CLUSTERING EFFICIENCY OF SCENARIOS IN THE VSS CAPTURE FOR THE TRANSPORTATION PROBLEM WITH STOCHASTIC DEMAND

José Ernesto Agüero Gutiérrez

Universidad Católica de la Santísima Concepción

Caupolicán 491, Concepción, Chile

jeaguero@ing.ucsc.cl

Manuel Salvador Cepeda Junemann

Universidad Católica de la Santísima Concepción

Caupolicán 491, Concepción, Chile

mcepeda@ucsc.cl

ABSTRACT

The main objective of this paper is to determine the cluster approach of scenario that captures efficiently the value of stochastic solution (VSS) for the transportation problem with stochastic demand. There are two methods to solve this problem, the stochastic approach which gives the expected solution and considers all possible scenarios for it, which is complex to get. On the other hand, the deterministic approach gives a solution based on a scenario and it is easy to obtain. The focus of this proposal is oriented towards achieving an efficient balance between the two approaches resolutions of this problem, which is achieved by considering a number of scenarios that are formed through the clustering technique and are considerably lower than total number of stages and which are obtained close to the optimal solution efficiently.

Keywords: Transportation problem, Stochastic Programming, VSS, Cluster.

Introduction

The transportation problem with stochastic demand aims to determine the amount to send a product from a certain plant to a certain customer in order to satisfy the demand, which is not constant but follows a certain known probability distribution. This kind of problem belongs to an area of Operation Research called stochastic Programming, which focuses on treating mathematical programming problems where some of the parameters are random variables or follow a certain probability distribution.

The case to analyze is a stochastic programming problem in 2 stages, this means that in the first instance the customer demand is not known but this parameter can be known after the occurrence of a given scenario and therefore can satisfy the demand, if necessary, with the decisions of the second stage, so as to always satisfy the customer demand.

The current methodology for solving such problems is to determine the best strategy in order to minimize the expected costs through DELP (deterministic equivalent linear problem) which will give first-stage decisions, namely, the decisions performed before knowing the demands and the decisions of the second stage, which correspond to those made after knowing the demand.

This is the basic idea of the two-step process. In the first stage, before the realization of the random variables become known, are chosen decision variables of the first stage to optimize the expected value in the objective function, which is in turn, the optimum value of the second stage the optimization problem.

$$\begin{aligned} &\boldsymbol{Min_x c^T x} + \mathrm{E}[\boldsymbol{Q}(\boldsymbol{x}, \xi(\omega))] \\ &\boldsymbol{S.a.} \quad \boldsymbol{Ax} = \boldsymbol{b} \quad \boldsymbol{x} \geq 0 \end{aligned}$$

Where $\boldsymbol{Q}(\boldsymbol{x}, \xi)$ is the optimum value of the second stage

$$\begin{aligned} &\boldsymbol{Min_y q^T y} \\ &\boldsymbol{S.a.} \quad \mathrm{T}\boldsymbol{x} + \boldsymbol{Wy} = \boldsymbol{h} \quad \boldsymbol{x}, \boldsymbol{y} \geq 0 \end{aligned}$$

X and Y are vectors of the first and second stage decision variables respectively. The problem of second stage depends on the data where some or all of its elements can be random.

The transportation problem that will be analyze have demands that follows certain discrete probability with a finite number K scenarios, so the adaptation of DELP for the case of two stages is:

$$
\begin{aligned}
& \underset{x, y_1 \ldots y_k}{Min}\; c^T x + \sum_{k=1}^{K} p_k q_k^T y_k \\
& S.a. \quad Ax = b \\
& \qquad\quad T_k x + W_k y_k = h_k, \qquad k=1,\ldots\ldots,K \\
& \qquad\quad x \geq 0, y_k \geq 0, \qquad k=1,\ldots\ldots,K
\end{aligned}
$$

An alternative way of approaching the transportation problem with stochastic demand is to a simpler approach, which is to consider the expected value of the problem where the decision maker supersedes all random variables by their expected values and solver a deterministic problem

$$
\begin{aligned}
& \underset{x,y}{Min}\, c^T x + q^T y \\
& S.a. \quad Ax = b \\
& \qquad\quad Tx + Wy = h, \\
& \qquad\quad x \geq 0, y \geq 0,
\end{aligned}
$$

Where it obtain the expected values of the variables of first-stage and EEV, is the expected cost when the obtained solution is fixed occupying the expected values of the variables in the first stage to solve the problem of DELP.

