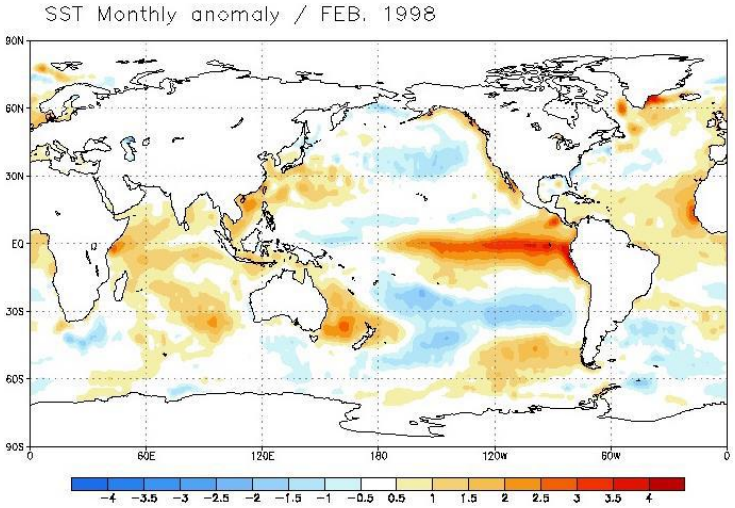
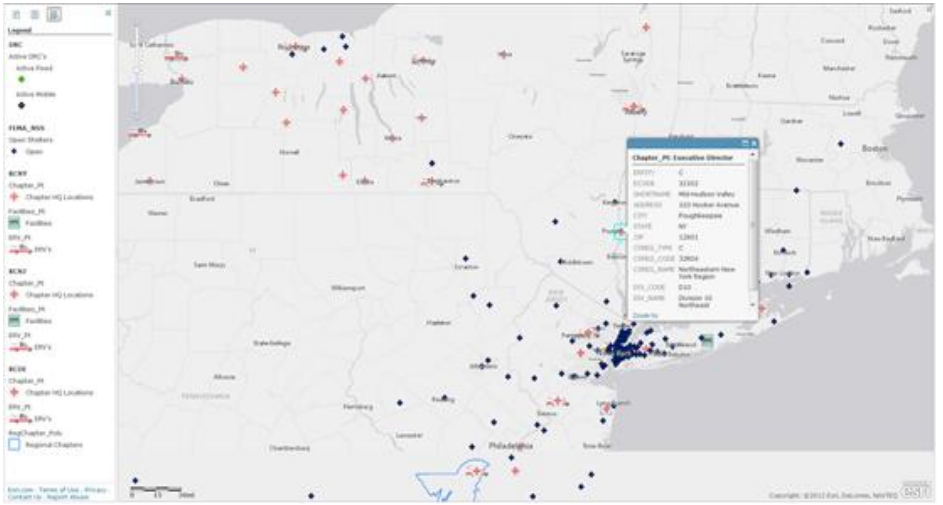
High Performance Computing of Points Generalization and its Potential Application in Risk and Disaster Management

Jie Shen, Jiemin Zhou
School of Geography Science, Nanjing Normal University
shenjie@njnu.edu.cn
July 2, 2017

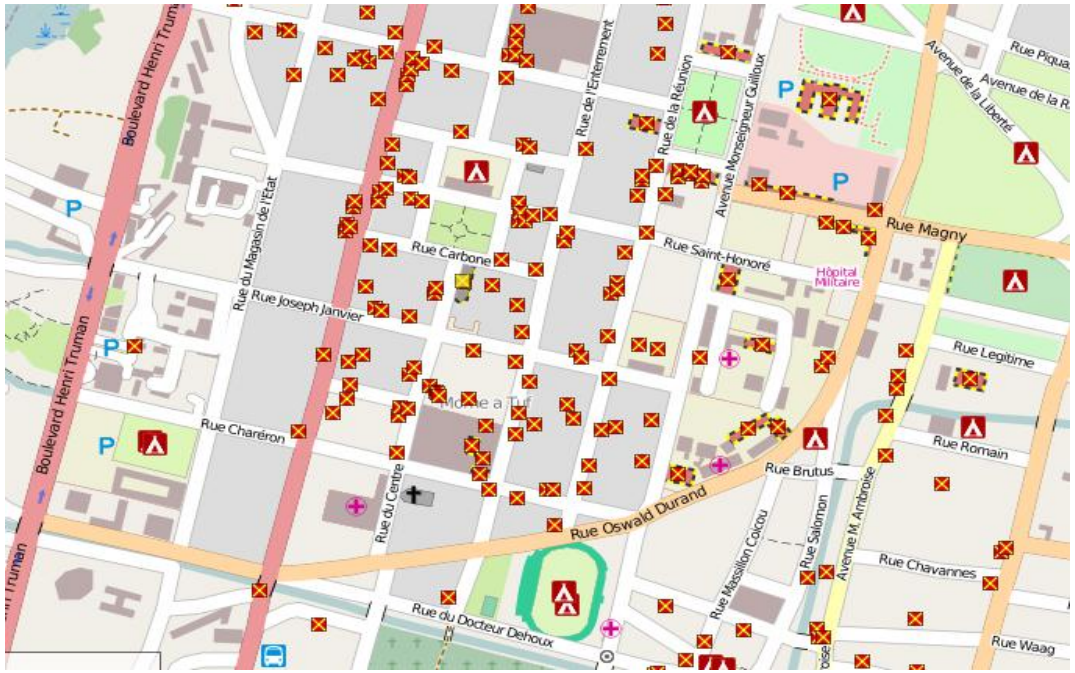
Outline


SST Monthly anomaly / FEB. 1998
90N
60N
30N
EQ
30S
60S
90S
0
60E
120E
180
120W
60W
0
−4 −3.5 −3 −2.5 −2 −1.5 −1 −0.5 0.5 1 1.5 2 2.5 3 3.5 4

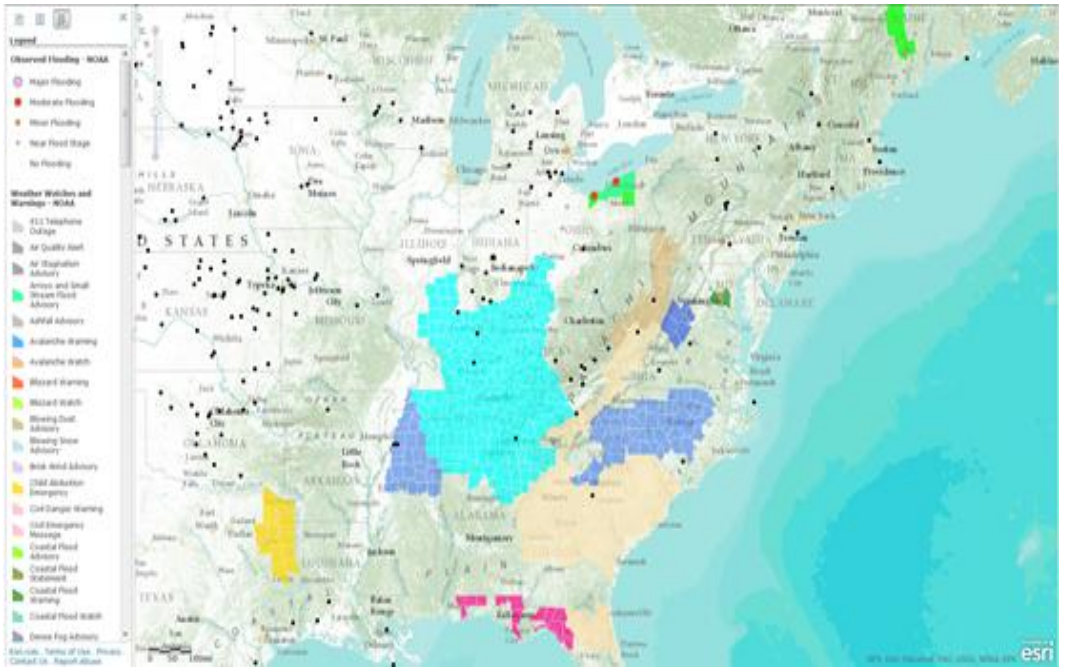
The picture shows emergency succor materials an area needed



rescue shelters distribution map

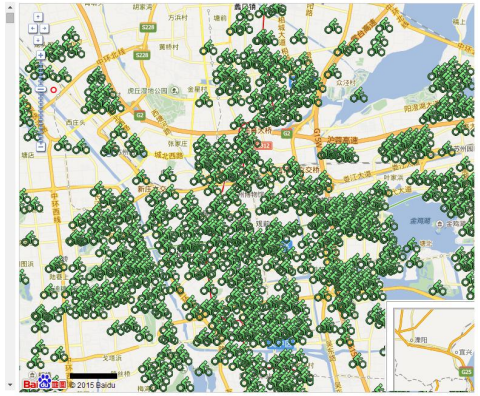


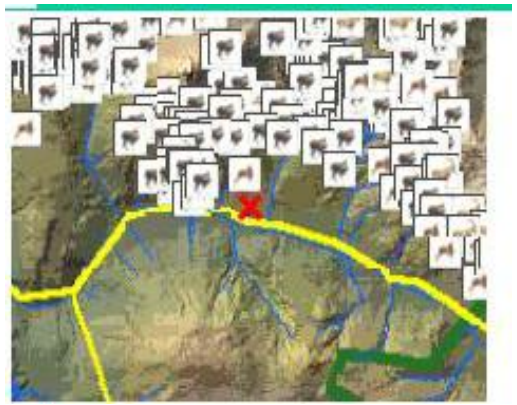
Haiti earthquake

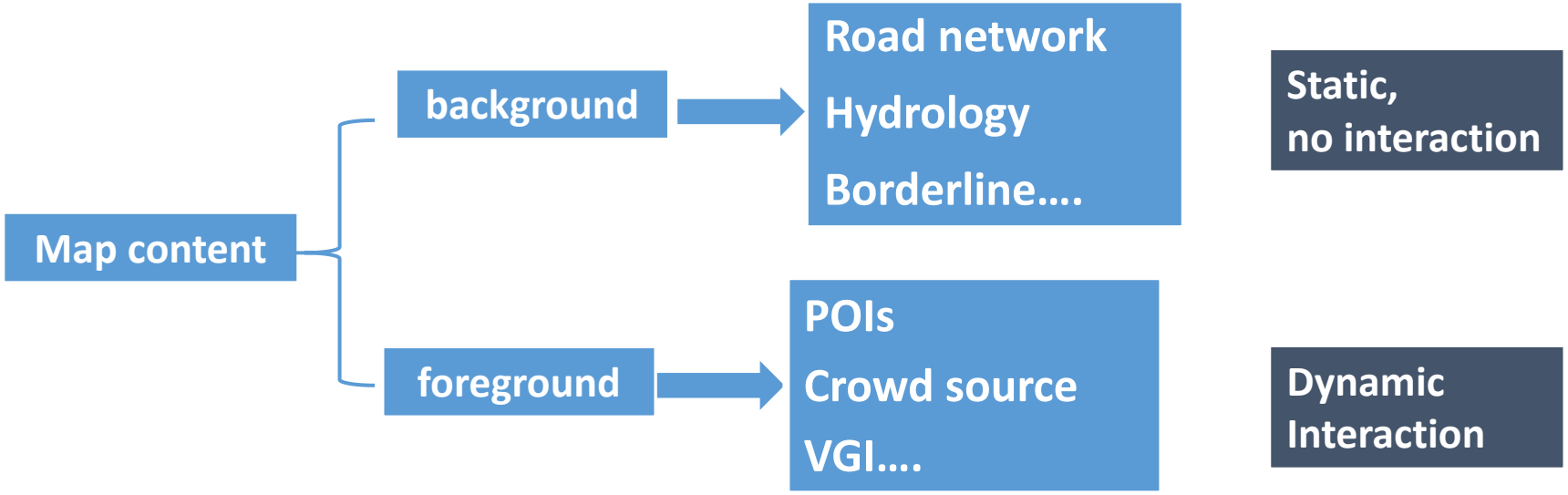


flood disaster map

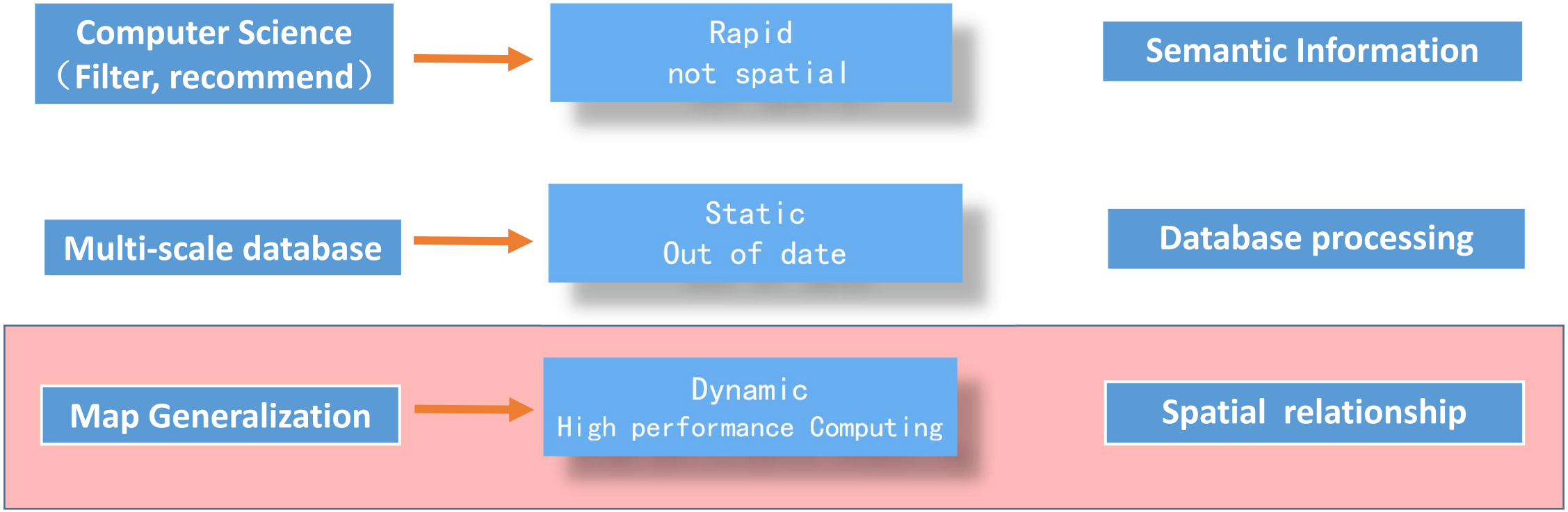
1. Point feature is the most **active** component among all spatial information.
2. The promotion of sensors makes the amount of point data to be huge. This has brought great challenges to the current sequential computing environment.



