

Bachelor Thesis

**Prognostische Faktoren für  
Wildverbiss**

von

Lisa Möst

Betreuer:

Prof. Dr. Torsten Hothorn, M.Sc. Nikolay Robinzonov

Institut für Statistik

Ludwig-Maximilians-Universität München

23. Juli 2009

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>4</b>
<b>2. Generalisierte Lineare Regressionsmodelle</b>	<b>6</b>
2.1. Generalisierte Lineare Modelle . . . . .	6
2.2. Modell für binäre Daten . . . . .	10
2.3. Modell für Zähldaten . . . . .	13
<b>3. Generalisierte Additive Modelle</b>	<b>15</b>
3.1. Nonparametrische Regression . . . . .	15
3.1.1. Univariate Glättung . . . . .	15
3.1.2. Basisfunktionsansätze . . . . .	16
3.1.3. Penalisierte KQ-Schätzung . . . . .	20
3.1.4. Bias-Varianz Trade-off und Glättungsparameterwahl . . . . .	21
3.2. Additive Modelle . . . . .	23
3.3. Generalisierte Additive Modelle . . . . .	25
3.3.1. Generalisierte nonparametrische Regression . . . . .	25
3.3.2. Inferenz . . . . .	27
3.3.3. Glättungsparameterwahl . . . . .	28
3.4. Generalisierte Additive Modelle in $\mathbb{R}$ . . . . .	29
<b>4. Boosting von Additiven Modellen</b>	<b>30</b>
4.1. Der FGD-Algorithmus . . . . .	30
4.2. Wahl der base-procedure . . . . .	32
4.3. Boosting für binäre Zielgrößen . . . . .	34
4.4. Boosting für Zähldaten . . . . .	35
4.5. Boosting für stetige Zielgrößen . . . . .	35
4.6. Tuning-Parameter . . . . .	36
4.7. Boosting in $\mathbb{R}$ . . . . .	37
<b>5. Datensituation</b>	<b>39</b>
<b>6. Prognostische Faktoren für Wildverbiss</b>	<b>43</b>
6.1. Allgemeine Vorgehensweise . . . . .	43
6.2. Vogelbeere . . . . .	45
6.2.1. Binomialmodell . . . . .	45
6.2.2. Poissonmodell . . . . .	50

## *Inhaltsverzeichnis*

6.3. Tanne . . . . .	53
6.3.1. Binomialmodell . . . . .	53
6.3.2. Poissonmodell . . . . .	57
6.4. Buche . . . . .	60
6.4.1. Binomialmodell . . . . .	60
6.4.2. Poissonmodell . . . . .	64
6.5. Bergahorn . . . . .	68
6.5.1. Binomialmodell . . . . .	68
6.5.2. Poissonmodell . . . . .	71
6.6. Fichte . . . . .	74
6.6.1. Binomialmodell . . . . .	74
6.6.2. Poissonmodell . . . . .	78
<b>7. Zusammenfassung</b>	<b>83</b>
<b>A. Graphiken Vogelbeere</b>	<b>86</b>
<b>B. Graphiken Tanne</b>	<b>89</b>
<b>C. Graphiken Buche</b>	<b>92</b>
<b>D. Graphiken Bergahorn</b>	<b>96</b>
<b>E. Graphiken Fichte</b>	<b>98</b>

# 1. Einleitung

„Natur Natur sein lassen“ lautet das Motto des Nationalparks „Bayerischer Wald“. Dadurch ergibt sich die Möglichkeit die Entwicklung der Natur nach ihren ureigensten Mechanismen zu beobachten.

Im Rahmen der Waldinventur 2002 wurden zahlreiche umweltspezifische Faktoren für die Region erhoben. Hierfür wurde das Gebiet des Nationalparks in so genannte Inventurpunkte (InvPkt) aufgeteilt, die durch ihre Koordinaten exakt identifizierbar sind. In den Inventurpunkten wurden sowohl geographische Eckdaten, wie z.B. die Höhe über NN oder die Hangneigung, als auch umweltspezifische bzw. ökologische Kenngrößen, z.B. die Bodenstruktur (Gras, Farn, Fels, usw.), erhoben. Des Weiteren liegen für die untersuchten Bereiche die Anzahl verbissener Bäume sowie die zugehörige Baumart vor. Zusätzlich ist eine Angabe der Reh-, Rothirsch- und Luchsdichte für die Inventurpunkte vorhanden. Gegenstand dieser Arbeit ist es, für die erfassten Inventurpunkte den Einfluss der erhobenen Kovariablen auf den Baumverbiss zu untersuchen. Dabei wurden getrennt für die fünf Baumarten Buche, Bergahorn, Tanne, Fichte und Vogelbeere geeignete Modelle angepasst. Zur Modellierung wurden einerseits logistische Regressionsmodelle, die es ermöglichen die *Chance* auf Verbiss zu schätzen, verwendet. Andererseits kann die erwartete Anzahl verbissener Bäume für die jeweilige Baumart mit Poissonmodellen untersucht werden. Neben der Identifizierung jener Einflussgrößen, die tatsächlich einen Einfluss auf den Verbiss der Baumart haben, war die Spezifizierung der Art des Einflusses eine entscheidende Fragestellung bei der Modellfindung.

Zu Beginn der Arbeit folgen einige theoretische Kapitel, worin die verwendeten Methoden zur Schätzung der endgültigen Regressionsmodelle vorgestellt werden. Hierbei beschäftigt sich Kapitel 2 mit Generalisierten Linearen Regressionsmodellen, wobei besonders auf das logistische Regressionsmodell sowie die Modellierung von Zähldaten im Poissonmodell eingegangen wird. Kapitel 3 enthält die weitere Verallgemeinerung der Generalisierten Linearen Regressionsmodelle zu den Generalisierten Additiven Regressionsmodellen. Die Modellierung nonparametrischer Einflussgrößen findet hier besondere Beachtung. Anschließend wird in Kapitel 4 Boosting zur Konstruktion eines Generalisierten Additiven Modells vorgestellt, wobei die darin bereits integrierte Variablenselektion besonders von Vorteil ist.

Im zweiten Teil der Arbeit werden die Ergebnisse für die Daten aus dem Nationalpark präsentiert. Hierzu wird in Kapitel 5 ein Überblick über die Datensituation gegeben. In Kapitel 6 wird zunächst die genaue Vorgehensweise bei der Modellierung des Einflusses der Kovariablen auf die Zielgröße beschrieben. An-

## *1. Einleitung*

schließlich werden sowohl das beste Binomial- sowie das beste Poissonmodell getrennt für die fünf Baumarten präsentiert. Zuletzt werden die gefundenen Ergebnisse noch einmal kurz zusammengefasst, sowie Gemeinsamkeiten und Unterschiede der Modelle herausgearbeitet.

## 2. Generalisierte Lineare Regressionsmodelle

In diesem Kapitel soll eine kurze Einführung in die Verwendung und die Eigenschaften der Generalisierten Linearen Regressionsmodelle gegeben werden. Dabei wird im Besonderen auf das logistische Regressionsmodell zur Modellierung binärer Zielgrößen, sowie das Poissonmodell zur Modellierung von Zähldaten eingegangen, da diese zur Identifikation der prognostischen Faktoren für Wildverbiss in Kapitel 6 verwendet werden.

### 2.1. Generalisierte Lineare Modelle

Das klassische lineare Regressionsmodell ist für stetige Zielvariablen geeignet, die (nach geeigneter Transformation) approximativ normalverteilt sind. In vielen Situationen trifft jedoch die Annahme der approximativen Normalverteilung für die Zielgröße nicht zu und das gewöhnliche lineare Regressionsmodell kann zur Modellanpassung nicht verwendet werden. Eine Alternative bildet die Anpassung eines Generalisierten Linearen Regressionsmodells (GLM), das bei vielen weiteren Verteilungen der Zielgröße geeignet ist. Das GLM ist grundsätzlich durch eine Verteilungs- und eine Strukturannahme definiert. Eine ausführliche Einführung wird beispielsweise in [McCullagh/Nelder 1989] gegeben.

#### Verteilungsannahme

Für gegebene Kovariablen  $x_i = (1, x_{i1}, \dots, x_{ip})$  sind die Zielvariablen  $y_i$  bedingt unabhängig und deren bedingte Dichte (Binomial-, Poissonverteilung) muss in Form einer *einparametrischen Exponentialfamilie*

$$f(y_i|\theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi}\omega_i + c(y_i, \phi, \omega_i)\right) \quad (2.1)$$

darstellbar sein. Dabei bezeichnet  $\theta_i$  den *natürlichen* bzw. *kanonischen Parameter*,  $\phi$  einen von  $i$  unabhängigen Dispersionsparameter und  $\omega_i$  üblicherweise Gewichte. Für die Funktion  $b(\theta_i)$  müssen die erste und zweite Ableitung existieren, da diese zur Spezifikation des Erwartungswerts und der Varianz dienen:

$$\mathbb{E}(y_i) = \mu_i = b'(\theta_i); \quad \mathbb{V}(y_i) = \sigma_i^2 = \phi b''(\theta_i) / \omega_i. \quad (2.2)$$

## 2. Generalisierte Lineare Regressionsmodelle

Die Funktion  $c(y_i, \phi, \omega_i)$  ist von sekundärem Interesse.

### Strukturannahme

Im klassischen linearen Modell wird angenommen, dass die Kovariablen linear auf den Erwartungswert  $\mathbb{E}(y|\mathbf{x})$  wirken. Ist die Zielgröße jedoch beispielsweise nicht stetig und normalverteilt, sondern binär, strikt positiv oder ein Zählprozess, so ist diese Modellierung häufig wenig sinnvoll. Deshalb wird der Erwartungswert  $\mathbb{E}(y_i|x_i)$  der Zielgröße  $y_i$  mit dem linearen Prädiktor  $\eta_i = x_i^T \beta = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$  über die Responsefunktion  $h$ , bzw. die Linkfunktion  $g = h^{-1}$  verknüpft:

$$\mu_i = h(\eta_i) = h(x_i^T \beta); \quad \eta_i = g(\mu_i). \quad (2.3)$$

Dabei leistet die Responsefunktion die Abbildung in den zulässigen Wertebereich. Beispielsweise garantiert bei einem Binomialmodell die Funktion  $h$ , dass für den Erwartungswert  $\mu_i \in [0, 1]$  gilt. Des Weiteren muss die Responsefunktion  $h$  invertierbar und zweimal stetig differenzierbar sein.

### Natürliche Linkfunktion

Jede Verteilung, die sich in der Form einer *einparametrischen Exponentialfamilie* darstellen lässt, besitzt eine *natürliche* bzw. *kanonische Linkfunktion*. Eine Linkfunktion heißt *kanonische Linkfunktion*, wenn

$$\theta = \mu \quad (2.4)$$

gilt, wobei  $\theta$  der *kanonische Parameter* ist.

Beispielsweise ist für binomial-verteilte Zielgrößen der *Logit-Link*, für poisson-verteilte Zielgrößen der *log-Link* der natürliche Link.

Durch die Wahl des kanonischen Links ergeben sich einige wünschenswerte Eigenschaften, z.B. die Vereinfachung der Score-Funktion (vgl. Kapitel ML-Schätzung), jedoch führt diese Wahl für  $g$  nicht immer zur besten Modellanpassung [McCullagh/Nelder 1989].

### Maximum-Likelihood-Schätzung

Ziel bei der Schätzung des Parametervektors  $\beta$  ist es, die Schätzer so zu bestimmen, dass eine möglichst gute Anpassung des Modells an die Daten resultiert. Die Parameterschätzer werden durch die Maximierung der vollständigen *Likelihood* bzw. der *log-Likelihood* der Daten ermittelt.

Auf Grund der Unabhängigkeit der Zielvariablen  $y|\mathbf{x}$  ergibt sich die vollständige

## 2. Generalisierte Lineare Regressionsmodelle

Likelihood als Produkt der Dichten  $f(y_i|\beta; x_i)$  [Fahrmeir et al 2007]

$$L(\beta) = \prod_{i=1}^n f(y_i|\beta; x_i). \quad (2.5)$$

Da für ein Generalisiertes Lineares Regressionsmodell lediglich Verteilungsannahmen in Form einer einparametrischen Exponentialfamilie zulässig sind, liegt die *log-Likelihood* allgemein in der Form

$$l_i(\beta) = \log(f(y_i|\beta)) = \frac{y_i\theta_i - b(\theta_i)}{\phi} \omega_i \quad (2.6)$$

vor. Die *log-Likelihood* hängt hierbei über den natürlichen Parameter  $\theta_i = \theta(\mu_i)$ ;  $\mu_i = h(x_i'\beta)$  vom Parametervektor  $\beta$  ab.

Zur Bestimmung des *Maximum-Likelihood-Schätzers* muss zunächst die erste Ableitung der *log-Likelihood*, die *Score-Funktion*, berechnet werden. Diese weist ebenfalls eine für alle Verteilungsannahmen typische Form auf:

$$s(\beta) = \frac{\partial l(\beta)}{\partial \beta} = \sum x_i \frac{d_i}{\sigma_i^2} (y_i - \mu_i); \quad (2.7)$$

$$\mu_i = \mathbb{E}(y_i|x_i) = h(x_i'\beta); \quad \sigma_i^2 = \mathbb{V}(y_i|x_i); \quad d_i = \frac{\partial h(\eta_i)}{\partial \eta_i}.$$

Die Maximum-Likelihood-Schätzer ergeben sich durch die Bedingung

$$s(\hat{\beta}) = 0$$

und werden durch den *Fisher-Scoring-Algorithmus* bestimmt, der auf dem *Newton-Raphson-Algorithmus* basiert. Im Gegensatz zum *Newton-Raphson-Algorithmus* wird jedoch die *erwartete Fisher-Matrix* anstelle der *beobachteten Fisher-Matrix* verwendet. Beim *Fisher-Scoring* Verfahren nähert man sich durch mehrere Iterationen einer *reweighted Least-Squares-Procedure* dem Parameterschätzer an. Nähere Details sind in [Fahrmeir et al 2007] zu finden.

Die Kovarianzmatrix von  $\beta$ , bzw. die *Fisher-Matrix* ist durch

$$F(\beta) = \sum x_i x_i' \omega_i; \quad \omega_i = \frac{d_i^2}{\sigma_i^2} \quad (2.8)$$

definiert.

Wählt man den kanonischen Link als Linkfunktion, so vereinfachen sich die *Score-Funktion* und die *Fisher-Matrix*. Darüber hinaus sind auch der *Newton-Raphson* und der *Fisher-Scoring-Algorithmus* in diesem Fall identisch.

Für den ML-Schätzer gilt unter Regularitätsbedingungen asymptotische Konsistenz und Normalität:

$$\hat{\beta}_n \stackrel{a}{\sim} N(\beta, F^{-1}(\beta)). \quad (2.9)$$

## 2. Generalisierte Lineare Regressionsmodelle

Hierbei bezeichnet  $\hat{\beta}_n$  den ML-Schätzer in Abhängigkeit vom Stichprobenumfang  $n$ . Für groß genug gewähltes  $n$ , ist der ML-Schätzer nicht nur erwartungstreu, sondern auch normalverteilt mit der inversen Fisher-Matrix als Kovarianzmatrix. Ersetzt man die Fisher-Matrix des wahren Parameters  $F(\beta)$  durch die Schätzung  $F(\hat{\beta})$ , so bleibt die asymptotische Normalität des ML-Schätzers erhalten [Fahrmeir et al 2007].

### Modellselektion

Ziel bei der Bestimmung des besten GLMs ist es, ein Modell zu finden, das gut an die Daten angepasst ist, aber gleichzeitig möglichst wenige Parameter enthält. Eine zu hohe Anzahl an Einflussgrößen führt dazu, dass das Modell unnötig kompliziert wird und erschwert eine gute, sinnvolle Interpretation. Somit sollen die Einflussgrößen identifiziert werden, die den größten Einfluss auf die Zielgröße haben um anschließend mit Hilfe dieser Variablen ein Modell anzupassen, das die vorliegenden Daten möglichst gut beschreibt.

Die Devianz ist ein Maß um die Güte der Anpassung des Modells an die Daten zu messen:

$$D(y; \mu) = -2\phi \{l(\hat{\mu}; y) - l(y; y)\}; \quad l(\hat{\mu}; y) = \sum_i \log f_i(y_i; \theta_i). \quad (2.10)$$

Dabei bezeichnet  $l(y; y)$  die maximal erreichbare log-Likelihood, da die gefitteten Werte exakt den beobachteten Werten entsprechen. Dieses Modell ist aber uninformativ, da es die Daten nicht zusammenfasst, sondern sie mit Hilfe von  $n$  Parametern schlichtweg wiederholt. Mit  $l(\hat{\mu}; y)$  wird der Betrag der log-Likelihood an der Stelle des gerade betrachteten Schätzers  $\hat{\mu}$  bezeichnet. Somit beschreibt die Devianz, ausgewertet für das gerade betrachtete Modell, die Abweichung der Anpassung durch dieses Modell von der maximal möglichen Anpassung. Generell werden die Werte von  $\mu$  nie exakt mit den  $y$ -Werten übereinstimmen, jedoch ist eine möglichst kleine Devianz erstrebenswert. Gesucht wird ein Modell mit  $p$  Parametern, das möglichst gut an die Daten angepasst ist, und dessen log-Likelihood damit möglichst nah an  $l(y; y)$  liegt [McCullagh/Nelder 1989].

Für das Poisson- und das Binomial-Modell ergeben sich folgende Devianzen:

$$\begin{aligned} \text{Binomial: } D(y; \mu) &= 2 \sum_{i=1}^n \left\{ n_i y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) + n_i (1 - y_i) \log \left( \frac{1 - y_i}{1 - \hat{\mu}_i} \right) \right\} \\ \text{Poisson: } D(y; \mu) &= 2 \sum_{i=1}^n \left\{ y_i \log \left( \frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right\} \end{aligned}$$

Eine geeignete Teststatistik zum Modellvergleich ist der *Likelihood-Quotienten-Test*. Bei der Suche nach dem „besten“ Modell ist besonders die Frage interessant, ob die jeweilige Kovariable überhaupt einen Einfluss auf die Zielgröße

## 2. Generalisierte Lineare Regressionsmodelle

hat, oder ob sie im Modell vernachlässigt werden kann.  
Die Hypothesen des Tests werden durch

$$H_0 : C\beta = d \quad H_1 : C\beta \neq d \quad (2.11)$$

formuliert, wobei die Matrix C und der Vektor d zur Modellierung der Restriktionen an den Parametervektor  $\beta$  benötigt werden.

Die *Likelihood-Quotienten-Statistik*

$$lq = -2 \left\{ l(\tilde{\beta}) - l(\hat{\beta}) \right\}$$

$\tilde{\beta}$ : Schätzer des restringierten Modells

$\hat{\beta}$ : Schätzer des vollen Modells

kann auch mittels der Devianzen der Modelle ausgedrückt werden:

$$\begin{aligned} D(\tilde{\beta}) &= -2\phi \left\{ l(\tilde{\beta}) - l(\bar{y}) \right\} \Rightarrow l(\tilde{\beta}) = \frac{D(\tilde{\beta})}{-2\phi} + l(\bar{y}) \\ D(\hat{\beta}) &= -2\phi \left\{ l(\hat{\beta}) - l(\bar{y}) \right\} \Rightarrow l(\hat{\beta}) = \frac{D(\hat{\beta})}{-2\phi} + l(\bar{y}) \\ lq &= \frac{D(\tilde{\beta}) - D(\hat{\beta})}{\phi} \end{aligned} \quad (2.12)$$

Sollen also zwei hierarchische Modelle hinsichtlich ihrer Güte miteinander verglichen werden, genügt es die Differenz ihrer Devianzen zu betrachten. Um einen signifikanten Unterschied zwischen den Modellen identifizieren zu können, wird die Teststatistik  $lq$  mit dem entsprechenden  $\chi^2$ -Quantil verglichen, wobei sich die Zahl der Freiheitsgrade  $r$  aus der Anzahl der Restriktionen an den Parametervektor  $\beta$  ergibt

$$lq \sim \chi_r^2. \quad (2.13)$$

Alternative Teststatistiken zum Modellvergleich sind die *Score-Statistik* und die *Wald-Statistik*, welche ebenfalls als  $\chi^2$ -verteilt angenommen werden. Die Definitionen dieser alternativen Teststatistiken können [Fahrmeir et al 2007] entnommen werden.

## 2.2. Modell für binäre Daten

Im Fall einer binären Zielgröße  $Y \in \{0; 1\}$  ist die Modellierung der Wahrscheinlichkeit  $\mathbb{P}(y = 1|\mathbf{x}) = \mathbb{E}(y|\mathbf{x}) = \pi$  in Abhängigkeit der Kovariablen  $x_1, \dots, x_p$  von primärem Interesse.

Die Verwendung eines klassischen linearen Regressionsmodells ist dabei aus

## 2. Generalisierte Lineare Regressionsmodelle

zwei Gründen ungeeignet. Zunächst ist die Annahme homoskedastischer Fehlervarianzen im Fall einer binären Zielgröße nicht angebracht. Da die Varianz von  $y_i$

$$\mathbb{V}(y_i) = \pi_i(1 - \pi_i)$$

von  $\pi_i$  und damit von den jeweiligen Kovariablen abhängt, spricht diese Struktur der Fehlervarianzen gegen eine vorliegende Homoskedastizität.

Außerdem ist die Modellierung des Erwartungswertes durch  $\mu_i = \eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$  denkbar ungeeignet, da für den Erwartungswert einer binären Zielgröße  $\mu \in [0, 1]$  gelten muss. Somit sollte die gewählte Responsefunktion  $h$  die Abbildung in den zulässigen Wertebereich  $[0, 1]$  für den Erwartungswert gewährleisten. Hierfür bieten sich Verteilungsfunktionen an, da diese nicht nur auf das Intervall  $[0, 1]$  beschränkt sind, sondern auch monoton steigen.

Die am häufigsten verwendete Linkfunktion, vor allem wegen der guten Interpretierbarkeit, ist der *Logit-Link*:

$$\eta = \log\left(\frac{\mu}{1 - \mu}\right) \quad (2.14)$$

Hierbei findet die Verknüpfung von Erwartungswert und linearem Prädiktor über die logistische Verteilungsfunktion statt.

Weitere mögliche Linkfunktionen sind der *Probit-Link*,  $\eta = \Phi^{-1}(\mu)$ , mit der Verteilungsfunktion der Standardnormalverteilung, sowie der *complementary log-log-Link*,  $\eta = \log(-\log(1 - \mu))$ , unter Benutzung der Verteilungsfunktion der Extremwertverteilung. Letztere Linkfunktion ist besonders bei der Modellierung zeitdiskreter Verweildauern geeignet.

### Parameterschätzung

Wie in Kapitel 2.1 beschrieben werden die Parameterschätzer  $\hat{\beta}$  berechnet, indem mittels dem Fisher-Scoring-Algorithmus iterativ eine Lösung für  $s(\hat{\beta}) = 0$  gesucht wird. Im Fall der Logit-Linkfunktion ergibt sich die vereinfachte Score-Funktion

$$s(\beta) = \sum x_i^l (y_i - \mu_i); \quad \mu_i = \frac{\exp(x_i^l \beta)}{1 + \exp(x_i^l \beta)}. \quad (2.15)$$

Der Fisher-Scoring-Algorithmus konvergiert lediglich dann nicht, wenn  $\hat{\beta} \rightarrow \infty$  gilt. Dies ist der Fall, falls die Daten exakt trennbar sind, d.h. es existiert ein Wert  $\nu$ , für den gilt:  $y_i = 0$ , falls  $x_i \leq \nu$  und  $y_i = 1$ , falls  $x_i > \nu$ .

### Interpretation der Parameter im Logit-Modell

Durch die Wahl der logistischen Verteilungsfunktion als Responsefunktion  $h$  gilt:

$$\pi_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}, \quad \eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}.$$

## 2. Generalisierte Lineare Regressionsmodelle

Äquivalent dazu ist die Annahme, dass die Kovariablen einen exponentiell- multiplikativen Einfluss auf die Chance  $\frac{\pi_i}{1-\pi_i}$

$$\frac{\mathbb{P}(y_i = 1)}{\mathbb{P}(y_i = 0)} = \exp(\beta_0) \cdot \exp(\beta_1 x_{i1}) \cdot \dots \cdot \exp(\beta_p x_{ip})$$

besitzen.

Beispielsweise wird nun die Kovariable  $x_{i1}$  um eine Einheit auf  $x_{i1} + 1$  erhöht. Daraus ergibt sich für die Chance

$$\frac{\mathbb{P}(y_i = 1|x_{i1} + 1, \dots)}{\mathbb{P}(y_i = 0|x_{i1} + 1, \dots)} = \exp(\beta_0) \cdot \exp(\beta_1 (x_{i1} + 1)) \cdot \dots \cdot \exp(\beta_p x_{ip}).$$

Für das Odds-Ratio der Chancen folgt daraus

$$\frac{\mathbb{P}(y_i = 1|x_{i1} + 1, \dots) / \mathbb{P}(y_i = 0|x_{i1} + 1, \dots)}{\mathbb{P}(y_i = 1) / \mathbb{P}(y_i = 0)} = \exp(\beta_1).$$

Erhöht sich die Kovariable  $x_{i1}$  um eine Einheit, so verändert sich die Chance  $y_i = 1$  zu beobachten um den Faktor  $\exp(\beta_1)$ , wenn alle anderen Kovariablen konstant bleiben.

## Überdispersion

Falls mehrere Zeilen der Kovariablen-Datenmatrix identisch sind, d.h. zu mindestens zwei Beobachtungen existieren genau die selben Kovariablenausprägungen, so lassen sich die Daten *gruppieren*. Nach der Gruppierung enthält die Datenmatrix nur noch Zeilen mit unterschiedlichen Ausprägungen für die Kovariablenvektoren  $x_i$ . Zusätzlich wird die Anzahl der Wiederholungen  $n_i$  für jedes  $x_i$  und die relative Häufigkeit  $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_j$  gespeichert. Für gruppierte

Daten lässt sich die Varianz in jeder Gruppe  $i$  durch die *empirische Varianz*  $\bar{y}_i(1 - \bar{y}_i) / n_i$  abschätzen. Ist diese Varianz der Daten größer als die vom binären Regressionsmodell unterstellte Varianz  $\hat{\pi}_i(1 - \hat{\pi}_i) / n_i$  mit  $\hat{\pi}_i = h(x_i^T \beta)$ , so spricht man von *Überdispersion*. Eine Möglichkeit diese erhöhte Varianz im Modell zu berücksichtigen ist der *Quasi-Likelihood-Ansatz*.

Die Wahl eines generalisierten linearen Regressionsmodells impliziert die Annahme einer bestimmten Erwartungswertstruktur,  $\mathbb{E}(y|\mathbf{x}) = \mu = h(\mathbf{x}^T \beta)$ , und einer bestimmten Varianzstruktur,  $\mathbb{V}(y|\mathbf{x}) = \phi b''(\hat{\mu})$ . Im Fall eines Überdispersionsproblems ist die Varianz jedoch durch diesen Ansatz falsch spezifiziert. Die Verwendung eines Quasi-Likelihood-Ansatzes erlaubt die separate Spezifikation von Erwartungswert- und Varianzstruktur. Die Varianz wird durch

$$\mathbb{V}(y_i|x_i) = \phi \frac{\pi_i(1 - \pi_i)}{n_i}; \quad \phi > 1 \tag{2.16}$$

beschrieben. Der multiplikative Überdispersionsparameter  $\phi$  erhöht hierbei die angenommene Varianz, da im gewöhnlichen binären Regressionsmodell  $\phi = 1$

## 2. Generalisierte Lineare Regressionsmodelle

angenommen wird.

Der Überdispersionsparameter kann beispielsweise durch die gemittelte Devianz

$$\hat{\phi} = \frac{1}{n-p} D; \quad D := \text{Devianz} \quad (2.17)$$

geschätzt werden. Die geschätzte Kovarianz muss anschließend mit  $\hat{\phi}$  multipliziert werden:

$$\widehat{\text{Cov}}(\hat{\beta}) = \hat{\phi} F^{-1}(\hat{\beta}) \quad (2.18)$$

Dabei entspricht  $F^{-1}(\hat{\beta})$  der inversen Fisher-Matrix ausgewertet am Parameterschätzer  $\hat{\beta}$ .

Hierbei wird von einem Quasi-Likelihood-Ansatz gesprochen, da durch die Veränderung der Varianzstruktur keine „echte“ Likelihood mehr vorgegeben ist, sondern lediglich eine Quasi-Likelihood  $Q(\beta, \phi)$ . Wichtig ist jedoch, dass die erste Ableitung dieser Quasi-Likelihood die übliche Form einer Score-Funktion besitzt.

Hierbei ist zu Beachten, dass durch die Wahl eines Quasi-Likelihood-Ansatzes die Parameterschätzer identisch bleiben, durch die größere angenommene Varianz verändern sich jedoch die p-Werte. Somit ergeben sich für die Einflussgrößen andere Signifikanzen, was somit die Auswahl anderer Einflussgrößen zur Spezifikation des Modells nach sich ziehen kann.

Ist die durch das Modell unterstellte Varianz **kleiner** als die empirische Varianz der Daten, so spricht man von *Underdispersion*. In diesem Fall wird das Modell analog zur *Überdispersion* durch einen Quasi-Likelihood-Ansatz spezifiziert, der Überdispersionsparameter  $\phi$  wird hier jedoch auf einen Wert kleiner als 1 geschätzt.

### 2.3. Modell für Zähldaten

Zur Modellierung von Zähldaten sind diskrete Verteilungen am besten geeignet, da diese deren spezielle Eigenschaften berücksichtigen. In der Praxis wird in solchen Situationen bevorzugt die Poissonverteilung verwendet. Diese besitzt als Träger die positiven ganzen Zahlen  $0, 1, \dots$ , sowie als einzigen Parameter den Erwartungswert  $\lambda$ , der die Verteilung exakt spezifiziert und positiv definiert sein muss.

#### Log-lineares Poisson-Modell

Da der Erwartungswert  $\lambda_i$  positiv sein muss, ist die Wahl einer einfachen linearen Verknüpfung von linearem Prädiktor und Erwartungswert,  $\lambda_i = \mathbb{E}(y_i | x_i) = \eta_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}$ , nicht zufriedenstellend. Aus negativen Werten von  $\eta_i$  würde ebenfalls ein negativer Erwartungswert  $\lambda_i$  resultieren. Um trotzdem

## 2. Generalisierte Lineare Regressionsmodelle

ausschließlich positive Ausprägungen für  $\lambda_i$  zu garantieren, wären zusätzliche Restriktionen an den Parameterschätzer  $\beta$  nötig. Deshalb werden  $\eta_i$  und  $\lambda_i$  bevorzugt durch die Exponentialfunktion miteinander verknüpft,

$$\lambda_i = \exp(\eta_i); \quad \eta_i = \log(\lambda_i) = x_i^T \beta \quad (2.19)$$

wodurch für alle Ausprägungen von  $\eta_i$  ein positiver Erwartungswert garantiert wird. Des Weiteren werden durch die Wahl dieser Linkfunktion multiplikative Effekte modelliert.

### Überdispersion

Die Annahme einer Poisson-Verteilung impliziert, dass Erwartungswert und Varianz die gleichen Werte besitzen bzw. durch den selben Parameter modelliert werden:

$$\mathbb{E}(y_i|x_i) = \mathbb{V}(y_i|x_i) = \lambda_i. \quad (2.20)$$

Wie bereits für binäre Daten beschrieben, kommt es in Anwendungen jedoch häufig vor, dass diese vom Modell unterstellte Varianz deutlich kleiner ist als die beobachtete empirische Varianz der Daten.

In diesem Fall eignet sich ebenfalls die Modellierung eines *Quasi-Likelihood-Ansatzes*, der die separate Spezifikation von Erwartungswert und Varianz erlaubt. Die folgende Definition der Varianz

$$\mathbb{V}(y_i|x_i) = \phi \lambda_i; \quad \phi > 1 \quad (2.21)$$

wird durch den Einbezug des Überdispersionsparameters  $\phi$  (und die damit verbundene Vergrößerung der Varianz) der vorliegenden Datensituation besser gerecht [Fahrmeir et al 2007].

Wie in Kapitel 2.2 beschrieben, kann auch hier der Überdispersionsparameter mit Hilfe der mittleren Devianz geschätzt werden.

Im Fall der *Unterdispersion* ist ein Dispersionsparameter  $0 < \phi < 1$  zur Spezifikation der Varianz angebracht.

# 3. Generalisierte Additive Modelle

In diesem Kapitel wird auf die Verallgemeinerung der Generalisierte Linearen Modelle zu den Generalisierten Additiven Modellen eingegangen. Hierbei ist besonders die Modellierung nonparametrischer Einflussgrößen von Bedeutung. Dadurch wird eine flexiblere Konstruktion des Einflusses der Kovariablen auf die Zielgröße ermöglicht. Zunächst wird die Funktionsweise der nonparametrischen Regression für normalverteilte Fehlerterme vorgestellt. Anschließend findet die Übertragung dieser Idee auf Generalisierte Modelle Beachtung.

## 3.1. Nonparametrische Regression

Im klassischen Linearen Regressionsmodell wird der Einfluss der Kovariablen auf die Zielgröße als linear angenommen. Eine rein lineare Modellierung ist jedoch in vielen praktischen Anwendungen nicht ausreichend. Die *Nonparametrische Regression* ermöglicht eine flexible Modellierung des Einflusses metrischer Kovariablen auf die abhängige Variable.

### 3.1.1. Univariate Glättung

In diesem Abschnitt wird der flexible Einfluss lediglich einer metrischen Kovariablen  $x$  auf die abhängige Variable  $y$  betrachtet. Man spricht von *Univariater Glättung* bzw. von einem *Streudiagrammglätter*, da sich die zugrundeliegende Datenlage am besten in einem Streudiagramm veranschaulichen lässt. Es soll nun eine möglichst glatte Funktion gefunden werden, die den im Streudiagramm dargestellten Zusammenhang zwischen  $x$  und  $y$  möglichst gut beschreibt [Fahrmeir et al 2007].

Es wird folglich ein Zusammenhang der Art

$$y_i = f(x_i) + \epsilon_i; \quad \mathbb{E}(\epsilon_i) = 0; \quad \mathbb{V}(\epsilon_i) = \sigma^2 \tag{3.1}$$

betrachtet. Die Fehlerterme  $\epsilon$  werden als unabhängig voneinander angenommen und sie weisen eine konstante Varianz  $\sigma^2$  auf.

Daraus folgt für die Zielgröße  $y$

$$\mathbb{E}(y_i) = f(x_i); \quad \mathbb{V}(y_i) = \sigma^2, \tag{3.2}$$

### 3. Generalisierte Additive Modelle

mittels der Funktion  $f$  soll demnach der Erwartungswert der Zielgröße modelliert werden.

Die Glattheitsanforderung ist eine wichtige Bedingung für die Anpassung der Funktion  $f$  an die vorliegenden Daten. Bei der Verwendung von *Polynom-Splines* werden, basierend auf einer vorgegebenen Knotenmenge  $\kappa_1, \dots, \kappa_k$ , stückweise Polynome vom Grad  $l$  angepasst. Damit die Polynomstücke anschließend glatt zusammengesetzt werden können, müssen die Polynome an den Knoten  $(l - 1)$ -mal stetig differenzierbar sein. Diese Anforderung ist essentiell für die Funktionsschätzung. Des Weiteren ist die Anzahl der Knoten der bestimmende Faktor für die Rauheit der Funktion. Werden sehr viele Knoten gewählt, so wird die Funktionsschätzung flexibler und es resultiert ein sehr unruhiger Funktionsverlauf. Werden hingegen sehr wenige Knoten gewählt ist die Funktion zwar glatt, aber möglicherweise wird der vorliegende Zusammenhang zwischen  $x$  und  $y$  nur unzureichend beschrieben. Auf diese Problematik der Glättungsparameterwahl wird in Abschnitt 3.1.4 noch genauer eingegangen. Besonders häufig werden kubische Regressionssplines verwendet. Diese besitzen einerseits einen niedrigen Grad  $l$  und damit eine geringere Anzahl zu schätzender Parameter, andererseits ermöglichen sie bereits eine glatte Modellierung der Funktion [Green/Silverman 1994].

#### 3.1.2. Basisfunktionsansätze

Die im vorherigen Abschnitt beschriebenen Polynom-Splines sind eine Möglichkeit die Funktion  $f$  darzustellen. Es gibt jedoch zahlreiche weitere Ansätze, die eine sinnvolle, flexible Modellierung der Funktion ermöglichen. Einige sollen im Folgenden dargestellt werden.

#### B-Splines

Eine erste Alternative bietet die B-Spline-Basis. Bei der Konstruktion der Funktion  $f(x)$  mit B-Spline-Basisfunktionen findet ebenfalls die Glattheitsanforderung an die Funktion besondere Beachtung. Deshalb werden hier, ähnlich wie bei den Polynom-Splines, Polynomstücke des gewünschten Grads konstruiert, die an den Knoten ausreichend glatt zusammengesetzt werden.

Die Zielfunktion  $f(x)$  wird anschließend als Linearkombination der Basisfunktionen dargestellt:

$$f(x) = \sum_{j=1}^d \gamma_j B_j(x). \quad (3.3)$$

Die Anzahl der benötigten Basisfunktionen wird durch die Anzahl der Knoten  $m$  und den gewünschten Grad  $l$  der Polynomstücke bestimmt,  $d = m + l - 1$ .

### 3. Generalisierte Additive Modelle

Folglich ergibt sich für die Designmatrix  $X$  eine Darstellung der Form

$$X = \begin{pmatrix} B_1(x_1) & \dots & B_d(x_1) \\ \vdots & \ddots & \vdots \\ B_1(x_n) & \dots & B_d(x_n) \end{pmatrix}.$$

Mit Hilfe dieser Darstellung kann die zugrundeliegende Struktur auch durch

$$y_i = \gamma^l x_i + \epsilon_i; \quad x_i = (B_1(x_i), \dots, B_d(x_i))^l \quad (3.4)$$

beschrieben werden. Aus Gleichung 3.4 wird ersichtlich, dass die Parameter  $\gamma_i$  mittels der *Kleinste-Quadrate-Schätzung* bestimmt werden können [Fahrmeir et al 2007], da die Gleichung mit der Darstellung des linearen Prädiktors im klassischen Regressionsmodell übereinstimmt.

