Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

Spatial cooperative simulation of land use-population-economy in the Greater Bay Area, China

Wei Tu, Wei Gao, Mingxiao Li, Yao Yao, Biao He, Zhengdong Huang, Jie Zhang & Renzhong Guo

To cite this article: Wei Tu, Wei Gao, Mingxiao Li, Yao Yao, Biao He, Zhengdong Huang, Jie Zhang & Renzhong Guo (27 Nov 2023): Spatial cooperative simulation of land use-population-economy in the Greater Bay Area, China, International Journal of Geographical Information Science, DOI: 10.1080/13658816.2023.2285459

To link to this article: https://doi.org/10.1080/13658816.2023.2285459

Published online: 27 Nov 2023.

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
https://doi.org/10.1080/13658816.2023.2285459

RESEARCH ARTICLE

Spatial cooperative simulation of land use-population-economy in the Greater Bay Area, China

Wei Tu[a,b,c,d], Wei Gao[a,b], Mingxiao Li[a,b], Yao Yao[d,e], Biao He[a,c,d], Zhengdong Huang[a,c,d], Jie Zhang[f] and Renzhong Guo[a,c,d]

[a]Shenzhen Key Laboratory of Spatial Smart Sensing and Service, Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China; [b]Ministry of Natural Resources (MNR) Key Laboratory for Geo-Environmental Monitoring of Great Bay Area, Shenzhen University, Shenzhen, China; [c]State Key Laboratory of Subtropical Building and Urban Science, Shenzhen, China; [d]Guangdong-Hong Kong-Macao Joint Laboratory for Smart Cities & Guangdong Key Laboratory of Urban Informatics, Shenzhen, China; [e]National Engineering Research Center of GIS, School of Geography and Information Engineering, China University of Geosciences, Wuhan, China; [f]National Innovation Center for Digital Agricultural Products Circulation, College of Information and Electrical Engineering, China Agricultural University, Beijing, China

ABSTRACT

Fast urbanization brings great challenges to sustainable development goals, such as excessive exploitation and population explosion. Classical cellular automata (CA) have been widely used to independently simulate the change of spatial features, i.e. land use, population, economic production, etc. However, most CA models rely on historical data as static driving factors to simulate future scenarios while ignoring the inter-wined influences among multiple features in the development process. To address this issue, this study proposes a spatial cooperative simulation (SCS) approach to simulate the land use, population, and economy changes. The SCS approach starts with a separate CA model to obtain the initial scenes of each feature. Then, the simulation results of each other two features are used as dynamically updated driving factors, rather than the static historical data, to capture the inter-wined influence of multiple features during the development process. This step is iteratively performed until the changes of each feature converge and the final simulation results will be reported. The simulation experiment in Greater Bay Area demonstrates that the SCS approach can well capture the simultaneous development process and outperforms baseline approaches. The SCS approach is capable of forecasting future development scenarios and facilitates spatial planning and infrastructure synergies.

ARTICLE HISTORY
Received 28 September 2022
Accepted 14 November 2023

KEYWORDS
Spatial simulation; cooperative influence; cellular automata; Greater Bay Area

1. Introduction

Cities are places gathering people and resources to provide well-living services, job opportunities, and scientific innovations (United Nations 2018). With worldwide

CONTACT Mingxiao Li limx@szu.edu.cn

urbanization, cities grow, connect with each other, and gradually merge into one mega-city region. In the past forty years, many mega-city regions have appeared, such as the greater London (Rydin *et al.* 2004), the greater Tokyo area (Du *et al.* 2018), and the Guangdong-Hong Kong-Macao Greater Bay Area (GBA) (Yeh and Chen 2020, Xu *et al.* 2021). These polarized urbanization processes have brought many challenges to land, population, economy, and ecology, such as limited developing land, insufficient housing, and environmental pollution (Shafizadeh-Moghadam *et al.* 2017, Liang and Yang 2019, Chen *et al.* 2022, Li *et al.* 2022). The GBA, including Hong Kong, Macao, Shenzhen, Guangzhou, and other seven cities in the Pearl River Delta, is cooperating to achieve high-quality sustainable development (Zhou *et al.* 2018, Shao *et al.* 2020, Weng *et al.* 2020). Policy analysis for regional coordinated development decisions highlights the consideration of many different inter-related features, including population, land use, transport, economy, and environment (White *et al.* 2015, Harvey *et al.* 2019, Usman *et al.* 2022, Zhang and Xia 2022). Spatial simulation considering the intertwined influences of multiple features is essential to understanding future development scenarios and supporting spatial policy-making (Arsanjani *et al.* 2013, Liu *et al.* 2017, Mustafa *et al.* 2018, Schwaab *et al.* 2018).

Spatial simulation has been widely used to forecast the varying geoprocessing by considering the interaction among transport, land use, natural resources, and human activities. Typical models include cellular automata (CA) (Benenson and Torrens 2004, Li *et al.* 2020, Jalayer *et al.* 2022), multi-agent system (MAS) (Ghavami *et al.* 2022, Lu *et al.* 2022), hybrid models (Bijandi *et al.* 2019, Jiang *et al.* 2022), etc. Especially, CA is a common cellular model to simulate spatial change by estimating the cell state, according to its initial state, neighborhood effects, and transition rules (Liu *et al.* 2017, Alaei Moghadam *et al.* 2018). Because of the ability to generate rich patterns and capture the spatially dynamic processes, CA-based simulation has been widely used to simulate spatial processes of land use and land cover (Shahfahad *et al.* 2022, Wang *et al.* 2022), urban growth (Alaei Moghadam *et al.* 2018), population dynamics (Sun *et al.* 2020, Crols *et al.* 2017), economic development (Yu *et al.* 2019, Liang *et al.* 2020), etc. As the most intuitive manifestation of urbanization, the simulation of urban growth and land use/land cover (LULC) has attracted much attention. For example, Du *et al.* (2018) combined the CA and tree-based machine learning methods to simulate multiple land use changes in the Greater Tokyo Area. Rahnama (2021) simulated land use change in Mashhad, Iran based on the CA Markov chain model to reveal the historical and future spatial-temporal conversion patterns of land use from 2016 to 2030. Geng *et al.* (2022) proposed the hybrid spatiotemporal convolution-based CA model named ST-CA, which used 3D-CNN to assimilate the latent information in spatiotemporal neighborhoods for accurate LULC simulation. These studies provide theoretical and practical insights for cooperative simulation of land use, population, and economy in megacity regions.

In terms of the simulation of population and economy, there are two main categories of studies generally. The first category of studies mainly models spatial change of population or economy from the top to the down. These studies simulate future population or economy volume with numerical computation methods or take the inner structure of population and economy into the projection (Harvey *et al.* 2019, Liang *et al.* 2020, Guzman *et al.* 2022). The spatial changes in population or economy volume are simulated, and the citywide population and economy are projected into grids with spatialization methods (Li *et al.* 2019, Georganos *et al.* 2021). These studies verified that economy and population are strongly related to natural characteristics (e.g. topographic, land surface temperature, etc.) and built environment (e.g. land cover, road network, etc.). However, these studies ignore the spatial diffusion and co-evolution characteristics in spatial neighbourhoods during the population growth process, thus reducing the simulation accuracy. To tackle this issue, the second category of studies mainly uses CA to simulate the population or economy from the bottom to the up. For example, Yu *et al.* (2019) presented a CA-Markov simulation model to estimate the green gross domestic production (GDP) by considering the influence of the ecological value, economic suitability, land use, and neighbourhood effect. Sun *et al.* (2020) presented a logistic probabilistic CA model to simulate regular pattern formation of population dynamics. While Guzman *et al.* (2022) estimate population density with simulated land use patterns from a CA-based model. These studies demonstrate that CA models can effectively simulate the complex change in population or economy.

The above advanced studies demonstrate the CA's capability to simulate different spatially varying features. They typically use historical economic, social, and transportation data as static driving factors to simulate future development scenarios while ignoring the dynamic synergistic influence among multiple features in the development process. However, land use, population, and economy concurrently evolve and intertwine with each other (National Research Council 2005). For example, because of the new economic development zone or high-speed railways, urban built-up areas will expand to forests or farming land. It will attract more people to migrate into cities and stimulate to build more public and infrastructure facilities, such as residential buildings, schools, and hospitals. Consequently, it also increases economic production and leads the encroachment on agricultural land (Lei *et al.* 2021). On the other hand, population boom and economic growth will further require more factories, roads, and schools; thus, more land will be deforested (Zhang *et al.* 2020). Whereas, the slumping economy will decrease the population and consequently reforest some built-up parcels for a sustainable environment. These concurrent spatial processes suggest that the regional coordinated development of land, population, and economy should be considered. Therefore, it highlights the necessity of cooperative simulation of multiple features, which will comprehensively portray future development patterns and benefit regional spatial policy-making.

To fill this gap, we propose a spatial cooperative simulation (SCS) approach to simulate concurrent land use-population-economy changes in the megacity region. The main idea is inspired by the co-evolution of land, population, and economic production. The basic CA is used to forecast the spatial process of land use, population, and economic production, respectively. Then, the simulation results of each other two features will be used as dynamically updating driving factors, rather than the historical data, to repeatedly train the simulation model. In this way, the inter-wined influence of multiple features during the development process is captured. The training process will be repeated until the overall simulation error is converged. The main contributions of this approach are summarized as follows:

1. The SCS approach is proposed to model the concurrent dynamics of land use, population, and economic production.
2. A step-wise cooperative simulation framework is designed to portray the cooperative interaction of multiple features by iteratively updating the driving factors.
3. An experiment on the GBA, China, demonstrates that the proposed SCS approach outperforms popular baseline methods.

