

## Clustering Logistics Facilities in a Metropolitan Area via a Hot-Spot Analysis

**İsmail Önden**  
Okan University  
Department of  
International Logistics,  
İstanbul, Turkey  
[ismail.onden@okan.edu.tr](mailto:ismail.onden@okan.edu.tr)

**Fahrettin Eldemir**  
Yildiz Technical University  
Industrial Engineering  
Department,  
İstanbul, Turkey  
[feldemir@yildiz.edu.tr](mailto:feldemir@yildiz.edu.tr)

**Metin Çancı**  
Okan University  
Department of  
International Logistics,  
İstanbul, Turkey  
[metin.canci@okan.edu.tr](mailto:metin.canci@okan.edu.tr)

### Abstract

Hot spots can be described as the high attraction points. Defining the hot spots and clustering approaches in a metropolitan area helps to provide solutions for balancing the freight flows between the sub-areas in the city center. It also helps to provide solutions for secondary problems, such as traffic congestion and air pollution. The technique assists decision-makers in making inferences about the city's future and taking precautions for sustainability. In this paper, a geographic information systems (GIS) analysis tool, spatial statistics based on the Getis Ord\* statistics, is used to illustrate which part of Istanbul has hot spots. The hot spot identification is based on logistics activities at the locations of the logistics facilities. The outputs of the analysis are discussed within the context of logistics costs and environmental effects. Anahtar Kelimeler: Mağaza yeri seçimi, delphi tekniği, makro analiz, mağaza yeri seçim kriterleri

**Keywords:** Sustainability; city logistics; clustering; geographic information systems (GIS); spatial statistics; spatial analysis

### 1. Introduction

Network models can be seen as hard-to-solve problems, even under restrictive assumptions. On the other hand, logistics networks in metropolitan areas are formed by even more sophisticated structures, due to a larger scale than the studied problems. This may cause problems, because of the uncertainties which are not taken into consideration in the theoretical studies. This is owed to the limitations, such as the amount of the demand, supply nodes, and regulations. The reasons for the regulations can be seen as reasonable, due to the problems faced by a metropolis, such as traffic congestion and environmental concerns.

Another aspect of the complexity is the expected service levels by citizens in a metropolis that stems from fast city life. The stated conditions produce the necessity for a good management of supply chains for companies. On the other hand, for sustainable cities, city management should be in charge of regulating the flow distributions in the

city borders to avoid the negative impacts mentioned previously. To control for these difficulties and to achieve sustainable city management, it is of great importance to better understand how these activities take place on the plane.

Companies should take cost issues into consideration to provide for sustainability, while city managers should prepare the infrastructure of the cities to meet these expectations. In the literature, logistics costs are taken into consideration with multi-dimensions, such as distance, time, composition, transshipments, and decomposition (Hesse and Rodrigue, 2004). The centralization literature also focuses on the same issues to minimize inventory costs and travelled distances (Eppen, 1979). Within the management literature, many studies have illustrated that the centralization of the inventories has positively affected the financial status of the companies.

Centralization studies first appeared in the 1960s in textbooks (Brown, 1967; Starr and Miller, 1962). Not long afterwards, the square root law was proposed to express how centralization decisions affect inventory costs (Heskett et al. 1973). According to the law, firms can save up to 78% of their inventory cost by centralizing their 20 facilities into 1 facility.

The formal proposition of the square root law was conducted by Maister in the mid-1970s (Maister, 1976). Following the Maister study, Eppen proposed the newsboy problem framework and found that the total inventory level was lower than the decentralization system (Eppen, 1979). Ballou found that the centralization effect illustrates variety according to the focused sectors (Ballou, 1981). The following studies focused on the risk pooling and lateral shipments concept (Jonsson and Silver, 1987; Lee, 1987; Tagaras, 1989). Later, portfolio effect is been studied as another centralization topic (Tallon, 1993). Within this context, centralization of the facilities can be taken as an evaluation topic of city management to reach the cost objectives of the companies. To reach this aim, optimization approaches and clustering algorithms can be taken as a solution tool for mapping the logistics activities and determining the logistics centers.

