Penalized structured additive regression for multicategorical space-time data

Thomas Kneib
University of Munich
joint work with Ludwig Fahrmeir and Stefan Lang

1. Forest health data
2. Multicategorical models for ordinal response
3. Penalized structured additive regression
4. Mixed model representation
5. Results
6. Software

16.03.2004

Forest health data

- Data collected in yearly forest health inventories carried out in a forest in northern Bavaria from 1983 to 2001.

- 83 observation points with beeches in an area extending 15 km from east to west and 10 km from north to south.

- y_{it}, the defoliation degree of beech i in year t, is measured in three ordered categories (multicategorical response):

  $y_{it} = 1$ no defoliation,
  $y_{it} = 2$ defoliation 25% or less,
  $y_{it} = 3$ defoliation above 25%.

- Covariates:

  t calendar time,
  s_i site of the beech,
  a_{it} age of the tree in years,
  u_{it} further (mostly categorical) covariates.

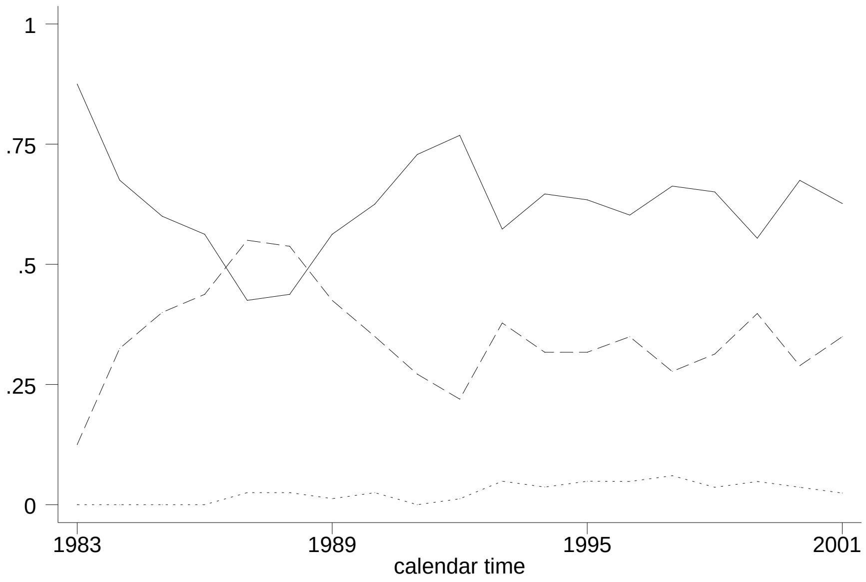

Figure 1: Temporal development of the frequency of the damage states:

— no damage,
- - - medium damage,
· · · severe damage.

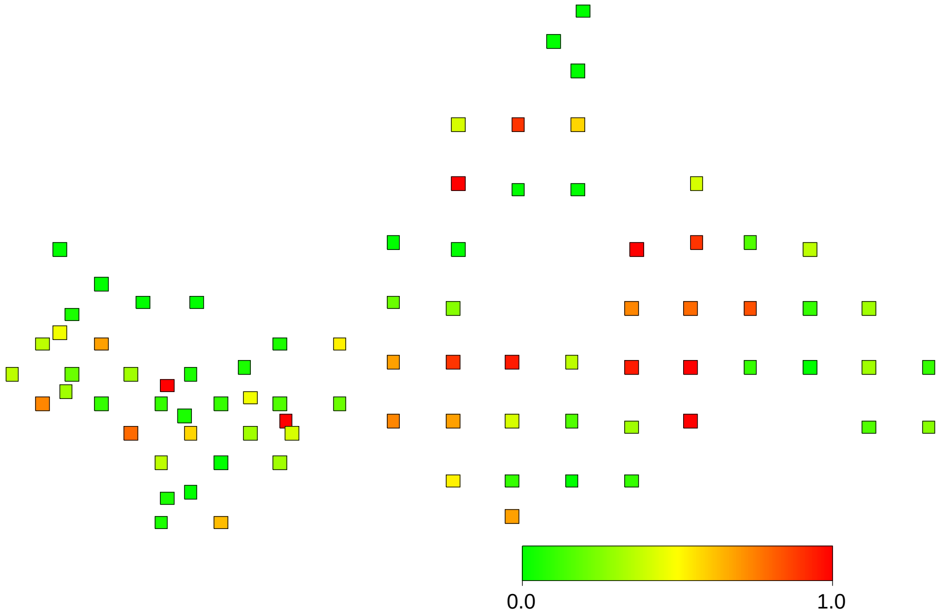

Figure 2: Spatial distribution of the beeches and percentage of time points for which a beech was classified to be damaged (damage state 2 or 3).

