Systematic Review”, Masters in Research Dissertation, Cranfield University.

Mladenovic, N., (2003), “Solving the p-Center Problem with Tabu Search and Variable Neighborhood Search”, *Networks,* vol. 42(1), pp. 48-64.

ReVelle, C. S. and Eiselt, H. A. (2005), "Location analysis: A synthesis and survey", *European Journal of Operational Research,* vol. 165, no. 1, p. 1.

ReVelle, C.S., Eiselt, H.A., and Daskin, M.S. (2008), “A Bibliography for Some FundamentalProblem Categories in Discrete Location Science”, *European Journal of Operational Research,* vol. 184, pp. 817-848.

Rushton, A., Croucher, P. and Baker, P. (2006), “The Handbook of Logistics and Distribution Management”, Kogan Page.

Simchi-Levi, D., Kaminsky, P. and Simchi-Levi, E. (2008). “Designing and Managing the Supply Chain: Concepts, Strategies, and Case Studies”, McGraw Hill Higher Education.

Sridharan, R. (1991), "A Lagrangian Heuristic for the Capacitated Plant Location Problem with Side Constraints", *The Journal of the Operational Research Society,* vol. 42, no. 7, p. 579.

Sridharan, R. (1993), "A Lagrangian heuristic for the capacitated plant location problem with single source constraints", *European Journal of Operational Research,* vol. 66, no. 3, p. 305.

Waters, D. (2006), “Inventory Control and Management”, Wiley.

Weber, A., (1909), “Uber den Standort der Industrien*”,* Tubingen, (translated by Friedrich, C.J., “Theory of the Location of Industries*”*, *University of Chicago Press,* Chicago, IL, USA.)

Wen, M. and Iwamura, K. (2008), "Fuzzy facility location-allocation problem under the Hurwicz criterion", *European Journal of Operational Research,* vol. 184, no. 2, p. 627.

Zhou, J. and Liu, B. (2003), "New stochastic models for capacitated location-allocation problem", *Computers & Industrial Engineering,* vol. 45, no. 1, p. 111.

APPENICES

APPENDIX A: CUSTOMER DATABASES

APPENDIX A-1: DATABASE 1

Location	Demand/Weight	Latitude	Longitude
Cadiz	29	36	(6)
Malaga	10	37	(4)
Cordoba	4	38	(4)
Alicante	9	38	(0)
Valencia	23	40	(1)
Barcelona	42	41	2
Madrid	55	40	(4)
Valladolid	20	41	(4)
Bilbao	11	43	(3)
San Sebastian	6	43	(2)
Coimbra	11	40	(8)
Porto	20	41	(8)
Lisbon	60	39	(9)
Toulouse	7	44	1
Bordeaux	19	45	(1)
Limoges	35	45	1
Montpellier	21	43	4
Lyon	57	46	5
Dijon	54	47	5
Tours	19	47	1
Nantes	33	47	(2)
Rennes	47	49	(1)
Lille	21	50	3
Paris	88	48	2
Brussels	160	51	4
Rotterdam	54	52	4
Antwerp	35	51	4
Geneva	39	46	6
Zurich	110	47	9

Location	Demand/Weight	Latitude	Longitude
Innsbruck	32	47	11
Vienna	42	48	16
Frankfurt	88	51	8
Stuttgart	108	49	9
Munich	81	48	12
Aarhus	41	56	10
Odense	20	55	10
Dresden	41	51	14
Essen	27	51	7
Dusseldorf	110	51	6
Leipzig	59	51	12
Hannover	40	52	10
Berlin	161	53	13
Palermo	15	38	13
Genoa	26	44	9
Verona	7	45	11
Florence	27	46	11
Rome	110	42	13
Taranto	29	40	17
Milan	136	46	9
Dublin	82	53	(6)
Portsmouth	57	50	(1)
Birmingham	126	52	(2)
Edinburgh	91	56	(3)
London	344	51	(0)

APPENDIX A-2: DATABASE 2

Location	Demand/Weight	Latitude	Longitude
Achavanich	2	58	(3)
Clara	8	53	(8)
Airport Southampton	4	51	(1)
Aldea Del Rey	56	39	(4)
Almolda (La)	46	42	(0)
Costa De Caparica	97	39	(9)
Arenas (Las)	8	43	(3)
Altinova (Izmit)	12	41	30
Calcinelli Di Saltarati	145	44	13
Anjan	57	64	13
Canitello	72	38	16
Barraca De Aguas Vivas	163	39	(0)
Helvoirt	1,037	51	5
Asele	90	64	17
Ceanannus Mor (Kells)	53	54	(7)
Auslikon	40	47	9
Phasli	156	35	32
Dahl Nw (Hagen)	17	51	8
Baggetorp	88	59	16
Balderton	57	53	(1)
Corbally	265	54	(9)
Borriol	20	40	(0)
Cardedu	306	40	10
Brie	27	50	3
Bromont Lamothe	420	46	3
Epfendorf	899	48	9
Gladbeck	76	52	7
Goppingen	31	49	10
Bicker	12	53	(0)
Tranby	8	60	10
Blackwater Hants	38	51	(1)
Negreni	4	47	23
Drage	108	44	16
Castrillo De La Vega	4	42	(4)
Chablis	174	48	4
Lovrec (Hrvatska)	56	43	17
Chaumont Sur Aire	7	49	5
Craon	177	48	(1)
Daumazan-Sur-Arize	294	43	1

Location	Demand/Weight	Latitude	Longitude
Staporkow	89	51	21
Mikleus	10	46	18
Burton Joyce	838	53	(1)
Esporoes	89	42	(8)
Castelfranco Emilia	599	45	11
Kato Akourdhalia	131	35	32
Chioggia	268	45	12
Cropani Marina	15	39	17
Pisoes (Vila Real)	265	42	(8)
Dubbione	96	45	7
Colorado (El)	80	36	(6)
Dunkerrin	53	53	(8)
Bourg St Leonard (Le)	537	49	0
Eaubonne Paris	268	49	2
Breidenbach	30	49	7
Grenoble-Meylan	68	45.21	5.78
Falerna Marina	368	39.00	16.12
Goce Delcev	30	41.58	23.75
Grange (Sligo)	219	54.39	-8.53
Coatham Mundeville	70	54.57	-1.55
Font De La Figuera	297	38.80	-0.86
Julianstown	23	53.67	-6.29
Portagem	15	39.38	-7.39
Jarnac	39	45.68	-0.17
Ganaceto	168	44.71	10.90
Cullaville	59	54.06	-6.65
Grobnobach	193	48.35	11.58
Lacaune	394	43.65	2.76
Potamitissa	59	34.92	32.99
Sadrazamkoy	119	35.39	32.96
Modran	163	44.95	17.96
Camlibel	1,055	35.32	33.07
Dore	191	53.31	-1.53
Pleternica	809	45.29	17.80
Gattorna	204	44.43	9.19
Tobercurry	111	54.05	-8.74
Lescure	1,248	43.00	1.23
Vaksdal	131	60.48	5.75
Pomarkko	97	61.70	22.00
Emirdag	31	39.02	31.16
Harb	88	50.43	8.99

