NETWORK DESIGN USING MIX INTEGER PROGRAMMING AND MONTECARLO SIMULATION IN AN INTERNATIONAL SUPPLY CHAIN NETWORK.

Eduardo Villareal[(a)], Idalia Flores[(b)]

(a) Universidad Nacional Autónoma de México, México
(b) Universidad Nacional Autónoma de México, México.

(a) eduardo_vp0914@yahoo.com.mx (b) idalia@unam.mx

ABSTRACT

In this paper a methodology was developed to combine optimization techniques whit Montecarlo Simulation to design an International Supply Chain Network in Mexico. The case of study was taken from a business where an optimal network is defined by choosing from a variety of distribution centers. For the defined network there are 2 kinds of products, 18 distribution centers and more than 3000 retailers that have been grouped in 73 locations. Also, the model has considered 6 types of transportation modes. The optimization model has an interface built on Microsoft Excel, a model formulation that was built in Lingo V.8.0 and a geographic interface built in Microsoft Map Point 2009. For the simulation model we considered stochastic demand and some cost parameters such as the fixed cost for the Distribution Centers and maritime cost for the international flow to Mexico.

Keywords: optimization techniques, montacarlo simulation, supply chain network

1. INTRODUCTION

According to Bramel and Levi (1997) one of the most important aspects of logistics is the location of a new distribution center to provide a service to one or more customers, warehouses or factories within the supply chain.

The problem of choosing the location of new distribution centers or plants within a logistics network consists of optimizing the transport cost throughout the network by incorporating new elements into the distribution network while complying with the constraints on demand and storage capacity (to name but a few).

2. BACKGROUND

Location problems can by looked at from three points of view:

- As an assignment problem
- As a minimum cost maximum flow problem
- As a combination of the above two

Different authors, including Bramel (1997) as well as Levi, Wu, Shen and Levi (2004), analyze location problems as mixed linear programming problems or else, as problems with nonlinear cost functions. The presence of nonlinearity in a location problem implies that economies of scale or fixed costs have been introduced into the cost function. According with Dalila (2006) in the paper entitled Heuristic Solutions for General Concave Minimum Cost Network Flow Problems, specifies that when the effect of economies of scale or fixed costs is added in, the cost function becomes concave and thus becomes a convex space bounded by a nonlinear hyperplane. Likewise Kouvelis (2004) incorporates economies of scale, fixed costs and nonlinear variable costs for a problem of network design.

The limitations involved in using an assignment problem to define the location of a new distribution center are:

- It only considers the rental costs of each store.
- Flows through the network are not taken into account.
- The demand of each one of the customers situated all along the network are not considered.
- The interaction between all the parts of the distribution network is not observed in the formulation of the model and are not considered in the decision-making process.

The cost function is linear Hillier and Liberman (2000).

On the other hand, the limitations involved in using a minimum cost flow model are:

1. It does not consider the fixed costs involved in opening and closing a distribution center.
2. The decision variables only determine the optimum flow within the network.
3. The cost function is linear.

2.1. The Weber Problem

The goal is to find a $P(x^*,y^*)$ point that minimizes the sum of the weighted Euclidean distances of n points with coordinates (a_i,b_i). The weightings associated with the n fixed points are denoted by w_i. When we transfer this problem to the Supply Chain, the problem consists of situating a distribution center with a w_i transport cost associated with the location of the customers on the fixed points (a_i,b_i). Therefore, (x^*,y^*) is the point that minimizes the distribution cost.

The problem can be expressed as follows:

$$\min Z = W(x,y) = \sum_{i=1}^{n} w_i \cdot d_i(x,y) \tag{1}$$

Where $d_i(x,y) = \sqrt{(x-a_i)^2 + (y-b_i)^2}$ is the Euclidean distance between (x,y) and (a_i, b_i)

The simplest way to solve this problem is using the Weiszfeld algorithm[1].

