Evaluation and Development of Sustainable Urban Land Use Plans Through Spatial Optimization

Jing Yao[1], Alan T. Murray[2], Jing Wang[3], Xiaoxiang Zhang[4*]

[1]Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow G12 8RZ, Scotland, UK.

[2]Department of Geography, University of California at Santa Barbara, Santa Barbara, CA 93106, USA.

[3]School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China.

[4]School of Earth Sciences and Engineering, Hohai University, Nanjing 210029, China.

Emails: JY Jing.Yao@glasgow.ac.uk

ATM amurray@ucsb.edu

JW wangjing-whu@whu.edu.cn

XZ xiaoxiang@hhu.edu.cn (*Corresponding author)

Evaluation and Development of Sustainable Urban Land Use Plans Through Spatial Optimization

Abstract

Along with rapid global urbanization, cities are challenged by environmental risks and resource scarcity. Sustainable urban planning is central to address the dilemma of economic growth and ecosystem protection, where the use of land is critical. Sustainable land use patterns are spatially explicit in nature, which can be structured and addressed using spatial optimization integrating geographic information systems (GIS) and mathematical models. This research discusses prominent sustainability concerns in land-use planning and suggests a generalized multi-objective spatial optimization model to facilitate conventional planning. The model is structured to meet land use demand while satisfying the requirements of the physical environment, society and economy. Unlike existing work relying on raster data due to its simple data structure and ease of spatial relationship evaluation, this research develops an approach for identifying land use solutions based on vector data that better reflects the actual shape and spatial layout of land parcels as well as the ways land-use information is managed in practice. An evolutionary algorithm is developed to find the set of efficient (Pareto) solutions given the complexity of vector-based representations of space. The proposed approach is applied in an empirical study of Dafeng, China in order to support local urban growth and development. The results demonstrate that spatial optimization can be a powerful tool for deriving effective and efficient land use planning strategies. A comparison to results using a raster data approach supports the superiority of land use optimization using vector data as part of planning practice.

Keywords: Land-use planning; Sustainable development; Evolutionary algorithm;

NSGA-II; Spatial optimization

1 Introduction

Urban landscape worldwide has undergone remarkable changes during the massive global urbanization over the last few decades. The urban area has increased by 58,000 km^2 between 1970 and 2000 (Seto et al., 2011), with a projected expansion of 1.2 million km^2 by 2030 given current trends in population and urban growth (Seto et al., 2012). The conversion to urban land from other land uses (e.g. farmland, woodland and pasture) has resulted in serious consequences like land degradation, reduced biodiversity, intensified soil erosion and fragmented habitat (Elmqvist et al., 2013). Also, rapid growth in urban populations has greatly increased the demand for services like housing and transport, leading to conflicting land uses. Sustainable urban planning is central to address the dilemma of urban growth and ecosystem protection, where the use of land is critical. Further, sustainable land-use planning has significant implications for quality of life, necessitating assessment and regulation of land uses in the context of society, economy, environment and ecosystem in order to mitigate land-use conflicts and promote long-term balanced development (Healey, 2006).

From an operational perspective, sustainable land-use planning involves arrangement of various land uses over geographic space in order to meet the demand of diverse activities, often constrained by economic, social and environmental conditions (Aerts et al., 2005; Ligmann-Zielinska et al., 2008; Stewart & Janssen, 2014). Land-use planning problems are therefore spatial in nature and can be structured and addressed using spatial optimization integrating GIS and mathematical models (Church, 2002; Murray, 2010, 2017). GIS not only facilitates the management, manipulation and analysis of land-use data, but also provides an environment for visualizing, exploring and evaluating alternative land-use scenarios. Further, various goals of planning practices, such as minimizing development costs and maximizing ecological benefits as well as constraining conditions with respect to economy, society and the physical environment, can be expressed through optimization models based on linear, integer or mix-integer programming.

With the advances in GIS and computing technologies, numerous spatial optimization approaches have been proposed for land-use planning over the last few decades (Yao et al., 2017). Spatial decision support systems (SDSS) have been developed for assisting interactive processes (Porta et al., 2013; Dai & Ratick, 2014; Santé et al., 2016a). Such methods and tools have been applied in a variety of contexts, ranging from reserve design (Önal et al., 2016) and forest management (Church et al., 2000) to general urban and regional planning (Caparros-Midwood et al., 2015). Various sustainable land use concerns have been considered for different applications, including compactness of selected regions, contiguity of equal land use, compatibility of different land uses, and environmental and ecological impacts, among others (Aerts et al., 2003; Ligmann - Zielinska et al., 2008; Stewart & Janssen, 2014; Önal et al., 2016).

Most approaches and applications, however, have relied upon raster data structured using regular grid cells, largely due to its simplicity and ease of assessing spatial relationships among land parcels, such as proximity and adjacency. With the exception of Chandramouli et al., (2009), Cao & Ye (2013), Masoomi et al. (2013) and Stewart & Janssen (2014), little has been done utilizing vector data in land-use modeling processes, yet this reflects actual decision-making units. That means land use information is generally vector based, and managed using this data. Structuring and solving land use optimization problems using vector data, however, can be more challenging. For instance, evaluation of the spatial relationships (e.g. adjacency) between parcels is often required for calculating measures of land-use patterns (e.g. compactness and contiguity), but this requires geometric assessment and derivation. In a raster structure, through row-column referencing, this is trivial as it is essentially part of the data structure. Similarly, attribute assessment in a raster representation, such as area, length of common boundary or perimeter of a parcel cluster, is generally a simple summation of cells. For vector data, however, this is complicated by spatial query and topological relationship evaluation. While the use of vector data in land use optimization can better reflect planning practice, it also requires more computational processing and evaluation involving polygon object geometries.

Heuristics, such as greedy approaches, genetic algorithms (GA) (or more generally evolutionary algorithms, EA), simulated annealing and particle swarm algorithms, are often adopted in land use optimization (Yao et al., 2017). One reason for this is that spatial optimization usually involves spatial query and evaluation of spatial relationships that adds to computational complexity (Murray, 2010). Exact solution approaches are often limited in practice (Porta et al. 2013; Santé et al., 2016b). Further, land-use planning generally involves multiple, often conflicting, objectives, so one best solution likely does not exist. This means compromise outcomes are a reality, and Pareto-optimal solutions are necessary for multi-objective land use optimization problems. Many heuristics have been specifically designed for raster data structures, and are not necessarily amenable to vector data. Therefore, modified or new heuristics are needed for implementing land use optimization using vector data given the challenges mentioned above.

The aim of this research is to develop an EA-based heuristic for generalized land use optimization models that account for sustainability concerns, applicable for vector land-use data, and identify efficient solutions for multi-objective problems. The next section reviews related work on sustainable land use optimization, focusing on approaches that account for patterns. Then, a generalized model and solution method are presented. These approaches are then applied in an empirical study of Dafeng, China, where a comparison with results based on raster data is detailed. The paper concludes with a discussion of the properties and wider applicability of the proposed modelling approach, the superiority of land use optimization using vector data over raster in practice, and areas for future research.

2 Related Research

Urban land-use planning is important for government to reconcile diverse and often competing interests, regulate activities and promote development. This usually involves assessing land potential and allocating various socioeconomic activities to land parcels (Berke & Godschalk, 2006). Given the challenges brought by worldwide urbanization, such as traffic congestion, air pollution, loss of farm land, food security and urban poverty, sustainable development has become the goal of urban land-use planning. A variety of strategies have been proposed to account for sustainability in practice, such as mixed land uses, compact communities, infill development, and decentralization (Leccese & McCormick, 2000). Godschalk (2004) developed a sustainability/livability prism to cope with conflicts in land-use planning. Berke & Godschalk (2006) discussed how to incorporate economy, environment, and equity (also known as three E's of sustainability) into urban land-use planning. United Nations (2015) explicitly specified 17 global sustainable development goals to be achieved by 2030, including poverty, social equity, economy, and sustainable cities. In this regard, land-use planning can be considered as a way to effectively utilize land resources to achieve social, economic and environmental objectives in a sustainable manner.

In the field of spatial optimization, the primary concerns in sustainable land-use planning are often the form and spatial arrangement of land parcels, as well as the spatial relationships between them. Thus, in addition to socioeconomic and environmental dimensions, spatial optimization approaches explicitly consider geographic configuration of land uses, often using compactness, contiguity and compatibility.

Compactness is related to the configuration of a land use type, where land parcels that cluster to form a circular shape are desired. Compact land uses can be more energy efficient and compact urban form can promote social equality in access to public services (Watson, 2016). Many shape indices have been proposed and utilized as the indicators of compactness in land-use planning, such as these of perimeter, area and perimeter-to-area ratio (Janssen et al., 2008; Porta et al. 2013). Other methods encourage parcels to be assigned the same land use as their neighbors, thereby forming spatial clusters. Common strategies include maximizing the number of adjacent parcels of the same land use, maximizing the largest cluster and minimizing the number of total clusters for each land use (Aerts et al., 2003; Stewart et al., 2004; Stewart & Janssen 2014).

Contiguity refers to the connectiveness of land parcels. A land unit is considered contiguous if one can move from one point to another point without leaving the same land use. Compact and contiguous land is considered more sustainable for habitats (Önal et al., 2016). Many scholars have formulated contiguity through network flow approaches that abstract parcels as nodes on a network (Shirabe, 2009; Duque et al., 2011). Also, it has been found that encouraging compactness will implicitly promote contiguity, thus the former is usually employed as a surrogate for the latter (Aerts et al., 2003).

Compatibility reflects the coexistence of different land uses in an area without negative effects on each other. For example, residential and certain public infrastructure land uses are considered highly compatible as facilities like schools and parks can serve nearby neighborhoods. Compatibility is usually obtained through Delphi or analytic hierarchy process (AHP) techniques that are subsequently optimized (Ligmann-Zielinska et al., 2008; Cao et al., 2011; Masoomi et al., 2013).

Given the inherent complexity of identifying sustainable land use patterns as discussed above, a number of heuristics have been developed to solve land use optimization problems, such as GA (Chandramouli et al., 2009; Cao et al., 2011, 2012; Cao & Ye, 2013; Schwaab et al., 2017), simulated annealing (Caparros-Midwood et al., 2015; Santé et al., 2016b), particle swarm (Masoomi et al., 2013) and ant colony algorithms (Mi et al., 2015). For example, Aerts et al. (2005) applied both GA and simulated annealing to solve a goal-programming model for land-use allocation and found that the former had better performance in terms of both computational efficiency and quality of solutions. Porta et al. (2013) and Santé et al. (2016b) sought to improve the efficiency of GA and simulated annealing using parallel computing, respectively.

