Automated identification and characterization of parcels (AICP) with OpenStreetMap and Points of Interest

(Prepared for the seminar at Martin Centre, 6 Dec 2013)

Dr. Ying Long

Beijing Institute of City Planning

Martin Centre, University of Cambridge

Paper and data of this study

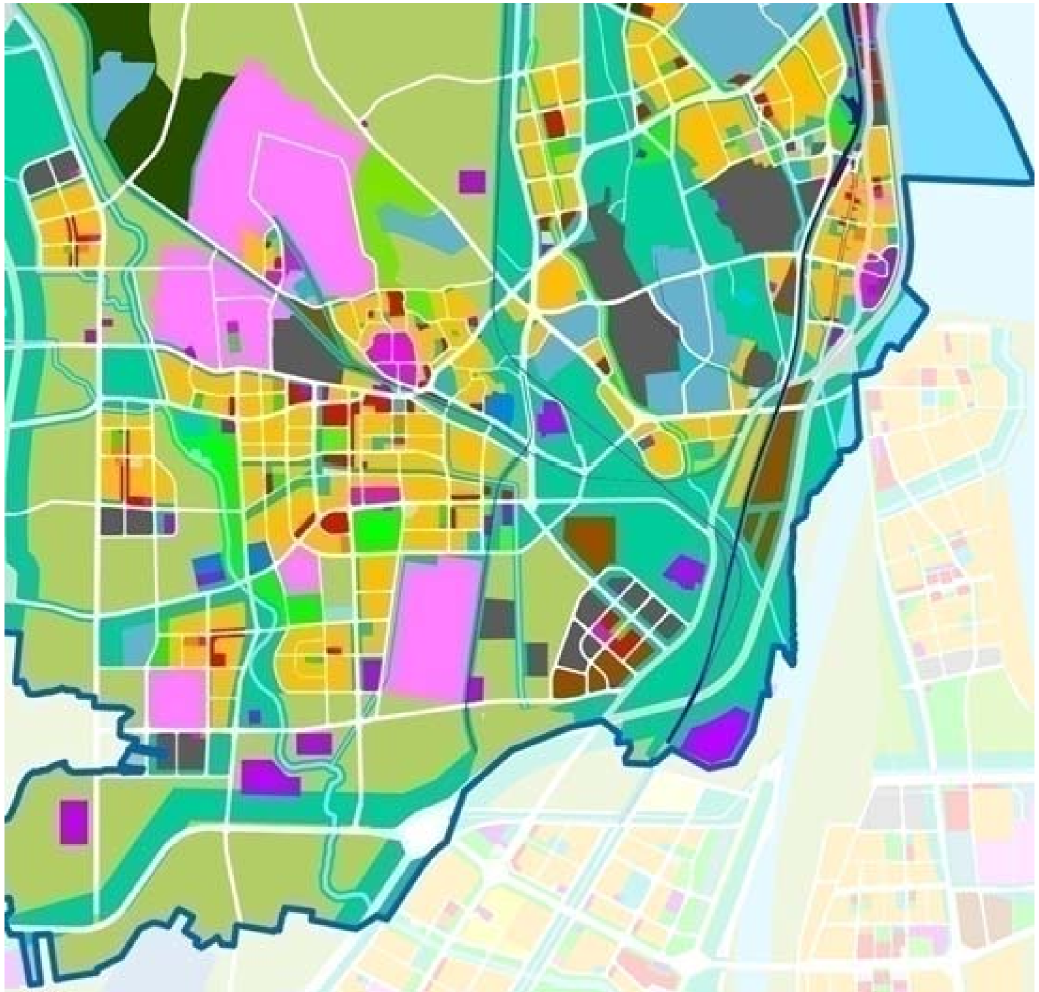
- Paper
  - http://arxiv.org/abs/1311.6165 (submitted to EPB)
- Data download
  - Data 15 at Beijing City Lab http://longy.jimdo.com/data-released/
  - **Over 1k downloads in the first week of data release**
  - Over 100 comments at Sina Weibo, Chinese version Twitter
- Density visualisation:
  - https://a.tiles.mapbox.com/v3/jianghaowang.gcnng3cg/page.html?secure=1#5/36.014/105.996
- Urban function visualisation
  - https://a.tiles.mapbox.com/v3/jianghaowang.ge1lmn67/page.html?secure=1#5/37.78808138412046/106.7431640625
  - OSM China: http://www.openstreetmap.org/#map=9/39.5295/116.8698

Parcel maps are essential for

both planning practices and academic research

- Urban planning and management
  - Spatial plans, zoning, building permits
- Urban studies
  - Urban form and its impact (travel behaviour, energy consumption, health, quality-of-life, etc.)
- Applied urban modeling
  - Vector-based simulation (CA/ABM/Microsimulation)
- The parcel map
  - Geometry, land use type, density

The parcel map in Beijing (partial)



- **Existing parcel map**
- Planned parcel map (based on existing one)

We do not have parcel maps in developing cities!

- Poor developed digital infrastructure
  - Big cities
    - Beijing, one of the most technologically advanced and rapidly developing cities in the erstwhile Third World –dated in 2010 (parcel density limited to six ring road)
  - Medium- and small-sized cities:
    - Not well prepared / digitalized
- Institutional barriers (according to our interview with over 50 professionals)
  - Parcel maps are confidential/classified, and constrained within plan bureaus and official planning institutes like BICP
  - Foreign and private planning agencies, NO
  - Professors and students in universities, **NO**

This condition has limited the progress of quantitative urban studies, urban planning compilation as well as urban management in developing countries in general, and in China in particular.

