北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

Application and challenge of big data in urban park ecosystem service recognition and evaluation

Zhongwei Zhu, Yinglu Miao, Zhifang Wang, Sirui Wang, Yaojin Zhou
College of Architecture and Landscape Architecture, Peking University
2019/07/02, Milan

• Introduction-Literature Review of ES evaluation

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

Method		Observation	Delphi method	In-depth interview	Questionnaire	Big Data method
		• Participant Observation • Unstructured Observation • Structured Observation	• Delphi interview	• Semi-Structured Interview • Unstructured Interview	• Structured Interview • Q Methodology • Internet Survey • Mail Survey	• Textual Analysis • Image Recognition • PPGIS
Sample	Individual	✓		✓	✓	✓
	Group	✓				✓
	Expert		✓			
Assess type	Qualitative	✓	✓	✓		✓
	Quantitative	✓	✓		✓	✓
Advantage		• Posed in natural environment • Flexible in experiment implementation	• Low Cost • Adapt to situation lack of data • Available of usage of academic words	• Learn more about people's ideas • Flexibility during interview	• Available of collecting quantitative data	• low cost • Correctable • Large temporal and spatial span
Limitation		• Hard to explain and analyze	• Result relies on experts chosen	• Small sample	• Limited in standardized question • Not flexible • High cost	• Lack of availability • Lack of reliability and validity
Reference		Tzoulas and James (2010) and Jerneck and Olsson (2013)	Edwards et al. (2012)； Tengberg et al. (2012)	Swapan， 2017； Flurina M, 2018； Sagie et al. (2013) and Smith； Sullivan (2014)	Buche, 2015； Derek （2014）	Daniel R. Richards, 2018； Yoshimura, 2017； Yoshimura, 2017

- **ES in urban parks** are vitally **import** (Palomo,2018) for evaluation of its performance. While **traditional ES evaluation** has some **disadvantages** and **big data** makes up some of former method's limitations.

- Most big data methods try to utilize **images**(Seppelt, 2011) to assess ES, but almost no research focused on textual data. However **textual data** contains plenty of information including **emotional**(Zhang et al., 2011)**and opinion expression**(Pak and Paroubek, 2010; Sandifer, Sutton & Ward, 2015) .

• Introduction-UGC as data source for ES study

- With the advent of the era of big data(Seresinhe, Preis et al., 2015), the explosive growth of **UGC (user generated content)** data provides a **large-scale data** source(Rozas et al., 2017) and new research ideas for landscape ecosystem evaluation(Macolna et al., 2013) .

- Through the text data generated by user comments, not only can researchers obtain the **emotional perception** (Zhang, Zhou et al. 2018) of the park landscape system, but also can **identify and evaluate** (Daniel R. Richards, 2018; Yoshimura, 2017; Yoshimura, 2017) the supply of ecosystem services in the park.

• Introduction-Site selection of Beijing

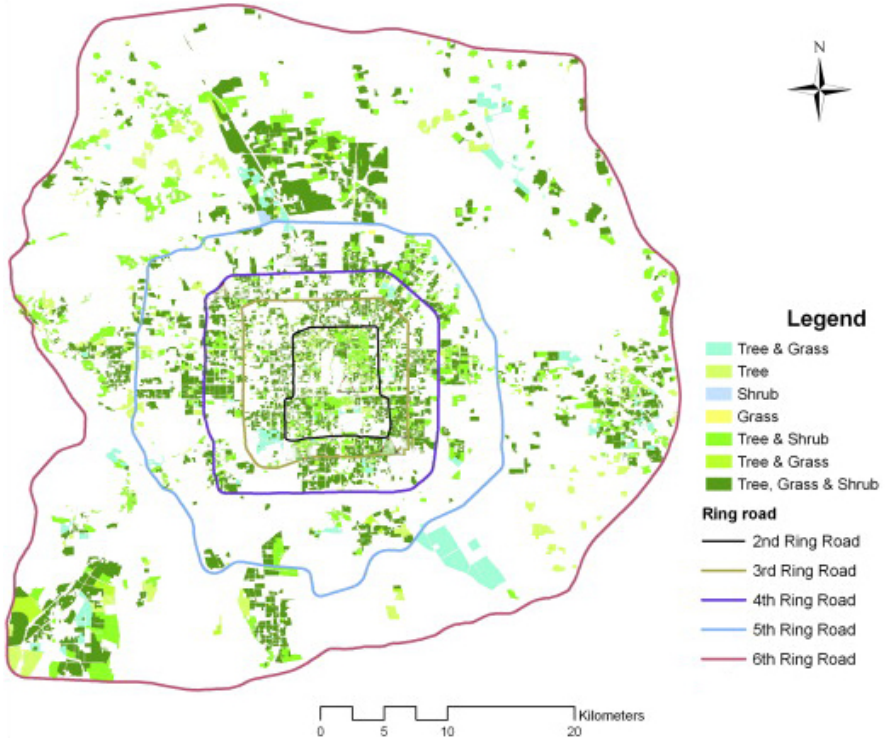
北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

- **Rapid development of park green space**

2001yr	2008yr	
70	398	Number
7139ha	12471ha	Area

- **Complete urban park system**

- A large number of cultural and historical gardens & natural parks
- Provide different ES
- A unique urban park system with the continuous development of the city.



Spatial distribution of urban green space within the 6th Ring Road of Beijing.

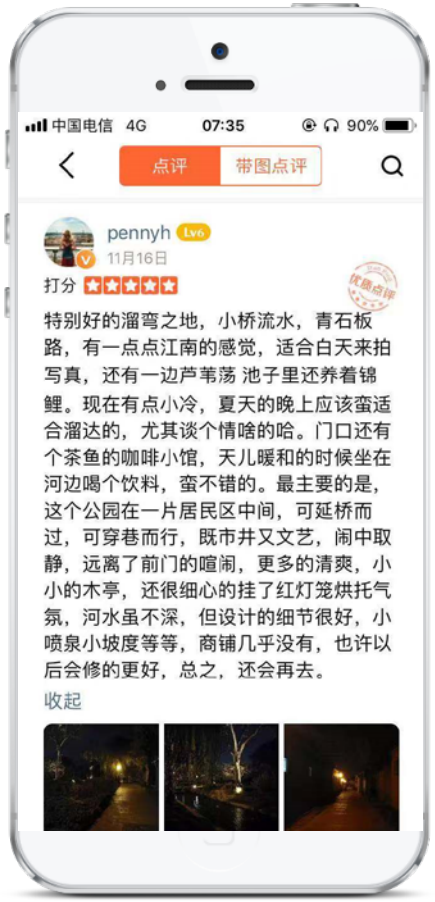
Data

- **Data Source: DA ZHONG DIAN PING Website** （http://www.dianping.com）
- **Reason:** Large User Sample & Targeted urban park textual data