Die B-Spline-Basisfunktionen weisen einige Besonderheiten auf. Zunächst sind die Basisfunktionen lokal beschränkt, da sie nur auf einem Bereich von  $l + 2$  benachbarten Knoten definiert sind. Des Weiteren sind die B-Splines positiv definiert und ihr nach oben beschränkter Wertebereich garantiert eine numerische Stabilität [Green/Silverman 1994]. Deshalb ist eine B-Spline-Basis aus numerischer Sicht den Polynom-Splines vorzuziehen. Für äquidistante Knoten besitzen alle Basisfunktionen die gleiche Form, sie sind lediglich auf dem Kontinuum verschoben. Für beispielsweise quantilsbasierte Knoten weisen die Basisfunktionen jedoch eine unterschiedliche Gestalt auf [Hastie/Tibshirani 1990].

### Penalisierte Splines (P-Splines)

Wie bereits erwähnt hängt die Güte der über Basisfunktionsansätze konstruierten Funktion  $f(x)$  stark von der Anzahl und der Position der gewählten Knoten ab. Eine mögliche Strategie dieses Problem zu beheben sind *Penalisierungsansätze*. Hierbei wird zur Schätzung der Funktion eine große Anzahl an Knoten gewählt, somit ist die Funktion grundsätzlich flexibel genug um auch sehr unruhige Funktionsverläufe modellieren zu können. Anschließend wird eine zu große Variabilität bzw. eine zu große „Rauheit“ der Funktionsschätzung mittels einem *Glättungsparameter*  $\lambda$  bestraft.

Soll allgemein die Glattheit einer Funktion bestimmt werden, so eignet sich dafür besonders das Integral über die quadrierte 2.Ableitung der Funktion. Die 2.Ableitung stellt ein Maß für die Krümmung der Funktion dar, folglich ist eine Funktion besonders unruhig, wenn das Integral einen großen Wert aufweist. Somit eignet sich

$$\lambda \int (f''(x))^2 dx \quad (3.5)$$

zur Bestrafung einer großen Variabilität der Funktionsschätzung.

Für die B-Spline-Basisfunktionen ergibt sich eine besonders einfache Darstellung dieses Penalierungsansatzes. In diesem Fall lassen sich beispielweise die 2.Ableitungen einfach durch die 2.Differenzen der Parameter  $\gamma_j$  approximieren [Fahrmeir et al 2007].

### 3. Generalisierte Additive Modelle

#### Natürliche Kubische Splines

In diesem Fall wird die Problemstellung aus einer anderen Sichtweise betrachtet. Gesucht wird eine Funktion  $f : [a; b] \mapsto \mathbb{R}$  mit  $a \leq x_1 < \dots < x_n \leq b$ , für die

$$\sum_{i=1}^n (y_i - f(x_i))^2 + \lambda \underbrace{\int_a^b \{f''(t)\}^2 dt}_{\text{Strafterm}} \quad (3.6)$$

minimal wird. Die Funktion soll also einerseits die Residuenquadratsumme minimieren und dadurch eine gute Datennähe garantieren, andererseits soll das Integral über die quadrierte zweite Ableitung klein sein. Die zweite Ableitung ist hierbei ein Maß für die Glattheit der Funktion. Mittels dem Parameter  $\lambda$  kann gesteuert werden, welches Gewicht die Glattheitsanforderung bei der Funktionsschätzung erhalten soll.

Trotz der sehr allgemeinen Problemformulierung ergibt sich ein erstaunlicherweise sehr konkretes Ergebnis. Der Ausdruck 3.6 wird von einem *natürlichen kubischen Spline* minimiert, mit einem Knoten in jeder Beobachtung  $x_i$ . „Natürlich“ bedeutet in diesem Zusammenhang, dass der kubische Spline außerhalb der Grenzen  $a$  und  $b$  linear fortgesetzt wird. Der natürliche kubische Spline besitzt also die Optimalitätseigenschaft, dass er die kleinste Krümmung aller glatten Kurven aufweist, die die Daten interpolieren [Green/Silverman 1994]. Im Vergleich zu allen anderen glatten interpolierenden Funktionen weist der natürliche kubische Spline den kleinsten Wert für das penalisierte KQ-Kriterium auf.

#### Thin-Plate Regression Splines

Obwohl in diesem Abschnitt lediglich Glättungsverfahren für eine einzige metrische Einflussgröße besprochen werden, soll nun der Vollständigkeit halber auch ein Basisfunktionsansatz zur Konstruktion von mehrdimensionalen Glättungssplines behandelt werden. Die Thin-Plate-Splines eignen sich den Einfluss mehrerer Kovariablen  $x_1, \dots, x_p$  flexibel bzw. glatt zu modellieren. Eine Generalisierung der Glättungssplines in mehrere Dimensionen wird also notwendig.

Im Folgenden wird das Problem auf eine zweidimensionale Glättung reduziert, da diese Problemstellung in der Praxis am häufigsten auftritt, beispielsweise wenn der Einfluss von Koordinaten im Modell berücksichtigt werden soll. Die weitere Verallgemeinerung ins  $p$ -Dimensionale ergibt sich analog [Green/Silverman 1994].

Gegeben sind zweidimensionale Punkte  $z_1, z_2, \dots, z_n$  und zugehörige Werte  $g_1, g_2, \dots, g_n$ . Das Ziel ist es eine ausreichend glatte Funktion  $g$  zu finden, so dass

$$y_i = x_i^T \beta + g(z_i) + \epsilon_i. \quad (3.7)$$

Dabei beschreibt  $x_i^T \beta$  den linearen Einfluss weiterer Kovariablen,  $g(z_i)$  den Einfluss der bivariaten Einflussgröße. Im Gegensatz zum eindimensionalen Fall, in

### 3. Generalisierte Additive Modelle

dem eine glatte Kurve gefittet werden soll, soll nun im Zweidimensionalen eine glatte Oberfläche angepasst werden.

Analog zum eindimensionalen Fall muss nun ebenfalls ein Funktional  $J(g)$  definiert werden, das eine zu unruhige Schätzung der Oberfläche bestraft:

$$J(g) = \int \int \left\{ \left( \frac{\partial^2 g}{\partial x_1^2} \right)^2 + \left( \frac{\partial^2 g}{\partial x_1 \partial x_2} \right)^2 + \left( \frac{\partial^2 g}{\partial x_2^2} \right)^2 \right\} dx_1 dx_2; \quad z = (x_1, x_2). \quad (3.8)$$

$J(g)$  erhöht sich, wenn die 2. Ableitungen von  $g$  bzgl.  $x_1$  und  $x_2$  groß sind. Somit wird tatsächlich eine starke Krümmung der Oberfläche bestraft. Ein weiterer Vorteil des Strafterms ist, dass eine Rotation der Koordinaten im  $\mathbb{R}^2$   $J(g)$  nicht beeinflusst [Green/Silverman 1994]. Des Weiteren gilt  $J(g) = 0$  genau dann, wenn  $g$  linear ist.

#### Definition der Thin-Plate-Splines

Gegeben seien zweidimensionale Beobachtungen  $z_1, \dots, z_n \in \mathbb{R}^2$ , sowie die Funktion  $\eta(r)$  mit

$$\eta(r) = \frac{1}{16\pi} r^2 \log r^2, \quad \text{für } r > 0; \quad \eta(0) = 0. \quad (3.9)$$

Des Weiteren werden folgende Funktionen im  $\mathbb{R}^2$  definiert:

$$\begin{aligned} \phi_1(x_1, x_2) &= 1 \\ \phi_2(x_1, x_2) &= x_1 \\ \phi_3(x_1, x_2) &= x_2. \end{aligned}$$

Jede lineare Funktion des  $\mathbb{R}^2$  kann als Linearkombination der Funktionen  $\phi_j$  dargestellt werden.

Außerdem wird die  $3 \times n$  Matrix  $Z$

$$Z = \begin{pmatrix} 1 & \dots & 1 \\ x_{11} & \dots & x_{1n} \\ x_{21} & \dots & x_{2n} \end{pmatrix} \quad (3.10)$$

definiert. Eine Funktion  $g(z)$  ist ein Thin-Plate-Spline basierend auf den Daten  $z_1, \dots, z_n$ , wenn  $g$  die Form

$$g(z) = \sum_{i=1}^n \delta_i \eta(\|z - z_i\|) + \sum_{j=1}^3 a_j \phi_j(z) \quad (3.11)$$

aufweist. Gilt des Weiteren  $Z\delta = 0$ , so ist  $g$  ein natürlicher Thin-Plate-Spline. In diesem Fall kann der Strafterm in der üblichen Form  $J(g) = \delta^t E \delta$  mit  $E_{ij} = \eta(\|z_i - z_j\|)$  geschrieben werden, die Einträge der Matrix  $E$  stellen folglich eine Approximation des Straffunktionals 3.8 dar [Green/Silverman 1994].

### 3. Generalisierte Additive Modelle

Wie die natürlichen kubischen Splines im Eindimensionalen, besitzen die natürlichen Thin-Plate-Splines im  $p$ -dimensionalen die Eigenschaft, dass sie eine Oberfläche anpassen, die alle Punkte  $z_i$  enthält und die kleinste Krümmung aufweist.

Der Strafterm  $J(g)$  kann zur Glättung verwendet werden. Durch Minimierung der penalisierten Residuenquadratsumme der Oberfläche

$$S(g) = (g_i - g(z_i))' (g_i - g(z_i)) + \lambda J(g),$$

mit  $g_1, \dots, g_n$  als Responsevariablen und dem Glättungsparameter  $\lambda$ , erhält man eine Schätzung für die zweidimensionale Funktion  $g$ .

Eine Alternative univariate Glättungssplines zu generalisieren bieten die *tensor product splines* [Hastie/Tibshirani 1990].

#### 3.1.3. Penalisierte KQ-Schätzung

Wird der flexible Einfluss der Kovariablen  $x$  auf die Zielgröße  $y$  mit einem univariaten Basisfunktionsansatz modelliert, so ergibt sich ein Zusammenhang der Form

$$y = X\gamma + \epsilon. \quad (3.12)$$

Die Parameter  $\gamma$  können also durch Minimierung der Residuenquadratsumme geschätzt werden.

Berücksichtigt man jedoch zusätzlich einen Penalisierungsansatz, der eine große Varianz der resultierenden Funktionsschätzung bestraft, erweitert sich der KQ-Ansatz um einen Strafterm. Dieser Strafterm kann folgendermaßen dargestellt werden:

$$\int (f''(x))^2 dx = \sum_{i=1}^d \sum_{j=1}^d \gamma_i \gamma_j \int B_i''(x) B_j''(x) dx = \gamma' K \gamma. \quad (3.13)$$

Dabei stellt  $K$  die *Strafmatrix* dar, die Restriktionen für die Parameterschätzer  $\gamma$  enthält, so dass eine gewünschte Glattheit der Funktionsschätzung erzielt wird. Mittels dem Glättungsparameter  $\lambda$  kann der Anwender bestimmen, welches Gewicht der Glattheit beigemessen werden soll.

Damit ergibt sich das *Penalisierte KQ-Kriterium*

$$PKQ(\gamma) = (y - X\gamma)' (y - X\gamma) + \lambda \gamma' K \gamma \xrightarrow{\gamma} \min. \quad (3.14)$$

Durch Ableiten und Nullsetzen erhält man den *PKQ-Schätzer*

$$\begin{aligned} \frac{\partial PKQ(\gamma)}{\partial \gamma} &= -X'y + X'X\gamma + \lambda K\gamma \stackrel{!}{=} 0 \\ \gamma (X'X + \lambda K) &= X'y \\ \hat{\gamma} &= (X'X + \lambda K)^{-1} X'y. \end{aligned} \quad (3.15)$$

### 3. Generalisierte Additive Modelle

Aus dieser Darstellung wird besonders deutlich, dass nun der Glättungsparameter  $\lambda$  die Form der Schätzung beeinflusst. Hat die Glattheit der Funktion keine Bedeutung, also  $\lambda = 0$ , so resultiert der gewöhnliche KQ-Schätzer, der zu einer sehr unruhigen Funktionsschätzung führt. Der andere Extremfall  $\lambda \rightarrow \infty$  bewirkt eine sehr glatte Funktionsschätzung, die üblicherweise den Zusammenhang zwischen  $x$  und  $y$  nur ungenügend beschreibt.

#### 3.1.4. Bias-Varianz Trade-off und Glättungsparameterwahl

Bei der Streudiagrammglättung gibt es einen fundamentalen trade-off zwischen Bias und Varianz. Werden beispielsweise bei den Polynom-Splines viele Knoten gesetzt, im Extremfall ist jede Beobachtung ein Knoten, so ergibt sich eine (exakte) Interpolation der Daten und damit stimmen die Schätzungen exakt mit den tatsächlichen Beobachtungen überein. In diesem Fall existiert kein Bias, dafür hat die Schätzung jedoch eine große Varianz. Wird eine geringe Anzahl an Knoten gewählt, so resultiert eine sehr glatte Funktion mit geringer Varianz, jedoch ist in diesem Fall ein Bias vorhanden, da die geschätzten Werte nicht exakt mit den beobachteten  $y$ -Werten übereinstimmen.

Die entscheidende Frage ist somit, wie der Glättungsparameter  $\lambda$  zu wählen ist, so dass sich eine optimale Funktionsschätzung ergibt. „Optimal“ bedeutet in diesem Fall, dass der bestmögliche Kompromiss zwischen Verzerrung und Variabilität des Schätzers gefunden werden muss. Demzufolge wäre eine Methode sinnvoll, die automatisch aus den gegebenen Daten den optimalen Glättungsparameter berechnet.

Das *Kreuzvalidierungs-Kriterium* ist eine Möglichkeit einen optimalen Glättungsparameter auszuwählen. Dabei wird ein Schätzer  $\hat{f}(x)$  als gut eingestuft, wenn die quadratische Abweichung von Fit und wahrem Wert,  $\{y - \hat{f}(x)\}^2$ , für eine neue Beobachtung  $y$  am Punkt  $x$  möglichst klein ist. Da im Regelfall keine solchen neuen Beobachtungen zur Verfügung stehen, werden diese „künstlich“ erzeugt, indem schlichtweg ein Teil der Daten bei der Schätzung nicht berücksichtigt wird. Bei der *Leave-one-out Kreuzvalidierung* wird jeweils eine der Beobachtungen  $(x_i, y_i)$  bei der Schätzung nicht berücksichtigt ( $\hat{f}^{(-i)}(x)$ ) und anschließend für diese Beobachtung, basierend auf der Funktionsschätzung der restlichen  $(n - 1)$  Beobachtungen, die Güte der Anpassung gemessen. Die Güte der Anpassung wird mittels dem quadratischen Abstand von Schätzung und tatsächlicher Beobachtung berechnet. Anschließend wird die *Kreuzvalidierungsquadratsumme* berechnet:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left( y_i - \hat{f}^{(-i)}(x_i; \lambda) \right)^2. \quad (3.16)$$

### 3. Generalisierte Additive Modelle

Der Glättungsparameter für die Schätzung von  $f(x)$  wird so gewählt, dass  $CV(\lambda)$  minimal wird. Bei der Suche nach dem Minimum wird üblicherweise folgendermaßen vorgegangen: Anhand eines vorgegebenes Gitters an Werten für den Glättungsparameter  $\lambda$  wird jeweils  $CV(\lambda)$  berechnet und das  $\lambda$  mit dem niedrigsten zugehörigen Kreuzvalidierungswert wird für die Schätzung ausgewählt.

Problematisch bei diesem Kriterium ist jedoch, dass die Funktion  $n$ -mal geschätzt werden muss, was besonders bei großen Datenmengen sehr aufwendig sein kann. Unter Verwendung der Glättungsmatrix  $S$  kann das Kreuzvalidierungskriterium jedoch auf eine effizientere Form umgeschrieben werden:

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - \hat{f}(x_i)}{1 - s_{ii}} \right)^2. \quad (3.17)$$

Dabei bezeichnet  $\hat{f}(x_i)$  die Schätzung basierend auf allen Beobachtungen, sowie  $s_{ii}$  das  $i$ -te Diagonalelement der Glättungsmatrix.

Eine weitere Vereinfachung bietet die Verwendung einer Modifikation der Kreuzvalidierung, die *Generalisierten Kreuzvalidierung*:

$$GCV = \frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - \hat{f}(x_i)}{1 - sp(S)/n} \right)^2. \quad (3.18)$$

Statt jedes einzelne Diagonalelement von  $S$  zu betrachten, werden diese durch ihren Mittelwert  $sp(S)/n$  ersetzt. Das liefert vor allem einen computationalen Vorteil, da die Spur einer Matrix bestimmt werden kann, ohne jedes einzelne Diagonalelement zu kennen [Green/Silverman 1994].

Die Glättungsmatrix ist mit der *Hat-Matrix* im klassischen linearen Regressionsmodell vergleichbar. Ähnlich wie im linearen Regressionsmodell gilt der folgende lineare Zusammenhang

$$\hat{f} = Sy. \quad (3.19)$$

Mittels der Spur der Glättungsmatrix kann die Anzahl der effektiven Parameter im Modell bestimmt werden. Alle bisher vorgestellten Basisfunktionsansätze sind lineare Glätter, folglich wird ein Zusammenhang wie in (3.19) angenommen.

### 3. Generalisierte Additive Modelle

## 3.2. Additive Modelle

Die Form des linearen Prädiktors wird nun insoweit weiter verallgemeinert, dass der Einfluss mehrerer Kovariablen nonparametrisch modelliert werden kann. Da das Modell additiv in den Effekten der Prädiktoren ist, wird von einem **Additiven Modell** gesprochen. Im Folgenden wird ein Zusammenhang der Form

$$y_i = \sum_{j=1}^p f_j(x_{ij}) + \epsilon_i = f_1(x_{i1}) + \dots + f_p(x_{ip}) + \epsilon_i \quad (3.20)$$

angenommen. Für die Fehlerterme  $\epsilon_i$  gilt weiterhin, dass sie unabhängig von  $x_{ij}$  sind, sowie  $\mathbb{E}(\epsilon_i) = 0$  und  $\mathbb{V}(\epsilon_i) = \sigma^2$ . Die  $f_j$  sind univariate Funktionen, jeweils für einen Prädiktor  $x_j$ . Die Funktionen  $f_j(x_j)$  können mit Hilfe eines der vorgestellten Basisfunktionsansätzen modelliert werden

$$f_j(x_j) = \sum_{l=1}^{d_j} \gamma_{jl} B_l(x_j) = X_j \gamma_j. \quad (3.21)$$

Werden die Funktionen beispielsweise mit B-Spline-Basisfunktionen konstruiert, so kann für jede Funktion ein anderer Grad der Basisfunktionen sowie eine andere Anzahl an zugrundeliegenden Knoten gewählt werden. Jeder dieser Funktionsansätze muss wieder mittels einer Strafmatrix  $K_j$  und einem Glättungsparameter  $\lambda_j$  penalisiert werden, um eine zu raue Funktionsanpassung zu vermeiden [Fahrmeir et al 2007]. Für den Moment wird angenommen, dass die Funktionen  $f_j$  den Einfluss der Prädiktoren flexibel modellieren. Das ist aber keineswegs eine Voraussetzung für ein additives Modell. Diese können sowohl metrische lineare Effekte, sowie kategoriale Variablen enthalten, deren Einfluss wie im klassischen linearen Modell parametrisch gefittet wird. Ein solches Modell bestehend aus nonparametrischen und parametrischen Einflussgrößen wird auch als *semiparametrisch* bezeichnet.

Um die Identifizierbarkeit des Modells zu garantieren, müssen des Weiteren die Funktionen „um 0 zentriert werden“:

$$\sum_{i=1}^n f_1(x_{i1}) = \dots = \sum_{i=1}^n f_p(x_{ip}) = 0. \quad (3.22)$$

Eine wichtige Eigenschaft zur Interpretierbarkeit des Modells erbt das additive Modell vom linearen Modell. Ist das additive Modell angepasst, können die  $p$  Funktionen separat graphisch dargestellt und interpretiert werden, unter der Bedingung, dass die anderen Kovariableneinflüsse konstant bleiben. Diese graphische Darstellung erlaubt ebenfalls eine erste Einschätzung, ob die entsprechende Kovariable tatsächlich nonparametrisch ins Modell aufgenommen werden sollte, oder ob eine einfachere parametrische Modellierung den bestehenden Zusammenhang ausreichend beschreibt [Hastie/Tibshirani 1990].

### 3. Generalisierte Additive Modelle

#### Der Backfitting-Algorithmus

Der *Backfitting-Algorithmus* ist ein genereller, iterativer Algorithmus, der es ermöglicht ein additives Modell anzupassen.

Dabei ist der folgende Zusammenhang der partiellen Residuen ein grundlegender Baustein:

$$\begin{aligned}\mathbb{E}(y|\mathbf{x}) &= \beta_0 + f_1(x_1) + \dots + f_p(x_p) \\ \Rightarrow \mathbb{E}\left(y - \beta_0 - \sum_{i \neq j} f_i|\mathbf{x}\right) &= f_j(x_j).\end{aligned}\tag{3.23}$$

Diese Möglichkeit die  $k$ -te Funktion  $f_k(x_k)$  zu schätzen wird im Backfitting-Algorithmus iterativ angewendet. Des Weiteren wird zur Schätzung die lineare Glättungsmatrix  $S_j$  benötigt, die für jede der  $p$  Funktionen unterschiedlich ist. Die Form der jeweiligen Glättungsmatrix hängt vom gewählten Basisfunktionsansatz zur Konstruktion der Funktion ab und sie besitzt einen individuellen Glättungsparameter  $\lambda_j$ .

#### Der Backfitting-Algorithmus

(i) Initialisierung:  $\beta_0 = \bar{y}$ ,  $f_j = f_j^{(0)}$ ,  $j = 1, \dots, p$

(ii) Zyklus:  $s = 1, \dots, n_{stop}$

Für  $j = 1, \dots, p$

$$\hat{f}_j^{(s)} = S_j \left( y - \beta_0 - \sum_{i < j} \hat{f}_i^{(s)} - \sum_{i > j} \hat{f}_i^{(s-1)} \right)$$

(iii) Wiederhole (ii), bis sich die individuellen Funktionen nicht mehr verändern

Die nötigen Startwerte für den Algorithmus werden in Schritt (i) zugewiesen. Es ist zu beachten, dass im Schritt (ii) zur Bestimmung der Funktionsschätzungen immer die aktuellsten Schätzer benutzt werden. Wird beispielsweise die Funktion  $f_j$  im Iterationsschritt (3) geschätzt, so können für die Funktionen  $f_1, \dots, f_{j-1}$  bereits die Schätzungen aus Schritt (3) benutzt werden, für die Funktionen  $f_{j+1}, \dots, f_p$  müssen jedoch noch die Schätzungen aus Schritt (2) verwendet werden. In jedem Iterationsschritt werden also die Schätzungen für alle  $p$  Funktionen aktualisiert. Als Abbruchkriterium ist beispielsweise

$$\frac{\left\| \sum_j f_j^{(s)} - \sum_j f_j^{(s-1)} \right\|}{\left\| \sum_j f_j^{(s-1)} \right\|} < \epsilon\tag{3.24}$$

geeignet. Ist der relative Fehler zwischen den Funktionsschätzungen aus Schritt (s-1) und Schritt (s) kleiner als eine zuvor definierte Toleranzschranke  $\epsilon$ , so bricht der Algorithmus ab und gibt die Funktionsschätzungen aus Schritt (s)

### 3. Generalisierte Additive Modelle

als endgültige Schätzungen zurück [Hastie/Tibshirani 1990]. Die maximale Anzahl an Iterationsschritten wird mit  $n_{stop}$  bezeichnet. Diese kann im Vorfeld vom Anwender festgelegt werden, oder ergibt sich automatisch, wenn die Toleranzschranke  $\epsilon$  unterschritten wird.

## 3.3. Generalisierte Additive Modelle

Analog zum Übergang von einem Linearen Modell zu einem Additiven Modell verläuft der Übergang des Generalisierten Linearen Modells zum Generalisierten Additiven Modell. Bei dem bisher betrachteten Additiven Modell wurden die Fehlerterme  $\epsilon$  als homogen angenommen. Im Generalisierten Additiven Modell ist die Erweiterung der Verteilung der Fehlerterme auf sämtliche Verteilungen, die sich in Form einer einparametrischen Exponentialfamilie darstellen lassen, erlaubt.

Folglich lässt sich der lineare Prädiktor im Generalisierten Additiven Modell folgendermaßen darstellen:

$$\eta_i = \beta_0 + \sum_j f_j(x_{ij}) + x_i^T \beta. \quad (3.25)$$

Demnach ergibt sich nun eine Verknüpfung von Erwartungswert  $\mathbb{E}(y|\mathbf{x}) = \mu$  und linearem Prädiktor  $\eta$  der Form

$$g(\mu_i) = \beta_0 + \sum_j f_j(x_{ij}) + x_i^T \beta, \quad (3.26)$$

wobei  $g$  die Link-Funktion ist. Dabei umfasst der erste Teil die nonparametrischen Einflussgrößen, im parametrischen Teil  $x_i^T \beta$  sind vor allem kategoriale Variablen enthalten. Hier werden die glatten Funktionen  $f_j$  als univariat angenommen, dies ist aber keinesfalls eine Voraussetzung in Generalisierten Additiven Modellen. In den linearen Prädiktor können ebenfalls mehrdimensionale Funktionen zum Anpassen einer Oberfläche für  $y$  eingebunden werden.

Zur Schätzung der Parameter bzw. der Funktionen zur Repräsentation der glatten Terme wird in Generalisierten Additiven Modellen die *Local-Scoring-Prozedur* verwendet. Der Name rührt daher, dass eine lokale Mittelung benutzt wird, um die im Fall eines Generalisierten Linearen Modells verwendete *Fisher-Scoring-Prozedur* zu verallgemeinern. Es wird also ein Fisher-Scoring-Algorithmus mit einer integrierten Backfitting-Schleife zur Bestimmung der Funktionsschätzungen verwendet.

### 3.3.1. Generalisierte nonparametrische Regression

Bei der bisherigen Betrachtung der nonparametrischen Regression wurde implizit

$$y|x \sim N(f(x), \sigma^2)$$

### 3. Generalisierte Additive Modelle

angenommen, was die Verwendung einer (penalisierten) KQ-Schätzung zur Bestimmung der Koeffizientenschätzer gerechtfertigt hat.

Die Verteilungsannahme von  $y|x$  soll durch Zulassung jeder einfachen Exponentialfamilie flexibler gestaltet werden:

$$\mu_i = h(f(x_i)). \quad (3.27)$$

Die Responsefunktion  $h$  ist weiterhin dafür zuständig, die Ausprägungen des linearen Prädiktors in den für den Erwartungswert  $\mu$  zulässigen Bereich abzubilden. Hier wird lediglich von der Modellierung eines glatten Terms ausgegangen. Die Vorgehensweisen erweitern sich aber analog für die Modellierung mehrerer glatter Einflussgrößen. Zur Konstruktion der Funktion  $f$  sind die aus Abschnitt 3.1 bekannten Basisfunktionsansätze zulässig, die gewährleisten, dass weiterhin ein linearer Prädiktor vorliegt

$$\eta = f(x) = \sum_{j=1}^d \gamma_j B_j(x); \quad (3.28)$$

$$\eta_i = (B_1(x_i), \dots, B_d(x_i)) \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_d \end{pmatrix} = x_i^T \gamma. \quad (3.29)$$

Zur Schätzung des Parametervektors  $\gamma$  ist nun die Maximierung der **penalisierten** log-Likelihood notwendig:

$$l_p(\gamma) = \sum_{i=1}^n l_i(y_i; \eta_i) - \frac{\lambda}{2} \gamma^T K \gamma \xrightarrow{\gamma} \max. \quad (3.30)$$

Die penalisierte log-Likelihood setzt sich aus der gewöhnlichen log-Likelihood der zugrundeliegenden Verteilung und einem Strafterm zusammen. Dabei sind  $l_i$  die einzelnen log-Likelihood-Beiträge der Beobachtungen. Wie bei der bisherigen nonparametrischen Regression bestraft der Penalisierungsterm eine zu große Variabilität der Schätzung. Die Strafmatrix  $K$  enthält Restriktionen an die Parameterschätzer, so dass eine glatte Funktionsschätzung resultiert. Ein möglicher Strafterm ist auch in dieser Situation das Integral über die quadrierte zweite Ableitung der zu schätzenden Funktion  $f$ . Die Penalisierung der log-Likelihood ist notwendig, da sich andererseits als Schätzung der Funktion  $g$  eine exakte Interpolation der Daten ergeben würde, da somit eine Maximierung der gewöhnlichen log-Likelihood resultieren würde.

Der Parametervektor  $\gamma$  wird wie üblich durch die Nullstellensuche der penalisierten Score-Funktion geschätzt. Hierzu wird der Fisher-Scoring-Algorithmus verwendet, wobei die Scorefunktion und die Fisher-Matrix durch die penalisierte Score-Funktion und die penalisierte Fisher-Matrix ersetzt werden. Die penalisierte Score-Funktion ergibt sich durch Ableitung der penalisierten log-

### 3. Generalisierte Additive Modelle

Likelihood nach  $\gamma$ , die penalisierte Fisher-Matrix ist die erwartete zweite Ableitung der penalisierten log-Likelihood:

$$s_p(\gamma) = \frac{\partial l_p(\gamma)}{\partial \gamma} = \sum_{i=1}^n \frac{\partial h(\eta_i)}{\partial \eta} \frac{y_i - \hat{\mu}_i}{\sigma_i^2} - \lambda K \gamma;$$

$$F_p(\gamma) = \mathbb{E} \left( -\frac{\partial^2 l_p(\gamma)}{\partial \gamma \partial \gamma^T} \right) = X^T D(\gamma) \Sigma(\gamma)^{-1} D(\gamma) X + \lambda K$$

mit

$$\Sigma(\gamma) = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{pmatrix}; D(\gamma) = \begin{pmatrix} \frac{\partial h(\eta_1)}{\partial \eta} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial h(\eta_m)}{\partial \eta} \end{pmatrix}.$$

#### 3.3.2. Inferenz

Wie in Kapitel 2 beschrieben, eignet sich für Generalisierte Lineare Modelle die Devianz bzw. die Likelihood-Quotienten-Statistik um zwei hierarchische Modelle miteinander zu vergleichen. Im Fall eines Generalisierten Additiven Modells macht der Modellvergleich mittels der Betrachtung der Differenz der Devianzen immer noch Sinn. Es gibt jedoch keine gesicherte Annahme über die (asymptotische) Verteilung dieser Teststatistik. Obwohl in diesem Fall die Devianz nicht einmal asymptotisch  $\chi^2$ -verteilt ist, haben Simulationsstudien gezeigt, dass die  $\chi^2$ -Verteilung immer noch eine sinnvolle Approximation darstellt [Hastie/Tibshirani 1990]. Die für die Testentscheidung benötigten Freiheitsgrade ergeben sich wie üblich als Differenz der Freiheitsgrade der beiden Modelle. Mit Hilfe des Devianzvergleichs kann für die gefitteten Modelle einerseits überprüft werden, ob die betrachteten Kovariablen einen signifikanten Einfluss besitzen, oder ob sie besser aus dem Modell entfernt werden sollten. Andererseits kann aber zusätzlich überprüft werden, ob eine parametrische Aufnahme der bisher nonparametrisch angepassten Variablen den Zusammenhang ebenfalls zufriedenstellend beschreibt. Das heißt im Gegensatz zum Generalisierten Linearen Modell muss nicht nur überprüft werden ob die Kovariablen überhaupt einen Einfluss haben, sondern für die glatten Terme muss beurteilt werden, wie glatt sie tatsächlich sein sollten. Folglich wird für einen glatten Term zunächst überprüft, ob ein linearer Term den Zusammenhang auch ausreichend beschreibt und anschließend ob die Kovariable nicht ganz aus dem Modell entfernt werden kann. Dabei besitzen lineare Terme 1 Freiheitsgrad, entfernte Variablen natürlicherweise 0 Freiheitsgrade. Für die glatten Terme gilt, je höher die Anzahl der Freiheitsgrade, desto unruhiger ist der Funktionsverlauf. Weitere Möglichkeiten des Modellvergleichs liefert der Vergleich der GCV- bzw. AIC-Werte der Modelle. Die entsprechenden näheren Definitionen werden in Kapitel 3.3.3 gegeben. Für beide Kriterien gilt, je niedriger der GCV- bzw. AIC-Wert, desto besser ist das Modell. Speziell der AIC berücksichtigt hierbei nicht nur die

### 3. Generalisierte Additive Modelle

Anpassung des Modells an die Daten sondern auch die Modellkomplexität. Bei der *Rückwärtsselektion* wird zunächst ein Modell mit allen vorhandenen Einflussgrößen angepasst, wobei jede metrische Variable zunächst nonparametrisch ins Modell aufgenommen wird. Anschließend wird zuerst für die glatten Terme überprüft, ob eine parametrische Aufnahme ins Modell nicht ausreichend wäre. Im letzten Schritt wird für alle Kovariablen überprüft, ob sie überhaupt einen Erklärungswert für die Zielgröße liefern, oder aus dem Modell entfernt werden sollten.

#### 3.3.3. Glättungsparameterwahl

Im Folgenden wird davon ausgegangen, dass zur Konstruktion des Einflusses der  $p$  nonparametrischen Einflussgrößen  $f_1, \dots, f_p$  im Modell die Glättungsmatrizen  $S_1, \dots, S_p$ , sowie die zugehörigen Glättungsparameter  $\lambda_1, \dots, \lambda_p$  benutzt werden.

In Kapitel 3.1 wurde zur Glättungsparameterwahl das Kreuzvalidierungskriterium vorgestellt. Hierbei wurde die Güte der Anpassung mittels der Abweichung der auf  $(n - 1)$  Beobachtungen basierenden Funktionsschätzung  $\hat{f}^{-i}(x_i)$  vom tatsächlichen Wert  $y_i$  gemessen. Im Fall eines Generalisierten Additiven Modells wird diese Abweichung mittels der Devianz  $D(y; \mu) = \sum_{i=1}^n D(y_i; \mu_i)$  bestimmt. Für die Leave-one-out Kreuzvalidierung wird analog die Schätzung  $\hat{\mu}_i^{-i}$  betrachtet, die auf allen Beobachtungen außer der  $i$ -ten beruht. Für das Kreuzvalidierungskriterium ergibt sich somit

$$CV = \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i^{-i}). \quad (3.31)$$

Das Kreuzvalidierungskriterium muss nun für  $\lambda_1, \dots, \lambda_p$  minimiert werden, was computational aufwendig ist, da für jeden zu testenden Wert von  $\lambda_1, \dots, \lambda_p$   $n$  Anwendungen der Local-Scoring-Prozedur durchgeführt werden müssen. Analog zum Fall der univariaten Glättung existieren auch hier Approximationen des Kreuzvalidierungskriteriums, die eine computational günstigere Glättungsparameterwahl ermöglichen.

Die *Generalisierte Kreuzvalidierung* wird durch

$$GCV = \frac{\frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i)}{\{1 - sp(R)/n\}^2} \quad (3.32)$$

definiert. Dabei enthält die Gesamtglättungsmatrix  $R$  die einzelnen Glättungsmatrizen  $S_j$  für die Funktionsschätzungen  $f_j$ . Dabei ist  $sp(R)$  wieder eine Abschätzung der Anzahl der effektiven Parameter im Modell.

Eine Alternative bildet *Akaike's Informationskriterium AIC*:

$$AIC = \frac{1}{n} \sum_{i=1}^n D(y_i; \hat{\mu}_i) + 2sp(R) \phi/n. \quad (3.33)$$

### 3. Generalisierte Additive Modelle

Sowohl GCV als auch AIC benötigen lediglich eine einzige Anwendung des Local-Scoring-Algorithmus für jeden Wert von  $\lambda = (\lambda_1, \dots, \lambda_p)$  und bieten somit eine praktikable Möglichkeit zur Glättungsparameterwahl [Hastie/Tibshirani 1990].

## 3.4. Generalisierte Additive Modelle in R

Generalisierte Additive Modelle können im Statistikprogramm R im add-on-package **mgcv** mittels der Funktion **gam()** angepasst werden. Dabei können bei der Modellformulierung vom Anwender sowohl glatte als auch parametrische Einflussgrößen bestimmt werden.

Glatte Terme werden durch **s()** als solche gekennzeichnet. Dabei wird die Dimension des anzupassenden Regressionssplines automatisch durch die Anzahl der übergebenen Variablen bestimmt. Des Weiteren werden die nonparametrischen Einflussgrößen mittels penalisierten Regressionssplines bestimmt, was die Wahl eines Glättungsparameters notwendig macht. Hierfür steht einerseits das GCV-Kriterium (bei unbekanntem Scale-Parameter  $\phi$ , z.B. Quasi-Likelihood-Modellen) und der UBRE/AIC-Score (bei bekanntem Scale-Parameter  $\phi$ , z.B. Binomialmodell mit  $\phi = 1$ ) zur Verfügung. Andererseits kann die Glättung auch durch eine fest vorgegebene Anzahl an Freiheitsgraden gesteuert werden. Bei ersterem wird also das Maß der Glättung als Teil der Anpassung automatisch durch den Algorithmus bestimmt.

Mittels der **bs()**-Option in der **s()**-Funktion kann eine spezielle *penalisierte Glättungsbasis* zur Anpassung ausgewählt werden:

1. **bs="tp"**: Thin-Plate-Regression-Splines  
default, da optimale Glätter für jegliche Dimension;  
computational zwar langsamer, aber bestes MSE-Ergebnis;  
automatische Wahl der Anzahl effektiver Freiheitsgrade
2. **bs="cr"**: penalisierte kubische Splines  
Benutzung der quadrierten 2. Ableitung zur Modellierung des Strafterms
3. **bs="ps"**: P-Splines mit B-Spline-Basis

Zur Festlegung der Verteilung der Zielgröße wird die **family()**-Option benutzt. Hier kann der Anwender zusätzlich die gewünschte Linkfunktion bestimmen. Als Verteilungsannahmen stehen auch Quasi-Likelihood-Ansätze (z.B. `family=quasibinomial(link="logit")`) zur Verfügung.

Da für **gam()** keine **step()**-Funktion zur Vorwärts- bzw. Rückwärtsselektion vorhanden ist, muss die Modellselektion über Vergleiche der jeweiligen GCV- bzw. UBRE/AIC-Werte erfolgen.