Satisfying the customer service level and maximizing the total profit of the firms can be seen as conflicting objectives; thus, companies and city managers should take advantage of all of the analyzing tools to better understand and solve the conflict. There are different approaches available to overcome the mentioned trade-off between objectives, such as multi-objective approaches and information technologies. Different approaches can be modeled to represent the real system to determine how to build the companies' logistics network. They can also be used to determine how the specialized packages of the known geographic information systems (GIS) software can track the vehicles of the companies and manage the existing structure. Before deciding on how to approach the problem, the decision environment should be evaluated.

In addition to the stated studies which used mathematical approaches to evaluate the logistics costs, information systems, such as GIS, can be used with clustering algorithms for city management issues. There are different spatial calculation abilities available, such as the distance based algorithms. There are also different statistical analyses available, such as the calculation of the densities and hot spot analysis. Euclidean analysis approaches are generally used in a suitability analysis to determine the locations of the facilities (Sen, et al. 2011); on the other hand, hot-spot analysis is used to determine hot-spots statistically under Moran's I and Getis Ord's statistical

background to order the sub-areas of the study area. These sub-areas can be used to locate the hot-spots of the focused activities, such as logistics freight.

As mentioned in the previous paragraphs, there are different approaches to manage logistics activities. In this study, we focus on the centralization of the logistics facilities and its cons and pros. There is also an example to illustrate how logistics facilities can be clustered with GIS; thus, spatial statistics approaches are taken into consideration to better understand how the logistics activities are sprawled on the plane in the metropolitan area and truck parking spots are accepted as the origin of the logistics flows. These facilities are used for city planning to determine where to site a logistics center. The application stands as the first example of the clustering of the logistics facilities in a metropolitan area in a centralization concept. Thus, the paper proposes an information technique for both a city management and logistics management issue.

The structure of the rest of the paper is as follows. In the following section, the specialties of the study area and the data are taken into consideration in detail. In addition, why this area and the selected dataset was used for the analysis is discussed. In the third section, the used technique's background and its steps are expressed. The fourth section is devoted to illustrate how the analysis is performed in the study area. The final chapter, the conclusion, includes the final discussion and summarization of the study.

## **2. Study Area and Data**

This chapter provides information about the focused study area and what kind of data is needed to be used for a hot spot analysis. In following section, the characteristics of the transportation and freight in Istanbul will be explained. In the following section, the data types used and the data gathering process will be discussed.

### **2.1. About the Study Area**

Cities can be classified into tourism destinations and industrial centers. These classifications have different decision environments for decision-makers, due to their logistics activities. In the case of cities that fit into the mentioned classification, it is relatively easy for policy-makers to discuss what the goals of city management are and determine the regulations. There are also mediocre cities, which cannot fit into a category, and include both tourism and industry functions in their borders, which makes their structures relatively sophisticated, according to others. Decision-makers in these cities deal with the environmental consequences of the industrial facilities, such as air pollution.

Within this particular study, Istanbul, the most crowded metropolis in Europe, according to its population size, is selected as the study area. Istanbul is a mediocre city which attracts millions of tourists every year with its historical background. Istanbul is also the center of industrial activities in its region. In addition, there are many ongoing multi-billion dollar projects aimed to increase the economical attractiveness of the city. As a result of these investments, it is expected that the complexity of city management will increase.

The distribution of the goods in metropolitan borders requires city logistics to answer the demands of over thirteen million people. This is done with firms' own vehicles or third party logistics firms with decentralized warehouse network structures, which negatively affects the existing traffic problem in Istanbul.

On the other hand, the European Union (EU) is the biggest economic partner of Turkey. Its demand is mostly met by road transportation and maritime transportation with RO-ROs. In addition to Istanbul's economic relationship with the EU, Istanbul is a natural bridge between Asia and Europe, and transportation cannot be conducted without using inner city highway lines. Owing to the high transportation demand on the motorway network, the city is faced with high traffic congestion and air pollution.

