

地理要素的尺度、分区效应，可变面积单元问题与空间统计的挑战

戴劭劼

第十一届中国R语言会议, 人民大学, 2018.05.27

大纲

- 1 地理要素尺度、分区效应与可变面积单元问题
- 2 空间分层异质性
- 3 地理探测器原理
- 4 模型不确定性的定性分析
- 5 案例研究
- 6 挑战与未来发展

1 地理要素尺度、分区效应与可变面积单元问题

- 尺度效应通常是指由于研究进行的尺度变化而导致的研究结果的变化。
- 在地理学方面，在可变面积单元问题（MAUP）中，尺度效应已经研究了几十年。
- 尺度效应可以用尺度多样性，特征尺度和等级来解释，但也可能是由于采样和测量误差，数据重采样失真以及统计分析和建模中的缺陷造成的伪像。

Jianguo WU, Jones KB, Harbin LI, et al. SCALING AND UNCERTAINTY ANALYSIS IN ECOLOGY. 2007.

1 地理要素尺度、分区效应与可变面积单元问题

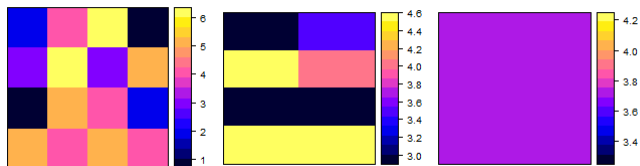
- 尺度效应通常是指由于研究进行的尺度变化而导致的研究结果的变化。
- 在地理学方面，在可变面积单元问题（MAUP）中，尺度效应已经研究了几十年。
- 尺度效应可以用尺度多样性，特征尺度和等级来解释，但也可能是由于采样和测量误差，数据重采样失真以及统计分析和建模中的缺陷造成的伪像。

Jianguo WU, Jones KB, Harbin LI, et al. SCALING AND UNCERTAINTY ANALYSIS IN ECOLOGY. 2007.

- 由于地球表面空间作为一个巨系统的复杂性，在某一尺度上人们观察到的性质、总结出的原理或规律，在另一尺度上可能有效、可能相似，也可能需要修正。

李小文, 王祎婷. 定量遥感尺度效应刍议. 2013.

1 地理要素尺度、分区效应与可变面积单元问题

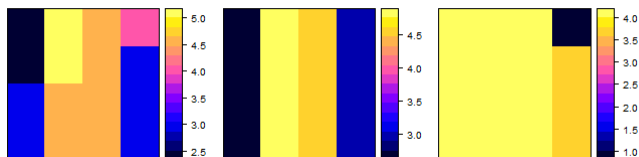


Mean Value: 3.75, 3.75, 3.75

Variance: 2.6, 0.5, 0

1 地理要素尺度、分区效应与可变面积单元问题

- 分区效应描述了同一尺度下由于数据组合和配置的不同所造成的相关统计结果的差异性，即区划方案所导致的结果偏差。



Mean Value: 3.75, 3.75, 3.17 Variance: 0.93, 1.04, 2.11

李海萍. 空间统计分析中的MAUP及其影响. 2009.

Dennis E. Jelinski, Jianguo Wu . The modifiable areal unit problem and implications for landscape ecolog. 1996.

1 地理要素尺度、分区效应与可变面积单元问题

- 定义a: 可变面积单元问题(Modified Areal Unit Problem, MAUP)——在将数据汇总到不同尺度的面状单元的过程中, 数据细节变化会损失, 面状数据的统计分析结果会随着面状数据的空间单元大小和区域空间单元划分方法不同而不同。
- 定义b: 由于对连续的地理现象的空间单元的人为划分而产生空间模式的变化所引起的问题。
- 可变面积单元问题效应包括尺度效应和分区效应

陈江平, 张瑶, 余远剑. 空间自相关的可塑性面积单元问题效应. 2011.