Der Modellfit in **gam()** wird durch Maximierung der penalisierten Likelihood bestimmt. Im Gegensatz zur GLM-Anpassung, bei der eine einfache Iteratively Reweighted Least Square (IRLS)-Schleife benutzt wird, wird hier das Problem durch eine penalisierte IRLS-Schleife gelöst [R-Hilfeseite zu **gam()**].

## 4. Boosting von Additiven Modellen

In diesem Kapitel wird das Boosting als weitere Möglichkeit Generalisierte Additive Modelle zu schätzen vorgestellt. Komponentenweises Boosting ist vor allem für komplexe parametrische bzw. nonparametrische Modelle, beispielsweise Generalisierte Lineare oder Generalisierte Additive Modelle mit vielen Einflussgrößen, geeignet. Ein besonderer Vorteil liegt in der integrierten Variablenselektion, sowie Regularisierungsansätzen, die eine glatte Funktionsschätzung im Fall von nonparametrischen Einflussgrößen ermöglichen. Außerdem kann Boosting auch in Situationen mit  $p > n$  angewendet werden. Andere Schätzverfahren hingegen (beispielsweise die schrittweise lineare Regression) können nicht angewendet werden, wenn die Anzahl der Kovariablen größer als die Anzahl der Beobachtungen ist.

Boosting kann ebenfalls als ein *Functional-Gradient-Descent-Algorithmus* in Funktionenräumen interpretiert werden, wie er im folgenden Abschnitt erläutert wird. Das ermöglicht die direkte Interpretation von Boosting als Methode zur Funktionsschätzung.

Das Boosting-Verfahren ist für mehrere Verteilungen der Zielgrößen geeignet. Im Fall einer binären Zielgröße spricht man von *Binomial Boosting*, bei Zähldaten von *Poisson Boosting* und bei stetigen Zielgrößen von  *$L_2$ -Boosting*. Die Verfahren werden in den entsprechenden Kapiteln näher beschrieben.

### 4.1. Der FGD-Algorithmus

Seien  $(x_1, y_1), \dots, (x_n, y_n)$  Realisationen unabhängig und identisch verteilter Zufallsvariablen, wobei  $x_1, \dots, x_n$   $p$ -dimensionale Einflussgrößen und  $y_1, \dots, y_n$  eindimensionale Responsevariablen sind. Ziel des *Functional-Gradient-Descent-Algorithmus* ist es die reellwertige Funktion  $f^*(\cdot)$  zu minimieren:

$$f^*(\cdot) = \underset{f(\cdot)}{\operatorname{argmin}} \mathbb{E}[\rho(y, f(x))]. \quad (4.1)$$

Dabei bezeichnet  $\rho(\cdot, \cdot)$  die Verlustfunktion, die bezüglich  $f(x)$  differenzierbar sein muss. Die Schätzung der Funktion  $f^*(\cdot)$  erfolgt mittels Minimierung des *empirischen Risikos*  $\sum_{i=1}^n \rho(y_i, f(x_i))$  bezüglich  $f$ . Der Algorithmus folgt dabei iterativ dem steilsten Abstieg in Funktionenräumen.

#### 4. Boosting von Additiven Modellen

##### Allgemeiner FGD oder Boosting-Algorithmus

1. Initialisiere  $\hat{f}^{(0)}(\cdot)$  mit einem offset-Wert, z.B.  $\hat{f}^{(0)}(\cdot) \equiv 0$ ;  
setze  $m = 0$ ;  
spezifiziere eine *base-procedure*
2. Erhöhe  $m$  um 1;  
Berechne den negativen Gradienten  $-\frac{\partial}{\partial f}\rho(y, f)$  und werte diesen an der Stelle  $\hat{f}^{(m-1)}(x_i)$ ,  $i = 1, \dots, n$  aus.

Dadurch erhält man den negativen Gradienten Vektor

$$U^{(m-1)} = \left( U_i^{(m-1)} \right)_{i=1, \dots, n} := \left( -\frac{\partial}{\partial f}\rho(y, f) \Big|_{y=y_i, f=\hat{f}^{(m-1)}(x_i)} \right)_{i=1, \dots, n}$$

3. Passe den negativen Gradienten  $U^{(m-1)}$  mittels der *base-procedure* an jede der  $p$  Komponenten von  $X$  an. Dadurch erhält man  $p$  Vektoren, wobei jeder Vektor eine Schätzung für den negativen Gradienten-Vektor  $U^{(m-1)}$  ist:

$$\left( x_{ij}, U_i^{(m-1)} \right)_{i=1}^n \xrightarrow{\text{base-procedure}} \hat{g}_j^m(\cdot); j = 1, \dots, p$$

4. Selektiere die Komponente von  $X$ , die  $U^{(m-1)}$ , gemäß einem im Vorfeld definierten Gütekriterium, am besten anpasst. Setze  $U^{(m-1)} = \hat{g}_k^m(\cdot)$ , wobei die Kovariable  $k$  die beste Kovariable ist.
5. Update:  $\hat{f}^{(m)} = \hat{f}^{(m-1)} + \nu U^{(m-1)}$ ;  $\nu$  ist eine Schrittweite mit  $0 < \nu \leq 1$
6. Iteriere Schritte 2-5 bis  $m = m_{\text{stop}}$

Die in Schritt (1) zu wählende *base-procedure* ist essentiell für den Algorithmus. Welche *base procedure* für welche Situation adäquat ist, wird in Kapitel 4.2 näher erläutert. Der Algorithmus wird als *Functional-Gradient-Descent-Algorithmus* bezeichnet, da in jedem Iterationsschritt mit Hilfe der sogenannten *base-procedure* der negative Gradient der Verlustfunktion angepasst wird und anschließend diese Schätzung des negativen Gradienten zur Aktualisierung der Funktionsschätzung (Schritt 5) verwendet wird. Diese Vorgehensweise resultiert automatisch in einer additiven Modellanpassung [Schmid/Hothorn 2007]. Aus Schritt 4 des Algorithmus wird außerdem ersichtlich, dass eine Variablen-selektion stattfindet. Da hier lediglich die Kovariable mit der besten Anpassung ausgewählt wird, wird die Funktion in Schritt 5 auch nur für diese Kovariable aktualisiert. Diese Eigenschaft macht das Boosting vor allem in hochdimensionalen Modellen mit vielen Einflussgrößen sinnvoll, da man in diesen Fällen lediglich an den Kovariablen interessiert ist, die tatsächlich die Zielgröße beeinflussen.

#### 4. Boosting von Additiven Modellen

Außerdem ermöglicht diese Eigenschaft die Anwendung des FGD-Algorithmus für  $p > n$ .

Vor allem der Parameter  $m_{\text{stop}}$  muss adäquat gewählt werden, da nur so ein Overfitting der Daten vermieden werden kann [Schmid/Hothorn 2007].

### 4.2. Wahl der base-procedure

Jeder Boosting-Algorithmus benötigt die Spezifikation einer sogenannten base-procedure (vgl. FGD-Algorithmus Schritt 1). Diese base-procedure wird im Algorithmus jeweils dazu verwendet, den Gradienten der Verlustfunktion möglichst gut anzupassen. Dadurch erhält man eine Schätzung des negativen Gradientenvektors, der zur Aktualisierung der Funktionsschätzung benötigt wird.

Den allgemeinen Boosting-Schätzer

$$\hat{f}^{(m)}(\cdot) = \nu \sum_{k=1}^m \hat{g}^{(k)}(\cdot) \quad (4.2)$$

erhält man als Summe der base-procedure-Schätzungen. Die Schätzung  $\hat{g}(\cdot)$  ist demnach eine Schätzung mittels der base-procedure, die auf den Daten  $(x_1, U_1), \dots, (x_n, U_n)$  basiert. Damit bestimmen strukturelle Eigenschaften der base-procedure die resultierende Funktionsschätzung.

### Komponentenweise KQ-Schätzung für Lineare Einflussgrößen

Für parametrische Einflussgrößen ist die *komponentenweise Kleinste-Quadrate-Schätzung* als base-procedure geeignet [Bühlmann/Hothorn 2008]. Mittels der gewöhnlichen KQ-Schätzung wird die Kovariable gesucht, die im Sinne des KQ-Kriteriums am besten geeignet ist. Folglich wird die Kovariable gesucht, die die entsprechende Residuenquadratsumme

$$\hat{S} = \underset{1 \leq j \leq p}{\operatorname{argmin}} \sum_{i=1}^n \left( U_i - \hat{\beta}^{(j)} x_i^{(j)} \right)^2 \quad (4.3)$$

minimiert. Die Parameterschätzungen ergeben sich gemäß

$$\begin{aligned} \frac{\partial \hat{S}}{\partial \beta^{(j)}} &= -2 \sum_{i=1}^n \left( U_i - \beta^{(j)} x_i^{(j)} \right) x_i^{(j)} \stackrel{!}{=} 0 \\ \hat{\beta}^{(j)} &= \frac{\sum_{i=1}^n x_i^{(j)} U_i}{\sum_{i=1}^n \left( x_i^{(j)} \right)^2} \end{aligned}$$

#### 4. Boosting von Additiven Modellen

Sei nun  $\hat{\beta}^{(\hat{S})}$  der Schätzer der „besten“ Kovariable, mit  $\hat{S} \in \{1, \dots, p\}$ , also der Kovariablen, die die Residuenquadratsumme am meisten vermindert. Damit ergibt sich für die Schätzung durch die base-procedure

$$\hat{g}(x) = \hat{\beta}^{(\hat{S})} x^{(\hat{S})}. \quad (4.4)$$

Durch das Kriterium 4.3 wird folglich immer genau eine Kovariable selektiert und anschließend wird die Funktionsschätzung im 5. Schritt des Algorithmus auch nur für diese Kovariable aktualisiert. In jedem Iterationsschritt wird also eine Variablenselektion vorgenommen, es muss aber keineswegs in jedem Iterationsschritt eine andere Variable ausgewählt werden. Neben der Variablenselektion ist eine weitere Eigenschaft, dass die Schätzer  $\hat{\beta}^{(m)}$  deutlich kleiner als die KQ-Schätzer sind. Das ist von Vorteil, da bei der KQ-Schätzung die Parameter tendentiell zu groß geschätzt werden. Diese base-procedure kann sowohl beim  $L_2$ -Boosting als auch beim BinomialBoosting angewendet werden.

### Komponentenweise Glättungssplines für nonparametrische Einflussgrößen

Die Flexibilität, die das additive Modell erhält, indem man metrische Einflussgrößen nonparametrisch berücksichtigt, kann auch beim Boosting erreicht werden. In diesem Fall muss eine nonparametrische base-procedure zur Funktionsschätzung verwendet werden. Dabei geht man wieder von einer Minimierung der Residuenquadratsumme aus, berücksichtigt aber zusätzlich einen Strafterm, der eine ausreichende Glattheit der resultierenden Funktionsschätzung garantiert:

$$\hat{f}^{(j)}(\cdot) = \sum_{i=1}^n \left( U_i - f \left( x_i^{(j)} \right) \right)^2 + \lambda \int (f''(x))^2 dx. \quad (4.5)$$

Für die geschätzte base-procedure ergibt sich

$$\hat{g}(x) = \hat{f}^{\hat{S}} \left( x^{(\hat{S})} \right), \quad (4.6)$$

wobei  $\hat{f}^{(\hat{S})}(\cdot)$  die Funktionsschätzung bezeichnet, die das Kriterium 4.5 minimiert. Hierbei wird der Glättungsparameter  $\lambda$  so gewählt, dass die Spur der Hat-Matrix der Anzahl der vorgegebenen Freiheitsgrade entspricht. Die Freiheitsgrade sollten klein gewählt werden, beispielsweise  $df = 4$  [Bühlmann/Hothorn 2008]. Durch die Verwendung von Glättungssplines als base-procedures können Additive Modelle geschätzt werden. Dabei konnte gezeigt werden, dass Boosting nicht nur eine Alternative zu Standardschätzverfahren wie beispielsweise dem Backfitting-Algorithmus ist, sondern diese sogar übertreffen kann [Schmid/Hothorn 2007].

#### 4. Boosting von Additiven Modellen

Ein weiterer Ansatz ersetzt die Glättungssplines der base-procedure durch P-Splines, was zu ähnlichen Vorhersagegenauigkeiten führt, aber deutlich weniger Rechenzeit in Anspruch nimmt. Dabei bezeichnen P-Splines penalisierte Regressionssplines basierend auf einer B-Spline-Basis (vgl. Kapitel 3.1.1). Die Formulierung des Strafterms durch das Integral über die quadrierte zweite Ableitung bei gewöhnlichen Glättungssplines nimmt besonders viel Rechenzeit in Anspruch, da es aufwendig ist diesen Term auszuwerten. P-Splines hingegen verwenden  $d$ -te Differenzen benachbarter Koeffizienten als diskrete Approximation der  $d$ -ten Ableitungen, welche wesentlich einfacher zu berechnen sind (vgl. Kapitel 3.1.1). Werden Differenzen 2. Ordnung verwendet, so ergibt sich eine diskrete Approximation des Strafterms der Glättungssplines. Die Dimensionalität des Strafterms wird jedoch deutlich reduziert, was zu einer größeren Effektivität des Boosting-Algorithmus führt. Die Simulationsstudien, die in [Schmid/Hothorn 2007] vorgestellt werden zeigen, dass die P-Spline base-procedure hinsichtlich des Mean-Squared-Errors eine gute Approximation der Glättungsspline base-procedure darstellen. Des Weiteren resultiert eine Wahl von wenigen Freiheitsgraden in kleineren Vorhersagefehlern und die Anzahl an äquidistanten Knoten für den P-Spline scheint keinen großen Einfluss zu haben (vgl. Kapitel 4.6).

### 4.3. Boosting für binäre Zielgrößen

Bei binären Zielgrößen entspricht die Anwendung eines Boosting-Algorithmus einem Klassifikationsverfahren. Durch die Definition verschiedener Verlustfunktionen  $\rho(\cdot, \cdot)$  werden unterschiedliche Boosting-Algorithmen durchgeführt. Hier soll für den Fall einer binären Zielgröße  $y \in \{-1, 1\}$  die negative log-Likelihood einer Binomialverteilung als Verlustfunktion angenommen werden:

$$\rho(y, p) = -(y \log(p) + (1 - y) \log(1 - p)). \quad (4.7)$$

Außerdem wird für den Parameter  $p$

$$p = \frac{\exp(f)}{\exp(f) + \exp(-f)} \quad (4.8)$$

angenommen. Daraus ergibt sich für den *population-minimizer*

$$f^*(x) = \frac{1}{2} \log \left( \frac{p(x)}{1 - p(x)} \right); \quad p(x) = \mathbb{P}[y = 1|x] \quad (4.9)$$

die Hälfte der logarithmierten Chance.

Die Anwendung des FGD-Algorithmus mit der Verlustfunktion 4.7 wird als BinomialBoost bezeichnet. Werden als base-procedure *komponentenweise lineare Kleinste-Quadrate* verwendet, so resultiert ein Logistisches Lineares Modell.

#### 4. Boosting von Additiven Modellen

Werden Glättungssplines als base-procedure verwendet, so ergibt sich ein Logistisches Additives Modell.

Als Informationskriterium kann beispielsweise Akaike's Informationskriterium

$$AIC(m) = -2 \sum_{i=1}^n [y_i \log(\hat{p}^{(m)}(x_i)) + (1 - y_i) \log(1 - \hat{p}^{(m)}(x_i))] + 2df(m) \quad (4.10)$$

verwendet werden, wobei  $df(m)$  die Anzahl der Freiheitsgrade in Abhängigkeit von der  $m$ -ten Boostingiteration darstellt. Die Anzahl dieser Freiheitsgrade kann über die Spur der Hat-Matrix der  $m$ -ten Boostingiteration ermittelt werden. Eine optimale Anzahl an Iterationen  $m_{\text{stop}}$  kann beispielsweise durch Minimierung von  $AIC(m)$  ermittelt werden.

Eine weitere mögliche Verlustfunktion ist die exponentielle Verlustfunktion [Bühlmann/Hothorn 2008].

### 4.4. Boosting für Zähldaten

Ein Boosting-Algorithmus für Zähldaten bzw. eine poisson-verteilte Zielgröße  $y \in \{0, 1, 2, \dots\}$  ist durch das *PoissonBoosting* gegeben. Dabei wird die negative log-Likelihood der Poisson-Verteilung als Verlustfunktion definiert:

$$\rho(y, f) = -yf + \exp(f); \quad f = \log(\lambda). \quad (4.11)$$

Die Schätzung der Funktion erfolgt auch hier mittels dem FGD-Algorithmus mit der in 4.11 definierten Verlustfunktion. Je nach Wahl der base-procedure ergibt sich ein Lineares oder ein Additives Modell.

Zur Wahl der optimalen Anzahl an Boosting-Iterationen kann neben der Kreuzvalidierung analog zum BinomialBoosting Akaike's Informationskriterium verwendet werden:

$$AIC(m) = -2 \sum_{i=1}^n [y_i \log(\hat{\lambda}^{(m)}(x_i)) + \hat{\lambda}^{(m)}(x_i)] + 2df(m). \quad (4.12)$$

Dabei kann die Anzahl an Freiheitsgraden mittels der Spur der Hat-Matrix im  $m$ -ten Iterationsschritt ermittelt werden. Optimales  $m$  ergibt sich durch Minimierung von  $AIC(m)$ .

### 4.5. Boosting für stetige Zielgrößen

Für die Regression mit einer stetigen Zielgröße  $y \in \mathbb{R}$  wird meistens der quadratische Fehler als Verlustfunktion verwendet, was den bekannten population-minimizer

$$\rho_{L_2}(y, f) = \frac{1}{2} |y - f|^2 \quad (4.13)$$

$$f_{L_2}^*(x) = \mathbb{E}[y|x] \quad (4.14)$$

#### 4. Boosting von Additiven Modellen

ergibt. Der zugehörige Boosting-Algorithmus wird als  $L_2$ -Boosting bezeichnet. Der FGD-Algorithmus verändert sich durch den quadratischen Fehler als Verlustfunktion insoweit, dass der negative Gradient in jedem Iterationsschritt jeweils mit den Residuen,  $U_i = y_i - \hat{f}^{(m-1)}(x_i)$ , übereinstimmt. Als Startwert wird  $\hat{f}^{(0)} \equiv \bar{y}$  gewählt.  $L_2$ -Boosting betreibt also ein iteratives Fitting der gewöhnlichen Residuen.

### 4.6. Tuning-Parameter

#### $m_{\text{stop}}$

Die maximale Anzahl an Iterationsschritten  $m_{\text{stop}}$  ist der hauptsächliche Tuning-Parameter und kann beispielsweise durch Kreuzvalidierung oder die Minimierung von Akaike's Informationskriterium bestimmt werden. Wird  $m_{\text{stop}}$  zu groß gewählt, resultiert ein Overfitting der Daten. Deshalb ist es notwendig die maximale Anzahl an Iterationen so zu wählen, dass ein solches Overfitting vermieden wird.

#### Schrittweite $\nu$

Der Parameter  $\nu$  kann als Schrittweite des Gradientenabstiegs, sowie als Shrinkage-Parameter für die Koeffizientenschätzungen interpretiert werden. Die Wahl der Schrittweite  $\nu$  ist hinsichtlich der Vorhersage des FGD-Algorithmus weniger bedeutsam, sie muss lediglich „klein“ sein. In [Bühlmann/Hothorn 2008] wird beispielsweise eine Schrittweite  $\nu = 0,1$  oder  $\nu = 0,01$  vorgeschlagen. Die Wahl eines zu kleinen Wertes für  $\nu$  resultiert lediglich in einer größeren notwendigen Anzahl an Iterationsschritten, trägt aber nicht sonderlich zur Verbesserung des Ergebnisses bei [Schmid/Hothorn 2007].

#### Wahl der Freiheitsgrade

Die Problematik, welche Anzahl an Freiheitsgraden gewählt werden soll, stellt sich natürlicherweise nur, wenn man Glättungs- bzw. P-Splines als base-procedure benutzt. Damit der FGD-Algorithmus eine gute Annäherung an die wahre Funktion  $f^*$  darstellt, sollte die Glättungsspline-base-procedure einen großen Bias und eine kleine Varianz aufweisen [Schmid/Hothorn 2007]. Diese Eigenschaft der base-procedure kann erreicht werden, indem eine geringe Anzahl an Freiheitsgraden gewählt wird, da so eine besonders glatte Funktion angenommen wird. Der damit verbundene „große“ Bias kann durch eine Erhöhung der Anzahl an Iterationen  $m_{\text{stop}}$  anschließend verkleinert werden. In [Schmid/Hothorn 2007] wird eine konstante Wahl von  $df = 4$  in jedem Iterationsschritt vorgeschlagen.

## 4. Boosting von Additiven Modellen

### Knotenwahl

Die Wahl der Anzahl  $p$  an äquidistanten Knoten ist lediglich bei der Verwendung von P-Splines als base-procedure von Bedeutung. Bei der Wahl von Glättungssplines als base-procedure entspricht jede Beobachtung einem Knoten was dazu führt, dass die Knoten nicht äquidistant sind und nicht vom Benutzer definiert werden müssen. In der Theorie über P-Splines wird betont, dass die Anzahl und die Positionierung der Knoten keinen großen Einfluss auf die Qualität der resultierenden Funktionsschätzung besitzt [Fahrmeir et al 2007]. Es wird lediglich empfohlen eine große Anzahl an Knoten zu wählen, so dass die Funktion generell flexibel genug ist, die gewünschte Glattheit wird anschließend mit dem Strafterm reguliert (vgl. Kapitel 3.1). In [Schmid/Hothorn 2007] wird die Wahl von 20–50 äquidistanten Knoten vorgeschlagen.

### 4.7. Boosting in R

Im Statistikprogramm **R** steht der Boosting-Algorithmus im add-on Paket **mboost** mittels der Funktion **gamboost ()** zur Verfügung. Verwendet wird das Prinzip des FGD-Algorithmus, also Gradienten-Boosting.

Bei der Modellformulierung stehen mehrere base-procedures zur Auswahl, die hier kurz vorgestellt werden:

1. **bss**: Glättungsspline base-procedure zur Modellierung einer glatten Funktionsschätzung; hier kann vom Benutzer zusätzlich die gewünschte Zahl an Freiheitsgraden definiert werden;  
**default:**  $df = 4$
2. **bbs**: P-Splines mit zugrundeliegender B-Spline-Basis; der Benutzer kann die gewünschte Anzahl an Freiheitsgraden, Anzahl an äquidistanten Knoten, sowie den Grad der Differenzen für den Strafterm definieren;  
**default:**  $df = 4$ ,  $knots = 20$ , kubische B-Splines mit 2. Differenzen als Strafterm
3. **bns**: P-Splines mit natürlichen Splines als Basis  
**default:**  $df = 4$ ,  $knots = 20$ ,  $differences = 2$
4. **bo1s**: für metrische lineare Kovariablen und Faktoren; Verwendung komponentenweiser KQ-Schätzung
5. **bspatial**: bivariate penalisierte Tensor-Produkt-Splines; zur Modellierung zweidimensionaler Einflussgrößen geeignet;  
**default:**  $df = 5$ , jeweils 20 Knoten für die  $x$ - und  $y$ -Komponente, kubische Splines mit 2. Differenzen für den Strafterm

Des weiteren wählt man mit dem **family ()**-Objekt die Verteilungsannahme für die Zielgröße aus. Per **default** wird von einer stetigen, normalverteilten

#### 4. Boosting von Additiven Modellen

Zielgröße ausgegangen, man kann jedoch durch die Auswahl der Binomial()- , bzw. Poisson()-Familie Modelle für binäre Zielgrößen oder Zähldaten schätzen. Hierbei ist dem Benutzer freigestellt selbst eine Verlustfunktion bzw. einen negativen Gradienten zu definieren, per Standardeinstellung werden aber beispielsweise für die Binomial- oder die Poisson-Verteilung die entsprechenden negativen log-Likelihoods verwendet.

Eine weitere wichtige Funktion ist die Funktion `boost.control()`. Hier können die Tuning-Parameter für den Boosting-Algorithmus festgelegt werden. Beispielsweise ist die Auswahl der maximalen Anzahl an Iterationsschritten  $m_{\text{stop}}$  oder der Schrittweite  $\nu$  möglich [[R-Hilfeseite zu `gamboost\(\)`](#)].

## 5. Datensituation

In dieser Arbeit soll der Einfluss einiger umweltspezifischer Kovariablen auf die Verbissrate der fünf Baumarten Buche, Bergahorn, Tanne, Fichte und Vogelbeere untersucht werden. Die Datei „2002 Datentabelle mit Koordinaten.xls“, beinhaltet die Ergebnisse der Erhebung im Rahmen der Waldinventur 2002. Bei dieser Waldinventur wurde das Gebiet des Nationalpark Bayerischer Wald in Inventurpunkte unterteilt, die jeweils eine Größe von ca.  $25m^2$  umfassen. Jede Datensatzzeile enthält die Kovariablenausprägungen für genau einen dieser Inventurpunkte. Des Weiteren sind die Anzahl verbissener Buchen, Bergahorne, Tannen, Fichten und Vogelbeeren für jeden Inventurpunkt aufgeführt.

Zur Modellierung der Verbissraten werden zwei unterschiedliche Ansätze verwendet.

Mittels logistischer Regressionsmodelle kann der Einfluss der Kovariablen auf die *Chance auf Verbiss* untersucht und quantifiziert werden. Hierfür wird ein zusätzlicher Datensatz benötigt, wobei jede Zeile genau einen Baum repräsentiert. Für jeden dieser aufgeführten Bäume wurde der Verbissstatus erhoben, sowie weitere Kovariablen wie beispielsweise die Baumhöhe, die in „2002 Datentabelle mit Koordinaten.xls“ natürlicherweise nicht enthalten sind. Zur Modellierung der logistischen Regressionsmodelle wurde deshalb der neue Datensatz „daten\_bin.Rda“ erzeugt, der in jeder Zeile genau einen Baum enthält. Zusätzlich zu den baumspezifischen Einflussgrößen wurden die Kovariablenausprägungen für den Inventurpunkt, in dem sich der jeweilige Baum befindet, hinzugefügt. Um die logistischen Regressionsmodelle getrennt für die fünf Baumarten anzupassen, wurden lediglich die Bäume aus „daten\_bin.Rda“ berücksichtigt, die der betrachteten Baumart zugeordnet werden.

Durch Poissonmodelle kann die *erwartete Anzahl* an verbissenen Bäumen der jeweiligen Baumart modelliert werden. Hierfür kann direkt der Datensatz „2002 Datentabelle mit Koordinaten.xls“ verwendet werden. Für das Poissonmodell einer bestimmten Baumart werden nur die Inventurpunkte berücksichtigt, in denen die Baumart überhaupt vorhanden ist.

In beiden Datensätzen mussten einige Korrekturen vorgenommen werden, da eine bestimmte Datensatzstruktur von dem Statistikprogramm R vorausgesetzt wird. Beispielsweise mussten fehlende Werte mit NA kodiert werden.

Bei der Waldinventur wurden zunächst Kovariablen für jeden Inventurpunkt erhoben. Dabei können die Inventurpunkte durch die Standortkoordinaten exakt identifiziert werden. Neben der Anzahl verbissener Buchen, Bergahorne, Tan-

## 5. Datensituation

nen, Fichten und Vogelbeeren ist auch deren Bestand im Inventurpunkt erfasst, sowie der Nadelholz- und Laubholzzaltbestand. Des Weiteren ist die Verjüngungsdichte (Bäume mit einer Höhe unter 2m) pro Hektar enthalten. Neben einigen geographischen Gegebenheiten, z.B. die Höhe über NN oder die Hangneigung, wird die Bodenbeschaffenheit näher beschrieben. Der Deckungsgrad an beispielsweise Gras oder Beeren wird in Prozentklassen angegeben. Die Rothirsch-, Reh- und Luchsdichte wurde mit Hilfe eines Kerndichteschätzers basierend auf Spuren der Tiere im Winter abgeschätzt. Zusätzlich wurden baumspezifische Kovariablen erhoben. Neben dem Verbissstatus ist vor allem die Baumhöhe bedeutend. Diese kann maximal einen Wert von 200cm annehmen, da nur verjüngte Bäume in der Analyse berücksichtigt werden. Eine ausführliche Übersicht der erhobenen Einflussgrößen ist in der folgenden Tabelle dargestellt, die sowohl den Variablennamen, als auch den Datentyp sowie genauere Erläuterungen zu dieser Einflussgröße beinhaltet.

### Kovariablen der Waldinventur 2002

#### Kovariablen für die Inventurpunkte

Variable	Datentyp	Erläuterungen
KOORD	numerisch	Koordinate des Inventurpunktes (Identifikation des Inventurpunktes)
X_BEST	numerisch	x-Koordinate des Inventurpunktes
Y_Best	numerisch	y-Koordinate des Inventurpunktes
Bergahorn_verb	numerisch	Anzahl verbissener Bergahorne im Inventurpunkt
Buche_verb	numerisch	Anzahl verbissener Buchen im Inventurpunkt
Fichte_verb	numerisch	Anzahl verbissener Fichten im Inventurpunkt
Tanne_verb	numerisch	Anzahl verbissener Tannen im Inventurpunkt
Vogelbeere_verb	numerisch	Anzahl verbissener Vogelbeeren im Inventurpunkt
Verbiss	numerisch	Anzahl verbissener Bäume im Inventurpunkt
Bergahorn	numerisch	Anzahl Bergahorn-Bäume im Inventurpunkt
Buche	numerisch	Anzahl Buchen im Inventurpunkt
Fichte	numerisch	Anzahl Fichten im Inventurpunkt
Tanne	numerisch	Anzahl Tannen im Inventurpunkt
Vogelbeere	numerisch	Anzahl Vogelbeeren im Inventurpunkt
Buche_alt	numerisch	Anzahl Buchen mit einer Höhe > 200cm

## 5. Datensituation

Variable	Datentyp	Erläuterungen
Fichte_alt	numerisch	Anzahl Fichten mit einer Höhe > 200cm
LH_alt	numerisch	Anzahl Laubhölzer mit einer Höhe > 200cm
NH_alt	numerisch	Anzahl Nadelhölzer mit einer Höhe > 200cm
Verjüngung	numerisch	Dichteschätzung der Bäume mit 20–200cm Höhe pro Hektar
h_nn	numerisch	Höhe über NN
expos_neu	kategorial	Exposition/Ausrichtung des Hanges <b>warm:</b> südlich geprägte Richtungen (SO-S-SW-W) <b>kalt:</b> nördlich geprägte Richtungen (NW-N-NO-O)
Neigung	numerisch	Neigung des Hanges in Grad gegenüber der Horizontalen
RANDBEREICH	binär	<b>1:</b> ja <b>0:</b> nein
STADIUM_neu	kategorial	<b>offen:</b> unbestockte Flächen <b>strukturreich</b> <b>strukturarm</b> <b>Totholz</b>
BEST_DIST	numerisch	Distanz des Inventurpunktes zu Bestandesgrenzen in [m]
Tot_ausger	binär	<b>1:</b> Inventurpunkt ist eine ausgeräumte Totholzfläche <b>0:</b> Inventurpunkt ist keine ausgeräumte Totholzfläche
Wgatr_Dist	numerisch	Distanz zum Wintergatter in [m]
Jagd_dis2	numerisch	Distanz zu bejagten Flächen in [m]
Str_dist	numerisch	Distanz zur nächsten Straße in [m]
Wege_dist	numerisch	Distanz zu Wanderwegen in [m]
nass	numerisch	Deckungsgrad von Nass- und Feuchtflächen in Prozentklassen <b>0:</b> < 0,1% <b>1:</b> 0,1-5% <b>2:</b> 6-10% <b>3:</b> 11-25% <b>4:</b> 26-50% <b>5:</b> 51-75% <b>6:</b> 76-100%
freibo	numerisch	Deckungsgrad von freiliegendem Boden in Prozentklassen pro Inventurpunkt

## 5. Datensituation

Variable	Datentyp	Erläuterungen
freife	numerisch	Deckungsgrad von freiliegendem Fels in Prozentklassen pro Inventurpunkt
fliessg	numerisch	Deckungsgrad von Fließgewässern in Prozentklassen pro Inventurpunkt
gras	kategorial	Deckungsgrad von Gras in Prozentklassen pro Inventurpunkt
farn	numerisch	Deckungsgrad von Farn in Prozentklassen pro Inventurpunkt
beer	numerisch	Deckungsgrad von beerentragenden Zwergsträuchern in Prozentklassen pro Inventurpunkt
him_brom	numerisch	Deckungsgrad von Him- und Brombeere in Prozentklassen pro Inventurpunkt
moos	numerisch	Deckungsgrad von Moos in Prozentklassen pro Inventurpunkt
Reh	numerisch	Dichteschätzung aus Punktbeobachtungen der Rehe
Rothirsch	numerisch	Dichteschätzung aus Punktbeobachtungen der Rothirsche
Luchs	numerisch	Dichteschätzung aus Punktbeobachtungen der Luchse

### Baumspezifische Kovariablen

Variable	Datentyp	Erläuterungen
damage	binär	Verbissstatus des Baums <b>1</b> : Baum ist verbissen <b>0</b> : Baum ist nicht verbissen
region	kategorial	<b>R-L</b> : Rachel-Lusen-Gebiet <b>F-R</b> : Falkenstein-Rachel-Gebiet
species	kategorial	Baumart <b>beech</b> : Buche <b>sycamore</b> : Bergahorn <b>fir</b> : Tanne <b>spruce</b> : Fichte <b>rowanbery</b> : Vogelbeere
height	numerisch	Baumhöhe in cm
level	kategorial	Hanglage: Mischwald, v.a. Buche, Tanne, Bergahorn Auflichte: Tallage, sumpfiger Boden Hochlage: v.a. Fichte, Vogelbeere

## 6. Prognostische Faktoren für Wildverbiss

In diesem Kapitel werden nun die Modelle mit der besten Anpassung zur Schätzung der Verbissrate für die verschiedenen Baumarten vorgestellt. Dazu wird zunächst das allgemeine Vorgehen zur Bestimmung des besten Modells beschrieben. Hierbei ist vor allem der Vergleich der unterschiedlichen Modellierungsstrategien von großer Bedeutung. Bei der Modellwahl spielen sowohl die Variablenselektion als auch die Spezifizierung der Art des Kovariableneinflusses eine wichtige Rolle. Des Weiteren werden die Wahl der Basisfunktionsansätze zur Modellierung der glatten Terme, die Wahl der Verteilungsfamilie, sowie die Parameterwahl für die Algorithmen näher erläutert. Anschließend wird für jede Baumart das Binomialmodell zur Modellierung der *Chance auf Verbiss* vorgestellt, das die beste Anpassung besitzt. Zusätzlich folgt die Präsentation des Poissonmodells zur Modellierung der *erwarteten Anzahl verbissener Bäume*, das ebenfalls die beste Anpassung liefert.

### 6.1. Allgemeine Vorgehensweise

#### Vergleich der Modellierungsstrategien

Wie in den Kapiteln 3 und 4 beschrieben, gibt es zwei grundsätzliche Möglichkeiten additive Binomial- bzw. Poissonmodelle zu konstruieren.

Bei der Verwendung der Funktion `gam()` aus dem Paket `mgcv` können sowohl parametrische als auch nonparametrische Einflussgrößen berücksichtigt werden, jedoch besitzt diese Funktion keine integrierte Variablenselektion. Das macht eine Variablenselektion per Hand notwendig. Hierfür wurde zunächst ein Modell mit allen möglichen Einflussgrößen angepasst, wobei alle metrischen Kovariablen nonparametrisch aufgenommen wurden. Anschließend wurden mit Hilfe der `summary`-Option die nonparametrischen Einflussgrößen identifiziert, die auch parametrisch ins Modell aufgenommen werden könnten. In Frage kommen hierbei die Einflussgrößen, deren Anzahl an geschätzten Freiheitsgraden ca. 1 ist, da 1 Freiheitsgrad einem linearen Einfluss entspricht. Im nächsten Schritt wurde jeweils für die am wenigsten signifikante Einflussgröße überprüft, ob diese im Modell vernachlässigt werden kann. Hierfür wurde die `anova()`-Funktion verwendet, die die Devianzen zweier hierarchischer Modelle miteinander vergleicht (vgl. Kapitel 3.2.2).

## 6. Prognostische Faktoren für Wildverbiss

Dieses Vorgehen wird so lange wiederholt, bis die Entfernung jeder noch vorhandenen Einflussgröße zu einer signifikanten Modellverschlechterung führt.

Da diese Methode sehr aufwendig sein kann, bietet das Boosting eine hilfreiche Alternative. Bei der Verwendung der Funktion **gamboost ()** aus dem Paket **mboost** können ebenfalls parametrische sowie nonparametrische Einflussgrößen berücksichtigt werden. Beim Boosting findet jedoch durch die Funktionsweise des Algorithmus automatisch eine Variablenselektion statt. Daher wurde diese Methode dazu genutzt, die Kovariablen zu identifizieren, die einen Einfluss auf die Zielgröße besitzen. Anschließend wurde eine Modellanpassung mit der **gam ()**-Funktion durchgeführt, wobei lediglich die vom Boosting-Verfahren ausgewählten Einflussgrößen berücksichtigt wurden.

Für jede Baumart wurden jeweils zwei Binomial- und zwei Poissonmodelle angepasst, wobei die beiden oben genannten Verfahren zur Modellfindung benutzt wurden. Anschließend wurde das „bessere“ der beiden Modelle ausgewählt, wobei hierfür der GCV-Score als Auswahlkriterium verwendet wurde. Das resultierende Modell wies also den geringsten GCV-Score auf.

### Schätzung mit **gam ()**

Bei der Schätzung mittels der **gam ()**-Funktion ist zunächst die Wahl der Basisfunktionen für die Modellierung der nonparametrischen Einflussgrößen notwendig. Die zweidimensionale räumliche Komponente, die den Koordinatenstandpunkt der Bäume bzw. Inventurpunkte berücksichtigt wird mit einem Thin-Plate-Regression-Spline konstruiert. Für die eindimensionalen glatten Einflussgrößen werden jedoch kubische Splines als Basisfunktionen verwendet, da diese computationally gesehen weniger aufwendig zu konstruieren sind und trotzdem ein adäquates Ergebnis liefern. Da der geschätzte Dispersionsparameter teilweise stark von dem für Binomial- und Poissonverteilungen festgesetzten Wert 1 abweicht, ist die Verwendung von Quasilikelihoodansätzen eine sinnvolle Wahl. Bei den Poissonmodellen musste zusätzlich die logarithmierte Anzahl der Bäume der jeweiligen Baumart im Inventurpunkt als **offset** berücksichtigt werden. Eine gute Schätzung der erwarteten Anzahl verbissener Bäume macht die Beachtung der Anzahl der Bäume dieser Baumart im Inventurpunkt notwendig.

