PHILIPPINE ENGINEERING JOURNAL
PEJ 2006; **Vol. 27 No. 2**:1–26

AN OPPORTUNITY COST-BASED GENETIC ALGORITHM FOR A MODIFIED CAPACITATED p-MEDIAN PROBLEM

Joanna Zomil Resurreccion*

Department of Industrial Engineering and Operations Research
University of the Philippines Diliman
Quezon City 1101 PHILIPPINES

ABSTRACT

Median problems are combinatorial problems of searching for p facility locations (medians) that will serve a network of n demand nodes at a minimum cost. A Capacitated p-Median Problem (CPMP) allocates the demand of all nodes to the p facilities subject to the service capacity of each facility. In view of a future increase in demand of the network, this study presented a modified CPMP, called the CPkMP, which incorporates the network's existing number of k facilities in search of new and additional $p-k$ facility locations. This study evaluated the performance and applicability of a recently developed genetic algorithm (GA) that was used for a non-capacitated p-Median Problem (PMP) in order to solve the CPkMP. The proposed GA completes a combination of the needed $p-k$ medians by implementing a child generation procedure that is based on *opportunity costs*. When the closest facility location that can satisfy the demand of a node is removed from the list of candidate locations, the demand is reassigned to another facility farther from the node, and the lost opportunity of not choosing the closest facility to minimize the distance traveled is called the *opportunity cost (OC)*. The opportunity cost genetic algorithm (OC-GA) removes from the set of candidate locations a candidate with the lowest total opportunity costs arising from the additional distance of reassigning the network demands to farther candidate facilities. Computational tests of the OC-GA on ten CPMP problems from literature with known optimal solutions showed a 4.5% relative error from the known optimal solution. When the locations for the k existing facilities are selected from the optimal solution of CPMP, the needed additional $p-k$ facilities in the CPkMP tend to converge to the remaining members of this optimal solution. More so, when the k existing facilities are different from the medians in the optimal CPMP solution, 50 to 70 percent of the searched $p-k$ locations are part of the CPMP optimal solution.
Key Words: genetic algorithm, capacitated p-median problems, opportunity cost

1. RATIONALE

A p-median problem (PMP) is a combinatorial problem of finding the locations of p facilities or medians that will serve a network of n demand nodes at a minimum cost. This cost is the sum of the weighted distances of each demand node to a nearest facility. The search for p medians in a PMP assumes no existing facilities to cover any demand. The PMP, in general,

*Correspondence to: Department of Industrial Engineering and Operations Research, University of the Philippines Diliman, Quezon City 1101 PHILIPPINES. email: joanna.zomil@up.edu.ph

Received February 13, 2008
Revised May 7, 2008
Accepted June 3, 2009

refers to a network having infinite capacities for its service facilities, which is described as noncapacitated. When the demand that can be allocated to a facility is constrained by the capacity of the facility, the PMP is referred to as a capacitated p-median problem (CPMP). PMP and CPMP solutions have potential applications in telecommunications, transportation, manufacturing, and distribution networks where the number of demand points is relatively larger than the number of service facilities (Hakimi, 1964, 1965; Maranzana, 1964).

A network whose total demand is currently satisfied by an existing number of k facilities may experience changes in its demand over time. As this demand increases, it is essential to create new additional facilities to satisfy the higher demand, raising the number of available facilities from k to p. Furthermore, decisions regarding the locations for the new facilities and the retention of some, if not all, of the existing ones have to be made (e.g., because of provisions in law, investment, and other network constraints). In some networks where there are no existing k facilities at such stage and p new medians are required, a number of k identified prime locations set by some criteria may be assumed to represent "'existing"' facilities so that the additional $p-k$ locations that go best with these assumed k facilities can be found. This problem of searching for medians or new locations in addition to an existing number of k facilities as constrained by the service capacity of each facility is presented in this study as the capacitated *$p-k$ median problem* (CPkMP).

The high investment and operational costs make it impractical to build a service facility on every demand node even if this means achieving a zero distance between the service facility and the demand node. Thus, in a CPMP, only p candidate medians ($p < n$) are chosen to serve the demand of n nodes. In the CPkMP solution, $p-k$ candidates are combined with the known existing k facilities to complete the set of p medians. The demand of each node is assigned to the closest median among the p medians of the solution to minimize cost for both cases of no existing facilities for the CPMP and with k existing facilities for the CPkMP.

The CPMP differs from the CPkMP on two aspects. First, the number of possible candidate medians for the CPkMP solution is smaller than the number of possible candidate medians for the CPMP solution. There are n candidates to choose from in a CPMP whereas there are only $n-k$ candidates to choose from in the CPkMP. Second, the CPkMP chooses only $p-k$ facilities out of the $n-k$ candidate locations while the CPMP selects p facility locations from n candidates. The CPMP has a total number of solutions equal to the combination of n taken p. For example, the solution space of a 30 node network with a 10 median requirement results in approximately 30 million possible solutions (i.e., the combination of 30 taken 10). As for a CPkMP, if a network has k existing facilities, the solution space for searching additional $p-k$ facilities from $n-k$ candidates reduces to the combination of $n-k$ taken $p-k$. If five facilities already exist in the previous example, the solution set is reduced to around 53,000 solutions (i.e., the combination of 25 taken 5). The differences in the number of possible candidates and the number of candidates required by each solution resulted in the larger solution space and greater search complexity of the CPMP as compared to the CPkMP. These differences in solution space and search complexity brought about by considering the existing facilities of a given network are in agreement with the observations of Resurreccion and Resurreccion (2004) for the noncapacitated PMP and PkMP (a noncapacitated PMP with k existing facilities).

Furthermore, the CPkMP is also different from a PkMP since the search for the $p-k$ medians for the CPkMP is dependent on the capacity of the existing facilities to accommodate the demand of the network (i.e., the capacity constraint of the existing facilities).

On the basis of the aspects mentioned above, the CPkMP deviates from the CPMP in searching for new locations to complete a set of p medians. The CPMP and the CPkMP are two entirely different problems when it comes to the size of the solution space, capacity constraint and problem complexity. Thus, the CPkMP can not be simply reduced to a CPMP assuming that no facilities exist in searching for the additional $p - k$ medians. Due to the inclusion of the capacity constraint, CPkMP cannot be treated similar to the noncapacitated PkMP. The lack of studies in literature related to the CPkMP to address these concerns has led the author into this study.

2. REVIEW OF RELATED LITERATURE

A significant component of this study is the general complexity of the PMP, CPMP and CPkMP. While the PMP can be solved in polynomial time on a tree network, it is NP-hard on a general graph (Kariv and Hakimi, 1979; Garey and Johnson, 1979). Hakimi (1964) was the first to present the formulation of the PMP. Daskin (1995) discussed three classes of heuristic algorithms in solving the PMP, namely myopic, exchange heuristic and neighborhood search algorithms. When coupled with any of the three heuristic algorithms, the Lagrangean relaxation approach (Beasley, 1985) was found out to be superior to many proposed algorithms with respect to the quality of solution but requiring a large computational time (Daskin, 1995), which is in agreement with Senne and Lorena (2000).

The computational complexity of the PMP can be extended to the CPMP where the number of combinations is also polynomial in n for constant values of p (Daskin, 1995; Mirchandani and Francis, 1990). Results comparing the solution set sizes for varying values of p and k in Resurreccion and Resurreccion (2004) also confirmed this complexity over a noncapacitated PkMP (i.e., PMP with k existing facilities). The difference of a CPkMP with PkMP is that the amount of demand that can be assigned to a facility is limited to the facility's capacity. However, for the same values of p and k, the CPkMP and PkMP have the same number of candidate medians and require the same number of additional medians. Therefore, their solutions have the same composition and solution set sizes. Further, the complexity of a PkMP is also polynomial in n for constant values of p and k (Resurreccion and Resurreccion, 2004).

Ideal for the complexity of median problems is the use of genetic algorithm (GA). The GA maintains and improves a subset of solutions from an otherwise very large complex set of possible median combinations (Holland, 1975; Goldberg, 1989). The subset consists of solutions that are naturally selected under a controlled and well-defined environment (Holland, 1975). For the survival and reproduction of better solutions, GA randomly mixes and recombines the attributes (medians) of the solutions from this subset resulting into a new solution or *child*. The child is evaluated based on a certain criterion or *fitness* to determine if it is better than any other solution from the current subset. A series of recombination eventually yields better children that are accepted as members of an improved subset of solutions, from which one of the solutions satisfies the total demand of the network at a minimum cost.

Though several authors have applied GA to solve the p-Median problem (Dvorett, 1999; Bozkaya et al., 2001; Lorena and Fortado, 2001; Berman et al., 2002; Alp et al., 2003), only a few studies have used GA on a CPMP (Correa et al., 2004; Goseiri and Ghannadpour, 2007). These studies, however, have assumed no existing facilities to satisfy any fraction of the network's demand.

