

# **AN EFFICIENT HEURISTIC FOR THE SINGLE-FACILITY EUCLIDEAN LOCATION PROBLEM**

**Ching-Chung Kuo**

Department of Management, College of Business, University of North Texas, Denton, TX 76203, U.S.A.

It is well known in location theory that the modified gradient procedure (MGP) is called for to solve the Euclidean mini-sum single facility location problem (MSSFLP). Since it is often tedious to implement the MGP, some researchers have proposed the use of the center of gravity (COG) as an approximation to the optimal solution to the Euclidean MSSFLP to reduce the computational efforts. They argue that, aside from the advantage of being easy to calculate, the COG usually results in a transportation cost that is close to the true minimum albeit no evidence is provided. The main thrust of this paper is to conduct a simulation experiment to empirically substantiate the claim. In addition, data collected from the study is subject to regression analysis to identify the key factors affecting the effectiveness of the COG as a substitute for the MGP-based optimal solution to the Euclidean MSSFLP in practical applications.

**Keywords:** Facility location; Center of gravity; Heuristics; Simulation; Regression analysis.

*(Received 1 May 2009; Accepted in revised form 1 June 2010)*

## **1. INTRODUCTION**

Location problems arise in a broad array of contexts and have found widespread applications in both the private and public sectors of our society, including agriculture (Lucas and Chhaged, 2004), healthcare (Daskin and Dean, 2004), land use (Chan, 2005), logistics (Taylor, 2008), manufacturing (Steenhuis and De Bruijn, 2004), and government services (Narasimhan et al., 2005) among others. They differ in their characteristics and may be categorized by the number of facilities to be located or relocated, the objective to be achieved, the distance measure to be used, the spatial space to be involved, or any other attribute (Francis et al., 1983; Brandeau and Chiu, 1989). Of particular interest in this paper is the class of mini-sum single facility location problems (MSSFLPs), where a new facility is to be established among a number of existing facilities. The volume of goods to be shipped between the new establishment and any of the existing facilities is known and the transportation cost is linearly proportional to an appropriately defined distance. The aim of the MSSFLP is to determine the location of the new facility so that the sum of the weighted transportation costs is as low as possible.

The history of the MSSFLP dates back to a problem posed by Pierre de Fermat in the early 17<sup>th</sup> century and subsequently solved by Evangelista Torricelli prior to 1640 (Love et al., 1988). The results were generalized by Jacob Steiner in 1837 and Alfred Weber incorporated socioeconomic factors in addition to distance into the location model in the early 20<sup>th</sup> century (Sivazlian and Stanfel, 1975). Since then, many extensions as well as variations of the problem have been studied and solution algorithms have been proposed. Excellent surveys of theoretical developments and practical applications in this area can be found in Drezner (1995), Drezner and Hamacher (2002), Francis et al. (1992), as well as Nickel and Puerto (2005).

Despite the large body of current location literature, very few researchers have explored the possibility of using the center of gravity (COG) as a heuristic solution to the MSSFLP involving Euclidean distances, which can only be solved optimally by applying the modified gradient procedure (MGP). The advantage of the surrogate is two-fold. Firstly, the COG may be computed quite easily whereas it is time-consuming to implement the MGP. Secondly, some authors have observed that although the COG is not optimal to the Euclidean MSSFLP, it is often in close proximity to the MGP-based optimum and, consequently, the resulting transportation cost does not differ significantly from the true minimum. Nevertheless, no theoretical or empirical evidence has been provided to substantiate the observation. The purpose of this paper is to bridge the gap by performing a simulation experiment and regression analysis to shed some lights on this seemingly interesting issue.

## **2. PROBLEM FORMULATION AND SOLUTION**

MSSFLPs can take place in a one-, two-, or three-dimensional space. The one-dimensional location problem is useful in situations ranging from siting railroad stations for a transportation network (Singhal, 1990) to establishing rest areas along a section of interstate highway involving multiple points of entry (Mirchandani et al., 1996) to locating load/unload stations in a flexible manufacturing system where machines are laid out in a single row and serviced by automated guided vehicles (Kouvelis and Chiang, 1992). On the other hand, the two-dimensional model is applicable to, for instance, the

determination of the optimal site for a distribution center in a large geographic region (Brimberg and Love, 1998), the location of a service hub for an airline on a continent (Drezner and Hamacher, 2002), the positioning of a load extractor for a storage carousel (Egbelu and Wu, 1998), as well as the establishment of public facilities (Suzuki et al., 1991). Further, a large number of problems encountered in manufacturing and telecommunication industries may be formulated as three-dimensional location models and solved, such as the selection of the dwell point of an idle machine in an automated storage/retrieval system with minimum expected response time (Chang and Egbelu, 1997) and the placement of transmitters in a radio communication system to minimize path losses (Sherali et al., 1996).

Three distance measures are commonly used in location research: rectilinear, Euclidean, and squared Euclidean. Rectilinear distances are appropriate in urban cities with orthogonal streets, industrial warehouses where storage shelves are arranged as a grid, and modern service organizations with rectangular hallways to facilitate the travel of employees and customers. A handy example involving Euclidean distances is the location of a power-generating facility to minimize the total length of electrical cable that must be laid to connect the plant to the customers at different towns in a large level area. One also notes that air travel tends to follow a straight-line path, so does the movement of materials on a conveyor system or along a pipeline (Kuo and White, 2004). As for squared Euclidean distances, one of the most interesting applications is in the relocation of a blood donor clinic and transfusion center in Quebec, Canada (Price and Turcotte, 1986).