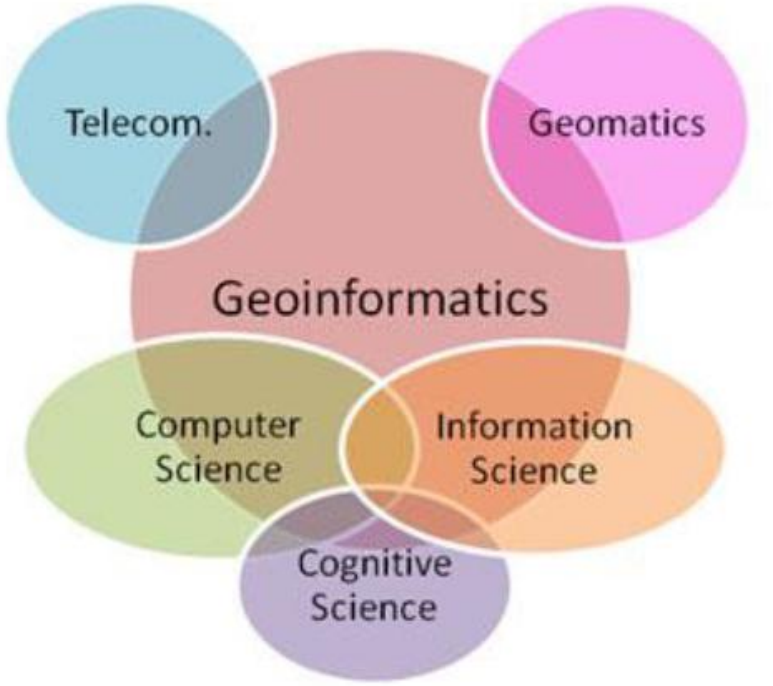




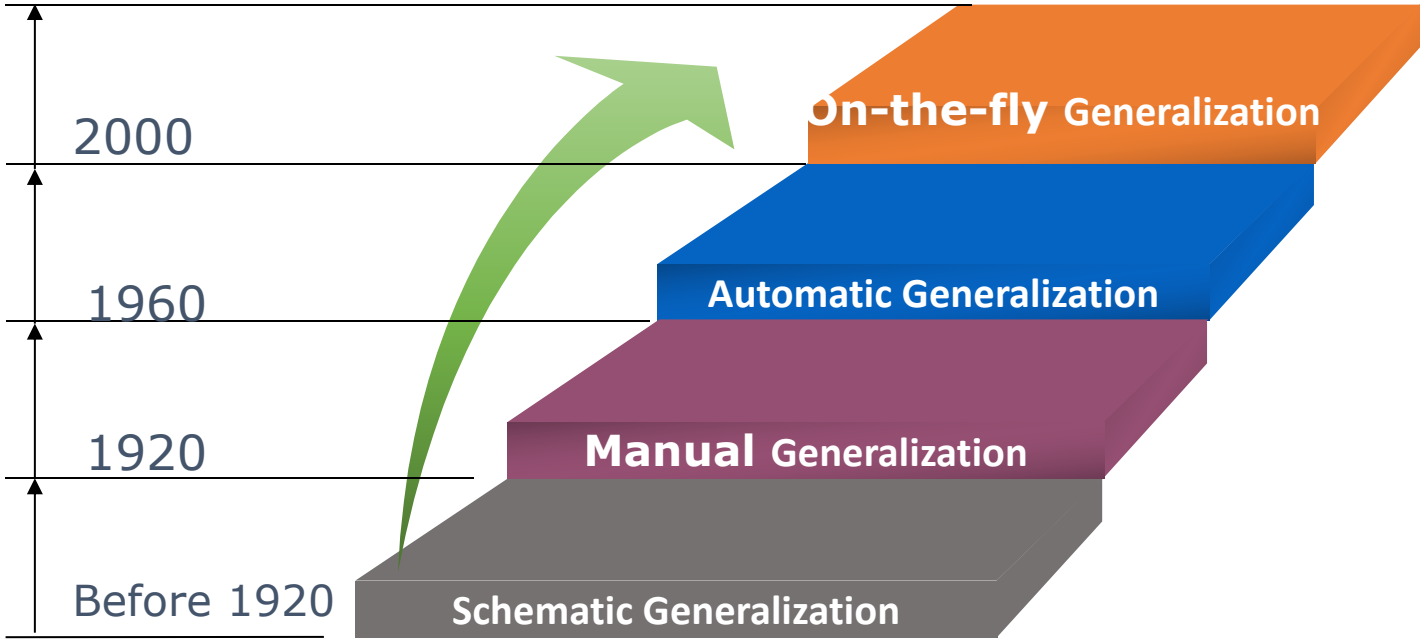
Methods for solving points clustering, overlaying

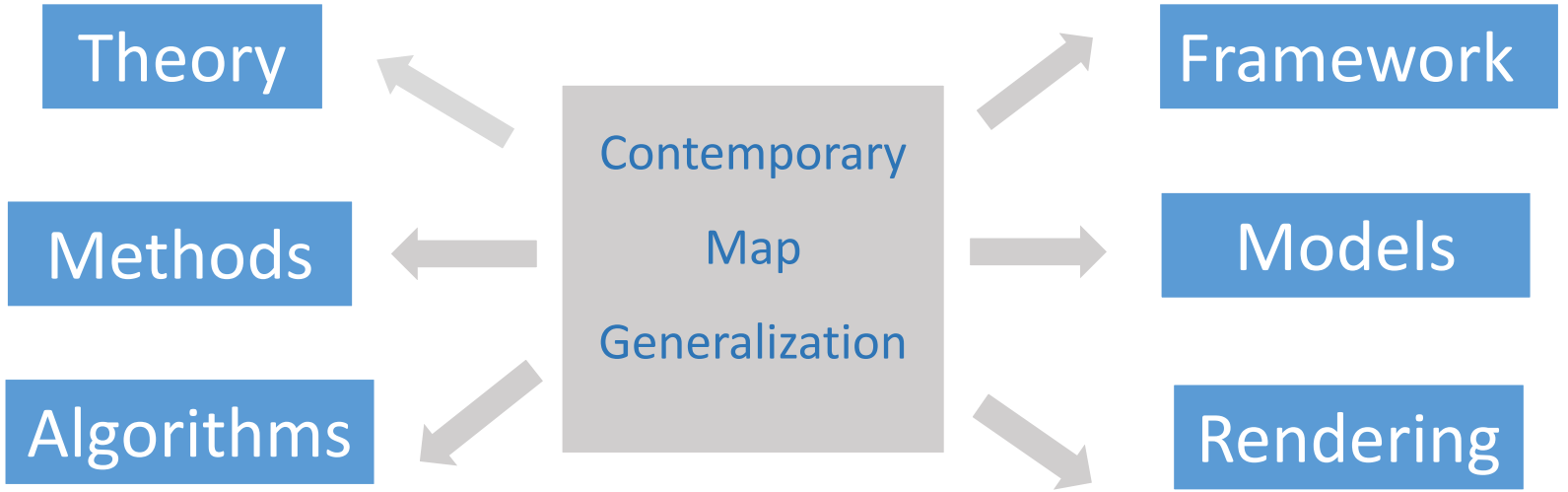


The development of map generalization



Relationships between Geoinformatics and other disciplines

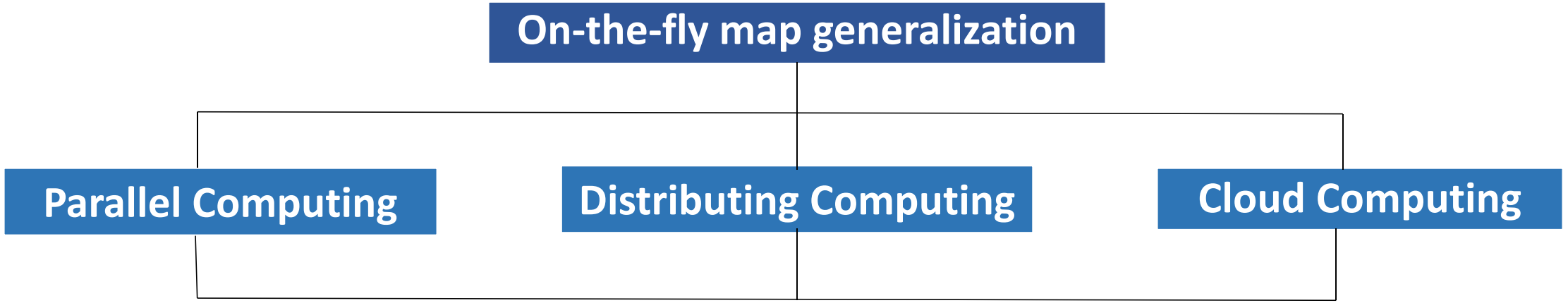




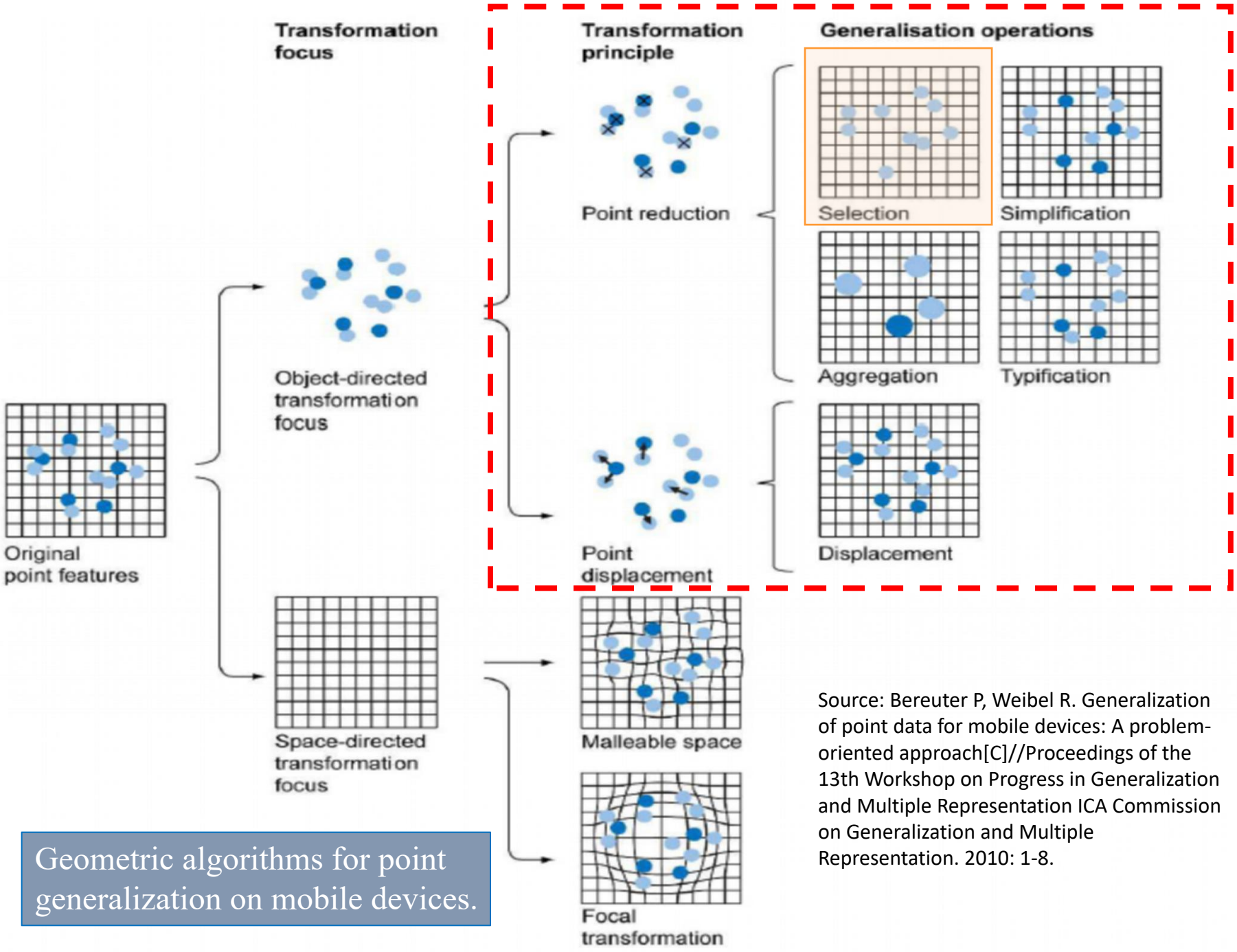
Usability: measured by the efficiency, effectiveness and satisfaction in the use of the given set of data by particular users for a specified application (purpose).

Efficiency: High Performance computing of map generalization

Effectiveness and satisfaction: Dynamic Model, User Model for map generalization



- The evaluation of decomposability of map generalization oriented parallel computing.
- Data and task decomposition methods for the parallel computation of map generalization.
- Realization of parallel computing of map generalization in different parallel computing environment(Open MP, MPI, Pthread).
- Realization of Cloud computing of map generalization



Geometric algorithms for point generalization on mobile devices.

Source: Bereuter P, Weibel R. Generalization of point data for mobile devices: A problem-oriented approach[C]//Proceedings of the 13th Workshop on Progress in Generalization and Multiple Representation ICA Commission on Generalization and Multiple Representation. 2010: 1-8.

Usability analysis of point generalization in mobile environment

No.	Operator	Algorithms	Time complexity
1	Selection	Selection based on the attribute information	$O(n)$
2		Selection based on the relation	$O(n)$
3	Simplification	settlement-spacing ratio algorithm	$O(n^2)$, Voronoi$O(n\log n)$
4		distribution-coefficient algorithm	$O(n^2)$
5		gravity-modeling algorithm	$O(n^2)$， Delaunay$O(n\log n)$
6		Circle growth algorithm	$O(n^2\log n)$；$O(n\log n)$， $O(n^2)$
7		ε逼近算法	$O(n^2\sqrt{n}\log n)$

Langran G E, Poiker T K. Integration of Name Selection and Name Placement: Proceedings of 2nd International Symposium on Spatial Data Handling, Seattle, Washington, USA, 1986[C].

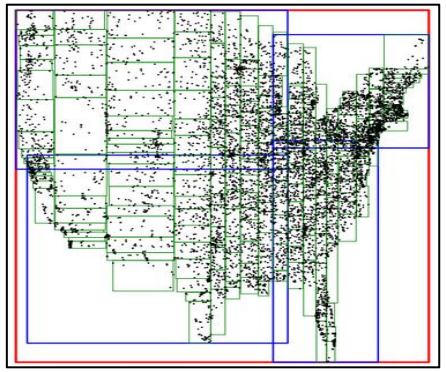
Kreveld V, et al. Efficient settlement selection for interactive display. Processing of Auto Carto 1995, 12, Bethesda, Md: 287-296

De Berg M, Bose P, Cheong O, et al. On Simplifying Dot Maps[J]. Computational Geometry-Theory and Applications, 2004, 27(1):43-62.

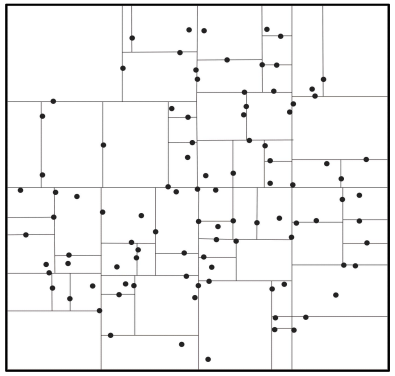
Outline

Traditional point data decomposition method

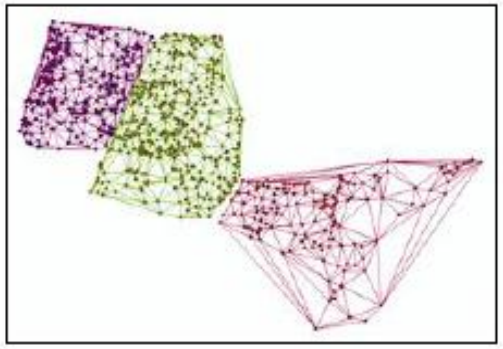
R*Tree



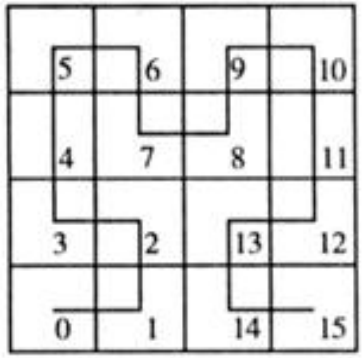
N-KD Tree



K-Means++



Hilbert Code

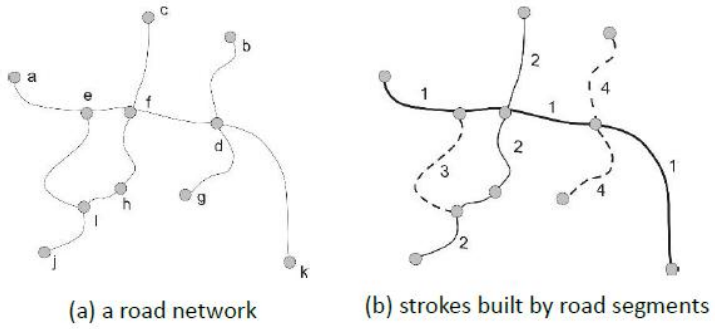


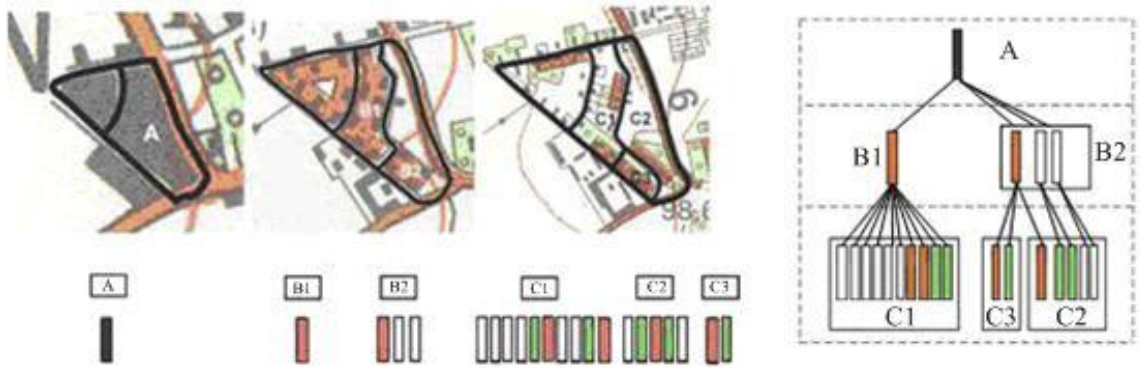
Antonin Guttman,1984； Kap S. Bang， 1996； Fu X D， 1997； Zhu Qing, 2007

