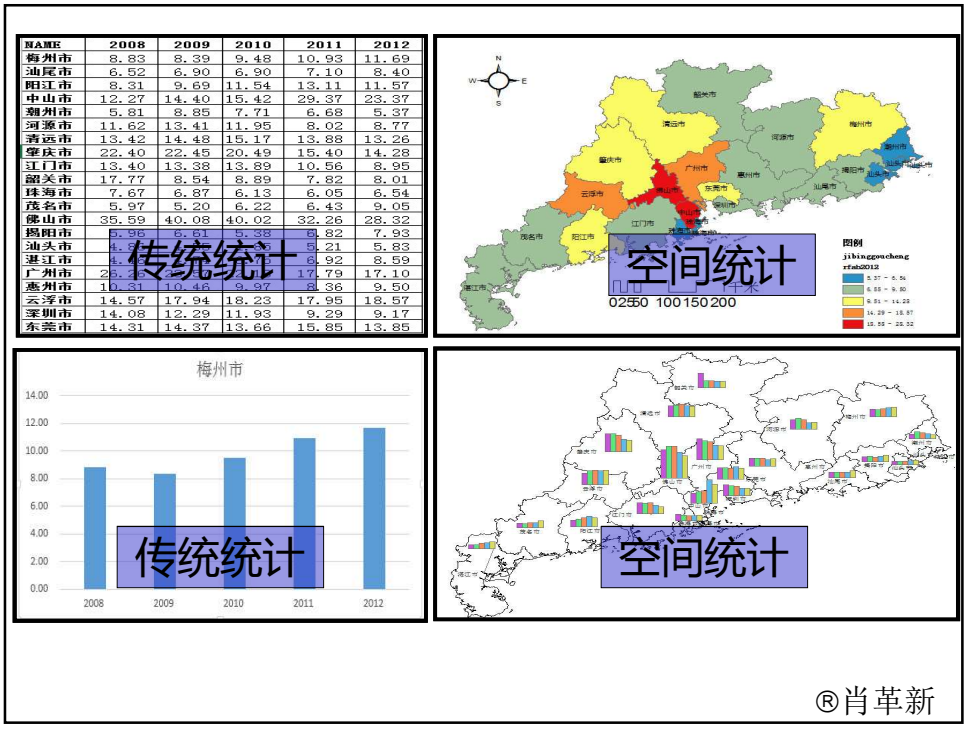
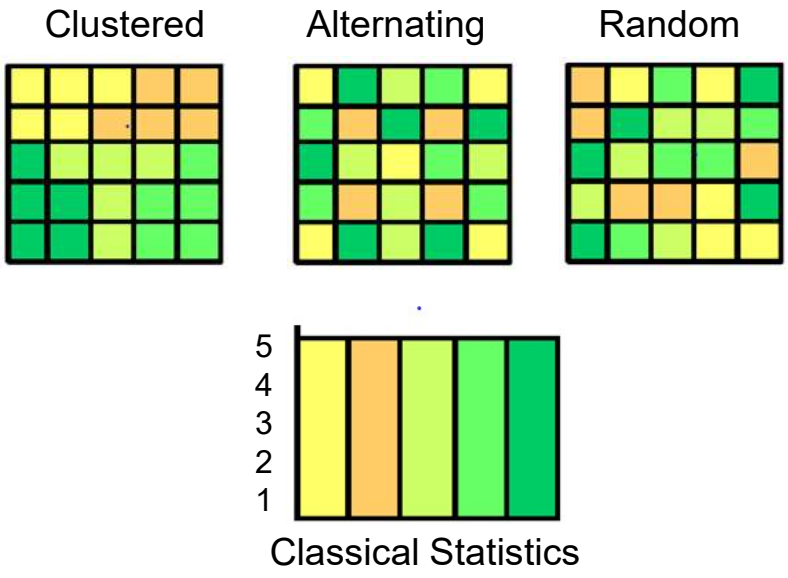
中国地理信息产业协会教育与科普工作委员会
2019年全国高校青年教师GIS教学研修班（2019年7月2-7日）
地点：南京，河海大学国际交流中心(隽恒酒店)，南京市西康路1号
时间：2019年7月3日(星期三) 下午14:00-15:30

空间数据三大特性与空间统计学
Spatial Properties & Spatial Statistics

王劲峰
wangjf@Lreis.ac.cn
资源与环境信息系统国家重点实验室
中国科学院地理科学与资源研究所
2019-7-3（星期三）

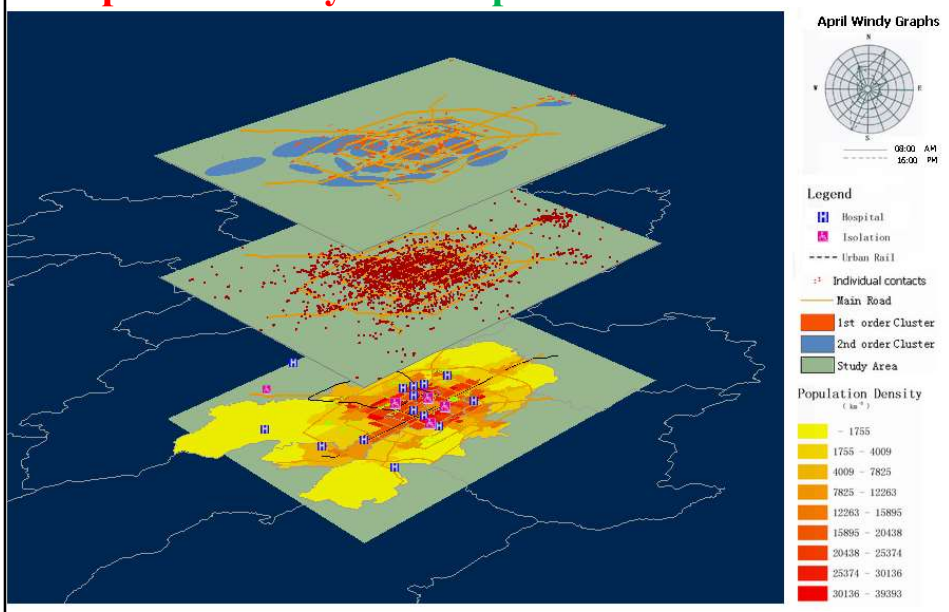


Spatial Pattern → Classical iid



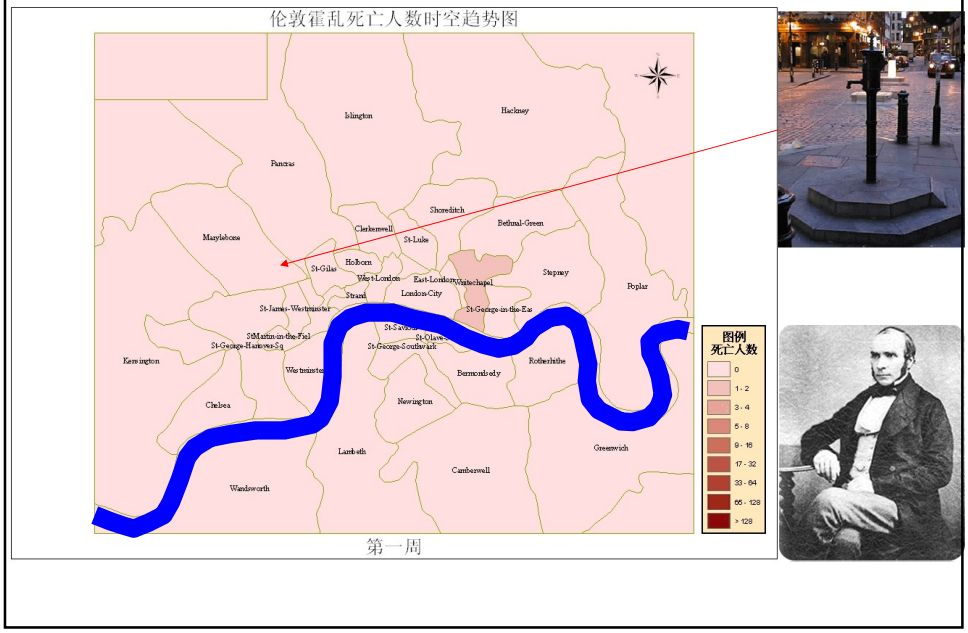
Haining. Issues in analysis spatial data, Chapter 2 Haining. SIE 512 Lecture 4. September 11, 2003

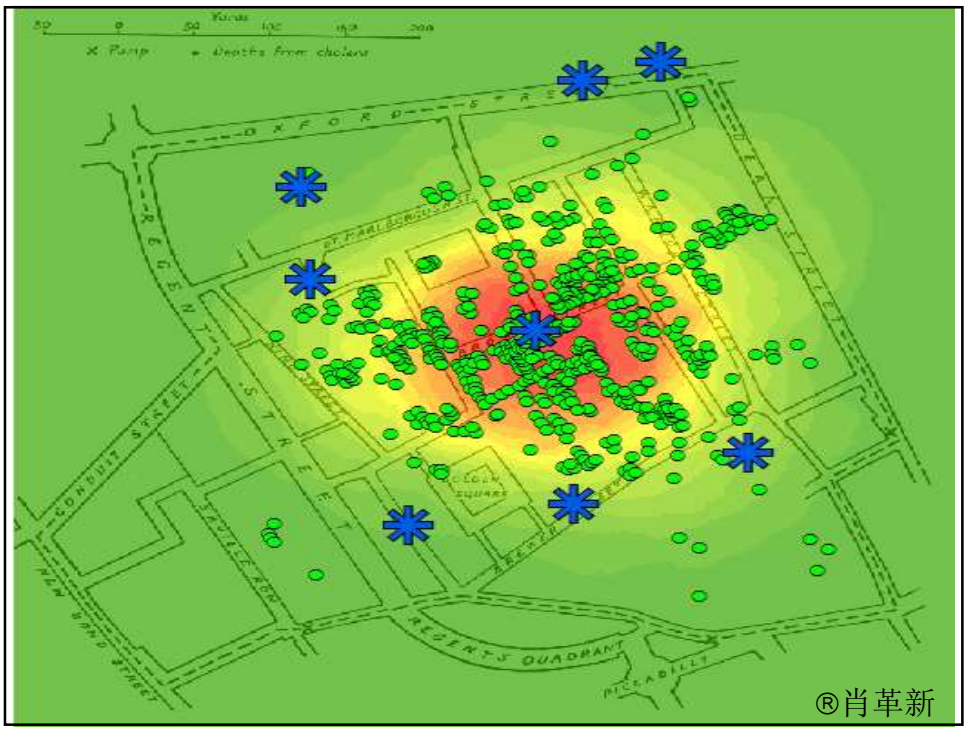
Spatial Seemly iid → Spatial Pattern



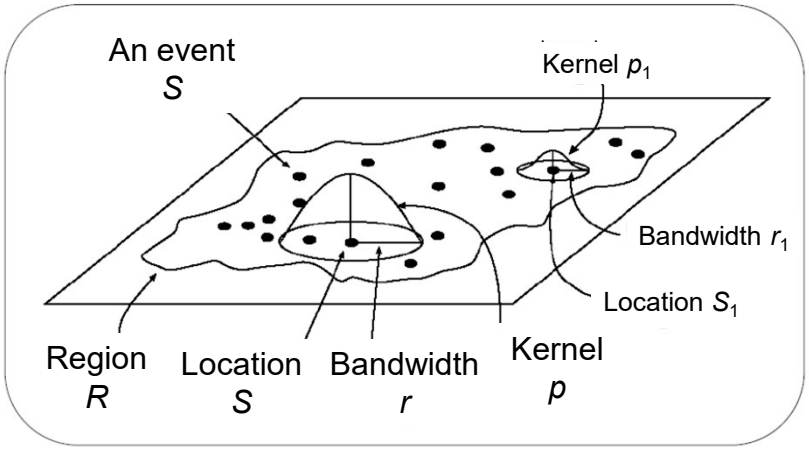
Wang JF, McMichael AJ, Meng B, *et al*. 2006. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. **Bulletin of the World Health Organization** 84(12): 965-968

Spatial Causation





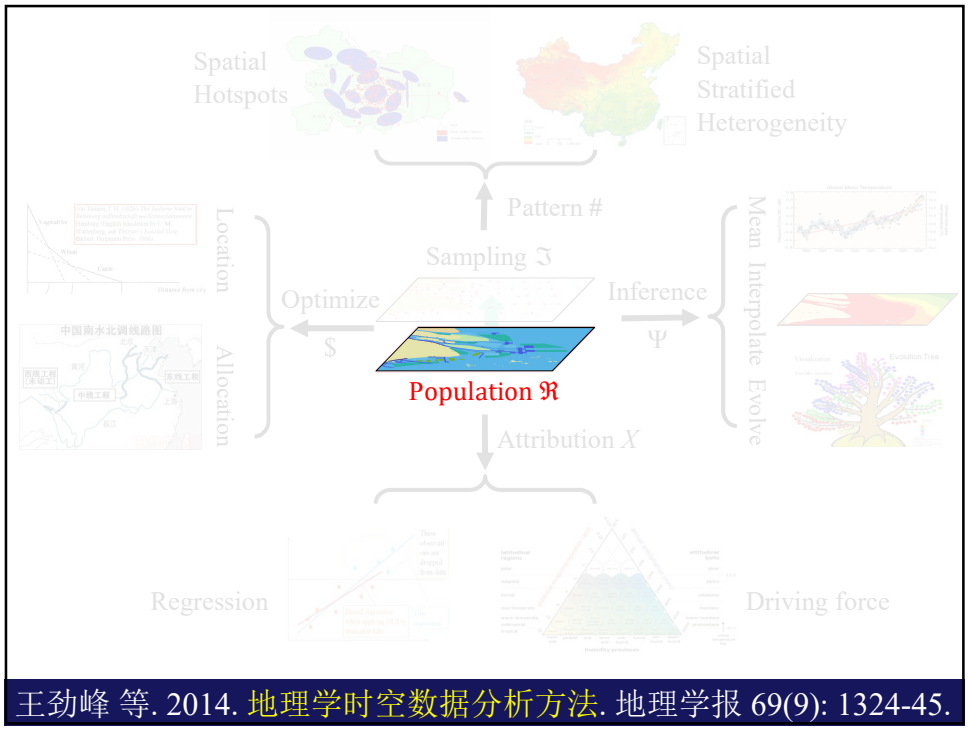
Kernel Density Estimation



$$\hat{\lambda}_p = \frac{\#[S \in C(p, r)]}{\pi r^2}$$

If *r* is too large or too small, then …… ?

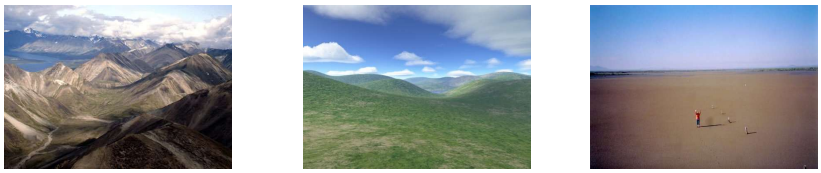
Classic Statistics iid (Independent + Identical Distribution)	Spatial population
Simple random (Ross, p52) Let $X_1, \ldots, X_n$ be rvs, and $\bar{X} = \sum_{i=1}^{n} X_i/n$ If the rvs are iid, then $E\bar{X} = \mu$; $V\bar{X} = \sigma^2$	
Linear regression (Gujiaradi p93-97) Let $Y_{ik} = \beta_1 + \beta_2 X_i + u_{ik}$ If $E(u_{ik}\|X_i) = 0$, $V(u_{ik}\|X_i) = \sigma^2$, $C(u_{ik}, u_{jk}\|X_i, X_j) = 0$, + → iid then OLS $\hat{\beta}$ is BLUE of β	
χ^2 pdf (FD2, p26) Let $X_1, \ldots, X_n$ be iid ~ $N(0, 1)$ then $\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2$ **pdf**	
Poisson Process (FD1, p100-101) If (1) $p(s) = p$; (2) $p[r(s)] = p(r)$; (3) $p(s\|t) = p(s)$, → iid then $p_k(r) = \frac{(\lambda r)^k}{k!} e^{-\lambda r}$	



王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.

Spatial Autocorrelation

The coincidence of locational and attribute similarity
(Anselin, L. 1988. Spatial econometrics. Methods and models. Dordrecht: Kluwer)



low high maximum

- Human: movement
- Disease: contact, airborne, waterborne;
- Economy: movement of goods, labor;
- Atmosphere: diffusion, transfer, radiation;
- Payment for ecosystem services (PES) (in China named as “eco-compensation”

王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 科学出版社.

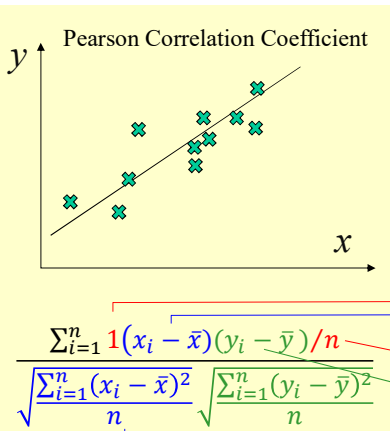
Spatial Autocorrelation: Consequences

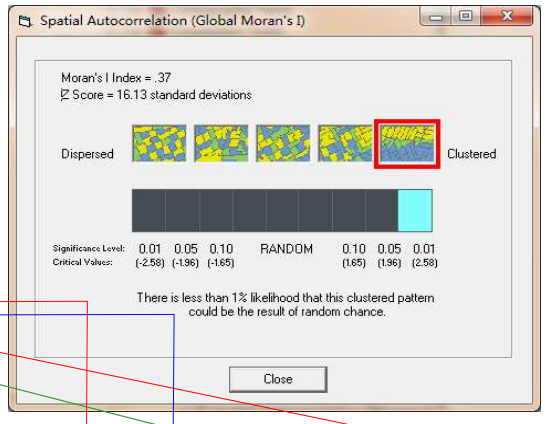
iid population → autocorrelated population

- Population: iid → autocorrelated: Moran I, G
- Regression: CLRM → SLRM
- Mean: simple ave Σy/n → Bkriging Σwy
- Mapping: complete sample → Kriging
- Sampling: *n*(*iid*)↘*n*(ACR); *v*(*iid*)↘*v*(ACR)
- Cause other ACR: air pollution → lung cancer

王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 科学出版社.