Multicategorical models for ordinal response

- Response y_{it} follows multinomial distribution with three ordered categories $r = 1, 2, 3$.

- Model the cumulative probabilities

$$P(y_{it} \leq r) = F(\theta_r - \eta_{it})$$

with thresholds $-\infty = \theta_0 < \theta_1 < \theta_2 < \theta_3 = \infty$ and linear predictor η_{it}.

- $F(\cdot)$ can be any cumulative distribution function:

standard normal	$\Longrightarrow$	cumulative probit model,
logistic	$\Longrightarrow$	cumulative logit model.

- Consider a random variable with density $f = F'$ and expectation η_{it}.

  $\Longrightarrow$ Linear predictor determines shift on latent scale.

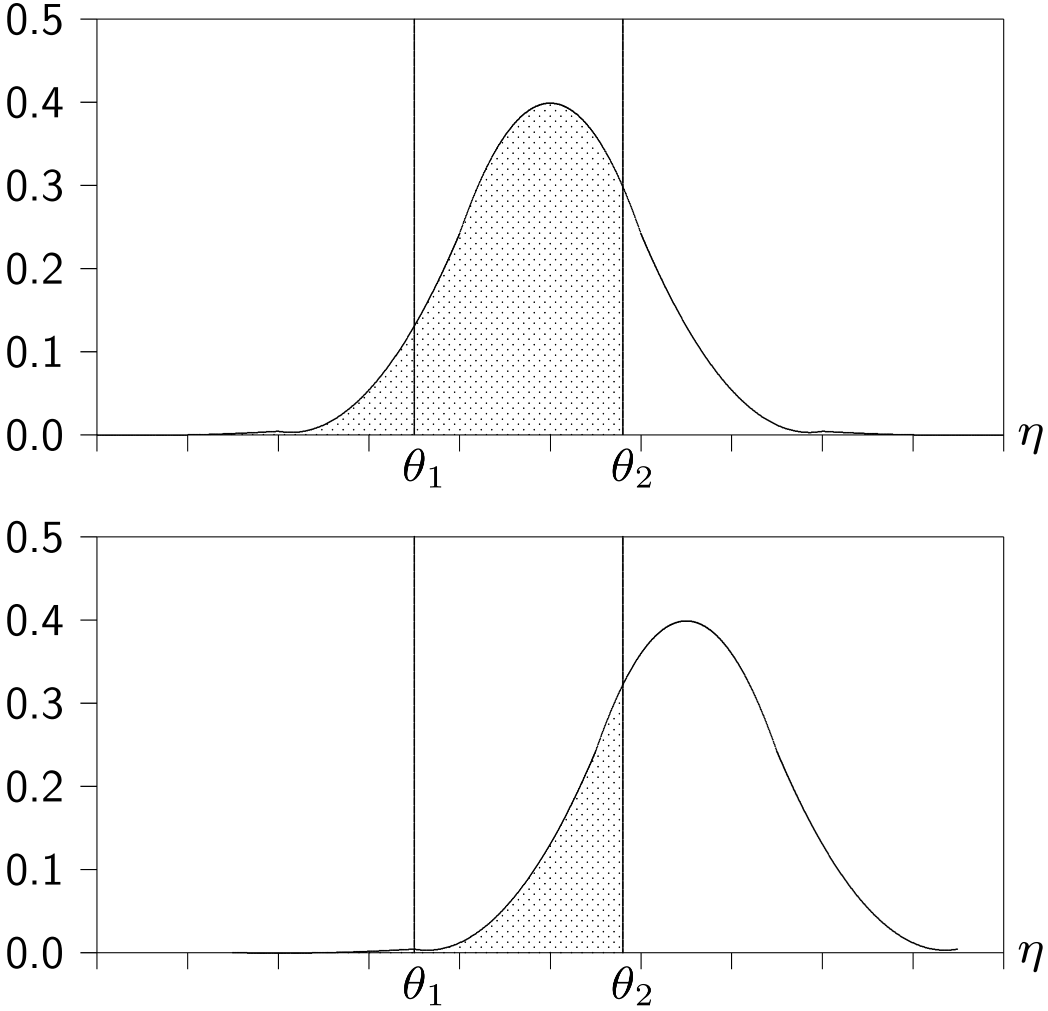

Figure 3: The shaded areas represent $P(y_{it} \leq 2)$ for different values of η_{it}.