Partially deriving the objective function and making it equal to zero, we get the first-order conditions for ensuring optimality:

$$\frac{\partial}{\partial x} W(x,y) = \sum_{i=1}^{n} \frac{w_i \cdot (x-a_i)}{d_i(x,y)} = 0 \tag{2}$$

$$\frac{\partial}{\partial y} W(x,y) = \sum_{i=1}^{n} \frac{w_i \cdot (y-b_i)}{d_i(x,y)} = 0$$

It can be shown that $W(x,y)$ is convex and the system of equations (2) defines a minimum. However, these derivatives do not exist when (x,y) coincide with the fixed point i because $d_i(x,y)=0$ and therefore the equations in (2) cannot be solved for (x,y) if $n>3$

We can extract x from the first equation in (2) and extract y from the second equation in (2). The result is an iterative procedure if we consider the extracted pair (x,y) as a new iteration $(k+1)$. To be precise:

$$(x_{k+1}, y_{k+1}) = \left(\frac{\sum_{i=1}^{n} \frac{w_i \cdot a_i}{d_i(x_k,y_k)}}{\sum_{i=1}^{n} \frac{w_i}{d_i(x_k,y_k)}}, \frac{\sum_{i=1}^{n} \frac{w_i \cdot b_i}{d_i(x_k,y_k)}}{\sum_{i=1}^{n} \frac{w_i}{d_i(x_k,y_k)}} \right) \tag{3}$$

This is an iterative method for solving the location problem.

2.2. Model for the warehouse location problem.

Let us consider a set of customers spread out within a particular geographic region. The problem is to determine the location of a number p of available warehouses. We assume that there are $m \geq p$ places that have been previously selected as possible locations. Once the location of warehouse p has been determined, each one of the n customers shall be supplied by the nearest warehouse. In this model we assume.

- There is no fixed cost for locating warehouse p.
- There is no constraint on the capacity of any warehouse to meet the demand.

This example was taken from a real company that imports wood from Chile and has some distribution centers as Tampico where the wood arrives,and then the wood is transported for all over the country.

Let:

Set of customers $N = \{1, 2, 3,..., n\}$
Set of possible warehouses $M = \{1, 2, 3 ..., m\}$
Let w_i = the flow of demand between the customer i and their warehouse for the entire $i \in N$
Let c_{ij} = the cost of transporting w_i units from the warehouse j to the customer i for each $i \in N$ y $j \in M$

The problem is to locate p of the m available warehouses in such a way that the transport cost is minimized.

Let:

$$y_j = \begin{cases} 1 \text{ if storage is located at j place} \\ 0 \text{ o. c.} \end{cases}$$

$$x_{ij} = \begin{cases} 1 \text{ if client i is served by storage j} \\ 0 \text{ o.c.} \end{cases}$$

For all $i \in N$ and for all $j \in M$

Thus the problem is formulated as follows:

$$\min Z = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} \cdot x_{ij} \tag{4}$$

Subject to:

$$\sum_{j=1}^{m} x_{ij} = 1 \;\; \forall \; i \in N \tag{5}$$

$$\sum_{j=1}^{m} y_j = p \tag{6}$$

$$x_{ij} \leq y_j \tag{7}$$

for

$$x_{ij}, y_j \in \{0,1\} \;\; \forall \; i \in N \;\; \forall \; j \in M$$

The constraint on (5) ensures that a warehouse is assigned to every customer. The constraint on (6) ensures that a *p* warehouse are assigned and the constraint on (7) ensures that each customer chooses a single warehouse. The problem is integer linear.

2.3. Location of capacitated-constrained warehouses. Let us consider the model for the p-median algorithm under the following hypotheses:

1. The number of warehouse to be located is not a fixed value *p*.
2. A fixed cost f_j is incurred by locating a warehouse in place *j*.

Minimize:

$$Z = \sum_{i=1}^{n}\sum_{j=1}^{m} c_{ij} \cdot x_{ij} + \sum_{j=1}^{m} f_j \cdot y_j \qquad (8)$$

Subject to:

$$\sum_{j=1}^{m} x_{ij} = 1 \quad \forall\ i \in N \qquad (9)$$

$$\sum_{i=1}^{n} w_i \cdot x_{ij} \le q_j \cdot y_j \quad \forall\ j \in M \qquad (10)$$

$$x_{ij}, y_j \in \{0, 1\}\ \forall\ i \in N\ \ \forall\ j \in M$$

The constraint on (9) ensures that a warehouse is assigned to every customer. The constraint on (10) ensures that the capacity of a warehouse is not exceeded and also if there is no warehouse located in j no customer can be assigned to that place.