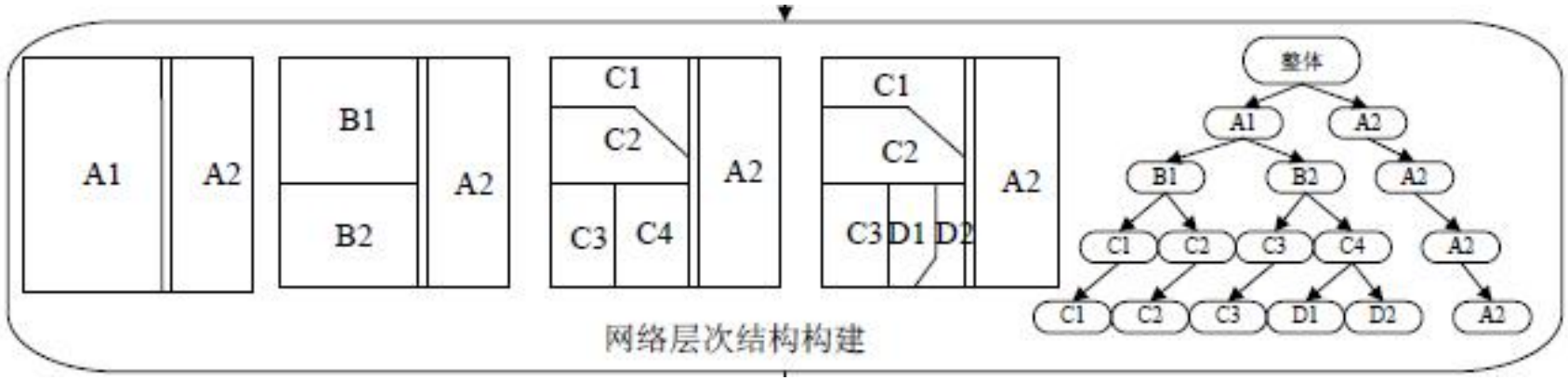
Methods	Type	Efficiency	Characteristic
R*Tree	Spatial Index	middle	It is difficult to take into account spatial clustering and load balancing.
N-KD Tree		high	
K-Means++	Spatial Cluster	high	Fast but the data is inclined, it is difficult for load balancing.
Hilbert Code	Mapping function	low	It can keep the spatial proximity but the decomposition cost a long time.

Data Decomposition Method based on Hierarchical Stroke Mesh (DMHSM)

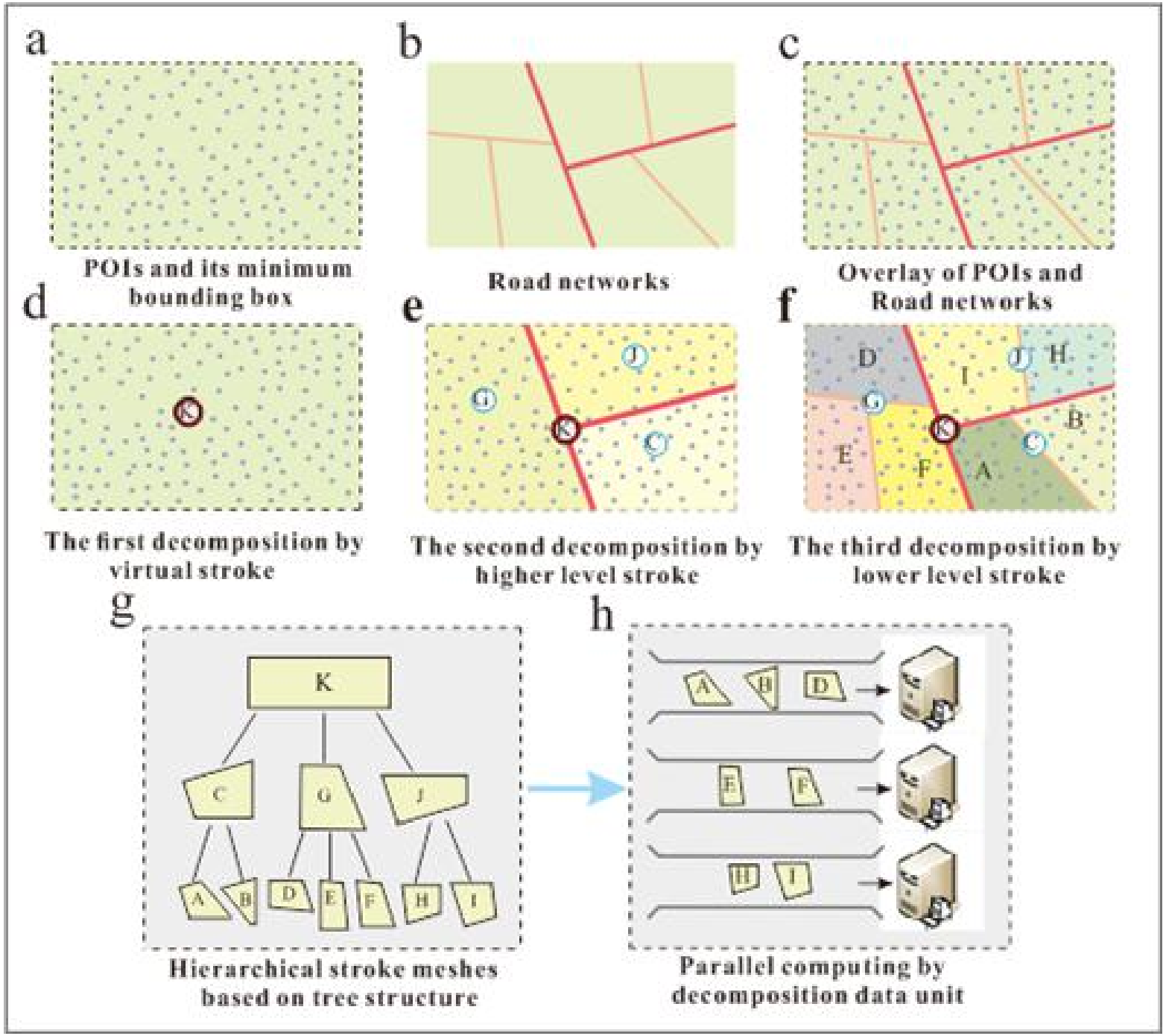
Hierarchical Stroke Mesh

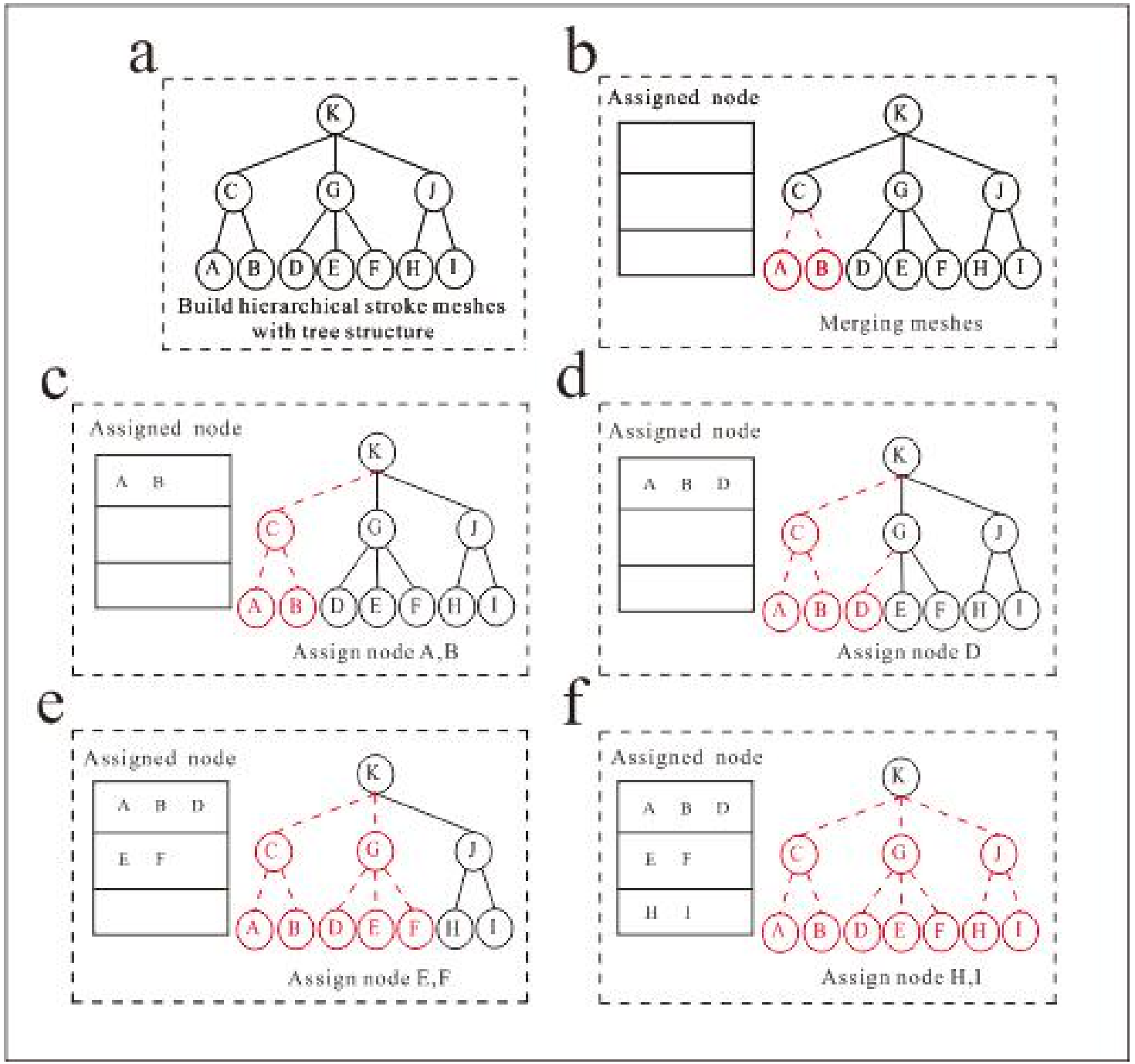






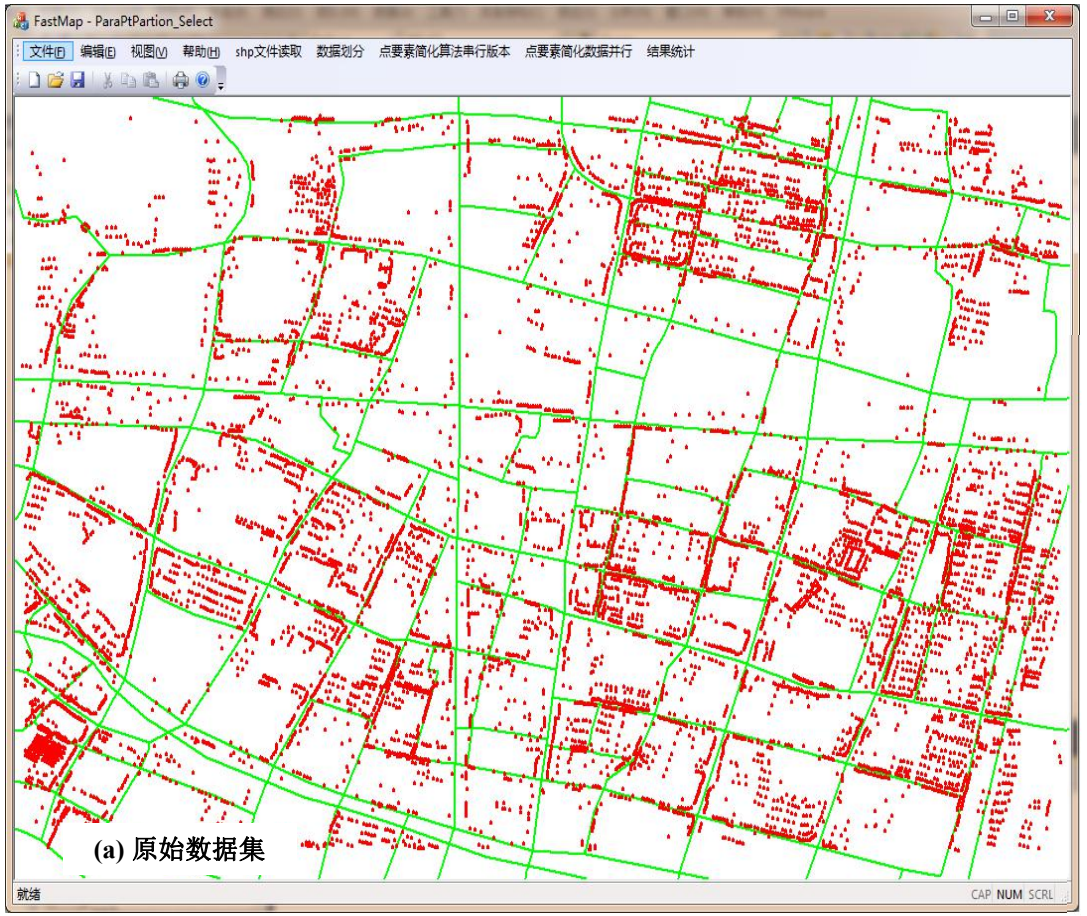
Thomson and Richardson , 1999; Zhou and L i, 2012 ,B Jiang, X Liu.2004,2007, Z Li, H Yan.2004, T Ai., OZ Chaudhry,WA Mackaness,2008


a
POIs and its minimum bounding box
b
Road networks
c
Overlay of POIs and Road networks
d
K
The first decomposition by virtual stroke
e
J
G
K
C
The second decomposition by higher level stroke
f
D
I
J
H
G
K
B
C
E
F
A
The third decomposition by lower level stroke
g
K
C
G
J
A
B
D
E
F
H
I
Hierarchical stroke meshes based on tree structure
h
A
B
D
E
F
H
I
Parallel computing by decomposition data unit

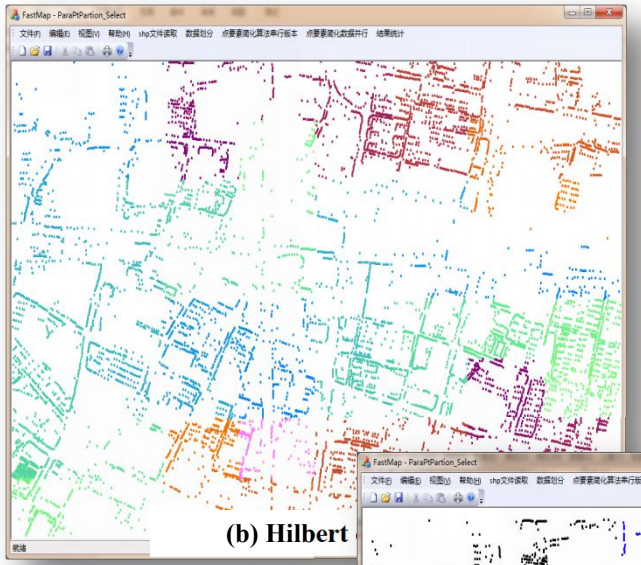

a
K
C G J
A B D E F H I
Build hierarchical stroke meshes with tree structure
b
Assigned node
Merging meshes
c
Assigned node
A B
Assign node A,B
d
Assigned node
A B D
Assign node D
e
Assigned node
A B D
E F
Assign node E,F
f
Assigned node
A B D
E F
H I
Assign node H,I

