

AGGREGATION AND NON AGGREGATION TECHNIQUES FOR LARGE FACILITY LOCATION PROBLEMS - A SURVEY

Chandra Ade IRAWAN

*Centre for Operational Research and Logistics, Department of Mathematics, University
of Portsmouth, UK and Department of Industrial Engineering, Institut Teknologi
Nasional, Bandung, Indonesia
chandra.irawan@port.ac.uk*

Said SALHI

*Centre for Logistics and Heuristic Optimization (CLHO), Kent Business School,
University of Kent
S.Salhi@kent.ac.uk*

Received: September 2014 / Accepted: January 2015

Abstract: A facility location problem is concerned with determining the location of some useful facilities in such a way so to fulfil one or a few objective functions and constraints. We survey those problems where, in the presence of a large number of customers, some form of aggregation may be required. In addition, a review on conditional location problems where some (say q) facilities already exist in the study area is presented.

Keywords: Large Location Problem, p -median and p -centre Problems, Point Representation, Aggregation.

MSC: 90B06, 90C05, 90C08.

1. INTRODUCTION

Research in location theory formally started in 1909 by Alfred Weber [110] known as the father of modern location theory (Eilon *et al.* [35]). He studied the problem of locating a single warehouse in order to minimise the total travel distance between the warehouse and a set of customers. Since then, many researchers

have observed this problem in many different areas. These include Hotelling [65], who considered the problem of locating two competing vendors along a straight line. The first powerful iterative approach to deal with the single facility location problem in the plane so to minimise the sum of the weighted distances from a single facility to all the points (i.e., continuous space) is put forward by Weiszfeld [111].

Modern location theory arose during the 1950's when several researchers investigated some problems in the area of location analysis. These include Valinsky [109], who determined the optimal location for fire fighting vehicles, Miehle [81], who investigated the problem of minimizing link length in networks, Mansfield and Wein [80], who presented a model for the location of a railroad classification yard, and Young [113], who determined the optimum location for checking stations on a rail line.

The study of location theory began to grow when Hakimi [54] published the seminal paper about location problems. In this paper, he wanted to locate one or more switching centres in a communication network and police stations in a highway system to minimise the sum of distances or the maximum distance between facilities and points on a network. These models are known as the p -median and p -centre problems respectively, where p denotes the number of facilities to be located. This will be reviewed later. For more information or references, chapter 1 of Drezner and Hamacher [34] gives a brief review of the history of location analysis. Farahani *et al.* [40] provided a recent review on hierarchical facility location problem.

There are many books and papers that provide a review of location theory. For books, which briefly describe the taxonomy of location problems and a variety of techniques to solve location problems, see Eilon *et al.* [35], Handler and Mirchandani [56], Love *et al.* [77], Mirchandani and Francis [84], Francis *et al.* [43], Daskin [25], Drezner and Hamacher [34], and Nickel and Puerto [91]. Moreover, there are several interesting papers that review location problems, including Francis *et al.* [42], Tansel *et al.* [106] [107], Aikens [1], Brandeau and Chiu [12], Eiselt *et al.* [36], Sridharan [104], Hale and Moberg [55], Daskin [27], and Brimberg *et al.* [13].

Location problems may be classified by their objective functions, including the minimax, the maximin, or the minisum. Based on these objectives, location problems can be divided into three groups as follows:

- Median Problems (minisum)
The median problems are those where one or more facilities are to be located in order to minimise the average cost (average time) between the customer and the nearest facility. The problem is known as the minisum problem or the p -median problem, p denoting the number of facilities to be located.
- Centre Problems (minimax/maximin)
Centre problems arise when a given number of facilities needs to be found with the objective of minimizing the maximum travel cost (travel time) between customers and the nearest facility. The problem is known as the

minimax problem or the p -centre problem. In the case of locating obnoxious facilities such as nuclear/chemical station and waste disposal sites, the objective function reverses to a maximin instead of a minimax.

- Covering Problems

Covering problems occur when there is a given critical coverage distance or cost or time within customers and facilities. The number of facilities is deemed sufficient if the distance between the customer and the nearest facility does not exceed some critical value, but deemed insufficient otherwise. This introduces the notion of coverage. Note that the p -centre can also be considered as a version of covering where the coverage value becomes a decision variable instead of an input.

The conditional location problems occur if some (say q) facilities already exist in the study area, and the aim is to locate p new facilities given the existing q facilities. This problem is also known as the (p, q) median/centre problem (Drezner [33]) where a customer can be served by the existing or the new open facilities, whichever that is closest to the customer. When $q = 0$, the problem reduces to the unconditional problem (the p -median/centre problem for short).

The purpose of this paper is to survey methods for solving large discrete location problems, and the review is classified into two main categories, namely a review with and without the incorporation of aggregation. In addition, a review on the conditional location problems is presented. This survey could also be very useful for researchers and students to find questions that identify research gaps. The paper is organized as follows. The review on solving large location problems using aggregation is described in Section 2, followed by the one without aggregation in Section 3. The review on conditional location problems is given in Section 4. The last section provides a conclusion and some highlights for possible research avenues.

2. A REVIEW ON SOLVING LARGE LOCATION PROBLEMS USING AGGREGATION

In special cases, facility location problems may consist of a large number of demand points (customers). These problems arise, for example, in urban or regional area where the demand points are individual private residences. It may be time consuming or even impossible to solve optimally the location problems involving a large number of demand points. It is quite common to aggregate demand points when solving large scale location problems. The idea behind the aggregation is to reduce the number of demand points to be small enough so an optimiser can be used. In this case, the location problems are partitioned into smaller problems and can be solved within a reasonable amount of computing time. However, this aggregation may reduce the accuracy of the model. In other words, this aggregation introduces error in the data used by location models and models output. Many researchers have studied the effects of aggregation on the solution of location problems. Note that in this review we do not discuss the

case of covering a complete region, such as land for irrigation, nature reserve, and weather radar equipments. Approximating such areas by point may not be appropriate because the errors due to approximation will occur. One way is to partition the entire area into smaller areas (polygons), where each polygon needs to be covered, see Murray and Wei [89] and Murray *et al.* [90].

In this section, first we give a brief introduction to aggregation by describing an aggregation scheme on the p -median, the p -centre, and the Set Covering problems. This is followed by the description of the aggregation error measurements, aggregation literature on median problems and on centre/covering problems, and related aggregation work on other location problems.

2.1. An Introduction to Aggregation

The idea behind the aggregation is to reduce the number of demand points so to be small enough that an optimiser can be used. In this case, the location problems are partitioned into smaller problems, and hence they can be solved within a reasonable amount of computing time. However, this aggregation may reduce the accuracy of the model. In other words, this aggregation introduces error in the data used by location models and models output. Many researchers have studied the effects of aggregation on the solution of location problems.

Current and Schilling [23] define demands point as Basic Spatial Unit (BSU) and aggregated demands point as Aggregated Spatial Unit (ASU). The right number of ASUs to be generated to solve location problems is a challenging issue. Until now, there is no a unique answer how to trade-off the benefits and costs of aggregation. The process of determining an aggregation scheme with a minimum error is an NP-hard problem, see Francis and Lowe [44].

Table 1 describes our notation in location models, which is focused on an aggregation approach. We assume that there are n BSUs, $i = 1, \dots, n$. Let C be the list of BSUs, $C = (c_1, c_2, \dots, c_n)$, and $I = \{1, 2, 3, \dots, n\}$ the set of all BSUs. Each BSU usually has a demand or a weight, say w_i . Conducting aggregation, n BSUs are replaced by m ASUs, where $m \ll n$. Let $I' = \{1, 2, 3, \dots, m\}$ represent the set of all ASUs, and each ASU denotes one or more BSUs (subset). Let C' be the list of ASUs, $C' = (c'_1, c'_2, \dots, c'_m)$ and A_k denotes the set of BSU in the subset k , $k = 1, \dots, m$. It is common that the centroid of BSU is used in each part of the subset as the ASU location. Let F represent the set of locations of the p facilities. In the original model, we denote the distance between c and the closest element in F by $D(F, c)$ with $D(F, c) = \min\{d(F, c), c \in C\}$, whereas in the aggregated model $D(F, c')$ represents the distance between c' and the nearest facility. Let the objective function with the given original BSUs be $f(F : C)$, whereas the one with ASUs be $f(F : C')$. The difference between $f(F : C)$ and $f(F : C')$, i.e. $|f(F : C) - f(F : C')|$, is known as the aggregation error.

Table 2 shows example formulations of the objective function on the p -median, the p -centre, and the Set Covering problems. These example formulations explain how to represent n BSUs with aggregation of m ASUs. It can also be argued that the aggregation error will decrease as m increases.

Table 1: Notation in location model

| Notation | Description |
|----------------------------------|---|
| $I = \{1, 2, 3, \dots, n\}$ | the set of all BSUs |
| $C = (c_1, c_2, \dots, c_n)$ | the list of BSU |
| w_i | demand/weight for the i^{th} BSU, $i = 1, 2, \dots, n$ |
| $I' \in \{1, 2, 3, \dots, m\}$ | the set of all ASUs |
| $C' = (c'_1, c'_2, \dots, c'_m)$ | the list of ASUs |
| A_k | the set of BSU in the k^{th} (ASU), $k = \{1, 2, 3, \dots, m\}$ with $\bigcup_{k=1}^m A_k = I$ |
| p | the number of facilities to be located |
| F | the optimal locations of the p facilities (i.e. optimal facility configuration) using the original formulation of the location model (i.e., full model) |
| F' | the optimal locations of the p facilities found with the aggregated location model |
| $D(F : c)$ | the distance between a BSU $c \in C$ and the closest element in F (i.e. $\text{Min}\{d(c, e), e \in F\}, c \in C$) |
| $D(F : c')$ | the distance between a BSU $c' \in C'$ and the closest element in F (i.e. $\text{Min}\{d(c', e), e \in F\}, c' \in C'$) |
| $D(F' : c)$ | the distance between a BSU $c \in C$ and the closest element in F' (i.e. $\text{Min}\{d(c, e'), e' \in F'\}, c \in C$) |
| $D(F' : c')$ | the distance between a BSU $c' \in C'$ and the closest element in F' (i.e. $\text{Min}\{d(c', e'), e' \in F'\}, c' \in C'$) |
| $f(F : C)$ | objective function evaluated using F and $D(F : c)$ |
| $f(F : C')$ | objective function evaluated using F and $D(F : c')$ |
| $f(F' : C)$ | objective function evaluated using F' and $D(F' : c)$ |
| $f(F' : C')$ | objective function evaluated using F' and $D(F' : c')$ |
| r | coverage distance of customers from a facility (for covering problems) |

Table 2: Objective functions for the original and the aggregated location models

| | Original location model | Aggregated location model |
|--------------|--|---|
| p -median | $f(F : C) = \sum_{i=1}^n w_i D(F, c_i)$ | Let $\hat{w}_k = \sum w_i : i \in A_k$ $f(F : C') = \sum_{k=1}^m \hat{w}_k D(F, c'_k)$ |
| p -centre | $f(F : C) = \max\{w_i D(F : c_i) : i \in I\}$ | Let $w'_k = \max\{w_i : i \in A_k\}$ $f(F : C') = \max\{w'_k D(F, c'_k) : k \in I'\}$ |
| set covering | Minimize $ F'' $ $st : D(F : c_i) \leq r_i : i \in I$ | Let $\rho_k = \min\{r_i : i \in A_k\}$ Minimize $ F $ $st : D(F : c'_k) \leq \rho_k : \forall k \in I'$ |

Francis *et al.* [47] provided a comprehensive review of aggregation error for location models. They described aggregation error measurements and surveyed some of the principal papers about aggregation errors. They also arranged the review into two main categories, namely median problems and centre/covering problems.