The value of the stochastic solution is defined as

$$VSS = EEV - DELP$$

And measures the expected increase of the value to solve the stochastic version of a model instead of only the deterministic.

So far is only been discussed on the current techniques of problem resolution stochastic programming and in case the transportation problem with stochastic demand of 2 stages, which has a stochastic approach PRSP and deterministic approach CLP. Comparing the two approaches can be seen that the first solution delivers better quality than the second, as the second option only considers an average scenario for each of the stochastic demands. But on the other hand, the deterministic form of resolution has the great advantage that it is easy to resolve from the computational point of view, it reduces the number of variables and constraints of the problem.

Therefore, Will there be any way to find a middle point between both approaches in order to obtain a good solution considering only a low number of scenarios for the problem under consideration?

The hypothesis is that you can save a considerable percentage of VSS occupying only a small number of scenarios with respect to all possible scenarios. This hypothesis is expressed graphically through a concave curve where the value obtained is in direct relation to the number of scenarios employed. The chart below shows

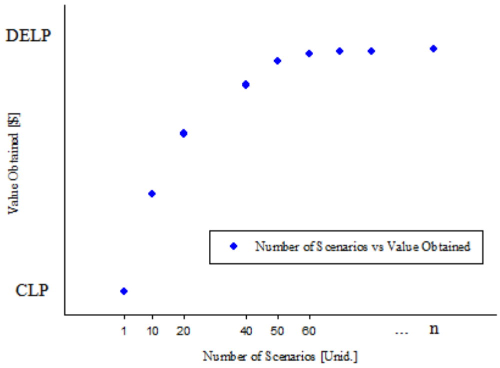

Fig 1 Value Obtained graph v/s Number of Scenarios

The way to group these scenarios in order to maximize the percentage rescued from the VSS is the technique of Cluster Analysis. This technique aims to bring together individuals and objects into clusters so as to maximize the homogeneity of objects within clusters while at the same time maximizing the heterogeneity between the aggregates.

After analysis of the purpose and methodology of cluster analysis, concludes that this technique of grouping of observations in the case of the transportation problem with stochastic demands are scenarios, has great utility for the hypothesis put forward, as it seeks to bring together scenarios that are very similar to each other, because these settings provide the same information, but on the other side wants to separate scenarios that are different between them, so as to obtain more information by maintaining the diversity of scenarios.

To perform the cluster technique is necessary to answer the following questions:

- How can measure the similarity between the observations?
- How are the clusters?
- How many groups are formed?

Is for this reason that there are more than one type of cluster, since each type of resolve these unknowns are different.

There are two approaches to clustering, the hierarchical and not hierarchical, the first is employed when it is not known in advance the number of clusters to be used, and the second is for the case there is a priori fixed number of clusters to be filled.

Due to the fact that in this work you need to set the number of clusters to use, clustering methods that are occupied will not hierarchical type, which are:

- K-Means
- K-Medoids

Both algorithms belong to Center-based clustering algorithms, which are very efficient for large databases and clustering for high-dimensional database.

Both clustering methods take as distance measure the Euclidean norm, in order to quantify the proximity between each of the scenarios.

Finally it aims to determine the cluster approach of scenario that capture efficiently the VSS to the transportation problem with stochastic demand.

Methodology

In the first instance to solve the transportation problem with stochastic demand, if possible, to get the value of the DELP. Thus finding the best value solution to the problem. If there is no way to get the expected value of the solution, using the stochastic approach, because the problem analyzed is large, a projection will be made according to the trend curve showing the solution, this is built by the results obtained solve the problem for different numbers of scenarios.

