Optimal siting of rural settlement through a GIS-based assessment: A Case study in China

Chuanhua Zhu[1,2], Dunhui Xiao[3], QinYao Sun[1]

[1.]School of Environment and Energy Engineering, Anhui Jianzhu University, Hefei 230601, China
[2.]Anhui Engineering and Technology Research Center of Smart City, Hefei 230601, China
[3.]ZCCE, College of Engineering, Swansea University, SA1 8EN, UK
chzhu.cug@163.com

Abstract. The optimization of spatial allocation of rural settlements in China requires an evaluation of rural land suitability and location optimization for the convenience of rural residents living and working. The main idea behind this study was to propose a new methodology to locate the most suitable sites for rural residents as considering their potential demand and supply. A geographic information system (GIS) based multiple criteria decision-making (MCDM) model was used to identify the most suitable areas and the unsuitable parts. The weighted information value method was used to assigning weight for influential factors. A case study was conducted for Tangjiahui, a hilly area in the central of China, by performing a town-wide suitability assessment. The very high suitability area is 4621.05 hm² where is close to the road and with a moderate elevation and low slope, which contains 70.29% of residents. Finally, the maximal covering model was performed to determine the most suitable locations. Six sites were selected from 312 potential sites, which cover whole potential demand settlements with an average travel distance of 2.9 kilometers. The result of location optimization is reasonable and therefore the methodology is applicable.

Keywords: Hilly area, Rural settlement, Weighted information value model, Land suitability analysis, Location optimization, Maximal covering.

1   Introduction

In recent years, land use planning of rural areas are becoming popular because of the rapid expansion of urban area leading the shortage of farming land in China. As a consequence, many studies have analyzed the relationship between the spatial distribution of rural settlement and related influencing factors ,including the natural environment, economic development and social culture, on the basis of location theory [1-2], to determine the land suitability and relocate famers' houses. while western studies have focused on the pattern evolution of rural settlements, the landscape evolution of rural settlements, and the relationship between rural industries and rural settlements [3-5]. China is a country has spacious territory, rural settlements are located in various

type of land, such as hilly, plain and delta. However, in order to utilize the potential land, whatever the types of rural residential areas, the land suitability evaluation is first carried out, and then the optimization analysis. In the wider domain of geo-computation and spatial analysis, adopting the Multiple criteria decision-making [6-7] approach for land suitability analyses and a number of different methods are well-established for weight of criteria. The most popular method for site selection is the Analytic Hierarchy Process (AHP), Delphi and their combination with other methods [8-13], such as combination of Delphi and weighted entropy, combination of the AHP and weighted entropy, which are all objective that depend on the expert opinion in some extent. In addition, they produced too many potential sitings for decision-makers to choose from. Optimization analysis generally also use the qualitative methods, which just give the suggestion from macro view without the blueprint for resettlement [10-11].

However, limited research has been conducted using quantitative methods. Xie and Fan used the weighted Voronoi tessellation to represent the influenced area of rural settlement by using the comprehensive index value as the weight value [12-13], so as to determine the relocation and retention of settlement. But the analysis of Vorinoi tessellation is based on Euclidean distance and does not take the actual demand of settlements and capacity of supply into account. Maximal covering model is a kind of location-allocation model ,which considers not only the needs of demand points but also the capacity of supply, which had been successfully applied in the location optimization of police patrol areas, alternative-fuel stations, school, biogas plants, health care organizations, bioenergy facilities and bike-sharing station[14-22]. This model will be introduced in our study to analyze the location optimization of rural settlement.

The overall objective of this study work is to develop a methodology to allocate the optimal settlement site for rural residents considering their demand and capacity of candidate sites based on real road networks. The specific objectives include:(1) Developing a framework to assess the suitable location for rural settlement using a set of environmental factors;(2) Integrate GIS and weighted information value method for the development of the framework;(3) Implement a location-allocation analysis to select the most suitable sites for rural settlements in the study area.

This paper has four sections. Section 1 proceeds with a review of land suitability analyses that have been used for rural settlements and applications of location-allocation algorithms in this context. Before introducing the specific case study used to demonstrate this methodology(Section 2), Section 3 provides a detailed description of the proposed weighted information value method and maximal covering model. The results and discussion are presented in Section 4.

2   Methodology

There are two stages of the proposed method, that is land suitability analysis and location optimization, for identifying optimal sites for settlements. First, it is necessary to know the distribution of the potential demand using land suitability analysis, also

known as Multiple Criteria Decision-Making, which use different types of spatial analyze tools such as reclassify and map algebra to determine the potential supply and demand locations, and then location optimization ,which employ search heuristics to satisfy location-allocation problem such as the maximal covering location problem and seek to match spatially distributed supply and demand.

