BayesX - Software for Bayesian Inference in Structured Additive Regression

Thomas Kneib

Faculty of Mathematics and Economics, University of Ulm
Department of Statistics, Ludwig-Maximilians-University Munich

joint work with
Stefan Lang (University of Innsbruck) & Andreas Brezger (University of Munich)

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

2.7.2007

Structured Additive Regression

- Regression in a general sense:
  - Generalised linear models,
  - Multivariate (categorical) generalised linear models,
  - Regression models for duration times (Cox-type models, multi-state models).

- Common structure: Model a quantity of interest in terms of categorical and continuous covariates, e.g.

$$\mathbb{E}(y|u) = h(u'\gamma) \qquad \text{(GLM)}$$

or

$$\lambda(t|u) = \lambda_0(t)\exp(u'\gamma) \qquad \text{(Cox model)}$$

- General idea of structured additive regression: Replace usual parametric predictor with a flexible semiparametric predictor containing
  - Nonparametric effects of time scales and continuous covariates,
  - Spatial effects,
  - Interaction surfaces,
  - Varying coefficient terms (continuous and spatial effect modifiers),
  - Random intercepts and random slopes.

- Example: Car insurance data from two insurance companies in Belgium.
- Sample of approximately 160.000 policyholders.
- Aims: Separate risk analyses for claim size and claim frequency to predict risk premium from covariates.
- Variables of primary interest: Claim size y_i or claim frequency h_i of policyholders.
- Covariates:

$vage$	vehicles age
$page$	policyholders age
hp	vehicles horsepower
bm	bonus-malus score
s	district in Belgium
v	Vector of categorical covariates

- Geoadditive models:
  - Gaussian model for log-costs $\log(y)$:

$$\log(y) \sim N(\eta, \sigma^2)$$

with

$$\eta = f_1(vage) + f_2(page) + f_3(bm) + f_4(hp) + f_{spat}(s) + v'\zeta.$$

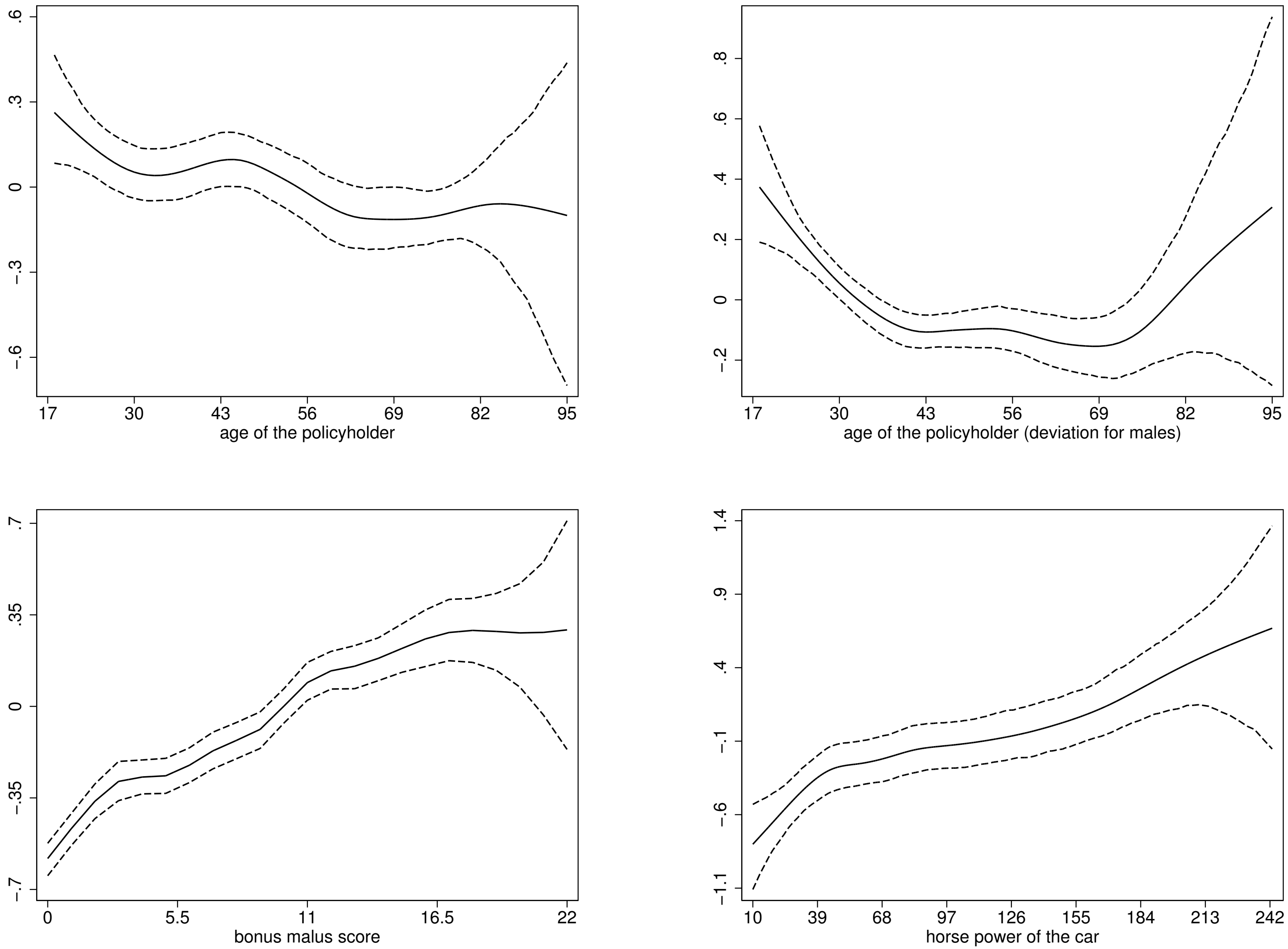
  - Poisson model for frequencies h_i:

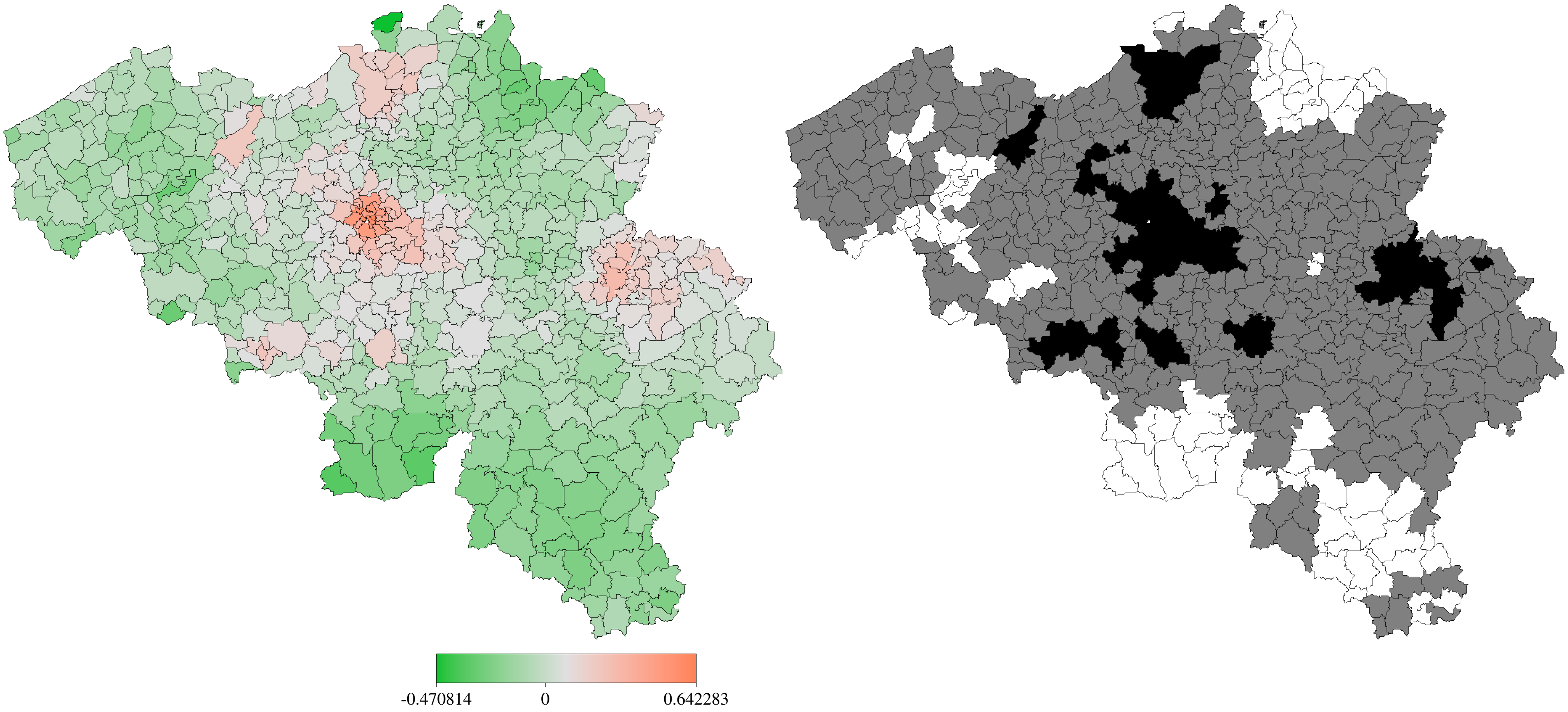
$$h \sim Po(\exp(\eta))$$

with

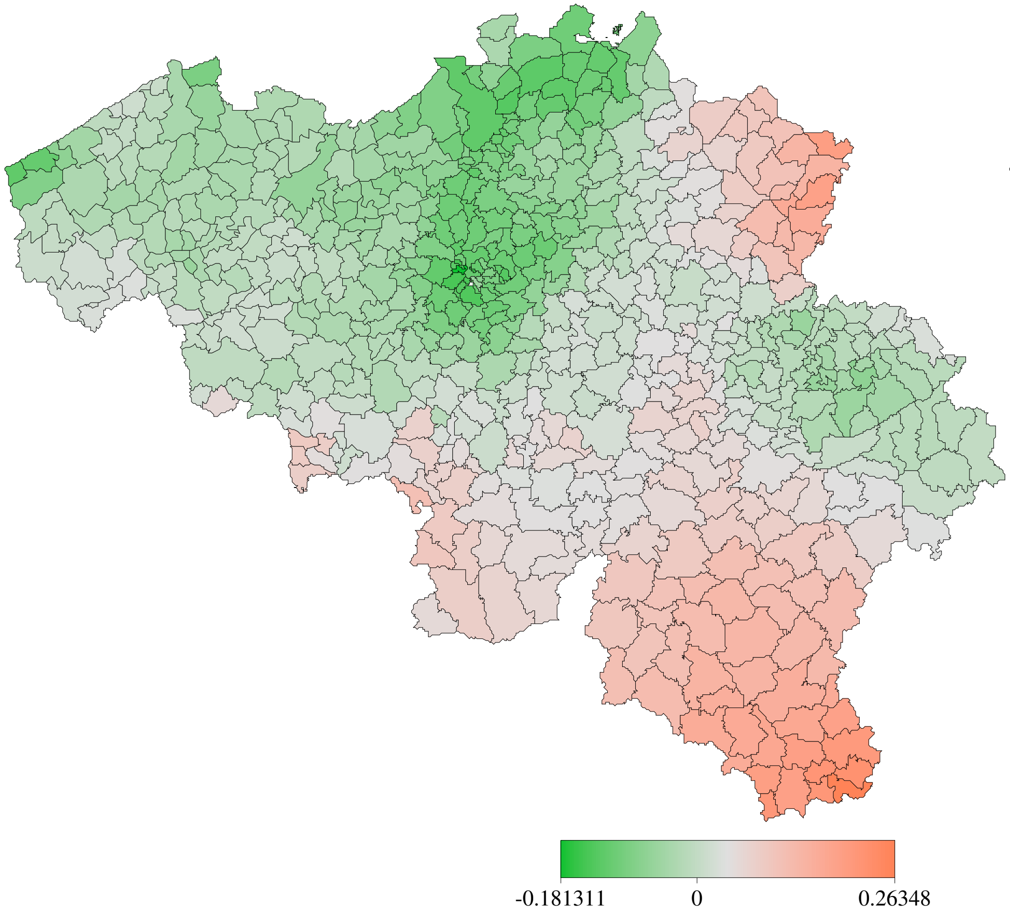
$$\eta = f_1(vage) + f_2(page) + f_3(page)sex + f_3(bm) + f_4(hp) + f_{spat}(s) + v'\zeta.$$