Location	Demand/Weight	Latitude	Longitude
Ersnas	601	65.53	21.79
Sao Miguel De Machede	19	38.64	-7.73
Houten	39	52.04	5.17
Lamezia	584	38.91	16.28
Felbridge	50	51.14	-0.05
Ffostrasol	38	52.10	-4.38
Peilstein	8	48.62	13.90
Forzanes	160	42.37	-8.45
Guardia (La)	212	39.80	-3.48
Valjok	396	69.75	25.94
Ustek	131	50.58	14.34
Ayios Konstantinos	27	34.87	33.07
Latisana	40	45.78	12.99
Lodi	4	45.31	9.50
Gempenach	51	46.94	7.20
Halsteren	280	51.53	4.27
Hedemunden	461	51.39	9.76
Demirhan	8	35.23	33.48
Herrenberg	40	48.61	8.90
Gotlunda	336	59.35	15.67
Ampezzo	134	46.42	12.79
Marseille En Beauvaisis	79	49.57	1.95
Jeserig Bb (Belzig)	8	52.09	12.45
Higuera De La Serena	52	38.65	-5.74
Perchau	70	47.11	14.46
Viborg	1,037	56.45	9.39
Purmerend	254	52.51	4.95
Kindelbruck	249	51.26	11.08
Hatfield Woodhouse	89	53.57	-0.98
Klietz	33	52.67	12.06
Veenwouden	89	53.24	5.99
Konigsfeld Sn	85	51.06	12.76
Leitariegos	35	42.99	-6.41
Ekenas/Tammisaari	54	59.98	23.43
Hollingworth	26	53.46	-1.99
Haderup	89	56.39	8.99
Wageningen	273	51.97	5.67
Podoleni	133	46.79	26.61
Librazhd	41	41.18	20.33
Incirliova	269	37.85	27.73
Pucioasa	95	45.07	25.44

Location	Demand/Weight	Latitude	Longitude
Hobro	177	56.64	9.80
Meze	57	43.43	3.60
Kreuztal	6	50.96	7.99
Aprica	164	46.16	10.16
Rantakyla	205	61.68	27.22
Lillerod	593	55.87	12.34
Nagyatad	38	46.22	17.37
Yayla	235	35.23	32.94
Kelso	131	55.60	-2.43
Venaja	461	60.96	23.31
Loose Sh	75	54.52	9.89
Kirkwall	42	58.98	-2.96
Lugau Sn	17	50.74	12.74
Lutzen	46	51.26	12.14
Cerilly (Chatillon S S)	59	47.87	4.50
Mechterstadt	291	50.95	10.53
Flattach	441	46.93	13.13
Mirabel Et Blacons	169	44.71	5.09
Marina Di Montemarciano	8	43.65	13.34
Langley Mill	156	53.02	-1.33
Nola	45	40.93	14.49
Zornica	131	42.39	26.93
Llagostera (Gerona)	2	41.83	2.89
Morigny	83	48.45	2.17
Neutal	20	47.55	16.45
Kajaani	235	64.23	27.73
Swidnik	391	51.19	22.66
Maqueda	164	40.04	-4.22
Palleronte	7	44.21	10.01
Mellendorf	40	52.55	10.72
Sip	336	44.68	22.51
Merkstein	88	50.89	6.12
Muhlhausen Im Tale	119	48.57	9.66
Isaccea	111	45.27	28.46
Malpartida De La Serena	57	38.68	-5.64
Mataro	90	41.55	2.43
Pontedazzo	177	43.49	12.62
Mouchan	12	43.91	0.30
Montcada	17	41.48	2.19
Obervieland	277	53.04	8.82
Ottersberg	59	53.11	9.15
Raesfeld	8	51.77	6.84

Location	Demand/Weight	Latitude	Longitude
Prat Bonrepaux	17	43.03	1.02
Smederevo	22	44.66	20.93
Quimper	294	47.99	-4.10
Stremski Karlovci	1,037	45.20	19.94
Montoro	42	38.02	-4.38
Murguia	19	42.93	-2.79
Richelieu	108	47.01	0.20
Rosans	1	44.39	5.47
Reichenbach Sn (Gorlitz)	6	51.14	14.80
Olias Del Rey	1,055	39.95	-3.99
Athlone	39	53.42	-7.96
Nasviken (Stromsund)	79	63.85	15.52
Razvad	1,479	44.93	25.54
Sankt Polten	28	48.20	15.63
Tiffen	269	46.71	14.07
Newtownbutler	294	54.18	-7.36
Primolano	105	45.96	11.71
Nottingham	30	52.95	-1.15
Aghamore (Longford)	277	53.90	-7.96
Rethem	191	52.78	9.38
Of	16	40.93	40.26
Ossa De Montiel	7	38.97	-2.74
Alhos Vedros	17	38.65	-9.03
S Heer Arendskerke	4	51.49	3.82
Paravadella	268	43.07	-7.15
Rinteln	96	52.19	9.09
Tapioszele	26	47.33	19.89
Rocca S Casciano	53	44.06	11.84
Paterna De Rivera	102	36.52	-5.87
Pilar Del Prado	80	36.72	-4.53
Boraja	4	43.62	16.08
Brajkovici	45	45.11	13.77
Frydlant N Ostravici	215	49.60	18.35
Rojales	268	38.09	-0.73
Vila Nova Da Barquinha	40	39.46	-8.43
Zabari	750	44.35	21.22
Tusnad	235	46.21	25.92
Chapelle Vendomoise (La)	132	47.67	1.24
Taglio Corelli	177	44.54	12.01
Madaras (Satu Mare)	12	47.69	22.85
Horazdovice	571	49.34	13.69
Rouen-Deville Les Rouen	54	49.46	1.05

