RESEARCH

Clustered and dispersed: exploring the morphological evolution of traditional villages based on cellular automaton

Xi Yang[1]* and Fuan Pu[2]

Abstract

The spatial pattern of traditional villages can be generally divided into two main types: clustered and dispersed. In order to explore and compare the spatial evolutionary characteristics of different village patterns, and provide a reliable basis for spatial planning, a universal Cellular Automaton (CA) model was built and applied in different spatial simulations. Through model comparison, it was established that: (1) both types of villages have developed in the same cyclical changing mode of "outlying + edge-expansion", which was probably rooted in the inherent spatial sense of the ethnic group inhabiting village types; (2) the spatial growth of the clustered village was more relevant to the distribution structure of pre-existing buildings, whereas the spatial sprawl of a dispersed one was more connected to external natural factors; and (3) the development of every economic unit in a dispersed village was strictly restricted to the building area, and to the proportion of population inhabiting farmland area. Although village patterns developed under the same logical framework rooted in ethnic culture, their development tendency varied, with different dynamic mechanisms and constraints.

Keywords: Clustered and dispersed, Cellular Automaton, Dynamic space, Spatial simulation and analysis, Traditional villages

Introduction

Traditional villages refer to villages that originated in earlier times and have rich cultural and natural resources [1]. These villages need to be protected because of their specific values in history, culture, science, art, economy and society [2, 3]. Instead of the concept of the "Ancient Village", the concept of "Traditional Village" in China originated and was legalized in September 2012, when the first national list was published [4, 5]. For a long time before this, and even in the following few years, the concepts of "Traditional Village", "Ancient Village" [6, 7] and "Historical Village" [8, 9] were used in a mixed way until 2017. The legalization of the concept of "Traditional Village" actually emphasizes the identification standard, narrows the affirmation scope, and puts more attention to the long history of the defined villages, the preciousness of their different values, and the necessity of their sustainable revitalization in conservation planning [10]. Since then, efforts to protect traditional villages have attracted national attention and gained legal status [4]. By the end of 2019, there were five lists including a total of more than 6800 villages designated as national traditional villages in China. Some of these villages are World Heritage sites such as Fujian Tulou, Kaiping Diaolou, Xidi and Hongcun [11]. The protection of traditional villages should maintain their authenticity, integrity and sustainability, respect the traditional rules of spatial growth and its dependence on the surrounding environment. In this context, it is important to investigate the spatial mechanisms of the historical development of traditional villages.

*Correspondence: xi-yang12@tsinghua.org.cn
[1] School of Architecture, Harbin Institute of Technology (Shenzhen), 518055 Shenzhen, China
Full list of author information is available at the end of the article

As an important carrier of rural culture, economy, and ecology, the spatial pattern of traditional villages is usually gradually generated by self-organization without any interference of planning activities [12–15]. Therefore, once the planning activities intervene in the rural spatial development process, protection or development actions of people in the area need a certain historical reference. History can teach us about two important issues: one is the appearance of the spatio-temporal changes of the village pattern, and the other is the logic behind that [16]. Before making conservation plan [17], decision makers would be highly concerned with the following issues: "How does the growth of patches containing buildings occur?", "What are the factors driving such growth?", and "What are the differences in the driving force of various factors?" [18]. Analyzing the characteristics of the spatial process and getting acquainted with its basic logic is the first step in planning [19, 20]. Only by understanding the process and dynamic mechanisms of the rural morphological structure, decision makers are able to understand certain characteristics of the civilization and spiritual core of the village, and grasp the rules of spatial organization, based on which they can propose reliable expectations about the spatial development [21]. In order to interpret the spatial process and extract the dynamic mechanism of traditional villages, the following questions need to be answered: firstly, how to describe a village pattern and its expansion modes; secondly, how to explore the driving factors of spatial evolution; and thirdly, how to effectively extract, express and compare the dynamic mechanism of traditional villages with different spatial patterns.

Description of spatial patterns and expansion

From the perspective of static structure, the spatial pattern of the villages has been defined in the description of a system called Euclidean geometry, in terms of a strip, ring, clump or other shapes [22]. This type of classification is simple and intuitive, but no effective principles have been established to limit the infinite classification of shape types. In fact, the classification of the spatial pattern based on the geometric form ignores the links between the environmental factors and the village form. For example, if the site selection principle of each house in a village is to avoid arable land and flooding hazards, and be close to the foothills, then the village will have a circular or semi-circular layout if it is in hilly areas, and a strip layout if it is in valleys. In other words, even under the same rules of site selection, villages will develop according to different morphological trends in different geographical environments [23]. In this case, the geometric form can hardly be the basis for the morphological identification of the villages. Therefore, we believe that the distribution of villages in the environment should be controlled by relational factors in two mutually independent dimensions: one is the spatial relationship between the village and the natural environment, i.e., the environmental attributes of the village; the second is the internal morphological organizational relationship of the village, i.e., the spatial organization of the village. The appropriate method for spatial morphological classification and identification should not be correlated with the natural attributes of the village but should reflect the logic of the internal morphological organization of the village. Therefore, when taking into account the organization of buildings that reflect the internal morphological logic of the village, there are only two types of spatial patterns, namely clustered and dispersed type [24]. The dispersed type can be further subdivided into random discrete type and uniform discrete type [20].

If observed through a coarse temporal resolution [25] (covered only the initial stage, the intermediate stage and the current stage of the development process), the modes of rural spatial growth can be divided into two types: edge-expansion growth and outlying growth [26, 27]. Then the growth of clustered villages approximately belongs to the edge-expansion type of growth (including outward spreading and inward filling), and the growth of dispersed villages appears to belong to the outlying type of growth. Clustered villages exhibit a tendency of agglomeration, where settlements and farmlands are relatively more distant from one another. On the contrary, settlements in dispersed villages are significantly far away from each other, and each residential group is located as close as possible to its proprietary land which it inhabits and its natural resources [24, 28]. Based on the comparison of characteristics of the two types of villages, scholars have hypothesized about the original development tendency of villages [24, 28–30], and they further explored whether the course of the development of villages throughout history was from gathering process to the scattering one, or vice versa [31–33]. The focus of the paper is not on the sequential development order between these two types, but on the connection and intersection of the mentioned types. In other words, since the transition from one spatial pattern to another happened during the historical development, there was no absolute boundary between these two types, where a similar logical framework of development may connect them.

We selected several traditional villages in different geographical environments from different regions of China based on data availability of spatial process, including Xidi Village [34] in the intermountain basin of Anhui Province, Tanshui Village in the island area of Fujian Province, QuanFu Village [35] in the plateau area of

Yunnan Province, Qiaoxiang Village [36] in the mountain valley and Kengzi Village [37] in the hilly basin of Guangdong Province. All these villages have a history of more than 100 years. So far, these villages have retained most of their historic buildings and local folk customs. If the development process of these villages is reconstructed with a fine temporal resolution that includes five temporal sections in the development process (see Fig. 1), the outcome would be that both types of villages, i.e., clustered and dispersed villages, display the similar combinatorial growth logic of "outlying + edge-expansion". Specifically, the development of clustered villages did not exhibit the spatial growth of the edge-expansion type. On the one hand, the boundary of the village was expanded by outlying development, while on the other hand, the space between each enclave was filled by edge-expansion until a compact cluster was formed. The same process was also present in the development of dispersed villages. The essential difference between spatial growth processes of these two types of villages lies in their distance of outlying expansion and their scale of edge-expansion. Under this logic, a village can be seen as a composite system consisting of numerous economic units [38]. Each of these units agglomerated towards their core building which can be regarded as a spatial node of the village structure. The growth of a village is the proliferation of economic units in outlying mode, with edge-expansion of each economic unit [37].

In addition, traditional village studies need to be conducted at a relatively fine spatial resolution [39], making it difficult to draw absolute boundaries between outlying and edge growth. In particular, due to the influence of some historical, political and economic factors, the history of rural development in southern China is much longer than that of northern China. Moreover, their difference is also in the structure of social organization. Under the influence of these factors, villages in the south are generally larger and more scattered, while those in the north are smaller and more concentrated [40]. This is also reflected in the patterns of villages in different regions (see Fig. 1). Landscape evolution of villages cannot be clearly shown at the macroscopic scale level of land use. On the contrary, it should be observed at the mesoscopic scale level that can reflect the spatial array, connectivity, and their interaction when it comes to the landscape of villages [41]. The decrease in the scale of observation means the increase in spatial resolution, where we can see more subtle changes [42]. Given that the width of a village house (or an enclosed village house composed of several buildings and courtyards) in the study settlement is about 10–150 m, the spatial resolution in this study is 5–10 m, while the specific values for different villages

Fig. 1 Spatial development process of some villages in China
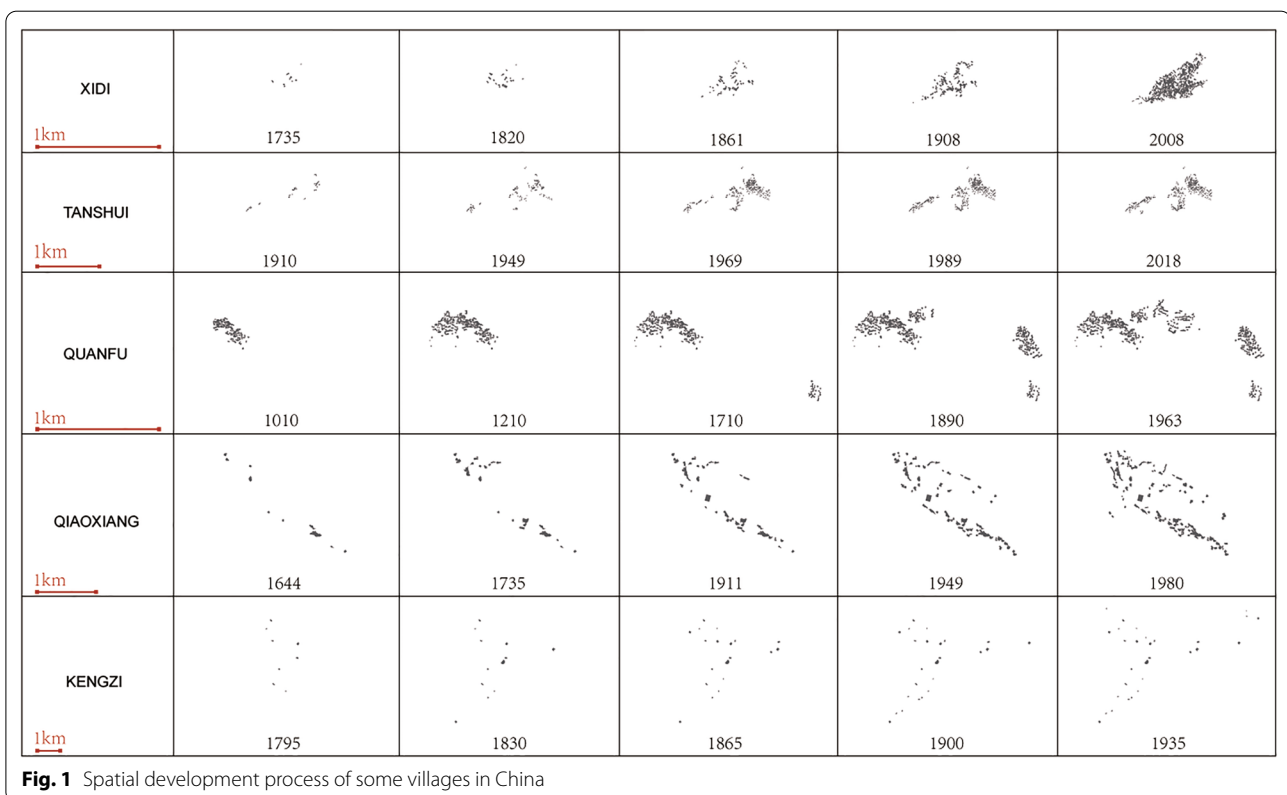
depend on the scale of the village and the simulation efficiency. At this spatial resolution, we can observe individual patches containing buildings, rather than continuous patches of land. The differentiation of spatial growth mode does not depend on whether the patch-growth is continuous, but on the minimum distance between new patches and original patches. Edge-growth can be defined as near-spaced or non-spaced outlying growth.

