

# 城市时空大数据分析可视化

## Urban Spatio-temporal Data Analysis and Visualization

王江浩

([wangjh@lreis.ac.cn](mailto:wangjh@lreis.ac.cn))

中国科学院地理科学与资源研究所  
资源与环境信息系统国家重点实验室

2016-11-08 @ Tsinghua

# Contents

---

- 引言
- 时空数据挖掘
- 人群足迹建模
- 人口流动分析
- 城市活力研究
- 总结

# 1 INTRODUCTION



# 大数据制图产生的背景

- **Approximately 80%** of the informational needs of a local government policymaker is related to a geographical location. ( *Robert E. Williams, 1987* )
- **数字地球——Digital Earth** . (Al Gore , 1998).
- **智慧城市——Smart City**: a smart economy, mobility, environment, people, living, governance. (Caragliu et al. 2009)
- **大数据时代——Big Data Era**
- .....



# 现代科学技术的发展与推动



Silver Partner



Premiere Value Partner



Premiere Business Partner



Premiere Certified Partner



Solutions Advisor



Authorized Reseller



Platinum Partner



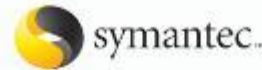
Gold Partner



Silver Partner



Advanced Velocity Partner



Authorized Reseller



Gold Certified Partner



Solutions Partner



Certified Partner



Authorized Reseller



Business Partner



Enterprise Solutions Provider



Authorized Reseller



Gold Partner



100K Club Partner



Authorized Reseller



Authorized Reseller



Authorized Reseller



Authorized Reseller



Business Partner



VIP Partner



Authorized Reseller



Authorized Reseller



Authorized Reseller



Bronze Partner



Authorized Reseller



Authorized Reseller



Authorized Reseller

# Spatio-temporal Data

## □ Spatial

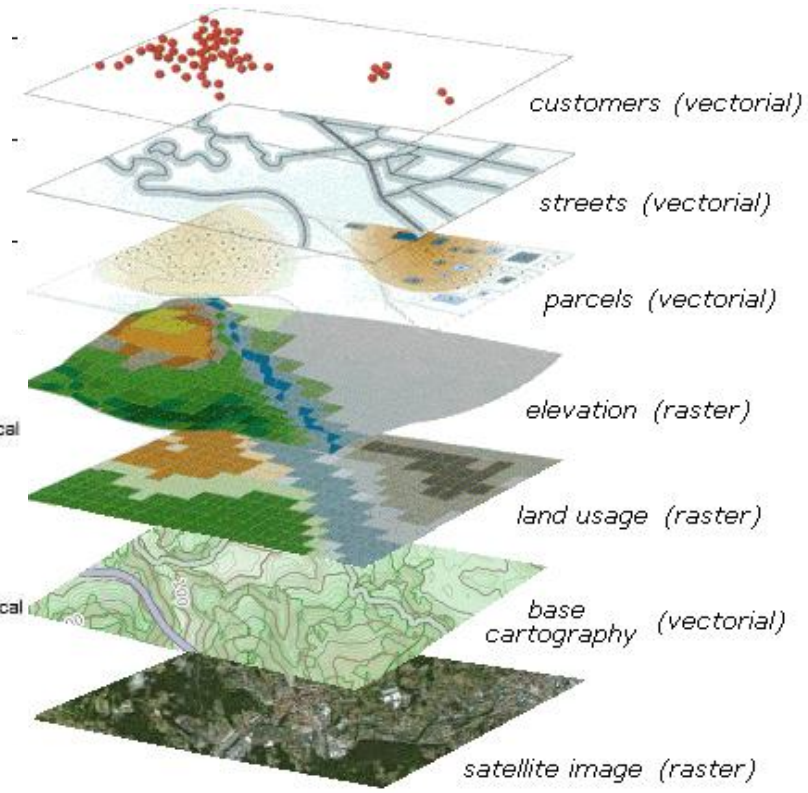
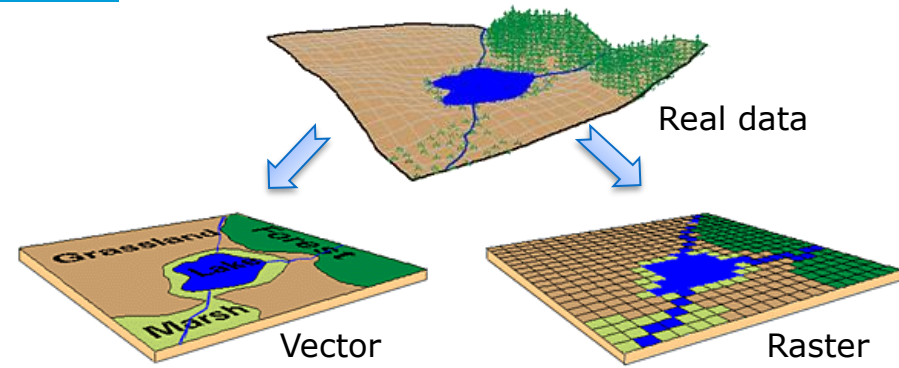
- Vectorial: Point, Polyline, Polygon, Volume
- Raster
- Spatial Reference

## □ Temporal

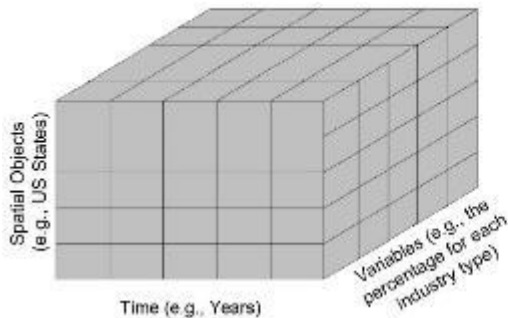
- Date + Time

## □ Attribute

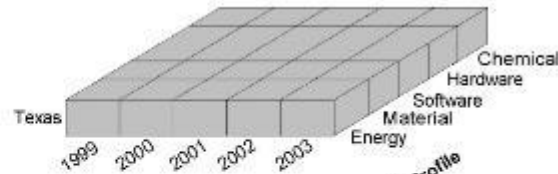
- Data Frame



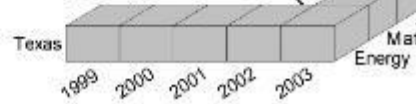
(A) Space-Time-Attribute Cube



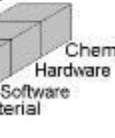
(B) Time-Attribute Slice (e.g., for Texas)



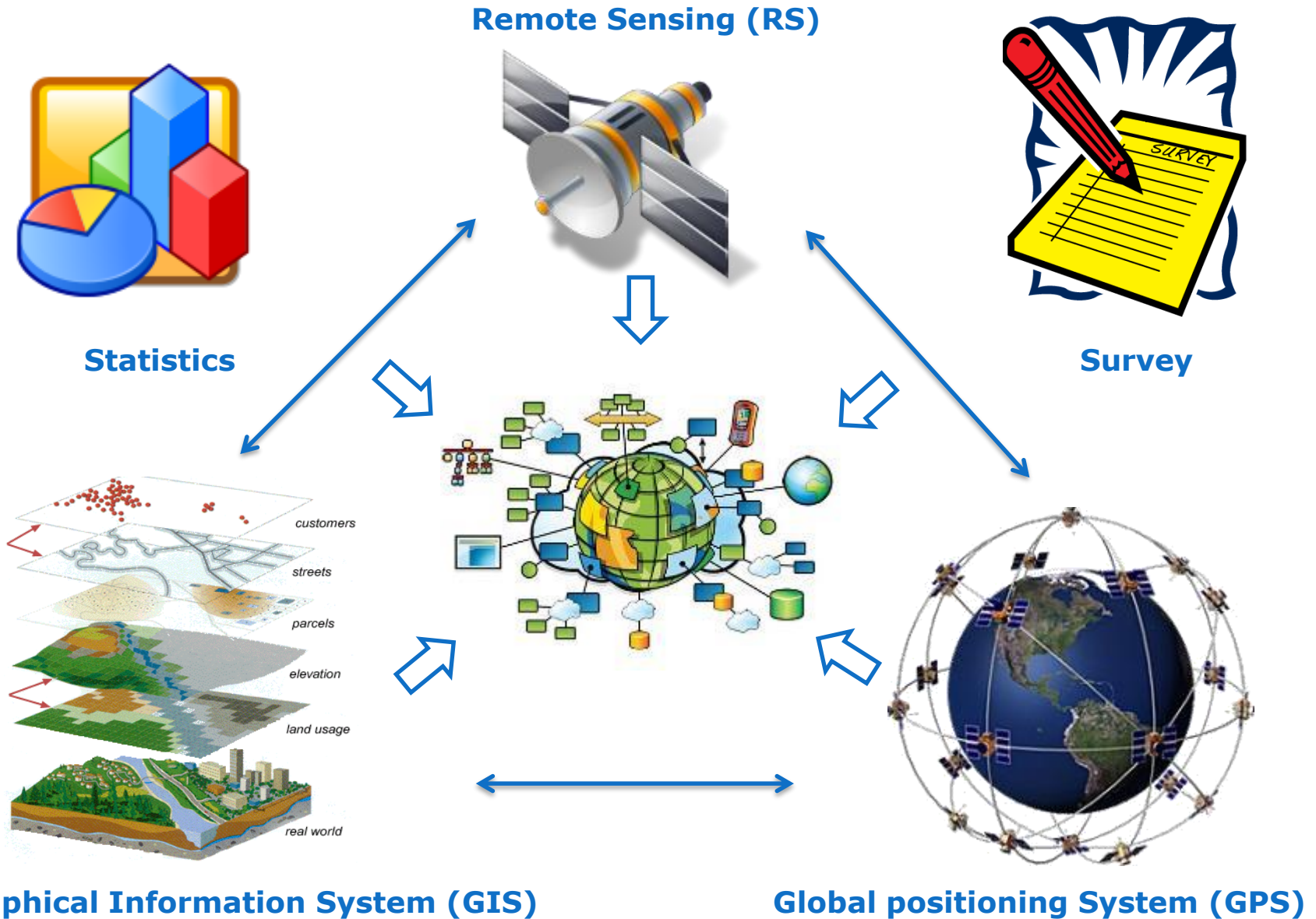
(D) Temporal Series



(C) Multivariate Profile



# Modeling Our World



# New Spatio-temporal (Big) Data

---

## (1) Social Media Data





# New Spatio-temporal (Big) Data

---

## (2) User Generated Contents (UGC)



20 million POIs in China

# New Spatio-temporal (Big) Data

(3) GPS-enabled smart phones generated data



**Where People Run**



**Taxi Tracking**  
(by Yu Zheng)

# New Spatio-temporal (Big) Data

---

(4) Remote Sensing, Wireless Sensors Network, VGI, Smart Card...

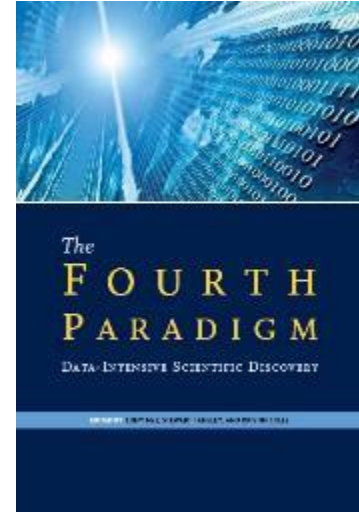
**WSN**   
Wireless Sensor Networks



OpenStreetMap, OSM

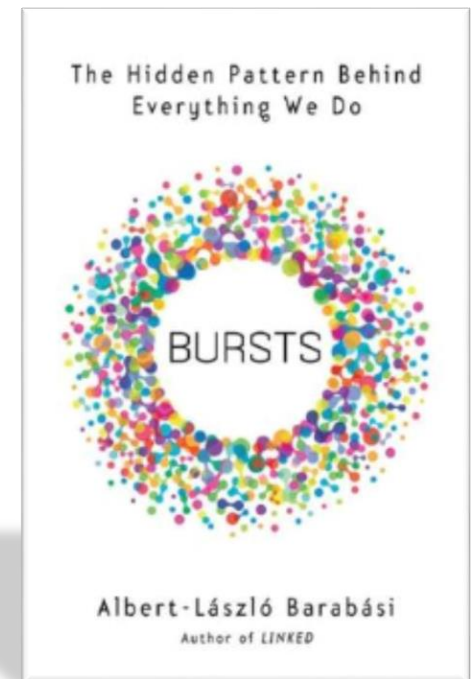
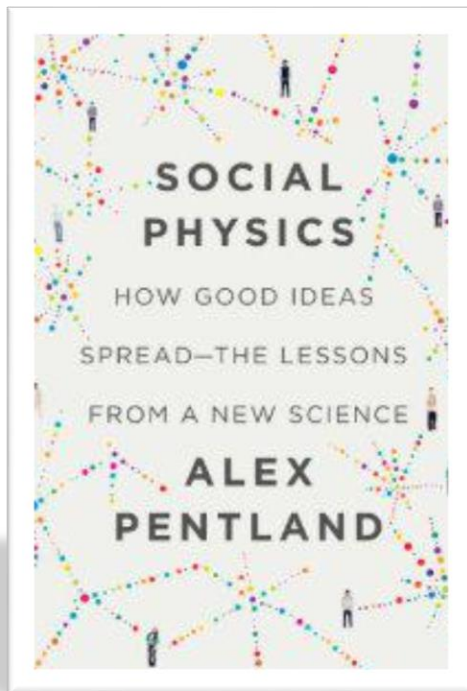
# New Way to Model Our World

- **Big data & Open data**
- **VGI : Citizens as sensors:** the world of volunteered geography.  
— Michael F. Goodchild , 2007
- **Fourth paradigm of Scientific**
- **Physical => Social**



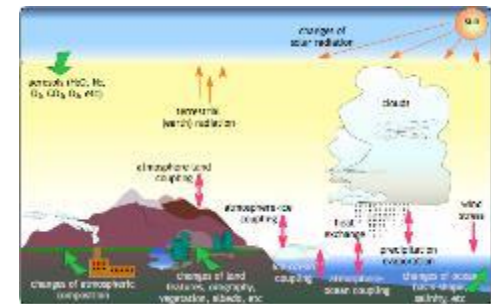
# “以人为本” 的研究

- **Social Physics / Social Geography / Social Sensing**
  - Social Computing
  - Complex Network
  - Human Dynamics



# ST modeling and data mining

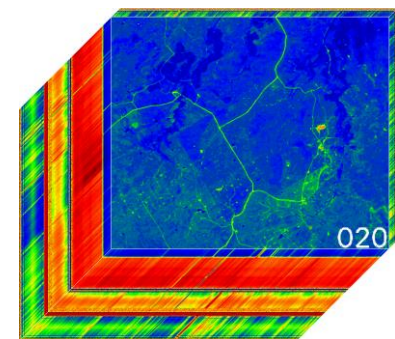
- Description and Visualization
- Separate model
  - Spatial analysis
  - Time series analysis
- Spatio-temporal model
  - **Statistical model**
  - Physical model for specific field
  - Statistical + physical model
- Spatio-temporal data mining
  - Machine learning
  - Statistical deep learning
  - .....



Physical model

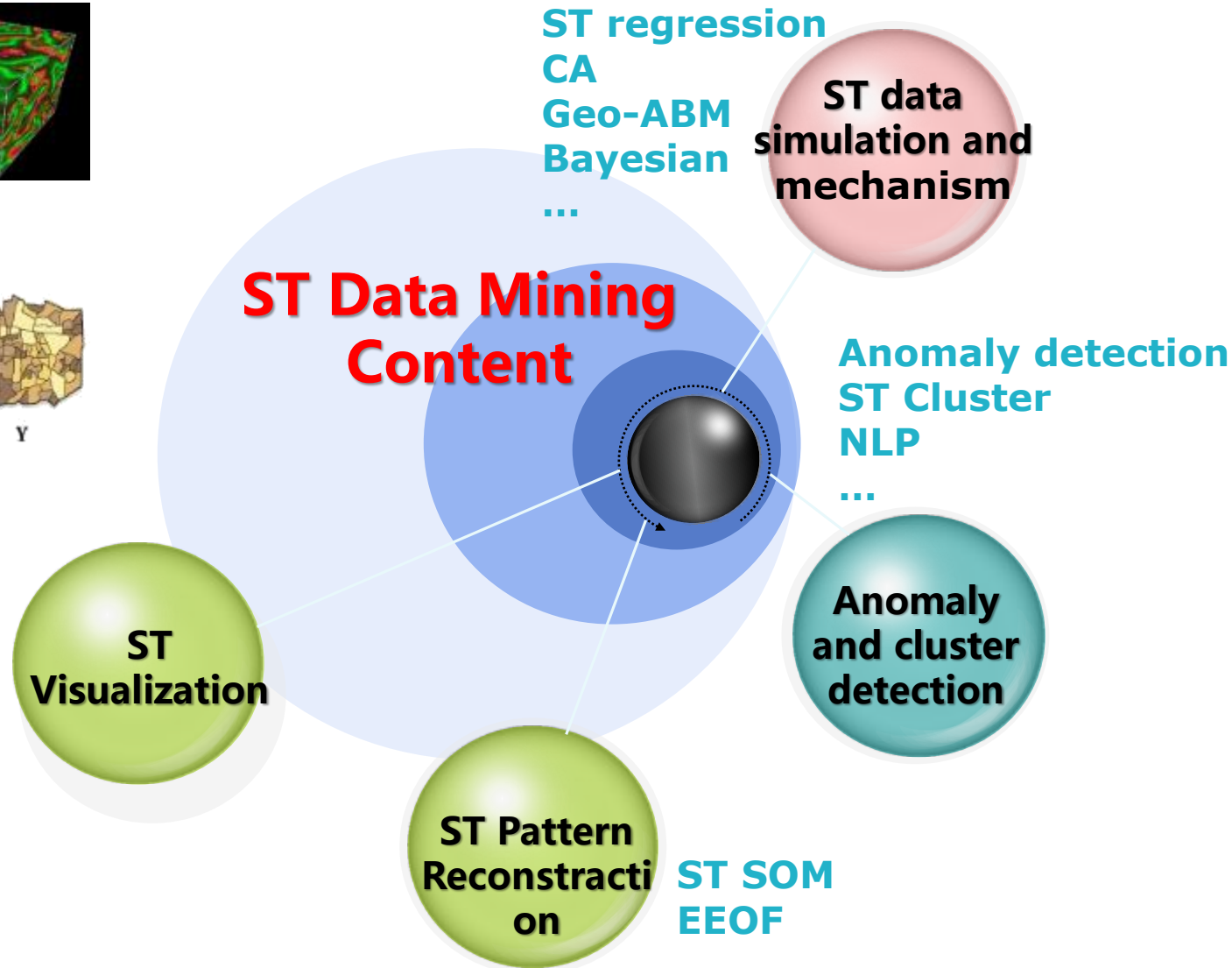
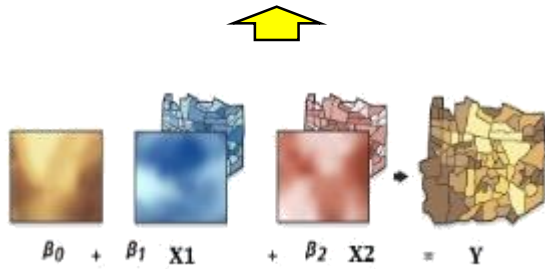
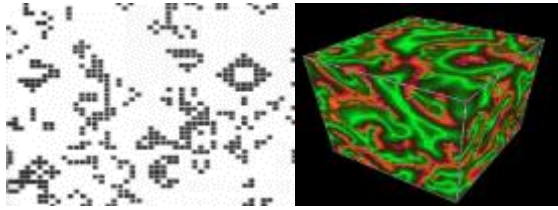


Statistics model

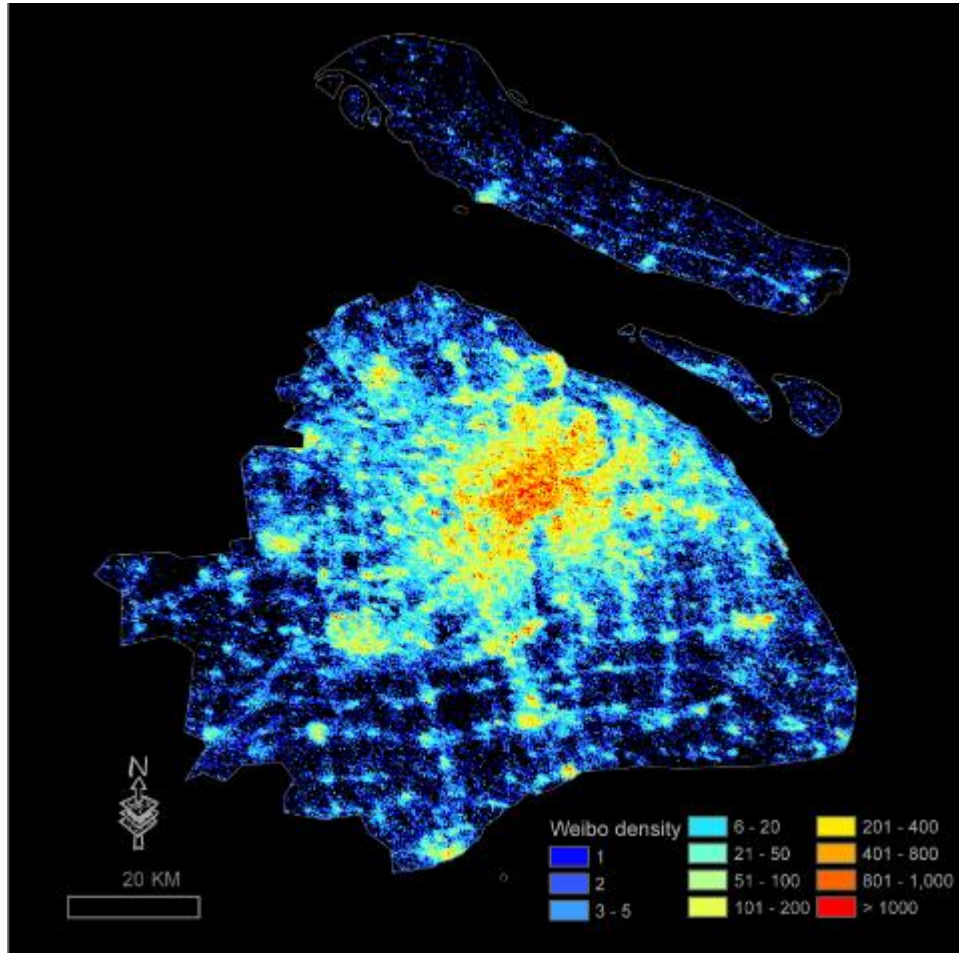


Data mining

# ST big data mining



# Application of ST data mining



Social Media in Shanghai

## Smart City

- Transportation management
- Urban planning
- City management
- Social life
- .....

## Scientific Research

- Human behavior
- Urban computing
- Public health
- Social-economic
- Geographic study
- .....



# 2 ST Data Analysis



# Techniques for ST data analysis

## Road to be a Data Scientist

### □ Acquisition



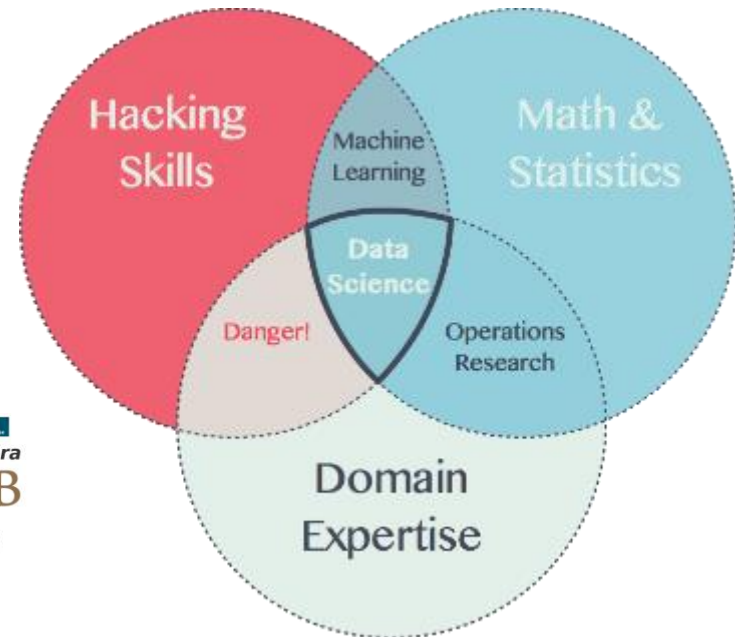
### □ Database



### □ Analysis, data mining



### □ Specialized tool



# Development of ST statistics

---

- classical statistics
  
- spatial statistics
  - Spatial target
  - Spatial autocorrelation , spatial heterogeneity
  
- Spatio-temporal statistics
  - Spatiotemporal observations
  - Spatial temporal autocorrelation;
  - Spatial temporal interaction *et al.*



# Spatial statistics

## □ Research contents

**View of data model** (Cressie 1993)

1. Point pattern
2. Lattice data
3. Geostatistical data

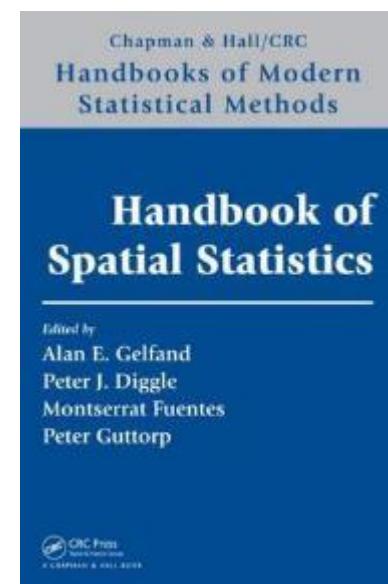
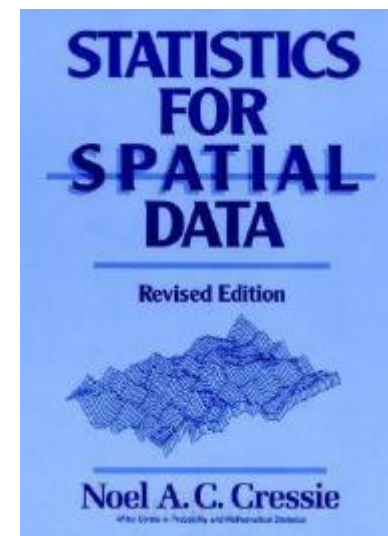
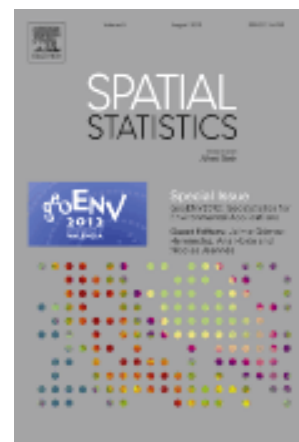
**View of topics** (Gelfand 2010)

1. Continuous spatial variation—including geostatistics , hierarchical model , non-stationary Gaussian random fields
2. Discrete spatial variation —spatial autocorrelation、 disease mapping and spatial econometrics
3. Spatial point pattern—point pattern
4. Spatio-temporal process—ST model and assimilation
5. Addition topic—multivariate, change of support, aggregation and disaggregation

## □ BOOKS

Cressie, N. A. C. (1993). *Statistics for spatial data*. New York, Wiley.

