

## A NOTE ON THE WAREHOUSE LOCATION PROBLEM WITH DATA CONTAMINATION

XUEHONG GAO<sup>1,2,\*</sup> AND CAN CUI<sup>3</sup>

**Abstract.** To determine the optimal warehouse location, it is usually assumed that the collected data are uncontaminated. However, this assumption can be easily violated due to the uncertain environment and human error in disaster response, which results in the biased estimation of the optimal warehouse location. In this study, we investigate this possibility by examining these estimation effects on the warehouse location determination. Considering different distances, we propose the corresponding estimation methods for remedying the difficulties associated with data contamination to determine the warehouse location. Although data can be contaminated in the event of a disaster, the findings of the study is much broader and applicable to any situation where the outliers exist. Through the simulations and illustrative examples, we show that solving the problem with center of gravity lead to biased solutions even if only one outlier exists in the data. Compared with the center of gravity, the proposed methods are quite efficient and outperform the existing methods when the data contamination is involved.

**Mathematics Subject Classification.** 90B25.

Received December 23, 2019. Accepted March 10, 2021.

### 1. INTRODUCTION

After large-scale disasters such as earthquakes, floods, tropical storms, and nuclear leaks, many people are affected and injured severely. In the event of such a disaster, rapid and effective responses to a disaster are required to relieve the detrimental situation [8, 16, 33, 35, 37, 38, 69]. Among them, the most expeditious response is to help all those injured people who require medical attendance through life-saving operations [7]. In this case, a number of temporary medical service centers that are considered as demand points are needed to be quickly identified and set up to meet the urgent needs of medical services for reducing mortality and preventing health deterioration from evacuees [2]. To guarantee reliability and avoid an unbalanced medical supply distribution, a warehouse is needed to be set up to stock the medical supplies in advance and distribute the medical supplies from the warehouse to the demand points possibly multiple times.

To determine the optimal warehouse location with respect to the objective function, it is usually assumed that the collected data are uncontaminated. However, the previous conventional methods may not work properly

---

*Keywords.* Facility location problem, robust, center of gravity, weighted median.

<sup>1</sup> Research Institute of Macro-Safety Science, School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing, P.R. China.

<sup>2</sup> Department of Industrial Engineering, Pusan National University, Busan, Republic of Korea.

<sup>3</sup> Viewshare Technology Co., Ltd, Beijing, P.R. China.

\*Corresponding author: [gao2016@pnu.edu](mailto:gao2016@pnu.edu), [gaoxh2020@ustb.edu.cn](mailto:gaoxh2020@ustb.edu.cn)

when the data are collected under volatile operating conditions such as uncertain environment and human error in disaster response. Once this assumption is violated in a serious manner, the estimated optimal warehouse location may be located far away from the true optimal location. Thus, a statistically robust approach that is less sensitive to the outliers is clearly warranted in this study.

To address the aforementioned questions and research gap, this study focuses on the warehouse location problem under data contamination in disaster response. This study considers three different methods for three distances (norms), namely, Manhattan distance, Euclidean distance, and squared Euclidean distance, to determine the optimal warehouse locations. The comparison results showed that the proposed location estimation methods outperformed the CG method both in accuracy and robustness. Then a simulation study and several illustrative examples are conducted to present that the proposed location estimation methods outperformed the CG method in determining the optimal warehouse locations under data contamination.

The rest of this study is organized as follows. Section 2 reviews the previous studies including the facility location problems and the methodologies to solve the problem, and then we highlight the main differences and contributions from previous studies. In Section 3, the methodologies used to determine the optimal warehouse location are described. Then extensive Monte Carlo simulations are carried out in Section 4 and an illustrative case study from a benchmark is provided in Section 5. Finally, Section 6 summarizes this study and discusses potential future works.

## 2. LITERATURE REVIEW

The facility location problem is an important issue in supporting the physical distribution as it contributes significantly to the travel time or cost in logistics systems [9, 20, 34, 39]. Because the travel time or cost can be analyzed by discrete and continuous aspects in the space, the facility location problem is commonly divided into two categories; namely, discrete-space and continuous-space facility location problems. For the discrete-space facility location problem, the location of a facility can be only located at one of the specific points. On the other hand, for the continuous-space facility location problem, the facility is allowed to be located anywhere in the planning area [4]. Thus, we briefly review those two categories of facility location problems and introduce the methodologies used for determining the facility location. Finally, the necessity and novelty of this study are discussed and presented.

In the discrete-space facility location problem, the facility location needs to be selected from a pre-specified set of potential locations, typically the nodes of a network. Several studies [1, 11, 19, 23, 41, 45, 49] have been developed to address this issue. Unfortunately, because the facility can be only located at one of the specific points, the optimal discrete facility locations are worse compared to the optimal continuous solution for a given metric. For the continuous-space facility location problem, anywhere in the plane can be considered as the facility location. Thus, there is an infinite number of potential locations for the facility. So far, the continuous-space facility location problem has received much attention in the literature and many studies have been developed to identify the continuous-space facility location from different perspectives [28, 48, 51, 52].

To determine the facility location, many methods have been developed to determine the single-facility location [22, 46]. Among them, the center-of-gravity (CG) method is one of the most widely used approaches for its simplicity in previous studies [30, 53] as it can be deemed the optimal facility location. In the studies using the CG methods, the goals are to minimize an objective function involving squared Euclidean distance, Euclidean distance, or Manhattan distance between the facility and demand points. For instance, Ohsawa [55] aimed to determine the CG so that the average squared Euclidean distance can be minimized from the facility to the demand points, but he did not derive any distance-related travel time or cost.

Besides, the CG method is also generally applied in the clustering-based multi-facility location problems, where the CG is severed as the optimal facility location within each cluster for its simplicity. Esnaf and Küçükdeniz [27] applied Euclidean distance with a correction factor to determine the optimal plant location within each cluster, and they used the weighted mean to obtain the CG. Nadizadeh *et al.* [54] used the mean to obtain the CG that was considered as the location of the depot to minimize the total Euclidean distance

to customers. Unfortunately, the (weighted) mean is not the optimal solution to minimize the total (weighted) Euclidean distance between facilities. Besides, Gao *et al.* [36] developed a clustering-based genetic algorithm for a multi-facility location problem, where the weighted mean (CG) was served as the optimal location when the weighted Manhattan distance is used within each cluster. However, the weighted mean does not minimize the total weighted Manhattan distance between facilities. The above three studies are not appropriate because they used the same (weighted) mean to obtain the CG but under different distances. In urban areas, due to the street configurations, the distance is usually modeled as  $L^1$  norm (Manhattan distance), while in rural areas the travel is faster, the distance is usually modeled as  $L^p$  norm with  $p > 1$  [70]. Thus, in this study, we focus on identifying the optimal facility locations according to their specific distances.

Against the uncertain information after a disaster, there has been a great deal of research effort to develop robust optimization models in their facility location problems. Most robust optimization models use different scenarios ranging from the most optimistic to the worst case. For more detailed information on the scenario-based robust optimization models, the readers are referred to previous studies, including Baron *et al.* [10], Ashtiani *et al.* [5], Gülpınar *et al.* [42], De Rosa *et al.* [25] and An *et al.* [3] among others. In many practical problems, the data can often be collected under volatile operating conditions [57, 58], which results in contaminated data and variability in parametric value in disaster response. Note that one possible reason for the existence of an outlier is that an observation is contaminated. Another possible reason for the existence of an outlier is that the underlying distribution of the sample has fat tails so that the observation is not contaminated but just an extreme observation from one of the heavy-tailed ends of the underlying distribution. For more details on the outliers and their effect on the estimator behavior, one should refer to Tukey [67], Rousseeuw and Croux [63], Park [60] and Gao and Jin [34]. This data contamination and variability in parametric value have been receiving insufficient attention in disaster response. Under such situations, the scenario-based robust optimization models could be inappropriate or lead to infeasible solutions. It is required to obtain a robust solution when the data contamination is involved in the facility location problem.

To address this challenge, a statistically robust approach that is less sensitive to the outliers or noises is clearly warranted. Hence, in this paper, we develop several location estimation methods for different distances so that the optimal warehouse location can be determined against outliers or noises. Besides, the facility location problem with data contamination has never received any attention in the previous studies. As a consequence, it is necessary to fill this research gap by developing several reasonable location estimation methods according to their corresponding distances in this continuous-space warehouse location problem with data contamination. In particular, given three different distances, namely Manhattan distance, Euclidean distance, and squared Euclidean distance, we develop the corresponding location estimation methods so that the optimal warehouse location can be obtained. We also compared their performance with previous methods to illustrate the superiority. Finally, extensive Monte Carlo simulation studies are carried out and several illustrative examples are provided.

### 3. METHODOLOGY

In this section, we develop the warehouse location estimation methods based on three different distances. Here, there are two common assumptions that need to be declared: (i) the demand points with different priority levels (weights) and locations are known, and (ii) the objective function is linearized with the distance. Before the methods are presented, the notations used in the proposed model are given as follows.

#### *Parameters*

- $n$  Number of demand points ( $i = 1, 2, \dots, n$ ).
- $X_i$   $X$ -coordinate of demand point  $i$ .
- $Y_i$   $Y$ -coordinate of demand point  $i$ .
- $W_i$  Weight (Priority level) of demand point  $i$ .

### Decision variables

- $u$   $X$ -coordinate of the warehouse.
- $v$   $Y$ -coordinate of the warehouse.

In the continuous-space facility location problem, three different distances are usually applied, Manhattan distance ( $L^1$  norm), Euclidean distance ( $L^2$  norm), and squared Euclidean distance (squared  $L^2$  norm). These three distances can easily be computed using geographic coordinates. The objective function is to minimize the total weighted distance-related travel time or cost. We derive those warehouse location estimation methods for the above three different distances into models 1, 2, and 3 in this study. Before developing the formulation for the warehouse-location models, we normalize the weights first using the following equation

$$P_i = \frac{W_i}{\sum_{i=1}^n W_i}, \quad (3.1)$$

where  $P_i$  is the normalized weight of demand point  $i$ .

