



Studienabschlussarbeiten

Fakultät für Mathematik, Informatik
und Statistik

Bauer, Alexander:

Auswirkungen der Erdbebenquellendynamik auf den
zeitlichen Verlauf der Bodenbewegung

Masterarbeit, Wintersemester 2016

Fakultät für Mathematik, Informatik und Statistik

Ludwig-Maximilians-Universität München

<https://doi.org/10.5282/ubm/epub.31976>



Masterarbeit

Auswirkungen der Erdbebenquelldynamik auf den
zeitlichen Verlauf der Bodenbewegung

Alexander Bauer

Projektpartner:

Dr. Alice-Agnes Gabriel

Department für Geo- und Umweltwissenschaften (Geophysik)

LMU München

Betreuer:

Prof. Dr. Helmut Küchenhoff, Dr. Fabian Scheipl



Institut für Statistik

Ludwig-Maximilians-Universität München

28. November 2016

Abstract

Im Rahmen dieser Arbeit wurde in Zusammenarbeit mit dem Department für Geophysik der Ludwig-Maximilians-Universität München untersucht, auf welche Art die auf der Bruchfläche eines Erdbebens vorherrschenden physikalischen Bedingungen die durch das Beben verursachten Bodenbewegungen an der Erdoberfläche beeinflussen. Ziel ist es dabei, ein besseres Verständnis der großräumigen Prozesse zu erlangen, welche sich bei der Ausbreitung von Erdbeben abspielen. Die Datengrundlage bilden 200 mithilfe der Software SeisSol durchgeführte Simulationen auf Basis eines 1994 in Northridge (Südkalifornien) stattgefundenen Erdbebens.

Die über die Zeit gemessenen Bodenbewegungen werden durch ein funktionales Regressionsmodell unter Annahme einer Gammaverteilung für die Zielgröße modelliert. Insbesondere aufgrund der Datengröße wird dabei ein besonders effizienter Algorithmus zur Modellschätzung verwendet. Für eine adäquate Schätzung der Effekte wird eine mehrstufige Grid Search zur Modellwahl angewandt, unter anderem um die Knotenzahlen glatter Effekte optimal zu bestimmen. Um des Weiteren die Topographie im Modell zu berücksichtigen wird ein Ansatz verwendet, welcher eine genauere Kategorisierung des Gebietes nach Geländeformen ermöglicht. Auch aufgrund der Flexibilität der Methode kann dieser Ansatz für zukünftige Anwendungen weiterempfohlen werden. Zentrale Annahme bei der Modellierung ist, dass alle Erdbeben dem Hypozentrum als Punktquelle entspringen.

Von den evaluierten Einflussgrößen weisen der dynamische Reibungswiderstand sowie der Abstand zur Punktquelle den stärksten Einfluss auf die Bodenbewegungen auf. Je höher beide Parameter sind, desto geringer fallen die Bodenbewegungen dabei aus. Werden die Simulationsparameter bereits im Modell berücksichtigt, so bringt die zusätzliche Aufnahme der skalaren Momentenmagnitude als zentralem Parameter zur Beschreibung der Erdbebenquellendynamik keinen Mehrwert bezüglich der Vorhersage der Bodenbewegungen. Letzterer Befund stimmt mit dem weiteren Ergebnis überein, dass die Magnitude durch die Simulationsparameter sehr gut vorhergesagt werden kann. Bezüglich der Topographie wurde beobachtet, dass in Ebenen insgesamt etwas geringere Bodenbewegungen gemessen werden als in gebirgigeren Gebieten.

Im Rahmen der Modellierung zeigte sich auch, dass die Wahl der Punktquelle einen durchaus relevanten Einfluss auf die Modellgüte aufweist. Eine optimale Bestimmung der pro Beben verwendeten Punktquelle könnte für die Zukunft einen wichtigen Ansatzpunkt zur Verbesserung des Modells bieten.

Inhaltsverzeichnis

1	Einleitung	1
2	Datenbasis und Deskription	2
2.1	Simulationsparameter	3
2.2	Quellparameter	4
2.3	Topographie	7
2.4	Bodenbewegungen	8
3	Kategorisierung der Topographie	18
3.1	Topographic Position Index	18
3.2	Wahl geeigneter Nachbarschaften	21
4	Funktionale Regression	25
4.1	Notation	26
4.2	Funktion-auf-Skalar Regression	26
4.3	Schätzung	27
4.4	Effiziente Schätzung	31
5	Anwendung	35
5.1	Vorbereitungen	36
5.2	Modellwahl	38
5.3	Erklärung der Bodenbewegungen	42
5.3.1	Modell inklusive Quellparameter	42
5.3.2	Modell ohne Quellparameter	48
5.3.3	Modellevaluation	50
5.3.4	Auswirkung der Wahl der Punktquelle	55
5.4	Erklärung der Momentenmagnitude durch Simulationsparameter	59
5.5	Nebenfragestellungen	62
5.5.1	Unterschiedliche Ausbreitung nach Geländeform	62
5.5.2	Nicht-isotroper Richtungseffekt	64
6	Fazit und Ausblick	65
7	Literatur	66
8	Anhang	68
A	Daten (digital)	68
B	R-Codes (digital)	68
C	Finale Modelle zu Haupt- und Nebenfragestellungen (digital)	68
D	Evaluation der Kovariablen-Effekte	68
E	Residualplots Finales Modell inklusive Quellparameter	75
F	Modellergebnisse alternative Punktquelle	76
9	Eigenständigkeitserklärung	86

1 Einleitung

In Zusammenarbeit mit dem Department für Geophysik der Ludwig-Maximilians-Universität München wurde im Rahmen dieser Arbeit untersucht, auf welche Art die auf der Bruchfläche eines Erdbebens vorherrschenden physikalischen Bedingungen die durch das Beben verursachten Bodenbewegungen an der Erdoberfläche beeinflussen. Dies ist insbesondere von Interesse, da die physikalischen Gegebenheiten der Erdbebenquelle in aktuellen Erdbebenmodellen nur selten einbezogen werden. Generelles Ziel ist es, ein besseres Verständnis der großräumigen Prozesse zu erlangen, welche sich bei der Ausbreitung von Erdbeben abspielen. In Zukunft könnten so eventuell Rückschlüsse aus den gemessenen Bodenbewegungen auf die Bedingungen an der Bruchfläche gezogen werden oder bestehende Modelle zur Risikoanalyse, welche den auf Erdbeben bezogenen Gefährungsgrad einzelner Gebieten vorhersagen, könnten verbessert werden.

Die Datengrundlage bilden 200 mithilfe der Software SeisSol durchgeführte Simulationen eines Erdbebens unter jeweils veränderten physikalischen Bedingungen. Grundlage für die Analysen ist ein reales Erdbeben, welches 1994 in Northridge, einem Stadtteil von Los Angeles, stattfand. Unter anderem aufgrund des sich nah unter der Erdoberfläche befindlichen Hypozentrums sowie des dichter besiedelten Gebietes ist dieses Erdbeben Stand 2016 weltweit die sechstgrößte Naturkatastrophe seit 1980 (NatCatSERVICE 2016).

Dieses Projekt unterscheidet sich von der Arbeit von Biek und Park (2015) dahingehend, dass zum einen weitere Parameter zur Beschreibung der unterirdischen Dynamik der Erdbebenquelle in die Analysen einfließen. Zum anderen wird als Zielgröße nicht die maximal gemessene, skalare Bodenbewegung betrachtet, sondern die pro Seismograph gemessene Bodenbewegung über die Zeit wird als Zielgröße angesehen.

Für die Modellierung der über die Zeit gemessenen Bodenbewegungen werden funktionale Regressionsmodelle verwendet, wobei der Fokus aufgrund des Umfangs der Daten hierbei insbesondere auf der Verwendung eines effizienten Schätzalgorithmus liegt. Um die Topographie des betrachteten Gebietes in die Modellierung einzubeziehen wird ein Ansatz vorgestellt, welcher eine Kategorisierung derselbigen nach verschiedenen Geländeformen ermöglicht. Zentral für die Analysen ist darüber hinaus die vereinfachende Annahme, dass die gemessenen Erdbeben einer Punktquelle entstammen. Für alle Analysen wurde die Statistik-Software R (Version 3.3.1) verwendet (R Core Team 2016).

Eine Einführung in die Datenbasis sowie eine Deskription aller relevanten Größen findet sich in Kapitel 2. In den Kapiteln 3 und 4 wird anschließend die verwendete statistische Methodik eingeführt. Kapitel 5 motiviert den konkret verwendeten Modellierungsansatz und enthält die Ergebnisse bezüglich den untersuchten Fragestellungen. Ein Fazit sowie ein Ausblick auf mögliche weitere Ansatzpunkte bei der Modellierung der Bodenbewegungen wird schließlich in Kapitel 6 gegeben.

2 Datenbasis und Deskription

Die Daten, welche dieser Arbeit zugrunde liegen, stammen aus vom Geophysik-Department der Ludwig-Maximilians-Universität München in Zusammenarbeit mit dem Institut für Informatik der Technischen Universität München durchgeführten Erdbeben-Simulationen. Grundlage hierfür war ein reales Erdbeben, welches sich 1994 in Northridge (Südkalifornien) ereignete. Für die Untersuchung wurde die Ausbreitung des Erdbebens 200 mal unter jeweils veränderten physikalischen Bedingungen simuliert und die über die Zeit variierenden Messungen der Bodenbewegung wurden an 6146 künstlich aufgestellten Seismographen aufgezeichnet. Das Hypozentrum des Bebens - welches im Gegensatz zum Epizentrum die genaue Lokation der Erdbebenquelle im Untergrund angibt und nicht den senkrecht darüber liegenden Ort auf der Erdoberfläche - lag in ca. 18km Tiefe. Eine Darstellung der Bruchfläche, welche in den Simulationen verwendet wurde, findet sich in Abbildung 1. Die schräg im Untergrund liegende, gekrümmte Fläche hat dabei eine ungefähre Größe von $32\text{km} \times 24\text{km}$ und befindet sich in einer Tiefe von etwa 7km bis 23km.

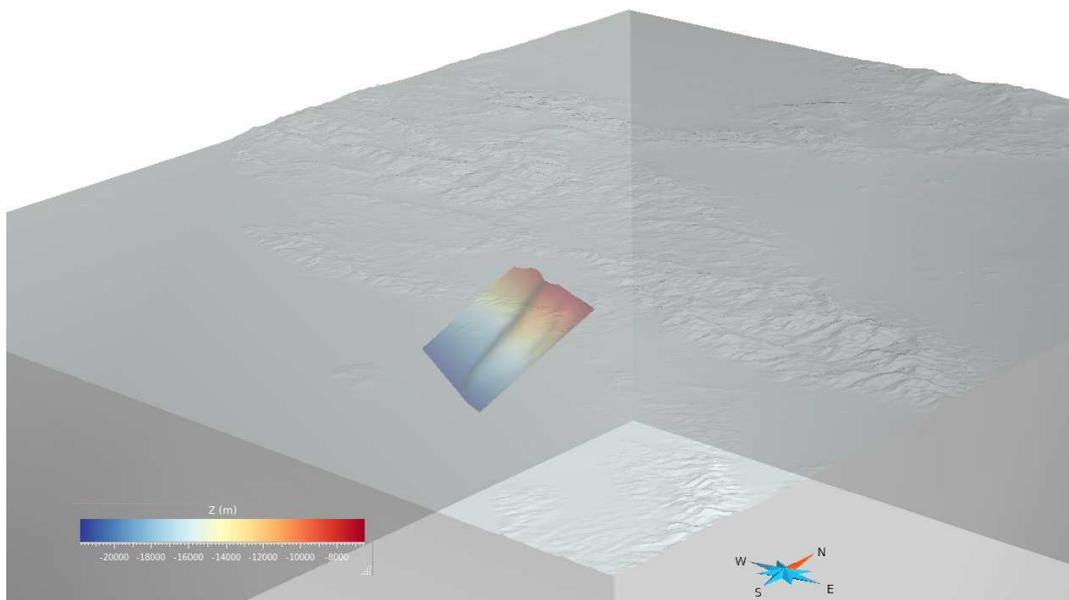


Abbildung 1: Lage der in den Simulationen verwendeten Bruchfläche im Untergrund des betrachteten Gebietes

Die Simulationen wurden mithilfe der Open-Source-Software SeisSol (www.seissol.org) durchgeführt. Das betrachtete dreidimensionale Gebiet wird dabei mit einem tetraedrischen Netz an Simulationspunkten überspannt. Basierend auf der Lösung elastischer Wellengleichungen, in welchen der auftretende Bruch als spezielle interne Randbedingung einfließt, ergeben sich anschließend die simulierten Bodenbewegungen pro Simulationspunkt (Dumbser und Käser 2006; Pelties et al. 2012; Pelties, Gabriel und Ampuero 2013; Heinecke et al. 2014).

Im Folgenden werden alle für die Analysen relevanten Variablen eingeführt und deskriptiv dargestellt. Neben den Simulationsparametern, welche die den Simulationen zugrunde liegenden und künstlich variierten physikalischen Parameter enthalten, wird auf die ebenfalls über die Simulationen variierenden Quellparameter eingegangen. Letztere beschreiben die Bedingungen an der Erdbebenquelle, also der Bruchfläche des Bebens im Untergrund. Zu betonen ist, dass die Quellparameter dabei nicht wie die Simulationsparameter künstlich variiert wurden. Sie ergeben sich stattdessen jeweils durch die im konkreten Simulationssetting vorherrschenden, durch die Simulationsparameter definierten physikalischen Gegebenheiten. Nachdem im Anschluss daran die topographischen Daten, die für das Untersuchungsgebiet vorliegen, eingeführt wurden, wird schließlich ein Überblick der über die Zeit gemessenen Bodenbewegungen gegeben, welche in der Modellierung durch die übrigen Parameter erklärt werden sollen.

Da die Quellparameter, die eine zentrale Rolle in den Analysen spielen, für die Simulationen Nr. 1 bis 50 nicht vorliegen, wurden diese Simulationen in den kompletten Untersuchungen aussen vor gelassen. Weiterhin waren die Messungen der Bodenbewegungen bei 15 weiteren Simulationen fehlerhaft und folglich wurden diese Simulationen ebenfalls ausgeschlossen. Insgesamt flossen also die Daten von 135 der 200 Simulationen in die folgenden Analysen ein. Eine (negative) Beeinflussung der Analyseergebnisse durch einen möglichen systematischen Ausschluss von Simulationen kann ausgeschlossen werden, da die Beschädigung der Datensätze lediglich auf Fehler bei einem Kopiervorgang der Dateien zurückzuführen ist.

Alle für die vorliegende Arbeit relevanten Daten finden sich im digitalen Anhang A.

2.1 Simulationsparameter

Anhand eines Versuchsplanungsdesigns wurden in den durchgeführten Simulationen des Northridge-Erdbebens insgesamt fünf Parameter variiert, welche die physikalischen Bedingungen der Simulationen definieren. Die Koeffizienten wurden dabei durch das verwendete Latin Hypercube Design derart variiert, dass eine optimale Abdeckung des Parameterraums erzielt wird (McKay und Beckman 1979). Diese Parameter werden in der folgenden Arbeit als Simulationsparameter bezeichnet. Die Parameter unterscheiden sich dabei nur zwischen verschiedenen Simulationen - innerhalb einer Simulation wurden die Parameterwerte über alle Messpunkte konstant gehalten.

Eine Übersicht der fünf Simulationsparameter findet sich in Tabelle 1. Neben der Beschaffenheit des Untergrunds, welcher pro Simulation einheitlich über das komplette Untersuchungsgebiet entweder aus Gestein oder Sediment besteht, umfassen die variierten Simulationsparameter dabei Koeffizienten, die verschiedene Reibungsparameter zu Beginn der Simulation definieren (l , μ_s , μ_d). Darüber hinaus wird durch den Parameter *angle*

der Winkel der Hintergrundspannung vorgegeben, welcher die Initialspannungen auf der Bruchfläche beeinflusst.

Parameter	Beschreibung	Werte
<i>struct</i>	Untergrundstruktur/Bodenbeschaffenheit	{rock, sediment}
<i>l</i>	lineares slip-weakening in m; über diese Distanz fällt der Reibungswiderstand vom statischen auf den dynamischen Wert ab	0.1 - 1.5
μ_s	statischer Reibungskoeffizient	0.6 - 0.8
μ_d	dynamischer Reibungskoeffizient	0.1 - 0.5
<i>angle</i>	Winkel der (konstanten) Hintergrundspannung	25° - 35°

Tabelle 1: Übersicht über die Simulationsparameter; Quelle: Stöcker und Kühn (2015)

2.2 Quellparameter

Die Erdbebenquellendynamik, welche die auf der Bruchfläche im Untergrund vorliegenden Gegebenheiten bezeichnet, wird durch verschiedene Parameter beschrieben. Wie erläutert wird für diese Arbeit der Einfachheit halber davon ausgegangen, dass der Ausgangspunkt des Bebens einer Punktquelle entspricht und keiner *Bruchfläche*. Darüber hinaus werden die physikalischen Zustände durch zeitkonstante, skalare Parameter beschrieben. Da die in den Rohdaten vorliegenden Quellparameter über die gesamte Bruchfläche räumlich aufgelöst sind stellen die hier verwendeten Versionen der Quellparameter räumliche Aggregationen dar.

Eine Übersicht über alle Quellparameter findet sich in Tabelle 2. Neben der skalaren Momentenmagnitude - welche sich insbesondere daraus berechnet, wie viele Meter ein spezifischer Punkt auf der Bruchfläche verglichen mit einem zu Beginn des Erdbebens benachbarten Punkt insgesamt durch das Beben verschoben wurde - sind weiterhin die *slip rate* sowie der *rupture speed* (bzw. *rupture velocity*) relevante Kenngrößen bzgl. der Queldynamik. Erstere slip rate steht in ihrer unaggregierten Version dabei dafür wie hoch die maximale absolute Partikelgeschwindigkeit in m/s ist, mit der sich die zwei Seiten eines Bruchs relativ zueinander über den Verlauf des Bebens verschieben. Der rupture speed hingegen gibt pro Messpunkt auf der Bruchfläche die über die Zeit gemittelte Bruchgeschwindigkeit in m/s an und ist somit ein Maß dafür, wie schnell sich die Bruchfront auf der Bruchfläche ausbreitet.

Das zur Umrechnung der slip-Werte in Momentenmagnituden benötigte *Schermodul* stellt eine Materialkonstante dar, welche sich auf die Verformung des Materials infolge einer

bestehenden Spannung bezieht. Allgemein variiert das Schermodul je nach bestehendem Untergrund (*rock / sediment*), hier wurde in den Simulationen bei beiden Untergründen der Wert $3,6288 \cdot 10^{10}$ verwendet.

Parameter	Beschreibung
$magnitude_{sum}$	Summe der skalaren Momentenmagnituden in Nm; Berechnung der Einzel-Magnituden aus den slips (der relativen Verschiebungen benachbarter Punkte auf der Bruchfläche in m): $magnitude [Nm] = slip [m] \cdot \text{Bruchfläche} [m^2] \cdot \text{Schermodul} [N/m^2]$ $= slip \cdot 8,6 \cdot 10^8 \cdot 3,6288 \cdot 10^{10}$
sr_{mean}	Mittelwert der slip rates, d.h. der maximalen absoluten Partikelgeschwindigkeiten in m/s, mit der sich die zwei Seiten eines Bruchs relativ zueinander verschieben
vr_{mean}	Mittelwert der rupture speeds, d.h. der punktweisen mittleren Bruchgeschwindigkeiten in m/s, mit der sich die Bruchfront auf der Bruchfläche ausbreitet

Tabelle 2: Übersicht über die Quellparameter, welche räumliche Aggregationen über alle sich auf der Bruchfläche befindlichen Messpunkte darstellen

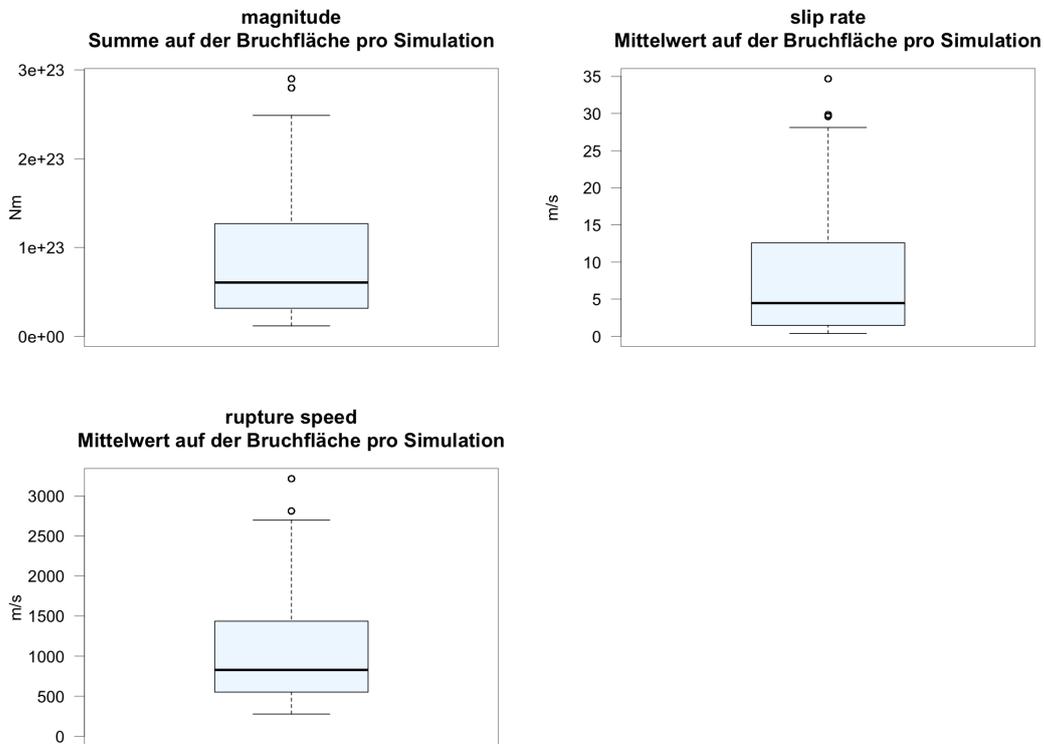


Abbildung 2: Verteilungen der Quellparameter

Für die Aggregation der Quellparameter werden die Parameter auf unterschiedliche Arten über die Bruchfläche zusammengefasst: Bzgl. der Magnitude wird die Summe über alle Punkte auf der Bruchfläche gebildet. Die dadurch erhaltene skalare Momentenmagnitude stellt dadurch ein Maß für die allgemeine Stärke eines Erdbebens dar. Bzgl. der slip rate wird der Mittelwert der einzelnen maximalen absoluten Partikelgeschwindigkeiten verwendet. Die räumlich aufgelösten rupture speeds schließlich werden ebenfalls per Mittelwertbildung zusammengefasst.

