Spatial Generalized Linear Mixed Models with Application to Prevalence Mapping

空间广义线性混合模型及其在预测流行病中的应用
2015 级硕士学位论文答辩

学生：黄湘云　　导师：李再兴

专业：统计学　方向：数据分析与统计计算

中国矿业大学（北京）

China University of Mining & Technology, Beijing

2017 年 11 月 15 日/ 逸夫楼

Outline

例例 例 例 例

1. radionuclide concentrations on Rongelap Island
2. childhood malaria in the gambia
3. Loa loa prevalence in Cameroon and surrounding areas

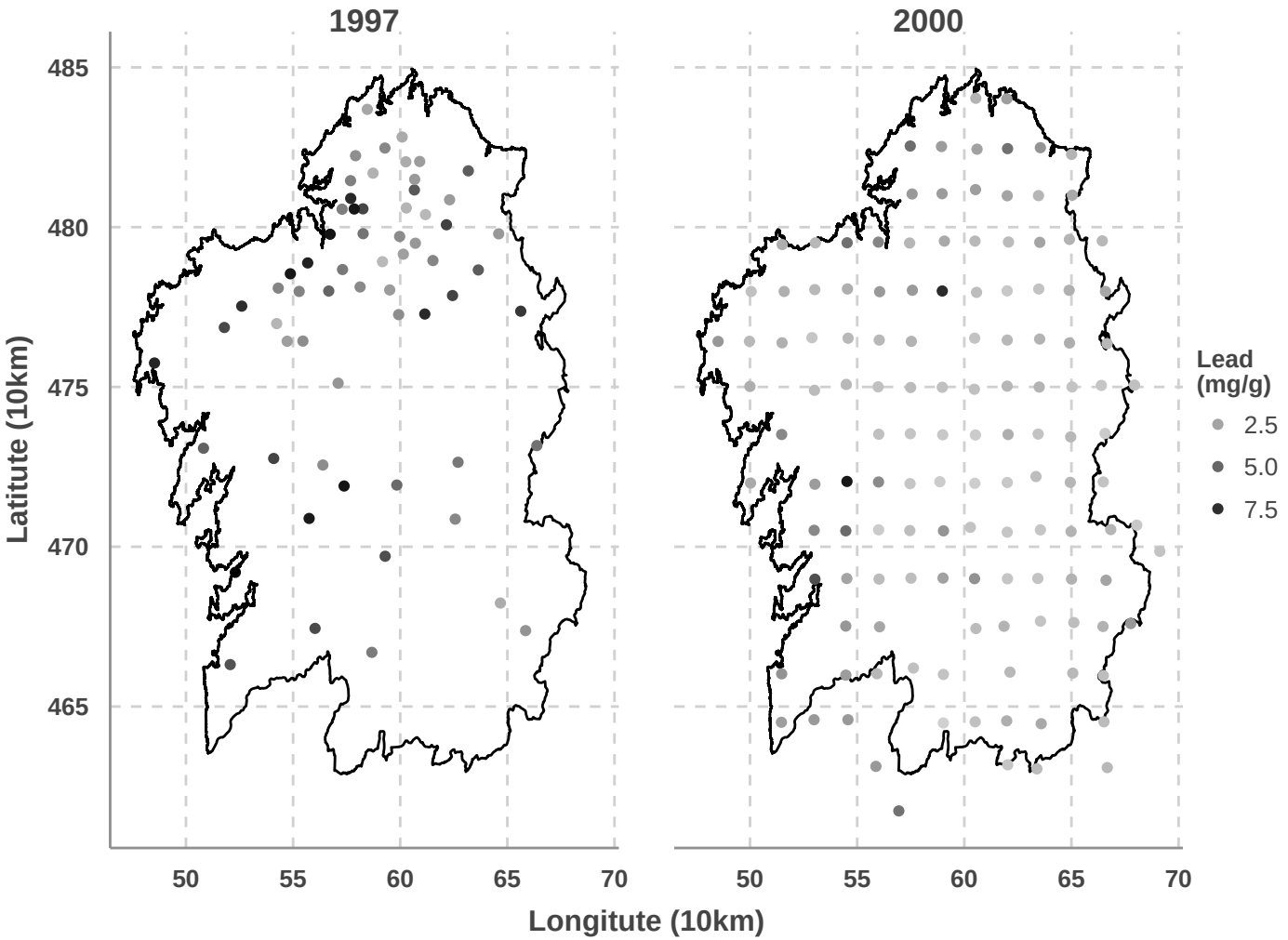
Introduction

Diggle *et al.* (2002)

- First item in the list
- Second item
- and so on
  - First item in the list
  - Second item
  - and so on
- the effects of child level covariates (age and bed net use)
- village level covariates (the primary health care and greenness of surrounding vegetation)
- separate components for residual spatial
- non-spatial extrabinomial variation

$\mathbb{R}^n$

$$\log\{p_{ij}/(1-p_{ij})\} = \alpha + \beta' z_{ij} + U_i + S(x_i)$$


1997
2000
485
480
475
470
465
Latitute (10km)
50
55
60
65
70
Longitute (10km)
Lead (mg/g)
2.5
5.0
7.5

The function f is given by

$$f(x) = 2x + \frac{x-7}{x^2+4}$$

for all real numbers x.
The roots of a quadratic polynomial $ax^2 + bx + c$ with $a \neq 0$ are given by the formula

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The roots of a cubic polynomial of the form $x^3 - 3px - 2q$ are given by the formula

$$\sqrt[3]{q + \sqrt{q^2 - p^3}} + \sqrt[3]{q - \sqrt{q^2 - p^3}}$$

where the values of the two cube roots must are chosen so as to ensure that their product is equal to p.

Multiple prevalence surveys

