Lineare gemischte Modelle zur Schätzung von strukturiert additiven Regressionsmodellen

Thomas Kneib
Institut für Statistik, LMU München

1. Strukturiert additive Regression
2. Modellkomponenten und Priori-Verteilungen
3. Reparametrisierung als lineares gemischtes Modell
4. Inferenz in linearen gemischten Modellen
5. Beispiel: Modellierung von Waldschäden
6. Vergleich mit MCMC-Verfahren
7. Erweiterungen I + II

LMU

12.05.2004

SFB 386

Strukturiert additive Regression

- Generalisierte lineare Modelle:

$$E(y_{it}|u_{it},\gamma) = \mu_{it} = h(\eta_{it}) \qquad \eta_{it} = u_{it}'\gamma$$

  mit Regressionsparametern γ und Responsefunktion h.

- Probleme einer rein parametrischen Modellierung:
  - nicht-lineare Effekte metrischer Kovariablen,
  - zeitliche Korrelationen,
  - räumliche Korrelationen,
  - unbeobachtete Heterogenität,
  - komplexe Interaktionen zwischen Kovariablen.

$\Longrightarrow$ Strukturiert additive Regressionsmodelle

- Idee: Ersetze den linearen Prädiktor durch einen flexiblen, semiparametrischen Prädiktor.

- Raum-Zeit-Modell mit Haupteffekten

$$\eta_{it} = f_1(x_{it1}) + \ldots + f_l(x_{itl}) + f_{time}(t) + f_{spat}(s_i) + u_{it}'\gamma$$

  - $f_1, \ldots, f_l$ glatte Funktionen der metrischen Kovariablen $x_1, \ldots, x_l$,
  - f_{time} nichtlinearer Effekt der Zeit,
  - f_{spat} glatter räumlicher Effekt,
  - $u'\gamma$ üblicher parametrischer Teil des Prädiktors.

- Häufig ist es sinnvoll des zeitlichen Effekt aufzuspalten in Trend- und Saison-Komponente:

$$f_{time}(t) = f_{trend}(t) + f_{season}(t).$$

- Analog lässt sich der räumliche Effekt in einen strukturierten und einen unstrukturierten Anteil zerlegen:

$$f_{spat}(s) = f_{str}(s) + f_{unstr}(s).$$

- Erweiterungen des Haupteffekt-Modells
  - Individuenspezifische Effekte:

$$\eta_{it} = f_1(x_{it1}) + \ldots + u'\gamma + w'_{it}b_i$$

    mit u.i.v. zufälligen Effekten b_i.
  - Modelle mit Interaktionen

$$\eta_{it} = \ldots + f_{time}(t) + g(t)u_{it} + \ldots$$

$$\eta_{it} = \ldots + f_1(x_{it1}) + f_2(x_{it2}) + f_{1|2}(x_{it1}, x_{it2}) + \ldots$$

- Einheitliche Schreibweise

$$\eta_{it} = f_1(z_{it1}) + \ldots + f_p(z_{itp}) + u_{it}'\gamma$$

wobei

- $f_1, \ldots, f_p$ Funktionen verschiedenen Typs,
- $z_1, \ldots, z_p$ generische Kovariablen.
- Beispiele:

$f(v) = f(x)$	$v = x$	nonparametrische Funktion einer metrischen Kovariablen
$f(v) = f_{spat}(s)$	$v = s$	räumlicher Effekt
$f(v) = g(x)u$	$v = (x, u)$	Effekt mit variierenden Koeffizienten
$f(v) = f(x_1, x_2)$	$v = (x_1, x_2)$	Interaktionsoberfläche

etc.

Modellkomponenten und Priori-Verteilungen

- Alle Effekte können als Produkt einer Designmatrix Z_j und eines Vektors von Regressionsparametern β_j beschrieben werden:

$$f_j = Z_j \beta_j.$$

- Bayesianischer Ansatz: Priori-Verteilung für β_j.

- Allgemeine Form:

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

  wobei K_j eine Strafmatrix ist und τ_j^2 ein Glättungsparameter.