Resurreccion and Resurreccion (2004) implemented an opportunity cost-based GA (called

OC-GA) to the non-capacitated p-median test problems of Galvao and Revelle (1996) to investigate the inclusion of a k number of existing facilities on a PMP (referred to as the PkMP). Since there is a lack of related studies in PkMP and CPkMP, the OC-GA can only be compared with the available competing methods on cases when k is equal to zero. However, Resurreccion and Resurreccion (2004) have provided comparison between OC-GA against competing methods for PMP and showed that the OC-GA performed at 0.52% deviation from the Lagrangean relaxation solution with no greater than 1.06% difference in minimum cost values. Similarly, OC-GA performed better than the combined myopic and neighborhood search algorithms for values of p and k that are higher than 10 and 5, respectively (Resurreccion and Resurreccion, 2004).

3. OBJECTIVES

The objectives of this study are (1) to further develop an opportunity cost-based genetic algorithm (hereby called OC-GA) that was recently used to solve a noncapacitated PMP with k existing facilities (PkMP) to solve the CPkMP, (2) to evaluate the performance of this OC-GA on CPMP test problems that have known optimal solutions, (3) to investigate the reliability and convergence of the OC-GA when applied to CPkMP with the k medians selected from the optimal solution of the corresponding CPMP, and (4) to use OC-GA as a tool in providing the optimal possible combination of existing and new facility locations in serving the demand of a network.

The study provides a formulation for the CPkMP and verifies the applicability of a recently-developed opportunity cost child generation procedure for a genetic algorithm (OC-GA) in solving the CPkMP. In this study, 10 50-node and one 100-node network test problems with known solutions were used for model validation purposes. A practical application of the OC-GA is introduced through a case study of a distribution company engaged in the delivery of food and beverage to the various stores of a food chain in Metro Manila. As there have been changes in the demand of the network both in terms of quantity and roster of stores over the past five to eight years, the study uses OC-GA to investigate the suitability of the existing distribution centers in satisfying the demand of the larger network and evaluates the best candidates for placing new additional facilities in the future.

4. MODEL FORMULATION

4.1. Capacitated $p - k$ Median Problem Formulation

In a CPMP, the cost associated in satisfying the demand of a node from a facility is proportional to the distance between the demand node and the nearest facility (Daskin, 1995). To minimize the cost, the facility closest to a node must serve the demand of the node without exceeding the service capacity of the facility. This cost is given by the product of the demand and the distance of the node from the facility. The integer (0-1) programming formulation of the CPMP adopted from Revelle and Swain (1970) with notations for the inputs and decision variables from Daskin (1995) is given as follows,

$$\text{Minimize} \quad z = \sum_{i=1}^{n} \sum_{j}^{n} h_i d_{ij} y_{ij} \tag{1}$$

$$\text{Subject to :} \quad \sum_{j=1}^{n} y_{ij} = 1 \qquad \forall i \tag{2}$$

$$\sum_{j=1}^{n} x_j = p \tag{3}$$

$$y_{ij} - x_i \leq 0 \qquad \forall i, j \tag{4}$$

$$\sum_{i=1}^{n} h_{ij} y_{ij} \leq c_j x_j \qquad \forall j \tag{5}$$

$$x_j = 0, 1 \qquad \forall j \tag{6}$$

$$y_{ij} = 0, 1 \qquad \forall i, j \tag{7}$$

where n = the total number of demand nodes in the network
i = index of demand node location ranging from 1 to n
j = index of candidate facility location ranging from 1 to n
h_i = demand at node i
d_{ij} = distance between demand node i and candidate location j
c_j = capacity of facility at candidate location j
p = number of facilities to locate

The binary decision variables are represented as,

$$x_j = \begin{cases} 1, & \text{if a facility is to be located at location } j \\ 0, & \text{if not} \end{cases}$$

$$y_{ij} = \begin{cases} 1, & \text{if the demand of node } i \text{ is served by a facility at location } j \\ 0, & \text{if not} \end{cases}$$

The *cost function* z in Eq. (1) is the sum of the demand weighted distance between each of the demand nodes and the nearest median to each node. Eq. (2) guarantees that exactly one median serves the demand of each node. Eq. (3) requires that a total of p medians serve the entire demand of the network. Eq. (4) prevents the allocation of demand to a location that is not a median. Eq. (5) constraints the total demand assigned to median j to be no more than the service capacity if median j exists. The last two sets of constraints, Eq. (6) and Eq. (7) are standard binary constraints for the decision variables. When the fixed capacity of a median is exceeded, the demands of the remaining nodes in the network are assigned to the nearest median with a capacity not yet entirely consumed.

Since the existing k facilities are already part of the needed p medians in CPkMP, the variable x_j in Eq. (6) is set to a value of 1 for every location index j where a facility exists. The existing facilities satisfy their own demands such that values of y_{ij} are set to 1 in Eq. (7) whenever $x_j = 1$ and $i = j$.

The classical assignment of demand to a median can be described with the following additional notations and by the following pseudo code of Resurreccion and Resurreccion, (2004) and Ghoseiri and Ghannadpour, (2007).

Let $\mathbf{X_p} = \{X_{p,1}, X_{p,2}, X_{p,3}, ..., X_{p,p}\}$ be the median set of facility location indices
$\mathbf{C_p} = \{c_{p,1}, c_{p,2}, c_{p,3}, ..., c_{p,p}\}$ be the set of median capacities
$\mathbf{H_p} = \{h_{p,1}, h_{p,2}, h_{p,3}, ..., h_{p,p}\}$ be the set of demands of the medians
$\mathbf{H} = \{h_1, h_2, h_3, ..., h_{n-p}\}$ be the demand set (i.e., demand points) of all nodes in the network excluding the demands of the p medians

Read

$\mathbf{X_p} = \{X_{p,1}, X_{p,2}, X_{p,3}, ..., X_{p,p}\}$
$\mathbf{C_p} = \{c_{p,1}, c_{p,2}, c_{p,3}, ..., c_{p,p}\}$
$\mathbf{H_p} = \{h_{p,1}, h_{p,2}, h_{p,3}, ..., h_{p,p}\}$
$\mathbf{H} = \{h_1, h_2, h_3, ..., h_{n-p}\}$

do

$c_{p,i} := c_{p,i} - h_{p,i}$
while $i := 1...p$

do

$$m = \text{index}\left(\min_{i=1...p} \{h_j d_{ij} \text{ such that } c_{p,i} - h_j \geq 0\}\right)$$

(i.e., a demand point is assigned to a median m)

$c_{p,m} := c_{p,m} - h_j$
while $j := 1...n$
output := *cluster of median r and demand points serviced by median m,*
where $m = 1..p$

This pseudo code of the classical assignment of demands begins with the input of data of the identified facility location indices as set $\mathbf{X_p}$, the corresponding capacities of the facilities as set $\mathbf{C_p}$, and the demand of the serving facilities as set $\mathbf{H_p}$. A separate set, $\mathbf{H}$, defines the demands of the nodes where no facility is to be located (i.e., $x_j = 0$). Since the facilities must first serve their own demand to minimize distance travel, each demand from the set $\mathbf{H_p}$ must be assigned to the corresponding facility in the set $\mathbf{X_p}$. This leaves a remaining unassigned capacity for each median as $c_{p,i} = c_{p,i} - h_{p,i}$ (i.e., capacity of ith median from the set $\mathbf{X_p}$ minus the demand of the median). Then, the demands of the nodes from the set $\mathbf{H}$ are allocated to the nearest facility in the set $\mathbf{Xp}$ that has a remaining unassigned capacity greater than or equal to the demand (i.e., $c_{p,i} - h_j \geq 0$). Every time a demand is assigned to a facility, the unassigned capacity of that facility is reduced and updated as $c_{p,i} = c_{p,i} - h_j$. The assignment of demands ends when all demand nodes have been assigned to a particular median in $\mathbf{X_p}$.

The median set $\mathbf{X_p}$ that corresponds to the set of p required facility locations form one solution. Since there are a total of n location indices in the network, the decision variables, x_j, of the integer programming (IP) model represents the solution as a collection of n binary numbers. Eq. (6) and Eq. (7) show that for each value of location index j, x_j is either 0 (i.e. no facility at j) or 1 (i.e. there is a facility at location j). For example, a solution identifies the candidate location indices of j = 3, 6, and 10 to be the facility locations for a 10-node network. Thus, $x_3 = x_6 = x_{10} = 1$ and $x_j = 0$ for other values of j. The solution can be represented as a collection of the variables x_j where $\mathbf{X_P} = \{0, 0, 1, 0, 0, 1, 0, 0, 0, 1\}$ in the sequence from $j = 1$ to 10. With only p of these x_j variables having a value of 1, a more compact representation of the solution, $\mathbf{X_P}$, can be used by considering only the location index numbers, j, corresponding to locations where there will be facilities (i.e., j whose $x_j = 1$). Using the compact representation, the collection of facility location indices in the previous example can be written as $\mathbf{X_P} = \{3, 6, 10\}$. Further, the use of this median set, $\mathbf{X_P}$, in compact form provides a faster recognition of which facilities to choose from in order to satisfy the demand of a certain node. In this study, the solution is represented using the compact form, $\mathbf{X_P}$.

