



城市时空大数据与交通

李清泉（乐阳）

空间信息智能感知与服务深圳市重点实验室
深圳大学



目 录

1. 城市时空大数据
2. 城市交通时空数据一体化管理
3. 城市时空大数据与交通应用
4. 总结与展望



城市时空大数据

时空大数据

80%数据均和空间位置相关。现实世界中的各种现象几乎都包含空间、时间和属性三个基本特征。绝大部分数据都是时空数据(Spatial-temporal data)



空间



时间

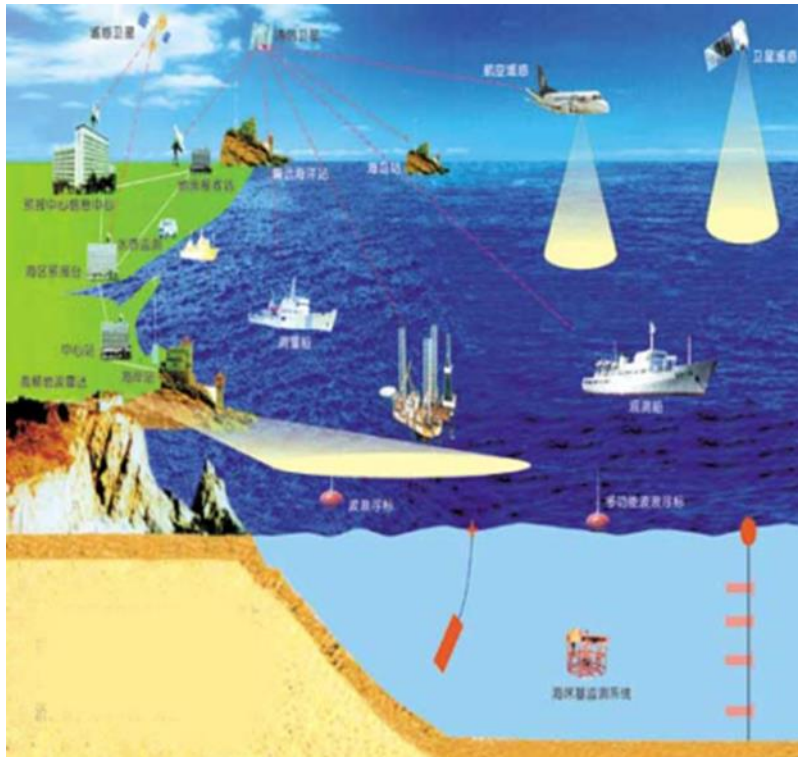
查询结果：共查询到 17 条记录

选中	快照图	定位目标	影像名称	景编号	拍摄时间	分辨率	容量	高度角
☒			WBT1~216817	WBT1~2168174	2009-05-06, 6	0.6000000038	0	53.68
☒			LED-ALAY2A15	ALAY2A15492	20081221	1	0	26.14
☒			LED-ALAY2A15	ALAY2A15667	20090302	1	0	24.12
☒			LED-ALAY2A15	ALAY2A15740	20090307	1	0	26.73
☒			LED-ALP3W14	ALP3W14952	20081114	1	1	29.54
☒			LED-ALP3W14	ALP3W14921	20081105	2.420129459	0	33.2
☒			LED-ALP3W14	ALP3W14921	20081105	2.4202790900	0	32.97
☒			LED-ALP3W09	ALP3W090278	20071022	2.4470443929	0	42.23
☒			LED-ALP3W16	ALP3W16119	20090202	2.422568507608	0	28.62
☒			LED-ALP3W16	ALP3W16119	20090202	2.4216067448	0	29.46
☒			LED-ALP3W14	ALP3W14967	20081115	2.435027143	0	32.7

属性

时空大数据的特点---动态/多粒度

时空数据既有长期积累的矢量/栅格数据，还有无人机、移动测量车、GPS实时定位等产生的大量动态数据，以及视频、微博等非结构化数据，动态化是时空数据的显著



空天地一体化监测网络



视频

#庭审直播#经交警调查证实：白某驾驶机动车行经交叉路口时，没确保安全行车，违反《中华人民共和国道路交通安全法》第二十一条第一款的规定，简某驾驶机动车通过有标志标线控制的交叉路口没让优先通行的一方先过。因此，白某与简某各承担此事故的同等责任。

7月23日 17:09 来自专业版微博

转发 | 收藏 | 评论

#庭审直播#案情情况：白某驾驶粤JEGXXX小汽车从恩平市恩东路往西方向行驶，行驶至恩平市体育馆门前路段时，遇简某驾驶摩托车乘载朱某从体育馆往小东方向行驶，两车发生碰撞，造成简、朱两人倒地受伤、两车损坏的交通事故。

7月23日 17:08 来自专业版微博

转发 | 收藏 | 评论

#庭审直播#各位网友大家好，恩平法院以“阳光司法，公正审判”为主题，举办法院开放日活动，邀请市人大代表、人大代表、政协委员和廉政监督员共25人旁听一案机动车交通事故责任纠纷案的审理，现就案件进行庭审直播。

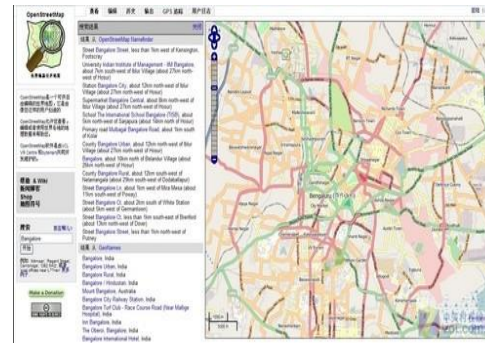
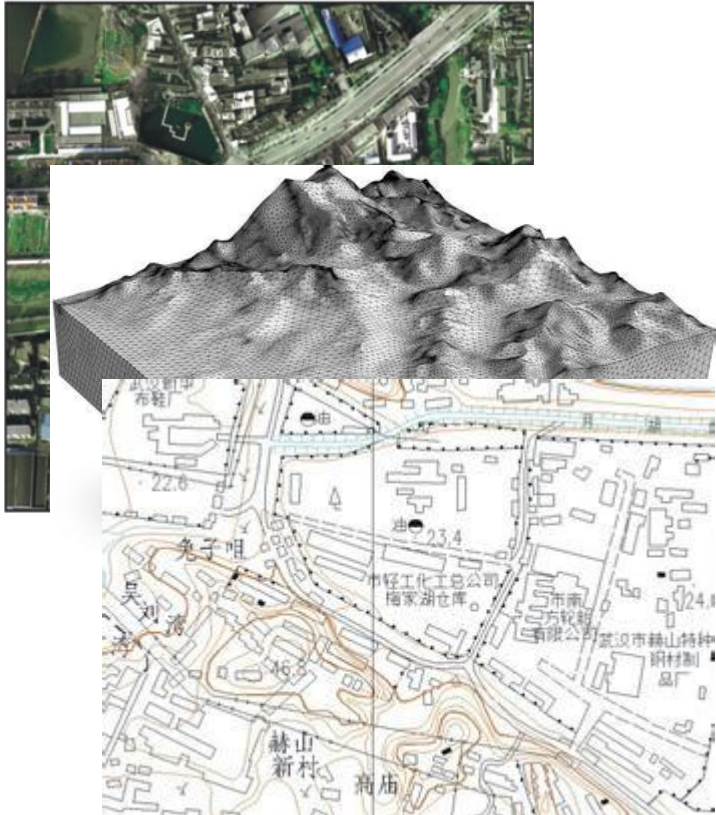
7月23日 17:08 来自专业版微博

转发 | 收藏 | 评论

微博

时空大数据的特点---泛在/异构

时空大数据的形式多样。既包括按照测绘规范生产的大量空间数据，如DOM，DEM，DRG和DLG；也包括更大体量的自发的数据，如文字，图片，语音等。既有实体空间中的数据，也有虚拟空间中的数据



OpenStreetMap

- ①北京将主办2014年APEC会议；
- ②地点在北京雁栖湖，位于北京市东北部、怀柔新城以北的雁栖镇；
- ③已完成官方网站建设、会标设计等筹备工作；^[2]
- ④开放购票，参会者可到活动家官方网站购买。^[3]

文本描述



腾讯街景



新兴交通时空数据

ICT技术特别是移动互联网技术的发展带来了更多的**新兴的、非交通专业的时空数据**。

● 手机数据



● WiFi定位数据

● 网络数据



● 社交媒体数据

● 定制公交数据



● 嘀嘀快的



空间



时间

面向个人，同时记录
时间及位置

交通时空数据的机遇

不断增长的交通时空数据为解决复杂交通问题如：出行调查，交通排放、交通-能源问题等，带来了新机遇。

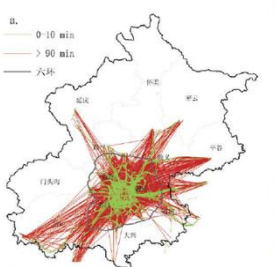
专业
交通
时空
数据



非
交通
专业
时空
数据



交通状态探测



出行调查



交通排放

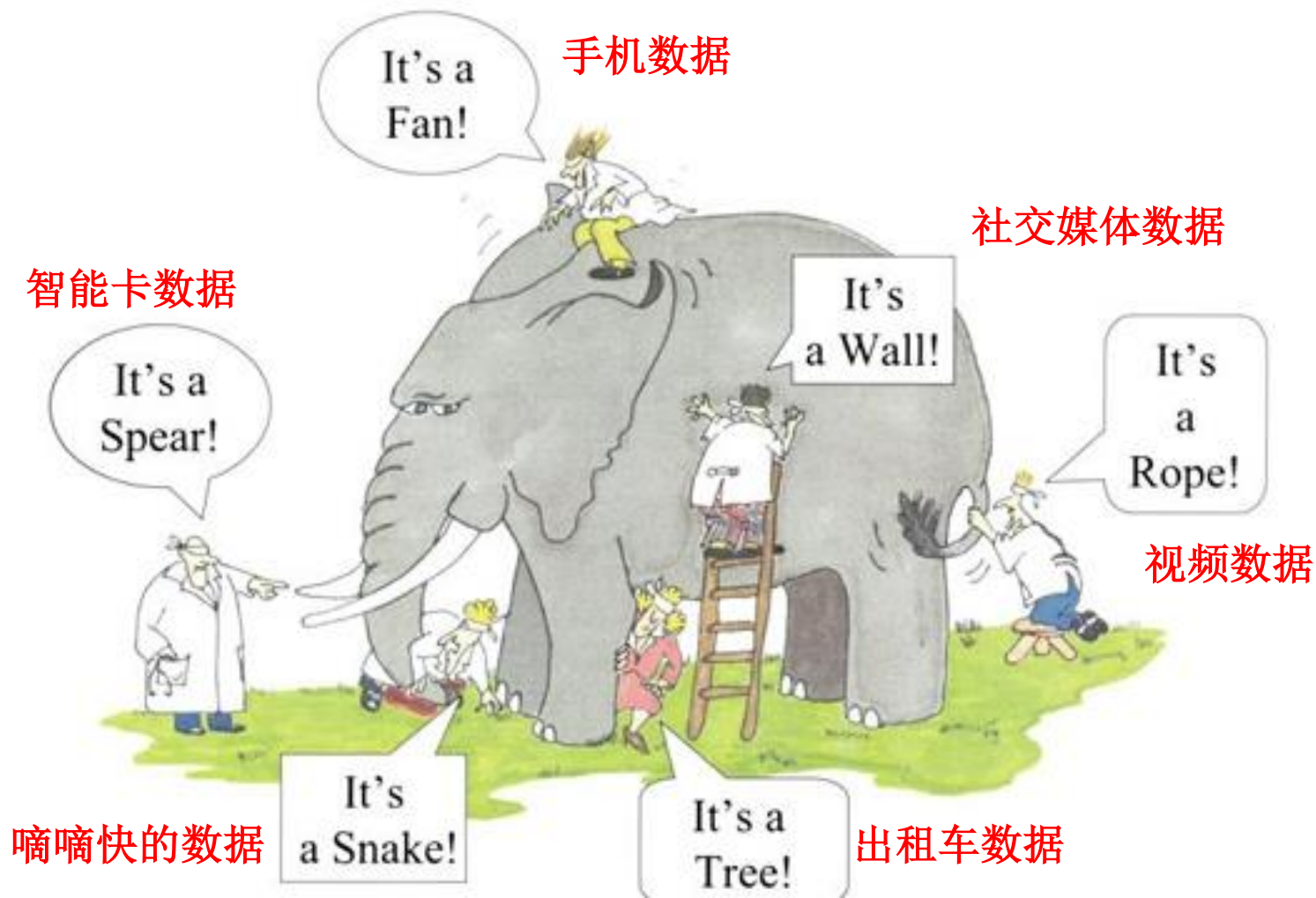


交通-能源

.....

