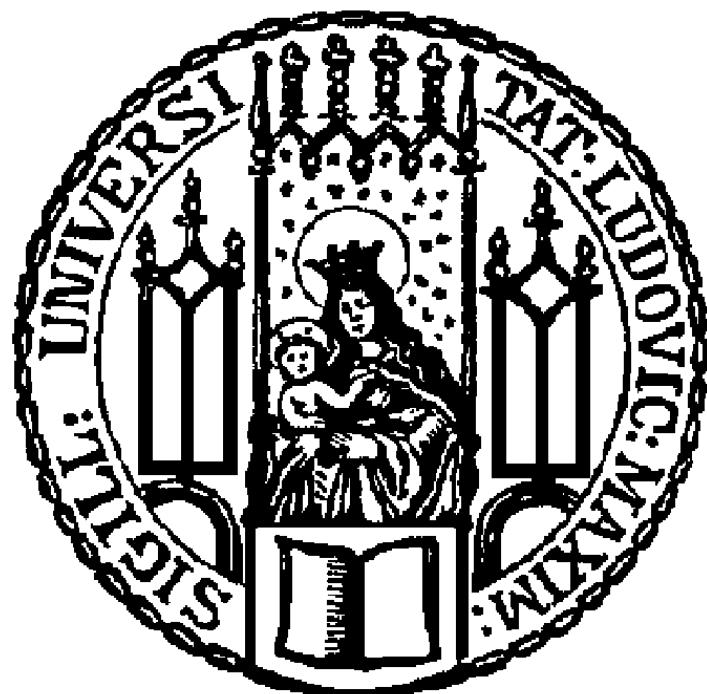


INSTITUT FÜR STATISTIK  
LUDWIG-MAXIMILIANS-UNIVERSITÄT  
MÜNCHEN

# Adaptive Disease Mapping

Masterarbeit  
betreut von Prof. Dr. Volker Schmid



Autor: Christian Schnell  
Matrikelnummer: 8077540

München, den 21. Oktober 2013



## Zusammenfassung

Bei der analytischen Betrachtung einer Disease Mapping Problemstellung ist die Annahme von strukturellen Abhängigkeiten zwischen den Ausprägungen einzelner Beobachtungseinheiten eine übliche Herangehensweise. Auf dieser Grundlage bieten sogenannte Gauß-Markov-Zufallsfelder die Möglichkeit räumliche Strukturen in eine Modellierung aufzunehmen. Das zentrale Ziel der vorliegenden Masterarbeit liegt in diesem Zusammenhang in der adaptiven Schätzung der Gewichtungsparameter eines Gauß-Markov-Zufallfeldes unter der Berücksichtigung der jeweiligen Nachbarschaftsstruktur. Als Ausgangssituation wird hierfür ein auf einem generalisierten linearen gemischten Modell basierender Bayes-Ansatz mit mehreren hierarchischen Stufen gewählt. Für die Umsetzung einer solch komplexen Modellstruktur wird die Methodik eines computerintensiven Markov-Chain-Monte-Carlo Verfahrens benötigt, welche im Rahmen dieser Arbeit in einem eigens implementierten Hybrid-Sampler ihre Berücksichtigung findet. Die erfolgten Anwendungen beziehen sich in diesem Kontext sowohl auf empirische Krebsdaten aus Schottland und Deutschland als auch auf rein simulationsbasierte Daten. Im Fokus der Analysen steht dabei abzuwegen, inwiefern die geschätzten Verteilungskarakteristika der Gewichtungsparameter die tatsächliche Stärke einer nachbarschaftlichen Beziehung zweier Beobachtungseinheiten repräsentieren. Zudem ist von Interesse, wie sich die Schätzungen auf Basis der schottischen Nachbarschaftsstruktur im Vergleich zu der wesentlich komplexeren räumlichen Anordnung der deutschen Landkreise verhält. Eine nähere Betrachtung der geschätzten Posteriori-Mittelwerte zeigt, dass die räumlichen Effekte durch die zusätzliche Berücksichtigung von anpassungsfähigen Gewichten kaum beeinflusst wird, unabhängig von der zugrundeliegenden Struktur. Die geschätzten mittleren Gewichte stehen jedoch in einem deutlichen Zusammenhang mit der Anzahl an Nachbarn der betroffenen Beobachtungseinheiten.

# Inhaltsverzeichnis

<b>1 Einleitung</b>	<b>1</b>
1.1 Disease Mapping . . . . .	1
1.2 Zielsetzung und Aufbau der Arbeit . . . . .	2
<b>2 Theoretischer Hintergrund</b>	<b>3</b>
2.1 Generalisiertes lineares gemischtes Modell (GLMM) . . . . .	3
2.1.1 Generalisiertes lineares Modell . . . . .	3
2.1.2 Gemischtes Modell . . . . .	7
2.1.3 Modellnotation . . . . .	8
2.1.4 Zusammenfassung: Generalisiertes lineares gemischtes Modell . . . . .	9
2.2 Bayes-Inferenz . . . . .	10
2.2.1 Grundlegendes . . . . .	10
2.2.2 Hierarchische Modellstruktur . . . . .	13
2.2.3 Modellwahlkriterien . . . . .	15
2.2.4 Zusammenfassung: Bayes-Inferenz . . . . .	18
2.3 Allgemeines zur räumlichen Statistik . . . . .	19
2.3.1 Räumliche Effekte – strukturiert und unstrukturiert . . . . .	19
2.3.2 Räumliche Daten – diskret und stetig . . . . .	21
2.3.3 Räumliche Gitter – regulär und irregulär . . . . .	24
2.3.4 Zusammenfassung: Allgemeines zur räumlichen Statistik . . . . .	25
2.4 Gauß-Markov-Zufallsfelder (GMZF) . . . . .	26
2.4.1 Nachbarschaftsstrukturen über ungerichtete Graphen . . . . .	26
2.4.2 Räumliche Markov-Eigenschaft . . . . .	30
2.4.3 Verteilung der Gauß-Markov-Zufallsfelder . . . . .	35
2.4.4 Adaptive Modellierung . . . . .	42
2.4.5 Zusammenfassung: Gauß-Markov-Zufallsfelder . . . . .	45
<b>3 Anwendung</b>	<b>46</b>
3.1 Datengrundlage . . . . .	46
3.1.1 Daten zu Lippenkrebs in Schottland . . . . .	47
3.1.2 Daten zu Kehlkopfkrebs in Deutschland . . . . .	50
3.2 Umsetzung der hierarchischen Bayes-Struktur . . . . .	54
3.2.1 Verteilungsannahmen aller hierarchischen Stufen . . . . .	54
3.2.2 Reparametrisierung des Modells . . . . .	57
3.2.3 Aufstellung der vollständig bedingten Dichten . . . . .	58
3.3 Implementierung . . . . .	64
3.3.1 Markov-Chain-Monte-Carlo (MCMC) Verfahren . . . . .	65
3.3.2 Verteilungsapproximation über Taylorentwicklung . . . . .	67
3.3.3 Skizze des Hybrid-Samplers . . . . .	69
3.3.4 Optimierung von Rechenoperationen . . . . .	73
3.4 Anwendung auf empirische Daten ohne Gewichtsadaption . . . . .	77
3.4.1 Modellanpassungen . . . . .	77
3.4.2 Grundlagen der grafischen Samplingbewertung . . . . .	78
3.4.3 Schätzergebnisse aus den schottischen Lippenkrebsdaten . . . . .	81
3.4.4 Schätzergebnisse aus den deutschen Kehlkopfkrebsdaten . . . . .	89

3.5	Anwendung auf simulierte Daten mit Gewichtsadaption . . . . .	94
3.5.1	Ausgangslage für das Sampling simulierter Daten . . . . .	94
3.5.2	Schätzergebnisse der simulierten Daten . . . . .	97
3.5.3	Bewertung der Schätzergebnisse . . . . .	104
3.6	Anwendung auf empirische Daten mit Gewichtsadaption . . . . .	107
3.6.1	Konvergenzanalyse nach Gelman und Rubin . . . . .	108
3.6.2	Schätzergebnisse aus den schottischen Lippenkrebsdaten . . . .	110
3.6.3	Schätzergebnisse aus den deutschen Kehlkopfkrebsdaten . . . .	113
<b>4</b>	<b>Abschließende Bemerkungen</b>	<b>115</b>
4.1	Zusammenfassung der wichtigsten Ergebnisse . . . . .	115
4.2	Grenzen der Analyse und Ausblick . . . . .	116
<b>Abbildungsverzeichnis</b>		<b>117</b>
<b>Referenzen</b>		<b>119</b>
<b>Elektronischer Anhang</b>		<b>122</b>



# 1 Einleitung

Räumliche Analysen gewinnen in der heutigen Zeit immer mehr an Bedeutung. Ein ausschlaggebender Grund für diese Entwicklung ist die kontinuierliche Weiterentwicklung und Verbreitung von Computern seit Anfang der 1990er Jahre. In vielen Forschungsgebieten, aber auch explizit in vielen Teilgebieten der Statistik, hat die Entwicklung von leistungsstarken und zugleich preiswerten Computern eine Wende in Bezug auf die Möglichkeiten der Statistiker und deren Methoden bewirkt. Die simple deskriptive Betrachtungen von Daten, welche die Grundlage für eine ausführliche Datenanalyse bildet, war vor dieser computationalen Wende kaum denkbar bzw. mit einem hohen Zeitaufwand verbunden. Neben der Anwendung von computerintensiven Methoden ermöglichen die neuen Rechenleistungen zusätzlich das Sammeln und Abspeichern von riesigen Datenmengen. In sogenannten „Smart Cities“ ist der neueste Trend die Verknüpfung dieser Daten mit geografischen Informationen über die Internetanbindung von Smartphones und den Messungen verschiedenster Sensoren. Auf diesem Wege können unter anderem Werte zur Luftverschmutzung, zur Lärmbelastung oder aber auch zur Ausbreitung von Krankheiten in einzelnen Stadtgebieten erhoben und überwacht werden. Dazu fokussieren diese Projektstädte<sup>1</sup> eine strategische Vernetzung ihrer Stadt für eine Aus- und Bewertung der ortsgebundenen Daten nahezu in Echtzeit (vgl. Spiegel, 2013). Die Vielzahl an Möglichkeiten der räumlichen Statistik, welche sich aus diesen neuen Errungenschaften ergeben können, sind momentan noch nicht abzusehen. Fest steht jedoch, dass für eine optimale Ausnutzung der Informationen eine Weiterentwicklung der Methoden zur Analyse von räumlichen Daten unerlässlich ist. Diese Erkenntnis dient als Motivation der vorliegenden Arbeit zu dem Thema „Adaptive Disease Mapping“, deren zentrales Ziel das Bilden von adaptiven Methoden bereits bestehender räumlicher Analyseverfahren in Bezug auf Krankheitsdaten ist.

## 1.1 Disease Mapping

Die Bezeichnung „Disease Mapping“ wird oftmals in verschiedenen Zusammenhängen verwendet, wodurch eine gewisse Unklarheit über den Begriff entstehen kann. Grundsätzlich werden damit statistische Methoden bezeichnet, die auf Daten in Verbindung mit Landkarten angewendet werden, die in einem epidemiologischen Zusammenhang stehen. In der Regel handelt es sich dabei um Häufigkeiten einer Krankheit, welche für die jeweiligen Beobachtungseinheiten kumuliert und abgetragen werden. Etwas spezieller ist damit aber auch die Schätzung eines relativen Risikos in Verbindung mit Krankheitskarten gemeint (vgl. Melnick und Everitt, 2008, S. 514). Nach Rytkönen (2004) ist allerdings in jedem Fall die Identifizierung von bedeutungsvollen Informationen bezüglich einer Krankheit das zentrale Ziel:

„Regardless of whether simple mapping methods or advanced spatial models are used, the ultimate goal must be differentiation between meaningful and spurious information“ (Rytkönen, 2004, S. 15).

Anfänglich wurden Disease Mapping Methoden eingesetzt, um den Ursprung einer Infektion zu ermitteln oder die allgemeine Ausbreitung einer Krankheit zu beschreiben.

---

<sup>1</sup>Die spanische Stadt Santander kann diesbezüglich als Vorreiter bezeichnet werden.

Die Einbeziehung zusätzlicher Faktoren, die das Auftreten von Krankheiten verstärken können, wie bestimmte Umwelteinflüsse oder die geografische Lage einer Region, stellt deshalb einen enormen Informationsgewinn für das öffentliche Gesundheitswesen dar. Die Entwicklung von komplexeren Disease Mapping Verfahren stößt somit auf ein breites Interesse, da durch deren Anwendung Maßnahmen zur Bekämpfung von Krankheiten geschaffen werden können (vgl. Rytkönen, 2004, S. 11).

## 1.2 Zielsetzung und Aufbau der Arbeit

Bei der analytischen Betrachtung einer Disease Mapping Problemstellung ist die Annahme von strukturellen Abhängigkeiten zwischen den Ausprägungen einzelner Beobachtungseinheiten eine übliche Herangehensweise. Auf dieser Grundlage bieten sogenannte Gauß-Markov-Zufallsfelder die Möglichkeit räumliche Strukturen in einer Modellierung zu berücksichtigen. Im Rahmen eines bayesianischen Modellaufbaus geschieht das in der Regel unter der Annahme eines globalen Glättungsparameters. Oftmals kann es jedoch sinnvoll sein kantenerhaltende Verfahren zu verwenden, die beispielsweise bei Bildgebungsanalysen wie dem „Brain Mapping“ eingesetzt werden. Hier wird eine adaptive Schätzung der lokalen Glättungsgewichte durchgeführt, d. h. die Gewichtung jeder einzelnen Nachbarschaft des betrachteten Systems wird individuell geschätzt. Eine direkte Umsetzung auf Krankheitsdaten ist jedoch nicht möglich, da sich die Ausgangssituation zu Brain Mapping Daten grundlegend unterscheidet. Im Allgemeinen liegen dort für jede einzelne Beobachtungseinheit wiederholte Messungen über die Zeit vor, womit ein deutlicher Informationsgewinn für die Modellierung der Zielgröße einhergeht. Zudem sind die Verfahren im Gegensatz zu einer Disease Mapping Datensituation auf normalverteilte Zielgrößen zugeschnitten. Die zentrale Zielsetzung dieser Arbeit ist deshalb die Modellierung flexibler Nachbarschaftsgewichtungen auf räumlich gebundene Krankheitsdaten anzupassen.

Der Aufbau der vorliegenden Arbeit gliedert sich dabei in zwei große Blöcke. In Kapitel 2 wird zunächst der theoretische Hintergrund der Modellierung räumlicher Effekte auf Basis eines generalisierten linearen gemischten Modells in einem bayesianischen Kontext aufbereitet (siehe Abschnitte 2.1 und 2.2), bevor nach der Einführung grundlegender Aspekte der räumlichen Statistik (vgl. Unterkapitel 2.3) die Eigenschaften von Gauß-Markov-Zufallsfeldern bis hin zu den theoretischen Grundlagen einer Modellierung mit adaptiven Nachbarschaftsgewichten erläutert wird (siehe Abschnitt 2.4). Die Anwendung der erlangten Hintergrundkenntnisse ist schließlich Bestandteil von Kapitel 3, welches eingeleitet wird mit einer deskriptiven Aufbereitung der empirischen Datengrundlage. Zudem stellen die Umsetzung der hierarchischen Struktur des bayesianischen Modellaufbaus sowie deren Implementierung im Rahmen eines sogenannten Hybrid-Samplers zwei zentrale Aspekte der Anwendung dar (siehe Unterkapitel 3.2 und 3.3). Im Vordergrund steht jedoch die Aufbereitung der erzielten Schätzergebnisse unter der Einbeziehung von empirische Daten (ohne Gewichtsadaption in Abschnitt 3.4 und mit Gewichtsadaption in Abschnitt 3.6) und rein simulierten Daten (siehe Unterkapitel 3.5). Kapitel 4 beinhaltet abschließend eine Zusammenfassung der wichtigsten Ergebnisse und Erkenntnisse sowie eine Aufarbeitung weiterführender Analysemöglichkeiten im Ausblick.

## 2 Theoretischer Hintergrund

Die folgende Aufarbeitung des theoretischen Hintergrunds im Rahmen der Analyse von räumlichen Daten bildet die Grundlage der adaptiven Möglichkeiten, die in Kapitel 3 ihre Anwendung finden. Der Ansatz, der im Rahmen dieser Arbeit verwendet wird, basiert auf der Einbindung von sogenannten Gauß-Markov-Zufallsfeldern, die durch eine bayesianische Schätzung von generalisierten linearen gemischten Modellen ermöglicht wird. Nach dem Bayes-Prinzip kann für diese Zufallsfelder die gewünschte Priori-Verteilung festgelegt werden, die eine Modellierung der strukturellen Begebenheiten einer Landkarte erlaubt.

Für eine schrittweise Aufarbeitung der theoretischen Grundlagen, ist es deshalb unerlässlich zunächst die Grundzüge eines generalisierten linearen gemischten Modells als Ausgangssituation zu erläutern (siehe Unterkapitel 2.1) und ferner das Prinzip der Bayes-Inferenz aufzuarbeiten (siehe Abschnitt 2.2). Bis zu diesem Aspekt unterscheidet sich die Vorgehensweise im Rahmen einer gewöhnlichen statistischen Problemstellung nicht von der Modellierung in Verbindung von räumlichen Zusammenhängen. Da das Hauptelement der Gauß-Markov-Zufallsfelder die Abbildung der strukturellen Anordnung von vordefinierten Regionen beinhaltet, soll die Zusammenfassung wichtiger Eigenschaften hinsichtlich einer räumlichen Datensituation in Unterkapitel 2.3 die Grundlage für die Ausführungen in Bezug auf Zufallsfelder mit Gauß-Markov-Eigenschaften in Abschnitt 2.4 liefern.

### 2.1 Generalisiertes lineares gemischtes Modell (GLMM)

Ein generalisiertes lineares gemischtes Modell (kurz: GLMM (engl.)) erweitert den standardmäßigen linearen Regressionsansatz durch die Hinzunahme von zufälligen Effekten (Random Effects) und der Verallgemeinerung auf normal- oder andersverteilte Zielgrößen. Das folgende Unterkapitel behandelt in diesem Zusammenhang die theoretische Basis eines GLMMs, die inhaltlichen Begründungen für die einzelnen Bestandteile sowie eine Aufbereitung der üblichen Modellnotation. Darüber hinaus beinhaltet diese Darstellung eine Aufarbeitung der Ausgangssituation für die Modellierung von räumlichen Daten in Verbindung mit Gauß-Markov-Zufallsfeldern. Deren Struktur kann auf der Grundlage von GLMMs durch verschiedene Vorgehensweisen erfasst werden, die in diesem Kontext eine kurze Erwähnung finden sollen, ohne tiefer auf deren Anwendung einzugehen. Abschnitt 2.1.4 fasst abschließend die zentralen Eigenschaften einer solchen Modellierung zusammen.

#### 2.1.1 Generalisiertes lineares Modell

Der klassischen linearen Normalregression liegt die Annahme einer Normalverteilung der Residuen (und demzufolge der Zielgröße) und ein linearer Einfluss der Kovariablen zugrunde. Die registrierten Krankheitsfälle, welche in der vorliegenden Problemstellung im Rahmen von Disease Mapping Analysen als Zielgröße fungieren, sind als Zähldaten zu betrachten und somit nicht normalverteilt, da eine solche Datensituation in der Regel näherungsweise durch eine Poissonverteilung beschrieben werden kann. Hierfür gilt es zu berücksichtigen, dass diese zu der Klasse der Exponentialfamilien gehört, womit eine Modellierung mit einem generalisierten Ansatz ermöglicht bzw. unerlässlich wird.

## Poissonverteilung

Die Poissonverteilung eignet sich demnach für die Modellierung von Zählvorgängen. Dabei werden bestimmte Ereignisse gezählt, die innerhalb eines fest vorgegebenen zeitlichen und/oder räumlichen Intervalls eintreten können. Die mögliche Anzahl der Ereignisse ist dabei nach oben nicht beschränkt. Beispiele für Zählvorgänge dieser Art sind die Anzahl von Schadensmeldungen bei einer Sachversicherung innerhalb eines Jahres, die Anzahl von Krankheitsfällen einer (seltenen) Krankheit in einem Monat oder die Anzahl von Kunden, die in einem Monat bei einer Bank einen Kredit beantragen (vgl. Fahrmeir *et al.*, 2007, S. 260 f.). In der vorliegenden Analyse wird das Auftreten von Sterbefällen, resultierend aus bestimmten Krebserkrankungen, auf Landkreisebene in Deutschland bzw. Schottland gezählt. Dieser Zählvorgang bezieht sich somit auf die Grenzen der einzelnen Landkreise, d. h. es ist von Interesse, wie viele Krankheitsfälle in den einzelnen Landkreisen registriert wurden. Zusätzlich wurde im Zuge der damaligen Datenerhebung eine zeitliche Eingrenzung vorgenommen, indem sich die beobachteten Anzahlen in den Landkreisen auf bestimmte Jahre beziehen (siehe hierfür Abschnitt 3.1).

Im Folgenden ist die Größe  $Y$  von Bedeutung, welche die Anzahl der Ereignisse, die innerhalb der vorgegebenen zeitlichen und räumlichen Intervalle eintreten, beinhaltet und die somit als Zielgröße fungiert. Eine Zufallsvariable  $Y$  mit folgender Wahrscheinlichkeitsverteilung heißt schließlich poissonverteilt mit dem Parameter  $\lambda > 0$ ; in abgekürzter Schreibweise  $Y \sim Po(\lambda)$ :

$$Po_\lambda(y) = \frac{\lambda^y}{y!} \exp(-\lambda) \quad (1)$$

Für die Poissonverteilung gilt im Allgemeinen eine Äquidispersionseigenschaft, d. h. dass der Erwartungswert und die Varianz identisch sind mit  $\mathbb{E}(Y) = \lambda$  und  $\mathbb{V}(Y) = \lambda$ . Außerdem werden als Wertebereich die Zahlen  $\mathbb{N}_0 = \{0, 1, 2, \dots\}$  angenommen (vgl. Fahrmeir *et al.*, 2007, S. 262).

## Zusammenhang zwischen Poisson- und Binomialverteilung

In der Epidemiologie findet neben der Poisson- auch die Binomialverteilung eine häufige Verwendung. Die Ausprägung in einer Region wird dabei als binomialverteilt aufgefasst mit einer unbekannten Wahrscheinlichkeit  $\pi$ , die sich auf die Anzahl der Personen  $n$  bezieht, die in der jeweiligen Region unter Risiko stehen (vgl. Schmid, 2004, S. 7 f.). Die Wahrscheinlichkeitsverteilung  $Y \sim B(n, \pi)$  ist diesbezüglich folgendermaßen zusammengesetzt (vgl. Fahrmeir *et al.*, 2007, S. 254):

$$B_{n,\pi}(y) = \binom{n}{y} p^y (1-p)^{n-y} \quad \text{für } y = 0, 1, \dots, n$$

Das Auftreten einer Krankheit in der Bevölkerung kann für jede unter Risiko stehende Person als ein, voneinander unabhängiges, Bernoulli-Experiment interpretiert werden, das mit einer extrem kleinen Erfolgswahrscheinlichkeit behaftet ist. Da die numerische

Berechnung bei großem  $n$  und sich oft wiederholenden Rechenvorgängen trotz der heutigen Rechenleistungen relativ aufwendig werden kann, stellt die Poissonverteilung eine sinnvolle Approximation der Binomialverteilung dar. Grundsätzlich gilt demnach, dass für  $n \rightarrow \infty$  und  $n \cdot \pi \rightarrow \lambda$  die Werte dieser beiden Dichten konvergieren (Nach einer Faustregel bereits ab  $n > 30$  und  $\pi \leq 0,05$  (vgl. Fahrmeir et al., 2007, S. 320)). Im Zuge dessen ist ein weiterer Vorteil der Approximation durch die Poissonverteilung, dass  $\lambda$  als alleiniger Parameter auftritt, weshalb viele unterschiedliche Binomialverteilungen durch eine einzelne Poissonverteilung angenähert werden können (vgl. Hübner, 2009, S. 64 f.).

Die Konstellation von großen  $n$ 's und kleinen  $\pi$ 's tritt in der Epidemiologie nahezu immer auf, weshalb der Ansatz über eine Poissonverteilung als die gängigere Variante zu betrachten ist. Ein weiterer Vorteil ist außerdem, dass die Häufigkeiten der Zielgröße über Kovariablen, wie beispielsweise der Altersstruktur einer Region, standardisiert werden können (vgl. Schmid, 2004, S. 8). Diese Vorgehensweise wird auch als Einbindung eines sogenannten Offsets in die Modellierung bezeichnet, auf welche an späterer Stelle in diesem Abschnitt genauer eingegangen werden soll. Aufgrund der genannten Vorteile, erscheint die Verwendung der Poissonverteilung bezüglich der epidemiologischen Datenlage im Rahmen der vorliegenden Arbeit als sinnvoller Ansatz (vgl. Unterkapitel 3.1).

### Poissonverteilung als Bestandteil der Exponentialfamilie

Die Zugehörigkeit der Poissonverteilung zur Exponentialfamilie ermöglicht die Modellierung mit einem generalisierten Modell. Die allgemeine Exponentialform einer solchen Verteilung stellt sich wie folgt dar (vgl. Tutz, 2011, S. 21):

$$f(y | \theta, \phi) = \exp \left( \frac{y\theta - b(\theta)}{\phi} + \underbrace{c(y, \phi)}_{\text{Normierung}} \right)$$

mit  $\theta \hat{=} \text{natürlicher/kanonischer Parameter}$   
und  $\phi \hat{=} \text{Dispersionsparameter}$

Durch anschließende Umformung der ursprünglichen Darstellung der Poissonverteilung, kann dieser Zusammenhang bestätigt werden (vgl. Tutz, 2011, S. 29):

$$\begin{aligned} f(y) &= \frac{\lambda^y}{y!} \exp(-\lambda) \\ &= \exp \left( \log \left( \frac{\lambda^y}{y!} \right) - \lambda \right) \\ &= \exp \left( \frac{y \log(\lambda) - \lambda}{1} - \log(y!) \right) \end{aligned}$$

mit  $\theta = \log(\lambda)$ ,  
 $b(\theta) = \exp(\theta) = \lambda \Rightarrow \mathbb{E}(y) = \lambda$  und  $\mathbb{V}(y) = \lambda$ ,  
 $\phi = 1 \Rightarrow$  Ein-Parameter-Verteilung,  
und  $c(y, \phi) = -\log(y!)$

Demzufolge ist die Verwendung eines generalisierten Modellansatzes in Zusammenhang mit der vorliegenden Poissonverteilung gerechtfertigt.

### Log-lineares Poisson-Modell

Es wird allgemein angenommen, dass die Zielgröße  $y_i$  Werte aus  $\{0, 1, 2, \dots\}$  annimmt und bei gegebenen Einflussgrößen  $x_{i1}, \dots, x_{ik}$  (bedingt) unabhängig ist, wobei der Index  $i$  die einzelnen Beobachtungseinheiten repräsentiert. Der Effekt der Einflussgrößen auf die Zielgröße wird hierbei als linear angenommen. Die Aufnahme von nicht-linearen Einflussgrößen wird im Rahmen der vorliegenden Datenlage und Modellierung nicht berücksichtigt. Dementsprechend wird das sogenannte log-lineare Poisson-Modell verwendet, welches als gängigste Variante des Poissonansatzes gilt. In diesem Modell für Zähldaten wird die Rate  $\lambda_i = \mathbb{E}(y_i | \mathbf{x}_i)$  der Poissonverteilung mit dem linearen Prädiktor  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$  über

$$\lambda_i = \exp(\eta_i) = \exp(\beta_0) \cdot \exp(\beta_1 x_{i1}) \cdot \dots \cdot \exp(\beta_k x_{ik})$$

bzw. in log-linearer Form durch

$$\log(\lambda_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}$$

verbunden. Der Effekt von Einflussgrößen auf die Rate  $\lambda$  ist demnach exponentiell-multiplikativ, während der Effekt auf die logarithmierte Rate linear ist (vgl. Fahrmeir *et al.*, 2009, S. 210). Da grundsätzlich die Wirkung auf die nicht-logarithmierte Rate  $\lambda$  intuitiver zu interpretieren ist, werden in der Regel die exponentiierten Effekte der Einflussgrößen betrachtet.

Ferner wird  $h$  als Responsefunktion (Antwortfunktion) und  $g = h^{-1}$  als Linkfunktion (Verknüpfungsfunktion) bezeichnet. Der Erwartungswert der Zielgröße wird dann mit Hilfe der Link-Funktion mit den Daten verbunden:

$$\begin{aligned} \mathbb{E}(y_i) &= \lambda_i = \mathbf{x}_i^T \boldsymbol{\beta} = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) \\ \Rightarrow g(\lambda_i) &= \mathbf{x}_i^T \boldsymbol{\beta} = \lambda_i \end{aligned}$$

Der Effekt der Kovariablen kann außerdem durch den linearen Prädiktor  $\eta$  erfasst werden. Dabei gilt der Zusammenhang  $\eta_i = g(\lambda_i)$  (vgl. Fahrmeir *et al.*, 2009, S. 190).

## Poissonverteilung mit Offset

Die Verwendung der Poissonverteilung kann darüber hinaus durch eine mögliche Standardisierung der Daten motiviert werden. So können beispielsweise die Altersstruktur, aber auch andere Kovariablen, innerhalb des Modells als sogenannter Offset ihre Berücksichtigung finden (vgl. Schmid, 2004, S. 9). Grundannahme ist dabei, dass in Landkreisen mit einer älteren Zusammensetzung der Einwohner grundsätzlich mehr Fälle einer bestimmten Krankheit zu erwarten sind und es somit einer Korrektur der tatsächlich beobachteten Anzahlen bedarf. Je nach Krankheit könnte diese Einschätzung jedoch auch gegenläufig ausfallen. Um die Relevanz des Offsets zu verdeutlichen, kann folgendes Beispiel herangezogen werden: Angenommen man möchte die Anzahl an Autos schätzen, die eine bestimmte Kreuzung passieren. Allein das reine Auszählen der vorbeifahrenden Autos ist in keinem Falle ausreichend, da zusätzlich die Länge eines jeweiligen Beobachtungsintervalls berücksichtigt werden muss. Bei einer Beobachtungsdauer von 10 Minuten ist in aller Regel eine niedrigere Anzahl an vorbeifahrenden Autos zu erwarten als bei 60 Minuten. Ein Offset der Zeit stellt in diesem Fall ein sinnvolles Korrekturverfahren dar.

Geht man im Rahmen des verwendeten Poisson-Modells davon aus, dass es sich um eine seltene nicht ansteckende Krankheit handelt (z. B. Krebs), dann kann die Anzahl an Erkrankungen näherungsweise durch eine Poissonverteilung mit der Intensität  $\lambda_s$  dargestellt werden (vgl. mit Dichte (1)), wobei der Index  $s$  die betrachteten Regionen repräsentiert. Um eine Vergleichbarkeit zwischen den beobachteten Häufigkeiten in den einzelnen Bereichen gewährleisten zu können, wird ein Offset  $e_s$  in Form des erwarteten Risikos pro Beobachtungseinheit in die Modellierung eingebunden, wodurch die Verteilungsannahme für  $y_s$  folgendermaßen erweitert wird (vgl. Fahrmeir und Kneib, 2010, S. 303 f.):

$$y_s \sim Po(e_s \lambda_s).$$

Sollte zwischen den Regionen kein Unterschied bezüglich des erwarteten Risikos vorliegen entspricht  $e_s = n_s \cdot \pi$ , wobei  $n_s$  die Einwohnerzahl der jeweiligen Beobachtungseinheit und  $\pi$  das durchschnittliche globale Risiko für eine einzelne Person zu erkranken darstellt. Liegen jedoch, beispielsweise in Bezug auf das Alter, strukturelle Unterschiede vor, so können diese über das erwartete Risiko  $e_s$  in die Modellierungen mit aufgenommen werden (Fahrmeir und Kneib, 2010, S. 304).

### 2.1.2 Gemischtes Modell

Gemischte Modelle berücksichtigen für den Prädiktor neben den bisher betrachteten festen Effekten der linearen Einflussgrößen zusätzliche zufällige Effekte (Random Effects). Klassischerweise werden gemischte Modelle bei der Analyse von wiederholten Messungen eingesetzt. Ein typisches Beispiel dafür ist die Messung des Blutdrucks im Tagesverlauf. Man vermutet, dass wiederholte Messungen an einer Person ähnlicher sind als die Messungen verschiedener Personen. Jedes Individuum weist als subjektspezifische Eigenschaft ein eigenes Level auf, um das die einzelnen Messungen schwanken. Für diesen Effekt wird eine eigene Verteilung angenommen (in der Regel eine Normal-

verteilung), da davon ausgegangen wird, dass diese Eigenschaft rein zufällig ist (vgl. Rieger, 2011, S. 49 f.). Neben den wiederholten Messungen lässt sich die Idee der subjektspezifischen Eigenschaften auch auf Cluster übertragen. In Bezug auf die vorliegenden Daten entsprechen diese den einzelnen Landkreisen. In einem gemischten Modell kann somit die Annahme, dass die einzelnen Beobachtungseinheiten einen spezifischen Effekt auf die Zielgröße aufweisen, passend dargestellt werden (vgl. Fahrmeir *et al.*, 2009, S. 253). Die Random Effects drücken letztlich die Heterogenität, d. h. die Uneinheitlichkeit der Landkreise aus, gegebenenfalls unter der Berücksichtigung von Einflussgrößen. Somit wird es möglich Landkreise zu entdecken, in welchen verhältnismäßig viele bzw. wenige Krankheitsfälle beobachtet wurden. Überdies ist es praktikabel im Rahmen der Bayes-Inferenz a priori eine andere Verteilungsannahme für die ursprünglich zufälligen Effekte zu treffen (vgl. Abschnitt 2.2). Den Daten kann beispielsweise eine räumliche Struktur zu Grunde liegen, die durch die Verwendung von Gauß-Markov-Zufallsfelder modelliert werden kann (vgl. Abschnitt 2.4). Im Zuge dessen erweist es sich als sinnvoll sowohl strukturierte als auch unstrukturierte Effekte in das Modell aufzunehmen (siehe hierfür Teilabschnitt aus 2.3.1).

### 2.1.3 Modellnotation

In diesem Abschnitt werden die eben vorgestellten Bestandteile des verwendeten Modells in einer Notation zusammengeführt. Für einen generalisierten Ansatz sind diesbezüglich sowohl die Verteilungsannahme als auch die Strukturannahme von zentraler und bestimmender Bedeutung. Wie bereits beschrieben, wird für die Verteilung der Zielgröße eine Poissonverteilung angenommen. Daraus resultiert bezüglich der Verteilungsannahme folgende Notation, während  $s$  die einzelnen Landkreise repräsentiert und den bisherigen Index  $i$  ersetzt:

$$y_s \mid \boldsymbol{x}_s, b_s \sim Po(\lambda_s)$$

Das verwendete GLMM setzt sich dabei aus den folgenden Komponenten zusammen:

1. Lineare Einflussgrößen:  $\boldsymbol{x}_s$
2. Landkreisspezifischer Intercept bzw. Random Effect:  $b_s$

Zusammengefasst kann folgende Strukturannahme für das Modell getroffen werden:

$$E(y_s \mid \boldsymbol{x}_s, b_s) = \lambda_s = \exp(\eta_s)$$

Mit dem Prädiktor:

$$\eta_s = \boldsymbol{x}_s^T \boldsymbol{\beta} + b_s$$

Zusätzlich wird für die Random Effects unter der Voraussetzung einer unabhängigen und identischen Verteilung eine Normalverteilung angenommen:

$$b_s \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$$

### 2.1.4 Zusammenfassung: Generalisiertes lineares gemischtes Modell

#### 1. Generalisierte Komponente:

- Zielgröße  $y_s$ : „Krankheitsfälle in einem Landkreis“ als Zählvorgang
- Poissonverteilung mit  $y_s \sim Po(e_s \lambda_s)$
- $\lambda_s \hat{=} \text{Intensitätsrate}$ ;  $e_s \hat{=} \text{erwartetes Risiko (Offset)}$
- Korrektur der Beobachtungen durch den Offset  $e_s$
- Poissonverteilung gehört zu der Familie der Exponentialverteilungen
- Generalisierte Modellierung dadurch unumgänglich bzw. ermöglicht
- Verwendung eines log-linearen Modellansatzes mit  $\lambda_s = \exp(\eta_s)$

#### 2. Lineare Komponente:

- Feste Effekte mit linearem Einfluss auf die Zielgröße
- Nicht-lineare Einflussgrößen werden nicht berücksichtigt

#### 3. Gemischte Komponente:

- Ermöglicht Aufnahme von landkreisspezifischen Effekten  
→ auch zufällige Effekte oder Random Effects genannt
- Allgemein mit Normalverteilungsannahme:  $b_s \stackrel{\text{i.i.d.}}{\sim} N(0, \sigma^2)$
- Random Effects drücken die Heterogenität (Uneinheitlichkeit) der Landkreise unter Berücksichtigung der Einflussgrößen aus
- Identifizierung individueller Eigenschaften einzelner Landkreise
- Modifiziert durch Annahme einer Nachbarschaftsstruktur  
→  $b_s \sim \text{Gauß-Markov-Zufallsfeld}$

## 2.2 Bayes-Inferenz

Durch die, in der Einleitung beschriebene, computationale Wende Anfang der 1990er Jahre hat gleichermaßen auch die Anwendung von simulationsbasierten statistischen Methoden im Rahmen der Bayes-Inferenz eine neue Entwicklung genommen. Solche Methoden sind von großem Interesse, da sich der bayesianische Ansatz von der Likelihood-Inferenz dadurch abgrenzt, „dass die unbekannten Parameter  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^T$  keine festen deterministischen Größen sind, sondern als zufällig angenommen werden und eine Priori-Verteilung besitzen“ (Fahrmeir *et al.*, 2009, S. 478). In diesem Zusammenhang gilt es die sogenannten Markov-Chain-Monte-Carlo (MCMC) Verfahren hervorzuheben. Diese computerintensiven Methoden ermöglichen es aus unbekannten (Posteriori-)Verteilungen Zufallszahlen zu ziehen, wodurch die Grundlage sämtlicher Inferenzschritte gebildet wird (vgl. Fahrmeir *et al.*, 2009, S. 478).

Die Betrachtung eines hierarchischen Modellaufbaus ist, unter anderem in Verbindung mit Gauß-Markov-Zufallsfeldern, von besonderem Interesse, um komplexe räumliche Zusammenhänge möglichst realitätsbezogen abbilden zu können. Rue und Held beschreiben diese Situation folgendermaßen:

„Complex hierarchical models are at the core of modern statistics, and [...] [Gaussian Markov random fields] play a central role in this framework to describe the spatial [...] dynamics of nature and real systems“ (Rue und Held, 2005, Vorwort S. 1).

Ferner soll an dieser Stelle erwähnt werden, dass die bayesianische Schätzung, im Vergleich zur Likelihood-Inferenz, das Potential bietet, die Gewichte der Gauß-Markov-Zufallsfelder individuell zu schätzen (vgl. hierfür Abschnitt 2.4.4). Die Entwicklung adaptiver Verfahren wird durch diesen bedeutenden Unterschied grundlegend geprägt.

Die Anwendung der Bayes-Inferenz wird somit durch die Möglichkeit einer hierarchischen Modellierung und der Einbeziehung von Priori-Wissen motiviert. In den anschließenden Abschnitten werden diesbezüglich zunächst die Grundlagen eines bayesianischen Modellaufbaus, basierend auf den Ansätzen nach Thomas Bayes, und die Zusammenhänge bis hin zur Posteriori-Verteilung erläutert (siehe Unterkapitel 2.2.1). Des Weiteren dient eine beispielhafte Einführung von hierarchischen Modellstrukturen in Zusammenhang mit sogenannten Hyperparametern einem besseren Verständnis der daran anschließenden Betrachtung der vollständig bedingten Dichten (vgl. Abschnitt 2.2.2). Passage 2.2.3 verschafft zusätzlich einen Überblick über mögliche Modellwahlkriterien im bayesianischen Kontext und soll die Verwendung des sogenannten Devianz Informations Kriteriums in Verbindung mit den adaptiven Anwendungen in Kapitel 3 motivieren. Die Zusammenfassung der wichtigsten Aspekte der Bayes-Inferenz im Kontext des adaptiven Disease Mappings (vgl. Abschnitt 2.2.4) vervollständigt schließlich dieses Unterkapitel.

### 2.2.1 Grundlegendes

Der ursprüngliche Satz von Bayes beschreibt den bedingten Zusammenhang zwischen den Wahrscheinlichkeiten der beiden Ereignissen  $A$  und  $B$  (vgl. Robert, 2007, S. 8 f.):

$$\mathbb{P}(A | B) = \frac{\mathbb{P}(B | A) \cdot \mathbb{P}(A)}{\mathbb{P}(B)}$$

Übertragen auf zwei stetige Zufallsvariablen  $x$  und  $y$  mit einer bedingten Verteilung  $p(y | x)$  und der Randverteilung  $p(x)$ , ist die bedingte Verteilung von  $x$  bei gegebenem  $y$  schließlich folgendermaßen definiert (vgl. Robert, 2007, S. 9):

$$p(x | y) = \frac{p(y | x) \cdot p(x)}{p(y)}$$

Betrachtet man nun nicht mehr die beiden Zufallsvariablen  $x$  und  $y$ , sondern den Parametervektor  $\boldsymbol{\theta}$  bei gegebenen Beobachtungen  $\mathbf{y} = (y_1, \dots, y_n)^T$ , so gilt folgende Proportionalität:

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}) \quad (2)$$

⇓

$$\text{Posteriori-Verteilung} \propto \text{Likelihood} \cdot \text{Priori-Verteilung}$$

Hierbei ist zu berücksichtigen, dass die Randverteilung  $p(\mathbf{y})$  unabhängig von den Parametern  $\boldsymbol{\theta}$  ist, wodurch die Betrachtung des proportionalen Zusammenhangs aus (2) in Bezug auf die Posteriori-Verteilung ausreicht (vgl. Fahrmeir *et al.*, 2009, S. 478).

### 1. Priori-Verteilung:

Wie bereits erwähnt, werden als Besonderheit der Bayes-Inferenz die unbekannten Parameter  $\boldsymbol{\theta}$  als zufällig angenommen. Um ein gewisses subjektives Vorwissen über die Parameter, beispielsweise in Bezug auf deren Verteilung, in die Modellierung aufzunehmen zu können, wird die Priori-Verteilung bestimmt, im Folgenden bezeichnet mit  $p(\boldsymbol{\theta})$ . Sollte kein spezielles Vorwissen über einen Parameter vorliegen, werden an dieser Stelle oftmals möglichst uninformative Verteilungen gewählt. Konsequenz einer solchen Vorgehensweise ist, dass die beobachteten Daten stärker in die Gewichtung einfließen (vgl. ebd.).

### 2. Likelihood:

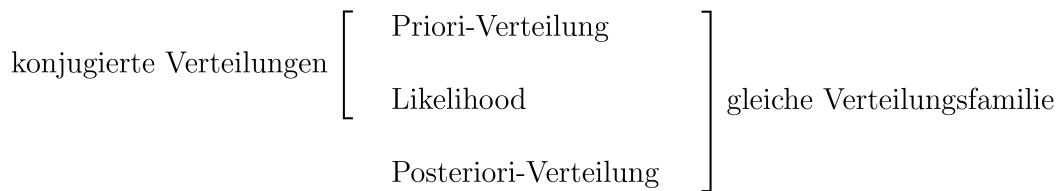
Durch die Hinzunahme der Likelihood werden eben diese Erkenntnisse der Daten  $\mathbf{y}$  unter der Berücksichtigung der unbekannten Parameter spezifiziert. Diesbezüglich muss korrekterweise erwähnt werden, dass es sich bei der Dichte der Verteilung  $p(\mathbf{y} | \boldsymbol{\theta})$  lediglich um eine zur Likelihood  $L(\boldsymbol{\theta})$  proportionalen Form handelt, die unter anderem als „Beobachtungsmodell“ bezeichnet wird (vgl. ebd.). In den folgenden Ausführungen soll jedoch der Ausdruck „Likelihood“ hinreichend genau sein.

### 3. Posteriori-Verteilung:

Die Posteriori-Verteilung beinhaltet schließlich eine Kombination aus dem a priori Wissen über den Parameter(vektor)  $\theta$  und den Erkenntnissen aus den beobachteten Daten in Form der Likelihood. Auf Basis dieser Verteilung können Punkt- und Intervallschätzer wie der Posteriori-Erwartungswert, der Posteriori-Median und der Posteriori-Modus sowie das Kredibilitätsintervall berechnet werden (vgl. Fahrmeir *et al.*, 2009, S. 480).

## Konjugierte Verteilungen

In vielen Fällen kann eine Posteriori-Verteilung nicht explizit berechnet werden, d. h. es ergibt sich eine unbekannte Verteilung, für die nicht ohne Weiteres die zuvor genannten Charakteristika bestimmt werden können. Wählt man allerdings die Priori-Verteilung und die Verteilung der Likelihood zueinander passend, so kann zumindest der dafür ausreichende Kern einer bekannten Verteilung gebildet werden. Aus mathematischer Sicht können Priori-Verteilungen demnach am besten bei der Bildung von Posteriori-Verteilungen eingebunden werden, wenn es sich um eine zueinander konjugierte (passende) Verteilungsfamilie handelt. Hierbei heißen eine Priori-Verteilung und eine Likelihood konjugiert, wenn die Posteriori-Verteilung und die Priori-Verteilung zur gleichen Verteilungsfamilie gehören (vgl. Cramer und Kamps, 2008, S. 312). Anschließendes Schaubild verdeutlicht diesen Zusammenhang:



**Abbildung 1:** Zusammenhang konjugierte Verteilungen und Verteilungen der gleichen Familie

Liegt eine solche Situation vor, kann die Posteriori-Verteilung verhältnismäßig einfach bestimmt werden. Beispiele für konjugierte Verteilungsklassen sind:

Verteilung	unbekannter Parameter $\theta$	konjugierte Verteilungsklasse
$Bin(n, \pi)$	$\pi$	$Be(a, b)$
$Geom(\pi)$	$\pi$	$Be(a, b)$
$\mathcal{N}(\mu, \sigma^2)$	$\mu$	$\mathcal{N}(\nu, \tau^2)$
$\mathcal{N}(\mu, \sigma^2)$	$\sigma^2$	$IG(a, b)$
$\mathcal{N}(\mu, \kappa = (\sigma^2)^{-1})$	$\kappa$	$G(a, b)$
$Po(\lambda)$	$\lambda$	$G(a, b)$
$Exp(\lambda)$	$\lambda$	$G(a, b)$

**Tabelle 1:** Paare konjugierter Verteilungen (vgl. Cramer und Kamps, 2008, S. 312)

Neben der Bestimmung von Posteriori-Verteilungen ist ferner die Verwendung konjugierter Verteilungen bei der Berechnung vollständig bedingter Dichten von besonderem Interesse, welche in Zusammenhang des hierarchischen Modellaufbaus des nächsten Unterkapitels beschrieben werden. Darüber hinaus finden konjugierte Verteilungen in Abschnitt 3.2 bei der Umsetzung der Bayes-Inferenz in Bezug auf Gauß-Markov-Zufallsfelder ihre Verwendung.

### 2.2.2 Hierarchische Modellstruktur

Ein hierarchischer Modellaufbau wird immer dann benötigt, wenn eine stochastische Abhängigkeit zwischen den unbekannten Parametern berücksichtigt werden soll. Im Rahmen der Betrachtung von Gauß-Markov-Zufallsfeldern ist ein dreistufiger Aufbau ausreichend, da bereits die unbekannten Hyperparameter der dritten Stufe ein Zufallsfeld näher bestimmen. Für die folgende beispielhafte Erläuterung soll ein Modell mit einer einfachen Zusammensetzung betrachtet werden. Für die Erklärung der Zielgröße  $\mathbf{y}$  wird demnach ausschließlich ein Gauß-Markov-Zufallsfeld  $\boldsymbol{\gamma}$  mit dem Hyperparameter  $\theta_{Hyp}$  in die Modellierung miteinbezogen. Einflussgrößen oder andere Effekte werden in diesem Beispiel nicht berücksichtigt. Es wird immer dann von einem Hyperparameter gesprochen, wenn eine Priori-Verteilung bis auf einen Parameter (den Hyperparameter) bestimmt ist. Das Feld  $\boldsymbol{\gamma}$  steht somit in einer Beziehung zu den Daten  $\mathbf{y}$ , für welche im Normalfall angenommen wird, dass diese bedingt unabhängig sind bei gegebenem  $\boldsymbol{\gamma}$  (für eine detailliertere Ausführung für Gauß-Markov-Zufallsfelder im hierarchischen bayesianischen Kontext vgl. Abschnitt 3.2). Im einfachsten Fall hängt jede Beobachtung lediglich von einem zugehörigen Element aus  $\boldsymbol{\gamma}$  ab. Das bedeutet, dass  $\mathbf{y}$  und  $\boldsymbol{\gamma}$  die gleiche Dimension haben. Demzufolge können die drei Stufen des hierarchischen Modellaufbaus folgendermaßen spezifiziert werden (vgl. Rue und Held, 2005, S. 8):

- 1.)  $\mathbf{y} \stackrel{\text{iid}}{\sim} p(\mathbf{y} | \boldsymbol{\gamma})$  Verteilung der Daten
- 2.)  $\boldsymbol{\gamma} \sim p(\boldsymbol{\gamma} | \theta_{Hyp})$  Verteilung der Gauß-Markov-Zufallsfelder
- 3.)  $\theta_{Hyp} \sim p(\theta_{Hyp})$  Verteilung der Hyperparameter

Für die Posteriori-Verteilung ergibt sich dann:

$$p(\boldsymbol{\gamma}, \theta_{Hyp} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\gamma}) \cdot p(\boldsymbol{\gamma} | \theta_{Hyp}) \cdot p(\theta_{Hyp})$$

Bei der Durchführung von Simulationen in einem bayesianischen Modell ist die Betrachtung der vollständig bedingten Dichten der einzelnen Parameter aus  $\boldsymbol{\theta}$  von elementarer Bedeutung (vgl. Abschnitt 3.2). Diese können durch eine proportionale Betrachtung der Posteriori-Verteilung in Bezug auf den jeweiligen Parameter gebildet werden. Die nähere Bedeutung der vollständig bedingten Dichten soll anschließend erläutert werden.

## Vollständig bedingte Dichten

Wie bereits beschrieben, basiert das Konzept der bayesianischen Inferenz auf der Annahme, dass die unbekannten Parameter  $\boldsymbol{\theta}$  als zufällig angenommen werden. In der Posteriori-Verteilung werden schließlich die Informationen über das a priori Wissen in Bezug auf diese Parameter sowie das Datenwissen vereint, weshalb jegliche Inferenzschlüsse auf diese gemeinsame Verteilung zurückgehen. Meist kann jedoch über den Verteilungstypen der Posteriori-Verteilung keine Aussage getroffen werden, wodurch die Bestimmung von Verteilungscharakteristika analytisch nur schwer möglich ist. Durch die Generierung von Zufallszahlen kann man dieser Problematik allerdings Herr werden (vgl. Fahrmeir *et al.*, 2009, S. 480).

In diesem Zusammenhang wurden bereits in den 1950er Jahren MCMC Verfahren entwickelt, die es ermöglichen aus unzugänglichen Verteilungen zufällige Zahlen zu ziehen. Da diese jedoch sehr rechenintensiv sind, hat erst die computationale Wende Anfang der 1990er Jahre eine breite Anwendungsmöglichkeit dieser Verfahren und somit der Bayes-Inferenz bewirkt (vgl. Fahrmeir *et al.*, 2009, S. 482). Oftmals ergibt sich jedoch das Problem, dass aus der Posteriori-Dichte nicht direkt Zufallszahlen gezogen werden können. Ist das der Fall, werden die vollständig bedingten Dichten der einzelnen Parameter aller hierarchischen Stufen des Modells benötigt (vgl. Fahrmeir *et al.*, 2009, S. 487). Vollständig bedingte Dichten (engl. full conditionals) können gleichermaßen sowohl einer bekannten Verteilung folgen als auch völlig unbekannt sein. Durch diese Begebenheit wurden verschiedenste Algorithmen (zusammengefasst unter der Bezeichnung „MCMC Verfahren“) entwickelt, die trotz allem Zufallsziehungen aus den vollständig bedingten Dichten und somit aus der Posteriori-Verteilung ermöglichen. Eine ausführliche Aufbereitung solcher Algorithmen ist Bestandteil des Unterkapitels 3.3.1. Zudem beinhaltet Abschnitt 3.2 diesbezüglich die Bestimmung der vollständig bedingten Dichten in Zusammenhang mit Gauß-Markov-Zufallsfeldern.

Für die Bildung der Posteriori-Verteilung bzw. der vollständig bedingten Dichten leistet das Konzept der bedingten Unabhängigkeit einen hilfreichen Dienst. In diesem Zusammenhang soll ein Zufallsvektor  $\mathbf{x} = (x_1, x_2, x_3)^T$  betrachtet werden, in dem  $x_1$  und  $x_2$  bedingt unabhängig sind, gegeben  $x_3$ . Bei einem bekannten Wert für  $x_3$  erkennt man in Bezug auf die Verteilung von  $x_1$ , dass  $x_2$  keinerlei relevante neue Informationen beinhaltet. Unter diesen Bedingungen setzt sich die gemeinsame Verteilung  $p(\mathbf{x})$  folgendermaßen zusammen (vgl. Rue und Held, 2005, S. 1):

$$p(\mathbf{x}) = p(x_1 | x_3) \cdot p(x_2 | x_3) \cdot p(x_3)$$

Was einer Vereinfachung der generellen Zusammensetzung entspricht:

$$p(\mathbf{x}) = p(x_1 | x_2, x_3) \cdot p(x_2 | x_3) \cdot p(x_3)$$

Diese verkürzte Darstellungsmöglichkeit von gemeinsamen Dichten verschafft einen direkten Zugang zu den jeweiligen Verteilungen.

### 2.2.3 Modellwahlkriterien

Für die verschiedenen Inferenzkonzepte existiert eine Vielzahl von, an den jeweiligen Modellaufbau angepassten, Modellwahlkriterien. Meist wird neben der Anpassung der jeweiligen Schätzung an die Daten die Komplexität des Modells berücksichtigt. Im Rahmen der Bayes-Inferenz ist die Posteriori-Verteilung für die Festlegung auf das „beste“ Modell von zentraler Bedeutung. Wie im vorhergehenden Abschnitt erläutert, kann diese entweder durch die Anwendung von MCMC Verfahren oder, wenn möglich, durch eine direkte analytische Berechnung bestimmt werden.

#### Bayes-Faktor über Posteriori-Wahrscheinlichkeiten

Für den Vergleich zweier Modelle wird in diesem Zusammenhang üblicherweise die Posteriori-Wahrscheinlichkeit herangezogen, um deren Güte verifizieren zu können. Analog zu der Zusammensetzung der Posteriori-Verteilung der (proportionalen) Gleichung in (2) von Seite 11, ergibt sich für die Posteriori-Wahrscheinlichkeit eines Modells  $M_j$  folgender proportionaler Zusammenhang (vgl. Fahrmeir *et al.*, 2009, S. 488):

$$p(M_j | \mathbf{y}) \propto p(\mathbf{y} | M_j) \cdot p(M_j) \quad \text{mit } j = 1, 2$$

Um die beiden Modelle wirklich vergleichen zu können, erweist sich die Betrachtung des Verhältnisses der jeweiligen Posteriori-Wahrscheinlichkeiten als sinnvoll (vgl. ebd.):

$$\frac{p(M_1 | \mathbf{y})}{p(M_2 | \mathbf{y})} = \frac{p(\mathbf{y} | M_1)}{p(\mathbf{y} | M_2)} \cdot \frac{p(M_1)}{p(M_2)}$$

Dieses Verhältnis verkürzt sich zu dem sogenannten Bayes-Faktor, sollte die a priori Wahrscheinlichkeit für die Wahl einer der beiden Modelle gleich sein. Für den anschließenden Bayes-Faktor gilt somit  $p(M_1) = p(M_2) = 1/2$ :

$$BF(\mathbf{y}) = \frac{p(\mathbf{y} | M_1)}{p(\mathbf{y} | M_2)}$$

Die Verwendung einer anderen a priori Wahrscheinlichkeit ist eher selten, wodurch dieser Faktor häufig zum Einsatz kommt. Dessen Ergebnis muss auf die Art und Weise interpretiert werden, dass bei einem Verhältnis größer als eins die Entscheidung zugunsten von Modell 1 und andernfalls zugunsten von Modell 2 gefällt wird (vgl. ebd.).