To sum up, the spatial expansion mode, whether in clustered or dispersed villages, is usually a combination of edge-growth and outlying growth, which shows the similarity and complexity of spatial development among different villages. If observed on the meso-scale, there would be no absolute boundary between the edge-growth and the outlying growth in the rural spatial expansion process.

Judgement on the driving factors of spatial evolution

Explaining the spatial distribution characteristics, and elucidating dynamic mechanisms are not only the core issues in the study of the historical evolution of traditional villages, but also the prerequisite for spatial planning. When human awareness of the relationship between man and environment develops from environmental determinism to today's complex bifunctional feedback, that relationship is no longer confrontational, but interactional. When it comes to the interaction between man and environment, people are not individualized, but organized and socialized. Furthermore, the environment is not a pure natural environment, but a broad concept, including visible artificial environment, natural environment, and abstract social environment. Through continuous practical activities, socialized human beings create artificial environments and social environments in the process of utilizing and transforming the natural environment, or they bring about the utilization and transformation of the natural environment in their creation of artificial environments and social environments [43]. As an artificial product wrapped in a natural environment, spatial patterns of villages reflect not only the organizational structure of human society, but also the economic development potential provided by the natural environment. Therefore, the driving forces of spatial evolution are both cultural organization characteristics and geographical environment characteristics. However, scholars have different opinions with respect to the function of cultural and natural factors. For example, by analyzing the spatial distributional characteristics of settlements, some scholars believed that natural factors such as topography, vegetation and rivers are the basis for rural settlement distribution, while social factors such as population, industrial structure, economic development level, and policy system are the driving forces for spatial

evolution [44]. Contrary to that opinion, other scholars found that the same ethnic group may adopt similar spatial organization mode in different environments, while different ethnic groups mainly adopted different modes even in the same environment. Therefore, it was believed that inherent spatial concepts within an ethnic group were more decisive to the spatial morphology than the local natural environment [45]. What roles did the two types of contextual factors versus environmental factors play in the village spatial evolution? It is an issue that is going to be further discussed.

In the early stage of geographical research, work on factor recognition and explaining mechanisms was mainly completed by qualitative analysis. Since the 1990s, researchers have been more inclined to use quantitative models to assist their judgment [21, 46]. In fact, these two methods should not be used separately, because both of them, when used individually, have defects. According to human's cognition, qualitative analysis is realized by logical reasoning, starting from conditions and ending with results. Although the basic explanation can be provided, it is difficult to clearly measure the weights of various factors influencing spatial evolution, and thus impossible to infer the comprehensive effect of multi-factors. In a word, qualitative analysis is a useful method for induction, but not for deduction. On the other hand, quantitative deductive research based on model usually needs to predict driving factors beforehand [21, 46]. Having more factors involved is not necessarily better [47]. High dimension factors can easily lead to insufficiency of samples or difficulty in model operation. If directly entered into quantitative research, without any qualitative analysis, a large amount of irrelevant information will be dealt with in the quantitative research phase. Nevertheless, important information can still be omitted. Furthermore, the judgment can easily be clouded by these interfering instances [48]. Therefore, the combination of qualitative analysis and quantitative model is a relatively effective means for explaining, demonstrating and deducing the spatial dynamic mechanisms.

Methods of dynamic spatial analysis for a traditional village

The spatial expansion of rural settlements is a process in which patches containing buildings sprawl or spread out with different spacings between them. If the mentioned processes were to be represented with the model, the model needs to be able to simulate the outlying expansion. On the other hand, if acceptable and reliable projections about the past and the future growth are needed in order to analyze the dynamic mechanisms, rather than just seemingly suitable results, then the model that

we create should have a clear physical meaning, whose modus operandi is readily understood.

Cellular Automaton (CA) model has been widely used in the field of geographical research, especially in the simulation of urban land use. In the 1980s, some researchers began investigating urban expansion by applying CA-based models [49, 50]. Pure CA model is driven by neighboring functions in a grid of cells. CA models simulate geographic space as arrays of cells, in which each cell assumes a particular state based on its previous state and the state of neighboring cells, according to a set of transitional rules [51]. A CA model has the advantages of spatial explicitness defined by rules and powerful computation ability obtained from artificial intelligence, thus it can be developed through a combined method [52].

However, the CA model has not been widely used in the study of rural spaces, mainly because rural settlements are more spatially scattered than urban areas [53], which makes the simulation more complicated. Among various applications of CA, transition rules have obtained much attention since they are the core components for CA modeling [54]. Previous urban CA researchers have made an extensive use of the algorithm embedded in the state transition rules, and also made a comparison of the simulation accuracy among various algorithms [55]. Although most CA models used at present can simulate edge-growth well, it is difficult to effectively show outlying development [53]. Based on the Landscape Expansion Index (LEI), the spatial development modes have been classified, and the CA models using case-based reasoning algorithm had been constructed for the simulation of "outlying + edge-expansion"[26]. However, this method is not applicable for researching rural areas. Firstly, those models were based on LEI; however, it has been proven that LEI has a high risk of leading to misinterpretations when it comes to spatial development classification; moreover, it is not an accurate classification method [56, 57]. Secondly, the case-based reasoning algorithm embedded in the transition rules of the CA model is a type of analogy algorithm found in artificial intelligence. Moreover, the reliability of the judgement made according to the given conditions depends on the similarity between the conditions and the information found in the database. Therefore, CA models based on this algorithm could not work in a new situation-based deduction. Furthermore, there are no parameters to be learnt in this model, so the interpretability on the modular level is not possible. In addition, there is also a lack of global model interpretability, because the model is inherently local, and there are no global weights or structures explicitly learned [58]. Consequently, although the LEI binding case-based reasoning CA model may provide us with good results from simulations, it is not flexible and legible enough to provide us with sufficient knowledge in order to assist us in decision making process.

Dynamic theories of the CA model make it suitable for geospatial simulation. However, for the application of the CA model in traditional villages, given the demand of flexible applicability and interpretability of the model, it is not proper to incorporate black box algorithms [59] (e.g., case-based reasoning algorithm, artificial neural network algorithm, etc.) in the model.

Research purpose

In order to explore the characteristics and dynamic mechanisms of the spatial evolution of traditional clustered and dispersed villages, and discover their similarities and differences, we designed a universal CA model structure to simulate the spatial process of these two kinds of villages. Through the calibration of the model with historical data, the choice of model attributes and parameter adjustments were made. Based on that, the dynamic mechanisms of clustered and dispersed villages were analyzed and compared.

This study attempts to introduce the CA model into the spatial process analysis of traditional villages. To resolve the limitation of the application of CA model in small and scattered villages, we assume that both cluster and dispersed villages developed in a mixed mode of "outlying + edge-expansion", and added rules for switching the two growth modes to the CA model.

Since spatial analysis requires interpretability of the model, spatial constraints were preset through qualitative analysis of the development context of the village samples to obtain interpretable model variables. Furthermore, a probabilistic algorithm model, i.e., the Gaussian mixture model, was embedded in the transition rules to simulate the comprehensive effect of multi-spatial constraints on the growth of villages. Thus, the calculation principle of the model can be understood, and the simulation accuracy can be guaranteed.

Methodology

Study area and research period

Two villages in China, Hounan (in Dabu County, Meizhou City) and Zhoutian (in Huiyang District, Huizhou City), located in different regions but under the same homologous cultural system, were selected for spatial analysis and comparison.

The Hakka migration in Guangdong Province experienced two important processes. During the Song and Yuan Dynasties, the Hakka people mainly lived in southern Jiangxi, western Fujian and northeast Guangdong. The first major flow of population occurred during 1360-1530 of the Ming Dynasty. Under the influence of war, famine, and the evil of banditry, the

Hakka population at the junction of these three provinces frequently interflowed. The second major population flow occurred around 1680-1860. Attracted by the government's immigration and land reclamation policies, large numbers of the Hakka population in northeastern and northern Guangdong moved to the hilly areas near the Pearl River Delta [60, 61]

According to Hakka's migration process in Guangdong Province from Ming to Qing Dynasty (see Fig. 2), Hounan village is located in the upstream of the migration route, in a small basin surrounded by mountains on the southern bank of the Meitan River. The Yang clan has settled here since the 1500s and has developed over 500 years. At present, the village covers an area of 7.9 km^2 with a total population of about 5400. The Meitan River is a rather busy river due to heavy transportation of people and goods. Its downstream leads east to Fujian province and the Chaoshan area, along the coast of Guangdong Province. Historically, residents inhabiting the area mainly lived on growing, harvesting, and selling tobacco plants. At present, the village is that of a clustered type of spatial pattern.

Zhoutian Village is located in the valley of hilly area at the northeast edge of the Pearl River Delta, which represents the lower reaches of the migration route. The Ye clan settled here in the 1660s. With a history of more than 350 years, the village covers an area of 17.8 km^2. Its total population reached more than 4800. Throughout history, its residents mainly cultivated land on which they grew rice for living. Currently, the village is that of a dispersed type of spatial pattern (see Fig. 3).

Hounan Village is a village with a concentrated layout, while Zhoutian village has a scattered layout. These two villages were chosen as research samples for the following reasons. First, both villages have been on the list of "Chinese Traditional Villages" since 2013 due to their high historical, contextual and architectural significance. Second, the two villages have the same Hakka culture system in Guangdong Province, thus avoiding the interference of cultural attributes in a comparative study of different spatial forms. In addition, given the feasibility of model calibration, both villages are on a considerable spatial scale to provide architectural samples that meet the requirements of this study.