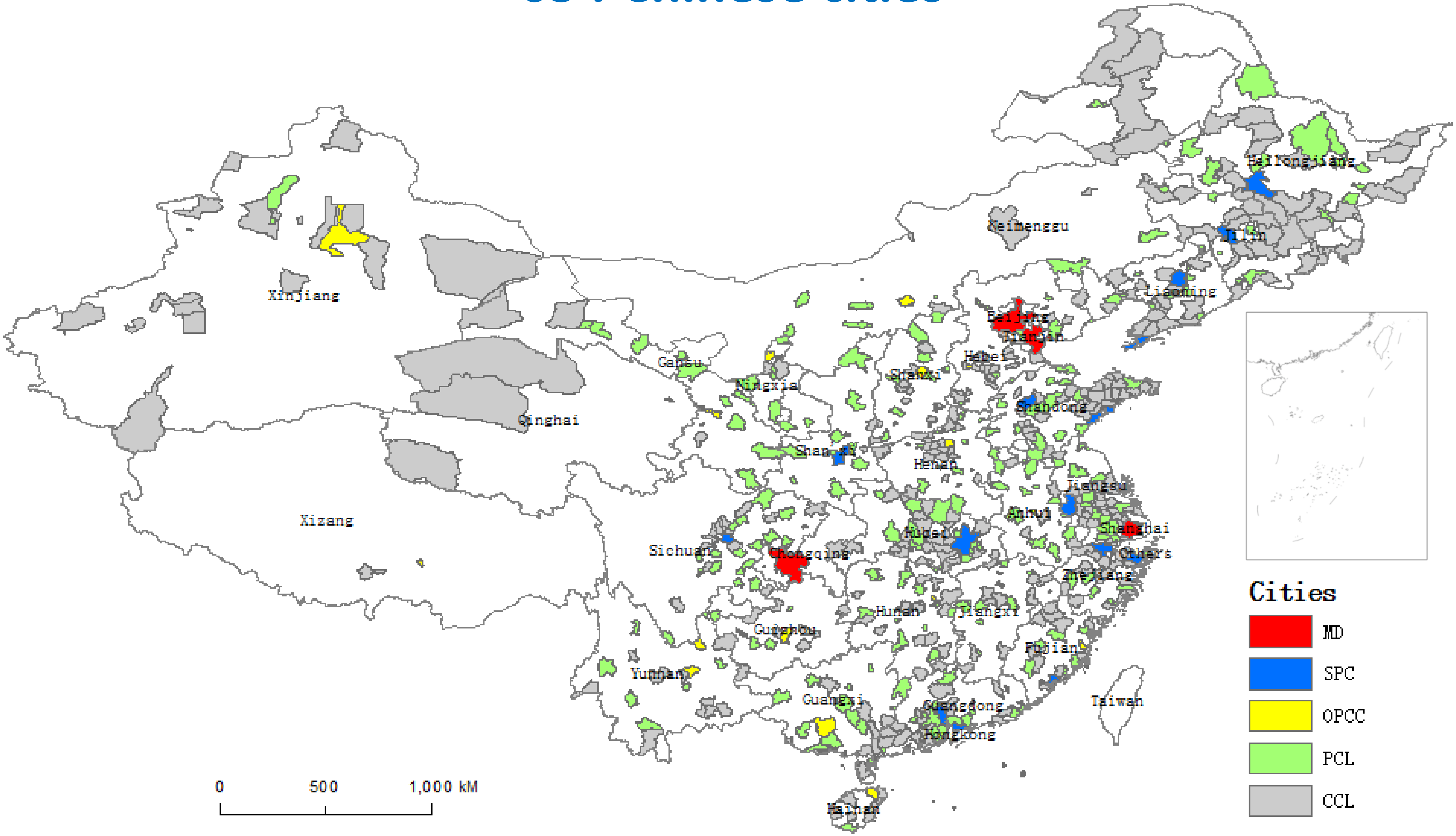
Conventional approaches

- Manual interpretation of remote sensing images, ordnance maps and field surveys
- Time-consuming, expensive, and labor-intensive, thus being not easy for longitudinal update
    1. E.g. it would take an experienced operator 3-5 hours to draw parcel geometries and infer their land use for 35-50 urban parcels in 1 km^2
    2. Many medium and small cities cannot afford in terms of financial and intelligent resources cost
    3. Expensive building data for calculating parcel density
    4. Land use mix never measured in the parcel level
    5. Generally 3-5 years in booming expanding Chinese cities

Our open-data solution in this paper

- A method for automatic identification and characterization of parcels (AICP), based on freely-available Open Street Map (OSM) and crowd-sourced Point-of-Interest (POI) data
    1. Provide quick and robust delineation of land parcels
    2. Select urban parcels from all generated parcels
    3. Infer urban functions, development density and mixed land uses for urban parcels