Die Verteilungen der einzelnen Quellparameter sind in Abbildung 2 dargestellt. Zu sehen ist, dass alle Parameter eine rechtsschiefe Verteilung aufweisen. Die Wertebereiche aller drei Quellparameter sind physikalisch sinnvoll, extreme Ausreißer liegen also nicht vor.

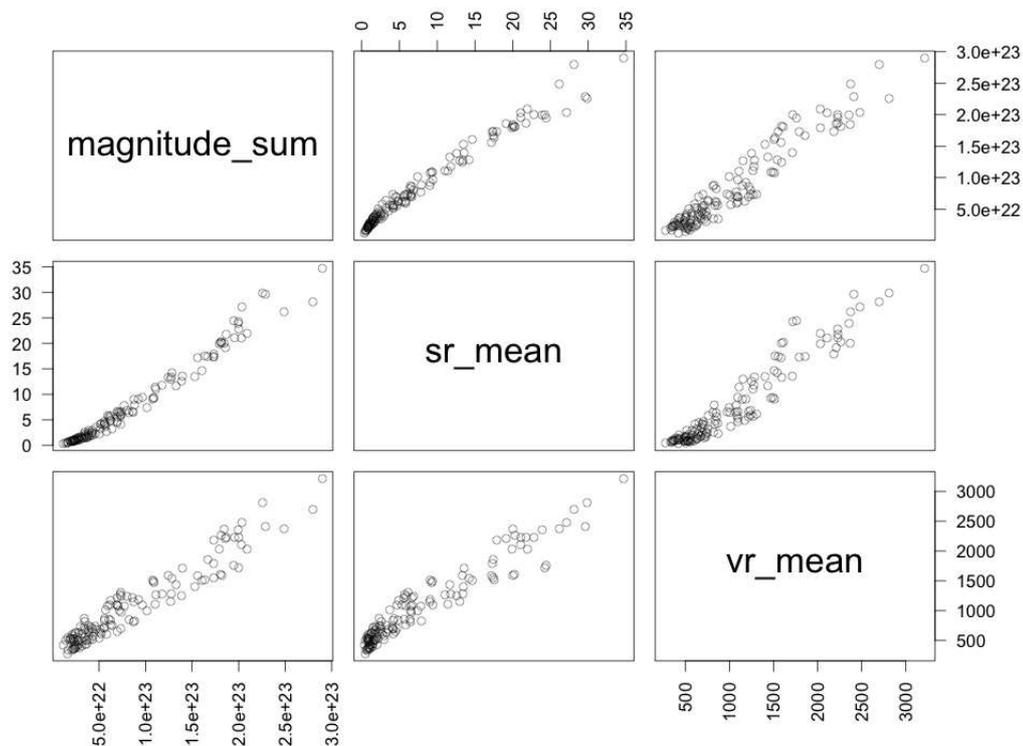


Abbildung 3: Paarweise Scatterplots der Quellparameter

Bei Betrachtung der paarweisen Scatterplots zwischen den Quellparametern in Abbildung 3 sowie der paarweisen Korrelationen in Tabelle 3 fällt auf, dass die Quellparameter durchweg sehr hoch korrelieren. Neben Korrelationen von 0,96 zwischen $magnitude_{sum}$ und vr_{mean} bzw. zwischen sr_{mean} und vr_{mean} ist zwischen $magnitude_{sum}$ und sr_{mean} eine fast perfekte lineare Korrelation von 0,99 zu beobachten. Wie in der Analyse mit der durch diese sehr hohen Korrelationen verursachten Multikollinearität umgegangen wird, wird in Kapitel 5 erläutert.

	$magnitude_{sum}$	sr_{mean}	vr_{mean}
$magnitude_{sum}$	1	0,99	0,96
sr_{mean}	-	1	0,96
vr_{mean}	-	-	1

Tabelle 3: Paarweise Korrelationen der Quellparameter

2.3 Topographie

Zur Untersuchung eines möglichen Effektes der Topographie des betrachteten Gebietes auf die Ausbreitung des Erdbebens liegen zusätzlich zu den genauen Lokationen und Höheninformationen der Messstationen hochaufgelöste topographische Informationen bzgl. des kompletten Gebietes vor. Die mittlere Auflösung in x-Richtung (nach dem UTM-Koordinatensystem, siehe Tabelle 4) beträgt dabei ca. 78m, die mittlere Auflösung in y-Richtung ca. 91m.

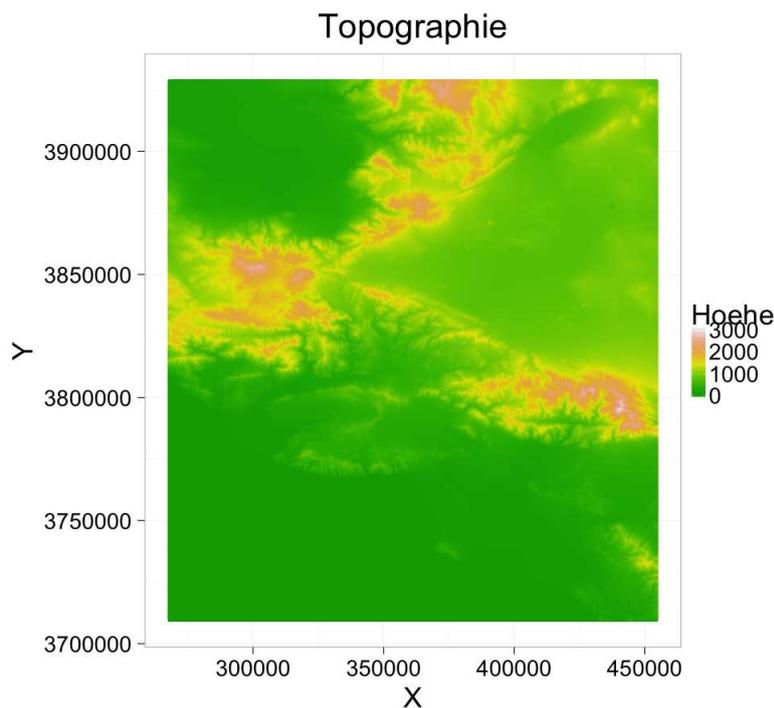


Abbildung 4: Hochaufgelöste Topographie; Koordinaten entstammen dem UTM-Koordinatensystem (siehe Tabelle 4)

Eine Darstellung der kompletten Topographie findet sich in Abb. 4. Anzumerken ist, dass für das Gebiet im Südwesten - welches von Meer bedeckt ist - keine Informationen über die Topographie vorlagen und die Höhe aller Punkte entsprechend auf 0m festgesetzt

wurde. Auf einen Ausschluss dieses Gebietes aus den Analysen wurde bewusst verzichtet, da trotz dieser nicht realen Höhen durch die Simulationen sinnvoll untersucht werden kann, wie sich die simulierten Beben auf dieser künstlichen Ebene verhalten. Auch soll nochmals betont werden, dass die Untergrundbeschaffenheit für spezifische Simulationen jeweils über das komplette Gebiet konstant ist (siehe Kapitel 2.1). Mögliche Effekte der Höhe oder der Geländeform sind damit wirklich reine Topographie-Effekte und sind nicht durch unterschiedliche Böden bzw. Erdschichten beeinflusst.

Von dem eigentlichen Meeresgebiet abgesehen umfasst die Umgebung von Northridge eine Vielzahl unterschiedlicher Geländeformen wie etwa Bergketten, schmale Täler oder auch weitere Ebenen. Auf welche Art die Topographie in die Modellierung einfließt wird in Kapitel 3 beschrieben.

2.4 Bodenbewegungen

Die durch die simulierten Erdbeben hervorgerufenen Bodenbewegungen wurden in jeder der Simulationen durch 6146 künstlich aufgestellte Seismographen aufgezeichnet. Alle Messungen beginnen dabei bei Ausbruch des jeweiligen Erdbebens und umfassen ein Zeitintervall von 30 Sekunden, wobei in den Rohdaten eine zeitliche Auflösung von 0,005s bzw. 200Hz vorlag. Dies entspricht pro 30-sekündiger Messung ca. 5500 Einzel-Beobachtungen.

In Abb. 5 ist die räumliche Verteilung aller Seismographen dargestellt. Für die Simulationen wurden die Messstationen auf einem regelmäßigen Gitter aufgestellt, welches ca. 198km x 166km groß ist. Hierdurch liegen zwischen zwei benachbarten Seismographen auf dem regelmäßigen Gitter ca. 2km in x-Richtung bzw. 2,5km in y-Richtung. Eine Ausnahme bilden lediglich die Seismographen Nr. 1 bis 62. Diese stehen an den Positionen 62 real existierender Seismographen und fallen nicht in das regelmäßige Gitter. Reale Messungen der Seismographen Nr. 1 bis 62 flossen jedoch nicht in die Daten mit ein, alle Daten entstammen den durchgeführten Computersimulationen. Seismograph Nr. 62 steht am Epizentrum der simulierten Erdbeben bzw. direkt über dem sich ca. 18km unter der Erdoberfläche befindlichen Hypozentrum.

Anzumerken ist, dass für die Durchführung der Simulationen alle Messstationen 50m unter der Erdoberfläche positioniert wurden, um durch die Simulationen hervorgerufene mögliche Artefakte in den Messungen zu vermeiden. Diese simulierten Messungen können letztendlich jedoch als Messungen von Seismographen an der Oberfläche angesehen werden. In den Analysen werden als Höhendaten der Messstationen die unveränderten Höhen der „eingegrabenen“ Seismographen verwendet. Eine systematische Verzerrung der Ergebnisse ergibt sich dadurch nicht, da wie erläutert alle 6146 Seismographen gleich weit unter der Erdoberfläche positioniert wurden.

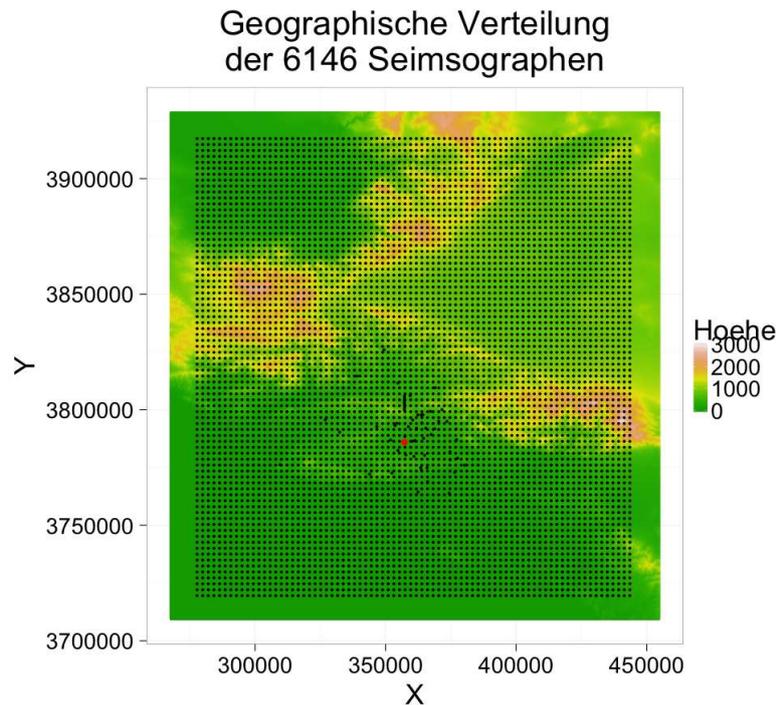


Abbildung 5: Geographische Verteilung der Seismographen; jeder Punkt repräsentiert einen Seismographen, in rot eingezeichnet ist das Epizentrum

Eine Übersicht aller die Seismographen-Lokationen bzw. -Messungen betreffenden Daten findet sich in Tabelle 4. Neben den Koordinaten und der Höhe der Messstationen sowie den Distanzen zum Hypo- und Epizentrum umfassen diese Daten auch die über die Zeit gemessenen Partikelgeschwindigkeiten an den Seismographen. Diese Geschwindigkeiten der Bodenbewegung liegen sowohl in horizontaler x- und y-Richtung als auch in vertikaler z-Richtung vor und stellen in den später gerechneten Modellen die zu erklärende Zielgröße dar.

Da von primärem Interesse ist, wie stark sich der Boden an einem Seismographen *allgemein* bewegt, werden die verschiedenen Partikelgeschwindigkeiten zu einer Zielgröße verrechnet. Durch Bildung der euklidischen Norm der drei Größen pro Zeitpunkt erhält man die richtungsunabhängige *absolute Geschwindigkeit der Bodenbewegung*, welche im Folgenden einfach als *Bodenbewegung* bezeichnet wird:

$$\text{Bodenbewegung} = \sqrt{u^2 + v^2 + w^2}$$

Als für die Ausbreitung von Erdbeben relevanteres Distanzmaß wird in den weiteren Analysen mit der hypo- und nicht mit der epizentralen Distanz gearbeitet. Grund dafür ist, dass sich Erdbeben im Allgemeinen auf direktem Wege von der Quelle zu den Messstationen an der Oberfläche ausbreiten.

Parameter	Beschreibung
x, y, z	Ost- und Nordwert-Koordinaten (Entfernung in m zum Mittelmeridian bzw. zum Äquator) sowie Höhe in m über NN
<i>hypo.dis</i>	Distanz zum Hypozentrum in m
<i>epi.dis</i>	Distanz zum Epizentrum in m
u, v, w	Zeitlich aufgelöste Partikelgeschwindigkeiten in x -, y - bzw z -Richtung in m/s
<i>Bodenbewegung</i>	Richtungsunabhängige absolute Partikelgeschwindigkeit, $\sqrt{u^2 + v^2 + w^2}$

Tabelle 4: Übersicht über die Informationen bzgl. der einzelnen Seismographen; verwendet wird das UTM-Koordinatensystem

Downsampling

Aufgrund der hohen zeitlichen Auflösung der Seismogramme war es als Vorbereitung auf die Analysen notwendig die Seismogramme zeitlich downzusamplen, um die Daten mittels komplexerer statistischer Modelle bei zur gleichen Zeit begrenzten computationalen Ressourcen analysieren zu können. Letztendlich wurde mit einer zeitlichen Auflösung von 0,5s bzw. 2Hz die größte Auflösung verwendet, welche gleichzeitig noch eine sinnvolle Repräsentierung der physikalischen Prozesse sicherstellt. Alle weiteren Analysen in dieser Arbeit basieren folglich auf diesen geringer aufgelösten Seismogrammen. Die für die Analyse verwendeten Daten weisen dadurch eine Größe von ca. 0,5GB auf.

Das Downsampling der Messungen erfolgte auf direkte Weise, indem lediglich jeder einhundertste Wert in den Daten behalten und die übrigen Messungen verworfen wurden. Diese Methode wurde verwendet, da typische seismologische Datenprozessierungsmethoden wie beispielsweise die Anwendung eines Aliasing-Filters zu instabilen Resultaten führten, unter anderem auch aufgrund der vergleichsweise kurzen Messreihe.

Deskription der Bodenbewegungen

Einen allgemeinen Eindruck über das Niveau der durch die Erdbeben verursachten Bodenbewegungen vermittelt Abbildung 6. Dargestellt wird darin pro Messstation der über die Zeit sowie über alle Simulationen gebildete Mittelwert. Wie erwartet fallen die mittleren Bodenbewegungen umso niedriger aus, je weiter entfernt vom Erdbebenzentrum die Messungen aufgenommen werden. Der mit ansteigender hypozentraler Distanz zunehmende Abfall verläuft dabei relativ gleichmäßig und weist nur insofern eine Abweichung

von einer isotropen, richtungsunabhängigen Entwicklung auf, als dass die Bodenbewegungen in Nord-Süd Richtung geringfügig schneller abfallen als in Ost-West Richtung. Auffällig ist darüber hinaus, dass das Epizentrum - welches in den Analysen als Punktquelle der Erdbeben angenommen wird - nicht übereinstimmt mit dem Seismographen, an welchem die größte mittlere Bodenbewegung registriert wurden. Dieser Punkt wird ab S.16 grundlegender diskutiert.

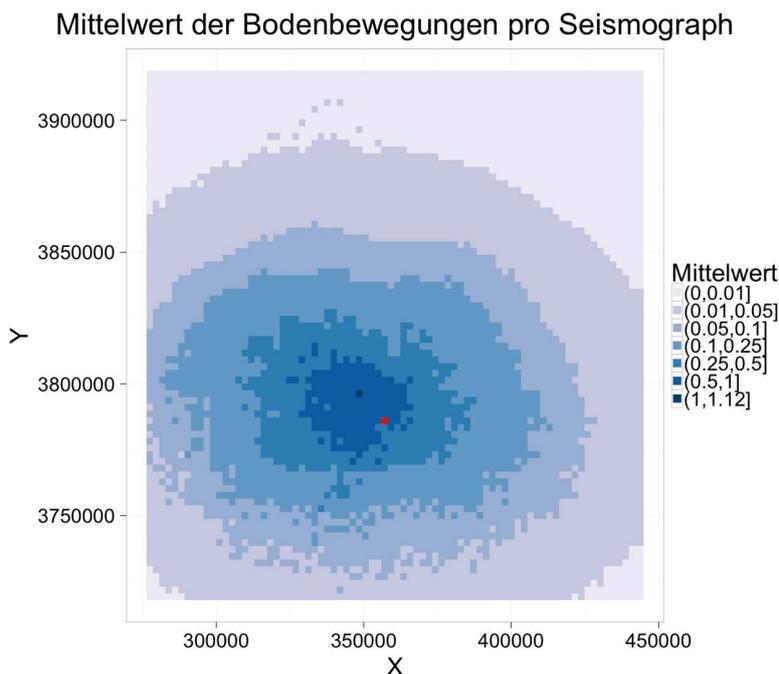


Abbildung 6: Darstellung der über die Zeit sowie über alle Simulationen gemittelten Bodenbewegungen pro Seismograph; ein Punkt stellt einen Seismographen dar; in rot eingezeichnet ist das Epizentrum

Eine Übersicht über die allgemeine Form der über die Zeit gemessenen Seismogramme findet sich in Abbildung 7, in der die Bodenbewegungen aller Seismographen für eine zufällig ausgewählte Simulation dargestellt sind. Der Übersichtlichkeit halber sind die Seismogramme hierbei nach der hypozentralen Distanz gruppiert. Die Messungen der übrigen Simulationen weisen eine vergleichbare Form auf.

Betrachtet man die Grafik der Messungen, welche in maximal 25km Entfernung vom Hypozentrum entstanden sind, so fällt eine klare Struktur der Seismogramme auf: Nach etwas kleineren Ausschlägen zu Beginn lassen sich die größten Bodenbewegungen nach ca. 5-10 Sekunden beobachten. Anschließend nehmen die Bodenbewegungen deutlich ab und stagnieren schließlich auf einem niedrigeren Niveau. Bei den Messungen weiter entfernter Seismographen verschiebt sich die Struktur der gemessenen Bodenbewegungen. Zum einen sinken sowohl das allgemeine Niveau als auch die Streuung der Messungen

mit steigendem Abstand zum Hypozentrum. Des Weiteren treten die durch die Erdbeben verursachten Bodenbewegungen mit steigender hypozentraler Distanz ebenfalls erst später ein, wodurch sich der zu Beginn der Seismogramme zu beobachtende Abschnitt an (Quasi-)Nullmessungen vergrößert. Dieser Abschnitt an führenden Nullmessungen trägt für die Analyse keine relevante Information, da dies lediglich die Zeit kennzeichnet, bevor sich das Erdbeben bis zur Messstation ausgebreitet hat. Schließlich lässt sich auch eine klare Struktur beobachten, dass, je weiter man sich von der Erdbebenquelle entfernt, ein längerer Abschnitt an kleineren Bodenbewegungen zu beobachten ist, bevor die Seismogramme auf ihren Maximalwert ansteigen.