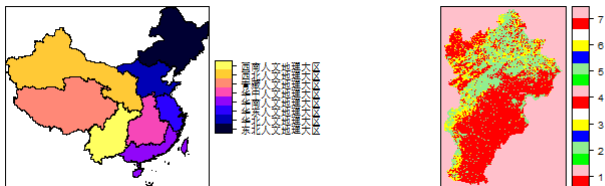
齐丽丽, 柏延臣. 社会经济统计数据热点探测的MAUP效应. 2012.

Openshaw S, Taylor P. A million or so correlation coefficients: Three experiments on the modifiable area unit problem. 1979.

Openshaw S. Concepts and Techniques in Modern Geography. 1984.

2 空间分层异质性

空间分层异质性(Spatial Stratified Heterogeneity)是一种广泛存在的地理现象(层内方差小于层间方差)。



除了空间自相关性之外的另一种地理要素空间分布规律。测度? 意义?

方创琳, 刘海猛, 罗奎, 于晓华. 中国人文地理综合区划. 2017.

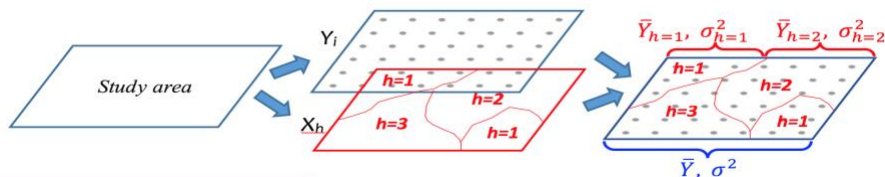
Yong Ge, Yu Jiang, Yuehong Chen, et al. Designing an Experiment to Investigate Subpixel Mapping as an Alternative Method to Obtain Land Use/Land Cover Maps. 2016.

3 地理探测器原理

3.2 核心思想

地理探测器是探测空间分异性，以及揭示其背后驱动力的一组统计学方法。其核心思想是基于这样的假设：如果某个自变量对某个因变量有重要影响，那么自变量和因变量的空间分布应该具有相似性。**地理探测器=空间叠加+统计检验**

- 作用1: 探测某地理/生态要素的空间分异性 Y
- 作用2: 揭示某地理/生态要素的背后驱动力 Y~X



王劲峰, 徐成东. 地理探测器原理及展望. 2017.