Then, it will solve the transportation problem with stochastic demand, using the expected values for the demands, so to solve the problem from the deterministic approach, and thus find the expected value of the solution only occupies a scenario (CLP)

Below is calculated the VSS, which is the difference between the value of the solution obtained using the stochastic approach minus the value of the solution obtained using the deterministic approach.

After having solved the problem, and have calculated the VSS, will proceed to perform an analysis of the different types of groupings of scenarios under the cluster analysis approach, which identify the data grouping techniques that give feasible solutions and observe the behavior of each of the techniques for the transportation problem with stochastic demand.

Is important to note that cluster analysis will be made by fixing the number of scenarios, this means that the obtained solutions depend of the amount of scenery to be grouped,

Finally, establishing what type of grouping to capture the largest share of VSS with a low number of scenarios with respect to the total number of possible scenarios that have the problem.

The analysis will determine, which is the number of scenarios efficient to use, under the judgment that the cost or effort to find that solution is less than the percentage rescued from VSS. The efficiency measure to measure is given by:

$$\boldsymbol{E}(\boldsymbol{k}) = \frac{\%\boldsymbol{VSS}}{\boldsymbol{F}(\boldsymbol{k})}$$

Where F (k) is a function that depends on the number of scenarios and measures the effort to solve the problem, given a certain number of scenarios. Notably, both ends of the efficiency curve has a value 0.

Similarly , determines whether a cluster analysis technique, is better in some particular circumstance to another technique, or whether to have a clustering technique for any number of scenarios used..

The following is a case study solves particular for the transportation problem with stochastic demand, with the aim of verifying the hypothesis posed for the curve of VSS, where for different numbers of scenarios, this curve is concave behaves in such way to rescue a significant percentage VSS only solve the problem with using a smaller number of scenarios

Study case

Format of instance type for the transportation problem with stochastic demand is:

Some plants i and j industry has customers located in various locations. The demand for some customers is known, however the other will be known after sending the goods. Each unit of output is produced at a cost of $ C and sold to a P value to any customer. If a customer receives less than that requested, the company must decide between 3 options to fill the demand gap.

- Send from any of the plants (with excess) at a cost of shipping is two times faster than the cost of regular shipping.
- Send from a location where it had been addressed more units than the client ultimately required.
- Buy local products on the market to satisfy demand shortfall.

If the plants had to send more units than required by the customer, they can be returned to the originating facility at a cost of normal transport or shipped to another customer with a demand shortfall.

The units that remain on the plants and that are returned to them are valued at cost price.

Each plant has a limited supply of products to make, each customer's demand and cost of shipping from plants to customers, Transport cost between locations where there are customers who can generate surplus or have shortfall and purchase costs of products in the local market.

To solve the problem is necessary to model, whereby in the first instance variables defined problem, these are:

$\boldsymbol{x}_{ij}$: Quantity to make a product in plant **i** sent to the client j. $\forall \boldsymbol{i}:1..\boldsymbol{I}$; $\forall \boldsymbol{j}:1..\boldsymbol{J}$

$\boldsymbol{s}_{ijt}$: Quantity to send of a product of the city **j** to the plant **i** given the t scenario.
$\forall \boldsymbol{i}:1..\boldsymbol{I}$; $\forall \boldsymbol{j}:\boldsymbol{k}...\boldsymbol{J}$ $\forall \boldsymbol{t}:1..\boldsymbol{T}$

$\boldsymbol{y}_{ijt}$: Quantity to send of a product of the plant i that has excess, in quick sale for the customer j in the t scenario. $\forall \boldsymbol{i}:1..\boldsymbol{I}$; $\forall \boldsymbol{j}:\boldsymbol{k}...\boldsymbol{J}$ $\forall \boldsymbol{t}:1..\boldsymbol{T}$