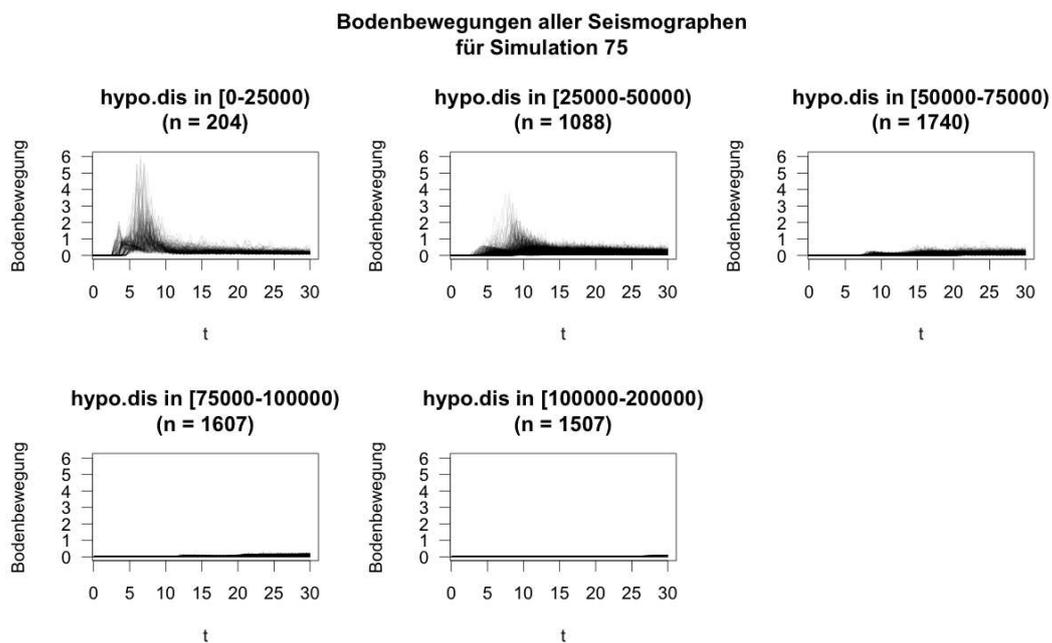


Abbildung 7: Darstellung der 30-sekündigen Seismogramme aller 6146 Seismographen für eine zufällig ausgewählte Simulation; die Einzelgrafiken enthalten jeweils nur Seismographen, welche in das jeweilige Intervall der hypozentralen Distanz fallen; n gibt die Anzahl an Seismographen pro Kategorie an

Die in Abbildung 7 zu beobachtende Struktur der Seismogramme einer spezifischen Simulation, dass der Wertebereich der Bodenbewegungen mit steigender hypozentraler Distanz der Seismographen abnimmt, findet sich - wie in Abbildung 8 zu sehen - in allen Simulationen wieder. Pro Seismogramm ist hierbei die größte gemessene Bodenbewegung gegen die hypozentrale Distanz der zugehörigen Messstation abgetragen. Der Klarheit halber soll nochmals betont werden, dass die maximalen Messungen in Abbildung 8 gegen die hypozentrale Distanz und *nicht* wie in Abbildung 7 gegen die Zeit abgetragen sind. Neben der bereits hervorgehobenen klaren Struktur, dass der Wertebereich der Bodenbewegungen mit steigender Entfernung deutlich abnimmt, lässt sich auch erkennen,

dass es über alle Seismographen und Simulationen keine extremen Ausreißer nach oben gibt. Wie in Abbildung 6, so ist auch hier auffällig, dass die stärksten Bodenbewegungen nicht an den Seismographen registriert wurden, welche die geringste Distanz zum Hypo- bzw. Epizentrum aufweisen, sondern erst 5-10 Kilometer weiter entfernt, nahe dem Punkt der größten mittleren Bodenbewegung. Eine mögliche Ursache dafür liegt wie erwähnt in der vereinfachenden Annahme einer Punktquelle als Ausgangspunkt der Erdbeben und wird auf S.16 weiter ausgeführt.

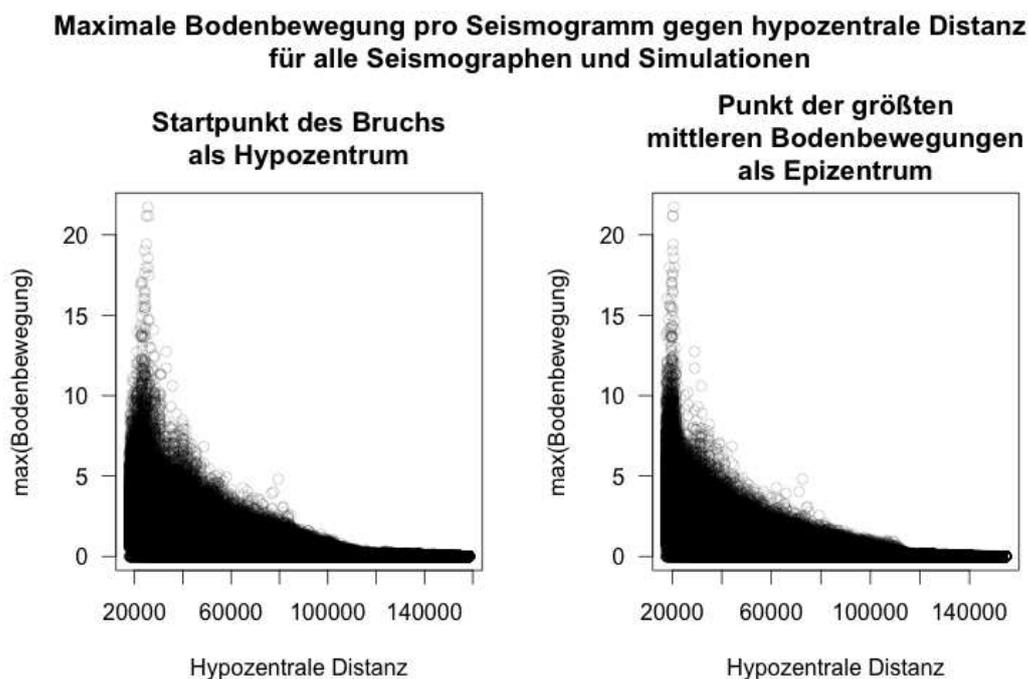


Abbildung 8: Maximale Bodenbewegung pro Seismogramm gegen die hypozentrale Distanz; Ein Punkt in der Grafik stellt die maximale Bodenbewegung eines 30-sekündigen Seismogramms gegen die hypozentrale Distanz des zugehörigen Seismographen dar

Zum Abschluss der Deskription wird nun noch untersucht, ob es möglicherweise einen Effekt dahingehend gibt, dass sich Erdbeben in bestimmte Richtungen schneller ausbreiten als in andere. Zur Veranschaulichung sind in Abbildung 9 pro Seismogramm für eine zufällig ausgewählte Simulation die Zeitpunkte dargestellt, zu welchen die erste Bodenbewegung von mindestens 0,01 gemessen wurde, bzw. zu welchen für das jeweilige Seismogramm der maximal gemessene Wert eintrat. Alle Bodenbewegungen, welche unter die Schwelle von 0,01 fallen, werden analog zur späteren Analyse als nicht durch ein Erdbeben verursachtes Grundrauschen angesehen. Bei den übrigen, hier nicht dargestellten Simulationen, lassen sich vergleichbare Strukturen erkennen.

Betrachtet man die räumliche Ausbreitung des Erdbebens diesbezüglich, wann pro Seismograph die erste Bodenbewegung $\geq 0,01$ registriert wurde (Abbildung 9, linke Grafik),

fallen durchaus unterschiedliche Ausbreitungsgeschwindigkeiten auf. Klar zu erkennen ist, dass die erste gemessene Bodenbewegung $\geq 0,01$ auf einer von Südwest nach Nordost verlaufenden Achse sowie auf einer zweiten, von Nordwest nach Südost verlaufenden Achse teilweise deutlich später zu beobachten ist, als dies an anderen Messstationen, welche eine vergleichbare Distanz zum Hypozentrum aufweisen, der Fall ist. Abgesehen von diesen beiden Achsen breitet sich das Beben jedoch relativ gleichmäßig in die verschiedenen Himmelsrichtungen aus.

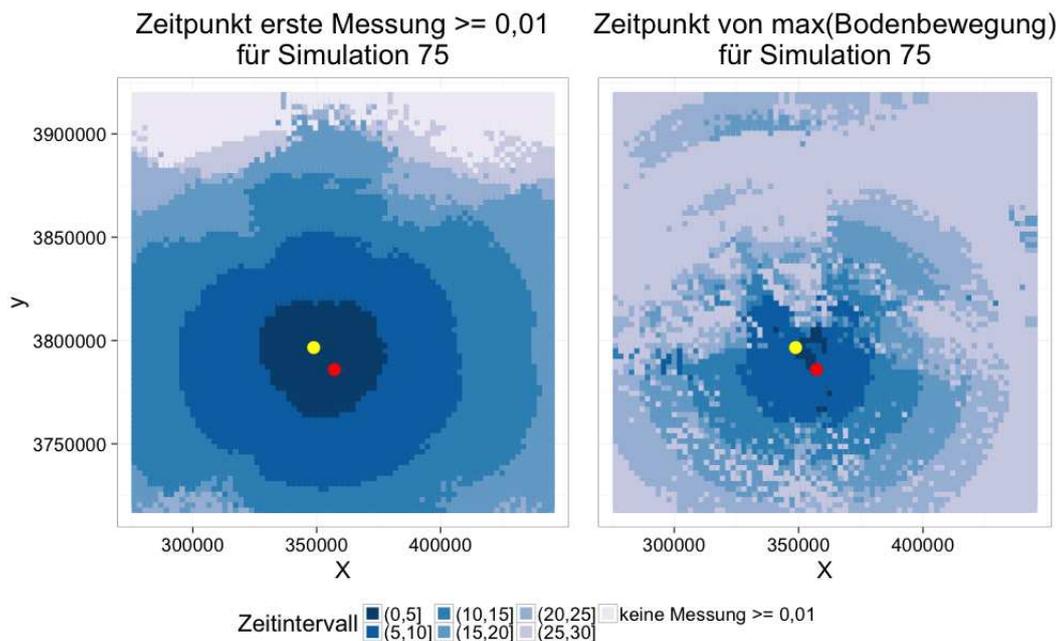


Abbildung 9: Darstellung der Zeitpunkte, zu welchen für die Seismogramme einer zufällig ausgewählten Simulation die erste gemessene Bodenbewegung $\geq 0,01$ auftritt (linke Grafik) bzw. die für das jeweilige Seismogramm maximale Bodenbewegung erreicht wird (rechte Grafik); in rot eingezeichnet ist das Epizentrum, in gelb der Punkt der größten mittleren Bodenbewegung

Auf Basis dieses Befundes ist eine tiefere Untersuchung möglicher nicht-isotroper Ausbreitungsgeschwindigkeiten der simulierten Erdbeben zu empfehlen. Aufgrund der Nichttrivialität dieser Analysen - u.a., da dazu die Richtung des Bruchs auf der unterirdischen Bruchfläche pro Erdbeben einbezogen werden müsste - war dies im Rahmen der vorliegenden Arbeit jedoch nur begrenzt umsetzbar. Die vorgenommenen Analysen diesbezüglich finden sich in Kapitel 5.5.2.

Bei Betrachtung der Ausbreitungsgeschwindigkeit von Erdbeben ist nicht nur die bis hierhin diskutierte Dimension wichtig, wie schnell eine erste relevante Bodenbewegung registriert wird, sondern unter anderem auch mit welchem zeitlichen Abstand sich die größte Bodenbewegung an einem Ort ereignet. Diese zweite Dimension wird in der rechten Grafik von Abbildung 9 untersucht. Obwohl sich erneut in viele Richtungen recht gleichmäßige Ausbreitungsgeschwindigkeiten erkennen lassen, liegen im Vergleich zur

Ausbreitung des Bebens bzgl. der ersten gemessenen Bodenbewegung $\geq 0,01$ hierbei deutlich unregelmäßigere Strukturen vor. Insbesondere vom Erdbebenzentrum aus Richtung Westen, Norden und Osten fallen Gebiete auf, welche keine allzu große Entfernung zum Epi- bzw. Hypozentrum aufweisen, bei welchen die maximal gemessene Bodenbewegung jedoch erst deutlich später eintritt als bei vergleichbaren, ähnlich weit vom Erdbebenzentrum entfernt aufgezeichneten Seismogrammen. Ganz im Norden ist darüber hinaus ein Gebiet auffällig, in welchem die maximal gemessene Bodenbewegung im Vergleich zu Nachbargebieten deutlich früher einsetzte.

**Zeitpunkt von max(Bodenbewegung)
für Simulation 75**

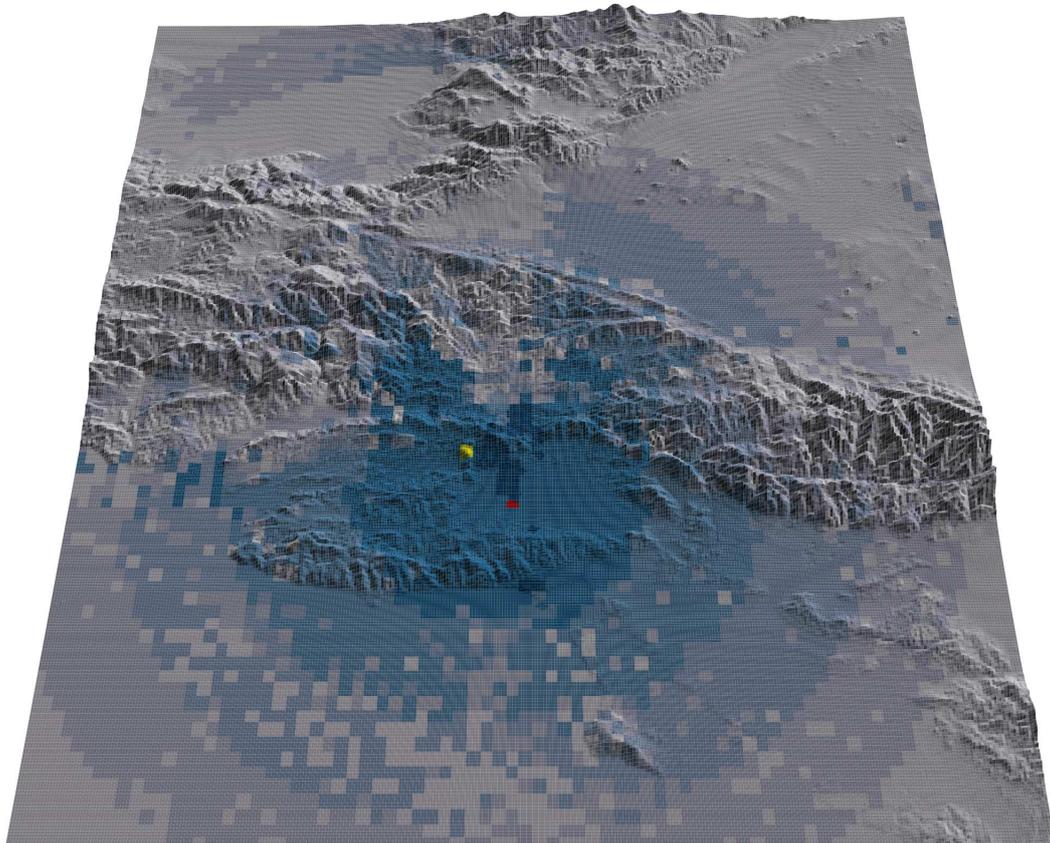


Abbildung 10: Darstellung der Zeitpunkte, zu welchen für die Seismogramme einer zufällig ausgewählten Simulation die für das jeweilige Seismogramm maximale Bodenbewegung erreicht wird; unterlegt ist die dreidimensionale Topographie des betrachteten Gebietes; die verwendete Farblegende ist identisch zu Abbildung 9

Eine mögliche Ursache für die unregelmäßigen Strukturen dieser räumlichen Ausbreitung könnte die Geländetopographie darstellen. Aus physikalischer Sicht scheint dies plausi-

bel, da die durch ein Erdbeben verursachten Bodenbewegungen durchaus von der Topographie beeinflusst werden. Im Gegensatz zu auf einer Ebene positionierten Seismographen - welche von den vom Hypozentrum ausgehenden Druckwellen auf direktem Wege erreicht werden - werden auf einem Berg gelegene Seismographen beispielsweise erst von den Wellen erreicht, welche durch den Kontakt mit der gefalteten, gebirgigen Erdoberfläche bereits des Öfteren reflektiert und dadurch abgeschwächt wurden. Durch die Simulationsdaten lässt sich diese Vermutung bestätigen (vgl. Abbildung 10): Ein Großteil der Gebiete, in welchen die maximal gemessene Bodenbewegung später eintrat als dies bei ähnlich weit vom Erdbebenzentrum entfernten Gebieten der Fall war, ist sehr gebirgig. Zu betonen ist hierbei nochmals, dass pro Simulation von einer einheitlichen Bodenbeschaffenheit des kompletten Gebietes ausgegangen wurde. Mögliche unterirdische Reflektionen bzw. Abschwächungen der Druckwellen aufgrund unterschiedlicher Bodenschichten sind hier also nicht möglich!

Unabhängig von den untersuchten Ausbreitungsgeschwindigkeiten von Erdbeben in die verschiedenen Richtungen ist eine wichtige Auffälligkeit in Abbildung 9, dass das mit rot markierte Epizentrum der betrachteten Simulation bezogen auf die Grafiken nicht komplett zentral liegt und dass das zugehörige Hypozentrum gemittelt über alle Simulationen somit wohl nicht dem Punkt auf der Bruchfläche entspricht, an welchem die größten Bewegungen stattfinden. Diese Vermutung lässt sich auch durch die Abbildungen 6 und 8 bestätigen (siehe Diskussionen bzgl. den Grafiken). Ein möglicher Grund für diesen Befund ist die Tatsache, dass die größten Bewegungen an der unterirdischen Bruchfläche abhängig vom genauen Verlauf des Bruches an einem beliebigen, im Vorhinein nicht trivial bestimmbar Punkt stattfinden. Im Gegensatz zum in den Simulationen festgesetzten Hypozentrum als Startpunkt des Bruches, welches sich über alle Simulationen an der gleichen Position befindet, variiert dieser Punkt dabei über die Simulationen.

Wie bei möglicherweise bestehenden unterschiedlichen Ausbreitungsgeschwindigkeiten der Erdbeben ist die genaue Lokalisierung des Gebietes auf der Bruchfläche mit der größten Bewegung für die Analysen prinzipiell sehr relevant. Insbesondere auf die Modellierung der nah am Erdbebenzentrum aufgezeichneten Bodenbewegungen könnte es sich möglicherweise positiv auswirken, wenn anstelle des Startpunktes des Bruches eben jener Punkt als Punktquelle verwendet wird. Für eine optimale Modellierung der Ausbreitung eines Erdbebens unter Annahme einer Punktquelle stellt dieser Punkt dabei die optimale Punktquelle dar. Ist dieser wie im aktuellen Fall nicht bekannt so sollte zumindest ein möglichst naher Alternativpunkt auf der Bruchfläche als Punktquelle angenommen werden.

Basierend auf der Betrachtung aller bis hierhin gezeigten Deskriptionen der Bodenbewegungen lässt sich insgesamt vermuten, dass der Punkt der größten mittleren Bodenbewe-

gung über alle Simulationen gesehen durchschnittlich wohl näher am (über die Simulationen variierenden) Punkt der größten auf der Bruchfläche stattfindenden Bewegung liegt als das wahre Epizentrum der Erdbeben. Abbildung 11 zeigt, dass der senkrecht auf die Bruchfläche projizierte Punkt der größten mittleren Bodenbewegung etwas mittiger auf der Bruchfläche liegt als das wahre Hypozentrum. Die getroffene Vermutung erscheint basierend auf dieser Beobachtung - und ohne Einbezug weiterer Informationen über die Bruchfläche - also durchaus realistisch. Insgesamt erscheint es ebenfalls realistisch, dass die Wahl einer alternativen Punktquelle über alle Simulationen gesehen zu einer Verbesserung der Modellierung beitragen könnte im Vergleich zu auf dem Hypozentrum basierenden Modellen. Letztendlich basieren die hier gerechneten Modelle jedoch alle auf dem Hypozentrum als Punktquelle, da insbesondere für kleine Erdbeben, bei welchen sich der Bruch nicht weit ausbreitet, nur für das Hypozentrum sichergestellt ist, dass die Punktquelle auch wirkliche Bodenbewegungen zur Folge hat. Um einen besseren Eindruck über die Auswirkung dessen zu ermöglichen, welche Punktquelle letztendlich verwendet wird, wird das finale, auf dem Hypozentrum basierende Modell in Kapitel 5.3.4 demselbigen Modell basierend auf einer alternativen Punktquelle gegenübergestellt. Als alternative Punktquelle wird dabei der senkrecht auf die Bruchfläche projizierte Punkt verwendet, für welchen die größte mittlere Bodenbewegung beobachtet wurde (vgl. Abbildung 6).

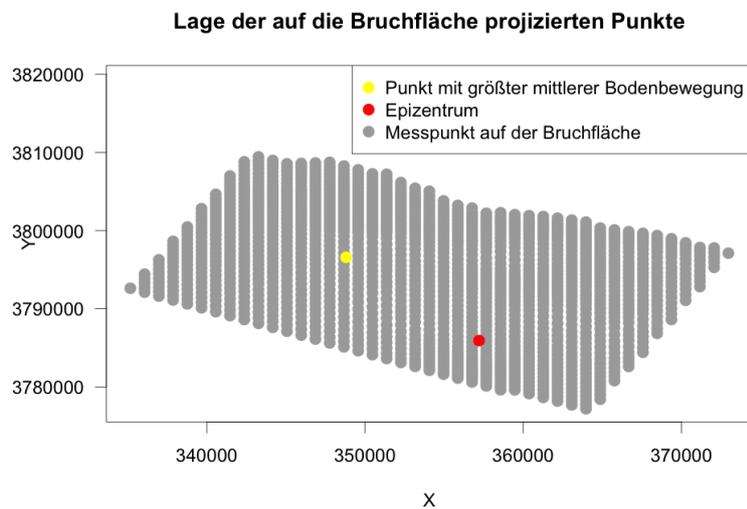


Abbildung 11: Darstellung der Lage des senkrecht auf die Bruchfläche projizierten Punktes der größten mittleren Bodenbewegung sowie des Hypozentrums (senkrecht auf die Bruchfläche projiziertes Epizentrum) auf der Bruchfläche

3 Kategorisierung der Topographie

Zur Einbeziehung der lokalen Topographie an den Seismographen-Messstellen - welche in den Modellen als Kontrollvariable auf einen möglichen Effekt hin überprüft werden soll - ist es sinnvoll, zusätzlich zu den vorliegenden rohen Topographie-Daten, welche die Ost- und Nordwert-Koordinaten sowie die Höhe der Punkte auf einem hochaufgelösten Gitter enthalten, weitere Informationen aus den Daten zu berechnen und in die Analysen einzubeziehen. Ziel dabei ist, für jeden Messpunkt angeben zu können, ob dieser beispielsweise in einem Tal, auf einer Ebene oder auf einem Berggipfel liegt. Hierzu werden die Punkte jeweils anhand ihrer Höhe mit allen Punkten in einer sinnvoll zu wählenden Nachbarschaft verglichen und darauf basierend in verschiedene Kategorien klassifiziert. Im Folgenden wird nun zuerst die grundsätzliche Methodik eingeführt, durch welche die Kategorisierung der Punkte in verschiedene Landform-Klassen vorgenommen wird. Im Anschluss daran wird eine für die aktuelle Anwendung geeignete Definition der verwendeten Nachbarschaften diskutiert.

