



# LBS大数据和城市功能及国家扶贫效率

---

姚尧 博士，副教授，高级工程师

地理与信息工程学院，地图制图学与地理信息工程

东京大学，空间情报科学研究中心

阿里巴巴集团，达摩院，访问学者

Email: [yaoy@cug.edu.cn](mailto:yaoy@cug.edu.cn)

办公地点：未来城校区地信楼522办公室





# 主要内容



- 1 LBS大数据简介
- 2 LBS大数据关键技术
- 3 LBS大数据与城市功能
- 4 LBS大数据与贫困



# 主要内容



- 1 LBS大数据简介
- 2 LBS大数据关键技术
- 3 LBS大数据与城市功能
- 4 LBS大数据与贫困

# 01 | LBS大数据简介



怎样去估计/模拟/计算/分析数十亿人的流动?



# 01 | LBS大数据简介



LBS (Location Based Service) 数据：是利用各类型的**定位技术**来获取定位设备当前的所在位置，通过**移动互联网**收集定位设备所获取的数据。



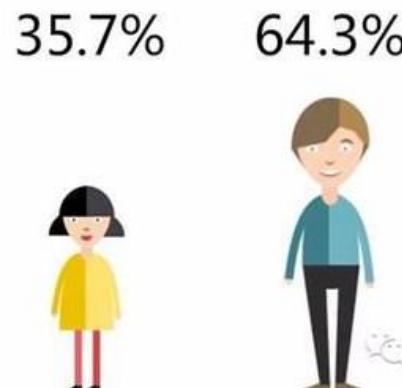
# 01 | LBS大数据简介



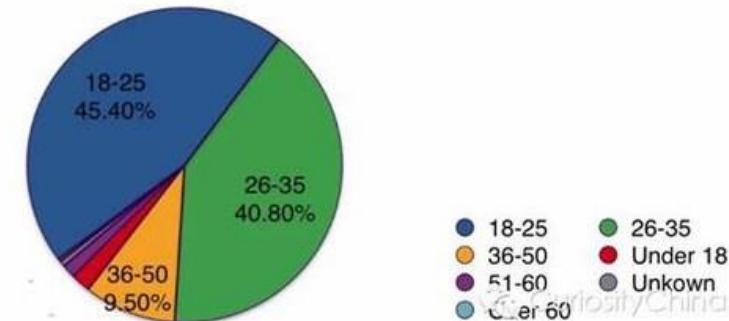
## ● 腾讯公司 Tencent

- ✓ 中国最大的互联网公司之一；
- ✓ 2015年1季度用户8.08亿，其中手机客户端用户超过6亿（微信用户5.49亿）；
- ✓ 旗下手机客户端：手Q，微信，腾讯地图，手机QQ空间，腾讯邮件等。

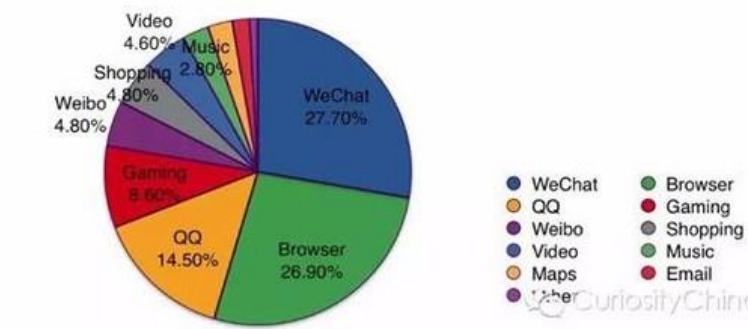
Gender



WeChat users by age



Mobile data consumption by activity



# 01 | LBS大数据处理



## ● 腾讯实时用户分布数据 Real-Time Tencent User Density (RTUD)

✓ 数据提供：宜出行（微信公共服务） <http://easygo.qq.com>

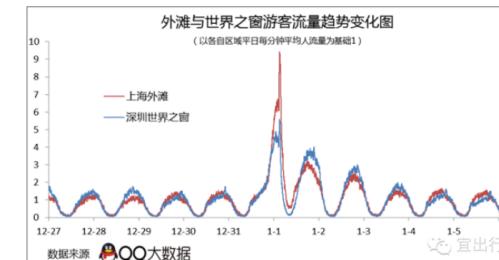
✓ 时间分辨率1 hour, 空间分辨率 $0.00025^\circ$  (约26.5米)



- 最近一周，一般10点人多
- 一般周一多，周五少



1、晚上游客暴增，达平时8倍  
12月31日18: 00-1月1日03: 00，上海外滩游客是30日同时段的8.3倍(总量上，锁定区域人流量估测达10万人，整个外滩加南京路总计30万人)；相对地，同样有跨年活动的深圳世界之窗，该比例为4.2；而同样在外滩，2014年国庆期间，该比例为3.8倍；可以看出，上海外滩跨年夜当晚“超载”程度的确要严重更多。



注: 用户地理信息来自QQ“附近的人”等lbs服务，通过数学模型拟合“附近的人”业务用户数与全量用户的方程式，估算该地点的全量用户数。

2、地理分布不均匀，局部密度高  
下左图和右图分别是上海外滩与深圳世界之窗在12月31日18:00至1月1日3:00的人流密



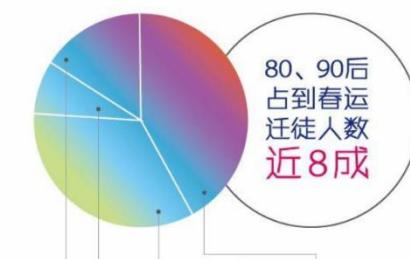
春节去哪儿(2014)

2015-03-19 宜出行



2014.1.16 - 2.16  
16.52亿人次  
加入春运的人潮人海中

人次统计方法：小明从A地到C地，如果在中间B地停留超过1天，则算迁徙2次。



# 01 | LBS大数据处理



## 宜出行8天(192h)数据网络API:

<http://easygo.qq.com/neighborship?>

openid=wx519ab479708c8cf1&gps\_type=1&latitude=22.5659533&longitude=114.2517972&time=1435145717

openid: 用户唯一标识

gps\_type: gps坐标类型, 1为GCJ02坐标系

latitude: 纬度

longitude: 经度

time: unix时间戳 (从1970年1月1日零时起的秒数)

## 腾讯地图中心网格和GCJ02坐标 (经纬度) 换算方法:

$$\text{longitude} = (\text{grid\_x} \cdot 250 + 125) \times 0.000001$$

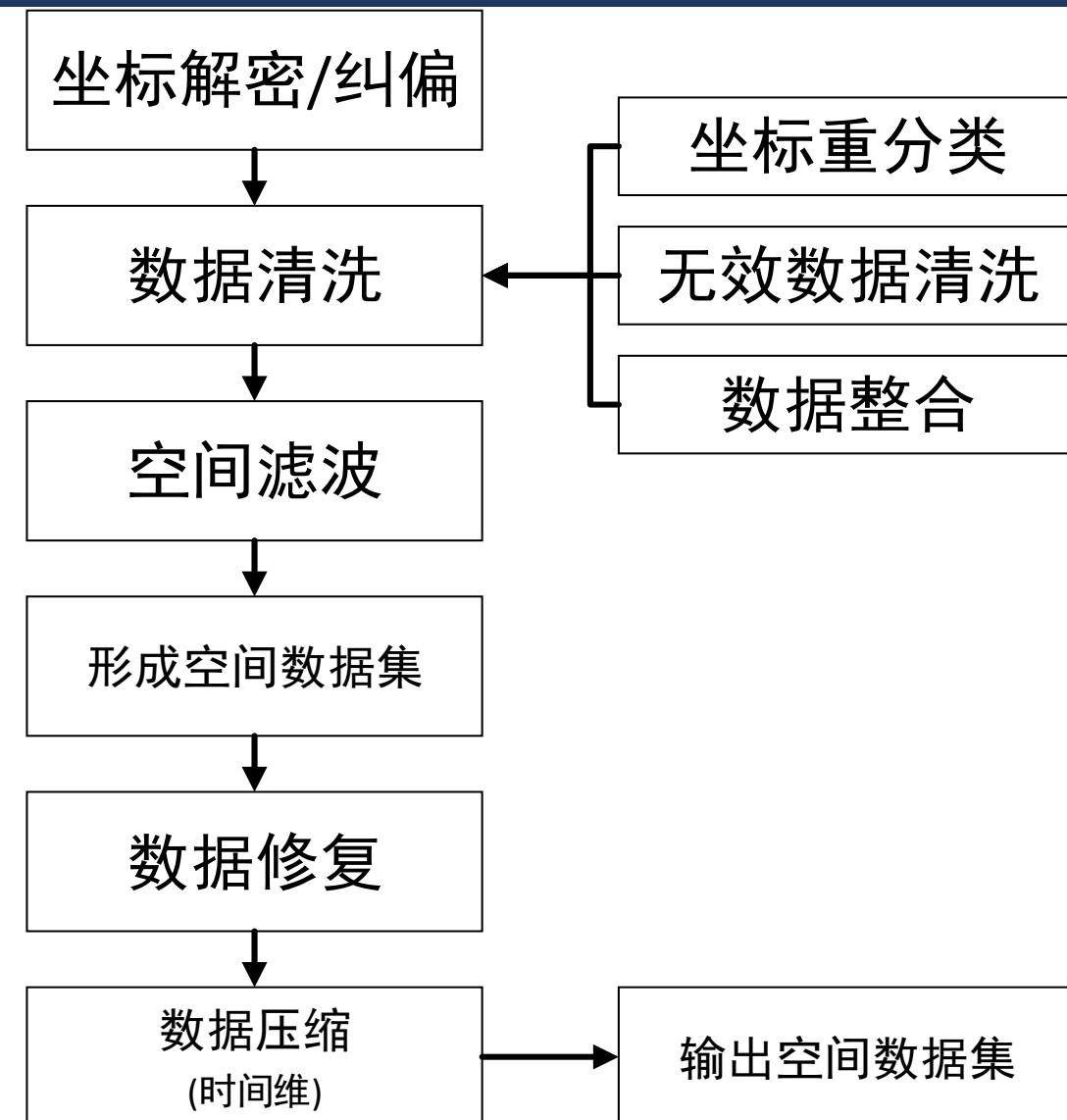
$$\text{latitude} = (\text{grid\_y} \cdot 250 + 125) \times 0.000001$$

The screenshot shows a Notepad++ window displaying a large text file named 'guangzhou0623'. The file contains approximately 2754 lines of data, each consisting of a single number followed by a series of commas and digits. The numbers range from 113.606875 to 113.606875, with some variations in the decimal places. The data appears to be a list of grid identifiers or coordinates.

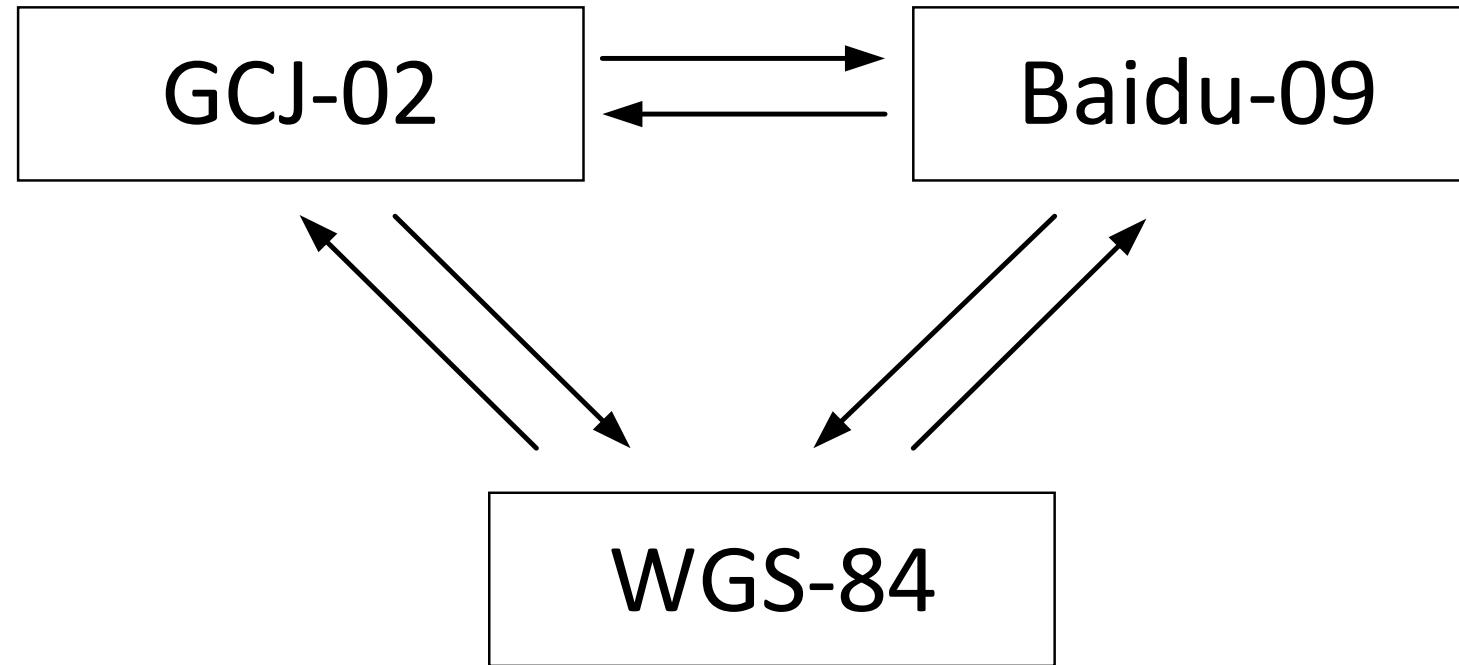
源码: [https://172.22.30.144/svn/haplib/trunk/haplib\\_vs2010/TencentEasyGoGrab](https://172.22.30.144/svn/haplib/trunk/haplib_vs2010/TencentEasyGoGrab)

- ✓ 数据坐标被测绘部门加密，存在500米~5km的**系统误差**；
- ✓ 因为腾讯API限制的问题，数据是分区域抓取，对于大型城市没有一个统一的时间段完整数据（**时空数据缺失**）；
- ✓ 因为网络抓取和服务器的问题，数据丢包和连接重置导致数据存在缺失（**时空数据缺失**）；
- ✓ （网络大数据普遍存在的问题）**文件破碎、数据冗余、数据量巨大、预处理过程复杂**等。

# 01 | LBS大数据处理



# 01 | LBS大数据处理



**GCJ-02坐标系：**国家测绘局加密坐标系，强制对外公开的WGS84坐标加入误差，俗称“火星坐标系”。常见于非百度系产品，如腾讯数据、高德地图、谷歌数据等内置非百度地图服务的网站和App；

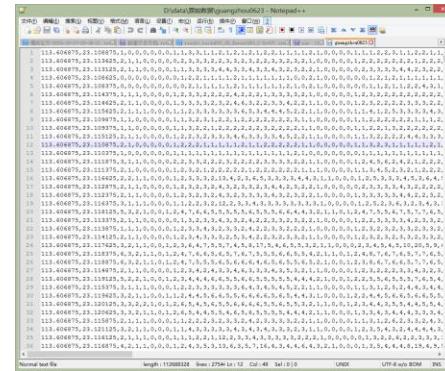
**Baidu-09坐标系：**在GCJ-02的基础上二次加密的坐标系。常见于百度系产品，如百度地图、搜房网等。

坐标纠偏模型非常复杂。现在我们实现的纠偏模型是基于大地测量模型和二分法可以得到近似结果，广州地区纠偏误差可由500米降低到5米，全国误差 $\leq 10m$ 。

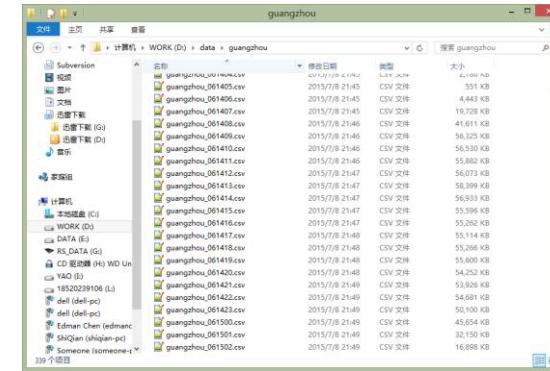
纠偏模型SVN下载地址：

[https://172.22.30.144/svn/hplib/trunk/hplib\\_vs2010/GCJ02\\_WGS84\\_Transform](https://172.22.30.144/svn/hplib/trunk/hplib_vs2010/GCJ02_WGS84_Transform)

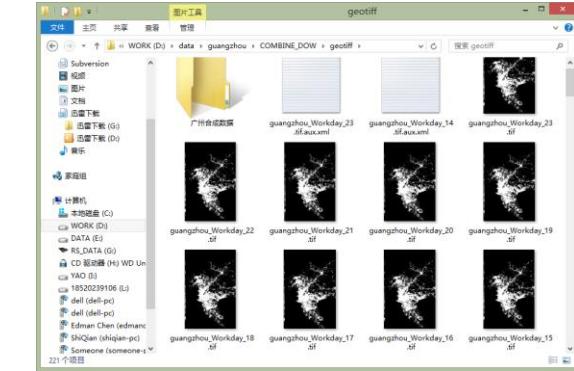
1. 将所有数据纠偏后，以时间（以小时为单位）基准重新归类到新的文件；
2. 在归类过程中，将出现错误或不全的数据条给删除（时序数据总和为0或出现无效值nan/负值），以达到数据清洗和节省存储空间的目的；
3. 基于GDAL和C++，构建坐标和投影系统，将点数据文件**形成栅格/矢量数据集**。遍历数据，每个坐标位置数值累加，并统计空间中每个位置数据出现的频率，进行**均值滤波**；
4. 此时形成了**初步的栅格/矢量空间数据集**。



Step 1  
&  
Step 2



Step 3  
&  
Step 4

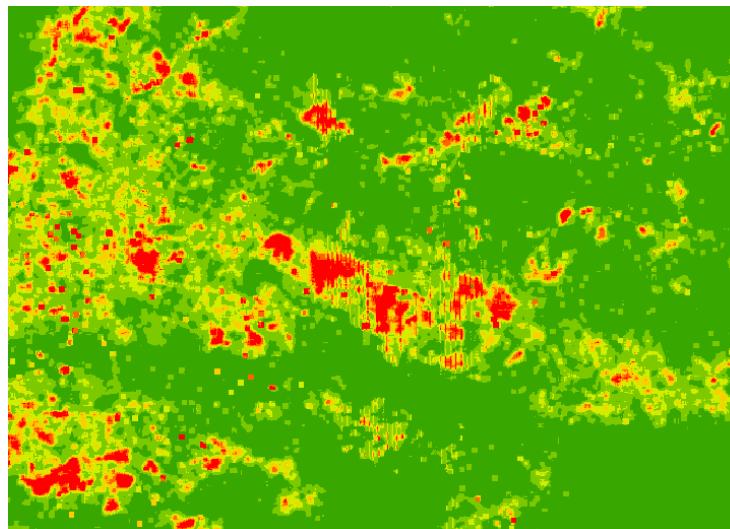


# 01 | LBS大数据处理

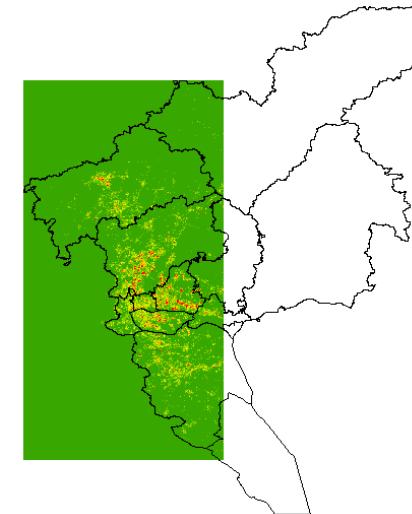


时空缺失：因为API权限限制或网络问题导致时间和空间有效数据缺失

- 寻找到对应空间位置的时序腾讯用户密度数据值进行插值（时间序列插值）；
- 寻找到对应的工作日/休息日数据插值（优先同一个周历day of week）；
- 寻找缺失数据区域周围 $3*3$ 或 $5*5$ 邻域有效数据均值滤波（空间插值法）；
- 如果以上都无法解决（比如数据在某个时间点大面积数据缺失），则在数据压缩步骤解决。



小范围时空缺失



大面积数据缺失

# 01 | LBS大数据处理



初级空间数据集问题：数据量大、维度非常高、数据冗余、不同的数据覆盖的时空范围不一致

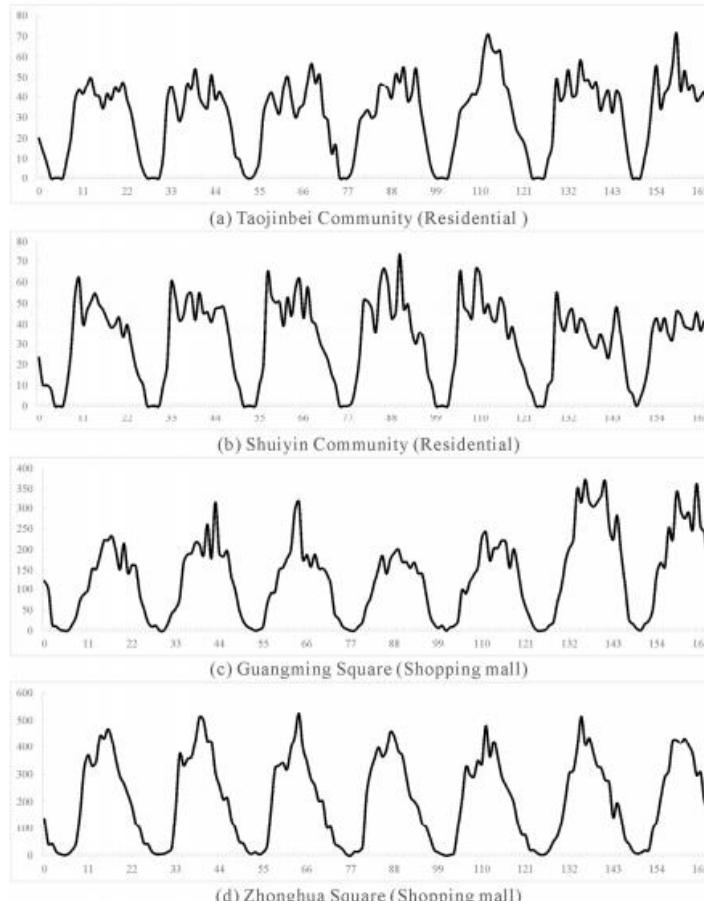


Figure 3. Temporal changes of the Tencent user density (y axis) by hour (x axis) in the four building samples.