Location	Demand/Weight	Latitude	Longitude
Taverna Fr	39	41.45	13.92
Vermes	537	45.53	21.66
Zevenbergen	131	51.65	4.61
Saint Claude	89	46.39	5.87
Rudesheim	31	50.02	7.95
Bebing	17	48.71	7.00
Voislova	39	45.52	22.47
Schwarza Th (Rudolstadt)	92	50.70	11.33
Steinheim	461	51.87	9.10
Saint Maurice De Beynost	368	45.83	4.98
Untergrombach	70	49.09	8.56
Torriglia	39	44.52	9.16
Sanchonuno	297	41.33	-4.30
Saint Sernin Sur Rance	50	43.89	2.60
Sartilly	271	48.76	-1.45
Warin	269	53.80	11.69
Neudorf Ob Wildon	45	46.90	15.49
Iza	51	47.76	18.23
Sillans	31	45.34	5.39
Saint Helen Auckland	27	54.63	-1.71
Thevet St Julien	273	46.64	2.07
Treveray	56	48.61	5.40
Salen	539	61.16	13.27
Nyirpazony	51	48.02	21.82
San Lorenzo Di Sebato	441	46.79	11.89
Susana	52	42.81	-8.49
Bukovac (Crna Gora)	51	43.05	18.87
Santa Marina Salina	215	38.56	14.87
Boldu	73	45.32	27.24
Varangeville	249	48.64	6.32
Scarisbrick	168	53.61	-2.93
Schmerikon	204	47.23	8.94
Weibenburg By	111	49.02	10.98
Guntesdorf	22	48.65	16.05
Villevallier	177	48.02	3.32
Ciergnon	102	50.13	5.02
Kihlanki	396	67.58	23.55
Lebbeke	475	51.00	4.13
Solhan	89	38.96	41.05
South Petherton	1,209	50.94	-2.80
Cabeco	277	40.45	-8.72
Wernberg-Koblitz	51	49.56	12.14

Location	Demand/Weight	Latitude	Longitude
Stone Glos	537	51.66	-2.46
Dolna Mitropolija	6	43.47	24.53
Sudbury G Lon	40	51.56	-0.33
Toreno	51	42.69	-6.52
Carvalhal (Castro Daire)	338	40.85	-7.93
Gharghur	733	35.92	14.46
Ballyshannon	10	54.50	-8.19
Arnhem	79	51.99	5.91
Tercan	102	39.78	40.39
Voutenay Sur Cure	102	47.56	3.78
Divci	1,479	44.30	20.03
Battaglia Terme	132	45.29	11.79
Villafranca De Los Caballeros	52	39.42	-3.35
Europaplein	27	50.85	5.75
Brzeg	234	50.87	17.48
Bad Berka	1	50.90	11.28
Tryde	23	55.57	13.93
Zauan	70	47.23	22.66
Ugrak	750	40.31	40.11
Wittmar	56	52.13	10.64
Skinnerup	212	56.98	8.67
Lipsko	16	51.16	21.67
Opatow (Otrowiec Swiet)	338	50.81	21.43
Bendorf	183	50.42	7.60
Biri	90	60.95	10.61
Belgodere	142	42.58	9.02
Korpilahti	145	62.02	25.57
Hurezani	102	44.81	23.65
Villasandino	391	42.37	-4.11
Rydal	601	57.56	12.70
Bergen Ni (Celle)	42	52.81	9.96
Albaida	75	38.83	-0.52
Besenfeld	83	48.60	8.43
Egersund	1	58.47	6.02
Igoumenitsa	273	39.49	20.27
Kardamila	70	38.51	26.09
Samothraki	38	40.48	25.53
Whitchurch Shrops	39	52.97	-2.68
Wilkieston	8	55.90	-3.41
Ozd	174	48.25	20.29
Wotton Bridge	23	50.73	-1.23
Kamenne Zehrovice	750	50.13	14.02

Location	Demand/Weight	Latitude	Longitude
Gabrovo	271	42.87	25.33
Fredrikstad	28	59.21	10.97
Dhierona	205	34.83	33.11

APPENDIX A-3: DATABASE 3

Location	Demand/Weight	Latitude	Longitude
Modica	21.555	36.84	14.76
Ragusa	190168.318	36.92	14.72
Comiso	4191.303	36.95	14.60
Gela	1788.736	37.07	14.24
San Benedetto	79873.333	37.37	13.64
Catania	101907.141	37.51	15.10
Reggio di Calabria	11442.834	38.10	15.64
Palermo	202860.096	38.12	13.36
Capo d'Orlando	490.79	38.16	14.75
Tagliatore	3701.986	38.19	15.30
Roccella Ionica	40902.11	38.32	16.41
Santa Eufemia Lamezia	257.515	38.91	16.26
Catanzaro	181304.502	38.91	16.59
Vila Franca De Xira	6.635	38.95	-9.00
Cagliari	7.825	39.23	9.10
Cagliari	220599.795	39.23	9.10
Amendola	106838.586	39.34	16.12
Santa Giusta	4526.465	39.88	8.61
Birori	53258.629	40.27	8.81
Nuoro	31065.2	40.32	9.33
Lecce	335.048	40.35	18.16
Lecce	46999.629	40.35	18.16
San Pancrazio Salentino	11832.323	40.42	17.84
Taranto	750.803	40.48	17.23
Polla	21246.072	40.53	15.50
Potenza	4403.769	40.64	15.80
Pontecagnano Faiano	80090.973	40.65	14.85
San Vito dei Normanni	591.021	40.66	17.71
Salerno	64147.46	40.67	14.78
Matera	3.275	40.68	16.59
Castellamare di Stabia	9.036	40.70	14.49
Pagani	26001.408	40.75	14.61
Pozzuoli	21199.951	40.83	14.12
Altamura	9211.661	40.83	16.55
Napoli	186236.141	40.85	14.25
Napoli	1321.384	40.86	14.27
Casalnuovo di Napoli	243514.579	40.91	14.32
Avellino	166549.444	40.91	14.79
Melito Di Napoli	279646.565	40.92	14.23
Melito Di Napoli	10.1	40.92	14.23