- Limitations of a purely parametric approach:
  - Spatio-temporal structure of the data implies spatial and temporal correlations.
  - Nonlinear effects of continuous covariates?
  - Complex interactions between covariates?

$\Longrightarrow$ Structured additive regression models.

Penalized structured additive regression

- Replace usual parametric predictor with a flexible semiparametric predictor

$$\eta_{it} = f_1(t) + f_2(a_{it}) + f_3(t, a_{it}) + f_{spat}(s_i) + u_{it}'\gamma,$$

  where

  – f_1 and f_2 are nonparametric functions of calendar time and age,

  – f_3 is an interaction surface between calendar time and age,

  – f_{spat} is a spatial function, and

  – u is a vector of further covariates with parametric effects.

- Structured additive regression extends (and combines) generalized additive mixed models, geoadditive models and varying coefficient models.

- Allows unified treatment of all effects within a Bayesian framework.

- $f_1(t)$, $f_2(a_{it})$: P-splines
  - Approximate f_j by a B-spline of a certain degree (basis function approach).
  - Penalize differences between adjacent parameters of basis functions to ensure smoothness.
  - Alternatives: Random walks, autoregressive priors.
- $f_3(t, a_{it})$: Two-dimensional extensions of P-splines
  - Define two-dimensional basis functions based on tensor products of one-dimensional B-splines.
  - Use priors from spatial statistics for penalization.
  - Alternative: Varying coefficient models, if one of the interacting variables is categorical.

- $f_{spat}(s_i)$: Markov random fields
  - Consider two trees as neighbors if their distance is less than (e.g.) 1.2 km.
  - Assume that the expected value of $f_{spat}(s)$ is the average of the function evaluations of adjacent sites.
- $f_{spat}(s_i)$: Stationary Gaussian random fields (kriging)
  - Spatial effect follows a zero mean stationary Gaussian stochastic process.
  - Correlation of two arbitrary sites is defined by an intrinsic correlation function.
- Split up spatial effect into structured and unstructured part. The unstructured effect can be modelled by i.i.d. random effects, the structured effect by a MRF or a GRF.

Mixed model representation

- All nonparametric effects can be expressed as the product of a design matrix Z and a vector of regression coefficients β.

- Rewrite the structured additive predictor in matrix notation as

$$\eta = Z_1\beta_1 + Z_2\beta_2 + Z_3\beta_3 + Z_{spat}\beta_{spat} + U\gamma.$$

- Bayesian approach: Assign an appropriate prior to β_j.

- All priors can be cast into the general form

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

  where K_j is a penalty matrix and τ_j^2 is a smoothing parameter.

- Type of the covariate and prior beliefs about the smoothness of f_j determine special Z_j and K_j.

- Accuracy of the estimation relies heavily on the choice of the smoothing parameters.
- Idea: Reexpress the structured additive regression model as a multicategorical mixed model and use mixed model methodology.
- Each parameter vector β_j can be partitioned into an unpenalized part (with flat prior) and a penalized part (with i.i.d. Gaussian prior) yielding a variance components model

$$\eta = X^{unp}\beta^{unp} + X^{pen}\beta^{pen}$$

with

$$p(\beta^{unp}) \propto const \qquad \beta^{pen} \sim N(0, \Lambda)$$

and

$$\Lambda = \text{blockdiag}(\tau_1^2 I, \dots, \tau_4^2 I).$$

- Regression coefficients are estimated via modified Fisher scoring.
- The mixed model representation allows for restricted maximum likelihood / marginal likelihood estimation of the variance components:

$$L(\Lambda) = \int L(\beta^{unp}, \beta^{pen}, \Lambda) p(\beta^{pen}) d\beta^{pen} d\beta^{unp} \rightarrow \max_{\Lambda}.$$

- From a Bayesian perspective, we get empirical Bayes / posterior mode estimates.
- Closely related to penalized likelihood.
- Fahrmeir, Kneib and Lang (2004) derive numerically efficient formulae that allow the computation even for fairly large data sets.