654 Chinese cities

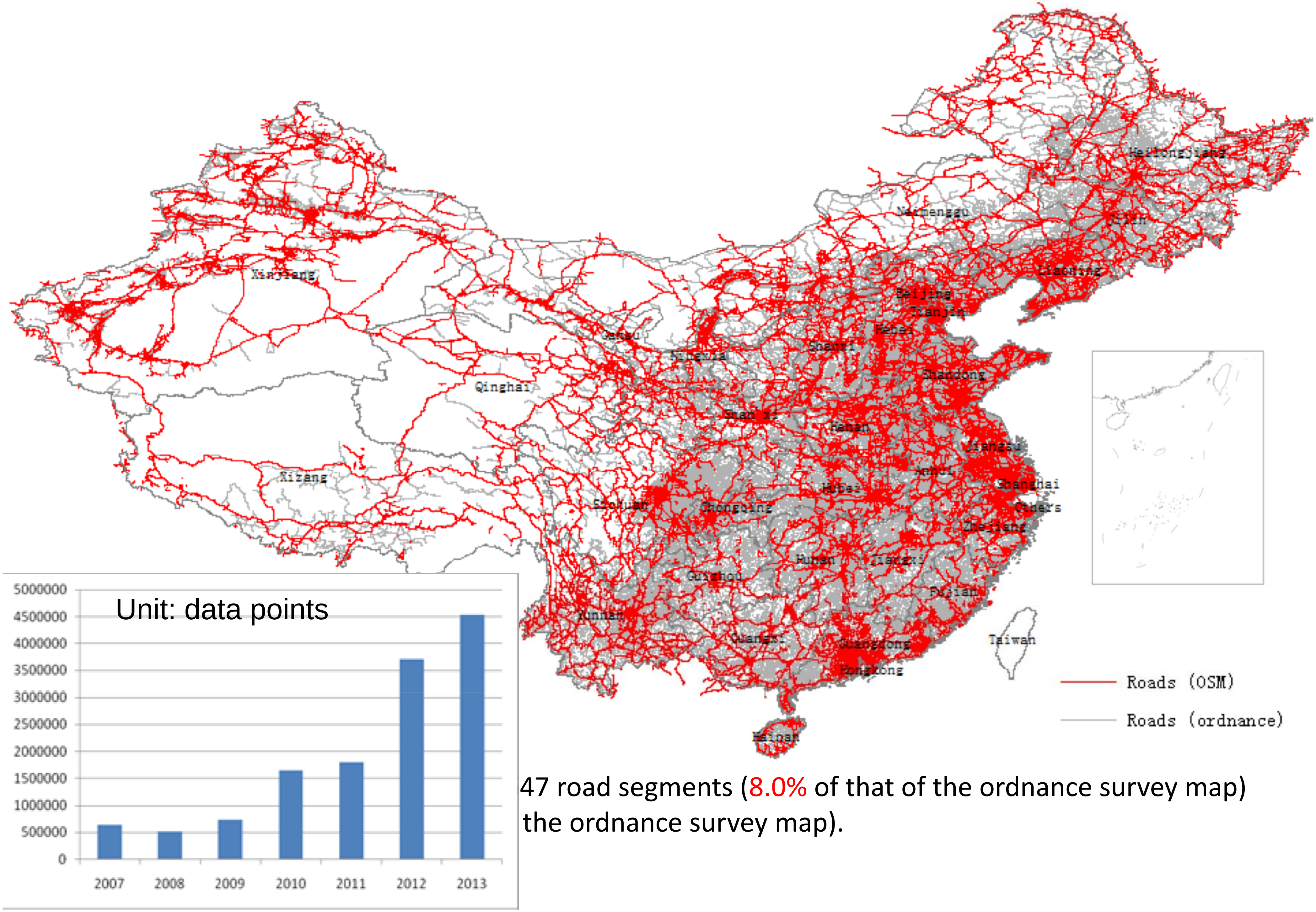


- Five levels of cities in China:
  - municipalities directly under the Central Government (MD, 4 cities), sub-provincial cities (SPC, 15), other provincial capital cities (OPCC, 16), prefecture-level cities (PLC, 251), and county-level cities (CLC, 368)

Data

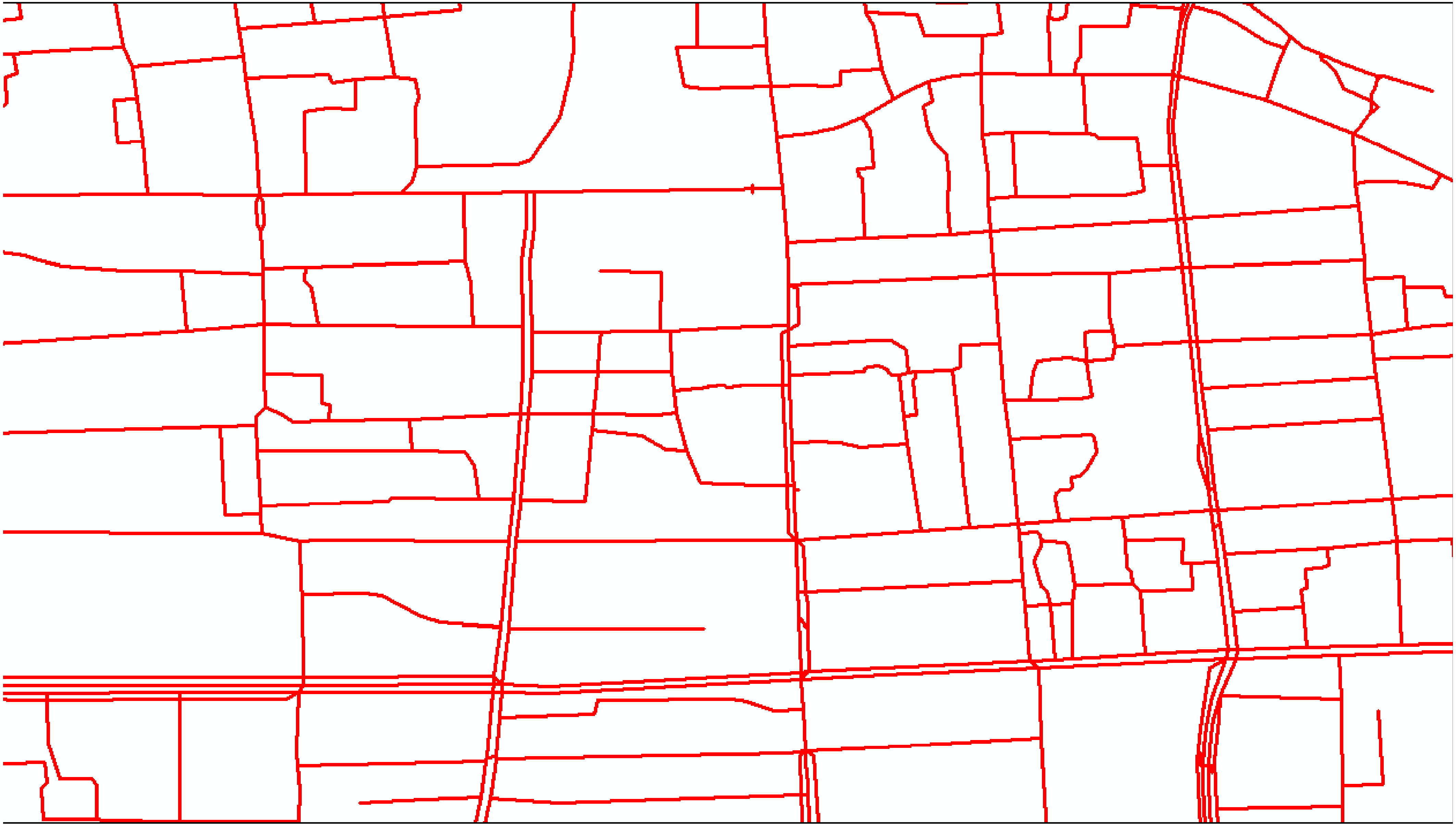
- Total urban land area for each city in 2012 from the *Chinese City Construction Statistics Yearbook 2012*
- 5 m POIs gathered from and geocoded by business cataloging websites
  - 9 categories, including commercial, transport, government, education, residence, green space, etc
- For model validation
  - The ordnance survey map with all detailed road networks in China
  - Urban area of China interpreted from
    - DMSP/OLS (1-km spatial resolution, night light images; Yang et al. 2013)
    - GLOBCOVER (300-m spatial resolution; Bontemps 2009)
  - Manually generated parcel map for Beijing in 2010 gathered from BICP