### 3.1. Model 1 ( $L^1$ norm)

When the Manhattan distance is considered in the model, the objective is to minimize the weighted  $L^1$  norm, which is given by

$$\text{Obj}_1 = \sum_{i=1}^n P_i (|X_i - u| + |Y_i - v|), \quad (3.2)$$

where  $|X_i - u| + |Y_i - v|$  is the distance measured along the axis at right angles, which is usually applied in the urban area of interest due to the street configurations [21, 37]. For instance, the determination of the bicycle-sharing stations or the ambulance stations needs to consider the Manhattan distance. Let  $(\tilde{u}, \tilde{v})$  be the optimal warehouse location based on the  $\text{Obj}_1$ . Then we have

$$(\tilde{u}, \tilde{v}) = \underset{(u,v)}{\text{argmin}} (\text{Obj}_1). \quad (3.3)$$

The minimizer of  $\text{Obj}_1$ , denoted by  $(\tilde{u}, \tilde{v})$ , can be obtained by the following estimating equations that need to be solved for  $u$  and  $v$  [44].

$$\frac{\partial \text{Obj}_1}{\partial u} = \sum_{i=1}^n P_i \text{sgn}(X_i - u) = 0 \quad (3.4)$$

$$\frac{\partial \text{Obj}_1}{\partial v} = \sum_{i=1}^n P_i \text{sgn}(Y_i - v) = 0. \quad (3.5)$$

Then the optimal values  $\tilde{u}$  and  $\tilde{v}$  can be calculated separately. It is easily seen that  $\tilde{u}$  is the weighted median of the  $x$ -axis observations and  $\tilde{v}$  is the weighted median of the  $y$ -axis observations, which will be detailed later. Note that the weighted median was first suggested by Edgeworth [26] and since then it has been widely used in many applications [68]. As an illustration, we briefly introduce the conventional median. Then we introduce the weighted median. As the values  $\tilde{u}$  and  $\tilde{v}$  can be obtained separately, we consider the weighted median for the  $x$ -axis observations. Then the weighted median for the  $y$ -axis observations is easily obtained using the same method. Because the median can be obtained by using the empirical cumulative distribution function, we briefly introduce the definition of the empirical cumulative distribution function which was introduced in Definition 2.1 of Owen [59]. Then we propose a new definition that can consider the weights.

**Definition 3.1.** Given a set of observations  $X_1, X_2, \dots, X_n$ , the empirical cumulative distribution function  $F_n$  is defined as

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(X_i \leq x), \quad x \in \mathbb{R}, \quad (3.6)$$

where  $I(A)$  represents the indicator function defined as

$$I(A) = \begin{cases} 1 & \text{if } A \text{ is satisfied} \\ 0 & \text{otherwise} \end{cases}. \quad (3.7)$$

Csorgo [24] defined the sample quantile function (inverse cumulative distribution function) as below.

$$F_n^{-1}(p) = \inf \{x : F_n(x) \geq p\}. \quad (3.8)$$

Using this, the conventional median  $F_n^{-1}(1/2)$  is obtained as:

$$F_n^{-1}\left(\frac{1}{2}\right) = \inf \left\{x : F_n(x) \geq \frac{1}{2}\right\} = X_{(k)} \quad \text{if } \frac{k-1}{n} < \frac{1}{2} \leq \frac{k}{n}, \quad k = 1, 2, \dots, n, \quad (3.9)$$

where  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ .

**Definition 3.2.** Given a set of observations  $X_1, X_2, \dots, X_n$  with corresponding positive weights  $P_1, P_2, \dots, P_n$  such that  $\sum_{i=1}^n P_i = 1$ , we have the empirical cumulative distribution function  $G_n(x)$  with weights, which is defined as

$$G_n(x) = \sum_{i=1}^n P_i I(X_i \leq x). \quad (3.10)$$

Note that the above  $G_n$  includes the conventional empirical cumulative distribution function  $F_n$ , as a special case when  $P_i = 1/n$ . Similar to the definition of the sample quantile function in Csorgo [24], we define the sample quantile function with weights as below.

$$G_n^{-1}(p) = \inf \{x : G_n(x) \geq p\}. \quad (3.11)$$

Next, our goal is to obtain the weighted median for the  $x$ -axis observations, which is given as

$$G_n^{-1}\left(\frac{1}{2}\right) = \inf \left\{x : G_n(x) \geq \frac{1}{2}\right\} = X_{(k)} \quad \text{if } \sum_{j=1}^{k-1} P_{(j)} < \frac{1}{2} \leq \sum_{j=1}^k P_{(j)}, \quad k = 1, 2, \dots, n, \quad (3.12)$$

where  $P_{(j)}$  is the weight for  $X_{(j)}$ .

The above  $X_{(k)}$  minimizes the weighted norm.

### 3.2. Model 2 ( $L^2$ norm)

When the Euclidean distance is considered in this model, the objective is to minimize  $\text{Obj}_2$  that includes the weighted  $L^2$  norm, which is given by

$$\text{Obj}_2 = \sum_{i=1}^n P_i \sqrt{(X_i - u)^2 + (Y_i - v)^2}, \quad (3.13)$$

where  $\sqrt{(X_i - u)^2 + (Y_i - v)^2}$  is the straight-line distance between two points, which is widely applied in the rural area of interest, long-distance cases, and signal propagation cases [18, 27]. For instance, refer to the location determination of 4/5 G stations or plants within a large-scale area. The goal is to minimize  $\text{Obj}_2$  so that the optimal location  $(\check{u}, \check{v})$  is given by the minimizer of  $\text{Obj}_2$ , which is denoted by

$$(\check{u}, \check{v}) = \underset{(u,v)}{\operatorname{argmin}} (\text{Obj}_2). \quad (3.14)$$

Unlike the Manhattan distance, the optimal location based on the  $\text{Obj}_2$  is not in explicit form. Thus, we have to obtain it by using numerical methods. Here, we use the quasi-Newton method that is also called BFGS

[14, 29, 40, 65] to determine the optimal warehouse location so that the total weighted Euclidean distance can be minimized. For specific details on the BFGS method, the reader can be referred to Byrd *et al.* [15]. The BFGS algorithm is representative of the typical quasi-Newton method, which has been implemented in the R language by R core Team [66]. Note that the BFGS method is able to find the optimal warehouse location due to the convexity of  $\text{Obj}_2$  (for more detail, see Appendix A). Because of the convexity of  $\text{Obj}_2$ , we can obtain the global optimal location by solving the following system of equations:

$$\frac{\partial \text{Obj}_2}{\partial u} = \sum_{i=1}^n \frac{P_i (u - X_i)}{\sqrt{(X_i - u)^2 + (Y_i - v)^2}} = 0 \quad (3.15)$$

$$\frac{\partial \text{Obj}_2}{\partial v} = \sum_{i=1}^n \frac{P_i (v - Y_i)}{\sqrt{(X_i - u)^2 + (Y_i - v)^2}} = 0. \quad (3.16)$$

Then, the optimal location  $(\tilde{u}, \tilde{v})$  can be calculated by using numerical methods. In this study, we use the BFGS method to determine the warehouse location.

### 3.3. Model 3 (squared $L^2$ norm)

Next, we consider the squared Euclidean distance for the warehouse location problem. The objective is to minimize the weighted squared  $L^2$  norm, which is given by

$$\text{Obj}_3 = \sum_{i=1}^n P_i \left[ (X_i - u)^2 + (Y_i - v)^2 \right], \quad (3.17)$$

where  $(X_i - u)^2 + (Y_i - v)^2$  is the *squared value* of the straight-line distance between two points. The goal is to minimize  $\text{Obj}_3$  so that the optimal location  $(\tilde{u}, \tilde{v})$  is given by the minimizer of  $\text{Obj}_3$ , which is denoted by

$$(\tilde{u}, \tilde{v}) = \underset{(u,v)}{\text{argmin}} (\text{Obj}_3). \quad (3.18)$$

Taking the derivative of the  $\text{Obj}_3$  in (3.17) with respect to  $u$  and  $v$ , we have the following equations:

$$\frac{\partial \text{Obj}_3}{\partial u} = \sum_{i=1}^n 2P_i (u - X_i) = 0 \quad (3.19)$$

$$\frac{\partial \text{Obj}_3}{\partial v} = \sum_{i=1}^n 2P_i (v - Y_i) = 0. \quad (3.20)$$

The above system of equations is easily solved so that we obtain the optimal warehouse location

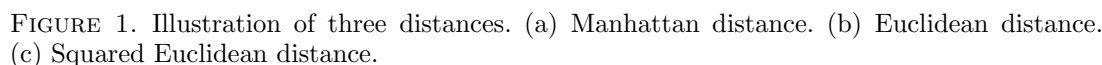
$$\tilde{u} = \sum_{i=1}^n P_i X_i \quad (3.21)$$

$$\tilde{v} = \sum_{i=1}^n P_i Y_i, \quad (3.22)$$

which are known as the weighted means. Note that the optimal warehouse location based on the weighted mean is also known as the CG because  $(\tilde{u}, \tilde{v})$  is the location of fulcrum when weights are located at the demand points.

Here, we provide three different distances in Figure 1. The Manhattan distance ( $D_1$ ) in Figure 1a is usually applied in urban areas due to the street configurations. As an illustration, the travel distance  $D_1(A, F)$  from points  $A$  to  $F$  via points  $B, C, D$ , and  $E$  in Figure 1a is the summation of street distances, which is given by

$$D_1(A, F) = D_1(A, B) + D_1(B, C) + D_1(C, D) + D_1(D, E) + D_1(E, F). \quad (3.23)$$


$$D_2(A, F) = D_2(A, B) + D_2(B, C) + D_2(C, D) + D_2(D, E) + D_2(E, F). \quad (3.24)$$
$$D_3(A, F) = D_2(A, F)^2. \quad (3.25)$$
$$\begin{aligned} D_3(A, F) &= D_3(A, C) + D_3(C, F) \\ &= D_3(A, B) + D_3(B, C) + D_3(C, E) + D_3(E, F) \\ &= D_3(A, B) + D_3(B, C) + D_3(C, D) + D_3(D, E) + D_3(E, F) \\ &= \dots \end{aligned} \tag{3.26}$$

Here, we summarize the properties of the above three distances and their optimal locations in Table 1. As shown in Table 1, the (weighted) median is able to minimize the total (weighted) Manhattan distance from the warehouse to the demand points. The (weighted) mean is able to minimize the total (weighted) squared Euclidean distance from the warehouse to the demand points. Unlike the (weighted) median and (weighted) mean, to minimize the total (weighted) Euclidean distance, the optimal warehouse location is not in explicit form so that we need to obtain it by using numerical methods. Unfortunately, many studies [17, 27, 36, 54, 56] applied the (weighted) mean to minimize the total (weighted) Manhattan distance or squared (weighted) Euclidean distance, which is unreasonable because the squared Euclidean distance is not a metric. In addition, the travel time or cost generally is linearized with Manhattan distance or Euclidean distance between two nodes [10], which



TABLE 1. Summary of optimal location estimation methods given different distances.

