

# A general mixed model approach for spatio-temporal regression data

Thomas Kneib, Ludwig Fahrmeir & Stefan Lang  
Department of Statistics, Ludwig-Maximilians-University Munich

1. Spatio-temporal regression data
2. Examples I: Categorical forest health data
3. Structured additive regression models
4. Mixed model based inference
5. Software
6. Examples II: Leukemia survival times



16.12.2004



## Spatio-temporal regression data

- Regression in a general sense: Generalized linear models, mult categorical models, survival models.
- Usual regression data: Some response with categorical and continuous covariates.
- Spatio-temporal regression data: in addition **spatial** and **temporal information**.
- Special issues:
  - **Spatial and spatio-temporal correlations**,
  - **Time- or space-varying effects**,
  - Non-linear effects,
  - Complex interactions,
  - Unobserved heterogeneity.

## Example I: Categorical forest health data

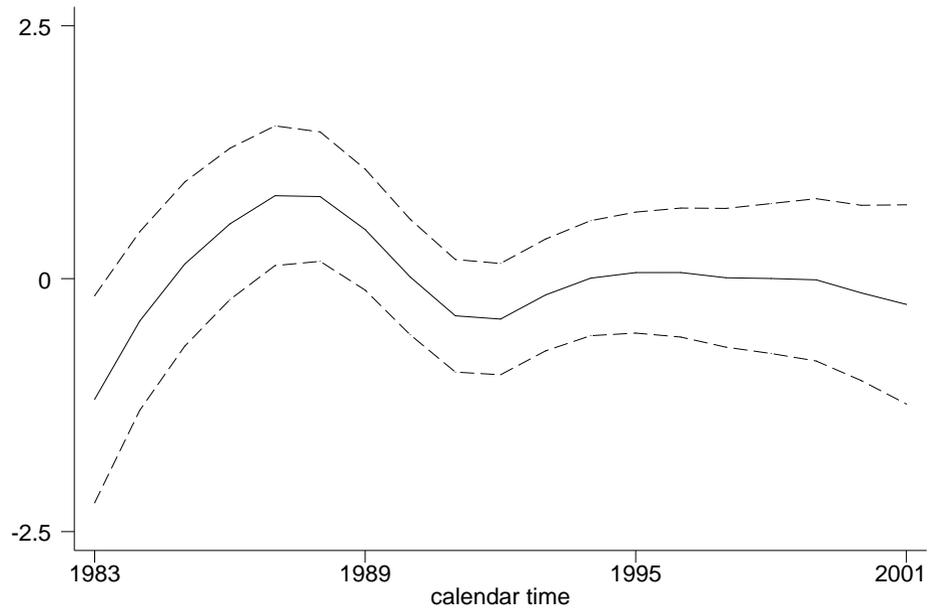
- Yearly forest health inventories carried out from 1983 to 2001.
- 83 beeches within a 15 km times 10 km area.
- Response: defoliation degree of beech  $i$  in year  $t$ , measured in three ordered categories:

$$\begin{aligned}
 y_{it} = 1 & \quad \text{no defoliation,} \\
 y_{it} = 2 & \quad \text{defoliation 25\% or less,} \\
 y_{it} = 3 & \quad \text{defoliation above 25\%.}
 \end{aligned}$$

- Cumulative probit model:

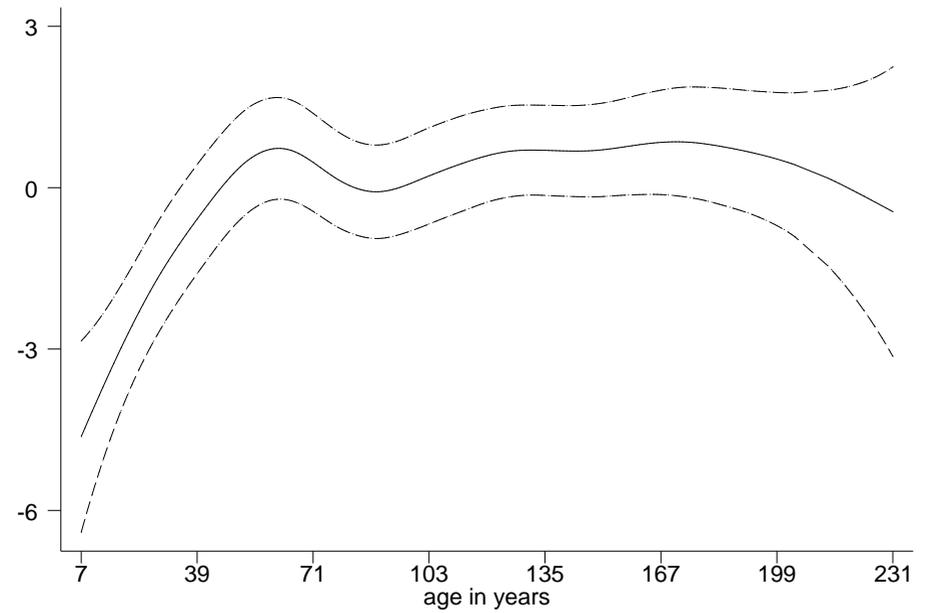
$$P(y_{it} \leq r) = \Phi [\theta_r - f_1(t) - f_2(\text{age}_{it}) - f_3(t, \text{age}_{it}) - f_{\text{spat}}(s_i) - u'_{it}\gamma]$$