Gelfand, A. E. (2010). *Handbook of spatial statistics*. Boca Raton, CRC Press.



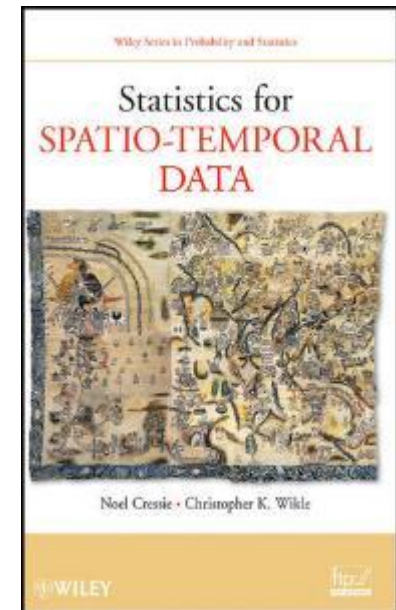
# Spatio-temporal statistics

## □ Research contents

- Based on spatial statistics and time series analysis
- Exploratory spatial data analysis (ESDA)
- Basic models
  - Spatio-temporal covariance model
  - Spatio-temporal kriging
  - Stochastic difference model
  - Spatial process with time series analysis
  - Spatio-temporal point pattern
- Hierarchical dynamic space-time model
- .....

## □ BOOKS

- Cressie, N. A. C. and C. K. Wikle (2011). Statistics for spatio-temporal data. Hoboken, N.J., Wiley.
- Finkenstädt, B., L. Held, et al. (2007). Statistical methods for spatio-temporal systems. Boca Raton, FL, Chapman & Hall/CRC.



# ST statistics application fields

## □ Earth system sciences

- Climate, weather simulation and forecast
- Geology, ocean, atmosphere, etc

## □ GIS, RS, and GPS ( 3S )

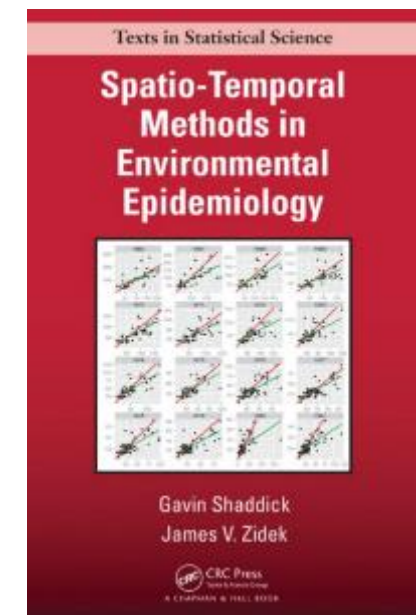
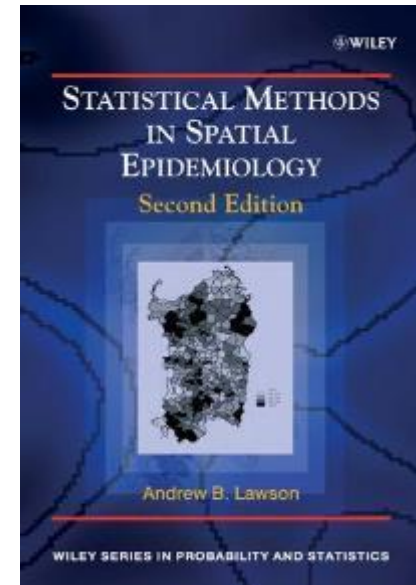
- GIS : Collection, storage, management, analysis and visualize geographic information of time and spatial information
- RS : Satellite and aerial remote sensing, Wireless sensor networks
- GPS : position and tracking

## □ Global change and the ecological environment

## □ Environmental Health

## □ Socio-economic field

- ST regression, STARMA *et al.*



# ST statistics in R

---

- Classified by task view

- TimeSeries

- <http://cran.r-project.org/web/views/TimeSeries.html>

- Spatial

- <http://cran.r-project.org/web/views/Spatial.html>

- Spatiotemporal

- <http://cran.r-project.org/web/views/SpatioTemporal.html>

```
install.packages("ctv")  
library("ctv")  
install.views(" Spatial ")  
update.views(" Spatial ")
```

# GIS Interface

- ❑ rgdal、 rgeos
- ❑ plotKML、 ggmap、 RgoogleMaps
- ❑ OpenStreetMap、 leafletR
- ❑ spgrass6、 rgrass7
- ❑ RPyGeo、 RArcInfo
- ❑ RSAGA



**GRASS GIS**  
The world's leading Free GIS software



**GEOS** Geometry  
Engine  
Open  
Source



**Google**  
Maps



*Leaflet* 



# 互联网地图平台

□ Google Map or Earth : <https://maps.google.com/>



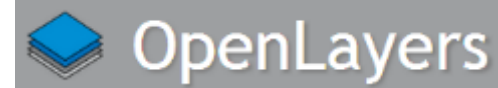
□ Baidu Map : <http://developer.baidu.com/map/>



□ Leaflet : <http://leafletjs.com/>



□ OpenLayers : <http://openlayers.org/>



□ OpenStreetMap : <http://www.openstreetmap.org/>



□ ArcGIS online : <http://www.arcgis.com/home/>



□ Cloudmate : <http://cloudmade.com/>



□ Ploymaps : <http://polymaps.org/>



□ Mapbox: <https://www.mapbox.com/>

□ 地图慧: <http://c.dituhui.com/aboutus>

# 空间数据库技术 — PostgreSQL

## PostgreSQL(Postgres)



- <http://www.postgresql.org>
- 世界上最先进的开源数据库

Limit	Value
Maximum Database Size	Unlimited
Maximum Table Size	32 TB
Maximum Row Size	1.6 TB
Maximum Field Size	1 GB
Maximum Rows per Table	Unlimited
Maximum Columns per Table	250 - 1600 depending on column types
Maximum Indexes per Table	Unlimited

## 空间扩展 : PostGIS

- <http://www.postgresql.org/docs/9.2/static/index.html>
- [http://wiki.postgresql.org/wiki/Main\\_Page](http://wiki.postgresql.org/wiki/Main_Page)

# 空间数据库技术 — PostgreSQL特点

## ●支持多平台

- Linux, UNIX (AIX, BSD, HP-UX, SGI IRIX, Mac OS X, Solaris, Tru64), Windows

## ●丰富的开发接口

- C/C++, Java, .Net, Perl, Python, Ruby, Tcl, ODBC

## ●完全兼容ACID

## ●支持SQL 92和SQL 99

## ●可扩展的数据类型

- geometry

## ●多语言存储过程

- PL/pgSQL , C/C++ , JAVA....

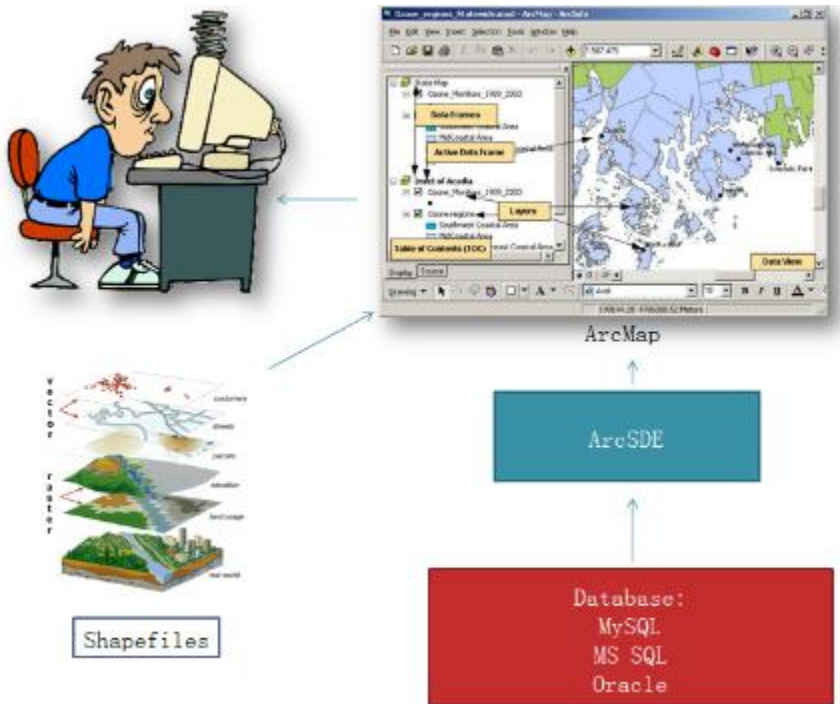
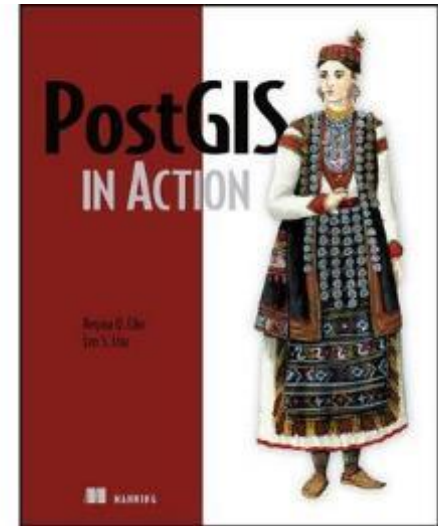
## ●高级特性

- 多版本并发控制 ( MVCC ),异步同步 , 在线热备份 ,

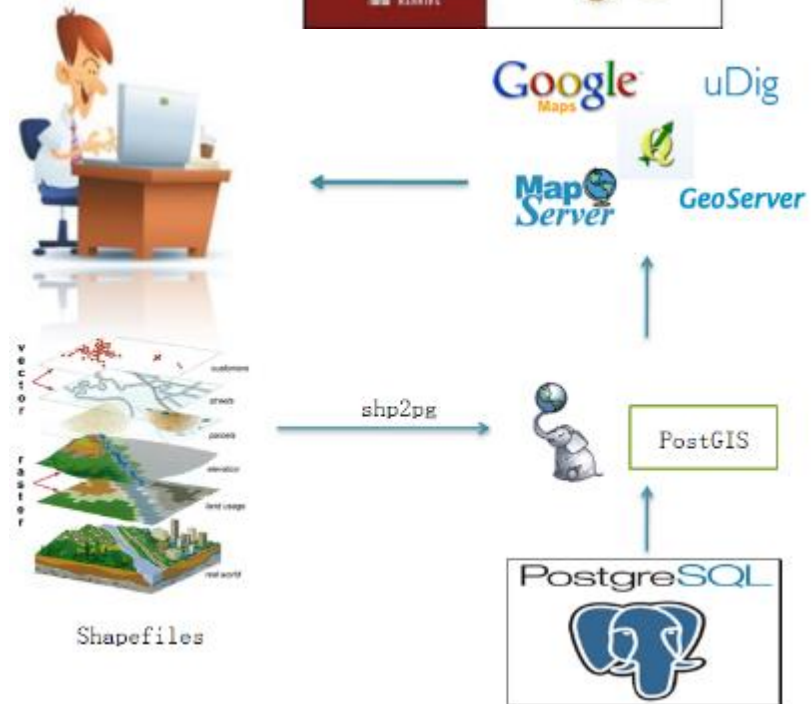


# 空间数据库技术 — PostGIS

- PostGIS (<http://postgis.refractory.net/>) 是对象关系型数据库系统PostgreSQL对空间数据的一个拓展。
- PostGIS提供如下空间信息服务功能：空间对象、空间索引、空间操作函数和空间操作符。



桌面端



网络端

# Cloud Cartography



<https://www.mapbox.com/>



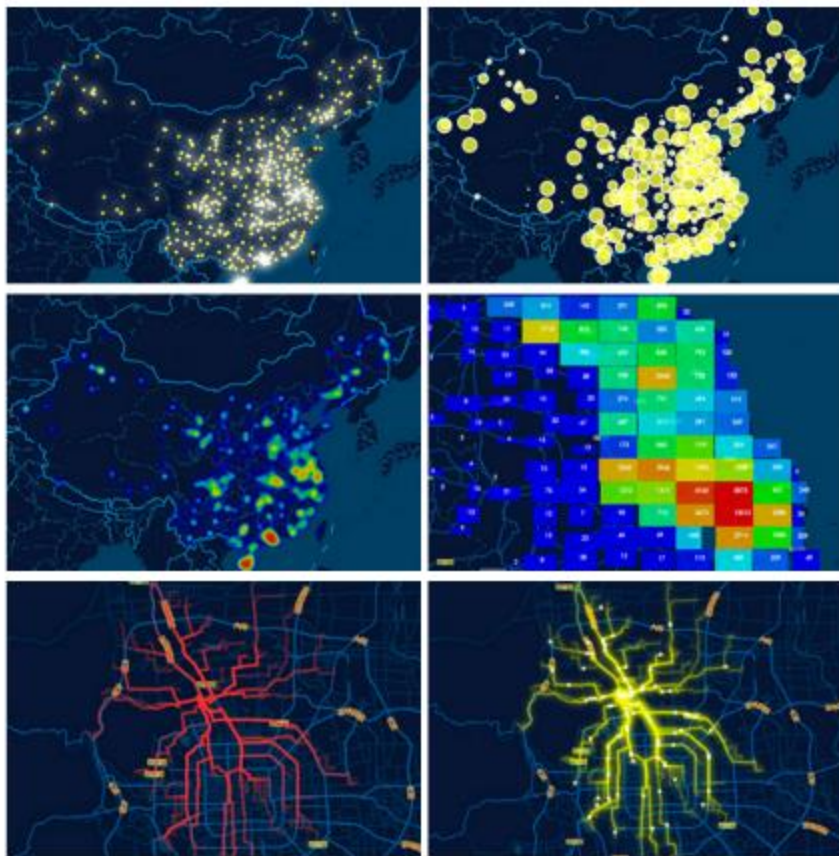
<https://cartodb.com/>



<https://geohey.com/>



# 互联网地图开发




<https://github.com/huiyan-fe/mapv>

## DEVELOPER TOOLS

Use Mapbox to add responsive vector maps with Instant on device rendering to your web and mobile apps. Pick a platform, grab your access token, and get started in minutes.

[Need Help?](#)



### Platforms

- Mapbox GL JS**  
GL maps for web
- Mapbox iOS SDK**  
GL maps for iOS devices
- Mapbox Android SDK**  
GL maps for Android devices
- Mapbox JS**  
Image tile maps for web

### Web services

[View all →](#)

- Styles API**  
Create and edit styles, fonts and colors, and request static images and tiles
- Uploads API**  
Upload custom data directly to your account
- Geocoding API**  
Convert between addresses and geographic coordinates
- Directions API**  
Get driving, walking and biking routes with turn-by-turn directions
- Surface API**  
Inspect and analyze vector data
- Distance API**  
Create a distance matrix between groups of points
- Static API**  
**CLASSIC STYLES ONLY**  
Get static image tiles from classic styles
- Tiles API**  
**CLASSIC STYLES ONLY**  
Get image tiles to use with libraries like Leaflet

### Open source tools

[Learn more →](#)

- Turf.js**  
Geospatial analysis tools in JavaScript
- Tippecanoe**  
Build vector tile datasets from large collections of GeoJSON features
- Rasterio**  
Read and write geospatial raster datasets
- Carmen**  
Vector tile-based geocoder with support for shapable data sources

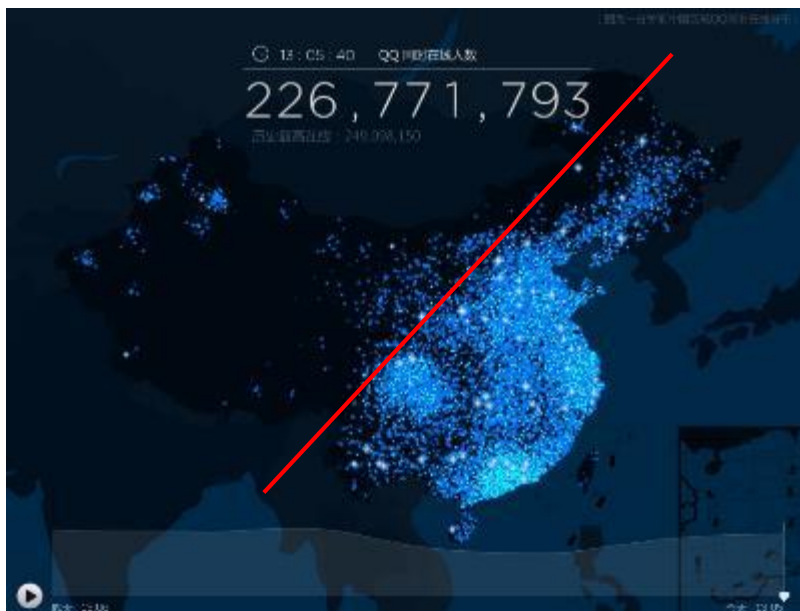
<https://www.mapbox.com/developers/>

# 3 TRACKING CHINESE FOOTPRINT



# 人群足迹与流动

人群足迹与流动的空间特征定量化分析、挖掘与建模，有助于对受人类空间运动驱使的各种复杂现象的深入理解，并在交通、人口、区域发展、城市规划与管理等方面产生重要价值。



腾讯在线用户分布

人地关系



百度迁徙 春运人口流动

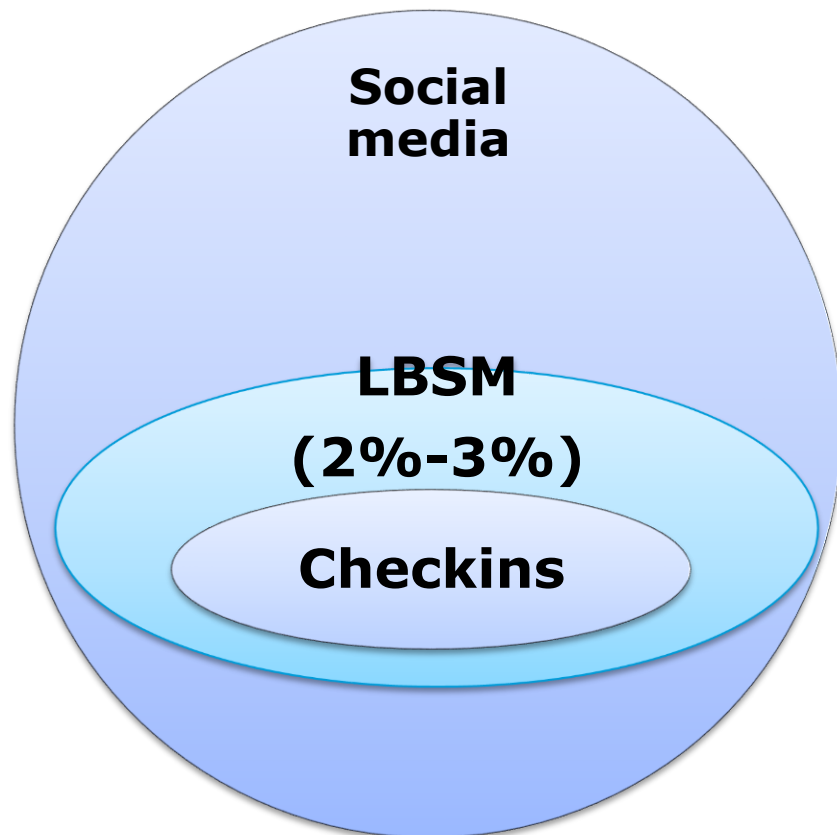


# 中国人空间足迹 — 方法对比

	普查数据空间化	夜晚灯光遥感	地理微博	手机定位
组织方式	↓	↓	↑	↑
数据易获取性	😊	😊	😐	😞
数据代表性	😐	😐	😊	😊
时空粒度	😞	😐	😊	😊
语义特征	😊	😞	😊	😞

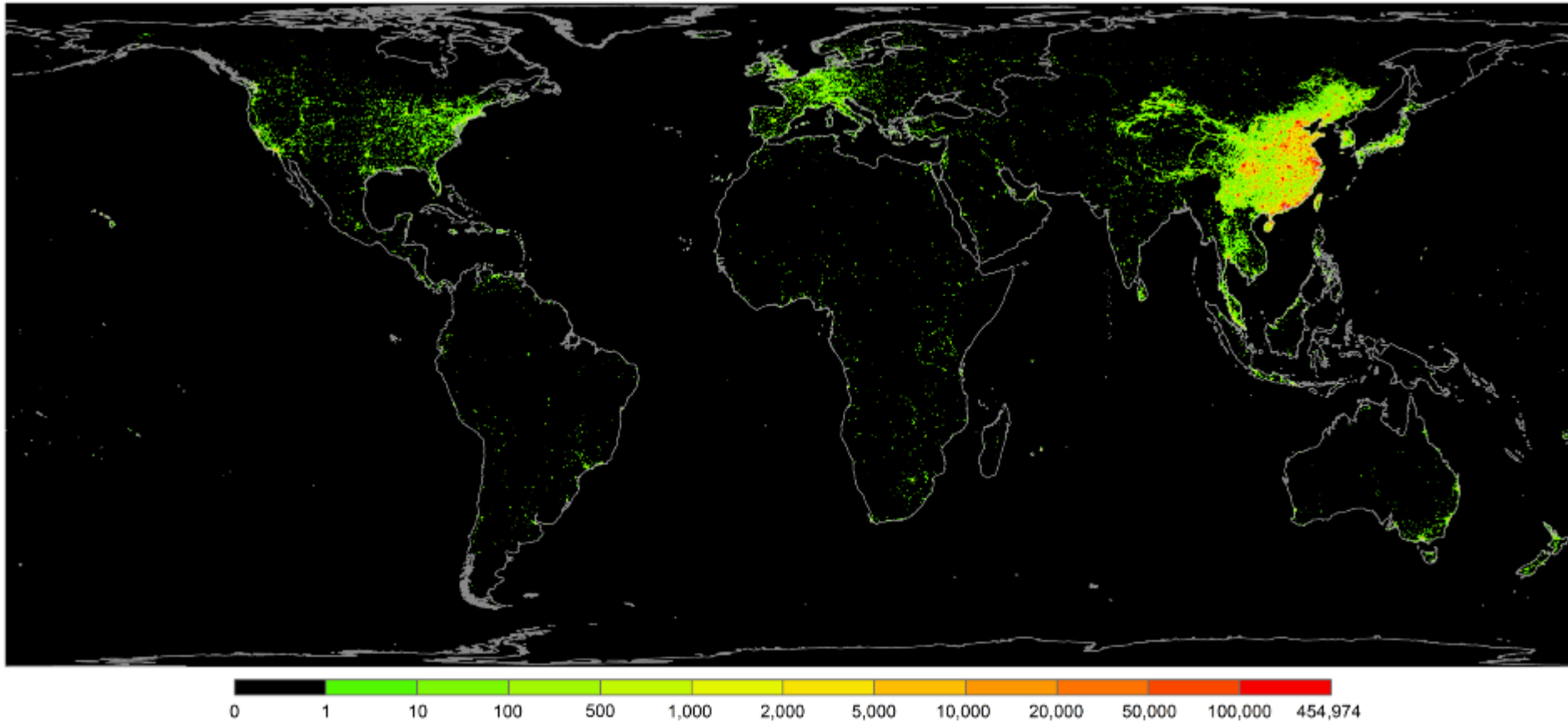
# 数据1：地理微博 (LBSM)