Spatial Autocorrelation: Moran's I





$$\frac{\sum_{i=1}^{n} 1(x_i-\bar{x})(y_i-\bar{y})/n}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}}}$$

$$\frac{\sum_{i=1}^{n} w_{ij}(x_i-\bar{x})(x_j-\bar{x})/\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}}{\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}}$$

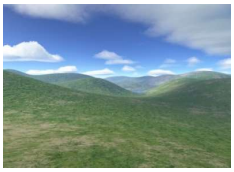
© wangjf@Lreis.ac.cn

Moran PAP. 1950. Notes on continuous stochastic phenomena. Biometrika 37: 17-23

Spatial Autocorrelation: Semivariogram

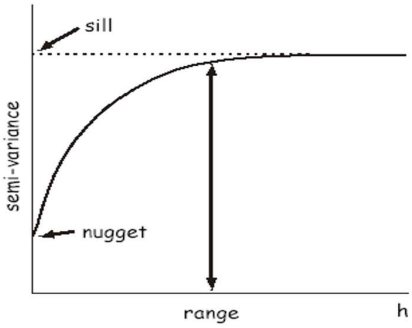
Low

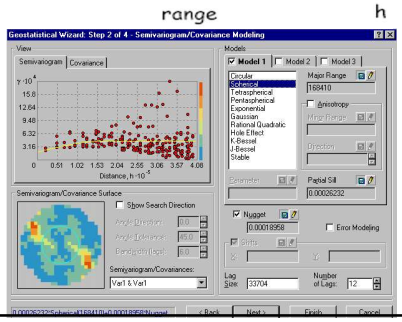
High



Maximum

Longley et al., Ch 14,15

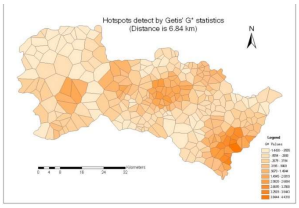




Spatial Heterogeneity: Phenomena

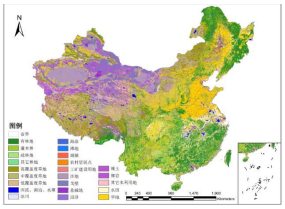
The variation which exceeds random (Haining 2017.9.28 CAS)

Hotspots

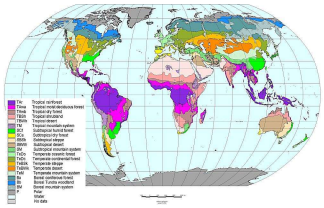


Spatial Local Heterogeneity

Classification



Regionalization



Spatial Stratified Heterogeneity (SSH)

- Nature: climate zone, landcover, watershed, plates
- Human: states, classes, labor division, urban function zones
- Biology: niche, footprint, fauna
- Culture: religions, behavior, food
- Miscellanies: east, middle and east

Wang JF, Zhang TL, Fu BJ. 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 7: 250-256.

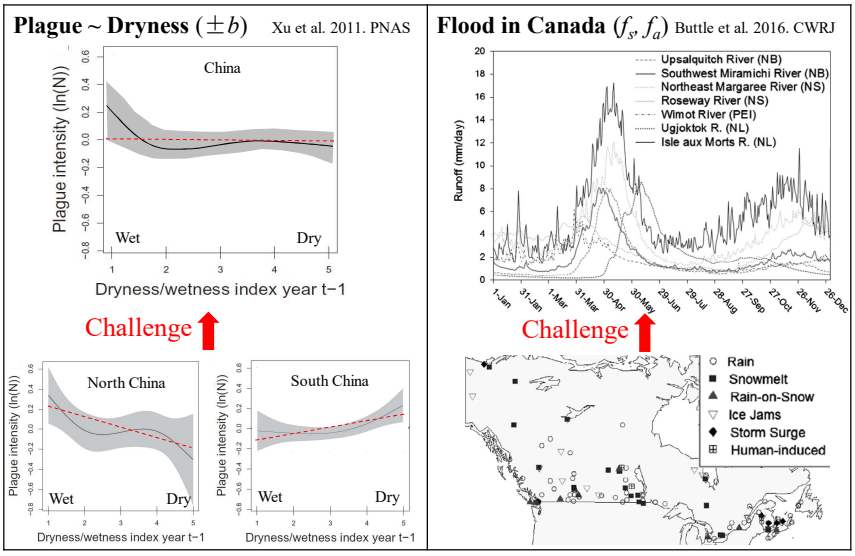
Spatial Heterogeneity: Consequences

Population → Local, Strata

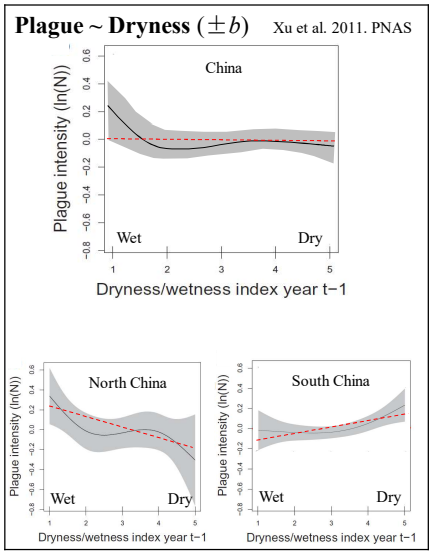
- Spatial Relate: Moran I → LISA, G → Gi
- Spatial stratified heterogeneity: info; confound
- Regression: Spatial regres → GWR, BHM
- Total: Bkriging → MSN, B-shade, SPA
- Mapping: Kriging → Sandwich
- Sampling: Sample is biased when $n < L$
- Cause other heterogeneity

王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 科学出版社.

SSH → Confounding



wangjf@Lreis.ac.cn

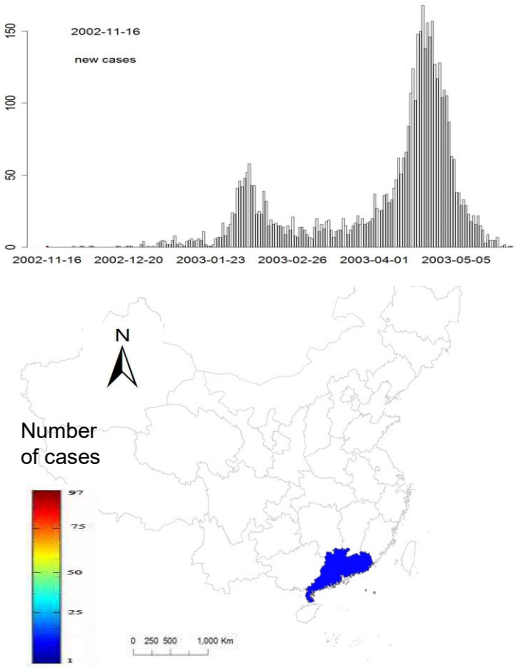


Insignificance

SSH

Significance

wangjf@Lreis.ac.cn



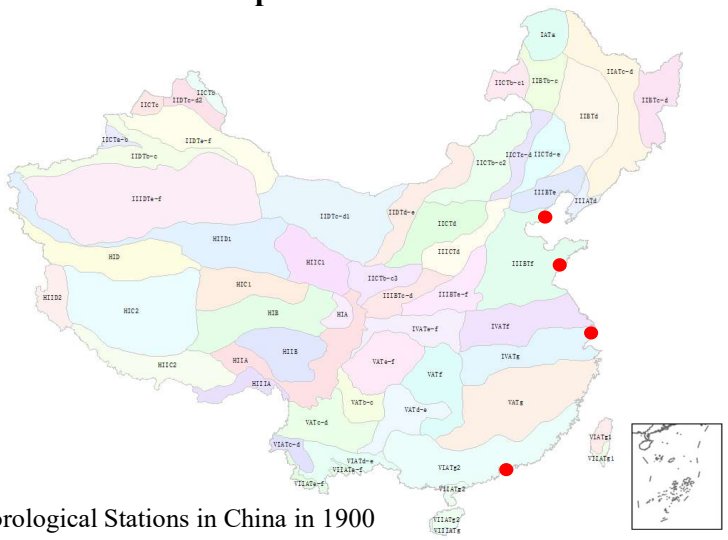
Complex

SSH

Simple

SSH → Sample Bias

Annual Temperature in China in 1900 ?



Meteorological Stations in China in 1900
Climate zones

wangjf@Lreis.ac.cn

Spatial Local Heterogeneity: Measures

Gi, LISA, SatScan

$$G_i(d) = \frac{\sum_{j\neq i}^{n} w_{ij}(d)x_j}{\sum_{j\neq i}^{n} x_j}$$

Getis A, Ord JK. 1992. The analysis of spatial association by use of distance statistics. Geographical Analysis 24(3): 189-206.

$$I_i = z_i \Sigma_j w_{ij} z_j, \quad z_i = \frac{x_i - \bar{x}}{SD_i},$$

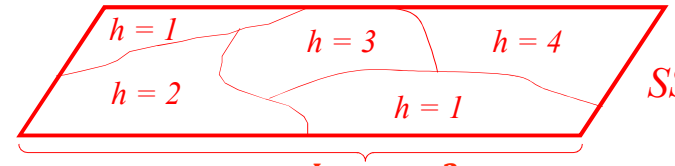
Q2(X- , WX+): LH Q1(X+,WX+):HH
Q3(X- , WX-): LL Q4(X+, WX-): HL

Anselin L. 1995. Local indicators of spatial association – LISA. Geographical Analysis 27(2): 93-115

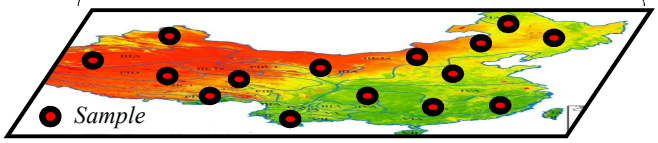
$$\mathrm{LR} = [c/\mu]^c \times [(C-c)/(C-\mu)]^{C-c}$$

Kulldorff M. 1997. A spatial scan statistic. Communications in Statistics – Theory and Methodology 26(6): 1481-1496

Spatial Stratified Heterogeneity (SSH): Measure



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

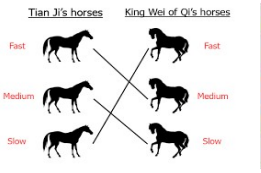


$q \in [0, 1]$, Y has $100q\%$ degree of SSH
0, if Y has no SSH
1, if Y is fully SSH

$$\frac{N-L}{L-1}\frac{q}{1-q} \sim F(L-1, N-L; \lambda)$$

Wang JF, Zhang TL, Fu BJ. 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 7: 250-256.

MAUP



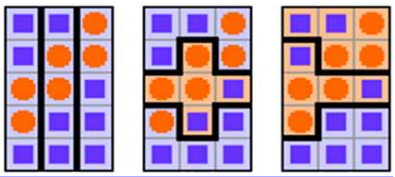


Table 3. Individual and ecological correlations (after Robinson, 1950)

level of aggregation	number of units	correlations between: negroandilliteracy	nativity and illiteracy
individual	98 million	.203	.118
state	48	.773	-.526
census division	9	.946	-.619

Gerrymandering
http://en.wikipedia.org/wiki/Gerrymandering

Mean, Variance, Correlation (*r*), Autocorrelation (*I*): different results upon
(1) at different levels of spatial resolution (The scale effect)
(2) regrouping zones at a given scale (The zoning effect)

Solutions (Fotheringham et al. 2000. Quantitative Geography: Perspectives on Spatial Data Analysis. London: SAGE. P.238)
(1) Using spatially disaggregrated data;
(2) Report the results at the most spatially disaggregated level with sensitive results
(3) Construction of the optimal zoning system;
(4) Add grouping variables, relate individual and aggregate levels.

Openshaw S. 1984. The Modifiable Areal Unit Problem. CATMOG 38. Norwich: GeoAbstracts

References for the Properties of Spatial Population

Anselin L. 1988. **Spatial Econometrics: Methods and Models**. Kluwer Academic Publishers.

Cressie N. 1992. **Statistics for Spatial Data**. Blackwell Publishing Ltd

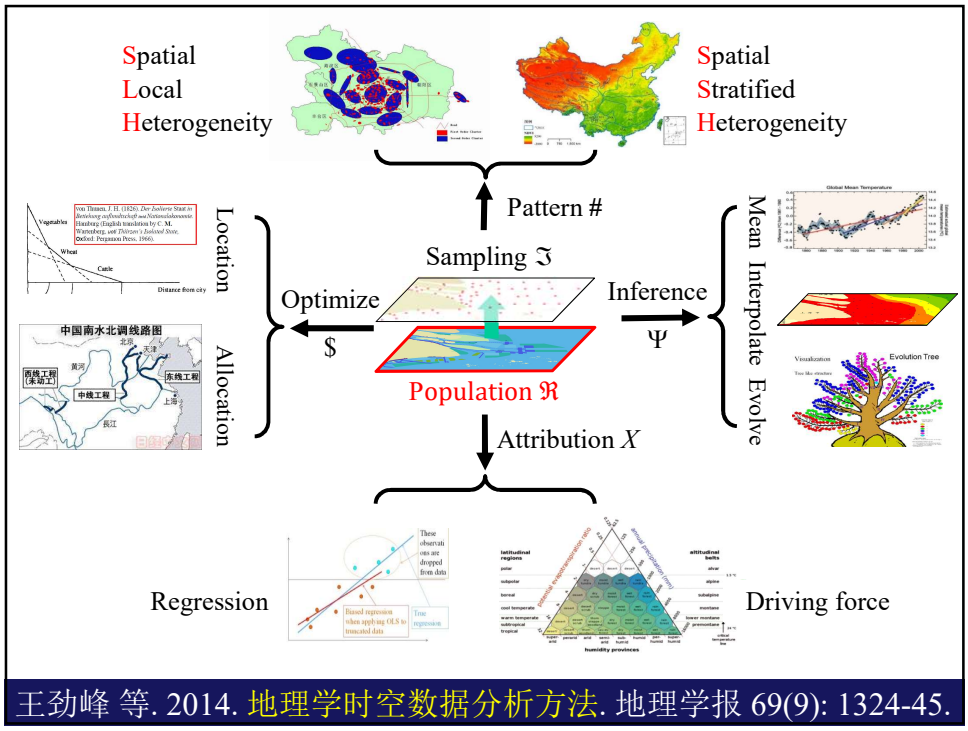
Fischer M, Wang JF. 2010. **Spatial Data Analysis: Models, Methods and Techniques**. Springer

Fotheringham A, Brunsdon C, Charlton M. 2000. **Quantitative Geography**. Sage

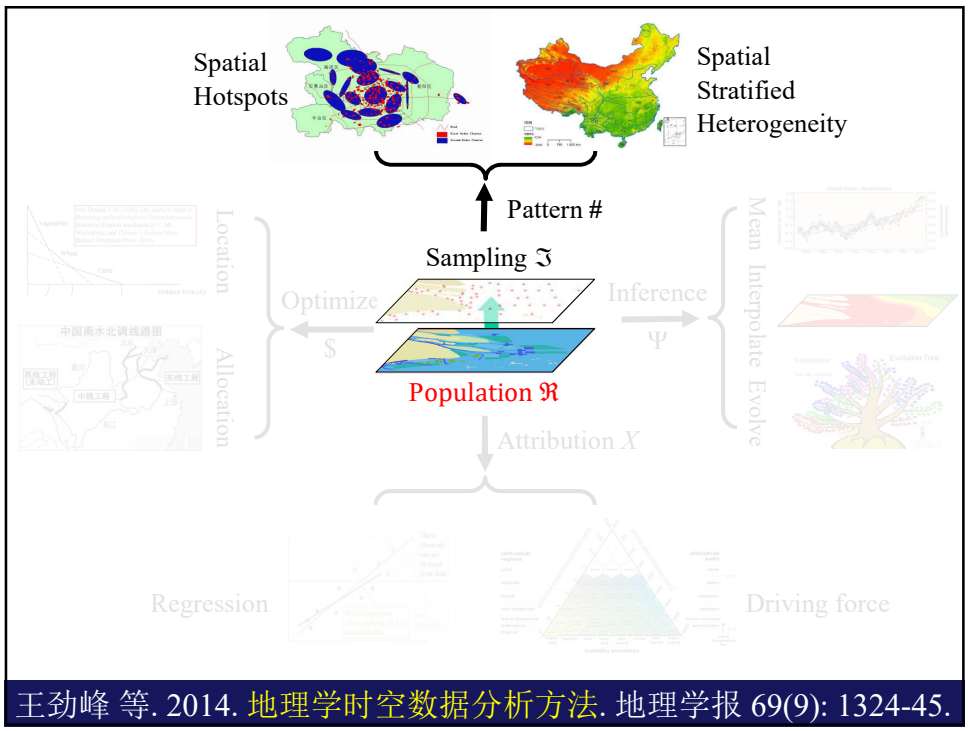
Haining R. 2003. **Spatial Data Analysis: Theory and Practice**. Cambridge University Press

王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 科学出版社

Classic Statistics iid (Independent + Identical Distribution)	Spatial population SAC + SSH
Simple random (Ross, p52) Let $X_1, \ldots, X_n$ be rvs, and $\bar{X} = \sum_{i=1}^{n} X_i/n$ If the rvs are iid, then $E\bar{X} = \mu$; $V\bar{X} = \sigma^2$	Flow: human, materials, money, information, etc. Spatial Autocorrelation (SAC) → Not Independent
Linear regression (Gujiaradi p93-97) Let $Y_{ik} = \beta_1 + \beta_2 X_i + u_{ik}$ If $E(u_{ik}\|X_i) = 0$, $V(u_{ik}\|X_i) = \sigma^2$, $C(u_{ik}, u_{jk}\|X_i, X_j) = 0$, + → iid then OLS $\hat{\beta}$ is BLUE of β	
χ^2 pdf (FD2, p26) Let $X_1, \ldots, X_n$ be iid ~ $N(0, 1)$ then $\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2$ **pdf**	Landuse types; climate zones, Adm border, etc. Spatial Stratified Heterogeneity (SSH) → Not Identical Distribution
Poisson Process (FD1, p100-101) If (1) $p(s) = p$; (2) $p[r(s)] = p(r)$; (3) $p(s\|t) = p(s)$, → iid then $p_k(r) = \frac{(\lambda r)^k}{k!} e^{-\lambda r}$	



王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.



王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.

点格局（Points' pattern）：登革热病例空间聚集吗？

- 样方分析
- 最近邻居
- 层次聚集
- Ripley K

→ 几何：聚集、分散、随机？

热点（Hotspot）：手足口病发病率空间聚集吗？

- Gi
- LISA
- SatScan

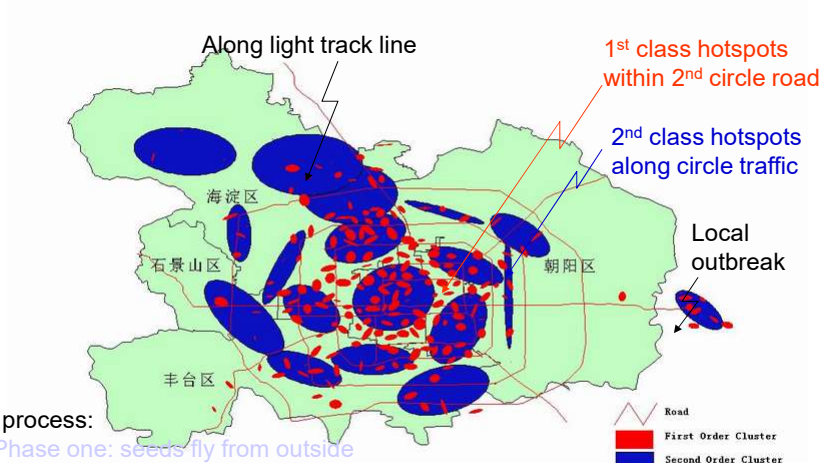
→ 属性|几何：热点、聚集？

分异（Stratified heterogeneity）：肝癌率城乡差异吗？

- Geodetector *q*-statistic：层间差异（属性|几何；几何|属性）

王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 科学出版社.