1. Contrast for the point data decomposition

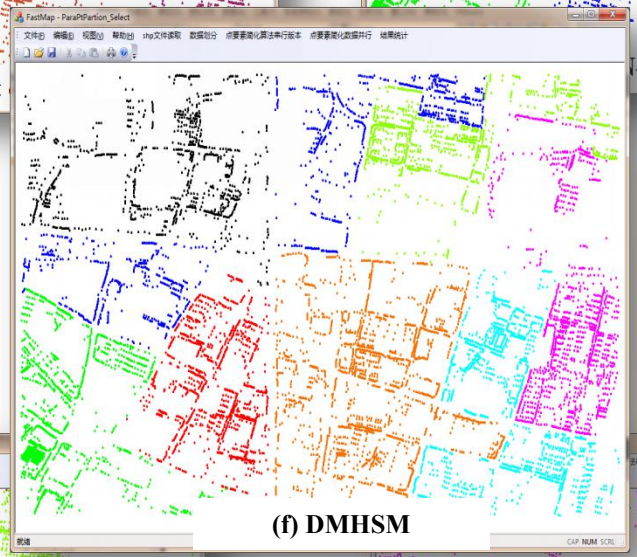
- Microsoft Visual Studio .NET 2010 + .NET Framework 4.0
- CPU: Intel(R) Core(TM) i7-3770 CPU @ 3.4GHz
- Study area: Nanjing Xinjiekou Area
- POI :12006 Stroke:69

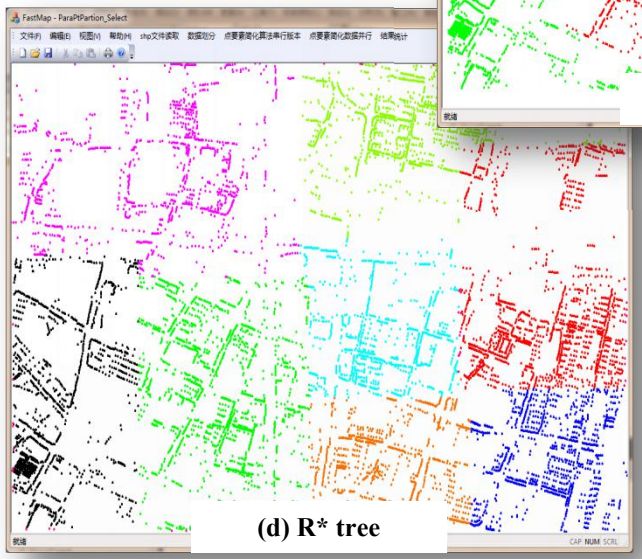


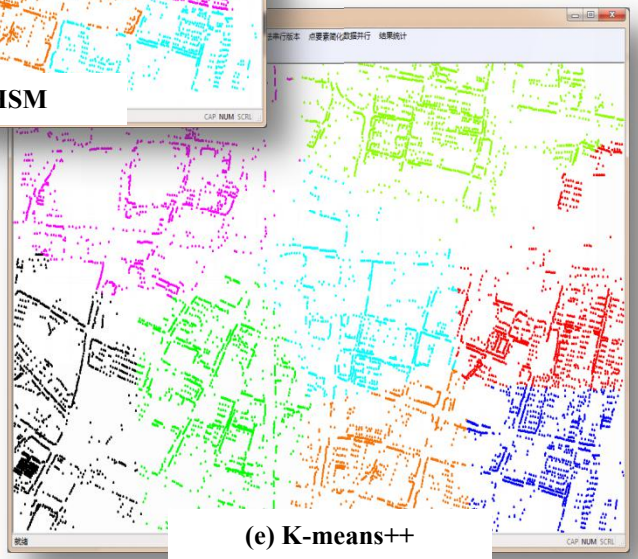
(a) 原始数据集


(b) Hilbert

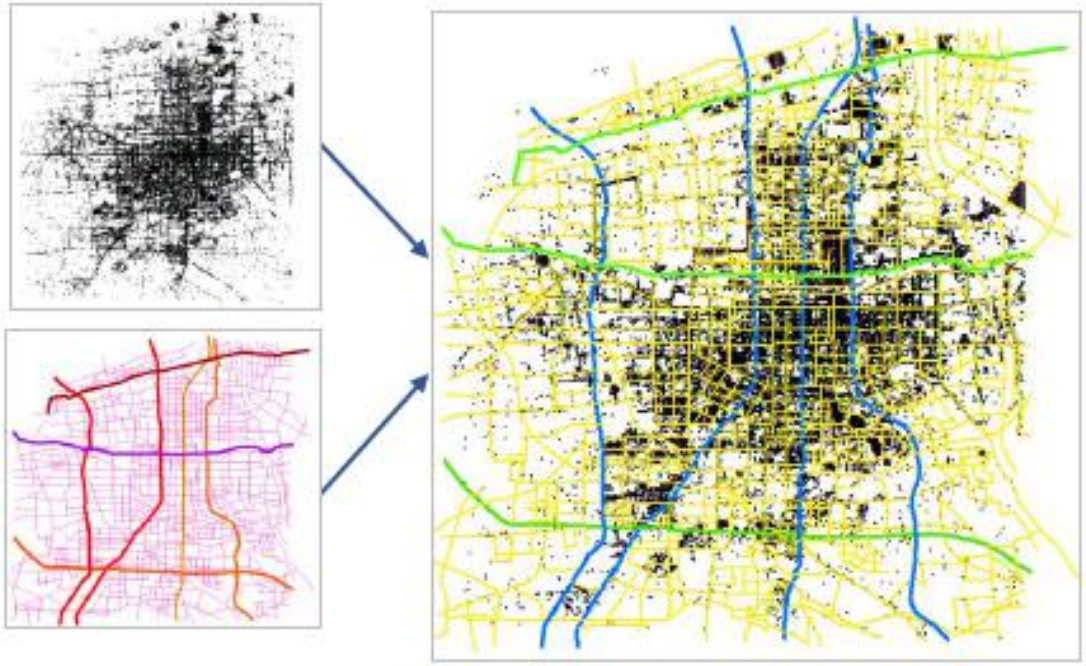

-KD tree


(f) DMHSM


(d) R* tree


(e) K-means++

Decomposition methods	Decomposition result								time
	1	2	3	4	5	6	7	8	(s)
Hilbert code	1639	1552	1525	1813	1521	1595	1552	809	6.039
N-KD tree	1500	1500	1500	1500	1501	1501	1502	1502	0.221
K-means++	1526	1765	1977	1076	1793	1513	1082	980	0.082
Road mesh tree	1274	1908	1990	1177	1950	1097	1544	1066	1.573
R star tree	1296	1577	1466	1382	1412	1333	1285	2242	2.584



(a) Test data 1 (Xian)

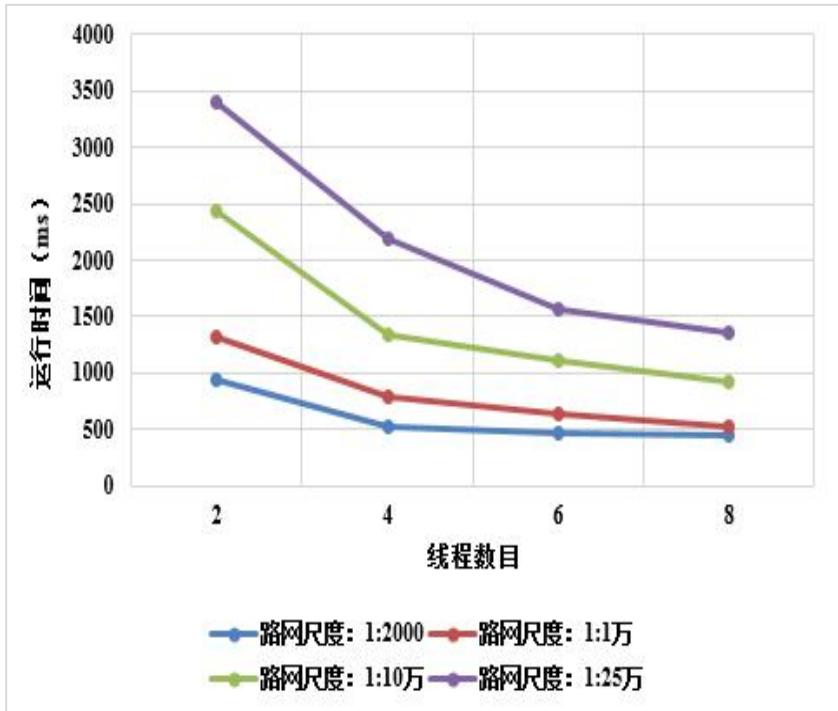
(b) Test data 2 (Beijing)

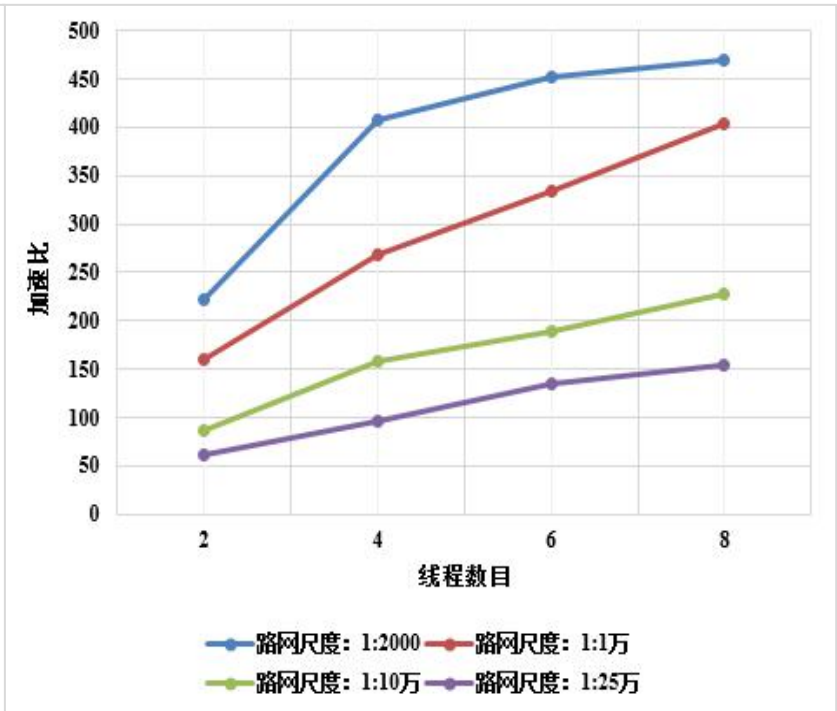
Parallel computing of POI selection

Area	POI	Scale	Stroke
Beijing	207710	1:2000	2042
		1:10 k	649
		1:100k	361
		1:250k	225
Xian	59650	1:2000	1408
		1:10 k	816
		1:100k	389
		1:250k	241

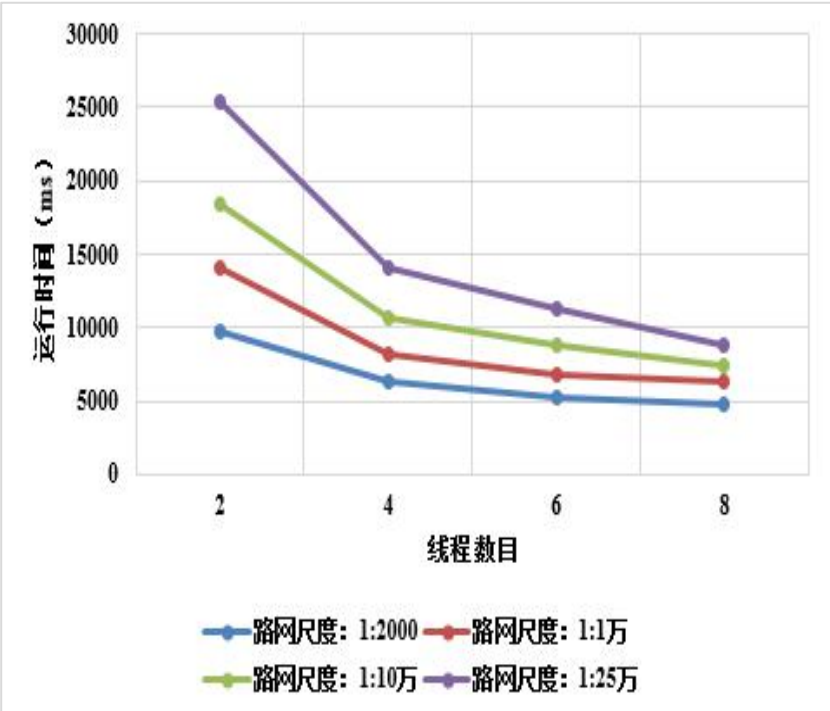
POI generalization platform: GDAL2.0.0+MFC

