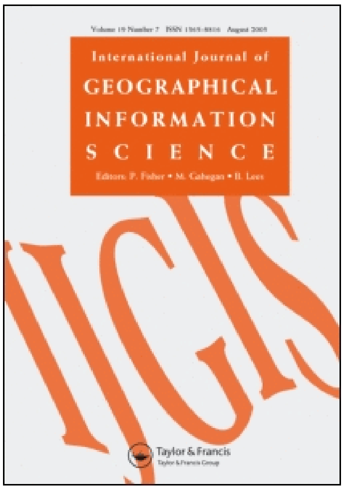

International Journal of Geographical Information Science
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713599799

Intelligent GIS for solving high-dimensional site selection problems using ant colony optimization techniques

Xia Li [a]; Jinqiang He [a]; Xiaoping Liu [a]
[a] School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

Online Publication Date: 01 April 2009

To cite this Article Li, Xia, He, Jinqiang and Liu, Xiaoping(2009)'Intelligent GIS for solving high-dimensional site selection problems using ant colony optimization techniques',International Journal of Geographical Information Science,23:4,399 — 416

To link to this Article: DOI: 10.1080/13658810801918491

URL: http://dx.doi.org/10.1080/13658810801918491

International Journal of Geographical Information Science
Vol. 23, No. 4, April 2009, 399–416

Research Article

Intelligent GIS for solving high-dimensional site selection problems using ant colony optimization techniques

XIA LI*, JINQIANG HE and XIAOPING LIU
School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

(*Received 19 October 2007; in final form 20 December 2007*)

This paper presents a new method to solve site selection problems using ant colony optimization (ACO) techniques. Optimal spatial search for siting public facilities is a common task for urban planning. The objective is to find *N* optimal sites (targets) for sitting a facility so that the total benefits are maximized or the total costs are minimized. It is straightforward to use the brute-force method for identifying the optimal solution by enumerating all possible combinations. However, the brute-force method has difficulty in solving complex spatial search problems because of a huge solution space. Ant colony optimization can provide a useful tool for site selection. In this study, the integration of ACO with geographic information systems is proposed to include various types of spatial variables in the optimization. A number of modifications have also been introduced so that ACO can fit spatial allocation problems. The novelty of this research includes the adoption of the strategies of neighborhood pheromone diffusion, tabu table adjusting, and multi-scale optimization. This method has been applied to the allocation of a hypothetical facility in Guangzhou City, China. The experiment indicates that the proposed model has better performance than the single search and the genetic algorithm for solving common site search problems.

Keywords: Ant colony optimization; GIS; Artificial intelligence; Site selection

1. Introduction

Geographic information systems (GIS) is evolving by incorporating more artificial intelligence in modelling for solving various decision-making problems. Intelligent algorithms can be used in GIS to enhance its efficiency in formulating decisions, such as complex spatial planning and resource optimization (Birkin *et al.* 1996). For example, multi-objective hybrid meta-heuristic algorithms can be used to develop a combinatorial optimization model for spatial zoning (Bong and Wang 2004). Intelligent GIS can provide an effective tool for addressing many challenges of including large solution space and multi-objectives in spatial simulation and modelling using GIS.

There are numerous studies on solving the problems of site selection, which is an important task for urban planning (Kariv and Hakimi 1979, Toregas *et al.* 1971, Hansen and Mladenovic 1997). Siting facilities have long been a subject of interest in the field of landscape architecture, although such problems can be quite 'wicked' (Church 2002). Location-allocation models that attempt to find the best sites for

*Corresponding author. Email: lixia@mail.sysu.edu.cn

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2009 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810801918491

facilities are attractive for public facilities planning (Yeh and Chow 1996). The objectives of site selection may include maximizing population coverage and minimizing total travel distance (Rushton 1984). The facility location problem and its variants, including most location-allocation and p-median problems, are known to be NP-hard combinatorial optimization problems.

Site selection usually involves a large set of spatial variables. It is a complex decision-making task because the solution space of identifying N optimal sites is usually huge. Numerous possible combinations should be examined before finding the best solution. Moreover, most of the traditional methods cannot easily handle thousands of demand points and sites, especially when GIS is used to provide detailed spatial locations and attributes (Church 1999). Some heuristic methods have been developed to tackle these hard problems. They include local optimization algorithms (e.g. the Gauss-Newton and the Levenberg-Marquart) (Zhan *et al.* 2003), and global optimization algorithms (e.g. simulated annealing and genetic algorithms) (Li and Yeh 2005).

Siting analysis with GIS began as early as in the 1970s (Kiefer and Robbins 1973, Dobson 1979). Site selection, which is a basic task of GIS functionalities, is to find the best sites for sitting a facility or a number of facilities. The search problems become very complex because all these selected sites are related and should be evaluated simultaneously for the optimization. The combined effects of all these selected sites can be assessed using utility functions, which are subject to various goals. The objective is to maximize these utility functions under the efficiency criterion. Spatial optimization becomes difficult when a richer set of GIS and high-resolution remote sensing data are available recently. For example, detailed population and transportation data can be used to calculate the utilities using GIS databases. The exhaustive blind (brute-force) method for solving optimization problems with high-dimensional spatial data is infeasible because of a huge combinatorial solution space.

Heuristic search methods have been developed to deal with the problem of huge solution space. The Monte Carlo optimization is a simple and efficient algorithm among these methods. However, this method could get stuck on local suboptimal solutions because a move is only made when a better solution is found (Openshaw and Openshaw 1997). Recently, Li and Yeh (2005) have proposed a method to solve the spatial allocation problems using genetic algorithms (GA). Unfortunately, GA has problems in reaching the convergence if the number of targets is large. For example, the optimal locations of more than 15 targets can hardly be found using GA because the required length of chromosome is too long.

Intelligent algorithms should be developed within GIS to solve complex spatial allocation problems. This study is to examine the potential of using simple ant intelligence for solving complex site selection problems. Ant colony optimization (ACO), which was first proposed by Dorigo *et al.* (1991) as a computation algorithm, is composed of a set of cooperating artificial ants with simple intelligence. Complex tasks, such as finding optimized route for seeking foods, are effectively accomplished by the mutual cooperation between individual ants. During the optimization processes, each ant agent makes a random choice according to local information (pheromone). Although there is typically no centralized control dictating the behaviour of the artificial ants, the accumulation of local interactions in time often gives arise to a global pattern (Colorni *et al.* 1991). An ant colony system is robust, and the integrity of the overall system is not easily affected by the failure of one or several agents (Dorigo *et al.* 1996).

Studies indicate that ACO has better performance than other nature-inspired algorithms, such as simulated annealing and evolutionary computation, in solving the travelling salesman problem (TSP) (Dorigo and Gambardella 1997). ACO is a compromising optimization technique for solving complex combinatorial optimization problems because of using swarm intelligence. TSP needs to examine the combinations of various possible paths for constructing the shortest route. The optimization is complex because the solution space is very huge. The mechanism of pheromone updating is effective for finding the optimal solutions using cooperating artificial ants.

Although ACO is very attractive for tackling complex and ill-defined spatial decision-making problems, it has been mainly applied to the solution of travelling salesman problems so far. In this research, a new method is proposed to integrate ACO with GIS for solving high-dimensional site selection problems. Like TSP, site selection is a kind of NP-hard combinatorial optimization problems. The identification of the optimal combination of various conditions in site selection is similar to the optimal path finding in TSP. Ant intelligence of collective behaviours can be used to handle complex heterogeneous spatial data for the optimal site search. Significant modifications will be made by incorporating the strategies of neighbourhood pheromone diffusion, tabu table updating, and multi-scale search into ACO. This is crucial for ACO to suit spatial optimization problems. This ACO method will be also compared with other conventional methods, such as the single search (SS) method and the GA method (Li and Yeh 2005), regarding their performance in site selection.

