Annals of the American Association of Geographers

ISSN: 2469-4452 (Print) 2469-4460 (Online) Journal homepage: http://www.tandfonline.com/loi/raag21

Mapping Block-Level Urban Areas for All Chinese Cities

Ying Long, Yao Shen & Xiaobin Jin

To cite this article: Ying Long, Yao Shen & Xiaobin Jin (2016) Mapping Block-Level Urban Areas for All Chinese Cities, Annals of the American Association of Geographers, 106:1, 96-113, DOI: 10.1080/00045608.2015.1095062

To link to this article: http://dx.doi.org/10.1080/00045608.2015.1095062

Published online: 16 Nov 2015.

Submit your article to this journal

Article views: 122

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
http://www.tandfonline.com/action/journalInformation?journalCode=raag21

Download by: [Hong Kong Polytechnic University] **Date:** 28 December 2015, At: 22:18

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

Mapping Block-Level Urban Areas for All Chinese Cities

Ying Long,*,† Yao Shen,‡ and Xiaobin Jin¶

*Beijing Institute of City Planning
†Beijing Key Lab of Capital Spatial Planning and Studies
‡Space Syntax Laboratory, Bartlett School of Architecture, University College London
¶School of Geographic and Oceanographic Sciences, Nanjing University

As a vital indicator for measuring urban development, urban areas are expected to be identified explicitly and conveniently with widely available data sets, thereby benefiting planning decisions and relevant urban studies. Existing approaches to identifying urban areas are normally based on midresolution sensing data sets, low-resolution socioeconomic information (e.g., population density) in space (e.g., cells with several square kilometers or even larger towns or wards). Yet, few of these approaches pay attention to defining urban areas with high-resolution microdata for large areas by incorporating morphological and functional characteristics. This article investigates an automated framework to delineate urban areas at the block level, using increasingly available ordnance surveys for generating all blocks (or geounits) and ubiquitous points of interest (POIs) for inferring density of each block. A vector cellular automata model was adopted for identifying urban blocks from all generated blocks, taking into account density, neighborhood condition, and other spatial variables of each block. We applied this approach for mapping urban areas of all 654 Chinese cities and compared them with those interpreted from midresolution remote sensing images and inferred by population density and road intersections. Our proposed framework is proven to be more straightforward, time-saving, and fine-scaled compared with other existing ones. It asserts the need for consistency, efficiency, and availability in defining urban areas with consideration of omnipresent spatial and functional factors across cities. *Key Words: China, points of interest (POIs), road network, urban block, vector cellular automata.*

城市地区由于作为评估城市发展的生动指标，因而被期待能够以广泛可及的数据集明确且便利地进行指认，藉此加惠规划决策和相关的城市研究。指认城市地区的既有方法，一般是根据中度辨识率的遥测数据集、空间中(例如具有数平方公里的区块，甚至是更大的乡镇或行政区）低度辨识率的社会经济信息(例如人口密度)。但这些方法鲜少关注透过纳入形态与功能之特徵，以高度辨识率的微观数据为大型区域界定城市区域。本文探讨自动架构以描绘街廓层级的城市地区，使用逐渐可及的地形测量以生产所有街廓(或地理单位）与普遍存在的兴趣点（POIs）来推断各街廓的密度。本文採用向量细胞自动机模型，从所有生成的街廓中指认城市街廓，并将各街廓的密度、邻里条件与其他空间变异纳入考量。我们将此方法应用于绘製中国共六百五十四座城市的城市地区地图，并将其与从中度辨识率的遥测影像、以及从人口密度与道路交口推断的地区进行比较。与其他既有的方法相较之下，我们所提出的架构证实更为直接、省时、且尺度精密。该架构主张界定城市地区时必须具有一致性、效率与可及性，并考量广佈各城市的无所不在的空间与功能因素。 *关键词： 中国，兴趣点 (POIs)，路网，城市街廓，向量细胞自动机。*

Como indicador vital para medir el desarrollo urbano, se espera que las áreas urbanas sean identificadas explícita y convenientemente con conjuntos de datos ampliamente disponibles, fortaleciendo así las decisiones de planificación y los estudios urbanos relevantes. Los enfoques existentes para identificar las áreas urbanas normalmente se basan en conjuntos de datos de sensores a resolución intermedia, información socioeconómica de baja resolución (e.g., densidad de población) en el espacio (e.g., celdas de varios kilómetros cuadrados o incluso pueblos más grandes o distritos). No obstante, pocos de estos enfoques le prestan atención a definir las áreas urbanas con datos micro de alta resolución para áreas grandes, incorporando características morfológicas y funcionales. Este artículo investiga un marco automático para delinear áreas urbanas a nivel de manzanas o cuadras, usando los cada vez más comunes servicios de cartografía para generar todas las manzanas (o geounidades) y ubicuos punto de interés (POIs) para inferir la densidad de cada manzana. Se adoptó un modelo *vector cellular* autómata para identificar manzanas urbanas a partir de todas las manzanas generadas, tomando en cuenta la

Annals of the American Association of Geographers, 106(1) 2016, pp. 96–113 © 2016 by American Association of Geographers
Initial submission, December 2014; revised submissions, April and July 2015; final acceptance, July 2015
Published by Taylor & Francis, LLC.

densidad, la condición del vecindario y otras variables espaciales de cada manzana. Aplicamos este enfoque para cartografiar las áreas urbanas de todas las 654 ciudades chinas, y las comparamos con las interpretadas de imágenes de percepción remota a resoluciones intermedias y deducidas de la densidad de población y las intersecciones de carreteras. El esquema que proponemos es probado como más directo, económico en tiempo y de escala fina en comparación con otros disponibles. Este esquema reivindica la necesidad de consistencia, eficiencia y disponibilidad para definir áreas urbanas con la consideración de los factores espaciales y funcionales presentes por doquier a través de las ciudades. *Palabras clave: China, puntos de interés, red de carreteras, manzana o cuadra urbana, modelo vector celular autómata.*

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

A universal difficulty for urban studies is how to properly define a city (Zipf 1949; Krugman 1996; Batty 2006). Urban areas play a strong role in representing urban spatial development for planning decisions, management, and urban studies. They not only illustrate spatial patterns, such as the development levels and scales of the built environment, but also reveal socioeconomic unevenness within built-up areas, thereby representing how a city evolves in a complex manner (Batty 2012).

Conventional methods of capturing the borders of a built-up area from the top down have been applied in major cities around the world, mainly relying on mid-resolution sensing data sets or socioeconomic distributions (e.g., population density) associated with low-resolution settings (e.g., cells with several square kilometers or even larger towns or wards). Although there is growing evidence that more accurate mapping results of urban areas can be generated when following the progress of remote sensing technologies and availability of census data, the applicability of these mapping approaches has been debated. First, such methods cannot be applied to most cities in developing countries due to lack of the necessary data or fine digital equipment (Long and Liu 2013). Moreover, these existing methods still require multiple steps according to unique conditions if a fine-scaled result is expected. Furthermore, these existing approaches seem to isolate the spatial characteristics from the functional ones; therefore, the real urban activities seem to be absent in capturing the urban areas by existing methods.

In the past thirty years, cellular automata (CA) has increasingly attracted attention in understanding the growth of urban areas, providing a perspective to simulate urban change from the bottom up. By using the CA model, mapping of urban areas can successfully simulate the continuity of urban areas based on spatial proximity and allow for modeling the interaction between urban lands defined by urban activities. Furthermore, vector cellular automata (VCA) is able to use the defined urban blocks as basic cells during the simulation process; hence, the results of urban areas will provide the same spatial units adopted in urban planning and real practice. In other words, the adopted unit in the VCA model best matches the applied component in Chinese planning. Besides, the fine-scaled data sets of road networks and points of interest (POIs) covering all geographic areas in China secure the consistency of generative resolutions of urban blocks among all Chinese cities through the proposed VCA model. Thus, supported by the ubiquitous data, mapping the urban areas of a VCA model can provide an understanding of the urban usage of space in reality instead of a virtual environment.

This study is a manual for making a "block-up" understanding of urban areas for all Chinese cities. Urban areas here are defined as the merging of urban blocks enclosed by roads with more urban activities than nonurban areas. Dissecting the existing body of influential approaches to defining urban areas and examining cases identifies a link between two dominant factors that are instrumental in achieving the goal. The first is the urban block demarcated by roads, which could be considered the basic unit of urban areas; the second is urban density, which is the key function characteristic, manifested in density of POIs. Our findings indicate that the proposed framework could be used as an open and direct approach by inputting high-resolution, ubiquitous data to capture small block-based urban areas. This study aims to provide a straightforward and time-saving way to gain insight into complex urban systems across cities from the bottom up and give the common phenomena for measuring urban sprawl consistently during Chinese rapid urbanization, along with its policy implications.