Due to the loss of rural population caused by unbalanced regional development after the Chinese reform and opening up, these two villages did not have significant spatial expansion after the 1970s. Therefore, the research

Fig. 2 Hakka's migration routes in Guangdong Province, and the location of Hounan and Zhoutian
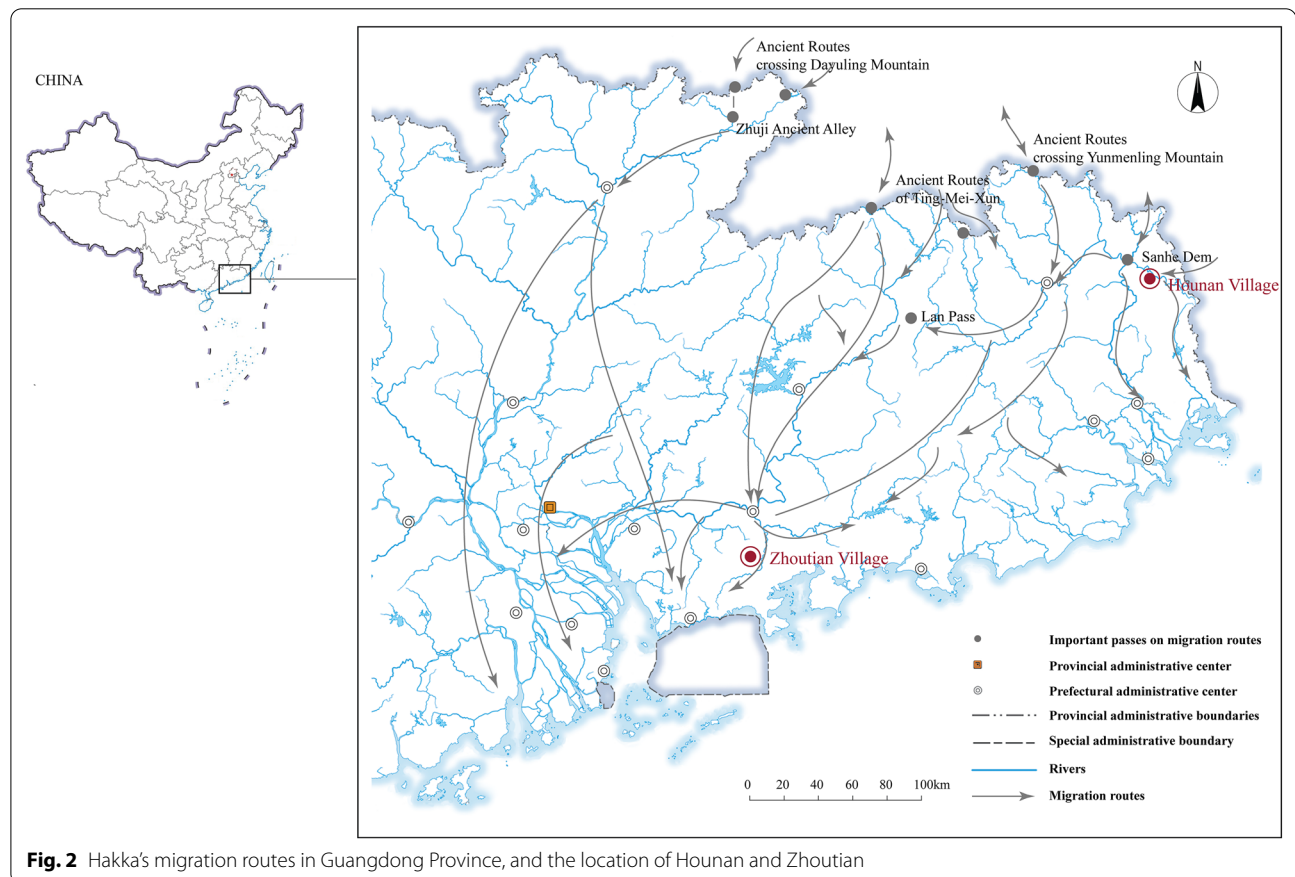
Fig. 3 Satellite images of Hounan and Zhoutian in 2021. (Google Earth satellite imagery data in Level 19, the data comes from the map provider of TuxinGIS)

period of this study focuses on the spatial process from the initial stage of village foundation to the middle of the twentieth century.

Data processing

First of all, for the two villages in question, the patches containing buildings, rivers, canals, ponds, and roads were extracted from the satellite image coming from the map provider TuxinGIS, taken in 2021 with a resolution

of 5 m and mapped onto a spatial lattice. We compared the current satellite images of the selected villages with the earliest geodetic survey maps (around 1930) that could be found (see Fig. 4). Although the measurement accuracy of historical maps is not high enough, it is found that the natural environment of the village has not changed, and the infrastructure of the village, such as natural channels (adapted to the terrain), main roads and ponds, has no obvious changes. However, the current
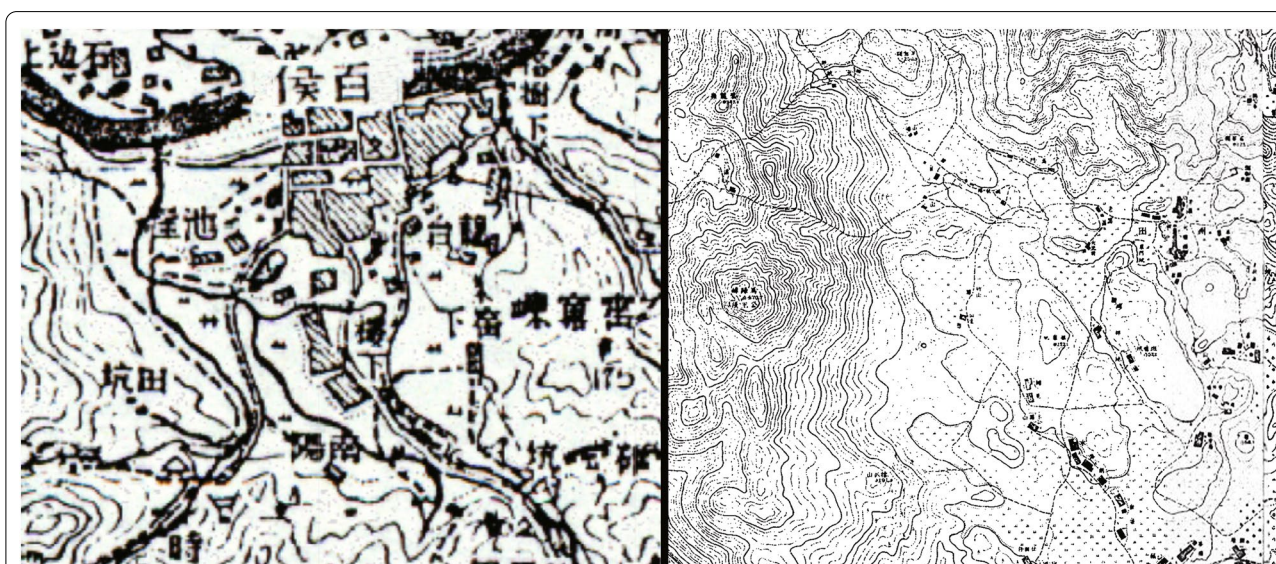

Fig. 4 Hounan and Zhoutian Village, as recorded in historic geodetic survey maps, circa 1900. (From the Atlas of mainland China at a scale of 1:25,000, produced by the Japanese Land Survey Bureau, published by Academy of Science in Tokyo, 1990.)

layout of the branch roads in the village has increased, i.e., the evolution of branch roads in the village tends to increase with the proliferation of buildings. Other historical changes in the infrastructure are the lack of documentation. However, based on the comparison of the image of the village in two periods, we can roughly estimate that the changes in infrastructure in different periods may not be drastic. Due to the limitations of historical data and in order to reduce the complexity of the spatial simulation, this study ignored the small changes of the main infrastructure of the village in the historical period. In addition, in the analysis of the spatial driving factors, the less influential and destabilizing factor, i.e., the factor of branch roads, was not taken into account.

The distances between those mapped ground objects and the cells in the spatial lattice were calculated using ArcMap 10.7 (see the raster data in columns 4–7 in Fig. 5), which is primarily used to view, edit, create, and analyze geospatial data. Secondly, we obtained the rasterized elevation data from the 5-m resolution Digital Elevation Model (DEM) generated by TuxinGIS and used them to obtain the slope information via deduction process (see the raster data in columns 2–3 in Fig. 5). Then, based on the information of building construction time in village annals, the multi-temporal spatial patterns of the two villages were reconstructed (Fig. 6). Finally, the data where the buildings were located were extracted. The data about distances between buildings and about
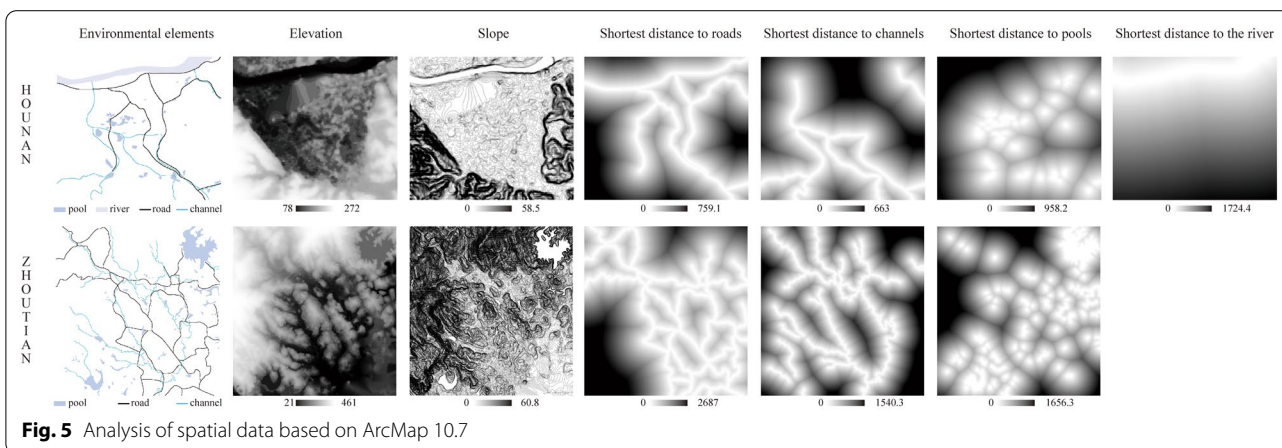
Fig. 5 Analysis of spatial data based on ArcMap 10.7
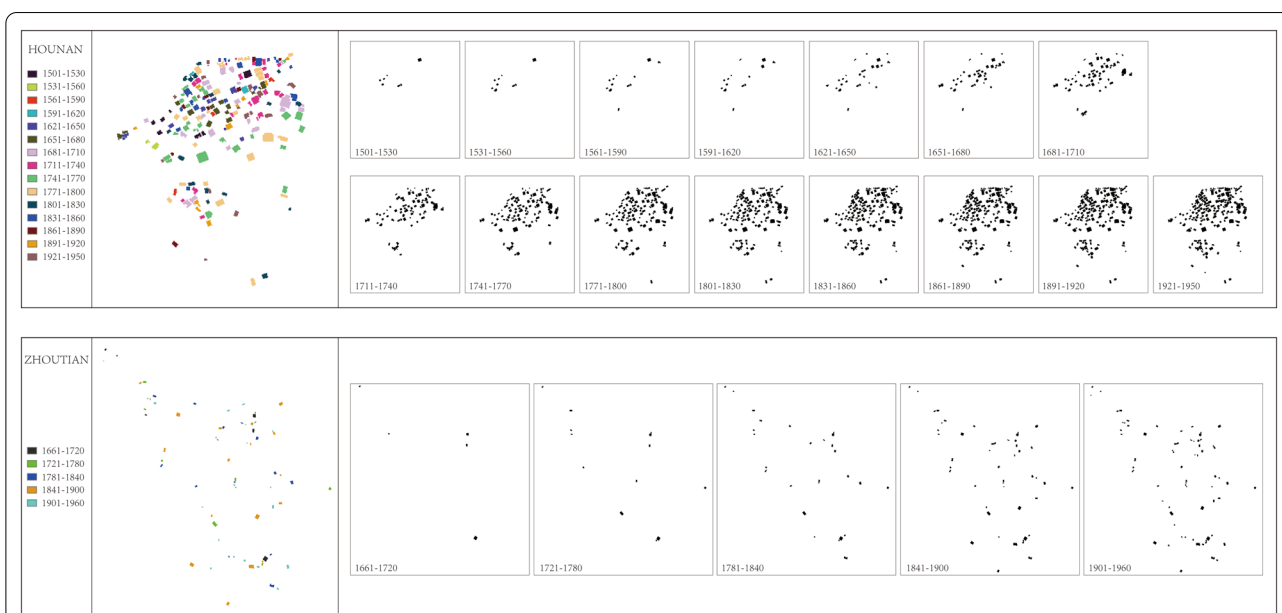
Fig. 6 Multi-temporal spatial patterns of Hounan and Zhoutian

arable land area were calculated through SCILAB programming method. All of the above represent the basic data for model training.