Increasing OSM in China



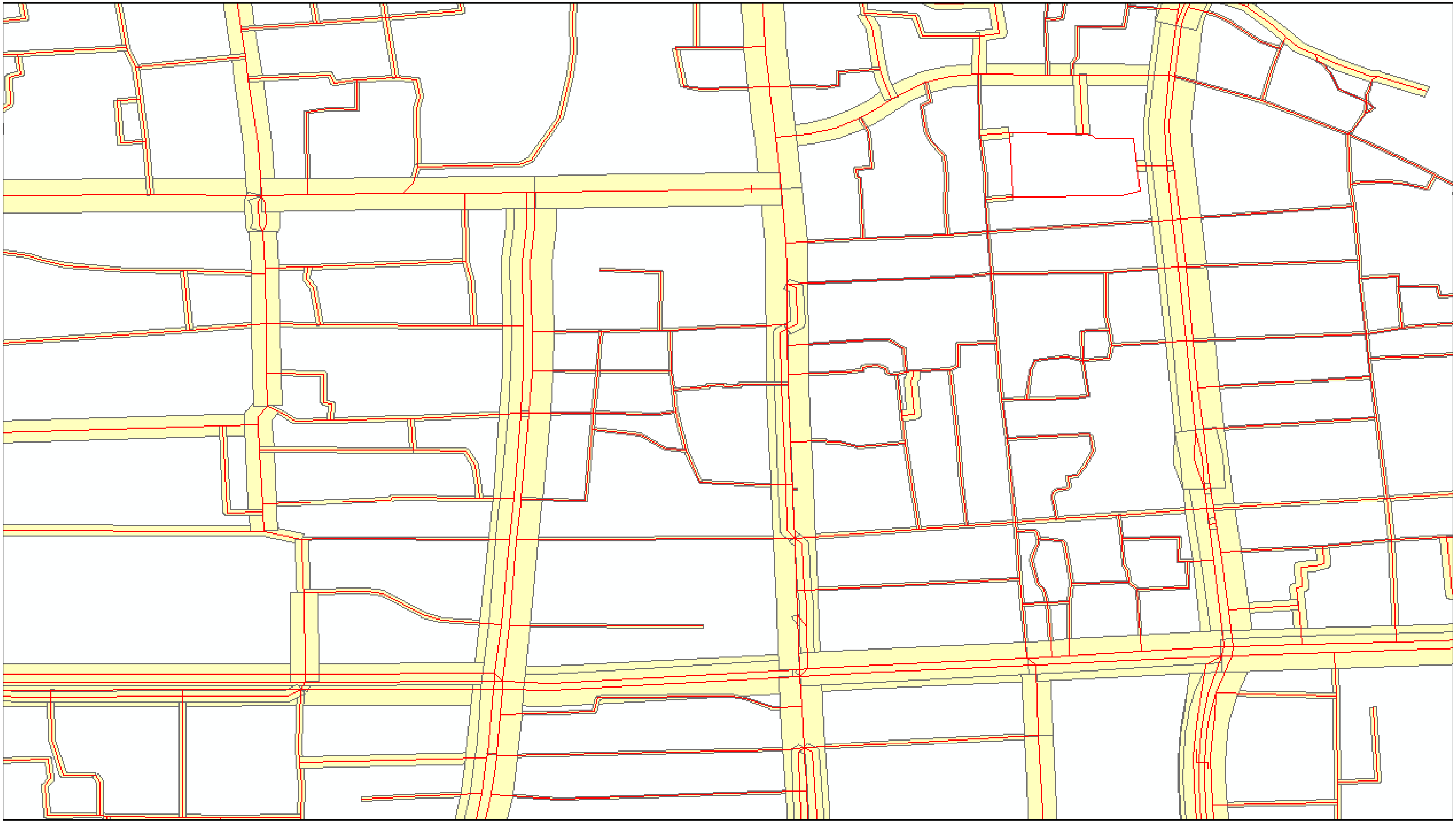
47 road segments (8.0% of that of the ordnance survey map) the ordnance survey map).

Delineating parcel boundaries



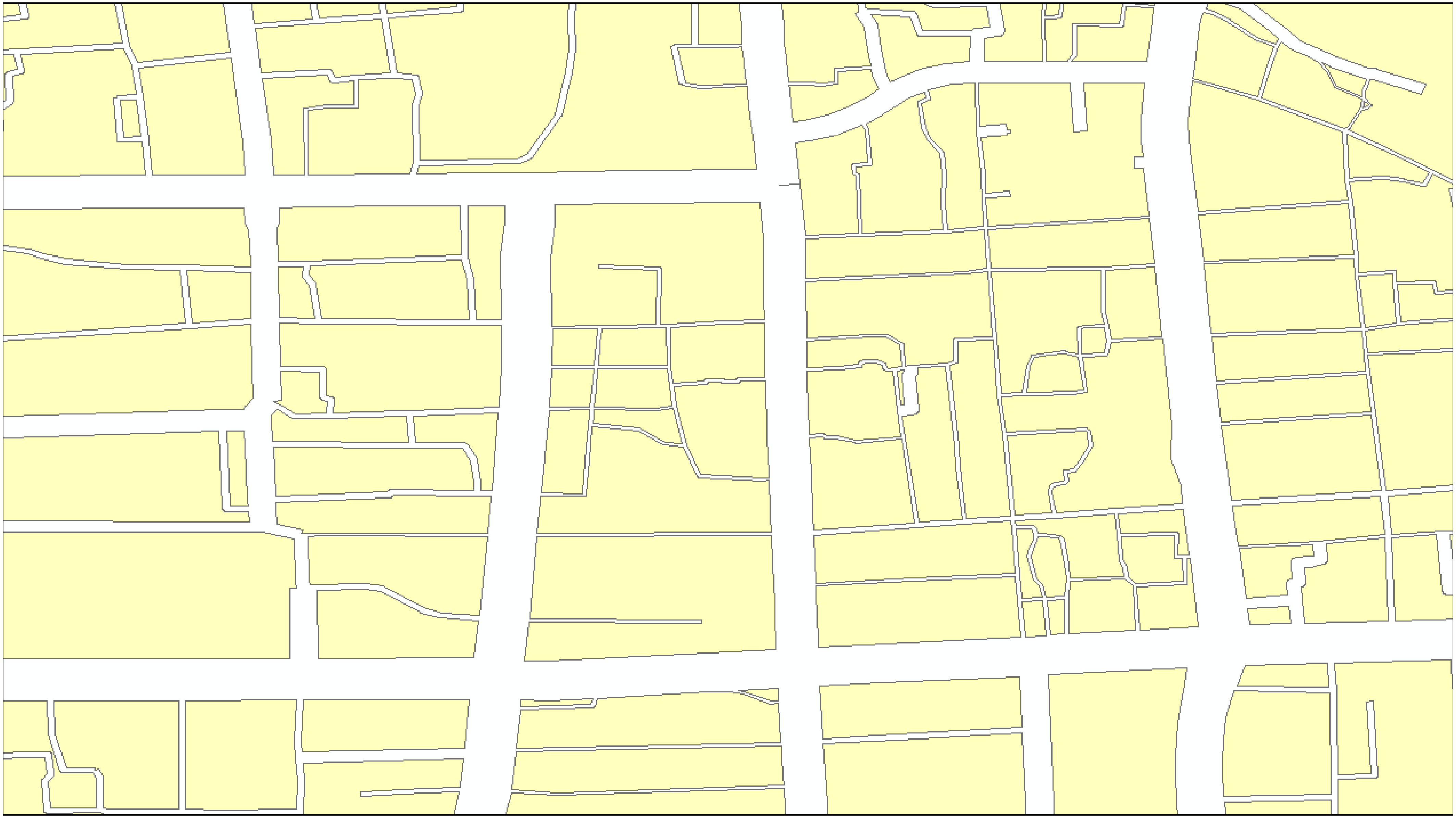
- Raw OSM roads
- Various of road types: primary, secondary, footpath, etc

