Analysing geoadditive regression data: a mixed model approach.

Thomas Kneib
Department of Statistics, Ludwig-Maximilians-University Munich

Due to the increasing availability of spatial or spatio-temporal regression data, models that allow to incorporate the special structure of such data sets in an appropriate way are highly desired. A flexible modelling approach should not only be able to account for spatial and temporal correlation, but also to model further covariate effects in a semi- or nonparametric fashion.

We propose a general class of geoadditive regression models suitable for univariate responses, multicategorical responses and survival times, comprising the following features:

- estimation of nonparametric effects of continuous covariates and time scales based on penalised splines,
- estimation of spatial effects based on Markov random fields, stationary Gaussian random fields or bivariate extensions of penalised splines,
- inclusion of cluster- or individual-specific random effects (or frailties),
- modelling of interactions via varying coefficients terms or interaction surfaces.

For the analysis of survival times, additional functionality extending the Cox model includes:

- the ability to deal with arbitrary combinations of left, right and interval censoring as well as left truncation,
- joint estimation of covariate effects and baseline hazard rate,
- the possibility to include (piecewise constant) time-varying covariates,
- relaxation of the proportional hazards assumption via the inclusion of time-varying effects.

Inference is based on a unifying Bayesian formulation of the model components that endues all effects with appropriate priors of different degrees of smoothness but of one general form. This general form of the prior allows to rewrite the model as a simple mixed model where regression coefficients can be estimated based on penalised likelihood. The smoothing parameters of the original model formulation transform to variance components in the mixed model and are estimated jointly with the regression coefficients using (approximate) marginal likelihood.

Two complex examples will be discussed to emphasise the generality of the proposed model. In the first example we consider a regression model, where the health status of trees is assessed in three ordered categories. Yearly information on 83 beeches within a northern Bavarian forest district is available and, hence, the data set has both a spatial and a longitudinal arrangement. We will compare several spatial smoothing techniques and discuss appropriate modelling of the trend function.

The second example is an analysis of childhood undernutrition in Nigeria, where the survival time of children is obtained from a retrospective questionnaire of the parents. Due to memory effects, most of the uncensored survival times are rounded and have to be regarded as being interval censored. Covariates in the data set include spatial information on the district the mother lives in, a number of continuous covariates with possibly non-linear effects and also time-varying covariates such as whether the child is currently breastfed.

References:

- Fahrmeir, L., Kneib, T. and Lang, S. (2004): Penalized structured additive regression for space-time data: a Bayesian perspective. Statistica Sinica, 14, 731-761.
- Kneib, T. and Fahrmeir, L. (2005): Structured additive regression for multicategorical space-time data: A mixed model approach. Biometrics, to appear.
- Kneib, T. (2005): Geoadditive hazard regression for interval censored survival times. Discussion Paper 447, SFB 386, LMU München.