

Journal of
Applied Remote Sensing

**Simulating urban expansion using an
improved SLEUTH model**

Xinsheng Liu
Rui Sun
Qingyun Yang
Guiwu Su
Wenhua Qi

Simulating urban expansion using an improved SLEUTH model

Xinsheng Liu,^{a,b,c} Rui Sun,^{a,b,c} Qingyun Yang,^d Guiwu Su,^e and Wenhua Qi^e

^aBeijing Normal University and IRSA, CAS, State Key Laboratory of Remote Sensing Science, Beijing 100875, China

sunrui@bnu.edu.cn

^bBeijing Normal University, School of Geography, Beijing 100875, China

^cBeijing Normal University, Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing 100875, China

^dBeijing Normal University, School of Foreign Languages and Literature, Beijing 100875, China

^eInstitute of Geology, China Earthquake Administration, Beijing 100029, China

Abstract. Accelerated urbanization creates challenges of water shortages, air pollution, and reductions in green space. To address these issues, methods for assessing urban expansion with the goal of achieving reasonable urban growth should be explored. In this study, an improved slope, land use, exclusion, urban, transportation, hillshade (SLEUTH) cellular automata model is developed and applied to the city of Tangshan, China, for urban expansion research. There are three modifications intended to improve SLEUTH: first, the utilization of ant colony optimization to calibrate SLEUTH to simplify the calibration procedures and improve their efficiency; second, the introduction of subregional calibration to replace calibration of the entire study area; and third, the incorporation of social and economic data to adjust the self-modification rule of SLEUTH. The first two modifications improve the calibration accuracy and efficiency compared with the original SLEUTH. The third modification fails to improve SLEUTH, and further experiments are needed. Using the improvements to the SLEUTH model, forecasts of urban growth are performed for every year up to 2020 for the city of Tangshan under two scenarios: an inertia trend scenario and a policy-adjusted scenario. © 2012 Society of Photo-Optical Instrumentation Engineers (SPIE). [DOI: [10.1117/1.JRS.6.061709](https://doi.org/10.1117/1.JRS.6.061709)]

Keywords: urban expansion; ant colony optimization; cellular automata; SLEUTH, Tangshan.

Paper 12192SS received Jun. 26, 2012; revised manuscript received Nov. 26, 2012; accepted for publication Nov. 29, 2012; published online Dec. 27, 2012.

1 Introduction

Urban areas are complex systems, with characteristics of openness, nonlinearity, self-organization, randomness, hierarchy, and dissipation. The cellular automata (CA) method was first introduced by Tobler¹ to model the rapid urban sprawl of the city of Detroit in 1979, and it remains one of the most important methodologies used to study complex systems. In the 1980s, Couclelis^{2,3} provided a generalization of the theory of CA and outlined the interaction between urban structure and individual behavior. Beginning in the 1990s, a series of typical CA models of urban expansion were developed: dynamic urban evolution model, by Batty and Xie;^{4,5} slope, land use, exclusion, urban, transportation, hillshade (SLEUTH), by Clarke and colleagues,^{6,7} which was named after its model inputs, namely, slope, land use, exclusion, urban, transportation, hillshade; and land use evolution and impact assessment model by Deal and Sun.⁸ Among these three models, SLEUTH is preferred for urban expansion studies due to its more readily available input data, its few parameters for calibration, and its relatively good simulation accuracy.^{9–17}

CA models are discrete in time, space, and state. The classic CA model consists of four parts:¹⁸ (1) a cell space with grid cell as the basic unit; (2) cell states, i.e., the land use type (urban or nonurban) of a cell; (3) the neighborhood, which is composed of a set of cells in a defined area that surround some central cell (it is the neighborhood cells that have an influence

on the state of their center cell at the next time step); and (4) transition rules, the core of CA models, which determine how cells change from one state to another at each time step. In actual applications, the classic CA model is adapted to better simulate real urban systems by adjusting attributes such as the homogeneity of the cell space, the uniformity of the neighborhood interactions, and universal transition functions.¹²

SLEUTH is characterized by a series of rules in a nested loop to simulate urban growth, focusing on the transition from nonurban to urban land use. SLEUTH's inputs require four or more historic urban seed images, two or more transportation network images, one slope image, one hillshade (background) image, and an exclusion image. Urban images, even transportation images, can be derived easily from remote-sensing data; slope and hillshade layers can be produced by a digital elevation model (DEM) with a geographic information system (GIS). Only five calibration parameters are needed for SLEUTH. Despite these advantages, three major problems associated with SLEUTH can be summarized as follows: (1) the calibration procedure is time-consuming, and although the calibration accuracy is gradually improved by using the three-level calibration (coarse, fine, and final), the resulting parameters are generally suboptimal; (2) there is only one set of parameters for a whole urban area, which is insufficient for multicore cities; and (3) the model predicts urban growth based on a simple extrapolation of inert historic growth trends calibrated by at least four time control points, without taking account of influences caused by macro driving forces, such as social and economic factors.

To address these three issues, this paper introduces three modifications to SLEUTH. First, ant colony optimization (ACO) is adopted to calibrate SLEUTH and improve its calibration efficiency and accuracy. ACO is a type of artificial intelligence algorithm that simulates the food-hunting behaviors of ants to optimize parameters. Second, the whole study area is divided into several subareas, and different parameter sets are obtained for each subarea. Third, macro social economic factors are incorporated into SLEUTH to adjust the self-modification rule of SLEUTH.

2 Study Area

The city of Tangshan, China, is selected for the study to validate the improved SLEUTH model. Tangshan was completely destroyed by a massive earthquake in 1976. Rapid development in the 30 + years since the earthquake has led to the recovery of Tangshan, which is now a modern city with an urban population of >3 million. The city is also an important part of the Beijing-Tianjin-Tangshan metropolitan area, the third largest metropolitan area in China, which means Tangshan may undergo considerable further development in the future. The study area is illustrated by the blue-bordered rectangle box in Fig. 1, which covers an area of 3445 km².

3 Data and Methods

3.1 Introduction of SLEUTH

Urban dynamics are simulated by four growth rules in SLEUTH: spontaneous growth, new spreading center growth, edge growth, and road-influenced growth. Spontaneous growth simulates the random urbanization phenomenon. New spreading center growth is oriented to the generation of new urban centers, which results in further growth. Edge growth characterizes the tendency of cities to expand outward while experiencing infilling growth. Road-influenced growth simulates the tendency of urbanization to occur next to and along transportation lines. Each growth rule is controlled by the five model coefficients of diffusion, breed, spread, road gravity, and slope resistance. All five parameters are represented by integers ranging from 0 to 100 and are dimensionless. With the exception of slope resistance, each of the parameters reflects the relative contribution to urban growth. The last important rule of SLEUTH is the self-modification rule for parameter adjustment. According to the rule, the model parameters will be multiplied by a value (BOOM) >1 when the simulation growth rate of an urban area is greater than the threshold CRITICAL_HIGH, and the model parameters will be multiplied by a value (BUST) <1 when the simulation growth rate is lower than the threshold CRITICAL_LOW. Without the self-modification rule, SLEUTH would produce only linear or exponential growth.¹⁴



Fig. 1 Study area. The location of the study area in China is outlined in red in the sub-map at the top left corner.

SLEUTH uses a brute-force method, making attempts on sufficient combinations of model parameters to perform a calibration and derive a set of parameters that can best capture the historic growth trend of a study area. The calibration procedure includes three phases: coarse, fine, and final. Each phase includes several Monte Carlo iterations to ensure the robustness of the resulting parameters. During the three-level calibration, the value ranges of the parameters are narrowed level by level, and the calibration accuracy is improved step by step.