2.2. Aggregation Error Measurements

Replacing demand points (BSUs) by aggregated demand points (ASUs) on location problems introduce demand point aggregation error. This error affects model output, such as facility locations and inaccurate value of the objective function. There are several commonly used aggregation error measurements but no agreement on which measurement is the best. In the subsequent section, we review them.

Table 3 shows a variety of aggregation errors in location models. These are given in Francis *et al.* [47], Farinas and Francis [38], Hodgson *et al.* [64], and Casillas [16]. Calculating the distance between ASU and BSU (ASU-BSU distance) is a simple way to measure the aggregation error. If the set of location of the facilities has already been found, we can measure the difference in distance between the facilities and BSU, and between the facilities and ASU as another aggregation error (distance error). According to Table 1, we consider $f(F : C)$ as the objective of the original facility model and $f(F : C')$ as its approximation. From these objective functions, there are three basic aggregation error types, namely the absolute error, the relative error, and the maximum absolute error. The absolute error is defined as $ae(F) = |f(F : C') - f(F : C)|$. As the error can be negative or positive, the absolute function is adopted to avoid misinterpretation of the errors that could be caused by the cancellation of the negative against the positive errors. Moreover, the absolute function is useful in calculating the total error. Without the absolute function, the total error can even be zero or close to zero, which could be misleading by giving a wrong signal. The relative error allows us to know how far the errors from the objective function are spread. The relative error is usually converted to a percentage for simplicity. By the maximum error, the facility that provides the biggest error is determined, which represents the worst case scenario. One of the effects of aggregation error could lead to incorrect location of facilities. The easiest way to measure this error is to calculate the difference in the distance between F' and F .

Casillas [16] proposed the concept of cost error and optimality error on aggregation. These errors arise as a result from the ABC errors introduced by Hillsman and Rhoda [60], which we describe in the next subsection. The cost error can be defined by $ce = f(F' : C) - f(F' : C')$, the difference between the objective function evaluated using F' and $D(F', c)$ and the one using F' and $D(F', c')$. The difficulty to solve the original location problem lies behind this error. The optimality error (oe) is the difference between the objective function evaluated by using F and $D(F, c)$, and the one using F' and $D(F', c)$. It means that the original location problem has to be solved before the optimality error can be computed. Both of these errors usually convert to percent (ce and oe are divided by $f(F' : C)$ and $f(F : C)$

Table 3: Aggregation error measurements

| Aggregation Error Type | Aggregation Error Formulation |
|---|--|
| For general location problem: | |
| ASU-BSU distances | $d(c'_k, c_i), i \in N, k \in M, c_i \in C, c'_k \in C'$ |
| Distance error | $D(F, c'_k) - D(F, c_i), i \in N, \text{all } F, k \in M, c_i \in C, c'_k \in C'$ |
| Absolute error | $ae(F) = f(F : C') - f(F : C) , \text{all } F$ |
| Relative error | $rel(F) = ae(F)/f(F : C), \text{all } F$ |
| Maximum absolute error | $mae(f', f) = \max\{ae(F) : F, F \subset S, F = p\}$ |
| Location error | Difference between F' and $F, diff(F', F)$ |
| Cost error | $ce = f(F' : C) - f(F : C')$ |
| Optimality error | $oe = f(F : C) - f(F' : C)$ |
| Error bound eb | a number eb with $ae(F) \leq eb$ for all F |
| Ratio error bounds (when $f(F : C), f(F : C') > 0$) | $ f(F : C')/f(F : C) - 1 \leq eb/f(F : C)$ for all F $ f(F : C)/f(F : C') - 1 \leq eb/f(F : C')$ for all F |
| For median problems: | |
| BSU error | $e_i(F) = w_i[D(F, c'_k) - D(F, c_i)], i \in N, \text{all } F, k \in M$ |
| Total BSU error | $e(F) = \sum \{e_i(F) : i \in N\}$ all F |
| ABC error | $ebc_k(F) = w_k D(F, c'_k) - \sum \{w_i D(F, c_i) : i \in N_k\}$ N_1, \dots, N_m is a subset of $N = \{1, \dots, n\}$ for all $F, w_k \equiv \sum \{w_i : i \in N_k\}, N_k \subset N, k \in M$ decrease the potential facility number |
| D error | |
| For covering problems: | |
| Violation error | $VE_i(F) = (1/r)([D(F, c_i) - r]^+, i \in N$ where $[D(F, c_i) - r]^+ \equiv \max\{0, D(F, c_i) - r\}$ |
| Average violation error | $AVE(F) = \frac{\sum_{i=1}^n VE_i(F)}{n}$ |
| Maximum violation error | $MVE(F) = \max\{VE_i : i \in N\}$ |
| Coverage error | $CE(F) = \frac{ U(F) }{n}, U(F) = \{i \in N : D(F, c) > r\}$ |
| Conditional average violation error | $CAVE(F) = \begin{cases} \frac{\sum_{i=1}^n VE_i(F)}{ U(F) } & \text{if } U(F) \neq \phi \text{ (i.e. } U(F) \geq 1) \\ 0 & \text{otherwise } U(F) = \phi \end{cases}$ |

respectively). This can be used to evaluate the performance of a new approach on smaller instances or when the exact method is able to run a long time without computer failure.

Francis et al. [46] proposed error bounds for facility location models. These error bounds are a guide for demand point aggregation to keep the error small. Error bound (eb) is a given number such that $|f(F : C') - f(F : C)| \leq eb$ or $ae(F) \leq eb$ for all F . Ratio error bound can also be used instead of the error bound as the latter is easier to describe (i.e., 5% accuracy).

2.2.1. Aggregation error on the p -median problem

In the p -median problem, the easiest way to measure the error is to measure the distance between each BSU location and its weight ASU location. This is also known as the BSU error and is defined as $e_i(F) = w_i[D(F, c'_k) - D(F, c_i)]$. Hillsman and Rhoda [60] classify errors caused by aggregation in the location problems into three types, namely source A, B, and C errors. Later on, Hodgson et al.

[64] introduced another type of error occurring in discrete location problems, which they called Source D error. These four types of error, which will be used throughout this review, are described as follows.

- Source A error

This error occurs because of the loss of location information due to aggregation. It appears when, instead of the true average distance between a BSU and a facility to solve a facility location problem, the distance between an ASU and a facility is used. Figure 1 demonstrates the existence of Source A error. In the figure, it is assumed that the demand at BSU i , $i + 1$ and $i + 2$ has been aggregated as ASU k . Allocating ASU k to facility j means that all BSU i , $i + 1$ and $i + 2$ are allocated to facility j . Source A error is then defined as $|d(k, j)\hat{w}_k - \sum_{r=1}^{i+2} w_r d(r, j)|$, where $\hat{w}_k = w_i + w_{i+1} + w_{i+2}$. This error occurs when the distance between ASU k and facility j is not equal to the distances between BSU i , $i + 1$, and $i + 2$ and facility j .

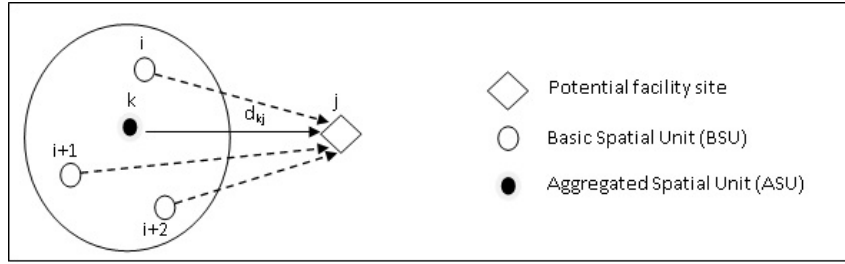


Figure 1: Existence of Source A Error

- Source B error

The loss of location information due to aggregation also leads to Source B error. This is a special case of Source A error. This error occurs when a facility is located at an aggregate spatial unit (ASU) (i.e., site $j \equiv$ site k). Figure 2 shows the existence of Source B Error where the demand at BSU i , $i + 1$ and $i + 2$ has been aggregated as ASU k . The figure also shows that facility j has been located at ASU k . However, the true distance from BSU i , $i + 1$ and $i + 2$ to facility j must be greater than zero. This is formally defined as $\sum_{r=i}^{i+2} w_r d(r, j) > \hat{w}_k d(k, j) = 0$ as $d(k, j) = 0$.

- Source C error

Source C error is also a direct result of the loss of location information because of aggregation. This happens when a basic spatial unit (BSU) is assigned to the wrong facility. For example, a given BSU is not assigned to the nearest facility but its corresponding ASU is. Figure 3 shows the

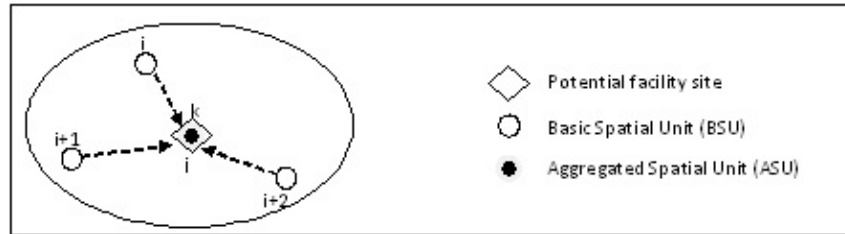


Figure 2: Existence of Source B Error

existence of Source C error where there are two ASUs, say ASU_k and ASU_{k+1} . Demand at BSU_r ($r = i, i+1$ and $i+2$) has been aggregated at the k^{th} ASU. On the other hand, demand at BSU_s ($s = i+3, i+4$ and $i+5$) has been aggregated at the $(k+1)^{th}$ ASU. ASU_k and ASU_{k+1} are then allocated to facility j and facility $j+1$, respectively. The k^{th} ASU is assigned to facility j , which therefore forces the $(i+2)^{th}$ BSU to be assigned to facility j , although this BSU is closer to facility $j+1$ than to facility j .