2003 SARS Transmission in Beijing

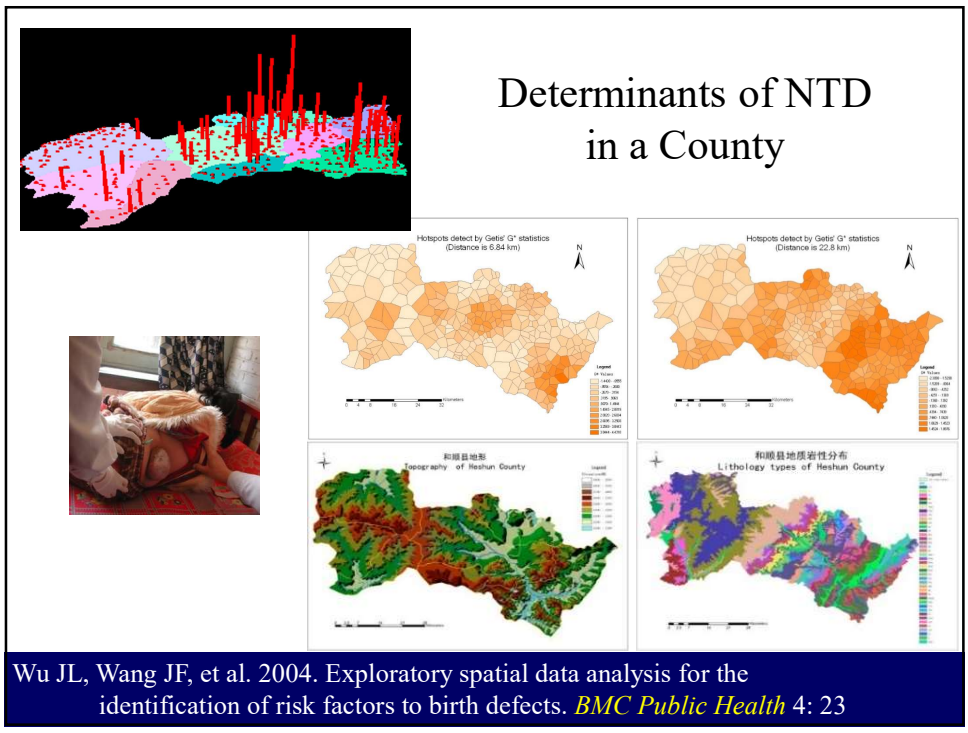


Spatial process:

- Phase one: seeds fly from outside
- Phase two: spread within city randomly
- Phase three: cluster in space

Wang JF, *et al*. 2006. Spatial dynamics of an epidemic of severe acute respiratory syndrome in an urban area. **Bulletin of the World Health Organization** 84(12): 965-968.


NTD prevalence in China


Determinants of NTD
in a County
Hotspots detect by Getis' G* statistics
(Distance is 6.84 km)
Hotspots detect by Getis' G* statistics
(Distance is 22.8 km)
和顺县地形
Topography of Heshun County
和顺县地质岩性分布
Lithology types of Heshun County
Wu JL, Wang JF, et al. 2004. Exploratory spatial data analysis for the identification of risk factors to birth defects. *BMC Public Health* 4: 23

Hotspots: SatScan

A window at a site

– c is # cases in a window, C=# cases in whole area

μ is the expected # cases in the window = $n(C/N)$

– LR = $[c/\mu]^c \times [(C-c)/(C-\mu)]^{C-c}$

Windows scan at the site

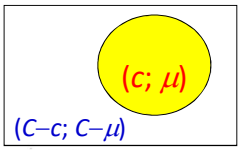
– The highest LR window is Most Likely Cluster

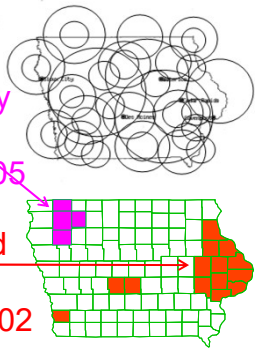
Reference

– 999 spatial random distributions of C: LR(rand)

– Rank the LR(real) in the 999 LR(rand): p

Rank	LR
1/1000	15.22
2/1000	14.34
3/1000	14.01
4/1000	13.88
5/1000	13.15
6/1000	13.02
7/1000	11.34
…	
101/1000	9.27
102/1000	8.91
…	
1000/1000	2.79



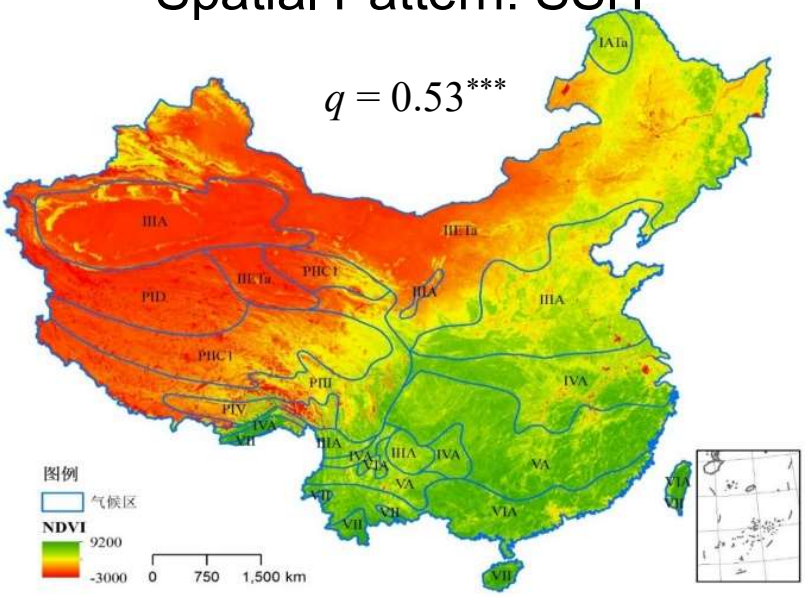


Primary cluster
p =0.005

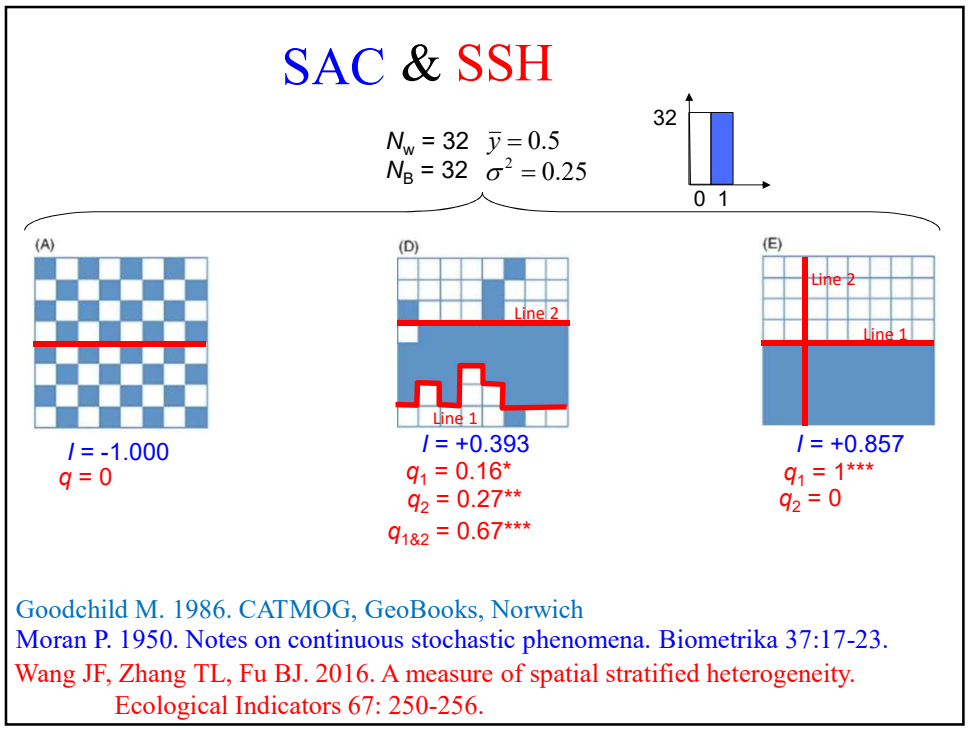
Second cluster
p = 0.102

Spatial Pattern: SSH

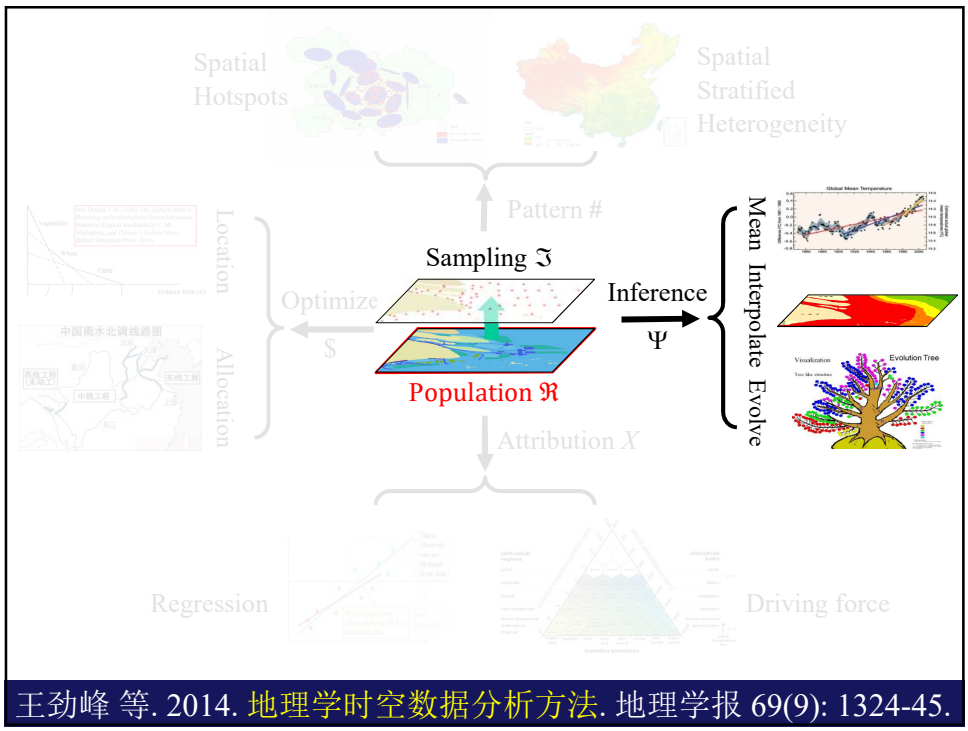
$q = 0.53^{***}$



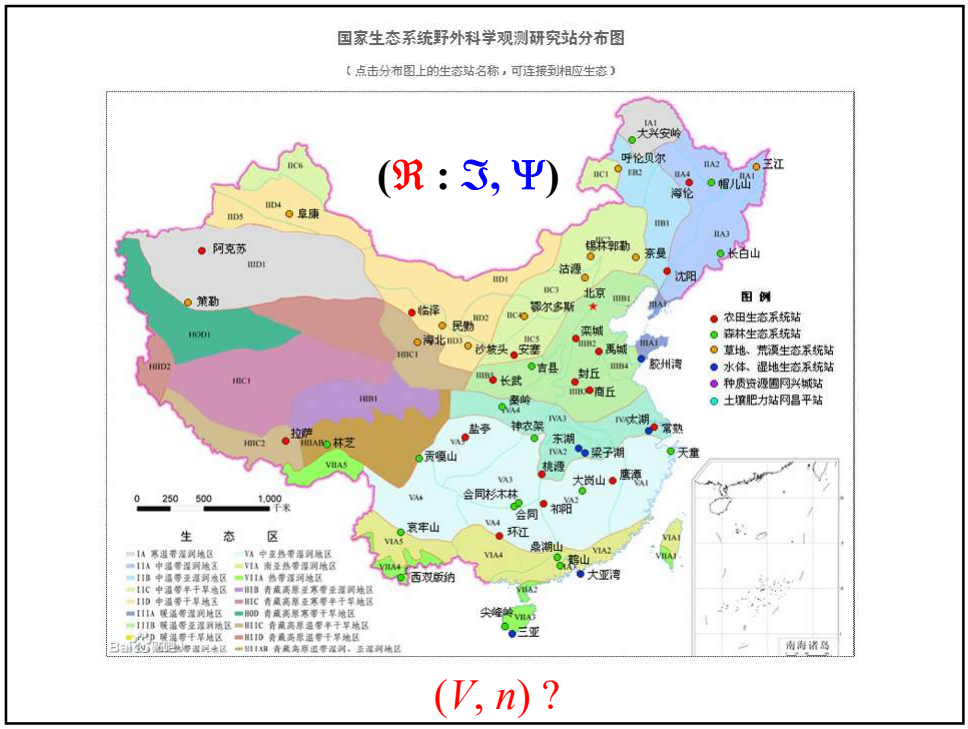
Wang JF, Zhang TL, Fu BJ. 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 7: 250-256.

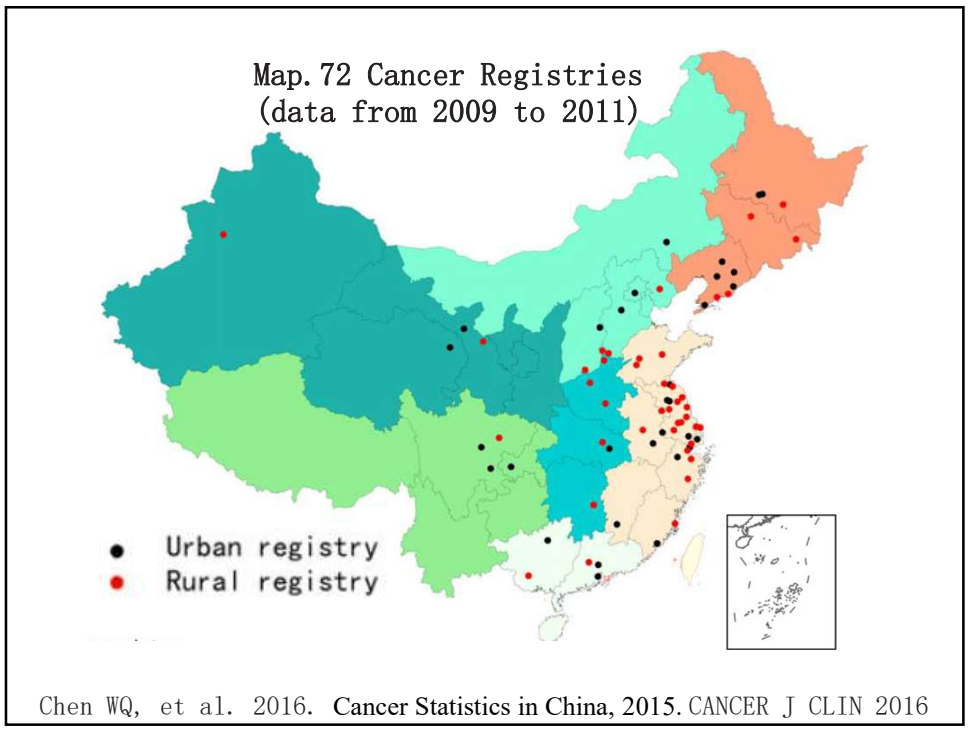


Goodchild M. 1986. CATMOG, GeoBooks, Norwich
Moran P. 1950. Notes on continuous stochastic phenomena. Biometrika 37:17-23.
Wang JF, Zhang TL, Fu BJ. 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 67: 250-256.

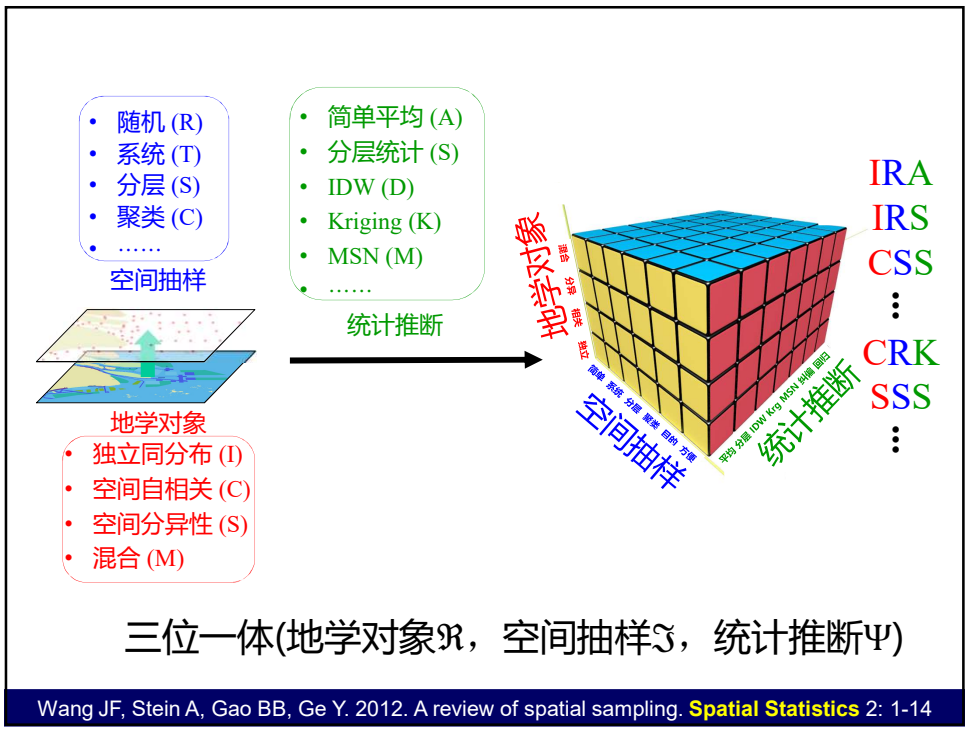


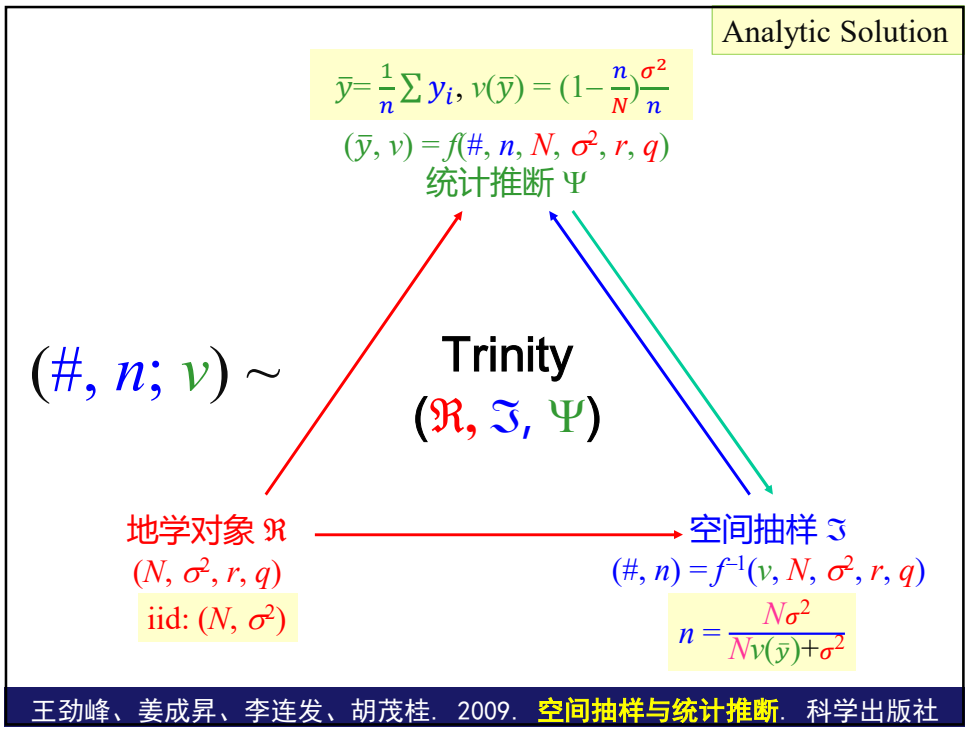
王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.

