

Georgia Southern University

Digital Commons@Georgia Southern

---

11th IMHRC Proceedings (Milwaukee,  
Wisconsin. USA – 2010)

Progress in Material Handling Research

---

9-1-2010

## Determining Warehouse Storage Location Assignments Using Clustering Analysis

Carl Egas  
*Ohio University*

Dale Masel  
*Ohio University*

Follow this and additional works at: [https://digitalcommons.georgiasouthern.edu/pmhr\\_2010](https://digitalcommons.georgiasouthern.edu/pmhr_2010)



Part of the [Industrial Engineering Commons](#), [Operational Research Commons](#), and the [Operations and Supply Chain Management Commons](#)

---

### Recommended Citation

Egas, Carl and Masel, Dale, "Determining Warehouse Storage Location Assignments Using Clustering Analysis" (2010). *11th IMHRC Proceedings (Milwaukee, Wisconsin. USA – 2010)*. 19.  
[https://digitalcommons.georgiasouthern.edu/pmhr\\_2010/19](https://digitalcommons.georgiasouthern.edu/pmhr_2010/19)

This research paper is brought to you for free and open access by the Progress in Material Handling Research at Digital Commons@Georgia Southern. It has been accepted for inclusion in 11th IMHRC Proceedings (Milwaukee, Wisconsin. USA – 2010) by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact [digitalcommons@georgiasouthern.edu](mailto:digitalcommons@georgiasouthern.edu).

# DETERMINING WAREHOUSE STORAGE LOCATION ASSIGNMENTS USING CLUSTERING ANALYSIS

Carlos Egas  
Dale Masel  
Department of Industrial and Systems Engineering  
Ohio University  
270 Stocker Center  
Athens, Ohio 45701

## Abstract

The methodology used to assign products to a storage location in a warehouse can have a significant impact on the amount of time required to retrieve all of the items needed to fill an order. This paper describes a methodology that uses a clustering approach to determine storage assignments, where the metric of the strength of the relationship between two stock-keeping units (SKUs) is the number of times that the SKUs appear in the same order. Clustering is performed to maximize the frequency with which SKUs in the same cluster are ordered together. In testing, the clustering assignments were compared to a demand-based assignment strategy and showed a reduction of 20-30% in the number of aisles visited to retrieve orders.

## 1 INTRODUCTION

Labor for order picking is one of the largest costs in a distribution center, and travel time is generally considered to be the largest components of this labor [1]. A variety of strategies have been developed to reduce the amount of travel required for retrieving the items needed for an order. These strategies include the use of automated equipment (such as carousels) that bring items to the picker; routing methods to minimize the length of a picker's path; and wave picking, to retrieve multiple orders on a single trip.

Another strategy for reducing the distance traveled by pickers involves how products are assigned to storage locations in a facility. Approaches to storage assignment that have been developed include assigning stock-keeping units (SKUs) in decreasing order of demand (so that the most popular items are closest to the input/output point) or placing the SKUs in classes based on demand and then putting the class with the most-frequently-ordered items nearest to the I/O point.

Another approach for assigning SKUs to storage locations is to group together items that frequently appear in the same order. This reduces travel distance by allowing pickers to retrieve most or all of the items for an order from the same aisle or region of the storage area, without having to travel through the entire storage area to retrieve items. Clustering can be used to create groups of SKUs that are ordered together.

The storage assignments that are determined using clusters would be most appropriate for the forward storage area of a distribution center, where items are picked for filling orders. Typically, only a fraction of the total inventory is stored in the forward area, so that the area has a small footprint, which reduces the travel distance for pickers. The majority of the inventory is stored in a separate reserve storage area, which holds items that are used to replenish the forward area as necessary.

Rosenwein [2] described a method for clustering using  $m$  vectors each with  $Q$  entries, where  $m$  is the number of SKUs being evaluated and  $Q$  is the number of orders in the database. The  $q$ th element in vector  $v^i$  is 1 if SKU  $i$  appears in order  $q$  and 0 if it does not. The distance between two SKUs  $i$  and  $j$ ,  $w_{ij}$ , is calculated using these vectors, as shown in Equation 1:

$$d_{ij} = \sum_{k=1}^p |v_k^i - v_k^j| \quad (1)$$

so that SKUs that frequently both appear in the same order ( $v^i - v^j = 1-1 = 0$ ) or both do not appear in the same order ( $v^i - v^j = 0-0 = 0$ ) have low distance values. Therefore, the objective is to assign each SKU to one of  $p$  clusters to minimize the total distance across all clusters.

