Abstract

The Supply Network Design represents one of the high impact strategic decisions in competitiveness for companies. An optimal location of facilities in relation to the capacity of supply and demand, allows a high level service to attend the markets. In this paper, a methodological framework for designing supply networks by joining application of clustering techniques and mathematical programming are presented. The proposed methodology has been tested with real data obtained from a company of non-alcoholic beverages in Colombia. The approach considers three main stages. First, the customers clustering process is performed by K-means in order to obtain the location for potential Distribution Centers (DC’s). In the second stage, the model for supply network design is performed using a Mixed Integer Programming (MIP) by considering different options to assign DC’s, and finally the valuation of the proposed methodology on a real case. A distribution scheme, which allows enter to new market areas with an efficient strategy to penetrate products to big cities such as Bogotá in Colombia, was found.

Key words: Clustering, Mixed Integer Programming, Supply Chain Management, Supply Network Design

1. INTRODUCTION

The growth of an organization is subject to market conditions and capabilities, which must be managed for the long term. The previous statement is distinguished as a key strategic issue in managing the supply chain. The Supply Network Design (SND) represents the impact of strategic decisions on the competitiveness of an organization, defining the number, location, and capacity of the warehouses and manufacturing plants, or the flow of material along the logistics network. [1]. This article addresses the issue of the Supply Network Design (SND) for industries, by considering the growth and aspiration to enter new markets, and also by changing their distribution strategies and schemes of operation, with the inclusion of new warehouse locations, outlets and a heterogeneous fleet for product delivery. The proposed methodology has been tested on real data obtained of the case of an industry dedicated to the regional production and distribution of Non-Alcoholic Beverages (NAB) in Colombia. Non-Alcoholic Beverages in Colombia remain at a constant increase due to two primary forces: increased income and poverty reduction. This second element has been one of the principal reasons to generate an increase in Non-Alcoholic Beverages (NAB) consumed in the previous years, despite a considerable decrease between 2010 and 2013 of about 37%. In other words, approximately 1.7 millions of people have become part of the middle class and have converted into active market consumers [2]. As a result of the previous fact, there exists benefits and challenges for the big companies of the sector, but also for the medium-sized industries which are not part of the national distribution in a big scale, but that look for space to sell their products in small and big cities.

This article addresses the problem of designing a distribution network for companies that sell soft drinks (NAB). The problem has been split into three stages: i) generation of clusters through grouping of mathematical techniques to determine the potential customers of distribution centers locations; ii) development of a mathematical model of mixed integer programming for network design; iii) validation of the proposed methodology in a company dedicated to the production...
and distribution of NAB. The proposed model considers a distribution system - inventory of three links in the supply chain (factories, distribution centers and areas of consumption), in which decisions must be made for expansion or contraction of distribution centers, allowing tactical decisions and strategic logistics network.

The main contribution of the paper is to examine the applicability and effectiveness of a mixed methodology by combining clustering and Mixed Integer Linear Programming in a real case study associated with a supply chain of consumer products NAB. According to the literature reviewed, there are few articles that include methodologies for design distribution networks with clustering techniques to solve real cases.

Section 2 presents relevant literature with supply chain management and clustering technique. The section 3 offers the proposed methodology. Finally, computational results and conclusions are presented in the section 4.

2. LITERATURE REVIEW

Part of the planning process of the Supply Chain Management (SCM), is oriented to find the best possible configuration of the network. The process considers decisions related to the location of facilities, the production and the inventory decisions, as well the distribution systems by considering routing decisions [3].

2.1 Supply network design and management

Many studies related to the supply network design SND problem, or SCND (Supply Chain Network Design) have been published [4–13]. From the conceptualization and scope of the supply network design, these works consider three levels: strategic, tactical and operational [1], [8], [9] and [10]. In particular, the fulfillment of demand by a set of performed routes, the pricing, the service level, are elements related with the level of operational decisions. The amount to be transported from the network locations, the transport mode, volume, type of inventory, quantities to buy from suppliers, the type of information technology and knowledge management are considered as problems of the tactical level.

