Using cluster analysis for location decision problem

P Chain and S Arunyanart*
Supply Chain and Logistics System Research Unit, Department of Industrial Engineering, Faculty of Engineering, Khon Kean University, Khon Kaen 40002, Thailand

*Email: sirawadee@kku.ac.th

Abstract. Numbers of cattle farms have been increasing throughout the production regions. The aim of this paper is to investigate appropriate locations of regional food distribution centers in Northeast Thailand for distributing feed to cattle farms. A geographic information system (GIS) was used to map spatial data locations from 421 standardized cattle farms in 63 districts from Northeast Thailand. Hierarchical agglomerative clustering was applied as a statistical technique to sort datasets into homogeneous groups and three clusters were identified. The center of gravity (COG) method was then used to find the optimal locations of food distribution centers, with results suggesting Chatturat district, Pa Tio district, and Wang Sam Mo district as most suitable for distributing feed to cattle farms.

1. Introduction
In Thailand, the major area of beef cattle farming is located in the Northeast of the country with suitable tropical weather, humidity, dryness, and moderate annual rainfall [1]. Cattle farms normally use animal feed supplied by animal fodder production companies in the central region of Thailand. The financial crisis in 1997 resulted in many agricultural farms changing from traditional cattle farming to industrial systems, while demand for animal food products has increased along with development to ensure adequate animal feed availability for each farm. Prices of ingredients used to produce animal feed have increased along with higher transportation costs [2]. Thus, appropriate locations for regional centers to distribute animal feed products to each farm efficiently and cost-effectively are urgently required. This paper presents a solution to identify locations of appropriate animal feed distribution centers using the method of group creation (cluster) to identify groups of datasets and then apply the center of gravity (COG) method to determine specific locations for these distribution centers.

2. Literature Review
Beef cattle farming is a major agricultural industry in Thailand, which according to Thai Cattle Strategy was mainly maintained in the past by agricultural labor. However, beef cattle farming has now become a trading environment, with productivity depending on the consumption needs of the population [3]. Moreover, beef cattle farming has changed from natural free-range farming to penned livestock farming. Demand for animal food products is also increasing. Thus, distribution centers at proper location to provide animal food to the farms could make efficient supply network. Few academic papers are available which use cluster analysis to determine optimal location positioning.

Toro et al. [4] determined the Capacitated Location-Routing Problem (CLRP) which is a problem related to the location selection of distribution center and vehicle routing to customers. The objective is to minimize costs and environmental effect by using the mathematical model that presented a mixed-
The location for one cluster and the product transportation route to minimize cost. Hierarchical clustering and non-hierarchical clustering techniques which are agglomerative clustering and division clustering with similar variables were used to calculate an appropriate cluster for a suitable location. Domingues et al. [7] presented mathematical models to determine suitable locations for slaughterhouses which assist farmers to reduce the cost of beef transport distribution for export. Khalid and Herbert-Hansen [8] presented K-means Clustering method for investigate the International Location Decision (ILD). Widaningrum et al. [9] determined the suitable location of Distribution Center (DC) for foodservice industry in Jakarta. The objective is to minimize supply chain risk and environmental risk of Jakarta by using the Cluster Analysis. The results in this method show the location of the additional DC. Bosona and Gebresenbet [10] considered the development of local food supply chains and determined optimal product collection center (CC) locations using the geographic information system (GIS). The objectives were to improve efficiency in logistics and transportation networks of agricultural products using data collection of geographical locations to plan transportation routing. The COG technique was applied to determine CC locations. Wood and Brown [11] used GIS and COG to resolve location decision-making problems of UK food retailers. Fuente and Lozano [12] determined suitable locations of warehouses for food manufacturers in Spain using cluster analysis. Bosona et al. [13] presented a food distribution network using GIS analysis and COG to locate distribution centers and improve logistics and supply chain efficiency.

3. Methodology

Cluster analysis is a statistical technique to group similar observations into a number of clusters [14]. Clustering algorithms can be hierarchical or non-hierarchical. This study focused on the hierarchical cluster analysis (HCA) method which is used to identify relatively homogeneous groups of similar objects or distance between objects in multivariate datasets collected from fields. Two types of HCA cluster methods are agglomerative hierarchical and division hierarchical. This study used an agglomerative hierarchical cluster analysis with SPSS software because the agglomerative method has various processes that can be alternated between-group linkage, within-group linkage, nearest neighbor, etc. Between-group linkage was considered to be the most appropriate method because it can group data using the average distance among groups to cluster the smallest distant values in the same group.

An agglomerative hierarchical cluster algorithm starts by treating each data point as an individual cluster by considering the nearest distance measured from all the pairwise between the data point. The pairs of similar clusters are then successively merged to form new groups and the distance between the new groups is recalculated until one cluster is formed. Statistics are applied to calculate similar objects or distance between objects. The distance measure used for continuous variables was the Euclidean Squared distance metric using the following formula:

\[ D_{ij} = d + c ; D_{ij} > 0 \] (1)

Where \( D_{ij} \) is the number of variables at \( i \) and \( j \), both \( d \) and \( c \) are different value of data.

The Euclidian Distance formula is proposed to use between two points, which is the square root of the sum of squared distances. Since the Euclidian Distance is squared, it increases the importance of large distances, while reduces the importance of small distances.