- Verbindung zu penalisierter ML-Schätzung:

$$P(\beta_j) = \log\left[p(\beta_j|\tau_j^2)\right] = -\frac{1}{2}\lambda_j \beta_j' K_j \beta_j, \qquad \lambda_j = 1/\tau_j^2$$

Stetige Kovariablen und Zeitskalen

- Bayesianische P-Splines:
  - Approximiere $f_j(x_j)$ durch einen B-Spline mit großer Knotenzahl, d.h.

$$f_j(x_j) \approx \sum_m \beta_{jm} B_m(x_j).$$

  - Die Designmatrix Z_j enthält die Auswertungen der Basisfunktionen an den beobachteten Werten von x_j.
  - Random Walk-Priori für die B-Spline-Koeffizienten β_j, d.h.

$$\beta_{jm} = \beta_{j,m-1} + u_{jm} \quad \text{oder} \quad \beta_{jm} = 2\beta_{j,m-1} - \beta_{j,m-2} + u_{jm}$$

    mit $u_{jm} \sim N(0, \tau_j^2)$.
  - Die Strafmatrix hat dann die Form $K_j = D'D$ mit Differenzenmatrix D.

- Random Walk-Prioris
  - Populär zur Modellierung von Zeittrends.
  - Können als P-Splines vom Grad 0 mit Knoten an allen verschiedenen Beobachtungswerten aufgefasst werden.

- Autoregressive Prioris
  - Zur Modellierung flexibler Saisonkomponenten.
  - Priori für Saisonkomponente $f_{season}(t) =: \beta_t$ mit Periode per:

$$\beta_t = -\sum_{j=1}^{per-1} \beta_{t-j} + u_t$$

wobei $u_t \sim N(0, \tau^2_{season})$.

Räumliche Kovariablen

- Stationäre Gauss-Felder
  - Exakte räumliche Information $s = (s_x, s_y)$.
  - Annahme: $f_{spat}(s)$ folgt einem stationären Gauss-Prozess mit

$$f_{spat}(s) \sim N(0, \tau^2_{spat})$$

    und isotroper Korrelationsfunktion

$$C(s, s') = C(||s - s'||).$$

  - Die Strafmatrix K wird durch die Korrelationsfunktion bestimmt.
  - Stationäre Gauss-Felder können als Oberflächenschätzer mit speziellen Basisfunktionen betrachtet werden.
  - Alternative: Zweidimensionale P-splines.

- Markov Zufallsfelder
  - Keine exakten Lokationen, sondern $s \in \{1, \ldots, S\}$ gibt Zugehörigkeit zu Regionen an.
  - Definiere geeignete Nachbarschaften.
  - Annahme: $f_{spat}(s)$ ist das gewichtete Mittel der Funktionswerte der benachbarten Regionen.
  - Die Strafmatrix hat die Form einer Nachbarschaftsmatrix.

Zufällige Effekte

- Können verwendet werden zur Modellierung von
  - unbeobachteter Heterogenität zwischen Clustern,
  - individuenspezifischen Effekten,
  - unstrukturierten räumlichen Effekten.

Interaktionen

- Variierende Koeffizienten
  - Modellierung von Interaktionen der Form

$$f(z_1, z_2) = g(z_1)z_2$$

    d.h. der Effekt von z_2 variiert über den Wertebereich von z_1.
  - Häufig ist z_2 kategorial.
  - $g(z_1)$ kann prinzipiell jede beliebige Funktion sein, z.B. ein P-Spline, eine Saisonkomponente, ein Markov-Zufallsfeld, . . .

- Oberflächenschätzer: Zweidimensionale P-Splines
  - Verwende Tensorprodukte eindimensionaler B-Splines als Basisfunktionen.
  - Definiere zweidimensionalen Random Walk als Priori für die Koeffizienten.

Reparametrisierung als lineares gemischtes Modell

- Für gegebene Glättungsparameter können die Regressionsparameter über modifiziertes Fisher-Scoring bestimmt werden.

- Problem: Wie bestimmt man die Glättungsparameter?
  - Generalisierte Kreuzvalidierung,
  - Voller Bayes-Ansatz mit MCMC-Verfahren,
  - Verwendung von Methoden für lineare gemischte Modelle.