164x207mm (300 x 300 DPI)

1. 时序用户分布数据具备周期性；
2. 工作日的曲线波形相似，休息日的曲线波形也相似；但工作日和休息日的曲线波形不一致。
3. 通过matlab的FFT分析形成周期-功率谱图，我们计算得到工作日的固定周期是 $T = 24$ ，休息日的曲线波动固定周期 $T = 24$ 。

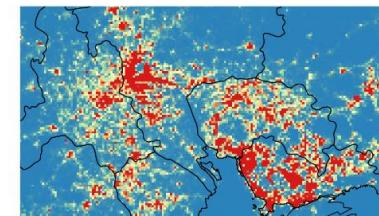
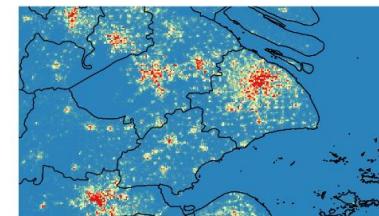
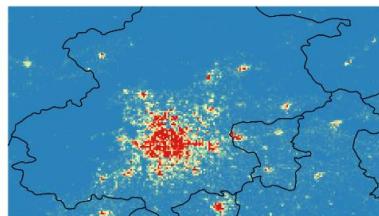
## 数据压缩方法：

- ① 将数据分为工作日（周一~周五，不包含节假日），休息日（周六、周日、节假日）；
- ② 将工作日（或休息日）相同时间、相同位置的数据进行求均值处理（均值滤波）。

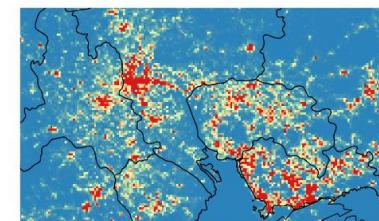
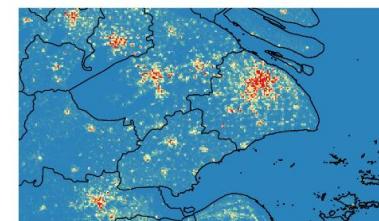
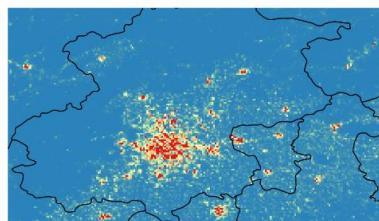
# 01 | LBS大数据处理



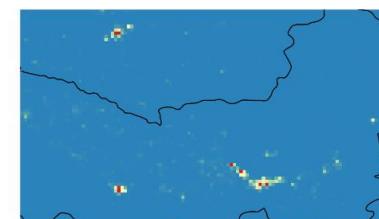
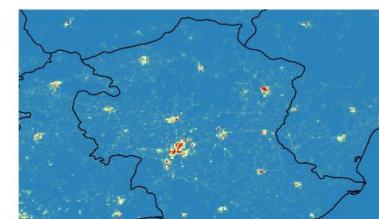
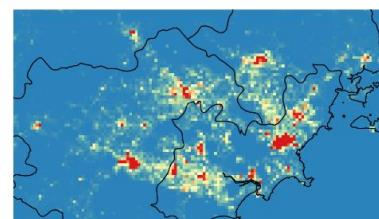
Working Days



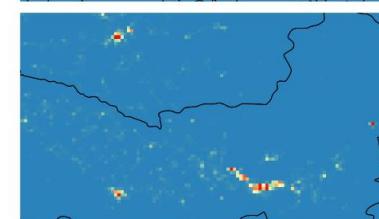
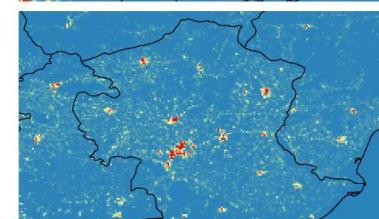
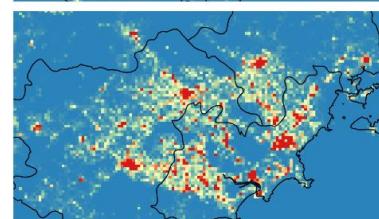
Spring Festival



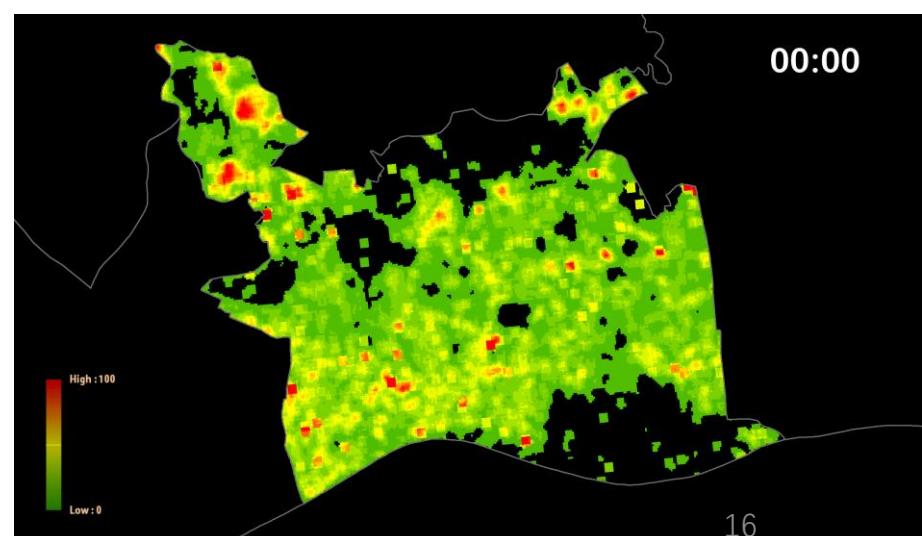
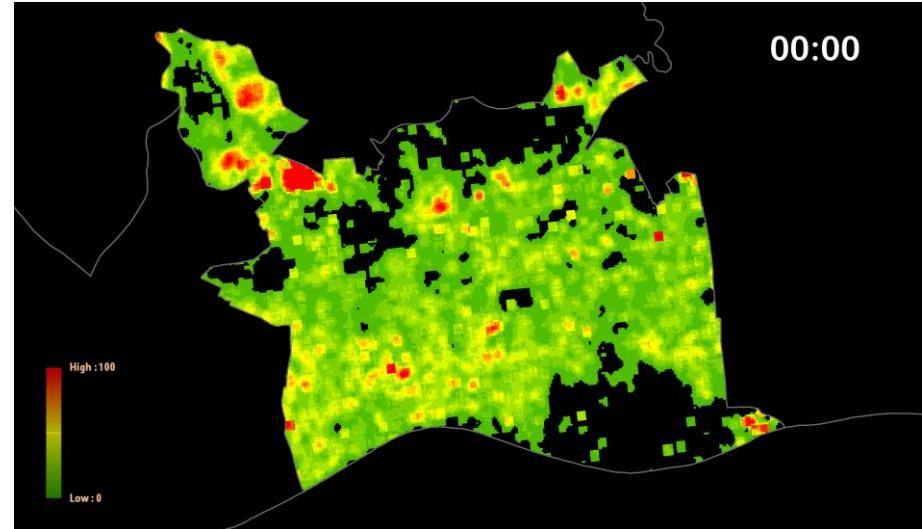
Working Days



Spring Festival



0 4000  
Low : 0 High : 100



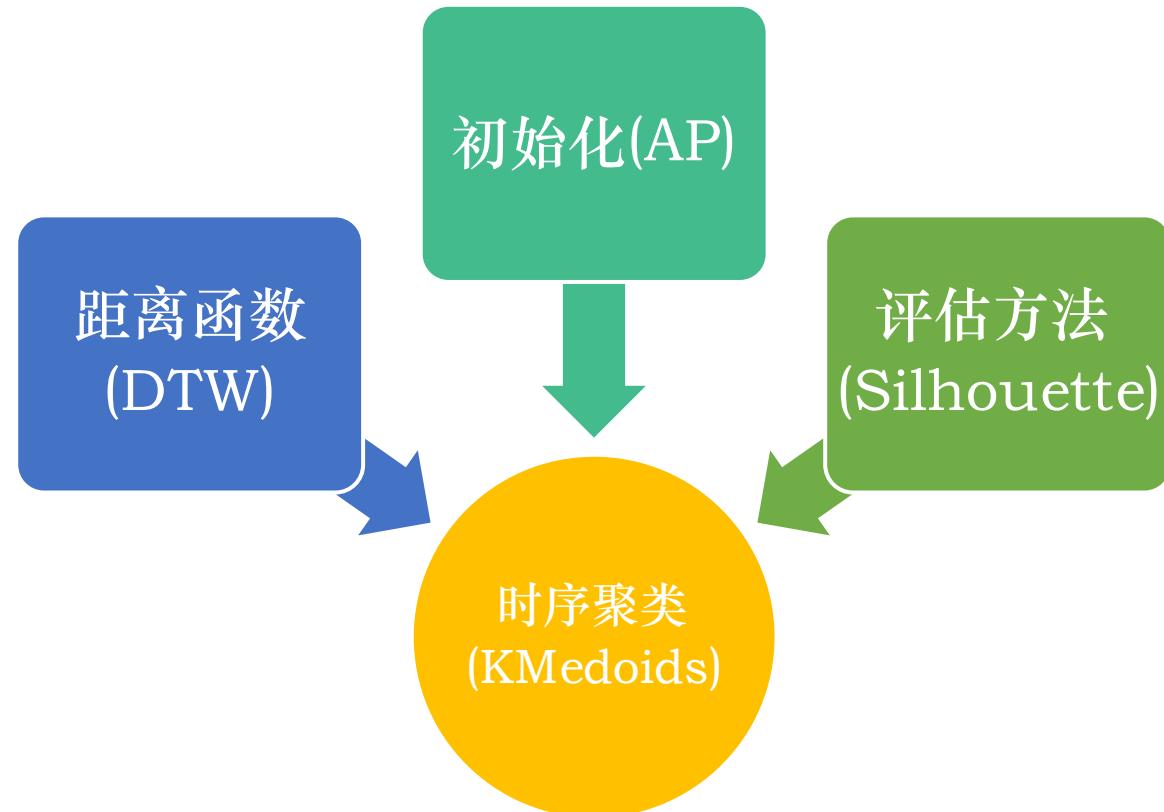


# 主要内容



- 1 LBS大数据简介
- 2 LBS大数据关键技术
- 3 LBS大数据与城市功能
- 4 LBS大数据与贫困

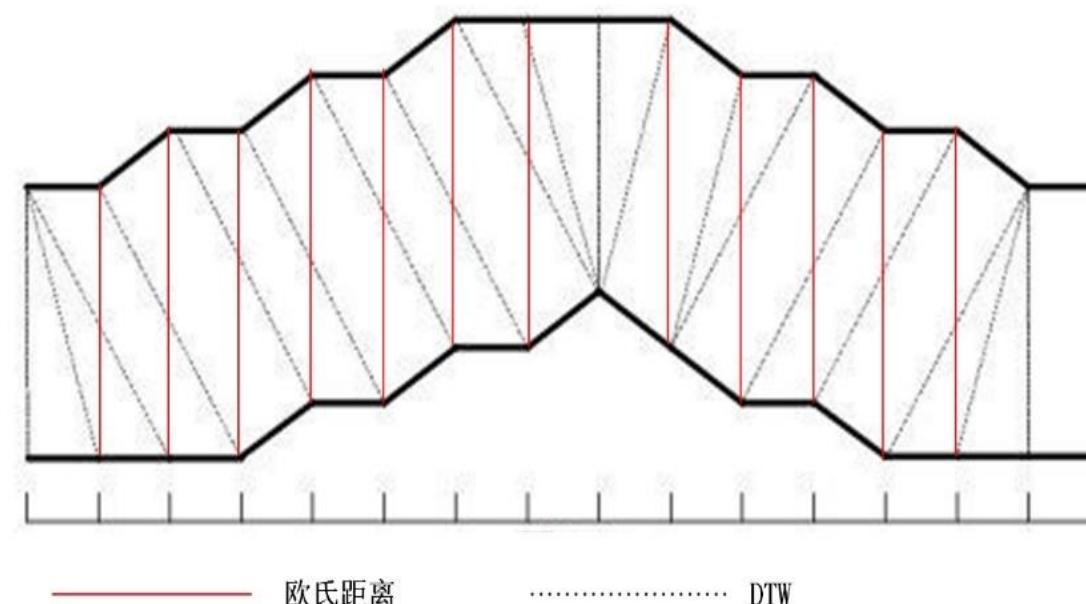
## 02 | LBS大数据关键技术



**聚类：**通过判断事物之间的**差异性**（距离），将相似的事物聚集，不同的事物加以区分。

**聚类方法：**基于**DTW**距离的**K-Medoids**聚类算法

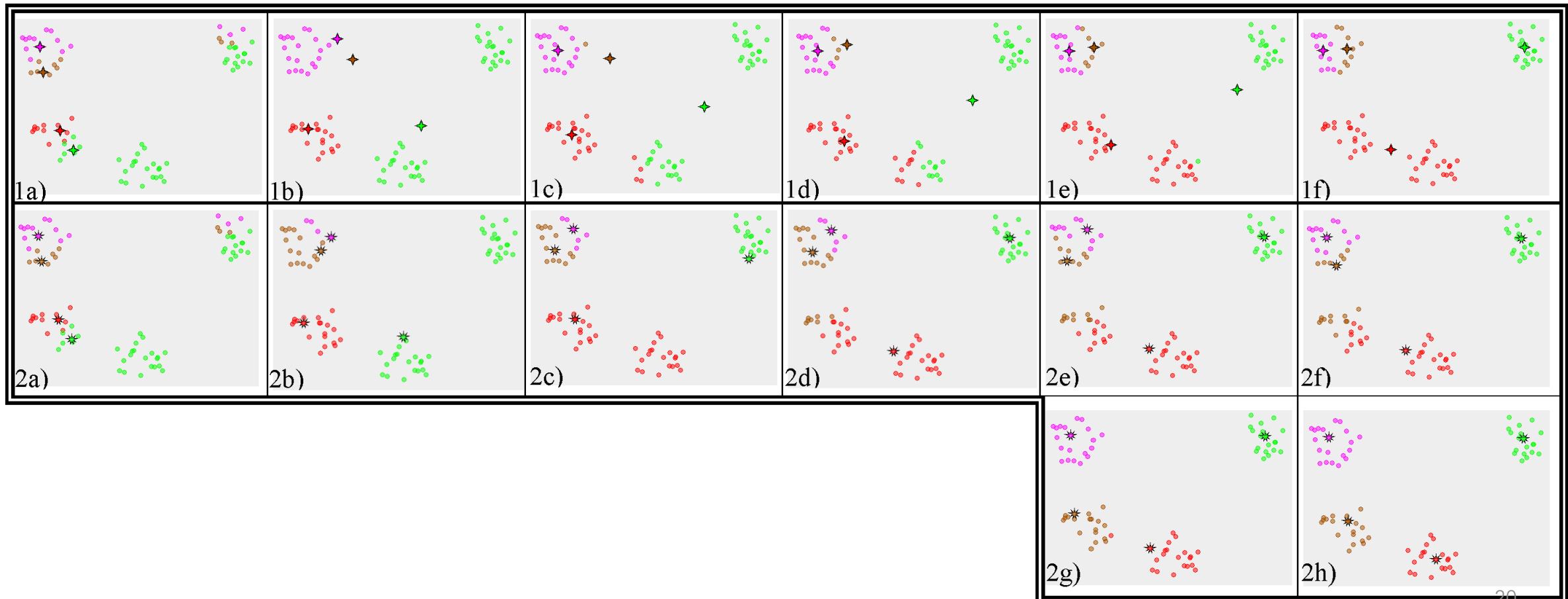
**距离：**动态时间规整(Dynamic Time Warping, **DTW**)：DTW是求两段**时间序列**之间距离的最优方法。DTW可以找到两段时间序列中的对应位置并将其对齐，从而**计算出真正的距离**。



## 02 | LBS大数据关键技术



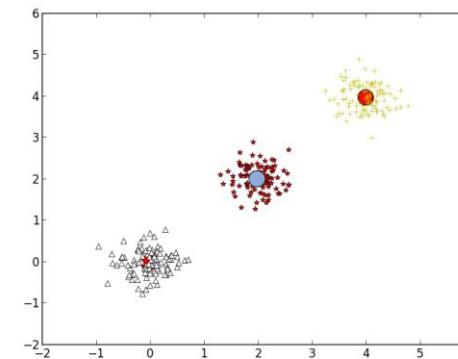
**聚类模式：**根据采取的聚类方法，在聚类中心选择、类别划分等方面所采取的策略。



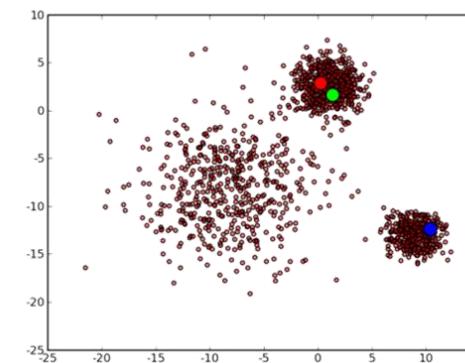
## 聚类效果评价： Silhouette

Silhouette 是用于评价聚类结果的一个系数，可应用在不同聚类算法对同一数据集的聚类情况进行评价。

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$



Sihouette = 0.8

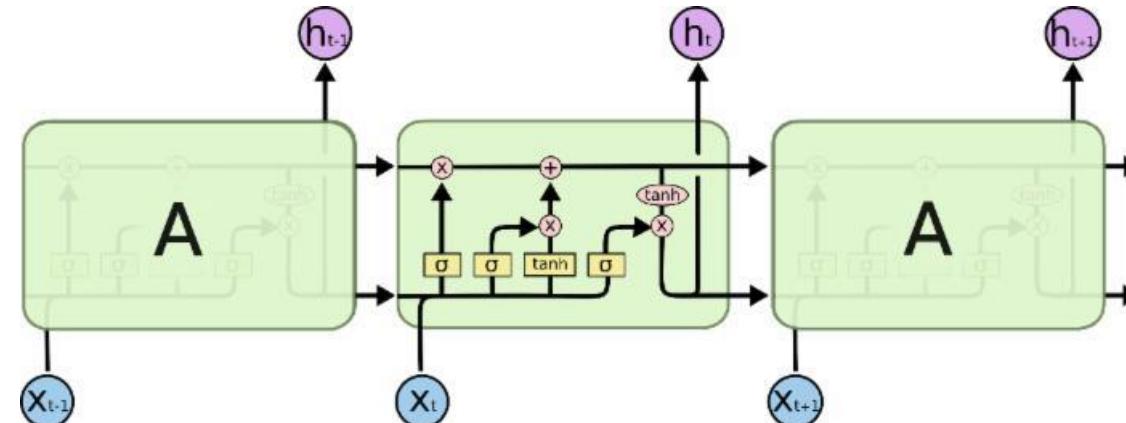


Sihouette = -0.2

其中， $s(i)$  表示 $i^{th}$  向量的Silhouette系数， $a(i)$  为 $i^{th}$  向量所属类别内的平均距离， $b(i)$  为 $i^{th}$  向量到其它类别向量的均值中的最小值

## LSTM

Hochreiter和Schmidhuber提出了长短期记忆网络（LSTM），改进了传统的RNN模型，有效地解决了RNN训练中处理长时序数据的梯度消失问题。在音频和视频数据分析、图像主题建模和道路交通流量预测等领域成功的运用。利用LSTM模型提取时序数据中所包含信息是目前新兴的研究方向之一。