### Bayesianisches Informationskriterium (BIC)

Die Berechnung des Bayes-Faktors ist oftmals jedoch nicht ohne weiteres möglich. Aus diesem Grund findet das, dem Bayes-Faktor durch eine Laplace-Approximation angenäherte, bayesianische Informationskriterium (engl. Bayesian Information Criterion [BIC]) seine Anwendung. Das BIC verrechnet diesbezüglich die Log-Likelihood des ML-Schätzers über den Parametervektor  $\boldsymbol{\theta}$  mit  $l(\hat{\boldsymbol{\theta}})$  und die Komplexität des Modells repräsentiert durch die Anzahl der zu schätzenden Parameter  $p$  und der Stichprobengröße  $n$  (vgl. Fahrmeir *et al.*, 2009, S. 488 f.):

$$BIC = -2 \cdot l(\hat{\boldsymbol{\theta}}) + \log(n) \cdot p$$

Das Modell mit dem geringsten BIC im Vergleich zu allen anderen betrachteten Modellen hat schließlich approximativ die größte Posteriori-Wahrscheinlichkeit, weshalb dieses dann als geeignetes Modell gewählt werden sollte. Das BIC wird vergleichsweise jedoch selten eingesetzt. Ein Grund dafür ist, dass die Voraussetzungen für eine approximative Herleitung gerade in komplexeren Modellen meist nicht moderat erfüllt werden. Wird außerdem ein Modell mit MCMC Verfahren geschätzt, ist das BIC von Grund auf nicht anwendbar (vgl. Fahrmeir *et al.*, 2009, S. 489). Hinzu kommt, dass in komplexen hierarchischen Modellen mitunter uneigentliche Priori-Verteilungen verwendet werden, wie beispielsweise die intrinsischen Gauß-Markov-Zufallsfelder in Verbindung mit räumlichen Analysen (vgl. hierfür Abschnitt 2.4.4), für die der Bayes-Faktor bzw. das BIC nicht definiert sind (vgl. Gelman *et al.*, 2004, S. 190). In einem solchen Fall muss eine andere Methode für den Modellvergleich herangezogen werden.

### Bayesianisches Devianz Informationskriterium (DIC)

Das Auftreten der beschriebenen Problempunkte des BIC kann, besonders in Verbindung mit einer komplexen hierarchischen Modellstruktur, durch die Verwendung des Devianz Informations Kriteriums (engl. Deviance Information Criterion [DIC]) umgangen werden. Das von Spiegelhalter *et al.* (2002) begründete DIC findet deshalb eine breite Anwendung, da es „im Rahmen eines MCMC Algorithmus besonders einfach und schnell, quasi nebenbei berechenbar [ist]“ (Fahrmeir *et al.*, 2009, S. 489 f.)<sup>2</sup>. Die Grundlage für die Berechnung des DIC wird durch die Stichprobe der Zufallszahlen  $\theta^{(1)}, \dots, \theta^{(T)}$  gestellt, welche sich im Zuge der Anwendung eines MCMC Verfahrens ergeben. Ferner sind die sogenannte unstandardisierte Devianz

$$\begin{aligned} D(\boldsymbol{\theta}) &= -2 \cdot l(\boldsymbol{\theta}) \\ &= -2 \cdot \log[p(\mathbf{y} \mid \boldsymbol{\theta})] \end{aligned}$$

---

<sup>2</sup>Die gesamten Ausführungen bzgl. des DIC basieren auf Fahrmeir *et al.* (2009) (vgl. S. 489 f.).

sowie die effektive Anzahl an Parametern  $p_D$  des jeweiligen Modells von besonderem Interesse. Dabei kann  $p_D$  folgendermaßen geschätzt werden:

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$$

Wobei  $\overline{D(\boldsymbol{\theta})}$  die mittlere Posteriori-Devianz und  $D(\bar{\boldsymbol{\theta}})$  die Devianz, berechnet im Durchschnitt der gezogenen Zufallszahlen  $(\boldsymbol{\theta}^{(t)})$ , repräsentieren. Dementsprechend setzt sich  $p_D$  aus den Bestandteilen

$$\overline{D(\boldsymbol{\theta})} = \frac{1}{T} \sum_{t=1}^T D(\boldsymbol{\theta}^{(t)})$$

und

$$D(\bar{\boldsymbol{\theta}}) = D\left(\frac{1}{T} \sum_{t=1}^T \boldsymbol{\theta}^{(t)}\right)$$

zusammen. Insgesamt ergibt sich somit folgende Gleichung für die Bestimmung des DIC:

$$\text{DIC} = \overline{D(\boldsymbol{\theta})} + p_D = 2 \cdot \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}).$$

Bei der Anwendung des DIC ist zu berücksichtigen, dass die Ergebnisse bei unterschiedlichen MCMC Durchläufen nicht exakt gleich sein müssen, da diesem Kriterium lediglich eine Stichprobe aus der Posteriori-Verteilung zugrunde liegt. Besitzen zwei Modelle eine ähnlich gute Anpassungsgüte, kann es somit passieren, dass die Modellwahl basierend auf dem DIC von Lauf zu Lauf unterschiedlich bewertet wird.

### 2.2.4 Zusammenfassung: Bayes-Inferenz

#### 1. Grundlegendes zur Bayes-Inferenz:

- Basierend auf dem Satz von Bayes ergibt sich:

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) \cdot p(\boldsymbol{\theta})$$

$\Updownarrow$

Posteriori-Verteilung  $\propto$  Likelihood  $\cdot$  Priori-Verteilung

- Posteriori-Verteilung  $\hat{=}$  Verknüpfung von Vor- und Datenwissen
- Wahl von konjugierten Verteilungen vereinfacht die Bestimmung der Posteriori-Verteilung bzw. der vollständig bedingten Dichten

#### 2. Hierarchische Modellstruktur:

- Modellierte Abhängigkeiten zwischen unbekannten Parametern
- Komplexe Strukturen ergeben oft unbekannte Posteriori-Verteilungen  
→ MCMC Algorithmen ermöglichen das Ziehen von Zufallszahlen
- Oftmals schrittweise über die vollständig bedingten Dichten

#### 3. Modellwahlkriterien:

- Bayes-Faktor: Verhältnis von Posteriori-Wahrscheinlichkeiten  
→ Berechnung ist analytisch oft nur schwer zugänglich
- BIC: Annäherung an Bayes-Faktor durch Laplace-Approximation  
→ Nicht geeignet bei Anwendung von MCMC Verfahren
- DIC: Berechnung im MCMC Kontext ist schnell und einfach  
→ Modellwahlergebnisse können schwanken

## 2.3 Allgemeines zur räumlichen Statistik

Die Aufarbeitung der Theorie von generalisierten linearen gemischten Modellen und deren Modellierung über einen bayesianischen Ansatz hat hervorgebracht, dass die räumliche Struktur von Daten über die „gemischte“ Komponente des GLMMs und den Priorannahmen an die Gauß-Markov-Zufallsfelder im Rahmen der Bayes-Inferenz erfasst werden können. Bevor jedoch auf deren Zusammensetzung eingegangen werden kann, müssen vorerst einige Begrifflichkeiten der räumlichen Statistik erläutert werden. In diesem Zusammenhang ist es sinnvoll neben den räumlichen Effekten, die in strukturierte und unstrukturierte Elemente aufgeteilt werden können (siehe Abschnitt 2.3.1), die Beschaffenheit von räumlichen Daten zu beleuchten. Deren diskreter oder stetiger Aufbau (vgl. Unterkapitel 2.3.3) kann sich entweder auf ein reguläres oder irreguläres Gitter beziehen (siehe Abschnitt 2.3.2).

### 2.3.1 Räumliche Effekte – strukturiert und unstrukturiert

Im Zuge dessen gibt es verschiedene Möglichkeiten einen räumlichen Effekt in ein Modell mit aufzunehmen. Eine mögliche Variante ist ein semi-parametrischer Ansatz, bei dem in einem sogenannten geoadditiven Modell neben den Kovariablen  $x_1, \dots, x_k$  mit linearem und den Kovariablen  $z_1, \dots, z_q$  mit nicht-linearem Einfluss, zusätzlich die geografische Lokation  $s$  als Einflussgröße für das Modell berücksichtigt wird. Dieses wird somit erweitert um  $f_{geo}(s_i)$  und setzt sich folgendermaßen zusammen (vgl. Fahrmeir *et al.*, 2009, S. 407):

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + f_1(z_{i1}) + \dots + f_q(z_{iq}) + f_{geo}(s_i) + \varepsilon_i$$

Werden bei der Modellierung neben dem räumlichen Effekt ausschließlich lineare Einflussgrößen berücksichtigt, vereinfacht sich das Modell zu:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + f_{geo}(s_i) + \varepsilon_i$$

An dieser Stelle soll das sogenannte Kriging seine Erwähnung finden, welches als räumlicher Glätter ein geeignetes Verfahren zur Modellierung von glatten räumlichen Effekten darstellt (vgl. Fahrmeir *et al.*, 2009, S. 405). Für diese Art der Modellierung müsste das in Abschnitt 2.1 vorgestellte GLMM um nicht-lineare Komponenten zu einem geoadditiven Modell erweitert werden, während die zufälligen Effekte ausgespart werden könnten.

Eine andere Möglichkeit der Modellierung erlaubt die Erfassung von geografischen Strukturen eben über diese zufälligen Effekte des gemischten Modellansatzes. Vorteil dieser Interpretation ist eine potentielle Trennung in einen strukturierten und einen unstrukturierten räumlichen Effekt. Unterschieden wird dabei durch die Festlegung der Prior-Verteilungen der einzelnen Bestandteile im Rahmen der Bayes-Inferenz. Diese Vorgehensweise wird berücksichtigt bei der im Zuge dieser Arbeit verwendeten Modellierungen.

## Unstrukturierter räumlicher Effekt

Wie die Bezeichnung bereits eröffnet, werden für unstrukturierte räumliche Effekte keinerlei Strukturannahmen in Bezug auf die geografische Zusammensetzung der Daten getroffen. Unter diesem Verständnis können die im Rahmen des gemischten Modellansatzes vorgestellten Random Effects ihre Verwendung finden. Dieser einkalkulierte zufällige Effekt entspricht, übertragen auf eine räumliche Betrachtungsweise, einem spezifischen Effekt der jeweiligen Region (vgl. Abschnitt 2.1.2). In der Notation werden im Folgenden die Individuen  $i$  aus dem klassischen gemischten Modell durch die Regionen  $s$  ersetzt. Wie bereits in Unterkapitel 2.1.2 zu den gemischten Modellen beschrieben, beziehen sich somit die zufälligen Effekte nicht mehr auf die wiederholten Beobachtungen eines Individuums, sondern auf das spezielle Niveau der einzelnen Cluster, die in dem vorliegenden Fall den einzelnen Regionen entsprechen (vgl. Rieger, 2011, S. 53). Allgemein ergibt sich für ein generalisiertes lineares gemischtes Modell folgende Notation:

$$y_s = \mathbf{x}_s^T \boldsymbol{\beta} + \alpha_s$$

Die Random Effects sind hierbei unabhängig und identisch normalverteilt mit  $\alpha_s \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \tau^2)$  und können als Abweichung einer Beobachtungseinheit vom globalen Mittelwert aufgefasst werden. Die allgemeine Notation der zufälligen Effekte mit  $b_s$  aus Abschnitt 2.1.3 wird demnach ersetzt durch  $\alpha_s$ . Die Einführung dieser neuen, inhaltlich jedoch identischen, Bezeichnung soll die Unterscheidung zwischen unstrukturierten und strukturierten Effekten verdeutlichen.

## Strukturierter räumlicher Effekt

Indes wird bei den strukturierten räumlichen Effekten explizit eine Annahme für die Struktur zwischen den einzelnen Regionen  $s$  eines Gitters (regulär oder irregulär) angenommen. Eine gängige Variante ist hierbei eine Definition über die Priorannahme von Gauß-Markov-Zufallsfeldern, die in Abschnitt 2.4 detailliert erläutert wird. Im Rahmen des GLMMs wird für die ursprünglichen zufälligen Effekte  $b_s$  keine Normalverteilungsannahme mehr, sondern die eines Gauß-Markov-Zufallsfelds getroffen ( $\gamma \sim \text{Gauß-Markov-Zufallsfeld}$ ). Die Struktur wird dann durch eine sogenannte Nachbarschaftsmatrix repräsentiert (vgl. Fahrmeir und Kneib, 2010, S. 302). Für gewöhnlich kann durch den strukturierten räumlichen Effekt ein Großteil der Variabilität der Daten erklärt werden. Im Gegensatz dazu werden durch den unstrukturierten räumlichen Effekt meist lediglich größere Städte mit einer offensichtlich anderen Risikozusammensetzung aufgedeckt (vgl. Fahrmeir und Kneib, 2010, S. 323).

## Strukturiert und unstrukturierter räumlicher Effekt im Modell

Es ist immer dann sinnvoll strukturierte und unstrukturierte räumliche Effekte simultan zu berücksichtigen, wenn man zum einen davon ausgeht, dass die Lokalisierung

in dem Verbund bestimmter Regionen (z. B. repräsentiert durch eine Nachbarschaftsstruktur) einen Einfluss auf das Auftreten von Krankheiten oder anderen Zielgrößen hat und man zum anderen annimmt, dass aufgrund unbeobachteter Kovariablen zusätzlich regionenspezifische Effekte auftreten, die a priori nicht beschrieben werden können. Es muss jedoch allgemein berücksichtigt werden, dass eine strikte Trennung zwischen dem strukturierten und dem unstrukturierten Effekt charakteristischerweise nicht möglich ist, was zur Folge hat, dass bei simultaner Verwendung die Effekte ausschließlich in Kombination sinnvoll zu interpretieren sind. Für die Erfassung der gesamten geografischen Struktur, ist somit die Aufnahme beider Effekte durchaus erstrebenswert (vgl. Fahrmeir und Kneib, 2010, S. 303). Unterstellt man beispielsweise eine poissonverteilte Zielgröße und ein simples Regressionsmodell ohne jeglichen Einflussgrößen, so kann folgende kombinierte Zusammensetzung betrachtet werden (vgl. Fahrmeir und Kneib, 2010, S. 305):

$$y_s \sim Po(e_s \lambda_s) \quad \text{mit} \quad \lambda_s = \exp(\gamma_s + \alpha_s)$$

Dabei bezeichnet weiterhin  $\gamma_s$  den strukturierten oder auch räumlich korrelierten Effekt mit einer Gauß-Markov-Zufallsfeld Verteilungsannahme und  $\alpha_s$  den unstrukturierten Effekt einer Region  $s$  mit einer Normalverteilungsannahme, analog zu den Random Effects aus dem herkömmlichen gemischten Modell.

In ihrem Artikel über bayesianische Methoden im Zusammenhang mit Verfahren der Bildwiederherstellung beschreiben Besag *et al.* (1991) eben diese Unterteilung des räumlichen Effekts:

„The additional terms,  $u$  [(hier:  $\gamma_s$ )] and  $v$  [(hier:  $\alpha_s$ )], can be interpreted as surrogates for unknown or unobserved covariates; the  $u_i$ 's represent variables that, if observed, would display substantial spatial structure in that the values for a pair of contiguous zones would be generally much more alike than for two arbitrary zones, whereas the  $v_i$ 's represent unstructured variables“ (Besag *et al.*, 1991, S. 7).

Von noch größerem Interesse ist die Aufnahme beider Effekte für Daten mit räumlichen und zugleich zeitlichen Komponenten. Bei mehreren Beobachtungen über die Zeit kann dann das individuelle Level innerhalb einer Region durch eine solche Modellierung optimal erfasst werden (vgl. Fahrmeir und Kneib, 2010, S. 326). Im Zuge dieser Arbeit sind longitudinale Daten jedoch nicht von Bedeutung.

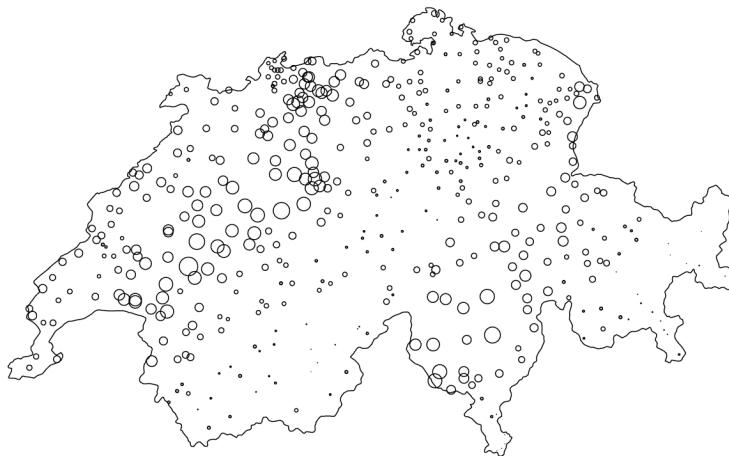
### 2.3.2 Räumliche Daten – diskret und stetig

Grundlage für die Analyse von räumlichen Zusammenhängen ist außerdem die Beschaffenheit der dafür notwendigen geografischen Informationen. Zentral ist hierbei typischerweise die Differenzierung darin, ob die Lokationsvariable  $s \in \{1, \dots, d\}$  stetig oder diskret erhoben wird, wobei die Menge  $D = \{1, \dots, s, \dots, d\}$  alle möglichen Gitterpunkte bezeichnet. Die weiteren Ausführungen in Kapitel 2 können mit dieser

Klassifizierung motiviert werden, da eine solche Unterscheidung die verwendeten Methoden und Modellierungen stark beeinflusst.

### Stetige räumliche Daten

Liegen die Informationen zu den Ortsangaben in Form von Koordinaten vor, dann repräsentiert  $s = \{s_1, s_2\}$ , bestehend aus den zwei Komponenten geografische Länge und geografische Breite, die Lage der jeweiligen Orte. Voraussetzung ist, dass sich  $s$  beliebig und somit stetig aus dem Raum  $D \subset \mathbb{R}^2$  zusammensetzen kann. In der Regel wird dieser auf einen regionalen Bereich beschränkt, auf dem sich die einzelnen Messpunkte verteilen. Eine für diesen Fall typische Situation ist die Betrachtung der Schweizer Regenfallmenge (siehe Abbildung 2) an vielen verschiedenen Messstationen (vgl. Fahrmeir und Kneib, 2010, S. 292 f.).

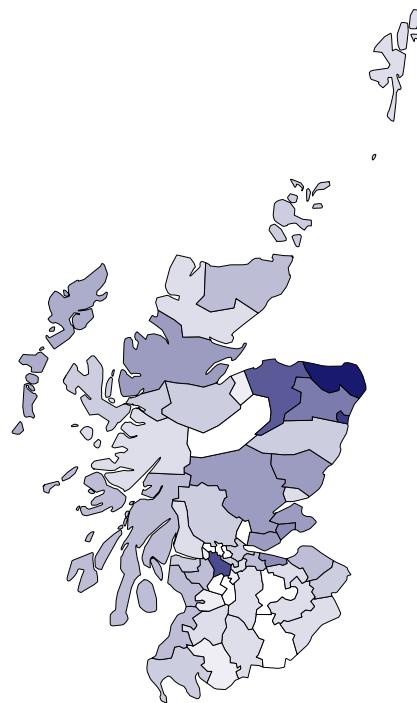


**Abbildung 2:** Stetige räumliche Daten: Schweizer Regenfallmengen (Grafik übernommen aus Fahrmeir und Kneib, 2010, S. 294)

Die Größe der einzelnen Kreise spiegelt hierbei die Regenmenge an den jeweiligen Messpunkten wider. In Zusammenhang von stetigen räumlichen Daten spricht man auch von räumlichen Punktprozessen (vgl. Schmid, 2004, S. 3).

### Diskrete räumliche Daten

Im diskreten Fall wird die Lokationsvariable  $s$  als räumlicher Index mit  $s \in \{1, \dots, d\}$  aufgefasst. Beispielsweise kann dieser Index den  $s$ -ten Pixel in einem regulären Gitter aus  $D$  Pixeln oder die  $s$ -te Region in einem irregulären Gitter aus  $D$  Regionen repräsentieren (vgl. Fahrmeir und Kneib, 2010, S. 293). Die ortsgebundenen Informationen liegen dann nicht exakt in Form von Koordinaten vor, sondern können lediglich den Clustern der jeweiligen Gitterform zugeordnet werden, in denen sie liegen. In den einzelnen Clustern werden die Beobachtungen dann aggregiert. Typischerweise ist die Anzahl an möglichen Regionen von Natur aus beschränkt (vgl. Fahrmeir und Kneib, 2010, S. 301).



**Abbildung 3:** Diskrete räumliche Daten: Beobachtete Todesfälle in Schottland verursacht durch Lippenkrebs abgetragen nach Distrikten

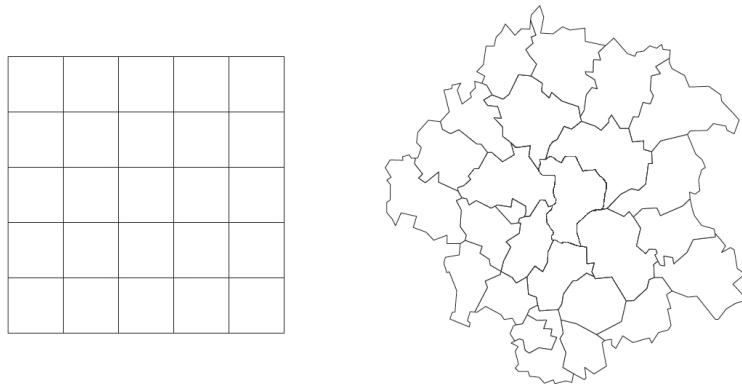
### Diskrete räumliche Daten $\Leftrightarrow$ Stetige räumliche Daten

Wie eben beschrieben, können räumliche Daten sowohl in stetiger als auch in diskreter Form vorliegen. Diese Trennung ist in der Praxis jedoch nicht so eindeutig, wie es zunächst erscheint. Es ist jederzeit eine Umformung in die jeweilige andere Datenform möglich. Betrachtet man beispielsweise diskrete räumliche Daten auf Landkreisebene in Deutschland, so besteht die Möglichkeit diese in eine stetige Form umzuwandeln, indem für jeden Landkreis repräsentativ die Koordinaten der zugehörigen Zentroide verwendet werden. Praktisch relevant wäre diese Vorgehensweise bei einer extrem großen Anzahl an Regionen, da die computationale Auslastung dadurch reduziert werden könnte. Gleichermassen kann eine Diskretisierung von stetigen räumlichen Daten durchgeführt werden. Eine Möglichkeit bildet die Überlagerung des Beobachtungsbereichs mit einem regulären oder irregulären Gitter. Die vorliegenden Erhebungen können dann den einzelnen Regionen des Gitters zugeordnet und darüber aggregiert werden. Eine andere Option wäre die Definition von Nachbarschaftsstrukturen anhand der Entferungen der Zentroide (vgl. Fahrmeir und Kneib, 2010, S. 305). Welche Rolle die Festlegung von Nachbarschaftsstrukturen einnimmt, wird ausführlich in Abschnitt 2.4.1 aufgearbeitet.

Wie zu Beginn des Unterkapitels bereits angedeutet, motiviert die diskrete bzw. stetige Form der räumlichen Daten den Modellierungsaufbau von  $f_{geo}$ . Bei stetigem  $s$  können Kriging oder ein Ansatz mit einer zweidimensionalen Basisfunktion als Glätter verwendet werden, wohingegen ein diskretes  $s$  die Anwendung von (Gauß-)Markov-Zufallsfeldern erfordert (vgl. Fahrmeir *et al.*, 2009, S. 407).

### 2.3.3 Räumliche Gitter – regulär und irregulär

Die Analyse von räumlichen Daten verlangt überdies die Auseinandersetzung mit den verschiedenen Formen der eigentlichen Beobachtungsfläche. Relevant ist die Unterscheidung zwischen einem regulären und irregulären Gitter, wobei für beide Formen gilt, dass sich die einzelnen Cluster bzw. Regionen, gekennzeichnet mit  $s$ , nicht überlappen. Die Cluster entsprechen den einzelnen Partitionen des jeweiligen Gitters. Reguläre Gitter zeichnen sich durch eine symmetrische und gleichmäßige Form aus. Sie werden beispielsweise eingesetzt bei bildgebenden Verfahren, wie dem sogenannten „Brain Mapping“. Dort entsprechen die Cluster den einzelnen Pixeln des Bildes. Unter irregulären Gittern versteht man die räumliche Anordnung von den Regionen  $s$ , die untereinander zumindest eine oder mehrere gemeinsame Grenzen haben oder in einem inhaltlichen Bezug zueinander stehen. Im klassischen Sinne werden diese Regionen häufig durch administrative Zuständigkeiten, wie etwa Landkreise oder Regierungsbezirke, bestimmt (vgl. Rue und Held, 2005, S. 16 f.). Abbildung 4 veranschaulicht den Unterschied dieser beiden Formen.



**Abbildung 4:** Illustrierung eines regulären (links) und eines irregulären (rechts) Gitters (Grafiken abgeändert übernommen aus Fahrmeir et al., 2009, S. 388)

Da in der vorliegenden Arbeit ausschließlich Methoden in Bezug auf diskrete räumliche Daten in irregulären Gittern betrachtet und angewendet werden, soll nach der anschließenden Zusammenfassung der wichtigsten Informationen des Unterkapitels, die theoretische Grundlage der Gauß-Markov-Zufallsfelder erläutert werden.

### 2.3.4 Zusammenfassung: Allgemeines zur räumlichen Statistik

#### 1. Räumliche Effekte – strukturiert und unstrukturiert:

- Räumlicher Effekt als Komponente des GLMMs
- Unstrukturierter räumlicher Effekt:  $\alpha_s \sim \mathcal{N}(0, \tau^2)$
- Strukturierter räumlicher Effekt:  $\gamma_s \sim \text{Gauß-Markov-Zufallsfeld}$
- Kombination im log-linearen Poissonmodell:  $\lambda_s = \exp(\gamma_s + \alpha_s)$

#### 2. Räumliche Daten – diskret und stetig:

- Stetige Daten:  $s = \{s_1, s_2\} \hat{=} \text{Koordinaten der Beobachtung}$
- Diskrete Daten:  $s \in \{1, \dots, d\} \hat{=} \text{räumlicher Index}$   
→ Beobachtungen werden über die Regionen  $s$  aggregiert
- Relevant für vorliegende Modellierung: diskrete räumliche Daten

#### 3. Räumliche Gitter – regulär und irregulär:

- Reguläres Gitter:  $s \hat{=} \text{z. B. Pixel eines Bilds}$
- Irreguläres Gitter:  $s \hat{=} \text{z. B. administrative Regionen einer Landkarte}$
- Relevant für vorliegende Modellierung: irreguläres Gitter

⇒ Modellierung der räumlichen Struktur über **Gauß-Markov-Zufallsfelder**

## 2.4 Gauß-Markov-Zufallsfelder (GMZF)

Durch den hierarchischen Bayes-Ansatz aus Abschnitt 2.2 wird es über die Priori-Verteilungen ermöglicht bestimmte priori Annahmen in die Modellierung mit aufzunehmen. Im Kontext einer räumlich strukturierten Datenlage ist dabei die Verwendung von Markov-Zufallsfeldern und explizit von Gauß-Markov-Zufallsfeldern ein beliebtes und probates Mittel, um die Nachbarschaftsstruktur von Beobachtungseinheiten angemessen abbilden zu können. Die Grundannahme besteht diesbezüglich darin, dass die Konstellation von räumlichen Einheiten in einem Gitternetz mit ausschlaggebend für das Ausmaß der Zielgröße ist, wobei angenommen wird, dass sich die jeweilige Ausprägung von benachbarten Einheiten nicht besonders stark voneinander unterscheidet (vgl. Schmid, 2004, S. 10). Im Folgenden wird dabei für den Ausdruck „Gauß-Markov-Zufallsfeld“ die Abkürzung GMZF verwendet. Um eine bessere Lesbarkeit zu gewährleisten, ersetzt diese verkürzte Bezeichnung sowohl die Singular- als auch die Pluralform sowie alle weiteren grammatischen Formen. In der englischsprachigen Literatur wird vorzugsweise die Abkürzung GMRF für „Gaussian Markov random field“ eingesetzt.

Rue und Held (2005) geben in ihrem Buch über GMZF einen umfangreichen Überblick über deren Zusammensetzung und Möglichkeiten. Sie beschreiben einleitend ein GMZF auf folgende Art und Weise:

„A GMRF is really a simple construct: It is just a (finite-dimensional) random vector following a multivariate normal (or Gaussian) distribution“  
(Rue und Held, 2005, S. 1).

Inhalt der nachkommenden Ausführungen soll schließlich sein, die einzelnen methodischen Bestandteile von GMZF vorzustellen und zu motivieren. Es muss zunächst eine Grundlage in Bezug auf Nachbarschaften und deren Darstellung über ungerichtete Graphen geschaffen werden (vgl. Abschnitt 2.4.1), um dann in Unterkapitel 2.4.2 auf die Verallgemeinerung der eindimensionalen Markov-Eigenschaft aus dem zeitlichen Kontext hin zu einer räumlichen Markov-Eigenschaft eingehen zu können. Dabei werden nicht mehr die Zustände einer eindeutigen zeitlichen Abfolge als Bedingung der Markov-Eigenschaft herangezogen, sondern die Zustände aus der jeweiligen Nachbarschaft einer räumlichen Einheit. Die gemeinsame Verteilung der GMZF kann dann über die bedingte Verteilung der einzelnen Bestandteile des jeweiligen Zufallsfelds bestimmt werden. Ein Zufallsfeld bezeichnet dabei einen mehrdimensionalen Zufallsprozess, der durch die Verbindung einzelner Zufallsvariablen bestimmt wird (vgl. Meier und Borkowski, 2011, S. 25). Durch die Normalverteilungsannahme der vollständig bedingten Dichten, ergibt sich die von Rue und Held beschriebene multivariate Normal- oder auch Gauß-Verteilung für das Markov-Zufallsfeld. Ferner sind auch andere Verteilungsannahmen denkbar, die jedoch kein Bestandteil der vorliegenden Arbeit sein sollen.

### 2.4.1 Nachbarschaftsstrukturen über ungerichtete Graphen

Aus der Erhebung von diskreten räumlichen Daten ergibt sich die Schwierigkeit, dass keine Messung der Entfernung zwischen den Regionen bzw. Messpunkten mehr möglich ist. Um trotzdem die Anordnung der einzelnen Regionen modellieren zu können, wird

bei der Verwendung von Zufallsfeldern die Nachbarschaftsstruktur des jeweiligen Beobachtungsgitters in das Modell mit aufgenommen. In diesem Zusammenhang muss zunächst einmal festgelegt werden, wann zwei Regionen als Nachbarn bezeichnet werden. Die von Rue und Held (2005) definierten ungerichteten Graphen spiegeln außerdem die bedingten Unabhängigkeitsbeziehungen zwischen den einzelnen Nachbarn wider, die zentral sind für das Verständnis der weiteren Modellierung und der Zusammensetzung der Verteilung von GMZF.

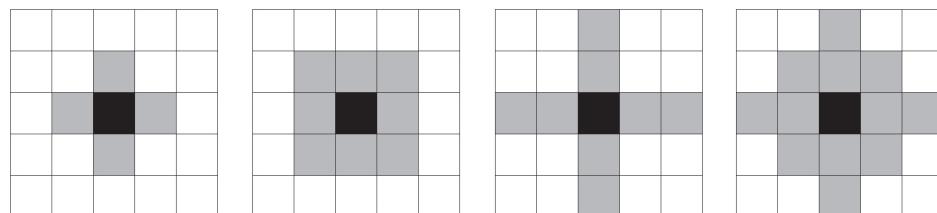
### Definition von Nachbarschaft

Unter dem Begriff „Nachbarschaft“ würde man intuitiv Regionen mit einer gemeinsamen geografischen Grenze als benachbart einstufen. In diesem Fall ist von einer direkten Nachbarschaft 1. Ordnung die Sprache, welche in Abbildung 5 beispielhaft für ein reguläres bzw. irreguläres Gitter aufgezeigt wird. Die Nachbarn der schwarz gekennzeichneten Region sind jeweils grau hervorgehoben.



**Abbildung 5:** Nachbarschaften 1. Ordnung bei regulärem Gitter (links) und bei irregulärem Gitter (rechts) (Grafiken übernommen aus Fahrmeir et al., 2009, S. 388)

Zusätzlich kann die Annahme getroffen werden, dass nicht nur direkte Nachbarn, sondern auch komplexere Zusammenhänge für die Abbildung der Nachbarschaftsstruktur einer Region  $s$  relevant sein könnten. Für diesen Fall beachtet man für die Modellierung auch Nachbarschaften höherer Ordnung. Unter anderem sind folgende Konstellationen bezüglich eines regulären Gitters möglich:



**Abbildung 6:** Nachbarschaften höherer Ordnung in einem regulären Gitter (Grafiken abgeändert übernommen aus Schmid, 2013, S. 32)

Von links nach rechts betrachtet illustrieren die Grafiken eine Nachbarschaft 1. Ordnung, eine 2. Ordnung, eine diagonale Nachbarschaft mit einer Nachbarschaft 1. Ordnung

nung sowie eine diagonale Nachbarschaft mit einer Nachbarschaft 2. Ordnung (vgl. Schmid, 2013, S. 32). Ferner ist an dieser Stelle festzuhalten, dass eine Einheit weder in einem regulären noch in einem irregulären Gitter ein Nachbar von sich selbst sein kann.

Je nach Datenlage kann es durch inhaltliche Überlegung sinnvoll sein Nachbarschaften anders zu definieren. Bei stetigen räumlichen Daten könnte die Nachbarschaftsstruktur beispielsweise über Umkreise um die einzelnen Beobachtungspunkte mit einem festen Radius definiert werden. All jene Beobachtungen, die innerhalb des jeweiligen Kreises liegen, gelten dann als benachbart zu der Beobachtung im Mittelpunkt (vgl. Fahrmeir und Kneib, 2010, S. 306). Bei der Analyse der Ausbreitung von Infektionskrankheiten muss in Bezug auf die Definition der Nachbarschaft der zusätzliche Aspekt berücksichtigt werden, dass sich die Krankheiten in der Regel nicht nur über direkte Grenzen verbreiten können. Regionen mit einer direkten Flugverbindungen sollten unter diesen Umständen ebenfalls als benachbart spezifiziert werden. Im Folgenden werden jedoch ausschließlich direkte Nachbarschaftsstrukturen 1. Ordnung für irreguläre Gitter in Betracht gezogen. Es kann allgemein festgehalten werden, dass die Definition des Nachbarschaftsbegriffs keinen Einfluss auf die Art des Modellierungsaufbaus und der Konstruktion von GMZF hat.

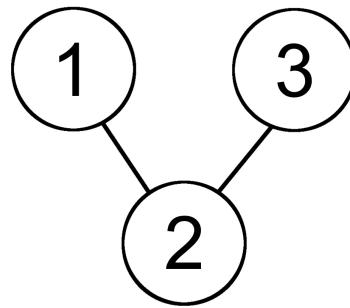
Für eine deutlichere Erfassung der weiteren Ausführungen müssen einige generelle Notationen und Abkürzungen eingeführt werden. Sind per Definition die beiden Regionen  $s$  und  $r$  Nachbarn, dann beschreibt  $s \sim r$  diese Begebenheit. Des Weiteren verkörpert  $\delta_r$  alle Indizes der von  $s$  benachbarten Regionen ohne der Region  $s$  selbst. Wird außerdem ein Vektor mit dem Index  $-s$  ausgewiesen, so beschreibt dieser die Menge aller betrachteten Regionen abzüglich der Region  $s$ .

## Ungerichtete Graphen

Ein zentraler Aspekt für das Verständnis von GMZF ist die bedingte Unabhängigkeit in Zusammenhang mit der Verteilung der räumlichen Effekte. Es ist bekannt, dass zwei Zufallsvariablen  $x$  und  $y$  unabhängig sind, sobald der folgende Zusammenhang  $\pi(x, y) = \pi(x) \pi(y)$  gilt; kurz geschrieben als  $x \perp y$ . Diese sind bedingt unabhängig bei gegebenem  $z$ , wenn  $\pi(x, y | z) = \pi(x | z) \pi(y | z)$  Gültigkeit hat; kurz  $x \perp y | z$ . Das Konzept der bedingten Unabhängigkeit kann außerdem problemlos auf den multivariaten Fall übertragen werden (vgl. Rue und Held, 2005, S. 17 f.). Für die Modellierung des Effekts eines GMZF wird dessen gemeinsame Verteilung benötigt, die auf dem Prinzip der bedingten Unabhängigkeit basiert. Man nimmt dabei an, dass die Verteilung des räumlichen Effekts einer Region  $s$  bedingt wird durch die räumlichen Effekte aller benachbarten Regionen. Die genauere Zusammensetzung der Verteilung der räumlichen Effekte von GMZF wird schließlich in Unterkapitel 2.4.3 näher beleuchtet.

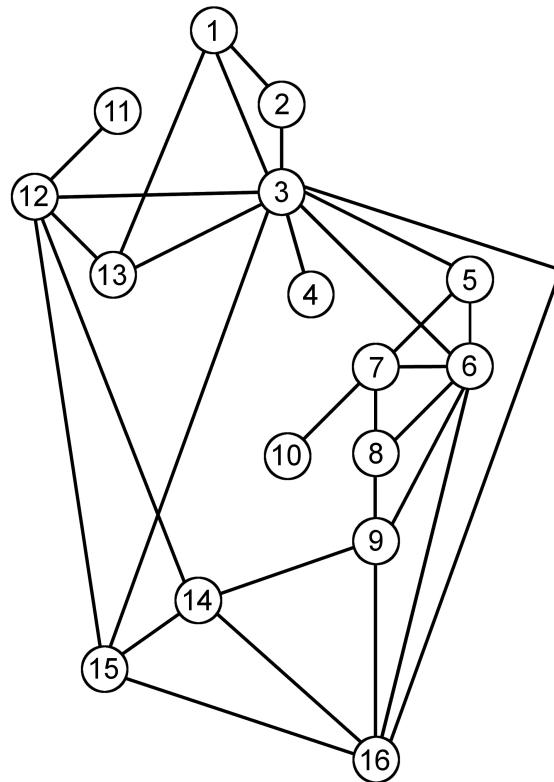
Um die bedingte Unabhängigkeitsstruktur eines Nachbarschaftssystems in Zusammenhang mit GMZF repräsentieren zu können, definieren Rue und Held in ihrem Buch über GMZF die Klasse der ungerichtete Graphen. Demnach entspricht ein ungerichteter Graph  $\mathcal{G}$  einem Tupel aus den Mengen  $\mathcal{V}$  und  $\mathcal{E}$ , d. h.  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ . Die Menge  $\mathcal{V}$  beinhaltet hierbei alle Knotenpunkte eines ungerichteten Graphens. In den vorliegenden Betrachtungen von irregulären Gittern entsprechen diese beispielsweise den Regio-

nen auf einer Landkarte, wie etwa den 16 Bundesländern von Deutschland. Ebenfalls in Form von Tupeln umfasst zusätzlich die Menge  $\mathcal{E}$  alle ungerichteten Verbindungen zwischen zwei Knotenpunkten aus  $\mathcal{V}$ . Stehen dabei die Knotenpunkte 1 und 2 in einer direkten Verbindung (siehe Abbildung 7), dann ist das Tupel  $\{1, 2\}$  Bestandteil der Menge  $\mathcal{E}$ . Im Rahmen der Betrachtungen dieser Arbeit ist eine direkte Verbindung zwischen zwei Knotenpunkten als direkte Grenze definiert.



**Abbildung 7:** Beispiel für einen ungerichteten Graphen (Grafik nachempfunden: Rue und Held, 2005, S. 18)

Bestimmt man die Teilmengen von  $\mathcal{G}$  aus Abbildung 7 vollständig nach dieser Definition, dann ergibt sich für  $\mathcal{V} = \{1, 2, 3\}$  und für  $\mathcal{E} = \{\{1, 2\}, \{2, 3\}\}$  (vgl. Rue und Held, 2005, S. 18). Analog kann ein Nachbarschaftsgraph für die 16 deutschen Bundesländer aufgestellt werden:



**Abbildung 8:** Graph der 16 Bundesländer von Deutschland (Grafik nachempfunden: Rue und Held, 2005, S. 5)

Nach der eben beschriebenen Definition von ungerichteten Graphen, sollen anhand des Nachbarschaftsgraphen der 16 deutschen Bundesländer aus Abbildung 8 beispielhaft dessen Bestandteile  $\mathcal{V}$  und  $\mathcal{E}$  bestimmt werden. Die Menge der Knoten setzt sich aus den 16 verschiedenen Bundesländern zusammen, die jeweils einen Knotenpunkt repräsentieren, d. h.  $\mathcal{V} = \{1, 2, \dots, 16\}$ . Die Verbindungen der einzelnen Knotenpunkte symbolisieren zudem eine gemeinsame Grenze von benachbarten Bundesländern. Somit ergibt sich für die Menge der Verbindungen zwischen den Knoten:

$$\mathcal{E} = \{\{1, 2\}, \{1, 3\}, \{1, 13\}, \{2, 3\}, \{3, 4\}, \{3, 5\}, \{3, 6\}, \{3, 12\}, \{3, 13\}, \{3, 15\}, \{3, 16\}, \{5, 6\}, \{5, 7\}, \{6, 7\}, \{6, 8\}, \{6, 9\}, \{6, 16\}, \{7, 8\}, \{7, 10\}, \{8, 9\}, \{9, 14\}, \{9, 16\}, \{11, 12\}, \{12, 13\}, \{12, 14\}, \{12, 15\}, \{14, 15\}, \{14, 16\}, \{15, 16\}\}$$

Durch die Verbindungen der Menge  $\mathcal{E}$  und der Abbildung 8 zeichnet sich bereits eine Nachbarschaftsstruktur in Bezug auf die 16 deutschen Bundesländer ab. Für eine nähere Betrachtung wird dieses Beispiel in Abbildung 13 auf Seite 41 nochmals aufgegriffen.

#### 2.4.2 Räumliche Markov-Eigenschaft

Ein besonderes Kennzeichen der Gauß-Markov-Zufallsfelder ist deren Markov-Eigenschaft. Die Zusammensetzung der GMZF kann daher durch die Betrachtung von autoregressiven Modellen motiviert werden. Im Allgemeinen basieren Modelle dieser Art auf einer Markov-Eigenschaft mit zeitlichem Aspekt.

#### Autoregressiver Prozess und zeitliche Markov-Eigenschaft

Ein oft genanntes Beispiel ist in diesem Kontext ein autoregressiver Prozess 1. Ordnung, für den im Folgenden der Index  $t$  die Zeit repräsentiert (vgl. Rue und Held, 2005, S. 1):

$$y_{t+1} = \phi y_t + \varepsilon_{t+1}, \quad \text{mit } |\phi| < 1. \quad (3)$$

Diese Darstellung ermöglicht es eine, um den Faktor  $\phi$  gewichtete, zeitliche Abhängigkeitsstruktur zu modellieren. Eigenschaften bezüglich des Störterms  $\varepsilon_{t+1}$  sollen an dieser Stelle unkommentiert bleiben, da in Unterkapitel 2.4.3 näher darauf eingegangen wird. Die zukünftige Entwicklung  $t + 1$  hängt in diesem Fall von der Gegenwart  $t$  ab. Die Beobachtungen der Vergangenheit spielen unter diesen Umständen für  $y_{t+1}$  keine Rolle mehr (vgl. Meinstrup und Schäffler, 2005, S. 227). Meinstrup und Schäffler halten allgemein Folgendes fest:

„Die Kunst der stochastischen Modellbildung besteht darin, genug Abhängigkeit zu berücksichtigen, um ein realistisches Modell zu erhalten, jedoch nicht zu viel, um explizite Berechnungen zu ermöglichen. Markov-Ketten bilden einen solchen Kompromiss“ (Meinstrup und Schäffler, 2005, S. 227).

Markov-Ketten bilden aufgrund ihrer diskreten Aufnahme von Zeitpunkten einen Spezialfall der Markov-Prozesse. Dieser Ansatz soll jedoch den Anforderung der anschließenden räumlichen Umsetzung genügen. Zudem ist der Umgang mit Markov-Ketten im Vergleich wesentlich weniger kompliziert. Allgemein zeichnet sich in Bezug auf die bedingte Dichte eines zum Zeitpunkt  $t+1$  auftretenden Zustands  $j$  die Markov-Eigenschaft von zeitlich diskreten Ketten folgendermaßen aus (vgl. Meinstrup und Schäffler, 2005, S. 229):

$$p(Y_{t+1} = j_{t+1} | Y_t = j_t, Y_{t-1} = j_{t-1}, \dots, Y_0 = j_0) = p(Y_{t+1} = j_{t+1} | Y_t = j_t) \quad (4)$$

Die Wahrscheinlichkeit für einen zukünftigen Zustand  $j$  zum Zeitpunkt  $t+1$  ist demzufolge lediglich durch den Zustand des gegenwärtigen Zeitpunkts  $t$  bedingt. Ältere Zustände sind in diesem Zusammenhang irrelevant. Diese zeitliche Eigenschaft kann nun für den räumlichen Fall verallgemeinert werden.

### Verallgemeinerung zur räumlichen Markov-Eigenschaft

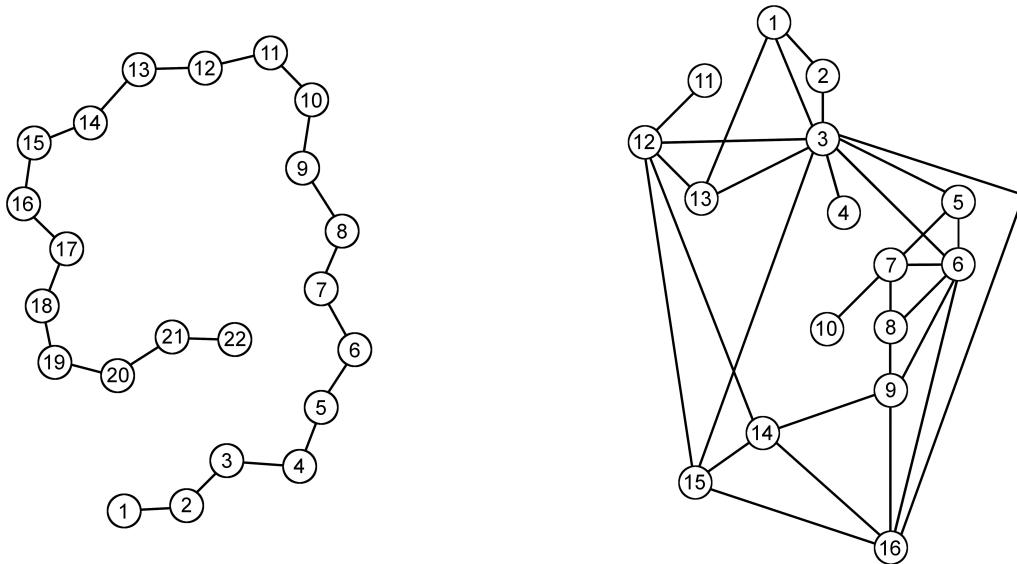
Analog zu dem Aufbau des autoregressiven Prozesses ist die Grundidee der GMZF, dass der räumliche Effekt von benachbarten Regionen nicht zu stark voneinander abweicht. Die Verallgemeinerung der zeitlichen Markov-Eigenschaft erfolgt nun in dem Sinne, dass der gegenwärtige Zustand, der die Wahrscheinlichkeit des zukünftigen Zustands bedingt (vgl. die bedingte Dichte in (4)), durch die Nachbarn der Region  $s$  ersetzt wird. Diesbezüglich hält Besag (1974) nachstehendes fest:

„Any system of  $n$  [(here:  $d$ )] sites, each with specified neighbours, clearly generates a class of valid stochastic schemes. We call any member of this class a Markov field“ (Besag, 1974, S. 197).

Darüber hinaus wird im folgenden Kontext ein Zustand durch den räumlichen Effekt der jeweiligen Region beschrieben – repräsentiert durch  $\gamma_s$ . Wie bereits im Rahmen der Definition von Nachbarschaft in Unterkapitel 2.4.1 erläutert, stellt die Region  $r$  einen Nachbarn der Region  $s$  dar und  $\delta_r$  verkörpert die Indizes aller Nachbarn der Region  $s$ . An dieser Stelle ist noch einmal festzuhalten, dass eine Region nicht mit sich selbst benachbart sein kann, das bedeutet es gilt  $s \neq r$ . Die zugehörige bedingte Dichte kann somit folgendermaßen formuliert werden (vgl. Fahrmeir *et al.*, 2009, S. 389 f.):

$$p(\gamma_s | \boldsymbol{\gamma}_{-s}, s \in D) = p(\gamma_s | \boldsymbol{\gamma}_{\delta_r}, r \in D).$$

Wie bei einer eindimensionalen, zeitlichen Markov-Kette, lässt sich auch hier erkennen, dass die Wahrscheinlichkeit für den räumlichen Effekt der Region  $s$  lediglich auf die Effekte der Nachbarregionen bedingt ist. Regionen, die in keiner Nachbarschaftsbeziehung zu  $s$  stehen, haben dahingegen keinen Einfluss. Der Unterschied zwischen einem zeitlichen und einem räumlichen Prozess kann folgendermaßen illustriert werden:



**Abbildung 9:** Linearer Graph eines autoregressiven Prozesses 1. Ordnung (links) und Graph der 16 Bundesländer von Deutschland (rechts) (Grafiken nachempfunden: Rue und Held, 2005, S. 5)

Der autoregressiver Prozess aus Gleichung (3) kann schließlich auf eine mehrdimensionale Betrachtungsweise übertragen werden. Analog zu  $\phi$  charakterisiert  $\beta_{sr}$  die Gewichtung einer Nachbarschaft zwischen  $s$  und  $r$  (vgl. Schmid, 2013, S. 33):

$$\gamma_s = \sum_{r \in \delta_s} \beta_{sr} \gamma_r + \varepsilon_s \quad (5)$$

Aufgrund der Darstellungsmöglichkeit über Markov-Ketten liegen verschiedene räumliche Markov-Eigenschaften in Bezug auf die GMZF vor. Diese sollen im folgenden Abschnitt veranschaulicht werden.

### Räumliche Markov-Eigenschaft der Gauß-Markov-Zufallsfelder

Im Folgenden wird angenommen, dass der Vektor  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_s, \dots, \gamma_d)^T$  ein GMZF in Bezug auf den ungerichteten Graphen  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  darstellt (vgl. Teilabschnitt aus 2.4.1 zu ungerichteten Graphen). Um eine bedingte Unabhängigkeitsstruktur zwischen den Beobachtungen der Region  $i$  und der Region  $j$  überprüfen zu können, ist es erforderlich, dass förmlich gezeigt wird, ob der Zusammenhang  $\gamma_i \perp \gamma_j | \boldsymbol{\gamma}_{-ij}$  gilt; respektive muss dargelegt werden, ob  $\gamma_i$  und  $\gamma_j$  bedingt unabhängig sind, gegeben allen restlichen Beobachtungen des GMZF. Nach der in spätere Betrachtungen einbezogenen Äquivalenzrelation (13) von Seite 39 entspricht das der Überprüfung, ob die zugehörigen Diagonaleinträge der Präzisionsmatrix  $Q_{ij}$  der GMZF ungleich Null sind. Die ungerichteten Graphen erweisen in diesem Kontext einen wertvollen Dienst, da sich ein Graph  $\mathcal{G}$  aus dem von Null verschiedenen Bereich von  $\mathbf{Q}$  zusammensetzt und somit die bedingte Unabhängigkeitsstruktur abbildet. Eine nützliche Eigenschaft der GMZF ist in diesem Zusammenhang, dass aus den ungerichteten Graphen, in Bezug

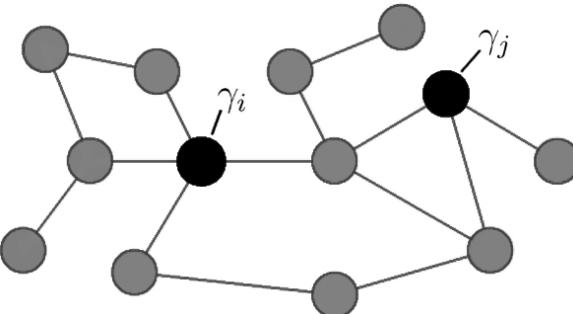
auf die bedingte Unabhängigkeit, zusätzliche Informationen extrahiert werden können. Diesbezüglich ergibt die folgende Betrachtung der paarweisen (a), der lokalen (b) und der globalen (c) Markov-Eigenschaft in Verbindung mit dem Graphen  $\mathcal{G}$ , dass diese drei Eigenschaften für GMZF äquivalent sind (vgl. Rue und Held, 2005, S. 24 f.):

(a) Paarweise Markov-Eigenschaft:

$$\gamma_i \perp \gamma_j \mid \boldsymbol{\gamma}_{-ij}, \quad \text{wenn } \{i, j\} \notin \mathcal{E} \quad \text{und} \quad i \neq j.$$

Diese Eigenschaft beschreibt den Zusammenhang zwischen  $\gamma_i$ ,  $\gamma_j$  und allen restlichen räumlichen Effekten  $\boldsymbol{\gamma}_{-ij}$ , dargestellt durch die Knotenpunkte der Menge  $\mathcal{V}$ . Hierbei ist zentral, dass die Knotenpunkte von  $i$  und  $j$  zwar Bestandteil der Menge  $\mathcal{V}$ , jedoch nicht benachbart sind ( $\{i, j\} \notin \mathcal{E}$ ), weshalb in Abbildung 10 keine Verbindungsgeraden zwischen diesen Punkten bestehen. Kernaussage ist somit, dass zwei beliebige nicht benachbarte räumliche Effekte aus der Menge  $\mathcal{V}$  bedingt unabhängig sind, gegeben allen anderen Effekten.

Der zugehörige ungerichtete Graph  $\mathcal{G}_{\text{paar}}$  wird durch die anschließende Abbildung visualisiert. Die gesamte Menge  $\mathcal{V}$  wird demzufolge durch die Knotenpunkte und die Tupel der Menge  $\mathcal{E}$  durch deren Verbindungsgeraden repräsentiert.

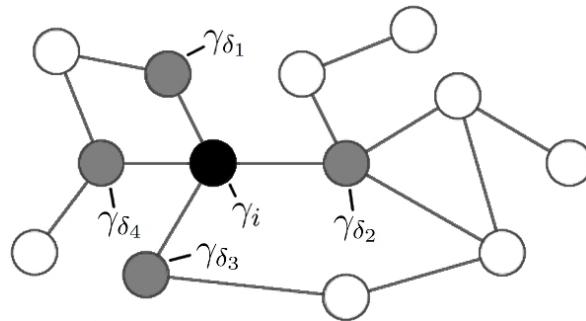


**Abbildung 10:** Paarweise räumliche Markov-Eigenschaft: die beiden schwarzen Knotenpunkte sind bedingt unabhängig, gegeben den grauen Knotenpunkten (Grafik abgeändert übernommen aus Rue und Held, 2005, S. 25)

(b) Lokale Markov-Eigenschaft:

$$\gamma_i \perp \boldsymbol{\gamma}_{-\{i, \delta_r\}} \mid \boldsymbol{\gamma}_{\delta_r}, \quad \text{für jedes } i \in \mathcal{V}$$

Die lokale Markov-Eigenschaft kann als Informationsgewinn der GMZF zur paarweisen Eigenschaft betrachtet werden. Inhalt dieses Konstruktions ist, dass ein beliebiges  $\gamma_i$  und alle weiteren Knotenpunkte aus  $\mathcal{V}$ , abzüglich der von  $\gamma_i$  benachbarten Effekte (dargestellt durch  $\boldsymbol{\gamma}_{-\{i, \delta_r\}}$ ), bedingt unabhängig sind, gegeben der Menge eben dieser benachbarten Knotenpunkte  $\boldsymbol{\gamma}_{\delta_r}$ . Analog zu Abbildung 10 veranschaulicht Abbildung 11 den zugehörigen Graphen  $\mathcal{G}_{\text{lokal}}$ . In diesem Zusammenhang bilden die Punkte  $\gamma_{\delta_1}, \gamma_{\delta_2}, \gamma_{\delta_3}, \gamma_{\delta_4}$  gemeinsam die Menge der von  $\gamma_i$  benachbarten Knotenpunkte  $\boldsymbol{\gamma}_{\delta_r}$ .

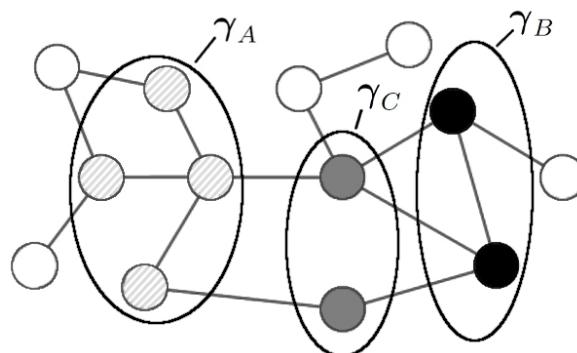


**Abbildung 11:** Lokale räumliche Markov-Eigenschaft: die schwarzen und die weißen Knotenpunkte sind bedingt unabhängig, gegeben den grauen Knotenpunkten (Grafik abgeändert übernommen aus Rue und Held, 2005, S. 25)

(c) Globale Markov-Eigenschaft:

$$\gamma_A \perp \gamma_B \mid \gamma_C$$

Die globale Markov-Eigenschaft gilt für alle disjunkten Zusammenstellungen aus den Mengen A, B und C als Teilmengen von  $\mathcal{V}$  ( $\{A, B, C\} \subseteq \mathcal{V}$ ), bei denen C die Mengen A und B voneinander trennt und zusätzlich A und B keine leeren Mengen sind. Sobald C eine leere Menge darstellt, sind  $\gamma_A$  und  $\gamma_B$  unabhängig. Andernfalls besteht eine bedingte Abhängigkeitsstruktur zwischen  $\gamma_A$  und  $\gamma_B$  bei gegebenem  $\gamma_C$ . Anschließende Abbildung illustriert den Graphen  $\mathcal{G}_{global}$  mit einer möglichen Konstellation der genannten Mengen.