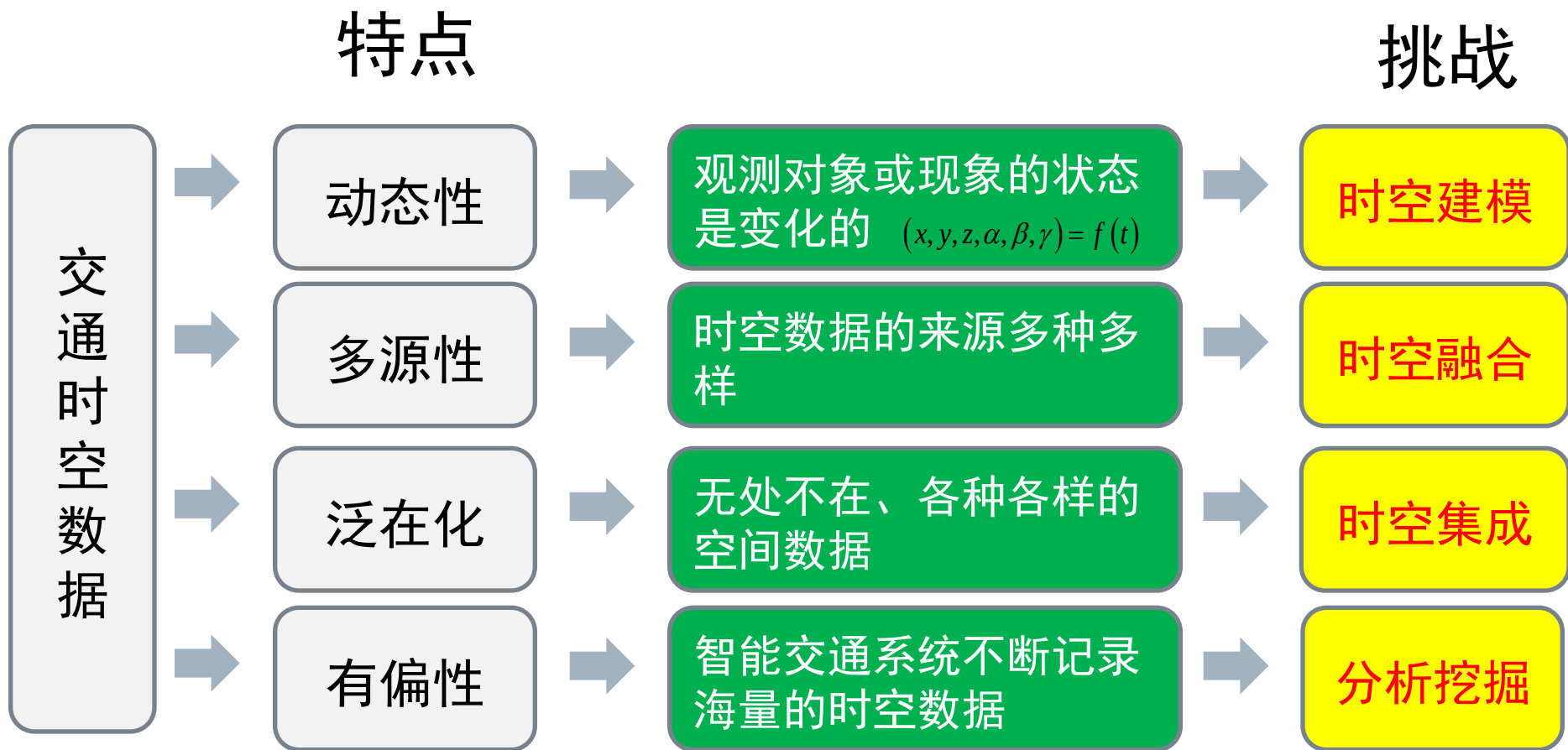
交通时空数据的挑战

■ 交通时空数据的**代表性**问题



交通时空数据的挑战(2)

海量多源交通时空数据也带来了更多挑战



科学研究范式的变革

图灵奖得主Jim Gray和Jnan Dash
在《科学的第四个范式》中，
将人类科学研究的历史分为
四个阶段

科学范式的变革

实验科学

描述自然现象

几千年前

理论科学

牛顿定律、麦克斯韦方程等

几百年前

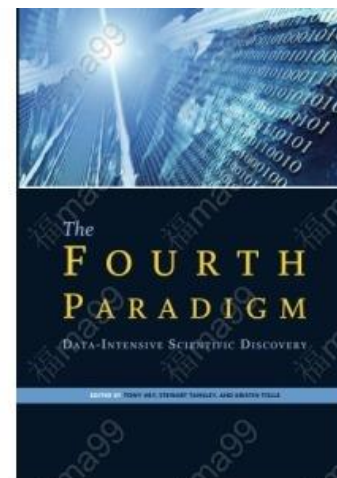
计算科学

模拟复杂现象

几十年前

数据密集型科学
理论和实验及模拟的统一

今天

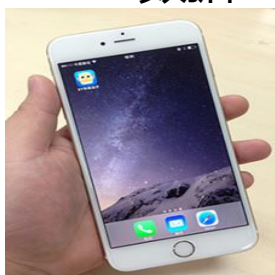


研究目标

时空数据获取技术



GPS数据



手机数据



移动社交网络数据

...

海量多源时空数据



从**时空数据分析**
到**位置智能服务**

GIS理论与方法



空间数据管理



二维



三维

研究思路

基础空间数据

空间数据查询

电子地图展示

动态时空信息

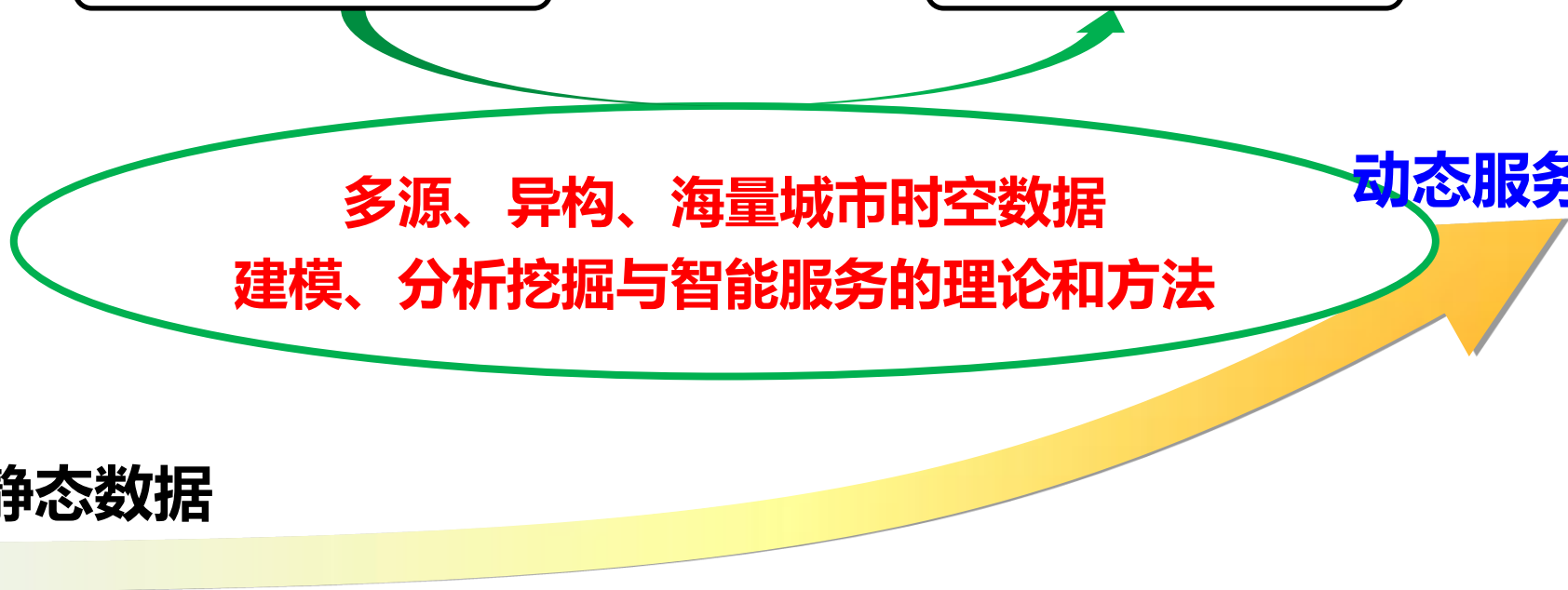
知识挖掘发现

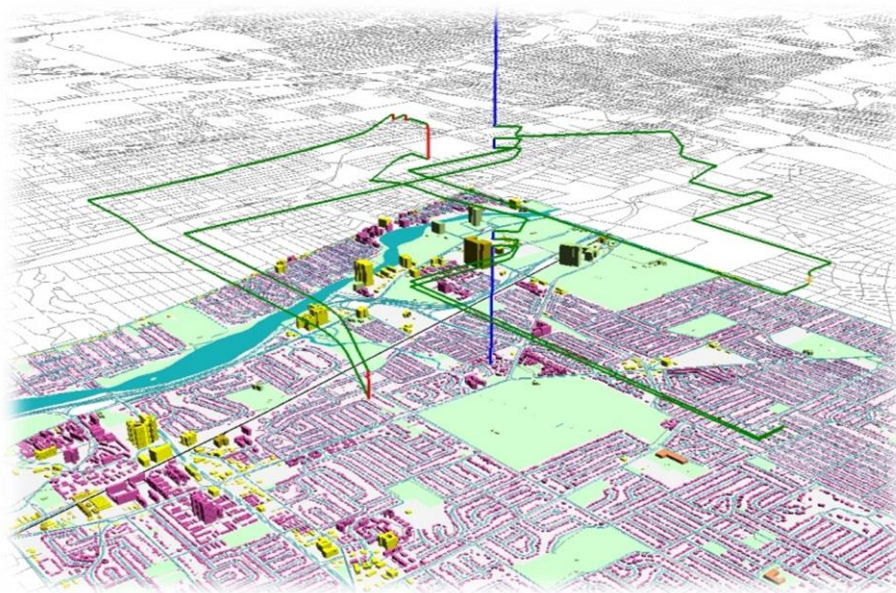
位置智能服务

多源、异构、海量城市时空数据
建模、分析挖掘与智能服务的理论和方法

动态服务

静态数据





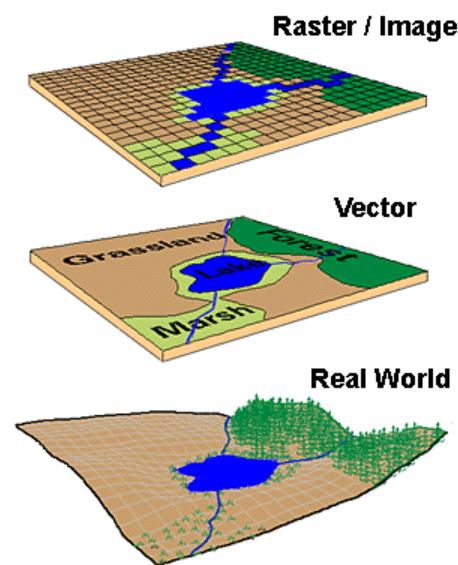
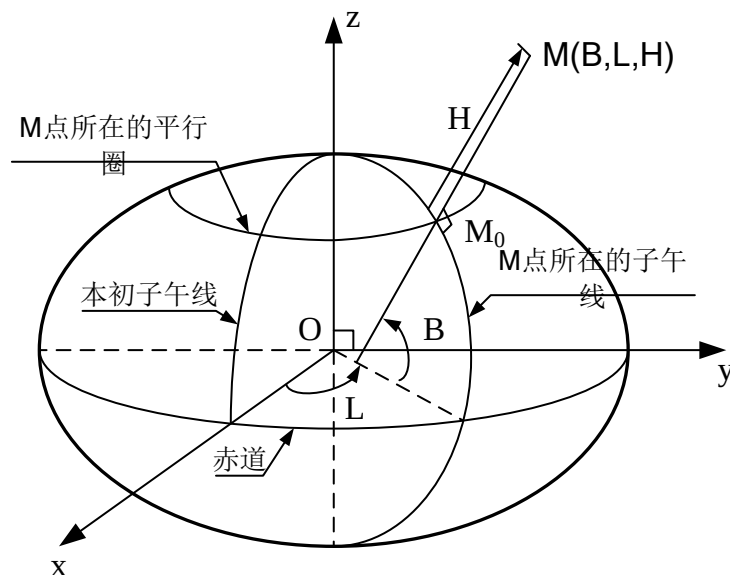
2

交通时空数据一体化管理

时空基准

大地坐标系作为对地观测系统空间基准，通常采用大地纬度(B)、大地经度(L)和大地高(H)表达。常用的坐标大地坐标系统有北京54，西安80，2000国家大地坐标系统，WGS84坐标系。

时间基准通常采用GPS时间进行时间同步。



多源空间数据

空间参考系

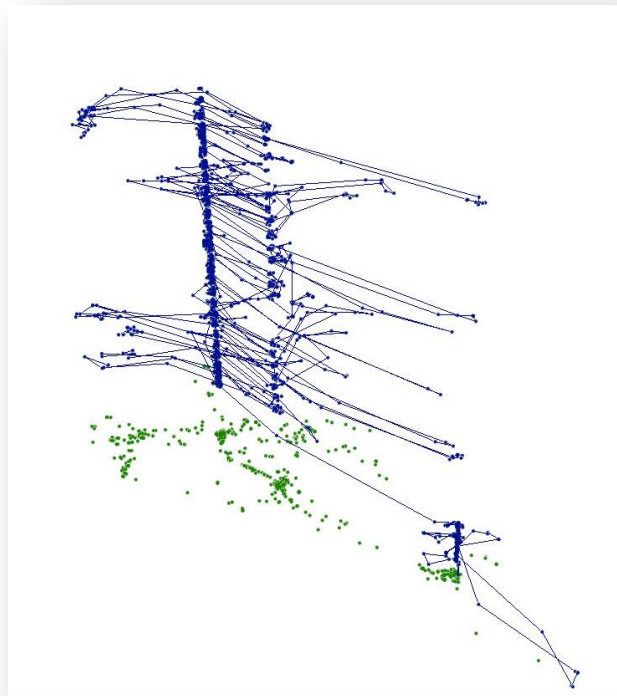
坐标系转换



时空轨迹

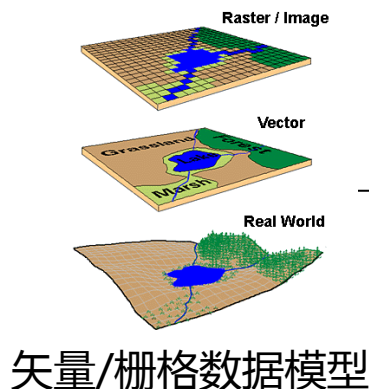
时空轨迹在时空坐标系中描述运动对象的时空变化，其中空间维度描述位置/形状的变化；时间维度描述时间的变化。在数学上，时空轨迹由一序列的时空元组组成，每个时空元组表示一个时空记录。

$$SpacetimeTrajectory = \{ \langle x_0, y_0, t_0 \rangle, \langle x_1, y_1, t_1 \rangle, \dots, \langle x_n, y_n, t_n \rangle \}$$



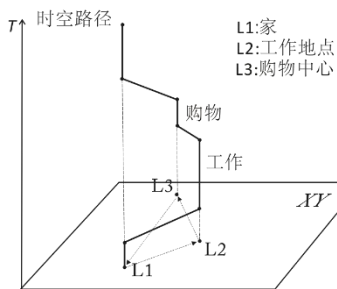
时空一体化数据模型

静态空间数据

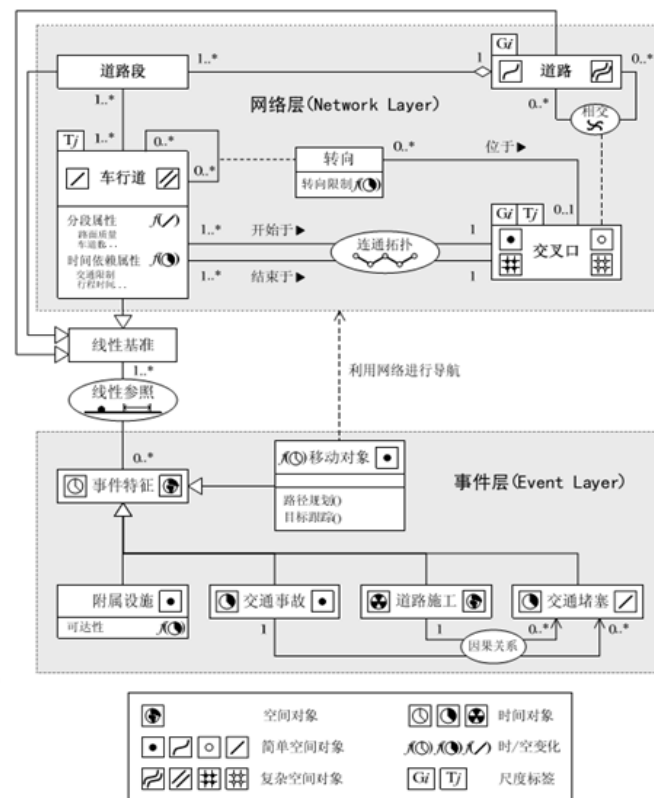


空间参考系

移动对象模型



线性参考系



基于特征的时空数据时空一体化模型(以GIS-T为例)