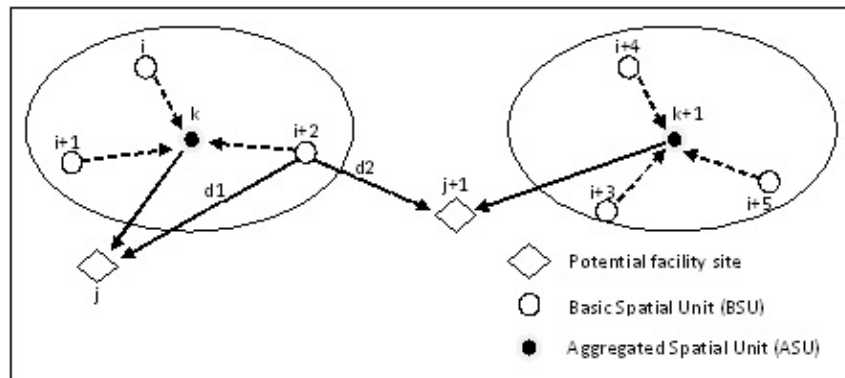


Figure 3: Existence of Source C Error

- Source D error

In discrete facility location problems, Hodgson et al. [64] introduced another error and named it Source D error. This occurs when a BSU happens to be also at a potential facility location. In other words, this error arises when the BSU locations themselves are potential sites, and hence the optimal configuration will be part of these sites. Conducting aggregation will decrease the number of potential facilities, but using an ASU as a potential location in facility location problems could lead to Source D error.

2.2.2. Aggregation error on covering problems

The covering problem aims at finding the minimum number of facilities such that each customer is covered by at least one facility. It means that facilities have a covering area, usually represented by a given radius (r). Figure 4 demonstrates the error on the covering problems.

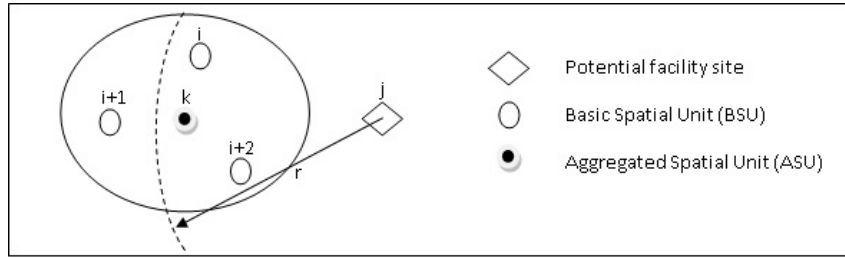


Figure 4: Example of error on the covering problems

The figure indicates that demand at BSU i , $i + 1$ and $i + 2$ has been aggregated at the k^{th} ASU. On the aggregated model, the k^{th} ASU is assigned to facility j . The figure shows that facility j can cover the k^{th} ASU because it is within r from the facility j . However, the error will occur when the k^{th} ASU is disaggregated (BSU i , $i + 1$ and $i + 2$ are also allocated to facility j). There is no error at the i^{th} BSU and the $(i + 2)^{th}$ BSU, but at the $(i + 1)^{th}$ BSU there exists an error $d(j, BSU_{i+1}) > r$. Farinas and Francis [38] define this error as the violation error at the $(i + 1)^{th}$ BSU, denoted by $VE_i(X)$, which is given as follows:

$$VE_i(F) = (1/r)([D(F, c_i) - r]^+ \text{ where } [D(F, c_i) - r]^+ \equiv \max(0, D(F, c_i) - r)$$

If the i^{th} BSU is covered by facility F within r then, $VE_i(X)$ is obviously zero.

On the covering problems, Farinas and Francis [38] also proposed other types of errors including the average violation error, the maximum violation error, the coverage error, and the conditional average error as defined in Table 3. It can be noted that these coverage based errors are highly likely to exist if an ASU is tightly covered and have some BSUs that are located on the opposite side of the facility that could not be easily covered, and hence generate such errors.

2.3. Aggregation Literature on Median Problems

In this section, we give an overview of some papers dealing with aggregation literature on the p -median problem, see Table 4 for a summary. Aggregation error was first formally defined by Hillsman and Rhoda [60], who aggregated the BSUs by constructing a grid of regular polygon over a planar distribution of BSUs by using the centroid in each polygon as the ASU position. The experiment showed that if a few ASUs were assigned to each server then the aggregation error was bigger. These errors are also usually used in many papers to measure the aggregation scheme performance.

Table 4: Aggregation literature on the p -median problem

| Authors | Journal | Year | Range of BSUs | Range of ASUs | Range of p | Setting |
|----------------------------|---|------|-----------------------|---------------|--------------|----------------------|
| Hillsman and Rhoda | Annals of Regional Science | 1978 | Uniformly distributed | 1 to 50 | N/A | Planar |
| Goodchild | Geographical Analysis | 1979 | 900 at most | 1 to 900 | 1 to 4 | Planar/ Network |
| Bach | Environment and Planning A | 1981 | 286 | 1262171 | 1 to 12 | Discrete |
| Mirchandrani and Reilly | Annals of Operations Research | 1986 | N/A | 1 | N/A | Planar |
| Current and Schilling | Geographical Analysis | 1987 | 681 | 3070 | 5,7,9 | Discrete |
| Casillas | Spatial analysis and location-allocation models | 1987 | 500 | 50-200 | 1,2,4,6 | Discrete |
| Ohsawa <i>et al.</i> | Geographical Analysis | 1991 | N/A | N/A | 1 | Planar |
| Francis and Lowe | Annals of Operations Research | 1992 | N/A | N/A | N/A | Network |
| Hodgson and Neuman | Location Science | 1993 | 27 | 6-25 | 3 | Discrete |
| Ballou | Transportation Journal | 1994 | 900 | 25-900 | 1-100 | Planar |
| Fotheringham <i>et al.</i> | Geographical Analysis | 1995 | 871 | 25-800 | 10 | Discrete |
| Francis <i>et al.</i> | Transportation Science | 1996 | 5500-69960 | 250-1225 | 1,3,5,7 | Planar |
| Hodgson <i>et al.</i> | The Canadian Geographer | 1997 | 688 | 158 | 1-25 | Discrete |
| Murray and Gottsegen | Geographical Analysis | 1997 | 913 | 100-800 | 10 | Discrete |
| Andersson <i>et al.</i> | Location Science | 1998 | 49,000 to 1,000,000 | 216, 376, 832 | 5,7 | Planar/ Network |
| Bowerman <i>et al.</i> | Computers and Operations Research | 1999 | 2630 | 209 | 5-50 | Discrete/ Network |
| Erkut and Bozkaya | Computers and Operations Research | 1999 | 668 | 50, 200, 300 | 2-6 | Discrete |
| Zhao and Batta | European Journal of Operational Research | 1999 | 5032 | 50 | 1,5 | Discrete |
| Francis <i>et al.</i> | Operations Research | 2000 | N/A | N/A | N/A | General |
| Zhao and Batta | Networks | 2000 | N/A | N/A | N/A | Network |
| Plastria | TOP | 2001 | 2000 | 400 | 1, 2, 5 | Planar |
| Hodgson | Annals of Operations Research | 2002 | N/A | N/A | 5-50 | Discrete |
| Hodgson and Hewko | Annals of Operations Research | 2003 | 722 | N/A | 1-25 | Discrete |
| Francis <i>et al.</i> | Naval Research Logistics | 2003 | 1193-250000 | 25-900 | 1, 3, 5 | Planar |
| Francis <i>et al.</i> | Annals of Operations Research | 2009 | N/A | N/A | N/A | General |
| Qi and Shen | European Journal of Operational Research | 2010 | 10k, 20k, and 30k | 3000, 5000 | N/A | Planar |
| Avella <i>et al.</i> | Computers and Operations Research | 2012 | Up to 89600 | 5000 | Up to 200 | Discrete |
| Irawan and Salhi | Journal of Global Optimization | 2013 | Up to 89600 | Up to 8960 | 25-100 | Discrete |
| Irawan <i>et al.</i> | European Journal of Operational Research | 2014 | Up to 89600 | Up to 8960 | 25-100 | Discrete |
| Salhi and Irawan | Computers and Operations Research | 2015 | Up to 71009 | N/A | 5-30 | Discrete |

The effects of aggregation error in median and centre problems were investigated by Goodchild [52]. He showed that on the median problems the effects of aggregation error are significant and include inaccurate value of the objective function, and inaccurate location of the facilities. He stated that 'aggregation tends to produce more dramatic effects on location than on the values of the objective function' (Goodchild, p. 253). Moreover, he also highlights that there is no aggregation scheme without a possible resulting error.

Bach [6] investigated the effects of different levels of aggregation and different types of distance measures for the discrete median problem, the centre problem, and the covering problem. He used data sets for the cities of Dortmund, Kleve, and Emmerich in Germany to analyse the location error and the objective function error. He concluded that 'the level of aggregation exerts a strong influence on the optimal locational patterns as well as on the values of the locational criteria'.

Mirchandani and Reilly [85] examined the effect of replacing the distances to demand points (BSUs) in a region by the distance to a single point (ASU), representing the region in a discrete location model. The continuously distributed demand points are used in their experiments.

Current and Schilling [23] proposed a method for eliminating source A, and source B errors. They introduced a novel way of measuring aggregated weighted travel distances for p -median problems. Let $d(i, j)$ denote the distance between the i^{th} and the j^{th} BSUs and $\tilde{d}(k, j)$ the distance between the representative point of the k^{th} ASU and the j^{th} BSU. The distance between the k^{th} ASU and the j^{th} facility is traditionally defined as:

$$\hat{d}(k, j) = \hat{w}_k \tilde{d}(k, j) \quad (1)$$

where $\hat{w}_k = \sum_{i \in A_k} w_i$ with A_k being the set of aggregated BSUs at the k^{th} ASU. To eliminate source A and B errors, the distance proposed in Current and Schilling [23] is set as:

$$\hat{d}(k, j) = \sum_{i \in A_k} w_i d(i, j) \quad (2)$$

Equation (2) measures the true weighted travel distance to the potential facility from all BSUs, aggregated at ASU k . This measurement method can also eliminate source B errors. For example, when the k^{th} ASU is also the j^{th} potential facility, the traditional measurement method would set $d(k, j) = 0$, whereas the improved method gives $d(k, j) \neq 0$, and hence measures the true weighted travel distance from all BSUs in the subset A_k to the facility located at the k^{th} ASU. Unfortunately, this method cannot eliminate source C errors.

