Supplementary Material

1. Data

Table S1. Dataset description

Data name	Year	Data format	Coordinate system	Data source
China provincial boundary	2017	Polygon shp	GCS_WGS_1984	Socioeconomic Data and Applications Center (SEDAC)
Nanjing boundary	2017	Polygon shp	GCS_WGS_1984	NSTI-GEODATA.CN
Nanjing DEM	2015	Geotiff	GCS_WGS_1984	United State Geological Survey Website (USGS)
Nanjing 1:250,000 inhabitants	2015	Point shp	GCS_WGS_1984	NSTI-GEODATA.CN
Distribution of employment population in Nanjing	2015	Point shp	GCS_WGS_1984	National Bureau of Statistics of China (NBS)
Historical data of land use classification in Nanjing	2000, 2005, 2010, 2015	Raster	Krasovsky_1940_Albers	NSTI-GEODATA.CN
Nanjing road network	2017	Polyline shp	WGS_1984_UTM_Zone_50N	Openstreetmap (OSM)
Restricted construction areas	2019	Polygon shp	WGS_1984_World_Mercator	NSTI-GEODATA.CN

Note: All the input data are transformed to WGS_1984_UTM_Zone_50N for modelling process.

Table S2. Development indicator and goal transformation for Nanjing's Master Plan for 2035

Category	Indicator	Development goal	Goal transformation
Construction land development	Total size of urban land construction (km^2)	≤2150	Determine the minimum value of land conversion probability
	Urban development boundary area (km^2)	≤1998	Determine the maximum amount of urban growth land
Ecological land development	Protected area of permanent farmland (km^2)	≥1960	Determine the basic farmland location and area
	Forest coverage (%)	≥31.5	Determine the minimum forest cover area
	Water coverage (%)	≥11.4	Determine the minimum water cover area
	Basic ecological protection area (km^2)	≥1403	Determine the minimum ecological protection area
Shared land development	Per capita land for urban and rural construction (m^2)	≥134	Determine the minimum urban land area demand per capita under each scenario

Note: Indicators and goal data were obtained from Master Plan of Nanjing 2035.

2. LEAM model description

The Land Use Evolution and Impact Assessment Model (LEAM) is a robust tool for urban spatial simulation. LEAM is generally a network analysis-based dynamic spatial model that uses cellular automata (CA) as a core. Similar to the operational theory of CA, LEAM consists of a simulation environment in a grid space where the cell properties would be transformed according to defined transformation rules and vicinities. A set of initial drivers and projections will be imported as referenced factors for the land use change simulation before the establishment of cell transformation rules. The drivers cover physical geographic (water, soil, slope and other landforms) and socio-economic (residence, employment, road network, administrative boundaries and other planning areas) aspects. The urban land use change drivers indicate the complex interaction between the urban system and the surrounding environment. All factors combine and interact in a variety of ways in the model to assign probabilities of potential land use changes to each 30 m * 30 m cell in the studied grid space. The distribution of this urban built-up land change probability values is the result of the general superposition of local and global effects. The distribution includes the causal change mechanisms, such as the accessibility of cells to city attractors, the constraints of the ecological environment, the state of the regional economy, the social-economic impact and the stochastic disturbance. Typically, the final change possibility of each land use cell from urban unbuilt area to built-up area is defined as:

$$P_{i,t} = \alpha(A_{i,t} + \mathrm{N}(\theta_{i,t-1}))\mathrm{F}(R_{i,t}), \tag{1}$$

where $P_{i,t}$ is the land use change probability for land use cell i at time step t, α is a stochastic disturbance parameter that makes the generated patterns to be closer to reality, $\mathrm{F}(R_{i,t})$ is the function of multiple growth restrictions and planning zoning effects on land use types for land use cell i at time step t and $A_{i,t}$ is the accessibility value for land use cell i at time step t and is defined as:

$$A_{i,t} = \frac{\sum_{j=1}^{n} a_{i,j} w_{i,j}}{n}, \tag{2}$$

where $a_{i,j}$ is the attraction power of land use cell i to urban attractor cell j, whereas $w_{i,j}$ is the corresponding weight and n is the total number of the attractors.

This study aims to quantify how urban built-up areas are shaped by location-choice factors including population and employment centres, highway and major streets, forest and water resources and to compare the agglomeration and dispersion of developed lands in different urban areas. The location-choice factors are defined as attractors in this study under the assumption that they determine the surrounding development in a gravity-type function in which the attraction power decays with increased distances. This gravity function can be determined through the shortest distance algorithm and data value from various sources. This study uses Pan et al.'s (2018) parallel Stochastic Greedy Algorithm (SGA) to find the shortest distance and the inverse distance model to determine the attraction value for population, employment and transportation attractors. In this inverse model, for each attractor j, its attraction to land use cell i is noted as $a_{i,j}$ and is calculated as:

$$a_{i,j} = \sum_{k \in S_j} \frac{p_k}{d_{ij}}, \tag{3}$$

where S_k is the set of attractors of type j. p_k is the attraction value of the k_{th} attractor in S_k, d_{ij} is the distance between the k_{th} attractor in S_k and land use cell i calculated by SGA.

The attractor types are described below:

- Population attraction (S_1). p_k is the total population at place k according to the 2010 Chinese Census. A large a_{ij} indicate additional population or a close distance to the population centre.
- Employment attraction (S_2). p_k is the total employment at place k according to the Hoover data directory. A large a_{ij} indicate additional employment or a close distance to the establishment.
- Transportation attraction (S_3). p_k is the posted speed at major road k. A large a_{ij} indicate a high level of road classification or a close distance to major road.