Objective function	Optimal location	Metric	Robust
Obj <sub>1</sub>	(Weighted) median	Yes	Yes
Obj <sub>2</sub>	Numerical method	Yes	Yes
Obj <sub>3</sub>	(Weighted) mean	No	No

TABLE 2. Probability distributions under consideration.

Case ID	$n$	$X_i$	$Y_i$	$P_i$
1	100	Normal (0,1)	Normal (0,1)	0.5
2	100	Normal (0,1)	Normal (0,1)	Beta (1,1)
3	100	$2 * \text{Beta}(1, 1) - 1$	$2 * \text{Beta}(1, 1) - 1$	0.5
4	100	$2 * \text{Beta}(1, 1) - 1$	$2 * \text{Beta}(1, 1) - 1$	Beta (1,1)
5	200	Normal (0,1)	Normal (0,1)	0.5
6	200	Normal (0,1)	Normal (0,1)	Beta (1,1)
7	200	$2 * \text{Beta}(1, 1) - 1$	$2 * \text{Beta}(1, 1) - 1$	0.5
8	200	$2 * \text{Beta}(1, 1) - 1$	$2 * \text{Beta}(1, 1) - 1$	Beta (1,1)

indicates that the (weighted) mean cannot be used to minimize the total travel time or cost as the squared  $L^2$  norm is not a metric. It is widely known that the median is a robust outlier-resistant location estimator, whereas the mean is not [43]. Consider the weights at demand points, we investigate the robust property of the weighted median and the other location estimation methods. Interestingly, we find that the robustness is an additional gain of using  $L^1$  and  $L^2$  norms. To validate the proposed corresponding location estimation methods, they are compared with the previous studies to illustrate their priorities in the accurateness and robustness.

#### 4. SIMULATION STUDY

In this section, a simulation study is carried out to show the effectiveness of the proposed location estimation methods. The proposed location estimation methods are compared with the conventional CG method. Regarding the weighted median, we also develop its R function as shown in Appendix B. All the simulations are performed on a computer with an Intel(R) Core(TM) i7-7700 CPU@3.6 GHz and 8 GB memory under Windows 10 Pro system. Those location estimation methods are implemented in R language, which is a non-commercial and open-source software for statistical computing and graphics originally developed by Ihaka and Gentleman [47].

In this simulation study, as shown in Table 2, we use eight cases with randomly generated demand-point locations and weights to obtain their optimal locations. To evaluate the performances of the proposed location estimation methods, we apply the efficiency [64] to measure the quality of the solution. Here we replicate each case 10 000 times, which results in 10 000 estimated locations for each of the distances.

Efficiency [64] is a method to measure the quality of an estimator in the experimental design. In this study, the statistical efficiency of an estimator is related to the variance of the estimator. The relative efficiency of the two procedures is the ratio of their variances, which is used as a metric for comparing the effectiveness of the two estimators. For more details, see Section 2.2 of Lehmann [50] and Park and Leeds [61]. The relative efficiency  $RE_M$  of  $\text{Obj}_1(u_{CG}, v_{CG})$  to  $\text{Obj}_1(u_w, v_w)$  with respect to the Manhattan distance is defined as

$$RE_M = \frac{\text{Var}[\text{Obj}_1(u_w, v_w)]}{\text{Var}[\text{Obj}_1(u_{CG}, v_{CG})]} \times 100\% \quad (4.1)$$



TABLE 3. Relative efficiency in different methods for eight cases.

Case ID	1	2	3	4	5	6	7	8
RE <sub>M</sub>	99.43%	99.24%	99.34%	99.13%	99.72%	99.62%	99.67%	99.56%
RE <sub>E</sub>	99.72%	99.63%	99.75%	99.66%	99.86%	99.82%	99.87%	99.83%

where  $(u_w, v_w)$  is the obtained warehouse location using weighted median and  $\text{Obj}_1(u_w, v_w)$  is the total weighted Manhattan distance from  $(u_w, v_w)$  to the demand points. And  $(u_{CG}, v_{CG})$  is the obtained warehouse location using the CG method and  $\text{Obj}_1(u_{CG}, v_{CG})$  is the total weighted Manhattan distance from  $(u_{CG}, v_{CG})$  to the demand points.

The relative efficiency RE<sub>E</sub> of  $\text{Obj}_2(u_{CG}, v_{CG})$  to  $\text{Obj}_2(u_B, v_B)$  with respect to the Euclidean distance is defined as

$$\text{RE}_E = \frac{\text{Var}[\text{Obj}_2(u_B, v_B)]}{\text{Var}[\text{Obj}_2(u_{CG}, v_{CG})]} \times 100\% \quad (4.2)$$

where  $(u_B, v_B)$  is the obtained warehouse location using the BFGS method and  $\text{Obj}_2(u_B, v_B)$  is the total weighted Euclidean distance from  $(u_B, v_B)$  to the demand points. And  $\text{Obj}_2(u_{CG}, v_{CG})$  is the total weighted Euclidean distance from  $(u_{CG}, v_{CG})$  to the demand points.

As shown in Table 3, the relative efficiencies are calculated and presented for Cases 1–8. Specifically, RE<sub>M</sub> is always smaller than 100%, which indicates that the (weighted) median is more reasonable to minimize the total Manhattan distance than the CG method. Besides, RE<sub>E</sub> is also smaller than 100% in Cases 1–8, which shows that the numerical method such as the BFGS method is able to find the optimal warehouse location to minimize the total Euclidean distance. As a consequence, the CG method is not preferred to determine the warehouse location as its worse performance.

As the data can be often collected under volatile operating conditions, which results in contaminated data and variability in parametric values in disaster response. To evaluate the robustness of the proposed optimal locations against an outlier, those eight cases with different outlier values are tested. We suppose that the outliers exist in the customer's (outlier) coordinates. To evaluate the robust performances of different methods, we run the simulation given different values to the outlier in Cases 1–4. We use  $M$  to represent a customer's (outlier) coordinates. The comparison results are presented in Figures 2–5, which clearly shows the robustness of the (weighted) median and BFGS methods in determining the optimal warehouse location, whereas the CG method is strongly sensitive to the outlier. It is also obvious that the optimal warehouse location coordinate shows a directly upward tendency with equal weights when  $M$  goes from 0 to 300 (see Figs. 2 and 4). Both Figures 3 and 5 also present general upward trends but with some fluctuations as the random weights are involved when  $M$  goes from 0 to 300.

In addition, to further illustrate their robust performance in determining the optimal warehouse location against different outlier numbers  $B$ , we replicate Cases 5–8 randomly for 50 times to avoid the extreme cluttering, which results in 50 estimated warehouse locations. Let  $M$  be 50 that is considered as the outlier value. Here, we consider three situations that have 0, 5, and 10 outliers in each of the cases. The optimal warehouse locations using different distances are plotted in Figures 6–9 where the red circles, the green triangles, and the blue squares are the optimal warehouse locations with contamination using  $L^1$ ,  $L^2$ , and squared  $L^2$  norms, respectively.

As shown in Figures 6–9, they provide a clear visual illustration of the outlier resistance behavior in three different distances. Note that Figures 6a–6c have different coordinate intervals, which is the same as Figures 7–9. It is obvious that the warehouse locations without considering data contamination are around the optimal location  $(x = 0, y = 0)$  for all the distances. However, with the data contamination, the warehouse locations obtained using squared  $L^2$  norm extreme shifts away from the true optimal location  $(x = 0, y = 0)$ . Conversely, those obtained optimal warehouse locations using  $L^1$  and  $L^2$  norms remain near the true optimal location. This

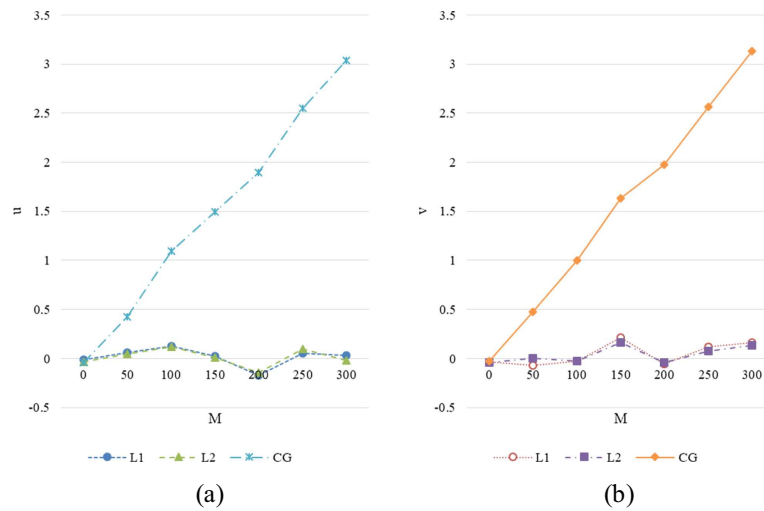


FIGURE 2. Warehouse locations in different methods given different outlier values in Case 1. (a)  $u$ . (b)  $v$ .

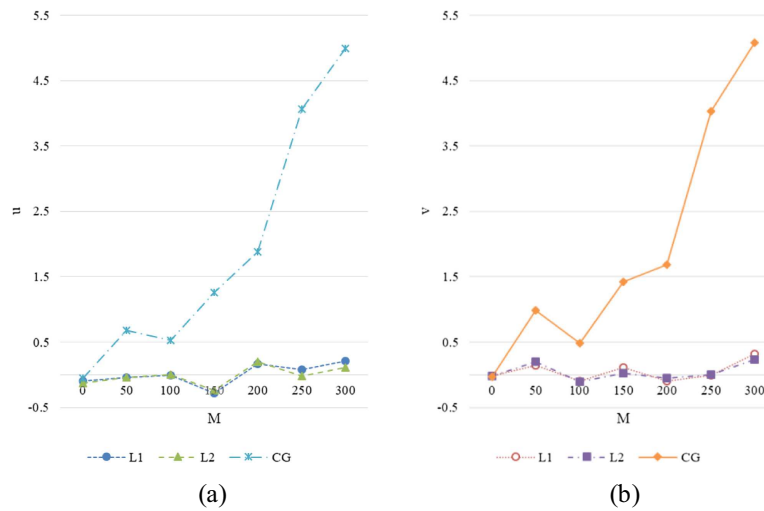


FIGURE 3. Warehouse locations in different methods given different outlier values in Case 2. (a)  $u$ . (b)  $v$ .

is a piece of definitely strong evidence to show that the proposed using  $L^1$  and  $L^2$  norms are much more reliable compared with the CG method.