采集了世界范围内**2014**年全年，**超过3亿**条带有地理位置的微博数据  
属性：**Lon、Lat、Time、Text、People attribute, et al.**



# 研究1 : Chinese Footprint on Our Planet

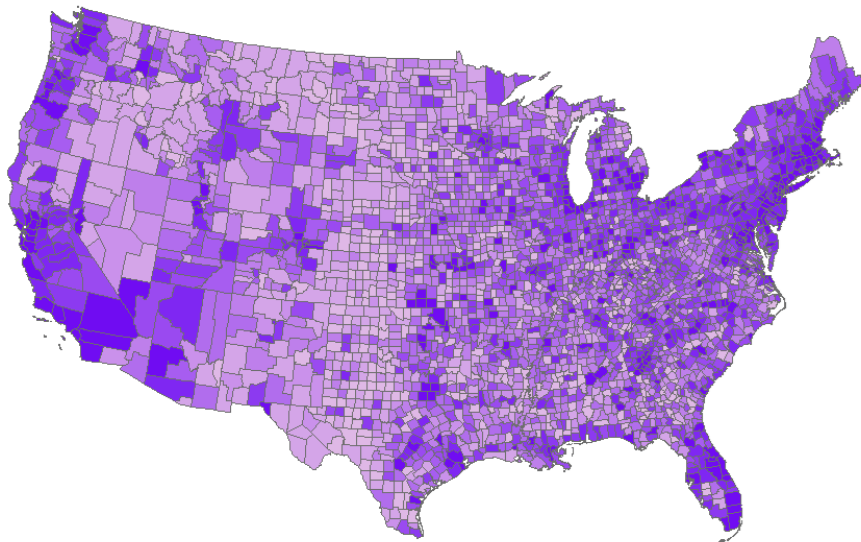
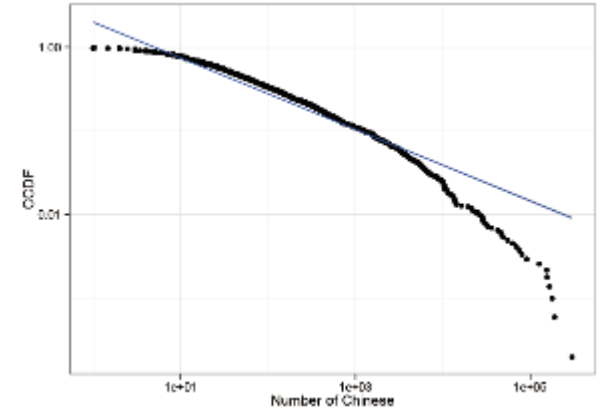
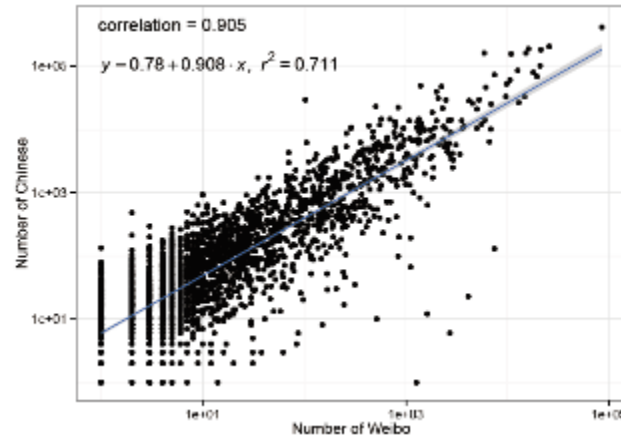
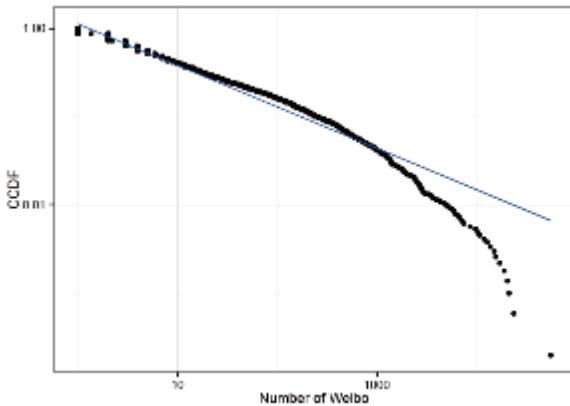
利用微博只有华人使用这一系统有偏性，回答：**Where are the Chinese?**



Liu X, Wang J H\*, *The geography of Weibo*. Environment and Planning A, 2015

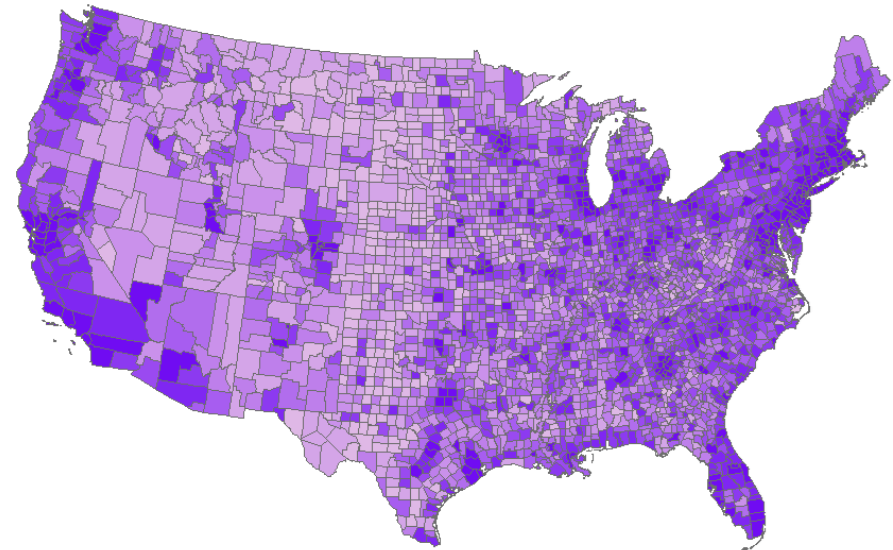
# 地理微博数据 VS. 美国人口普查

**Corr = 0.905, P < 0.001**



**Chinese American**

(Pew Research, based on 2010 Census)

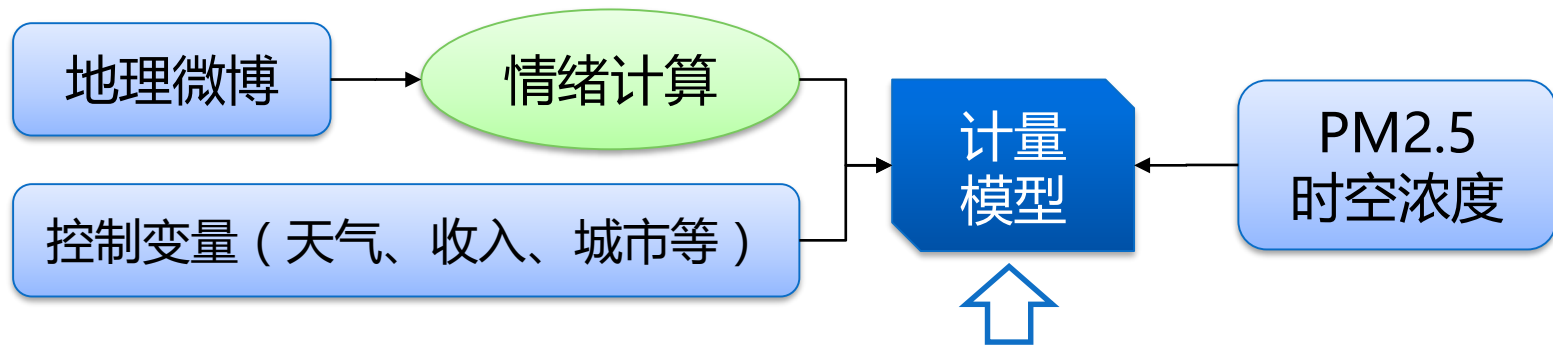


**Chinese American**

(Estimated based on geotagged Weibo)

# 研究2：微博情感与PM<sub>2.5</sub>

- 每升高一个PM<sub>2.5</sub> 浓度标准差将降低 0.05-0.07 个标准差情绪指数
- 每升高一个PM<sub>2.5</sub> 浓度标准差带来的情绪影响将需要提高0.56个标准差的  
城市平均工资来补偿

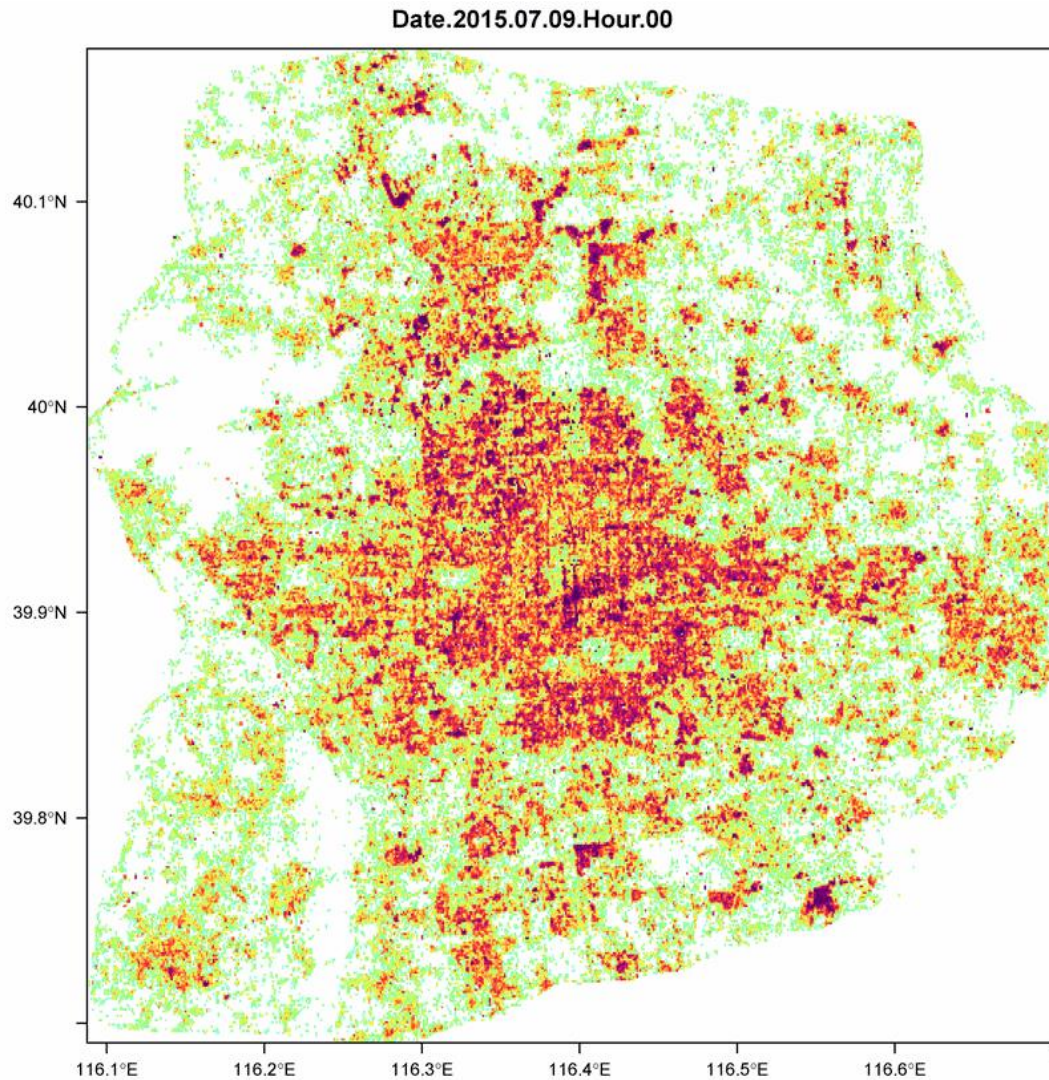


$$SENTIMENT_{it} = \alpha_0 + \alpha_1 PM2.5_{it} + \alpha_2 X_{it} + T_t + \gamma_i + \varepsilon_{it}$$

Zheng, S., **Wang, J.H.**, Kahn, M.E., Sun, C. & Zhang, X. **The Impact of Air Pollution on Subjective Well Being in Urban China: Evidence from Social Media Content**

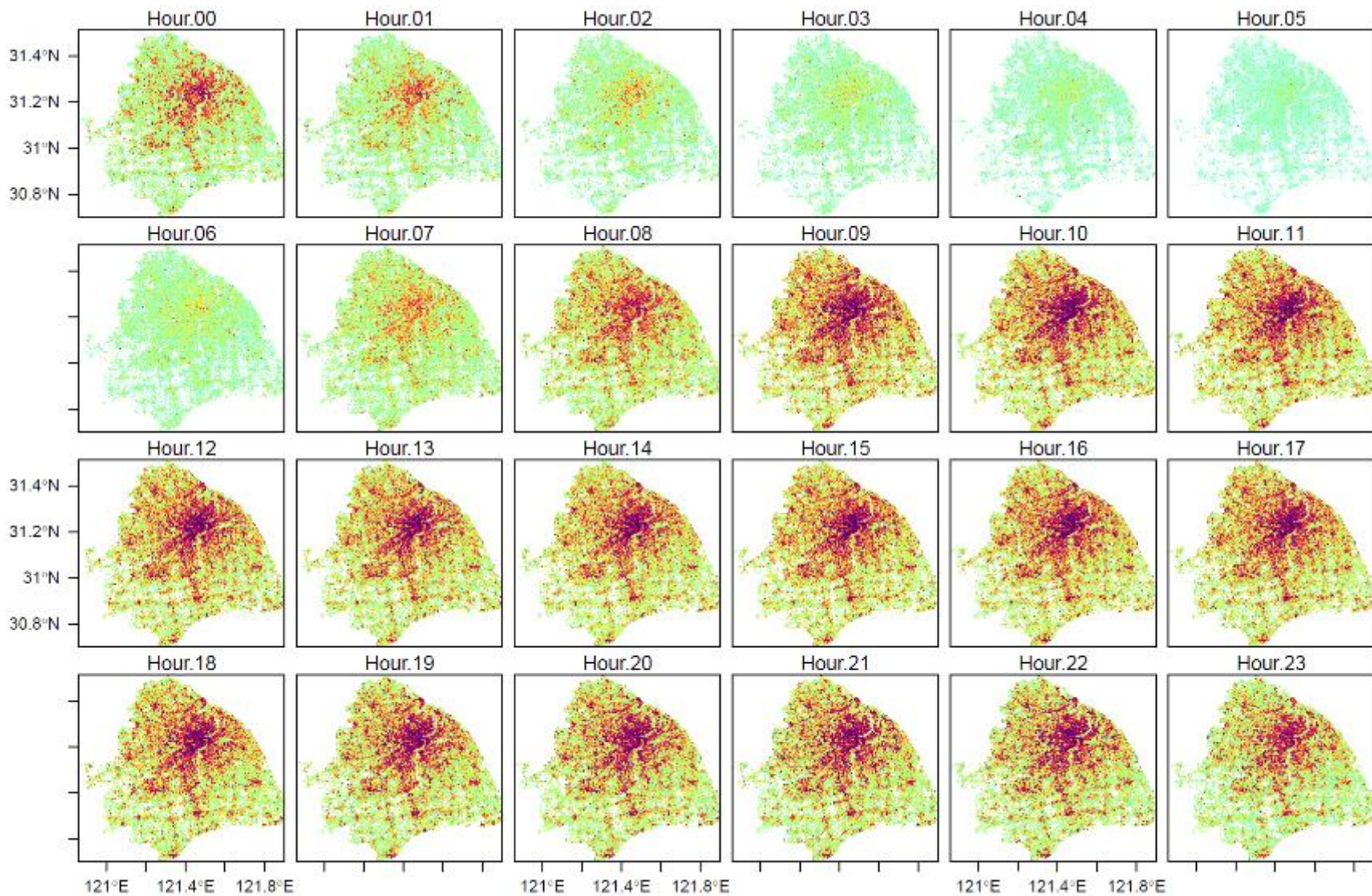


# 北京手机定位强度时空变化



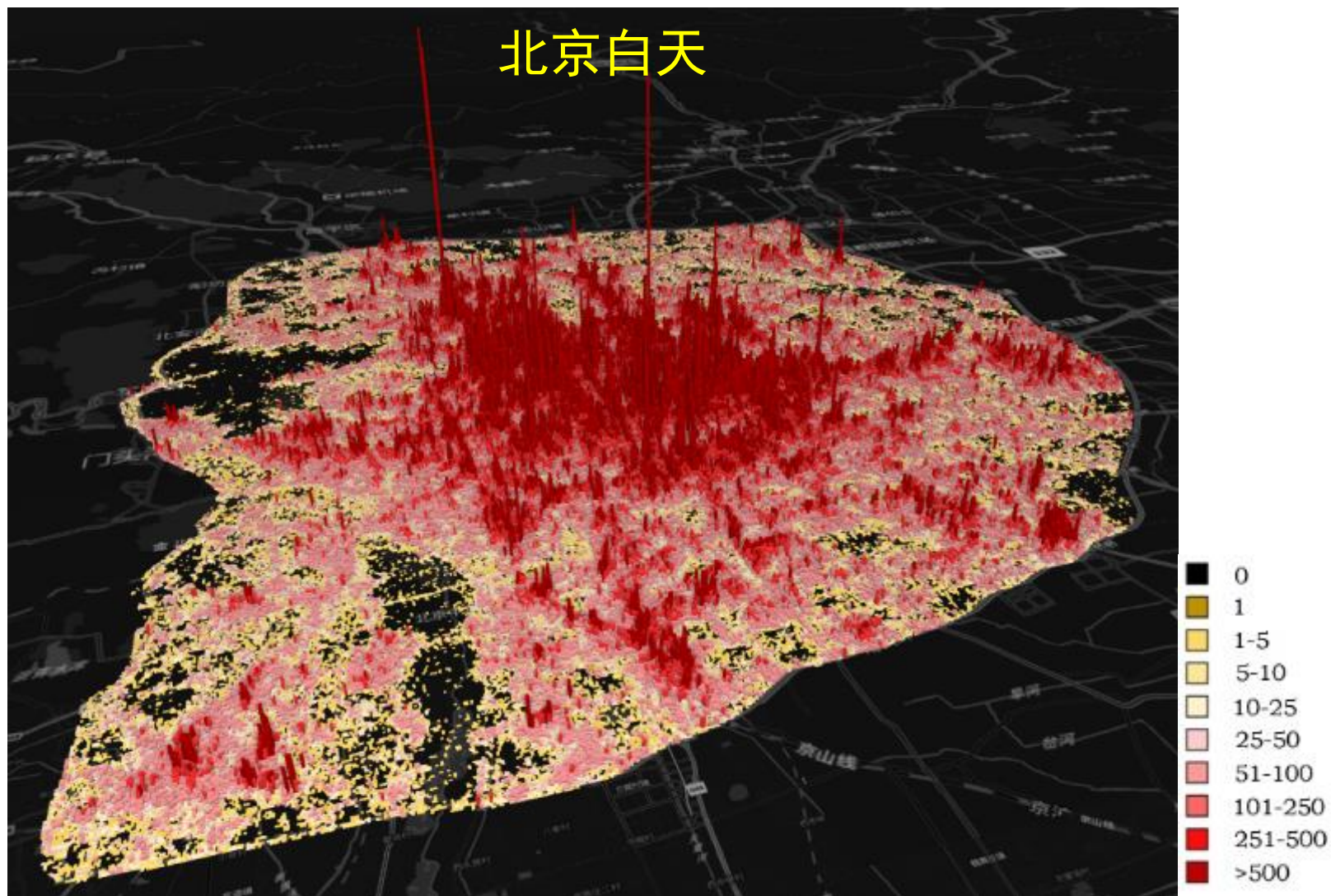
MP Positioning Dynamics (100 m)