The specific problems to be examined in this paper are MSSFLPs with Euclidean distances in one-, two-, and three-dimensional spaces. In the remainder of this section, however, the exposition will focus on analysis of issues involving two spatial dimensions since location models defined on a plane are most popular and have been extensively studied. Let

- $n$  = Number of existing facilities
- $(a_i, b_i)$  = Planar coordinates of the location of existing facility  $i$ ,  $i = 1, 2, \dots, n$
- $(x, y)$  = Planar coordinates of the location of the new facility to be established
- $w_i$  = Transportation cost per unit of distance between  $(a_i, b_i)$  and  $(x, y)$ ,  $i = 1, 2, \dots, n$

The Euclidean MSSFLP is concerned with finding the values of  $x$  and  $y$  so that the total transportation cost represented by the following function is minimized:

$$f(x, y) = \sum_{i=1}^n w_i [(a_i - x)^2 + (b_i - y)^2]^{1/2} \quad \dots \quad (1)$$

This problem turns out to be rather difficult to wrestle with and there is no simple algebraic approach to solving it. Instead, the tedious MGP due to Kuhn (1967), which is also known as the Weiszfeld procedure (Drezner, 1995), is called for to obtain the optimal solution  $(x^*, y^*)$ . The algorithm normally starts with the COG located at  $(x_0, y_0)$ , which is defined in Step 0 below, and the solution is improved and getting closer to the optimum in each iteration until two successive solutions are identical. The implementation steps are summarized below:

Step 0: Let  $k = 0$  and  $(x_0, y_0) = \left( \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}, \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i} \right)$ .

Step 1: Let  $k = k + 1$  and  $g_i(x_{k-1}, y_{k-1}) = w_i / [(x_i - x_{k-1})^2 + (y_i - y_{k-1})^2]^{1/2}$ ,  $i = 1, 2, \dots, n$ ,  $(x_k, y_k) = \left( \frac{\sum_{i=1}^n g_i(x_{k-1}, y_{k-1})(x_i)}{\sum_{i=1}^n g_i(x_{k-1}, y_{k-1})}, \frac{\sum_{i=1}^n g_i(x_{k-1}, y_{k-1})(y_i)}{\sum_{i=1}^n g_i(x_{k-1}, y_{k-1})} \right)$ .

Step 2: If  $(x_k, y_k) = (x_{k-1}, y_{k-1})$ , then  $(x^*, y^*) = (x_{k-1}, y_{k-1})$  and stop; otherwise, go to Step 1.

While the eventual termination of the iterative process is guaranteed under some conditions (Ostresh, 1978; Chandrasekaran and Tamir, 1989), the series of solutions  $(x_k, y_k)$  could converge very slowly and, hence, a great deal of computational efforts might be required to reach the optimality especially in large problems (Drezner, 1995). Additionally, as pointed out by Kuhn (1967) and Francis et al. (1992), the MGP will fail if the least-cost location overlaps with one of the existing facility locations.

In contrast, the COG may be found efficiently by using the closed-form formulas presented in Step 0 of the MGP. Moreover, Krajewski et al. (2007) argue that although the COG does not exactly solve the Euclidean MSSFLP, it serves as an excellent starting point to find the optimum. Martinich (1997) notes that the COG is usually near the lowest-cost location for the new facility. Along the same line, Goetschalckx (2008) suggests the use of the COG as an approximate solution in designing distribution systems when the shipping cost is linearly proportional to the Euclidean distance traveled. All of these observations make the COG an attractive substitute for the MGP-based optimal solution in practice, but there is a lack of theoretical or empirical evidence in the existing literature of location research to substantiate them. In what follows, we

will take a simulation approach to tackling the problem of how good the COG is as a heuristic solution to the Euclidean MSSFLP in one-, two-, and three-dimensional spaces. The evaluation criterion to be used is the average percent deviation of the COG-based total transportation cost from the MGP-based total transportation cost.

### 3. SIMULATION DESIGN AND EXPERIMENT

In Section 2, the unconstrained total transportation cost function in Equation (1) is to be minimized for solving the two-dimensional Euclidean MSSFLP. Note that its one- and three-dimensional counterparts may be represented, respectively, by Equations (2) and (3) below:

$$f(x) = \sum_{i=1}^n w_i | a_i - x |, \quad \dots \quad (2)$$

$$f(x, y, z) = \sum_{i=1}^n w_i [(a_i - x)^2 + (b_i - y)^2 + (c_i - z)^2]^{1/2}. \quad \dots \quad (3)$$

To carry out a simulation study of the performance of the COG as a suboptimal solution to the Euclidean MSSFLP, a brief account of the experimental design is in order. First of all, the number of spatial dimensions ( $m$ ) is set at three different levels: 1, 2, and 3. In each dimensional space, the number of existing facilities ( $n$ ) in the Euclidean MSSFLP is set at 10 different levels: 5, 10, 15, 20, 25, 30, 35, 40, 45, and 50. For each number of existing facilities, the magnitude of problem parameter values ( $r$ ) is set at five different levels, 1, 2, 3, 4, and 5. In other words, the values assumed by such parameters as  $a_i$ ,  $b_i$  ( $a_i$  in a one-dimensional space and  $a_i, b_i, c_i$  in a three-dimensional space) as well as  $w_i$  will be taken from the range  $[0, 10^r - 1]$ ,  $r = 1, 2, \dots, 5$ . Finally, for each range of parameter values, the number of replications ( $s$ ) is 10. As such, the entire data set consists of  $(3)(10)(5)(10) = 1,500$  location problem instances.

