Example: Stomach cancer in Slovenia

Data: - Slovenia is partitioned into 194 municipalities
- For each municipality, we have (years 1995-2001)
  - observed count of stomach cancer cases (O_i)
  - Expected # cases (E_i), using indirect standardization
  - A measure of socioeconomic status, centered and scaled (SEC_i)

Reich et al (2006)

Model: $O_i \sim \text{Poisson}(E_i e^{\mu_i})$

$$\mu_i = \beta_{SEC} SEC_i + S_i + H_i$$

H_i: iid $N(0, \frac{1}{\tau_h})$

S_i: CAR, precision τ_s; "neighbors" = share a boundary

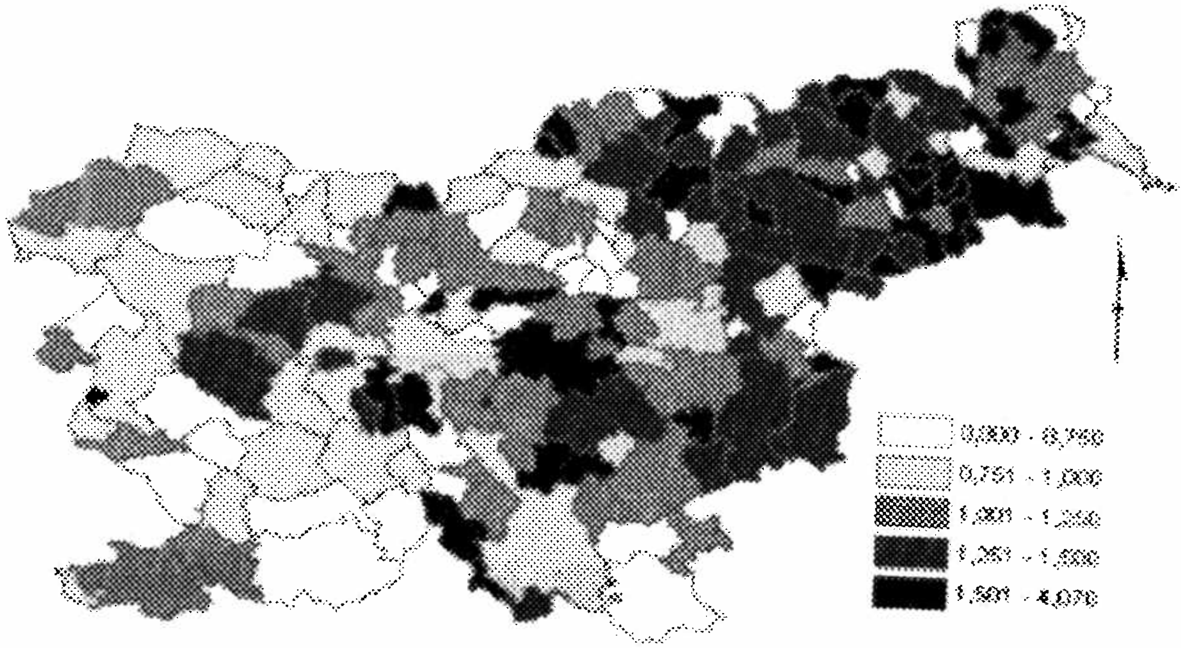
Model	DIC	p_D	β_{SEC} post'r 95% interval
No CAR, no Hetero	1153.0	2.0	(-0.175, -0.098)
CAR, Hetero	1081.5	62.3	(-0.100, 0.057)