3. DEFINITION OF VARIABLES AND PARAMETERS

The parameters of the model are defined below.

3.1. Seaport costs

κ_{itw}= Storage capacity d in m^3 for product type *t* for type of vessel *w*.

C_{ijtw} = Total cost per m^3 of transport from place of origin *i* to port *j* for product type *t* in type of vessel *w*.

The cost of sending a boat depends on the following parameters:

- Carriage cost from plant to port in Chile per m^3
- Port expenses per m^3 in Chile
- Cost for secure sea freight per m^3 Chile-Mexican port
- Port expenses per m^3 in Mexican port

3.2. Rail transport

ν_{tw} = Storage capacity in m^3 for product type *t* for freight car type *w*.

In the case of containerized freight, a sea container has the same capacity as a railroad container since it is the same container being transported.

C_{jktw} = Cost per m^3 for transporting from port *j* to Distribution Center *k* for product type *t* for freight car type *w*.

3.3. Storage cost

υ_k = Storage capacity in m^3 of the Distribution Center located in port k.

ω_l = Storage capacity in m^3 of the Distribution Center located in *l*

C_k = Fixed annual cost of the warehouse located in port *k*

C_l = Fixed annual cost of the warehouse located in *l*

The fixed cost of the warehouse is made up by the following categories:

- Costs of management services (cost of the annual salary of the Distribution Center manager).
- Rental cost per m^2.
- Annual cost of equipment depreciation and amortization.
- Workers wages (annual)
- Wages of administrative staff (annual)
- Electric power and other expenses

3.4. Cost of Single-Trailer and Double-Trailer transport

D_{nm} = distance in km from place of origin *n* to destination *m*.

η_{nmtw} = price or cost of carriage from place of origin *n* to destination *m* of product type t in transport type *w* (single-trailer or double-trailer).

φ_w = Transport capacity in m^3 for mode *w*.

ξ_{nmtw} = Cost per m^3 of transport from place of origin *n* to destination *m* of product type *t* in mode of transport *w*

Finally, the demand of each customer is defined as:
d_{mt} = demand of each customer *l* for product type *t*

3.5. Variables of the model

X_{ijtw} = quantity in m^3 of product type *t* delivered from country *i* to port *j* in transport type *w*. This variable is positive continuous.

X_{jktw} = quantity in m^3 of product type *t* delivered from port *j* to the Distribution Center in port *k* in mode of transport *w*. This variable is positive continuous.

X_{jltw} = quantity in m^3 of product type *t* delivered from port *j* to the inland Distribution Center *l* in mode of transport *w*. This variable is positive continuous.

X_{kltw} = quantity in m^3 delivered from Distribution Center at port *k* to the inland Distribution Center *l* from product type *t* in mode of transport *w*.

X_{kmtw} = decision on merchandise of product type *t* delivered from the Distribution Center in port *k* to customer *m* in mode of transport w. This variable is binary (0,1).

X_{lmtw} = decision on merchandise of product type *t* delivered from the inland Distribution Center *l* to customer *m* in means of transport *w*. This variable is binary (0,1).

Y_k = decision on opening or closing the CD located at port *k*

Y_l = decision on opening or closing the CD located at *l*

3.6. Performance measurements

For this problem we propose 3 performance measurements:

1. Z_{mar} = Cost of sea transport.
2. Z_{ter} = Cost of land transport.
3. Z_{alm} = Cost of storage.

Overall we have:

Z = Total cost of operating the supply chain.

4. INFORMATION GATHERING.

- Information was directly provided by the company's ERP system (SAS) about the shipments made from January 2007 to December 2007 and organized in a database with the following fields:
- Order number: used to locate the release details.
- Linked dispatch: this number identifies each dispatch order with the deliveries made. Used for the trackability of the purchase order.
- Date of shipment: date when the shipment was made.
- Name of carrier: Name under which the carrier is registered in the system.
- Place of origin: City of origin (City, State)
- Destination: Name of the city of destination (City, State)
- M^3 transported: quantity in m^3 of material transported.
- Carriage cost: cost incurred by taking the shipment from the place of origin to the required destination.
- Product type: classification of the product shipped to the customer.