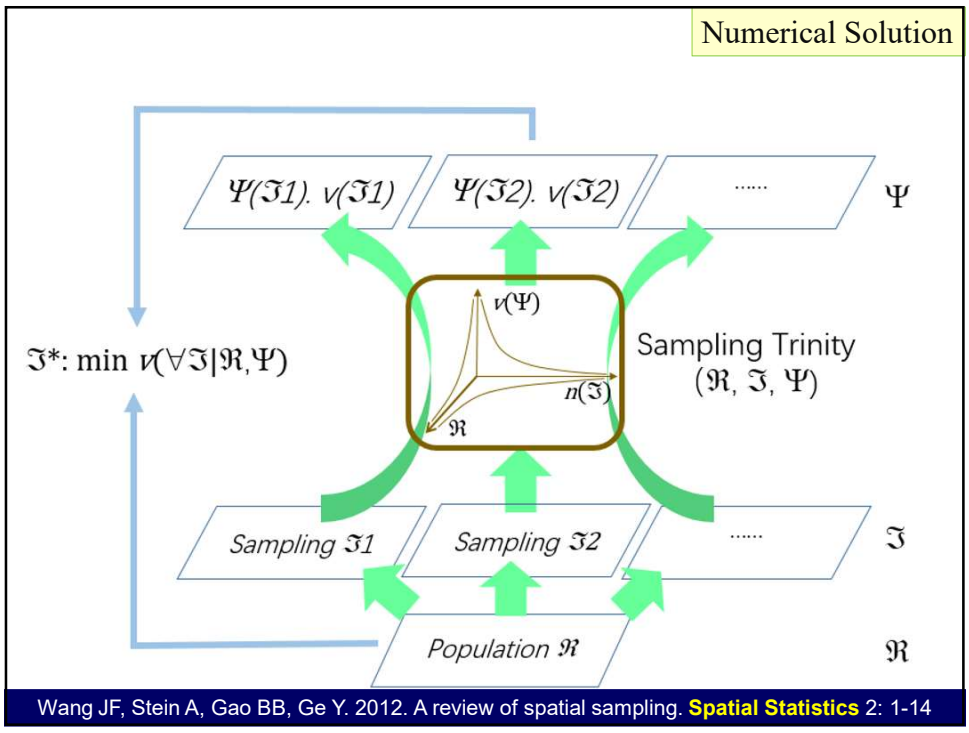

国家生态系统野外科学观测研究站分布图
（点击分布图上的生态站名称，可连接到相应生态）
$(\Re : \Im, \Psi)$
(V, n) ?

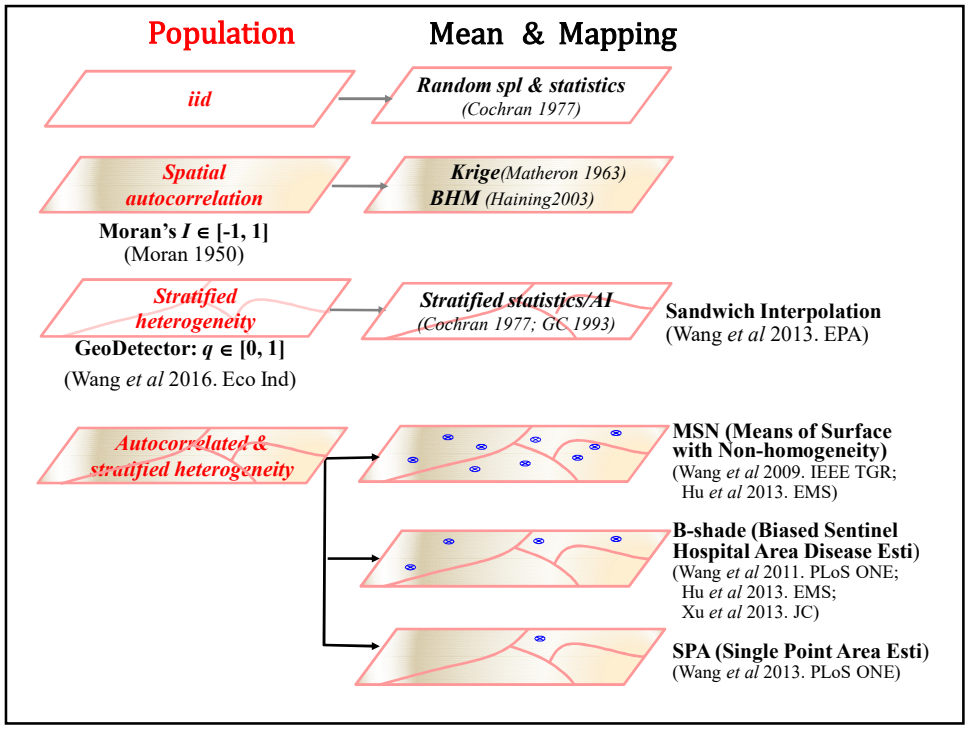

Map. 72 Cancer Registries
(data from 2009 to 2011)
Urban registry
Rural registry
Chen WQ, et al. 2016. Cancer Statistics in China, 2015. CANCER J CLIN 2016

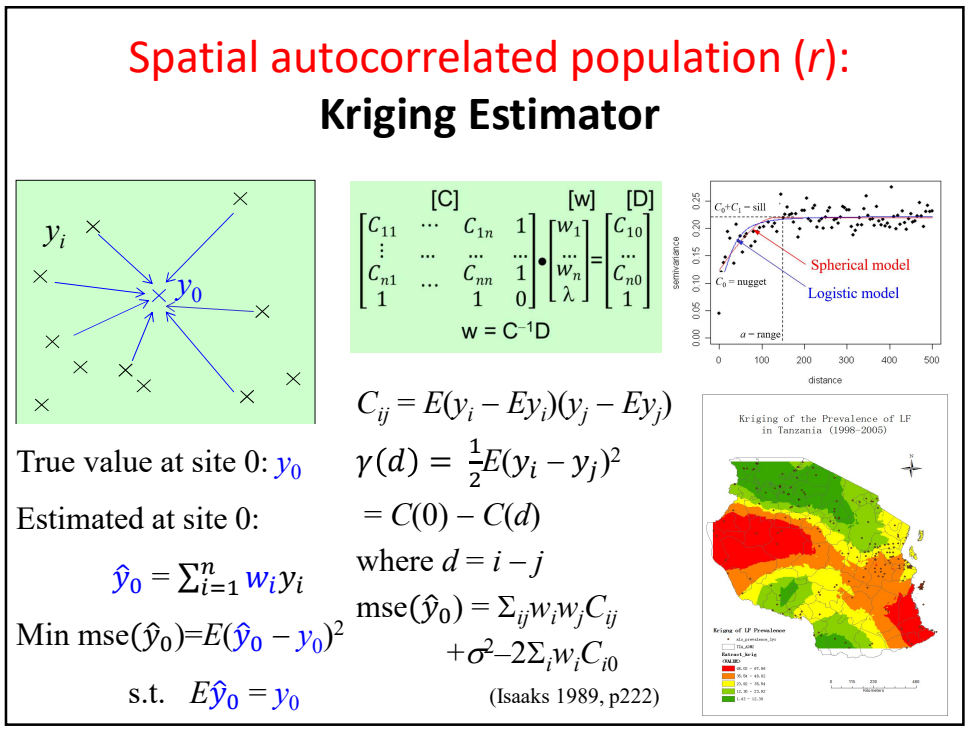

空间抽样与统计推断是地学调查和分析的基本手段
土壤
土地
作物
大气
大数据
交通流
疾病
林业
灾害
生态
地下水
水资源
污染
流动人口
高效监测
ℜ
高精度 ~ 低费用
mse(Ψ) ~ n(ℑ)
王劲峰、姜成昇、李连发、胡茂桂. 2009. 空间抽样与统计推断. 科学出版社


随机 (R)
系统 (T)
分层 (S)
聚类 (C)
……
空间抽样
简单平均 (A)
分层统计 (S)
IDW (D)
Kriging (K)
MSN (M)
……
统计推断
地学对象
独立同分布 (I)
空间自相关 (C)
空间分异性 (S)
混合 (M)
地学对象
空间抽样
统计推断
IRA
IRS
CSS
CRK
SSS
三位一体(地学对象ℜ，空间抽样ℑ，统计推断Ψ)
Wang JF, Stein A, Gao BB, Ge Y. 2012. A review of spatial sampling. Spatial Statistics 2: 1-14

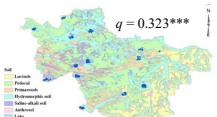

Analytic Solution
$\bar{y}=\frac{1}{n}\sum y_i$, $v(\bar{y})=(1-\frac{n}{N})\frac{\sigma^2}{n}$
$(\bar{y}, v)=f(\#, n, N, \sigma^2, r, q)$
统计推断 Ψ
$(\#, n; v)\sim$
Trinity
$(\Re, \Im, \Psi)$
地学对象 ℜ
(N, σ^2, r, q)
iid: (N, σ^2)
空间抽样 ℑ
$(\#, n)=f^{-1}(v, N, \sigma^2, r, q)$
$n=\frac{N\sigma^2}{Nv(\bar{y})+\sigma^2}$
王劲峰、姜成昇、李连发、胡茂桂. 2009. 空间抽样与统计推断. 科学出版社


Numerical Solution
Ψ(ℑ1). v(ℑ1)
Ψ(ℑ2). v(ℑ2)
.....
Ψ
v(Ψ)
n(ℑ)
ℜ
ℑ*: min v(∀ℑ|ℜ,Ψ)
Sampling Trinity
(ℜ, ℑ, Ψ)
Sampling ℑ1
Sampling ℑ2
.....
ℑ
Population ℜ
ℜ
Wang JF, Stein A, Gao BB, Ge Y. 2012. A review of spatial sampling. Spatial Statistics 2: 1-14

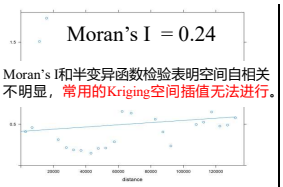

Population
Mean & Mapping
iid
Random spl & statistics
(Cochran 1977)
Spatial autocorrelation
Krige(Matheron 1963)
BHM (Haining2003)
Moran's I ∈ [-1, 1]
(Moran 1950)
Stratified heterogeneity
Stratified statistics/AI
(Cochran 1977; GC 1993)
Sandwich Interpolation
(Wang et al 2013. EPA)
GeoDetector: q ∈ [0, 1]
(Wang et al 2016. Eco Ind)
Autocorrelated & stratified heterogeneity
MSN (Means of Surface with Non-homogeneity)
(Wang et al 2009. IEEE TGR;
Hu et al 2013. EMS)
B-shade (Biased Sentinel Hospital Area Disease Esti)
(Wang et al 2011. PLoS ONE;
Hu et al 2013. EMS;
Xu et al 2013. JC)
SPA (Single Point Area Esti)
(Wang et al 2013. PLoS ONE)


Spatial autocorrelated population (r):
Kriging Estimator
y_i
y_0
[C] [w] [D]
$\begin{bmatrix} C_{11} & \cdots & C_{1n} & 1 \\ \vdots & \cdots & \cdots & \cdots \\ C_{n1} & \cdots & C_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \bullet \begin{bmatrix} w_1 \\ \cdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} C_{10} \\ \cdots \\ C_{n0} \\ 1 \end{bmatrix}$
w = C⁻¹D
C_0+C_1 = sill
C_0 = nugget
a = range
Spherical model
Logistic model
semivariance
distance
$C_{ij} = E(y_i - Ey_i)(y_j - Ey_j)$
$\gamma(d) = \frac{1}{2}E(y_i - y_j)^2$
$= C(0) - C(d)$
where $d = i - j$
$\text{mse}(\hat{y}_0) = \Sigma_{ij} w_i w_j C_{ij}$
$+\sigma^2 - 2\Sigma_i w_i C_{i0}$
(Isaaks 1989, p222)
True value at site 0: y_0
Estimated at site 0:
$\hat{y}_0 = \sum_{i=1}^{n} w_i y_i$
Min mse($\hat{y}_0$)=$E(\hat{y}_0 - y_0)^2$
s.t. $E\hat{y}_0 = y_0$
Kriging of the Prevalence of LF in Tanzania (1998-2005)

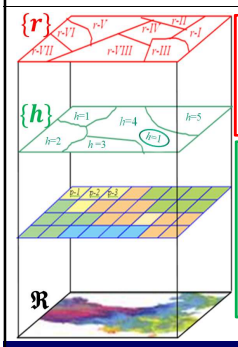
Spatial stratified heterogeneous population (q): **Sandwich Estimator**



全国鼠疫布氏菌病防治基地，吉林白城



Moran's I和半变异函数检验表明空间自相关不明显，常用的Kriging空间插值无法进行。



From SSH units to reporting units,

$\bar{y}_r = \frac{1}{N_r}\sum_{h=1}^{L_{rh}} N_{r\cap h}\bar{y}_h = \sum_{h=1}^{L_{rh}} W_{r\cap h}\bar{y}_h$

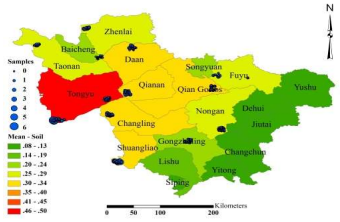
$V\bar{y}_r = \sum_{h=1}^{L_{rh}} W_{r\cap h}^2 V\bar{y}_h$

From population to SSH units,

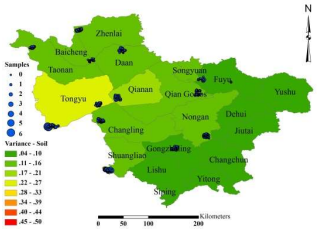
$\bar{y}_h \triangleq \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$

$V\bar{y}_h = \frac{1}{n_h}(1-\frac{n_h}{N_h})\sigma_h^2$

where $\sigma_h^2 = \frac{1}{n_h-1}\sum_{i=1}^{n_h}(y_{hi}-\bar{y}_h)^2$



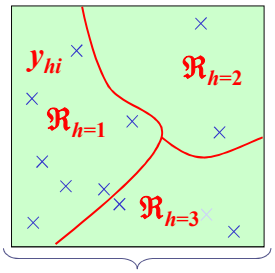
三明治插值鼠密度



三明治插值标准差

Wang JF, *et al.* 2010. Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. **International Journal of Geographical Information Science** 24(4): 523-543.

(r + q) & Stratified sample: **MSN Estimator**



$\Re$

$\bar{Y}_\Re = \Re^{-1}\int_\Re y(s)ds$

$\bar{y}_\Re = \sum_{i=1}^{n} w_i y_i$

$= \sum_{h=1}^{L}\sum_{i=1}^{n} a_h w_{hi} y_{hi}$

Min mse($\bar{y}_\Re$)

st $E\bar{y}_\Re = \bar{Y}_\Re$

$a_g \sum_{h=1}^{H}\sum_{i=1}^{n_h} a_h w_{hi} cov(y_{hi}, y_{gj}) + \mu_g$

$= a_g \Re^{-1}\int_R cov\left(y_{gj}, y(s)\right)ds$

$g = 1, \ldots, H; j = 1, \ldots, n_g;$

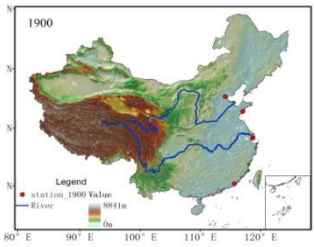
$a_g = \Re_g\Re^{-1}; a_h = \Re_h\Re^{-1}$

Annual temperature, sample estimation error $v \le v_0 = 0.4^0$C (existing stations, stations to be added)

Wang JF, Christakos G, Hu MG. 2009. Modeling spatial means of surfaces with stratified non-homogeneity. **IEEE Transactions on Geoscience and Remote Sensing** 47(12): 4167-4174

(*r* + *q*) & Biased sample: **B-Shade Estimator**

China Temperature Anomaly Series in 1900-2006

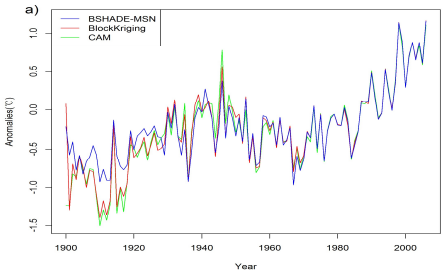


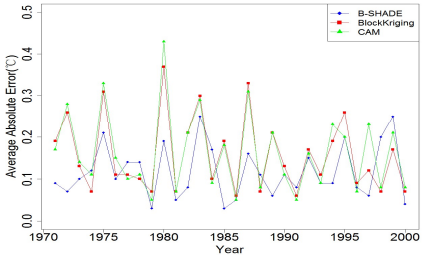
$\bar{y}(\boldsymbol{w}) = \sum_{i=1}^{n} w_i y_i$

to min $v\bar{y}(\boldsymbol{w}) = E(y(\boldsymbol{w}) - Y)^2$

s.t. $Y = E\bar{y}(\boldsymbol{w}) = \Sigma_{i=1}^{n} w_i E y_i$,

i.e $\Sigma_{i=1}^{n} w_i b_i = 1$, $b_i = Ey_i/EY$

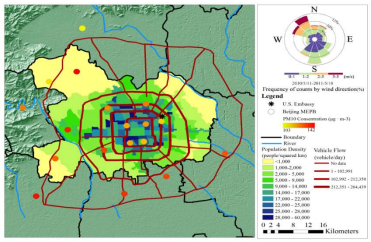




Wang JF, Xu CD, Hu MG, Li QX, Yan ZW, Zhao P, Jones P. 2014. A new estimate of the China temperature anomaly series and uncertainty assessment in 1900-2006. **Journal of Geophysical Research** 119(1): 1-9.