Delineating parcel boundaries



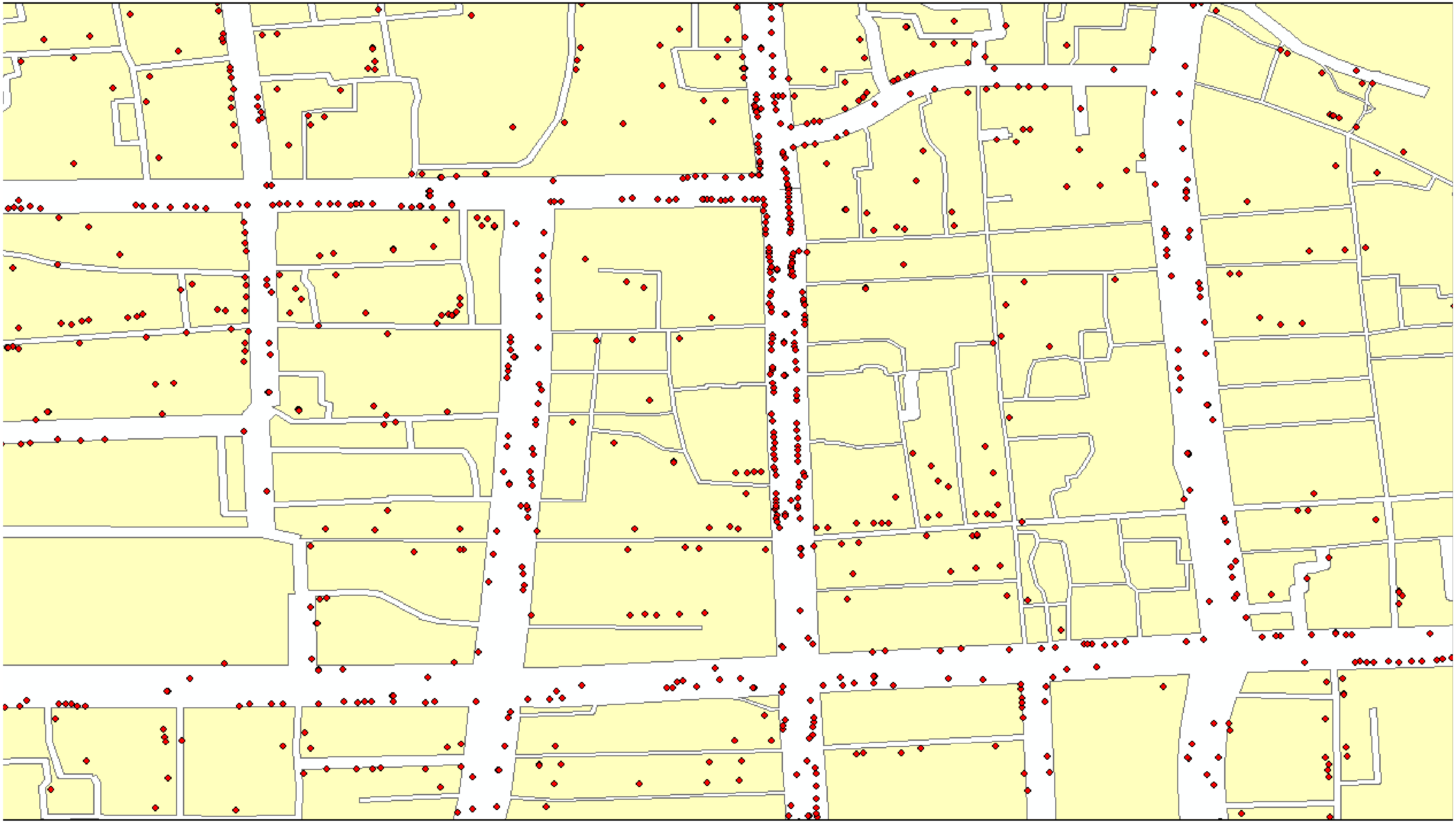
- Buffer OSM roads
- Buffer width varies from road types (2 - 30 m)

Delineating parcel boundaries



- Erase road space from the study area
- Road space retained as the land use "Transport"