4.2. Computational Complexity and solution space of CPkMP

Similar to a non-capacitated p-median problem, the total number of solutions (solution space S) of the capacitated p-median problem is equal to the number of ways of choosing p facilities out of the n possible locations given as,

$$S = C_n^p = \frac{n!}{p!(n-p)!} = \begin{pmatrix} n \\ p \end{pmatrix} \tag{8}$$

where n is the number of nodes of the network and p is the number of facilities to be located (Daskin, 1995). For a constant value of p, the number in Eq. (8) is $O\left(n^P\right)$ which is polynomial in n. It should be noted, however, that certain combinations of p facilities render the solution to be infeasible due to the constraints in the total capacity of the median combinations.

In a CPkMP, the search for the $p - k$ additional locations reduces the size of the solution space into,

$$S = C_{n-k}^{p-k} = \frac{(n-k)!}{(p-k)!\,(n-p)!} = \begin{pmatrix} n-k \\ p-k \end{pmatrix} \tag{9}$$

Similar to Eq. (8), the total number of solutions of the $p - k$ median problem is polynomial in n for constant values of p and k. For a constant network size n, the solution space becomes larger as k becomes smaller.

5. GENETIC ALGORITHM

5.1. Individuals and Fitness

In this study, genetic algorithm (GA) which is based on the natural processes of selection and evolution is used to solve both CPMP and CPkMP. GA works well in any search space by randomly generating a subset of solutions (called *sample population*) from a very large and complex set of candidate solutions, making it suitable to solve combinatorial problems such as CPMP and CPkMP.

Changes in the members (called *individuals*) of the sample *population* are made based on the survival (in terms of *fitness* as a criterion) of an individual that performs better than other individuals. Also, an individual with a high fitness value has a greater chance to be selected as a parent to reproduce a child or a new solution. As a result, better individuals reproduce children that will improve the future generation of solution sets eventually leading to a more suitable (thus, optimal) solution.

For the GA of this study, an individual represents one solution and the individual's attributes (called *alleles*) represent the facility locations or medians. The solution is given as a compact form of the median set of p facility location indices, $\mathbf{X_p}$. The set of alleles of an individual is therefore the median set of location indices,

$$\mathbf{X_p} = \{X_{K,1}, X_{K,2}, X_{K,3}, ..., X_{K,k}, X_{P-K,k+1}, X_{P-K,k+2}..., X_{P-K,p}\}$$

Each allele from $\mathbf{X_p}$ is represented as $X_{status,m}$, where *status* has a code of K when the location has an existing facility and $P-K$ when location is a candidate for a new facility. The value of the allele, $X_{status,m}$, is the location index value, j, defined from the IP model whose range is from 1 to n. The first k members of the set, $\{X_{K,1}, X_{K,2}, X_{K,3}, ..., X_{K,k}\}$, correspond to the location index values of the existing k facilities. The remaining $p-k$ members are the location indices of the additional locations for new $p-k$ facilities. The m code in $X_{status,m}$ is a count from 1 to p of each member of the set $\mathbf{X_p}$. For example, an individual for a 10-median problem with 50 nodes and 3 existing facilities at locations $j = 12, 27$, and 40 may be represented as $\mathbf{X_P} = \{12, 27, 40, X_{P-K,k+1}, X_{P-K,k+2}..., X_{P-K,p}\}$, where $X_{P-K,k+1}, X_{P-K,k+2}..., X_{P-K,p}$ are the seven possible candidate location indices to be chosen from the 50-node network excluding indices 12, 27, and 40. The fitness of the individual, $\mathbf{X_P}$, is evaluated in terms of the cost function, z, according to Eq. (1).

The general procedure of genetic algorithm can be summarized into the following steps: (1) generation of initial population, (2) random selection of parents, (3) generation of a child, (4) mutation of randomly selected individual, (5) updating and improvement of the current population, (5) termination test and convergence to a solution.

5.2. Generation of the Initial Members of the Population

The initial members of the population are generated such that all nodes (i.e., candidate facility locations) in the network are uniformly represented. This diverse representation of all nodes as alleles in the initial population helps prevent premature convergence to a local optimal solution. The number of members, D, comprising a population is dependent on the size of the solution space, S. Following Alp et al. (2003) and Resurreccion and Resurreccion (2004), the size D of the sample population can be determined for both CPMP and CPkMP as,

$$D\left(n-k, p-k\right) = max\left\{2, \frac{n-k}{100}\frac{ln\left(S\right)}{d}\right\}d \tag{10}$$

where $d = (n-k)/(p-k)$ and S is the solution space given in Eq. (9) with $k = 0$ for CPMP.

All members of the population automatically have the k facility indices as alleles. All other $n-k$ potential median indices are then uniformly distributed across the initial D members of the population in order to preserve the genetic diversity of the initial population. The paper of Resurreccion and Resurreccion (2004) is referred to for the detailed explanation of the generation of initial members of a sample population.

5.3. Parent Selection

To generate a child, two different individuals from the population are selected as parents using the roulette wheel selection method (Dvorette, 1999). The individuals in the population are ranked in ascending order according to fitness value z. The probability of the i^{th} individual, $\mathbf{X_{pi}}$, in a population of size D to be selected as a parent is directly proportional to its fitness value z_i and is given as,

$$P\left(X_{pi}\right) = \frac{z_i^{-1}}{\sum_{i=1}^{D} z_i^{-1}} \quad for\ i = 1\ to\ D \tag{11}$$

Two random numbers between 0 and 1 are generated and compared with the cumulative probability distribution $P\left(X_{pi}\right)$ (i.e., calculated probability values between 0 and 1) in order to select two individuals from the population as parents. The random numbers are generated following Park and Miller (1988). This biased random selection of giving higher probability to a better individual (with a high fitness value) is a feature of the roulette wheel selection. The probability distribution changes as cost functions change when new members are introduced into the population.

5.4. Opportunity Cost-Based Child Generation Procedure

The child generated from two selected parents with facility median sets $\mathbf{X_{p1}}$ and $\mathbf{X_{p2}}$, will inherit parts of the attributes or alleles (location indices) of each parent. These parent median sets can be represented as:

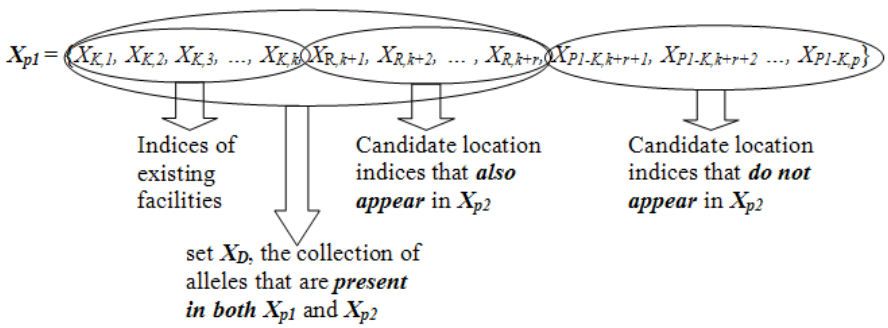
$$\begin{aligned}
\mathbf{X_{p1}} &= \{X_{K,1}, X_{K,2}, X_{K,3}, ..., X_{K,k}, X_{R,k+1}, X_{R,k+2}, ..., \\
&\quad X_{R,k+r}, X_{P1-K,k+r+1}, X_{P1-K,k+r+2} ..., X_{P1-K,p}\} \\
\mathbf{X_{p2}} &= \{X_{K,1}, X_{K,2}, X_{K,3}, ..., X_{K,k}, X_{R,k+1}, X_{R,k+2}, ..., \\
&\quad X_{R,k+r}, X_{P2-K,k+r+1}, X_{P2-K,k+r+2} ..., X_{P2-K,p}\}
\end{aligned}$$

where

$X_{status,m}$ = the location index value, j, where a facility is to be assigned ranging from 1 to n
m = the count for the median or location that is included in the set $\mathbf{X}$ ranging from the 1 to p
k = total number of existing facilities
r = total number of candidate locations for the new facilities that appear in both $\mathbf{X_{p1}}$ and $\mathbf{X_{p1}}$

Status =
- K status code for existing facility
- R status code for a candidate location for a new facility that appear in both $\mathbf{X_{p1}}$ and $\mathbf{X_{p2}}$
- $P1 - K$ code for a candidate facility location in $\mathbf{X_{p1}}$ that is not present in $\mathbf{X_{p2}}$
- $P2 - K$ code for a candidate facility location in $\mathbf{X_{p2}}$ that is not present in $\mathbf{X_{p1}}$

A child generated at this stage of the GA is assumed to only inherit alleles from the allele sets of the parents. The alleles $\{X_{K,1}, X_{K,2}, X_{K,3}, ..., X_{K,k}\}$ representing the k existing facilities that are present in both parents are automatically inherited by the child. Any number of alleles, r, that do not represent existing facilities but belong in the median sets of both parents are treated as dominant alleles and are also automatically inherited by the child. Thus, the child inherits these $k + r$ alleles that appear in both $\mathbf{X_{p1}}$ and $\mathbf{X_{p2}}$. The set of alleles corresponding to the existing locations and candidate locations that can be found in both parents, $\mathbf{X_{p1}}$ and $\mathbf{X_{p2}}$ form the set $\mathbf{X_D}$, as illustrated in the following diagram.