- Ziel: Strukturiert additive Regressionsmodelle umschreiben zu linearen gemischten Modellen (genauer: zu Varianzkomponenten-Modellen).

- Jeder Parametervektor β_j kann zerlegt werden in einen unpenalisierten Teil (mit flacher Priori) und einen penalisierten Teil (mit i.i.d. Normalverteilungspriori):

$$\beta_j = X_j^{unp}\beta_j^{unp} + X_j^{pen}\beta_j^{pen}.$$

- Fall 1: Strafmatrix mit vollem Rang
  - Kein unpenalisierter Anteil.
  - $X_j^{pen}\beta_j^{pen} = K_j^{-1/2}\beta_j^{pen}$, d.h. β_j^{pen} erhält man durch Standardisierung von β_j.
- Fall 2: Strafmatrix mit Rangabfall
  - X_j^{unp} enthält eine Basis des Nullraums von K_j. (Was wird nicht von K_j penalisiert?)
  - X_j^{pen} enthält eine orthonormale Basis der Abweichungen von diesem Nullraum (orthogonal zur Basis in X_j^{unp}).
- Die unpenalisierten Anteile von f_j entsprechen der Funktionsschätzung für $\lambda_j \to \infty$, z.B.
  - einem Polynom vom Grad $k-1$ für P-splines mit RW-k Priori,
  - einer starren Saisonkomponente für einen flexiblen saisonalen Effekt,
  - einem konstanten Effekt für Markov Zufallsfelder.

- Aus der Zerlegung von β_j folgt eine Zerlegung von f_j:

$$\begin{aligned} f_j &= Z_j X_j^{unp} \beta_j^{unp} + Z_j X_j^{pen} \beta_j^{pen} \\ &= Z_j^{unp} \beta_j^{unp} + Z_j^{pen} \beta_j^{pen}. \end{aligned}$$

- Insgesamt erhält man ein Varianzkomponenten-Modell mit

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

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

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

Inferenz in linearen gemischten Modellen

- Es werden abwechelnd neue Schätzungen für die Regressionsparameter und die Varianzparameter bestimmt.

- Neue Regressionsparameter erhält man durch Lösen des Gleichungssystems

$$\begin{pmatrix} X'WX & X'WP \\ P'WX & P'WP + \Lambda^{-1} \end{pmatrix} \begin{pmatrix} \beta^{unp} \\ \beta^{pen} \end{pmatrix} = \begin{pmatrix} X'W\tilde{y} \\ P'W\tilde{y} \end{pmatrix}$$

  mit den üblichen GLM-Gewichten W und Arbeitsbeobachtungen $\tilde{y}$.

- Varianzparameter werden durch Maximieren der marginalen Likelihood (Restricted Likelihood) gewonnen:

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

- Man erhält empirische Bayes-Schätzer / Posteriori-Modus-Schätzer.

Beispiel: Modellierung von Waldschäden

- Daten stammen aus jährlichen Waldschadens-Beurteilungen in Nordbayern in den Jahren 1983 bis 2001.

- 83 Beobachtungspunkte mit Buchen.

- y_{it} Entlaubungsgrad von Baum i im Jahr t in drei geordneten Kategorien:

  $y_{it} = 1$ keine Entlaubung,
  $y_{it} = 2$ 25% Entlaubung oder weniger,
  $y_{it} = 3$ mehr als 25% Entlaubung.

- Kovariablen:

  t Kalenderzeit,
  s_i Standort der Buche,
  a_{it} Alter in Jahren,
  u_{it} weitere (hauptsächlich kategoriale) Kovariablen.