## 5. ILLUSTRATIVE EXAMPLE

In addition to the simulation study, we also provide some illustrative examples from a benchmark to evaluate the performance of the proposed models. Here, we first introduce the set of benchmark instances that are used for the application of facility location in disaster management. We use these instances to (i) show the priorities of

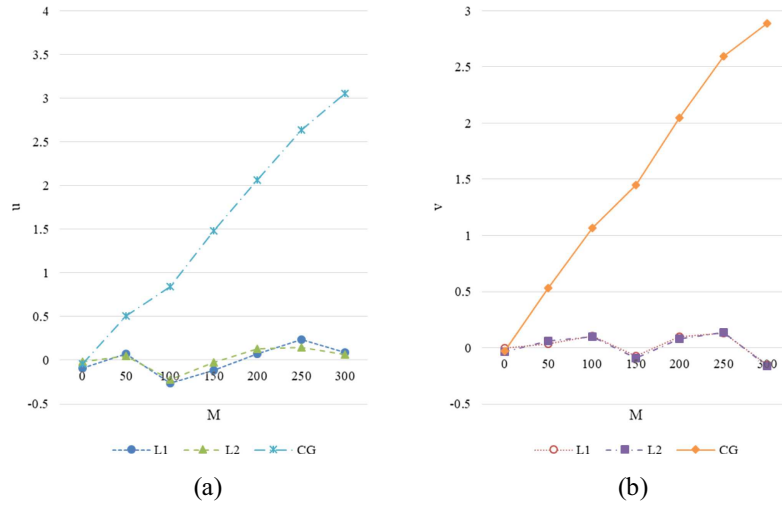


FIGURE 4. Warehouse locations in different methods given different outlier values in Case 3. (a)  $u$ . (b)  $v$ .

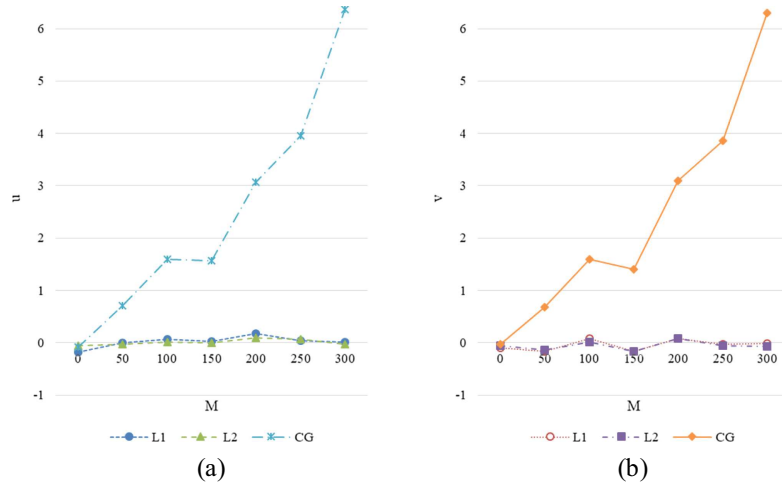


FIGURE 5. Warehouse locations in different methods given different outlier values in Case 4. (a)  $u$ . (b)  $v$ .

the proposed locations in accurateness given different distances (norms) and (ii) assess the robust performance compared with the conventional CG method.

To validate the effectiveness of the proposed models, we use a benchmark (<http://neo.lcc.uma.es/vrp/>) that is composed of 27 instances that are referred to Augerat *et al.* [6] (see Tab. 4). Each instance provides customers' (demand point) geographical coordinates and demand quantities. Here, the demand quantities are deemed weights. Given the coordinates and weights associated with each demand point, the optimal warehouse location can be obtained with the proposed location estimation methods. Note that we enlarged the last few customers' coordinates 10 times that are considered as outliers.

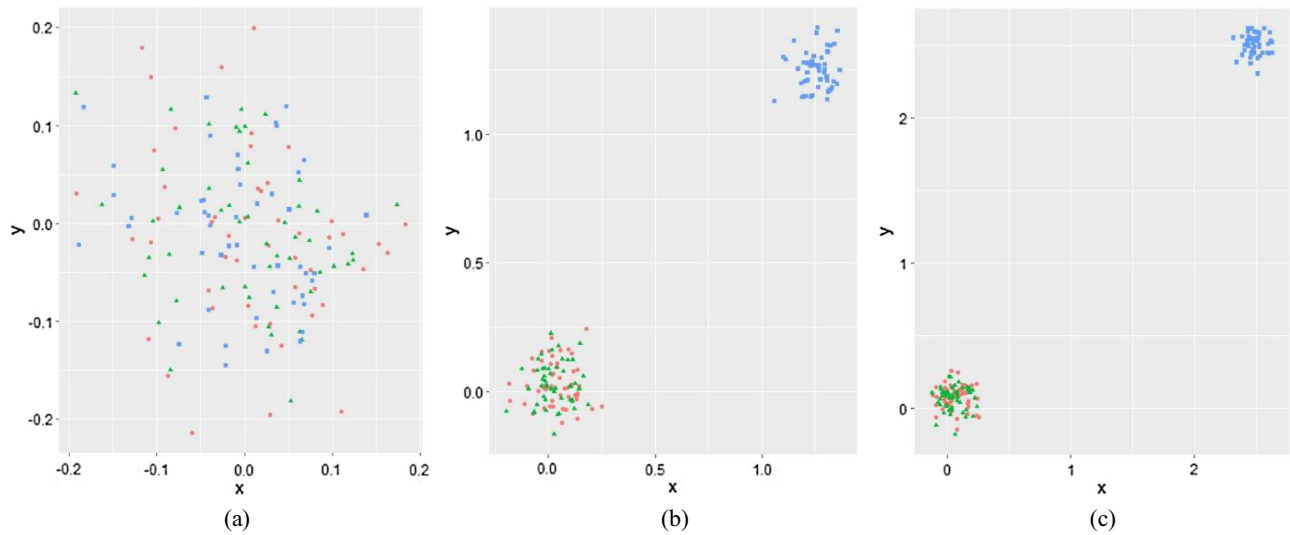


FIGURE 6. Optimal warehouse locations obtained using different norms in Case 5. (a)  $B = 0$ . (b)  $B = 5$ . (c)  $B = 10$ .

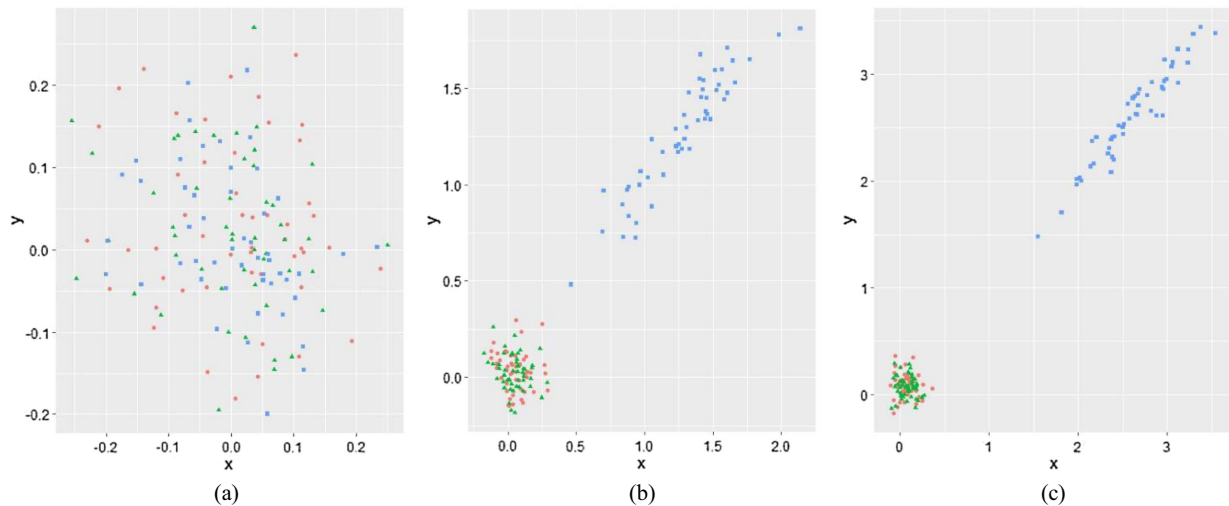


FIGURE 7. Optimal warehouse locations obtained using different norms in Case 6. (a)  $B = 0$ . (b)  $B = 5$ . (c)  $B = 10$ .

Here, we use Instances 1–18 to compare the proposed optimal locations given different distances. Since  $\text{Obj}_3$  is not a metric, there is no need to compare and calculate the squared Euclidean distance. The obtained warehouse location and the total travel distance are shown in Table 5. Specifically, it is obvious that the weighted median provides a smaller  $\text{Obj}_1$  value than the CG method in Instances 1–18, which indicates that the weighted median is more reasonable to minimize the total weighted Manhattan distance. Besides, the BFGS method also outperforms the CG method to minimize  $\text{Obj}_2$  value in Instances 1–18. As a consequence, the CG method is not preferred to determine the warehouse location as its worse performance in minimizing both  $\text{Obj}_1$  and  $\text{Obj}_2$ .

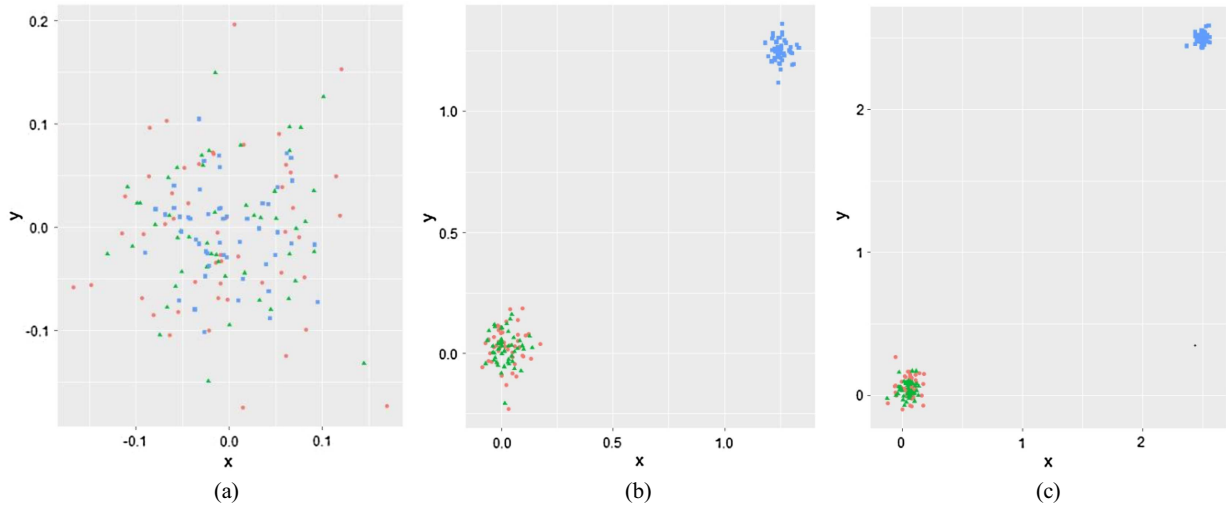


FIGURE 8. Optimal warehouse locations obtained using different norms in Case 7. (a)  $B = 0$ . (b)  $B = 5$ . (c)  $B = 10$ .