Thirteen indices can be used to assess the calibration results of SLEUTH modeling. All indices range from 0 to 1. The greater an index, the higher the calibration accuracy. Indices $\text{pop-}r^2$, $\text{edges-}r^2$, $\text{clusters-}r^2$, and leesalee were chosen for this study, with leesalee used as the metric to measure calibration accuracy. Pop is the amount of urbanized cells, and $\text{pop-}r^2$ is the least square regression score of the modeled urban area compared with the actual urban area for the control years. The same is true for $\text{edges-}r^2$ and $\text{clusters-}r^2$, the total number of edges and urban clusters. Leesalee is a shape index that measures the spatial fitness between the modeled growth and the known urban extent for control years and can be formulated as

$$\text{leesalee} = (A \cap B) / (A \cup B), \quad (1)$$

where A is the modeled urban extent, and B is the observed urban extent.

3.2 Model Input and Macro Factors

3.2.1 SLEUTH input

Landsat TM and ETM+ images of 1992, 1999, 2005, and 2009 were collected to derive four urban layers, four transport layers, and one exclusion layer. The urban layers were extracted from

land cover maps of Landsat images, with a support vector machine classification method and an overall accuracy >80%; the transport layers were produced using on-screen digitization of the same images, with references of two history transportation maps,^{19,20} and the resulting vector layers were converted into raster images. The exclusion layer, which covers lands that are partially or fully protected from urban development, consists of local parks and water bodies from the 2009 land cover map. The exclusion-layer value ranged from 0 (no constraints) to 100 (100% excluded). Slope and hillshade layers were created from an ASTER GDEM (http://lpdaac.usgs.gov/get_data_with_a_spatial_resolution_of_1_arc_s), which is a product of Japan Ministry of Economy, Trade, and Industry and U.S. National Aeronautics and Space Administration (NASA). All SLEUTH input layers have a spatial resolution of 30 × 30 m in this study.

3.2.2 Macro factors

The macro factors of society and economy drive urban sprawl.²¹ The behavior of a city is directly dependent on its economy and its changing internal mix of industry, housing, and population.²² It is often the case that higher urban population concentration promotes the urban economy. Due to urban economic development, more and more people are attracted to the cities for higher income, which leads to more urban land demand. This pattern fuels the need to introduce macro factors to the SLEUTH model.

Population (POP), gross domestic product (GDP), investment in fixed assets (FAI), urban per-capita income (UPI), and rural per-capita income (RPI) were chosen to reflect the influence of macro factors on urban sprawl in this study. The macro factor data span from 1992 to 2007 and cover areas designated as urban center (UC), Fengrun District (FRD), and Fengnan District (FND) (Fig. 1). Data from Luan County and Luannan County were excluded because their built-up areas are not within the study area.

The derivatives per-capita GDP (GDP_P), per-capita FAI (FAI_P), and the ratio between UPI and RPI (INCOME_RATIO) were the final macro factors for this study. For an easy usage of macro factors, logistic regression equations for the five raw statistics were established (Table 1).

3.3 Improvements to SLEUTH

3.3.1 ACO calibration algorithm

In the wild, a colony of ants has the ability to establish the shortest path from their nest to food without any clues. An ant can secrete a chemical called a pheromone on its way to search for food and on the return path. As the ant moves, it can sense the existence and concentration of pheromones, which help it make decisions about which direction to take; the ant always tends to choose the direction with the highest pheromone concentration. A second ant can follow the same pheromone-covered path to find the food source and will also secrete pheromones. The path taken by a single ant may not be the shortest, but the food hunting behavior of a crowd of ants has the capacity to find the shortest route. Inspired by the food hunting behavior of ants, Dorigo and Di Caro²³ were the first to systematically utilize the ACO algorithm and apply it to a series of combinatorial optimization problems.

It is possible to use ACO to calibrate SLEUTH and explore a set of optimal parameters. The set of optimal parameters is the shortest path, and the calibration accuracy corresponds to the food source. The key is to find a way to map ACO to calibrate SLEUTH, and the transformation can be summarized in the four steps described below:²⁴

1. Walk path. The walk path of ant k can be expressed as [diffusion, breed, spread, slope, road gravity] $_k$, and each parameter ranges from 0 to 100. For a detailed realization of ACO, there are five variables on behalf of the five parameters, and each variable is expressed as a three-decimal-digit number (one decimal place). There exists a 10-by-10 array to indicate pheromone concentrations on all possible paths between the previous digit and current digit. The digit preceding the first is set to zero for each variable, and the last digit of each variable has no subsequent digit.

Table 1 Regression equations for five raw statistics of each subregion.

| Subregion | Data | Equation | R ² |
|-----------|------|--|----------------|
| FRD | POP | $y = 1/(5 \times 10^{-7} + 7.201 \times 10^{-7} \times 0.989^{(t-1991)})$ | 0.987 |
| | GDP | $y = 1/(1 \times 10^{-7} + 5.893 \times 10^{-6} \times 0.831^{(t-1991)})$ | 0.954 |
| | FAI | $y = 1/(1 \times 10^{-6} + 2.308 \times 10^{-5} \times 0.834^{(t-1991)})$ | 0.968 |
| | UPI | $y = 1/(1 \times 10^{-5} + 3.575 \times 10^{-4} \times 0.914^{(t-1991)})$ | 0.946 |
| | RPI | $y = 1/(1.25 \times 10^{-5} + 8.334 \times 10^{-4} \times 0.900^{(t-1991)})$ | 0.838 |
| FND | POP | $y = 1/(1 \times 10^{-6} + 9.946 \times 10^{-7} \times 0.985^{(t-1991)})$ | 0.979 |
| | GDP | $y = 1/(1 \times 10^{-7} + 5.035 \times 10^{-6} \times 0.834^{(t-1991)})$ | 0.962 |
| | FAI | $y = 1/(1 \times 10^{-6} + 1.838 \times 10^{-5} \times 0.835^{(t-1991)})$ | 0.950 |
| | UPI | $y = 1/(1 \times 10^{-5} + 3.410 \times 10^{-4} \times 0.913^{(t-1991)})$ | 0.939 |
| | RPI | $y = 1/(1.25 \times 10^{-5} + 7.265 \times 10^{-4} \times 0.904^{(t-1991)})$ | 0.833 |
| UC | POP | $y = 1/(2 \times 10^{-7} + 5.612 \times 10^{-7} \times 0.99^{(t-1991)})$ | 0.987 |
| | GDP | $y = 1/(5 \times 10^{-8} + 3.519 \times 10^{-6} \times 0.867^{(t-1991)})$ | 0.984 |
| | FAI | $y = 1/(5 \times 10^{-7} + 1.890 \times 10^{-5} \times 0.829^{(t-1991)})$ | 0.947 |
| | UPI | $y = 1/(1 \times 10^{-5} + 3.218 \times 10^{-4} \times 0.915^{(t-1991)})$ | 0.959 |
| | RPI | $y = 1/(1.25 \times 10^{-5} + 6.342 \times 10^{-4} \times 0.914^{(t-1991)})$ | 0.844 |

Note: Levels of significance of all equations are <0.001. *t* stands for year since 1992.

- Local update of pheromone concentration. At time *t*, ant *k* moves from digit *a* to *b* and leaves an amount of $\Delta\tau$ pheromone on path *ab*. Meanwhile, the pheromone concentration at time *t* - 1 of *ab* evaporates at a ratio of ρ (value range [0, 1]), so pheromone τ_{ab}^t on *ab* at time *t* can be expressed as

$$\tau_{ab}^t = (1 - \rho)\tau_{ab}^{t-1} + \Delta\tau. \tag{2}$$

- Rule of decision making for movement. The greater the pheromone concentration on a path, the greater the probability the path will be picked, which makes the ants converge to a current local optimal path quickly from the beginning of their search. To explore more paths, a random method is adopted to decide whether the path with the highest concentration is selected. This rule is formularized as

$$b = \begin{cases} \max(\tau_{ab}), & \text{if}(\text{randnum} \leq \lambda) \\ p_{ab} = \tau_{ab} / \sum_{i=0}^9 \tau_{ai}, & \text{else} \end{cases}, \tag{3}$$

where randnum and λ are uniform random numbers ranging from 0 to 1, and *i* represents the possible digit to be picked.