The remaining parts of this article are organized as follows. Section 2 defines the spatial cooperative simulation problem. Section 3 presents the details of the proposed SCS approach. Section 4 reports the experiment setting and the result analysis. Section 5 discusses the implications of this study and the vision for future work. Finally, this study is concluded in Section 6.

2. Problem statement

The spatial cooperative simulation problem can be expressed as the synchronous simulation of multiple spatial features by considering the cooperative influence among them. It can be defined as follows:

Given a set of spatial units $S = [U_1,\ U_2,\ \ldots,\ U_n]$ seamlessly covering the study area, each spatial unit with multiple features, i.e. land use, population, economic production, etc., simulates the spatial change of multiple features such that the differences between the simulated and observed results are minimized.

Taking the simulation of land use-population-economy changes as an example, each kind of simulation error at time t can be described as $E^t_{LU},\ E^t_{POP},\ E^t_{ECO}$, and the aim is to minimize the overall simulation error, as Equation (1):

$$OSE = w_l * E^t_{LU} + w_p * E^t_{POP} + w_e * E^t_{ECO} \tag{1}$$

where OSE represents the overall simulation error, $w_l,\ w_p,$ and w_e represent the corresponding weights, and the sum of these three weights is set to 1.

3. Methodology

Following the cyclical interaction among land, population, and economic production, the proposed SCS approach step-wisely simulates the change of land, population, and economic production. It extends the basic urban CA to forecast the co-evolving process of multiple spatial features. It generally includes two parts: initial simulation and step-wise cooperative simulation. In the initial simulation part, a normal CA-based feature simulation (CAFS) is used to coarsely simulate the spatial process of land, population, and economic production, respectively. In the step-wise cooperative simulation part, the interactions of multiple features are considered to refine the simulation by iteratively updating interacted features as dynamic driving factors. The overall workflow is illustrated in Figure 1.

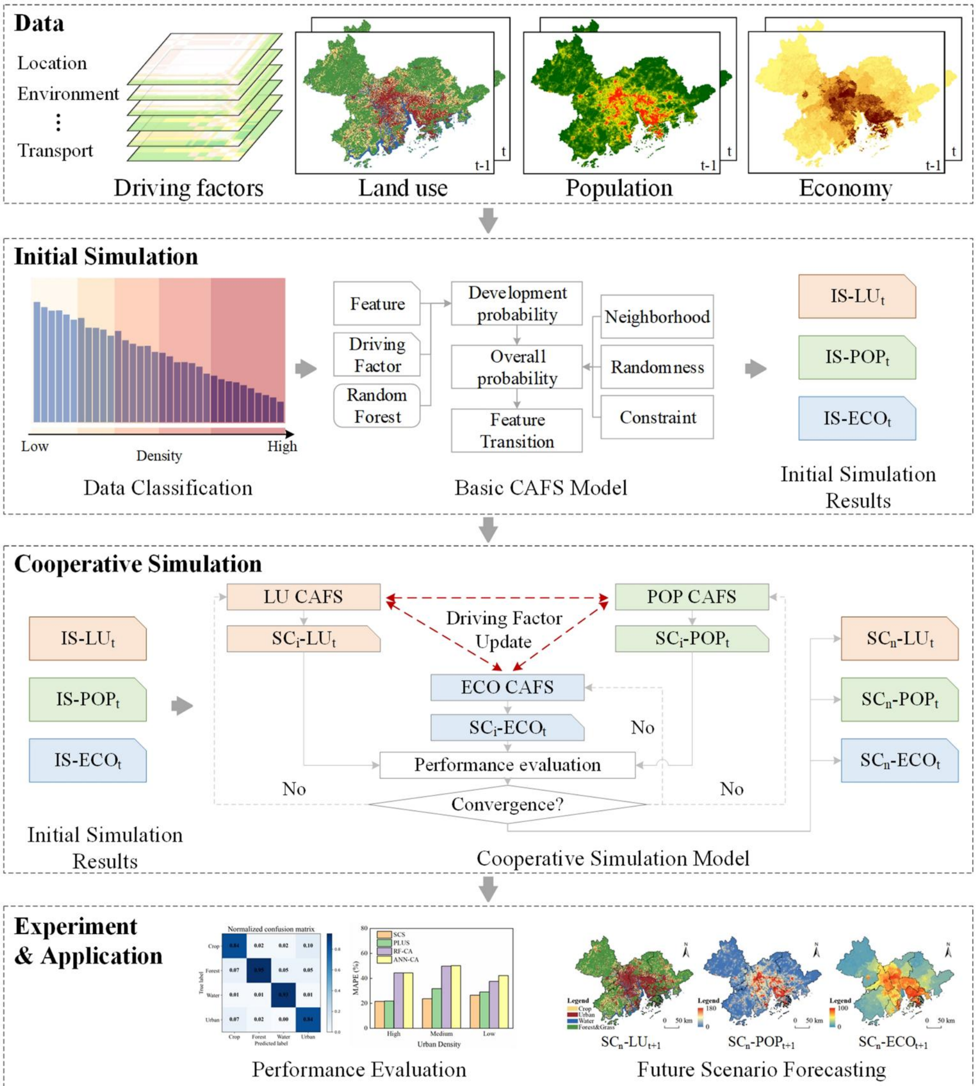

Figure 1. The flowchart of spatial cooperative simulation.

3.1. Cellular automaton-based feature simulation

CA has been widely used to simulate the change of urban phenomenon, such as land use, population, or economic production, by considering the effect of physical environment, transport network, location, socio-economic, and neighborhood effects. Hence, this study takes them as common driving factors for CAFS. Following the patch-generating land use simulation (PLUS) model (Liang *et al.* 2021), the basic CAFS model computing is sequentially divided into four parts: development probability, neighborhood effects, adaptive parameter, and overall probability.

The development probability estimates the future feature by considering the impact of spatial variables, such as physical environment, transport network, and socio-economic. Because of the good performance of random forest (RF) (Du *et al.* 2018), CAFS utilizes the RF for mining the relationship between the development probabilities of features (e.g. land-use patterns, population density, economic production density, etc.) and multiple associated factors. Data from two time periods are used for modeling the impact of spatial variables on feature changes. During the training process, a portion of the areas whose feature state changes at the premier time *t-1* and the current time t are selected as training samples. The spatial variables of the selected area are used as the input vector and the corresponding feature state at time t is used as the label. RF is used to accurately mine the feature conversion rules and output development probabilities $Pg^d_{u,k}$, as follows Equation (2):

$$Pg^d_{u,k}(x) = \frac{\sum_{m=1}^{M} I(h_m(x) = d)}{M} \tag{2}$$

where $Pg^d_{u,k}$ is the development probability of cell u shifting to feature type k; d can be taken as 0 or 1, 1 means other feature types can be transformed to feature type k, 0 means no transformation; x is a vector consisting of multiple driver forces, i.e. the distance to railways, distance to rivers, distance to first and second grade roads, distance to second grade roads, distance to major cities in the Greater Bay Area, the slope degree. $I(\cdot)$ is the indicative function of the decision trees in RF; $h_m(x)$ is the prediction type of the m-th decision tree for vector x; M is the total count of decision trees.

Neighborhood impose great effects on the cell conversion. Neighborhood interaction rules play a key role in the calculation of cellular conversion probabilities. The neighborhood effect function is given as follows:

$$\Omega^r_{u,k} = \frac{\sum_{l\times l} I\left(c_u^{r-1} = k\right)}{l \times l - 1} \times w_k \tag{3}$$

where $\Omega^r_{u,k}$ is the neighborhood effect of cell u subject to feature type k at the *r-th* iteration; $\sum_{l\times l} I\left(c_u^{r-1} = k\right)$ denotes the total number of grid cells occupied by feature type k at the last iteration $r-1th$ within the $l \times l$ grids; w_k denotes the weights between different feature types, where different types have different neighborhood effects.

The adaptive parameter A^r_k depends on the difference in the number of cells between current development and future demand F_k as follows:

$$A^r_k = \begin{cases} A_k^{r-1}, & if \ \left|F_k^{r-1}\right| \le \left|F_k^{r-2}\right| \\ A_k^{r-1} \times \dfrac{F_k^{r-2}}{F_k^{r-1}}, & if \ 0 > F_k^{r-2} > F_k^{r-1} \\ A_k^{r-1} \times \dfrac{F_k^{r-1}}{F_k^{r-2}}, & if \ 0 < F_k^{r-2} < F_k^{r-1} \end{cases} \tag{4}$$

where F_k^{r-1} and F_k^{r-2} are the differences between the current amount and future demand for feature type k at the $r-1$th and $r-2$th iteration.

The final overall probability of CAFS model can be estimated as follows:

$$P^{d,r}_{u,k} = \begin{cases} Pg^d_{u,k} \times (\varphi \times \mu_k) \times A^r_k, & if \ \ \Omega^r_{u,k} = 0 \ and \ \varphi < \ Pg^d_{u,k} \\ Pg^d_{u,k} \times \Omega^r_{u,k} \times A^r_k, & all \ others \end{cases} \tag{5}$$

where $P_{u,k}^{d,r}$ is the of cell u developing into feature type k at the r-th iteration, and φ is a random value from 0 to 1, μ_k determined by model users is the threshold value to generate the new patches for feature type k. Finally, according to the overall probabilities of all the feature types, a roulette wheel is used to select the feature state in the next iteration.

