Structured additive regression for space-time data: A mixed model approach
Ludwig Fahrmeir, Thomas Kneib and Stefan Lang

Abstract

We propose extensions of penalized spline generalized additive models for analyzing space-time regression data and study them from a Bayesian perspective. Non-linear effects of metrical covariates and time trends are modelled through Bayesian versions of penalized splines, while smooth spatial effects follow a Markov random field prior. This allows to treat all functions and effects within the same general framework by assigning appropriate priors with different forms and degrees of smoothness. InfERENCE is based on a generalized linear mixed model representation. This approach can be viewed as posterior mode estimation and is closely related to penalized likelihood estimation in a frequentist setting. Variance components, corresponding to inverse smoothing parameters, are then estimated by using marginal quasi-likelihood.

Bayesian structured additive regression

Consider regression situations, where observations \((y_i, x_i, u_i), i = 1, \ldots, n\), on a response variable \(x\), a vector \(u\) of \(m\) metrical covariates, time scales or spatial covariates and a vector \(u\) of further covariates are given. Generalized additive and semiparametric models (Hastie and Tibshirani, 1990) assume, that given, \(x\) and \(u\), the distribution of \(y\) belongs to an exponential family, with mean \(\mu\)

\[ \mu = \exp(\mathbf{X} \beta + \mathbf{u} \gamma) \]

able to write evaluations at the observed values of \(x\).

We may estimate only a structured spatially correlated effect or split up the spatial effect into a structured (correlated) and an unstructured (un correlated) component:

\[ f_i(s) = f_{unstr}(s) + f_{str}(s) \]

where \(f_{unstr}(s)\) is the unstructured part of the spatial term and \(f_{str}(s)\) the structured part. For both cases the design matrix \(\mathbf{X}\) contains a basis of the nullspace of the penalty matrix, the penalty matrix

\[ \mathbf{X}' \Omega \mathbf{X} = \mathbf{0} \]

is that \(\mathbf{X}_{unstr} \Omega \mathbf{X}_{unstr} = \mathbf{I}\) and \(\mathbf{X}_{str} \Omega \mathbf{X}_{str} = \mathbf{0}\), where \(\mathbf{X}_{unstr}\) is the matrix for the unstructured part and \(\mathbf{X}_{str}\) for the structured part.

for \(\beta_{unstr}\) to be i.i.d. Gaussian.

\[ \beta_{unstr}, \tau_{unstr} \sim N(0, \tau^2) \]

and from the general prior (2) it follows that

\[ p(\beta_{unstr}) = \text{const} \]

Unstructured spatial effects are of special interest here, especially in the context of geostatistical applications. The smoothness of the spatial part is controlled by the penalty parameter \(\tau^2\). For large values of \(\tau^2\), the smoothing is small and vice versa.

Possible Extensions

Decompose the vectors of regression coefficients \(\beta\) into an unpenalized and a penalized part:

\[ \beta = \beta_{unstr} + \beta_{str} \]

with Gaussian errors \(\nu_{unstr} \sim N(0, \tau^2)\) and diffuse priors for initial values.


Spatial covariates

The values of \(x\) in \(\{1, \ldots, S\}\) represent the location or site in geographical regions.

Each site is associated with one parameter, i.e.

\[ f_{unstr}(s) + f_{str}(s) \sim N\left(\beta_{unstr}, \tau_{unstr}\right) \]

Unstructured group indicators

Suppose \(x\) is now a grouping variable with values \(x = 1, \ldots, M, M\). The values of \(x\) may denote a unit or cluster index, or the location in geographical maps.

To account for unobserved unit or group specific heterogeneity one possible way is to include an additive random effect into the predictor. Then we assume the group specific effects \(\beta_j\) to be i.i.d. Gaussian,

\[ \beta_j, \tau_j \sim N(0, \tau^2) \]

Adam is the design matrix for a 0/1 incidence matrix where the number of columns is equal to the number of different sites. If observation \(s\) belongs to site \(x\), then the element in the \(x\)-th column is one, otherwise.

To model interactions between covariates we consider the following model, which is a special case of (5) with the regions of a geographical map as the grouping variable.

Possible Extensions

• Interactions between covariates can be modelled through Varying Coefficient Models. This also allows to incorporate random slopes in the model, since models with unstructured group indicators as effect modifiers are equivalent to models with random slopes.

• A more flexible approach for modelling interactions between metrical covariates can be based on two dimensional surface fitting. Two dimensional P-splines, defined as the tensor product of two one dimensional B-Splines with a spatial smoothness prior, are described in Lang and Brezger (2003).

• All extensions can be cast into the general form (2) and may therefore be treated using the same methodology as presented here.

Mixed model representation

To rewrite the model (1) as a generalized linear mixed model (GLMM) we proceed as follows:

• Decompose the vectors of regression coefficients \(\beta\) into an unpenalized and a penalized part:

\[ \beta = \beta_{unstr} + \beta_{str} \]

with Gaussian errors \(\nu_{unstr} \sim N(0, \tau^2)\) and diffuse priors for initial values.

Bayesian structured additive regression

Consider regression situations, where observations \((y_i, x_i, u_i), i = 1, \ldots, n\), on a response variable \(x\), a vector \(u\) of \(m\) metrical covariates, time scales or spatial covariates and a vector \(u\) of further covariates are given. Generalized additive and semiparametric models (Hastie and Tibshirani, 1990) assume, that given, \(x\) and \(u\), the distribution of \(y\) belongs to an exponential family, with mean \(\mu\)

\[ \mu = \exp(\mathbf{X} \beta + \mathbf{u} \gamma) \]

able to write evaluations at the observed values of \(x\).

