

# Predicting the Metro Scenario-based Spatiotemporal Evolution of Land Use Using CA-Kalman Filter Model: a Case Study of Nanjing City, China

Jialing Dai (✉ [jjialingdai@163.com](mailto:jjialingdai@163.com))

NJU-ECE Institute for Underground Space and Geo-environment, Nanjing University

Xiaozhao Li

Chinese Academy of Geological Sciences

Wei Zhang

NJU-ECE Institute for Underground Space and Geo-environment, Nanjing University

Ke Liao

NJU-ECE Institute for Underground Space and Geo-environment, Nanjing University

Tao Liu

NJU-ECE Institute for Underground Space and Geo-environment, Nanjing University

Peng Zhao

NJU-ECE Institute for Underground Space and Geo-environment, Nanjing University

Wentao Xu

NJU-ECE Institute for Underground Space and Geo-environment, Nanjing University

---

## Research Article

**Keywords:** Land use change, Metro scenario analysis, Cellular automata, Kalman filter, Spatiotemporal evolution

**Posted Date:** October 22nd, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-970357/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

1 **Predicting the metro scenario-based spatiotemporal evolution of land use**  
2 **using CA-Kalman filter model: A case study of Nanjing City, China**

3

4 Jialing Dai<sup>1\*</sup>, Xiaozhao Li<sup>2,3</sup>, Wei Zhang<sup>1</sup>, Ke Liao<sup>1</sup>, Tao Liu<sup>1</sup>, Peng Zhao<sup>1</sup>, Wentao Xu<sup>1</sup>

5 <sup>1</sup> *NJU-ECE Institute for Underground Space and Geo-environment, School of Earth Sciences*  
6 *and Engineering, Nanjing University, Nanjing 210023, China*

7 <sup>2</sup> *Chinese Academy of Geological Sciences, Beijing, 100037, China*

8 <sup>3</sup> *State Key Laboratory for Geomechanics and Deep Underground Engineering, China*  
9 *University of Mining & Technology, Xuzhou 221116, China*

10

11 Corresponding author:

12 Jialing Dai\*

13 163 Xianlin Road, Qixia District, Nanjing, 210023 Jiangsu Province, P. R. China

14 Email: jialingdai@163.com

15

16

17

18

19

20

21

22

23 **Abstract**

24 The expansion of metro system can bring varying degrees of impact to the surrounding  
25 environment. To study this complex system problem, this paper discusses the temporal and  
26 spatial impact by metro system from the perspective of land use change simulation and scenario  
27 analysis. The traditional cellular automata (CA) model can realize the simulation of land use  
28 change under various scenarios through system dynamics or Markov chain to control the long-  
29 term demand forecasting. However, this type of model ignores the filtering of noise data from  
30 imageries and increases uncertainty of the system. Therefore, based on the Future Land Use  
31 Simulation (FLUS) model, this paper integrates Kalman filter to control the stochastic process  
32 of the state-space system, and predicts the spatio-temporal evolution of land use change  
33 impacted by metro system in Nanjing from 2019 to 2035. The results show that: (1) The  
34 proposed CA-Kalman filter model can realize the optimized simulation of land use change with  
35 good accuracy; (2) Urban patches impacted by metro system will emerge from the existing  
36 urban boundaries at the cost of occupation of cultivated land, although there is still significant  
37 expansion of urban land and construction land, it will reach the upper limit in 2050.

38 **keywords:** Land use change; Metro scenario analysis; Cellular automata; Kalman filter;  
39 Spatiotemporal evolution

40

41

42

43

44

## 45 **1 Introduction**

46 At the end of year 2020, Caojiawan Station near the city of Chongqing in China suddenly  
47 attracted great attention across the globe, for its contrasting nearby image against five years ago  
48 <sup>1</sup>. The place where it has been constructed is embracing flourishing developments since the  
49 opening of metro exits, which arouses people's concern about how the neighboring  
50 environment responds to metro construction and development? Will this process exacerbate  
51 urban sprawl in the long run?

52 It is estimated that 70% of the world's population will live in urban areas by 2050 <sup>2</sup>. As a  
53 result, metro system is increasingly viewed as one of solutions to address the demand for more  
54 resilient space and dependable infrastructure <sup>3,4</sup>. Metro system intertwines through urban areas  
55 and expands to the fringe of city, burden with population migration and intangible capital flow.  
56 As the urban traffic axes and development axes of Urban Underground Space (UUS), metro  
57 system should make city compact and better <sup>5</sup>. Thus, it appeals to scholars and stakeholders into  
58 its planning and governance <sup>6,7</sup>. Account for direct relationship with land use planning, current  
59 researches pertaining to metro system are mainly focused on socioeconomic aspects of urban  
60 center or downtown area. Of which <sup>8</sup> quantified the volumes and depths of functional use in  
61 Alexanderplatz, Germany; <sup>9</sup> specified the driving factors and predicted the demand of UUS in  
62 Xinjiekou, China; <sup>10</sup> identified the quantitative relationship of UUS area, passenger flow rate  
63 and land price in Osaka, Japan.

64 Nevertheless, there are two points need to be emphasized: (1) the investigation extent of  
65 metro system can be considered broader to the city suburb. Meanwhile, it shows that urban area  
66 is not constantly stationary, instead it expands simultaneously in accordance with city

67 development <sup>11,12</sup>. (2) It remains mystery about the metro system's influence on the built  
68 environment. Apparently, the rapid developments of cities promote the formation of metro  
69 networks, which in turn affect the environment. However, with unprecedented construction of  
70 metro in China <sup>13</sup>, the invisible aspects of metro construction and incidental impacts to cities'  
71 environment are far less researched. Therefore, for the sake of sustainable development,  
72 quantitative characterization of large-scale impacts of metro system on city's environment  
73 become necessary and imminent.

74       Researches have shown that land use change is directly associated with environmental and  
75 ecological effects manifested by remote sensing imagery, reflecting environmental crises and  
76 regional ecological safety like carbon emission <sup>14</sup>, soil degradation <sup>15</sup>, climate change <sup>16</sup>, etc. Its  
77 dynamic simulation is useful in elaborating correlated driving factors, including metro networks,  
78 and its scenario prediction will aid decision-makers in multiple land use planning. The  
79 spatiotemporal analysis of metro scenario -based land use change is crucial not only in  
80 quantitative geography, but also in environmental, spatial planning and sustainability studies <sup>17</sup>.

81       In addition, it is widely acknowledged that cellular automata (CA) are a prudent way to  
82 achieve this exploration, considering both the spatial and temporal extent of vast land. Cellular  
83 automata was proposed by Von Neumann in the late 1940s, for the first time it exhibited self-  
84 reproduction characteristics in computer, and gradually established fundamental role in the  
85 theory of complex system. CA-based models have been applied in the simulation of urban  
86 growth since the 19th century <sup>18,19</sup>. Taking advantage of GIS platform, CA model can simulate  
87 non-linear and complex results close to reality from local rules <sup>20</sup>. It discretizes time and space,  
88 assigns values to cells in the bottom-up framework. In recent years, many scholars have

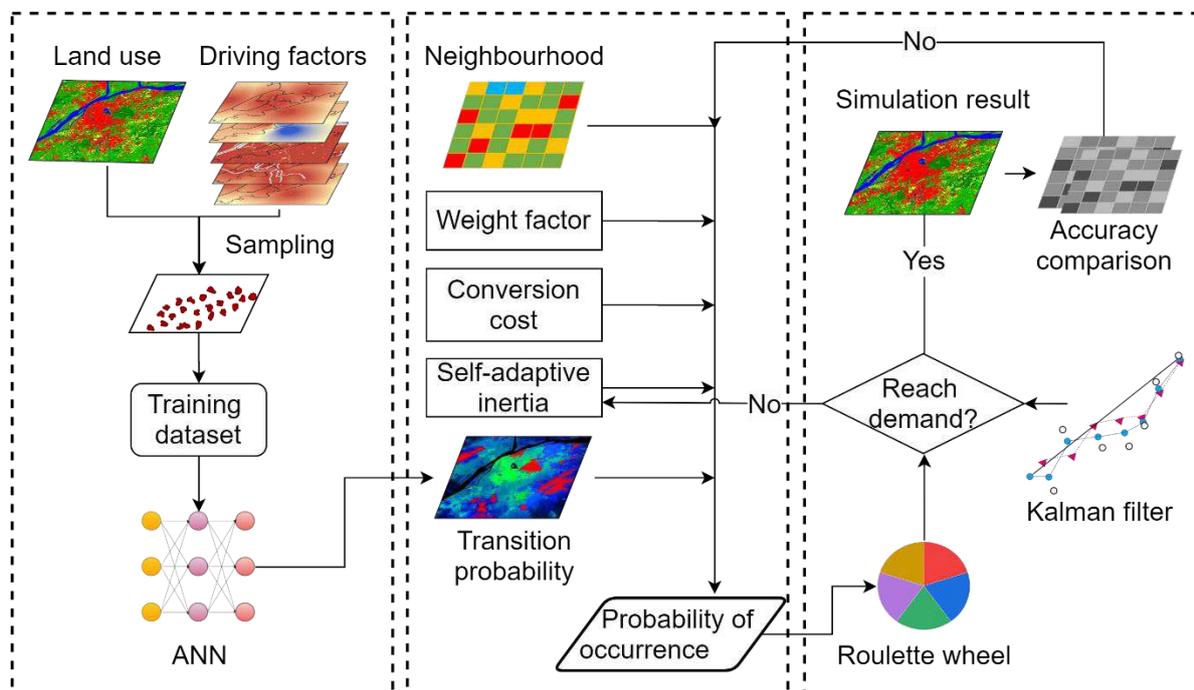
89 combined CA models with artificial neural network <sup>21</sup>, data mining <sup>22</sup>, logistic regression <sup>23</sup>, etc.  
90 Some representative CA models are as follows: DUEM <sup>24</sup>, CLUE-S <sup>25</sup>, Dinamica EGO <sup>26</sup>,  
91 Metronamica <sup>27</sup>, SLEUTH <sup>28</sup>, and so forth.

92 Merging metro system into CA model will enhance simulation accuracy and make the  
93 prediction more reliable. Since Nanjing metro system has been planned for long term until 2035  
94 (www.njmetro.com.cn), it is an ideal research area for exploring its impact on land use. To that  
95 end, Kalman filter, derived from the optimal state estimation method, is integrated with CA  
96 model to simulate land use change from 2005 to 2019. Then the model verifies the land use  
97 pattern of the study area in 2015 and 2019, and predicts the two scenarios of Nanjing city (with  
98 new planned metro lines or not) in 2035, providing reference for exploring the impact of metro  
99 planning.

## 100 **2 Methodology**

101 In this paper, the model integrates top-down Kalman filter with bottom-up ANN-based CA  
102 model to predict land use change of Nanjing in 2035. This paper chooses Future Land Use  
103 Simulation (FLUS) model as a framework of CA model. The FLUS model <sup>16</sup> is a hybrid model  
104 mainly consisting of artificial neural network (ANN) and cellular automata (CA). FLUS model  
105 can simulate the interaction and transition of multiple land use types at the grid cell. It employs  
106 self-adaptive inertia and competition mechanism to guide the simulation of multiple land uses  
107 on the right trajectory. However, as a CA model, FLUS model requires an external source to  
108 regulate evolution of cells, such that the amount of land use types can be in line with the real  
109 macro-scale demands and city planning. The top-down Kalman filter therein plays an important  
110 role from a macro-scale perspective. The Kalman filter is used to identify future demand for

111 land uses for each year, and then allocates the amounts of land use types to the FLUS model.  
 112 Only when the cells of land use types in the FLUS model meets the demanded amounts  
 113 determined by Kalman filter, can the simulation result be generated at one certain iteration. In  
 114 a word, the CA-Kalman filter model is integrated in a bottom-up and top-down framework  
 115 where land use quantities and local allocations mutually feed back in the time series. The  
 116 schematic workflow of proposed integrated model is illustrated in Fig. 1.