3.1 Topographic Position Index

Für die Klassifizierung der Topographie an den einzelnen Messstationen wird die auf dem *Topographic Position Index* basierende Kategorisierung verwendet, wie sie von Weiss (2001) eingeführt wurde. Der TPI eines bestimmten Ortes gibt als metrischer Score dabei an, wie hoch die Differenz zwischen der Höhe des betrachteten Punktes und der mittleren Höhe aller in der Nachbarschaft liegenden Punkte ist. Für einen Punkt s lässt sich der TPI bzgl. einer Nachbarschaft NB folglich berechnen durch

$$TPI(s) = z(s) - \text{mean}\{z(r) | r \in NB\}, \quad (3.1)$$

wobei $z(r)$ die Höhenkoordinate eines Ortes r angibt.

Als Nachbarschaft werden üblicherweise alle Punkte gewählt, welche maximal eine gewisse Distanz vom betrachteten Punkt s entfernt sind. Dass die Wahl einer passenden Nachbarschaft großen Einfluss auf die Qualität der letztendlichen Kategorisierung der Topographie hat wird in Abbildung 12 verdeutlicht. Je größer die dem TPI zugrunde liegende Nachbarschaft gewählt wird, desto größer wird gleichzeitig die Topographie des betrachteten Gebietes abgebildet. Abhängig von den Geländeformen der konkret betrachteten Gebiete können dabei unterschiedlich grobe Klassifizierungen und somit die Wahl unterschiedlich großer Nachbarschaften sinnvoll sein. Die für die vorliegende Anwendung geeignete Nachbarschaft wird in Abschnitt 3.2 diskutiert.

Um die Topographie eines Ortes möglichst genau abbilden zu können werden pro betrachtetem Ort zwei TPI-Werte berechnet. Neben einem *small-scale* TPI, welcher basierend auf

einem eher geringen Radius berechnet wird und wodurch die sehr lokale Topographie des Gebietes abgebildet werden kann, wird zusätzlich noch ein *large-scale* TPI verwendet, wofür ein etwas größerer Radius verwendet wird. In der Anwendung von Weiss (2001) stellten sich beispielsweise ein small-scale Radius von 300m und ein large-scale Radius von 2000m als am geeignetsten heraus. Grundsätzlich ist die Idee dabei, dass eine Verwendung zweier verschieden großer Nachbarschaften sowohl dem entgegenwirkt, dass man eine zu kleinteilige, lokale Topographie betrachtet, als auch dass lediglich die grobe Topographie eines Gebietes in die Analysen einfließt, wodurch ebenfalls wichtige topographische Informationen verloren gehen könnten.

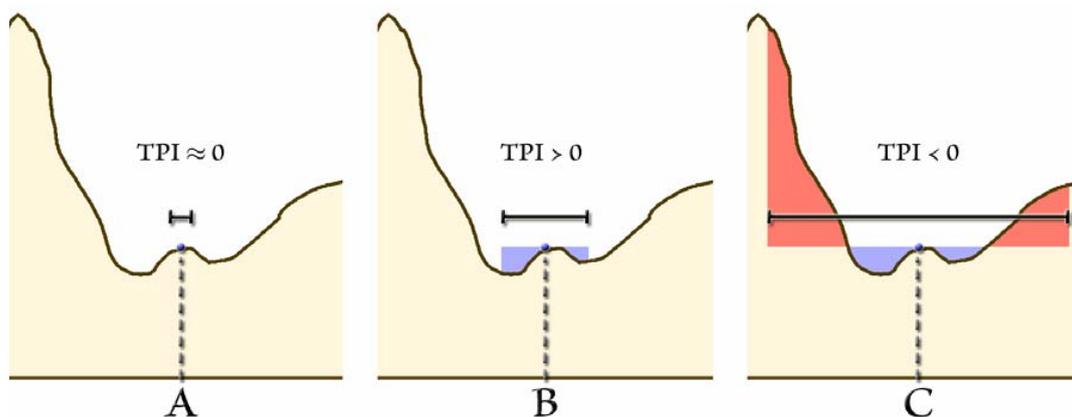


Abbildung 12: Auswirkung der Nachbarschaftswahl auf die TPI-Klassifizierung: Bei einer sehr kleinen Nachbarschaft wird der betrachtete Punkt als Ebene klassifiziert (Fall A), eine etwas größere Nachbarschaft führt zu einer Klassifizierung als Gipfel (Fall B). Bei einer sehr großen Nachbarschaft wird der Punkt dagegen als Tal klassifiziert (Fall C). Quelle: Jenness (2006)

Zur Bildung der Grenzen für die Klassifizierung der auf obige Weise berechneten metrischen TPIs werden jeweils die Standardabweichungen des small-scale bzw. des large-scale TPI verwendet. Dies hat den Vorteil, dass es als datenbasierte Wahl der Kategorisierungsgrenzen eine flexible Anpassung der Grenzen an die für eine Anwendung konkret vorliegende Datensituation bietet. Diese Flexibilität ist notwendig, da zu erwarten ist, dass sich die Wertebereiche des TPI nicht nur je nach gewähltem Radius der Nachbarschaft, sondern darüber hinaus auch je nach der Topographie des betrachteten Gebietes unterscheiden werden. Durch die datenbasierte Wahl der Grenzen wird außerdem der Anwender dahingehend entlastet, als dass sich dieser auf die Wahl sinnvoller Radien konzentrieren kann, ohne gleichzeitig die Grenzen selbst festsetzen zu müssen.

Insgesamt führt die Klassifizierung der metrischen TPI-Werte anhand Standardabweichungs-Einheiten zu einer fast komplett eindeutigen Klassifizierung. Der einzig uneindeutige Fall liegt vor, wenn sowohl der small-scale TPI als auch der large-scale TPI in ihrem jeweiligen Intervall $(-SE, SE)$ liegen. Dieser Fall kommt sowohl vor, wenn das be-

trachtete Gebiet eine Ebene ist, als auch wenn das Gebiet ein offener Hang ist, also eine konstante Steigung aufweist, da die mittlere Höhe der Nachbarschaft in diesem Fall ebenfalls ähnlich der Höhe des betrachteten Punktes ist. Um zwischen einer Ebene und einem offenen Hang zu unterscheiden wird in diesen Fällen deshalb die Steigung des Gebietes als zusätzliches Kriterium herangezogen. Beträgt die Steigung maximal 5° , so wird der betrachtete Punkt als Ebene klassifiziert, andernfalls als offener Hang.

Zur Berechnung der Steigung eines Gebietes werden im ersten Schritt die Steigungen in Grad zwischen dem betrachteten Punkt und allen sich in der Nachbarschaft befindlichen Punkten berechnet. Im zweiten Schritt wird anschließend das arithmetische Mittel der Absolutbeträge aller auf diese Weise berechneten Steigungen gebildet. Insgesamt ergibt sich dadurch die Formel

$$\text{Steigung}(s) = \text{mean} \left\{ \text{abs} \left(\frac{180^\circ}{\pi} \tan^{-1} \left(\frac{z(t) - z(s)}{\sqrt{(x_t - x_s)^2 + (y_t - y_s)^2}} \right) \right) \mid t \in NB \right\}. \quad (3.2)$$

Da in die Klassifizierung wie beschrieben letztendlich sowohl ein small-scale TPI als auch ein large-scale TPI einfließen muss erwähnt werden, dass für die Berechnung der Steigung immer die large-scale Nachbarschaft verwendet wird.

Eine Übersicht über die komplette TPI-basierte Klassifizierung findet sich in Tabelle 5. Abbildung 13 bietet eine Visualisierung der durch die Klassifizierung erzielten topographischen Kategorien.

Kategorie	small-scale TPI	large-scale TPI	Steigung
Tiefes Tal		$\leq -SE$	-
Lokales Tal in Ebene/Hang	$\leq -SE$	$\in (-SE, SE)$	-
Lokales Tal in Hochebene		$\geq SE$	-
U-förmiges Tal		$\leq -SE$	-
Ebene	$\in (-SE, SE)$	$\in (-SE, SE)$	$\leq 5^\circ$
Offener Hang		$\in (-SE, SE)$	$> 5^\circ$
Flacher Gipfel		$\geq SE$	-
Lokaler Gipfel in Tal		$\leq -SE$	-
Lokaler Gipfel in Ebene/Hang	$\geq SE$	$\in (-SE, SE)$	-
Steiler Gipfel		$\geq SE$	-

Tabelle 5: TPI-basierte Klassifizierung nach Weiss (2001); Die Klassifizierung erfolgt anhand von small-scale- und large-scale-spezifischen Standardabweichungs-Einheiten (SE) sowie anhand der mittleren Steigung in der large-scale Nachbarschaft

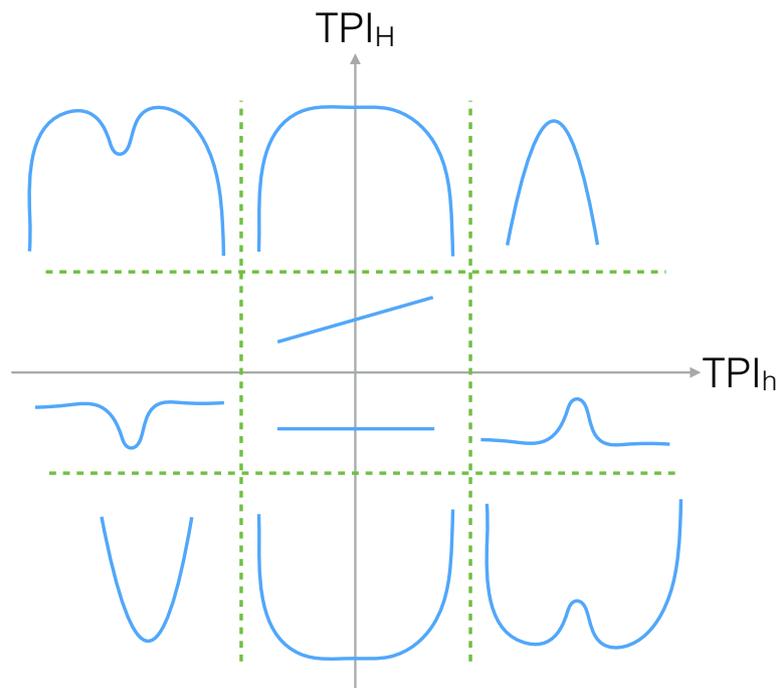


Abbildung 13: Visualisierung der in Tabelle 5 dargestellten Klassifizierung anhand des small-scale TPI TPI_h und des large-scale TPI TPI_H

3.2 Wahl geeigneter Nachbarschaften

Elementar für eine gute Klassifizierung der Geländeformen mittels der TPI-basierten Methodik ist die Definition sinnvoller Nachbarschaften. Wie im vorigen Abschnitt erläutert wurde wird für die letztendliche Klassifizierung dabei sowohl eine kleinere *small-scale* Nachbarschaft als auch eine größere *large-scale* Nachbarschaft benötigt, bzgl. welchen die TPI-Werte eines Ortes berechnet werden. Die Nachbarschaften werden jeweils mittels eines Radius definiert, welcher die maximale Entfernung angibt, die ein Punkt vom betrachteten Ort entfernt sein darf, um noch in der Nachbarschaft zu liegen.

Je nach Art der Topographie kann es sinnvoll sein unterschiedliche Radien als Grundlage für die Klassifizierung zu wählen. Weiss (2001) wählte für seine Anwendung die Radien 300m und 2000m. Im folgenden Abschnitt werden einige verschiedene Radien bzgl. der Güte ihrer Klassifizierung miteinander verglichen, wodurch schließlich die für die vorliegende Anwendung geeignetsten Nachbarschaften definiert werden können. Da die Topographie des hier betrachteten Gebietes sehr ähnlich zu der Topographie in Weiss (2001) ist - beide Gebiete umfassen eine gebirgige Region mit Höhenprofilen zwischen 0m und knapp über 3000m über NN - werden dabei ausschließlich Radien in die Evaluierung miteinbezogen, welche ähnlich den von Weiss verwendeten Radien sind. Als Radien für die small-scale Nachbarschaft werden deshalb 200m, 300m und 400m in den Vergleich mit-

einbezogen, bzgl. der Definition der large-scale Nachbarschaft dagegen 1000m, 2000m sowie 3000m.

Die Güte der Klassifizierung unter Benutzung der verschiedenen Radien wird auf einem Testgebiet evaluiert, welches ein 10×10 km großer Ausschnitt aus dem dieser Arbeit zugrunde liegenden Gebiet ist. Das Gebiet wurde dahingehend ausgewählt, dass es alle unterschiedlichen Geländeformen umfasst, welche in den Daten vorkommen - sowohl hohe Bergketten, als auch tiefe Täler und Übergänge von einem gebirgigen Gebiet hin zu einer Ebene sind im Evaluierungsgebiet enthalten. Eine Darstellung des verwendeten Evaluierungsgebietes sowie ein graphischer Vergleich der Klassifizierungen basierend auf den unterschiedlichen Radien findet sich in Abbildung 14. Verglichen werden hierbei alle möglichen Kombinationen der oben definierten small-scale und large-scale Radien.

Vergleicht man die Klassifizierungen basierend auf den verschiedenen small-scale Radien von 200m, 300m sowie 400m miteinander so lässt sich erkennen, dass bei einem kleineren verwendeten Radius die Klassifizierung insgesamt deutlich kleinteiliger wird. Unter Verwendung eines small-scale Radius von 400m lassen sich dadurch deutlich klarere Strukturen erkennen als unter der Verwendung von 200m als Radius. Alles in allem sind die erkennbaren Unterschiede zwischen den verschiedenen Radien jedoch relativ gering und es kann keine klare Aussage darüber getroffen werden, welcher der Radien das Gebiet besser oder schlechter klassifiziert. Auf dieser Grundlage wird letztendlich der 300m Radius als mittlere Kategorie verwendet. Dies erscheint auch sinnvoll, da Weiss (2001) wie erläutert für ein topographisch ähnliches Gebiet ebenfalls diesen small-scale Radius als geeignetsten Radius verwendet hat.

Deutlichere Unterschiede sind beim Vergleich der Klassifizierungen basierend auf den verschiedenen large-scale Radien zu erkennen. Bei Betrachtung der Klassifizierungen unter Verwendung von 1000m als large-scale Radius fällt negativ auf, dass auf dem Gebirgsgrat - welcher ausgehend vom Westen in nordöstlicher Richtung verläuft - verglichen mit der wahren Topographie des Evaluierungsgebietes deutlich zu viele Gipfelstellen als offener Hang klassifiziert werden. Darüber hinaus werden im nordwestlich verlaufenden Tal zu viele Talstellen ebenfalls als offener Hang klassifiziert. Bei der Klassifizierung basierend auf einem Radius von 3000m wird der Gebirgsgrat sinnvoll abgebildet. Im nordwestlichen Tal verkehrt sich das Problem, welches bei der Verwendung von einem 1000m Radius auffällt, in die andere Richtung, d.h. es werden nun zu viele Punkte als Tal klassifiziert, welche eigentlich einem offenen Hang entsprechen. Nimmt man den large-scale Radius von 2000m als Grundlage für die TPI-basierte Klassifizierung der Topographie, so werden sowohl der Gebirgsgrat als auch das Tal sinnvoll klassifiziert. Die Gebiete, welche im Richtung Süden verlaufenden Übergang vom Gebirge in die Ebene liegen, werden von allen drei large-scale Radien auf sinnvolle Weise klassifiziert.

Basierend auf den obigen Ausführungen wird die realitätsnächste Klassifizierung des Evaluierungsgebietes durch die Verwendung eines small-scale Radius von 300m sowie eines large-scale Radius von 2000m erreicht. Dies bestätigt die Radienwahl von Weiss (2001) basierend auf dessen topographisch ähnlichem Gebiet.

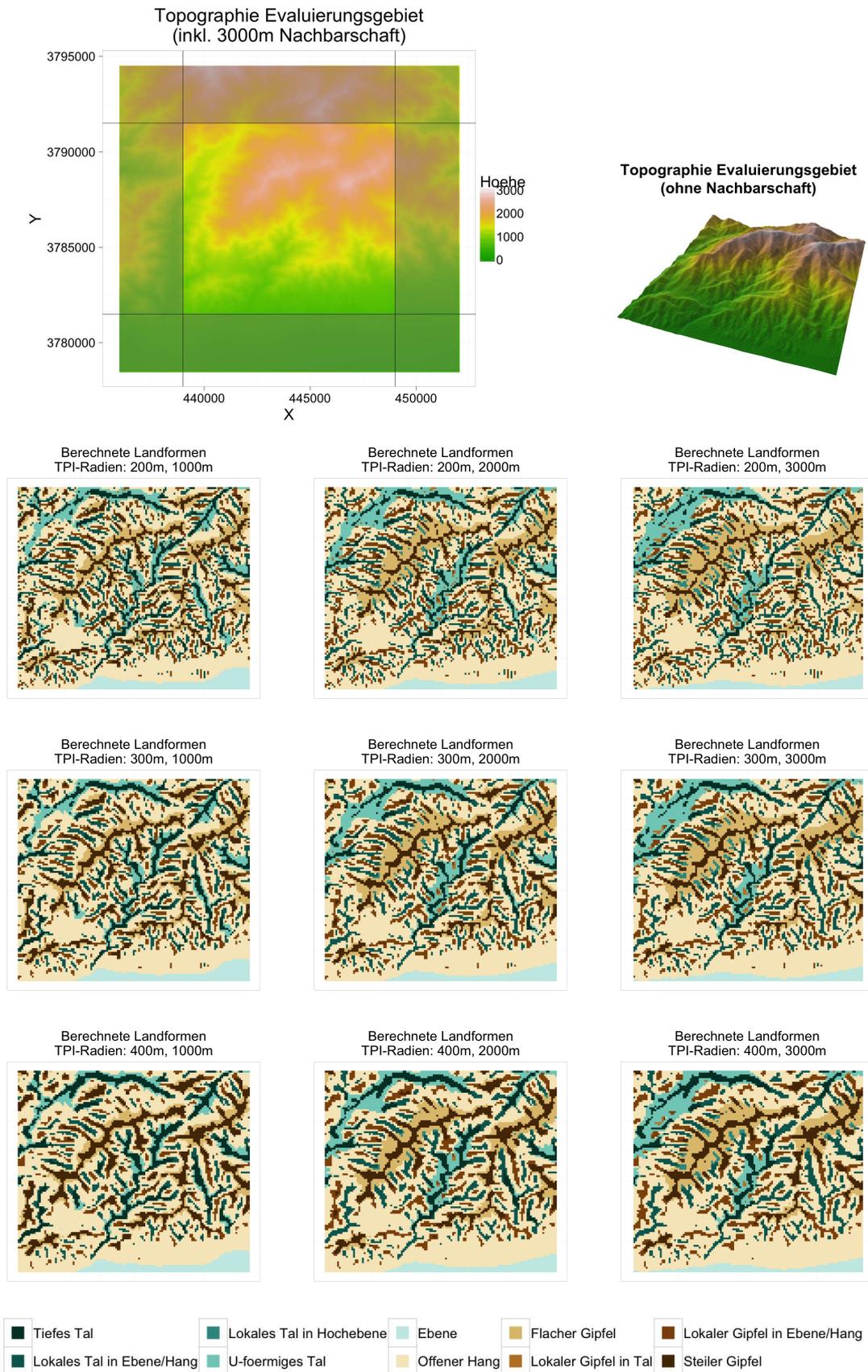


Abbildung 14: Vergleich der TPI-Klassifizierungen unter Benutzung verschiedener small-scale und large-scale Radien anhand eines 10 × 10km großen Evaluierungsgebietes

4 Funktionale Regression

Das multiple Regressionsmodell basiert unter anderem auf der Annahme, dass alle Messungen bzw. Beobachtungen von einzelnen Ziel- oder Einflussgrößen skalar sind. Wie bei den in Kapitel 2 beschriebenen Messungen der Bodenbewegung erläutert wurde ist es hier jedoch sinnvoll, die an einzelnen Untersuchungseinheiten durchgeführten Einzelmessungen nicht als (korrelierte) skalare Beobachtungen aufzufassen, sondern jeweils als *eine* funktionale Beobachtung. Implizit wird dadurch angenommen, dass hinter den gemessenen Datenpunkten eine glatte Funktion steht, welche nur auf einem diskreten (in diesem Fall zeitlichen) Gitter sowie nur mit einem Messfehler erhoben werden kann. Wird diese Annahme getroffen, so ist eine Anwendung der rein skalaren multiplen Regression nicht mehr möglich. Stattdessen muss das Modell je nach vorliegender Datensituation an entsprechenden Stellen erweitert werden, um mit funktionalen Daten umgehen zu können. Bei der Modellierung von Regressionsmodellen unter dem Einbezug von funktionalen Daten sind drei Fälle zu unterscheiden. Diese sind danach definiert, welche Rolle die funktionalen Größen in der Regression spielen: Bei der *Skalar-auf-Funktion* Regression ist die Zielgröße ein Skalar, unter den Einflussgrößen befinden sich aber neben skalaren Größen auch funktionale Variablen. Ist dagegen die Zielgröße funktionaler Art und die Einflussgrößen sind ausnahmslos skalar, so spricht man von einer *Funktion-auf-Skalar* Regression. Der dritte Fall, eine *Funktion-auf-Funktion* Regression, liegt dann vor, wenn die Zielgröße eine funktionale Größe ist und sich unter den Einflussgrößen ebenfalls funktionale Variablen befinden (Morris 2014).

Im Anwendungsteil dieser Arbeit werden Daten modelliert, bei welchen der Response als funktionale Zielgröße aufgefasst wird. Die Kovariablen hingegen sind skalar. Da für diese Arbeit folglich nur der Fall der *Funktion-auf-Skalar* Regression relevant ist wird auf die anderen beiden Regressionsmethoden nicht weiter eingegangen. Des Weiteren wird aufgrund der praktischen Anwendung dieser Arbeit explizit nur vom *zeitlich* funktionalen Charakter der Daten gesprochen. Eine Anwendung der Modelle auf funktionalen Daten, welche über andere Dimensionen als der Zeit erhoben wurden ist selbstredend jedoch auch möglich.

