

城市空间数据分析方法

Research and Develop Project Scenarios Using the Urban Spatial
Data Analysis Method

Landscape Architecture

宝 瑞清
Ruiqing Bao



- 1. Python在设计领域**
01_Python在设计领域
- 2. python解释器**
python解释器
- 3. 列表的知识结构-A/百度POI数据采集**
百度POI数据采集
列表的知识结构-A
- 4. 列表的知识结构-B/建筑外环境测量数据可视化**
建筑外环境测量数据可视化
列表的知识结构-B
- 5. 元组/基于GPS调研与数据读取**
基于GPS调研与数据读取
基于GPS调研与数据读取
元组
- 6. 字典/异常值处理初步**
异常值处理初步
字典
- 7. numpy-A/城市空间-色彩-A**
城市空间-色彩-A-1
城市空间-色彩-A-2
numpy-A
- 8. numpy-B/基于百度POI数据生成.kml文件与描述性统计**
数据采集+数据格式
.kml文件生成
描述性统计
numpy-B
- 9. 循环语句/机器学习-聚类-城市色彩-B-印象**
机器学习概述-A
示例数据集训练/学习与预测+模型保存(参数更新+类别数据
聚类
城市色彩-印象
itertools库+warnings库
基本语句-循环
- 10. 协方差|卡方检验/poi空间分布结构**
协方差/卡方检验
批量采集+poi分布A-DBSCAN密度聚类
poi空间分布结构B
基本语句-条件
- 11. 函数|GDAL/OGR/SIR模型与卷积|森林的蔓延**
卷积
- SIR模型**
森林的蔓延(
- 12. pandas-A|基于RF高程预测地形生成(参数化结合)与回归预测**
回归预测
地形生成<随机森林回归模型|参数化>
pandas-A
- 13. pandas-B|基于回归预测NDVI修复地生态廊道的构建**
修复地生态廊道构建
pandas-B
- 14. pandas-C|基于DBSCAN密度空间聚类林缘线生成设计**
DBSCAN密度空间聚类林缘线生成设计
pandas-C
- 15. pycharm/虚拟环境/SQLite数据库|将规划设计信息写入数据库**
景观评估项目-实验:技术路线
pycharm|虚拟环境搭建
数据库
SQLite数据库
- 16. flask构建实验用网络应用平台|Jinja2模板**
flask轻量级 Web 应用框架
Jinja2模板
阶段B-目录结构
- 17. _class类|SIFT|BOW|ERF|景观质量视觉评估预测**
图像特征提取
BOW视觉词袋模型
ERF图像分类器
-图像识别器
class类|OPP
- 18. ubuntu|VNC|Nginx|uWSGI/实验用网络应用平台的部署**
实验用网络应用平台的部署
- 19. ROS-Kinetic深度相机3DSLAM三维点云建筑空间**
点云/三维模型
rtabmap|ROS
- 20. NLP_LDA主题建模_提取主题关键字**
实验素材/与计算结果
NLP-自然语言处理
有用的库方法
- 21. 提取.jpg(.png)图像 (热力图) 格式的数据为地理信息数据**
.jpg图像格式数据预处理
- 色彩空间**
.jpg数据提取
- 22. 生活圈_01_连续聚类与数据分析和数据保存|collections 库**
读取与组织POI数据, 为skleran的bunch存储方式
使用DBSCAN聚类POI数据
将聚类结果存储为.shp的点文件格式, 字段包括“poi”存储行业列表信息, “cluster”存储聚类簇信息
计算POI的业态空间结构分布并存储为图片格式文件
数据统计分析(箱型图/小提琴图/折线图)
存储计算数据/Numpy数据的保存与读取
在GIS中编写python脚本定义.shp文件投影, 并转换为依据距离为单元大小的栅格数据
numpy的ravel() 和 flatten()函数
collections 库
缩放带有outlier异常值的数据
- 23. 生活圈_02_通过计算曲线拐点找到特征层级|kneed库**
kneed库
寻找曲线拐点
box plot 箱型图/violin plot 小提琴图
pandas读取excel(.xlsx)数据
- 24. 生活圈_03_信息熵与均衡度**
信息熵与均衡度公式
计算信息熵/均衡度
均衡度与距离距离的关系_箱型图
如何在GitHub上下载单独的文件
- 25. 生活圈_04_相关系数热力图(簇行业类)与批量图片自动排版**
相关系数热力图(簇行业类)
批量图片自动排版
numpy.transpose()
- 26-30. 城市热环境_01-05_基于LST**
(26)01_LST反演地表温度
(27)02_基于地表温度差值数量变化界定全区植被降温效应的空间分布
(28)03_基于卷积温度梯度变化界定冷区和热区的空间分布结构
(29)04_地表覆盖与精度评价
(30)05_基于机器学习回归算法建立用于绿地规划评估的地表温度预测模型

1. Python Language
2. Baidu Maps POI Capture and .csv File Read and Data Display
3. Environmental Measurement and Data Diagram outside Buildings
4. Read GPS-based Research Data
5. <outliers> Outlier Processing
6. Colour Space and Distribution of Images
7. Urban Colour Impressions and Machine Learning/Cluster
8. Descriptive Statistics and .kml Generation Program of Baidu Maps POI
9. Living Circle——POI Data Cluster, Space Distribution Structure and Its Information Entropy at Different Distances
10. Spread, Convolution and SIR Models of Forest
11. Terrain Generation (Combined with Parameterization) based on RF Elevation Prediction and Regression Prediction
12. Construction of Ecological Corridor in Rehabilitated Place based on Regression Prediction NDVI
13. DBSCAN-based Forest Edge Generation Design
14. pycharm |Virtual Environment|SQLite Database
15. Flask-based Development of Network Application Platform for Experiment
16. Visual Landscape Quality Evaluation (Machine Learning Prediction)
17. ROS-Kinetic Depth Camera|3DSLAM Point-cloud Architectural Space
18. NLP-Topic Modelling_Find Important Contents
19. Generation of Terrain Selection Schemes (based on Generative Adversarial Network)
20. Temporal-spatial Data Analysis and Thermodynamic Chart Data Extraction
21. Hourly Meteorological Data Processing and Spatial Interpolation-Urban Climate Analysis
22. Geographic Informatization of Wind Simulation Data and Urban Wind Environmental Prediction Modelling
23. Cluster of Vegetation Coverage and Ecological Network Building

Study_1_Spatial Distribution Structure of Urban Life Circle Under Continuous Distance Clustering Hierarchy Based on Baidu POI_{Points-Of-Interest}

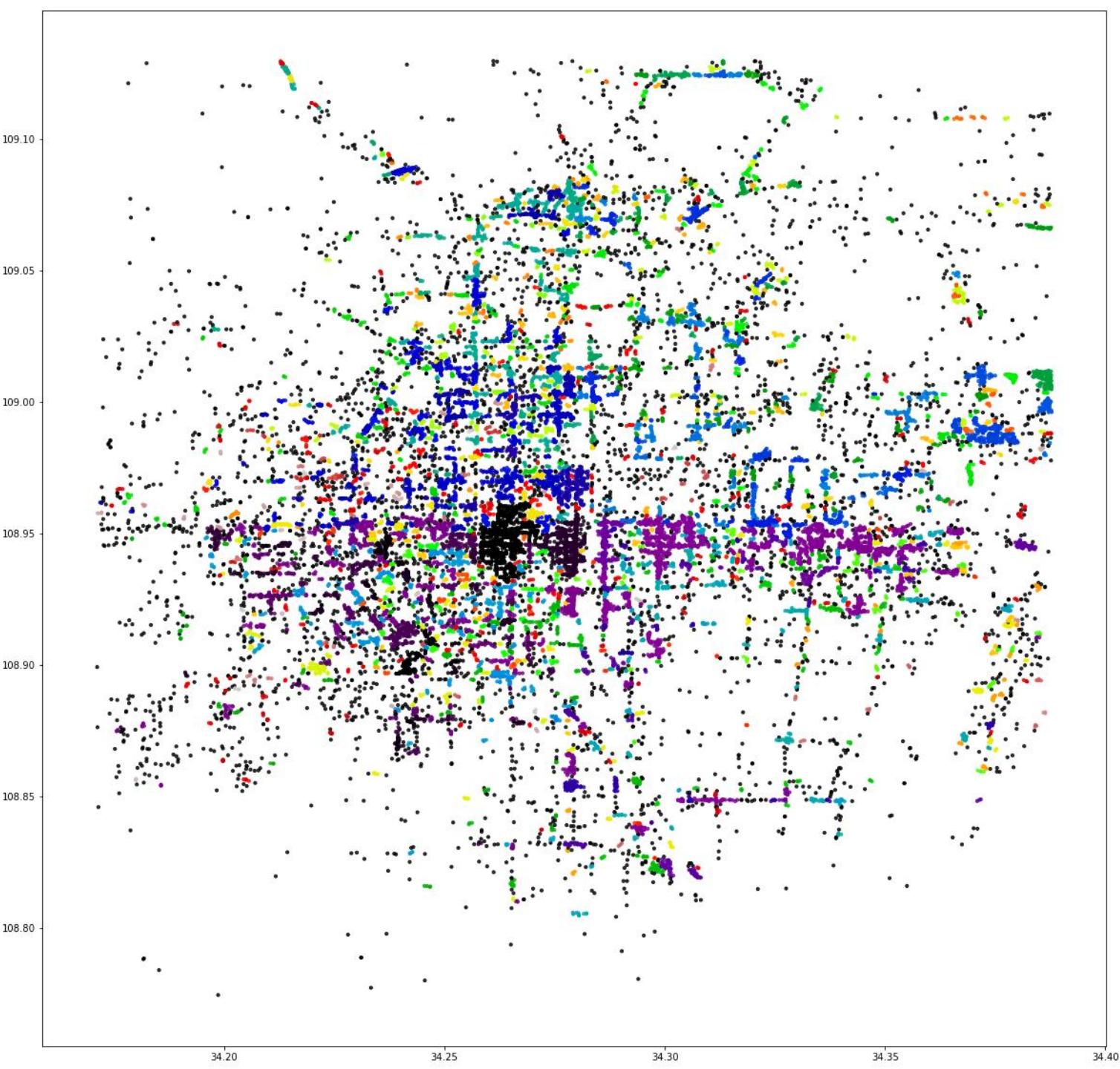
基于百度POI连续距离聚类层级下城市生活圈的空间分布结构

- 数据/Data
- 方法/Method
- 结果/Results
- 讨论/Discussion

• 数据/data

 poi_5_spot.csv

钟
楼,108.94237005134782,34.26105005572973,1bf85b2519cd8ea5a1a93414,"{'name': '钟楼',
'location': {'lat': 34.265725, 'lng': 108.95346},
'address': '东西南北四条大街交汇处', 'street_id':
'1bf85b2519cd8ea5a1a93414', 'telephone':
'(029)87214665', 'detail': 1, 'uid':
'1bf85b2519cd8ea5a1a93414', 'detail_info': {'tag':
旅游景点;风景区, 'navi_location': {'lng':
108.95347392631, 'lat': 34.265274911505}, 'type':
'scope', 'detail_url':
http://api.map.baidu.com/place/detail?uid=1bf85b2519cd8ea5a1a93414&output=html&source=placeapi_v2, 'price': '70', 'overall_rating': '4.7',
'comment_num': '500'}}"



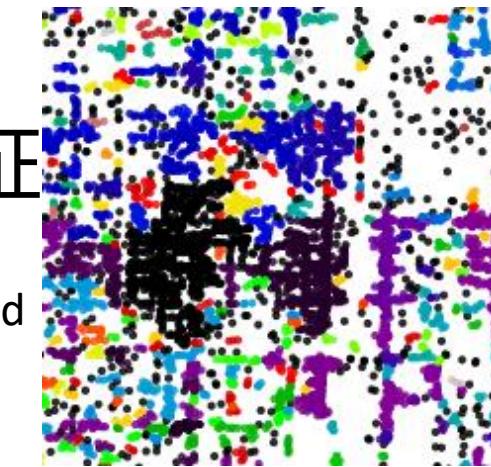
• 数据/data

一级行业分类 Level 1 industry classification		二级行业分类 Level 2 industry classification
美食 food	中餐厅、外国餐厅、小吃快餐店、蛋糕甜品店、咖啡厅、茶座、酒吧	
酒店 hotel	星级酒店、快捷酒店、公寓式酒店	
购物 shopping	购物中心、百货商场、超市、便利店、家居建材、家电数码、商铺、集市	
生活服务 life service	通讯营业厅、邮局、物流公司、售票处、洗衣店、图文快印店、照相馆、房产中介机构、公用事业、维修点、家政服务、殡葬服务、彩票销售点、宠物服务、报刊亭、公共厕所	
丽人 beauty	美容、美发、美甲、美体	
旅游景点 tourist attractions	公园、动物园、植物园、游乐园、博物馆、水族馆、海滨浴场、文物古迹、教堂、风景区	
休闲娱乐 entertainment	度假村、农家院、电影院、KTV、剧院、歌舞厅、网吧、游戏场所、洗浴按摩、休闲广场	
运动健身 sports fitness	体育场馆、极限运动场所、健身中心	
教育培训 education and training	高等院校、中学、小学、幼儿园、成人教育、亲子教育、特殊教育学校、留学中介机构、科研机构、培训机构、图书馆、科技馆	
文化传媒 culture media	新闻出版、广播电视台、艺术团体、美术馆、展览馆、文化宫	
医疗 medical	综合医院、专科医院、诊所、药店、体检机构、疗养院、急救中心、疾控中心	
汽车服务 auto service	汽车销售、汽车维修、汽车美容、汽车配件、汽车租赁、汽车检测场	
交通设施 traffic facilities	飞机场、火车站、地铁站、长途汽车站、公交车站、港口、停车场、加油加气站、服务区、收费站、桥、充电站、路侧停车位	
金融 financial	银行、ATM、信用社、投资理财、典当行	
房地产 real estate	写字楼、住宅区、宿舍	
公司企业 corporation	公司、园区、农林园艺、厂矿	
政府机构 government agency	中央机构、各级政府、行政单位、公检法机构、涉外机构、党派团体、福利机构、政治教育机构	
出入口 entrance	高速公路出口、高速公路入口、机场出口、机场入口、车站出口、车站入口、门（备注：建筑物和建筑物群的门）、停车场出入口	
自然地物 natural features	岛屿、山峰、水系	

• 方法/Method

1 study the spatial hierarchy variation character of urban life circle based on POI location information.

基于POI位置信息研究城市生活圈的空间层级变化特征



DBSCAN (Density-Based Spatial Clustering of Applications with Noise.) /Clustering method

Set the distance range (20, 510) meter, calculated every 10 meters, totally 50 times.

2 study the spatial industry/amenities structure of urban life circle based on POI first-level industry classification.