**Abbildung 12:** Globale räumliche Markov-Eigenschaft: die gestreiften und die schwarzen Knotenpunkte sind bedingt unabhängig, gegeben den grauen Knotenpunkten (Grafik abgeändert übernommen aus Rue und Held, 2005, S. 25)

Aus der globalen ergibt sich umgehend die lokale und die paarweise Markov-Eigenschaft. Für die gegenläufige Betrachtung ist zu beachten, dass die Vereinigung von A,B und C nicht zwingend die gesamte Menge von  $\mathcal{V}$  beinhalten muss (vgl. in Abbildung 12 die weißen Knotenpunkte). Aus diesem Grund können Eigenschaften der Randverteilungen gleichermaßen von  $\mathcal{G}$  abgeleitet werden (Rue und Held, 2005, S. 25).

### 2.4.3 Verteilung der Gauß-Markov-Zufallsfelder

Um die gemeinsame Verteilung eines gaußverteilten Markov-Zufallsfelds  $p(\boldsymbol{\gamma})$  mit  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)^T$  bestimmen zu können, ist es hilfreich bedingte autoregressive (kurz CAR (engl.)) Modelle zu betrachten. Diesen Ansatz führte Besag in zwei grundlegenden Artikeln zu diesem Thema in den Jahren 1974 und 1975 das erste Mal ein. Solche Modelle schaffen demnach die Möglichkeit, die Verteilung eines GMZF über vollständig bedingten Dichten zu spezifizieren. Konkret angewendet auf die Modellierung bedingter räumlicher Unabhängigkeitsstrukturen wird die Verteilung des räumlichen Effekts der Region  $s$  bedingt durch alle weiteren räumlichen Effekte ohne der Region  $s$ , wobei  $s \in D$  gilt (vgl. Rue und Held, 2005, S. 28).

Wie bei der räumlichen Verallgemeinerung der Markov-Eigenschaft bereits festgestellt, hat die vollständig bedingte Dichte einer Region  $s$  schließlich folgende Form (vgl. Fahrmeir und Kneib, 2010, S. 311):

$$p(\gamma_s | \boldsymbol{\gamma}_{-s}) = p(\gamma_s | \boldsymbol{\gamma}_{\delta_r}),$$

wobei sich die Darstellungsweise in dem Sinn reduziert, dass für die bedingte Verteilung einer Region lediglich die Menge der per Definition benachbarten Regionen eine Rolle spielen, repräsentiert durch  $\delta_r$ . Allgemein ist es jedoch nicht in jedem Fall möglich aus vollständig bedingten Dichten eine gültige gemeinsame Verteilung zu bilden. Um deren Existenz zu gewährleisten, müssen bestimmte Voraussetzungen vorliegen, die in diesem Unterkapitel erläutert werden sollen. Im Folgenden wird zunächst die Bildung der vollständig bedingten Dichten der  $d$  Regionen über die Anwendung von CAR Modellen hergeleitet, woraufhin in Verbindung mit Brook's Lemma aufgezeigt wird, dass unter bestimmten Voraussetzungen die gemeinsame Verteilung eines GMZF bis auf Proportionalität bestimmt werden kann. Der Zusammenhang zwischen den ungerichteten Graphen aus Unterkapitel 2.4.1 und der bedingten Unabhängigkeitsstruktur innerhalb eines GMZF dient als Überleitung zur Präzisionsmatrix und den Bedingungen, welche erfüllt sein müssen, um die Gültigkeit der gemeinsamen Verteilung zu gewährleisten.

#### Spezifizierung über ein bedingt autoregressives Modell

Nimmt man für den autoregressiven Prozess 1. Ordnung aus Gleichung (3) des vorhergehenden Unterkapitels 2.4.2 einen normalverteilten Störterm mit  $\varepsilon_{t+1} \sim \mathcal{N}(0, \sigma^2)$  an, so kann die Verteilung der bedingten Form folgendermaßen dargestellt werden (vgl. Schmid, 2013, S. 33):

$$y_{t+1} | y_1, \dots, y_t \sim \mathcal{N}(\phi y_t, \sigma^2) \quad \text{mit } t = 1, \dots, n$$

Die Verteilung des Zustands zum Zeitpunkt  $t + 1$  ist jedoch ausschließlich durch den einen vorherigen Zustand bedingt, weshalb die bedingte Verteilung von  $y_{t+1}$  auch verkürzt dargestellt werden kann mit:

$$y_{t+1} | y_t \sim \mathcal{N}(\phi y_t, \sigma^2) \tag{6}$$

Wie bei der Verallgemeinerung der Markov-Eigenschaft bereits festgestellt wurde, fällt für GMZF die natürliche Anordnung der einzelnen Elemente weg, die bei einem autoregressiven Prozess noch vorliegen (vgl. Abbildung 9). Eine übliche Annäherung ist dann die Spezifizierung der gemeinsamen Verteilung eines GMZF über die vollständig bedingten Dichten jeder einzelnen Region  $s$  (vgl. Rue und Held, 2005, S. 3). Für eine vereinfachte Darstellung der bedingten Verteilung einer Region  $s$  werden, analog zu der verkürzten bedingten Verteilung des autoregressiven Prozesses in (6), lediglich die Menge der benachbarten Regionen  $\delta_r$  berücksichtigt. Gilt somit für zwei Regionen  $s_i$  und  $s_j$ , dass sie nicht benachbart sind ( $s_i \not\sim s_j$ ), wird  $\beta_{s_i s_j}$  als 0 definiert (vgl. Fahrmeir und Kneib, 2010, S. 319). Demzufolge kann folgende Ersetzung stattfinden:

$$\sum_{-s \in D} \beta_{s,-s} \quad \Rightarrow \quad \sum_{r \in \delta_r} \beta_{sr}$$

Ausgehend von einem bedingt autoregressiven Modell mit normalverteilten vollständig bedingten Dichten und einem Mittelwert  $\mu_s \neq 0$  gilt schließlich (vgl. Fahrmeir und Kneib, 2010, S. 318):

$$\gamma_s | \boldsymbol{\gamma}_{\delta_r} \sim \mathcal{N} \left( \mu_s + \sum_{r \in \delta_r} \beta_{sr} (\gamma_r - \mu_r), \sigma_s^2 \right) \quad (7)$$

Als Dichte formuliert ergibt sich dann:

$$p(\gamma_s | \boldsymbol{\gamma}_{\delta_r}) = \frac{1}{\sqrt{2\pi\sigma_s^2}} \cdot \exp \left( -\frac{1}{2\sigma_s^2} \left[ \gamma_s - \mu_s - \sum_{r=s+1}^d \beta_{rs} (\gamma_r - \mu_r) \right]^2 \right)$$

mit bedingtem Erwartungswert und bedingter Varianz (vgl. Schmid, 2013, S. 35):

$$\mathbb{E}(\gamma_s | \boldsymbol{\gamma}_{\delta_r}) = \mu_s + \sum_{r=s+1}^d \beta_{rs} (\gamma_r - \mu_r) \quad (8)$$

$$\mathbb{V}(\gamma_s | \boldsymbol{\gamma}_{\delta_r}) = \sigma_s^2 > 0 \quad (9)$$

Der bedingte Erwartungswert ist zusammengesetzt aus der Differenz des unbedingten Erwartungswerts und einem gewichteten Mittelwert aus den Abweichungen der benachbarten Beobachtungen von ihren jeweiligen Erwartungswerten. Die Gewichtung basiert dabei auf der Nachbarschaftsstruktur der Regionen, welche sich in der Präzisionsmatrix widerspiegelt, die in einer späteren Passage des Abschnitts näher betrachtet werden soll. Da nun die vollständig bedingten Dichten spezifiziert sind, kann die Bestimmung der gemeinsamen Verteilung erfolgen.

### Gemeinsame Verteilung über Brook's Lemma

Die gemeinsame Dichte von  $\boldsymbol{\gamma}$  kann auf Basis von autoregressiven Modellen mit normalverteilten vollständig bedingten Dichten über Brook's Lemma bis auf eine Normierungskonstante berechnet werden. Eine Grundvoraussetzung für die Gültigkeit von Brook's Lemma ist die Positivität. Die gilt immer dann, wenn der Träger der gemeinsamen Verteilung von  $\boldsymbol{\gamma}$  durch das kartesische Produkt der Träger der Randverteilungen gebildet werden kann, d. h. wenn  $p(\gamma_s) > 0$ , dann gilt auch  $p(\boldsymbol{\gamma}) > 0$ . Im Folgenden repräsentiert  $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)^T$  einen Referenzpunkt im Träger der gemeinsamen Dichte  $\boldsymbol{\gamma}$  (vgl. Fahrmeir und Kneib, 2010, S. 313). Ausgangspunkt ist die Identität der Dichten (vgl. Rue und Held, 2005, S. 29):

$$\frac{p(\gamma_d | \gamma_1, \dots, \gamma_{d-1}) \cdot p(\gamma_1, \dots, \gamma_{d-1})}{p(\mu_d | \gamma_1, \dots, \gamma_{d-1}) \cdot p(\gamma_1, \dots, \gamma_{d-1})} = \frac{p(\gamma_1, \dots, \gamma_{d-1}, \gamma_d)}{p(\gamma_1, \dots, \gamma_{d-1}, \mu_d)}$$

Durch eine einfache Umstellung der Gleichung kann die anschließende Form erreicht werden (vgl. ebd.):

$$p(\gamma_1, \dots, \gamma_d) = \frac{p(\gamma_d | \gamma_1, \dots, \gamma_{d-1})}{p(\mu_d | \gamma_1, \dots, \gamma_{d-1})} \cdot p(\gamma_1, \dots, \gamma_{d-1}, \mu_d)$$

Drückt man den letzten Term der rechten Seite der Gleichung auf die gleiche ausführliche Art und Weise aus, so ergibt sich (vgl. ebd.):

$$p(\gamma_1, \dots, \gamma_d) = \frac{p(\gamma_d | \gamma_1, \dots, \gamma_{d-1})}{p(\mu_d | \gamma_1, \dots, \gamma_{d-1})} \cdot \frac{p(\gamma_{d-1} | \gamma_1, \dots, \gamma_{d-2})}{p(\mu_{d-1} | \gamma_1, \dots, \gamma_{d-2})} \cdot p(\gamma_1, \dots, \gamma_{d-2}, \mu_{d-1}, \mu_d)$$

Wiederholt man diesen Prozess  $d - 1$  mal, so erhält man folgende proportionale Darstellung der Faktorisierung (vgl. Fahrmeir und Kneib, 2010, S. 319):

$$p(\boldsymbol{\gamma}) \propto \prod_{s=1}^d \frac{p(\gamma_s | \gamma_1, \dots, \gamma_{s-1}, \mu_{s+1}, \dots, \mu_d)}{p(\mu_s | \gamma_1, \dots, \gamma_{s-1}, \mu_{s+1}, \dots, \mu_d)} \quad (10)$$

Die gemeinsame Verteilung ist definiert in Bezug auf den gewählten Referenzpunkt. Anders als bei einer zeitlichen und eindimensionalen Abfolge, bei der die erste Beobachtung als Referenzpunkt gewählt wird, können die Elemente in der vorliegenden räumlichen Betrachtung beliebig permuiert werden. Das hat zur Folge, dass für das Erlangen der gleichen gemeinsamen Verteilung einige Einschränkungen in Bezug auf die vollständig bedingten Dichten berücksichtigt werden müssen. Diese beschränken sich im Wesentlichen auf die Zusammensetzung der Nachbarschaftsmatrix, worauf an späterer Stelle in diesem Unterkapitel explizit eingegangen wird. Eine andere Möglichkeit der Bestimmung der gemeinsamen Verteilung kann über das Hammersley-Clifford-Theorem erfolgen. Hierbei spielen sogenannte Cliques und die Negpotential-Funktion eine wichtige Rolle (vgl. Fahrmeir und Kneib, 2010, S. 315). Im Rahmen der vorliegenden Arbeit soll jedoch nicht weiter auf diesen Ansatz eingegangen werden.

Mit der Normalverteilungsannahme der vollständig bedingten Dichten aus (7) erhält man folgende gemeinsame Verteilung (vgl. Fahrmeir und Kneib, 2010, S. 319):

$$\begin{aligned}
 \log(p(\boldsymbol{\gamma})) &\propto \sum_{s=1}^d \left( \frac{p(\gamma_s | \mu_1, \dots, \mu_{s-1}, \gamma_{s+1}, \dots, \gamma_d)}{p(\mu_s | \mu_1, \dots, \mu_{s-1}, \gamma_{s+1}, \dots, \gamma_d)} \right) \\
 &\propto \sum_{s=1}^d \frac{1}{2\sigma_s^2} \left( \left[ \sum_{r=s+1}^d \beta_{rs}(\gamma_r - \mu_r) \right]^2 - \left[ \gamma_s - \mu_s - \sum_{r=s+1}^d \beta_{rs}(\gamma_r - \mu_r) \right]^2 \right) \\
 &= \sum_{s=1}^d \frac{1}{2\sigma_s^2} \left( -(\gamma_s - \mu_s)^2 + 2(\gamma_s - \mu_s) \sum_{r=s+1}^d \beta_{rs}(\gamma_r - \mu_r) \right) \\
 &= -\sum_{s=1}^d \frac{1}{2\sigma_s^2} (\gamma_s - \mu_s)^2 + \sum_{s=1}^d \sum_{r=s+1}^d \frac{\beta_{rs}}{\sigma_s^2} (\gamma_s - \mu_s)(\gamma_r - \mu_r)
 \end{aligned}$$

Nach einer Exponentierung des gesamten Terms kann die anschließende Matrixschreibweise hergeleitet werden (vgl. ebd.):

$$p(\boldsymbol{\gamma}) \propto \exp \left( -\frac{1}{2} (\boldsymbol{\gamma} - \boldsymbol{\mu})^T \underbrace{\boldsymbol{\Sigma}^{-1} (\mathbf{I}_d - \mathbf{B})}_{\mathbf{P}} (\boldsymbol{\gamma} - \boldsymbol{\mu}) \right)$$

Bei der Verteilung von  $\boldsymbol{\gamma}$  handelt es sich somit um eine multivariate Normalverteilung mit Erwartungswert  $\boldsymbol{\mu}$  und Kovarianz  $\boldsymbol{\Sigma}(\mathbf{I}_d - \mathbf{B})^{-1}$  unter der Voraussetzung, dass  $(\mathbf{I}_d - \mathbf{B})$  positiv definit und  $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_d^2)$  ist.

In den folgenden Darstellungen wird die Nachbarschaftsmatrix durch  $\mathbf{Q}$  repräsentiert und die Varianz der Normalverteilung durch deren Präzision ersetzt, d. h. es gilt  $\sigma^2 = \kappa^{-1}$  (vgl. Schmid, 2004, S. 11). Die Kombination aus Präzision und Nachbarschaftsmatrix wird außerdem als Präzisionsmatrix bezeichnet ( $\mathbf{P} = \kappa \mathbf{Q}$ ). Darüber hinaus ergibt sich durch die Kenntnis der Präzisionsmatrix  $\mathbf{P}$  die Normierungskonstante der gemeinsamen Verteilung über  $|\mathbf{P}|^{1/2}$ . In der folgenden Darstellungsform wird zusätzlich  $\boldsymbol{\mu} = 0$  gesetzt:

$$p(\boldsymbol{\gamma} | \kappa) \propto |\mathbf{P}|^{1/2} \cdot \exp \left( -\frac{1}{2} \boldsymbol{\gamma}^T \mathbf{P} \boldsymbol{\gamma} \right) \quad (11)$$

In diesem Kontext ist die richtige Wahl der Normierungskonstante der gemeinsamen Verteilung ausschlaggebend für eine unverzerrte Schätzung. Diese kann mit  $|\mathbf{P}| = |\kappa \mathbf{Q}|$  dargestellt werden, wodurch das Produkt der Eigenwerte, welche ungleich Null sind bezeichnet wird, da die Determinante nicht explizit berechnet werden kann. Es können somit  $rg(\kappa \mathbf{Q}) = rg(\mathbf{Q}) = d - 1$  Eigenwerte gebildet werden, weshalb  $\kappa$  dann  $rg(\mathbf{Q})$ -mal vorgezogen werden kann und sich  $\kappa^{rg(\mathbf{Q})}$  als erster Term ergibt (vgl. Schmid, 2004, S. 11 f.).

Für eine detailliertere Bestimmung der  $\beta_{sr}$ 's können diese als Anteile am Gesamtgewicht formuliert werden, sodass sich anschließende Formen ergeben (vgl. Fahrmeir und Kneib, 2010, S. 319):

$$\beta_{sr} = p \frac{w_{sr}}{w_{s+}} \quad \text{mit} \quad w_{s+} = \sum_{r \in \delta_r} w_{sr} \quad \text{und} \quad p \in [0, 1]$$

Zusätzlich wird die Varianz kleiner bzw. die Präzision größer, je mehr Nachbarn eine Region  $s$  hat bzw. je höher die Summe der Gewichtung der Nachbarschaften ist:

$$\sigma_s^2 = \frac{\sigma^2}{w_{s+}} \quad \text{bzw.} \quad \kappa_s = \frac{1}{\kappa w_{s+}}$$

Für die vollständig bedingten Dichten gilt somit (vgl. Fahrmeir und Kneib, 2010, S. 320):

$$\gamma_s | \boldsymbol{\gamma}_{\delta_r} \sim \mathcal{N} \left( p \sum_{r \in \delta_r} \frac{w_{sr}}{w_{s+}} \gamma_r, \frac{1}{\kappa w_{s+}} \right) \quad (12)$$

mit den symmetrisch angeordneten Gewichten  $w_{sr} = w_{rs}$  und  $w_{s+} = \sum_{r \in \delta_r} w_{sr}$ . Der bedingte Erwartungswert für den räumlichen Effekt  $\gamma_s$  einer Region  $s$  ist somit gegeben durch das gewichtete Mittel der Koeffizienten der jeweiligen Nachbarn (vgl. Fahrmeir et al., 2009, S. 391).

### Verbindung zwischen ungerichteten Graphen und GMZF

Die Präzisionsmatrix spielt eine zentrale Rolle für den Umgang mit GMZF. Bevor näheres zu dieser Matrix und deren Anforderungen erläutert wird, stellt die Klärung des Zusammenhangs eines ungerichteten Graphen (vgl. Abschnitt 2.4.1) und einem GMZF eine wichtige Ergänzung dar.

In Bezug auf ein GMZF  $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)^T$  wird im Folgenden der Erwartungswert  $\boldsymbol{\mu}$  und die Kovarianzmatrix  $\boldsymbol{\Sigma}$  angenommen. Wie bereits in Verbindung mit der räumlichen Betrachtungsweise der paarweisen Markov-Eigenschaft aus Unterkapitel 2.4.2 erläutert, eignen sich ungerichtete Graphen hervorragend, um die bedingte Unabhängigkeitsstruktur eines GMZF zu erfassen, da ein Graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  mit  $\mathcal{V} = \{1, \dots, d\}$  und  $\mathcal{E}$  so konstruiert ist, dass zwischen den Knotenpunkten  $i$  und  $j$  keine Verbindung besteht, wenn  $\gamma_i \perp \gamma_j | \boldsymbol{\gamma}_{-ij}$  Gültigkeit hat. Der Vektor  $\boldsymbol{\gamma}$  kann dann als GMZF bezeichnet werden in Bezug auf den ungerichteten Graphen  $\mathcal{G}$ . Laut einem Theorem aus Rue und Held (2005) gilt unter der Voraussetzung  $P > 0$  für  $i \neq j$  folgender Zusammenhang (vgl. Rue und Held, 2005, S. 21):

$$\gamma_i \perp \gamma_j | \boldsymbol{\gamma}_{-ij} \Leftrightarrow P_{ij} = 0. \quad (13)$$

Die Einträge der Präzisionsmatrix  $P_{ij}$  sind demnach Null, sobald zwei Regionen nicht benachbart sind, wodurch die bedingte Unabhängigkeitsstruktur der Regionen abgebildet wird. Aus der Präzisionsmatrix kann somit abgeleitet werden, ob  $\gamma_i$  und  $\gamma_j$  bedingt unabhängig sind bzw. ob sie benachbart sind oder nicht. Zusammengefasst sagt dieses Theorem außerdem aus, dass die Einträge von  $\mathbf{P}$ , die ungleich Null sind, den Graphen  $\mathcal{G}$  bestimmen, da diese die bedingte Unabhängigkeitsstruktur abbilden, wobei entgegen gesetzt aus einem gegebenen Graphen die Einträge für  $P_{ij}$  abgeleitet werden können. Ferner ist festzustellen, dass der Erwartungswert  $\boldsymbol{\mu}$  keinen Einfluss auf die paarweise bedingte Unabhängigkeitseigenschaft von  $\boldsymbol{\gamma}$  hat. Dadurch kann man darauf schließen, dass diese Information ausschließlich in der Präzisionsmatrix integriert sein muss (vgl. Rue und Held, 2005, S. 21).

## Präzisionsmatrix

Damit nach Brook's Lemma eine gemeinsame Verteilung aus den vollständig bedingten Dichten gebildet werden kann, müssen bestimmte Anforderungen eingehalten werden, welche auf die Eigenschaft der Symmetrie und der positiven (semi) Definitheit der Präzisionsmatrix reduziert werden können (vgl. Rue und Held, 2005, S. 4).

Die Betrachtung der Nachbarschaftsmatrix  $\mathbf{Q}$  ist in diesem Kontext aufgrund ihrer Einträge von besonderer Bedeutung. Nach der Verteilung der vollständig bedingten Dichten aus (12) setzen sich diese für  $\mathbf{Q}$  im Allgemeinen folgendermaßen zusammen (vgl. Fahrmeir *et al.*, 2009, S. 392):

$$q_{sr} = \begin{cases} -pw_{sr}, & s \neq r, s \sim r \\ 0, & s \neq r, s \not\sim r \\ pw_{s+}, & s = r \end{cases}$$

Immer dann, wenn zwei Regionen benachbart sind, beinhaltet die Matrix an der zugehörigen Stelle von  $\mathbf{Q}$  die negative Gewichtung der jeweiligen Nachbarschaft multipliziert mit dem globalen Gewichtungsfaktor  $p$ . Sind die beiden Regionen nicht benachbart, dann wird diese Begebenheit mit einer Null repräsentiert. Die Diagonaleinträge der Matrix beinhalten zudem das Produkt aus  $p$  und der Summe der Gewichte der von  $s$  benachbarten Regionen.

Für die Symmetrie von  $\mathbf{Q}$  kann festgehalten werden, dass die Nachbarschaftsstruktur per Definition bereits eine symmetrische Eigenschaft aufweist und dadurch unmittelbar diese Voraussetzung für Brook's Lemma eingehalten wird, da somit  $q_{sr} = q_{rs}$  gilt. Eine übliche Methode die positive (semi) Definitheit zu belegen besteht darin aufzuzeigen, dass  $\mathbf{Q}$  diagonaldominant ist unter der Voraussetzung  $\kappa > 0$ . Das bedeutet, dass in jeder Zeile bzw. Spalte der Nachbarschaftsmatrix die Diagonaleinträge größer (gleich) sind als die Summe der Absolutwerte aller restlichen Zeilen- bzw. Spalteneinträge (vgl. Rue und Held, 2005, S. 4). Eine symmetrische Matrix wird schließlich dann als (schwach) diagonaldominant bezeichnet, wenn Folgendes gilt (vgl. Voigt und Adamy, 2007, S. 81):

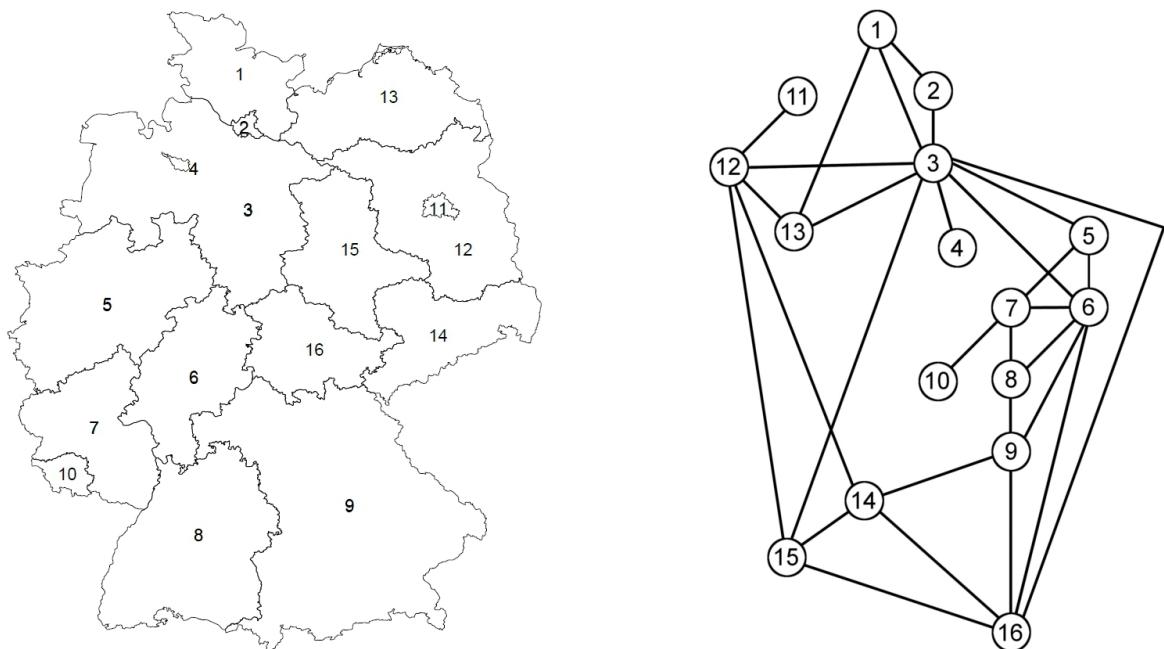
$$|q_{ss}| \geq \sum_{r \in \delta_r}^d |q_{sr}|$$

Handelt es sich bei  $\mathbf{Q}$  um eine symmetrische, positiv (semi-)definite Matrix, dann gilt das mit  $\kappa > 0$  auch für die Präzisionsmatrix  $\kappa\mathbf{Q} = \mathbf{P}$ .

Um an dieser Stelle eine anschauliche Illustrierung der Nachbarschaftsmatrix zu ermöglichen, wird nun von intrinsischen GMZF (vgl. anschließendes Unterkapitel 2.4.4) und zusätzlich einer einheitlichen Gewichtung aller Nachbarschaften ausgegangen. Allgemein ergeben sich hierfür folgende Einträge (vgl. Fahrmeir *et al.*, 2009, S. 389):

$$q_{sr} = \begin{cases} -1, & s \neq r, s \sim r \\ 0, & s \neq r, s \not\sim r \\ N_s, & s = r \end{cases}$$

Mit  $N_s$  als Anzahl der Nachbarn der Region  $s$ . Anschließende Grafik zeigt einen Vergleich der Aufgliederung der Bundesrepublik Deutschland in die 16 Bundesländer und dem Deutschlandgraphen aus Abbildung (8):



**Abbildung 13:** links: Karte der deutschen Bundesländer (Grafik übernommen aus Schmid, 2004, S. 12), rechts: Deutschlandgraph (Grafik nachempfunden: Rue und Held, 2005, S. 5)

Die Nummerierung der Bundesländer spiegelt sich in den Zeilen- bzw. Spaltenindizes der Nachbarschaftsmatrix wider (vgl. Schmid, 2004, S. 12):

$$Q = \begin{pmatrix} 3 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 9 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & -1 & -1 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 3 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & -1 & 6 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 5 & -1 & -1 & -1 \\ -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & 0 & 4 & -1 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 \\ 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & -1 & -1 & 5 \end{pmatrix}$$

Diese Herangehensweise und weitere Möglichkeiten der Anpassung werden in nachstehendem Unterkapitel näher beleuchtet.

#### 2.4.4 Adaptive Modellierung

Nachdem nun die Grundvoraussetzungen für eine Modellierung von räumlich strukturierten Daten in Verbindung mit GMZF erläutert wurde, soll als abschließender Aspekt der Aufbereitung des theoretischen Hintergrunds auf verschiedene Möglichkeiten der Zusammensetzung von GMZF eingegangen werden.

Eine große Schwäche der direkten räumlichen Glättung über GMZF ohne jeglichen Anpassungen ist, dass über ein unerwartet hohes Auftreten der Zielgröße hinweg geglättet wird, wodurch diese meist unterschätzt werden (vgl. Brezger *et al.*, 2007, S. 1). Allgemeines Ziel ist es deshalb eine möglichst adaptive bzw. anpassungsfähige Modellierung zu entwickeln, um eine flexiblere Glättung der Daten zu erhalten. Um das gewährleisten zu können, ist es hilfreich die Gewichtungen  $w_{sr}$  der Nachbarschaften im Zuge des hierarchischen Modellaufbaus über eine eigene Prioriannahmen zu berücksichtigen. Brezger *et al.* (2007) beschreiben diese Möglichkeit folgendermaßen:

„To improve spatial adaptivity, we introduce a class of inhomogeneous Markov random fields with stochastic interaction weights in a space-varying coefficient model“ (Brezger *et al.*, 2007, S. 1).

Bevor jedoch explizit auf diese Vorgehensweise eingegangen wird, sollen zunächst die sogenannten intrinsischen Gauß-Markov-Zufallsfelder (IGMZF) betrachtet werden, welche in dieser Arbeit die Grundlage für die Aufbereitung der adaptiven Modellierungen darstellen.

### Intrinsische Gauß-Markov-Zufallsfelder (IGMZF)

Yan (2006) hält in einem Artikel fest, dass gerade in neuerer Zeit IGMZF eine breite Anwendung im Rahmen von räumlichen hierarchischen Modellierungen finden:

„In the last decade, intrinsic GMRFs (IGMRFs), a special type of GMRF with a degenerate precision matrix, have been widely used as improper priors for random effects in hierarchical spatial modeling“ (Yan, 2006, S. 388).

Diese zeichnen sich dadurch aus, dass der globale Mittelwertsparameter  $p$  auf 1 festgelegt wird. Die bereits im Zuge der Ausführungen zur Präzisionsmatrix des vorhergehenden Unterkapitels erwähnten IGMZF, mit zusätzlich gleicher Gewichtung aller Nachbarschaften, stellen dabei die knappste Form der GMZF dar. Durch die genannten Annahmen vereinfachen sich die vollständig bedingten Dichten aus (12) von Seite 39 dann zu (vgl. Fahrmeir und Kneib, 2010, S. 302):

$$\gamma_s | \boldsymbol{\gamma}_{\delta_r} \sim \mathcal{N} \left( \frac{1}{N_s} \sum_{r \in \delta_r} \gamma_r, \frac{1}{\kappa N_s} \right) \quad (14)$$

Im Einzelnen wird für die Gewichte  $w_{sr} = 1$  angenommen, woraus folgt, dass sich für  $w_{s+} = N_s$  ergibt. Bei einer solchen Konstruktion wird allerdings lediglich die räumliche Nachbarschaft unter den Regionen per se verwendet. Für eine adaptive Modellierung ist es jedoch unerlässlich die Definition für die Gewichtungen der Nachbarschaften zu verallgemeinern. Die vollständig bedingten Dichten für IGMZF stellen sich somit folgendermaßen dar (vgl. Fahrmeir *et al.*, 2009, S. 390 f.):

$$\gamma_s | \boldsymbol{\gamma}_{\delta_r} \sim \mathcal{N} \left( \sum_{r \in \delta_r} \frac{w_{sr}}{w_{s+}} \gamma_r, \frac{1}{\kappa w_{s+}} \right)$$

Die gemeinsame Verteilung eines GMZF ist immer dann eigentlich, solange die Bedingung  $0 \leq p < 1$  erfüllt ist. Für IGMZF gilt es allerdings allgemein zu beachten, dass durch die Festlegung des globalen Mittelwerts  $p$  auf 1 für die vollständig bedingten Dichten eine uneigentliche gemeinsame Verteilung gebildet wird. Aufgrund der vorliegenden Gauß-Eigenschaft wird in diesem Fall auch von einer singulären multivariaten Normalverteilung gesprochen (vgl. Schmid, 2004, S. 11). Die uneigentliche Verteilung entsteht dadurch, dass  $(\mathbf{I}_d - \mathbf{B})$  keinen vollen Rang mehr besitzt und aus diesem Grund auch nicht mehr invertierbar – aber immer noch positiv semidefinit – ist. Diese uneigentliche Version der GMZF hat den Vorteil, dass bei der Berechnung des globalen Effekts die Abweichung von einem konstanten Niveau nicht bestraft wird. Im Gegensatz dazu wird bei einer eigentlichen Verteilung erzwungen, dass dieser Effekt gegen den priori Erwartungswert  $\mu$  läuft. Ferner stellt es einen Vorteil dar, dass es nicht notwendig ist für den zusätzlichen Hyperparameter  $p$  eine Priorannahme zu treffen (vgl. Fahrmeir und Kneib, 2010, S. 320).

## Anpassung der Gewichte

Bei einer globalen Bestimmung der Glättungsparameter wie in (14) wird in einem GMZF über alle Beobachtungseinheiten gleich geglättet. Eine solche globale Glättung ist oftmals jedoch nicht zielführend, wodurch die Verwendung anpassungsfähigere Modelle motiviert werden kann (vgl. Schmid, 2004, S. 17). Ein erster Schritt ist eine von Grund auf andere Definition der einzelnen Gewichte. Diese können beispielsweise als invers proportional zum Abstand der Zentroide von benachbarten Regionen  $s$  und  $r$  bestimmt werden. Das bedeutet  $w_{sr} \propto \exp(-d(s, r))$ , wobei  $d(s, r)$  die euklidische Distanz kennzeichnet. Eine weitere Herangehensweise kann dargestellt werden durch die Festlegung der Gewichte proportional zur Länge der gemeinsamen Grenze zweier benachbarter Regionen  $s$  und  $r$ . Andere Arten der Bestimmung sind ebenso denkbar. Wichtig ist allerdings in jedem Fall, dass es sich um eine symmetrische Definition der Gewichte handeln muss, um zulässige GMZF zu spezifizieren (vgl. Fahrmeir und Kneib, 2010, S. 320).

Um jedoch eine wirklich anpassungsfähige Modifikation von GMZF modellieren zu können, ist es erforderlich, dass die einzelnen Gewichtungen  $w_{sr}$  im Verbund der hierarchischen Modellierung stochastisch variieren können (vgl. Brezger *et al.*, 2007, S. 331). Grundsätzlich kann in diesem Zusammenhang die Anpassung der Gewichte zwischen den benachbarten Beobachtungseinheiten auf zwei verschiedene Arten erfolgen. Bei einer möglichen Variante wird für die Hyperpriori der einzelnen Gewichte eine Gamma-Verteilung angenommen mit  $w_{sr} \sim G(\nu/2, \nu/2)$  und einer kleinen Anzahl an Freiheitsgraden, vornehmlich mit  $\nu = 1$ :

$$p(w_{sr}) \propto w_{sr}^{(\nu/2-1)} \exp\left(-\frac{\nu}{2}w_{sr}\right)$$

Diese Annäherung erlaubt die Aufnahme der Beziehungen in unterschiedlicher Stärke, jedoch ohne einer *a priori* festgelegten räumlichen Struktur (vgl. Brezger *et al.*, 2007, S. 334).

Eine weitere, etwas komplexere, Herangehensweise kann über die Annahme entwickelt werden, dass die logarithmierten Gewichte  $\log(w_{sr}) = v_{sr}$  ebenso eine räumliche Struktur aufweisen. In diesem Kontext werden zusätzlich die Gewichte selbst über ein GMZF in die Modellierung mit aufgenommen mit der Präzision  $\phi$  und der Nachbarschaftsmatrix  $L$  (vgl. ebd.):

$$p(\mathbf{v} | \phi) \propto \exp\left(-\frac{1}{2}\mathbf{v}^T \phi \mathbf{L} \mathbf{v}\right)$$

Der Hauptgrund dafür, warum sich die Inferenz bei stochastisch variierende Gewichten zusehends erschwert, ist die Einbindung der Eigenwerte der Präzisionsmatrix, die nicht Null sind, in die Normierungskonstante der gemeinsamen Verteilung eines GMZF in Verbindung mit der Abhängigkeit von den Gewichten in einer komplizierten und nicht-linearen Art und Weise (vgl. Brezger *et al.*, 2007, S. 332). Diese Abhängigkeit der Präzisionsmatrix  $\mathbf{P}$  von dem Vektor der Gewichte  $\mathbf{w}$  wird im Folgenden repräsentiert durch  $\mathbf{P}_w$ , kann aber auf die Abhängigkeit der Nachbarschaftsmatrix  $\mathbf{Q}$  von den Gewichten reduziert werden, geschrieben als  $\mathbf{Q}_w$ .

### 2.4.5 Zusammenfassung: Gauß-Markov-Zufallsfelder

#### 1. Bedingte Unabhängigkeitsstruktur in GMZF:

- Nachbarschaftsstruktur wird aufgefasst als bedingte Unabhängigkeit
- Geeignete Darstellung über ungerichtete Graphen  $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$
- Verwendung von räumlichen Markov-Eigenschaften

#### 2. Verteilung der vollständig bedingten Dichten über CAR Modelle:

$$\gamma_s | \boldsymbol{\gamma}_{\delta_r} \sim \mathcal{N} \left( p \sum_{r \in \delta_r} \frac{w_{sr}}{w_{s+}} \gamma_r, \frac{\sigma^2}{w_{s+}} \right) \quad \text{mit} \quad p \in [0, 1]$$

- $\gamma_s \hat{=} \text{räumlicher Effekt von Region } s$
- $r \hat{=} \text{benachbarte Region von Region } s$  ( $s$  und  $r$  sind Nachbarn:  $s \sim r$ )
- $\delta_r \hat{=} \text{Indizes aller Nachbarn von Region } s$
- $w_{sr} \hat{=} \text{Gewichtung einer Nachbarschaft mit } w_{s+} = \sum_{r \in \delta_r} w_{sr}$
- $p \hat{=} \text{globaler Mittelwert} \rightarrow \text{intrinsische GMZF (IGMZF) mit } p = 1$
- Vereinfachung über die Präzision der Normalverteilung:  $\sigma^2 = \kappa^{-1}$

#### 3. Gemeinsame Verteilung eines GMZF über Brook's Lemma:

$$p(\boldsymbol{\gamma} | \kappa, \mathbf{w}) \propto |\mathbf{P}_w|^{1/2} \cdot \exp \left( -\frac{1}{2} \boldsymbol{\gamma}^T \mathbf{P}_w \boldsymbol{\gamma} \right)$$

- $\mathbf{P}_w = \kappa \mathbf{Q}_w \hat{=} \text{Präzisionsmatrix}$
- $\mathbf{Q}_w \hat{=} \text{Nachbarschaftsmatrix mit den Einträgen:}$

$$q_{sr} = \begin{cases} -pw_{sr}, & s \neq r, s \sim r \\ 0, & s \neq r, s \not\sim r \\ pw_{s+}, & s = r \end{cases}$$

#### 4. Adaptive Modellierung von IGMZF:

- Abhängigkeit der Präzisions- bzw. Nachbarschaftsmatrix von Vektor  $\mathbf{w}$
- Berücksichtige stochastisches Variieren der einzelnen Gewichte  $w_{sr}$
- Ohne räumlicher Struktur der Gewichte mit  $w_{sr} \sim G(\nu/2, \nu/2)$
- Mit räumlicher Struktur der log-Gewichte  $\log(w_{sr}) = v_{sr}$  mit  $\mathbf{v} \sim \text{GMZF}$

### 3 Anwendung

Nach der erfolgten Aufbereitung des theoretischen Hintergrunds der Modellierung von räumlich strukturierten Daten in Zusammenhang mit (intrinsischen) GMZF und deren adaptiven Möglichkeiten, sollen diese im Zuge des nachstehenden Kapitels umgesetzt werden. Im Fokus stehen diesbezüglich eine Anwendung auf empirische Daten über das Auftreten von Lippenkrebs in Schottland und Kehlkopfkrebs in Deutschland sowie auf rein simulierte Daten auf Basis der schottischen Grenz- und Nachbarschaftsstruktur. In diesem Kontext werden in Unterkapitel 3.1 zunächst die jeweiligen Erhebungen deskriptiv aufbereitet, bevor auf die Umsetzung der hierarchischen Struktur des bayesianischen Modellaufbaus eingegangen wird (vgl. Abschnitt 3.2). Eine Übersicht der Verteilungsannahmen an die einzelnen Modellbestandteile bildet dabei die Grundlage für die Aufbereitung der durchgeführten Reparametrisierung, welche für eine günstigere Ausgangslage in Hinblick auf die Verwendung eines MCMC Verfahrens sorgt. Dieses erfordert außerdem die Aufstellung der vollständig bedingten Dichten aller hierarchischen Stufen des Modells (siehe Teilabschnitt 3.2.3). In Unterkapitel 3.3 werden im Zuge der Implementierung eines sogenannten Hybrid-Samplers einleitend die theoretischen Bestandteile eines MCMC Verfahrens dargestellt. Im Weiteren steht die Erläuterung zusätzlich notwendiger Schritte, wie die Verteilungsapproximation über eine Taylorentwicklung für die vollständig bedingte Dichte des Prädiktors sowie die Optimierung der Laufzeit des in Abschnitt 3.3.3 skizzierten Samplers, im Fokus. Die Aufbereitung und die Interpretation der geschätzten Verteilungscharakteristika der jeweiligen Parameter ist schließlich Inhalt der daran anschließenden Unterkapitel 3.4, 3.5 und 3.6. Diesbezüglich wird unterschieden in eine Betrachtung der Schätzergebnisse der empirischen Daten ohne und mit anpassungsfähigen Nachbarschaftsgewichtungen. Zusätzlich dient eine Anwendung auf simulierte Daten unter der Berücksichtigung verschiedener Settings einer verbesserten Beurteilungsmöglichkeit der generierten Samples.

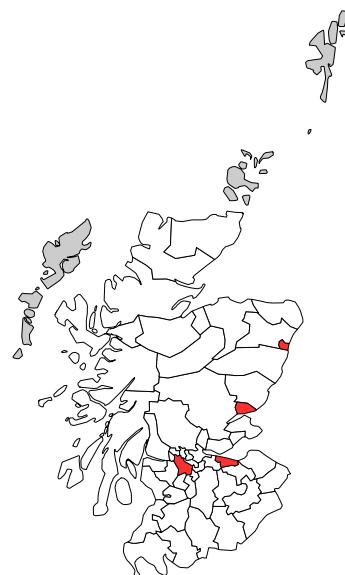
#### 3.1 Datengrundlage

Im Rahmen der vorliegenden Arbeit werden Daten zu Lippenkrebs in Schottland und Kehlkopfkrebs in Deutschland für eine Anwendung herangezogen. Beide Datensätze beinhalten auf Landkreis- bzw. auf Distriktebene Variablen zu beobachteten und den erwarteten Sterbefällen ( $y_s$  und  $e_s$ ), welche durch die jeweilige Krebserkrankung begründet sind. Die erwarteten Fälle können auf die Berechnungsmethoden nach Mantel und Stark (1968) zurückgeführt werden, für welche beispielsweise die Altersstruktur, die Geschlechterzusammensetzung oder die Einwohnerzahlen der einzelnen Landkreise eine Berücksichtigung finden (vgl. Clayton und Kaldor, 1987, S. 677). Für die Einbindung in die Modellierung wird  $e_s$  schließlich als Offset aufgefasst (vgl. Teilabschnitt aus 2.1.1).

Eine aussagekräftige Betrachtungsmöglichkeit stellt in diesem Zusammenhang der SMR (standardized morbidity ratio (engl.)) dar, der sich aus dem Quotienten von beobachteten und erwarteten Todesfälle bildet und als relative Risikorate zu sterben interpretiert werden kann, mit  $\text{SMR}_s = y_s/e_s$  (vgl. Fahrmeir und Kneib, 2010, S. 304). Die detaillierte deskriptive Aufbereitung der beiden Datensätze im Anschluss verschafft einen allgemeinen Überblick über die Ausgangslage der weiteren Anwendungen.

### 3.1.1 Daten zu Lippenkrebs in Schottland

Die schottischen Lippenkrebsdaten wurden über die Jahre von 1975 bis 1980 hinweg für die damaligen 56 schottischen Distrikte erhoben, die in etwa gleichzusetzen sind mit den Landkreisen in Deutschland. Da im Jahr 1996 für Schottland eine neue Verwaltungsgliederung eingeführt wurde, beziehen sich alle Darstellungen von Schottland auf die Aufteilung zum Zeitpunkt der Datenerhebung. Bivand *et al.* (2008) bilden mit dem Code aus ihrem Buch über die Analyse räumlicher Daten in R die Grundlage für die Visualisierung der schottischen Distriktgrenzen (vgl. dort S. 90 f.)<sup>3</sup>.

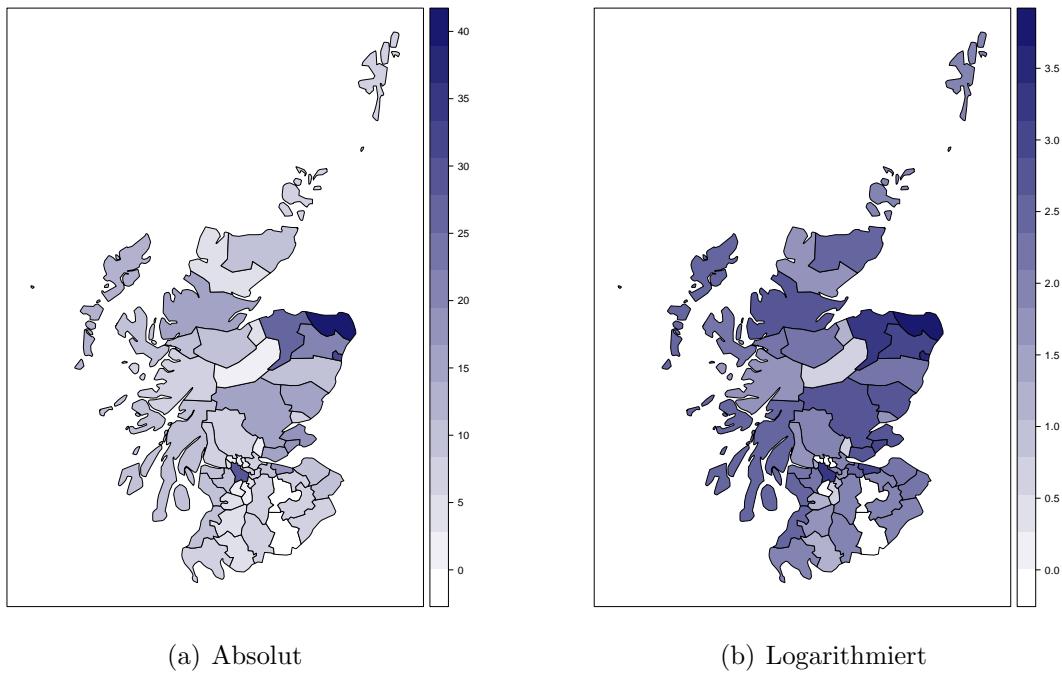


**Abbildung 14:** Umriss der 56 schottischen Distriktgrenzen zum Zeitpunkt der Datenerhebung.  
Hervorgehoben: Inselgruppen ohne direkte Nachbarschaft (grau) und Distrikte der wichtigsten Städte (rot).

Die drei grau hervorgehobenen Inselgruppen stellen Distrikte dar, die nach der Nachbarschaftsdefinition über direkte Grenzen keine Nachbarn haben. Deshalb werden diese außer bei der nachfolgenden deskriptiven Aufbereitung nicht weiter berücksichtigt. Es handelt sich dabei im Westen Schottlands um die äußereren Hybriden, im Nordosten um die Orkney Inseln und weiter nordöstlich davon um die Shetland Inseln. Alle weiteren Inseln gehören Regionen an, die über direkte Nachbarschaften verfügen. Die rot markierten Distrikte beinhalten außerdem die vier größten Städte Schottlands, welche zur Orientierung bei der anschließenden Aufbereitung dienen. Von Westen nach Osten betrachtet sind das die Beobachtungseinheiten mit den Städten Glasgow, Edinburgh, Dundee und Aberdeen.

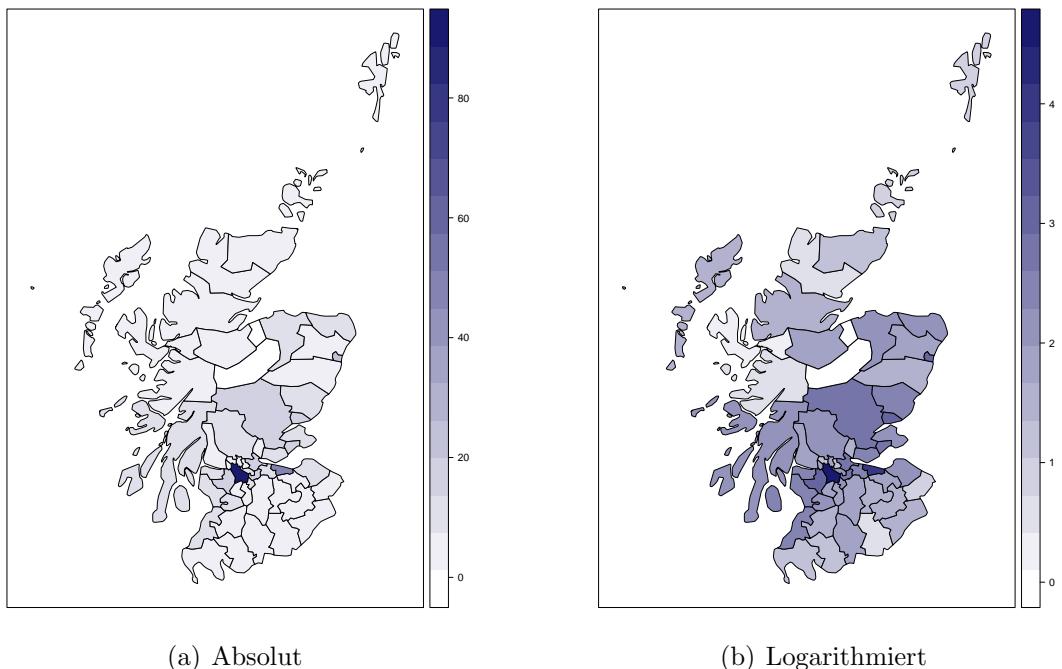
Aufgrund von Ausreißern ist es sinnvoll neben den absoluten zusätzlich auch die logarithmierten Häufigkeiten zu betrachten. Im Folgenden ist dabei zu beachten, dass die Einfärbungen für jede Grafik separat festgelegt wird, in dem die Region mit der jeweils höchsten Ausprägung am dunkelsten ist. Aus diesem Grund können die Farbabstufungen zwischen den Grafiken nicht verglichen werden.

<sup>3</sup>Zudem wird auf der Projekt-Homepage zu ihrem Buch eine ausführliche Beschreibung sowie zusätzliche R-Codes zur Verfügung gestellt: siehe <http://www.bias-project.org.uk/ASDARcourse/>.



**Abbildung 15:** Beobachtete Sterbefälle aus den schottischen Lippenkrebsdaten

Die Regionen nördlich und westlich von Aberdeen weisen die höchsten Anzahlen an beobachteten Sterbefällen auf (maximal 39 in Distrikt Banff-Buchan). Zusätzlich ist im Vergleich zu den angrenzenden Bereichen eine Konzentration auf den Glasgower Distrikt sowie ein geringere Anzahl in einigen südlichen Regionen zu erkennen.



**Abbildung 16:** Erwartete Sterbefälle aus den schottischen Lippenkrebsdaten

Aufgrund der verhältnismäßig hohen Einwohnerzahl von Glasgow werden dort mit 89 Sterbefällen deutlich mehr erwartet als beobachtet (28 Fälle). Ähnlich verhält es sich in Edinburgh mit 51 erwarteten zu 19 beobachteten tödlichen Erkrankungen. Die Abstufungen der anderen Distrikte sind mit einer logarithmierten Skala besser zu erkennen, befinden sich bei einer absoluten Betrachtung jedoch aller unterhalb von 23 erwarteten Fällen. Hervorzuheben ist außerdem die Region mit den meisten beobachteten Sterbefällen Banff-Buchan, für die aufgrund der Zusammensetzung der Bevölkerung lediglich 4 Fälle erwartet werden.

Anschließend soll neben der Verteilung des einleitend beschriebenen SMR zusätzlich die einzige Kovariable betrachtet werden. Diese beinhaltet auf Distriktebene den prozentualen Anteil der Beschäftigten in einem der Bereiche aus Agrarwirtschaft, Fischerei oder Forstwirtschaft (kurz: AFF).

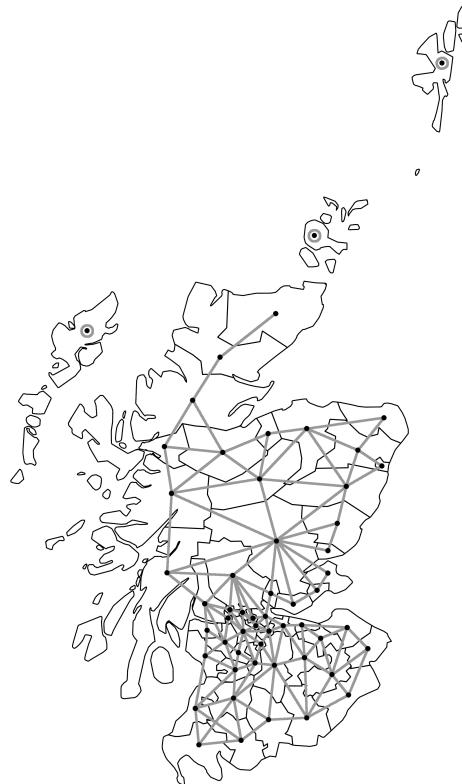


**Abbildung 17:** Zusätzliche Merkmale aus den schottischen Lippenkrebsdaten

Das relative Risiko an Lippenkrebs zu sterben ist in dem dunkel eingefärbten Distrikt Skye-Lochalsh im Nordwesten Schottlands mit einem SMR von 6,4 am höchsten. Gefolgt von der Region Banff-Buchan im Nordosten mit einem Ratio von 4,5. Für alle anderen Beobachtungseinheiten liegt der Quotient unter 3,5. Auf den Orkney Inseln und in drei südlichen Distrikten wird ein Anteil der Beschäftigten im AFF Bereich von 24% registriert. In dem Areal um Aberdeen ist ebenfalls ein hoher Anteil zu beobachten (jeweils 16%), wohingegen im Ballungsraum um Glasgow nahezu gar keine Einwohner in diesem Sektor eine Anstellung haben.

Wie in den theoretischen Grundlagen bereits herausgearbeitet, spielen die Nachbarschaften zwischen den einzelnen Beobachtungseinheiten eine entscheidende Rolle bei der Modellierung von räumlichen Strukturen mittels GMZF. Die in Teilabschnitt 2.4.1 kennen gelernte Darstellung über ungerichtete Graphen wird in anschließender Grafik

auf die schottische Verwaltungsgliederung angewendet. Diese Struktur findet für die Modellierung schließlich als Nachbarschafts- bzw. Präzisionsmatrix ihre Berücksichtigung.



**Abbildung 18:** Struktur der 117 direkten Nachbarschaften der 56 schottischen Distrikte

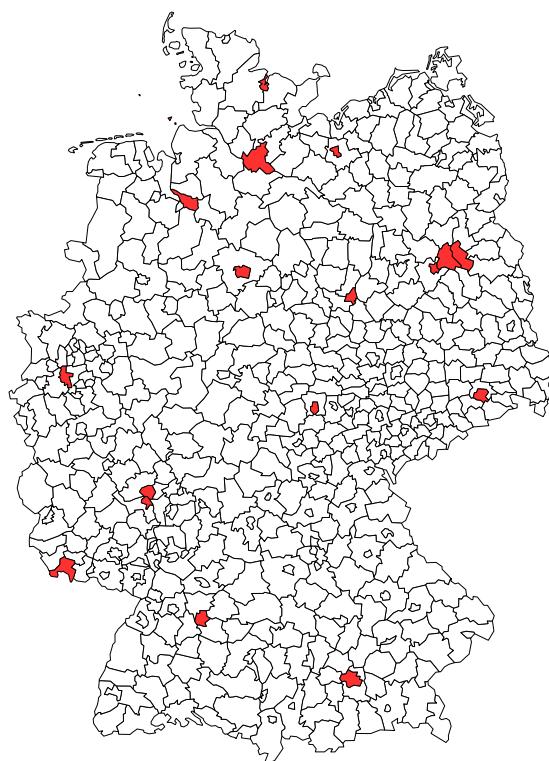
Analog zu der Beschreibung von Abbildung 14 geht von den drei Inselgruppen im Nordwesten und Nordosten Schottlands keine Nachbarschaftsbeziehung zu anderen Distrikten aus.