The child inherits all the $(k + r)$ location indices from the set $\mathbf{X_D}$. To complete the allele requirements of the child, the remaining $p-(k + r)$ indices of the child are needed to be chosen from the alleles that appear distinctively from each of the parents. There are $p - k - r$ alleles of the first parent that are distinct from the second parent and, in the same way, the second parent also has a number of $p-k-r$ alleles that are different from the first parent. Altogether, the combined number of alleles that appear distinctively from the two parents is $2\,(p - k - r)$. This collection of the distinct alleles from the two parents forms the set $\mathbf{X_U}$. Hence, the child, $\mathbf{X_C}$, inherits all the members of the set $\mathbf{X_D}$ and searches from the set $\mathbf{X_U}$ a number of $p-k-r$ indices to complete the p median requirement. In summary, the following notations are applied for this search:

$\mathbf{X_D}$ = the set of alleles that are present in both parents $\mathbf{X_{p1}}$ and $\mathbf{X_{p2}}$ or the intersection $(\mathbf{X_{p1}} \bigcap \mathbf{X_{p2}})$ with $(k + r)$ elements

$\mathbf{X_U}$ = the set that combines the alleles that are distinct from each parent, $\mathbf{X_{p1}}$ and $\mathbf{X_{p2}}$ (alleles that are uncommon between $\mathbf{X_{p1}}$ and $\mathbf{X_{p2}}$)
= $(\mathbf{X_{p1}} \bigcup \mathbf{X_{p2}}) - \mathbf{X_D}$ with $2(p - (k + r))$ elements

$\mathbf{X_C}$ = the p-median set of facility locations of the child with p elements

A child generation procedure called *opportunity cost-based genetic algorithm* (OC-GA) was developed by Resurreccion and Resurreccion (2004) for a PkMP that reduces the set $\mathbf{X_U}$ to the required number of $p - k - r$ alleles using an opportunity cost criterion. The set $\mathbf{X_U}$ with $2(p - k - r)$ member indices is reduced by removing one index at a time until only half of the members remain in the set $\mathbf{X_U}$. The remaining $p - k - r$ indices in $\mathbf{X_U}$ combined with the indices from $\mathbf{X_D}$ form the set of p-medians of the child, $\mathbf{X_C}$. Hence, the child generation procedure determines which members in $\mathbf{X_U}$ must be chosen as part of the medians of the set $\mathbf{X_C}$ that minimizes total cost.

The OC-GA procedure works by assuming that all of the current members in $\mathbf{X_U}$ except for one member, X_N, are chosen as medians. For the demand of a node to be assigned to X_N, X_N must be the closest facility to the node among the facilities in the union set $\mathbf{X_U} \bigcup \mathbf{X_D}$ provided that the unassigned capacity of X_N is greater than or equal to the demand of the node. In effect, the removal of X_N requires the demand of the node to be assigned to another facility in the set $\mathbf{X_U} \bigcup \mathbf{X_D}$, which is farther from the demand node than X_N. Therefore, the change in the cost function (Eq. 1) as a result of the removal of X_N is an additional distance travel per unit of satisfied demand.

The concept of opportunity cost can be used to quantify the effect of not placing a facility in X_N on the cost function z (Eq. 1). *Opportunity cost* (OC) is defined as the cost of a foregone opportunity (DeGarmo et al., 1993). The total additional cost of not having a facility in X_N is the cost of reassigning the demand of all nodes that were supposedly assigned to X_N to a farther facility from the node than X_N. Retaining X_N as a median prevents incurring such additional cost which can also be interpreted as an opportunity for further lowering the cost function (Eq. 1). This additional cost is the opportunity cost of not making X_N a median. Thus, each member of the set $\mathbf{X_U}$ is assumed as X_N and is considered for removal from $\mathbf{X_U}$ based on its opportunity cost.

To help in evaluating which nodes in $\mathbf{X_U}$ must be eliminated from $\mathbf{X_U}$, the following notations used in the OC-GA are introduced:

X_N	= candidate facility location considered for removal from $\mathbf{X_U}$
$\mathbf{X_{U-N}}$	= set of alleles from $\mathbf{X_U}$ excluding X_N
$\mathbf{X_{UD}}$	= $\mathbf{X_D} \bigcup \mathbf{X_U}$ (set of facility locations including X_N)
$\mathbf{X}$	= $\mathbf{X_D} \bigcup \mathbf{X_{U-N}}$ (set of facility locations excluding X_N)
CUD	= set of the capacities of the facilities in the set $\mathbf{X_{UD}}$
C	= set of the capacities of the facilities in the set X
$X_{i,UD}$	= facility location index from $\mathbf{X_{UD}}$ that is closest to demand node i
X_i	= facility location index from $\mathbf{X}$ that is closest to demand node i
$d_{i,UD}$	= distance between node i and $X_{i,UD}$
d_i	= distance between node i and X_i
h_i	= demand of node i
$OC(i, X_N)$	= opportunity cost incurred when the demand of node i cannot be satisfied by X_N
OC_{X_N}	= sum of opportunity costs of not having a facility in location X_N

Eq. (1) associates cost minimization with the shortest distance traveled. If X_N is the closest facility to a demand node i, (i.e., $X_N = X_{i,UD}$), X_N has the best opportunity of minimizing the cost to satisfy the demand of node i with a value of $h_i^* d_{i,UD}$. By not placing a facility in X_N, demand node i will be required to seek service from another facility which is farther than X_N. This service facility is defined to be X_i, the closest facility to node i from the set $\mathbf{X}$ assuming no facility can be placed in X_N. Thus, with a different facility serving node $i(X_i \neq X_N)$ when X_N should have been the best choice $(X_N = X_{i,UD})$, there is a higher cost of satisfying the demand of node i, which is given by the distance $h_i d_i$ where $d_i > d_{i,UD}$. The increase in the distance traveled (i.e., the additional distance $d_i - d_{i,UD}$) by having the facility in X_i serve the demand of node i instead of X_N is the opportunity cost (OC) of not placing a facility in X_N to satisfy the demand of node i. This OC is equal to the additional distance traveled per unit of demand times the demand of node i,

$$OC(i, X_N) = (d_{i,UD} - d_i)h_i \tag{12}$$

The removal of the candidate location, X_N, from $\mathbf{X_U}$ has no effect, however, on the cost function (Eq. 1) if X_N is not the closest facility to the demand node $i(X_N \neq X_{i,UD})$ since the closest facility to node i is included in both sets $\mathbf{X_{UD}}$ and $\mathbf{X}(X_i = X_{i,UD})$. No opportunity cost is incurred in satisfying the demand of i since the facility closest to i will be still be available even after the removal of X_N. For the same reason, the distances $d_{i,UD}$ and d_i will be equal $(d_{i,UD} - d_i = 0)$ and the opportunity cost, $OC(i, X_N)$, is equal to zero for the node i under consideration. Further, if the closest facility to the demand node i is in set XD, the opportunity cost $OC(i, X_N)$ is always equal to zero with respect to node i regardless of which member in $\mathbf{X_U}$ is being considered for removal.

The evaluation of the opportunity cost of a candidate location, X_N, is therefore considered only for the demand nodes having X_N as their closest facility. In the CPkMP, the OC-GA identifies a facility location to be the serving facility to a node when it is the nearest location to the node and has an unassigned capacity to satisfy the node's demand. The opportunity cost of not placing a facility in a location index X_N is equal to the sum of all opportunity costs from all nodes that should have X_N as the closest facility to serve their demands. This is given as,

$$OC_{X_N} = \sum_{X_{i,UD}=X_N} OC(i, X_N) = \sum_{X_{i,UD}=X_N} (d_i - d_{i,UD})^* \, h_i \quad \forall \text{ such that } X_{i,UD} = X_N \tag{13}$$

Among the potential facilities in the set $\mathbf{X_U}$, the opportunity cost genetic algorithm (OC-GA) eliminates the candidate location, X_N, which has the lowest sum of opportunity costs, OC_{X_N}. Choosing to remove the X_N with the lowest OC_{X_N} from the list of candidate locations in $\mathbf{X_U}$ prevents the foregoing of better opportunities in minimizing cost. The procedure is repeated with one less member in the set $\mathbf{X_U}$ updating the indices in the new $\mathbf{X_{UD}}$, $\mathbf{X}$, $X_{i,UD}$ and X_i until $p - k - r$ facilities remain in $\mathbf{X_U}$. Finally, $\mathbf{X_C}$ is the union of the set $\mathbf{X_D}$ and the $\mathbf{X_U}$ after the application of the OC-GA. A detailed computational example of an opportunity cost-based child generation procedure is given by Resurreccion and Resurreccion (2004).