Die im folgenden Kapitel erläuterte Schätzmethode für funktionale Regression basiert auf Scheipl, Staicu und Greven (2015) sowie Scheipl, Gertheiss und Greven (2016), welche die in der verwendeten *pffr*-Funktion aus dem R-Paket *refund* (Goldsmith et al. 2016) implementierte Methodik erläutern. Darüber hinaus wird eine von Wood et al. (2016) entwickelte und im *mgcv*-Paket (Wood 2006) implementierte besonders effiziente Schätzmethode erläutert, welche später bei der Berechnung der Regressionsmodelle verwendet wird.

4.1 Notation

Die in dieser Arbeit verwendete Notation orientiert sich an gängigen notationellen Praktiken aus der Statistik. So werden Zufallsvariablen durch Großbuchstaben und die beobachteten Ausprägungen von Zufallsvariablen durch Kleinbuchstaben gekennzeichnet. Handelt es sich um Vektoren, werden diese im Gegensatz zu Skalaren fett gedruckt. Matrizen werden durch fett gedruckte Großbuchstaben dargestellt.

Der wichtigste konzeptionelle Unterschied bei der Arbeit mit Daten mit funktionalem Charakter im Vergleich zu rein skalaren Daten ergibt sich dadurch, dass mehrere Messungen über die Zeit, welche an einer Untersuchungseinheit erhoben wurden, als *eine* Beobachtung aufgefasst werden und nicht als ein Vektor mehrerer Einzel-Beobachtungen. Analog zur oben eingeführten Notation werden einzelne Funktionen ebenfalls durch nicht fett gedruckte Kleinbuchstaben gekennzeichnet. Die Unterscheidung, ob es sich bei der verwendeten Schreibweise in einem konkreten Fall um einen Skalar oder um eine Funktion handelt wird jeweils durch den Kontext deutlich gemacht. Vektoren von Funktionen werden durch fett gedruckte Kleinbuchstaben dargestellt.

Soll auf die skalare Ausprägung einer Funktion x zu einem bestimmten Zeitpunkt referenziert werden, so wird die Notation x_l verwendet, wobei $x_l = x(t_l)$ der Beobachtung von x zum l -ten Messzeitpunkt entspricht ($l \in \{1, \dots, T\}$). Da die dieser Arbeit zugrunde liegenden Daten balanciert sind, ist die Anzahl an Messzeitpunkten T dabei über alle Beobachtungen konstant und alle Beobachtungen basieren auf Erhebungen an den gleichen Messzeitpunkten. Über die n funktionalen Beobachtungen ergeben sich dadurch insgesamt $N = nT$ skalare Einzel-Messungen.

4.2 Funktion-auf-Skalar Regression

Das allgemeine additive Modell der Funktion-auf-Skalar Regression im Fall einer Normalverteilungsannahme für den Response sowie unter Verwendung des Identitätslinks lässt sich auf folgende Weise darstellen (Scheipl, Staicu und Greven 2015):

$$y_{il} = \sum_{r=1}^R f_r(\mathbf{x}_{ri}, t_{il}) + \varepsilon_{il} \quad (4.1)$$

Jeder Term des additiven Prädiktors enthält dabei mit \mathbf{x}_{ri} eine Teilmenge der in der kompletten Designmatrix \mathbf{X} enthaltenen Kovariablen. Neben eindimensionalen glatten Effekten einer einzelnen Kovariable ist auf diese Weise auch die Einbeziehung glatter Interaktionseffekte möglich. Durch die Abhängigkeit der Funktionen f_r von der Zeitkomponente t_{il} wird darüber hinaus der prinzipiell mögliche Einbezug zeitvariierender Effekte ermöglicht.

Um obiges Modell zu motivieren werden nur vergleichsweise geringe Anpassungen an das rein mit Skalaren arbeitende additive lineare Regressionsmodell benötigt. Alle Modellkomponenten werden dabei durch den zeitlichen Index l erweitert: y_{il} und ε_{il} entsprechen dadurch der i -ten funktionalen Beobachtung y_i bzw. dem i -ten funktionalen Residuum ε_i evaluiert zum Zeitpunkt t_l . Die glatten Effekte f_r werden im Vergleich zum skalaren additiven Modell lediglich um eine mögliche Interaktion mit der Zeit erweitert, wodurch die Schätzung zeitvariierender Effekte hervorgehoben wird. Die Schätzung einfacherer linearer Effekte beispielsweise der Form $f_r(\mathbf{x}_{ri}, t) = \mathbf{x}_{ri}\beta$ bzw. $\mathbf{x}_{ri}\beta(t)$ ist mit Modell 4.1 ebenfalls möglich.

Eine Verallgemeinerung des obigen Modells auf den generalisierten Fall, in welchem nicht notwendigerweise von einer Normalverteilung der Zielgröße sowie einer Verwendung des Identitätslinks ausgegangen wird, führt zu einem Modell der Form

$$y_{il} \sim F(\mu_{il}, \nu)$$

$$g(\mu_{il}) = \sum_{r=1}^R f_r(\mathbf{x}_{ri}, t_{il}), \quad (4.2)$$

in welchem F eine beliebige Verteilung und μ_{il} den konditionalen Erwartungswert dieser Verteilung darstellen. ν ist darüber hinaus ein optionaler Vektor von Störparametern, welcher beispielsweise die Varianzparameter der betrachteten Verteilung enthält (Scheipl, Gertheiss und Greven 2016).

4.3 Schätzung

Bevor auf die genaue Methodik zur Schätzung in funktionalen Regressionsmodellen eingegangen wird ist es notwendig, eine alternative, auf Tensorprodukten basierende Darstellung der in Modell 4.2 enthaltenen glatten Effekte $f_r(\mathbf{x}_{ri}, t_{il})$ einzuführen, insbesondere um die Schätzung von ein- und mehrdimensionalen Effekten in einem einheitlichen Framework zu ermöglichen. Die Darstellung erfolgt dabei analog zur Darstellung glatter Effekte in Generalisierten additiven Modellen.

Tensorprodukt-Darstellung der glatten Effekte

Eindimensionale glatte Effekte - welche in der Funktion-auf-Skalar Regression auftreten, falls der Effekt einer Kovariablen über den Wertebereich derselbigen variieren soll bzw. falls ein einfacher zeitvariierender Effekt geschätzt werden soll - werden wie üblich als Linearkombinationen von Basisfunktionen repräsentiert. Dadurch wird erreicht, dass die Problematik der Schätzung nichtlinearer Effekte auf direkte Weise in ein Schätzproblem überführt wird, welches ausschließlich lineare Koeffizienten enthält. Bei der Verwendung

von Tensorprodukt-Basen, wie es hier der Fall ist, lässt sich die Repräsentierung mehrdimensionaler Effekte auf direkte Weise aus der Aufstellung solcher eindimensionaler Basis-Systeme verallgemeinern. Grundlage einer mehrdimensionalen Basis bildet dabei die Aufstellung von marginalen, auf den Wertebereichen der univariaten Kovariablen definierten Spline-Basen. Durch die Verknüpfung mehrerer univariater Spline-Basen auf die im Folgenden vorgestellte Art wird eine multivariate Spline-Basis definiert, wobei durch die Art der Verknüpfung mittels Tensorprodukten insbesondere sichergestellt wird, dass der komplette mehrdimensionale Wertebereich adäquat abgedeckt ist.

Der zentrale Punkt bei der flexiblen Darstellung mehrdimensionaler Effekte durch mehrere marginale Basis-Systeme ist die Bildung von zeilenweisen Tensorprodukten, wie es in Wood (2006, Kap. 4.1.8) eingeführt wird. Definiert man $\mathbf{1}_d = (1, \dots, 1)^T$ als einen Spaltenvektor der Länge d , \otimes als das Kroneckerprodukt sowie \cdot als den Operator zur elementweisen Multiplikation zweier Matrizen, so lässt sich das zeilenweise Tensorprodukt zweier Matrizen \mathbf{A} (Dimension $m \times a$) und \mathbf{B} ($m \times b$) definieren als $\mathbf{A} \odot \mathbf{B} = (\mathbf{A} \otimes \mathbf{1}_b^T) \cdot (\mathbf{1}_a^T \otimes \mathbf{B})$. Ein zweidimensionaler Effekt einer skalaren Kovariable, welcher sowohl über den Wertebereich der Kovariable als auch über die zeitliche Dimension variiert, lässt sich dadurch beispielsweise darstellen als

$$f_r(\mathbf{x}_r, t) = (\Phi_{xr} \odot \Phi_{tr})\boldsymbol{\theta}_r = \Phi_r\boldsymbol{\theta}_r, \quad (4.3)$$

wobei Φ_{xr} (Dimension $N \times K_{xr}$) die evaluierten marginalen Basisfunktionen der Kovariable enthält und Φ_{tr} ($N \times K_{tr}$) die evaluierten marginalen Basisfunktionen, welche auf dem Wertebereich der Zeit t definiert sind. K_{xr} und K_{tr} entsprechen jeweils der Anzahl an verwendeten Basisfunktionen. Der Spaltenvektor $\boldsymbol{\theta}_r$ enthält die Koeffizienten, mit welchen die Basisfunktionen skaliert werden und welche in der Regression letztendlich zu schätzen sind.

Der entsprechende Penalisierungsterm zur durch obige Weise aufgestellten multivariaten Basis lässt sich ebenfalls mithilfe des Kroneckerproduktes als Zusammensetzung aus den Penalisierungstermen der verwendeten marginalen Basen definieren. Betrachtet man erneut das Beispiel des zweidimensionalen Effektes aus 4.3, so ergibt sich der Penalisierungsterm der bivariaten Basis als

$$\begin{aligned} \text{pen}(\boldsymbol{\theta}_r | \lambda_{xr}, \lambda_{tr}) &= \boldsymbol{\theta}_r^T \mathbf{P}_r(\lambda_{xr}, \lambda_{tr}) \boldsymbol{\theta}_r \\ \text{mit } \mathbf{P}_r(\lambda_{xr}, \lambda_{tr}) &= \lambda_{xr} \mathbf{P}_{xr} \otimes \mathbf{I}_{K_{tr}} + \lambda_{tr} \mathbf{I}_{K_{xr}} \otimes \mathbf{P}_{tr}. \end{aligned} \quad (4.4)$$

\mathbf{P}_{xr} sowie \mathbf{P}_{tr} stellen dabei die positiv semi-definiten Penalisierungsmatrizen der jeweiligen marginalen Basen dar. λ_{xr} und λ_{tr} sind die entsprechenden Glättungsparameter.

Schätzung

Grundlage für die Parameterschätzung ist das Aufstellen der penalisierten Log-Likelihood. Definiert man $\mathbf{y} = (y_1^T, \dots, y_n^T)$ und $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \dots, \boldsymbol{\theta}_R^T)^T$ als Spaltenvektoren, welche alle Messungen bzw. Koeffizienten enthalten, $\boldsymbol{\lambda} = (\lambda_{x1}, \lambda_{t1}, \dots, \lambda_{xR}, \lambda_{tR})$ als den Vektor aller Glättungsparameter, $\Phi = [\Phi_1 | \dots | \Phi_R]$ als die Matrix aller nebeneinander geschriebenen Φ_r und $\boldsymbol{\mu} = g^{-1}(\Phi\boldsymbol{\theta})$ entsprechend als den erwarteten Response-Vektor, so lässt sich die penalisierte Log-Likelihood darstellen als

$$\ell_p(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu} | \mathbf{y}) = \ell(\boldsymbol{\mu}, \boldsymbol{\nu} | \mathbf{y}) - \frac{1}{2} \sum_{r=1}^R \text{pen}(\boldsymbol{\theta}_r | \lambda_{xr}, \lambda_{tr}). \quad (4.5)$$

$\ell(\boldsymbol{\mu}, \boldsymbol{\nu} | \mathbf{y}) = \sum_{i=1}^n \sum_{l=1}^T \ell(\mu_{il}, \nu | y_{il})$ entspricht dabei der Log-Likelihood, welche aus der angenommenen Response-Verteilung $F(\boldsymbol{\mu}_{il}, \boldsymbol{\nu})$ resultiert.

Auf Basis der so aufgestellten Likelihood werden die Parameter anschließend mittels der von Wood (2011) entwickelten, REML-basierten Schätzmethode berechnet. Wie von Reiss und Ogden (2009) und Wood (2011) gezeigt wurde ist eine solche REML-basierte Schätzung der Glättungsparameter, bei der diese als Teil der Varianzkomponenten von zufälligen Effekten angesehen und mit einem REML-Ansatz geschätzt werden, einer Schätzung basierend auf einer Optimierung beispielsweise des GCV-Kriteriums vorzuziehen.

Die marginale REML-Likelihood, welche zur Schätzung der Glättungsparameter $\boldsymbol{\lambda}$ sowie der Störparameter $\boldsymbol{\nu}$ verwendet wird, ergibt sich durch Herausintegrieren von $\boldsymbol{\theta}$ aus Likelihood 4.5. Dieses Integral $\int \ell_p(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu} | \mathbf{y}) d\boldsymbol{\theta}$ wird mithilfe der Laplace-Approximation näherungsweise berechnet. Insgesamt ergibt sich schließlich die folgende approximative marginale REML-Likelihood:

$$\begin{aligned} \ell_{LA}(\boldsymbol{\lambda}, \boldsymbol{\nu} | \mathbf{y}) = & \ell(\tilde{\boldsymbol{\mu}}, \boldsymbol{\nu} | \mathbf{y}) - \frac{1}{2} \sum_{r=1}^R \left(\text{pen}(\tilde{\boldsymbol{\theta}}_r | \lambda_{xr}, \lambda_{tr}) - \log |P_r(\lambda_{xr}, \lambda_{tr})|^+ \right) \\ & - \frac{1}{2} \log |H_p| + \frac{1}{2} d_\theta \log(2\pi), \end{aligned} \quad (4.6)$$

mit H_p der Summe aus der negativen Hesse-Matrix $\mathbf{H} = -\partial^2 \ell_p(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu} | \mathbf{y}) / (\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T)$ der penalisierten Log-Likelihood 4.5 sowie einer die $P_r(\lambda_{xr}, \lambda_{tr})$ enthaltenden Block-Diagonalmatrix, $\tilde{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \ell_p(\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\nu} | \mathbf{y})$ dem für feste $\boldsymbol{\lambda}$ und $\boldsymbol{\nu}$ im aktuellen Schritt optimalen $\boldsymbol{\theta}$ -Schätzer, $\tilde{\boldsymbol{\mu}} = g^{-1}(\Phi \tilde{\boldsymbol{\theta}})$, $|\mathbf{A}|^+$ der generalisierten Determinante, welche dem Produkt aller positiven Eigenwerte von \mathbf{A} entspricht, sowie mit d_θ der Summe der Rangabfälle, also der Dimension der Nullräume der Penalisierungsmatrizen $P_r(\lambda_{xr}, \lambda_{tr})$, $r = 1, \dots, R$. Im Normalverteilungsmodell mit Identitätslink entspricht obige Likelihood ex-

akt der REML-Likelihood.

Basierend auf der approximativen marginalen Log-Likelihood 4.6 wird anschließend ein iterativer Schätzalgorithmus angewandt, um λ und ν zu bestimmen. Diese von Wood (2011) entwickelte, numerisch stabile Schätzmethode basiert dabei auf der Optimierung der marginalen (REML-)Likelihood. Pro Iterationsschritt wird jeweils zuerst $\tilde{\theta}$ gegeben dem aktuellen λ durch eine *Penalized Iteratively Reweighted Least Squares* (PIRLS)-Schätzung bestimmt und anschließend wird die λ -Schätzung geupdated. Zusammengefasst werden bei der Schätzung der Glättungs- sowie der Störparameter pro Iterationsschritt die folgenden vier Schritte durchlaufen:

1. Schätzung von $\tilde{\theta}$ gegeben dem aktuellen λ per PIRLS-Schätzung
2. Berechnung des Gradienten sowie der Hesse-Matrix von $\ell_{LA}(\lambda, \nu | \mathbf{y})$ bzgl. $\log(\lambda)$ durch implizite Differentiation
3. Update von $\log(\lambda)$ per Newton-Raphson-Algorithmus
4. Schätzung von ν erfolgt entweder im Zuge der λ -Schätzung (Optimierung der marginalen Likelihood) oder basierend auf der Pearson-Statistik des gefitteten Modells (Wood 2006)

Implementiert ist die aufgeführte Schätzmethode im Paket `mgcv` und kann durch den Befehl `gam(..., method="REML")` aufgerufen werden. Die im Anwendungsteil dieser Arbeit verwendete Funktion `pffr` basiert auf der `gam`-Funktion und verwendet die obige Schätzmethode per default.

Anzumerken ist, dass der aufgeführte Iterationsalgorithmus nicht der per default verwendeten Schätzmethode in `gam(..., method="REML")` entspricht, in welcher die Glättungsparameter λ nicht in jeder Iteration geupdated werden, sondern welche jeweils den kompletten Algorithmus für eine Menge von davor festgelegten λ -Werten durchführt. Anschließend wird anhand des REML-Kriteriums das optimale λ bestimmt. Diese von `gam` per default verwendete „äußere“ Optimierung von λ wird hier jedoch nicht weiter aufgeführt, da die im Rahmen dieser Arbeit verwendete Schätzung ausschließlich auf der im folgenden Abschnitt eingeführten effizienteren Version des obigen Algorithmus basiert, welche die vorgestellte „innere“ Optimierung von λ anwendet (Wood 2006). Diese „innere“ Optimierung entspricht darüber hinaus dem „performance iteration“-Ansatz von Gu (1992) bzw. dem äquivalenten PQL-Ansatz nach Breslow und Clayton (1993) und wird in der R-Hilfe zu `?gam` entsprechend als *performance iteration*-Methode bezeichnet.

Auch soll angemerkt werden, dass sich die oben beschriebene, in der `pffr`-Funktion verwendete Methode zur Schätzung von Funktion-auf-Skalar Modellen von jenem Ansatz unterscheidet, welcher im Standardwerk von Ramsay und Silverman (2005) beschrie-

ben wird. Ein wesentlicher Unterschied besteht dabei darin, dass Ramsey und Silverman alle funktionalen Einzel-Beobachtungen jeweils durch Spline-Basen repräsentieren, wodurch die Beobachtungen vor der Schätzung bereits zu einem gewissen Grad geglättet werden und Information verloren geht. Beim Ansatz von Scheipl, Gertheiss und Greven (2016) werden die funktionalen Beobachtungen dagegen ohne vorherige Glättung in die Schätzung einbezogen.

4.4 Effiziente Schätzung

Aufgrund der großen Datenmenge, auf Basis welcher die Modelle geschätzt werden sollen, liegt ein Hauptaugenmerk dieser Arbeit auf einer computationally effizienten Schätzung der funktionalen Regressionsmodelle. Wie in der Datenbeschreibung aufgeführt wurde, umfassen die kompletten Daten aus den Simulationen ca. 2TB. Betrachtet man die Seismogramme nicht mit voller zeitlicher Auflösung, sondern lediglich - wie in Kap. 2.4 beschrieben - die durch Downsampling verkleinerten Messungen, so umfassen die Seismogramm-Daten ca. 600MB bzw. 5GB. Die zeitliche Auflösung, welche in den 600MB-Daten verwendet wird, stellt dabei die minimal erforderliche Auflösung dar, um die untersuchten geophysikalischen Prozesse adäquat abbilden zu können.

Ein wesentlicher praktischer Vorteil der Schätzung von funktionalen Modellen, in der Art in der sie in der `pfrr`-Funktion implementiert ist, liegt darin, dass das funktionale Regressionsproblem lediglich als Spezialfall bzw. Erweiterung von Generalisierten additiven Modellen dargestellt wird und es dadurch möglich wird auf bestehende Forschungsergebnisse sowie Software-Implementierungen der GAM-Schätzalgorithmen aufzubauen. Im konkreten Fall der `pfrr`-Funktion, welche lediglich einen Wrapper der `gam`-Funktion aus dem `mgcv`-Paket darstellt, kann dadurch sowohl auf die bereits gut erforschten robusten Schätzalgorithmen als auch auf kürzlich für GAM-Modelle neu entwickelte effiziente Schätzmethoden zurückgegriffen werden.

Die effiziente Schätzmethode, welche im Rahmen dieser Arbeit verwendet wird, basiert auf einem ersten Ansatz aus Wood, Goude und Shaw (2015) und wurde in Wood et al. (2016) schließlich weiterentwickelt. Durch die Weiterentwicklung lassen sich GAMs - bzw. auf dem GAM-Framework aufbauende funktionale Regressionsmodelle - nun auf sehr großen Datenmengen (Wood spricht plakativ von „Gigadaten“) mit deutlich geringerem computacionalem Aufwand schätzen. Konkret handelt es sich bei der von Wood neu entwickelten iterativen Schätzmethode um eine Anpassung an die in Abschnitt 4.3 erläuterte Methode. Grundlage für eine effizientere Gestaltung des Schätzalgorithmus sowie dessen Implementierung in der `bam`-Funktion boten dabei verschiedene Ansatzpunkte.

Ein offensichtliches Problem bei der Arbeit mit sehr hochdimensionalen Daten ist der mit der Datenmenge steigende Speicherplatzbedarf, welcher für bestimmte Berechnungen benötigt wird. Bei der Schätzung von Regressionsmodellen - beispielsweise anhand der im vorigen Abschnitt beschriebenen Schätzmethode - bezieht sich dieser Punkt insbesondere auf die intern gebildete Modellmatrix, welche z.B. aufgrund von mehrkategorialen Variablen, welche im Modell durch Dummy-Variablen repräsentiert werden, oft auch deutlich größer sein kann als die eigentliche Datentabelle. Üblicherweise wird in Software-Implementierungen von Regressionsmodellen - und so auch in der `gam`-Funktion - die volle Modellmatrix gebildet. Bei sehr großen Datenmengen resultiert dies dann bereits bei vergleichsweise einfachen Modellen oft in sehr zeitintensiven Berechnungen. Um die Schätzung basierend auf solchen sehr umfangreichen Daten praktikabler zu gestalten umgeht Wood dieses Problem durch Benutzung einer blockweisen Cholesky-Zerlegung, welche die üblicherweise verwendete QR-Zerlegung ersetzt. Die Tatsache, dass es sich um eine blockweise arbeitende Methode handelt, ermöglicht die Schätzung von Modellen, ohne dass die volle Modellmatrix während der Schätzung gebildet werden muss. Positive Nebeneffekte dieser Methode sind darüber hinaus eine i.A. deutlich bessere Skalierung der Cholesky-Zerlegung im Vergleich zur QR-Zerlegung auf große Daten sowie die Parallelisierbarkeit der Methode, welche aus der blockweisen Anwendung resultiert.