As mentioned in the previous section, Casillas [16] showed that the A, B, and C errors cause two other types of error, namely the cost error ($ce = f(F' : C) - f(F' : C')$), and the optimality error. Aggregation effects are investigated based on 500 BSUs, which are randomly generated using $m = 50, 100, 150$, and 200 , and $p = 1, 2, 4$, and 6 . The results showed that the optimality error was small for small values of p , but the error increased when the values of p and m were larger.

Oshawa *et al.* [92] studied the location error and the cost error due to rounding (either rounded up or rounded down) in the unweighted 1-median and 1-centre problems in the one-dimensional continuous space. They denote aggregate data as rounded data. They also investigated the effects of aggregating BSUs into the midpoints of intervals of equal width. The main conclusions of their experiment are (i) rounding tends to exert more serious influence on the median problem than on the centre problem, and (ii) for median and centre problems, there was a pattern that the bigger location error implies smaller cost error.

Aggregation error bounds for the median and the centre problems were developed by Francis and Lowe [44]. Their study was focused on the network location problem. The error obtained from the worst objective function is used as the minimal error bound.

Hodgson and Neuman [62] introduced a Geographical Information System (GIS) method for eliminating source C error. The method spatially disaggregates data as needed during the solution procedure ('on the fly'). The method also uses Thiessen (Voronoi) overlay polygon, where every point within such a polygon is nearer to that polygon's centroid than to the centroid of any other polygon. It means that the disaggregation process is based on the membership in a polygon. Their method was applied to estimate the magnitude of cost estimate and optimality errors.

Transport costing error was investigated by Ballou [7] for the median problem. The transport costing error refers to the cost error as defined by Casillas [16], $ce = f(F' : C) - f(F' : C')$. Ballou used 900 three-digit zip codes as initial BSUs to cover a population of 248,000,000 people in the U.S. (Hawaii, Alaska, Puerto Rico, and APOs) in 1990. The weight of each BSU (zip code) is based on the population size. Coopers [22] location/allocation heuristic was also used to solve the median problem. Ballou found that the cost error increases as p and m increase.

Fotheringham *et al.* [41] examined the sensitivity of the median procedure (the objective function value and optimal locations) to the definition of spatial units for which the demand is measured (aggregation schemes). Data of 871 BSUs from Buffalo and New York census block was used to test the method. They aggregated it into 800, 400, 200, 100, 50, and 25 ASUs and used $p = 10$ to solve the median problems. The results showed that the level of aggregation affects the location error more significantly than the objective function value.

A median row-column aggregation method was introduced by Francis *et al.* [45] to find an aggregation that gives a small error bound, initially introduced by Francis and Lowe [44]. The authors deal with median problems with rectilinear distances and weight normalized to a total of unity. For given values on the number of rows (r) and the number of columns (c), the method, which is based on Hassin and Tamir [59], constructs an rc aggregation that minimizes the error bound. The value of r and c may not be equal, moreover the width of each column or row may also be different.

Hodgson *et al.* [64] studied the aggregation error effects on the discrete-space p -median model. The Canada census data for Edmonton is used in their experiment. By varying p , they calculate cost error ($ce = f(F' : C) - f(F' : C')$) and

optimality error ($oe = f(F : C) - f(F' : C)$). The results show that when p increases, the cost error decreases while the optimality error increases.

Murray and Gottsegen [88] investigated the influence of data aggregation on the stability of facility locations and objective function for the planar p -median model. Various levels of aggregation and various aggregation schemes for a fixed level of aggregation for the planar median problem were conducted. The result indicated that the value of the objective function ($f(F' : C)$) did not seem to vary significantly, although the facility locations varied as a result of the level of aggregation and aggregation method used. Like other researchers, they found that smaller values of m gives poorer results.

Demand point aggregation procedures for both the p -median and the p -centre for network location models were studied by Andersson *et al.* [2]. As the first step, they use 'row-column' method of Francis *et al.* [45] to obtain a coarse aggregation structure (the spacing of rows and columns of the grid). The next step is to locate the ASUs points on the subnetworks induced by the cells of the grid (using 1-median or 1-centre). They also use the concept of a network Voronoi diagram to find improved ASUs. They found that the level of aggregation affected the street network structure, and that the error estimates were not too sensitive to the value of p .

Bowerman *et al.* [10] investigated the demand partitioning method for reducing aggregation errors (source A, B, and C errors) in the p -median problems. They used the method of Current and Schilling [23] to eliminate source A and B errors, and the method of Hodgson and Neuman [62] to eliminate source C error. Data from the Central Valley of Costa Rica were used to examine their approach. The result showed that their method was better than that of Current and Schilling [23] in reducing error, however the computation time needed to solve the problems was recorded to be relatively higher.

A good review of aggregation errors for the p -median problem was provided by Erkut and Bozkaya [39]. They introduced six type of source errors, namely UD (assumption of uniform demand data), RA (use of a random aggregation method), FL (focusing on location errors), EC (emphasis on cost errors), DF (use of a different feasible solution set due to aggregation), and OA (aggregation level / over aggregation). For the p -median problem, they also proposed some guidelines (dos and don'ts) for aggregating spatial population data.

Zhao and Batta [115] performed a theoretical analysis of aggregation effects for the planar median problems. The worst and average case errors were also investigated with respect to centroid aggregation scheme and Euclidean distance. They produced the approximate distribution of the cost error for the 1-median problem, while the effect of source C error was closely examined for $p > 1$. Data location of houses in Buffalo, New York, Ontario, and California were used to test their analytical results.

Francis *et al.* [46] proposed a general model structure for location models and provided a theory to derive error bounds for all location models. They applied the idea of the triangle inequality with the SAND (SA short for subadditive and ND for nondecreasing).

Zhao and Batta [116] studied the p -median problem on a discrete and a continuous network where demand could be on the links of the network. They showed that the optimal solution can be approximated by a nodal solution. They also demonstrated that a model with continuous link demand can be transformed into an equivalent discrete link demand model. A method to aggregate demands on each link was also introduced, where they argued that their method did not introduce any aggregation errors to the problem solution.

Plastria [94] investigated how to minimise the aggregation error when selecting the ASUs location at which to aggregate given groups of BSUs. He studied the p -median problem with various distance measures derived from gauge functions. Some of his experiments focused on aggregating data at the centroid of the sub data set. He generated randomly 2000 BSUs in the plane and aggregated them to 400 ASUs. He measured source A and B errors by varying $p = 1, 3$, and 5. For the 1-median problem, he found that the aggregation error decreases as the distance between the facility and the ASU location (the centroid) increases.

Data surrogation error in p -median models was introduced by Hodgson [61]. This error occurs when an inappropriate variable is used to stand in for a target population's demand. He demonstrated the concept of this error for 25 Canadian cities where the total population data is used (in place of children or elderly citizens). He also identified the correlation of surrogation error. The conclusion of his research was that the level of surrogation error is related to the value of p , the dissimilarity of the target and surrogate distributions, city size, and the size of the service areas.

Hodgson and Hewko [63] studied aggregation and surrogation error in the p -median model using Edmonton, Canada data. The results showed that the surrogation error was a more serious problem than the aggregation error (Source A, B, and C error). They also proposed a disaggregation method to reduce aggregation and surrogation errors.

Francis et al. [48] developed the theory and algorithms to construct an aggregation that minimises the maximum of aggregation error for rectilinear distance for the 1-median problem. The method is based on row-column aggregations and uses the centroid as ASU location. The method does the aggregation for 1-median problems in the plane using aggregation results for 1-median problems on the line. The method for the 1-median problem is used to solve median problems for $p > 1$. By varying the value of $p = 1, 3$, and 5, they found that the error can be well-defined by a function a/mb , where a is a positive constant, $b \geq 1$, and m is the number of ASUs.

Qi and Shen [98] studied the worst-case analysis of demand point aggregation for the Euclidean p -median problem on the plane. They utilised a 'honeycomb heuristic' algorithm introduced by Papadimitriou [93] to develop a 'multi-pattern tiling' to attain smaller worst-case aggregation error bounds. The honeycomb heuristic works by partitioning a study area. The study area in which the demand points are distributed is partitioned into k sub-areas (hexagon polygons). All polygons are regular and congruent. The demand points (BSU) in each sub-area are represented by one median (ASU) at the centre of each polygon. They found

that the worst-case error bounds from the 'multi-pattern tiling' algorithms are smaller than that of the 'honeycomb heuristic' for arbitrarily distributed demand points.

An aggregation heuristic for large scale p -median problems was proposed by Avella *et al.* [4]. They introduced a new heuristic approach based on Lagrangean relaxation to deal with large-scale median problems. In this paper, they proposed three main procedures, namely sub-gradient column generation, combining sub-gradient optimization with column generation (core heuristic), and an aggregation heuristic. The main idea of sub-gradient column generation procedure is solving the LP over a subset of variables, implicitly considering other variables, and dynamically adding those variables when optimality conditions are violated. The core heuristic is defined by a subset of the most promising variables found according to the Lagrangean reduced costs, associated with the open facilities as well as those associated with the allocation variables. They implement Cplex with specific options to deal with larger instances. The experiments empirically show that the core heuristic does not obtain good solution when the value of p is relatively small. To overcome this drawback, an aggregation heuristic was introduced. It is based on solving the original problem with much larger p first. The obtained locations are then considered as centres for aggregation. The computational experiments show that their procedure provides better results compared to the solutions found by Hansen *et al.* [57], and Resende and Werneck [97].

Irawan and Salhi [67] utilised an aggregation technique and Variable Neighbourhood Search (VNS) for solving large-scale discrete p -median problems. A multi-stage methodology is designed, where learning from previous stages is taken into account when tackling the next stage. Each stage is made up of several aggregated problems that are solved by a mini VNS. In each stage, the solutions obtained from the aggregated problems are put together to make up a promising subset of potential facilities. VNS is used to solve this augmented p -median problem. This multi-stage process terminates when a certain criterion is met. The last stage is a post optimisation stage applied to the original (disaggregated) problem using the best solution from the previous stages as an initial solution.