117  
 118 **Fig. 1.** Workflow chart of CA-Kalman filter model

119 **2.1 Framework of CA model of FLUS**

120 **2.1.1 Artificial neural network**

121 Artificial neural network (ANN) is one member of machine learning (ML) family. It is well  
 122 acknowledged that ANN can deal with uncertainties emerged from non-linear system, without  
 123 designating model structure, conversion rule, and other parameters. It is comprised of an input  
 124 layer, a hidden layer and an output layer, where multiple neurons connecting inputs and outputs

125 learn features and relationships of a number of variables. In FLUS model, each neuron in the  
 126 input layer represents driving factors that need training, while each neuron in the output layer  
 127 corresponds to a specific land use type. The signal in the input layer can be written as:

$$128 \quad X = [x_1, x_2, \dots, x_m]^T \quad (1)$$

129 Where  $x_i$  is the  $i$ th neuron in the input layer. The signal received by neuron  $j$  in the hidden  
 130 layer on cell  $p$  at time  $t$  is estimated as an adaptive weight based equation:

$$131 \quad ann_j(p, t) = \sum w_{ij} \times x_i(p, t) \quad (2)$$

132 Where  $ann_j(p, t)$  is the signal received by neuron  $j$  in the hidden layer;  $x_i(p, t)$  is the  $i$   
 133 th variable associated with the input neuron  $i$  on cell  $p$  at training time  $t$ ; and  $w_{ij}$  is an  
 134 adaptive weight between the hidden layer and output layer. The connection between hidden  
 135 layer and output layer is controlled by a sigmoid function, mapping values of variables to the  
 136 range between 0 and 1. The sigmoid activation function adds to non-linear characteristics in the  
 137 output layer, which is expressed as follows:

$$138 \quad sigmoid(ann_j(p, t)) = \frac{1}{1 + e^{-ann_j(p, t)}} \quad (3)$$

139 At last, the probability of occurrence  $P(p, k, t)$  at cell  $p$  for land use type  $k$  at training  
 140 time  $t$  is calculated in the output layer as follows:

$$141 \quad P(p, k, t) = \sum_j w_{j,k} \times sigmoid(ann_j(p, t)) = \sum_j w_{j,k} \times \frac{1}{1 + e^{-ann_j(p, t)}} \quad (4)$$

142 Where  $w_{j,k}$  is an adaptive weight between the hidden layer and the output layer.

### 143 **2.1.2. Self-adaptive inertia and competition mechanism**

144 The land use change is booting up under the condition of various factors which exert varying

145 degree of impacts. In order to reflect the real state and process of cell conversion for different  
 146 land use types, namely remain unchanged or convert to another, the model takes into  
 147 consideration neighborhood effect, the conversion cost, and competition mechanism.

148 There are many neighborhood windows proposed by scholars in their CA models, such as  
 149 Von Neumann, Moore, Margolus, etc. The final patterns and emergence are found largely  
 150 derived from the initial set of window shape<sup>18,29</sup>. The neighborhood dominance at cell  $p$  for  
 151 land use type  $k$  is defined as:

$$152 \quad \Omega_{p,k}^t = \frac{\sum_{N \times N} \text{con}(c_p^{t-1} = k)}{N \times N - 1} \times w_k \quad (5)$$

153 Where  $\sum_{N \times N} \text{con}(c_p^{t-1} = k)$  represents the total number of cells occupied by the land use type  $k$   
 154 at the last iteration time  $t-1$  within the  $N \times N$  window.  $w_k$  is the variable weight among the  
 155 different land use types, which is mainly determined based on expert knowledge.

156 Meanwhile, it is necessary to adjust the transition state of cell in accordance with future  
 157 demand of land use. As a result, a self-adaptive inertia is used to enhance the inheritance of  
 158 previous land use types when there exists contradiction. For instance, if more urban land parcels  
 159 are required according to master plan, whereas the cell inclines to avoiding being urban land,  
 160 the inertia will regulate this unfavorable trend onto the right trajectory. The inertia coefficient  
 161 is expressed as follows:

$$162 \quad Inertia_k^t = \begin{cases} Inertia_k^{t-1} & \text{if } |D_k^{t-1}| \leq |D_k^{t-2}| \\ Inertia_k^{t-1} \times \frac{D_k^{t-2}}{D_k^{t-1}} & \text{if } D_k^{t-1} < D_k^{t-2} < 0 \\ Inertia_k^{t-1} \times \frac{D_k^{t-1}}{D_k^{t-2}} & \text{if } 0 < D_k^{t-2} < D_k^{t-1} \end{cases} \quad (6)$$

163 Where  $Inertia_k^t$  denotes the inertia coefficient for land use type  $k$  at iteration time  $t$ .  $D_k^{t-1}$   
 164 denotes the difference between the macro demand and designated amount of land use type  $k$

165 at last iteration time  $t-1$ , and so is  $D_k^{t-2}$ .

166 The inertia coefficient can be altered with the dynamic trend of developing contradiction: (1)  
167 If the macro demand for the specific land use type  $k$  equals to the current allocated amount,  
168 then the inertia coefficient at iteration time  $t$  will stay unchanged; (2) If the macro demand  
169 for the specific land use type  $k$  is less than the current allocated amount, then the inertia  
170 coefficient at iteration time  $t$  will decrease slightly by multiplying the previous coefficient by  
171  $D_k^{t-2} / D_k^{t-1}$ ; (3) If the macro demand for the specific land use type  $k$  is greater than the current  
172 allocated amount, then the inertia coefficient at iteration time  $t$  will increase slightly by  
173 multiplying the previous coefficient by  $D_k^{t-1} / D_k^{t-2}$ .

174 Besides, the model takes in the conversion cost from the perspective of nature of land, i.e.  
175 the intrinsic difficulty to convert land use type. It is determined based on expert experience,  
176 works the same as the inertia to promote or inhabit the growing trend of current land use type.  
177 For each land use pair  $c$  and  $k$ , the cost of the land use change from  $c$  to  $k$  is denoted as  
178  $sc_{c \rightarrow k}$ . The value of the conversion cost  $sc_{c \rightarrow k}$  varies between the range of  $[0,1]$ . Larger values  
179 indicate a greater conversion difficulty, and a value of 1 means that the conversion is nearly  
180 impossible. For example, the urban land is unlikely to convert to forest, because the region with  
181 intense economic activities is unlikely to develop backward in common sense, so the conversion  
182 cost would be 0<sup>30</sup>.

### 183 **2.1.3. Roulette wheel selection**

184 By considering aspects above, the combined probability of a cell being occupied by a specific  
185 land use type is estimated. The combined probability of a cell being occupied by a specific land

186 use type is estimated using the following equation:

$$187 \quad TP_{p,k}^t = P_{p,k} \times \Omega_{p,k}^t \times \text{Inertia}_k^t \times (1 - sc_{c \rightarrow k}) \quad (7)$$

188 Where  $TP_{p,k}^t$  denotes the combined probability of grid cell  $p$  to covert from the original land  
189 use type to the target type  $k$  at iteration time  $t$ ;  $P_{p,k}$  denotes the probability-of-occurrence  
190 of land use type  $k$  on grid cell  $p$ ;  $\Omega_{p,k}^t$  denotes the neighborhood effect of land use type  
191  $k$  on grid cell  $p$  at iteration time  $t$ ;  $\text{Inertia}_k^t$  denotes the inertia coefficient of land use  
192 type  $k$  at iteration time  $t$ ; and  $sc_{c \rightarrow k}$  denotes the conversion cost from the original land use  
193 type  $c$  to the target type  $k$ .

194 The most important part of the model is the roulette selection, because it endows allocation  
195 opportunity to the land use types, which possess lower combined possibilities. This equal treat  
196 is designed out of the perception of competition. That is, even though the specific land use type  
197 is eligible to occupy the cell, it may well not be capable of eliminating other land use type when  
198 it comes to similar competence or the fortune factor in reality. Therefore, a roulette wheel is  
199 proposed where the area of a sector representing a land use type is proportional to the combined  
200 probability. With a uniformly distributed random number ranging from 0 to 1 falling into the  
201 area of a sector, the corresponding land use type is determined. Generally speaking, a land use  
202 type with a higher combined probability is still more likely to be selected at iteration time, but  
203 those with lower probabilities have a chance to be allocated.

## 204 **2.2 Land use demand prediction using Kalman filter**

205 The FLUS model calculates the probability of transition within the bottom-up framework,  
206 however it needs a top-down method to predict future amount of land use types for each year.  
207 The commonly utilized methods are Markov chain and system dynamics, both of which are

208 from macro-level perspective. System dynamics (SD) is another tool originated from complex  
209 system, could have been an optimal choice to be integrated. However, SD requires a variety of  
210 massive data across time, which would be a hurdle for pooling data, let alone complex processes  
211 of adjusting parameters<sup>16,31</sup>. As to Markov chain, it is an incidence-based method commonly  
212 used to predict the variance of future phase. It deploys state transition probability to estimate  
213 the transition trend of certain pixels, but cannot distinguish noises from actual signals. The noise  
214 of simulation is initially derived from the classification of land use in the satellite imagery most  
215 of the time. The cells may be mistaken as other kind of land use by the Deep Learning  
216 techniques. It is crucial to identify real amount of grid cells for typical land use types, and then  
217 incorporate it into the CA model, to minimize the accumulated error and uncertainty during  
218 simulated time series. Therefore, this paper improves predictive methods by integrating Kalman  
219 filter with the state transition probability of Markov chain, governed by the premise that the  
220 amount of pixels remain the same during transition.

221 Kalman filter is an algorithm of data fusion in the theory of optimal state estimation. It is  
222 based on minimizing the mean-square error, can support estimations of past, present, and future  
223 states, which has been widely used in the field of dynamics and control over the past three  
224 decades.<sup>32</sup> proposed to use extended Kalman filter and unscented filter to calibrate  
225 magnetometer.<sup>33</sup> used adaptive Kalman filter algorithm (AKF) to improve the performance of  
226 the robot's speed and heading angle.<sup>34</sup> estimated the target scattering coefficients in an adaptive  
227 radar system based on Kalman filter (KF) with waveform optimization.<sup>35</sup> developed a new  
228 extended interval Kalman filter (EIKF) for tracking the missile system.<sup>36</sup> implemented the  
229 integration of GPS with INS using an extended Kalman filter.<sup>37</sup> proposed a fuzzy multi-sensor

230 data fusion Kalman model to help reduce integrated vehicle health maintenance system  
 231 (IVHMS) failure risk. <sup>38</sup> estimated the state space model of time varying parameter approach  
 232 (TVP) by Kalman filter, which is used for estimating the effect of energy consumption on  
 233 economic growth over period 1967-2009 in Iran. In a word, Kalman filter in its various forms  
 234 is clearly established as a fundamental tool for analyzing and solving a broad class of estimation  
 235 problems.

236 Kalman filter involves the process state and a sequence of noisy measurements, in which the  
 237 modeled system is filled with uncertainty. The equations for Kalman filter fall into two parts:  
 238 time update equations and measurement update equations <sup>39,40</sup>. The time update equations are  
 239 responsible for projecting forward the current state and error covariance estimates to obtain the  
 240 *a priori* estimates for the next time step (Eqs. (19) and (21)). The measurement update equations  
 241 are responsible for the feedback, in other words, for incorporating a new measurement into the  
 242 *a priori* estimate to obtain an improved *a posteriori* estimate (Eqs. (15), (17), and (18)).

243 It begins by assuming the random process to be estimated in the form:

$$244 \quad x_{k+1} = \Phi_k x_k + w_k \quad (8)$$

245 Where  $x_k$  is a  $n \times 1$  process state vector at time  $t_k$ ;  $x_{k+1}$  is a  $n \times 1$  process state vector at  
 246 time  $t_{k+1}$ ;  $\Phi_k$  is a  $n \times n$  matrix relating  $x_k$  to  $x_{k+1}$  in the absence of a forcing function;  
 247  $w_k$  is a  $n \times 1$  vector assumed to be a white sequence with known covariance structure. It is  
 248 the input white noise contribution to the state vector for the time interval  $(t_k, t_{k+1})$ .

249 The measurement of the process is assumed to occur at discrete time points in accordance  
 250 with the linear relationship:

$$251 \quad z_k = H_k x_k + v_k \quad (9)$$

252 Where  $z_k$  is a  $m \times 1$  vector measurement at time  $t_k$ ;  $H_k$  is a  $m \times n$  matrix giving the ideal  
 253 (noiseless) connection between the measurement and the state vector at time  $t_k$ ;  $v_k$  is a  $m \times$   
 254 1 measurement error assumed to be a white sequence with known covariance structure.