For each of the  $(3)(10)(5) = 150$  distinct combinations of  $m, n$ , and  $r$ , the random number generation function built in MS Excel 2007 is employed to generate 10 sets of parameter values for  $a_i, b_i$  ( $a_i$  in a one-dimensional space and  $a_i, b_i, c_i$  in a three-dimensional space) as well as  $w_i$ , each characterizing a specific Euclidean MSSFLP. The computer program WinQSB (Chang and Desai, 2003) is then run to solve each of these 10 problems to find the MGP-based optimum  $S^* = (x^*, y^*)$  ( $(x^*)$  in a one-dimensional space and  $(x^*, y^*, z^*)$  in a three-dimensional space) along with the minimum total transportation cost  $f(S^*)$ . Next, the COG corresponding to each of the 10 location problems is computed based on Step 0 in the MGP and denoted by  $S$ . Plug  $S$  in the function defined in Equation (1) (Equation (2) in a one-dimensional space and Equation (3) in a three-dimensional space) and we obtain the suboptimal total transportation cost  $f(S)$ . The percent deviation of cost is represented by  $p\%$ , where

$$p = 100[f(S) - f(S^*)]/f(S^*). \quad \dots \quad (4)$$

To illustrate the procedure of computing the percent deviation, let us consider a Euclidean MSSFLP defined below with  $m = 3, n = 5$ , and  $r = 2$ :

$i$	$a_i$	$b_i$	$c_i$	$w_i$
1	96	40	65	16
2	23	55	93	52
3	31	96	81	60
4	45	49	70	13
5	1	78	32	24

It is easy to prove that  $S^* = (29.26, 75.50, 80.41)$  and  $f(S^*) = 5,531.53$ . In addition,  $S = (31.52, 71.33, 75.24)$  and  $f(S) = 16[(96 - 31.52)^2 + (40 - 71.33)^2 + (65 - 75.24)^2]^{1/2} + \dots + 24[(1 - 31.52)^2 + (78 - 71.33)^2 + (32 - 75.24)^2]^{1/2} = 5,635.76$ . It follows that  $p = 100(5,635.76 - 5,531.53)/5,531.53 \approx 1.88$  and the percent deviation is 1.88%.

The percent deviation for each of the other nine problem instances is determined in the same way and the average of them, which is termed the average percent deviation and denoted by  $\underline{d}$ , is calculated. This procedure is repeated for each of the other 149 three tuples ( $m, n, r$ ). All of the  $\underline{d}$ 's are summarized in Tables 1, 2, and 3 for one-dimensional (1D), two-dimensional (2D), and three-dimensional (3D) problems, respectively. Based on these, the mean percent deviation ( $d$ ) over all three types of location problems for every ( $m, n$ ) pair is computed and shown in Table 4. Likewise, those for ( $m, r$ ) and ( $n, r$ ) pairs can be respectively found in Tables 5 and 6.

**Table 1. Average percent deviation ( $\underline{d}$ ) - 1D.**

n	r					Row Mean
	1	2	3	4	5	
5	7.11	9.65	6.32	11.16	10.72	8.99
10	6.61	5.35	2.28	4.51	3.22	4.39
15	3.27	2.99	2.01	1.78	4.75	2.96
20	2.30	0.67	1.85	2.09	3.64	2.11
25	1.80	1.56	1.08	4.09	2.55	2.22
30	1.28	0.55	1.56	1.14	1.05	1.12
35	1.55	0.64	0.93	1.43	1.11	1.13
40	1.97	1.21	0.92	0.66	1.34	1.22
45	1.45	0.57	1.29	1.23	1.10	1.13
50	1.15	1.03	0.63	0.39	0.29	0.70
Column Mean	2.85	2.42	1.89	2.85	2.98	2.60

**Table 2. Average percent deviation ( $\underline{d}$ ) - 2D.**

n	r					Row Mean
	1	2	3	4	5	
5	2.60	5.70	4.35	6.35	3.67	4.53
10	3.80	3.06	0.89	2.07	2.59	2.48
15	1.61	0.53	1.16	0.84	0.99	1.03
20	0.60	0.75	0.99	1.32	0.72	0.88
25	0.71	1.48	0.77	0.87	0.60	0.89
30	0.60	0.96	0.50	0.38	0.50	0.59
35	0.52	0.59	0.40	0.63	0.34	0.50
40	0.47	0.45	0.53	0.63	0.49	0.51
45	0.26	0.61	0.37	0.37	0.61	0.44
50	0.21	0.30	0.33	0.56	0.15	0.31
Column Mean	1.14	1.44	1.03	1.40	1.07	1.22

**Table 3. Average percent deviation ( $\underline{d}$ ) - 3D.**

n	r					Row Mean
	1	2	3	4	5	
5	2.75	3.33	4.13	4.09	2.81	3.42
10	0.96	0.84	0.74	1.18	0.59	0.86
15	0.63	0.70	0.85	0.66	0.71	0.71
20	0.57	0.58	0.54	0.39	0.34	0.48
25	0.38	0.25	0.38	0.37	0.55	0.39
30	0.27	0.27	0.41	0.26	0.29	0.30
35	0.25	0.42	0.26	0.55	0.23	0.34
40	0.17	0.17	0.15	0.16	0.26	0.18
45	0.32	0.23	0.24	0.24	0.19	0.24
50	0.22	0.16	0.19	0.11	0.24	0.18
Column Mean	0.65	0.70	0.79	0.80	0.62	0.71

**Table 4. Mean percent deviation (d).**

n	m			Row Mean
	1	2	3	
5	8.99	4.53	3.42	5.65
10	4.39	2.48	0.86	2.58
15	2.96	1.03	0.71	1.57
20	2.11	0.88	0.48	1.16
25	2.22	0.89	0.39	1.17
30	1.12	0.59	0.30	0.67
35	1.13	0.50	0.34	0.66
40	1.22	0.51	0.18	0.64
45	1.13	0.44	0.24	0.60
50	0.70	0.31	0.18	0.40
Column Mean	2.60	1.22	0.71	1.51