3.2. Initial simulation with CAFS

The spatial distribution of land use, population, and economic production at premier time *t-1* and the current time *t* are used as inputs to get the initial conversion rules with the previously introduced CAFS model, respectively. To keep a uniform and clear CA framework for following complex simulations, the population and economic production data are first transformed into categorical variables. As both the population and economic production data show positive skewness distribution, the natural breaks method was used to ensure the largest inter-group spacing and the smallest intra-group spacing (Jenks 1967). The values of population and economic production are divided into several groups, and the mean values in each group were chosen as the representative values for the following calculation. Worth noting that the transformation method and the group number are crucial. In terms of the transformation method, the equal interval classification and the quantile interval classification are considered. Because of the long tails of population and economic production, the equal interval method is unable to well portray clustering trends. The quantile interval method fails to capture the variations in some areas with extremely large populations and economic production (e.g. the area in urban center), resulting in the underestimated simulation. In terms of group numbers, a small value will not help distinguish the differences between groups. A large value will make some groups with few samples thus difficult to learn the conversion rules appropriately. After examining the distribution of population and economic production data, we set the group number to 30 with the help of the Caliński-Harabasz analysis (Calinski and Harabasz 1974).

The changes in land use, population, and economic production are affected by the natural environment, transport, and locations (Lippe *et al.* 2022, Wang *et al.* 2022). Hence, three categories of factors are considered, including natural resources (rivers, forests, slopes, etc.), transport elements (roads, railways, airports, etc.), and the location factor (city centers, etc.). With the support of the feature important test with RF and the grid search strategy, the distance to the rivers, the slope, and the elevation, the distance to railways, the distance to first and second-grade roads, the distance to the centers of major cities in the GBA (Guangzhou, Shenzhen, and Hongkong), are selected as common driving factors.

Using the categorical features and those driving factors, the initial simulation is conducted. Three CAFSs, including the LU CAFS, the POP CAFS, and the ECO CAFS are implemented separately, and the initial simulation results at current time t, IS-LU$_t$, IS-POP$_t$, and IS-ECO$_t$ are obtained.

3.3. Step-wise cooperative simulation

Regional land use, population, and economic production development are inter-wined and demonstrated with significant cooperative development characteristics. Here, we

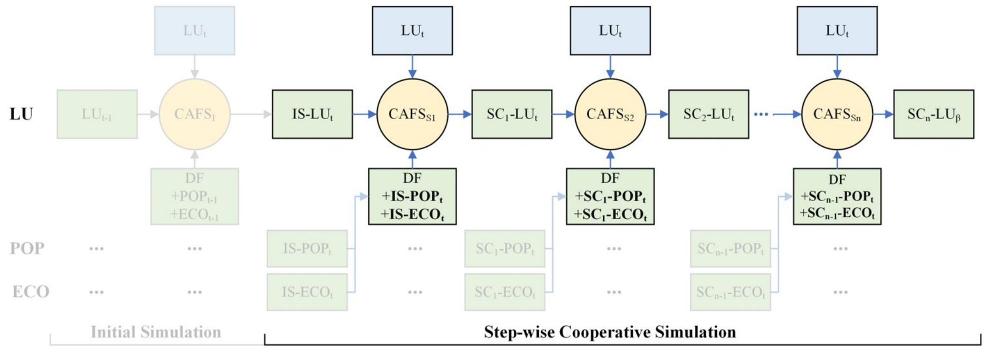

Figure 2. Illustration of step-wise cooperative simulation in land use.

develop a step-wise dynamic driving factor updating strategy that follows the cyclical interaction among land, population, and economic production. The core idea is to train the CAFS models repeatedly using the simulated features, rather than the historical data, to capture the cooperative influence of features. The details of the step-wise cooperative simulation are described as follows. Firstly, the initial simulation results are used as part of the driving factors to learn the conversion rules between the initial simulated results and the observed results and output the cooperative simulation result in step 1. Then, the cooperative simulation results are used as the dynamically updating driving factors to re-train the corresponding CAFS model, thus improving the simulation performance of the remained two simulation features. This step will be iteratively performed until the changes of each feature are converged.

Taking land use as an example, Figure 2 illustrates the step-wise cooperative simulation process. The initial simulated population (labelled as IS-POP$_t$) and economic production (labelled as IS-ECO$_t$) are first used as driving factors to learn the conversion rules from the initial simulated land use (labelled as IS-LU$_t$) and the observed land use (labelled as LU$_t$) at time t. Then, output the result of cooperative simulation land use in step 1 (labelled as SC$_1$-LU$_t$). Similarly, the cooperative simulated results of the population (labelled as SC$_1$-POP$_t$) and economic production (labelled as SC$_1$-ECO$_t$) in step 1 are obtained. They are used as part of the driving factors to generate the cooperative land use simulation result in step 2 (labelled as SC$_2$-LU$_t$). The same cycle will be iteratively performed until the changes of each feature tend to converge. The final step of the land use simulation result (labelled as SC$_n$-LU$_t$) is reported as part of the SCS results. A complete technical flowchart of the step-wise cooperative simulation is shown in Figure A2 in the Appendix file. When performing future scenario forecasting, the land use, population, and economic production scenes and the common driving factors at current time t will be organized into the same format and transmitted sequentially to the series of trained CAFS models, and the final output SC$_n$-LU$_{t+1}$, SC$_n$-POP$_{t+1}$, and SC$_n$-ECO$_{t+1}$, will be treated as the forecasted scenario.

Note that the number of simulation steps significantly affects the performance. As the problem statement section suggests, the aim is to minimize the overall error *OSE* (which was defined in Equation 1). In this study, the weights w_l, w_p, and w_e were set equally which was widely used in the multi-criteria studies (Tran *et al.* 2020). The land use simulation error E_{LU}^t was set to the difference between one and the overall accuracy (OA). As the sub-district is usually the basic spatial unit that holds the local authority for urban governance, the mean absolute percentage error (MAPE) at the administrative sub-district level is set as the population simulation error E^t_{POP} and the economic production simulation error E^t_{ECO}. When the *OSE* does not decrease for *k* steps, the SCS model is converged. The result with the lowest *OSE* is output as the final result. The *k* value affects the final result. A larger *k* value can improve the confidence of model convergence, but it will cost more computational effort.

3.4. Performance evaluation

As both the categorical variable (land use) and the continuous variables (population and economic production) are simulated in the SCS approach, four kinds of metrics are used to evaluate the performance. In particular, the OA and the (FOM) coefficient are used to measure the inter-rater reliability for land use simulation, which are defined as follows:

$$OA = 1 - \frac{A + C + D}{N} \tag{6}$$

$$FOM = \frac{B}{A + B + C + D} \tag{7}$$

where *N* indicates the number of cells in the GBA, China. *A* denotes the area of error due to the observed changes being simulated as unchanged, *B* denotes the area of observed changes being simulated as changing into a right category, *C* denotes the area of error due to the observed changes being simulated as changing into a wrong category, *D* denotes the area of error due to the unchanged being simulated as changes.

The simulation of population and economic production were evaluated using the metrics of MAPE and root mean square error (RMSE). The metrics can be defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}\left(v_{i,t} - v'_{i,t}\right)^2}{M}} \tag{8}$$

$$MAPE = \frac{\sum_{i=1}^{M} {|v_{i,t} - v'_{i,t}|}\big/{v_t}}{M} * 100\% \tag{9}$$

where *M* indicates the number of spatial units in the GBA, $v_{i,t}$ denotes the density of the observed feature value in sub-district *i* at time *t*, and $v'_{i,t}$ denotes the density of the simulated value of the feature in sub-district *i* at time *t*. Worth noting that the performance evaluation on population and economic production were evaluated with the raw values to ensure comprehensibility.

4. Experiment and result analysis

4.1. Study area and datasets

The GBA, which includes nine cities in Guangdong Province, and two special administrative regions (Hong Kong and Macao), was selected to investigate the performance

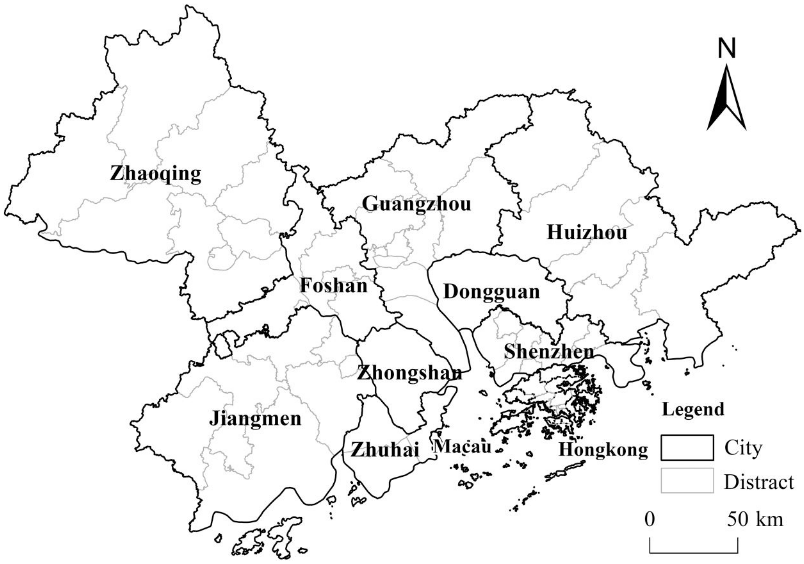

Figure 3. The map of the Greater Bay Area.

of the presented SCS approach. The GBA is one of the regions with the highest degree of openness and economic vitality in China, which holds an important strategic role in national development, as Figure 3 shows. The area of the GBA was ~56,000 km^2, with a population of 86.17 million, and a GDP of 11.59 trillion CNY in 2020 (Xu *et al.* 2021). According to the national GBA development plan, it will be built as the international financial, transportation, innovation, and trade centers.