当前时刻t 的LSTM单元内部运行结构 (Yu et al. 2017)

$$\text{forget gate: } F_t = \sigma(W_f \odot [H_{t-1}, X_t] + b_f)$$

$$\text{input gate: } I_t = \sigma(W_i \odot [H_{t-1}, X_t] + b_i)$$

$$\text{output gate: } \tilde{C}_t = \tanh(W_C \odot [H_{t-1}, X_t] + b_C)$$

$$\text{cell input: } C_t = F_t \odot C_{t-1} + I_t \cdot \tilde{C}_t$$

$$\text{cell output: } O_t = \sigma(W_o \odot [H_{t-1}, X_t] + b_o)$$

$$\text{hidden layer output: } H_t = O_t \odot \tanh(C_t)$$

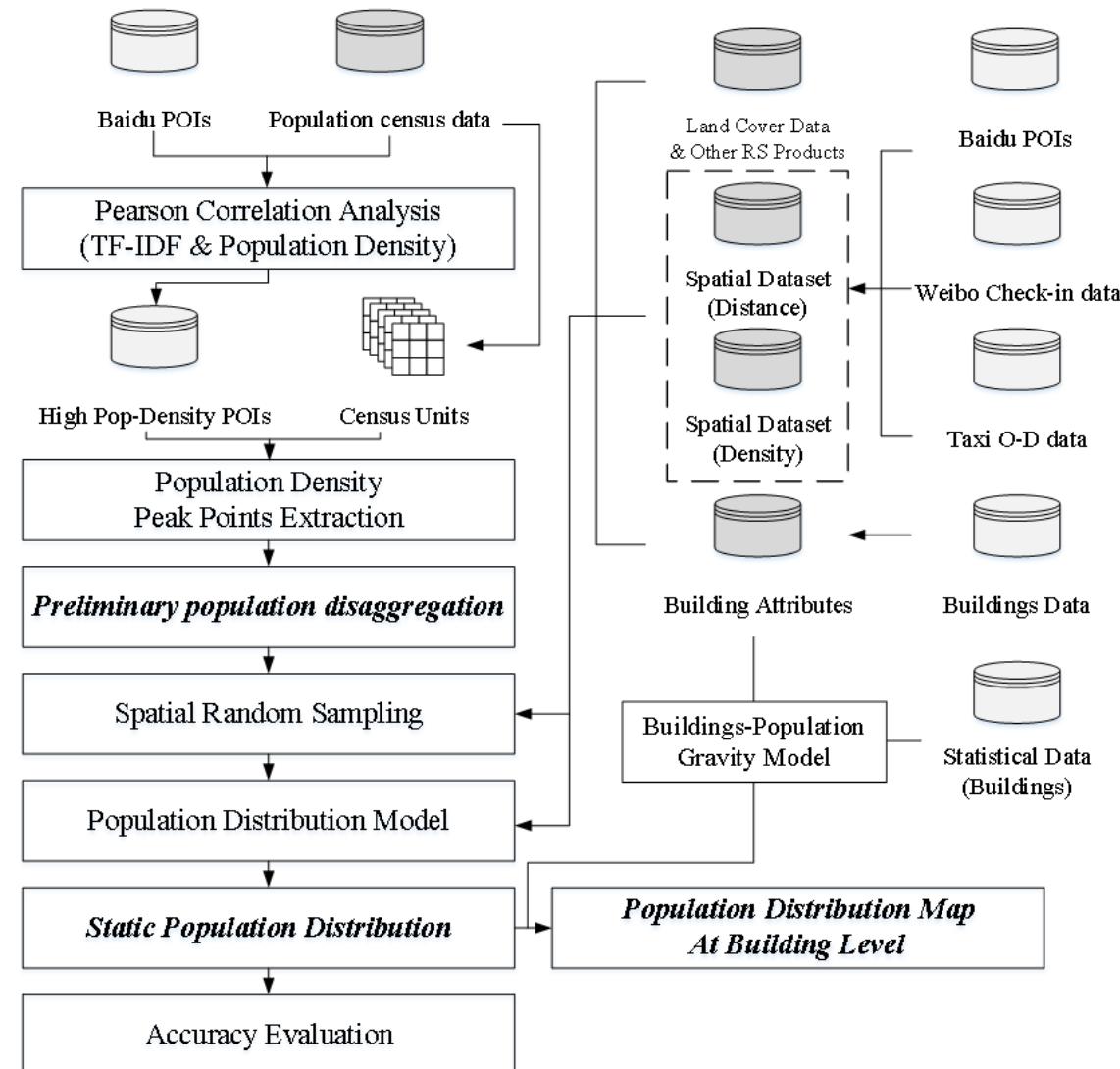


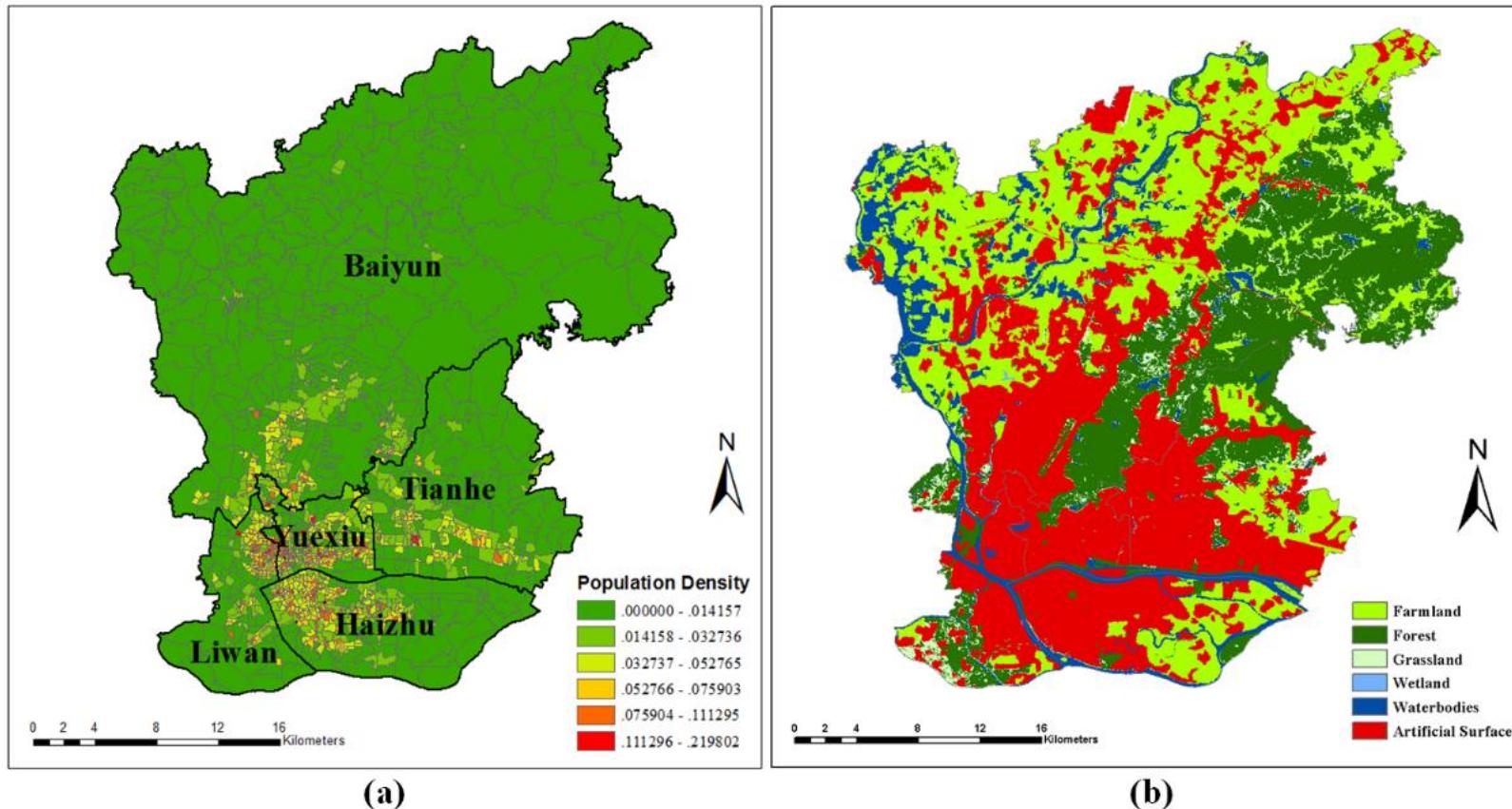
# 主要内容



- 1 LBS大数据简介
- 2 LBS大数据关键技术
- 3 LBS大数据与城市功能
- 4 LBS大数据与贫困

# 03 | LBS大数据与建筑物尺度人口识别





Based on the statistical data of the Sixth China's National Population Census in 2010, the total area of all 5 districts in our study area is 984.8012 square kilometers (sq. km), about 13.61% of the total area of Guangzhou.

The number of administrate street units is 1,278, and permanent resident population is 686.2246 thousands (2010), and this number increase to 694.9637 in 2014, accounting for 58.75% of the total resident population of Guangzhou.

## Auxiliary spatial data sets (37): Geospatial (big) data:

- Baidu POIs (237,402 records)
- Weibo Check-In
- OpenStreetMap Road data
- GPS trajectories of taxies
- Basic GIS data
- Governmental building data

## RS applications:

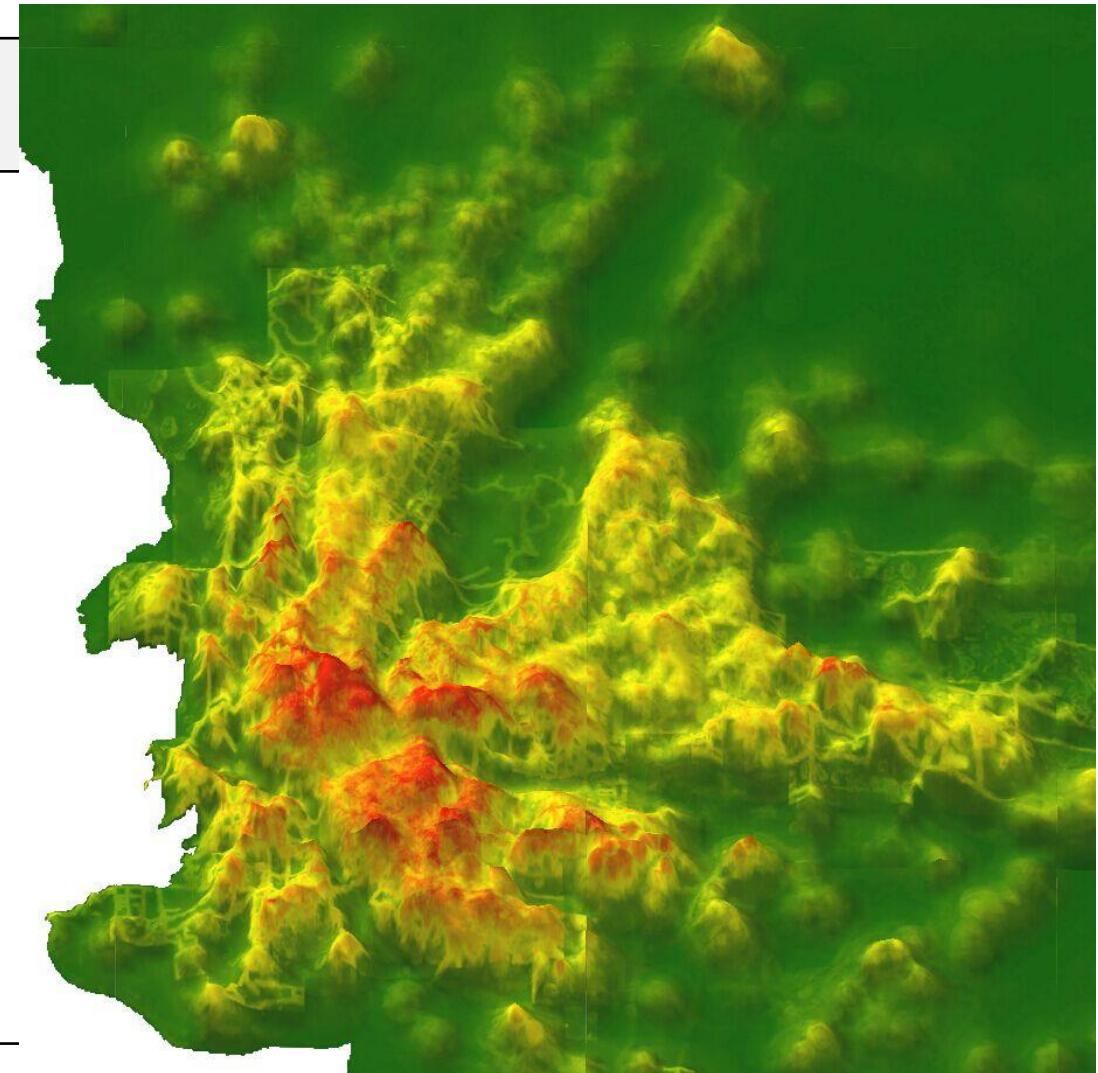
- ASTER DEM/SLOPE
- DMSP-OLS nigh-light data
- Landscan population data
- NDVI (Landsat)

etc.

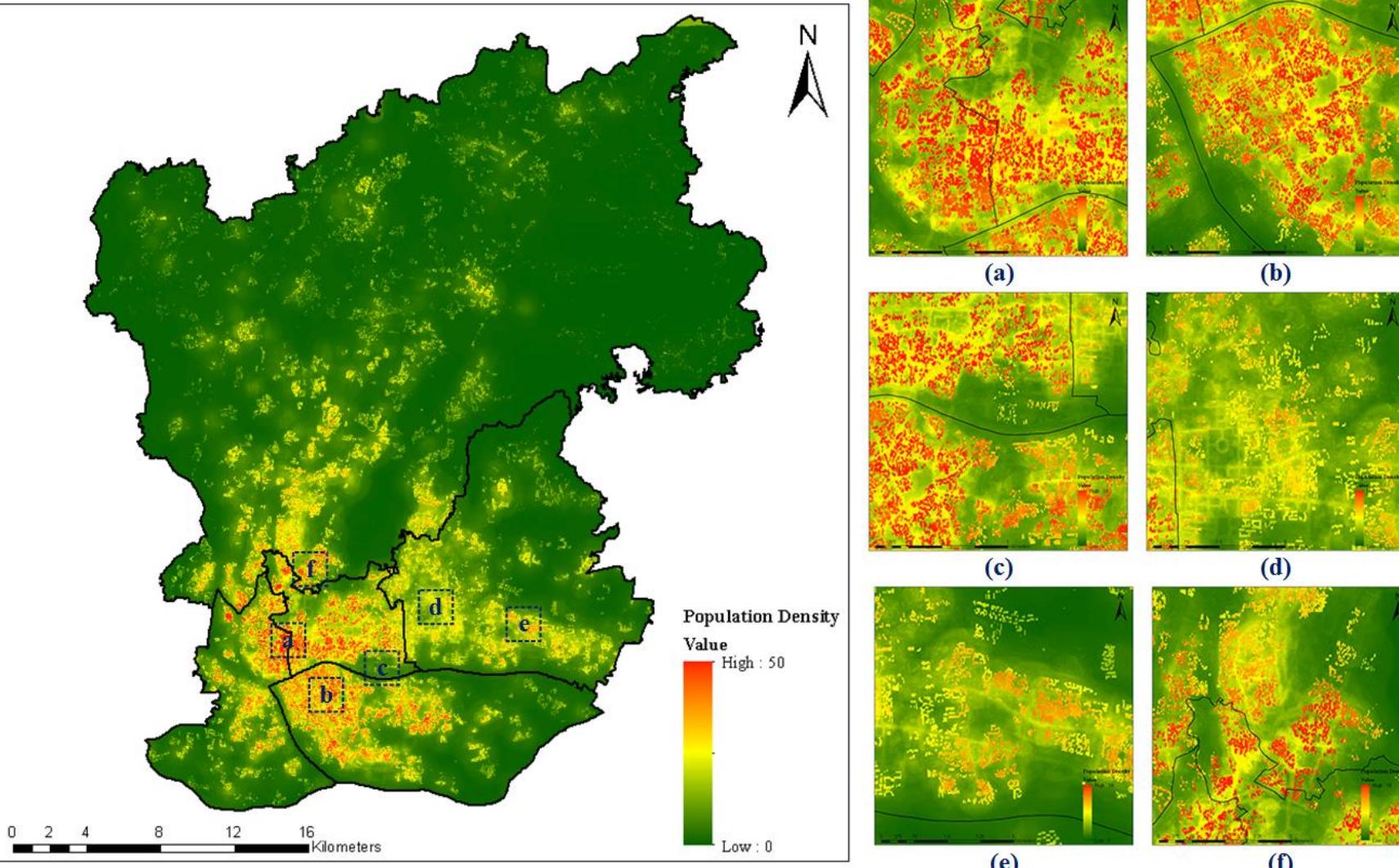
# 03 | LBS大数据与建筑物尺度人口识别



Categories of POIs	Correlation
Clinic facility	0.7874
Residential Community	0.6756
Education	0.6561
Restaurant	0.6536
Automobile service	0.6396
Traffic Facility	0.6262
Life service	0.6261
Road label	0.6215
Shopping	0.5539
Corporation	0.5425

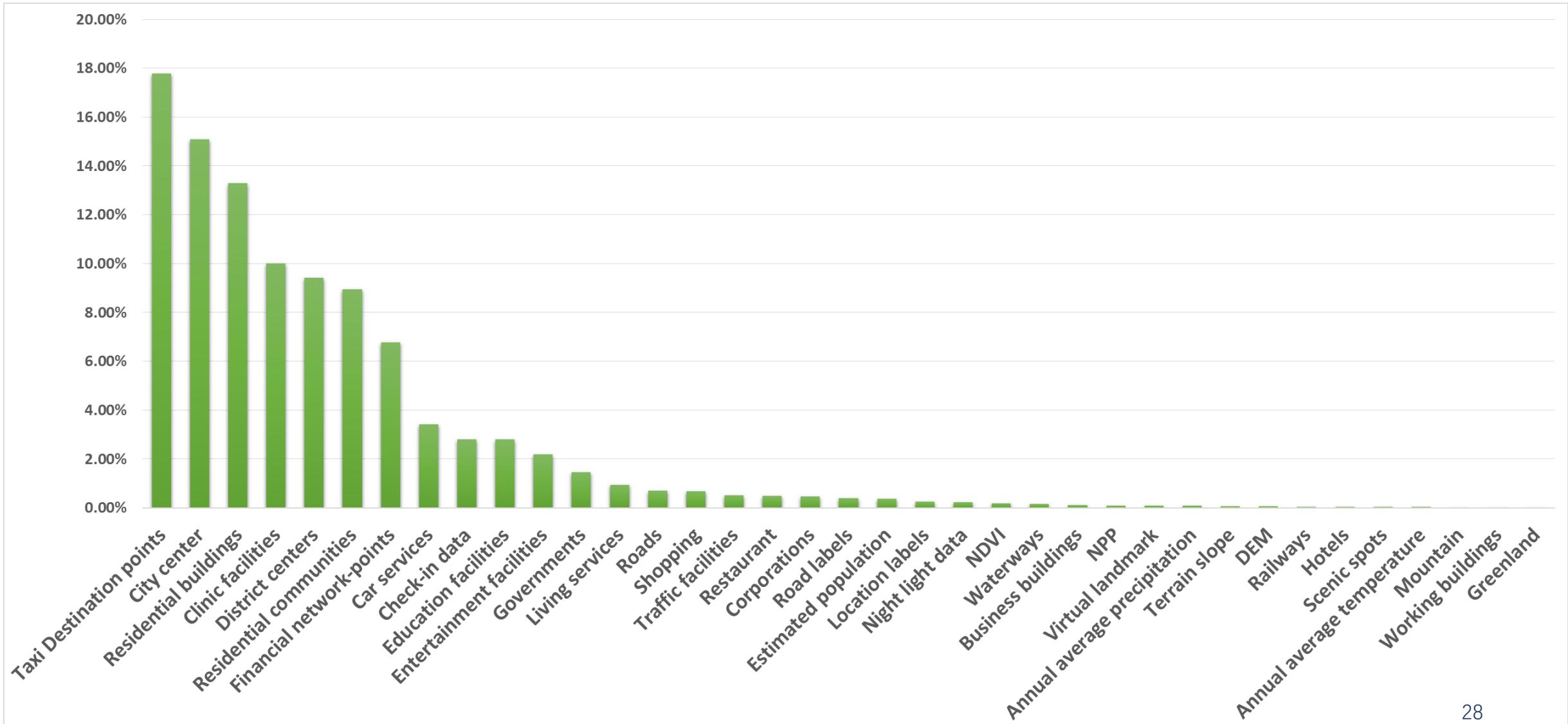


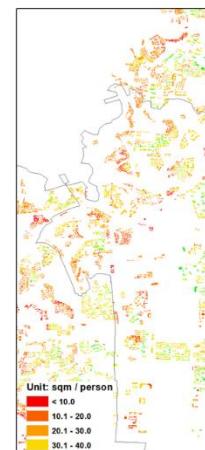
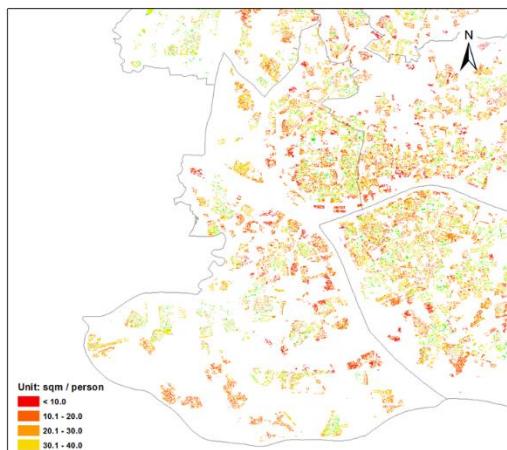
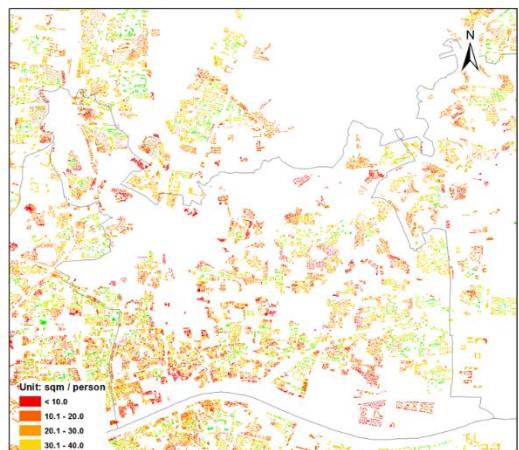
# 03 | LBS大数据与建筑物尺度人口识别



Fine-scale static population distribution map (25 meters).