Results

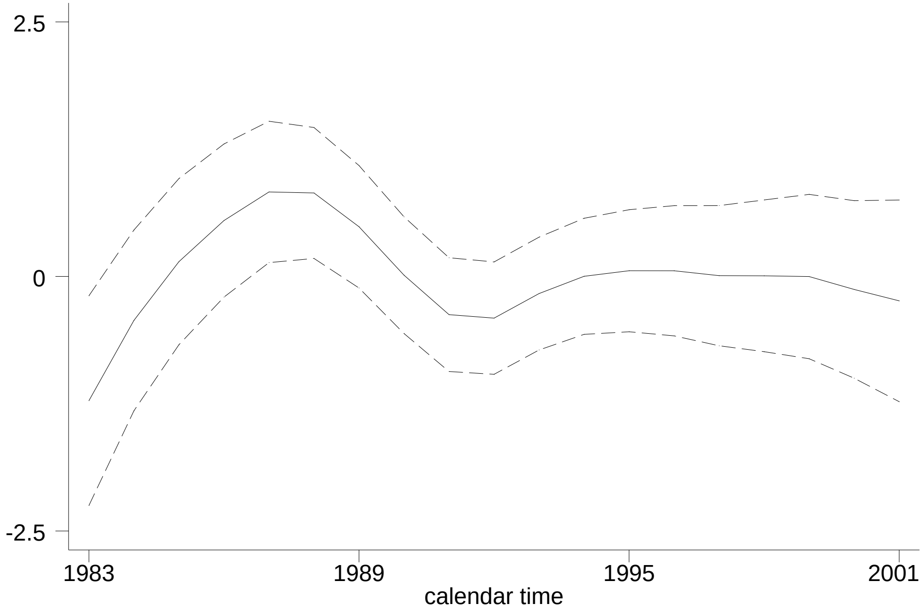

Figure 4: Time trend.

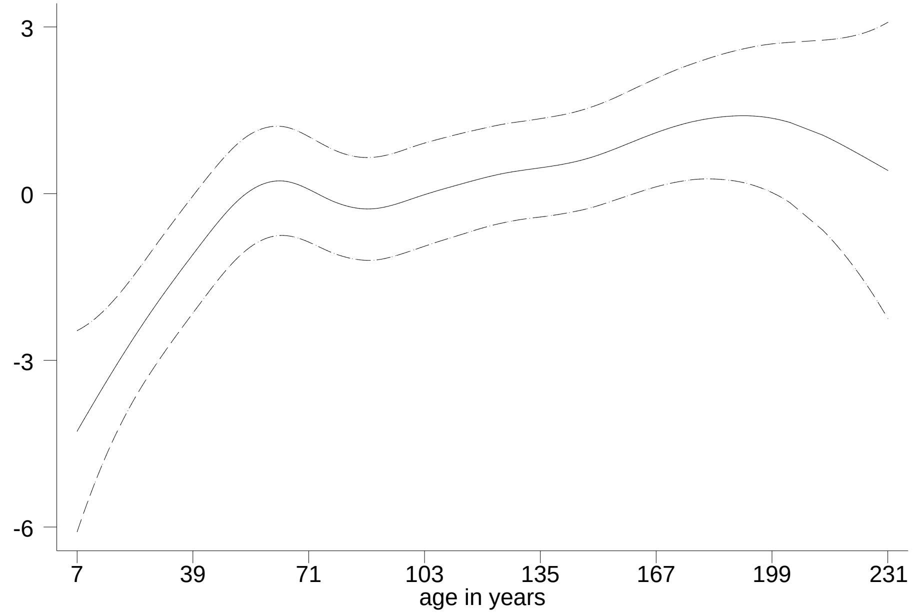

Figure 5: Age effect.