- Global update of pheromone concentration. Global update occurs when a search of all ants ends, and it only updates the pheromone concentration of the current best path of this search. The global update rule and the strategy for movement together ensure that the ants implement a sufficient number of searches around the current optimal path, largely accelerating the optimization speed. The global update rule is expressed as

$$\tau_{ab} = \alpha \cdot \tau_{ab} + (1 - \alpha) / \text{leesalee}_{\max}, \tag{4}$$

where α is the global evaporation ratio, a constant ranging from 0 to 1, and leesalee_{\max} is the calibration accuracy of the current optimal path.

The parameter initialization of ACO depends on the specific problem to be solved and has a large influence on ACO performance. The greater the number of ants, the greater the probability of obtaining the shortest path and the more time is required; the higher or lower the local evaporation ratio ρ , the poorer the efficiency to converge, which is the same condition for global evaporation ratio α ; and the greater the probability of random decision-making λ , the faster the convergence speed. In this study, the total number of one group of ants was set to 40; $\rho = 0.2$; $\lambda = 0.6$; $\alpha = 0.2$; and pheromone increment $\Delta\tau = 0.01$.

3.3.2 Introduction of subregional calibration

SLEUTH uses only one set of parameters for a whole study area, which may be suitable for cities with a single core. For multicore cities, however, in which each core region may have different natural conditions and macro driving factors, the subregions have distinct urban growth rates. The aim of subregional calibration is to produce multiple sets of parameters for multicore cities, making the calibration of SLEUTH zone-based. It is also important to introduce a zoning mechanism for single-core cities, since there exists spatial heterogeneity in their urban growth rates as well.

The zoning mechanism used here does not actually divide an entire study area into several subareas and then run SLEUTH subarea by subarea. Specifically, the zoning was achieved by addition of a new input data, the subarea layer (Fig. 2). Multiple sets of optimal parameters for different subregions were obtained by having ACO employ multiple groups of ants, meaning that one group was in charge of one subregion.

The study area was divided into four subregions according to administrative boundaries (Fig. 2): urban center (UC), Fengrun District (FRD), Fengnan District (FND), and suburban area (referred to as SA, which is the sum of Luan County and Luannan County).

3.3.3 Modification of the self-modification rule of SLEUTH

The self-modification rule of SLEUTH calls for the multipliers of BOOM and BUST to be constant, so the modification of this rule occurs by making constants into variables. The modified self-modification rule can be described with pseudocode as follows.

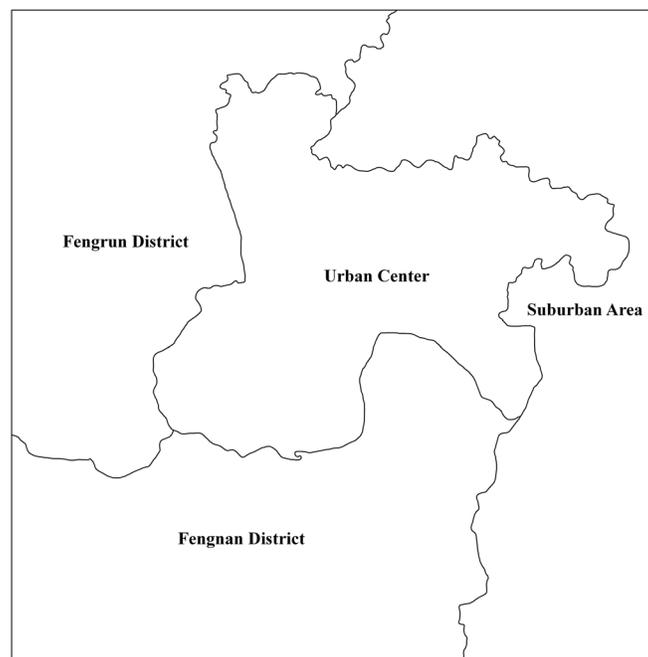


Fig. 2 Regional division of Tangshan study area.

```

X = GDP_P * FAI_P * INCOME_RATIO * GROWTH_RATE/10
IF(GROWTH_RATE > CRITICAL_HIGH)
  X = X/CRITICAL_HIGH
  BOOM = 1 + C1 * (1 - exp(-X))/(1 + exp(-X))
  Coeff = Coeff * BOOM
IF(GROWTH_RATE < CRITICAL_LOW)
  X = X/CRITICAL_LOW
  BUST = 1 - C2 * (1 - exp(-X))/(1 + exp(-X))
  Coeff = Coeff * BUST

```

In the above code, GDP_P, FAI_P and INCOME_RATIO are macro factors (introduced in Sec. 3.2.2); GROWTH_RATE is the current urban growth rate of a region; and Coeff represents the SLEUTH parameters. Influences on urban growth stemming from macro factors are based on the assumption that an urban growth rate for the next year is positively correlated with the GDP_P, FAI_P, and INCOME_RATIO of the current year. The product of the macro factors is mapped into the value range [0, 1] by a hyperbolic function, and that constants C_1 and C_2 , ranging from 0 to 1, limit the maximum value of BOOM and the minimum value of BUST, respectively. In this study, taking the default value of BOOM 1.01 and BUST 0.09 as base values, C_1 and C_2 were determined by means of trial and error: Taking C_1 as an example, C_2 was fixed first, then C_1 varied between its base value and some effective upper bound, e.g., 0.05. Every time a better result was obtained than the last time, C_1 increased slightly; otherwise, C_1 decreased. Specifically, C_1 was set to 0.03, and C_2 was set to 0.91, that is, $BOOM_{max} = 1.03$ and $BUST_{min} = 0.09$.

The hyperbolic function is formularized as follows:

$$f(x) = (1 - e^{-x})/(1 + e^{-x}). \quad (5)$$

This function monotonically increases in domain $[0, +\infty]$. When $x = 7$, the function value becomes very close to 1. Thus, in the above code, x is divided by 10 to expand the convergence range of the hyperbolic function.

3.4 Scenario Predictions to 2020

Scenario prediction takes into account a range of factors that restrict future urban development, and it predicts the corresponding urban morphologies that result from those diverse backgrounds. Specific influencing factors include resources (water, land), policies, topography, geological environment, ecosystem protection, and urban planning.

Two different scenarios were developed for Tangshan. The first scenario, the inertia trend scenario, assumed no change in the exclusion layer. The second scenario, the policy-adjusted scenario, took account of another three constrained factors for Tangshan: the reservoir's protected buffer zone, a polluted industry and mining area that had little appeal for urbanization, and a fragile area around the 1976 earthquake center. Changes were added to the exclusion layer to reflect these factors. For the modified exclusion layer, the middle and northern parts of a 5-km buffer zone around the largest reservoir of Tangshan was given a value of 50 considering the reservoir's great importance as a clean water source for the city. The contaminated area and a 10-km buffer zone around the historic earthquake center were both set to 20 due to their equivalent lower appeal for urbanization. A comparison between the original and modified exclusion layer can be seen in Fig. 3.

Two scenarios were run on two improved versions of SLEUTH: one improved by ACO and subregional calibration (referred to as AS), the other improved by ACO, subregional calibration, and the introduction of macro factors (referred to as ASM), making a total of four scenario prediction results for Tangshan.

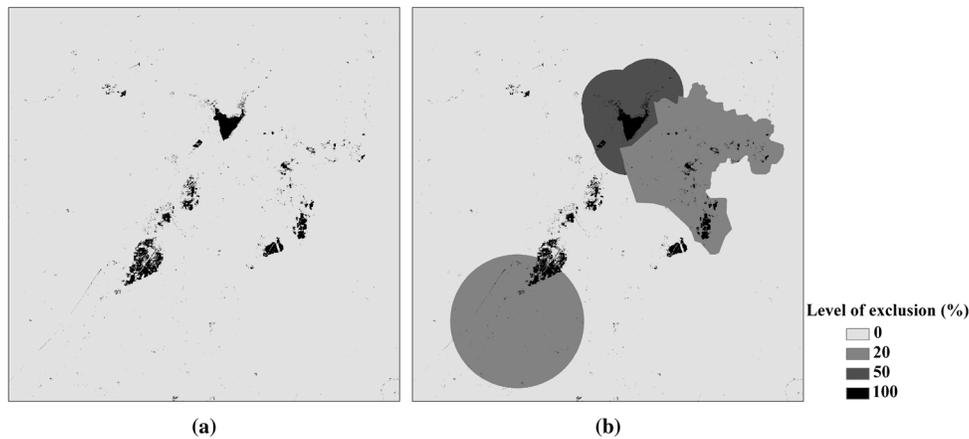


Fig. 3 Exclusion layers for two prediction scenarios. (a) Exclusion layer for the inertia scenario. (b) Exclusion layer for the policy-adjusted scenario.