2.1 Land suitability analysis

In this work, a land suitability analysis was performed to identify suitable and unsuitable areas. In order to do so, some environmental and social factors were considered. According to the references [8,10-12] and data availability, several influence factors have been selected for land use suitability analysis of rural settlements, those are slope, aspect, elevation, distance to river, distance to road, distance to town and central villages. Each influence factors, the continuous variable, are reclassified into discrete variable by quantifying the influence on the suitability, such as the slope is divided into five categories, which are 0°-5°, 5°-10°, 8°-15°, 15°-22.5° and 22.5° or more; the aspect is divided into five categories, which are south, flat, southeast, southwest, east, west, northeast, northwest and north; the elevation is divided into five categories, which are 0-200m, 200-400m, 400-600m, 600-800m and 800m or more; the distance to river is divided into five categories, which are 0-100m, 100-200m, 200-400m, 400-600m and 600m or more; the distance to road is divided into five categories, which are 0-100m, 100-200m, 200-500m, 500-1000m and 1000m or more; the distance to market town and central villages is divided into five categories, which are 0-200m, 200-500m, 500-1000m, 1000-2000m and 2000m or more.

The weighted information value method was used to combine all these raster maps to produce the final suitability map. The method proposed by Van Westen is also called the statistical index method, was employed in landslide susceptibility mapping under a GIS environment[23]. In this method, a weighted value for a parameter class is defined as the natural logarithm of the rural settlement density class, divided by the settlement density in the entire map. The formula was given below that forms the basis of the approach:

$$Wi = \ln \frac{Densclass}{Densmap} = \ln \frac{\dfrac{Npix(Si)}{Npix(Ni)}}{\dfrac{SNpix(Si)}{SNpix(Ni)}} \tag{1}$$

Where

W_i =The weight given to a certain parameter class;

$Densclass$ =Settlements density within the parameter class;

$Densmap$ =Settlements density within the entire map;

$Npix(Si)$ =Number of pixels that contain settlements in a certain parameter class;

$Npix(Ni)$ =Total number of pixels in a certain parameter class;

4

$SNpix(Si)$ =Number of pixels that contain settlements in the entire map;

$SNpix(Ni)$ =Total number of pixels in the entire map.

The W_i method is based on statistical correlation of the settlements inventory map with attributes of different parameter maps. In this study, every parameter map was crossed with the existed settlements map, and the density of the settlements in each class was calculated. Correlation results were stored in resultant raster and the density of the settlements per parameter class was calculated. Then the W_i value of each attribute was calculated. Finally, all layers were summed up and a resultant suitability map was obtained. The final suitability map was divided into five classes by the Natural Break method. The classes are very low suitability, low suitability, moderate suitability, high suitability and very high suitability.

2.2 Location-allocation model

Maximal covering model can be applied to the problem of generating optimal concentrated-settlements with the following formula:

$$Maximize\ Z = \sum_{i\in I} a_i y_i \tag{2}$$

Subject to：

$$\sum_{j\in N_i} x_j \geq y_i\ for\ all\ i \in I \tag{3}$$

$$\sum_{j\in J} x_j = P \tag{4}$$

$$x_j = (0,1)\ for\ all\ j \in J \tag{5}$$

$$y_i = (0,1)\ for\ all\ i \in I \tag{6}$$

$$\sum_{i\in N_j} a_i x_j \leq M_j\ \ for\ all\ j \in J \tag{7}$$

Where:
Z=the total demand of rural residents serviced
I=the set of known settlements locations
J=the set of potential locations for concentrated-settlements
x_j =1 if a concentrated-settlements covers settlements locations j, and 0 otherwise
y_i=1 if an settlement locations at i is covered by at least one 1 concentrated-settlement, and 0 otherwise
P=the number of concentrated-settlements to be located
$N_i = \{j \in J | d_{ij} \leq S\}$
d_{ij}=the shortest distance from settlement location i to concentrated-settlement location j
$N_j = \{i \in I | d_{ij} \leq S\}$
S=the acceptable service distance or time
a_i =the demand of settlements location i
M_j =the maximum capacity that each concentrated-settlement at j can serve

In this model, it is assumed that the daily work of rural residents needs to go back and forth between the work zone and the existing settlements, and the distance between the existing settlements and the work zone is negligible, that is the rural residents need to go back and forth between the existing settlements and the concentrated-settlements area after location optimization.