### 3.1.2 Daten zu Kehlkopfkrebs in Deutschland

Ähnlich zu der schottischen Verwaltungsgliederung in Distrikte, hat sich die deutsche Aufteilung in Landkreise seit der Erhebung der Kehlkopfkrebsdaten verändert. Während dem Beobachtungszeitraum von 1986 bis 1990 war Deutschland (West- und Ostdeutschland zusammen) noch in 544 Landkreise unterteilt – heute sind es 402. Die gezeigten Grafiken berücksichtigen aus diesem Grund die Aufgliederung zu dem damaligen Zeitpunkt mit den heutigen Auslandsgrenzen von Deutschland. Die jeweilige Darstellung der Grenzen basiert dabei auf einem R-Code bereitgestellt auf der Homepage des INLA-Projektes (Rue, 2009).

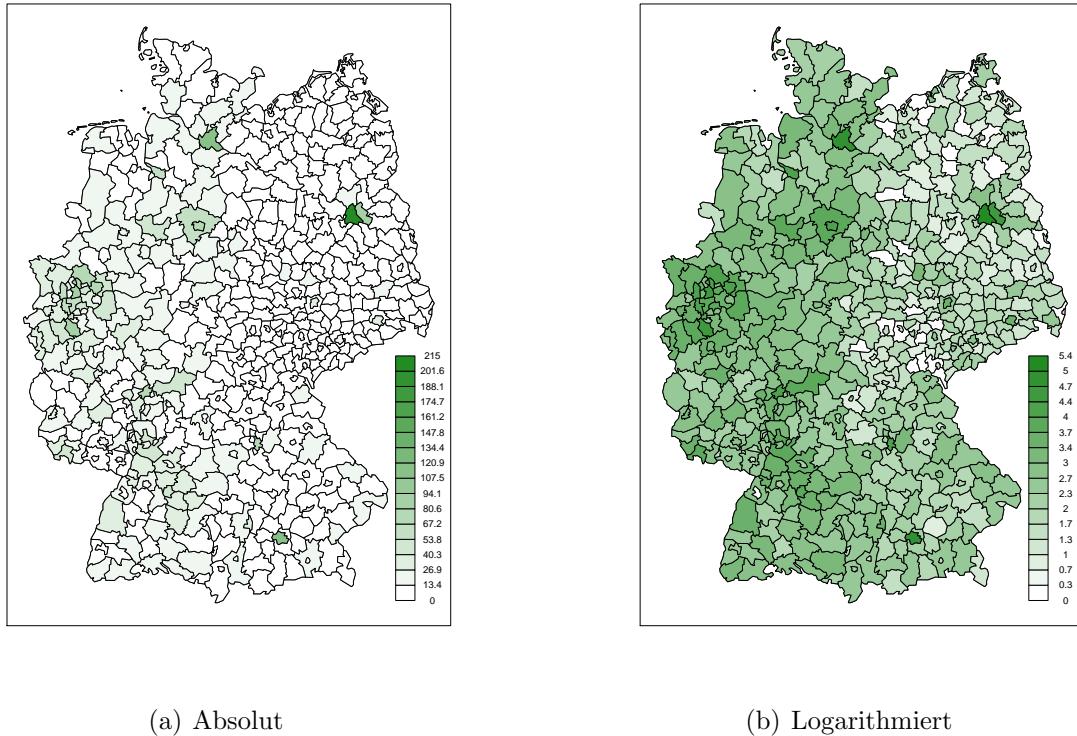
In der anschließenden Abbildung sind die Umrisse aller deutschen Beobachtungseinheiten abgetragen. Bei den rot markierten Regionen handelt es sich um die Landkreise der heutigen Landeshauptstädte der 16 deutschen Bundesländer. Analog zu den schottischen Betrachtungen, dienen diese wichtigen Städte als Orientierung bei den späteren Interpretationen. Von Süden nach Norden betrachtet lauten die Namen der Städte in den hervorgehobenen Landkreisen mit dem zugehörigen Bundesland

in Klammern folgendermaßen: München (Bayern), Stuttgart (Baden-Württemberg), Saarbrücken (Saarland), Mainz (Rheinland-Pfalz), direkt daran angrenzend Wiesbaden (Hessen), nordwestlich davon Düsseldorf (Nordrhein-Westfalen), zentral Erfurt (Thüringen), östlich Dresden (Sachsen), nördlich davon Ostberlin und Westberlin (Berlin), direkt im Westen daran angrenzend Potsdam (Brandenburg), westlich davon Magdeburg (Sachsen-Anhalt), weiter westlich Hannover (Niedersachsen), nordwestlich davon Bremen (Bremen), zentral Hamburg (Hamburg), östlich Schwerin (Mecklenburg-Vorpommern) und als nördlichste Landeshauptstadt Kiel (Schleswig-Holstein).

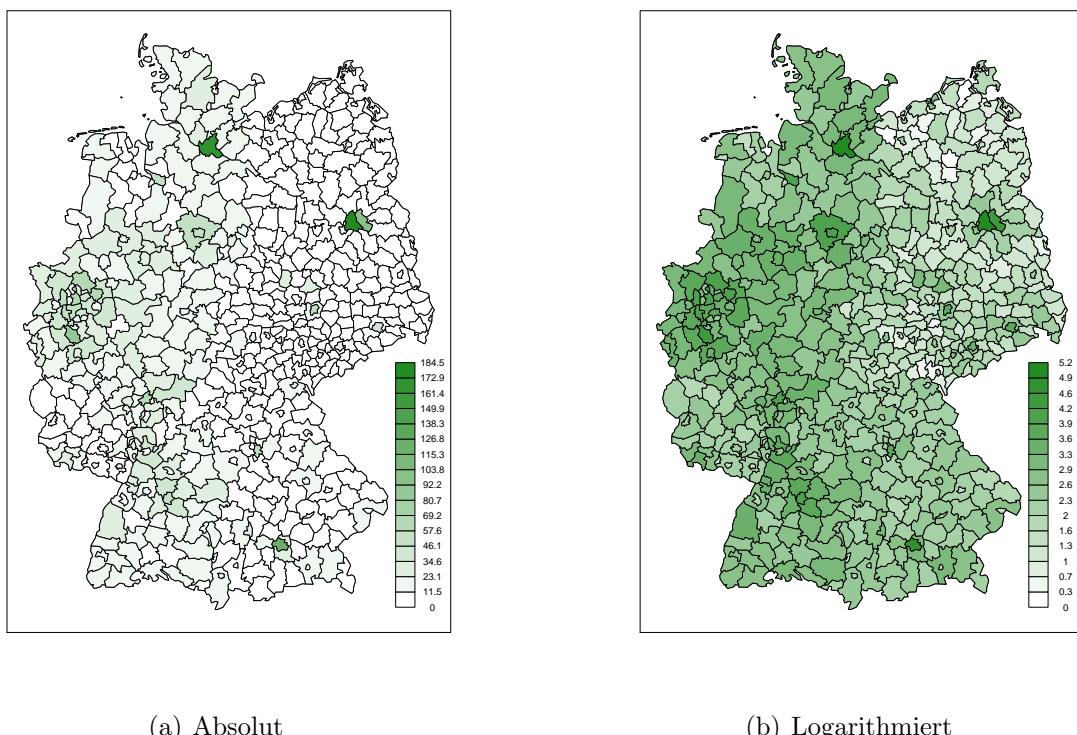


**Abbildung 19:** Umriss der 544 deutschen Landkreisgrenzen zum Zeitpunkt der Datenerhebung. Hervorgehoben: Landkreise der heutigen Landeshauptstädte (rot).

In den anschließenden Grafiken sind die beobachteten und die erwarteten Sterbefälle sowohl in absoluter als auch in logarithmierter Form dargestellt. In den Darstellungen von Abbildung 20 ist zu erkennen, dass in Westberlin die Ausprägung mit einem Niveau von 215 registrierten Todesfälle am höchsten ist. Die Regionen um Hamburg und München folgen mit großem Abstand bei einer Anzahl von 119 bzw. 111 beobachteten Fällen. Für Ostberlin, den Bereich um Hannover, um Stuttgart und um Mainz/Wiesbaden sowie für die Großstädte in der Nähe von Düsseldorf kann ebenfalls ein höheres Aufkommen festgestellt werden. Die Häufigkeiten befinden sich jedoch allesamt unterhalb der Grenze von 100 beobachteten Todesfällen. Die verhältnismäßige Einfärbung des Ausreißers Westberlin kann durch eine logarithmierte Skala geeignet angepasst werden.



**Abbildung 20:** Beobachtete Sterbefälle aus den deutschen Kehlkopfkrebsdaten

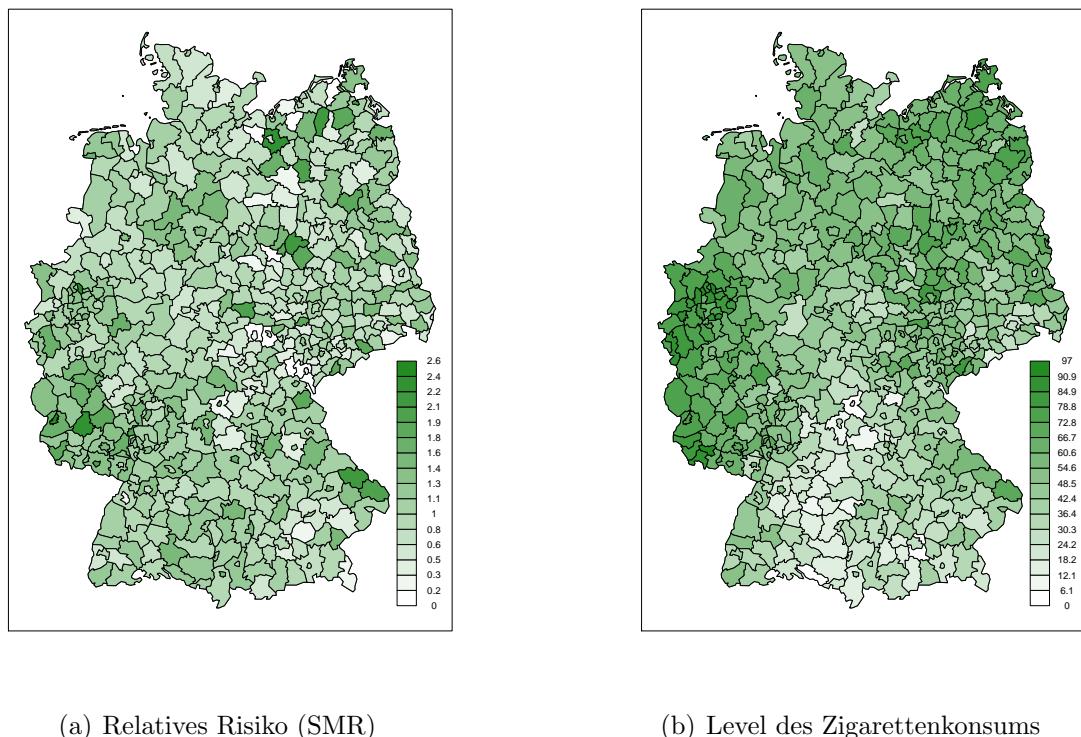


**Abbildung 21:** Erwartete Sterbefälle aus den deutschen Kehlkopfkrebsdaten

Allgemein verzeichnen sowohl die beobachteten als auch die erwarteten Sterbefälle aus Abbildung 21 ein ähnliche Aufteilung der Ausprägungen. Zudem ist für beide Betrachtungen zu erkennen, dass ein West-Ost-Gefälle vorliegt, da in den Landkreisen der ehemaligen DDR weniger Todesfälle aufgrund von Kehlkopfkrebs verzeichnet wurden. Daraus lässt sich möglicherweise schließen, dass das Auftreten von Kehlkopfkrebs bereits gut durch die Zusammensetzung der Bevölkerung eines Landkreises erklärt werden kann, welche in die Bestimmung der erwarteten Häufigkeiten einfließt.

Aufgrund dieser Begebenheit fällt der SMR, abgebildet in Abbildung 22(a), für die einzelnen Landkreise in Hinblick auf die schottischen Verhältnisse mit einem Maximalwert von 2,6 (Landkreis bzw. kreisfreie Stadt Trier nordwestlich von Saarbrücken) vergleichsweise gering aus. Des Weiteren ist keine Häufung um bestimmte Punkte der Karte zu erkennen, viel eher variiert die Höhe des SMR über gesamt Deutschland hinweg. Für die Ausprägungen der Kovariable „Level des Zigarettenkonsums“, welche als Risikofaktor für das Auftreten von Kehlkopfkrebs zu betrachten ist, stellt sich das anders dar, da in den Bundesländern Saarland, Rheinland-Pfalz, Nordrhein-Westfalen und in Teilen von Hessen sowie mit ansteigender Tendenz in Richtung Nordosten der ehemaligen DDR eine erhöhtes Level zu verzeichnen ist.

Dieses Level setzt sich aus verschiedenen Faktoren zusammen. Unter anderem werden die Anzahl an verkauften Zigaretten pro Erwachsenen und der jeweilige Teergehalt bei der Bestimmung berücksichtigt (vgl. Knorr-Held und Rainer, 2001, S. 125).



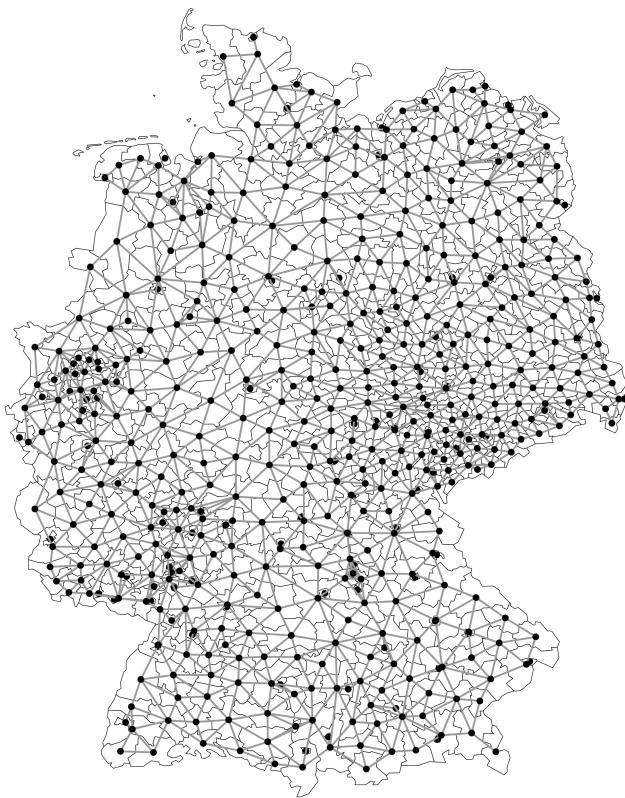
(a) Relatives Risiko (SMR)

(b) Level des Zigarettenkonsums

**Abbildung 22:** Zusätzliche Merkmale aus den deutschen Kehlkopfkrebsdaten

Analog zu der schottischen Aufbereitung kann für die damalige deutsche Verwaltungsgliederung in 544 Landkreise ein Nachbarschaftsgraph erstellt werden. Anders als bei

den schottischen Distrikten hat jedoch jeder Landkreise mindestens eine direkte Grenze zu einer anderen Beobachtungseinheit, wodurch sich folgender Graph ergibt:



**Abbildung 23:** Struktur der 1416 direkten Nachbarschaften der 544 deutschen Landkreise (Grafik übernommen aus Fahrmeir und Kneib, 2010, S. 304)

## 3.2 Umsetzung der hierarchischen Bayes-Struktur

In diesem Unterkapitel wird die Zusammensetzung für den hierarchischen Modellaufbau im Zuge der Bayes-Inferenz inklusive der einzelnen Elemente aufbereitet. Diesbezüglich dient eine Zusammenfassung der Verteilungsannahmen aller hierarchischen Stufen der bayesianischen Modellierung sowie eine Erläuterung zu den jeweiligen Bestandteile im Detail einer besseren Übersicht über die verwendete Modellstruktur. Um eine möglichst zugängliche Umsetzung zu gewährleisten, stellt in diesem Zusammenhang eine Umstrukturierung der Parameter ein geeignetes Mittel dar. Diese soll im Rahmen der Betrachtung der Posteriori-Verteilung und der vollständig bedingten Dichten zunächst eingeführt und daran anschließend umgesetzt werden. Dieses Unterkapitel bildet somit die Grundlage für die spätere Implementierung der vorliegenden Modellsituation (siehe Abschnitt 3.3).

### 3.2.1 Verteilungsannahmen aller hierarchischen Stufen

Für alle Betrachtungen steht eine poissonverteilte Zielgröße  $y_s$  – korrigiert um den Offset  $e_s$  – im Fokus der Modellierung. Zusätzlich findet sowohl für die deutschen als auch für die schottischen Daten lediglich eine Einflussgröße  $x_s$  mit dem Koeffizienten

$\beta$  eine Berücksichtigung (vgl. Unterkapitel 3.1 zur Datengrundlage). Die Parameter  $\gamma_s$  und  $\alpha_s$  repräsentieren zudem den strukturierten und den unstrukturierten Effekt, für welche die zusätzlichen Hyperparameter  $\kappa$  und  $\tau$  sowie  $w_{sr}$  eine wichtige Rolle einnehmen. Des Weiteren wird der Intercept  $\mu$  mit einer eigenen Verteilungsannahme in die Modellierung mit aufgenommen. Über den gesamten Parameterraum ergeben sich somit im Einzelnen folgenden Verteilungsannahmen:

$$y_s | \mu, x_s \beta, \gamma_s, \alpha_s \stackrel{\text{iid}}{\sim} Po(e_s \lambda_s) \quad (15)$$

$$\mu \sim \mathcal{N}(v_\mu = 0, \sigma_\mu^2) \quad \text{mit} \quad \sigma_\mu^2 \rightarrow \infty \quad (16)$$

$$\beta \sim \mathcal{N}(\mu_\beta = 0, \sigma_\beta^2) \quad \text{mit} \quad \sigma_\beta^2 \rightarrow \infty \quad (17)$$

$$\gamma \sim \mathcal{MVN}(0, \kappa^{-1} \mathbf{Q}_w^{-1}) \quad (18)$$

$$\kappa \sim G(a = 1, b = 10^{-3}) \quad (19)$$

$$w_{sr} \sim G(e = \nu/2, f = \nu/2) \quad (20)$$

$$\alpha \stackrel{\text{iid}}{\sim} \mathcal{MVN}(0, \mathbf{T}^{-1}) \quad (21)$$

$$\tau \sim G(g = 1, h = 10^{-5}) \quad (22)$$

Vorzugsweise sollen bei einer bayesianischen Modellierung ausschließlich tatsächlich vorliegende a priori Informationen berücksichtigt werden. Aus diesem Grund wird für Parameter, über die vorab keine expliziten Hintergrundinformationen vorliegen in der Regel eine möglichst uninformativen bzw. flache Priori-Verteilung angenommen mit der zusätzlichen Berücksichtigung einer kompakten Modellierung durch die Wahl von konjugierten Verteilungen.

Da bei einer bayesianischen Auffassung alle Parameter des Modells stochastischer Natur sind, wird sowohl für den Intercept  $\mu$  als auch für den Kovariablen-Koeffizienten  $\beta$  ebenso eine Priori-Verteilungen angenommen (vgl. Verteilung (16) und (17)). In Anlehnung an die frequentistische Sichtweise erweist sich an dieser Stelle meist eine Normalverteilung mit den festen Parametern  $v_\mu$  und  $\sigma_\mu^2$  bzw.  $\mu_\beta$  und  $\sigma_\beta^2$  als geeignete Wahl. Durch den Grenzfall  $\sigma_\mu^2 \rightarrow \infty$  (bzw.  $\sigma_\beta^2 \rightarrow \infty$ ) erhält man schließlich näherungsweise die gewünschten flachen Verteilungen mit  $p(\mu), p(\beta) \propto$  konstant, welche a priori somit keine Zusatzinformation in die Modellierung einbringt. Bei der Aufstellung der vollständig bedingten Dichten erweist es sich zudem als günstig für diese beiden Parameter eine Normalverteilungsannahme getroffen zu haben (vgl. Fahrmeir *et al.*, 2009, S. 272).

Die Entstehung der multivariaten Normalverteilung eines GMZF  $\gamma$ , aufgeführt in (18), wurde bereits ausführlich in Abschnitt 2.4.3 erläutert. Dort fließen als Hyperparameter die Präzision  $\kappa$ , welche als globaler Glättungsparameter betrachtet werden kann, und die Gewichte  $w_{sr}$  über die Nachbarschaftsmatrix  $\mathbf{Q}_w$  in den Aufbau mit ein. Analog zu der Intention bei der Wahl der uninformativen Priori-Verteilung für  $\mu$  und  $\beta$  wird die Verteilung von  $\kappa$  gewählt (siehe Verteilung (19)). Als konjugierter Partner zur Normalverteilung von  $\gamma$  gilt dabei eine Gamma-Verteilung mit den Parametern  $a = 1$  und  $b = 10^{-3}$  in der Praxis als übliche Wahl (vgl. Schmid, 2004, S. 16 f.). In Abschnitt 2.4.4 wurde zudem auf die beiden verschiedenen Möglichkeiten bei der Bestimmung der Priori-Verteilung der Gewichte  $w_{sr}$  bzw. der logarithmierten Gewichte  $v_{sr}$  eingegangen.

Brezger *et al.* (2007) schlagen in diesem Zusammenhang bei der Modellierung von  $w_{sr}$  über eine Gamma-Verteilung eine kleine Anzahl an Freiheitsgraden vor, beispielsweise über  $\nu = 1$  (siehe (20)). Als Alternative hierzu erläutern sie in ihren Ausführungen die Annahme einer strukturierten Verteilung von  $v_{sr}$ , repräsentiert durch ein weiteres GMZF und einer zusätzlichen hierarchischen Stufe mit:

$$p(\mathbf{v} | \phi) \propto \exp\left(-\frac{1}{2}\mathbf{v}^T \phi \mathbf{L} \mathbf{v}\right)$$

Fahrmeir *et al.* (2009) schlagen für den Vektor der unstrukturierten Effekte  $\boldsymbol{\alpha}$  eine multivariate Normalverteilung mit Erwartungswert 0 und eine Diagonalmatrix  $\mathbf{T}$  für die Kovarianz (siehe Verteilung in (21)) sowie eine schwach informative Gamma-Verteilung für den Präzisionsparameter  $\tau$  vor (siehe (22)). Durch die Beschränkung auf die diagonalen Einträge der Kovarianzmatrix kann umgangen werden, dass für  $\mathbf{T}$  eine Wishart-Verteilung gebildet werden muss, da die gemeinsame Verteilung der einzelnen Hyperparameter  $\tau$  letztlich in das Produkt unabhängiger Gamma-Verteilungen zerfällt (vgl. Fahrmeir *et al.*, 2009, S. 276). Um die Anzahl der zu schätzenden Parameter zu reduzieren, ist es außerdem sinnvoll die Diagonalmatrix  $\mathbf{T}$  mit einem einzigen zufällig gezogenen  $\tau$  zu befüllen. Demzufolge entspricht  $\mathbf{T} = \text{diag}(\tau, \dots, \tau)$ .

Grundsätzlich ist es erstrebenswert, dass Schwankungen in den Daten, die nicht aufgrund der Ausprägungen der Kovariable beschrieben werden können, eher durch einen strukturierten, als durch einen zufälligen (unstrukturierten) Effekt erklärt werden. Deshalb wird die Gamma-Verteilung der Präzision  $\tau$  a priori noch uninformativer gewählt als die von  $\kappa$ , wodurch die Festlegung der Parameter auf  $b = 10^{-3}$  und  $h = 10^{-5}$  motiviert werden kann. Da es sich bei den unstrukturierten Effekten  $\boldsymbol{\alpha}$  um unabhängig und identisch verteilte Effekte handelt und  $\tau$  für alle  $\alpha_s$  identisch ist, kann neben der Matrixschreibweise der multivariaten Normalverteilung:

$$p(\boldsymbol{\alpha} | \mathbf{T}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \cdot |\mathbf{T}|^{\frac{1}{2}} \cdot \exp\left(-\frac{1}{2}\boldsymbol{\alpha}^T \mathbf{T} \boldsymbol{\alpha}\right)$$

auch eine Darstellung über das Produkt der einzelnen Verteilungen für jede Region verwendet werden:

$$\begin{aligned} p(\boldsymbol{\alpha} | \mathbf{T}) &= \prod_{s=1}^d \frac{\tau^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \cdot \exp\left(-\frac{1}{2}\tau\alpha_s^2\right) \\ &\stackrel{\text{iid}}{=} \frac{\tau^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}} \cdot \exp\left(-\frac{\tau}{2} \sum_{s=1}^d \alpha_s^2\right). \end{aligned}$$

Abschließend ist festzuhalten, dass sich die beschriebenen Annahmen zunächst auf alle möglichen Bestandteile bis inklusive der 3. hierarchischen Stufe mit den üblichen Parameterwerten für  $a, b, g, h$  und  $\nu$  beziehen. Die späteren Analysen werden jedoch zeigen, dass die Modellzusammensetzung für die Schätzung adaptiver Gewichte angepasst werden muss. Ausschlaggebend ist hierfür die hohe Menge der zu schätzenden Parameter im Verhältnis zu der relativ geringen Anzahl an Beobachtungen.

### 3.2.2 Reparametrisierung des Modells

Nachdem im vorangegangenen Unterkapitel die Annahmen an die Verteilungen der einzelnen Parameter(blöcke) der Modellierung motiviert wurden, soll im Folgenden die Umsetzung eines sogenannten Hybrid-Samplers vorbereitet werden (siehe hierfür Unterkapitel 3.3.1 zu den MCMC Verfahren). Diesbezüglich wird eine Reparametrisierung des Modells durchgeführt, die für eine Erleichterung bei dem späteren Ziehen von Zufallszahlen sorgt. Bereits bei der Aufstellung der vollständig bedingten Dichten in Abschnitt 3.2.3 ist zu erkennen, dass sich die Umstrukturierung der Parameter auszahlt. Zunächst soll jedoch auf die Veränderung bei der Zusammensetzung der Posteriori-Verteilung eingegangen werden.

Für die ursprüngliche Posteriori-Verteilung ergibt sich nach dem proportionalen bayesianischen Zusammenhang aus Gleichung (2) von Seite 11 und unter der Berücksichtigung aller beschriebenen Parameter folgende Konstellation:

$$\begin{aligned} p(\mu, \beta, \boldsymbol{\gamma}, \kappa, \mathbf{w}, \boldsymbol{\alpha}, \mathbf{T} | \mathbf{y}) \\ \propto p(\mathbf{y} | \mu, \beta, \boldsymbol{\gamma}, \kappa, \mathbf{w}, \boldsymbol{\alpha}, \mathbf{T}) \cdot p(\mu) \cdot p(\beta) \cdot p(\boldsymbol{\gamma}, \kappa, \mathbf{w}) \cdot p(\boldsymbol{\alpha}, \mathbf{T}) \\ = p(\mathbf{y} | \mu, \beta, \boldsymbol{\gamma}, \kappa, \mathbf{w}, \boldsymbol{\alpha}, \mathbf{T}) \cdot p(\mu) \cdot p(\beta) \cdot p(\boldsymbol{\gamma} | \kappa, \mathbf{w}) \cdot p(\kappa) \cdot p(\mathbf{w}) \cdot p(\boldsymbol{\alpha} | \mathbf{T}) \cdot p(\mathbf{T}) \end{aligned}$$

Es lässt sich schnell erkennen, dass daraus keine bekannte Verteilung gebildet werden kann. Aus diesem Grund ist es notwendig, die vollständig bedingten Dichten der einzelnen Parameter(blöcke) zu betrachten.

Bevor dies im Einzelnen geschieht, erweist es sich als sinnvoll, die erwähnte Reparametrisierung über den Prädiktor  $\boldsymbol{\eta}$  und den unstrukturierten Effekt  $\boldsymbol{\alpha}$  durchzuführen. Dabei wird der Parametervektor  $\boldsymbol{\theta} = (\mu, \beta, \boldsymbol{\gamma}^T, \boldsymbol{\alpha}^T)^T$  zu  $\boldsymbol{\theta}_{\text{Rep}} = (\mu, \beta, \boldsymbol{\gamma}^T, \boldsymbol{\eta}^T)^T$  umgeformt, wodurch sich aus der ursprünglichen Form des Prädiktors  $\boldsymbol{\eta} = \mu \mathbf{1}_{[d]} + \mathbf{x}\beta + \boldsymbol{\gamma} + \boldsymbol{\alpha}$  schließlich  $\boldsymbol{\alpha} = \boldsymbol{\eta} - \mu \mathbf{1}_{[d]} - \mathbf{x}\beta - \boldsymbol{\gamma}$  für den unstrukturierten Effekt ergibt (vgl. Rue und Held, 2005, S. 174). Der große Vorteil dieser Reparametrisierung ist die Zusammensetzung der vollständig bedingten Dichten. Wie an späterer Stelle zu sehen sein wird, ergeben sich dadurch für den Koeffizienten  $\beta$  der Einflussgröße und für den Parameterblock  $\boldsymbol{\gamma}$ , als Repräsentant für das GMZF, bekannte (multivariate) Normalverteilungen, aus welchen über einen Gibbs-Sampler direkt Zufallszahlen generiert werden können (siehe Teilabschnitt aus Unterkapitel 3.3.1). Lediglich für die Ziehung der Prädiktorwerte  $\eta_s$  und der Gewichte  $w_{sr}$  werden Metropolis-Hastings-Schritte benötigt (vgl. Schmid, 2004, S. 34). Knorr-Held und Rue (2002) fassen diesen Vorteil folgendermaßen zusammen:

„One of the advantages of the reparametrized model is that the full conditional for the spatially structured parameters is now multivariate Gaussian, so updating can be done with a simple (multivariate) Gibbs step“ (Knorr-Held und Rue, 2002, S. 603 f.).

Die Randverteilung des Prädiktors  $\eta_s$  bildet sich dann aus der Verteilung der unstrukturierten Effekte  $\alpha_s$  und der weiteren Zusammensetzung der Parameter zu (vgl. Rue und Held, 2005, S. 174):

$$\eta_s | \mu, x_s \beta, \gamma_s \sim \mathcal{N}(\mu + x_s \beta + \gamma_s, \tau^{-1})$$

Aufgrund der Reparametrisierung ergibt sich außerdem für die Posteriori-Verteilung folgende, neue Zusammensetzung:

$$\begin{aligned}
& p(\mu, \beta, \gamma, \kappa, \mathbf{w}, \boldsymbol{\eta}, \mathbf{T} | \mathbf{y}) \\
& \propto p(\mathbf{y} | \boldsymbol{\eta}) \cdot p(\mu) \cdot p(\beta) \cdot p(\gamma | \kappa, \mathbf{w}) \cdot p(\kappa) \cdot p(\mathbf{w}) \cdot p(\boldsymbol{\eta}) \cdot p(\mathbf{T}) \\
& = \prod_{s=1}^d \frac{[e_s \exp(\eta_s)]^{y_s}}{y_s!} \cdot \exp \left[ -e_s \exp(\eta_s) \right] \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\mathbf{y} | \boldsymbol{\eta}) \\
& \quad \frac{1}{(2\pi\sigma_\mu^2)^{\frac{1}{2}}} \cdot \exp \left[ -\frac{1}{2\sigma_\mu^2} (\mu - v_\mu)^2 \right] \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\mu) \\
& \quad \frac{1}{(2\pi\sigma_\beta^2)^{\frac{1}{2}}} \cdot \exp \left[ -\frac{1}{2\sigma_\beta^2} (\beta - \mu_\beta)^2 \right] \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\beta) \\
& \quad |Q_w|^{1/2} \cdot \kappa^{rg(Q_w)/2} \cdot \exp \left[ -\frac{1}{2} \gamma^T \kappa Q_w \gamma \right] \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\gamma | \kappa, \mathbf{w}) \\
& \quad \frac{b^a}{\Gamma(a)} \cdot \kappa^{(a-1)} \cdot \exp \left[ -b\kappa \right] \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\kappa) \\
& \quad \prod_{s \sim r}^{\delta_r} \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \cdot w_{sr}^{(\nu/2-1)} \cdot \exp \left[ -\frac{\nu}{2} w_{sr} \right] \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\mathbf{w}) \\
& \quad \prod_{s=1}^d \frac{\tau^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \cdot \exp \left[ -\frac{\tau}{2} (\eta_s - \mu - x_s \beta - \gamma_s)^2 \right] \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\boldsymbol{\eta}) \\
& \quad \prod_{s=1}^d \frac{h^g}{\Gamma(g)} \cdot \tau^{(g-1)} \cdot \exp(-h\tau) \cdot \left. \begin{array}{l} \\ \\ \end{array} \right\} p(\mathbf{T})
\end{aligned}$$

Da es sich hierbei wiederum um eine unbekannte gemeinsame Verteilung handelt, werden für die einzelnen Parameterschätzungen anhand von MCMC Verfahren die vollständig bedingten Dichten aller hierarchischen Stufen benötigt. Diese lassen sich über eine proportionale Betrachtungsweise aus der soeben dargestellten gemeinsamen Posteriori-Verteilung herleiten.

### 3.2.3 Aufstellung der vollständig bedingten Dichten

Die Betrachtung der gemeinsamen Poisson-Verteilung des Beobachtungsmodells, angedeutet in Verteilungsannahme (15) auf Seite 55, ist elementar bei der Bildung der jeweiligen vollständig bedingten Dichten. Nach der Reparametrisierung stellt sie sich wie folgt dar:

$$\begin{aligned}
p(\mathbf{y} | \boldsymbol{\eta}) & = \prod_{s=1}^d \frac{[e_s \exp(\eta_s)]^{y_s}}{y_s!} \cdot \exp \left[ -e_s \exp(\eta_s) \right] \\
& \stackrel{\text{iid}}{=} \frac{1}{y_1! \cdot \dots \cdot y_d!} \cdot [e_s \exp(\eta_s)]^{\sum_{s=1}^d y_s} \cdot \exp \left\{ -d \left[ e_s \exp(\eta_s) \right] \right\}
\end{aligned} \tag{23}$$

Die Likelihood ist ein fester Bestandteil bei der Bildung der vollständig bedingten Dichten. In den meisten Fällen ermöglicht es die erfolgte Umstrukturierung der Parameter jedoch, dieses Element aufgrund einer proportionalen Betrachtungsweise unberücksichtigt zu lassen, wodurch bekannte Verteilungen konstruiert werden können. Im Folgenden sollen die vollständig bedingten Dichten aller Parameter des Modells gebildet werden.

### Vollständig bedingte Dichte des Intercepts $\mu$

$$\begin{aligned} p(\mu \mid \cdot) &\stackrel{\text{bzgl.}}{\propto} p(\boldsymbol{y} \mid \boldsymbol{\eta}) \cdot p(\mu) \cdot p(\boldsymbol{\alpha} \mid \boldsymbol{T}) \\ &\stackrel{\mu}{\propto} \frac{1}{(2\pi\sigma_\mu^2)^{\frac{1}{2}}} \cdot \exp\left[-\frac{1}{2\sigma_\mu^2}(\mu - v_\mu)^2\right] \cdot \left. \frac{\tau^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}} \cdot \exp\left[-\frac{\tau}{2}\sum_{s=1}^d \alpha_s^2\right] \right\} p(\mu) \\ &\quad \left. \left. \right\} p(\boldsymbol{\alpha} \mid \boldsymbol{T}) \right. \end{aligned}$$

mit Reparametrisierung:  $\alpha_s = \eta_s - \mu - x_s\beta - \gamma_s$

$$\begin{aligned} &\stackrel{\text{bzgl.}}{\propto} \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 - 2\mu v_\mu + v_\mu^2}{\sigma_\mu^2}\right]\right\} \cdot \exp\left\{-\frac{1}{2}\left[\frac{\sum_{s=1}^d (\eta_s - \mu - x_s\beta - \gamma_s)^2}{\tau^{-1}}\right]\right\} \\ &\stackrel{\mu}{\propto} \exp\left\{-\frac{1}{2}\left[\frac{\mu^2 - 2\mu v_\mu}{\sigma_\mu^2}\right] - \frac{1}{2}\left[\frac{\sum_{s=1}^d (\mu^2 - 2\mu\eta_s + 2\mu x_s\beta + 2\mu\gamma_s)}{\tau^{-1}}\right]\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\left(\mu^2\sigma_\mu^{-2} - 2\mu v_\mu\sigma_\mu^{-2}\right) + \tau\left(\mu^2 d + \sum_{s=1}^d 2\mu(x_s\beta + \gamma_s - \eta_s)\right)\right]\right\} \\ &= \exp\left\{-\frac{1}{2}\left[\mu^2\left(\sigma_\mu^{-2} + \tau d\right) - 2\mu\left(\underbrace{v_\mu\sigma_\mu^{-2} - \tau\sum_{s=1}^d (x_s\beta + \gamma_s - \eta_s)}_{\mathbf{I}}\right)\right]\right\} \\ &\stackrel{\text{bzgl.}}{\propto} \underbrace{\exp\left\{-\frac{1}{2}\underbrace{\left(\sigma_\mu^{-2} + \tau d\right)}_{\tilde{\sigma}_\mu^{-2}}\left[\mu - \underbrace{\left(\sigma_\mu^{-2} + \tau d\right)^{-1}}_{\tilde{v}_\mu}\right]^2\right\}}_{\text{Kern einer Normalverteilung}} \\ &\quad \Downarrow \\ &\mu \mid \boldsymbol{x}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \kappa, \boldsymbol{w}, \boldsymbol{\eta} \sim \mathcal{N}(\tilde{v}_\mu, \tilde{\sigma}_\mu^2) \end{aligned}$$

Vollständig bedingte Dichte des Kovariablen-Koeffizienten  $\beta$

mit Reparametrisierung:  $\alpha_s = \eta_s - \mu - x_s\beta - \gamma_s$

$$\begin{aligned}
& \text{bzgl. } \beta \propto \exp \left\{ -\frac{1}{2} \left[ \frac{\beta^2 - 2\beta\mu_\beta + \mu_\beta^2}{\sigma_\beta^2} \right] \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \frac{\sum_{s=1}^d (\eta_s - \mu - x_s \beta - \gamma_s)^2}{\tau^{-1}} \right] \right\} \\
& \text{bzgl. } \beta \propto \exp \left\{ -\frac{1}{2} \left[ \frac{\beta^2 - 2\beta\mu_\beta}{\sigma_\beta^2} \right] - \frac{1}{2} \left[ \frac{\sum_{s=1}^d (x_s^2 \beta^2 + 2x_s \beta \gamma_s + 2x_s \beta \mu - 2x_s \beta \eta_s)}{\tau^{-1}} \right] \right\} \\
& = \exp \left\{ -\frac{1}{2} \left[ \left( \beta^2 \sigma_\beta^{-2} - 2\beta\mu_\beta \sigma_\beta^{-2} \right) + \tau \left( \sum_{s=1}^d x_s^2 \beta^2 + \sum_{s=1}^d 2\beta x_s (\gamma_s + \mu - \eta_s) \right) \right] \right\} \\
& = \exp \left\{ -\frac{1}{2} \left[ \beta^2 \left( \sigma_\beta^{-2} + \tau \sum_{s=1}^d x_s^2 \right) - 2\beta \underbrace{\left( \mu_\beta \sigma_\beta^{-2} - \tau \sum_{s=1}^d x_s (\gamma_s + \mu - \eta_s) \right)}_{II} \right] \right\} \\
& \text{bzgl. } \beta \propto \underbrace{\exp \left\{ -\frac{1}{2} \underbrace{\left( \sigma_\beta^{-2} + \tau \sum_{s=1}^d x_s^2 \right)}_{\tilde{\sigma}_\beta^{-2}} \left[ \beta - \underbrace{\left( II \right) \cdot \left( \sigma_\beta^{-2} + \tau \sum_{s=1}^d x_s^2 \right)^{-1}}_{\tilde{\mu}_\beta} \right]^2 \right\}}_{\text{Kern einer Normalverteilung}} \\
& \Downarrow \\
& \beta \mid \mu \mathbf{1}_{[d]}, \boldsymbol{\gamma}, \kappa, \mathbf{w}, \boldsymbol{\eta} \sim \mathcal{N}(\tilde{\mu}_\beta, \tilde{\sigma}_\beta^2)
\end{aligned}$$

## Vollstndig bedingte Dichte des GMZF $\gamma$

Für die Aufstellung der vollständig bedingten Dichte des GMZF, ist die Verwendung einer kanonischen Darstellung der multivariaten Normalverteilung hilfreich. Dabei wird ein GMZF durch  $\gamma$  und die Präzisionsmatrix durch  $P$  repräsentiert. Der bisher nicht verwendete Mittelwertsvektor wird in diesem Zusammenhang über den Parameter  $\mu_\gamma$  dargestellt. Folgende Umformung verwendet zu Beginn den gewöhnlichen Kern einer

multivariaten Normalverteilung mit  $\boldsymbol{\gamma} \sim \mathcal{MVN}(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \mathbf{P}^{-1})$  (vgl. Schmid, 2004, S. 19):

$$\begin{aligned} p(\boldsymbol{\gamma} | \cdot) &\stackrel{\text{bzgl.}}{\propto} \exp \left[ -\frac{1}{2} (\boldsymbol{\gamma} - \boldsymbol{\mu}_{\boldsymbol{\gamma}})^T \mathbf{P} (\boldsymbol{\gamma} - \boldsymbol{\mu}_{\boldsymbol{\gamma}}) \right] \\ &\stackrel{\text{bzgl.}}{\propto} \exp \left[ -\frac{1}{2} (\boldsymbol{\gamma}^T \mathbf{P} \boldsymbol{\gamma} - 2 \boldsymbol{\gamma}^T \mathbf{P} \boldsymbol{\mu}_{\boldsymbol{\gamma}} + \boldsymbol{\mu}_{\boldsymbol{\gamma}}^T \mathbf{P} \boldsymbol{\mu}_{\boldsymbol{\gamma}}) \right] \\ &\stackrel{\text{bzgl.}}{\propto} \exp \left[ -\frac{1}{2} \boldsymbol{\gamma}^T \mathbf{P} \boldsymbol{\gamma} + \boldsymbol{\gamma}^T \mathbf{b} \right] \quad \text{mit } \mathbf{b} = \mathbf{P} \boldsymbol{\mu}_{\boldsymbol{\gamma}} \end{aligned}$$

Die Verteilung von  $\boldsymbol{\gamma}$  kann somit durch  $\boldsymbol{\gamma} | \cdot \sim \mathcal{MVN}_C(\mathbf{b}, \mathbf{P})$  dargestellt werden. In die standardmäßige Darstellung übertragen entspräche das  $\boldsymbol{\gamma} | \cdot \sim \mathcal{MVN}(\mathbf{P}^{-1} \mathbf{b}, \mathbf{P}^{-1})$ . Für die vollständig bedingte Dichte von  $\boldsymbol{\gamma}$  ergibt sich dann:

$$\begin{aligned} p(\boldsymbol{\gamma} | \cdot) &\stackrel{\text{bzgl.}}{\propto} p(\mathbf{y} | \boldsymbol{\eta}) \cdot p(\boldsymbol{\gamma} | \kappa, \mathbf{w}) \cdot p(\boldsymbol{\alpha} | \mathbf{T}) \\ &\stackrel{\text{bzgl.}}{\propto} |\mathbf{Q}_{\mathbf{w}}|^{1/2} \cdot \kappa^{rg(\mathbf{Q}_{\mathbf{w}})/2} \cdot \exp \left( -\frac{1}{2} \boldsymbol{\gamma}^T \kappa \mathbf{Q}_{\mathbf{w}} \boldsymbol{\gamma} \right) \cdot \quad \left. \right\} p(\boldsymbol{\gamma} | \kappa, \mathbf{w}) \\ &\quad \frac{\tau^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}} \cdot \exp \left( -\frac{\tau}{2} \sum_{s=1}^d \alpha_s^2 \right) \quad \left. \right\} p(\boldsymbol{\alpha} | \mathbf{T}) \end{aligned}$$

mit Reparametrisierung:  $\alpha_s = \eta_s - \mu - x_s \beta - \gamma_s$

$$\begin{aligned} &\stackrel{\text{bzgl.}}{\propto} \exp \left\{ -\frac{1}{2} \boldsymbol{\gamma}^T \kappa \mathbf{Q}_{\mathbf{w}} \boldsymbol{\gamma} \right\} \cdot \exp \left\{ -\frac{1}{2} \left[ \frac{\sum_{s=1}^d (\eta_s - \mu - x_s \beta - \gamma_s)^2}{\tau^{-1}} \right] \right\} \\ &\stackrel{\text{bzgl.}}{\propto} \exp \left\{ -\frac{1}{2} \boldsymbol{\gamma}^T \kappa \mathbf{Q}_{\mathbf{w}} \boldsymbol{\gamma} - \frac{1}{2} \left[ \frac{\sum_{s=1}^d (\gamma_s^2 - 2\gamma_s(\eta_s - \mu - x_s \beta))}{\tau^{-1}} \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \boldsymbol{\gamma}^T \kappa \mathbf{Q}_{\mathbf{w}} \boldsymbol{\gamma} - \frac{1}{2} \left[ \frac{\boldsymbol{\gamma}^T \boldsymbol{\gamma} - 2\boldsymbol{\gamma}^T (\boldsymbol{\eta} - \mu \mathbf{1}_{[d]} - \mathbf{x} \beta)}{\tau^{-1}} \right] \right\} \\ &= \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\gamma}^T \kappa \mathbf{Q}_{\mathbf{w}} \boldsymbol{\gamma} + \boldsymbol{\gamma}^T \tau \mathbf{1}_{[d \times d]} \boldsymbol{\gamma} \right] + \boldsymbol{\gamma}^T \tau [\boldsymbol{\eta} - \mu \mathbf{1}_{[d]} - \mathbf{x} \beta] \right\} \\ &= \underbrace{\exp \left\{ -\frac{1}{2} \underbrace{\boldsymbol{\gamma}^T \left[ \underbrace{\tau \mathbf{1}_{[d \times d]} + \kappa \mathbf{Q}_{\mathbf{w}}}_{\tilde{\mathbf{P}}_{\mathbf{w}}} \right] \boldsymbol{\gamma}}_{b} + \boldsymbol{\gamma}^T \underbrace{\tau [\boldsymbol{\eta} - \mu \mathbf{1}_{[d]} - \mathbf{x} \beta]}_{b} \right\}}_{\text{Kern einer multivariaten Normalverteilung}} \\ &\Downarrow \\ &\boldsymbol{\gamma} | \mu \mathbf{1}_{[d]}, \mathbf{x} \beta, \kappa, \mathbf{w}, \boldsymbol{\eta} \sim \mathcal{MVN}_C(\mathbf{b}, \tilde{\mathbf{P}}_{\mathbf{w}}^{-1}) \end{aligned}$$

### Vollständig bedingte Dichte des Hyperparameters $\kappa$ des GMZF

$$\begin{aligned}
p(\kappa | \cdot) &\stackrel{\text{bzgl.}}{\propto} p(\gamma | \kappa) \cdot p(\kappa) \\
&= |\mathbf{Q}_w|^{1/2} \cdot \kappa^{rg(\mathbf{Q}_w)/2} \cdot \exp\left(-\frac{1}{2}\gamma^T \mathbf{Q}_w \gamma \kappa\right) \cdot p(\gamma | \kappa) \\
&\quad \frac{b^a}{\Gamma(a)} \cdot \kappa^{(a-1)} \cdot \exp\left(-b\kappa\right) \cdot p(\kappa) \\
&= \underbrace{\kappa^{(a-1)} \cdot \kappa^{rg(\mathbf{Q}_w)/2} \cdot \exp\left[-b\kappa - \frac{1}{2}\gamma^T \mathbf{Q}_w \gamma \kappa\right]}_{\text{Kern einer Gamma-Verteilung}} \\
&\quad \underbrace{\kappa^{(a+rg(\mathbf{Q}_w)/2-1)} \cdot \exp\left[-\left(b + \frac{1}{2}\gamma^T \mathbf{Q}_w \gamma\right)\kappa\right]}_{\text{Kern einer Gamma-Verteilung}} \\
&\quad \Downarrow \\
&\kappa | \mu \mathbf{1}_{[d]}, \mathbf{x}\beta, \gamma, \mathbf{w}, \boldsymbol{\eta} \sim G\left(a + rg(\mathbf{Q}_w)/2, b + \frac{1}{2}\gamma^T \mathbf{Q}_w \gamma\right)
\end{aligned}$$

### Vollständig bedingte Dichte des Hyperparameters $\tau$ von $\alpha$

$$\begin{aligned}
p(\tau | \cdot) &\stackrel{\text{bzgl.}}{\propto} p(\boldsymbol{\alpha} | \mathbf{T}) \cdot p(\tau) \\
&= \frac{\tau^{\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}} \cdot \exp\left(-\frac{\tau}{2} \sum_{s=1}^d \alpha_s^2\right) \cdot p(\alpha_s | \tau) \\
&\quad \frac{h^g}{\Gamma(g)} \cdot \tau^{(g-1)} \cdot \exp\left(-h\tau\right) \cdot p(\tau)
\end{aligned}$$

mit Reparametrisierung:  $\alpha_s = \eta_s - \mu - x_s\beta - \gamma_s$

$$\begin{aligned}
&\stackrel{\text{bzgl.}}{\propto} \tau^{(g-1)} \cdot \tau^{\frac{d}{2}} \cdot \exp\left\{-h\tau - \frac{\tau}{2} \sum_{s=1}^d (\eta_s - \mu - x_s\beta - \gamma_s)^2\right\} \\
&= \underbrace{\tau^{(g+\frac{d}{2}-1)} \cdot \exp\left\{-\left[h + \frac{1}{2} \sum_{s=1}^d (\eta_s - \mu - x_s\beta - \gamma_s)^2\right]\tau\right\}}_{\text{Kern einer Gamma-Verteilung}} \\
&\quad \Downarrow \\
&\tau | \mu \mathbf{1}_{[d]}, \mathbf{x}\beta, \gamma, \kappa, \mathbf{w}, \boldsymbol{\eta} \sim G\left(g + \frac{d}{2}, h + \frac{1}{2} \sum_{s=1}^d (\eta_s - \mu - x_s\beta - \gamma_s)^2\right)
\end{aligned}$$

### Vollständig bedingte Dichte des Prädiktors $\eta_s$

Aufgrund der durchgeführten Reparametrisierung ist die Likelihood, aufgestellt in Gleichung (23), lediglich bei der Bildung der vollständig bedingten Dichten der  $\eta_s$  von Bedeutung. Ebenso wie bei der nachfolgenden proportionalen Betrachtung der Gewichte, folgt diese Zusammensetzung keiner bekannten Verteilung, wodurch bei der späteren Umsetzung die Anwendung von Metropolis-Hastings-Schritten notwendig wird.

$$\begin{aligned}
 p(\eta_s | \cdot) &\stackrel{\text{bzgl.}}{\underset{\eta_s}{\propto}} p(y_s | \eta_s) \cdot p(\eta_s) \\
 &= \frac{1}{y_s!} \cdot \left[ e_s \exp(\eta_s) \right]^{y_s} \cdot \exp \left[ -e_s \exp(\eta_s) \right] \cdot \underbrace{\frac{\tau^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}} \cdot \exp \left[ -\frac{\tau}{2}(\eta_s - \mu - x_s \beta - \gamma_s)^2 \right]}_{p(\eta_s)} \\
 &\stackrel{\text{bzgl.}}{\underset{\eta_s}{\propto}} \exp(\eta_s)^{y_s} \cdot \exp \left[ -e_s \exp(\eta_s) \right] \cdot \exp \left[ -\frac{\tau}{2}(\eta_s - \mu - x_s \beta - \gamma_s)^2 \right] \\
 &= \underbrace{\exp \left[ y_s \eta_s - e_s \exp(\eta_s) - \frac{\tau}{2}(\eta_s - \mu - x_s \beta - \gamma_s)^2 \right]}_{\text{Kern einer unbekannten Verteilung}} \quad (24)
 \end{aligned}$$

↓

$$\eta_s | \mu, x_s \beta, \gamma_s, \kappa, w_{sr} \sim ?$$

Eine mögliche Annäherung an diese unbekannte Verteilung stellt eine Taylorentwicklung zweiten Grades dar, welche in Unterkapitel 3.3.2 ausführlich betrachtet wird.

### Vollständig bedingte Dichte der Hyperparameter $w_{sr}$ des GMZF

Da es sich bei den Gewichten  $w_{sr}$  um Hyperparameter handelt, fließt in deren vollständig bedingte Dichte die Verteilungsannahme des GMZF mit ein. Für die spätere Anwendung der MCMC Verfahren ist in diesem Zusammenhang der Umgang mit der Determinante von  $\mathbf{Q}_w$  von Interesse. Brezger *et al.* (2007) zeigen diesbezüglich in ihrem Paper über adaptive GMZF, dass  $|\mathbf{Q}_w|$  durch das Produkt der Eigenwerte von  $\mathbf{Q}_w$ , die ungleich Null sind, ersetzt werden kann, welche anschließend durch  $\lambda^{eig}$  repräsentiert werden. Daraus ergibt sich für die vollständig bedingte Dichte von  $p(w_{sr} | \cdot)$  folgender Zusammenhang:

$$\begin{aligned}
p(w_{sr} | \cdot) &\stackrel{\text{bzgl.}}{\propto} p(\boldsymbol{\gamma} | w_{sr}) \cdot p(w_{sr}) \\
&= |\mathbf{Q}_w|^{1/2} \cdot \kappa^{rg(\mathbf{Q}_w)/2} \cdot \exp\left(-\frac{1}{2}\boldsymbol{\gamma}^T \kappa \mathbf{Q}_w \boldsymbol{\gamma}\right) \cdot \\
&\quad \left. \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \cdot w_{sr}^{(\nu/2-1)} \cdot \exp\left(-\frac{\nu}{2}w_{sr}\right) \right\} p(\boldsymbol{\gamma} | w_{sr}) \\
&\stackrel{\text{bzgl.}}{\propto} |\mathbf{Q}_w|^{1/2} \cdot w_{sr}^{(\nu/2-1)} \cdot \exp\left[-\frac{\nu}{2}w_{sr} - \frac{1}{2}\boldsymbol{\gamma}^T \kappa \mathbf{Q}_w \boldsymbol{\gamma}\right] \\
&\stackrel{\text{bzgl.}}{\propto} \underbrace{\left(\prod_{s=2}^d \lambda_s^{eig}\right)^{1/2} \cdot w_{sr}^{(\nu/2-1)} \cdot \exp\left[-w_{sr} \left(\frac{\nu}{2} + \frac{\kappa \boldsymbol{\gamma}^T \boldsymbol{\gamma}}{2}\right)\right]}_{\text{Kern einer unbekannten Verteilung}} \quad (25) \\
&\qquad \qquad \qquad \text{Kern einer Gamma-Verteilung} \\
&\Downarrow \\
w_{sr} | \mu \mathbf{1}_{[d]}, \mathbf{x}\beta, \boldsymbol{\gamma}, \kappa, \boldsymbol{\eta} &\sim ?
\end{aligned}$$

### 3.3 Implementierung

Die Aufstellung der vollständig bedingten Dichten hat sowohl bekannte als auch unbekannte Verteilungen hervorgebracht. Aus diesem Grund wird im Folgenden für die komplette Erfassung der vorliegenden Modellsituation ein sogenannter Hybrid-Sampler verwendet. Dessen Zusammensetzung erfordert eine erweiterte Kenntnis über die Möglichkeiten von MCMC Verfahren (siehe Abschnitt 3.3.1). Der Aufbau des implementierten Samplers wird schließlich in Unterkapitel 3.3.3 skizziert und bildet die Grundlage für die späteren Schätzungen der Parametereigenschaften in Verbindung mit den empirischen und den simulierten Daten. Handelt es sich um unbekannte Verteilungen wird im Rahmen des Metropolis-Hastings Algorithmus (Hauptbestandteil des Hybrid-Samplers) eine passende Vorschlagsdichte benötigt. Für die Parameter des Prädiktors einer jeden Region eignet sich diesbezüglich eine Taylorentwicklung als Approximation. Aufgrund der verhältnismäßig aufwendigen Herleitung wird diese in Abschnitt 3.3.2 umfassend dokumentiert. Die optimierte Einbindung einzelner Bestandteile in Bezug auf die Reduzierung der Laufzeit des Hybrid-Samplers wird abschließend in Unterkapitel 3.3.4 behandelt. Dabei stellen das blockweise Ziehen von Zufallszahlen aus einer multivariaten Normalverteilung und die Ausnutzung der geringen Besetzung einer Nachbarschaftsmatrix zwei entscheidende Aspekte dar. Hervorzuheben ist außerdem, dass die Umsetzung der folgenden Ausführungen ausnahmslos über eine Programmierung in R (R Development Core Team, 2013) erfolgt, deren Code dem elektronischen Anhang dieser Arbeit angefügt ist.

### 3.3.1 Markov-Chain-Monte-Carlo (MCMC) Verfahren

Die Grundidee von Markov-Chain-Monte-Carlo (MCMC) Verfahren wurde bereits im Rahmen der Erläuterungen zur Bayes-Inferenz vorgestellt (vgl. Teilabschnitt aus 2.2.2). Zusammengefasst sind MCMC Methoden demnach von zentraler Bedeutung, da sie „das Ziehen von Zufallszahlen aus einer in der Regel analytisch unzugänglichen Posteriori-Verteilung (und prinzipiell auch aus jeder anderen Verteilung) [erlauben]“ (Fahrmeir *et al.*, 2009, S. 482). Im Gegensatz zu anderen Methoden der Ziehung von Zufallszahlen, ist ein großer Vorteil von MCMC Verfahren, dass es keine Einschränkung auf Dichten gibt, die mit einer niedrigen Dimension behaftet sind. Zusätzlich ist es eine erhebliche Erleichterung, dass die Normierungskonstante der jeweiligen Verteilung nicht bekannt sein muss, da deren Berechnung oftmals numerisch nur sehr schwer oder gar nicht durchführbar ist (vgl. Fahrmeir *et al.*, 2009, S. 482). Können aus einer Posteriori-Dichte nicht direkt Zufallszahlen generiert werden, benötigt man die vollständig bedingten Dichten der einzelnen Parameter aller hierarchischen Stufen des Modells, aus welchen dann zufällig gezogen und eine approximative Realisation des jeweiligen Parameters erzeugt wird (siehe Abschnitt 3.2).

