BIG-DATA AND URBAN METABOLISM

Cases in China

Dong LI, PhD
Deputy Director
Innovation Center for Technology
Beijing Tsinghua Tongheng Urban Planning & Design Institute (THUPDI)
BEIJING TSINGHUA TONGHENG URBAN PLANNING & DESIGN INSTITUTE

Large State-owned Planning and Design institute

- In 1993, Beijing Tsinghua Urban Planning and Design Institute was established, attached to Tsinghua University.
- In Aug. 2012, from the ownership by the whole people to limited liability company, the overall company transferred to Tsinghua Holdings, officially renamed Beijing Tsinghua Tongheng Urban Planning and Design Institute.
- In Mar. 2013, Tsinghua Tongheng and several companies have undergone strategic reorganization, integrated into Tsinghua Holdings Human Settlement Group.
- The only company in industrial groups of Tsinghua University specialized in research and practice on urban planning and design.

Class A qualification in town and country planning
Class A qualification in national tourism planning and design
Class A qualification in survey and design on cultural relic protection project
Class A qualification in special design on landscape architecture
Class A qualification in construction industry (architecture)
Class B qualification in land planning
Certified by ISO9001:2008 quality management system.

Innovation Center for Technology
Established by Tsinghua Tongheng Planning and Design Institute in 2014

- The first department in domestic planning and design companies to set up Big Data R&D department
- Rooted in profound academic accumulation of Tsinghua University, and Tsinghua Tongheng’s deep understanding of city, utilizing multi-source data to develop urban innovative think tank business
- Provide emerging service model and technique product in the field of urban planning and management
- ~50 professionals from multiple disciplines

Tsinghua Tongheng provides comprehensive technical service of whole industry chain from strategy, planning and design to implementation and operation, and forms service mode of cross-disciplinary collaboration covering town and country planning, landscape architecture, architectural design, municipal traffic as well as science and technology media. Combining advanced concept and technology support with local characteristics, Tsinghua Tongheng offers customized service to government at all levels, such as national ministry, province, municipality, county (district and city), street, town and village, provincial and municipal departments, and park management committee as well as large enterprises (urban construction investment and operation company, and industrial operation company, etc.)

Transforming planning, design and research results into the driving force to supporting urban and rural development, Tsinghua Tongheng persists on paying back through planning and technology, and actively devotes to public welfare, volunteering to take on industrial mission and social responsibility.

Innovation Center for Technology

www.thupdi.com

Innovation Center for Technology

ict.thupdi.com
Providing chains of solutions for urban planning, operation, management and services in the context of big data.
Social-economic and urbanization trends in China

- Decreasing growth rates of economy and urbanization since 2011
Forwarding a better city life

**From urban sprawls to urban quality:** Traffic congestion, housing difficulties, environmental degradation, resource constraints and so on "big city disorders" is becoming more and more serious
A starting big-data era: data, computing and algorithms

- More than 3 billion people and 170 billion devices are connected to the net
  - The total amount of data before 2003: 5 exabytes
  - Before 2007: 300 exabytes
  - Before 2015: 966 exabytes
  - Till 2025: 170,000 exabytes (200 times forecasted)
  - ......

- Powerful devices / computing capability
  - Mobile phones now vs. PCs 10 years ago
    - cloud computing
  - Smarter algorithms
    - alphaGo in go game
    - alphaCity?
Big-data analytics matters in urban governance

Higher frequency, finer granularity and richer property

- In terms of time, space, subject, attribute, relationships and so on

Data: human ↔ Web ↔ IoT ↔ Data: city

New-type data

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>手机信令</td>
<td>企业注册信息</td>
</tr>
<tr>
<td>一卡通</td>
<td>滴滴出行</td>
</tr>
<tr>
<td>互联网房价</td>
<td>摩拜单车数据</td>
</tr>
<tr>
<td>互联网 LBS</td>
<td>路网矢量</td>
</tr>
<tr>
<td>健身APP</td>
<td>出租车GPS</td>
</tr>
<tr>
<td>大众点评</td>
<td>P0I设施</td>
</tr>
</tbody>
</table>