5.5. Improvement of the Population

A candidate solution or child is accepted as a new member of the next generation provided that its fitness value is better than the worst fitness value of an individual in the current population. If the child is accepted, new worst and best values of fitness and parent-selection probabilities of the members in the new population are recomputed. If the generated offspring is rejected, another parent-pair is randomly obtained for the generation of a different child. The best and worst values of z are kept as basis of comparison for the acceptance of any candidate child into the new generation.

5.6. Mutation

Genetic diversity can be maintained using the process of mutation where a randomly selected allele of an individual is changed (mutated) when necessary. In this study, a 10 percent uniform mutation rate is used to prevent premature convergence to a local optimal solution.

5.7. Termination

The generation of offspring is terminated until a number of children have failed to replace the best solution of the population. Termination is done when there is no change in the current value of the best solution for 500 iterations. The OC-GA used to solve the CPkMP was coded in C using the built-in C/C++ compiler in Linux system.

6. RESULTS AND DISCUSSION

This section presents the summary of the performance of the recently developed OC-GA tested against ten 50-node and one 100-node network capacitated p-median problems (CPMP) from literature and applied on the corresponding capacitated $p - k$ median problem (CPkMP).

6.1. Test Problems and Model Performance

The OC-GA was applied on the 50-node and 100-node networks of the capacitated p-median test problems of Osman and Christofides (1994) that are available in the standard OR library (http://people.brunel.ac.uk/ mastjjb/jeb/orlib/files) of Beasley (1990). The downloaded data input files were modified to suit the input file format required by the program.

Table I showed the performance of the OC-GA tested against the ten 50-node 5-median CPMP that assumed no existing facilities in the original network (i.e., $k = 0$). The OC-GA solution to the CPMP for the 10 test problems has a 4.48% deviation from the known Lagrangean optimal solution (Table I). However, the results obtained are in very close agreement with the results of Ghoseiri and Ghannadpour (2007) in applying a classical constructive GA on the same test problems. In comparison with other computing methods, the OC-GA gave relatively better values of z than the constructive heuristic algorithm used in Osman and Christofides (1994) for the same test problems. The few studies on the application of GA and other algorithms to solve the CPMP were mostly improvement algorithms (Senne and Lorena, 2000; Ghoseiri and Ghannadpour 2007; Osman and Christofides 1994) while the OC-GA presented in this study is a constructive algorithm to solve the CPkMP which is a modified CPMP. In perspective and as a recommendation for further study, an improvement algorithm can be imposed on the constructive OC-GA such as prioritization of the assignment of the demand nodes to the facilities to further reduce the z-value toward the optimal solution.

For each of the ten test problems, the number of existing facilities k was varied and different location indices were selected from the optimal combination of medians in the CPMP to represent the k existing facilities for the CPkMP. The performance of the proposed OC-GA on the CPMP was evaluated by comparing the generated remaining set of $p - k$ medians with the optimal median combination of the corresponding original CPMP.

For a 50-node network with $p = 5$, the CPkMP was solved considering all the possible combinations of k facilities taken from the optimal CPMP solution for each value of k from 1 to 4 (Table I and Fig 1a - d). The results of the 5-median CPkMP for the ten 50-node networks are summarized in Table I. When the locations for the k existing facilities are derived from the optimal Lagrangean solution of the CPMP, the additional $p - k$ medians found by the OC-GA for the CPkMP usually tend to converge to the remaining medians of the optimal CPMP solution. This can be observed from the low relative percentage errors of 0.03% to 1.52% between the OC-GA solutions to the CPMP and the CPkMP for the 10 test problems

Average of Z best values										
k	test1	test2	test3	test4	test5	test6	test7	test8	test9	test10
0	736	758	789	684	717	814	824	858	739	860
1	753	762	806	685	724	830	840	883	746	872
2	739	759	793	684	721	815	824	870	741	885
3	737	758	790	684	718	814	824	859	740	872
4	736	758	789	684	717	814	824	858	739	860
mean	741.08	759.30	794.33	684.23	719.99	818.18	828.13	867.33	741.51	872.05
mean %error	0.69	0.17	0.67	0.03	0.42	0.51	0.50	1.09	0.34	1.40
optimal Z value	713	740	751	651	664	778	787	820	715	829
deviation from the optimal Z value	3.23	2.43	5.06	5.07	7.98	4.63	4.70	4.63	3.36	3.74

Standard deviation of Z best values										
k	test1	test2	test3	test4	test5	test6	test7	test8	test9	test10
0										
1	9.92	2.86	21.86	0.84	6.14	11.99	17.64	10.33	4.97	11.69
2	5.48	1.63	4.58	0.32	4.92	2.02	0.32	13.36	3.14	32.63
3	2.21	1.26	1.58	0.00	2.83	0.00	0.00	2.53	1.58	26.72
4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
mean	4.40	1.44	7.00	0.29	3.47	3.50	4.49	6.55	2.42	17.76

Table I. Fitness values of the calculated optimal solutions for the ten 50-node network $p(=5)$ -median problems at different k number existing facilities. Mean and standard deviations of the fitness values are computed for all possible combinations of k taken from p.

(Table I). It should be noted that the difference in the solution set sizes of the CPkMP and CPMP resulted in finding some medians for the CPkMP solution that do not belong to the optimal solution of the CPMP. The probability of finding a median different from the optimal set of CPMP is proportional to the size of the solution space where the GA performs its search, as illustrated by the greater dispersion of z values for lower values of k (Fig 1 a - d) in agreement with Eq. (8) and Eq. (9). The dispersion (width of the range of z values) reduces as k increases. Also, as k increases, the average z values decreases markedly approaching the optimal z value of the corresponding CPMP. Thus, it is more likely that the remaining $p-k$ candidate facilities would converge to the optimal medians with more existing facilities forced to take on the locations of thee known optimal medians.

However, for a 100-node network with ten required p medians, not all combinations of k medians representing the fixed facilities were evaluated for higher values of k and p as the number of combinations increase as p and k increase, according to

$$S = C_p^k = \frac{p!}{k!\,(p-k)!} = \begin{pmatrix} p \\ k \end{pmatrix}$$

Thus, for the 100-node network with ten required p-medians, only 20% of the possible combinations that were randomly generated were tested in this study for each considered value of k ($k = 1$ to $k = 9$) due to the large number of possible combinations. This can be accepted only to reveal the general trend in z values by varying k.

The effect of increasing the number k of existing facilities in the solution to the CPkMP is shown in Fig 2 for a 10-median 100-node problem. It is apparent that the range of z values becomes narrower as k increases at a given value of p, in agreement with Fig 1 a-d. Although the z value approaches the optimal solution, the rate of convergence to the optimal solution is slower in the 100-node network than the 50-node network because of the significant increase in size of the solution space for the 100-node network, following Eq. (9) (compare Fig 1 and Fig 2).

6.2. *Sensitivity Analysis*

Fig 3 illustrates the effect of increasing the required number of facilities, p, and increasing the existing facilities k at a specified value of p on the z value. As p increases, the z apparently decreases since more medians are able to serve the overall demand of the network. Having more accessible available medians in minimizing z counterbalances the effect of the increase in the size of the solution space as p is increased. Consistent with Fig 1 and 2, increasing the number of k existing facilities has a marked effect on the z value especially at low values of p (Fig 3). The range of z values at different k values for $p = 5$ is magnified to show that z values are higher at low values of k and closer to the optimal solution at high values of k. Also, the z values for all values of k become significantly narrower as p increases. This is because of the expected increase in the size of the solution space as k decreases, as described by Eq. (9) since a lower value of k provides a larger number ways for choosing $p - k$ median combinations.

The OC-GA was also applied to one of the test problems (test problem 2) using k existing facilities that are not at all part of the known optimal solution of the CPMP. Fig 4 shows the result if the existing facility locations were selected from the median set 10, 20, 30, 40, 50 for different values of k for a CPkMP with $p = 5$. Forcing a median that is different from the optimal median combination modifies the final solution that was being searched for. When specified medians are included as part of the solution, the remaining $p - k$ medians found by the OC-GA were the same medians from the optimal solution of CPMP within the range of 50% to 70% of the time. When the number of existing facilities whose locations are not taken from the CPMP is increased, the minimum cost value, z, obtained by the OC-GA moves farther from the optimal z value and closer to $\mathbf{X_S} = 10, 20, 30, 40, 50$. This means that CPkMP is a totally different problem from the CPMP but the optimal solution for the CPkMP tries to become similar to the optimal solution of the CPMP to guarantee the minimization of the cost function z.