Irawan *et al.* [68] introduced a multiphase approach that incorporates demand points aggregation, VNS, and an exact method for large unconditional and conditional p -median problems. The approach is made up of four phases. The first phase is similar to the first stage proposed by Irawan and Salhi [67] except that a more efficient implementation of the local search is adopted to generate promising facility sites, which are then used to solve a reduced problem in Phase 2 using VNS or an exact method. Phase 3 is an iterative learning process which tackles the aggregated problem using as the initial solution, the solution obtained from the previous phase. The last phase is a post optimisation phase where local search is applied for the original problem. The proposed approach is also adapted to cater for the conditional p -median problem with interesting results.

Salhi and Irawan [100] implemented a special data compression method using a quadtree-based method for allocating very large demand points to their nearest facilities while eliminating aggregation error. This allocation approach is effective

when solving large p -median problems in the Euclidean space. The allocation method aggregates demand points by eliminating aggregation-based allocation error, and disaggregating them if necessary. TSP datasets up to 71009 points are used for testing the method with encouraging results.

2.4. Aggregation Literature on Centre and Covering Problems

Some papers dealing with aggregation on centre and covering problems are discussed in this section. A list of aggregation literature on this topic is summarized in Table 5. Daskin *et al.* [32] investigated the aggregation effects for discrete planar maximum covering models. They measured the aggregation errors with the three different aggregation schemes, namely scheme A, scheme B, and Scheme C. The type of errors includes the optimality error, the coverage error introduced by Church and ReVelle [21], and the location error. Scheme A was based on the relative demands of BSUs only, whereas scheme B was solely based on the distances between the BSUs. Scheme C was based on both the demands and the distances between BSUs. All the three aggregation schemes were tested on 335 BSUs representing demand areas in the U.S. The results showed that for scheme A and scheme C, aggregation on demand and candidate locations produced small coverage or optimality errors. For all schemes and any level of aggregation, location errors are found to be big. Their finding confirms Goodchild's [52] results that 'aggregation has a greater effect on location decisions than on the values of the objective function'.

Current and Schilling [24] studied aggregation errors for the planar set covering and maximal covering location models. They proposed three rules on data aggregation, which reduce the aggregation errors. The rules were examined using 681 BSUs representing Baltimore City, Maryland. The result show that the aggregation rules reduce both problem size and aggregation error. They also observed source A, B, and C errors introduced by Hillsman and Rhoda [60], and defined their coverage counterparts. Source A errors occur if the facility covers the ASU but not the BSU, or the BSU is covered by a facility but not its associated ASU. Source B errors arise when a facility is located at an ASU. There might be some BSUs represented by this ASU that are not covered by this facility. In the covering problem, only source C error is not present.

Rayco *et al.* [96] studied a grid-positioning aggregation procedure for the centre problem with rectilinear distance. This procedure can also be utilised to estimate the maximum error, so letting the aggregation error be kept within tolerable limits. The procedure recognized an imposed grid structure on the plane. The cells of the grid structure were diamond-shaped and all of the same user-specified dimensions. The grid position was determined by minimizing an upper bound (eb). They used both computer-generated data sets and a real-world data set instances. The result showed that the rate of improvement in the error measures decreases as the number of ASUs increases.

Francis *et al.* [49] investigated a demand point aggregation analysis for a class of constrained location models. Here, the nearest distances of BSUs to facilities are used in the objective function, as well as in the constraints. They utilised and

improved the error bound introduced by Francis *et al.* [46], and observed the effect of aggregation errors in both the constraints and the objectives. The method was tested on the centre location models.

table Aggregation literature on the p -centre and set covering problems

| Authors | Journal | Year | Range of BSUs | Range of ASUs | Range of p | Setting |
|----------------------------|-----------------------------------|------|-----------------|--|--------------|----------|
| Daskin <i>et al.</i> | Annals of Operations Research | 1989 | 355 | 67, 201 | 4 | Discrete |
| Current and Schilling | Geographical Analysis | 1990 | 681 | 185, 415 | 30, 70 | Discrete |
| Rayco <i>et al.</i> | Location Science | 1997 | 5000 - 10000 | 25-2500 | 1, 3, 5, 7 | Planar |
| Rayco <i>et al.</i> | Computers and Operations Research | 1999 | 5000-10000 | 25-2500 | 1, 3, 5, 7 | Planar |
| Francis <i>et al.</i> | IIE Transactions | 2004 | N/A | N/A | N/A | General |
| Francis <i>et al.</i> | Geographical Analysis | 2004 | 50000 and 69960 | 50 - 900 in increment. of 50 | N/A | Planar |
| Emir-Farinas and Francis | Annals of Operations Research | 2005 | 50000 and 69960 | 50 - 900 in increment. of 50 | N/A | Planar |
| Plastria and Van-haverbeke | Network Spatial Economics | 2007 | 3337 and 19781 | 103, 111, and 128 also 340, 442, and 594 | 3 | Discrete |
| Francis <i>et al.</i> | Annals of Operations Research | 2009 | N/A | N/A | N/A | General |
| Irawan and Salhi | Journal of Heuristics | 2014 | Up to 71009 | Up to 7100 | 10-100 | Discrete |
| Salhi and Irawan | Computers and Operations Research | 2015 | Up to 71009 | N/A | 5-30 | Discrete |

Aggregation decomposition and aggregation guidelines for a class of minimax and covering location problems were studied by Francis *et al.* [50]. They used various distance types in the models on the plane. They proposed a method to find an aggregation to attain a small error bound value. The 'square root' formulas were introduced to support the aggregation procedure. The method can also accommodate aggregation decomposition for location problems involving multiple 'separate' communities. The method was tested on computer-generated data and real data. Firstly, they examined the method with 50,000 BSUs uniformly distributed in a square of dimensions 1,000 by 1,000 and varied the number of ASUs between 50 and 900. Secondly, the method was tested on 69,960 BSUs of power transformer locations in Palm Beach County, Florida. For the latter, they varied the number of ASUs between 50 and 3,250.

Farinas and Francis [38] studied aggregation methods with a priori error bounds for planar covering location models. They observed four types of aggregation error in the covering location problems, namely the average violation error (*AVE*), the maximum violation error (*MVE*), the coverage error (*CE*), and the conditional average violation error (*CAVE*). They pointed out which of the four errors should be considered most meaningful for a given situation. They also established three aggregation schemes, called Independent Projection Algorithm (IPA), Pick the Farthest (PTF), and Random Selection (RS). The schemes were tested on the data used in Francis *et al.* [50]. They found that the PTF scheme produced the smallest *AVE*, *CAVE*, and *CE*, whereas IPA performed better than PTF and RS with *MVE*.

Plastria and Vanhaverbeke [95] proposed a pre-processing aggregation method for competitive location models. The method prevents the loss of information of BSUs while aggregating BSUs, and hence avoids the possible loss of optimality. This method was applied to find the best location for a new hypermarket chain in Belgium. The experiment was first conducted for Brabant dataset, which is of a medium scale (3,337 BSUs), and then on the large scale Belgium dataset (19,781 BSUs). They concluded that their aggregation method besides being faster has a crucial influence on the size of BSUs.

Irawan and Salhi [69] proposed two meta-heuristics for large-scale unconditional and conditional vertex p -centre problems incorporating aggregation approach, Variable Neighbourhood Search, and exact method. Salhi and Irawan [100] also applied a quadtree-based approach for solving large p -centre problems in the Euclidean space.

2.5. Related Aggregation Work on Other Location Problems

Some interesting papers dealing with aggregation on other location problems do not fit within our classification schemes. They are just briefly discussed in this section, and their list is given in Table 6.

Sankaran [102] solved large instances of the capacitated facility location problem and proposed two types of methods. The first one relates to customer aggregation, while the second concerns the judicious selection of variable-upper-bounding constraints to be included in the initial integer-programming formulation. The results showed that both methods could be relevant in solving these large scale problems.

Limbourg and Jourquin [76] investigated aggregation errors and best potential locations on large networks in rail-road terminal locations when studying the p -hub median location problem. They proposed a method to separate the best potential locations in a hub-and-spoke network rail-road terminal location. The method uses two types of input, namely the flows and clustering based approaches to determine a set of potential locations for hub terminals. These potential locations are then utilised as an input in optimal location method. Data from trans-European networks was used to test their methods.

Gavriliouk [53] studied a method on aggregation to reduce aggregation errors in hub location problems. Moreover, the author proposed a heuristic (meta-

Table 5: Aggregation literature on others related location problems

| Authors | Journal | Year | Setting | Description |
|-----------------------|---|------|----------|---------------------------------------|
| Sankaran | European Journal of Operational Research | 2007 | Discrete | Capacitated Facility Location Problem |
| Limbourg and Jourquin | European Journal of Transport and Infrastructure Research | 2007 | Discrete | p -Hub Median Problem |
| Gavriliouk | Computers and Operations Research | 2009 | Network | Hub Location Problem |
| Zeng et al. | Geographical Analysis | 2010 | Network | Flow-Intercepting Problem |

algorithm) based on aggregation for p -hub centre problems and errors measure. The method was tested using generated large data sets uniformly distributed on a square of size $1,000,000 \times 1,000,000$, where total numbers of BSUs are 300, 400, 500, 600, and 1000. By varying $p = 2, 3, 5, 7, 10$, and 20, the number of ASUs is set to 10% of the number of BSUs. It was found that for each data set, the exact procedure (using CPLEX software) did not find a feasible solution within 5 minutes of CPU time, whereas the heuristic (meta-algorithm) method obtained solutions in each case, though the quality of the solution can not be judged.

Aggregating data for the flow-intercepting location model was studied by Zeng et al. [114]. Their research utilised GIS, optimization, and heuristic technologies to establish a method and a framework of aggregating data for the standard flow-intercepting location model. The authors applied the method to a real-world transportation system of Edmonton, Alberta, involving 395 traffic zones, 2,211 network nodes, 6,211 links, and 149,644 nonzero origin-destination (O-D) flow pairs for the afternoon traffic peak in 2001. This framework/method proved to be efficient in solving this real life problem.

3. A REVIEW ON SOLVING LARGE LOCATION PROBLEMS WITHOUT AGGREGATION

A review on solving large p -median problems without the use of aggregation techniques is presented here. This is followed by some studies on the p -centre problem.