Simulation process based on CA

The spatial resolution used in the simulation should not exceed the minimum scale of a house. In addition, considering data accuracy and model operation efficiency of villages with different area and building scales, a two-dimensional lattice with 5-m resolution for Hounan, and a lattice with 10-m resolution for Zhoutian, were used to construct the CA space. Cellular state included the main road, river, canal, pond, building, and land yet to be developed. Among them, the state about transportation and water source was fixed, and there is only one condition for state transformation, i.e., for land to be developed into land used as building area.

Figure 7 shows the CA simulation process created in reference to the common logic framework of "outlying + edge-expansion". In the site selection stage of the simulation process, buildings generated in the outlying expansion mode are marked as nodal buildings, while the ones generated in the edge-expansion mode are marked as ordinary buildings.

Selection of spatial constraints

Owing to the inhomogeneity of geographical space, CA models usually work with constraints when applied to geographical spatial simulation [62]. The constraints could be reflected in the cell attributes, which can be expressed by *Boolean attribute variables* (using 1 or 0 to mark whether the cell has some attributes or not) and *floating-point attribute variables* (using decimals to mark the cell's value of certain attributes). All the attribute values will exert a comprehensive influence on the cellular state transformation.

Moreover, too many constraints will reduce the working efficiency of the model [47], whereas too few of them will lower the outcome accuracy of the simulation. Balancing increased explanatory power with a reasonable number of cell attribute is important. In order to ensure that the explanatory power of the model is met, first we analyzed the historical, social, and geographical contexts of the region, which was helpful for the primary selection and exclusion of attributes. For further verification
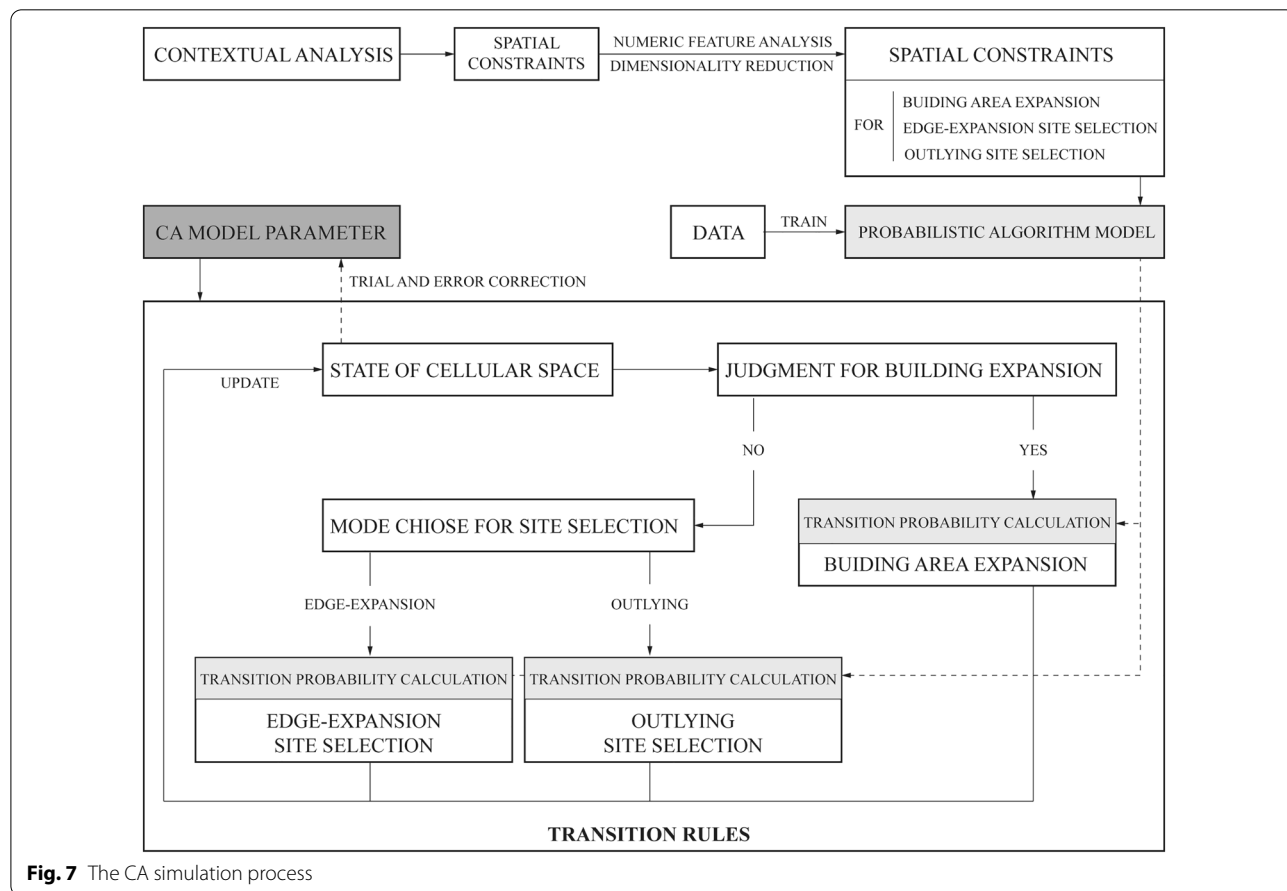
Fig. 7 The CA simulation process

and dimensionality reduction of attributes, the univariate Gaussian mixture model (UniGMM) was introduced in order to analyze the probability density distribution of building-cell samples in each attribute:

$$\text{UniGMMEval}(x) = \sum_{j=1}^{k} \omega_j N(x|\mu_j, \sigma_j^2), \sum_{j=1}^{k} \omega_j = 1 \tag{1}$$

where, there are certain k components. Each component is a Gaussian distribution parameterized by μ_j and σ_j^2, with ω_j as the weight of component j.

The peaks of the GMM curve indicate the feature value of the building-cell attributes, which can also be considered as the spatial preference for village growth. The dispersion degree of the attribute value distribution is expressed as γ, which indicates the sensitivity of a settlement distribution to each attribute:

$$\gamma = \frac{\sum_{i=1}^{k} \omega_i \sigma_i}{x_{max} - x_{min}}, \sum_{i=1}^{k} \omega_i = 1 \tag{2}$$

where, $\sum_{i=1}^{k} \omega_i \sigma_i$ is the standard deviation of the attribute values of building cells, and $x_{max} - x_{min}$ is the domain of the attribute values of building cells. The lower γ in the formula is, the more concentrated the attribute values of building cells, which indicates that the attribute has the stronger impact.

Calculation of the transition probability
Spatial state transition involves three cases: the growth of single building area, the site selection for new buildings in edge-expansion mode, and the counterpart in outlying mode. Accordingly, the state transition probability of candidate cells was also calculated in all three cases.

When it was considered that the size of the building area was going to increase, the neighboring function would be taken into account, and the attributes about building interval would not be. The calculation formula of transition probability is the following:

$$\text{Blocks\_prob}(i) = \text{GMMEval\_ATTRI}(i) \times \text{NeighborEval}(i) \tag{3}$$

In the case of site selection for new buildings in edge-expansion mode, the transition probability was calculated as follows:

$$\text{Blocks\_prob}(i) = \text{GMMEval\_ATTRI}(i) \times \text{GMMEval\_Ordi}(i) \tag{4}$$

In the case of site selection for new buildings in outlying mode, the transition probability was calculated as follows:

$$\text{Blocks\_prob}(i) = \text{GMMEval\_ATTRI}(i) \\ \times \text{GMMEval\_Nod}(i) \tag{5}$$

In the above three formulas, Blocks_prob(i) is the transition probability of the candidate cell, while GMMEval_ATTRI(i) is the contribution of environmental attributes (elevation, slope, the shortest distance to traffic facilities, water source, etc.) to the transition probability. GMMEval_Ordi(i) and GMMEval_Nod(i) are the contribution of social attributes (the average distance to the two nearest pre-existing nodal buildings and the nearest distance to the pre-existing ordinary building, etc.) to the transition probability of the candidate cell in edge-expansion mode and in outlying expansion mode. Finally, NeighborEval(i) is the transition probability contribution of the neighboring cell state to the central cell.

The above three formulas touch upon the issue of inferring the transition probability from the cell attribute value. Compared to cities, villages are on a smaller scale, and the data available for machine learning is lacking. In other words, the spatial data of a village for model calibration are not "large" enough. In order to obtain better simulation results with smaller data, to effectively circumvent overfitting [63], and to articulate the complex comprehensive effects of multiple factors, GMMEval_ATTRI(i), GMMEval_Ordi(i), and GMMEval_Nod(i) are calculated through the multivariable Gaussian mixture model (Multi-GMM) as follows:

$$\text{MultiGMMEval}\left(\vec{X}\right) = \sum_{j=1}^{k} \omega_j N(\vec{X}|\vec{\mu}_j, \vec{\Sigma}_j), \sum_{j=1}^{k} \omega_j = 1 \tag{6}$$

A Multi-GMM model is composed of k Multi-Gaussian distributions, where each Multi-Gaussian distribution is considered as a component j with the mixture weight ω_j and mean vector $\vec{\mu}_j$; $\vec{\Sigma}_j$ is a $d \times d$ covariance matrix of component j, which captures the correlation between different variables; and $\vec{X}$ is the vector formed by the multivariate attributes of cells.

The neighborhood of the CA model was defined by Moore (8 cells) neighborhood configuration mode. Supposing that the neighborhood effect on the central cell i (represented by NeighborEval(i)) was related to its adjacent cellular number (represented by NEI_i), then the neighborhood effect could be defined as follows:

$$\text{NeighborEval}(i) = \frac{1}{8}|\text{NEI}_i| \tag{7}$$

Judgement on the area growth and relevant parameters

In the CA model of this study, the growth of a building area was determined by the current size of the building and the covered area of the surrounding constructions. We set two prerequisites, and only if both of them were satisfied, can the building continue to grow. The first prerequisite states that, in order to present the importance of the size of a building, we compare the reciprocal of the building area with a random number from 0 to 1. If the former is smaller than the latter, the building will stop growing; otherwise, it will continue to grow. The second prerequisite states that, in order to simulate the influence of the surrounding construction amount, we set a module scale for area control, namely, the scale of an economic unit (a circular region with the centroid of the building as the center, and R as the radius), as well as the area threshold (A_{Ex}) of the construction in the module. If the current construction areas around the building within the control module were smaller than the threshold, the building was marked to grow; if not, it stopped growing. To simulate the uncertainty of the real world, we introduce random perturbations to the threshold.

Judgement on the expansion mode and relevant parameters

The modes of new building site included selection of edge-expansion site and outlying site. The feature boundary of the edge-expansion (normal building) and the outlying expansion (nodal building) is the shortest distance from the new building to the pre-existing buildings. Therefore, we distinguished two modes according to the distance threshold value (D_{Thres}). Furthermore, we found that the behavior of location choice in the outlying mode occurred approximately periodically. In addition, the number of spatial nodes in the early stage of village development is higher than in the late stage. The occurrence period of outlying expansion was set as T_{Inter} and different occurrence probabilities of outlying site selection as P_{pro} and P_{ana} in different periods, in order to control the spatial expansion in different developmental stages.