$\boldsymbol{z}_{hjt}$: Quantity to send of a product from the locality t that has an exceed to the locality j that has shortfall of products. $\forall \boldsymbol{h}:\boldsymbol{k}..\boldsymbol{J}$; $\forall \boldsymbol{j}:\boldsymbol{k}..\boldsymbol{J}$; $h \neq j$; $\forall \boldsymbol{t}:1..\boldsymbol{T}$

$\boldsymbol{M}_{jt}$: Quantity to purchase of a product in the intern market of the locality **j** that has a shortfall given the **t** scenario. $\forall \boldsymbol{j}:\boldsymbol{k}...\boldsymbol{J}$; $\forall \boldsymbol{t}:1..\boldsymbol{T}$

$\boldsymbol{E}_{it}$: Quantity of a product that in the end stays in the inventory at the plant **i** given the **t** scenario
$\forall \boldsymbol{i}:1..\boldsymbol{I}$; $\forall \boldsymbol{t}:1..\boldsymbol{T}$

Besides, set the following parameters:

$\boldsymbol{U}_i$: Utility for product sold was manufactured in the plant **i**. $\forall \boldsymbol{i}:1...\boldsymbol{I}$

$\boldsymbol{P}_j$: Sales Price for of a product to the customer **j**. $\forall \boldsymbol{j}:1...\boldsymbol{J}$

$\boldsymbol{c}_{ij}$: Cost of shipping a unit of a product from the plant **i** to the customer **j**. $\forall \boldsymbol{i}:1...\boldsymbol{I}$; $\forall \boldsymbol{j}:1...\boldsymbol{J}$

$\boldsymbol{d}_{hj}$: Cost of transporting a unit of a product from the locality **h** to the locality **j**.
$\forall \boldsymbol{h}:\boldsymbol{k}...\text{T}$; $\forall \boldsymbol{j}:\boldsymbol{k}...\text{T}$

$\boldsymbol{CM}_j$: Cost of buying a unit of product in the local market of the locality **j**. $\forall \boldsymbol{j}:\text{k}...\boldsymbol{T}$

$\boldsymbol{D}_{jt}$: Demand for a product of the customer **j** given **t** scenario. $\forall$j:k...T ; $\forall$t:1..T

$\boldsymbol{O}_i$: Supply of a product of the plant i. $\forall \boldsymbol{i}:1...\boldsymbol{I}$

Then, raises the stochastic linear programming model, which is:

Objective function:

$$\boldsymbol{Max}\ \mathrm{Z}: \boldsymbol{U}_i * (\sum_{i=1}^{I}\sum_{j=1}^{T} \boldsymbol{x}_{ij}) - \sum_{i=1}^{I}\sum_{j=1}^{T} \boldsymbol{c}_{ij}\boldsymbol{x}_{ij} + \ldots$$

$$+ \sum_{t=1}^{T} \boldsymbol{p}_t * (-\sum_{i=1}^{I}\sum_{j=k}^{J} \boldsymbol{c}_{ij}\boldsymbol{s}_{ijt} - \boldsymbol{U}_i \sum_{i=1}^{I}\sum_{j=k}^{J} \boldsymbol{s}_{ijt} + \boldsymbol{U}_i \sum_{i=1}^{I}\sum_{j=k}^{J} \boldsymbol{y}_{ijt} - 2 * \sum_{i=1}^{I}\sum_{j=k}^{J} \boldsymbol{c}_{ij}\boldsymbol{y}_{ijt} - \sum_{h=k}^{J}\sum_{j=k}^{J} \boldsymbol{d}_{hj}\boldsymbol{Z}_{hjt} + ..$$

$$+ \boldsymbol{P}_j \sum_{j=k}^{J} \boldsymbol{M}_{jt} - \sum_{j=k}^{J} \boldsymbol{CM}_j \boldsymbol{M}_{jt})$$

Subject to:

Demand constraints:

$$\sum_{i=1}^{T} \boldsymbol{x}_{ij} = \boldsymbol{D}_j \quad \forall \boldsymbol{j}: 1 \ldots \boldsymbol{k} - 1$$

$$\sum_{i=1}^{I} \boldsymbol{x}_{ij} - \sum_{i=1}^{I} \boldsymbol{s}_{ijt} - \sum_{h=k, k\neq j}^{J} \boldsymbol{z}_{hjt} + \sum_{i=1}^{I} \boldsymbol{y}_{ijt} + \sum_{h=k, k\neq j}^{J} \boldsymbol{z}_{jht} + \boldsymbol{M}_{jt} = \boldsymbol{D}_{jt} \quad \forall \mathrm{t}:1\ldots\mathrm{T}\ \forall \boldsymbol{j}:\mathrm{k}\ldots\mathrm{J}$$

Supply constraints:

$$\sum_{j=1}^{J} \boldsymbol{x}_{ij} + \sum_{j=k}^{J} \boldsymbol{y}_{ijt} - \sum_{j=k}^{J} \boldsymbol{s}_{ijt} + \boldsymbol{E}_{it} = \boldsymbol{O}_i \quad \forall \mathrm{t}:1\ldots\mathrm{T}\ ;\ \forall \boldsymbol{i}:1\ldots\mathrm{I}$$

Balance constraints:

$$\sum_{j=1}^{J} \boldsymbol{x}_{ij} + \sum_{j=k}^{J} \boldsymbol{y}_{ijt} \leq \boldsymbol{O}_{it} \quad \forall \mathrm{t}:1\ldots\mathrm{T}\ ;\ \forall \boldsymbol{i}:1\ldots\mathrm{I}$$

$$-\sum_{i=1}^{I} \boldsymbol{x}_{ij} + \sum_{i=1}^{I} \boldsymbol{s}_{ijt} + \sum_{\substack{h=k \\ j\neq h}}^{J} \boldsymbol{z}_{jht} \leq 0 \quad \forall \mathrm{t}:1\ldots\mathrm{T}\ ;\ \forall \boldsymbol{j}:\mathrm{k}\ldots\mathrm{J}$$

No negativity constrains

$$\boldsymbol{x}_{ij} \geq 0 \quad \forall \boldsymbol{i}: 1\ldots\boldsymbol{T}\ ;\ \forall \boldsymbol{j}: 1\ldots\boldsymbol{J}$$
$$\boldsymbol{s}_{ijt} \geq 0\ \forall \boldsymbol{i}: 1\ldots\boldsymbol{I}\ ;\ \forall \boldsymbol{j}: \boldsymbol{k}\ldots\boldsymbol{J}\ \forall \boldsymbol{t}: 1\ldots\boldsymbol{T}$$
$$\boldsymbol{y}_{ijt} \geq 0\ \forall \boldsymbol{i}: 1\ldots\boldsymbol{I}\ ;\ \forall \boldsymbol{j}: \boldsymbol{k}\ldots\boldsymbol{J}\ ;\ \forall \boldsymbol{t}: 1\ldots\boldsymbol{T}$$
$$\boldsymbol{z}_{hjt} \geq 0\ \forall \boldsymbol{h}: \boldsymbol{k}\ldots\boldsymbol{J}\ ;\ \forall \boldsymbol{j}: \boldsymbol{k}\ldots\boldsymbol{J}\ ;\ \mathrm{h} \neq \mathrm{j}\ ;\ \forall \boldsymbol{t}: 1\ldots\boldsymbol{T}$$
$$\boldsymbol{M}_{jt} \geq 0\ \forall \boldsymbol{j}: \boldsymbol{k}\ldots\boldsymbol{J}\ ;\ \forall \boldsymbol{t}: 1\ldots\boldsymbol{T}$$

Results and Analysis

The results obtained for two instances of the problem are presented below. Scenarios sizes are 12 and 3125 respectively.

First problems will be solved from the stochastic point of view where the results show the decisions to be taken both in the first stage and second stage for each of the possible scenarios in both cases. Also shows the quantity of product remaining in inventory (Eit) at the end of the second stage.