**Table 5. Mean percent deviation (d).**

r	m			Row Mean
	1	2	3	
1	2.85	1.14	0.65	1.55
2	2.42	1.44	0.70	1.52
3	1.89	1.03	0.79	1.24
4	2.85	1.40	0.80	1.68
5	2.98	1.07	0.62	1.56
Column Mean	2.60	1.22	0.71	1.51

**Table 6. Mean percent deviation (d).**

n	r					Row Mean
	1	2	3	4	5	
5	4.15	6.23	4.93	7.20	5.73	5.65
10	3.79	3.08	1.30	2.59	2.13	2.58
15	1.84	1.41	1.34	1.09	2.15	1.57
20	1.16	0.67	1.13	1.27	1.57	1.16
25	0.96	1.10	0.74	1.78	1.23	1.17
30	0.72	0.59	0.82	0.59	0.61	0.67
35	0.77	0.55	0.53	0.87	0.56	0.66
40	0.87	0.61	0.53	0.48	0.70	0.64
45	0.68	0.47	0.63	0.61	0.63	0.60
50	0.53	0.50	0.38	0.35	0.23	0.40
Column Mean	1.55	1.52	1.24	1.68	1.55	1.51

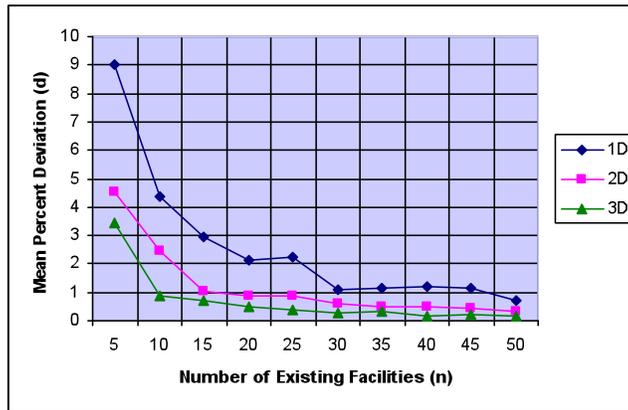
#### 4. RESULTS AND DISCUSSION

A close look at Tables 1, 2, and 3 reveals that all of the average percent deviations ( $\bar{d}$ 's) fall within 0.11% and 11.16% and most of them are quite small. In particular, for the 1D location problems, the overall average is 2.60% with the largest being 11.16% and the smallest being 0.29%. In comparison, the maximum and the minimum average percent deviations for the 2D cases are, respectively, 6.35% and 0.15% with an overall average of 1.22%. Lastly, the average percent deviations range from 0.11% to 4.13% and the overall average is 0.71% for the 3D problems. Taking into account the entire set of

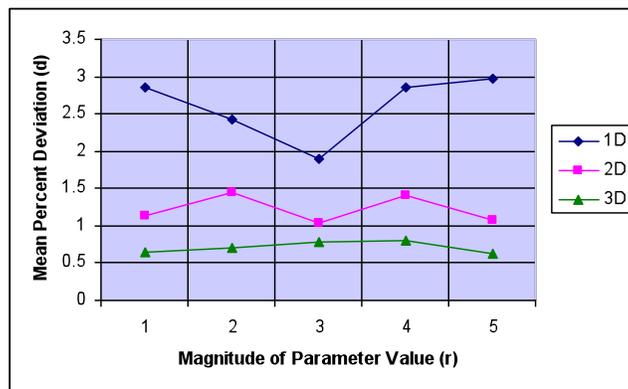
1,500 instances tested, the grand average percent deviation is  $\bar{d} = (2.60\% + 1.22\% + 0.71\%)/3 \approx 1.51\%$ , which can also be seen from any of Tables 4, 5, and 6.

Since locating the new facility at the COG in a Euclidean MSSFLP results in a total transportation cost higher than the minimum one by 1.51% on average, it is immediately concluded that the easy-to-calculate COG is an excellent substitute for the MGP-based optimal solution for practical purposes. Not only does this empirical finding lend support to the observations made by Krajewski et al. (2007), Martinich (1997), as well as Goetschalckx (2008), it is also very similar to the result reported by Kuo et al. (2002) with  $\bar{d} = 1.58\%$ .

To gain an in-depth insight into the outcome of the simulation analysis, we use the information contained in Tables 4, 5, and 6 to plot three charts to uncover possible relationships between the mean percent deviation ( $d$ ) and each of the following: the number of spatial dimensions ( $m$ ), the number of existing facilities ( $n$ ), and the magnitude of model parameters ( $r$ ). These are labeled as, respectively, Figures 1, 2, and 3.



**Figure 1. Mean percent deviation versus number of existing facilities.**



**Figure 2. Mean percent deviation versus magnitude of parameter value.**

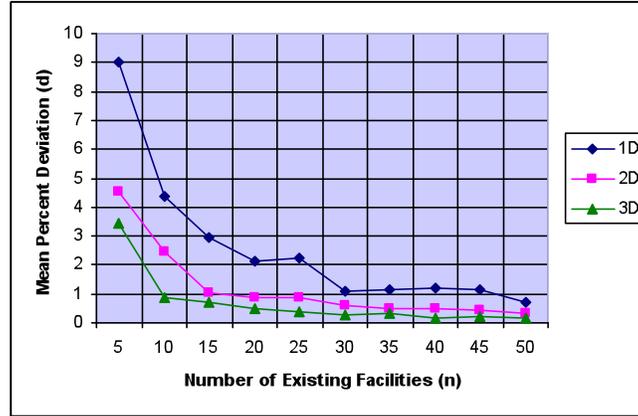


Figure 3. Mean percent deviation versus number of existing facilities.