Ein weiterer Ansatzpunkt zur effizienteren Gestaltung der Schätzung findet sich im pro Iterationsschritt der PIRLS-Schätzung durchgeführtem Update von $\log(\lambda)$. In der Methode von Wood (2011) wird für das Update jeweils der Newton-Raphson-Algorithmus verwendet, welcher bis zur Erreichung von Konvergenz durchlaufen wird. Dieser Update-Schritt wurde in Wood et al. (2016) dahingehend angepasst, dass jeweils lediglich der erste Schritt der Newton-Methode durchlaufen wird. Da die äußere PIRLS-Iteration - von welcher das Update von $\log(\lambda)$ Teil ist - weiterhin bis zur Konvergenz gerechnet wird lässt sich durch diese Maßnahme unnötiger Rechenaufwand vermeiden, ohne die Güte der Schätzung zu stark negativ zu beeinflussen.

Der dritte Teil der Anpassung der Schätzmethode, welcher auf eine Verbesserung der computationalen Effizienz der Methode abzielt, ist schließlich eine der Schätzung vorgeschaltete Diskretisierung aller in der Modellmatrix enthaltenen Spalten, auf Basis welcher ein- oder mehrdimensionale glatte Effekte geschätzt werden sollen. Dadurch wird erneut der zur Schätzung benötigte Speicherplatz verringert und bestimmte Operationen - wie beispielsweise die bei großen Datenmengen sehr rechenintensive Bildung bestimmter Kreuzprodukte - sind auf Basis der Diskretisierung weniger rechenaufwendig. Verwendet wird für die Diskretisierung ein von Wood et al. (2016) neu entwickelter Ansatz, welcher auf einer *marginalen* Diskretisierung der einzelnen Kovariablen aufbaut. Der Vorteil dieses marginalen Ansatzes wird bei der Verwendung von mehrdimensionalen glatten Ef-

fechten im Modell deutlich: Eine Diskretisierung des gemeinsamen, mehrdimensionalen Raums hat entweder den Nachteil, dass eine auf einer adäquaten Auflösung des Raums basierende Diskretisierung sehr speicherintensiv ist, oder dass - um den Speicherbedarf niedrig zu halten - eine zu grobe, ungenaue Diskretisierung des Raums verwendet wird. Die verwendete marginale Methode diskretisiert dagegen lediglich die einzelnen Kovariablen und bildet die mehrdimensionalen Basisräume entsprechender glatter Effekte erst im Rahmen der Schätzung und basierend auf den marginal diskretisierten Variablen.

Im Falle der Betrachtung eines univariaten glatten Effektes lässt sich der zugehörige Teil \mathbf{X}_j der vollen Modellmatrix - welcher für den aktuell betrachteten j -ten glatten Effekt die Spalten aller p_j evaluierten Basisfunktionen umfasst und somit der Matrix Φ_j entspricht - nach der Diskretisierung darstellen als

$$\mathbf{X}_j(i, l) = \bar{\mathbf{X}}_j(k_j(i), l). \quad (4.7)$$

Die Matrix $\bar{\mathbf{X}}_j$ hat dabei nur mehr die Dimension $m_j \times p_j$, wobei m_j der Anzahl an Kategorien entspricht, in welche die Variable diskretisiert wird, und k_j einen Indexvektor darstellt, durch welchen in der richtigen Reihenfolge auf die jeweiligen Matrixeinträge zugegriffen wird. Anstatt die volle Matrix \mathbf{X}_j zu speichern reicht es nach Anwendung der Diskretisierung folglich aus die $m_j \times p_j$ -Matrix $\bar{\mathbf{X}}_j$ sowie den entsprechenden Indexvektor zu speichern. Mehrdimensionale glatte Effekte, welche auf Tensorprodukten basieren, lassen sich darstellen als

$$\mathbf{X}_j = \Phi_0^j \odot \Phi_1^j \odot \dots \odot \Phi_{d_j-1}^j. \quad (4.8)$$

Die Φ_q^j entsprechen dabei den die evaluierten marginalen Basisfunktionen enthaltenden Ausschnitten aus der Modellmatrix, d_j beschreibt die Dimensionalität des glatten Effektes und der Operator \odot entspricht dem auf S.28 definierten zeilenweisen Tensorprodukt. Durch die Diskretisierung lassen sich in diesem mehrdimensionalen Fall die einzelnen marginalen Komponenten Φ_k^j sparsamer darstellen als

$$\Phi_q^j(i, l) = \bar{\Phi}_q^j(k_q^j(i), l). \quad (4.9)$$

Auf eine notationelle Einbeziehung von Restriktionsmatrizen wurde in der obigen Aufstellung verzichtet. Selbstredend ist es jedoch notwendig, die Matrizen \mathbf{X}_j jeweils mit einer solchen Restriktionsmatrix zu multiplizieren, um die Identifizierbarkeit des zu schätzenden Modells zu gewährleisten - üblicherweise indem die (mehrdimensionalen) glatten Effekte dadurch um Null zentriert werden.

In der Praxis lässt sich die auf den vorgestellten Anpassungen basierende effiziente Schätz-

methode von Wood et al. (2016) durch Verwendung der bam-Funktion (unter Angabe des Arguments `'discrete = TRUE'`) verwenden. Mit dieser Funktion lassen sich analog zur Schwesterfunktion gam ebenfalls Generalisierte additive Modelle fiten. In pffr lässt sich die bam-Schätzmethode intern durch das Argument `algorithm='bam'` aufrufen.

5 Anwendung

Neben den beiden Hauptfragestellungen, auf welche Art die Variablen die gemessenen Bodenbewegungen beeinflussen und auf welche Art sich die skalare Momentenmagnitude als zentraler Quellparameter aus den Simulationsparametern ergibt, werden im folgenden Kapitel zwei kleinere Nebenfragestellungen untersucht. Zum einen kam durch die deskriptiven Untersuchungen aus Kapitel 2.4 die Vermutung auf, dass die Bodenbewegungen bei Betrachtung von in einer Ebene sowie in einem gebirgigen Gebiet vorgenommenen Messungen bei ansonsten gleichen Bedingungen strukturell unterschiedlich ausfallen. Als zweite Hypothese wird überprüft, ob die Ausbreitung der simulierten Erdbeben möglicherweise nicht-isotrop ist und ein Richtungseffekt vorliegt.

Grundsätzlich besteht bei der Modellierung unter Einbezug sowohl der Simulations- als auch der Quellparameter als Einflussvariablen die Problematik, dass sich letztere kausal durch erstere ergeben. Zur Untersuchung des Einflusses der einzelnen Simulationsparameter auf die Bodenbewegungen muss folglich nicht nur der *direkte Effekt* dieser Parameter einbezogen werden, sondern auch die für die Quellparameter geschätzten Effekte stellen in gewissem Rahmen *indirekte Effekte* der Simulationsparameter dar. Um trotz dieser Problematik eine möglichst umfassende Einschätzung der *insgesamten* Effekte der Simulationsparameter geben zu können wird auf eine generelle Vorgehensweise der Pfadanalyse zurückgegriffen (Loehlin 2004). Zusätzlich zum vollen Modell zur Modellierung der Bodenbewegungen, welches sowohl die Simulations- als auch die Quellparameter als Kovariablen umfasst, wird dafür in der in Abschnitt 5.4 behandelten Fragestellung der kausale Einfluss der Simulationsparameter auf den zentralen Quellparameter mittels eines Regressionsmodells quantifiziert. Darüber hinaus wird das volle Modell zusätzlich unter Ausschluss des Quellparameters berechnet. Für die Praxis ist letzteres Modell insbesondere auch relevant, um eine gute Prognose der Bodenbewegungen auch in Situationen zu ermöglichen, in welchen der im Modell verwendete Quellparameter nicht bekannt ist.

Die in den Regressionsmodellen geschätzten glatten Effekte basieren durchweg auf kubischen P-Splines, wobei für die Penalisierung eine Differenzierung zweiten Grades verwendet wird. Die für glatte Terme angegebenen p-Werte basieren auf einem Wald-ähnlichen Test, welcher die bayesianisch hergeleitete Kovarianzmatrix der geschätzten Koeffizienten verwendet (Wood 2013).

Nachdem im nächsten Abschnitt einige wichtige Entscheidungen - sowohl in statistischer als auch inhaltlicher Hinsicht - motiviert wurden, welche im Rahmen der Modellierung getroffen wurden, wird in Abschnitt 5.2 der verwendete Ansatz zur Modellwahl vorgestellt und diese wird durchgeführt. Die Ergebnisse bezüglich der finalen Modelle der ersten Hauptfragestellung finden sich in Abschnitt 5.3. Am Ende wird dabei ebenfalls evaluiert, wie stark sich eine alternativ gewählte Punktquelle auf der Bruchfläche auf die

Ergebnisse beziehungsweise die Güte der Modelle auswirkt. Das als in allen übrigen Modellen als Punktquelle dienende Hypozentrum wird dabei mit dem in der Deskription eingeführten, senkrecht auf die Bruchfläche projizierten Punkt der größten mittleren Bodenbewegung als alternative Punktquelle verglichen. In Kapitel 5.4 wird untersucht, auf welche Art die Simulationsparameter auf die skalare Momentenmagnitude als zentralen Quellparameter wirken. Die beiden Nebenfragestellungen werden in Abschnitt ?? analysiert. Der den Auswertungen zugrunde liegende R-Code findet sich im digitalen Anhang B.

5.1 Vorbereitungen

Die als Zielgröße dienenden Bodenbewegungen wurden in der Modellierung grundsätzlich nicht ab dem zeitlichen Simulationsbeginn betrachtet. Da in funktionalen Regressionsmodellen primär die Form der zeitvariierenden Messungen modelliert wird ist es stattdessen sinnvoll die Bodenbewegungen jeweils erst ab dem Zeitpunkt zu betrachten, ab dem diese durch das sich in der jeweiligen Simulation ausbreitende Beben beeinflusst wurden. Auf diese Weise wird sichergestellt, dass die beobachteten Seismogramme grundsätzlich eine vergleichbare Form aufweisen. Als Indikator, ab welchem Zeitpunkt die Bodenbewegungen durch das Erdbeben verursacht wurden, wurde der Schwellenwert 0,01 verwendet. Bei Betrachtung eines spezifischen Seismogramms werden dadurch alle anfänglich gemessenen Bodenbewegungen, welche vor der ersten Messung $\geq 0,01$ auftreten, als nicht durch das Beben verursachtes „Hintergrundrauschen“ betrachtet und entsprechend aus den Messungen entfernt. Grundlage hierfür ist, dass diese Beobachtungen, welche Messungen noch vor Eintreffen der Druckwellen des jeweiligen Erdbebens darstellen, für die bearbeiteten Fragestellungen nicht informativ sind. Die Seismogramme werden - auf Basis lediglich dieser ersten Maßnahme - insgesamt letztlich ab der ersten Messung $\geq 0,01$ bis Ende des 30-sekündigen Messintervalls betrachtet und stellen somit Messungen der „zeitlichen Bodenbewegung ab Eintreffen relevanter Druckwellen der Erdbeben“ dar. Da für die Modellierung mittels der `pfrr`-Funktion funktionale Beobachtungen benötigt werden, welchen ein gleich großes Zeitintervall zugrunde liegt, wurden alle zeitlichen Messungen am Ende mit fehlenden Werten aufgefüllt, falls führende Nullen abgeschnitten wurden.

Letztendlich machte die Betrachtung der Bodenbewegungen erst ab dem Zeitpunkt der ersten Messung $\geq 0,01$ eine weitere Einschränkung der Daten notwendig. In allen geschätzten Modellen, welche auf den derart aufbereiteten Daten basierten, war zu beobachten, dass die glatt geschätzten zeitvariierenden Effekte, sowie entsprechend ebenfalls die dadurch prognostizierten Bodenbewegungen, bei Betrachtung späterer Zeitpunkte nahe dem Ende der 30-sekündigen Messintervalle unrealistische, in die Extreme gehende Wer-

te Annahmen. Möglicher Auslöser hierfür ist die durch das Ignorieren der führenden Messungen $< 0,01$ auftretende Datenstruktur, dass Messungen weiter vom Erdbebenzentrum entfernter Seismographen lediglich gemessen über eine kürzere Zeitspanne vorliegen als Seismogramme mit einer geringen hypozentralen Distanz. Diese von einer Kovariablen - der hypozentralen Distanz - abhängige regelmäßige Datenstruktur bzgl. der fehlenden Werte stellt vermutlich die Ursache für die Schätzprobleme dar. Auszuschließen ist, dass die beschriebenen extremen Effekte aufgrund der geringeren Zahl an echten Messungen bei fortgeschrittener Zeit beziehungsweise einer anderweitig ausgelösten höheren Unsicherheit in der Schätzung rühren, da die entsprechenden Konfidenzbänder gegen Ende des betrachteten Zeitintervalls nicht strukturell größer wurden.

Da obige Schwierigkeiten bei der Schätzung nicht gelöst werden konnten wurden als Alternative statt des 30-sekündigen Messintervalls nur jeweils die ersten 15 Sekunden der Seismogramme ab der ersten Messung $\geq 0,01$ betrachtet und es wurden alle Seismogramme aus den Daten genommen, welche keine vollständige 15-Sekunden-Messung umfassten. Durch diese Maßnahme wurde das Problem umgangen und es konnten sinnvoll geschätzte Effekte sichergestellt werden. Zur Veranschaulichung der Einschränkung sind in Abbildung 15 alle Seismographen dargestellt, welche in mindestens einer der Simulationen eine vollständige 15-sekündige Messung aufweisen. Die in den verbleibenden Daten enthaltenen Seismographen weisen letztendlich eine maximale Entfernung zum Epizentrum von etwa 100km auf.

Weiterhin musste mit der hohen paarweisen Korrelation der drei Quellparameter umgegangen werden, wie sie in Kapitel 2.2 diskutiert wurde. Da eine Zusammenfassung der drei Parameter beispielsweise mittels Mittelwertbildung oder Hauptkomponentenanalyse aus geophysikalischer Sicht einer sinnvollen Interpretierbarkeit der Ergebnisse entgegensteht, wurde letztendlich lediglich die über die Bruchfläche aggregierte Momentenmagnitude $magnitude_{sum}$ als Einflussvariable in den Modellen berücksichtigt. Um möglichen negativen Auswirkungen des extremen Wertebereichs von $magnitude_{sum}$ ($1,17 \cdot 10^{22}$ bis $2,90 \cdot 10^{23}$) auf die numerische Schätzung vorzubeugen wird die Variable dabei skaliert mit dem Wert $\frac{1}{10^{22}}$ in die Modellierung aufgenommen.

Ein bei der praktischen Modellierung immer zu beachtender Punkt besteht in etwaigen, durch die verwendete Implementierung der statistischen Methoden aufgetragenen Einschränkungen. Hier äußerte sich dieser Umstand derart, dass bei der Modellierung auf die Größe der Daten geachtet werden musste. Ursache ist, dass die verwendete `bam`-Funktion für die Berechnung einiger Matrixprodukte auf FORTRAN-Algorithmen zurückgreift, welche nicht auf den Umgang mit sehr hochdimensionalen Vektoren bzw. Matrizen angepasst sind. Im Speziellen betrifft dies den nicht implementierten Umgang mit Vektoren bzw. Matrizen mit $\geq 2^{31}$ Einträgen, welche in der verwendeten R-Version 3.3.1 auf 64-Bit

Systemen intern als `long vector`-Objekte dargestellt werden. Insbesondere wirkte sich dieser Umstand auf die dadurch notwendig gewordene Entscheidung aus, dass die Bodenbewegungen lediglich mit einer Auflösung von 2Hz betrachtet wurden.

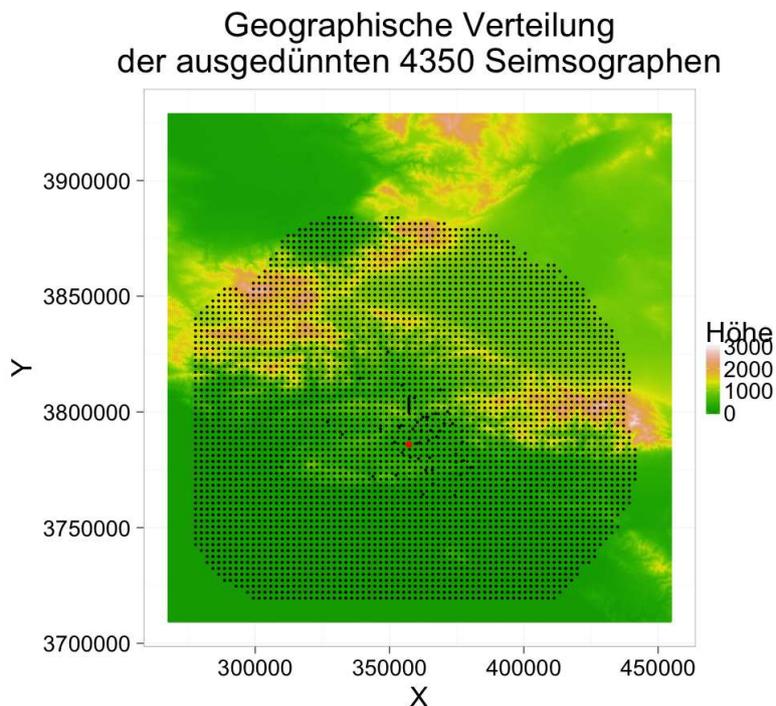


Abbildung 15: Darstellung aller Seismographen, welche in mindestens einer Simulation eine vollständige 15-sekündige Messung aufweisen; in rot eingezeichnet das Epizentrum

5.2 Modellwahl

Grundsätzlich dient die Modellwahl der möglichst genauen Bestimmung der einzelnen Effekte der Kovariablen, indem in erster Linie die Knotenzahlen der pro Kovariable geschätzten glatten Effekte als Hyperparameter optimiert werden. Darüber hinaus wird auch evaluiert, ob ein über den Wertebereich der jeweiligen Kovariable glatter Effekt überhaupt notwendig ist sowie ob der jeweilige Effekt zeitkonstant oder -variierend aufgenommen werden sollte. Am Ende dieses Unterkapitels wird die Modellwahl für verschiedene Verteilungsfamilien angewandt und die dadurch erhaltenen Modelle werden miteinander verglichen.

Bei der Modellierung fiel auf, dass die zu penalisierten (Tensorprodukt-)Splines geschätzten Koeffizienten in den funktionalen Regressionsmodellen letztendlich meist nur wenig penalisiert und die glatten Effekte teilweise unrealistisch rau geschätzt wurden. Dies machte eine Optimierung der Anzahl an Knotenpunkten pro Spline überhaupt notwendig. Die grundsätzliche Problematik ist auf die Größe der verwendeten Daten zurückzuführen und lässt sich mittels der bayesianischen Darstellung glatter Effekte erklären. Durch die

Penalisierung wird eine (Normalverteilungs-)Priori auf die Regressionskoeffizienten gelegt, deren Varianz proportional zum inversen Glättungsparameter λ ist und welche somit die Stärke der Penalisierung beeinflusst (Wood 2006). Allgemein ergeben sich die Regressionsparameter in der bayesianischen Darstellung als gewichtetes Mittel des Erwartungswertes der Priori-Verteilung sowie des aus den Daten berechneten Mittels. Je größer die verwendete Stichprobe dabei ist, desto geringer ist der Einfluss der verwendeten Priori auf die letztendlichen Koeffizienten (Gelman et al. 2014). Um dadurch möglicherweise entstehendem Overfitting vorzubeugen wird die Anzahl der pro glattem Effekt verwendeten Knotenpunkte optimal bzgl. des mittleren quadratischen Fehlers auf einem Testdatensatz bestimmt.

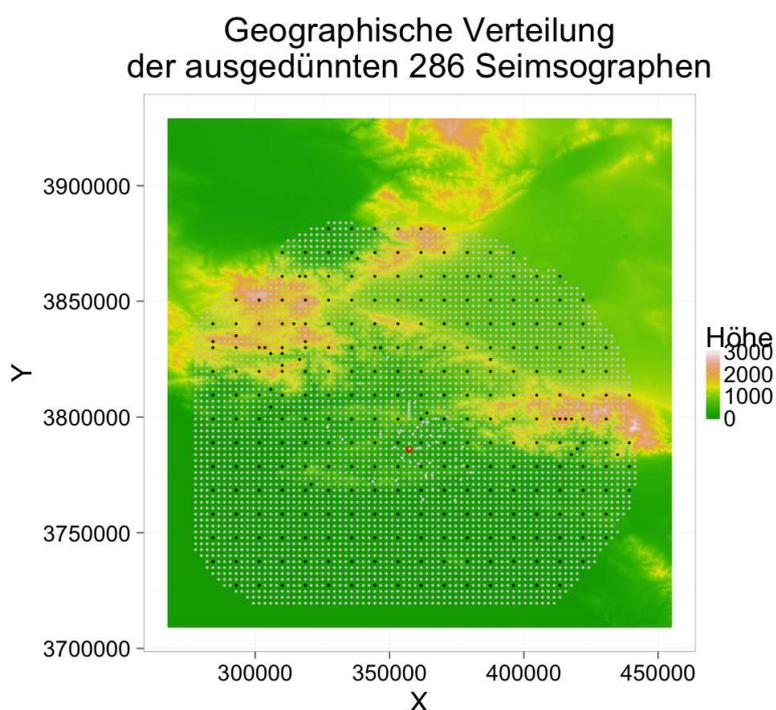


Abbildung 16: Aufteilung der Seismographen in Trainingsdaten (schwarze Punkte) und Testdaten (graue Punkte); Grundlage sind die bzgl. vollständiger 15-sekündiger Messungen ausgedünnten Daten; in rot eingezeichnet das Epizentrum, an dessen Stelle sich ein Seismograph aus den Trainingsdaten befindet

Bei der Aufteilung der Daten in Trainings- und Testdaten wurden die Trainingsdaten deutlich kleiner gewählt als die Testdaten. Dieser Punkt war insbesondere wichtig, um die Modellwahl mit vertretbarem zeitlichen Aufwand durchführen zu können. In erster Linie wurden die vollen Daten dabei räumlich ausgedünnt, wobei das regelmäßige Gitter Grundlage war, auf welchem die meisten Seismographen in den Simulationen aufgestellt wurden. Letztendlich wurden hierbei die Messungen all jener Seismographen in die Trainingsdaten aufgenommen, welche auf jeder vierten x-Koordinate sowie jeder vierten y-

Koordinate des regelmäßigen Gitters lagen, wodurch die Trainingsdaten nur etwa jede 16. Messung umfassen. Von den Seismographen mit Nummern 1-62 wurde lediglich Nr. 62 in die Trainingsdaten aufgenommen, da dieser Seismograph direkt am angenommenen Epizentrum der Simulationen lokalisiert ist. Da der größte Teil des betrachteten Gebietes als Ebene klassifiziert wurde, liegt ein besonderes Augenmerk darauf, dass alle Landformen in den Trainingsdaten adäquat repräsentiert sind, um die Schätzung der Landform-Effekte nicht negativ zu beeinflussen. Falls sich in den Trainingsdaten nach der obigen Einteilung weniger als zehn Seismographen mit einer bestimmten Landform-Kategorie befanden wurden deshalb daraufhin eine entsprechende Anzahl an Seismographen mit eben jener Landform-Kategorie wieder aus den Testdaten gezogen und nachträglich in die Trainingsdaten übernommen. Eine graphische Darstellung, welche Seismographen sich aufgrund dieser Einteilung im Trainings- und welche im Testset befinden, findet sich in Abbildung 16. Neben diesem räumlichen Ausdünnen der Daten enthalten die Trainingsdaten darüber hinaus nur die auf einer Zufallsauswahl basierende Hälfte der Simulationen.