The publicly accessible land use, population, and economic production were retrieved for the experiment. The land use data were derived through the GlobeLand30 project in the years 2010 and 2020 (Chen *et al.* 2014). Manual corrections were conducted with the aid of concurrent high-resolution remote sensing images. The classification evaluation was carried out by an independent group with an overall accuracy of 85.72%. The population data were derived through the open GPWv4 project (CIESIN 2018). It consists of human population density based on counts consistent with national censuses and population registers. Population counts are assigned to 1×1 km grid cells with the allocation gridding algorithm. In terms of economic production, we took the GDP as the indicator. The GDP data with a spatial resolution of 1×1 km was derived through the Resource and Environment Science Data Centre of China (Xu 2017). All these three features were calculated. The overall distributions of land use, population, and GDP in the year 2020 are shown in Figure 4.

The natural environment, transportation, and location are considered as the common driving factors of the SCS approach. The railways, rivers, first and second-grade roads, and major cities in the Greater Bay Area (Guangzhou, Shenzhen, and Hongkong), were obtained by using the AMAP API (https://lbs.amap.com/). The digital elevation model (DEM) was downloaded from the ASTER GDEM project (https://www.jspacesystems.or.jp/ersdac/GDEM/E/). Using these data, driving factors in Section 3.2 were calculated. All these spatial variables were reclassified with a spatial resolution of 1×1 km. The distributions of driving factors are shown in Figure A1 in the Appendix file.

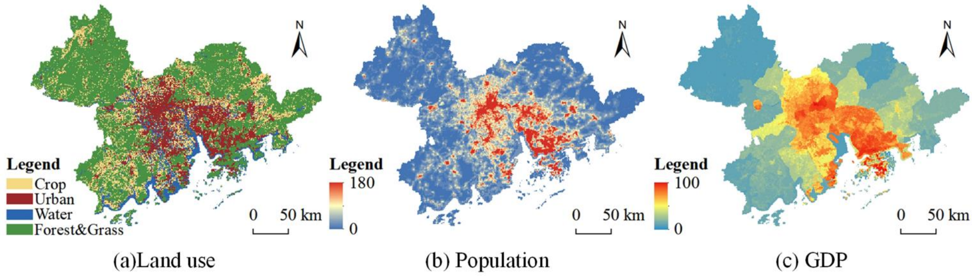

Figure 4. The spatial distribution of (a) land use, (b) population (thousand people per km^2), and (c) GDP (billion CNY per km^2) for the GBA 2020.

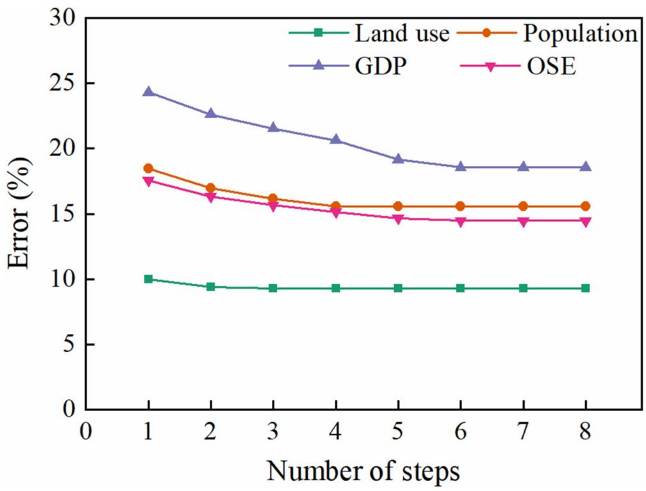

Figure 5. Influences of the number of steps on simulation error.

4.2. Results of the spatial cooperative simulation approach

To verify the performance of the presented SCS approach, the simulation errors of land, population, and GDP were calculated. The OSE was summed. The corresponding errors are shown in Figure 5. It suggests that the simulation errors of all four indices decrease at the beginning, and converge after the 3rd, 4th, 6th, and 6th steps, respectively. It demonstrates the effectiveness of the proposed step-wise dynamic driving factor updating strategy as these indices fast converge to a lower value. Worth noting that the simulation of land use reaches convergence at the early step. We speculate that it is because the number of categories in land use is smaller (four categories *vs.* 30 categories), making it easier for the LU CAFS model to learn the conversion rules. More categories of population and economy increase the complexity of the CAFS model therefore it is not easy to learn the conversion rules. As the simulation results have initially converged in the first part, the final results demonstrate that the step-wise cooperative simulation can still reach a good and reliable performance.

The OA and the FOM of the SCS approach in land use simulation are 90.71% and 0.267, respectively. The MAPE of the population and the GDP are 15.56 and 18.56%, respectively. The repeatability test is conducted to test the reliability of the simulation performance. The SCS approach was run five times repeatedly with the same dataset and hyperparameter settings. The OA of land use ranges from 90.39 to 90.79%, the MAPE for population ranges from 15.56 to 16.51%, and the MAPE for GDP ranges from 18.56 to 23.84%. The uncertainty was quite acceptable. A detailed simulation result is shown in Table A1 in the Appendix file.

The spatial comparison analysis was conducted to examine the spatial trends of the simulation results. The simulation results of land use, population, and GDP for the GBA in 2020 are shown in Figure 6, respectively. Compared to the actual land use, population, and GDP in 2020 in Figure 4, it suggests that the simulated spatial patterns are well correlated with the actual pattern at the whole GBA scale. Most urban built-up areas and populations are distributed in the area along the Pearl River, such as Guangzhou-Foshan, Shenzhen-Hong Kong, and Zhuhai. Regarding the economy, Guangzhou, Shenzhen, and Hong Kong are with high GDP. Moreover, a typical high-density sub-district, Shatian in Hong Kong, and a low-density sub-district, Kaiping in Jiangmen, are selected to show more details of the simulated results. The enlarged views of these two areas also showed high spatial consistencies with actual and simulated patterns of land use, population, and GDP. It further demonstrates the versatility of the proposed SCS approach in mega-city regions with different development levels.

4.3. Comparison with baseline approaches

To further evaluate the simulation performance of the proposed SCS approach, three popular approaches were selected as baselines:

- **Artificial neural network-CA (ANN-CA):** This approach applies an artificial neural network to learn the transition rules and simulate the evolution of feature changes based on the integration of neural networks and CA (Li and Yeh 2002).
- **Random forest-CA (RF-CA):** This approach combines random forest and CA for feature simulation. (Kamusoko and Gamba 2015).
- **Patch-generating simulation model (PLUS):** This approach adopts a land expansion analysis strategy and CA with multi-type random patch seeds to portray the drivers of feature changes (Liang *et al.* 2021).

For all baseline methods, the data used and the static parameters were consistent with the SCS approach to ensure comparability. Specific parameters of baseline methods were set with a grid search algorithm. The detailed parameters were reported in Table A2 in the Appendix file.

The results of all methods are shown in Table 1. The OA and the FOM of the SCS approach in land use simulation are 90.71% and 0.267, respectively, which outperform three baseline methods with the best OA of 90.01% (PLUS) and 0.200 (PLUS). The MAPE and RMSE of the SCS in population and GDP simulation are also the lowest. PLUS reported good population and GDP simulation results with the MAPE 21.08 and 26.12%, respectively. RF-CA and ANN-CA reported similar simulation results of the population and the GDP with the MAPE around 29 and 44%, respectively. These