Traditional data

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>统计数据</td>
<td>普查数据</td>
</tr>
<tr>
<td>抽样数据</td>
<td>调查数据</td>
</tr>
<tr>
<td>……</td>
<td>……</td>
</tr>
</tbody>
</table>
Case 1: Landfills and social media

Landfill odors have created a major concern for the Chinese public. Based on the combination of a first order decay (FOD) model and a ground-level point source Gaussian dispersion model, the impacts from odors emitted from the 1955 landfills in China are evaluated in this paper. Our bottom-up approach uses basic data related to each landfill to achieve a more accurate and comprehensive understanding of impact of landfill odors. Results reveal that the average radius of impact of landfill odors in China is 796 m, while most landfills (46.85%) are within the range of 400~1000 m, in line with the results from previous studies. The total land area impacted by odors has reached 837,476 ha, accounting for 0.09% of China's land territory. Guangdong and Sichuan provinces have the largest land areas impacted by odors, while Tibet Autonomous Region and Tianjin Municipality have the smallest. According to the CALPUFF (California Puff) model and an analysis of social big data, the overall uncertainty of our calculation of the range of odor impacts is roughly 32.88% to 32.67%. This type of study is essential for gaining an accurate and detailed estimation of the affected human population and will prove valuable for addressing the current Not In My Back Yard (NIMBY) challenge in China.
Case 1: Landfills and social media

- H₂S emission estimation
- Landfill odor diffusion
  - Gaussian dispersion model
  - CALPUFF model
- Geo-tagged Social media data
  - Locations and contents

First, we identified all the microblogs with geographical coordinates and selected those containing the words “landfill” and “odors.” The selected microblogs were then assessed individually to make sure they actually reflected landfill odor impacts. The distance from where the microblog was submitted to the nearest landfill was set as the odor impact range of that landfill. For data from news reports, we selected those containing the specific odor impact distance and specific landfill. These data from social media and news reports were then compared with the results from our physical models.

It should be noted that the distance determined by this method might be smaller than the actual affected range of landfill odor, so the data could underestimate the actual influence range of landfill odors to a certain extent.
Case 1: Landfills and social media

Results reveal that the average radius of impact of landfill odors in China is 796 m, while most landfills (46.85%) are within the range of 400~1000 m, in line with the results from previous studies. The total land area impacted by odors has reached 837,476 ha, accounting for 0.09% of China’s land territory. Guangdong and Sichuan provinces have the largest land areas impacted by odors, while Tibet has the smallest.