with standard normal cdf  $\Phi$  and thresholds  $-\infty = \theta_0 < \theta_1 < \theta_2 < \theta_3 = \infty$

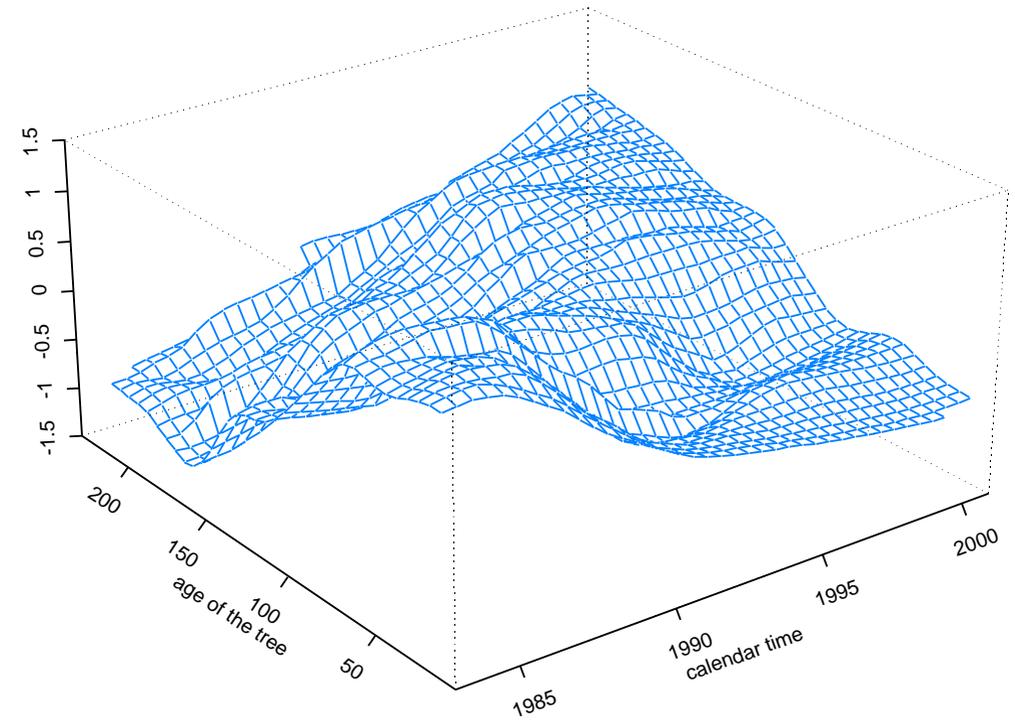
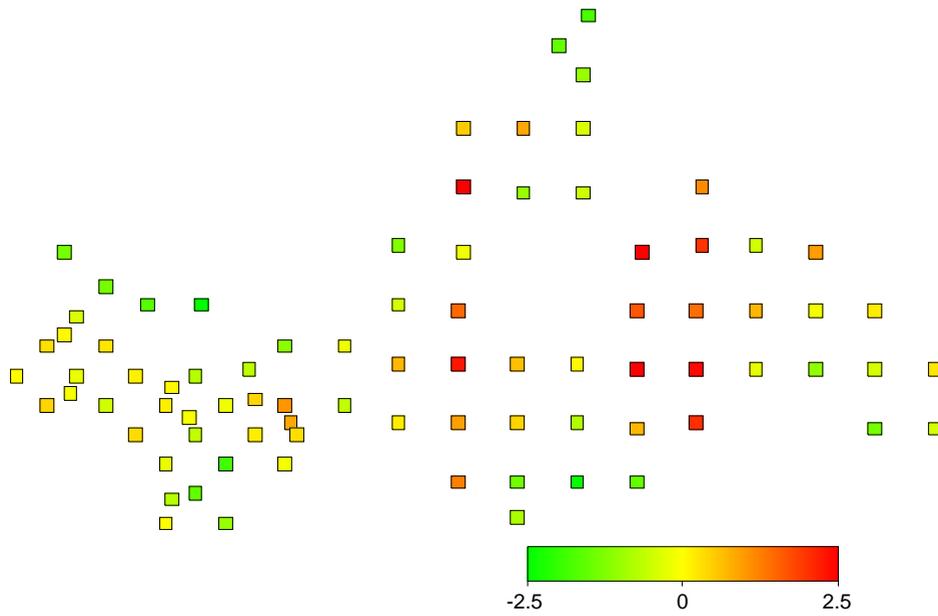


Time trend.