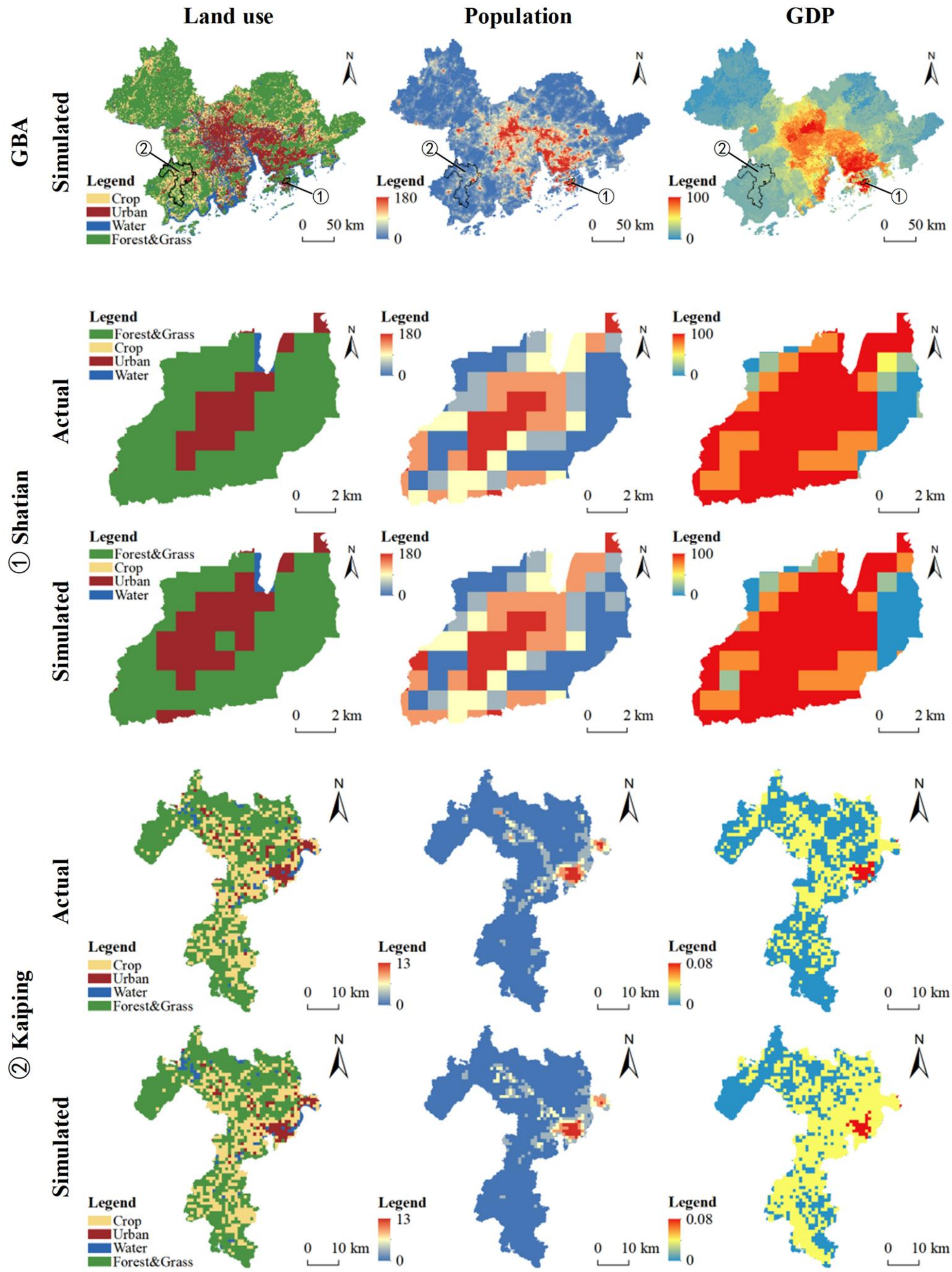

Figure 6. The simulation results of land use, population (thousand people per km^2), and GDP (billion CNY per km^2) in the GBA, 1 Shatian, and 2 Kaiping in 2020.

results thus demonstrate the superiority of the presented SCS approach. We acknowledge that the performance of the GDP simulation doesn't look good enough. Considering that the GBA China is a large region with significant development heterogeneity and the inherent uncertainties arising from the spatialization process of the

Table 1. Comparison of simulation performance with baseline methods (population in thousand people per km^2, GDP in billion CNY per km^2).

Feature	Approaches	OA (%)	FOM	MAPE (%)	RMSE
Land Use	SCS	90.71	0.267	–	–
	PLUS	90.01	0.200	–	–
	RF-CA	85.95	0.158	–	–
	ANN-CA	86.00	0.155	–	–
Population	SCS	–	–	15.56	2.205
	PLUS	–	–	21.08	4.222
	RF-CA	–	–	28.26	4.311
	ANN-CA	–	–	29.15	4.485
GDP	SCS	–	–	18.56	0.980
	PLUS	–	–	26.12	1.222
	RF-CA	–	–	45.94	1.716
	ANN-CA	–	–	43.89	1.715

Table 2. The Wilcoxon signed rank test results for the simulation performance comparison.

	Land use		Population		GDP	
Compared pairs	Statistics	*p*-Values	Statistics	*p*-Values	Statistics	*p*-Values
SCS and PLUS	90	$p < 0.001$	662	0.0018	743	0.0134
SCS and RF-CA	170	$p < 0.001$	429	$p < 0.001$	91	$p < 0.001$
SCS and ANN-CA	99	$p < 0.001$	422	$p < 0.001$	91	$p < 0.001$

population and GDP, the simulation performance of land use, population, and GDP in the GBA was quite acceptable. Moreover, since the metrics of OA and MAPE are more intuitive and easier to understand, they are used as the main evaluation metrics for the following analysis.

A Wilcoxon signed rank test was conducted to verify whether the simulation results of the proposed approach were significantly different from the baseline methods (Wilcoxon 1945). The simulation errors were aggregated into the sub-district scale with the evaluation metric of OA and MAPE, respectively. The hypothesis testing result is shown in Table 2. Compared to the expected values of the null hypothesis, all compared pairs achieved larger statistics values and showed relatively small *p*-values ($p < 0.05$). They indicated that the simulation performance of our proposed SCS approach for land use, population, and GDP were all significantly better than baseline approaches.

To further examine the superiority of the proposed SCS approach, the simulation performance was analyzed from the perspectives of the land use type and the density of the cities. In terms of the land use simulation, the errors of different land use types have different meanings. As the water area was regarded as a limiting factor that wasn't allowed to be converted, the simulation errors of the other three types of land use in each approach were calculated. The comparison results are shown in Figure 7. We observed that all simulation approaches showed better simulation performance on forest and grassland than on urban land and crops. It could be expected that the forest and grassland were less likely to change and the conversion rules were relatively simple. The SCS approach performed well for each land use type. Worth noting that the simulation performance of our approach on urban land, the most important type in land use simulations, was significantly better than baseline methods.

In terms of the simulations of population and GDP, population density, as an important factor in economic and urban planning studies, affects the simulation

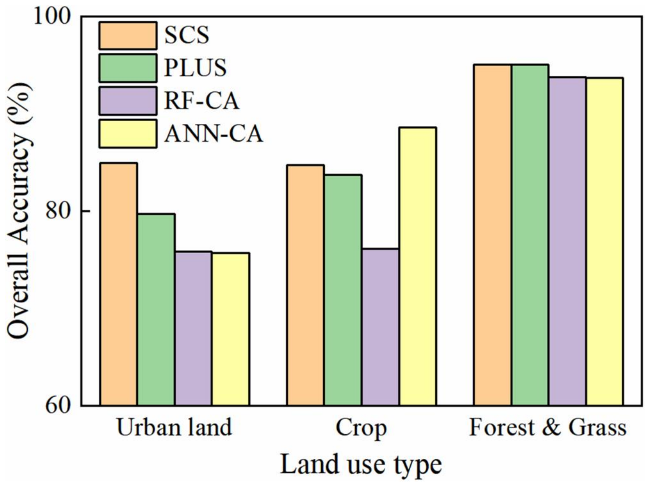

Figure 7. Simulation performance on different land use types.

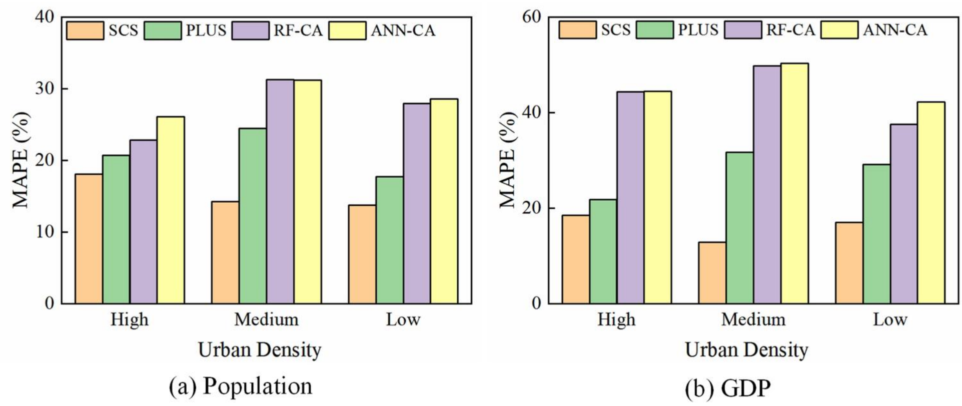

Figure 8. Influence of urban density on (a) population simulation and (b) GDP simulation.

performance (Hui 2001). Referring to the cumulative distribution values of 40 and 80% of the population density, we classified the sub-districts into three groups (the high, the medium, and the low group) with the density thresholds of about 2500 people and 12,500 people per km^2, respectively. The comparison results are shown in Figure 8. It is noted that the MAPE of our proposed SCS approach varied from 13.78 to 18.12% in population simulation and from 17.06 to 18.57% in GDP simulation in three groups. Compared with baseline methods, it still achieved the lowest simulation error.

4.4. Future land use-population-economics scenario in the GBA

The future scenario of land use, population, and GDP in the GBA in 2030 was forecasted using the SCS approach. In this study, the natural development scenario (which is based on the past and current development in the GBA) was set to project future land use, population, and economic production. Current trends for land use,

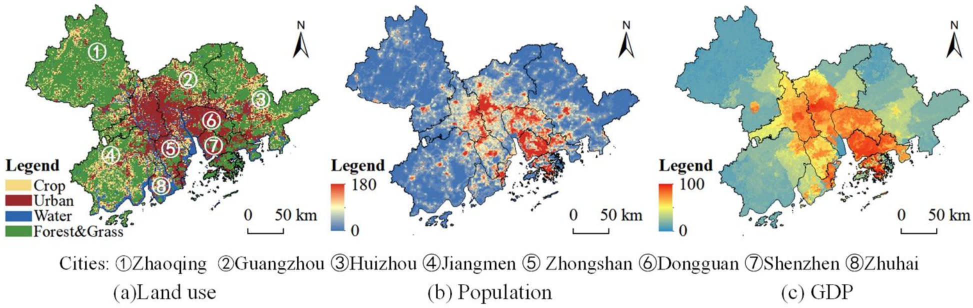

Figure 9. The forecasting results of (a) land use, (b) population (thousand person per km^2), and (c) GDP (billion CNY per km^2) in the GBA in 2030.

population, and economics and the influence of natural environment, transport, and location factors are assumed to remain consistent.