<table>
<thead>
<tr>
<th>Provinces</th>
<th>H$_2$S emissions (kg)</th>
<th>Range of impacts (ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>8537</td>
<td>14391</td>
</tr>
<tr>
<td>Tianjin</td>
<td>2425</td>
<td>1410</td>
</tr>
<tr>
<td>Hebei</td>
<td>6961</td>
<td>28055</td>
</tr>
<tr>
<td>Shanxi</td>
<td>4944</td>
<td>12733</td>
</tr>
<tr>
<td>Inner Mongolia</td>
<td>5259</td>
<td>5646</td>
</tr>
<tr>
<td>Liaoning</td>
<td>10353</td>
<td>21735</td>
</tr>
<tr>
<td>Jilin</td>
<td>6757</td>
<td>17464</td>
</tr>
<tr>
<td>Heilongjiang</td>
<td>5513</td>
<td>6546</td>
</tr>
<tr>
<td>Shanghai</td>
<td>11164</td>
<td>45086</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>10543</td>
<td>21964</td>
</tr>
<tr>
<td>Zhejiang</td>
<td>19901</td>
<td>61268</td>
</tr>
<tr>
<td>Anhui</td>
<td>8726</td>
<td>19227</td>
</tr>
<tr>
<td>Fujian</td>
<td>4689</td>
<td>15465</td>
</tr>
<tr>
<td>Jiangxi</td>
<td>8170</td>
<td>41356</td>
</tr>
<tr>
<td>Shaanxi</td>
<td>11002</td>
<td>6962</td>
</tr>
<tr>
<td>Henan</td>
<td>7166</td>
<td>25983</td>
</tr>
<tr>
<td>Hubei</td>
<td>7094</td>
<td>48234</td>
</tr>
<tr>
<td>Hunan</td>
<td>10650</td>
<td>58136</td>
</tr>
<tr>
<td>Guangdong</td>
<td>27749</td>
<td>112203</td>
</tr>
<tr>
<td>Guangxi</td>
<td>4909</td>
<td>14934</td>
</tr>
<tr>
<td>Hainan</td>
<td>849</td>
<td>3337</td>
</tr>
<tr>
<td>Chongqing</td>
<td>4816</td>
<td>30963</td>
</tr>
<tr>
<td>Sichuan</td>
<td>8929</td>
<td>78245</td>
</tr>
<tr>
<td>Guizhou</td>
<td>3158</td>
<td>14529</td>
</tr>
<tr>
<td>Yunnan</td>
<td>5565</td>
<td>23344</td>
</tr>
<tr>
<td>Tibet</td>
<td>347</td>
<td>1634</td>
</tr>
<tr>
<td>Shaanxi</td>
<td>6404</td>
<td>47321</td>
</tr>
<tr>
<td>Gansu</td>
<td>3523</td>
<td>27379</td>
</tr>
<tr>
<td>Qinghai</td>
<td>2334</td>
<td>5575</td>
</tr>
<tr>
<td>Ningxia</td>
<td>1328</td>
<td>7712</td>
</tr>
<tr>
<td>Xinjiang</td>
<td>6258</td>
<td>18639</td>
</tr>
<tr>
<td>Total</td>
<td>226623</td>
<td>837476</td>
</tr>
</tbody>
</table>
Case 1: Landfills and social media

Social media relevant to landfills per province

字体大小代表词频的高低；“垃圾填埋场”不作为关键词进行分析。

垃圾填埋场微博语义和关键词分析结果
Case 1: Landfills and social media

According to the CALPUFF (California Puff) model and an analysis of social big data, the overall uncertainty of our calculation of the range of odor impacts is roughly 32.88% to 32.67%.

CALPUFF

Geo tagged social media: yellow dots

Fig. 5. Comparison of the simulated results between the CALPUFF and Gaussian dispersion models. Landfill names (and locations): (a) Siyuan (Beijing); (b) Langang (Shanghai); (c) Mayuan (Nanchang); (d) Xingfeng (Guangzhou); (e) Xiaoying (Shenzhen); (f) Changchuan (Chongqing); (g) Chongli (Shanghai); (h) Jiangtou (Xi’an); (i) Shanyang (Xinjiang).
Case 1: Landfills and social media

Demographic structure relevant to landfills (landscan)

12,280,000 ppl in total (landscan)

- Adult (15-65 years): 964万
- Child (<15 years): 164万
- Aging (>65 years): 100万
- Male: 628万
- Female: 600万

- Population affected Static (landscan)
  - Land covering: 0.09%
  - Population affected: 0.90%

- Population affected dynamic (social media)
  - 1.82%
Case 1: Landfills and social media

Study of Facility Distribution Indicators of Typical Landfills Based on a Distance-density Relationship

Issues for planning and management:
- to better evaluate the location of landfills based on a higher resolution and dynamic dataset of population
  - Data: LandScan (static) vs. geo-tagged Social media (dynamic)
  - for overall pattern in a city and for specific facilities
  - Comparing different cities within same dataset and framework
Case 1: Landfills and social media

Landfills and geo-tagged social media in Beijing and Shanghai
Case 1: Landfills and social media

IDW: inverse distance weighted

\[ P_i = \frac{S_{id} \times (d_{\text{max}} - d)}{\sum_d S_{id}} \]
Case 1: Landfills and social media

LandScan

Geo-tagged social media

上海老港（浦东）（编号 409）布局指数 P 最高，潜在的环境影响暴露水平最大，其次是 北京丰台垃圾场（编号 3）、北京六里屯（编号 5）等处理设施，而上海崇明（编号 426）布局指数 P 最低
Case 2: City GHG grid management platform