With 62.3 DF in the fit (out of 194), the fit must be pretty close to the data

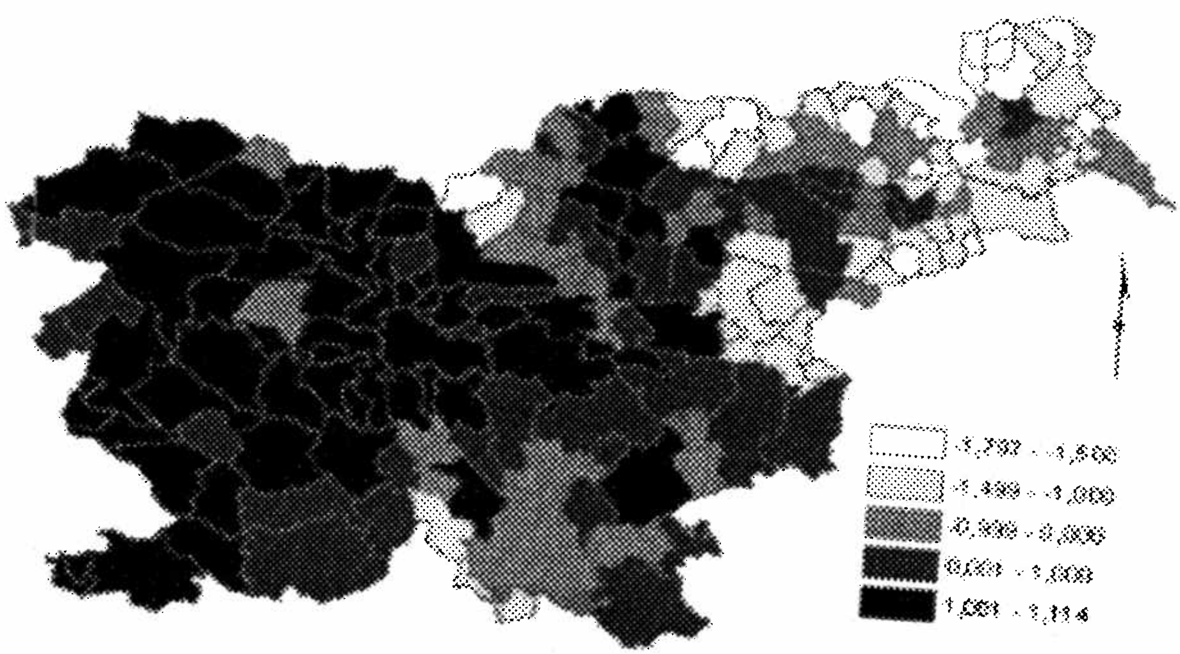
VAM T2 add 1 8/11/08

Here's a picture of the data:

O_i / E_i



SE_i



VAM T2 add 2 8/11/08

Penalized spline model (Nick Salkowski class project):

$O_i \sim \text{Poisson}(E_i e^{\mu_i})$

$\mu_i = \beta_0 + \beta_{SEc} SEc_i + \beta_{X1c} X1c_i + \beta_{X2c} X2c_i$

centroid of municipality i
$\begin{cases} X1c = \text{E-W, centered} \\ X2c = \text{N-S, centered} \end{cases}$

$+ \sum_{k=1}^{K} u_k \| x - K_k \|^2 \log \| x - K_k \|$

penalized spline part

Result (using SemiPar package)

	df	Knots
$f(X1c, X2c)$	2	48

i.e. the u_k are all smoothed to zero!

Here's the fit with 2 df in the spline, and one forced to have 26 df in the spline (of 50 max)

