Kriging Interpolation Theory

Xiangyun Huang

China University of Mining & Technology,Beijing

June 2, 2016

Contents

Motivation

Gold is found in the conglomerate strata of the younger members of the Supergroup. The abundance of this gold is without equal anywhere else in the world. Over 50 000 tons have been mined from these rocks since this precious metal was first discovered here in 1886. This accounts for approximately 50% of all the gold ever mined on earth. [1]

Background

Danie G. Krige (1919.8-2013.3) was a South African Mining Engineer and professor at the University of the Witwatersrand who pioneered the field of geostatistics [2] and practiced the science of evaluating mineral resources for mining purposes. The technique of Kriging is named after him.

Herbert S. Sichel (1915-1995) was a statistician who developed the Sichel-t Estimator for the Log-normal distribution's t-statistic and also made great leaps in the area of the Generalized Inverse Gaussian Distribution. [3] He pioneered the science of geostatistics [4] with **Danie G. Krige** in the early 1950s and also was well recognised in the field of statistical linguistics. [5]

Georges Matheron (1930.12-2000.8) was a French mathematician and geologist who well known as the founder of geostatistics and Kriging. [6]

Two distributions

Log-normal distribution

$$\mathcal{N}(\ln x;\mu,\sigma)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}},x>0$$

Generalized inverse Gaussian distribution

$$f(x)=\frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})}x^{p-1}e^{-(ax+b/x)/2},x>0$$

where K_p is a modified Bessel function of the second kind,$a>0,b>0$ and p is a real parameter.

$$I_\alpha(t)=i^{-\alpha}J_\alpha(it)=\sum_{m=0}^{\infty}\frac{1}{m!\Gamma(m+\alpha+1)}(\frac{t}{2})^{2m+\alpha}$$

$$K_\alpha(t)=\frac{\pi}{2}\frac{I_{-\alpha}(t)-I_\alpha(t)}{\sin(\alpha\pi)}$$

文献综述

- Theories
  - Weighted Average(closely related to regression analysis)
    Best Linear Unbiased Estimator(BLUE)
    Gauss-Markov Theorem
  - Polynomial Trend Surfaces
    Generalized Least Squares Polynomial Curve Fitting
  - Bayesian Inference [7]
- Applications
  - Environmental science [8]
  - Hydrogeology [9]
  - Mining [10]
  - computer experiments and optimization [11]

国内研究现状

引进国外的研究方法，对国内的具体情况进行实证研究和方法的比较研究，少量在方法、算法、模型上有初创性的工作.(中国知网)

Contents

kriging 插值原理

设 x_0 为未观测的需要估值的点，$x_1, x_2, ..., x_n$ 为其周围的观测点，相应的观测值为 $y(x_1), y(x_2), ..., y(x_n)$. x_0 处的估计值记为 $\tilde{y}(x_0)$，它由相邻观测点的已知观测值加权求得:

$$\tilde{y}(x_0) = \sum_{i=1}^{N} \lambda_i y(x_i) \tag{1}$$

这里，λ_i 为待定的加权系数.
两个约束条件：无偏估计和估计方差最小
设估值点的真值为 $y(x_0)$，由于空间变异性的存在，$y(x_i), \tilde{y}(x_0), y(x_0)$ 都可视为随机变量.

$$E[\tilde{y}(x_0) - y(x_0)] = 0 \tag{2}$$

推导出

$$\sum_{i=1}^{N} \lambda_i = 1 \tag{3}$$

kriging 插值原理 (续 I)

估计的方差最小

$$\min_{\lambda_i} D[\tilde{y}(x_0) - y(x_0)]$$

经推导

$$D[\tilde{y}(x_0) - y(x_0)] = -\sum_{i=1}^{N}\sum_{j=1}^{N}\lambda_i\lambda_j\gamma(x_i, x_j) + 2\sum_{i=1}^{N}\lambda_i\gamma(x_i, x_0) \quad (4)$$

其中 $\gamma(x_i, x_j)$ 表示以 x_i 和 x_j 两点间的距离作为间距 h 时参数的半方差值，$\gamma(x_i, x_0)$ 表示以 x_i 和 x_0 两点间的距离作为间距 h 时参数的半方差值. 观测点和估计点的位置是已知的，相互之间的距离已知，只要有所求参数的半方差 $\gamma(h)$，便可以求得各个 $\gamma(x_i, x_j)$ 和 $\gamma(x_i, x_0)$ 值.

kriging 插值原理 (续 II)

在满足 (3) 式的约束条件下，求目标函数 (4) 式达到最小，用拉格朗日乘数法，可导出优化问题的解

$$\sum_{j=1}^{N} \lambda_i \gamma(x_i, x_j) + \mu = \gamma(x_i, x_0), i = 1, 2, ..., N \tag{5}$$

其中 μ 是拉格朗日乘子. 求解由 (3) 和 (5) 式组成的 $n+1$ 阶线性方程组，可得到 n 个加权系数 λ_i 和拉格朗日乘子 μ.