(*r* + *q*) & Single point sample: **SPA Estimator**

Single Point Area Estimation of PM2.5

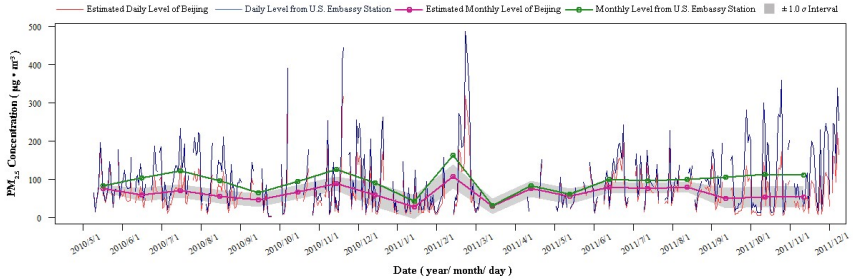


$Y = \sum_{i=1}^{N} g_i y_i$

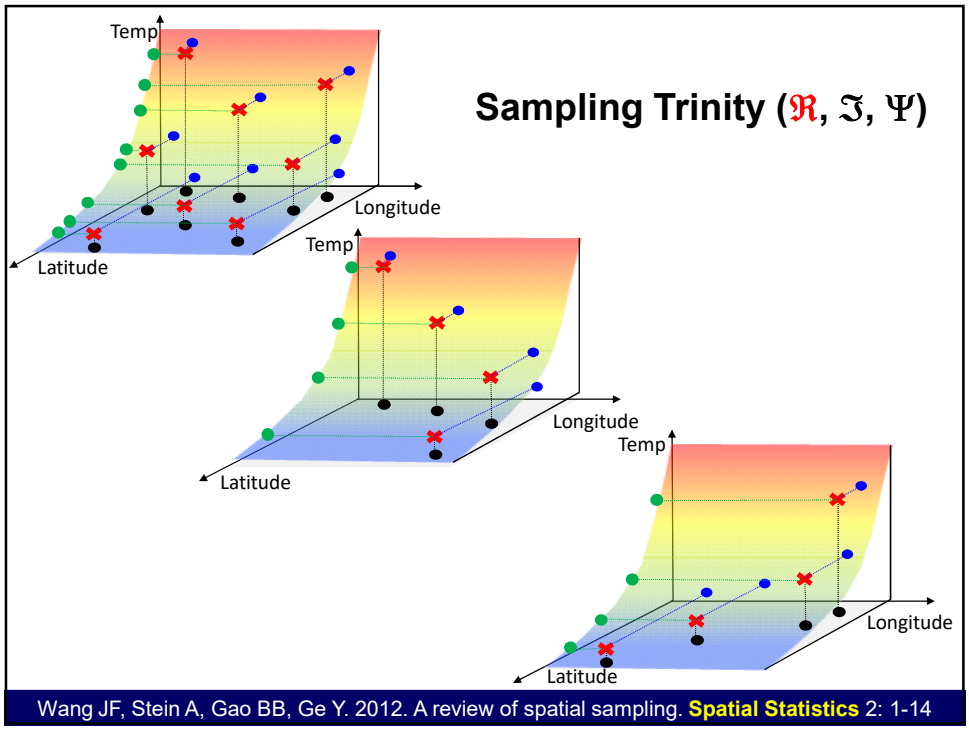
$\hat{Y} = w_0 y_0$

$w_0 = \text{argmin}\,[v(Y) = E(\hat{Y} - Y)^2]$

s.t. $E\hat{Y} = Y$



Wang JF, Hu MG, Xu CD, Christakos G, Zhao Y. 2013.
Estimation of Citywide Air Pollution in Beijing. **PLoS ONE** 8(1): e53400.



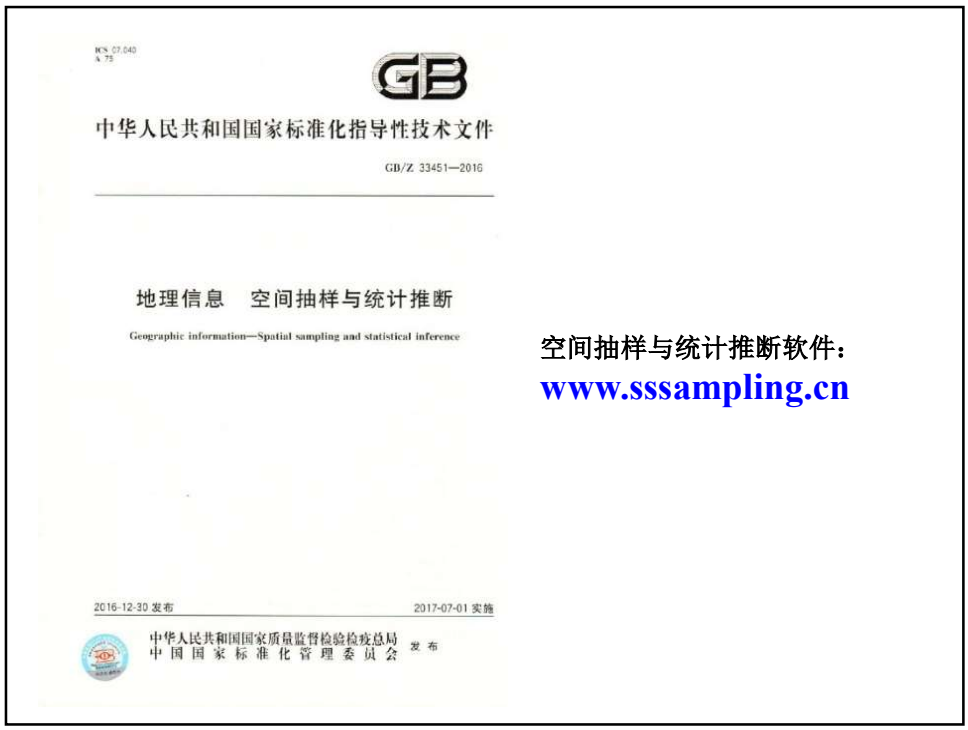
Sampling Trinity (ℜ, ℑ, Ψ)

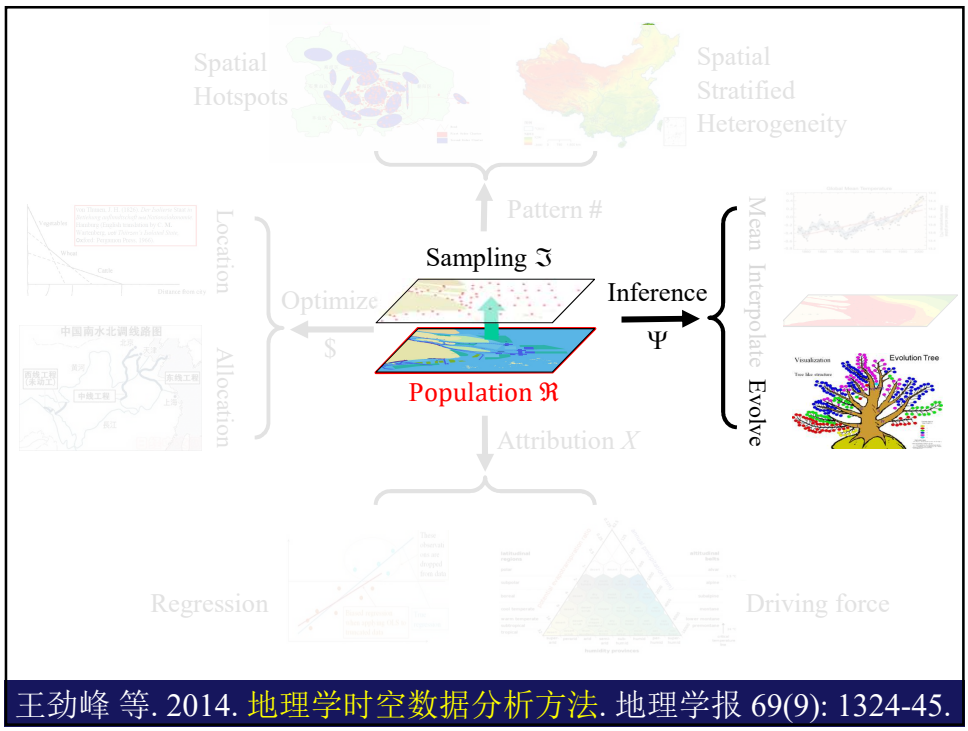
On a **planned** monitoring network ℜ (agriculture, demography, economy, environment, or epidemics):
- Calculation of the optimum distribution and density of sample sites to form a highly efficient spatial sampling project or monitoring network;

On an **existing** monitoring network ℑ (weather observation network, epidemic surveillance network, satellite monitoring scheme):
- Recommend the best overall valuation methods and recommendations to improve the monitoring network (based on monitoring the target characteristics and distribution of observations);

On **published** statistics Ψ (regional greenhouse gas emissions; prevalence of a disease in a region, areas of contaminated soil in a region)
- Evaluation of the statistical errors (through the study of its sample distribution, density, statistical methods)

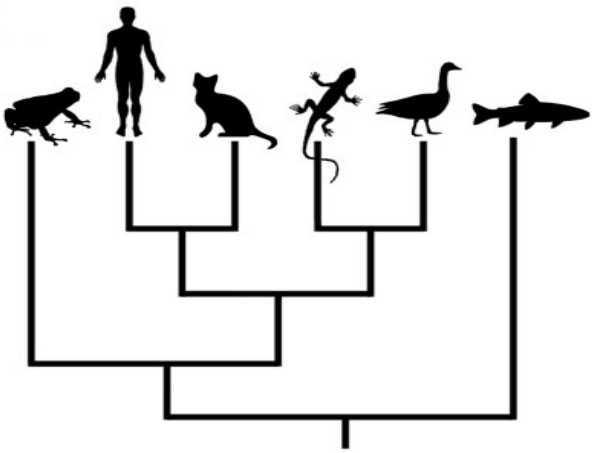
王劲峰 等. 2010. 空间抽样与统计推断. 科学出版社





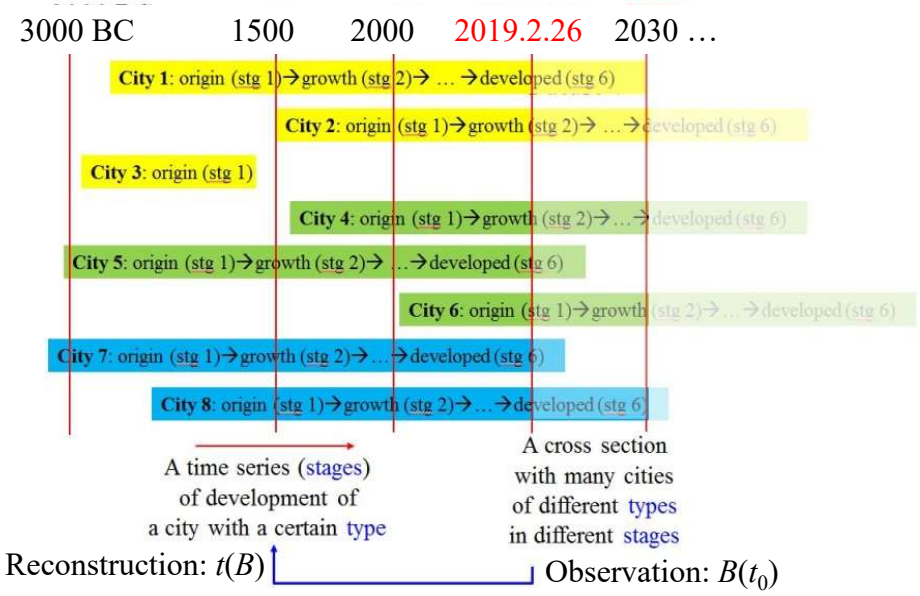
王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.

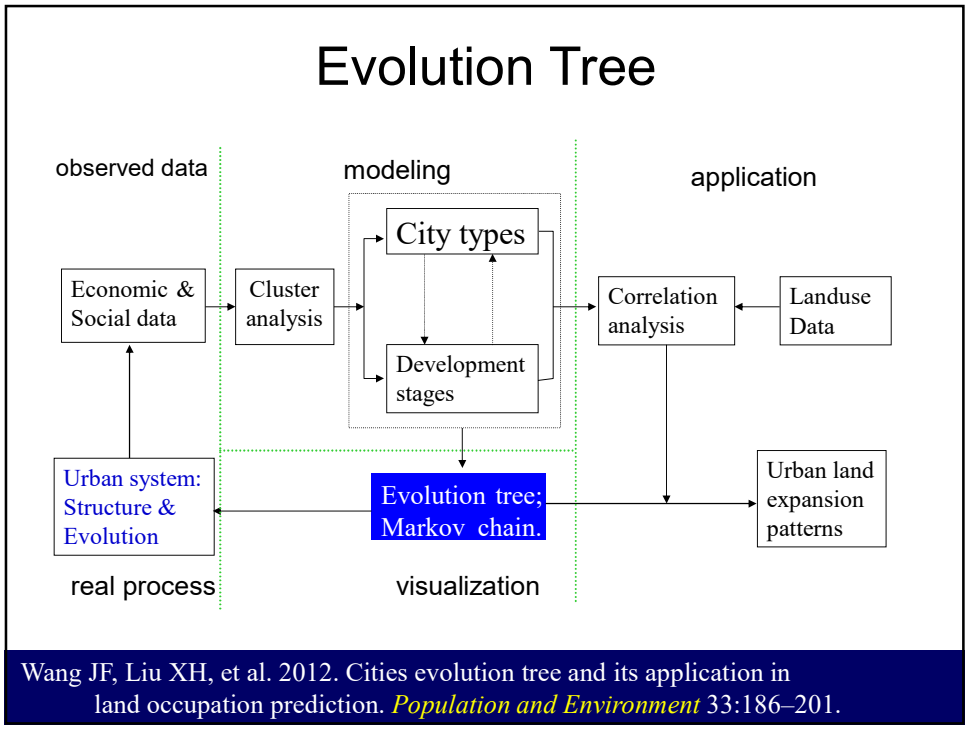
Biological Evolution



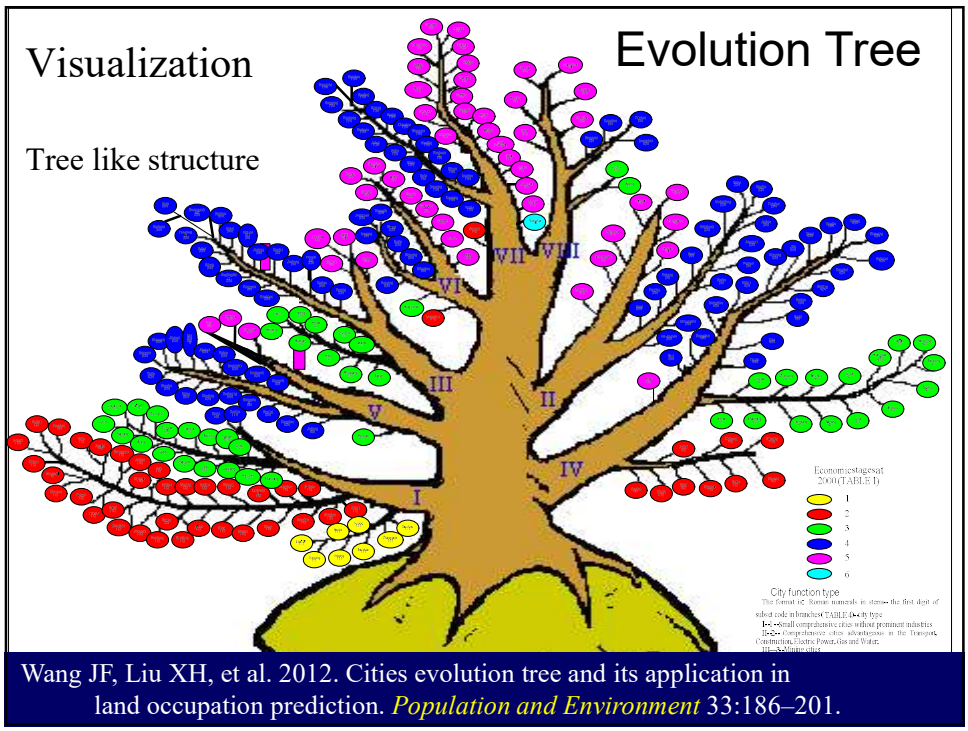
Evolution: $B(t)$
Observation: $B(t_0)$
Reconstruction: $t(B)$

Cities Evolution



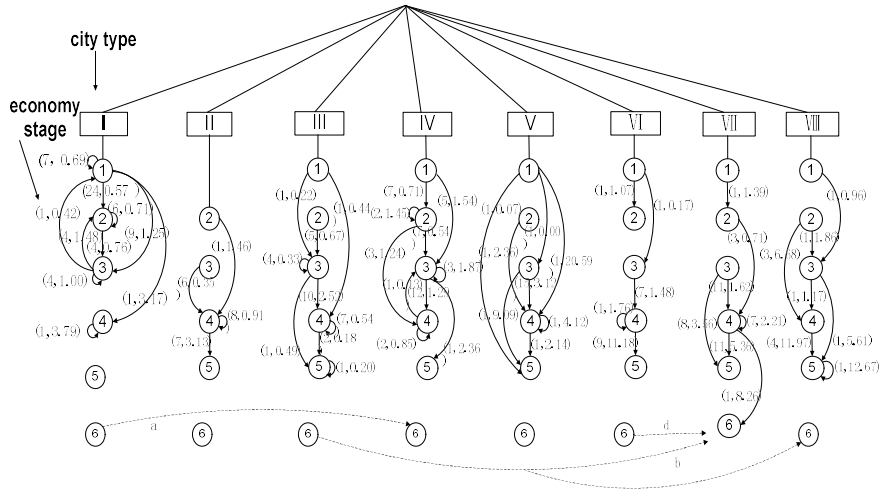


Wang JF, Liu XH, et al. 2012. Cities evolution tree and its application in land occupation prediction. *Population and Environment* 33:186–201.



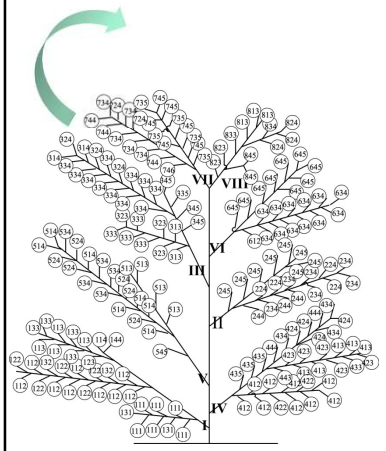
Wang JF, Liu XH, et al. 2012. Cities evolution tree and its application in land occupation prediction. *Population and Environment* 33:186–201.

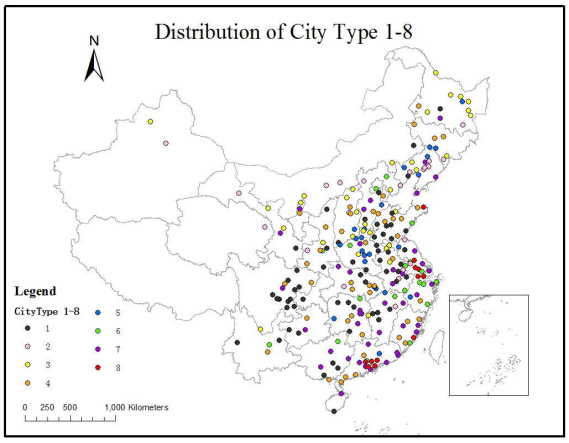
Evolution Tree
Application II: City Land Occupation



Wang JF, Liu XH, et al. 2012. Cities evolution tree and its application in land occupation prediction. *Population and Environment* 33:186–201.

