De Wijs Process and Modeling Disease Risk

DEBASHIS MONDAL

Department of Statistics,

University of Chicago,

Chicago

Acknowledgement: NSF award number: DMS-0906300

IMS NRC 2009

Outline for the talk

- **overview**
- Two approaches to disease mapping
  - **Gaussian Markov random fields**
  - **Geostatistical**
- Combining the above two approaches via **de Wijs Process**
- End with an example on **disease mapping**

Overview

Interest on **maps of risks** of non-infectious diseases over a **geographic region**.

These maps

- reveal heterogeneity
- indicate areas of high relative risk that may require screening and intervention
- help detect effects of socio-economic conditions and exposure to environmental hazards
- provide clues to the etiology of diseases
- use counts aggregated over zip-codes, counties or states to protect individual identities.

Examples: **Clayton and Kaldor (1987), Besag et al. (1991), Best et al. (2001), Kelsall and Wakefield (2002), Banerjee et al. (2004) and Rue and Held (2005)**....

Two statistical approaches are **Markov random fields** and **geostatistics**.

Example of irregular regions

Markov random field approach (Besag et al., 1991, Rue and Held, 2005)

Consider the **conditional Poisson model**

$$Y(A_i) \sim \text{independent Poisson}\Big(c(A_i)R(A_i)\Big),\ i = 1, \ldots, n, \tag{1}$$

with

$$\log R(A_i) = T(A_i)\beta + U(A_i) + \epsilon(A_i),$$

	$y(A_i)$	=	**observed number** of cases of disease in region (e.g., county) A_i
	$c(A_i)$	=	expected number of cases of disease in region A_i
	$R(A_i)$	=	**log relative risk** in region A_i
where	$T(A_i)$	=	area-specific covariate,
	β	=	covariate effect
	$\epsilon(A_i)$	=	**unstructured component** or residual
	$U(A_i)$	=	A Markov random field

Assumes two areas are **adjacent** or **neighbors** when they share a common boundary.

Conditionally, $U(A_i)$ follows $N\Big(\sum_{j\sim i} U(A_j)/l_i, \sigma_U^2/l_i\Big)$.

$j \sim i$ implies that area A_j is adjacent to A_i and l_i is number of neighbors of A_i.

Geostatistical approach (Kelsall and Wakefield (2002)

Consider the **conditional Poisson model** driven by a Cox process

$$Y(A_i) \sim \text{independent Poisson}\Big(\sum_k \int_{A_i} c_k(s) R_k(s)\, ds\Big), \; i = 1, \ldots, n. \tag{2}$$

with

$$R_k(s) = R(s), \quad R(A_i) = \int_{A_i} f_i(s) e^{T(s)\beta + U(s) + \epsilon(s)}\, ds, \tag{3}$$

where

k	$=$	population stratum, $R_k(s)$ log relative risk in stratum k, location s
$f_i(s)$	$=$	weighted average of the stratum-specific population density
$c(A_i)$	$=$	$\sum_k \int_{A_i} c_k(s)\, ds$ is expected number of cases of disease in region A_i
$T(s)$	$=$	location-wise covariate information
β	$=$	covariate effect
$\epsilon(s)$	$=$	Gaussian white noise process
$U(s)$	$=$	A continuum spatial process

Includes **area** and **population information**, but computationally challenging.

Combining Markov random fields with geostatistical approach

Let $\{X_{u,v} : (u,v) \in \mathcal{Z}^2\}$ be Gaussian with **conditional** means and variances

$$\mathrm{E}\,(X_{u,v}|\ldots) = \gamma(x_{u-1,v} + x_{u+1,v} + x_{u,v-1} + x_{u,v+1}), \quad \mathrm{var}\,(X_{u,v}|\ldots) = \sigma_X^2, \qquad (4)$$

(i) $2\gamma < \frac{1}{2} \;\Rightarrow$ stationary autoregression e.g. Lévy (1948), Besag (1974).

(ii) $2\gamma = \frac{1}{2} \;\Rightarrow$ **intrinsic** case: Künsch (1987), Besag & Kooperberg (1995).

Limit diagram for stationary and intrinsic case

$2\gamma \to \frac{1}{2}$

Stationary autoregression ——————> intrinsic autoregression

| |

| |

$\vee$ $\vee$

Bessel_0 process ——————————-> De Wijs process

Idea is to use exponential functional of the de Wijs process.

In other words, use $e^{X_{u,v}}$ to **approximate** $U(s)$!!!