Age effect



Spatial effect.



Interaction effect.

## Structured additive regression

- General Idea: Replace usual parametric predictor with a **flexible semiparametric** predictor containing
  - Nonparametric effects of **time scales** and continuous covariates (P-splines, random walks),
  - **Spatial effects** (Markov random fields, stationary Gaussian random fields),
  - Interaction surfaces (2d P-splines),
  - Varying coefficient terms (continuous and **spatial effect modifiers**),
  - Random intercepts and random slopes.
- In a Bayesian context, all effects can be cast into **one general framework**.

## Mixed model based inference

- Each term in the predictor is associated with a vector of regression coefficients with **improper multivariate Gaussian prior**:

$$p(\beta_j | \tau_j^2) \propto \exp \left( -\frac{1}{2\tau_j^2} \beta_j' K_j \beta_j \right)$$

⇒ Reparametrize the model to a **proper mixed model**.

- Obtain empirical Bayes estimates via iterating
  - Penalized maximum likelihood for regression coefficients.
  - Restricted Maximum / Marginal likelihood for variance parameters.

## Software

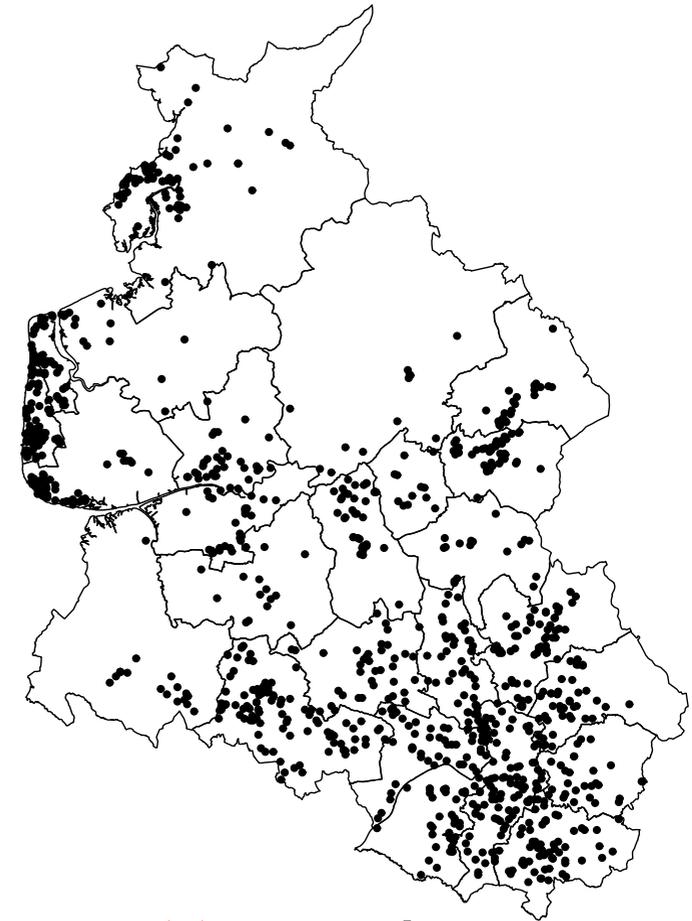
- Implemented in the public domain software package BayesX.
- Allows for spatio-temporal regression in the context of
  - Univariate responses from exponential families,
  - **Multicategorical responses** with ordered and unordered categories,
  - **Continuous time survival analysis.**
- Available from