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

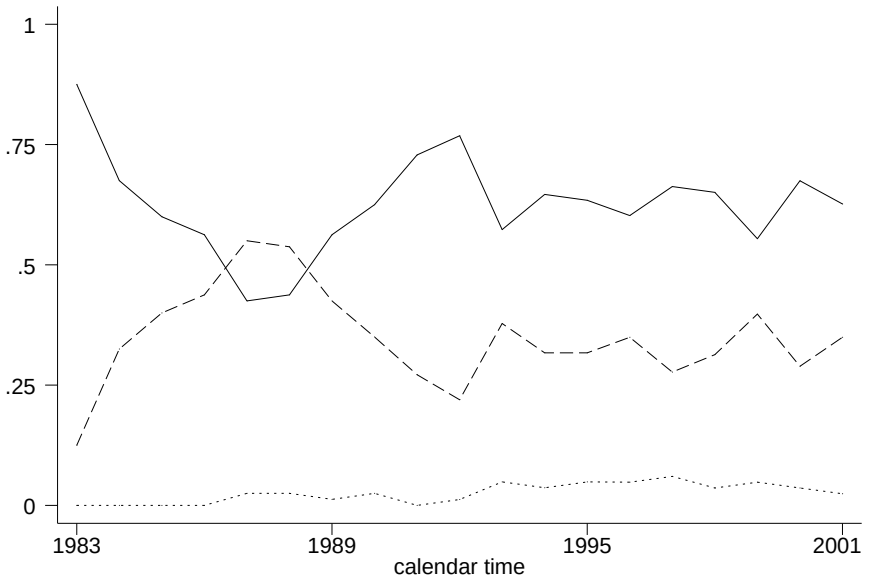

Abbildung 1: Zeitliche Entwicklung des Anteils der verschiedenen Schädigungsstufen .

— keine Schädigung,

- - - mittlere Schädigung,

· · · schwere Schädigung.

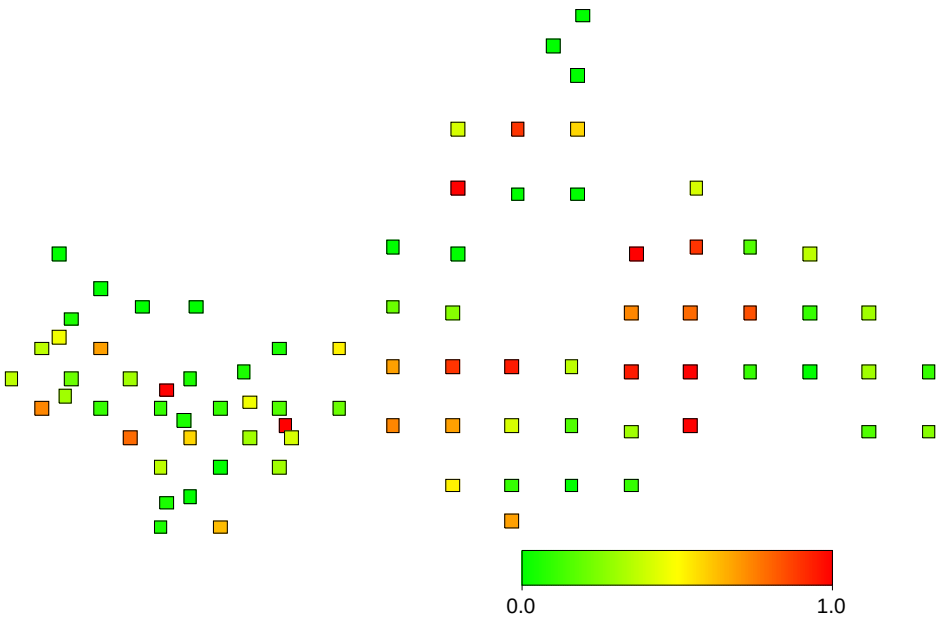

Abbildung 2: Räumliche Verteilung der Bäume und Anteil der Zeitpunkte zu denen ein Baum als geschädigt eingestuft wurde.