Model calibration

Rural building patches are scattered patches of low density. In addition, the scale of building patches is small. Therefore, compared to urban CA models, the simulation results of the village models are difficult to accurately coincide with the real situation [64]. In essence, it was important that the locations were similar and that the overall location patterns resembled each other in relevant ways [65]. CA models should be assessed on the basis of plausibility rather than on the one-to-one correspondence or correlation measures [66]. We believe that a simulation is valid if the geometric distance deviation

between the simulation results and the real situation is within an acceptable range. Hence, the Matching Index with Tolerance [67] (denoted as I_T) was adopted in the model assessment. The acceptable deviation is theoretically related to the scale of the economic unit. Being aware of that, we have delimited the acceptable deviation range as the union of circular regions with the centroid of the real building patches as the center, and the radius of the economic unit (R) as the radius. Simulated cells falling within the acceptable deviation range were considered effective (see Fig. 8). The percentage of effective cells was calculated as simulated accuracy.

Model comparison

Using the same simulation logic made it easier to compare the models of clustered villages and dispersed villages, in order to find differences and similarities between them. Model comparison was performed from cellular attribute variables and model parameters.

Regarding the cellular attribute variables, we were able to obtain two pieces of information. One was the sensitivity of the settlement distribution to each attribute, which was used for the validation of the constraints; the other was the feature value of each attribute variable that indicated the settlement distribution tendency.

The important model parameters need to be preset through context analysis and to be finalized through the process of trial and error. These parameters include the threshold to distinguish the outlying expansion and the edge-expansion (D_{Thres}), the radius of an economic unit (R), the area threshold within an economic unit (A_{Ex}), the occurrence period of outlying expansion (T_{inter}), and the occurrence probabilities of outlying site selection (P_{pro} and P_{ana}).

Results

Selection of cellular attributes

Context analysis and primary attribute selection

The clustered village of Hounan is nearly a single-family-name village, with the vast majority of the villagers having the same surname - Yang. Its development history can be traced back to the middle of the ruling Ming dynasty (approximately at the beginning of the sixteenth century). The village is located in the basin with high mountains to the east, south and west, and the Meitan River to the north. The arable land resources are relatively concentrated in the basin, but in shortage. Villagers mainly rely on tobacco cultivation for land development, and on the Meitan River and riverside markets to conduct local and out-of-town business, i.e., exportation. During the Kangxi-Qianlong period of the Qing dynasty, some villagers began going on long journeys in order to do business, mainly in Chaozhou, Suzhou, Hangzhou,

Fig. 8 Determination of effective simulation
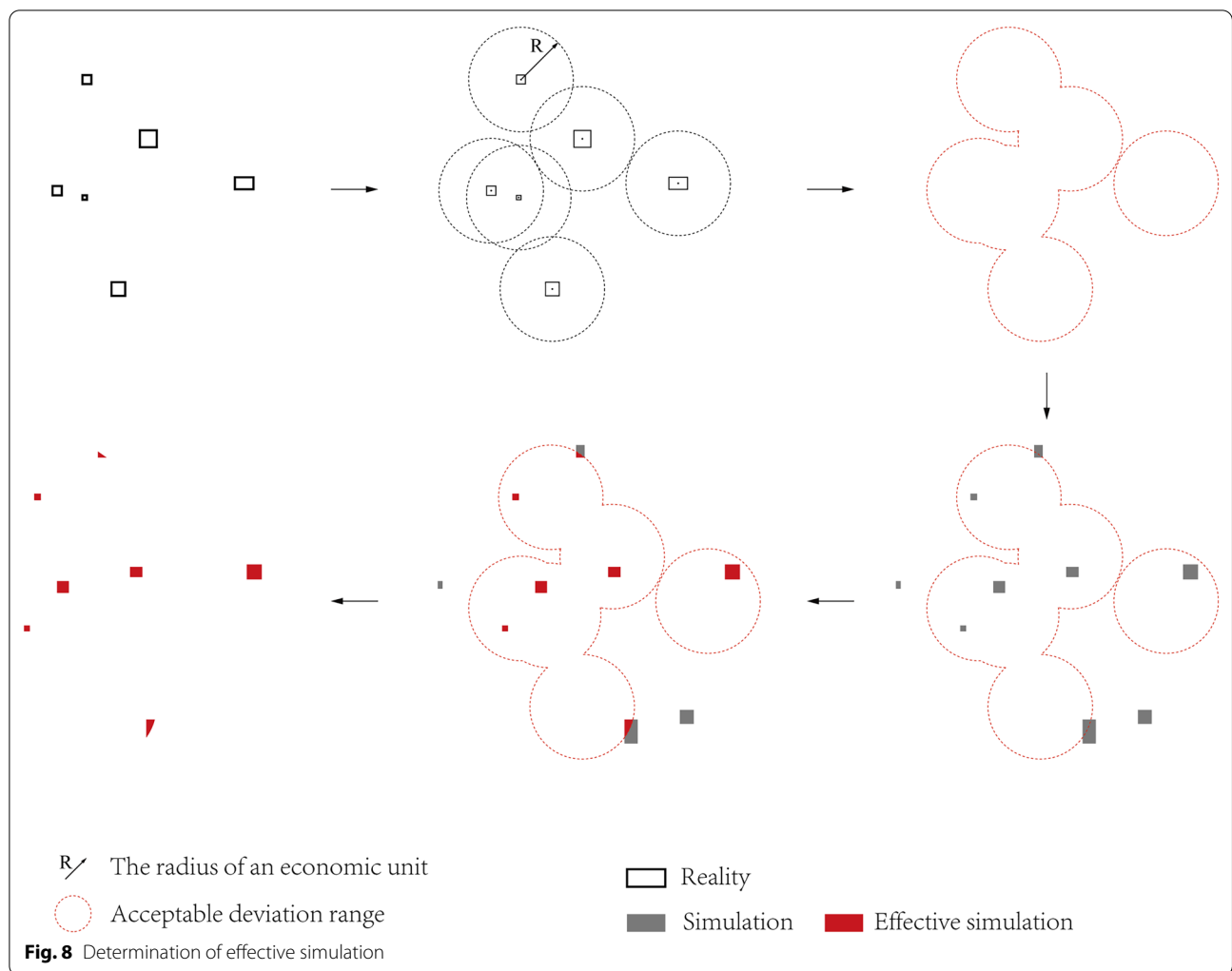
etc. After the Daoguang period of the Qing dynasty, they were mainly concentrated in Shanghai, Shantou and even Southeast Asia. In addition, a large number of families in the village have historically stepped into educational and political circles.

The dispersed Zhoutian village is also a single-family-name village, surnamed Ye. Founded in Kangxi period of the Qing dynasty (about 150 years later than Hounan village), Zhoutian village is located in the lower reaches of the Hakka migration route, a mountainous region without any broad river crossing, but with relatively more open space and more arable land resources. The arable land is scattered, forming an irregular shape. Many gullies carry springs down the hills, which became the main source of water for living and production. Throughout the history, Zhoutian was purely an agricultural village. Farmers lived on plantations of rice, sweet potato, peanut and beans. In the late Qing dynasty, villagers also began making a living overseas. Since then, the overseas capital

became the main support for the village construction and land development.

From the social and natural environment, it can be seen that Hounan is a commercial economic society, whereas Zhoutian is an agricultural economic society. Theoretically, the spatial distribution of the former might be influenced by factors related to commercial operations, such as market, water and land transport, etc. For the latter, factors related to agricultural operations, such as terrain, cultivating radius, arable land, water source, population, etc., could play an important role in space molding.

By observing the morphological development of the two villages, it was found that the positional relationship between the new buildings and the pre-existing buildings was not limited to the point-to-point relationship. In general, the development of things usually starts with structure generation, and then expands from that basic structure, and so does the growth of villages. If the new buildings far away from the pre-existing buildings were

Fig. 9 Generation of the spatial development skeleton
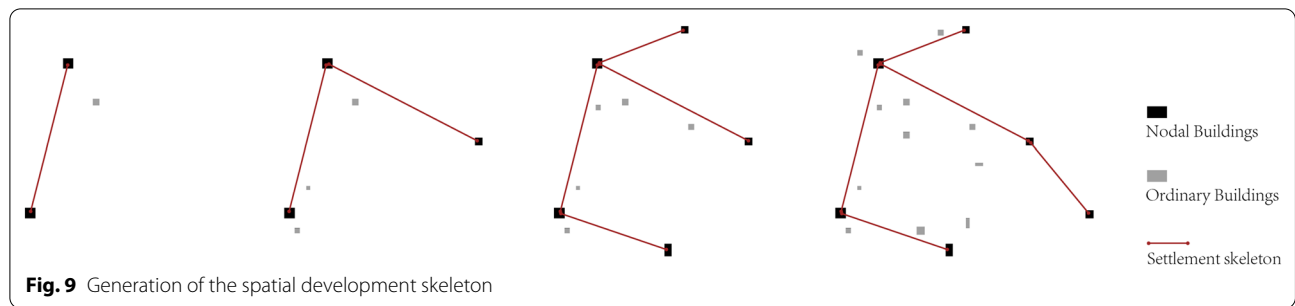
marked as nodal buildings, we could connect the new-born nodal building in each cycle of development with its nearest pre-existing nodal building, thus generating the spatial development skeleton (Fig. 9). Then it could be concluded that the choice of building locations seems to be related to the skeleton. The edge-expansion site selection was likely to be closer to the pre-existing framework, while the outlying site selection seemed to be further away from it.

Based on the above analysis, the cellular spatial attributes were preset for the two villages (Table 1).

Spatial attributes comparison

The attribute data of the cells in building patches of these two villages were extracted, respectively. Their probability density distributions were fitted by the UniGMM (Figs. 10, 11). The feature value *F* and the sensitivity value γ of each attribute are shown in Table 2.