Among various heuristics, GA has proven effective and efficient for land use planning (Stewart et al., 2004; Aerts et al., 2005; Janssen et al., 2008; Chandramouli et al., 2009; Cao et al., 2011, 2012; Cao & Ye, 2013; Stewart & Janssen, 2014; Li & Parrott, 2016; Schwaab et al., 2017). It is a type of search method built on the theories of natural selection and genetics, seeking solutions of high quality ("fitness") through an evolutionary process where child solutions are generated from parent solutions through an iterative process including sequential operations – selection, crossover and mutation (Sastry et al., 2014). The applications of GA in land use optimization thus have focused on how to design specialized operators for spatial data and sustainable land use patterns. For example, in terms of crossover, Stewart et al. (2004) assigned two land uses to each 50% of the chosen cells within the two parent solutions. Porta et al. (2013) employed a two-point crossover operator that exchanges parent land uses. Regarding mutation, Stewart & Jansen (2014) swapped the land uses of two random subsets of proximate units from two parents. Schwaab et al. (2017) compared several crossover and mutation operators for land-use allocation problems and found that combining diverse mutation operators is helpful to identify representative Pareto-optimal solutions.

Most applications using GA for solving land use optimization problems combine several objectives into one. Common strategies include weighted-sum (Demetriou et al., 2013; Porta et al., 2013) and goal programming (or reference point) (Stewart et al., 2004; Janssen et al., 2008; Chandramouli et al., 2009; Cao et al., 2012; Cao & Ye, 2013; Stewart & Janssen, 2014) approaches. However, the former cannot find all Pareto-optimal solutions in a non-convex solution space. The later has the potential for finding the full Pareto frontier, if properly designed and implemented, but there is no guarantee. In practice a group of diverse Pareto-optimal solutions are often preferred, and there has been increasing interest in EA like non-dominated sorting GA (NSGA) and elitist NSGA (NSGA-II) designed for multi-objective decision-making (Deb, 2014). They have been applied in various land use planning contexts (Cao et al., 2011; Schwaab et al., 2017). Further, previous work utilizing GA for vector-based land use optimization have either employed non-spatial operators (Chandramouli et al., 2009; Cao & Ye, 2013) or adopted certain model simplifications so that procedures designed for raster data could be used (Stewart & Janssen, 2014). Therefore, this research attempts to develop an EA for land use optimization by extending NSGA-II (Deb et al., 2002) to explicitly account for spatial characteristics of vector data, incorporate sustainable land use patterns and explore the impacts of various parameter settings.

3 Spatial Optimization Model

Without loss of generality, a spatial optimization model for sustainable land-use planning would include two types of objectives, spatial and non-spatial, subject to basic limits for each land use. Consider the following notation:

i = index of land parcels;

k, k' = index of land use types;

N, N' = total number of spatial and non-spatial objectives;

U_k, L_k= upper and lower bounds on total area for land use type k;

a_i = area of parcel i;

$$x_{ik} = \begin{cases} 1 & \text{if parcel } i \text{ is assigned land-use type } k \\ 0 & \text{otherwise} \end{cases}$$

A generalized model can be formulated as follows:

$$Minimize \qquad F = \{f_1, f_2, \dots, f_N\} \tag{1}$$

$$Minimize \qquad F' = \{f'_1, f'_2, \dots, f'_{N'}\} \tag{2}$$

$$Subject\ to \qquad \textstyle\sum_k x_{ik} = 1 \qquad \forall i \tag{3}$$

$$\textstyle\sum_i \sum_k a_i x_{ik} \leq U_k \qquad \forall k \tag{4a}$$

$$\sum_i \sum_k a_i x_{ik} \geq L_k \qquad \forall k \qquad (4b)$$

$$x_{ik} = \{0,1\} \qquad \forall i, k \qquad (5)$$

Objectives (1) and (2) include a set of functions to achieve spatial (e.g. compactness and contiguity) and non-spatial goals (e.g. various costs, economic benefits and ecological impacts), respectively. Thus, a typical spatial land use optimization problem includes at least one objective in each set F and F'. Although minimization is adopted in (1) and (2), multiplication by -1 for any function f or f' indicates maximization. Constraints (3) require only one type of land use allocated to each parcel. Constraints (4a) and (4b) set the range of the desired acreage for each land use. Constraints (5) indicate that the values of the decision variables are binary (0 or 1).

Figure 1 details the terminology for the heuristic solution approach employed based on GA. The GA usually start with a generation of population (a set of solutions or land-use plans), representing the parents. A child in the offspring generation is obtained from two parents by means of selection, crossover and mutation operations. The procedure usually stops when a certain number of generations are obtained, or other evaluation criteria are met.

<Figure 1 about here>

The proposed solution procedure for this model, (1) - (5), is based on the NSGA-II of Deb et al. (2002) and is presented in Figure 2. Associated parameters are summarized in Table 1. The two distinct strategies adopted by NSGA-II are non-dominated sorting (Step 2b) and crowding sorting (Step 3), where the former encourages convergence to the Pareto frontier and the latter promotes diversity of solutions. The solution procedure starts with an initial population that includes N_P individuals (solutions), from which non-dominated solutions are copied into A – the archive set containing all the non-dominated solutions throughout the entire solution process. For the first iteration, a random set of N_P individuals from the combination of P and A are selected to create the next generation Q through crossover (Step 5) and mutation (Step 6). The feasibility of the individuals in Q is checked and modified to be feasible if necessary. Then, A is updated by comparing the individuals in Q and existing solutions in A with respect to their non-domination levels and crowding distances. For the second and subsequent iterations, P is updated by non-dominated sorting (Step 2b) and crowding sorting (Step 3) before being used to create the offspring generation through selection, crossover, mutation and feasibility amendment (Steps 4-7). The procedure terminates when it is run N_iter times or the same archive set A is obtained in N_sameA consecutive iterations. Unique to the implementation reported here, Steps 1, 5, 6, and 7 are explained as follows:

<Figure 2 about here>

<Table 1 about here>

Step 1 Population initialization:

1) For the first individual, a set (G) of $\lfloor N_{par} * r_0 \rfloor$ (the largest integer smaller than $N_{par}*r_0$) parcels are randomly selected (including the parcels with fixed land-use), for which the current land uses are to be retained. Thus, each member in the complementary set $\bar{G}$ is to be assigned a land use.
2) For each member g in G, use it as a seed to build a region by including the neighboring unallocated parcels and assign the same land use g_k to them until the desired area range $[L_{g_k}, U_{g_k}]$ is achieved. The neighbors are defined by

adjacency so that the first-order neighbors of g (directly adjacent to g) are check first, then second-order neighbors (adjacent to the first-order neighbors), etc.

3) Repeat 2) until all the land uses involved in G achieve their allowance.
4) If every parcel in $\bar{G}$ is allocated a land-use type, go to 8). Otherwise, continue with 5).
5) If there are unassigned land uses, randomly select an unallocated parcel as a seed and give it an unassigned land use. Again, similar to 2), use that seed to build a region by assigning the same land use to the adjacent/proximate unallocated parcels until the desired allowance is achieved.
6) If there are still unallocated parcels, for each of them, randomly assign a land use of their neighboring parcels.
7) Repeat 4)-6) until every parcel in $\bar{G}$ is assigned a land use.
8) Repeat 1)-7) $(N_p - 1)$ times to create the rest of the individuals.

Step 5 Crossover: randomly select a parcel and use it as a seed to select a block of $\left|N_{par} * r_c\right|$ adjacent or proximate parcels. For the two parents selected for crossover, the land uses of the parcels within that same block are switched and the others remain unchanged.

Step 6 Mutation: similar to Step 5, randomly select a block of $\left|N_{par} * r_m\right|$ adjacent or proximate parcels. Randomly exchange the land uses among them. In addition, in order to promote the diversity of the solutions, we use p_c instead of p_m to increase the probability of mutation for the individuals which have multiple copies in the population.

Step 7 Feasibility amendment: the purpose of this step is to modify infeasible solutions in Q to promote their feasibility according to constraints (4a) and (4b) through following procedure:

1) For every infeasible individual in Q, all land-use types are grouped into two categories: $K = \{k_1, k_2, \ldots, k_i\}$ and $K' = \{k'_1, k'_2, \ldots, k'_j\}$, so that the land uses in K and K' have a shortage and an excess of area, respectively.
2) Randomly choose a land-use type k_i from K and a parcel with type k_i.
3) Check the neighbors of that parcel (again, first check the first-order neighbors and then the second-order neighbors, etc). If a neighboring parcel has type k'_j belonging to K', change the land use of that neighbor to k_i. Update the overall area of k_i and k'_j accordingly.
4) Repeat 3) until the area of k_i meets the allowance. Remove k_i from K.
5) Repeat 2)-4) until K is empty.

The fixed-use parcels are kept unchanged during Steps 5-7. In addition, feasibility is also considered in Step 2b when determining the non-domination level for each solution. That is, infeasible solutions are always dominated by feasible solutions. Two infeasible solutions, s_1 and s_2, are compared using the violation values defined as in (6):

$$V_s = \sum_k \frac{diff_{s,k}}{U_k - L_k} \quad \text{with} \quad diff_{s,k} = \begin{cases} L_k - Area_{s,k} & if\ L_k > Area_{s,k} \\ Area_{s,k} - U_k & if\ Area_{s,k} > U_k \end{cases} \tag{6}$$

where $Area_{s,k}$ is the total area of land use k for solution s. Therefore, s_1 is dominated by s_2 if $V_1 > V_2$.

Steps 2b, 3, 4 and 8 are rather standard for NSGA-II implementation. Details regarding such steps can be found in Deb et al. (2002), among many others.

4 Empirical Study

4.1 Study area and the planning context

The study area is along the urban fringe of Dafeng, Jiangsu Province, China, covering part of Chuandong Farm and Caodianmiao Town (see Figure 3). The area consists of 680 parcels with a total area 6.52 km^2, and are grouped into six land-use categories according to national regulations: arable land, green land, construction land, water, transportation and other (undeveloped) land. During the last three decades, like many Chinese cities, Dafeng has experienced extensive changes in the physical forms and functions of land use caused by rapid urbanization, and there has been increasing demand in land for urban construction and transportation due to population and economic growth. The primary planning goal is to increase urban construction and transportation land by 3-12% and 4-6%, respectively, mainly through transforming undeveloped land which can be reduced by 50-90%. The area of all the other land uses can vary by ±10%. In total 160 parcels were selected for retaining current land uses, most of which are roads, rivers and key constructions. Sustainable urban development goals require certain patterns of spatial layout for different land uses, such as compactness and compatibility. In this context, spatial optimization approaches are developed and applied to assist decision-making for local urban land-use planning.

<Figure 3 about here>

4.2 Model settings

To support application and analysis, the following additional notation is used:

k_i = current land-use type of parcel i;

$c_{kk'}$ = cost/km^2 for conversion between land-use types k and k';

$C_{kk'}$ = compatibility of land-use types k and k';

Ω_i= {j; j is a neighbor of i in space, that is, i and j are adjacent}

Three objectives – two spatial and one non-spatial - are considered for the empirical study, which can be defined as follows:

$$\text{Compactness/Contiguity} \quad \textit{Maximize} \quad \sum_k \sum_i \sum_{j \in \Omega_i} x_{ik} x_{jk} \tag{7}$$

$$\text{Compatibility} \quad \textit{Maximize} \quad \sum_k \sum_{k'} \sum_i \sum_{j \in \Omega_i} C_{kk'} x_{ik} x_{jk'} \tag{8}$$

$$\text{Cost} \quad \textit{Minimize} \quad \sum_i \sum_k c_{k_i k} a_i x_{ik} \tag{9}$$

Considering the generalized model in section 3, the first two objectives, (7) and (8), belong to the spatial objective set F in (1), and Objective (9) belongs to the non-spatial objective set F' in (2). Objective (7) is to maximize the compactness by encouraging the same land use to be assigned to adjacent parcels, which would implicitly promote contiguity of the same land use (see Aerts et al., 2003). Objective (8) aims to maximize the overall compatibility of different land uses. Objective (9) attempts to minimize total costs. Two models are solved: Scenario I includes objectives (7) and (9), and Scenario II includes all the three objectives (7)-(9), both subject to constraints (3)-(5).