Other distribution center decisions that clustering has been used for include setting up storage zones to balance workload across zones [3] and determining how to batch orders for efficient picking. [4, 5]

## 2 METHODOLOGY

In this paper, the metric that is used to represent the distance between SKUs  $i$  and  $j$  is the number of orders that contain both of the SKUs. Therefore, the objective in this case is to maximize the cumulative distance of the SKUs in the cluster. By grouping SKUs that are ordered together into the same cluster, these SKUs can be located close to each other in the warehouse to reduce the travel distance incurred during retrieval.

The clustering methodology is an iterative procedure, which begins by randomly assigning each SKU to a cluster. Then, in each iteration, a centroid is determined for each cluster, where the centroid represents the SKU that has the most orders in common with the other SKUs currently in the cluster. Then, each of the remaining SKUs is assigned to the centroid with which it is ordered most frequently, which results in modified clusters. This iterative process is repeated until a steady state is reached.

Once the clusters have been formed, each cluster is assigned to a contiguous space in the storage area in order to keep SKUs that are frequently ordered together in the same picking area. No limit was placed on the number of SKUs in each cluster, so the number of SKUs in a cluster is not necessarily the same as the number of SKUs that can be stored in an aisle. Therefore, small clusters are combined to get as many groups as possible that are equal to the capacity of an aisle; large clusters are split among multiple aisles.

### 2.1 Step 1: Data Collection

The procedure begins with historical data that lists the SKUs that are requested in each order. The quantity of each SKU is not considered because the travel distance does not depend on how many units of each SKU are picked. A picker has to travel the same distance to reach a storage location, regardless of whether they are picking 1 unit or 100 units once they reach that location.

### 2.2 Step 2: Data Preprocessing

Before the clustering starts, preprocessing of the data is done. The SKUs in the data set that have the highest and the lowest demand are removed from consideration. The highest demand SKUs are removed because they have orders in common with so many of the SKUs that they tend to attract almost all of the other SKUs, concentrating most of them in just a small number of clusters so that there is a significant imbalance in the size of the final clusters. These SKUs could be placed separately after the clustered SKUs have been assigned. For example, they could be assigned to the locations closest to the input/output point, or in locations near the end of multiple aisles, so that they are located close to wherever the other SKUs in an order are located.

The lowest demand items are typically not stored in the forward picking area since this would require a larger storage space. Low-demand SKUs are often only stored in the reserve area, which means that they require more time for retrieval than if they were in the forward area. However, since these SKUs are ordered infrequently, the added time is not incurred very often and this is preferable that making the forward area large enough to hold all of these SKUs.

The order data is used to create a matrix that describes the relationship between each pair of SKUs, a portion of which is shown in Table 1. The values in the matrix represent the number of orders that contain both SKUs—that is,

the value  $w_{ij}$  represents the number of times that SKUs  $i$  and  $j$  appeared in the same order. In the matrix, values for  $w_{ij}$  are included only where  $i < j$ , since  $w_{ij} = w_{ji}$ , orders would be double counted.

**Table 1: Frequency of SKU pairs appearing in the same order**

	1001	1002	1003	1004	1005
1001	0	355	42	0	250
1002	0	0	498	17	0
1003	0	0	0	344	825
1004	0	0	0	0	55
1005	0	0	0	0	0

Once the matrix has been populated, the values are normalized by dividing by the largest number in the matrix, to produce values that are between 0 and 1 for each entry  $w_{ij}$ .

### 2.3 Step 3: Assign SKUs to Clusters

To begin the clustering procedure, SKUs are assigned to clusters. The methodology was tested with random assignments and by assigning sequentially to clusters in order of decreasing demand, but the method of initial assignment did not have a significant impact on the effectiveness of the method.