# 上海手机定位强度时空变化

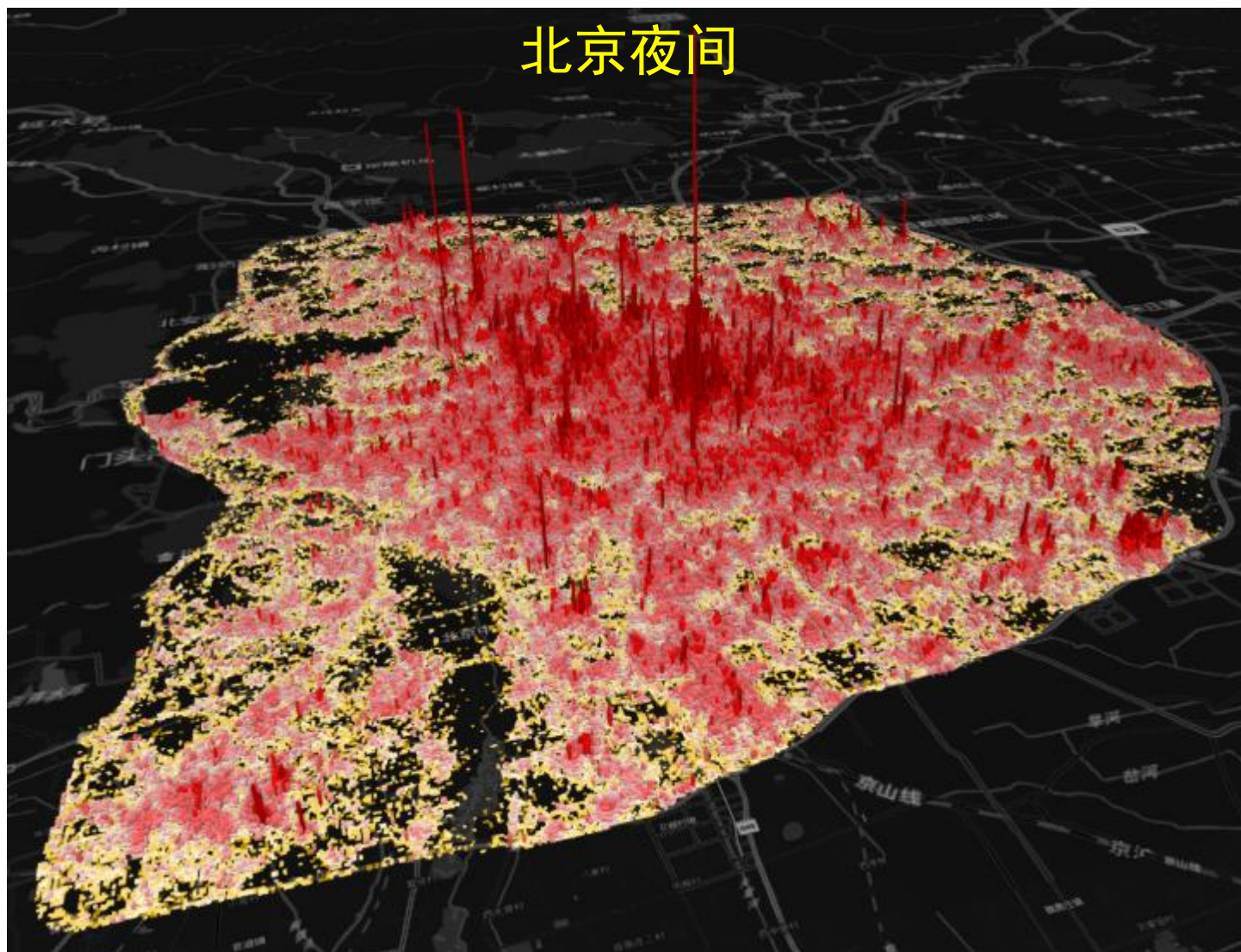




# 研究3：人口分布动态制图

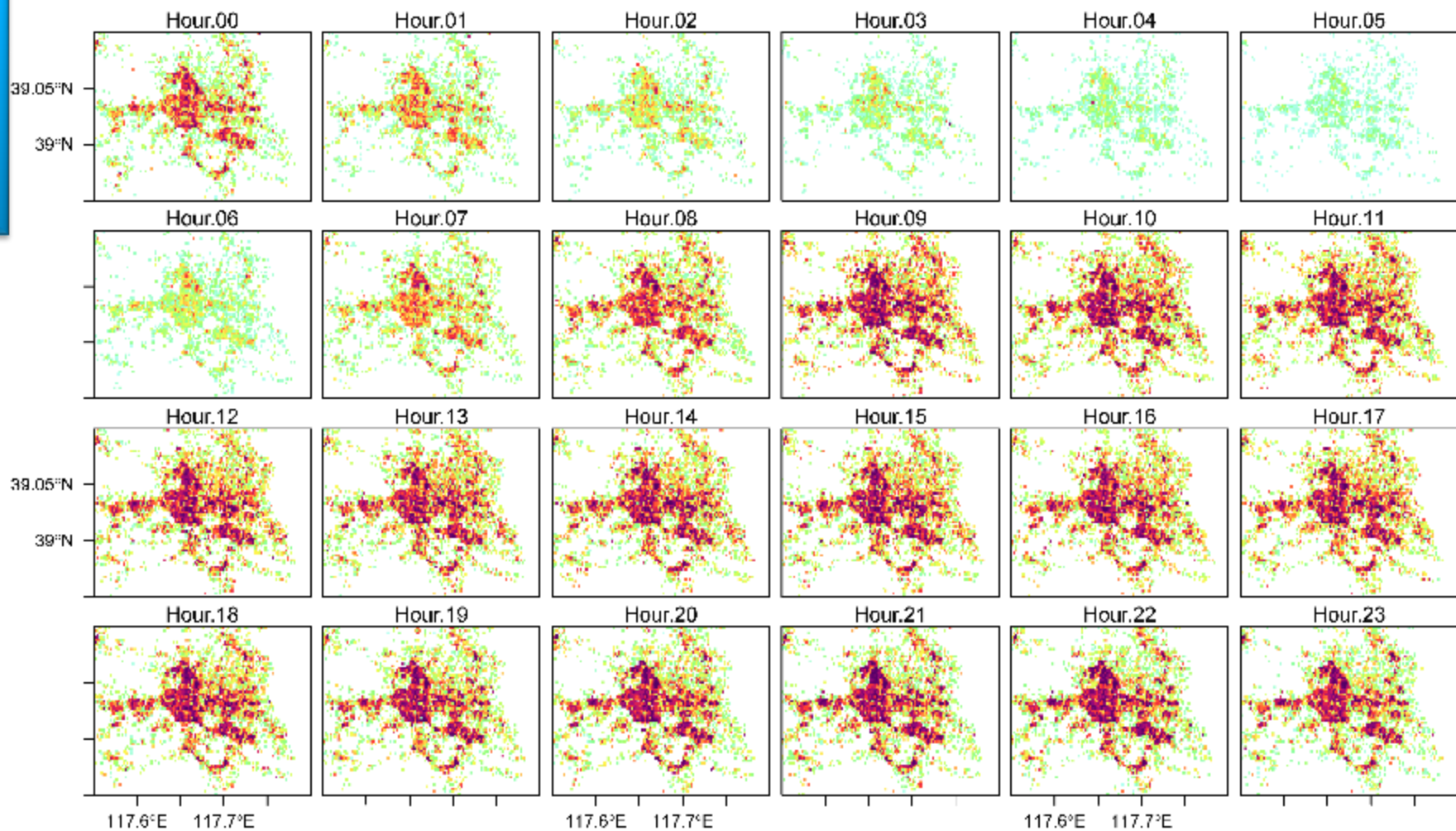


# 研究3：人口分布动态制图

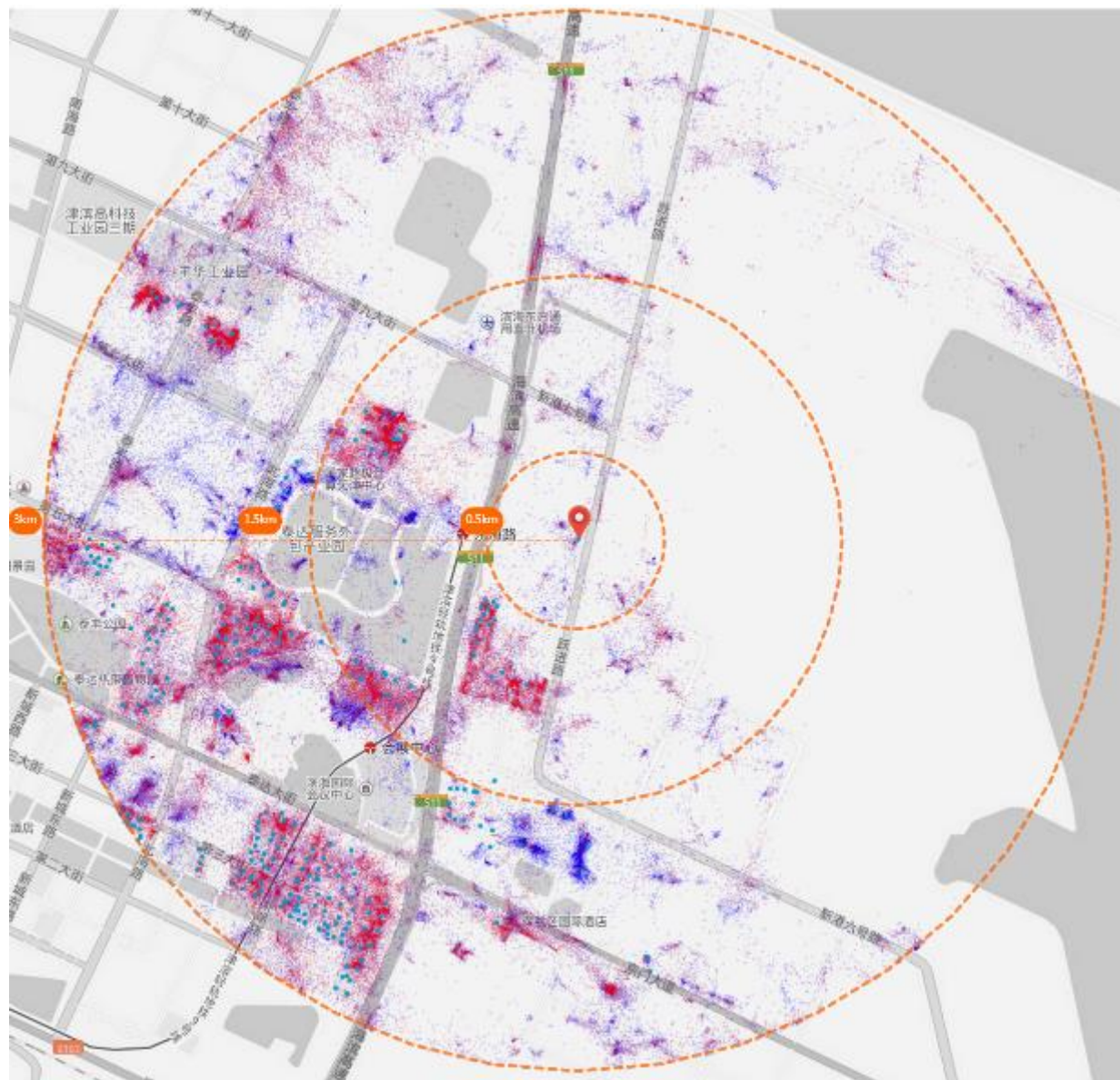


# 手机定位数据估算人群分布

天津滨海新区



# 天津爆炸案之前手机定位请求分布



利用手机GPS定位请求数据获取的天津滨海爆炸案之前的手机定位数据分布。

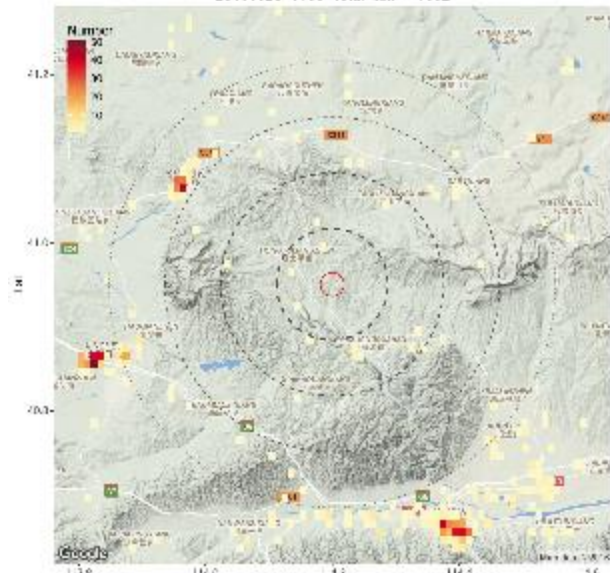
其中蓝色为该区域的工作人口、红色为该区域的常住人口。

## 图例

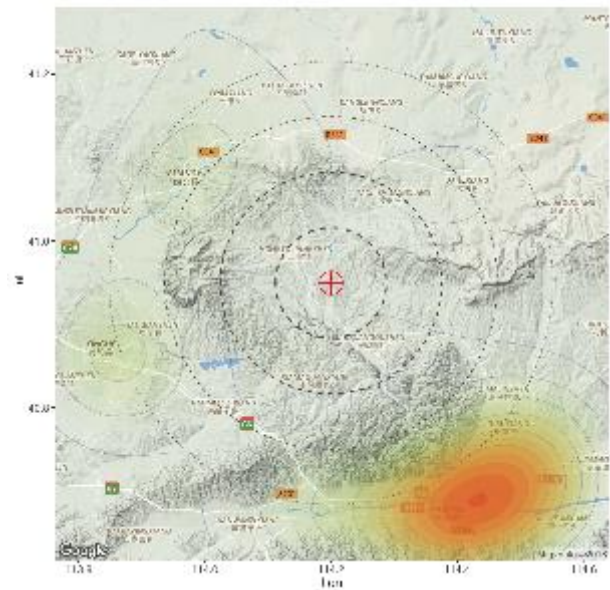
-  爆炸点
-  工作人口
-  居住人口
-  住宅区
-  辅助线

# 基于手机定位的地震灾情发现与评估

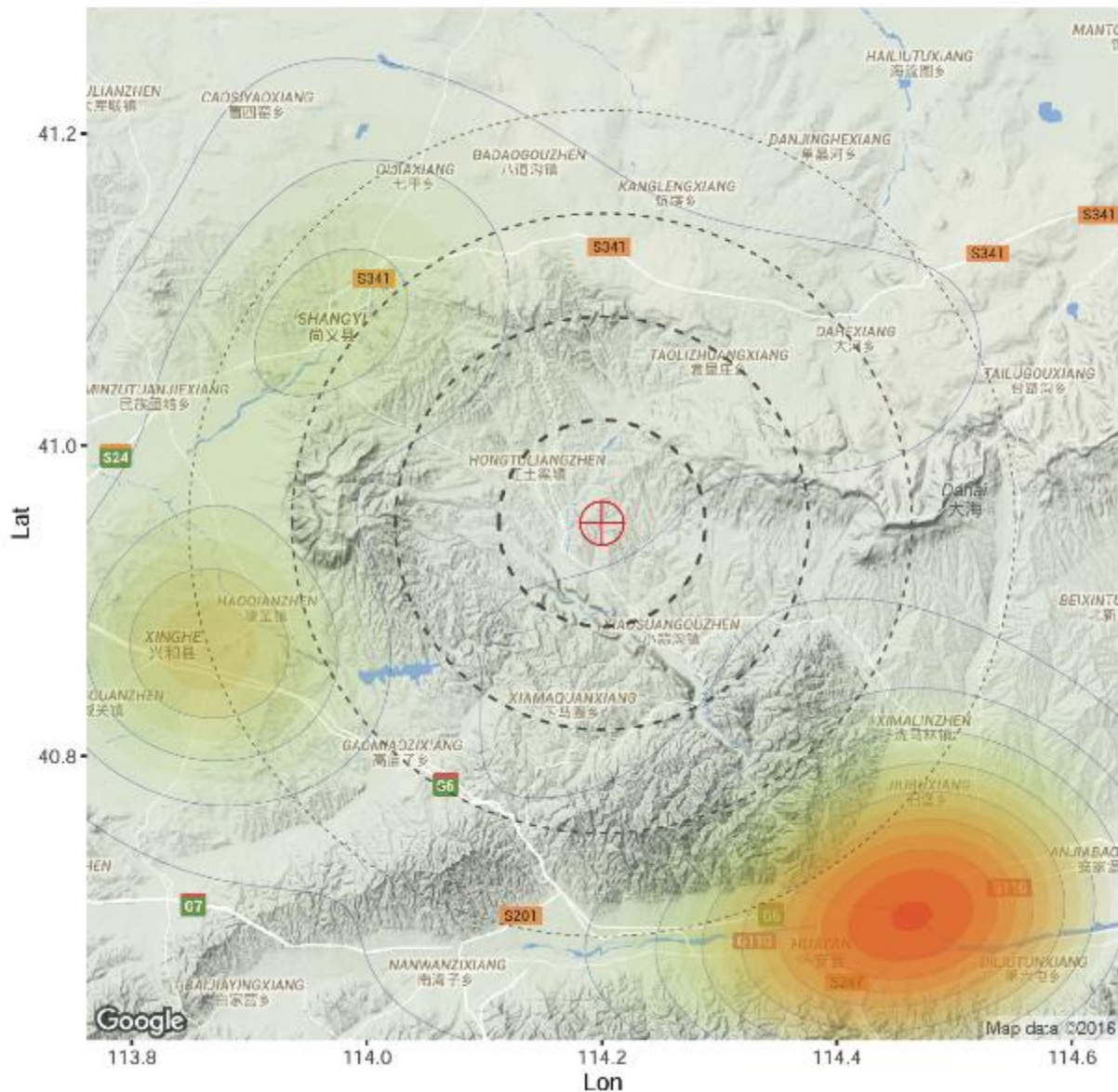
20160623-0700 Total Num = 1002



20160623-0700 Total Num = 1002



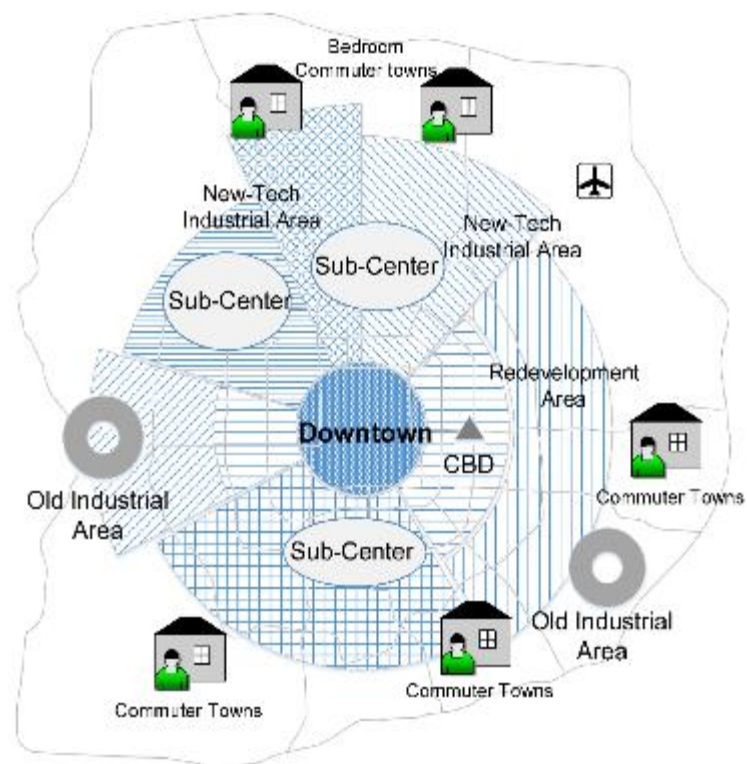
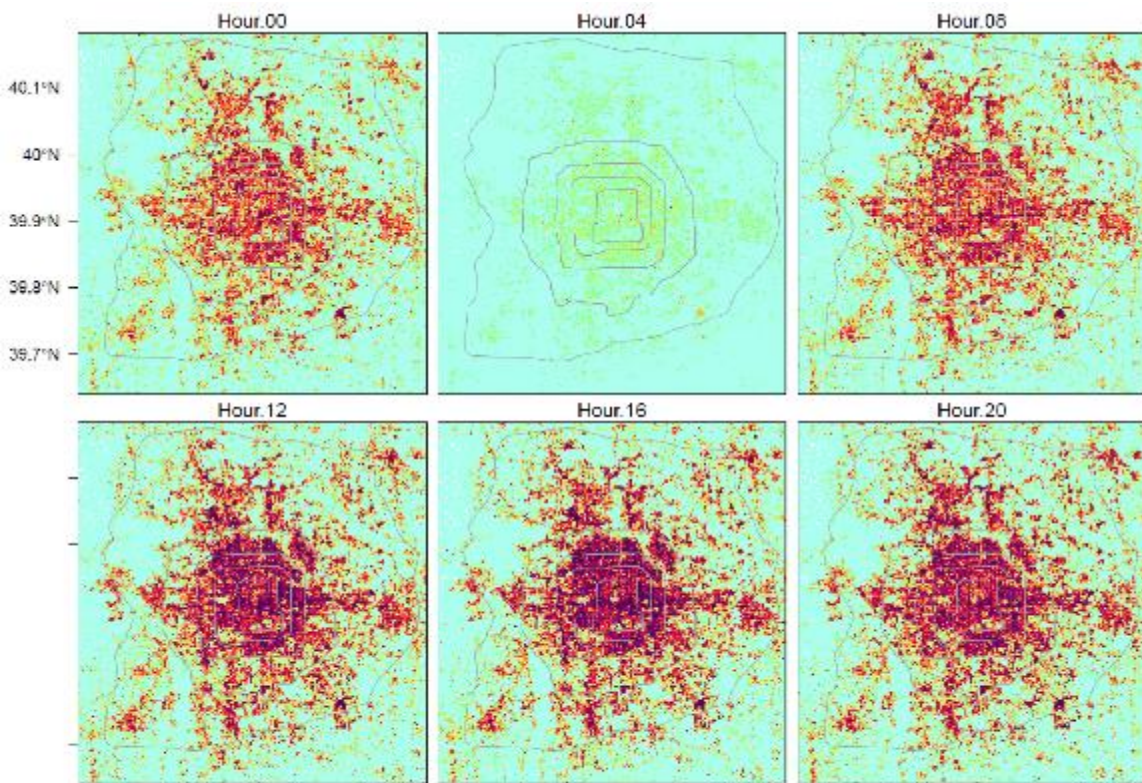
20160623-1000 Total Num = 1106



# 研究4：人类活动、土地利用与城市活力

1. Jacobs (1961); Yue et al. (2016) 混合土地利用与城市活力
2. Glaeser et al. (2001) 消费城市与城市活力
3. Jacobs-Crisioni et al., (2014) 人类活动多样性与城市活力

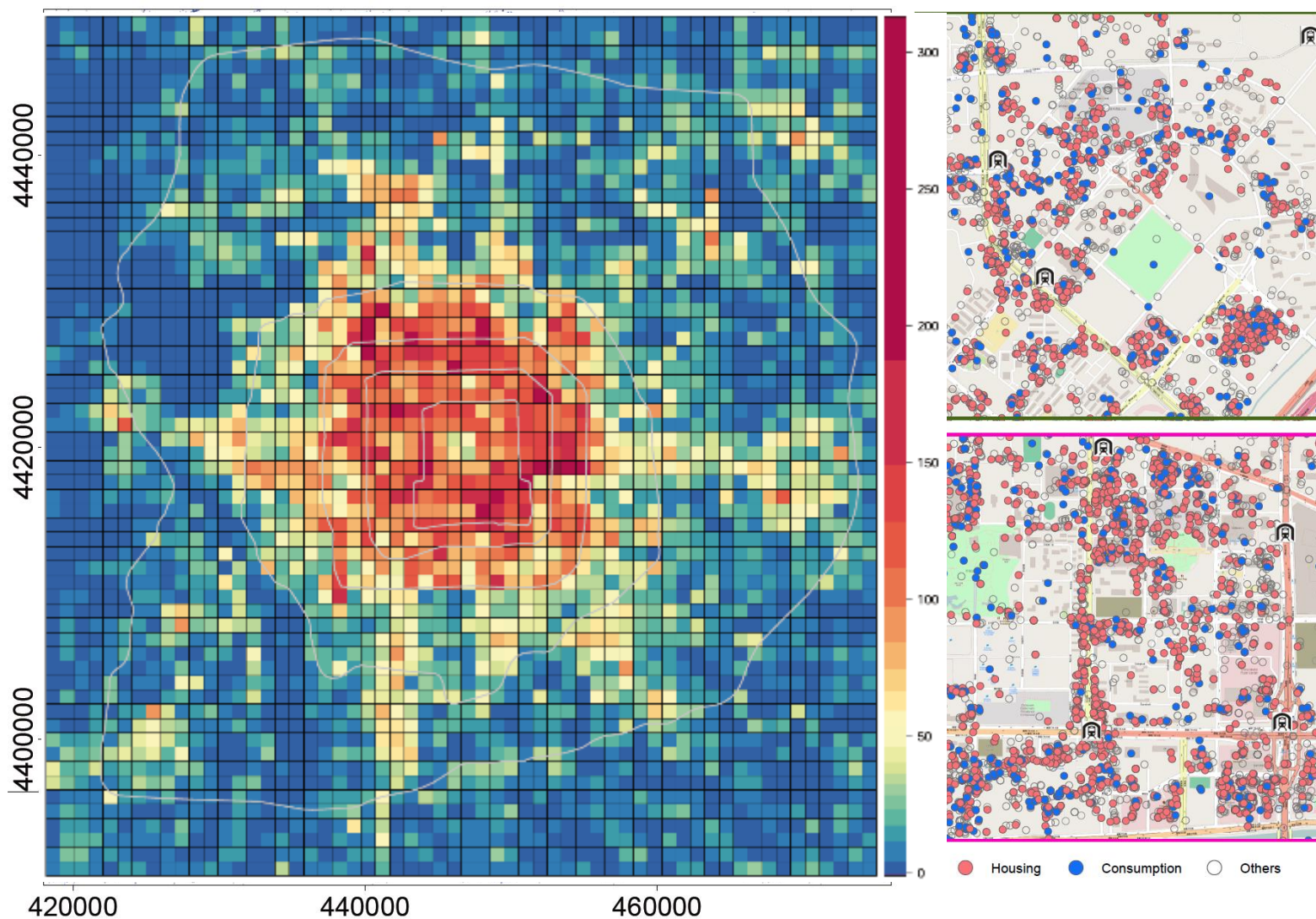
交叉  
研究



Wu W, Wang J H\*, Li C, Wang X. The Geography of City Liveliness and Land Use Configurations: Evidence from Location-based Big Data in Beijing. LSE Working Paper

# 土地功能类型

利用大众点评网的POIs类型等信息表征城市土地功能



# 时空计量模型构建

## □ 时空变系数模型 ( Spatially and Temporally Varying Coefficient Model )

$$y_{i,t} = \sum_{k_1=1}^{p_1} x_{i,t,k_1} \beta_{k_1}(a_i, b_i, \tau_t) + \sum_{k_2=1}^{p_2} z_{i,t,k_2} \gamma_{k_2}(a_i, b_i) + \sum_{k_3=1}^{p_3} g_{i,t,k_3} \theta_{k_3}(\tau_t) + \sum_{k_4=1}^{p_4} h_{i,t,k_4} \zeta_{k_4} + \varepsilon_{i,t}$$

时空系数项      纯空间系数      纯时间系数      非变化系数

$\gamma = \theta = \zeta = 0$  1. Geographically and Temporally Weighted Regression (**GTWR**)

$\beta = \theta = \zeta = 0$  2. Geographically Weighted Regression (**GWR**)

$\beta = \gamma = \zeta = 0$  3. Time varying coefficient model (**TWR**)

$\beta = \theta = 0$  4. Semi spatially varying coefficient model

$\beta = \gamma = \theta = 0$  5. linear regression model (**OLS**)

模型设定

$$\ln y_{it} = \beta_0(a_i, b_i, \tau_t) + \ln C_{it} \beta_1(a_i, b_i, \tau_t) + \ln H_{it} \beta_2(a_i, b_i, \tau_t) + \sum_k control_{it,k} \alpha_k(a_i, b_i, \tau_t) + \varepsilon_{it}$$



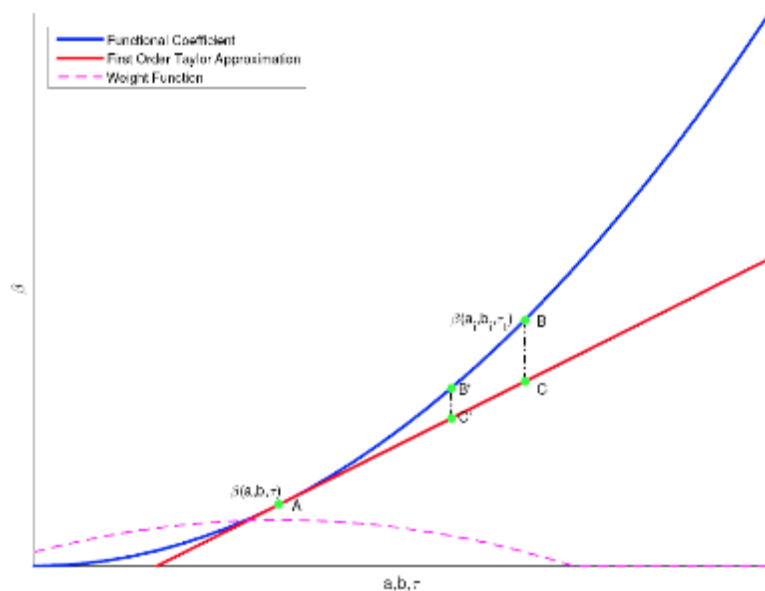
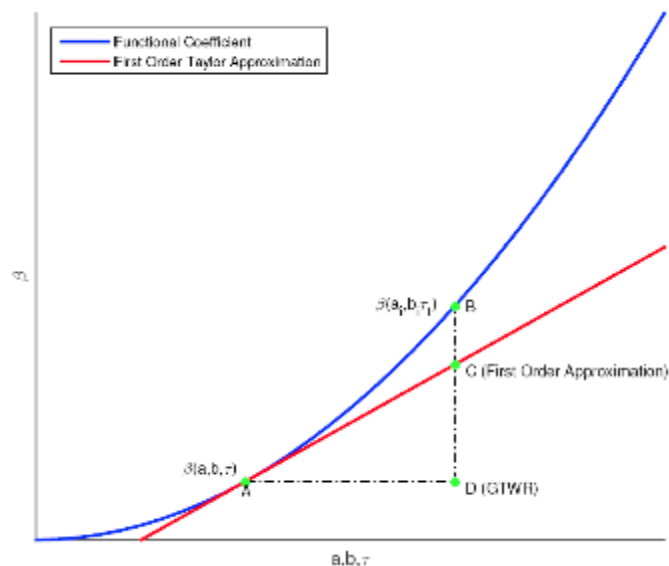
# 时空计量模型构建