Combining Markov random fields with geostatistical approach

Embed study region in a fine mesh

$$\mathcal{L} = \{(u, v) : u = 1, \dots I, v = 1, \dots, J\}, \tag{5}$$

Each grid point (u, v) is associated with a regular cell $D(u, v)$. Use

$$A_i = \cup_{(u,v)\in A_i} D(u, v). \tag{6}$$

Let $m(A_i)$ be equal to the number of cells $D(u, v)$ in A_i. Consider

$$f_i(u, v) = \int_{D(u,v)} f_i(s), \, ds.$$

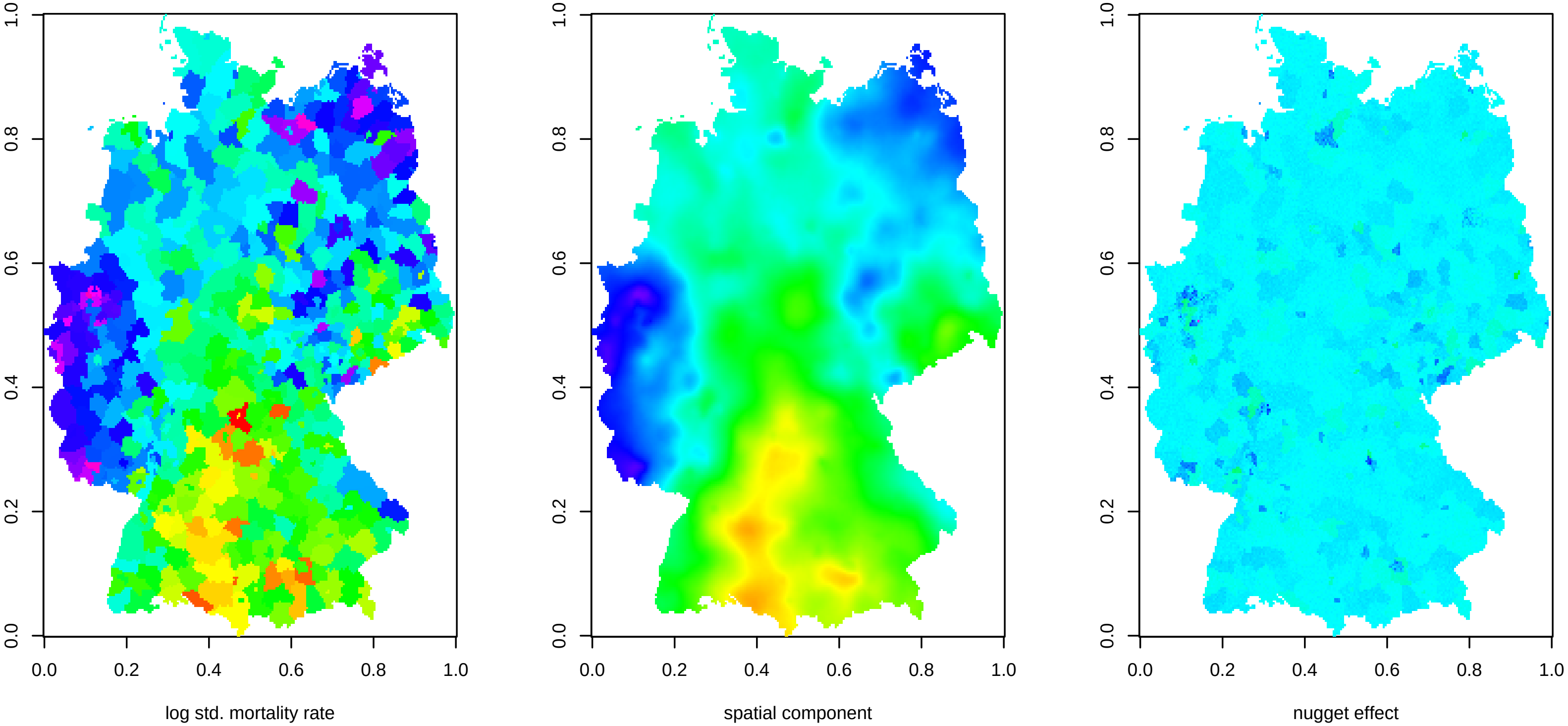
Finally, use

$$R(A_i) = \sum_{(u,v)\in A_i} f_i(u, v) e^{T(u,v)\beta + X_{u,v} + \epsilon_{u,v}}. \tag{7}$$

The **continuum random field theory** for another day!

This theory relies on **generalized Gaussian process**, **subordinate random fields** and the famous **Dobrushin's expansion**.