3 地理探测器原理

3.3 关键公式

分异及因子探测：探测Y的空间分异性以及探测某因子X多大程度上解释了属性Y的空间分异。用q值度量，表达式为：

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} = 1 - \frac{SSW}{SST}$$

$$SSW = \sum_{h=1}^L N_h \sigma_h^2, SST = N \sigma^2$$

q的值域为[0, 1]，值越大说明Y的空间分异性越明显；如果分层是由自变量X生成的，则q值越大表示自变量X对属性Y的解释力越强，反之则越弱。

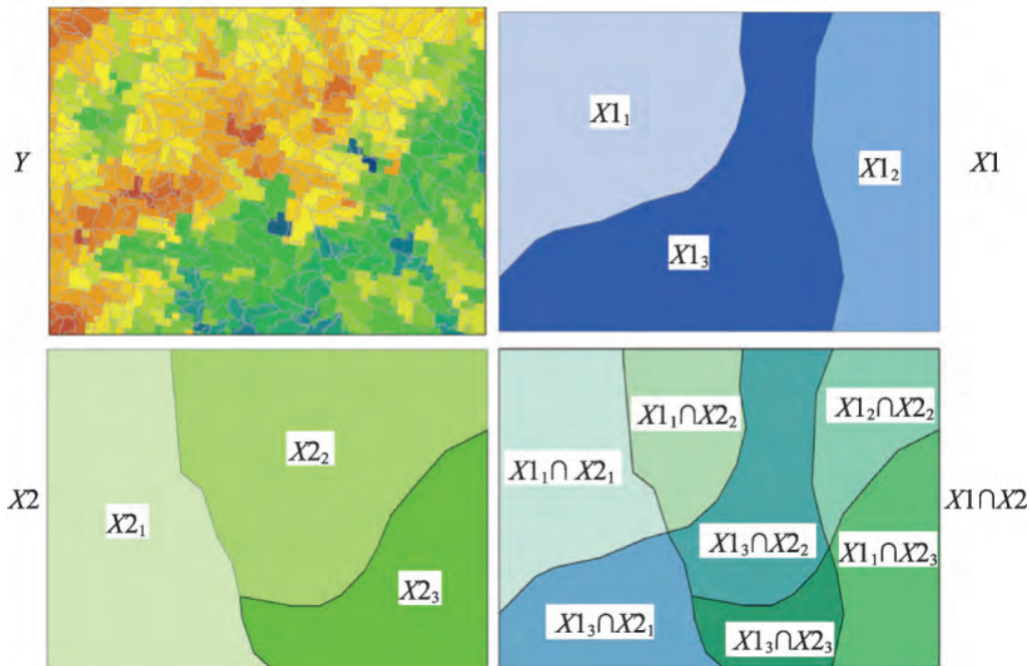
3 地理探测器原理

3.3 关键公式

交互作用探测：识别不同风险因子 X_s 之间的交互作用，即评估因子 X_1 和 X_2 共同作用时是否会增加或减弱对因变量 Y 的解释力，或这些因子对 Y 的影响是相互独立的。评估的方法是首先分别计算两种因子 X_1 和 X_2 对 Y 的 q 值： $q(X_1)$ 和 $q(X_2)$ ，并且计算它们交互（叠加变量 X_1 和 X_2 两个图层相切所形成的新的多边形分布，图2）时的 q 值： $q(X_1 \cap X_2)$ ，并对 $q(X_1)$ 、 $q(X_2)$ 与 $q(X_1 \cap X_2)$ 进行比较。