The sensitivity value indicates that most of the spatial attributes could exert a strong constraint on the spatial evolution. In comparison, the development pattern of Zhoutian was more sensitive to the water source factor

(Dc\Dp), whereas the spatial development of Hounan was more sensitive to the cultural factors revealed by the construction condition (Dnj\Dmc\Dnc). Both villages were similarly sensitive to the transportation factors (Dr). Only Dmj has a weak constraint on the development of Zhoutian, and this attribute was removed in the next step of the spatial simulation of this village.
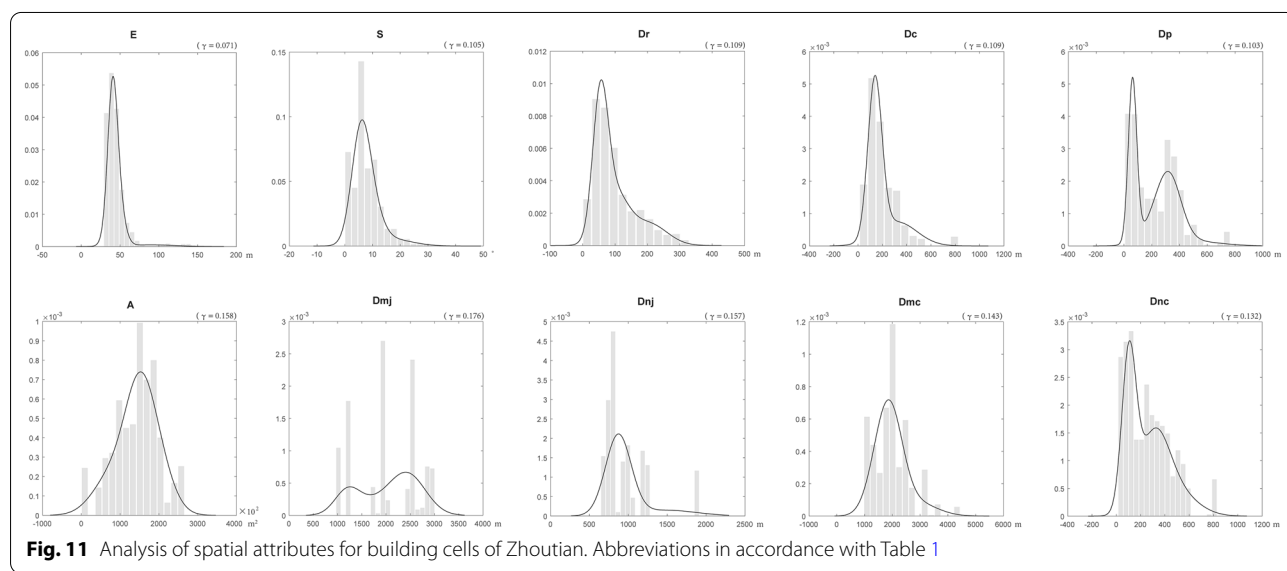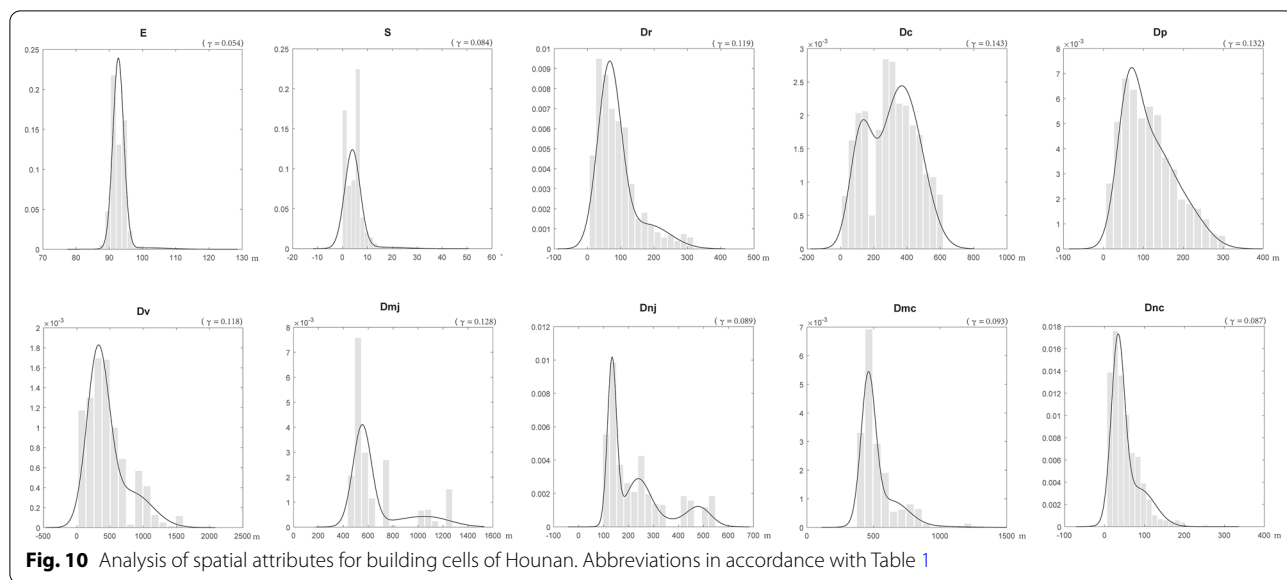
The feature value shows that, in terms of natural factors, the building distribution in Zhoutian tended to be closer to water sources and main roads, and its tolerable slope gradient for constructions was higher. In terms of social factors, the scale of leap distance of the outlying expansion in Zhoutian was about 5 times larger than that in Hounan.

Comparison of CA model parameters

In order to obtain a better simulation effect, we had to preset the CA parameter values based on the historical multi-temporal data observation. Through several processes of trial and error, the parameter values leading to a relatively optimal fitting result (see Fig. 12) were stored (Table 3).

Table 1 Pre-set cellular spatial attributes of the CA model

Common attributes	Individual attributes	
	Hounan (clustered)	**Zhoutian(dispersed)**
E—elevation	Dv—shortest distance to the river	A—area of the arable land within the cultivating radius
S—slope		
Dr—shortest distance to main roads		
Dc—shortest distance to canals		
Dp—shortest distance to ponds		
Dmj—the average distance between a new nodal building and two nearest pre-existing nodal buildings		
Dnj—the nearest distance between a new nodal building and a pre-existing ordinary building		
Dmc—the average distance between a new ordinary building and two nearest pre-existing nodal buildings		
Dnc—the nearest distance between a new ordinary building and a pre-existing ordinary building		

Fig. 10 Analysis of spatial attributes for building cells of Hounan. Abbreviations in accordance with Table 1

Fig. 11 Analysis of spatial attributes for building cells of Zhoutian. Abbreviations in accordance with Table 1

Under the control of the parameter values, the simulation accuracy (the Matching Index with Tolerance, denoted as I_T) of the two villages was 80.54% for Hounan and 74.19% for Zhoutian compared to the real-life situation.

It should also be noted that in the urban spatial simulation conducted from the middle stage of development, if the absolute matching index or the Matching Index with Tolerance reaches 70–90% of correspondence, the simulation can be considered effective[55, 67, 68]. However, rural spatial simulation adopts a more micro scale level. Furthermore, in order to display the whole spatial process lasting for hundreds of years, the spatial process on a long temporal scale was simulated from the early stage. Therefore, compared to the simulation of large-scale land use over several decades, the simulation might inevitably be largely flawed. In our opinion, an ideal simulation result obtained from the Matching Index with Tolerance is acceptable if its correspondence is not lower than 60%, and more effective if it is higher than 70%.

Discussions

We have attempted to use a common logical structure to build a CA model for the simulation of different types of villages. Through cellular attribute selection and model

Table 2 The sensitivity value and the feature value of each attribute of building cells

Spatial attribute	Sensitivity value *γ*		Feature value *F*	
	Hounan	Zhoutian	Hounan	Zhoutian
Dr	0.119	0.109	67.3 m	58.2 m
Dv	0.118	–	331.1 m	–
Dc	0.143	0.109	369.9 m	140.6 m
Dp	0.132	0.103	71.8 m	62.4 m
E	0.054	0.071	92.7 m	41.0 m
S	0.084	0.105	4.0°	6.3°
A	–	0.158	–	153,050.0 m²
Dmj	0.128	0.179	551.2 m	2404.1 m
Dnj	0.089	0.157	134.5 m	875.0 m
Dmc	0.093	0.143	463.6 m	1867.0 m
Dnc	0.087	0.132	34.1 m	110.3 m

Abbreviations in accordance with Table 1

parameter calibration, we found the similarities and differences between clustered and dispersed villages.

Analysis of similarities

In general, both clustered and dispersed villages were founded on the blood relations of the Hakka people. Their spatial development followed a common repetitive process of: site selection for nodal buildings—defining the economic sphere of influence—new ordinary buildings filled in the sphere until approaching saturation point. The main driving force for the spatial splitting and dispersing was the periodic separation of family registration and family property in Chinese traditional families [69]. In the earlier stage of the development, a village mainly presented outlying expansion; in the later stage, the outlying expansion gradually slowed down and edge-expansion began to increase.

At the level of settlement pattern, the newborn "ordinary buildings" in each growth cycle of the two types of villages would retreat towards the already existing spatial skeleton which connect the adjacent "nodal buildings" (see Fig. 9). In consanguineous villages (like Hounan and Zhoutian), although the descendant economic unit had broken away from the mother unit and became independent, it still demanded that the connection with its mother unit and neighboring relative units stay continuous. Once this connection was created, the subsequently generated dependent entities within each economic unit would utilize and further strengthen it, which is manifested in the formation of new ordinary buildings near the structural skeleton. Secondly, at the level of the relationship between settlements and the environment, transportation factors had a strong control over the site

selection. Both in the clustered and dispersed villages, the settlement distribution was close to main roads, which reflects the influence of land transportation on people's production and life.

Analysis of differences

Although the logical framework of the periodic development of clustered and dispersed villages is similar, there are some obvious differences.

In respect to the settlement pattern of the dispersed village, the time of the spatial expansion cycle is longer, the building scale is larger, the outlying growth occurs more often, the scale of geometric distance between economic units is higher, and the proportion of people and land is smaller when the economic unit reaches saturation. The clustered village is completely opposite to it. However, the spatial skeleton and distribution of pre-existing buildings in the clustered village were significantly more important than those in the dispersed village.

The above differences were actually influenced by local industry. The architectural scale reflects the organizational scale of the family unit, and the geometric distance between the nodal buildings reflects the scale of the family spatial sphere of influence. The clustered village had relatively few land resources and was supported by commercial capital. The dispersed village was relatively rich in arable land resources and its population mostly survived from the agricultural economy. Agricultural operations require more manpower than business managements, so a family unit in the dispersed village could absorb more cooperative labor, leading to larger building scale and a longer spatial expansion time cycle. Under the influence of the locality of agricultural activities, the capital strength of the agricultural economy was directly reflected in the area of arable land occupied by a family unit (composed of several nearby buildings). Therefore, in order to retain the capital power and have the opportunity to expand, the family unit would consciously control the proportion of people and land in its economic unit, which would lead to more frequent outlying expansions and greater distances between buildings and between economic units. In other words, the frequent outlying expansion and the large leap distance enabled each economic unit of the agricultural village to maintain stronger independence. Moreover, the correlation in spatial growth between economic units was significantly weaker than in commercial villages.