Calculating density for all parcels

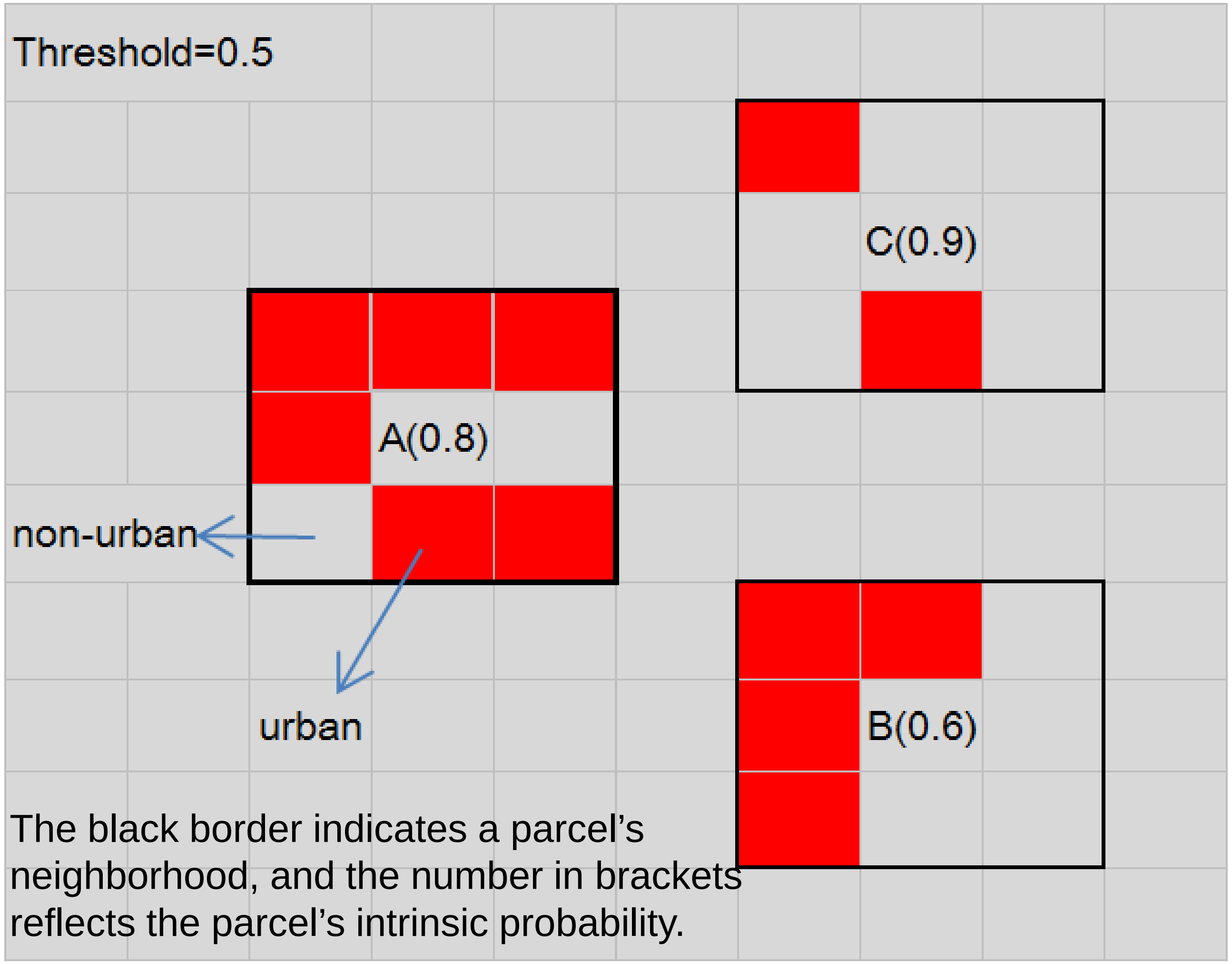


- Density = (The counts of POIs in/close to a parcel) / (The parcel area)
- Other measures (e.g., online check-ins and floor area ratio) can substitute POIs and approximate the intensity of human activities

Calculating density for all parcels

- Density = (The counts of POIs in/close to a parcel) / (The parcel area)
- Other measures (e.g., online check-ins and floor area ratio) can substitute POIs and approximate the intensity of human activities

Selecting urban parcels using vector cellular automata



$$P_{ij}^{t} = (P_l)_{ij} \times (P_\Omega)_{ij}$$

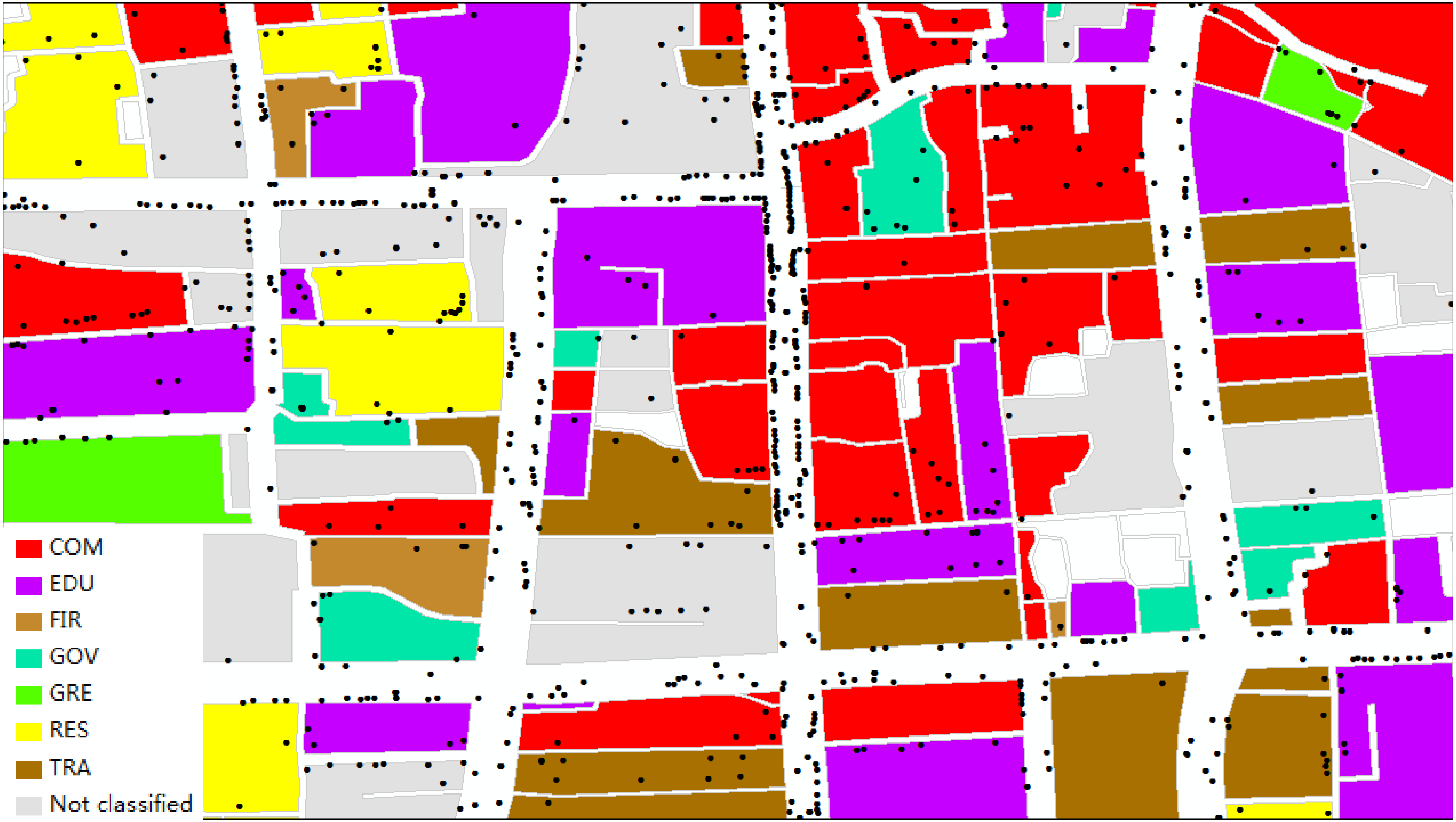
$$(P_l)_{ij} = \frac{1}{1+\exp[-(a_0 + \sum_{k=1}^{m} a_k c_k)]}$$

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

$$S_{ij}^{t+1} = \begin{cases} Urban \text{ for } P_{ij}^{t} \succ P_{thd} \\ NonUrban \text{ for } P_{ij}^{t} \le P_{thd} \end{cases}$$

- We developed one vector cellular automata model for each city, to allocate the urban area total in the yearbook into parcels.
- Neighborhood configuration: 500 m radius of each parcel
- Constraints: size, compactness, and POIs density (parameters calibrated using the BICP parcels)