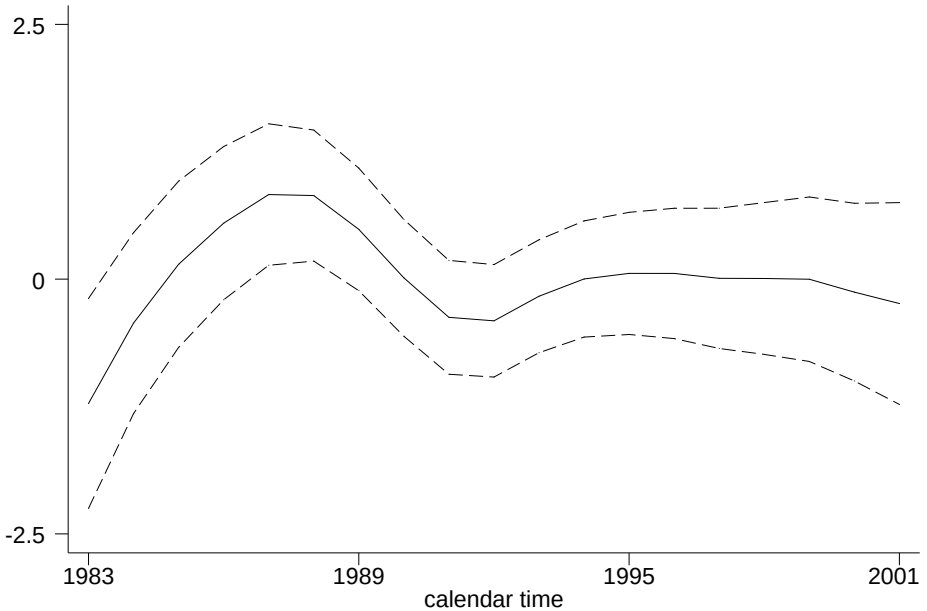

Abbildung 3: Effekt der Kalenderzeit (Trend).