The compatibility indicators are shown in Table 2. Four levels of compatibility were determined using the AHP method by consulting practitioners in the local land-use planning bureau, with higher values indicating better compatibility. The parameter values for the solution procedure are presented in Table 1, obtained by trial and error. It is well known that GA involves many parameters, among which are the probabilities of crossover and mutation. Deb (2014) indicated that a rule of thumb was to start with

p_c 0.6 and p_m 0.05. Since crossover usually occurs with a high probability and mutation happens with a low probability, we systematically tested a set of p_c and p_m values for Scenario I while keeping other parameters fixed. The former was varied from 0.55 to 1.0 and the latter was varied from 0.01 to 0.10. As it becomes difficult to visually compare the results from different tests when the model includes more than two objectives, we present one instance for Scenario II using the same parameters as those in Scenario I.

<Table 2 about here>

The models defined above were solved using the above detailed method, implemented using Python with spatial data/relationships processed and evaluated carried out using PySAL, an open-source Python library for spatial analysis (see Anselin & Rey, 2014). The commercial GIS software ArcGIS (version 10.6, by ESRI, Redlands, CA, USA) was employed for spatial data management, processing and visualization.

4.3 Results

Both models for Scenarios I and II were run on a Mac computer with 16 GB memory and 3.1 GHz Intel Core i7 processor. The tests for a range of parameter values in Scenario I took 20-40 sec and the running time for Scenario II is about 60 sec. The final Pareto-optimal solutions for Scenarios I are summarized in Figure 4, with opposite values shown for Objectives (7) and (8) for the corresponding minimization problems. Figures 4(a) and 4(b) depict the variations in the solution set for different crossover and mutation probabilities, respectively.

<Figure 4 about here>

It can be observed that the solutions are generally well spread in Figures 4(a) and 4(b) although the diversity of solutions vary across different crossover or mutation probabilities. Both graphs in Figure 4 suggest that there is no obvious association between the values of p_c or p_m and the quality of obtained solutions. In other words, an increase of p_c or p_m, when keeping all other parameters constant, does not necessarily guarantee a set of better Pareto-optimal solutions. For example, in Figure 4(b), most solutions are dominated by those with $p_m = 0.07$. Interestingly, Figure 4(a) indicates that the Pareto-optimal solutions derived with the highest p_m value 1.0 dominates most of those obtained with other p_m values.

Further, Figure 5 describes the set of non-dominated solutions across generations for Scenarios I and II, both with $p_c = 0.60$ and $p_m = 0.05$. It can be observed that the quality of non-dominated solutions increases with the progress of generation evolution, but the speed of improvement varies. Figure 5(a) suggests that the convergence of non-dominated solutions for Scenario I became much quicker than before after 80 generations. For Scenario II, the improvement of solutions was relatively slow until the 120th generation, and the speed of convergence greatly increased since the 180th generation.

<Figure 5 about here>

Figure 6 shows the corresponding land-use plans and the parcels involved in the conversion for two solutions in Figure 4(a): one with the best compactness (p_c = 0.55) (Figure 6(a)) and the other with the least cost (p_c = 1.0) (Figure 6(b)). As can be seen, the compactness of arable and construction land in Figure 6 are improved compared to the original land-use layout in Figure 3. However, compared to Figure 6(b), the newly added arable land (on the top right) and construction land (on the top) makes the plan in Figure 6(a) has better compactness. Regarding the land conversion, the increase in construction and transportation land are mainly achieved by transforming arable and undeveloped land. For example, the construction land is expanded mainly through transforming the surrounding arable land. There are also conversions between other types of land uses, such as arable land and water, and arable and green land.

<Figure 6 about here>

Figure 7 presents the land-use plan with the best compatibility objective value in Figure 5(b). Again, the increase of construction and transportation land is primarily achieved by reducing the undeveloped and arable land. Compared to the plans in Figure 6, the land-use layout is very different from the original one in the sense that it involves a lot of changes to water and arable land. For example, lots of water area on the top right is converted to arable land. The variety of land uses on the bottom left makes that region more fragmented than the original plan, where the same region mainly contains arable land and water.

<Figure 7 about here>

Finally, the area of different land uses for the plans in Figures 6 and 7 are summarized in Figure 8. Compared to the original area, arable, construction and transportation land are increased in all the three plans. In contrast, the amount of undeveloped land is decreased as expected. For green land, it is decreased in both plans in Figure 6 by 9.3% and 4.0% for Figures 6(a) and 6(b), respectively, while increased by 8.7% in the plan in Figure 7. Regarding the land for water resources, there is only a slight growth in the plan in Figure 6(a).

<Figure 8 about here>

4.4 Comparison with results using raster data

To provide a comparative assessment, the proposed approach was also applied for the study area represented using raster data. As about 0.9% of the land parcels have an area less than 100 m^2 and 10% with an area less than 400 m^2, a spatial resolution of 10 m * 10 m was selected in order to generate a raster surface representing the current land use layout with sufficient accuracy and detail. In total, the study area consists of 65,211 raster grid cells, with the land use in each cell determined by the type having the largest area within that cell, as shown in Figure 9. Accordingly, the land use quantity, U_k and L_k in (4a) and (4b), was rounded to the nearest integer.

<Figure 9 about here>

Using the same parameters as those adopted by the instances summarized in Figures 6 and 7, the land use plans generated from the raster representation (Figure 9) are presented in Figure 10. Figures 10(a) and 10(b) show two plans for Scenario I, most compact and least cost, respectively. Figure 10(c) depicts the plan with the best compatibility for Scenario II. The three plans in Figure 10 are very different from their vector counterparts in Figures 6 and 7. Obviously, the land use layouts in the former are more fragmented. Also, most of the area in the southeast of the study area is converted to construction and green space, with a mixture of scattered arable land and water. Three plans in Figure 10 are very similar visually, but there are some important differences. For instance, a road in the northwest is converted to arable land in Figure 10(a) but part of it remains in Figures 10(b) and 10(c). The plan for Scenario II (Figure 10(c)) has more green and construction land in the southeast than the other two plans for Scenario I (Figures 10(a) and 10(b)).

<Figure 10 about here>

Figure 11 further describes the allocation of each land use in the three plans in Figure 10, compared to the original plan in Figure 9. Similar to the results in Figure 8, arable, construction and transportation land is increased, largely by reducing the other (undeveloped) land. However, the amount of converted land is different. For example, the three plans generated with raster data all have less arable land (< 3.5 km^2) and more undeveloped land (over 45% retained) compared to the corresponding plans using vector data. In addition, some land has a loss in a vector-generated plan but a gain in the corresponding raster-generated plan. For instance, green space in both plans in Figure 6 is reduced but is increased in Figures 10(a) and 10(b). Compared to the decrease of 4.0% in Figure 6(b), water in the plan represented in Figure 10(b) increase 6.4%.

<Figure 11 about here>

5 Discussion and Conclusions

Spatial optimization has been widely applied to support land-use planning. Most studies, however, have utilized raster data due to the simple data structure and the ease of spatial relationship assessment. This research proposed an evolutionary algorithm for a generalized spatial land use optimization problem using vector data that better reflects how data are stored and organized in cadastral management and land-use planning. The empirical study demonstrated the effectiveness of the proposed approach in finding good quality and diverse land-use plans, with a focus on the sustainability concerns. The comparison with the results from raster data also suggests that vector-based land use optimization is more feasible and more desirable in the practice of land-use planning.

Although the generalized model defined by (1) - (5) can be applied to both raster and vector land-use data, implementation procedures can be very different. One challenge in using vector data for spatial land use optimization is constraints on each land use. It is straightforward for raster data since it consists of same-size grid cells, making the area calculation equivalent to counting the number of total cells of same land use. Stewart & Janssen (2014) adopted this approach for vector data by transforming constraints 4(a) and 4(b) to a parcel quantity constraint (i.e. the number of parcels to be allocated for each land-use type), but this is only applicable for cases where all parcels have similar sizes. Further, for raster data, Steps 5 and 6 might bring no area changes to each land use if the number of parcels of each type remains the same even though their spatial layout might be altered. However, in the case of vector representation, Steps 5 and 6 might lead to very different area allocations for each land use due to the diverse size of parcels, as shown in the empirical study in this research. Thus, it is more complex to assess and meet the land-use quantity constraints if using vector data, so the feasibility amendment in Step 7 is crucial.

In addition to simplification in implementation, several disadvantages of using raster data in land use optimization were observed when comparing results using vector data. First, compared to the original dataset, the amount of raster cells is almost 96 times that of land parcels, which can greatly increase the demand for data storage and computing. Also, a trade-off between the dataset size and the details of represented land use layout is often necessary when selecting the spatial resolution for raster data. For example, the proportion of transportation land will decrease from 1.9% to 0.6% if the cell size is increased from 10 m * 10 m to 30 m * 30 m, which will inevitably introduce errors or uncertainty in the obtained land use plans. Again, as partitioning a land parcel into a set of raster cells provides more flexibility of land use conversion, the obtained land use plans tend to be more fragmented, as shown in Figure 10, which is often undesirable in practice. Therefore, land use optimization using vector data can provide higher quality land use plans that better reflect the reality and needs of land-use planning.

The proposed approach has wide applicability as the major steps in Figure 2 are not problem specific or subject to any particular objectives and constraints. The only requirement for crossover and mutation operations is to select a contiguous block of parcels, which attempts to avoid fragmented land-use layout. For substantive applications, the generalized model in (1)-(5) can be modified by incorporating other objectives and constraints. Accordingly, the violation value in (6) needs to be adjusted by considering other constraints.

Evolutionary algorithms like GA do have some limitations. One shortcoming is that they usually involve many parameters (e.g. population size, crossover and mutation probability), where numerical tests are often required to find a good set of parameter values that can generate solutions of desired quality for specific problems. In this research, a set of common options for crossover and mutation probabilities were examined, indicating that increasing those probabilities did not necessarily result in an improvement or degradation of the solution quality (see Figure 4). Another limitation is that evolutionary multi-objective optimization approaches like NSGA-II are often difficult to apply to large-dimensional problems containing more than three objectives. This is mainly due to high computational costs required to preserve solution diversity, difficulty in visualization of large-dimensional solutions, and limited search capacity for new solutions because a solution would more easily become non-dominated (Deb, 2014). It also should be noted that multi-objective land use optimization is only part of the overall decision-making process, requiring as well other quantitative/qualitative considerations to select the final preferred plan.