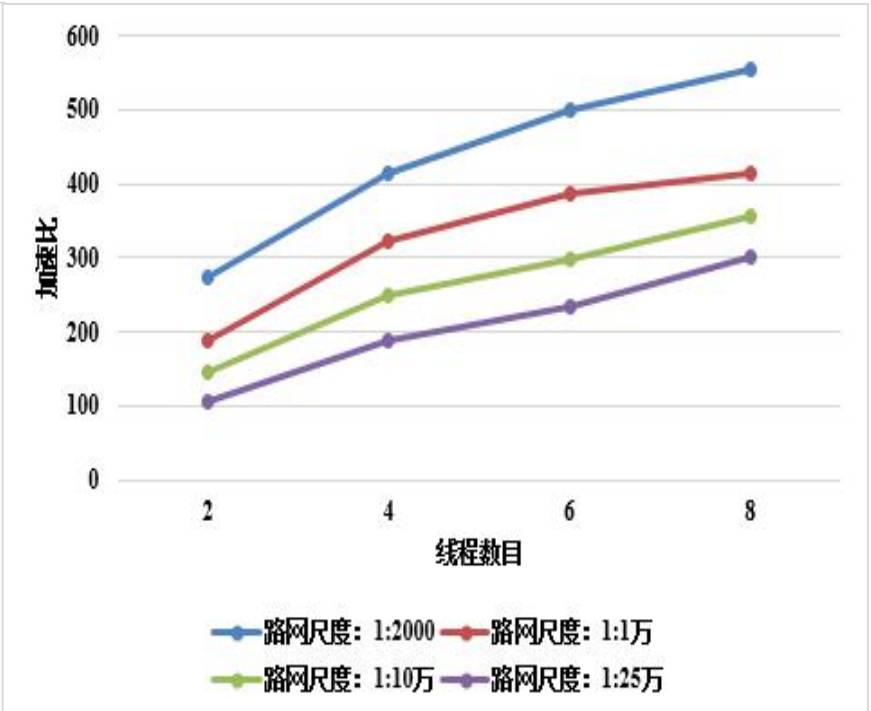
Results of Xian



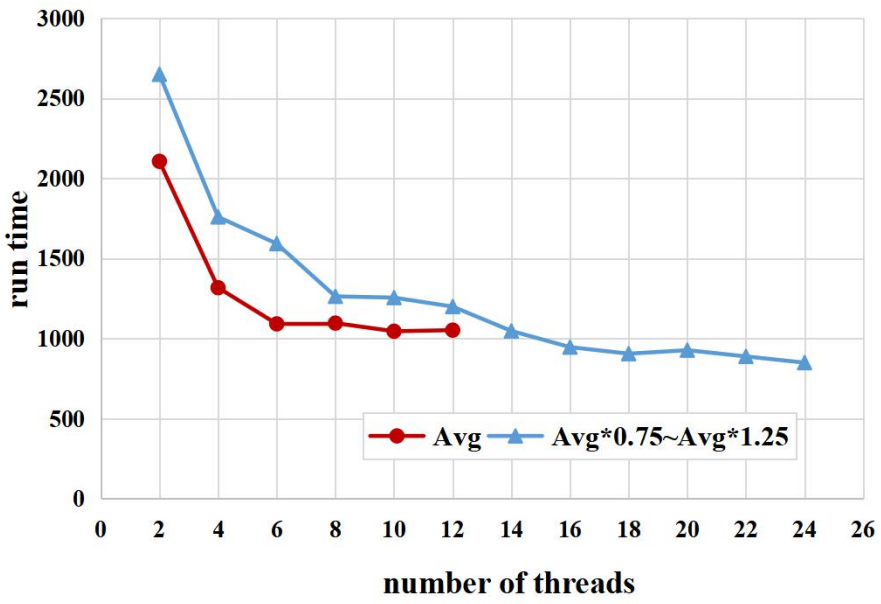


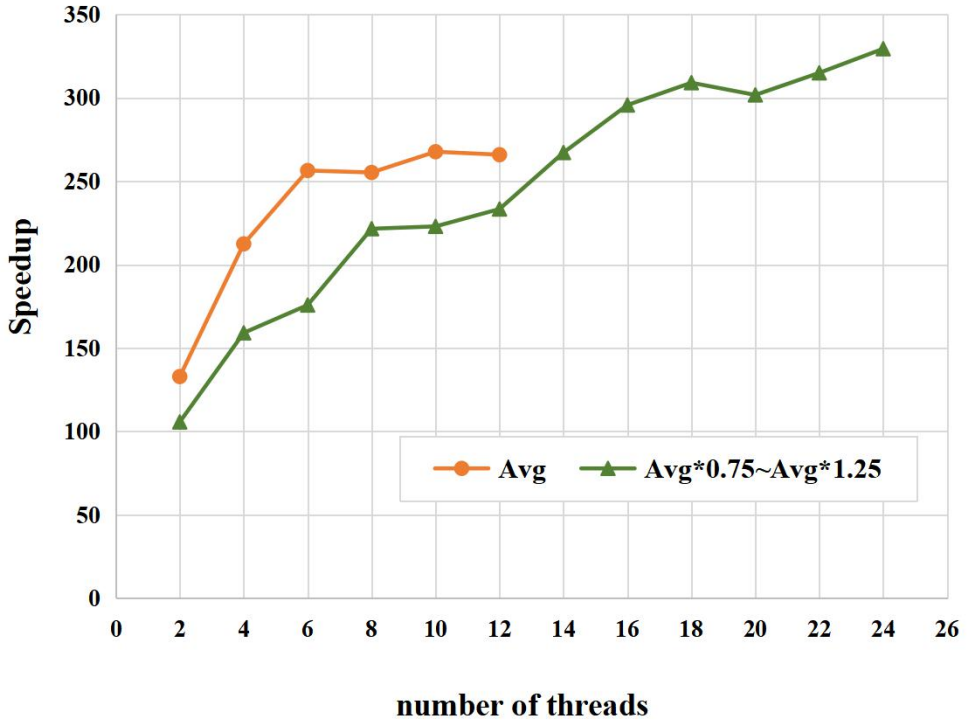
Results of Beijing





Results of Xian data on 24 threads

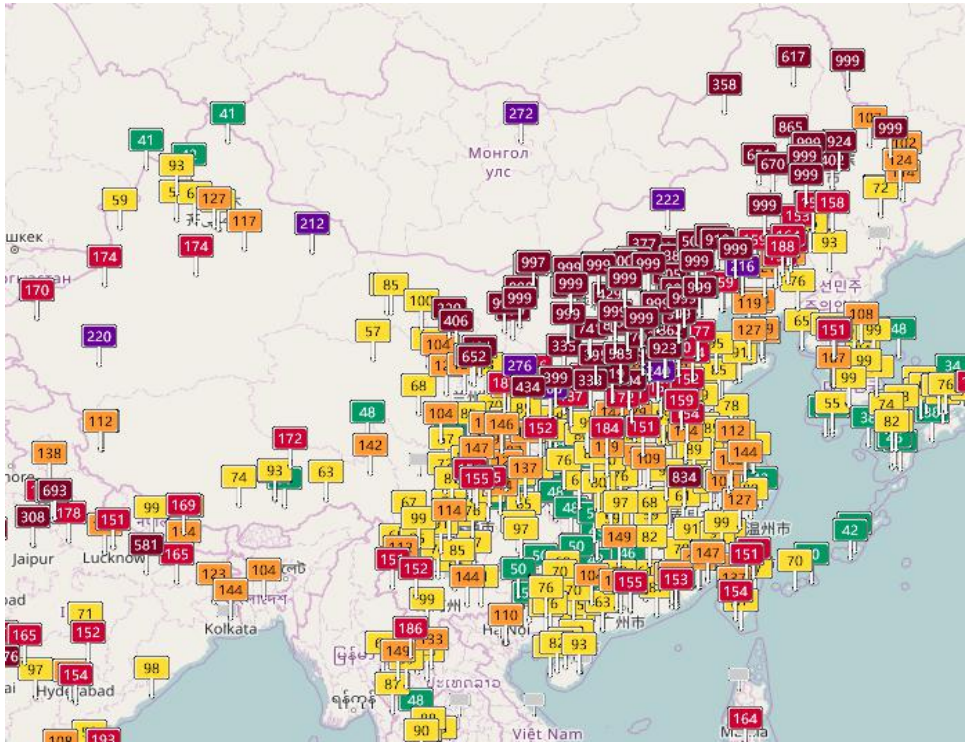


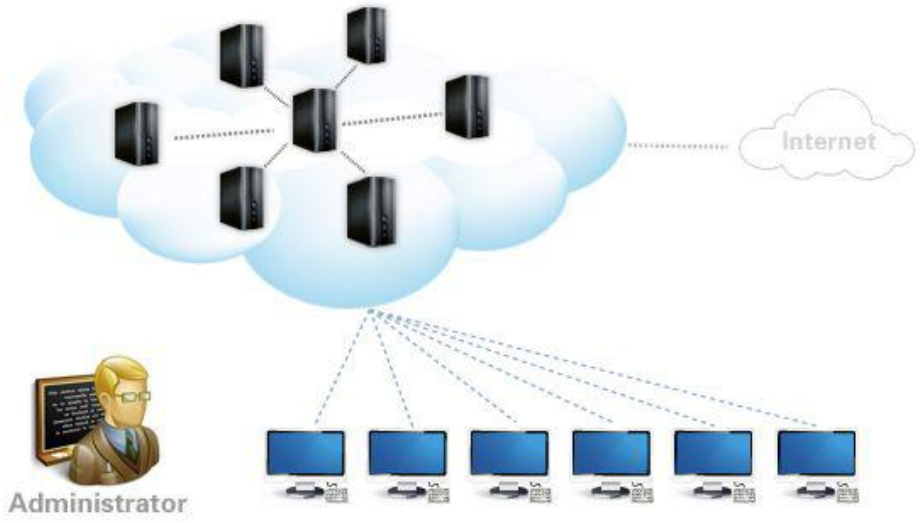


Outline

Background

With the proposal of **cloud computing** method, the efficiency of map generation has a new computing platform.





Research objectives

1. Proposing a method for calculating map load which is more effective and suitable for cloud computing environment.
2. Building a cloud-based points generalization platform.
3. Based on the MapReduce parallel computing model, some frequently-used points generalization operators are implemented.
4. The validity and efficiency of the functions on the platform are verified with the real point data.

A method of calculating the map load on cloud (1/2)

Map load is an indicator to measure the account of map content, and also is the main constraint to control map generalization.

The traditional cartographic generalization relies on the Square Root Model to calculate the map load. The formula is as follows:

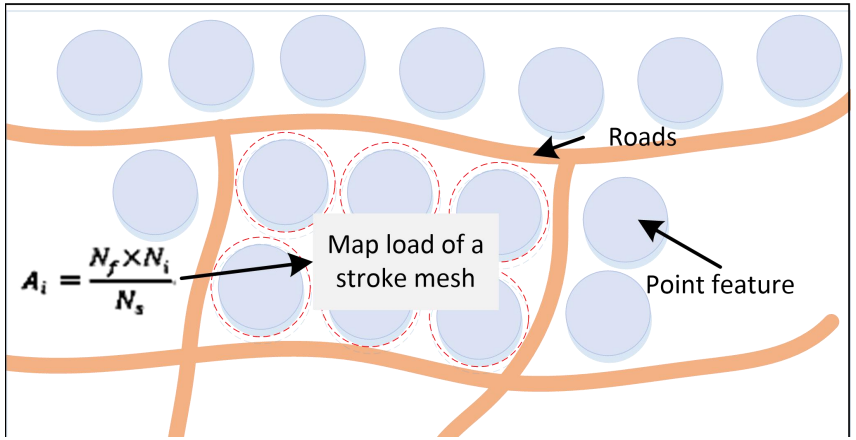
$$N_f = N_s \sqrt{\frac{M_s}{M_f}}$$

No.	Operator	Algorithms	Time complexity
1	Selection	Selection based on the attribute information	$O(n)$
2		Selection based on the relation	$O(n)$
3	Simplification	settlement-spacing ratio algorithm	$O(n^2)$; Voronoi$O(n\log n)$
4		distribution-coefficient algorithm	$O(n^2)$
5		gravity-modeling algorithm	$O(n^2)$ ； Delaunay$O(n\log n)$
6		Circle growth algorithm	$O(n^3\log n)$； $O(n\log n)$； $O(n^3)$
7		综述算法	$O(n^3\sqrt{n}\log n)$

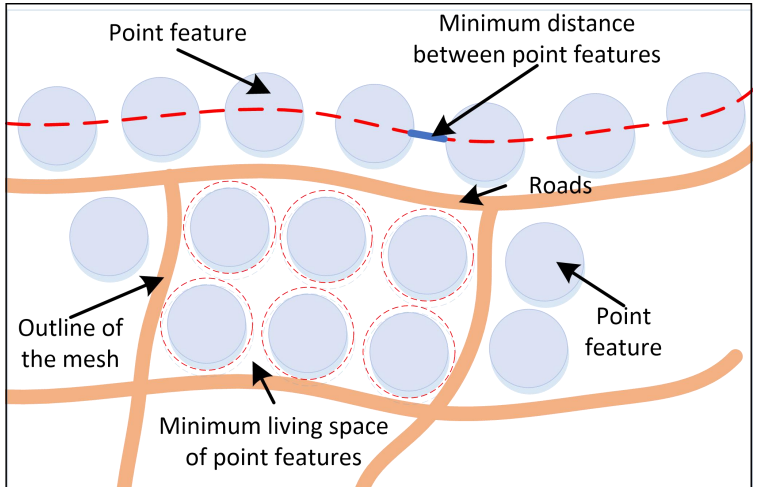
- The traditional method of calculating the map load only considers the overall map load.
- It is difficult to maintain the spatial distribution of data and is not suitable for the multi-nodes environment of the cloud platform.

A method calculating of map load on cloud (2/2)

The Method of Calculating the Point Load Considering the Constraints of the Square Root Model and the Hierarchical Stroke Mesh



Constraints of the Square Root Model



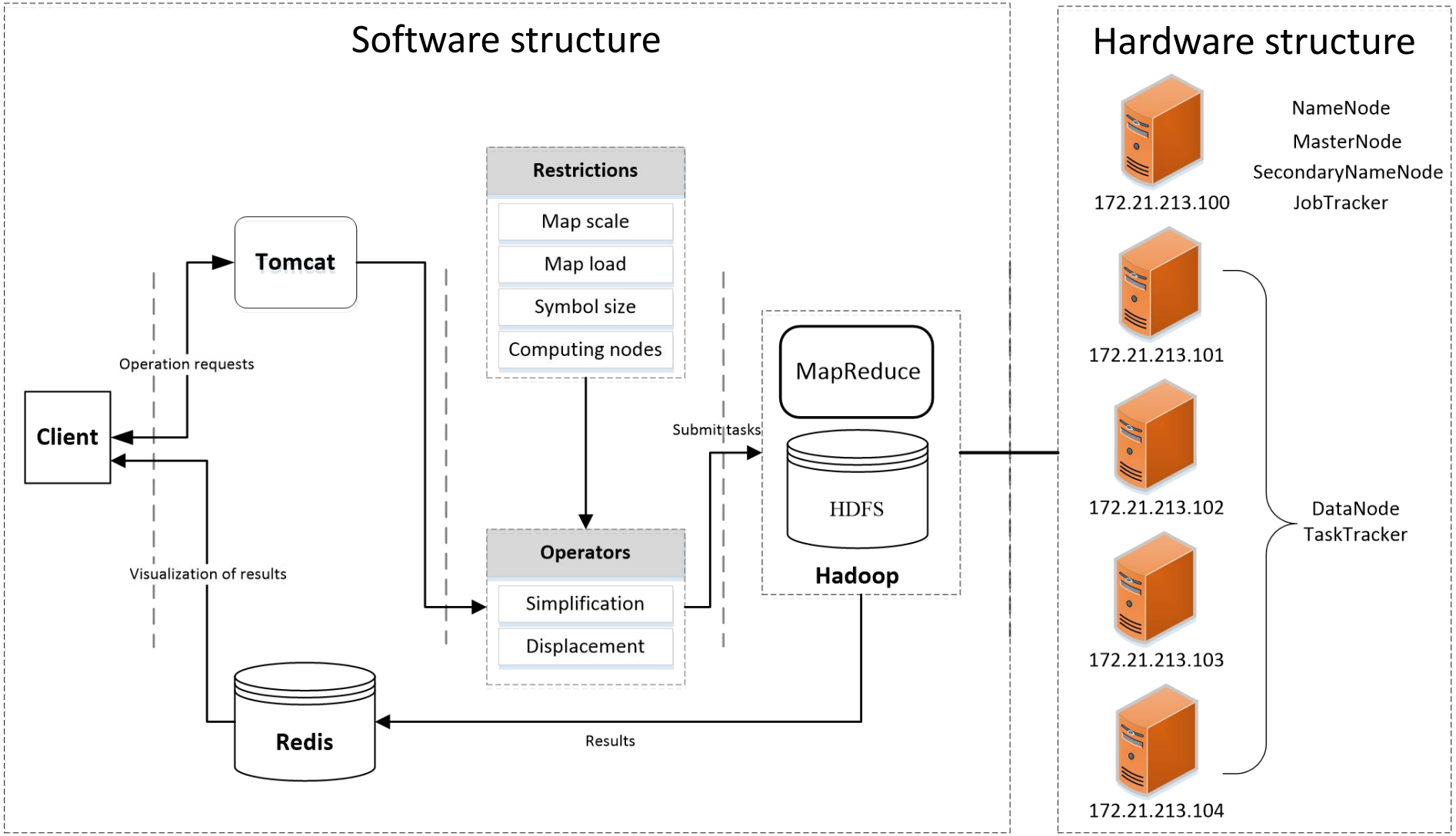
Constraints of the Hierarchical Stroke Mesh

Compared the results of the two constraints in a mesh, the **smaller one** is the **main restriction** in the mesh.

1、This method considers the **load balancing**, and is suitable for the **cloud platform**.

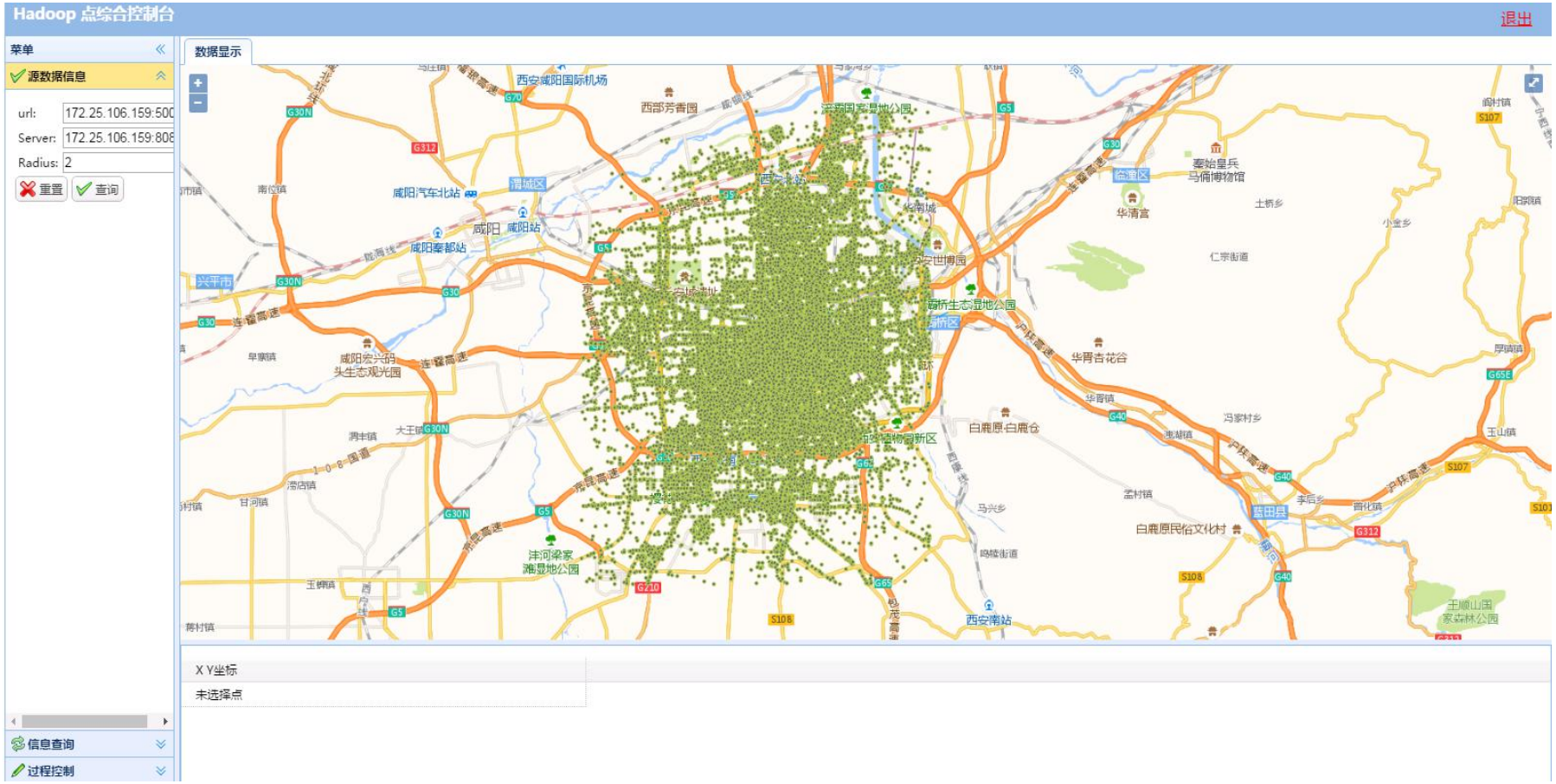
2、Maintaining the **spatial distribution** of the points.