Background

Existing Definitions and Methods for Mapping Urban Areas

The concept of the urban area, although widely applied, discussed, and referred to, is simultaneously ambiguous. In the existing literature, it encompasses

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

various descriptions, measurements, and applications, spanning various issues and spatial scales in different nations. Urban areas in the United States are identified as urbanized areas (UAs) in a typical administrative model for spatial statistics containing the incorporated places and census-designated places in central places and urban fringes controlling for the population density (Morrill, Cromartie, and Hart 1999). One similar term in Japan is *densely inhabited district* (DID), with a population density over 4,000 people per square kilometer. In China, the records of "one book and two certificates"[1] within administrative areas are widely accepted. Furthermore, UAs in the United Kingdom are derived from construction-built areas where certain real estate densities are detected through satellite images or other data sets (Y. Hu et al. 2008). On the other hand, socioeconomic factors are also adopted to describe active urban areas; for example, labor force markets and commuter sheds used to represent metropolitan areas (MAs) in the United States (Berry, Goheen, and Goldstein 1969). Consequently, the empirical and theoretical literature seems to rule out the possibility that an urban area is the geographical field where the real urban activities happened. Current definitions, however, fail to identify urban areas through explicitly bonding the spatial and functional dimensions. The real hidden mechanism that drives urban developments should be uncovered to illustrate the complexity of defining urban areas.

Parallel to the ambiguity of its definitions, there are various distinguished methods for mapping urban areas. From the morphological perspective, remote sensing images and road networks have received increasing academic attention. Remote sensing and nighttime satellite images help to filter rural areas on the basis of transferring land cover information or scanned light brightness to indexes (Henderson et al. 2003; He et al. 2006). In addition, various geometric characteristics of road networks have been introduced to identify the spatial organization of cities as physical entities, including road intersection density (Masucci, Stanilov, and Batty 2012), fractal indexes (Shen 2002; Tannier et al. 2011; Jiang and Yin 2013; Tannier and Tomas 2013), and size of urban blocks (Jiang and Liu 2012). In terms of the functional aspects, applying socioeconomic statistics such as demographic densities (Rozenfeld et al. 2009), effective employment density (COAG Reform Council 2012), and infrastructure accessibility (Y. Hu et al. 2008) have emerged as a standard method of defining urban areas. When more precise outcomes with higher resolution are expected, however, the disadvantages of these approaches are evident. First, remote sensing data–based approaches are limited by time-consuming interpretation steps and image resolution. Second, to accurately define cities, using the geometrical approach to directly link to specific spatial units, such as a block or tile, is difficult despite its advantage in cities with diverse sizes at an extremely large scale. Finally, spatial statistics methods are very time-consuming to prepare, limited by fine-scaled censuses, and the results are most likely to be altered significantly if the survey frequency is low.

Consequently, many challenges need to be overcome before a universal model can be established from the methodological perspective. The foremost challenge is the question of how fine-scaled spatial units can be set and which one will be suitable for urban studies and planning practice. Furthermore, the data quality for different cities might not permit the development of a universal approach. In addition, a difficult task is finding a straightforward way to generate an urban indicator describing urban areas while considering functional aspects. The consistency of methodological applicability for various cities is always limited by the refined data sets covering very large geographic areas. The timeliness of conventional approaches should be improved to generate temporal results so that instant urban growth can be monitored.

Computational Urban Areas Mapping

Computer simulations have been used for about fifty years, aiming for scientific investigations about modeling urban areas and their changes. Most of the aforementioned relevant studies adopt complexity theory and methods including CA modeling, artificial neural network (ANN), or agent-based modeling (ABM). Since the 1990s, these two models have been employed in the fields that study land use and land cover change (LULCC), urban extension, and urban morphology (White and Engelen 1993; Cecchini 1996; Batty, Xie, and Sun 1999). They significantly develop the accuracy of simulation and improve the weakness of spatial extrapolation methods (e.g., Markov chain model). In comparison with ABM, CA modeling is argued to be more efficient for computing urban areas and their spatial continuity based on more simplified principles from the bottom up, although the individual decision process is not factored (Couclelis 1985).

One core theme in defining urban areas through the CA model is the lattice setting. For solving the modifiable areal unit problem (MAUP), the VCA model is

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

argued to be closer to reality than the pure CA model (Stevens and Dragicevic 2007). On the other hand, if the statues transition rules in the CA model can be well supported by the fine data sets, urban areas and their dynamics would be reflected explicitly by CA simulation (Chen and Mynett 2003; S. Hu and Li 2004). This evidence implies the strengths of VCA models for building up more accurate and realistic knowledge about urban areas.

Data-Driven Attempts

Recently, these challenges of urban area delineation gave rise to methodological developments that address the same issue from the bottom up based on detailed street network maps and volunteered geographic information (VGI). Several studies concentrated on extracting a block-based urban area from the transport layer in OpenStreetMap (OSM; Jiang and Liu 2012; Jiang, Liu, and Tao 2013). Yet, a pure road network–based approach is hardly effective for generating fine-grained blocks and inferring urban blocks using the head–tail division rule (globally applied) to reflect temporal urban activities. With this background taken into account, some studies have been conducted to use POIs for inferring the function performance of auto-generated blocks so that urban blocks could be selected locally. Yuan, Zheng, and Xie (2012) segmented Beijing into disjointed blocks through the raster-based model, and their functional characteristics were inferred by incorporating POIs and taxi trajectories. Long and Liu (2013) proposed an approach called automatically identifying and characterizing parcels (AICP) by using OSM and POIs in 297 Chinese cities at the national scale. They compared the efficiency and accuracy of the approach with those of other methods. Apart from previous research, these two studies shed light not only on the autogeneration of blocks but also on the functional qualification in terms of online volunteered data (e.g., POIs). The unevenness of the resolution of OSM among various cities (Hagenauer and Helbich 2012), however, still limits the applicability of these methodologies for all cities and the resolution of the results.

In summary, the new data sets in very high resolution and the CA models show strong prospects by addressing the issues of defining urban areas in terms of data availability and methodological advances. Although the overall effectiveness of conventional methods of capturing urban areas has been proven in many studies, its accuracy, adoptability, immediacy, and consistency are restricted by methodological convenience and data availability. In this article, we intend to fill this gap of knowledge by delimiting urban areas based on block-based patterns of urban functions.

Method

Redefining Urban Area

Urban area in this article is defined as the continuous urban built-up blocks (known as *dikuai* in Chinese) within an administrative context, where dense urban activities are agglomerated and more likely to interact temporally. In this sense, several critical dimensions of urban areas are well focused on and further incorporated: administrative identity, spatiality, functionality, continuity, and temporality.

Specifically, administrative identity refers to the locality of urban areas, which infers governmental ownership of urban areas; spatiality denotes that the areas should be reflected by spatial entities—for instance, the blocks are enclosed by built-up roads; functionality stands for the fact that urban areas should be recognized to carry a certain density or intensity of urban activities; continuity refers to the phenomenon that urban areas tend to be developed as continuous fields or a patch-like group due to geographic interaction; and temporality here refers to a requirement that the urban areas should replicate the fields where real activities happen temporally. All of these factors will be reflected in the process of determining whether land is urbanized or not.

The Proposed Framework

The empirical framework for delineating urban areas contains three steps based on well-propagated data: block generation, urban block selection (vector CA module), and urban area production (block mergence; Figure 1). In the first step, all possible blocks are defined based on the fine-scaled road layers in the ordnance survey. In the next stage, blocks are inferred with their geometrical and geographical properties and POIs density to automatically select the urban blocks in a VCA approach. Finally, all of the urban blocks are dissolved and mapped, thereby clearly generating urban areas. All of these steps are illustrated fully in the following sections.

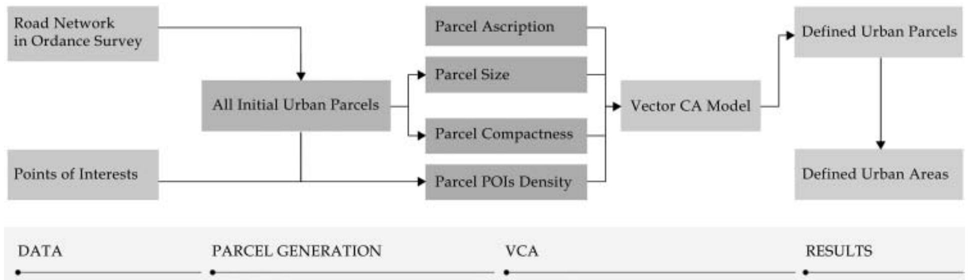

Figure 1. Flowchart of the proposed framework. *Note:* VCA = vector cellular automata.

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

Steps

Step 1: Generating Blocks and Inferring Their Density

Blocks are important spatial units for contemporary urban planning and design and urban studies. In this subsection, a block is defined as a continuous built-up area enclosed by roads. Supporting this idea, all of the possible urban blocks are generated by using the road layer in an ordnance survey. Before generating blocks, the road layers are processed according to their hierarchy before being merged as a single layer. More specifically, all segments are connected with a 20-m tolerance, and street segments shorter than 200 m are trimmed to avoid cul-de-sacs. This threshold selection is reliant on the basic judgment of collected spatial data sets. Moreover, the width of all roads is also defined, relying on their hierarchy. Finally, all initial blocks are presented when the roads are removed from the study areas.

Four properties are further calculated for each block. The first two refer to the geometric characteristics of each block, including the size and compactness determined by the block's shape. In addition, accessibility of each block is taken into consideration as a locational variable for describing a block. Another characteristic is the functional attribute of a block for reflecting its actual use. The number of POIs within or close to a block are featured as its urban density. Due to the natural unevenness of urban density between big cities and the small ones, POI density is further normalized and placed between 0 and 1 to release the heterogeneity among cities. Because of lack of further attributes of the importance of POIs, the popularity of each POI is assumed as the same in this study. When any substitutions are available, they can be expected to approximate the intensity of urban activities explicitly. Hence, in the way of using the road network and POIs to identify initial blocks, the spatial and functional features are incorporated together to present the suitability of blocks.