### Schätzung mit Boosting

Wie in Kapitel 4.2 beschrieben ist die Wahl der base-procedure für das Boosting essentiell. Für die zweidimensionale Modellierung der räumlichen Koordinaten wird die base-procedure **bspatial** verwendet, die für eine solche Kovariable geeignet ist. Eindimensionale nonparametrische Einflussgrößen werden mit der Funktion **bbs** ins Modell aufgenommen, was eine Modellierung mit komponentenweisen P-Splines als base-procedures bedeutet. Für die linearen Einflussgrößen ist eine komponentenweise Kleinste-Quadrate-Schätzung als base-

## 6. Prognostische Faktoren für Wildverbiss

procedure sinnvoll, die mittels der Funktion `bo1s` durchgeführt wird. Des Weiteren kann über die `family`-Option die Art des Boostings, also Binomial- oder PoissonBoosting ausgewählt werden. Dabei kann im Fall des Poissonmodells ebenfalls ein Offset berücksichtigt werden, der hier aber in der Verteilungsfamilie spezifiziert werden muss und nicht wie im GAM als Einflussgröße hinzugefügt werden kann. Mittels der `boost.control()`-Funktion kann eine feste Anzahl an durchzuführenden Iterationen für den Boosting-Algorithmus vorgegeben werden. Diese Anzahl wurde zunächst auf 500 gesetzt. Anschließend wurde mit Hilfe von Bootstrapping, eine Alternative zur Kreuzvalidierung, die optimale Anzahl an Boosting-Iterationen bestimmt. Die für diese optimale Anzahl an Iterationen resultierende Variablenselektion wurde anschließend als Grundlage der Modelldefinition für ein GAM benutzt.

### Darstellung der nonparametrischen Einflussgrößen

Die Funktionsschätzungen der nonparametrischen Einflussgrößen werden für die jeweiligen Modelle graphisch dargestellt. Dabei wurde zur besseren Interpretierbarkeit für jede glatte Komponente individuell ein Ausschnitt der  $y$ -Achse gewählt. Dadurch wird der genaue Funktionsverlauf besser sichtbar. Bei dieser Darstellung ist jedoch ein Vergleich des Einflusses der einzelnen nonparametrischen Kovariablen auf die Zielgröße nicht möglich. Deshalb sind im Anhang die Plots der Funktionsschätzungen mit einer einheitlichen  $y$ -Achse für jedes Modell zu finden. Die rote Linie kennzeichnet die 0-Linie.

## 6.2. Vogelbeere

### 6.2.1. Binomialmodell

Bei der Waldinventur im Jahr 2002 wurde der Verbissstatus sowie die baumspezifischen Kovariablen von 4738 Vogelbeeren erhoben. Zusätzlich liegen umweltspezifische Einflussgrößen für 1261 Inventurpunkte vor, in denen sich Vogelbeeren befinden.

Von den beiden in Kapitel 6.1 vorgestellten Methoden liefert das Quasibinomialmodell mit anschließender Variablenselektion per Hand eine etwas bessere Anpassung gegenüber dem Quasibinomialmodell, das aus dem Boosting-Verfahren resultiert.

Für das resultierende Quasibinomialmodell ergibt sich folgende Zusammenfassung:

```
Family: quasibinomial
Link function: logit
Parametric coefficients:
(Intercept)                Estimate Std. Error Pr(>|t|)
                        -2.95554      0.27759 < 2e-16 ***
```

## 6. Prognostische Faktoren für Wildverbiss

```

freife                -0.22717      0.08482 0.007424 **
gras                  0.11911      0.03218 0.000217 ***
farn                  0.25786      0.05553 3.51e-06 ***
him_brom              0.31160      0.06923 6.93e-06 ***
STADIUM_neustrukturarm -0.10111      0.26247 0.700089
STADIUM_neustrukturreich -0.05483      0.25466 0.829545
STADIUM_neuTotholz   0.80047      0.25921 0.002026 **
levelhanglage        0.56294      0.17082 0.000990 ***
levelhochlage        0.35171      0.29878 0.239185

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(x,y)	28.688	28.688	6.985	< 2e-16	***
s(height)	7.287	7.287	8.461	1.17e-10	***
s(Verjüngung)	8.185	8.185	7.280	8.00e-10	***
s(h_nn)	7.233	7.233	4.024	0.000172	***
s(Neigung)	8.829	8.829	6.461	4.79e-09	***
s(BEST_DIST)	6.468	6.468	3.104	0.003814	**
s(Wgatr_Dist)	8.985	8.985	10.202	1.09e-15	***
s(Jagd_dist)	5.950	5.950	6.146	2.15e-06	***
s(Str_dist)	8.250	8.250	2.485	0.010079	*
s>Wege_dist)	8.034	8.034	10.344	1.66e-14	***
s(Reh)	8.514	8.514	5.467	3.33e-07	***
s(Rothirsch)	7.549	7.549	10.364	8.42e-14	***
s(Luchs)	9.000	9.000	8.172	3.88e-12	***
s(Buche_alt)	1.000	1.000	7.048	0.007958	**
s(Fichte_alt)	8.599	8.599	12.811	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.338    Deviance explained = 32%  
GCV score = 0.80092    Scale est. = 0.77681    n = 4738

Wie aus der Zusammenfassung ersichtlich wird, besitzen alle ins Modell aufgenommenen Kovariablen einen signifikanten Einfluss zu einem Signifikanzniveau von 5%. Lediglich bei den kategorialen Variablen STADIUM\_neu und level sind nicht alle Faktorstufen von der Referenzkategorie signifikant verschieden.

Der geschätzte Scale- bzw. Dispersionsparameter ist kleiner als 1, das Modell weist also eine Unterdispersion auf. Die dem Modell zugrundeliegende empirische Varianz ist demnach kleiner als die vom theoretischen Modell unterstellte Varianz.

Von diesem Modell werden jedoch nur 32% der Devianz erklärt, was ein niedriger Wert ist.

## Parameterinterpretation

Der glatte Funktionsverlauf der nonparametrischen Einflussgrößen ist in Abbildung 6.1 dargestellt. Bei der Interpretation der glatten Einflussgrößen ist folgendes zu beachten: Ist die 0-Linie in den Konfidenzbändern enthalten, so ist keine Aussage über die Wirkungsweise der betrachteten Kovariablen auf die Zielgröße möglich. Verläuft die Funktion oberhalb der 0-Linie, so ist die Chance auf Verbiss erhöht, verläuft sie unterhalb der 0-Linie, so ist die Chance auf Verbiss als vermindert zu interpretieren.

Für die **räumliche Komponente** ergibt sich am östlichen bzw. südlichen Randbereich des Nationalparks eine erhöhte Chance auf Verbiss, hingegen im westlichen Randbereich eine verminderte Chance auf Verbiss.

Eine **Baumhöhe** kleiner als 30cm bzw. größer als 150cm hat eine Verminderung der Chance auf Verbiss zur Folge. Die größte Chance auf Verbiss ergibt sich für Bäume mit einer Höhe zwischen 50cm und 110cm. Die beiden Verbisshochpunkte bei 50cm und 100cm sind vermutlich durch die bevorzugten Verbisshöhen von Rehen und Rothirschen zu erklären. Von den Tieren werden vorwiegend Bäume verbissen, deren Größe mit ihrer eigenen Kopfhöhe übereinstimmt.

Die **Verjüngungsdichte pro Hektar** bezeichnet die Dichte an Bäume mit einer Höhe zwischen 20cm und 200cm auf einer Fläche von einem Hektar. Die Chance auf Verbiss ist für eine Verjüngungsdichte von ca. 100 Bäumen pro Hektar erhöht. Für eine weiter ansteigende Verjüngungsdichte sinkt die Chance wieder ab. Der Verlauf des Graphen ab einer Verjüngungsdichte von ca. 300 Bäumen basiert nur noch auf wenigen Beobachtungen und sollte deshalb nicht überbewertet werden.

Mit zunehmender **Höhe über NN** erhöht sich die Chance auf verbissene Vogelbeeren.

Für gar keine **Hangneigung** bzw. für eine Hangneigung von ca. 25° folgt eine leicht verminderte Chance auf Verbiss. Ab einer Hangneigung von 25° nimmt die Chance auf Verbiss von Vogelbeeren zu.

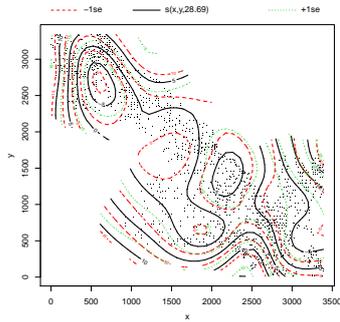
Scheinbar nimmt bis auf eine Entfernung von 100m von der **Bestandesgrenze** die Chance auf Verbiss leicht ab, auf Grund der relativ breiten Konfidenzbänder ist aber eine gesicherte Aussage nicht möglich. Ab 100m Enternung beginnt die Chance auf Verbiss zu steigen.

Mit zunehmender **Entfernung vom Wintergatter** vermindert sich die Chance auf Vogelbeerenverbiss, mit zunehmender **Entfernung von bejagten Flächen** erhöht sich tendenziell die Chance auf Vogelbeerenverbiss.

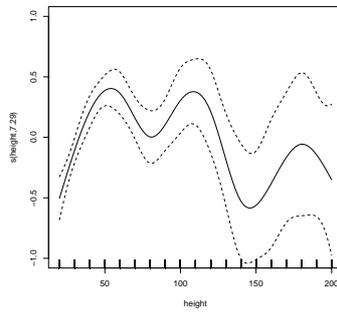
Über die Entwicklung des Vogelbeerenverbisses mit zunehmender **Entfernung von Straßen** ist keine Aussage möglich. Die Konfidenzbänder könnten auch einen waagrechten Verlauf des Graphen enthalten.

Die Chance auf Vogelbeerenverbiss erhöht sich tendenziell mit zunehmender **Entfernung von markierten Wanderwegen**. Ab einer Entfernung von ca. 500m sinkt die Chance ab.

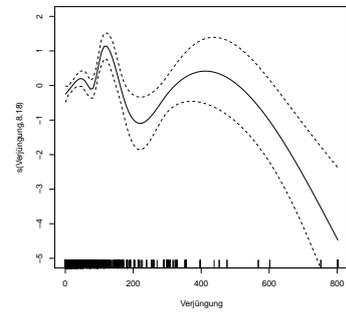
## 6. Prognostische Faktoren für Wildverbiss



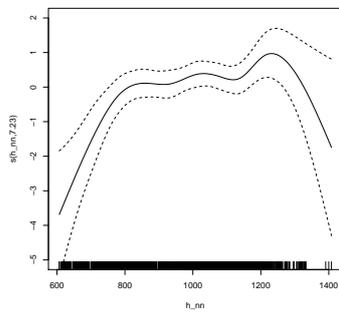
(a) Räumliche Komponente



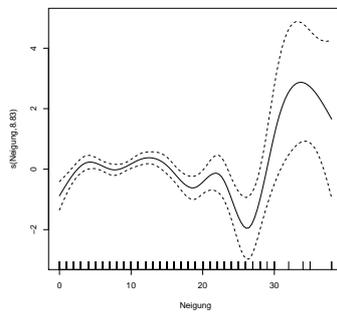
(b) Baumhöhe



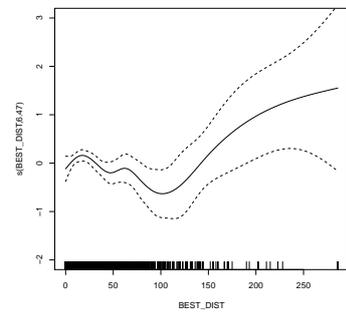
(c) Verjüngungsdichte



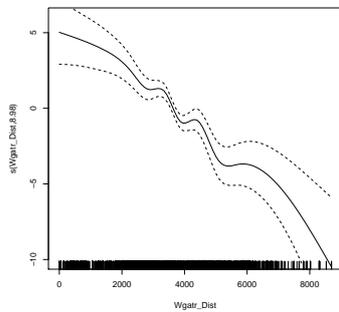
(d) Höhe über NN



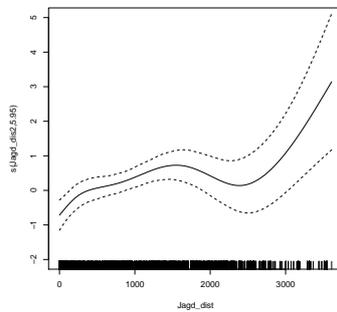
(e) Neigung



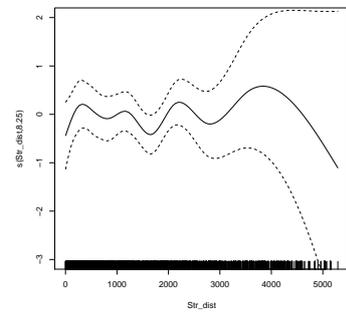
(f) Distanz Bestandesgrenze



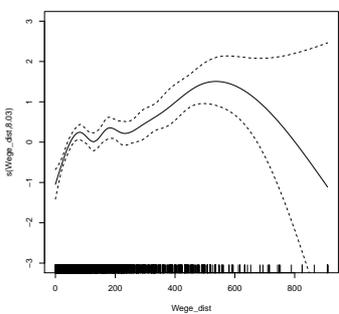
(g) Distanz Wintergatter



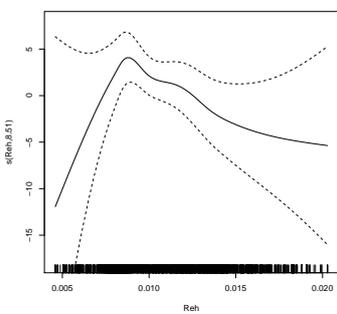
(h) Distanz Jagdfläche



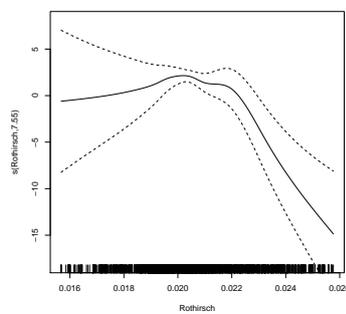
(i) Distanz Straße



(j) Distanz Wanderwege



(k) Rehdichte



(l) Rothirschdichte

## 6. Prognostische Faktoren für Wildverbiss

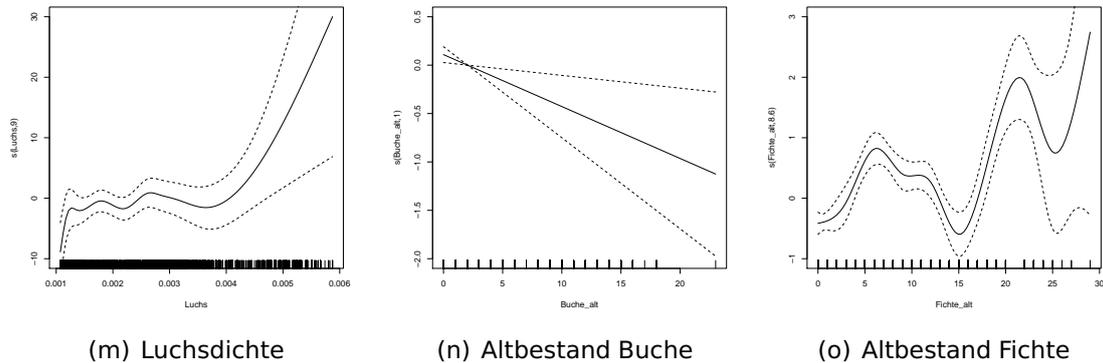


Abbildung 6.1.: Nonparametrische Einflussgrößen Binomialmodell Vogelbeere

Eine Aussage über einen Zusammenhang des Vogelbeerenverbisses und der **Rehdichte** ist ebenfalls nicht möglich, da die Konfidenzbänder einen waagrechteten Verlauf des Graphen enthalten könnten. Lediglich für eine Dichte von ca. 0.08 ergibt sich eine leicht erhöhte Chance.

Für eine **Rothirschdichte** von ca. 0.02 resultiert eine leicht erhöhte Chance auf Vogelbeerenverbiss, im weiteren Verlauf vermindert sich die Chance. Hingegen erhöht sich die Chance auf Verbiss für eine zunehmende **Luchsdichte**.

Mit steigender Anzahl an **Buchen** mit einer Höhe größer als 2m, vermindert sich die Chance auf Vogelbeerenverbiss.

Für den **Fichtenaltbestand** ergibt sich ein unruhiger Verlauf. Bis zu einem Fichtenaltbestand von 5 Bäumen nimmt die Chance auf Verbiss zu, dann sinkt sie bis zu einem Fichtenaltbestand von 15 Bäumen ab um anschließend wieder anzusteigen.

Für die parametrischen Einflussgrößen ergeben sich folgende Interpretationen: Erhöht sich der Deckungsgrad an **freiliegendem Fels** um eine Prozentklasse, so vermindert sich die Chance um das 0,8-fache ( $\exp(-0.22) = 0.8$ ).

Steigt der Deckungsgrad an **Gras** um eine Prozentklasse an, so erhöht sich die Chance auf Vogelbeerenverbiss um das 1,12-fache, für **Farn** um das 1,29-fache und für **Him- und Brombeeren** um das 1,37-fache.

Die kategoriale Variable **Stadium\_neu** hat vier Faktoren, wobei der Faktor „offen“ die Referenzkategorie bildet. Signifikant unterschiedlich von der Referenzkategorie ist die Kategorie „Totholz“. Für diese Faktorstufe ergibt sich eine 2,2-fach höhere Chance auf Vogelbeerenverbiss im Vergleich zur Referenzkategorie.

Die kategoriale Variable **level** besitzt drei Faktorstufen, wobei die Kategorie „Aufichte“ die Referenzkategorie bildet. Einen signifikanten Unterschied liefert die Kategorie „Hochlage“, die im Vergleich zur Referenzkategorie eine 1,8-fach höhere Verbisschance aufweist.

## 6. Prognostische Faktoren für Wildverbiss

### 6.2.2. Poissonmodell

Im Poissonmodell für die Vogelbeere wird die erwartete Anzahl verbissener Vogelbeeren für die 1261 Inventurpunkte, in denen Vogelbeeren vorhanden sind, modelliert. Das Quasipoissonmodell mit anschließender Variablenselektion per Hand liefert ein signifikant besseres Modell als das aus dem Boosting-Algorithmus resultierende Modell. Eine Bestätigung dieses signifikanten Unterschieds konnte mittels der `anova()`-Funktion durchgeführt werden, da es sich bei den beiden „Endmodellen“ um hierarchische Modelle handelt. Des Weiteren weist dieses Quasipoissonmodell einen niedrigeren GCV-Score auf, sowie einen höheren Prozentsatz der erklärenden Devianz.

Für das Modell ergeben sich folgende Parameterschätzungen:

```
Family: quasipoisson
Link function: log
```

Formula:

```
Vogelbeere_verb ~ s(x, y) + s(Verjüngung, bs = "cr") + s(h_nn,
  bs = "cr") + s(Neigung, bs = "cr") + s(Wgatr_Dist, bs = "cr") +
  s(Wege_dist, bs = "cr") + s(Fichte_alt, bs = "cr") + s(Buche_alt,
  bs = "cr") + freife + gras + farn + him_brom + s(Reh, bs = "cr") +
  s(Rothirsch, bs = "cr") + s(Luchs, bs = "cr") + STADIUM_neu +
  offset(log_Vogelbeere)
```

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-3.05390	0.65812	-4.640	3.88e-06	***
freife	-0.14651	0.07532	-1.945	0.05198	.
gras	0.06612	0.02661	2.485	0.01311	*
farn	0.14497	0.04417	3.282	0.00106	**
him_brom	0.13396	0.06117	2.190	0.02872	*
STADIUM_neustrukturarm	0.05482	0.23042	0.238	0.81201	
STADIUM_neustrukturreich	0.12242	0.22262	0.550	0.58249	
STADIUM_neuTotholz	0.59431	0.22761	2.611	0.00914	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(x, y)	27.748	28.248	3.919	4.45e-11	***
s(Verjüngung)	6.162	6.662	3.139	0.003272	**
s(h_nn)	5.891	6.391	2.839	0.008005	**
s(Neigung)	7.336	7.836	3.655	0.000363	***
s(Wgatr_Dist)	8.853	9.353	5.240	3.30e-07	***
s(Wege_dist)	4.861	5.361	9.706	1.39e-09	***
s(Fichte_alt)	7.224	7.724	5.154	3.46e-06	***
s(Buche_alt)	7.626	8.126	1.922	0.052248	.
s(Reh)	6.787	7.287	3.413	0.001077	**

## 6. Prognostische Faktoren für Wildverbiss

```
s(Rothirsch)    6.648   7.148  5.408  3.32e-06 ***
s(Luchs)       8.745   9.245  3.518  0.000222 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.861   Deviance explained = 45.8%
GCV score = 1.2117   Scale est. = 1.1099     n = 1261
```

Für dieses Quasipoissonmodell ergibt sich ein Prozentsatz der erklärenden Devianz von 45.8%. Der geschätzte Scale- bzw. Dispersionsparameter ist mit ca. 1.1 etwas größer als 1. Das Modell weist folglich eine geringe Überdispersion auf.

### Parameterinterpretation

Der Funktionsverlauf der nonparametrischen Einflussgrößen ist in Abbildung 6.2 dargestellt.

Die **räumliche Komponente** weist eine verminderte Anzahl an verbissenen Vogelbeeren im nördlichen Bereich des Nationalparks, sowie eine erhöhte Anzahl für die Randbereiche auf.

Bis zu einer **Verjüngungsdichte** von ca. 100 verjüngten Bäumen pro Hektar steigt die Anzahl verbissener Vogelbeeren. Dann sinkt die Anzahl verbissener Vogelbeeren bis zu einer Dichte von ca. 200 verjüngten Bäumen. Über den weiteren Verlauf ist wegen den breiten Konfidenzbändern keine Aussage möglich. Die Anzahl verbissener Vogelbeeren ist bis zu einer **Höhe** von 750m **über NN** vermindert. Der weitere Verlauf ist relativ konstant.

Für eine geringe bzw. überhaupt keine **Hangneigung** ergibt sich eine geringere Anzahl verbissener Vogelbeeren, im Bereich um 5° und 12° Hangneigung ergibt sich eine leicht erhöhte Anzahl.

Mit zunehmender **Distanz vom Wintergatter** vermindert sich die Anzahl verbissener Vogelbeeren, mit zunehmender **Distanz von markierten Wanderwegen** erhöht sich die Anzahl verbissener Vogelbeeren hingegen.

Der **Fichtenaltbestand** weist einen sehr unruhigen Verlauf auf. Bis zu einem Fichtenaltbestand von 5 Fichten steigt die Anzahl verbissener Vogelbeeren an, danach sinkt die Anzahl bis zu einem Altbestand von 15 Fichten ab. Anschließend beginnt die Anzahl verbissener Vogelbeeren mit zunehmendem Fichtenaltbestand wieder zu steigen.

Aufgrund der breiten Konfidenzbänder ist keine Aussage über die Wirkungsweise des **Buchenaltbestands** auf die Anzahl verbissener Vogelbeeren möglich. Somit ist auch dem starken Abfall bei einer Anzahl von 13 alten Buchen keine Bedeutung beizumessen.

Über den Zusammenhang der **Rehdichte** und der **Rothirschdichte** mit der Anzahl verbissener Vogelbeeren ist ebenfalls keine Aussage möglich, da die breiten Konfidenzbänder einen konstanter Verlauf ermöglichen.

Für die **Luchsdichte** hingegen sinkt die Anzahl verbissener Vogelbeeren mit

## 6. Prognostische Faktoren für Wildverbiss

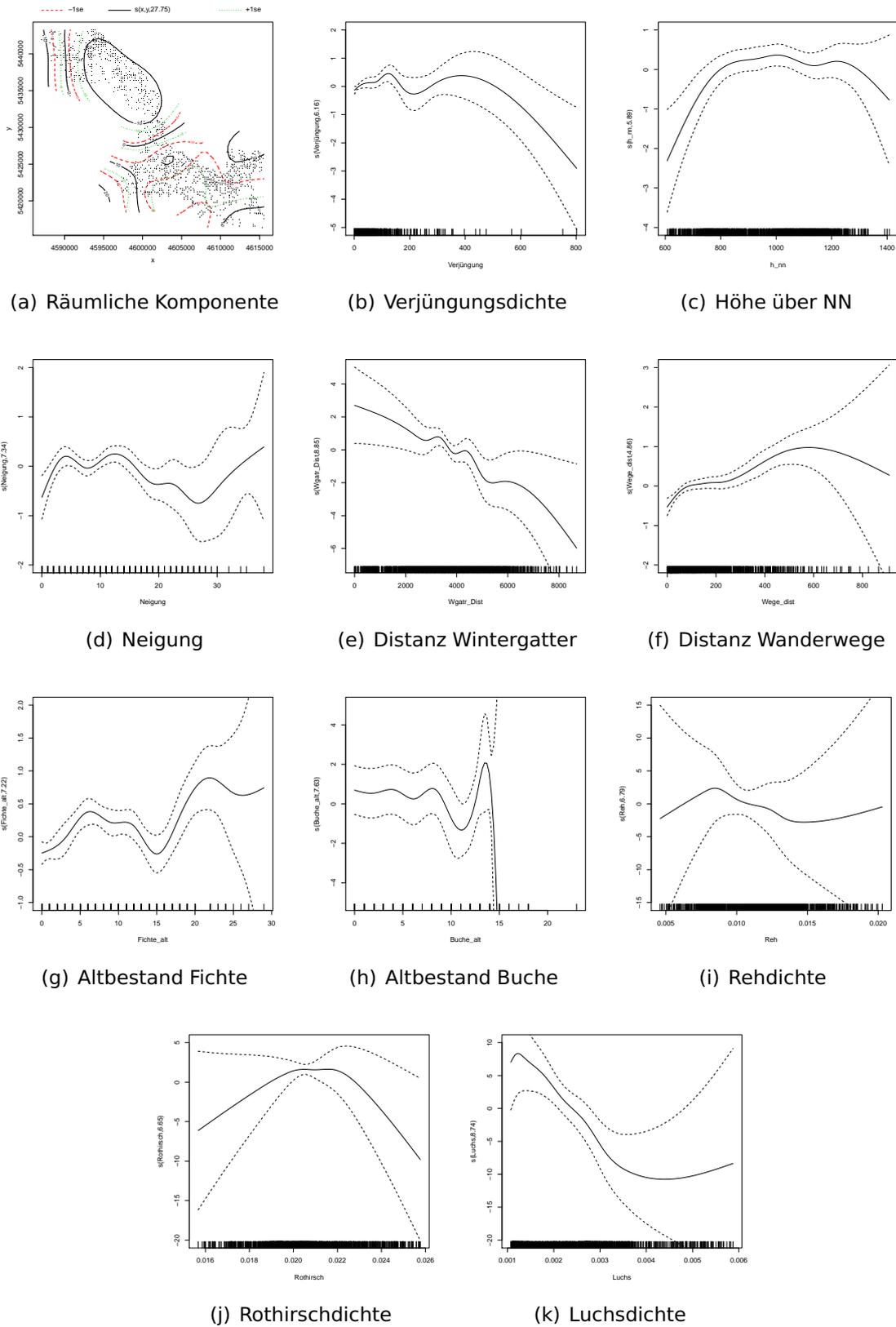


Abbildung 6.2.: Nonparametrische Einflussgrößen Poissonmodell Vogelbeere

## 6. Prognostische Faktoren für Wildverbiss

zunehmender Luchsdichte.

Bei den parametrischen Einflussgrößen weisen die Kovariablen **Gras**, **Farn**, sowie **Him- und Brombeeren** mit steigendem Deckungsgrad eine erhöhte Anzahl verbissener Vogelbeeren auf. Mit zunehmendem Deckungsgrad an **freilegendem Fels** hingegen vermindert sich die Anzahl verbissener Vogelbeeren. Für die kategoriale Variable **Stadium\_neu** ist die Anzahl verbissener Vogelbeeren für die Ausprägung „Totholz“ im Vergleich zur Referenzkategorie „offen“ signifikant erhöht. Die anderen beiden Ausprägungen „struktureich“ und „strukturarm“ bewirken ebenfalls eine Erhöhung, der Unterschied zur Referenzkategorie ist jedoch nicht signifikant.

### 6.3. Tanne

Im Rahmen der Waldinventur 2002 wurde der Verbissstatus von 2498 Tannen erfasst, sowie die umweltspezifischen Einflussgrößen von 649 Inventurpunkten, die einen Tannenbestand aufweisen.

#### 6.3.1. Binomialmodell

Das „beste“ Binomialmodell für die Tanne ist das Quasibinomialmodell mit anschließender Variablenselektion per Hand. In diesem Fall ergibt sich ein deutlich niedrigerer GCV-Score und ein deutlich höherer Prozentsatz der erklärenden Devianz im Vergleich zum aus dem Boosting-Algorithmus resultierenden Quasibinomialmodell.

Für die Parameter ergeben sich folgende Schätzwerte:

Family: quasibinomial

Link function: logit

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-4.150e+00	6.000e-01	-6.916	5.96e-12	***
nass	2.333e-01	7.361e-02	3.169	0.001550	**
freife	4.268e-01	1.677e-01	2.545	0.011001	*
fliessg	-1.288e+00	2.441e-01	-5.276	1.44e-07	***
farn	2.275e-01	1.239e-01	1.837	0.066343	.
beer	2.076e-01	5.863e-02	3.541	0.000407	***
him_brom	-4.022e-01	1.261e-01	-3.190	0.001442	**
NH_alt	-1.738e-01	5.926e-02	-2.932	0.003400	**
STADIUM_neustrukturarm	5.878e-01	5.875e-01	1.000	0.317192	
STADIUM_neustruktureich	1.250e+00	5.572e-01	2.244	0.024927	*
STADIUM_neuTotholz	1.304e+00	6.053e-01	2.154	0.031364	*
Tot_ausgerl	-3.041e+00	8.127e-01	-3.742	0.000187	***
levelhanglage	-6.365e-01	2.321e-01	-2.743	0.006132	**

## 6. Prognostische Faktoren für Wildverbiss

```
levelhochlage          -1.298e+02  3.265e+07 -3.97e-06  0.999997  
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(x,y)	28.685	29.185	5.369	< 2e-16	***
s(height)	8.878	9.378	10.410	< 2e-16	***
s(Verjüngung)	8.566	9.066	3.933	5.23e-05	***
s(h_nn)	7.832	8.332	7.403	4.30e-10	***
s(BEST_DIST)	8.386	8.886	3.339	0.000488	***
s(Wgatr_Dist)	6.220	6.720	4.542	6.43e-05	***
s(Jagd_dist)	3.487	3.987	2.724	0.028155	*
s(Str_dist)	7.836	8.336	5.242	1.03e-06	***
s>Wege_dist)	8.715	9.215	7.911	8.13e-12	***
s(Rothirsch)	8.998	9.498	8.579	2.57e-13	***
s(Luchs)	6.906	7.406	1.893	0.062526	.
s(Fichte_alt)	7.482	7.982	7.933	1.43e-10	***
s(Reh)	1.001	1.501	1.646	0.199118	

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.246   Deviance explained = 30.2%  
GCV score = 0.4988   Scale est. = 0.47345   n = 2498
```

Das endgültige Quasibinomialmodell weist eine erklärende Devianz von 30.2% auf, sowie einen geschätzten Dispersionsparameter von ca. 0.47, was einer Unterdispersion entspricht.

### Parameterinterpretation

In Abbildung 6.3 sind die Funktionsverläufe der glatten Komponenten dargestellt.

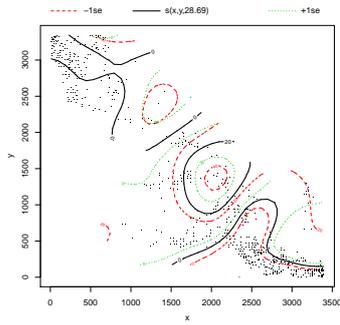
Eine erhöhte Chance auf Tannenverbiss ergibt sich für die **südlichere** Hälfte des Nationalparks. Hiervon sind die Randbereiche jedoch nicht betroffen.

Für eine **Baumhöhe** bis 30 cm resultiert eine verminderte Chance auf Verbiss, sowie für Bäume mit einer Höhe von ca. 160cm. Eine erhöhte Chance auf Verbiss ergibt sich für Bäume mit einer Höhe zwischen 30cm und 150cm. Die Verbisshochpunkte bei 90cm und 130cm lassen sich vermutlich durch die bevorzugte Verbisshöhe der Rehe und Rothirsche erklären.

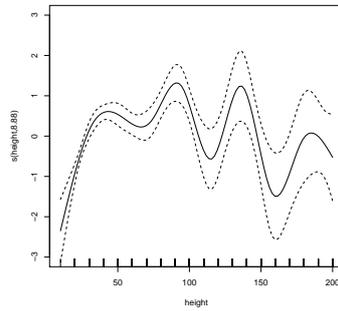
Mit zunehmender **Höhe über NN** bzw. zunehmender **Entfernung vom Wintergatter** erhöht sich die Chance auf Tannenverbiss.

Mit ansteigender **Entfernung von Straßen** und ansteigender **Luchsdichte** hingegen vermindert sich die Chance auf Tannenverbiss. Für eine Luchsdichte ab 0,0035 ist wegen der breiten Konfidenzbänder über den weiteren Verlauf

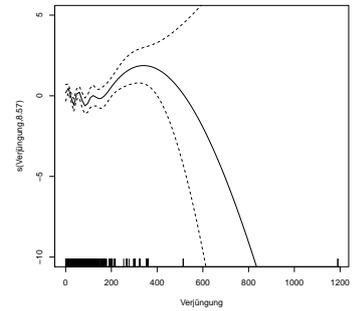
## 6. Prognostische Faktoren für Wildverbiss



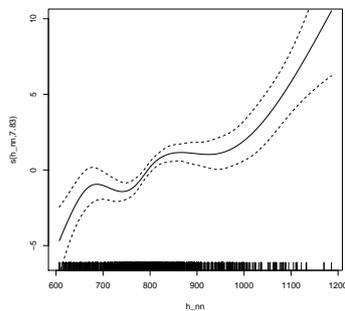
(a) Räumliche Komponente



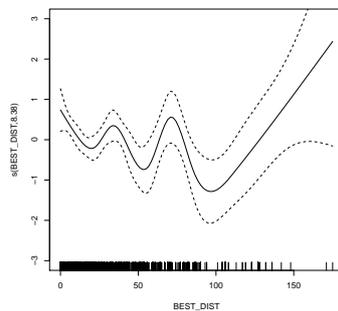
(b) Baumhöhe



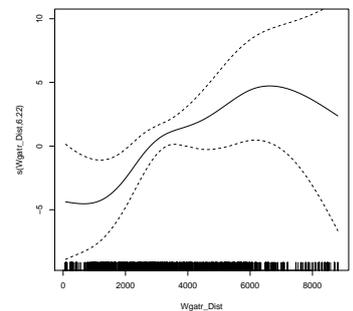
(c) Verjüngungsdichte



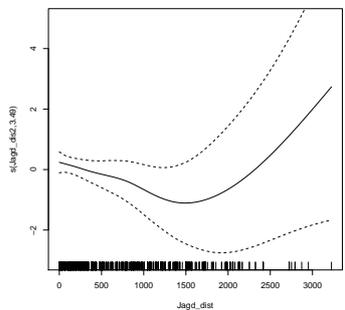
(d) Höhe über NN



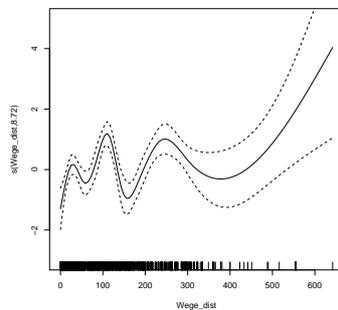
(e) Distanz Bestandesgrenze



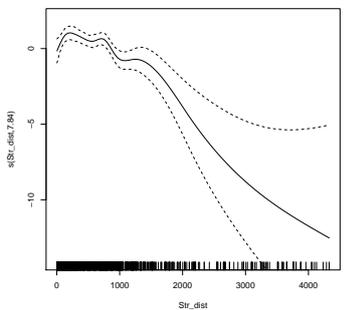
(f) Distanz Wintergatter



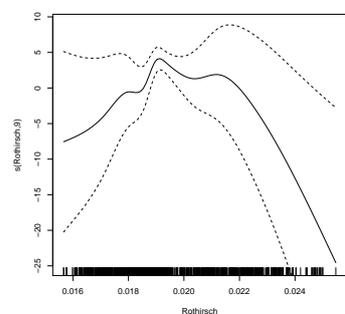
(g) Distanz Jagdflächen



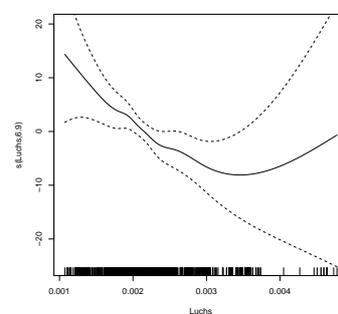
(h) Distanz Wanderwege



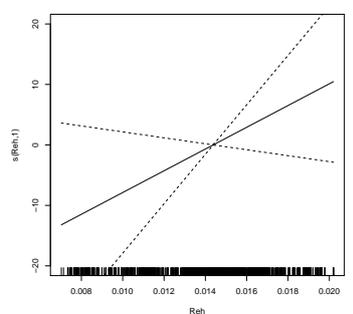
(i) Distanz Straßen



(j) Rothirschdichte

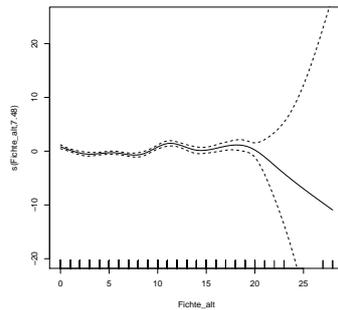


(k) Luchsdichte



(l) Rehdichte

## 6. Prognostische Faktoren für Wildverbiss



(m) Altbestand Fichte

Abbildung 6.3.: Nonparametrische Einflussgrößen Binomialmodell Tanne

keine Aussage möglich.

Die Einflussgrößen **Verjüngungsdichte**, **Distanz zur Bestandesgrenze**, **Distanz von markierten Wanderwegen** und der **Buchenaltbestand** weisen einen unruhigen Verlauf auf. Sie schwanken unregelmäßig um die 0-Linie und sind damit kaum sinnvoll interpretierbar.

Die breiten Konfidenzbänder der **Distanz von bejagten Flächen**, der **Rothirschkichte** und der **Rehdichte** enthalten die 0-Linie. Deshalb ist keine gesicherte Aussage über die Wirkungsweise dieser Kovariablen auf die Verbisschance möglich. Die Rehdichte weist einen linearen Verlauf auf, eine parametrische Aufnahme ins Modell würde jedoch zu einer signifikanten Modellverschlechterung führen.

Für die parametrischen Einflussgrößen ergeben sich folgende Interpretationen: Erhöht sich der Deckungsgrad von **Nass- und Feuchtflächen** um eine Prozentklasse, so erhöht sich die Chance für Tannenverbiss um das 1,26-fache, für **freiliegenden Fels** um das 1,53-fache, für **Farn** um das 1,26-fache und für **beerentragende Kleinsträucher** erhöht sich die Chance um das 1,23-fache. Steigt hingegen der Deckungsgrad von **Fließgewässern** um eine Prozentklasse, so vermindert sich die Chance auf Tannenverbiss um das 0,28-fache, für **Him- und Brombeeren** um das 0,67-fache.

Nimmt der Nadelholzalbestand um einen Baum zu, so vermindert sich die Chance auf Tannenverbiss um das 0,84-fache. Bei der kategorialen Einflussgröße **Stadium\_neu** ergeben sich für die Ausprägungen „struktureich“ und „Totholz“ signifikante Unterschiede im Vergleich zur Referenzkategorie „offen“. Für „struktureich“ erhöht sich die Chance auf Tannenverbiss um das 3,49-fache, für „Totholz“ um das 3,68-fache im Vergleich zur Referenzkategorie. Die dritte Ausprägung „strukturarm“ weist eine Erhöhung des Risikos um das 1,8-fache auf, dieser Unterschied ist jedoch nicht signifikant.

Für eine **ausgeräumte Totholzfläche** ergibt sich eine um das 0,048-fache verminderte Chance auf Tannenverbiss im Vergleich zu keiner ausgeräumten

## 6. Prognostische Faktoren für Wildverbiss

Totholzfläche.

Befindet sich der Baum in „Hanglage“ so vermindert sich die Chance auf Tannenverbiss signifikant um das 0,53-fache im Vergleich zur Referenzkategorie „Aufichte“, für die Ausprägung „Hochlage“ ergibt sich noch eine viel stärkere Verminderung, der p-Wert ist aber fast 1 und damit ist dieser großen Abweichung keine Bedeutung beizumessen.

### 6.3.2. Poissonmodell

Für die 649 Inventurpunkte mit Tannenbestand ergibt sich die beste Modellierung der erwarteten Anzahl an verbissenen Tannen mittels eines Quasipoissonmodells mit anschließender Variablenselektion per Hand. Im Vergleich zu dem aus der Boosting-Methode resultierenden Endmodell ist mit dem Quasipoissonmodell ein deutlich niedrigeren GCV-Score verbunden. Außerdem besitzt das Modell mit einem Wert von 39.4% einen deutlich höheren Prozentsatz der erklärenden Devianz. Des Weiteren weist der geschätzte Dispersionsparameter auf eine Unterdispersion hin. Die Parameterschätzungen sind in der folgenden Zusammenfassung gegeben:

```
Family: quasipoisson
Link function: log
```

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-3.44141	0.74585	-4.614	4.87e-06	***
nass	0.20486	0.06535	3.135	0.001807	**
freie	0.30688	0.12147	2.526	0.011790	*
fliessg	-0.87725	0.24473	-3.585	0.000366	***
beer	0.14591	0.05548	2.630	0.008773	**
Tot_ausger1	-2.34893	0.96860	-2.425	0.015612	*

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(x,y)	12.805	12.805	1.765	0.046189	*
s(h_nn)	1.000	1.000	3.207	0.073837	.
s(BEST_DIST)	8.295	8.295	3.079	0.001803	**
s(Wgatr_Dist)	6.250	6.250	3.756	0.000935	***
s(Str_dist)	3.666	3.666	2.770	0.030726	*
s>Wege_dist)	8.678	8.678	5.569	3.37e-07	***
s(Rothirsch)	9.000	9.000	2.893	0.002361	**
s(Buche_alt)	5.736	5.736	2.023	0.063995	.
s(Fichte_alt)	6.965	6.965	3.986	0.000295	***
s(NH_alt)	4.846	4.846	0.373	0.861957	.

---

## 6. Prognostische Faktoren für Wildverbiss

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.652 Deviance explained = 39.4%

GCV score = 0.88903 Scale est. = 0.7887 n = 649

### Parameterinterpretation

Die Funktionsschätzungen der nonparametrischen Einflussgrößen sind in Abbildung 6.4 dargestellt. Für die **räumliche Komponente** ergeben sich keine offensichtlichen Regionen mit erhöhtem Tannenverbiss.

Mit steigender **Entfernung von Straßen** vermindert sich die Anzahl verbissener Tannen. Die Konfidenzbänder sind ab einer Entfernung von 1500m so breit, dass eine genaue Aussage über den weiteren Verlauf nicht möglich ist.

Die Konfidenzbänder der Kovariablen **Distanz vom Wintergatter** enthalten über weite Teile die 0-Linie. Lediglich für eine Entfernung von ca. 6000m ist die Anzahl verbissener Tannen leicht erhöht.

Für den Kurvenverlauf der **Rothirschdichte** ist nur eine leichte Erhöhung der Anzahl verbissener Tannen für eine Dichte von ca. 0,019 festzustellen.

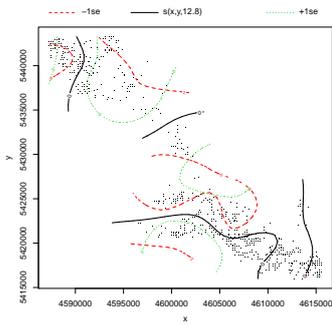
Die Einflussgrößen **Distanz zur Bestandesgrenze**, **Distanz zu markierten Wanderwegen** und der **Fichtenaltbestand** weisen einen sehr unruhigen Verlauf auf. Die Funktionsschätzungen schwanken unregelmäßig um die 0-Linie und sind damit nur schwer sinnvoll interpretierbar. Beispielsweise ergeben sich bei der Distanz von markierten Wanderwegen lokale „Hochpunkte“ der Anzahl verbissener Tannen für Entfernungen von 100 und 250m. Lokale „Tiefpunkte“ hingegen resultieren für Entfernungen von 50 und 150m.

Die 0-Linie ist in den Konfidenzbändern der Kovariablen **Höhe über NN**, **Buchenaltbestand** und **Altbestand an Nadelhölzern** komplett enthalten. Eine Aussage über den Einfluss dieser Größen auf die Anzahl verbissener Tannen ist also nicht möglich. Der Verlauf der Höhe über NN ist linear, eine parametrische Aufnahme ins Modell führt jedoch zu einer signifikanten Verschlechterung.

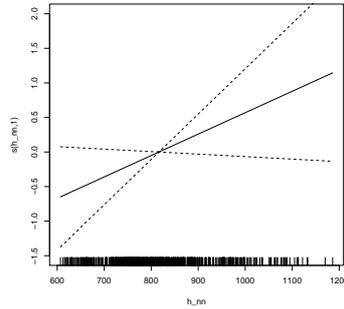
Für die parametrischen Einflussgrößen ergeben sich die folgenden Parameterinterpretationen:

Ein erhöhter Deckungsgrad an **Nass- und Feuchtflächen, freiliegendem Fels** und **beerentragenden Kleinsträuchern** hat eine erhöhte Anzahl verbissener Tannen zur Folge. Ein erhöhter Deckungsgrad an **Fließgewässern** hingegen bewirkt eine niedrigere Anzahl verbissener Tannen. Für ausgeräumte Totholzflächen resultiert ebenfalls eine niedrigere Anzahl verbissener Tannen im Vergleich zu keinen ausgeräumten Totholzflächen.

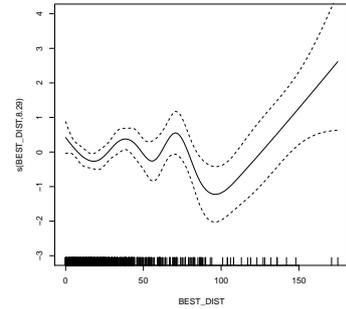
## 6. Prognostische Faktoren für Wildverbiss



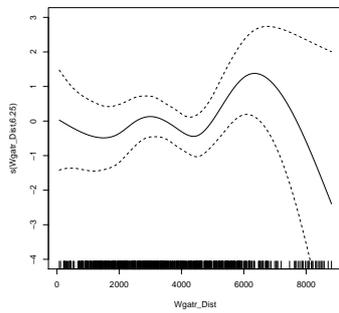
(a) Räumliche Komponente



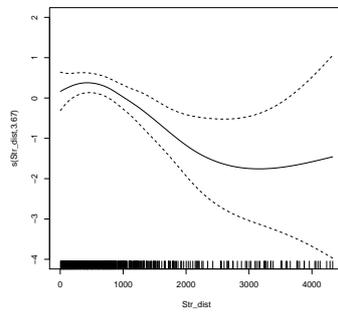
(b) Höhe über NN



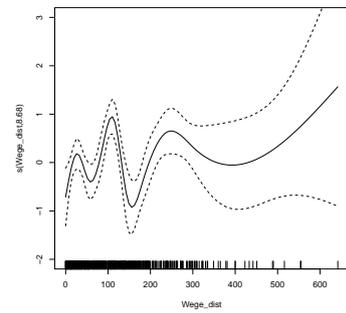
(c) Distanz Bestandesgrenze



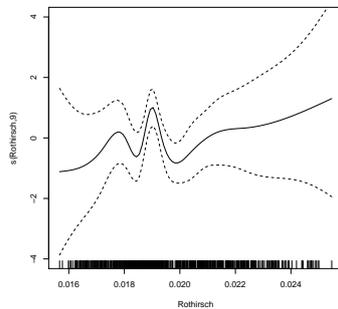
(d) Distanz Wintergatter



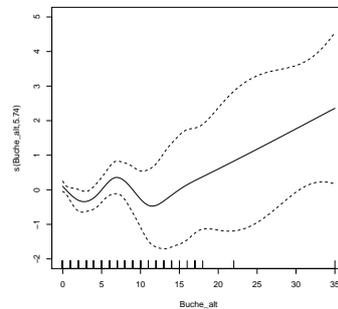
(e) Distanz Straße



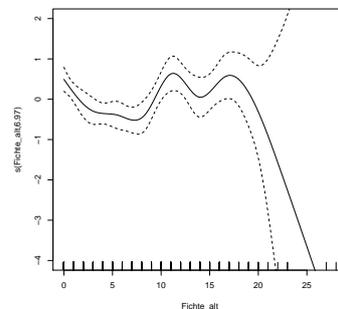
(f) Distanz Wanderwege



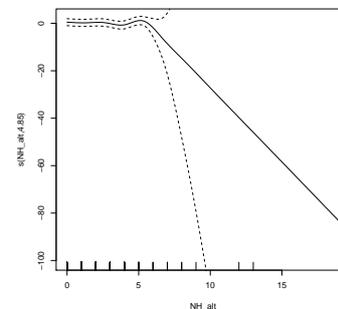
(g) Rothirschdichte



(h) Altbestand Buche



(i) Altbestand Fichte



(j) Altbestand Nadelhölzer

Abbildung 6.4.: Nonparametrische Einflussgrößen Poissonmodell Tanne

## 6. Prognostische Faktoren für Wildverbiss

### 6.4. Buche

Bei der Waldinventur wurde der Verbissstatus von 32461 Buchen erhoben, sowie die umweltspezifischen Kovariablen für 1701 Inventurpunkte, die einen Buchenbestand aufweisen.

#### 6.4.1. Binomialmodell

Da der Datensatz zur Analyse der Verbissraten für Buchen mit einer Länge von 32461 sehr viele Bäume beinhaltet, sind die in Kapitel 6.1 vorgestellten Modellierungsverfahren nicht anwendbar. Das Statistikprogramm R kann die entsprechenden Algorithmen für eine so große Datenmenge nicht durchführen. Deshalb musste die Anzahl der im Modell berücksichtigten Buchen verkleinert werden. Dazu wurde zunächst die maximale Anzahl an Buchen, die aus einem einzelnen Inventurpunkt berücksichtigt werden sollen, auf 30 festgesetzt. Anschließend wurden aus jedem Inventurpunkt maximal 30 Buchen durch „Sampling“ zufällig ausgewählt. Für Inventurpunkte, in denen der Verbissstatus von weniger als 30 Buchen erhoben wurde, wurden wie üblich alle erhobenen Buchen in der Analyse berücksichtigt. Durch dieses Vorgehen wird garantiert, dass weiterhin Buchen aus dem gesamten Nationalparksgebiet zur Analyse herangezogen werden. Bei einer rein zufälligen Auswahl aus allen Buchen könnten Fälle auftreten, in denen ganze Inventurpunkte bzw. ganze Regionen des Nationalparks nicht in der Modellierung berücksichtigt werden. Durch die Vorgabe von maximal 30 Buchen pro Inventurpunkt ergab sich insgesamt eine Anzahl von 16365 Buchen zur Anpassung des Quasibinomialmodells, sowie eine Reduzierung der ursprünglichen Datensatzgröße um den Faktor 0.476, es wurden also ca. die Hälfte aller erhobenen Buchen in die Analyse eingeschlossen.

Des Weiteren ergab sich durch die Anpassung eines Quasibinomialmodells mit anschließender Variablenselektion per Hand ein besseres Modell als durch den Boosting-Algorithmus. Die Schätzungen der Parameter sind in der folgenden Zusammenfassung dargestellt:

Family: quasibinomial

Link function: logit

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.87088	0.23212	-12.368	< 2e-16	***
nass	0.38398	0.04424	8.679	< 2e-16	***
freibo	0.11161	0.05584	1.999	0.045649	*
freife	0.08142	0.04602	1.769	0.076898	.
fliessg	-0.92729	0.12870	-7.205	6.06e-13	***
farn	-0.45412	0.04177	-10.872	< 2e-16	***
beer	-0.14809	0.03496	-4.236	2.29e-05	***
him_brom	0.36385	0.03831	9.498	< 2e-16	***

## 6. Prognostische Faktoren für Wildverbiss

```
moos                -0.14804    0.04007   -3.694  0.000221 ***
expos_neuwarm       0.28851    0.07701    3.746  0.000180 ***
RANDBEREICH1        0.22739    0.12744    1.784  0.074405 .
STADIUM_neustrukturarm -2.41271    0.16953  -14.232 < 2e-16 ***
STADIUM_neustrukturreich -1.81026    0.16476  -10.987 < 2e-16 ***
STADIUM_neuTotholz  -2.12728    0.18410  -11.555 < 2e-16 ***
Tot_ausger1         -1.03249    0.43669   -2.364  0.018074 *
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(x, y)	28.989	28.989	29.593	< 2e-16 ***
s(height)	8.663	8.663	81.473	< 2e-16 ***
s(Verjüngung)	8.966	8.966	29.590	< 2e-16 ***
s(h_nn)	8.673	8.673	41.823	< 2e-16 ***
s(Neigung)	8.743	8.743	14.851	< 2e-16 ***
s(BEST_DIST)	8.963	8.963	12.755	< 2e-16 ***
s(Wgatr_Dist)	8.857	8.857	26.737	< 2e-16 ***
s(Jagd_dist)	8.211	8.211	23.760	< 2e-16 ***
s(Str_dist)	8.641	8.641	29.366	< 2e-16 ***
s>Wege_dist)	8.984	8.984	42.434	< 2e-16 ***
s(Reh)	8.098	8.098	23.339	< 2e-16 ***
s(Rothirsch)	8.999	8.999	22.910	< 2e-16 ***
s(Luchs)	8.810	8.810	30.635	< 2e-16 ***
s(Buche_alt)	6.959	6.959	20.012	< 2e-16 ***
s(Fichte_alt)	8.852	8.852	8.944	2.15e-13 ***
s(LH_alt)	7.190	7.190	11.718	2.64e-15 ***
s(NH_alt)	5.549	5.549	0.956	0.449

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.286 Deviance explained = 36.1%

GCV score = 0.28131 Scale est. = 0.27827 n = 16365

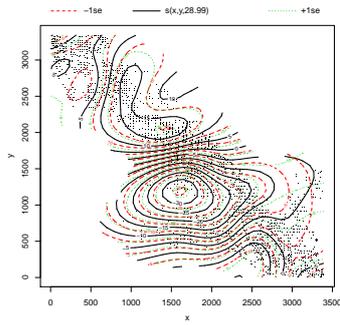
Das Quasibinomialmodell besitzt einen Prozentsatz der erklärenden Devianz von 36.1% und einen geschätzten Dispersionsparameter von ca. 0.28, was einer Unterdispersion entspricht.

### Parameterinterpretation

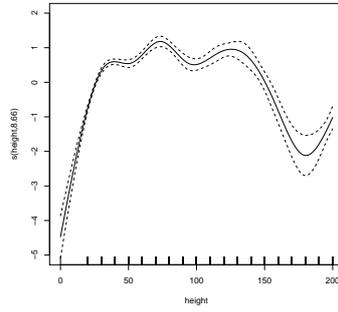
Die geschätzten Funktionen der glatten Einflussgrößen sind in Abbildung 6.5 dargestellt.

Die **räumliche Komponente** zeigt eine Erhöhung der Chance auf Verbiss im nördlichen Gebiet des Nationalparks, im südlichen Gebiet hingegen ist die Chance auf Verbiss vermindert.

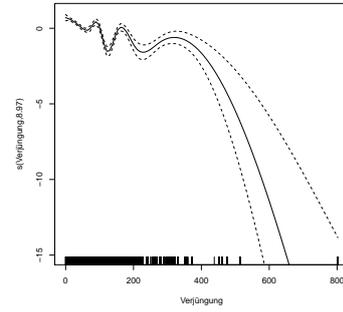
## 6. Prognostische Faktoren für Wildverbiss



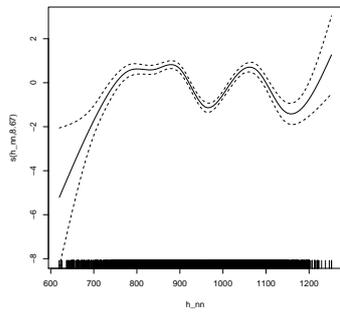
(a) Räumliche Komponente



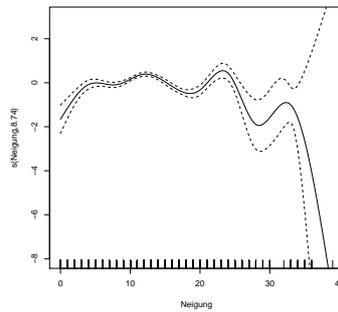
(b) Baumhöhe



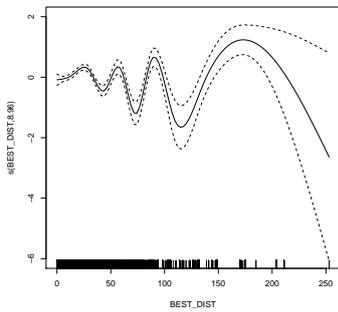
(c) Verjüngungsdichte



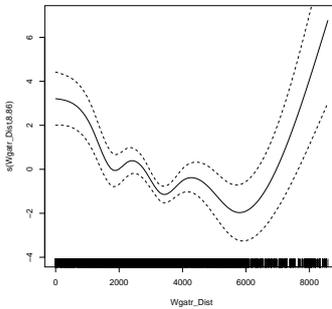
(d) Höhe über NN



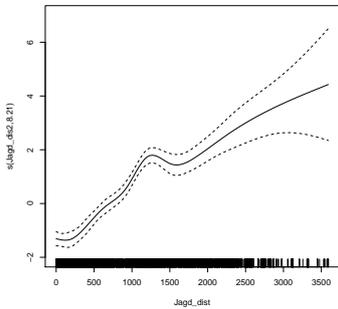
(e) Neigung



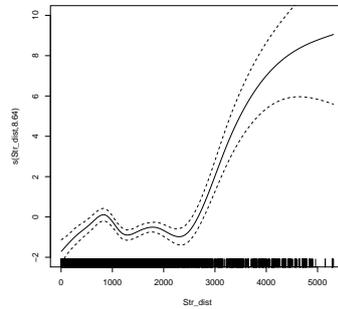
(f) Distanz Bestandesgrenze



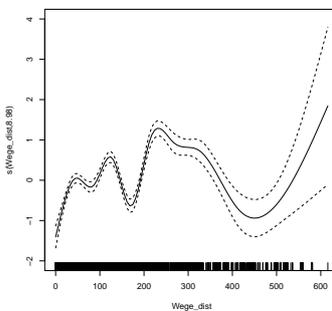
(g) Distanz Wintergatter



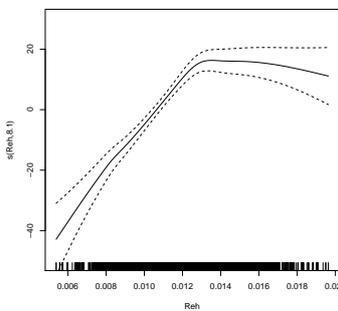
(h) Distanz Jagdflächen



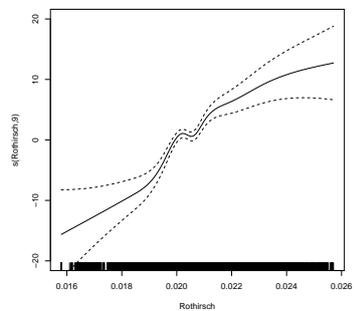
(i) Distanz Straße



(j) Distanz Wanderwege



(k) Rehdichte



(l) Rothirschdichte

## 6. Prognostische Faktoren für Wildverbiss

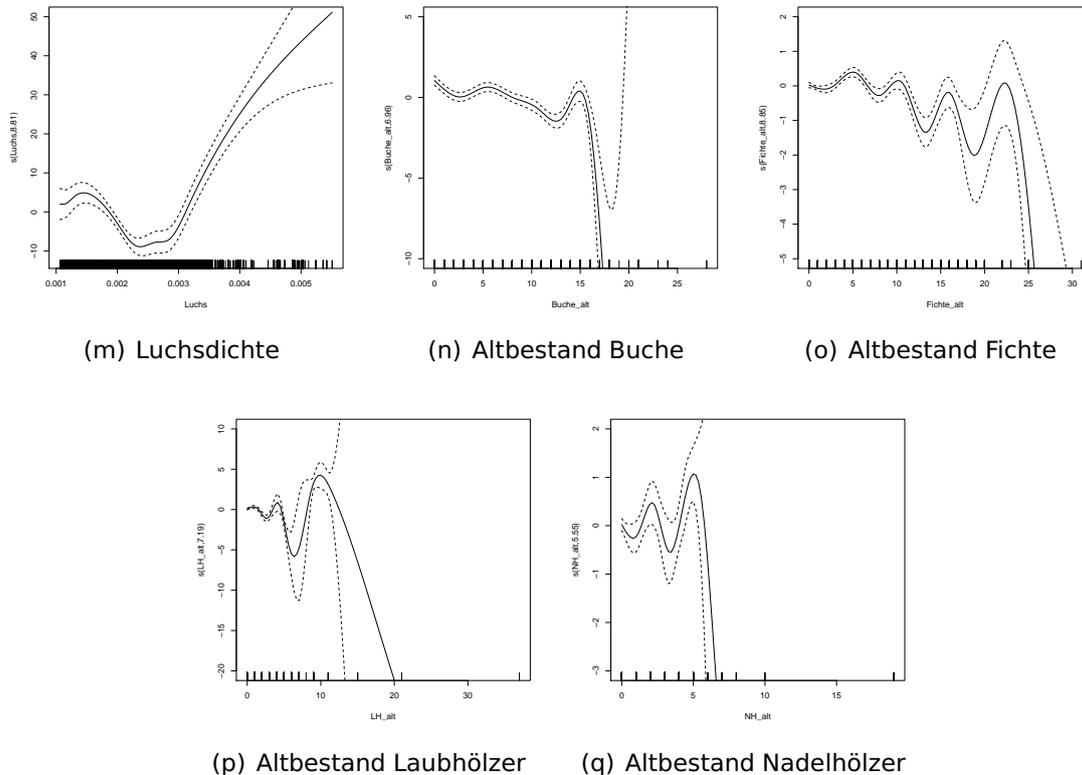


Abbildung 6.5.: Nonparametrische Einflussgrößen Binomialmodell Buche

Eine **Baumhöhe** kleiner als 30cm, bzw. größer als 150cm führt zu einer Verminderung der Chance auf Buchenverbiss. Für die dazwischenliegenden Baumhöhen resultiert eine Erhöhung der Verbisschance. Dabei ergeben sich „Verbisshochpunkte“ bei etwa 70cm und 120cm, was vermutlich durch die bevorzugte Verbisshöhe der Rehe und Rothirsche zu erklären ist.

Mit zunehmender **Verjüngungsdichte pro Hektar** verringert sich tendenziell die Chance auf Buchenverbiss. Ein lokaler „Tiefpunkt“ resultiert für eine Verjüngungsdichte von ca. 100 Bäumen.

Die **Höhe über NN** weist einen unruhigen Funktionsverlauf auf. Für sehr niedrig gelegene Regionen, sowie für eine Höhe über NN von ca. 950m und 1150m ergibt sich eine Verminderung der Chance auf Buchenverbiss.

Mit steigender **Entfernung von einem Wintergatter** vermindert sich die Chance auf Buchenverbiss, ab einer Entfernung von ca. 6000m beginnt die Chance wieder anzusteigen.

Für die **Entfernung von Straßen** und die **Entfernung von bejagten Flächen** gilt, dass mit zunehmender Entfernung ebenfalls die Chance auf Buchenverbiss zunimmt.

Nimmt die **Reh-** bzw. die **Rothirschdichte** zu, so erhöht sich die Chance auf

## 6. Prognostische Faktoren für Wildverbiss

Buchenverbiss.

Für die **Luchsdichte** hingegen ergibt sich zunächst ein Abfall der Verbisschance bis zu einer Luchsdichte von ca. 0.0025, anschließend steigt die Verbisschance stark an.

Der geschätzte Funktionsverlauf der Kovariablen **Neigung, Distanz von der Bestandesgrenze** und **Distanz von Wanderwegen** schwankt stark und sehr unruhig um die 0-Linie. Deshalb lässt sich der Einfluss dieser Einflussgrößen auf die Chance für Buchenverbiss nur schwer interpretieren. So ergeben sich beispielsweise für die Distanz von markierten Wanderwegen lokale „Verbisshochpunkte“ für Entfernungen von ca. 150m und für den Bereich von 200 bis 350m. Lokale „Verbisstiefpunkte“ hingegen resultieren für keine Distanz bzw. für Distanzen von ca. 170m und 450m.

Ein ähnliches Verhalten zeigen die Kovariablen **Buchenaltbestand, Fichtenaltbestand**, der **Altbestand an Laubhölzern** und der **Altbestand an Nadelhölzern**. Dabei ist zu beachten, dass für den Buchenaltbestand, sowie den Altbestand an Laub- und Nadelhölzern mit zunehmendem Altbestand die Konfidenzbänder sehr breit werden und somit eine Aussage über den weiteren Funktionsverlauf nicht möglich ist.

Nun folgt die Interpretation der parametrischen Einflussgrößen:

Erhöht sich der Deckungsgrad an **Nass- und Feuchtflächen** um eine Prozentklasse, so erhöht sich die Chance auf Buchenverbiss um das 1,47-fache, für **freiliegenden Boden** um das 1,12-fache, für **freiliegenden Fels** um das 1,08-fache und für **Him- und Brombeeren** um das 1,44-fache.

Erhöht sich andererseits der Deckungsgrad von **Fließgewässern** um eine Prozentklasse, so vermindert sich die Chance auf Buchenverbiss um das 0,4-fache, für **Farn** um das 0,64-fache, für **beerentragende Kleinsträucher** um das 0,86-fache und für **Moos** ebenfalls um das 0,86-fache.

Für eine südliche **Hangausrichtung** erhöht sich die Verbisschance um das 1,33-fache im Vergleich zur Referenzkategorie „nördliche Hangausrichtung“.

Befindet sich die Buche im Randbereich des Nationalparks, so erhöht sich die Verbisschance um das 1,26-fache im Vergleich zu Buchen, die nicht im Randbereich situiert sind. Die Chance auf Buchenverbiss vermindert sich um das 0,36-fache, wenn sich die Buche in einer ausgeräumten Totholzfläche befindet. Bei der kategorialen Variablen **Stadium\_neu** ergeben sich für die Ausprägungen „struktureich“, „strukturarm“ und „Totholz“ verminderte Buchenverbisschancen im Vergleich zur Referenzkategorie „offen“. Diese Verminderung beträgt für „strukturarm“ das 0,09-fache, für „struktureich“ das 0,16-fache und für „Totholz“ das 0,12-fache.

### 6.4.2. Poissonmodell

Dem Quasipoissonmodell der Buche liegen die Daten von 1701 Inventurpunkten zugrunde, die einen Buchenbestand aufweisen. Das Quasipoissonmodell mit anschließender Variablenselektion per Hand ist dem mittels dem Boosting-

## 6. Prognostische Faktoren für Wildverbiss

Algorithmus ausgewählten Modell gegenüber zu bevorzugen. Das Modell weist einen niedrigeren GCV-Score sowie einen höheren Prozentsatz der erklärenden Devianz auf.

Das resultierende Endmodell erklärt 66.6% der Devianz und weist mit einem geschätzten Dispersionsparameter von 1.44 eine leichte Überdispersion auf. Darüber hinaus ergeben sich die folgenden Parameterschätzungen:

```
Family: quasipoisson
Link function: log
Parametric coefficients:

```

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-3.48673	0.55491	-6.283	4.29e-10	***
nass	0.35986	0.06851	5.253	1.71e-07	***
fliessg	-0.74088	0.19816	-3.739	0.000192	***
farn	-0.24408	0.06494	-3.759	0.000177	***
him_brom	0.36052	0.06328	5.697	1.45e-08	***
moos	-0.14252	0.06318	-2.256	0.024224	*
expos_neuwarm	0.40179	0.13053	3.078	0.002119	**
STADIUM_neustrukturarm	-1.89006	0.26130	-7.233	7.37e-13	***
STADIUM_neustrukturreich	-1.09065	0.25808	-4.226	2.52e-05	***
STADIUM_neuTotholz	-1.56156	0.29326	-5.325	1.16e-07	***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Approximate significance of smooth terms:

```

	edf	Ref.df	F	p-value	
s(x,y)	28.719	28.719	10.186	< 2e-16	***
s(Verjüngung)	8.854	8.854	15.700	< 2e-16	***
s(h_nn)	8.167	8.167	9.321	7.70e-13	***
s(Neigung)	7.380	7.380	2.788	0.00592	**
s(BEST_DIST)	8.893	8.893	5.527	1.87e-07	***
s(Wgatr_Dist)	8.617	8.617	11.194	< 2e-16	***
s(Jagd_dist)	6.638	6.638	11.222	2.44e-13	***
s(Str_dist)	7.253	7.253	9.655	3.83e-12	***
s>Wege_dist)	8.855	8.855	10.747	4.16e-16	***
s(Reh)	7.694	7.694	5.925	2.72e-07	***
s(Rothirsch)	8.930	8.930	6.444	5.54e-09	***
s(Luchs)	7.208	7.208	5.877	6.78e-07	***
s(LH_alt)	6.639	6.639	1.538	0.15422	
s(Buche_alt)	6.934	6.934	11.859	1.12e-14	***
s(Fichte_alt)	7.267	7.267	1.465	0.17271	

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
R-sq.(adj) = 0.89   Deviance explained = 66.6%
GCV score = 1.3573  Scale est. = 1.2392   n = 1701
```

## Parameterinterpretation

Die graphische Darstellung der nonparametrischen Einflussgrößen ist in Abbildung 6.6 gegeben.

Eine **räumlich** bedingte Verminderung der Anzahl verbissener Buchen ergibt sich im östlichen Randbereich des Nationalparks, eine Erhöhung ist im südlichen Randbereich zu finden.

Der Funktionsverlauf der **Verjüngungsdichte** ist sehr unruhig. Tendenziell nimmt die erwartete Anzahl verbissener Buchen jedoch mit zunehmender Verjüngungsdichte ab. Lokale „Verbisstiefpunkte“ treten für eine Verjüngungsdichte von ca. 100 und ca. 200 Bäumen auf. Ab einer Verjüngungsdichte von 200 Bäumen nimmt die erwartete Anzahl jedoch wieder zu.

Tendenziell erhöht sich die erwartete Anzahl verbissener Buchen mit zunehmender **Höhe über NN**. Für eine Höhe von ca. 950m ergibt sich eine leichte Verminderung der Anzahl.

Mit steigender **Entfernung von bejagten Flächen** erhöht sich die Anzahl verbissener Buchen.

Bis zu einer **Distanz** von 2500m von **Straßen** sinkt die erwartete Anzahl verbissener Buchen leicht ab, ab 2500m Entfernung steigt diese erwartete Anzahl mit zunehmender Entfernung.

Ab einer **Luchsdichte** von ca. 0.0035 steigt die erwartete Anzahl verbissener Buchen mit zunehmender Luchsdichte. Eine so hohe Luchsdichte tritt jedoch nur in wenigen Inventurpunkten auf.

Für einen zunehmenden **Buchenaltbestand** vermindert sich die Chance auf Buchenverbiss leicht. Ab einem Altbestand von ca. 18 Buchen sind die Konfidenzbänder jedoch so breit, dass eine Aussage über den weiteren Funktionsverlauf nicht möglich ist.

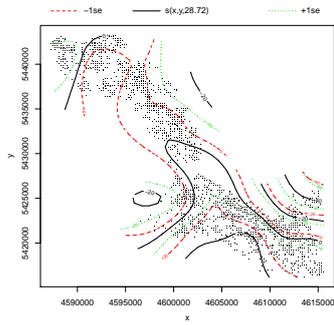
Der geschätzte Funktionsverlauf der Kovariablen **Höhe über NN**, **Neigung**, **Distanz von der Bestandesgrenze** und **Distanz von markierten Wanderwegen** ist sehr unruhig, schwankt um die 0-Linie und ist deshalb nur schwer interpretierbar. Beispielsweise ergeben sich für Distanzen von 50m, 80m und 120m von der Bestandesgrenze lokale „Verbisstiefpunkte“, für Entfernungen von ca. 70m, 100m und ab 150m resultiert eine erhöhte Anzahl verbissener Buchen.

Ein ähnliches Verhalten weisen die geschätzten Funktionsverläufe des **Fichtenaltbestands** und der **Altbestand an Laubhölzern** auf. Dabei werden die Konfidenzbänder mit zunehmendem Altbestand jedoch so breit, dass eine Aussage über den weiteren Funktionsverlauf nicht möglich ist.

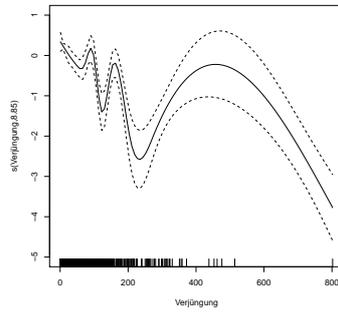
Die Konfidenzbänder für die Einflussgrößen **Distanz vom Wintergatter**, **Rehdichte** und **Rothirschdichte** sind so breit, dass sie die 0-Linie enthalten. Deshalb ist keine Aussage über den Einfluss dieser Variablen auf die erwartete Anzahl verbissener Buchen möglich.