基于POI一级行业分类，研究城市生活圈的业态结构

function

协方差逆矩阵

covariance.GraphicalLassoCV() + affinity_propagation()AP

聚类算法

均质性/同质性

信息熵

$$H = - \sum_{i=1}^N P_i \log P_i$$

To analysis the industry structure **homogeneity** of urban life circle using **entropy** and **equilibrium degree** formulas.

$$J = H/H_m = - \sum_i^N P_i \log P_i / \log N$$

均衡度

参考：9_循环语句/机器学习-聚类-城市色彩-B-印象
10_协方差|卡方检验/poi空间分布结构

1: 读取与组织POI数据，为skleran的bunch存储方式



2

参考“10_独立性检验(列联表)与poi空间分布结构”

2: 使用DBSCAN聚类POI数据



1

3: 将聚类结果存储为.shp的点文件格式，字段包括“poi”存储行业列表信息，“cluster”存储聚类簇信息



1

参考 “10_独立性检验(列联表)与poi空间分布结构”

4: 计算POI的业态空间结构分布并存储为图片格式文件



2

5: 数据统计分析

4

6: 存储计算数据，用于日后分析，避免重复计算，尤其较大计算量会花费较多时间。



7: 在GIS中编写python脚本定义.shp文件投影，并转换为依据距离为单元大小的栅格数据



2

8: 读取存储的数据，数据分析-通过计算曲线拐点找到特征层级

3

9: 计算信息熵与均衡度，分析业态均质性/相关系数

2

10: 自动排版海量分析图表

1

建立批量计算的循环程序

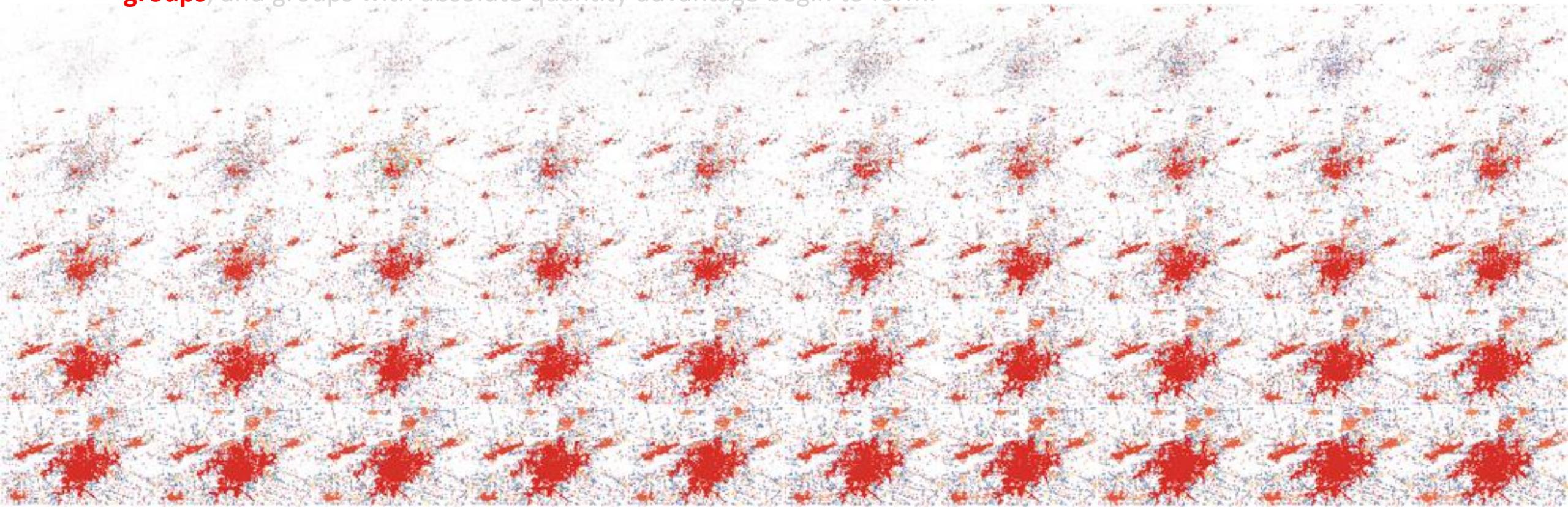


•结果/Results

1 study the spatial hierarchy variation character of urban life circle based on POI location information.

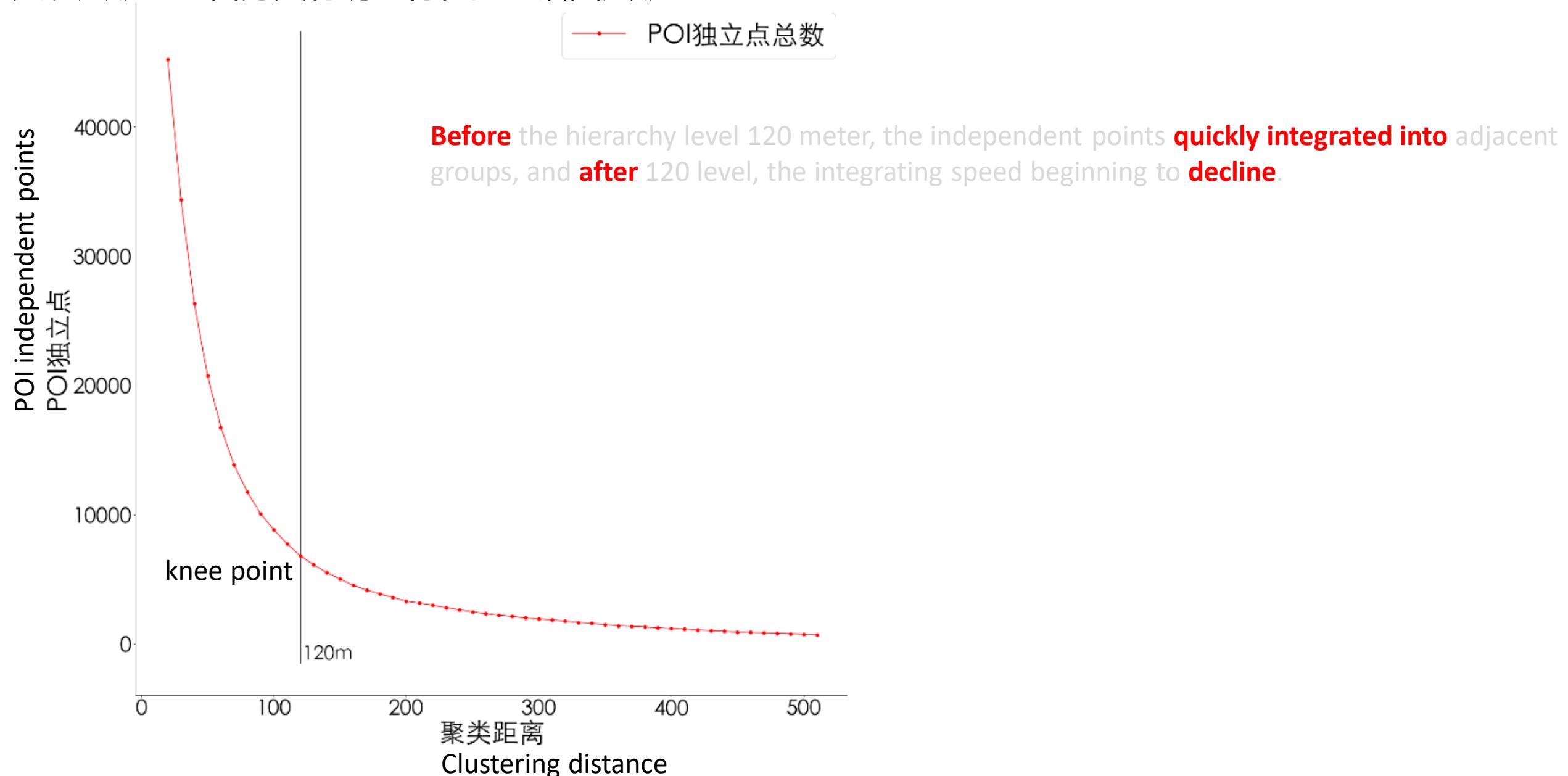
基于POI数据连续聚类，城市生活圈的空间层级变化特征

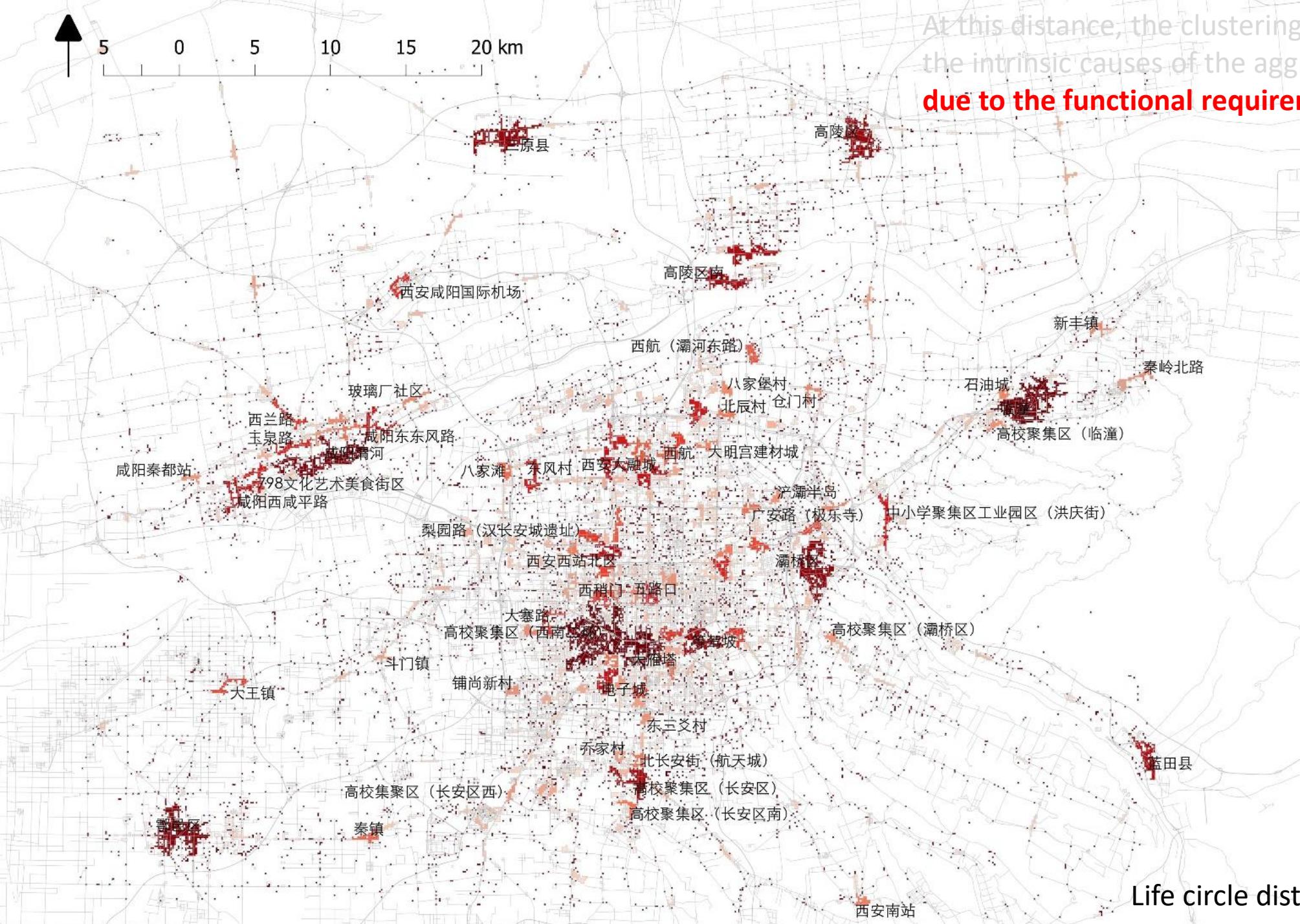
In the process of **continuous clustering** at 50 levels, adjacent regions are reunited to form **a lot of scattered groups**, and groups with absolute quantity advantage begin to form.



Critical level 1: the hierarchy of life circle with distinct characteristics

关键层级一：自身具有明显特征的生活圈层级



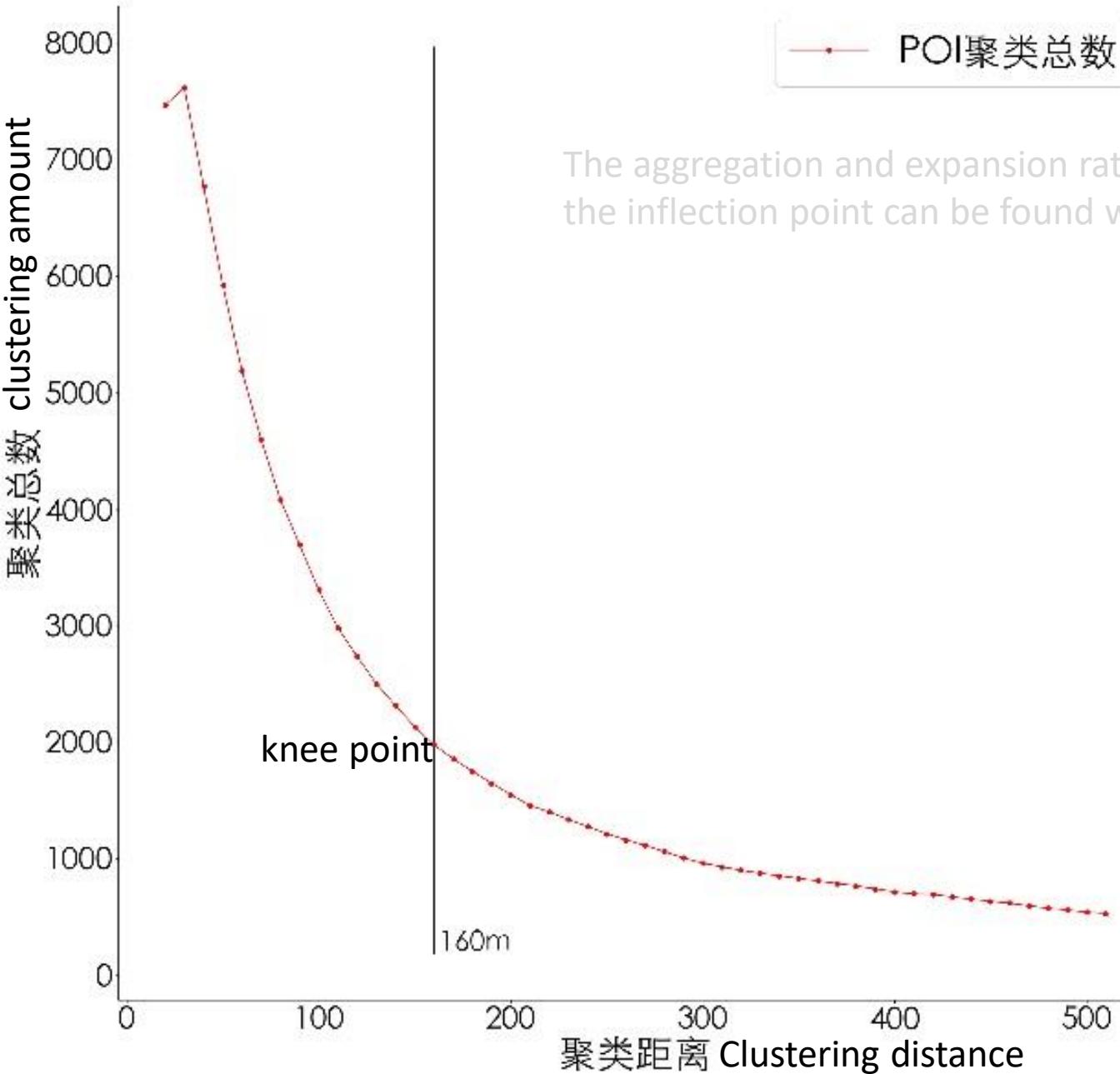


At this distance, the clustering clusters can clearly reflect the intrinsic causes of the aggregation of various groups **due to the functional requirements** of the city.