Step 2: Selecting Urban Blocks by Using VCA

Vector-based constrained CA models are used for selecting urban blocks from the initial ones generated by road networks in diverse cities. It is assumed in this article that this process is similar to modeling urban expansion, which extensively uses CA applications. Apart from the conventional raster CA model (Batty, Xie, and Sun 1999), vector-based CA models depend on irregular polygons rather than regular cells. In this research, each block is regarded as a cell with a valve that is 0 (urban) or 1 (nonurban). Based on existing studies (Li and Yeh 2000; Li, Yang, and Liu 2008; X. Liu et al. 2008; J. Liu, Zhang et al. 2010), this can be illustrated with the following formula:

$$S_{ij}^{t+1} = f\left(S_{ij}^{t},\ \Omega_{ij}^{t}, Con,\ N\right). \tag{1}$$

Here, a block's status at $t+1$ is considered as a function f of the block's statues and other proposed factors at t. In this function, $S_{ij}{}^{t}$ and S_{ij}^{t+1} denote the statues of blocks at time points of t and $t+1$, respectively; $\boldsymbol{\Omega}_{ij}{}^{t}$ is the neighboring situation; Con refers to the constraints, and N is the number of all blocks. This function can be further transferred to a detailed probability formula:

$$P_{ij}^{t} = (P_l)_{ij} \times (P_\Omega)_{ij} \times con(.) \times P_r. \tag{2}$$

In this function, the possibility of transforming a block's state at t is illustrated as the multiplied product of probabilities of factors. Specifically, $(P_l)_{ij}$ stands for the local potential that a block would convert its status from the nonurban to the urban, and $(P_\Omega)_{ij}$ denotes the conversion possibility in terms of the neighboring definition; $con(\cdot)$ stands for constraints; and P_r is the stochastic term.

The proposed spatial and functional characteristics are reflected in measuring the local potential.

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

This could be explained in the following formula using a logistic regression model (Wu 2002):

$$(P_l)_{ij} = \frac{1}{1+\exp\left[-\left(a_0+\sum_{k=1}^{m} a_k c_k\right)\right]}, \quad (3)$$

where a_0 is a constant, a_k is an estimated coefficient responding to the spatial variable c_k, and m is the total number of spatial variables. As a result, spatial and functional factors are bonded to reflect a block's state in this study. Block size is measured in the logarithm for calculating the polygon's area. Compactness of each block is calculated as the rate of the perimeter square subdivided by its area. Accessibility is abstracted using the minimum Euclidian distance to the city center. On the other hand, the functional factor is presented by applying the standardized POIs density, which is measured through calculating the rate of raw density in the maximum density within the study area.

The neighboring potential for a block is measured by the amount of peripheral urban blocks around it. This can be defined as

$$(P_\Omega)_{ij} = \frac{\sum con\left(S_{ij}^t = urban\right)}{n} \quad (4)$$

For block ij, $con(S_{ij}{}^t = urban)$ stands for the urban blocks within fixed areas, and n is the sum of all accessible blocks. The adjacent relation is defined as 500 m around the block ij.

Two layers—the steep area (a slope over 25 degrees) and various water bodies—are included as restriction conditions. Urban expansion is impossible in these areas. The constraints are expressed as $con(cell_{ij}{}^t = suitable)$ with a value of 0 or 1, where 1 indicates that there is no restriction on the block's development as urban, whereas 0 indicates that the block is restricted from development.

The stochastic disturbance P_r in the model stands for any possible change of local policies and accidental errors. It is calculated using

$$P_r = 1 + (-\ln\gamma)^\beta, \quad (5)$$

where γ is a random number ranging from 0 to 1, and β, ranging from 0 to 10, controls the effect of the stochastic factor.

Furthermore, by comparing the measured probability $(P_l)_{ij}$ with a calculated threshold value P_{thd}^t in the tth interation, the block's status at $t + 1$ can be detected. The threshold value in the tth iteration equals the final potential of the nth block (P_n^t) in the case that the accumulated urban area for the initial n blocks with greatest transit potential reaches to the defined total amount of urban area in the tth iteration. In other words, the urban blocks are captured from highest to the lowest potential until the accumulative urban area reach as the area limitation for each iteration in each city. The number of iterations (Num_{itr}) is predefined by the user, and then the total area of urban blocks ($TotalArea$) for each iteration ($stepArea^t$) is calculated accordingly. In this study, the statistics for urban area in 2012 for each city are obtained from Ministry of Housing and Urban-Rural Development (MOHURD) records (MOHURD 2013) and project that the number of iterations run in the simulation process for each city is 100. This method can balance the computation time and the quality of results. If the measured value is greater than the threshold, the block is considered to be urban; if not, the block will stay as nonurban. This progress can also be presented as shown here:

$$S_{ij}^{t+1} = \left\{ Urban \ for \ P_{ij}^t > P_{thd}^t \quad NonUrban \ for \ P_{ij}^t \leq P_{thd}^t \right. \quad (6)$$

$$P_{thd}^t = P_n^t \left(\sum_n Area_n^t \leq stepArea^t \right) \quad (7)$$

$$stepArea^t = \frac{TotalArea}{Num_{itr}}. \quad (8)$$

The whole simulation process of the CA model for every city in this study is shown in Figure 2, in which a general picture of parameter calibration and transit rule generation is illustrated. Accordingly, the iteration maps are also showcased in Figure 3 to present the typical transit process in the projected CA model in this study. First, the prerequired data sets are input to the model as basic variables. Specifically, before the iteration process starts, a group of initial urban blocks is defined according to their ranks of local potentials and the total area of urban blocks in each step ($stepArea$) is calculated according to the reported urban area in official documents and the setting of iteration times. Spatial variable selection and coefficients are further computed based on the historic urban development situation of typical cities. In every iteration, the local potential, conversion possibility, and stochastic disturbance are calculated and further packaged as the final transit potential. In the step of allocation, the block with the maximum final transit probability is converted from nonurban to urban in the allocation process until the cumulative area

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

of defined urban blocks reaches the proposed amount of urban area for every step. The iteration process is conducted for the proposed times, ensuring that the reported total area of urban blocks is reached. The whole simulation is ended when a block-based urban area pattern is generated.

Step 3: Mapping Urban Areas Using Selected Urban Blocks

Because street spaces and small unselected urban blocks surrounded by urban ones are also included in urban areas in planning practices, the selected urban blocks need to be transformed to urban areas. To map the urban areas of all cities in China, the selected urban blocks are remerged into the integrated urban lands in ESRI ArcGIS (Version 10.2, ESRI, Redlands, CA) using the toolbox function Aggregate Polygons to present the urban areas for each city. This tool is used for moderate scale reduction and aggregation on selected urban blocks. Aggregation will only occur where two blocks are within the specified aggregation distance to each other. According to the requirements of Chinese urban blocks, the distance to be satisfied between block boundaries for aggregation to happen is set to 500 m and the minimum area for an aggregated block to be retained is 1 ha. In addition, orthogonally shaped output urban areas are created for preserving the geometric characteristic of anthropogenic urban blocks. The projected approach is conducted in all 654 reported cities individually to speed up the block aggregation process. Urban areas of each city can then be mapped based on selected urban blocks.

Step 4: Model Validation

For validating our proposed model, self-validation and external justification are performed separately for