In terms of the relationship between the settlement and the environment, the two types of villages showed different performance on different economic bases. Based on agriculture, the attributes of the dispersed village showed a strong attachment to the terrestrial environment. First of all, the condition of the cultivated land was the basis
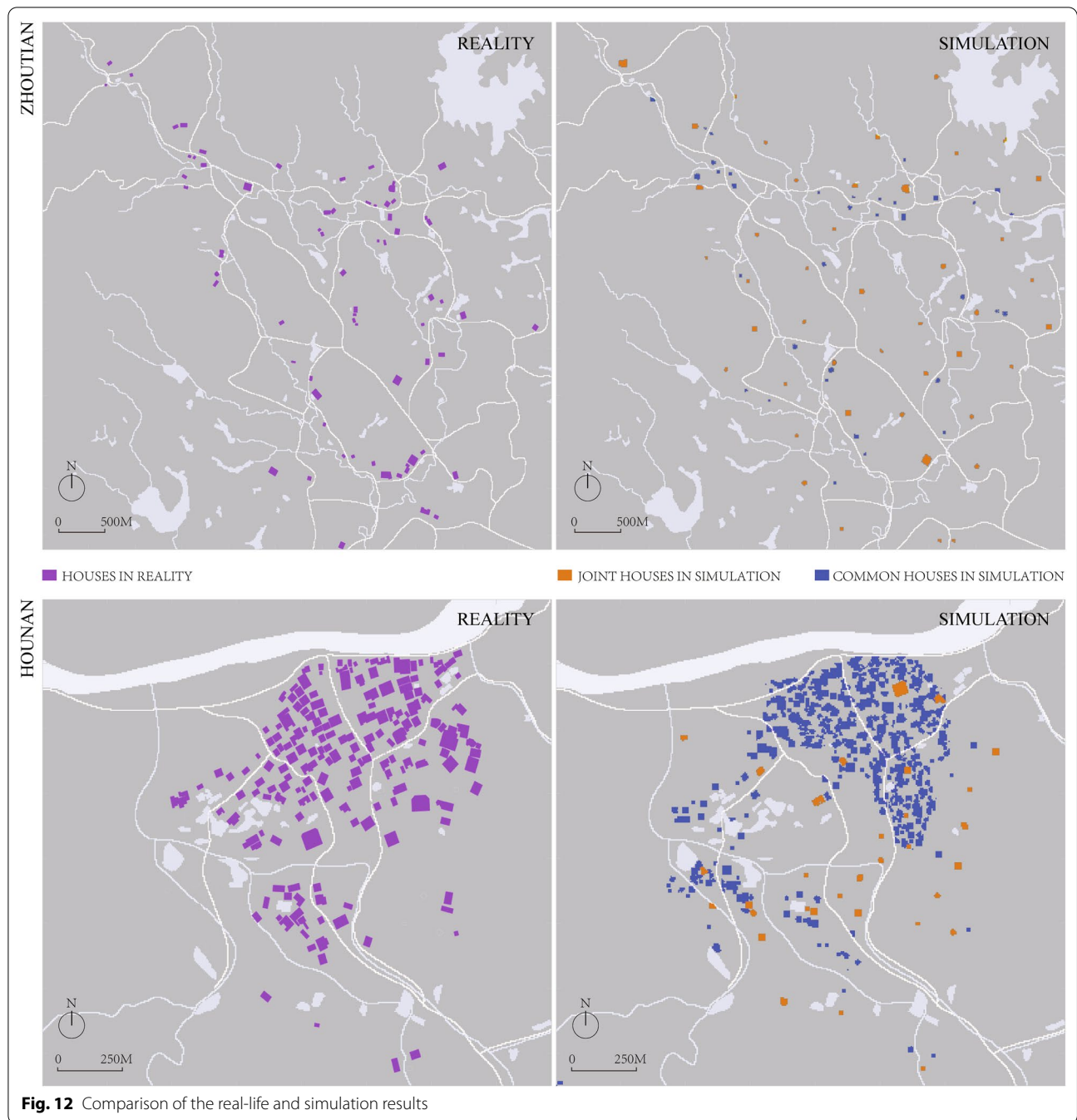
Fig. 12 Comparison of the real-life and simulation results

Table 3 The parameter values of the model with relatively optimal fitting results

Parameter	Meaning	Hounan value	Zhoutian value
D_{Thres}	The threshold to distinguish the outlying expansion and the edge expansion	120 m	600 m
R	The radius of an economic unit	60 m	300 m
A_{Ex}	The area threshold within an economic unit	4750 m^2	12,800 m^2
T_{Inter}	The occurrence period of outlying expansion	30 years	60 years
$P(P_{pro}|P_{ana})$	The occurrence probabilities of outlying site selection (early | late)	5%|10%	63%|32%

for its location choice, and there was plenty of arable land around its settlement. In order to save farmland, the dispersed village could develop its construction on relatively steep slopes, while clustered villages generally avoided such areas. For the same reason, the density of the road network of the dispersed village was very low. Therefore, in order to improve the convenience of transportation, settlements in the dispersed village were closer to the main road than those in the clustered village. Second, the water source has become another dominant factor in site selection for dispersed villages. The model data reveal that water sources were more important in the dispersed village, and the settlements were spatially closer to it.

Separated from external economic bodies and rooted in the local farming environment, the dispersed village focused on the operation of the internal world. It was more like a village union formed by numerous small economic groups through weak kinship ties, in which the population and economic capacity of each group were close to a micro-village.

Discussions on the questions raised in the introduction
Uncertainties in the transition of spatial patterns
In discussing the way traditional villages expand, previous studies have usually looked at this issue from the perspective of external action over the centuries, revealing that there is a transitional relationship between clustered and dispersed villages caused by social and environmental changes or sudden disasters (see in Section "Description of spatial patterns and expansion") [31–33]. These studies have divided the village evolution into two different stages before and after the interference of external actions. In other words, the village can gradually change from clustered to dispersed, and vice versa.

During the course of this research, the morphological changes of villages were explored from the perspective of the internal force of the subject in spatial constructions when both natural and social environments were relatively stable. The simulation results supported our hypothesis. Even if there are no significant external disturbances, the expansion mode of villages will be a mix of "outlying + edge-expansion". The mode change is not a one-way transition, but a continuous cycle of switching between the two modes. The transitional relationship between clustered and dispersed is not linear in nature and the transition between spatial patterns could occur in different periods of the same spatial unit. Accordingly, the final state of the village can be clustered or dispersed. As this study shows, Hounan continued to have a dispersed pattern in the early and middle stages of development. Due to the short outlying distance between economic units and the weak control of the population growth in each economic unit, the dispersed village gradually developed into a clustered one with several scattered units on the periphery. On the contrary, due to the much longer outlying distance and stronger control of population growth, the "dispersed" pattern of Zhoutian was relatively stable with an open spatial boundary and a stable economic mode.

Functions of cultural and natural factors
What role did contextual factors play in relation to environmental factors in the spatial evolution of the villages? Previous studies have differing opinions on this issue (see in Section "Judgement on the driving factors of spatial evolution") [44, 45]. Based on this study, we believe that people of the same ethnic origin may stick to their inherent common sense when constructing and developing living spaces, which is reflected in a similar logical framework of spatial evolution. However, in a specific spatial practice, people will choose individual economic modes according to the local site conditions, and then they will move towards a characteristic space under different intensity restrictions of different factors.

Guiding value of the semi-intelligent CA model
We had to base our understanding and predictions on a reliable reason. No matter how accurate the black box algorithm may be [59], it is ill-founded and it cannot make us believe that it is not a "shot in the dark" that happened to be accurate. The semi-intelligent CA model provides not only results, but also logical reasoning behind the results. Furthermore, the logical rules can be manually adjusted to make the model applicable to the simulation of different types of spaces, to compare and explore similarities and differences between different villages. Model flexibility and analyzability are especially important for historical spatial research. Moreover, the fully-intelligent model needs to obtain reliable simulation results based on a large amount of data. A large database is not a problem for urban space, but for rural space, where the available data is often not large enough to capture the rules by simple data mining. It makes sense to preset reliable rules and then use data to correct them.

The algorithm of the Gaussian mixture model embedded into the CA is not only a probabilistic model with strong explanatory power, but also suitable for adjusting characteristics based on limited data. However, we must recognize that although the CA model in this research was based on several village samples with available historical data and verified in the case of two villages, the universality of the model still needs to be further validated with more cases in the future. For villages with different patterns and spatial processes, our study provided a common analysis framework, which is also the construction logic of the CA model proposed in the study.

We believe that based on this understandable and logical framework, ideas and methods for recognizing and comparing features through spatial simulation of traditional villages can become clearer. This model can potentially be used for the division of village types and for the systematic zoning of traditional villages, which is the direction of our future research.

Application and limitations of the model

The model proposed in this study is based on the principle of bottom-up growth. It is applicable to the recognition, comparison and conservation planning of traditional villages that have naturally evolved according to this principle. Models are nothing more than auxiliary tools that can refine, integrate prior knowledge, and transfer it to future decisions. However, the model we created to extract, analyze and compare the dynamic mechanisms of historical spatial processes is an important tool for providing historical references for spatial planning in the new era, but not the only basis.

If the historical spatio-temporal model continues to function, evolution will continue according to historical rules, providing a spatial scenario guided by traditional values. However, this scenario is nothing more than an ideal state. The contemporary morphological development of traditional villages in accordance with the conservation planning laws and guidelines is a top-down process, which will be improved by planning constraints. For example, in the historical period the boundary between villages were vague, and the spatial expansion of clans was relatively free. However, the development of contemporary villages has clear boundaries that cannot be cross discretionarily. Another example is that during the period of the self-subsistence peasant economy, land transaction was free and the site selection for the construction of housing was relatively free and independent. However, nowadays, collectivization and nationalization of land ownership limit the arbitrary change of land attribute. In addition, when the cultural relic value of a certain building or districts in the village receives legal accreditation, there will be limited spatial constructions or deconstruction within the protection area and the development control area.

Therefore, the model given in this study can be used to make spatial analysis and generate scenarios on historical bases before planning. Based on historical-logically oriented scenarios, planners can make further adjustments according to the laws and guidelines of spatial protection in the new era.

Conclusions

By reconstructing the rural settlement evolution, we realized that the two villages of the same ethnic group located in different regions could have certain similarities in the logical framework of spatial development, which was significantly done in the outlying expansion of the nodal building with edge-expansion of ordinary buildings. Based on such assumptions, we proposed a common logical structure to create the CA model. Taking into consideration the characteristics of villages in different regions, we selected specific cellular attributes as constraints on spatial growth and calibrated the model parameters to apply the model in different types of villages.

Based on the simulation results, we proved our hypothesis on a common spatial logic and determined that the change of spatial development mode is a continuous switching cycle, which leads to a nonlinear transition relationship between clustered and dispersed patterns. For the two types of villages with different patterns, the similarities are that the expansion process was under strong control of a transportation skeleton constructed by the main roads. The differences lie in the fact that the dispersed village was more strictly constrained to the building area and the proportion of the population on farmland area within the economic units. Meanwhile, its spatial relationship between settlements and external natural elements was comparatively closer. On the contrary, the spatial growth of the clustered village was more closely related to the spatial structure of the pre-existing buildings. In fact, these differences are the result of different economic modes chosen on the basis of different site conditions.

This research also reveals that when people of the same ethnic group deal with rural spatial management problems in different areas, the endogenous ethnic cultural factors dominate the relatively stable spatial logic rules, while the differences in regional environmental factors further cause variations in spatial evolution.

Declarations

Author details
[1] School of Architecture, Harbin Institute of Technology (Shenzhen), 518055 Shenzhen, China. [2] School of Software, Tsinghua University, Beijing 100084, China.