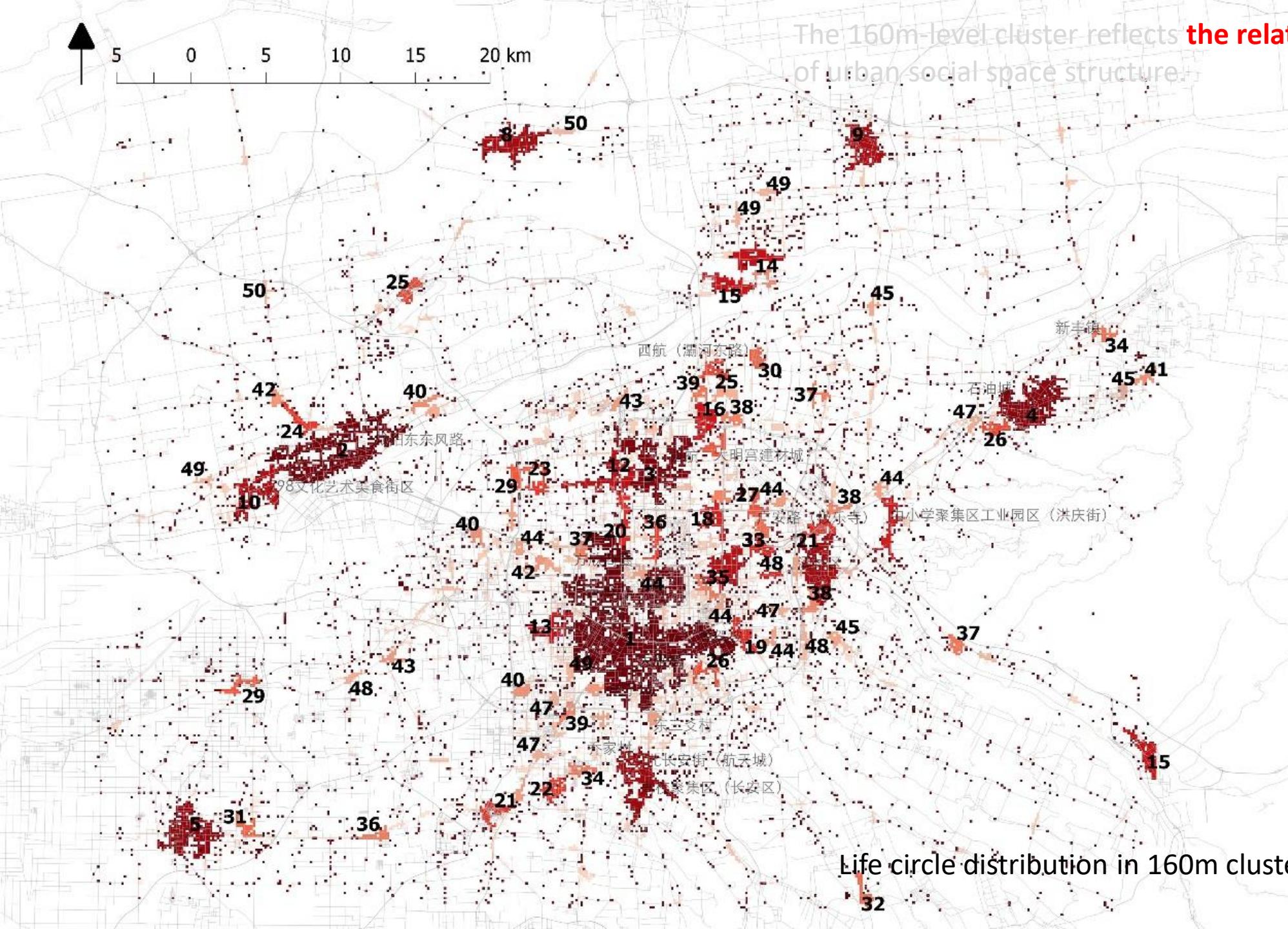
Life circle distribution in 120m clustering

Critical level 2: the urban life circle in **a relatively stable state**

关键层级二：处于相对稳定状态下的城市生活圈



The aggregation and expansion rate of a group is also a process of first fast and then slow, and the inflection point can be found which reflects the changes of a group.



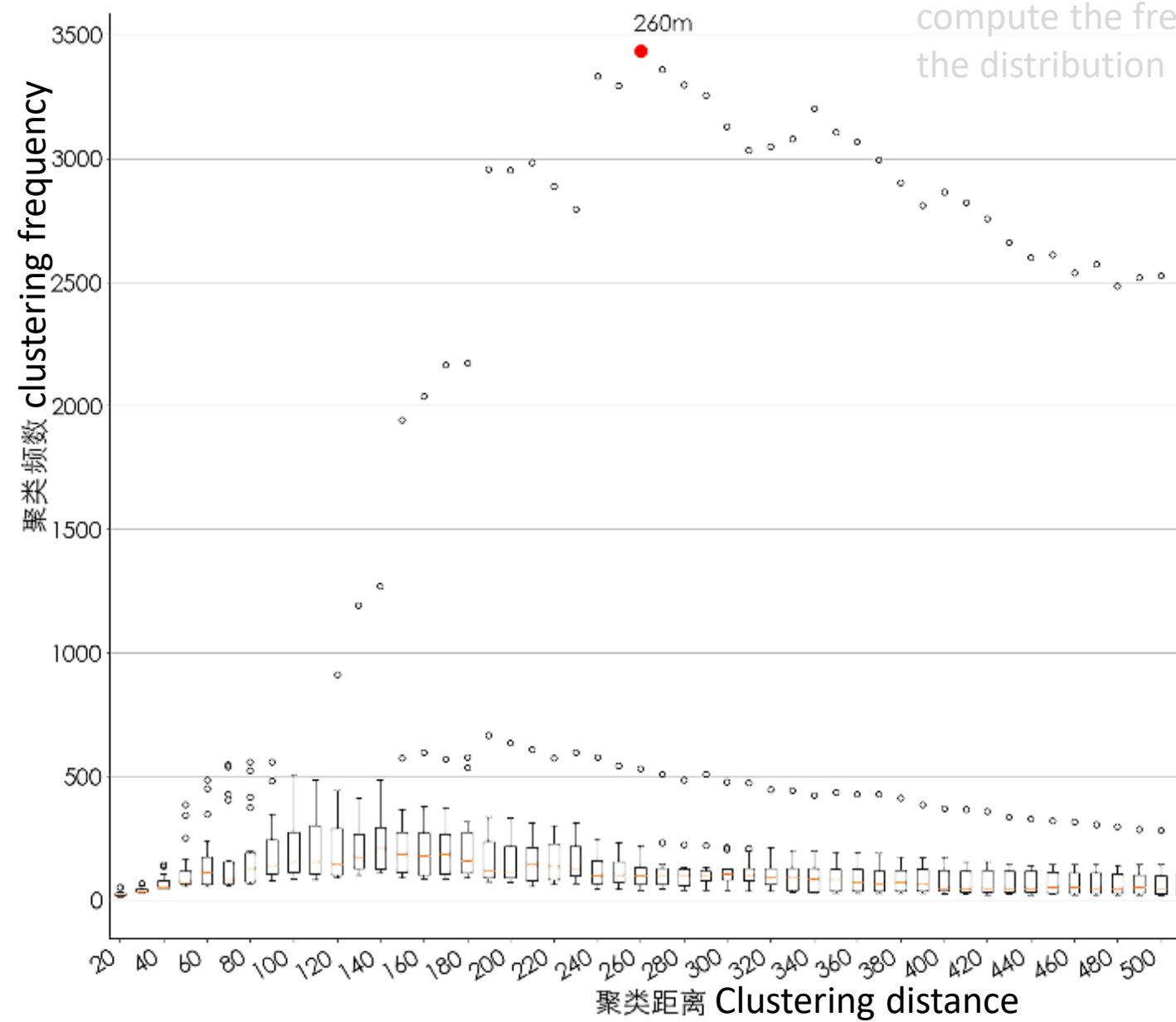
5 0 5 10 15 20 km

The 160m-level cluster reflects **the relatively stable basic pattern** of urban social space structure.

Critical level 3: the hierarchy and fringe of **the largest life circle**

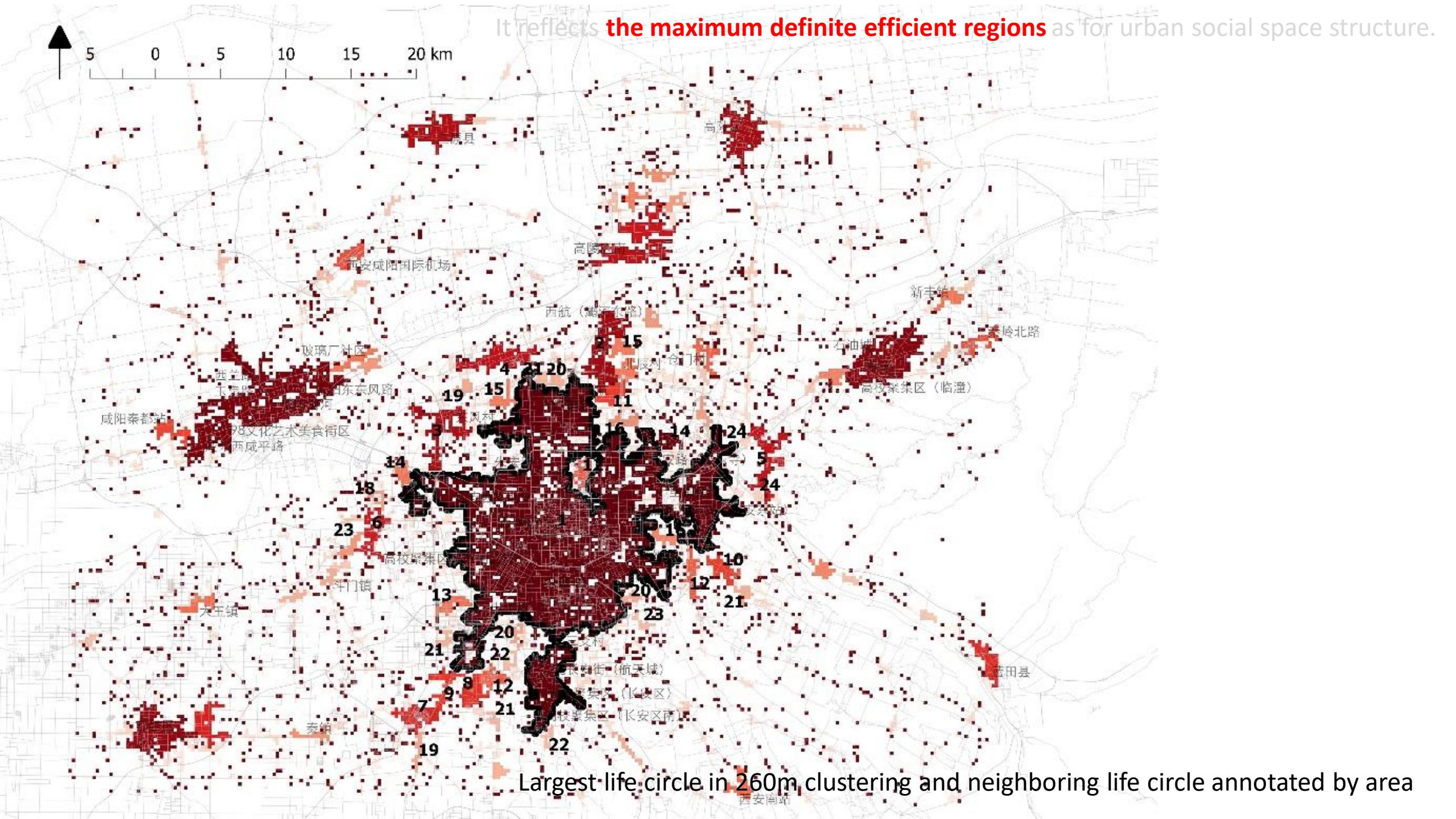
关键层级三：城市最大生活圈的层级与边缘

knee point



compute the frequency of cluster at each level of 50 layers, and analyze the distribution with a box figure

It reflects **the maximum definite efficient regions** as for urban social space structure.

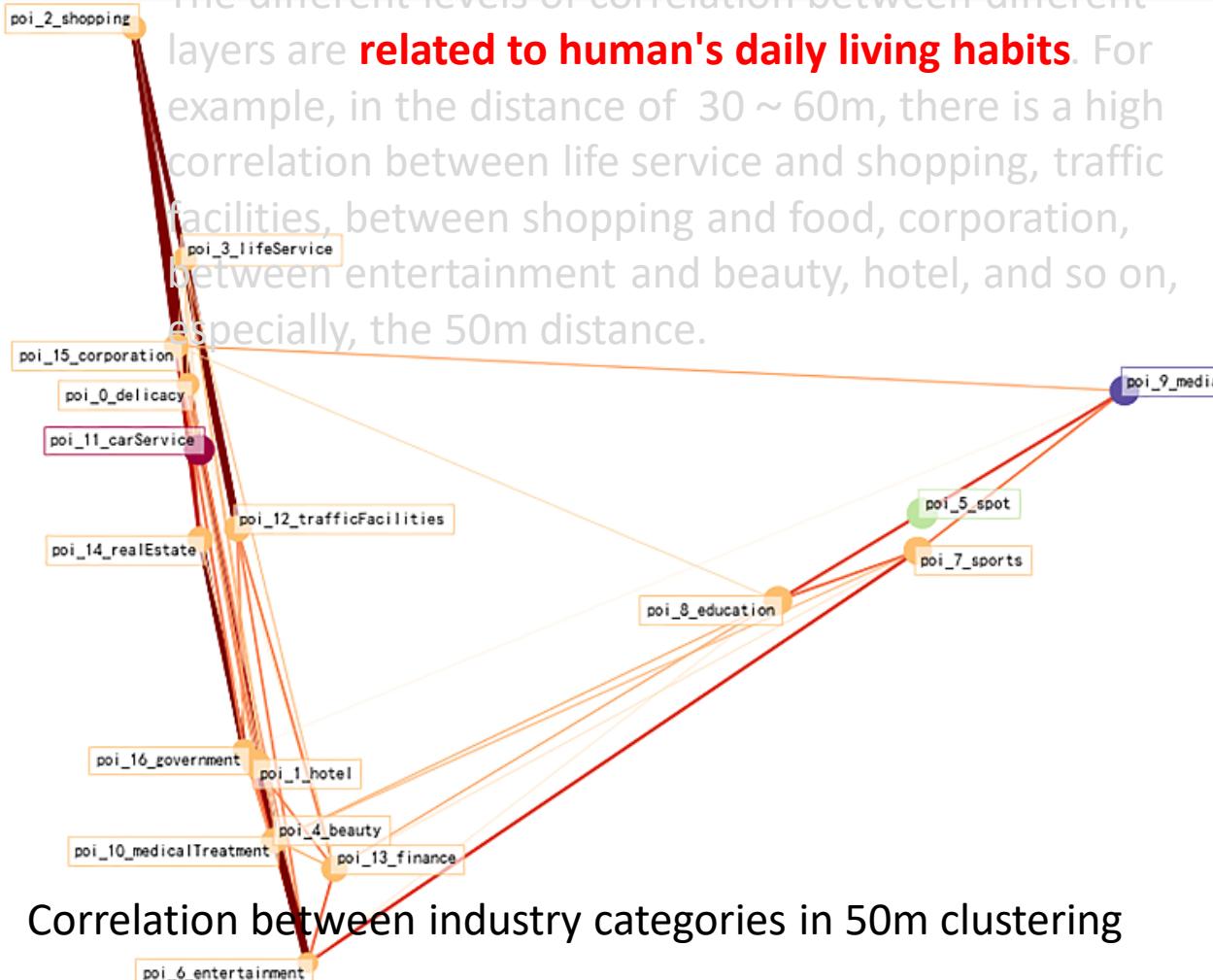


2 study the spatial industry structure of urban life circle based on POI first-level industry classification.