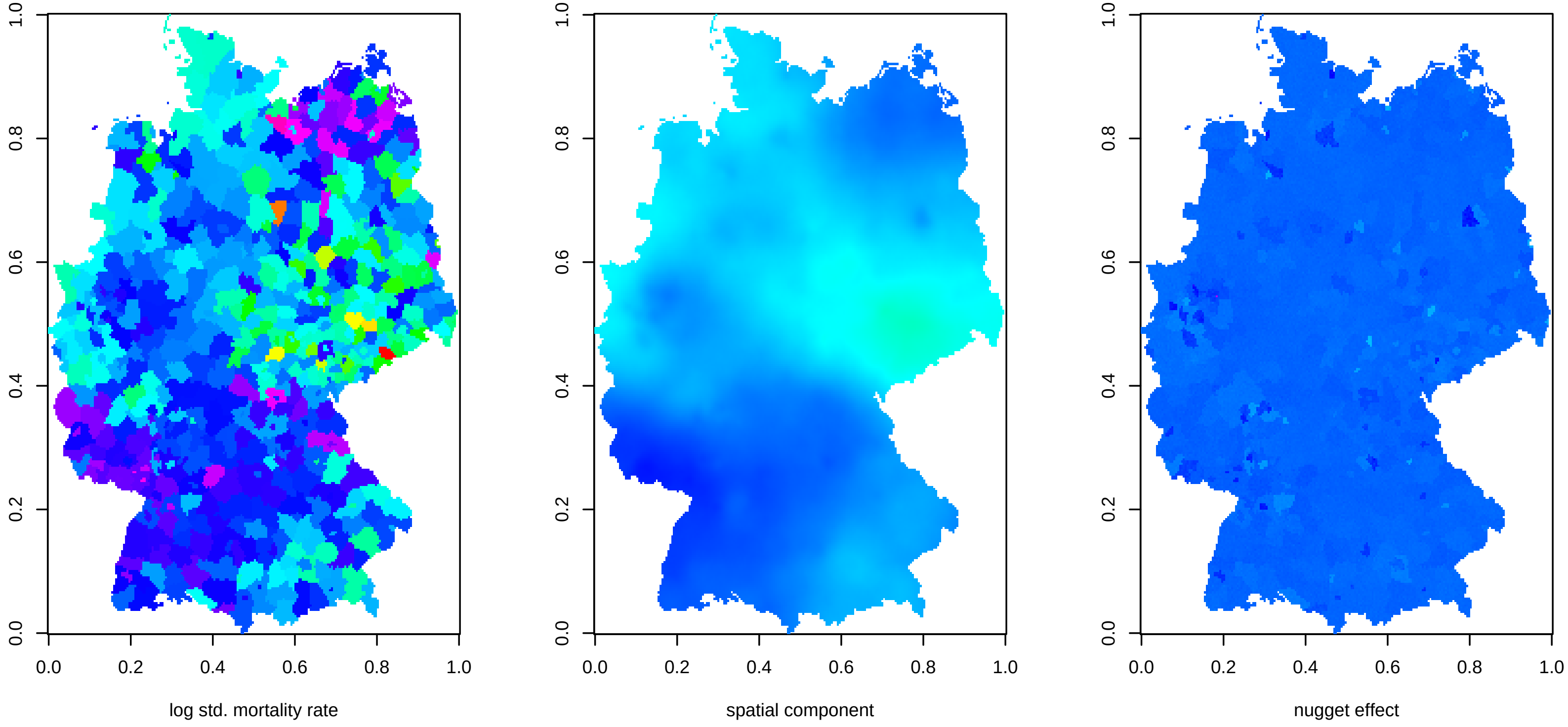
Lung cancer in Germany



Data in the period 1986-1990. 544 districts embeded in 289×214 girds.

See also Rue and Held, 2005.

Oral cancer in Germany



Data during the years 1986-1990. 544 districts embeded in 289 $\times$ 214 girds.

Joint disease mapping can also be done!

REFERENCES

Banerjee, S., Carlin, B. P. and Gelfrand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data.* Chapman and Hall.

Besag, J. E. and Mondal, D. (2005). First–order intrinsic autoregressions and the de Wijs process. *Biometrika*, **92**, 909–920.

Besag, J. E., York, J. and Molli, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.*, **43**, 1–59.

Matheron, G. (1971) *The Theory of Regionalized Variables and its Applications.* Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau, no. 5.

Mondal, D. (2009). Generalized Gaussian Markov random fields and modeling disease risk. Submitted.

Mondal, D. (2005). Variogram calculations for first-order intrinsic autoregressions and the de Wijs process. Technical Report 479, Department of Statistics, University of Washington.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields. Theory and applications.* Chapman and Hall.