PGPC (Points Generalization Platform based on Cloud)



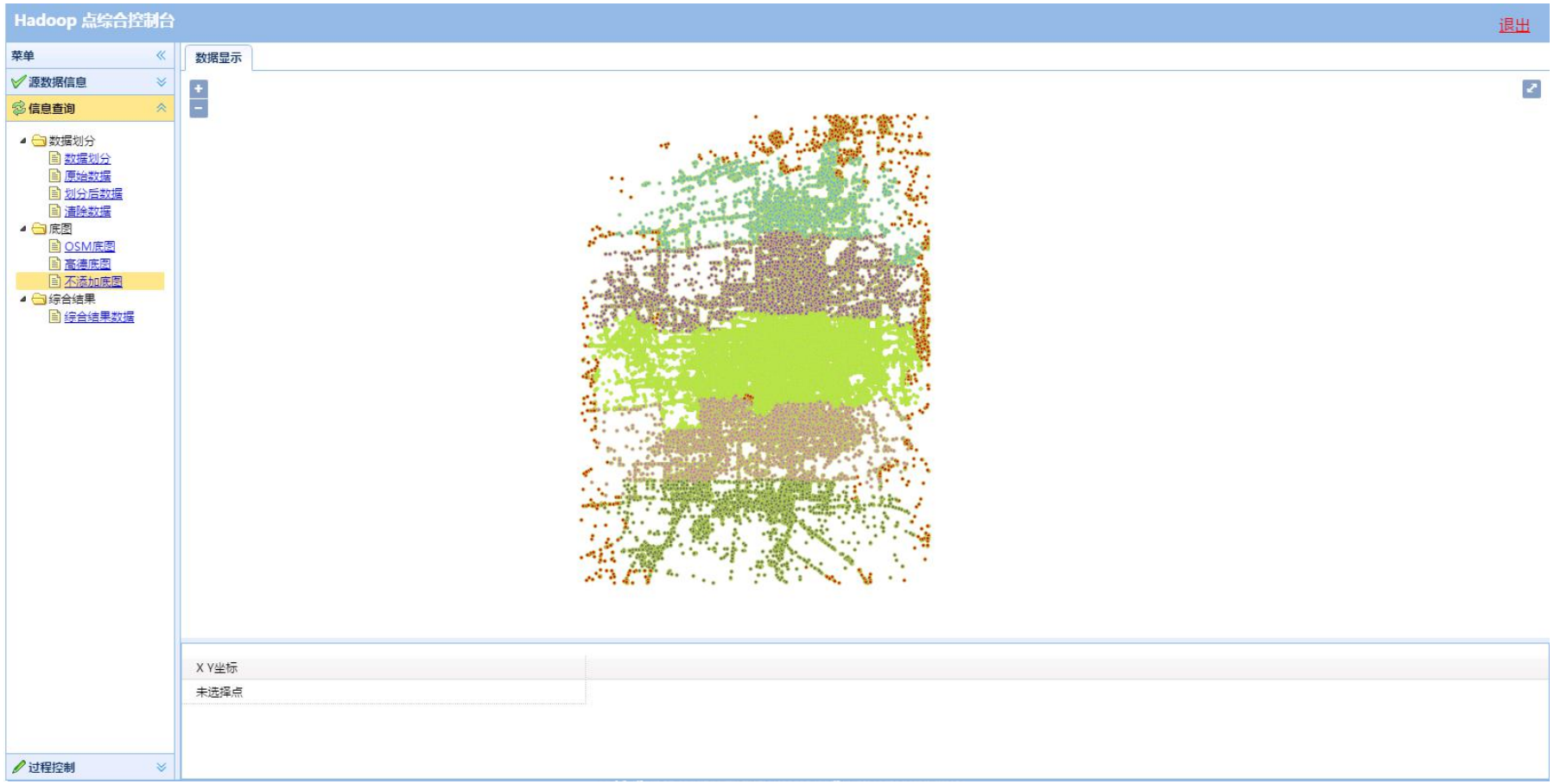
Functions of PGPC (1/4)

- Introduction
- Data query
- Data decomposition
- Points simplification
- Points displacement



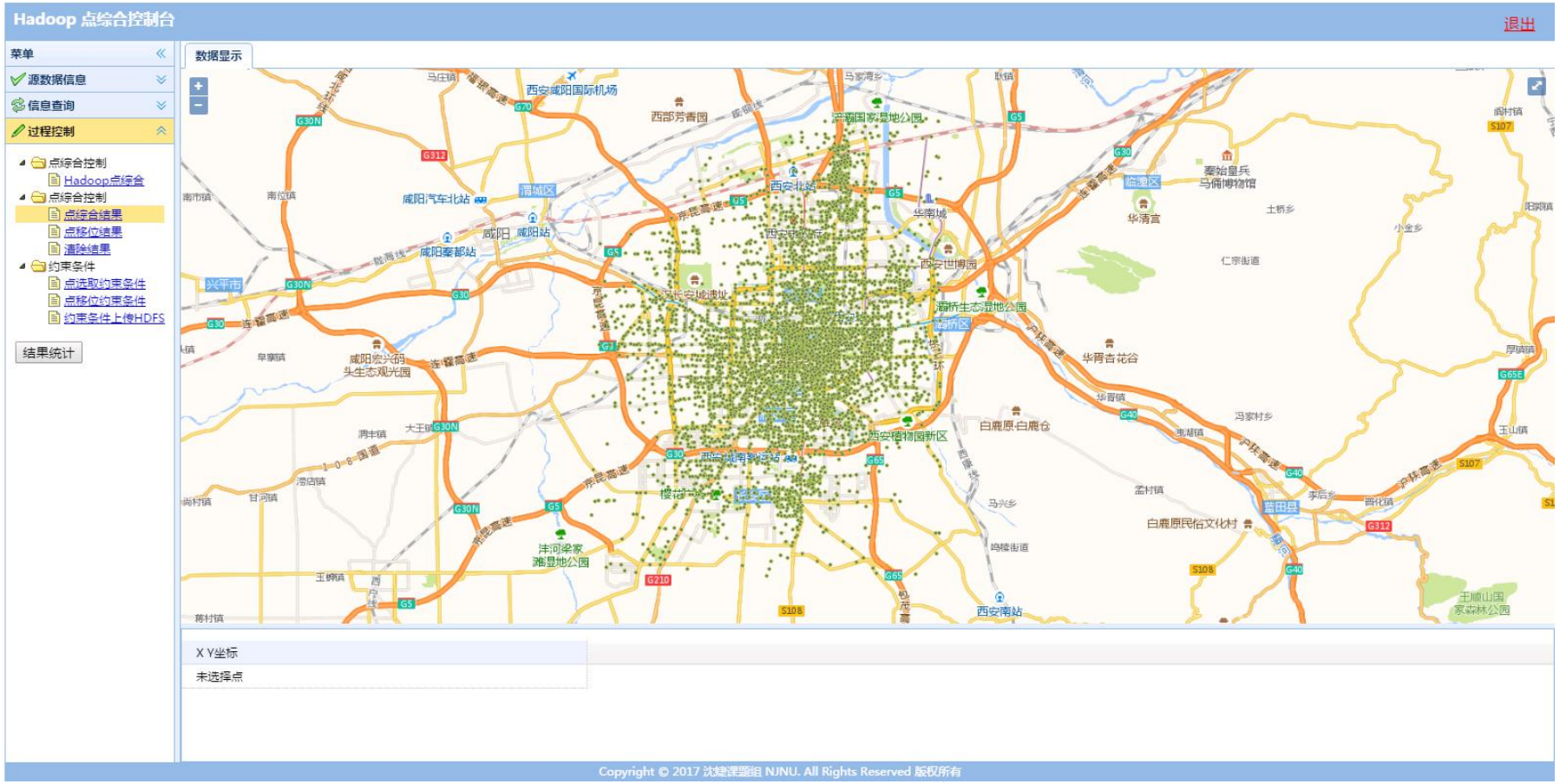
Functions of PGPC (2/4)

- Introduction
- Data query
- Data decomposition
- Points simplification
- Points displacement



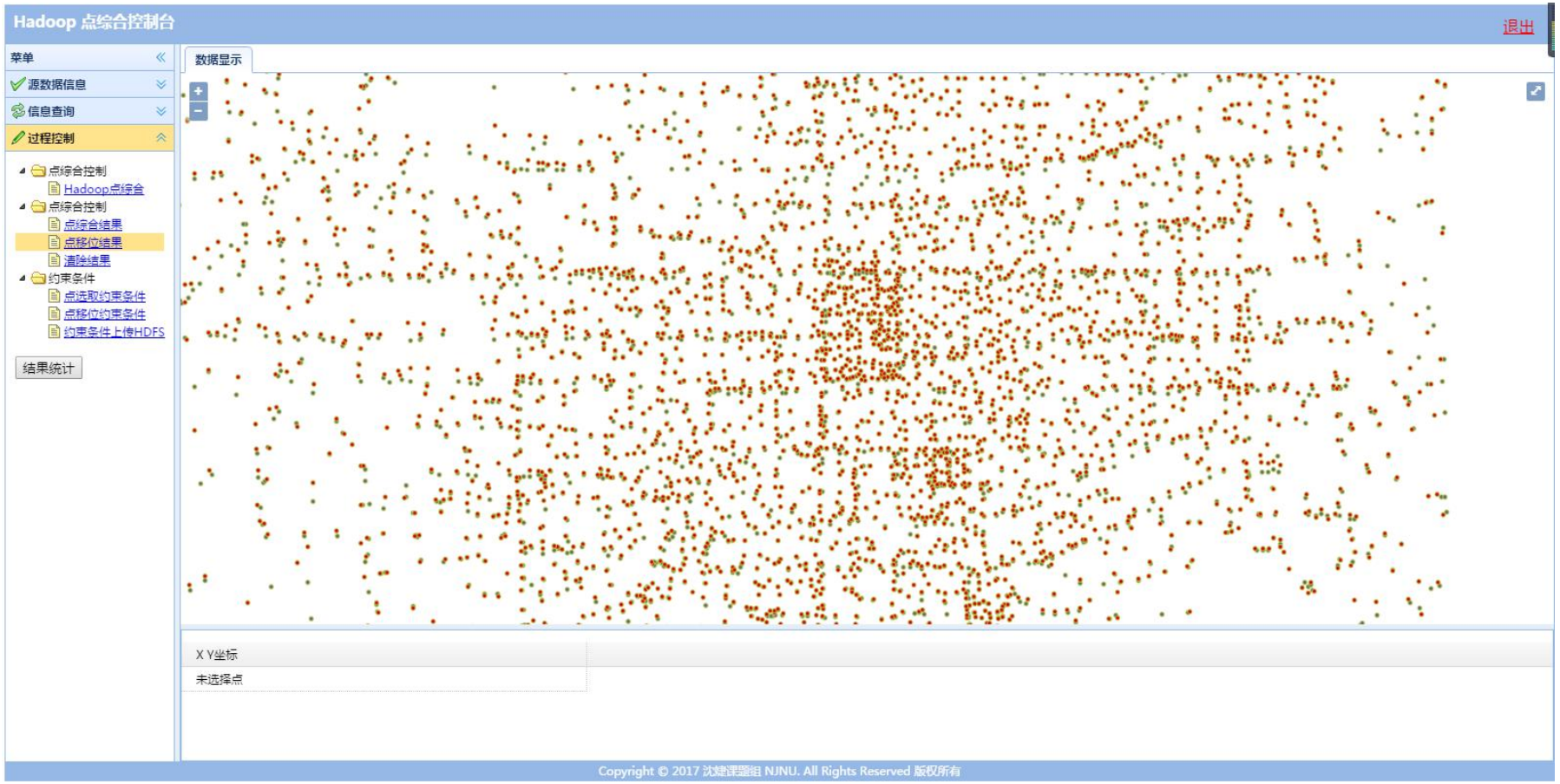
Functions of PGPC (3/4)

- Introduction
- Data query
- Data decomposition
- Points simplification
- Points displacement



Functions of PGPC (4/4)

- Introduction
- Data query
- Data decomposition
- Points simplification
- Points displacement



Red dots are the points after **displacement**, **green** dots are **original data**.

Experiment and data

Experimental design

- Comparing the results of the circle-growth algorithm on PGPSM and PGPC

  Data: Point data and road data of Xian

  Purpose: Verify quality of PGPC

- Comparing the results of points selection operator using different number of nodes on PGPC with that on sequential calculation environment

  Data: Data of Nanjing, data of Beijing, data of Xian and simulation data 1~5

  Purpose: Verify efficiency of PGPC

Experimental Data

Layers	Number of points or roads	Amount of data
Point data of Xinjiekou Nanjing	12,006	1.14M
Road data of Xinjiekou Nanjing	70	26.5K
Point data of Xian	59,650	22.8M
Road data of Xian	513	362K
Point data of Beijing	207,710	171M
Road data of Beijing	1,292	314K
Simulation point data of Xian_1	300,000	189M
Simulation point data of Xian_2	500,000	240M
Simulation point data of Xian_3	1,000,000	495M
Simulation point data of Xian_4	2,000,000	1.3G
Simulation point data of Xian_5	3,000,000	5.3G

- **PGPSM：** Points generalization platform based on single machine environment

Results of the circle-growth algorithm

Raw data (Stroke mesh + points of Xian)

Result in PGPSM

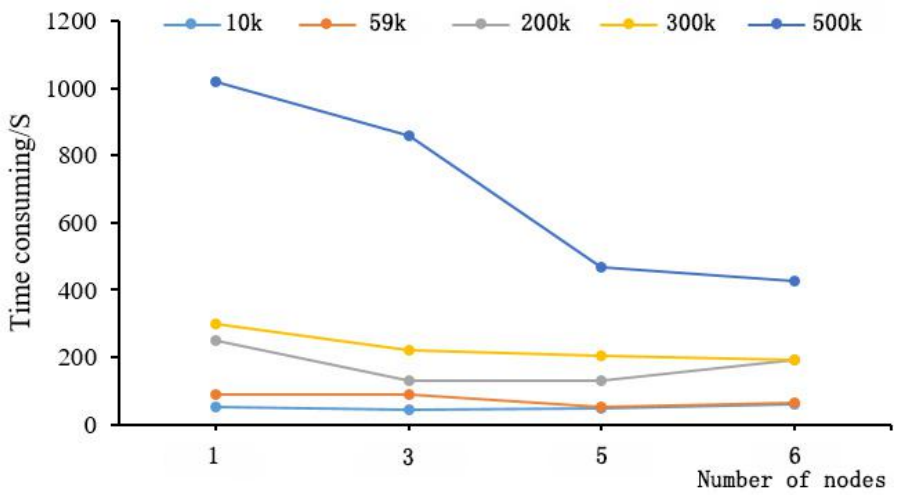
Result in PGPC

Comparing result of PGPSM with that of PGPC, we can find that:

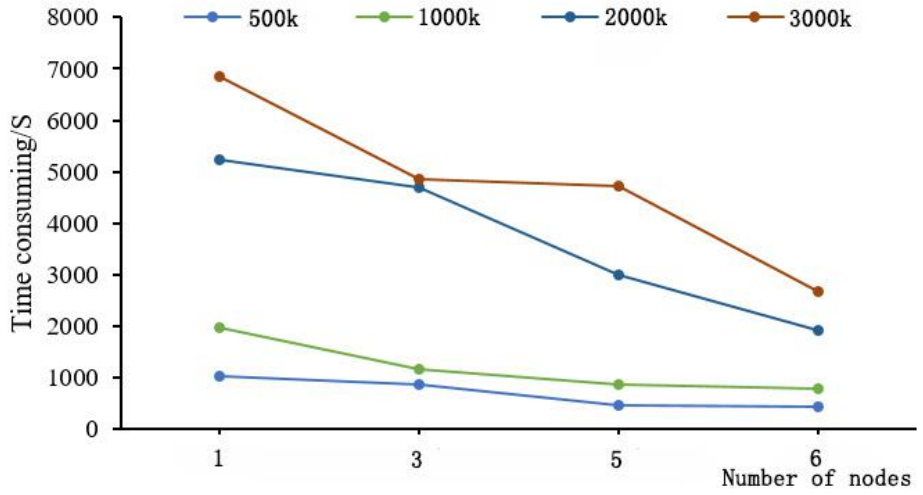
- The spatial distribution of result in PGPSM is scattered, which is different from the spatial distribution of the original point data set;
- The result in PGPC is very good to maintain the points spatial distribution characteristics.