通过时空基准、线性参考和时空分段等技术，解决了运动对象一体化建模与表达，现实/虚拟空间时空数据集成等关键问题，奠定了多源数据时空分析与智能服务理论基础

数据管理及分析框架

分析挖掘

个体/群体
行为模式

城市空间结构
合理性分析

设施优化
选址

时空建模、可视化

模型

融合

平台

数据

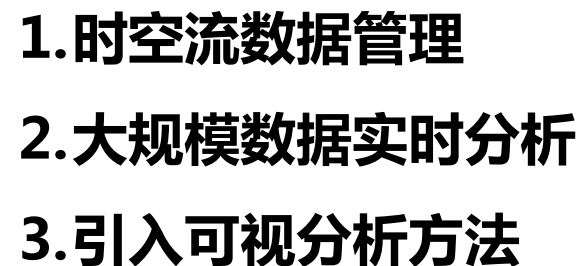
车辆轨迹
(出租车、
客运车辆)

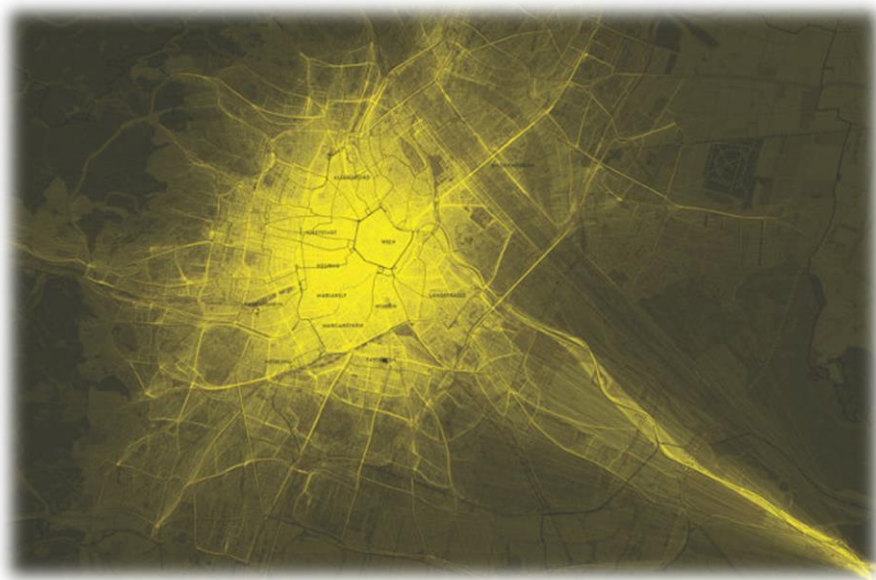
个体轨迹
(手机、
公交卡)

社交数据
(QQ,
Weibo
Check-ins)

底图
(POI、建
筑普查、
土地利用)

.....





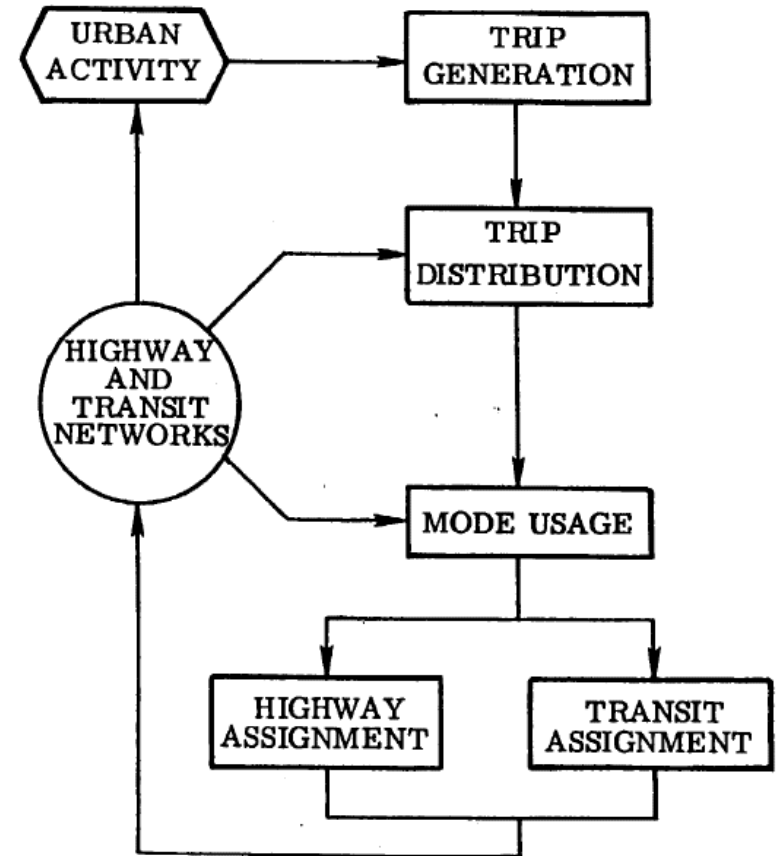
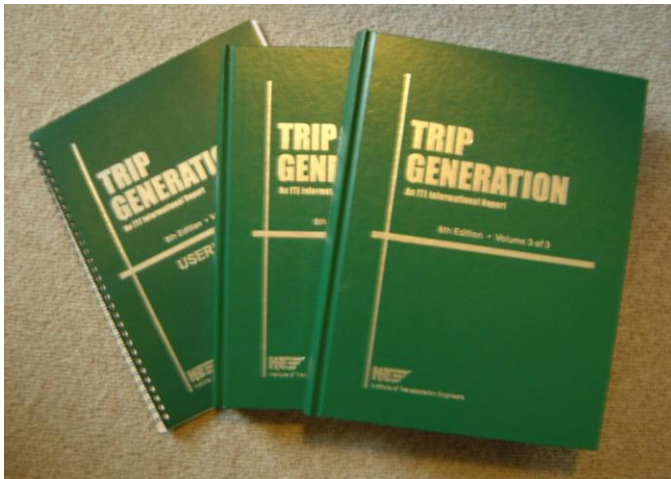
3