2. Ant intelligence for solving spatial search problems

2.1 *Ant intelligence for solving the TSP*

Ant colony optimization is a type of computer algorithms for solving combinatorial optimization problems using artificial intelligence and is devised by simulating ants' behaviours of selecting the best route from a food source to their nest (Dorigo and Gambardella 1997). This method is based on the positive feedback of artificial ants, in which the coordination among ants is achieved by exploiting the stigmergic communication mechanism. The pheromone provides the basic local information for the communication between ants. Ants release and deposit pheromone on the ground along their way, to guide others in finding foods efficiently. The amount of deposited pheromone will become larger on a path if it is shorter (easier) for finding foods. The larger the amount of pheromone, the more the ants will be attracted to select this path. A more amount of pheromone is in turn deposited on this path. Since the pheromone evaporates with time, the pheromone on a long path will decrease when fewer ants select it. The communication between ants based on the pheromone plays a key role in solving the path-optimization problem.

Figure 1 illustrates the process of seeking food by an ant colony, demonstrating that simple swarm intelligence can solve complex optimization problems. This swarm intelligence involves collective behaviour in decentralized systems, which are made up by a population of simple individuals interacting locally with each other and with their environment (Colorni *et al.* 1991). If there are no obstacles between ant nests and food sources, the shortest route is in a straight line (figure 1(*a*)). The attraction is that an ant colony with simple intelligence not only fulfills complex tasks but also adapts to the changes of surrounding environment. For example, if

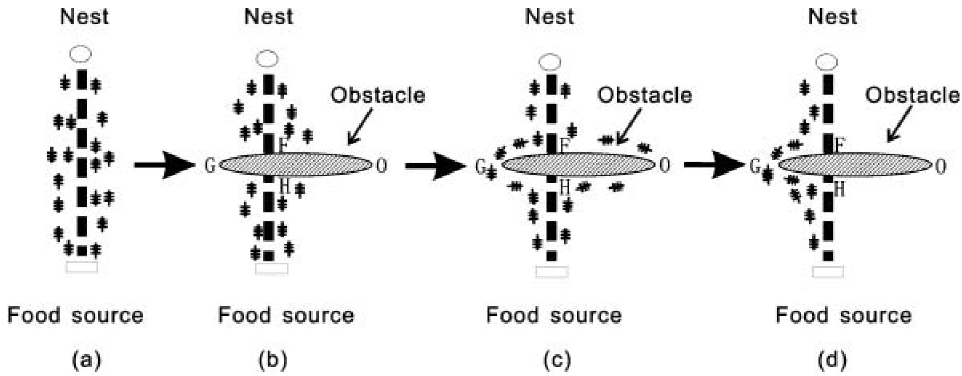

Figure 1. Ant intelligence of finding optimal route in seeking foods.

there exists an obstacle occurring on the route, ants can find the optimal solution based on simple swarm intelligence. At the beginning, ants select various routes by identical probability (figure 1(*b*)). During their movement, ants will deposit pheromone on routes that they passed by. Since the route F-G-H is shorter than F-O-H, the ants selecting the route F-G-H will reach the food source sooner than those selecting the route F-O-H. As a result, a larger amount of pheromone will be deposited on H-G-F than on H-O-F. This will attract more ants to select the route H-G-F (figure 1(*c*)). At the final stage, all the ants will choose the route H-G-F because the pheromone on the longer route gradually disappears (figure 1(*d*)). The above food seeking process based on positive feedback indicates that ACO is self-adaptive.

The first experiment for testing ACO was the TSP (Dorigo *et al.* 2000). TSP is to search for a closed tour of minimal length connecting N given cities. This ACO method is to add new cities to a partial solution by exploiting both information gained from past experience and a greedy heuristic (Dorigo and Gambardella 1997). Each ant constructs a TSP solution in an iterative way. An ant chooses a city to visit with a probability that is related to the amount of pheromone trail $\tau_{uv}(t)$ on the path and the travel distance. A tabu list, $tabu_k$, is used to prevent an ant going to the visited cities again. This transition probability from city u to city v for the kth ant at time t is thus given as follows

$$p_{uv}^{k}(t)=\begin{cases}\dfrac{[\tau_{uv}(t)]^{\alpha}\cdot[\eta_{uv}(t)]^{\beta}}{\sum\limits_{v\in allowed_k}[\tau_{uv}(t)]^{\alpha}\cdot[\eta_{uv}(t)]^{\beta}} & \text{if } v\in allowed_k \\ 0 & \text{otherwise}\end{cases} \tag{1}$$

where $\tau_{uv}(t)$ is the amount of pheromone trail on edge (u,v), $\eta_{uv}(t)$ is a heuristic function related to the visibility (distance). The set, $allowed_k=\{C-tabu_k\}$, represents the cities that can be visited next time without repetition.

The parameters of α and β control the relative importance of trail versus visibility (distance). A larger value of α indicates that the trail intensity has more effect on the probability. As a result, it is highly desirable for the probability if there is a lot of traffic on edge (u,v). A larger value of β means that more effect will be put on the visibility (distance) factor.

An ant has a higher probability of selecting the shorter route between two cities. The heuristic function $\eta_{uv}(t)$ is defined as the inverse of the distance between cities u

and v

$$\eta_{uv}(t) = \frac{1}{d_{uv}} \tag{2}$$

where d_{uv} is the distance between city u and city v.

At each iteration t, the trail density is updated according to the following formula

$$\tau_{uv}(t+1) = (1-\rho)\tau_{uv}(t) + \Delta\tau_{uv}(t) \tag{3}$$

$$\Delta\tau_{uv}(t) = \sum_{k=1}^{m} \Delta\tau_{uv}^{k}(t) \tag{4}$$

where ρ is a coefficient such that (1–ρ) represents the evaporation of trail between t and $t+\mathrm{n}$. $\Delta\tau_{uv}^{k}(t)$ is the quantity per unit of length of trail substance laid on path (u,v) by the kth ant between time t and $t+\mathrm{n}$.

$\Delta\tau_{uv}^{k}(t)$ is calculated using the following equation

$$\Delta\tau_{uv}^{k}(t) = \begin{cases} \frac{Q}{L_k} & \text{if the } k\text{th ant visits } (u,\, v) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where Q is a constant, and L_k is the tour length of the kth ant.

2.2 *Modified ACO for solving facility sitting problems*

In this study, distributed ants' intelligence is used to solve the hard optimization problems of sitting facilities. The proposed method is devised according to the ACO algorithm for TSP. In TSP, the question is to find a closed tour of minimal length connecting N given cities. ACO is dependent on two terms, the trail density and the visibility (distance), to choose the optimal route. The same concept is used to develop the site search algorithm.

The optimal sites are heuristically identified based on the trail density and the visibility (distance). The objective is to find the best N locations (targets) that can produce the largest value of a utility function. The optimal site selection is fulfilled using pheromone updating of ACO. In the initial stage, each cell will have an equal amount of pheromone. A certain amount of pheromone will be deposited on the cells visited by an ant. The combination of cells for siting the facility is evaluated according to a utility function. It is expected that ants are likely to visit the selected cells of higher utility values according to the greedy criterion. As a result, more amount of pheromone will be deposited on these potential cells. This will in turn attract more ants to visit them.