Location	Demand/Weight	Latitude	Longitude
Cozze	116567.925	41.03	17.15
Casagiove	152350.493	41.07	14.30
Modugno	62558.066	41.09	16.77
Bari	308680.244	41.12	16.85
Barletta	54154.332	41.32	16.28
Bocca di Fiume	5154.065	41.45	13.02
Foggia	50250.888	41.46	15.53
Foggia	13522.672	41.47	15.55
Campobasso	39240.328	41.56	14.66
Campobasso	1681.822	41.56	14.66
Cisterna di Latina	9025.385	41.60	12.83
Rome	135340.001	41.88	12.38
Roma	77.083	41.89	12.47
Roma	654628.157	41.89	12.47
Ajaccio	24522.542	41.92	8.74
Avezzano	12939.03	42.03	13.42
Vasto Ch	19863.208	42.12	14.71
Borgata Marina	48398.045	42.23	14.55
Caldare	97423.769	42.45	12.40
Pescara	241144.816	42.46	14.22
Perpignan	39456.559	42.70	2.89
Grosseto	55105.092	42.76	11.11
Grosseto	9551.116	42.76	11.12
Marconi	52217.136	42.88	13.90
Le Conie	3.275	42.91	11.72
Querce al Pino	10.1	43.02	11.91
Il Sardo	59131.972	43.08	12.43
Six Fours Les Plages	15255.2	43.10	5.83
Perugia	201177.904	43.11	12.37
La Velette-du-Var	39464.664	43.13	5.98
Narbonne	20359.135	43.19	3.00
Carcassonne	12337.491	43.22	2.36
Sinalunga	47663.666	43.23	11.74
Semeac	16481.265	43.23	0.10
Montecorsaro	86174.336	43.28	13.65
Aubagne	17925.27	43.29	5.57
Marcerata	44791.659	43.30	13.44
Billere	54385.275	43.30	-0.40
Béziers	28520.185	43.35	3.22
Marseille	94018.791	43.36	5.36
Flassans-sur-Issole	157.568	43.37	6.20

Location	Demand/Weight	Latitude	Longitude
Vitrolles	44539.513	43.45	5.25
Puget Sur Argens	27441.951	43.45	6.69
Poggibonsi	4491.2	43.47	11.15
Arezzo	4611.91	43.47	11.87
Arezzo	33554.974	43.47	11.87
Bayonne	39840.8	43.49	-1.47
Jesi	10.789	43.52	13.24
Portet Sur Garonne	114223.335	43.53	1.40
Livorno	120999.216	43.54	10.32
Livorno	5.55	43.54	10.32
Sansepolcro	26.025	43.57	12.14
Saint-Jean-de-Vedas	53632.17	43.58	3.83
Venelles	22292.801	43.59	5.48
Lavaiano	41508.943	43.64	10.57
Rocca Priora	115414.592	43.64	13.36
Villeneuve-Loubet	79816.872	43.66	7.12
Saint-Jean	136108.514	43.67	1.51
Empoli	146841.475	43.72	10.95
Firenze	161291.328	43.78	11.24
Baccane	95201.329	43.81	10.85
Saint-Cesaire	57594.851	43.81	4.33
Osmannoro	102299.538	43.81	11.18
Zone	74090.723	43.86	10.60
Prato Fi	47280.451	43.87	11.09
Les Vigneres	42911.495	43.88	5.04
Pesaro	140443.974	43.91	12.90
Le Pontet	78659.336	43.96	4.86
Bressols	1199.33	43.97	1.34
Massa	39649.823	44.04	10.14
Ramini	80448.768	44.06	12.56
Sarrians	2.592	44.08	4.98
Sarzana	91237.126	44.11	9.96
Larnac	22005.522	44.11	4.11
Layrac	60820.986	44.13	0.66
Cesena	121734.708	44.14	12.24
Faenza	300164.363	44.29	11.89
Savona	83787.717	44.30	8.46
Villetta	45524.409	44.34	7.08
Rodez	26776.765	44.35	2.56
Villefranche De Rouergue	39260.27	44.35	2.03
Imola	81688.85	44.36	11.71
Mondovi	92174.39	44.40	7.82

Location	Demand/Weight	Latitude	Longitude
Genova	278.875	44.41	8.95
Sampierdarena	142352.157	44.42	8.89
San Lazzaro	161910.409	44.47	11.41
Bologna	737.373	44.50	11.33
Bologna	16250.474	44.50	11.33
Mende	30798.121	44.52	3.50
Gap	20479.771	44.56	6.08
Modena	297896.331	44.64	10.92
Modena	223767.809	44.64	10.92
Savigliano	168.517	44.65	7.65
Reggio Emilia	191557.498	44.70	10.62
Canove	3.275	44.78	8.09
Parma	700.104	44.80	10.32
San Pancrazio	41284.967	44.81	10.27
San Pancrazio Pr	99243.569	44.81	10.27
Borgo	24208.66	44.82	10.89
Ferrara	376079.609	44.83	11.62
Bordeaux	117774.026	44.84	-0.57
Borgo Tanaro	96314.372	44.89	8.21
Tortona	5954.545	44.90	8.86
Valence	93451.914	44.90	4.88
Alessandria	88746.214	44.91	8.60
Aurillac	18338.121	44.93	2.45
Tetti Nina	24.446	45.00	7.53
Moncalieri	164792.475	45.00	7.68
Casteggio	88277.007	45.01	9.13
La Verza	377138.923	45.03	9.68
Romans Sur Isère	11183.579	45.04	5.05
Piacenza	114136.994	45.05	9.68
Les Sallins	84680.124	45.05	3.96
Mulino	245228.257	45.06	7.58
Torino	7117252.228	45.07	7.68
Rovigo	118354.946	45.07	11.79
Governolo	3383.853	45.09	10.96
Le Ridelet	40259.147	45.12	5.69
Larche	38427.627	45.12	1.41
Cremona	148141.428	45.13	10.01
Mantova	242202.823	45.16	10.80
Périgueux	42023.68	45.18	0.72
Spino d'Adda	701.344	45.40	9.49
Palazzina	3339.896	45.40	11.01
Padova	238228.113	45.42	11.87