There are several areas worth further investigation. First, the impact of the parameter values other than crossover and mutation probabilities, such as the population size and the proportion of parcels retaining original land uses, on the quality of solutions needs further exploration. Second, the crossover and mutation operators can be adapted to account for particular spatial objectives/constraints. For example, in Step 5 mutation, instead of random exchange, the land uses within the selected block can be reallocated based on their compatibility with adjacent land uses, or the dominant land use can be assigned to the whole block to promote compactness/contiguity. Third, when amending infeasible solutions, rather than choosing a single parcel, multiple parcels with a land use having insufficient area can be selected and then their neighbors are checked and modified in a region-growing manner. Forth, the measure of compactness/contiguity adopted in Objective (7) was originally designed for raster data, which might not work in some cases when using vector-based data as two adjacent parcels may only share a small proportion of boundary or may be very different in size. Therefore, other compactness and contiguity measures may be useful. Finally, land parcels might be divided into two or more smaller parcels in practice, which is a challenge for vector-based land use optimization.

Land-use planning has long been an active application area of spatial optimization. Given the ongoing global urbanization and consequent transformations in urban spatial and social structures, sustainable urban land-use plans are essential to achieve long-term balanced urban development. Spatial land use optimization explicitly considering sustainability concerns can be a valuable tool to generate a variety of land-use plans to be further evaluated by decision-makers with other ancillary information.

Acknowledgements

We are grateful for the financial support from the National Natural Science Foundation of China (Grants 41330750, 41201117 and 41201394).

References

Aerts, J. C., Eisinger, E., Heuvelink, G., & Stewart, T. J. (2003). Using linear integer programming for multi-site land-use allocation. *Geographical Analysis*, 35(2), 148-169.

Aerts, J., Van Herwijnen, M., Janssen, R., & Stewart, T. (2005). Evaluating spatial design techniques for solving land-use allocation problems. *Journal of Environmental Planning and Management*, 48(1), 121-142.

Anselin, L., & Rey, S. J. (2014). *Modern spatial econometrics in practice: A guide to GeoDa, GeoDaSpace and PySAL*. GeoDa Press LLC.

Cao, K., Batty, M., Huang, B., Liu, Y., Yu, L., & Chen, J. (2011). Spatial multi-objective land use optimization: extensions to the non-dominated sorting genetic algorithm-II. *International Journal of Geographical Information Science*, 25(12), 1949-1969.

Cao, K., Huang, B., Wang, S., & Lin, H. (2012). Sustainable land use optimization using Boundary-based Fast Genetic Algorithm. *Computers, Environment and Urban Systems*, 36(3), 257-269.

Cao, K., & Ye, X. (2013). Coarse-grained parallel genetic algorithm applied to a vector based land use allocation optimization problem: the case study of Tongzhou Newtown, Beijing, China. *Stochastic Environmental Research and Risk Assessment*, 27(5), 1133-1142.

Caparros-Midwood, D., Barr, S., & Dawson, R. (2015). Optimised spatial planning to meet long term urban sustainability objectives. *Computers, Environment and Urban Systems*, 54, 154-164.

Chandramouli, M., Huang, B., & Xue, L. (2009). Spatial change optimization: Integrating GA with visualization for 3D scenario generation. *Photogrammetric Engineering & Remote Sensing*, 75(8), 1015-1022.

Church, R. L. 2002. Geographical information systems and location science. *Computers & Operations Research*, 29: 541-562.

Church, R. L., Murray, A. T., Figueroa, M. A., & Barber, K. H. (2000). Support system development for forest ecosystem management. *European Journal of Operational Research*, 121(2), 247-258.

Dai, W., & Ratick, S. J. (2014). Integrating a Raster Geographical Information System with Multi-Objective Land Allocation Optimization for Conservation Reserve Design. *Transactions in GIS*, 18(6), 936-949.

Deb, K. (2014). Multi-objective optimization. In E.K.Burke and G. Kendall (eds.), *Search methodologies: Introductory Tutorials in Optimization and Decision Support Techniques* (pp. 403-449). Springer, Boston, MA.

Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. A. M. T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182-197.

Demetriou, D., See, L., & Stillwell, J. (2013). A spatial genetic algorithm for automating land partitioning. *International Journal of Geographical Information Science*, 27(12), 2391-2409.

Duque, J. C., Church, R. L., & Middleton, R. S. (2011). The p-Regions Problem. *Geographical Analysis*, 43(1), 104-126.

Elmqvist, T., Fragkias, M., Goodness, J., Güneralp, B., Marcotullio, P. J., McDonald, R. I., ... & Wilkinson, C. (Eds.). (2013). *Urbanization, biodiversity and ecosystem services: challenges and opportunities: a global assessment*. Springer.

Godschalk, D. R. (2004). Land use planning challenges: Coping with conflicts in visions of sustainable development and livable communities. *Journal of the American Planning Association*, *70*(1), 5-13.

Healey, P. (2006). *Urban complexity and spatial strategies: Towards a relational planning for our times*. Routledge.

Janssen, R., van Herwijnen, M., Stewart, T. J., & Aerts, J. C. (2008). Multiobjective decision support for land-use planning. *Environment and Planning B: Planning and Design*, 35(4), 740-756.

Berke, P. R., & Godschalk, D. R. (2006). Urban Land Use Planning (5th Ed). Urbana: University of Illinois Press.

Leccese, M., & McCormick, K. (2000). *Charter of the New Urbanism*. McGraw-Hill Professional.

Li, X., & Parrott, L. (2016). An improved Genetic Algorithm for spatial optimization of multi-objective and multi-site land use allocation. *Computers, Environment and Urban Systems*, 59, 184-194.

Ligmann-Zielinska, A., Church, R. L., & Jankowski, P. (2008). Spatial optimization as a generative technique for sustainable multiobjective land-use allocation. *International Journal of Geographical Information Science*, 22(6), 601-622.

Masoomi, Z., Mesgari, M. S., & Hamrah, M. (2013). Allocation of urban land uses by Multi-Objective Particle Swarm Optimization algorithm. *International Journal of Geographical Information Science*, 27(3), 542-566.

Mi, N., Hou, J., Mi, W., & Song, N. (2015). Optimal spatial land-use allocation for limited development ecological zones based on the geographic information system and a genetic ant colony algorithm. *International Journal of Geographical Information Science*, 29(12), 2174-2193.

Murray, A. T. (2010). Advances in location modeling: GIS linkages and contributions. *Journal of Geographical Systems*, 12(3), 335-354.

Murray, A. T. (2017). Regional analytics. *The Annals of Regional Science*, 59(1), 1-13.

Önal, H., Wang, Y., Dissanayake, S. T., & Westervelt, J. D. (2016). Optimal design of compact and functionally contiguous conservation management areas. *European Journal of Operational Research*, 251(3), 957-968.

Porta, J., Parapar, J., Doallo, R., Rivera, F. F., Santé, I., & Crecente, R. (2013). High performance genetic algorithm for land use planning. *Computers, Environment and Urban Systems*, 37, 45-58.

Santé, I., Pacurucu, N., Boullón, M., García, A. M., & Miranda, D. (2016a). An open source GIS-based Planning Support System: Application to the land use plan of La Troncal, Ecuador. *Transactions in GIS*, 20(6), 976-990.

Santé, I., Rivera, F. F., Crecente, R., Boullón, M., Suárez, M., Porta, J., ... & Doallo, R. (2016b). A simulated annealing algorithm for zoning in planning using parallel computing. *Computers, Environment and Urban Systems*, 59, 95-106.

Sastry, K., Goldberg, D. E., & Kendall, G. (2014). Genetic algorithms. In E.K.Burke and G. Kendall (eds.), *Search methodologies: Introductory Tutorials in Optimization and Decision Support Techniques* (pp. 93-117). Springer, Boston, MA.

Schwaab, J., Deb, K., Goodman, E., Lautenbach, S., van Strien, M. J., & Grêt-Regamey, A. (2017). Improving the performance of genetic algorithms for land-use allocation problems. *International Journal of Geographical Information Science*, 1-24.

Seto, K. C., Fragkias, M., Güneralp, B., & Reilly, M. K. (2011). A meta-analysis of global urban land expansion. *PloS One*, 6(8), e23777.

Seto, K. C., Güneralp, B., & Hutyra, L. R. (2012). Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences*, 109(40), 16083-16088.

Shirabe, T. (2009). Districting modeling with exact contiguity constraints. *Environment and Planning B: Planning and Design*, 36(6), 1053-1066.

Stewart, T. J., & Janssen, R. (2014). A multiobjective GIS-based land use planning algorithm. *Computers, Environment and Urban Systems*, 46, 25-34.

Stewart, T. J., Janssen, R., & van Herwijnen, M. (2004). A genetic algorithm approach to multiobjective land use planning. *Computers & Operations Research*, 31(14), 2293-2313.

United Nations. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. https://sustainabledevelopment.un.org/post2015/transformingourworld (Accessed 31 October 2018)

Watson, V. (2016). Locating planning in the New Urban Agenda of the urban sustainable development goal. *Planning Theory*, 15(4), 435-448.

Yao, J., Zhang, X., & Murray, A. T. (2017). Spatial Optimization for Land-use Allocation: Accounting for Sustainability Concerns. *International Regional Science Review*, 0160017617728551.

Figure 1 Meanings of EA terms in the context of optimization models and land-use planning

Figure 2 A solving procedure based on the NSGA-II

Figure 3 Study area and spatial layout of current land uses

Figure 4 Pareto-optimal solutions for Scenario I: (a) $p_c \in [0.55, 1.0]$; (b) $p_m \in [0.01, 0.10]$

Figure 5 Non-dominated solutions across generations with $p_c = 0.60$ and $p_m = 0.05$: (a) Scenario I; (b) Scenario II (Gen 10 = the 10^{th} generation, Pop = population)

Figure 6 Two land-use plans corresponding to two non-dominant solutions in Figure 4(a): (a) the solution with best compactness (p_c=0.55); (b) the solution with least cost (p_c=1.0)

Figure 7 The land-use plan for the Pareto-optimal solution with the best compatibility in Figure 5(b)

Figure 8 Area of different land uses for the plans in Figures 6 and 7

Figure 9 Raster representation of the study area

Figure 10 Land use plans generated by using raster representation: (a) the solution in Scenario I with best compactness; (b) the solution in Scenario I with least cost; (c) the solution in Scenario II with best compatibility

Figure 11 Area of different land uses for the plans in Figure 10

Table 1 Parameters for the solving procedure

Table 2 Compatibility between different land uses

Figure 1 Meanings of EA terms in the context of optimization models and land-use planning