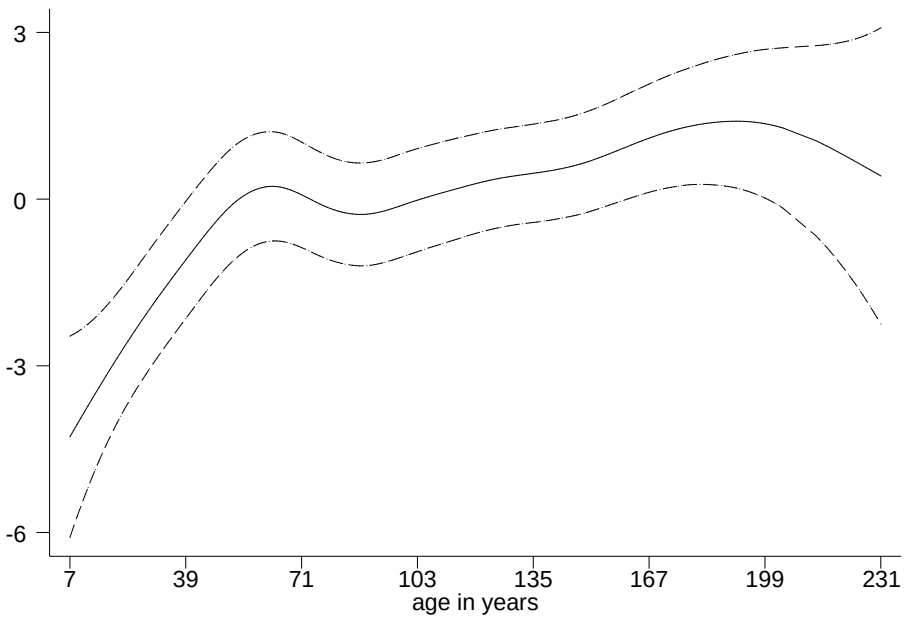

Abbildung 4: Alterseffekt

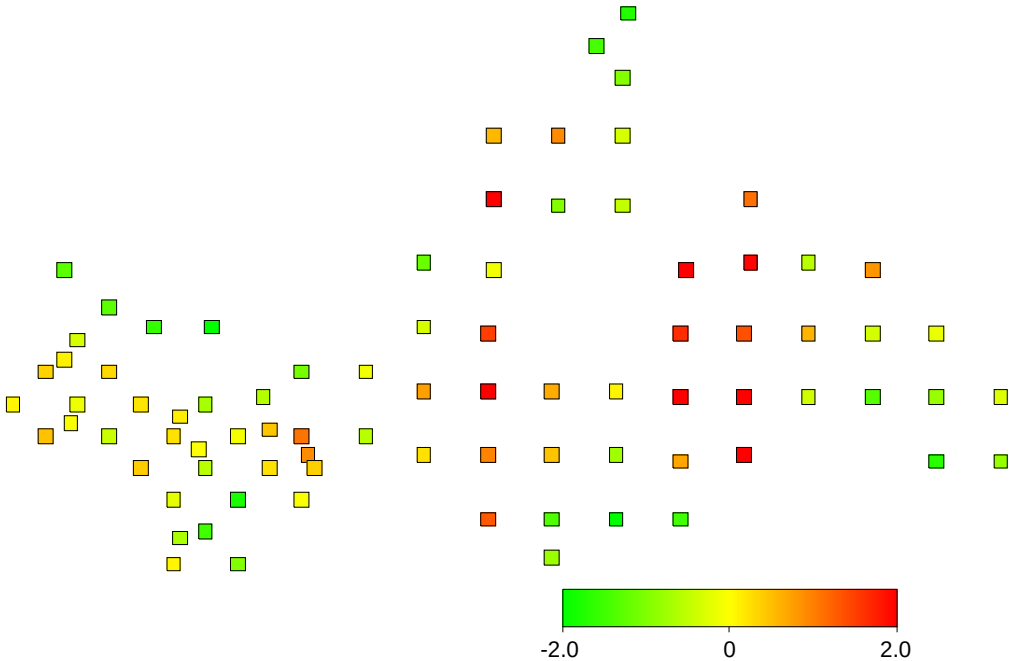

Abbildung 5: Strukturierter räumlicher Effekt.

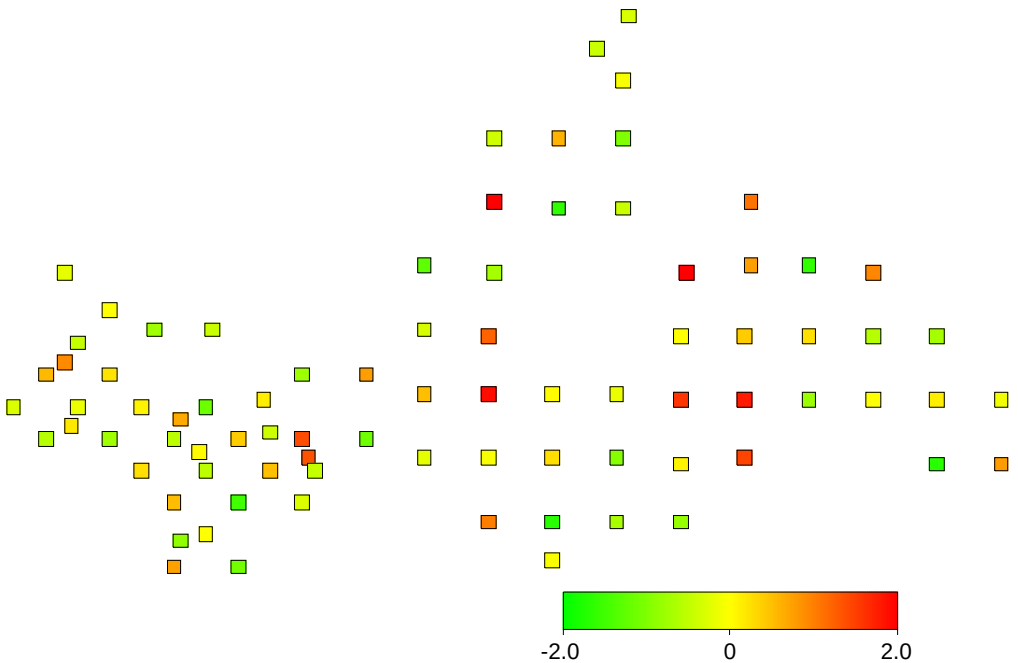

Abbildung 6: Unstrukturierter räumlicher Effekt.