## 2.2. Data

In this study, the clustering of the logistics activities in the metropolitan area, considering the geographic aspects, is the focus. It is also aimed at focusing on the logistics costs, traffic congestion and air pollution. To decrease the different logistics costs, such as transportation, cycle stock, and safety stock (Chandrasekhar Das and Tyagi, 1997), centralization can be used (Eppen, 1979; Heskett et al. 1973; Maister, 1976). Consequently, a hot spot analysis is applied to map the hot spots of the logistics activities in the metropolitan area. To perform a hot spot analysis and map the logistics activities in the study area, the origins of the logistics flow-ins are accepted as truck parking locations.

Based on the proposed methodology, discussed in the next section, there is a data requirement to represent real transportation network systems in GIS. The Istanbul Metropolitan Planning (IMP) department provided the raw data for the truck parking addresses, company names, and number of owned trucks by each company, as well as the parking capabilities of each truck parking spot. The gathered data was converted into vector data to better represent the real system on GIS and a geographic database, which contains the geographic locations of the truck parking spots and the other non-spatial features of the point data.

In addition to the logistics facilities, different data were needed, such as the districts of the study area, the borders of the freshwater reserves and the sea borders, to illustrate the logistics densities in the study area. The Istanbul region was separated into 967 sub-areas, according to the hotspot analysis and the borders of the sub-regions of the city. The spatial and non-spatial data of the logistics facilities were then spatially joined in the created city map. The sub-regions of the study area and the locations of the truck parking areas are represented in Figure 1.

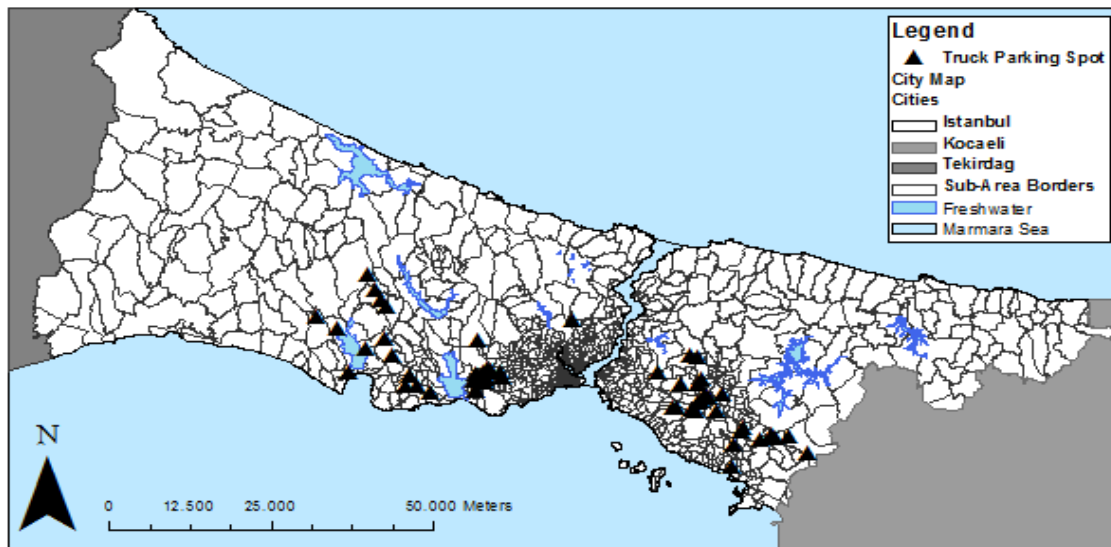


Figure 1: The study area and the collected data

### 3. Methodology

Hot spot analysis is a spatial statistics method used for statistically finding the hot or cold spots on the plane or for clustering the separated sub-groups of the study area based on a spatial autocorrelation and Moran's I structures. In addition to the harmony of this technique with the proposed methodology, there are existing studies based on hot spot analysis for different research applications. The technique used the GIS abilities to determine the traffic accident hot spots of Afyonkarahisar city (Erdogan et al. 2008).