- a) *Zhongshan Seventh Road* (Old city center);
- b) *Jiangnan Road* (Shopping Center);
- c) *Dongshan lake park* (North, leisure park) and *Sun Yat-sen university* (South, university);
- d) *Tianhe Sports Center* (CBD);
- e) *Tangxia village* (Urban village);
- f) *Guangzhou Railway Station* and *Sanyuanli village* (Urban village)



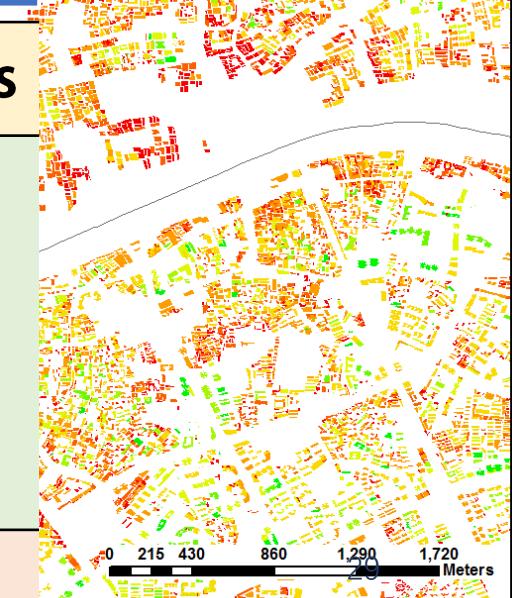


Estimated housing area per capita (unit: sqm) in study area

(a)

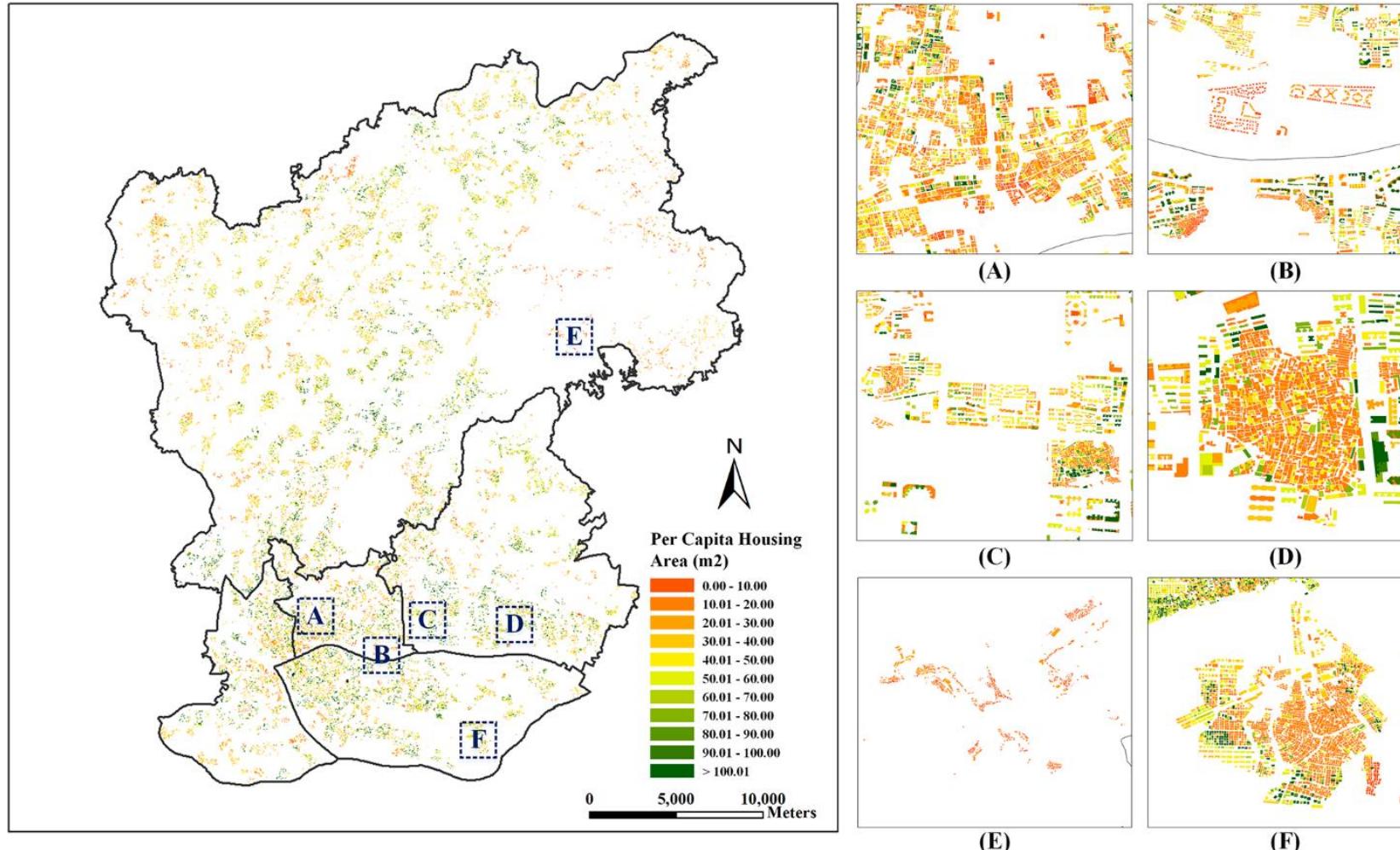
(b)

(c)



Types	Residential Community	Urban Village	Whole Buildings
Yuexiu District	18.8469	9.3940	16.5331
Liwan District	24.0715	15.9917	17.3849
Haizhu District	29.8832	20.6783	21.9969
Tianhe District	37.1856	28.2310	30.1035
Baiyun District	55.1041	35.2407	36.3031
<b>Whole Area</b>	<b>34.9973</b>	<b>30.5654</b>	<b>31.0923</b>

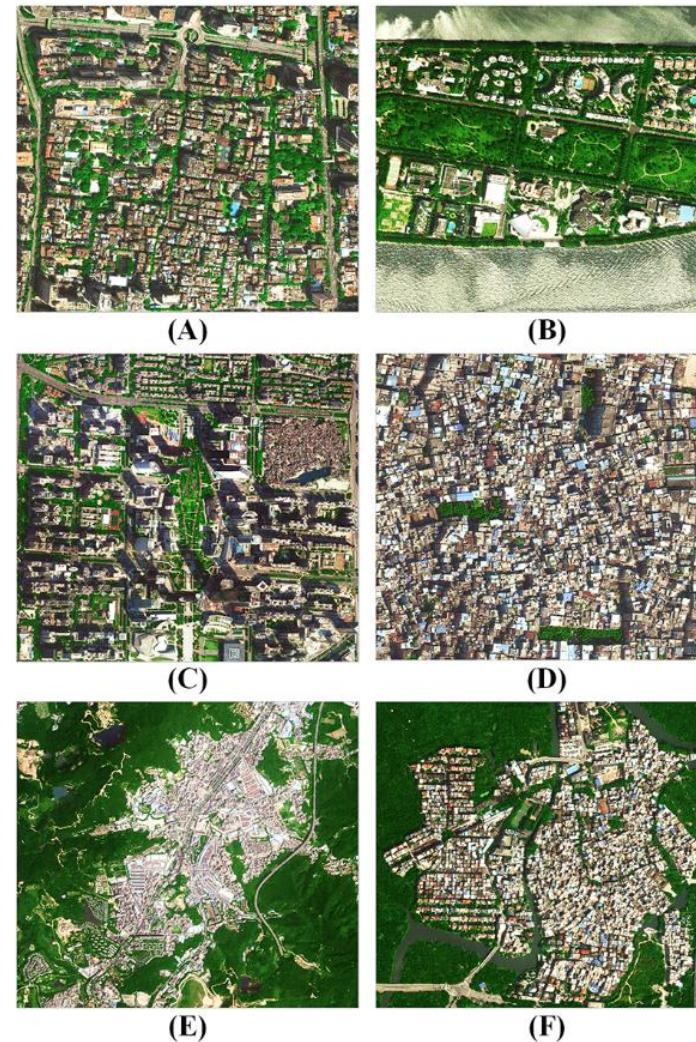
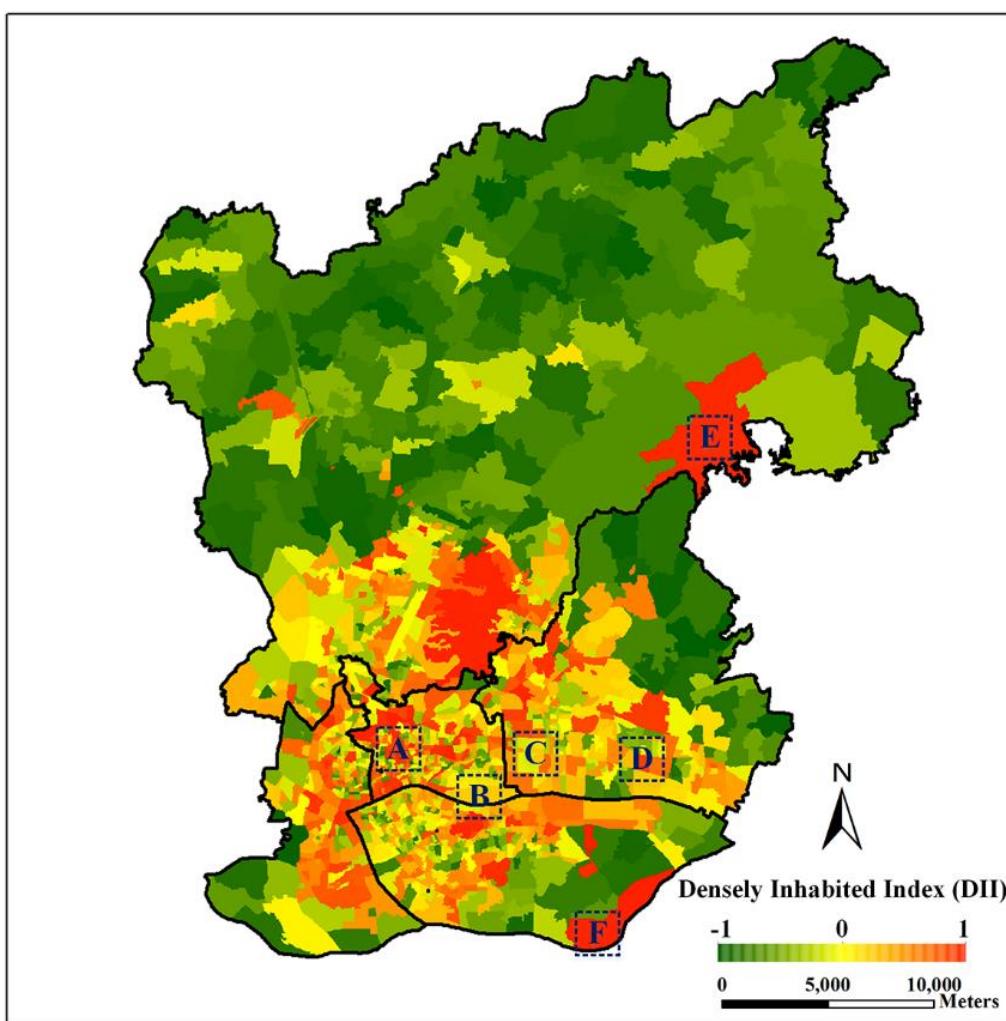
# 03 | LBS大数据与建筑物尺度人口识别



**Housing area per capita** at the building level from the proposed population-building gravity model.

- (A) Guangdong Province's government (city center of Guangzhou)
- (B) Er-sha Island (luxury residential area)
- (C) Flower City Square (CBD)
- (D) Tangxia Village (urban village)
- (E) Taihe Town (satellite town)
- (F) Xiaozhou Village (rural area)

# 03 | LBS大数据与建筑物尺度人口识别



**Densely inhabited index (DII)** in each administrative unit at the street level from the simulated population distribution at the building level.  
(A) Guangdong Province' s government building (city center of Guangzhou)  
(B) Er-sha Island (luxury residential area)  
(C) Flower City Square (CBD)  
(D) Tangxia Village (urban village)  
(E) Taihe Town (satellite town)  
(F) Xiaozhou Village (rural area)

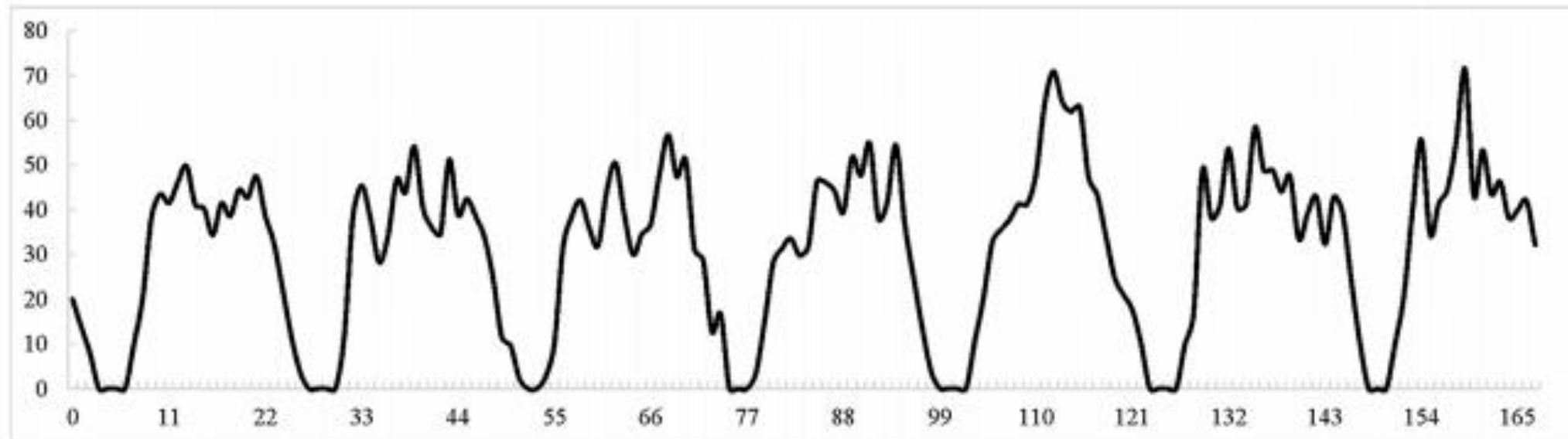
# 03 | LBS大数据与建筑物功能



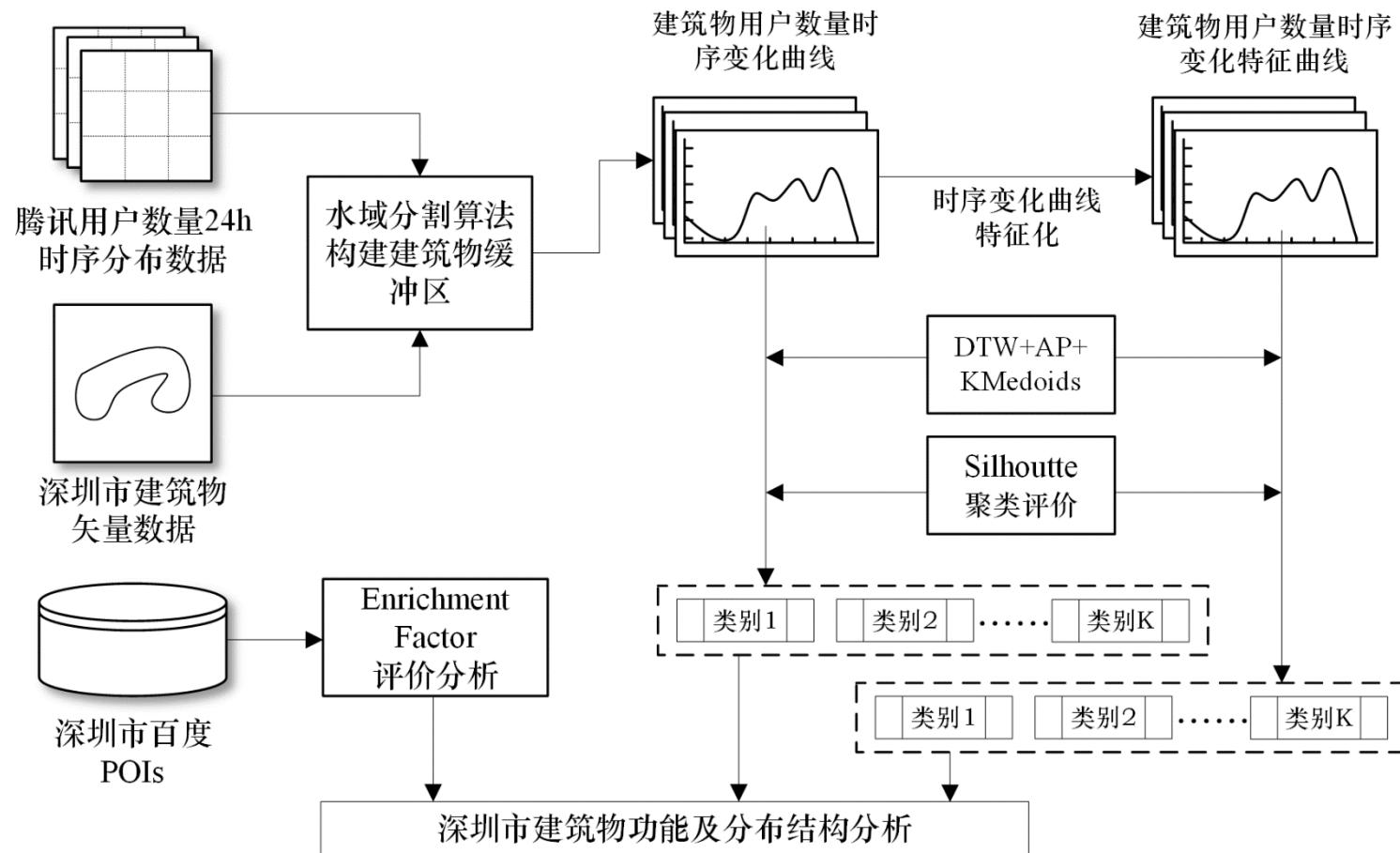
早上八点的你和晚上十点的你有可能出现的地方?

中午十二点学校哪里人最多?

周一到周五生活节奏是不是相同? 周六周日生活节奏是不是相同?

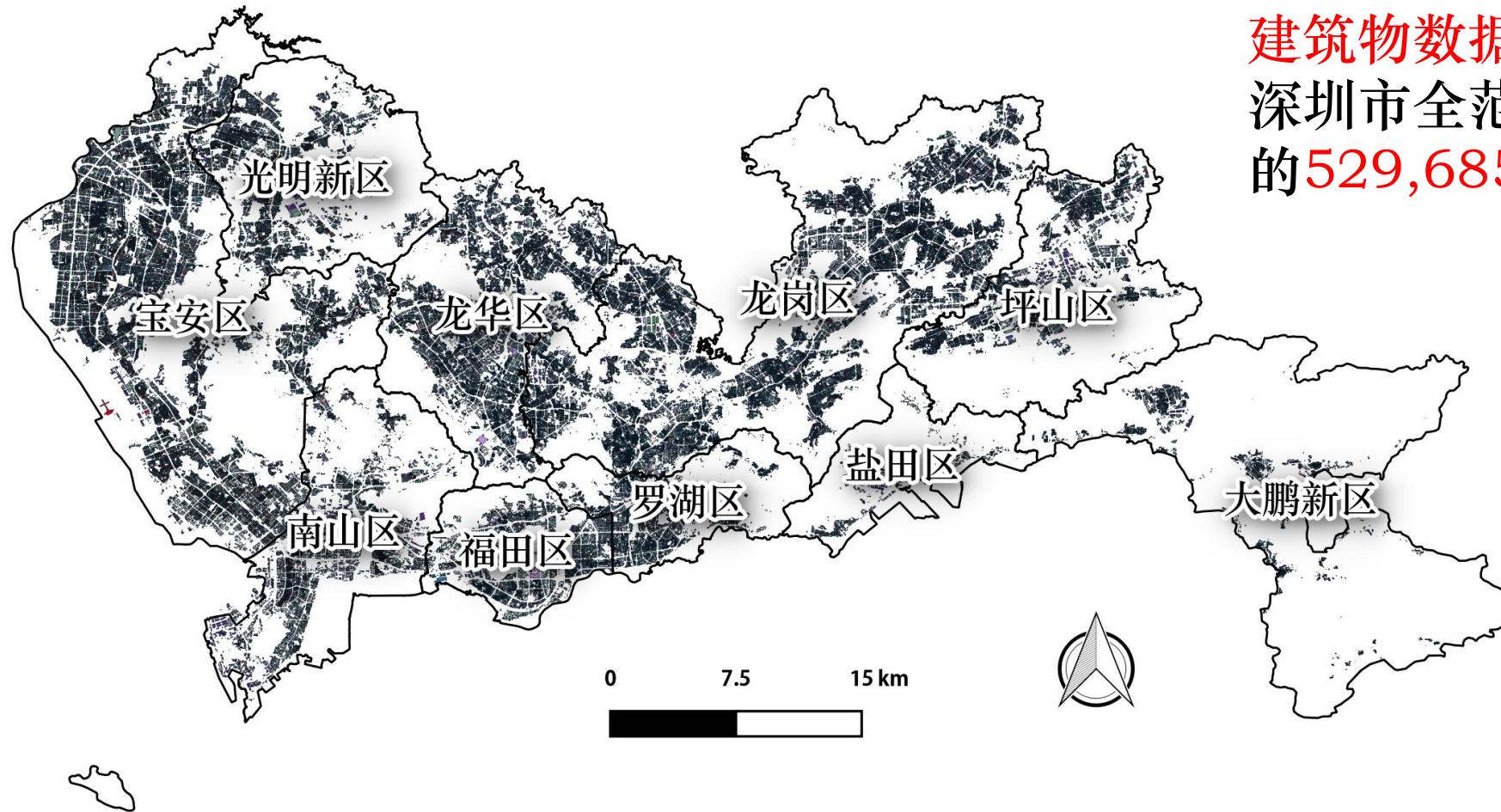


# 03 | LBS大数据与建筑物功能



1. 通过水域分割算法，构建建筑物采样区，将栅格尺度上的RTUD分配到邻近建筑物中，构建**建筑物尺度RTUD**；
2. 将建筑物尺度上的RUTD进行**归一化处理**，为了在相同功能的建筑物之间，消除由不同楼层、占地面积等因素引起的RTUD数量上的差异；
3. 利用**基于DTW距离的K-Medoids聚类算法**进行聚类，并通过Silhouette方法进行聚类评价；
4. 通过POI辅助数据计算聚类类别上POI的**富集程度**，对结果进行归纳分析。

# 03 | LBS大数据与建筑物功能



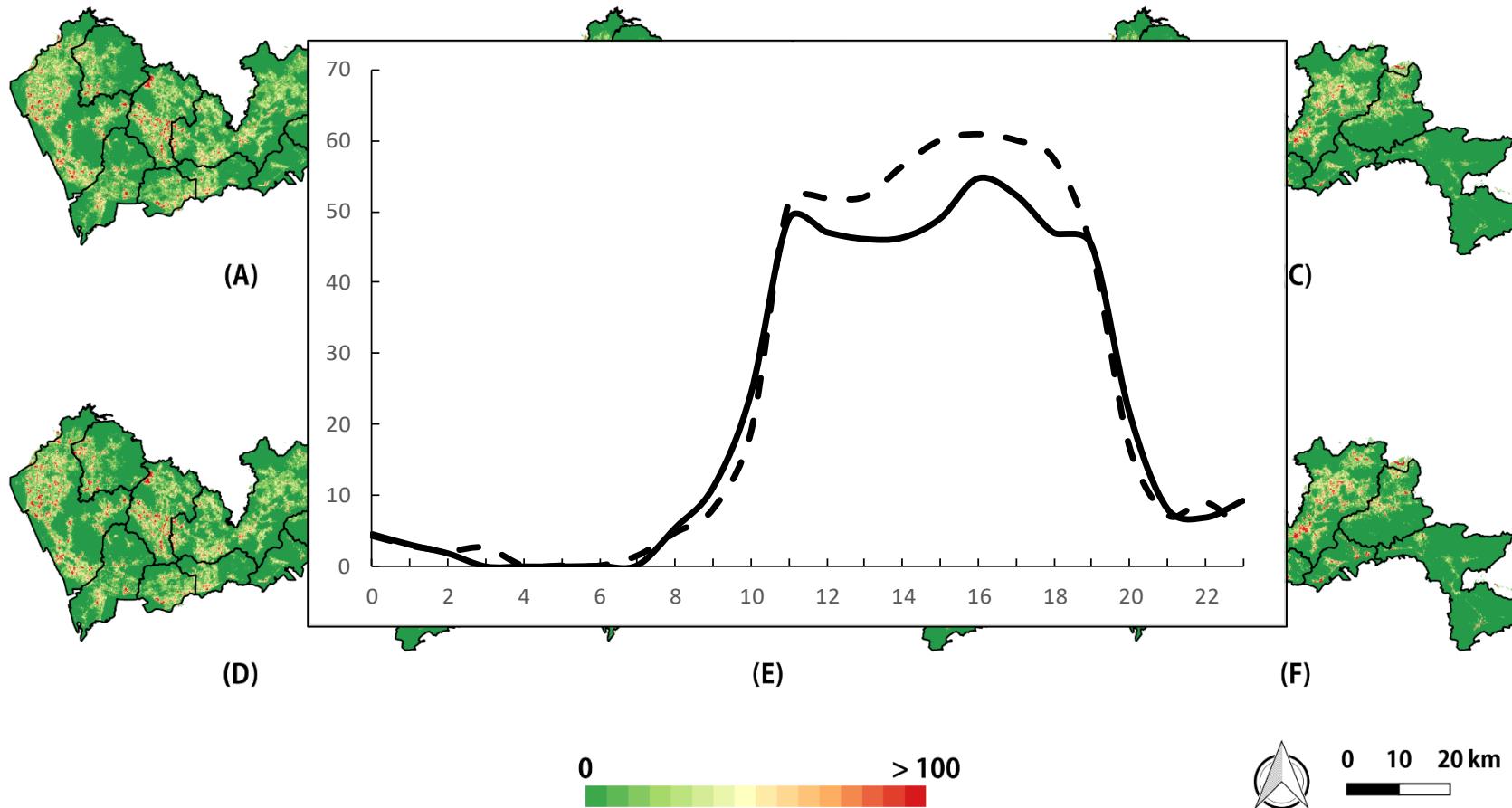
研究区为深圳市全范围，包含南山区、福田区、罗湖区和宝安区等10个发展状况和规划情况不同的区划。