3 Case study

3.1 Study site

The proposed method has been applied in Tangjiahui, a town of Jinzhai county in China. Jinzhai is located in the hinterland of Dabie Mountain area where is located in the centre China. And Tangjiahui is in the west part of Jinzhai and has 11 administrative villages, 1 street community, and 306 natural village, with a total area of 26850 hectares and more than 50000 inhabitants. The study area is characterized by hilly area that is about 80.4% and the forest coverage is 65.2% . Because of the natural terrain and environment, rural settlements are scattered in the more flat and less elevation area but more difficult to the road, which is the main reason for vulnerable people difficult to access the public facilities, as were shown in Fig 1.
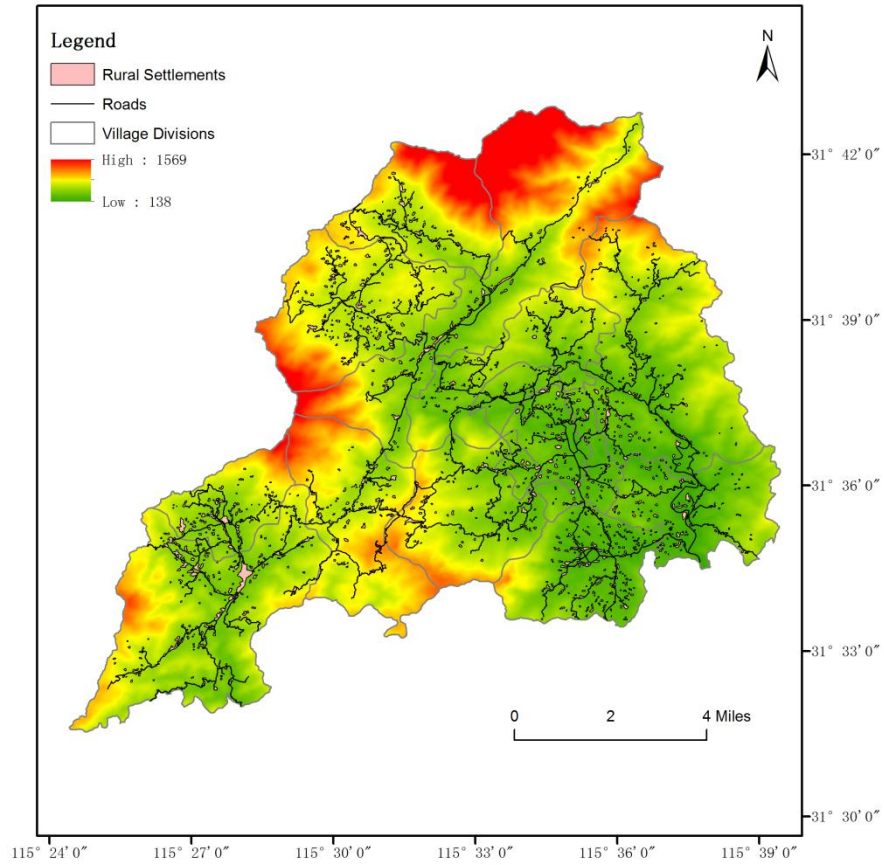
6

Fig. 1 The study area

3.2 Data used

The data used in this study involves: digitized data interpreted from Google images, DEM data, land use planning, demographic and socioeconomic data. The digitized data includes rural residential area, road, river, centers (governmental town and village). The DEM dataset, with a grid resolution of 30 meters, is provided by Geospatial Data Cloud site, Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn). The slope and aspect data are derived from the DEM data via spatial analysis, in particular, the surface analysis method. The land use planning, demographic and socioeconomic data is obtained from local governmental information website (http://www.ahjinzhai.gov.cn).

### 3.3 Evaluation of single data layer

In ArcGIS, Euclidean distance and reclassify tools are used to process each data layer, and the derived data layer interact with the settlements layer (binary raster layer) by using zonal statistics tool to calculate the information value W_i, as shown in Table 1.