In Levine's study (Levine, 2006), they used a spatial autocorrelation structure and hot spot analysis for crime mapping. GIS was used to identify hot spots for pedestrian-vehicle crashes with 13 years of gathered data under spatial correlation, Moran's I, and Getis-Ord  $G_i^*$  rules (Truong and Somenahalli, 2011).

Moran's I index is needed to measure the spatial autocorrelation to examine the spatial patterns of the study area. Moran's I measures the similarities of the attributes and calculates an index which consists of location proximities. There are different approaches to measure the location proximities; the zone of indifference is one of the options for a calculation. The stated index can be calculated using Equation 1, where  $w_{ij}$  is the proximity weight of location i and location j (where  $w_{ii}=0$ ;  $x_i$  is the severity index at location j;  $\bar{x}$  is the global mean value; and n is the total number of focused locations) (Truong and Somenahalli, 2011).

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x}) (x_j - \bar{x})}{(\sum_{i=1}^n \sum_{j=1}^n w_{ij}) (\sum_{i=1}^n (x_i - \bar{x})^2)} \quad (1)$$

Moran's I's statistical significance can be calculated via z-score methods. In Equation 2, the expected values (E [I]) for a random pattern, the variances (VAR [I]) are used. The mathematical representation of the Z-score is:

$$Z = \frac{I - E(I)}{\sqrt{VAR(I)}} \quad (2)$$

The appropriate distance threshold value should be found where the spatial autocorrelation is maximized, since each data point is analyzed in terms of its neighboring data points, as defined by a distance threshold. To reach the maximum value, the spatial autocorrelation tool should be performed several times. The changes in the z value should be observed empirically. The Z-value illustrates whether the data is clustered or randomly distributed on the plane with a determined significance level.

The Getis-Ord statistic (Getis and Ord, 1992) is used for specifying hot spots. A high value Getis-Ord statistic expresses the hot spots; whereas a low value illustrates the cold spots. The mathematical definition of the Getis-Ord  $G_i^*$  statistic is given in Equations 3 and 4:

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d) x_j}{\sum_{j=1}^n x_j} \quad (3)$$

$$Z (G_i^*) = \frac{I - E (I)}{\sqrt{VAR (I)}} \quad (4)$$

#### 4. Application

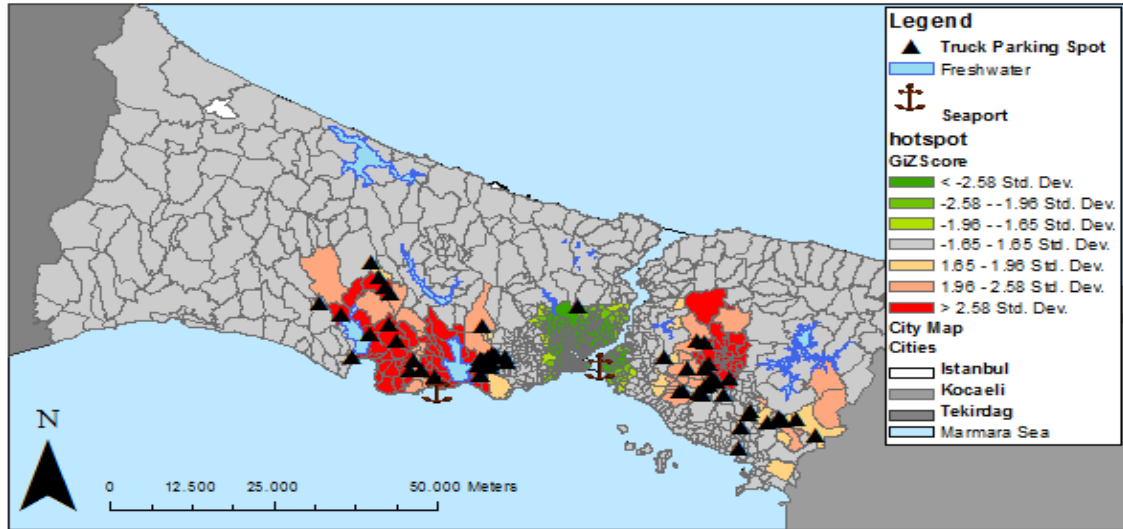
There are several steps and different calculations to reach the hot spots in GIS/SA; they are described in the methodology section. For this step, ESRI's ArcGIS software's spatial statistics toolbox is used for the analysis. The study area is separated into 967 parts, according to the borders of the districts of Istanbul, to determine which districts are the hot or cold spots, as related to their logistics activities. To cluster the separated districts according to the volumes of the logistics activities, a shape file is created. The shape file consists of the number of the logistics facilities and the number of trucks of these companies in each district. After creating the geographic data for the analysis, the z values and the Moran's I for the threshold distances are calculated to determine the maximum z score, where the spatial autocorrelation is maximized. The changes of the z scores are searched for through 23 iterations of Moran's I; these findings are presented in Table 1. The proper threshold distance was found to be 9150 meters; this value yields the best z score, 8.174. It also illustrates the statistically clustered pattern of the facilities.