建筑物数据：

深圳市全范围内在2016年采集的**529,685**个建筑物矢量数据。

由于改革开放政策需要，曾在1985~2005年间在深圳市划定边防管理线，对罗湖区、福田区、南山区和盐田区实行有管理的出入（俗称“**关内关外**”政策）。

其对社会发展情况仍未消失。

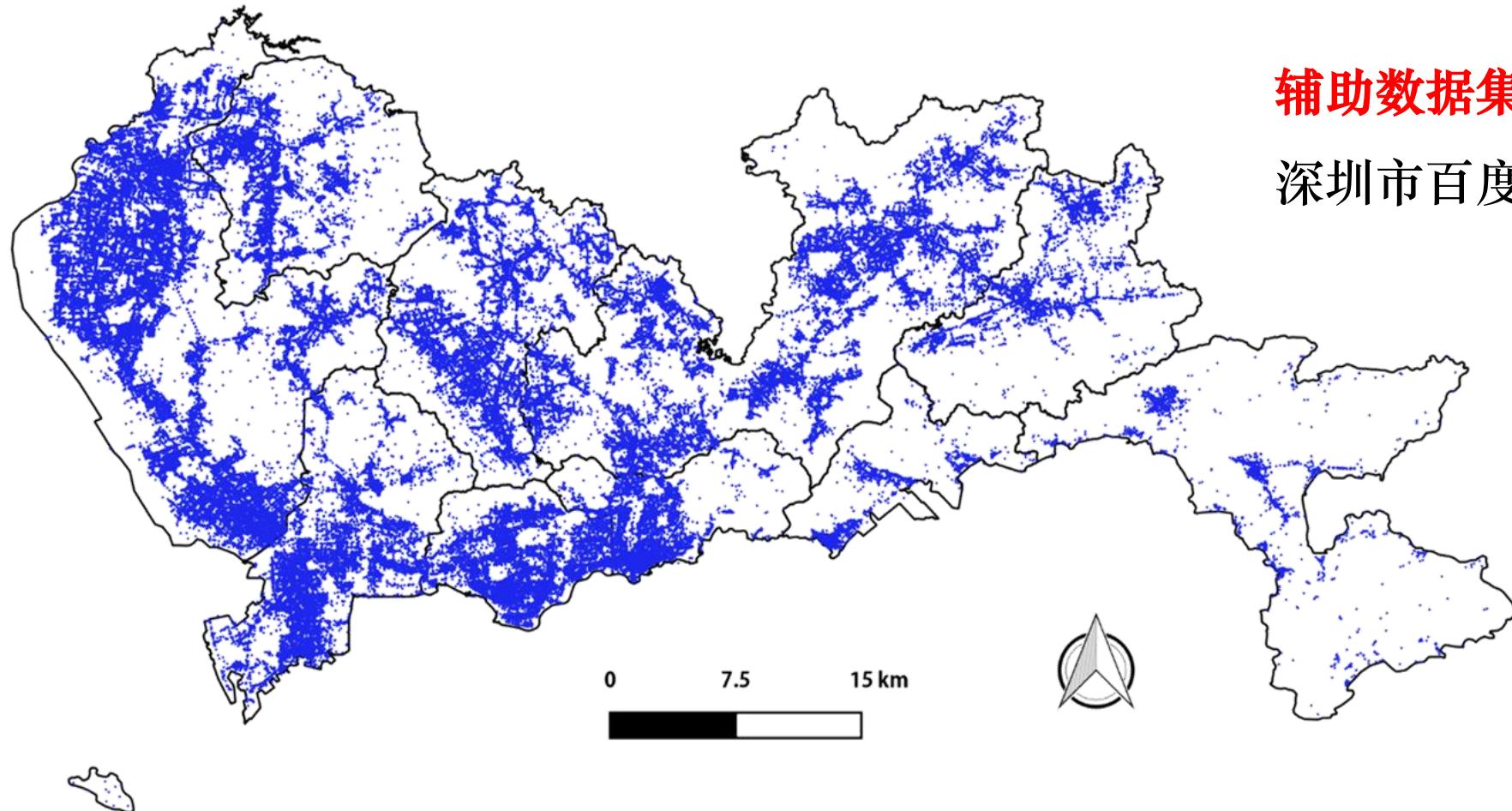


深圳市实时腾讯用户定位数据：工作日的 (A) 8:00 am, (B) 16:00 pm 和 (C) 22:00 pm；休息日的 (D) 8:00 am, (E) 16:00 pm 和 (F) 22:00 pm

**主要数据集：**  
深圳市腾讯用户时序分布  
数据 (RTUD)

采集自2015年6月15日至  
21日，以1h为间隔，包含  
工作日和休息日

25m分辨率的栅格数据



深圳市百度POIs分布情况

辅助数据集：

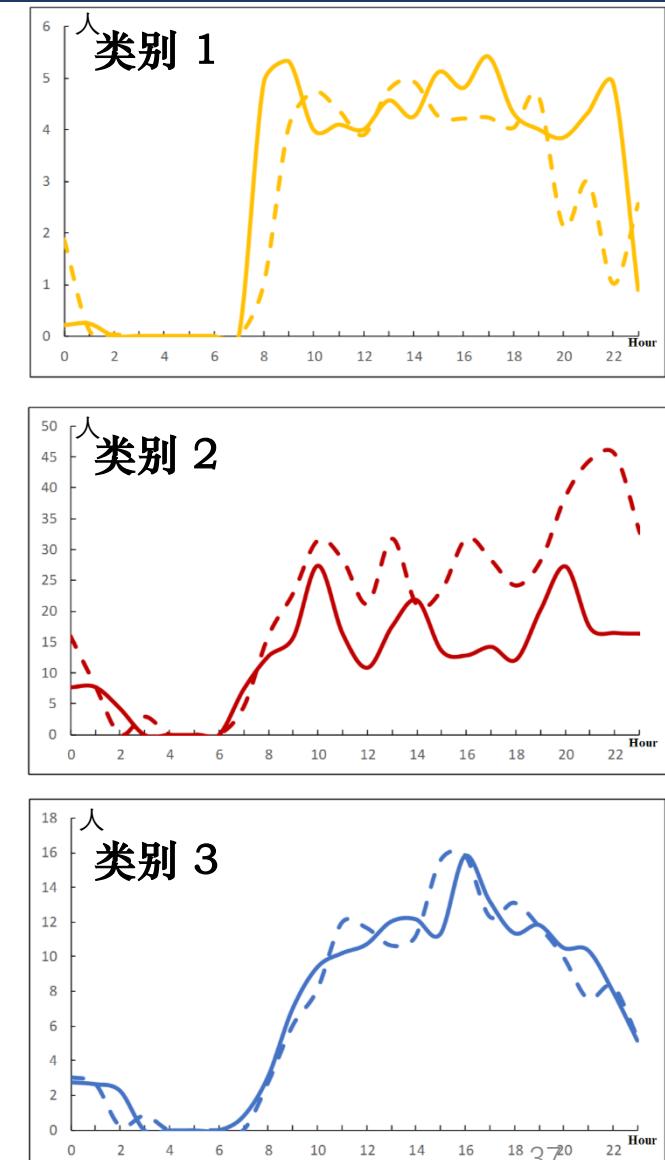
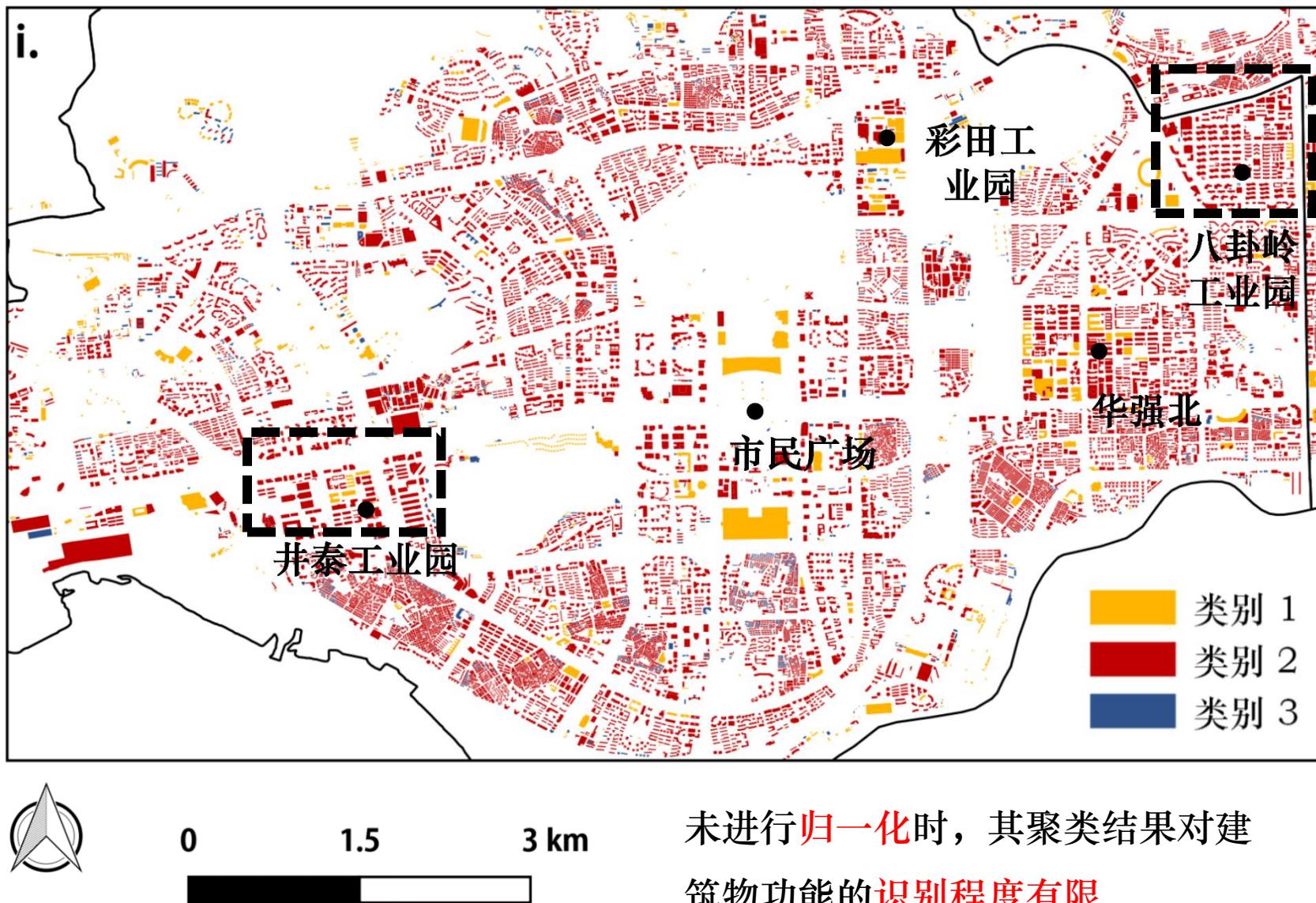
深圳市百度POI数据共**278,245**条

分为**17**类：

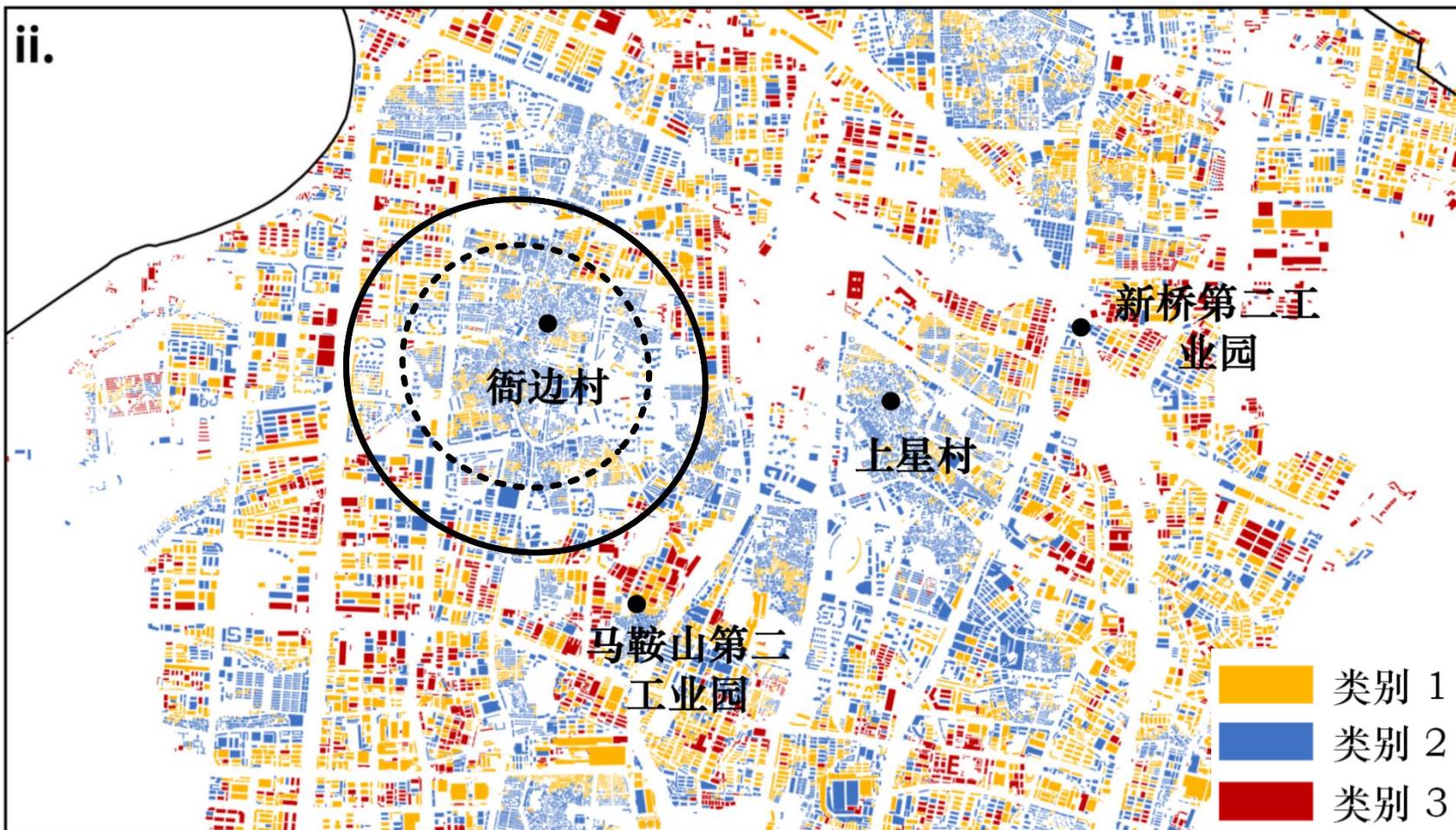
包含居民小区、小学、初高中、技校、大学、研究所、企业公司、金融机构、政府机关、诊所、医院、公园、广场、餐饮、零售店、娱乐和旅馆



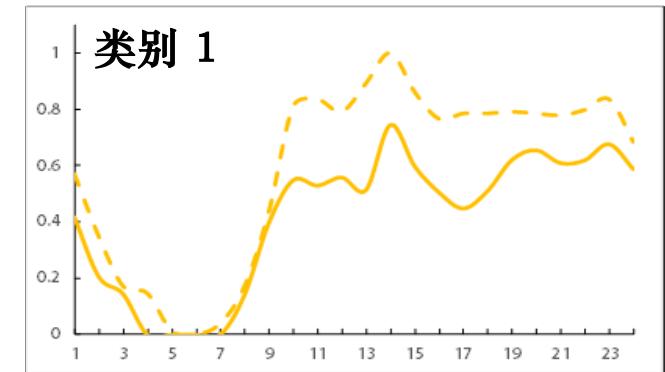
— 工作日 — 休息日



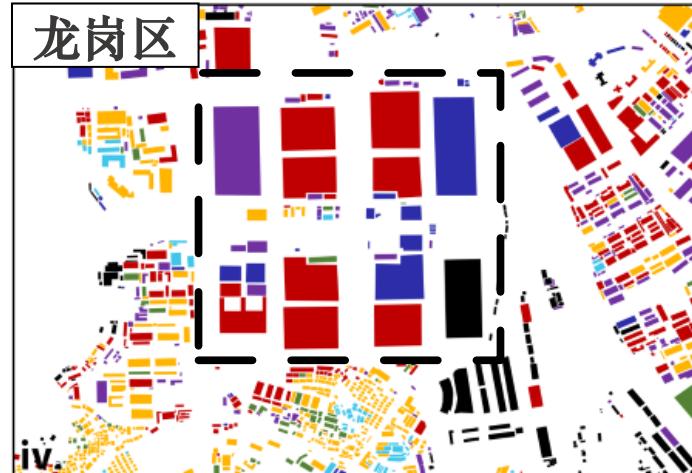
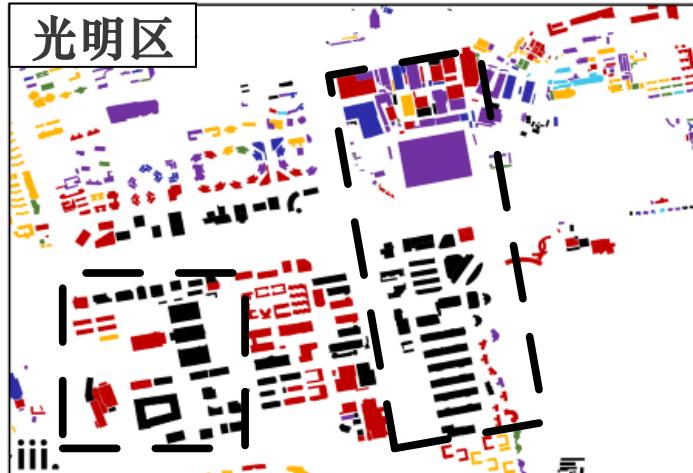
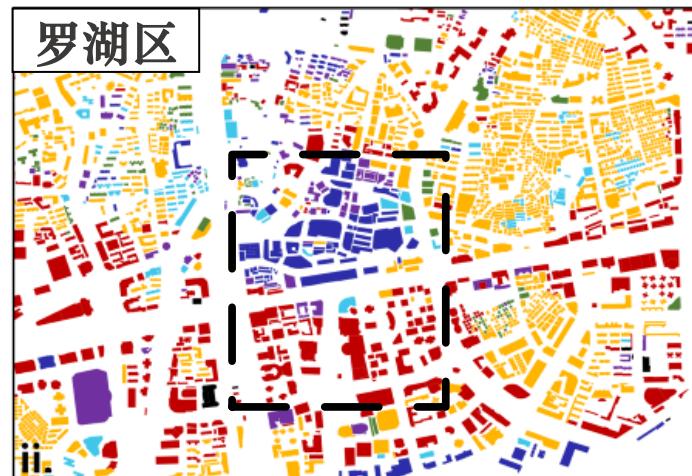
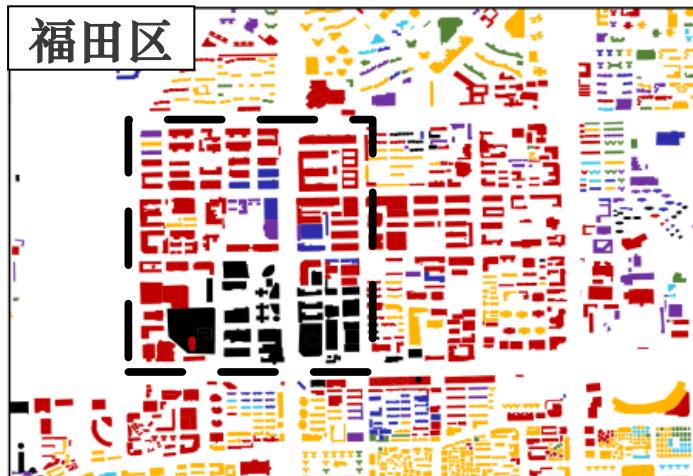
k	类别	居民小区	小学	初中/高中	技校	大学	研究所	企业公司	金融机构	政府机关	诊所	医院	公园	广场	餐饮	零售店	娱乐	旅馆	
用户总数 时序变化 曲线聚类	2	1	0.98	0.74	0.43	0.82	0.67	0.58	0.83	1.08	0.91	1.22	0.78	1.45	1.23	1.17	1.10	1.48	1.29
		2	1.01	1.08	1.17	1.05	1.10	1.12	1.05	0.98	1.03	0.94	1.07	0.87	0.93	0.95	0.97	0.86	0.91
	3	1	0.72	0.46	0.62	1.26	1.59	0.94	0.84	1.44	0.91	0.98	1.02	2.24	1.19	1.28	0.83	1.63	1.02
		2	1.09	0.98	0.21	0.64	0.45	0.32	1.09	0.50	0.90	1.30	0.69	0.00	1.19	0.98	1.18	0.55	0.76
		3	1.02	1.05	1.06	0.99	0.97	1.03	1.01	0.98	1.01	0.99	1.01	0.93	0.98	0.98	1.01	0.96	1.01
	8	1	0.62	0.23	0.42	1.27	1.76	1.22	0.81	1.68	0.87	0.86	1.26	3.50	1.55	1.34	0.77	1.77	0.94
		2	0.62	2.24	0.86	0.69	0.00	0.52	1.64	0.34	1.29	0.33	0.28	0.00	1.12	0.60	0.37	0.40	0.52
		3	0.66	0.53	0.80	0.96	0.57	0.00	0.88	0.91	0.82	1.36	0.17	0.00	0.35	1.33	0.97	1.53	1.10
		4	0.39	0.00	0.00	0.00	0.00	2.43	0.93	0.71	0.68	0.51	0.00	0.00	0.00	2.00	0.57	1.17	0.25
		5	1.28	0.62	0.00	0.66	0.00	0.58	0.89	0.51	0.91	1.89	0.41	0.00	1.25	1.08	1.52	0.56	0.80
		6	1.05	0.74	0.25	0.67	1.34	0.23	1.08	0.51	0.90	1.14	0.49	0.00	1.32	1.01	1.28	0.43	0.65
		7	0.98	1.27	0.68	0.53	0.48	0.63	1.14	0.56	0.98	0.86	0.75	0.00	1.04	0.89	1.24	0.48	0.66
		8	1.03	1.03	1.12	1.06	1.04	1.08	0.99	1.04	1.02	1.01	1.06	1.05	0.96	0.99	0.97	1.04	1.06
用户总数 时序变化 特征曲线 聚类	2	1	1.18	1.07	0.81	0.93	0.61	0.38	0.85	0.93	0.96	1.20	0.85	0.89	1.13	1.12	1.19	1.16	1.13
		2	0.46	0.81	1.57	1.22	2.17	2.83	1.46	1.20	1.13	0.41	1.43	1.32	0.61	0.63	0.42	0.53	0.62
	3	1	1.08	1.10	0.87	0.90	0.60	0.51	0.91	0.91	0.94	1.14	0.85	1.15	0.98	1.09	1.17	1.04	1.04
		2	1.20	0.99	0.94	1.13	1.24	0.59	0.88	1.04	1.04	1.08	1.05	1.29	1.34	1.08	1.01	1.26	1.19
		3	0.37	0.64	1.57	1.17	2.06	3.47	1.54	1.26	1.14	0.35	1.45	0.00	0.53	0.56	0.36	0.44	0.52
	7	1	1.10	1.05	0.76	0.87	0.32	0.39	0.82	0.94	0.87	1.24	0.93	1.66	0.88	1.15	1.31	1.12	1.11
		2	1.25	0.95	0.98	1.33	1.36	0.58	0.76	1.12	1.15	1.03	1.33	3.42	1.75	1.20	0.93	1.51	1.34
		3	0.47	0.74	1.21	1.11	1.92	2.22	1.46	1.08	1.03	0.45	1.37	0.80	0.43	0.68	0.47	0.52	0.65
		4	0.43	0.51	2.19	1.38	3.65	5.42	1.50	1.80	1.52	0.26	0.68	0.00	0.67	0.46	0.22	0.37	0.42
		5	1.30	0.67	0.44	0.88	0.57	0.31	0.73	0.95	0.88	1.46	0.62	0.00	1.24	1.20	1.30	1.34	1.36
		6	1.50	1.94	1.22	0.96	0.20	0.40	0.82	0.84	1.39	1.11	0.81	0.00	1.79	1.02	1.29	0.87	0.84
		7	0.74	1.92	2.12	0.94	1.15	0.63	1.40	0.67	0.94	0.51	1.30	0.00	0.86	0.75	0.64	0.69	0.52