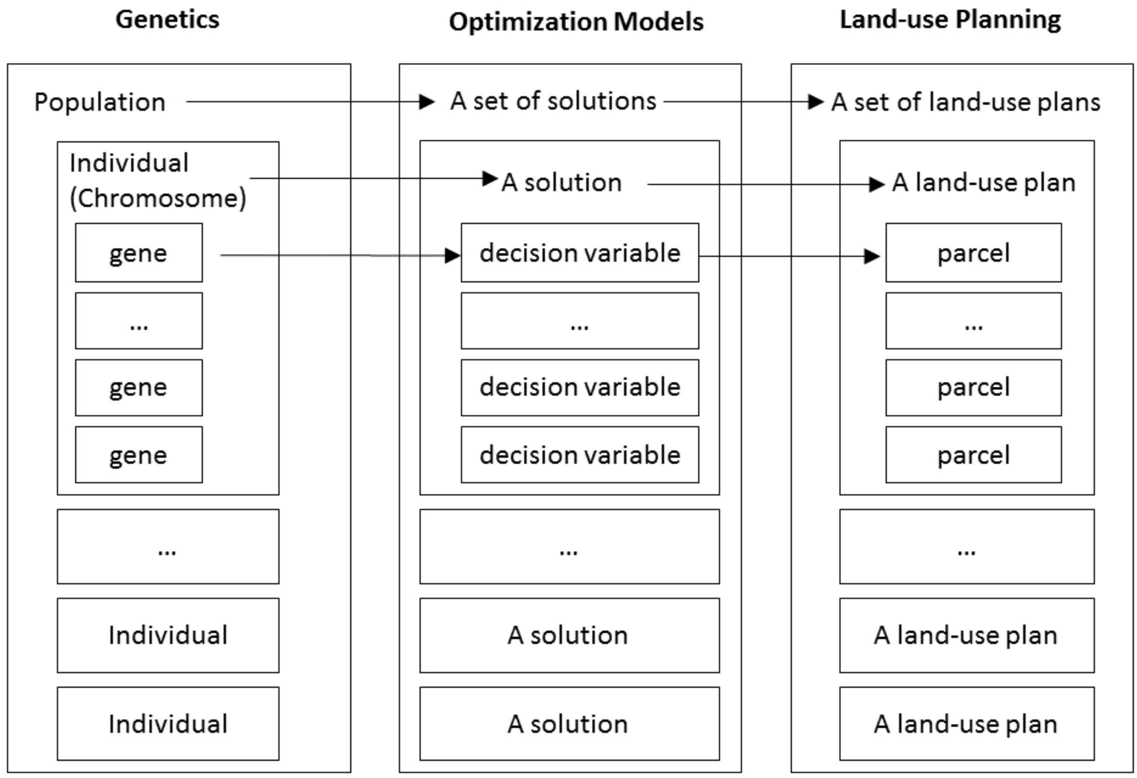


Figure 2 A solving procedure based on the NSGA-II

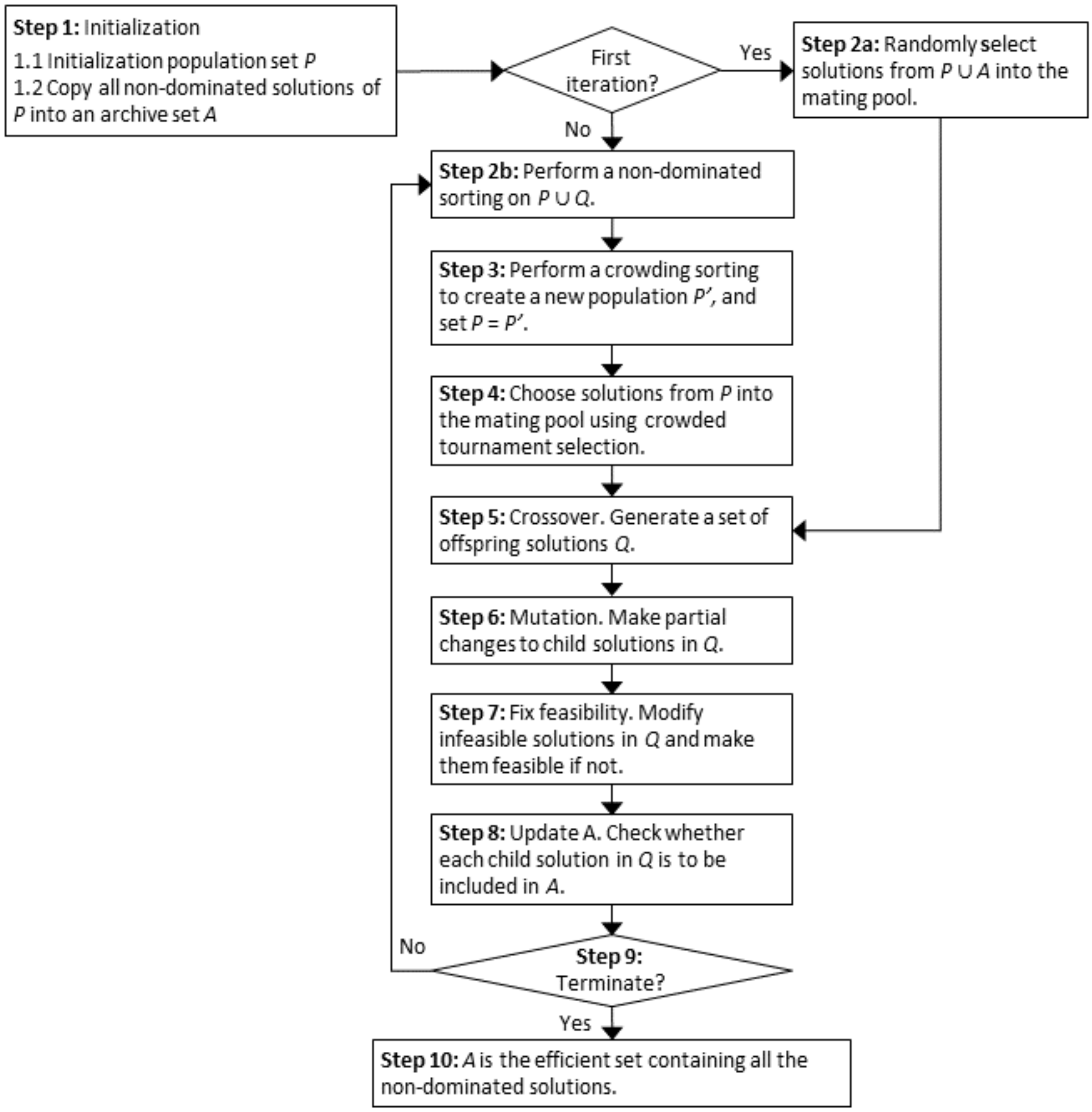


Figure 3 Study area and spatial layout of current land uses

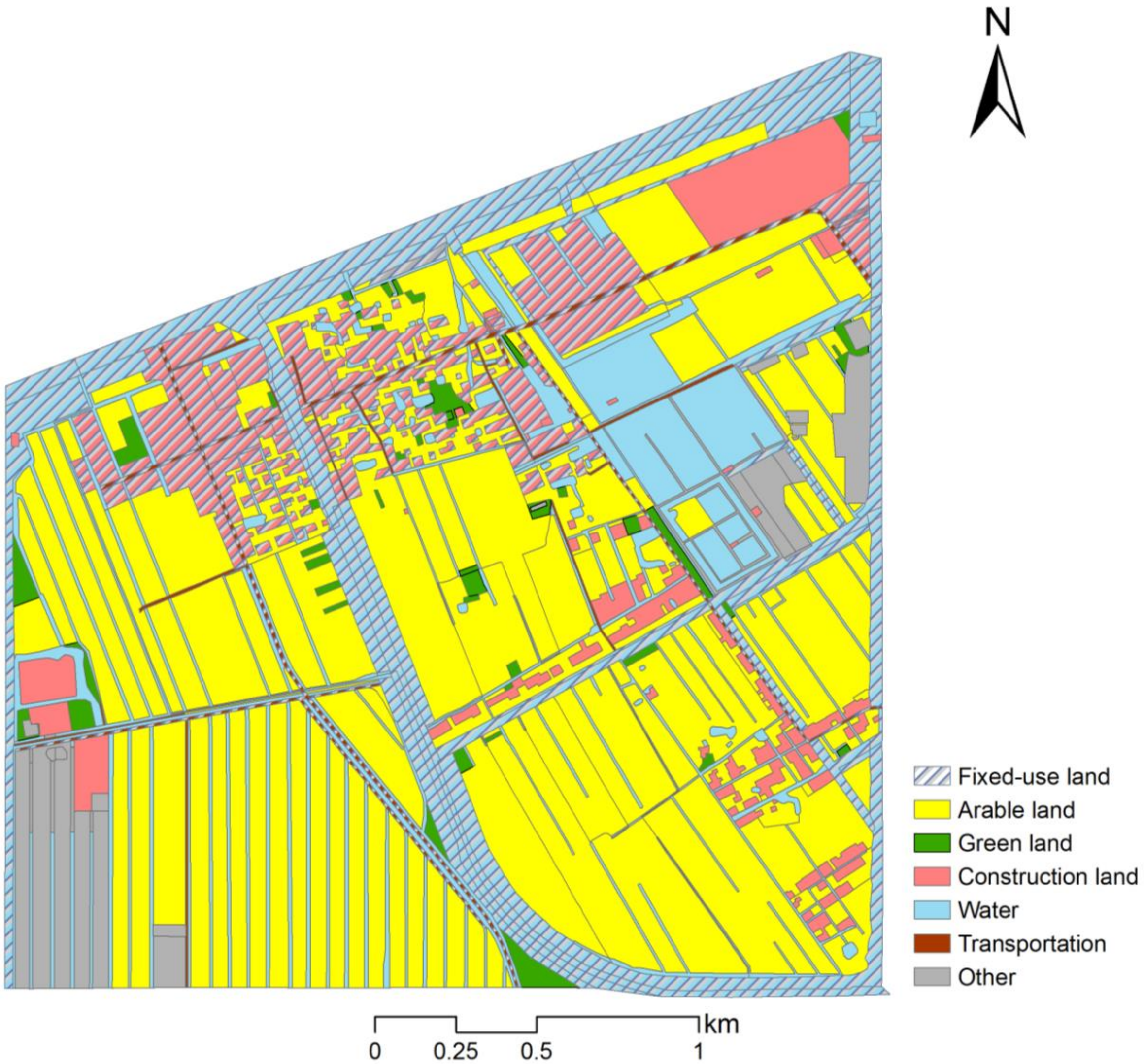
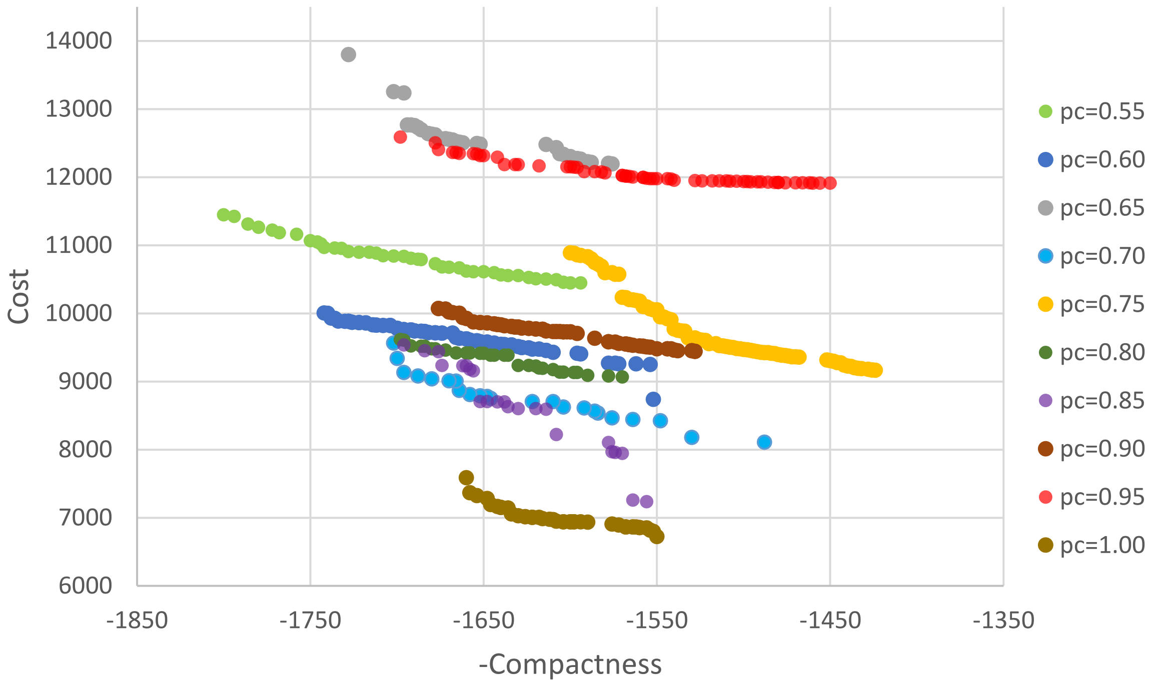