Im Folgenden repräsentiert  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)^T$  den gesamten Parameterraum, der sich aus den  $k$  verschiedenen Parametern der hierarchischen Modellstruktur zusammensetzt. Oftmals bietet es sich an, die vollständig bedingten Dichten für ganze Blöcke  $\boldsymbol{\theta}_r$  aus dem Parametervektor anzugeben, in Abhängigkeit von den Daten und den restlichen Parametern. Der Vektor  $\boldsymbol{\theta}$  lässt sich dann zerlegen in die  $R$  Blöcke  $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_R$ . In diesem Zusammenhang wird dann von sogenannten Blockverfahren gesprochen (vgl. Schmid, 2004, S. 18). Für die folgenden Beschreibungen der verschiedenen Algorithmen bezeichnet  $p()$  die Zieldichte (z. B. eine Posteriori-Verteilung oder eine vollständig bedingte Dichte) und  $q()$  die zugehörige Vorschlagsdichte. Zusätzlich sind die beiden Schreibweisen  $p(\theta_i | \cdot)$  und  $p(\theta_i | \boldsymbol{\theta}_{-i})$  als äquivalent zu betrachten. Beide stehen für die Dichte von  $\theta_i$  unter der Berücksichtigung aller restlichen Parameter von  $\boldsymbol{\theta}$  abzüglich  $\theta_i$ . Ferner beschreibt  $\theta^{(t-1)}$  den aktuellen und  $\theta^{(t)}$  den neuen Zustand eines Parameters.

### Metropolis-Hastings-Algorithmus

Der Aufbau des Metropolis-Hastings-Algorithmus kann im Folgenden als Ausgangssituation betrachtet werden, da andere Algorithmen, wie beispielsweise der Gibbs-Sampler, eine abgewandelte bzw. vereinfachte Version davon darstellen. Vorab ist hervorzuheben, dass die Zieldichte für die Berechnung der Akzeptanzwahrscheinlichkeit lediglich als Quotient mit den Auswertungen an den Stellen  $\theta^{(*)}$  und  $\theta^{(t-1)}$  berücksichtigt wird und sich somit sämtliche konstante Faktoren gegenseitig aufheben, wodurch es ausreichend ist den Kern der jeweiligen Verteilung zu kennen. Die Vorschlagsdichte, die ab dem 2. Schritt benötigt wird, hängt von dem aktuellen Parameterzustand  $\theta^{(t-1)}$  ab und kann grundsätzlich beliebig bestimmt werden. Es ist allerdings sinnvoll diese so zu wählen, dass Zufallszahlen einfach generiert werden können. Zudem sollte durch die richtige Wahl sichergestellt werden, dass die Akzeptanzwahrscheinlichkeit  $\alpha$  – welche benötigt wird, da die tatsächliche und die vorgeschlagene Dichte von Grund auf nicht deckungsgleich sind – die richtige Höhe aufweist (vgl. Fahrmeir *et al.*, 2009, S. 483). Die Abhängigkeit der nacheinander gezogenen Zufallsziehungen steht dabei im

Vordergrund. Diese sollte möglichst niedrig sein, da der allgemeine Zusammenhang gilt, dass „je geringer die Abhängigkeit, desto geringer ist der erforderliche Stichprobenumfang an Zufallszahlen zur Schätzung von Charakteristika der [...] Verteilung“ (Fahrmeir *et al.*, 2009, S. 483). Eine zusätzliche Möglichkeit der Vermeidung von zu starker Abhängigkeit liefert eine sogenannte Ausdünnung der gezogenen Zufallszahlen. So wird je nach Ausdünnungsfaktor  $a$  lediglich jede  $a$ -te Zahl gespeichert bei gleichzeitiger Verwerfung der dazwischen liegenden Ziehungen, ohne die eigentliche Schätzung der Parameter entscheidend zu verändern. Unter diesem Aspekt wird eine Ausdünnung der Ziehung oftmals dazu verwendet Speicherplatz einzusparen, der gerade bei Modellen mit einer hohen Anzahl an Parametern und einem großen Iterationsumfang schnell ausgereizt sein kann (vgl. Fahrmeir *et al.*, 2009, S. 485). Da im Zuge des Verfahrens eine Markov-Kette gebildet wird, deren Kern sich mit jedem Iterationsschritt dem der interessierenden Verteilung annähert, ist es außerdem notwendig eine sogenannte „Burn-In“ Phase abzuwarten, bis bei den gezogenen Zahlen von einer tatsächlichen Zufallsziehung aus der jeweiligen Randverteilung gesprochen werden kann (vgl. Fahrmeir *et al.*, 2009, S. 482). Zusammengefasst setzt sich der Algorithmus schließlich folgendermaßen zusammen (vgl. Fahrmeir *et al.*, 2009, S. 482 ff.):

### Metropolis-Hastings-Algorithmus

1. Wähle einen Startwert  $\theta^{(0)}$  sowie die Anzahl der Iterationen  $T$ .  
Setze  $t = 1$ .
2. Ziehe eine Zufallszahl  $\theta^{(*)}$  gemäß der Vorschlagsdichte  $q(\theta^{(*)} | \theta^{(t-1)})$ .
3. Akzeptiere  $\theta^{(*)}$  mit der Wahrscheinlichkeit:

$$\alpha(\theta^{(*)} | \theta^{(t-1)}) = \min \left\{ \frac{p(\theta^{(*)} | \cdot)}{p(\theta^{(t-1)} | \cdot)} \cdot \frac{q(\theta^{(t-1)} | \theta^{(*)}, \cdot)}{q(\theta^{(*)} | \theta^{(t-1)}, \cdot)}, 1 \right\},$$

als neuen Zustand  $\theta^{(t)}$ , ansonsten setze  $\theta^{(t-1)} = \theta^{(t)}$ .

4. Ist  $t = T$  beende den Algorithmus, andernfalls erhöhe  $t$  um 1 und wiederhole ab Schritt 2.

### Gibbs-Sampler und Hybrid-Algorithmen

In der Praxis werden oftmals sogenannte Hybrid-Algorithmen verwendet, meist in Verbindung mit hochdimensionalen Parametervektoren. Deren Grundidee basiert auf der zu Beginn des Abschnitts dargestellten Zerlegung von  $\boldsymbol{\theta}$  in mehrere Parameterblöcke. Bei der Bestimmung der vollständig bedingten Dichten der einzelnen Bestandteile kommt es häufig vor, dass bekannte Verteilungen gebildet werden, aus denen über den Gibbs-Sampler direkt Zufallszahlen gezogen werden können. Die Bezeichnung „Hybrid“

ist schließlich deshalb gerechtfertigt, da für die vollständig bedingten Dichten der Parameter(blöcke), die keiner standardmäßigen Verteilung folgen, der Metropolis-Hastings-Algorithmus aus dem vorherigen Unterkapitel angewendet werden muss und somit ein Algorithmus in Mischform entsteht (vgl. Schmid, 2004, S. 18). Der Gibbs-Sampler ist dabei folgendermaßen aufgebaut, wobei sich diesbezüglich der Parameterraum  $\boldsymbol{\theta}$  auf die Parameter beschränkt, deren vollständig bedingte Dichten einer bekannten Verteilung folgen (vgl. Fahrmeir *et al.*, 2009, S. 487):

### Gibbs-Sampler

1. Wähle Startwerte  $\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_R^{(0)}$  für die einzelnen Blöcke und die Anzahl der Iterationen  $T$ . Setze  $t = 1$ .
2. Für  $r = 1, \dots, R$ :  
Ziehe Zufallszahlen  $\boldsymbol{\theta}_r^{(t)}$  aus der jeweiligen vollständig bedingten Dichte

$$p(\boldsymbol{\theta}_r | \boldsymbol{\theta}_1^{(t)}, \dots, \boldsymbol{\theta}_{r-1}^{(t)}, \boldsymbol{\theta}_{r+1}^{(t-1)}, \dots, \boldsymbol{\theta}_R^{(t-1)}, \mathbf{y})$$

Hervorzuheben ist, dass die Dichte auf die zuletzt upgedateten Zustände der verschiedenen Parameter bedingt ist.

3. Ist  $t = T$  beende den Algorithmus, andernfalls erhöhe  $t$  um 1 und wiederhole ab Schritt 2.

Bei diesem Sampler werden in vielen Iterationen hintereinander zufällige Zahlen aus den vollständig bedingten Dichten des jeweiligen Parameterblocks gezogen. Diese werden mit der Wahrscheinlichkeit 1 als neuer Zustand der generierten Markov-Kette akzeptiert. Analog zu dem Metropolis-Hastings-Algorithmus muss zunächst eine gewisse Burn-In Phase abgewartet werden, bevor die Ziehungen als tatsächliche Realisation der Randdichten betrachtet werden können (vgl. Fahrmeir *et al.*, 2009, S. 487).

#### 3.3.2 Verteilungsapproximation über Taylorentwicklung

Die Ausführung zu MCMC Verfahren hat gezeigt, dass bei der Anwendung von Metropolis-Hastings-Algorithmen neben der Zieldichte eine Vorschlagsdichte benötigt wird, aus der zufällige Zustände des zu schätzenden Parameters gezogen werden können. Die verschiedenen Zieldichten wurden in Unterkapitel 3.2.3 hergeleitet und entsprechen in diesem Fall den vollständig bedingten Dichten der einzelnen Parameter(blöcke) des Modells. Aufgrund der Bildung einer unbekannten Verteilung wird bei der Schätzung der Nachbarschaftsgewichte  $w_{sr}$  und den Prädiktorwerten der einzelnen Regionen  $\eta_s$  jeweils ein Metropolis-Hastings-Schritt benötigt. Nach dem Vorschlag von Brezger *et al.* (2007) bietet sich im Zuge der Schätzung der Einzelgewichte die Verwendung einer

Gamma-Verteilung als Bestandteil der Dichte  $p(w_{sr} | \cdot)$  an (vgl. Dichte (25) auf Seite 64). Für die vollständig bedingte Dichte des Prädiktors für eine Region  $s$  ergibt sich aufgrund der Reparametrisierung durch  $\alpha_s$  folgende komplexere Konstellation (vgl. Herleitung in Gleichung (24)):

$$\begin{aligned} p(\eta_s | \cdot) &\stackrel{\text{bzgl.}}{\underset{\eta_s}{\propto}} p(y_s | \eta_s) \cdot p(\eta_s) \\ &\stackrel{\text{bzgl.}}{\underset{\eta_s}{\propto}} \exp \left\{ \underbrace{y_s \eta_s - e_s \exp(\eta_s) - \frac{\tau}{2}(\eta_s - z_s)^2}_{f(\eta_s) \cong \text{Kern einer unbekannten Verteilung}} \right\} \quad \text{mit} \quad z_s = \mu + x_s \beta + \gamma_s \quad (26) \end{aligned}$$

In der vorliegenden Situation für  $p(\eta_s | \cdot)$  ist es dennoch möglich eine Normalverteilung als Vorschlagsdichte zu wählen (vgl. Schmid, 2004, S. 21 f.). Hierfür wird üblicherweise die Funktion  $f(\eta_s)$  aus Dichte (26) durch eine Taylorentwicklung zweiten Grades um einen geeigneten Wert  $\eta_{s_0}$  approximiert (vgl. Rue und Held, 2005, S. 167):

$$\begin{aligned} f(\eta_s) &\approx f(\eta_{s_0}) + f'(\eta_{s_0})(\eta_s - \eta_{s_0}) + \frac{1}{2}f''(\eta_{s_0})(\eta_s - \eta_{s_0})^2 \\ &= a + b\eta_s - \frac{1}{2}c\eta_s^2 \end{aligned}$$

Als relevante Bestandteile dieser Umformung entspricht dabei  $b = f'(\eta_{s_0}) - f''(\eta_{s_0})\eta_{s_0}$  und  $c = -f''(\eta_{s_0})$ , wohingegen der Wert von  $a$  für die folgenden Betrachtungen unbedeutend bleibt. Der Vorschlag  $q(\eta_s | \cdot)$  zur vollständig bedingten Dichte  $p(\eta_s | \cdot)$  kann somit durch den normalverteilten Kern  $\tilde{q}(\eta_s | \cdot)$  angenähert werden (vgl. Rue und Held, 2005, S. 168):

$$\tilde{q}(\eta_s | \cdot) \stackrel{\text{bzgl.}}{\underset{\eta_s}{\propto}} \exp \left( -\frac{1}{2}c\eta_s^2 + b\eta_s \right) \quad (27)$$

Für den Prädiktor einer Region  $s$  ergeben sich an der Stelle  $\eta_{s_0}$  für die 1. und 2. Ableitung somit:

$$\begin{aligned} f'(\eta_{s_0}) &= \frac{\partial f(\eta_{s_0})}{\partial \eta_{s_0}} \\ &= \frac{\partial}{\partial \eta_{s_0}} \left( y_s \eta_{s_0} - e_s \exp(\eta_{s_0}) - \frac{\tau}{2}(\eta_{s_0} - z_s)^2 \right) \\ &= y_s - e_s \exp(\eta_{s_0}) - \tau(\eta_{s_0} - z_s) \end{aligned}$$

$$\begin{aligned} f''(\eta_{s_0}) &= \frac{\partial^2 f(\eta_{s_0})}{\partial^2 \eta_{s_0}} = \frac{\partial f'(\eta_{s_0})}{\partial \eta_{s_0}} \\ &= \frac{\partial}{\partial \eta_{s_0}} (y_s - e_s \exp(\eta_{s_0}) - \tau(\eta_{s_0} - z_s)) \\ &= -e_s \exp(\eta_{s_0}) - \tau \end{aligned}$$

Daraus können die Faktoren  $b$  und  $c$  mit  $z_s = \mu + x_s\beta + \gamma_s$  gebildet werden mit:

$$\begin{aligned} b &= f'(\eta_{s_0}) - f''(\eta_{s_0}) \eta_{s_0} \\ &= y_s - e_s \exp(\eta_{s_0}) - \tau(\eta_{s_0} - \mu - x_s\beta - \gamma_s) - (-e_s \exp(\eta_{s_0}) - \tau) \eta_{s_0} \\ &= y_s + e_s \exp(\eta_{s_0}) \cdot (\eta_{s_0} - 1) + \tau(\mu + x_s\beta + \gamma_s) \end{aligned}$$

und

$$c = -f''(\eta_{s_0}) = e_s \exp(\eta_{s_0}) + \tau.$$

Die Normalverteilung in Dichte (27) kann mit den Ergänzungen für  $b$  und  $c$  als geeignete Vorschlagsdichte im Rahmen des Metropolis-Hastings-Algorithmus angesehen werden. Diese kanonische Darstellungsweise kann übertragen werden auf die übliche Form der Normalverteilung mit Erwartungswert  $\mu = b/c$  und Varianz  $\sigma^2 = c^{-1}$ . Für die Stelle  $\eta_{s_0}$ , an der  $\tilde{q}(\eta_s | \cdot)$  durch die Taylorentwicklung an die Normalverteilung angenähert wird, stellt in diesem Zusammenhang der aktuelle Zustand nach jedem Iterationsschritt einen geeigneten Vorschlag dar (vgl. Rue und Held, 2005, S. 168).

### 3.3.3 Skizze des Hybrid-Samplers

Die Bestimmung der vollständig bedingten Dichten in Unterkapitel 3.2.3 hat in der Gesamtbetrachtung hervorgebracht, dass sowohl bekannte als auch unbekannte Verteilungen für das MCMC Verfahren berücksichtigt werden müssen. Explizit kann dabei für den Intercept  $\mu$ , den Kovariablen-Koeffizienten  $\beta$  sowie für die beiden Hyperparameter des strukturierten bzw. unstrukturierten Effekts  $\kappa$  und  $\tau$  eine direkte Ziehung neuer Parameterwerte über den Gibbs-Sampler erfolgen. Als Repräsentant des GMZF folgt zudem der gesamte Parameterblock  $\gamma$  einer bekannten multivariaten Normalverteilung, wodurch ebenfalls eine direkte Generierung von Parameterupdates über diesen Sampler durchgeführt werden kann. Wie bereits im vorhergehenden Abschnitt erläutert, stellt sich die Ausgangssituation für die Prädiktorwerte der einzelnen Regionen  $\eta_s$  und für die Gewichte der jeweiligen Nachbarschaften  $w_{sr}$  komplizierter dar, weshalb deren Updates jeweils über Metropolis-Hastings-Schritte erfolgen müssen. Aufgrund dieser Vermischung von Gibbs-Sampler und Metropolis-Hastings-Algorithmus wird die Implementierung eines Hybrid-Samplers unumgänglich.

Im Folgenden dient die Skizzierung des verwendeten Hybrid-Samplers als Grundlage für die anschließenden Unterkapitel. Deren Inhalte können als eine Art Werkzeug für die Umsetzung des angedeuteten Samplers betrachtet werden. Der eigentliche R-Code findet sich im elektronischen Anhang dieser Arbeit unter der Bezeichnung `hybrid.sample` wieder. Dessen Grundzüge sind folgendermaßen aufgebaut:

## Vorbereitung

1. Festlegung des Samplingumfangs  $S$   
→ Entspricht der Anzahl an Iterationen
2. Festlegung des Ausdünnungsfaktor und der Burn-In Phase  
→ Richtet sich in der Regel nach dem Samplingumfang und der Konvergenz

3. Initialisierung der Ergebnisbereiche (Vektoren, Matrizen, Listen):
- Vektoren der Länge  $S + 1$  für die Parameter  $\mu, \beta, \kappa$  und  $\tau$
  - Matrizen mit  $[d \times (S + 1)]$  Einträgen für die Parameterblöcke  $\boldsymbol{\gamma}$ ,  $\boldsymbol{\eta}$  und  $\boldsymbol{\alpha}$ , wobei  $d$  der jeweiligen Anzahl an Regionen entspricht
  - Liste mit  $S + 1$  Matrizen als Listenelemente und jeweils  $[d \times d]$  Matrixeinrägen für die gewichtete Nachbarschaftsmatrix  $\mathbf{Q}_w$
4. Bestimmung der Startwerte/Startvektoren/Startmatrix:
- Startwerte für  $\mu^{(0)}, \beta^{(0)}, \kappa^{(0)}$  und  $\tau^{(0)}$  festgelegt auf 1
  - Startvektoren für  $\boldsymbol{\gamma}^{(0)}$  und  $\boldsymbol{\eta}^{(0)}$ 
    - $\boldsymbol{\gamma}^{(0)}$  als Einheitsvektor  $\mathbf{1}_{[d]}$
    - $\boldsymbol{\eta}^{(0)}$  als  $\log(\mathbf{y})$  mit  $\mathbf{y} \hat{=} \text{Zielgrößenvektor}$
  - Startmatrix für  $\mathbf{Q}_w^{(0)}$ :  
 $\rightarrow$  Ziehe  $w_{sr}$  symmetrisch für alle  $-1$  Einträge der Nachbarschaftsmatrix  $\mathbf{Q}$  aus der Gamma-Verteilung  $G(\nu/2, \nu/2)$  (siehe Annahme (20))

$\Rightarrow$  Anmerkung: Der anschließende Gibbs-Sampler und die beiden Metropolis-Hastings-Schritte werden je nach Samplingumfang  $S$ -mal wiederholt.

### Gibbs-Sampler (vgl. Box auf Seite 67)

5. Betrifft alle vollständig bedingten Dichten mit bekannten Verteilungen  
 $\rightarrow p(\mu | \cdot), p(\beta | \cdot), p(\kappa | \cdot), p(\tau | \cdot)$  und  $p(\boldsymbol{\gamma} | \cdot)$

(a)  $p(\mu | \cdot) : \mu | \cdot \sim \mathcal{N}(\tilde{v}_\mu, \tilde{\sigma}_\mu^2)$

$\Rightarrow$  Ziehe  $\mu^{(t)}$  gegeben  $\beta^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}, \kappa^{(t-1)}, \tau^{(t-1)}, \boldsymbol{\eta}^{(t-1)}, \mathbf{Q}_w^{(t-1)}$  und  $\mathbf{x}$  mit  
 $\tilde{\sigma}_\mu^2 = \left( \sigma_\mu^{-2} + \tau d \right)^{-1}$  und  $\tilde{v}_\mu = \left( v_\mu \sigma_\mu^{-2} - \tau \sum_{s=1}^d (x_s \beta + \gamma_s - \eta_s) \right) \cdot \tilde{\sigma}_\mu^2$ ,  
wobei  $\sigma_\mu^2 = 10\,000$  und  $v_\mu = 0$  fest.

(b)  $p(\beta | \cdot) : \beta | \cdot \sim \mathcal{N}(\tilde{\mu}_\beta, \tilde{\sigma}_\beta^2)$

$\Rightarrow$  Ziehe  $\beta^{(t)}$  gegeben  $\mu^{(t)}, \boldsymbol{\gamma}^{(t-1)}, \kappa^{(t-1)}, \tau^{(t-1)}, \boldsymbol{\eta}^{(t-1)}, \mathbf{Q}_w^{(t-1)}$  und  $\mathbf{x}$  mit  
 $\tilde{\sigma}_\beta^2 = \left( \sigma_\beta^{-2} + \tau \sum_{s=1}^d x_s^2 \right)^{-1}$  und  $\tilde{\mu}_\beta = \left( \mu_\beta \sigma_\beta^{-2} - \tau \sum_{s=1}^d x_s (\gamma_s + \mu - \eta_s) \right) \cdot \tilde{\sigma}_\beta^2$ ,  
wobei  $\sigma_\beta^2 = 10\,000$  und  $\mu_\beta = 0$  fest.

$$(c) \ p(\boldsymbol{\gamma} | \cdot) : \ \boldsymbol{\gamma} | \cdot \sim \mathcal{MVN}_C(\mathbf{b}, \tilde{\mathbf{P}}_{\mathbf{w}}^{-1})$$

$\Rightarrow$  Ziehe Vektor  $\boldsymbol{\gamma}^{(t)}$  gegeben  $\mu^{(t)}, \beta^{(t)}, \kappa^{(t-1)}, \tau^{(t-1)}, \boldsymbol{\eta}^{(t-1)}, \mathbf{Q}_{\mathbf{w}}^{(t-1)}$  und  $\mathbf{x}$  mit  
 $\tilde{\mathbf{P}}_{\mathbf{w}} = \tau \mathbf{1}_{[d \times d]} + \kappa \mathbf{Q}_{\mathbf{w}}$  und  $\mathbf{b} = \tau \cdot (\boldsymbol{\eta} - \mu \mathbf{1}_{[d]} - \mathbf{x} \beta)$ .

$\rightarrow$  Siehe hierfür Teilabschnitt aus Unterkapitel 3.3.4 zu Sampling aus einer multivariaten Normalverteilung.

$$(d) \ p(\kappa | \cdot) : \ \kappa | \cdot \sim G(\tilde{a}, \tilde{b})$$

$\Rightarrow$  Ziehe  $\kappa^{(t)}$  gegeben  $\mu^{(t)}, \beta^{(t)}, \boldsymbol{\gamma}^{(t)}, \tau^{(t-1)}, \boldsymbol{\eta}^{(t-1)}, \mathbf{Q}_{\mathbf{w}}^{(t-1)}$  und  $\mathbf{x}$  mit  
 $\tilde{a} = a + rg(\mathbf{Q}_{\mathbf{w}})/2$  und  $\tilde{b} = b + \frac{1}{2} \boldsymbol{\gamma}^T \mathbf{Q}_{\mathbf{w}} \boldsymbol{\gamma}$ , wobei  $a = 1$  und  $b = 10^{-3}$  fest.

$$(e) \ p(\tau | \cdot) : \ \tau | \cdot \sim G(\tilde{g}, \tilde{h})$$

$\Rightarrow$  Ziehe  $\tau^{(t)}$  gegeben  $\mu^{(t)}, \beta^{(t)}, \boldsymbol{\gamma}^{(t)}, \kappa^{(t)}, \boldsymbol{\eta}^{(t-1)}, \mathbf{Q}_{\mathbf{w}}^{(t-1)}$  und  $\mathbf{x}$  mit

$\tilde{g} = g + \frac{1}{2}$  und  $\tilde{h} = h + \frac{1}{2} \sum_{s=1}^d (\eta_s - \mu - x_s \beta - \gamma_s)^2$ , wobei  $g = 1$   
und  $h = 10^{-5}$  fest.

### Metropolis-Hastings-Algorithmus (vgl. Box auf Seite 66)

6. Betrifft alle vollständig bedingten Dichten mit unbekannten Verteilungen  
 $\rightarrow p(\eta_s | \cdot)$  und  $p(w_{sr} | \cdot)$

$$(a) \text{ Zieldichte: } p(\eta_s | \cdot) \stackrel{\eta_s}{\propto} \exp \left[ y_s \eta_s - e_s \exp(\eta_s) - \frac{\tau}{2} (\eta_s - \mu - x_s \beta - \gamma_s)^2 \right]$$

$\Rightarrow$  Ziehe  $\eta_s^{(t)}$  gegeben  $\mu^{(t)}, \beta^{(t)}, \boldsymbol{\gamma}^{(t)}, \kappa^{(t)}, \tau^{(t)}, \mathbf{Q}_{\mathbf{w}}^{(t-1)}$  und  $\mathbf{x}$ .

Vorschlagsdichte:  $\tilde{q}(\eta_s | \cdot) \stackrel{\eta_s}{\propto} \exp \left( -\frac{1}{2} c \eta_s^2 + b \eta_s \right)$ , d. h.  $\eta_s | \cdot \sim \mathcal{N}\left(\frac{b}{c}, c^{-1}\right)$

mit  $b = y_s + e_s \exp(\eta_{s_0}) \cdot (\eta_{s_0} - 1) + \tau(\mu + x_s \beta + \gamma_s)$  und  $c = e_s \exp(\eta_{s_0}) + \tau$ .

$\rightarrow$  Verteilung wird angenähert um den Zustand  $\eta_{s_0}$  (vgl. Abschnitt 3.3.2).

Folgende Schritte werden für jede der  $d$  Region wiederholt:

- i. Ziehe Zufallszahl für  $\eta_s^{(*)}$  gemäß der Vorschlagsdichte  $\tilde{q}(\eta_s^{(*)} | \eta_s^{(t-1)})$ .

ii. Berechne Akzeptanzwahrscheinlichkeit:

→ Werte Zieldichte  $p(\eta_s | \cdot)$  an den Stellen  $\eta_s^{(*)}$  und  $\eta_s^{(t-1)}$  aus.

→ Werte Vorschlagsdichte  $\tilde{q}(\eta_s | \cdot)$  an der Stelle  $\eta_s^{(*)}$  mit  $\eta_{s_0} = \eta_s^{(t-1)}$  und an der Stelle  $\eta_s^{(t-1)}$  mit  $\eta_{s_0} = \eta_s^{(*)}$  aus.

$$\Rightarrow \text{Berechne } \alpha(\eta_s^{(*)} | \eta_s^{(t-1)}) = \min \left\{ \frac{p(\eta_s^{(*)} | \cdot)}{p(\eta_s^{(t-1)} | \cdot)} \cdot \frac{\tilde{q}(\eta_s^{(t-1)} | \eta_s^{(*)}, \cdot)}{\tilde{q}(\eta_s^{(*)} | \eta_s^{(t-1)}, \cdot)}, 1 \right\}$$

iii. Akzeptiere  $\eta_s^{(*)}$  als neuen Zustand  $\eta_s^{(t)}$  mit der Wahrscheinlichkeit  $\alpha$ , andernfalls setze  $\eta_s^{(t-1)} = \eta_s^{(t)}$ .

$$(b) \text{ Ziieldichte: } p(w_{sr} | \cdot) \stackrel{\text{bzgl. }}{\propto} \left( \prod_{s=2}^d \lambda_s \right)^{1/2} \cdot w_{sr}^{(\nu/2-1)} \cdot \exp \left[ -w_{sr} \left( \frac{\nu}{2} + \frac{\kappa \boldsymbol{\gamma}^T \boldsymbol{\gamma}}{2} \right) \right]$$

$$\Rightarrow \text{Ziehe } w_{sr}^{(t)} \text{ und gegeben } \mu^{(t)}, \beta^{(t)}, \boldsymbol{\gamma}^{(t)}, \kappa^{(t)}, \tau^{(t)}, \boldsymbol{\eta}^{(t)} \text{ und } \mathbf{x} \text{ mit } \nu = 1.$$

$$\text{Vorschlagsdichte: } q(w_{sr} | \cdot) \propto w_{sr}^{(\nu/2-1)} \cdot \exp \left[ -w_{sr} \left( \frac{\nu}{2} + \frac{\kappa \boldsymbol{\gamma}^T \boldsymbol{\gamma}}{2} \right) \right]$$

Verwende  $w_{sr} | \cdot \sim G \left( \frac{\nu}{2}, \frac{\nu}{2} + \frac{\kappa \boldsymbol{\gamma}^T \boldsymbol{\gamma}}{2} \right)$  als Bestandteil der Ziieldichte.

Folgende Schritte werden für jede der Nachbarschaften wiederholt:

i. Ziehe Zufallszahl für  $w_{sr}^{(*)}$  gemäß der Vorschlagsdichte  $q(w_{sr}^{(*)} | w_{sr}^{(t-1)})$ .

ii. Berechne Akzeptanzwahrscheinlichkeit:

$$\begin{aligned} \alpha(w_{sr}^{(*)} | w_{sr}^{(t-1)}) &= \min \left\{ \left( \prod_{s=2}^d \lambda_s^{(*)} / \prod_{s=2}^d \lambda_s^{(t-1)} \right)^{\frac{1}{2}}, 1 \right\} \\ &= \min \left\{ |\mathbf{Q}_{w_{11}}^{(*)}|^{\frac{1}{2}} / |\mathbf{Q}_{w_{11}}^{(t-1)}|^{\frac{1}{2}}, 1 \right\} \end{aligned}$$

→  $\mathbf{Q}_{w_{11}}$  repräsentiert die jeweilige gewichtete Nachbarschaftsmatrix ohne der letzten Zeile und der letzten Spalte.

iii. Akzeptiere  $w_{sr}^{(*)}$  als neuen Zustand  $w_{sr}^{(t)}$  mit der Wahrscheinlichkeit  $\alpha$  und bestimme  $\mathbf{Q}_{\mathbf{w}}^{(*)} = \mathbf{Q}_{\mathbf{w}}^{(t)}$ , andernfalls setze  $\mathbf{Q}_{\mathbf{w}}^{(t-1)} = \mathbf{Q}_{\mathbf{w}}^{(t)}$ .

## Nachbereitung

7. Die Zustände werden je nach Ausdünnung und Burn-In Phase abgespeichert.  
→ Wichtig: bei einem großen  $S$  muss auf die Speicherkapazität geachtet werden.

### 3.3.4 Optimierung von Rechenoperationen

Da eine Vielzahl an Iterationen des Hybrid-Samplers notwendig sind, um repräsentative Ziehungen aus den Randverteilungen der einzelnen Parameter zu erhalten, ist eine effiziente Programmierung ein zentraler Aspekt bei der Umsetzung des vorgestellten Modells. Bereits kleinste Zeiteinsparungen haben durch die häufige Wiederholung des Samplers eine deutliche Verringerung der Gesamtlaufzeit zur Folge. Vor diesem Hintergrund sollten deshalb die verwendeten Bestandteile möglichst Effektiv in den Sampler eingebunden werden.

#### Sampling aus einer multivariaten Normalverteilung

Bei der Aufstellung der vollständig bedingten Dichte des GMZF wurde eine multivariate Normalverteilung in kanonischer Form gebildet. Für die blockweise Generierung neuer Zustände der strukturierten Effekte aus dieser Verteilung stellt der sogenannte Rue-Block-Algorithmus eine effiziente Vorgehensweise dar, welche sich wie folgt zusammensetzt (vgl. Schmid, 2004, S. 19 f.):

#### Rue-Block-Algorithmus

1. Berechne die Cholesky-Zerlegung  $\mathbf{P}_w = \mathbf{L}\mathbf{L}^T$ .
2. Löse  $\mathbf{L}\mathbf{u} = \mathbf{b}$  nach  $\mathbf{u}$  auf mit  $\mathbf{u} = \mathbf{b}\mathbf{L}^T$ .
3. Löse  $\mathbf{L}^T\mathbf{v} = \mathbf{u}$  nach  $\mathbf{v}$  auf mit  $\mathbf{v} = \mathbf{u}\mathbf{L}$ .
4. Ziehe  $\mathbf{z}$  als Zufallsvektor mit Länge  $d \Rightarrow \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{1}_{[d]})$ .
5. Löse  $\mathbf{L}^T\mathbf{m} = \mathbf{z}$  nach  $\mathbf{m}$  auf mit  $\mathbf{m} = \mathbf{z}\mathbf{L}$ .
6. Berechne  $\boldsymbol{\gamma} = \mathbf{v} + \mathbf{m}$ .

Rue präsentiert diese Methode in seinem Paper über das schnelle Sampling aus GMZF (vgl. Rue, 2001, S. 328 ff.). In diesem Zusammenhang ist anzumerken, dass es sich hierbei um einen an die kanonische Form der multivariaten Normalverteilung angepassten Algorithmus handelt. Im Kontext der GMZF entspricht diese typischerweise:

$$p(\boldsymbol{\gamma} | \cdot) \propto \exp\left(-\frac{1}{2}\boldsymbol{\gamma}^T \mathbf{P}_w \boldsymbol{\gamma} + \mathbf{b}^T \boldsymbol{\gamma}\right)$$

Bei der Betrachtung des schrittweisen Ablaufs ist zu erkennen, dass die Cholesky-Zerlegung der (gewichteten) Präzisionsmatrix die Grundlage für diesen Algorithmus bildet. Im Allgemeinen nimmt dieser Schritt mit zunehmender Dimensionsgröße der

betreffenden Matrix verhältnismäßig viel Zeit in Anspruch. Nutzt man jedoch die geringe Besetzung der Nachbarschaftsmatrix aus, lässt sich zusätzlich eine Zeitersparnis erwirken, zumal die Cholesky-Zerlegung ebenso für das Update der Nachbarschaftsgewichte eine entscheidende Rolle spielt. Zusätzlich gilt es an dieser Stelle festzuhalten, dass die Nulleinträge von Präzisions- und Nachbarschaftsmatrix für ein bestimmtes Nachbarschaftssystem zu jeder Zeit identisch sind.

## Reduzierung der Bandweite von Matrizen

Um die knappe Besetzung tatsächlich ausnutzen zu können, ist eine Reduzierung der jeweiligen Bandweite nötig. Für eine Nachbarschaftsmatrix ist diese definiert als das Maximum aller absoluten Abstände der Indizes zweier benachbarter Regionen, d. h.  $b(\mathbf{Q}) = \max_{s \sim r} |s - r|$  (vgl. Rue, 2001, S. 328). Hierfür ist zu beachten, dass ein Eintrag  $q_{sr}$  Null ist, wenn zwei Beobachtungseinheiten nicht benachbart sind ( $q_{sr} = 0$ , wenn  $s \not\sim r$  und  $s \neq r$ ). Cuthill und McKee (1969) entwickelten diesbezüglich einen Algorithmus, der verhältnismäßig einfach ist und sehr gute Ergebnisse erzielt. In der Literatur sind oftmals lediglich oberflächliche Darstellungen für dessen genauen Ablauf zu finden (siehe beispielsweise Quarteroni *et al.* (2007)). Zavoianu (2009) stellt jedoch auf seiner Internetseite sowohl eine ausführliche Beschreibung des umgekehrten Cuthill-McKee-Algorithmus als auch eine praktische Umsetzung in Form eines Delphi-Programms zur Verfügung<sup>4</sup>. Bei der „umgekehrten“ Form handelt es sich um eine Anpassung nach Alan George, bei der lediglich die Indexreihenfolge der reduzierten Matrix nach Beendigung des eigentlichen Cuthill-McKee-Algorithmus umgedreht wird, wodurch Zeitersparnisse bei dem Lösen von linearen Gleichungssystemen erzielt werden können.

Die neue Position eines Knotens in dem Ergebnisvektor entspricht dabei dem neuen Index der bandweitenreduzierenden Reihenfolge. Die jeweilige Zahl repräsentiert dagegen den Index der ursprünglichen Matrix. Zusätzlich gilt es zu beachten, dass in manchen Szenarien die Möglichkeit besteht, dass ein Graph nicht vollständig verbunden ist, d. h. ein(e) oder mehrere Knoten(gruppen) sind weder direkt noch indirekt miteinander verbunden. In einer solchen Situation bleiben einzelne Knoten(gruppen) nach einem kompletten Durchlauf des Algorithmus übrig, da sie nicht zu dem Ergebnisvektor zugeordnet wurden. Ist das der Fall, muss der Algorithmus für diese Knoten ab dem 1. Schritt solange wiederholt werden, bis alle Knoten dem Ergebnisvektor zugeordnet worden sind. Für die relevanten Nachbarschaftssituationen von Schottland und Deutschland tritt diese Konstellation jedoch nicht auf.

Für die folgende Aufbereitung ist der Grad einer Region von zentraler Bedeutung. Dieser ist definiert als deren jeweilige Anzahl an benachbarten Beobachtungseinheiten. Ausgehend von einem Nachbarschaftssystem, welches durch den Graphen  $\mathcal{G}$  mit seinen Komponenten  $\mathcal{V}$  (alle Regionen bzw. Knoten eines Graphens) und  $\mathcal{E}$  (alle zugehörigen Nachbarschaften bzw. Verbindungen) repräsentiert wird, setzt sich der umgekehrte Cuthill-McKee-Algorithmus durch folgende Schritte zusammen (Zavoianu, 2009):

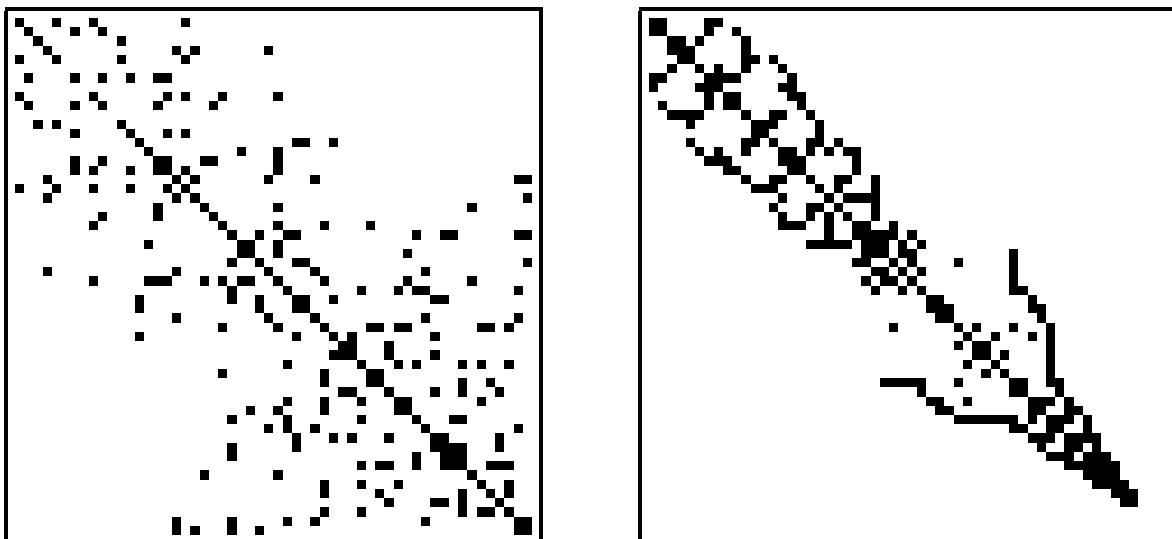
---

<sup>4</sup>Download von <https://app.box.com/shared/o9m4qh19b9> am 12.08.2013. Das Programm befindet sich außerdem im elektronischen Anhang.

### Umgekehrter Cuthill-McKee-Algorithmus (UCMA)

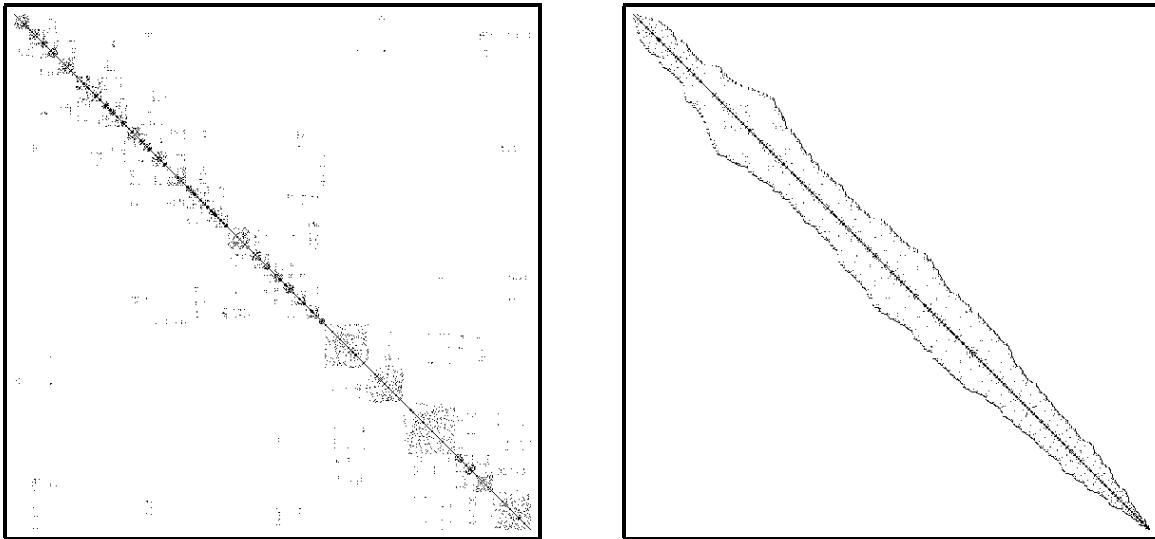
1. Erstelle einen leeren Zwischenspeicher  $Z$  und einen leeren Ergebnisvektor  $E$ .
2. Wähle aus der Menge  $\mathcal{V}$  den Knoten mit dem niedrigsten Grad.  
→ Insofern dieser noch kein Bestandteil von  $E$  ist.
3. Ordne diesen Knoten der ersten freien Position von  $E$  zu.
4. Füge dessen benachbarte Knoten zum Zwischenspeicher hinzu.  
→ Ordne alle Knoten in  $Z$  nach aufsteigendem Grad.
5. Betrachte den ersten Knoten von  $Z$ : Ist er noch kein Bestandteil des Ergebnisvektors, dann ...
  - (a) ... hänge ihn am Ende von  $E$  an und
  - (b) füge alle Nachbarn von diesem Knoten, die nicht in  $E$  enthalten sind, zu  $Z$  hinzu und ordne alle Knoten in  $Z$  aufsteigend nach deren Grad.
6. Entferne diesen Knoten in jedem Fall aus  $Z$ .
7. Ist  $Z$  danach nicht leer, dann wiederhole ab Schritt 5.
8. Umkehrung der Elementenreihenfolge in  $E$ .  
→ Element  $E_i$  wird getauscht mit Element  $E_{g+1-i}$ .  
→  $g$  entspricht der Länge von  $\mathcal{E}$  und  $i$  ist Laufindex von 1 bis  $g$ .

Durch die Anwendung des UCMA kann die Bandweite der Nachbarschaftsmatrix der 56 schottischen Distrikte von ursprüngliche 38 auf 14 reduziert werden.



**Abbildung 24:** Einträge der Nachbarschaftsmatrix der 56 schottischen Distrikte:  
ursprünglich (links) und reduziert mit UCMA (rechts)

Jeder schwarz markierte Pixel repräsentiert dabei einen Eintrag in der Matrix. Analog dazu kann die Bandweite der 544 deutschen Landkreise von 522 auf 59 mit der folgenden Neuordnung der Landkreisindizes reduziert werden:



**Abbildung 25:** Einträge der Nachbarschaftsmatrix der 544 deutschen Landkreise:  
ursprünglich (links) und reduziert mit UCMA (rechts)

Um die Vorteile dieser schmäleren Form voll ausschöpfen zu können, empfiehlt es sich zudem eine Umformung der jeweiligen Matrix mit der Funktion `sparseMatrix` aus dem R-Paket `Matrix` vorzunehmen (siehe Bates und Maechler, 2013). Es gilt dabei jedoch hervorzuheben, dass die Verwendung dieser Funktion in Verbindung mit dem in Unterkapitel 3.3.3 skizzierten Hybrid-Sampler lediglich bei der höherdimensionalen, deutschen Nachbarschaftsmatrix zu einer Zeitersparnis führt.

### Determinantenberechnung über Cholesky-Zerlegung

Der Sampler hat außerdem gezeigt, dass bei der Berechnung der Akzeptanzwahrscheinlichkeit für das Update einzelner Nachbarschaftsgewichte die Determinante der jeweilige Matrix benötigt wird. In diesem Zusammenhang kann eine erhebliche Zeiteinsparung erreicht werden, indem die Determinante über die jeweilige Cholesky-Zerlegung berechnet wird, da „die Determinante einer Matrix in Dreiecksform [...] das Produkt der Diagonalelemente [ist]“ (Fahrmeir *et al.*, 2009, S. 452). Dabei entsprechen die Faktoren der Cholesky-Zerlegung eben dieser Dreiecksform mit  $\mathbf{Q}_w = \mathbf{L}\mathbf{L}^T$ , bei der für den Cholesky-Faktor  $\mathbf{L}$  alle Elemente oberhalb bzw. unterhalb der Hauptdiagonale gleich Null sind. Demzufolge kann die Determinante von einer symmetrischen und positiv (semi-)definiten Nachbarschaftsmatrix  $\mathbf{Q}_w$  folgendermaßen berechnet werden:

$$|\mathbf{Q}_w| = \prod_{s=1}^d |\mathbf{L}_{ss}|^2$$

Um eine doppelte Berechnung zu vermeiden, sollte die jeweilige Determinante des aktuellen Zustands der gewichteten Nachbarschaftsmatrix zwischengespeichert werden.

## Neubildung von gewichteten Nachbarschaftsmatrizen

Bei der Verwendung des Hybrid-Samplers in seiner adaptiven Ausführung ist es für die Updates der einzelnen Nachbarschaftsgewichte erforderlich bei jedem Metropolis-Hastings-Schritt eine eigene gewichtete Vorschlagsmatrix zu generieren mit einer zufällig gezogenen Veränderung des betrachteten Gewichts (siehe Hybrid-Sampler Skizze im vorherigen Unterkapitel). Bei diesem Vorgang kann unter der Verwendung der Funktion `sparseMatrix` eine enorme zusätzliche Zeitersparnis erzielt werden. Für die Konstruktion einer solchen „sparse Matrix“ wird jeweils ein Vektor mit den Zeilen- und Spaltenindizes der Nicht-Null-Einträge der betrachteten Matrix sowie ein Vektor mit den zugehörigen Werten benötigt. Die Änderung an der betreffenden Stelle in diesem Vektor und eine zusätzliche Anpassung der Diagonaleinträge ist, im Zuge einer kompletten Neuerstellung der gewichteten Nachbarschaftsmatrix, wesentlich schneller, als die Neubestimmung dieser Einträge in einer bereits bestehenden Matrix. Hierbei ist wiederum zu berücksichtigen, dass bei kleineren Nachbarschaftssystemen, wie beispielsweise dem schottischen, die Verwendung von `sparseMatrix` in Verbindung mit der beschriebenen, alternativen Berechnungsmethode zu keiner Zeitoptimierung führt. Bei der Anwendung auf die deutschen Landkreise ist sie dahingegen besonders effektiv.

## 3.4 Anwendung auf empirische Daten ohne Gewichtsadaption

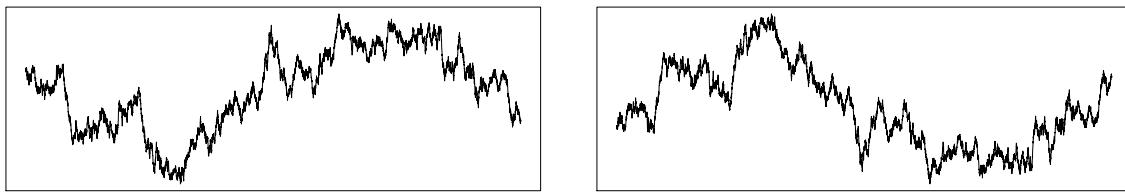
Für das anschließende Unterkapitel stehen zunächst die Schätzergebnisse der schottischen Lippenkrebs- und der deutschen Kehlkopfkrebsdaten ohne der Berücksichtigung von Gewichtsadaptionen im Fokus. Demzufolge ist die Zusammensetzung der Nachbarschaftsmatrix  $\mathbf{Q}_w$  für deren Modellierung unveränderlich und mit einheitlichen Gewichtungen der jeweiligen Nachbarschaften versehen, wie unter anderem in Abschnitt 2.4.3 beschrieben. Bevor jedoch auf die Aufbereitung der Schätzungen eingegangen wird, sollen zunächst einige Modellanpassungen erläutert werden, die für ein stabiles Sampling der vorliegenden Problemstellung notwendig sind. Der darauf folgende Einschub zu den grafischen Möglichkeiten der Samplingbeurteilung bildet zudem die Grundlage für eine umfassende Bewertung der Schätzergebnisse. Die Einbeziehung flexibler Gewichtungen unter der Anwendung auf simulierte sowie auf die bekannten empirischen Daten ist schließlich Bestandteil der Unterkapitel 3.5 und 3.6.

### 3.4.1 Modellanpassungen

In den Abschnitten 3.2 und 3.3 wurde, neben der Umsetzung der hierarchischen Bayes-Struktur des Modells, die Implementierung des verwendeten Hybrid-Samplers unter der Annahme aufbereitet, dass bis einschließlich der hierarchischen Stufe der Hyperparameter  $\kappa$  und  $\tau$  Updates im Zuge des Samplers durchgeführt werden. Sowohl die Anwendung auf empirische als auch auf simulierte Daten wird jedoch zeigen, dass die Festsetzung von mindestens einem der beiden Hyperparameter notwendig ist, um stabile Samplingdurchläufe zu gewährleisten. Die Begründung liegt diesbezüglich in der hohen Anzahl an Parametern, die aus einem vergleichsweise geringen Beobachtungsumfang geschätzt werden müssen. Für die folgende Betrachtung der empirischen Daten wird daher der Hyperparameter  $\tau$  der unstrukturierten Effekte auf 1 festgelegt, um ein sinnvolles Sampling der restlichen Parameter zu erhalten. Grundsätzlich

wäre die Wahl einer größeren Präzision und somit einer niedrigeren Varianz der unstrukturierten Effekte durchaus wünschenswert, um im Gegenzug die strukturierten Effekte zu verstärken. Bereits die Festsetzung auf  $\tau = 10$  hat jedoch zur Folge, dass sämtliche Parameter der zweiten hierarchischen Stufe ( $\gamma$  und  $\alpha$ ) extrem klein werden, wodurch keine sinnvollen Schätzungen der jeweiligen Verteilungscharakteristika mehr möglich sind. Um die strukturierten und die unstrukturierten Effekte in ein ausgewogenes Verhältnis zu bringen, sollte zudem die Priori-Verteilung des Hyperparameters  $\kappa$  informativer in die Modellierung mit einfließen. Dies kann erreicht werden, indem die Rate  $b$  der zugehörigen Gamma-Verteilung ebenfalls auf 1 gesetzt wird, anstatt der üblichen Wahl aus einem Bereich der Größenordnung  $10^{-3}$  (vgl. Unterkapitel 3.2.1 zu den ursprünglichen Verteilungsannahmen aller hierarchischen Stufen).

Als nicht praktikabel hat sich außerdem das direkte Sampling des Intercepts  $\mu$  über eine gesonderte Aufnahme in den Modellaufbau gezeigt. Bereits erste Testdurchläufe des Hybrid-Samplers haben in diesem Zusammenhang verdeutlicht, dass der Intercept aufgrund einer Identifizierbarkeitsproblematik vollends in die strukturierten und unstrukturierten Effekte eingeht, wodurch vollkommen gegenläufige Samplingpfade entstehen können (vgl. Grafiken in Abbildung (26)). Eine Schätzung des Intercepts über die Zentrierung der strukturierten und unstrukturierten Effekte im Nachgang des Samplings, erweist sich deshalb als sinnvollere Herangehensweise<sup>5</sup>.

(a) Samplingpfad  $\mu$ (b) Samplingpfad  $\gamma_s$ 

**Abbildung 26:** Gegenüberstellung beispielhafter Samplingpfade: Intercept  $\mu$  (a) vs. strukturierter räumlicher Effekt  $\gamma_s$  (b) aus dem gleichen MCMC Durchlauf

### 3.4.2 Grundlagen der grafischen Samplingbewertung

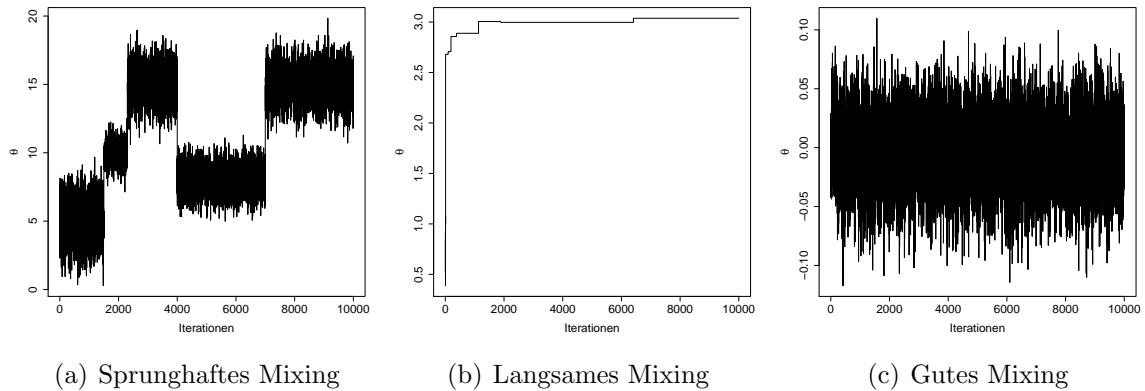
Als Kernannahme bei der Verwendung eines MCMC Verfahrens, können die generierten Parameterwerte nach der Burn-In Phase als zufällige Ziehungen aus der tatsächlichen Randverteilung des jeweiligen Parameters betrachtet werden, wodurch es ermöglicht wird, die zugehörigen Verteilungscharakteristika zu schätzen. Problematisch wird es immer dann, sollte der gewählte Iterationsumfang nicht ausreichend groß sein, um eine Konvergenz der generierten Markov-Ketten zu gewährleisten und unabhängige Zustände zu erhalten. In diesem Zusammenhang gibt es verschiedenste Diagnosewerkzeuge, durch welche entweder grafisch oder analytisch die Konvergenz sowie die Autokorrelation eines Samplingpfades überprüft werden können. Da die individuelle Schätzung von Nachbarschaftsgewichten im Fokus dieser Arbeit stehen, soll für die Betrachtungen der gesampelten Parameterzustände aus den empirischen Daten, ohne einer

<sup>5</sup>Über die Funktionseinstellungen des implementierten Hybrid-Samplers können einzelne Modellbestandteile für das Sampling ein- bzw. ausgeschlossen werden.

Gewichtsadaption, eine auf Grafiken basierende Samplingbewertung ausreichen. Bei der späteren Anwendung, inklusive anpassungsfähiger Gewichte, wird zusätzlich eine analytische Konvergenzanalyse nach Gelman und Rubin durchgeführt, unter besonderer Berücksichtigung der Gewichtungsparameter (siehe Teilabschnitt 3.6.1).