<http://www.stat.uni-muenchen.de/~lang/bayesx>

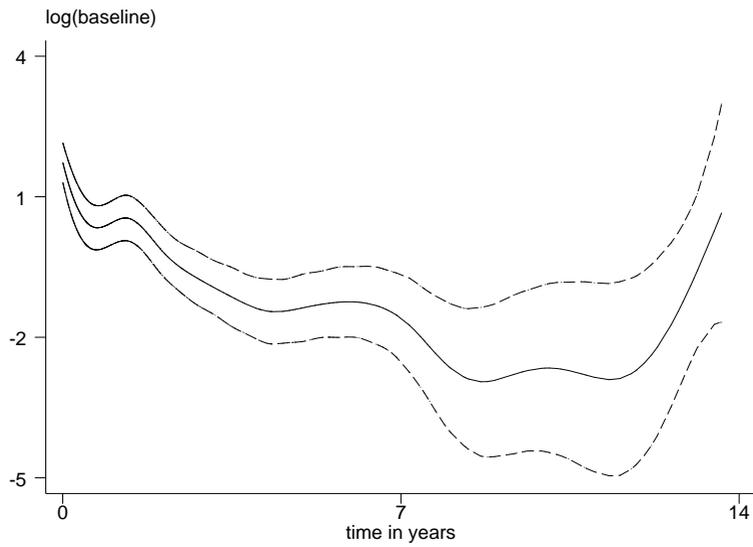
## Examples II: Leukemia survival times

- Survival time of adults after diagnosis of acute myeloid leukemia.
- 1,043 cases diagnosed between 1982 and 1998 in Northwest England.
- **Spatial information in different resolutions.**
- Model:

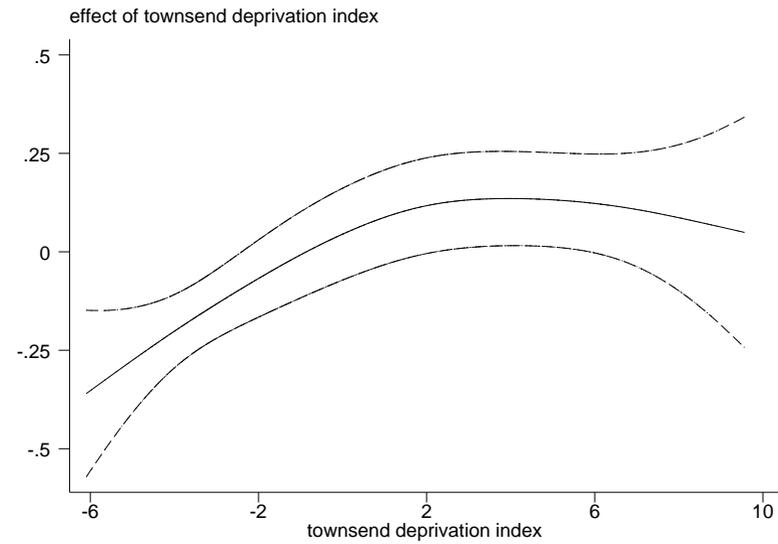


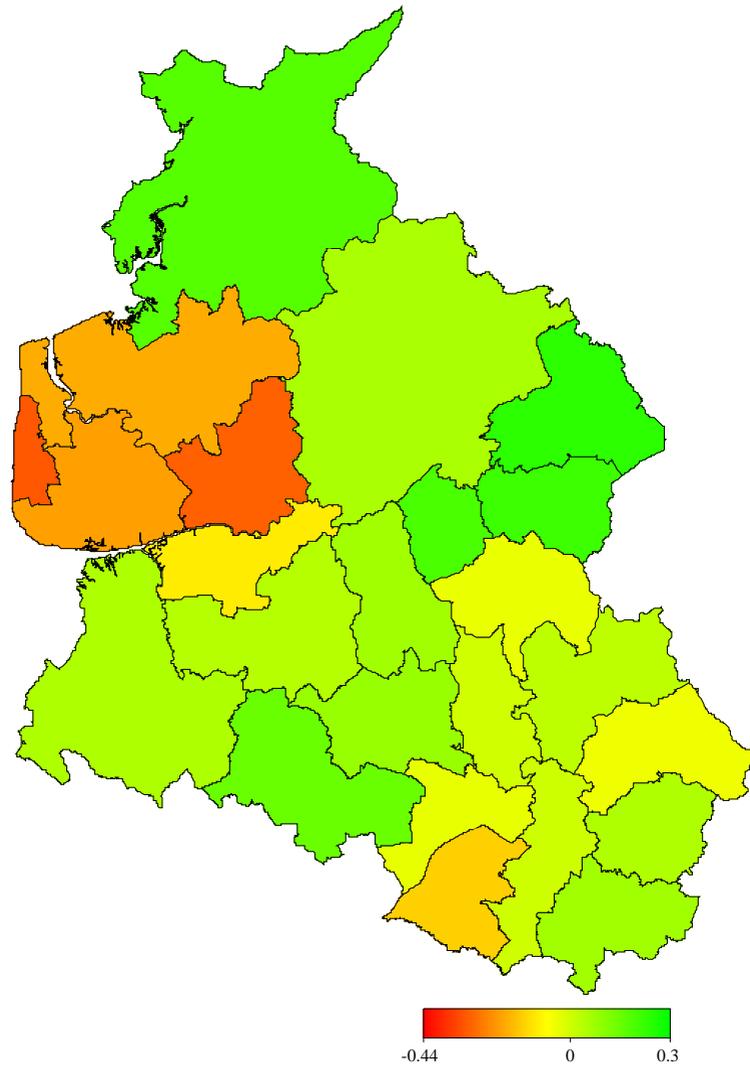
$$\lambda(t; \cdot) = \lambda_0(t) \exp[f_1(\text{age}) + f_2(\text{wbc}) + f_3(\text{tpi}) + f_{\text{spat}}(s) + \gamma \text{sex}]$$