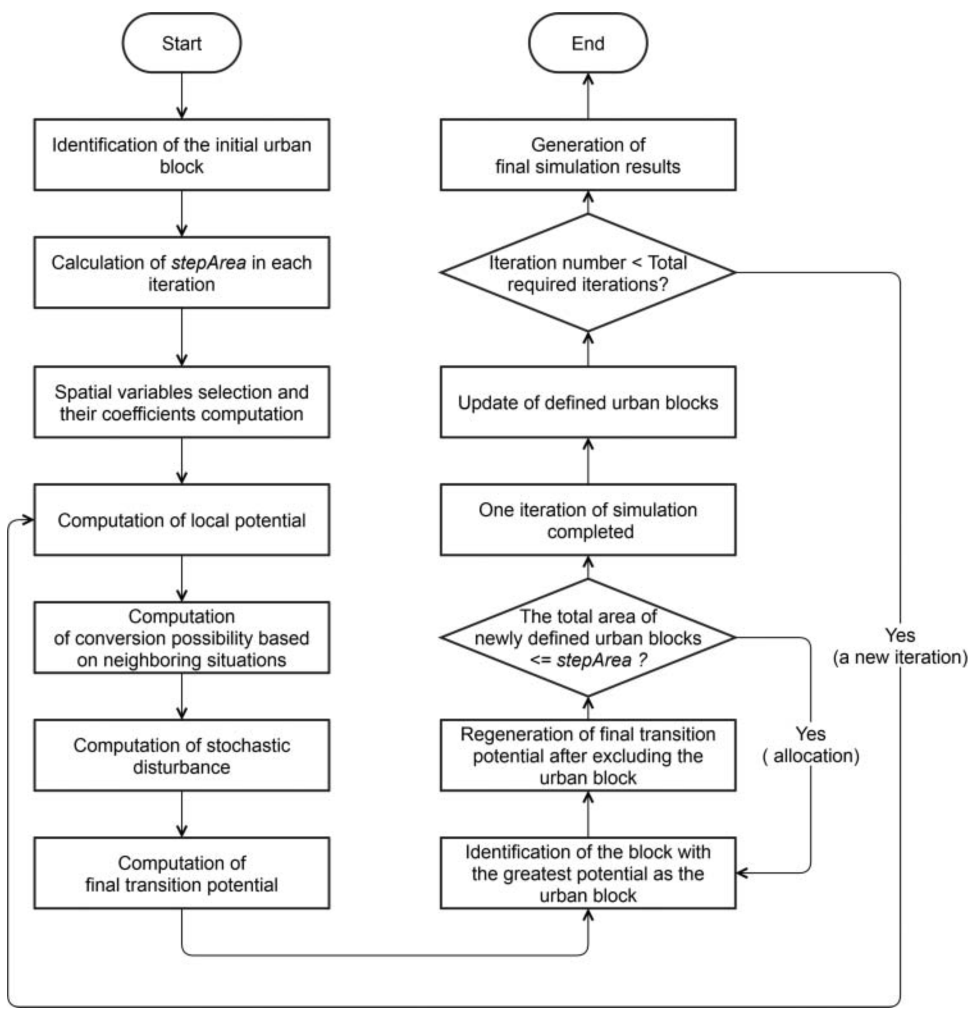

Figure 2. Flowchart of proposed cellular automata model.

Figure 3 The interaction maps to select urban blocks at t and $t + 1$ time: (A) the initial blocks enclosed by urban streets; (B) identification of initial urban blocks (in red); (C) captured urban blocks at t time (in red); (D) selection of new urban blocks at $t + 1$ time (in blue). (Color figure available online.)

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

cross-verifying the applicability in delineating urban areas among various cities. For the self-validation, all cities are ranked to detect the scaling law of city size horizontally on the one hand; urban blocks within typical cities at various administrative levels are also rated for finding the linear relation of a logarithmic scale rules of block sizes on the other hand. In addition, the automatic generated result of urban areas for all Chinese cities is compared with the outcomes produced by several classical methods reviewed in the Introduction section based on geographical coverage data sets containing remote sensing images, census-based population density, and road intersection density. Additionally, the results are evaluated with other data, including online check-in data sets and POIs, proving their effectiveness at reflecting temporal urban intensity of activities.

Process as a Constrained Inversion

The whole process of automatic identification of urban areas could be described as holistic constrained inversion of urban areas (Figure 4), which is a method for speculating very large and complex urban areas based on a relatively small amount of typical observed data. To avoid subjective factors setting, this model requires the verified effective constraints or parameters in some typically observed cases at the first stage. This process can be considered a partial inversion, a way to generate key constraints in defining urban areas. In this study, the urban density measured by POIs density is proven to be a key factor in identifying urban blocks in each city. To better define it in so-called holistic constrained inversion, there are two steps: The first part is about block segmentation and the second part is for blocks' mergence. These two steps interact with each other using POIs density and other factors to select

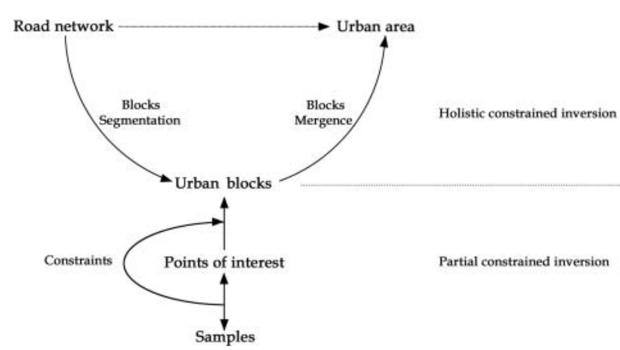

Figure 4. Constraints inversion of identifying urban areas.

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

urban blocks. This proposed open framework has several potentials to address the questions of identifying urban areas: First, it offers a method to generate block-based urban areas in a large-scale manner, relying on universal rules discovered in typical samples; moreover, it can be used as a reference to validate the surveyed urban areas; and, finally, it could further imply the potential role of an omnipresent data set in duplicating urban areas from a vast scope. Therefore, not only for promoting our model of delineating urban areas based on POIs and road networks, this study advocates an open framework for presenting block-up distributions of urban areas by combining holistic and partial inversions at different levels. That is to say, the model discussed in this article is an open system combining local equation-based analysis and global simulation, which is ensured for future development within the context that location-based data are increasingly available.

Data

Administrative Boundaries of Chinese Cities

Administrative boundaries of 654 Chinese cities[2]—the limitations of local geographies—are applied to partition the research areas into city boundaries so that ordnance survey maps and POIs can be curved off accordingly (Figure 5). According to the Chinese administrative system (MOHURD 2013), there are five levels of cities classified in this way: municipalities directly led by the nation (MD, 4 cities), subprovincial cities (SPC, 15 cities), other provincial cities (OPCC, 17 cities), prefecture-level cities (PLC, 250 cities), and county-level cities (CLC, 368 cities). This system also reflects the hierarchy of these cities in terms of city size and population. By doing so, the research scopes are specified to administrative areas of urban lands by keeping the national scale in mind. The principles of defining the administrative boundary of each city are not the same for the local governments, however, as the consequence of differentiation between cities in terms of urban land protection (the ratio of administrative land to all land cover). To achieve an objective spatial statistical result of reported urban area for every city, the applied administrative boundaries are geocoded according to a statistical yearbook with a well-calculated urban area (National Bureau of Statistics 2013).

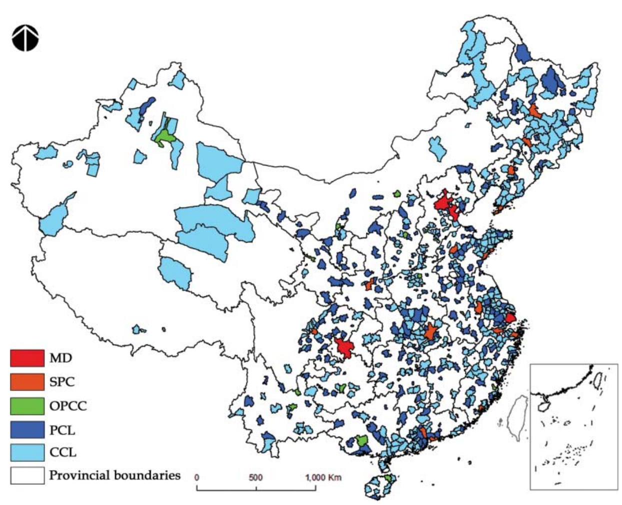

Figure 5. Administrative areas of Chinese cities. *Note:* MD = municipalities directly led by the nation; SPC = subprovincial cities; OPCC = other provincial cities; PCL = prefecture-level cities; CCL = county-level cities. (Color figure available online.)

Total Urban Area of Chinese Cities

Based on defined city administrative boundaries, statistics of urban areas are extracted from MOHURD (2013) to allocate the total into urban blocks in each city. Through 2012, the total urban area of 654 cities in China reached 46,744 km^2. An individual city is inferred by its statistical area accordingly (Figure 6). Consequently, our research areas in all of the cities are specifically featured with their administrative subordination and total urban area.

Road Network in Ordnance Survey and POIs in 2012

The ordnance map is considered to be the authorized map reflecting the most urban information (Haklay 2010). Urban streets, regional roads, and many other detailed streets are encompassed in the Chinese ordnance survey map, which is the most reliable data set obtained from a national institution. The applied data set of road networks in this research is derived from the ordnance survey data set (2013) published by Chinese geographical institutions, which has been compared with online data sets (e.g., Google Maps and Baidu Map) to prove its accuracy. In the study conducted by Long and Liu (2013), the Chinese ordnance survey map was compared with Open Street Map (OSM). The ordnance survey map was found to be far more detailed regarding the total segments and length of roads. The employed database in this study is