Finally, the decision of the number and the location of the echelons in the network, its capability, the quality, the technology type, the number of contracted suppliers, and the reserve capacity of suppliers and decompositions points are decisions of strategy level [10] and [12].

The previously published works present a base of the decision levels in SND environment; however, they can specific decisions between the combination of levels (especially in the tactical and operational level), where the route and fleet definition results are a high impact on the performance of the distribution [14]. The considered problem in this work is related to decision of strategic and tactical level.

2.2 Supply network design by using mathematical programming and clustering techniques

Given the nature of the mathematical models of the supply network design and the impact of their decisions throughout the supply system for a company; the field related to the supply network design is quite extensive. Some specific areas related to the network design are: deterministic models of echelons, deterministic models of multiples echelons and multiple product, design methodology by using exact methods, design methodology by considering simulation design methodology using heuristics and metaheuristics, dynamics network design models, and stochastic models [7]. Many of these areas have been integrated into the study of the problem by experts and academics, achieving important developments in the evolution of the supply network design theory [4–13].

In addition, different objective functions have been considered to address different network design problems [7 – 8]. Works [11] and [15] show a complete classification of modeling approaches used in the network design, including models with one or more objectives, and deterministic and stochastic approaches. In the literature reviewed, it is common to find deterministic applications with linear or nonlinear mixed formulation such as [16–20]. In addition, applications including uncertainty of parameters [4–8] and [21–23]. Generally binary variables are used to determine the opening or closing decisions of the supply network design [23].

In the last decade, different approaches have been proposed to group up installations in the supply networks [24], where the coordination between agents demonstrates synergies that allow improve the performance of the supply system, [25], specifically for emergent markets [26]. In summary, the clustering technique brings benefits from the point of view of SCM as well as from their joint application to the mathematical programming to respond to the SND problem [27–29].

3. METHODOLOGY

The development of the investigation was conducted into three stages: i) determine a possible number of groups of customers to select the potential location of the distribution centers. It is possible to determine a superior and inferior limit to find the optimal number of clusters of a considered solution area, when the number of groups or clusters is unknown, ii) development of a deterministic mathematical model of mixed integer programming for the network design; and iii) validation of the proposed methodology in the study case for a company.

The case studied company is dedicated to NAB production and distribution in the department of Cundinamarca in Colombia. The growing projection of the company is venturing into new consumption areas.
3.1 Grouping and Modelling for a supply network design

In this paper, the clustering algorithm k-means has been used, given its simplicity and effectiveness to the problem to generate clusters in similar logistics problems. The general idea of the algorithm is determining the best location of distribution centers that reflect a customer group. The algorithm description is shown below:

### 3.1.1 Clustering K-Means algorithm

Since group of objects is $X$ and an integer is $k \leq n$, the algorithm K-means tries to find the participation of $X$ in $k$ groups that minimize the sum of errors squared in the groups (WGSS), by its acronym in English *Within groups sum of Scared errors* [30]. This process is commonly formulated as show below:

$$
\text{Min} \ P(W, Q) = \sum_{i=1}^{k} \sum_{l=1}^{n} w_{il}d(X_i, Q_l)
$$

Subject to

$$
\sum_{l=1}^{k} w_{il} = 1 \quad 1 \leq i \leq n
$$

$$
\sum_{i=1}^{n} w_{il} \in \{0,1\} \quad 1 \leq i \leq n; \quad 1 \leq l \leq k
$$

$w_{il}$ is a binary variable represented with the value of 1 if object $i$ is assigned to group $l$, otherwise the variable is represented with the value of 0. $W$ is a matrix partition $x k$, $Q = \{Q_1, Q_2, ..., Q_k\}$ is a set of objects in the same domain, and $d(.,.)$ is the Euclidian distance between two objects. The restriction (2) assures that one object is assigned to one group; meanwhile the restriction (3) indicates the integrality of the variables. Once the K-means algorithm is applied, the potential location of each distribution center (assuming that one distribution center serves a group of customers) will be equivalent to the geographic coordinates of the center of gravity for each cluster. In the case that the center of gravity corresponds to a prohibited geographic site, the distribution center is then located in the nearest possible site. This way, the potential number and location of the distribution centers (subscript $k$ of the mathematical model) to be considered in the mathematical model is described below:

### 3.1.2 Mathematical model proposed for the SND problem

**General characteristics and assumptions**

The general characteristics and assumptions of the deterministic mathematical model of the supply network design are as follows:

- The entire physical network infrastructure is assumed inside of one unique country, without considering international physical distribution.
- The objective is minimize total logistics costs, emphasizing storage costs, fixed operating costs, opening and closing distribution centers, transportation costs and manipulation product cost.
- The model is designed to make relative consideration for a unique period of planning and
accepts the consideration of multiple products (beverage).
- The model includes as decision variables, the closure and consolidation of distribution centers and manufacturing and shipping flow of products through the network.
- We consider a multi-echelon distribution system. One of the echelons is constituted by the center distribution (CD’s), which is of sizable magnitude and has a considerable amount of inventory.
- The proposed model starts from an infrastructure of plants and center of distribution, which are already established, and it seeks to review the closure, opening and consolidating the distribution operation in the CD’s.
- We consider the capability and storage constraints for each echelon.
- The plants could send the finished product to the CD’s, that is to say that they are not considered direct shipments between plants and customers.
- It is considerate as a unique mode of transportation overland by truck (the proposed model not include the decision of mode of transportation), and to not include decisions on truck types.
- The model does not explicitly take into account financial considerations relative to taxes and tax benefits typical to marketing processes.

Index

- \( i \) Beverage type \((i = 1 \ldots m)\)
- \( j \) Fabric type \((j = 1 \ldots n)\)
- \( k \) Type of Intermediate CD \((k = 1 \ldots K)\)
- \( l \) Consume zone type \((l = 1 \ldots L)\)
- \( v \) Truck type \((v = 1 \ldots V)\)

Parameters

The parameters consider all the information (data) required for the development of the proposed model.

- \( CAIm_k \) Maximum storage capacity in the intermediate distribution center of type \( k \)
- \( CAIm\text{\textsubscript{min}}_k \) Minimum storage capacity in the intermediate distribution center of type \( k \)
- \( NMAX_v \) Maximum permitted number of trips per month by truck of type \( v \)
- \( NCam_v \) Available number of trucks type \( v \)
- \( CCag_v \) Capacity in units per truck of type \( v \)
- \( CF_k \) Fixed cost of operating store type \( k \)
- \( Cm_k \) Variable cost for operating a unit in the store \( k \)
- \( S_v \) Enlistment cost of one trip in the truck type \( v \)
- \( CKM_v \) Cost per traveled kilometer for the truck type \( v \)
- \( CDESC_v \) Discharge cost for the truck type \( v \) when it reaches the target site
- \( absCD_k \) X coordinate for the intermediate distribution center of type \( k \)
- \( absCP_j \) X coordinate for the plant type \( j \)
- \( OrdCP_k \) Y coordinate for the plant type \( k \)
- \( absZC_l \) X coordinate for the consume zone type \( l \)
- \( OrdZC_l \) Y coordinate for the consume zone type \( l \)
- \( Dem_{ij} \) Estimate demand for the beverage type \( i \) in the consume zone type \( l \)
- \( Cof_{ij} \) Supply capacity for the beverage type \( i \) in the plant type \( l \)
- \( CPr_{ijk} \) Unit cost production of the beverage type \( i \) in the plant type \( j \), which is taken to the intermediate distribution center type \( k \)
- \( D1_{jk} \) Distance between the plant type \( j \) and the intermediate distribution center type \( k \).