3 地理探测器原理

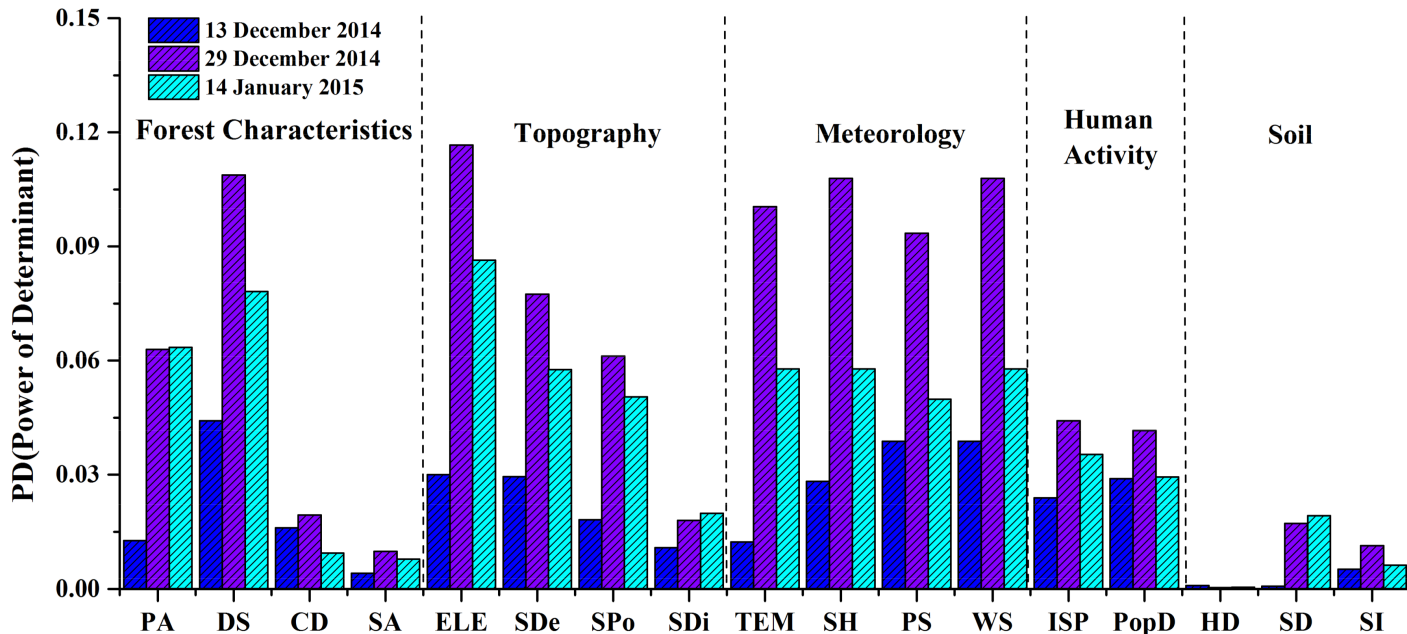
3.3 关键公式



3 地理探测器

3.4 应用

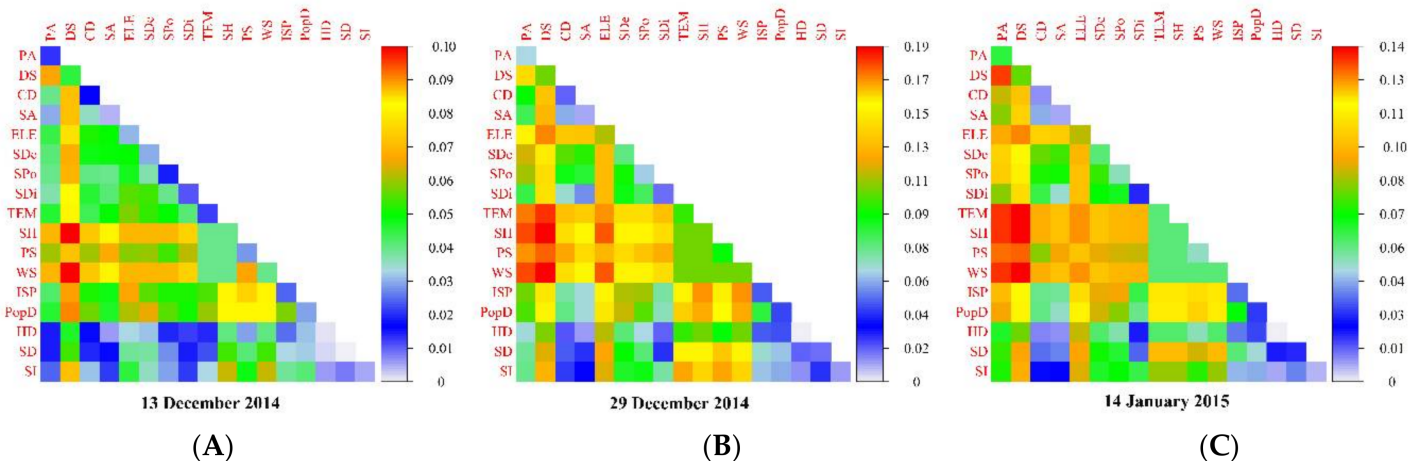
关于影响森林区域PM2.5的生态因子的研究。



3 地理探测器

3.4 应用

交互作用探测:



GuoLiang Yun, Shudi Zuo, Shaoqing Dai, et al. Individual and Interactive Influences of Anthropogenic and Ecological Factors on Forest PM_{2.5} Concentrations at an Urban Scale. 2018.

4 模型不确定性的定性分析

空间数据的不确定性与误差。

1. 计算误差

2. 观测的误差

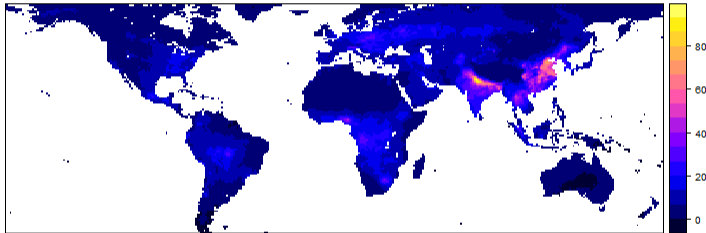
3. 采样的误差

4. 模型算法的误差

5. 随机误差

相同数据，不同软件之间结果是否可以重现？

4 模型不确定性的定性分析



4 模型不确定性的定性分析

```
sw<-matrix(c(1,1,1,1,0,1,1,1,1),nrow=3)  
Moran(pm25degree, w = sw)
```

```
## [1] 0.9674364
```