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1N} & 1 \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2N} & 1 \\ \vdots & \vdots & \ddots & \vdots & \\ \gamma_{N1} & \gamma_{N2} & \cdots & \gamma_{NN} & 1 \\ 1 & 1 & \ldots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu \end{pmatrix} = \begin{pmatrix} \lambda_{10} \\ \lambda_{20} \\ \vdots \\ \lambda_{N0} \\ 1 \end{pmatrix} \tag{6}$$

其中 γ_{ij} 为 $\gamma(x_i, x_j)$ 的简写

kriging 插值原理 (续 III)

求得各 λ_i 和 μ 值后，由 (1) 式可求得 x_0 点的最佳线性无偏估计值 $\tilde{y}(x_0)$，由 (4) 式可求得该处估计的方差最小值 σ^2_{min}，另外最小方差值还可以由下式求得

$$\sigma^2_{min} = \sum_{j=1}^{N} \lambda_i \gamma(x_i, x_0) + \mu \tag{7}$$

Contents

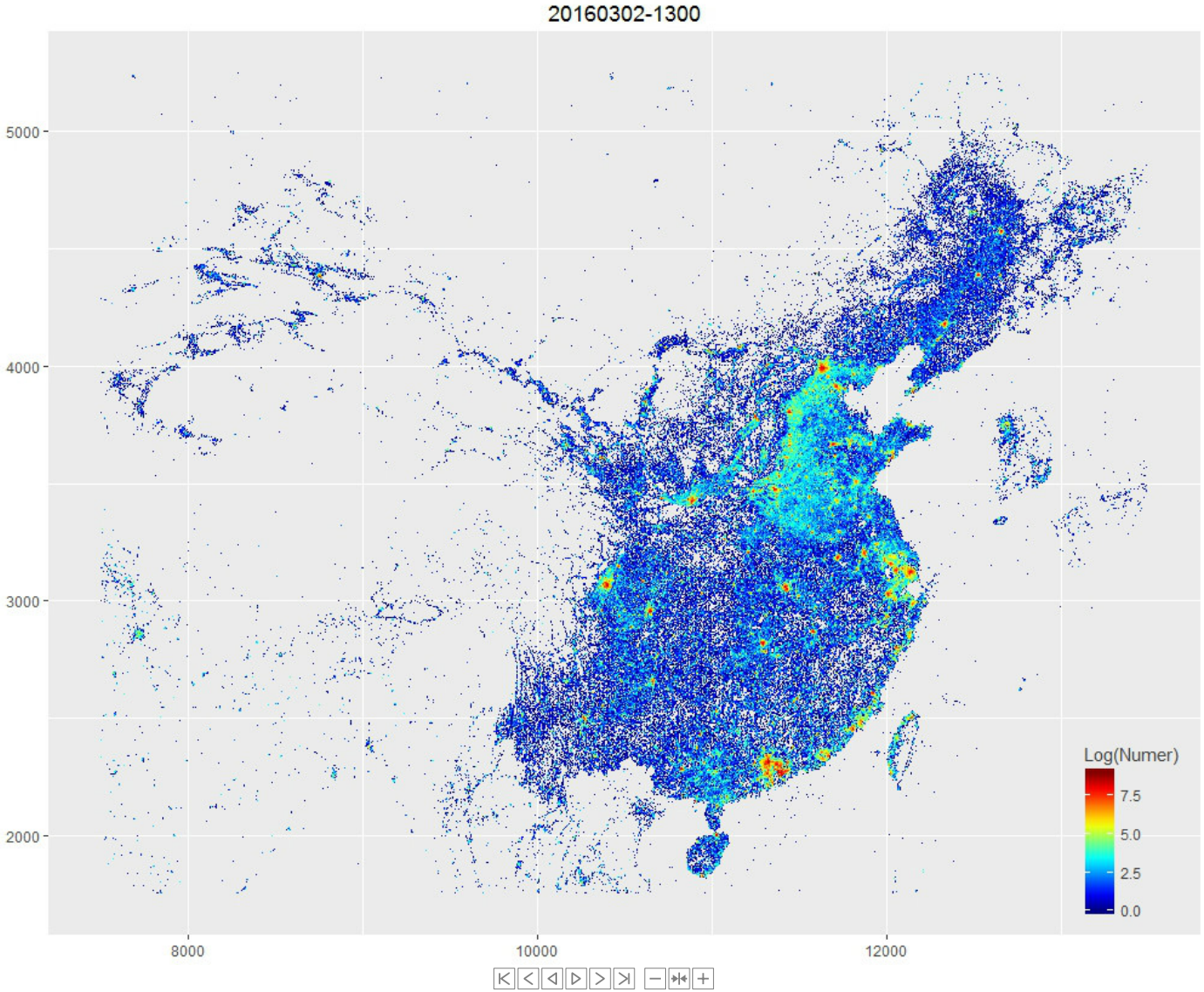


Figure: Where are the Chinese[1]

[1] 源自微博统计人的世界 @ 王江浩 CAS

更多例子

- 王江浩 (中国科学院地理科学与资源研究所)
  研究方向：GIS、RS、Spatio-temporal statistics、Geo-big data、Visualization[2]
  利用 NASA GMAO 资料，可视化了 2014 年 1 月逐小时的地表 PM2.5 浓度，可以直观地看出来 PM2.5 的传播过程[3].
- 袁晓如 (北京大学)
  研究方向：可视化和可视分析
  利用北京市各监测站点数据，可视化北京空气质量[4].
- 陈为 (浙江大学) 研究方向：可视化和可视分析
- 更多

[2] http://jianghao.wang/
[3] http://weibo.com/p/230444f8bab0c90e81cf0d2335fd0f6cc55e17
[4] http://vis.pku.edu.cn/wiki/