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

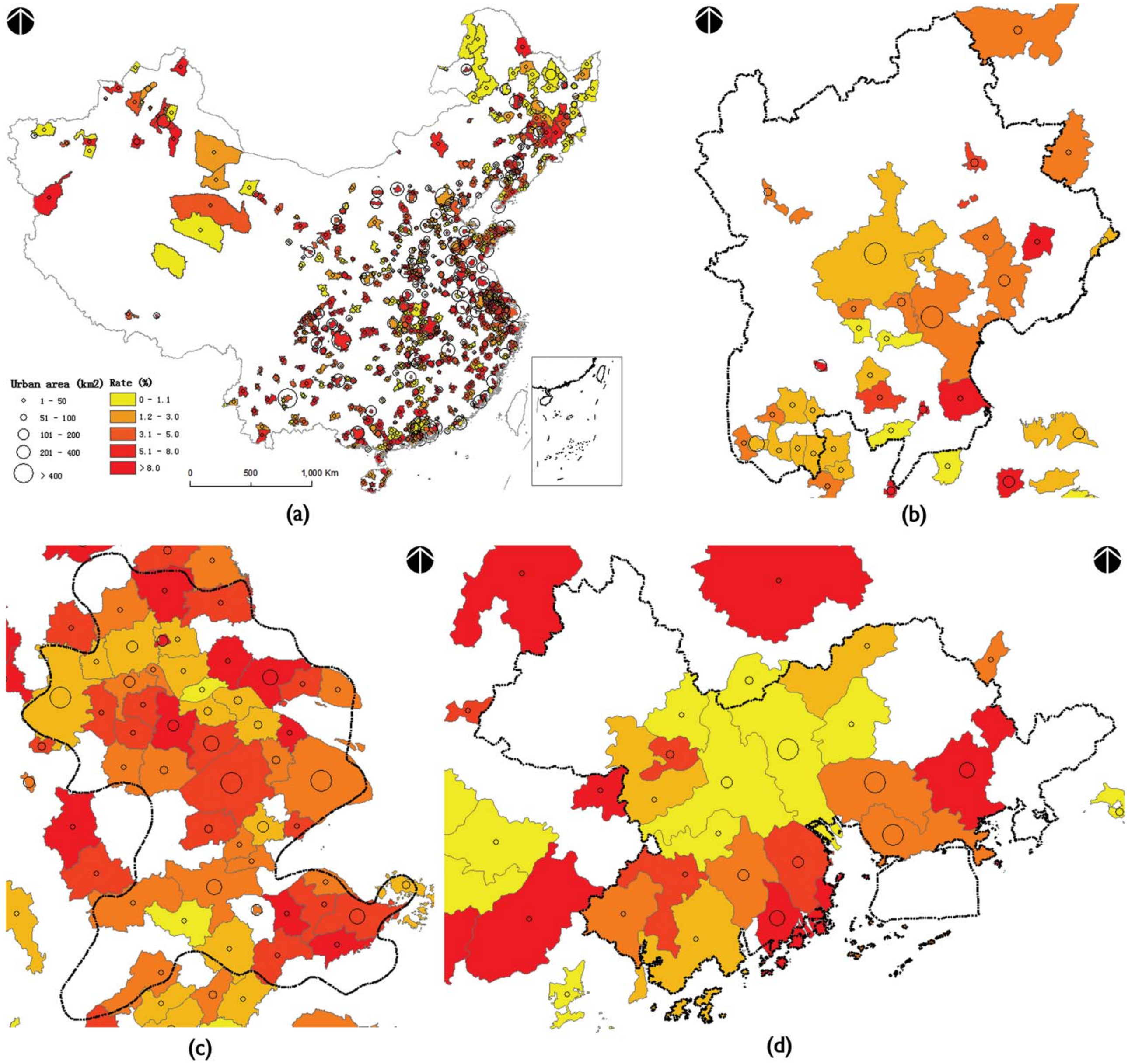

Figure 6. Total urban area in 2012 at the city level. (A) China; (B) Beijing-Tianjin-Hebei (BTH); (C) Yangtze River Delta (YRD); (D) Pearl River Delta (PRD). *Note:* The urban expansion rate during 2007–2012 for each city is also mapped in this figure to show the historical urban expansion of Chinese cities. (Color figure available online.)

made up of 6,026,326 segments with a total length of 2,623,867 km (Figure 7).

POIs containing a total of 5,281,382 points were gathered from business cataloging web sites—Sina Weibo. There are eight main types of POIs in the initial data set, and each type refers to one specified type of urban activity in the data set. The detail contents of POIs are shown in Table 1. All of the POIs are adopted in this empirical study to measure the land use density through calculating their total amount of activities for each generated block. These randomly selected sample data sets have been manually checked to ensure the overall data quality. It is noteworthy that the proposed empirical research framework is extendable in the way that the POIs data set can be replaced by other information regarding the distribution of urban activities.

Results

Model Calibration for VCA Model

Logistic regression is conducted for calibrating the weights for constraints in the proposed VCA models.

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

Figure 7. Ordnance roads of China in 2012: (A) China; (B) a part of the central city of Beijing.

Due to data availability, it is nearly impossible to calculate the weights of controlling factors for each city, thereby reflecting the spatial heterogeneity between cities. Hence, the 2010 blocks data set in Beijing City manually prepared by urban planners in Beijing Institute of City Planning (BICP) is applied as a typical example of all other cities (BICP 2010). It covers an area of 12,183 km^2 at a very fine scale of urban blocks (Yanqing and Miyun counties in the Beijing

Table 1 Classification of point of interest types

Type	Abbreviation	Count
Commercial sites	COM	2,573,862
Office building/office space	OBS	677,056
Transport facilities	TRA	561,236
Government	GOV	468,794
Education	EDU	285,438
Residence communities	RES	167,598
Green space	GRE	13,041
Others	OTH	534,357

Table 2 Binary logistic regression results for BICP blocks

Name	Coefficient	*SE*	Significance
Constant	5.359	0.058	0.000
Natural logarithm of block size	−0.306	0.006	0.000
Distance to the city center	−0.099	0.001	0.000
Points of interest density	3.431	0.085	0.000

Note: BICP = Beijing Institution of City Planning.

Metropolitan Area are excluded from Beijing City). There are 52,330 blocks reported, of which 36,914 blocks are identified as urban.

According to the result of a binary logistic regression (Table 2), 78.9 percent of all blocks can be explained by the generated function. All factors except compactness passed the *p* test, revealing that they are significantly related to the differences between nonurban and urban blocks. These logistic regression results have been employed in VCA models for all Chinese cities. To test the accuracy of this model, the generated results for Beijing City from the CA model were compared with the BICP data set again, and an overall explanatory ability of 81.5 percent on the real urban areas indicates the applicability of our model in delineating urban areas in terms of urban blocks.

Selected Urban Blocks

The proposed constrained VCA model was conducted on all 654 reported cities in China, for which a sum of 707,330 urban blocks with 51,286 km^2 in area was detected and labeled as urban from among all 851,054 initial blocks (Figure 8). The average numbers of urban blocks in cities on various administrative

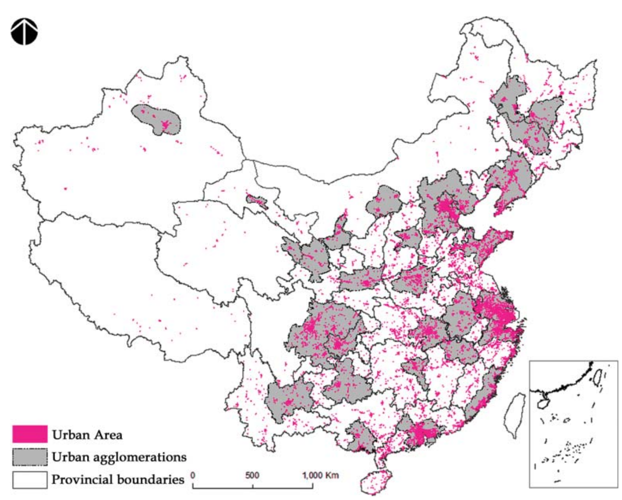

Figure 8. Selected urban blocks for all Chinese cities. (Color figure available online.)

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

levels differ from each other significantly. There are 1,411 urban blocks in an average MD, followed by 407 in SPC, 199 in OPCC, 79 in PLC, and 26 in CLC. When scrutinizing these statistics, the greater the population or higher administrative ranks the cities occupied, the greater number of urban blocks they will have. In other words, the scaling laws of the population or city size can be observed in terms of the amount of urban blocks in each city.

Scaling law is a universal rule not only for natural phenomenon but also for urban areas (Arcaute et al. 2013). On a logarithmic scale, this relationship between the size of urban areas and their frequency distribution should be linear. It also self-validates the proposed VCA models for each city in this article (Vliet, White, and Dragicevic 2009). To verify the performance of the model between cities, the size of all cities (in terms of the number of urban blocks in