Figure 4 Pareto-optimal solutions for Scenario I: (a) $p_c \in [0.55, 1.0]$; (b) $p_m \in [0.01, 0.10]$

(a)



(b)

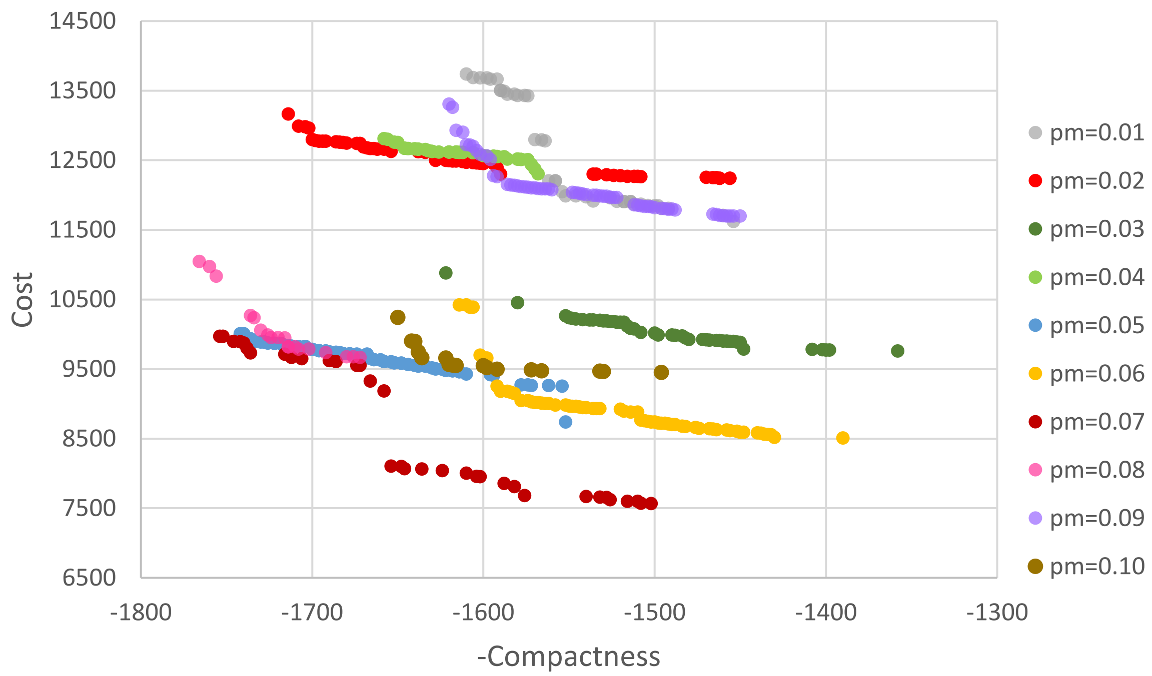
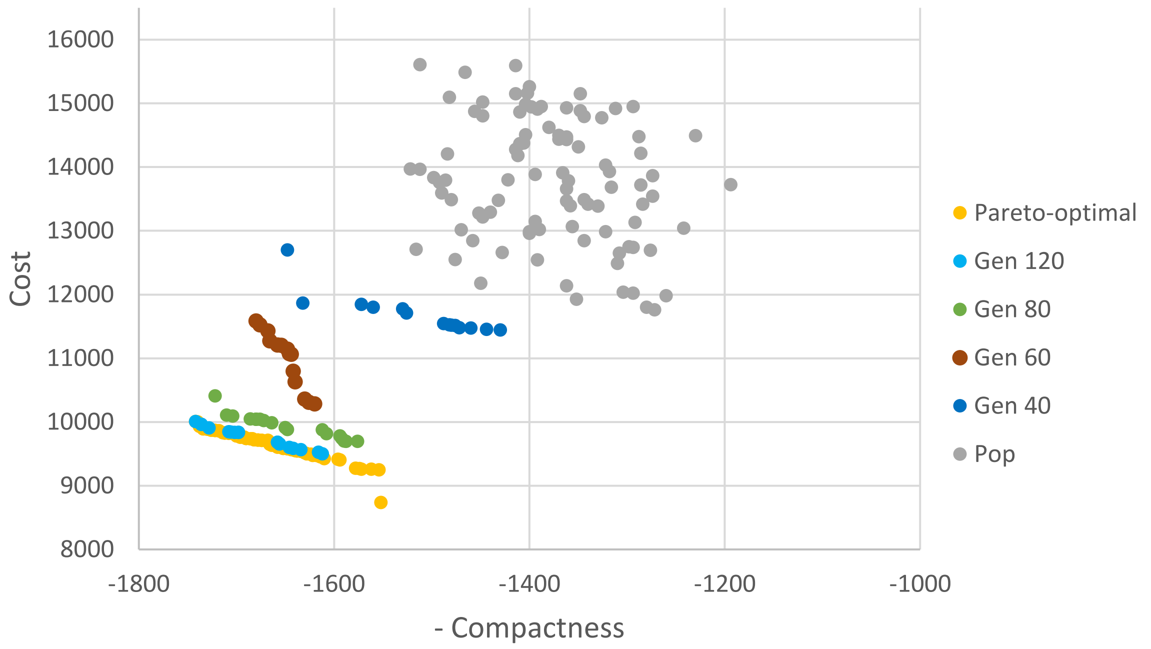


Figure 5 Non-dominated solutions across generations with $p_c = 0.60$ and $p_m = 0.05$: (a) Scenario I; (b) Scenario II (Gen 10 = the 10th generation, Pop = population)

(a)



(b)

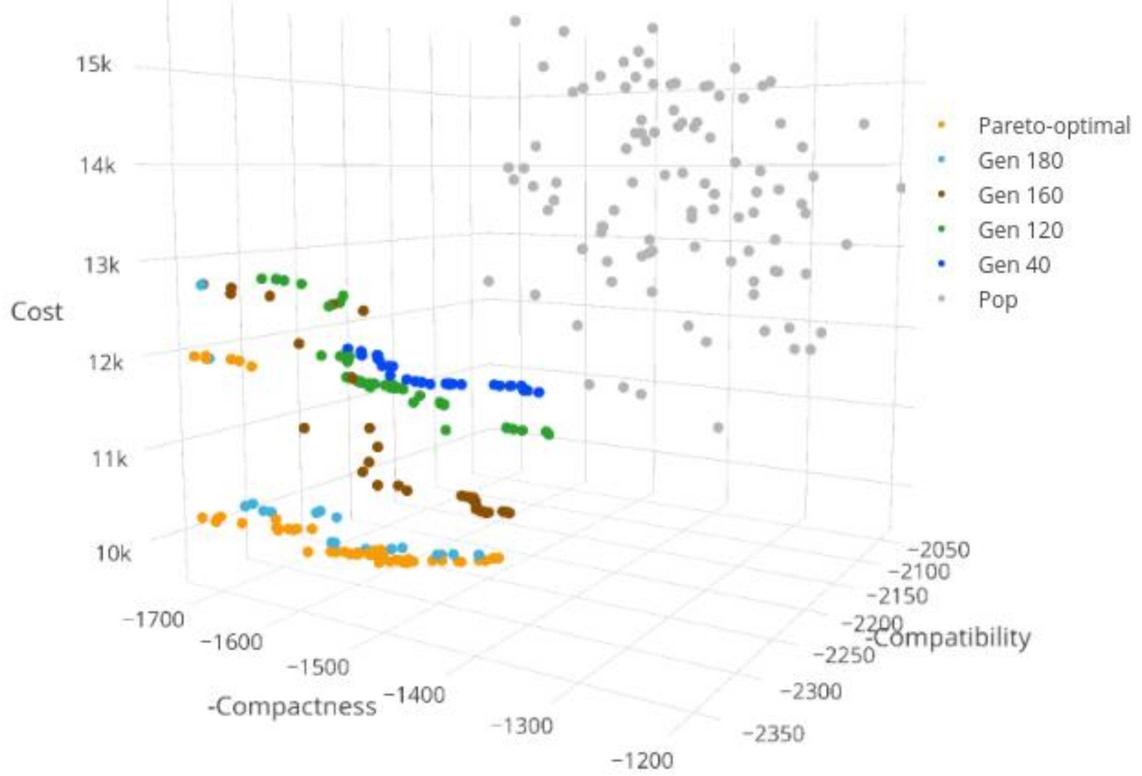


Figure 6 Two land-use plans corresponding to two non-dominant solutions in Figure 4(a): (a) the solution with best compactness ($p_c = 0.55$); (b) the solution with least cost ($p_c = 1.0$)