- 参数估计：采用局域方法与一阶泰勒展开近似联合估计方法

$$\begin{aligned}\beta(a_i, b_i, \tau_t) &\approx \beta(a, b, \tau) + \frac{\partial\beta(a, b, \tau)}{\partial a}(a_i - a) + \frac{\partial\beta(a, b, \tau)}{\partial b}(b_i - b) + \frac{\partial\beta(a, b, \tau)}{\partial \tau}(\tau_t - \tau) \quad (2.3) \\ &:= \beta(a, b, \tau) + \beta^{(a)}(a_i - a) + \beta^{(b)}(b_i - b) + \beta^{(\tau)}(\tau_t - \tau)\end{aligned}$$

$$y_{i,t} = \beta_0(a_i, b_i, \tau_t) + \sum_{k=1}^p x_{i,t,k} \beta_k(a_i, b_i, \tau_t) + \varepsilon_{i,t}, \quad \text{for } i \in S_{h_1}, t \in I_{h_2} \quad (2.5)$$

$$\approx \sum_{k=0}^p x_{i,t,k} \beta_k(a, b, \tau) + \beta_k^{(a)} x_{i,t,k} (a_i - a) + \beta_k^{(b)} x_{i,t,k} (b_i - b) + \beta_k^{(\tau)} x_{i,t,k} (\tau_t - \tau) + \varepsilon_{i,t}$$



# 关键变量描述

- **目标变量**：人类活动强度：工作日24小时变化；周末24小时变化
- **主要解释变量**：居住类设施、消费类设施
- **主要控制变量**：公共交通便利度、路网密度、到就业中心的距离、土地利用混合度等

Table 3.1: Descriptive statistics of the key variables

Variable Name	Definition	Mean	SD	Min	Max
Activity density at weekday(ywd)	mobile positioning data per area unit at weekday (in 1000 people)	5.415	10.075	0.004	225.924
Activity density at weekend(ywn)	mobile positioning data per area unit at weekend,(in 1000 people)	5.322	9.459	0.004	218.127
Housing amenities (H)	Number POIs related to residential complex compounds in each unit	1.372	3.209	0	32
Consumption amenities(C)	Number of POIs related to entertainment and leisure services	75.762	174.954	0	2177
Stations	Number of subway stations per unit	0.121	0.833	0	42
Road Density	Total road lengths per unit (in kilometers)	5.711	4.425	0	23.281
Distance to CBD	Straightline distance from each grid unit to the CBD (in kilometers)	22.441	9.000	0	44.926
Other amenities	Number POIs that do not fall in the category of Housing and Consumption	18.149	41.359	0	546

# 主要结果与结论

Table 4.1: The relationship between human activity intensity and land use configurations (Pooled Regression Estimates)

	(1)	(2)	(3)	(4)	(5)	(6)
Housing*Transport					0.0202*** (6.32)	0.0180*** (7.38)
Consumption*Transport					-0.0901*** (-18.74)	-0.0880*** (-20.49)
Housing*central city					-0.204*** (-29.95)	-0.201*** (-28.65)
Housing*Inner Suburb					0.0460*** (4.95)	0.0495*** (5.45)
Consumption*central city					-0.290*** (-29.09)	-0.299*** (-36.91)
Consumption*Inner Suburb					-0.247*** (-18.34)	-0.266*** (-19.26)
Constant	-1.612*** (-8.39)	-1.610*** (-8.29)	-0.247 (-1.17)	-0.205 (-1.00)	-0.310 (-1.49)	-0.270 (-1.33)
N	83520	83520	83520	83520	83520	83520

*t* statistics in parentheses(Cluster Robust Error)

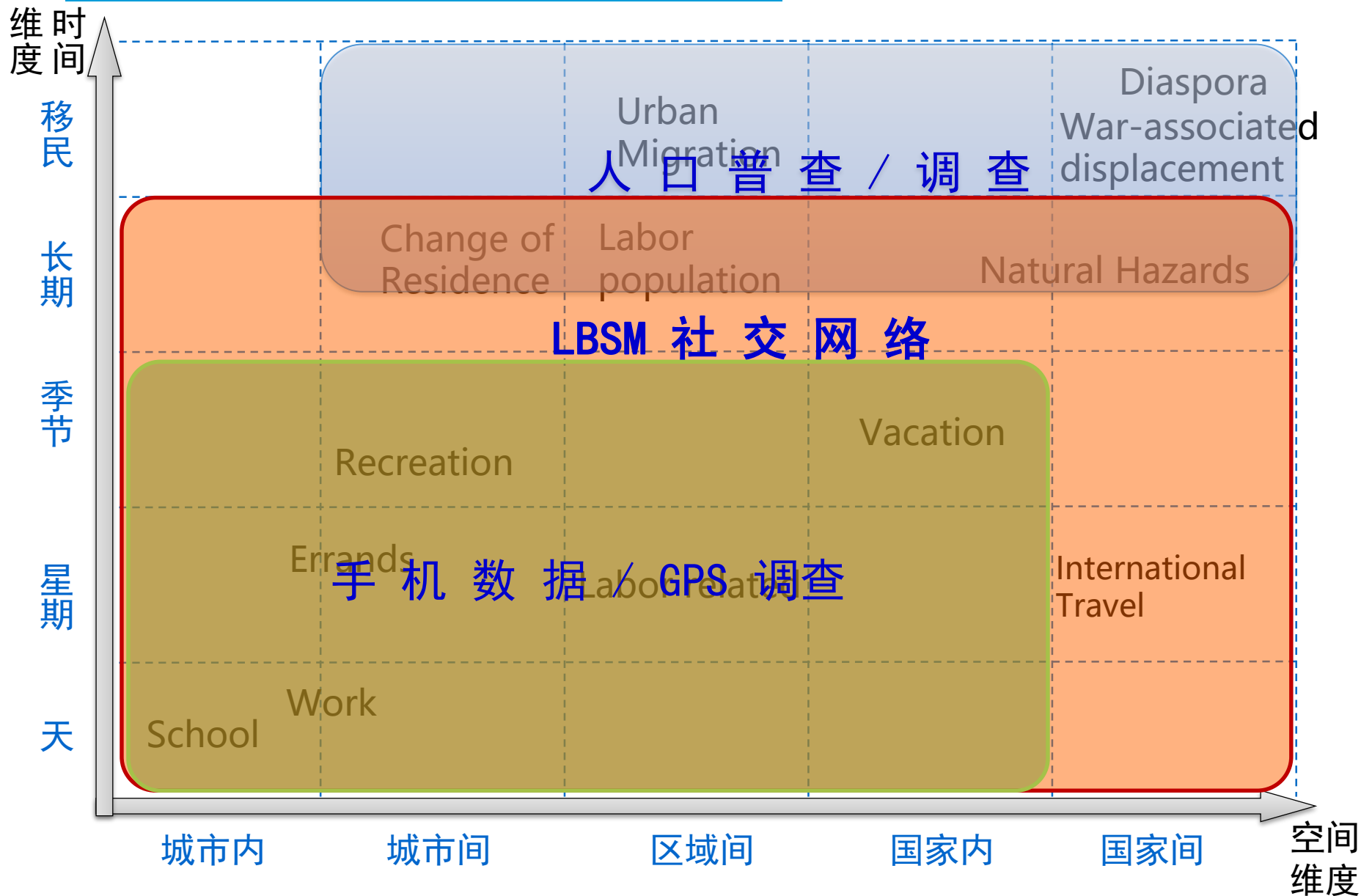
\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

1. 实证简·雅各布斯的城市活力理论
2. 混合土地利用对消费的正面影响作用
3. 封闭的社区对城市活力的负面影响作用与机制
4. 城市内部人类活动(生活-工作-娱乐)的时空不确定性

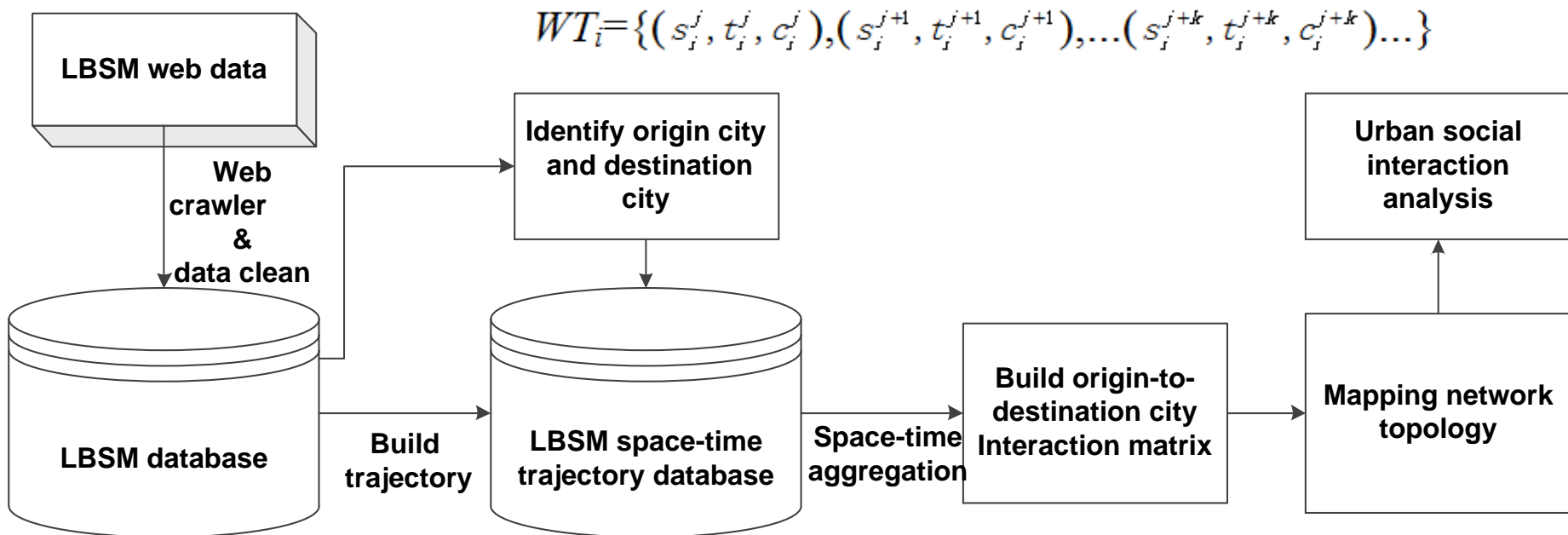
# 4 MODELING CHINESE MOBILITY



# 人群移动研究方法



# 面向LBSM人群移动计算框架

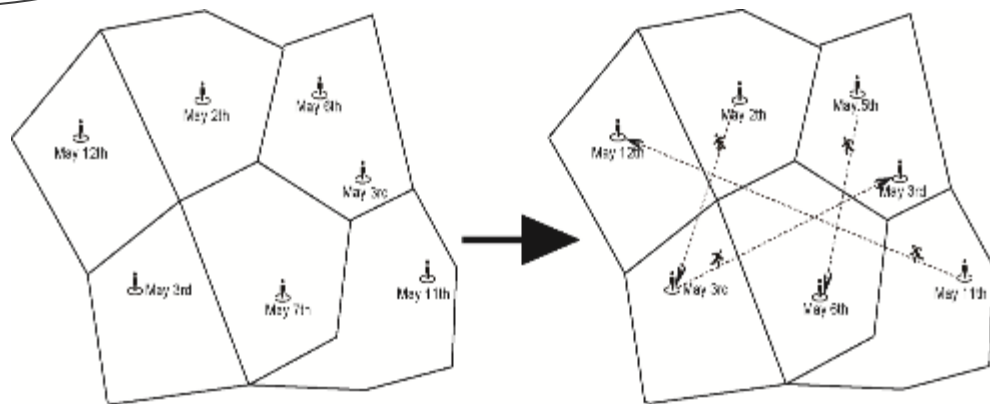


人口流动

$$OutC(p_m) = \sum_{i=1}^k OutC(p_{m,i}) - \sum_{i=1}^k F(p_{m,i})$$

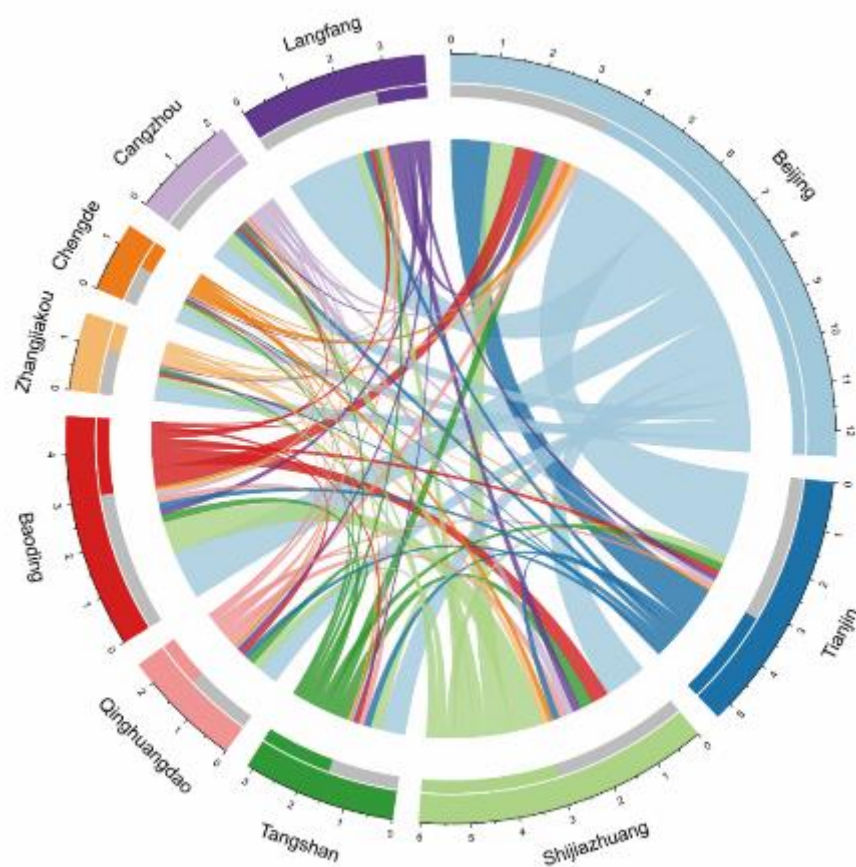
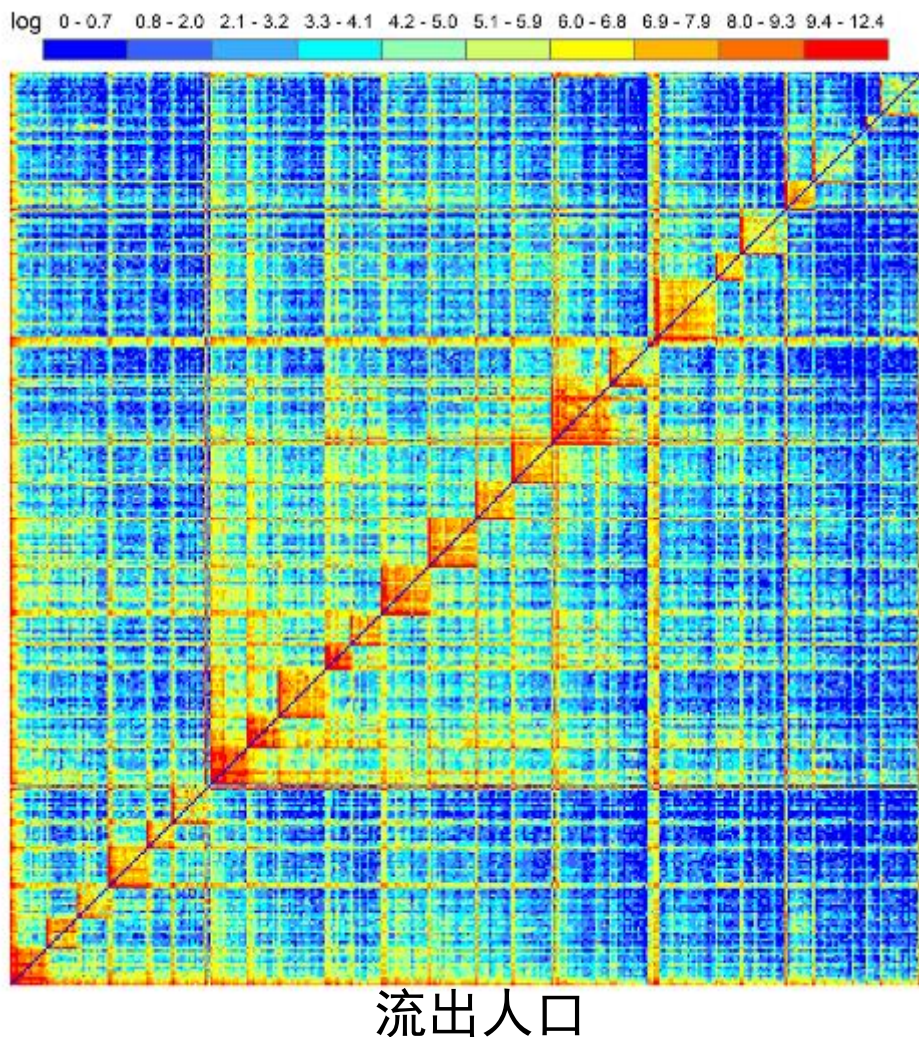
$$InC(p_m) = \sum_{i=1}^k InC(p_{m,i}) - \sum_{i=1}^k F(p_{m,i})$$