Location	Demand/Weight	Latitude	Longitude
Verona	534674.171	45.44	10.99
Cesano Boscone	497750.084	45.44	9.10
Novara	1.121	45.45	8.61
Banchette	60770.779	45.45	7.85
Novara	25870.245	45.45	8.62
La Terrasse	67747.976	45.47	4.37
Rivolta d'Adda	25531.713	45.48	9.51
Rivolta d'Adda	108260.316	45.48	9.51
Chirignago	128811.613	45.49	12.22
Cornaredo	408654.907	45.50	9.03
Saint-Romain-en-Gal	36941.246	45.53	4.86
Vienne	64.818	45.53	4.87
Brescia	106709.711	45.54	10.21
Brescia	186557.653	45.54	10.21
Vicenza	83158.768	45.55	11.53
Nerviano	81022.515	45.56	8.98
Chambéry	77250.657	45.57	5.92
Monza	65447.18	45.58	9.27
Castellanza	23863.815	45.61	8.89
Royan	1873.719	45.63	-1.02
Sabbio	98815.564	45.64	9.61
La Jard	7080.977	45.65	-0.59
Vedelago	200378.315	45.69	12.01
Le Maine	66073.551	45.69	0.18
Cognac	4130.694	45.70	-0.33
Saint Priest	135901.844	45.70	4.94
Thiene	278497.983	45.71	11.48
Sesto Calende	76774.707	45.73	8.63
Chassieu	38903.041	45.74	4.97
Perego	93342.468	45.74	9.36
Saintes	10714.59	45.75	-0.64
Barzago	247865.418	45.76	9.31
Clermont Ferrand	84952.566	45.78	3.08
Limoges	39392.521	45.83	1.26
Susegana	72881.612	45.85	12.25
Lecco	194511.757	45.85	9.39
Seynod	84573.494	45.88	6.10
Ornavasso	183114.967	45.96	8.43
Ponte Tagliamento	120187.353	45.97	12.93
Roanne	5881.137	46.05	4.07
Trento	191314.861	46.08	11.12
Cusset	42746.584	46.13	3.47

Location	Demand/Weight	Latitude	Longitude
Aytré	6430.742	46.13	-1.12
Belluno	2469.091	46.14	12.21
Sion	8338.517	46.23	7.36
Viriat	90222.431	46.23	5.19
Niort	19843.142	46.32	-0.47
Raccolana	97015.123	46.41	13.32
Lucon	760.966	46.45	-1.17
Fontenay Le Comte	1796.036	46.46	-0.81
Bolzano	40188.226	46.50	11.34
Morges	45414.692	46.51	6.50
Yzeure	28028.194	46.57	3.35
Poitiers	54515.08	46.59	0.35
Villach	24674.545	46.61	13.84
Bourbon Lancy	12411.824	46.62	3.78
Parthenay	8994.216	46.65	-0.24
Saint Marcel	41763.614	46.78	4.89
Yverdon Les Bains	15751.159	46.79	6.65
Luns	113552.064	46.79	11.98
Heimberg	16676.773	46.79	7.61
Villedieu-sur-Indre	10386.196	46.85	1.54
Coulanges-les-Nevers	19101.926	47.01	3.19
Kriens	24170.165	47.03	8.28
Graz	37730.285	47.06	15.42
Pougues-les-Eaux	2.958	47.07	3.11
Bourges	45482.135	47.09	2.41
Biel/bienne	1325.798	47.14	7.24
Vierzon	7195.228	47.22	2.07
Besançon	58498.172	47.25	6.02
Orvault	162858.897	47.26	-1.62
Innsbrück	29344.665	47.27	11.40
Steinbichl (kitzbühel)	3529.492	47.28	12.65
Chenôve	66540.176	47.30	5.02
Chambray Lès Tours	74264.913	47.34	0.72
Hendschiken	63774.591	47.39	8.21
Zürich	3959.25	47.39	8.54
Lustenau	16301.288	47.44	9.66
Kloten	215673.815	47.45	8.58
Bel-Air	28486.608	47.47	-0.68
Muttenz	30961.767	47.53	7.64
Staad	61511.015	47.58	9.37
Blois	37849.136	47.59	1.33
Lauchringen	8004.923	47.63	8.32

Location	Demand/Weight	Latitude	Longitude
Friedrichshafen	1515.013	47.65	9.48
Vannes	11627.663	47.66	-2.75
Weindorf	16262.338	47.69	11.20
Steaufzgen	8713.93	47.72	10.29
Singen	8385.648	47.76	8.82
Lanester	19070.517	47.77	-3.35
Illzach	41059.142	47.77	7.35
Ravensburg	30672.832	47.79	9.61
Louailles	1765.284	47.79	-0.25
Auxerre	21571.446	47.80	3.58
Winkl bei Grabenstatt	17868.317	47.83	12.52
Seethal	31241	47.84	12.49
Kaufbeuren	8952.553	47.88	10.62
Fleury-Les-Aubrais	61497.494	47.94	1.92
Montargis	50835.526	47.99	2.73
Freiburg	97425.667	47.99	7.84
Saint Florentin	15028.865	48.00	3.72
Le Mans	485.568	48.01	0.20
Mans (le)	77230.899	48.01	0.20
Soulge-sur-Ouette	23806.375	48.06	-0.57
Colmar	2323.73	48.08	7.35
La Hallerais	84407.664	48.08	-1.64
Loudéac	1.9	48.18	-2.75
Plomodiern	45713.024	48.18	-4.23
Milbertshofen	102029.541	48.18	11.59
Furstenfeldbruck	16979.981	48.19	11.25
Marchtrenk	32302.864	48.19	14.11
Wien	274566.639	48.20	16.37
Golbey	29361.23	48.20	6.44
Herbolzheim	7793.499	48.23	7.75
Teising	40040.595	48.23	12.61
Thal	80746.43	48.28	10.11
Lohhof	87520.656	48.30	11.58
Creney-pres-Troyes	22130.393	48.33	4.13
Illerbrucke	166183.994	48.37	9.99
Neu-ulm	389.828	48.39	9.99
Fontainebleau	218.371	48.40	2.70
Gersthofen	32587.823	48.42	10.87
Alençon	7497.086	48.43	0.09
Guipavas	37406.943	48.43	-4.40
Chartres	45919.332	48.44	1.49
Oberelchingen	1043024.979	48.46	10.07