**Table 1. Morans'I Z Scores**

| ITERATION | DISTANCE (M.) | Z SCORE | SIGNIFICANCE |
|-----------|---------------|---------|--------------|
| 1         | 25            | 0.307   | 0.01         |
| 2         | 500           | 0.567   | 0.01         |
| 3         | 1500          | 3.212   | 0.01         |
| 4         | 2500          | 5.939   | 0.01         |
| 5         | 3500          | 6.696   | 0.01         |
| 6         | 4500          | 6.881   | 0.01         |
| 7         | 5500          | 7.502   | 0.01         |
| 8         | 6500          | 7.325   | 0.01         |
| 9         | 7500          | 7.288   | 0.01         |
| 12        | 8500          | 7.644   | 0.01         |
| 11        | 9000          | 7.744   | 0.01         |
| 12        | 9150          | 8.174*  | 0.01         |
| 13        | 9200          | 8.169   | 0.01         |
| 14        | 9250          | 8.134   | 0.01         |
| 15        | 9350          | 8.069   | 0.01         |
| 16        | 9500          | 7.996   | 0.01         |
| 17        | 9750          | 8.092   | 0.01         |
| 18        | 10500         | 7.893   | 0.01         |
| 19        | 12500         | 7.563   | 0.01         |
| 20        | 15000         | 5.628   | 0.01         |
| 21        | 17500         | 4.143   | 0.01         |
| 22        | 20000         | 3.521   | 0.01         |
| 23        | 25000         | 0.924   | 0.01         |

After reaching the threshold distance with the pattern analyzed, ArcGIS' Getis Ord  $G_i^*$  tool is used to calculate the statistical significance based on the districts' cluster orders. The results of the Getis Ord  $G_i^*$  analysis are presented in Figure 2, which

is performed with the threshold distance values and is 9150 meters when the Euclidean distances are used for the analysis and for the conceptualization of the spatial relationships based on the zone of indifferences. In the map, the clustered districts illustrate the performances of the alternative locations, whose convenience is researched for the new facility.



**Figure 2: Illustration of the hot spots in the study area**

## 5. Conclusions

To build a sustainable city management structure, an accurate data evaluation should be conducted by the policy-makers of the regional management. Accordingly, all possible analytic techniques are needed to better understand the effects of the sprawling of the facilities. Within this context, spatial relationships can be used for logistics planning, due to its effects on traffic congestion, air pollution, company cost issues and competitiveness issues. In this paper, we used a spatial statistics technique in mapping logistics activities for the first time in the literature to illustrate how logistics activities are actualizing.

The study had two primary goals. The first goal focused on mapping the logistics activities and determining the hot-spots in a metropolitan area, where policy-makers need quantitative tools when they face difficulties in measuring actual activities in the region due to a lack of data technology, such as creating an O-D matrix between regions. The other goal was evaluating a spatial statistics analysis tool, such as a centralization methodology, instead of following the existing literature, which is based on optimization models under limited suppliers, origins or demands, and destination constraints.