### Samplingbewertung über das Mixing

Das Mixing der Zustände eines Parameters kann unter anderem als Indikator für die Konvergenz einer Markov-Kette verwendet werden, für dessen Bewertung der zugehörige Samplingpfad betrachtet werden muss. In diesem Zusammenhang wird ein Mixing immer dann als schlecht eingestuft, wenn in der Kette Sprünge zu erkennen sind oder sich die Zustände nur sehr langsam verändern. Im Gegensatz dazu sind bei einem gut gemischten Sampling die zufälligen Ziehungen gleichmäßig auf den Parameterraum verteilt (vgl. Lam, 2008, S. 10). Anschließende Abbildung illustriert sowohl Szenarien einer schlechten Mischung (siehe Grafiken (a) und (b)) als auch das einer guten Mischung (siehe Grafik (c)):



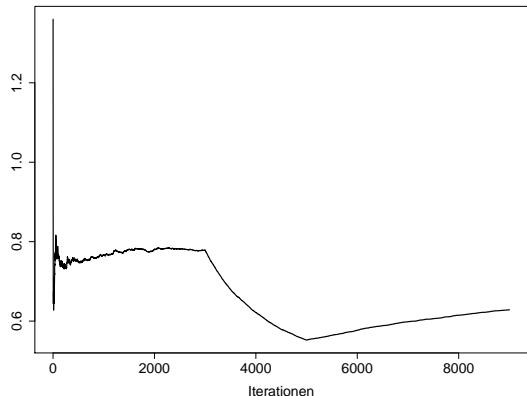
**Abbildung 27:** Beispielhafte Samplingpfade beliebiger Parameter  $\theta$  mit schlechtem (a) und (b) sowie gutem (c) Mixing (Grafik (a) nachempfunden: Lam, 2008, S. 11)

Als zusätzliche Visualisierung ist es zudem sinnvoll sich über einen Kerndichteschätzer die Verteilung der Zustände ausgeben zu lassen. Die Auswirkungen einer schlechten Mischung auf dessen Aussagekraft und somit auf die Qualität der geschätzten Verteilungscharakteristika kann folgendermaßen beschrieben werden:

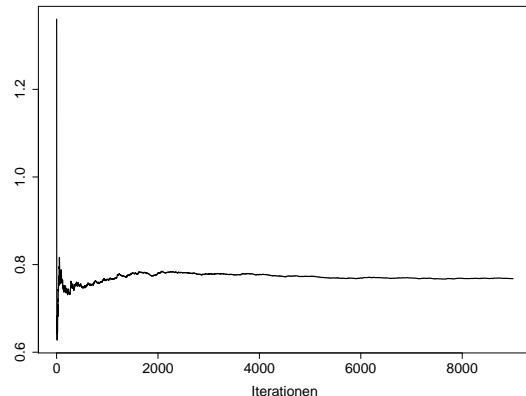
„Poor mixing invalidates the kernel density estimate, as it implies that the MCMC output is not a representative sample from the posterior distribution“ (Plummer *et al.*, 2006, S. 9).

Darüber hinaus ist die Entwicklung des Mittelwerts über einen Samplingverlauf hinweg ein sinnvolles grafisches Diagnosewerkzeug für die Beurteilung der Mischung und somit

der Konvergenz. Für die folgende beispielhafte Visualisierung wird für ein schlechtes und ein gutes Mixing der Mittelwert bis hin zu der jeweiligen Iteration berechnet und abgetragen.



(a) Mittelwertsentwicklung (schlechtes Mixing)



(b) Mittelwertsentwicklung (gutes Mixing)

**Abbildung 28:** Vergleich der Mittelwertsentwicklung bei schlechtem (a) und gutem (b) Mixing

Ein schlechtes und insbesondere ein langsames Mixing kann außerdem ein Hinweis darauf sein, dass aufeinanderfolgende Zustände stark miteinander korrelieren. Für die explizite Betrachtung der Autokorrelation innerhalb eines Samplings gibt es weitere grafische Analysemöglichkeiten.

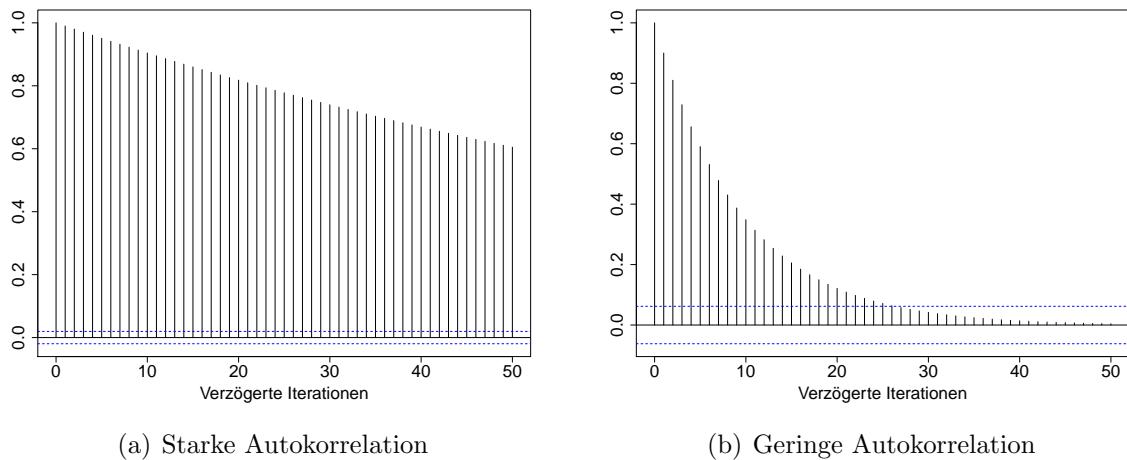
### Samplingbewertung über die Autokorrelation

Für die Berechnung der Autokorrelation zwischen den Parameterzuständen innerhalb eines Samplings kann folgende Funktion verwendet werden (vgl. Lam, 2008, S. 14):

$$\rho_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (28)$$

In diesem Zusammenhang wird eine sogenannte Verzögerung  $k$  (engl. lag) für die Berechnung der Autokorrelation  $\rho_k$  festgelegt. Dadurch repräsentiert  $\rho_k$  die Korrelation zwischen allen Zuständen und denen in einem Abstand von  $k$  Iterationen, wobei  $n$  die jeweilige verzögerte Stelle der Markov-Kette darstellt. Für ein konvergierendes Sampling nimmt man dabei an, dass die Zustände bei einer größer werdenden Verzögerung ( $k$  wird größer) weniger korreliert sind (vgl. Lam, 2008, S. 14). Des Weiteren ist bei Konvergenz zu erwarten, dass die Autokorrelationsfunktion relativ zügig sehr klein wird, d. h. umso schneller sich die Autokorrelation einer Markov-Kette mit fester Länge in einem Bereich um Null bewegt, desto besser können die Verteilungscharakteristika der zugehörigen Randverteilung eines Parameters geschätzt werden. Zusätzlich kann die Betrachtung

der Funktion (28) einen Anhaltspunkt für die Festlegung des Ausdünnungsfaktors geben (Fahrmeir *et al.*, 2009, S. 484 f.). Liegt eine starke Abhängigkeit der Zustände vor (siehe Grafik (a) in Abbildung 29) könnte dieser möglicherweise angepasst werden. Allgemein ist eine Entwicklung der Autokorrelation, wie sie in Grafik (b) abgebildet wird zielführend, um repräsentative Ziehungen zu erhalten.



**Abbildung 29:** Vergleich beispielhafter Autokorrelationsfunktionen mit starker (a) und geringer (b) Abhängigkeit von aufeinanderfolgenden Ziehungen

Autokorrelationswerte zwischen den blau gestrichelten Linien sind, im Gegensatz zu den Werten außerhalb, als nicht signifikant von Null verschieden einzustufen. Befindet sich die Autokorrelationsfunktion in diesem Intervall, kann demzufolge von unabhängigen Ziehungen gesprochen werden. Allgemein ist zu der Samplingbewertung außerdem anzumerken, dass die genannten grafischen Diagnosemethoden für jeden einzelnen Parameter des Modells betrachtet werden müssen, um ein MCMC Verfahren umfassend beurteilen zu können.

### 3.4.3 Schätzergebnisse aus den schottischen Lippenkrebsdaten

Wie bereits erwähnt, können nach der bayesianischen Interpretationsauffassung die generierten Parameterwerte des Hybrid-Samplers nach einer Burn-In Phase als zufällige Ziehungen aus der jeweiligen Randverteilung angesehen werden. Aus dieser Betrachtungsweise resultiert zudem, dass die Parameter, anders als beispielsweise bei der Likelihood-Inferenz, nicht direkt geschätzt werden können, sondern sich lediglich Schätzungen für die zugehörigen Verteilungscharakteristika ergeben, wie etwa der Posteriori-Mittelwert oder der Posteriori-Median. Angewendet auf die vorliegenden räumlichen Problemstellungen hat das zur Folge, dass für jede einzelne Beobachtungseinheit des Modells ein eigener Mittelwert oder Median aus dem jeweiligen Samplingpfad sowohl für den strukturierten als auch für den unstrukturierten Effekt sowie für alle weiteren Parameter gebildet werden muss. In diesem Zusammenhang soll noch einmal wiederholt werden, dass sich die unstrukturierten Effekte aus der Verrechnung von  $\eta$ ,  $\mathbf{x}\beta$  und  $\gamma$  ergeben (vgl. hierfür Unterkapitel 3.2.2) und sich somit indirekt aus den Samplings dieser Bestandteile bilden.

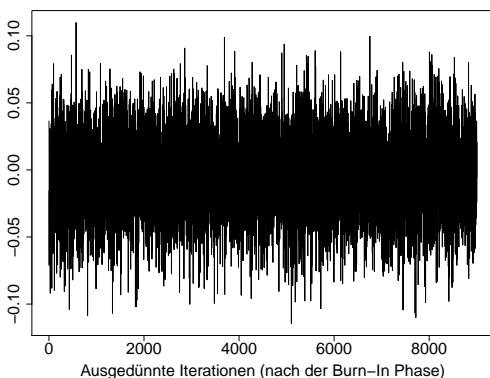
Wie im Rahmen der Modellanpassungen in Teilabschnitt 3.4.1 beschrieben soll die Schätzung des Intercepts  $\hat{\mu}$  nach dem eigentlichen Sampling erfolgen. Unter diesem Aspekt ist es deshalb sinnvoll die mittleren, zentrierten Effekte  $\hat{\gamma}_{\text{zent}}$  und  $\hat{\alpha}_{\text{zent}}$  zu betrachten, die sich folgendermaßen zusammensetzen:

$$\hat{\gamma}_{\text{zent}} = \hat{\gamma} - \frac{1}{d} \sum_{s=1}^d \hat{\gamma}_s \quad \text{bzw.} \quad \hat{\alpha}_{\text{zent}} = \hat{\alpha} - \frac{1}{d} \sum_{s=1}^d \hat{\alpha}_s$$

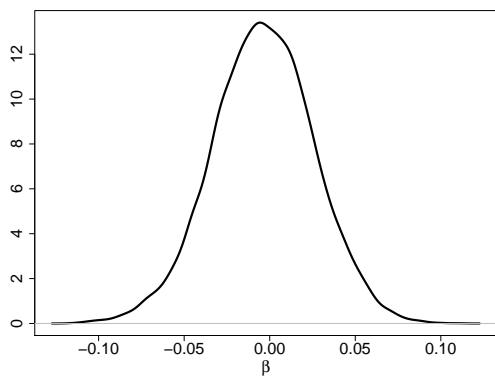
Hierbei werden die jeweiligen Effekte um deren Mittelwert über alle Beobachtungseinheiten hinweg bereinigt, wobei  $\hat{\gamma}$  und  $\hat{\alpha}$  die Vektoren der Posteriori-Mittelwerte des jeweiligen regionenbezogenen Parameters repräsentieren. Deren gemittelte Summe kann schließlich für die Schätzung von  $\hat{\mu}$  verwendet werden:

$$\hat{\mu} = \frac{1}{d} \sum_{s=1}^d \hat{\gamma}_s + \frac{1}{d} \sum_{s=1}^d \hat{\alpha}_s \quad (29)$$

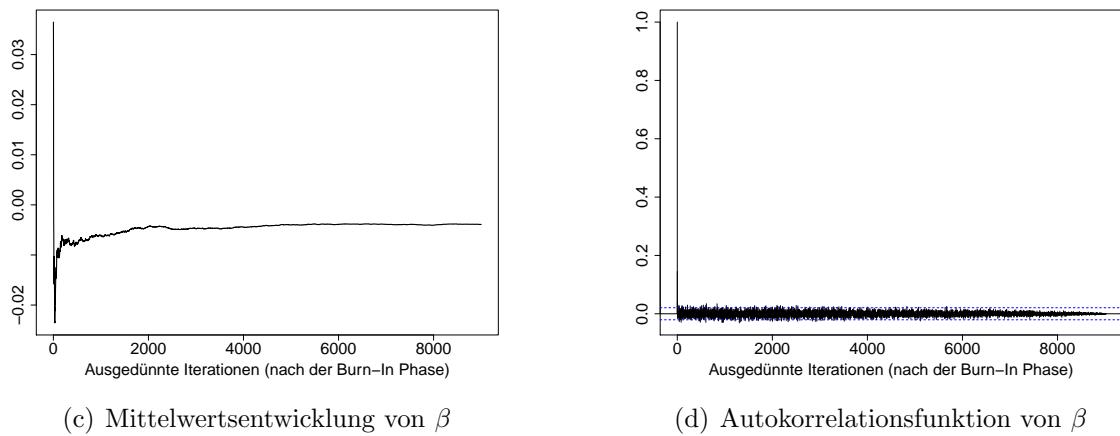
Für die schottischen Lippenkrebsdaten ergeben sich bei einem Iterationsumfang von 100 000, einem Ausdünnungsfaktor von 10 und nach einer Burn-In Phase von 10 000 ein geschätzter Intercept  $\hat{\mu}$  von 1.36. Des Weiteren wird für den Koeffizienten  $\beta$  der Kovariablen des prozentualen Anteils der Beschäftigten in einem der Bereiche aus Agrarwirtschaft, Fischerei oder Forstwirtschaft (AFF) ein Posteriori-Mittelwert von  $-0.0039$  geschätzt, d. h. bei einem Anstieg der Kovariablen AFF um ein Prozent und bei Konstanthaltung aller anderen Effekte verringert sich die geschätzte Anzahl an Sterbefällen in dem entsprechenden Landkreis im Mittel lediglich um den Faktor  $\exp(-0.0039) = 0.996$ . Aus der zusätzlichen Betrachtung des 95%-Kredibilitätsintervalls kann man zudem schließen, dass der Einfluss der Kovariablen auf die Anzahl der beobachteten Sterbefälle auf einem 5%-Niveau nicht signifikant von Null verschieden ist. Ferner resultierte aus dem Sampling für die Präzision  $\kappa$  des GMZF ein geschätzter Mittelwert von 0.77. Im Folgenden kann für die beiden Parameter  $\beta$  und  $\kappa$  eine gleichermaßen positive Bewertung des jeweiligen Samplingpfades (a), der zugehörigen Dichte (b), der Entwicklung des Mittelwerts (c) sowie der Autokorrelationsfunktion (d) in den Abbildungen 30 und 31 erfolgen.



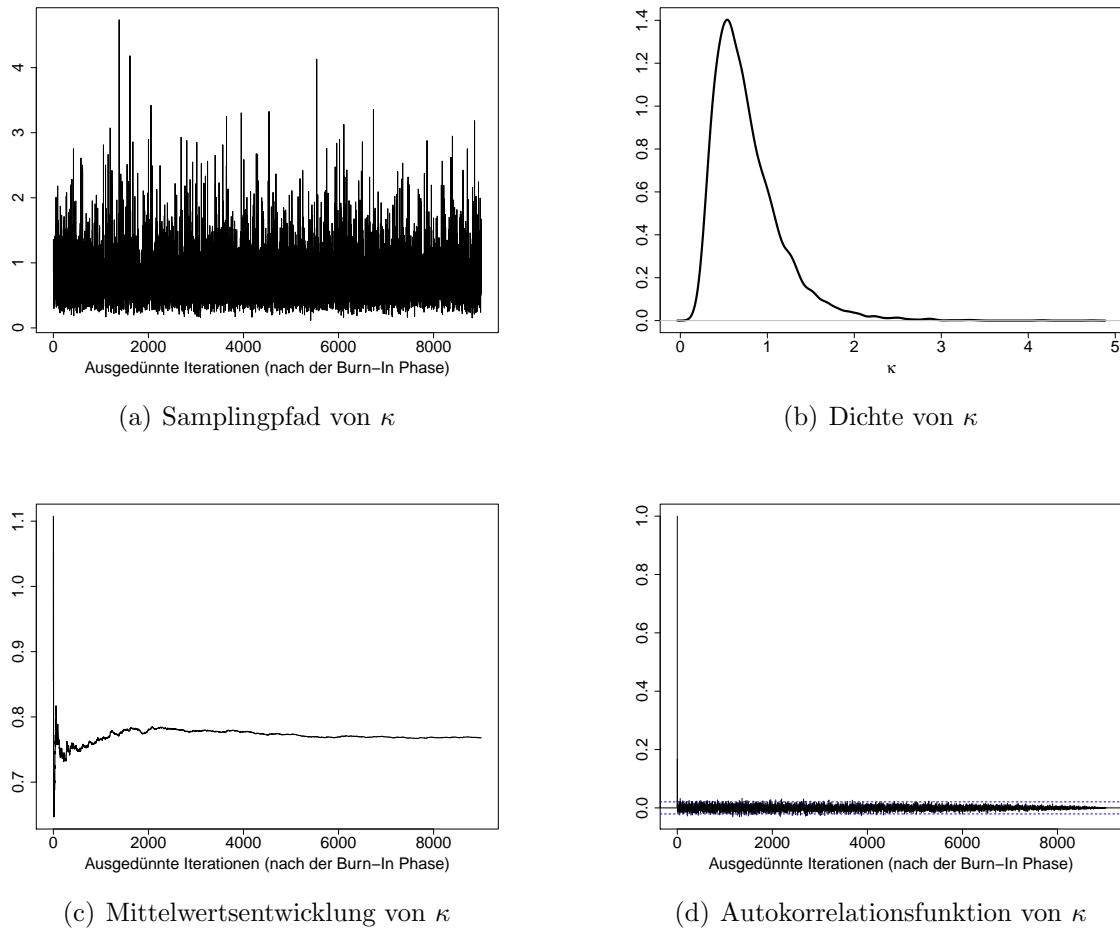
(a) Samplingpfad von  $\beta$



(b) Dichte von  $\beta$

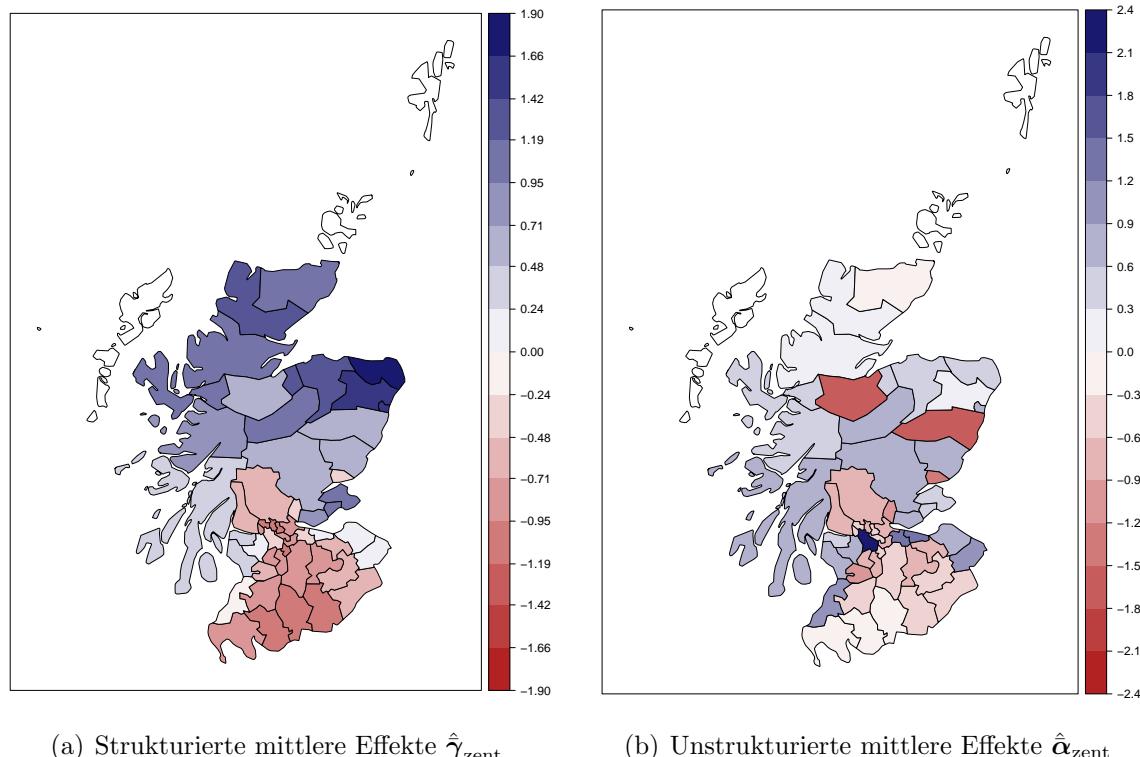
**Abbildung 30:** Samplingaufbereitung von  $\beta$  (Kovariablen-Koeffizient)

Die Betrachtung der Grafiken (d) zeigt zudem, dass die beiden Samplings bereits bei einem niedrigen Ausdünnungsfaktor eine geringe Autokorrelation aufweisen.

**Abbildung 31:** Samplingaufbereitung von  $\kappa$  (Hyperparameter des strukturierten Effekts)

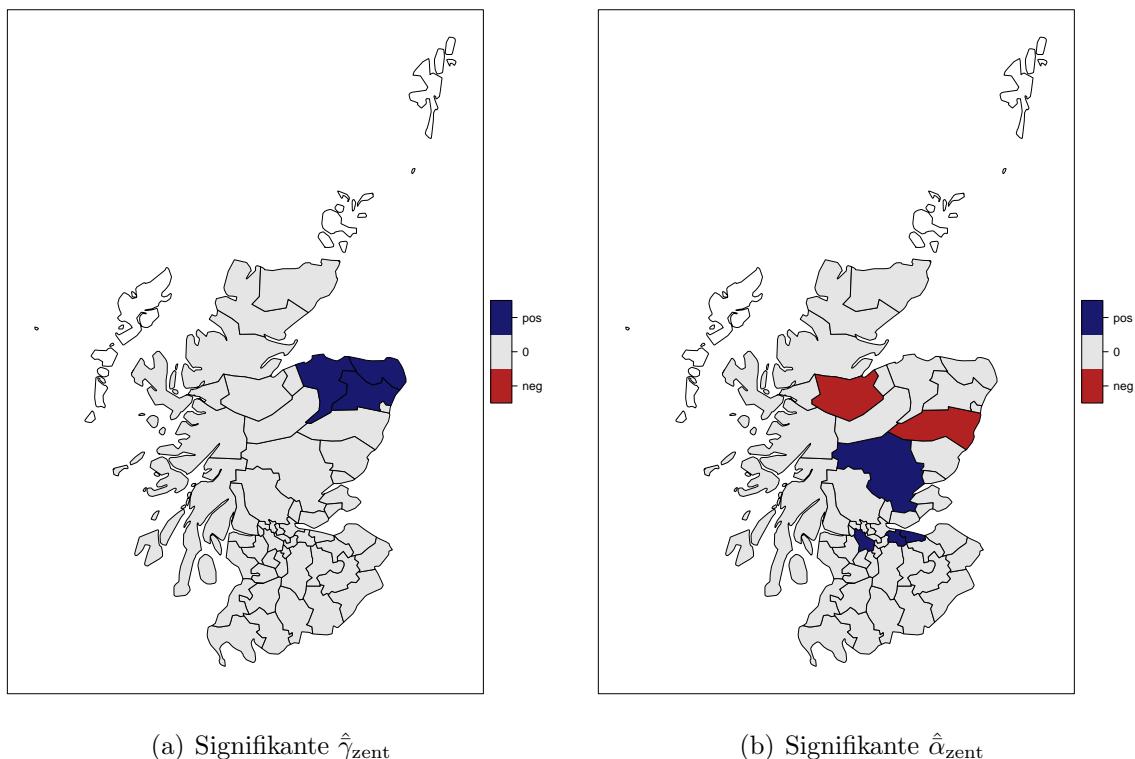
Um einer Speicherplatzproblematik entgegenzuwirken, wurde diesbezüglich bereits im Vorfeld des Samplings ein Faktor von 10 gewählt, welcher gleichzusetzen ist mit der Verzögerung  $k$  der Autokorrelationsfunktion aus (28).

Analog zu der deskriptiven Aufbereitung der Daten in Unterkapitel 3.1, gilt für die weiteren grafischen Veranschaulichungen, dass die Skala der Farbabstufung individuell für jede Grafik festgelegt wird und sich nach den Maximalwerten des jeweiligen Parameters richtet. Farbvergleiche zwischen verschiedenen Abbildungen sind deshalb nicht möglich. Zudem ist zu beachten, dass die Farbverläufe der Skalen jeweils symmetrisch um Null angeordnet sind. Wie bereits in Unterkapitel 2.3.1 zu den theoretischen Grundlagen der räumlich strukturierten und unstrukturierten Effekte beschrieben, können diese nicht eindeutig voneinander getrennt werden, weshalb für eine vollständige Erfassung der räumlichen Struktur eine gemeinsame Betrachtung erforderlich ist. In Abbildung 32 werden in diesem Zusammenhang die Posteriori-Mittelwerte der strukturierten Effekte  $\hat{\gamma}$  und der unstrukturierten Effekte  $\hat{\alpha}$  für die 53 schottischen Festland-Distrikte in ihrer zentrierten Form gegenübergestellt. Aufgrund der Festsetzung auf  $\tau = 1$  als Hyperparameter des unstrukturierten Effekts und einem geschätzten Posteriori-Mittelwert von 0.77 für  $\kappa$ , ergeben sich ähnliche Größenordnungen für die beiden Effekte. Der Beitrag für die Erklärung der Variabilität der beobachteten Sterbefälle ist somit relativ ausgeglichen. Aufgrund des log-linearen Modellaufbaus ist für eine direkte multiplikative Interpretation eine Exponentierung der einzelnen Effekte durchzuführen. Die Betrachtung expliziter Ausprägungen von bestimmten Beobachtungseinheiten steht jedoch nicht im Fokus der folgenden Aufbereitungen.



**Abbildung 32:** Geschätzte Posteriori-Mittelwerte der zentrierten räumlichen Effekte aus den schottischen Lippenkrebsdaten – mit positiver (blau) oder negativer (rot) Ausrichtung

Die theoretischen Grundlagen aus Kapitel 2 haben dargelegt, dass die Verwendung von GMZF eine glatte Modellierung des räumlichen Effekts ermöglichen. Infolgedessen können Schätzungen für Beobachtungseinheiten, in welchen eine niedrige Anzahl an Ausprägungen der jeweiligen Zielgröße dokumentiert wurden, über die Einbeziehung von Nachbarschaftsinformationen stabilisiert werden. Darüber hinaus verschafft die Betrachtung der geglätteten Effekte die Möglichkeit eine sinnvolle Einteilung in größere Bereiche mit über- oder unterdurchschnittlichem Auftreten der Zielgröße durchzuführen (vgl. Fahrmeir *et al.*, 2009, S. 391 f.). Die glatte Schätzung der strukturierten Effekte in Grafik (a) der Abbildung 32 weist in diesem Zusammenhang ein deutliches Nordost-Südwest Gefälle auf. Hervorzuheben ist hierbei der Bereich um Aberdeen im Nordosten Schottlands, in dem die höchsten Posteriori-Mittelwerte der strukturierten Effekte im Vergleich zum landesweiten Durchschnitt auftreten. Zudem ist anzumerken, dass die unstrukturierten Effekte ein verhältnismäßig gutes Gesamtbild aufzeigen, da die Trennung in glatte und zufällige Effekte für diese Betrachtung nicht gut funktioniert. Die spätere Anwendung auf empirische Daten mit Gewichtsadaptionen in Unterkapitel 3.6 wird diesbezüglich ein deutlicheres Bild ergeben. Der Distrikt um die Großstadt Glasgow im Zentrum der Insel sticht bei der Betrachtung der mittleren unstrukturierten Effekte besonders hervor. Daraus kann man schließen, dass dort trotz der Berücksichtigung eines Offsets und der Einflussgröße AFF eine andere Risikozusammensetzung vorliegt, die anderweitig nicht erklärt werden konnte.



**Abbildung 33:** Auf einem 5%-Niveau signifikant von Null verschiedene räumliche Effekte aus den schottischen Lippenkrebsdaten – mit positiver (blau) oder negativer (rot) Ausrichtung

Als Pendant zu den Konfidenzintervallen aus der Likelihood-Inferenz, eignet sich die Betrachtung sogenannter Kredibilitätsintervalle, um räumliche Strukturen im Rahmen einer bayesianischen Modellierung zu verdeutlichen. Diese können direkt interpretiert werden, da nach dem Bayes-Prinzip die Parameter als Zufallsvariablen aufgefasst werden. Ein Parameterwert liegt demnach mit einer Wahrscheinlichkeit von  $1 - \alpha$  in einem Intervall zwischen dem  $(1 - \frac{\alpha}{2})$ - und dem  $(\frac{\alpha}{2})$ -Quantil der zugehörigen Posteriori-Verteilung. Für ein Konfidenzintervall wäre eine solche Interpretation in keinem Fall zulässig (vgl. Holling und Gediga, 2013, S. 278 f.). Umschließt das 95%-Kredibilitätsintervall einer Markov-Kette die Null, so ist der jeweilige Effekt, aufgrund der vorausgegangenen Zentrierung, auf einem 5%-Niveau nicht signifikant von dem landesweiten, durchschnittlichen Effekt verschieden. In Abbildung 33 werden diesbezüglich die entsprechenden Distrikte mit einer grauen Hinterlegung dargestellt. Nach dieser Vorgehensweise können ebenso auch signifikant positive oder negative Effekte aufgedeckt werden. Für die strukturierten Effekte tritt diesbezüglich der bereits erwähnte Bereich im Nordosten der Insel in den Vordergrund, gleichwie die Distrikte der Städte Glasgow und Edinburgh sowie der zentrale Distrikt nördlich davon, in Bezug auf die zufälligen Effekte. Mit einer negativen Ausrichtung sind zudem weitere Regionen im Nordosten der Insel auffällig. Ferner ist anzumerken, dass für alle Samplings dieser insgesamt 106 Parameter ein gutes Mixing sowie schnelle Konvergenzraten beobachtet werden können, vergleichbar mit dem Erscheinungsbild für den Kovariablen-Koeffizienten  $\beta$  und den Hyperparameter  $\kappa$  aus den Abbildungen 30 und 31. Aufgrund dieser Tatsache und hinsichtlich der Vielzahl an Parametersamplings, die diesbezüglich aufbereitet werden müssten, sind die zugehörigen Grafiken für eine Beurteilung des Samplingverlaufs lediglich dem elektronischen Anhang dieser Arbeit beigefügt.

Im Folgenden soll der Prädiktor des Modells betrachtet werden, der aufgrund der vorgenommenen Reparametrisierung über einen eigenen Metropolis-Hastings Schritt in den Hybrid-Sampler eingebunden ist. Die Akzeptanzwahrscheinlichkeit der vorgeschlagenen Zustände, welche aus der Taylorentwicklung von Abschnitt 3.3.2 resultieren, wird dabei über den gesamten Prädiktorvektor  $\eta$  hinweg berechnet.

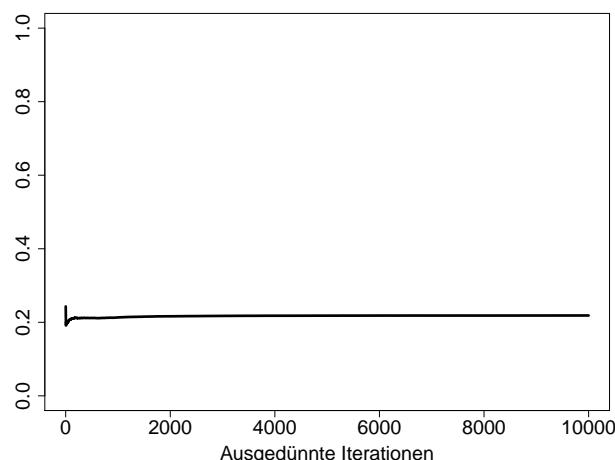
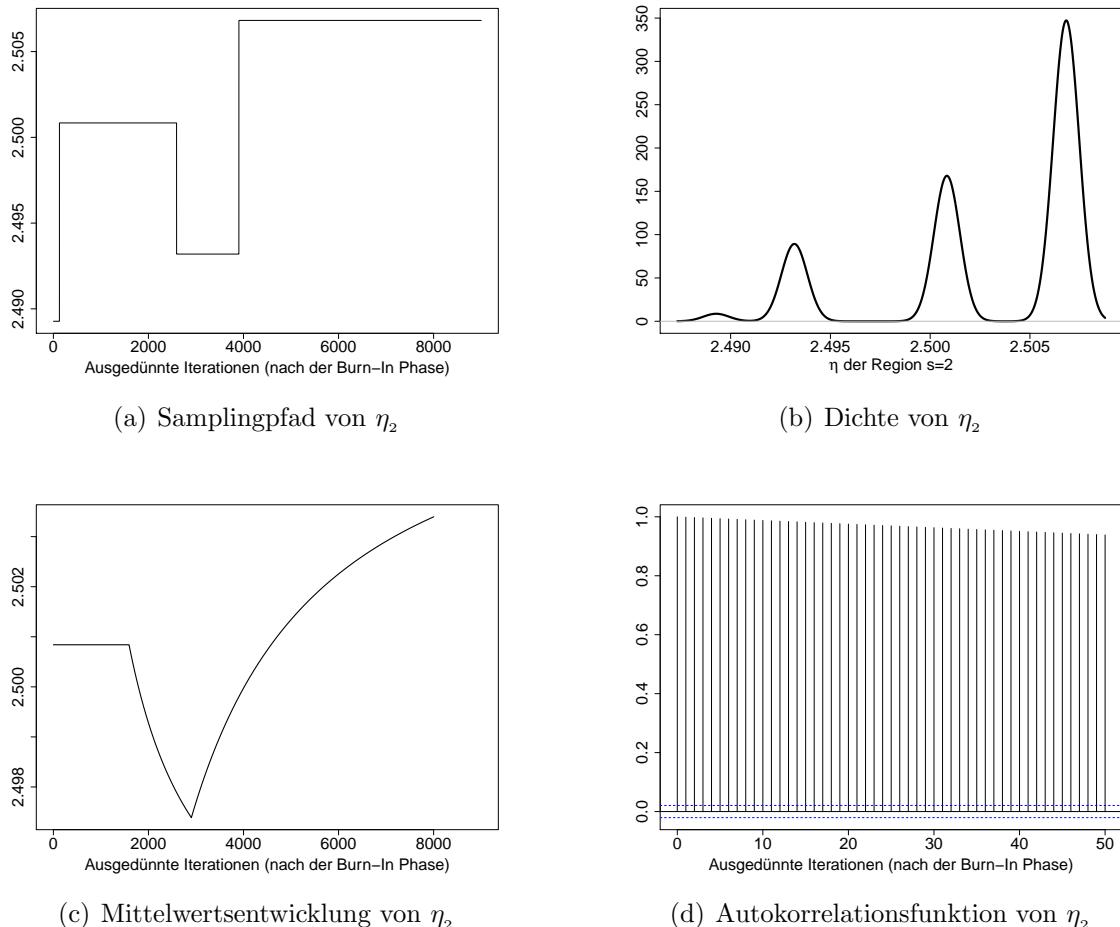


Abbildung 34: Akzeptanzwahrscheinlichkeit von  $\eta$  (schottische Lippenkrebsdaten)

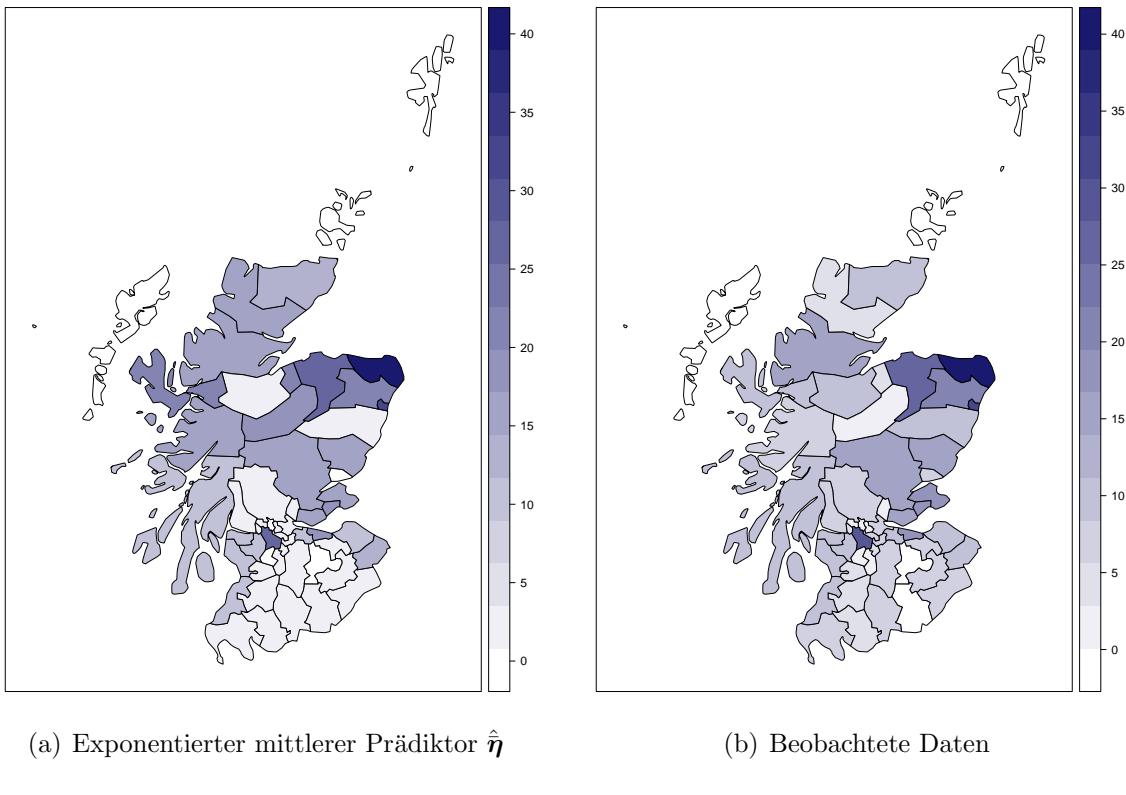
Abbildung 34 dokumentiert, dass selbst bei einer Abtragung aller (ausgedünnten) Iterationsschritte, inklusive der Burn-In Phase, diese Wahrscheinlichkeit schnell gegen einen Wert von ca. 22% konvergiert. Um jedoch eine umfassende Bewertung des gesamten Samplings zu erhalten, ist es notwendig die Samplingverläufe für jeden einzelnen Parameter zu betrachten. Ein Großteil der regionenbezogenen Prädiktorwerte weisen diesbezüglich zufriedenstellende Mischungen und Konvergenzverläufe auf. Im Folgenden soll das Sampling für einen Prädiktorwert aufbereitet werden, für den eine langsame Mischung des Samplings vorliegt, wodurch die Akzeptanzwahrscheinlichkeit für  $\eta$  verringert wird. Anschließende Grafiken veranschaulichen exemplarische den Zustandsverlauf von  $\eta_2$  nach der Burn-In Phase.



**Abbildung 35:** Samplingaufbereitung von  $\eta_2$  (Prädiktorwert der Region  $s = 2$ )

Die Grafiken (a) und (b) aus Abbildung 35 illustrieren in diesem Zusammenhang die erwähnte langsame Mischung der zugehörigen Markov-Kette. Zudem wird bei der Entwicklung des Mittelwerts in Grafik (c) und der Abtragung der Autokorrelationsfunktion in Grafik (d) verdeutlicht, dass kein konvergierender Verlauf vorliegt und aufeinanderfolgende Zustände, trotz einer enormen Verzögerung mit einem Ausdünnungsfaktor

von 2 000, eine starke Abhängigkeit aufweisen. Dieser Sachverhalt ist der Tatsache geschuldet, dass die Taylorentwicklung für vereinzelte Prädiktorwerte nicht besonders gut funktioniert und daher die Akzeptanzwahrscheinlichkeit für eine Veränderung der Parameterzustände sehr gering ausfällt. In einzelnen Extremfällen bleibt das Sampling nach der Burn-In Phase für alle Iterationsschritte unverändert. Eine alternative Herangehensweise, mit einem individuellen Tuning der Akzeptanzwahrscheinlichkeiten, soll in den abschließenden Bemerkungen dieser Arbeit in Unterkapitel 4.2 in den Ausblick gestellt werden. Abschließend erfolgt für die Aufbereitung der Schätzergebnisse aus den schottischen Lippenkrebsdaten, ohne der Einbeziehung anpassungsfähiger Nachbarschaftsgewichtungen, in Abbildung 36 der Vergleich der exponentierten Posteriori-Mittelwerte des Prädiktors (Grafik (a)) und den beobachteten Daten (Grafik (b)).



**Abbildung 36:** Exponentierte Posteriori-Mittelwerte des Prädiktors (a) im Vergleich zu den beobachteten schottischen Lippenkrebsdaten (b)

Trotz der vereinzelt auftretenden Problematik der Modellanpassung in Bezug auf die Akzeptanz neuer Vorschläge für den Prädiktor, ist dieser in seinem distriktspezifischen Ausmaß und in exponentierter Form gut vergleichbar mit den tatsächlich Beobachtungen. Die Betrachtung des globalen Mittelwerts der Daten mit 9.58 und des mittleren exponentierten Posteriori-Mittelwerts des Prädiktors mit 8.96, unterstützt diese Erkenntnis zusätzlich.

### 3.4.4 Schätzergebnisse aus den deutschen Kehlkopfkrebsdaten

Für die Modellierung der deutschen Kehlkopfkrebsdaten werden, mit einem Iterationsumfang von 100 000, einem Ausdünnungsfaktor von 10 sowie einer Burn-In Phase von 10 000, die gleichen Rahmenbedingungen wie bei den schottischen Schätzungen verwendet. Analog dazu wird der Hyperparameter  $\tau$  des unstrukturierten Effekts und die Rate  $b$  der Gamma-Verteilung des strukturierten Effekts auf 1 festgelegt (siehe Modellanpassungen in Unterkapitel 3.4.1). Neben  $\beta$  und  $\kappa$  müssen, im Vergleich zu den weiteren 106 Parametern der schottischen Modellierung, für die deutsche Betrachtung mit 1088 Parametern eine wesentlich größere Anzahl in die Schätzungen einbezogen werden. Aufgrund dieser hohen Gesamtzahl verschlechtert sich die Anpassung durch die Taylorentwicklung zusätzlich. Ohne ein weiteres Eingreifen würde bei dieser Konstellation die Akzeptanzwahrscheinlichkeit für  $\eta$  schnell auf 0% abfallen. Um dem entgegenzuwirken, werden jeweils die ersten fünf Vorschläge für die einzelnen landkreisbezogenen Prädiktorwerte mit einer Wahrscheinlichkeit von 100% akzeptiert. Diese Vorgehensweise führt zumindest zu einer Akzeptanz von 2%, die sehr schnell gegen diesen Wert konvergiert.

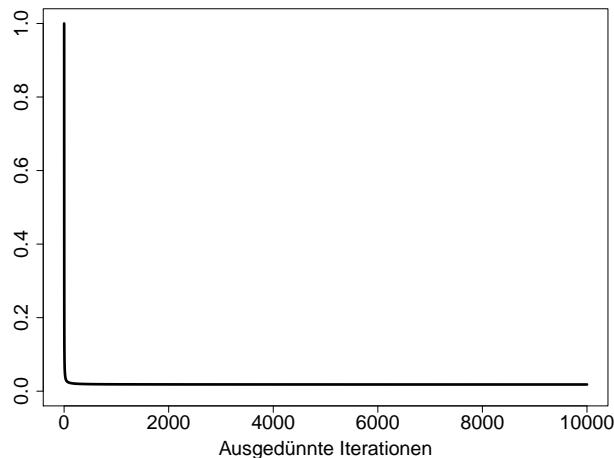
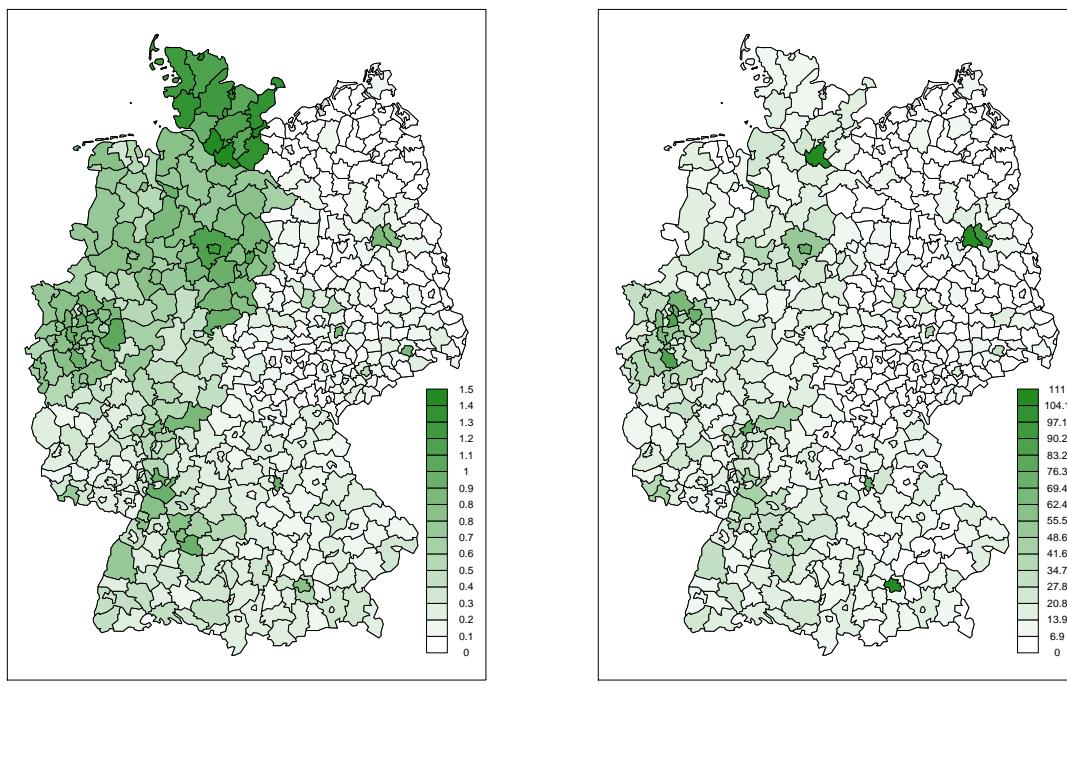


Abbildung 37: Akzeptanzwahrscheinlichkeit von  $\eta$  (deutsche Kehlkopfkrebsdaten)

Dieses niedrige Akzeptanzniveau entsteht, da etwa die Hälfte der Prädiktorsamplings, d. h. ein wesentlich größerer Anteil als bei der schottischen Modellierung, nach der Burn-In Phase einen konstanten Wert einnimmt. Viele der weiteren Markov-Ketten weisen zudem ein langsames Mixing auf, vergleichbar mit der beispielhaften Betrachtung in Abbildung 35. An dieser Stelle wird aufgrund der Ähnlichkeit zu bereits gezeigten Grafiken auf eine Aufbereitung der einzelnen Samplings verzichtet. Für eine detaillierte Betrachtung aller regionenbezogener Samplingverläufe sind wiederum sämtliche Grafiken dem elektronischen Anhang dieser Arbeit angefügt.

Bei dem Vergleich der beobachteten Daten mit den mittleren Prädiktorwerten in ihrer exponentieller Form fällt auf, dass sich deren landesweiter Mittelwert mit 0.48 auf einem deutlich anderen Niveau befindet, als der Mittelwert der beobachteten Daten mit 13.39. Trotz unterschiedlichem Skalenniveau ist jedoch ein Großteil der Aus-

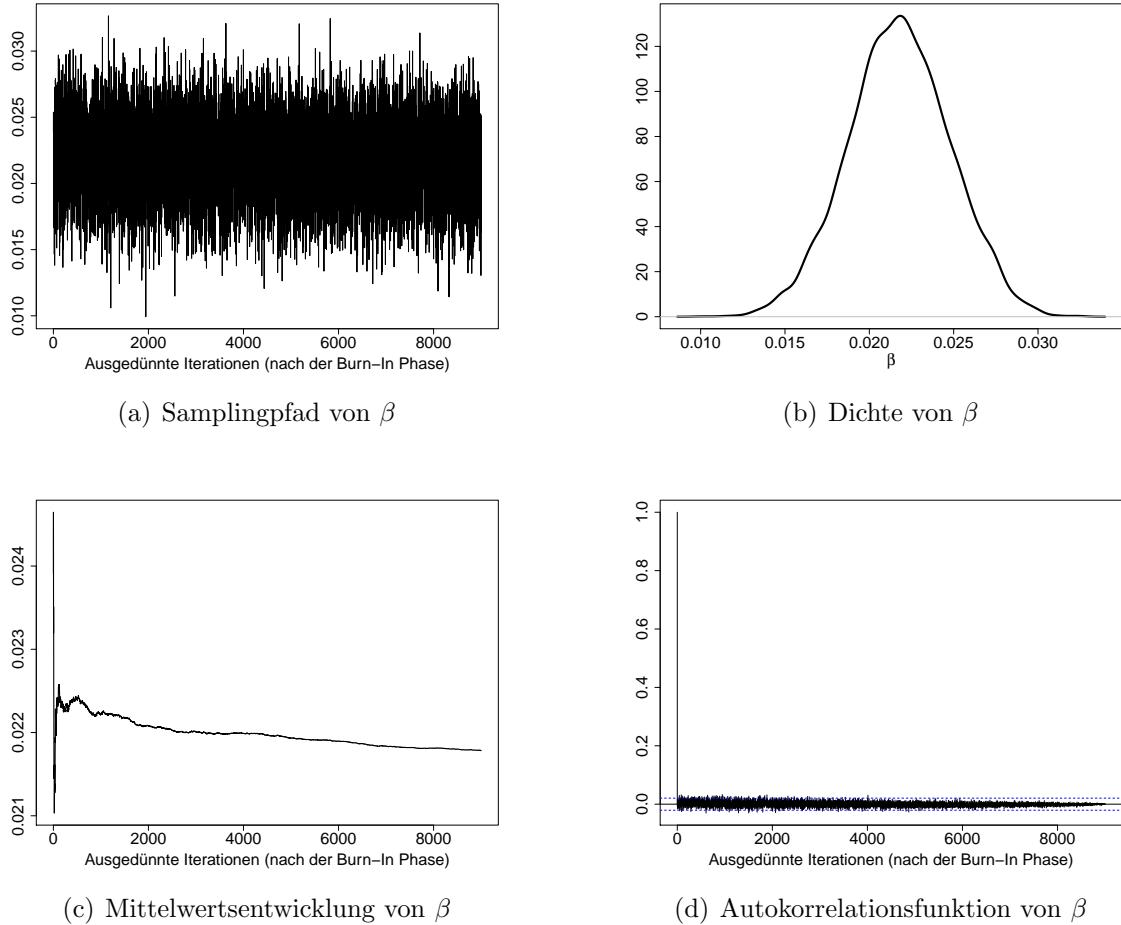
prägungsstrukturen in beiden Grafiken der Abbildung 38 wiedererkennbar. Um eine vernünftige Visualisierung zu ermöglichen, sind in diesem Zusammenhang einzelne Ausreißer bei den jeweiligen Farbskalenbildungen ausgeschlossen worden. Davon betroffen sind lediglich zwei kleine Landkreise in Schleswig-Holstein in Grafik (a) sowie die Landkreise der beiden Städte Westberlin und Hamburg in Grafik (b), durch deren Berücksichtigung keine weiteren Strukturen erkennbar wären (vgl. hierfür Abbildung 20 auf Seite 52). Die entsprechenden Beobachtungseinheiten sind dabei in der dunkelsten Einfärbung dargestellt, wobei demzufolge deren tatsächliche Werte nicht dem abgetragenen Skalenniveau entsprechen.



**Abbildung 38:** Exponentierter Prädiktor (a) im Vergleich zu den beobachteten deutschen Kehlkopfkrebsdaten (b)

Ähnlich zu den beobachteten Daten sind bei den Prädiktorwerten einzelne Landkreise, aber auch größere Bereiche zu erkennen, in denen eine Häufung an Sterbefällen aufgrund von Kehlkopfkrebs erwartet bzw. beobachtet wurde. Zudem fällt in beiden Betrachtungen auf, dass ein Großteil der ostdeutschen Landkreise tendenziell eher weniger Sterbefälle zu verzeichnen hat. In den nördlichen Regionen in und um das Bundesland Schleswig-Holstein werden im Vergleich höhere Posteriori-Mittelwerte des exponentierten Prädiktors geschätzt. Dabei ist anzumerken, dass für diese Landkreise ein deutlich besseres Mixing beobachtet werden kann.

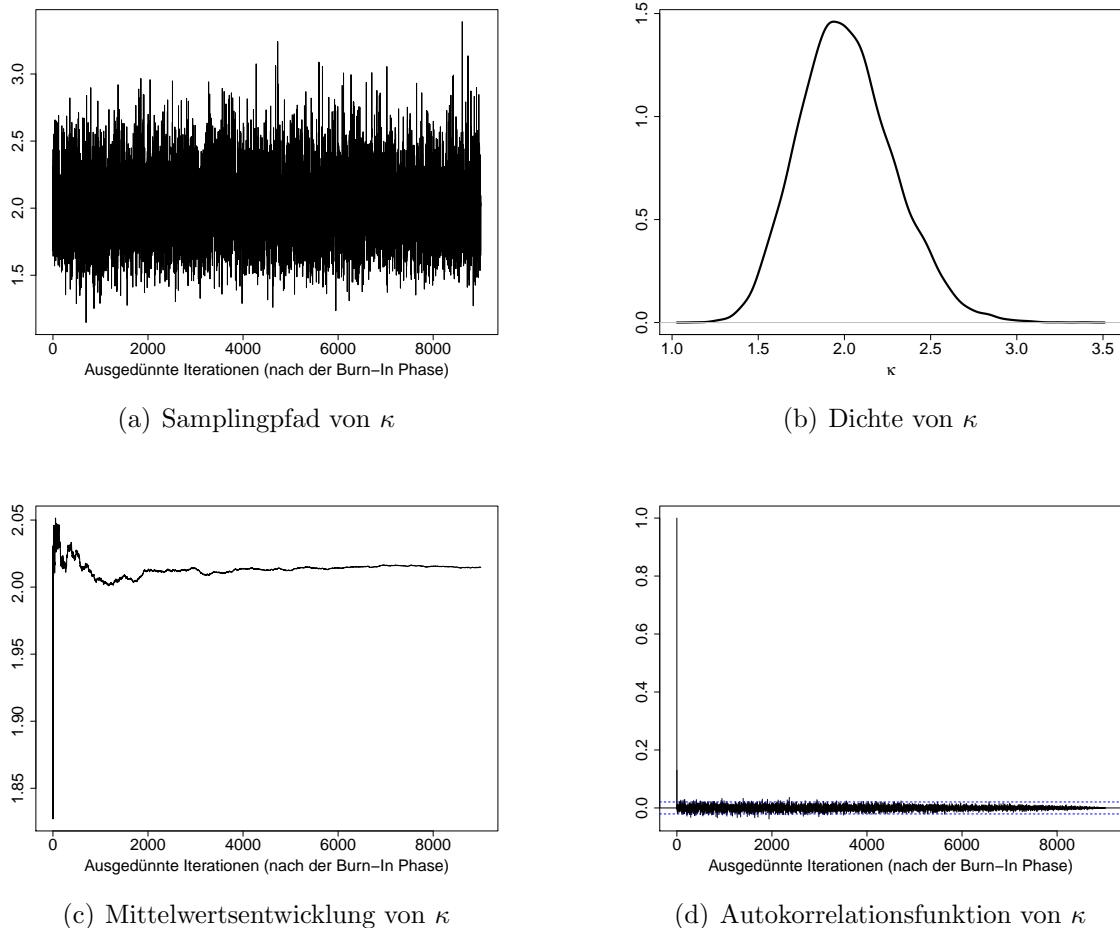
Die Aufbereitung weiterer Schätzergebnisse, insbesondere die der räumlichen Effekte, soll zusätzliche Strukturen aufzeigen. Zunächst stehen jedoch der Intercept, der Kovariablen-Koeffizient sowie der Hyperparameter des strukturierten Effekts im Fokus der Betrachtungen. Analog zu der Berechnung in Gleichung 29 kann der Intercept mit  $\hat{\mu} = -1.76$  über die mittleren Posteriori-Mittelwerte der räumlichen Effekte nach dem eigentlichen Sampling geschätzt werden. Des Weiteren ergibt sich für  $\beta$  ein Wert von 0.0218, d. h. bei einem ansteigenden Level des Zigarettenkonsums um eine Einheit und bei Konstanthaltung aller anderen Effekte, erhöht sich die geschätzte Anzahl an Sterbefällen in dem entsprechenden Landkreis im Mittel um den Faktor  $\exp(0.0218) = 1.022$ . Abbildung 39 zeigt außerdem, dass sowohl das Mixing (Grafiken (a) und (b)) als auch die Unabhängigkeit des Samplings (Grafik (d)) positiv zu bewerten sind. Alleinig die Konvergenz der Mittelwertsentwicklung, abgetragen in Grafik (c), ist nicht zu 100% zufriedenstellend.



**Abbildung 39:** Samplingaufbereitung von  $\beta$  (Kovariablen-Koeffizient)

Bei einer näheren Betrachtung der Skalierung der Ordinatenachse fällt jedoch auf, dass sich die Variierung der Mittelwertsentwicklung bereits in einem sehr engen Bereich abspielt. Aus diesem Grund ist die noch nicht vollständig erfolgte Konvergenz akzeptabel.