According to equation (1), the probability that a cell (x) can be selected for a visit by the kth ant at time t is estimated as follows

$$p_{x}^{k}(t) = \begin{cases} \dfrac{[\tau_x(t)]^{\alpha}\cdot[\eta_x(t)]^{\beta}}{\sum\limits_{x \in allowed_k} [\tau_x(t)]^{\alpha}\cdot[\eta_x(t)]^{\beta}} & \text{if } x \in allowed_k \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

A heuristic function is defined to obtain better convergence rates during the spatial search. It is assumed that a selected site should have a large population density. This can minimize the total transportation cost. Therefore, the heuristic function $\eta_x(t)$ is defined as follows

$$\eta_x(t) = p_{den}(x) \tag{7}$$

where $p_{den}(x)$ is the population density at cell x.

In TSP, L_k represent the tour length of the solution for the kth ant. The tour length of the solution is replaced by the total cost for solving site selection problems. It is expected that the optimal solution should be able to produce the minimum amount of transportation cost for visiting the N sites of the facility. Therefore, the total transportation cost for visiting the N sites (targets) of the facility is given by the following equation

$$L_{ktrans} = \sum_x d(x) p_{den}(x) A \tag{8}$$

where L_{ktrans} is the total transportation cost for the kth ant, $d(x)$ is the Euclidian distance between cell x and the closest target, and A is the area of each cell.

$1/L_{ktrans}$ can be regarded as the total utility for siting the N sites (targets) of the facility. More spatial variables can be included to define the utility function based on the domain knowledge. These variables may be related to the benefits and costs for siting the facility in these locations. Various proximity variables (e.g. distances to urban centres, roads, and airports) can be defined using GIS. Multi-criteria evaluation could be used to combine all these various spatial factors. In this study, the modified total transportation cost (L'_{ktrans}) for siting N targets is calculated by the following equation

$$L'_{ktrans} = w_1 \sum_x d(x) p_{den}(x) A + w_2 \sum_{d=1}^{N} e^{-\rho D_{road}} \tag{9}$$

where D_{road} is the distance between the selected site (target) and the closest roads. The parameter of ρ is set to 0.1, and w_1 and w_2 are set to 0.8 and 0.2 respectively. The first item is given with more importance.

It is also essential to define a diffusion strategy for pheromone updating for fitting the site selection problems. The pheromone will evaporate very fast because the selected cells only amount to a small percentage of all the cells. The positive feedback will be too weak to play a role in the optimization. A modification is to incorporate the strategy of neighborhood pheromone diffusion in defining pheromone updating. This is conducted by incorporating a distance decay function in equation (5). The revised equation is presented as follows

$$\Delta\tau_x^k(t) = \begin{cases} \frac{Q}{[d(x)+1]\cdot L'_{ktrans}} & \text{If } x \text{ falls within } 5\times 5 \text{ window of a closest target at time } t \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Like TSP, this method uses a tabu list, $tabu_k$, to mask out the selected cells which should not be visited again next time. Moreover, this list also marks the restricted cells (e.g. water and hills) which are excluded in site selection. This type of constraints is prepared using the ArcGIS package.

Another important modification is to adopt a multi-scale approach, which can alleviate the computational demand in large-scale spatial search. This includes two phases of optimization. First, the identification of approximate locations of targets is carried out in a coarser resolution. It is expected that the exact locations should be around these initial locations. Then the next round of optimization is implemented by just searching the neighborhood around these initial locations in the original resolution. This two-phase procedure of optimization can thus significantly reduce the computation time.

The following is the computation algorithm for implementing this site selection procedure:

ALGORITHM - A High-level Description of ACO for site selection
Set $t=0$
For every cell(x) set an initial value $\tau(0)=C$ and $\Delta\tau_x(0)=0$
/* The amount of pheromone is set to be the same for all the cells at the beginning */
Do while ($t<$ *Maxnum_iterations* and $M<$ *Num_convergence*)
/* *Maxnum_iterations* represents the maximum number of iterations, and *Num_convergence* is the total number of iterations for examining the convergence */
For $K=1$ to *Num_of_ants*
$tabu_k$=NULL
For $x=1$ to *Num_of_sites* /* Obtain N optimal site selection result*/
From all the cells, choose a cell as a potential sitting site with the probability of $p_x^k(t)$ defined by:

$$p_x^k(t)=\begin{cases}\frac{[\tau_x(t)]^\alpha\cdot[\eta_x(t)]^\beta}{\sum\limits_{x\in allowed_k}[\tau_x(t)]^\alpha\cdot[\eta_x(t)]^\beta}, & \text{if } x\in \text{allowed}_k\\ 0 & \text{otherwise}\end{cases}$$

Include this cell into $tabu_k$ if selected
Next x
Evaluate the selection result found by the kth ant
If (This result is better than the best previous found result) then
Replace the best previous result with this one
End if
If (The result is the same as the previous one)
/* to examine the convergence rate*/
then
$M=M+1$
Else
$M=0$
End if
Next K
For every cell(x) /* Pheromone update */
For $k=1$ to *Num_of_ant*

$$\Delta\tau_x^k(t)=\begin{cases}\frac{Q}{(d(x)+1)\cdot L'_{ktrans}}, & \text{If } x \text{ falls within } 5\times5 \text{ window of a closest target at time } t\\ 0, & \text{otherwise}\end{cases}$$

$$\Delta\tau_x(t)=\Delta\tau_x(t)+\Delta\tau_x^k$$

Next x

For every cell(x)
$\tau_x(t+1)=\rho\cdot\tau_x(\mathrm{t})+\Delta\tau_x(t)$
$\Delta\tau_x(t)=0$
$t=t+1$
Loop

Figure 2 illustrates the process of this heuristic spatial search based on pheromone updating. In the initial stage, the amount of pheromone is set to the same for all the cells. During the optimization process, a higher amount of pheromone will be deposited on the cells if the siting can produce higher values of the total utility ($1/L'_{ktrans}$). At the final stage, the optimal N locations for siting the facility are identified by these artificial ants according to the pheromone updating.

3. Implementation of ACO for site selection

3.1 *Study area and model parameters*

The study area is situated in the urban districts of Guangzhou, including the districts of Haizhu, Yuexiu, Liwan, Tianhe and Baiyun. Since these districts are densely populated, the provision of adequate facilities is essential for serving the basic needs of the population. This proposed method is used to identify the N optimal sites of siting a hypothetical facility (e.g. schools) for serving the population. This optimization problem considers two spatial variables (population distribution and transportation conditions) which are prepared using the ArcGIS package. The population data available at street-blocks are obtained from the 2003 census of Guangzhou. The pattern of population density in the study area is shown in figure 3. Road networks are also retrieved from the GIS database (figure 4). These spatial