Für die parametrischen Einflussgrößen ergeben sich folgende Parameterinterpretationen:

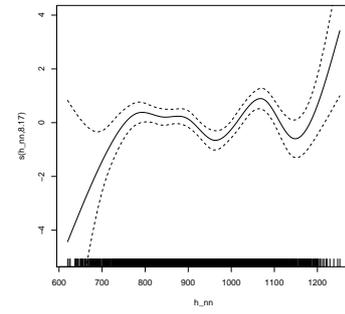
## 6. Prognostische Faktoren für Wildverbiss



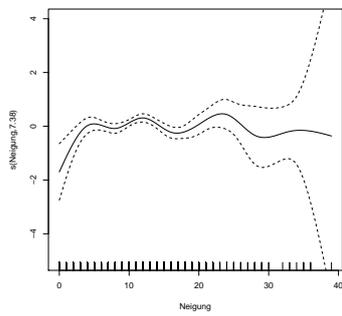
(a) Räumliche Komponente



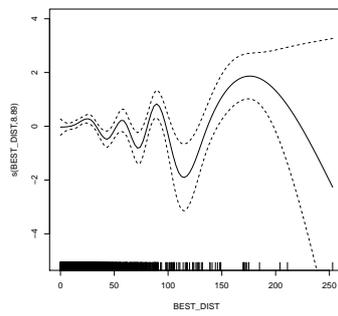
(b) Verjüngungsdichte



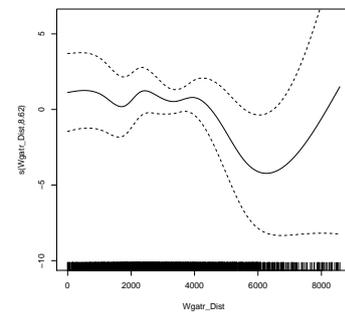
(c) Höhe über NN



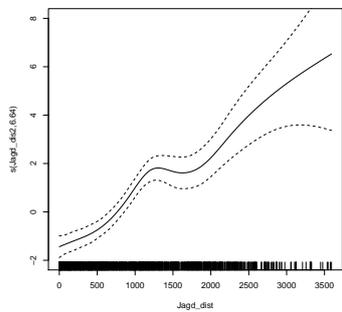
(d) Neigung



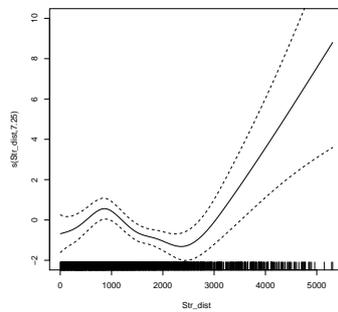
(e) Distanz Bestandesgrenze



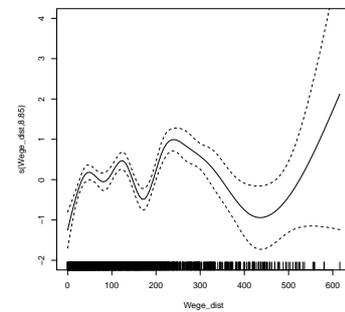
(f) Distanz Wintergatter



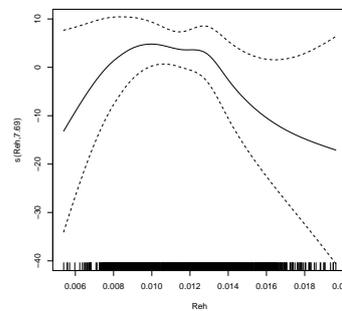
(g) Distanz Jagdflächen



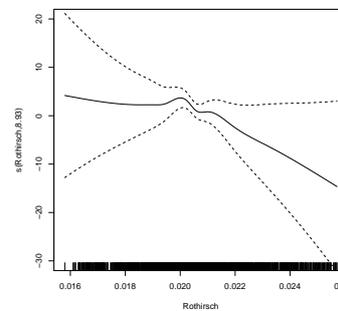
(h) Distanz Straßen



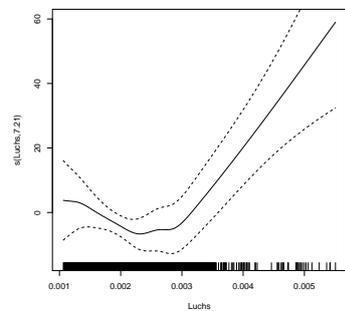
(i) Distanz Wanderwege



(j) Rehdichte



(k) Rothirschdichte



(l) Luchsdichte

## 6. Prognostische Faktoren für Wildverbiss

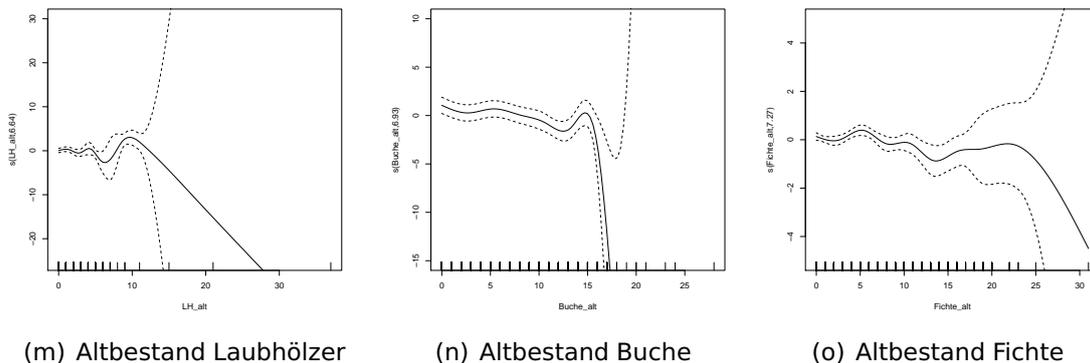


Abbildung 6.6.: Nonparametrische Einflussgrößen Poissonmodell Buche

Ein steigender Deckungsgrad von **Nass- und Feuchtflächen** sowie von **Him- und Brombeeren** bewirkt einen Anstieg der erwarteten Anzahl an verbissenen Buchen. Hingegen hat ein steigender Deckungsgrad von **Fließgewässern, Farn** und **Moos** eine Verminderung der erwarteten Anzahl zur Folge.

Für eine südliche **Hangausrichtung** resultiert eine Erhöhung der erwarteten Anzahl im Vergleich zu einer nördlichen Hangausrichtung.

Bei der kategorialen Variablen **Stadium\_neu** weisen die Ausprägungen „struktureich“, „strukturarm“ und „Totholz“ eine signifikante Verminderung der Anzahl verbissener Buchen im Vergleich zur Referenzkategorie „offen“ auf.

## 6.5. Bergahorn

Bei der Waldinventur wurde der Verbissstatus von 1071 Bergahornen erhoben. Des Weiteren liegen die umweltspezifischen und ökologischen Kovariablenausprägungen für 177 Inventurpunkte vor, die einen Bergahornbestand aufweisen.

### 6.5.1. Binomialmodell

Für das Quasibinomialmodell des Bergahorns ergibt sich das beste Endmodell mittels der Variablenselektion durch den Boosting-Algorithmus. Anschließend musste zusätzlich eine Variablenselektion per Hand durchgeführt werden, da nicht alle durch das Boosting ausgewählten Kovariablen zu einer Modellverbesserung beitrugen. Es resultieren die folgenden Parameterschätzungen:

```
Family: quasibinomial
Link function: logit
```

## 6. Prognostische Faktoren für Wildverbiss

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-8.017e+04	3.144e+05	-0.255	0.79881
Wgatr_Dist	2.133e-02	8.595e-03	2.481	0.01327 *
freife	-1.241e+02	2.511e+01	-4.942	9.09e-07 ***
fliesssg	-3.078e+02	5.326e+01	-5.780	1.00e-08 ***
him_brom	4.690e+01	1.170e+01	4.010	6.53e-05 ***
LH_alt	1.243e+01	4.054e+00	3.067	0.00222 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(height)	9.000	9.000	5.041	1.13e-06 ***
s(Verjüngung)	9.000	9.000	5.139	7.92e-07 ***
s(h_nn)	8.607	8.607	5.671	1.94e-07 ***
s(Neigung)	7.795	7.795	5.597	7.84e-07 ***
s>Wege_dist)	8.998	8.998	5.331	3.91e-07 ***
s(Jagd_dist)	8.194	8.194	5.370	9.70e-07 ***
s(Reh)	8.961	8.961	5.316	4.34e-07 ***
s(Rothirsch)	8.994	8.994	5.317	4.13e-07 ***
s(Buche_alt)	8.938	8.938	5.255	5.59e-07 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.557    Deviance explained = 59.3%  
GCV score = 0.51869    Scale est. = 0.4797    n = 1071

Das Endmodell weist einen Prozentsatz der erklärenden Devianz von 59.3% auf. Der geschätzte Dispersionsparameter von ca. 0.48 deutet auf eine Unterdispersion hin.

### Parameterinterpretation

Die Funktionsschätzungen für die nonparametrischen Einflussgrößen sind in Abbildung 6.7 dargestellt.

Für Bergahorne mit einer **Baumhöhe** im Bereich von 50cm ist die Chance auf Verbiss leicht erhöht. Für die anderen Baumhöhen ist keine Aussage möglich, da die 0-Linie enthalten ist.

Der Funktionsverlauf der **Verjüngungsdichte** pro Hektar ist sehr unruhig. Tendenziell vermindert sich die Chance auf Bergahornverbiss mit zunehmender Verjüngungsdichte. Für eine geringe Verjüngungsdichte ergibt sich eine erhöhte Chance auf Bergahornverbiss. Im speziellen resultiert für eine Verjüngungsdichte von ca. 70 Bäumen ein „Verbisstiefpunkt“.

Bis zu einer **Hangneigung** von 20° ist die Chance auf Bergahornverbiss erhöht. Ab 20° Hangneigung beginnt die Chance auf Verbiss stark abzufallen.

## 6. Prognostische Faktoren für Wildverbiss

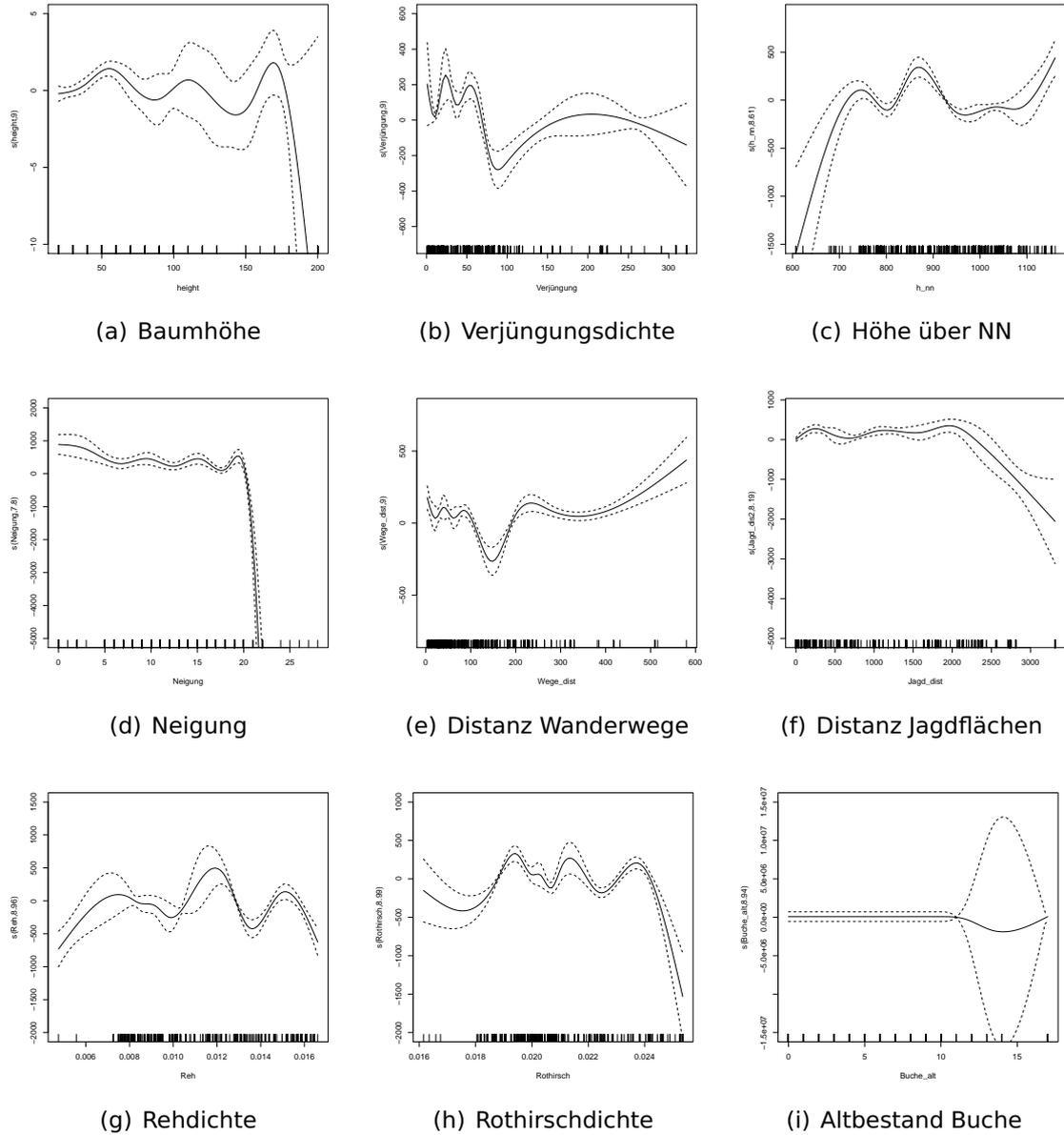


Abbildung 6.7.: Nonparametrische Einflussgrößen Binomialmodell Bergahorn

## 6. Prognostische Faktoren für Wildverbiss

Der Funktionsverlauf der Kovariablen **Höhe über NN**, **Distanz von markierten Wanderwegen** und **Distanz von bejagten Flächen** ist sehr unruhig und schwankt um die 0-Linie. Damit ist deren Einfluss auf die Chance für Bergahornverbiss nur schwer interpretierbar. Beispielsweise vermindert sich die Chance auf Bergahornverbiss für eine geringe Höhe über NN, für eine Höhe von etwa 850m ist die Verbisschance leicht erhöht. Bei der Distanz von Wanderwegen resultiert für eine Entfernung von ca. 150m eine verminderte Verbisschance.

Für den geschätzten Funktionsverlauf der **Reh-** und der **Rothirschdichte** ergibt sich ein ähnliches Ergebnis. Der Kurvenverlauf schwankt unregelmäßig um die 0-Linie und ist damit kaum sinnvoll interpretierbar.

Über den Zusammenhang des **Buchenaltbestands** mit der Chance auf Bergahornverbiss ist keine Aussage möglich, da die sehr breiten Konfidenzbänder die 0-Linie vollständig enthalten.

Für die parametrischen Einflussgrößen ergeben sich folgende Interpretationen: Erhöht sich die **Distanz vom Wintergatter** um 1m, so erhöht sich die Chance auf Bergahornverbiss um das 1,02-fache.

Die Erhöhung des Deckungsgrads an **freiliegendem Fels** bzw. an **Fließgewässern** bewirkt eine starke Verminderung der Chance auf Bergahornverbiss. Für freiliegenden Fels resultiert eine Verminderung um das  $1.27 \cdot 10^{-54}$ -fache, für Fließgewässer eine Verminderung um das  $2.1 \cdot 10^{-134}$ -fache.

Eine Erhöhung des Deckungsgrads von **Him- und Brombeeren** hat hingegen eine Vergrößerung der Chance auf Bergahornverbiss um das  $2.3 \cdot 10^{20}$ -fache zur Folge.

Wird der **Altbestand an Laubhölzern** um einen Laubbaum erhöht, so erhöht sich die Chance auf Bergahornverbiss um das 250196-fache.

### 6.5.2. Poissonmodell

Die erwartete Anzahl an verbissenen Bergahornen in den 177 Inventurpunkten mit Bergahornbestand wird am besten durch ein Quasipoissonmodell mit anschließender Variablenselektion per Hand erklärt. Dabei wird mit 81.6% ein sehr hoher Prozentsatz der erklärenden Devianz erzielt. Der geschätzte Dispersionsparameter weist mit 0.71 auf eine leichte Unterdispersion hin. In der folgenden Zusammenfassung sind die Parameterschätzungen sowie die selektierten Kovariablen dargestellt:

Family: quasipoisson

Link function: log

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-2.9449	0.3757	-7.838	1.87e-12	***
freife	-1.2859	0.2543	-5.056	1.52e-06	***
fliessg	-1.5556	0.5210	-2.986	0.00342	**

## 6. Prognostische Faktoren für Wildverbiss

```
moos          0.5571      0.1698    3.281  0.00135 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df      F  p-value
s(Verjüngung) 8.621  8.621  2.127 0.034185 *
s(h_nn)       5.143  5.143  3.094 0.010802 *
s(BEST_DIST)  8.937  8.937  3.423 0.000893 ***
s(Wgatr_Dist) 1.000  1.000  0.802 0.372323
s(Jagd_dist)  7.565  7.565  4.994 3.25e-05 ***
s(Neigung)    5.285  5.285  3.106 0.009894 **
s(Wege_dist)  5.987  5.987  3.213 0.005801 **
s(Rothirsch)  7.756  7.756  3.822 0.000555 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.986   Deviance explained = 81.6%
GCV score =  1.024   Scale est. = 0.70989   n = 177
```

### Parameterinterpretation

Eine Übersicht der nonparametrischen Einflussgrößen ist in Abbildung 6.8 gegeben.

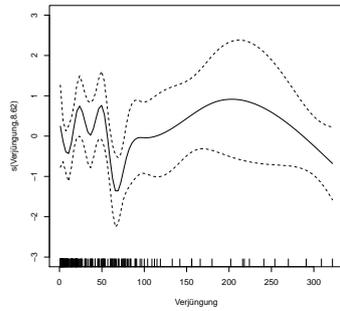
Bei der **Höhe über NN** ergibt sich eine erhöhte Anzahl verbissener Bergahorne im Bereich von 850m Höhe. Wegen der breiten Konfidenzbänder kann keine Aussage über den Funktionsverlauf in Regionen mit einer geringen Höhe über NN getroffen werden. Ab einer Höhe von ca. 900m über NN ist die 0-Linie in den Konfidenzbänder enthalten.

Für die **Distanz von markierten Wanderwegen** erhöht sich tendenziell mit zunehmender Distanz die Anzahl verbissener Bergahorne. Die breiten Konfidenzbänder enthalten jedoch weitgehend die 0-Linie.

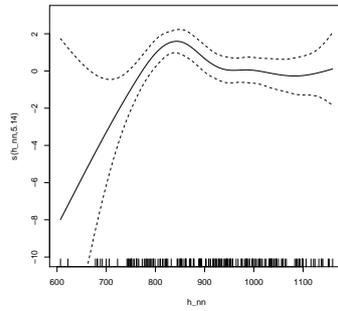
Die Funktionsverläufe der Kovariablen **Verjüngungsdichte**, **Distanz von der Bestandesgrenze** und **Distanz von bejagten Flächen** schwanken um die 0-Linie bzw. enthalten die 0-Linie über weite Strecken. Beispielsweise resultiert für den Bereich von eine Entfernung von etwa 1600m von bejagten Flächen eine Erhöhung der Anzahl verbissener Bergahorne. Für kleinere bzw. größere Entfernungen von bejagten Flächen ist die 0-Linie in den Konfidenzbändern enthalten. Eine **Rothirschdichte** von etwa 0.021 führt zu einer leichten Erhöhung der erwarteten Anzahl verbissener Bergahorne.

Die Konfidenzbänder der Einflussgrößen **Distanz vom Wintergatter** und **Neigung** enthalten fast komplett die 0-Linie. Somit ist keine Aussage über deren Einfluss auf die Zielgröße möglich. Obwohl der Funktionsverlauf der Distanz vom Wintergatter als linear geschätzt wurde, führt eine parametrische Aufnahme ins Modell zu einer signifikanten Modellverschlechterung.

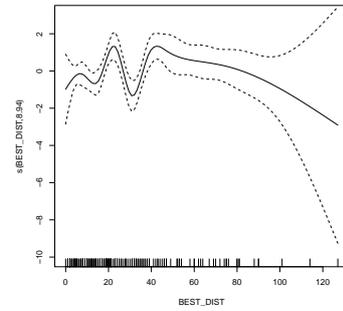
## 6. Prognostische Faktoren für Wildverbiss



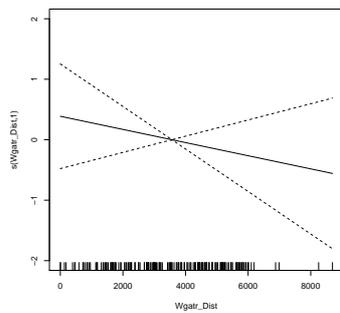
(a) Verjüngungsdichte



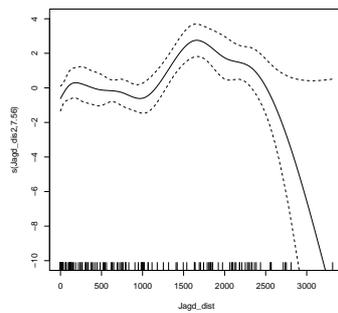
(b) Höhe über NN



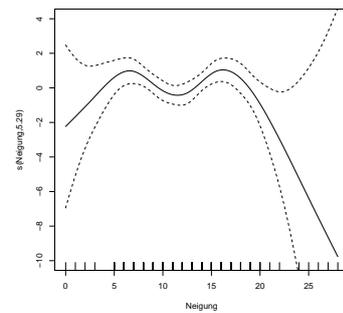
(c) Distanz Bestandesgrenze



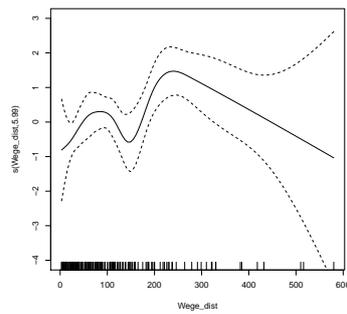
(d) Distanz Wintergatter



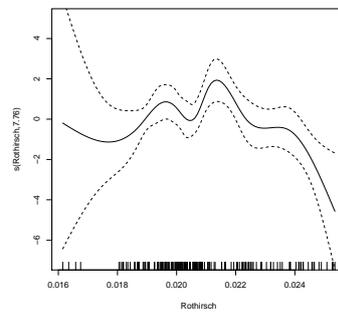
(e) Distanz Jagdflächen



(f) Neigung



(g) Distanz Wanderwege



(h) Rothirschdichte

Abbildung 6.8.: Nonparametrische Einflussgrößen Poissonmodell Bergahorn

## 6. Prognostische Faktoren für Wildverbiss

Bei den parametrischen Einflussgrößen bewirkt ein steigender Deckungsgrad an **Moos** eine Erhöhung der erwarteten Anzahl, ein steigender Deckungsgrad an **freiliegendem Fels** und **Fließgewässern** eine Verminderung der erwarteten Anzahl.

### 6.6. Fichte

Bei der Waldinventur 2002 wurde der Verbissstatus von 60183 Fichten erhoben, sowie die umweltspezifischen Kovariablen von 2534 Inventurpunkten mit Fichtenbestand.

#### 6.6.1. Binomialmodell

Für eine Anzahl von 60183 Fichten sind die Methoden zur Modellselektion nicht anwendbar. Analog zum Binomialmodell der Buche kann auch für das Binomialmodell der Fichte nur ein Anteil der erhobenen Fichten zur Modellanpassung verwendet werden. Die dazu verwendete Sampling-Methode wurde bereits für das Binomialmodell der Buche näher beschrieben. Bei der Auswahl der Fichten wurden aus jedem der Inventurpunkte mit Fichtenbestand zufällig 10 Fichten ausgewählt. Das führt zu einer Berücksichtigung von 15473 Fichten für die Modellanpassung. Als „Endmodell“ wurde das Quasibinomialmodell mit anschließender Variablenselektion per Hand ausgewählt.

Die folgende Zusammenfassung enthält die zugehörigen Parameterschätzungen:

Family: quasibinomial

Link function: logit

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-590.43954	116.85773	-5.053	4.41e-07	***
nass	0.60465	0.10120	5.975	2.36e-09	***
frei	-1.86903	0.18131	-10.309	< 2e-16	***
fliessg	4.66643	0.32786	14.233	< 2e-16	***
gras	0.80587	0.08889	9.066	< 2e-16	***
farn	-0.84133	0.13413	-6.272	3.65e-10	***
beer	-0.16381	0.05661	-2.894	0.00381	**
him_brom	-1.54294	0.16387	-9.416	< 2e-16	***
regionF-R	-62.88420	7.54936	-8.330	< 2e-16	***
NH_alt	-1.67770	0.26776	-6.266	3.81e-10	***
expos_neuwarm	3.72660	0.39252	9.494	< 2e-16	***
RANDBEREICH1	-4.32564	0.51025	-8.477	< 2e-16	***
STADIUM_neustrukturarm	-3.68625	0.70512	-5.228	1.74e-07	***
STADIUM_neustrukturreich	-8.17114	0.78168	-10.453	< 2e-16	***

## 6. Prognostische Faktoren für Wildverbiss

```
STADIUM_neuTotholz      -1.41277      0.70494     -2.004     0.04508 *
Tot_ausgerl              8.47396      0.63558     13.333     < 2e-16 ***
levelhanglage           -2.16659      0.42444     -5.105     3.35e-07 ***
levelhochlage           19.55837      1.42244     13.750     < 2e-16 ***
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