In the paper, an actual evaluation was conducted based on spatial statistics methods to better understand how logistics activities are sprawled on the plane. GIS' spatial statistics tool was used to determine the hot spots and cold spots in the focused region. This was done according to the locations of truck parking spots and the capacity of parking areas; these locations are related to their effects on traffic and logistics flows. Accordingly, a geographic database was created for the analysis. The necessary data was gathered from the regional authorities, the logistics companies' locations, and the logistics companies' non-spatial specialties, such as parking capacities, number of



trucks, and number of employees. After the data was collected, a hot spot analysis was performed. The hot spots were then determined (Figure 2). In the analysis, the z values and Moran's I for the threshold distances were searched to determine the maximum spatial autocorrelation result. While searching for the highest value of the z, the analysis was performed 22 times. The highest z was found with a 9150 meter threshold distance, where it reached the value of 7996, a statistically clustered pattern.

There were many goals for this study. The first goal of the study was to explore how the hot-spots can be detected in metropolitan areas via GIS abilities. This goal is related to create information about flows between regions, which is generally too expensive to track for decision-makers. Using the hot-spot analysis, it is possible to make some interference to discuss how the activities are sprawled on the plane. With the calculated statistical hot-spots, it is possible to rank the sub-areas. The ranking values can be used in multi-criteria approaches for decision-making applications; the results of the hot-spot analysis can be integrated with an overlay tool in GIS for the different analysis outputs to create new suitability maps with priority levels for the calculation of each map (Onden et al. 2012; Onden et al. 2012; Sen et al. 2011).

The second objective is that the paper discusses how the hot-spot analysis can be used as a centralization tool for logistics activities. The existing literature illustrates that the centralization of the inventories provides benefits to company costs (Ballou, 1981; C. Das, 1975; Eppen, 1979; Maister, 1976). Due to this proven positive impact, this approach has also been applied to the city management with centralization of the main logistics facilities to decrease the total traveled distances by trucks. This can result in cost minimization. It can also decrease the relative impacts, such as carbon emissions and traffic congestion. The primary structure of the analysis is to detect the main hot-spots in the study area with the selected weight values. The output of the analysis illustrates that the logistics activities are clustered, according to the Getis Ord  $G_i^*$  and Moran's I approaches.

The study area consists of 7 sub-areas, according to the statistical rules (Figure 2). The number of classifications can be changed easily. The highest level of the logistics activities found can be changed statistically, if necessary. That is why the application can be used as a location selection and evaluation technique. The study focused on the logistics activities, according to the strategic aspect. However, if a location of the facility is searched by a company, a similar study could be conducted to determine the suitability background.

The discussed method is useful for mapping the current status of the cities. However, there might be a necessity to make interferences to better understand how the sprawl changes. At that point, a sensitivity analyses should be performed. These approaches generally discuss the what-if questions with different scenarios. Within this context, due to the similarity in their outputs, a sensitivity analysis (Delgado and Sendra, 2004; Panell, 1997) for a multi-criteria spatial analysis (Chakhar and Mousseau, 2007; Jankowski and Richard, 1994; Karnatak et al. 2007; Vahidnia et al. 2009) can be taken into consideration to create the scenarios. For our study, creating expected parking lots with their weights might be an example of the previously discussed evaluation tool. The output of the sensitivity analysis can create the opportunity to determine how the logistics activities in the cities might illustrate a variety with their new locations.



All in all, in the paper, how logistics activities can be illustrated and how the output of the discussed method used in decision-making processes is expressed. One example case for the method was analyzed for a better understanding. Then, the outputs of the analysis were discussed. The usability of the method in centralization studies and multi-criteria decision-making processes is then argued.