Inferring dominating urban function for urban parcels

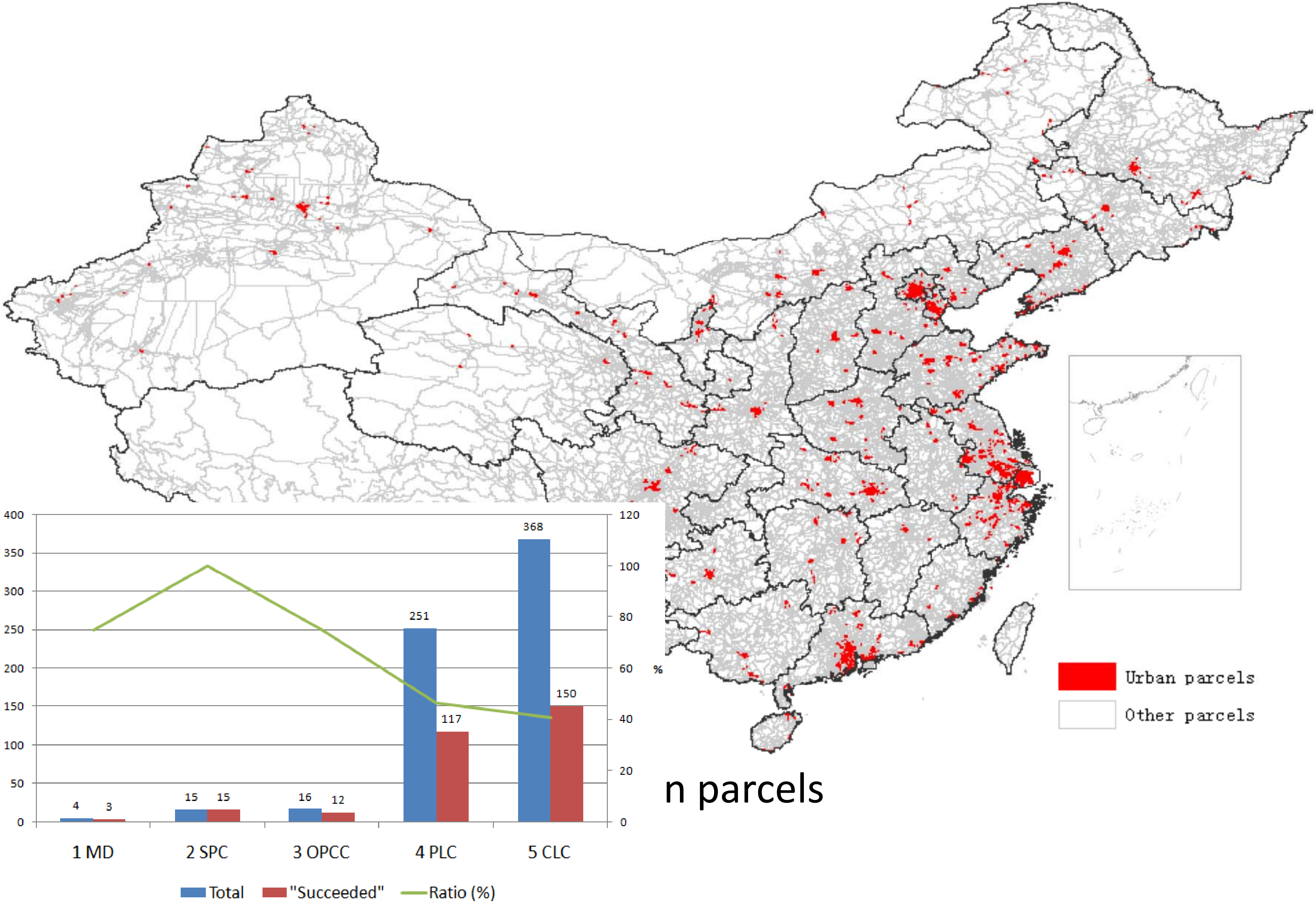


- A dominant POI type within a parcel is defined as the POI type that has accounted for more than 50% of all POIs within the parcel.
  - For example, if 31 out of 60 POIs within a parcel are labeled as “business establishment”, the urban function for that parcel will be assigned as “business”.

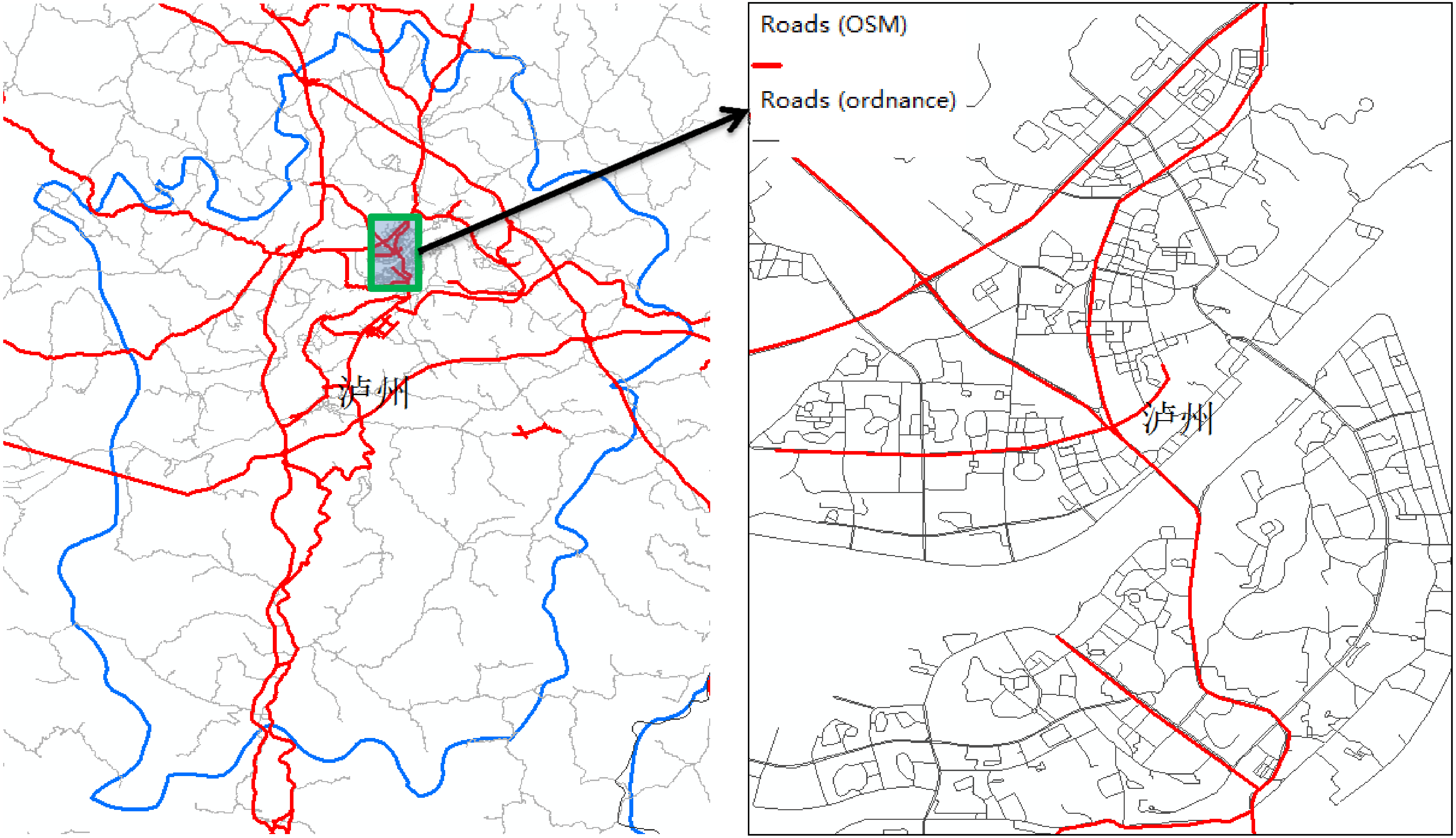
Inferring land use mix for urban parcels

