Spatio-temporal disease mapping using INLA

Birgit Schrödle, Leonhard Held

Biostatistics Unit
Institute of Social and Preventive Medicine
University of Zurich

Trondheim, May 14th, 2009

Overview

Data

- Data: Reported cases of 8 animal diseases
  - → Time: 1991 - 2008
  - → Space: 26 cantons/184 regions of Switzerland and the Principality of Liechtenstein
- Problem: Underreporting due to biased case reporting
- Aim of the analysis:
  - → Identify unusual spatial incidence and time trends
  - → Explore influence of cantons on case reporting
- Statistical methods: Fit spatio-temporal models using INLA

Spatial analysis: Standard BYM model

- Data Y: Number Y_i of reported cases in region i
- Cases are Poisson distributed given a region-specific rate λ_i:

$$Y_i|\lambda_i \sim \textit{Poisson}$$

- λ_i is modelled using a linear predictor η_i:

$$\eta_i = \log(\lambda_i) = \log(m_i) + \mu + \psi_i + \nu_i \quad i = 1, ..., 185$$

with

  - $\log(m_i)$: Offset, log. number of herds m_i in region i
  - μ: Intercept
  - ψ_i: Structured spatial effect in region i, modelled as a GMRF
  - ν_i: Unstructured spatial effect in region i, i.i.d. for each region

Extension of the standard BYM model

- Linear predictor of the standard BYM model:

$$\eta_i = \log(\lambda_i) = \log(m_i) + \mu + \psi_i + \nu_i \quad i = 1, ..., 185$$

- Aim of the analysis: Explore influence of cantons on case reporting
  $\rightarrow$ Additionally include an i.i.d. cantonal effect $\alpha_{j(i)}$ to account for cantonal heterogeneity
- Linear predictor of the extended BYM model:

$$\eta_i = \log(m_i) + \mu + \psi_i + \nu_i + \alpha_{j(i)} \quad j = 1, ..., 27$$

Spatio-temporal models

Aim: Develop a set of spatio-temporal models, which are able

- to detect unusual spatial incidence and time trends
- to explore influence of cantons on case reporting

Components of these models:

- Regional effects
- Cantonal effect
- Linear or nonparametric time trend
- Interaction term between time trend and space

Estimation of parameters: INLA (Rue *et al.*, 2009)

Model I: Linear time trend, inclusion of a cantonal effect

Assumptions:

- a linear time trend
- that cantonal heterogeneity is present
- that the increase in reported cases over time is differently steep for each canton

Linear predictor for region i at time t:

$$\eta_{it} = \mu + \psi_i + \nu_i + \alpha_{j(i)} + (\beta + \delta_{j(i)}) \cdot t$$

- Spatial effects: ψ_i, ν_i and $\alpha_{j(i)}$
- Linear time trend: $\beta \cdot t$
- Interaction term between time trend and canton: $\delta_{j(i)} \cdot t$
  $\longrightarrow$ modelled as a random slope (Bernardinelli *et al.*, 1995)
- Allow for correlation between $\alpha_{j(i)}$ and $\delta_{j(i)}$!

Specification of Model I using `inla`

- Allow for correlation between $\alpha_{j(i)}$ and $\delta_{j(i)}$:
  $\longrightarrow$ $\alpha_{j(i)}$ and $\delta_{j(i)}$ form a bivariate normal distribution with a Wishart prior specified for their precision components
- Using `inla`:

```
[canton]
type = ffield
model = 2diidwishartpart0
n = 27
covariates = canton1.txt
prior = wishart
parameters = 4   1   1   0
```

```
[interaction]
type=ffield
model = 2diidwishartpart1
n = 27
covariates = canton2.txt
weights = time.txt
```

Model II: Nonparametric time trend, inclusion of a cantonal effect

Assumptions:

- a nonparametric time trend
- that cantonal heterogeneity is present
- an interaction between nonparametric time trend and cantonal effect (Knorr-Held, 2000)