255 The covariance matrices for  $w_k$  and  $v_k$  vectors are given by:

$$256 \quad E[w_k \quad w_j^T] = Q_k \delta_{kj} \quad (10)$$

$$257 \quad E[v_k \quad v_j^T] = R_k \delta_{kj} \quad (11)$$

$$258 \quad E[w_k \quad v_j^T] = 0 \quad (12)$$

259 If  $k = j$ , then  $\delta_{kj} = 1$ . If  $k \neq j$ , then  $\delta_{kj} = 0$ .

260 We assume we have an initial estimate of the process at the point  $t_k$ , and it is based on all  
 261 our knowledge about the process prior to  $t_k$ . This *a priori* estimate will be denoted as  $\hat{x}_k^-$   
 262 where the “hat” denotes estimation, and the “super minus” is a reminder that this is our best  
 263 estimate prior to assimilating the measurement at  $t_k$ . We also assume that we know the error  
 264 covariance matrix associated with  $\hat{x}_k^-$ . That is, we define the estimation error to be:

$$265 \quad e_k^- = x_k - \hat{x}_k^- \quad (13)$$

266 In addition, the associated error covariance matrix is:

$$267 \quad P_k^- = E[e_k^- e_k^{-T}] = E[(x_k - \hat{x}_k^-)(x_k - \hat{x}_k^-)^T] \quad (14)$$

268 With the assumption of *a priori* estimate  $\hat{x}_k^-$ , we now seek to use the measurement  $z_k$  to  
 269 improve it. We choose a linear blending of the noisy measurement and the *a priori* estimate in  
 270 accordance with the equation

$$271 \quad \hat{x}_k = \hat{x}_k^- + K_k (z_k - H_k \hat{x}_k^-) \quad (15)$$

272 Where  $\hat{x}_k$  is a updated estimate;  $K_k$  is Kalman gain (yet to be determined).

273 The problem now is to find Kalman gain  $K_k$  that yields an updated estimate which is

274 optimal because it minimizes the trace of *a posteriori* error covariance  $P_k$ .

$$275 \quad P_k = E[e_k e_k^T] = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T] \quad (16)$$

276 Next, we substitute the resulting expression for  $\hat{x}_k$  into Eq. (16)

$$277 \quad P_k = (I - K_k H_k) P_k^- (I - K_k H_k)^T + K_k R_k K_k^T \quad (17)$$

278 We proceed to differentiate the trace of  $P_k$  with respect to  $K_k$ :

$$279 \quad K_k = P_k^- H_k^T (H_k P_k^- H_k^T + R_k)^{-1} \quad (18)$$

280 The updated estimated  $\hat{x}_k$  is easily projected ahead via the transition matrix. We are

281 justified in ignoring the contribution of  $w_k$  in Eq. (16) because it has zero mean and is not

282 correlated with any of the previous  $w$ 's. Thus, we have

$$283 \quad \hat{x}_{k+1}^- = \Phi_k \hat{x}_k \quad (19)$$

284 The error covariance matrix associated with  $\hat{x}_{k+1}^-$  is obtained by first forming the expression

285 for the *a priori* error

$$286 \quad e_{k+1}^- = x_{k+1} - \hat{x}_{k+1}^- = \Phi_k e_k + w_k \quad (20)$$

287 We now note that  $w_k$  and  $e_k$  have zero cross correlation, because  $w_k$  is the process

288 noise for the step ahead of  $t_k$ . Thus, we can write the expression for  $P_{k+1}^-$  as

$$289 \quad P_{k+1}^- = E[e_{k+1}^- e_{k+1}^{-T}] = \Phi_k P_k \Phi_k^T + Q_k \quad (21)$$

290 Equations (15), (18), (19), and (21) comprise the Kalman filter recursive equations. The time

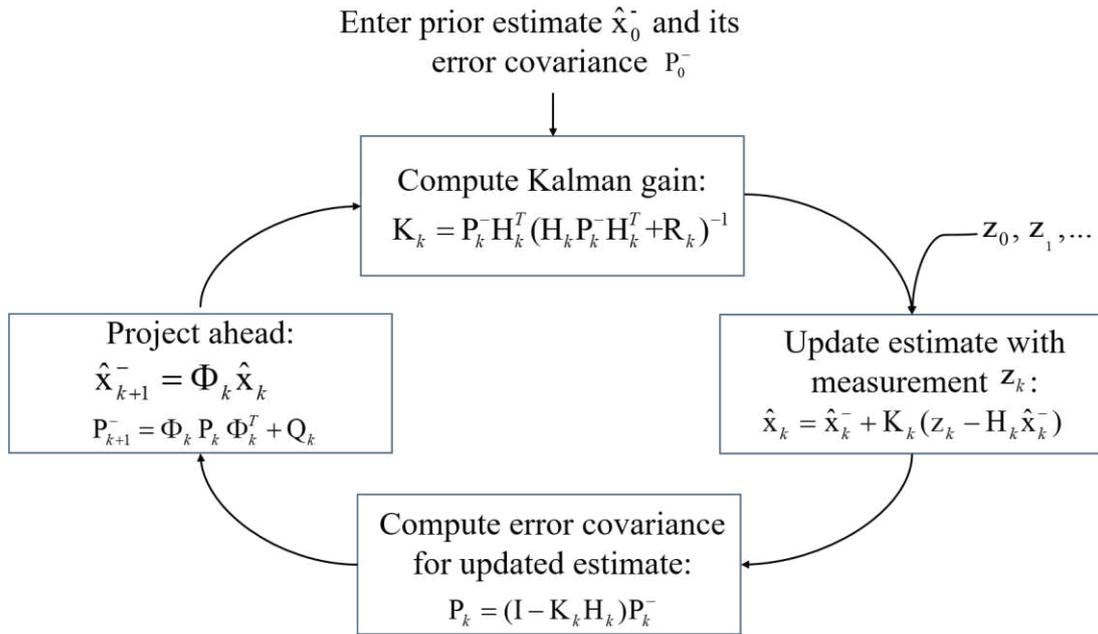
291 update equations can also be regarded as predictor equations, while the measurement update

292 equations can be thought of as corrector equations. Once the loop is entered, it can proceed on

293 *ad infinitum* recursively. With the aid of updated Kalman gain, the updated estimation is

294 resulted from the *a priori* estimate of process state assimilating the measurement as shown in

295 Fig. 2.



296

297

**Fig. 2.** Kalman filter loop

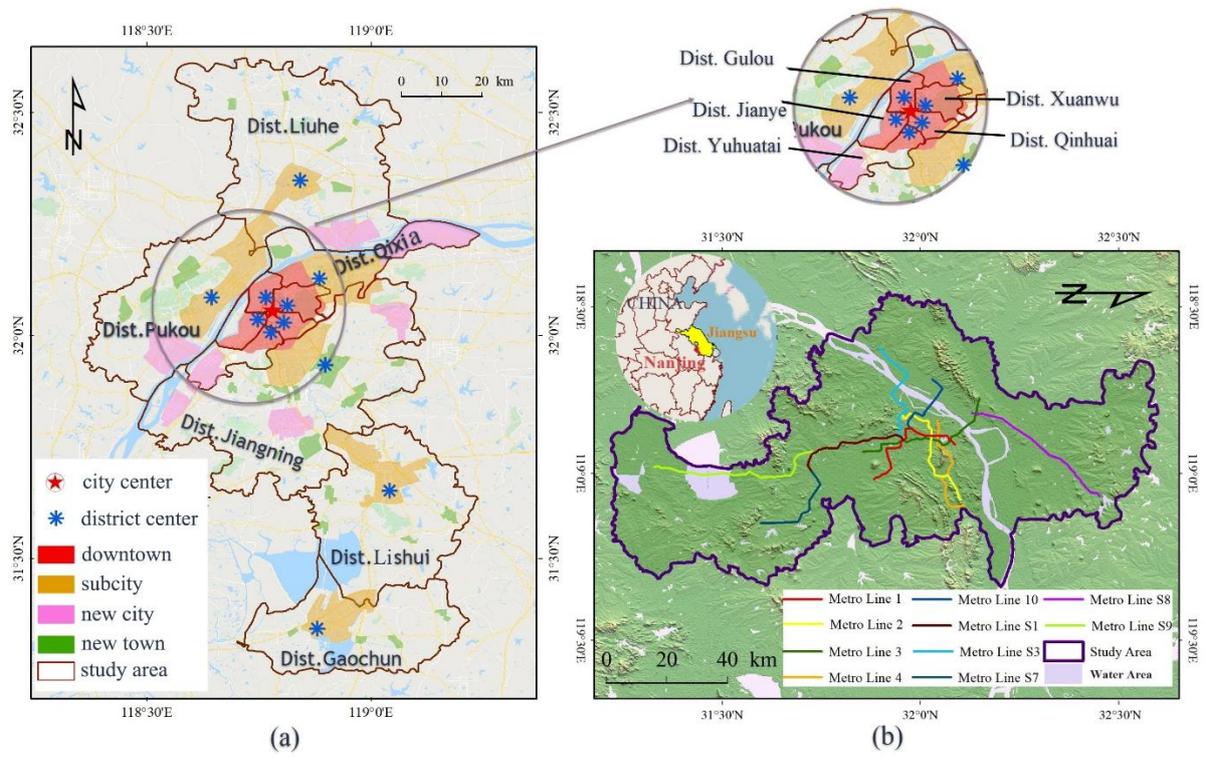
### 298 3 Case study

#### 299 3.1 Study area

300 Nanjing is the capital city of Jiangsu Province in East China, situated in the lower reaches of  
 301 the Yangtze River. It is located at the range between  $31^\circ 14'' \sim 32^\circ 37''$  N and  $118^\circ 22'' \sim$   
 302  $119^\circ 14''$  E, with the total area of 6,587.02 square kilometers, and the residential population of  
 303 8.335 million. Known as a historic and cultural city to the world, it boasts a picturesque scenery  
 304 integrating with ancient relics across six ancient dynasties. As the largest city in Yangtze River  
 305 Delta, Nanjing ranked 11th among national cities with the GDP of 143 billion RMB in 2019  
 306 (<http://www.stats.gov.cn>). In the same year, it was rated as one of 12 new first-tier cities of  
 307 China (<https://www.cbnweek.com/articles/magazine/23576>), showing its nimble  
 308 competitiveness in lots of international rankings.

309 Nanjing consists of 11 municipal districts, with its downtown area occupying 6 main districts,

310 and has district centers distributed in downtown, sub-cities, new cities, and new towns, as Fig.  
 311 3 (a) depicts. The booming economy of whole city flows through the transportation vessels of  
 312 highways, railways, and metro lines that have bolstered population migration for years. Fig. 3  
 313 (b) illustrates existing metro lines that are principally built along riverside districts. The density  
 314 of metros in different administrative districts indicates unbalanced development and short-term  
 315 planning emphasis, which will pose different impacts on land use pattern from the perspective  
 316 of territory spatial planning.



317 (a) 318 **Fig. 3.** Study area and layout of existing metro lines

319 **3.2 Data sources for spatial database**

320 **3.2.1 Data acquisition**

321 The data of existing metro lines come from the website of Amap, and data for the planned  
 322 lines come from the website of mapbar. Both of them are in the form of attribute table including

323 the name and location of metro stations that could be precisely projected in the study area. The  
 324 operation years of metro lines are classified into four periods by three year-sections, as  
 325 mentioned in Section 2. Table 1 shows all metro lines of Nanjing City, including the planned  
 326 ones. Based on each operation year, three year-sections (2010, 2015, and 2019) are determined,  
 327 dividing time span into four phases (2005-2010, 2010-2015, 2015-2019, and 2019-2035). The  
 328 ultimate year turns out to be 2035, for two reasons: (1) as Table 1 indicates, the clear operation  
 329 year for planned lines is 2030, it is better to leave a 5-year interval for the metro fusion with  
 330 surrounding land; (2) the official master plan of Nanjing City envisions the future to the year  
 331 2035.

332 **Table 1**

333 The operating and planned metro lines of Nanjing

NO.	Operation year	Length (km)	NO.	Operation year	Length (km)
1	2005	38.90	S6	2021	26.31
2	2010	37.95	S4	2022	8.15
10	2014	21.60	6	2023	32.40
S1	2014	37.30	9	2023	19.67
S8	2014	45.20	11	2023	27.00
3	2015	44.90	S5	2025	14.15
4	2017	33.80	13	2030	36.40
S3	2017	36.22	14	2030	34.20
S9	2017	52.42	15	2030	31.40
S7	2018	30.16	16	2030	25.16
5	2020	37.40	8	NAN	62.30
7	2021	35.49	12	NAN	24.30

Note: (1) The planned extensions of existing lines are not considered in this paper.

(2) For trans-provincial lines, the length of which beyond Nanjing is excluded.

334 Remotely sensed data have drawn considerable attention in the analysis and modelling of  
 335 land use change. In this regard, the remote sensing imageries are acquired from the open-source  
 336 platform EarthExplorer. In order to match the broad extent of area, to maintain the same

337 resolution of images in time span, and to be of relatively high quality among other types of  
338 satellites, Landsat TM5 imageries are chosen to present land use change. With a resolution of  
339 30m, Landsat TM5 imageries are widely deployed in the modelling and simulation of  
340 geological cellular automata <sup>41,42</sup>. Because of long-term sequence, this paper makes two  
341 validations after simulation based on the imageries of 2015 and 2019.