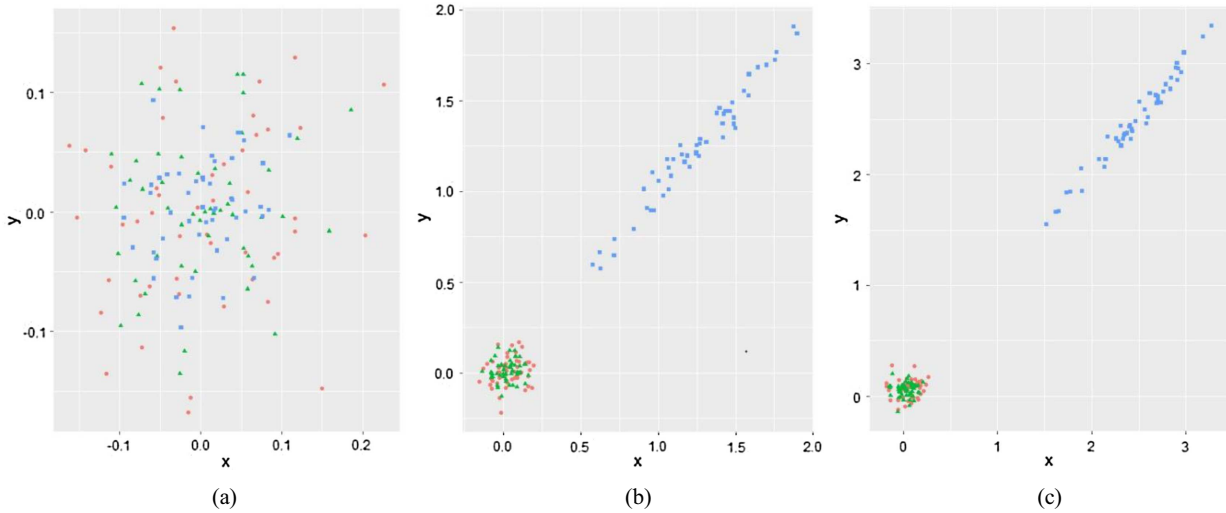


FIGURE 9. Optimal warehouse locations obtained using different norms in Case 8. (a)  $B = 0$ . (b)  $B = 5$ . (c)  $B = 10$ .

To evaluate the robustness of the proposed optimal location estimation methods and the CG method against the outliers, instances 19–27 with different outlier numbers  $H$  are tested. To evaluate the robust performances of different methods, the biases are evaluated and compared. In this study, we calculate the distance between the obtained warehouse location  $(\tilde{u}_o, \tilde{v}_o)$  without outliers and the location  $(\tilde{u}, \tilde{v})$  with outliers, which are defined as

$$\text{Bias}_M = |\tilde{u}_o - \tilde{u}| + |\tilde{v}_o - \tilde{v}| \quad (5.1)$$

and

$$\text{Bias}_E = \sqrt{(\tilde{u}_o - \tilde{u})^2 + (\tilde{v}_o - \tilde{v})^2} \quad (5.2)$$

TABLE 4. Summary of the benchmark instances.

Instance ID	Problem size	Instance ID	Problem size	Instance ID	Problem size
1	32	10	39	19	60
2	33	11	44	20	61
3	33	12	45	21	62
4	34	13	45	22	63
5	36	14	46	23	63
6	37	15	48	24	64
7	37	16	53	25	65
8	38	17	54	26	69
9	39	18	55	27	80

TABLE 5. Optimal warehouse locations based on Manhattan and Euclidean distances along with their corresponding values of the objective functions.

Instance ID	Manhattan distance				Euclidean distance			
	CG1		Weighted median		CG2		BFGS	
	$(\bar{u}, \bar{v})$	Obj <sub>1</sub>	$(\bar{u}, \bar{v})$	Obj <sub>1</sub>	$(\bar{u}, \bar{v})$	Obj <sub>2</sub>	$(\bar{u}, \bar{v})$	Obj <sub>2</sub>
1	(54.7, 39.4)	23,390.1	(58.0, 39.0)	23,262.0	(54.7, 39.4)	18,092.8	(61.7, 38.4)	17,938.5
2	(47.0, 45.1)	21,423.1	(47.0, 43.0)	21,372.0	(47.0, 45.1)	16,826.2	(48.3, 48.7)	16,793.9
3	(39.6, 49.7)	24,106.3	(45.0, 46.0)	24,036.0	(39.6, 49.7)	18,500.5	(42.5, 53.4)	18,434.2
4	(49.9, 59.6)	23,360.5	(55.0, 75.0)	22,532.0	(49.9, 59.6)	17,667.9	(54.3, 70.7)	17,251.8
5	(43.6, 50.4)	20,431.6	(33.0, 49.0)	20,020.0	(43.6, 50.4)	15,407.6	(45.0, 50.1)	15,401.3
6	(53.5, 50.8)	18,793.9	(53.0, 49.0)	18,777.0	(53.5, 50.8)	14,820.1	(52.1, 46.7)	14,734.0
7	(54.8, 19.3)	26,428.2	(57.0, 50.0)	26,351.0	(54.8, 49.3)	20,335.0	(55.0, 49.2)	20,334.8
8	(53.6, 48.5)	23,182.7	(47.0, 43.0)	22,902.0	(53.6, 48.5)	17,571.2	(47.7, 46.4)	17,406.2
9	(48.7, 41.3)	22,486.1	(45.0, 35.0)	21,726.0	(48.7, 41.3)	17,553.6	(48.0, 36.6)	17,461.3
10	(48.5, 45.3)	25,214.8	(41.0, 31.0)	24,340.0	(48.5, 45.3)	19,581.1	(43.3, 35.1)	19,083.5
11	(46.8, 50.6)	25,863.0	(42.0, 47.0)	25,277.0	(46.8, 50.6)	19,646.3	(42.5, 50.3)	19,520.3
12	(43.6, 48.8)	31,458.1	(36.0, 44.0)	31,165.0	(43.6, 48.8)	23,795.1	(41.5, 46.7)	23,759.0
13	(55.0, 42.8)	28,739.1	(49.0, 35.0)	28,080.0	(55.0, 42.8)	22,494.6	(53.4, 38.9)	22,402.3
14	(46.6, 42.2)	30,027.5	(43.0, 41.0)	29,844.0	(46.6, 42.2)	23,390.2	(46.3, 37.2)	23,275.0
15	(48.2, 49.1)	32,851.1	(49.0, 47.0)	32,818.0	(48.2, 49.1)	25,461.2	(48.2, 50.4)	25,453.2
16	(47.9, 47.1)	35,100.4	(53.0, 40.0)	34,842.0	(47.9, 47.1)	26,754.0	(48.9, 48.5)	26,732.8
17	(49.2, 43.7)	34,074.9	(51.0, 53.0)	33,944.0	(49.2, 43.7)	26,528.0	(50.8, 56.4)	26,479.0
18	(50.6, 50.3)	41,093.9	(50.0, 46.0)	40,877.0	(50.6, 50.3)	32,430.5	(45.9, 49.6)	32,284.5

**Notes.** CG1: based on the study of Gao *et al.* [37]; CG2: based on the study of Esnaf and Küçükdeniz [27] and Nadizadeh *et al.* [54].

for Manhattan distance and Euclidean distance, respectively.

As shown in Tables B.1–B.9, the warehouse locations and biases are summarized given different outlier numbers in each of the instances. It is obvious that the CG method is extremely sensitive to the outliers, and even a single outlier is enough to greatly influence the warehouse location in both distances. However, the weighted median and BFGS methods are not sensitive to the outliers in Manhattan and Euclidean distances respectively, which indicates that they have better performances than the CG method to estimate the facility location against the outliers. Thus, we can conclude that the weighted median and BFGS methods are not only able to minimize the total Manhattan distance and Euclidean distances respectively but also reliable to estimate the facility location against the outliers.

## 6. CONCLUDING REMARKS

This study focused on the warehouse location problem under data contamination in disaster response. We applied three different methods for three distances (norms), namely, Manhattan distance, Euclidean distance, and squared Euclidean distance, to determine the optimal warehouse locations. The locations were compared to illustrate the effectiveness of the proposed location estimation methods. The comparison results showed that the proposed location estimation methods outperformed the CG method both in accuracy and robustness. Then a simulation study was conducted to present the robust performance of the proposed location estimation methods. Finally, we carried out an illustrative example based on 27 instances in a benchmark to investigate the effectiveness and efficiency of the proposed methods. The results clearly showed that the proposed location estimation methods outperformed the CG method in determining the optimal warehouse locations when the data contamination is involved.

According to the results obtained in Sections 5 and 6, the warehouse locations based on different distances were cleared. As mentioned earlier, the CG method (weighted mean) was widely used in previous continuous-space facility location problems. However, the weighted mean is unable to minimize the total weighted Manhattan distance and Euclidean distance. Indeed the appropriate locations are very significant for decision-makers as they are strongly related to the objectives. On one hand, in terms of Manhattan distance, we found that the weighted median is the optimal solution to minimize the total weighted Manhattan distance in the urban area of interest and then we developed the R function of the weighted median. On the other hand, as for the Euclidean distance in the rural area of interest and long distance cases, we could not find the optimal warehouse location in explicit form so that we need to use numerical methods. In this study, we applied the BFGS method to obtain the optimal warehouse location as the convexity of  $\text{Obj}_2$ . As a consequence, we conducted this study to help decision-makers determine the optimal warehouse location based on the specific distance that they encountered.

In future work, some directions are meaningful that could be explored deeply from the following three perspectives. This study considers a single-facility location in disaster response. However, it would be interesting to develop a robust optimization model for the multi-facility location problem with data contamination. Besides, minimizing the maximum distance between the facilities is also meaningful in the multi-facility location problem. In addition, it is also meaningful to derive the breakdown point regarding the  $L^2$  norm. These three questions will be considered in future research.