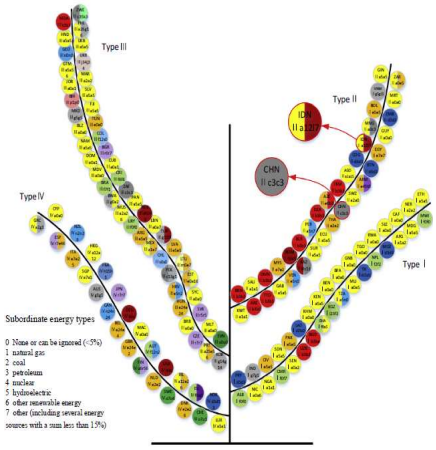
Attribute Space ⟺ Geographical Space

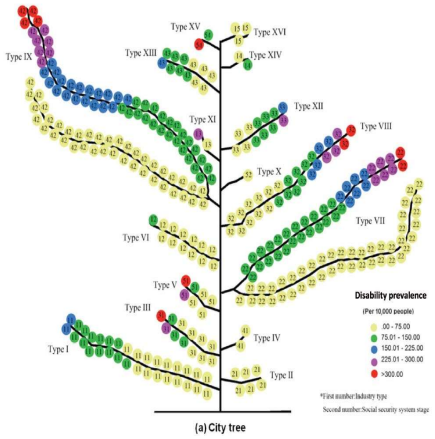




- 253 cities are separated in geographic space;
- 253 cities are highly correlated in the tree. (Wang et al 2012)

Evolution Tree

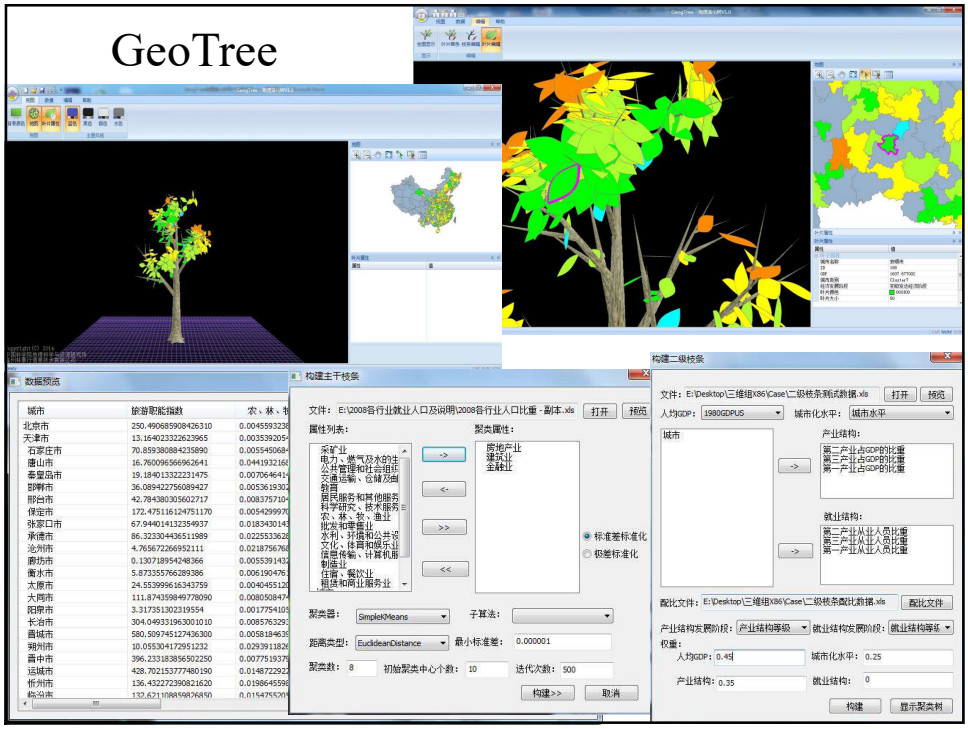




Hu et al. 2018. A novel evolution tree for analyzing the global energy consumption Structure. **Energy** 147: 1177-1187.

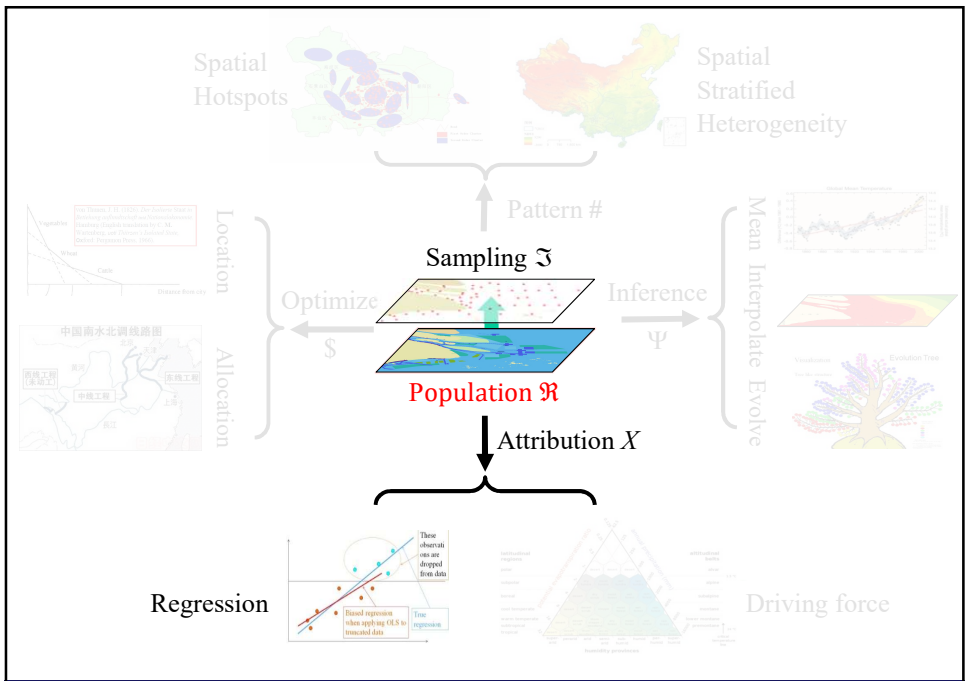
Liao YL, et al. 2013. Clustering of disability caused by unintentional injury amongst the population aged from 15–60 years: a challenge in rapidly developing China. **Geospatial Health** 8(1): 13-22.

GeoTree



Reference in Spatial Sampling & Inference

Classic：Cochran W. 1977. **Sampling Techniques**. Wiley
Spatial Autocorrelation & Interpolation
Matheron G . 1963. Principles of geostatistics. **Economic Geology** 58 (8): 1246-1266.
Spatial Stratified Heterogeneity & Interpolation：**Sandwich** (www.sssampling.cn)
Wang JF, Zhang TL, Fu BJ. 2016. A measure of spatial stratified heterogeneity. **Ecological Indicators** 7 (2016): 250-256.
Wang JF, Haining R, Liu TJ, Li LF, Jiang CS. 2013. Sandwich estimation for multi-unit reporting on a stratified heterogeneous surface. **Environment and Planning A** 45(10): 2515-2534.
Wang JF, Haining R, Cao ZD. 2009. Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. **International Journal of Geographic Information Science** 24(4): 523-543.
Wang JF, Liu JY, Zhuang DF, Li LF, Ge Y. 2002. Spatial sampling design for monitoring the area of cultivated land. **International Journal of Remote Sensing** 23(2): 263-284.
Li LF, Wang JF, Cao ZD, Feng XL, Zhang LL & Zhong ES. 2008. An information-fusion method to regionalize spatial heterogeneity for improving the accuracy of spatial sampling estimation. **Stochastic Environmental Research and Risk Assessment** 22(6): 689-704
Spatial Interpolation under Spatial Autocorrelation + Spatial Stratified Heterogeneity：**MSN** (www.sssampling.org)
Wang JF, Christakos G, Hu MG. 2009. Modeling spatial means of surfaces with stratified non-homogeneity. **IEEE Transactions on Geoscience and Remote Sensing** 47(12): 4167-4174.
Hu MG, Wang JF. 2011. A meteorological network optimization package using MSN theory. **Environmental Modeling & Software** 26: 546-548.
Biased sample remedy：**B-Shade** (www.sssampling.cn)
Wang JF, Reis BY, Hu MG, Christakos G, Yang WZ, et al. 2011. Area disease estimation based on sentinel hospital records. **PLoS ONE** 6(8): e23428.
Hu MG, Wang JF, Zhao Y, Jia L. 2013. A B-SHADE based best linear unbiased estimation tool for biased samples. **Environmental Modelling & Software** 48: 93-97.
Xu CD, Wang JF, Hu MG, Li QX. 2013. Interpolation of missing temperature data at meteorological stations using P-BSHADE. **Journal of Climate** 26: 7452-7463.
Xu CD, Wang JF, Li QX. 2018. A new method for temperatures spatial interpolation based on sparse historical stations. **Journal of Climate** 31: 1757-1770.
Wang JF, Xu CD, Hu MG, Li QX, Yan ZW, Zhao P, Jones P. 2014. A new China temperature anomaly series in 1900-2006. **Journal of Geophysical Research – Atmospheres** 119(1): 1-9.
Wang JF, Xu CD, Hu MG, Li QX, Yan ZW, Jones P. 2018. Global land surface air temperature dynamics since 1880. **International Journal of Climatology** 38: e466-e474.
Single point areal estimation: SPA (www.sssampling.cn)
Wang JF, Hu MG, Xu CD, Christakos G, Zhao Y. 2013. Estimation of citywide air pollution in Beijing. **PLoS ONE** 8(1): e53400.
Trinity (ℜ, ℑ, Ψ)
Wang JF, Stein A, Gao BB, Ge Y. 2012. A review of spatial sampling. **Spatial Statistics** 2: 1-14
Wang JF, Jiang CS, Hu MG, Cao ZD, Guo YS, Li LF. 2013. Design based spatial sampling: theory and implementation. **Environmental Modeling & Software** 40: 280-288
王劲峰、姜成晟、李连发、胡茂桂. 2009. **空间抽样与统计推断**. 科学出版社



王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.

Classic vs Spatial Linear Regression

Rob Haining 2003

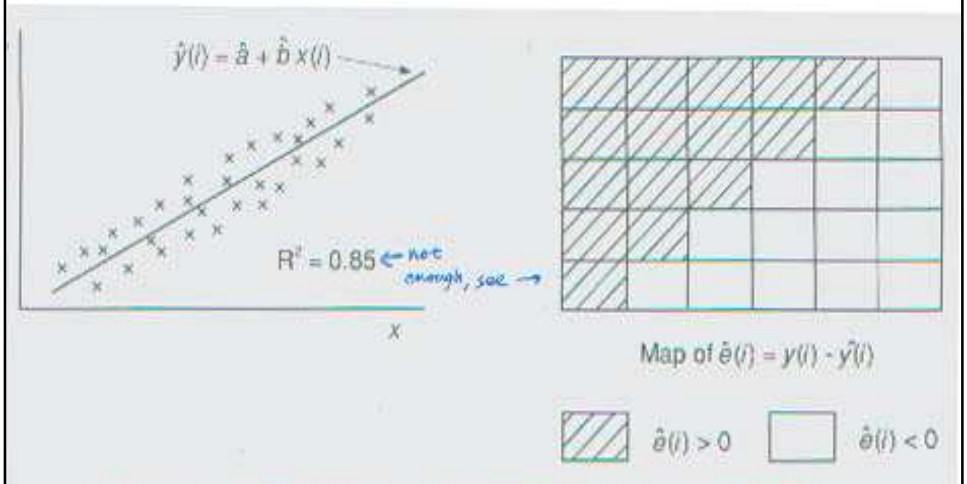
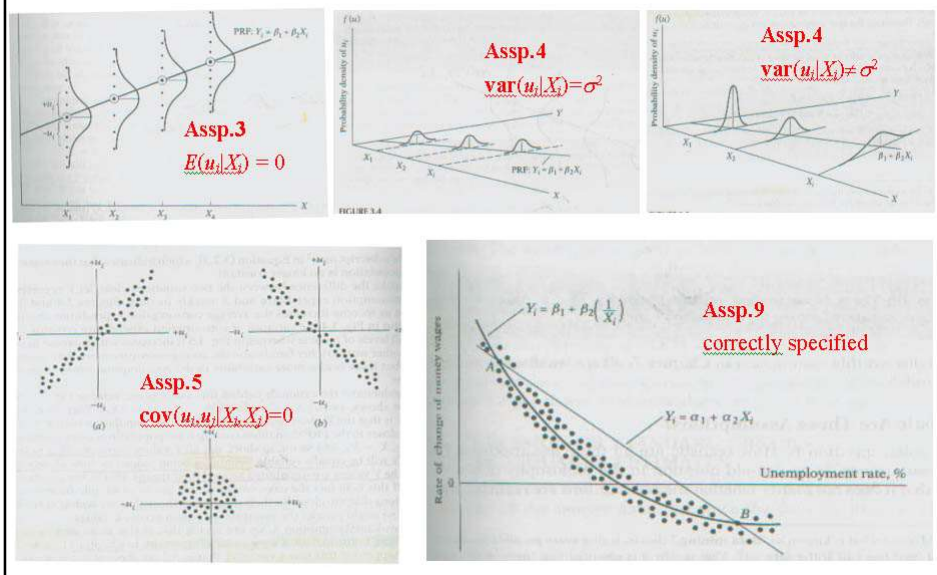


Figure 0.1 Evidence in assessing the adequacy of fit of a regression model

Classic Linear Regression



Spatial Linear Regression

(Luc Anselin 1988)

General model

$y = \rho W_1 y + X\beta + e$

$e = \lambda W_2 e + \mu$

$\mu \sim N(0, \Omega), \Omega_{ii} = h_i(za), h_i > 0$

β: K by 1; X: N by K

ρ: coeff. of the spatially lagged dependent variable

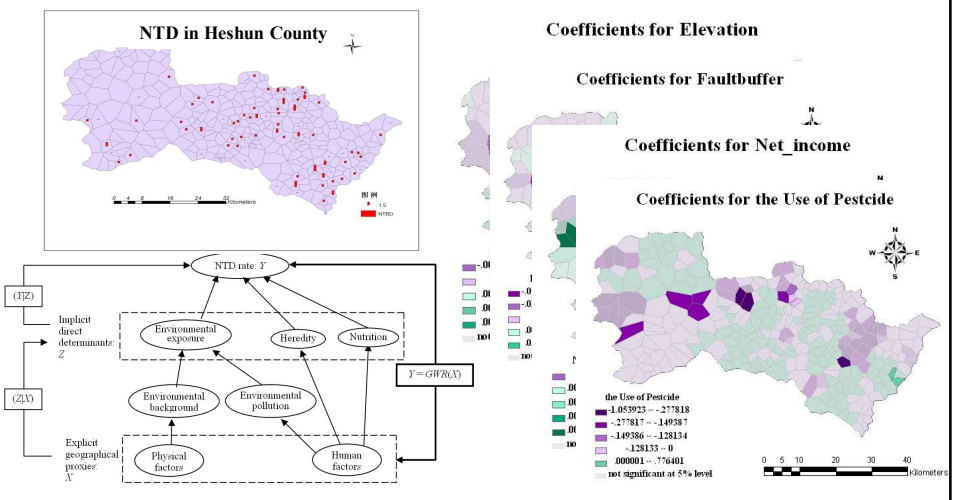
λ: coeff. in SAR structure for the disturbance e

μ: normally distributed with digonal covariance matrix Ω

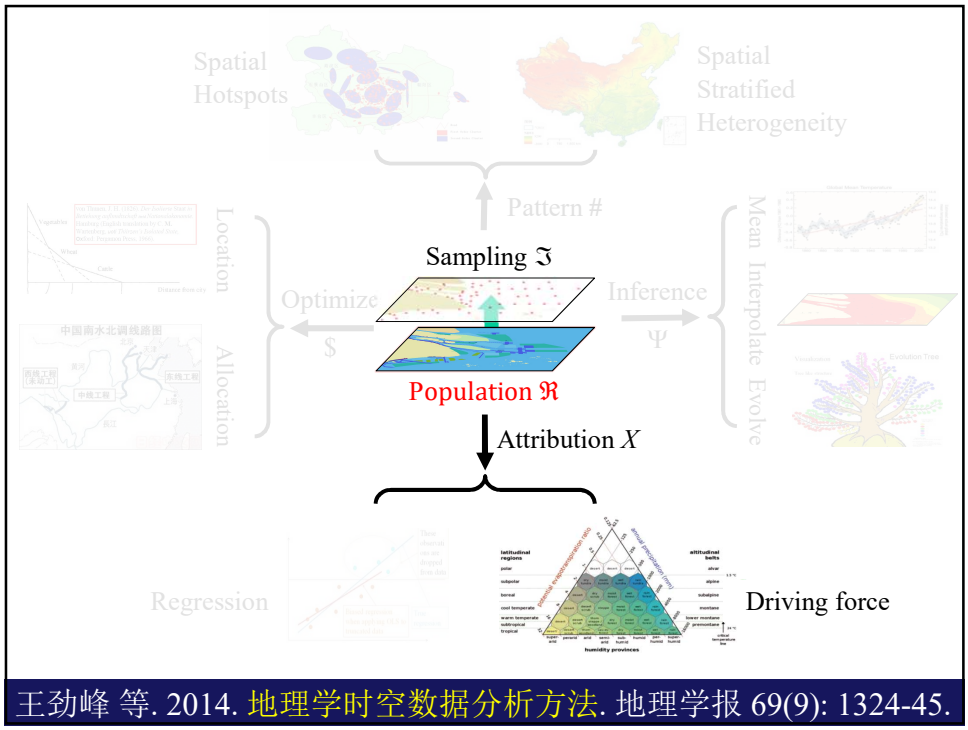
W_1, W_2: spatial weight matrices

GWR (Fothringham 1997)

$$y_i = a_0(u_i, v_i) + \sum_k a_k(u_i, v_i) x_{ik} + \varepsilon_i$$

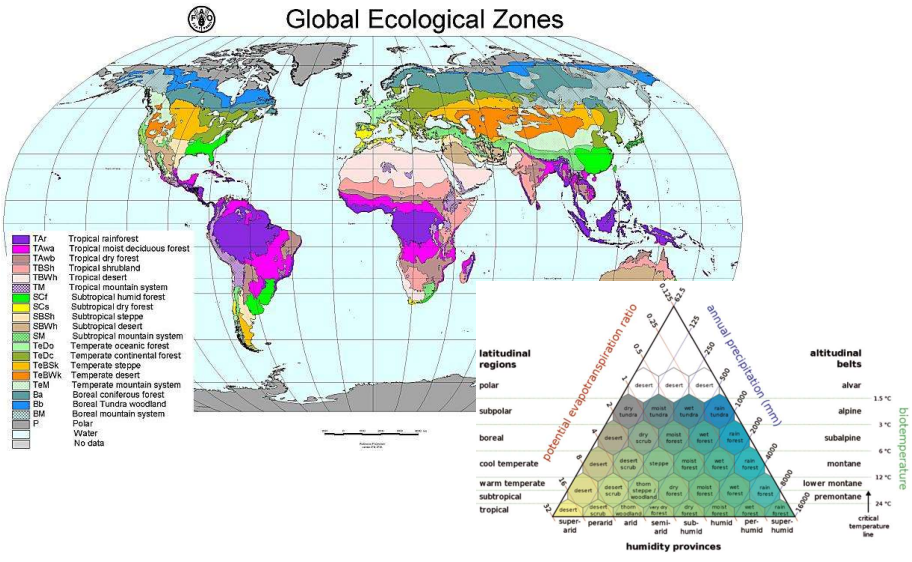


Wang JF, Liu X, Christakos G, Liao YL, Gu X, Zheng XY. 2010. Assessing local determinants of neural tube defects in the Heshun Region, China. *BMC Public Health* 10:52.



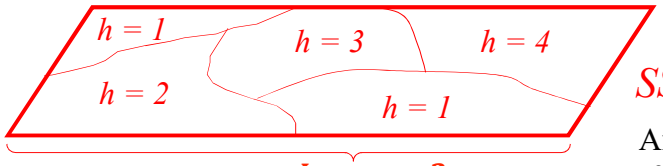
王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.