城市时空大数据与交通应用

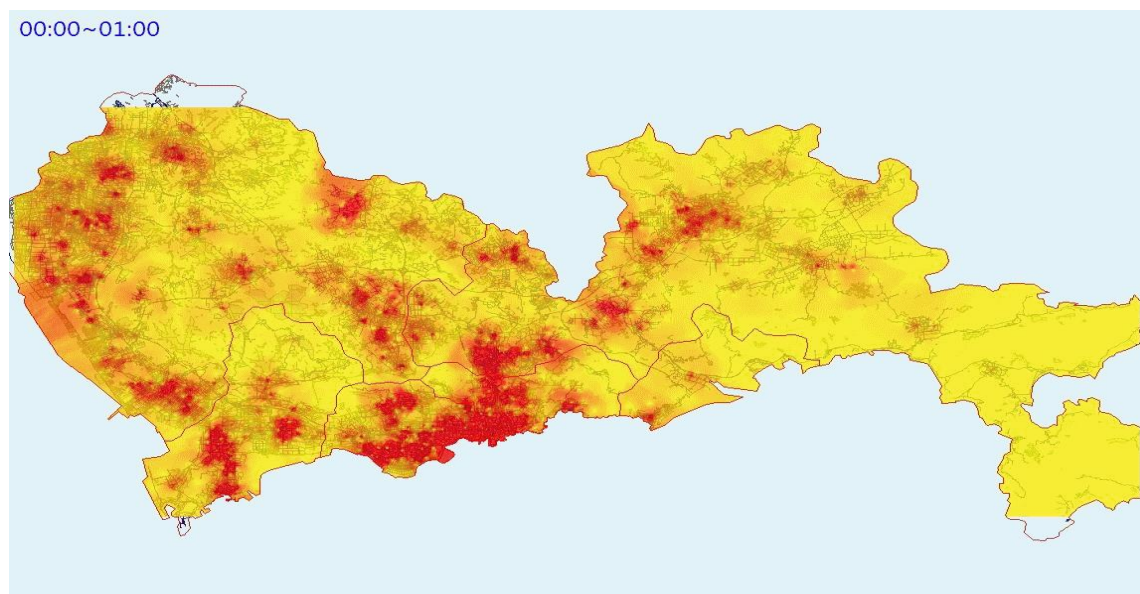
1. 基于多源时空数据的“四阶段”法

Forecast travel demand

- 交通生成
- 交通分布
- 出行方式划分
- 交通分配

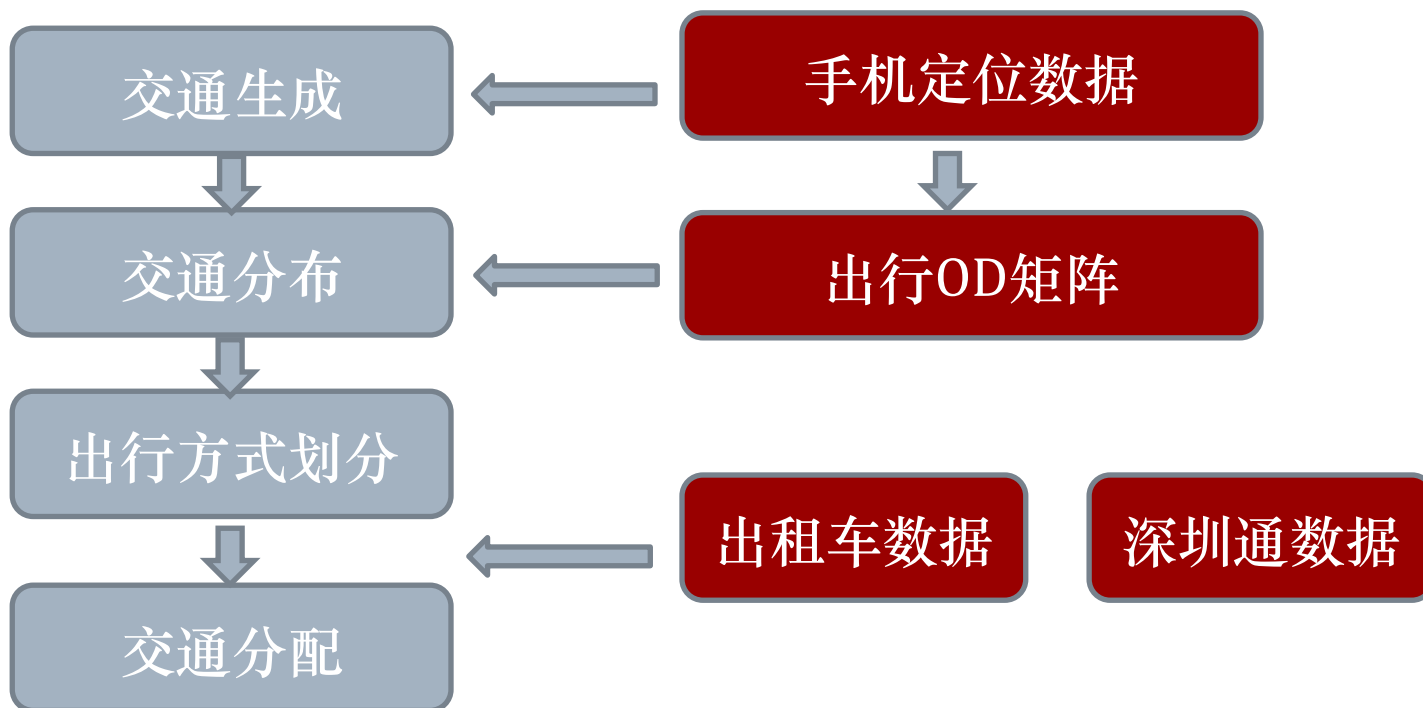


手机定位数据

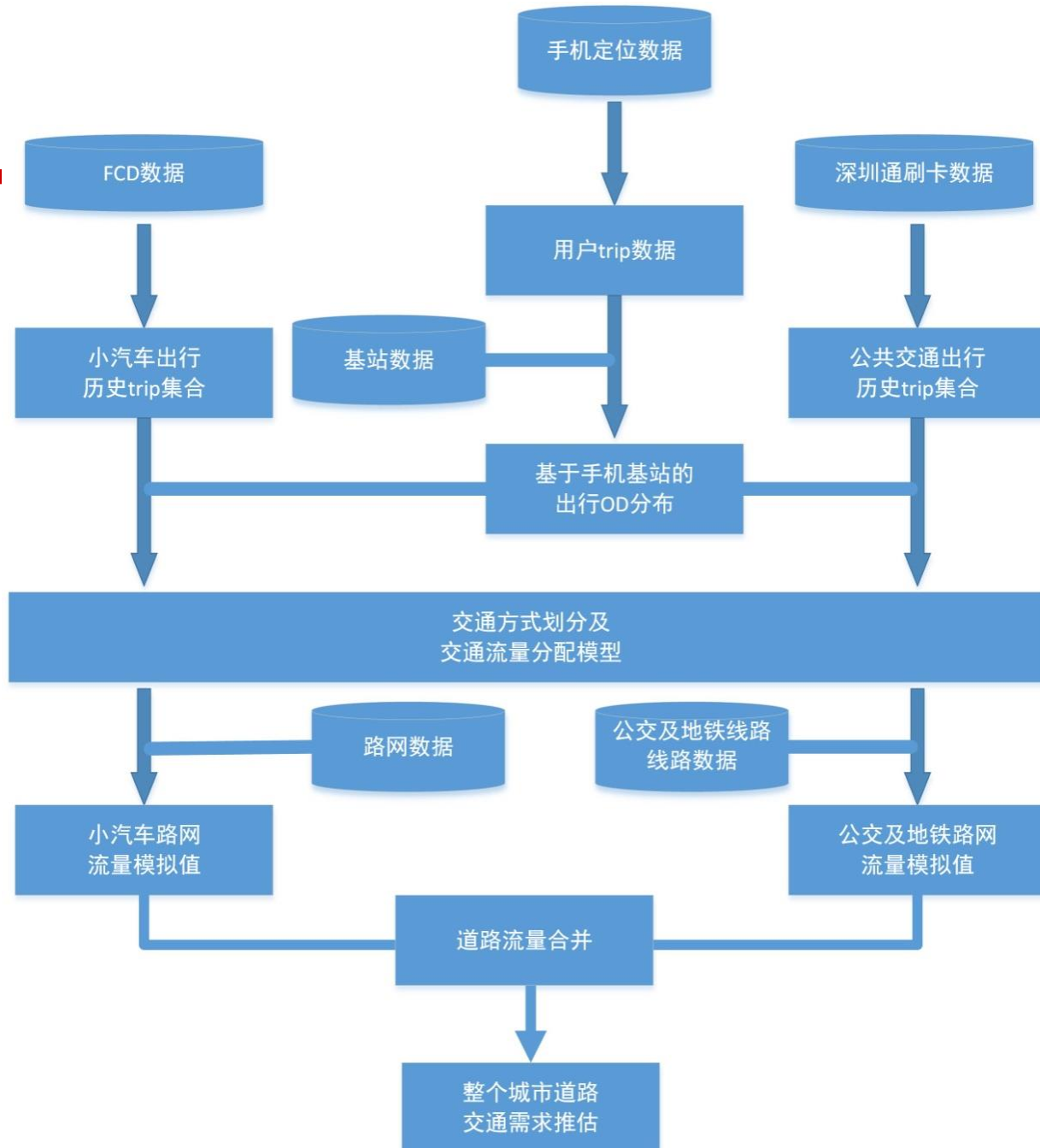


- 时间：A Friday, Mar. 2012
- 用户数：16,300,083
- 记录数：559,797,413
- 基站数：5,902

思路——数据驱动



框架

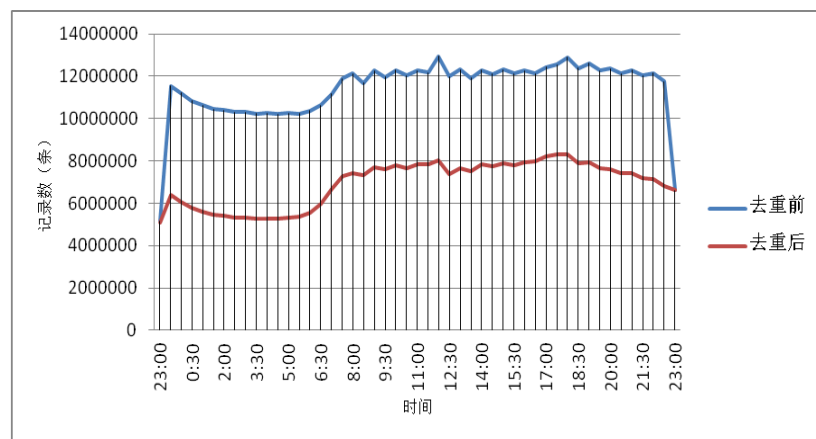


数据预处理

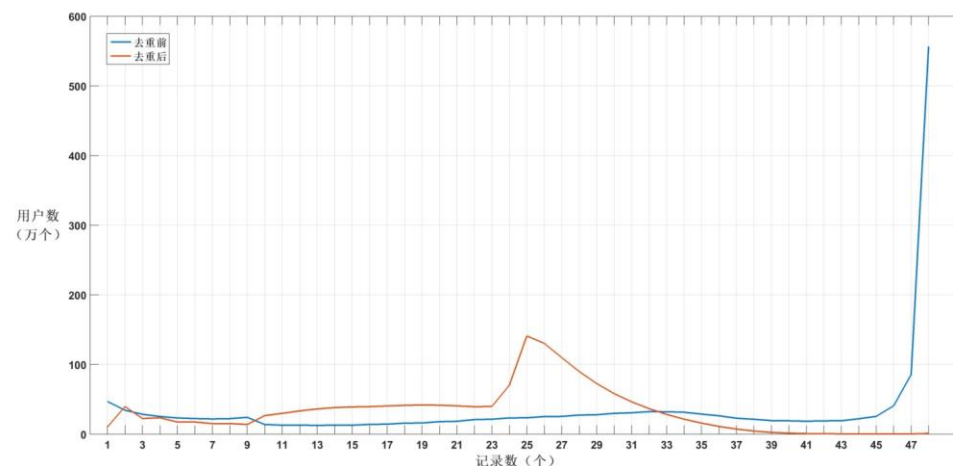
用户数：16,300,083

位置记录数：39,913,673

考虑该手机公司市场占有率**70%**，一人多机率约**30%**，
推估深圳市**2012**年人口约**1600**万。



去重前后不同时间段下的记录分布对比

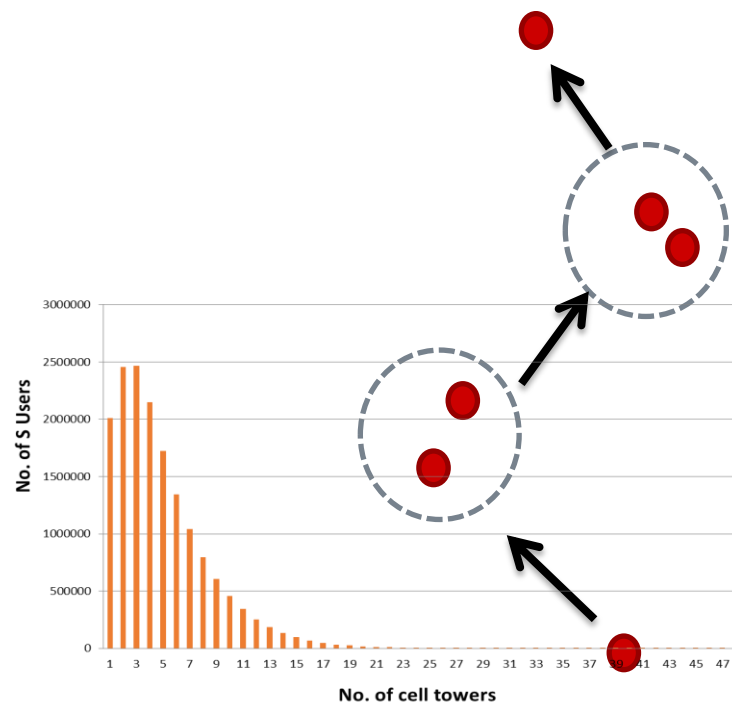
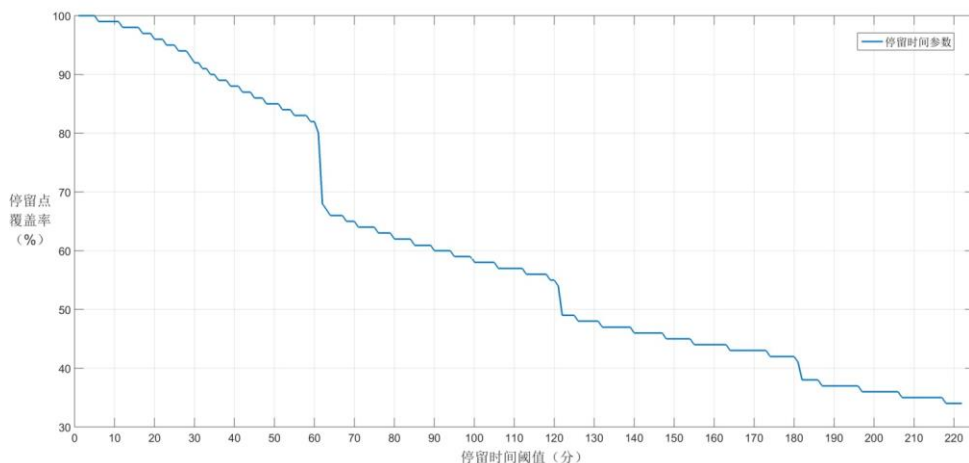


去重前后不同记录数下的用户数分布图

“根据深圳市出租屋综管办的最新统计，2011年10月底深圳登记在册的非户籍人口达到了1280万人，加上户籍人口274.7万，全市实际管辖人口已超过1500万人。”