Location	Demand/Weight	Latitude	Longitude
Offenburg	33404.526	48.47	7.95
Schirmeck	31.523	48.48	7.22
Melun	37693.017	48.54	2.66
Moniberg	12280.983	48.55	12.18
Pordic	42190.471	48.57	-2.82
Passau	16158.576	48.57	13.46
Morlaix	2473.126	48.58	-3.83
Montlhery	49513.536	48.64	2.27
Evry	70032.845	48.64	2.43
Houdemont	79542.37	48.65	6.14
Reichstett	62204.787	48.65	7.74
Saint Malo	2776.375	48.65	-2.01
Avranches	2343.468	48.68	-1.36
Nattheim	9982.795	48.69	10.21
Echterdingen	53685.507	48.69	9.16
Pichl	11671.299	48.71	11.48
Donauworth	17533.956	48.71	10.78
Dreux	18476.36	48.73	1.36
Boissy St Léger	54107.078	48.75	2.51
Flers	3267.11	48.75	-0.56
Chevilly-Larue	8.295	48.77	2.37
Bar Le Duc	25229.377	48.77	5.16
Trappes	200417.256	48.77	1.99
Villejuif	89375.126	48.79	2.37
Saint-Cyr-l'Ecole	28.364	48.80	2.07
Versailles	408.023	48.81	2.12
Deggendorf	31918.961	48.83	12.96
Wasseralfingen	2049.149	48.87	10.10
Straubing	54880.661	48.88	12.57
Moglingen	14611.403	48.89	9.14
Bobigny	293.222	48.91	2.44
Villeneuve-la-Garenne	111897.144	48.94	2.33
Chalons-en-Champagne	79046.583	48.95	4.37
Les Meuniers	19479.486	49.00	1.68
Herblay	89609.055	49.00	2.16
La Madeleine	24326.658	49.00	1.15
Karlsruhe	67393.8	49.01	8.40
Regensburg	29292.288	49.01	12.09
Dammartin-en-Goele	84853.445	49.05	2.68
Wousterviller	13.816	49.07	7.02
Saint Lô	1141.288	49.12	-1.10
Heilbronn	93382.533	49.14	9.22

Location	Demand/Weight	Latitude	Longitude
Cagny	91805.347	49.15	-0.25
Verdun	7480.265	49.16	5.39
Obereisesheim	109501.153	49.19	9.20
Maizières Lès Metz	123971.528	49.21	6.16
Reims	74083.956	49.26	4.04
Lacroix St Ouen	14787.232	49.35	2.79
Gouy	112886.504	49.35	1.13
Rothenbach	228935.57	49.43	11.05
Kaiserslautern	196183.283	49.44	7.77
Neckarau	118927.2	49.45	8.49
Le Havre	35554.205	49.50	0.12
Virton	10136.231	49.57	5.53
Cherbourg-Octeville	8852.747	49.64	-1.61
Neubau	29510.725	49.65	12.15
Arlon	168.312	49.68	5.82
Sien	29785.76	49.70	7.50
Kleinheubach	8174.797	49.71	9.20
Trier	36224.38	49.76	6.65
Charleville Mézières	6856.285	49.77	4.72
Würzburg	55611.415	49.79	9.93
Creuben	24079.377	49.84	11.63
Saint-Quentin	11013.785	49.86	3.28
Bamberg	36110.539	49.90	10.89
Amiens	46718.642	49.90	2.30
Aschaffenburg	28925.876	49.97	9.15
Mainz	7373.697	50.00	8.26
Bergrheinfeld	16791.843	50.00	10.18
Dettingen	22179.727	50.04	9.05
Wiesbaden	13946.502	50.08	8.25
Neufeld	111854.324	50.11	8.60
Lichtenborn	58463.864	50.11	6.29
Ruckingen	28370.616	50.16	9.01
Beaurains	65439.282	50.26	2.80
Moschendorf	66399.009	50.29	11.92
Auberchicourt	55181.164	50.33	3.24
Koppelsdorf	12208.137	50.35	11.19
Coblence	2292.72	50.36	7.59
Douai	15.374	50.37	3.07
Neuhausel	109081.597	50.38	7.71
Gosselies	31412.012	50.47	4.43
Béthune	7782.311	50.53	2.65
Kunzell	37178.268	50.55	9.70

Location	Demand/Weight	Latitude	Longitude
Meiningen	11197.816	50.57	10.41
Schlema	3923.582	50.60	12.68
Lezennes	125158.922	50.60	3.11
Condette	18863.194	50.66	1.65
Herstal	14944.402	50.66	5.63
Marcq En Baroeul	1.97	50.67	3.08
Altenkirchen Westerwald	25730.554	50.69	7.64
Breitungen	20588.153	50.76	10.34
Aachen	77263.994	50.77	6.09
Chursdorf	66897.093	50.78	12.25
Chemnitz	64209.048	50.83	12.92
Bad Hersfeld	12098.399	50.87	9.71
Zellik	27866.399	50.88	4.28
Buhl	146088.721	50.91	7.91
Freiberg	10943.679	50.92	13.34
Jena	58417.465	50.92	11.58
Marck	20031.885	50.95	1.96
Kreuztal	5081.54	50.96	7.99
Geilenkirchen Nw	69710.904	50.96	6.12
Mulheim	5.48	50.97	7.00
Buchen	17616.245	50.97	7.72
Mengenich	132055.28	50.97	6.86
Erfurt	23184.588	50.98	11.04
Eisenach	14051.913	50.98	10.31
Weimar	11297.81	50.98	11.33
Bergisch Gladbach	28069.688	50.99	7.13
Fort-Mardyck	32881.378	51.02	2.32
Houthalen Helchteren	46689.049	51.03	5.38
Rochlitz	7488.5	51.05	12.80
Gent	1826.322	51.05	3.72
Dresden	69723.312	51.06	13.73
Bad Berleburg	4611.072	51.06	8.39
Obersteinbach	16374.517	51.11	13.18
Bischofswerda	6824.363	51.14	14.18
Golitz	16731.112	51.14	14.99
Hilden	2018.651	51.17	6.93
Balen	55541.602	51.17	5.17
Bautzen	12002.372	51.17	14.42
Remscheid	21217.287	51.18	7.19
Herentals	28192.854	51.18	4.83
Antwerpen	38435.65	51.22	4.41
Morsenbroich	139219.847	51.25	6.81

Location	Demand/Weight	Latitude	Longitude
Schnarum	18663.338	51.26	7.61
Dingelstadt	33290.771	51.31	10.32
Kassel	51009.632	51.32	9.51
Krefeld	16050.288	51.33	6.57
Leipzig	102763.266	51.34	12.37
Oestrich	2401.759	51.36	7.64
Hagen	16553.519	51.36	7.48
Venlo	23457.699	51.37	6.17
Duisburg	22990.016	51.44	6.81
Frillendorf	8261.885	51.46	7.05
Essen	9246.186	51.46	7.01
Nordhausen	11905.286	51.50	10.80
Son	27659.436	51.51	5.49
Elisenhof	32784.503	51.54	8.84
Tilburg	8422.356	51.55	5.08
Etten Leur	56138.351	51.57	4.64
Erwitte	36926.853	51.61	8.34
Watford Herts	2286158.757	51.66	-0.40
Cottbus	49556.701	51.76	14.34
Rees	37236.348	51.77	6.39
Almkerk	29329.876	51.77	4.96
Bernburg	25710.141	51.80	11.74
Holzminden	53889.234	51.83	9.45
Dessau	28481.861	51.83	12.24
Nijmegen	27404.471	51.84	5.86
Spijkenisse	37309.898	51.85	4.33
Spexard	9385.533	51.89	8.43
Tiel	4490.453	51.89	5.44
Clarholz	21948.205	51.90	8.19
Rotterdam	38431.987	51.92	4.50
Munster	125488.525	51.96	7.67
Alfeld	31432.289	51.99	9.82
Bielefeld	29635.3	52.02	8.53
Utrecht	17160.964	52.10	5.11
Nordwest	22167.81	52.14	11.62
Barneveld	63447.45	52.15	5.59
Hildesheim	16349.503	52.15	9.94
Amersfoort	18497.932	52.16	5.39
Ebendorf	9713.648	52.18	11.57
Bad Oeynhausen	9537.708	52.21	8.80
Apeldoorn	14402.337	52.21	5.96
Wiesenau	9812.845	52.23	14.60