The result of an agglomerative hierarchical cluster is a tree-based representation of the objects, named dendroogram.
The center of gravity method is used for locating single facilities that considers existing facilities, the distances between them, and the volumes of goods to be shipped between them [6]. The methodology involves formulas used to compute the coordinates of the two-dimensional point as following:

\[
X = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}, \quad Y = \frac{\sum_{i=1}^{n} y_i w_i}{\sum_{i=1}^{n} w_i}
\]  

Where \(X\) and \(Y\) are coordinate for the new facility, \(x_i\) is the coordinate of the existing location \(i\) on horizontal axis, \(y_i\) is the coordinate of the existing location \(i\) on vertical axis, and \(w_i\) is volume of goods moved to or from the location.

4. Result and Discussion
Data of 421 cattle farms in 63 districts and 9 provinces of Northeast Thailand were used to group the farms into clusters. The latitude and longitude, distance from animal feed production company, number of cattle, and number of farms were input to SPSS software to calculate individual clusters. A dendrogram graph was then analyzed to determine the appropriate number of clusters.

The cluster technique cannot show the result of statistics or hypotheses because there is no correct or incorrect solution of clustering. The researcher has to select the parameters which show the most accurate underlying structure of the data. Combining the dataset into clusters can be evaluated from the dendrogram as a graph showing distance or similarity of the data. Distance between clusters was rescaled into a range of 0-25 along the X axis.

As shown in Figure 1, it can be seen that the dataset can be grouped into several clusters depending on distance between the groups used. For example, if distances between the groups is set to 10 units, the dataset will be grouped into 4 clusters, and 9 clusters at 5 units. Therefore, it depends on the researcher or decision maker to decide the proper distances between the groups to be used.

In this study, by considering the data of latitude-longitude and the number of cattle in farms in each district, distance between the groups was set to 15 units. The hierarchical clustering dendrogram in Figure 1 shows that the dataset was grouped into 3 clusters by considering distance or similarity of location. The first cluster consisted of 17 districts in Chaiyaphum and Nakhon Ratchasima Provinces. The second cluster consisted of 4 districts in Roi Et and Ubon Ratchathani Provinces, and the last cluster consisted of 42 districts in Nakhon Phanom, Sakon Nakhon, Mukdahan, Udon Thani and Loei Provinces. Figure 2 shows the 3 clusters on a map of Northeast Thailand.

The hierarchical clustering is easy to understand and easy to do, however, it rarely provides the best solution because it involves lots of decisions, and the dendrogram can be misinterpreted. In addition, it does not work well on very large datasets and does not work when values are missing in the data.

After the clusters were determined, the COG method was applied to identify a single location of animal feed distribution center for each cluster using equation (2). Results showed the COG of cluster 1 at (15.58,101.93), cluster 2 at (15.81,104.29), and cluster 3 at (17.12,103.45) as Chatturat district, Pa Tio district and Wang Sam Mo district, respectively. These three districts were suitable for the locations of feed distribution centers as shown in Figure 2.
Figure 1 Dendrogram showing the clustering of 63 districts.
5. Conclusion
A method was proposed to determine suitable locations for regional centers to distribute animal feed to cattle farms in Northeast Thailand. Agglomerative hierarchical cluster analysis methods were applied to cluster the dataset of 421 standard farms in 63 districts from 9 provinces. The center of gravity method was then applied to find a suitable location. Results showed that three districts as Chatturat, Pa Tio and Wang Sam Mo were suitable locations for distribution centers of animal feed in Northeast Thailand.

This study forms part of my research to improve beef logistics and supply chain network design in Thailand.

Acknowledgments
This research is supported by supply chain and logistics system research unit department of industrial engineering Khon Kaen, Thailand.

figure 2 Locations of three food distribution centers and clusters.
References

[1] Osothongs M, Khemsawat J, Sarakul M, Jattawa D, Suwanasopee T and Koonawoottririon S 2016 Proc. of Int. Symp.: Dairy Cattle Beef Up. Beef Industry in Asia: Improving productivity and Environmental Sustainability (Bangkok, Thailand, 19 August 2016) Vol 19 pp 5-8

[2] Chanjula P 1997 Sci. and Technol. 5 8-19

[3] Department of Livestock Development 2016 Thai Cattle Strategy 2012-2016 Report (Thailand)

[4] Toro EM, Franco JF, Echeverri MG and Guimarães FG 2017 Comp. & Ind. Eng. 110 114-25

[5] Stević Ž, Vesković S, Vasiljević M and Tepić G 2015 2nd Logistics International Conference (Belgrade, Serbia) pp 86-91

[6] Barreto S, Ferreira C, Paixao J and Santos BS 2007 European J. Oper. Res. 179 968-77

[7] Domingues Zucchi J, Zeng AZ and Caixeta-Filho JV 2011 Int. J. Log. Res. and Appl. 14 134-48

[8] Khalid W and Herbert-Hansen ZN 2018 J. Glob. Oper. and Stra. Sour. 11 274-300

[9] Widaningrum L, Andika A and Murphiyanto J 2017 IOP Conf. Ser.: Earth and Envi. Sci. 109 012021

[10] Bosona TG and Gebresenbet G 2011 Bio. eng. 108 293-302

[11] Wood S and Browne S 2007 Int. J. Retail & Dist. Manage. 35 233-55

[12] de La Fuente D and Lozano J 1998 Int. J. Phys. Dist. & Log. Manage. 28 68-79

[13] Bosona T, Nordmark I, Gebresenbet G and Ljungberg D 2013 Int. J. Busi. and Manage. 8 13

[14] Yim O and Ramdeen KT 2015 The quanti. methods for psyc. 11 8-21