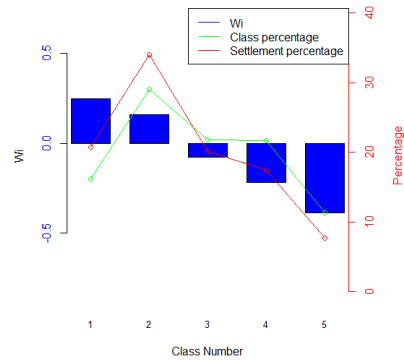**Table 1.** Influence factors for suitability analysis of rural settlements

Data layers	Classes number	Classes	Class area (%)	Settlements areas (%)	W_i
Slope	1	0°~5°	11.07	25.52	0.835
	2	5°~10°	23.40	56.48	0.881
	3	10°~15°	22.06	19.44	-0.127
	4	15°~22.5°	24.32	9.08	-0.985
	5	>22.5°	19.15	3.12	-1.814
Aspect	1	South、Flat	16.15	20.72	0.249
	2	Southeast、Southwest	29.06	34.03	**0**.158
	3	East、West	21.80	20.13	-0.079
	4	Northeast、Northwest	21.68	17.44	-0.217
	5	North	11.30	7.68	-0.387
Elevation	1	0-200m	2.66	5.92	0.799
	2	200-400m	17.00	30.96	0.600
	3	400-600m	20.96	31.85	0.418
	4	600-800m	20.83	17.66	-0.165
	5	>800m	38.55	13.61	-1.041
Distance to river	1	0-100m	12.00	22.33	0.621
	2	100-200m	8.18	15.93	0.667
	3	200-400m	15.37	17.16	0.110
	4	400-600m	13.78	14.05	0.019
	5	>600m	50.67	30.54	-0.506
Distance to	1	0-100m	21.50	87.41	1.403

road	2	100-200m	15.43	11.10	-0.329
	3	200-500m	11.41	1.17	-2.275
	4	500-1000m	7.27	0.27	-3.283
	5	>1000m	44.39	0.04	-6.989
Distance to center	1	0-200m	0.63	5.11	2.101
	2	200-500m	3.38	9.57	1.042
	3	500-1000m	11.33	21.19	0.626
	4	1000-2000m	31.73	42.01	0.280
	5	>2000m	52.93	22.11	-0.873

The W_i calculated by weighted information value method indicate the correlation of settlements with data layer, as shown in Table1and Fig 2 that the parameter layers with strong positive correlation with the settlements layer are as followed: 0 °-5 ° and 5 °-10 ° of slope class, south and flat class of aspect, 0-200m and 200-300m of elevation class, 0-100m and 100-200m class of distance to river, 0-50m class of distance to road and 0-200m and 200-500m class of distance to center.
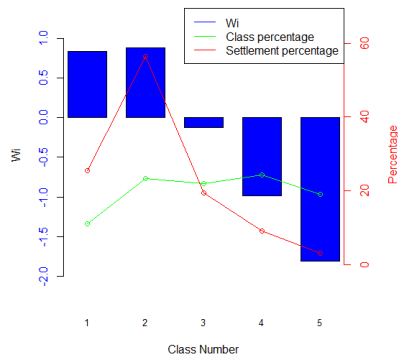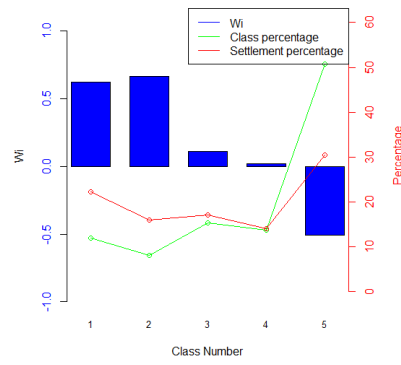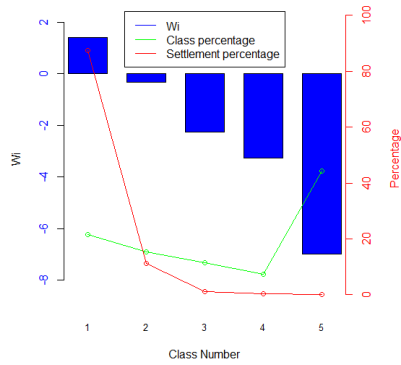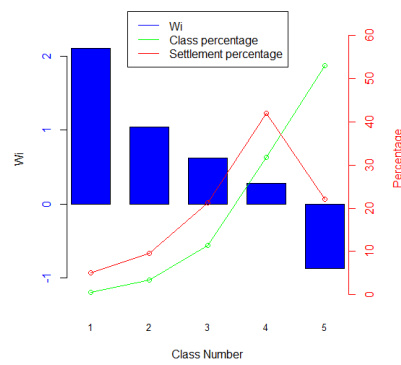

a. Reclassify of slope   **b.** Reclassify of aspect
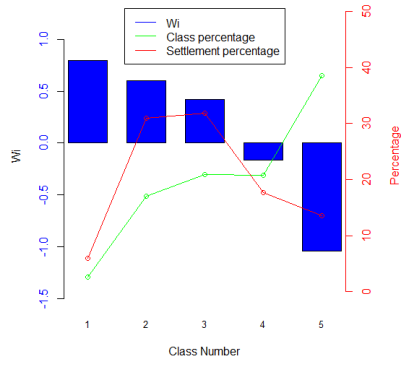
c. Reclassify of elevation