$$F(p_1, p_2) = \sum_{i=1}^k \sum_{j=1}^k F(p_{1,i}, p_{2,j})$$

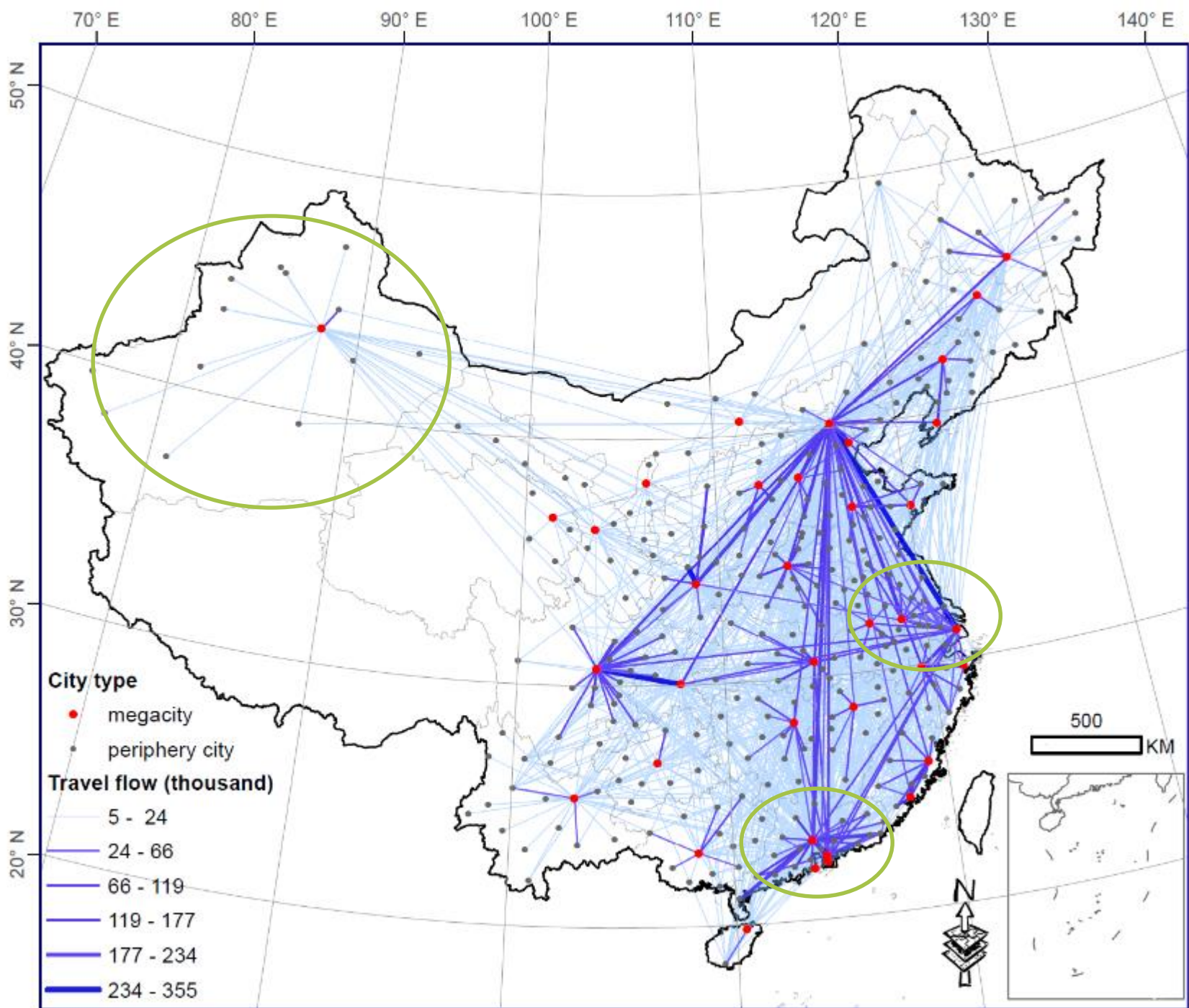


# 城际人口流动矩阵

基于以上框架，利用LBSM提取中国地级以上城市人口流动OD矩阵

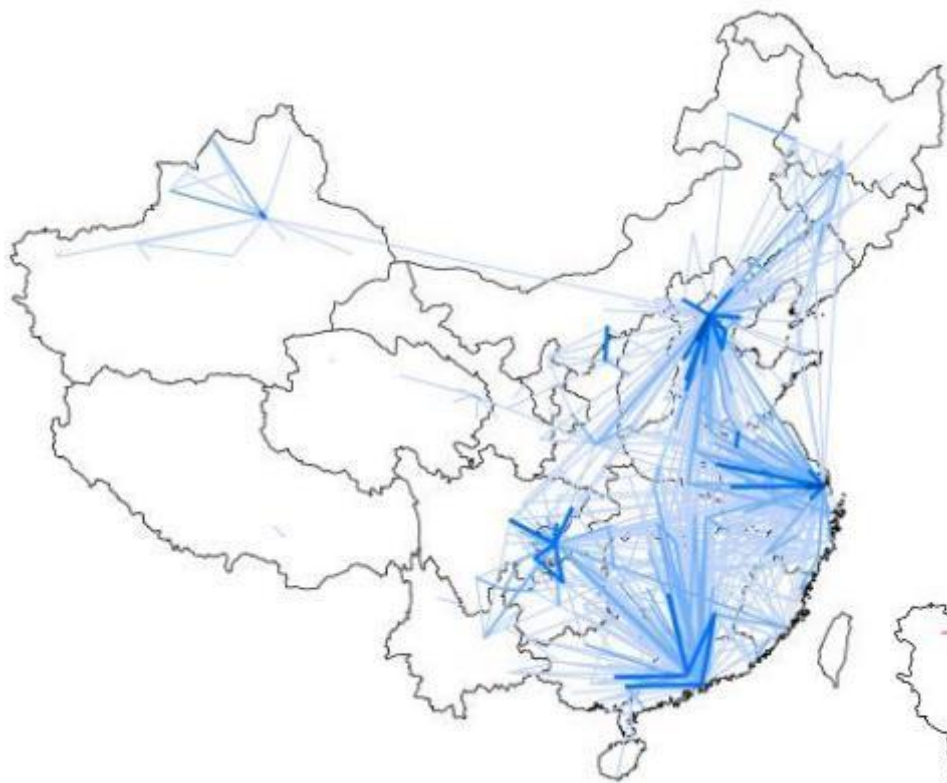


弦图：京津冀人口流动



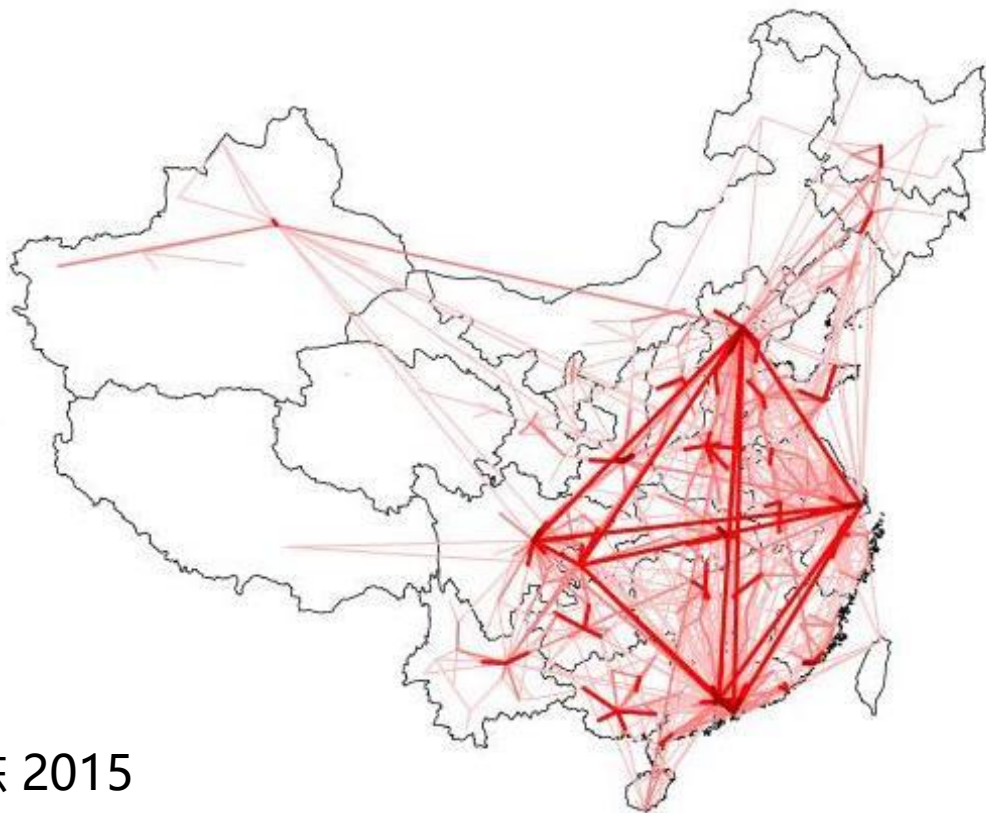


# 与其他人流数据比对



百度迁徙 人流网络

腾讯春运 人流网络

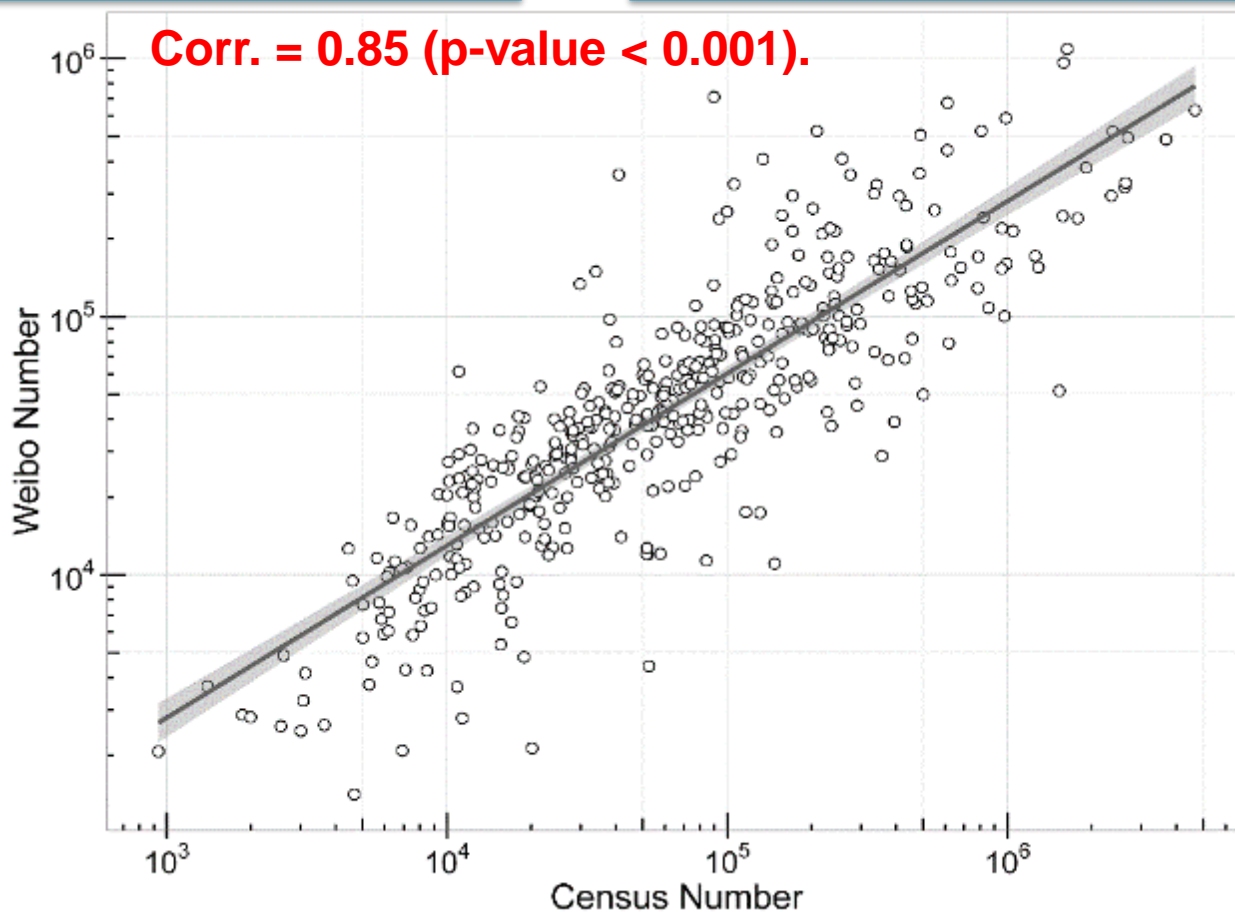
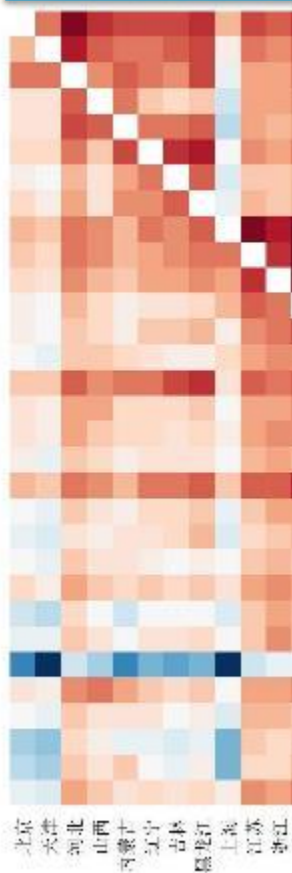


# 与人口普查中流动人口交叉检验

数据代表性 (可信度) 交叉检验

6<sup>th</sup> 人口普查省级流动人口

Weibo提取省级人口流动



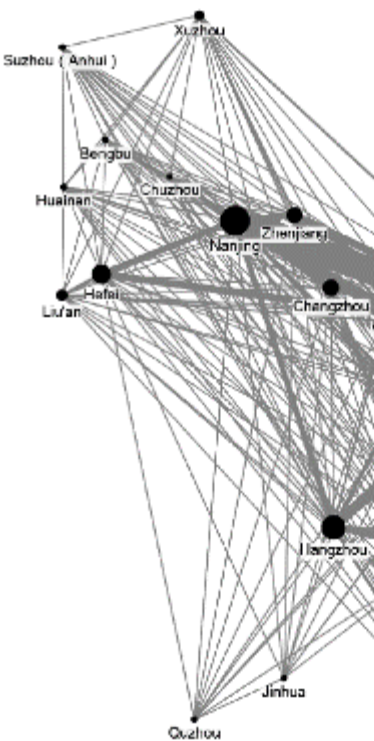
# 与基于公共交通网络的交叉检验

基于长三角地区的公共交通网络，推算出的城际人流数据(Zhang & Wang 2016);  
基于微博提取的城际人流网络，进行交叉验证。

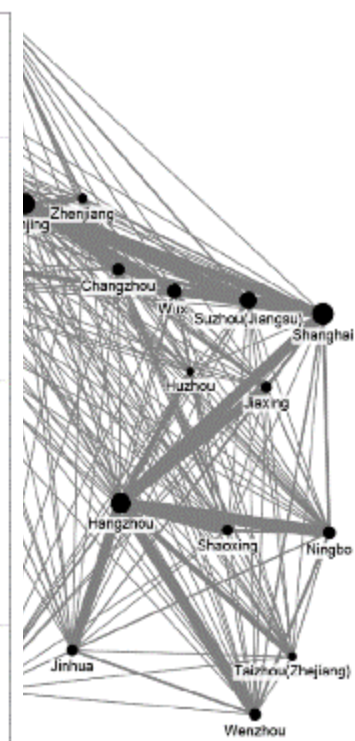
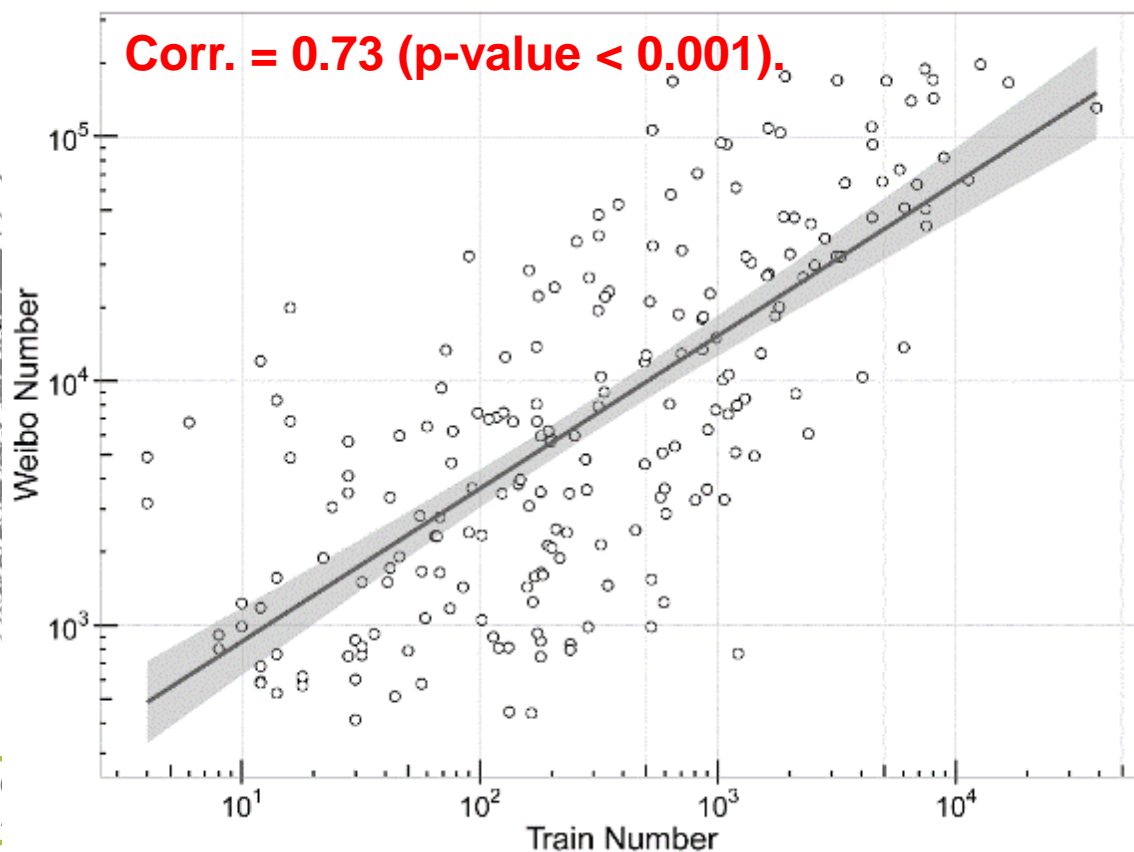
基于交通的城际流动网络

微博空间分布

基于微博的城际流动网络



Zhang, Wang, et al. 2016  
speed railway network, ...



Yangtze River Delta high-

# Model Chinese Mobility

---

- Cross-sectional OLS regression (in logs), city pairs

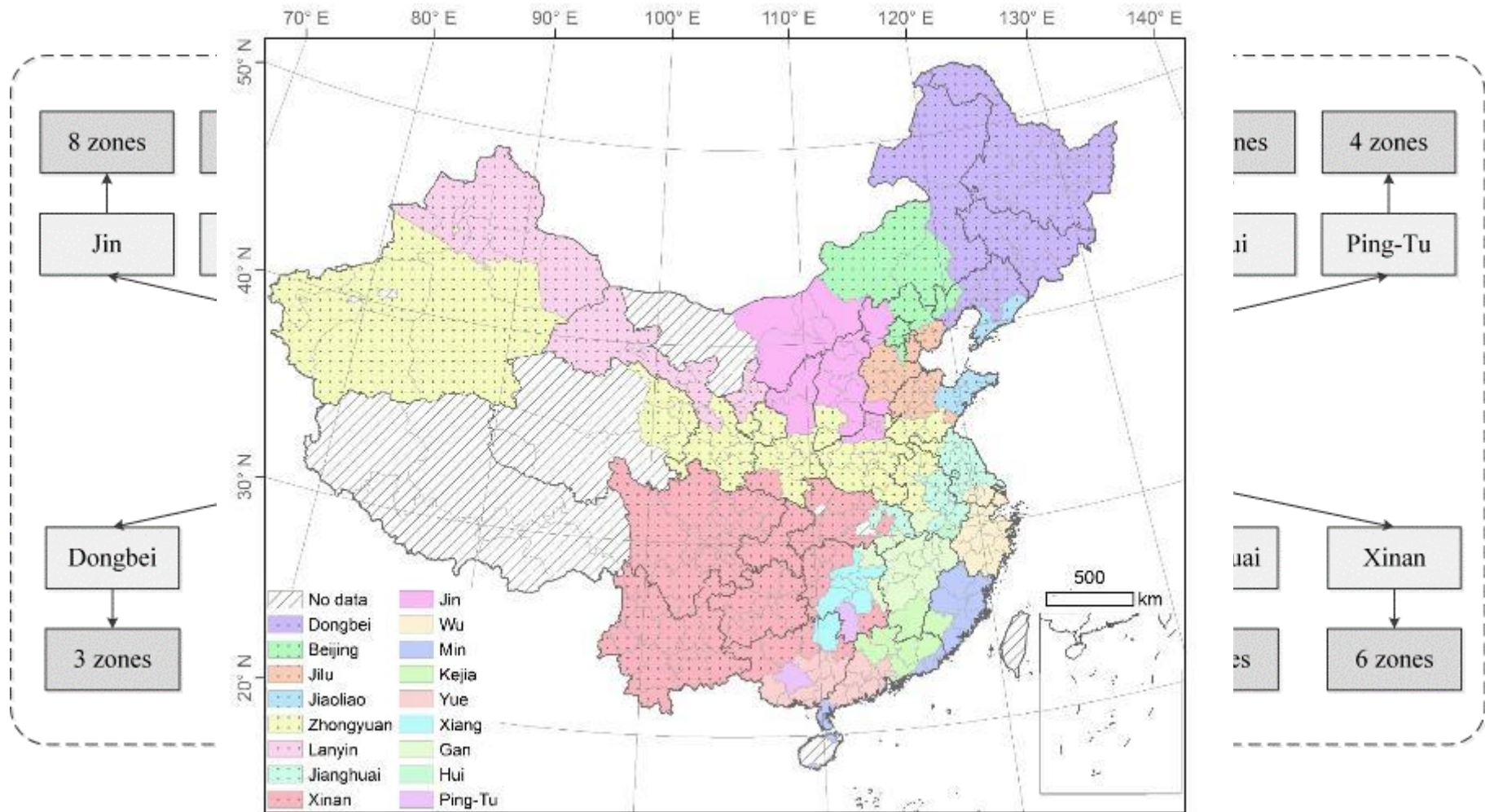
$$\log(T_{od} / L_o) = \beta_1 \cdot \log[\textit{Commuting}_{od}] + \beta_2 \cdot \log[\textit{Dialect}_{od}] + F_o + F_d + \textit{controls} + \varepsilon_{od}$$

$T_{od}/L_o$  : mobility flows between city pairs/total Weibo user flows of the origin city

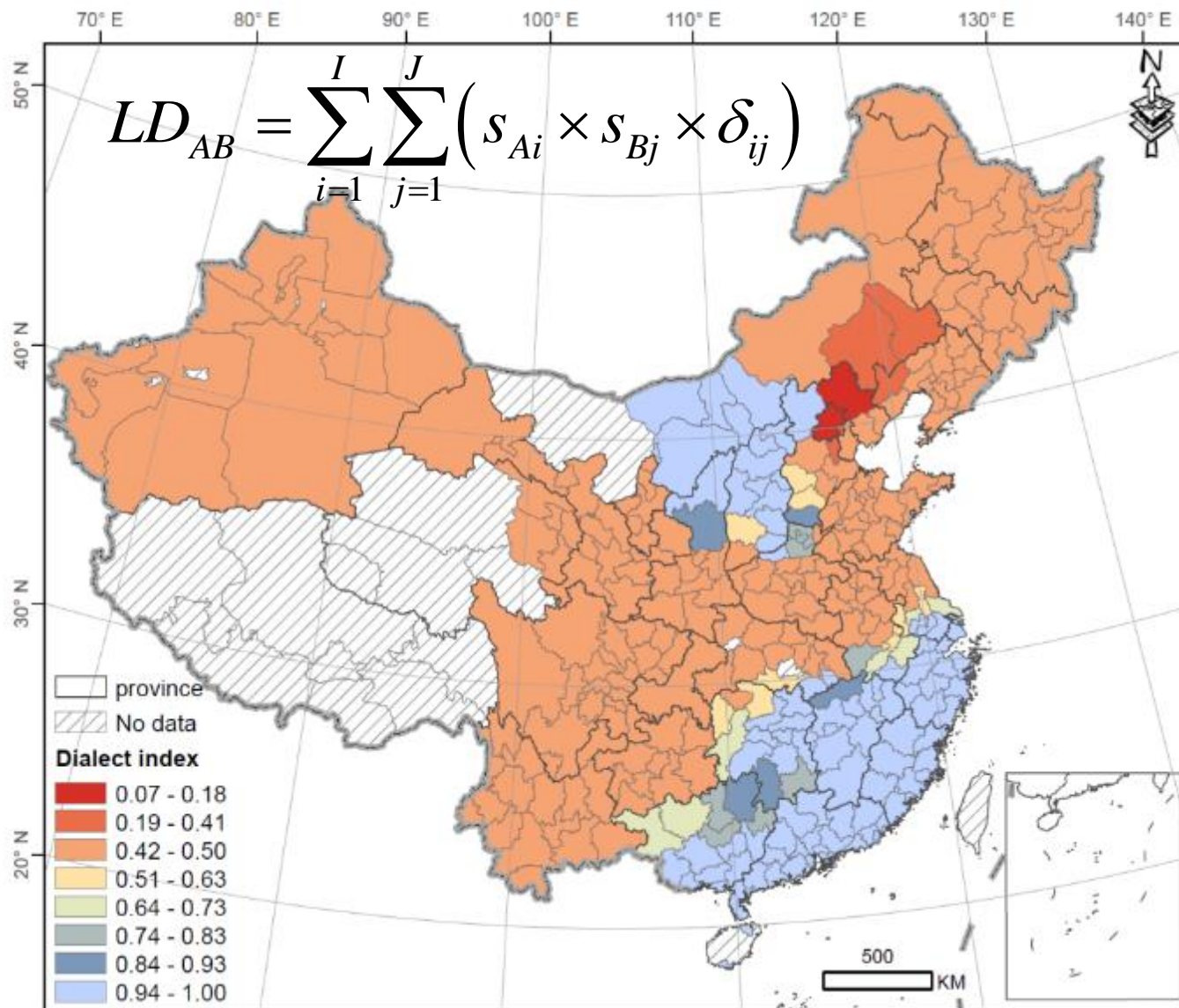
- [Commuting]: commuting distance&time-(pecuniary mobility costs)
- [Dialect]: dialect distance- (non-pecuniary mobility costs)
- $F_o$ : origin city **fixed effect**
- $F_d$ : destination city **fixed effect**
- $IV$ : historical dialect distance

# 主要因子1：方言文化

方言作为地方文化的综合载体，对区域人口流动格局起决定性作用

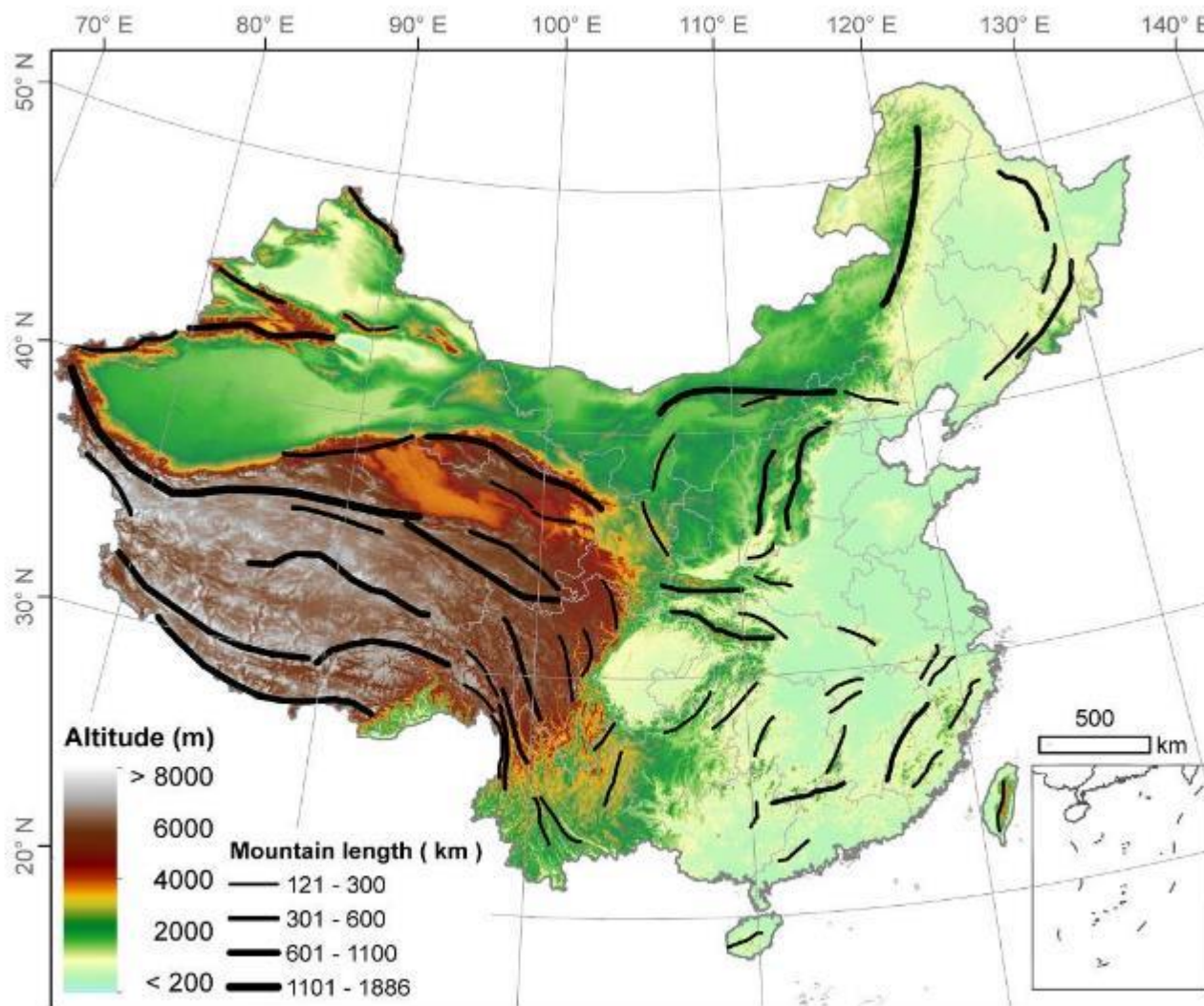


# 中国方言文化相似性度量



# 主要因子3：自然因子的影响

山体作为历史上区域流动主要空间障碍，是否在今天依然起影响作用？



# Mountains and Chinese Human-Cultural Geography

## 模型设定

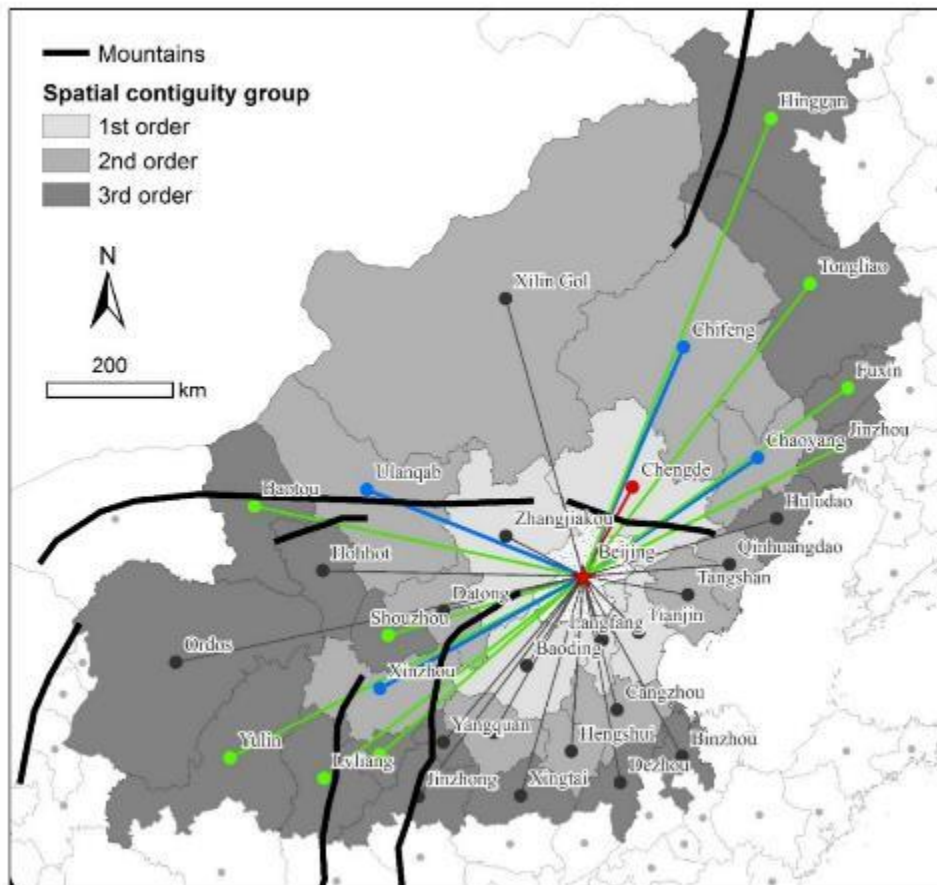
$$\begin{aligned} \log(Mobility_{od}) = & \alpha + \delta_1 \cdot [Mountain_{od}] + \delta_2 \cdot [Near_g] + \\ & \delta_3 \cdot [Mountain_{od}] * [Near] + \\ & F_o + F_d + controls + \varepsilon_{od} \end{aligned}$$