Competition proposal for Wuhan Urban Simulation Lab

- state-of-art GHG inventory for cities
- Quality control of data results
- Finer granularity GHG grid in Wuhan: spatial (1km) and time (every month)
- Management system with a modern and easy interface: analyzing and visualizing

武汉城市仿真实验室

碳排放模块

团队编号：02
团队名称：城市碳排放管理系统工作组

研究思路

目标和技术路线

- 建立武汉市高空间（1 km）、高时间（1月）分辨率二氧化碳排放清单体系；
- 建立武汉市基于网格的城市低碳管理综合模型；
- 建立武汉城市低碳评估可视化决策、分析系统。
Case 2: City GHG grid management platform

Competition proposal for Wuhan Urban Simulation Lab
- state-of-art GHG inventory for cities
- Quality control of data results
- Finer granularity GHG grid in Wuhan: spatial (1km) and time (every month)
- Management system with a modern and easy interface: analyzing and visualizing

主要步骤:
- 活动水平质量控制和典型企业第三方核查；
- 排放因子精度控制和抽样校对；
- 基于蒙特卡罗方法的不确定性分析。

考虑关键排放源、验证成本、准确性、验证方案和实施的复杂性、数据可获得性等。
Case 2: City GHG grid management platform

Competition proposal for Wuhan Urban Simulation Lab

- state-of-art GHG inventory for cities
- Quality control of data results
- Finer granularity GHG grid in Wuhan: spatial (1km) and time (every month)
- Management system with a modern and easy interface: analyzing and visualizing
Case 2: City GHG grid management platform

Competition proposal for Wuhan Urban Simulation Lab

- State-of-art GHG inventory for cities
- Quality control of data results
- Finer granularity GHG grid in Wuhan: spatial (1km) and time (every month)
- Management system with a modern and easy interface: analyzing and visualizing

Ad-hoc and interactive online analysis:
hot-cold area of GHG emissions
other cases

Data Fusion Analysis: for Public Services and healthy cities

Internet street View

Dock-less floating bike usages

Computer vision analysis of Street View image segmentation

Space-time analysis of bicycle riding trajectory hotspot
Data-driven Healthy cities: the movement trajectory and street view

The running trajectory is very significantly distributed around the Greenfield Park in Beijing. The Olympic Forest Park is one of the most popular places to run. Yuyuantan Park, Chaoyang Park, the Forbidden City moat is also a popular running place. On weekends, the above popular places will attract more people to come for a run.

Running track of weekday and distribution of green space in Beijing

The trajectory of weekend running and the distribution of green space in Beijing
other cases

Data Fusion Analysis: for Public Services and healthy cities

Data-driven Healthy cities: the movement trajectory and street view

- Sports Fitness Big Data
- Street View Map Big Data

Collect more than 1.6 million Street View data in Beijing six ring, use deep learning computer vision algorithm, semantic segmentation of streetscape image

Keep

User sample approx. 30,000
More than 200,000 behavior records
other cases

Data Fusion Analysis: for Public Services and healthy cities

Data-driven Healthy cities: the movement trajectory and street view

**Geo-detector**: a statistical method for detecting spatial differentiation and revealing the driving force behind indicators, the degree to which a factor X is identified by using the geo-factor detection method explains the spatial separation level of attribute Y.

Factors that significantly affect the amount of trajectory (descending):

- Proportion of the sky
- Distance from green space
- Distance from a settlement
- Road density
- Proportion of roads
- Nearby area of green space within 100 meters
- Proportion of buildings
- Percentage of cars
- Number of residential areas within 100 meters of the vicinity
other cases
Data Fusion Analysis: for Public Services and healthy cities

- **Dynamic operation monitoring of the functional core district of Beijing**
  
  - Outdoor environment sensors, human and traffic flow monitoring sensors were installed in the streets.
  - Analysis of spatial and temporal characteristics, human and traffic flow characteristics, job-housing characteristics based on data collection by the monitors.
  - More detailed understanding of urban operation, as well as scientific basis for further environment improvement.