- The mixed index (M) of a land parcel is calculated as M =-sum(pi *ln pi) (i = 1, ..., n)
  - where n denotes the number of POI types, and pi is the proportion of POI type i among all POIs in the parcel.

All generated parcels and urban parcels China



A case of “failed” city

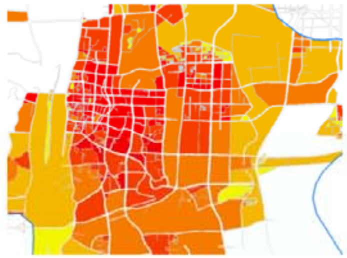
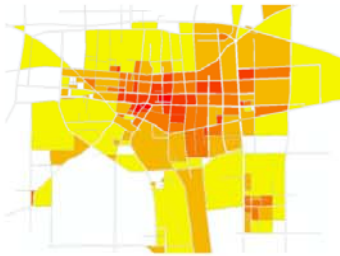
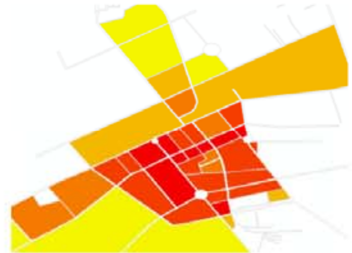
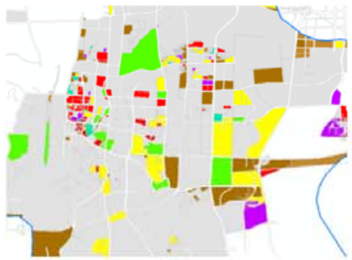


- 297 cities with ten or more urban parcels as “successfully” processed by our algorithm.
- Maximum size of an urban parcel is set as 10 km^2 according to the parcels in BICP (95%)

A profile of the results

- 232,145 parcels identified for 297 cities, and 82,645 labeled as “urban”(total urban area 25,905 km^2).
- The average number of urban parcels for MD, SPC, OPCC, PLC and CLC cities are 1411, 407, 199, 79 and 26, respectively.
- 55,728 (67.3%) urban parcels have “dominant” urban functions, including 16,018 residential parcels, 16,381 commercial parcels, 18,351 firm parcels, and 10,018 government parcels.
- The average land mix degree for all urban parcels in 297 cities is approximately 0.66

The profile of typical cities

	Beijing	Nanjing	Changsha	Weifang	Gongzhuling
City level	1 MD	2 SPC	3 OPCC	4 PLC	5 CLC
Density Low High					
Function COM EDU FIR GOV GRE RES TRA Not classified					
Land use mix Low High					

Validation at the parcel level (limited to Beijing)

Comparison of selected urban parcels in BICP and OSM in Beijing (R=ring road)

Parcels	Parcel count	Average size (ha)	Overlapped with BICP	Spatial distribution (in terms of area, km^2)					
				Within R2	R2-R3	R3-R4	R4-R5	R5-R6	Beyond R6
OSM	7,130	17.2	1194.2 km2 (71.2%)	42.5	74.0	113.4	263.5	666.5	519.9
BICP	57,818	2.9	-	48.6	69.7	99.8	229.5	687.9	544.4
OSM/BICP	0.12	5.93	-	0.87	1.06	1.14	1.15	0.97	0.95

- 71.2% area intersected
- Spatial pattern (see the above table)
- Size: both log-normal distribution with similar mean value
- Density: the correlation coefficient = 0.858 between density inferred by POIs and calculated by building floor space
- Function: 56.3% residential parcels by OSM appear in BICP
- Mix, not validated due to no data
- It is worth noting that these online visualizations serve as crowd-source validations for our method.

Validation at the regional level

- Urban parcels in ORDNANCE were generated and selected using the same parcel generation and selection methods like “OSM”.
- Among 1,184,524 parcels generated, we successfully selected 350,102 urban parcels in 627 cities in ORDNANCE.

Table 4 The comparison of urban parcels in OSM and ORDNANCE for 297 cities

Data	Urban area (km^2)	Parcel count	Average parcel/patch size (ha)	Intersected with ORDNANCE (km^2)
OSM	25,905	82,645	31.3	15,053
ORDNANCE	25,670	260,098	10.0	-

The match degree between urban land by OSM and ORDNANCE was 58.1%, calculated as the ratio of the area of overlapping urban parcels to the area of all OSM-based urban parcels.

The ratio for MC, SPC and OPCC was around 70% and the ratio for FLC and CLC was around 45%.

Our contributions

1. Propose a robust and straightforward approach to delineating parcels, identifying urban parcels, and characterizing parcel features using open data
2. Incorporate a vector-based cellular automata model into the identification of urban parcels.
3. Applied to hundreds of cities in China, and could possibly be extended to generate parcel data for other developing countries.

Potential applications

1. Urban planning and studies in places where digital infrastructure development is weak
2. Inter-urban study based on inner-city datasets
   – Quality of life
3. As spatial units to incorporate other ubiquitous and spatially referenced (big) data
   – "Big" parcel in the "big data" era
4. Vector-based urban modeling
   – Simulating urban expansion in the parcel level for all Chinese cities using a mega-vector-parcels cellular automata model (MVP-CA)
   – Each city, big or small, would have an urban expansion model for **decision making / planning support** after this study

Potential bias

1. OSM data quality is not sound enough for generating parcels in medium and small cities in China
   – Hope time could solve it, thanks for increasing contributors
2. Use POIs as a proxy of urban density
   – To enrich by online check-ins, taxi trajectories, and public transport smartcard records (in progressing)
3. Over-large urban parcels
   – Parcel subdivision techniques (Aliaga et al. 2008)

Thank you for your attention!

Open data + Open data

= Data releasable

Open Data Initiative

BCL

Beijing City Lab