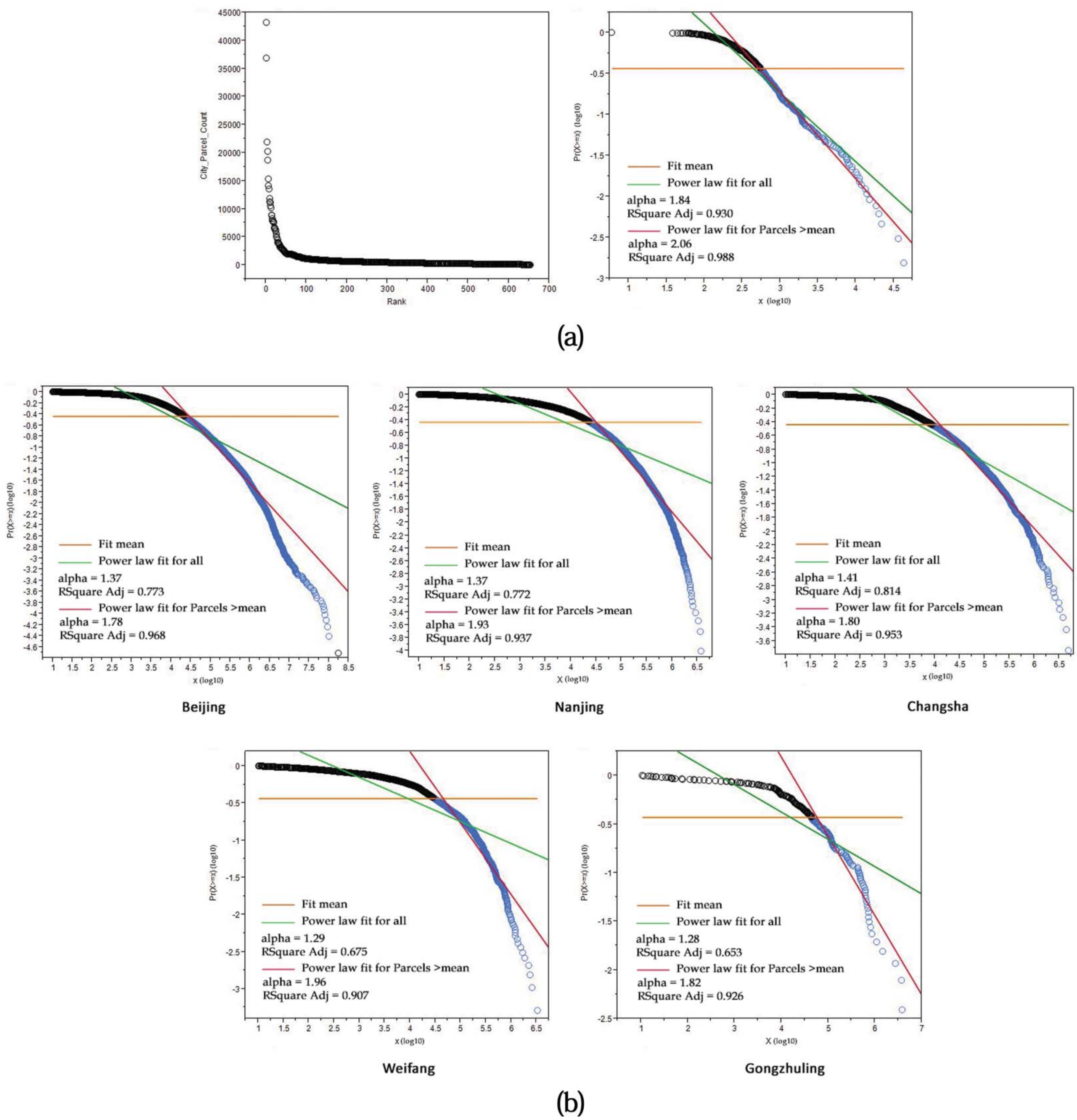

Figure 9. Power law distributions (A) in terms of the block numbers for all 654 cities; (B) in terms of block size for typical cities. (Color figure available online.)

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

each city) is plotted against their ranks (Figure 9A). A shape that reveals a long tail distribution recognizes that there are far more cities with fewer urban blocks than those with a large number of urban blocks (Jiang 2013). When it comes to the log-log distribution, a perfect power law fit (R^2 is 0.988 and alpha is 2.06) can be observed by considering the cities with more urban blocks than the average level. Thus, the significant rank-size pattern with high R^2 values indicates the applicability of our models for all cities.

On the other hand, the power law fit is also adopted in analyzing the ranks of urban blocks' size in typical cities at various governmental levels for understanding the applicability of this proposed approach in each kind of city internally. Generally, the power law fit for all blocks can explain around 70 percent of urban areas. Better regressions are implied in the cities occupying higher administrative levels than those at lower levels. More specifically, the alpha values for Beijing, Nanjing, and Changsha are all above 1.37 and the adjusted R^2 is greater than 0.77, whereas Weifang and Gongzhuling have alpha values of 1.28 and a smaller R^2, which is about 0.65. Removing blocks less than the mean size results in better power law fits (all R^2 increased above 0.9). Shared similar trends emerged, suggesting that our models can be applied in modeling the urban blocks within different kinds of cities.

Urban Areas of All Chinese Cities

Merging all selected urban blocks automatically illustrates urban areas of all Chinese cities. For gaining more insight into these results, the typical cities (e.g., Beijing for MD, Nanjing for SPC, Changsha for OPCC, Weifang for PLC, and Gongzhuling for CLC) on different administrative levels are listed and compared with the results from other data sets including the Defense Meteorological Satellite Program/Operational Linescan System (DMSP/OLS), census-based population density, and road intersections (Figure 10).

Compared with the results from other data sets, the urban area generated through our approach has generally higher resolution than other databases. The outputs captured using the other three data sets are highly correlated to the ones detected in our projected framework in relatively developed cities in terms of initial judgment, which might be mainly on the basis of the fact that there are better digital infrastructures and small census settings for survey in the big cities. This assumption can be verified by comparing the results of developing cities. Thus, our approach can produce more detailed

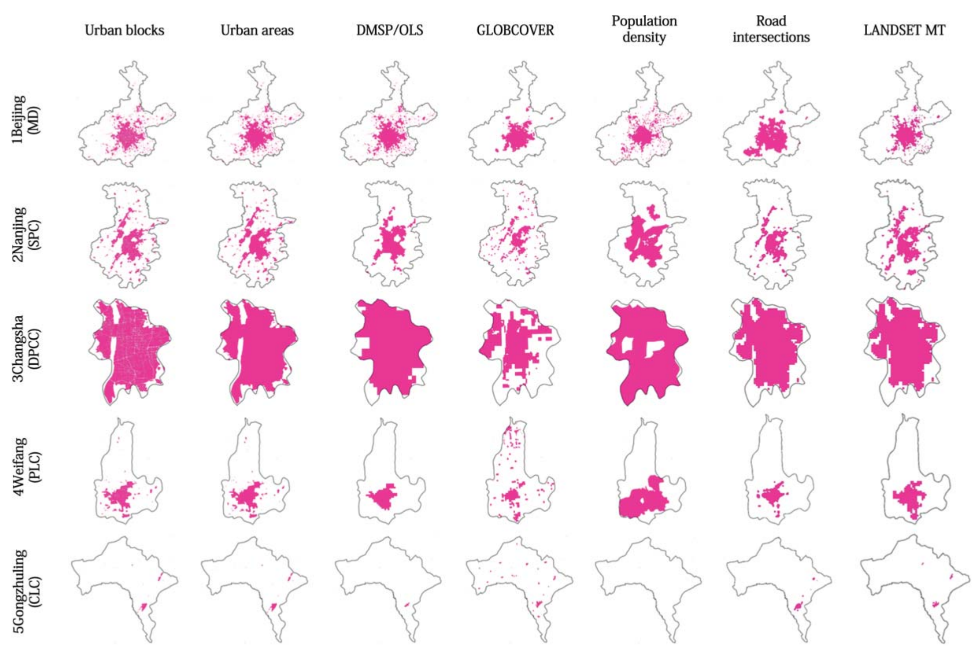

Figure 10. The profile of urban areas for typical cities. *Note:* MD = municipalities directly led by the nation; SPC = subprovincial cities; OPCC = other provincial cities; PLC = prefecture-level cities; CLC = county-level cities; DMSP/OLS = the Defense Meteorological Satellite Program/Operational Linescan System. (Color figure available online.)

results than the other two. The generated urban area was further overlapped with other results to detect the overlapping rates for higher precision. More details can be found in the Model Validation section.

Discussion

Model Validation

Our model validation is conducted by comparing urban areas with five data sets for all 654 Chinese cities: (1) the urban area's space defined at the 300-m resolution in GLOBCOVER (Bontemps et al. 2011); (2) the urban areas presented in the 1-km resolution retrieved from DMSP/OLS in 2008 (Yang et al. 2013); (3) the urban areas represented by subdistricts with population DENSITY greater than the mean density (977 people per square kilometer) of all 39,007 subdistricts of China in 2010 using Jiang's (2013) head–tail division rule and the 2010 population census of China (Wu et al. 2015); (4) the urban areas presented by road INTERSECTION density using the ordnance survey applied in this study. Urban areas are selected in each city by sorting all grids' estimated kernel densities of road intersections while considering the observed total area of a city; (5) the LANDSET TM data set, another authorized land use map, obtained from a recent study regarding LULCC in China (J. Liu et al. 2014).

All of the results are shown in Table 3. In terms of the captured size of urban blocks, the urban blocks in this study (average size is about 300 m * 400 m) are far smaller than the ones reported in the other four data sets, reflecting that better scaled outputs are achieved through our approach. From the perspective of overlapping rate, our study detected that 65.5 percent of common urban areas (a total of 30,606 km^2) in our outputs intersected with DMSP/OLS. With consideration of time mismatch between these two data sets, our suggested approach can be expected to produce accurate results for all Chinese cities depending on the preceding evaluation. The results of this study and the data of GLOBCOVER are not as good as initially expected. There are only 20,801 km^2 of urban area (44.5 percent) in our result that intersect with GLOBCOVER. This might be the result of noncorrespondence between these two data sets regarding time and resolution. A total of 81.9 percent of urban areas fall into the urban category represented by population density, indicating that most of our results are associated with high population density. The overlapped ratio over 80 percent is partially due to the overestimated urban areas in DENSITY, which is nearly three times that stated by MOHURD (2013). The comparison of results between ours and the ones generated using road intersections are highly acceptable (76.8 percent), which could be attributed to the same data source being used in both methods. In our results, 74.2 percent of the urban areas overlapped as urban land cover in LANDSET TM. The correlating rate is higher as the result of the inconsistency between a grid-based map and vector-based patterns.

Admittedly, it is hard to determine which result could be most accurate because each method could reflect one kind of possible answer to the same question. The overall correspondence, however, between the output produced by the introduced approach and the different existing data sets of urban areas informs the effectiveness of the method applied in this article.