Outdoor environment sensor  human and traffic flow monitoring sensor  Real-time visualization and analysis of monitoring data
Institute for China Sustainable Urbanization of Tsinghua University was co-established by National Development and Reform Commission and Tsinghua University in March 2016. Targeted at the theme of new urbanization, the Institute provides theoretical, intellectual and technical support for the new urbanization construction with Chinese characteristics. A comprehensive quality assessment system for the new urbanization construction is being developed. The Institute also researches on key engineering technologies and standards in the new urbanization field, and takes parts in a number of relevant evaluation systems and standards. Through the system based on multi-dimensional big data, the researchers are able to reveal the operational characteristics, trends and problems of urban agglomerations, and provide quantitative analysis support for the new urbanization research.
other cases
Big Data Platform for Cities

- **Next generation platform in Beijing**

  Forthcoming in 2019, hopefully

1. **Principal Method Framework of Urban Policy Evaluation**
2. **An Intelligent Platform of Urban Spatial Governance and Policy Evaluation for Beijing**
Loop 1: The business loop

- **Observation**: is the organization and interpretation of external sensory information by the human brain. Includes access to sensory information, understanding of information, screening of information, and organization of information.

- **Orient**: in psychology refers to the process of acquiring knowledge by forming psychological activities such as concept, perception, judgment or imagination, that is, the psychological function of individual thinking to information processing.

- **Decision Making**: is a cognitive process, after which individuals can decide to act on the basis of their personal beliefs or the reasoning of a combination of factors in a variety of options. Each decision-making process aims to produce a final decision and choose the final choice. And the form of these choices can be an action or a selection of opinions.

- **Action**: refers to the way an organism behaves, as well as a reaction to the environment in which it is located and other organisms or objects.
Loop 2: The data loop (workflow)

- **Data Perception**: the collection of data is cleaned, and the reality is more refined.
- **Data Measurement**: The analysis and evaluation of the data makes it more reasonable to judge the situation.
- **Data Mining**: a comprehensive interpretation of the data, understanding the reasons more deeply.
- **Data Simulation**: Deduction, extension and application of data mining rules.
Remarks 1: CPS - bridging between virtual and reality

Cyber-Physical System

Border closure, clear mechanism  Open border, fuzzy mechanism

machine → calculation  indicators → understanding
feedbacks  features  control

cyber  Algorithms modeling  physical

data → city

human
Remarks 2: IT and DT are different

- **IT (Information Technology)** Information technology, is a technology based on computer and internet to enhance people's information dissemination ability.

- **DT (Data Technology)** Data technology, the essence of data technology is the "processing" of data technology
  - The essence of data technology is also the technology of "cognition", the Technology of "thinking" and the technology of "decision making", which eventually forms the equipment of "artificial intelligence".
Remarks 3: data, and AI, is not Almighty

DARPA’s perspective on AI: three waves

That can be clustered into three waves of AI development that may be described as - *handcrafted knowledge, statistical learning and contextual adaptation*

1. **Handcrafted Knowledge**
   - Systems that have established sets of rules to represent knowledge in well-defined domains
   - **Features:** Enables reasoning over narrowly defined problems. No learning capability and poor handling of uncertainty
   - **Challenges:** The structure of the knowledge is defined by humans. The specifics are explored by the machine. Failure of the autonomous cars in the DARPA Grand Challenge

2. **Statistical Learning**
   - Systems based on statistical models developed to address specific challenges and trained using big data
   - **Examples:** Voice recognition, face recognition
   - **Features:** Advanced classification and prediction capabilities. No contextual capability and minimal reasoning ability
   - **Challenges:** Statistically impressive but individually unreliable. Inherent flaws can be exploited. Skewed training data creates maladaptation, “Blackbox”

3. **Contextual Adaptation**
   - Systems that construct *contextual explanatory models* for classes of real world phenomena
   - **Examples:** Image recognition
   - **Features:** Ability to perceive, learn, abstract and reason
   - **Models:** That generate explanations of how an object might have been created to explain and drive decisions

X Artificial intelligence
√ Augmented intelligence
Thank you!

Looking forward to future cooperation

清华同衡规划播报

同衡创新