## APPENDIX A. PROOF OF THE CONVEXITY OF $\text{Obj}_2$

In this study, we use the BFGS method that is able to find the optimal warehouse location due to the convex  $\text{Obj}_2$ . Before we minimize the objective function  $\text{Obj}_2$ , let  $f(x, y) = \sqrt{(x - u)^2 + (y - v)^2}$ , then we have the following Hessian matrix  $H$ :

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial u^2} & \frac{\partial^2 f}{\partial u \partial v} \\ \frac{\partial^2 f}{\partial v \partial u} & \frac{\partial^2 f}{\partial v^2} \end{bmatrix} = \frac{1}{f^3} \begin{bmatrix} (y - v)^2 & -(x - u)(y - v) \\ -(x - u)(y - v) & (x - u)^2 \end{bmatrix}. \quad (\text{A.1})$$

Then we can construct the following equation:

$$\begin{bmatrix} s & t \end{bmatrix} H \begin{bmatrix} s \\ t \end{bmatrix} = \frac{1}{f^3} \begin{bmatrix} s & t \end{bmatrix} \begin{bmatrix} (y - v)^2 & -(x - u)(y - v) \\ -(x - u)(y - v) & (x - u)^2 \end{bmatrix} \begin{bmatrix} s \\ t \end{bmatrix} = \frac{1}{f^3} [s(y - v) - t(x - u)]^2 \geq 0. \quad (\text{A.2})$$

The above equation is always nonnegative for any values of  $(s, t)$  and  $(x, y)$ , which indicates that  $H$  is positive semi-definite for any  $(s, t)$ . Thus,  $f(x, y)$  is always convex, which results in that the summation of convex functions  $\text{Obj}_2$  is also convex. There can be only one optimal solution for (3.19) and (3.20), which is the global minimum. For more details, see Section 3.4 of Boyd and Vandenberghe [13] and Section 2.1 of Beck [12]. As a consequence, it is easy for the BFGS method to determine the optimal warehouse location.



## APPENDIX B. R FUNCTION OF WEIGHTED MEDIAN

We provide the R function for the weighted median. Note that the R language includes the weighted mean as a basic function which is easily used by `weighted.mean()` command. The R function for the weighted median is given by

```

1. weighted.median <- function(x, w) {
2. # Preparation
3. if (missing(w)) w = rep(1L,length(x))
4. if (length(w) != length(x)) stop("'x' and 'w' must have the same length")
5. x = as.double(as.vector(x))
6. w = as.double(as.vector(w))
7. ok= complete.cases(x,w); xx=x[ok]; ww=w[ok]
8. stopifnot(all(w >= 0))
9. if(all(w <= 0)) stop("All weights are zero", call.=FALSE)
10. orderorderx = order(x)
11. xx = x[orderx]
12. ww = w[orderx] / sum(w)
13. Fn = cumsum(w)
14. tiny = sqrt(.Machine$double.eps)
15. # Main part
16. if (all(abs(Fn-0.5)>tiny)) { # any values of Fn is not 1/2.
17. k = sum(Fn < 0.5)
18. return(x[k+1])
19. } else {
20. k = which.min (signif(abs(Fn-0.5),digits=12)) # Find k with Fn=0.5
21. if (w[k+1] < tiny) { # check if w[k+1] == 0
22. return(x[k+1])
23. } else {
24. return(0.5*(x[k]+x[k+1]))
25. }
26. }
27. }
```

TABLE B.1. Optimal warehouse locations under different outlier numbers for different distances in Instance 19.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(51.3, 51.6)	0.0	(49.0, 57.0)	0.0	(51.3, 51.6)	0.0	(49.3, 52.4)	0.0
1	(67.6, 74.9)	39.6	(49.0, 57.0)	0.0	(67.6, 74.9)	28.4	(49.6, 52.2)	0.4
2	(71.9, 79.5)	48.5	(53.0, 57.0)	4.0	(71.9, 79.5)	34.7	(50.7, 53.5)	1.8
3	(93.2, 98.8)	89.1	(53.0, 57.0)	4.0	(93.2, 98.8)	63.1	(50.6, 53.7)	1.8
4	(110.0, 106.0)	113.1	(53.0, 59.0)	6.0	(110.0, 106.0)	80.0	(50.8, 55.0)	3.0
5	(116.0, 112.0)	125.1	(53.0, 59.0)	6.0	(116.0, 112.0)	88.5	(50.8, 54.9)	2.9
6	(121.0, 113.0)	131.1	(53.0, 61.0)	8.0	(121.0, 113.0)	92.9	(52.0, 56.0)	4.5
7	(122.0, 120.0)	139.1	(53.0, 61.0)	8.0	(122.0, 120.0)	98.4	(54.0, 58.6)	7.8
8	(133.0, 138.0)	168.1	(53.0, 61.0)	8.0	(133.0, 138.0)	119.0	(54.9, 58.5)	8.3
9	(158.0, 139.0)	194.1	(53.0, 61.0)	8.0	(158.0, 139.0)	138.0	(56.2, 60.1)	10.3
10	(160.0, 139.0)	196.1	(53.0, 65.0)	12.0	(160.0, 139.0)	139.0	(56.2, 60.5)	10.6

TABLE B.2. Optimal warehouse locations under different outlier numbers for different distances in Instance 20.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(45.9, 48.6)	0.0	(47.0, 45.0)	0.0	(45.9, 48.6)	0.0	(42.7, 46.6)	0.0
1	(59.8, 54.8)	20.1	(47.0, 45.0)	0.0	(59.8, 54.8)	15.2	(42.5, 47.0)	0.5
2	(60.9, 59.3)	25.7	(47.0, 45.0)	0.0	(60.9, 59.3)	18.4	(43.1, 47.3)	0.8
3	(75.0, 77.2)	57.7	(47.0, 45.0)	0.0	(75.0, 77.2)	40.8	(43.5, 47.0)	0.9
4	(85.8, 80.0)	71.3	(47.0, 45.0)	0.0	(85.8, 80.0)	50.8	(43.4, 47.5)	1.1
5	(89.3, 85.2)	80.0	(47.0, 45.0)	0.0	(89.3, 85.2)	56.8	(43.5, 47.5)	1.2
6	(99.0, 90.1)	94.6	(47.0, 47.0)	2.0	(99.0, 90.1)	67.4	(43.8, 48.9)	2.6
7	(100.0, 108.0)	113.5	(47.0, 47.0)	2.0	(100.0, 108.0)	80.3	(44.9, 49.3)	3.5
8	(105.0, 125.0)	135.5	(47.0, 47.0)	2.0	(105.0, 125.0)	96.6	(46.4, 49.5)	4.7
9	(121.0, 129.0)	155.5	(47.0, 47.0)	2.0	(121.0, 129.0)	110.0	(46.5, 50.3)	5.3
10	(130.0, 135.0)	170.5	(47.0, 49.0)	4.0	(130.0, 135.0)	121.0	(46.3, 51.0)	5.7

TABLE B.3. Optimal warehouse locations under different outlier numbers for different distances in Instance 21.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(50.1, 53.6)	0	(52.0, 53.0)	0.0	(50.1, 53.6)	0.0	(49.2, 57.4)	0.0
1	(52.3, 55.1)	3.7	(53.0, 55.0)	2.0	(52.3, 55.1)	2.7	(50.2, 58.2)	1.3
2	(53.7, 56.4)	6.4	(53.0, 55.0)	2.0	(53.7, 56.4)	4.6	(50.2, 58.5)	1.5
3	(71.7, 62.1)	30.1	(53.0, 61.0)	8.0	(71.7, 62.1)	23.2	(51.5, 60.8)	4.1
4	(101.0, 64.3)	61.6	(53.0, 69.0)	16.0	(101.0, 64.3)	52.0	(52.8, 63.0)	6.7
5	(111.0, 64.7)	72.0	(56.0, 69.0)	16.0	(111.0, 64.7)	61.9	(54.7, 64.7)	9.1
6	(121.0, 78.5)	95.8	(64.0, 70.0)	29.0	(121.0, 78.5)	75.1	(58.4, 68.0)	14.0
7	(122.0, 79.1)	97.4	(64.0, 70.0)	29.0	(122.0, 79.1)	76.3	(58.5, 68.2)	14.3
8	(124.0, 80.9)	101.2	(64.0, 70.0)	29.0	(124.0, 80.9)	78.8	(58.5, 68.2)	14.3
9	(125.0, 82.2)	103.5	(64.0, 70.0)	29.0	(125.0, 82.2)	80.2	(58.9, 68.5)	14.7
10	(126.0, 87.5)	109.8	(64.0, 70.0)	29.0	(126.0, 87.5)	83.1	(59.7, 69.0)	15.6

TABLE B.4. Optimal warehouse locations under different outlier numbers for different distances in Instance 22.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(48.4, 49.8)	0.0	(45.0, 47.0)	0.0	(48.4, 49.8)	0.0	(47.1, 48.3)	0.0
1	(52.1, 50.8)	4.7	(45.0, 47.0)	0.0	(52.1, 50.8)	3.8	(47.6, 49.0)	0.9
2	(60.1, 61.6)	23.5	(49.0, 47.0)	4.0	(60.1, 61.6)	16.6	(49.8, 49.0)	2.8
3	(68.4, 74.8)	45.0	(57.0, 47.0)	12.0	(68.4, 74.8)	32.0	(51.0, 48.7)	3.9
4	(68.7, 90.3)	60.8	(57.0, 47.0)	12.0	(68.7, 90.3)	45.3	(52.4, 49.6)	5.5
5	(70.4, 90.5)	62.7	(57.0, 47.0)	12.0	(70.4, 90.5)	46.3	(52.5, 49.7)	5.6
6	(80.4, 90.7)	72.9	(59.0, 47.0)	14.0	(80.4, 90.7)	51.9	(53.8, 51.0)	7.2
7	(99.4, 109.0)	110.2	(59.0, 47.0)	14.0	(99.4, 109.0)	78.1	(53.8, 51.0)	7.2
8	(103.0, 109.0)	113.8	(59.0, 59.0)	26.0	(103.0, 109.0)	80.5	(53.9, 51.4)	7.5
9	(113.0, 132.0)	146.8	(59.0, 59.0)	26.0	(113.0, 132.0)	105.0	(55.3, 51.2)	8.7
10	(121.0, 136.0)	158.8	(59.0, 59.0)	26.0	(121.0, 136.0)	113.0	(55.3, 52.1)	9.0