The forecasting results are shown in Figure 9. In terms of land use, we find that urban land expansion is mainly located in places around the metropolitan areas. The cities of Guangzhou, Shenzhen, Dongguan, and Zhongshan will form large contiguous areas of urban land. Most of the forest and grasslands are distributed in cities, such as Zhaoqing, Jiangmen, and Huizhou, which is consistent with the actual pattern in 2020. This result agrees with that of Jiao *et al.* (2019). As for the population and GDP, the cities of Shenzhen and Guangzhou still are the leaders in both the total amount and the growth rate. The city of Zhuhai has a low GDP growth rate despite a significant population growth rate. Zhuhai will attract more people to live but the economic growth is still limited by its industrial infrastructures. This suggests that the development pattern envisioned in the GBA development plan, with Zhuhai as the fourth-largest regional growth center, has not yet taken shape. Moreover, the surrounding cities, such as Zhaoqing are still developing at a slower pace. It reflects that, in the current development situation, the role of core GBA cities to drive neighboring cities still needs to be further strengthened.

5. Discussion

5.1. Empirical basis for cooperative simulation in regional development

Regional coordinated development is the result of the cooperation of natural, social, and economic resources (Wilson 1981, Alberti 2008, Zhao *et al.* 2022). The mutual influence and co-evolution phenomena of spatial features, such as land use, population, and economy in the regional development process provide an empirical basis for our proposed spatial cooperative simulation approach. Take land use change as an example, the growth of urbanization drives population and industry to a few cities. In the early stages of regional development, the combined effect of population and economy produced the rapid expansion of urban built-up areas. Consequently, it further accelerated the population and economic growth in the long term. With the further development of the region, the expansion of built-up area has squeezed the space of agricultural and ecological land, intensifying the contradiction between urban, water, forest, and grass. The distribution of high-quality arable land typically overlaps highly with urban built-up areas, which pressure on arable land protection. Nevertheless, the limited built-up area controlled urban sprawl and affected the growth rate of the urban population and economy to some extent. This study demonstrates the urban built-up area of Shenzhen is nearing the ceiling, which has driven up land prices. It has led to the migration of manufacturing industries to the north of Shenzhen, such as Huizhou and Dongguan. Local industries are transforming into high-value-added industries, such as Internet and financial companies. This has caused an out-migration of workers and affected the economic growth in the short term. While cities with population and industrial outflows show the opposite situation. The loss of population and economy has caused a significant gap between the urban built-up expansion rate in these cities and the core GBA cities. Its reserved surplus of urban built-up area can provide good conditions for industrial transfer from the core city. The Shenzhen-Shantou Cooperation Zone also provides a new paradigm for such cross-city cooperation and congenerous development. These phenomena pointed to the existence of the cooperative influence among land use-population-economy development, thus helping to demonstrate the rationality of our proposed SCS approach.

5.2. Insight of the step-wise simulation

This study designed a step-wise SCS framework to simulate the regional coordinated development of land, population, and economy. It generally has two implications. First, the step-wise SCS framework dynamically uses the predicted future feature as part of the driving factors. It is quite different from existing urban CAs that only use static and historical data to calculate driving factors. Taking land use as an example, when simulating the land use change in 2020, most methods used the 2010 population data and economic data as driving factors (e.g. Liu *et al.* 2017). The latent question is whether land use changes in 2020 are determined by the properties surrounding this patch in 2010. However, in the real-world development process of the city, for example, by 2015, the population and economic data have changed. If we can simulate the population and economic data in 2015 first, and then simulate the land use in 2020 with the simulated results, it will effectively improve the performance of the simulation approach. This is also the core idea of our proposed approach. From this perspective, the SCS approach can capture not only the natural trend of a whole time period, but also the interaction among land use, population, and economy in a long time period. Second, the step-wise SCS approach draws on a similar idea of the gradient descent decision tree (Friedman 2001). Since the CAFS model builds conversion rules by focusing only on cells with different categories, in the step-wise framework, the cells that are simulated correctly will no longer participate in the next step of conversion rule extraction. That is, in the next step, the SCS model pays more attention to the wrong simulation results by incorporating better results of land use, population, and economy. This reduces the complexity of the following simulation to a certain extent, thereby helping to improve the final simulation. The good performance of the step-wise SCS approach was verified by comparing it with three baseline methods including RF-CA, ANN-CA, and PLUS. We also compare our simulated 2020 population and economy results with two commonly used population datasets, GPWv4 (CIESIN 2018) and WorldPop (Tatem 2017), and two economy datasets, a dataset published by Resource and Environment Science and Data Center, China (Xu 2017) and a dataset published in the journal of Scientific data (Chen *et al.* 2022). The MAPEs of these two population datasets are 25.68 and 26.43%, respectively. While the MAPEs of these two economy datasets are 33.03 and 57.29%. All are worse than our proposed SCS approach (15.56% for the population and 18.56% for the economy). These results also demonstrated the superiority of the proposed SCS approach. Furthermore, the proposed step-wise framework is extensible. The simulated features can be replaced by other phenomena, such as carbon footprints and energy consumption.

5.3. Limitations and future work

There are several limitations to be further explored. First, in the current SCS approach, the continuous variables (population and GDP) were discretized firstly for the input requirement of the used feature simulation model. It allows a uniform CA base for the SCS approach but also limits the upper boundary of the simulation performance. Simulation models for continuous variables, such as Grayscale CA and machine learning-based models, may help solve this issue in future work. Second, limited by the obtained data, the used driving factors are mainly derived from basic geographic data. More driving factors (e.g. local climate, detailed economic factors, etc.) and important policies should be considered to further improve the performance. Third, the spatial scale of CA also affects the simulation performance. When increasing the size of the cell from 1 to 2 km, the FOM of the SCS approach on land use decreases to 0.181, and the MAPE on population and GDP simulation increases to 21.19 and 26.34%, respectively. A reasonable spatial scale selection scheme will be developed for the spatial cooperative simulation approach. Finally, since the aim of this study is to provide a novel cooperative simulation approach, only the natural development scenario, was implemented to forecast the future development patterns of the GBA. Move development scenarios, such as fast development scenarios, energy conservation and emission reduction scenarios, and ecological conservation scenarios, should be implemented to forecast the regional dynamics of GBA to provide technical support for policy decisions.

6. Conclusion

Classical CAs have been widely used to independently simulate the change in land use or population, which fosters spatial planning and policy-making. But they ignore the intertwined influences among land, population, and economy. This study proposes the SCS approach to simulate the land use-population-economy changes in the megacity region. Basic CA is used to forecast the spatial process of one feature. The interactions among multiple features are captured by taking one feature as the dynamic driving factor. The CA model is iteratively trained until the overall simulation error is converged. An experiment in the GBA, China was conducted to examine the performance of the SCS approach. The results demonstrate that the SCS outperforms baseline approaches with an OA of 90.71% on land use simulation, a MAPE of 15.56% on population simulation, and a MAPE of 18.56% on GDP simulation. Future GBA 2030 scenarios were obtained to support spatial planning and regional coordinated development policy-making.

This study provides an effective approach to comprehensively simulate the land use-population-economy in the mega-city region. It enables us to forecast the future regional development scenario under different development goals, which provides technical support for megacity region applications, such as regional planning and facility locations. In addition, the feature variation processes can help reveal the patterns of regional development, thus enriching the relevant theories and practices of sustainable development goals in the mega-city region.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Key Research and Development Program [2019YFB2103104, 2021YFD1901001]; the National Natural Science Foundation of China [42071360, 42101463]; the Shenzhen Science and Technology Program [JCYJ20220818100200001, ZDSYS20210623101800001]; the Guangdong Science and Technology Strategic Innovation Fund (the Guangdong–Hong Kong–Macau Joint Laboratory Program) [2020B1212030009], and the Major Science and Technology Projects of Yunnan Province [202202AE090034].

Notes on contributors

Wei Tu is a professor at the Department of Urban Informatics, Shenzhen University. His research interests include automatics recognition of human activity-mobility from multi-source urban data, trajectory modeling, and analysis and optimization. He contributed to the conception of the research idea, the supervision of the experiments, and the writing of the manuscript.

Wei Gao is a master's student at the Department of Urban Informatics, Shenzhen University. His research interests include urban simulation and spatiotemporal modeling. He contributes to the data collection, algorithm implementation, visualization, validation, and interpretation.

Mingxiao Li is an assistant professor at the College of Civil and Transportation Engineering, Shenzhen University. His research interests focus on urban computing, human mobility, GeoAI, and spatiotemporal data mining. He contributed to the idea conceptualization, methodology design, original draft, manuscript revision, experiment supervision, and visualization.

Yao Yao is a professor at China University of Geosciences (Wuhan), a researcher from the Center for Spatial Information Science at the University of Tokyo, and a visiting scholar at Alibaba Group. His research interests are geospatial big data mining, analysis, and computational urban science. He contributed to the model development and experiment supervision.

Biao He is an associate professor in the department of urban spatial information engineering, Shenzhen University. His research interests include spatial-temporal big data analysis and smart city. He contributed to the manuscript revision and project administration.

Zhengdong Huang is a professor at the Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University. His research interests include GIS and urban traffic, transportation, environment, urban spatial analysis, and smart city. He contributed to the methodology development and supervision.

Jie Zhang is an assistant professor at the College of Information and Electrical Engineering, China Agricultural University. Her research interests include sustainable development, environment changes monitoring and modelling from remote sensing, multi-source data, models, and system analysis. She contributed to the review and editing of this manuscript.

Renzhong Guo is a member of the Chinese Academy of Engineering and a professor at the Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University. His research includes smart cities, geographical information engineering, 3D city modeling, simulation of virtual urban environments, resource management of national territories, and spatial planning of national territories. He contributed to the methodology supervision and overall administration.