Efficiency of PGPC (1/2)

Time consuming of points selection with 10k~500k points



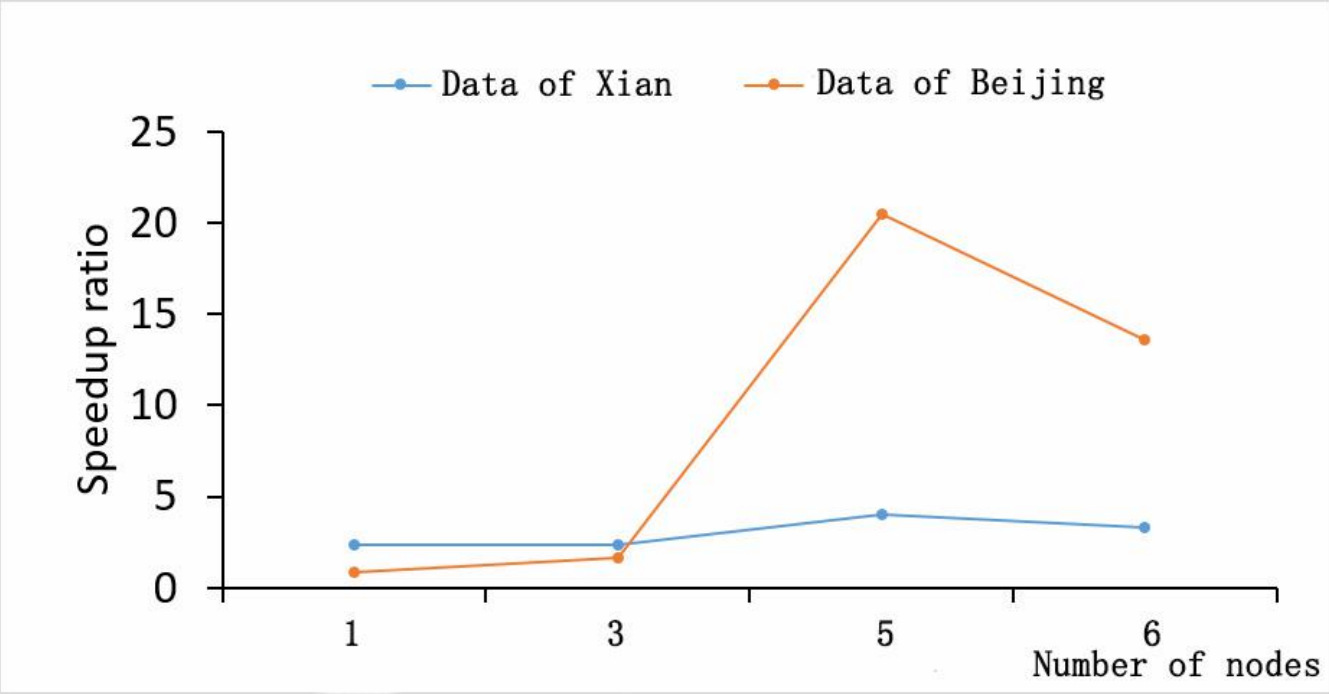
Time consuming of points selection with 500k~3000k points



- When the amount of data is low, the efficiency of PGPC is not obvious. But when the data volume is large, PGPC can greatly improve the computational efficiency.
- In the case of large number of points, we need to use as many computing nodes as possible; Otherwise, we should use fewer nodes.
- The number of nodes has a significant constraint on the point generalization process.

Efficiency of PGPC (2/2)

The speedup ratio compared with sequential calculation



- When the amount of data is large enough and the number of calculating nodes is reasonable, the cloud platform will greatly enhance efficiency of the points generalization.

Patents

证书号第2378202号

发明专利证书

发明名称：一种顾及空间邻近关系的矢量等高线数据划分方法

发明人：沈婕;朱伟;吴鹏;尹子鹤

专利号：ZL 2014 1 0004177.9

专利申请日：2014年01月03日

专利权人：南京师范大学

授权公告日：2017年02月15日

本发明经过本局依照中华人民共和国专利法进行审查，决定授予专利权，颁发本证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。

本专利的专利权期限为二十年，自申请日起算。专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年01月03日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。

局长
申长雨

中华人民共和国国家知识产权局
2017年02月15日

证书号第2006943号

发明专利证书

发明名称：基于道路网眼层次结构划分的POI简化并行计算方法

发明人：沈婕;郭立帅;吴银丽

专利号：ZL 2013 1 0197049.6

专利申请日：2013年05月23日

专利权人：南京师范大学

授权公告日：2016年03月30日

本发明经过本局依照中华人民共和国专利法进行审查，决定授予专利权，颁发本证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。

本专利的专利权期限为二十年，自申请日起算。专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年05月23日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。

局长
申长雨

中华人民共和国国家知识产权局
2016年03月30日

证书号第2109031号

北京市海淀区地方税务局 印花税代扣专用章 国家知识产权局

发明专利证书

发明名称：一种面向并行POI简化的任务拆分与分发方法

发明人：沈婕;郭立帅;朱伟;吴银丽

专利号：ZL 2013 1 0197054.7

专利申请日：2013年05月23日

专利权人：南京师范大学

授权公告日：2016年06月15日

本发明经过本局依照中华人民共和国专利法进行审查，决定授予专利权，颁发本证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。

本专利的专利权期限为二十年，自申请日起算。专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年05月23日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。

局长
申长雨

中华人民共和国国家知识产权局
2016年06月15日

Software copyright

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第0550110号

软件名称：基于OPENMP的线要素光滑系统软件 V1.0

著作权人：南京师范大学

开发完成日期：2012年09月01日

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第1470928号

软件名称：基于OpenMP的点要素并行简化系统软件 V1.0

著作权人：南京师范大学

开发完成日期：2015年12月01日

首次发表日期：2015年12月01日

权利取得方式：原始取得

权利范围：全部权利

登记号：2016SR292311

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第1470928号

软件名称：基于OpenMP的点要素并行简化系统软件 V1.0

著作权人：南京师范大学

开发完成日期：2015年12月01日

首次发表日期：2015年12月01日

权利取得方式：原始取得

权利范围：全部权利

登记号：2016SR292311

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第0550110号

软件名称：基于OPENMP的线要素光滑系统软件 V1.0

著作权人：南京师范大学

开发完成日期：2012年09月01日

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第1470928号

软件名称：基于OpenMP的点要素并行简化系统软件 V1.0

著作权人：南京师范大学

开发完成日期：2015年12月01日

首次发表日期：2015年12月01日

权利取得方式：原始取得

权利范围：全部权利

登记号：2016SR292311

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第0550110号

软件名称：基于OPENMP的线要素光滑系统软件 V1.0

著作权人：南京师范大学

开发完成日期：2012年09月01日

中华人民共和国国家版权局

计算机软件著作权登记证书

证书号：软著登字第1470928号

软件名称：基于OpenMP的点要素并行简化系统软件 V1.0

著作权人：南京师范大学

开发完成日期：2015年12月01日

首次发表日期：2015年12月01日

权利取得方式：原始取得

权利范围：全部权利

登记号：2016SR292311

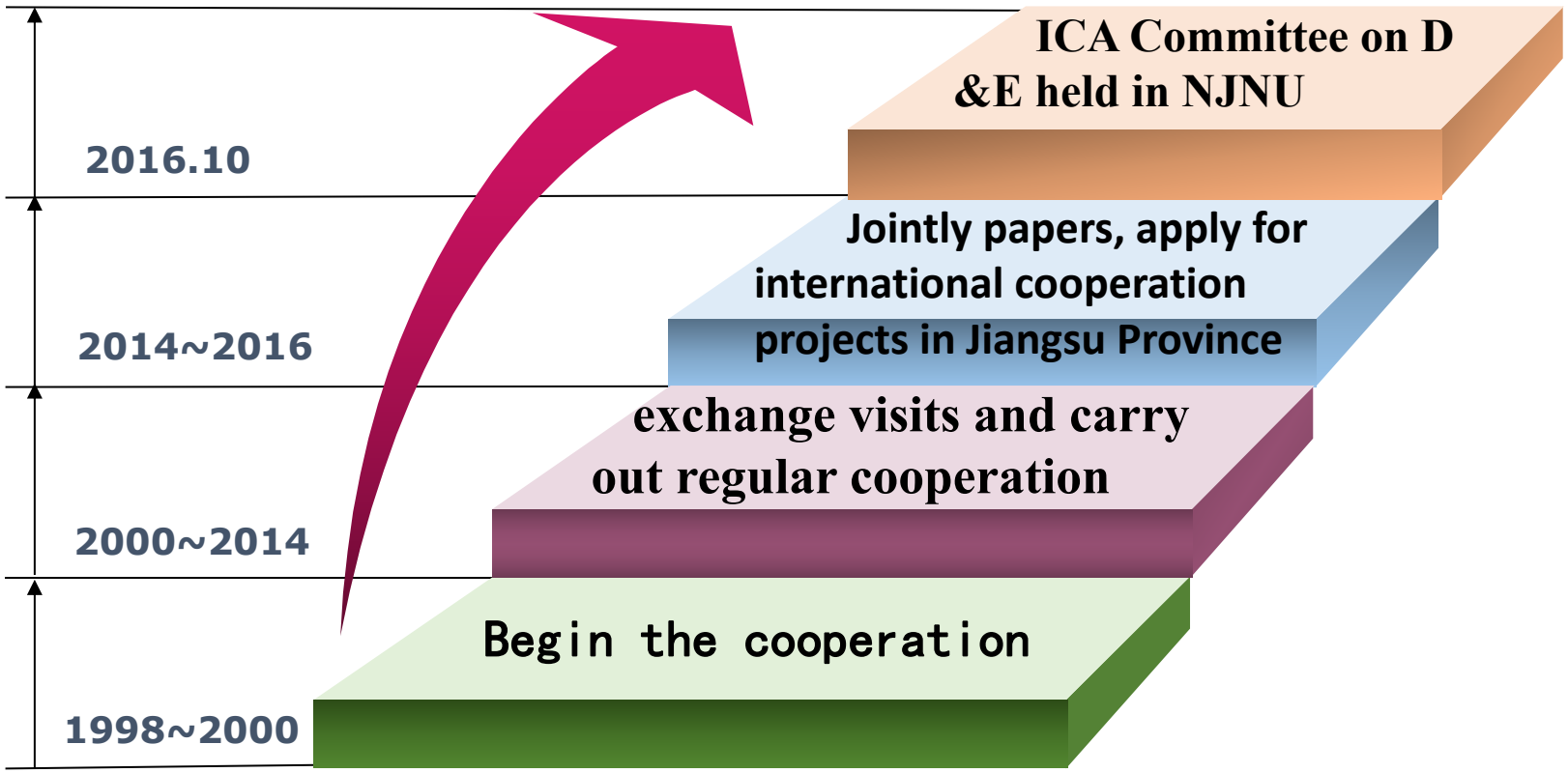
Outline

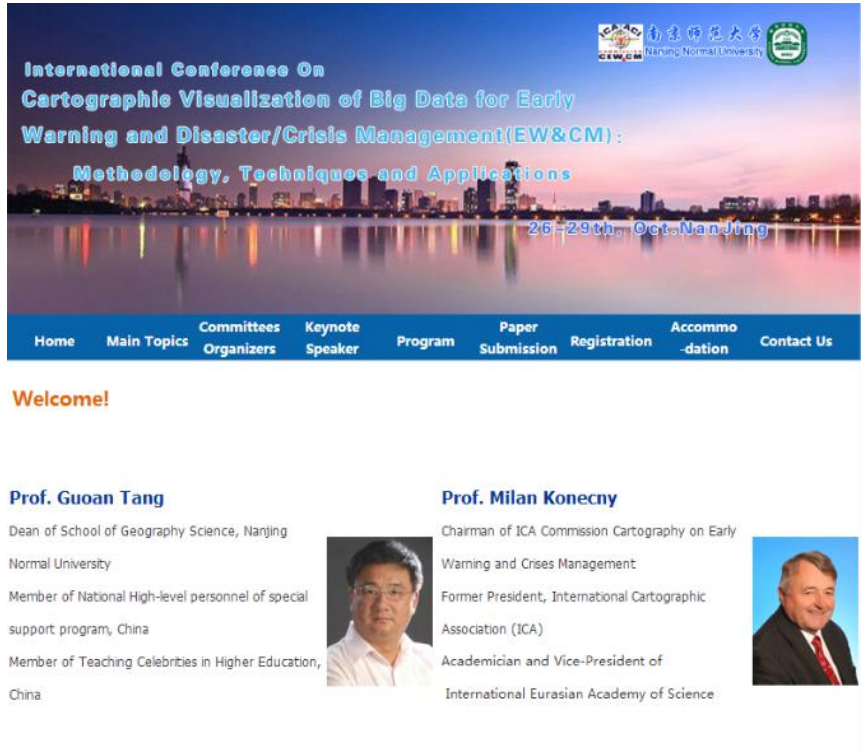
China-Czech Intergovernmental Science and Technology Cooperation Project 2017.4—2019.12

Dynamic mapping for risk and crisis management in big data era

- Now the staff has **50** members, including **1** academician, **11** doctor's tutors, **20** professors, **20** associate professors, **8**instructors, **1** secretary and **1** lab assistant.
- National Science and Technology Platform: Data-sharing Network of China Earth System Science—Network branch of middle and lower reaches of Yangtse River
  - Environment and disaster
  - Geography
  - Economy
  - Land use……

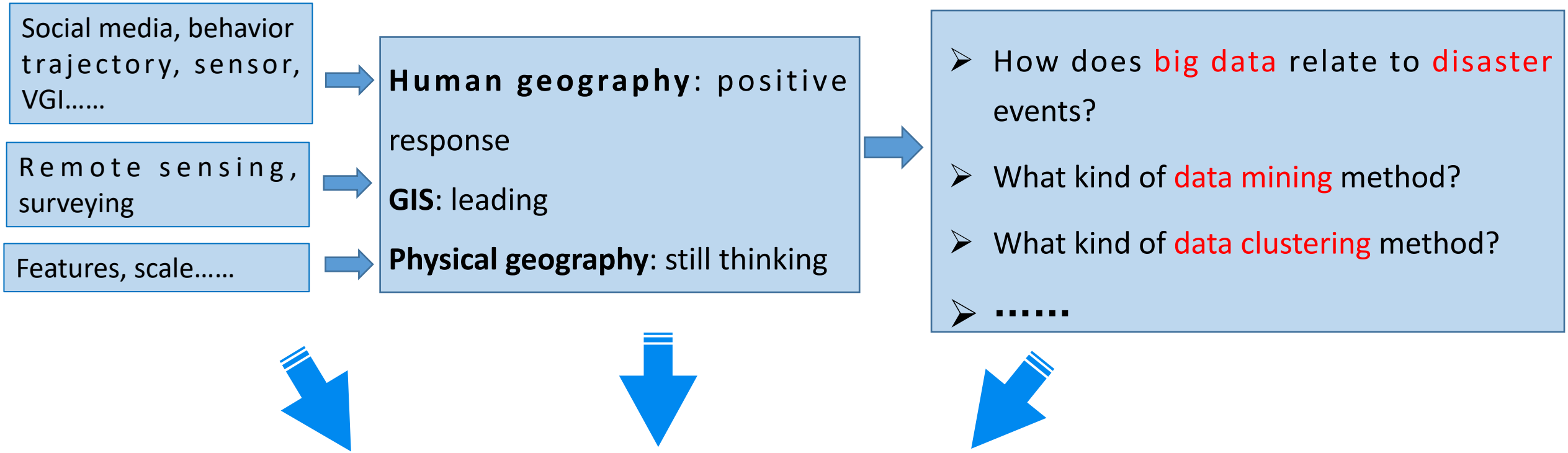
Cooperation basis





Background

Geography encounters big data: coping and thinking



The core of geographic research: Earth Surface System

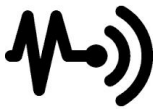
Scientific data: the basis for scientific research

The Core issues of geography: pattern, process, mechanism

Select data source for different types of disasters

Normal data source

Basic Geodatabase



Sensor

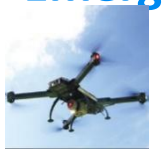
New data source

Social media

VGI

Business data

Emergency data source



Drone

Emergency RS

Surveying

- Map generalization model
- Map generalization method
- High performance computing

- Disaster type
- Scale, process, mechanism
- Disaster map design

- Map symbol
- Map hierarchical elements
- Map scale
- Map color

For different users, different disasters online map

- Area match
- Integration
- Map interaction
- Online edit

Public help map

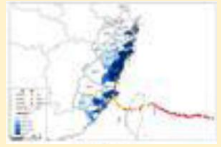
Government decision map

Disaster analysis map

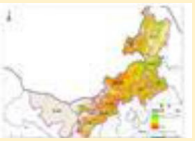
...