3.1. A Review on solving p -median problems with a focus on large problems

The p -median problem is categorized as NP-hard (Kariv and Hakimi [72]). For relatively large problems, optimal solutions may not be found, and hence heuristic or metaheuristic methods are usually considered to be the best way to solve such problems. Mladenovic et al. [86] provided an excellent review on the p -median problem focusing on metaheuristic methods.

Table 6: Papers dealing with large p -median problem

| Authors | Journal | Year | Description |
|----------------------|-------------------------------------|------|---|
| Taillard E.D. | Journal of Heuristics | 2003 | Centroid Clustering Problem based heuristics |
| Avella <i>et al.</i> | Mathematical Programming | 2007 | Branch-and-Cut-and-Price Algorithm with reduction schemes |
| Hansen <i>et al.</i> | Data Mining and Knowledge Discovery | 2009 | Primal-dual variable neighbourhood search and decomposition/reduced VNS |
| Garcia <i>et al.</i> | INFORMS Journal on Computing | 2010 | Covering based with a radius formulation |

The interchange method is one of the most commonly used heuristic for solving the p -median problem. This method can be applied either alone or as a subroutine, as part of more complex methods (e.g. within metaheuristics). Whitaker [112] introduced a focal method known as the fast interchange heuristic. This method was applied by Hansen and Mladenovic [58] as a local search within a Variable Neighbourhood Search (VNS). The interchange local search using large neighbourhood structure was also suggested by Kochetov *et al.* [75]. An efficient implementation of the interchange method was produced by Resende and Werneck [97], who embedded an efficient data structure within the search to avoid recomputing already computed information. They used the interchange method within their proposed heuristics, which they refer to as the fast swap-based local search procedure. This heuristic is very efficient but could require an extra memory due to the use of a two dimensional matrix as part of its data structure.

In this section, we provide a few papers that deal with large p -median problems without using aggregation techniques. These papers are briefly summarised in Table 7.

Taillard [105] introduced heuristic methods to solve hard centroid clustering problems such as the p -median, the sum of square clustering, and the multi-source Webber problems. He proposed three methods for solving such problems, namely the candidate list search (CLS), local optimization (LOPT), and decomposition (DEC) procedures. CLS is based on a greedy procedure. This method randomly perturbs a solution that is locally optimal according to the alternate location-allocation procedure. LOPT optimizes the position of a given number of centres dynamically. The heart of LOPT is to choose a centre, a few of its closest centres and the set entities allocated to them to create a subproblem. DEC decomposes the problem into subproblems, which can be solved separately. His experiments show that these methods are very efficient and fast in producing better quality solutions for the medium size instances. The results for instances with more than 85,000 entities and 15,000 centres were also reported.

A computational study of large-scale p -median problems is conducted by Avella *et al.* [5]. They used Branch-and-Cut-and-Price algorithm to deal with such problems. The main components of this algorithm are delay column-and-row generation to avoid the excessive memory problem, and cutting planes to

strengthen the formulation. In the former component, this method exploits the special structure of the formulation to solve the LP-relaxation. The latter one aims to strengthen the formulation by limiting the size of the enumeration tree. The method provided good solutions for instances with the number of vertices being less than and equal to 3,795.

Hansen *et al.* [57] introduced a primal-dual variable neighbourhood search (VNS) metaheuristic for solving large p -median clustering problems. Within the search, decomposition is used to obtain better solutions and to reduce computational time. The authors used Reduced VNS to get good initial solutions, which are then used in their VNS with decomposition to tackle large problems. In addition, they provided good lower bounds via VNS to guarantee a small optimality gap. An efficient data structure based on Resende and Werneck procedure is implemented successfully within an existing local search. Their experiments show that VNS with decomposition is the best approach for solving very large instances. It is also observed that the difficulty of the problem depends not only on the value of n (the number of customers) but also on the value of p (the number of facilities).

Garcia *et al.* [51] investigated large p -median problems using a radius formulation. They proposed a model based on a covering-based formulation containing a small subset of constraints and variables. This method is found to be efficient due to a powerful branch-and-bound framework based on dynamic reliability branching within Cplex. Their experiments show that the method is able to solve large p -median problems ($n = 24,978$) especially when p is relatively large, as this tends to reduce the problem complexity within their formulation.

3.2. A review on solving p -centre problems

In this subsection, we review some papers that investigate the p -centre problem. Hakimi [54] initially introduced the p -centre problem where he investigated an absolute 1-centre problem on a graph. Minieka [82] proposed a basic algorithm based on solving a finite sequence of set covering problems for solving the problem when $p > 1$. The weighted case of the p -centre problem was investigated by Kariv and Hakimi [71], who concluded that the p -centre problem is NP-hard.

Polynomially bounded procedures for solving the p -centre and covering problems on a tree network were suggested by Tansel *et al.* [108]. Tansel *et al.* [106][107] provided an excellent review of network location problems including the p -centre problem. Two heuristics and an optimal algorithm to solve the p -centre problem for a given value of p in polynomial time in n were introduced by Drezner [29]. For relatively small p , algorithms for finding p -centres on a weighted tree were suggested by Jaeger and Kariv [70].

A useful and interesting recursive type algorithm using the Set Covering Problem (SCP) for attaining an optimal solution for the problem was designed by Daskin [25]. The approach is based on Minieka's method where the bisection technique is used to decrease the gap between upper and lower bounds. A spanning tree approach on cyclic networks was introduced by Bozkaya and Tansel [11]. Shaw [103] proposed a unified limited column generation approach for facility problems including the p -centre problem on trees.

Efficient exact methods for the vertex p -centre problem were studied by Daskin [26] and Ilhan and Pinar [66]. In the former, the problem was formulated as a maximum set covering sub-problem and then Lagrangean Relaxation is used to solve the problem. The latter designed an approach which comprises two stages, namely the LP-Phase and the IP-Phase, where in Stage 1 sub-problems with a certain covering distance are systematically discarded. A method called Dominant was introduced by Caruso *et al.* [15]. Efficient meta-heuristics (tabu search and variable neighbourhood search) were implemented by Mladenovic *et al.* [87] with excellent results. Minieka's approach is utilised by Elloumi *et al.* [37], incorporating a greedy heuristic and the IP formulation of the sub-problem for solving the problem optimally.

Al-Khedhairi and Salhi [3] proposed two enhancements to improve Daskin [25] and Ilhan and Pinar's [66] method. The objective of the enhancements is to decrease the number of calls to the SCPs. The first approach records the gaps in the distance matrix which are efficiently sorted, while the second approach explores appropriate jumps in the covering distance. An efficient approach by modelling the network as an interval graph was investigated by Cheng *et al.* [20]. Chen and Chen [18] suggested relaxation approaches for both the continuous and discrete p -centre problems. They solve optimally several smaller reduced problems first, then augmented them gradually by adding ' k ' customers at a time, where k is a parameter that needs to be defined, which raise the question of its value so as the choice of the ' k ' points to be added. The idea is that when the optimal solution of the subproblem happens to be feasible for the entire problem, the search terminates with the current solution as the optimal solution. The question is the choice of the value of k as well as the choice of the k points to be added.

Salhi and Al-Khedhairi [99] enhanced the method of Al-Khedhairi and Salhi even further by incorporating heuristic information into exact methods. Tight lower bounds are generated systematically once a good upper bound is found, making the search converge faster. Salhi and Sari [101] suggested a multilevel type meta-heuristic to obtain tight upper bounds which are then utilised to derive promising lower bounds. A bee colony optimization heuristic algorithm and a non-deterministic Voronoi diagram algorithm are investigated by Davidovic *et al.* [28] for the unconstrained and constrained p -centre problem respectively.

A double bounded method based on two-element restrictions was suggested by Calik and Tansel [14] to attain the optimal solution by solving a series of simple structured integer programs. Lu and Sheu [79] studied a robust vertex p -centre model for locating urgent relief distribution centres, while Lu [78] recently investigated a generalized weighted vertex p -centre model that represents uncertain nodal weights and edge lengths.

4. A REVIEW OF THE CONDITIONAL LOCATION PROBLEMS

Minieka [83] initially introduced the conditional location problem where he studied conditional centers and medians on a graph. Drezner [30] explained that conditional p -centre problems can be solved by solving $O(\log n)$ p -centre problems.

In other words, an effective algorithm for the p -centre problem can be adapted for the conditional problem.

An algorithm that requires the one-time solution of an unconditional $(p + 1)$ center or $(p + 1)$ median for solving the conditional $(p + 1)$ center or $(p + 1)$ median on networks was developed by Berman and Simchi-Levi [9]. Chen [17] designed a method for solving minisum and minimax conditional location-allocation problems with $p \geq 1$.

Drezner [33] developed a general heuristic for the conditional p -median problem on both network and the plane, where he introduced the term ' (p, q) median problem'. Let Q present the set of existing facilities where $Q \subset J$. Drezner [33] modified the objective function for the p -median problem as follow:

$$Z = \sum_{i \in I} w_i \left[\text{Min} \left\{ \min_{j \in Q} \{d(i, j)\}, \min_{j \in I, j \notin Q} \{d(i, j)\} \right\} \right] \quad (3)$$

Because $D_i = \min_{j \in Q} \{d(i, j)\}$ can be calculated for each $i \in I$ beforehand, equation (3) can be rewritten as :

$$Z = \sum_{i \in I} w_i \left[\text{Min} \left\{ D_i, \min_{j \in I, j \notin Q} \{d(i, j)\} \right\} \right] \quad (4)$$

The introduction of equation (4) makes computation more efficient as unnecessary calculations can be avoided.

A method for solving both the conditional p -median and p -centre problems was investigated by Berman and Drezner [8]. The method requires one-time solution of an unconditional p -median and p -centre problem incorporating the shortest distance matrix. Chen and Chen [19] proposed a relaxation-based algorithm for solving both the conditional discrete and continuous p -centre problems.

The conditional and unconditional p -centre problems using a modified harmony search algorithm is studied by Kaveh and Nasr [74]. Kaveh and Esfahani [73] also investigated a hybridization approach incorporating a harmony search and a greedy heuristic for solving conditional p -median problems. Recently, Irawan et al. [68] designed a multiphase approach using demand points aggregation, VNS, and an exact method for solving large conditional p -median problems. The method is tested on TSP datasets consisting of up to 71,009 points with various values of p .

5. CONCLUSIONS AND SUGGESTIONS

This paper presents a review of selected papers related to large-scale location problems focussing on the p -median, the p -centre, and related location problems. The division made in this review is mainly based on two categories of approaches, namely with and without aggregation. The former describes aggregation error measurements, papers related to aggregation on the p -median problem, aggregation on the p -centre, and the set covering problems, as well as other related

location problems. The latter discusses papers devoted to solving large p -median problems without aggregation and the p -centre problem in general. In addition, we also review some papers related to conditional location problems.