It is evident from Figure 1 that the mean percent deviation over the five possible ranges of parameter values generally decreases as the number of existing facilities increases in each of the three dimensional spaces. Besides, it declines from 1D to 2D to 3D problems for any of the 10 possible numbers of existing facilities. These imply that the COG serves as a better approximate solution to the Euclidean MSSFLP in a higher dimensional space and when there are more existing facilities.

Figure 2 shows that, regardless of the magnitude of parameter values, the mean percent deviation over the 10 different numbers of existing facilities goes down when the number of spatial dimensions goes up. But the trend does not seem to exhibit any particular pattern from value range to value range. The implications are that the COG generally performs better as the number of dimensions becomes larger and that the magnitude of parameter values has little effect on the mean percent deviation.

Finally, in Figure 3, the mean percent deviation over the three dimensional spaces is a decreasing function of the number of existing facilities for each of the five different ranges of parameter values. However, the curves intertwine with each other and none of them is consistently above or below the others from one number of existing facilities to another. This seems to suggest that the mean percent deviation is a decreasing function of the number of existing facilities but it has little to do with the magnitude of parameter values.

### 5. REGRESSION ANALYSIS

In this section, we aim to perform multiple regression analysis to determine if the associations between  $d$  and  $m$  (negative),  $d$  and  $n$  (negative), as well as  $d$  and  $r$  (none) observed in the previous section may be generalized from the random sample of 1,500 problem instances used in the above simulation experiment to the entire population of Euclidean MSSFLPs. Toward that end, the average percent deviation ( $\underline{d}$ ) is a natural choice for the dependent variable  $Y$ . Since the independent variable for  $m$  is qualitative and encompasses three dimensional categories, it is replaced by two dummy variables  $X_1$  and  $X_2$  with 2D being the baseline. More specifically, we assign  $(X_1, X_2) = (1, 0)$  to 1D,  $(X_1, X_2) = (0, 0)$  to 2D, and  $(X_1, X_2) = (0, 1)$  to 3D. Aside from defining  $X_3$  as the independent variable corresponding to  $n$ , we introduce  $X_4$  as the independent variable for  $r$ .

Our choice of a nonlinear model is based on a careful review of the scatterplots of  $\underline{d}$  against each of  $m$ ,  $n$ , and  $r$  as well as the residual plots (Neter et al., 2004). In particular, the charts reveal a seemingly natural logarithmic relationship between  $Y$  and  $X_3$  while  $Y$  appears to have little to do with  $X_4$ . Two terms combining  $X_3$  with  $X_1$  and  $X_2$ , respectively, are also included in the analysis to capture any impact that the interactions between independent variables might have on the dependent variable while not introducing too much multicollinearity. All of these discussions lead to the proposal of the following equation:

$$\ln Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_3^{-1} + \beta_6 X_1 X_3^{-1} + \beta_7 X_2 X_3^{-1} + \epsilon, \quad \dots \quad (5)$$

where  $\beta_0$  is the intercept,  $\beta_i$  is the coefficient of the  $i^{\text{th}}$  term involving any independent variable in the regression model,  $i = 1, 2, \dots, 7$ , and the random error term  $\epsilon$  is assumed to follow a normal distribution with a mean of 0 and a standard deviation of  $\sigma$ .

Based on the information contained in Tables 1, 2, and 3 along with the multiple nonlinear regression model in Equation

(5), we develop an input data set comprising 150 entries in the form of  $(\ln Y, X_1, X_2, X_3, X_4, X_3^{-1}, X_1X_3^{-1}, X_2X_3^{-1})$ . An incomplete listing can be found in Table 7. We see from the SPSS output that the seven independent variables and combinations of independent variables collectively explain  $R^2 = 0.879 = 87.9\%$  of the variation in the dependent variable. The regression model is extremely significant at the 1% level, which is confirmed by  $F = 147.868$  ( $***p < 0.001$ ). However, in addition to the insignificant intercept  $\beta_0 = 0.000$  ( $+p > 0.100$ ), the following are found to have no statistically significant impact on  $Y$ :  $X_4$  ( $\beta_4 = -0.010$ ,  $+p > 0.100$ ),  $X_1X_3^{-1}$  ( $\beta_6 = -0.721$ ,  $+p > 0.100$ ), and  $X_2X_3^{-1}$  ( $\beta_7 = 1.572$ ,  $+p > 0.100$ ). After these three terms have been dropped from the model sequentially (Cohen et al., 2003), Equation (5) reduces to:

$$\ln Y = \beta_0' + \beta_1'X_1 + \beta_2'X_2 + \beta_3'X_3 + \beta_5'X_3^{-1} + \epsilon' \quad \dots \quad (6)$$

The primary results from another run of SPSS based on the new regression model in Equation (6) indicate that the validity of the revised model is well established by either the new  $R^2 = 0.877$  or the new  $F = 257.895$  ( $***p < 0.001$ ). As evidenced by  $\beta_1' = 0.799$  ( $***p < 0.001$ ),  $\beta_2' = -0.604$  ( $***p < 0.001$ ),  $\beta_3' = -0.025$  ( $***p < 0.001$ ), and  $\beta_5' = 8.676$  ( $***p < 0.001$ ), all of  $X_1, X_2, X_3$ , and  $X_3^{-1}$  are highly significant independent variables. In contrast, the intercept  $\beta_0' = -0.048$  ( $+p > 0.100$ ) is not helpful in predicting  $Y$  and, hence, will be excluded from further consideration. Since the variance inflationary factor (VIF) for each explanatory variable is relatively small ( $< 3.000$ ) in this case, the moderate multicollinearity does not present a serious problem here. In sum, the multiple nonlinear regression model sought is:

$$\ln Y = 0.799X_1 - 0.604X_2 - 0.025X_3 + 8.676X_3^{-1} \quad \dots \quad (7)$$

or

$$Y = \exp(0.799X_1 - 0.604X_2 - 0.025X_3 + 8.676X_3^{-1}) \quad \dots \quad (8)$$