4.1. Adaptation of the sample

Using random sampling, 138 data for the m^3 transported during the period from January 2007 to December 2007 were selected in order to assess whether the sample from one year is enough for the study.

The annual mean of the population is 1006 m^3 a day shipped.

Table 1: Test mean-value of sample size

Hypothesized Value	1006
Actual Estimate	1024.07
df	137
Std Dev	760.259

The statistical hypotheses are defined as:

Ho: μ = 1006 m^3 a day
Ha: $\mu \neq$ 1006 m^3 a day

Value p is 0.78 in the test; assuming a significance α = 0.05, as $p>\alpha$ we are within the acceptance region and therefore cannot reject the null hypothesis that establishes that the mean of the sample is equal to the mean of the sample for all 2007. Therefore, the sample from January 2007 to December 2007 is representative. On the other hand, we get the same result with the Wilcoxon signed-rank test for non-parametric data, since the p value in from the test is 0.29 and greater than the 0.05 significance so we can say that there is statistical weight.

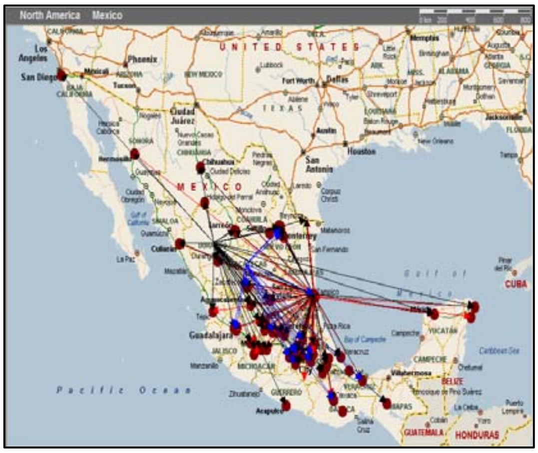

Figure 1: Base case for 2007 demand distribution

The 288 data corresponding to 2007 can be used for the distribution network model.

4.2. Estimating the cost of land transport

The cost of land transport depends on the following factors:

1. Distance travelled.
2. Carriage cost
3. Product type.
4. Capacity of the means of transport.

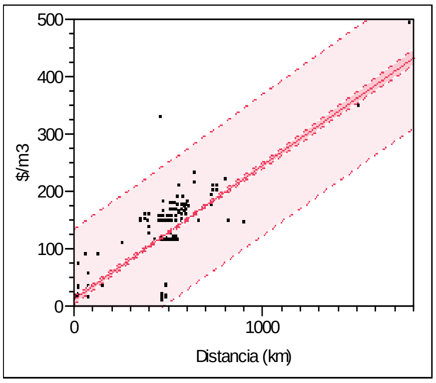

Figure 2: Linear regression for Wood.

Linear Fit for Wood
$/m3 = 13.563404 + 0.2344908*Distance (km)

The regression coefficient is 80.62%, the intercept is $13.56 that is interpreted as the fixed cost for operating one item and the slope is $0.23 which is the variable part of the rate.

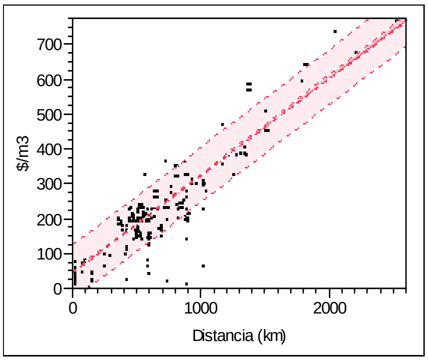

Figure 3: Linear regression for MDF

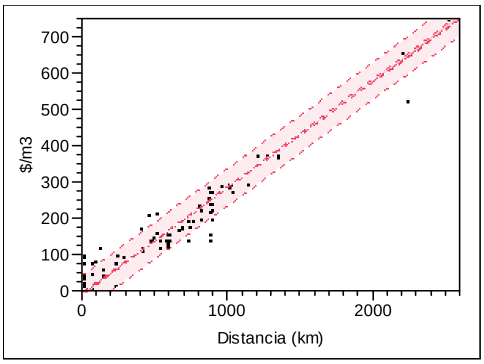

Figure 4: Linear regression for sheeting

Linear Fit MDF
$/m3 = 49.476191 + 0.2795796*Distance (km)