To determine the precision of our approach about mapping the real urban intensity of activities used several times in generated urban areas, online check-in data sets collected from one of the largest Chinese check-in Web sites and the POIs data set are adopted. As shown in Table 4, within our defined urban areas for all 654 Chinese cities, 85.9 percent of urban

Table 3 The comparison of urban areas in various data sets for 654 cities in China

Data	Year	Spatial resolution	Urban area (km^2)	No. of patches	Average patchsize (ha)	Intersected with ORDNANCE (km^2)
ORDNANCE	2012	269 m	46,713	18,404	312.5	N/A
DMSP/OLS	2008	300 m	45,834	1,345	3,407.7	30,606 (65.5%)
GLOBCOVER	2009	1 km	39,789	12,701	313.3	20,801 (44.5%)
DENSITY	2010	6.7 km	126,860	728	17,425.8	38,245 (81.9%)
INTERSECTION	2012	500 m	46,703	4,221	1,106.5	35,868 (76.8%)
LANDSET MT	2010	1 km	45,201	2,892	1,562.9	34,644 (74.2%)

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

Table 4 Coverage of urban areas in terms of urban intensity (measured by online check-in data set) and land uses

Data	Year	Covered	Uncovered	Coverage
Online check-in	2013	14,643,090	2,401,023	85.9%
Points of interest	2012	2,987,338	937,390	76.1%

intensity measured by online check-in data is mapped in the results of our urban areas, whereas 76.1 percent of surveyed POIs are for all Chinese cities. This further illustrates that the proposed method in this study can be successfully employed to generate digital urban areas that reflect actual urban activities.

Horizontal Evaluation on Methods for Delineating Urban Areas

In addition to quantitatively comparing our results with methods, there are six dimensions that are considered for qualitatively evaluating the strengths of existing methods, including practicality, geographical scale, result resolution, data availability, methodological convenience, and dynamics (Figure 11). Ten professionals in different planning institutes in China were interviewed and asked to rate the performance of every approach according to their working experience. Practicality here refers to the value for real planning and design projects. Due to the similarly basic spatial units setting, our urban block-based results are generally considered the most straightforward method to reflect actual developments in urban blocks. In the meantime, conventional approaches (e.g., DMSP/OLS, survey block maps, and population density maps) are also labeled as practical tools to help understand urban extent. Moreover, the approach proposed in this study is expected to better balance the dilemma between the cover scale and result resolution in traditional models. Relying on the open data sets in the background of development of VGI data sets, the method projected in this article is also well-thought-out to be a publicly accessible and temporally updatable data set for urban planning and studies. Regarding the methodological convenience, spatial survey and statistics-based methods are understandable, whereas our approach is understood as a direct way of packaging complex simulations in an automatic manner. All of these evaluations are generally based on the reality of urban planning in developing countries, particularly in China, which means that an assessment addressing the same issue would be different in developed nations where urban surveys and statistics have been conducted for many years. It is still worth promoting our produced method in Web 2.0,

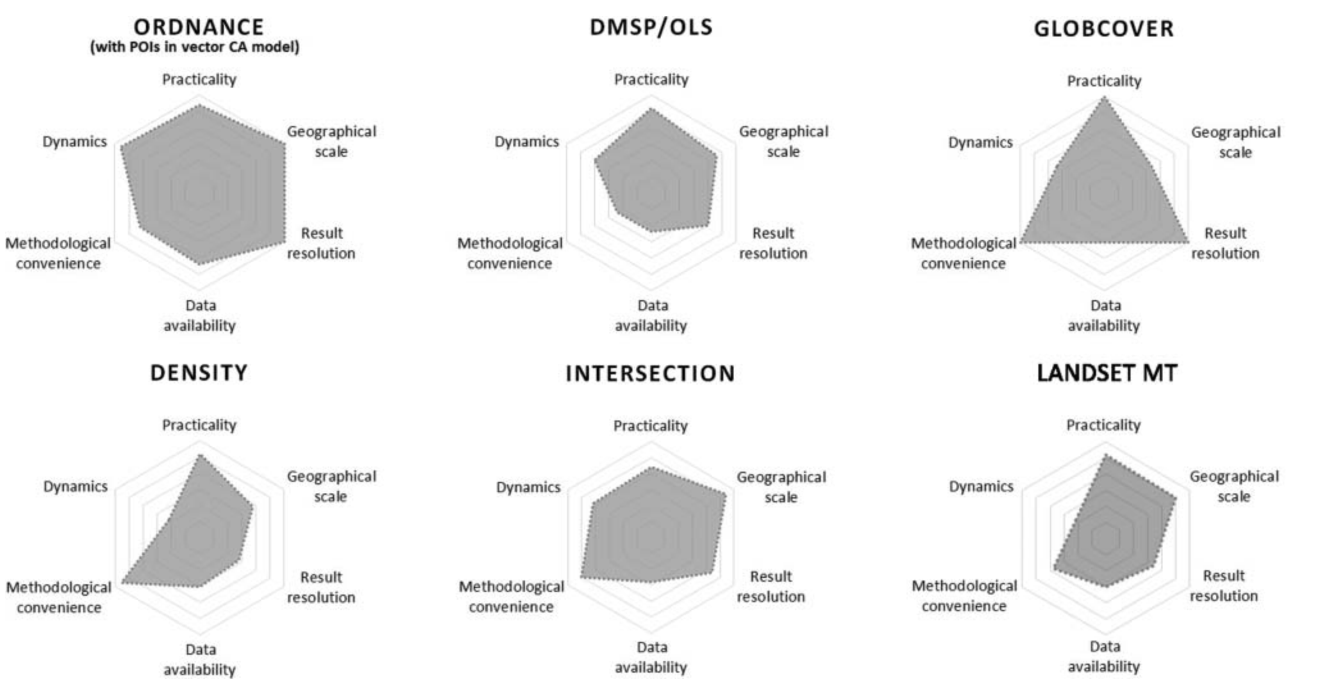

Figure 11. Comparison between existing methods of delineating urban areas. *Note:* POIs = points of interest; CA = cellular automata; DMSP/OLS = the Defense Meteorological Satellite Program/Operational Linescan System.

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

however. It is a large model (Long et al. 2014) in a direct, fine-scaled, and dynamic sense based on omnipresent open data, thereby benefiting the understanding of urban areas for city management and planning.

Potential Bias and Further Steps

This study proposes an automatic framework to generate urban area and provides examples with all Chinese cities. The increasingly available VGI in this framework also promotes the merits of this approach. Nevertheless, several limitations still exist in this study, which will be highlighted in our future research. First, current methodology could be directly improved based on the increase in open data availability. Location-based online information (e.g., check-ins) could infer the weights of POIs, thereby reflecting actual urban usage more accurately. Second, more samples of cities should be used for model calibration to enhance the precision of our approach, even if our methods have already proven the applicability and flexibility of applying local constraints rather than global ones for all cities. Finally, the presently applied city-level urban area statistics could be replaced by fine-scaled ones (e.g., districts or even subdistricts) to control the total area of urban blocks in a more detailed manner.

Conclusion

In this article, a VCA model is proposed based on road networks in ordnance survey and POIs for explicitly delineating block-based urban areas. Urban areas in all 654 cities in China were generated using our approach. The whole process contains several components, including block generation, urban block selection, and urban area production. In the first step, a road network layer in ordnance survey is applied to define blocks by removing buffered roads from the study area. In the following stage, all blocks are equipped with attributes such as size, compactness, accessibility, and POI density. The VCA model is then adopted for identifying urban blocks from among all generated blocks, taking into account the spatial variables of each block as well as conveyed total area in each city. Finally, the urban areas of each city are mapped by aggregating urban blocks. In the process of self-validation, power law fits are detected when analyzing the relationships between generated urban blocks and their ranks across cities and the correlations between blocks' size and the frequency distributions in five typical cities, indicating the applicability of our approach. The final results are also validated by comparing them with urban areas presented by DMSP/OLS, GLOBCOVER, population density, road intersection density, and LANDSET TM maps. After interviewing relevant urban planners, the proposed approach in this article was given a high ranking on various dimensions. In summary, our model is proven to be not only effective in modeling urban areas through incorporating spatial and functional features of urban blocks but also more straightforward, time-saving, and fine-scaled, when compared to other existing models.

The projected framework has the potential to benefit relevant urban studies and policy distributions. Through this study, the current status of urban development can be reflected at a standard level, thus feeding both intracity and inner-city academic studies. Our model would be more helpful for relatively small cities where digital infrastructures are poor and fine-level statistics are hardly secured. On the other hand, it can significantly lower the costs of collecting data temporally without a large investment. Second, the urban area simulation process could promote a deeper understanding of block-up urbanism that reflects the progression of a large site divided and sold off for development. Third, this model can be further developed into an advanced version simulating urban expansion. It might directly benefit predictive urban planning and evaluate the effectiveness of strategies and policies. Fourth, our model can be methodologically helpful in unifying the calibers of defining urban areas among diverse cities based on ubiquitous data and reclaiming the need for consistency, efficiency, and temporal renewability in defining urban areas with omnipresent spatial and functional factors across cities.

Acknowledgments

The authors would like to thank Dr. Kang Wu and Dr. Dong Li for providing China data sets. Our thanks are also given to Stewart Scales for his advice on polishing the text. The mapping results are available at the Beijing City Lab Web site (www.beijingcitylab.com). Yao Shen is the author to whom all correspondence should be addressed.