**Table 7. Incomplete input data set for regression analysis.**

Y	lnY	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>3</sub> <sup>-1</sup>	X <sub>1</sub> X <sub>3</sub> <sup>-1</sup>	X <sub>2</sub> X <sub>3</sub> <sup>-1</sup>
7.11	1.9615	1	0	5	1	0.2000	0.2000	0
:	:	:	:	:	:	:	:	:
1.80	0.5878	1	0	25	1	0.0400	0.0400	0
:	:	:	:	:	:	:	:	:
3.80	1.3350	0	0	10	1	0.1000	0	0
:	:	:	:	:	:	:	:	:
0.47	-0.7550	0	0	40	1	0.0250	0	0
:	:	:	:	:	:	:	:	:
0.25	-1.3863	0	1	35	1	0.0286	0	0.0286
:	:	:	:	:	:	:	:	:
0.16	-1.8326	0	1	50	2	0.0200	0	0.0200

Note: Y = Average percent deviation; X<sub>1</sub> = Dummy for one-dimensional space; X<sub>2</sub> = Dummy for three-dimensional space; X<sub>3</sub> = Number of existing facilities; X<sub>4</sub> = Magnitude of parameter value.

Since the exponential function  $\exp(\cdot)$  is an increasing function, a close examination of the signs of the four regression coefficients in equation 8 suggests that the average percent deviation ( $\underline{d}$ ) in the two-dimensional Euclidean MSSFLPs ( $X_1 = 0$  and  $X_2 = 0$ ), which is the baseline, is smaller than that in one-dimensional problems ( $X_1 = 1$  and  $X_2 = 0$ ) but larger than that in three-dimensional cases ( $X_1 = 0$  and  $X_2 = 1$ ). This indicates that  $\underline{d}$  decreases as  $m$  increases. One can also tell that  $\underline{d}$  will go down as  $n$  goes up because  $-0.025X_3 + 8.676X_3^{-1}$  is a decreasing function of  $X_3$ . Also, the absence of  $X_4$  in (8) is a clear indication that the magnitude of such problem parameters as  $a_i, b_i, c_i$ , and  $w_i, i = 1, 2, \dots, n$ , is irrelevant to  $\underline{d}$ , which makes intuitive sense due to the normalization effect of the definition of percent deviation in Equation (4). Finally,  $\underline{d}$  approaches 0 when  $X_3$  approaches  $+\infty$  regardless of the values of  $X_1$  and  $X_2$ . In other words, the COG will eventually coincide with the MGP-based optimal location when the number of existing facilities ( $n$ ) is exceedingly large irrespective of the dimensional space involved. All of these agree with our findings obtained in the simulation experiment.

As an illustration of the usefulness of the regression model in (8), suppose that we have decided to locate the new facility at the COG in a collection of randomly generated three-dimensional Euclidean MSSFLPs with 20 existing facilities each. With  $m = 3$  (i.e.,  $X_1 = 0$  and  $X_2 = 1$ ) and  $n = 20$  (i.e.,  $X_3 = 20$ ), the average percent deviation from the minimum total

transportation cost may be estimated as  $Y = \exp[0.799(0) - 0.604(1) - 0.025(20) + 8.676(20)^{-1}] = \exp(-0.6702) \approx 0.52$ , which falls between the smallest  $\underline{d}$  (0.34) and the largest  $\underline{d}$  (0.58) in the fourth row of Table 3 and is reasonably close to the row mean of 0.48. Consider another example in which the COG is to be used as an approximate solution to a set of two-dimensional Euclidean MSSFLPs randomly generated with 60 existing facilities each. Since  $m = 2$  (i.e.,  $X_1 = 0$  and  $X_2 = 0$ ) and  $n = 60$  (i.e.,  $X_3 = 60$ ), the average percent deviation of transportation cost is predicted to be  $Y = \exp[0.799(0) - 0.604(0) - 0.025(60) + 8.676(60)^{-1}] = \exp(-1.3554) \approx 0.26$ . This is slightly lower than the row mean of 0.31 in the last row of Table 2 with  $n = 50$  and is consistent with our expectation because, as shown previously, the average percent deviation is a decreasing function of the number of existing facilities.

## 6. INSIGHTS FOR PRACTITIONERS

We have demonstrated in this paper that the COG in any Euclidean MSSFLP can be found very efficiently whereas it is not only tedious but also time-consuming to obtain the MGP-based optimal solution. Besides, locating a new facility at the COG leads to a total transportation cost that is within 1.51% of the minimum cost. In view of the trade-off between computational effort and economic efficiency, a facility planner will obviously choose to use the COG as a heuristic approximation to the optimality for practical purposes. We have also shown that the COG-based transportation cost is getting even closer to the true minimum when the number of existing facilities grows. In addition, the difference between them could be predicted with high accuracy based on a nonlinear function. This knowledge provides a quick-and-dirty approach to estimating the deviation when a location decision needs to be made in a small time window without the aid of a computer.