342 The land use is strongly driven by geophysical, socioeconomic, and institutional conditions  
343 <sup>43</sup>. It is impossible to make simulation that absolutely resemble the reality, but only sufficiently  
344 close to that by choosing most representative driving factors. This paper considers topology,  
345 transportation, and socio-economy as strong driving factors, aims to incorporate them to arrive  
346 at reasonable simulation and prediction. The topology aspect includes DEM and slope that  
347 discern elevation of study area and correspond to physical geo-restrictions and development  
348 costs. It indicates the terrain's suitability for urban development. The transportation aspect  
349 shows the proximity to residential settlement and population flow, which are strongly associated  
350 with spatial planning of transportation infrastructure. The socio-economy aspect reveals  
351 population density and GDP density, reflecting the macro policy influence on urban  
352 development. It suggests the possibility of the nonurban-to-urban transformation. It should be  
353 emphasized that transportation factor here only contains metro information, excluding roads  
354 and railways. Because by ruling out other types of transportation, the metro effect could be  
355 ideally extracted (using scenario subtraction) in case of multiple redundant intersections. At the  
356 cost of overall accuracy declination, the idea of subtraction of scenarios proves worthy in the  
357 following sections. The raw data of DEM and slope can be obtained from the open-source  
358 platform Earthdata. The transportation aspect contains distance to city center, distance to district

359 center, and distance to metro lines, which are made out of original geo-points after Euclidean  
 360 distance analysis and resampling process. The city center refers to Xijiekou that located at the  
 361 intersection of Gulou District, Xuanwu District and Qinhuai District <sup>9</sup>. The district center refers  
 362 to the administrative center of each district. It should be noted that the administrative centers  
 363 (of city/district) went through changes in Feb. 2013 according to official documents, which  
 364 would subsequently influence the visualization of distance to district center and other factors  
 365 for year 2005, 2010, 2015, and 2019. For instance, Qinhuai District and Baixia District were  
 366 merged into new Qinhuai District; Gulou District and Xiaguan District were merged into new  
 367 Gulou District; Lishui and Gaochun were updated from county to district administratively.  
 368 However, the alteration concerning merge and rename would not cause significant changes to  
 369 DisDist visualization. Because the merged districts were adjacent to each other, moreover, they  
 370 covered quite small area in the city core area. Given that, we use the method called Euclidean  
 371 distance analysis to illustrate the distance to District centers without having to produce DisDist  
 372 (2005, 2010, 2015, 2019). As to the distance to metro lines, GDP and population, the same  
 373 processing is also applied. While it is important to count GDP and population according to the  
 374 changes of districts before and after year 2013, for GDP and population are not in the same  
 375 magnitude as the distance when it comes to map visualization using Euclidean distance analysis.  
 376 As a result, we need to produce GDP and Pop separately and differently for each year section  
 377 (2005, 2010, 2015, 2019). Table 2 sums up the preparation of data before simulation.

378 **Table 2**

379 Data preparation for land use change simulation

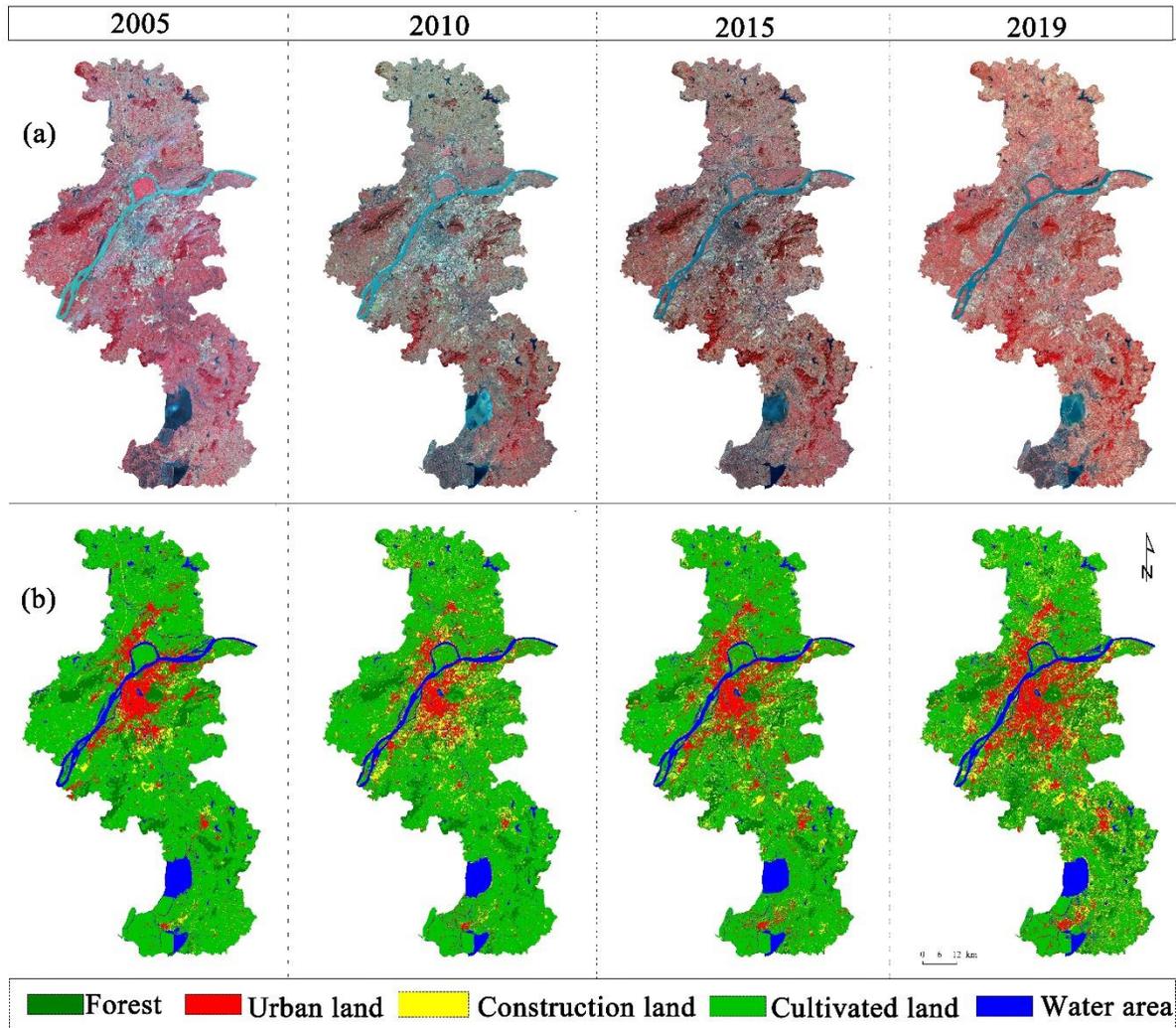
Type	Abbr.	Statement	Res.	Source	Purpose
Land use	LandUse	Remote sensing images	30m	earthexplorer.usgs.	Input

		classification		gov	
Driving factors	DEM	Elevation	30m	earthdata.nasa.gov	Topology impact
	Slope	Slope	30m		
	DisCity	Dis. to city center	30m	www.amap.com	
	DisDist	Dis. to district center	30m	www.mapbar.com	Transportation impact
	DisLine	Dis. to metro line in 2005-2019	30m	www.mapbar.com	
	Pop	Population density in 2005-2019	30m	www.worldpop.org	Socioeconomic impact
	GDP	GDP density in 2005-2019	30m	tjj.nanjing.gov.cn	

Note: the scope of study area is 5103x2793

### 380 3.2.2 Data visualization

381 To begin with, the scope of study area is clipped at 5103 rows and 2973 columns in ENVI  
382 software. After some necessary pre-processes, the land use is categorized into five types by  
383 means of supervised classification, namely forest, urban land, construction land, cultivated land,  
384 and water area. Supervised classification is a commonly used method based on subjective  
385 sampling of naked eye that demands prior experience and knowledge for each land use type.  
386 Then an amount of samples are trained according to maximum likelihood of bands combination,  
387 thus all cells find their own categories thanks to special spectrum feature, as is depicted in Fig.  
388 4.



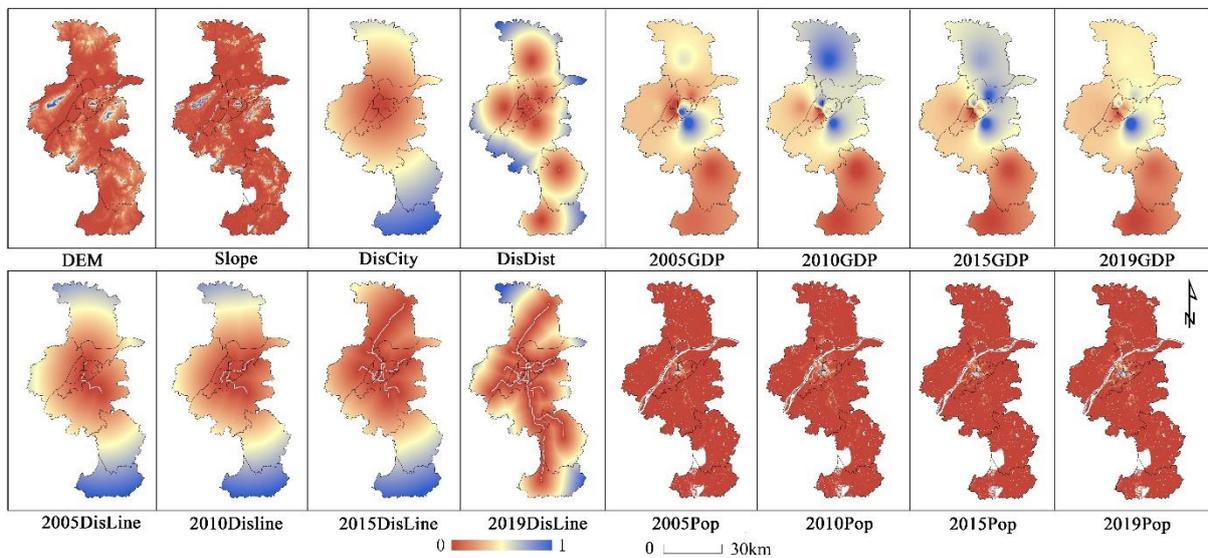
**Fig. 4.** Presentation and classification of Landsat imagery of Nanjing city

(a: Landsat remote sensing imagery; b: categorical map based on supervised classification)

Compared with 2005, the categorical map of 2010 reveals an increase in construction land and a decrease in urban land in downtown area. It can be ascribed to urban redevelopment in a way. As <sup>30</sup> points out, that urbanization is inexorable for no existing urban areas are subjected to deurbanization. The disappearance of urban patches can be observed during urbanization, owing to local redevelopment <sup>44</sup>. A land transformation from urban built area to demolished land can be identified as short-term decrease of impervious surface coverage by remote sensing. Theoretically speaking, such a phenomenon occurs at a point of time span, but soon phases out

399 or be masked in a long interval.

400 When pooling driving factors for simulation, quantification and visualization is always the  
401 priority. Factors like industrialization <sup>45</sup>, slope and elevation <sup>46</sup>, land use policy and urban  
402 planning <sup>47</sup>, infrastructure <sup>48</sup> are proved feasible with regard to data source and mapping.  
403 Meanwhile, for the sake of dimension effect, the normalization method of <sup>49</sup> is adopted, shown  
404 is the visualization of driving factors in Fig. 5.



405

406 **Fig. 5.** Driving factors for land use change in 2005, 2010, 2015, 2019

### 407 3.3 Model simulations

#### 408 3.3.1 Parameter adjustment

409 Before calculating the transition probability of land use types in the FLUS model, an ANN  
410 based training is executed where the number of hidden layer is set to 7, and the sampling rate  
411 is set to 0.2%, which is tested as optimal result after multiple trials. Next, by adding prepared  
412 raster files of seven driving factors, integrating with the categorical maps, the transition  
413 probabilities in the year 2010, 2015 are respectively generated. Note that the year 2005 is not

414 included because the data of the former year 2000 is unavailable here, so that the Kalman filter  
 415 could not produce corresponding data of year 2010 to estimate whether the transition has  
 416 reached the demand or not. In addition, the year 2019 is ruled out because it is merely for  
 417 prediction in Section 3.3.3.

418 Then, during the self-adaptation and competition of cells with three neighbors designated,  
 419 the model will iterate at most 300 times. In case of slow operation due to broad extent of area,  
 420 specified are the value of accelerator as 0.1, as well as the running thread as 8. A few restrictions  
 421 are posed taking into account of reality rules. The cost matrix for each land use type is as Table  
 422 3 shows. When one type of land use is able to convert to another type, the cost coefficient would  
 423 be 1, otherwise turns to be 0. In addition, the weights of neighborhood of forest, urban land,  
 424 construction land, cultivated land, and water area are 0.5, 1.0, 1.0, 1.0, and 0.1, respectively.  
 425 The parameter adjustment for neighborhood weights is based on trial-and-error, by comparing  
 426 the expansion performance of cells with real satellite imageries during multiple trials of  
 427 coefficient combinations. The greater the value of weights, the stronger the expansion ability  
 428 of specific type cells would be.