4 Results

4.1 Performance of Calibration Algorithms

4.1.1 Brute-force calibration

The calibration details of the brute-force method are illustrated in Table 2. As shown in the table, leesalee is improved level by level, and the final accuracy is 0.47491. In the resulting optimum set of parameters, diffusion has a low value of 1, indicating a small contribution of the spontaneous rule to urban growth. The three-level calibration took approximately 14 days and 9 h. Clearly, the brute-force method is extremely time-consuming.

The urban forms of the control years were simulated by initializing SLEUTH using the optimal set of parameters described above, with the urban extent of the starting year 1992 as the basis. Statistics of the simulation results for the four growth rules can be seen in Tables 3 and 4. In Table 3, the simulated pop is slightly greater than that of the observation for each control year, and the derived statistics are edges- $r^2 = 0.315$, clusters- $r^2 = 0.983$, and pop- $r^2 = 0.975$. Although edges- r^2 is low, the absolute difference of the edges between the simulation and observation is not that notable compared with the total, at 7456, 6684, and 10886 for 1999, 2005, and 2009, respectively. Table 4 shows that the edge growth rule dominates SLEUTH, which is in accordance with the findings of Jantz and Goetz.²⁵

4.1.2 ACO calibration

As seen in Table 5, the calibration accuracy leesalee is 0.47547, which is slightly better than that calibrated by the brute-force method in Table 2. The time cost for ACO is approximately 14 days and 21 h, suggesting no efficiency improvement. An in-depth analysis of the ACO performance is given in Sec. 4.1.3. For the resulting five optimal parameters, diffusion and spread are similar to the results in Table 2. The high values for breed and road gravity reveal the enhanced contributions of new spreading-center growth and road-influenced growth; the low value of the slope demonstrates a small slope resistance for urban growth.

The simulation results for the control years were generated by SLEUTH with initialization of the best parameters (Table 5). Table 3 shows the evaluating indices of simulation results. The pop index with ACO is slightly greater than that with the brute-force method for each year. The derived statistics are edge- $r^2 = 0.316$, clusters- $r^2 = 0.983$, and pop- $r^2 = 0.978$.

4.1.3 Comparison between ACO and brute force

For calibration accuracy, ACO is slightly better than the brute-force method, as shown in Tables 2 and 5. In fact, ACO also outperforms brute force in terms of calibration efficiency.

Table 2 Three-level calibration and detailed information for the brute-force method.

| Parameter | Coarse calibrate | Fine calibrate | Final calibrate | optimum |
|------------------------|-------------------|-------------------|--------------------|---------|
| Monte Carlo iterations | 5 | 8 | 9 | |
| Combination No. | 3125 | 4500 | 5400 | |
| leesalee | 0.47343 | 0.47484 | 0.47491 | |
| Time | 2d:10 h:32 m:59 s | 5d:01 h:37 m:02 s | 06d:21 h:18 m:28 s | |
| Diffusion | | | | 1 |
| Range | 0–100 | 0–20 | 1–5 | |
| Step | 25 | 5 | 1 | |
| Breed | | | | 40 |
| Range | 0–100 | 25–75 | 30–55 | |
| Step | 25 | 10 | 5 | |
| Spread | | | | 20 |
| Range | 0–100 | 15–35 | 18–23 | |
| Step | 25 | 5 | 1 | |
| Slope | | | | 39 |
| Range | 0–100 | 40–60 | 38–42 | |
| Step | 25 | 5 | 1 | |
| Road gravity | | | | 50 |
| Range | 0–100 | 25–75 | 45–70 | |
| Step | 25 | 10 | 5 | |

Note: Time means time taken by one calibration phase.

This better performance is demonstrated in two ways. First, Figs. 4 and 5 show that the five parameters and the consequent leesalee converged to their limits at the 15th of a total of 100 runs; that is, it took approximately 2 days and 6 h to obtain the final parameters by ACO. Second, the time cost of the subregional calibration by ACO, shown in Table 6, is less than 2 days and 4 h, also proving the advantage of ACO.

4.2 Subregional Calibration

The calibration results from the zoning mechanism are illustrated in Table 6. As shown in the table, each subregion has distinct parameters and acquires different calibration accuracies. The leesalee for the whole study area is 0.48067, which is a slight improvement compared with Table 5. The total calibration time taken with 60 runs is just under 2 days and 4 h.

The convergence process of the parameters and leesalees of the four subregions can be seen in Fig. 6. The figure shows that, with the exception of UC, the parameters of the subregions converge to limits in 60 runs.

Simulated urban forms of the control years were also produced by the improved SLEUTH, with parameter initialization for four subregions. Table 3 (subregional part) shows the simulation result statistics. Compared with ACO, the pop index is somewhat greater for each year than that shown in ACO; the leesalee is also greater than or equal to the leesalee in ACO. The derived statistics are $\text{edge-}r^2 = 0.200$, $\text{clusters-}r^2 = 0.978$, and $\text{pop-}r^2 = 0.980$.

Table 3 Morphological indices and calibration accuracy of the simulation results and corresponding indices of the observations of the control years.

| Metric | Brute force | | | ACO | | | Subregional | | | Modified self-modification | | | Observation | | |
|----------|-------------|--------|--------|--------|--------|--------|-------------|--------|--------|----------------------------|--------|--------|-------------|--------|--------|
| | 1999 | 2005 | 2009 | 1999 | 2005 | 2009 | 1999 | 2005 | 2009 | 1999 | 2005 | 2009 | 1999 | 2005 | 2009 |
| Edges | 322702 | 333009 | 336477 | 324298 | 334774 | 338316 | 320291 | 325970 | 326755 | 321589 | 328066 | 329508 | 330158 | 326325 | 347363 |
| Clusters | 42139 | 38150 | 35469 | 41954 | 37923 | 35230 | 41695 | 37538 | 34950 | 41781 | 37800 | 35371 | 49832 | 36665 | 22712 |
| pop | 705847 | 849257 | 954552 | 716224 | 864005 | 969110 | 724089 | 874758 | 978753 | 728707 | 877498 | 978169 | 598708 | 811449 | 895206 |
| leesalee | 0.48 | 0.48 | 0.47 | 0.47 | 0.48 | 0.47 | 0.47 | 0.49 | 0.48 | 0.47 | 0.49 | 0.48 | 0.48 | 0.49 | 0.48 |

Note: Brute force for Sec. 4.1.1, ACO for Sec. 4.1.2, subregional for Sec. 4.2, modified self-modification for Sec. 4.3.

Table 4 Simulation results using the four growth rules for the control years.

| Year | Spontaneous growth | New spreading-center growth | Edge growth | Road-influenced growth |
|------|--------------------|-----------------------------|-------------|------------------------|
| 1999 | 10 | 9 | 22416 | 35 |
| 2005 | 11 | 10 | 25021 | 41 |
| 2009 | 12 | 11 | 27246 | 33 |

Table 5 Calibration results and run information of ACO.

| Parameter | Optimum | ACO information |
|--------------|---------|-----------------------------|
| Diffusion | 1 | Monte Carlo iterations = 10 |
| Breed | 88 | Run no. = 100 |
| Spread | 19 | Ants population = 40 |
| Slope | 1 | leesalee = 0.47547 |
| Road gravity | 65 | time = 14d:21h:11m:38s |

Note: Run no. means the number of total searches executed by a group of ants.