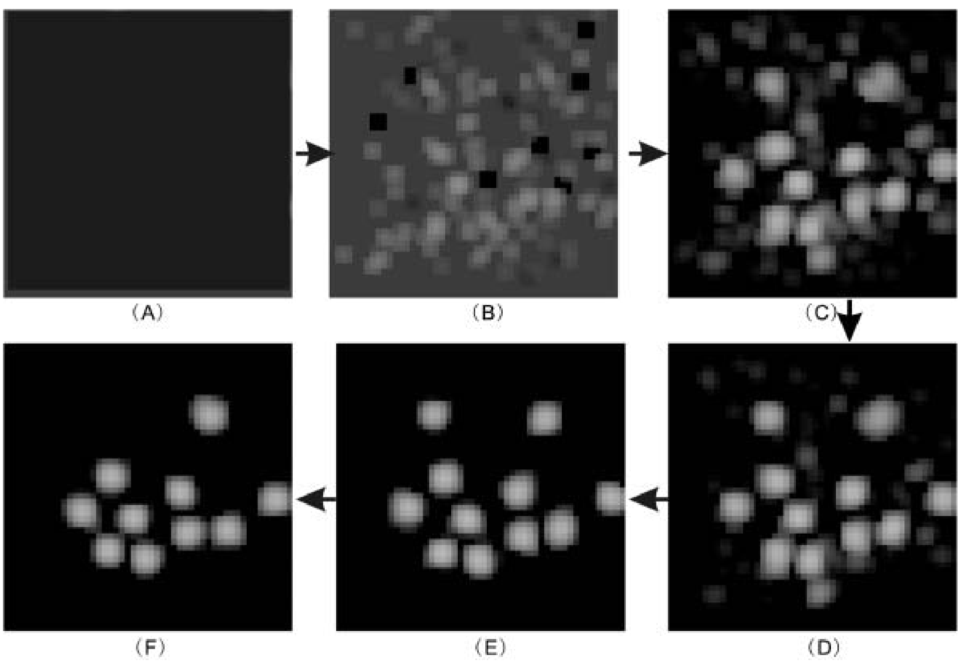

Figure 2. Optimal site selection according to the pheromone updating of ACO.

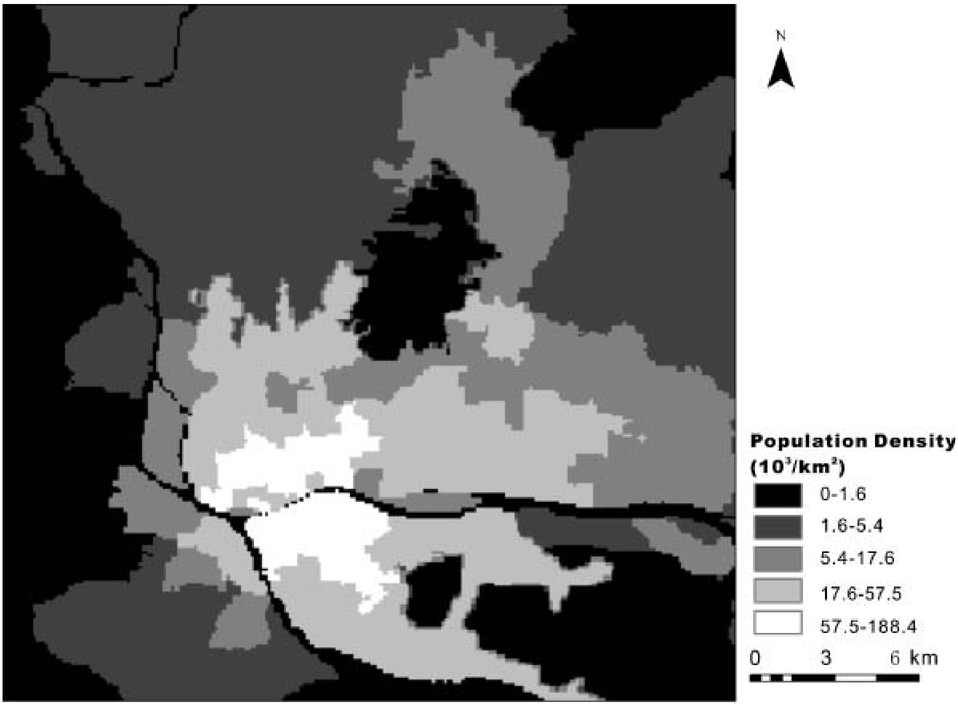

Figure 3. The population density for the districts of Guangzhou.

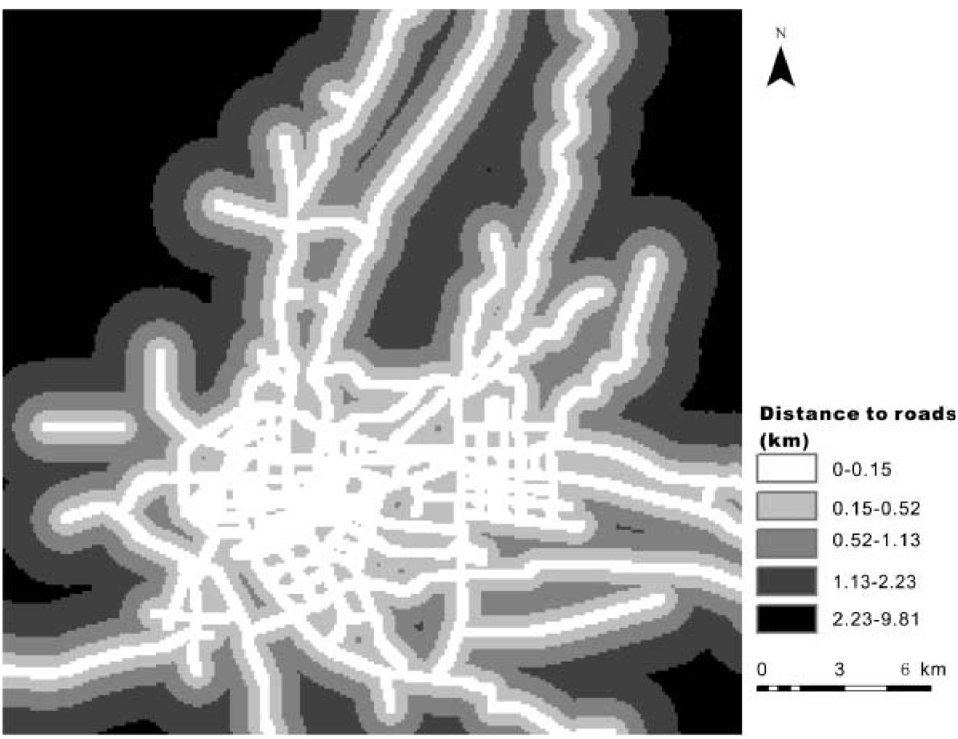

Figure 4. Road networks of the study area from the GIS database.

factors have formed a quite heterogeneous surface which requires sophisticated heuristic methods to find the optimal solution.

The proposed model is developed using Visual Basic and ArcObjects of ArcGIS. The use of ArcObjects can allow easy access to the spatial data and functionality of GIS. Like many GIS-based site selection models, this model is performed based on raster data format for modelling convenience. The original GIS vector (polygon) data of the study area are converted into raster layers as the inputs to the modelling. The raster layers have a resolution of 100×100 m with a size of 250×250 pixels. Brute-force algorithms have difficulties in finding the optimal solution since the process involves numerous possible combinations (Li and Yeh 2005). For example, the number of the combinations is as large as $\frac{(250 \times 250)!}{10! \times (250 \times 250 - 10)!} = 2.50 \times 10^{41}$ if a brute-force algorithm is used to find the optimal solution of identifying 10 targets.