The practical usefulness of the above-mentioned results is evidenced by the fact that a broad spectrum of real-world problems in the human society may be formulated as one-, two-, or three-dimensional Euclidean MSSFLPs and solved (Hale, 2000; Cera and Ortega, 2002; Davalbhakta and Sule, 2010). In particular, a large number of successful case studies have been reported in the location science literature. These include, but are not limited to, (1) planning for delivery of hot meals to the home-bound infirm and elderly to ensure adequate nutrition as well as independent living in Allegheny County in Pennsylvania (Johnson et al., 2002), (2) determining the optimal route and stops of the light rail transit system in the City of Maebashi, Japan, with an aim to minimize the total cost (Kawasaki and Kishimoto, 2005), (3) building a hospital in the Negev of southern Israel to meet the medical needs of about 400,000 immigrants (Mehrez et al., 1996), and (4) siting health care facilities in the central region of Ghana to improve patient access to them without requiring additional resources (Kissah-Korsah, 2008). It is conceivable that each of these and other projects involving Euclidean distances would have been completed more cost effectively had the center of gravity been used as a surrogate for the optimal location. Thus, the findings from this study have tremendous managerial implications for practitioners who are faced with the challenge of locating a manufacturing, service, or government facility.

## 7. CONCLUDING REMARKS

Some researchers have noted that the COG is often very close to the MGP-based optimum in the Euclidean MSSFLP so that the former may serve as a good substitute for the latter to increase computational efficiency in practice. This paper is motivated by the lack of theoretical or empirical evidence in the location literature to substantiate the observation. We set out to design and conduct a comprehensive simulation experiment involving a large number of problem instances with a wide variety of characteristics. The results show that using the COG as an approximate solution leads to a deviation of 1.51% from the minimum total transportation cost on average. Furthermore, the performance of the COG is a decreasing function of both the number of spatial dimensions and the number of existing facilities in the location problem.

It is worth pointing out that our findings generally echo those reported by Kuo et al. (2002). However, we go beyond the scope of their study and seek to statistically test the aforementioned associations between the average percent deviation and the key parameters in the Euclidean MSSFLP by taking a regression analysis approach. The final nonlinear model obtained not only confirms the relationships uncovered in the simulation exercise, but also allows facility managers to assess the trade-off between increase in transportation expenses and savings in computational costs as a consequence of locating the new facility at the COG.

A number of interesting issues remain to be addressed. For example, are there any other independent variables or interaction terms that should be introduced in the multiple nonlinear regression model in Equation (5) to help explain the  $100\% - 87.9\% = 12.1\%$  of variations in the dependent variable unaccounted for? Can the theoretical worst-case performance bound of the average percent deviation be derived? In other words, what is the largest possible gap between the COG-induced total transportation cost and the MGP-based minimum total transportation cost in any Euclidean MSSFLP? What are the implications of this research for other fields of study where the concept of COG has long been used such as military science (Janiczek, 2007)? Lastly, it would lend more credence to the results from this research if they could be validated by using industry data. These challenging problems will be part of our future investigation, and their solutions

should help us gain a deeper understanding of the role of COG in location planning as well as other disciplines.

## 8. REFERENCES

1. Brandeau, M. L. and Chiu, S. S. (1989). An Overview of Representative Problems in Location Research. *Management Science*, 35: 645-674.
2. Brimberg, J. and Love, R. F. (1998). Solving a Class of Two-Dimensional Uncapacitated Location-Allocation Problems by Dynamic Programming. *Operations Research*, 46: 702-709.
3. Cera, M. and Ortega, F. A. (2002). Locating the Median Hunter Among n Mobile Prey on the Plane. *International Journal of Industrial Engineering - Theory, Applications and Practice*, 9: 6-15.
4. Chan, Y. (2005). *Location, Transport and Land-Use: Modeling Spatial-Temporal Information*. Springer, New York, United States.
5. Chandrasekaran, R. and Tamir, A. (1989). Open Questions Concerning Weiszfeld's Algorithm for the Fermat-Weber Location Problem. *Mathematical Programming*, 44: 293-295.
6. Chang, S.-H. and Egbelu, P. J. (1997). Relative Pre-Positioning of Storage/Retrieval Machines in Automated Storage/Retrieval System to Minimize Expected System Response Time. *IIE Transactions*, 29: 313-322.
7. Chang, Y.-L. and Desai, K. (2003). *WinQSB (Second Edition)*. Wiley, New York, United States.
8. Cohen, J., Cohen, P., West, S. G. and Aiken, L. S. (2003). *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (Third Edition)*. Erlbaum Associates, Mahwah, New Jersey, United States.
9. Daskin, M. S. and Dean, L. K. (2004). Location of Health Care Facilities. In *Operations Research and Health Care: A Handbook of Methods and Applications* (Eds.: M. L. Brandeau, F. Sainfort and W. P. Pierskalla). Springer, New York, United States, pp. 43-76.
10. Davalbhakta, A. and Sule, D. R. (2010). A New Method to Solve Weighted Minimax Single Facility Location Problem with Euclidean Distance. *International Journal of Industrial Engineering - Theory, Applications and Practice*, forthcoming.
11. Drezner, Z. (1995). A Note on Accelerating the Weiszfeld Procedure. *Location Science*, 3: 275-279.
12. Drezner, Z. (Editor) (1995). *Facility Location: A Survey of Applications and Methods*. Springer, New York, United States.
13. Drezner, Z. and Hamacher, H. W. (Editors) (2002). *Facility Location: Applications and Theory*. Springer, New York, United States.
14. Egbelu, P. J. and Wu, C. T. (1998). Relative Positioning of a Load Extractor for a Storage Carousel. *IIE Transactions*, 30: 301-317.
15. Francis, R., McGinnis, L., Jr. and White, J. (1983). Locational Analysis. *European Journal of Operational Research*, 12: 220-252.
16. Francis, R., McGinnis, L., Jr. and White, J. (1992). *Facility Layout and Location: An Analytical Approach (Second Edition)*. Prentice Hall, Upper Saddle River, New Jersey, United States.
17. Goetschalckx, M. (2008). Distribution System Design. In *Logistics Engineering Handbook* (Ed.: G. D. Taylor). CRC, Boca Raton, Florida, pp. 13/1-13/18.
18. Hale, T. S. and Hale, L. C. (2000). An Aggregation Technique for Location Problems with One-Dimensional Forbidden Regions. *International Journal of Industrial Engineering - Theory, Applications and Practice*, 7: 133-139.
19. Janiczek, R. M. (2007). A Concept at the Crossroads: Rethinking the Center of Gravity. *US Army War College*, Carlisle, Pennsylvania, United States.
20. Johnson, M. P., Gorr, W. L. and Roehrig, S. F. (2002). Location/Allocation/ Routing for Home-Delivered Meals Provision: Models and Solution Approaches. *International Journal of Industrial Engineering - Theory, Applications and Practice*, 9: 45-56.
21. Kawasaki, S. and Kishimoto, T. (2005). Study on Optimal Location of Stop and Route of Urban Public Transportation. *Regional Community Facilities Planning*, 23: 1-8.
22. Kissah-Korsah, K. (2008). Spatial Accessibility to Health Care Facilities in the Ajumako-Enyan-Essiam and Upper Denkyira Districts in the Central Region of Ghana. *Norwegian Journal of Geography*, 62: 203-209.
23. Kouvelis, P. and Chiang, W.-C. (1992). A Simulated Annealing Procedure for Single Row Layout Problems in Flexible Manufacturing Systems. *International Journal of Production Research*, 30: 717-732.
24. Krajewski, L. J., Ritzman, L. P. and Malhotra, M. K. (2007). *Operations Management: Process and Value Chains (Eighth Edition)*. Prentice Hall, Upper Saddle River, New Jersey, United States.
25. Kuhn, H. W. (1967). On a Pair of Dual Non-Linear Problems. In *Non-linear Programming* (Ed.: J. Abadie). North-Holland, New York, United States, pp. 38-54.
26. Kuo, C.-C. and White, R. E. (2004). A Note on the Treatment of the Center-of-Gravity Method in Operations Management Textbooks. *Decision Sciences Journal of Innovative Education*, 2: 219-227.