429 **Table 3**

430 Cost matrix for each land use type

	Forest	Urban land	Constr. land	Cultiv. land	Water area
Forest	1	1	1	1	0
Urban land	0	1	0	0	0
Constr. land	0	1	1	0	0
Cultiv. land	1	1	1	1	1
Water area	0	0	0	1	1

431 As soon as the simulation reaches the maximum iteration, or the result gradually converges

432 to a value, the probabilities of occurrence in 2015 and 2019 are calculated which are originated  
 433 from initial maps of 2010, 2015 respectively. The probability of occurrence for each cell is then  
 434 selected by roulette wheel to determine final transition land use type. At this moment, Kalman  
 435 filter generates future demand for land use of the two years, to estimate whether the transition  
 436 amount has meet the need. If so, the model will generate simulated results. If not, the feedback  
 437 from Kalman filter will be given to self-adaptive inertia, to reproduce new probability of  
 438 occurrence in multiple iterations.

### 439 3.3.2 Validation

440 By comparing the actual imagery and simulated result, we can make two validations for the  
 441 year 2015 and 2019. In this part, three image indicators are adopted to quantify the accuracy of  
 442 simulation, namely Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure  
 443 (SSIM), and Feature Similarity Index for Image (FSIM). Among them, PSNR is calculated from  
 444 the gray value of the image, which can be used as a preliminary characterization of the  
 445 comparison between the predicted result and the actual result. In addition, SSIM and FSIM are  
 446 used to calculate the similarity between the two images.

447 The value of PSNR ranges from 20 to 40, the larger the value, the smaller the difference of  
 448 two images. It can be expressed as follows:

$$449 \quad PSNR = 10 \log_{10} \left( \frac{L^2}{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [I_1(i, j) - I_2(i, j)]^2} \right) \quad (22)$$

450 Where  $M, N$  represent the image size;  $I_1(i, j), I_2(i, j)$  represent the gray value of two  
 451 images used for comparison at coordinates  $(i, j)$ ;  $L$  is the peak signal, for an 8-bit gray  
 452 image,  $L=2^8 - 1=255$ .

453 In SSIM, structural information is incorporated, i.e. brightness, contrast, and structural  
 454 similarity<sup>50</sup>. If the value of SSIM is closer to 1, it means highest similarity. Assuming that X  
 455 and Y are two images for comparison, then we have:

$$456 \quad l(X, Y) = \frac{2\mu_X \mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \quad (23)$$

$$457 \quad c(X, Y) = \frac{2\sigma_X \sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \quad (24)$$

$$458 \quad s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X \sigma_Y + C_3} \quad (25)$$

459 Where  $\mu_X$ ,  $\mu_Y$  are the mean value of image X and Y respectively, denote brightness  
 460 information;  $\sigma_X$ ,  $\sigma_Y$  are the variance of image X and Y respectively, denote contrast  
 461 information;  $\sigma_X \sigma_Y$  is the correlation coefficient of image X and Y, denotes the similarity of  
 462 structural information;  $C_1, C_2, C_3$  are natural numbers close to 0 to prevent abnormal results  
 463 when the denominator is zero. Thus, SSIM can be expressed as follows:

$$464 \quad SSIM(X, Y) = [l(X, Y)]^\alpha + [c(X, Y)]^\beta + [s(X, Y)]^\gamma \quad (26)$$

465 Where  $\alpha, \beta, \gamma$  are used to adjust the proportions of three types of information.

466 FSIM is an image evaluation index based on traditional image evaluation method<sup>51,52</sup>, where  
 467 phase consistency and image gradient are considered for comparison. It bears the assumption  
 468 that the point with the most count of Fourier component is a feature point, instead of simply  
 469 taking the biggest gray value change as a feature to simulate human perception and observation  
 470 of images. If the value of FSIM is closer to 1, it means highest similarity. By obtaining two  
 471 images' phase consistency  $PC_1$ ,  $PC_2$ , and image gradient  $G_1$ ,  $G_2$ , we have:

$$472 \quad S_{pc}(X) = \frac{2PC_1(X) \cdot PC_2(X) + T_1}{PC_1^2(X) + PC_2^2(X) + T_1} \quad (27)$$

473 
$$S_G(X) = \frac{2G_1(X) \cdot G_2(X) + T_2}{G_1^2(X) + G_2^2(X) + T_2} \quad (28)$$

474 
$$S_L(X) = S_{PC}(X) \cdot S_G(X) \quad (29)$$

475 Thus, FSIM can be expressed as follows:

476 
$$FSIM = \frac{\sum_{X \in \Omega} S_L(X) \cdot PC_m(X)}{\sum_{X \in \Omega} PC_m(X)} \quad (30)$$

477 Where  $T_1, T_2$  are constant;  $PC_m = \max(PC_1(X), PC_2(X))$ ;  $\Omega$  represents the extent of  
 478 image.

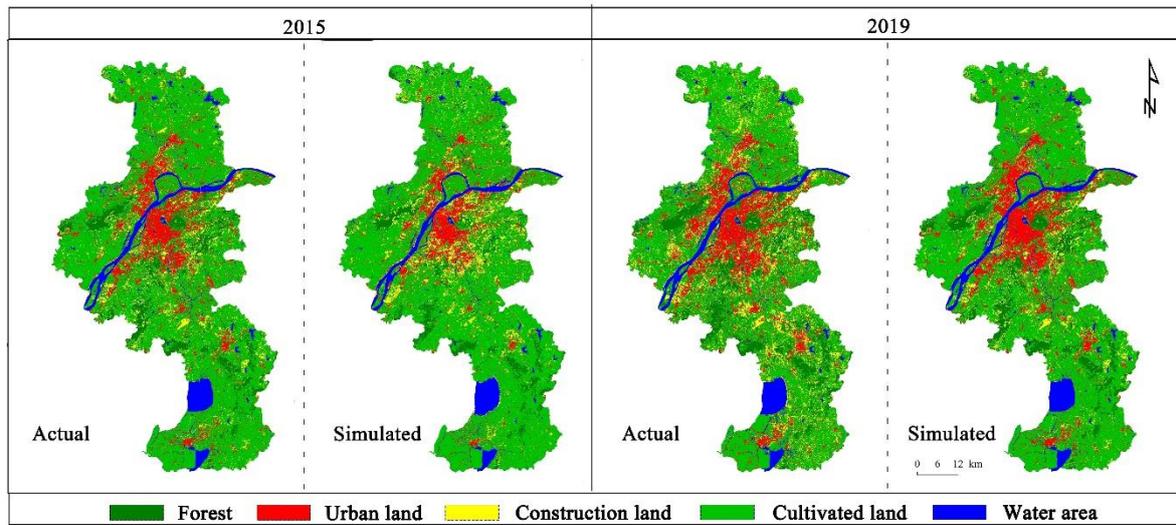
479 **Table 4**

480 The comparison of image indicators for model simulation accuracy

Image indicators	Accuracy	
	2015	2019
PSNR	38.0460	34.3977
SSIM	0.7002	0.6534
FSIM	0.8795	0.8554

481 As Fig. 6 illustrates, the simulated results have common urban tissues with actual ones, yet  
 482 fail to recognize a certain number of urban patches, as well as construction land. Table 4 shows  
 483 that three indicators (PSNR, SSIM, and FSIM) are respectively 38.0460, 0.7002, 0.8795 in 2015,  
 484 and are respectively 34.3977, 0.6534, 0.8554 in 2019. Notwithstanding the accuracy appears  
 485 slightly satisfactory, it matches other findings in literature <sup>42</sup> when it comes to broad extent of  
 486 study area with the resolution of 30 meters, let alone single transportation factor metro is  
 487 involved. Consequently, the cellular automata model can generate relatively real imagery across  
 488 long term, which will support the prediction of the land use in 2035 in Section 3.3.3, and

489 extraction of patches with regard to metro effect in Section 3.4.



490

491 **Fig. 6.** Comparison between actual and simulated land use change

### 492 3.3.3 Prediction

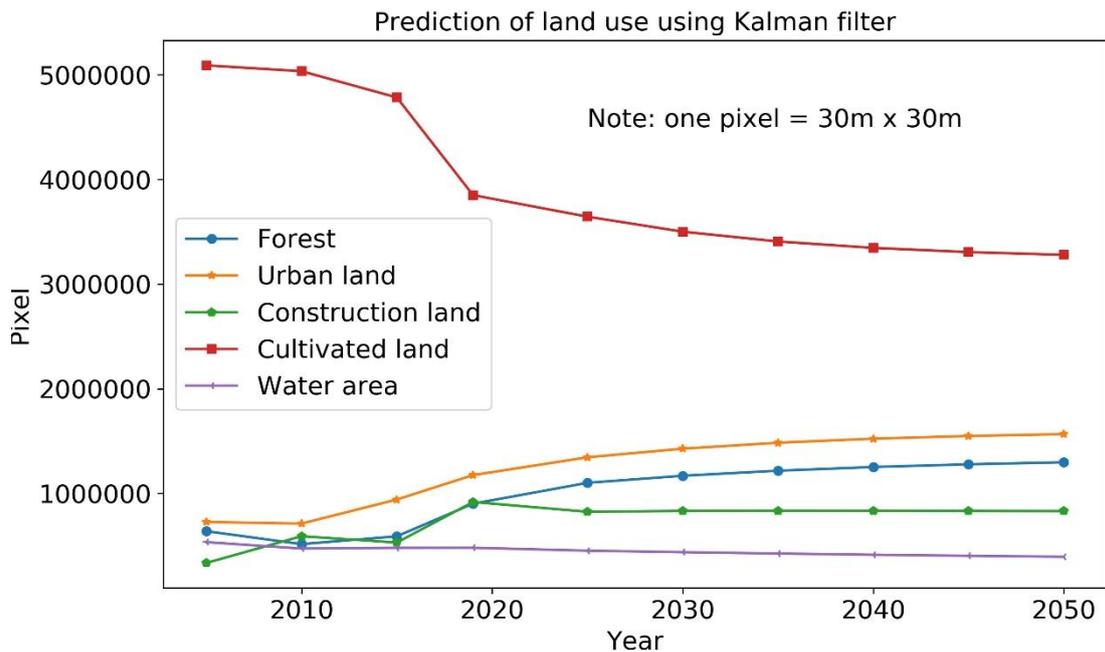
493 From 2019 to 2035, 14 new metro lines are planned on the basis of existing lines, some of  
494 which are currently under construction. The descriptive information and data are collected from  
495 the webpage of Nanjing metro in Baidu baike and mapbar, respectively. After projection,  
496 calibration, and vectorization in ArcGIS, the driving factor of merged metro lines, including  
497 existing lines as of 2019, are incorporated into the coupled model to predict land use change as  
498 of 2035.

499 The task of prediction here is for two future scenarios in 2035, namely metro scenario that  
500 includes 24 metro lines, non-metro lines that includes 10 metro lines. The scenario for non-  
501 metro also needs to be simulated in the coupled model, only by selecting of metro raster files  
502 as of 2019 when importing driving factors. By repeatedly following simulation processes  
503 mentioned in sections above, the model successfully predicts two land use scenarios of Nanjing

504 City as of 2035. In this case, the state transition probability is determined as the ratio of transited  
505 amount to total amount. Integrating the available transition matrices of four periods, a  
506 comprehensive state transition probability matrix is obtained, which then serves as the process  
507 state equation  $\Phi_{k+1,k}$  of Kalman filter. The matrix is as follows:

$$508 \quad \Phi_{k+1,k} = \begin{bmatrix} 0.5689 & 0.2473 & 0.0192 & 0.1591 & 0.0055 \\ 0.0675 & 0.8014 & 0.0604 & 0.0643 & 0.0064 \\ 0.0191 & 0.1308 & 0.8394 & 0.0069 & 0.0038 \\ 0.0675 & 0.1432 & 0.0045 & 0.7793 & 0.0056 \\ 0.0127 & 0.0877 & 0.0052 & 0.014 & 0.8805 \end{bmatrix} \quad (31)$$