TABLE B.5. Optimal warehouse locations under different outlier numbers for different distances in Instance 23.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(48.5, 52.3)	0.0	(41.0, 49.0)	0.0	(48.5, 52.3)	0.0	(44.4, 50.6)	0.0
1	(51.7, 56.9)	7.8	(46.0, 49.0)	5.0	(51.7, 56.9)	5.6	(44.4, 50.6)	0.0
2	(58.3, 65.0)	22.5	(46.0, 50.0)	6.0	(58.3, 65.0)	16.0	(46.7, 52.1)	2.8
3	(64.8, 73.7)	37.7	(46.0, 51.0)	7.0	(64.8, 73.7)	26.9	(49.5, 54.5)	6.4
4	(69.4, 75.8)	44.4	(55.0, 54.0)	19.0	(69.4, 75.8)	31.4	(49.5, 54.7)	6.5
5	(72.5, 86.8)	58.5	(60.0, 54.0)	24.0	(72.5, 86.8)	42.0	(51.0, 55.4)	8.2
6	(82.1, 110.0)	91.3	(60.0, 54.0)	24.0	(82.1, 110.0)	66.8	(52.2, 56.6)	9.8
7	(96.6, 123.0)	118.8	(60.0, 54.0)	24.0	(96.6, 123.0)	85.5	(52.2, 57.3)	10.3
8	(98.1, 136.0)	133.3	(60.0, 54.0)	24.0	(98.1, 136.0)	97.3	(54.7, 60.2)	14.1
9	(102.0, 141.0)	142.2	(60.0, 61.0)	31.0	(102.0, 141.0)	104.0	(56.1, 62.0)	16.3
10	(112.0, 146.0)	157.2	(60.0, 70.0)	40.0	(112.0, 146.0)	113.0	(57.7, 64.9)	19.5

TABLE B.6. Optimal warehouse locations under different outlier numbers for different distances in Instance 24.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(49.9, 50.0)	0.0	(51.0, 49.0)	0.0	(49.9, 50.0)	0.0	(44.4, 50.6)	0.0
1	(50.6, 52.0)	2.7	(51.0, 49.0)	0.0	(50.6, 52.0)	2.1	(44.4, 50.6)	0.0
2	(79.9, 80.0)	60.0	(55.0, 51.0)	6.0	(79.9, 80.0)	42.4	(49.5, 54.5)	6.4
3	(86.6, 81.4)	68.1	(55.0, 51.0)	6.0	(86.6, 81.4)	48.3	(49.5, 54.5)	6.4
4	(90.5, 81.8)	72.4	(57.0, 51.0)	8.0	(90.5, 81.8)	51.6	(49.5, 54.7)	6.5
5	(91.8, 84.7)	76.6	(57.0, 51.0)	8.0	(91.8, 84.7)	54.4	(51.0, 55.4)	8.2
6	(93.2, 90.3)	83.6	(59.0, 55.0)	14.0	(93.2, 90.3)	59.2	(52.2, 55.6)	9.3
7	(106.0, 104.0)	110.1	(59.0, 55.0)	14.0	(106.0, 104.0)	77.9	(52.2, 57.3)	10.3
8	(122.0, 106.0)	128.1	(59.0, 57.0)	16.0	(122.0, 106.0)	91.3	(54.7, 60.2)	14.1
9	(124.0, 112.0)	136.1	(59.0, 57.0)	16.0	(124.0, 112.0)	96.6	(56.1, 62.0)	16.3
10	(129.0, 119.0)	148.1	(59.0, 57.0)	16.0	(129.0, 119.0)	105.0	(57.7, 64.9)	19.5

TABLE B.7. Optimal warehouse locations under different outlier numbers for different distances in Instance 25.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(46.9, 47.5)	0.0	(49.0, 41.0)	0.0	(46.9, 47.5)	0.0	(47.3, 45.9)	0.0
1	(47.3, 56.9)	9.8	(49.0, 41.0)	0.0	(47.3, 56.9)	9.4	(48.1, 46.7)	1.1
2	(49.4, 61.4)	16.4	(51.0, 55.0)	16.0	(49.4, 61.4)	14.1	(49.6, 50.3)	5.0
3	(53.5, 66.8)	25.9	(53.0, 59.0)	22.0	(53.5, 66.8)	20.4	(51.0, 52.7)	7.7
4	(55.4, 67.8)	28.8	(53.0, 59.0)	22.0	(55.4, 67.8)	22.0	(50.9, 52.8)	7.8
5	(71.2, 85.5)	62.3	(53.0, 59.0)	22.0	(71.2, 85.5)	45.1	(51.1, 52.6)	7.7
6	(82.6, 88.0)	76.2	(53.0, 63.0)	26.0	(82.6, 88.0)	54.0	(51.2, 54.1)	9.1
7	(100.0, 95.6)	101.2	(53.0, 65.0)	30.0	(100.0, 95.6)	71.6	(51.0, 55.0)	9.8
8	(101.0, 95.9)	102.5	(55.0, 67.0)	32.0	(101.0, 95.9)	72.6	(51.0, 55.0)	9.8
9	(120.0, 100.0)	125.6	(55.0, 67.0)	32.0	(120.0, 100.0)	90.0	(52.2, 56.9)	12.0
10	(130.0, 106.0)	141.6	(59.0, 69.0)	38.0	(130.0, 106.0)	102.0	(54.0, 60.2)	15.8

TABLE B.8. Optimal warehouse locations under different outlier numbers for different distances in Instance 26.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(42.2, 45.5)	0.0	(37.0, 44.0)	0.0	(42.2, 45.5)	0.0	(39.5, 44.9)	0.0
1	(45.6, 58.5)	16.4	(37.0, 44.0)	0.0	(45.6, 58.5)	13.4	(40.5, 45.2)	1.0
2	(67.9, 70.5)	50.7	(37.0, 44.0)	0.0	(67.9, 70.5)	35.9	(40.3, 46.0)	1.4
3	(70.6, 75.3)	58.2	(38.0, 44.0)	1.0	(70.6, 75.3)	41.2	(40.6, 45.9)	1.5
4	(89.4, 81.2)	82.9	(38.0, 46.0)	3.0	(89.4, 81.2)	59.2	(40.7, 47.2)	2.6
5	(96.9, 85.8)	95.0	(41.0, 46.0)	6.0	(96.9, 85.8)	67.9	(40.6, 47.4)	2.7
6	(106.0, 85.8)	104.1	(41.0, 51.0)	11.0	(106.0, 85.8)	75.5	(42.3, 48.7)	4.7
7	(110.0, 96.5)	118.8	(41.0, 51.0)	11.0	(110.0, 96.5)	84.8	(44.2, 50.7)	7.5
8	(122.0, 103.0)	137.3	(41.0, 51.0)	11.0	(122.0, 103.0)	98.4	(44.0, 51.1)	7.7
9	(147.0, 108.0)	167.3	(41.0, 54.0)	14.0	(147.0, 108.0)	122.0	(44.2, 52.3)	8.8
10	(147.0, 114.0)	173.3	(41.0, 54.0)	14.0	(147.0, 114.0)	125.0	(44.7, 52.7)	9.4

TABLE B.9. Optimal warehouse locations under different outlier numbers for different distances in Instance 27.

$B$	Manhattan distance				Euclidean distance			
	CG		Weighted median		CG		BFGS	
	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_M$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$	$(\tilde{u}, \tilde{v})$	$\text{Bias}_E$
0	(55.4, 44.9)	0.0	(57.0, 42.0)	0.0	(55.4, 44.9)	0.0	(57.1, 43.6)	0.0
1	(75.3, 47.5)	22.5	(57.0, 42.0)	0.0	(75.3, 47.5)	20.1	(57.0, 44.0)	0.4
2	(76.3, 47.8)	23.8	(57.0, 42.0)	0.0	(76.3, 47.8)	21.1	(57.1, 44.1)	0.5
3	(77.5, 49.0)	26.2	(57.0, 42.0)	0.0	(77.5, 49.0)	22.5	(57.2, 44.1)	0.5
4	(81.2, 59.7)	40.6	(57.0, 42.0)	0.0	(81.2, 59.7)	29.7	(57.9, 44.2)	1.0
5	(83.1, 60.9)	43.7	(58.0, 43.0)	2.0	(83.1, 60.9)	32.0	(58.3, 44.5)	1.5
6	(93.5, 70.2)	63.4	(61.0, 43.0)	5.0	(93.5, 70.2)	45.7	(59.5, 44.1)	2.5
7	(102.0, 80.7)	82.4	(61.0, 43.0)	5.0	(102.0, 80.7)	58.8	(59.8, 44.0)	2.7
8	(103.0, 82.2)	84.9	(61.0, 43.0)	5.0	(103.0, 82.2)	60.5	(59.9, 44.0)	2.8
9	(113.0, 85.6)	98.3	(61.0, 45.0)	7.0	(113.0, 85.6)	70.5	(59.9, 44.4)	2.9
10	(116.0, 89.5)	105.2	(61.0, 45.0)	7.0	(116.0, 89.5)	75.2	(60.1, 44.3)	3.1

*Acknowledgements.* The work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (NRF-2017R1A2B4004169) and by the BK21 FOUR of the NRF funded by the Ministry of Education (No. 5199990914451).