7. ABC FOOD SERVICES: AN OC-GA APPLICATION TO A DISTRIBUTION NETWORK

ABC Food Services Incorporated is the sole distributor of food products and beverages of a food chain in the Philippines. The company provides delivery of food products to stores of the food chain where 9 trucks are devoted to deliveries within Metro Manila. The first distribution center has a delivery capacity of 702,000 cubic feet per week. In 2000, the company has decided to open its second distribution center which has a smaller capacity of 360,000 cubic feet per week in order to reduce the frequency of special trips that are supposedly scheduled for maintenance but are dispatched to satisfy the demand, to reduce hiring from external sources (i.e., outsourcing) which was proven to be more costly, and to prevent having unsatisfied demand for the day that are carried over to demands for other days.

From serving 98 stores in 2000, the number of demand points has now increased to 141 which is composed of restaurants, company owned stores, and franchises. A list of the 141 stores in Metro Manila including their respective coordinate locations, weekly demands and capacities is provided in table V in Appendix A. The existing distribution centers are located at nodes 84 and 137 with node 137 having the larger capacity of 702,000 cubic feet per week. The relative locations of the nodes are shown in Figure 5 in Appendix A.

The evaluation of the current assignment of the demands of the nodes to the two distribution centers is summarized as scenario 0 in Table II. An average weekly demand of 142,467.01 cubic feet is outsourced by the company at a cost of PhP 2500 per 200 cubic feet or PhP 1,780,837.63 per week. This cost is in addition to the cost of labor and transportation for traveling a distance of 1,632 kilometers per week for the deliveries made by the company.

SCENARIO 0:

Current system's main facility at node 137 with 117,000 cu. ft. capacity per day or 702,000 cu. ft. capacity per week

A second smaller facility is at node 84 with 360,000 cu. ft. capacity per week.

Current system performance		*delivery in cu. ft. per week*	*distance travelled (in km) per week*
Facility at node 137	=>	702,000.00	982.16
Facility at node 84	=>	333,558.74	450.55
Commissioned/Outsourced		142,467.01	199.32
	TOTAL =	1,178,025.75	**1,632.04**

Table II. Evaluation of current assignment of demands to the two distribution centers

The OC-GA is applied on five different scenarios as summarized in Table III. Scenario 1 is the noncapacitated PMP problem for different required p facilities while scenario 2 is the CPMP counterpart of scenario 1 where the capacity of each node is assigned to be 702,000 cubic feet per week which is the capacity of one of the existing distribution centers. Both scenario 1 and scenario 2 assume that there are no existing facilities. By comparing a noncapacitated PMP scenario 1 to a CPMP scenario 2, the capacity of the candidate nodes obtained the same OC-GA solution except when the required number of facilities is four (i.e., for $p = 3$). The difference in the distance traveled, z, as given by the OC-GA solution for between scenarios 1 and 2 when $p = 3$ is 1.49 km per week. Also, by comparing scenario 1 (or 2) with scenario 0 for $p = 2$, the reduction in distance traveled is at least 700 km per week than scenario 0 (i.e., current state). Further, greater savings amounting to at least 40 million a year is realized by no longer outsourcing every other week.

Scenario 3 is a CPMP similar to scenario 2 but with a lower capacity of 360,000 cubic feet per week assigned to the each facility node. Requiring only two or three facilities (i.e., $p = 2$ or 3) resulted in no solution as the capacities of the nodes are not sufficient to satisfy the total demand of the network. Thus, there is a need for external sources (outsourcing) or the company has to decide on building new facilities as distribution centers. When the required facility exceeds 4 (i.e., $p \geq 4$), the corresponding reduction in distance traveled when compared to scenario 0 is at least 963 km per week. This may be costly as three or more facilities are required to be built in addition to the existing facilities saving only 200 km more than having two facilities in scenario 2. However, the company may consider the savings from no longer hiring an external source as a means of putting up these new facilities. Thus, a balance in the number of facilities, increasing the capacity of the distribution centers, and operating expenses must be taken into consideration.

SCENARIO 1:

infinite capacity per week per candidate node
k=0 (no existing facilities)

P (required facilities)	optimal facility indices	minimum weekly distance (km)
P = 2	42, 47	**926.74**
P = 3	1, 68, 101	**757.31**
P = 4	18, 42, 68, 85	**596.29**
P = 5	17, 42, 44, 68, 85	**499.79**

SCENARIO 2:

702,000 cu. ft. capacity per week per candidate node
k=0 (no existing facility)

P (required facilities)	optimal facility indices	minimum weekly distance (km)
P = 2	42, 47	**926.74**
P = 3	1, 68, 101	**757.31**
P = 4	1, 42, 68, 85	**597.78**
P = 5	17, 42, 44, 68, 85	**499.79**

SCENARIO 3:

360,000 cu. ft. capacity per week per candidate node
k=0 (no existing facility)

P (required facilities)	optimal facility indices	minimum weekly distance (km)
P = 2	-	-
P = 3	-	-
P = 4	78, 84, 90, 110	**668.96**
P = 5	17, 42, 44, 68, 85	**499.79**

Table III. OC-GA SOLUTIONS for PMP and CPMP with no existing facilities ($k = 0$) applied to ABC Food Services, Inc. (Scenario 1 is a non-capacitated PMP, scenarios 2 and 3 are CPMP.)

The company that has two existing distribution centers at node 84 and 137 can further evaluate whether the replacement of the smaller facility at node 84 while retaining the larger facility at node 137 with capacity at 702,000 cubic feet per week is beneficial to the company. Scenarios 4 and 5, as shown in Table IV, uses the OC-GA with k equal to 1 corresponding to the remaining existing facility in node 137 (capacity of 702,000 cubic feet per week) while searching for new locations for the additional $p - k$ facilities. Scenario 4 is a CPkMP with capacity of capacity of 702,000 cubic feet per week assigned to each candidate node. Scenario 5 is similar to scenario 4 but with 360,000 cubic feet per week capacity assigned to each candidate node.

SCENARIO 4:

k=1 (Existing facility at node 137 with 702,000 cu. ft. capacity)

=> 702,000 cu. ft. capacity per week per candidate node

P (required facilities)	optimal facility indices	minimum weekly distance (km)
P = 2	79, 137	**1097.45**
P = 3	42, 47, 137	**858.47**
P = 4	5, 54, 85, 137	**692.81**

SCENARIO 5:

k=1 (Existing facility at node 137 with 702,000 cu. ft. capacity)

=> 360,000 cu. ft. capacity per week per candidate node

P (required facilities)	optimal facility indices	minimum weekly distance (km)
P = 3	62, 123, 137	**1001.97**
P = 4	8, 42, 68, 137	**719.53**
P = 5	24, 68, 85, 103, 137	**556.05**

Table IV. OC-GA SOLUTIONS for CPkMP with one existing facilities ($k = 1$) at node 137 applied to ABC Food Services, Inc.

Comparison of scenario 4 with scenario 0 shows a decrease of 534.59 km distance traveled (i.e., 1632.05 - 1097.45) by adding one more facility with a 720,000 cubic feet capacity to the existing facility at node 137. Between scenario 5 and scenario 0, a 630.07-km of distance traveled (i.e., 1632.05 - 1001.97) is saved per week when two new facilities at location nodes 62 and 123 that each has a 360,000-cubic feet weekly capacity are added to the network replacing the facility at node 84. The distance to be traveled decreases for every new facility added to the network but greater savings in distance traveled is realized by adding a new facility with a capacity of 360,000 than adding a new facility with a capacity of 702,000 cubic feet. The cost of building these additional facilities may be compared to the costs of hiring from outside sources and the additional income that will be generated by future increases in the demand of the network.

8. CONCLUSIONS

In this study, the recently-developed opportunity cost-based genetic algorithm (OC-GA) worked well when tested against ten 50-node and one 100-node network test problems from literature. The consideration of a k number of existing facilities chosen from the optimal medians at $k = 0$ that will be part of the p medians resulted into the searched $p - k$ facility locations that also form the optimal set of medians (at $k = 0$). When the k fixed facility locations were chosen from the optimal median solution of the CPMP, the OC-GA solution to the CPkMP had lower deviation from the optimal median combination for higher values of k. When the location indices assigned as fixed facilities are not part of the optimal solution of the CPMP, the OC-GA was still able to include medians from the optimal set with an increasing deviation as more fixed facilities become different from the optimal CPMP solution. In this study, the OC-GA proved to be a useful tool in providing facilities to satisfy the total demand of a food distribution company at a minimum costs. Thus, it is recommended that the proposed OC-GA be used to solve capacitated p-median problems with or without fixed existing facilities.