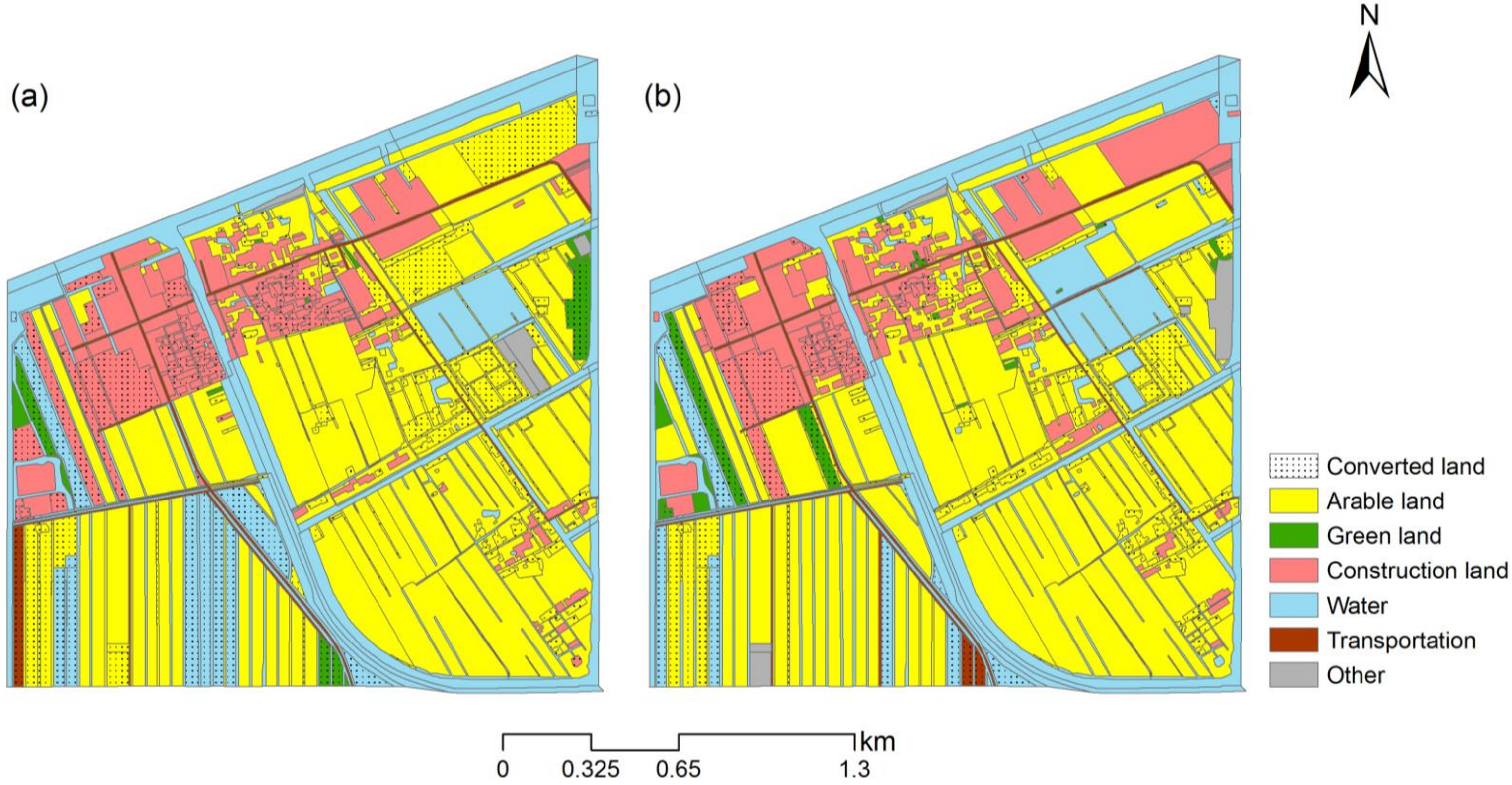


Figure 7 The land-use plan for the Pareto-optimal solution with the best compatibility in Figure 5(b)