From Home to Work

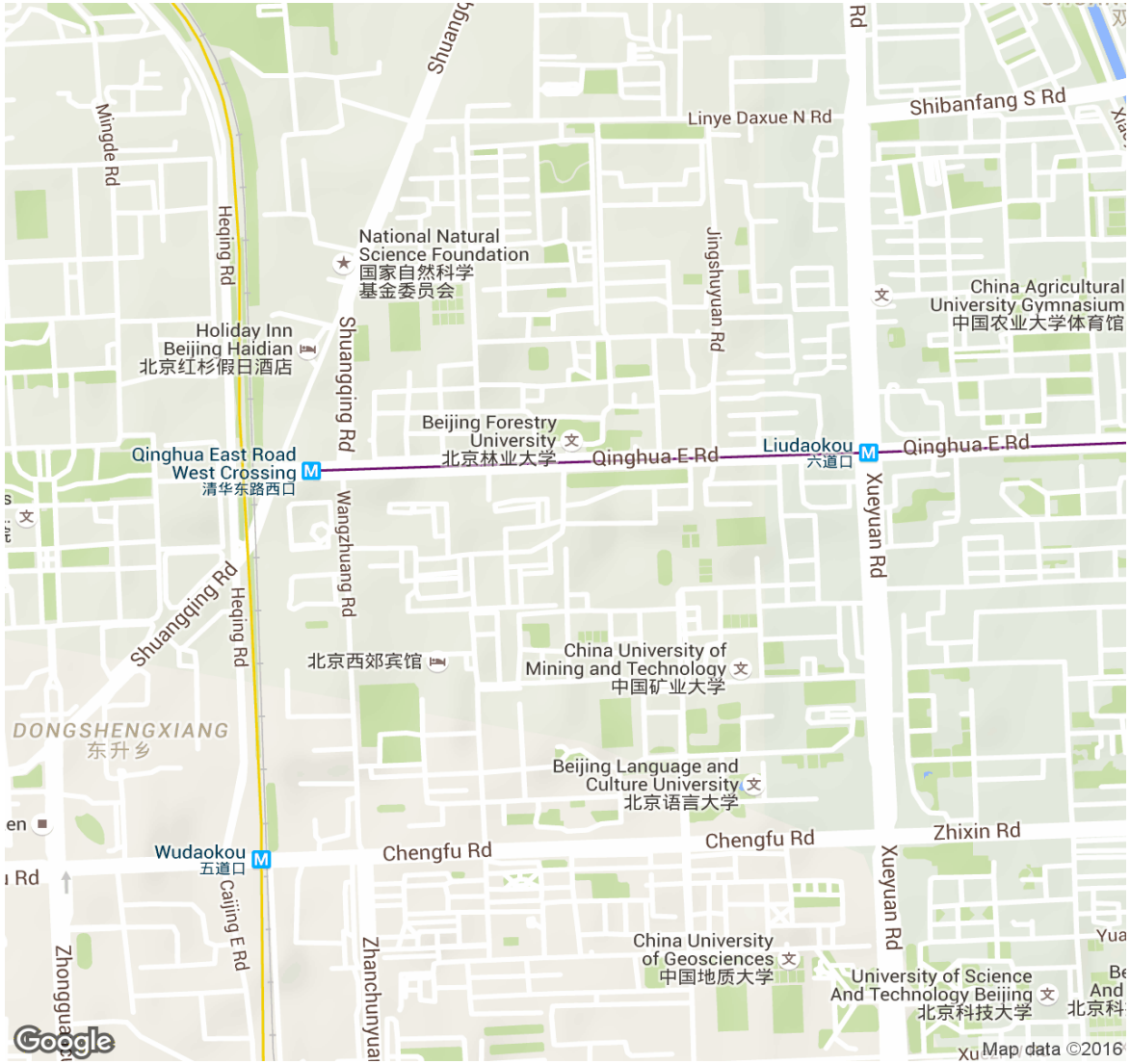
Figure: London: The Information Capital

Contents

空间数据可视化

- API(Google Maps,Baidu Maps,OpenStreetMap)
  - 定位 (坐标数据)
  - 描点或区域 (空间变量)
  - 配色等 (可视分析)
- Non-API
  - 网格或三维坐标系 (坐标数据)
  - 加数据层 (空间变量)
  - 配色等 (可视分析)

坐标定位


Mingde Rd
Heqing Rd
National Natural
Science Foundation
国家自然科学
基金委员会
Linye Daxue N Rd
Shibanfang S Rd
Jingshuyuan Rd
China Agricultural
University Gymnasium
中国农业大学体育馆
Holiday Inn
Beijing Haidian
北京红杉假日酒店
Shuangqing Rd
Beijing Forestry
University
北京林业大学
Qinghua East Road
West Crossing
清华东路西口
Qinghua E Rd
Liudaokou
六道口
Xueyuan Rd
Wangzhuang Rd
北京西郊宾馆
China University of
Mining and Technology
中国矿业大学
DONGSHENGXIANG
东升乡
Beijing Language and
Culture University
北京语言大学
Wudaokou
五道口
Chengfu Rd
Zhixin Rd
Caijing E Rd
Zhongguancun
Zhanchunyuan
China University
of Geosciences
中国地质大学
University of Science
And Technology Beijing
北京科技大学
Google
Map data ©2016

描点

Figure: 莆田系医院的位置可视化[5]