The regression coefficient is 88.89% with a fixed cost of $49.47 per m^3 and a variable cost of $0.28 per km /m^3.

Linear Fit Sheeting
$/m3 = -5.445904 + 0.2934067*Distance (km)

The regression coefficient is 97.46% with a fixed cost of -$5.44 per m^3 and a variable cost of $0.29 per km /m^3.

4.3. Base line

The annual cost of logistics is $245 million pesos of which $161 are for sea transport, $65 for land transport and $33 for storage.

As for the situation at the present time, we know that in 2007 the cost of transporting goods to the customer was $59 million pesos. This gives us an error of 2.39% in the model.

5. EXECUTION OF OPTIMIZATION RUNS.

5.1. Definition of scenarios

Together with the team of project directors, we defined the following scenarios to be assessed:

Scenario 1: Free consolidated.

- Opening and closing of Distribution Centers without any constraint on contracts with 3PL.
- Flows and modes of transport defined by the optimization model.
- Flow conservation constraints between nodes.
- Demand fulfillment constraints.
- The demand is assumed to be constant according to the data collected in the Jan-Dec 2007 period.

Scenario 2: Not closing Tampico.

- Tampico should not be closed because of issues to do with contracts with 3PL.
- Constraint of at least 77% on entry of bulk freight into ports.

- Constraint not higher than 23% on entry of containerized freight into ports.
- Flows and land modes defined by the model.
- Flow conservation constraints between nodes.
- Demand fulfillment constraints.
- Consolidation of bulk freight in a Distribution Center located in the port (there are no direct deliveries from the port to the customers).

This means that 485 containers would be required every month in the ports on the Pacific and, operationally speaking, it is not feasible for the factory in Chile to ship that quantity of containers only for Mexico. Furthermore, in the Mexican ports it is not possible to get the 485 trucks per month that would be required. Therefore, the best option is to be open a Distribution Center in Tampico for bulk freight (77% of what arrives in Mexico), close the one in Altamira, extend the Mexico City Distribution Center, have the containers come in through Manzanillo and Mazatlan to the factory in Durango. The factory in Durango should also operate as a Distribution Center for MDF in the north of the country.

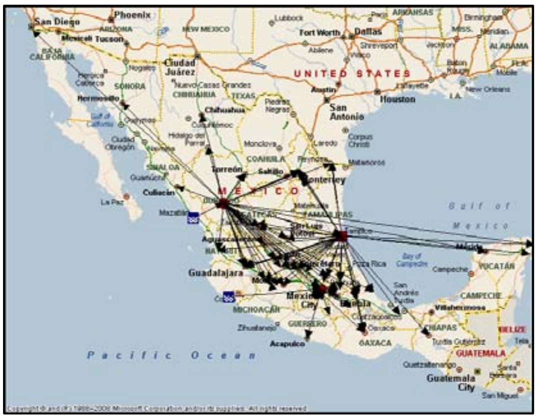

Figure 4: Optimal solution for Wood and MDF. Sheeting is distributed 100% from Durango.

6. SIMULATION MODEL

The simulation model was developed in the following stages:

1. Information gathering
2. Formulation of the model
3. Validation of the model
4. Programming of the model
5. Validation of the programmed model
6. Design and analysis of experiments

A sample of the dispatches to the end customer in one year (Jan-Dec 2007) was taken from the optimization model. The sample considers the three product types: wood, MDF and Sheeting.