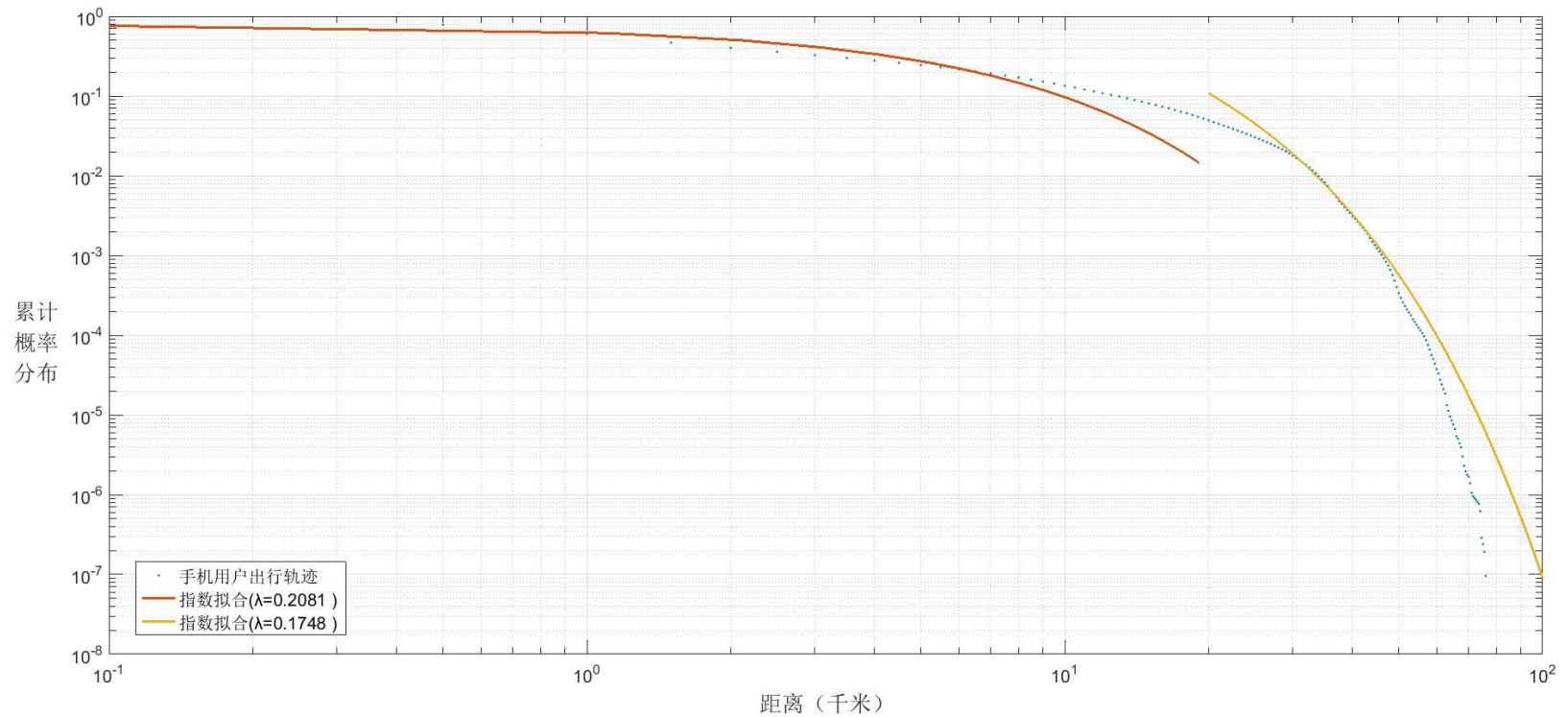
手机用户出行信息提取

- 停留时间阈值: 60 min
- 停留空间阈值: 500 m

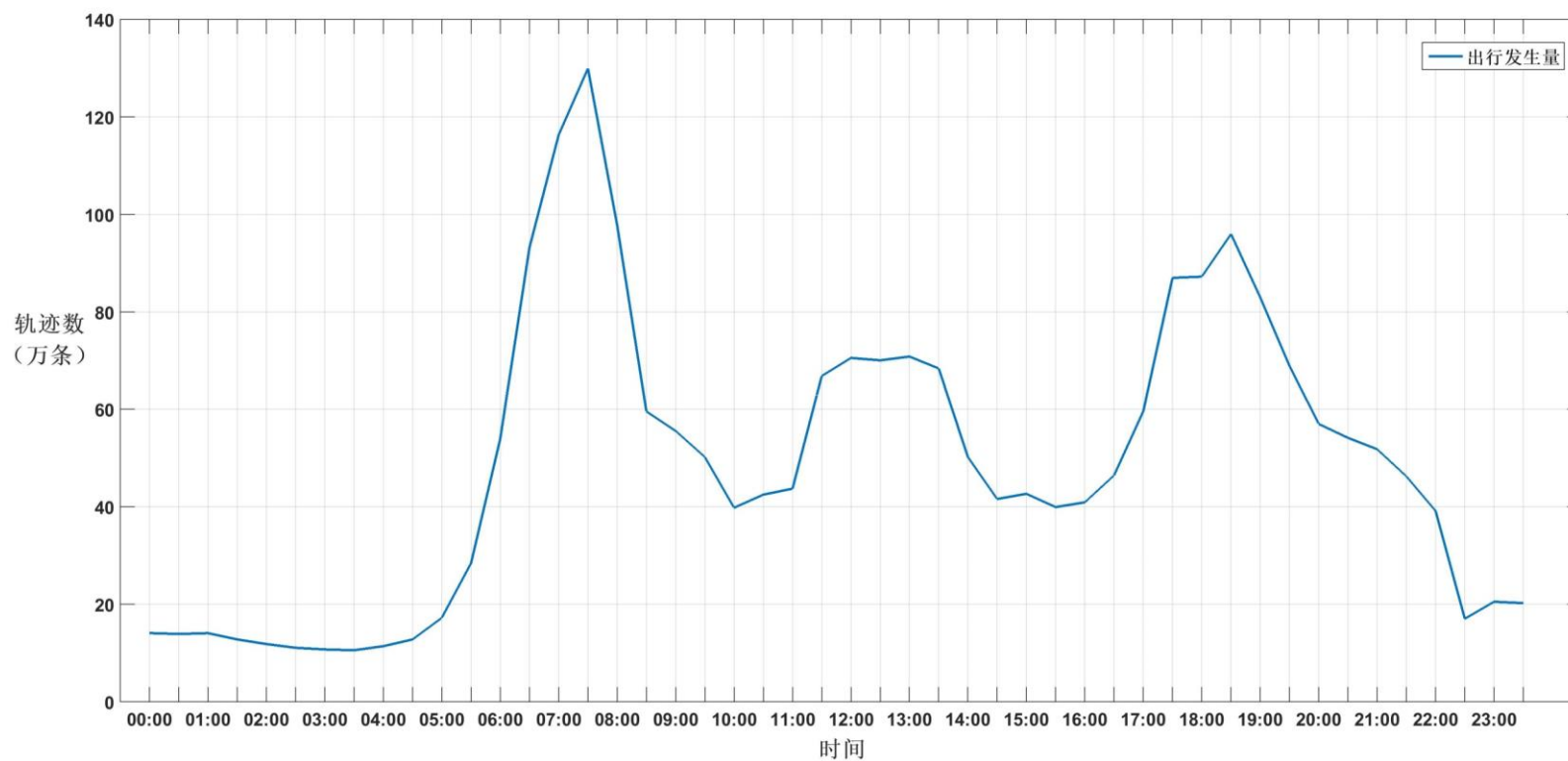


总出行数: 23,844,337次
平均次数: 1.5次

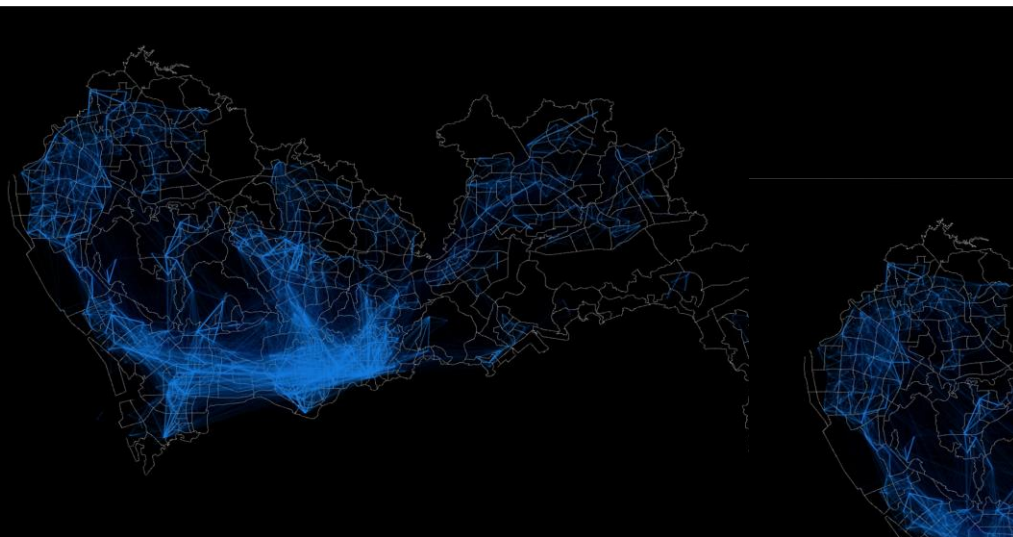
出行距离累计概率分布



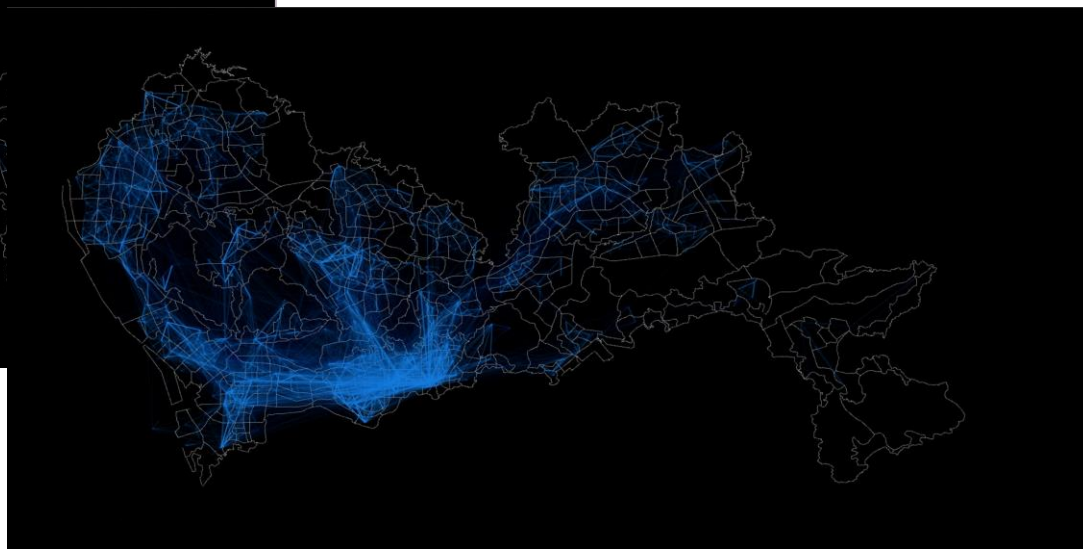
出行发生时间分布



出行空间分布

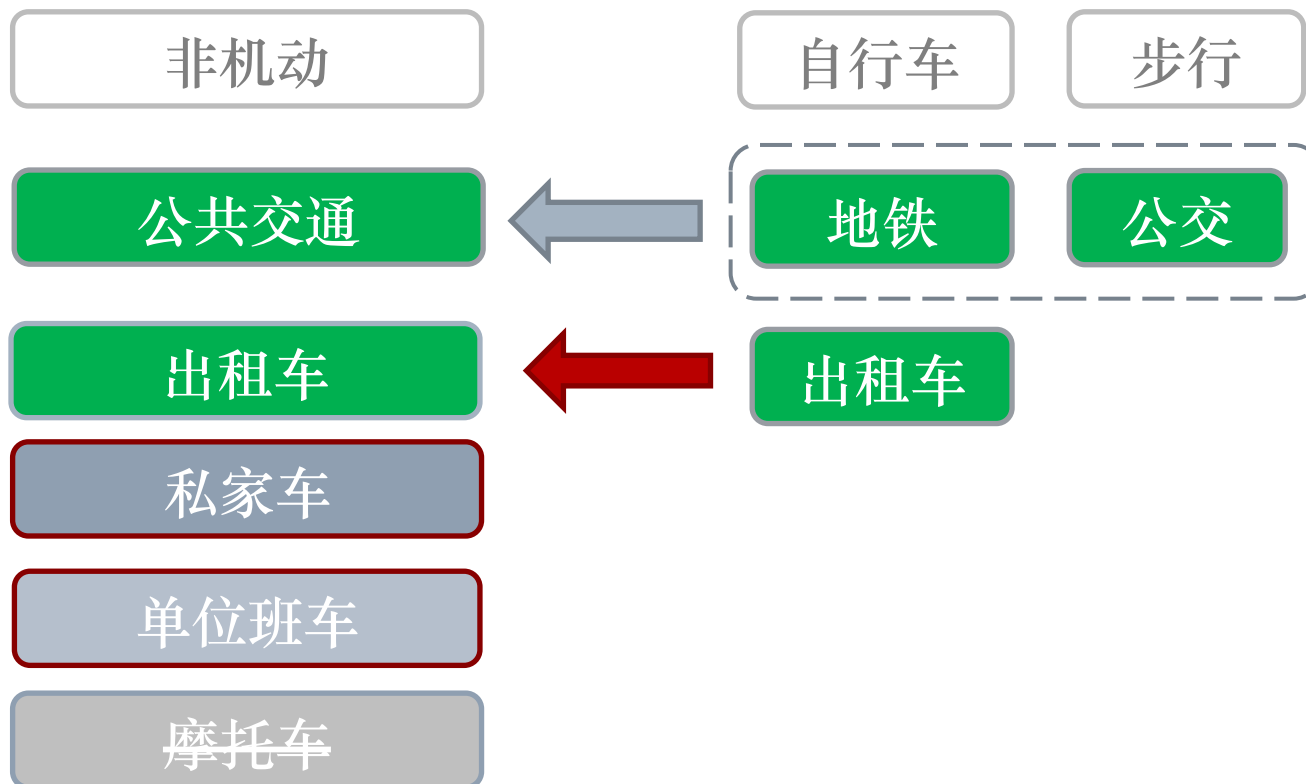


早高峰 (7:00-9:30)



晚高峰 (17:30-20:00)

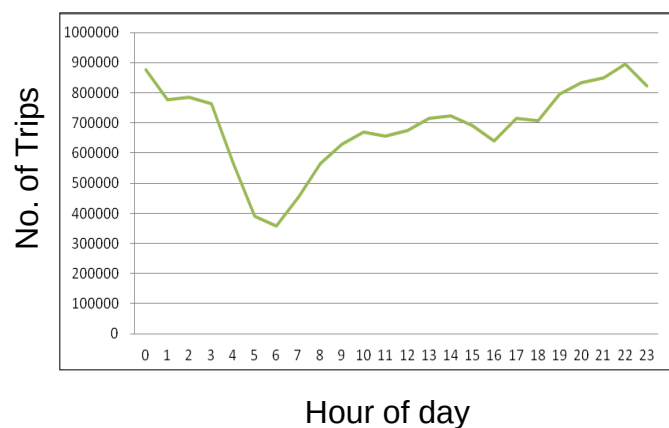
交通方式划分



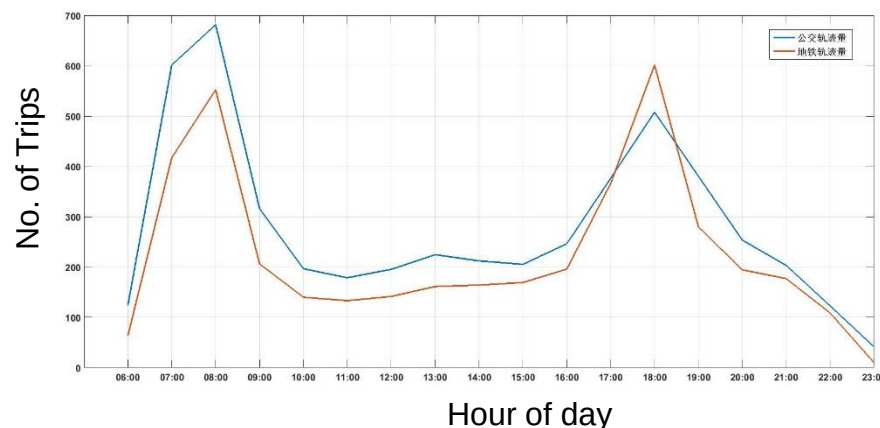
交通方式划分

Total no. of trips: 23,000,000

地铁	公交	的士	小汽车	(电动) 自行车	步行
7.70%	32.61 %	3.13%	?	/	/



Taxi: 720,000 trips/day



Metro: 1,770,000 trips/day;
Bus: 2,500,000 trips/day (x3)

交通方式划分

2010年深圳市居民出行调查与分析研究

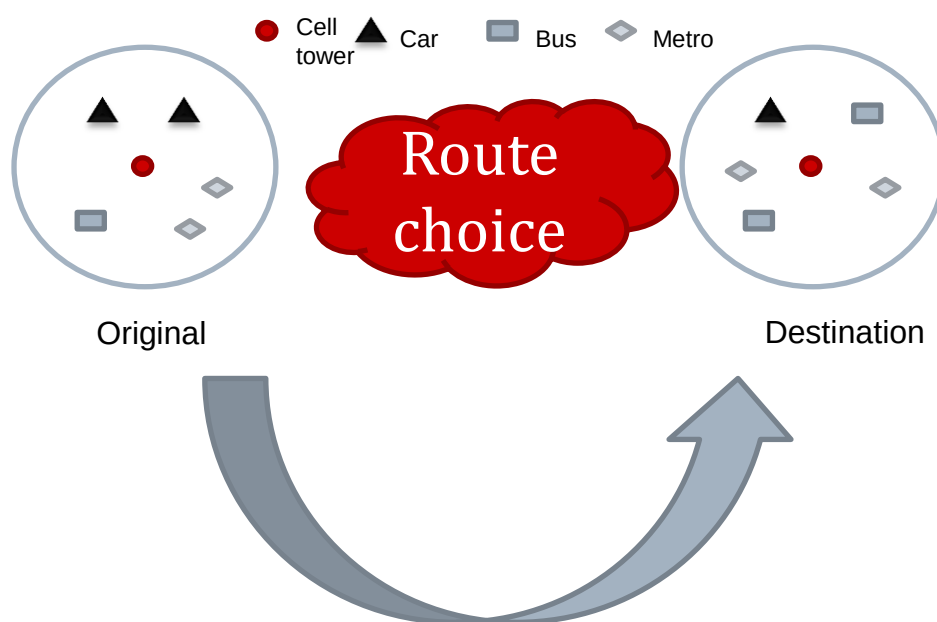
年份	摩托车	公共交通	的士	单位班车	小汽车
2000 年	35.3%	37.4%	2.2%	7.5%	17.6%
2005 年	7.1%	44.2%	3.3%	10.2%	35.2%
2010 年	4.9%	38.3%	3.7%	8.8%	44.2%

深圳私家车保有量:

2010: 170万 → 2012: 250万 → 2015: 320万

地铁	公交	出租车	私家车	自行车	步行
7.70%	32.61 %	3.13%	50.00%	/	/

出行路径分配



个人方式出行流量分配:

$$E_{n_p \cdot n_q} = ((N_i, C_{N_i}), (N_j, C_{N_j}), \dots)$$

$$F_{N_i} = \frac{C_{N_i}}{C_{N_i} + C_{N_j} + \dots} * F_{n_p \cdot n_q}$$

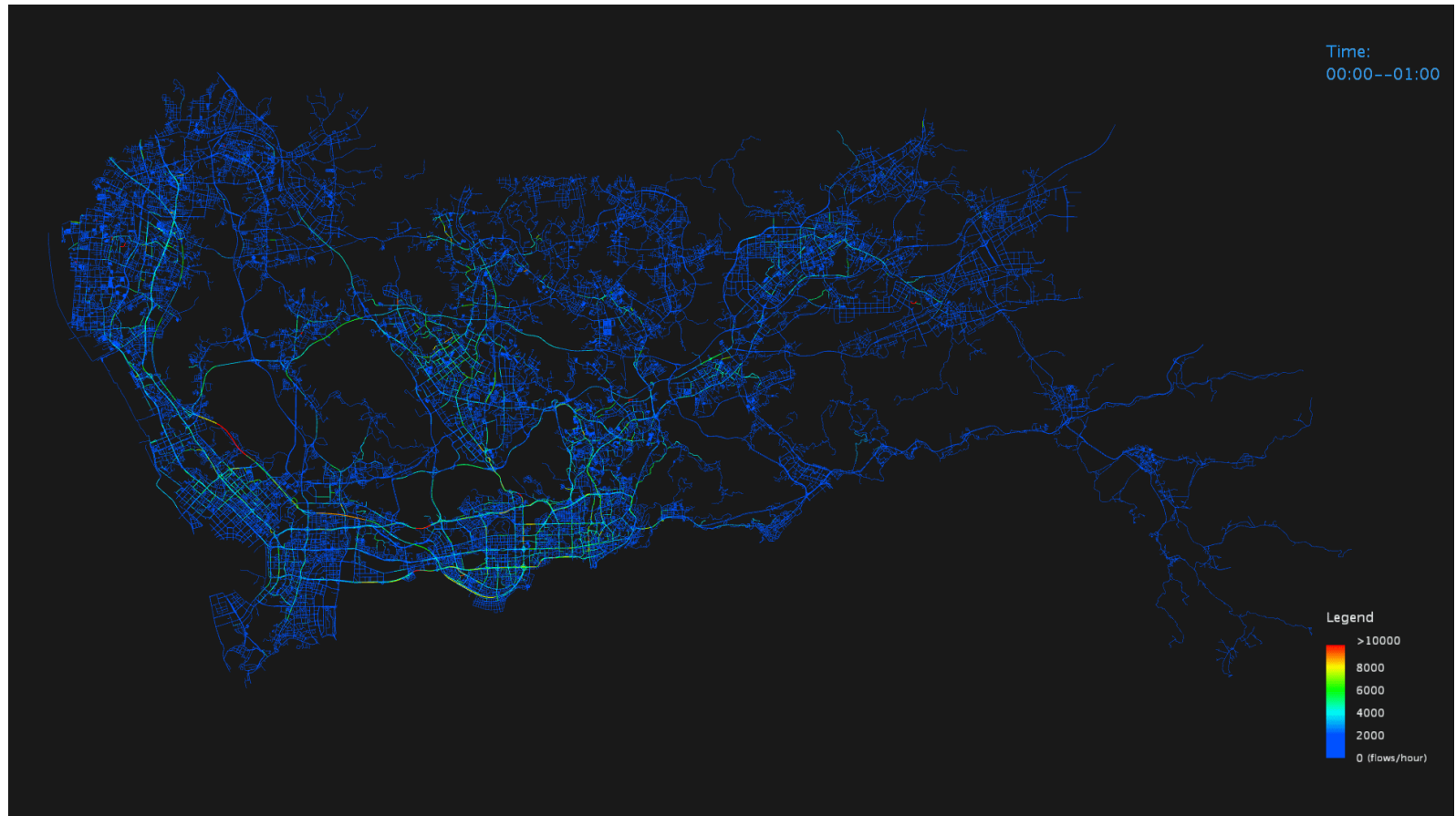
$$F_{n_{ij} \cdot n_{i(j+1)}} = F_{N_i}, 0 \leq j \leq k-1$$