We may estimate only a structured spatially correlated effect or split up the spatial effect into a structured (correlated) and an unstructured (un correlated) component:

\[ f_i(x) = f_{unstr}(x) + f_{str}(x) \]

Each site is associated with one parameter, i.e.

\[ f_{unstr}(s) + f_{str}(s) \sim N\left(\beta_{unstr}, \tau_{unstr}\right) \]

A common choice are Markov random fields (Besag, York and Mollié, 1991) for the structured effect, e.g.

\[ \beta_{unstr}, \tau_{unstr} \sim N(0, \tau^2) \]

where \(\beta\) denotes the sites, which are neighbours of site \(s\) and \(N_s\) are the number of neighbours.

For the unstructured effect \(f_{unstr}\) we assume that the parameters \(\beta_{unstr}\) are i.i.d. Gaussian.

\[ \beta_{unstr}, \tau_{unstr} \sim N(0, \tau^2) \]

Finally, defining

\[ \mathbf{X} = (X_1 X_2 X_3 X_4 \cdots X_n X_{str}) \]

\[ \beta = (\beta_{unstr}, \cdots, \beta_{str})' \]

yields a generalized linear mixed model with linear predictor

\[ \eta = \mathbf{X}\beta + \mathbf{u}\gamma \]

and

\[ \beta_{unstr}, \tau_{unstr} \sim N(0, \tau^2) \]

The GLMM representation allows to examine the identification problem inherent to nonparametric regression from a different angle. Except for i.i.d. random effects the matrix product \(\mathbf{X}' \Omega \mathbf{X}\) contains the identity vector and therefore \(\mathbf{U}\) has not full column rank. So all identity vectors have to be eliminated from \(\mathbf{U}\) to guarantee identifiability.

Now we can utilize GLMM methodology for simultaneous estimation of the smooth functions and the variance parameters \(\tau^2\).

Especially, variance parameters may be estimated via marginal likelihood.

For Gaussian response the maximization of \(l(\tau^2)\) yields restricted maximum likelihood (REML) estimates. For more general responses a Laplace approximation to \(l(\tau^2)\) has to be used.

Since \(\tau^2\) is estimated via marginal likelihood, the estimates \(\hat{\beta}\) can be seen as empirical Bayes / posterior mode estimates.


Simulation study

We carefully compared the presented empirical Bayes approach with a fully Bayesian approach that uses MCMC techniques for posterior analysis (see Fahrmeir, Lang, 2001a, 2001b, Lang, Brezger, 2003 and Brezger, Lang, 2003) through a simulation study. The results of this simulation study can be summarized as follows:
In general, the empirical Bayes approach yields better point estimates of the functions $f_j$ in terms of MSE.

The differences are most noticeably for Borelli distributed response and turn out to be smaller for Gaussian, Poisson or Binomial distributed response with at least 3 repeated binary observations.

The empirical Bayes approach tends to smoother function estimates. This can also be shown theoretically (Kauermann, 2002).

Coverage probabilities meet the nominal level for smooth functions of metrical covariates. This is not the case for spatial and random effects, where coverages are far from the nominal level.

Since no problems with coverage probabilities occur in the fully Bayesian analysis a combination of both approaches seems to be promising: The variance components are estimated via marginal likelihood while the function estimates and the credible intervals are obtained from an MCMC analysis, that uses these variance components.

The combination leads to estimates, that keep the smaller MSE of the empirical Bayes estimates, but inherit the better coverage properties from the fully Bayesian analysis.

**Applications**

**Rents for flats: A spatial study**

According to the German law, owners of apartments or flats can base an increase in the amount that they charge for rent on “average rents” for flats comparable in type, size, equipment, quality and location in a community. To provide information about these “average rents”, most larger cities provide annotated databases. These databases are comparable in type, size, equipment, quality and location in a community.

With the help of these “average rents”, most larger cities will turn out to be smaller for Gaussian, Poisson or Binomial distributed response with at least 3 repeated binary observations.

Since no problems with coverage probabilities occur in the fully Bayesian analysis a combination of both approaches seems to be promising: The variance components are estimated via marginal likelihood while the function estimates and the credible intervals are obtained from an MCMC analysis, that uses these variance components.

The combination leads to estimates, that keep the smaller MSE of the empirical Bayes estimates, but inherit the better coverage properties from the fully Bayesian analysis.

**Figure 1: Estimated effect of floor space with pointwise 95% intervals.**

**Figure 2: Estimated effect of year of construction with pointwise 95% intervals.**

**Figure 3: Estimated spatial effect.**

**Figure 4: Temporal development of the frequency of damaged trees.**

The estimated spatial effect is given in Figure 9. It reflects the raw spatial effect shown in Figure 5 but also illuminates a spatial pattern with increased damage state around the village of Rothenbuch.

<table>
<thead>
<tr>
<th>Year</th>
<th>Rent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1983</td>
<td>0.125</td>
</tr>
<tr>
<td>1989</td>
<td>0.25</td>
</tr>
<tr>
<td>1995</td>
<td>0.375</td>
</tr>
<tr>
<td>2001</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Table 1: Classification table with and without spatial effect.

The presented mixed model approach is implemented in GGMAM, a software package that includes several Splus/R-functions. The program allows the estimation of non-linear effects of metrical covariates (modelled as P-splines), structured effects of spatial covariates (modelled as Markov random fields) and uncorrelated random effects (random intercepts and random slopes) for Gaussian, gamma, Poisson and Binomial distributed response. GGMAM is available from

www.stat.uni-muenchen.de/~knob

Fully Bayesian analyses have been carried out with BayesX, a software for Bayesian inference based on MCMC techniques. BayesX is available from

www.stat.uni-muenchen.de/~lang

**References**