d. Reclassify of distance to river

e. Reclassify of distance to road

f. Reclassify of distance to town and village center

Fig. 2 Reclassify of influence factors of rural residential land in Tangjiahui

4 Result and discussion

4.1 Suitable areas suggestions

The suitability index map is derived from the influence layers. Add operator in the map algebra toolset was used to sum up all data layers, then the resultant suitability map was divided to five classes according to the information value stored in each single grid, as shown in Fig 3.
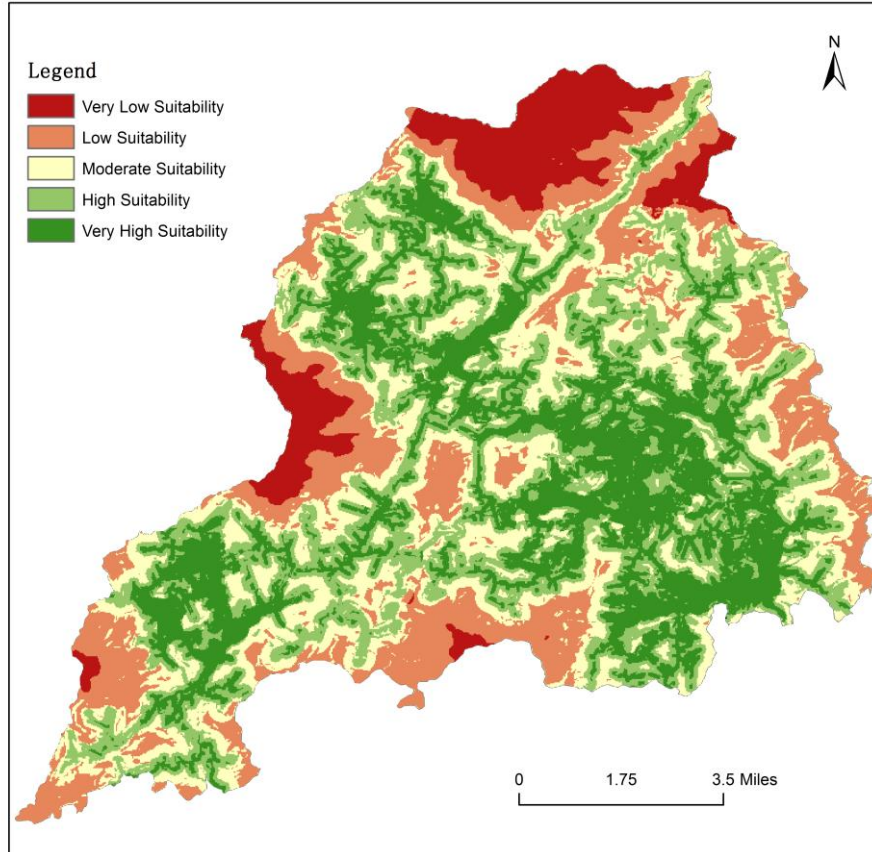
10

Fig. 3 Suitability map of Tangjiahui

4.2 Validation of the land suitability model

The zonal statistics tool was use to statistic the distribution of settlements layer in suitability classes. As shown in Table 2, the very high suitability area is only accounts for 18.03% of the total area, but contains about 70.29% of the existing settlements, and the b/a ratio of classes from unsuitable area to suitable area is gradually increased, which indicate the classification is good and model is useful.

Table 2. Model verification of suitability zoning

Suitability level	Area （hm²）	a(%)	b(%)	c(%)	b/a
Very low suita-bility	7237.35	28.24	0.04	0.004	0.001
Low suitability	4417.56	17.24	0.03	0.004	0.002

Moderate suitability	4068.9	15.88	3.10	0.50	0.20
High suitability	5283.81	20.62	26.53	3.31	1.29
Very high suitability	4621.05	18.03	70.29	10.04	3.90

Notation: a indicated the grid number of division as a percentage of the total grid number; b indicated the grid number of settlements in division as a percentage of the total grid number of settlements; c indicated the grid number of settlements in division as a percentage of the grid number of division.