## References

- Ballou, R. H. (1981). Estimating and auditing aggregate inventory levels at multiple stocking points. *Journal of Operations Management*, 1, 143–153.
- Brown, R. G. (1967). *Decision Rules for Inventory Management*. Hinsdale, IL: Dryden Press.
- Chakhar, S., & Mousseau, V. (2007). Spatial multicriteria decision making. *LAMSADE: Publications*. Retrieved from <http://www.lamsade.dauphine.fr/mcda/biblio/PDF/ChakharMousseauInbook2007b.pdf>
- Das, C. (1975). Supply and redistribution rules for two-location inventory systems: one period analysis. *Management Science*, 21, 765–776.
- Das, C., & Tyagi, R. (1997). Role of inventory and transportation costs in determining the optimal degree of centralization. *Transportation Research Part E: Logistics and Transportation Review*, 33(3), 171–179.
- Delgado, M. G., & Sendra, J. B. (2004). Sensitivity Analysis in Multicriteria Spatial Decision-Making: A Review. *Human and Ecological Risk Assessment: An International Journal*, 10(6), 1173–1187. doi:10.1080/10807030490887221
- Eppen, G. D. (1979). Effects of centralization on expected costs in a multi-location newsboy problem. *Management Science*, 25(5), 498–501.
- Erdogan, S., Yilmaz, I., Baybura, T., & Gullu, M. (2008). Geographical information systems aided traffic accident analysis system case study: city of Afyonkarahisar. *Accident Analysis and Prevention*, 40, 174–181.
- Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189–206.
- Heskett, J. L., Glaskowsky, N. A., & Ivie, R. M. (1973). *Business logistics: Physical distribution and materials management*. New York: Ronald Press.
- Hesse, M., & Rodrigue, J.-P. (2004). The transport geography of logistics and freight distribution. *Journal of Transport Geography*, 12, 171–184.
- Jankowski, P., & Richard, L. (1994). Integration of GIS-based suitability analysis and multicriteria evaluation in a spatial decision support system for route selection. *Environment and Planning B: Planning and Design*, 21, 323–340.
- Jonnsson, H., & Silver, E. A. (1987). Analysis of a two-echelon inventory control system with complete redistribution. *Management Science*, 33, 215–227.
- Karnatak, H. C., Saran, S., Bhatia, K., & Roy, P. S. (2007). Multicriteria Spatial Decision Analysis in Web GIS Environment. *GeoInformatica*, 11(4), 407–429. doi:10.1007/s10707-006-0014-8
- Lee, H. L. (1987). A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, 33, 1302–1316.
- Levine, N. (2006). Crime Mapping and the Crimestat Program. *Geographical Analysis*, 38(1), 41–56.

- Maister, D. H. (1976). Centralization of inventories and the “square root law.” *International Journal of Physical Distribution and Materials Management*, 6(124-134).
- Onden, I., Güngör, C., & Şen, A. (2012). Integration of Integer Programming with GIS Analyzing Abilities for Determining the Convenience Levels of Retail Stores. *Procedia-Social and Behavioral Sciences*, 62, 1144–1150.
- Onden, I., Tuzla, H., & Cobb, S. (2012). Evaluation of the retail store location alternatives for investment decisions using the delphi technique and geographic information systems. *International Business: Research, Teaching and Practice*, 6(2), 64–75.
- Panell, J. (1997). Sensitivity analysis of normative economic models: Theoretical framework and practical strategies. *Agricultural Economics*, 16, 139–152.
- Sen, A., Onden, I., Gokgoz, T., & Sen, C. (2011). A GIS approach to fire station location selection. In *GI4DM 2011 GeoInformation For Disaster Management Conference*. Antalya.
- Starr, M. K., & Miller, W. D. (1962). *Inventory Control: Theory and Practice*. Englewood Cliffs, NJ: Prentice-Hall.
- Tagaras, G. (1989). Effects of pooling on the optimization and service levels of two-location inventory systems. *IIE Transactions*, 21, 251–257.
- Tallon, W. J. (1993). The impact of inventory centralization on aggregate safety stock: the variable supply lead time case. *Journal of Business Logistics*, 14, 185–203.
- Truong, L. T., & Somenahalli, S. V. C. (2011). Using GIS to Identify Pedestrian-Vehicle Crash Hot Spots and Unsafe Bus Stops. *Journal of Public Transportation*, 14(1), 99–114.
- Vahidnia, M. H., Alesheikh, A. a, & Alimohammadi, A. (2009). Hospital site selection using fuzzy AHP and its derivatives. *Journal of Environmental Management*, 90(10), 3048–56. doi:10.1016/j.jenvman.2009.04.010