- 类别1为职住平衡好的建筑物
- 类别2为职住平衡差的建筑物，如城中村
- 类别3为工作为主的建筑物

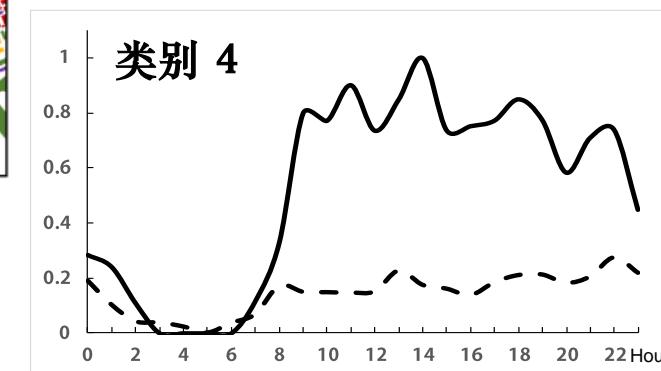
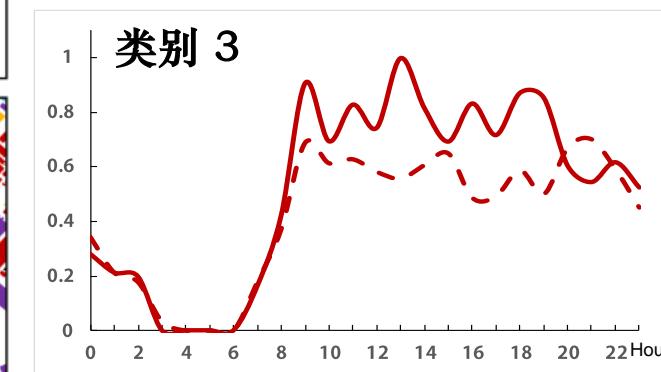
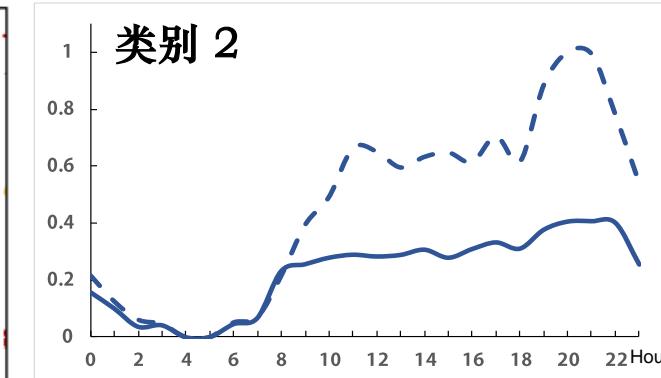


# 03 | LBS大数据与建筑物功能



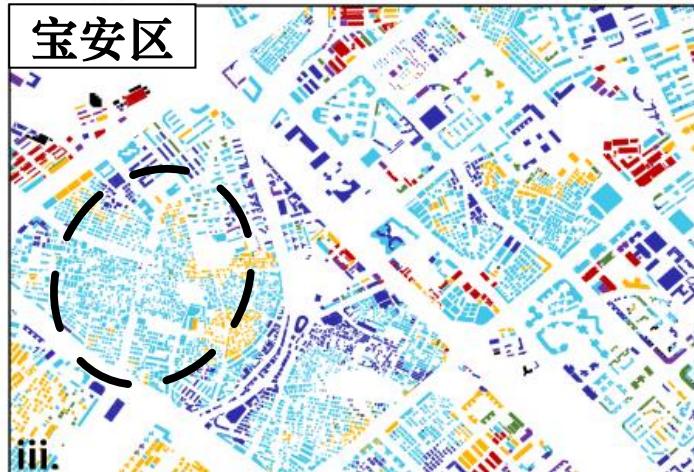
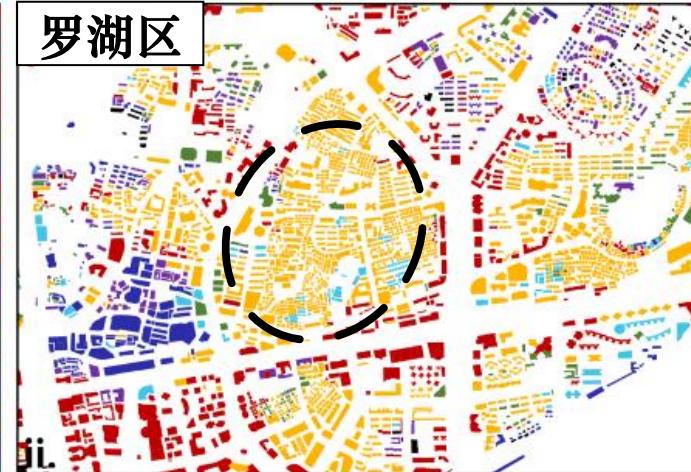
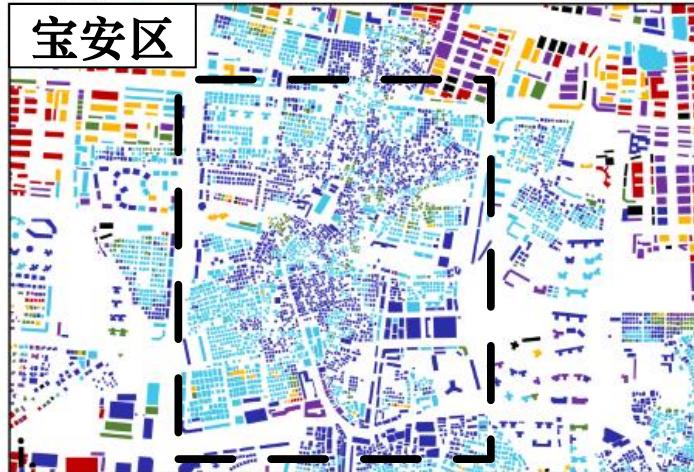
图例

类别 1	类别 3	类别 5	类别 7
类别 2	类别 4	类别 6	



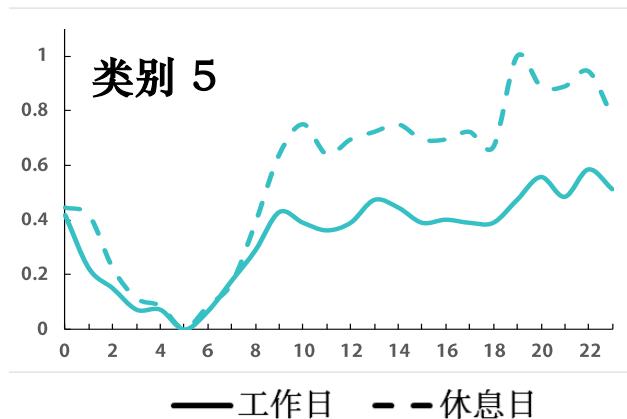
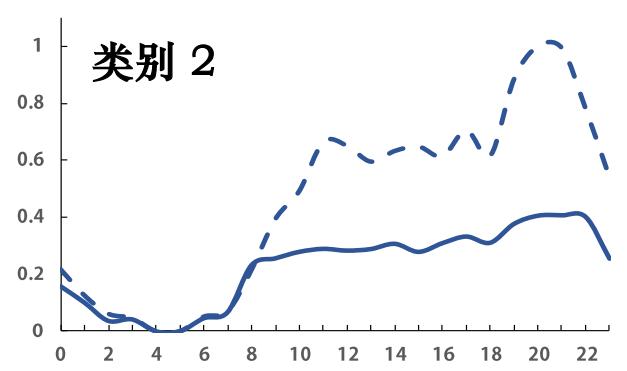
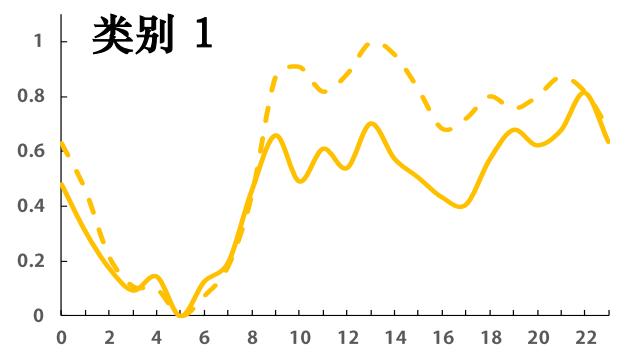
类别2以购物娱乐功能为主

类别4以商业功能为主



图例

类别 1	类别 3	类别 5	类别 7
类别 2	类别 4	类别 6	

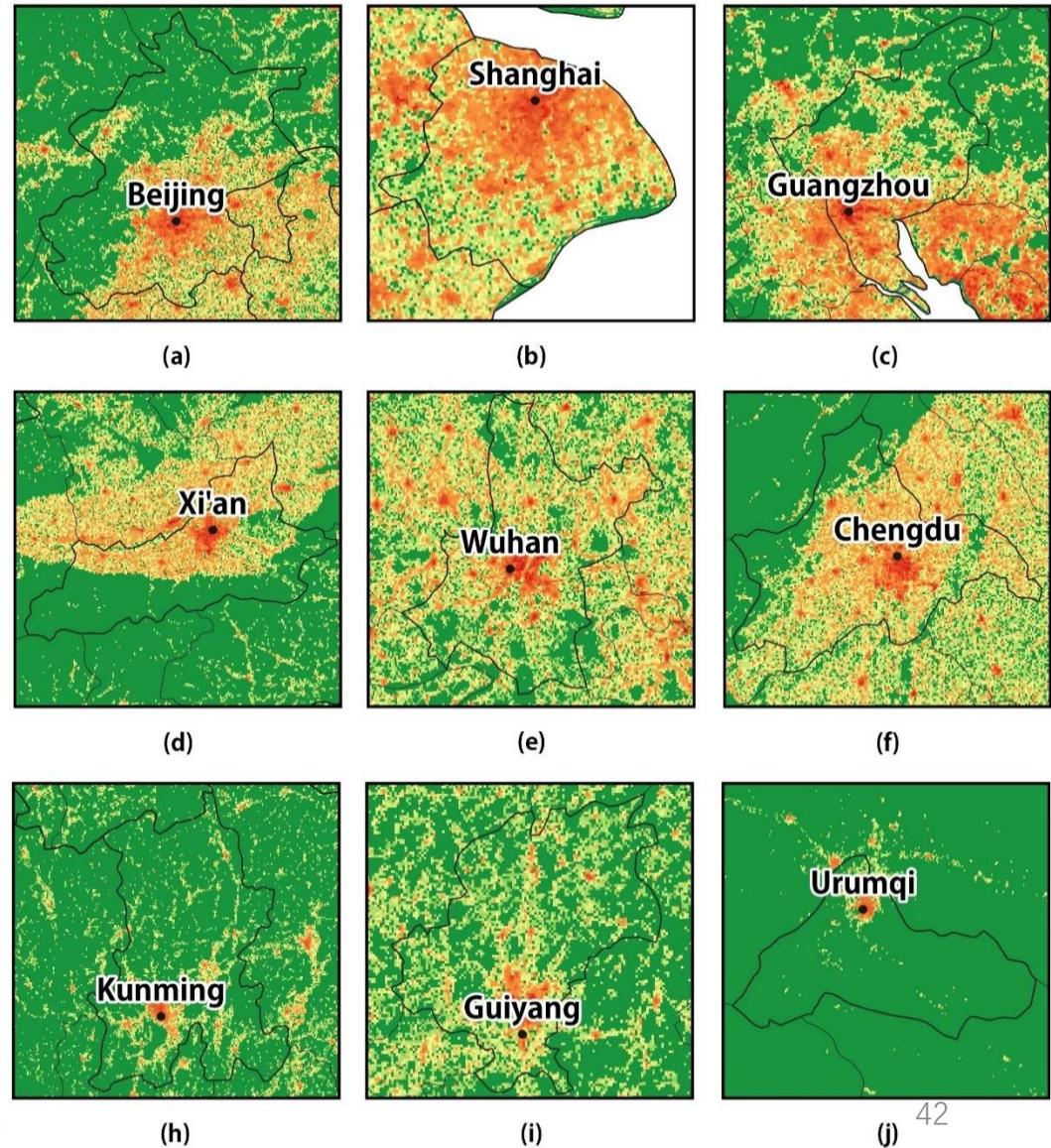
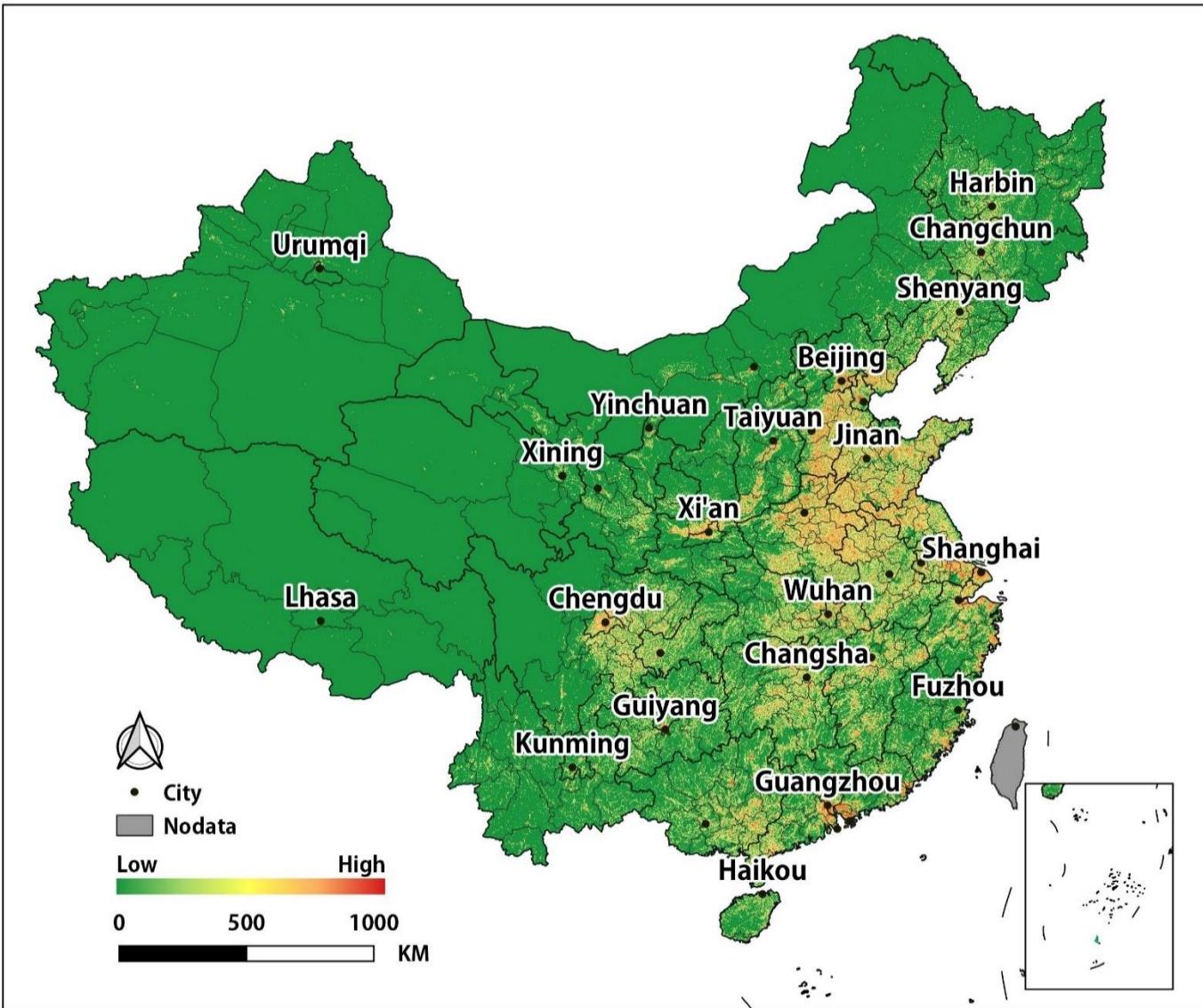


类别1的城中村主要分布在发展较好的地区，发展较差的地区还以居民区存在

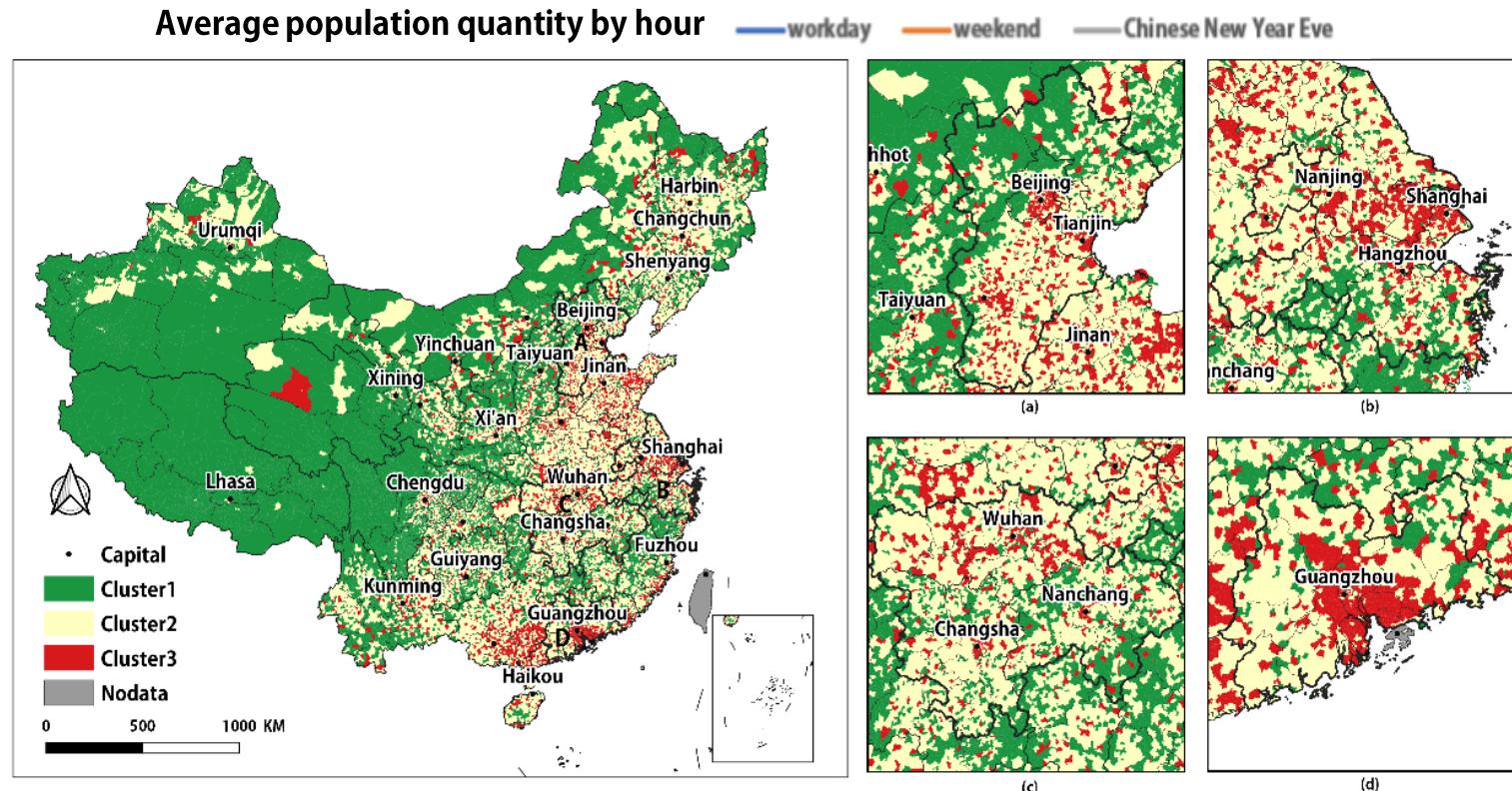
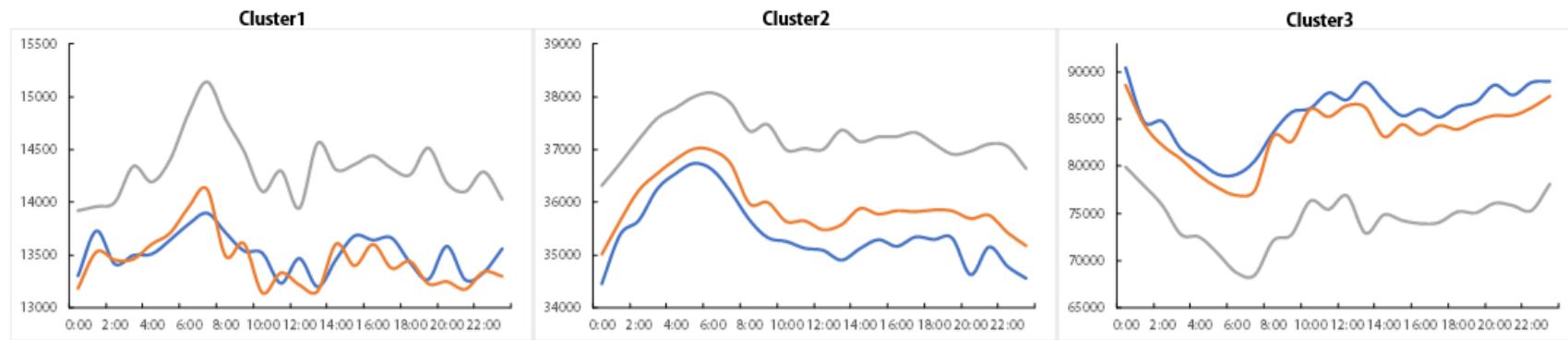
类别2在不同的地区具有不同的功能情况  
如发展较好的地区以购物娱乐存在，较差地区以城中村存在