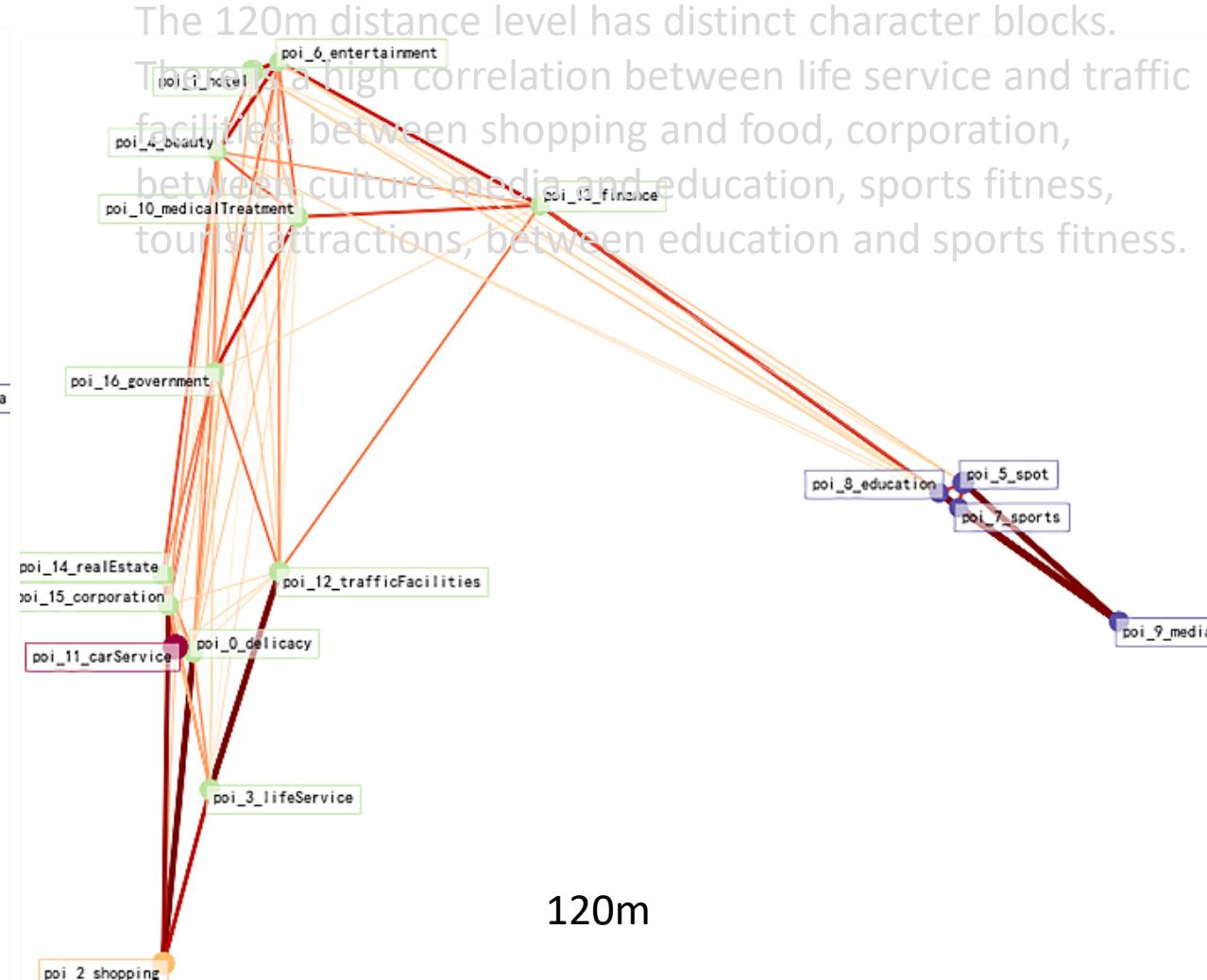
生活圈的业态结构空间变化特征

industry categories correlation and hierarchy effect 行业类的相关性与层级效应

The different levels of correlation between different layers are **related to human's daily living habits**. For example, in the distance of 30 ~ 60m, there is a high correlation between life service and shopping, traffic facilities, between shopping and food, corporation, between entertainment and beauty, hotel, and so on, especially, the 50m distance.



The 120m distance level has distinct character blocks. There is a high correlation between life service and traffic facilities, between shopping and food, corporation, between culture media and education, sports fitness, tourist attractions, between education and sports fitness.



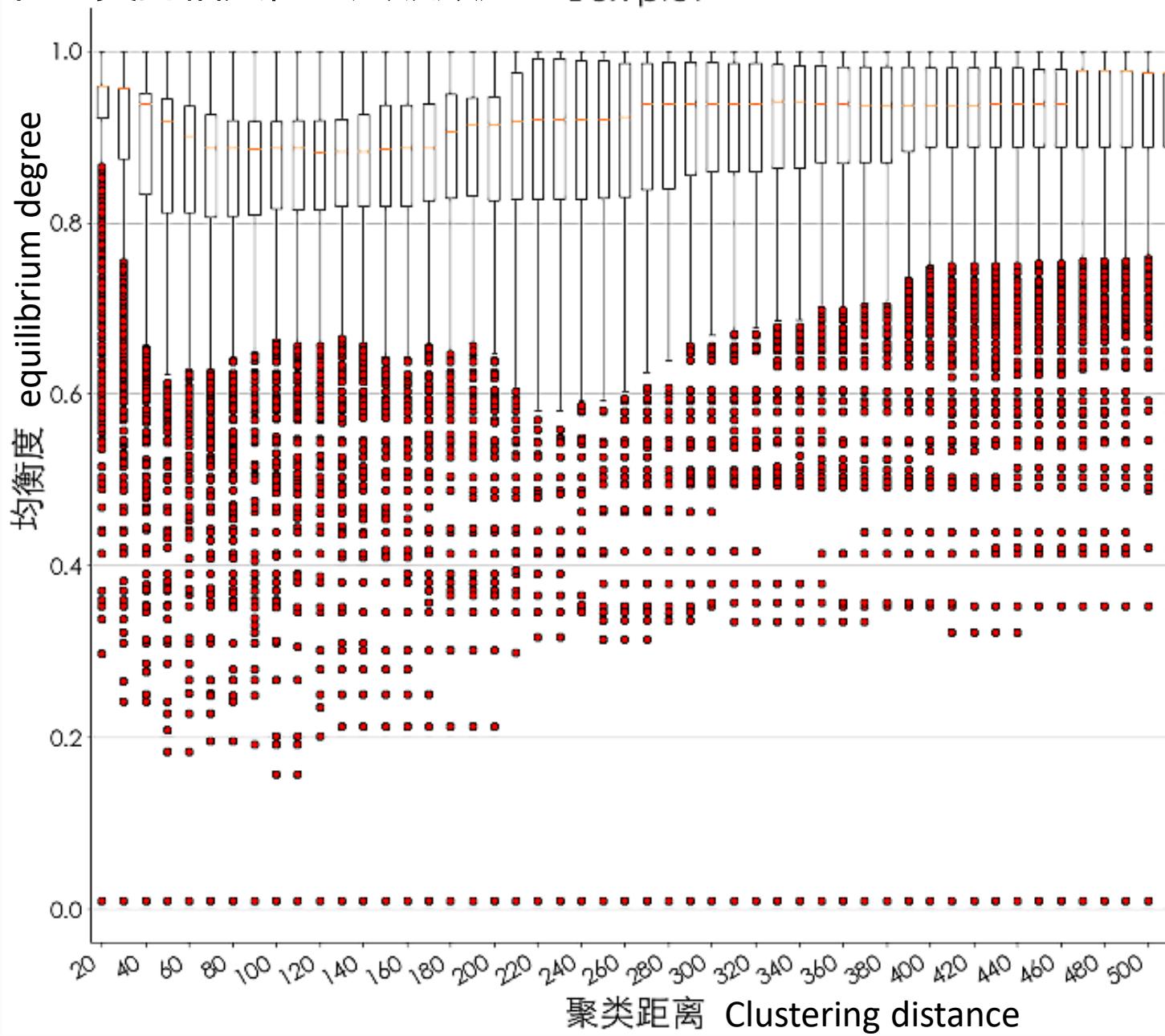
According to computing the correlation between different layers, we selected significant assemblies with higher correlation. The results show that **the more top associations between different categories situate in various layers**.

距离(m) /相关系数 (绝对值)	1.生活服务		2.购物		3.休闲娱乐				4.文化传媒			5.教育培训		
	购物	交通设	美食	公司企	丽人	酒店	金融	美食	教育培	运动健	旅游景	运动健	金融	旅游景
20	0.16													
30	0.38	0.24	0.29	0.18	0.20	0.19	0.03	0.13	0.09	0.09	0.08	0.12	0.00	0.00
40	0.29	0.25	0.27	0.27	0.32	0.21	0.08	0.14	0.10	0.08	0.04	0.10	0.04	0.00
50	0.23	0.31	0.31	0.24	0.31	0.21	0.12	0.11	0.15	0.11	0.03	0.12	0.08	0.00
60	0.20	0.30	0.30	0.26	0.27	0.21	0.16	0.09	0.21	0.17	0.04	0.15	0.10	0.00
70	0.18	0.27	0.28	0.23	0.20	0.19	0.14	0.08	0.18	0.20	0.01	0.15	0.10	0.00
80	0.18	0.25	0.26	0.24	0.18	0.20	0.15	0.07	0.20	0.24	0.05	0.17	0.09	0.04
90	0.17	0.23	0.26	0.25	0.18	0.21	0.16	0.07	0.21	0.30	0.07	0.18	0.11	0.08
100	0.19	0.30	0.26	0.24	0.19	0.20	0.18	0.06	0.24	0.33	0.13	0.24	0.16	0.09
110	0.19	0.29	0.25	0.24	0.18	0.18	0.18	0.08	0.24	0.33	0.21	0.25	0.19	0.10
120	0.18	0.26	0.26	0.20	0.17	0.17	0.17	0.07	0.28	0.30	0.26	0.26	0.16	0.15
130	0.16	0.24	0.25	0.20	0.15	0.16	0.15	0.06	0.25	0.27	0.29	0.22	0.14	0.18
140	0.16	0.22	0.22	0.21	0.14	0.15	0.16	0.06	0.25	0.27	0.28	0.22	0.12	0.19
150	0.19	0.19	0.23	0.20	0.1	0.13	0.13	0.05	0.21	0.24	0.26	0.19	0.12	0.19
160	0.19	0.18	0.22	0.21	0.12	0.13	0.12	0.06	0.21	0.23	0.27	0.18	0.12	0.19
170	0.19	0.17	0.21	0.22	0.11	0.13	0.12	0.06	0.20	0.23	0.27	0.17	0.12	0.19
180	0.17	0.17	0.23	0.17	0.11	0.12	0.12	0.06	0.20	0.23	0.27	0.17	0.11	0.19
190	0.15	0.14	0.19	0.14	0.08	0.09	0.09	0.06	0.17	0.20	0.22	0.15	0.12	0.16
200	0.16	0.14	0.19	0.15	0.08	0.09	0.09	0.06	0.17	0.19	0.22	0.15	0.12	0.16
210	0.15	0.13	0.19	0.14	0.08	0.09	0.09	0.06	0.17	0.19	0.21	0.15	0.12	0.16
220	0.14	0.13	0.19	0.14	0.08	0.09	0.09	0.06	0.17	0.19	0.21	0.15	0.12	0.15
230	0.14	0.13	0.18	0.14	0.08	0.09	0.09	0.06	0.17	0.19	0.21	0.15	0.12	0.16
240	0.12	0.11	0.17	0.12	0.07	0.08	0.07	0.06	0.14	0.16	0.18	0.12	0.10	0.13

entropy and equilibrium degree formulas based on POI life circle information

行业类的相关性与层级效应

Box plot



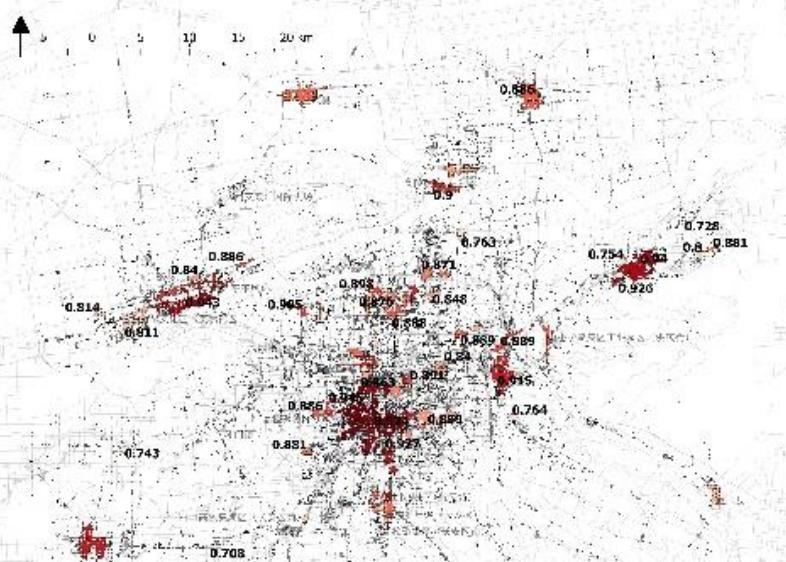
From the box plot, we can see that the equilibrium degree shows **an upward trend with the increase of distance**.

Equilibrium degree in 120m, 160m and 260m clustering hierarchies

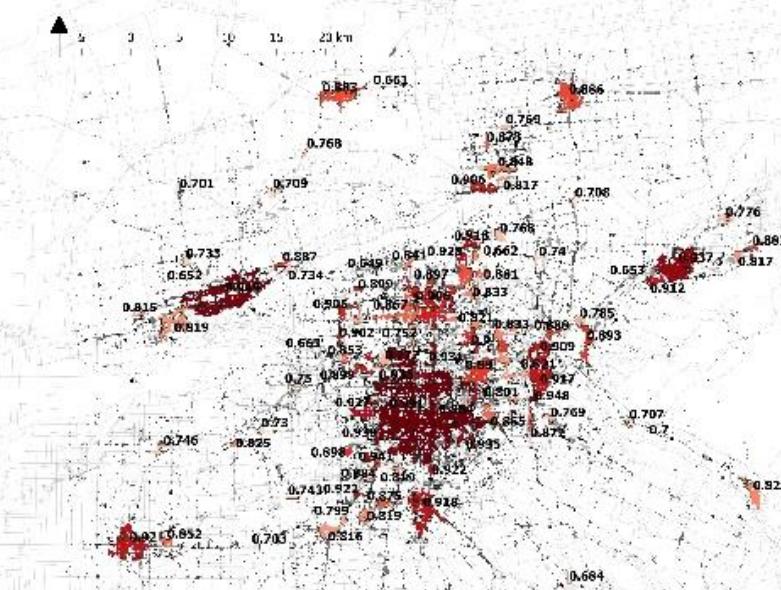
120、160、260m关键聚类层级的均衡度

By analyzing the changes of the equilibrium degree of each life circle at the critical hierarchies, to compare the homogeneity of the spatial distribution of industry categories at the same level, and **a higher equilibrium degree has a higher homogeneous**. We can explore the closeness of the connections between life circles from the aspects that the services provided by industry types can meet people's living needs.