This ACO method involves some parameters which could have impacts on simulation results. In equation (1), the parameters of α and β control the relative importance of trail versus visibility (distance). These parameters should be defined before running the simulation. Since the simulation itself will take time to complete, the enumeration of all possible combination of α and β is impossible for examining their effects. Alternatively, a two-step procedure is proposed to determine the suitable values of these parameters for site selection. The first step is to examine the influences of α (α=0.5, 1.0, 1.5, 2.0, 2.5) when β is initially set to 0.5. Table 1 lists the average utility value for each combination by repeatedly running this ACO model 10 times. It is found that the optimal average utility value is obtained when α falls within the range of [1, 2]. A large value of α will let the model get stuck on suboptimal solutions because the convergence rate is too quick. The suitable value of α is then set to 1.5. The second step is to examine the influences of β (β=0.5, 1.0, 1.5, 2.0, 2.5) when α is set to its optimal value (α=1.5). It is found that the optimal average utility value is obtained when β falls within the range of [0.2, 1.0]. A large value of β will also let the model get stuck on suboptimal solutions because the convergence rate is too quick. Therefore, the best combination of α and β is 1.5 and 0.6 according to this experiment.

3.2 *Spatial search using a multi-scale ACO approach*

Site selection usually needs to handle a large size of spatial data for solving a practical problem. The solution space is still too large for the distributed ACO algorithm. A multi-scale search is required for reducing the computation time. This procedure can be illustrated in the search for six optimal sites for siting a facility (e.g. schools) using raster GIS data (figure 5). The raster GIS data include the grids

Table 1. Search for suitable values of importance of trail (α) and importance of visibility (β).

Step 1	Importance of trail (α)	Importance of visibility (β)=0.5					
		0.5	1.0	1.5	2.0	2.5	3.0
	Average total utility value (10^{-6})	7.77	8.19	8.18	8.17	8.04	7.93
	Stabilized iterations	374	459	377	387	483	188
Step 2	Importance of trail (β)	Importance of visibility (α)=1.5					
		0.0	0.2	0.5	1.0	1.5	2.0
	Average total utility value (10^{-6})	8.05	8.15	8.16	8.17	8.08	7.80
	Stabilized iterations	442	388	377	396	310	141

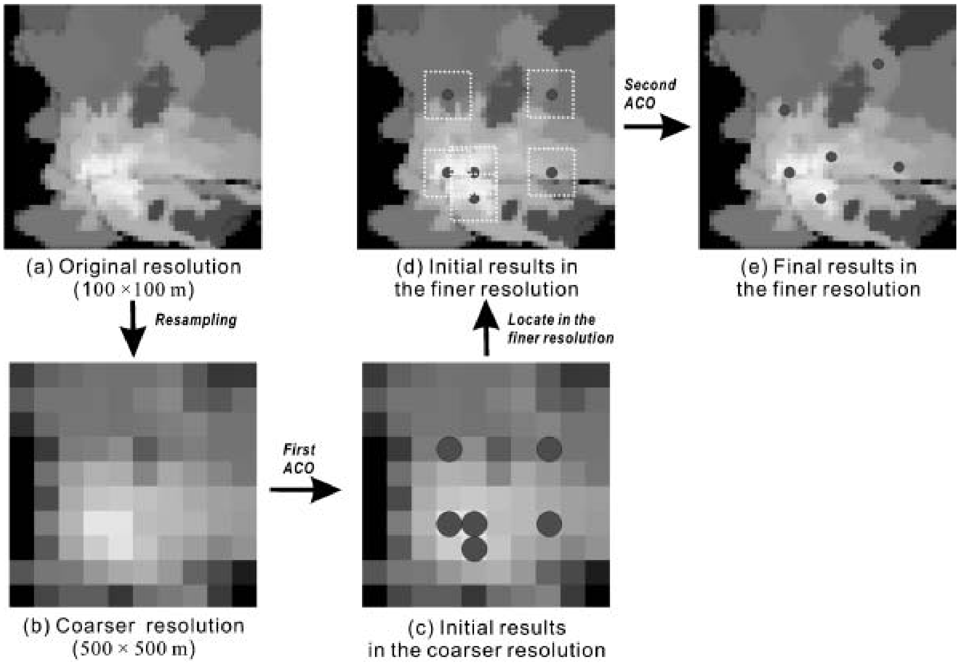

Figure 5. A multi-scale approach of using ACO for site selection.

of population density and transportation with the resolution of 100×100 m. The original resolution was converted to the coarser resolution of 500×500 m. As a result, the data volume of the new layers only amounts to 1/25 of the original one. ACO can thus have much faster computation speed using these down-scaling data.

The initial coordinates for the six targets were identified on the grids of 500×500 m resolution. The exact locations for these targets need a further step of the search on the grids of 100×100 m resolution. However, the new search is only limited to a small neighborhood around these initial targets. A window of 5×5 pixels around the coarse locations was used for executing this ACO algorithm again. This experiment indicates that this multi-scale search can effectively reduce the data volume and make the computation feasible for solving practical problems.

This ACO method has been compared with the SS algorithm and GA to evaluate its performance in site selection. The SS algorithm is similar to the bombing strategy described by Openshaw and Steadman (1982). This algorithm can be used to find *N* sites (targets) for siting a facility by assessing each site separately. Although the maximum utility value can be obtained for each independent site, the total utility value is not maximal without considering the combined effects of these sites. As a result, this SS method cannot guarantee that the solution will produce the maximum population coverage. The only advantages of this method are its model simplicity and fast calculation speed. However, its effectiveness of producing the optimal results is in doubt for solving high-dimensional problems.

An alternative method is to apply the GA method for siting facilities optimally (Li and Yeh 2005). This GA method has the advantages of dealing with the combined effects of all targets using chromosomes. For example, the GA method can generate the best utility values with 18.2% larger than the SS method for allocating 10 targets of a facility (Li and Yeh 2005).

Table 2. Comparison of the improvement of the total utility value between the SS, GA and ACO methods in site selection based on the resolution of 500×500 m.

Number of targets	2	4	6	8	10
	Total utility value (10^{-6})				
SS	3.7	5.0	6.0	6.9	7.7
GA	3.9	5.2	6.2	7.1	8.0
ACO	3.9	5.3	6.3	7.2	8.2
(ACO-SS)/SS	5.3%	5.5%	2.7%	3.9%	2.8%
(ACO-GA)/GA	0.0%	0.4%	1.3%	1.2%	2.5%

The performance of these three methods, the SS, GA and ACO, is compared based on the total utility $(1/L'_{ktrans})$ using the same set of data. It is expected that a better method should generate a higher utility value. Site selection usually involves very large volumes of spatial data and needs a very long time for the optimal search. For example, the SS algorithm needs to calculate 625000 times of the utility function if the image size of the study area is 250 columns $\times$ 250 rows and the number of the targets is 10. The search takes 83 h using a computer with a Pentium IV 3.2 GHz CPU. Therefore, the comparison was carried out using a coarser resolution of 500×500 m.

Table 2 compares the total utility values of these three methods at this coarser resolution. It is found that the proposed ACO method has 2.7–5.5% improvement of the total utility value over the SS method. However, the ACO method only has slight improvement of the total utility value over the GA method. Figure 6 further compares the convergence rates of the GA method and the ACO method for finding the optimal sites in terms of the utility function. Both methods can quickly reach the convergence, but the ACO method has better performance.