Linear predictor for region i at time t:

$$\eta_{it} = \mu + \psi_i + \nu_i + \alpha_{j(i)} + \beta_t + \delta_{j(i)t}$$

- Nonparametric time trend (RW1): β_t
- Interaction term between time trend and canton: $\delta_{j(i)t}$

$$P(\delta_{jt}|\lambda_{\delta_{jt}}) \propto \exp\left(-\frac{\lambda_{\delta_{jt}}}{2}\sum_{j=1}^{n_j}\sum_{t=2}^{T}(\delta_{jt} - \delta_{j,t-1})^2\right)$$

Specification of Model II using `inla` - Issue 1

Issue 1: Structure matrix of the interaction term

- Interaction term is a random effect indep. of the main effects
- Structure matrix is obtained by the Kronecker product of interacting effects
- Model II $\longrightarrow$ blockwise random walk (RW1) for each canton
- In `inla`: Define a file `Qmat.txt` containing the non-zero entries of the structure matrix

```
0   0    1
0   1   -1
1   0   -1
1   1    2
1   2   -1
⋮   ⋮    ⋮
```

and set options `model=generic`, `Qmatrix=Qmt.txt`

Specification of Model II using `inla` - Issue 2

Issue 2: Extra constraints on the interaction term

- To ensure identifiability, sum-to-zero constraints on subsets of the $\delta_{j(i)t}$'s have to be specified
- Model II $\rightarrow$ 27 constraints, one per canton
- Can be passed to `inla` in a file `extra.txt` using option `extraconstraint=extra.txt`
- Rank deficiency is computed automatically, but can be passed to `inla` using option `rankdef=27`
- Redundant constraints should be left out to speed up computations

Specification of Model II using `inla` - Final tag

Final tag for the interaction term:

```
[interaction]
type = ffield
model = generic
Qmatrix = Qmat.txt
extraconstraint = extra.txt
covariates = interaction.txt
parameters = 1 0.01
initial = 0.5
diagonal = 0.0001
quantiles = 0.025 0.975
```

Model III: Nonparametric time trend, no cantonal effect

Assumptions:

- a nonparametric time trend
- that cantonal heterogeneity is **not** present
- an interaction between nonparametric time trend and regions (Knorr-Held, 2000)

Linear predictor for region i at time t:

$$\eta_{it} = \mu + \psi_i + \nu_i + \beta_t + \delta_{it}$$

- Nonparametric time trend (RW1): β_t
- Interaction term between time trend and regions: δ_{it}

$$P(\delta_{it}|\lambda_{\delta_{it}}) \propto \exp\left(-\frac{\lambda_{\delta_{it}}}{2}\sum_{t=2}^{T}\sum_{i\sim l}(\delta_{it} - \delta_{lt} - \delta_{i,t-1} + \delta_{l,t-1})^2\right)$$

Model summary

Model I: $\eta_{it} = \mu + \psi_i + \nu_i + \alpha_{j(i)} + (\beta + \delta_{j(i)})\, t$

- ψ_i, ν_i: Regional, GMRF and i.i.d.
- $\alpha_{j(i)}$: Cantonal, i.i.d.
- $\beta \cdot t$: Temporal, linear
- $\delta_{j(i)}$: Interaction between time and canton, random slope

Model II: $\eta_{it} = \mu + \psi_i + \nu_i + \alpha_{j(i)} + \beta_t + \delta_{j(i)t}$