ORCID

Wei Tu http://orcid.org/0000-0002-4749-6037
Wei Gao http://orcid.org/0000-0002-7769-830X
Mingxiao Li http://orcid.org/0000-0003-0779-4432
Yao Yao http://orcid.org/0000-0002-2830-0377
Biao He http://orcid.org/0000-0002-2608-9731
Zhengdong Huang http://orcid.org/0000-0001-7264-4567
Jie Zhang http://orcid.org/0000-0002-9888-5665

Data and codes availability statement

The data and codes used in this research are available on figshare.com with the unique identifier at the link https://doi.org/10.6084/m9.figshare.21218372, and the software can be downloaded from https://www.urbancomp.net/archives/coca-v100.

References

Alaei Moghadam, S., Karimi, M., and Habibi, K., 2018. Simulating urban growth in a megalopolitan area using a patch-based cellular automata. *Transactions in GIS*, 22 (1), 249–268.

Alberti, M., 2008. *Advances in urban ecology: integrating humans and ecological processes in urban ecosystems*. New York, NY: Springer.

Arsanjani, J.J., *et al.*, 2013. Integration of logistic regression, Markov chain and cellular automata models to simulate urban expansion. *International Journal of Applied Earth Observation and Geoinformation*, 21, 265–275.

Benenson, I. and Torrens, P., 2004. *Geosimulation: automata-based modeling of urban phenomena*. Hoboken, NJ: John Wiley & Sons.

Bijandi, M., Alesheikh, A.A., and Sadeghi Niaraki, A., 2019. Integrating multi-agent model and irregular cellular automata to simulate urban growth. *Scientific-Research Quarterly of Geographical Data (SEPEHR)*, 27 (108), 61–76.

Calinski, T. and Harabasz, J., 1974. A dendrite method for cluster analysis. *Communications in Statistics-Theory and Methods*, 3 (1), 1–27.

Chen, D., *et al.*, 2022. A hierarchical approach for fine-grained urban villages recognition fusing remote and social sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 106, 102661.

Chen, J., Ban, Y., and Li, S., 2014. Open access to Earth land-cover map. *Nature*, 514 (7523), 434–434.

CIESIN, Center for International Earth Science Information Network, Columbia University, 2018. *Gridded population of the world, version 4 (GPWv4): basic demographic characteristics, revision 11*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Crols, T., *et al.*, 2017. High-resolution simulations of population-density change with an activity-based cellular automata land-use model. *In*: *Proceedings of GeoComputation 2017*, 1–4.

Du, G., *et al.*, 2018. A comparative approach to modelling multiple urban land use changes using tree-based methods and cellular automata: the case of Greater Tokyo Area. *International Journal of Geographical Information Science*, 32 (4), 757–782.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29 (5), 1189–1232.

Geng, J., *et al.*, 2022. A hybrid spatiotemporal convolution-based cellular automata model (ST-CA) for land-use/cover change simulation. *International Journal of Applied Earth Observation and Geoinformation*, 110, 102789.

Georganos, S., *et al.*, 2021. Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. *Geocarto International*, 36 (2), 121–136.

Ghavami, S.M., Taleai, M., and Arentze, T., 2022. An intelligent web-based spatial group decision support system to investigate the role of the opponents' modeling in urban land use planning. *Land Use Policy*, 120, 106256.

Guzman, L.A., *et al.*, 2022. Modeling population density guided by land use-cover change model: a case study of Bogota. *Population and Environment*, 43 (4), 553–575.

Harvey, E.P., *et al.*, 2019. Developing integrated models by coupling together existing models; land use, economics, demographics and transport in Wellington, New Zealand. *Computers, Environment and Urban Systems*, 74, 100–113.

Hui, S., 2001. Low energy building design in high density urban cities. *Renewable Energy*, 24 (3–4), 627–640.

Jalayer, S., *et al.*, 2022. Modeling and predicting land use land cover spatiotemporal changes: A case study in Chalus watershed, Iran. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 5496–5513.

Jenks, G.F., 1967. The data model concept in statistical mapping. *International Yearbook of Cartography*, 7, 186–190.

Jiang, X., *et al.*, 2022. Examining the spatial simulation and land-use reorganisation mechanism of agricultural suburban settlements using a cellular-automata and agent-based model: six settlements in China. *Land Use Policy*, 120, 106304.

Jiao, M., Hu, M., and Xia, B., 2019. Spatiotemporal dynamic simulation of land-use and landscape-pattern in the Pearl River Delta, China. *Sustainable Cities and Society*, 49, 101581.

Kamusoko, C. and Gamba, J., 2015. Simulating urban growth using a Random Forest-Cellular Automata (RF-CA) model. *ISPRS International Journal of Geo-Information*, 4 (2), 447–470.

Lei, W., *et al.*, 2021. Scaling of urban economic outputs: insights both from urban population size and population mobility. *Computers, Environment and Urban Systems*, 88, 101657.

Li, M., *et al.*, 2022. Fine-grained crowd distribution forecasting with multi-order spatial interactions using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 144, 103908.

Li, M., Zhang, H., and Chen, J., 2019. Fine-grained dynamic population mapping method based on large-scale sparse mobile phone data. In *2019 20th IEEE International Conference on Mobile Data Management (MDM)*. IEEE, 473–478.

Li, X. and Yeh, A.G.-O., 2002. Neural-network-based cellular automata for simulating multiple land use changes using GIS. *International Journal of Geographical Information Science*, 16 (4), 323–343.

Li, X., *et al.*, 2020. *Geographical simulation system: cellular automata and GeoAI*. Beijing: Science Press.

Liang, H., *et al.*, 2020. GDP spatialization in Ningbo City based on NPP/VIIRS night-time light and auxiliary data using random forest regression. *Advances in Space Research*, 65 (1), 481–493.

Liang, W. and Yang, M., 2019. Urbanization, economic growth and environmental pollution: evidence from China. *Sustainable Computing: Informatics and Systems*, 21, 1–9.

Liang, X., *et al.*, 2021. Understanding the drivers of sustainable land expansion using a patch-generating land use simulation (PLUS) model: a case study in Wuhan, China. *Computers, Environment and Urban Systems*, 85, 101569.
Lippe, M., Rummel, L., and Günter, S., 2022. Simulating land use and land cover change under contrasting levels of policy enforcement and its spatially-explicit impact on tropical forest landscapes in Ecuador. *Land Use Policy*, 119, 106207.
Liu, X., *et al.*, 2017. A future land use simulation model (FLUS) for simulating multiple land use scenarios by coupling human and natural effects. *Landscape and Urban Planning*, 168, 94–116.
Lu, Y., Song, W., and Lyu, Q., 2022. Assessing the effects of the new-type urbanization policy on rural settlement evolution using a multi-agent model. *Habitat International*, 127, 102622.
Mustafa, A., *et al.*, 2018. Modelling built-up expansion and densification with multinomial logistic regression, cellular automata and genetic algorithm. *Computers, Environment and Urban Systems*, 67, 147–156.
National Research Council, 2005. *Population, land use, and environment: research directions. The national academies collection: reports funded by National Institutes of Health*. Washington, DC: National Academies Press (US).
Rahnama, M.R., 2021. Forecasting land-use changes in Mashhad Metropolitan area using Cellular Automata and Markov chain model for 2016–2030. *Sustainable Cities and Society*, 64, 102548.
Rydin, Y., *et al.*, 2004. The Greater London Authority—a case of conflict of cultures? Evidence from the planning and environmental policy domains. *Environment and Planning C: Government and Policy*, 22 (1), 55–76.
Schwaab, J., *et al.*, 2018. Improving the performance of genetic algorithms for land-use allocation problems. *International Journal of Geographical Information Science*, 32 (5), 907–930.
Shafizadeh-Moghadam, H., *et al.*, 2017. Coupling machine learning, tree-based and statistical models with cellular automata to simulate urban growth. *Computers, Environment and Urban Systems*, 64, 297–308.
Shahfahad, *et al.*, 2022. Prediction of land use changes at a metropolitan city using integrated cellular automata: past and future. *Geology, Ecology, and Landscapes*. Advance online publication. doi: 10.1080/24749508.2022.2132010
Shao, Q., Liu, X., and Zhao, W., 2020. An alternative method for analyzing dimensional interactions of urban carrying capacity: case study of Guangdong-Hong Kong-Macao Greater Bay Area. *Journal of Environmental Management*, 273, 111064.
Sun, J., *et al.*, 2020. Regular pattern formation regulates population dynamics: Logistic growth in cellular automata. *Ecological Modelling*, 418, 108878.
Tatem, A.J., 2017. WorldPop, open data for spatial demography. *Scientific Data*, 4 (1), 170004.
Tran, P.T.M., *et al.*, 2020. Cyclists' personal exposure to traffic-related air pollution and its influence on bikeability. *Transportation Research Part D: Transport and Environment*, 88, 102563.
United Nations, 2018. *World urbanization prospects: The 2018 revision*. Geneva: United Nations.
Usman, M., *et al.*, 2022. How do financial development, energy consumption, natural resources, and globalization affect Arctic countries' economic growth and environmental quality? An advanced panel data simulation. *Energy*, 241, 122515.
Wang, J., *et al.*, 2022. Machine learning in modelling land-use and land cover-change (LULCC): current status, challenges and prospects. *The Science of the Total Environment*, 822, 153559.
Weng, H., Kou, J., and Shao, Q., 2020. Evaluation of urban comprehensive carrying capacity in the Guangdong–Hong Kong–Macao Greater Bay Area based on regional collaboration. *Environmental Science and Pollution Research International*, 27 (16), 20025–20036.
White, R., Engelen, G., and Uljee, I., 2015. *Modeling cities and regions as complex systems. Fromtheory to planning applications*. Cambridge, MA; London: The MIT Press.
Wilcoxon, F., 1945. Individual comparisons by ranking methods. *Biometrics Bulletin*, 1 (6), 80–83.
Wilson, A.G., 1981. *Geography and the environment: systems analytical methods*. Chichester: John Wiley & Sons, Ltd.
Xu, C., *et al.*, 2021. *Report of reform and innovation of Guangdong-Hongkong-Macao greater bay area in China (2020)*. Beijing: Social Sciences Academic Press.