## REFERENCES

- [1] M. Acar and O. Kaya, A healthcare network design model with mobile hospitals for disaster preparedness: a case study for Istanbul earthquake. *Transp. Res. Part E: Logistics Transp. Rev.* **130** (2019) 273–292.
- [2] O.I. Alsalloum and G.K. Rand, Extensions to emergency vehicle location models. *Comput. Oper. Res.* **33** (2006) 2725–2743.
- [3] Y. An, B. Zeng, Y. Zhang and L. Zhao, Reliable  $p$ -median facility location problem: two-stage robust models and algorithms. *Transp. Res. Part B: Methodol.* **64** (2014) 54–72.
- [4] A.B. Arabani and R.Z. Farahani, Facility location dynamics: an overview of classifications and applications. *Comput. Ind. Eng.* **62** (2012) 408–420.
- [5] M.G. Ashtiani, A. Makui and R. Ramezani, A robust model for a leader–follower competitive facility location problem in a discrete space. *Appl. Math. Model.* **37** (2013) 62–71.
- [6] P. Augerat, J.M. Belenguer, E. Benavent, A. Corberán, D. Naddef and G. Rinaldi, Computational Results with a Branch and Cut Code for the Capacitated Vehicle Routing Problem (IMAG) (1995).
- [7] J. Badal, M. Vázquez-Prada and Á. González, Preliminary quantitative assessment of earthquake casualties and damages. *Nat. Hazards* **34** (2005) 353–374.
- [8] X. Bai, Two-stage multiobjective optimization for emergency supplies allocation problem under integrated uncertainty. *Math. Prob. Eng.* **2016** (2016) 2823835.
- [9] M. Baïou and F. Barahona, A polyhedral study of a two level facility location model. *RAIRO:OR* **48** (2014) 153–165.
- [10] O. Baron, J. Milner and H. Naseraldin, Facility location: a robust optimization approach. *Prod. Oper. Manage.* **20** (2011) 772–785.
- [11] M. Bastian and M. Volkmer, A perfect forward procedure for a single facility dynamic location/relocation problem. *Oper. Res. Lett.* **12** (1992) 11–16.
- [12] A. Beck, Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB. SIAM (2014).
- [13] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press (2004).
- [14] C.G. Broyden, Quasi-Newton methods. In: Numerical Methods for Unconstrained Optimization. Academic Press, New York (1972).
- [15] R.H. Byrd, P. Lu, J. Nocedal and C. Zhu, A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **16** (1995) 1190–208.
- [16] C. Cao, Y. Liu, O. Tang and X. Gao, A fuzzy bi-level optimization model for multi-period post-disaster relief distribution in sustainable humanitarian supply chains. *Int. J. Prod. Econ.* (2021) 108081.
- [17] E. Cetin and L.S. Sarul, A blood bank location model: a multiobjective approach. *Eur. J. Pure Appl. Math.* **2** (2009) 112–124.
- [18] S. Chanta, M.E. Mayorga and L.A. McLay, Improving emergency service in rural areas: a bi-objective covering location model for EMS systems. *Ann. Oper. Res.* **221** (2014) 133–159.

- [19] C.-H. Chen and C.-J. Ting, Combining Lagrangian heuristic and ant colony system to solve the single source capacitated facility location problem. *Transp. Res. Part E: Logistics Transp. Rev.* **44** (2008) 1099–1122.
- [20] Y. Chen, Q. Zhao, L. Wang and M. Dessouky, The regional cooperation-based warehouse location problem for relief supplies. *Comput. Ind. Eng.* **102**, 259–267.
- [21] W.-Y. Chiu and B.-S. Chen, Mobile location estimation in urban areas using mixed Manhattan/Euclidean norm and convex optimization. *IEEE Trans. Wireless Commun.* **8** (2009) 414–423.
- [22] R. Church and C.R. Velle, The maximal covering location problem. *Papers Reg. Sci.* **32** (1974) 101–118.
- [23] I.A. Contreras and J.A. Díaz, Scatter search for the single source capacitated facility location problem. *Ann. Oper. Res.* **157** (2008) 73–89.
- [24] M. Csorgo, Quantile Processes with Statistical Applications. SIAM (1983).
- [25] V. De Rosa, E. Hartmann, M. Gebhard and J. Wollenweber, Robust capacitated facility location model for acquisitions under uncertainty. *Comput. Ind. Eng.* **72** (2014) 206–216.
- [26] F.Y. Edgeworth, XXII. On a new method of reducing observations relating to several quantities. *London Edinburgh Dublin Phil. Mag. J. Sci.* **25** (1888) 184–191.
- [27] Ş. Esnaf and T. Küçükdeniz, A fuzzy clustering-based hybrid method for a multi-facility location problem. *J. Intell. Manuf.* **20** (2009) 259–265.
- [28] R.Z. Farahani, Z. Drezner and N. Asgari, Single facility location and relocation problem with time dependent weights and discrete planning horizon. *Anna. Oper. Res.* **167** (2009) 353–368.
- [29] R. Fletcher, A new approach to variable metric algorithms. *Comput. J.* **13** (1970) 317–322.
- [30] R.L. Francis, J.A. White and L.F. McGinnis, Facility Layout and Location: An Analytical Approach. Prentice-Hall Englewood Cliffs, NJ (1974).
- [31] X. Gao, A bi-level stochastic optimization model for multi-commodity rebalancing under uncertainty in disaster response. *Ann. Oper. Res.* (2019) 1–34.
- [32] X. Gao, A location-driven approach for warehouse location problem. To appear in: *J. Oper. Res. Soc.* (2020). DOI: [10.1080/01605682.2020.1811790](https://doi.org/10.1080/01605682.2020.1811790).
- [33] X. Gao and C. Cao, Multi-commodity rebalancing and transportation planning considering traffic congestion and uncertainties in disaster response. *Comput. Ind. Eng.* **149** (2020) 106782.
- [34] X. Gao and X. Jin, A robust multi-commodity rebalancing process in humanitarian logistics. In: Vol. 591 of *Advances in Production Management System, APMS 2020*. Springer, Cham (2020).
- [35] X. Gao and G.M. Lee, A stochastic programming model for multi-commodity redistribution planning in disaster response. In: IFIP International Conference on Advances in Production Management Systems. Springer (2018) 67–78.
- [36] X. Gao, Y. Zhou, M.I.H. Amir, F.A. Rosyidah and G.M. Lee, A hybrid genetic algorithm for multi-emergency medical service center location-allocation problem in disaster response. *Int. J. Ind. Eng.: Theory App. Pract.* **24** (2017) 663–679.
- [37] X. Gao, M.K. Nayeem and I.M. Hezam, A robust two-stage transit-based evacuation model for large-scale disaster response. *Measurement* **145** (2019) 713–723.
- [38] X. Gao, X. Jin, P. Zheng and C. Cui, Multi-modal transportation planning for multi-commodity rebalancing under uncertainty in humanitarian logistics. *Adv. Eng. Inf.* **47** (2021) 101223.
- [39] F. Gayraud, N. Grangeon, L. Deroussi and S. Norre, Notation and classification for logistic network design models. *RAIRO:OR* **49** (2015) 195–214.
- [40] D. Goldfarb, A family of variable-metric methods derived by variational means. *Math. Comput.* **24** (1970) 23–26.
- [41] G. Guastaroba and M.G. Speranza, A heuristic for BILP problems: the single source capacitated facility location problem. *Eur. J. Oper. Res.* **238** (2014) 438–450.
- [42] N. Gülpınar, D. Pachamanova and E. Çanakoglu, Robust strategies for facility location under uncertainty. *Eur. J. Oper. Res.* **225** (2013) 21–35.
- [43] F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel, Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons (2011).
- [44] T.P. Hettmansperger and J.W. McKean, Robust Nonparametric Statistical Methods. CRC Press (2010).
- [45] S.C. Ho, An iterated tabu search heuristic for the single source capacitated facility location problem. *Appl. Soft Comput.* **27** (2015) 169–178.
- [46] K. Hogan and C. ReVelle, Concepts and applications of backup coverage. *Manage. Sci.* **32** (1986) 1434–1444.
- [47] R. Ihaka and R. Gentleman, R: a language for data analysis and graphics. *J. Comput. Graphical Stat.* **5** (1996) 299–314.
- [48] C.A. Irawan, M. Luis, S. Salhi and A. Imran, The incorporation of fixed cost and multilevel capacities into the discrete and continuous single source capacitated facility location problem. *Ann. Oper. Res.* **275** (2019) 367–392.
- [49] M. Körkel, Discrete facility location with nonlinear facility costs. *RAIRO:OR* **25** (1991) 31–43.
- [50] E.L. Lehmann, Elements of Large-Sample Theory. Springer Science & Business Media (2004).
- [51] S. Manzour-al-Ajdad, S.A. Torabi and S. Salhi, A hierarchical algorithm for the planar single-facility location routing problem. *Comput. Oper. Res.* **39** (2012) 461–470.
- [52] E. Moradi and M. Bidkhor, Single facility location problem. In: Facility Location. Springer (2009).
- [53] J. Morris, R. Love and G. Wesolowsky, Facilities Location: Models and Methods. North-Holland, New York (1988).
- [54] A. Nadizadeh, R. Sahraeian, A.S. Zadeh and S.M. Homayouni, Using greedy clustering method to solve capacitated location-routing problem. *Afr. J. Bus. Manage.* **5** (2011) 8470–8477.



- [55] Y. Ohsawa, A geometrical solution for quadratic bicriteria location models. *Eur. J. Oper. Res.* **114** (1999) 380–388.
- [56] N. Onnela, *Determining the optimal distribution center location*. Master thesis (2015).
- [57] L. Ouyang, Y. Ma, J. Wang and Y. Tu, A new loss function for multi-response optimization with model parameter uncertainty and implementation errors. *Eur. J. Oper. Res.* **258** (2017) 552–563.
- [58] L. Ouyang, J. Chen, Y.Z. Ma, C. Park and J. Jin, Bayesian closed-loop robust process design considering model uncertainty and data quality. *IIE Trans.* (2020) 288–300.
- [59] A.B. Owen, Empirical Likelihood. Chapman and Hall/CRC (2001).
- [60] C. Park, Determination of the joint confidence region of the optimal operating conditions in robust design by the bootstrap technique. *Int. J. Prod. Res.* **51** (2013) 4695–4703.
- [61] C. Park and M. Leeds, A highly efficient robust design under data contamination. *Comput. Ind. Eng.* **93** (2016) 131–142.
- [62] M. Parnas and D. Ron, Testing metric properties. *Inf. Comput.* **187** (2003) 155–195.
- [63] P.J. Rousseeuw and C. Croux, Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **88** (1993) 1273–1283.
- [64] R. Serfling, Asymptotic relative efficiency in estimation. *Int. Encycl. Stat. Sci.* **23** (2011) 68–72.
- [65] D.F. Shanno, Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* **24** (1970) 647–656.
- [66] R.C. Team, R: A language and environment for statistical computing (2019) 201.
- [67] J.W. Tukey, A survey of sampling from contaminated distributions. *Contrib. Prob. Stat.* (1960) 448–485.
- [68] I. Vazler, K. Sabo and R. Scitovski, Weighted median of the data in solving least absolute deviations problems. *Commun. Stat.-Theory Methods* **41** (2012) 1455–1465.
- [69] L. Yajie, L. Hongtao, W. Zhiyong and Z. Dezhi, A robust model predictive control approach for post-disaster relief distribution. *Comput. Ind. Eng.* **135** (2019) 1253–1270.
- [70] M. Zaferanieh, H.T. Kakhki, J. Brimberg and G.O. Wesolowsky, A BSSS algorithm for the single facility location problem in two regions with different norms. *Eur. J. Oper. Res.* **190** (2008) 79–89.