4.3 Calibration of Modified Self-Modification Rule

Another calibration experiment was conducted using the modified self-modification rule with macro factors. Table 7 shows the calibration result. As is seen from the table, UC maintains the greatest spread value, in line with Table 6, but all subregional leesalees, except FND, are less than what is shown in Table 6. Consequently, the leesalee for the entire study area is not improved as expected. The accuracy difference is not significant, however, suggesting a need for further experiments with parameters BOOM and BUST to achieve confidence regarding the modified self-modification rule.

Urban extent for the control years was simulated using the additional improvements to SLEUTH, and the statistical metrics for the simulation results are shown in Table 3 (modified self-modification part). As shown, the pop index in the self-modification part is approximately equal to that in the subregional part for each year, and the leesalee is the same as that shown in

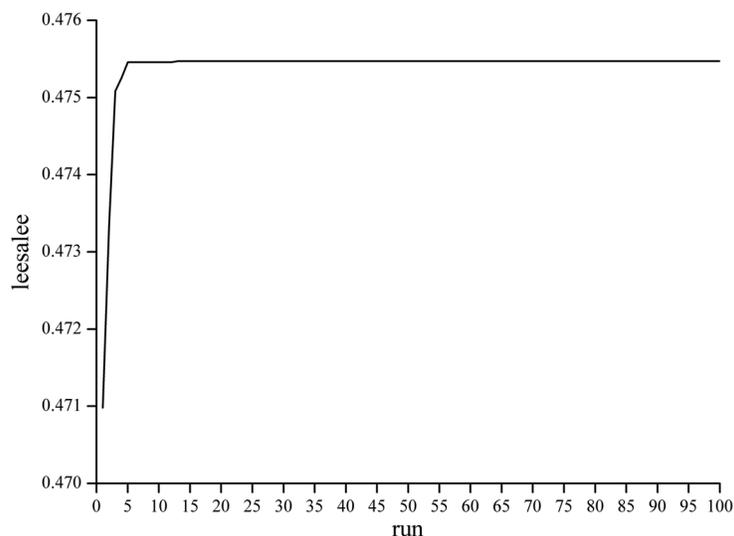


Fig. 4 Variation of calibration accuracy, leesalee index with ACO.

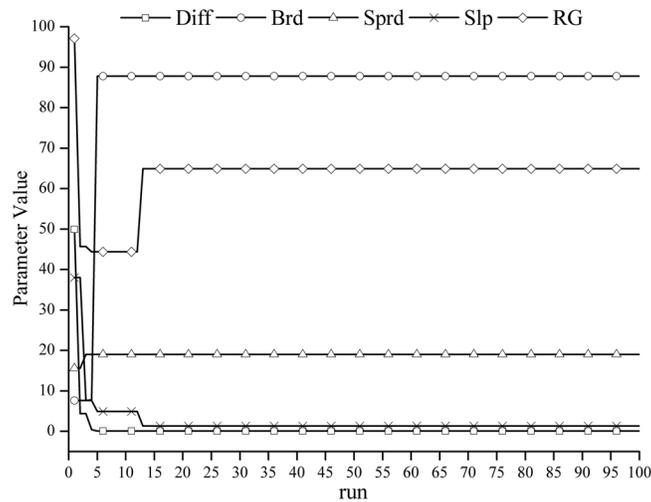


Fig. 5 Variations of the five model parameters with ACO calibration. Diff, Brd, Sprd, Slp, and RG are the abbreviations of diffusion, breed, spread, slope, and road gravity, respectively.

Table 6 Subregional calibration results by ACO and run information of ACO.

| ACO information | | | | |
|-----------------------------|--------------|---------------|------------------------|-----------------------|
| Monte Carlo iterations = 10 | Run no. = 60 | Ants pop = 40 | Time = 02d:03h:51m:42s | leesaleeall = 0.48067 |
| Subregion | Parameter | Optimum | leesalee | |
| SA | Diffusion | 1 | 0.38260 | |
| | Breed | 26 | | |
| | Spread | 15 | | |
| | Slope | 24 | | |
| | Road gravity | 44 | | |
| UC | Diffusion | 1 | 0.567833 | |
| | Breed | 16 | | |
| | Spread | 29 | | |
| | Slope | 2 | | |
| | Road gravity | 10 | | |
| FRD | Diffusion | 2 | 0.495857 | |
| | Breed | 6 | | |
| | Spread | 21 | | |
| | Slope | 28 | | |
| | Road gravity | 36 | | |
| FND | Diffusion | 1 | 0.396705 | |
| | Breed | 27 | | |
| | Spread | 18 | | |
| | Slope | 42 | | |
| | Road gravity | 22 | | |

Note: leesaleeall is leesalee for the whole study area.

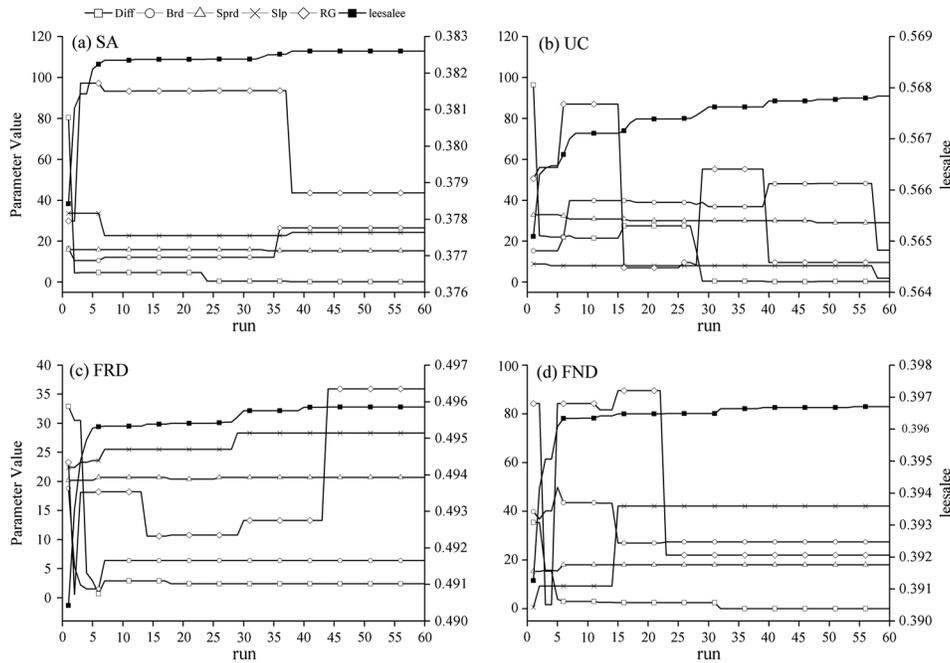


Fig. 6 Variations of the five model parameters and leesalee for the four subregions. Diff, Brd, Sprd, Slp, and RG are the abbreviations of diffusion, breed, spread, slope, and road gravity, respectively.

the subregional part. The derived statistics are edge- $r^2 = 0.250$, clusters- $r^2 = 0.976$, and pop- $r^2 = 0.982$.

A comparison between the simulated and corresponding observed urban extent is illustrated in Fig. 7. The red color in Fig. 7 represents area in which SLEUTH failed to simulate the actual growth. Green represents the common set of the observed and simulated urban extent. Blue represents commission errors by SLEUTH. To assess the simulation accuracy, the traditional image classification accuracy metric was adopted: the resultant overall accuracy is 75.3%, 73.7%, and 75.2% for 1999, 2005, and 2009, respectively, and the corresponding κ coefficients are 0.53, 0.49, and 0.52. The figure shows that the simulated result of each control year has a more compact internal form than the observed. In addition, the simulations of 2005 and 2009 maintain the shape of 1999, with infill at the urban center and no directional differences in urban growth. These SLEUTH simulation results are largely determined by the edge-growth rule.