Where:

\[
D1_{jk} = \sqrt{(absCD_k - absCP_j)^2 + (OrdCD_k - OrdCP_j)^2} \quad (4)
\]

\( D2_{kl} \) Distance between the Intermediate distribution center of type \( k \) and the Consume zone type \( l \).

Where:

\[
D2_{kl} = \sqrt{(absZC_l - absCD_k)^2 + (OrdZC_l - OrdCD_k)^2} \quad (5)
\]

\( CTrans_{jkv} \) Transport cost from the plant type \( j \) to the Intermediate distribution center of type \( k \) in the truck type \( v \).

Where:

\[
CTrans_{jkv} = S_v + (D1_{jk} * CKM_v) + CDESC_v \quad (6)
\]

\( CTrans_{klv} \) Transport cost from the Intermediate distribution center of type \( k \) to the Consume zone type \( l \) in the truck type \( v \).

Where:

\[
CTrans_{klv} = S_v + (D2_{kl} * CKM_v) + CDESC_v \quad (7)
\]

Scalar

- \( p \) Indicates the number of intermediate stores CD to wish to open

Variables

- \( X_{ijk} \) Variable that determines the beverage count type \( i \) made in fabric type \( j \) and sent to the distribution center type \( k \).
- \( Y_{ijk} \) Variable that determines the beverage count type \( i \) from the distribution center type \( k \) to the consumer zone type \( l \).
- \( BIN1_k \) Binary variable that determines the assignation of a fixed cost to open the distribution center type \( k \).
- \( BIN2_{kl} \) Binary variable that determines the opening (value=1) or closing (value=0) to the distribution center type \( k \) serving the consume zone type \( l \).
- \( NVA_{jkv} \) Number of trips between plant type \( j \) and the intermediate distribution center type \( k \) in the truck type \( v \).
Constraints that determine the amount of flows of product must be less than the capacity of each factory.

\[ X_{ijk} \leq Cof_{ijk} \ast BIN1_k \quad \forall i, j, k \]

Constraints ensuring the compliance of the demand of the beverage type \( i \) in the consumer zone \( l \), which must be served by the intermediate distribution center type \( k \).

\[ \sum_{k}^{K} Y_{kil} = \sum_{k}^{K} Dem_{il} \ast \frac{BIN2_{kl}}{p} \quad \forall i, l \]

Constraints that determine the number of open centers of distribution type \( k \).

\[ \sum_{k}^{K} BIN2_{kl} = p \quad \forall l \]

Constraints that determine the maximum storage capacity for each CD type \( k \).

\[ \sum_{l}^{L} Dem_{il} \ast \frac{BIN2_{kl}}{p} \leq Calm_k \ast BIN1_k \quad \forall k \]

Constraints that determine the minimum storage capacity for each CD type \( k \).

\[ \sum_{l}^{L} Dem_{il} \ast \frac{BIN2_{kl}}{p} \geq Calmin_k \ast BIN1_k \quad \forall k \]
measurement areas. The most common consumption zones in Colombia, where the product is distributed are: Girardot, Melgar, La Mesa, Mesitas, Fusagasugá, Carmen de Apicalá, Nariño, Agua de Dios and Flandes (Figure 2). The main products marketed are sodas, juices and water in presentations of 335ml, 600ml, 1 liter and 2 liters, for sodas and juices; water is produced and packed by PET bottle 600ml, 5.5 liter bags, 340ml, 600ml bag; 30ml and refreshments.

4.1. Results and Discussion

As shown in Fig.2 there are three possible areas of consumption in the city of Bogotá D.C., which is a decision generated after an economic study conducted previously. The new zones: Bosa, Fontibón and Engativá, are part of a plan to include product to the Colombian capital. Following the methodology proposed, the results of clustering and optimization are shown below.

4.1.1 K-means clustering

After identifying the consume centers X, it is necessary to identify a reference point in each zone, this way there can be a k number of groups depending on their distances d(·,·) so that later on it is possible to identify possible collection centers for product distribution, the software available for this use is Orange Canvas which has a free version available at (http://orange.biolab.si/) and runs on a computer with processor Corei5 to 1.8HGz and 6GB RAM. The number of clusters is 3 and it is a known priori in this case study.