Sample n_i individuals,observe Y_i positives,$i = 1, 2, \cdots, m$

$$Y_i \sim \text{Bin}(n_i, p_i)$$

Extra-binomial variation

Sample n_i individuals,observe Y_i positives,$i = 1, 2, \cdots, m$

$$Y_i|d_i, U_i \sim \text{Bin}(n_i, p_i) \quad \log\{p_i/(1-p_i)\} = d_i'\beta + U_i \quad U_i \sim N(0, \tau^2)$$

notations: Spatial Generalized Linear Mixed Models (SGLMM)

- Latent spatially correlated process
  Stationary Gaussian Process: $S(x) \sim \text{SGP}\{0, \sigma^2, \rho(u)\}$
  correlation function: e.g. $\rho(u) = \exp(-|u|/\phi)$
- Linear prediction (regression model)
  $d(x) =$ covariates at location x
  Linear prediction: $\eta(x) = d(x)'\beta + S(x)$
  Link function: logit $p(x) = \log\{\eta(x)/[1-\eta(x)]\}$
- Conditional distribution for positive proportion Y_i/n_i
  $Y_i|S(\cdot) \sim \text{Bin}(n_i, p(x_i))$ (binomial sampling)

Let **u**,**v** and **w** be three vectors in $\mathbf{R}^3$. The volume V of the parallelepiped with corners at the points **0**, **u**, **v**, **w**, $\mathbf{u}+\mathbf{v}$, $\mathbf{u}+\mathbf{w}$, $\mathbf{v}+\mathbf{w}$ and $\mathbf{u}+\mathbf{v}+\mathbf{w}$ is given by the formula

$$V = (\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}.$$

$$\cos(\theta + \phi) = \cos\theta \cos\phi - \sin\theta \sin\phi$$

$$M^{\perp} = \{f \in V' : f(m) = 0 \text{ for all } m \in M\}.$$

Thank You

参考文献 I

Diggle, Peter, Moyeed, Rana, Rowlingson, Barry, & Thomson, Madeleine. 2002. Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **51**(4), 493–506.

软件环境

R 3.4.2 rstan geoR geoRglm INLA