4.3 Location optimization

There are three steps to decide the optimal sites. Firstly, determine the potential demand (settlements located in unsuitability) and supply (concentrated-settlements). The grids in the suitability map with high suitability or above were converted to vector format ,and then intersect with the construction areas that allowed by government, therefore the candidate parcels with the area more than 1000 square meters were selected, which were used as potential supply. The raster grids with low suitability or below were also converted to vector format, and then intersect with the existing settlements, therefore the selected settlements were used as potential demand. Secondly, all the potential demand and supply parcels were converted to points and then connected to the nearest node in the road network, which were set as the demand point and supply point respectively. Finally, location-allocation model of network analyst module in ArcGIS is used to solve the capacitated maximum covering problem with constrains.

Fig.5 shows there are 312 parcels of potential supply ,of which the largest capacity is 965 (counted by 5 residents per 120 square meters) ,and 28 parcels of potential demand include 662 residents.
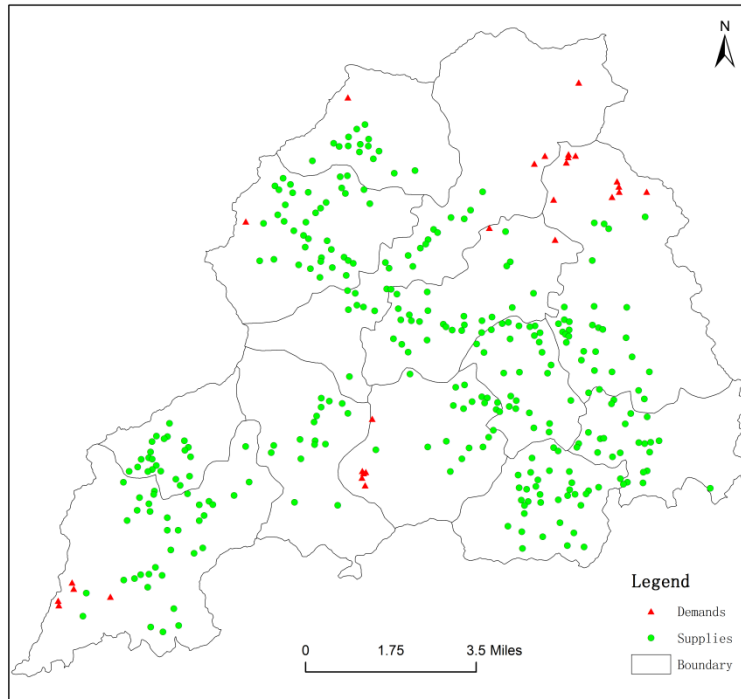
12

Fig.5 Candidate parcels and settlements to be relocated

Maximal covering models with capacity and without capacity were both used to identify optimal sites for concentrated-settlements. Indices include average travel distance, the longest single distance , the least single distance , the ratio of covering and the ratio of service were used to evaluate the model. Table 3 shows when choosing the uncapacitated maximal covering without capacity and just 1 candidate site to be chosen, the location-allocation map is show in Fig.6, that is the covering rate is 100%, but the single longest distance is 29.9 kilometers and average travel distance is 14.2 kilometers, of which result is unreasonable; while when choosing the model with capacity and 6 candidate site to be chosen with the acceptable distance of 10000 meters, the location-allocation map is show in Fig.7, and covering rate is 100%, and the single longest distance is 9.5 kilometers and average travel distance is 2.9 kilometers, of which result is reasonable.