Location	Demand/Weight	Latitude	Longitude
Knochenort	92357.678	52.27	7.96
Braunschweig	50992.181	52.28	10.52
Minden	16484.296	52.29	8.92
Waldheim	217129.735	52.35	9.79
Furstenwalde	12939.435	52.37	14.06
Potsdam	76373.734	52.40	13.06
Gummer	52714.854	52.41	9.52
Munchehagen	13990.902	52.44	9.19
Herzfelde	31265.264	52.48	13.85
Gifhorn	9892.21	52.48	10.55
Beverwijk	14802.975	52.49	4.66
Berlin	133732.765	52.53	13.41
Meppel	43570.904	52.70	6.20
Meppen	90.95	52.70	7.29
Oranienburg	40927.374	52.75	13.24
Cloppenburg	8192.6	52.85	8.04
Delmenhorst	16085.216	53.05	8.62
Bremen	84453.909	53.08	8.80
Veendam	14076.108	53.10	6.88
Winsford Ches	209331.487	53.19	-2.51
Groningen	1.091	53.22	6.56
Ochtmissen	16625.265	53.28	10.40
Neustrelitz	51163.684	53.36	13.06
Pasewalk	11671.402	53.51	14.00
Wilhelmshaven	8078.9	53.53	8.12
Hamburg	149909.107	53.55	10.01
Osteel	71145.884	53.56	7.25
Flottbek	17957.724	53.57	9.87
Schwerin	91837.612	53.63	11.41
Pinneberg	13860.652	53.66	9.79
Itzehoe	7371.131	53.93	9.51
Wittorf	10219.97	54.05	9.98
Reutershagen	32853.471	54.11	12.06
Krummendorf	48467.739	54.14	12.12
Stralsund	12261.867	54.30	13.09
Gaarden	56907.709	54.31	10.14
Maribo	6043.47	54.77	11.51
Friesischer Berg	27848.435	54.78	9.41
Flensburg	42462.829	54.78	9.44
Vøjens	12984.334	55.25	9.30
Odense	26095.153	55.40	10.39
Esbjerg	25279.809	55.46	8.44

Location	Demand/Weight	Latitude	Longitude
Vejen	11466.35	55.48	9.14
Kølding	18692.861	55.48	9.47
Glostrup	64388.25	55.65	12.39
Vejle	58542.379	55.70	9.54
Herning	20599.972	56.13	8.96
Århus	25242.624	56.15	10.21
Holstebro	15798.405	56.35	8.62
Gislaved	2158.138	57.30	13.54
Jönköping	2706.656	57.78	14.17
Norrköping	2619.763	58.60	16.16
Sandnes	8967.588	58.85	5.73
Stockholm	33389.937	59.30	18.07
Oslo	10425.123	59.91	10.75
Espoo	122765.713	60.22	24.67
Turku	1742.542	60.46	22.23
Kuopio	346.596	62.91	27.68
Trondheim	10160.498	63.43	10.40
Kokkola/karleby	2349.225	63.85	23.12
Steinkjer	3535.305	64.00	11.51
Tromsø	5606.698	69.70	18.95

UNIVERSIDAD DE LA SABANA
INSTITUTO DE POSTGRADOS- FORUM

RESUMEN ANALÍTICO DE INVESTIGACIÓN (R.A.I)

ORIENTACIONES PARA SU ELABORACIÓN:

El Resumen Analítico de Investigación (RAI) debe ser elaborado en Excel según el siguiente formato registrando la información exigida de acuerdo la descripción de cada variable. Debe ser revisado por el asesor(a) del proyecto. EL RAI se presenta (quema) en el mismo CD-Room del proyecto.