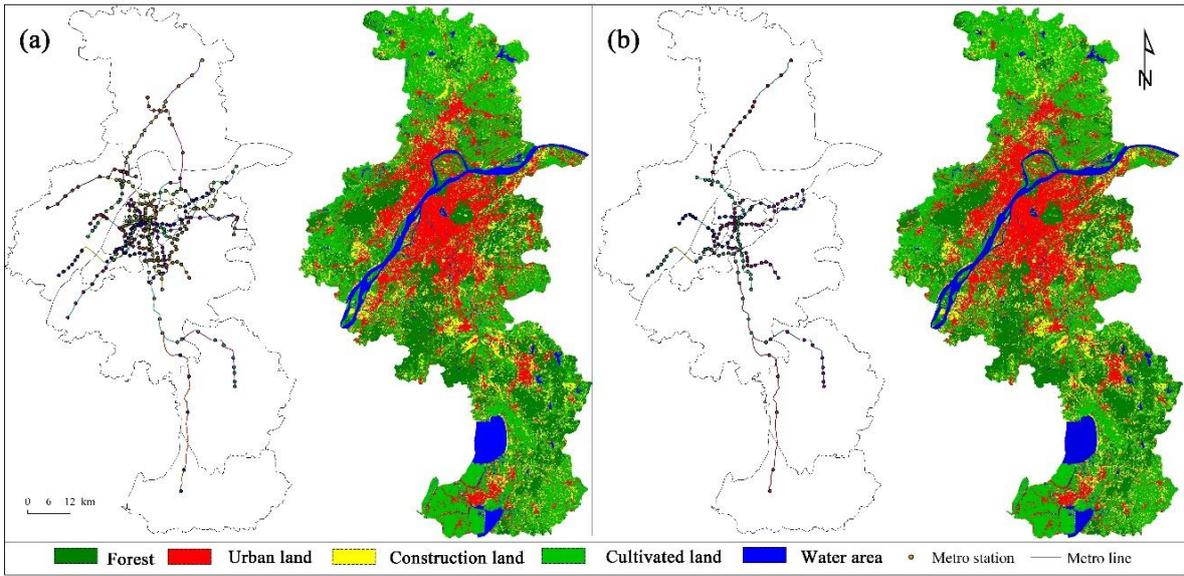
509 According to equation (21), (18), (17), (15) in Section 2.2, we predefine observation matrix  
510  $H_k$  as [1 1 1 1 1], the column number of which is determined by the number of land use types.  
511 In addition, both *a priori* estimation error covariance matrix  $P_k^-$  and process error covariance  
512 matrix  $Q_k$  are defined as a  $5 \times 5$ -identity matrix filled with 10. Here, 10 is thought as a mild  
513 increment for iterations. Measurement error covariance matrix  $R_k$  is set as  $10^5$ , therein we  
514 obtain Kalman Gain  $K_k$  through those necessary parameters. Subsequently, the estimation  
515 error covariance  $P_k$  is updated with the help of Kalman Gain  $K_k$ , the process of which is  
516 called from *a priori* to *a posteriori* in Section 2.2. Finally, we update the estimate  $\hat{x}_k^-$  with  
517 measurement  $z_k$ . The long-term prediction of land use evolution is therefore depicted in Fig.  
518 7.



519  
520 **Fig. 7.** Land use prediction using Kalman filter from 2005 to 2050

521 Based on the prediction of Kalman filter, the FLUS model proceeds to allocate the amount  
522 of land use types for 2035. As is shown in Fig. 8, five land use types are exhibiting more  
523 complex spatial characteristics. Particularly the urban land and construction land are replacing  
524 more other types of land. Nevertheless, these two scenarios look alike, as if there is no land use  
525 differences caused by newly added metro lines.

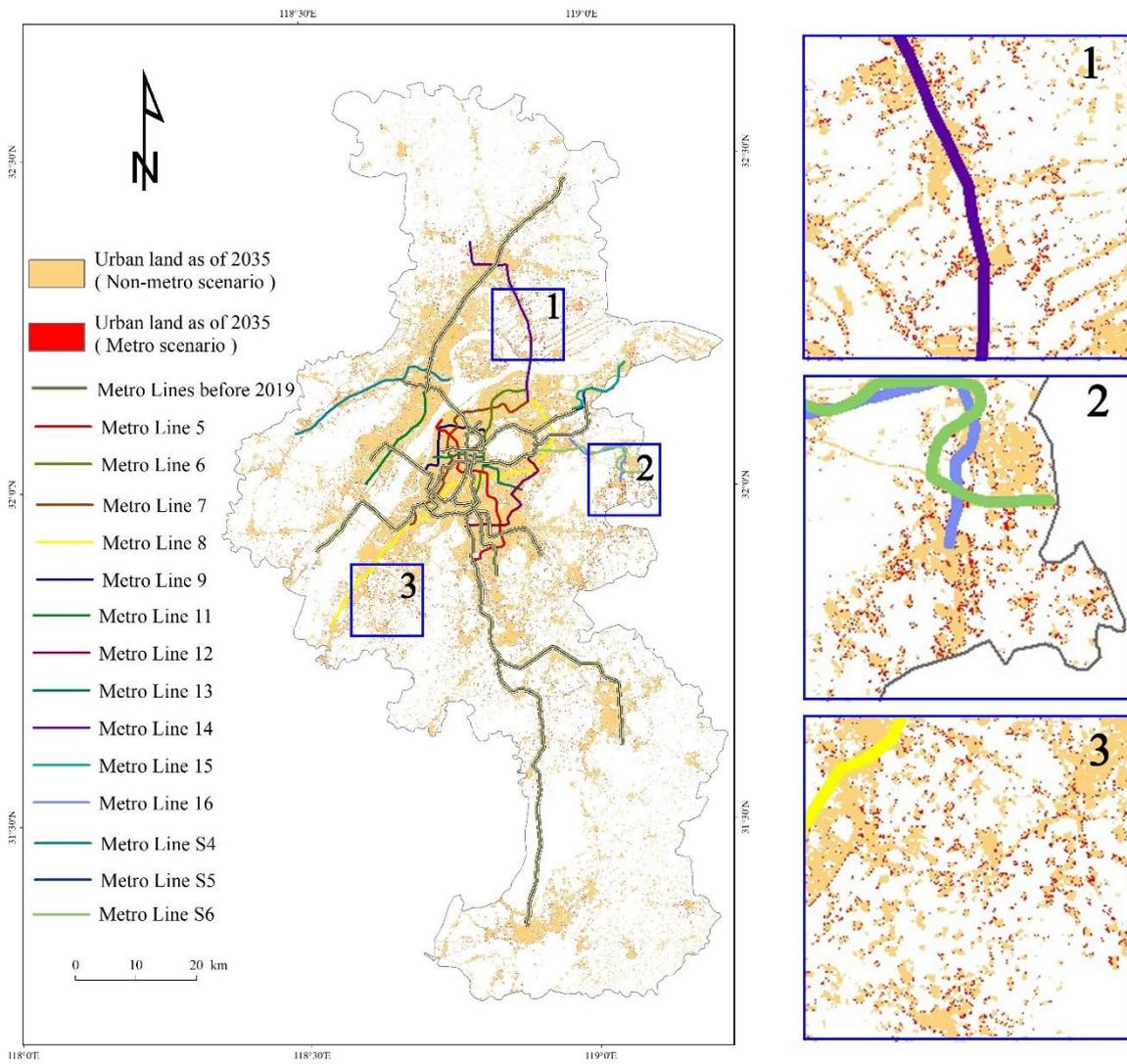
526 In light of that, more micro-level details are depicted in Fig. 9. Taking urban land as example,  
527 new urban patches are distributed approximate existing urban patches, not very crowded, but at  
528 a certain scale to form considerable settlements. The concealed information are nearly masked  
529 between two scenarios' maps, showing that the metro effects are in favor of local area. To study  
530 moderate increment of grain sizes, it requires enlarging specific area with one scenario as  
531 background. At three places of Fig. 9, it indicates the number of urban patches under metro  
532 scenario surpass that of non-metro scenario.



533

534

**Fig. 8.** Simulated results for 2035 (a: metro scenario; b: non-metro scenario)



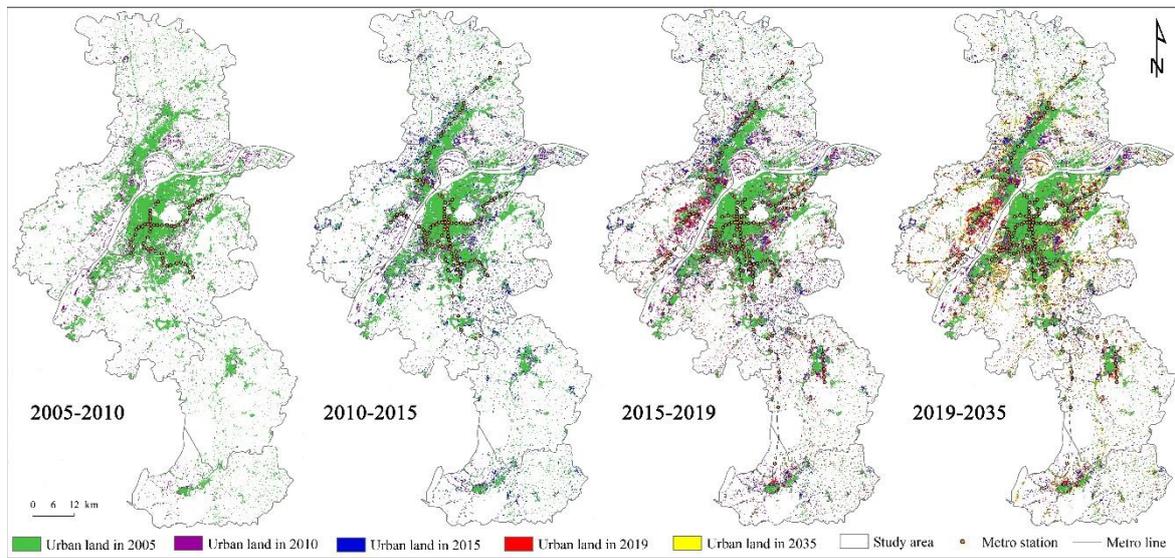
535

536 **Fig. 9.** Comparison of urban land under two scenarios for 2035  
537 (Left: Nanjing metro lines as planned since 2019; right: three enlarged area near metro lines)

538 Therefore, three questions emerge: how does land use respond to metro expansion? How to  
539 measure the relationship between urban land and metro lines across space and time? How the  
540 accumulated metro effects can cast light on territory development plan?

### 541 **3.4 Results and Discussion**

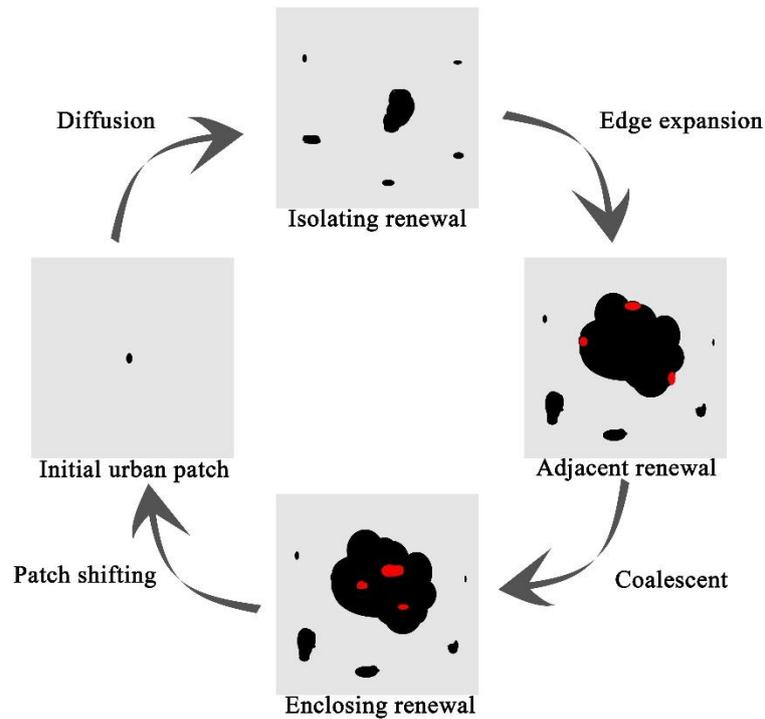
542 If separately examining the growth of urban land, we shall discover that the urban growth  
543 boundaries delineated in each four period are cluttered with patches dotted around the outskirts  
544 (Fig. 10). That absorption of neighboring land is meant to reduce the concentration of wealth  
545 and culture in the main city area.



546 **Fig. 10.** Urban growth boundary as of four year-sections  
547

548 According to <sup>44</sup>, the evolution of urban area actually includes simultaneous expansion and  
549 shrinkage, which can be characterized as oscillated processes of diffusion and coalescence (Fig.  
550 11). The process makes urban patches exhibit three kinds of shrinkage: isolating renewal,  
551 adjacent renewal, enclosing renewal. Among them, enclosing renewal is exactly in connection

552 with urban redevelopment, proving again the evidence found in Section 3.4.1. As is shown in  
 553 Fig. 10, colored urban patches of four periods have suffered both expansion and shrinkage.  
 554 Despite of “salt-and-pepper” effects, notable shrinkage happens in the periphery of existing  
 555 urban land, reflecting intensity of redevelopment near downtown area.