### 2.4 Step 4: Determine Centroids

The centroid of each cluster is the SKU in the cluster that has in the most orders with other SKUs in the cluster. That is, the centroid of cluster  $K$  is the SKU  $i \in K$  that has the maximum value of  $W_i$ , where

$$W_i = \sum_{j \in K} w_{ij} \quad (2)$$

This is done so that the centroid is a SKU that is frequently ordered with other SKUs, since this represents a SKU that's most important to be in the same location as other SKUs. This doesn't mean that all of the SKUs in the cluster have an important relationship with the centroid—in the next step, the SKUs may move to another cluster. But by selecting the SKU that has the strongest relationship with the other SKUs in the cluster, it has the greatest likelihood of retaining the SKUs that are already in the cluster as well as attracting SKUs from other clusters that have a weak relationship with their centroid.

### 2.5 Step 5: Reassign Non-Centroid SKUs

Once a centroid has been identified for each of the  $k$  clusters, the SKUs that weren't chosen as centroids are considered to see which of the centroids they should be associated with. This is done by assigning each SKU  $i$  to the centroid  $j$  with which it has the largest distance. That is, assign SKU  $i$  to the cluster with centroid  $j$  such that the value of  $w_{ij}$  is maximum.

Doing this places each SKU in a cluster with at least one other SKU (the centroid) with which it appears in the most orders. This attempts to achieve the goal of creating clusters that include SKUs that frequently appear together in orders.

### 2.6 Step 6: Repeat Steps 4 and 5

Steps 4 and 5 are repeated until no changes to the cluster assignments occur, or a threshold number of iterations is reached. Each time the clusters are reconfigured (Step 5), the centroids must be recalculated (Step 4) to identify the SKU that has the most orders in common with the rest of the cluster. If any of the centroids change as a result of these calculations, then some of the other SKUs may change clusters to join this new centroid. Once the centroids stop changing, however, the SKUs will not change clusters and the iterations will be complete.

### **2.7 Step 7: Assign Clusters to Aisles**

The purpose of creating the clusters is so that all of the SKUs in a cluster can be assigned to a contiguous region of the forward storage area. However, the size of each cluster is not controlled, so clusters may not contain enough items to fill an aisle or they may contain too many items to fit in an aisle. In the case of a small cluster, it should be combined with another small cluster so that the size of the combined clusters is approximately equal to the capacity of the aisle.

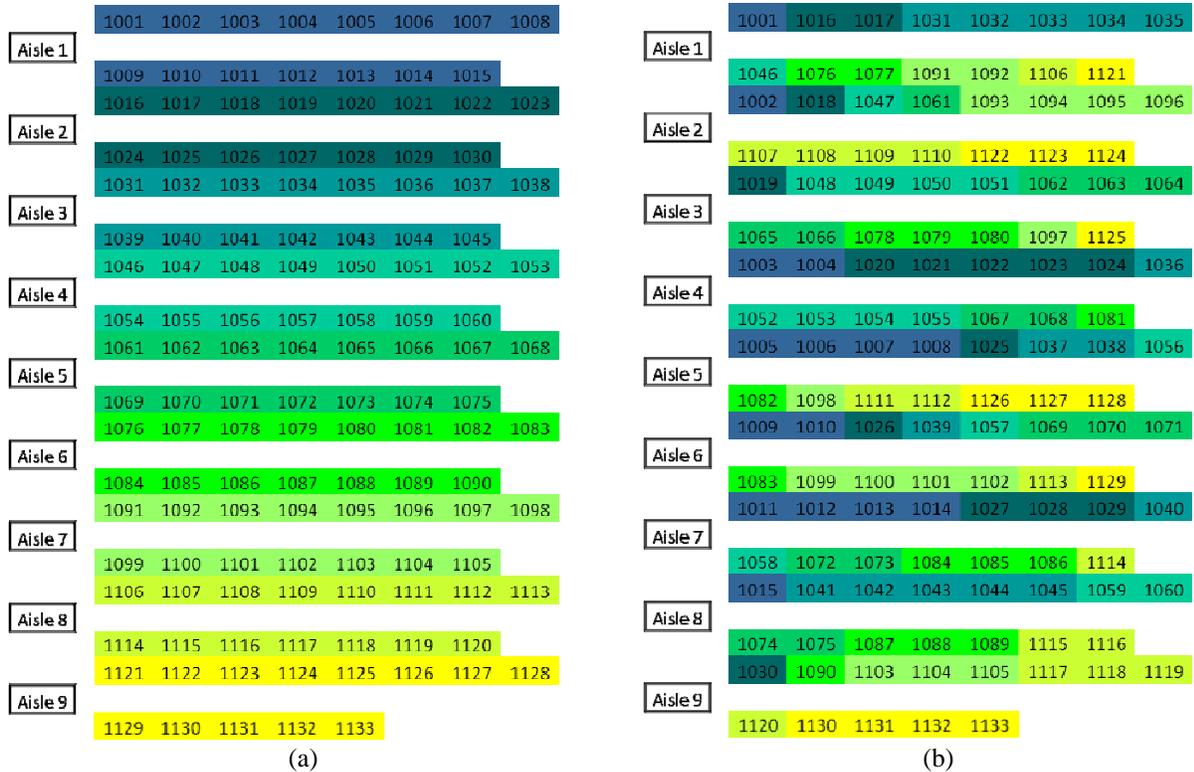
A large cluster should be split into two new clusters, where the size of one of the new clusters is equal to the capacity of an aisle, and the remainder is treated as a small cluster and combined with another small cluster. No study of the optimal method for splitting a cluster has been made, but determining how to do this is a potential area for further study.