No.	VARIABLES	DESCRIPCIÓN DE LA VARIABLE
1	NOMBRE DEL POSTGRADO	Especialización en Gerencia Logística y Cadena de Abastecimiento
2	TÍTULO DEL PROYECTO	ASSESSMENT OF THE CENTRE OF GRAVITY TECHNIQUE FOR THE SOLUTION OF THE FACILITY LOCATION PROBLEM
3	AUTOR(es)	Daniel Franco
4	AÑO Y MES	August 2010.
5	NOMBRE DEL ASESOR(a)	Dr Andrew Palmer
6	DESCRIPCIÓN O ABSTRACT	Español: El principal objetivo de esta tesis es identificar las principales características, fortalezas y debilidades del modelo de centro de gravedad (COG) utilizado para resolver The Facility Location Problem (FLP) y analizar si la agrupación de los consumidores por medio de closterizacion algorítmica, puede mejorar la solución dada por el modelo. Los resultados obtenidos muestran que el modelo COG puede ser usado como guía y no como una herramienta definitiva para la toma de decisiones, ya que debido a su simplicidad tiene fuertes debilidades. También se muestra que al utilizar modelos de closterizacion algorítmica los resultados del modelo pueden ser mejorados considerablemente. Para finalizar se realizan sugerencias para futuras investigaciones del modelo COG. English: The aim of this thesis is to identify the main characteristics, strengths and weaknesses of the Centre of Gravity (COG) model for solving the Facility Location Problem (FLP). Furthermore, it also looks at how by grouping the customer database using clustering algorithms the solution given by the model can be improved. The results show that the COG model can be used for solving the FLP, however, it should be used only as a guideline and not as a decision tool since it has some very critical weaknesses due to its simplicity. They also show that the solution given by the model can be
7	PALABRAS CLAVES	Clustering; Logistics and Supply Chain; Distribution Network Design; Gustafson-Kessel; Fuzzy c-mean.
8	SECTOR ECONÓMICO AL QUE PERTENECE EL PROYECTO	Sector financiero, comercio, educación, exportaciones, manufactura, servicios transporte, etc. Se recomienda consultar al DANE o Cámara de Comercio.
9	TIPO DE ESTUDIO	Proyecto de investigación.
10	OBJETIVO GENERAL	Can the performance of the COG model for the solution of the FLP be improved by the application of different clustering algorithms?
11	OBJETIVOS ESPECÍFICOS	Identify the main characteristics that encompass the Facility Location Problem within Logistics and Supply Chain Management. Define the Centre of Gravity technique to solve the Facility Location Problem and identify its strengths and weaknesses. Apply different Clustering Algorithms to client databases and analyse the solutions given in terms of improvement to distribution costs.
12	RESUMEN GENERAL	In a market where globalization is becoming a norm more than a trend the company's supply chains are expected to cover these new necessities and cope with the new demands that clients have. In order to achieve this, the use of decision tools is becoming more important as the number of variables increase exponentially. Clients located around the globe, products manufactured in various countries, numbers of parts, volumes, client's expectations, legislations, resource limitations are just some of the variables that supply chain managers have to take into consideration when planning a new supply chain or remodelling an existing one. With this new complexity, the tools used to help the decision making process should also be reviewed in order to make sure that they have optimum results under these new market conditions. One of the problems that supply chain managers have to consider is the location of the facilities that will support the distribution of the products. For this, the centre of gravity mode has been used as a simple and efficient way of solving the Facility Location Problem (FLP) throughout the years. The model is based on the weighted distances of the different clients that the company wishes to serve. However, as supply chains are becoming more complex than ever with the globalization of the markets, the simplicity of the model raises some questions on its effectiveness and true functionality as a stand-alone decision tool for supply chain managers under the new conditions. Some evidence has been found showing that the model can have under-optimal solutions when the problem to be solved presents some specific characteristics that are commonly present in today's supply chains. Literature has highlighted that the model can have problems when there is one dominant client that represents more than half of the total demand and also when clients are located a long way from the rest of the demand nodes. However, there may be ways of overcoming some of these issues, if the centre of gravity were to be used not as a stand-alone decision

13	CONCLUSIONES.	In concluding this thesis, the objectives of this research were: • Identify the main characteristics that encompass the Facility Location Problem within Logistics and Supply Chain Management. This has been done following a literature review on the FLP and how it impacts on the supply chain. The findings are that the FLP is a key part of the designing phase of the supply chain and it has a great impact on the overall performance of the future distribution network. The FLP impacts on decisions that are usually considered for long term and require high investments such as the opening or closing of a distribution centre or the reassignment of clients to existing DCs. The FLP was identified to be categorized depending on the characteristics of the problem and the variables that should be taken into account. However, due to the complexity of the problems, it is currently unclear on how to have models that can help solve multi-variable FLP. • Define the Centre of Gravity technique to solve the Facility Location Problem and identify its strengths and weaknesses. This has been done by an exhaustive literature review coming from seminal works such as Weber (1909) and Cooper (1961) to more recent works that utilize the COG as a hybrid method with other modelling techniques such as clustering techniques like that presented by Esnaf and Küçükdeniz (2009). The COG is identified as a linear equation defined in equation 1. The main strength that was found is the fact that the COG is a simple model that can help identify possible locations where to place the DCs in order to minimize the weighted distance to the customers. However, some weaknesses were also found that put to doubt the applicability of the COG as a standalone decision tool. The fact that the COG can fall anywhere in the plane irrespective of geographical limitations is an example of such weaknesses. • Apply different Clustering Algorithms to client databases and analyse the solutions given in terms of improvement to distribution costs. After going through some bibliography and identifying the characteristics of clustering algorithms, how they work and the different classifications, three different algorithms were chosen for applying onto the databases. Additionally, in order to review the impact that the hybrid model has on databases of different sizes, three databases of varying sizes were chosen. From the research done on clustering algorithms, it was identified that one of the main inputs for the algorithms is the number of clusters into which the dataset is to be divided. Therefore, in order to quantify the impact that this decision has on the final result, each dataset was analysed using each of the three algorithms with different number of clusters from 1 cluster up to 8 clusters. Each of the results of the algorithm was then passed through a model developed by the author in order to quantify the distribution costs
14	FUENTES BIBLIOGRÁFICAS	Alp, O., Erkut, E. and Drezner, Z. (2003), "An Efficient Genetic Algorithm for the p-Median Problem", Annals of Operations Research, vol. 122, no. 1/4, p. 21. Arostegui, M. A.,Jr, Kadipasaoglu, S. N. and Khumawala, B. M. (2006), "An empirical comparison of Tabu Search, Simulated Annealing, and Genetic Algorithms for facilities location problems", International Journal of Production Economics, vol. 103, no. 2, p. 742. Ballou, R. H. (1973), "Potential Error in the Center of Gravity Approach to Facility Location", Transportation Journal, vol. 13, no. 2, pp. 44-50. Ballou, R.H. (1999), “Business Logistics Management: Planning, Organization & Controlling the Supply Chain”, Modern Purchasing, vol. 42, no. 11, p. 14, Toronto Bezdek, J.C., (1981), “Pattern Recognition with Fuzzy Objective Function Algorithms”, New York: Plenum Press. Bose, I. and Chen, X. (2009), “A Method for Extension of Generative Topographic Mapping for Fuzzy Clustering”, Journal of the American Society for Information Science and Technology, vol. 60, no. 2, pp. 363-371. Chen, C. and Ting, C. (2008), "Combining Lagrangian heuristic and Ant Colony System to solve the Single Source Capacitated Facility Location Problem", Transportation Research.Part E, Logistics & Transportation Review, vol. 44, no. 6, p. 1099. Chistopher, M. (1992), “Logistics and Supply Chain Management”, Pitmans, London, UK. Cooper, L. (1961), “Location-Allocation Problems”, Operations Research, vol. 11, 331-343. Drezner, Z. (1995), “Facility Location: A Survey of Applications and Methods”, Springer Verlag. Drezner, T. and Drezner, Z. (2007), "The gravity p-median model", European Journal of Operational Research, vol. 179, no. 3, p. 1239. Eiselt, H.A., and Sandblom, C.-L. (2004), “Decision Analysis, Location Models, and Scheduling Problems”, Springer-Verlag. Erlenkotter, D. (1978), "A Dual-Based Procedure for Uncapacitated Facility Location", Operations Research, vol. 26, no. 6, p. 992.

Vo Bo Asesor y Coordinador de Investigación:

CRISANTO QUIROGA OTÁLORA