Spatial Stratified Heterogeneity & Attribution



Spatial pattern is a window for human to understand physical mechanism behind

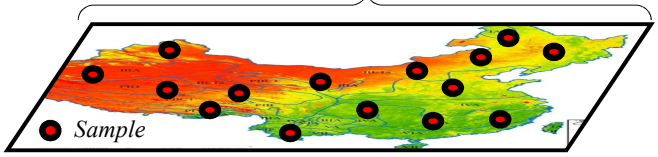
Spatial Stratified Heterogeneity (SSH): Attribute



SSH(X)

Axiom:
If *X* causes *Y*, their maps tend to be coupled

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



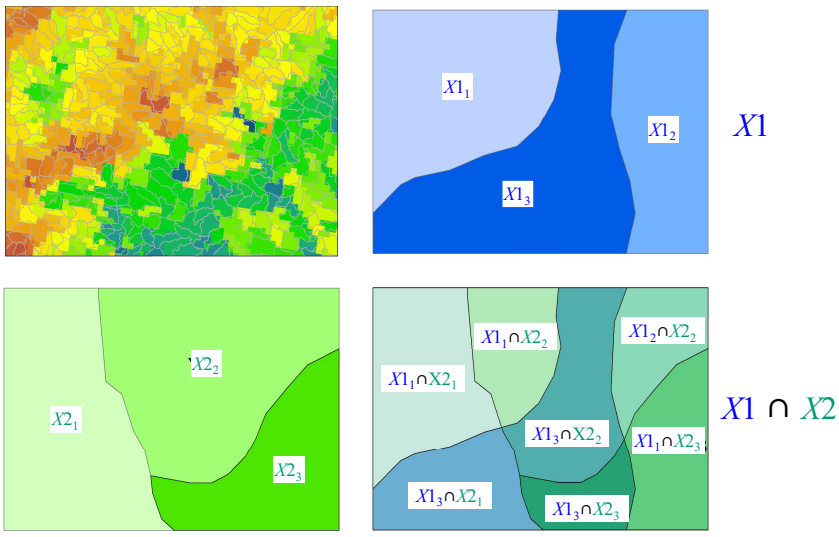
Y

$q \in [0, 1]$, $100q\%$ of *Y* is explained by *X*
0, if *Y* is independent to *X*
1, if *Y* is fully depended upon *X*

$$\frac{N-L}{L-1}\frac{q}{1-q} \sim F(L-1, N-L; \lambda)$$

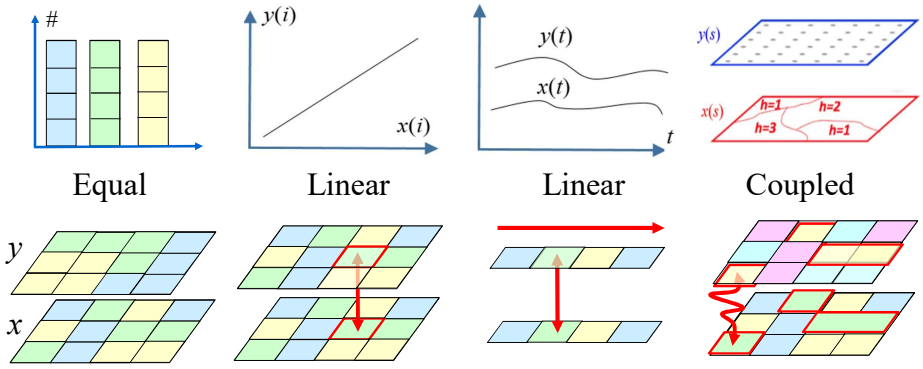
Wang JF, Zhang TL, Fu BJ. 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 7: 250-256.

Interaction: *Y* ← *X1* ∩ *X2*



Correlation → Causation

Histogram (#, 0D) → Pearson Correlate (+Pairs, 0D) → Two Time Series (+Order, 1D) → Geodetector (2D)

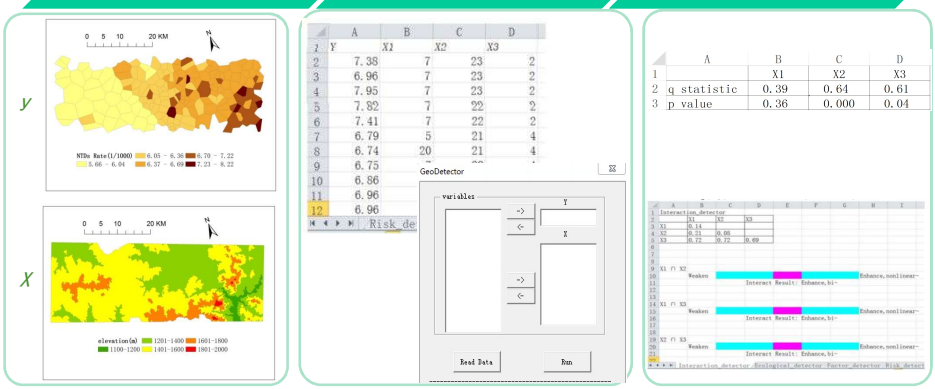


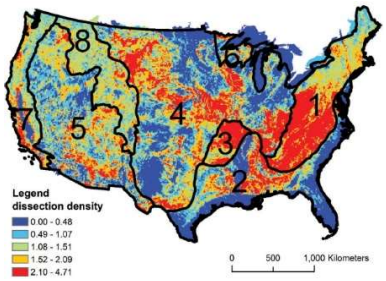
Wang JF, Zhang TL, Fu BJ. 2016. A measure of spatial stratified heterogeneity. Ecological Indicators 7: 250-256.

Software:

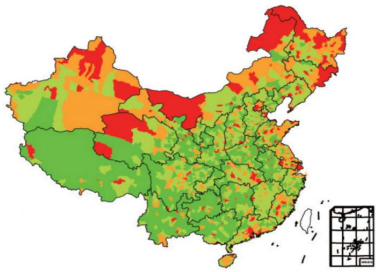
www.geodetector.cn

Prepare data → Run software → Output

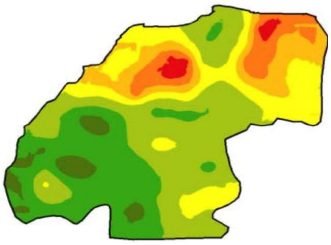




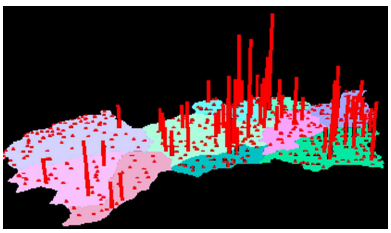
Nature: Dissection in USA



Social: Urbanization in China

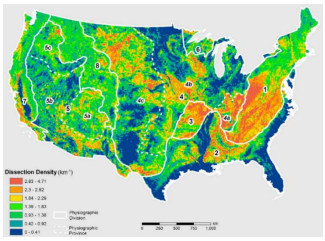


Pollution: Soil Antibiotic in a Town

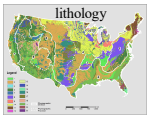


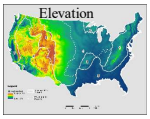
Health: NTD in a County

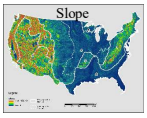
Determinants of Land Dissection Density in US

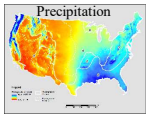


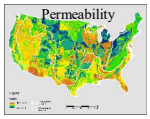
Y











X

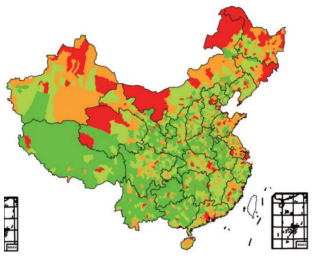
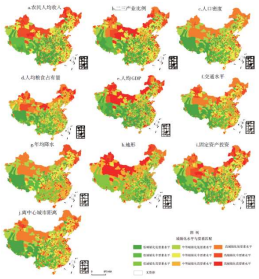
Y ← *Geodetecto* — *X*

Table 2. Factor or Factor Interaction With Maximum *q* Value

Physiographic Division/Province Name (#)	Dominant Factor	*q*	Dominant Interaction	*q*
Appalachian Highlands (1)	litho	0.49	litho ∩ elev	0.58
Atlantic Plain (2)	planc	0.46	elev ∩ precip[a]	0.62
Interior Highlands (3)	logk	0.31	logk ∩ precip	0.47
Interior Plains (4)	planc/tanc	0.29	litho ∩ planc/litho ∩ tanc	0.40
Interior Low Plateau (4a)	litho	0.21	logk ∩ litho/logk ∩ difelev	0.34
Central Lowlands (4b)	glaci	0.36	glaci ∩ planc/glaci ∩ tanc	0.52
Great Plains (4c)	planc	0.24	logk ∩ litho/logk ∩ tanc	0.34
Intermontane Plateaus (5)	litho	0.31	litho ∩ slp	0.37
Colorado Plateau (5a)	litho	0.13	litho ∩ planc	0.20
Basin and range (5b)	litho	0.35	Litho ∩ slp/litho ∩ tanc	0.41
Columbia Plateau (5c)	litho	0.33	litho ∩ precip	0.46
Laurentian Upland (6)	litho	0.23	litho ∩ tanc	0.32
Pacific Mountain System(7)	elev	0.28	elev ∩ precip[a]	0.43
Rocky Mountain System (8)	litho	0.10	litho ∩ tanc/litho ∩ slp	0.18

Luo W, Jasiewicz J, Stepinski T, et al. 2016. Spatial association between dissection density and environmental factors over the entire conterminous United States, *Geophys. Res. Lett.* 43: 692–700

Driving Forces of Urbanization in China

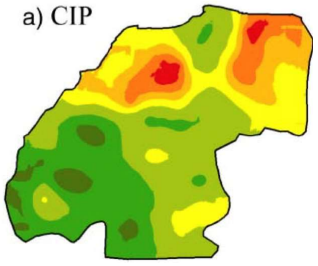
Y   X

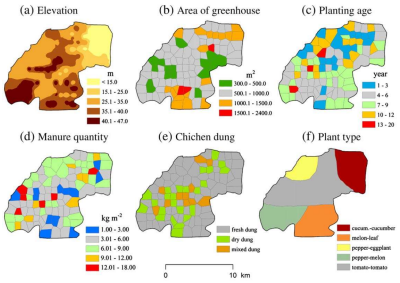
$Y \xrightarrow{Geodetector} X$

Area	Fixed invest	Dist from center city	Farmer income	Indus struct	Popula density	Crop per cpt	GDP Per cpt	Transport ation	Elev ation	Precip itation
106 Highway	0.30	0.34	**0.58**	0.18	0.10	0.45	0.53	0.13	0.00	0.11
North Border	0.06	0.03	0.13	0.05	0.01	**0.23**	0.02	0.21	0.22	0.18
Eastern Coast	0.16	0.17	0.26	0.10	0.28	0.09	**0.36**	0.08	0.05	0.01
Yangzi River	0.27	0.19	0.39	0.33	0.29	0.21	**0.53**	0.13	0.19	0.07
Belt & Road	**0.51**	0.28	0.41	0.22	0.04	0.22	0.29	0.03	0.01	0.17

刘彦随，杨忍. 2012.中国县域城镇化的空间特征与形成机理. 地理学报 67(8): 1011–1020

Determinants of Soil Antibiotic in a County

Y 



X

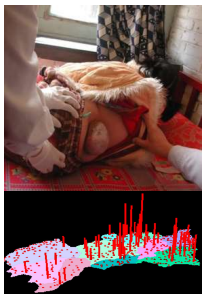
$Y \xrightarrow{Geodetector} X$

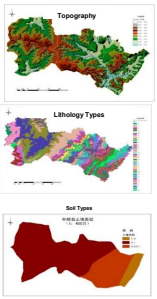
Vegetable planting models (0.28) > chicken dung quantity (0.20) > elevation (0.18) > planting age (0.09) > chicken dung types (0.06) > area of greenhouse (0.02).

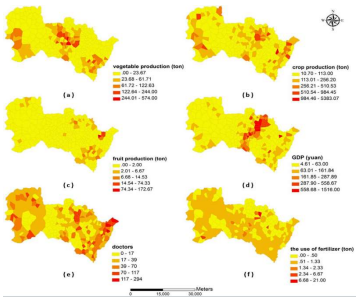
Li XW, et al. 2013. Influence of planting patterns on Fluoroquinolone residues in the soil of an intensive vegetable cultivation area in north China. **Science of the Total Environment** 458-460: 63-69.

Determinants of NTD in a County

Y







X

$Y \xrightarrow{Geodetector} X$

Primary environment (watershed, lithozone, and soil) strongly controls the NTD.
Basic nutrition (food) is more important than artificial environment (fertilizer) in controlling the spatial pattern of NTDs.
Ancient materials released from faults then spreading along slopes dramatically increase the risk of NTDs.

Wang JF, Li XH, Christakos G, et al 2010. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. International Journal of Geographical Information Science 24(1): 107-127.

References for Spatial Attribution

Spatial Regression:

Anselin L. 1988. **Spatial Econometrics: Methods and Models**. Kluwer Academic Publishers. GeoData

Spatial Interaction Modelling:

Fischer M, Wang JF. 2010. **Spatial Data Analysis: Models, Methods and Techniques**. Springer

GWR:

Fotheringham A, Brunsdon C, Charlton M. 2000. **Quantitative Geography**. Sage

BHM:

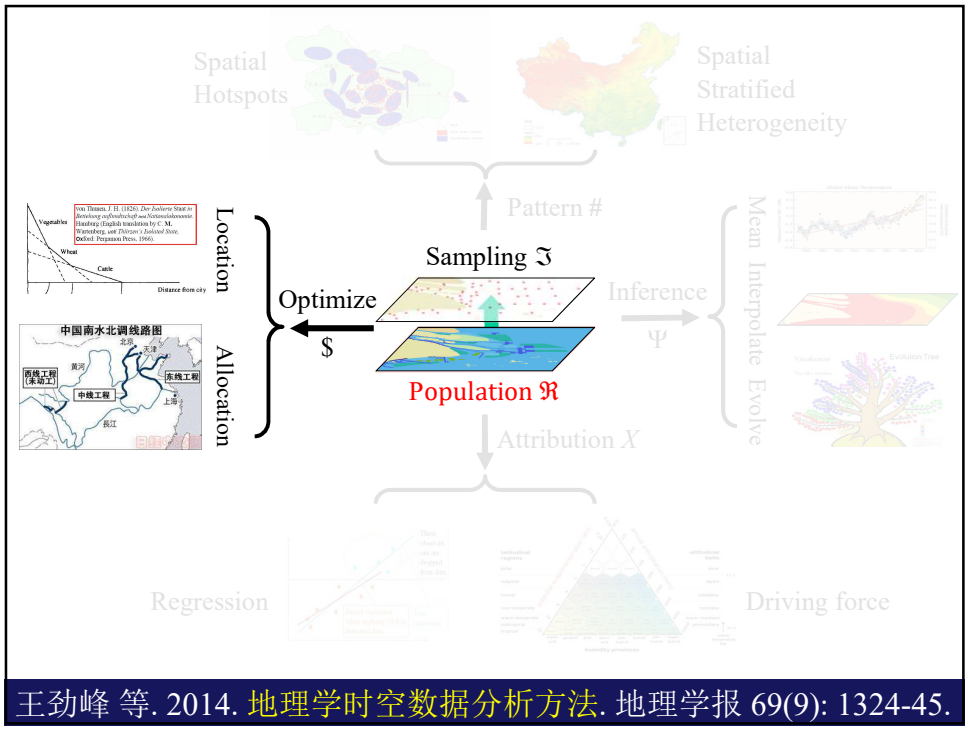
Haining R. 2003. **Spatial Data Analysis: Theory and Practice**. Cambridge University Press

Geographical Detector:

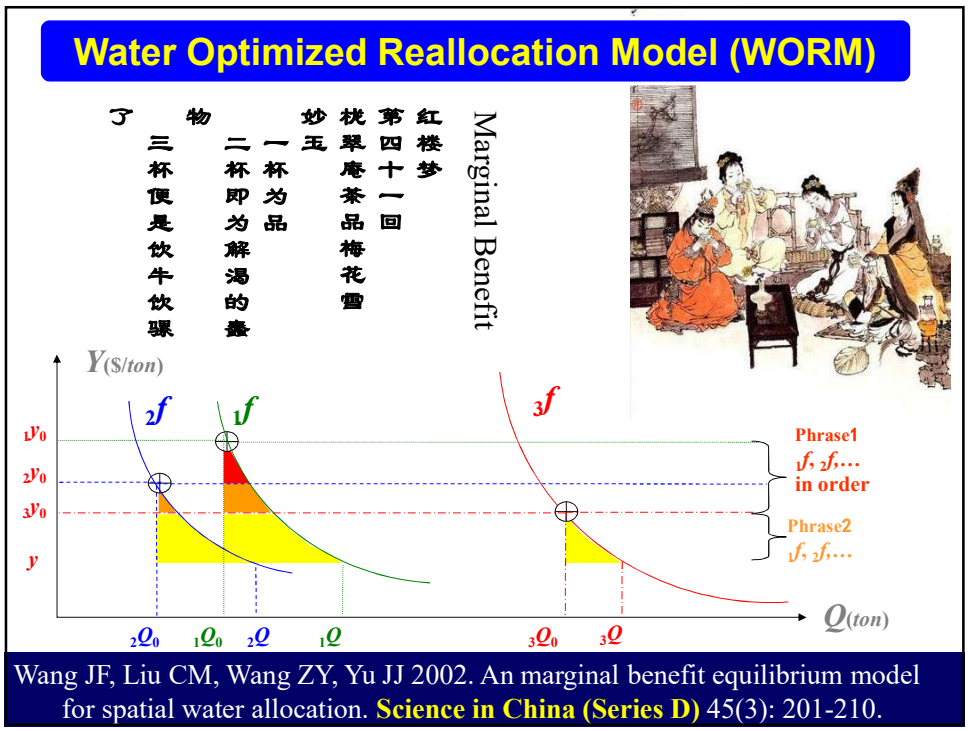
www.geodetector.cn

All:

王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程**（第二版）. 科学出版社

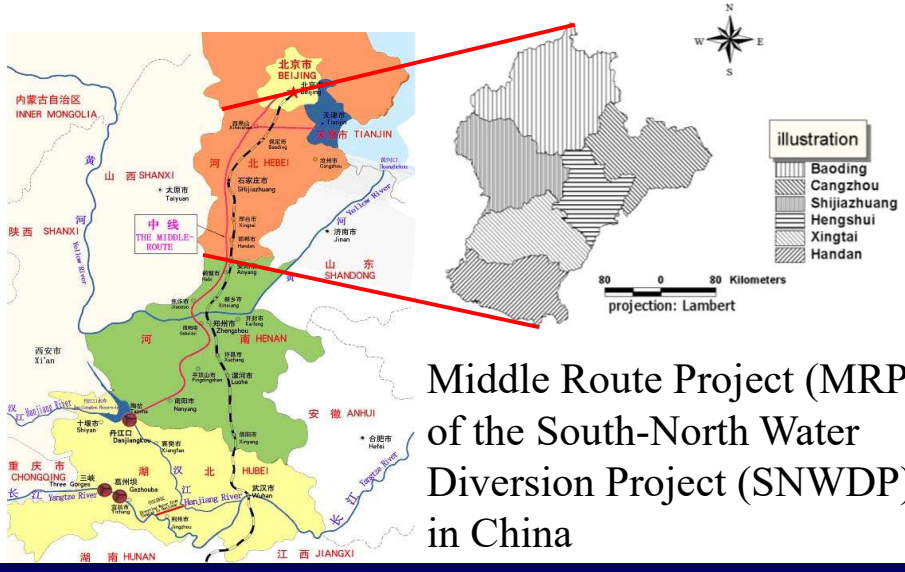


王劲峰 等. 2014. 地理学时空数据分析方法. 地理学报 69(9): 1324-45.



Wang JF, Liu CM, Wang ZY, Yu JJ 2002. An marginal benefit equilibrium model for spatial water allocation. **Science in China (Series D)** 45(3): 201-210.