[5] http://www.xueqing.tv/cms/article/199

Scientific Visualization of Spatio-Temporal Data

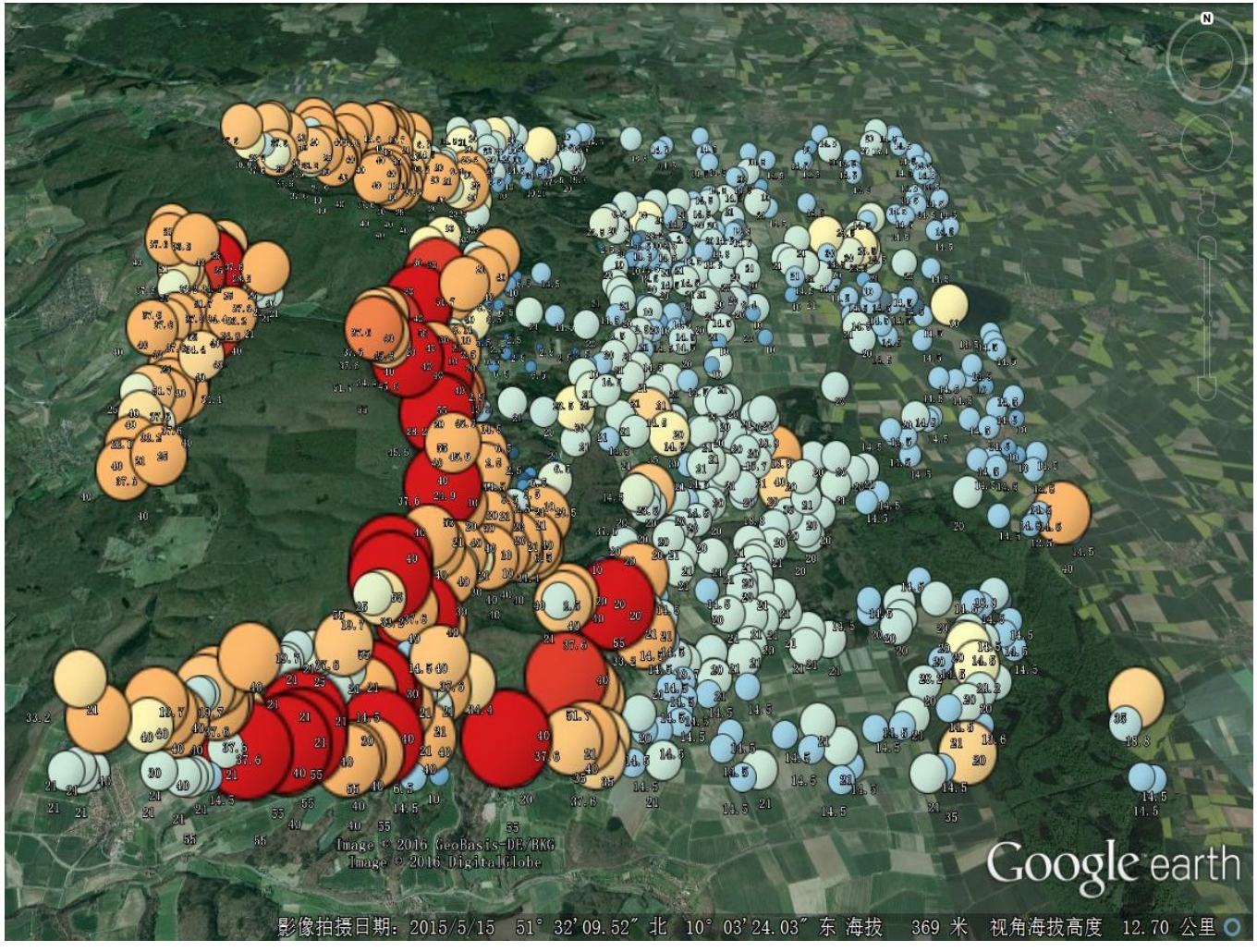
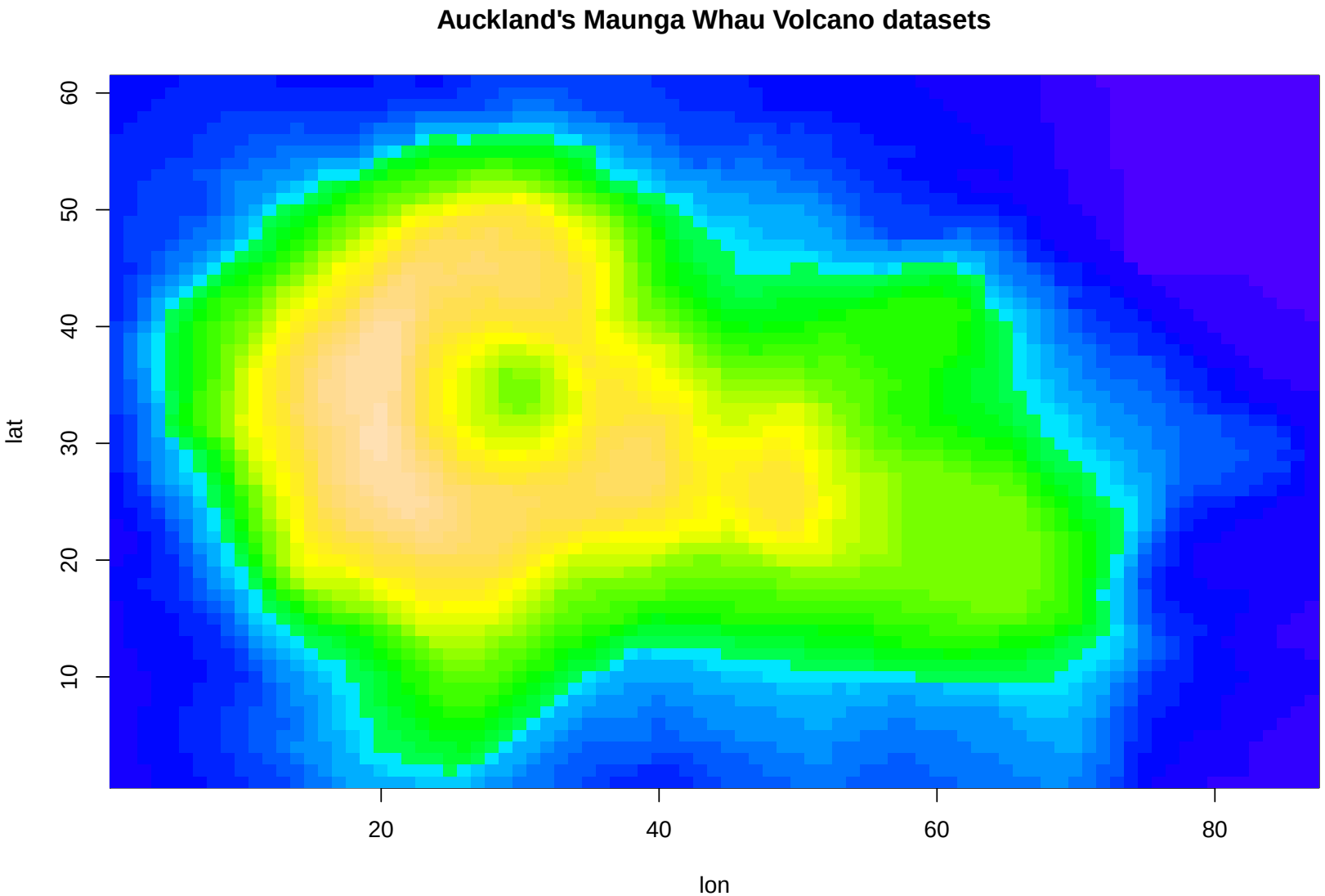