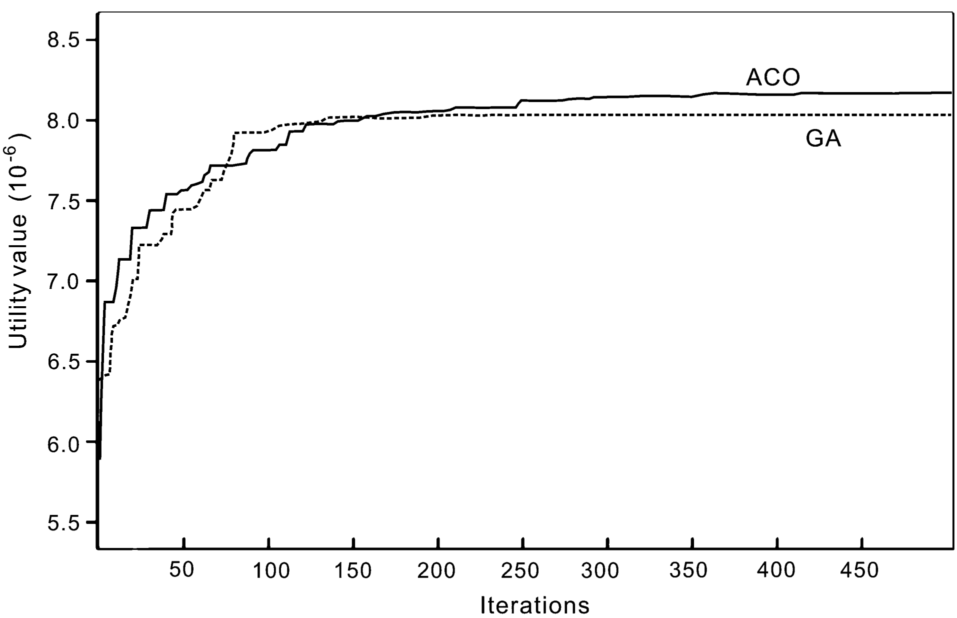

Figure 6. The convergence rate of the GA and ACO methods for finding the optimal sites in terms of the utility function.

Table 3. Comparison of the computation time between the SS, GA and ACO methods based on the resolution of 500×500 m.

Number of targets	2	4	6	8	10
			Computation time (s)		
SS	32	95	198	333	525
GA	45	138	247	419	832
ACO	4	28	41	54	73
ACO/SS	12.5%	29.5%	20.7%	16.2%	13.9%
ACO/GA	8.9%	20.3%	16.6%	12.9%	8.8%

Table 3 lists the computation time of these methods for the site selection. The ACO method needs much less computation time than both the SS method and the GA method. The computation time of this ACO method only amounts to 12.5–29.5% and 8.8–20.3% of those of the SS method and the GA method respectively.

The proposed multi-scale strategy can be adopted for solving real-world problems which usually involve a large size of spatial data. In this study, the original resolution of these raster data is 100×100 m and the targets to be identified range from 1 to 30. The initial site selection is carried out in the grids of 500×500 m resolution. The computation time is significantly reduced since the data volume is reduced to 1/25 of the original. It is expected that the accurate locations for N targets should be around these initial selected sites. Therefore, the exact locations can be further identified using a 5×5 window around these initial sites. The experiment indicates that the multi-scale ACO has many more advantages of reducing the computation time than the SS, genetic algorithms and the single ACO (Table 4). Although the SS is very simple, it still needs to take very long time to complete the cell-by-cell search if the space contains a large number of cells. Figure 7 illustrates the final results for siting 10 and 30 targets of the facility optimally under the complex distribution of population and transportation conditions.

A further experiment was carried out to examine if this multi-scale ACO has accuracy loss due to the data reduction. The validation is based on the difference of the utility value between the multi-scale ACO method and the single ACO method. It is expected that these two methods can produce the same value or very close values of the total utility. A small sub-area was used for the comparison with the number of targets ranging from 1 to 16. Table 5 is the comparison of their total utility values. It is found that these two methods can obtain almost the same result in terms of the utility value. When the number of targets is small (e.g. $N<5$), these two methods have the same values of the total utility. Therefore, this multi-scale search can produce a satisfactory accuracy for finding the optimal sites using much less computation time.

Table 4. Comparison of the computation time for identifying 10 targets using the SS, GA and ACO methods for the study area.

Algorithm	Computation time (h)	Total utility value (10^{-6})
SS	82.5	7.7
GA	5.2	8.0
single ACO	2.1	8.2
multi-scale ACO	0.5	8.2

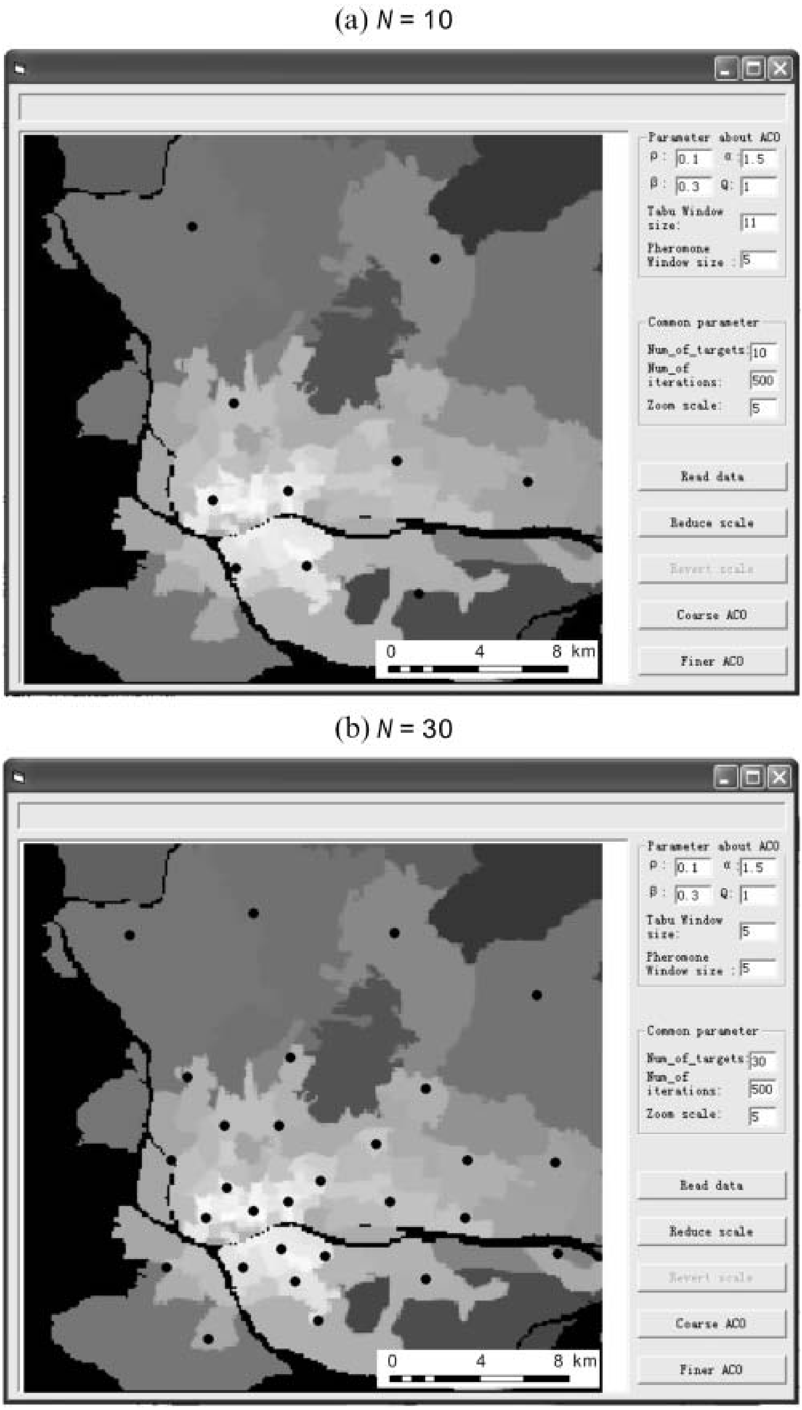

Figure 7. Identifying 10 and 20 optimal sites for siting a hypothetical facility using ACO.