REFERENCES

1. Alp, E., Erkut, E. and Drezner, Z., 2003. "'An Efficient Genetic Algorithm for the p-Median Problem,"' Annals of Operations Research, 122, pp. 21-42.
2. Beasley, J. E., 1985. "'A Note on Solving Large p-median Problems,"' European Journal of Operations Research, 21, pp. 270-273.
3. Beasley, J.E., 1990, "'OR-Library: distributing test problems by electronic mail"', Journal of the Operational Research Society. 41(11), pp. 1069-1072.
4. Berman, O. Drezner, Z. and Wesolowsky, G. O., 2003, "'Locating Service Facilities whose Reliability is Distance Dependent."' Computers and Operations Research, 20, 1683-1695.
5. Bozkaya, B., Zhang, J., Erkut, E., 1997. "'An effective genetic algorithm for the p-median problem."' Research Report No. 97-2, Research Papers in Management Science, University of Alberta, Canada.
6. Correa E.S., Steiner M.T.A., Freitas A.A. and Carnieri C. 2004, "'A Genetic Algorithm for Solving a Capacitated p-Median Problem: Theory and Practice in Optimization"' *In* Numerical Algorithms. Martnez, J. M.and Yuan, J. Y. (eds.). 35, pp. 373-388.
7. Daskin, M. S., 1995, *Network and Discrete Location: Models, Algorithms and Applications*. New York: John Wiley and Sons, Inc.
8. Dvorett, J., 1999, *Compatibility-Based Genetic Algorithm: A New Approach to the p-Median Problem*, Northwestern University, IL.
9. DeGarmo, E. P., Sullivan, W. and Bontadelli, J., 1993, *Engineering Economy*, 9th ed. Macmillan Publishing Company, New York.
10. Galvao, R. D. and ReVelle, C., 1996, "'A Lagrangean Heuristic for the Maximal Covering Location Problems,"' European Journal of Operations Research, 88, pp. 114-123.
11. Garey, M. R. and Johnson, D. S., 1979, *Computers and Interactibility: a Guide to the Theory of NP-Completeness*. Freeman, San Francisco, California.
12. Ghoseiri, K. and Ghannadpour, S.F. 2007. "'Solving capacitated p-median problem using genetic algorithm."' IEEE International Conference on Industrial Engineering and Engineering Management. Dec. 2-4, 2007. pp. 885-889.
13. Goldberg, D. E., 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley, Reading, MA.
14. Hakimi, S. L., 1964, "'Optimum Location of Switching Centers and the Absolute Centers and Medians of a Graph,"' Operations Research, 12, pp. 462-475.
15. Hakimi, S. L., 1965, "'Optimum Location of Switching Centers in a Communication Network and Some Rated Graph Theoretic Problems,"' Operations Research, 13, pp. 450-459.
16. Holland, J. H., 1975, Adaptation in Natural and Artificial Systems, University of Michigan Press, Ann Arbor, MI.
17. Lorena, L. A. N. and Furtado, J. C., 2001, "'Constructive Genetic Algorithm for Clustering Problems,"' Evolutionary Computation 9(3), Massachusetts Institute of Technology.
18. Kariv, O. and Hakimi, S. L., 1979, "'An Algorithmic Approach to Network Location Problems. II: The p-Medians,"' SIAM Journal on Applied Mathematics, 37, pp. 539-560.
19. Maranzana, F. E., 1964, "'On the Location of Supply Points to Minimize Transport Costs,"' Operational Research Quarterly, 15, pp. 261-270.
20. Mirchandani, P. B. and R. L. Francis, 1990. *Discrete Location Theory*, Wiley, New York.
21. Osman, I. H., and N. Christofides. 1994, "'Capacitated clustering problems by hybrid simulated annealing and tabu search"', International Transactions in Operations Research. 1(3), pp. 317-336.

22. Park, S. and Miller, K., 1988, "'Random Number Generators: Good Ones are Hard to Find,"' Communications of the ACM, vol.30, no.10, pp. 1192-1201.
23. Revelle, C. and Swain, R., 1970, Central Facilities Location, *Geographical Analysis*, 2, pp. 30-42.
24. Resurreccion, J. and A. Resurreccion. 2004, "'An opportunity cost-based modified genetic algorithm to solve p - k median problems."' Philippine Engineering Journal.
25. Senne, L. F. and Lorena L. A. N., 2000. Lagrangian/Surrogate Heuristics for p-Median Problems. In Laguna, M. and Gonsalez-Velarde, J. L. (eds.), *Computing Tools for Modeling, Optimization and Simulation: Interfaces in Computer Science and Operations Research*, 115-130, Kluwer Academic Publishers, London, UK.

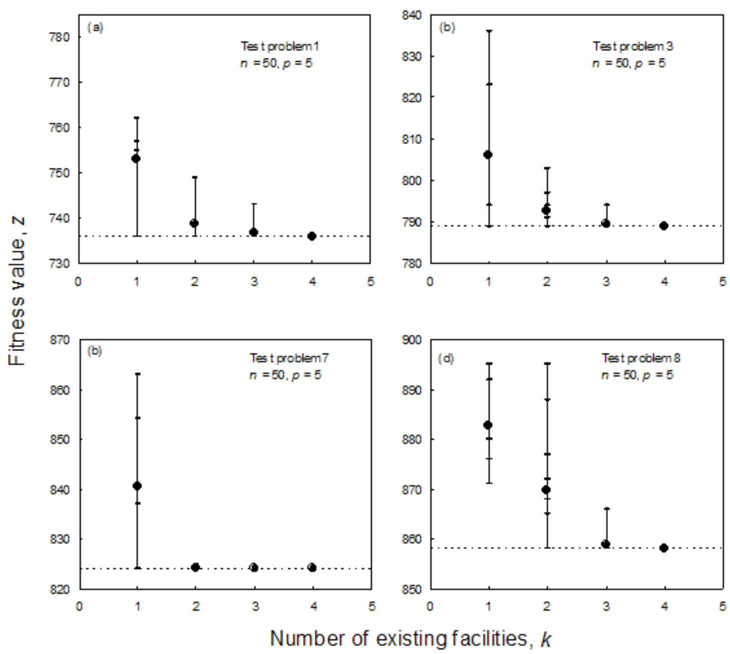


Figure 1. (a) - (d) Plot of the fitness value z against the number of existing facilities k for selected capacitated p-median test problems ($n = 50$ and $p = 5$). The k facilities were chosen from the optimal set of p medians while the remaining $p - k$ medians were searched from the $n - k$ candidate locations. Each filled out circle represents the average of the z-values at the specified number of k facilities. The dashed horizontal line is the z value of the optimal p median combinations solved by the opportunity cost-based genetic algorithm.

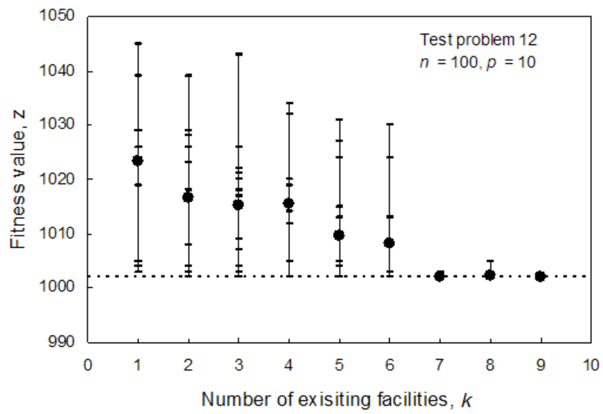


Figure 2. Plot of the z fitness value against the number of existing facilities k. The combination of location indices for the k facilities is taken from the optimal set of p medians of a selected capacitated 10-median 100-node test problem. Note that only 20 percent of the possible combinations for $k = 3$, 4, and 5 was shown. The dashed horizontal line is the z value of the optimal p median solution of the opportunity cost-based genetic algorithm.

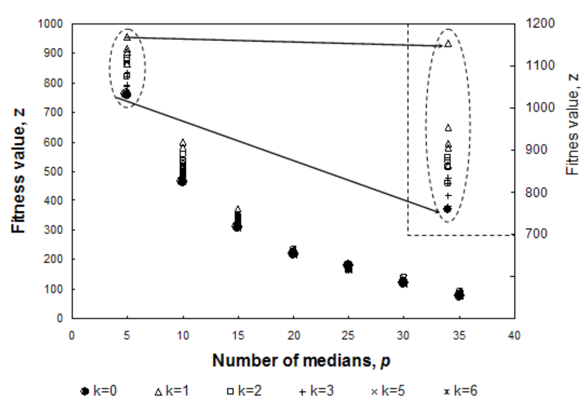


Figure 3. Plot of the fitness value z against the number of median facilities p for a selected capacitated p-median test problem (test problem 2; $n = 50$). For each case of p (= 10, 15, 20, 25, 30), certain combinations of existing facilities for $k = 1$ to 6 and for the case when $p = 5$, combinations of existing facilities for $k = 1$ to 3 were simulated with results as shown. The inset figure is for the z values at different values of k facilities for $p = 5$.