```
          edf Ref.df      F  p-value
s(x,y)      29.000 29.000 26.02 < 2e-16 ***
s(height)    7.994  7.994 63.43 < 2e-16 ***
s(Verjüngung) 6.984  6.984 39.27 < 2e-16 ***
s(h_nn)      9.000  9.000 30.92 < 2e-16 ***
s(Neigung)   7.000  7.000 45.57 < 2e-16 ***
s(BEST_DIST) 8.993  8.993 24.95 < 2e-16 ***
s(Wgatr_Dist) 9.000  9.000 32.33 < 2e-16 ***
s(Jagd_dist) 9.000  9.000 32.92 < 2e-16 ***
s(Str_dist)  9.000  9.000 30.53 < 2e-16 ***
s>Wege_dist) 9.000  9.000 35.39 < 2e-16 ***
s(Reh)       8.994  8.994 43.13 < 2e-16 ***
s(Rothirsch) 8.977  8.977 44.88 < 2e-16 ***
s(Luchs)     8.976  8.976 39.96 < 2e-16 ***
s(LH_alt)    3.122  3.122 55.32 < 2e-16 ***
s(Buche_alt) 4.688  4.688 10.73 8.34e-10 ***
s(Fichte_alt) 9.000  9.000 38.71 < 2e-16 ***
```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.328 Deviance explained = 59.8%

GCV score = 0.027462 Scale est. = 0.027166 n = 15473

Aufgrund der immer noch sehr hohen Fallzahl erweist sich eine ganze Menge von Einflussgrößen als signifikant. Das Modell besitzt einen Prozentsatz der erklärenden Devianz von 59.8% und der geschätzte Dispersionsparameter weist mit 0.03 auf eine starke Unterdispersion hin.

### Parameterinterpretation

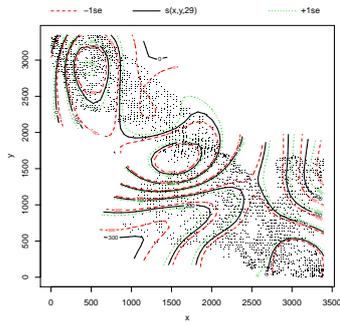
Die Funktionsschätzungen für die nonparametrischen Einflussgrößen sind in Abbildung 6.9 dargestellt.

Für den nördlichen Bereich des Nationalparks resultiert eine verminderte Chance für Fichtenverbiss, für den südlichen Teil hingegen eine Erhöhung der Chance für Fichtenverbiss.

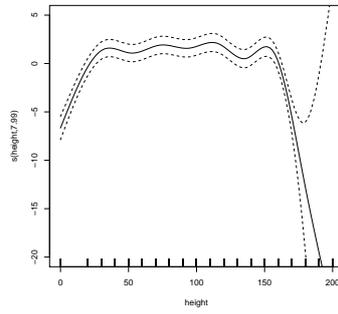
Bei der **Baumhöhe** ergeben sich für besonders kleine (<30cm) und besonders große (>150cm) Fichten eine verminderte Chance auf Verbiss.

Mit zunehmender **Distanz vom Wintergatter** vermindert sich die Chance auf

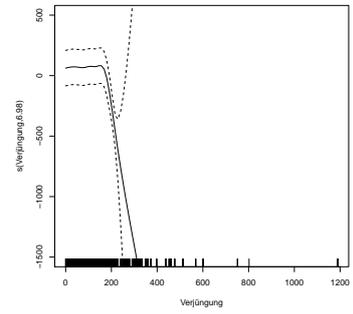
## 6. Prognostische Faktoren für Wildverbiss



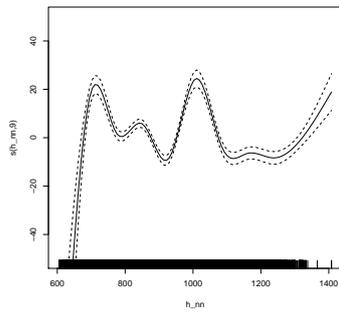
(a) Räumliche Komponente



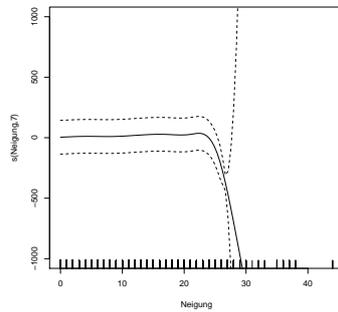
(b) Baumhöhe



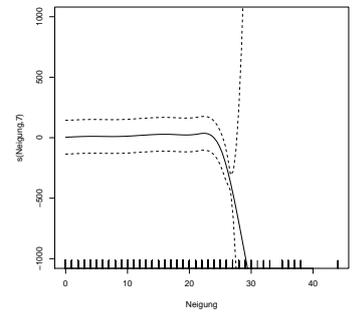
(c) Verjüngungsdichte



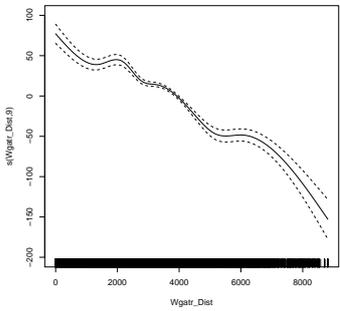
(d) Höhe über NN



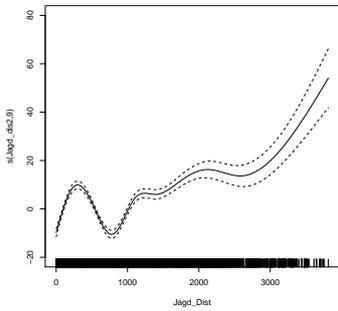
(e) Neigung



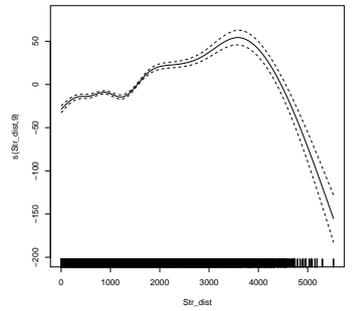
(f) Distanz Bestandesgrenze



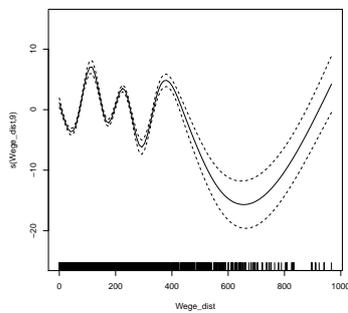
(g) Distanz Wintergatter



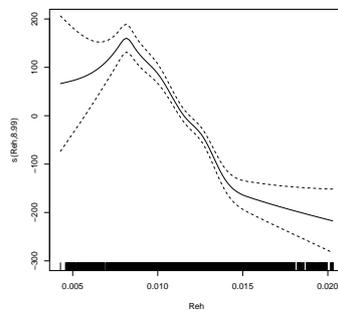
(h) Distanz bejagte Fläche



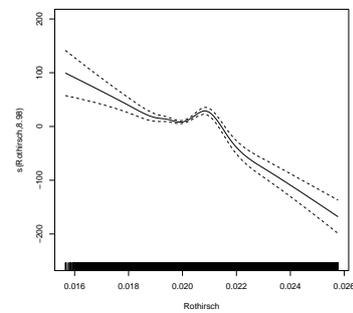
(i) Distanz Straße



(j) Distanz Wanderwege



(k) Rehdichte



(l) Rothirschdichte

## 6. Prognostische Faktoren für Wildverbiss

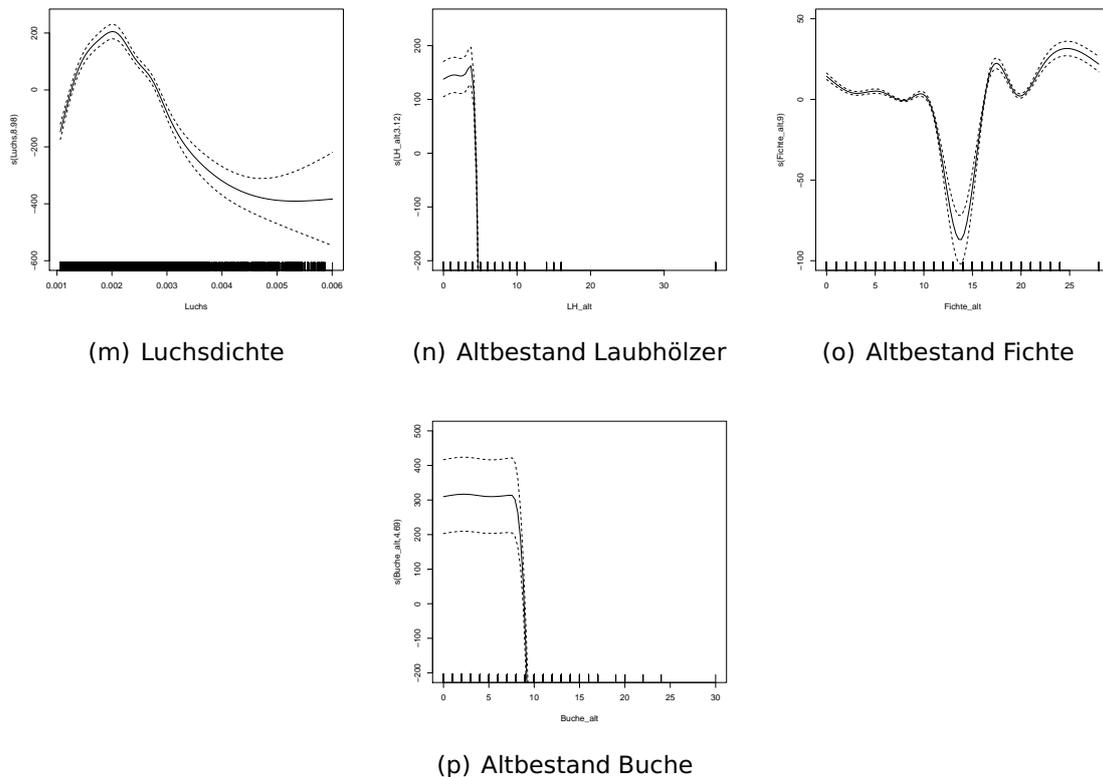


Abbildung 6.9.: Nonparametrische Einflussgrößen Binomialmodell Fichte

Fichtenverbiss.

Für eine zunehmende **Distanz von bejagten Flächen** erhöht sich tendenziell die Chance auf Fichtenverbiss. Für eine Entfernung von ca. 400m von bejagten Flächen ergibt sich ein „Verbisshochpunkt“.

Bis zu einer **Distanz** von 3500m von **Straßen** erhöht sich die Chance auf Fichtenverbiss mit zunehmender Entfernung. Ab der Distanz von 3500m sinkt die Chance auf Fichtenverbiss stark ab.

Für die **Reh-** und die **Rothirschdichte** gilt, dass mit zunehmender Dichte die Anzahl verbissener Fichten abnimmt. Dieses untypische Verhalten könnte dadurch erklärt werden, dass die Fichte generell selten verbissen wird. Bis zu einer **Luchsdichte** von 0.002 steigt die Chance auf Fichtenverbiss an, eine weitere Erhöhung der Luchsdichte führt jedoch zu einer Verminderung der Verbisschance.

Die **Höhe über NN** weist einen sehr unruhigen Verlauf auf, jedoch ergibt sich für Regionen mit einer geringen Höhe über NN eine verminderte Chance auf Fichtenverbiss. Lokale „Verbisshochpunkte“ resultieren für eine Höhe von 700m und 100m über NN.

Der Funktionsverlauf der **Distanz von Wanderwegen** ist sehr unruhig und

## 6. Prognostische Faktoren für Wildverbiss

schwankt stark um die 0-Linie. Deshalb ist der Einfluss dieser Variablen auf die Verbisschance nur schwer interpretierbar. Lokale „Verbisshochpunkte“ ergeben sich für Entfernungen von 100m, 250m und 400m, lokale „Verbisstiefpunkte“ für Distanzen von 50m, 170m, 280m und 600m.

Die Konfidenzbänder der Einflussgrößen **Verjüngungsdichte**, **Neigung** und **Distanz von der Bestandesgrenze** enthalten weitgehend die 0-Linie. Eine Aussage über den Einfluss dieser Kovariablen auf die Chance für Verbiss ist folglich nicht möglich.

Die Funktionsschätzungen des **Altbestands an Laubhölzern** und des **Buchenaltbestands** weisen eine sehr ähnliche Gestalt auf. Für einen geringen Altbestand ist die Chance auf Fichtenverbiss erhöht, ab einem Grenzwert beginnt die Chance steil abzufallen. Die unruhige Funktionsschätzung des **Fichtenaltbestands** weist eine deutliche Verminderung der Chance auf Fichtenverbiss für einen Altbestand von ca. 14 Fichten auf.

Für die parametrischen Einflussgrößen ergeben sich folgende Interpretationen: Erhöht sich der Deckungsgrad von **Nass- und Feuchtflächen** um eine Prozentklasse so erhöht sich die Chance auf Fichtenverbiss um das 1,83-fache, für **Fließgewässer** um das 106-fache, für **Gras** um das 2,24-fache und für eine südliche **Hangausrichtung** um das 42-fache.

Erhöht sich hingegen der Deckungsgrad von **freiliegender Boden** um eine Prozentklasse, so vermindert sich die Chance auf Fichtenverbiss um das 0,15-fache, für **Farn** um das 0,43-fache, für **beerentragende Kleinsträucher** um das 0,85-fache und für **Him- und Brombeeren** um das 0,21-fache. Eine Verminderung der Chance auf Fichtenverbiss ergibt sich des Weiteren für eine Lage im **Randbereich** (0,013-fach) und einer Lage im **Rachel-Lusen-Gebiet** ( $4.894845e-28$ -fach), eine Erhöhung der Verbisschance resultiert für die Lage in einer **ausgeräumten Totholzfläche** (4788-fach). Nimmt der **Nadelholzaltbestand** um einen Baum zu, so vermindert sich die Chance auf Fichtenverbiss um das 0,19-fache. Für die kategoriale Variable **Stadium\_neu** ergeben sich für die Ausprägungen „Totholz“ (0,24-fach), „struktureich“ (0,0002-fach) und „strukturarm“ (0,025-fach) verminderte Verbisschancen im Vergleich zur Referenzkategorie „offen“. Die Ausprägung „Hanglage“ der Kovariablen **level** führt zu einer Verminderung der Verbisschance um das 0,11-fache im Vergleich zur Referenzkategorie „Aufichte“, die Ausprägung „Hochlage“ zu einer Erhöhung der Verbisschance um das  $3 \cdot 10^8$ -fache im Vergleich zur Referenzkategorie.

### 6.6.2. Poissonmodell

Zur Modellierung der erwarteten Anzahl verbissener Fichten eignet sich ein Quasipoissonmodell mit anschließender Variablenselektion per Hand am besten. Der Prozentsatz der erklärenden Devianz ist mit 79.9% sehr hoch und der geschätzte Dispersionsparameter weist mit einem Wert von 0.13 auf eine starke Unterdispersion hin. Die Parameterschätzungen sind in der folgenden Zusammenfassung aufgelistet:

## 6. Prognostische Faktoren für Wildverbiss

Family: quasipoisson  
Link function: log

Parametric coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-208.72597	43.32596	-4.818	1.54e-06	***
freife	0.59514	0.07960	7.476	1.07e-13	***
fliessg	-0.13358	0.19588	-0.682	0.49533	
gras	0.26422	0.03677	7.185	8.93e-13	***
farn	-0.39150	0.07445	-5.259	1.58e-07	***
him_brom	-0.37340	0.07463	-5.003	6.04e-07	***
moos	0.10744	0.04487	2.394	0.01673	*
expos_neuwarm	0.76895	0.16041	4.794	1.74e-06	***
RANDBEREICH1	-2.06667	0.20990	-9.846	< 2e-16	***
STADIUM_neustrukturarm	-1.26272	0.43036	-2.934	0.00338	**
STADIUM_neustrukturreich	-0.80016	0.40544	-1.974	0.04855	*
STADIUM_neuTotholz	0.17213	0.44272	0.389	0.69746	
Tot_ausger1	1.28937	0.18794	6.861	8.72e-12	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value	
s(x,y)	29.000	29.000	20.437	<2e-16	***
s(Verjüngung)	8.781	8.781	26.337	<2e-16	***
s(h_nn)	8.999	8.999	20.498	<2e-16	***
s(Neigung)	6.988	6.988	30.947	<2e-16	***
s(BEST_DIST)	8.946	8.946	17.467	<2e-16	***
s(Wgatr_Dist)	8.660	8.660	15.802	<2e-16	***
s(Jagd_dist)	8.940	8.940	12.215	<2e-16	***
s(Str_dist)	8.322	8.322	20.889	<2e-16	***
s>Wege_dist)	9.000	9.000	27.155	<2e-16	***
s(NH_alt)	4.842	4.842	0.066	0.996	
s(Reh)	9.000	9.000	32.922	<2e-16	***
s(Rothirsch)	9.000	9.000	32.458	<2e-16	***
s(Luchs)	8.630	8.630	30.995	<2e-16	***
s(LH_alt)	3.766	3.766	23.149	<2e-16	***
s(Buche_alt)	3.930	3.930	1.025	0.392	
s(Fichte_alt)	8.549	8.549	35.884	<2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.885    Deviance explained = 79.9%  
GCV score = 0.13948    Scale est. = 0.13077    n = 2534

## Parameterinterpretation

Die geschätzten Funktionsverläufe der nonparametrischen Einflussgrößen sind in Abbildung 6.10 dargestellt.

Die **räumliche Komponente** ergibt eine erhöhte Anzahl verbissener Fichten für den östlichen Randbereich des Nationalparks, für den westlichen Randbereich hingegen ergibt sich eine verminderte Anzahl verbissener Fichten.

Der Funktionsverlauf der **Verjüngungsdichte** pro Hektar ist unruhig. Bis zu einer Verjüngungsdichte von 150 Bäumen pro Hektar erhöht sich die Anzahl verbissener Fichten. Für eine Verjüngungsdichte von ca. 200 Bäumen ergibt sich ein lokaler „Verbisstiefpunkt“. Anschließend steigt die Anzahl verbissener Fichten bis zu einer Verjüngungsdichte von 500 Bäumen wieder an. Die folgende Abnahme der Anzahl verbissener Fichten basiert nur noch auf wenigen Beobachtungen.

Die **Höhe über NN** weist ebenfalls einen unregelmäßigen Funktionsverlauf auf. Im Besonderen ergibt sich für Regionen mit einer niedrigen Höhe über NN eine verminderte Anzahl verbissener Fichten. „Verbisshochpunkte“ hingegen resultieren für Höhen von ca. 750m und 1300m über NN.

Eine zunehmende **Distanz von Straßen** führt zu einer Erhöhung der Anzahl verbissener Fichten.

Die Anzahl verbissener Fichten vermindert sich tendenziell mit zunehmender **Rehdichte**, mit steigender **Rothirschdichte** hingegen erhöht sich die Anzahl verbissener Fichten. Bis zu einer **Luchsdichte** von ca. 0.003 steigt die Anzahl verbissener Fichten. Eine weiter ansteigende Luchsdichte führt zu einer Verminderung der Anzahl verbissener Fichten.

Ein niedriger **Laubholzaltbestand** hat keinen Einfluss auf die Anzahl verbissener Fichten, ab einem Albestand von ca. 8 Bäumen vermindert sich die erwartete Anzahl.

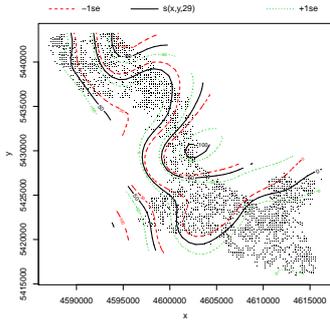
Eine niedrige **Hangneigung** hat ebenfalls keinen Einfluss auf die erwartete Anzahl verbissener Fichten. Einer Hangneigung von mehr als 28° führt zu einer Verminderung der erwarteten Anzahl.

Die Funktionsverläufe der Einflussgrößen **Distanz von der Bestandesgrenze** und **Distanz von bejagten Flächen** schwanken um die 0-Linie. Für eine Entfernung von der Bestandesgrenze von etwa 150m resultiert eine Verminderung der erwarteten Anzahl verbissener Fichten, für eine Distanz von ca. 200m ist die erwartete Anzahl erhöht. Bei der Entfernung von Jagdfächen führt keine Entfernung zu einer leichten Erhöhung der erwarteten Anzahl, eine Entfernung von mehr als 3000m bewirkt hingegen eine verminderte erwartete Anzahl.

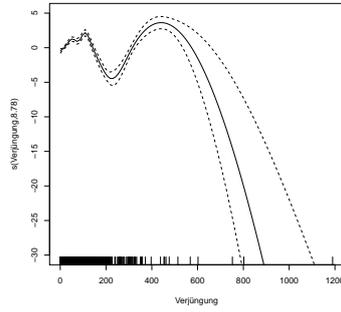
Die **Distanz von markierten Wanderwegen** und der **Fichtenaltbestand** weisen einen sinusförmigen, unruhigen Verlauf auf. Deshalb ist der Einfluss dieser Variablen auf die erwartete Anzahl nur schwer sinnvoll interpretierbar.

Die Konfidenzbänder der **Distanz vom Wintergatter**, des **Nadelholzaltbestands** und des **Buchenaltbestands** beinhalten die 0-Linie. Damit ist eine Aussage über den Zusammenhang dieser Kovariablen mit der Zielgröße nicht

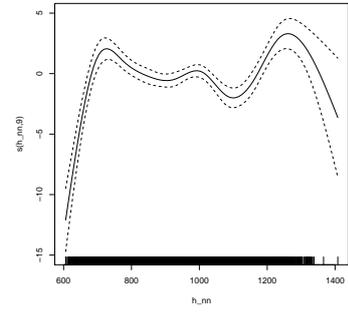
## 6. Prognostische Faktoren für Wildverbiss



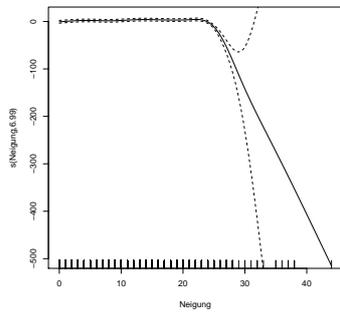
(a) Räumliche Komponente



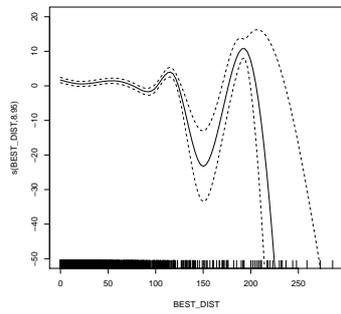
(b) Verjüngungsdichte



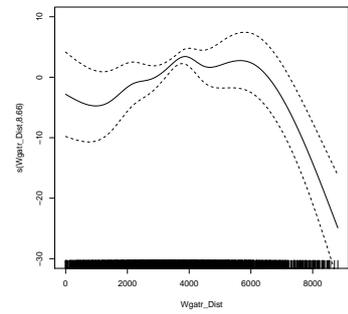
(c) Höhe über NN



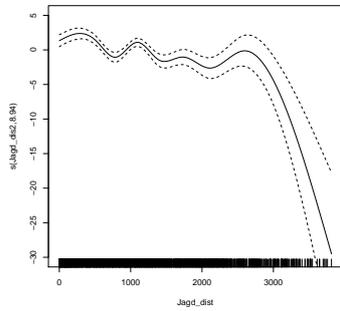
(d) Neigung



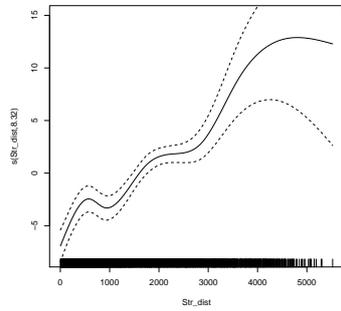
(e) Distanz Bestandesfläche



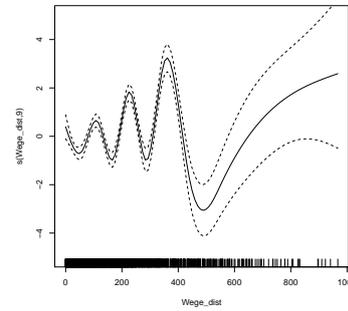
(f) Distanz Wintergatter



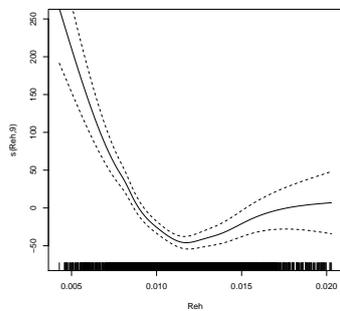
(g) Distanz bejagte Flächen



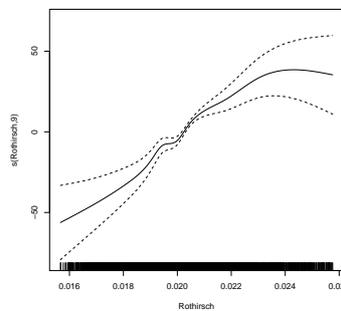
(h) Distanz Straße



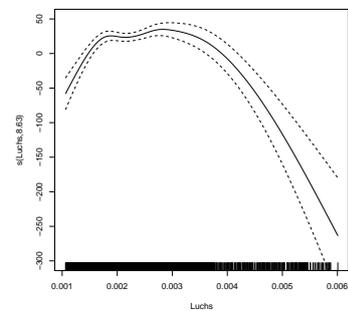
(i) Distanz Wanderwege



(j) Rehdichte

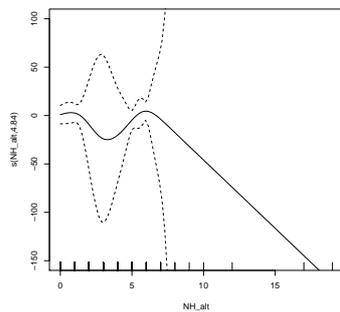


(k) Rothirschdichte

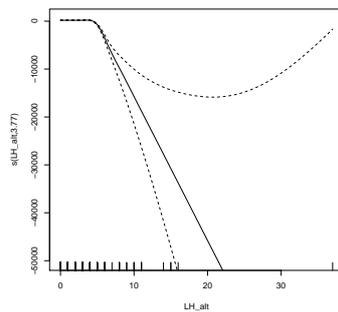


(l) Luchsdichte

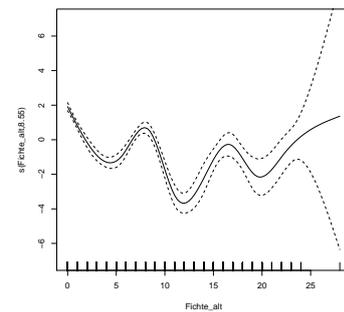
## 6. Prognostische Faktoren für Wildverbiss



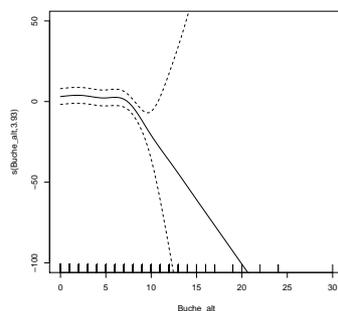
(m) Altbestand Nadelhölzer



(n) Altbestand Laubhölzer



(o) Altbestand Fichte



(p) Altbestand Buche

Abbildung 6.10.: Nonparametrische Einflussgrößen Poissonmodell Fichte

möglich.

Die Parameterinterpretationen für die parametrischen Einflussgrößen ergeben sich wie folgt:

Die Erhöhung des Deckungsgrads von **freiliegendem Fels**, **Gras** und **Moos**, sowie eine südliche **Hangausrichtung** und eine Lage in einer **ausgeräumten Totholzfläche** führen zu einer erhöhten Anzahl verbissener Fichten. Hingegen bewirkt ein erhöhter Deckungsgrad an **Fließgewässern**, **Farn** und **Him- und Brombeeren**, sowie eine Lage im **Randbereich** eine Verminderung der Anzahl verbissener Fichten. Bei der kategorialen Variablen **Stadium\_neu** vermindern die Ausprägungen „strukturarm“ und „struktureich“ die Anzahl verbissener Fichten im Vergleich zur Referenzkategorie „offen“. Die Ausprägung „Totholz“ erhöht die Anzahl, dieser Unterschied ist aber nicht signifikant.

## 7. Zusammenfassung

Ziel dieser Arbeit war die Identifikation prognostischer Faktoren für den Wildverbiss im Nationalpark Bayerischer Wald. Hierfür wurden getrennt für die fünf Baumarten Vogelbeere, Tanne, Buche, Bergahorn und Fichte jeweils ein logistisches Regressionsmodell zur Modellierung der *Chance auf Verbiss*, sowie ein Poissonmodell zur Modellierung der *erwarteten Anzahl* verbissener Bäume angepasst. Dabei erlaubte die Anpassung eines Generalisierten Additiven Modells die Einbeziehung nonparametrischer Einflussgrößen. Des Weiteren erwies sich eine Modellierung basierend auf einem Quasi-Likelihood-Ansatz als sinnvoll, da sowohl Über- als auch Unterdispersionsprobleme auftraten.

In den meisten Fällen lieferte die Modellierung eines Quasibinomial- bzw. Quasipoissonmodells mit anschließender Variablenselektion per Hand eine bessere Anpassung im Vergleich zu den Modellen, die aus dem Boosting-Verfahren resultierten.

Im Folgenden soll überprüft werden, ob sich für einige der Einflussgrößen eine einheitliche Wirkungsweise für alle fünf Baumarten ergab. Bei den parametrischen Einflussgrößen hatte ein erhöhter Deckungsgrad von Gras und ein erhöhter Deckungsgrad an Feucht- und Nassflächen einen erhöhten Baumverbiss zur Folge. Ein erhöhter Deckungsgrad an Fließgewässern hingegen bewirkte fast ausschließlich eine Verminderung des Baumverbisses. Für die Modelle mit signifikantem Einfluss der Hangausrichtung, ergab eine südliche Hangausrichtung eine erhöhte Verbissrate. Möglicherweise bevorzugen die Tiere den Aufenthalt in wärmeren bzw. sonnigeren Regionen. Die binäre Variable „region“ hatte lediglich in einem der Modelle einen signifikanten Einfluss.