- **Amount of parks: 176** parks in **Beijing**
- **Time range:** 2006.10-2018.9
- **Data size:** around **170,000** comments of parks
- **Data form:** text

Park	Time	Star	Comment
Beijing Olympic Forest Park	2018/10/13	5	奥林匹克森林公园很大，有好几个门，不过各个门交通都很方便，周末这里就是孩子们的天堂，深秋了，天气凉了，周日的孩子们也减少了，正好又可以一边锻炼，一边欣赏秋景了，只是跑起来老停下拍照，影响速度啊，不过能活动起来就是好猫，红叶银杏小菊花甚是漂亮啊

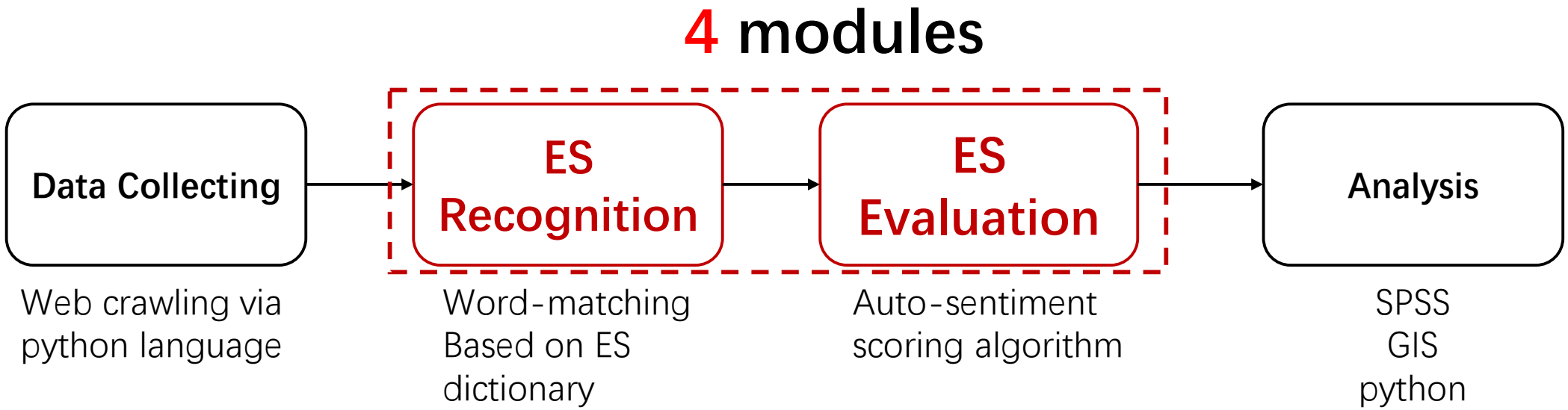
Data format



Method-Overview

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

- **Web crawling** via python.
- **ES recognition** with establishment of dictionary of ES.
- **ES evaluation** using Sentiment auto scoring based on machine learning.
- **Analysis** of ES performance of each park.



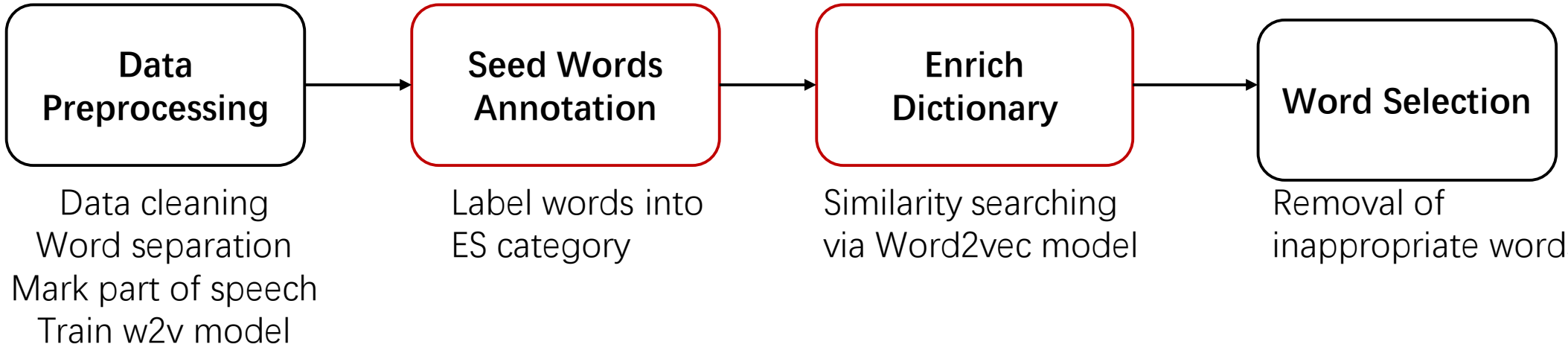
Methodology Framework proposed

• Method-ES Recognition

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

1. Establishment of dictionary of ES

- Annotate high frequency words as **seed words** and categorized them into **16 ES labels**.
- **Enrich dictionary** by similarity searching via Word2Vec model to add words in corpus similar to seed words.
- Manually remove inappropriate words from dictionary and **coding** the dictionary into format of json for further utilization and fixation.