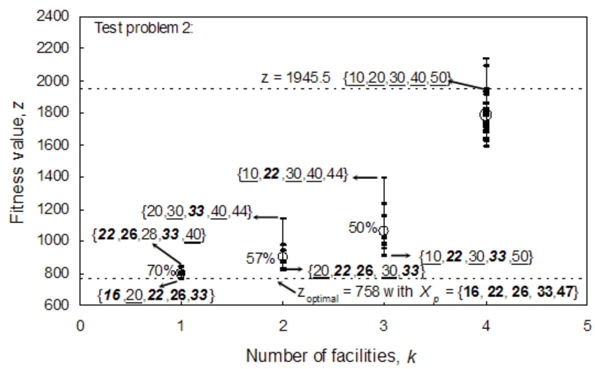


Figure 4. Plot of the fitness value z against the number of existing facilities k when the locations for the k facilities are obtained from the set of specified medians $\mathbf{X_S} = 10, 20, 30, 40, 50$ for test problem 2 ($n = 50$ and $p = 5$). The lower dashed horizontal line represents the z value for the optimal p-median solution of the OC-GA for the CPMP (i.e., $k = 0$) with $p = 5$ that is $\mathbf{X_P} = 16, 22, 26, 33, 47$. The upper dashed horizontal line is the z value obtained by assigning all the locations in $\mathbf{X_S}$ as fixed medians. The unfilled circles are the corresponding average values of z when setting the locations for the k fixed facilities using medians from $\mathbf{X_S}$. The percentage shown is the relative amount of medians in the OC-GA solution that converge or appear similar to the medians in $\mathbf{X_P}$. Bold indices are part of the optimal set of p-medians and underlined indices are part of the fixed medians.

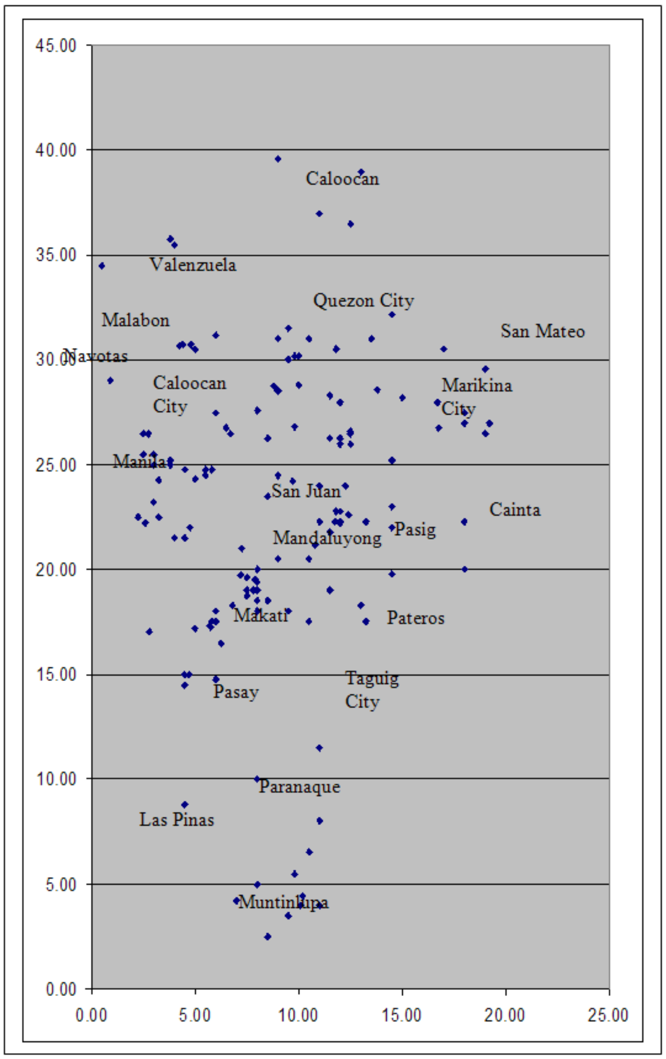


Figure 5. Plot of the Stores in Metro Manila

APPENDIX A

Store No.	weekly demand (cu. ft.)	position X (km)	coordinates Y (km)
1	22342.25	12	26
2	26408.97	11	24
3	7483.11	8.5	18.5
4	19517.51	3.25	22.5
5	10256.56	8	27.6
6	13559.48	8	19
7	11378.28	10	30.2
8	13422.21	10.2	4.4
9	10883.22	7.8	19
10	5003.49	7.5	18.75
11	6661.27	11.75	22.3
12	4133.62	4.5	8.8
13	12874.81	13.8	28.6
14	1792.45	12	22.75
15	10193.41	5.5	24.75
16	7355.73	16.75	26.75
17	21508.44	12	22.2
18	13492.24	12	26.25
19	4889.63	12.25	24
20	7632.83	3	25
21	12051.58	11.8	30.5
22	8380.95	11	4
23	7854.78	18	22.3
24	3801.21	11.5	26.25
25	5745.36	12.5	26.6
26	2729.3	19	29.6
27	3511.62	6.8	18.3
28	4425.2	17	30.5
29	6676.22	10.5	31
30	7410.89	3	23.2
31	5458.94	13	18.3
32	9092.89	12.5	36.5
33	5715.23	0.5	34.5
34	9670.2	2.5	25.5
35	4243.25	2.25	22.5
36	6605.16	7	4.2
37	5647.76	13.25	17.5
38	9400.3	6	14.75
39	8423.29	9	28.5
40	8574.41	4.75	22
41	15399.73	6.5	26.75
42	12543.82	6	17.5
43	17648.63	15	28.2
44	11353.14	11.5	28.3

Table V. Weekly Demand and Position Coordinates of the Nodes

Store No.	**weekly demand (cu. ft.)**	**position X (km)**	**coordinates Y (km)**
45	5445.12	9	31
46	6391.06	9.5	3.5
47	6773.35	9.8	26.8
48	6377.03	9.5	3.5
49	6492.74	7.9	19.5
50	4383.04	4.5	14.5
51	5393.61	6	31.2
52	13921.3	4.5	21.5
53	13443.52	10	28.8
54	6158.57	7.5	18.75
55	6282.41	8.8	28.75
56	9270.22	5.8	24.75
57	9513.61	4.5	24.75
58	6344.15	4.7	15
59	7534.87	8	20
60	6809.35	2.75	26.5
61	4425.99	7.25	21
62	5946.14	6.7	26.5
63	7125.6	3.25	24.25
64	8154.75	9.8	30.2
65	7560.46	18	20
66	6393.59	8	18
67	2548.92	4.4	30.75
68	7483.99	3.8	25
69	7982.11	8	10
70	5052.24	9.7	24.2
71	6205.62	11	37
72	5464.02	6	18
73	6777.8	19.2	27
74	6402.31	13	39
75	9438.52	11	11.5
76	6192.1	10.8	21.2
77	7941.3	8	18.5
78	6589.81	11.8	22.8
79	3952.36	7.2	19.75
80	3847.83	8	19.4
81	6948.49	3.25	24.25
82	7154.95	12	22.3
83	4234.51	14.5	25.2
84	*5770.09*	*10.5*	*6.5*
85	7845.07	9.8	5.5
86	6725.6	12.4	22.6
87	6397.41	11.5	21.8
88	5538.87	4	21.5
89	8080.81	11.5	19
90	6896.95	9.5	30
91	8341.03	9.5	30
92	5515.49	18	27.5

Store No.	weekly demand (cu. ft.)	position X (km)	coordinates Y (km)
93	6714.35	11	22.3
94	4816.16	12.5	26
95	5510.78	0.9	29
96	4773.04	13.25	22.3
97	6874.91	12.5	26.5
98	5303.79	14.5	23
99	5805.94	9	39.6
100	3979.3	5.5	24.5
101	10316.39	6.25	16.5
102	7736.79	12.5	26.5
103	20054.05	5.8	17.5
104	6800.11	13.5	31
105	4833.74	8.5	26.25
106	10758.35	3.8	25.2
107	6464.17	14.5	32.2
108	5932.84	8.5	23.5
109	6002.24	4	35.5
110	7797.09	5.75	17.25
111	3061.08	9	24.5
112	20983.3	5	17.2
113	9387.41	5	24.3
114	6213.04	7.5	19
115	4558.42	9.5	18
116	6447.08	4.8	30.75
117	8690.07	14.5	19.8
118	30385.68	4.5	15
119	8832.66	2.5	26.5
120	9353.09	3	25.5
121	5346.43	7.5	19.6
122	11767.97	2.8	17
123	35144.85	5.7	17.3
124	8038.03	3.8	35.75
125	6213.42	8	5
126	7968.12	9	20.5
127	9145.2	8.5	2.5
128	10210.83	2.6	22.2
129	2298.19	10.5	17.5
130	5308.2	18	27
131	7839.32	5	30.5
132	4427.61	4.25	30.7
133	5361.6	19	26.5
134	12041.93	11	8
135	7801.46	12	28
136	8060.67	14.5	22
137	*4076.27*	*16.7*	*28*
138	9274.58	9.5	31.5
139	13210.98	10.5	20.5
140	5666.38	10.1	4
141	11815.79	6	27.5