- **Control Group, Gray lines:**

city pairs with no mountain barriers between them

- **Treatment Group, Color lines:**

**red lines** : 1st order spatial contiguity group  
**blue lines** : 2nd order spatial contiguity group  
**green lines**: 3rd order spatial contiguity group



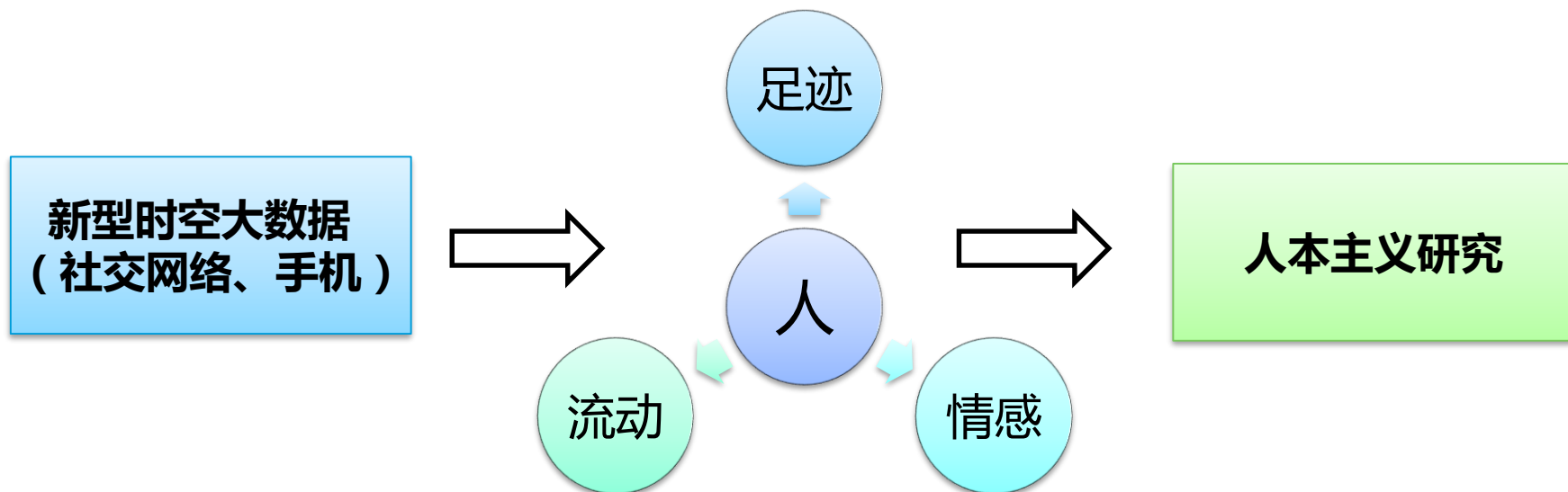


# 其他控制变量

	Definition	mean
通勤距离	Shortest journey time between city pairs by railroads (in logs)	2.2623
地理距离	The straightline geographical distance between two cities is calculated as Euclidean distance in kilometers between each pair of cities' centroids (in logs)	13.9316
驾驶距离	Shortest car driving distance between city pairs (in logs)	14.2210
是否同省	Binary indicator: 1= a city pair equals to the same province; 0=otherwise	0.0389
是否邻接	Binary indicator: 1= a city pair shares the administrative border; 0=otherwise	0.0159
工资收入差异	Difference in wages between city pairs (in logs)	0.2393
工业产值差异	Difference in employment share of non-agricultural sectors between city pairs (in logs)	18.1750
灯光指数差异	Difference in night light intensity level between city pairs (in logs)	1.3621

# 主要结论

- 在区域层面，方言控制着人口流动格局，但在跨区域层面，人们更愿意去文化差异大的地区短期流动；
- 高速铁路等基础设施的放在，促进城际的联系，增加了人群流动的时空异质性；
- 山体等自然要素至今依然对人流起直接影响作用（相对于无山体阻断的城市对，减少26% - 37%的人口流动量）



# 5 URBAN VALIDITY (城市活力)



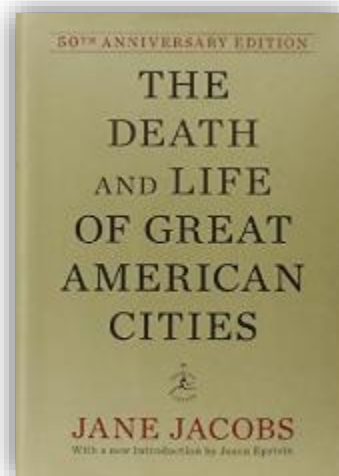
# 纪念简·雅各布斯诞辰100周年

美国城市的教母:

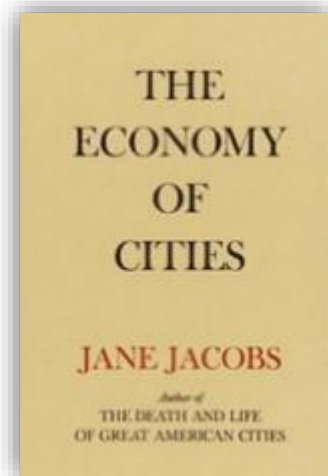
Jane Jacobs(1916-2016)

“Cities have the capability of providing something for everybody, only because, and only when, they are created by everybody.”

理想的城市应该是由所有居住者来参与其中的，城市的发展是自发的、基于人们活动的、而非设计得来。



《美国大城市的死与生》(1961)



《城市经济学》(1969)



2016年5月4日是美国记者-《美国大城市的死与生》作者简·雅各布斯 ( Jane Jacobs ) 的诞辰100周年。

她以对城市的研究而闻名，被认为过去半个世纪对美国乃至世界城市规划发展影响最大的人士之一。

# 城市充满活力的四个条件

- 雅各布斯认为一个城市只有当其具有多样性的物理环境时，才能变得繁荣和有活力。
  
- 多样性的四个条件
  - ① 应能具备多种主要功能：辖区必须具有至少两个以上的功能，这样才能吸引人们无论是白天还是夜晚，能有不同的目的、在不同的时间来到室外
  - ② 大多数街区应短小而便于向四处通行：城市的街廊必须够短，并且有足够多的路口，这样能给行人创造许多交流的机会
  - ③ 住房应是不同年代和状况的建筑的混合：城市内应该有不同年代和不同类型的多样性建筑，以满足低租金和高租金租户的要求
  - ④ 人口应比较稠密：一个城区必须有足够密集的人口与建筑

# 城市活力 Urban Vibrancy

- **【词典】活力**：1 ) 旺盛的生命力；2 ) 事物得以生存、发展的能力
- **城市规划——Live-work-play (LWP) places**
  - 具有活力的社区往往是紧凑、高密度、具有很好通达性、混合利用和易于步行的，可以是以就业为导向（活力就业中心）或以居住为导向（活力居住社区），它们更能吸引高技能年轻人、企业家和创业者，能够产生更高的生产效率和更多的创新机会，并具有环境友好性（Song, 2014）
- **城市经济学——The consumption value of urban density**
  - 高密度的城市，具有更多的社会互动与多元化的消费机会，人们愿意为此支付更高的租金和交通成本（Coutour, 2014）

## “消费城市” (Consumer Cities)

——城市经济学家哈佛大学Edward Glaeser教授

From Production Cities to Consumer Cities



# 城市活力研究内容

## 度量方法

- 微观尺度：街道、邻里、社区
- 中观尺度：行政区和新城（鬼城识别）
- 宏观尺度：城市、区域

## 形成机制

- 城市规划与设计
- 公共投资与乘数效应

## 更多研究

- 活力与交通拥堵
- 活力与消费城市
- 活力与房地产价值
- 网络冲击下的传统商业等

# 研究1：基于点评网的消费活力

- 对点评网上北京所有的零售店和餐饮点
- 两个度量维度：密度+多样性

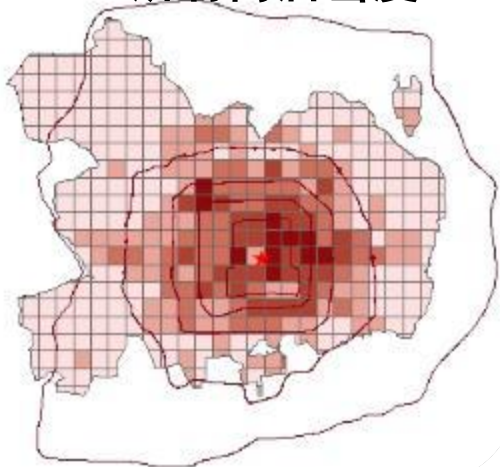
多样性的度量  
(借鉴生物多样性指标)

$$\text{diversity} = \sum_{i=1}^S p_i * \log p_i$$

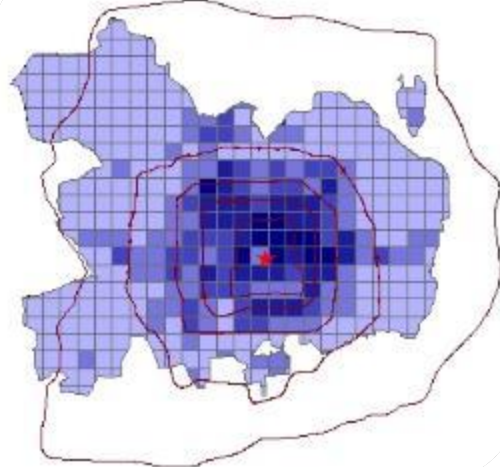


类别	亚类	说明
美食	北京菜	包含官府菜、烤鸭等小类
	川菜	包含烤鱼、香锅等小类
	江浙菜	包括上海菜、淮扬菜等小类
	.....	
购物		包括购物中心、超市便利店等
休闲娱乐		包括
运动健身		包括
.....		

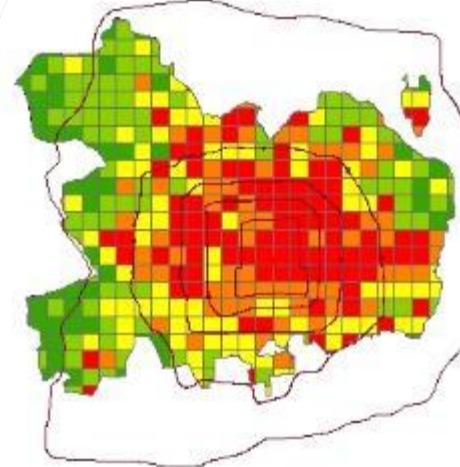
所有商店密度



餐馆密度



餐馆多样性



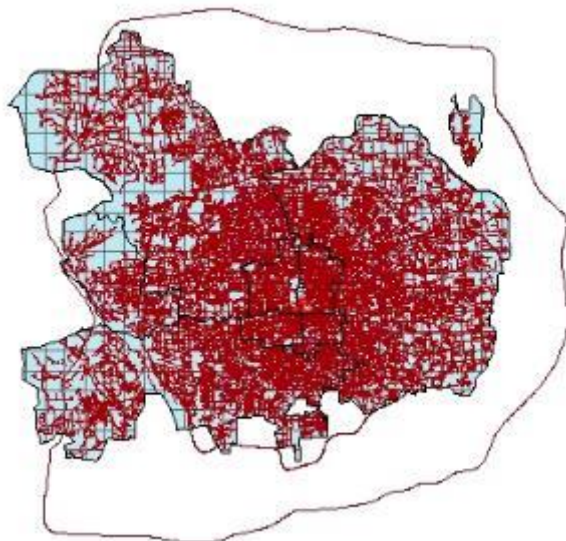


# 机制：城市规划与设计如何影响消费活力？

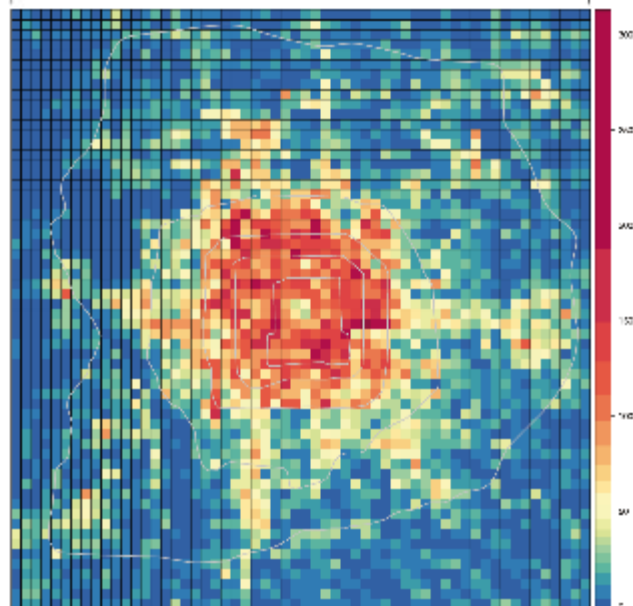
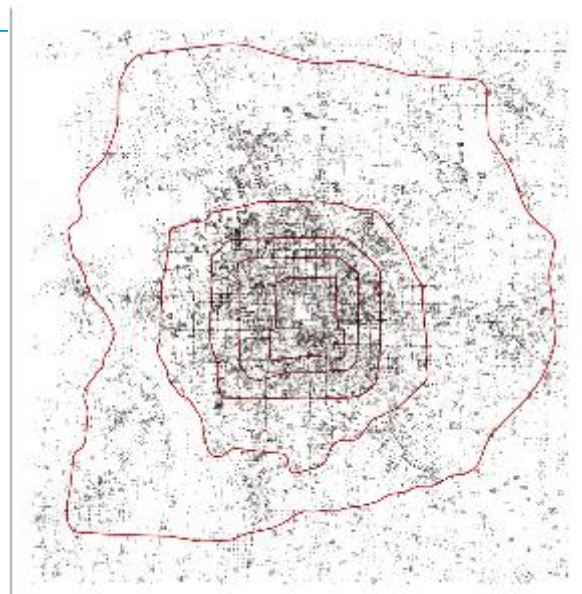
- **小路、十字路口**等利于行走的道路规划有利于消费活力的形成（较强内生性）
- **绿地**虽然占用土地，但有利于消费多样性的提升
- **轨道交通**便捷性的提升有利于商业的聚集，但与消费多样性相关度较低（较强内生性）
- **政府单位**所在地消费活力较高



主干道路



小道路

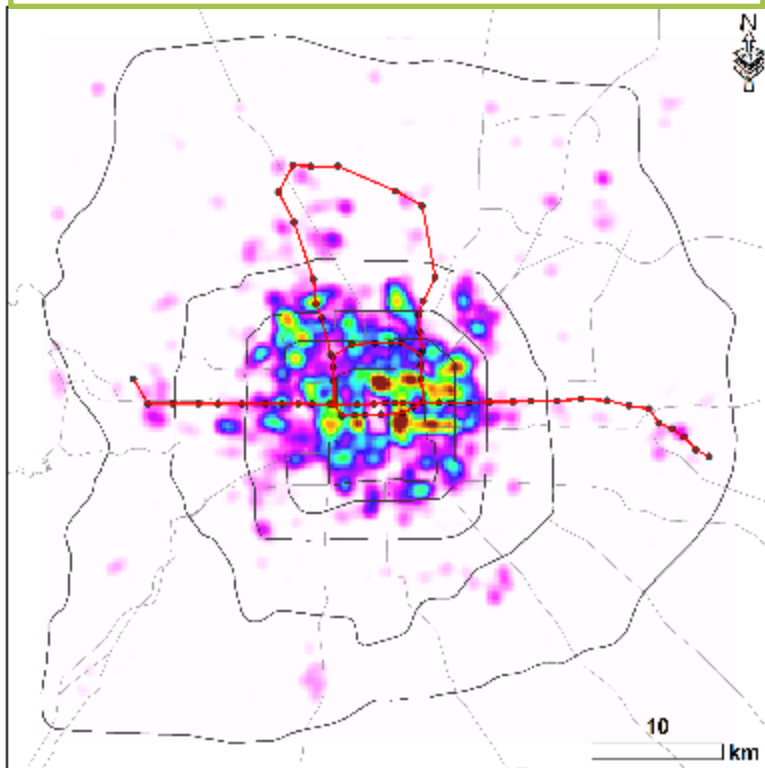


十字路口密度

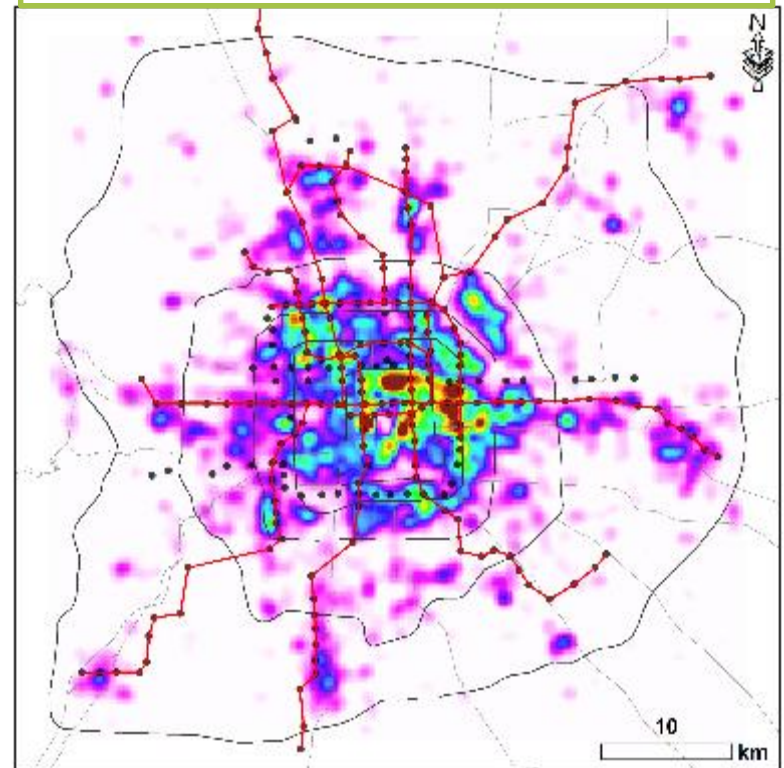
# 机制：基础设施建设如何影响消费活力？

北京市的奥林匹克公园以及地铁线路等**基础投资及建设**，带动了私人投资，共同促进了周边地区的升级，影响城市消费活力。

2006年北京地铁交通网络与消费热点



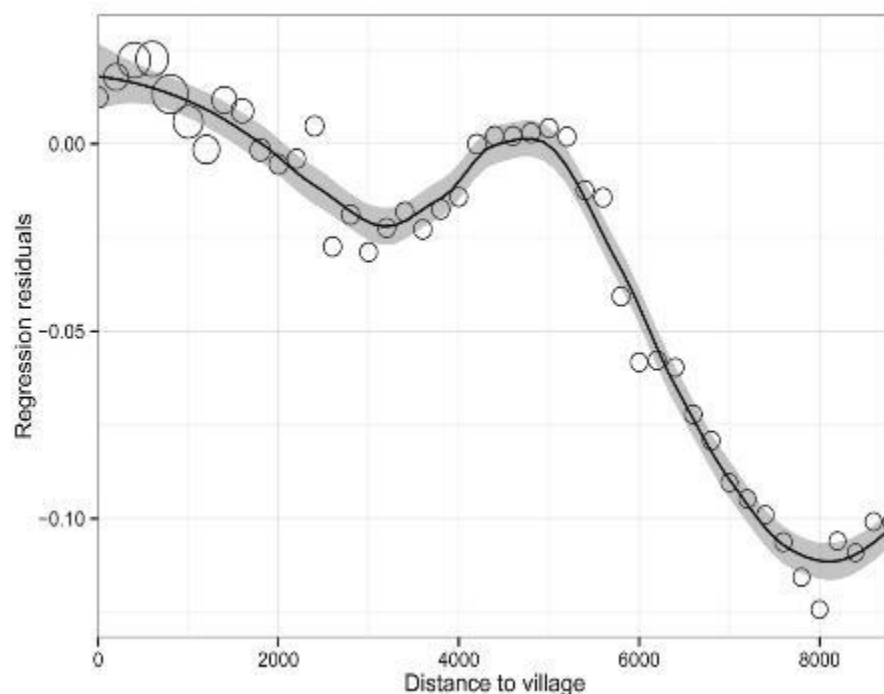
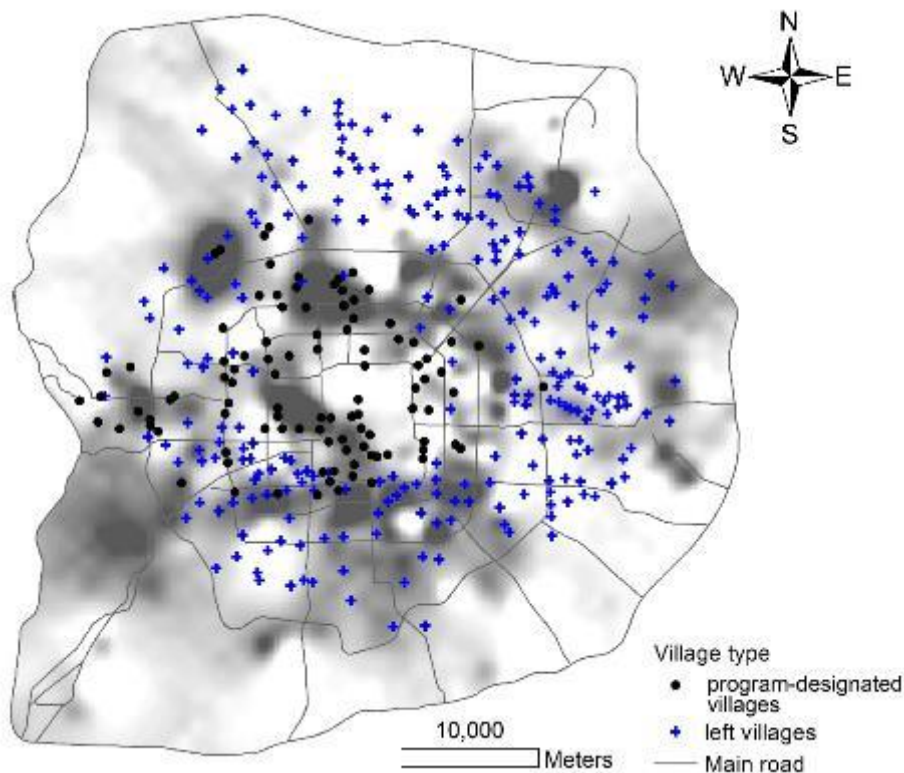
2014年北京地铁交通网络与消费热点



Zheng S Q, Hu X K, Wang J H. Wang R. 2016 *Subways near the subway: Rail transit and neighborhood catering businesses in Beijing*. *Transport Policy*.