Figure: Ebergötzen soil mapping data set[12]

网格


Auckland's Maunga Whau Volcano datasets
lat
10
20
30
40
50
60
lon
20
40
60
80

一些工具

- 入门级
  - RgoogleMaps [13]、plotKML [12] 和 ggmap [14] 包使用 Google Maps API
  - REmap [15] 和 baidumap [16] 包使用 Baidu Maps API
  - tmap [17](Thematic Maps 专题地图) 和 choroplethr [18](Choropleth Maps 区域地图)
- 进阶级
  - DiceKriging [19] 和 DiceOptim [20] 等
  - Echarts [6]和 D3 [7]等

[6] http://echarts.baidu.com/
[7] https://d3js.org/

Contents

Leaflet

Leaflet[8] is one of the most popular open-source JavaScript libraries for interactive maps. It's used by websites ranging from The New York Times[9] and The Washington Post[10] to GitHub[11] and Flickr[12], as well as GIS specialists like OpenStreetMap[13], Mapbox[14], and CartoDB[15].

[8] http://leafletjs.com/

[9] http://www.nytimes.com/projects/elections/2013/nyc-primary/mayor/map.html

[10] http://www.washingtonpost.com/sf/local/2013/11/09/washington-a-world-apart/

[11] https://github.com/blog/1528-there-s-a-map-for-that

[12] https://www.flickr.com/map

[13] http://www.openstreetmap.org/

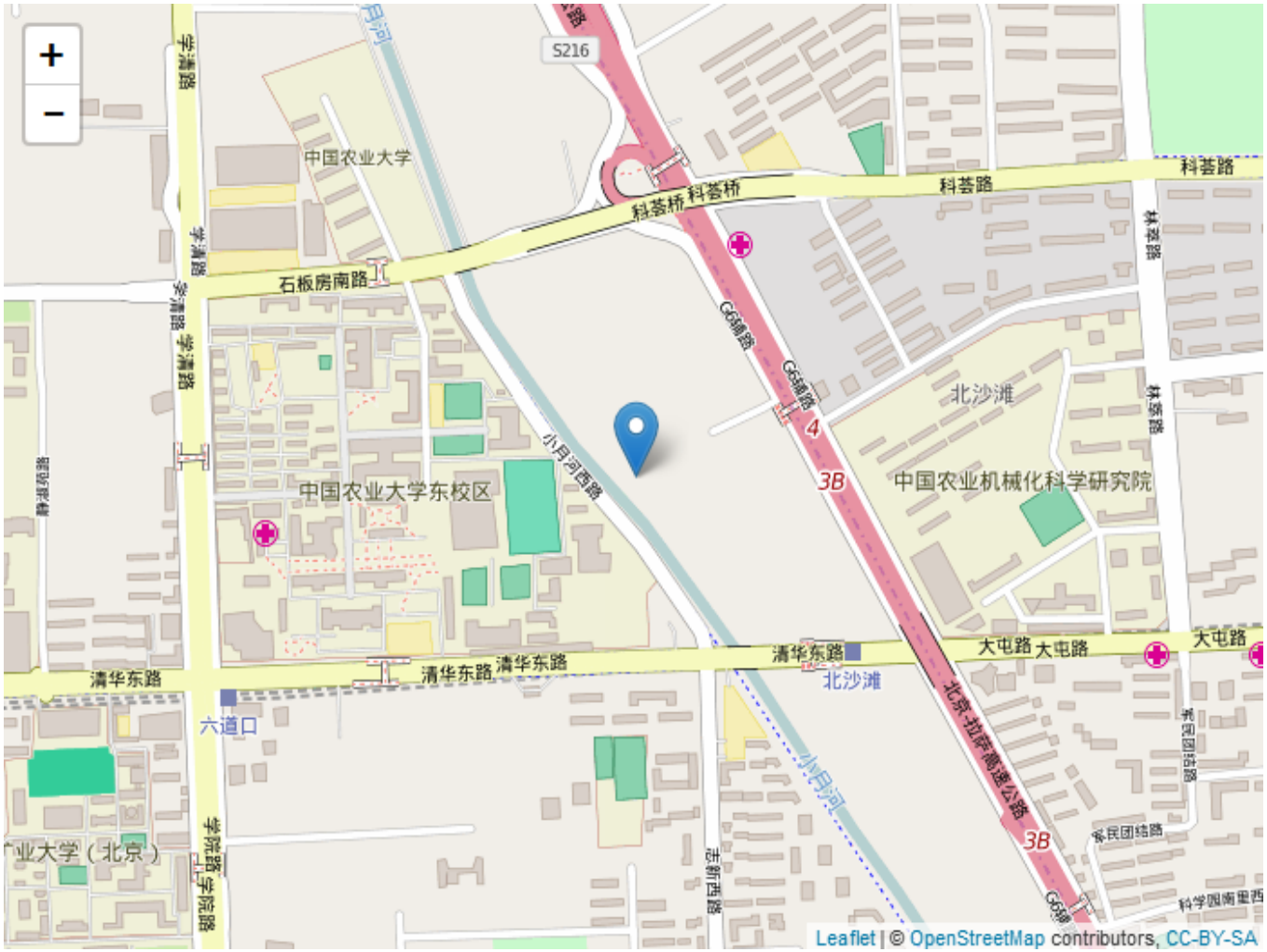
[14] https://www.mapbox.com/

[15] https://cartodb.com/

获取学校坐标

```
library(leaflet)
library(baidumap)
## getcoordinate return vector
loc<-getCoordinate('中国矿业大学(北京)', formatted = T) # character
library(ggmap)
## geocode return dataframe
ggloc<-geocode("China University of Mining and Technology,Beijing")
## plot
CUMTB <- leaflet() %>%
  addTiles() %>%  # Add default OpenStreetMap map tiles
  addMarkers(lng=loc[1], lat=loc[2],
             popup="China University of Mining and Technology,Beijing")
CUMTB  # Print
ggCUMTB <- leaflet() %>%
  addTiles() %>%  # Add default OpenStreetMap map tiles
  addMarkers(lng=ggloc$lon, lat=ggloc$lat,
             popup="China University of Mining and Technology,Beijing")
ggCUMTB  # Print
```