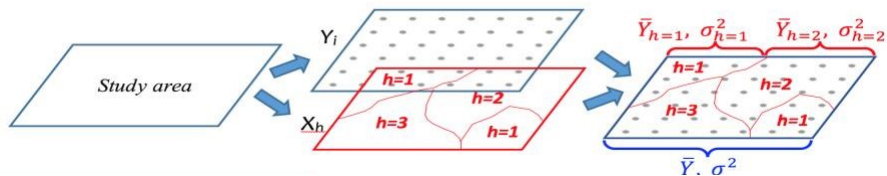
```
sw<-matrix(c(1,2,1,2,0,2,1,2,1),nrow=3)  
Moran(pm25degree, w = sw)
```

```
## [1] 0.9676104
```

```
sw<-matrix(c(1,3,1,3,0,3,1,3,1),nrow=3)  
Moran(pm25degree, w = sw)
```

```
## [1] 0.9676973
```

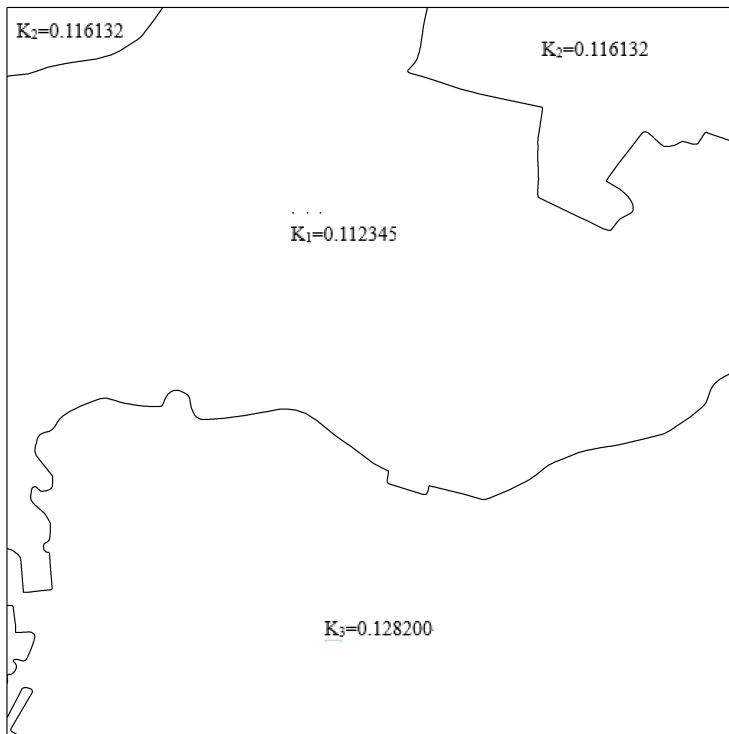
4 模型不确定性的定性分析



y~格网数据，x~分类型变量。空间叠加时必然会有MAUP的问题，而这个问题导致的地理探测器结果存在一定的不确定性。

X的分区数量？如何分区？

4 模型不确定性的定性分析



由于不同数据的尺度差异，势必要对Y值进行聚合，此时不同多边形的Y值如何取，按面积均分或者都取Y值？

5 案例研究

目标

MAUP效应造成的地理探测器结果(1.空间分异性探测; 2.驱动力探测)的不确定性，主要针对因子探测模块。

软件

由于地理探测器原本是基于Excel2003开发的，在数据读取上存在一定限制。这里我开发了一个R语言版本的地理探测器。当然由于本人拖延症懒癌晚期，至今仍未发布。而王劲峰老师团队也在近期更新了一个R包（[GD](#)）。当然我当时做R包的时候不仅仅是针对实现地理探测器的功能，会有一些其他附加的功能，后面仍会发布（欢迎发布后引用和使用）。

关于数据：

PM2.5数据：[Atmospheric Composition Analysis Group](#)