ES dictionary establishment process

• Method-ES Recognition

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

1. Establishment of dictionary of ESS

```
00 : ['瓜果', '瓜', '蔬菜', '农作物', '果蔬', '柿子', '柿', '油柿', '山楂', '柿子树', '果实', '果子', '桑葚', '桃树', '核桃', '山楂树',
01 : ['水库', '蓄水池', '塘堰', '塘坝', '十三陵水库', '水源地', '污水', '水源', '坝河']
02 : ['树木园', '木材', '木料', '木头', '木', '原木', '石材', '金丝', '中药材', '中药', '国药', '中医药', '名贵药材', '药材', '药草', '
03 : ['水能', '动能', '电能', '磁能', '光能', '化学能', '原子能', '机械能', '太阳能', '结合能', '内能', '高能', '官能', '体能', '风能',
10 : ['空气清新', '新鲜空气', '吸氧', '空气污染', '环境污染', '空气流通', '空气调节', '污染空气', '空气清新', '香', '呼吸', '透气', '深
11 : ['湿润', '气候宜人', '小气候', '季风气候', '湿润', '降温作用', '吸收温室气体', '很冷', '中暑', '晒黑', '冻死', '暖暖的', '暖洋洋',
12 : ['水质', '碧水', '清水', '污水', '活水', '死水', '池水', '淡水', '地面水', '海水', '江水', '井水', '水太绿', '嫌水脏', '汲水', '汲
13 : ['灾害', '灾患', '灾', '灾荒', '地震带', '地震台', '地震棚', '地震', '地动', '震', '震害', '劫难', '万劫不复', '灾祸', '灾难', '赈
20 : ['生机', '生机勃勃', '栖息地供给', '蓬勃', '旺盛']
30 : ['摄影', '留影', '拍摄', '拍照', '照相', '摄像', '摄录', '录像', '拍花', '赏景', '看景', '陶冶情操', '赏荷', '观景', '赏花', '踏青
31 : ['教育', '教导', '教化', '感化', '教诲', '训诲', '训迪', '启蒙', '春风化雨', '耳提面命', '施教', '有教无类', '科普', '熏陶', '学习
32 : ['散步', '溜达', '遛', '逛', '绕弯儿', '遛弯儿', '转悠', '转转', '走走', '遛弯', '闲逛', '跑步', '散散步', '晨练', '饭后', '锻炼身
33 : ['散散心', '散散步', '散心', '体验', '经验', '体会', '心得', '感受', '散心', '消遣', '清闲', '排解', '解闷', '散闷', '消闲', '闲逛
34 : ['文化', '知识', '学问', '学识', '人文', '艺术', '内涵', '文化底蕴', '底蕴', '博大精深', '深厚', '宗教', '佛教文化', '风俗', '唱戏
35 : ['祭祀', '祭', '祀', '祝福', '祭天', '祭拜', '腊', '祭地', '祭祖', '祈祷', '帝王', '牌位', '五谷丰登', '五谷', '祭天', '祭祀', '祭
36 : ['相亲', '聊聊天', '打打牌', '聊天', '看看书', '坐坐', '聊聊', '野餐', '吃吃喝喝', '散散步', '席地而坐', '小聚', '谈恋爱', '恋爱',
```

Part of the ES dictionary

ES Main Category	ES Sub Category	Word Amount
0 Provisioning Service	00 Food provisioning	119
	01 Water provisioning	10
	02 Material provisioning	21
	03 Energy provisioning	26
1 Regulating Service	10 Air quality Regulating	37
	11 Climate Regulating	203
	12 Water Resource Regulating	36
	13 Disaster Regulating	21
2 Supporting Service	20 Bio-diversity	5
3 Cultural Service	30 Aesthetic	581
	31 Education	39
	32 Recreation	917
	33 Spirit	3059
	34 Historic	201
	35 Religion	102
	36 Socio	194

ES dictionary basic information

• Method-ES Recognition

2. Word Recognition

- Match word in comment with dictionary to **recognize ES**.
- If the word appears in ES dictionary, **mark the service** mentioned as **1**, otherwise **0**.

comment	Key word	ESS remark
Olympic Forest Park in autumn is so beautiful! I can run in the park and enjoy the wonderful scene at the same time.	beautiful run scene	Aesthetic-1 Recreation-1

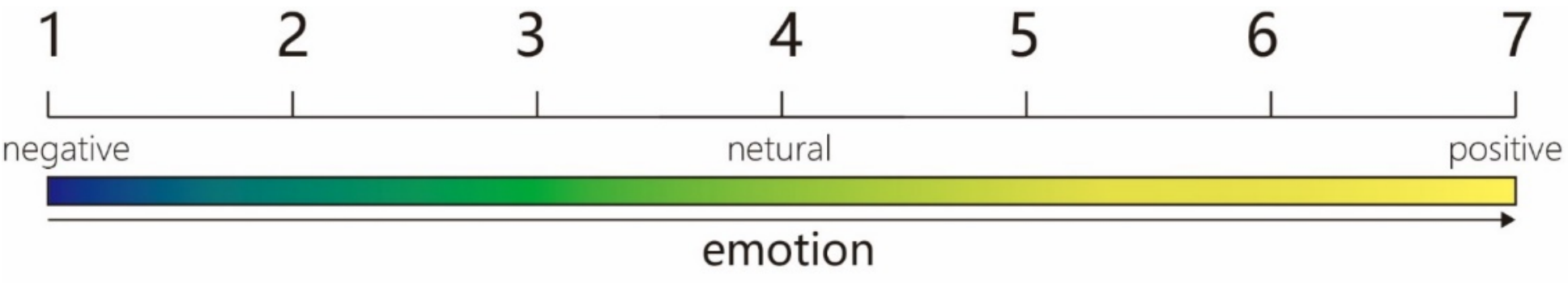
Example result of word match

Method-ES Evaluation

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

Sentiment auto scoring algorithm based on machine learning

- Manual labeling 3500 comment data into score (1-7) to represent the positivity of visitor's emotion on park.
- Vectorizing sentence with Word2Vec model.
- Using XGBregressor model to fit data vectorized and test model's performance.
- Auto scoring emotion of both ES and visitor's overall satisfaction of park .