- β_t: Temporal, RW1
- $\delta_{j(i)t}$: Interaction between time and canton

Model III: $\eta_{it} = \mu + \psi_i + \nu_i + \beta_t + \delta_{it}$

- $\alpha_{j(i)}$: *Not* present
- δ_{it}: Interaction between time and region

Model choice criteria

1. Deviance information criterion (DIC):

$$\text{DIC} = \bar{D} + p_D$$

   Trade-off between model fit and model complexity

2. Histogram of cross-validated PIT values (Czado *et al.*, 2009):

$$P(Y_{it} \leq y_{it} | y_{-it})$$

   Assess predictive quality with respect to calibration

3. Logarithmic score:

$$LS = -\log P(Y_{it} = y_{it} | y_{-it})$$

   Compare predictive performance of various models

$\rightarrow$ All required quantities are given to the user by `inla` setting options `dic=1` and `cpo=1`

Example: Salmonellosis

- Data: All reported cases of Salmonellosis in cattle, 1991-2008

Year	91	92	93	94	95	96	97	98	99
n	69	73	30	48	61	83	99	64	79

Year	00	01	02	03	04	05	06	07	08
n	56	42	52	39	32	16	22	27	17

- 1. Model Choice
- 2. Interpretation of the results

DIC

	$\bar{D}$	p_D	DIC
Model I	3302	104	3406
Model II	3034	163	3207
Model III	2979	294	3273

Table: DIC of Models I, II and III

- Fit is best for Model III, but it also has the highest number of effective parameters
- Best trade-off between model complexity and fit for Model II $\longrightarrow$ Lowest DIC

PIT histogram

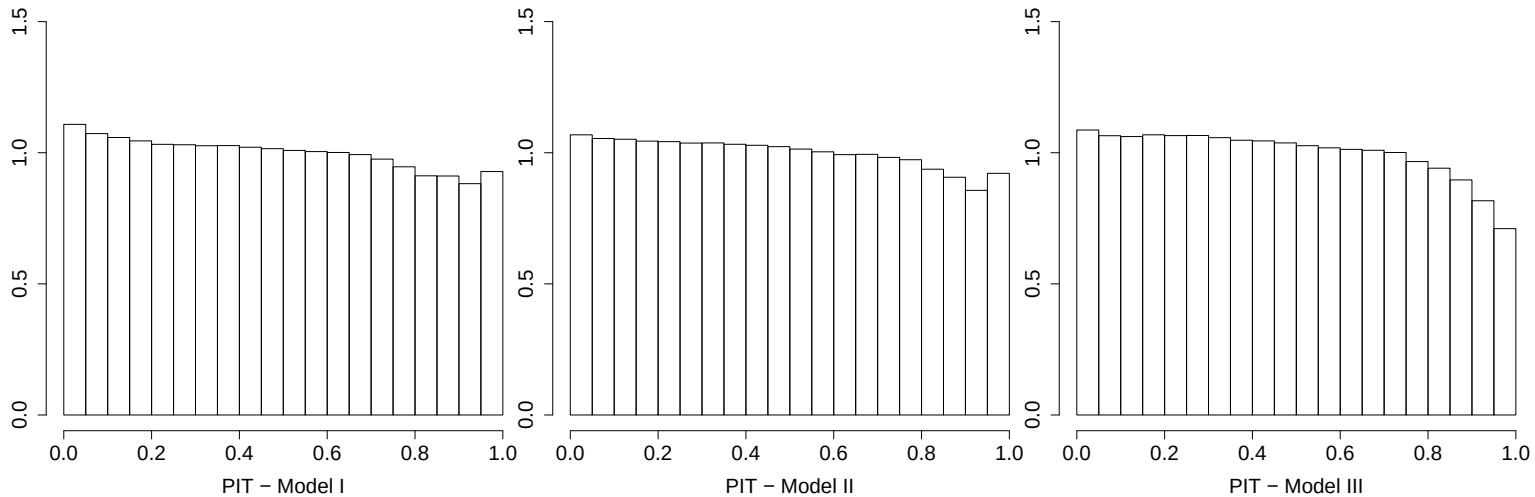


Figure: Cross-validated PIT histogram for Models I, II and III

- Calibration of all models is reasonable
- Histogram of Model II might be closest to uniform