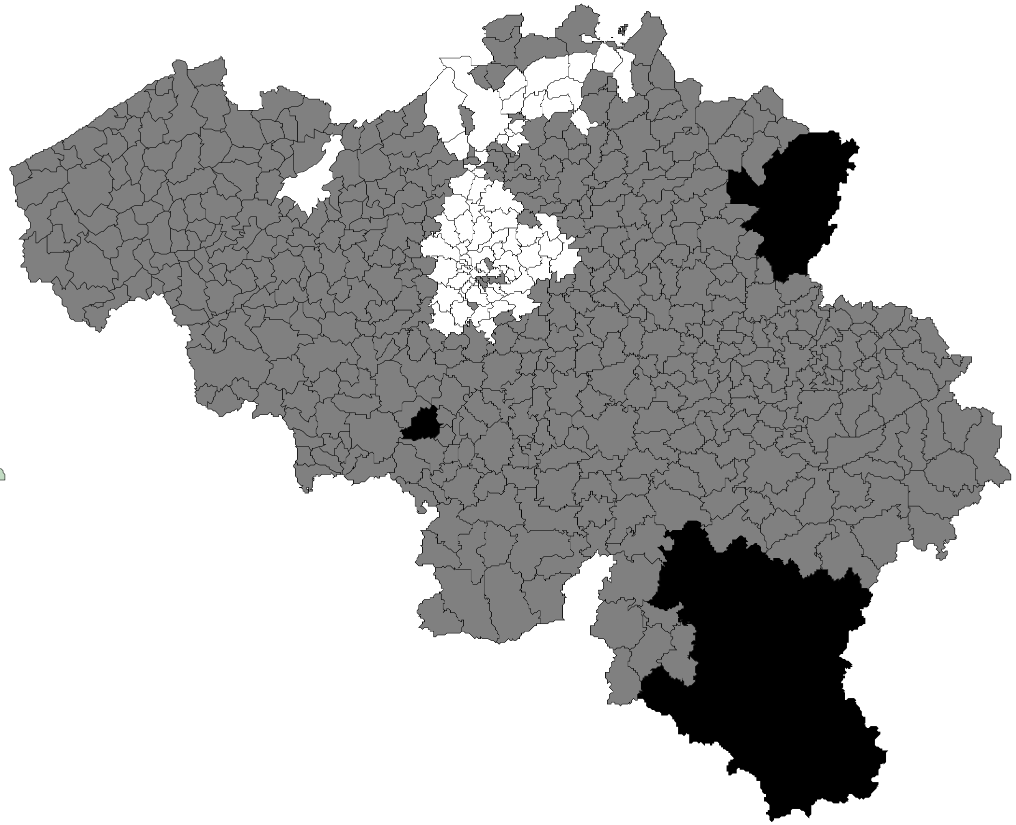
- Results for claim frequency:




-0.470814
0
0.642283

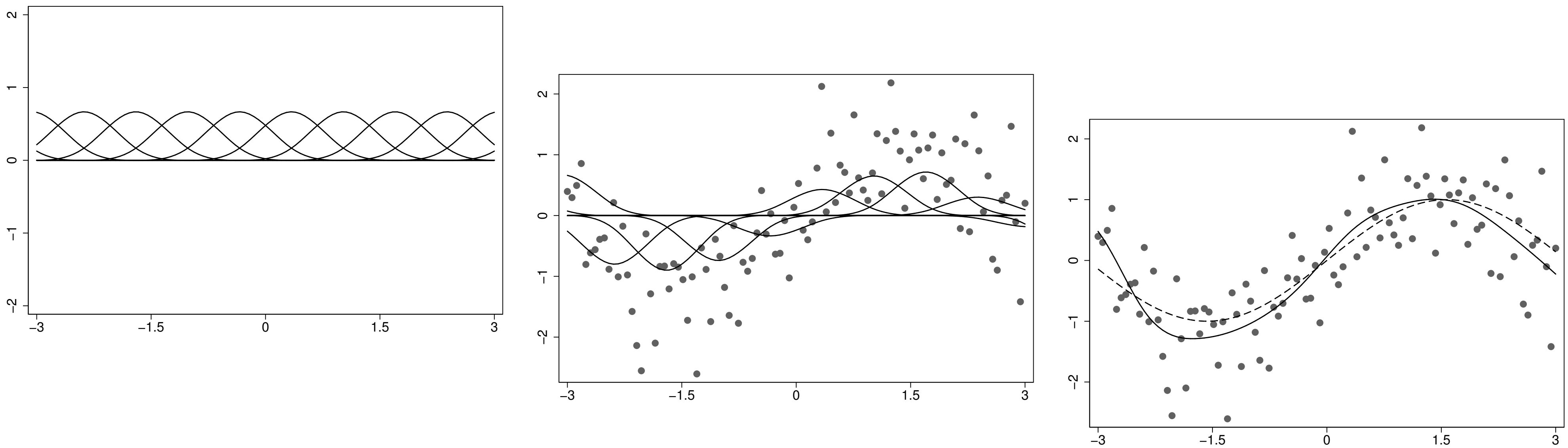
- Spatial effect for claim size:



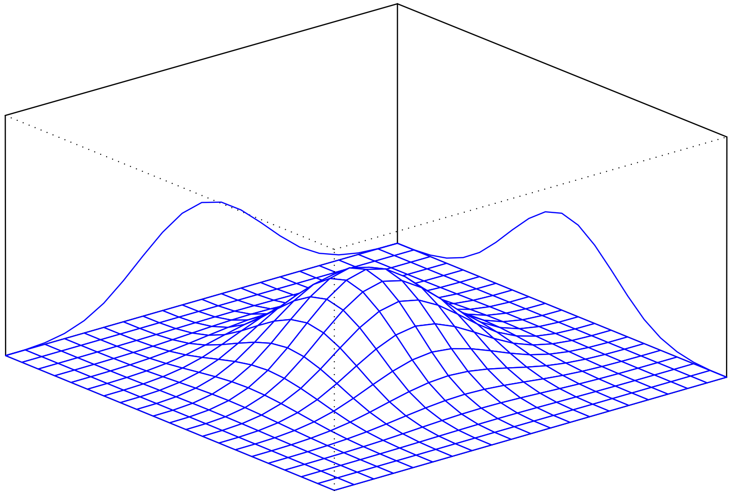


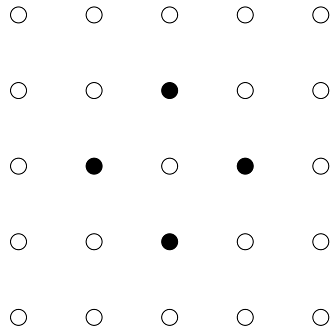
Model Components and Priors

- Penalised splines.
  - Approximate $f(x) = \sum \xi_j B_j(x)$ by a weighted sum of B-spline basis functions.
  - Employ a large number of basis functions to enable flexibility.
  - Penalise differences between parameters of adjacent basis functions to ensure smoothness.



- Bivariate penalised splines.



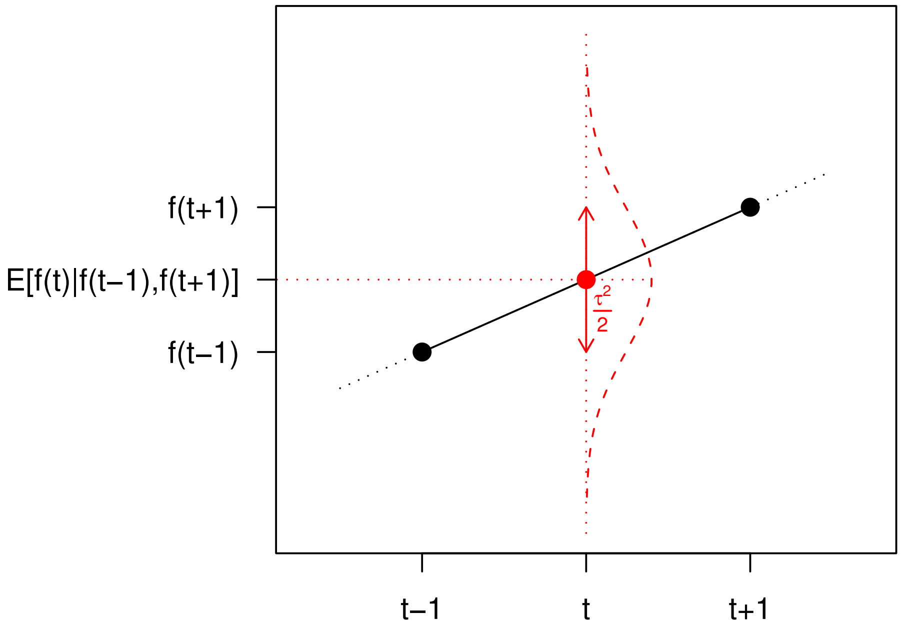


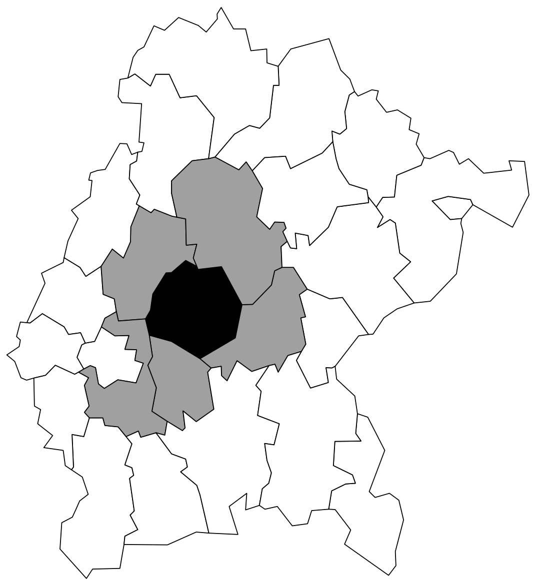
- Varying coefficient models.
  - Effect of covariate x varies smoothly over the domain of a second covariate z:

$$f(x,z) = x \cdot g(z)$$

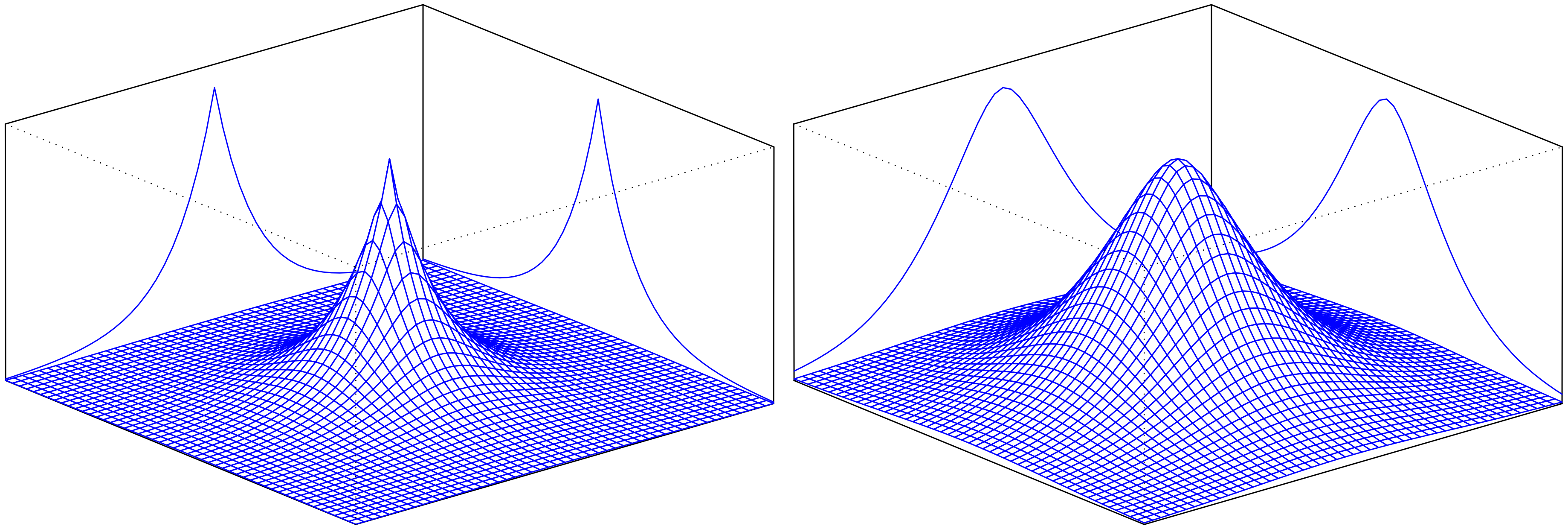
  - Spatial effect modifier $\Rightarrow$ Geographically weighted regression.

- Spatial effect for regional data: Markov random fields.
  - Bivariate extension of a first order random walk on the real line.
  - Define appropriate neighbourhoods for the regions.
  - Assume that the expected value of $f_{spat}(s)$ is the average of the function evaluations of adjacent sites.





- Spatial effect for point-referenced data: Stationary Gaussian random fields.
  - Well-known as Kriging in the geostatistics literature.
  - Spatial effect follows a zero mean stationary Gaussian stochastic process.
  - Correlation of two arbitrary sites is defined by an intrinsic correlation function.
  - Can be interpreted as a basis function approach with radial basis functions.



- All effects can be cast into one general framework.
- All vectors of function evaluations f_j can be expressed as

$$f_j = Z_j \xi_j$$

  with design matrix Z_j and regression coefficients ξ_j.

- Generic form of the prior for ξ_j:

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

- $K_j \geq 0$ acts as a penalty matrix, $\mathrm{rank}(K_j) = k_j \leq d_j = \dim(\xi_j)$.
- $\tau_j^2 \geq 0$ can be interpreted as a variance or (inverse) smoothness parameter.

Bayesian Inference

- Fully Bayesian inference:
  - All parameters (including the variance parameters τ^2) are assigned suitable prior distributions.
  - Typically, estimation is based on MCMC simulation techniques.
  - Usual estimates: Posterior expectation, posterior median (easily obtained from the samples).

- Empirical Bayes inference:
  - Differentiate between parameters of primary interest (regression coefficients) and hyperparameters (variances).
  - Assign priors only to the former.
  - Estimate the hyperparameters by maximising their marginal posterior.
  - Plugging these estimates into the joint posterior and maximising with respect to the parameters of primary interest yields posterior mode estimates.