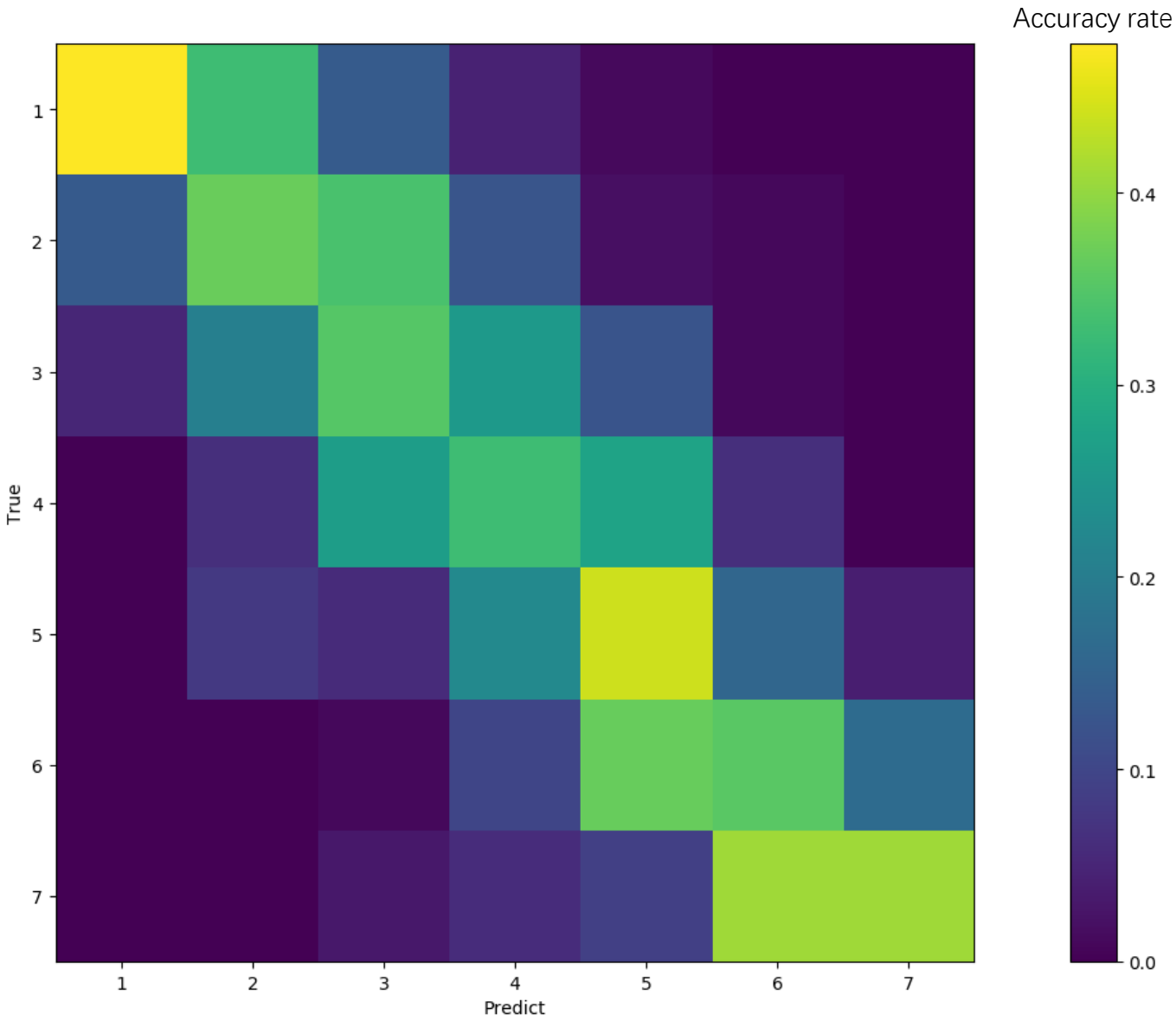


Emotion positivity measuring in 7 classes

• Method-ES Evaluation

Sentiment auto scoring algorithm based on machine learning

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture



confusion matrix in one of the tests

- Adjust the parameters of machine learning model to find the best choice.
- Separate data into train data and test data.
- Using the model trained to predict the test data and tests its performance.
- Comparing with other model/method 's performance.

model/method	SnowNLP	ROST CM6	Our method
R^2（5-fold cross validation average）	0.365	0.455	0.620

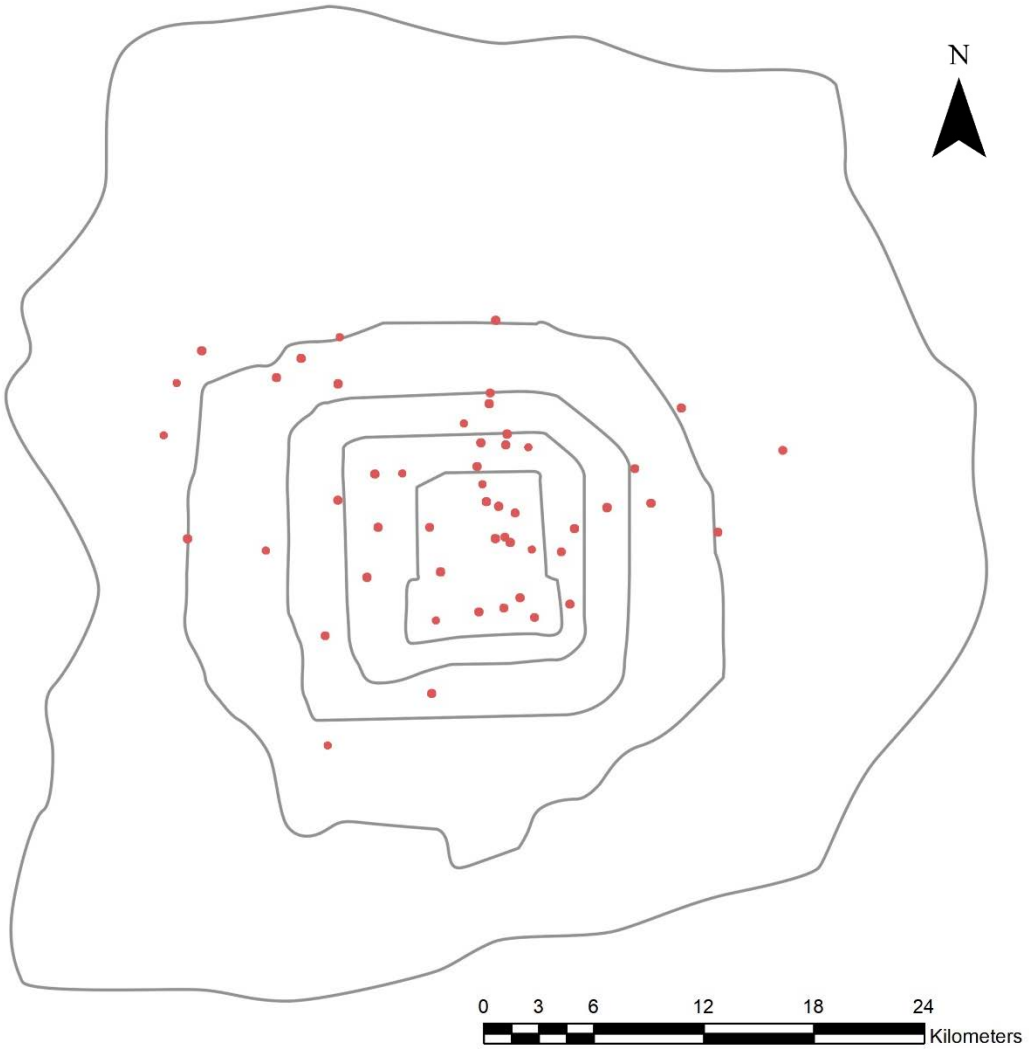
Label	1	2	3	4	5	6	7
True predict	1,2	1,2,3	2,3,4	3,4,5	4,5,6	5,6,7	6,7
True orientation predict	1,2,3	1,2,3	1,2,3,4	3,4,5	4,5,6,7	5,6,7	5,6,7
Predict result accuracy	R^2 =0.655（R=0.812, spearman）,True predict accuracy=0.83, True orientation predict=0.872						

Our model outperforms baselines !