Auch bei den nonparametrischen Einflussgrößen konnten ähnliche Funktionsverläufe für alle fünf Baumarten beobachtet werden. Bei der Baumhöhe, die nur in den logistischen Regressionsmodellen berücksichtigt werden konnte, zeigte sich, dass besonders kleine, sowie besonders große Bäume seltener verbissen werden. Außerdem ergaben sich zwei „Verbisshochpunkte“, die sich jeweils vermutlich durch die bevorzugten Verbisshöhen der Rehe und Rothirsche ergeben. Regionen, die eine geringe Höhe über NN aufweisen, weisen ebenfalls einen verminderten Wildverbiss auf. Intuitiv würde man einen Anstieg des Wildverbisses mit zunehmender Entfernung von Straßen, markierten Wanderwegen und bejagten Flächen vermuten. Diese Tendenz konnte in vielen, jedoch nicht in allen Modellen beobachtet werden. Im Besonderen die Distanz von markierten Wanderwegen und die Distanz von der Bestandesgrenze weisen häufig einen sehr unruhigen sinusförmigen Verlauf auf, der nur schwer interpretierbar ist. Eine weitere Vermutung wäre die Zunahme des Wildverbisses mit zunehmender

## 7. Zusammenfassung

Rothirsch-, Reh- und Luchsdichte. Dieser Verlauf konnte in einigen Modellen ermittelt werden, es ergaben sich aber auch genau entgegengesetzte bzw. sehr unruhige und damit nur schwer interpretierbare Verläufe.

Ein Vergleich der beiden Modelle für jeweils eine Baumart liefert die folgenden Erkenntnisse:

Bei den parametrischen Einflussgrößen der **Vogelbeere** ergeben sich nahezu identische Resultate. Zunächst werden exakt die selben Kovariablen selektiert, insoweit diese in beiden Datensätzen vorhanden sind. Des Weiteren weisen die ausgewählten Kovariablen identische Vorzeichen auf und beeinflussen demnach die Zielgröße in die selbe Richtung. Im Binomialmodell der Vogelbeere wird eine größere Anzahl an nonparametrischen Einflussgrößen berücksichtigt als im Poissonmodell. Eine ähnliche Form weisen die Funktionenschätzungen der Höhe über NN, Distanz vom Wintergatter und der Rehdichte auf. Der Einfluss der Variablen Verjüngungsdichte, Luchsdichte und der Buchenaltbestand unterscheiden sich hingegen.

Das Poissonmodell der **Tanne** selektiert weniger parametrische Einflussgrößen als das Binomialmodell. Wird aber eine Kovariable von beiden Modellen ausgewählt, dann stimmen die Vorzeichen überein, sie beeinflussen die Zielgröße also in die gleiche Richtung. Der glatte Verlauf der Höhe über NN, der Distanz von Straßen, der Distanz von der Bestandesgrenze und des Fichtenaltbestands weisen in beiden Modellen einen ähnlichen Verlauf auf.

Beim Vergleich der Modelle für die **Buche** ergibt sich ein richtungsmäßig gleicher Einfluss für die parametrischen Kovariablen Deckungsgrad an Nass- und Feuchtflächen, Farn, Fließgewässern, Him- und Brombeeren, sowie einer südlichen Hangausrichtung. Bei den nonparametrischen Einflussgrößen ergibt sich ein ähnlicher Verlauf für Verjüngung, Höhe über NN, Distanz von der Bestandesgrenze, Distanz von bejagten Flächen, Distanz von Straßen, Distanz von Wanderwegen und der Luchsdichte.

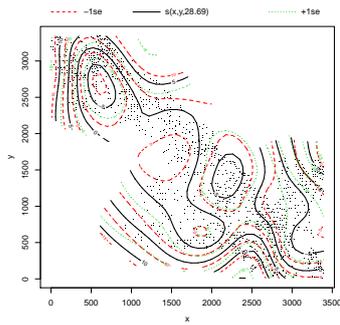
Für beiden Modelle des **Bergahorns** werden nur wenige parametrische Einflussgrößen ausgewählt. Die parametrischen Kovariablen Deckungsgrad an freiliegendem Fels und Fließgewässern, die in beiden Modellen ausgewählt werden, führen in beiden Modellen zu einer Verminderung des Wildverbisses. Ein ähnlicher Kurvenverlauf ergibt sich lediglich für die glatte Kovariable Distanz von Jagdflächen. In beiden Modellen für die **Fichte** ergaben sich aufgrund der hohen Beobachtungszahl eine große Anzahl signifikanter Einflussgrößen. Bei den parametrischen Einflussgrößen resultierte in vielen Fällen eine richtungsmäßig gleiche Beeinflussung der Zielgröße. Lediglich für den Deckungsgrad an Fließgewässern variiert das Vorzeichen. Bei den glatten Einflussgrößen weisen die Funktionsschätzungen für die Rehdichte, die Luchsdichte und die Distanzen von markierten Wanderwegen und vom Wintergatter eine ähnliche Gestalt auf.

Somit ergeben sich für die beiden Modellansätze der Baumarten relativ einheitliche Resultate für den Einfluss der prognostischen Faktoren. Für einige Faktoren, beispielsweise die Baumhöhe, konnte ein Funktionsverlauf geschätzt werden, der eine ökologisch sinnvolle Interpretation ermöglicht.

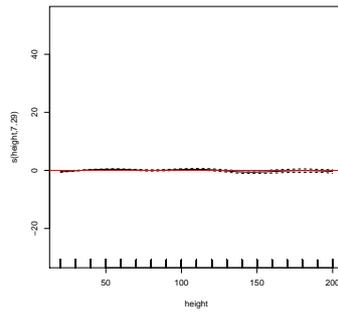
# Literaturverzeichnis

- [McCullagh/Nelder 1989] McCullagh and J.A. Nelder, 1989, *Generalized Linear Models*: CHAPMAN&HALL/CRC
- [Fahrmeir et al 2007] Ludwig Fahrmeir, Thomas Kneib & Stefan Lang, 2007, *Regression-Modelle, Methoden und Anwendungen*: Springer Verlag
- [Hastie/Tibshirani 1990] T.J. Hastie and R.J. Tibshirani, 1990, *Generalized Additive Models*: CHAPMAN&HALL/CRC
- [Green/Silverman 1994] P.J. Green and B.W. Silverman, 1994, *Nonparametric Regression and Generalized Linear Models – A roughness penalty approach*: CHAPMAN&HALL/CRC
- [Wood 2006] Simon N. Wood, 2006, *Generalized Additive Models – An Introduction with R*: CHAPMAN&HALL/CRC
- [Bühlmann/Hothorn 2008] P. Bühlmann and T. Hothorn, 2008, *Boosting algorithms: regularization, prediction and model fitting*: Statistical Science, accepted
- [Schmid/Hothorn 2007] M. Schmid and T. Hothorn, 2007, *Boosting Additive Models using Component-wise P-Splines*: Technical Report Number 002,2007 Department of Statistics, University of Munich
- [R-Hilfeseite zu **gamboost()**] *R-Hilfeseite zu **gamboost()***; R-Version R 2.9.0
- [R-Hilfeseite zu **gam()**] *R-Hilfeseite zu **gam()***; R-Version R 2.9.0

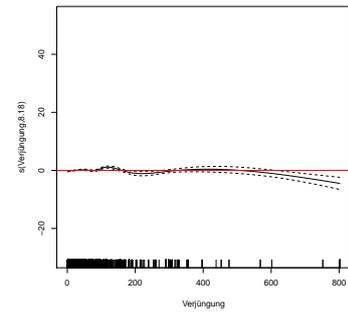
# A. Graphiken Vogelbeere



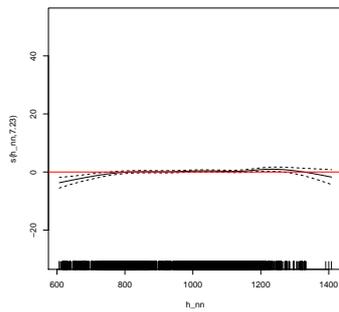
(a) Räumliche Komponente



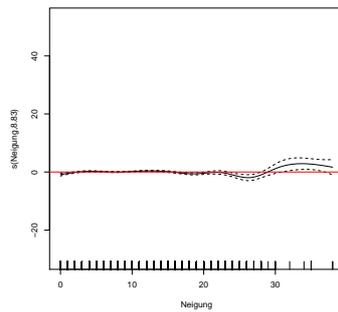
(b) Baumhöhe



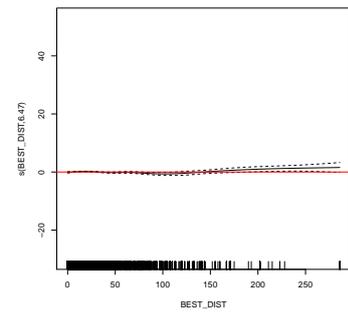
(c) Verjüngungsdichte



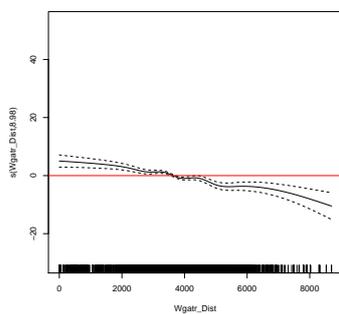
(d) Höhe über NN



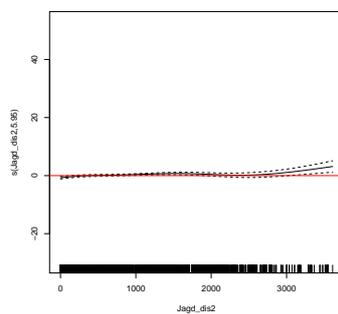
(e) Neigung



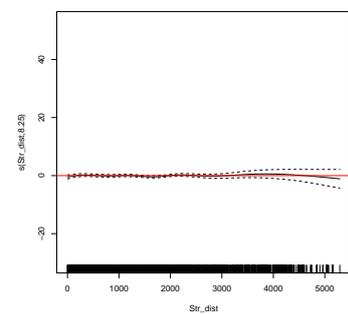
(f) Distanz Bestandesgrenze



(g) Distanz Wintergatter

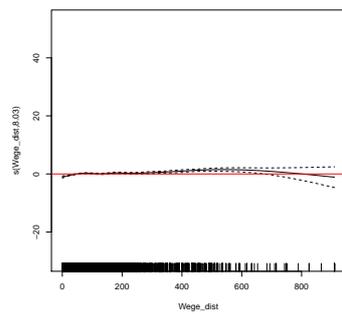


(h) Distanz Jagdfläche

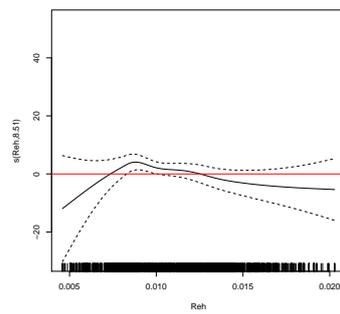


(i) Distanz Straße

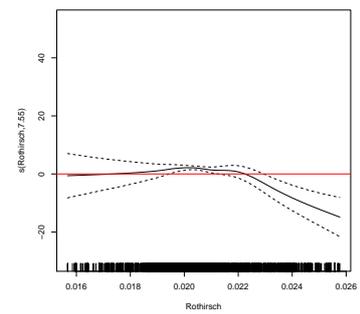
## A. Graphiken Vogelbeere



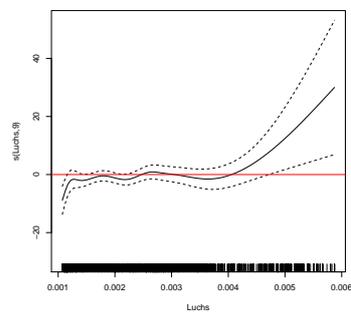
(j) Distanz Wanderwege



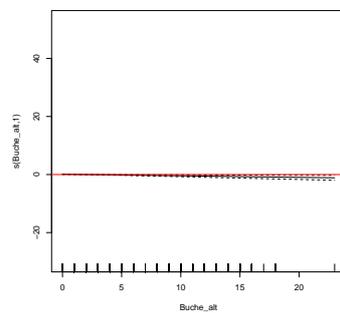
(k) Rehdichte



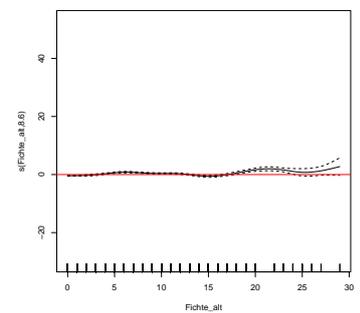
(l) Rothirschkichte



(m) Luchsdichte



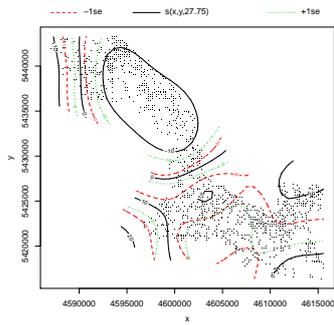
(n) Altbestand Buche



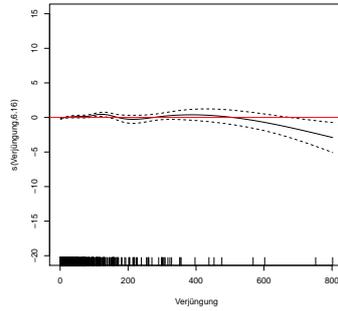
(o) Altbestand Fichte

Abbildung A.1.: Nonparametrische Einflussgrößen Binomialmodell Vogelbeere

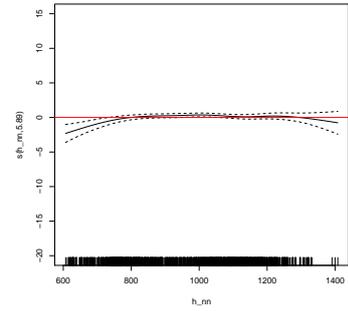
# A. Graphiken Vogelbeere



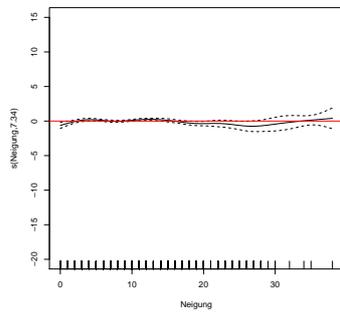
(a) Räumliche Komponente



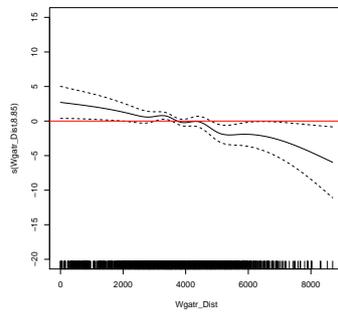
(b) Verjüngungsdichte



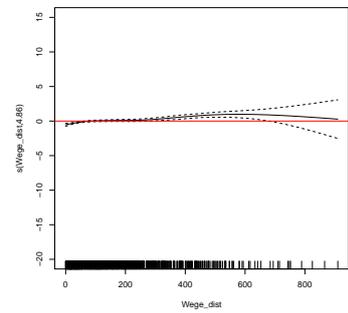
(c) Höhe über NN



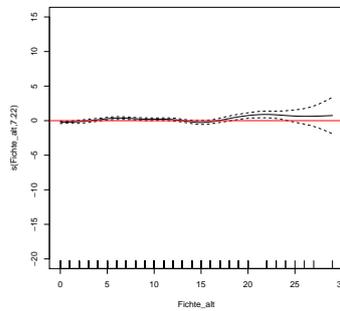
(d) Neigung



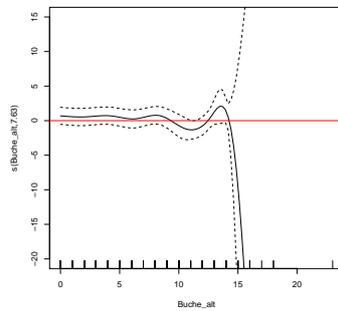
(e) Distanz Wintergatter



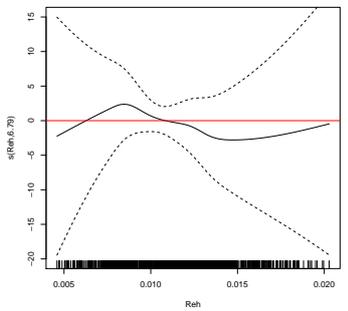
(f) Distanz Wanderwege



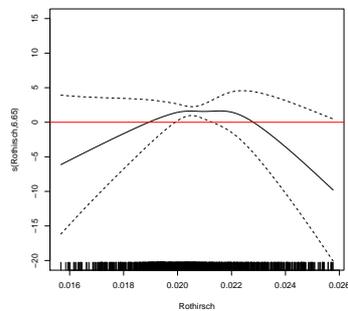
(g) Altbestand Fichte



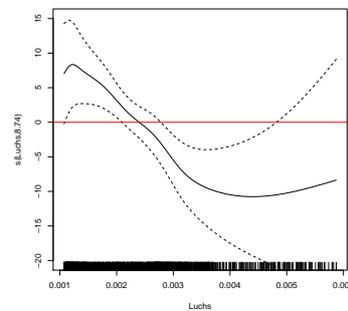
(h) Altbestand Buche



(i) Rehdichte



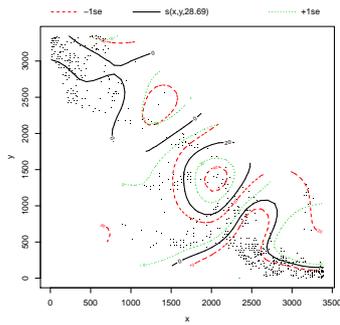
(j) Rothirschdichte



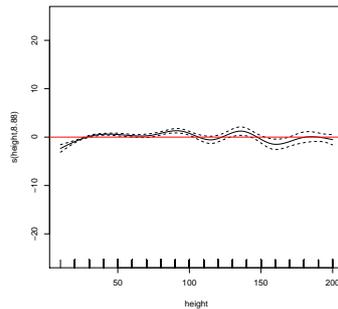
(k) Luchsdichte

Abbildung A.2.: Nonparametrische Einflussgrößen Poissonmodell Vogelbeere

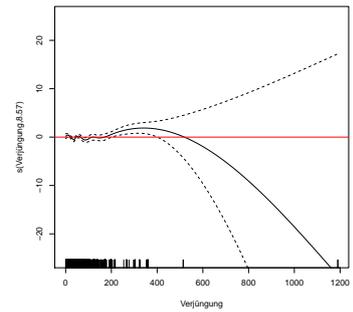
# B. Graphiken Tanne



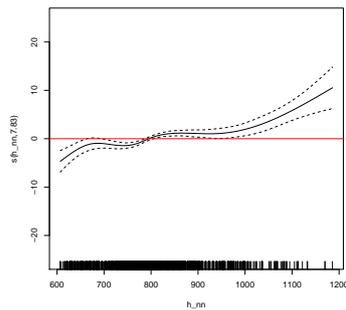
(a) Räumliche Komponente



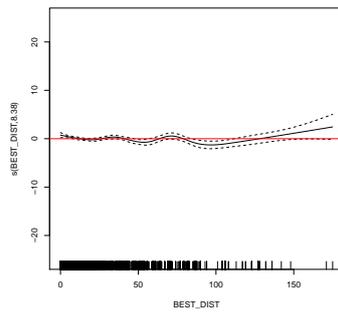
(b) Baumhöhe



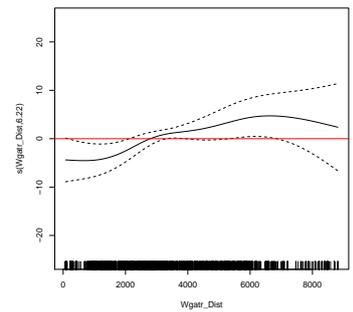
(c) Verjüngungsdichte



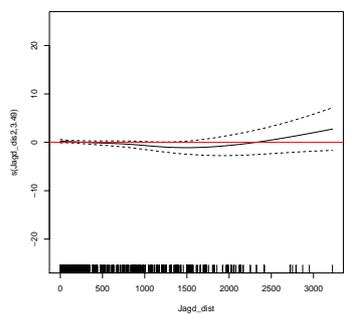
(d) Höhe über NN



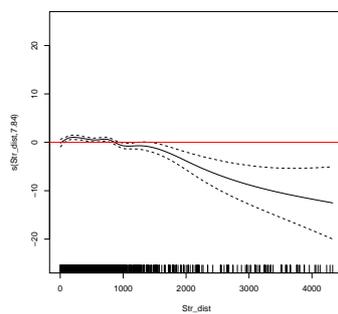
(e) Distanz Bestandesgrenze



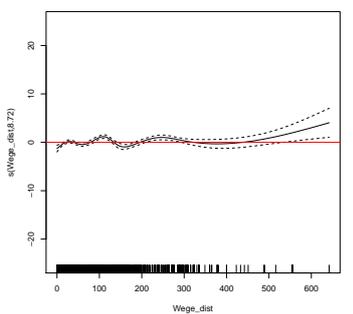
(f) Distanz Wintergatter



(g) Distanz Jagdflächen

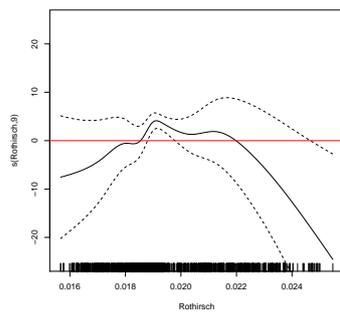


(h) Distanz Wanderwege

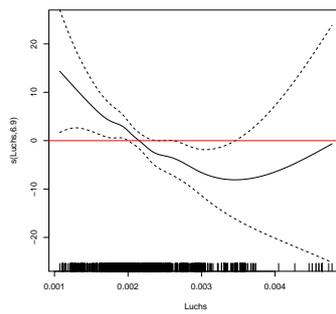


(i) Distanz Straßen

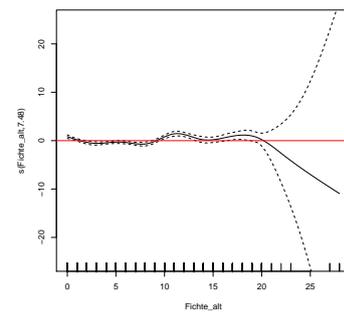
## B. Graphiken Tanne



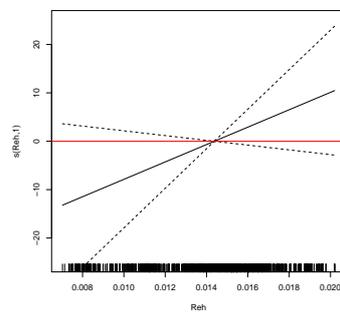
(j) Rothirschdichte



(k) Luchsdichte



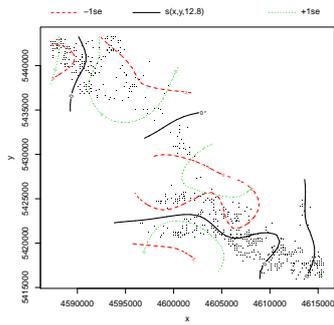
(l) Rehdichte



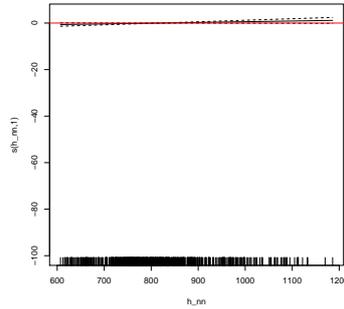
(m) Altbestand Fichte

Abbildung B.1.: Nonparametrische Einflussgrößen Binomialmodell Tanne

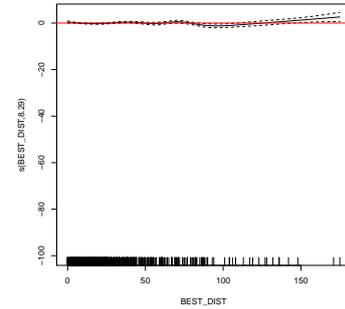
## B. Graphiken Tanne



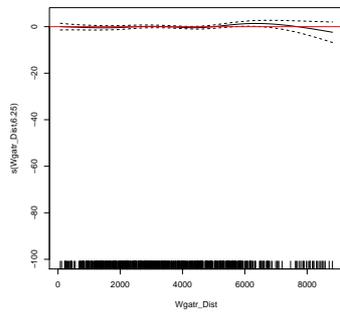
(a) Räumliche Komponente



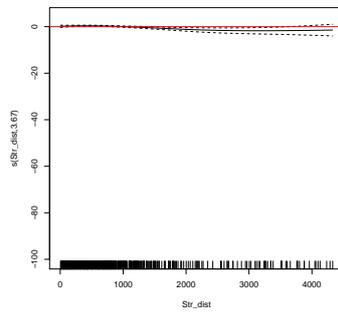
(b) Höhe über NN



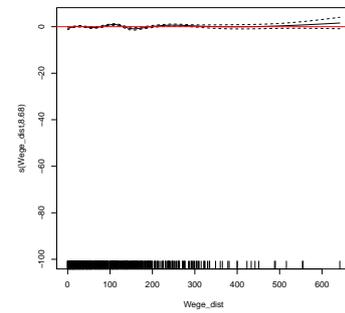
(c) Distanz Bestandesgrenze



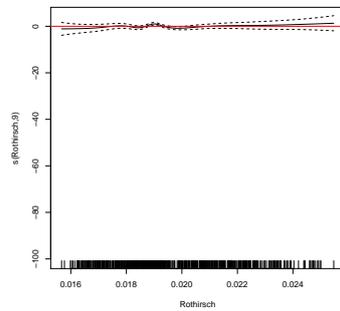
(d) Distanz Wintergatter



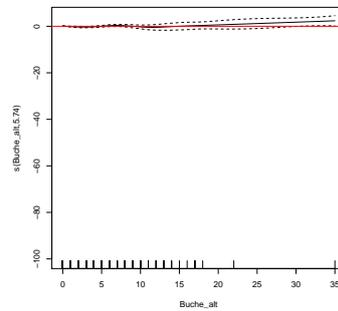
(e) Distanz Straße



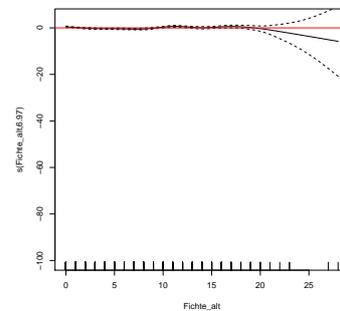
(f) Distanz Wanderwege



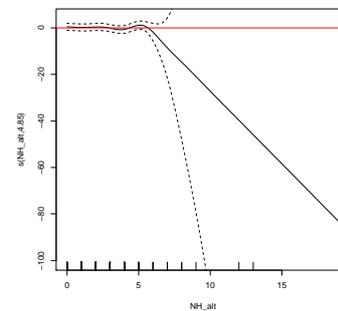
(g) Rothirschdichte



(h) Altbestand Buche



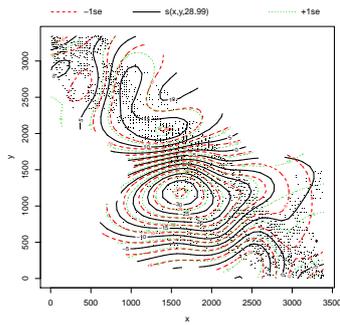
(i) Altbestand Fichte



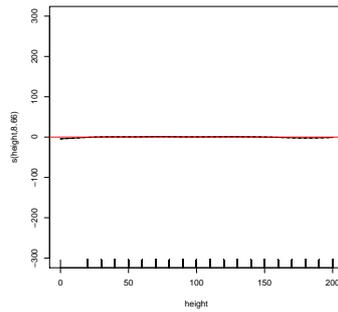
(j) Altbestand Nadelhölzer

Abbildung B.2.: Nonparametrische Einflussgrößen Poissonmodell Tanne

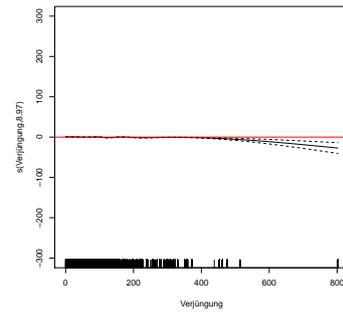
# C. Graphiken Buche



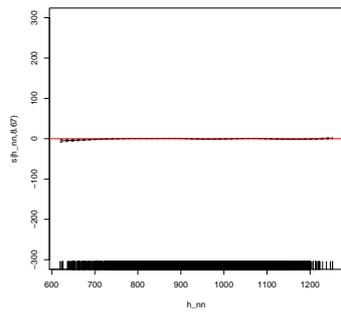
(a) Räumliche Komponente



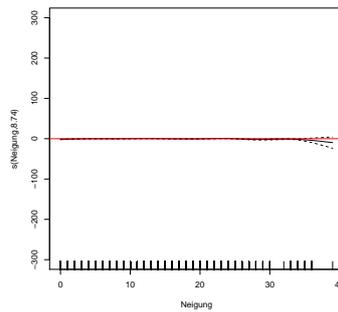
(b) Baumhöhe



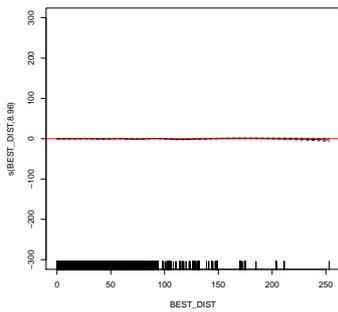
(c) Verjüngungsdichte



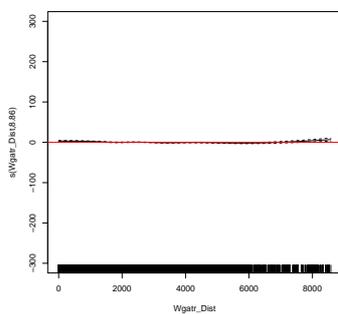
(d) Höhe über NN



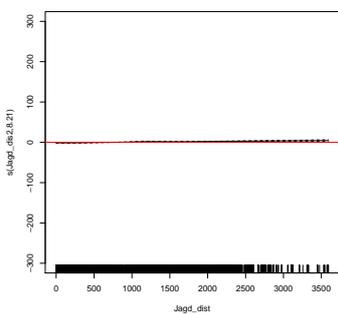
(e) Neigung



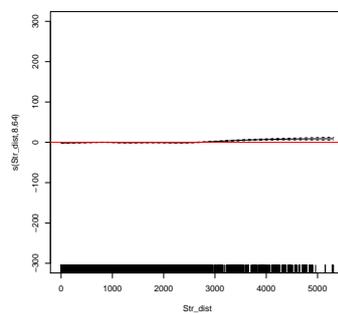
(f) Distanz Bestandesgrenze



(g) Distanz Wintergatter

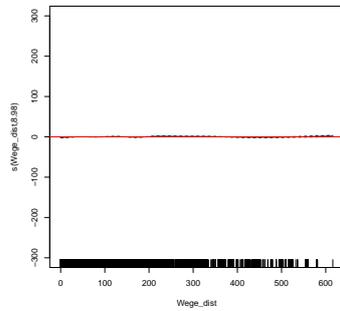


(h) Distanz Jagdflächen

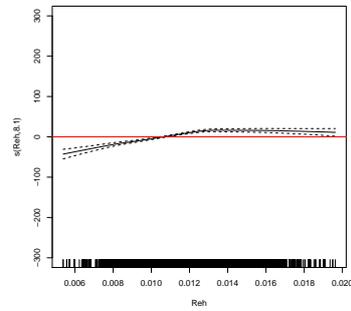


(i) Distanz Straße

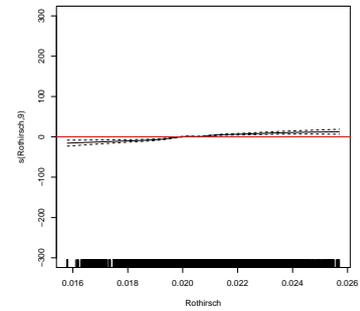
## C. Graphiken Buche



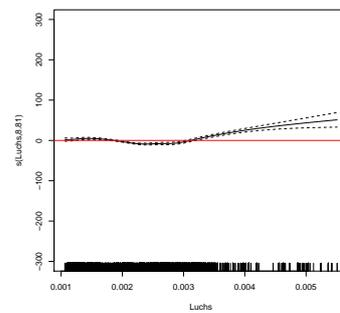
(j) Distanz Wanderwege



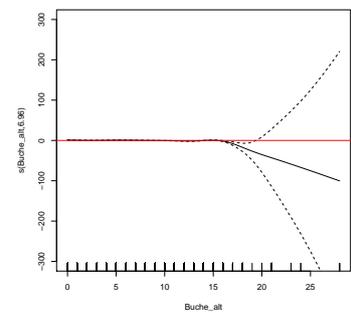
(k) Rehdichte



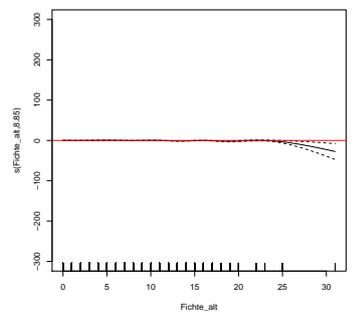
(l) Rothirschdichte



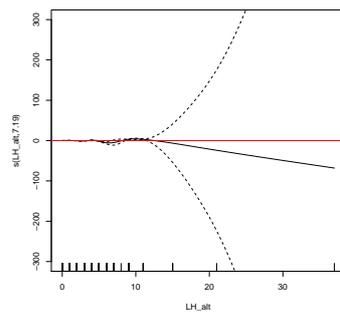
(m) Luchsdichte



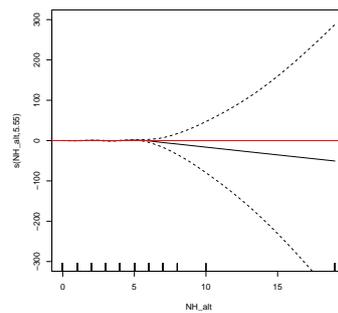
(n) Altbestand Buche



(o) Altbestand Fichte



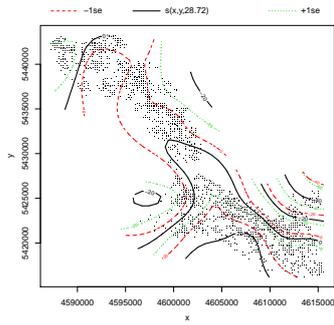
(p) Altbestand Laubhölzer



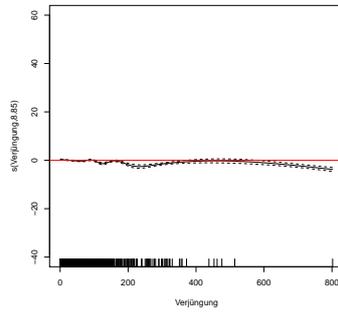
(q) Altbestand Nadelhölzer

Abbildung C.1.: Nonparametrische Einflussgrößen Binomialmodell Buche

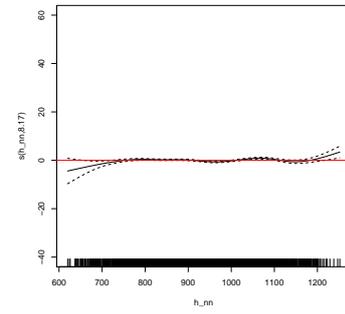
## C. Graphiken Buche



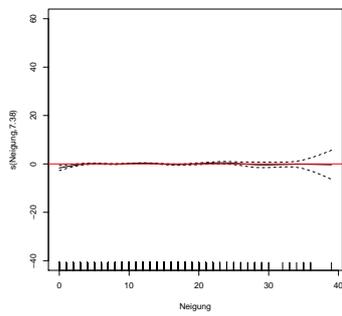
(a) Räumliche Komponente



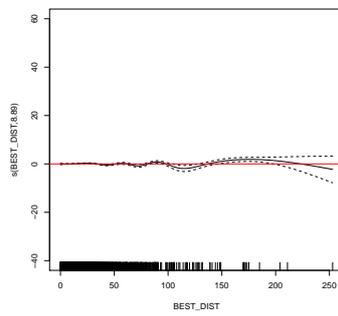
(b) Verjüngungsdichte



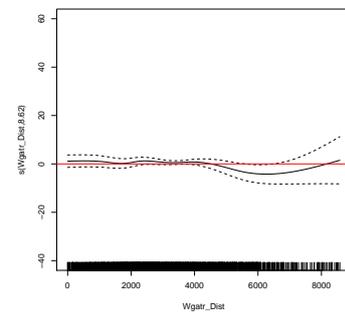
(c) Höhe über NN



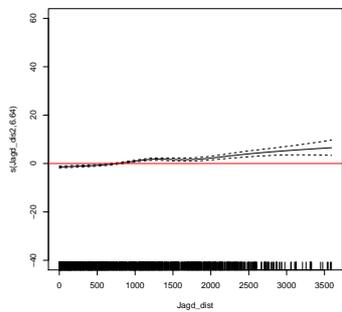
(d) Neigung



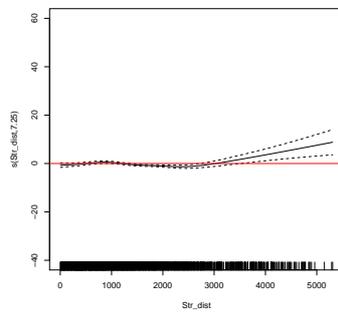
(e) Distanz Bestandesgrenze



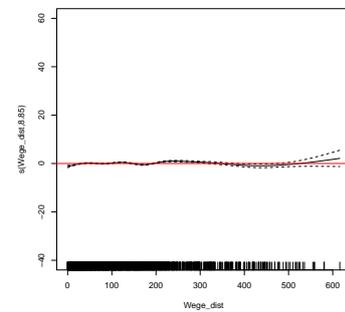
(f) Distanz Wintergatter



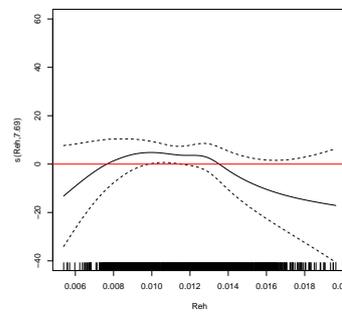
(g) Distanz Jagdflächen



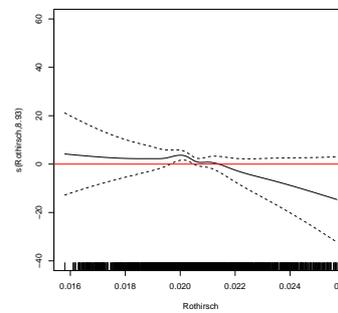
(h) Distanz Straßen



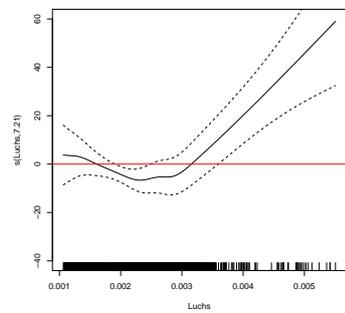
(i) Distanz Wanderwege



(j) Rehdichte

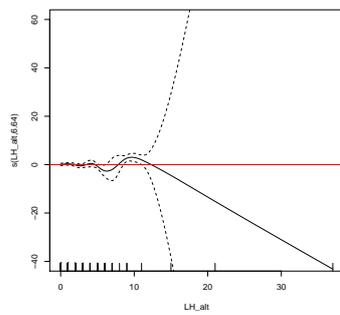


(k) Rothirschdichte

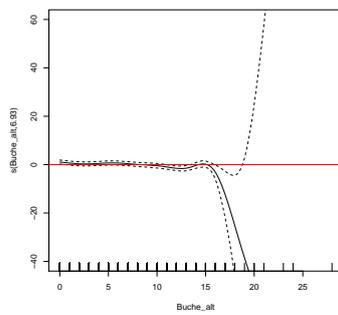


(l) Luchsdichte

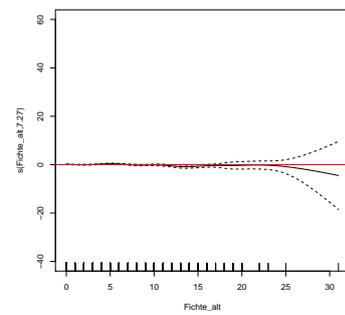
### C. Graphiken Buche



(m) Altbestand Laubhölzer



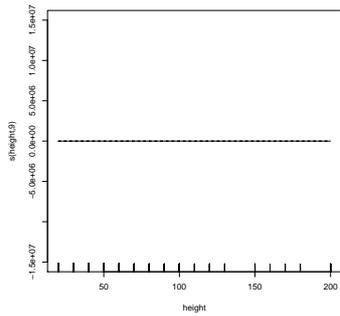
(n) Altbestand Buche



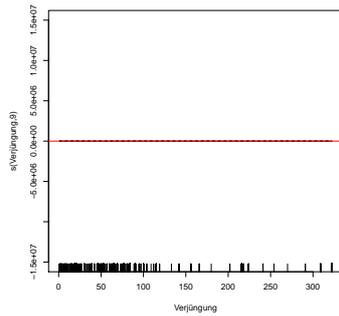
(o) Altbestand Fichte

Abbildung C.2.: Nonparametrische Einflussgrößen Poissonmodell Buche

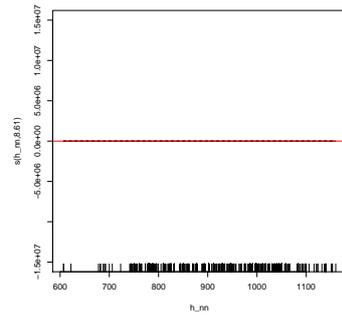
# D. Graphiken Bergahorn



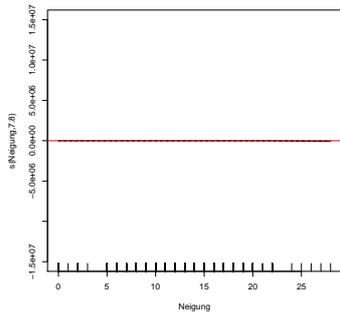
(a) Baumhöhe



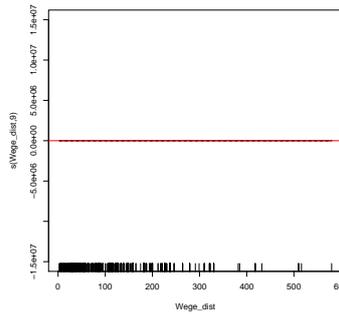
(b) Verjüngungsdichte



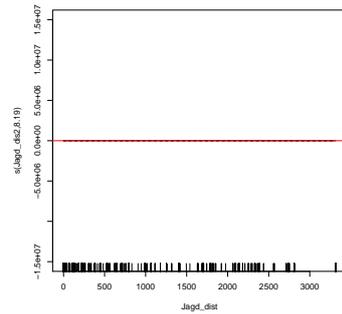
(c) Höhe über NN



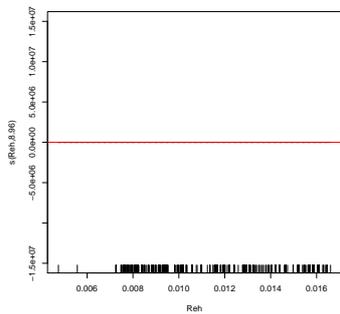
(d) Neigung



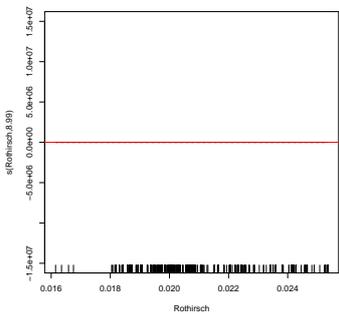
(e) Distanz Wanderwege



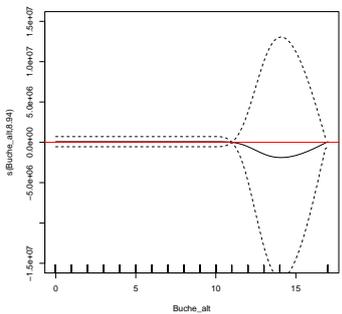
(f) Distanz Jagdflächen



(g) Rehdichte



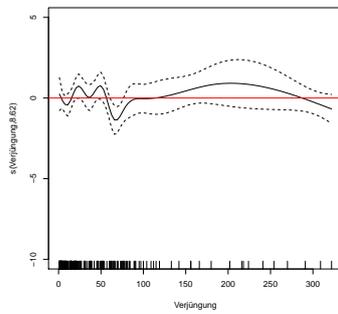
(h) Rothirschdichte



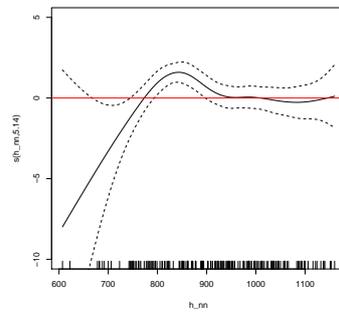
(i) Altbestand Buche

Abbildung D.1.: Nonparametrische Einflussgrößen Binomialmodell Bergahorn

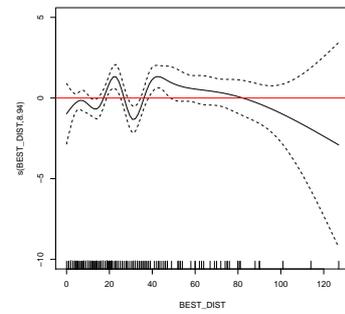
## D. Graphiken Bergahorn



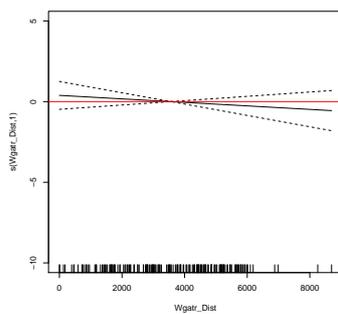
(a) Verjüngungsdichte



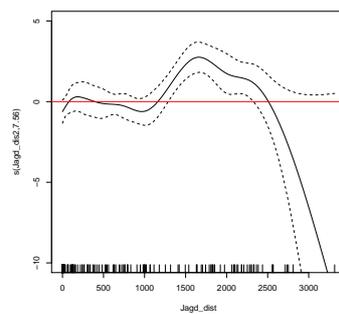
(b) Höhe über NN



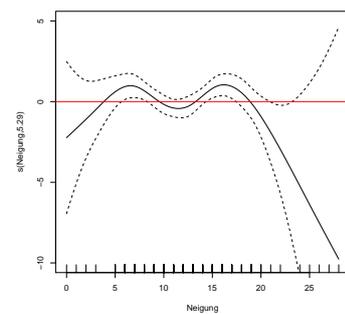
(c) Distanz Bestandesgrenze



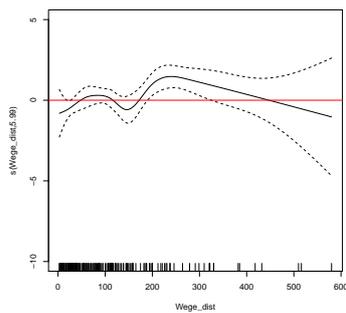
(d) Distanz Wintergatter



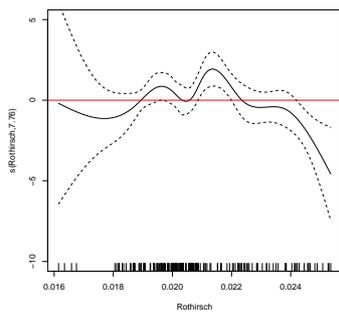
(e) Distanz Jagdflächen



(f) Neigung



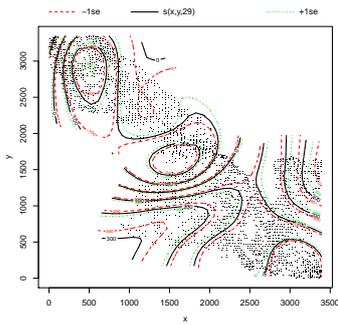
(g) Distanz Wanderwege



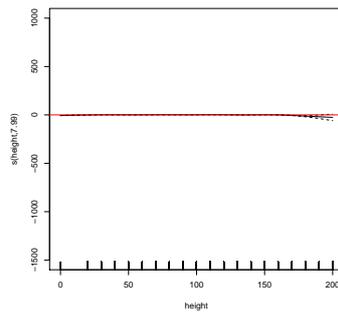
(h) Rothirschdichte

Abbildung D.2.: Nonparametrische Einflussgrößen Poissonmodell Bergahorn

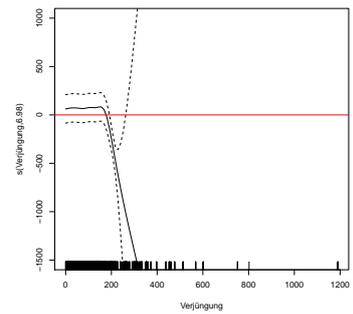
# E. Graphiken Fichte



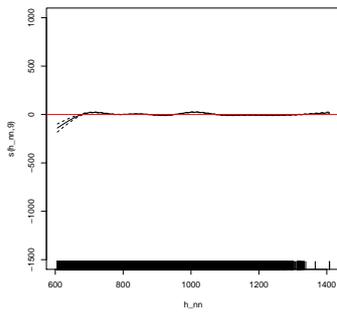
(a) Räumliche Komponente



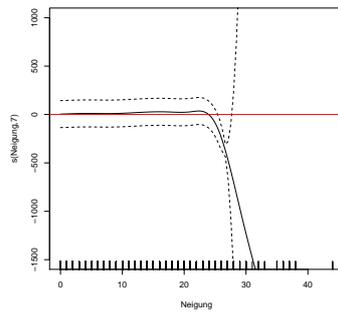
(b) Baumhöhe



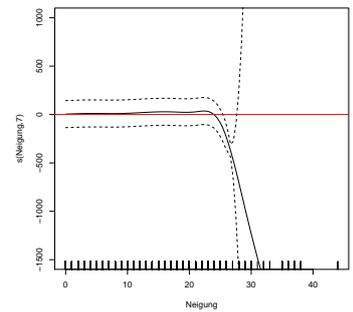
(c) Verjüngungsdichte



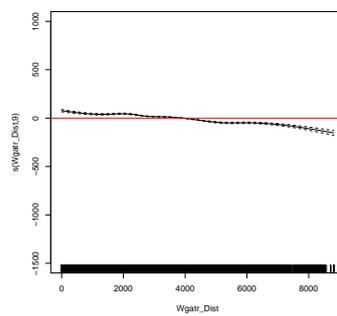
(d) Höhe über NN



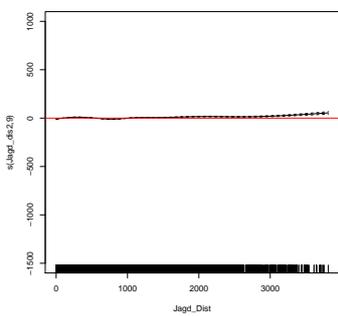
(e) Neigung



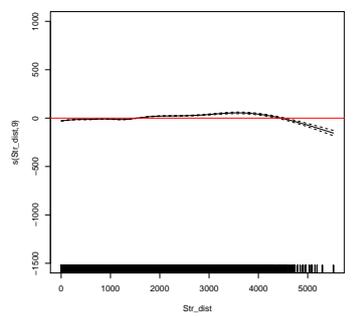
(f) Distanz Bestandesgrenze



(g) Distanz Wintergatter

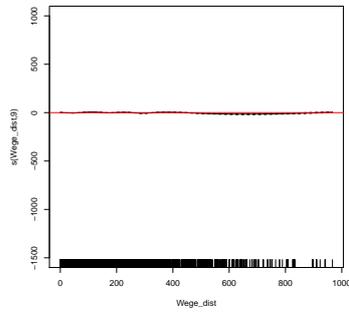


(h) Distanz bejagte Fläche

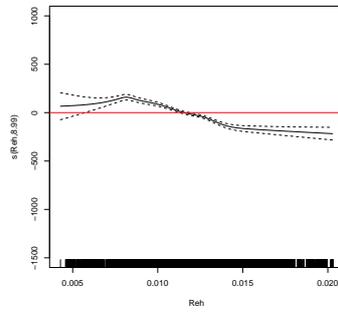


(i) Distanz Straße

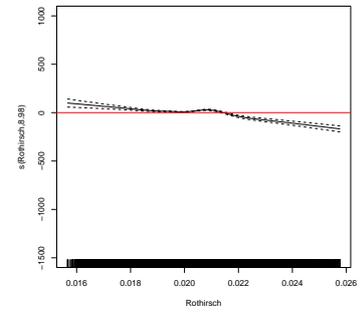
## E. Graphiken Fichte



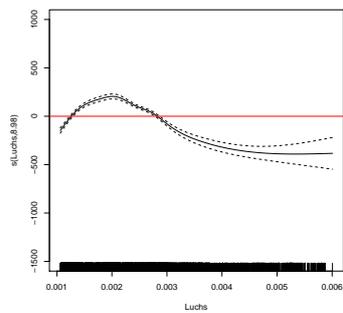
(j) Distanz Wanderwege



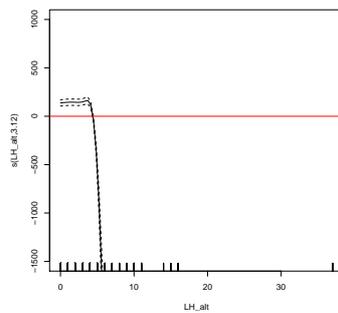
(k) Rehdichte



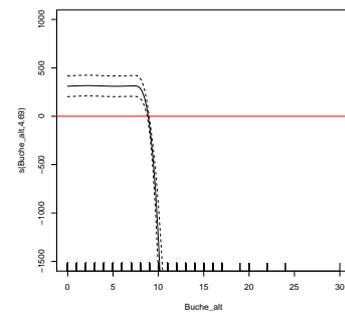
(l) Rothirschdichte



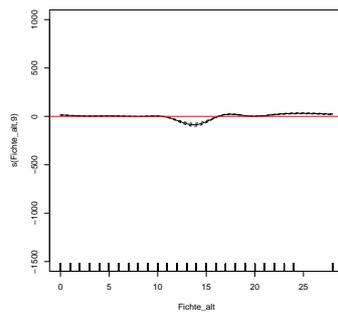
(m) Luchsdichte



(n) Altbestand Laubhölzer



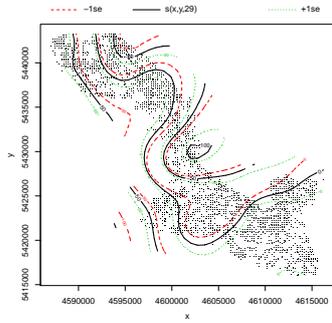
(o) Altbestand Fichte



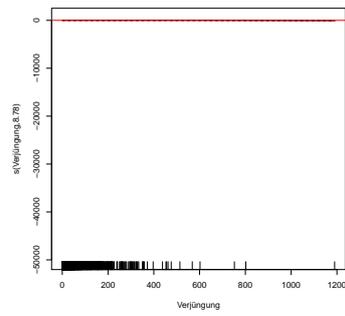
(p) Altbestand Buche

Abbildung E.1.: Nonparametrische Einflussgrößen Binomialmodell Fichte

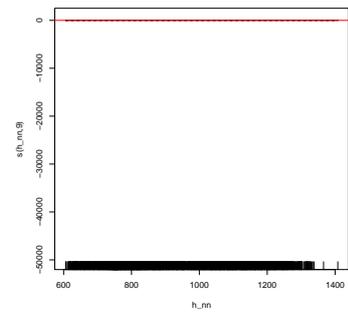
## E. Graphiken Fichte



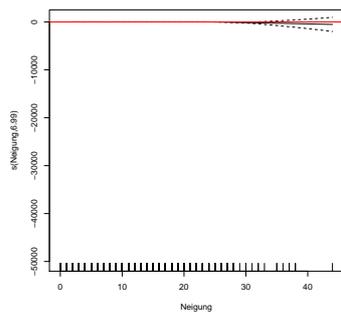
(a) Räumliche Komponente



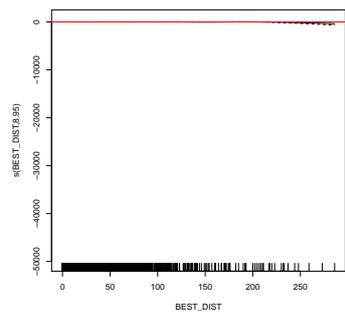
(b) Verjüngungsdichte



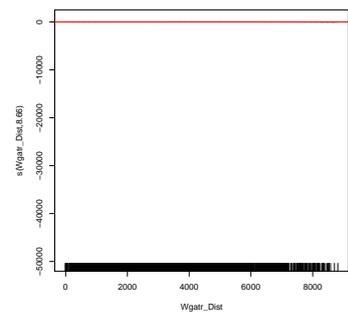
(c) Höhe über NN



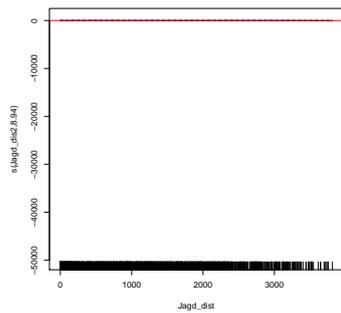
(d) Neigung



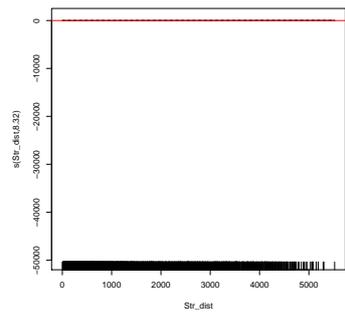
(e) Distanz Bestandesfläche



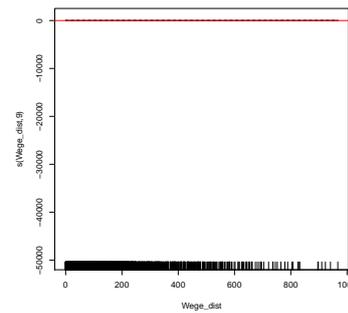
(f) Distanz Wintergatter



(g) Distanz bejagte Flächen

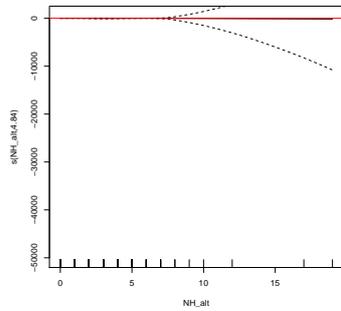


(h) Distanz Straße

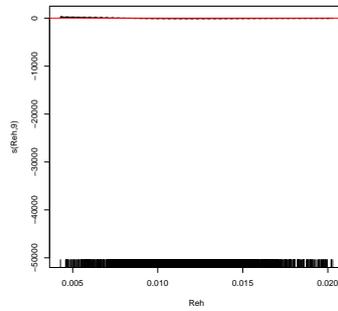


(i) Distanz Wanderwege

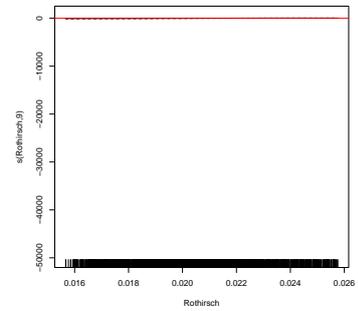
## E. Graphiken Fichte



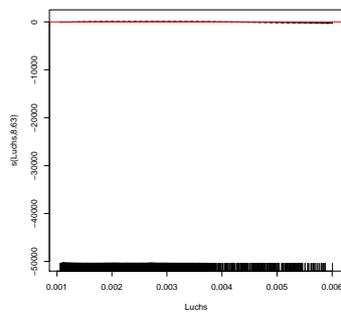
(j) Rehdichte



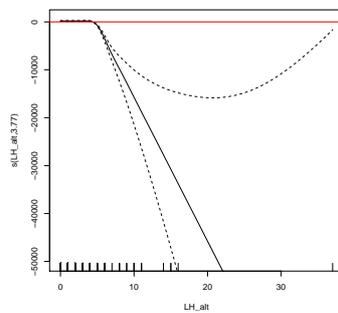
(k) Rothirschdichte



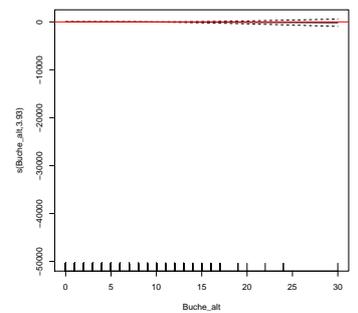
(l) Luchsdichte



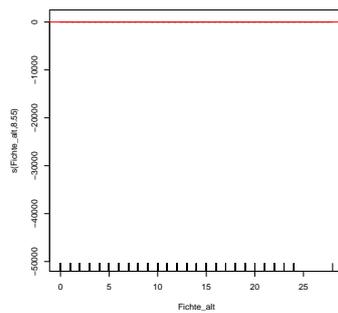
(m) Altbestand Nadelhölzer



(n) Altbestand Laubhölzer



(o) Altbestand Fichte



(p) Altbestand Buche

Abbildung E.2.: Nonparametrische Einflussgrößen Poissonmodell Fichte

# Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

München, den 23. Juli 2009

---

Lisa Möst