Logarithmic score

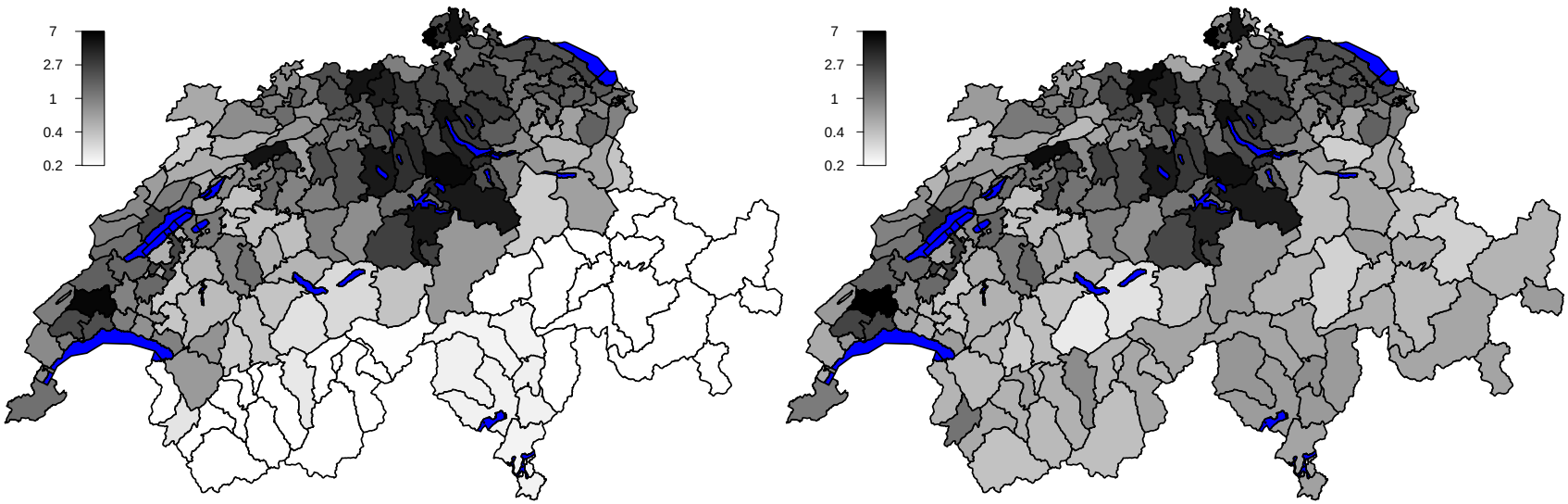
	Log-score
Model I	0.515
Model II	0.486
Model III	0.493

Table: Logarithmic score of Models I, II and III

Permutation test	p-value
M I vs. M II	< 0.001
M I vs. M III	< 0.001
M II vs. M III	0.022