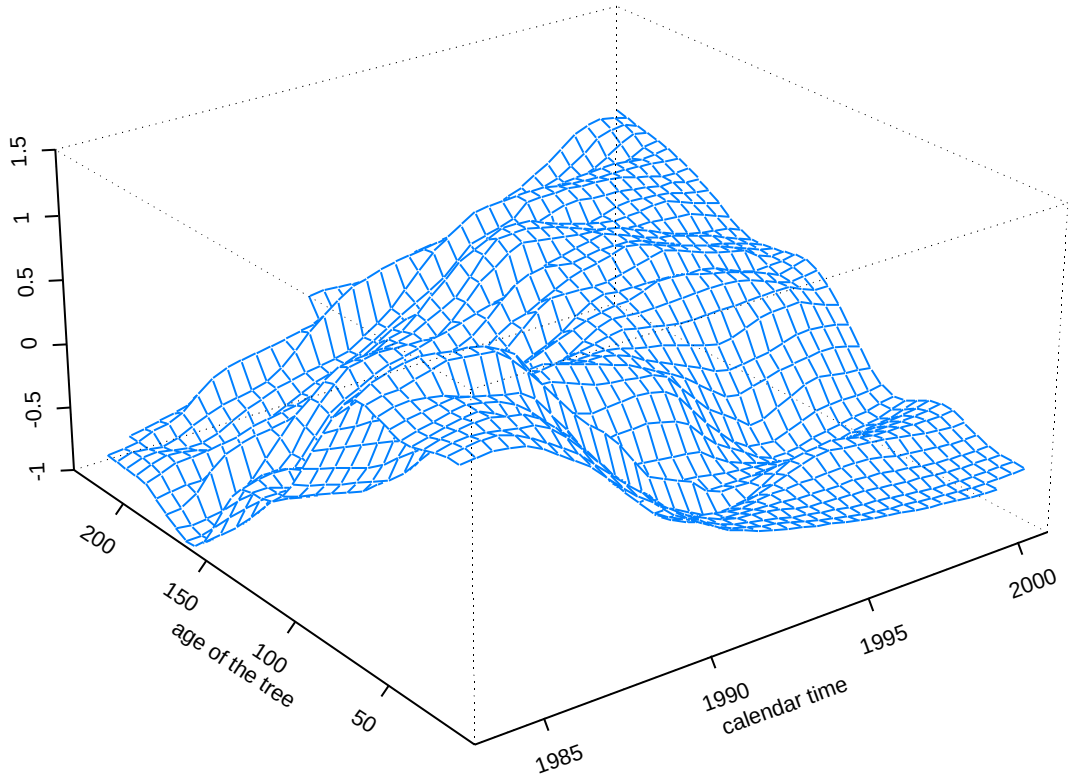

Abbildung 7: Interaktionseffekt.

Tabelle 1: Klassifikationen mit und ohne räumlichem Effekt .

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

12.5%

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

19.7%

Vergleich mit MCMC-Verfahren

Nachteile:

- Nicht modular aufgebaut. Der numerische Aufwand wächst schneller an.
- Kredibilitätsintervalle nur asymptotisch.
- Nur Plug-in Schätzungen für Funktionale der Parameter.
- Schwieriger zu erweitern.

Vorteile:

- Keine Fragen nach Konvergenz und Mixing.
- Liefert etwas bessere Punktschätzungen.
- In einfachen Modellen wesentlich schneller.
- Bayesianische Betrachtungsweise ist nicht notwendig (?)
- Weniger Vorwissen zur statistischen Theorie nötig (?)

Erweiterungen I: Multikategorialer Response

- Bereits entwickelt: Modelle für "einfache" multinomiale Logit-Modelle sowie ordinale Logit- und Probit-Modelle.
- Mögliche Weiterentwicklungen:
  - Multinomiale Probit-Modelle, insbesondere basierend auf korrelierten latenten Variablen.
  - Kategorienspezifische Effekte bei nominalem Response.
  - Kovariablenabhängige Schwellenwerte bei ordinalem Response.
- Probleme:
  - Bestimmung von Wahrscheinlichkeiten aus multivariaten Normalverteilungen.
  - Identifizierbarkeit bei ordinalen Modellen.

Erweiterungen II: Verweildaueranalyse

- Bereits entwickelt (nicht von uns): Nonparametrische Modellierung der Baseline-Hazardrate.
- Ziel: Erweiterung auf allgemeinen strukturiert additiven Prädiktor.
- Mögliche Methoden:
  - Numerische Maximierung der marginalen Likelihood bezüglich den Glättungsparametern.
  - Anpassung von Methoden für Frailty-Modelle im Cox-Modell.