Vorgehen

Die Modellwahl entspricht einer mehrstufigen Grid Search. Anfänglich wird davon ausgegangen, dass alle Einflussgrößen potenziell zeitvariierende Effekte aufweisen und sich dadurch nicht nur auf die Lage, sondern auch auf die Form der beobachteten Bodenbewegungen auswirken. In den unten dargestellten Schritten 1 und 2 werden folglich lediglich die Anzahlen an Knotenpunkten pro marginaler Basis der Kovariablen variiert. In Schritt 3 werden schließlich pro Einflussgröße auch zeitkonstante Effekte evaluiert. Das verwendete Gütemaß bei der durchgeführten Modellwahl bildet der mittlere quadratische Fehler (MSE) auf den Testdaten. Insgesamt werden folgende Schritte bei der Modellwahl vorgenommen, wobei pro Schritt eine Grid Search über alle Kombinationen der verschiedenen Parameter durchgeführt wird:

1. Durchlaufen der Knotenzahlen $k \in \{5, 10\}$ für alle Variablen sowie die marginale Zeitbasis;
Ausnahme: *angle* sowie *l* haben festes $k = 5$, da diese zu wenige unterschiedliche Ausprägungen aufweisen
2. Durchlaufen unterschiedlicher k je nach dem optimalen $k_1 = k$ von Schritt 1:
 - $k_1 = 5$: $k \in \{1, 5\}$ ($k = 1$: Über Wertebereich der Kovariablen konstanter Effekt)
 - $k_1 = 10$: $k \in \{10, 15\}$
 - marginale Zeitbasis: weiter $k \in \{5, 10\}$
3. k sind nun fest; Fitten aller Effekte als zeitvariierend und nicht-zeitvariierend

Modellvergleich

In der Anwendung der bis hierhin beschriebenen Modellwahl wird aus verschiedenen Modelltypen das optimale Modell ausgewählt. Evaluiert werden dabei ein Modell mit Normalverteilungsannahme (NV), ein NV-Modell mit log-transformierter Zielgröße sowie ein Modell mit einer Gammaverteilung für den Response. Die Modellwahl wurde für jeden der drei Modelltypen separat durchgeführt. Ein Vergleich der verschiedenen Modelle bezüglich dem MSE auf den Testdaten sowie dem Anteil der erklärten Devianz findet sich in Tabelle 6.

Da die Bodenbewegungen die Struktur aufweisen, dass die Daten bei einem höheren Niveau der Messungen stärker streuen als bei geringen Bodenbewegungen wurde das NV-Modell mit log-transformierter Zielgröße bewusst einem NV-Modell mit log-Link vorgezogen. Grundlage hierfür ist, dass die Fehler bei Betrachtung der Originalskala unter Verwendung einer log-Transformation multiplikativ - und nicht wie bei einem reinen log-Link additiv - einfließen und somit diese Eigenschaft der Daten besser wieder spiegeln sollten. Durch den quadratischen Zusammenhang von Erwartungswert und Varianz in der Gammaverteilung wird dieser Zusammenhang auch im letzten Modell implizit berücksichtigt. Auf eine theoretisch grundsätzlich sinnvolle Einbeziehung einer t -Verteilung für den Response in den Modellvergleich wurde verzichtet, da der effiziente Schätzalgorithmus in der `bam`-Funktion nicht für diese Verteilung implementiert ist und sich auch einzelne Modelle bei Benutzung der Funktion `gam` auf den für diese Arbeit verfügbaren Rechnern nicht in vertretbarer Zeit schätzen ließen.

	NV	logNV*	Gamma
MSE auf Testdaten	0,0866	0,0832	0,0757
Anteil erklärter Devianz	45,8%	57,7%	70,7%

*: NV mit log-Transformation

Tabelle 6: Vergleich der durch die Modellwahl jeweils erhaltenen Modelle bzgl. ihres MSE auf den Testdaten sowie dem Anteil der erklärten Devianz

Beim Vergleich der Modelltypen weist das Gamma-Modell einen deutlich kleineren MSE auf den Testdaten als die beiden anderen Modelle auf und ebenso ist der Anteil der erklärten Devianz deutlich höher. Dass eine Gammaverteilung zur Modellierung der Bodenbewegungen einer Normalverteilung vorgezogen werden kann bestätigt sich auch bei Betrachtung der QQ-Plots in Abbildung 17. Wie in den Normalverteilungsmodellen zeigt sich hierbei zwar eine negativ zu bewertende Struktur bei den größten Devianz-Residuen (welche aus den extremen Bodenbewegungs-Messungen resultieren), diese ist jedoch

deutlich weniger stark ausgeprägt. Als für die Hauptfragestellung dieser Arbeit finales Modell wird das Gamma-Modell im folgenden Abschnitt genauer interpretiert und evaluiert.

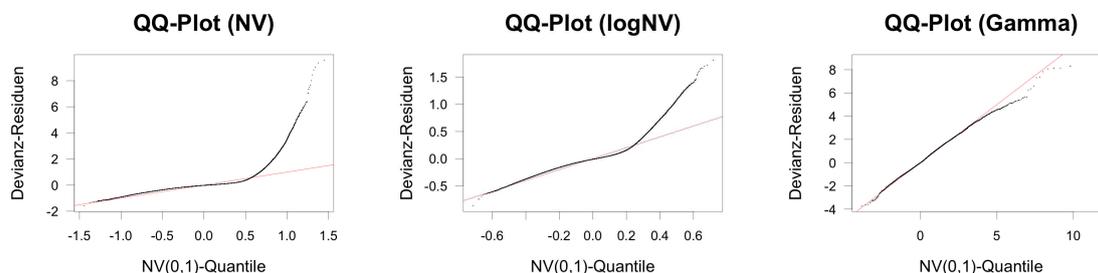


Abbildung 17: *QQ-Plots der Modelle; logNV beschreibt ein NV-Modell mit log-transformierter Zielgröße*

5.3 Erklärung der Bodenbewegungen

In diesem Abschnitt werden die geschätzten Effekte des durch die Modellwahl erhaltenen finalen Modells genauer dargestellt und interpretiert. Grundlage ist aufgrund des kleinsten MSE auf den Testdaten das Modell mit einer Gammaverteilung für den Response. Darüber hinaus wird das Modell in Abschnitt 5.3.2 zusätzlich unter Ausschluss des Quellparameters neu berechnet, um auch eine optimale Prognose der Bodenbewegungen in Situationen zu ermöglichen, in welchen der Quellparameter nicht bekannt ist, sowie um die insgesamten Effekte der Simulationsparameter auf die Bodenbewegungen besser einschätzen zu können. Im Anschluss an die Vorstellung der Modellergebnisse findet sich eine Evaluation der Residuen sowie eine Einschätzung, wie gut beide oben beschriebenen Modelle letztendlich zur Untersuchung sowie zur Prognose der durch die Erdbeben verursachten Bodenbewegungen dienen. Gegen Ende dieses Unterkapitels wird das auf dem Hypozentrum als angenommene Punktquelle basierende finale Modell testweise mit einer alternativen Punktquelle neu berechnet, um eine Einschätzung darüber geben zu können, wie stark die konkret gewählte Punktquelle die Modellergebnisse beeinflusst.

5.3.1 Modell inklusive Quellparameter

Die Ergebnisse der Modellschätzung bezüglich des vollen Modells inklusive des Quellparameters finden sich in Tabelle 7, eine Visualisierung der Effekte in Abbildung 18. Zu betonen ist, dass das finale Gamma-Modell unter Benutzung eines log-Links berechnet wurde. Die dargestellten Effekte wirken sich folglich nicht additiv auf die erwarteten Bodenbewegungen aus, sondern multiplikativ und nachdem sie durch die Exponentialfunktion transformiert wurden. Um einen besseren Eindruck darüber zu bekommen, wie stark

sich alle Einflussgrößen letztendlich auf die Bodenbewegungen auswirken wird auf die ausführlichere graphische Evaluation in Anhang D verwiesen. Da sich keine Interaktionen im Modell befinden gelten die folgenden Interpretationen jeweils bei Konstanthalten der übrigen Einflussgrößen. Auf Signifikanzen wird nicht weiter eingegangen, da insbesondere aufgrund der Größe der Daten ausschließlich p-Werte $< 0,0001$ auftreten.

	Variable	glatter Effekt	edf/Schätzer*	p-Wert
	Intercept (skalar)	✗	-2,32	-
	Intercept	✓	8,90	-
	<i>hypo.dis</i>	✓	85,43	$< 0,0001$
	<i>z</i>	✓	36,74	$< 0,0001$
	<i>magnitude_{sum}</i>	✓	60,25	$< 0,0001$
Simulation	<i>struct_{sediment}</i>	✓	9,90	$< 0,0001$
	<i>l</i>	✓	46,97	$< 0,0001$
	μ_s	✓	40,27	$< 0,0001$
	μ_d	✓	28,62	$< 0,0001$
	<i>angle</i>	✓	22,89	$< 0,0001$
Landform	<i>LF_{Tiefes Tal}</i>	✓	8,31	$< 0,0001$
	<i>LF_{Lokales Tal in Ebene/Hang}</i>	✓	8,06	$< 0,0001$
	<i>LF_{Lokales Tal in Hochebene}</i>	✓	9,50	$< 0,0001$
	<i>LF_{U-förmiges Tal}</i>	✓	8,26	$< 0,0001$
	<i>LF_{Offener Hang}</i>	✓	7,31	$< 0,0001$
	<i>LF_{Flacher Gipfel}</i>	✓	8,29	$< 0,0001$
	<i>LF_{Lokaler Gipfel in Tal}</i>	✓	9,19	$< 0,0001$
	<i>LF_{Lokaler Gipfel in Ebene/Hang}</i>	✓	9,60	$< 0,0001$
	<i>LF_{Steiler Gipfel}</i>	✓	9,41	$< 0,0001$

*: edf bei glatten Effekten, Schätzer bei skalaren Effekten

Tabelle 7: Ergebnisse des finalen Modells mit Gamma-Responseverteilung; Als Referenzkategorien werden LF_{Ebene} sowie $struct_{rock}$ verwendet; Für die zeitvariierenden Effekte wurde eine marginale Zeitbasis mit $k = 10$ Knotenpunkten verwendet; Für den Intercept wurden $k = 20$ Knoten verwendet; edf beschreibt die geschätzten Freiheitsgrade

Da der zeitvariierend geschätzte Intercept um Null zentriert ist enthält das Modell ebenfalls einen skalaren Intercept-Koeffizienten. Aufgrund der größeren Zahl an Einflussgrößen, zu welchen insbesondere jeweils ein glatter Effekt geschätzt wurde, ist die In-

terpretation der beiden Intercept-Schätzer jedoch nicht trivial. Stattdessen wird für eine umfassendere Interpretation der gesamten Effekte erneut auf die in Anhang D gezeigten Effekt-Grafiken verwiesen. Bezüglich der Effekte der Simulationsparameter soll nochmals angemerkt werden, dass diese aufgrund der Kausalitätsstruktur (siehe Kapitel 5.4) potenziell teilweise durch den Effekt des Quellparameters überdeckt werden. Falls die Effekte von Simulationsparametern inhaltlich nicht plausibel erscheinen wird deshalb insbesondere auf das in Abschnitt 5.3.2 berechnete Modell verwiesen, um einen umfassenderen Eindruck der Effekte der Simulationsparameter zu bekommen.

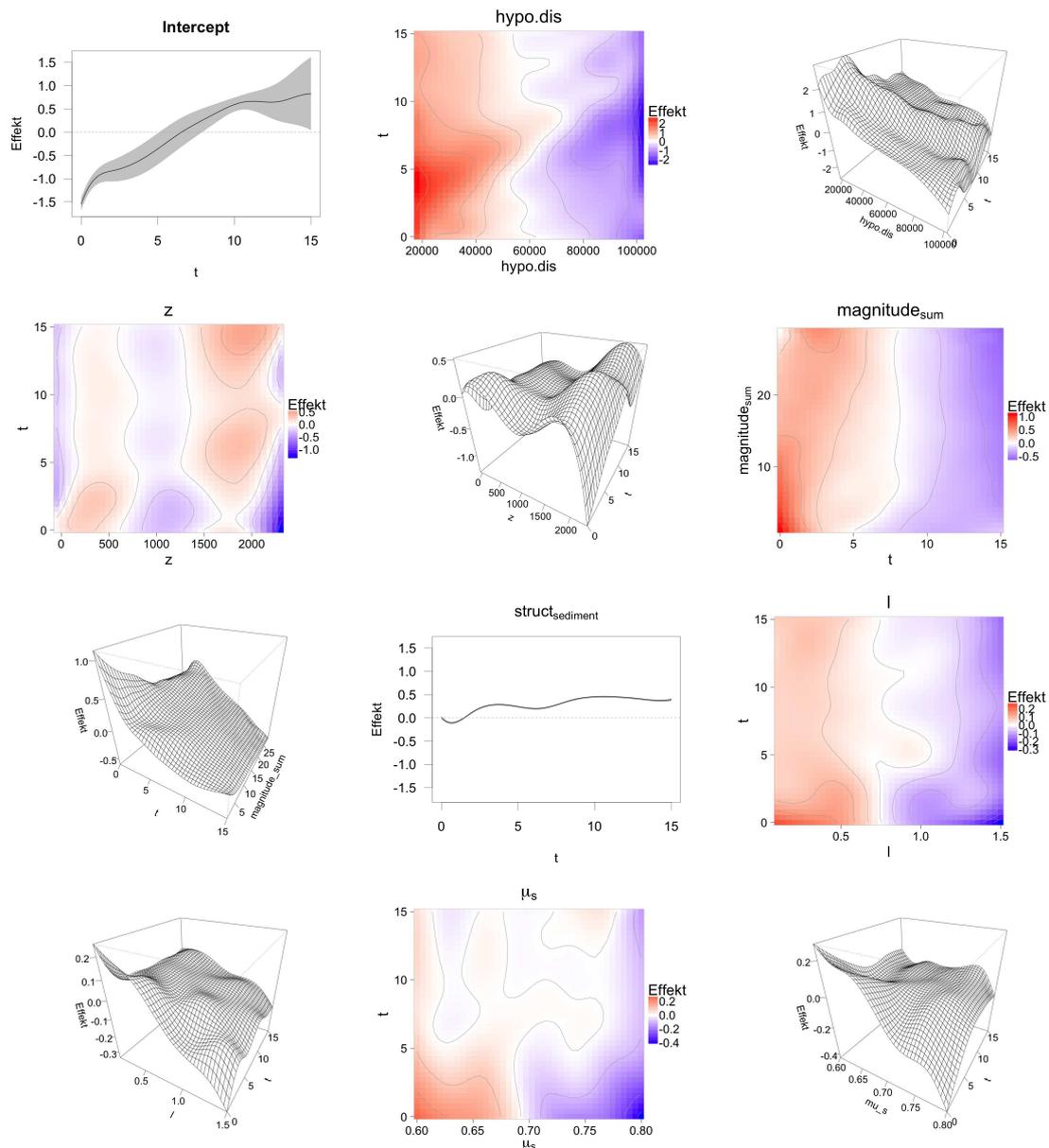


Abbildung 18: Visualisierung der glatten Effekte im finalen Modell (Fortsetzung siehe S.47); grau eingezeichnet sind bayesianische Konfidenzbänder (Wood 2006)

Der durch das Modell geschätzte Effekt der hypozentralen Distanz weist einen deutlichen Einfluss der Variable nach und bestätigt die entsprechenden Beobachtungen aus der Deskription. Die am nächsten zum Hypozentrum aufgenommenen Messungen haben die klare Form, dass vergleichsweise kurz nach Eintreten der ersten Messung $\geq 0,01$ eine stärkere Bodenbewegung eintritt. Anschließend nehmen die Bodenbewegungen wieder recht schnell ab, stagnieren jedoch auf einem noch vergleichsweise hohen Niveau. Je weiter entfernt vom Hypozentrum die Messung durchgeführt wird, desto geringer sind die allgemeinen Bodenbewegungen über die Zeit. Darüber hinaus tritt die maximale Bodenbewegung mit steigendem Abstand immer später ein und fällt immer niedriger aus.

Die Höhe der Messstation - welche lediglich als Kontrollvariable aufgenommen wurde - weist ebenfalls einen zeitlich und über den Wertebereich der Kovariable klar nichtlinearen Effekt auf. Die größten Bodenbewegungen werden hierbei bei Höhen zwischen 1500m und 2200m sowie zwischen 100m und 600m erwartet.

Bezüglich des Effektes der skalaren Momentenmagnitude lässt sich eine gewisse Struktur dahingehend erkennen, dass, je höher die Magnitude als Maß für die Erdbebenstärke ist, desto eher eine Form der Seismogramme vorliegt, dass die Bodenbewegungen ihr Maximum nach wenigen Sekunden erreichen und anschließend wieder abnehmen. Bei sehr kleinen Magnituden sieht man dagegen, dass der höchste Effekt direkt zu Messbeginn vorliegt und anschließend schneller abnimmt. Grundsätzlich erscheint es insgesamt fragwürdig, dass der höchste Effekt für die kleinste Magnitude geschätzt wurde. Da die Magnitude jedoch wie erläutert kausal von den Simulationsparametern beeinflusst wird und hier in erster Linie als *indirekter* Effekt derselbigen zu interpretieren ist ist eine wirkliche Einschätzung über die Plausibilität des Effektes letztendlich schwierig, da dieser *indirekte* Effekt immer auch zusammen mit den *direkten*, in diesem Modell geschätzten Effekten der Simulationsparameter betrachtet werden muss.

Der geschätzte Effekt der Bodenbeschaffenheit gibt den mittleren Unterschied an zwischen den Bodenbewegungen bei Simulationen mit einem Untergrund komplett aus Gestein und Simulationen mit einem Untergrund komplett aus Sediment. In Simulationen mit einem Sediment-Untergrund wurden dabei fast über das komplette Messintervall höhere Bodenbewegungen registriert. Der Effekt ist zeitvariierend, eine bessere Einschätzung wie stark die über die Zeit geschätzten Unterschiede des Effektes sind lässt sich an der zugehörigen Grafik in Anhang D erkennen. Zu Beginn der Messung unterscheiden sich die Bodenbeschaffenheiten kaum, die Unterschiede fallen jedoch etwas größer aus, wenn man das Maximum sowie das zeitliche Ende der gemessenen Bodenbewegungen betrachtet. Aus physikalischer Sicht bestätigt dieser Effekt die Vermutung, dass Erdbeben bei einem Sedimentboden als weicherem, „schwammigerem“ Untergrund größere Bodenbewegungen verursachen als bei einem Untergrund aus Stein.

Bezüglich dem linearen slip-weakening l - welches angibt, über welche Distanz der Reibungswiderstand vom statischen auf den dynamischen Wert abfällt - wurde ebenfalls ein zweidimensionaler glatter Effekt geschätzt. Insgesamt lässt sich die klare Struktur erkennen, dass die Bodenbewegungen bei steigendem l abnehmen. Da der vorherrschende Reibungswiderstand umso höher ist, je größer die Distanz l ist und je langsamer dadurch der Reibungswiderstand auf das niedrigere Niveau abfällt, erscheint dieser Effekt dabei sinnvoll. Auffällig ist darüber hinaus, dass die absolut stärksten Effekte bis ca. 4 Sekunden nach Messbeginn zu beobachten sind. Dies könnte darauf zurückzuführen sein, dass die Distanz, über welche der Reibungswiderstand abfällt, lediglich zu Beginn der Messung eine etwas größere Rolle spielt, bevor der Reibungswiderstand mit dem dynamischen Wert sein dauerhaftes Minimum erreicht.

Auch bezüglich des statischen Reibungskoeffizienten μ_s werden bei höheren Parameterwerten grundsätzlich niedrigere Bodenbewegungen erwartet. Dies erscheint auch hier sinnvoll, da eine größere vorherrschende Reibung einer schnellen Ausbreitung des Bruches und somit des Erdbebens entgegensteht. Insbesondere die Struktur, dass die absolut größten Effekte zu Beginn des Messintervalls geschätzt wurden, macht aus geophysikalischer Sicht Sinn, da μ_s wie beschrieben lediglich den Startwert des Reibungswiderstandes beschreibt, welcher über die Zeit abnimmt.