Table 3. The statistics indices of location optimization

| The number of located sites (P)	ID of Selected sites	The acceptable service distance S in meters	Is capacitated or not	The number of coverage settlements/ The ratio of coverage(%)	The number of people serviced/ The ratio of service(%)	Average travel distance in meters	Least single distance in meters	Longest single distance (m)
1	414	50000	not	28/100	662/100	14220	4153	29971
1	null	10000	yes	14/50	399/60.3	null	null	null
2	null	10000	yes	21/75	552/83.4	null	null	null
3	null	10000	yes	25/89.3	599/90.5	null	null	null
4	null	10000	yes	27/96.4	658/99.4	null	null	null
5	null	10000	yes	27/96.4	658/99.4	null	null	null
6	318,337,354, 356,523,621	10000	yes	28/100	662/100	2910	1155	9551
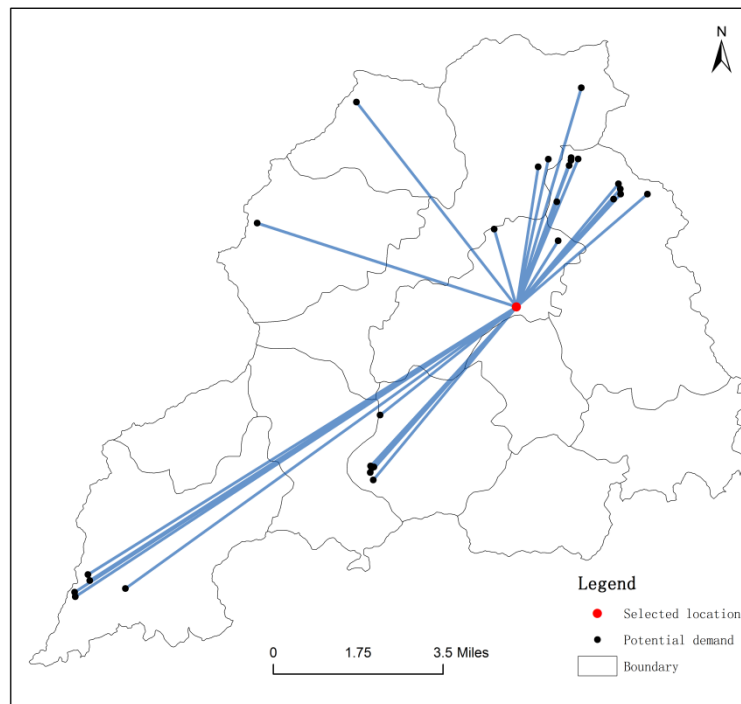
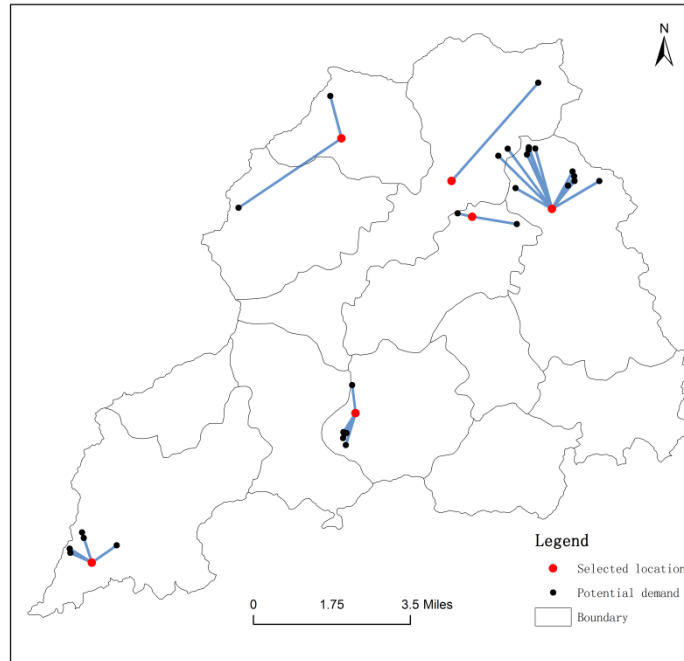**Fig.6** Allocation of location optimization subject to P=1,S=50000

14

Fig.7 Allocation of location optimization subject to P=6,S=10000

5 Conclusion

In this study, we have shown the possibility of the maximal covering model integrated with the land suitability evaluation model and applied to a proposal for the optimal location of rural concentrated-settlements in Tangjiahui, a town in Dabie mountain area of China. The result of land suitability evaluation showed the area of very high suitability is 4621.05 hm² , which is close to the road and located in zone with the moderate elevation and low slope, with the 70.29% of residents. 6 sites were selected from 312 potential sites by capacitated maximal covering model, which cover the whole potential demand settlements with an average travel distance of 2.9 kilometers, therefore result of optimal sitting is reasonable. The results produced by the proposed method are useful and applicable for local administration in reallocating the rural settlements. Nevertheless, it is clear that there are some limitations. Firstly, the influence factors of land suitability evaluation need to be considered thoroughly, which the distribution of rural settlements is not only influenced by the natural environment but also by the social, economic, ethical, and other facets. The assignment of settlements demand is the second limitation, which was distributed by the area of settlements rather than the actual data because of the difficulty of the collection at the village scale. In conclusion, the methodology proposed in this paper considers the actual

demand and supply capacity, which is a practicable and quantifiable way to determine the optimal location for rural settlements relocating.