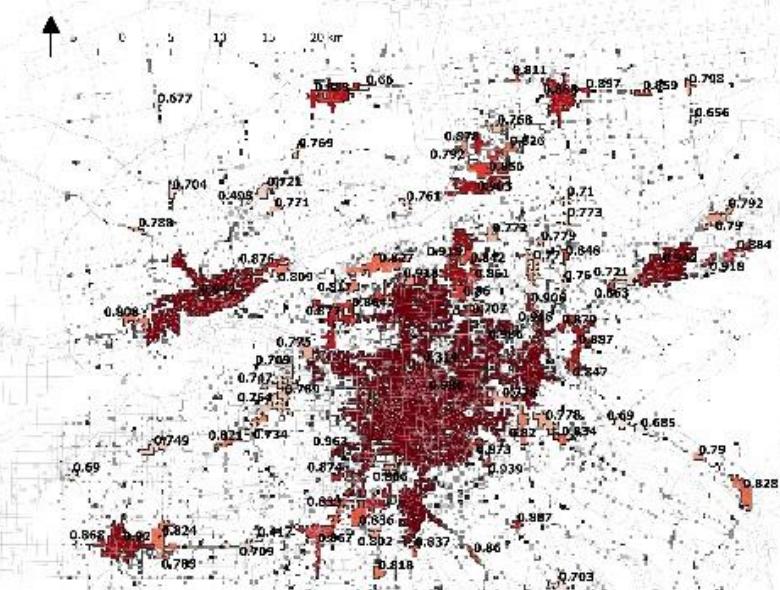
120m



160m



260m



• 讨论/Discussion

Implementation of life circle boundary under continuous dynamic hierarchy change 连续动态层级变化下生活圈边界的落实

As for the expression of life circles, it is extended from people's daily activities at a certain distance from single or multiple meaningful sites as the centers to the clustering distribution of POI points at different distances. First, it reflects the agglomeration process of social space attributes, which can clarify the evolution between levels. Second, it is of considerable significance to the research of urban spatial expansion at a certain level. Since quantifying the scope of each life circle, for different levels, especially for the meaningful levels, the study can **further analyze the degree of correlation between different life circles**, meanwhile, **explore the gaps between life circles** to dig the mechanism of urban spatial pattern formation.

social attribute structure of life circle 生活圈的社会属性结构

According to the calculated clustering frequency of the industry category, for one thing, it is possible to observe the clustering aggregation degree between urban functions under **self-organization** of various urban life circles. For another thing, the dynamic changes of industrial structure between successive levels can **provide references for the promotion of diversified social needs**. At the same time, the changes of entropy or equilibrium degree can **speculate the balanced development of the industry class** when the life circle at a certain level expands at the next level.

Study_2 _ The Relationship Between Surface Temperature and Urban Spatial Distribution Structure and the Establishment of a Prediction Model

地表温度与城市空间分布结构的关系及预测模型的建立

- 数据/Data
- 方法/Method
- 结果/Results
- 讨论/Discussion

• 数据/data United States Geological Survey, USGS

January 17, 2019 data:

LANDSAT_PRODUCT_ID = "LC08_L1TP_127036_20190117_20190131_01_T1", DATE_ACQUIRED = 2019-01-17,
SCENE_CENTER_TIME = "03:19:39.7352850Z";

August 10, 2018 data:

LANDSAT_PRODUCT_ID = "LC08_L1TP_127036_20180810_20180815_01_T1", DATE_ACQUIRED = 2018-08-10,
SCENE_CENTER_TIME = "03:19:13.1549220Z"

• 方法/Method

1 define the spatial distribution of vegetation cooling effect based on the quantitative variation of surface temperature difference
基于地表温度差值数量变化界定全区植被降温效应的空间分布

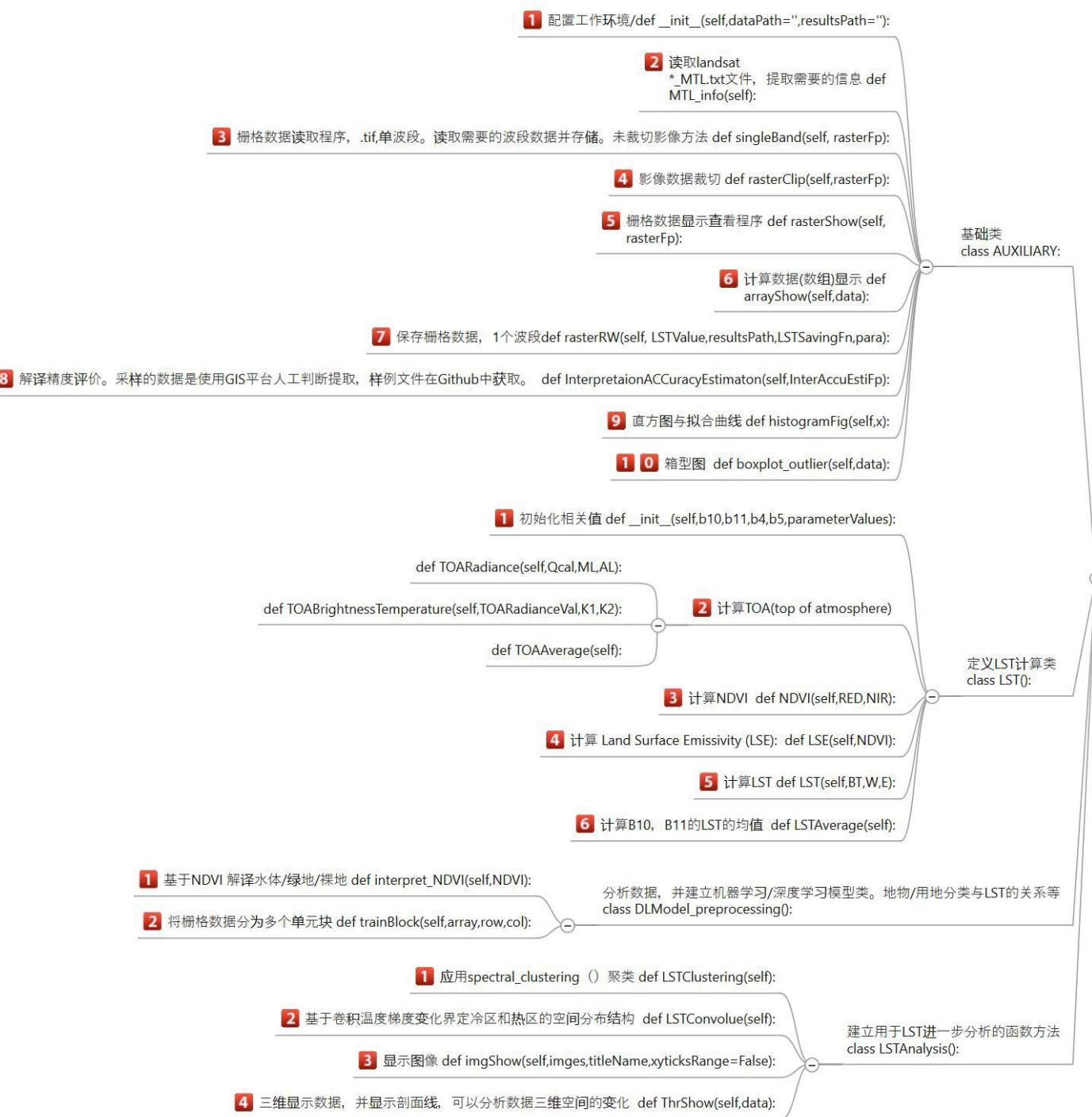
retrieve the typical **surface temperatures** of summer and winter → calculate the surface **temperature difference** → analyze the trait of the temperature difference histogram numerical fitting curve to **find the inflection points** → find numerical distribution classification and extract the **spatial distribution structure** according to the inflection points → observe the vegetation cooling effect in summer, heating effect in winter, and **spatial changes relationship** through the temperature profile

2 define the spatial structure of the cold and hot zones based on convoluting temperature gradient changes
基于卷积温度梯度变化界定冷区和热区的空间分布结构

calculate the difference between the cell and the mean value of the surrounding cells
Convolution kernel $\begin{bmatrix} 1/8, 1/8, 1/8 \\ 1/8, -1, 1/8 \\ 1/8, 1/8, 1/8 \end{bmatrix}$

3 establish the prediction model of surface temperature base on some machine learning regression algorithm for a greenbelt planning evaluation
基于机器学习回归算法建立用于绿地规划评估的地表温度预测模型

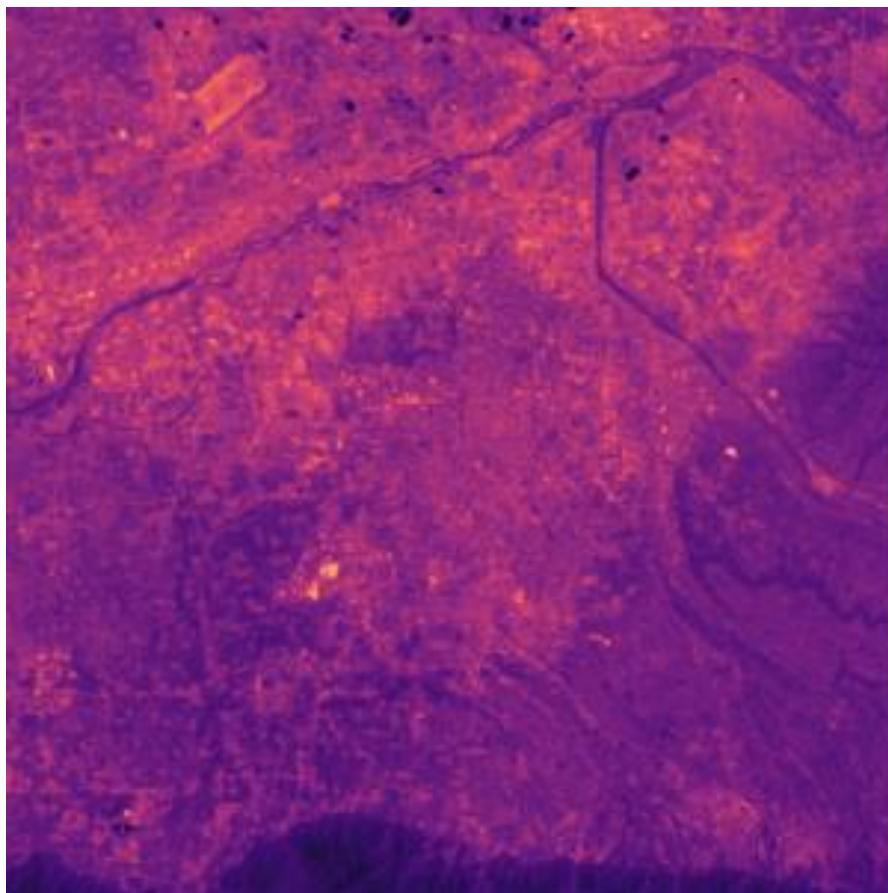
scikit-learn: Extra trees、K-nn、Linear regression and Ridge



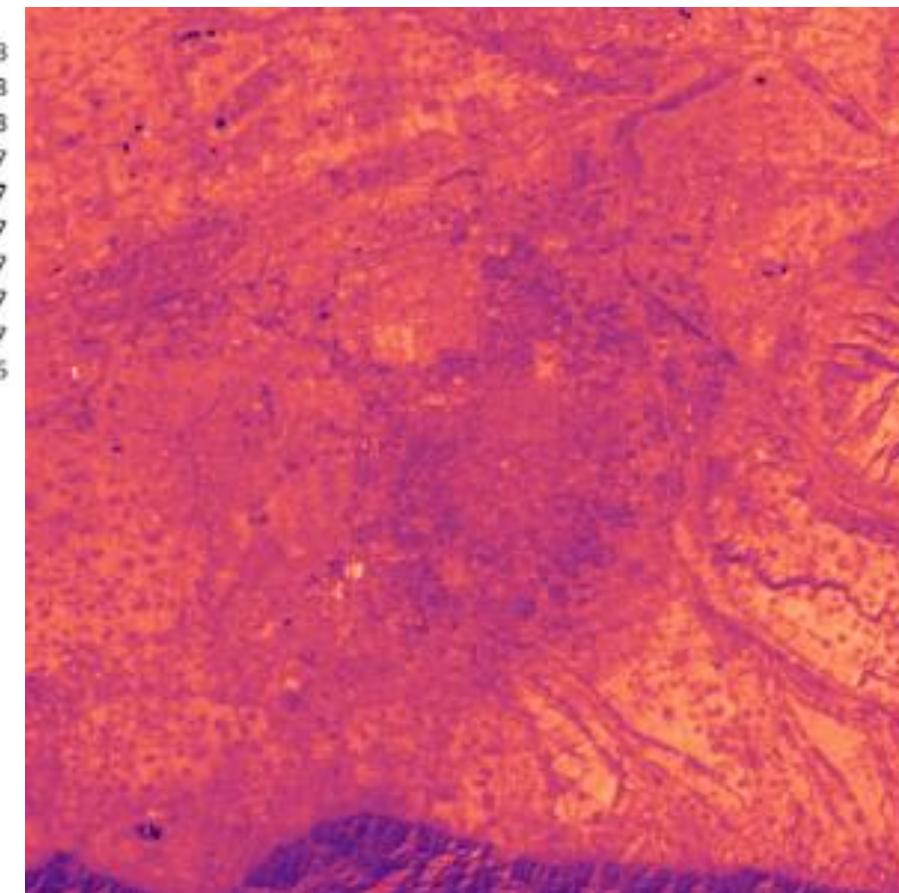
•结果/Results

1 define the spatial distribution of vegetation cooling effect based on the quantitative variation of surface temperature difference