VAM T2add3 8/11/08

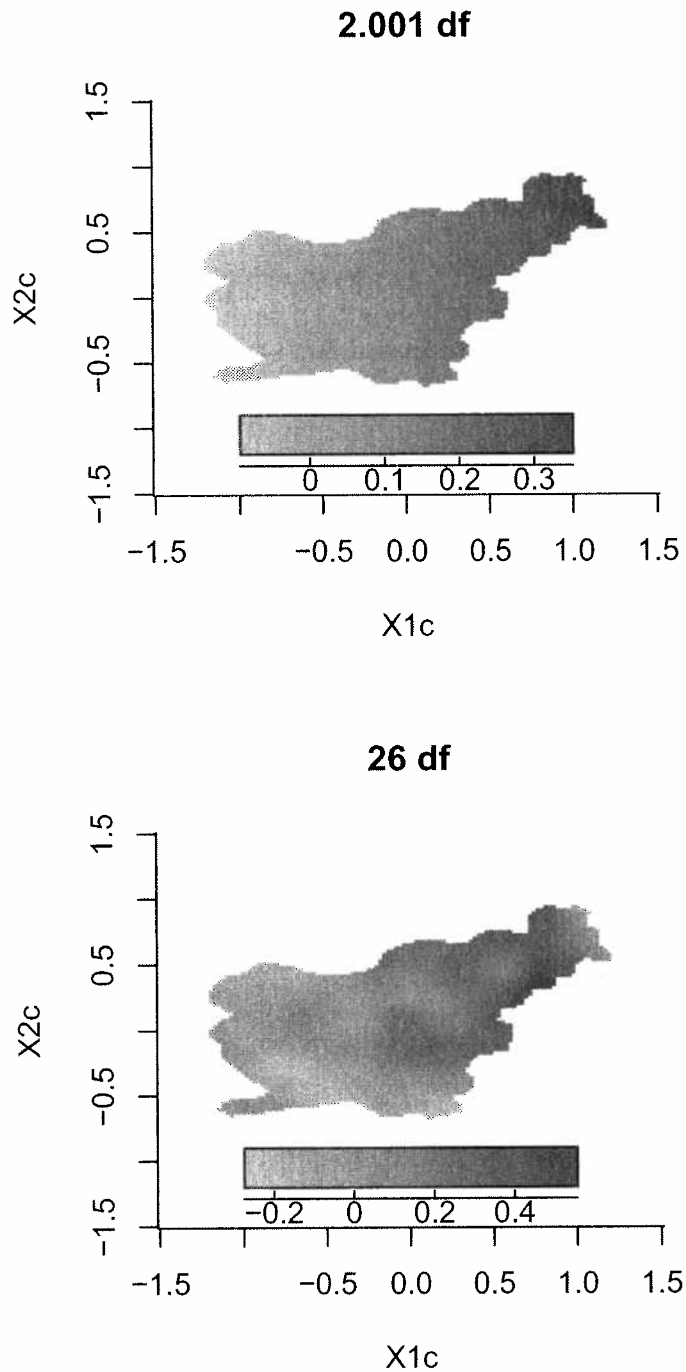
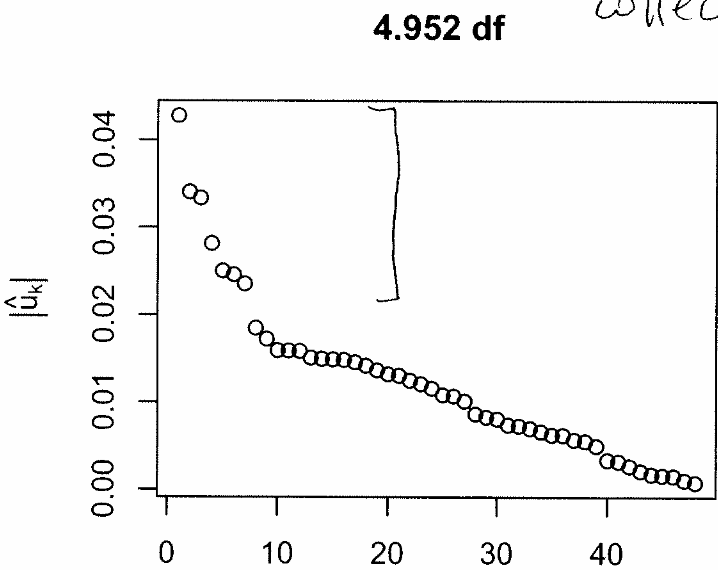


Figure 8: Plots of the bivariate smoothing function with 2.001 and 26 degrees of freedom.

VAM T2 add 4 8/11/08

If you force the u_k to have about 3df collectively about 7 of the 48 knots get "large" $|\hat{u}_k|$.



Here's where those knots are

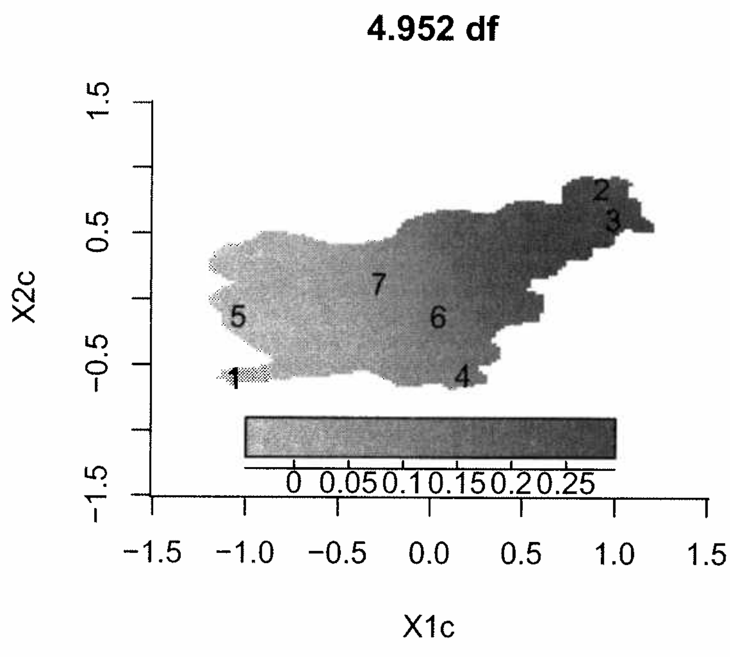


Figure 9: Ordered absolute estimated random effects with 4.952 degrees of freedom (top). Location of the knots associated with the 7 largest absolute estimated random effects.

VAM T2add 5 8/11/08