公共方式出行流量分配:

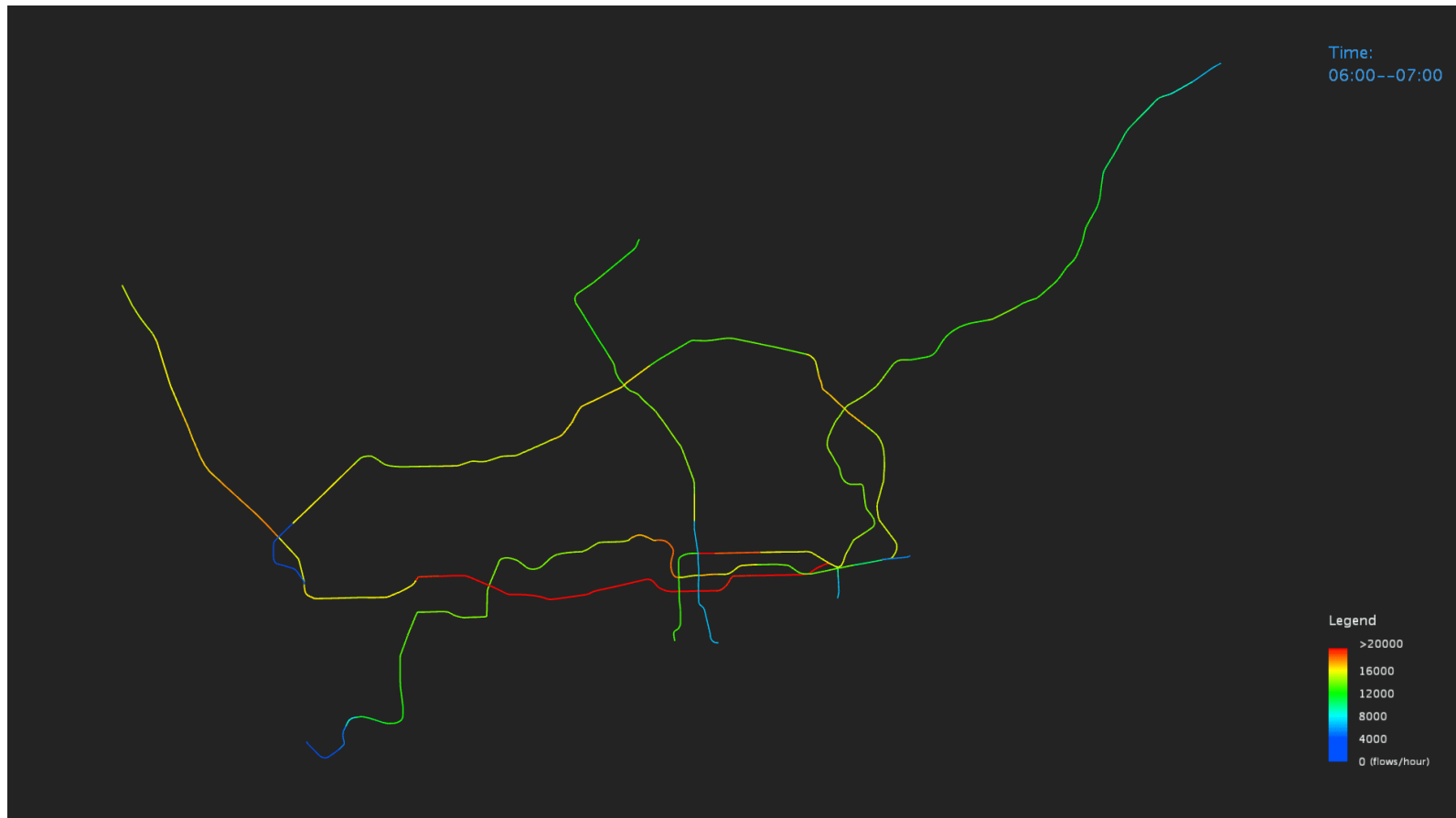
$$E_{s_p \cdot s_q} = (S, C_S) \quad F_S = F_{s_p \cdot s_q}$$

$$F_{s_i \cdot s_{i+1}} = F_S, 0 \leq i \leq k-1$$

路面交通需求预测结果



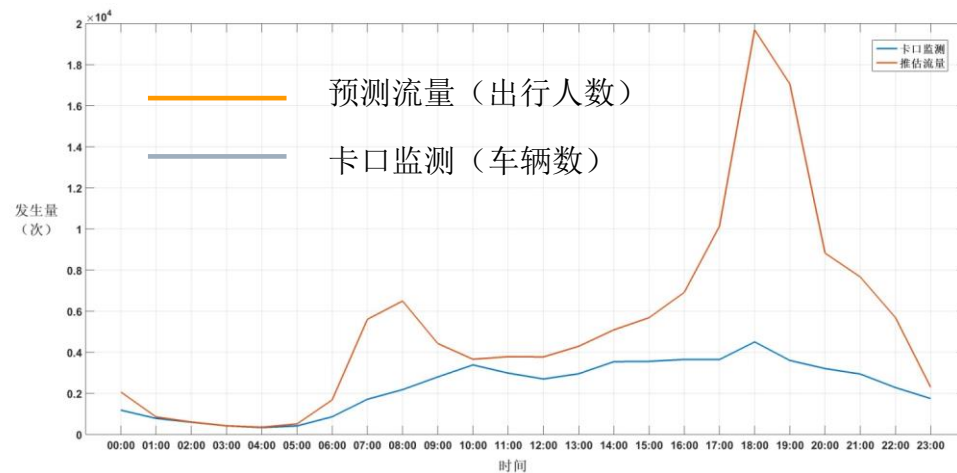
地铁线路交通需求预测结果



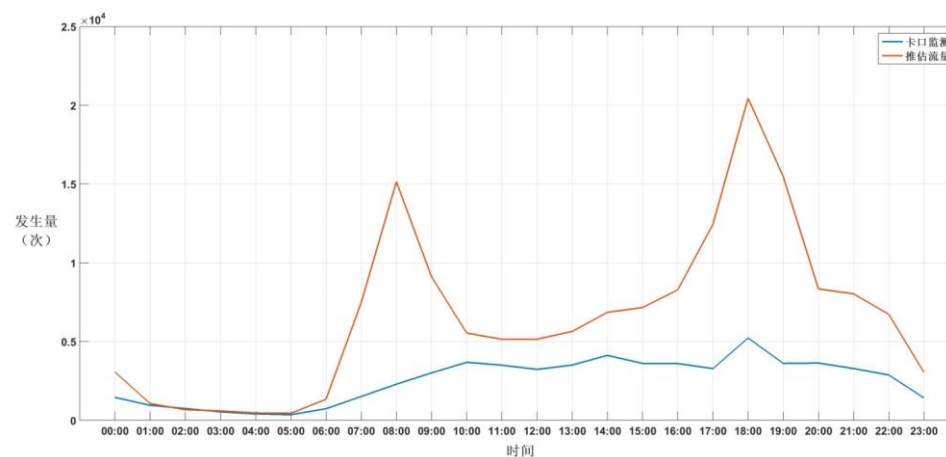
结果验证



卡口监测点: 260~
(2015年数据)



滨海大道沙河东立交西往东



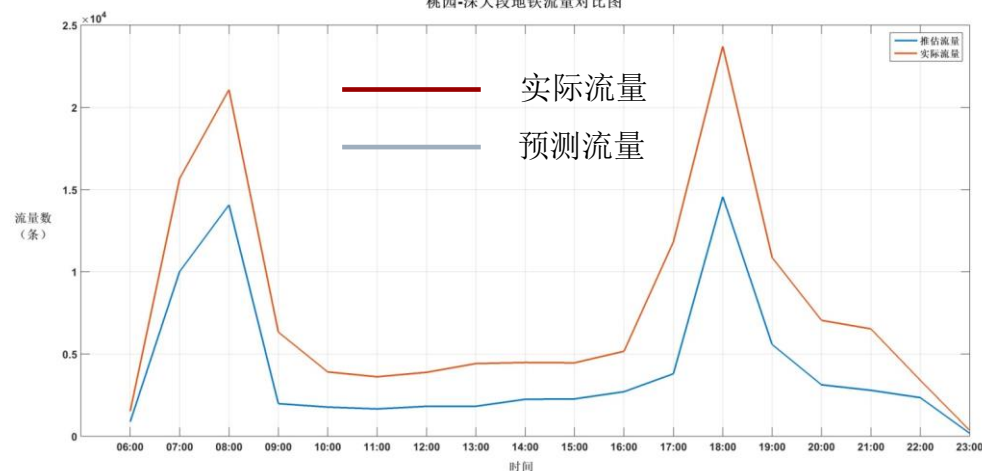
福龙路横龙山隧道入口南往北

结果验证

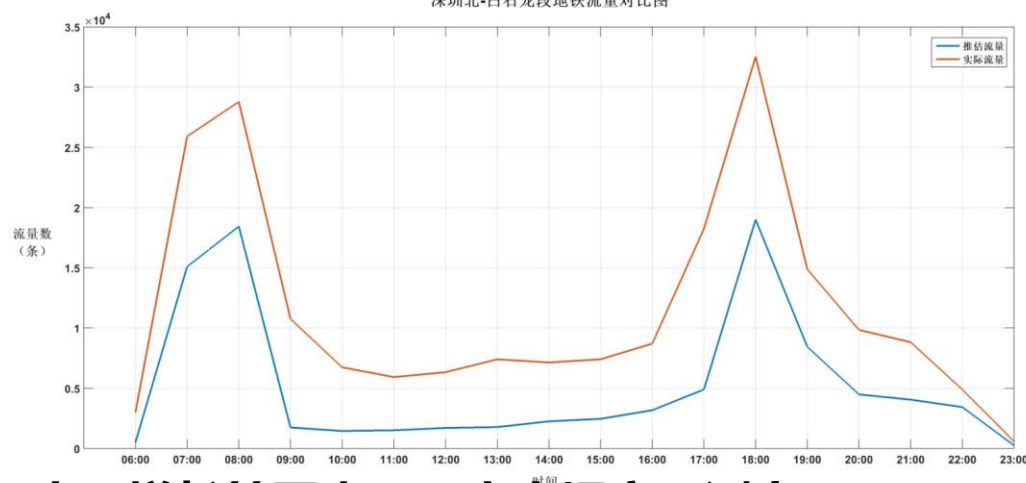


地铁路段数: 121

桃园-深大段地铁流量对比图



深圳北-白石龙段地铁流量对比图



若加入市场占有率、一人多机及人口增长等因素，可大大提高预测率

局限性

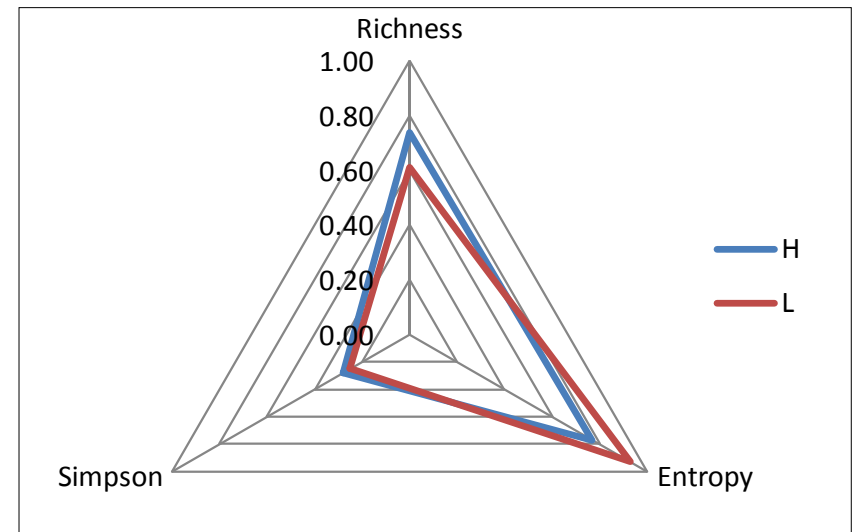
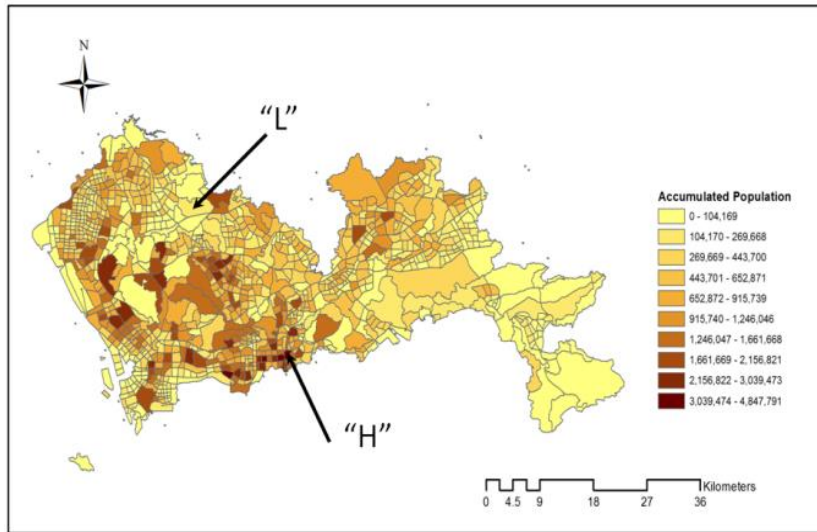
- 目前还不能对所有交通出行方式进行估算
- 手机用户与人口数量之间的偏差
- 此过程中尚未顾及人口的增长率