### Log-baseline hazard.

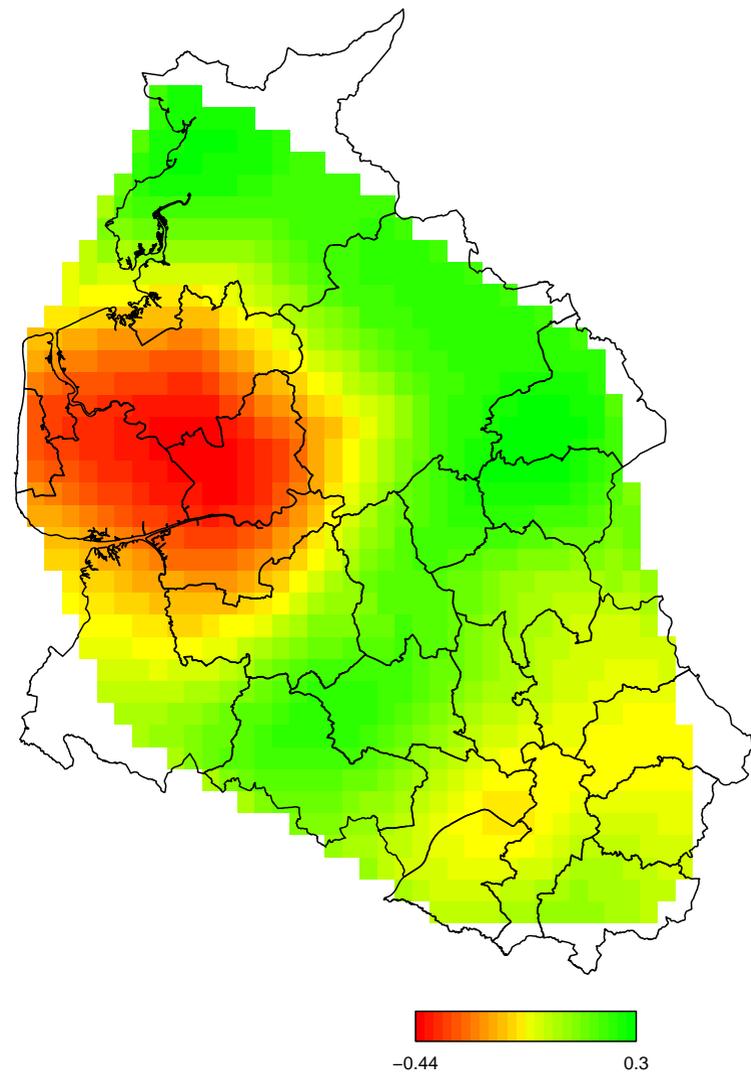


### Effect of deprivation.





District-level analysis



Individual-level analysis

## Discussion

- Bayesian treatment of complex regression models for spatio-temporal data **without relying on MCMC simulation techniques**.
- Closely related to penalized likelihood in a frequentist setting.
- **Future work:**
  - Extended modelling for categorical responses, e.g. with correlated latent utilities.
  - More general censoring schemes for survival times, e.g. interval censoring.
  - Anisotropic spatial effects.
  - 3d extensions of P-splines.

## References

- Kneib, T. and Fahrmeir, L. (2004): A mixed model approach for structured hazard regression. SFB 386 Discussion Paper 400, University of Munich.
- Kneib, T. and Fahrmeir, L. (2004): Structured additive regression for categorical space-time data: A mixed model approach. Under revision for *Biometrics*.
- Fahrmeir, L., Kneib, T. and Lang, S. (2004): Penalized structured additive regression for space-time data: A Bayesian perspective. *Statistica Sinica*, 14, 715-745.
- Available from

<http://www.stat.uni-muenchen.de/~kneib>