The solution of the problem for both instances indicates that the value of the objective function that maximizes the expected income is $ 60,398,810 for the case of 12 scenarios and $ 148,391,000 for the case of 3215 scenarios.

On the other hand, to solve the problems from the deterministic approach, occupying only a possible scenario where the expected values are used for demands, determines that the value of the solution which aims to maximize the income is $ 59,850,350 for the case of 12 scenarios and $ 148,391,000 for the case of 3215 scenarios.

The stochastic value of the solution is:

Instance	Number of Scenarios	Value Obtained [$] DELP	Value Obtained [$] CLP	VSS [$]
1	12	60,398,810	59,850,350	548,460
2	3125	148,391,000	145,461,000	2,930,000

Table 1 Stochastic value of the solution for the instances.

This shows that the values of the solutions found with the stochastic approach to the problem of transport are better than the values found by the deterministic approach, which considers the expected value for each of the demands.

Below are displayed and analyzed the results of clustering methods for both instances.

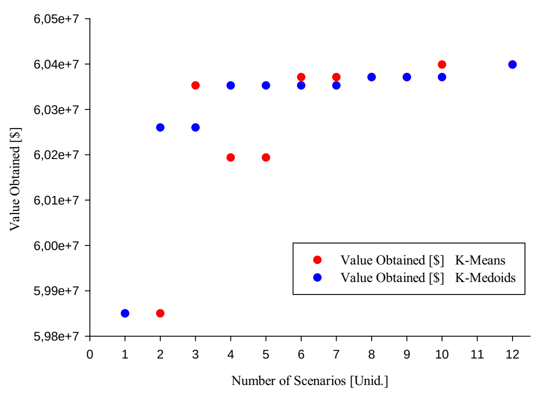

Fig 2. Value Obtained Graph Instance 12 Esc.

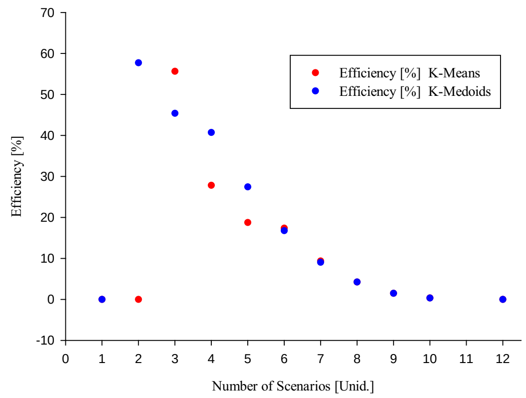

Fig 3. Efficiency Graph Instance 12 Esc

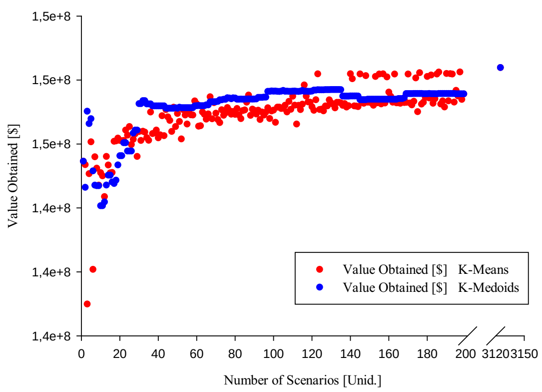

Fig 4. Value Obtained Graph Instance 3125 Esc.

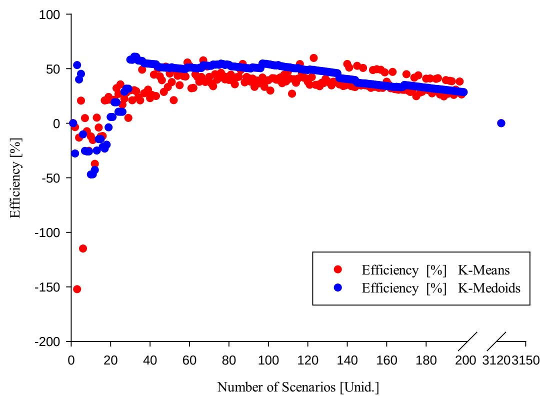

Fig 5. Efficiency Graph Instance 3125 Esc

The graphs show the behavior when using cluster techniques in relation to different amounts of scenarios for the transportation problem with stochastic demand.