4.4 Urban Growth Forecasts for Tangshan to 2020

4.4.1 Forecasts of inertia trend scenario

The inertia trend scenario was run on two versions of the improved SLEUTH, AS and ASM, to compare their simulation results. Using the urban extent of 2009 as the basis, forecasts from 2010 to 2020 were generated by the two improved SLEUTHs. The simulated urban forms of 2015 and 2020 were used to establish a comparison, as shown in Table 8 (inertia scenario part). As seen in the table, the urban area predicted by AS for 2015 will be 1.19 times that for 2009, and that of 2020 will be 1.35 times that of 2009. Correspondingly, for ASM, 2015 and 2020 will be 1.20 and 1.38 times, respectively. The introduction of macro factors produces a slight increase in urban growth.

Comparisons of urban growth from 2010 to 2020 can be seen in Fig. 8. The figure shows that for all prediction years the annual urban growth rate is higher for ASM than AS, leading to a greater slope for ASM pop. At the same time, the urban growth rate for both ASM and AS decreases from 2010 to 2020. Despite the incorporation of macro factors for ASM, the predicted urban growth maintains a linear trend.

Table 7 Calibration results with modified self-modification rule and run information.

| ACO information | | | | |
|-----------------------------|--------------|---------------|------------------------|-----------------------|
| Monte Carlo iterations = 10 | Run no. = 60 | Ants pop = 40 | Time = 02d:04h:36m:39s | leesaleeall = 0.48067 |
| Subregion | Parameter | Optimum | leesalee | |
| SA | Diffusion | 3 | 0.382261 | |
| | Breed | 1 | | |
| | Spread | 16 | | |
| | Slope | 19 | | |
| UC | Road gravity | 39 | 0.567622 | |
| | Diffusion | 14 | | |
| | Breed | 34 | | |
| | Spread | 30 | | |
| FRD | Slope | 5 | 0.495328 | |
| | Road gravity | 23 | | |
| | Diffusion | 2 | | |
| | Breed | 10 | | |
| FND | Spread | 20 | 0.396723 | |
| | Slope | 11 | | |
| | Road gravity | 20 | | |
| | Diffusion | 1 | | |
| | Breed | 10 | | |
| | Spread | 19 | | |
| | Slope | 33 | | |
| | Road gravity | 70 | | |

4.4.2 Forecasts of the policy-adjusted scenario

Two types of urban growth forecasts were produced by AS and ASM. Table 8 (policy-adjusted scenario part) shows the statistical metrics of the simulation for 2015 and 2020. The pop index in the policy-adjusted scenario is slightly lower than that in the inertia scenario for both reference years, indicating an insignificant influence on urban growth caused by the policies tested in this study.

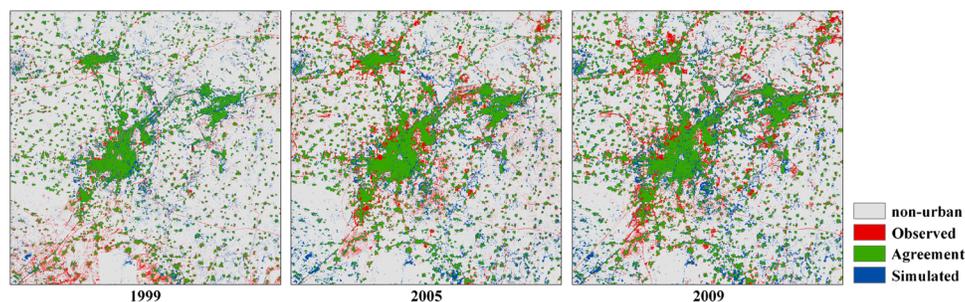


Fig. 7 Comparison between simulated and observed growth of Tangshan for control years 1999, 2005, and 2009.

Table 8 Morphological indices of simulated urban forms of 2015 and 2020.

| Metric | Inertia scenario | | | | Policy-adjusted scenario | | | |
|----------|------------------|---------|---------|---------|--------------------------|---------|---------|---------|
| | AS | | ASM | | AS | | ASM | |
| | 2015 | 2020 | 2015 | 2020 | 2015 | 2020 | 2015 | 2020 |
| Edges | 359395 | 357551 | 359303 | 354546 | 359257 | 359438 | 359430 | 357423 |
| Clusters | 26582 | 26696 | 25115 | 23504 | 26748 | 27098 | 25406 | 24085 |
| Pop | 1065990 | 1209481 | 1077347 | 1238644 | 1052662 | 1186125 | 1062618 | 1212749 |

Note: Inertia scenario for Sec. 4.4.1, policy-adjusted scenario for Sec. 4.4.2.

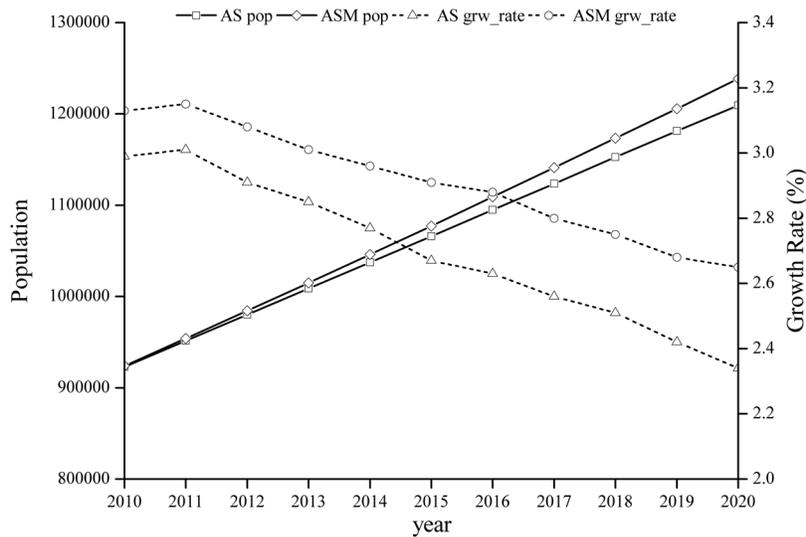


Fig. 8 Comparisons of urban growth of Tangshan from 2010 to 2020 for the inertia scenario, where pop is the population of urban pixels, and grw_rate is the annual growth rate of urban area.

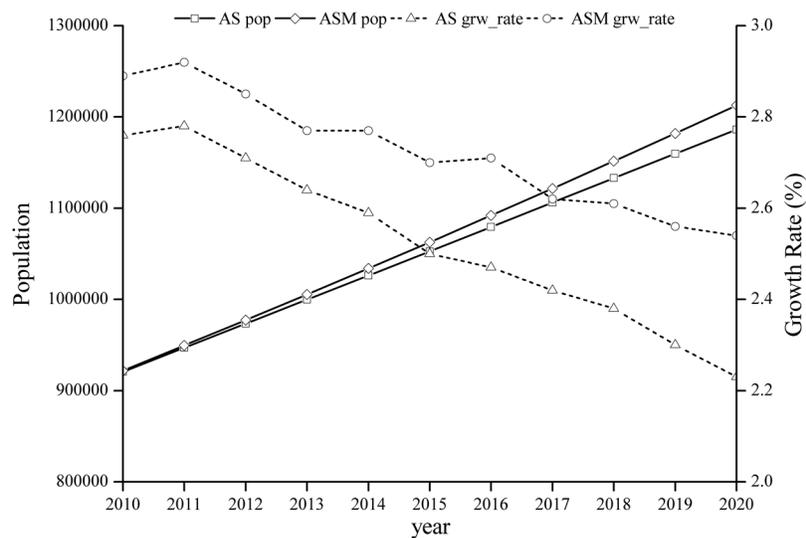


Fig. 9 Comparisons of urban growth of Tangshan from 2010 to 2020 for the policy-adjusted scenario, where pop is the population of urban pixels, and grw_rate is the annual growth rate of urban area.