• Case study

- Beijing urban parks (within 6th ring road & comment amount>=50)
- **Park Amount: 50**
- **Data Size: 113,994** comments

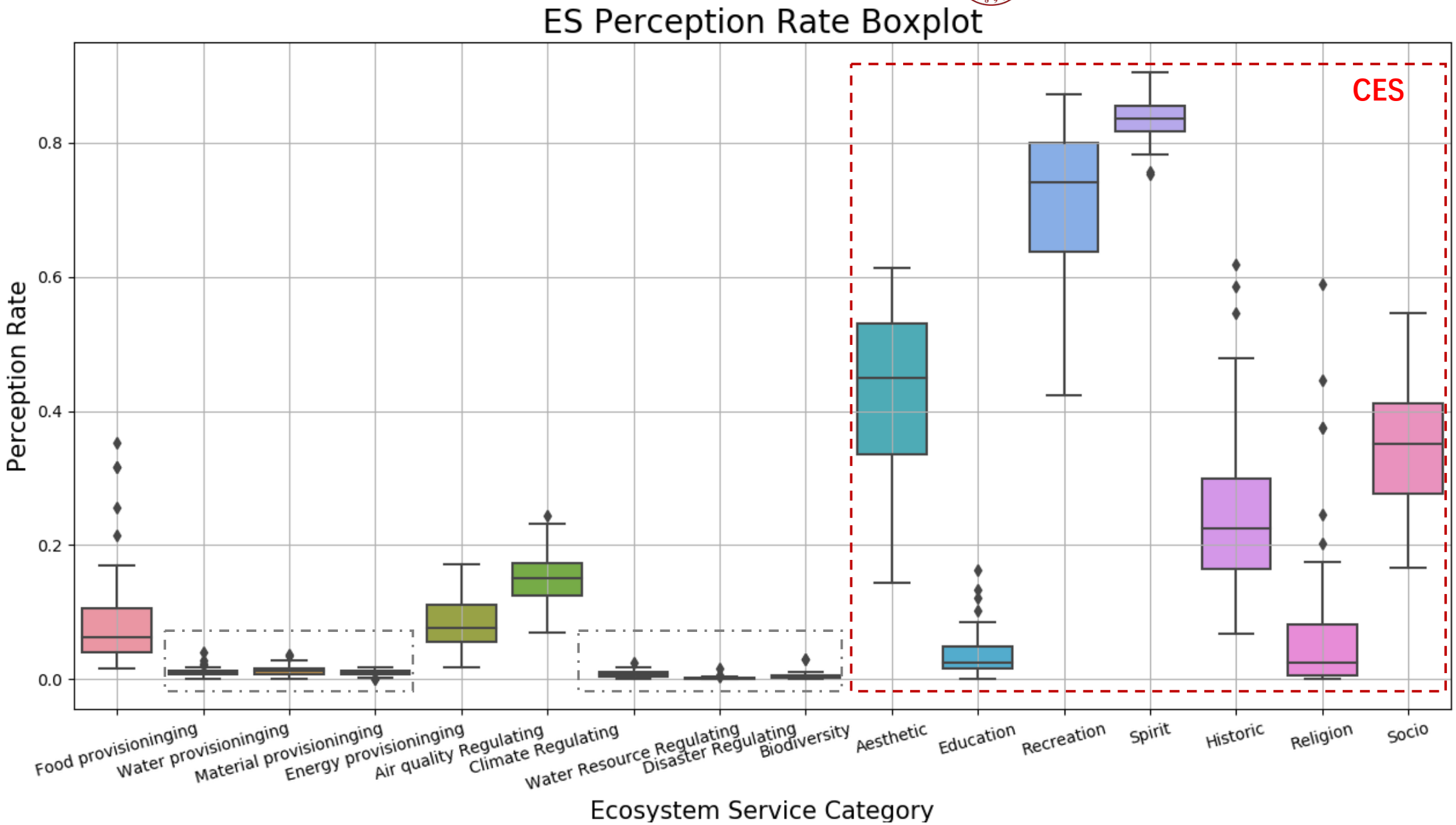
- Detect ES perception in each comment using our method.
- Evaluate the comment' s emotion positivity and ES performance in each comment.
- Analyze differences of 50 parks both in ES perception and ES performance.



Spatial distribution of urban parks in case study

- **Result-ES Recognition**

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture



Result-ES Recognition

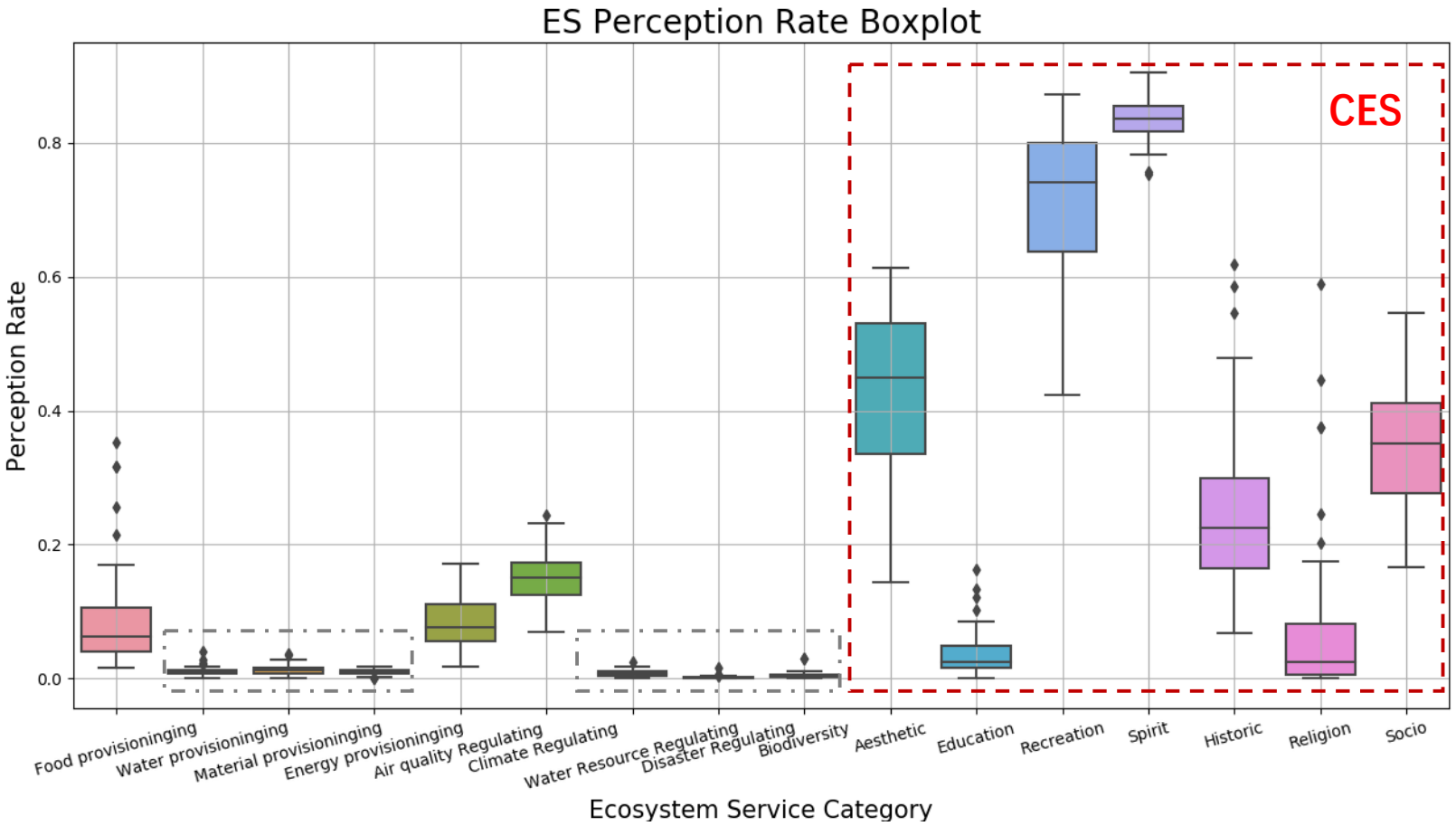
北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