Like other heuristic methods, this ACO method may be affected by some randomness. It is necessary to carry out many simulations to see if ACO can produce stable results. Figure 8 shows the 10 simulations of ACO and their overlaid

Table 5. Comparison of the total utility value between the multi-scale ACO method and the single ACO method.

Number of targets	multi-scale ACO	single ACO	multi-scale/ single	**Number of Targets**	multi-scale ACO	single ACO	multi-scale/ single
	Total utility values (10^{-6})				Total utility values (10^{-6})		
1	0.609	0.609	1.0000	**9**	1.772	1.771	1.0006
2	0.786	0.786	1.0000	**10**	1.858	1.860	0.9989
3	1.028	1.028	1.0000	**11**	1.967	1.967	1.0001
4	1.176	1.176	1.0000	**12**	2.064	2.061	1.0013
5	1.310	1.310	1.0000	**13**	2.149	2.148	1.0007
6	1.452	1.452	1.0003	**14**	2.230	2.230	0.9997
7	1.567	1.567	0.9998	**15**	2.322	2.323	0.9996
8	1.666	1.666	1.0000	**16**	2.400	2.400	0.9998

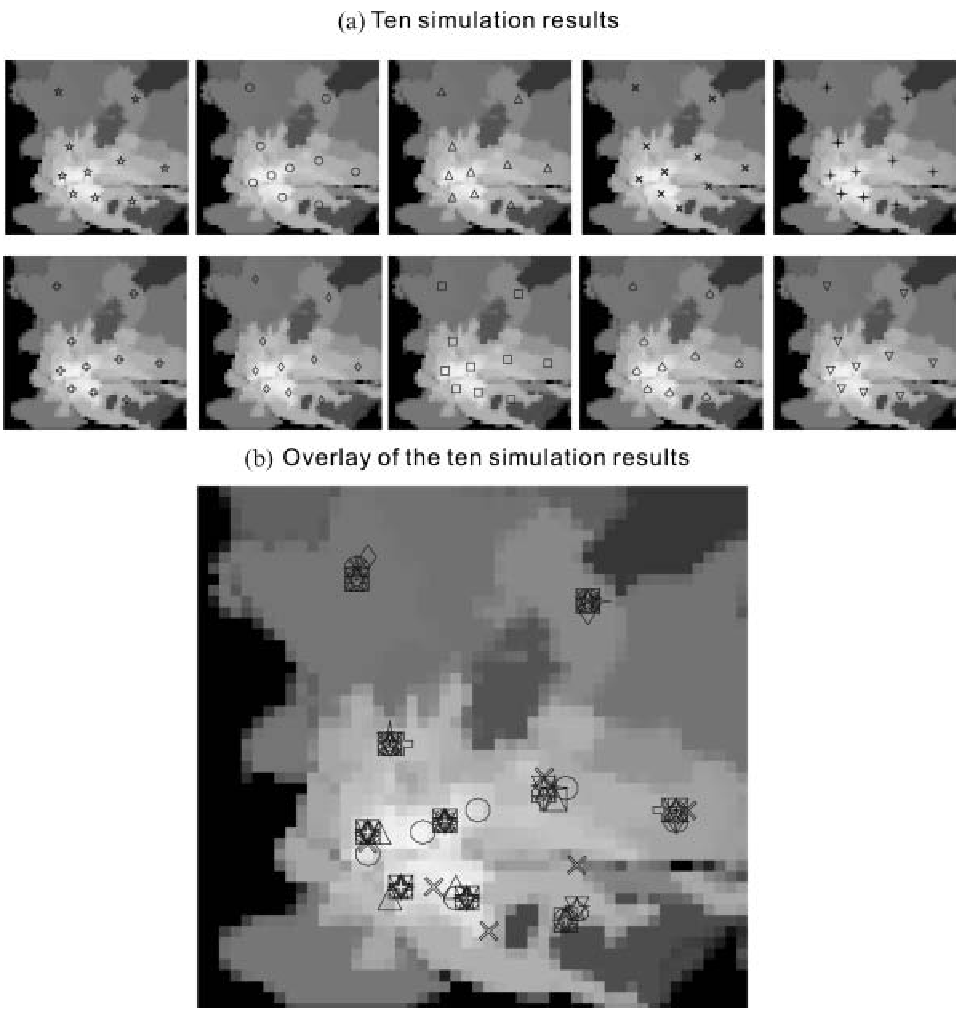

Figure 8. Overlay of the 10 simulation results of the ACO method.

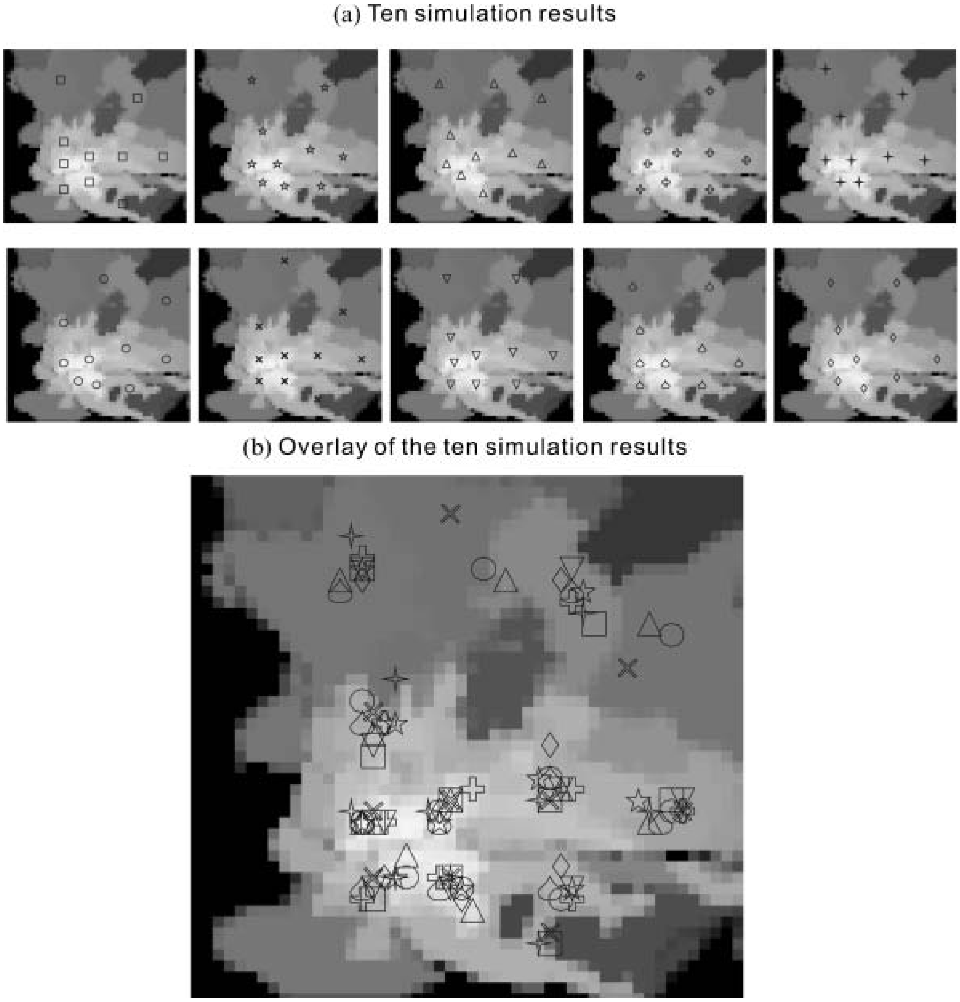

Figure 9. Overlay of the 10 simulation results of the GA method.

results. The good overlapping indicates that this ACO method can produce quite stable simulation results. Figure 9 is the 10 simulations of GA and their overlaid results. It has poorer overlapping results, compared with the proposed method. Therefore, ACO can produce much more stable simulation results than GA for site selection. Table 6 further compares the total utility value and the standard deviation of these 10 simulations for these two methods. ACO can produce higher utility value and smaller standard deviation (SD). This means that ACO can have better capability to avoid the trapping in suboptimal solutions.