Water Optimized Reallocation Model (WORM)

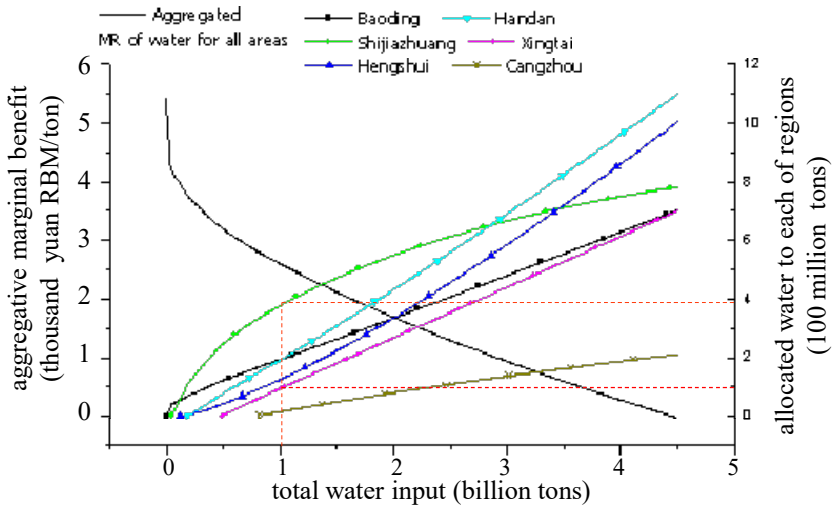


Middle Route Project (MRP) of the South-North Water Diversion Project (SNWDP) in China

Wang JF, Liu CM, Wang ZY, Yu JJ 2002. An marginal benefit equilibrium model for spatial water allocation. **Science in China (Series D)** 45(3): 201-210.

Water Optimized Reallocation Model (WORM)

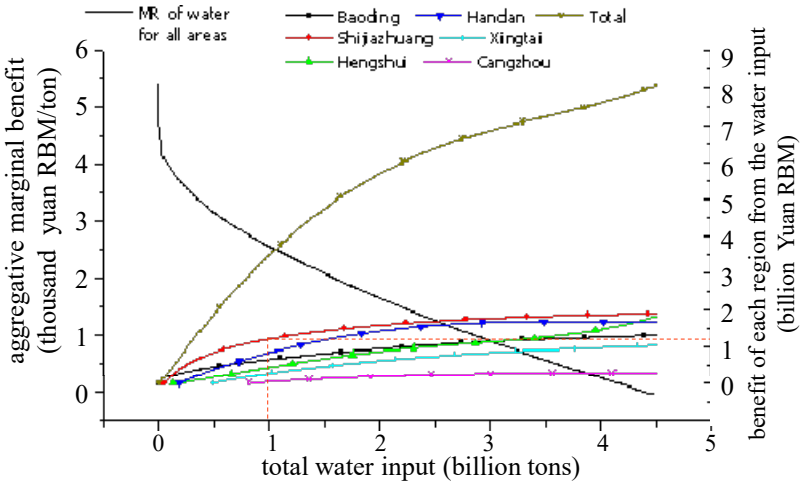
Distributing Water among Regions



Wang JF, Liu CM, Wang ZY, Yu JJ 2002. An marginal benefit equilibrium model for spatial water allocation. **Science in China (Series D)** 45(3): 201-210.

Water Optimized Reallocation Model (WORM)

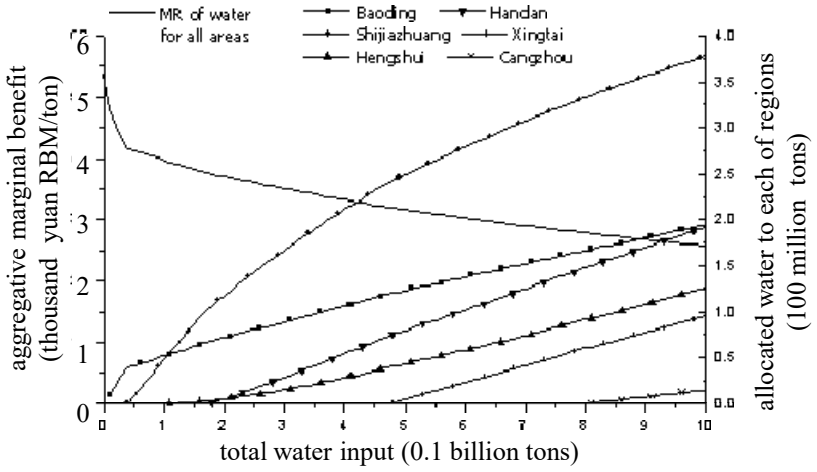
Benefit of Each Region from the Water Distribution



Wang JF, Liu CM, Wang ZY, Yu JJ 2002. An marginal benefit equilibrium model for spatial water allocation. **Science in China (Series D)** 45(3): 201-210.

Water Optimized Reallocation Model (WORM)

Time Series to Start to Distribute Water to Each of the Regions



Wang JF, Liu CM, Wang ZY, Yu JJ 2002. An marginal benefit equilibrium model for spatial water allocation. **Science in China (Series D)** 45(3): 201-210.

Water Optimized Reallocation Model (WORM)

Case 2：Heihe River

黑河流域的开发史可以上溯到两千多年前。汉武帝时代，骠骑将军霍去病两度率军西征，击败匈奴人，打通了河西走廊，开辟了中原王朝与中亚乃至欧洲贸易的“丝绸之路”。


明末清初
土尔扈特
Ming Dynasty
(1368-1644)

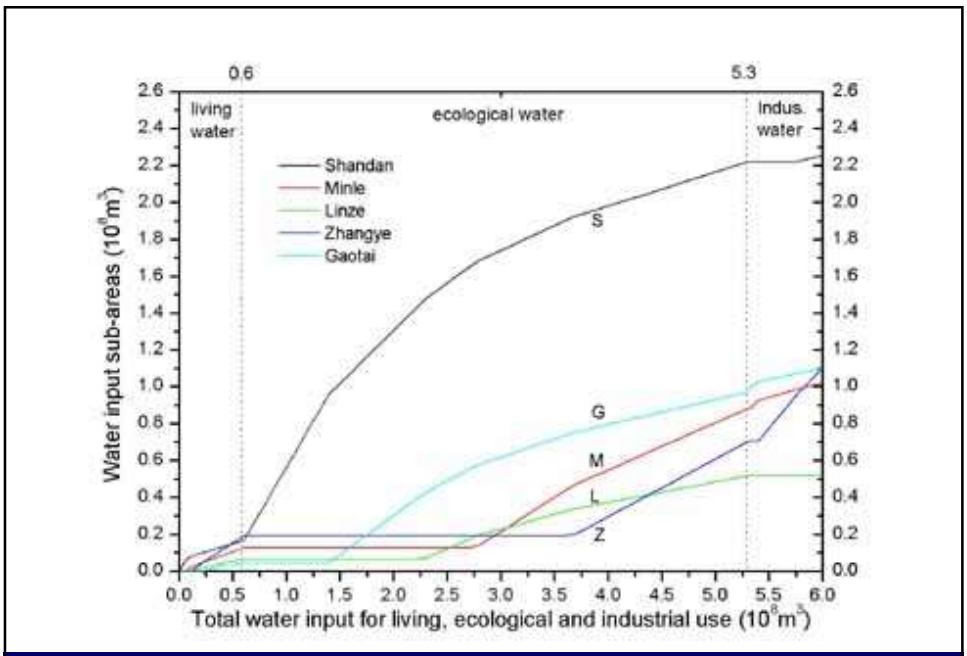

清初
均水制
Qing Dynasty
(1644-1912)


2001年
94次总理
办公会
PRC (1949-)

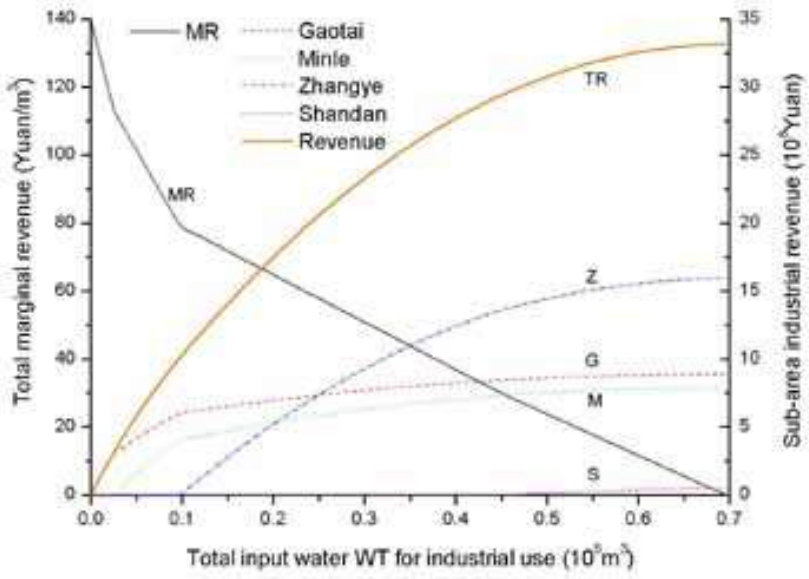
Fair & Efficient

2008
WORM

Wang JF, Chen GD, Gao YG, Long AH, Xu ZM, Li Xin. 2008. Optimal water resources allocation in arid and semi-arid areas. **Water Resources Management** 22: 239–258.



Wang JF, Chen GD, Gao YG, Long AH, Xu ZM, Li Xin. 2008. Optimal water resources allocation in arid and semi-arid areas. **Water Resources Management** 22: 239–258.



Wang JF, Chen GD, Gao YG, Long AH, Xu ZM, Li Xin. 2008. Optimal water resources allocation in arid and semi-arid areas. **Water Resources Management** 22: 239–258.

Summary

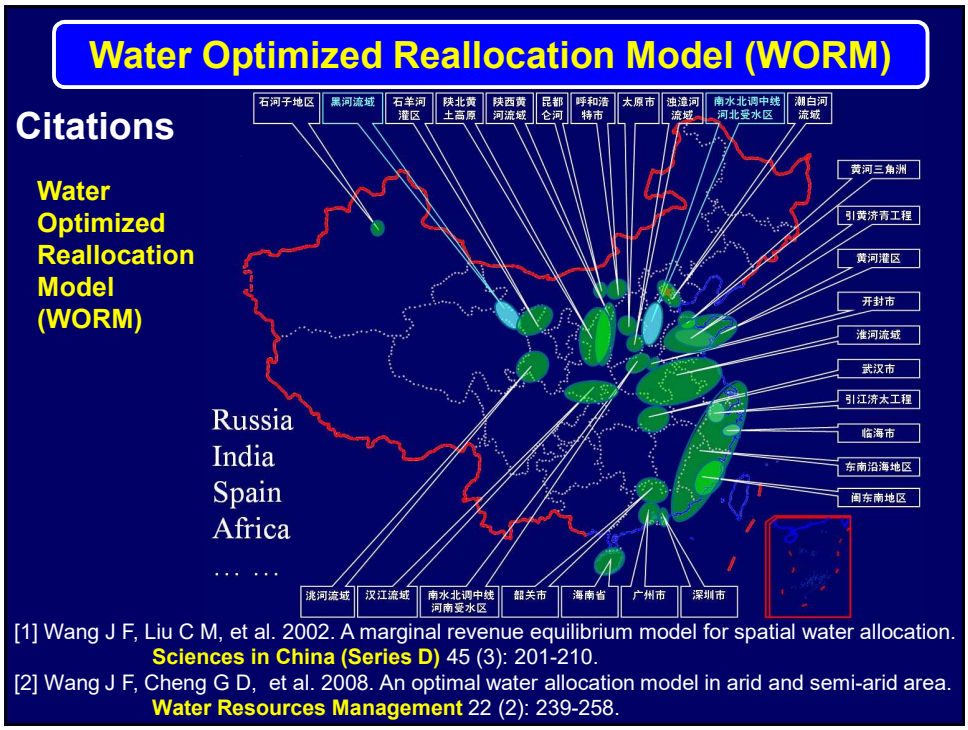
Input:

- MR_i: marginal benefit of *sub*-region *i*
- W: *total* water to be optimized

Output:

- W_i: water distr. to *sub*-regions
- S_i: benefit to *sub*- & *all*-region(s)
- T_i: sequence of *sub*-regions

Wang JF, Chen GD, Gao YG, Long AH, Xu ZM, Li Xin. 2008. Optimal water resources allocation in arid and semi-arid areas. **Water Resources Management** 22: 239–258.



Reference in Water Resources

Wang JF, Liu CM, Wang ZY, Yu JJ. 2002. An marginal benefit equilibrium model for spatial water allocation. **Science in China (Series D)** 45(3): 201-210.

Wang JF, Chen GD, Gao YG, Long AH, Xu ZM, Li X. 2008. Optimal water resources allocation in arid and semi-arid areas. **Water Resources Management** 22: 239–258.

王劲峰 刘昌明 王智勇 于静洁. 2001. 水资源空间配置的边际效益均衡模型. **中国科学** 31(5): 422-427.

王劲峰 刘昌明 于静洁 陈红焱. 2001. 区域调水时空优化配置理论模型探讨. **水利学报** 2001(4): 7-14.

Discussion

Benefit from Spatial Analysis

1. Geospace: an extra bonus
2. Spatial autocorrelation → less efficient regression
3. Spatial stratified heterogeneity → bias statistics
4. One realization → Spatial sampling trinity (SST)
5. Index: integrating human & environment factors
6. Spatial optimization: water, locations
7. Transform between geospacer & time → Geotree

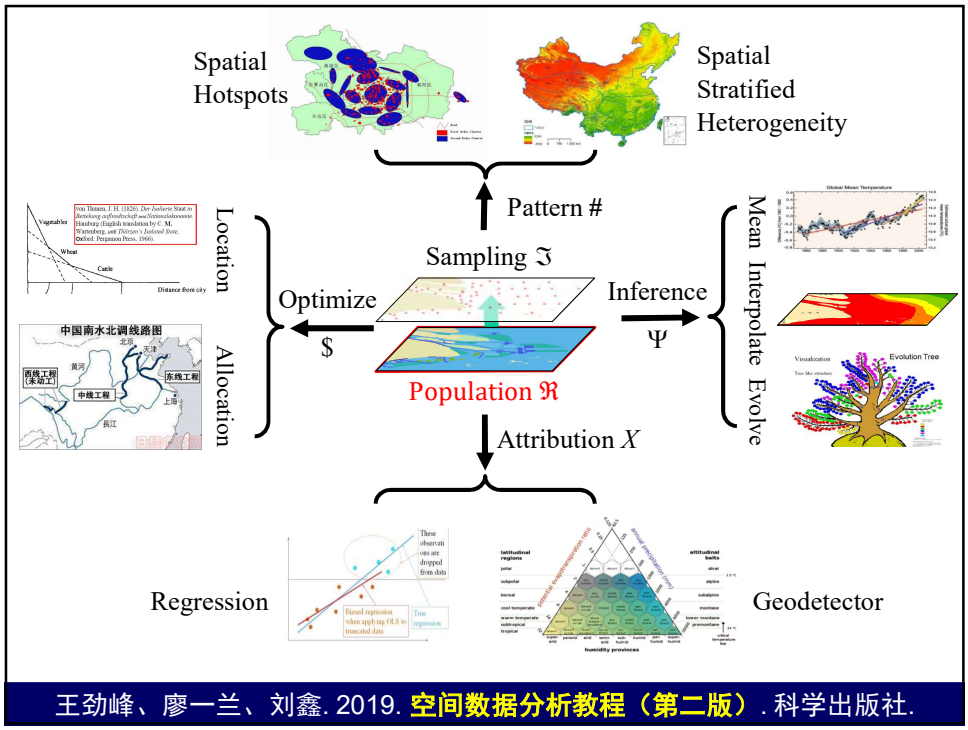
王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 科学出版社.

- Statistics vs. Mathematical model
- Statistics vs. Big data

Calculations are no longer an issue or that data sets can be “big,” there is nothing more to statistics than learning the syntax for your software or which options to “click.” (190213)

Methods for the reduction of data (Everitt BS, Skrondal A. 2010. The Cambridge Dictionary of Statistics, 4th Edition. Cambridge: Cambridge University Press)

Conclusion



王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 科学出版社.

空间统计学

地理特性 / 空间统计	空间**自相关** (Autocorrelation) (1950 ~ 1995)	空间**异质性** (Heterogeneity)	
		局域 (Local) (1992 ~ 2000)	**分异** (Stratified) (2010 ~)
度量	Moran's I；半变异函数	Gi; LISA; SatScan	Geodetector *q*
归因	SAR/MA/CAR	GWR	Geodetector *q*
插值	Kriging; IDW; Spline	N/A	Sandwich,MSN,Bshade,SPA
模式	“三位一体”空间统计模式		

General Reference for the Course

文章

[1] 王劲峰 等. 2014. 地理学时空数据分析方法. **地理学报** 69(9): 1326-1345.
[2] 王劲峰，徐成东. 2017. 地理探测器：原理与展望. **地理学报** 72(1): 116-134.
[3] Wang JF, et al. 2012. A review of spatial sampling. **Spatial Statistics** 2(1): 1-14.

书

[4] 盛　骤 等. 2008. **概率论与数理统计**. 北京: 高等教育出版社（基础）
[5] 王劲峰、廖一兰、刘鑫. 2019. **空间数据分析教程（第二版）**. 北京: 科学出版社
[6] Cressie N. 1993. **Statistics for Spatial Data**. Willey
[7] Anselin L. 1988. **Spatial Econometrics: Methods and Models**. Dordrecht.
[8] Haining R. 2003. **Spatial Data Analysis: Theory and Practice**. Cambridge Univ Press.

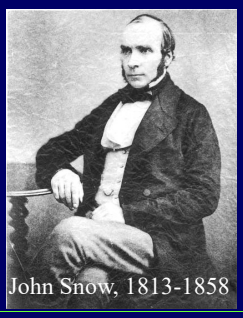
杂志（理论）

[9] Spatial Statistics (Elsevier)
[10] Geographical Analysis (Wiley)
[11] Journal of Geographical Systems (Springer)
[12] International Journal of Geographical Science (Taylor)

软件

[13] 经典统计: Excel, Matlab, SPSS, SAS, R
[14] 空间统计: ArcGIS, GeoDA, GeoBUGS, SatScan, CrimeStat, Geodetector, SSSampling
(kriging, 空间回归, BHM, 热点探测, 层次聚类, 非线性归因, 空间抽样)

Two Wheels

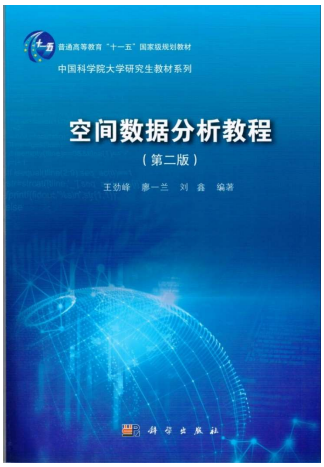

John Snow, 1813-1858

Spatial Epidemiology
- Physical determinants
- Man-made pollution
- Nutrition processes
- Heredity and habits
- Communicable


Apr 2006 London

Spatial Economics
- distance and cost
- Spillover


2010-8-21, Jonkoping, Sweden



网址请见书各章

首页
关于本书
数据下载
软件下载
购买此书
和第一版重要区别

数据集下载

第01章 第03章
第05章 第06章
第07章 第09章
第10章 第11章
第12章 第13章
第14章 第15章
第16章 第17章
第18章 第20章

Thanks