- **Perception portion:** Cultural Service （89% in total) > Regulating Service （8% in total) > Provision Service （3% in total) > Supporting Service(close to 0 in total)

- Social media data in urban parks tends to mention **cultural ecosystem service** more but other three types of service less.

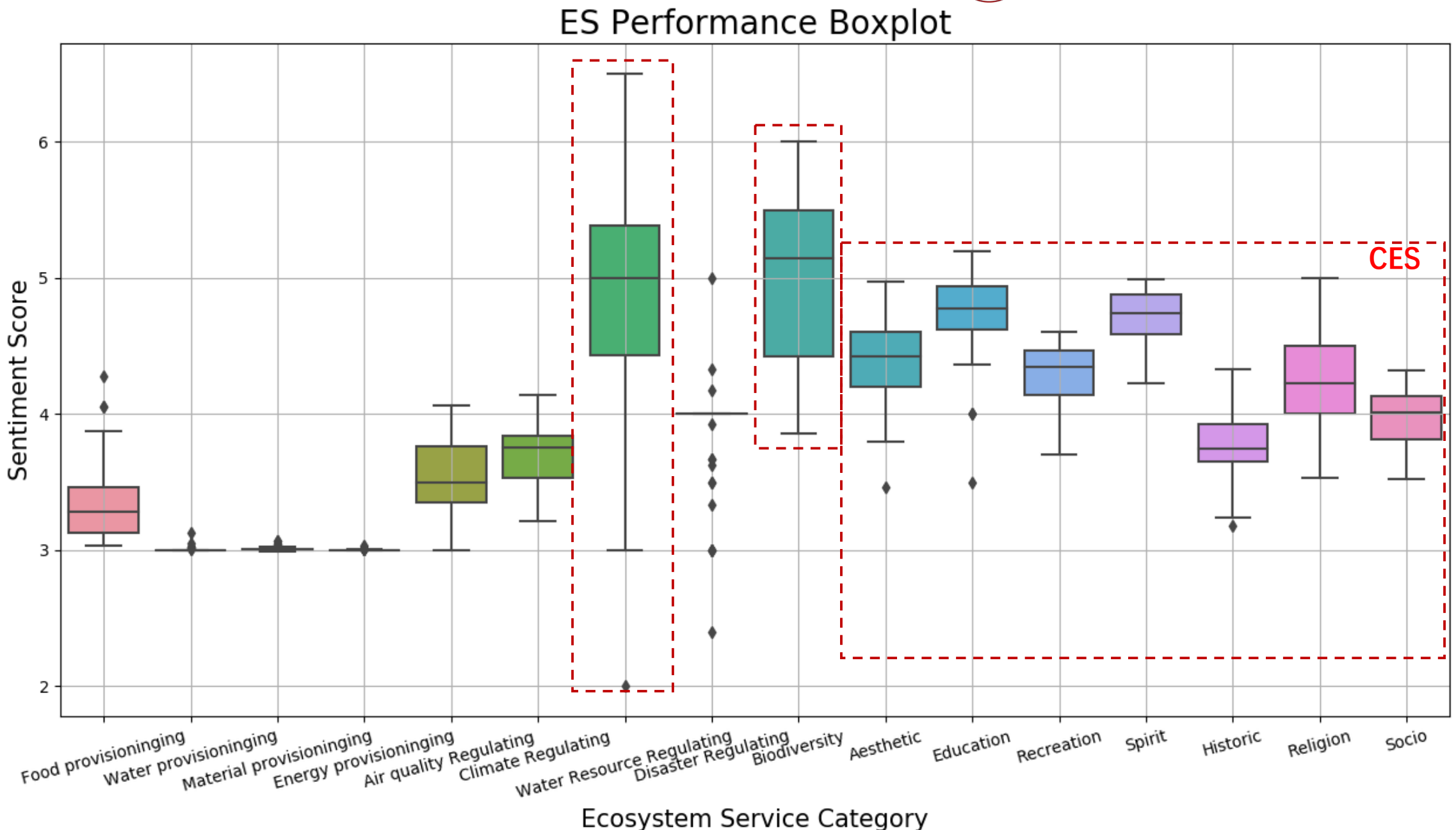
- **ES perception rate** of cultural service in 50 parks is relatively **high** compared to other 3 type of ES.

- There's no **significant difference** between 50 parks in Beijing of perception in provision, regulating, bio-diversity, but **vary much** in **cultural ecosystem service**.