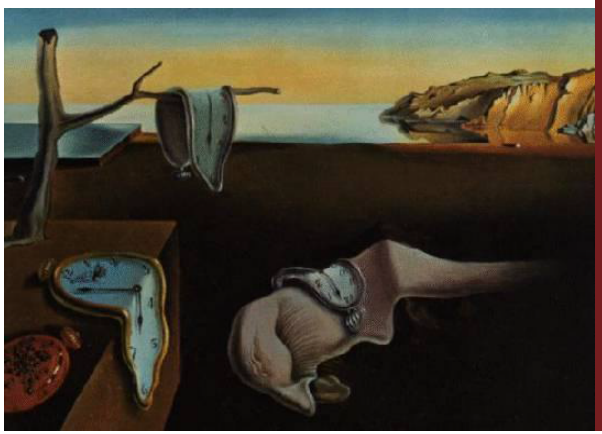
BaiduMap

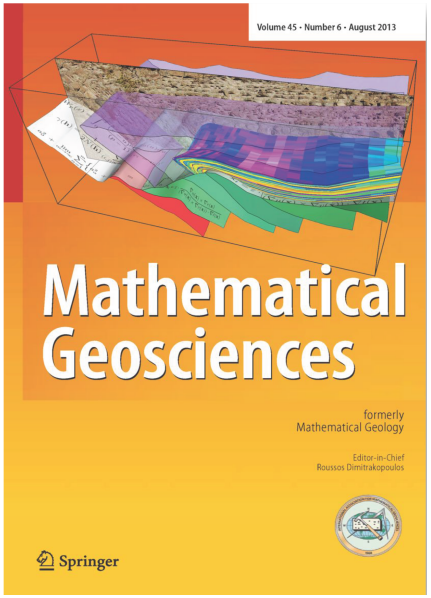

S216
中国农业大学
科荟桥 科荟路
科荟路
石板房南路
学清路
林荟路
北沙滩
中国农业机械化科学研究院
中国农业大学东校区
小月河西路
清华东路
六道口
大屯路
北京-拉萨高速公路
小月河
Leaflet | © OpenStreetMap contributors, CC-BY-SA

GoogleMap


林业大学北路
石板房南路
北林家属区
半导体所
中国农业大学东校区
白颐大道
清华东路 清华东路
清华东路西口
六道口
中国矿业大学（北京）
北京语言大学
成府路 成府路
志新路 志新路
Leaflet | © OpenStreetMap contributors, CC-BY-SA

References





Journal of Statistical Software [16] JSTOR [17] Mathematical Geosciences [18]

- ✏ http://www.sci-hub.cc/
- ✏ https://arxiv.org/
- ✏ http://gen.lib.rus.ec/

[16] http://www.jstor.org/
[17] http://www.jstatsoft.org/
[18] http://link.springer.com/journal/11004

References I

N. Norman and G. Whitfield.

Geological journeys: A traveller's guide to south africa's rocks and landforms.
2006.

D. G. Krige.

A statistical approach to some basic mine valuation problems on the witwatersrand.
Journal of the Chemical Metallurgical & Mining Society of South Africa, 94(3):95–111, 1951.

H. S. Sichel.

Asymptotic efficiencies of three methods of estimation for the inverse gaussian-poisson distribution.
Biometrika, 69(2):467–472, 1982.

Herbert S. Sichel.

The estimation of the parameters of a negative binomial distribution with special reference to psychological data.
Psychometrika, 16(1):107–127, 1951.

H. S. Sichel.

On a distribution law for word frequencies.
Journal of the American Statistical Association, 70(351):542–547, 1975.

Georges Matheron.

Principles of geostatistics.
Economic Geology, 58(8):1246–1266, 1963.

References II

Mark S.Handcock and Michael L.Stein.
A bayesian analysis of kriging.
Technometrics, 35(35):403–410, 1993.

Hanefi Bayraktar and F. Sezer. Turalioglu.
A kriging-based approach for locating a sampling site—in the assessment of air quality.
Stochastic Environmental Research & Risk Assessment, 19(4):301–305, 2005.

Chiles Jeanpaul, Delfiner Pierre, and John Wiley.
Geostatistics modeling spatial uncertainty.
Journal of the American Statistical Association, 95, 2000.

Andrew Richmond.
Financially efficient ore selections incorporating grade uncertainty.
Mathematical Geology, 35(2):195–215, 2003.

Jian Xia Zhang, Yi Zhong Ma, and Lian Yan Zhu.
Multiobjective Simulation Optimization Using Stochastic Kriging.
Atlantis Press, 2016.

Tomislav Hengl, Pierre Roudier, Dylan Beaudette, and Edzer Pebesma.
plotkml: Scientific visualization of spatio-temporal data.
Journal of Statistical Software, 63(5):1–25, 2015.

Markus Loecher and Karl Ropkins.
Rgooglemaps and loa: Unleashing r graphics power on map tiles.
Journal of Statistical Software, 63(4):1–18, 2015.

References III

David Kahle and Hadley Wickham.
ggmap: Spatial visualization with ggplot2.
The R Journal, 5(1):144–161, 2013.

Dawei Lang.
REmap: Create html maps by Echarts.
R package version 0.3.2.

yalei Du.
baidumap: A package for spatial visualization with Baidu.
R package version 0.2.

Martijn Tennekes.
tmap: Thematic Maps, 2016.
R package version 1.4-1.

Ari Lamstein and Brian P Johnson.
choroplethr: Simplify the Creation of Choropleth Maps in R, 2016.
R package version 3.5.2.

Olivier Roustant, David Ginsbourger, and Yves Deville.
Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization.
Journal of Statistical Software, 51(1):1–55, 2012.

References IV

D. Ginsbourger, V. Picheny, O. Roustant, with contributions by C. Chevalier, S. Marmin, and T. Wagner.
DiceOptim: Kriging-Based Optimization for Computer Experiments, 2015.
R package version 1.5.

Thank You