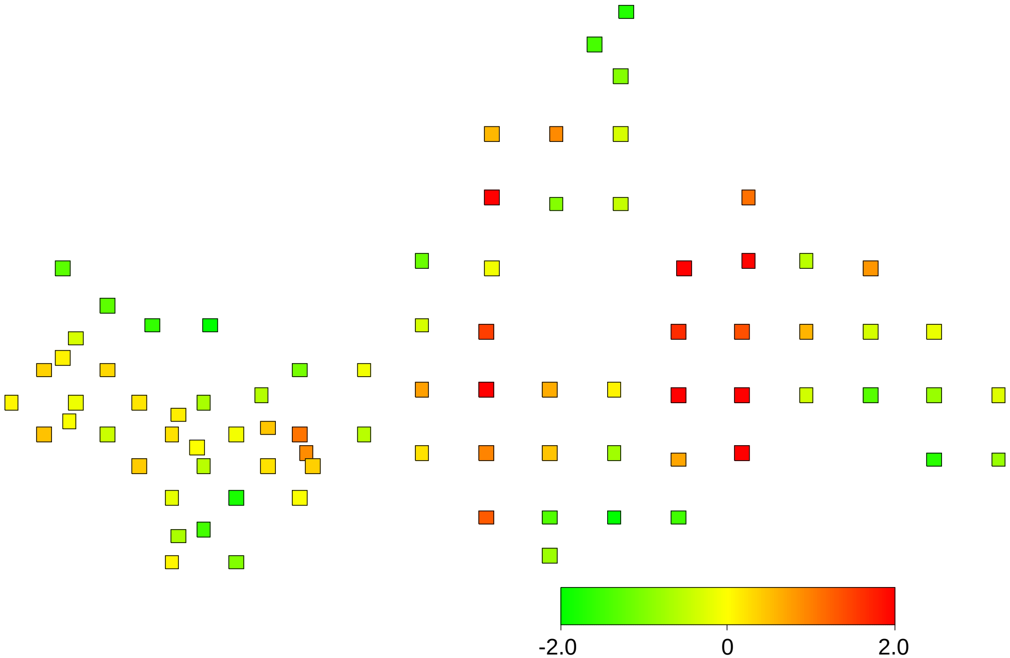

Figure 6: Structured spatial effect.

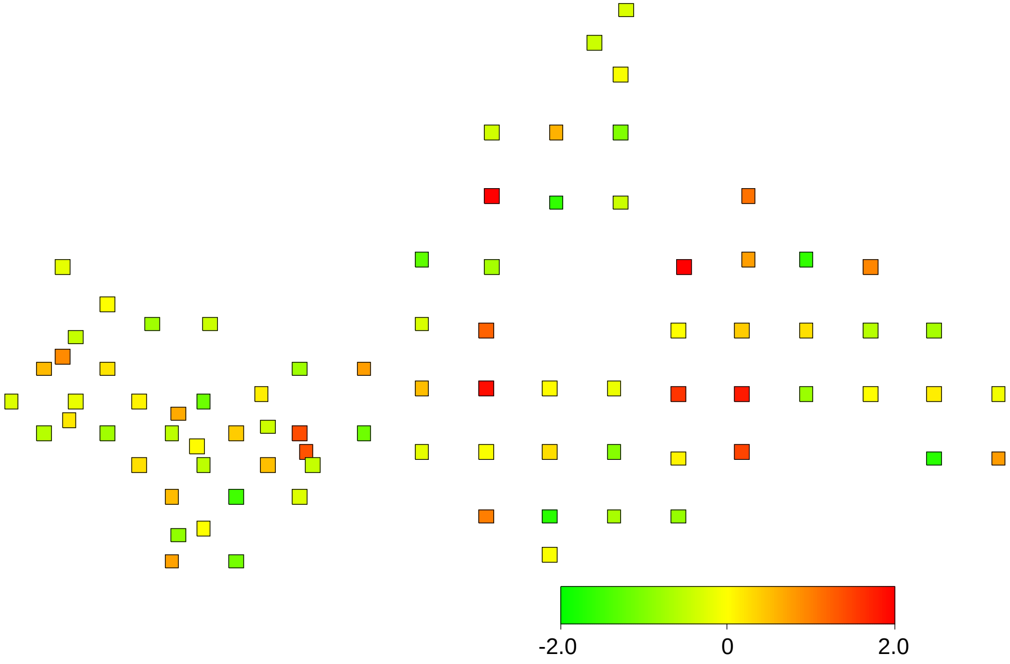

Figure 7: Unstructured spatial effect.