土地利用数据：[土地利用混合像元多尺度分解实验数据集](#)

5 案例研究

5.1 尺度效应造成空间分异性探测的不确定性

```
library(raster)
library(sp)
source('F:/R/GeographicalDetectorR/Factordetector.R')
```

```
pm25degreejjj <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/pm25degreepjjj.tif")
pm25kmjjj <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/pm25kmpjjj.tif")
pm25kmjjjagmean <- aggregate(pm25kmjjj, fact = 10, fun = mean)
pm25kmjjjagmedian <- aggregate(pm25kmjjj, fact = 10, fun = median)
```

5 案例研究

5.1 尺度效应造成空间分异性探测的不确定性

```
## q value: class 0.8433877 p value: 4.113783e-10
##   factor      q      p
## 1 class 0.8433877 4.113783e-10
## q value: class 0.8468863 p value: 5.489537e-10
##   factor      q      p
## 1 class 0.8468863 5.489537e-10
## q value: class 0.8456691 p value: 5.42443e-10
##   factor      q      p
## 1 class 0.8456691 5.42443e-10
## q value: class 0.8440829 p value: 5.98237e-10
##   factor      q      p
## 1 class 0.8440829 5.98237e-10
```

5 案例研究

5.2 分区效应造成空间分异性探测的不确定性

```
pm25degreejjjt$class2 <- ifelse(pm25degreejjjt$pm25degreepjjjj >= 250, 1,  
                                ifelse(pm25degreejjjt$pm25degreepjjjj >= 150, 2,  
                                        ifelse(pm25degreejjjt$pm25degreepjjjj >= 55, 3,  
                                              ifelse(pm25degreejjjt$pm25degreepjjjj >= 35, 4,  
                                                    ifelse(pm25degreejjjt$pm25degreepjjjj >= 12, 5, 6))))))
```

```
pm25degreejjjt$class3 <- kmeans(pm25degreejjjt$pm25degreepjjjj, centers = 6, nstart = 20)$cluster  
pm25degreejjjt$class4 <- kmeans(pm25degreejjjt$pm25degreepjjjj, centers = 7, nstart = 20)$cluster  
pm25degreejjjt$class5 <- kmeans(pm25degreejjjt$pm25degreepjjjj, centers = 8, nstart = 20)$cluster  
pm25degreejjjt$class6 <- kmeans(pm25degreejjjt$pm25degreepjjjj, centers = 9, nstart = 20)$cluster
```


5 案例研究

5.2 分区效应造成空间分异性探测的不确定性

Factordetector(pm25degreejjjt)

```
## q Value: class 0.8433877 p Value: 4.113783e-10
## q Value: class2 0.9389284 p Value: 1.232152e-10
## q Value: class3 0.978673 p Value: 8.316727e-10
## q Value: class4 0.985093 p Value: 3.169325e-10
## q Value: class5 0.9882919 p Value: 2.610105e-10
## q Value: class6 0.9904176 p Value: 2.960121e-10
```

```
## factor      q      p
## 1 class 0.8433877 4.113783e-10
## 2 class2 0.9389284 1.232152e-10
## 3 class3 0.9786730 8.316727e-10
## 4 class4 0.9850930 3.169325e-10
## 5 class5 0.9882919 2.610105e-10
## 6 class6 0.9904176 2.960121e-10
```

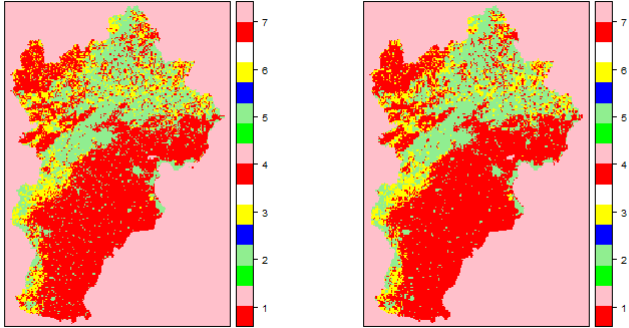
5 案例研究

5.3 尺度效应造成驱动力探测的不确定性

```
landuset1km <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/L1_tif/L1_1000.tif")
landuset500m <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/L1_tif/L1_500.tif")
landuset250m <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/L1_tif/L1_250.tif")
landuset100m <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/L1_tif/L1_100.tif")
```