Let:

d_{kij} = Quantity of m^3 of product i shipped every day k in locality j

D_i = Quantity of m^3 of product i shipped every year

Therefore:

$$D_i = \sum_{k=1}^{n} \sum_{j=1}^{m} d_{kj} \text{ for each product } i \quad (11)$$

$i = Madera, MDF, Placa$
$j = 1, 2, 3 \ldots, m$
$k = 1, 2, 3 \ldots, n$

If the demand for each locality and product is a random variable. Therefore the annual demand for each product is also a random variable. In order to simplify the simulation model, only 85% of the volume is considered to be stochastic and the remaining 15% considered a constant owing to the fact that the shipments are mostly sporadic.

According to Axëter (1996), a supposition for modeling the demand is to consider that it follows a Poisson distribution with a parameter λ which is the average number of units in m^3 shipped every day. We know that the Poisson distribution and the exponential distribution are intimately linked and that, for the continuous case, it works better if we assume that demand follows an exponential distribution.

Now then, the problem consists of finding a random number generator for an exponential distribution.

Let:

x = the quantity of m^3 to be shipped every day. Then x follows an exponential distribution with parameter λ:

$$f_x = \lambda \, e^{-\lambda x} \quad (12)$$

Let F(x) be the accumulated probability distribution, therefore:

$$F_x = \int_0^x \lambda \, e^{-\lambda x} \, dx = 1 - e^{-\lambda x} \quad (13)$$

Let r be a random number between 0 and 1. Therefore as F_x is between 0 and 1, we can write:

$$F_x = r = 1 - e^{-\lambda x} \quad (14)$$

Finding x we get:

$$x = -\frac{\ln(-r + 1)}{\lambda} \quad (15)$$

This is a random number generator with exponential distribution.

Let:

x_{kij} = quantity of m^3 of product i shipped on day k in locality j

Therefore, annual demand for locality j is:

$$d_j = \sum_{k=1}^{n} x_k = \sum_{k=1}^{n} \frac{\ln(1 - r_k)}{\lambda_j} \quad (16)$$

for each product i

Therefore the annual demand for each product i is:

$$D_i = \sum_{k=1}^{n} \sum_{j=1}^{m} \frac{\ln(1 - r_{kj})}{\lambda_j} \quad (17)$$

for each product i

For each product, the volumes shipped to the localities considered were 74,545, 114,979 and 54,827 for wood, MDF and sheeting respectively.

6.1. Results from the simulation model

On running the analysis of the sample, we found that it is necessary to have a sample of n=256 to obtain a power of 95%. However, we decided to execute the runs of n=300 to get a power of 0.97, as shown in the JMP results.

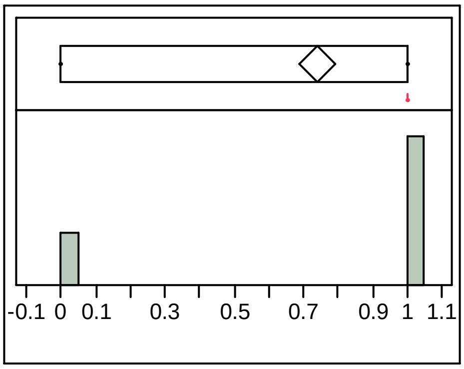

Figure 5: Simulation results for Guadalajara.

Therefore, we are able to say that on 74% of the occasions, under conditions of changing demand and changeable fixed storage costs, Guadalajara turned out to be the best option as a logistics plan. While on 26% of the occasions, Toluca was the best option. Therefore, there is 26% potential risk that if we open the Distribution Center in Guadalajara, we will not get optimum results. However, we are in a position to conclude that the current logistics of this company are wrong since they did not appear within the solutions proposed by the Optimization-Simulation Model we developed.

7. CONCLUSIONS AND FUTURE WORK.

Although it is possible to separate the techniques of mathematical programming and analysis, the optimization and simulation of the supply chain give an optimum solution while the uncertainty in the model's parameters is evaluated. This permits the construction of multi-scenari optimization models that make it possible to evaluate different alternatives and automatically carry out the sensitivity analysis of the network model. Thus, the optimization and simulation model presented throughout this thesis paper permitted:

1. Interaction between the optimization and simulation of the supply chain.
2. Finding an optimum solution.
3. The evaluation of different scenarios through sensitivity analysis.
4. Determining a metrics to measure the optimality of the solution proposed by the optimization model that can be translated into a risk metrics for the decision.
5. Determining a methodology for the development of large-scale optimization and simulation models.