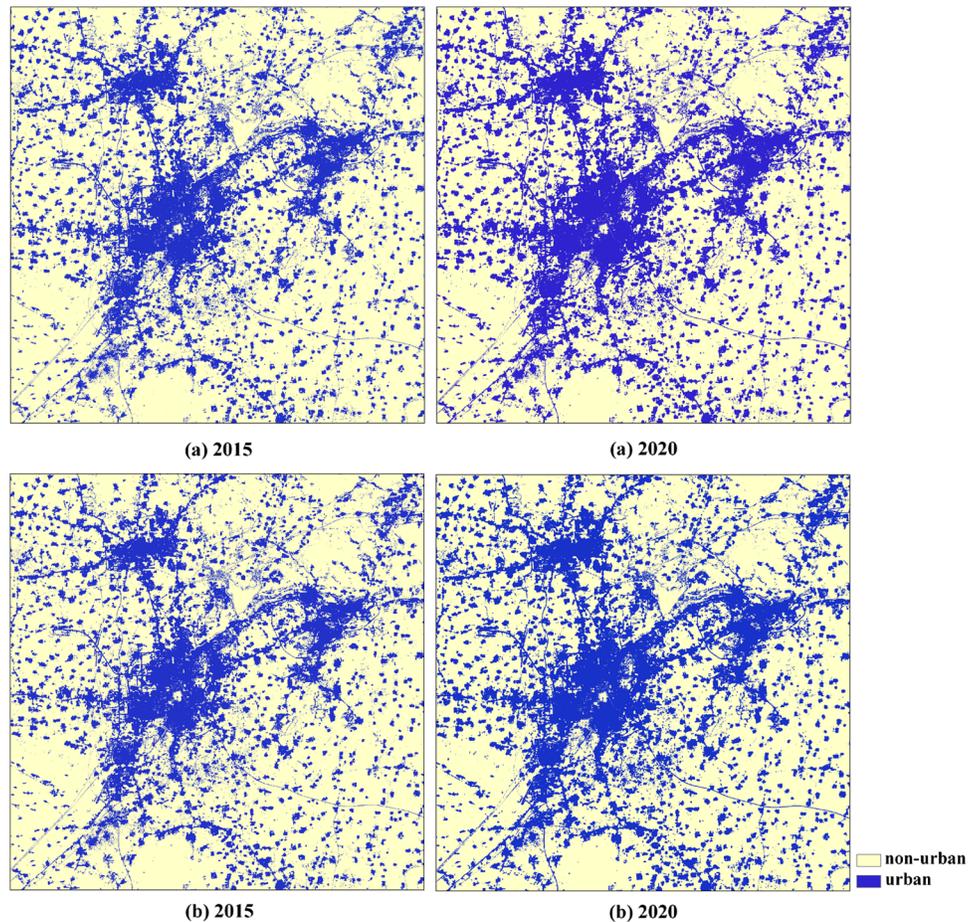


Fig. 10 Forecasts of urban extent for 2015 and 2020 under the policy-adjusted scenario. (a) Forecasts by AS SLEUTH. (b) Forecasts by ASM SLEUTH.

Detailed results of the two series of forecasts are shown in Fig. 9. There are no significant differences between Figs. 8 and 9. Urban extents for 2015 and 2020 as predicted by AS and ASM are shown in Fig. 10.

5 Discussion

This study demonstrates improvements to SLEUTH, but many important questions remain that require further study. Generally, these questions can be summarized as calibration efficiency, accuracy improvement, sensitivity analysis of parameters, and methodologies for introducing macro factors.

5.1 Calibration Efficiency

Two aspects require further study: the calibration algorithms and the model rules. This study tested the former for SLEUTH. Jantz et al.¹² attempted an alternate method to improve the road-search algorithm for the road-influenced growth rule of SLEUTH, resulting in a significant efficiency improvement for calibration. In terms of the algorithms, the ACO employed could be further improved, specifically regarding ways to achieve massively parallel processing.

5.2 Accuracy Improvement

Model rules contribute significantly to model performance. The simulation produced by SLEUTH is characterized by edge growth and infill growth, which makes the city morphology

more compact. These characteristics are determined by the edge-growth rule of SLEUTH. Suggested improvements to this rule should address two issues: how to control compact growth and how to reflect the direction difference of growth; an alternative solution for the latter is to take spatial autocorrelation into consideration.

5.3 Sensitivity Analysis of Parameters

In addition to the five basic parameters for controlling growth rules, SLEUTH has several other parameters, such as the upper bound of the urban growth rate (CRITICAL_HIGH), the lower bound of the urban growth rate (CRITICAL_LOW), the influence threshold of slope (CRITICAL_SLOPE), and BOOM and BUST in the self-modification rule. These additional parameters also contribute to the model accuracy, and the significance of their influence and the determination of a method to calibrate them need to be addressed.

5.4 Approaches to Incorporating Macro Factors

At present, there are two methods for combining urban growth models with macro factors. In the first, urban models are loosely coupled with models of macro information, establishing the quantitative relationship between the amount of urban growth and the macro factors (the early urban models of system dynamics);^{26,27} in the second, macro factors are spatialized to become inputs of urban CA models.²⁸ The first method involves a one-way interaction between macro factors and urban growth, where macro factors control urban growth. In fact, urban growth counteracts macro factors, thus creating the need to establish feedback mechanisms between the factors and urban growth. The second method is challenging because of the limited availability of high-resolution macro data, which means that spatialized macro inputs are usually calculated by a spatial distribution model, and the resultant macro layers generally have the feature of spatial symmetry. For this reason, the introduction of isotropic macro factors cannot provide valuable changes to the CA simulation results.

Aside from the general questions discussed above, four issues remain. First, for subdivisions of the Tangshan study area produced by administrative boundaries, it is preferable to divide the area along borders with gravity equilibrium among different urban cores, according to the law of universal gravitation. The term mass of the law can be replaced by an economic indicator, such as regional GDP. Second, for the modified self-modification rule, on one hand, SLEUTH is more sensitive to the adjustment of C_1 when simulating the urbanization of Tangshan because the growth rate is maintained at a high level, leading to BOOM growing faster or maintaining its maximum, which gives positive feedback to the growth rate in return. The higher the BOOM value, the larger the C_1 and the poorer the simulation accuracy will be. The same condition may hold true for C_2 if simulation growth rate remains low. The issue is when to meet the condition to transform the growth state from BOOM to BUST. On the other hand, the trial-and-error selection of constants C_1 and C_2 is complex, so automated calibration would be a good option for determining them. Third, again related to the modified self-modification rule, because the relationship between the macro factors and model parameters was bridged by a hyperbolic function, it is likely that this function did not match the practical situation of the urban growth trend due to the function's fast convergence rate. Other mapping functions should be the subject of further study. Fourth, the influence of the scenario prediction strategy on urban simulation in SLEUTH is static. Different exclusion layer inputs play the role of scenarios whose ultimate goal is to decrease the amount of nonurban land. If the percent ratio between the current available nonurban land and urban land is far beyond CRITICAL_HIGH, then exclusion will have no impact on the simulation regardless of any exclusion values. If the percent ratio is lower than CRITICAL_LOW, the simulation will enter the BUST or stagnation state; if neither, the simulation state may first be intermediate between BOOM and BUST and then BUST, or it may keep BUST.

6 Conclusion

This study is primarily devoted to the improvements of the SLEUTH model. The city of Tangshan, China, a rapidly urbanizing region, is chosen as the study area for the model

experiments. Scenario forecasts of urban growth to 2020 are also performed. The conclusions are as follows:

1. ACO outperforms the brute-force method for SLEUTH calibration in both efficiency and accuracy. ACO is equal to or faster than coarse calibration, and it achieves more accurate parameters, improving the calibration efficiency and accuracy of the SLEUTH model.
2. The introduction of a zoning mechanism is a further improvement for the SLEUTH calibration, according to the leesalee index. The urban growth amount of the subregional simulation, however, is slightly higher than that of the total area.
3. The incorporation of macro factors should theoretically improve the performance of SLEUTH, but in this study it fails to do so. The causes of this failure may be attributed to the mapping function, a hyperbolic function that converges rapidly. In addition, the urban growth rate decreases more slowly after modifying the self-modification rule with macro factors compared to the unmodified rule.
4. Constraint factors, such as environment, policy, terrain, and geological conditions, will not change the forecast trends of the SLEUTH model and can only produce a slight reduction of the urban growth amount.
5. The prediction results of ASM under inertia-trend scenario indicate that Tangshan will maintain a high annual rate of area growth, >2.6%, from 2010 to 2020 and that in 2015 and 2020, the urban area will be 1.20 and 1.38 times that of 2009, respectively.