## **3 RESULTS**

The effectiveness of this methodology was evaluated by applying the method to a large dataset of historical customer orders. After pre-processing of the order data, there were 133 SKUs that were being evaluated. The storage system for the items was assumed to consist of 9 aisles, each with a capacity of 15 SKUs (for a total capacity of 135 SKUs, so there will be some empty locations). To measure picking distance, the number of aisles visited to retrieve the items for an order was used as the metric. This assumes that a picker will travel completely down an aisle once the aisle has been entered.

As a benchmark for comparison, the SKUs were also assigned to storage locations on the basis of demand, with the most popular SKUs in the first aisle, the next most popular in the second aisle, etc. This approach is used in making storage assignments with the hope that pickers will not have to travel beyond the first few aisles to retrieve the items for an order. If the clustering assignments are effective, pickers will need to travel down fewer aisles to visit the locations of all of the needed items, compared to a demand-based assignment.

Different configurations of the methodology were tested to observe their impact on the quality of the resulting clusters. Parameters that were varied included the number of clusters and the number of SKUs being considered. An example of the layout is shown in Figure 1.



**Figure 1: Assignment of SKUs to aisles using (a) demand-based method; (b) clustering method**

The layout on the left shows the SKUs assigned based on demand and the layout on the right shows the assignment based on clustering. In the demand-based layout, the SKUs are shaded by aisle and each SKU is shaded the same in the clustering layout. This shows how the demand is distributed across all of the aisles, so that the reduction in travel distance is accomplished by grouping the SKUs in an effective way, not by recreating the demand-based layout.

The impact of the clustering on the layout is shown by the results in Table 2. The values in this table correspond to the layout shown in Figure 1. For each aisle, the values from the full matrix in Table 1 are summed for all pairs in each cluster.

**Table 2: Number of orders containing pairs of SKUs found in each aisle**

Aisle	Demand	Cluster
1	1,087	3,656
2	269	4,152
3	2,342	3,702
4	360	3,961
5	709	2,227
6	205	3,857
7	1,140	2,527
8	2,283	4,741
9	1,648	2,818
TOTAL	10,043	31,641

These results show that there is a significant increase in the number of pairs in each aisle using the clustering method, compared to the demand-based method. This leads to the opportunity to pick all of the items for an order by visiting fewer aisles.

Different numbers of clusters were tested to evaluate their impact on the effectiveness of the solution. A set of test orders was generated and the number of aisles to be visited for each order was calculated. SKUs in an order that were not stored in the forward area were ignored. The metric for each layout was the total aisles visited to pick all orders. The number of aisles visited is shown in Table 3 and the percent improvement over the demand-based assignment is shown in Table 4.