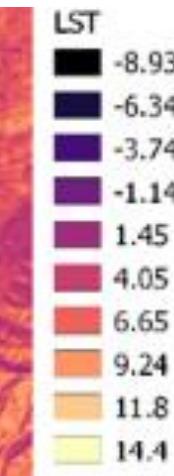
基于地表温度差值数量变化界定全区植被降温效应的空间分布



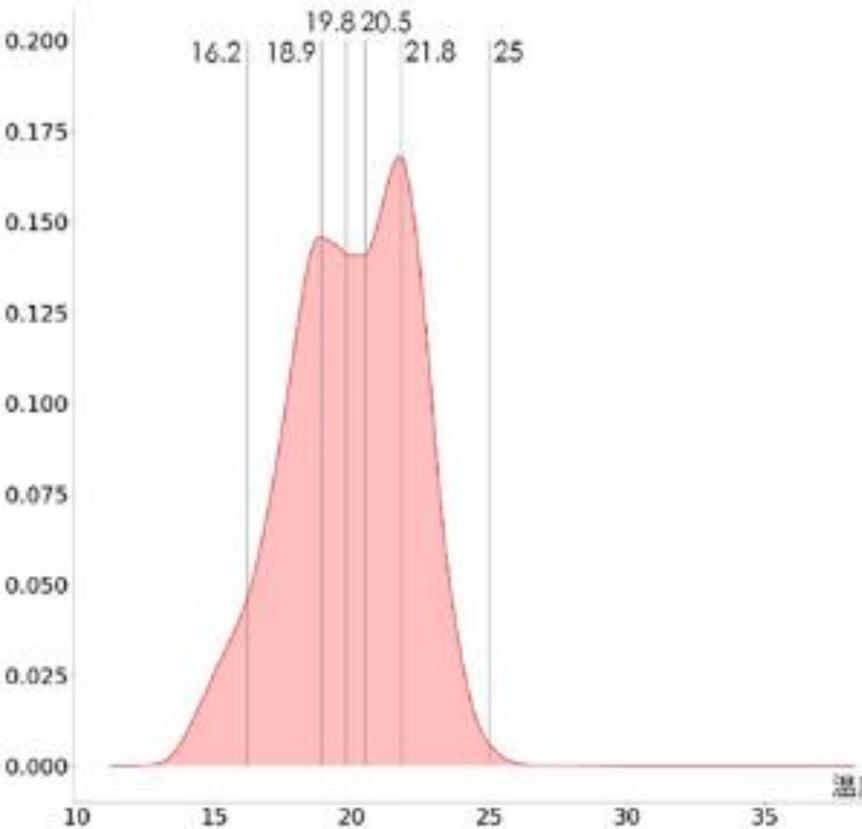
Surface temperature on August 10, 2018



Surface temperature on January 17, 2019

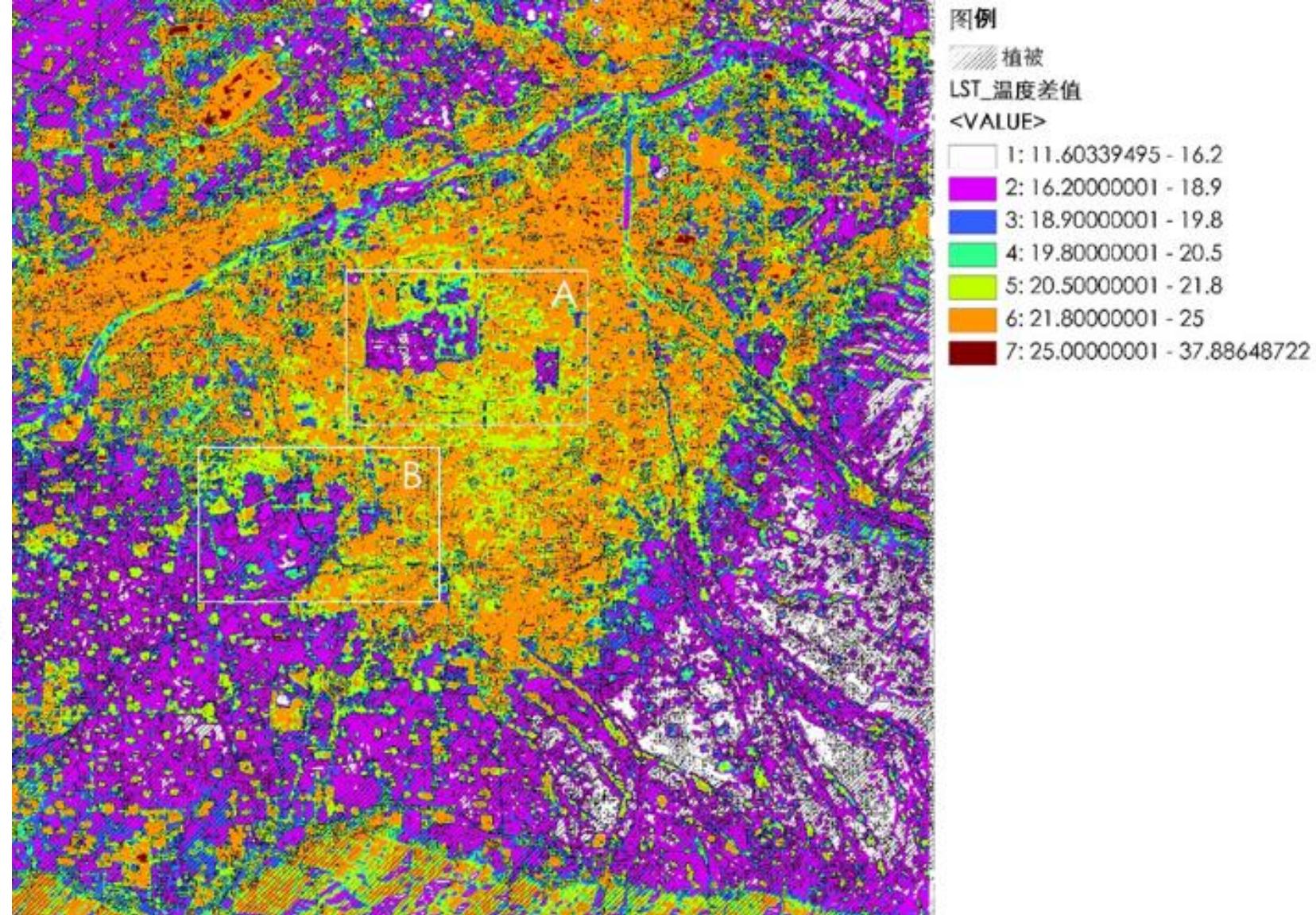


地表温度



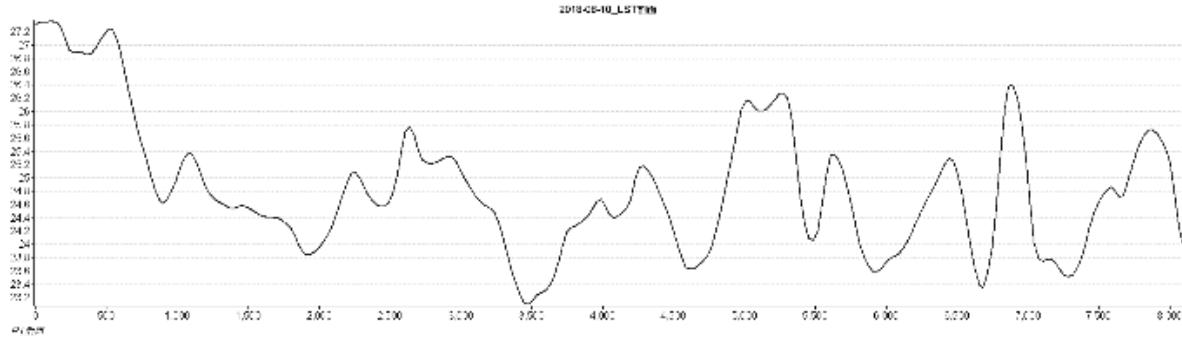
Histogram of temperature difference, fitting curve and inflection point

通过温度变化分布的直方图，提取拐点，并依次重分类地表温度分布图

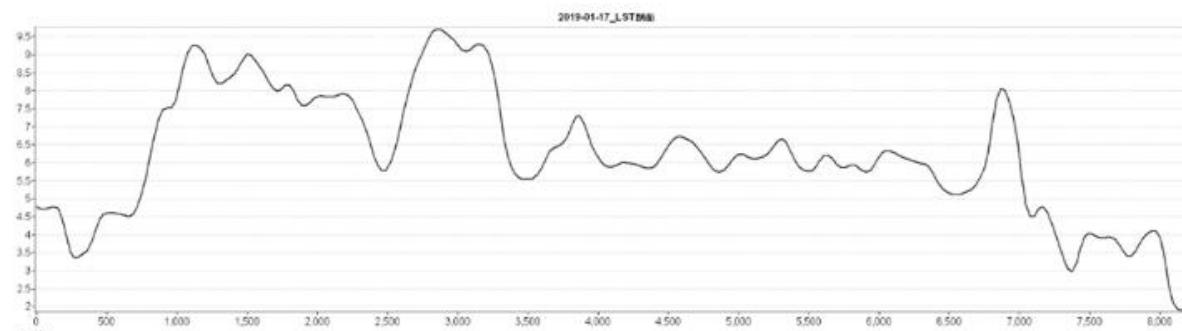


Surface temperature difference

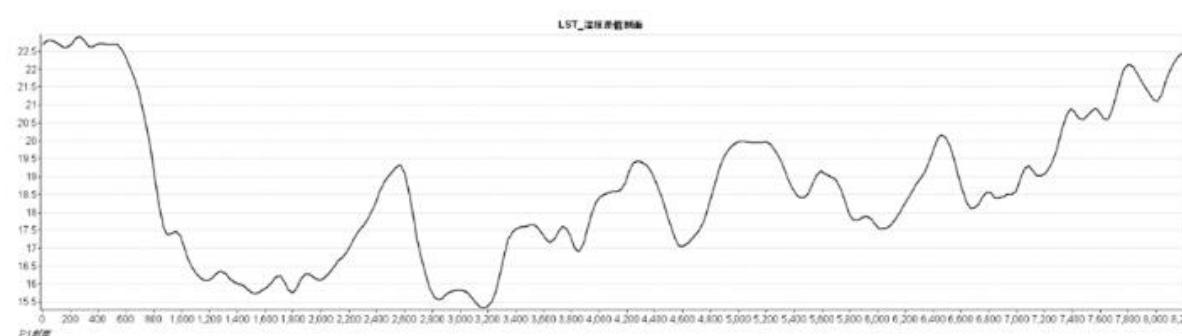
In order to **analyze the relationship between surface temperature change and the spatial distribution of vegetation** and construction, especially the cooling effect of plants, the histogram was used to observe the data distribution of temperature difference and find out the critical points through the fitting curve.



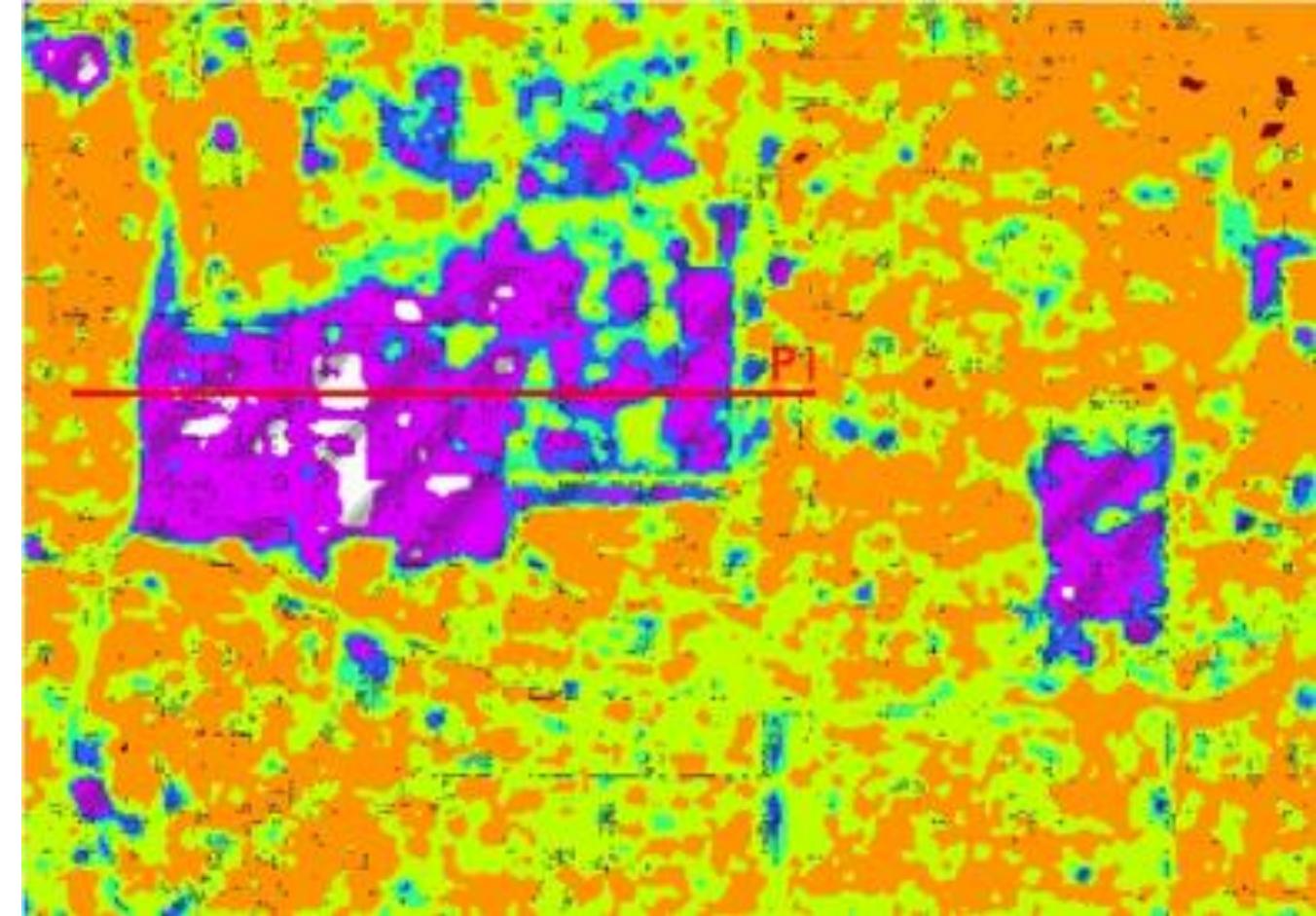
Surface temperature of P1 section on August 10, 2018



Surface temperature of P1 section on January 17, 2019



Temperature difference of P1 profile



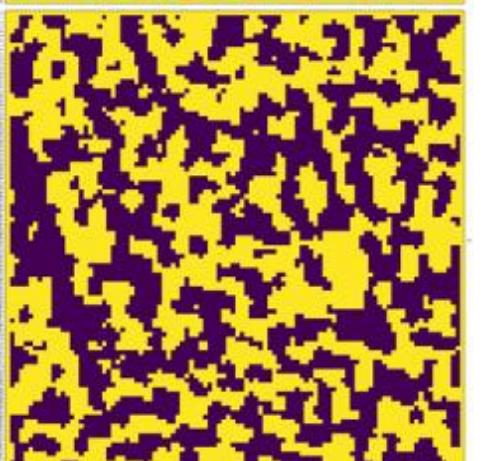
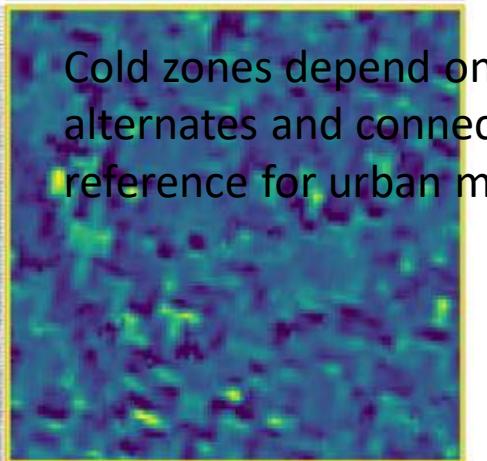
Urban greenbelt (WeiYangGong ruins of Chang'an city)

The spatial distribution of the cooling effect of urban green space provides **a quantifiable spatial range** for the study of urban development, green space construction, especially open space.