Acknowledgments

This work is supported by the Natural Scientific Foundation of China (Grant No. 4097122, 40771011), the Research and Development Program of the Institute of Geology, China Earthquake Administration (No. IGCEA0903), and project E0202/1112/0102. The AST14DEM data were obtained through the online Data Pool at the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD (https://lpdaac.usgs.gov/get_data). We also thank two anonymous reviewers for their valuable comments and suggestions that improved this paper.

References

1. W. R. Tobler, "Cellular Geography," in *Philosophy in Geography*, S. Gale and G. Olsson, Eds., pp. 379–386, D. Reidel, Dordrecht, Holland (1979).
2. H. Couclelis, "Cellular worlds, a framework for modelling micro-macro dynamics," *Environ. Plann. A* **17**(5), 585–596 (1985), <http://dx.doi.org/10.1068/a170585>.
3. H. Couclelis, "Macrostructure and microbehavior in a metropolitan area," *Environ. Plann. B Plann. Des.* **16**(2), 141–154 (1989), <http://dx.doi.org/10.1068/b160141>.
4. M. Batty and Y. Xie, "Modeling inside GIS: part 1: model structures, exploratory spatial analysis," *Int. J. Geogr. Inform. Syst.* **8**(3), 291–307 (1994), <http://dx.doi.org/10.1080/02693799408902001>.
5. M. Batty and Y. Xie, "Modeling inside GIS: part 2: selecting and calibrating urban models using ARC/INFO," *Int. J. Geogr. Inform. Syst.* **8**(5), 451–470 (1994), <http://dx.doi.org/10.1080/02693799408902013>.
6. K. C. Clarke, S. Hoppen, and L. Gaydos, "A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area," *Environ. Plann. B Plann. Des.* **24**(2), 247–261 (1997), <http://dx.doi.org/10.1068/b240247>.
7. K. C. Clarke and L. Gaydos, "Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore," *Int. J. Geogr. Inform. Sci.* **12**(7), 699–714 (1998), <http://dx.doi.org/10.1080/136588198241617>.
8. B. M. Deal and Z. Sun, "A spatially explicit urban simulation model: landuse evolution and impact assessment model (LEAM)," in *Smart Growth and Climate Change: Regional Development, Infrastructure and Adaptation*, M. Ruth, Ed., pp. 181–203, Edward Elgar, Northampton, Massachusetts (2006).

9. S. Berling-Wolff and J. Wu, "Modeling urban landscape dynamics: a case study in Phoenix, USA," *Urb. Ecosyst.* **7**(3), 215–240 (2004), <http://dx.doi.org/10.1023/B:UECO.0000044037.23965.45>.
10. M. Cagliioni, M. Pelizzoni, and G. A. Rabino, "Urban sprawl: a case study for project Gigalopolis using SLEUTH model," in *Cellular Automata, 7th International Conf. on Cellular Automata, for Research and Industry, Proceedings. Lecture Notes in Computer Science 4173 (ACRI 2006, LNCS 4173)*, S. El Yacoubi, B. Chopard, and S. Bandini, Eds., pp. 436–445, Springer-Verlag, Berlin Heidelberg (2006).
11. P. R. Claggett et al., "Assessing development pressure in the Chesapeake Bay Watershed: an evaluation of two land-use change models," *Environ. Monit. Assess.* **94**(1–3), 129–146 (2004), <http://dx.doi.org/10.1023/B:EMAS.0000016884.96098.77>.
12. C. A. Jantz et al., "Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model," *Comput. Environ. Urban Syst.* **34**(1), 1–16 (2010), <http://dx.doi.org/10.1016/j.compenvurbsys.2009.08.003>.
13. R. Rafiee et al., "Simulating urban growth in Mashad City, Iran through the SLEUTH model (UGM)," *Cities* **26**(1), 19–26 (2009), <http://dx.doi.org/10.1016/j.cities.2008.11.005>.
14. E. A. Silva and K. C. Clarke, "Calibration of the SLEUTH urban growth model for Lisbon and Porto, Portugal," *Comput. Environ. Urban Syst.* **26**(6), 525–552 (2002), [http://dx.doi.org/10.1016/S0198-9715\(01\)00014-X](http://dx.doi.org/10.1016/S0198-9715(01)00014-X).
15. A. D. Syphard, K. C. Clarke, and J. Franklin, "Using a cellular automaton model to forecast the effects of urban growth on habitat pattern in southern California," *Ecol. Complex.* **2**(2), 185–203 (2005), <http://dx.doi.org/10.1016/j.ecocom.2004.11.003>.
16. X. Wu et al., "Performance evaluation of the SLEUTH model in the Shenyang metropolitan area of Northeastern China," *Environ. Model. Assess.* **14**(2), 221–230 (2009), <http://dx.doi.org/10.1007/s10666-008-9154-6>.
17. G. Xian and M. Crane, "Assessments of urban growth in the Tampa Bay watershed using remote sensing data," *Remote Sens. Environ.* **97**(2), 203–215 (2005), <http://dx.doi.org/10.1016/j.rse.2005.04.017>.
18. M. Batty and Y. Xie, "Possible urban automata," *Environ. Plann. B Plann. Des.* **24**(2), 175–192 (1997), <http://dx.doi.org/10.1068/b240175>.
19. Hebei Surveying & Mapping Bureau, *Transportation Map of Tangshan*, SinoMaps Press, Beijing (1991).
20. Cartography Institute of Hebei, *Revised Tangshan Atlas*, Xi'an Cartography Press, Shanxi (2007).
21. X. Feng, R. Li, and H. Mo, "Study of urban expansion and driving factor using RS and GIS: a case study of Xi'an City (in Chinese)," *Remote Sens. Tech. Appl.* **25**(2), 202–208 (2010).
22. J. W. Forrester, *Urban Dynamics*, MIT Press, Cambridge, MA, pp. 14–20 (1969).
23. M. Dorigo and G. Di Caro, "Ant algorithms for discrete optimization," *Artif. Life* **5**(2), 137–172 (1999), <http://dx.doi.org/10.1162/106454699568728>.
24. Y. Chen, "Ant colony system for continuous function optimization (in Chinese)," *J. Sich. Univ. (Eng. Sci. Ed.)* **36**(6), 117–120 (2004).
25. C. A. Jantz and S. J. Goetz, "Analysis of scale dependencies in an urban land-use-change model," *Int. J. Geogr. Inform. Sci.* **19**(2), 217–241 (2005), <http://dx.doi.org/10.1080/13658810410001713425>.
26. C. He et al., "Modeling urban expansion scenarios by coupling cellular automata model and system dynamic model in Beijing, China," *Appl. Geogr.* **26**(3), 323–345 (2006), <http://dx.doi.org/10.1016/j.apgeog.2006.09.006>.
27. C. He et al., "Modelling dynamic urban expansion processes incorporating a potential model with cellular automata," *Landsc. Urban Plann.* **86**(1), 79–91 (2008), <http://dx.doi.org/10.1016/j.landurbplan.2007.12.010>.
28. L. Li, Y. Sato, and H. Zhu, "Simulating spatial urban expansion based on a physical process," *Landsc. Urban Plann.* **64**(1–2), 67–76 (2003), [http://dx.doi.org/10.1016/S0169-2046\(02\)00201-3](http://dx.doi.org/10.1016/S0169-2046(02)00201-3).

Biographies and photographs of the authors not available