- MCMC-based inference:
  - Assign inverse gamma prior to τ_j^2:

$$p(\tau_j^2) \propto \frac{1}{(\tau_j^2)^{a_j+1}} \exp\left(-\frac{b_j}{\tau_j^2}\right).$$

Proper for	$a_j > 0,\ b_j > 0$	Common choice: $a_j = b_j = \varepsilon$ small.
Improper for	$b_j = 0,\ a_j = -1$	Flat prior for variance τ_j^2,
	$b_j = 0,\ a_j = -\frac{1}{2}$	Flat prior for standard deviation τ_j.

  - Conditions for proper posteriors in structured additive regression are available.
  - Gibbs sampler for $\tau_j^2 | \cdot$:

    Sample from an inverse Gamma distribution with parameters

$$a_j' = a_j + \frac{1}{2}\text{rank}(K_j) \qquad \text{and} \qquad b_j' = b_j + \frac{1}{2}\xi_j' K_j \xi_j.$$

– Metropolis-Hastings update for $\xi_j|\cdot$:

Propose new state from a multivariate Gaussian distribution with precision matrix and mean

$$P_j = Z_j'WZ_j + \frac{1}{\tau_j^2}K_j \qquad \text{and} \qquad m_j = P_j^{-1}Z_j'W(\tilde{y} - \eta_{-j}).$$

IWLS-Proposal with appropriately defined working weights W and working observations $\tilde{y}$.

- Efficient algorithms make use of the sparse matrix structure of P_j and K_j.

- For binary or categorical regression models: Efficient implementation based on latent variable representation.

- Empirical Bayes inference.
  - Consider the variances τ_j^2 as unknown constants to be estimated from their marginal posterior.
  - Consider the regression coefficients ξ_j as correlated random effects with multivariate Gaussian distribution

    $\Rightarrow$ Use mixed model methodology for estimation.

- Problem: In most cases partially improper random effects distribution.

- Mixed model representation: Decompose

$$\xi_j = X_j\beta_j + V_j b_j,$$

where

$$p(\beta_j) \propto const \qquad \text{and} \qquad b_j \sim N(0, \tau_j^2 I_{k_j}).$$

$\Rightarrow$ β_j is a fixed effect and b_j is an i.i.d. random effect.

- This yields a variance components model with pedictor

$$\eta = X\beta + Vb$$

where in turn

$$p(\beta) \propto const \qquad \text{and} \qquad b \sim N(0, Q).$$

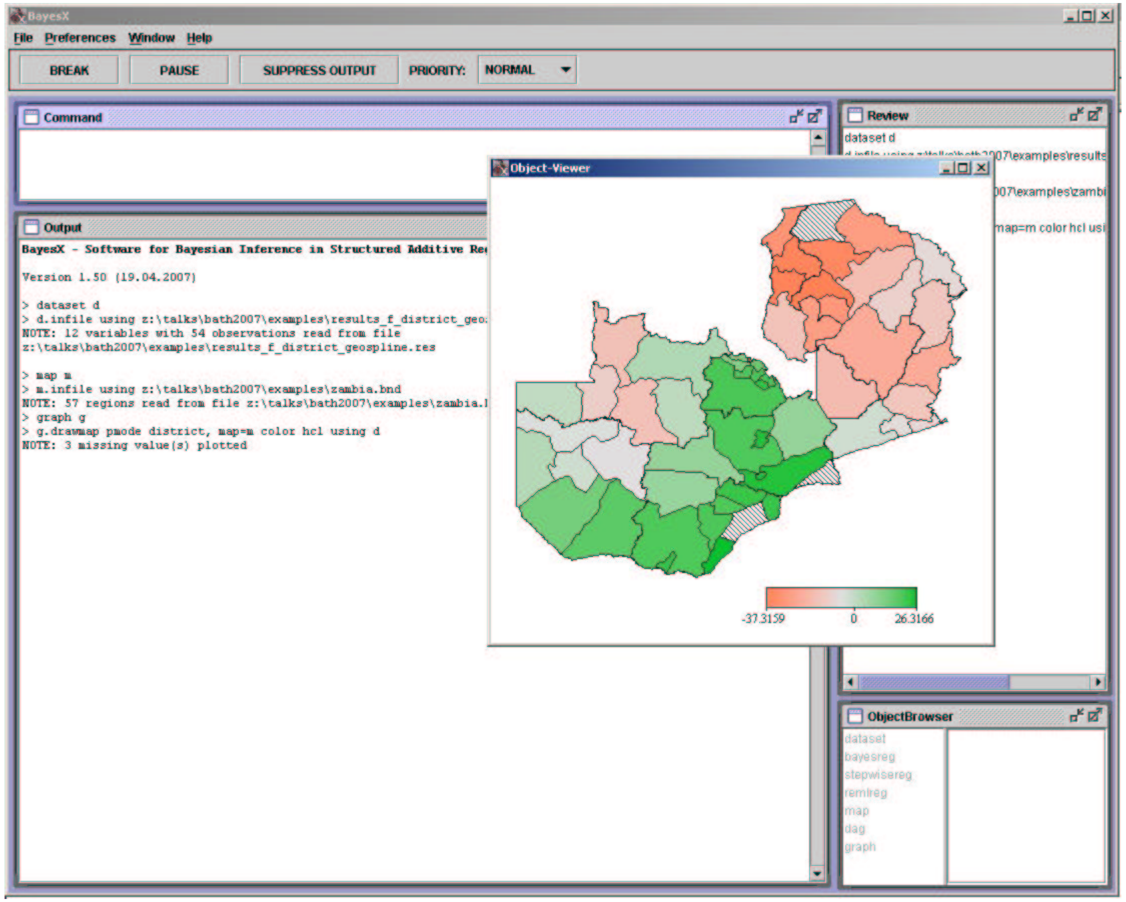
- Obtain empirical Bayes estimates / penalized likelihood estimates via iterating
  - Penalized maximum likelihood for the regression coefficients β and b.
  - Restricted Maximum / Marginal likelihood for the variance parameters in Q:

$$L(Q) = \int L(\beta, b, Q) p(b) d\beta db \to \max_Q .$$

- Involves a Laplace approximation to the marginal likelihood (corresponding to REML estimation of variances in Gaussian mixed models).

BayesX

- BayesX is a software tool for estimating structured additive regression models.



- Stand-alone software with Stata-like syntax.
- Developed by Andreas Brezger, Thomas Kneib and Stefan Lang with contributions of seven colleagues.
- Computationally demanding parts are implemented in C++.
- Graphical user interface and visualisation tools are implemented in Java.
- Currently, BayesX only runs under Windows, a Linux version as well as a connection to R are work in progress.

- Resources:
  - `http://www.stat.uni-muenchen.de/~bayesx`.
  - Reference Manual, Methodology Manual, Tutorial Manual.
  - Some publications:

BREZGER, KNEIB & LANG (2005). BayesX: Analyzing Bayesian structured additive regression models. *Journal of Statistical Software*, **14** (11).

BREZGER, A. & LANG, S. (2006). Generalized additive regression based on Bayesian P-splines. *Computational Statistics and Data Analysis* **50**, 967–991.

FAHRMEIR, L., KNEIB, T. & LANG, S. (2004). Penalized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica* **14**, 731–761.

- R provides functionality for generalised additive models and a number of extensions such as varying coefficient models, interaction surfaces, spatio-temporal effects, etc. What is the gain in using BayesX?

- BayesX provides functionality for more general response types:
  - Ordered and unordered categorical responses,
  - Continuous survival time models extending the Cox model,
  - Multi-state models.

- Inference can be based on mixed model methodology or MCMC.

- Not all possibilities and combinations are supported by both inferential concepts.

- Categorical responses with unordered categories:
  - Multinomial logit and multinomial probit models,
  - Category-specific and globally-defined covariates,
  - Non-availability indicators can be defined to account for varying choice sets.
- Response families in BayesX: `multinomial`, `multinomialprobit`.
- Categorical responses with ordered categories:
  - Ordinal as well as sequential models,
  - Logit and probit models,
  - Effects can be category-specific or constant over the categories.
- Response families in BayesX: `cumlogit`, `cumprobit`, `seqlogit`, `seqprobit`.

- Continuous survival times:
  - Cox-type hazard regression models,
  - Joint estimation of the baseline hazard rate and the covariate effects,
  - Time-varying effects and time-varying covariates,
  - Arbitrary combinations of right, left and interval censoring as well as left truncation.
- Response family: `cox`.
- Multi-state models:
  - Describe the evolution of discrete phenomena in continuous time,
  - Model in terms of transition intensities, similar as in the Cox model.
- Response family: `multistate`.

Summary & Future Plans

- Take home message:

  BayesX is a user-friendly software that allows for the routine estimation of a broad class of semiparametric regression models even in non-standard situations.

- Future plans:
  - Linux version and connection to R.
  - Interval censoring for multi-state models.
  - Bayesian regularisation-priors and model choice (with LASSO-type priors).
  - Bayesian treatment of measurement error in structured additive regressions models.