556

557 **Fig. 11.** The oscillated renewal of urban patch (after <sup>44</sup>)

558 In order to find out the transition mechanism of urban expansion, four transition matrices for  
 559 land use change in four periods are obtained (Table 5, 6, 7, 8). Derived from that, Sankey  
 560 diagram can vividly show flow of land use changes across time. As can be seen from Fig. 12,  
 561 all types of land use flow through year columns, where the amount of fluxes in and out keeps  
 562 constant. It is in the intervals that the land use transitions occur.

563 **Table 5**

564 Transition matrix for 2005-2010 Unit (km<sup>2</sup>)

Constr. land	Cultiv. land	Forest	Urban land	Water area	Total
--------------	--------------	--------	------------	------------	-------

Constr. land	112.5568	138.5992	2.483745	37.07076	1.888615	292.5991
Cultiv. land	278.5777	3993.911	98.89591	203.4437	25.86005	4600.689
Forest	15.12361	186.8649	355.5045	6.379493	6.200732	570.0732
Urban land	100.9727	174.8352	2.660908	368.0879	3.192459	649.7492
Water area	13.67253	61.72074	1.765433	15.30958	386.4265	478.8948
Total	520.9033	4555.931	461.3105	630.2914	423.5684	6592.005

565 **Table 6**

566 Transition matrix for 2010-2015 Unit (km<sup>2</sup>)

	Constr. land	Cultiv. land	Forest	Urban land	Water area	Total
Constr. land	132.6396	211.8874	3.916026	165.75	6.80069	520.9937
Cultiv. land	283.4857	3874.567	140.485	224.4975	32.76978	4555.805
Forest	11.2351	65.09388	380.1576	3.871004	0.976144	461.3337
Urban land	37.53482	144.4981	2.060864	439.2687	6.936454	630.2989
Water area	1.743369	31.43602	2.355814	6.195052	381.783	423.5133
Total	466.6386	4327.482	528.9753	839.5822	429.2661	6591.945

567 **Table 7**

568 Transition matrix for 2015-2019 Unit (km<sup>2</sup>)

	Constr. land	Cultiv. land	Forest	Urban land	Water area	Total
Constr. land	169.9787	156.32	27.44262	110.0352	2.860328	466.6369
Cultiv. land	545.8737	3104.223	311.6591	318.4189	46.909	4327.083
Forest	17.08505	50.48425	456.1169	3.637992	1.705446	529.0296
Urban land	70.42236	131.8793	7.642882	622.1443	7.503453	839.5924
Water area	6.929806	44.86672	4.084645	2.133693	371.2811	429.296
Total	810.2896	3487.773	806.9461	1056.37	430.2594	6591.638

569 **Table 8**

570 Transition matrix for 2019-2035 Unit (km<sup>2</sup>)

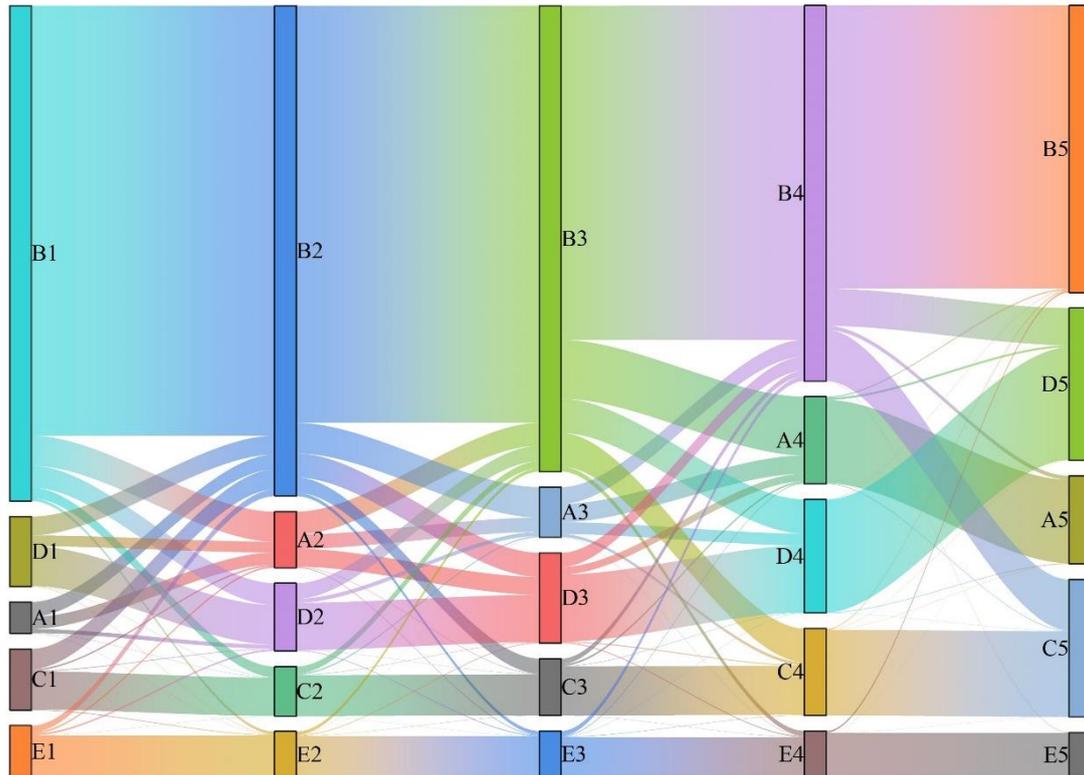
	Constr. land	Cultiv. land	Forest	Urban land	Water area	Total
Constr. land	774.2862	10.13516	6.211778	19.82076	0.025448	810.4793

Cultiv. land	37.71043	2628.713	473.4116	345.4758	3.381191	3488.692
Forest	1.868848	7.176006	795.6388	2.341234	0.1054	807.1303
Urban land	5.309822	3.51069	1.791607	1045.666	0.17246	1056.45
Water area	0.060381	16.47267	0.868985	0.998961	411.9678	430.3688
Total	819.2356	2666.008	1277.923	1414.303	415.6523	6593.121

571

572 In the four phases, the augmented urban land mainly comes from construction land and  
573 cultivated land. The construction land can be large or small sometimes, and inclines converting  
574 into urban land and cultivated land. Cultivated land, which has always accounted for the largest  
575 proportion in all phases, has been decreasing year by year. In 2015, a large part of it went to  
576 construction land, and in 2019, a large proportion of it went to forest and urban land. In general,  
577 the total amount of green space (forest and cultivated land) has not changed much, and the water  
578 area has basically remained unchanged, which has a lot to do with urban ecological protection  
579 policies.

580 The conversion focus of each phase is also different. Since 2005, although the conversion of  
581 cultivated land to urban land has always been the main theme, the focus of 2010-2015 is the  
582 conversion of cultivated land to construction land. The focus of 2015-2019 has added the  
583 conversion of cultivated land to forest. The focus of 2019-2035 will follow the same trend. This  
584 change of orientation shows that urban growth has shifted from increments to stocks, not  
585 because the vitality of the metro is gradually declining, but the invisible effects under the  
586 intervention of the master plan are gradually being guided.



587

588

A: Construction land; B: Cultivated land; C: Forest; D: Urban land; E: Water area

589

**Fig. 12.** Sankey diagram of land use transitions from 2005 to 2035

590

591

592

593

594

595

596

597

598

599

In effect, when using the Kalman filter to predict the amount of future land use, it is found that in 2050 or so, urban land parcels will become saturated, as will construction land. Perhaps a small amount of construction land (for industry, tourism) will contribute to urban land continually, and this balance will not be broken in the short run. On the basis of the land use in 2035 under the metro scenario (because there is no real reference in 2035), the urban land change brought about by metro is minimal compared with the previous years. It is basically a slight expansion at the edge of the existing urban periphery to enhance connectivity. To a certain extent, it shows that the metro's vitality to drive urban sprawl has an upper limit, and it is not an endless incursion of other land to achieve transformation.

Through calculations, the urban growth rate brought by the metro during four periods of

600 2005-2010, 2010-2015, 2015-2019, 2019-2035 are 0.27, 0.38, 0.29, 0.08 respectively.  
601 Assuming that the predicted amount under the metro scenario is 0.7 times the actual amount,  
602 then the rate of 0.08 will be converted to 0.11, which is still less than the ideal value 0.3. In  
603 other words, taking into account the forecast errors, the growth rate of urban brought by the  
604 metro from 2019 to 2035 is still less than the expected value, which has a lot to do with the  
605 upper limit of urban parcels mentioned above.

#### 606 **4 Conclusion**

607 This paper studies the metro impact on Nanjing city's environment from the perspective of  
608 land use change over 30 years. Based on FLUS model of CA, this paper incorporates kalman  
609 filter, originated from optimal state estimation theory to distinguish information from noise in  
610 state-space system. The simulation and prediction results of proposed integrated model can  
611 provide reference for policymaking and spatial planning. The shortcoming of this paper lies in  
612 only considering limited driving factors and no comparisons of other models, so the geo-  
613 simulation will be improved by incorporating up-to-date data and integrating more precise  
614 prediction methods.

615 The main conclusions of this study are as follows:

616 (1) Comparing the two scenarios in the future 2035 (with/without new metro lines), it shows  
617 that the newly expanded urban patches in the metro scenario are mostly scattered on the  
618 edge of existing urban land, such that promote patch connectivity and personnel mobility.  
619 The amount of urban parcel is predicted to reach saturation in 2050, which approximate that  
620 in 2035. Construction land and other land use types will be in a dynamic balance, and urban  
621 boundaries will basically be shaped. The development focus of future urban parcel is meant

622 to be in new city and new town along the Yangtze river and metro axis;

623 (2) The simulation and prediction results demonstrate that in the four phases of 2005-2010,  
624 2010-2015, 2015-2019, and 2019-2035, the transition fluxes of five land use types are  
625 flowing in and out mutually, and the transition focus of each phase gradually deviates  
626 towards more environmental friendly development. The prevailing transition trend is still  
627 the conversion of cultivated land into urban land. Metro stations have played an important  
628 role in enhancing the expansion intensity of nearby construction land, so it is necessary to  
629 protect the ecological red line, by implementing more stringent basic cultivated land  
630 protection policies;

631 (3) With the metro lines expanding year by year, the CA model can assist city planners in  
632 scenario prediction and decision-making. In addition, the simulation of long-term series (in  
633 a 5-year step) using Kalman filter, can analyze uncertainty of state-space system with good  
634 accuracy, which is beneficial to understand the temporal and spatial evolution of land use  
635 coupled with metro and sort out the symbiotic relationship between them.

### 636 **Acknowledgements**

637 This work is supported by Sino-Swiss Science and Technology Cooperation (SSSTC:  
638 IZLCZ2123929). Sincere thanks are given for the comments and contributions of anonymous  
639 reviewers and the editor for their valuable comments.