2 define the spatial structure of the cold and hot zone based on convoluting temperature gradient changes

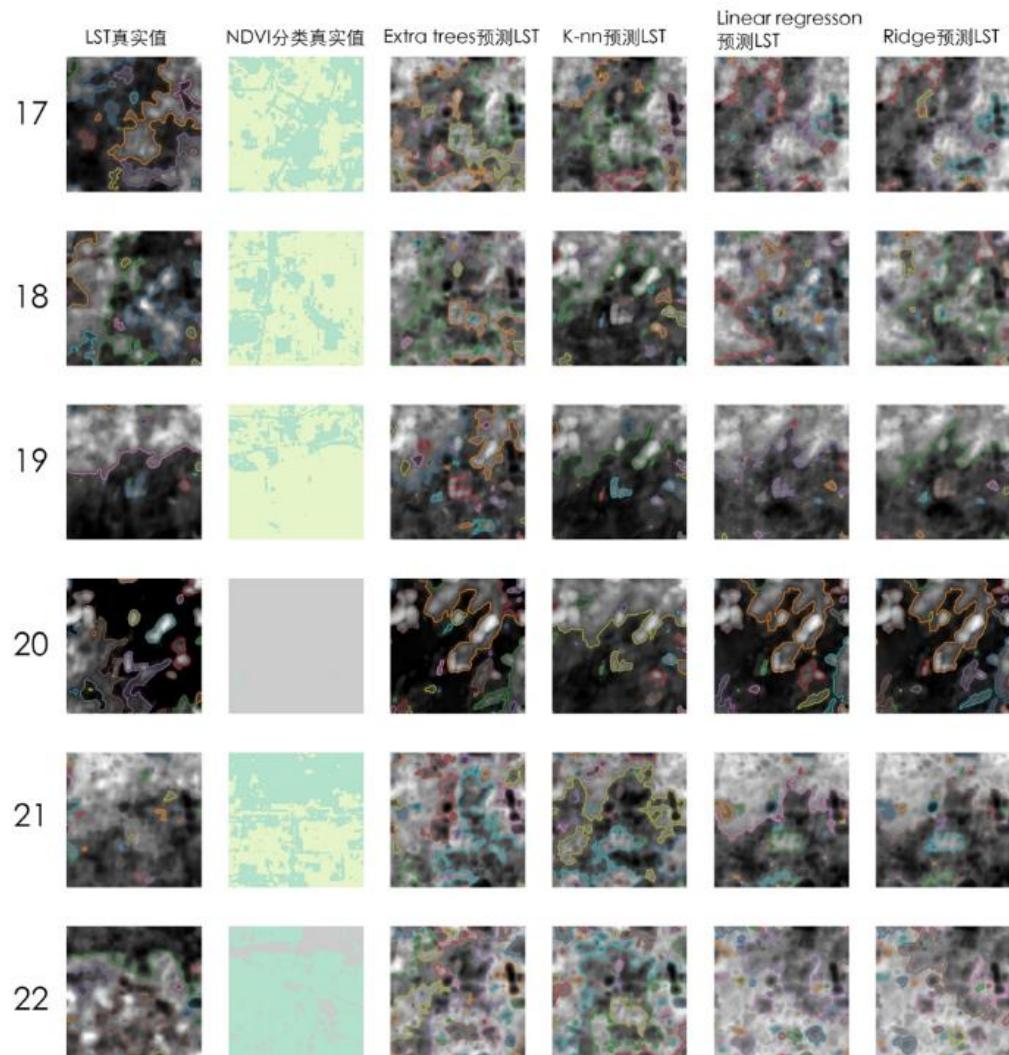
基于卷积温度梯度变化界定冷热区的空间分布结构

Surface temperature, K
Cold zones depend on the distribution of vegetation. The results show that the spatial distribution of cold and hot regions alternates and connects, forming a large number of uncertain local microclimate environments. It provides a quantitative reference for urban micro-renewal and improvement of the urban thermal environment.

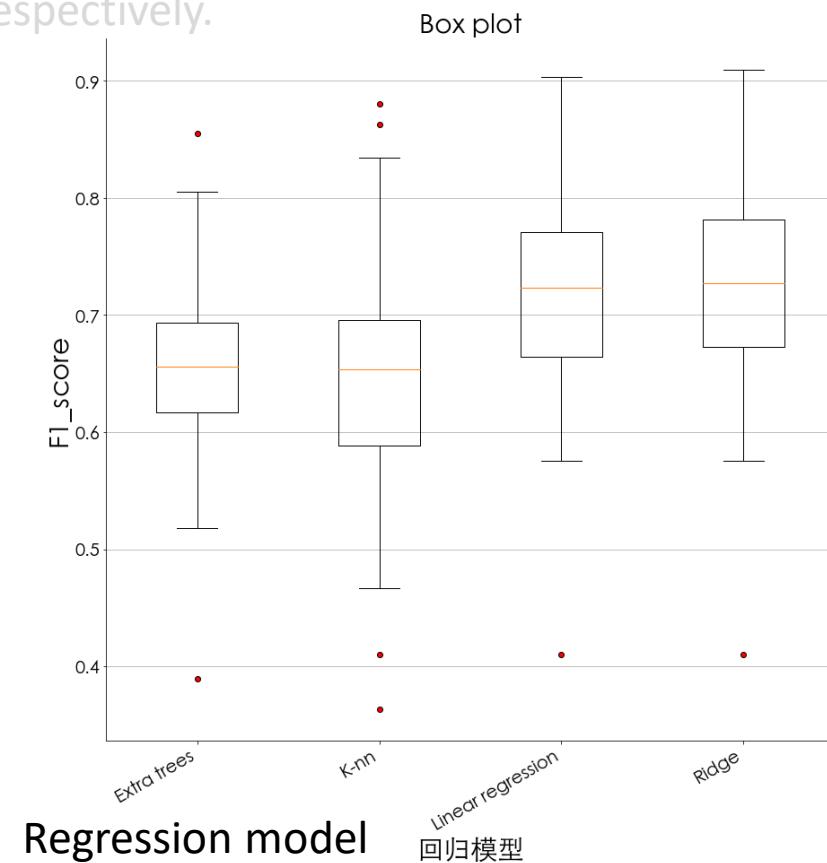


3 establish the prediction model of surface temperature base on machine learning regression algorithm for a greenbelt planning evaluation

基于机器学习回归算法建立用于绿地规划评估的地表温度预测模型



Construction land, green land, and water body are taken as explanatory variables of the regression model, and LST calculation results are taken as target variables. Regression models Extra, trees, K-nn, Linear regression are used to train the models, respectively.



• Discussion

urban spatial distribution structure reflected by quantifying surface temperature

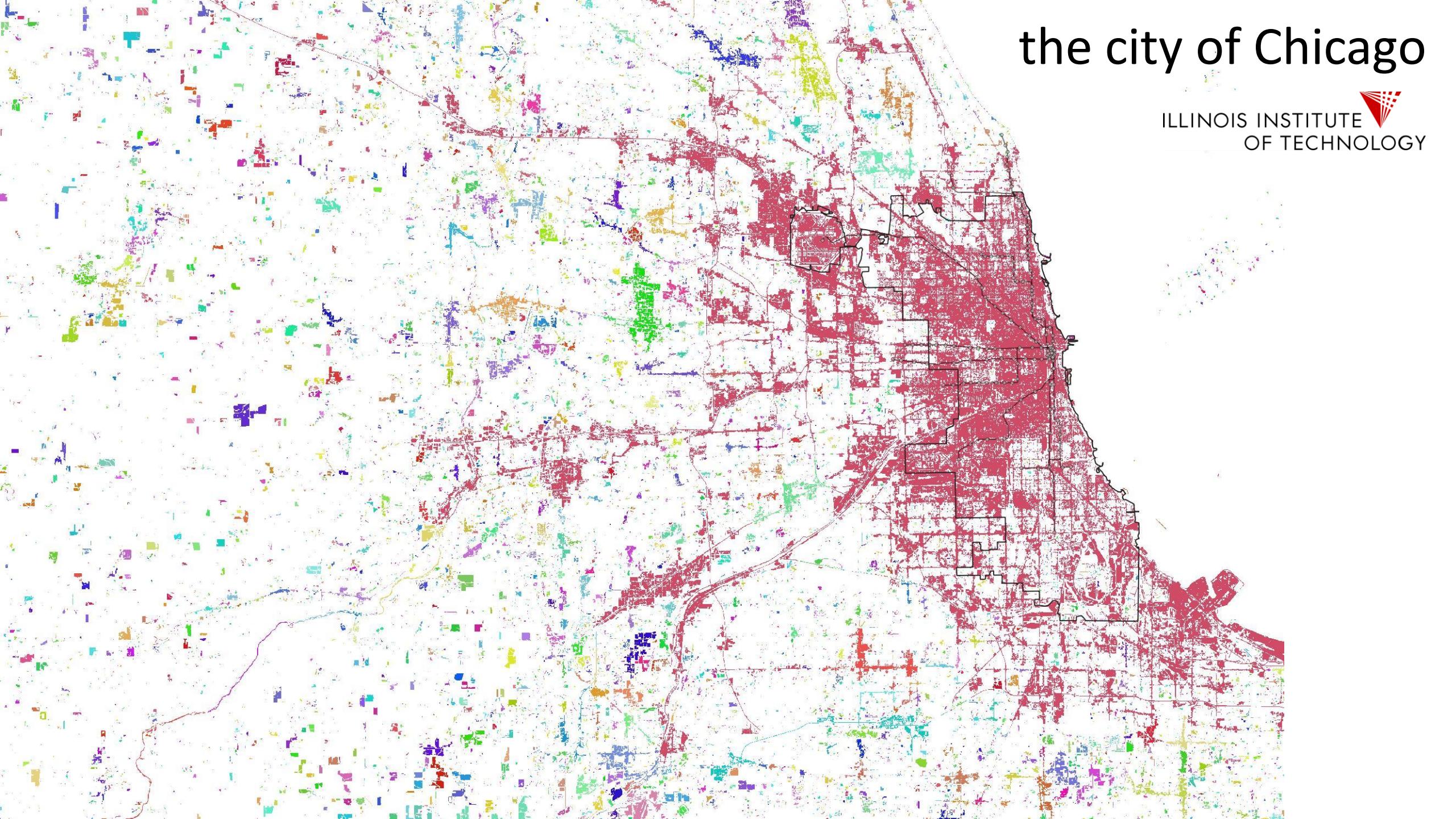
量化地表温度所反映的城市空间分布结构

It is meaningful to extract the specific range of surface temperature change in the transition area between vegetation and construction land at the urban scale, that is, the vegetation cooling effect zone. The determination of specific regions of the cooling effect is helpful to the implementation of urban space planning when **the surface temperature is taken as an indicator of urban thermal environment evaluation.**

to establish a prediction model when taking surface temperature as the indicator of the urban thermal environment evaluation

将地表温度作为城市热环境评价的指示因子建立预测模型

The prediction **accuracy** is affected by several factors, including the accuracy of remote sensing images, the classification of land cover, the distribution difference of the same land cover and the selection of model, etc. Therefore, it is usually used to predict the distribution of temperature change.