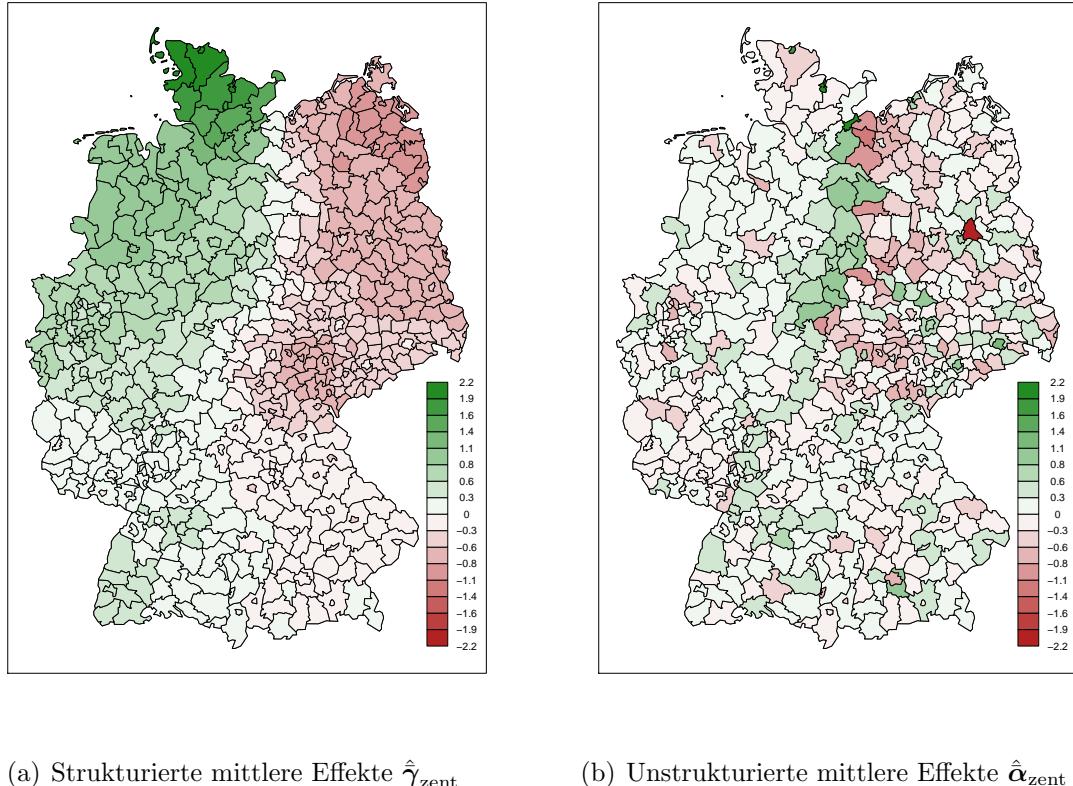
Im Weiteren ergibt sich für  $\kappa$  im Vergleich zu der schottischen Schätzung ein etwas höherer Posteriori-Mittelwert von 2.01. Die grafische Bewertung des Samplings im Rahmen der anschließenden Abbildung kann diesbezüglich als gut eingestuft werden.



**Abbildung 40:** Samplingaufbereitung von  $\kappa$  (Hyperparameter des strukturierten Effekts)

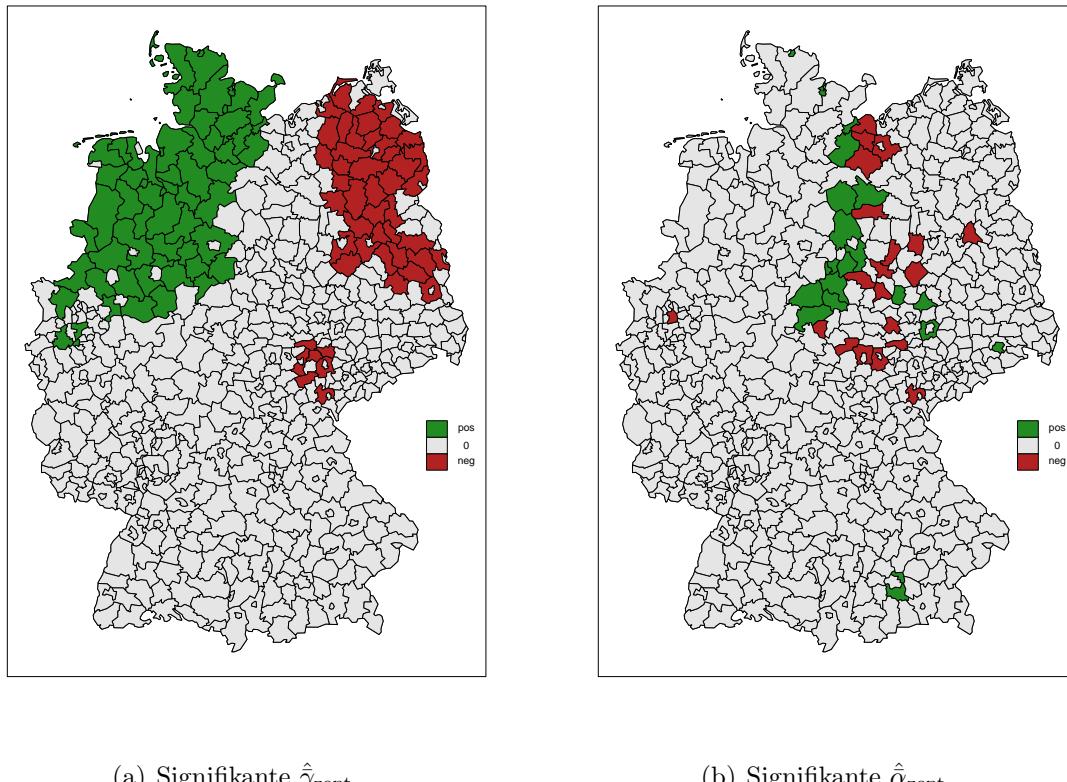
Wie bereits bei den schottischen Schätzergebnissen, sind die mittleren strukturierten und unstrukturierten Effekte in ihrer Größe und somit in ihrem Beitrag zur Erklärung der Variabilität der Zielgröße relativ ausgeglichen. Um eine verbesserte Gesamtübersicht über die 544 deutschen Landkreise in Bezug auf deren Posteriori-Mittelwerte zu erhalten, wird die Farbskalierung der Grafiken in Abbildung 41 wiederum geringfügig angepasst. Hierfür werden jeweils die beiden größten Ausreißer aus der landesweiten Farbbestimmung ausgenommen und schließlich mit dem dunkelsten Farbton eingefärbt. Die glatte Schätzung der Posteriori-Mittelwerte des strukturierten räumlichen Effekts in Grafik (a) weist dabei eine auffällige Erhöhung im Nordwesten Deutschlands sowie einen deutlichen Bruch an der damaligen Grenze von West- zu Ostdeutschland auf. Dieser wird zusätzlich durch die Betrachtung der unstrukturierten Effekte verstärkt (siehe Grafik (b)), da offensichtlich für Landkreise, die über diese Grenze hinweg benachbart sind, trotz der Berücksichtigung des Zigarettenkonsumlevels als Einflussgröße,

eine deutlich andere Risikozusammensetzung vorliegt. Zudem sticht der Landkreis von Westberlin hervor, für den, trotz der Berücksichtigung eines Offsets, wesentlich weniger Sterbefälle erwartet werden.



**Abbildung 41:** Geschätzte Posteriori-Mittelwerte der zentrierten räumlichen Effekte aus den deutschen Kehlkopfkrebsdaten – mit positiver (grün) oder negativer (rot) Ausrichtung

Darüber hinaus werden die erwähnten Landkreise durch die folgende Betrachtung der jeweiligen Kredibilitätsintervalle zusätzlich hervorgehoben. Beobachtungseinheiten, deren mittlere Effekte auf einem 5%-Niveau signifikant von Null verschieden sind, erhalten eine dementsprechende farbige Markierung, dargestellt in Abbildung 42. Dabei ist anzumerken, dass für alle grünen Regionen das zugehörige Intervall oberhalb und für alle roten Regionen unterhalb der Null liegt. Alle Landkreis, deren räumliche Effekte nicht signifikant von Null verschieden sind, werden grau eingefärbt. In Grafik (a) ist diesbezüglich im Nordwesten ein signifikant höherer und im Nordosten ein signifikant niedrigerer, mittlerer strukturierter Effekt im Vergleich zum deutschlandweiten Durchschnitt zu beobachten. Signifikante Posteriori-Mittelwerte der unstrukturierten Effekte (siehe Grafik (b)) treten dahingegen hauptsächlich entlang der ehemaligen innerdeutschen Grenze auf.



**Abbildung 42:** Auf einem 5%-Niveau signifikant von Null verschiedene räumliche Effekte aus den deutschen Kehlkopfkrebsdaten – mit positiver (grün) oder negativer (rot) Ausrichtung

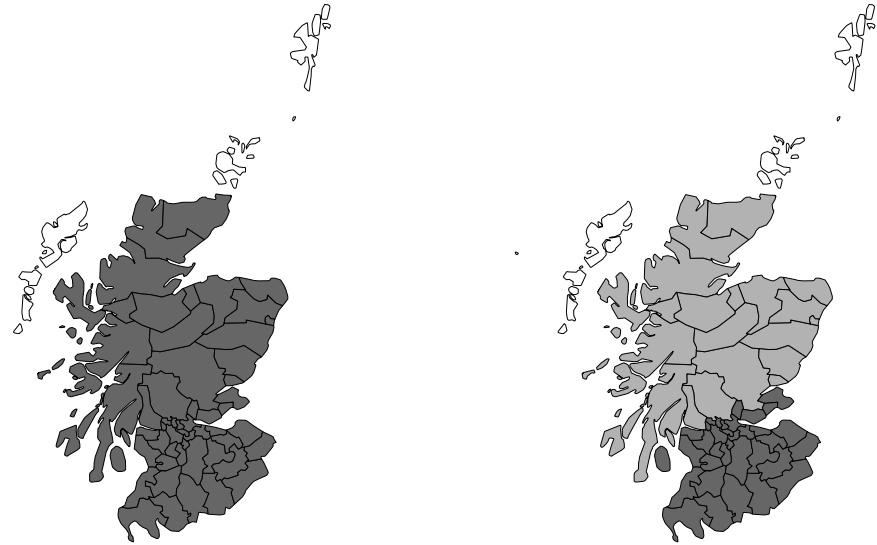
### 3.5 Anwendung auf simulierte Daten mit Gewichtsadaption

Die bisherige Einschränkung des Hybrid-Samplers durch die Ausklammerung von adaptiven Gewichten und somit einer Festsetzung auf einheitliche Nachbarschaftsbeziehungen, wird im Folgenden aufgehoben. Zunächst wird diesbezüglich eine Anwendung auf simulierte Daten durchgeführt, bevor die Berücksichtigung von adaptiven Gewichten in Unterkapitel 3.6 für die bekannten empirischen Daten umgesetzt wird. Für die Realisierung von simulationsbasierten Schätzungen nimmt die eigentliche Generierung der Daten eine entscheidende Rolle ein. Verschiedene Ausgangssituationen, definiert über unterschiedliche Grenzsettings, sowie die Verteilungsannahmen einzelner Parameter werden in diesem Zusammenhang in Abschnitt 3.5.1 betrachtet. Die auszugsweise Aufbereitung der erzielten Schätzergebnisse in Unterkapitel 3.5.2 verschafft außerdem einen ersten Überblick für deren Bewertung. Im Weiteren kann der mittlere quadratische Fehler als geeignetes Maß für eine detaillierte Beurteilung der durchgeführten Samplings von simulierten Daten betrachtet werden (siehe Unterkapitel 3.5.3).

#### 3.5.1 Ausgangslage für das Sampling simulierter Daten

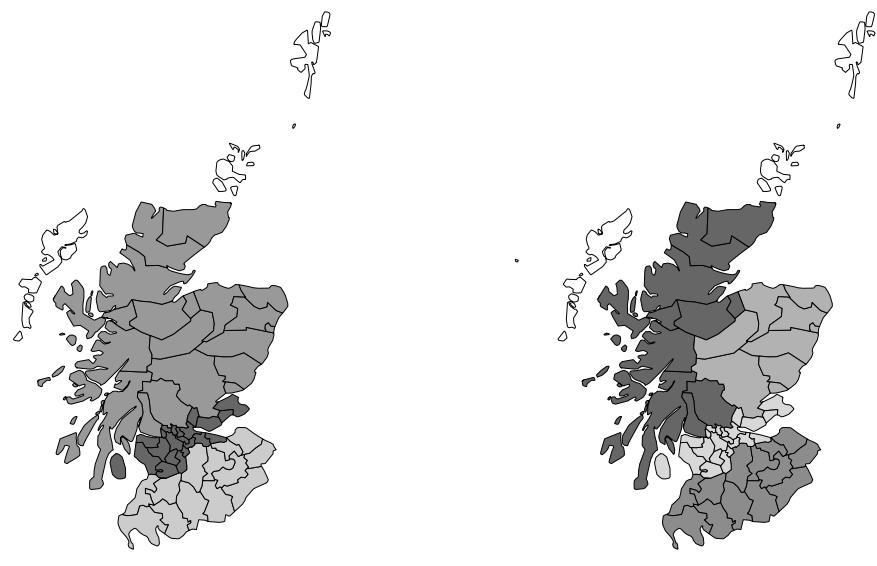
Bevor der eigentliche Sampler seine Anwendung findet, muss zunächst auf die Generierung der simulierten Daten eingegangen werden. Basierend auf der bereits bekannten Distriktstruktur von Schottland mit insgesamt 117 direkten Nachbarschaften werden

dafür im Folgenden vier verschiedene Grenzsettings bestimmt. Durch diese Vorgehensweise werden unterschiedliche Voraussetzungen für die Simulierung der Daten geschaffen. Folgende Grenzstrukturen sind dabei von Interesse:



(a) Setting 1 – ohne Grenze

(b) Setting 2 – mit einer Grenze



(c) Setting 3 – mit zwei Grenzen

(d) Setting 4 – mit vier Grenzen

**Abbildung 43:** Vier verschiedene Grenzsettings als Grundlage der Datensimulation

Für die Simulation der Daten spiegeln sich diese vier verschiedenen Grenzsettings in der jeweiligen Nachbarschaftsmatrix  $\mathbf{Q}_w$  wider. Im Weiteren werden diese durch  $\mathbf{Q}_w^{(j)}$  gekennzeichnet, mit  $j = 1, \dots, 4$  für die Unterscheidung nach der jeweiligen Konfiguration.

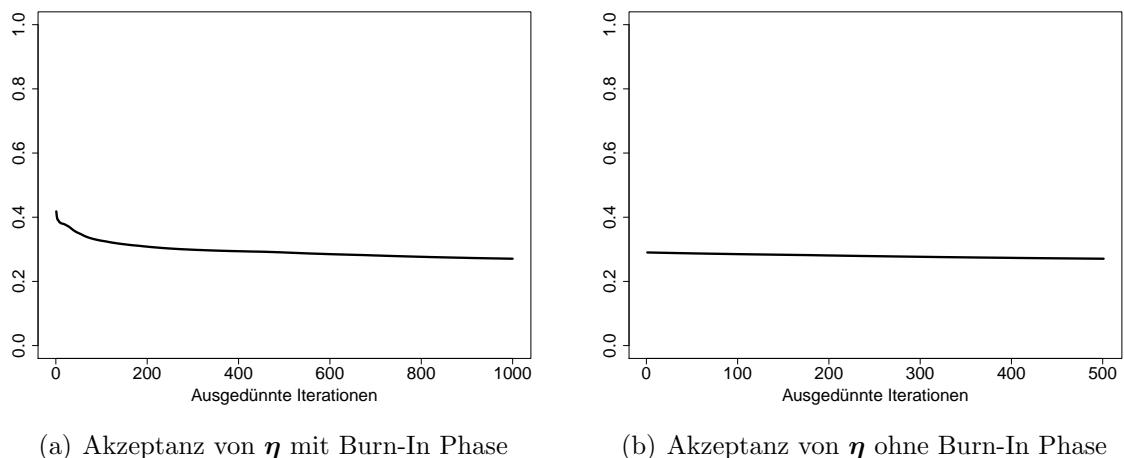
Da für das erste Setting keine einzige Grenze von Bedeutung sein soll (siehe Abbildung 43 Grafik (a)), werden in diesem Zusammenhang alle Gewichte einheitlich auf einen Wert von 1.5 festgesetzt. Für die zweite Ausgangssituation dient lediglich eine einzige künstliche Grenze für die Unterscheidung in einen nördlichen und einen südlichen Landesteil – in Grafik (b) gekennzeichnet mit Distrikten in hellem und dunklem Grau. Alle Nachbarschaften innerhalb der beiden jeweiligen Bereiche werden weiterhin mit einer Gewichtung von 1.5 versehen, im Gegensatz zu den grenzüberschreitenden Nachbarschaften, die mit einem deutlich kleineren Wert von 0.1 gewichtet werden. Grundidee ist dabei, dass über die imaginäre Grenze hinweg lediglich geringe strukturelle räumliche Einflüsse simuliert werden. Analog dazu setzen sich die Nachbarschaftsmatrizen des 3. und 4. Grenzsettings aus den Grafiken (c) und (d) zusammen, wiederum mit einer Gewichtung von 1.5 innerhalb der markierten Bereiche und einer Gewichtung von 0.1 bei grenzüberschreitenden Nachbarschaften.

In Bezug auf die Simulation ist außerdem zu beachten, dass zusätzlich keine Kovariable berücksichtigt wird und die Hyperparameterwerte von  $\kappa$  und  $\tau$  nicht simuliert werden, sondern mit  $\kappa = 1$  und  $\tau = 100$  fest gewählten Werten entsprechen. Die räumlichen Effekte, aufgeteilt in die beiden Komponenten  $\gamma$  und  $\alpha$ , finden über zufällige Ziehungen aus deren Priori-Verteilungen  $\gamma \sim \mathcal{MVN}(0, (\kappa Q_w^{(j)})^{-1})$  und  $\alpha \sim \mathcal{MVN}(0, T^{-1})$  ihre Berücksichtigung. Durch eine anschließende Standardisierung wird zudem sichergestellt, dass sich die generierten Effekte in einem sinnvollen Wertebereich befinden. Für die späteren Betrachtungen ist die explizite Trennung in strukturiert und unstrukturiert nicht von besonderer Bedeutung, vielmehr sind die Schätzergebnisse der summierten räumlichen Effekte sowie der einzelnen Gewichtungsparameter von Interesse. Für die Beobachtungseinheiten setzt sich der Prädiktor  $\eta_s$  des beschriebenen Modellaufbaus ausschließlich aus der erwähnten Summe von  $\alpha_s$  und  $\gamma_s$  zusammen. Diese bilden in ihrer exponentierten Form letztlich die Grundlage für die Intensität  $\lambda_s$ , über die aus der Poissonverteilung zufällige Zielgrößenwerte für die jeweilige Beobachtungseinheit generiert werden können. Für jedes der vier Settings wurden in diesem Zusammenhang jeweils 10 Datensätze gebildet, auf deren Grundlage über den Hybrid-Sampler die entsprechenden Verteilungscharakteristika geschätzt werden konnten.

Analog zu den festen Parametereinstellungen für das Sampling aus den empirischen Daten, ohne der Verwendung anpassungsfähiger Gewichte, wird für die folgenden Modellierungen der Hyperparameter  $\tau$  des unstrukturierten Effekts auf 1 festgesetzt. Aufgrund der vielen zusätzlichen Gewichtungsparameter ist es außerdem notwendig ebenso den Hyperparameter  $\kappa$  der strukturierten Effekte im Vorfeld zu bestimmen, um ein stabiles Sampling zu erhalten. Diese Fixierung kann durch die anpassungsfähige Modellierung der Gewichte aufgefangen werden, weshalb grundsätzlich die Zusammensetzung der Präzisionsmatrix in ihrer Flexibilität nicht eingeschränkt wird. Des Weiteren gilt es für den Verteilungsparameter  $\nu$ , bekannt aus der Priori-Verteilungsannahme der Gewichte, einen geeigneten Wert zu bestimmen. Brezger *et al.* (2007) schlagen für die Anwendung auf deren Brain Mapping Daten eine Festlegung auf 1 vor. Im Zusammenspiel mit  $\kappa = 0.001$  und  $\tau = 1$  ergibt sich, übertragen auf die vorliegende Disease Mapping Problematik, ein sinnvoller Wertebereich für die geschätzten Verteilungscharakteristika der einzelnen Gewichtungen, ohne einer zu großen Einschränkung der Flexibilität der restlichen Modellparameter.

### 3.5.2 Schätzergebnisse der simulierten Daten

Ein vordergründiges Problem bei dem Sampling aus den simulierten Daten ergibt sich durch die verhältnismäßig hohe Anzahl an Nullern, die über die Poissonverteilung für die Zielgröße generiert werden. Als Folge werden für die Prädiktorparameter, der entsprechenden uninformativen Beobachtungen, zum Teil sprunghafte Mischungen registriert. In diesem Kontext ist im Allgemeinen zu erkennen, dass trotz großer Iterationsumfänge die Akzeptanzwahrscheinlichkeit von  $\eta$  eine leicht abfallende Tendenz aufweist. In einer Abwägung, zwischen dem zeitlichen Aufwand für die Durchführung von 40 Samplingdurchläufen und der geringfügigen Verbesserung der Schätzergebnisse durch eine Vielzahl an zusätzlichen Iterationen, wurde für die folgenden Betrachtungen der Umfang auf 200 000 festgelegt, mit einem Ausdünnungsfaktor von 200. Abbildung 44 zeigt in Grafik (a) die beispielhafte Entwicklung einer solchen Akzeptanzwahrscheinlichkeit für den vollständigen Samplingverlauf im Vergleich zu Grafik (b) ohne einer Burn-In Phase von 100 000 Iterationen.

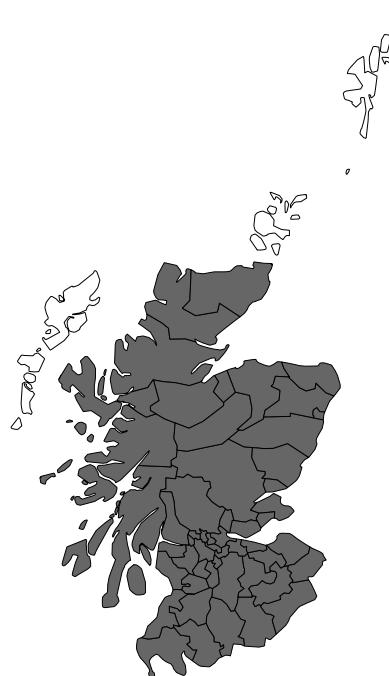


**Abbildung 44:** Beispielhafte Akzeptanzwahrscheinlichkeit von  $\eta$  mit (a) und ohne (b) Burn-In Phase (simulierte Daten mit Gewichtsadaption)

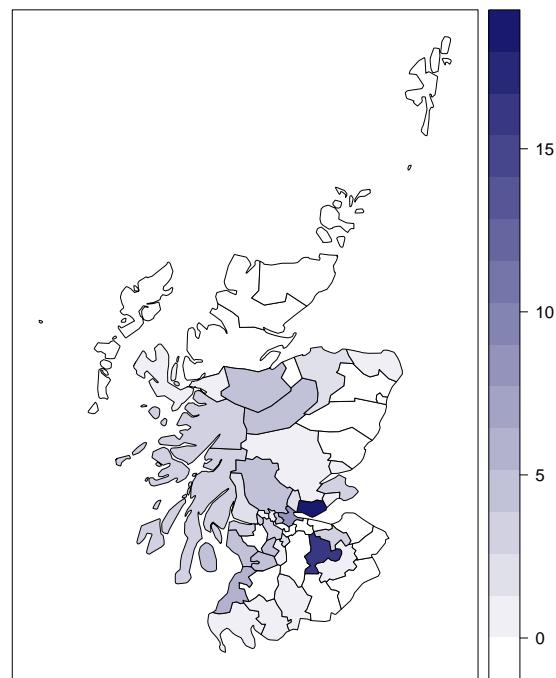
Durch diesen Vergleich kann die Wahl der verhältnismäßig großen Burn-In Phase motiviert werden, da zu erkennen ist, dass die Kurve in Grafik (b) nahezu konvergiert. Die Anwendung auf empirische Daten wird zeigen, dass die Anpassung dort besser funktioniert.

Im Folgenden werden beispielhaft für die vier verschiedenen Settings jeweils eine Simulation mit dem zugehörigen Samplingergebnis der summierten räumlichen Effekte dargestellt. Dabei zeigen die Abbildungen 45, 46, 47 und 48 sowohl die simulierten Daten (jeweils Grafik (b)) als auch die standardisierte Summe der simulierten räumlichen Effekte (jeweils Grafik (c)), im Vergleich zu der zentrierten Summe der entsprechenden geschätzten Posteriori-Mittelwerte (jeweils Grafik (d)). Zusätzlich wird in Grafik (a) das jeweilige Setting als Ausgangslage der Simulation aufgeführt.

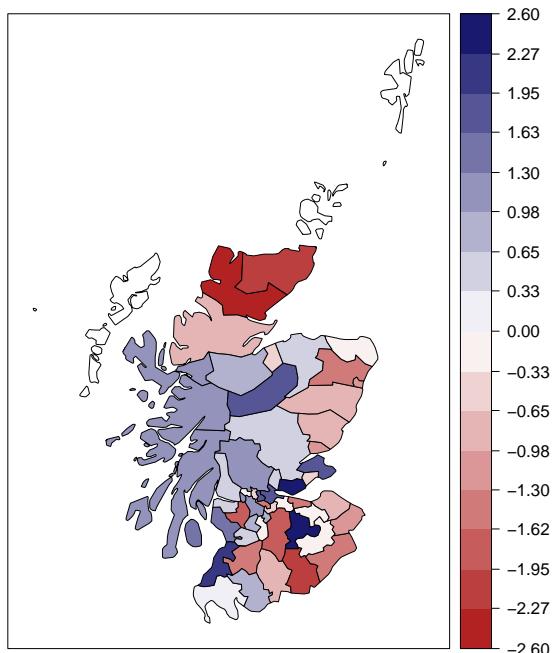
**Darstellung von Simulation und Schätzung: Setting 1 – Simulation 10**



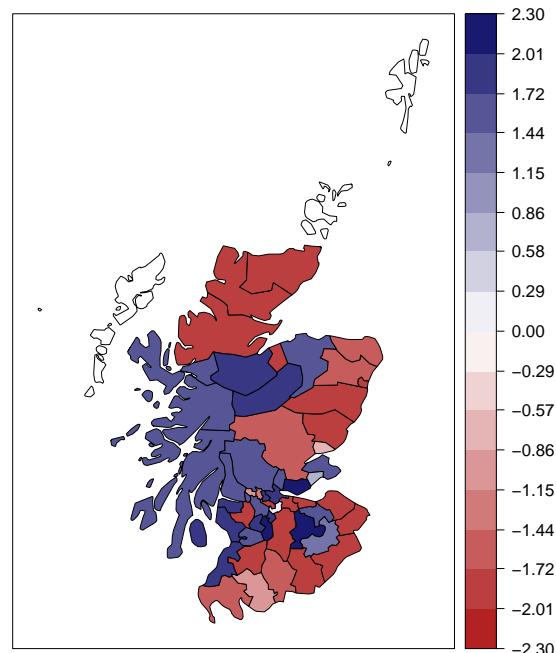
(a) Setting 1



(b) Simulierte Daten



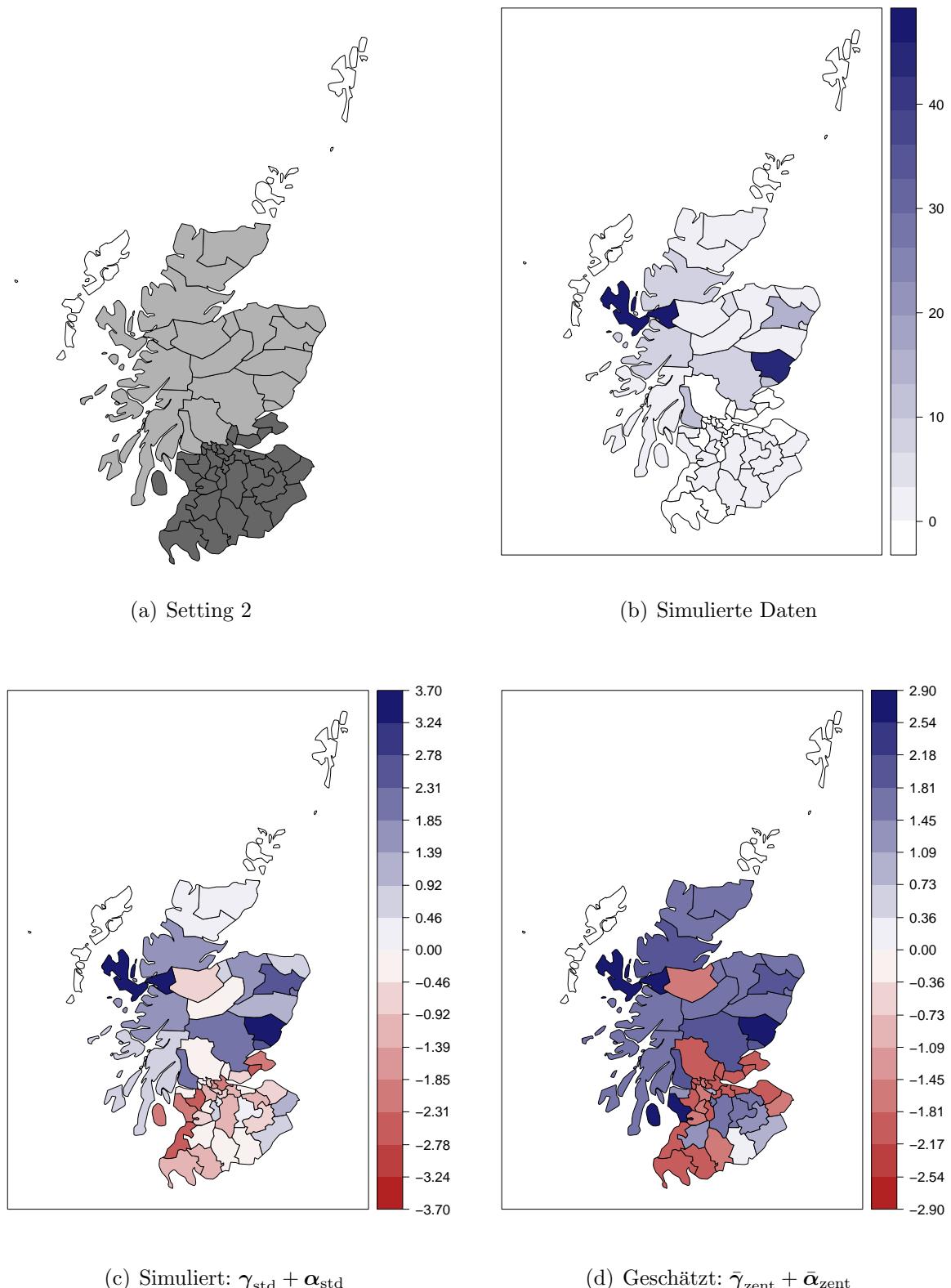
(c) Simuliert:  $\gamma_{\text{std}} + \alpha_{\text{std}}$



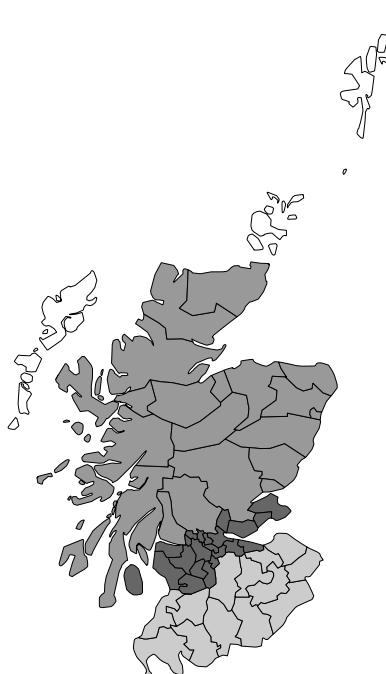
(d) Geschätzt:  $\bar{\gamma}_{\text{zent}} + \bar{\alpha}_{\text{zent}}$

**Abbildung 45:** Aufbereitung der simulierten Schätzergebnisse: Setting 1 – Simulation 10

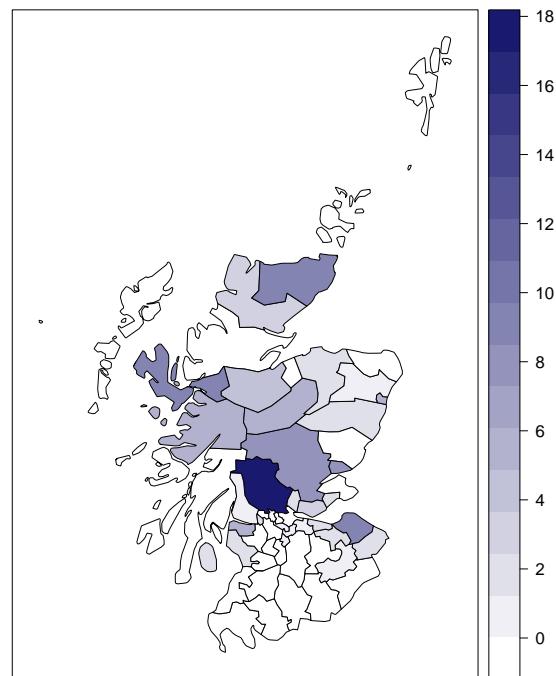
## Darstellung von Simulation und Schätzung: Setting 2 – Simulation 10

**Abbildung 46:** Aufbereitung der simulierten Schätzergebnisse: Setting 2 – Simulation 10

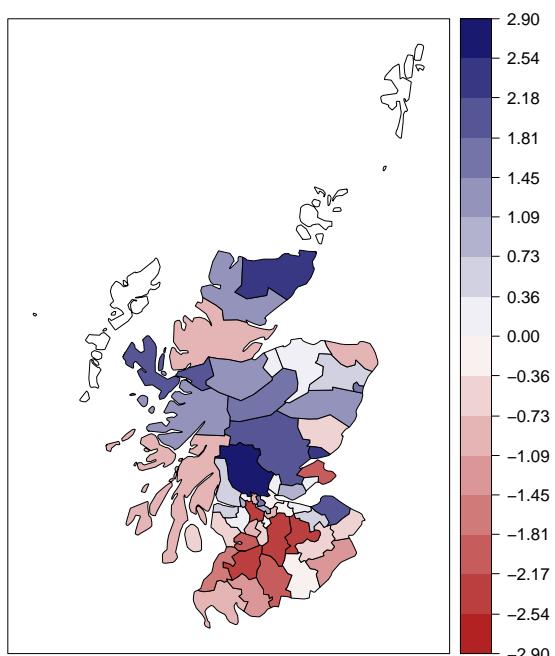
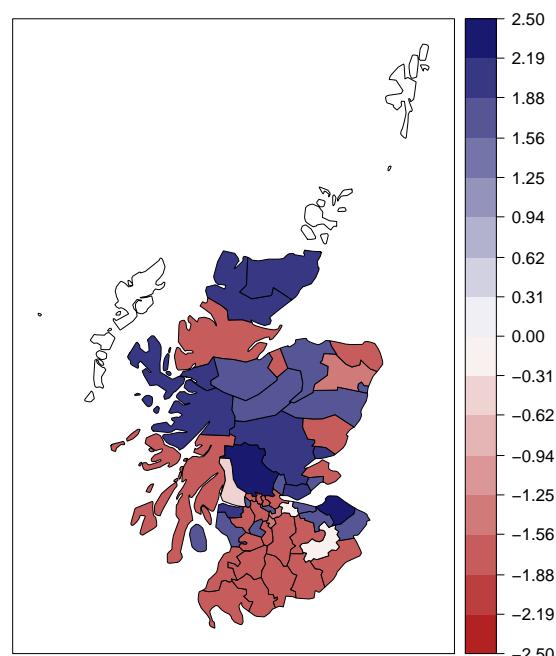
**Darstellung von Simulation und Schätzung: Setting 3 – Simulation 5**



(a) Setting 3

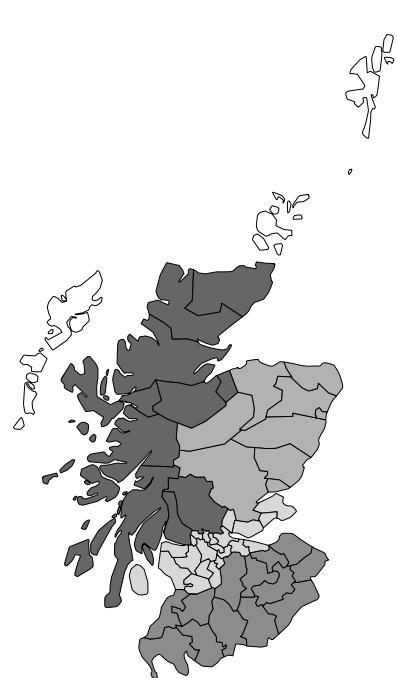


(b) Simulierte Daten

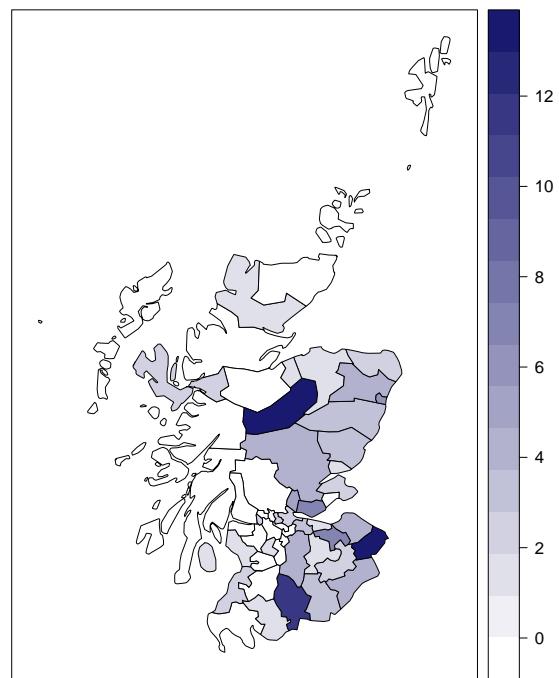
(c) Simuliert:  $\gamma_{\text{std}} + \alpha_{\text{std}}$ (d) Geschätzt:  $\bar{\gamma}_{\text{zent}} + \bar{\alpha}_{\text{zent}}$ 

**Abbildung 47:** Aufbereitung der simulierten Schätzergebnisse: Setting 3 – Simulation 5

## Darstellung von Simulation und Schätzung: Setting 4 – Simulation 4



(a) Setting 4



(b) Simulierte Daten

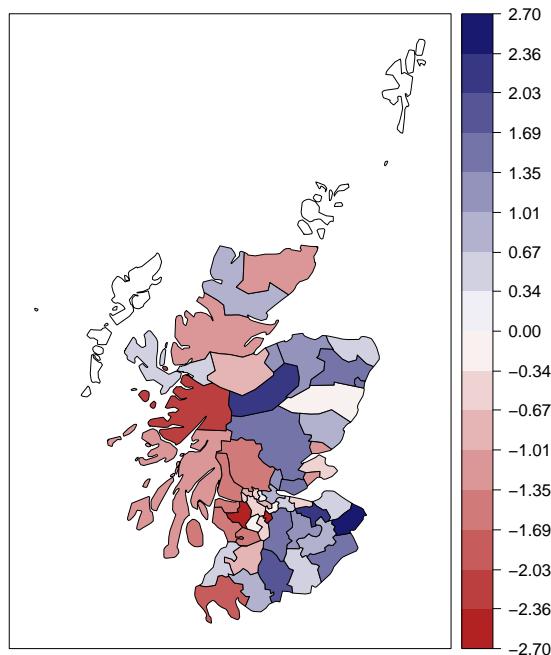
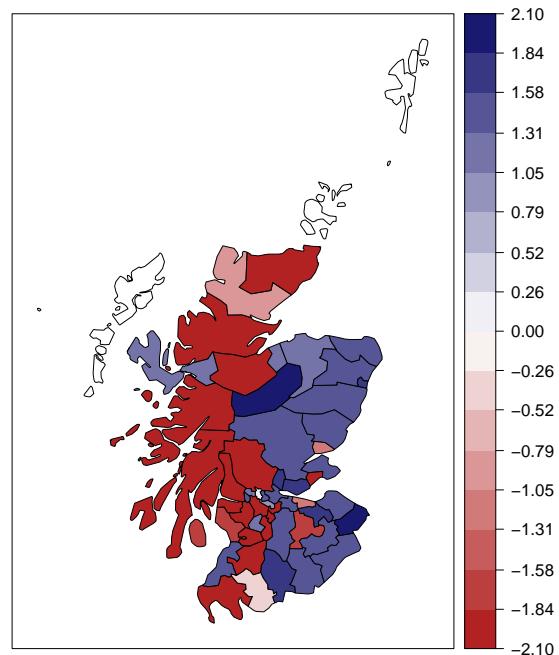
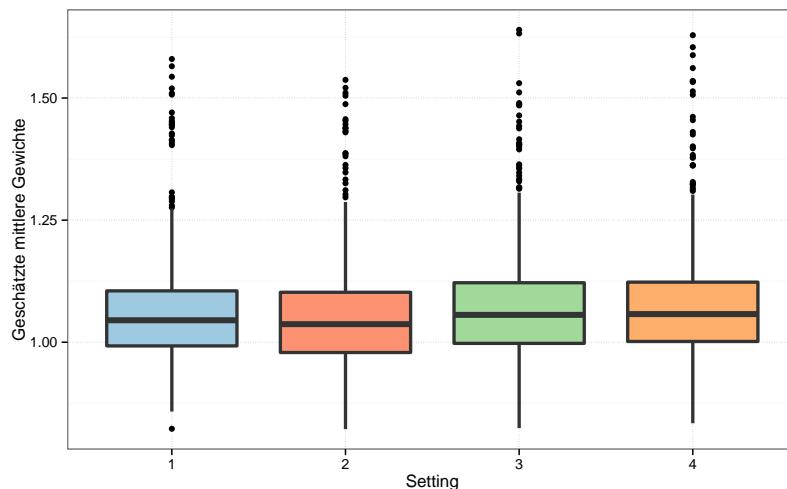
(c) Simuliert:  $\gamma_{\text{std}} + \alpha_{\text{std}}$ (d) Geschätzt:  $\bar{\gamma}_{\text{zent}} + \bar{\alpha}_{\text{zent}}$ 

Abbildung 48: Aufbereitung der simulierten Schätzergebnisse: Setting 4 – Simulation 4

Bei näherer Betrachtung der Abbildungen fällt auf, dass sich die zentrierte Summe der geschätzten Posteriori-Mittelwerte der räumlichen Effekte stark an den simulierten Daten orientiert. Die vorgegebene Struktur des jeweiligen Settings ist dahingegen in den insgesamt 40 Simulationen meist lediglich angedeutet wiederzuerkennen. Ein Vergleich der Schätzungen zwischen den Settings erweist sich im Folgenden trotzdem als sinnvoll. Eine detaillierte und umfassende Zusammenstellung der Grafiken aller simulierten Datensätze sowie sämtliche graphischen Aufbereitungen bezüglich der einzelnen Samplingbewertungen sind dem elektronischen Anhang zu entnehmen.

### Aufbereitung der geschätzten mittleren Gewichte

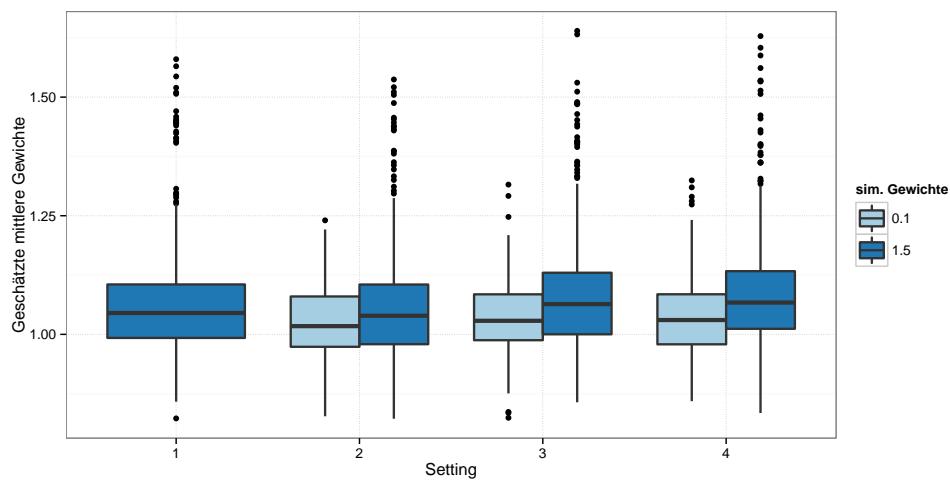
Die Schätzungen der Posteriori-Mittelwerte der einzelnen Nachbarschaftsgewichte ist von besonderem Interesse. Dabei gilt es zu klären, ob diese trotz einer verhältnismäßig uninformativen Datenstruktur sinnvoll geschätzt werden können. Die anschließenden deskriptiven Aufbereitungen sind hilfreich, um erste Erkenntnisse darüber zu gewinnen. Im Weiteren sollen die Verteilungen der jeweiligen Gruppierungen über einen nicht-parametrischen Wilcoxon-Rangsummen-Test miteinander verglichen und auf signifikante Lageverschiebungen untersucht werden. In anschließender Abbildung werden zunächst die geschätzten Posteriori-Mittelwerte der Nachbarschaftsgewichtungen gruppiert nach den vier verschiedenen Settings in Boxplots zusammengefasst dargestellt.



**Abbildung 49:** Geschätzte mittlere Nachbarschaftsgewichte gruppiert nach Settings

Da für jedes Setting insgesamt 10 simulierte Samplings mit jeweils 117 Gewichtungsparameter durchgeführt wurden, liegen jedem der vier Boxplots 1170 geschätzte mittlere Nachbarschaftsgewichte zugrunde. Für deren Veranschaulichung wird grundsätzlich angenommen, dass die vordefinierten Gewichtungsparameter der Nachbarschaftsmatrizen  $\mathbf{Q}_w^{(j)}$  für die 10 Simulationen einer Konfiguration gleich sind, weshalb bei einer korrekten Schätzung der Gewichte innerhalb eines Settings ähnliche Posteriori-Mittelwerte für die Nachbarschaftsbeziehungen zu erwarten wären. Im Gegensatz dazu sollte bei einem Vergleich zwischen den vier Grenzkonfigurationen ein Unterschied erkennbar werden,

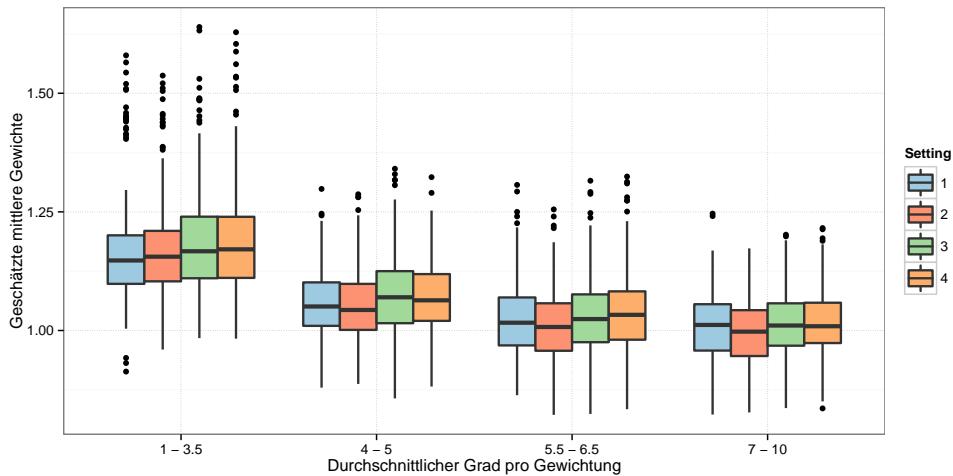
da für das erste Setting alle ursprünglichen Gewichte der Simulation bei 1.5 liegen und die Anzahl an grenzüberschreitenden Gewichtungen (versehen mit 0.1) über die Settings 2, 3 und 4 laufend zunimmt. Für Abbildung 49 ergibt die Analyse anhand des Wilcoxon-Rangsummen-Test, dass die Lage der Gewichtsverteilungen aus Setting 2 auf einem 5%-Niveau signifikant niedriger ist, als die Verteilung in allen anderen Settings. Gleichermaßen gilt für den Vergleich von Setting 1 zu Setting 4. Alle weiteren Betrachtungen haben keine signifikanten Unterschiede hervorgebracht, womit sich eher eine umgekehrte Bild abzeichnet, verglichen mit den ursprünglichen Erwartungen. Die Einbeziehung einer zusätzlichen Gruppierung der Settings nach den festgelegten Gewichtungen im Vorfeld der Simulationen zeigt wiederum eine eher zu erwartende Tendenz an.



**Abbildung 50:** Geschätzte mittlere Nachbarschaftsgewichte gruppiert nach Settings und ursprünglichen Gewichten der Simulation

Da für die erste Grenzkonfiguration alle Ausgangsgewichte bei 1.5 liegen, kann an dieser Stelle in Abbildung 50 keine Unterteilung erfolgen. Für die restlichen Settings setzen sich die 1170 Gewichtsmittelwerte jeweils aus der Aufteilung in die beiden Kategorien 1.5 und 0.1 (im Folgenden dargestellt in Klammern) als Gewicht der zugrundeliegenden Simulation zusammen. Es ergeben sich in diesem Zusammenhang die Fallzahlen 1070 (100) für Setting 2, 950 (220) für Setting 3 und 880 (290) für Setting 4. Die Analyse über den Wilcoxon-Rangsummen-Test bringt hervor, dass in allen Settings (Setting 1 ausgenommen) für die Verteilung der geschätzten mittleren Gewichte, mit einer ursprünglichen Nachbarschaftsbeziehung von 1.5, eine signifikant positive Lageverschiebung auf einem 5%-Niveau vorliegt, im Vergleich zu den Verteilungen bei einem Ausgangsgewicht von 0.1. Demzufolge weisen die Schätzungen eine sinnvolle Tendenz auf, die jedoch durch die anschließende Betrachtung an Aussagekraft einbüßt. In diesem Zusammenhang wurde eine kategoriale Größe gebildet, welche den durchschnittlichen Grad zweier benachbarter Beobachtungseinheiten der entsprechenden Gewichtung zuordnet. Der Grad einer Region ist diesbezüglich definiert als die jeweilige Anzahl an Nachbarn. Ausgerichtet an den 117 Einzelgewichten setzen sich die vier Kategorien dieser Variable in der Art zusammen, dass der Bereich von '1 bis 3.5' 19 Fälle, von '4 bis 5' 41 Fälle, von '5.5 bis 6.5' 37 Fälle und von '7 bis 10' 20 Fälle beinhaltet. Für

die 16 Boxplotzusammenfassungen der folgenden Abbildung müssen diese Fallzahlen jeweils verzehnfacht werden.



**Abbildung 51:** Geschätzte mittlere Nachbarschaftsgewichte gruppiert nach Kategorien des durchschnittlichen Grades pro Gewichtung und Settings

Für die Lagevergleiche der Verteilungen über den Wilcoxon-Rangsummen-Test, wurden ausschließlich die verschiedenen Kategorien innerhalb eines Settings berücksichtigt. Von Kategorie 1 bis hin zu Kategorie 4, können in diesem Kontext Lageverschiebungen auf einem Signifikanzniveau von 5% mit negativer Ausrichtung erkannt werden. Ausgenommen ist hiervon der Unterschied zwischen den Verteilungen der mittleren Gewichte von Kategorie 3 zu Kategorie 4 des ersten Settings. Dort liegt keine signifikante Verschiebung der Lage vor. Diese Erkenntnisse relativieren die vorherigen Schlüsse aus der Betrachtung von Abbildung 50, da scheinbar die Anzahl der Nachbarn einer Region den geschätzten Posteriori-Mittelwert einer Gewichtung beeinflussen kann. Die nähere Betrachtung des durchschnittlichen Grades pro mittlerem Gewicht, aufgesplittet nach den Settings und den Ausgangsgewichten der Simulation zeigt, dass keine einzige Nachbarschaft der untersten Kategorie '1 bis 3.5' mit 0.1 gewichtet wird. Der Anteil in der Kategorie '4 bis 5' ist zudem äußerst gering.

### 3.5.3 Bewertung der Schätzergebnisse

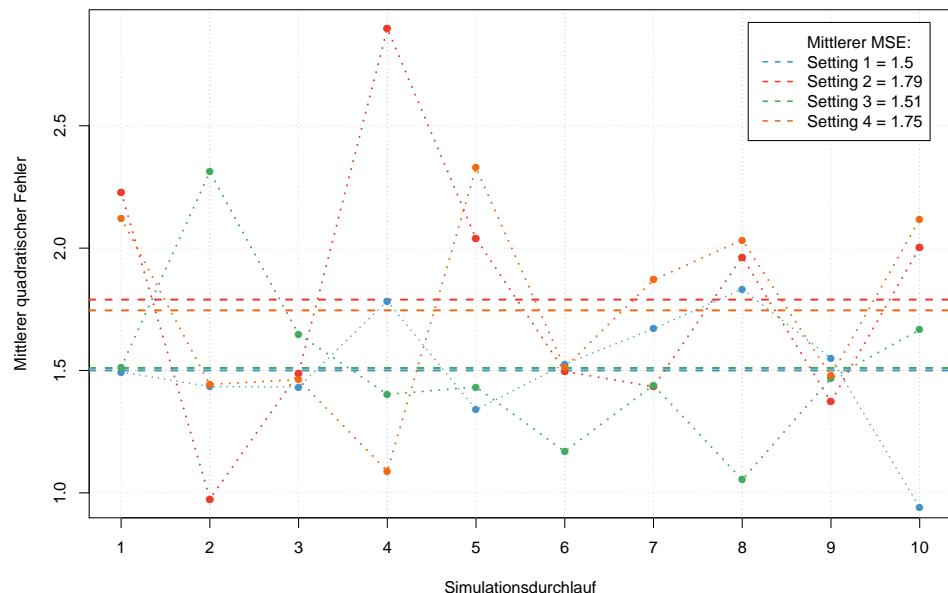
Der grundlegende Vorteil bei einem Sampling aus simulierten Daten besteht darin, dass die geschätzten Verteilungscharakteristika mit den ursprünglichen Parameterwerten verglichen werden können. Die Berechnung des mittleren quadratischen Fehlers (kurz: MSE (engl.)) kann in diesem Zusammenhang als geeignetes Maß für die Abweichung zwischen geschätztem und wahren Parameterwert betrachtet werden. Die durchschnittliche Abweichung eines kompletten Parametervektors  $\boldsymbol{\theta}$  setzt sich schließlich folgendermaßen zusammen:

$$MSE_{\boldsymbol{\theta}} = \frac{1}{n} \sum_{i=1}^n (\hat{\theta}_i - \theta_i)^2$$

Der Laufindex  $i = 1, \dots, n$  repräsentiert in diesem Kontext die jeweilige Anzahl der Parameter. Im Folgenden ist die Abweichung der zentrierten Summe der geschätzten Posteriori-Mittelwerte der räumlichen Effekte von der standardisierten Summe der dementsprechenden simulierten Werte von Interesse, wobei  $j$  das jeweilige Setting und  $l$  den jeweiligen Simulationsdurchlauf markiert.

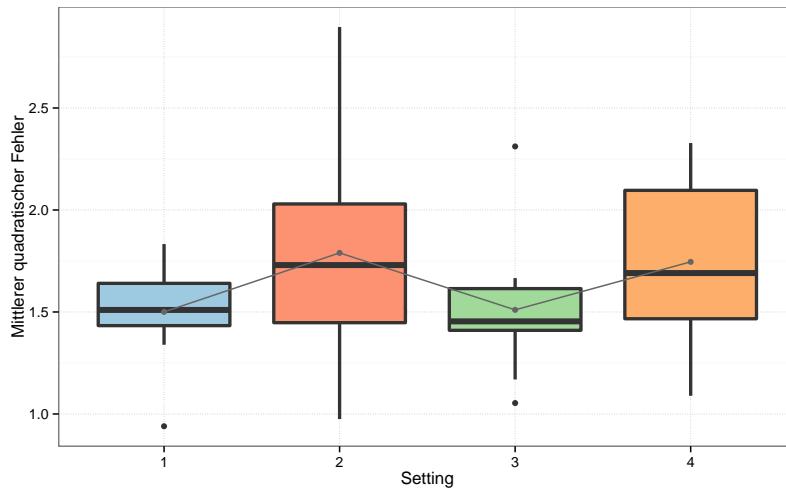
$$MSE_{\gamma_{\text{std}} + \alpha_{\text{std}}}^{(j,l)} = \frac{1}{53} \sum_{s=1}^{53} ((\hat{\gamma}_{\text{zent},s} + \hat{\alpha}_{\text{zent},s}) - (\gamma_{\text{std},s} + \alpha_{\text{std},s}))^2$$

Bei der Abtragung der berechneten Abweichungen für die einzelnen Simulationsdurchläufe zeigt sich ein stark vermischttes Bild ohne einer erkennbaren Tendenz. Abbildung 52 zeigt außerdem, dass die Durchschnittswerte innerhalb einer Grenzkonfiguration mit 1.50 für Setting 1, mit 1.79 für Setting 2, mit 1.51 für Setting 3 und mit 1.75 für Setting 4 in einem relativ engen Bereich liegen.