It must be emphasized that owing to the complexity of the mixed-integer programs (MIP), which tend to be considered as NP-complete, it is not possible nowadays for a program to do the post-optimality analysis of the optimization model. However, with the aid of the simulation model it is possible be able to determine the optimality parameters for these models.

From a business point of view, the optimization model for the distribution network allowed us to:

1. Identify strategies for inter-continental distribution through the use of different modes of sea transport.
2. Evaluate of the country's points of entry that would allow greater flexibility in the supply chain and minimize costs.
3. Determine the optimum logistics plan for each product in the distribution network.
4. Define marketing strategies.
5. Take a medium-term strategic decision about the logistics and marketing scheme to be followed in the next few years.

Throughout the development of the model, it was necessary to carry out a meticulous validation of each one of the parameters in order to guarantee, with a permissible estimation error, an optimum solution being rectified by the simulation model. Therefore, we can say that if the estimation of the parameters is good, the optimum solution will be the one that, under conditions of uncertainty, prevails over all the other possible feasible solutions.

This simulation and optimization model was used in the context of a real case scenario where, in order to analyze the efficiency of the distribution network as Preusser (2008) suggests, while the static characteristics of the linear model can be overcome by the motor de optimization, it is possible to use the simulation model to approximate a stochastic and possibly nonlinearity analysis of the problem in order to determine rules of decision that favor the operations of the business at a low cost.

Future research can focus on the integration of the simulation model with simulation languages that permit optimization while the simulation of the stochastic events is generated. In the case of the network models described in the case study, the subject of inventories and planning horizons could be included in order to get production, storage and distribution programs for the different products. An alternative means of solving these types of models, considering stochastic elements, is known as “ stochastic programming”. Nowadays some software packages such as CPLEX and GAMS are capable of solving such formulations by the inclusion of special algorithms.

We will obtain other branches for research as we reach the limits of optimization and simulation. If more complex models are used, new ways of solving them (metaheuristics, heuristics) will have to be developed.

REFERENCES

Ahuja, R.K. and Magnanti, T.L. 1993. *Network Flows.* New Jersey: Prentice Hall.

Bazaraa, S. and Sherali, D.2006. *Nonlinear Programming. Theory and Algoritms* New Jersey*.* Wiley Inter Science.

Chen An, C.A. and Hansjörg, F.R., 2005. *Supply Chain Management on Demand: Strategies, technologies and applications*. Netherlands: Springer Verlang

Dalila, B.M., Fontes and Goncalves, J.F. Heuristic Solutions for General Concave Minimun Cost Network flow Problems.*Wiley Interscience: Periodical Inc. Networks.* Vol. 50. Pag. 67-76.

Eiselt, H.A. and Sandblom. 2000. *Integer Programming and Network Models.* Berlin: Springer-Verlang.

Flores de la Mota, I.F. and Elizondo, M.E. 2006. *Apuntes de simulación*. México, D.F.: UNAM Editorial.

Hillier, F.H., 2000. *Introducción a la investigación de operaciones*. 8th Edition. México, D.F: Mc Graw Hill.

Jack, P.C.; Supply Chain Simulation: a Survey. *International Journal of Simulation and Process Modeling.*

Kok, A.G. and Graves, S.C. 2003. *Handbooks in operation research and management science Vol11: Supply Chain Management.* New York: Elsevier Editor.

Prreusser, M.P. 2008. *Simulation and optimization in Supply Chain Management.* 1st ed. Saardrüken: Dr. Muller editor.

Pistikopoulos and Georgiadis. 2008. *Supply Chain Optimization vol.* I. 1st ed.I Berlin: Wiley-VCH.

Simchi, L. and Bramel, 1997. *The logic of Logistics: Theory, Algorithms and Applications for Logistics Management.* New York: Springer-Verlang.

ACKNOWLEDGMENTS

This paper was written with the support of the Academic program DGAPA-UNAM throughout PAPIIT project number IN100608.