**Table 3: Number of Aisles Visited to Pick Items for All Orders in Dataset**

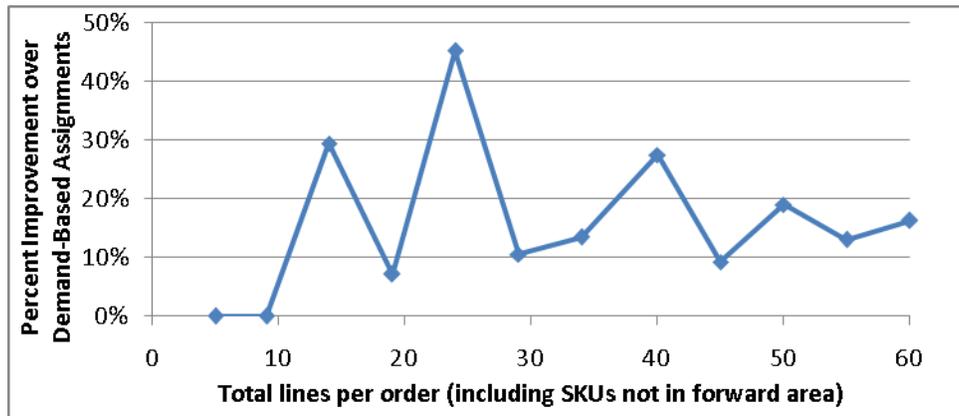
Problem Size	Number of aisles visited				
	Demand based	3 Clusters	5 Clusters	7 Clusters	11 Clusters
133 SKUs	67,852	58,450	57,871	58,783	58,351
90 SKUs	52,471	43,491	45,433	44,753	44,326

**Table 4: Number of Aisles Visited to Pick Items for All Orders in Dataset**

Problem Size	Percent improvement in number of aisles visited				
	Demand based	3 Clusters	5 Clusters	7 Clusters	11 Clusters
133 SKUs	0%	14%	15%	13%	14%
90 SKUs	0%	17%	13%	15%	16%

These results illustrate that there is a benefit from using clustering to determine the storage assignment, although the number of clusters is not significant in determining the amount of the benefit.

Testing was also conducted to evaluate whether the number of lines on an order is correlated with the amount of savings. That is, whether small orders would have less savings since there are only a few locations to be visited, which could mean that all of the items needed would be located in the same aisle with a demand-based layout if they consisted only of high-demand SKUs. Orders with more lines are more likely to have a mix of high and low demand SKUs, which may offer an opportunity for the clustering assignment to group the SKUs in the same aisle.



In examining the performance, the greatest benefit was for SKUs with a moderate number of line items. For small orders (5 lines or less) or large orders (45 lines or more), there was generally less benefit. Although it was not necessarily anticipated, the large orders may require pickers to visit almost all aisles in both layouts, again leaving little opportunity for either method to perform better.

## 4 CONCLUSIONS

The clustering method described in this paper provides a way to determine storage assignments that reduce the travel distance required in order picking. It is suitable for use in determining the configuration of the forward picking area of a facility and would result in a type of class-based assignment, since the assignments within an aisle are random.

The effectiveness depends on characteristics of the order fulfillment operations within the facility, such as the lines per order and the batching strategy being used. Although the methodology was testing assuming single-order picking, it could be utilized in batch picking while retaining the benefits. If batching is done by considering the aisles that need to be visited for each order (rather than randomly), a batch of orders requiring the picker to visit the same set of aisles can be generated.

Further testing is ongoing to validate the performance of the methodology and to better understand the situations in which it provides the most benefits. The impact of including more high-demand and low-demand SKUs in the analysis is being studied. In addition, the impact and feasibility of constraining the cluster size to match aisle capacity is also being studied. This would simplify the assignment of SKUs to aisles, since the number of SKUs in each cluster would match the capacity of an aisle.

## REFERENCES

1. de Koster, R., T. Le-Duc, and K. J. Roodbergen, 2007, "Design and control of warehouse order picking: a literature review," *European Journal of Operational Research*, 182(2), 481-501.
2. Rosenwein, Moshe B., 1994, "Application of cluster analysis to the problem of locating items within a warehouse," *IIE Transactions*, 26(1), 101-103.
3. Jane, C., Laih, Y., 2005, "A clustering algorithm for item assignment in a synchronized zone order picking system," *European Journal of Operational Research*, 166, 489-496.
4. Hwang, H., Baek, W., Lee, M., 1988, "Clustering algorithms for order picking in an automated storage and retrieval system," *International Journal of Production Research*, (26)2, 189-201.
5. Chen, M., Huang, C., Chen, K., Wu, H., 2005, "Aggregation of orders in distribution centers using data mining," *Expert systems with applications*, 28, 453-460.