Xu, X., 2017. China's GDP spatial distribution kilometer grid dataset. China GDP Spatial Distribution Kilometer Grid Dataset. Data Registration and Publication System of the Resource and Environmental Science and Data Center, Chinese Academy of Sciences. Available from: http://www.resdc.cn/DOI.

Yeh, A.G.-O. and Chen, Z., 2020. From cities to super mega city regions in China in a new wave of urbanisation and economic transition: issues and challenges. *Urban Studies*, 57 (3), 636–654.

Yu, Y., *et al.*, 2019. National green GDP assessment and prediction for China based on a CA-Markov land use simulation model. *Sustainability*, 11 (3), 576.

Zhang, B. and Xia, C., 2022. The effects of sample size and sample prevalence on cellular automata simulation of urban growth. *International Journal of Geographical Information Science*, 36 (1), 158–187.

Zhang, J., *et al.*, 2020. Exploring annual urban expansions in the Guangdong-Hong Kong-Macau Greater Bay Area: spatiotemporal features and driving factors in 1986–2017. *Remote Sensing*, 12 (16), 2615.

Zhao, T., *et al.*, 2022. Coupling graph deep learning and spatial-temporal influence of built environment for short-term bus travel demand prediction. *Computers, Environment and Urban Systems*, 94, 101776.

Zhou, Y., *et al.*, 2018. Emissions and low-carbon development in Guangdong-Hong Kong-Macao Greater Bay Area cities and their surroundings. *Applied Energy*, 228, 1683–1692.

Appendix A

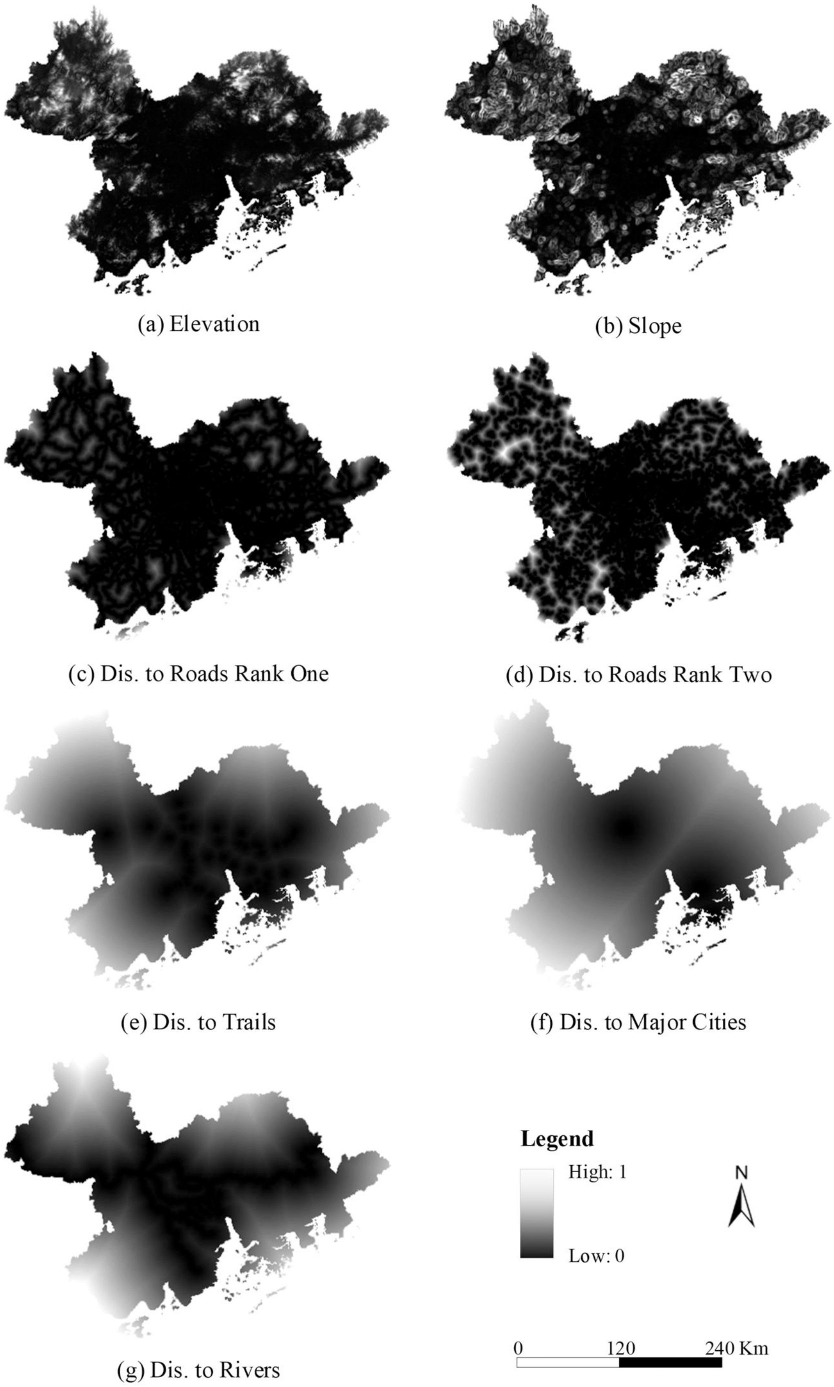

Figure A1. The maps of the driving factors. All the values were normalized into the range [0,1] with the min-max scale algorithm.

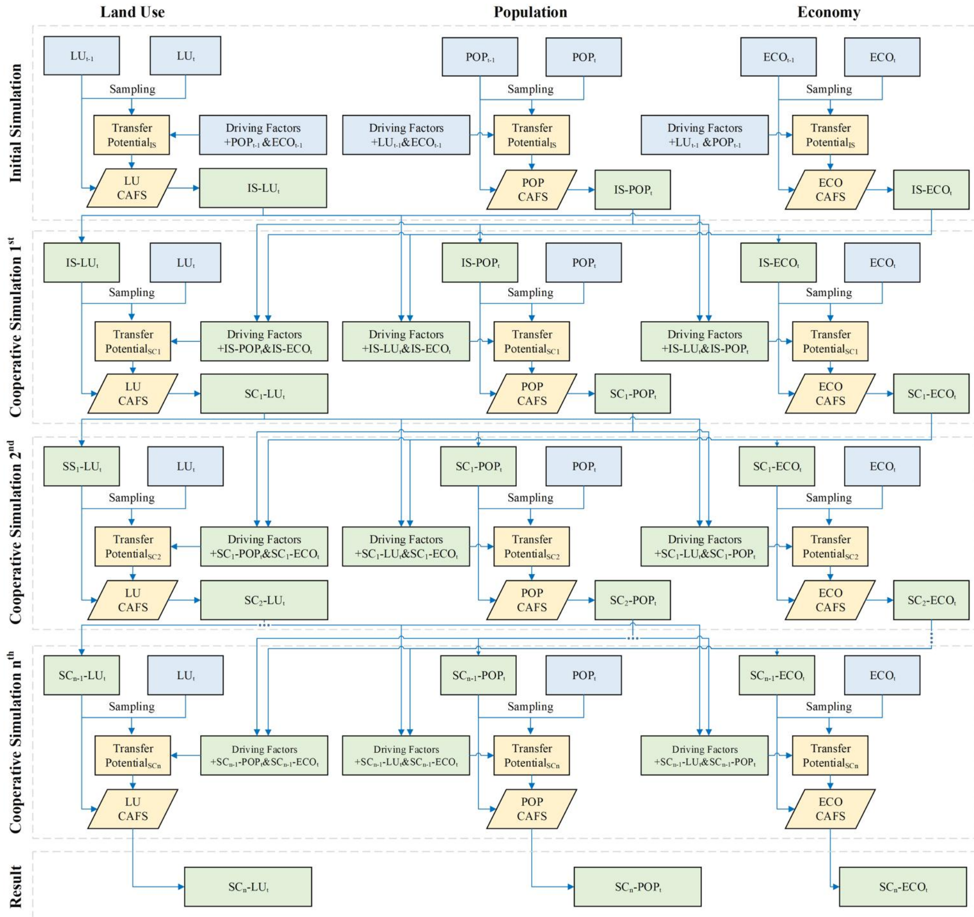

Figure A2. The complete technical flowchart of the step-wise cooperative simulation.

Table A1. The repeatability test result of SCS approach.

	Land use		Population		GDP	
No.	OA (%)	FOM	MAPE (%)	RMSE	MAPE (%)	RMSE
1	90.71	0.267	15.56	2.205	18.56	0.980
2	90.79	0.270	16.01	3.348	23.84	1.129
3	90.48	0.246	16.22	3.960	20.19	1.000
4	90.48	0.263	16.14	3.218	20.49	0.980
5	90.39	0.251	16.51	2.834	21.40	1.014

Table A2. Hyperparameter settings of the presented SCS approach and baseline methods.

Method	Hyperparameter	Value
Static setting	Neighborhood size	5*5
	Sampling method	Uniform sampling
	Sampling rate	30%
	Loss function	MSE
PLUS	Patch generate	0.5
	Expansion coefficient	0.5
	Percentage of seeds	0.01
ANN-CA	Number of input units	9
	Number of hidden layers	1
	Number of output units	4
RF-CA	Number of trees	1000
	Min samples split	2
	Min samples leaf	1
	Max depth	10