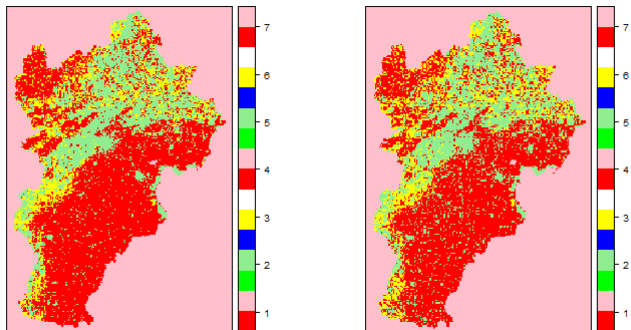
5 案例研究

5.3 尺度效应造成驱动力探测的不确定性



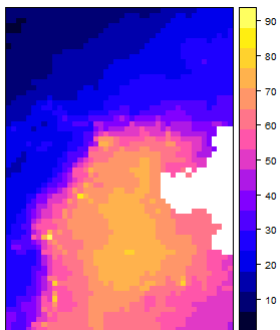
5 案例研究

5.3 尺度效应造成驱动力探测的不确定性



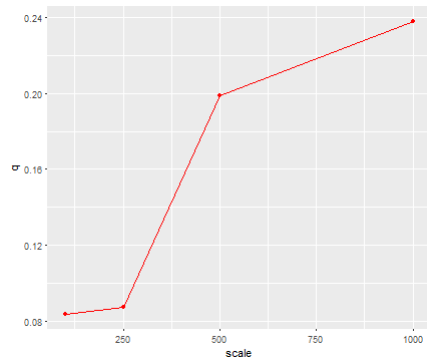
5 案例研究

5.3 尺度效应造成驱动力探测的不确定性



5 案例研究

5.3 尺度效应造成驱动力探测的不确定性



5 案例研究

5.3 尺度效应造成驱动力探测的不确定性

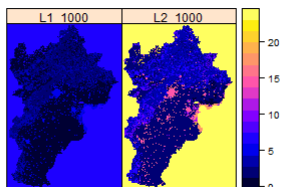
scaleqstat

```
##      factor      q      p  scale
## 1 landuse 0.08353797 1.303040e-10 100
## 2 landuse 0.08729653 2.275076e-10 250
## 3 landuse 0.19868498 8.245900e-10 500
## 4 landuse 0.23824758 3.188562e-10 1000
```

5 案例研究

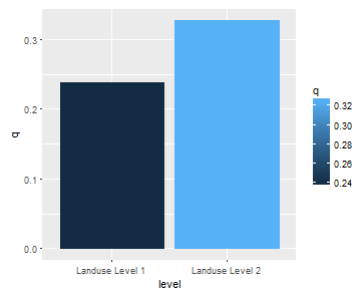
5.4 分区效应造成驱动力探测的不确定性

```
landuset1km <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/L1_tif/L1_1000.tif")  
landuset2km <- raster("F:/DataScience/COSR2018/GeographicalDetectorMAUP/data/L2_tif/L2_1000.tif")
```



5 案例研究

5.4 分区效应造成驱动力探测的不确定性



5 案例研究

5.4 分区效应造成驱动力探测的不确定性

unzone

```
##      factor      q      p      level
## 1 landuse 0.2382476 3.188562e-10 Landuse Level 1
## 2 landuse 0.3268270 4.504972e-10 Landuse Level 2
```

6 挑战与未来发展

- 1.两面性：MAUP既造成了地理探测器结果的不确定性，反过来，地理探测器也可以作为一个探究MAUP效应的有力工具。
- 2.普适性的时空尺度转换方法研究框架(李小文院士)。
- 3.强时空异质性要素的分析框架(Urban-Air的启示)。

谢谢