For different users

Earthquake



Typhoon



Drought

...

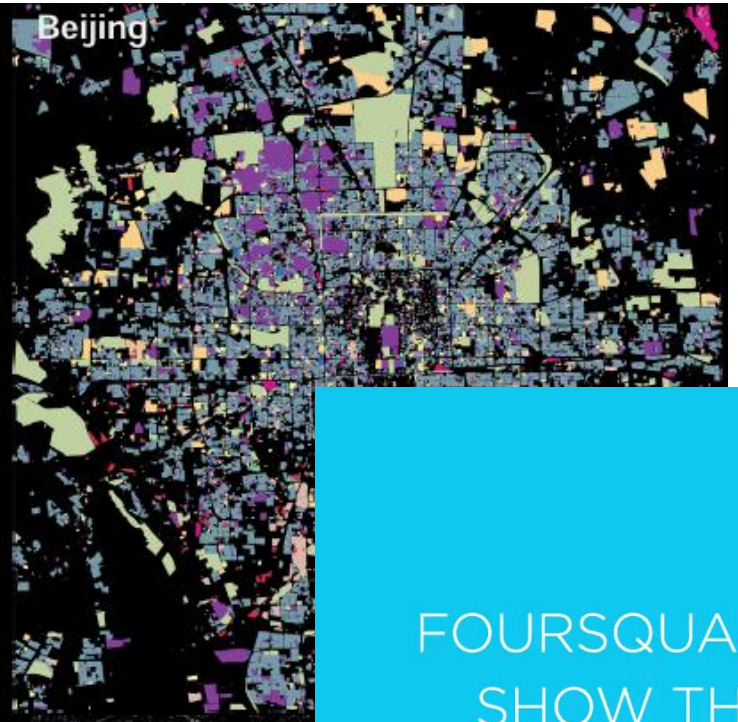
For different disasters

Outline

Urban Organs

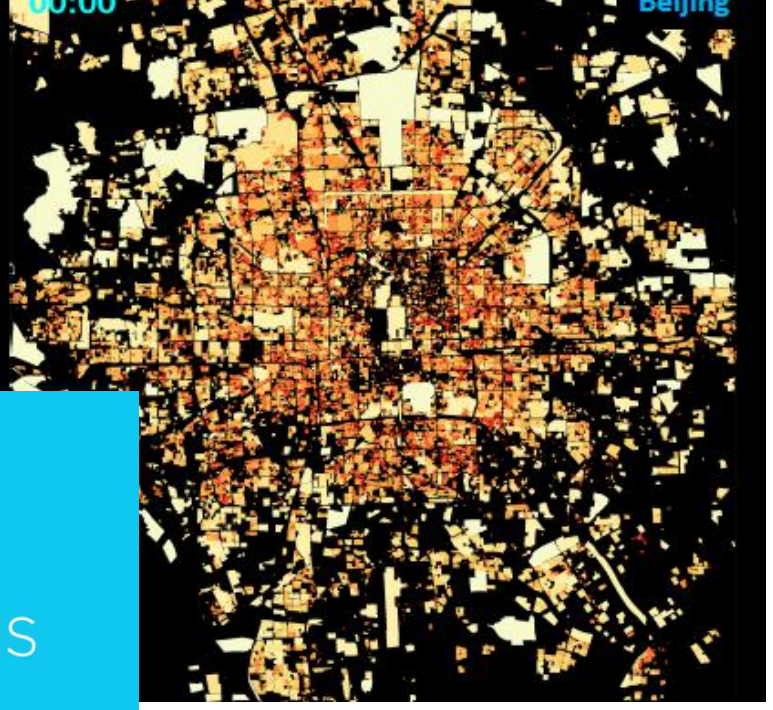
Inferring urban detailed land use functions through pan-space urban information with Tensor-based artificial intelligence algorithm.

- 20 classes
- Overall accuracy: 91%



Urban Rhythm

Using mobile phone positioning data, we map urban human dynamic at very fine spatio-temporal scales



FOURSQUARE CHECK-INS SHOW THE PULSE OF SAN FRANCISCO

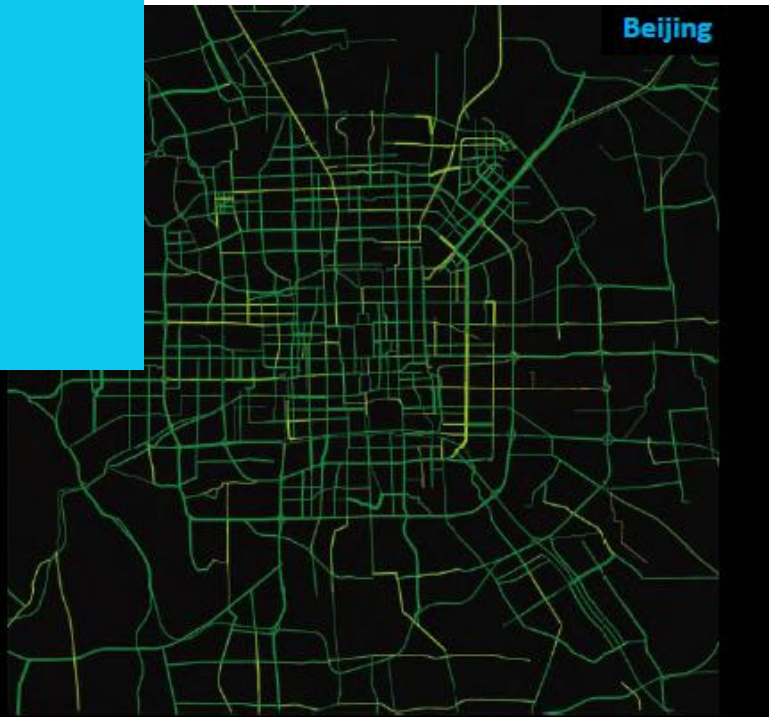
Urban Cells

Understanding detailed urban structure and behave through the point-of-interesting (POI) data.

Each point with different colors represent a urban unit (shop, school, hospital, or factory, etc.)

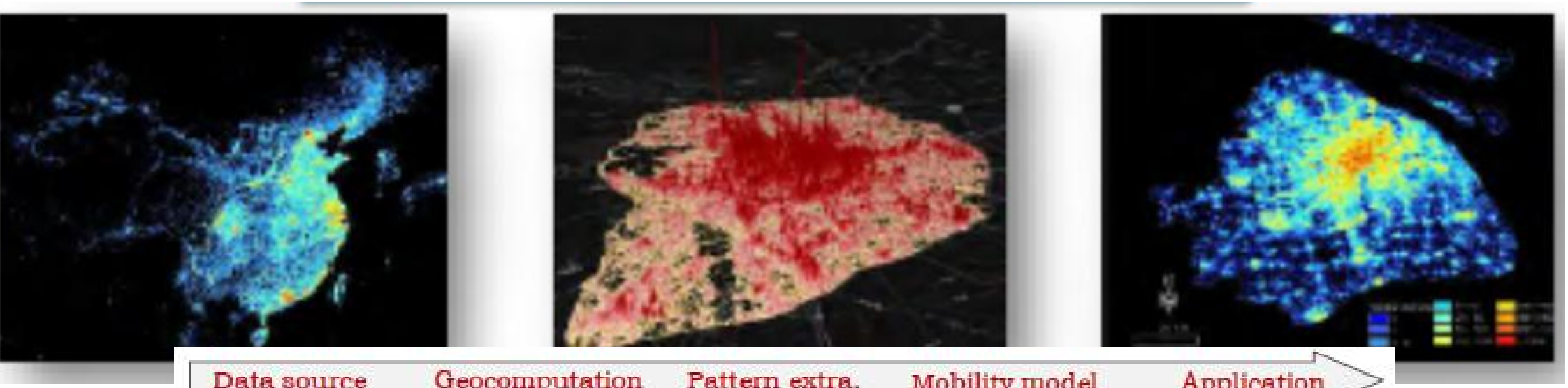
Algorithm: Bayesian deep learning with spatio-temporal correlations

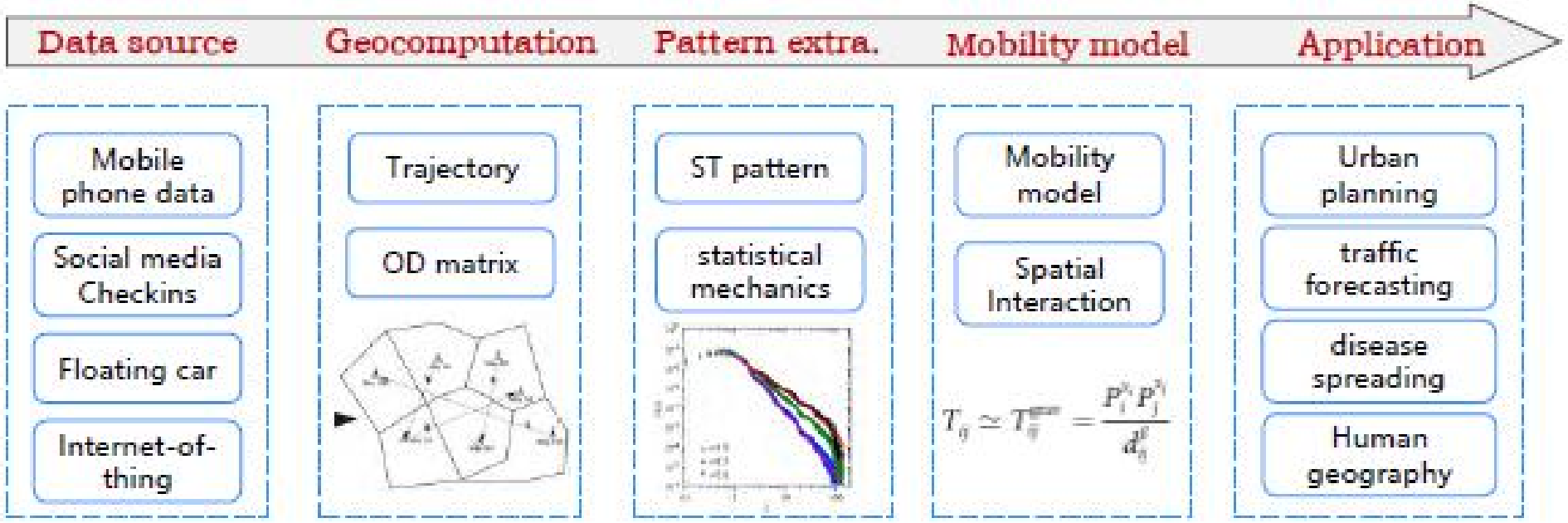
Forecasting accuracy: 85%



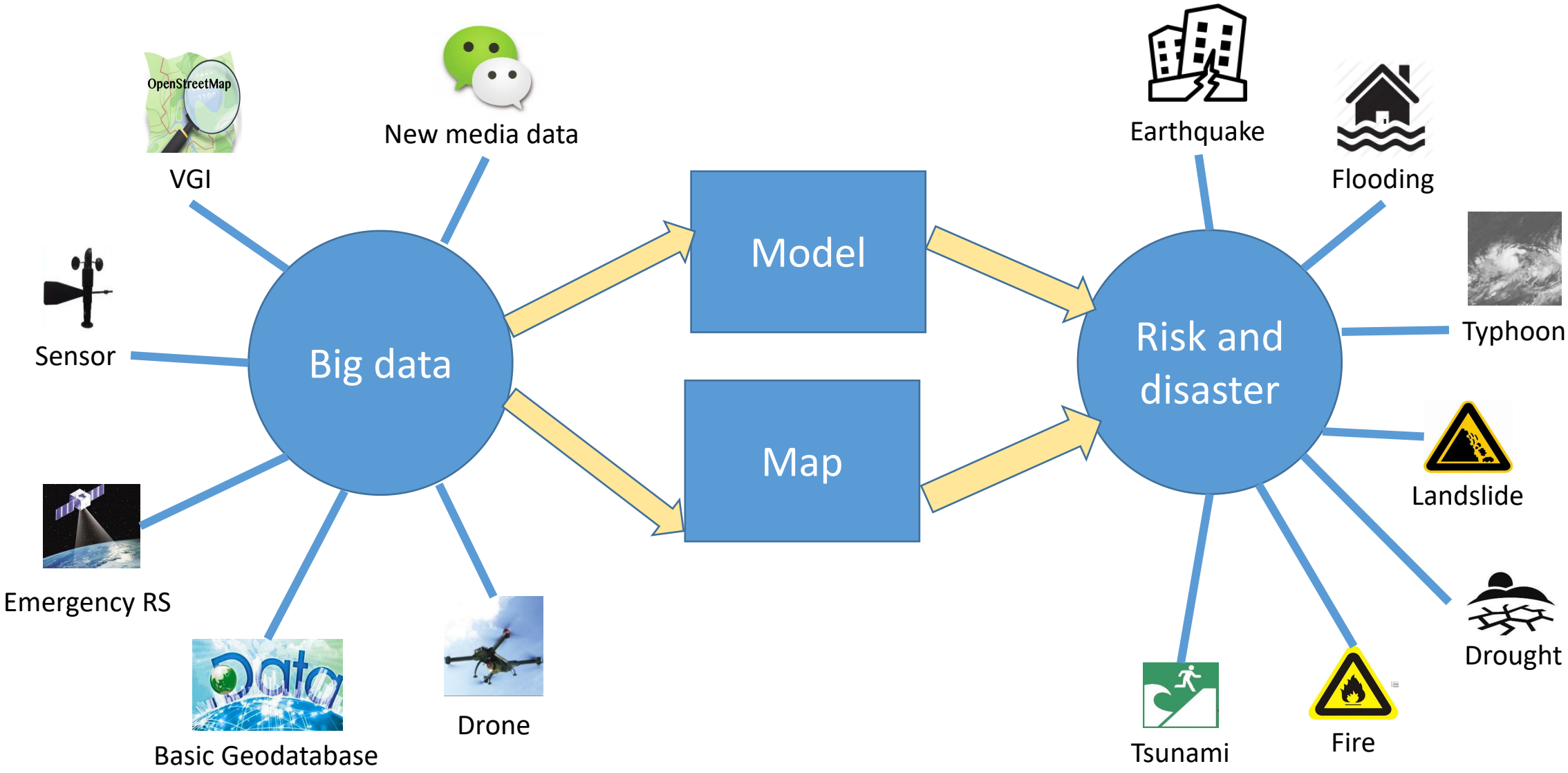
Dynamic people distribution at multiple scale

Location Based Service data from Tencent Big Data center:
800 million users/day
50 billion location request/day





Discussion


VGI
New media data
Sensor
Big data
Model
Map
Risk and disaster
Earthquake
Flooding
Typhoon
Landslide
Drought
Fire
Tsunami
Drone
Basic Geodatabase
Emergency RS

Thanks !

Q & A

Collaborators: Zhenguo Yu, Shuai Yang, Nanjing Normal university

Acknowledgement: China-Czech Intergovernmental Science and Technology Cooperation Project 2016