小结

- 利用数据驱动的方法，代替传统的四阶段模型，具有可行性
- 预测的范围受限于可获得的交通出行数据种类
- 随着电子车牌、公共自行车服务的推广，此方法预测精度会越来越高
- 进一步，利用大数据的优势进行细粒度的建模分析



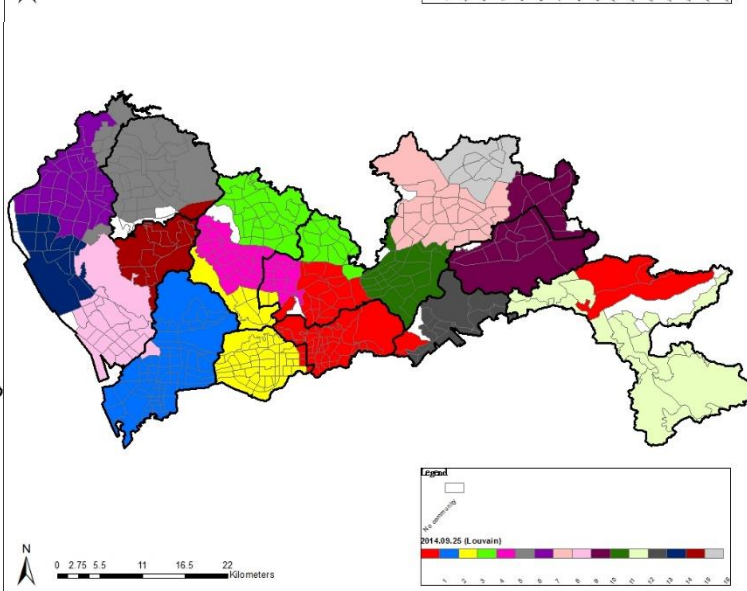
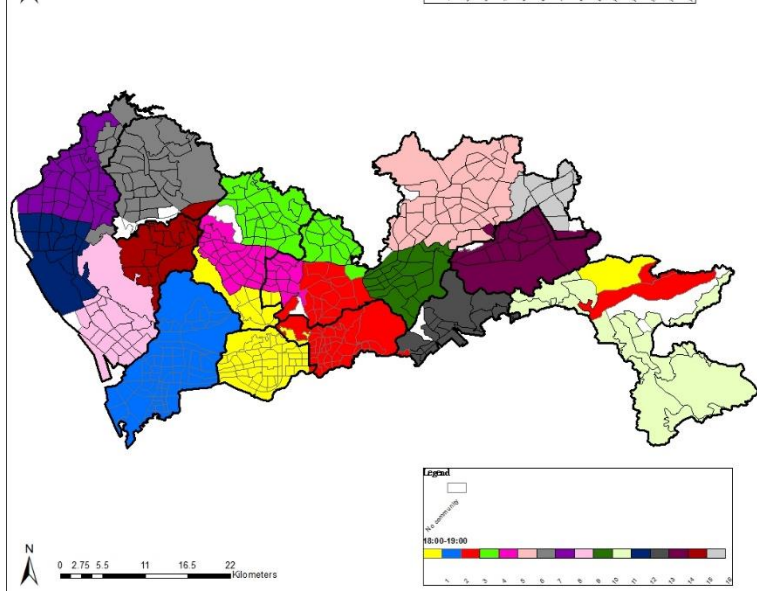
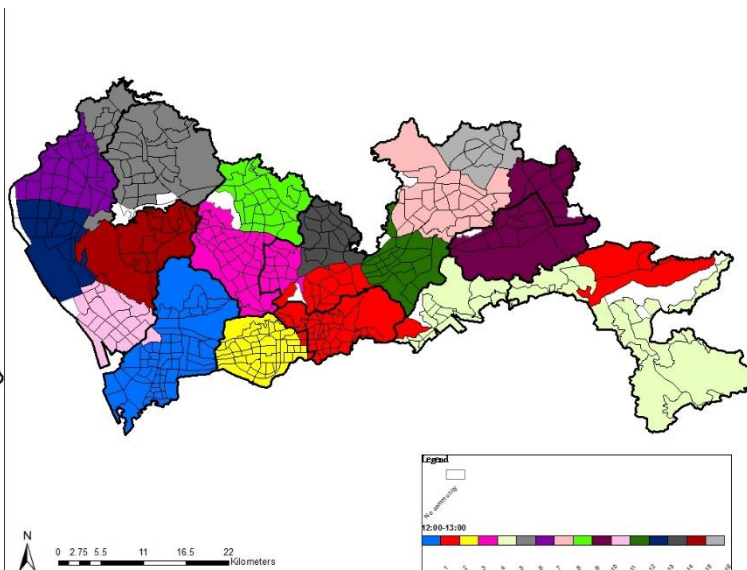
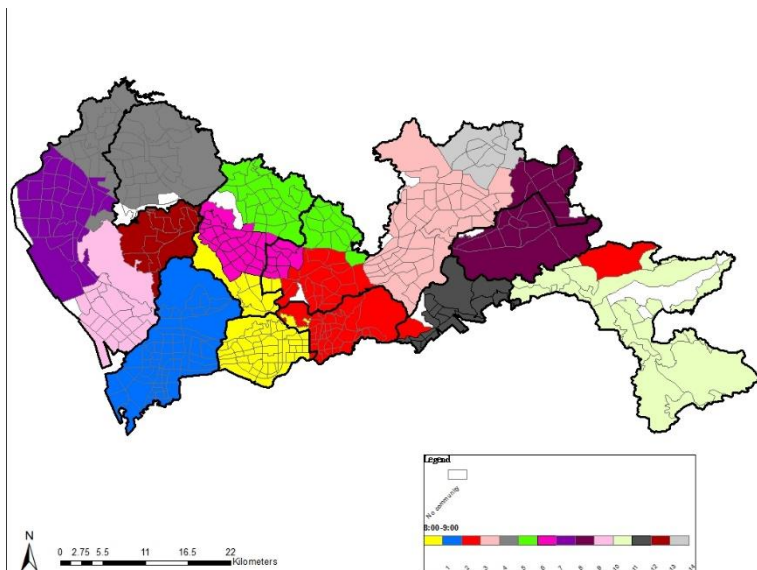
2. TAZ活力与土地混合利用的度量及关系



a) 基于手机用户量的TAZ活力度量;

b) 基于Hill numbers的土地混合度量

3. 基于个体活动轨迹的城市组团识别（手机）



上左：早高峰

上右：午高峰

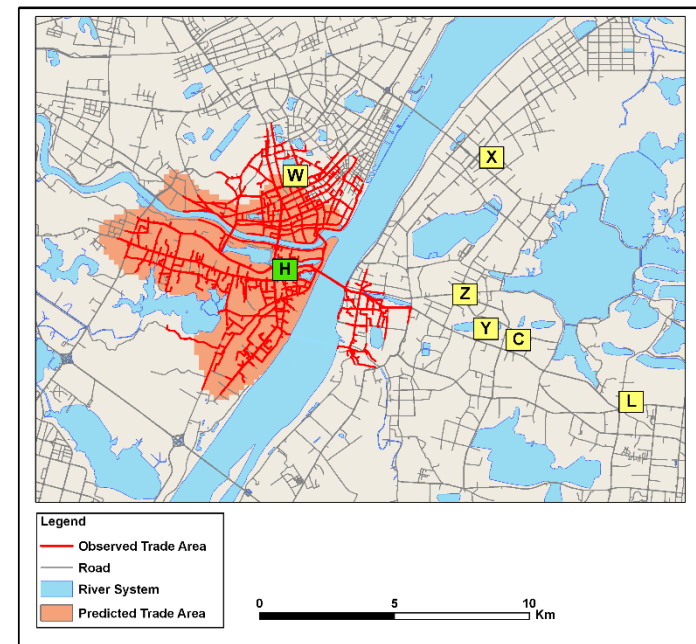
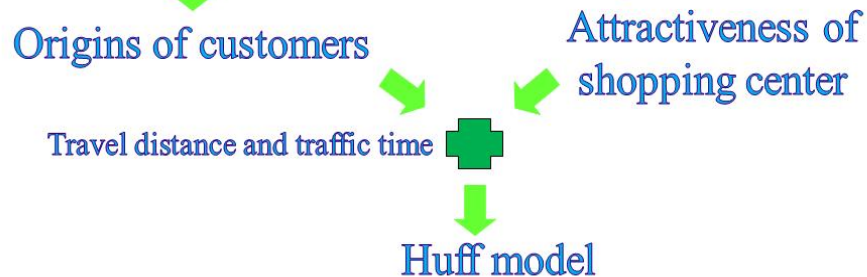
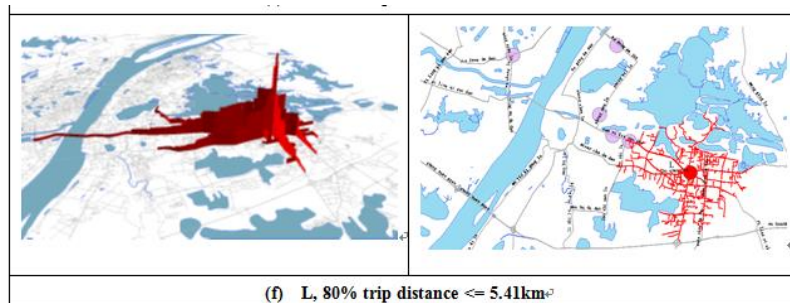
下左：晚高峰

下右：全天

4. 城市空间设施利用模式及影响力分析

商业区影响力分析可以用于发现商业区顾客来源、评价商业区的影响力、预测新商业区的位置。**传统的商业区影响分析**需要进行大范围的调查，**具有样本量少、费时费力等缺陷**。基于浮动车时空轨迹的商业区影响力分析方法能够真实捕捉商业空间的影响范围。

预测商圈范围

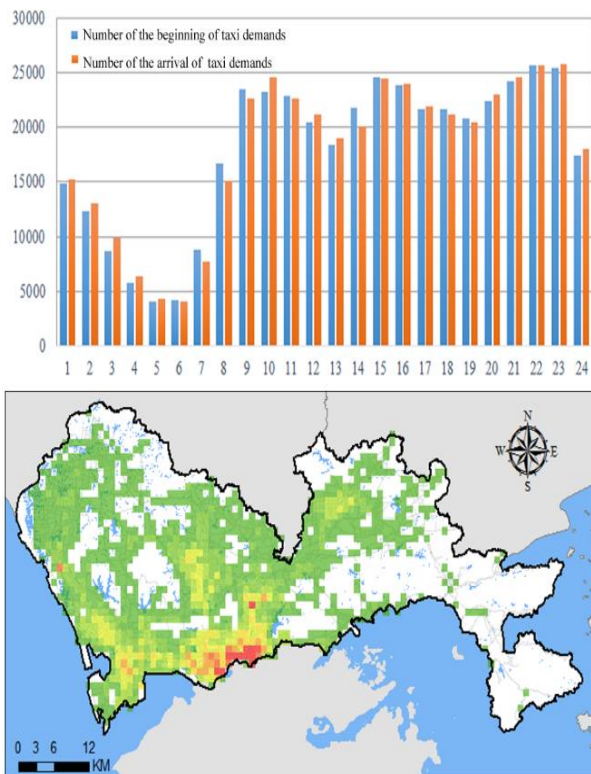


Yue, Y., Wang, H., Hu, B., **Li, Q.Q.**, Li, Y.G. and Yeh, A.G.O., 2012, Exploratory calibration of a spatial interaction model using taxi GPS trajectories. *Computers, Environment and Urban Systems*, 36, pp. 140-153.