类别5主要为城中村功能建筑物  
但发展一般的地区城中村以类别5为主，较差地区与类别2共同存在于城中村中

# 03 | LBS大数据与城市功能



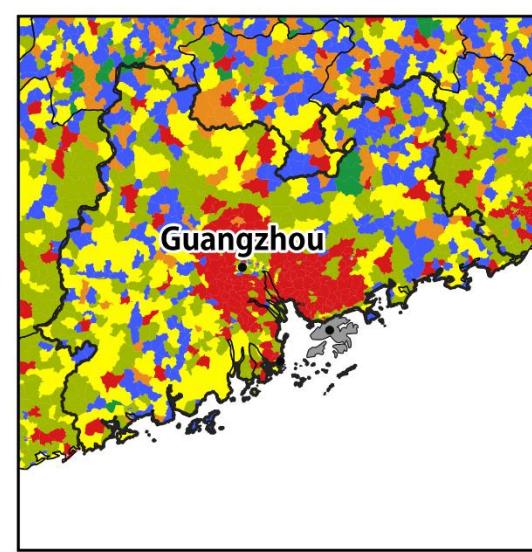
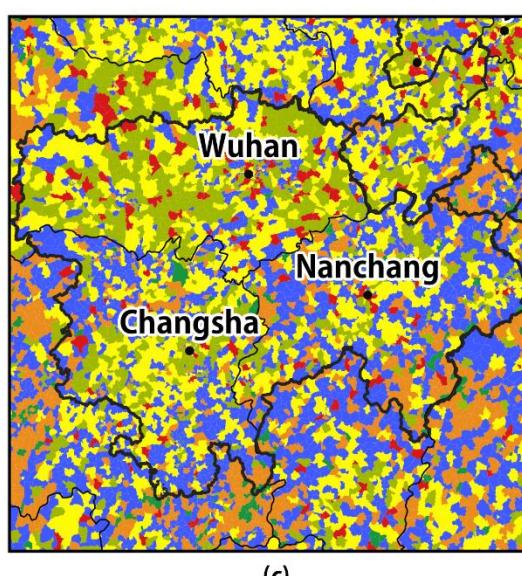
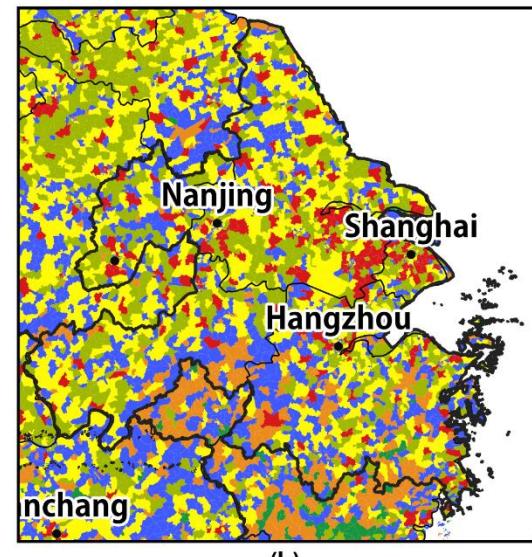
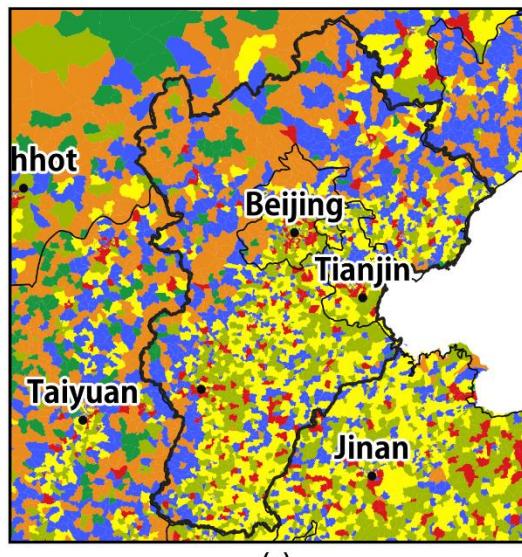
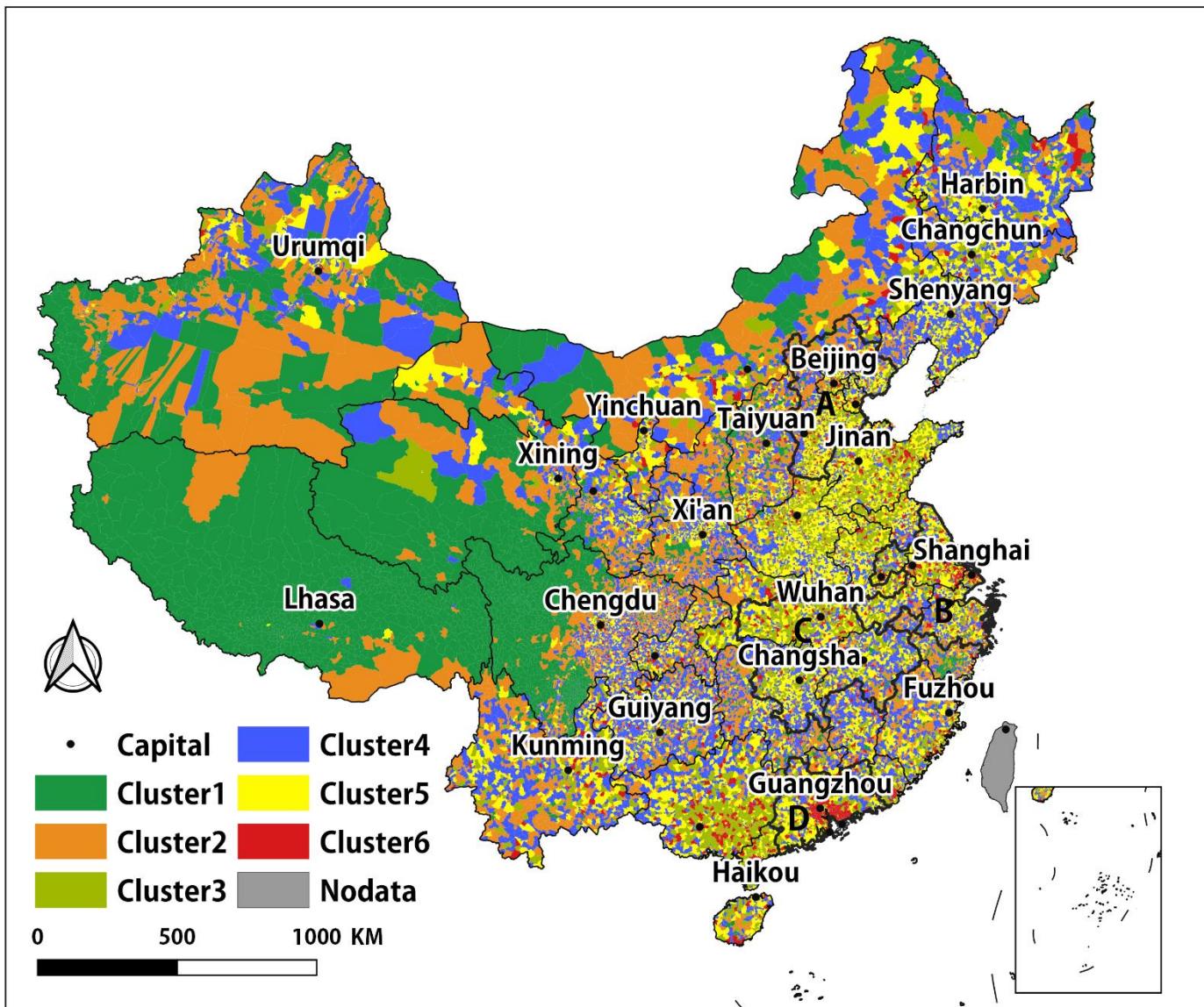
# 03 | LBS大数据与城市功能



	Cluster1	Cluster2	Cluster3
TR	0.974	1.384	1.495
LT	1.578	1.880	1.940
CF	0.968	1.134	1.119
ET	1.147	1.366	1.340
TS	0.046	0.167	0.661
RC	0.735	1.522	1.981
GA	2.083	1.868	1.646
TF	1.663	1.673	1.527
FS	1.618	2.091	2.102
RE	0.493	0.673	0.692

Note: TR=Transport LT=leisure tourism CF=communal facility  
 ET=enterprise RC=residential community TS=train station and airport  
 GA=government agency TF=technical facility FS= financial service RE=recreation

# 03 | LBS大数据与城市功能



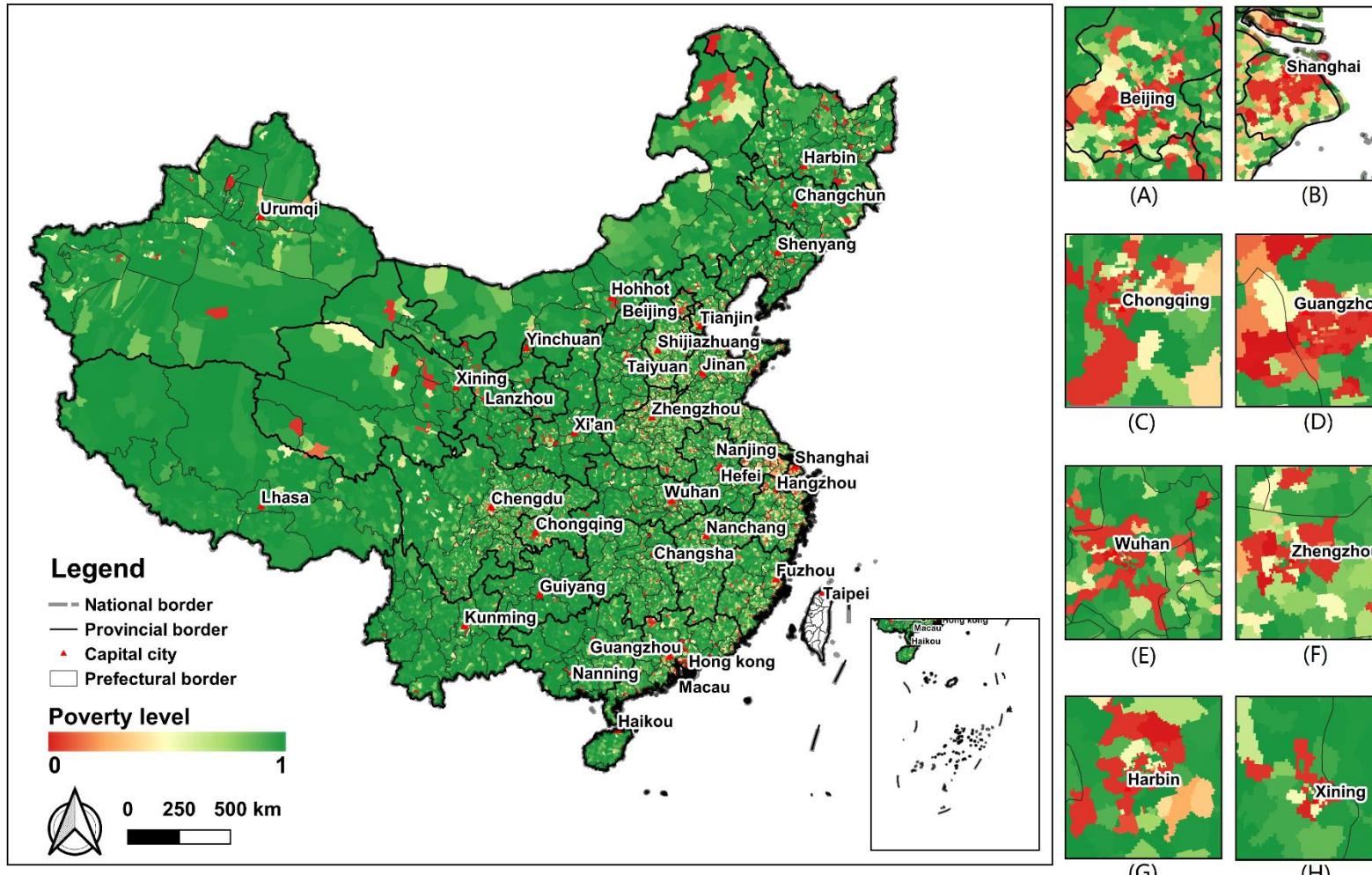


# 主要内容



- 1 LBS大数据简介
- 2 LBS大数据关键技术
- 3 LBS大数据与城市功能
- 4 LBS大数据与贫困

## ■ 乡镇尺度中国贫困区域分布



**富裕区域**

渤海湾城市群

长江三角洲

珠江三角洲

.....

**贫困区域**

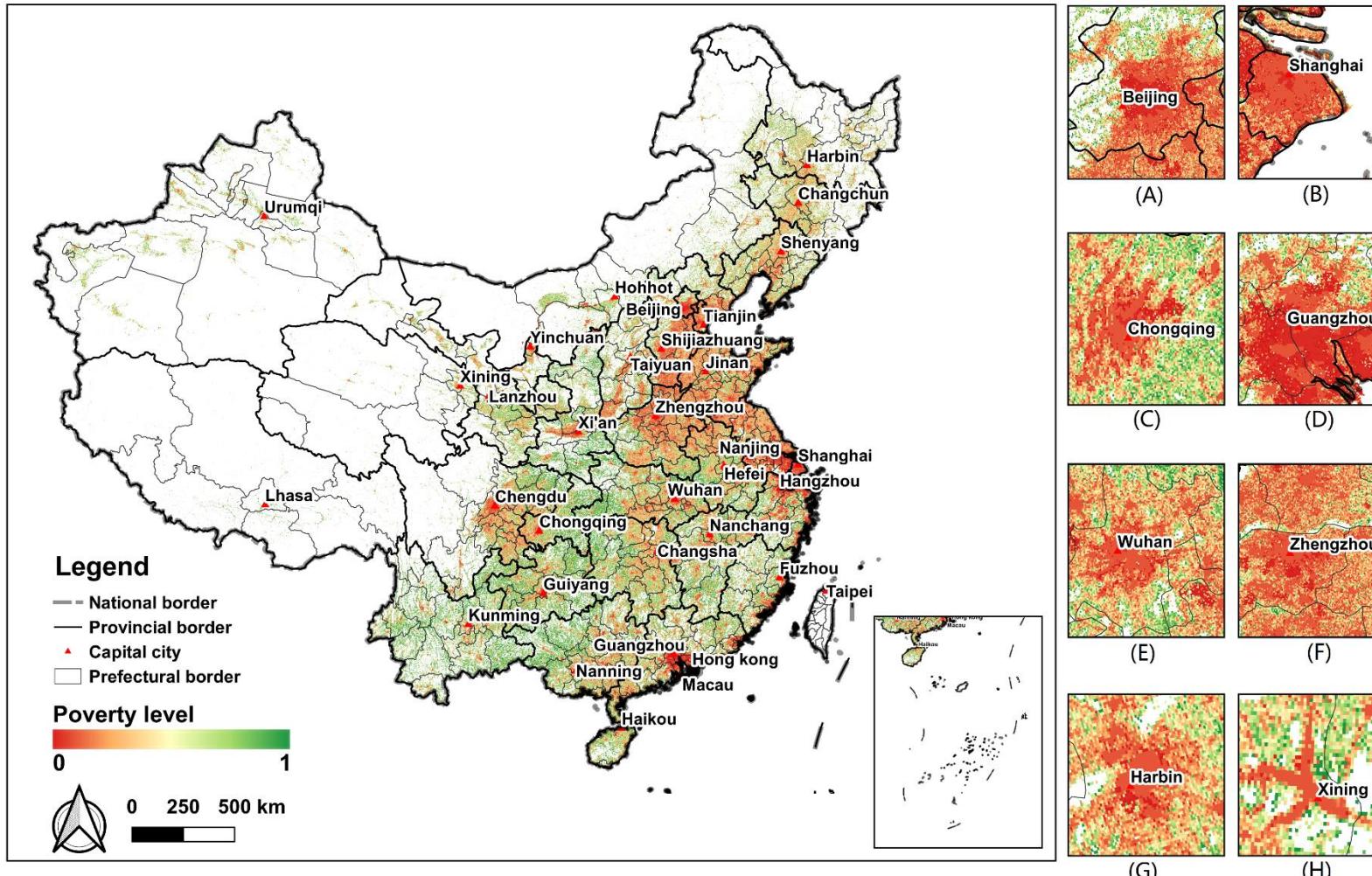
云贵高原

武夷山区

山西吕梁

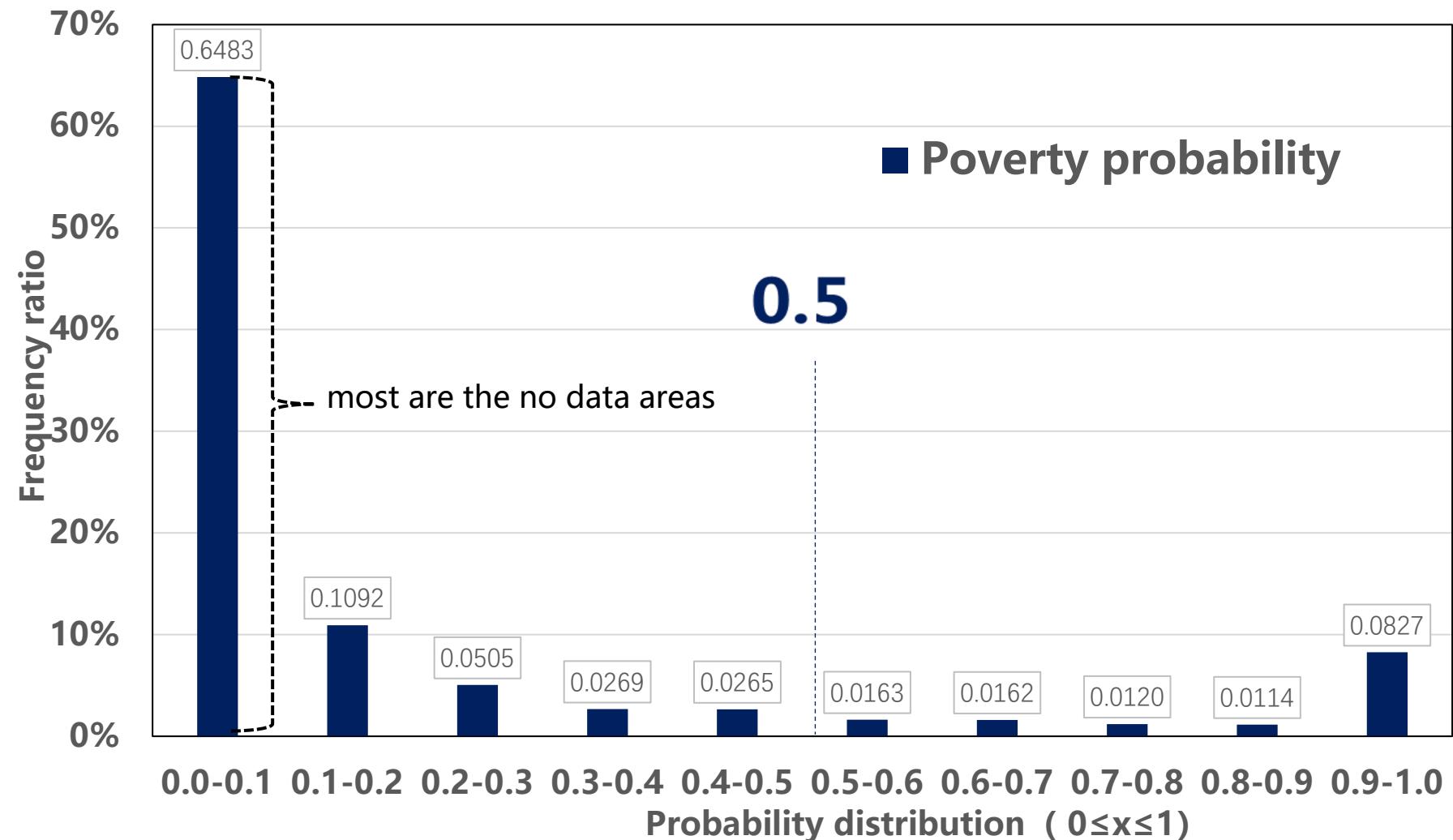
.....