References

1. Katapidi I. Heritage policy meets community praxis: widening conservation approaches in the traditional villages of central Greece. J Rural Stud. 2021;81(1):47–58.
2. He Y, Zhang T, Xiong D. Evaluation on cultural value of traditional villages and differential revitalization: a case study of Jiaozuo City Henan Province. Econ Geogr. 2020;40(10):230–9 (**in Chinese**).
3. Liu C, Xu M. Characteristics and influencing factors on the hollowing of traditional villages-taking 2645 villages from the Chinese traditional village catalogue (Batch 5) as an example. Int J Env Res Pub He. 2021;18(23):12759.
4. Gao J, Wu B. Revitalizing traditional villages through rural tourism: a case study of Yuanjia Village, Shaanxi Province China. Tourism Manage. 2017;63:223–33.
5. Luo P, Zheng Y. Visualization research on the literature review of chinese traditional village tourism development based on CiteSpace. Geogr Geo-Inf Sci. 2020;36(1):129–35 (**in Chinese**).
6. Guo Z, Sun L. The planning, development and management of tourism: the case of Dangjia, an ancient village in China. Tourism Manage. 2016;56(5):52–62.
7. Liu P, Dong S. Study of landscape-image of Chinese ancient village. Geogr Res-Aust. 1998;17(1):31–8 (**in Chinese**).
8. Qu Z, Peng Z, Zhou Z. The Winning Project of 4th Holcim Awards for sustainable construction: the post-earthquake reconstruction design of historic Xueshan village in Ya'an. Sichuan World Archit. 2015;300(6):135–7 (**in Chinese**).
9. Du F, Kobayashi H, Okazaki K, Ochiai C. Research on the disaster coping capability of a historical village in a mountainous area of China: case study in Shangli, Sichuan. Procedia Soc Behav Sci. 2016;218(3):118–30.
10. Liu S, Ge J, Bai M, Yao M, He L, Chen M. Toward classification-based sustainable revitalization: assessing the vitality of traditional villages. Land Use Policy. 2022;116(5): 106060.
11. Zhou R, Zhong L, Liu J. Research on rural world heritage sites: connotation and tourism utilization. Geogr Res-Aust. 2015;34(5):991–1000 (**in Chinese**).
12. Yang X, Pu F. Cellular automata for studying historical spatial process of traditional settlements based on gaussian mixture model: a case study of Qiaoxiang village in Southern China. Int J Archit Herit. 2020;14(4):568–88.
13. Akpomuvie BO. Self-Help as a strategy for rural development in Nigeria: a bottom-up approach. J Altern Perspect Soc Sci. 2010;2(1):88–111.
14. Altieri MA, Masera O. Sustainable rural development in Latin America: building from the bottom-up. Ecol Econ. 1993;7(2):93–121.
15. Huang H, Macmillan W. A generative bottom-up approach to the understanding of the development of rural societies. Agrifood Res Rep. 2005;68(1):296–312.
16. Guan Z, Wang T, Zhi X. Temporal-spatial pattern differentiation of traditional villages in central plains economic region. Econ Geogr. 2017;37(9):225–32 (**in Chinese**).
17. Luo Y, He J, He Y. A rule-based city modeling method for supporting district protective planning. Sustain Cities Soc. 2017;28(1):277–86.
18. Zhang D, Liu X, Wu X, Yao Y, Wu X, Chen Y. Multiple intra-urban land use simulations and driving factors analysis: a case study in Huicheng China. Gisci Remote Sens. 2019;56(2):282–308.
19. Yang R, Liu Y, Long H, Qiao L. Spatio-temporal characteristics of rural settlements and land use in the Bohai Rim of China. J Geogr Sci. 2015;25(5):559–72.
20. Yang R, Xu Q, Long H. Spatial distribution characteristics and optimized reconstruction analysis of China's rural settlements during the process of rapid urbanization. J Rural Stud. 2016;47(5):413–24.
21. Berling-Wolff S, Wu J. Modeling urban landscape dynamics: a review. Ecol Res. 2004;19(1):119–29.
22. Xiang Y, Cao M, Zhai Z, Yi C. The landscape genome maps construction and characteristics analysis of Shanxi traditional rural cave dwelling settlements. Hum Geogr. 2019;34(6):82–90 (**in Chinese**).
23. Yang X, Ma H, Zhang L, Song K. Inheritance, fusion, and variation during migration: an analysis of spatio-temporal morphological development of traditional Guangdong Hakka architecture and settlements. Geogr Res-Aust. 2021;40(4):958–76 (**in Chinese**).
24. Lu X. Dispersed and Clustered: Rural Settlement Patterns and Its Evolution in Traditional China. J Huazhong Norm Univ Humanit Soc Sci. 2013;52(4):113–30 (**in Chinese**).
25. Deng Y, Fan J, Zhang S, Fang X, Chen Z. Timing and patterns of the great Ordovician biodiversification event and Late Ordovician mass extinction: perspectives from South China. Earth-Sci Rev. 2021;220(9): 103743.
26. Liu X, Li X, Chen Y, Tan Z, Li S, Ai B. A new landscape index for quantifying urban expansion using multi-temporal remotely sensed data. Landscape Ecol. 2010;25(5):671–82.
27. García AM, Santé I, Boullón M, Crecente R. A comparative analysis of cellular automata models for simulation of small urban areas in Galicia, NW Spain. Comput Environ Urban Syst. 2012;36(4):291–301.
28. Demangeon A. Human Geography Problems. Beijing: The Commercial Press; 1993. (**in Chinese**).
29. Miyazaki I. The age of city-states in China. Hist Rev. 1950;33(2):144–63 (**in Japanese**).
30. Wang Q. Villages in North China in the late Qing dynasty: history and size. Hist Res. 2007;306(2):78–87 (**in Chinese**).
31. Huang Z. The Subdivision and concentration of villages in Huabei Plain in the Ming and Qing dynasties. Stud Qing Hist. 2005;58(2):21–31 (**in Chinese**).
32. Chen C, Xiao W. Settlement patterns and social transformation: the historical impact of local unrest in the Hanjiang river basin during the Ming and Qing dynasties. J Hist Sci. 2011;364(2):55–68 (**in Chinese**).
33. Carlos EC, Jeffrey RP. Geoarchaeology of an Aztec dispersed Village on the Texcoco Piedmont of Central Mexico. Geoarchaeology. 1997;12(3):177–210.
34. Duan J, Gong K, Chen X, Zhang X. Spatial Research (Spatial Analysis the World Cultural Heritage Village of Xidi). Nanjing: Southeast University Press; 2016. (**in Chinese**).
35. Zong L, Jiao Y, Li S, Zhang H, Zhang H, He Y, Niu L. The rural settlement landscape and its evolution in Hani rice terrace culture landscape areas: a case study of the Quanfuzhuang middle village, Yuanyang county, Yunnan. Trop Geogr. 2014;34(1):66–75 (**in Chinese**).
36. Chen Z, Li Q. Three villages in Meixian County. Beijing: Tsinghua University Press; 2007. (**in Chinese**).
37. Liu L. Researches of Hakka in Shenzhen. Shenzhen: Haitian Press; 2013. (**in Chinese**).
38. Hu S, Zheng Q. Living by Water, thriving by water – the influence of water system on the development and construction of basic economic units in Jiangnan from the perspective of the spatial pattern of some ancient water towns. China Anc City. 2019;211(4):54–9 (**in Chinese**).
39. Ma X, Qiu F, Li Q, Shan Y, Cao Y. Spatial pattern and regional types of rural settlements in Xuzhou City, Jiangsu province China. Chinese Geogr Sci. 2013;23(4):482–91.
40. He X. The logic of farmers' action and the regional difference in village administration. Open Times. 2007;187(1):105–21 (**in Chinese**).
41. Lipsky Z. The changing face of Czech rural landscape. Landscape Urban Plan. 1995;31(1):39–45.
42. Vliet VJ, Verburg PH, Grădinaru SR, Hersperger AM. Beyond the urban-rural dichotomy: towards a more nuanced analysis of changes in built-up land. Comput Environ Urban. 2019;74(2):41–9.

43. Mei X, On A. Demangeon's thoughts of human geography and environmental history. World History. 2004;166(3):13–24 (**in Chinese**).
44. Guo X, Zhang Q, Ma L. Analysis of the spatial distribution character and its influence factors of rural settlement in transition-region between mountain and hilly. Econ Geogr. 2012;32(10):114–20 (**in Chinese**).
45. Wang Y. Spatial concept in traditional settlement structure. Beijing: China building industry press; 2015. (**in Chinese**).
46. Wu F. An empirical model of intrametropolitan land-use changes in a Chinese city. Environ Plann B Plann Des. 1998;25(2):245–63.
47. Clarke KC. The limits of simplicity: toward geocomputational honesty in urban modeling. In: Atkinson P, Foody G, Darby S, Wu F, editors. GeoDynamics. Florida: CRC Press; 2004. p. 215–32.
48. Pedro D. The master algorithm: how the quest for the ultimate learning machine will remake our world. Beijing: Citic Publishing Housein; 2017. (**in Chinese**).
49. Batty M, Xie Y. From cells to cities. Environ planning B. 1994;21(7):31–48.
50. Couclelis H. Cellular worlds: a framework for modeling micro - macro dynamics. Environ Plan A. 2008;17(5):585–96.
51. Santé I, García AM, Miranda D, Crecente R. Cellular automata models for the simulation of real-world urban processes: a review and analysis. Landscape Urban Plan. 2010;96(2):108–22.
52. Liu Y, Feng Y, Pontius R. Spatially-explicit simulation of urban growth through self-adaptive genetic algorithm and cellular automata modelling. Land. 2014;3(3):719–38.
53. Liu Y, Kong X, Liu Y, Chen Y. Simulating the conversion of rural settlements to town land based on multi-agent systems and cellular automata. PLoS ONE. 2013;8(11): e79300.
54. Zheng HW, Shen GQ, Wang H, Hong J. Simulating land use change in urban renewal areas: a case study in Hong Kong. Habitat Int. 2015;46(2):23–34.
55. Süha B, Anıl A, Keith CC. Cellular automata modeling approaches to forecast urban growth for adana, Turkey: a comparative approach. Landscape Urban Plan. 2016;153(9):11–27.
56. Qian M, Pu L, Zhang J. Urban spatial morphology evolution in Suzhou-Wuxi-Changzhou region based on improved landscape expansion index. Scientia Geographica Sinica. 2015;35(3):314–21 (**in Chinese**).
57. Wu P, Zhou D, Gong H. A new landscape expansion index: definition and quantification. Acta Ecol Sin. 2012;32(13):4270–7 (**in Chinese**).
58. Molnar C. 2019. Interpretable machine learning. A Guide for Making Black Box Models Explainable.
59. Li X, Yeh AG. Calibration of cellular automata by using neural networks for the simulation of complex urban systems. Environ Plann a. 2001;33(8):1445–62.
60. Luo X. The origin and development of Hakka. Beijing: The Chinese Overseas Publishing House; 1989. (**in Chinese**).
61. Situ S. History and Human Geography of Lingnan. Guangzhou: Sun Yat-sen University Press; 2001. (**in Chinese**).
62. Long Y, Jin X, Yang X, Zhou Y. Reconstruction of historical arable land use patterns using constrained cellular automata: a case study of Jiangsu, China. Appl Geogr. 2014;52(8):67–77.
63. Tuggle K, Zanetti R. Automated splitting Gaussian mixture nonlinear measurement update. J Guid Control Dyn. 2018;41(3):725–34.
64. Gao X, Liu Y, Liu L, Li Q, Deng O, Wei Y, Ling J, Zeng M. Is big good or bad?: Testing the performance of urban growth cellular automata simulation at different spatial extents. Sustainability-Basel. 2018;10(12):4758.
65. White R, Engelen G, Uljee I, Lavalle C, Erlich D. Developing an Urban Land Use Simulator for European Cities. In: Fullerton, K. (Ed.), Proceedings of the the fifth EC GIS Workshop: GIS of tomorrow. Stresa, Italy: European Commission Joint Research Centre; 2000: 179–90.
66. Wu F. Calibration of stochastic cellular automata: the application to rural-urban land conversions. Int J Geogr Inf Sci. 2002;16(8):795–818.
67. Lagarias A. Urban sprawl simulation linking macro-scale processes to micro-dynamics through cellular automata, an application in Thessaloniki, Greece. Appl Geogr. 2012;34(5):146–60.
68. Liu D, Zheng X, Zhang C, Wang H. A new temporal–spatial dynamics method of simulating land-use change. Ecol Model. 2017;350(2):1–10.
69. Ai Y, Guo Y. Prohibition of family registration separation and family property division in Tang Code. Chin J Law. 2010;32(5):164–9 (**in Chinese**).

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.