Figure 3 and 4 show how the clusters are formed, thus providing three groupings: cluster one is conformed by the consume zone of Fusagasugá, La Mesa and Bosa; cluster two conformed of Agua de Dios Girardot, Carmen de Apicalá, Melgar, Flanders and Nariño; Finally, cluster three is formed with the remaining consume zones, such as, Mesitas, Engativá and Fontibón.

4.1.2 Identification of distribution centers

Once the clusters were obtained, the possible areas for new distribution centers to be located are identified, this way there is a benefit when the consume zones that are part of the group are highly populated. Figure 5 shows the available areas where the new locations could be located, which in summary are: for the first cluster, Granada (a), for second cluster Ricaurte (b) and for the last cluster (c) Mosquera. It is noteworthy that in the case of the third cluster, Mosquera was chosen as a potential location since it has an industrial potential and its proximity to the consume zones in that group.
Continuing the proposed methodology with the previous results, we have different possibilities of designs for the distribution network with the computational results shown below (see Table. 1). These results were obtained with the GAMS software (General Algebraic Modelling System, which is available in http://www.gams.com/download/) for use on a computer with the same characteristics mentioned in the previous section. It identifies that the given characteristics of a binary variable $BIN_{2k}$ is assignable to each CD $k$ opened, one consume area $l$ to attend, generating for each unit increase a value $p$, a decrease in the variable cost of the total network. In other words, the sum of the associated cost to produce, stored and carried out, given by $X_{ijk}$, $C_m$, $CTransA_{kiv}$, $CTransB_{klv}$, decrease primarily by reducing distances to travel by transport. On the contrary, given the unit increase of $p$, the total fixed cost $\sum_k CF_k$ increases (See. Table 2).

The final design recommended for the case study is given by the factory, the distribution center CD2 and CD5 Ricaurte and Bosa. Since it denotes lower logistics costs and better use of resources, as evidenced by Figure 7.

**Figure 5.** Georeferencing scheme for possible locations of new Distribution centers in (a) cluster one, (b) cluster two and (c) cluster three. Source: Authors

**Figure 6.** Geographical location for each facility. Source: Authors

**Objective Function values y variable costs for different network design**

<table>
<thead>
<tr>
<th>p Value</th>
<th>Total Cost</th>
<th>Variable Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ricaurte (CD2)</td>
<td>$24,044,844.805</td>
<td>$24,044,844.805</td>
</tr>
<tr>
<td>Ricaurte, Bosa (CD2, CD5)</td>
<td>$23,744,844.805</td>
<td>$22,724,530.432</td>
</tr>
<tr>
<td>Ricaurte, Bosa, Fontibón (CD2, CD5, CD6)</td>
<td>$23,356,502.139</td>
<td>$22,496,502.139</td>
</tr>
</tbody>
</table>

**Figure 7.** Values for the objective function and variable cost to different network designs Source: Authors
Table 2. Comparison of total costs in the network to values of $p$ between 1 and 3. Source: Authors.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Distribution Centers opened</th>
<th>Total Cost Distribution C. (Objective Function) ($/t)</th>
<th>Fixed Cost of CV</th>
<th>Variable Cost Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buriticá</td>
<td>Ref11.CD2</td>
<td>23.356.502.139</td>
<td>24.044.844.805</td>
<td>300.000.000</td>
</tr>
<tr>
<td>1</td>
<td>Ref11.CD5</td>
<td>23.356.502.139</td>
<td>24.044.844.805</td>
<td>300.000.000</td>
</tr>
<tr>
<td>2</td>
<td>Ref9.CD2</td>
<td>22.486.502.139</td>
<td>24.044.844.805</td>
<td>300.000.000</td>
</tr>
<tr>
<td>3</td>
<td>Ref9.CD2</td>
<td>22.486.502.139</td>
<td>24.044.844.805</td>
<td>300.000.000</td>
</tr>
</tbody>
</table>

Table 1. Computational results for possible network designs. Source: Authors.