## ■ 精细尺度中国贫困区域分布



- 与县级结果保持高度一致
- 精细尺度分布制图( 分辨率 1 km )
- 清晰的城市空间结构和范围
- 准确的贫困空间统计和分析

## ■ 贫困概率分布图

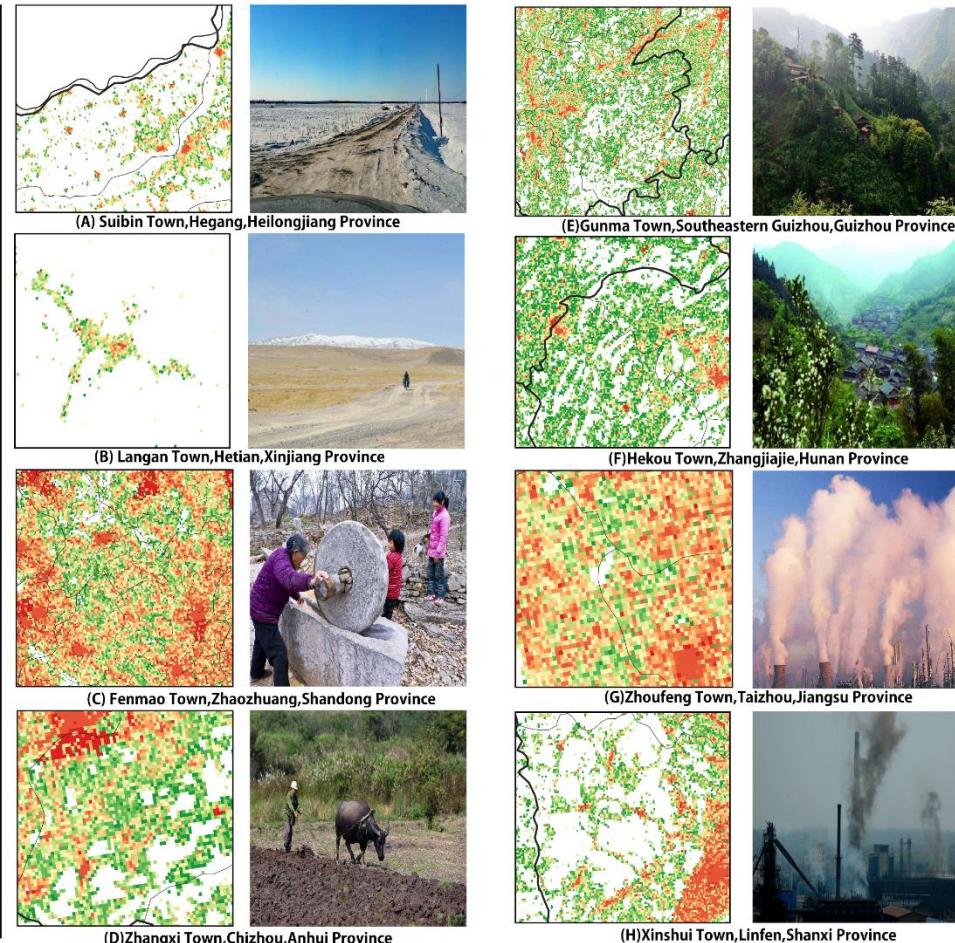
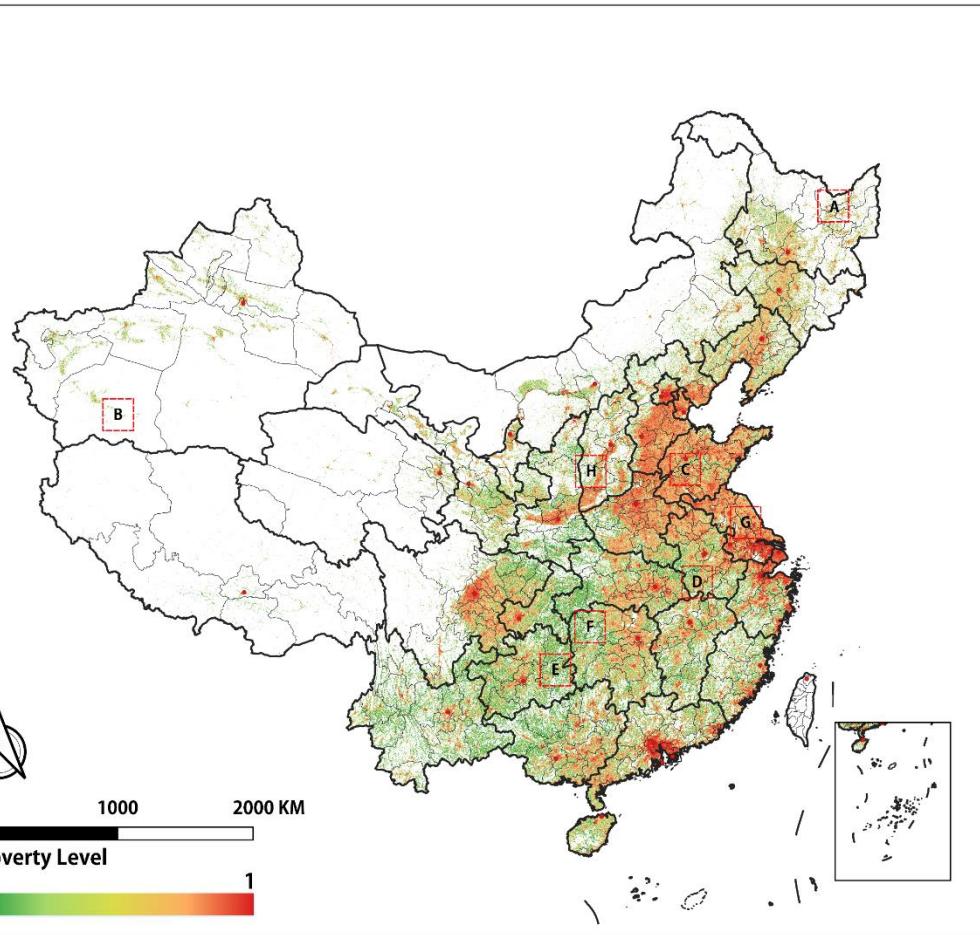


86.14% 的中国乡镇贫困概率  
低于0.5

(除去无数据区域这一比例为  
21.31%)

64.83% 的中国乡镇处于极度  
贫困 (绝大多数为无数据区域)

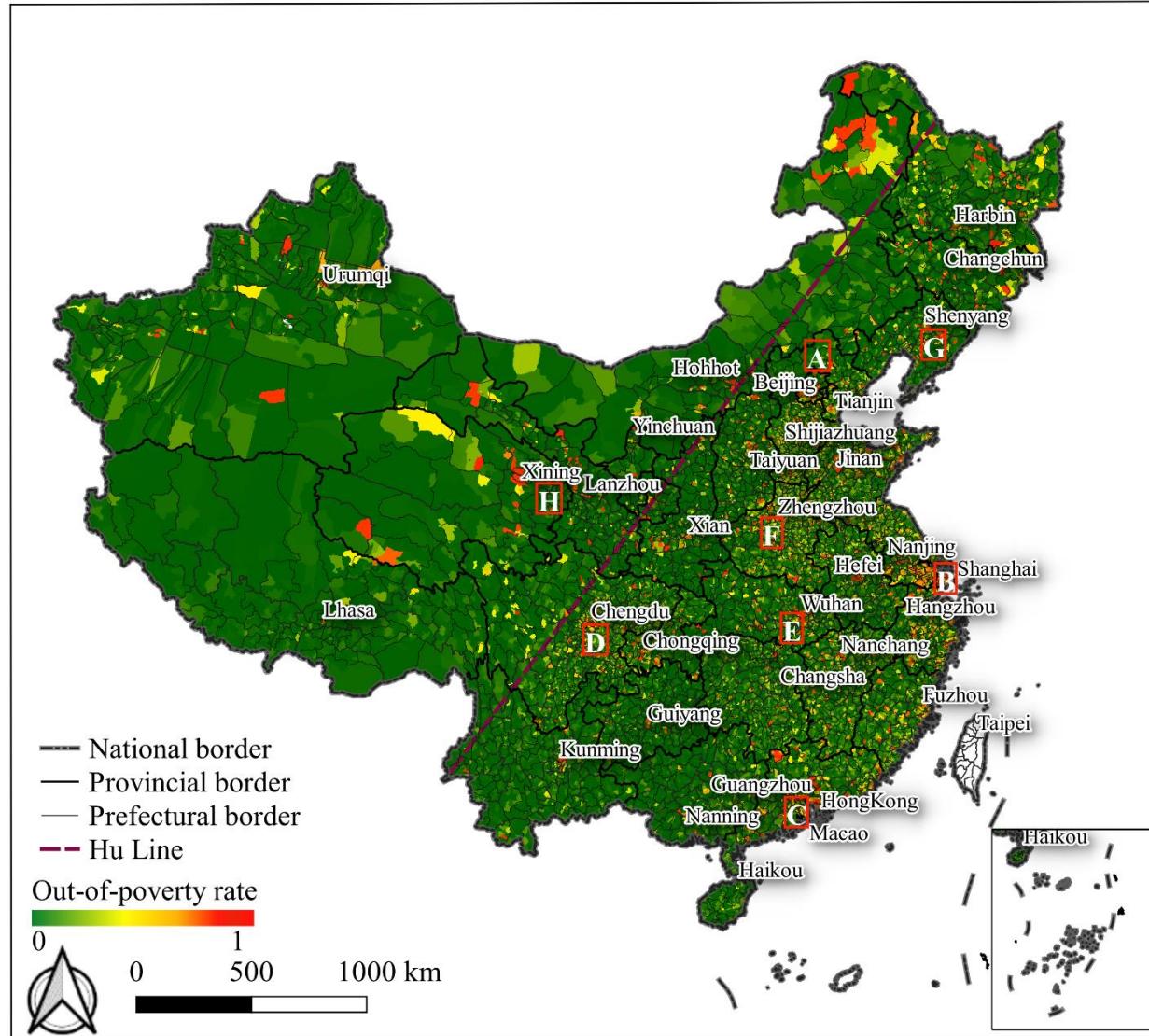
# 04 | LBS大数据与贫困



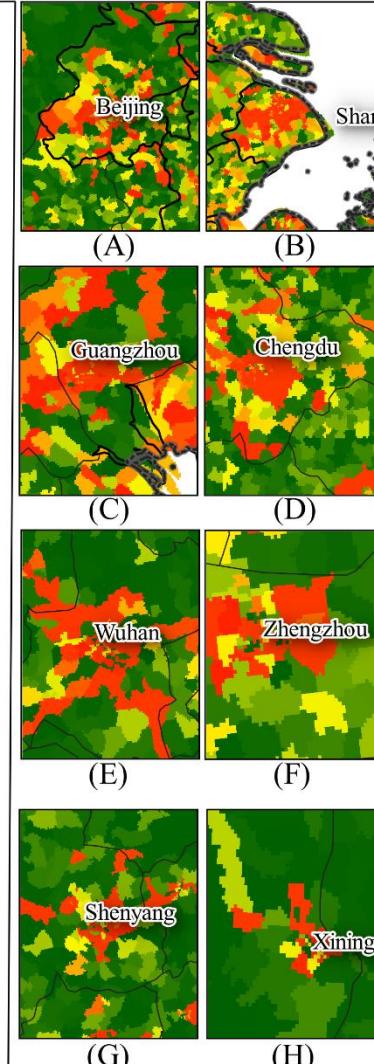
## 主要贫困类型

- 恶劣的自然环境
- 交通基础设施缺乏
- 劳动力流失
- 环境污染

# 04 | LBS大数据与扶贫效率



2016年乡镇街道尺度脱贫率空间分布



## 脱贫率低地区

云贵高原、青藏高原、  
吕梁山区、武夷山区、  
陕甘宁边区、南疆地区

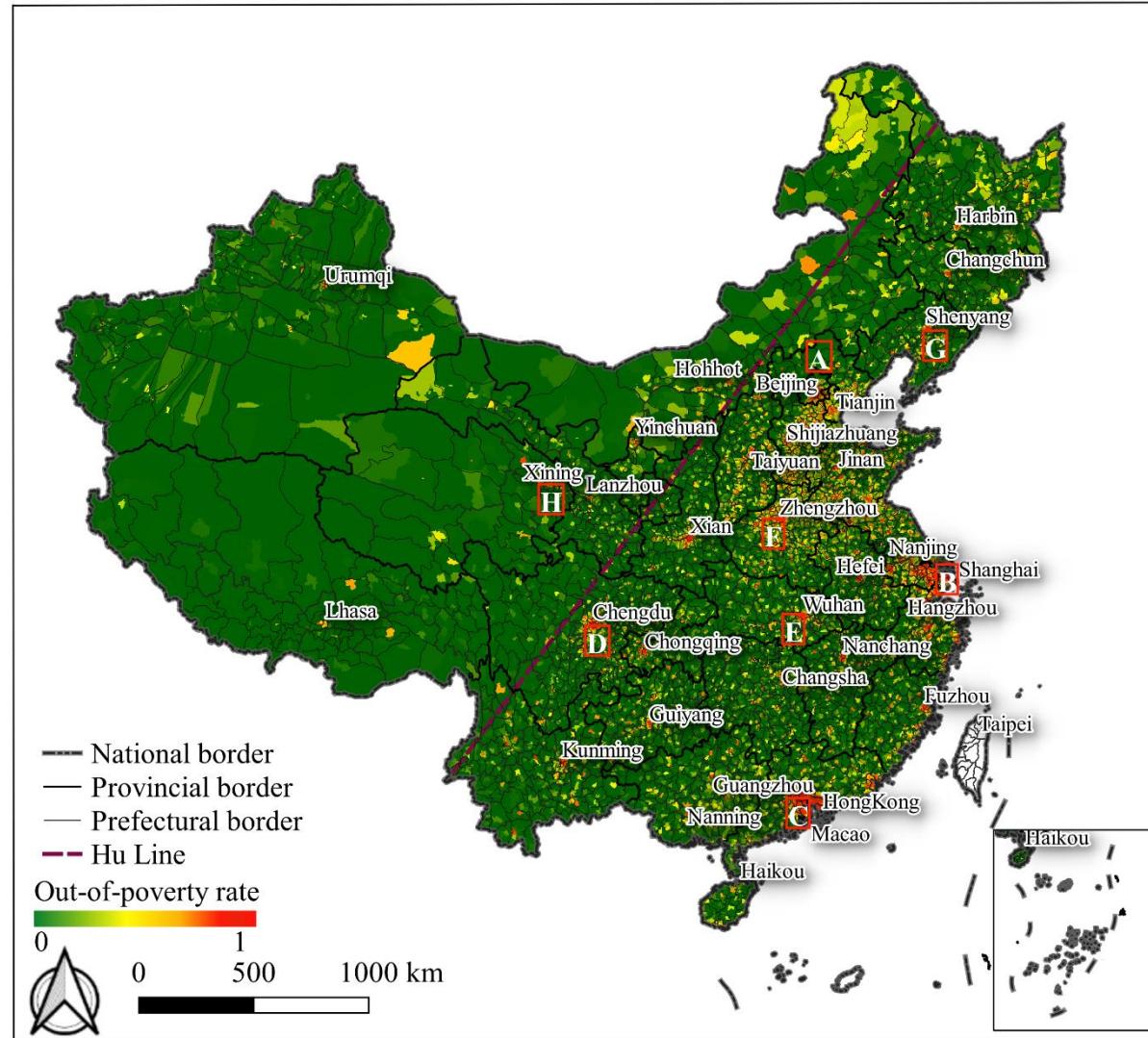
.....

## 脱贫率高地区

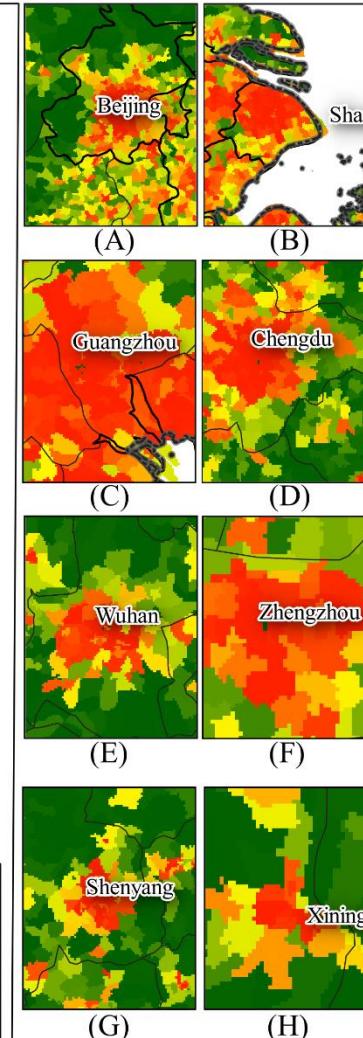
长三角、环渤海城市群、  
珠三角、华北平原北部、  
东南沿海、四川盆地腹地

.....

# 04 | LBS大数据与扶贫效率



2019年乡镇街道尺度脱贫率空间分布

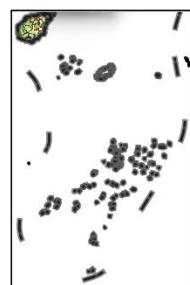
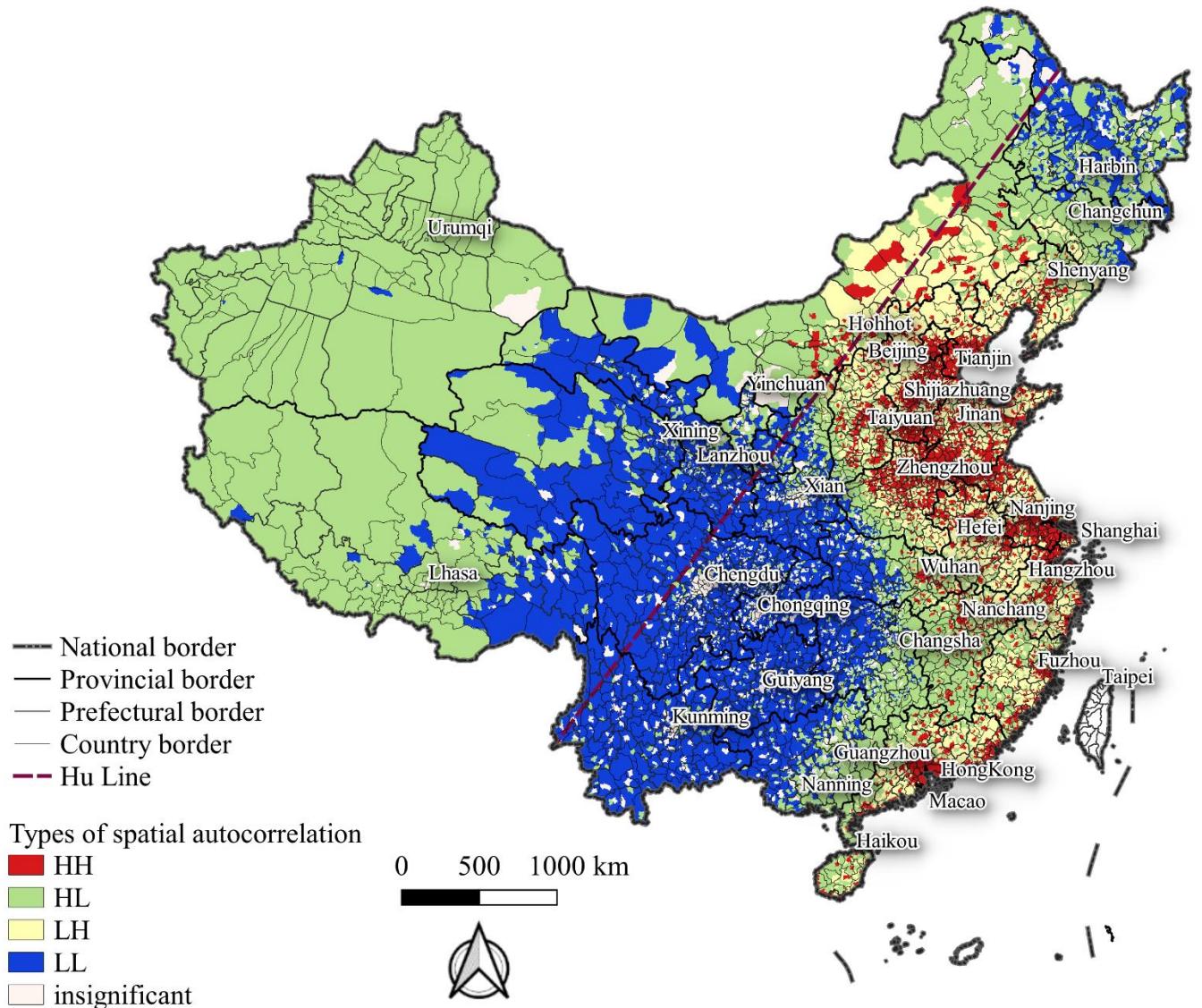


较2016年

主要城市群和中心城市  
脱贫率更高  
偏远贫困地区“更贫困”  
(可能与城镇化加速人口和生产要素的迁移和聚集有关)

中原地区、长三角地区、珠三角地区高脱贫率尤为明显

# 04 | LBS大数据与扶贫效率



自相关类型	数量	占比/%
高-高	8,896	21.559
高-低	3,365	8.155
低-高	6,941	16.821
低-低	11,293	27.368
不显著	10,768	26.096
总计	41,263	100.000



# 本章总结

本章介绍了LBS大数据的数据类型、数据处理过程、处理结果以及LBS大数据关键技术以及其在城市人口、城市功能、贫困等方面的应用。

LBS大数据主要形式为移动手机APP数据。

LBS大数据在建筑物尺度人口识别、建筑物功能分类、城市职能分类、贫困区域识别、扶贫效率分析等研究领域中都有应用。

LBS大数据在深入理解城市结构、定量分析城市社会经济情况等方面有广泛的应用前景，LBS大数据同样可以作为其他研究的辅助数据，提供城市内的社会经济信息。

LBS大数据对深入理解城市社会经济情况提供了帮助，为城市规划、评估城市发展状况等提供决策支持。