- **Result-ES Evaluation**

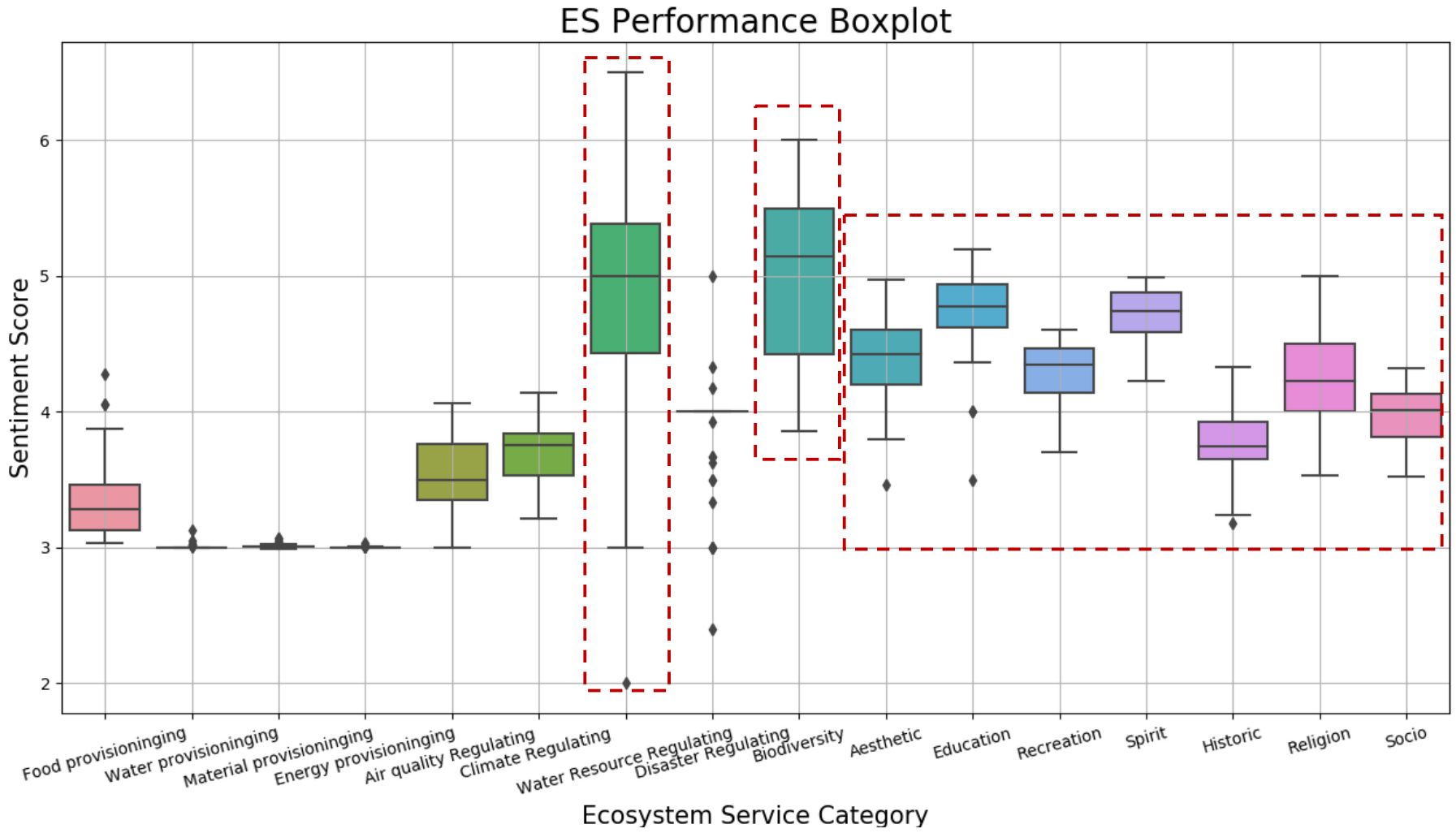
北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture



• Result-ES Evaluation

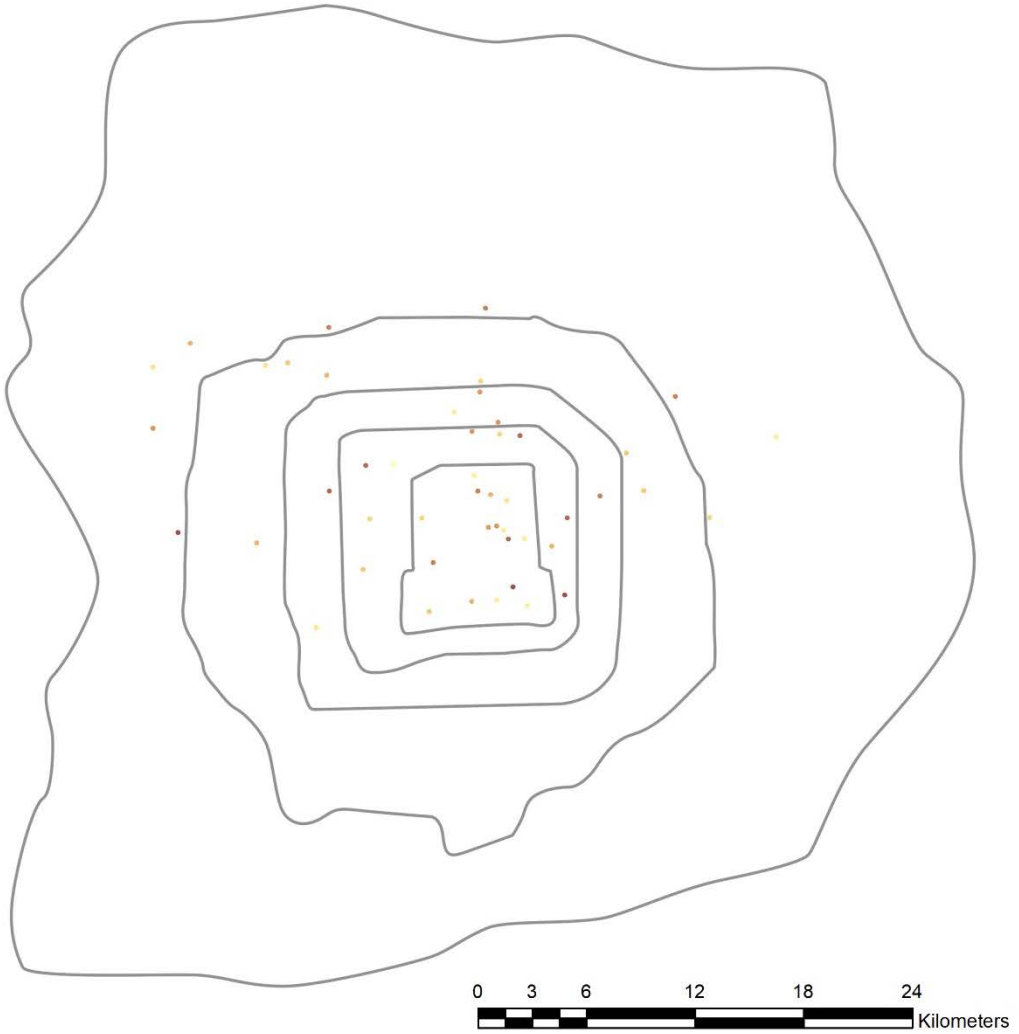
北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

- **Biodiversity** Service performs relatively better than other types of services; **Climate Regulating Service**'s average score ranks 2nd.
- **Cultural services** perform well and each parks' performance differ relatively little.