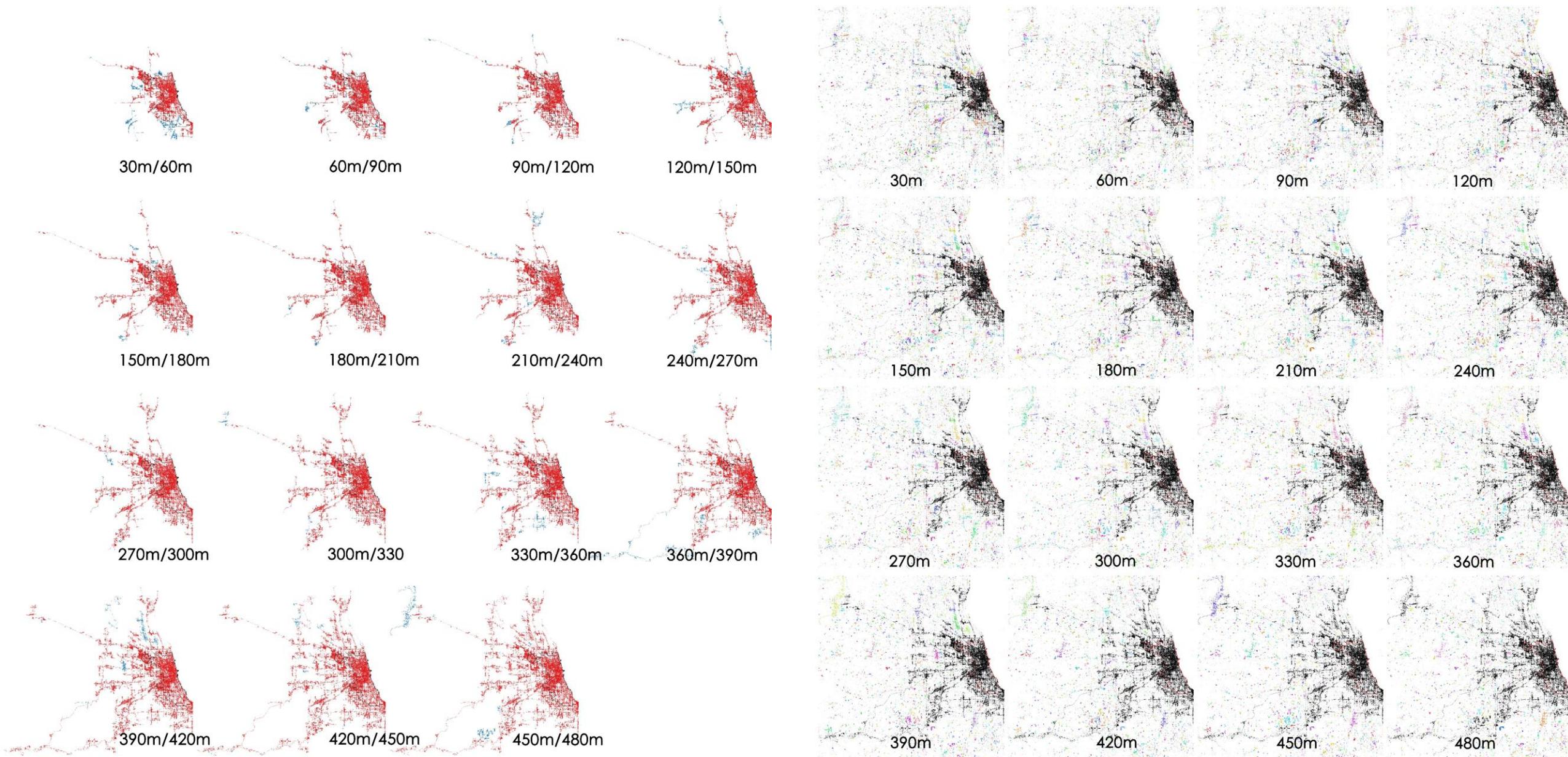


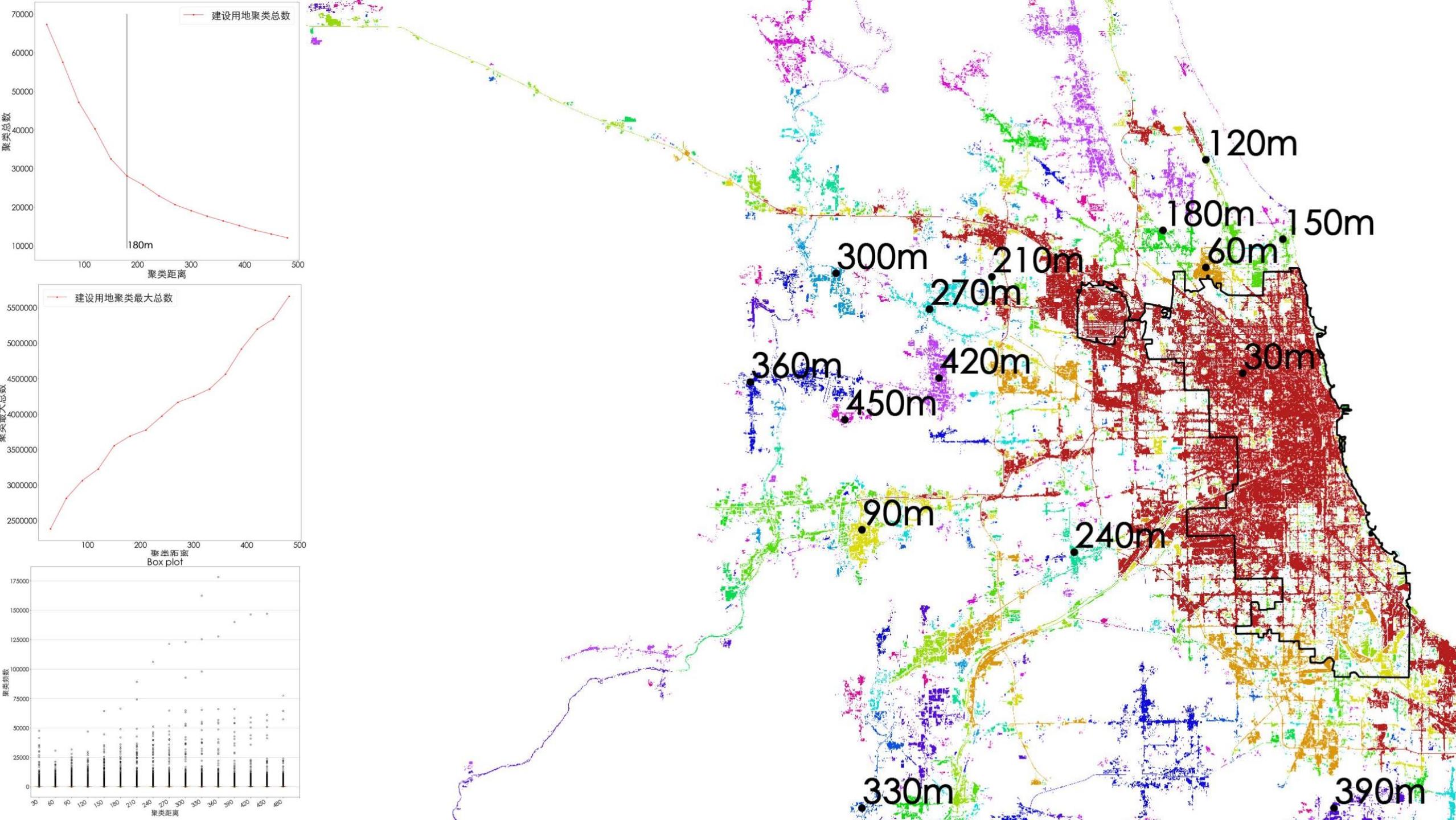
the city of Chicago

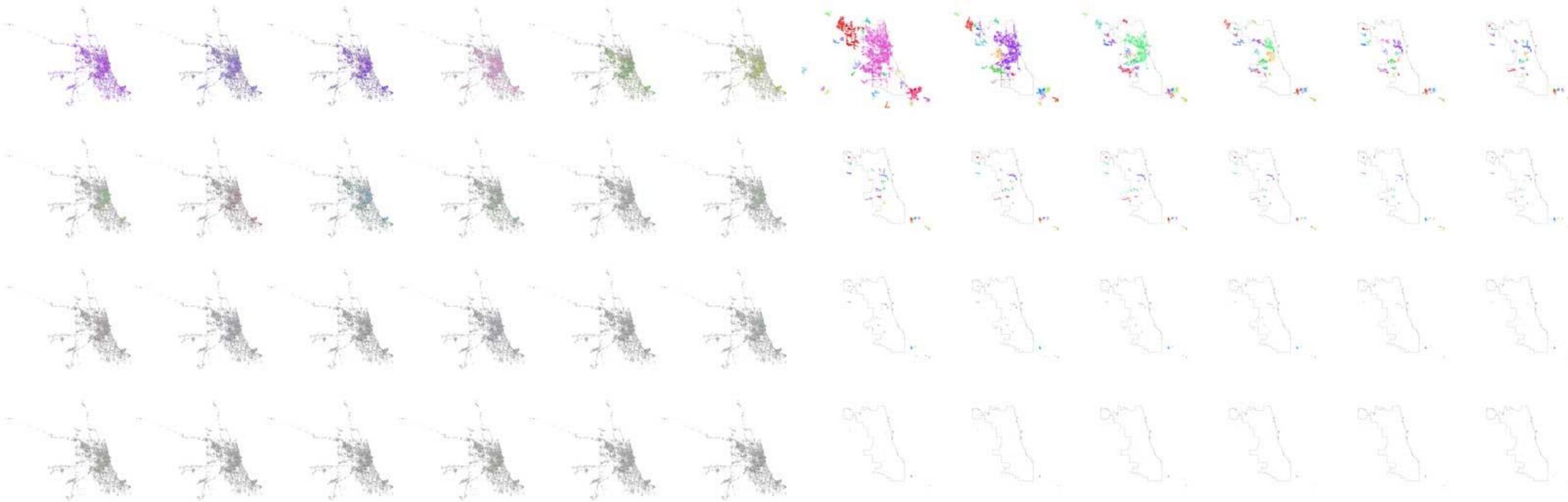


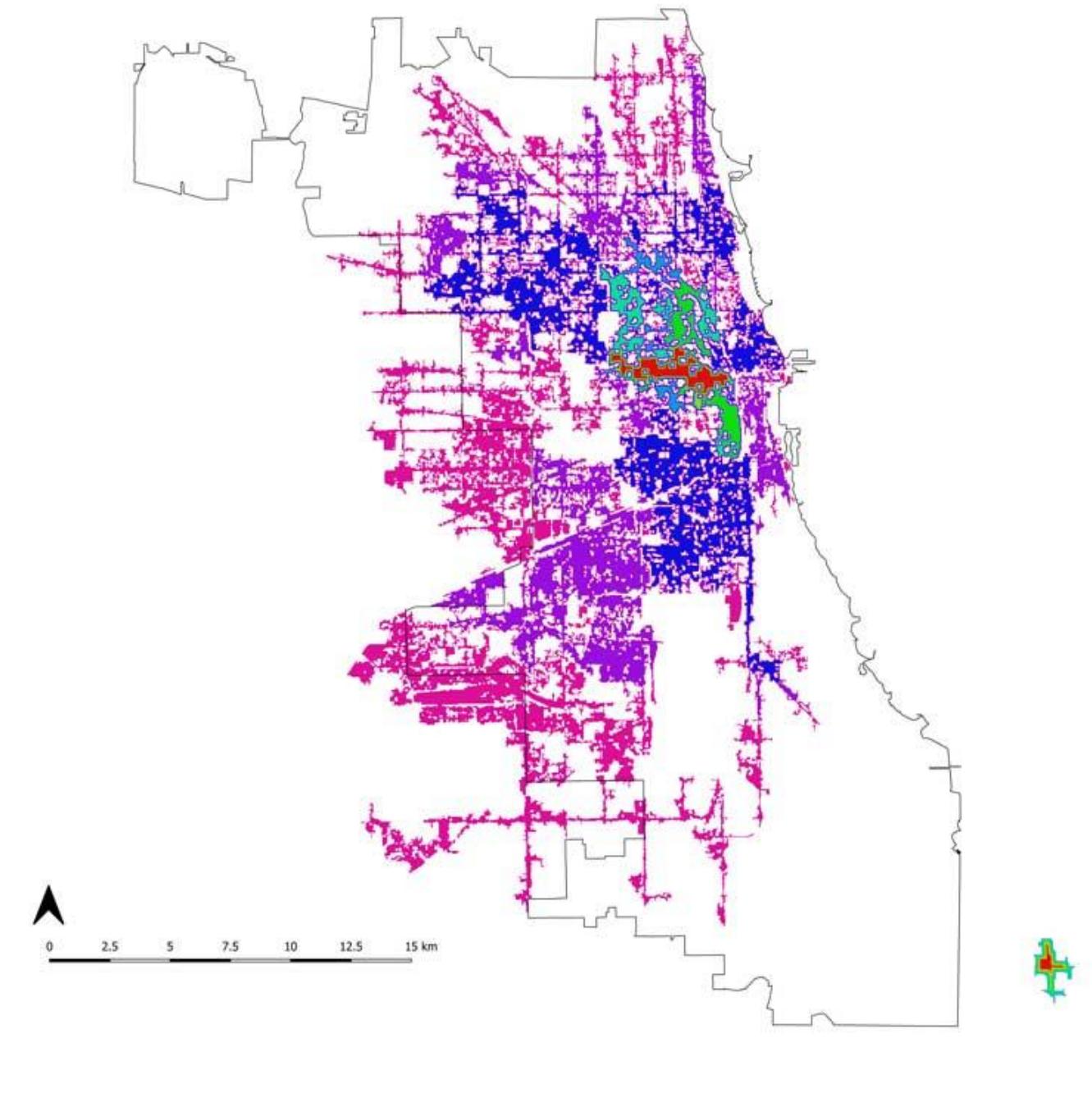
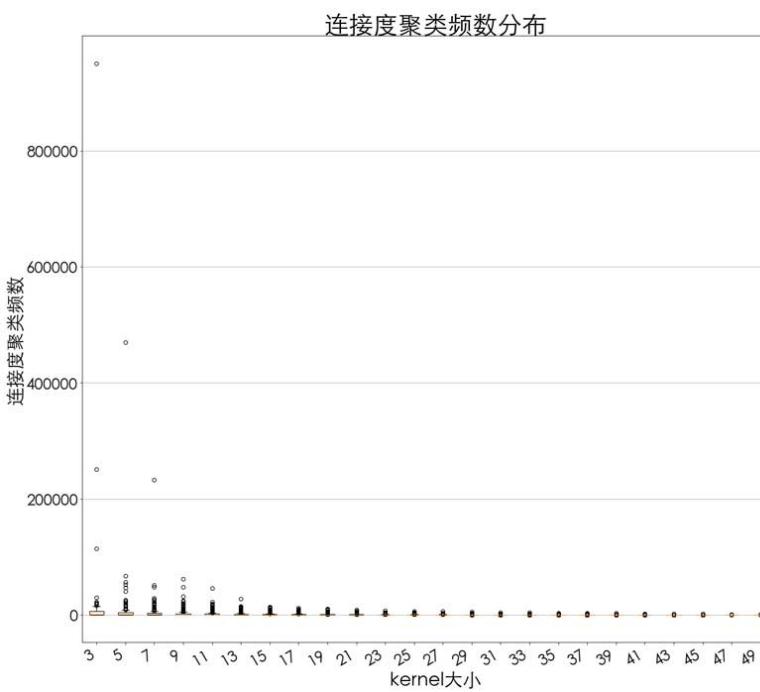
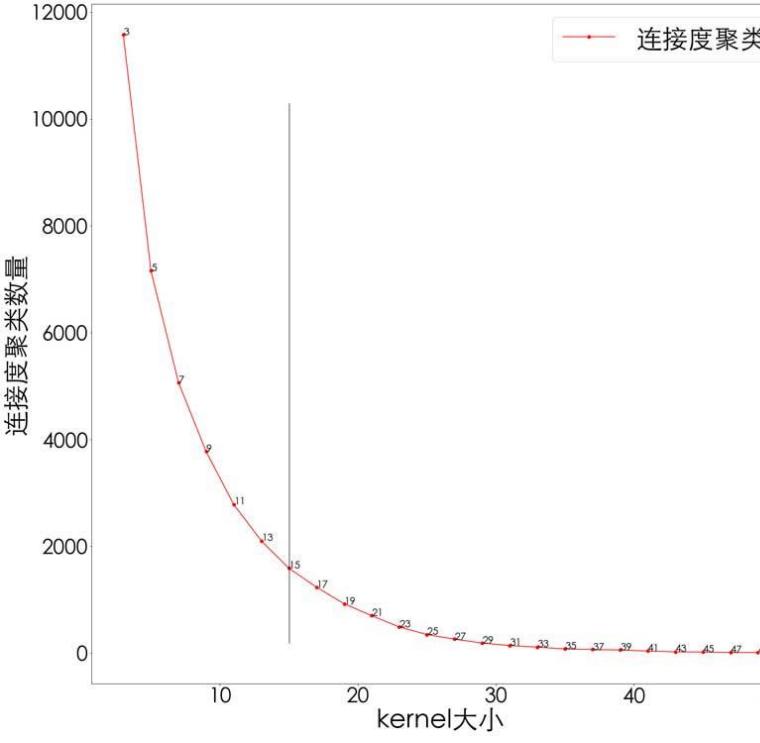


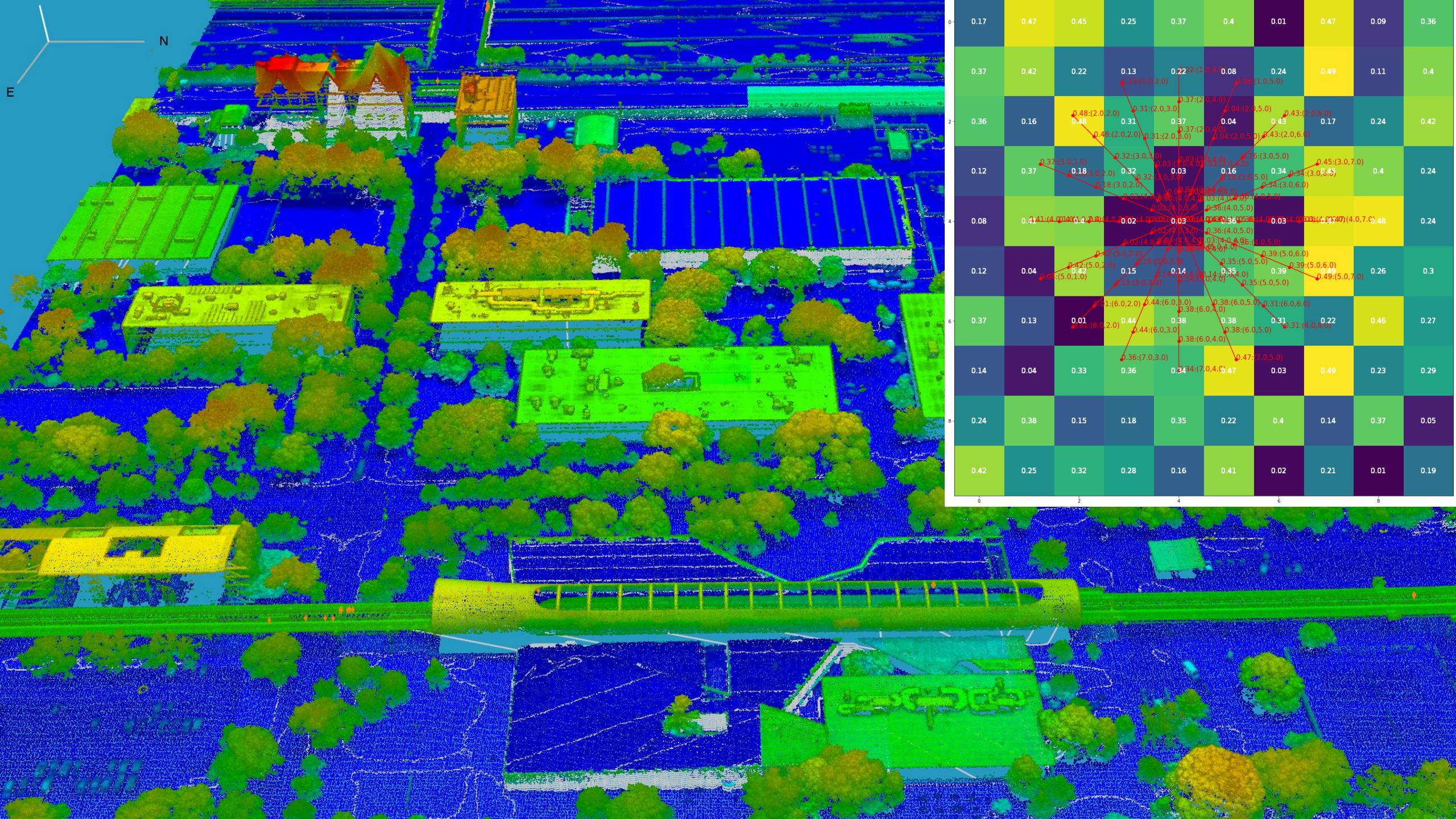
华晨宝马











open source

<https://github.com/richieBao/python-urbanPlanning>