We highlight some research aspects that we believe to be worth pursuing. As optimal solutions can be found more efficiently for the vertex p -centre problem using the SCP-based methodology, approaches that integrate heuristics and exact method regarding aggregation, known as matheuristics, could be worth exploring. The above approach could also be applied to the region coverage problem instead of using points coverage. The partition of the area into a suitable number of appropriate polygons is part of the challenge. This problem could also be investigated for the case of continuous location where the facilities do not have to be on potential sites. This problem is obviously more difficult but mathematically more challenging and rewarding. In this review we focussed on problems with two dimensions, but can be extended to three dimensions to cater for the depth or the height of the siting of the facilities, as in the case of locating electrical protection devices on the posts, speed cameras on the roads or even safety cameras on the rail networks, among others. The methodology would be even more challenging if adapted to data mining. This is one of the hot topics that our location experience could be made use of.

REFERENCES

- [1] Aikens, C. H., "Facility location models for distribution planning", *European Journal of Operational Research*, 22 (1985) 263-279.
- [2] Andersson, G., Francis, R.L., Normark, T., and Rayco, M.B., "Aggregation method experimentation for large-scale network location problems", *Location Science*, 6 (1998) 25-39.
- [3] Al-Khedhairi, A., and Salhi, S., "Enhancements to two exact algorithms for solving the vertex p -center problem", *Journal of Mathematical Modelling and Algorithms*, 4 (2005) 129-147.
- [4] Avella, P., Boccia, M., Salerno, S., and Vasilyev, I., "An aggregation heuristic for large scale p -median problem", *Computers and Operations Research*, 39 (2012) 1625-1632.
- [5] Avella, P., Sassano, A. and Vasilyev, I., "Computational study of large-scale p -median problems", *Mathematical Programming*, 109 (2007) 89-114.
- [6] Bach, L., "The problem of aggregation and distance for analysis of accessibility and access opportunity in location-allocation models", *Environment and Planning A*, 13 (1981) 955-978.
- [7] Ballou, R. H., "Measuring transport costing error in customer aggregation for facility location", *Transportation Journal*, 33 (1994) 49-59.
- [8] Berman, O., and Drezner, Z., "A new formulation for the conditional p -median and p -center problems", *Operations Research Letters*, 36 (2008) 481-483.
- [9] Berman, O., and Simchi-Levi, D., "The conditional location problem on networks", *Transportation Science*, 24 (1990) 77-78.
- [10] Bowerman, R.L., Calamai, P.H., and Brent Hall, G., "The demand partitioning method for reducing aggregation errors in p -median problems", *Computers and Operations Research*, 26 (1999) 1097-1111.
- [11] Bozkaya, B., and Tansel, B., "A spanning tree approach to the absolute p -center problem", *Location Science*, 6 (1998) 83-107.
- [12] Brandeau, M. L., and Chiu, S.S., "An Overview of Representative Problems in Location Research", *Management Science*, 35 (1989) 645-674.
- [13] Brimberg, J., Hansen, P., Mladenovi, N., and Salhi, S., "A Survey of Solution Methods for the Continuous Location-Allocation Problem", *International Journal of Operations Research*, 5 (2008) 1-12.

- [14] Calik, H., and Tansel, B.C., "Double bound method for solving the p -center location problem", *Computers and Operations Research*, 40 (2013) 2991-2999.
- [15] Caruso, C., Colomni, A., and Aloii, L., "Dominant, an algorithm for the p -center problem", *European Journal of Operational Research*, 149 (2003) 53-64.
- [16] Casillas, P., "Aggregation problems in location-allocation modeling. In Gosh, A., and Rushton, G. (Eds.), *Spatial analysis and location-allocation models* (327-344)", Van Nostrand Reinhold, New York, 1987.
- [17] Chen, R., "Conditional minisum and minimax location-allocation problems in Euclidean space.", *Transportation Science*, 22 (1990) 158-160.
- [18] Chen, D., and Chen, R., "New relaxation-based algorithms for the optimal solution of the continuous and discrete p -center problems", *Computers and Operations Research*, 36 (2009) 1646-1655.
- [19] Chen, D., and Chen, R., "A relaxation-based algorithm for solving the conditional p -center problem", *Operations Research Letters*, 38 (2010) 215-217.
- [20] Cheng, "An improved algorithm for the p -center problem on interval graphs with unit lengths", *Computers and Operations Research*, 34 (2007) 2215-2222.
- [21] Church, R. L., and Reville, C., "The maximal covering location problem", *Papers of the Regional Science Association*, 32 (1974) 101.
- [22] Cooper, L., "Solutions of generalized locational: equilibrium models", *Journal of Regional Science*, 7 (1967) 2-18.
- [23] Current, J.R., and Schilling, D.A., "Elimination of source A and B errors in p -median location problems", *Geographical Analysis*, 19 (1987) 95-110.
- [24] Current, J.R., and Schilling, D.A., "Analysis of Errors Due to Demand Data Aggregation in the Set Covering and Maximal Covering Location Problems", *Geographical Analysis*, 22 (1990) 116-126.
- [25] Daskin, M.S., *Network and discrete location: models, algorithms, and applications*, Wiley, New York, 1995.
- [26] Daskin, M.S., "A new approach to solving the vertex p -center problem to optimality: algorithm and computational results", *Communications of the Operations Research Society of Japan*, 45 (1995) 428-436.
- [27] Daskin, M.S., "What you should know about location modeling", *Naval Research Logistics*, 55 (2008) 283-294.
- [28] Davidovic, T., Ramljak, D., Selmic, M., and Teodorovic, D., "Bee colony optimization for the p -center problem", *Computers and Operations Research*, 38 (2011) 1367-1376.
- [29] Drezner, Z., "The p -centre problem-heuristic and optimal algorithms", *Journal of the Operational Research Society*, 35 (1984) 741-748.
- [30] Drezner, Z., "Conditional p -center problems", *Transportation Science*, 23 (1989) 51-53.
- [31] Drezner, Z., "On the conditional p -median problem", *Computers and Operations Research*, 22 (1995) 525-530.
- [32] Daskin, M. S., Haghani, A. E., Khand, M., and Malandraki, C., "Aggregation Effects in Maximum Covering Models", *Annals of Operations Research*, 18 (1989) 115-139.
- [33] Drezner, Z., "On the conditional p -median problem", *Computers and Operations Research*, 22 (1995) 525-530.
- [34] Drezner, Z., and Hamacher, H.W., *Facility location : applications and theory*, Springer, Berlin, 2002.
- [35] Eilon, S., Watson-Gandy, C.D.T., and Christofides, N., *Distribution Management*, Hafner, New York, 1971.
- [36] Eiselt, H.A., Laporte, G., and Thisse, J., "Competitive location models: A framework and bibliography", *Transportation Science*, 27 (1993) 44.
- [37] Elloumi, S., Labbe, M., and Pochet, Y., "A new formulation and resolution method for the p -center problem", *INFORMS Journal of Computing*, 16 (2004) 84-94.
- [38] Emir-Farinas, H., and Francis, R., "Demand point aggregation for planar covering location models", *Annals of Operations Research*, 136 (2005) 175-192.
- [39] Erkut, E., and Bozkaya, B., "Analysis of aggregation errors for the p -median problem", *Computers and Operations Research*, 26 (1999) 1075-1096.
- [40] Farahani, R.Z., Masoud Hekmatfar, M., Fahimnia, B., and Kazemzadeh, N., "Hierarchical facility location problem: Models, classifications, techniques, and applications", *Computers and Industrial Engineering*, 68 (2014) 104-117.
- [41] Fotheringham, A. S., Densham, P. J., and Curtis, A., "The zone definition problem in location-allocation modeling", *Geographical Analysis*, 27 (1995) 60-77.
- [42] Francis, R.L., McGinnis, L.F., and White, J.A., "Locational analysis", *European Journal of Opera-*

- tional Research*, 12 (1983) 220-252.
- [43] Francis, R.L., McGinnis, L.F., and White, J.A., *Facility Layout and Location: an Analytical Approach*, Prentice-Hall, Englewood Cliffs, 1992.
 - [44] Francis, R. L., and Lowe, T. J., "On worst-case aggregation analysis for network location problems", *Annals of Operations Research*, 40 (1992) 229-246.
 - [45] Francis, R.L., Lowe, T.J., and Rayco, M.B., "Row-column aggregation for rectilinear distance p -median problems", *Transportation Science*, 30 (1996) 160-174.
 - [46] Francis, R. L., Lowe, T. J., and Rayco, A., "Aggregation error bounds for a class of location models", *Operations research*, 48 (2000) 294.
 - [47] Francis, R. L., Francis, T. J., Rayco, M. B., and Tamir, A., "Aggregation error for location models: survey and analysis", *Annals of Operations Research*, 167 (2009) 171-208.
 - [48] Francis, R.L., Francis, T.J., Rayco, M.B., and Tamir, A., "Exploiting self-canceling demand point aggregation error for some planar rectilinear median location problems", *Naval Research Logistics*, 50 (2003) 614-637.
 - [49] Francis, R.L., Lowe, and T.J., Tamir, A., "Demand Point Aggregation Analysis for a Class of Constrained Location Models: A Penalty Function Approach", *IIE Transactions*, 36 (2004) 601-609.
 - [50] Francis, R.L., Lowe, T.J., and Tamir, A., "Aggregation Decomposition and Aggregation Guidelines for a Class of Minimax and Covering Location Models", *Geographical Analysis*, 36 (2004) 332-349.
 - [51] Garcia, S., Labbe, M., and Marin, A., "Solving large p -median problem with a radius formulation", *INFORMS Journal on Computing*, 23 (2010) 546-556.
 - [52] GoodChild, M.F., "The aggregation problem in location-allocation", *Geographical Analysis*, 11 (1979) 240-255.
 - [53] Gavrilouk, E.O., "Aggregation in hub location problems", *Computers and Operations Research*, 36 (2009) 3136-3142.
 - [54] Hakimi, S.L., "Optimum locations of switching centers and the absolute centers and medians of a graph", *Operations research*, 12 (1964) 450-459.
 - [55] Hale, T.S., and Moberg, C.R., "Location Science Research: A Review", *Annals of Operations Research*, 123 (2003) 21-35.
 - [56] Handler, G.Y., and Mirchandani, P.B., *Location on networks: Theory and algorithms*, Cambridge: MIT Press, Cambridge, (1979).
 - [57] Hansen, P., Brimberg, J., Urosevic, D., and Mladenovic, N., "Solving large p -median clustering problems by primal-dual variable neighborhood search", *Data Mining and Knowledge Discovery*, 19 (2009) 351-375.
 - [58] Hansen, P., and Mladenovic, N., "Variable neighbourhood search for the p -median", *Location Science*, 5 (1997) 207-225.
 - [59] Hassin, R., and Tamir, A., "Improved complexity bounds for location problems on the real line", *Operations Research Letters*, 10 (1990) 395-402.
 - [60] Hillsman, E.L., and Rhoda, R., "Errors in measuring distances from populations to service centers", *Annals of Regional Science*, 12 (1978) 74-88.
 - [61] Hodgson, M. J., "Data surrogation error in p -median models", *Annals of Operations Research*, 110 (2002) 153-1650.
 - [62] Hodgson, M.J., and Neuman, S., "A GIS approach to eliminating source C aggregation error in p -median models", *Location Science*, 1 (1993) 155-170.
 - [63] Hodgson, M.J., and Hewko, J., "Aggregation and surrogation error in the p -median model", *Annals of Operations Research*, 123 (2003) 53-66.
 - [64] Hodgson, M.J., Shmulevitz, F., and Krkel, M., "Aggregation error effects on the discrete-space p -median model: The case of Edmonton, Canada", *Canadian Geographer / Le Gographe canadien*, 41 (1997) 415-428.
 - [65] Hotelling, H., "Stability in competition", *Economic Journal*, 39 (1929) 41-57.
 - [66] Ilhan, T., Pinar, M., "An efficient exact algorithm for the vertex p -center problem", <http://www.ie.bilkent.edu.tr/mustafap/pubs/>, (2001).
 - [67] Irawan C. A., and Salhi S., "Solving large p -median problems by a multistage hybrid approach using demand points aggregation and variable neighbourhood search", *Journal of Global Optimization*, 63 (3)(2015) 537-554.
 - [68] Irawan C.A., Salhi S., and Scaparra, M.P., "An Adaptive Multiphase Approach for Large Unconditional and Conditional p -Median Problems", *European Journal of Operational Research*, 237 (2014) 590-605.
 - [69] Irawan C.A., and Salhi S., "Hybrid Meta-heuristics with VNS and Exact Methods: Application