Table: Results of permutation tests

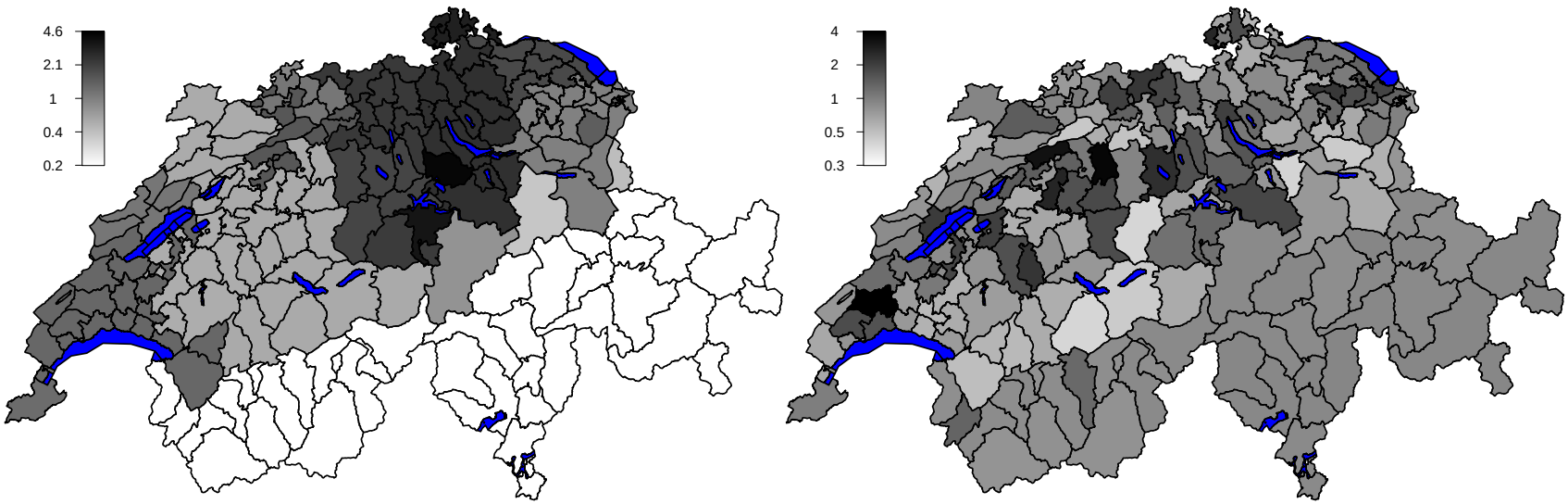
Model II and III - Fitted relative incidence