Table 6. Total utility value of 10 simulations of ACO and GA (10^{-6}).

No.	1	2	3	4	5	6	7	8	9	10	Mean	SD
ACO	8.20	8.09	8.19	8.11	8.18	8.19	8.18	8.19	8.20	8.20	*8.17*	*0.039*
GA	8.07	8.10	7.84	8.04	7.93	7.96	7.93	8.08	7.96	8.05	*8.00*	*0.084*

4. Conclusions

Facility planning usually needs to determine N optimal sites for siting a facility. One of the objectives is to generate the maximum utility value by considering the combined effects of these sites subject to various spatial variables. The solution space grows exponentially with the increase in the size of study areas and the number of spatial variables. Enumerative techniques, based on the enumeration of the partial solutions, have difficulties in solving this problem. Heuristic methods have been developed to improve the performances of spatial search. In this study, a heuristic method for site selection has been proposed based on ACO techniques. The novelty of this proposed method includes: the integration of ACO with GIS; the adoption of the strategies of neighborhood pheromone diffusion; the definition of tabu table with constraints and multi-scale searches.

First, integrating ACO with GIS is important for solving practical problems in site selection. The integration allows these two techniques to be mutually benefited from each other. ACO provides an efficient distributed computation algorithm while GIS provides useful spatial information. Second, ACO is modified to address spatial influences in updating pheromone. The strategy of neighborhood pheromone diffusion is adopted to keep pheromone from evaporating too fast. This strategy is important for site selection because the target cells are very few compared to all the cells. Moreover, a tabu table is defined to exclude not only the selected sites, but also some restricted ones based on GIS data. Last, a multi-scale search procedure is proposed to reduce computation time for handling real sets of spatial data. The procedure includes two phases of optimization. ACO is first used to find approximate locations for sitting a facility using a coarser resolution. Then the next search for the finer locations is just carried out within a window of these approximate locations using the original resolution.

This multi-scale ACO was applied to the spatial search of sitting a hypothetical facility in Guangzhou. Experiments indicate that this multi-scale ACO method can produce similar results but use much less computation time, compared with the single ACO method. Good position accuracies can be maintained although this approach is based on an approximation method. This method has better performance than conventional methods, such as the SS method and the GA method, for solving site search problems. ACO has yielded the utility improvement of 2.7–5.5%, compared with the SS method. Although ACO has slight improvement of the total utility value over the GA method, the former is able to reduce computation time significantly. Its computation time only amounts to 12.5–29.5% and 8.8–20.3% of those of the SS method and the GA method respectively.

Like other heuristic search, such as genetic algorithms, the proposed ACO method could have a chance to get stuck on suboptimal solutions. However, comparisons indicate that this ACO method is less likely to trap in suboptimal solutions than the GA method. Many simulations should be carried out to reduce the chance of getting stuck on suboptimal solutions. Further studies are also needed to examine if the proposed multi-scale approach makes reaching a suboptimal solution more likely.

Acknowledgement
This study was supported by the Key National Natural Science Foundation of China (Grant No. 40830532), the National Outstanding Youth Foundation of

China (Grant No. 40525002), and the Hi-tech Research and Development Program of China (863 Program) (Grant No. 2006AA12Z206).

References

BIRKIN, M., CLARKE, G., MARTIN, P. and CLARKE, A.W., 1996, *Intelligent GIS: Location Decisions and Strategic Planning*, p. 304 (New York: John Wiley & Sons, Inc.).

BONG, C.W. and WANG, Y.C., 2004, An intelligent GIS-Based spatial zoning system with multiobjective hybrid metaheuristic method. *Lecture Notes in Computer Science*, **3029**, pp. 769–778.

CHURCH, R.L., 1999, Location modeling and GIS. In *Geographical Information Systems*, P.A. Longley *et al.*, (Eds) Volume 1, pp. 293–303 (New York: John Wiley & Sons, Inc.).

CHURCH, R.L., 2002, Geographical information systems and location science. *Computers & Operations Research*, **29**, pp. 541–562.

COLORNI, A., DORIGO, M. and MANIEZZO, V., 1991, Distributed optimization by ant colonies. In *Proceedings of the 1st European Conference on Artificial Life*, pp. 134–142 (Paris: The MIT Press).

DOBSON, J.A., 1979, Regional screening procedure for land use suitability analysis. *The Geographical Review*, **69**, pp. 224–234.

DORIGO, M., BONABEAU, E. and THERAULAZ, G., 2000, Ant algorithms and stigmergy. *Future Generation Computer Systems*, **16**, pp. 851–871.

DORIGO, M. and GAMBARDELLA, L.M., 1997, Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transaction on Evolutionary Computation*, **1**, pp. 53–56.

DORIGO, M., MANIEZZO, V. and COLORNI, A., 1991, Positive feedback as a search strategy. *Technical Report* No. 91–106 (Milan: Politecnico di Milano).

DORIGO, M., MANIEZZO, V. and COLORNI, A., 1996, Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics – Part B*, **26**(1), pp. 29–41.

HANSEN, P. and MLADENOVIC, N., 1997, Variable neighborhood search for the *p*-median. *Location Science*, **5**, pp. 207–226.

KARIV, O. and HAKIMI, S.L., 1979, An algorithmic approach to network location problems. Part 1: The p-centers. *SIAM Journal Applied Mathematics*, **37**(3), pp. 513–538.

KIEFER, R.W. and ROBBINS, M.L., 1973, Computer-based land use suitability maps. *Journal of the Surveying and Mapping, Division-ASCE*, **99**, pp. 39–62.

LI, X. and YEH, A.G.O., 2005, Integration of genetic algorithms and GIS for optimal location search. *International Journal of Geographical Information Science*, **19**(5), pp. 581–601.

OPENSHAW, S. and OPENSHAW, C., 1997, *Artificial Intelligence in Geography* (Chichester: John Wiley & Sons).

OPENSHAW, S. and STEADMAN, P., 1982, On the geography of a worst case nuclear attack on the population of Britain. *Political Geography Quarterly*, **1**, pp. 263–278.

RUSHTON, G., 1984, Use of location-allocation models for improving the geographical accessibility of rural services in developing countries. *International Regional Science Review*, **9**, pp. 217–240.

TOREGAS, C., SWAIN, R., REVELLE, C. and BERGMAN, L., 1971, The location of emergency service facilities. *Operations Research*, **19**, pp. 1363–1373.

YEH, A.G.O. and CHOW, M.H., 1996, An integrated GIS and location-allocation approach to public facilities planning – an example of open space planning. *Computers, Environment and Urban Systems*, **20**, pp. 339–350.

ZHAN, H.G., LEE, Z.P., SHI, P., CHEN, C.Q. and CARDER, K.L., 2003, Retrieval of water optical properties for optically deep waters using genetic algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, **41**(5), pp. 1123–1128.