References

1. Li Jun, Li Xiaojian.: A Review on location of the rural settlements. Human Geography 23(4),23–27(2008).
2. Jiang Guanghui, Zhang Fengrong, Qin Jing, Zhang Lin, Gong Pan.: Relationship between distribution changes of rural residential land and environment in mountainous areas of Beijing. Transactions of the CSAE 22(11),85–92(2006).
3. Carrion-Flores C，Irwin E G.: Determinants of residential land use conversion and sprawl at the rural-urban fringe. American Journal of Agricultural Economics 86(4),889–904(2004).
4. Paquette S, Domon G.: Changing ruralities, changing landscapes: exploring social recomposition using a multi-scale approach. Journal of Rural Studies 19(3),425–444(2003).
5. Sevenant M, Antrop M.: Settlement models, land use and visibility in rural landscapes: two case studies in Greece. Landscape and Urban Planning 80(4),362–374(2007).
6. Jose M.C.Pereira, Lucien Duckstein.: A multiple criteria decision-making approach to GIS-based land suitability evaluation. International Journal of Geographical Information Systems 7(5),407–424(1993).
7. Mehmet Kabak, Mehmet Erbas, Cihan Cetinkaya, Eren Ozceylan. A GIS-based MCDM approach for the evaluation of bike-share stations. Journal of Cleaner Production 201, 49–60(2018).
8. Gao Hongjuan, Zhang Chaoqiong, Cai Guangpeng, Luo Xuqiang.: Land Suitability Evaluation of Rural Settlements in Krast Mountains Based on GIS. Research of Soil and Water Conservation 22(2), 200–203(2015).
9. Md.Mohib-UI-Haque Khan, Mahdi Vaezi, Amit Kumar.: Optimal siting of solid waste-to-value-added facilities through a GIS-based assessment. Science of the Total Environment 610–611,1065–1075(2018).
10. Qu Yanbo, Zhang Fengrong, Jiang Guanghui, Guan Xiaoke, Guo Lina.: Suitability evaluation and subarea control and regulation of rural residential land based on niche. Transactions of the CSAE 26(11),290–296(2011).
11. Jiang Lei, Lei Guoping, Zhang Jian, Zhang Yiran, Li jia.: Analysis of Spatial Distribution and Optimization of Rural Settlement. Research of Soil and Water Conservation 20(1),224–229(2013).
12. Xie Zuolun, Zhao Ruifeng, Jiang Penghui,Liu Haimeng, Zhu Wen.: The rural residential space reconstruction in Loess Hilly Regions: A case study of Yuzhong county in Lanzhou. Geographical Research 33(5),937–947(2014).

13. Fan Tianxiang, Yang Qingyuan, He Jian, Pan Fei, Deng Yongwang.: Spatial distribution optimization of rural residential land in hilly areas: A case study of Haitang town in Changshou district. Geographical Research 34(5),883–894(2015).
14. Curtin Kevin M, Hayslett-McCall Karen, Qiu Fang.: Determining Optimal Police Patrol Areas with Maximal Covering and Backup Covering Location Models. Netw Spat Econ 10,125–145(2010).
15. Upchurch Christopher, Kuby Michael, Lim Seow.: A Model for Location of Capacitated Alternative–Fuel Stations. Geographical Analysis 41,85–106(2009).
16. Kong Yunfeng,Wang Zhen.: Optimal Location-allocation for County-level Compulsory School Site Selection Using GIS and Integer Linear Programming. Journal of Geo-information science(Chinese Eidition) 14(3),299–304(2012).
17. Qie Huan, Li Yaguang.: GIS-based spatial distribution characteristics of settlements and its optimization in the tropical Lingshui County of Hainan Province.Science of Soil and Water Conservation 15(5),78–85(2017).
18. Kamalakanta Sahoo, Sudhagar Mani, Lalitendu Das, Pete Bettinger.: GIS-based assessment of sustainable crop residues for optimal siting of biogas plants. Biomass and Bioenergy 110 ,63–74(2018).
19. Siddhartha S. Syam, Murray J. Côté: A location-allocation model for service providers with application to not-for-profit health care organizations. Omega (38),157–166(2010).
20. Alexis Comber, Jennifer Dickie, Claire Jarvis, Martin Phillips, Kevin Tansey.: Locating bioenergy facilities using a modified GIS-based location-allocation-algorithm: Considering the spatial distribution of resource supply. Applied Energy 154,309–316(2015).
21. Aifa Sultana, Amit Kumar.: Optimal siting and size of bioenergy facilities using geographic information system. Applied Energy 94,192–201(2012).
22. Juan Carlos Garcia-Palomares,Javier Gutierrez,Marta Latorre.: Optimal the location of stations in bike-sharing programs: A GIS approach. Applied Geography 35,235–246(2012).
23. Cevik E, Topal T.: GIS-based landslide susceptibility mapping for a problematic segment of the natural gas pipeline, Hendek (Turkey). Environmental Geology 44,949–962(2003).