In addition, $\mathrm{N}(\theta_{i,t-1})$ in formula (1) is the function convert the nearest neighbouring effects to a probability value, and is defined as:

$$\mathrm{N}(\theta_{i,t-1}) = \frac{\sum_1^k (N_{pr,t-1} + \sigma_i)}{k}, \quad (4)$$

where $N_{pr,t-1}$ is the development possibility for neighbourhoods cells in time step $t-1$, whereas σ_i is a spread coefficient over the total surrounding cells ($k \leq 8$).

3. Scenario-based spatial dynamic modelling result

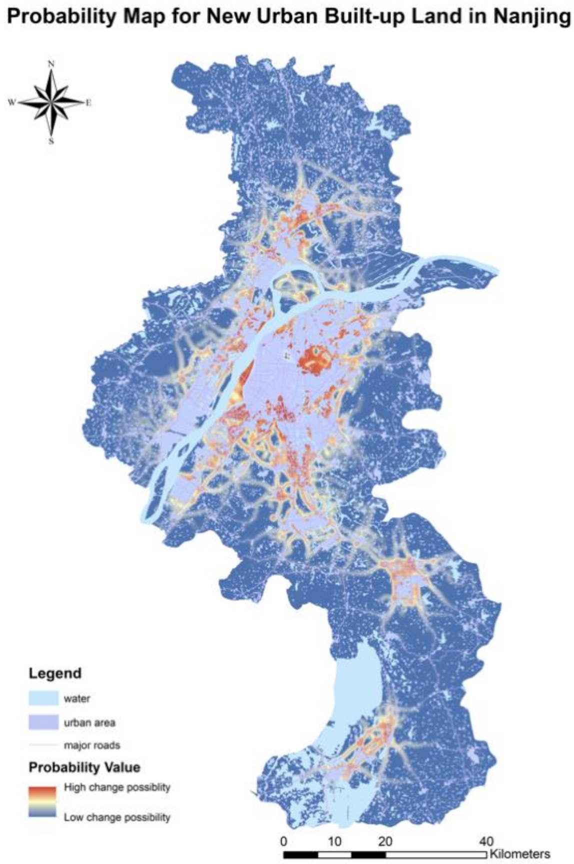

Figure S1. Land use change probability map for Nanjing 2035, example of the comprehensive and coordinated development mode.

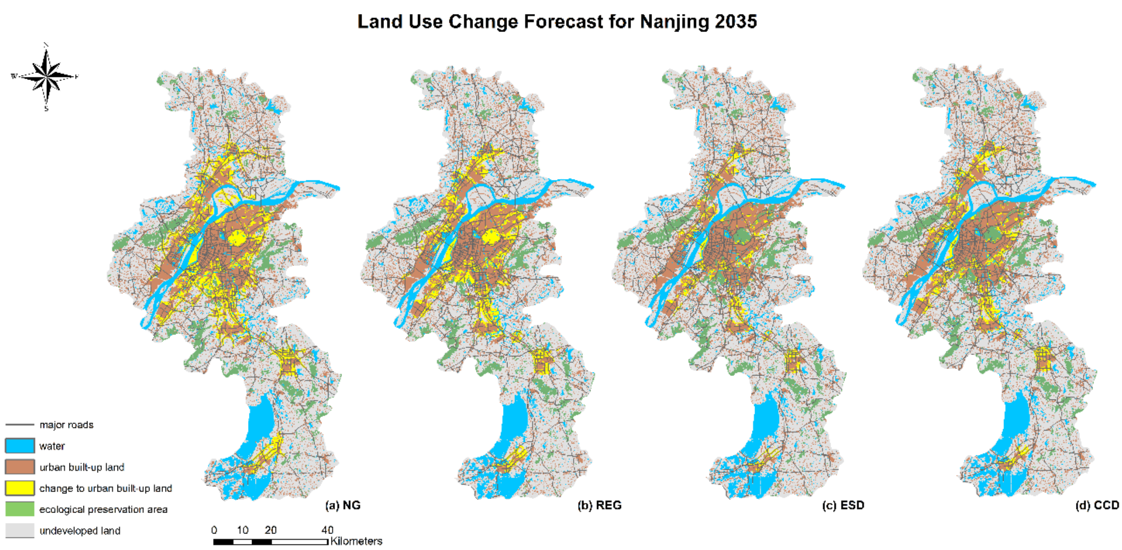

Figure S2. Land use changes forecast for Nanjing 2035 under four scenarios. (a) Natural growth mode. (b) Rapid economic growth mode. (c) Ecological and sustainable development mode. (d) Comprehensive and coordinated development mode.

Table S3. Results of changed area and direction of urban land use for the four scenarios

	Scenario	**EP-BL (km^2)**	**UD-BL (km^2)**	**Total changed area (km^2)**	**Total urban built-up area (km^2)**
a	Natural growth	72.36	517.52	589.88	2,236.66
b	Rapid economic growth	55.75	441.64	497.39	2,144.17
c	Ecological and sustainable development	0	309.68	309.68	1,956.46
d	Comprehensive and coordinated development	0	368.67	368.67	2,015.45

Note: EP-BL and UD-BL refer to the changed land area from ecological preservation area to urban built-up land and from undeveloped land to urban built-up land, respectively.