# 机制：城市更新如何影响房地产市场？

北京市奥运会前对城中村进行了改造，改善了周边基础设施和环境质量，从而提高周边房地产价值



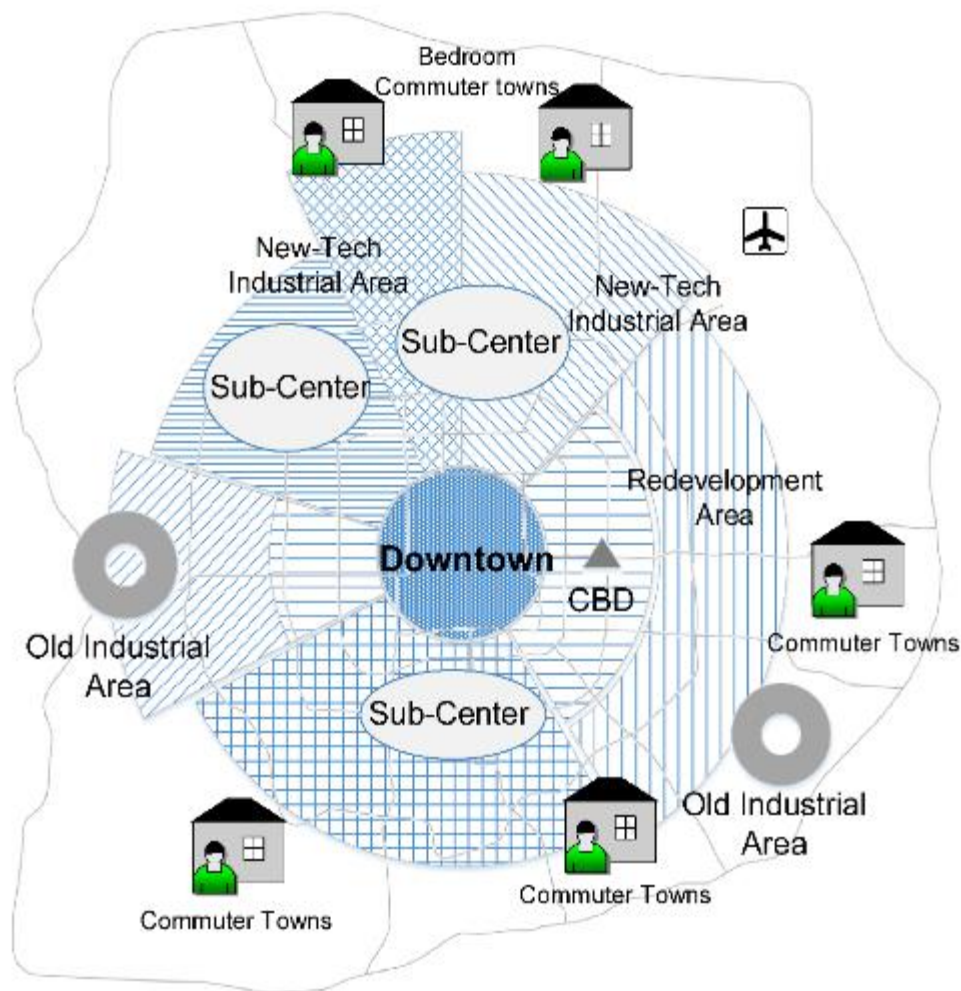
# 研究2：土地利用与城市活力

## □ 理论研究支持

1. **Jacobs (1961)** 混合土地利用与城市活力
2. **Glaeser et al. (2001)** 消费城市与城市活力
3. **Jacobs-Crisioni et al., (2014)** 人类活动多样性与城市活力

## □ 实证研究

1. 缺乏实证研究
2. 缺乏交叉研究 ( 1,2,3理论 )



北京城市结构概念图 (Wang J H)

# 土地利用与城市活力

---

## □ 研究目标

Do **land use configurations** for **housing and consumption amenities** influence the rise and decline of **city liveliness** ?

## □ 研究方法

构建了一种新的时空计量模型

( Spatially and Temporally Varying Coefficient Model )

## □ 主要发现

1. 城市对消费的正面影响作用
2. 封闭的社区对城市活力的负面影响作用与机制
3. 城市内部人类活动(生活-工作-娱乐)的时空不确定性

# 手机定位数据

手机定位数据表征人类活动与城市活力  
工作日、周末、24小时

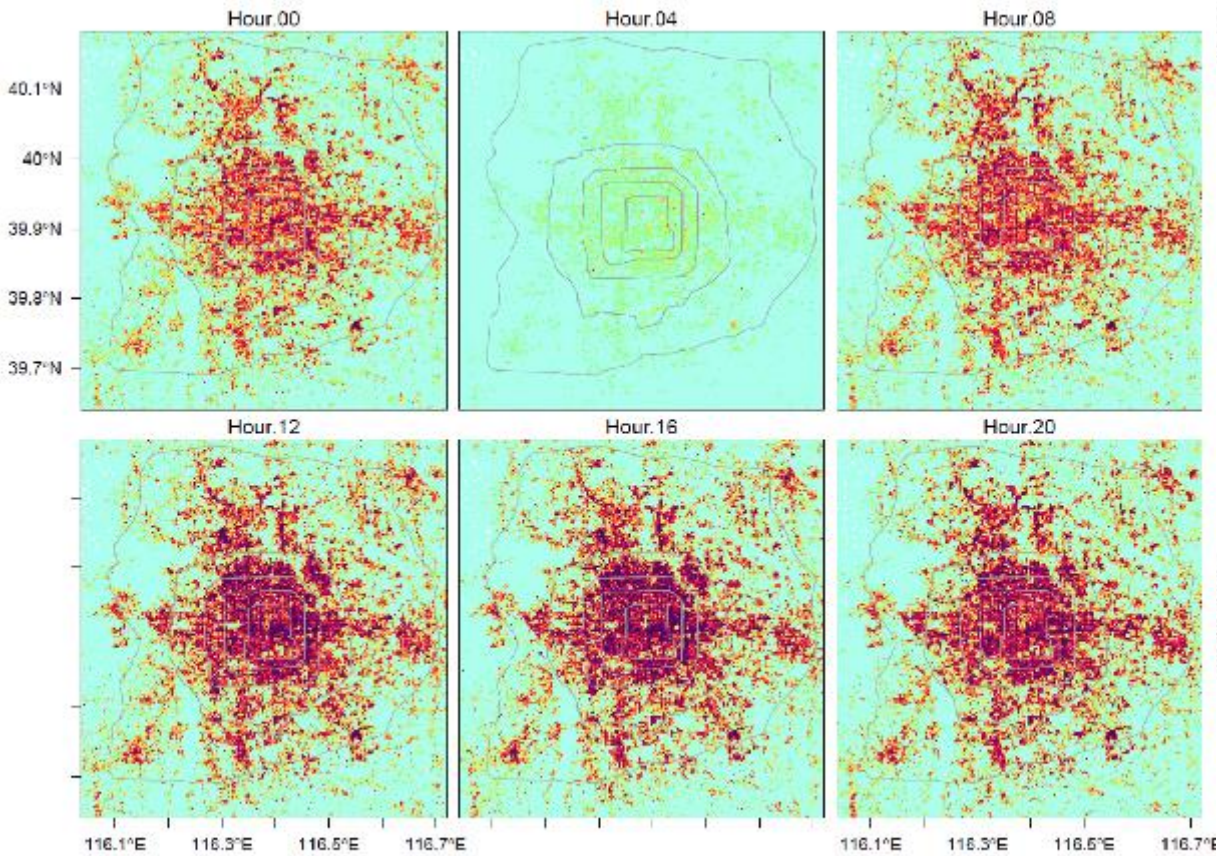
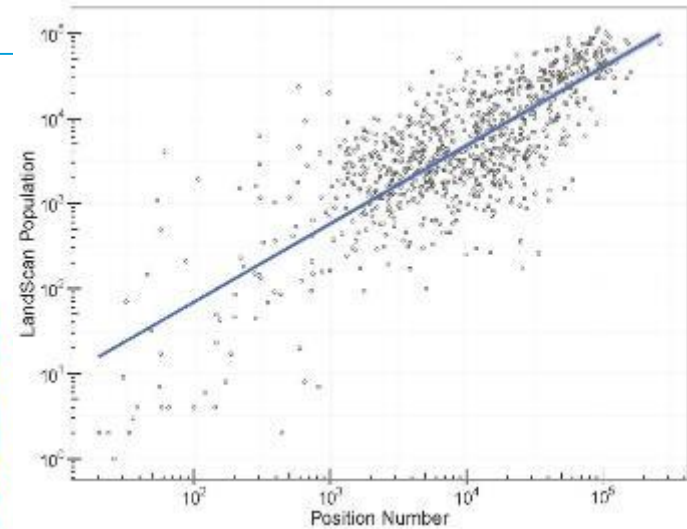
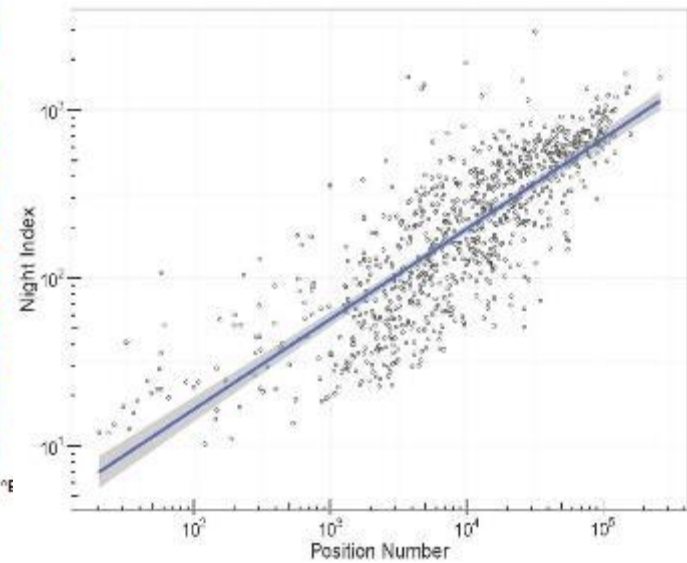


Figure 3.2: Spatio-temporal distribution of mobile phone positioning data.

## 数据代表性检验



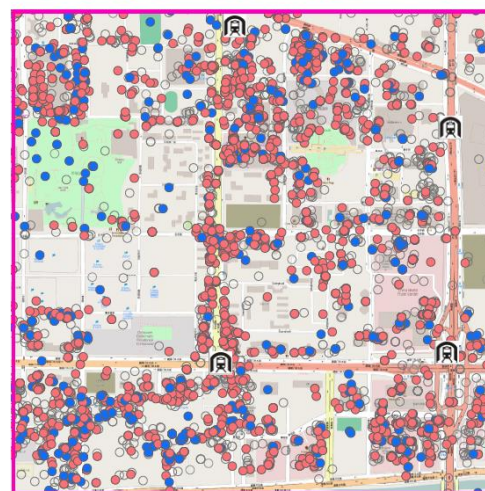
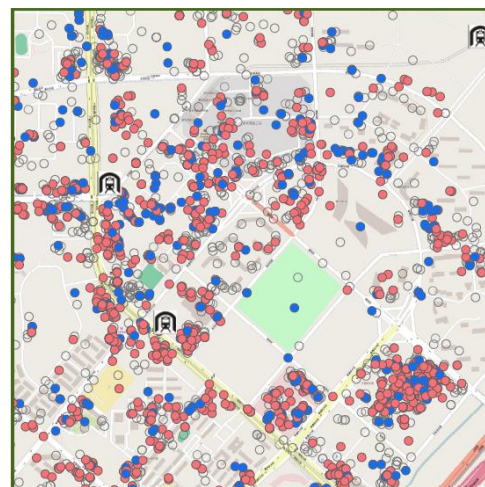
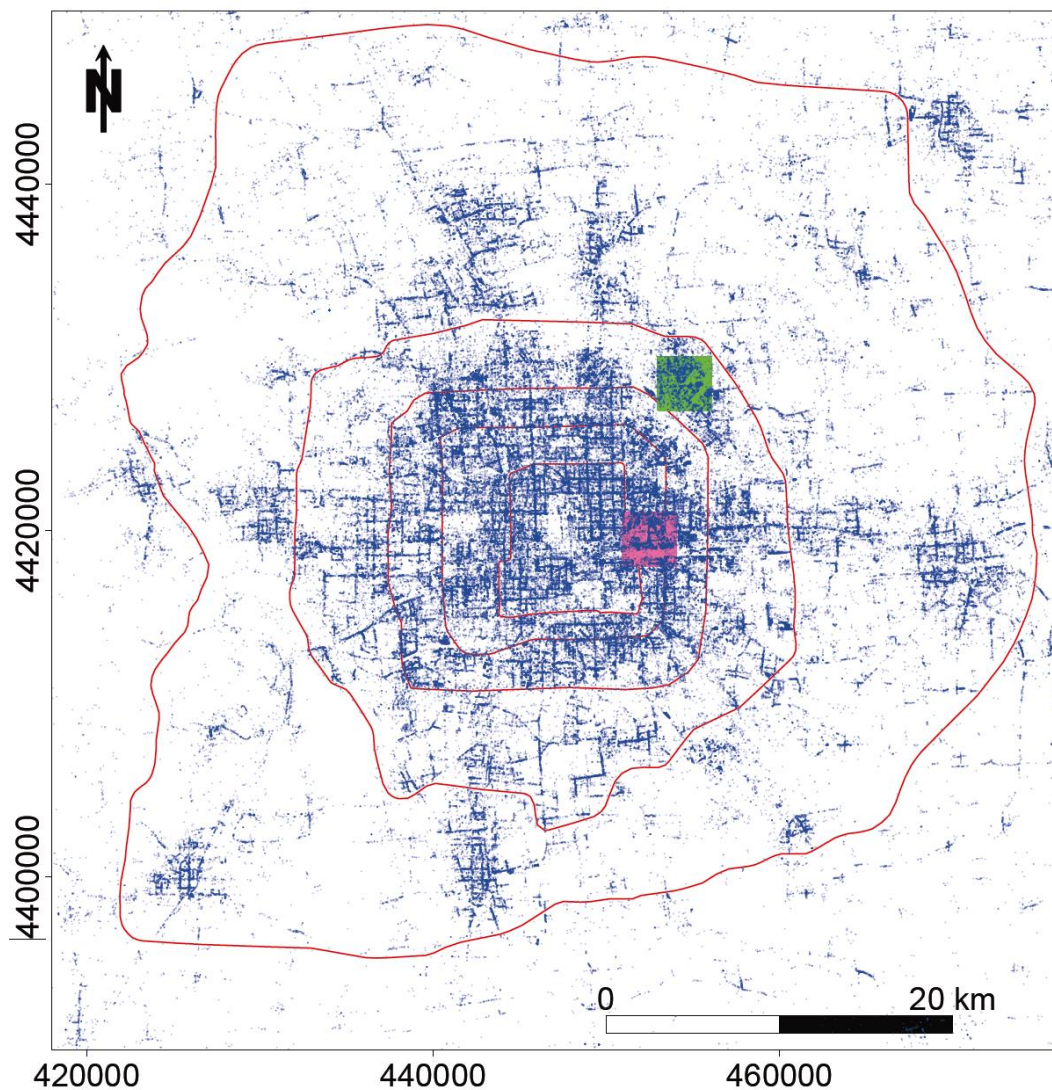
**LandScan population distribution  
(correlation coefficient: 0.76; p-value<0.001)**



**night light intensity  
(correlation coefficient: 0.71; p-value<0.001)**

# 土地功能类型

利用点评网的POIs类型等信息表征城市土地功能



● Housing ● Consumption ○ Others

# 关键变量描述

Table 3.1: Descriptive statistics of the key variables

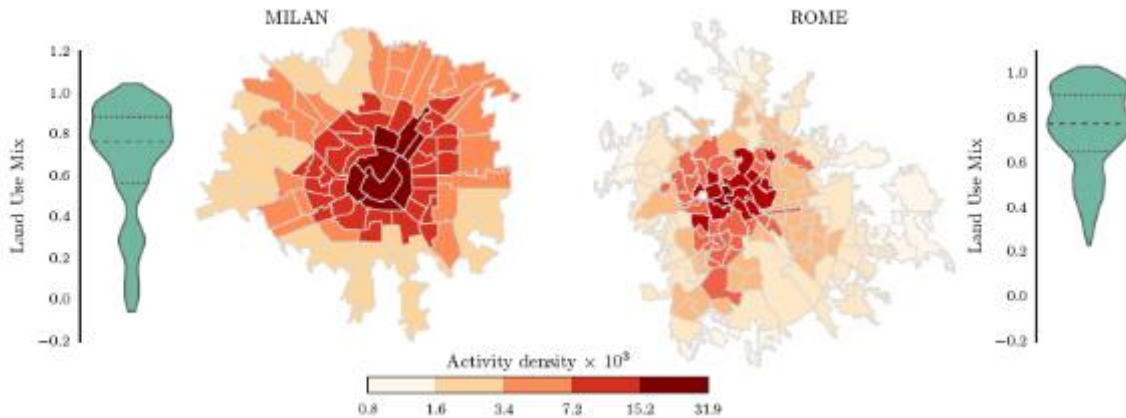
Variable Name	Definition	Mean	SD	Min	Max
Activity density at weekday(ywd)	mobile positioning data per area unit at weekday (in 1000 people)	5.415	10.075	0.004	225.924
Activity density at weekend(ywn)	mobile positioning data per area unit at weekend,(in 1000 people)	5.322	9.459	0.004	218.127
Housing amenities (H)	Number POIs related to residential complex compounds in each unit	1.372	3.209	0	32
Consumption amenities(C)	Number of POIs related to entertainment and leisure services	75.762	174.954	0	2177
Stations	Number of subway stations per unit	0.121	0.833	0	42
Road Density	Total road lengths per unit (in kilometers)	5.711	4.425	0	23.281
Distance to CBD	Straightline distance from each grid unit to the CBD (in kilometers)	22.441	9.000	0	44.926
Other amenities	Number POIs that do not fall in the category of Housing and Consumption	18.149	41.359	0	546

- **目标变量**：人类活动强度：工作日24小时变化；周末24小时变化
- **主要解释变量**：居住类设施、消费类设施
- **主要控制变量**：公共交通便利度、路网密度、到就业中心的距离、土地利用混合度等



# 城市活力研究展望

## □ 城市时空大数据+深度学习方法为定量城市活力提供广阔空间



- Jacobs-Crisioni et al., 2014
- Nadai et al., 2016
- Zheng et al., 2016
- Wu and Wang, 2016

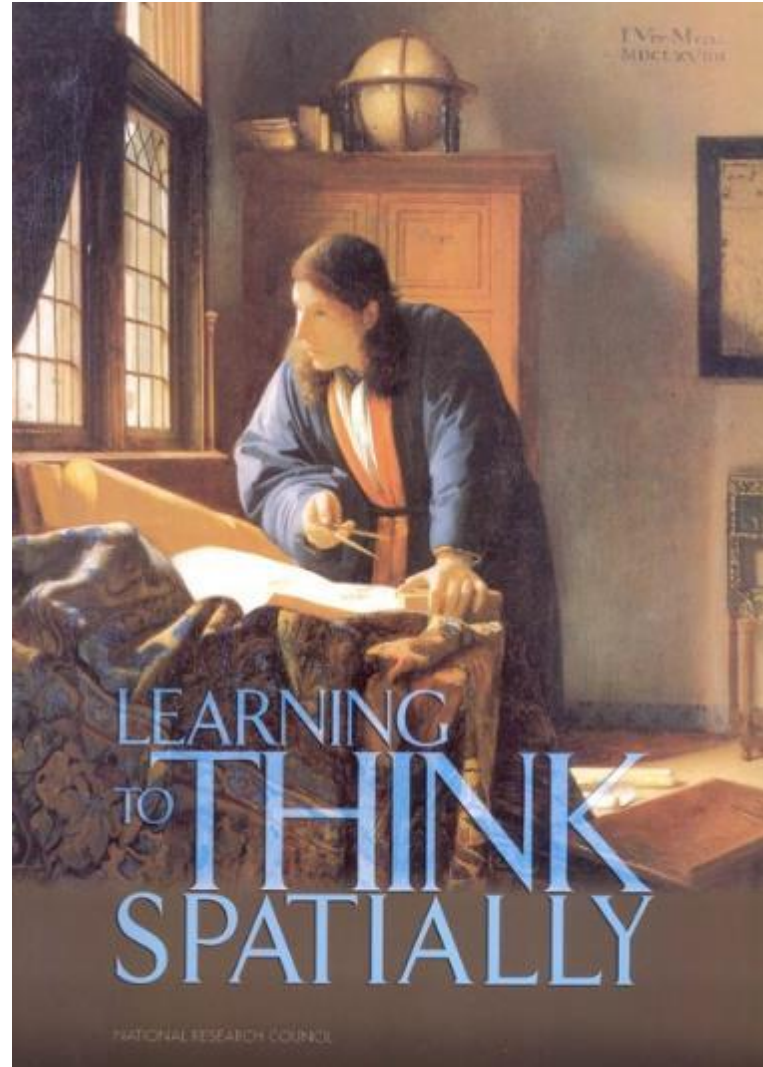
## □ 城市活力研究需要多学科交叉

- 数据采集-处理-分析：地理信息系统、计算科学（城市计算）
- 行为机制分析：城市规划、城市经济学
- 政策分析与设计：公共管理、城市治理

# Thinking in Spatial-temporal

---

- ◆ **Spatially**
- ◆ **Dynamically**
- ◆ **Interactively**
- ◆ **Autocorrelation**
- ◆ **Heterogeneity**
- ◆ **Scale**
- ◆ **Uncertainty**



# Thanks!

## Q & A



王江浩CAS

<http://jianghao.wang>

wangjh@lreis.ac.cn

# 配置云服务器环境

## □ [www.aliyun.com](http://www.aliyun.com)

- 云服务器ECS ( 按量付费 )
  - 地域: 华北 1
  - 可用区: 华北 1 可用区 B
  - 安全组ID: G1154587336553386
  - I/O 优化实例: 非 I/O 优化实例
  - 实例规格: 4 核 16GB
  - 网络类型: 经典网络
  - 带宽: 100Mbps ( 按使用流量 )
  - 操作系统: Ubuntu 14.04 64位
  - 系统盘: 40GB 普通云盘
  - 密码: 已设置
  - 实例名称: Tsinghua
- root@**115.29.39.169** (公) Abcd1234
- root@10.161.77.18 (内)

# linode

---

□ ssh root@45.33.94.61

# 辅助软件

□ **Chrome**

□ **ssh软件: PuTTY**

- <http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html>

□ **ftp软件: FileZilla**

- <https://filezilla-project.org/>



# 环境配置

---

## □ 更新系统

- `sudo apt-get update`

- `sudo apt-get upgrade`

## □ 安装**cURL**

- `sudo apt-get install libcurl4-  
openssl-dev`

- `apt-get install curl`

# 环境配置

---

## □ 安装R, For Linux Ubuntu 14.04

```
vim /etc/apt/sources.list
```

增加:

```
deb http://cran.rstudio.com/bin/linux/ubuntu  
trusty/
```

退出并保存

```
sudo apt-key adv --keyserver
```

```
keyserver.ubuntu.com --recv-keys E084DAB9
```

```
sudo apt-get update
```

```
sudo apt-get upgrade
```

```
sudo apt-get install r-base
```



# 环境配置

---

## □ 安装R包

```
install.packages("RCurl")
```

```
install.packages("rjson")
```

```
install.packages("plyr")
```

# cURL

---

```
curl "http://www.dianping.com/search/map/ajax/json" -H "Cookie:
_hc.v=""\\"77d4b41d-4087-46f6-bba2-e0930098f156.1477474568\\"";
PHOENIX_ID=0a01043e-158434024e3-67374c9; s_ViewType=10;
JSESSIONID=D8137B45FB48321FC3622C411F3B7A75; aburl=1; cy=2;
cye=beijing" -H "Origin: http://www.dianping.com" -H "Accept-Encoding:
gzip, deflate" -H "X-Request: JSON" -H "User-Agent: Mozilla/5.0 (Windows
NT 10.0; WOW64) AppleWebKit/537.36 (KHTML, like Gecko)
Chrome/53.0.2785.116 Safari/537.36" -H "Content-Type: application/x-
www-form-urlencoded;charset=UTF-8;" -H "Accept-Language: zh-
CN,zh;q=0.8,en;q=0.6" -H "Accept: application/json, text/javascript" -H
"Referer: http://www.dianping.com/search/map/category/2/10/r2580" -H
"X-Requested-With: XMLHttpRequest" -H "Connection: keep-alive" --data
"cityId=2&cityEnName=beijing&promoId=0&shopType=10&categoryId=10
&sortMode=2&shopSortItem=1&keyword=&searchType=1&branchGroupId
=0&shippingTypeFilterValue=0&page=1&glong1=116.3151554388428&glat
1=40.02384628560411&glong2=116.3927463812256&glat2=39.95769273
112198" --compressed -o test.json
```

```
library(rjson)
library(plyr)
json <- readLines("dianping.json", encoding = "UTF-8")
fdjson <- fromJSON(json, unexpected.escape = "skip")
str(fdjson)
list <- NULL
for(i in 1:length(fdjson)){
  list[[i]] <- data.frame(t(unlist(fdjson$shopRecordBeanList[[i]])),
stringsAsFactors=FALSE, row.names=NULL)
}
df <- do.call(rbind.fill, list)
write.csv(df, "data.csv")
```