640

641 **References**

- 642 1 Globe\_Times. *China's 'ghost towns' come to life thanks to urbanization and policies*  
643 <https://www.globaltimes.cn/content/1211265.shtml> (2020).
- 644 2 Bloom, D. E. 7 Billion and Counting. *Science* **333(6042)**, 562-569, doi:10.1126/science.1209290 (2011).
- 645 3 Broere, W. Urban underground space: Solving the problems of today's cities. *Tunnelling and*  
646 *Underground Space Technology* **55**, 245-248, doi:10.1016/j.tust.2015.11.012 (2016).
- 647 4 Hunt, D. V. L., Makana, L. O., Jefferson, I. & Rogers, C. D. F. Liveable cities and urban underground  
648 space. *Tunnelling and Underground Space Technology* **55**, 8-20, doi:10.1016/j.tust.2015.11.015 (2016).
- 649 5 Šipetić, N., Savić, M. & Furundžić, D. The invisible metro system: The case study of the Belgrade metro  
650 system planning. *Tunnelling and Underground Space Technology* **83**, 485-497,  
651 doi:10.1016/j.tust.2018.08.012 (2019).
- 652 6 Besner, J. Underground space needs an interdisciplinary approach. *Tunnelling and Underground Space*  
653 *Technology* **55**, 224-228, doi:10.1016/j.tust.2015.10.025 (2016).
- 654 7 von der Tann, L., Sterling, R., Zhou, Y. & Metje, N. Systems approaches to urban underground space  
655 planning and management – A review. *Underground Space* **5**, 144-166, doi:10.1016/j.undsp.2019.03.003  
656 (2020).
- 657 8 Bobylev, N. Underground space in the Alexanderplatz area, Berlin: Research into the quantification of  
658 urban underground space use. *Tunnelling and Underground Space Technology* **25**, 495-507,  
659 doi:10.1016/j.tust.2010.02.013 (2010).
- 660 9 Li, X., Xu, H., Li, C., Sun, L. & Wang, R. Study on the demand and driving factors of urban underground  
661 space use. *Tunnelling and Underground Space Technology* **55**, 52-58, doi:10.1016/j.tust.2016.02.010  
662 (2016).
- 663 10 Peng, J., Peng, F.-L., Yabuki, N. & Fukuda, T. Factors in the development of urban underground space  
664 surrounding metro stations: A case study of Osaka, Japan. *Tunnelling and Underground Space Technology*  
665 **91**, doi:10.1016/j.tust.2019.103009 (2019).
- 666 11 Pan, Z., Wang, G., Hu, Y. & Cao, B. Characterizing urban redevelopment process by quantifying thermal  
667 dynamic and landscape analysis. *Habitat International* **86**, 61-70, doi:10.1016/j.habitatint.2019.03.004  
668 (2019).
- 669 12 Bennett, M. M. & Smith, L. C. Advances in using multitemporal night-time lights satellite imagery to  
670 detect, estimate, and monitor socioeconomic dynamics. *Remote Sensing of Environment* **192**, 176-197,  
671 doi:10.1016/j.rse.2017.01.005 (2017).
- 672 13 Qi, Q. Present state, problems and development trends of urban underground space in China.  
673 *Tunnelling and Underground Space Technology* **55**, 280-289, doi:10.1016/j.tust.2015.11.007 (2016).
- 674 14 Zhu, E. *et al.* Carbon emissions induced by land-use and land-cover change from 1970 to 2010 in Zhejiang,  
675 China. *Sci Total Environ* **646**, 930-939, doi:10.1016/j.scitotenv.2018.07.317 (2019).
- 676 15 Jiang, P., Cheng, L., Li, M., Zhao, R. & Duan, Y. Impacts of LUCC on soil properties in the riparian zones  
677 of desert oasis with remote sensing data: a case study of the middle Heihe River basin, China. *Sci Total*  
678 *Environ* **506-507**, 259-271, doi:10.1016/j.scitotenv.2014.11.004 (2015).
- 679 16 Liu, X. *et al.* A future land use simulation model (FLUS) for simulating multiple land use scenarios by  
680 coupling human and natural effects. *Landscape and Urban Planning* **168**, 94-116,  
681 doi:10.1016/j.landurbplan.2017.09.019 (2017).
- 682 17 Jaeger, J. A. G. & Schwick, C. Improving the measurement of urban sprawl: Weighted Urban Proliferation  
683 (WUP) and its application to Switzerland. *Ecological Indicators* **38**, 294-308,

- 684 doi:10.1016/j.ecolind.2013.11.022 (2014).
- 685 18 Batty, M., & Xie, Y From cells to cities. *Environment and Planning B-planning & Design* **21(7)**, 31-48,  
686 doi:10.1068/b21S031 (1994).
- 687 19 Wolfram, S. Cellular automata as models of complexity. *Nature*, **311(5985)**, 419-424,  
688 doi:10.1038/311419a0 (1998).
- 689 20 Liu, G., Chen, S. & Gu, J. Urban renewal simulation with spatial, economic and policy dynamics: The  
690 rent-gap theory-based model and the case study of Chongqing. *Land Use Policy* **86**, 238-252,  
691 doi:10.1016/j.landusepol.2019.04.038 (2019).
- 692 21 Basse, R. M., Omrani, H., Charif, O., Gerber, P. & Bódis, K. Land use changes modelling using advanced  
693 methods: Cellular automata and artificial neural networks. The spatial and explicit representation of land  
694 cover dynamics at the cross-border region scale. *Applied Geography* **53**, 160-171,  
695 doi:10.1016/j.apgeog.2014.06.016 (2014).
- 696 22 Li, X. & Gar-On Yeh, A. Data mining of cellular automata's transition rules. *International Journal of*  
697 *Geographical Information Science* **18**, 723-744, doi:10.1080/13658810410001705325 (2004).
- 698 23 Jokar Arsanjani, J., Helbich, M., Kainz, W. & Darvishi Boloorani, A. Integration of logistic regression,  
699 Markov chain and cellular automata models to simulate urban expansion. *International Journal of*  
700 *Applied Earth Observation and Geoinformation* **21**, 265-275, doi:10.1016/j.jag.2011.12.014 (2013).
- 701 24 Xie, Y. A generalized model for cellular urban dynamics. *Geographical Analysis* **28(4)**, 350-373,  
702 doi:10.1111/j.1538-4632.1996.tb00940.x (1996).
- 703 25 Verburg, P. H. *et al.* Modeling the spatial dynamics of regional land use: the CLUE-S model. *Environ*  
704 *Manage* **30**, 391-405, doi:10.1007/s00267-002-2630-x (2002).
- 705 26 Davidson, E. A. *et al.* The Amazon basin in transition. *Nature* **481**, 321-328, doi:10.1038/nature10717  
706 (2012).
- 707 27 Lauf, S., Haase, D., Hostert, P., Lakes, T. & Kleinschmit, B. Uncovering land-use dynamics driven by  
708 human decision-making – A combined model approach using cellular automata and system dynamics.  
709 *Environmental Modelling & Software* **27-28**, 71-82, doi:10.1016/j.envsoft.2011.09.005 (2012).
- 710 28 Jantz, C. A., Goetz, S. J. & Shelley, M. K. Using the Sleuth Urban Growth Model to Simulate the Impacts  
711 of Future Policy Scenarios on Urban Land Use in the Baltimore-Washington Metropolitan Area.  
712 *Environment and Planning B: Planning and Design* **31**, 251-271, doi:10.1068/b2983 (2016).
- 713 29 Li, X. & Yeh, A. G.-O. Modelling sustainable urban development by the integration of constrained cellular  
714 automata and GIS. *International Journal of Geographical Information Science* **14**, 131-152,  
715 doi:10.1080/136588100240886 (2000).
- 716 30 Dietzel, C., Oguz, H., Hemphill, J. J., Clarke, K. C. & Gazulis, N. Diffusion and Coalescence of the  
717 Houston Metropolitan Area: Evidence Supporting a New Urban Theory. *Environment and Planning B:*  
718 *Planning and Design* **32**, 231-246, doi:10.1068/b31148 (2016).
- 719 31 Liu, D. *et al.* Interoperable scenario simulation of land-use policy for Beijing–Tianjin–Hebei region,  
720 China. *Land Use Policy* **75**, 155-165, doi:10.1016/j.landusepol.2018.03.040 (2018).
- 721 32 Crassidis, J. L., Lai, K.-L. & Harman, R. R. Real-Time Attitude-Independent Three-Axis Magnetometer  
722 Calibration. *Journal of Guidance, Control, and Dynamics* **28**, 115-120, doi:10.2514/1.6278 (2005).
- 723 33 Xiancui Wei, Z. S. Formation Control for Multi-robot System using Adaptive Kalman Filter Algorithm.  
724 *International Conference on Robotics and Biomimetics* 6(2012).
- 725 34 Chen, P., Qi, C., Wu, L. & Wang, X. Waveform Design for Kalman Filter-Based Target Scattering  
726 Coefficient Estimation in Adaptive Radar System. *IEEE Transactions on Vehicular Technology* **67**,  
727 11805-11817, doi:10.1109/tvt.2018.2875314 (2018).

728 35 Siouris, G. M., Chen, G., Wang, J.R. Tracking an incoming ballistic missile using an extended interval  
729 Kalman filter. *IEEE Trans. Aerosp. Electron. Syst.* **33(1)**, 232-240 (2002).

730 36 Maklouf, O. M., Halwagy, Y.E., Beumi, M., Hassan, S.D. Cascade Kalman filter application in GPS\INS  
731 integrated navigation for car like robot. *Radio Sci. Conf. IEEE, New Cairo*, 1-15 (2009).

732 37 Rodger, J. A. Toward reducing failure risk in an integrated vehicle health maintenance system: A fuzzy  
733 multi-sensor data fusion Kalman filter approach for IVHMS. *Expert Systems with Applications* **39**, 9821-  
734 9836, doi:10.1016/j.eswa.2012.02.171 (2012).

735 38 Mehrara, M. & Rafiei, F. Energy Consumption and Economic Growth: Kalman Filter Approach.  
736 *International Journal of Academic Research in Economics and Management Sciences* **3**,  
737 doi:10.6007/IJAREMS/v3-i3/1005 (2014).

738 39 Greg Welch, G. B. An Introduction to the Kalman Filter. *University of North Carolina at Chapel Hill*  
739 (1995).

740 40 Brown, R. G. a. P. Y. C. H. Introduction to Random Signals and Applied Kalman Filtering. *University of*  
741 *North Carolina at Chapel Hill* (1992).

742 41 Yang, X., Zheng, X.-Q. & Lv, L.-N. A spatiotemporal model of land use change based on ant colony  
743 optimization, Markov chain and cellular automata. *Ecological Modelling* **233**, 11-19,  
744 doi:10.1016/j.ecolmodel.2012.03.011 (2012).

745 42 Shu, B. *et al.* Modelling multi-regional urban growth with multilevel logistic cellular automata.  
746 *Computers, Environment and Urban Systems* **80**, doi:10.1016/j.compenvurbsys.2019.101457 (2020).

747 43 Li, G., Sun, S. & Fang, C. The varying driving forces of urban expansion in China: Insights from a spatial-  
748 temporal analysis. *Landscape and Urban Planning* **174**, 63-77, doi:10.1016/j.landurbplan.2018.03.004  
749 (2018).

750 44 Xia, C., Zhang, A. & Yeh, A. G.-O. Shape-weighted landscape evolution index: An improved approach  
751 for simultaneously analyzing urban land expansion and redevelopment. *Journal of Cleaner Production*  
752 **244**, doi:10.1016/j.jclepro.2019.118836 (2020).

753 45 Deng, X., Huang, J., Rozelle, S. & Uchida, E. Growth, population and industrialization, and urban land  
754 expansion of China. *Journal of Urban Economics* **63**, 96-115, doi:10.1016/j.jue.2006.12.006 (2008).

755 46 Li, X., Zhou, W. & Ouyang, Z. Forty years of urban expansion in Beijing: What is the relative importance  
756 of physical, socioeconomic, and neighborhood factors? *Applied Geography* **38**, 1-10,  
757 doi:10.1016/j.apgeog.2012.11.004 (2013).

758 47 Qu, S. *et al.* Interaction between urban land expansion and land use policy: An analysis using the DPSIR  
759 framework. *Land Use Policy* **99**, doi:10.1016/j.landusepol.2020.104856 (2020).

760 48 Lan, F., Gong, X., Da, H. & Wen, H. How do population inflow and social infrastructure affect urban  
761 vitality? Evidence from 35 large- and medium-sized cities in China. *Cities* **100**,  
762 doi:10.1016/j.cities.2019.102454 (2020).

763 49 Feng, Y. *et al.* Spatiotemporal characterization of megaregional poly-centrality: Evidence for new urban  
764 hypotheses and implications for polycentric policies. *Land Use Policy* **77**, 712-731,  
765 doi:10.1016/j.landusepol.2018.06.022 (2018).

766 50 Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: from error visibility  
767 to structural similarity. *IEEE Trans Image Process* **13**, 600-612, doi:10.1109/tip.2003.819861 (2004).

768 51 Zhang L, Z. L., Mou X, et al. FSIM: A Feature Similarity Index for Image Quality Assessment. *IEEE*  
769 *Transactions on Image Processing* **20(8)**, 2378-2386, doi:10.1109/TIP.2011.2109730 (2011).

770 52 Bhandari, A. K., Singh, V. K., Kumar, A. & Singh, G. K. Cuckoo search algorithm and wind driven  
771 optimization based study of satellite image segmentation for multilevel thresholding using Kapur's

773 **Figure captions**

774 Fig. 1. Workflow chart of CA-Kalman filter model

775 Fig. 2. Kalman filter loop

776 Fig. 3. Study area and layout of existing metro lines

777 Fig. 4. Presentation and classification of Landsat imagery of Nanjing city

778 Fig. 5. Driving factors for land use change in 2005, 2010, 2015, 2019

779 Fig. 6. Comparison between actual and simulated land use change

780 Fig. 7. Land use prediction using Kalman filter from 2005 to 2050

781 Fig. 8. Simulated results for 2035

782 Fig. 9. Comparison of urban land under two scenarios for 2035

783 Fig. 10. Urban growth boundary as of four year-sections

784 Fig. 11. The oscillated renewal of urban patch

785 Fig. 12. Sankey diagram of land use transitions from 2005 to 2035

786 **Table captions**

787 Table 1 The operating and planned metro lines of Nanjing

788 Table 2 Data preparation for land use change simulation

789 Table 3 Cost matrix for each land use type

790 Table 4 The comparison of image indicators for model simulation accuracy

791 Table 5 Transition matrix for 2005-2010

792 Table 6 Transition matrix for 2010-2015

793 Table 7 Transition matrix for 2015-2019

794 Table 8 Transition matrix for 2019-2035