**Abbildung 52:**  $MSE^{(j,l)}$  zwischen der zentrierten Summe der geschätzten Posteriori-Mittelwerte und der standardisierten Summe der simulierten räumlichen Effekte

Für die jeweiligen Settings können zudem folgende Minimalwerte (Maximalwerte) beobachtet werden: Setting 1 mit 0.94 (1.83), Setting 2 mit 0.98 (2.90), Setting 3 mit 1.05 (2.31) und Setting 4 mit 1.09 (2.33). In Abbildung 53 werden die wichtigsten Werte in Boxplots zusammengefasst. Ein paarweiser Vergleich der Settings über einen Rangsummen-Test ergibt, dass auf einem 5%-Niveau keine signifikanten Lageverschiebungen der jeweiligen Verteilungen vorliegen.

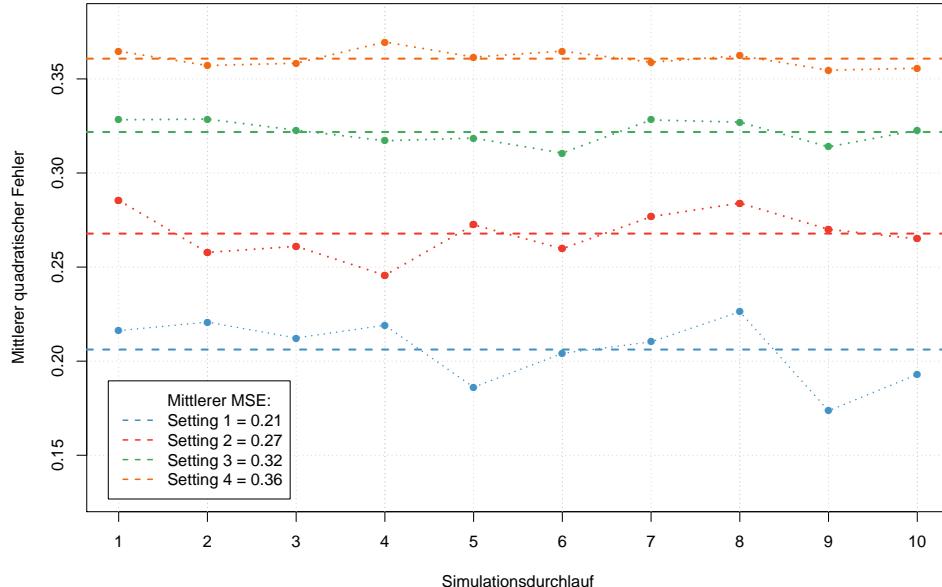


**Abbildung 53:**  $MSE^{(j,l)}$  der räumlichen Effekte gruppiert nach Settings – Mittelwert in grau

Im Weiteren steht der Fehler bei der Schätzung der mittleren Gewichtungen im Verhältnis zu den Ausgangsgewichten der Simulation im Fokus der Betrachtungen. Aus der jeweiligen Nachbarschaftsmatrix  $\mathbf{Q}_w^{(j)}$  bzw.  $\hat{\mathbf{Q}}_w$  wird diesbezüglich ein Vektor  $\mathbf{w}$  gebildet, welcher die entsprechenden Gewichte  $g = 1, \dots, 117$  beinhaltet. Daraufhin setzt sich der MSE folgendermaßen zusammen:

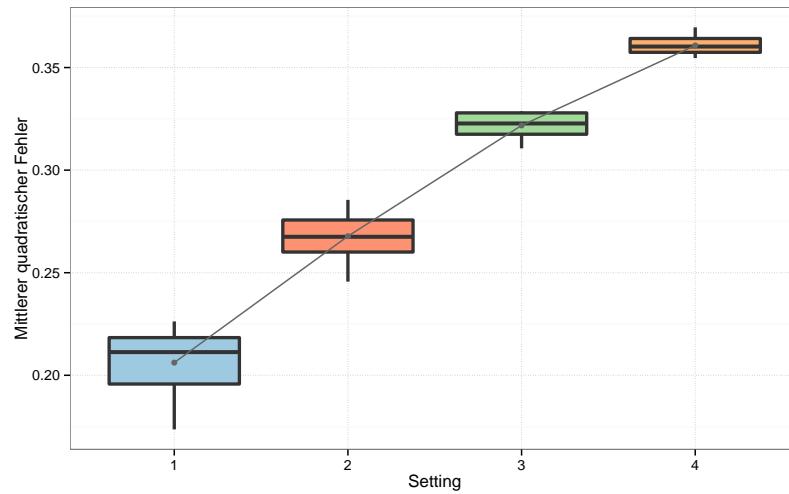
$$MSE_{\mathbf{w}}^{(j,l)} = \frac{1}{117} \sum_{g=1}^{117} (\hat{w}_g - w_g)^2$$

Abbildung 54 zeigt einen ersten Unterschied des  $MSE_{\mathbf{w}}^{(j,l)}$  zwischen den Settings auf.



**Abbildung 54:**  $MSE^{(j,l)}$  zwischen den geschätzten Posteriori-Mittelwerten der Nachbarschaftsgewichte und den Ausgangsgewichten der Datensimulierung

Die ansteigenden Mittelwerte von Setting 1 mit 0.21, Setting 2 mit 0.27, Setting 3 mit 0.32 zu Setting 4 mit 0.36 geben ein erstes Anzeichen dafür, dass Unterschiede für die verschiedenen Ausgangssituationen vorliegen. Die Minimalwerte (Maximalwerte) befinden sich dabei ebenso auf anderen Niveaus mit 0.17 (0.23) für Setting 1, 0.25 (0.29) für Setting 2, 0.31 (0.33) Setting 3 und 0.35 (0.37) für Setting 4. Bei einem paarweisen Vergleich der Verteilungen des MSE, unter Berücksichtigung der Settings über den Rangsummen-Test, bestätigt dieser signifikante Lageverschiebungen auf einem 5%-Niveau. Die Zusammenfassung über die Boxplots verdeutlicht diesen Sachverhalt zusätzlich.



**Abbildung 55:**  $MSE^{(j,l)}$  der Nachbarschaftsgewichte gruppiert nach Settings – Mittelwert in grau

Die deskriptive Aufbereitung der geschätzten Gewichte gruppiert nach den Settings in Abbildung 49 hat gezeigt, dass nahezu keine Unterschiede für die verschiedenen Ausgangssituationen vorliegen. Daraus kann für die Interpretation der Abbildung 55 möglicherweise geschlossen werden, dass der Fehler bei der Schätzung durch die Festlegung der Ausgangsgewichte bestimmt wird. Ebenso wie der MSE nimmt auch die Anzahl an Nachbarschaften mit einer ursprünglichen Gewichtung von 0.1 über die vier Settings hinweg zu.

### 3.6 Anwendung auf empirische Daten mit Gewichtsadaption

Die gewonnenen Erkenntnisse aus den simulationsbasierten Schätzungen über die Festsetzung einzelner Parameter im Vorfeld des Samplings, wird im Folgenden für die Anwendung auf empirische Daten, inklusive der Einbeziehung von anpassungsfähigen Gewichtungsparametern, übernommen. Demnach gehen die beiden Hyperparameter mit  $\kappa = 0.001$  und  $\tau = 1$  sowie der Parameter der Priori-Verteilung der Gewichte mit  $\nu = 1$  in den Hybrid-Sampler mit ein. Zudem wird im Weiteren auf die Berücksichtigung einer Kovariablen verzichtet. Als Folge der adaptiven Schätzung der Gewichte und der enormen Vergrößerung des Parametervektors treten diesbezüglich ähnliche Identifizierbarkeitsprobleme auf, wie in Bezug auf den Intercept (vgl. Unterkapitel 3.4.1). Bevor

jedoch die Schätzergebnisse aus den schottischen und den deutschen Krebsdaten aufbereitet werden, soll eine Konvergenzanalyse nach Gelman und Rubin anhand des schottischen Samplings durchgeführt werden.

### 3.6.1 Konvergenzanalyse nach Gelman und Rubin

Bei der Bewertung des Samplings aus einem MCMC Algorithmus, steht die Betrachtung der Konvergenz der gezogenen Markov-Ketten im Vordergrund. Um diese zu analysieren, gibt es neben den gezeigten grafischen Diagnosemöglichkeiten in Unterkapitel 3.4.2 eine Vielzahl an statistischen Kennzahlen und Methoden. Cowles und Carlin (1996) geben diesbezüglich in ihrem Paper über die MCMC Konvergenzdiagnose einen umfassenden Überblick über die verschiedenen Möglichkeiten. Eine der gängigsten Methoden ist in diesem Zusammenhang die Konvergenzanalyse nach Gelman und Rubin, die in dem R-Paket `coda` vollständig implementiert ist (vgl. Plummer *et al.*, 2006).

### Theoretische Grundlage

Unter der allgemeinen Annahme, dass die Markov-Kette eines Parameters gegen dessen entsprechende Randverteilung konvergiert, sollten die Ergebnisse durch eine Wiederholung des Samplings bei gleichen Rahmenbedingungen näherungsweise repliziert werden können. Aus diesem Grund schlagen Gelman und Rubin für die Überprüfung der Konvergenz vor, dass  $j = 1, \dots, m$  wiederholte Durchläufe des Samplings mit  $i = 1, \dots, n$  Iterationen durchgeführt werden, jeweils mit unterschiedlichen Startwerten für die relevanten Parameter. Als Kennzahl für die Güte der Konvergenz entwickelten sie diesbezüglich den sogenannten „Potential Scale Reduction Factor“ (im Folgenden repräsentiert durch  $R$ ). Um diesen berechnen zu können, wird die Varianz innerhalb der einzelnen Markov-Ketten (engl. *within-sequence variance*) sowie die Varianz zwischen den Markov-Ketten der verschiedenen Durchläufe (engl. *between-sequence variance*) benötigt. Die folgende Aufbereitung im Rahmen der Zusammensetzung der Kennzahl  $R$  basiert sowohl auf den Ausführungen nach Gelman und Rubin (1992) als auch auf der anschaulichen Zusammenfassung von Lam (2008).

Die „within“-Varianz für einen Parameter  $\theta$  stellt sich als die gemittelte Summe der Varianzen  $s_j^2$  innerhalb der  $m$  Markov-Ketten dar. Für die Zusammensetzung von  $W_\theta$  bedeutet das folgendes:

$$W_\theta = \frac{1}{m} \sum_{j=1}^m s_{\theta_j}^2, \quad \text{wobei} \quad s_{\theta_j}^2 = \frac{1}{n-1} \sum_{i=1}^n (\theta_{ij} - \bar{\theta}_j)^2$$

Da jede Parameterkette auf der Ziehung von  $n$  Zufallsziehungen basiert, wird die „between“-Varianz zwischen den  $m$  Samplingketten mit dem jeweiligen Iterationsumfang multipliziert. Für die Berechnung von  $B_\theta$  sind dabei der globale Mittelwert über alle wiederholten Ketten hinweg sowie der Mittelwerte jeder einzelnen Kette relevant.

$$B_\theta = \frac{n}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\bar{\theta}})^2, \quad \text{wobei} \quad \bar{\bar{\theta}} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_j$$

Aus den beiden Varianzen  $W_\theta$  und  $B_\theta$  kann schließlich eine gewichtete Gesamtvarianz als Hauptbestandteil von  $\widehat{R}_\theta$  geschätzt werden.

$$\widehat{Var}_\theta = \left(1 - \frac{1}{n}\right)W_\theta + \frac{1}{n}B_\theta \quad \Rightarrow \quad \widehat{R}_\theta = \sqrt{\frac{\widehat{Var}_\theta}{W_\theta}}$$

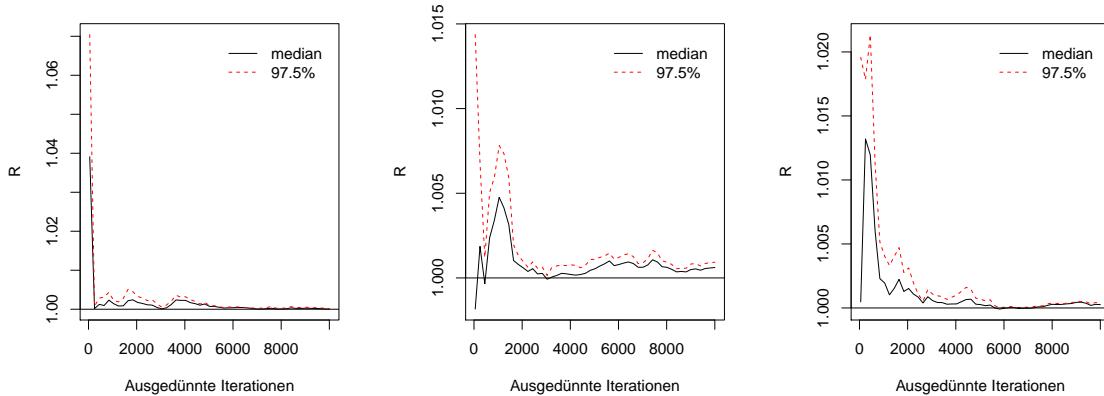
Der daraus resultierende Faktor ist folgendermaßen zu interpretieren:

„When the potential scale reduction is near 1, we conclude that each set of the  $m$  sets of  $n$  simulated values is close to the target distribution“ (Gelman und Rubin, 1992, S. 465).

Demnach kann bei einem berechneten  $\widehat{R}_\theta$  nahe bei 1 davon ausgegangen werden, dass die Markov-Ketten konvergieren und die gezogenen Parameterzustände nahezu Zufallsziehungen aus der wahren Randverteilung entsprechen. Allerdings ist zu berücksichtigen, dass bereits bei einem  $\widehat{R}_\theta$  höher als 1.1 eine Vergrößerung der Anzahl an Iterationen in Erwägung gezogen werden sollte. Ferner gilt es zu beachten, dass die Konvergenzanalyse für jeden einzelnen Parameter eines Modells durchgeführt werden muss, um den gewählten Samplingumfang verifizieren zu können (vgl. Lam, 2008, S. 23).

## Anwendung auf das schottische Sampling

Die beschriebene Konvergenzanalyse setzt voraus, dass wiederholte Samplings mit verschiedenen Parameterstartwerten durchgeführt werden. Da im Rahmen dieser Arbeit die Schätzung von adaptiven Gewichten im Fokus der Betrachtungen stehen, soll die zufällige Startwertgenerierung der Gewichte dieser Anforderung genügen (vgl. hierzu Vorbereitungspunkt 4c der Skizze des Hybrid-Samplers in Abschnitt 3.3.3 auf Seite 69). Das Sampling wird im Weiteren insgesamt fünf Mal wiederholt, jeweils mit identischen Rahmenbedingungen von 100 000 Iterationen, einem Ausdünnungsfaktor von 10 sowie den eingangs beschriebenen Parameterfestsetzungen. Neben der Betrachtung der Gewichtungsparameter wird zusätzlich auch die Konvergenz der strukturierten und der unstrukturierten Effekte überprüft. Die Funktionen `gelman.diag` und `gelman.plot` des `coda`-Packages erstellen in diesem Zusammenhang automatische Zusammenfassungen bzw. Ergebnisplots der jeweiligen Konvergenzdiagnose. Dabei wird für alle 117 Gewichtungsparameter der Faktor  $R = 1$  ausgewiesen, wovon in Abbildung 56 auszugsweise die Verläufe dreier Parametervergleiche, jeweils durch den Median (schwarz) und das 97.5%-Quantil (rot), abgetragen werden. Als Schlussfolgerung kann festgehalten werden, dass die Markov-Ketten der Nachbarschaftsgewichte gegen deren Randverteilungen konvergieren. Identische Ergebnisse werden bei der Betrachtung der Parameter der räumlichen Effekte erzielt. Sämtliche Analysegrafiken bezüglich der Konvergenzdiagnose sind außerdem dem elektronischen Anhang beigefügt.

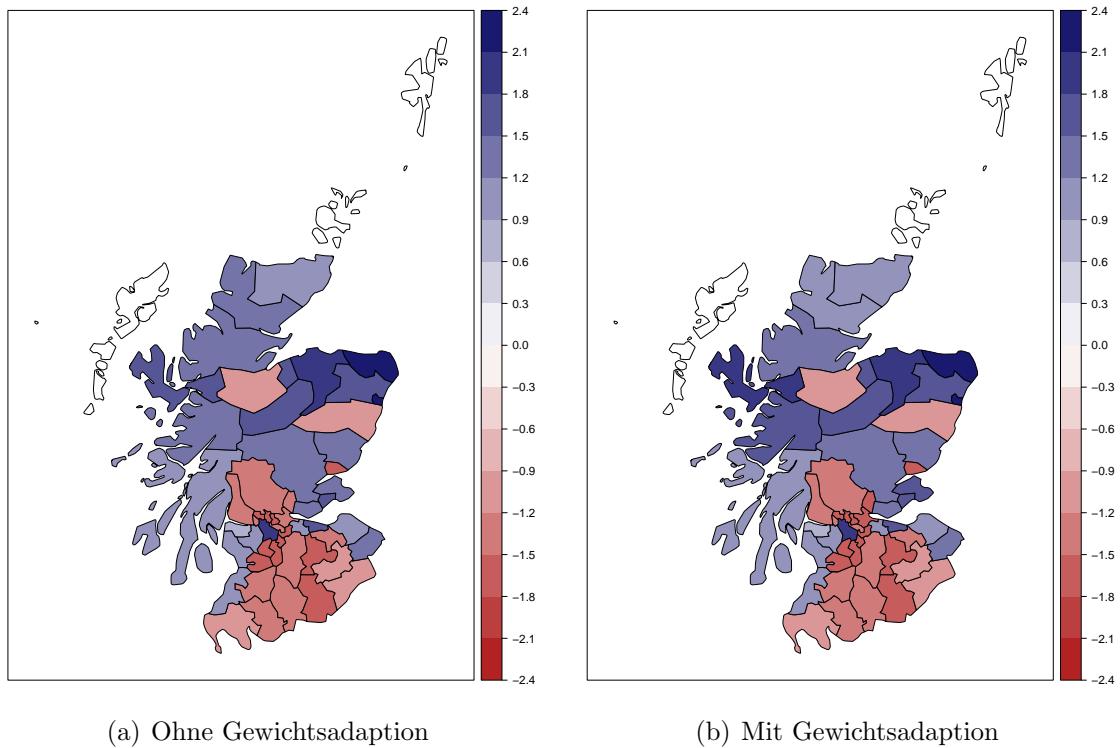


**Abbildung 56:** Auszugsweise Darstellung der Entwicklung des Faktors  $R$  in Bezug auf die Konvergenzanalyse der Gewichtungsparameter

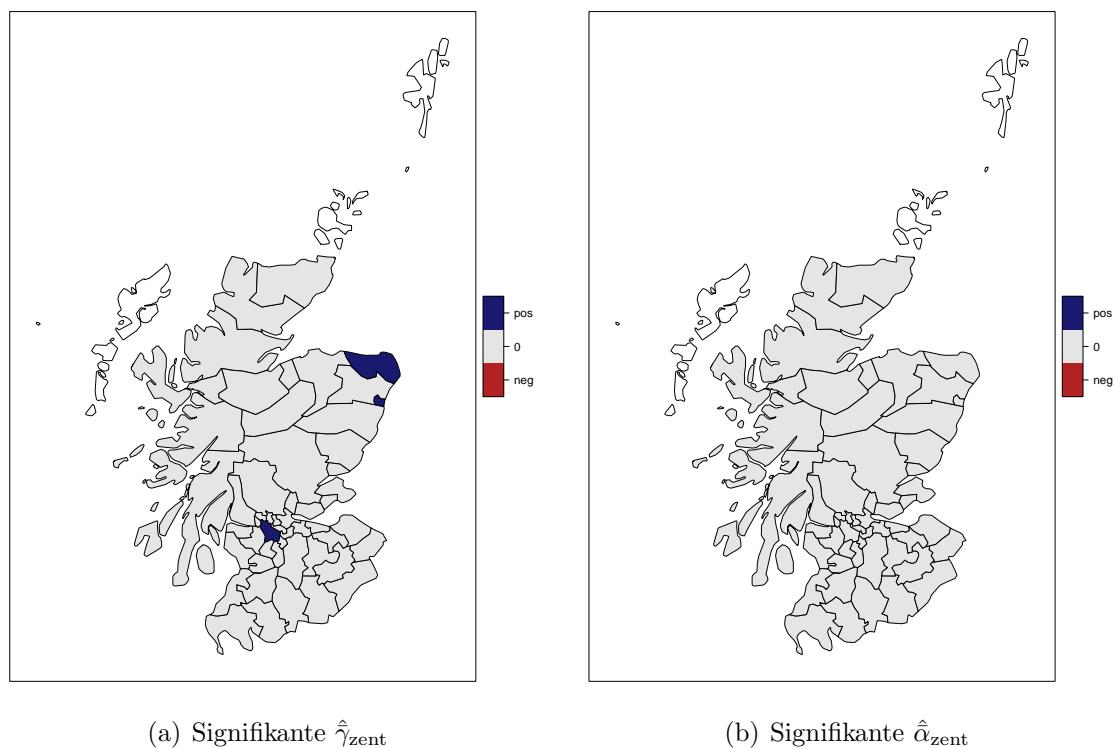
### 3.6.2 Schätzergebnisse aus den schottischen Lippenkrebsdaten

Unter der Berücksichtigung der genannten Paramtereinschränkungen müssen, aufgrund der Hinzunahme von 117 weiteren Gewichtungsparametern, bei der folgenden Betrachtung der 53 schottischen Distrikte zusammengefasst 223 Parametersamplings für die Schätzung der jeweiligen Verteilungscharakteristika durchgeführt werden. In Hinblick auf die vergleichsweise geringe Anzahl an Beobachtungen werden die folgenden Aufbereitungen zeigen, dass die Schätzergebnisse der räumlichen Effekte trotzdem kaum beeinflusst werden. In diesem Zusammenhang ist anzumerken, dass die Bewertung der einzelnen Samplingverläufe hinsichtlich ihrer Mischung und Konvergenz vergleichbare Ergebnisse hervorbringen, wie die Aufbereitung der schottischen Markov-Ketten ohne Gewichtsadaption. Aus diesem Grund wird an dieser Stelle auf eine erneute Veranschaulichung verzichtet. Für eine nähere Betrachtung einzelner Diagnoseplots, kann der elektronische Anhang eingesehen werden.

Da für den Koeffizienten der Kovariablen AFF ohnehin ein sehr niedriger Posteriori-Mittelwert geschätzt wurde, verändert sich die Schätzung des Intercepts  $\hat{\mu}$  nur minimal von 1.36 auf 1.31 im Vergleich zu dem Modell ohne anpassungsfähigen Gewichtungsparametern. Für die räumlichen Effekte ist festzuhalten, dass aufgrund der Festlegung der Hyperparameter auf  $\tau = 1$  und  $\kappa = 0.001$  die unstrukturierten Effekte extrem klein werden und der Großteil der räumlichen Effekte über das GMZF geschätzt wird. Im Rahmen der Gegenüberstellung in Abbildung 57 ist in diesem Zusammenhang zu erkennen, dass sich wie bereits angedeutet die Schätzungen trotz der Berücksichtigung variabler Gewichte kaum unterscheiden. Im Weiteren werden in Abbildung 58 die signifikanten räumlichen Effekte, aufgeteilt in strukturiert und unstrukturiert, visualisiert. Dabei ist zu erkennen, dass sich bei den zufälligen Effekten keine Beobachtungseinheit auf einem 5%-Niveau signifikant vom landesweiten Durchschnitt unterscheidet. Bei den strukturierten Effekten treten dagegen Distrikte mit einem signifikant positiven Unterschied in den Vordergrund, die bei der Modellierung ohne Gewichtsadaption noch auf beide Effekte aufgeteilt waren.

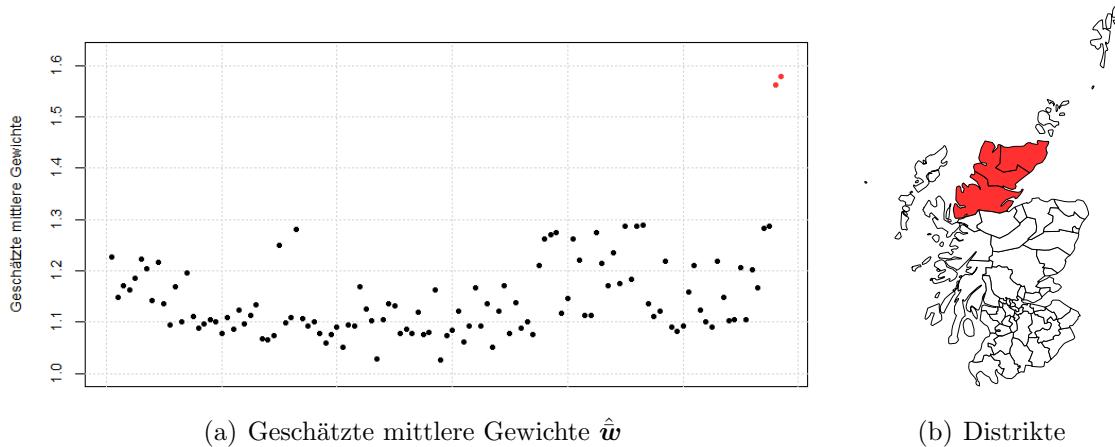


**Abbildung 57:** Zentrierte Summe der mittleren räumlichen Effekte im Vergleich



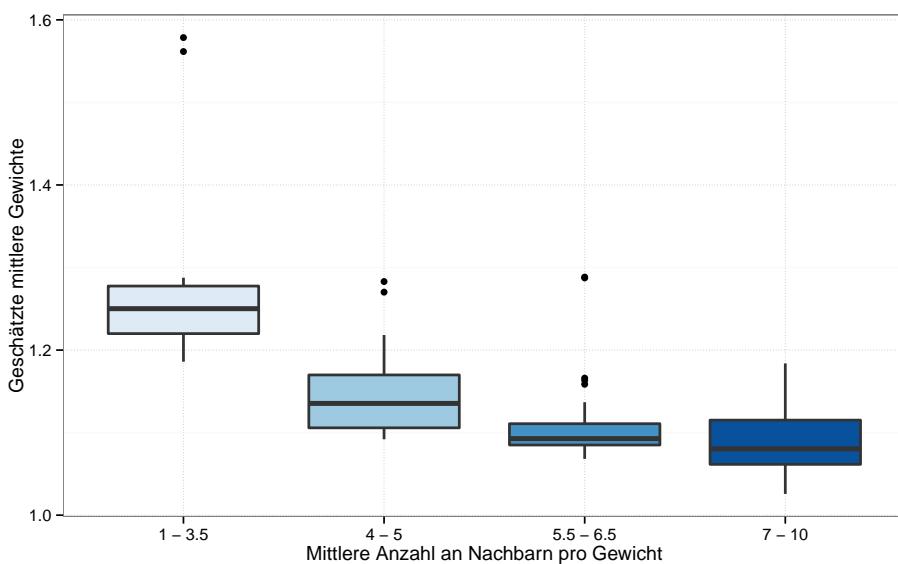
**Abbildung 58:** Auf einem 5%-Niveau signifikant von Null verschiedene räumliche Effekte aus den schottischen Lippenkrebsdaten – mit positiver (blau) oder negativer (rot) Ausrichtung

Die Abtragung der geschätzten Posteriori-Mittelwerte der Nachbarschaftsgewichte in Abbildung 59 Grafik (a) zeigt, von zwei Ausreißern abgesehen, ein gemischtes Bild. Bei diesen beiden mittleren Gewichten über 1.5, handelt es sich um die Nachbarschaften der drei Distrikte ganz im Norden der Insel (siehe Grafik (b)).



**Abbildung 59:** Auffällige Nachbarschaftsbeziehungen anhand der geschätzten Posteriori-Mittelwerte der Gewichtungen

Die Vermutung, welche aus der Analyse der geschätzten mittleren Gewichte basierend auf simulierten Daten in Unterkapitel 3.5.2 entstanden ist, kann durch die Gruppierung nach dem durchschnittlichen Grad einer Gewichtung bzw. dessen vier Kategorien in Abbildung 60 bestätigt werden. Demzufolge wird für Nachbarschaftsbeziehungen mit einem niedrigeren durchschnittlichen Grad tendenziell eine höhere mittlere Gewichtung geschätzt.

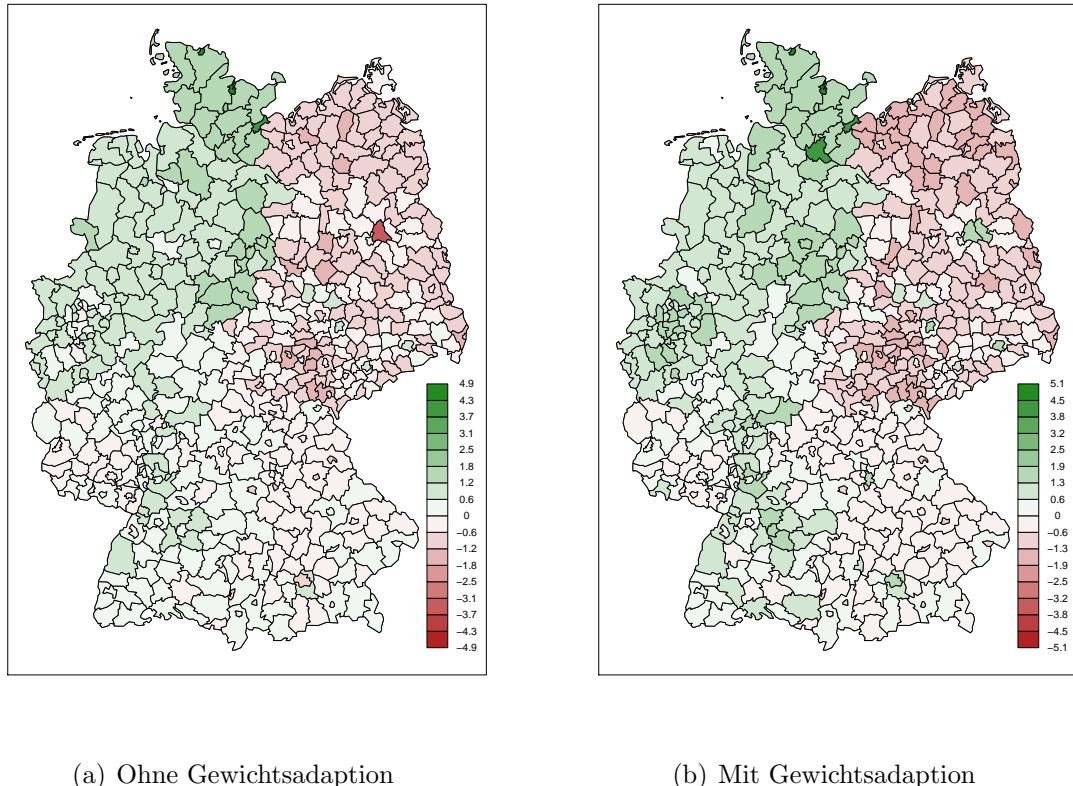


**Abbildung 60:** Geschätzte mittlere Nachbarschaftsgewichte gruppiert nach Kategorien des durchschnittlichen Grades pro Gewichtung

Eine erneute Durchführung des Wilcoxon-Rangsummen-Test bestätigt diesbezüglich eine signifikant negative Lageverschiebung der jeweiligen Verteilungen im Verlauf von Kategorie 1 hin zu Kategorie 4 auf einem 5%-Niveau. Ausgenommen der Vergleich zwischen Kategorie 3 und 4.

### 3.6.3 Schätzergebnisse aus den deutschen Kehlkopfkrebsdaten

Für eine Schätzung aus den deutschen Kehlkopfkrebsdaten mit der Berücksichtigung der Nachbarschaftsstruktur von 544 Landkreisen, erweitert sich die Anzahl der für das Sampling relevanten Parameter um 1416 Gewichtungen auf 2504. Hinsichtlich dieser umfangreichen Nachbarschaftsbeziehungen und einer dementsprechend großen Präzisionsmatrix, benötigt der Hybrid-Sampler wesentlich länger für die Durchführung einzelner Iterationen. Unter diesem zeitlichen Aspekt soll ein Iterationsumfang von 25 000 bei einem Ausdünnungsfaktor von 5 den weiteren Anforderungen genügen. Die gewählte Burn-In Phase liegt dabei unverändert bei 10 000. In Bezug auf die Samplinganpassungen im Vorfeld ist lediglich zu beachten, dass wiederum die ersten fünf Vorschläge für die Prädiktorwerte mit 100% akzeptiert werden. Alle a priori Festsetzungen von Parametern werden zudem aus den bisherigen Modellierungen, inklusive anpassungsfähigen Gewichten, übernommen.



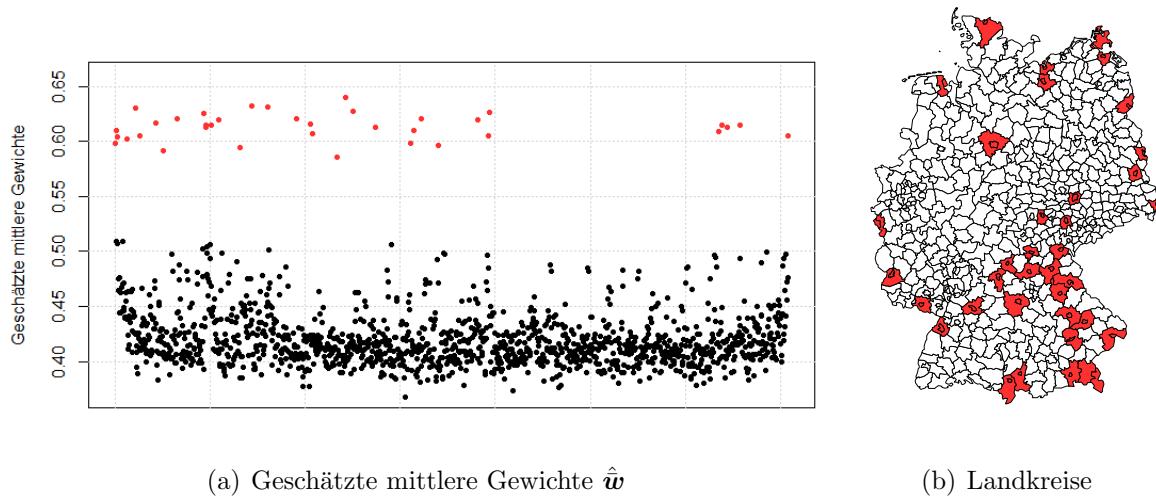
**Abbildung 61:** Zentrierte Summe der mittleren räumlichen Effekte im Vergleich

Analog zu den schottischen Betrachtungen ergeben sich für die Samplingverläufe der einzelnen Parameter, im Vergleich zu der deutschen Modellierung ohne Gewichtsadap-

tionen, ähnliche Bilder in Bezug auf Mixing und Konvergenz, weshalb diese ausschließlich im elektronischen Anhang hinterlegt sind.

Da die Schätzung des Posteriori-Mittelwerts  $\beta$  für die deutsche Modellierung ohne einer Gewichtsadaption im Vergleich zu der schottischen Schätzung etwas größer ausgefallen ist, liegt eine höhere Veränderung bei der Berechnung des Intercepts  $\hat{\mu}$  von -1.76 auf -1.36 in positiver Richtung vor. Für die Summe der räumlichen Effekte, aufbereitet in Abbildung 61, zeigen sich in der Gegenüberstellung der Schätzergebnisse aus der Modellierung ohne bzw. mit Gewichtsadaptionen, dass ähnliche Ergebnisse erzielt werden konnten. Alleinig für die Schätzungen der Landkreise um Berlin, Hamburg und München können Abweichungen erkannt werden. In Hinblick auf die Datenlage scheint dort die Anpassung ohne der Berücksichtigung einer Kovariable besser zu funktionieren. Die Betrachtung signifikanter Landkreise ergibt außerdem, dass lediglich für drei strukturierte Effekte ein von Null verschiedener Unterschied in positiver Richtung vorliegt. Betroffen sind die Landkreise der Städte Hamburg und Kiel sowie eine weitere Region in Schleswig-Holstein.

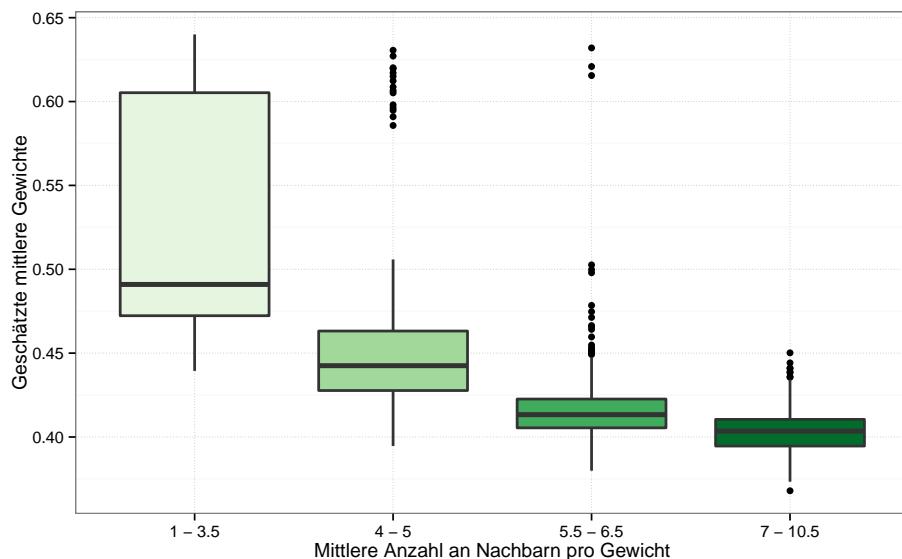
Bei der ersten Betrachtung der Posteriori-Mittelwerte zeichnet sich eine Bestätigung der bisherigen Vermutung ab, dass der Grad einer Beobachtungseinheit die jeweiligen geschätzten mittleren Nachbarschaftsgewichte in gewisser Weise beeinflussen kann. In Abbildung 62 Grafik (a) werden diesbezüglich alle 1416 geschätzten mittleren Nachbarschaftsgewichtungen abgetragen, wobei die 36 auffallend höheren Schätzungen rot eingefärbt sind. Eine Markierung der betroffenen Landkreise (siehe Grafik (b)) lässt wiederum erkennen, dass für Beobachtungseinheiten mit einem niedrigeren Grad tendenziell höhere mittlere Gewichte geschätzt werden.



**Abbildung 62:** Auffällige Nachbarschaftsbeziehungen anhand der geschätzten Posteriori-Mittelwerte der Gewichtungen

Die grafische Veranschaulichung in Abbildung 63 zeigt die geschätzten Posteriori-Mittelwerte der Nachbarschaftsgewichte gruppiert nach den Kategorien des zugehörigen

durchschnittlichen Grades. Der paarweise Vergleich über einen Rangsummen-Test bestätigt wiederholt die genannte Vermutung, da absteigend von Kategorie 1 zu Kategorie 4 auf einem 5%-Niveau signifikante Lageverschiebungen der Verteilungen der jeweiligen mittleren geschätzten Gewichte vorliegen.



**Abbildung 63:** Geschätzte mittlere Nachbarschaftsgewichte gruppiert nach Kategorien des durchschnittlichen Grades pro Gewichtung

## 4 Abschließende Bemerkungen

Im Folgenden werden die wichtigsten Ergebnisse der durchgeführten Schätzungen wiederholt und die zentralen Erkenntnisse zusammengefasst. Daran anschließend sollen die Grenzen der vorliegenden Analysen aufgezeigt werden, wodurch weiterführende bzw. andersartige Ansätze motiviert werden können.

### 4.1 Zusammenfassung der wichtigsten Ergebnisse

Die adaptive Schätzung der Gewichtungsparameter eines GMZF ist als zentrale Zielsetzung der vorgenommenen Untersuchungen zu betrachten. Die durchgeführten Analysen haben diesbezüglich gezeigt, dass a priori verhältnismäßig starke Annahmen getroffen werden müssen, wie beispielsweise die Festsetzung der Hyperparameter  $\kappa$  und  $\tau$ , um das Sampling der Gewichtungsparameter innerhalb eines sinnvollen Wertebereichs zu gewährleisten. Bei der Bewertung der Schätzergebnisse ohne einer Gewichtsadaption lässt sich im Vergleich zu dem Sampling mit Gewichtsadaption erkennen, dass diese zusätzliche Berücksichtigung die Schätzungen der Verteilungscharakteristika anderer Parameter, wie beispielsweise den räumlichen Effekten, nicht sonderlich beeinflusst. Im Fokus der Analysen steht außerdem abzuwägen, inwiefern die geschätzten Verteilungscharakteristika der Gewichtungsparameter die tatsächliche Stärke einer nachbarschaftlichen Beziehung zweier Beobachtungseinheiten repräsentieren. Die Analyse der geschätzten mittleren Gewichte lässt in diesem Zusammenhang deutlich erkennen,

dass deren Samplings mit der durchschnittlichen Anzahl an Nachbarn der betroffenen Beobachtungseinheiten zusammenhängen. Diese Erkenntnis kann sowohl aus der Betrachtung des Samplings aus simulierten Daten sowie auf Basis der schottischen und deutschen Krebsdaten gezogen werden.

## 4.2 Grenzen der Analyse und Ausblick

Ein einschränkender Faktor der vorliegenden Analysen ist die Schätzung einer großen Anzahl an Parametern aus einer verhältnismäßig uninformativen Datenlage. Als Konsequenz daraus sind die verwendeten Modellierungen äußerst sensitiv bezüglich der festgesetzten Hyperparameter. Bereits eine verhältnismäßig kleine Abänderung von  $\tau$  oder  $\kappa$  führt zu deutlich anderen geschätzten Verteilungscharakteristika der Gewichte. In der Regel befinden sich diese dann allesamt in einem Bereich Nahe bei Null. Allerdings ist „die mögliche Abhängigkeit von der Wahl der Hyperparameter [...] kein Nachteil der vollen Bayes-Inferenz“ (Fahrmeir *et al.*, 2009, S. 429).

Die Betrachtung der Samplingpfade für die Prädiktorenwerte hat außerdem ergeben, dass die Akzeptanz neuer Zustände über die durchgeführte Taylorentwicklung für einzelne Beobachtungseinheiten zum Teil sehr gering ausfällt. Für die komplexere Nachbarschaftsstruktur der deutschen Landkreise tritt diese Problematik im Vergleich zu den schottischen Distrikten sogar noch häufiger auf. Die Folge daraus ist, dass der MCMC Algorithmus, insbesondere der Metropolis-Hastings-Schritt, für die jeweilige Beobachtungseinheit nicht in vollem Umfang funktioniert. Eine alternative Herangehensweise ist die Verwendung einer sogenannten Random Walk Vorschlagsdichte in Verbindung mit dem Sampling des Prädiktors. Diese ermöglicht ein individuelles Tuning der Akzeptanzwahrscheinlichkeiten für jeden einzelnen Prädiktorwert. Limitierender Faktor kann in diesem Zusammenhang jedoch die Verlängerung der Laufzeit sein. Diese kann eine entscheidende Rolle spielen, bedenkt man die Samplingdauer von 70 Stunden für die Durchführung von 25 000 Iterationen im Rahmen des Samplings aus den deutschen Daten bei einer Berücksichtigung von adaptiven Gewichten.

Alternativ zu der verwendeten Gamma-Verteilung als Grundlage für die Priori-Dichte der Gewichtungsparameter schlagen Brezger *et al.* (2007) im Rahmen der Modellierung ihrer Brain Mapping Problematik die Annahme einer zusätzlichen, *a priori* festgelegten, räumlichen Struktur der Gewichte vor, modelliert über ein weiteres GMZF. Trotz der ansteigenden Komplexität der Modellstruktur kann es sinnvoll sein einen solchen Ansatz zu verwenden, wenn detailliertes Priori-Wissen über die strukturelle Stärke von nachbarschaftlichen Beziehungen vorliegt. Die genannten Ansätze können als Anregung für weiterführende Analysen und Forschungen in Bezug auf ein adaptives Disease Mapping aufgefasst werden.

## Abbildungsverzeichnis

1	Konjugierte Verteilungen . . . . .	12
2	Stetige räumliche Daten: Schweizer Regenfallmengen . . . . .	22
3	Diskrete räumliche Daten: Lippenkrebs in Schottland nach Distrikten .	23
4	Reguläres vs. irreguläres Gitter . . . . .	24
5	Nachbarschaften 1. Ordnung . . . . .	27
6	Nachbarschaften höherer Ordnung . . . . .	27
7	Ungerichteter Graph . . . . .	29
8	Graph der 16 Bundesländer von Deutschland . . . . .	29
9	Linearer Graph vs. Nachbarschaftsgraph . . . . .	32
10	Paarweise räumliche Markov-Eigenschaft . . . . .	33
11	Lokale räumliche Markov-Eigenschaft . . . . .	34
12	Globale räumliche Markov-Eigenschaft . . . . .	34
13	Karte der deutschen Bundesländer mit Nachbarschaftsgraph . . . . .	41
14	Umriss der 56 schottischen Distriktgrenzen . . . . .	47
15	Beobachtete Sterbefälle (schottische Lippenkrebsdaten) . . . . .	48
16	Erwartete Sterbefälle (schottische Lippenkrebsdaten) . . . . .	48
17	Zusätzliche Merkmale (schottische Lippenkrebsdaten) . . . . .	49
18	Direkte Nachbarschaftsstruktur von Schottland . . . . .	50
19	Umriss der 544 deutschen Landkreisgrenzen . . . . .	51
20	Beobachtete Sterbefälle (deutsche Kehlkopfkrebsdaten) . . . . .	52
21	Erwartete Sterbefälle (deutsche Kehlkopfkrebsdaten) . . . . .	52
22	Zusätzliche Merkmale (deutsche Kehlkopfkrebsdaten) . . . . .	53
23	Direkte Nachbarschaftsstruktur in Deutschland . . . . .	54
24	Schottische Nachbarschaftsmatrix: ursprünglich und reduziert . . . . .	75
25	Deutsche Nachbarschaftsmatrix: ursprünglich und reduziert . . . . .	76
26	Identifizierbarkeit: Intercept vs. räumlicher Effekt . . . . .	78
27	Beispiele für schlechtes/gutes Mixing . . . . .	79
28	Mittelwertsentwicklung bei schlechtem/gutem Mixing . . . . .	80
29	Beispielhaft Autokorrelationsfunktionen . . . . .	81
30	Samplingaufbereitung von $\beta$ (schottische Lippenkrebsdaten o. G.) . . .	83
31	Samplingaufbereitung von $\kappa$ (schottische Lippenkrebsdaten o. G.) . . .	83
32	Geschätzte $\hat{\gamma}_{\text{zent}}$ und $\hat{\alpha}_{\text{zent}}$ (schottische Lippenkrebsdaten o. G.) . . .	84
33	Signifikante $\hat{\gamma}_{\text{zent}}$ und $\hat{\alpha}_{\text{zent}}$ (schottische Lippenkrebsdaten o. G.) . . .	85
34	Akzeptanzwahrscheinlichkeit von $\eta$ (schottische Lippenkrebsdaten o. G.)	86
35	Samplingaufbereitung von $\eta_2$ (schottische Lippenkrebsdaten o. G.) . . .	87
36	$\exp(\hat{\eta})$ vs. beobachtete schottische Lippenkrebsdaten (o. G.) . . . .	88
37	Akzeptanzwahrscheinlichkeit von $\eta$ (deutsche Kehlkopfkrebsdaten o. G.)	89
38	$\exp(\hat{\eta})$ vs. beobachtete deutsche Kehlkopfkrebsdaten (o. G.) . . . .	90
39	Samplingaufbereitung von $\beta$ (deutsche Kehlkopfkrebsdaten o. G.) . . .	91
40	Samplingaufbereitung von $\kappa$ (deutsche Kehlkopfkrebsdaten o. G.) . . .	92
41	Geschätzte $\hat{\gamma}_{\text{zent}}$ und $\hat{\alpha}_{\text{zent}}$ (deutsche Kehlkopfkrebsdaten o. G.) . . .	93
42	Signifikante $\hat{\gamma}_{\text{zent}}$ und $\hat{\alpha}_{\text{zent}}$ (deutsche Kehlkopfkrebsdaten o. G.) . . .	94
43	Grenzsettings als Grundlage der Simulation . . . . .	95
44	Akzeptanzwahrscheinlichkeit $\eta$ mit/ohne Burn-In (simulierte Daten) .	97

---

45	Aufbereitung der Schätzergebnisse: Setting 1 – Simulation 10 . . . . .	98
46	Aufbereitung der Schätzergebnisse: Setting 2 – Simulation 10 . . . . .	99
47	Aufbereitung der Schätzergebnisse: Setting 3 – Simulation 5 . . . . .	100
48	Aufbereitung der Schätzergebnisse: Setting 4 – Simulation 4 . . . . .	101
49	Boxplots: $\hat{\boldsymbol{w}}$ nach Settings (simulierte Daten) . . . . .	102
50	Boxplots: $\hat{\boldsymbol{w}}$ nach Settings und Ausgangsgewichte (simulierte Daten) .	103
51	Boxplots $\hat{\boldsymbol{w}}$ nach Grad und Settings (simulierte Daten) . . . . .	104
52	MSE - räumliche Effekte (simulierte Daten) . . . . .	105
53	Boxplots: MSE - räumliche Effekte (simulierte Daten) . . . . .	106
54	MSE - Gewichte (simulierte Daten) . . . . .	106
55	Boxplots: MSE - Gewichte (simulierte Daten) . . . . .	107
56	Konvergenzanalyse Gewichte - Entwicklung Faktor $R$ . . . . .	110
57	Vergleich $\hat{\boldsymbol{\gamma}}_{\text{zent}} + \hat{\boldsymbol{\alpha}}_{\text{zent}}$ o. G. vs. m. G. (schottische Lippenkrebsdaten)	111
58	Signifikante $\hat{\boldsymbol{\gamma}}_{\text{zent}}$ und $\hat{\boldsymbol{\alpha}}_{\text{zent}}$ (schottische Lippenkrebsdaten m. G.) . . . . .	111
59	Auffällige $\hat{\boldsymbol{w}}$ (schottische Lippenkrebsdaten m. G.) . . . . .	112
60	Boxplots: $\hat{\boldsymbol{w}}$ nach Grad (schottische Lippenkrebsdaten m. G.) . . . . .	112
61	Vergleich $\hat{\boldsymbol{\gamma}}_{\text{zent}} + \hat{\boldsymbol{\alpha}}_{\text{zent}}$ o. G. vs. m. G. (deutsche Kehlkopfkrebsdaten)	113
62	Auffällige $\hat{\boldsymbol{w}}$ (deutsche Kehlkopfkrebsdaten m. G.) . . . . .	114
63	Boxplots: $\hat{\boldsymbol{w}}$ nach Grad (deutsche Kehlkopfkrebsdaten m. G.) . . . . .	115

## Referenzen

- Bates D, Maechler M (2013). “Matrix: Sparse and Dense Matrix Classes and Methods.” URL <http://cran.r-project.org/web/packages/Matrix/index.html>.
- Besag J (1974). “Spatial Interaction and the Statistical Analysis of Lattice Systems.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **36**(2), 192–236.
- Besag J, York J, Mollié A (1991). “Bayesian image restoration, with two applications in spatial statistics.” *Annals of the Institute of Statistical Mathematics*, **43**(1), 1–20.
- Bivand RS, Pebesma EJ, Rubio VG (2008). *Applied Spatial Data: Analysis with R*. Springer-Verlag New York.
- Brezger A, Fahrmeir L, Hennerfeind A (2007). “Adaptive Gaussian Markov random fields with applications in human brain mapping.” *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **56**(3), 327–345.
- Clayton D, Kaldor J (1987). “Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping.” *Biometrics*, **43**(3), 671–681.
- Cowles MK, Carlin BP (1996). “Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review.” *Journal of the American Statistical Association*, **91**(434), 883–904.
- Cramer E, Kamps U (2008). *Grundlagen der Wahrscheinlichkeitsrechnung und Statistik. Ein Skript für Studierende der Informatik, der Ingenieur- und Wirtschaftswissenschaften*. 2. Edition. Springer-Verlag Berlin Heidelberg.
- Cuthill E, McKee J (1969). “Reducing the bandwidth of sparse symmetric matrices.” In *Proceedings of the 1969 24th national conference*. 157–172. Association for Computing Machinery.
- Fahrmeir L, Kneib T (2010). *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. Oxford Statistical Science Series. Oxford University Press.
- Fahrmeir L, Kneib T, Lang S (2009). *Regression*. Springer-Verlag Berlin Heidelberg.
- Fahrmeir L, Künstler R, Pigeot I, Tutz G (2007). *Statistik. Der Weg zur Datenanalyse*. Springer-Verlag Berlin Heidelberg.
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004). *Bayesian Data Analysis*. 2. Edition. Chapman & Hall/CRC.
- Gelman A, Rubin DB (1992). “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical science*, **7**(4), 457–472.
- Holling H, Gediga G (2013). *Statistik – Wahrscheinlichkeitstheorie und Schätzverfahren*. Hogrefe Verlag.

- Hübner G (2009). *Stochastik. Eine anwendungsorientierte Einführung für Informatiker, Ingenieure und Mathematiker.* 5. Edition. Vieweg+Teubner Wiesbaden.
- Knorr-Held L, Rainer E (2001). “Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction.” *Biostatistics*, **2**(1), 109–129.
- Knorr-Held L, Rue H (2002). “On Block Updating in Markov Random Field Models for Disease Mapping.” *Scandinavian Journal of Statistics*, **29**(4), 597–614.
- Lam P (2008). “Bayesian Statistics in Political Science Research - Convergence Diagnostics (Titel von Homepage).” Online; Zugriff: 09.09.2013, URL <http://www.patricklam.org/teaching.html>.
- Mantel N, Stark CR (1968). “Computation of Indirect-Adjusted Rates in the Presence of Confounding.” *Biometrics*, **24**(4), 997–1005.
- Meier S, Borkowski A (2011). *Geometrie stochastischer Signale: Grundlagen und Anwendungen in der Geodaten-Verarbeitung.* Walter de Gruyter Berlin/Boston.
- Meintrup D, Schäffler S (2005). *Stochastik - Theorie und Anwendungen.* Springer-Verlag Berlin Heidelberg.
- Melnick EL, Everitt BS (2008). *Encyclopedia of Quantitative Risk Analysis and Assessment Volume 2.* West Sussex: John Wiley & Sons Ltd.
- Plummer M, Best N, Cowles K, Vines K (2006). “CODA: Convergence Diagnosis and Output Analysis for MCMC.” *R News*, **6**(1), 7–11. URL <http://cran.r-project.org/web/packages/coda/index.html>.
- Quarteroni A, Sacco R, Saleri F (2007). *Numerische Mathematik Band 1.* Springer-Verlag Berlin.
- R Development Core Team (2013). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- Rieger A (2011). *Spatio-temporale Aspekte des Screening-Programms für Kolonkarzinome in Bayern.* Masterarbeit, Ludwig-Maximilians-Universität München.
- Robert C (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.* 2. Edition. Springer-Verlag New York.
- Rue H (2001). “Fast sampling of Gaussian Markov random fields.” *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **63**(2), 325–338.
- Rue H (2009). “The R-INLA project: Code for Bym example (Titel von Homepage).” Online; Zugriff: 27.05.2013, URL <https://sites.google.com/a/r-inla.org/www/examples/volume-1/code-for-bym-example>.
- Rue H, Held L (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104. Chapman & Hall.

- Rytkönen MJ (2004). “Not all maps are equal: GIS and spatial analysis in epidemiology.” *International Journal of Circumpolar Health*, **63**(1).
- Schmid V (2004). *Bayesianische Raum-Zeit-Modellierung in der Epidemiologie*. Dissertation, Ludwig-Maximilians-Universität München.
- Schmid V (2013). “Skript der Veranstaltung Räumliche Statistik an der Ludwig-Maximilians-Universität München.” Online; Zugriff: 11.04.2013, URL <http://www.statistik.lmu.de/institut/ag/bioimg/lehre/rs2012/>.
- Spiegel (2013). “Spiegel Magazin Ausgabe 11 - Smart Cities.” Online; Zugriff: 27.05.2013, URL <http://www.spiegel.de/spiegel/print/d-91464895.html>.
- Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A (2002). “Bayesian measures of model complexity and fit.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), 583–639.
- Tutz G (2011). “Skript (Mitschrift) der Veranstaltung Generalisierte Regression an der Ludwig-Maximilians-Universität München.” Online; Zugriff: 10.06.2013, URL <http://www.statistik.lmu.de/~fachsch/vorlesungsdateien/GeneralisierteRegression.pdf>.
- Voigt C, Adamy J (2007). *Formelsammlung der Matrizenrechnung*. Oldenbourg Verlag München.
- Yan J (2006). “Gaussian Markov Random Fields: Theory and Applications.” *Journal of the American Statistical Association*, **101**(473), 388–389.
- Zavoianu C (2009). “Bandwidth Reduction - The CutHill-McKee Algorithm (Titel von Homepage).” Online; Zugriff: 12.08.2013, URL <http://ciprian-zavoianu.blogspot.de/2009/01/project-bandwidth-reduction.html>.

**Elektronischer Anhang**

## **Erklärung zur Urheberschaft**

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe.

München, den 21. Oktober 2013

Christian Schnell