Left: Fitted relative incidence obtained by **Model II**

Right: Fitted relative incidence obtained by **Model III**

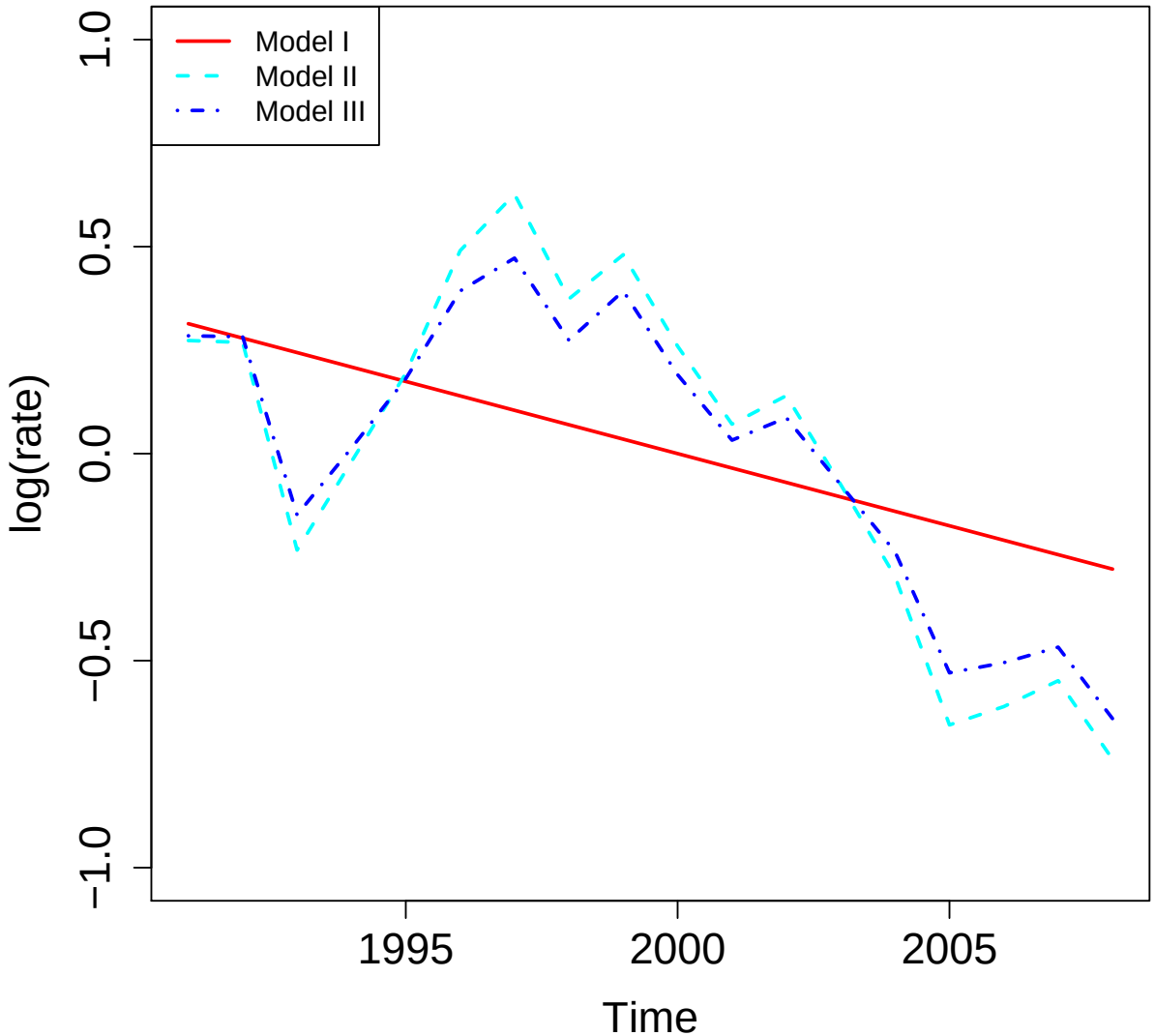
Model II - Cantonal and regional effect



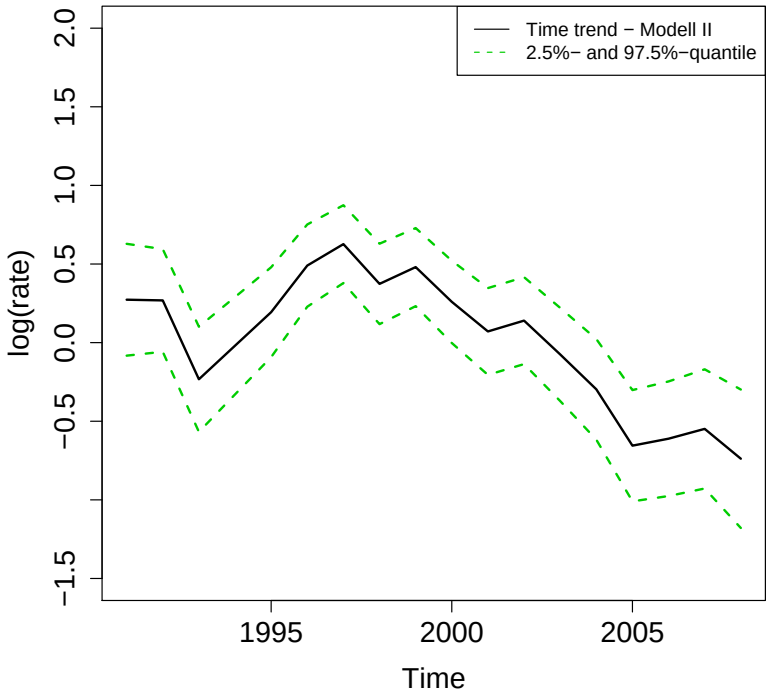
Left: Cantonal effect on exponential scale