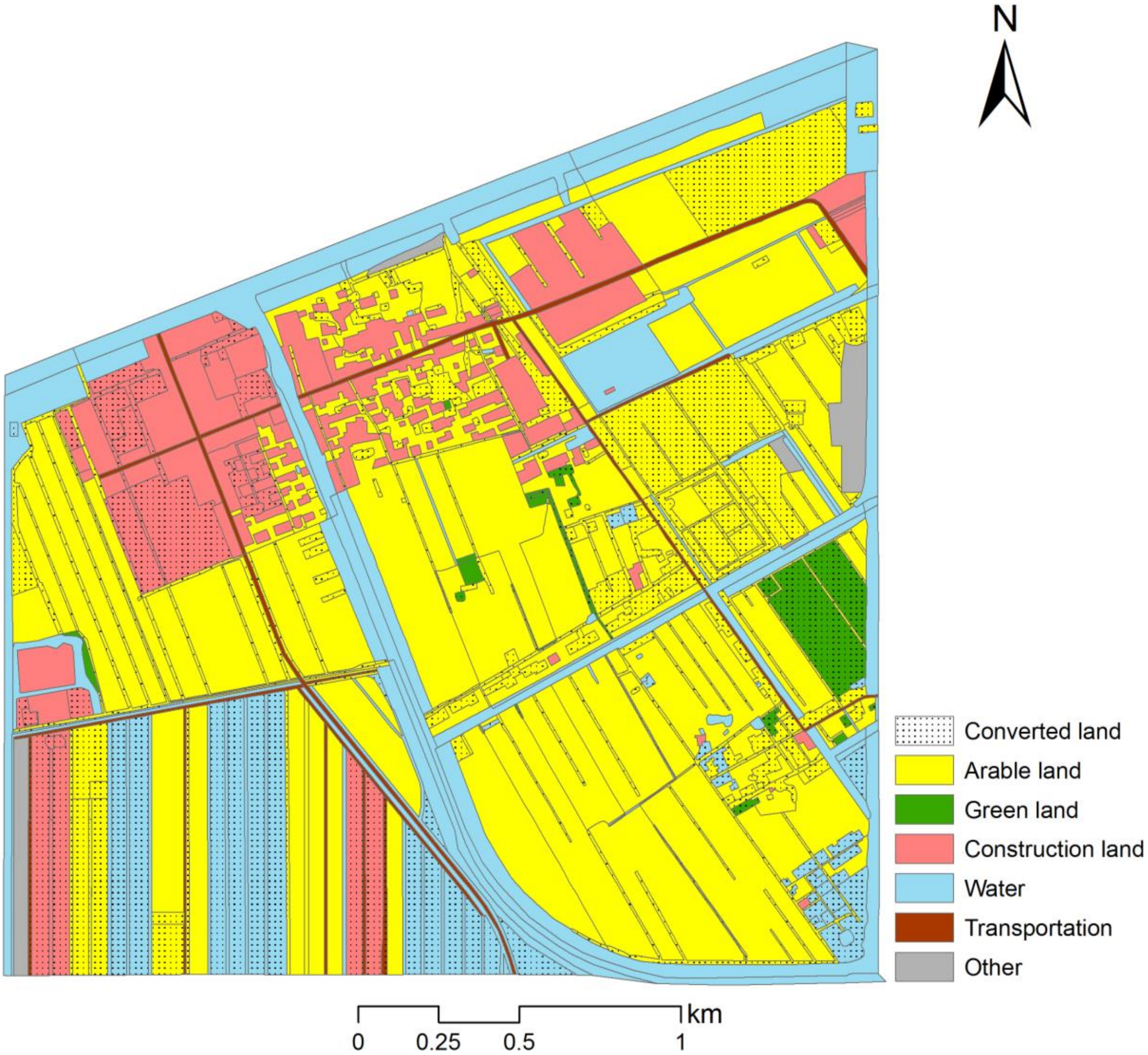


Figure 8 Area of different land uses for the plans in Figures 6 and 7

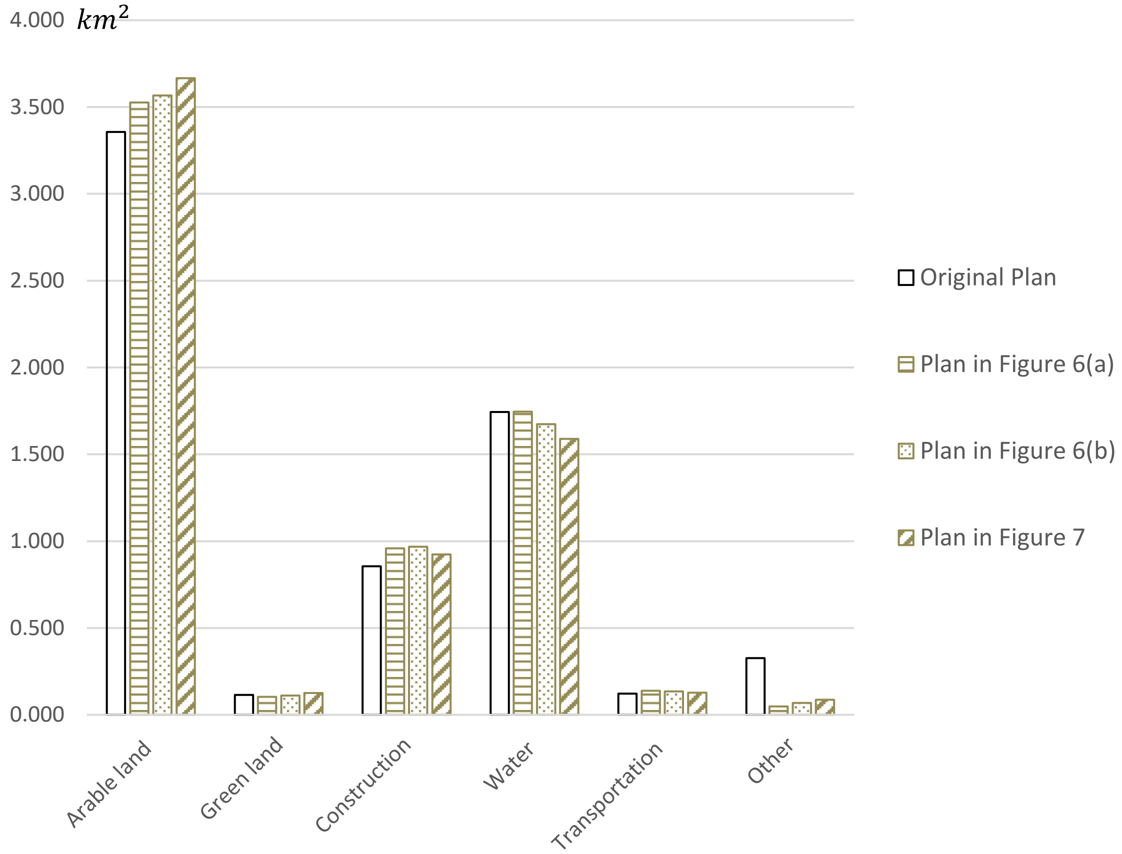


Figure 9 Raster representation of the study area

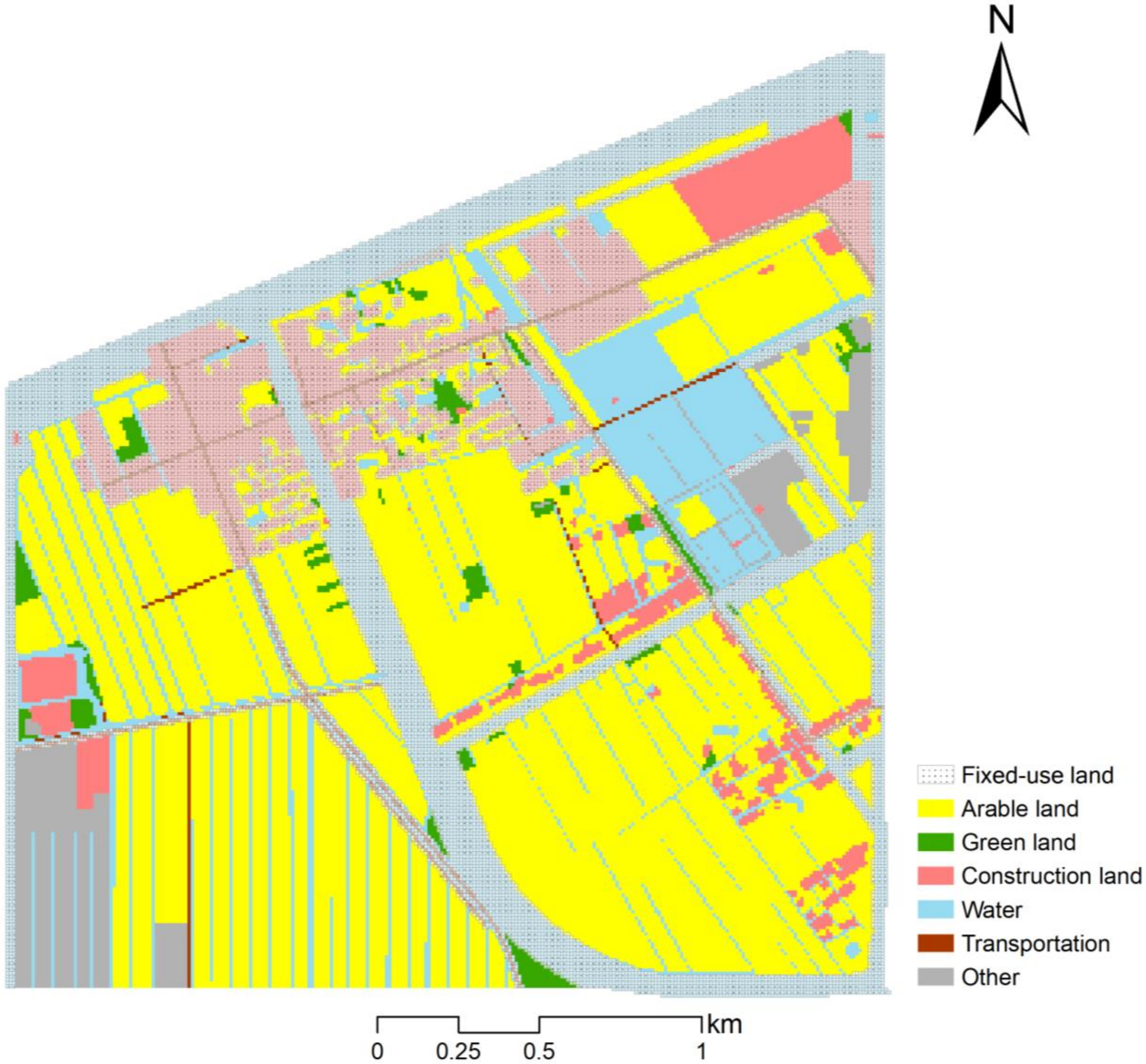


Figure 10 Land use plans generated by using raster representation: (a) the solution in Scenario I with best compactness; (b) the solution in Scenario I with least cost; (c) the solution in Scenario II with best compatibility

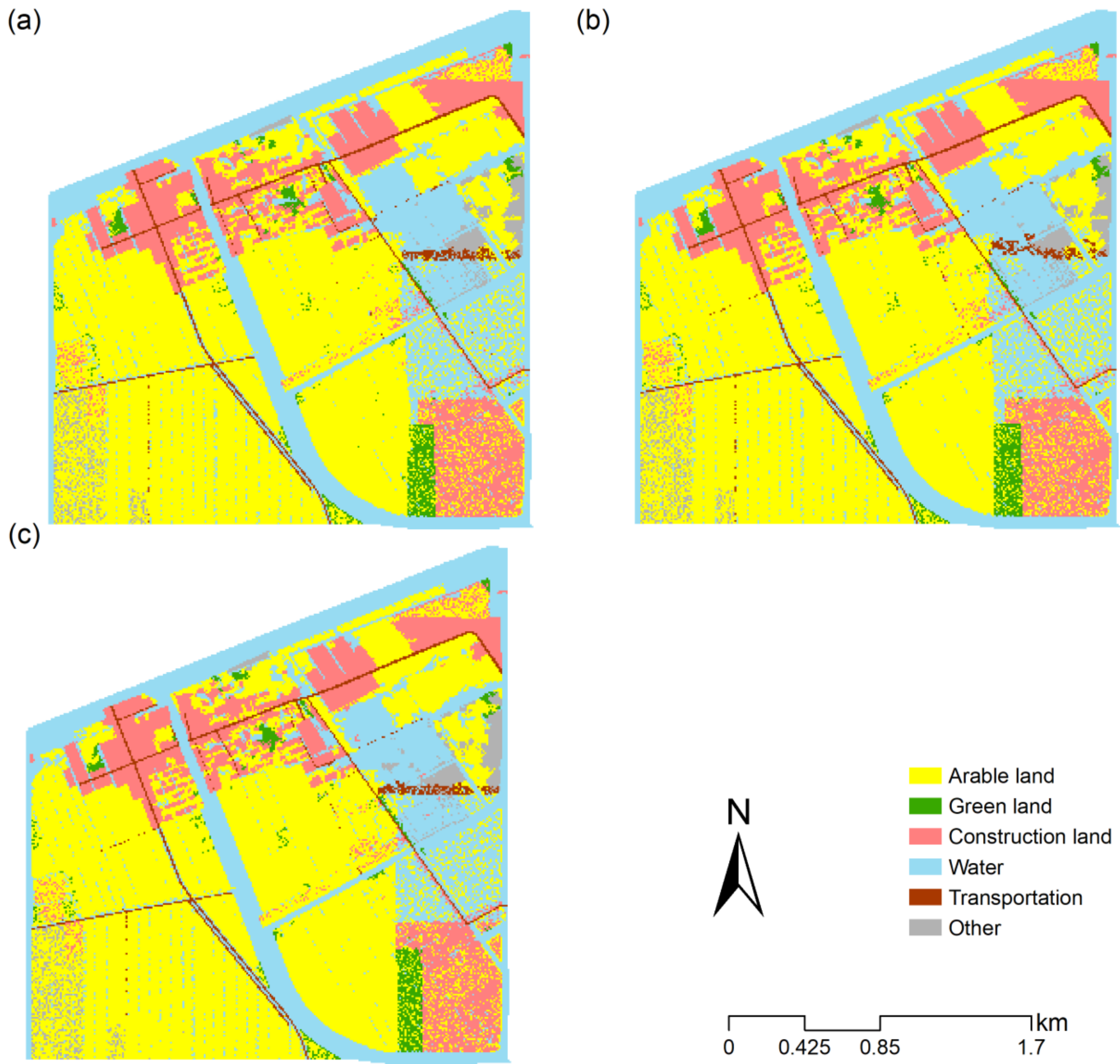


Figure 11 Area of different land uses for the plans in Figure 10

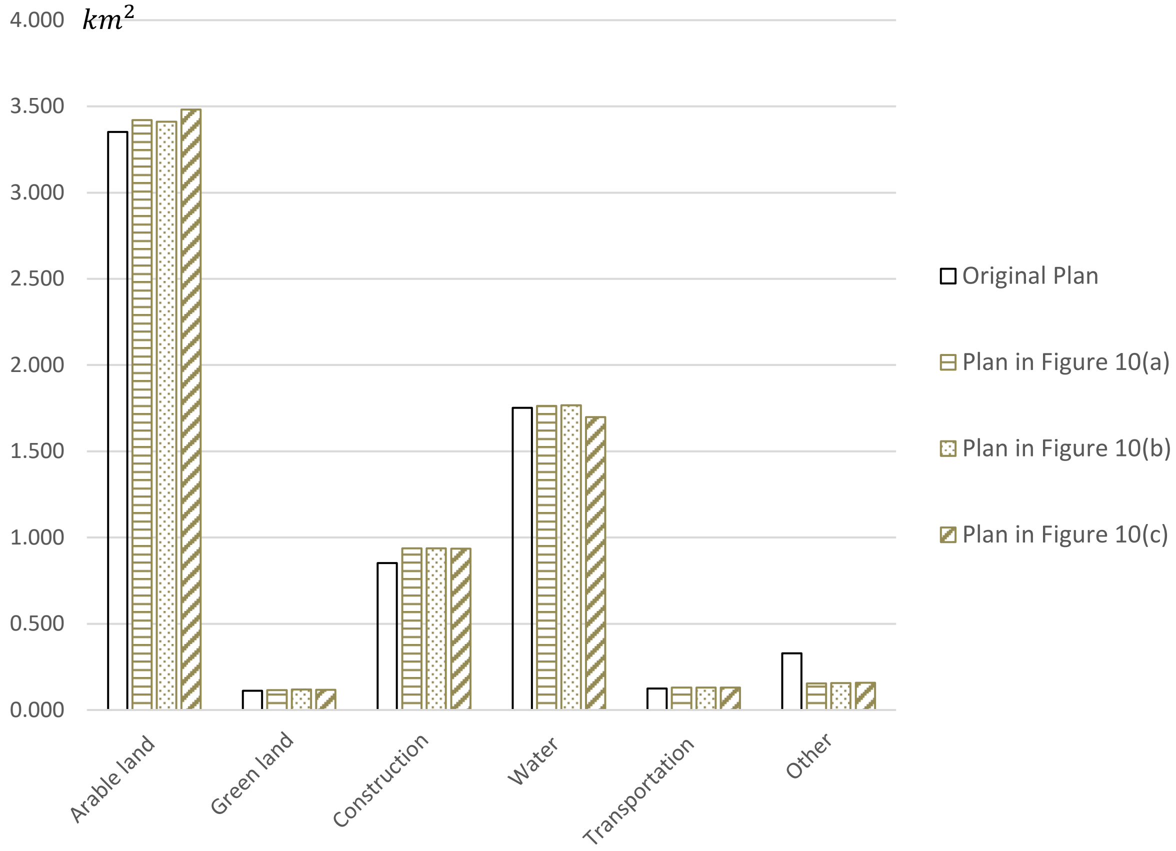


Table 1 Parameters for the solving procedure

Steps in Figure 2	**Parameter**	**Note**	**Value**	
			Scenario I	**Scenario II**
Step 1	N_P	Population size.	100	100
	N_A	The size of the archive set *A*.	100	100
	N_S	The size of each solution (i.e. the total number of parcels in a land-use plan).	680	680
	r_0	Proportion of parcels keeping current use in each solution within the initial population.	0.30	0.30
Step 5	p_c	Probability of crossover.	[0.55, 1.00] with a stepsize 0.05	0.60
	r_c	Proportion of parcels in a solution to be involved in crossover.	0.60	0.60
Step 6	p_m	Probability of mutation.	[0.01, 0.10] with a stepsize 0.01	0.05
	r_m	Proportion of parcels in a solution to be involved in mutation.	0.05	0.05
Step 9	*N_iter*	Number of iterations the algorithm runs.	200	200
	N_sameA	Number of consecutive iterations that generate the same archive sets.	10	10

Table 2 Compatibility between different land uses

	Type I	Type II	Type III	Type IV	Type V	Type VI
Type I	HH	HC	MC	HC	MC	LC
Type II		HH	HC	MC	HC	MC
Type III			HH	LC	HC	MC
Type IV				HH	MC	LC
Type V					HH	LC
Type VI						HH

* HH = 1, HC = 0.5, MC = 0.3, LC = 0.1