Funding

This work was supported by a grant from the National Natural Science Foundation of China (No. 51408039).

Notes

1. "One book and two certificates"—containing a proposal of project location, the permit of land planning, and the permit of construction planning—refers to the construction approval files delivered by the government based on urban planning law.
2. Sansha in Hainan and Beitun in Xinjiang appearing in MOHURD (2013) were not included due to spatial data availability. Taiwan was not included in all analysis and results in this article.

References

Arcaute, E., H. Erez, P. Ferguson, H. Youn, A. Johansson, and M. Batty. 2013. City boundaries and the universality of scaling laws. *arXiv*:1301.1674.

Batty, M. 2006. Rank clocks. *Nature* 444 (7119): 592–96.

———. 2012. Building a science of cities. *Cities* 29:S9–S16.

Batty, M., Y. Xie, and Z. Sun. 1999. Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban Systems* 23:205–33.

Beijing Institute of City Planning (BICP). 2010. Existing land use map of Beijing. Internal working report, Beijing Institute of City Plannng, Beijing, China.

Berry, B. J. L., P. G. Goheen, and H. Goldstein. 1969. *Metropolitan area definition: A re-evaluation of concept and statistical practice*. Washington, DC: U.S. Bureau of the Census.

Bontemps, S., P. Defourny, E. V. Bogaert, O. Arino, V. Kalogirou, and J. R. Perez. 2011. *GLOBCOVER 2009-products description and validation report*. Paris: ESA.

Cecchini, A. 1996. Approaching generalized automata with help online (AUGH). In *Artificial worlds and urban studies*, ed. E. Besussi and A. Cecchini, 231–48. Venice: DAEST.

Chen, Q., and A. E. Mynett. 2003. Effects of cell size and configuration in cellular automata based prey–predator modeling. *Simulation Modelling Practice and Theory* 11:609–25.

COAG Reform Council. 2012. *Productivity and agglomeration benefits in Australian capital cities*. Canberra: SGS Economics and Planning.

Couclelis, H. 1985. Cellular worlds: A framework for modeling micro–macro dynamics. *Environment and Planning A* 17 (5): 585–96.

Hagenauer, J., and M. Helbich. 2012. Mining urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. *International Journal of Geographical Information Science* 26 (6): 963–82.

Haklay, M. 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning & Design* 37 (4): 682–703.

He, C., P. Shi, J. Li, J. Chen, Y. Pan, J. Li, Z. Li, and T. Ichinose. 2006. Restoring urbanization process in China in the 1990s by using non-radiance calibrated DMSP/OLS nighttime light imagery and statistical data. *Chinese Science Bulletin* 51 (13): 1614–20.

Henderson, M., E. T. Yeh, P. Gong, C. Elvidge, and K. Baugh. 2003. Validation of urban boundaries derived from global night-time satellite imagery. *International Journal of Remote Sensing* 24 (3): 595–609.

Hu, S., and D. Li. 2004. Vector cellular automata based geographical entity. Paper presented at the 12th International Conference on Geoinformatics, Gavle, Sweden.

Hu, Y., Z. Wu, W. Xiong, and C. Pan. 2008. Study of identifying urban built-up area: Taking Wuhan as an example. *City Planning Review* 32 (4): 88–92.

Jiang, B. 2013. Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer* 65 (3): 482–94.

Jiang, B., and X. Liu. 2012. Scaling of geographic space from the perspective of city and field blocks and using volunteered geographic information. *International Journal of Geographical Information Science* 26 (2): 215–29.

Jiang, B., X. Liu, and J. Tao. 2013. Scaling of geographic space as a universal rule for map generalization. *Annals of the Association of American Geographers* 103 (4): 844–55.

Jiang, B., and J. Yin. 2013. Ht-Index to quantify the fractal or scaling structure of geographic features. *arXiv*:1305.0883.

Krugman, P. 1996. Urban concentration: The role of increasing returns and transport costs. *International Regional Science Review* 19 (1–2): 5–30.

Li, X., Q. Yang, and X. Liu. 2008. Discovering and evaluating urban signatures for simulating compact development using cellular automata. *Landscape and Urban Planning* 86 (2): 177–86.

Li, X., and A. G. O. Yeh. 2000. Modelling sustainable urban development by the integration of constrained cellular automata and GIS. *International Journal of Geographical Information Science* 14 (2): 131–52.

Liu, J., W. Kuang, Z. Zhang, X. Xu, Y. Qin, J. Ning, W. Zhou, et al. 2014. Spatiotemporal characteristics, patterns and causes of land use changes in China since the late 1980s. *Journal of Geographical Science* 69 (1): 3–14.

Liu, J., Z. Zhang, X. Xu, W. Kuang, W. Zhou, S. Zhang, R. Li, C. Yan, D. Yu, S. Wu, and N. Jiang. 2010. Spatial patterns and driving forces of land use change in China during the early 21st century. *Journal of Geographical Science* 20 (4): 483–94.

Liu, X., X. Li, L. Liu, J. He, and B. Ai. 2008. A bottom-up approach to discover transition rules of cellular automata using ant intelligence. *International Journal of Geographical Information Science* 22 (11–12): 1247–69.

Long, Y., and X. Liu. 2013. Automated identification and characterization of parcels (AICP) with OpenStreetMap and points of interest. *arXiv*:1311.6165.

Long, Y., K. Wu, J. Wang, and Z. Shen. 2014. Big models: From Beijing to the whole China. *arXiv*:1406.6417.

Masucci, A. P., K. Stanilov, and M. Batty. 2012. Limited urban growth: London's street network dynamics since the 18th century. *arXiv*:1206.5298.

Ministry of Housing and Urban-Rural Development of the People's Republic of China. 2013. *Chinese city construction statistics yearbook 2012*. Beijing: China Planning Press.

Morrill, R., J. Cromartie, and G. Hart. 1999. Metropolitan, urban, and rural commuting areas: Toward a better depiction of the United States settlement system. *Urban Geography* 20 (8): 727–48.

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015

National Bureau of Statistics. 2013. *China statistical yearbook 2013*. Beijing: China Statistics Press.
Rozenfeld, H. D., D. Rybski, X. Gabaix, and H. A. Makse. 2009. *The area and population of cities: New insights from a different perspective on cities*. Cambridge, MA: National Bureau of Economic Research.
Shen, G. 2002. Fractal dimension and fractal growth of urbanized areas. *International Journal of Geographical Information Science* 16 (5): 419–37.
Stevens, D., and S. Dragicevic. 2007. A GIS-based irregular cellular automata model of land-use change. *Environment and Planning B: Planning & Design* 34 (4): 708–24.
Tannier, C., and I. Thomas. 2013. Defining and characterizing urban boundaries: A fractal analysis of theoretical cities and Belgian cities. *Computers, Environment and Urban Systems* 41:234–48.
Tannier, C., I. Thomas, G. Vuidel, and P. Frankhauser. 2011. A fractal approach to identifying urban boundaries. *Geographical Analysis* 43 (2): 211–27.
Vliet, J. V., R. White, and S. Dragicevic. 2009. Modeling urban growth using a variable grid cellular automaton. *Computers Environment and Urban Systems* 33 (1): 35–43.
White, R., and G. Engelen. 1993. Cellular automata and fractal urban form: A cellular modelling approach to the evolution of urban land use patterns. *Environment and Planning A* 25 (8): 1175–99.
Wu, F. 2002. Calibration of stochastic cellular automata: The application to rural-urban land conversions. *International Journal of Geographical Information Science* 16 (8): 795–818.
Wu, K., Y. Long, Q. Mao, and X. Liu. 2015. Mushrooming Jiedaos, growing cities: An alternative perspective on urbanizing China. *Environment and Planning A* 47 (1): 1–2.
Yang, Y., C. He, Q. Zhang, L. Han, and S. Du. 2013. Timely and accurate national-scale mapping of urban land in China using Defense Meteorological Satellite Program's Operational Linescan System nighttime stable light data. *Journal of Applied Remote Sensing* 7 (1): 073535.
Yuan, J., Y. Zheng, and X. Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining*, ed. Q. Yang, D. Aqarwal, and J. Pei, 186–94. New York: Association for Computing Machinery.
Zipf, G. K. 1949. *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.

YING LONG is an interdisciplinary scholar with substantive planning experiences at the Beijing Institution of City Planning, Xicheng 100045, Beijing, China. E-mail: longying1980@gmail.com. In the past few years, his research has focused on quantitative urban studies and their applications in urban planning. Familiar with planning practices in China and versed in the international literature, his research creatively integrates international methods and experiences with local planning practices.

YAO SHEN is a PhD candidate in the Space Syntax Laboratory at the Bartlett School of Architecture, University College London, NW1 2BX, London, UK. E-mail: y.shen.12@ucl.ac.uk. His research interests include geospatial analysis and modeling, data-driven urban design and planning, spatial econometrics, and urban morphology.

XIAOBIN JIN is an Associate Professor in the School of Geographic and Oceanographic Sciences at Nanjing University, Jiangsu Province, China 210023. E-mail: jinxb@nju.edu.cn. His research interests include land use change modeling, ecological effects of land use, and land planning.

Downloaded by [Hong Kong Polytechnic University] at 22:18 28 December 2015