<table>
<thead>
<tr>
<th>Value of $p$</th>
<th>Núcleo</th>
<th>La Mesa</th>
<th>Carmen</th>
<th>Flandes</th>
<th>Nariño</th>
<th>Agua de D.</th>
<th>Finque</th>
<th>Tigmoid</th>
<th>Costo Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ref1.CD6</td>
<td>21.044.941.805</td>
<td>20.044.844.805</td>
<td>20.000.000</td>
<td>22.443.502.139</td>
<td>23.356.502.139</td>
<td>24.044.844.805</td>
<td>300.000.000</td>
<td>22.486.502.139</td>
</tr>
<tr>
<td>2</td>
<td>Ref1.CD5</td>
<td>21.044.941.805</td>
<td>20.044.844.805</td>
<td>20.000.000</td>
<td>22.443.502.139</td>
<td>23.356.502.139</td>
<td>24.044.844.805</td>
<td>300.000.000</td>
<td>22.486.502.139</td>
</tr>
<tr>
<td>3</td>
<td>Ref1.CD2</td>
<td>21.044.941.805</td>
<td>20.044.844.805</td>
<td>20.000.000</td>
<td>22.443.502.139</td>
<td>23.356.502.139</td>
<td>24.044.844.805</td>
<td>300.000.000</td>
<td>22.486.502.139</td>
</tr>
</tbody>
</table>

Total Cost: 660,000,000
5. CONCLUDING REMARKS

The development and implementation of the proposed methodology provide significant results to conclude that in the case study, different options for network designs are obtained and that after the computer runs with p = 1, p = 2 and p = 3. The obtained results are better when comparing network costs for the three distribution centers available.

In addition to determining the overall use of clustering to group consumption centers and mathematical programming, it turns out to be an option to organize the scheme layout and design of the network to different scenarios. Thus, the subsequent modeling process provides results that are close to the operational reality of the system and serves as support for the process of decision-making at a strategic and tactical scenario. Consequently, the strategy of entering new markets for the organization of the case study, dedicated to the production and regional distribution of NAB, is supported by the coverage of the paradigm of managing the supply chain SCM and important phase of the process planning supply network, known as network Design supply SND or SCND.

It is important to mention that in adopting the mathematical model together with K-means clustering, may propose additional changes in the conformation of groups using other clustering algorithm, which would entail changes to the distribution scheme found at the optimum network design.

Finally after reviewing the results, network designs with $p=1$ and $p=3$ are discarded due to its high logistics costs and due to dynamic distribution that could arise, for example in the case of $p=1$, if only the CD 2 (Ricaurte) is opened, it would be required to bring the product to the town of Ricaurte, store it and then distribute it to the surrounding municipalities of the original manufacturing site of NTBs; in this case the variable cost of operating in the network would be 7.62% higher than that of $p=2$. Similarly, in the case of $p=3$, if only CD2 (Ricaurte), CD5 (Bosa) and CD6 (Fontibón) are opened, the same situation would be presented, showing a variable cost of operating the network of 1.96% above the option $p=2$.

For future research, it is recommended that a process of experimentation with different clustering algorithms and the inclusion of alternative routes and sequencing of vehicles to make the modeling process more complete in its approach to the operational reality of companies looking to design their network and serve new consumption areas which are characterized by a demand of high level of service.

6. REFERENCES


Dizajn distributivne mreže primenom klastera i mešovitog programiranja celih brojeva

Nicolás Clavijo Buriticá, John Wilmer Escobar, Rafael Gutiérrez

Primljen (17.07.2017.); Recenziran (27.03.2018.); Prihvaćen (30.05.2018.)

Apstrakt


Ključne reči: Klasteriранje, mešovito programiranje celih brojeva, upravljanje lancima snabdevanja, dizajn distributivne mreže