By looking at Figure 2 which shows the results obtained for the instance of 12 scenarios, highlights that the curves formed by both algorithms tend to curve shape on the hypothesis posed, which is concave and where the value obtained is in relation to the number of stages employed on each of the resolutions. Similarly noteworthy that in both figure shows how the curve tends to level off in the section that goes from 3 to 5 stages, due to the case studied was a low number of possible scenarios. This entails that for these scenarios amounts of the clusters are not sufficiently heterogeneous to provide more information to solve the problem.

Figure 3 shows the behavior of the efficiency curve in case of the 12 scenarios. It is observed that the maximum efficiency using the algorithm of K-Means and K-Medoids is given to an amount of three two stages respectively. Where the pick of the efficiency ratio is 56% and 58% for K-Means and K-Medoids respectively.

Figure 4 shows, like Figure 1, the results obtained by both algorithms of cluster the instance of 3125 scenarios. Show that the respective curves tend to the shape on the hypothesis put forward. Note that the behavior is irregular along certain sections.

Where for the results obtained with K-Medoids shows that most of the results obtained by grouping to fewer than 20 clusters are worse than those found by the deterministic method. From a group of 30 or more clusters, solutions are obtained which tend to have a good quality, this means that recover a large amount of% of the value of stochastic solution.

Furthermore, the results delivered by K-Means show that there are a small number of groups of settings that provide a bad solution. On the other hand it is appreciated that since an amount of 50 or more scenarios used to solve this problem of transport values are obtained with the same quality, this means that rescue approximately the same % of the value of the stochastic solution. With regard to these solutions, it is observed that from 150 scenarios used there are solutions that tend to attract a larger percentage of the value of stochastic solution, this is due to the existence of certain groups of scenarios that allow the solution to be better than those found in the same range of number of scenarios.

Efficiency curves of both methods, shown in Figure 4, highlights that the pick efficiency is formed by grouping a number of stages equal to 1.8% in the case of K-Means and 1% if K-Medoids of all scenarios. In addition both methods achieve a similar efficiency value that rescue approximately the same stochastic% solution, which is an about 64%.

In conclusion, it appears that the use of a grouping method of scenarios to solve the problem of transport stochastic demand, where this method of grouping is clustering, either with K-Means algorithm or K-Medoids, can efficiently find a middle point between the solution found with the stochastic approach and found through the deterministic approach.

Finally, the results obtained in this case of study and the analyzes consequently, serve as a guide to further develop the hypothesis at the beginning and thus fulfill the purpose which is aimed at determining the cluster approach of scenario that capture efficiently the VSS to the transportation problem with stochastic demand.

Note that the results support the hypothesis.

Thereafter it is intended to work with different study cases for the transportation problem with stochastic demand, having a considerable number of possible scenarios, to be able to validate the hypothesis and draw conclusions in the principal objective.

Bibliography

Birge, J. (1982), The Value of the stochastic solution in stochastic linear programs with fixed recourse, Mathematical Programming, 24, 314-325.

Floudas, C. e Pardalos, P. Encyclopedia of Optimization, Springer, 2009.

Maggioni, F. e Wallace, S.(2010), Analyzing the quality of the expected value solution in stochastic programming, Annals of Operations Research, 2010, 1-18.

Gan, G., Ma, C. e Wu, J. Data Clustering Theory, Algorithms and Applications, Board, 2007.
Ruszczyński, A e Shapiro, A.(2003) Handbooks in OR & MS Vol. 10,Elsevier Science B.V,2003

Szwarc, W. (1964), The Transportation Problem With Stochastic demand, Management Science,11,1-19.