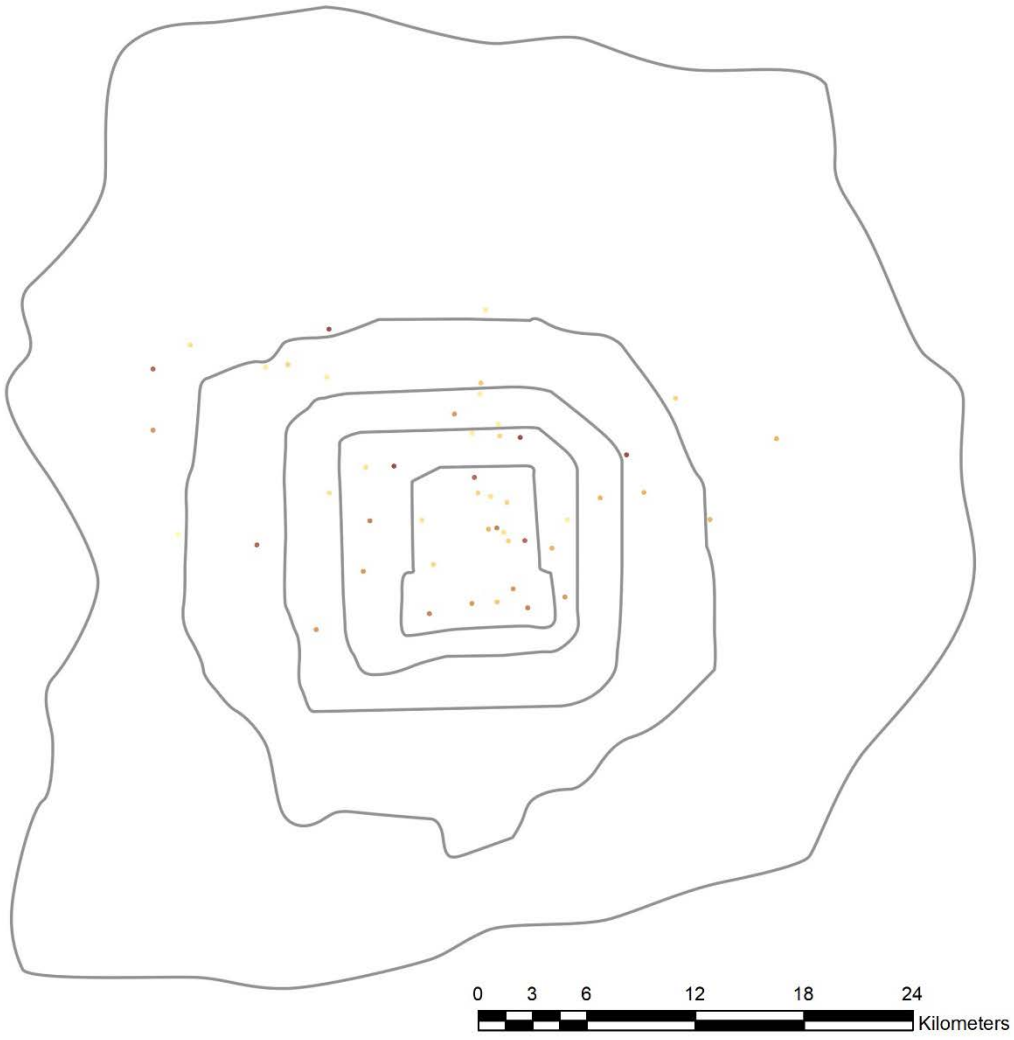


- **Result-mapping of CES**

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture



CES service perception mapping



CES service performance evaluation mapping

• Result-relation of satisfaction and ES performance

北京大学建筑与景观设计学院
College of Architecture and Landscape Architecture

- Set each park' s mean sentiment score as y, put 16 categories' ES sentiment score as x, using SPSS to do linear regression.

- The regression model' s R^2 reaches **0.772**(adjusted R^2 =0.751)

- The regression model indicates that visitors' sentiment feedback is strongly influenced mainly by CES, especially spirit service and aesthetic service.

$$Y=0.755SP+0.302AS-0.2ED-0.211FS+0.345$$

Y-sentiment score
SP-sprit score
AS-aesthetic score
ED-education score
FS-food provision

Conclusion & Discussion

- A new way to rapidly **evaluate ecosystem service of urban parks**.
- Perform well in CES recognition but proves to be less useful in other ES.
- ESS dictionary is coded in Chinese and cannot be applied to other language. But this method could be **borrowed to evaluate textual data in other language**.
- The ES recognition mainly relies on word-matching, but considering the complexity of word expression, some ES expression may be misrecognized while others may not be recognized. **Machine learning method** could be considered to apply in further study especially to improve the ESS recognition module.
- The **bias** in social media data should be examined and adjusted through **traditional survey approaches**.

References

- Wang Z, Jin Y, Liu Y, et al. Comparing social media data and survey data in assessing the attractiveness of Beijing Olympic Forest Park[J]. Sustainability, 2018, 10(2): 382.
- Zoé A. Hamstead, David Fisher, et al.Geolocated social media as a rapid indicator of park visitation and equitable park access. Computers, Environment and Urban Systems. Volume 72,2018,Pages 38-50
- Oteros-Rozas E, Martín-López B, Fagerholm N, et al. Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites[J]. Ecological Indicators, 2017:S1470160X17300572.
- Xu L , You H , Li D , et al. Urban green spaces, their spatial pattern, and ecosystem service value: The case of Beijing[J]. Habitat International, 2016, 56:84-95.
- Daniel R. Richards, Daniel A. Friess, A rapid indicator of cultural ecosystem service usage at a fine spatial scale: Content analysis of social media photographs, Ecological Indicators, Volume 53, 2015,Pages 187-195
- Yoshimura, N., & Hiura, T. (2017). Demand and supply of cultural ecosystem services: Use of geotagged photos to map the aesthetic value of landscapes in Hokkaido. Ecosystem Services, 24, 68-78. doi:10.1016/j.ecoser.2017.02.009
- Oteros-Rozas, E., Martín-López, B., Fagerholm, N., Bieling, C., & Plieninger, T. (2017). Using social media photos to explore the relation between cultural ecosystem services and landscape features across five European sites. Ecological Indicators. doi:10.1016/j.ecolind.2017.02.009
- Ghermandi, A. (2016). Analysis of intensity and spatial patterns of public use in natural treatment systems using geotagged photos from social media. Water Res, 105, 297-304. doi:10.1016/j.watres.2016.09.009
- V. Heikinheimo, E. Di Minin, H. Tenkanen, A. Hausmann, J. Erkkonen, T. Toivonen. User-generated geographic information for visitor monitoring in a National Park: A comparison of social media data and visitor survey. ISPRS International Journal of Geo-Information, 6 (3) (2017), p. 85
- Natural versus urban sciences: some psycho-physiological effects. Environ. Behav., 13 (1981), pp. 523-556
- Zhang F, Wu L, Zhu D, et al. Social sensing from street-level imagery: A case study in learning spatio-temporal urban mobility patterns[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 153: 48-58.

Thanks for listening!