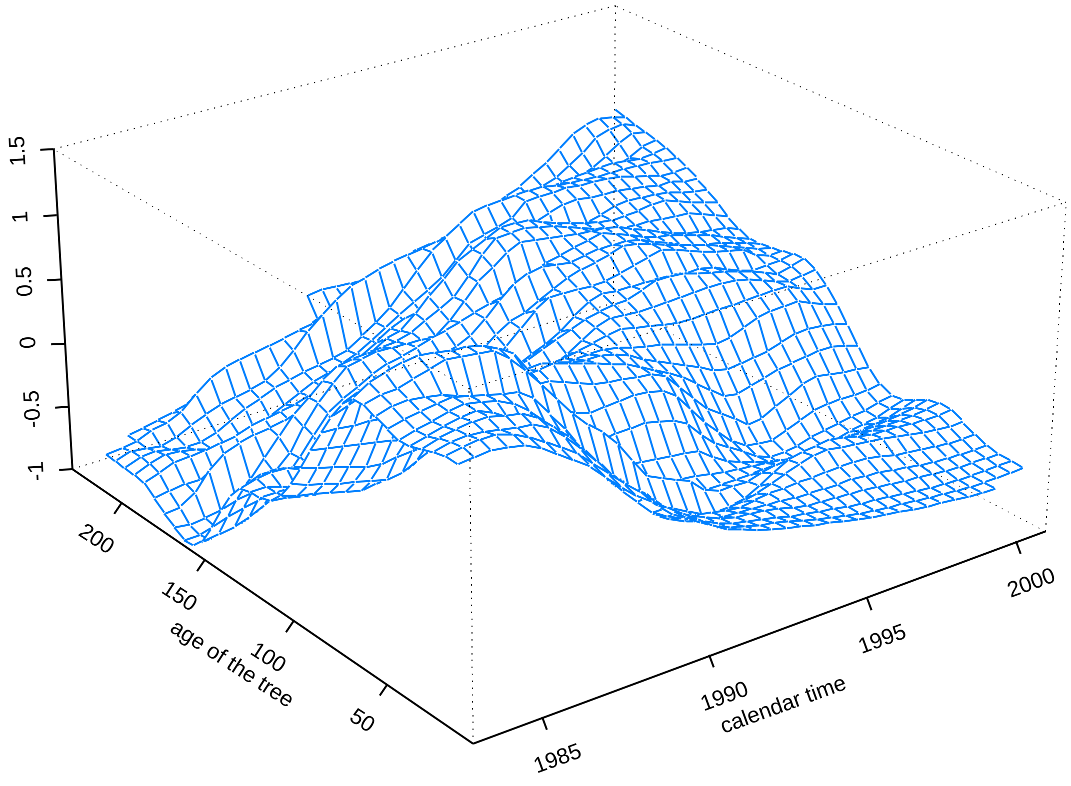

Figure 8: Interaction effect.

Table 1: Classifications with and without spatial effect.

	$\hat{y}$		
y	1	2	3
1	904	64	0
2	108	426	5
3	0	16	24

12.5%

	$\hat{y}$		
y	1	2	3
1	850	118	0
2	150	386	3
3	0	34	6

19.7%

Software

- Estimation was carried out using BayesX, a public domain software package for Bayesian inference.

- Available from

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

- Features (within a mixed model setting):
  - Responses: Gaussian, Gamma, Poisson, Binomial, ordered and unordered multinomial.
  - Continuous covariates and time scales: Random Walks, P-splines, autoregressive priors for seasonal components.
  - Spatial Covariates: Markov random fields, stationary Gaussian random fields, two-dimensional P-Splines.
  - Interactions: two-dimensional P-splines, varying coefficient models with continuous and spatial effect modifiers.
  - Random intercepts and random slopes.

References

- Fahrmeir, L., Kneib, T. and Lang, S. (2004): Penalized structured additive regression for space-time data: A Bayesian perspective. Under revision for Statistica Sinica.

- Kneib, T. and Fahrmeir, L. (2004): Structured additive regression for multicategorical space-time data: A mixed model approach. SFB 386 Discussion Paper 377, University of Munich.

- Both available from

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