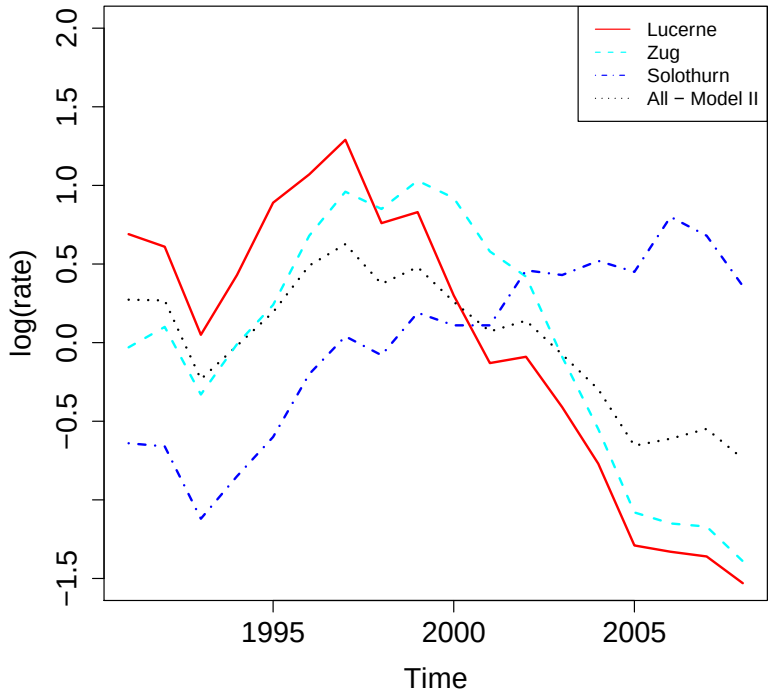
Right: Sum of regional effects on exponential scale

Model I, II and III - Estimated overall time trend



Model II - Overall time trend and differing cantons





Left: Estimated overall time trend and respective 2.5% - and 97.5% - quantile

Right: Cantons with significantly different time trend within some years

Constraints and Laplace approximation of $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$

- To ensure identifiability of interaction terms $\delta_{j(i)t}$ and δ_{it}:
  $\longrightarrow$ Constraints have to be formulated
  $\longrightarrow$ Specific subsets of effects must sum to zero
- Depending on the type of interaction (e.g. only temporally or spatially and temporally dependent) the number of constraints can become very high
- Example: Spatially and temporally dependent interactions δ_{it} between 185 regions and 18 years of reporting
  $\longrightarrow$ results in 202 constraints!

Constraints and Laplace approximation of $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$

- Problem:
  Simplified Laplace and full Laplace approximation for $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$ do not exactly correct for linear constraints on the x_i's

- Consequence:
  Parameters can be off the constraints

  $\longrightarrow$ Results for different models can hardly be compared
  $\longrightarrow$ If parameters are not centered around zero, no conclusions on significance can be drawn from quantiles

$\longrightarrow$ Solution
  - Use Gaussian approximation instead?

Failures in computation of cross-validated predictive measures by `inla`

- Option `cpo=1` returns cross-validated PIT and CPO values
  $\longrightarrow$ Derive PIT histogram and logarithmic score
- For 'hard' observations: Computation of cross-validated measures can face numerical problems
  $\longrightarrow$ File `failure.dat` reports a numerical failure

$\longrightarrow$ Output of `inla` cannot be used for assessment of cross-validated predictive quality of a model!

$\longrightarrow$ Solution

- Dawid-Sebastiani Score (Czado *et al.*, 2009)?

Discussion

+ INLA has proofed to be a useful and flexible tool for fitting spatio-temporal models with even a complex dependence structure
+ Running time is very fast and the results can be interpreted easily

– User needs some experience in choosing the most appropriate approximation technique and appropriate settings for the approximation routine
– Computation of cross-validated predictive measures by `inla` can fail and output cannot be used

Literature

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C. and Ghislandi, M. (1995). Bayesian analysis of space-time variation in disease risk, *Statistics in Medicine* **14**: 2433–2443.

Czado, C., Gneiting, T. and Held, L. (2009). Predictive model assessment for count data, *Biometrics*.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk, *Statistics in Medicine* **19**: 2555–2567.

Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion), *Journal of the Royal Statistical Society: Series B* **71**: 319–392.