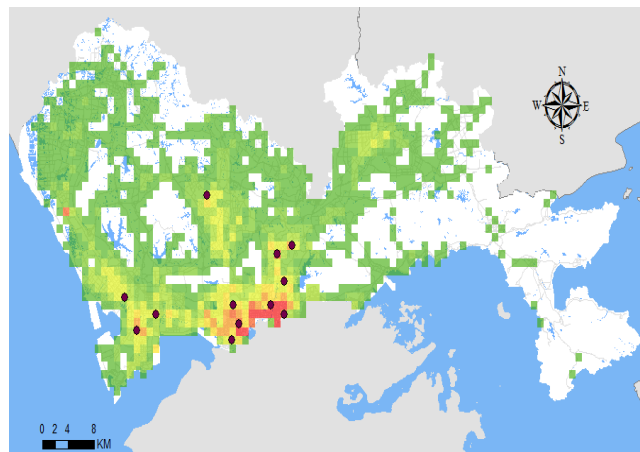
5. 基于时空轨迹数据的充电站选址

从出租车轨迹数据中提取附带时间标签的出行需求，考虑时空约束，顾及充电站容量，进行交通时空数据支撑下的电动汽车充电站选址。

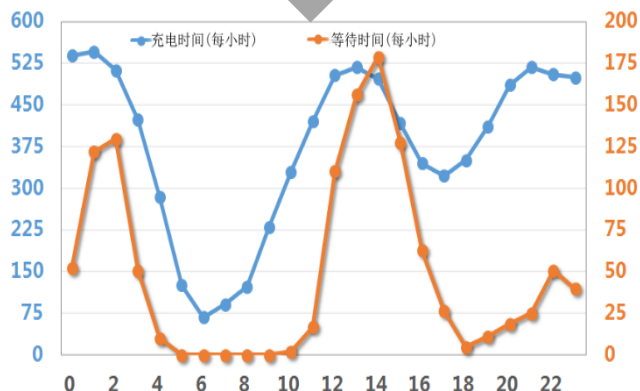
时空变化的出租出行行为



时空约束下的
选址模型



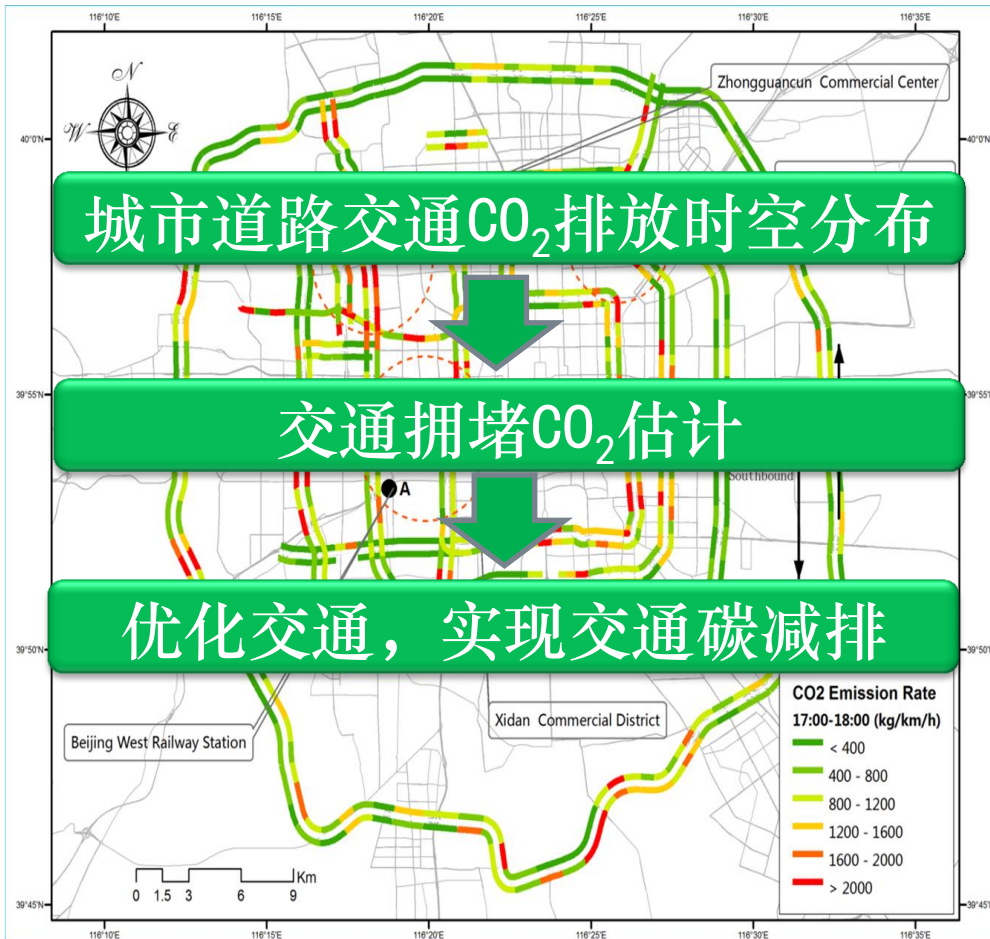
评估：时空变化的充电/等待行为



以深圳市电动出租车充电站选址为例

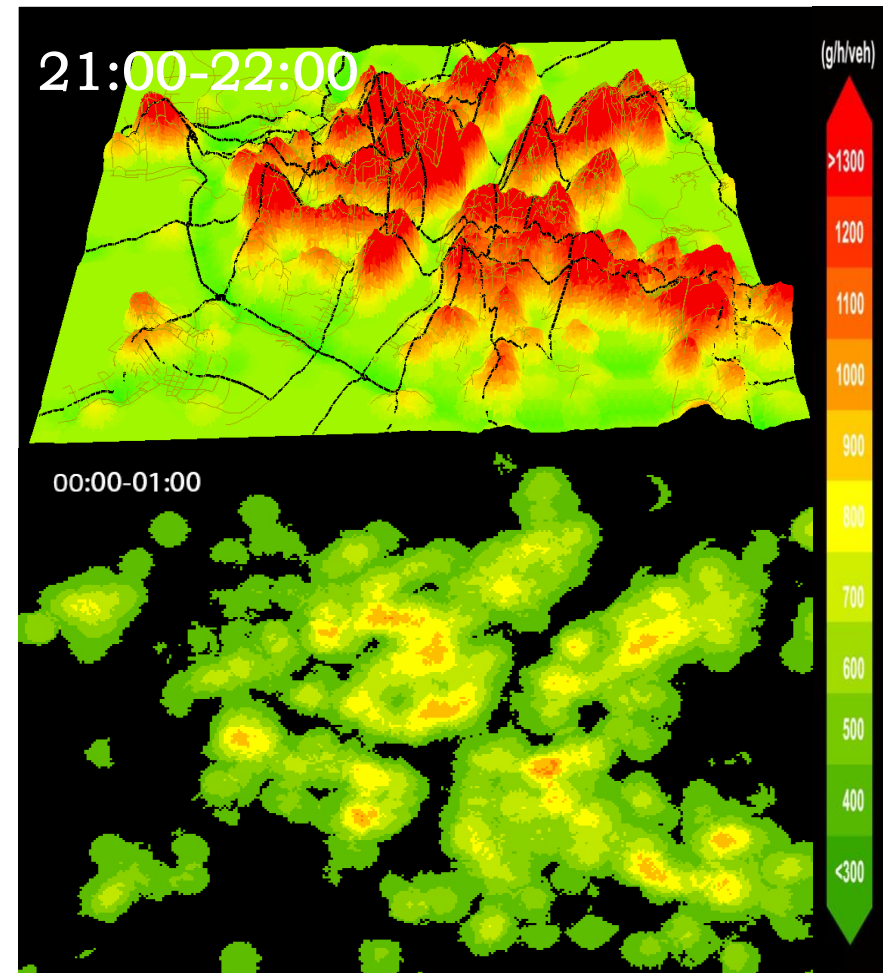
Wei TU*, Qingquan LI*, Zhixiang FANG, Shih-lung Shaw, Baoding ZHOU, Xiaomeng CHANG. Optimizing the locations of electric taxi charging stations: A spatial-temporal demand coverage approach. Transportation Research Part C: Emerging Technologies. 2015, 61.(SCI).

6. 城市交通碳排放制图

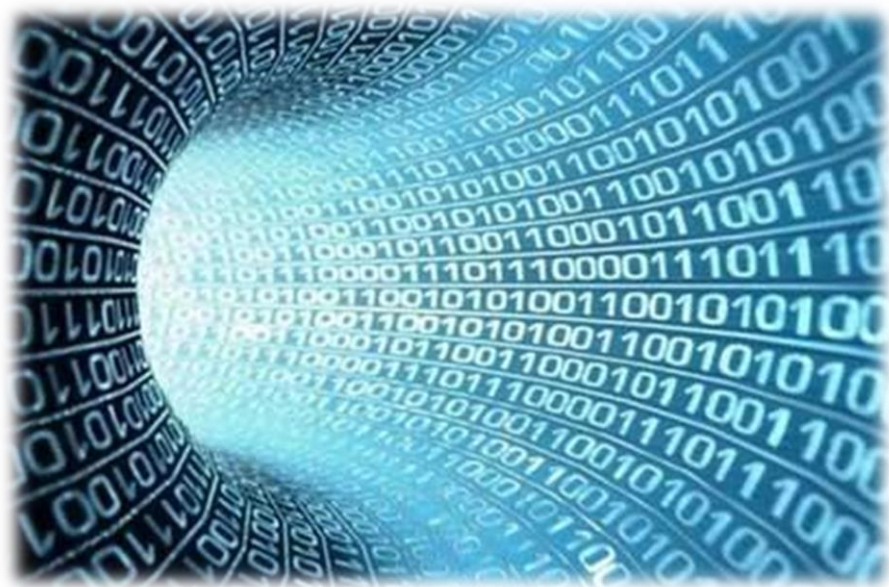


北京市环线CO₂排放量 (kg/h/km) 实时分布估计

Chang, X., Li, Q.*, et al. (2013). "Estimating Real-Time Traffic Carbon Dioxide Technologies." *Intelligent Transportation Systems, IEEE Transactions on* **14**(1): 469-479.



2013年，武汉机动车超过140万，假设平均每辆机动车每天运行1小时，每天因交通不畅造成的CO₂排放量超过1000吨



4

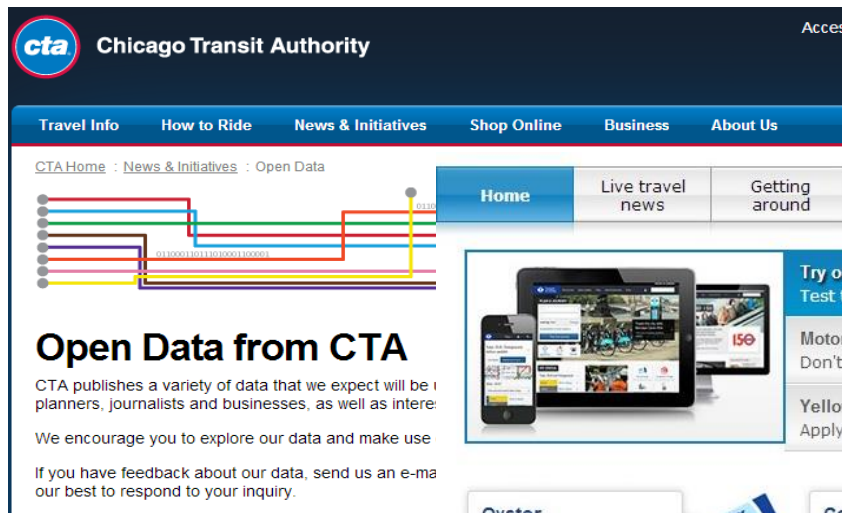
总结与展望

总结

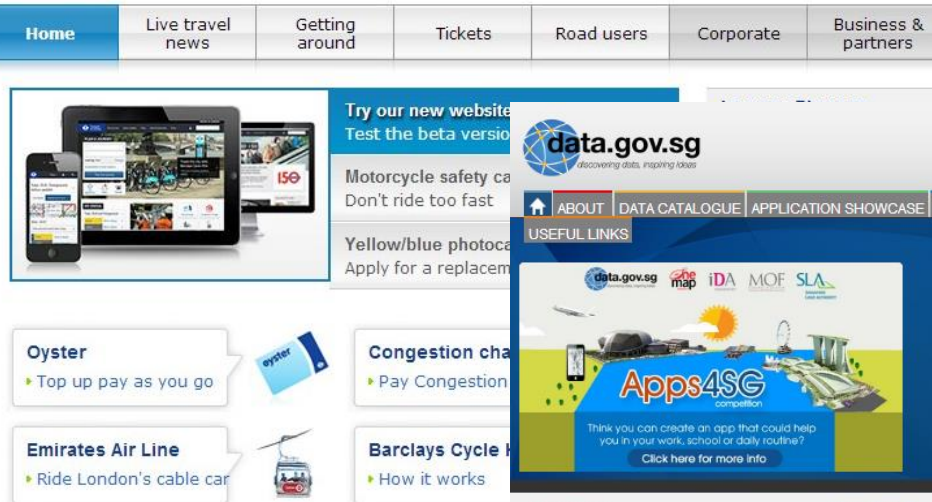
- **专业/非专业的多源交通时空数据带来了新的机遇，获得细时间分辨率、高空间分辨率的交通数据已经成为现实。**
- **多源交通时空数据伴随着多样化问题，不同应用环境下需考虑其适应性。**
- **以数据驱动为主，模型驱动为辅，是分析时空现象及交通问题的最佳途径。**

展望：数据开放

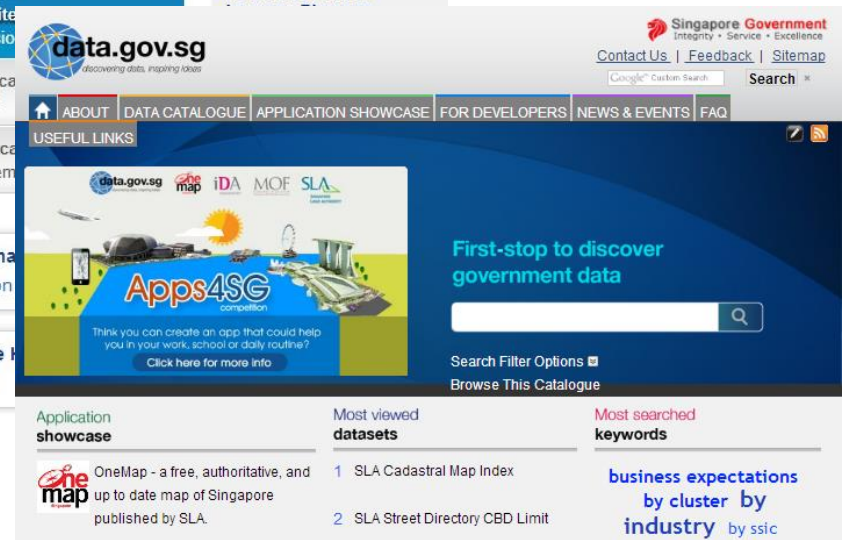
■ 开放数据、开放API、开放社区



The image shows the Chicago Transit Authority (CTA) website's 'Open Data' page. The header includes the CTA logo and navigation links: Travel Info, How to Ride, News & Initiatives, Shop Online, Business, and About Us. The main content area features a colorful map of the Chicago transit system and the heading 'Open Data from CTA'. Below this, it states: 'CTA publishes a variety of data that we expect will be of interest to planners, journalists and businesses, as well as interested citizens. We encourage you to explore our data and make use of it. If you have feedback about our data, send us an e-mail and we'll do our best to respond to your inquiry.'



The image shows the Singapore Government's data.gov.sg website. The header includes the data.gov.sg logo and navigation links: Home, Live travel news, Getting around, Tickets, Road users, Corporate, and Business & partners. The main content area features a large banner for 'Apps4SG competition' with the text 'Think you can create an app that could help you in your work, school or daily routine? Click here for more info'. Below the banner, there are sections for 'Oyster', 'Congestion charging', 'Emirates Air Line', and 'Barclays Cycle Hire'. The footer includes a search bar and links to 'Contact Us', 'Feedback', and 'Sitemap'.



The image shows the Singapore Government's data.gov.sg website, continuing from the previous block. The header includes the data.gov.sg logo and navigation links: ABOUT, DATA CATALOGUE, APPLICATION SHOWCASE, FOR DEVELOPERS, NEWS & EVENTS, and FAQ. The main content area features a large banner for 'Apps4SG competition' with the text 'Think you can create an app that could help you in your work, school or daily routine? Click here for more info'. Below the banner, there are sections for 'Oyster', 'Congestion charging', 'Emirates Air Line', and 'Barclays Cycle Hire'. The footer includes a search bar and links to 'Contact Us', 'Feedback', and 'Sitemap'.



[Home](#) [Computing Environment](#) [Dataset](#) [Submission](#)

ISPRS科学倡议数据公开大赛决赛

深圳大学空间信息

日程总览:

12月11日	
19:00-21:00	时空大数据分析、挖掘与可视化工作
2015/12/12	
上午：ISPRS科学倡议数据公开大赛决赛及	
9:00	开幕



时空数据分析与挖掘
条件下的社会治理体系和治理能力现代化路径探索
Understanding Human Dynamics with Mobile Phone Location Data
of Weather on Travel Behavior of Public Transit Users: from Mining Data from Different Sources
Decay in Healthcare Behavior and Delineation of Hospital 支持下的社会感知



敬请指正



 @深大_Spatial_Lab