27. Kuo, C.-C., White, R. E. and Chiang, W. K. (2002). On the Treatment of the Center-of-Gravity Method in the Existing Operations Management Textbooks. *Proceedings of the Annual Meeting of Decision Sciences Institute*, pp. 1957-1960.
28. Love, R., Morris, J. and Wesolowsky, G. (1988). *Facilities Location: Models & Methods*. North-Holland, New York, United States.
29. Lucas, M. T. and Chhajer, D. (2004). Applications of Location Analysis in Agriculture: A Survey. *Journal of the Operational Research Society*, 55: 561-578.
30. Martinich, J. (1997). *Production and Operations Management: An Applied Modern Approach*. Wiley, New York, United States.
31. Mehrez, A., Sinuany-Stern, Z., Arad-Geva, T. and Binyamin, S. (1996). On the Implementation of Quantitative Location Models: The Case of a Hospital in a Rural Region. *Journal of the Operational Research Society*, 47: 612-625.
32. Mirchandani, P., Kohli, R. and Tamir, A. (1996). Capacitated Location Problems on a Line. *Transportation Science*, 30: 75-80.
33. Narasimhan, R., Talluri, S., Sarkis, J. and Ross, A. (2005). Efficient Service Location Design in Government Services: A Decision Support System Framework. *Journal of Operations Management*, 23: 163-178.
34. Neter, J., Kutner, M. H., Nachtsheim, C. J. and Wasserman, W. (2004). *Applied Linear Regression Models (Fourth Edition)*. McGraw-Hill, New York, United States.
35. Nickel, S. and Puerto, J. (2005). *Location Theory: A Unified Approach*. Springer, New York, United States.
36. Ostresh, L. M. (1978). On the Convergence of a Class of Iterative Methods for Solving the Weber Location Problem. *Operations Research*, 26: 597-609.
37. Price, W. L. and Turcotte, M. (1986). Locating a Blood Bank. *Interfaces*, 16: 17-26.
38. Sherali, H. D., Pendyala, C. M. and Rappaport, T. S. (1996). Optimal Location of Transmitters for Micro-Cellular Radio Communication System Design. *IEEE Journal on Selected Areas of Communications*, 14: 662-673.
39. Singhal, J. (1990). Two-level Hierarchical Transportation Networks: A New Set of Problems and Practical Applications. *Decision Sciences*, 21: 171-182.
40. Sivazlian, B. D. and Stanfel, L. E. (1975). *Analysis of Systems in Operations Research*. Prentice-Hall, Englewood Cliffs, New Jersey, United States.
41. Steenhuis, H.-J. and De Bruijn, E. J. (2004). Assessing Manufacturing Location. *Production Planning and Control*, 15: 786-795.
42. Suzuki, T., Asami, Y. and Okabe, A. (1991). Sequential Location-Allocation of Public Facilities in One- and Two-Dimensional Space: Comparison of Several Policies. *Mathematical Programming*, 52: 125-146.
43. Taylor, G. D. (Editor) (2008). *Logistics Engineering Handbook*. CRC, Boca Raton, Florida, United States.

### BIOGRAPHICAL SKETCH

---



Ching-Chung Kuo is an associate professor of management in the College of Business at the University of North Texas. He received his Ph.D. in industrial engineering and management sciences from Northwestern University. His work has appeared in such journals as *Applied Mathematical Modelling*, *Decision Sciences*, *European Journal of Operational Research*, *International Journal of Advanced Manufacturing Technology*, *International Journal of Production Research*, and *Omega*. He co-received a Best Theoretical/Empirical Research Paper Award from the Decision Sciences Institute, of which he is a member. He is also affiliated with the Institute for Operations Research and the Management Sciences as well as the Institute for Supply Management.

---