- to Large Unconditional and Conditional Vertex p -Centre Problems", *Journal of Heuristics*, (2014) under revision.
- [70] Jaeger, M., and Kariv, O., "Algorithms for finding p -centers on a weighted tree (for relatively small p)", *Network*, 15 (1985) 381-389.
 - [71] Kariv, O., and Hakimi, S.L., "An algorithmic approach for the vertex p -center problem. Part I: The p -centers", *SIAM Journal of Applied Mathematics*, 37 (1979) 513-538.
 - [72] Kariv, O., and Hakimi, S.L., "An algorithmic approach to network location problems. Part II: The p -medians", *SIAM Journal of Applied Mathematics*, 37 (1979) 539-560.
 - [73] Kaveh, A., and Esfahani, H. N., "Hybrid harmony search for conditional p -median problems", *International Journal of Civil Engineering*, 10 (2012) 32-36.
 - [74] Kaveh, A., and Nasr, H., "Solving the conditional and unconditional p -center problem with modified harmony search: A real case study", *Scientia Iranica*, 18 (2011) 867-877.
 - [75] Kochetov, Y., Alekseeva, E., Levanova, T., and Loresh, N., "Large neighbourhood search for the p -median problem", *Yugoslav Journal of Operations Research*, 15(1)(2005) 53-63.
 - [76] Limbourg, S., and Jourquin, B., "Rail-Road terminal locations: aggregation errors and best potential locations on large networks", *European Journal of Transport and Infrastructure Research*, 7 (2007) 317-334.
 - [77] Love, R., Moris, J., and Wesolowsky, G., *Facility location: model and methods*, North-Holland, Amsterdam, 1988.
 - [78] Lu, C., "Robust weighted vertex p -center model considering uncertain data: An application to emergency management", *European Journal of Operational Research*, 230 (2013) 113-121.
 - [79] Lu, C., and Sheu, J., "Robust vertex p -center model for locating urgent relief distribution centers", *Computers and Operations Research*, 40 (2013) 2128-2137.
 - [80] Mansfield, E., and Wein, H.H., "A Model for the Location of a Railroad Classification Yard", *Management Science*, 4 (1958) 292-313.
 - [81] Miehle, W., "Link-Length Minimization in Networks", *Operations research*, 6 (1958) 232-243.
 - [82] Minieka, E., "The m -center problem", *SIAM Review*, 12 (1970) 138-139.
 - [83] Minieka, E., "Conditional Centers and Median on a Graph", *Network*, 10 (1980) 265-272.
 - [84] Mirchandani, P.B., and Francis, R.L., "Discrete location theory", New York: Wiley-Interscience, (1990).
 - [85] Mirchandani, P. B., and Reilly, J. M., "'Spatial nodes' in discrete location problems", *Annals of Operations Research*, 6 (1986) 203-222.
 - [86] Mladenovic, N., Brimberg, J., Hansen, P., and Moreno-Perez, J.A., "The p -median problem: A survey of metaheuristic approaches", *European Journal of Operational Research*, 179 (2007) 927-939.
 - [87] Mladenovic, N., Labbe, M., and Hansen, P., "Solving the p -center problem with tabu search and variable neighbourhood search", *Network*, 42 (2003) 48-64.
 - [88] Murray, A. T., and Gottsegen, J. M., "The influence of data aggregation on the stability of p -median location model solutions", *Geographical Analysis*, 29 (1997) 200-213.
 - [89] Murray, A.T., Wei, T., "A computational approach for eliminating error in the solution of the location set covering problem", *European Journal of Operational Research*, 224 (2013) 52-64.
 - [90] Murray, A.T., OKelly, M.E., Church, R.L., "Regional service coverage modeling", *Computers and Operations Research*, 35 (2008) 339-355.
 - [91] Nickel, S., and Puerto, J., *Location Theory: a unified approach*, Springer, Berlin, 2005.
 - [92] Oshawa, Y., Koshizuka, T., and Kurita, O., "Errors caused by rounded data in two simple facility location problems", *Geographical Analysis*, 23 (1991) 56-73.
 - [93] Papadimitriou, C.H., "Worst-case and probabilistic analysis of a geometric location problem", *SIAM Journal of Computing*, 10 (1981) 542-557.
 - [94] Plastria, F., "On the choice of aggregation points for continuous p -median problems: a case for the gravity center", *TOP*, 9 (2001) 217-242.
 - [95] Plastria, F., and Van Haverbeke, L., "Aggregation without Loss of Optimality in Competitive Location Models", *Networks and Spatial Economics*, 7 (2007) 3-18.
 - [96] Rayco, M.B., Francis, R.L., and Tamir, A., "A p -center grid-positioning aggregation procedure", *Computers and Operations Research*, 26 (1999) 1113-1124.
 - [97] Resende, M.G.C. and Werneck, R., "A fast swap-based local search procedure for location problems", *Annals of Operations Research*, 150 (2007) 205-230.
 - [98] Qi, L., and Shen, Z.M., "Worst-case analysis of demand point aggregation for the Euclidean p -median problem", *European Journal of Operational Research*, 202 (2010) 434-443.
 - [99] Salhi, S., and Al-Khedhairi, A., "Integrating heuristic information into exact methods: The case of the vertex p -centre problem", *Journal of the Operational Research Society*, 61 (2010) 1619-1631.

- [100] Salhi, S., and Irawan, C.A., "A quadtree-based allocation method for a class of large discrete Euclidean location problems", *Computers and Operations Research*, 55 (2015) 2335.
- [101] Salhi, S., and Sari, M., "A Multi-Level Composite Heuristic for the Multi-Depot Vehicle Fleet Mix Problem", *European Journal of Operational Research*, 103 (1997) 95-112.
- [102] Sankaran, J.K., "On solving large instances of the capacitated facility location problem", *European Journal of Operational Research*, 178 (2007) 663-676.
- [103] Shaw, D.X., "A unified limited column generation approach for facility location problems on trees", *Annals of Operations Research*, 87 (1999) 363-382.
- [104] Sridharan, R., "The capacitated plant location problem", *European Journal of Operational Research*, 87 (1995) 203-213.
- [105] Taillard, E.D., "Heuristic methods for large centroid clustering problems", *Journal of Heuristics*, 9 (2003) 51-73.
- [106] Tansel, B.C., Francis, R.L., and Lowe, T.J., "State of the art location on network: a survey. Part I: the p -center and p -median problems", *Management Science*, 29 (1983) 482-497.
- [107] Tansel, B.C., Francis, R.L., and Lowe, T.J., "State of the art location on network: a survey. Part II: exploiting tree network structure", *Management Science*, 29 (1983) 498-511.
- [108] Tansel, B.C., Francis, R.L., Lowe, T.J., and Chen, M., "Duality and distance constraints for the nonlinear p -center and covering problem on a tree network", *Operations Research*, 30 (1982) 725-744.
- [109] Valinsky, D., "A Determination of the Optimum Location of Fire-Fighting Units in New York City", *Journal of the Operations Research Society of America*, 3 (1955) 494-512.
- [110] Weber, A., "Über den Standort der Industrien 1", *Teil: Reine theorie des standortes*, Tübingen, Germany, 1909.
- [111] Weiszfeld, E., "Sur le point pour lequel la somme des distances de n points donnés est minimum", *Tohoku Mathematical Journal*, 43 (1937) 355-386.
- [112] Whitaker, R.A., "A fast algorithm for the greedy interchange for large-scale clustering and median location problems", *INFOR*, 21 (1983) 95-108.
- [113] Young, H.A., "On the Optimum Location of Checking Stations", *Operations research*, 11 (1963) 721-731.
- [114] Zeng, W., Castillo, I., and Hodgson, M.J., "Aggregating Data for the Flow-Intercepting Location Model: A Geographic Information System, Optimization, and Heuristic Framework", *Geographical Analysis*, 42 (2010) 301-322.
- [115] Zhao, and P., Batta, R., "Analysis of centroid aggregation for the Euclidean distance p -median problem", *European Journal of Operational Research*, 113 (1999) 147-168.
- [116] Zhao, and P., Batta, R., "An aggregation approach to solving the network p -median problem with link demands", *Networks*, 36 (2000) 233-241.