Neben der hypozentralen Distanz weist der dynamische Reibungskoeffizient μ_d klar den größten Effekt auf die Bodenbewegungen auf. Wie bei den übrigen Reibungsparametern l und μ_s ist die Struktur zu beobachten, dass höhere Werte von μ_d zu geringeren Messungen der Bodenbewegung führen. Wie erwähnt ist der Einfluss von μ_d jedoch deutlich größer als jener der anderen Simulationsparameter. Da der dynamische Reibungswiderstand den über die komplette Simulation mindestens vorherrschenden Reibungswiderstand angibt erscheint auch dieser Effekt plausibel.

Für den Winkel der (konstanten) Hintergrundspannung *angle* wurde ebenfalls ein zweidimensionaler Effekt geschätzt. Höhere Winkel lassen basierend auf dem Modell dabei geringere Bodenbewegungen erwarten. Ohne genauere Betrachtung der Beschaffenheit der Bruchfläche ist eine Einschätzung der Plausibilität dieses Effektes insgesamt schwierig. Grundsätzlich erscheint es jedoch möglich, dass sich der Winkel im betrachteten Wertebereich (welcher nur 10° aller 360° überspannt!) relativ monoton auswirkt.

Die geschätzten Effekte der Landformen sind jeweils im Vergleich zur Referenzkategorie Ebene zu interpretieren. Grundsätzlich ist auffällig, dass alle Landformen fast über das komplette zeitliche Messintervall mindestens gleiche, jedoch oft auch höhere erwartete Bodenbewegungen aufweisen als die Kategorie Ebene. Basierend auf den Effekt-Grafiken in Anhang D lassen sich jedoch darüber hinaus keine strukturellen Unterschiede beispielsweise beim Vergleich großräumiger Täler, Ebenen und Gipfel erkennen.

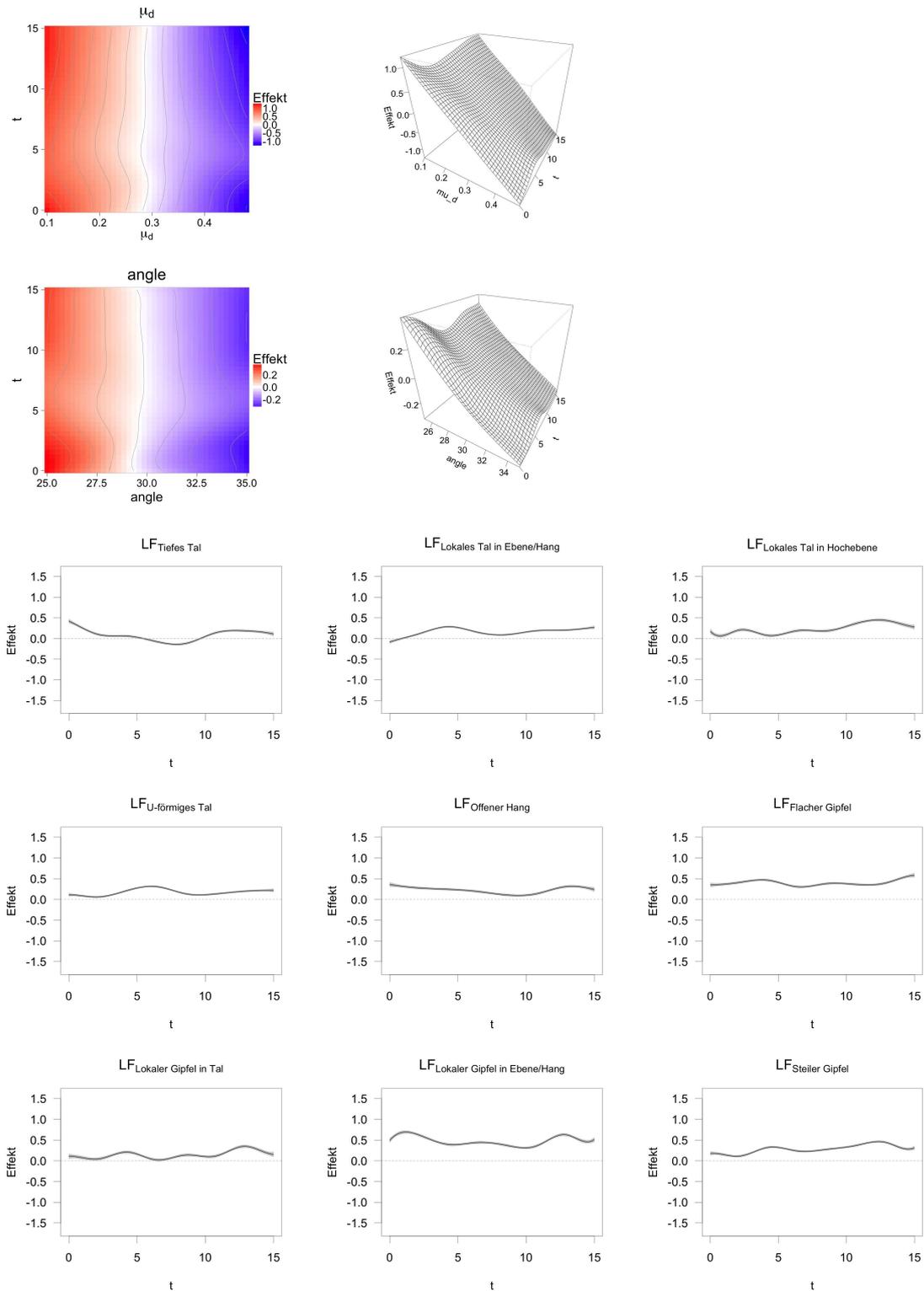


Abbildung 18: Visualisierung der glatten Effekte im finalen Modell (Fortsetzung von S.44); grau eingezeichnet sind bayesianische Konfidenzbänder (Wood 2006)

5.3.2 Modell ohne Quellparameter

Für die Berechnung des Modells ohne die skalare Momentenmagnitude als Einflussvariable wurde die Modellwahl teilweise nochmal neu berechnet. Für alle Variablen außer den Simulationsparametern wurde davon ausgegangen, dass die im vollen Modell optimalen Knotenzahlen die beste Repräsentierung der wahren Effekten der Variablen gewährleisten. Die Modellwahl wurde folglich lediglich bezüglich der Effekte der Simulationsparameter neu durchlaufen. Da der Effekt des Quellparameters - welcher einen indirekten Effekt der Simulationsparameter darstellt, da diese die Magnitude kausal beeinflussen - aus dem Modell genommen wurde ist es prinzipiell möglich, dass die Simulationsparameter potenziell auf etwas andere Weise wirken als im vollen Modell.

Aufgrund der Beobachtungen aus dem Modell aus Kapitel 5.4, welches untersucht, auf welche Weise die Simulationsparameter die Magnitude beeinflussen, wurde eine Interaktion zwischen *angle* und *struct* mit in das aktuelle Modell aufgenommen, da sich diese als relevante Einflussgröße zur Prädiktion der Magnitude herausstellte. Im Anschluss an die Modellwahl wurde die Interaktion testweise wieder aus dem Modell genommen. Da sich zeigte, dass das Modell inklusive der Interaktion einen größeren MSE auf den Testdaten aufwies wurde die Interaktion schließlich nicht im Modell behalten.

Für die Interpretation der Effekte wird im Folgenden lediglich auf die Simulationsparameter eingegangen, da diese von Haupt-Interesse sind und da die Effekte der übrigen Variablen ähnlich zu den Effekten im Modell inklusive des Quellparameters geschätzt wurden. Die geschätzten Effekte der Simulationsparameter finden sich in Tabelle 8, die glatten Effekte sind in Abbildung 19 dargestellt. Eine Übersicht über alle geschätzten Effekte dieses Modells sowie weitere graphische Evaluierungen der Kovariablen-Effekte findet sich in Anhang D.

	Variable	glatter Effekt	edf/Schätzer*	p-Wert
Simulation	<i>struct</i> _{sediment}	✓	9,90	< 0,0001
	<i>l</i>	✓	7,56	< 0,0001
	μ_s	✓	69,88	< 0,0001
	μ_d	✗	-5,48	< 0,0001
	<i>angle</i>	✓	6,64	< 0,0001

*: edf bei glatten Effekten, Schätzer bei skalaren Effekten

Tabelle 8: Geschätzte Effekte der Simulationsparameter aus dem finalen Modell unter Ausschluss des Quellparameters; edf beschreibt die geschätzten Freiheitsgrade

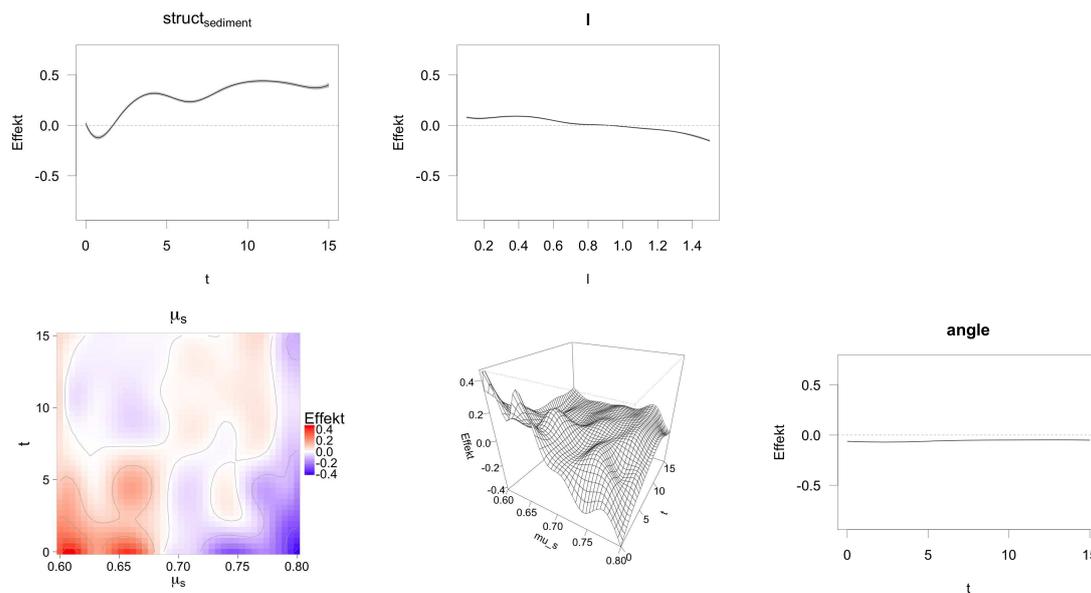


Abbildung 19: Glatte Effekte der Simulationsparameter aus dem finalen Modell ohne Quellparameter; grau eingezeichnet sind bayesianische Konfidenzbänder (Wood 2006)

Bezüglich der Untergrundbeschaffenheit wurde ein sehr ähnlicher Effekt wie im vollen Modell geschätzt. Simulationen mit Sedimentböden weisen also durchweg etwas höhere Bodenbewegungen auf als Simulationen mit einem Untergrund komplett aus Gestein.

Das lineare slip-weakening l weist einen zeitkonstanten, jedoch über den Wertebereich des Parameters variierenden glatten Effekt auf. Aufgrund der Zeitkonstanz dieses Effektes werden die Messungen bei verändertem l dabei über das komplette Messintervall mit dem gleichen Faktor skaliert. Die grundsätzliche Struktur, dass höhere Werte der Distanz, über welchen der Reibungswiderstand vom statischen auf den dynamischen Wert abfällt, zu geringeren Bodenbewegungen führen, wurde auch im vollen Modell gefunden. Unter Ausschluss des Quellparameters wirkt dieser Effekt jedoch nun zeitkonstant. Auffällig am geschätzten Effekt ist, dass für Werte des Parameters bis ca. 0,5 relativ gleiche Bodenbewegungen erwartet werden und eine größere Distanz in diesem Wertebereich nicht zu niedrigeren erwarteten Bodenbewegungen führt.

Der Effekt des statischen Reibungswiderstandes μ_s wurde vergleichbar zum Effekt im vollen Modell geschätzt und erscheint somit weiterhin plausibel. Höhere Werte von μ_s führen insgesamt zu abnehmenden Bodenbewegungen, jedoch wirkt sich dieser Effekt meist nur zu Beginn des Messintervalls etwas stärker aus.

Analog zum Modell inklusive des Quellparameters weist der dynamische Reibungswiderstand μ_d auch hier den stärksten Effekt aller Simulationsparameter auf. Unter Ausschluss des Quellparameters zeigt sich nun jedoch ein skalarer, linearer Effekt statt eines zwei-dimensionalen glatten Effektes. Steigt μ_d um 0,1, so nehmen die erwarteten Bodenbewe-

gungen dabei über das komplette Messintervall um 43% ab.

Für *angle*, den Winkel der Hintergrundspannung, wurde ein eindimensionaler glatter Effekt geschätzt, welcher über die Zeit variiert. Bei höheren Winkeln werden durch das Modell geringere Bodenbewegungen über das komplette Messintervall erwartet. Ca. bis 4 Sekunden nach Messbeginn wirkt sich der Parameter dabei minimal stärker aus als danach, insgesamt verläuft der Effekt über die Zeit jedoch sehr gleichmäßig. Wie bereits erwähnt ist eine Einschätzung der Plausibilität dieses Effektes schwierig ohne Einbezug genauerer Informationen über die Bruchfläche.

	Volles Modell	Modell ohne Quellparameter
MSE auf Testdaten	0,0757	0,0755
Anteil erklärter Devianz	70,7%	70,4%

Tabelle 9: Vergleich der beiden durch die Modellwahl erhaltenen finalen Modelle mit und ohne Quellparameter bzgl. des MSE auf den Testdaten sowie dem Anteil der erklärten Devianz

In Tabelle 9 wird das Modell unter Ausschluss des Quellparameters abschließend mit dem vollen Modell verglichen. Vergleicht man lediglich die Anteile der erklärten Devianzen zwischen den Modellen, so fällt grundsätzlich auf, dass sich die Werte nur minimal unterscheiden. Zusätzlich zu den bei beiden Modellen enthaltenen Simulationsparametern trägt der Quellparameter folglich nur sehr wenig weitere Information für die Modellierung der Bodenbewegungen bei. Dies bestätigt insbesondere die Ergebnisse von Kapitel 5.4. Basierend auf dem auf den Testdaten berechneten MSE schneidet das volle Modell sogar schlechter ab als das Modell ohne den Quellparameter. Da der Testdaten-MSE hier als zentrales Gütemaß verwendet wird, wird das Modell unter Ausschluss des Quellparameters im Folgenden als finales Modell für die Modellierung der Bodenbewegungen weiter evaluiert.

5.3.3 Modellevaluation

Die wichtigsten Residualplots, welche zur Evaluierung des finalen Modells (unter Ausschluss des Quellparameters, siehe vorigen Abschnitt) herangezogen werden, finden sich in den Abbildungen 20 und 21 und werden im Folgenden diskutiert. Im Sinne einer möglichst umfassenden Evaluation sind jeweils die zusammengefassten Daten, bestehend aus sowohl den Trainings- als auch den Testdaten, Grundlage für die Grafiken. Entsprechende Grafiken nur basierend auf den Trainingsdaten weisen eine vergleichbare Form auf. Residualplots für das Modell inklusive des Quellparameters weisen ebenfalls eine vergleichbare Form auf und finden sich in Anhang E.

Im Anschluss an die Evaluation der Residuen werden die Prognosen des finalen Modells mit den wahren Messungen einiger Simulationen und Seismographen verglichen, um darauf basierend einen Eindruck der allgemeinen Güte der Ergebnisse zu bekommen. Der Vollständigkeit halber werden hierbei auch die Prognosen des vollen Modells inklusive des Quellparameters mit verglichen.

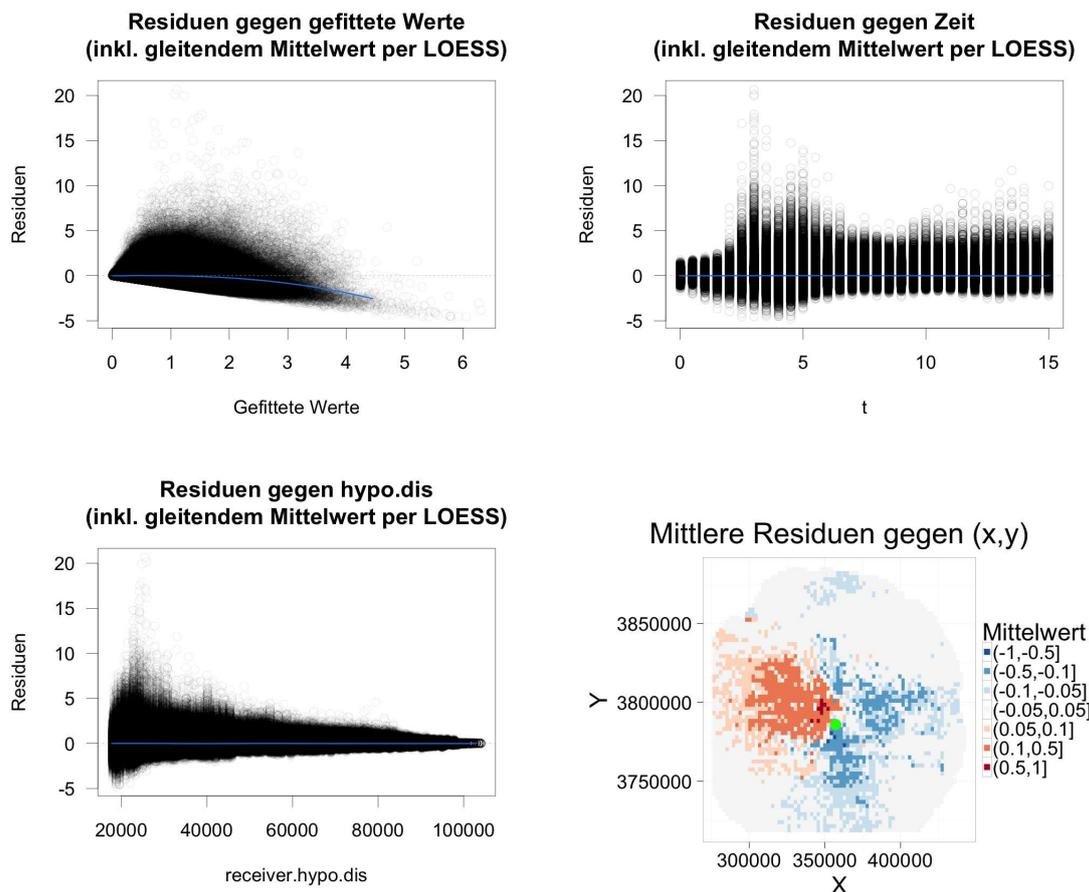


Abbildung 20: Residualplots des finalen Modells ohne den Quellparameter basierend auf den zusammengefassten Trainings- und Testdaten; LOESS-Schätzungen basieren jeweils auf 50.000 zufällig ausgewählten Residuen; in grün eingezeichnet ist das Epizentrum der Beben (Grafik unten rechts)

Basierend auf allen vier in Abbildung 20 dargestellten Grafiken kann gefolgert werden, dass das Modell eine noch annehmbare Residualstruktur aufweist. Insbesondere über die Wertebereiche der Zeit sowie der hypozentralen Distanz weichen die Modellprognosen nicht strukturell von den wahren Beobachtungen ab. Etwas unterschiedlich verhält sich dies bei Gegenüberstellung der Residuen gegen die gefitteten Werte. Je höher die vorhergesagten Werte dabei sind, desto größer fällt die strukturelle Überschätzung der Beobachtungen durch das Modell aus. Insgesamt lässt sich diese Überschätzung jedoch noch als

annehmbar einordnen.

Mittlere Residuen gegen (x,y)

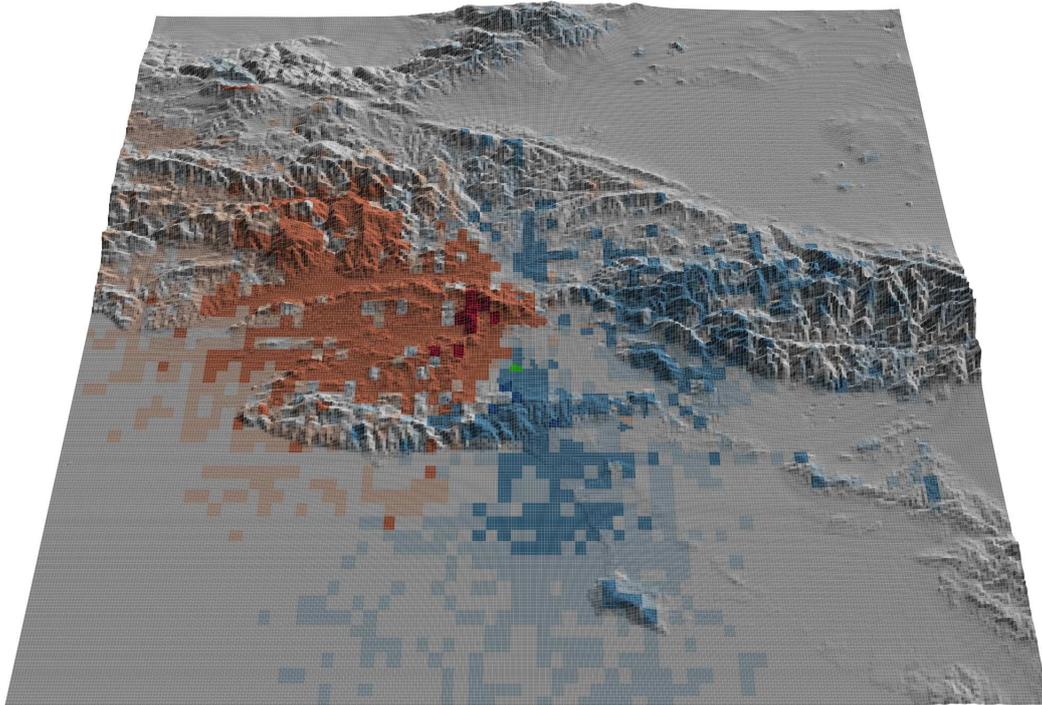


Abbildung 21: Residuen gegen (x,y) ; unterlegt ist die dreidimensionale Topographie des betrachteten Gebietes; die verwendete Farblegende ist identisch zu Abbildung 20; in grün eingezeichnet ist das Epizentrum der Beben

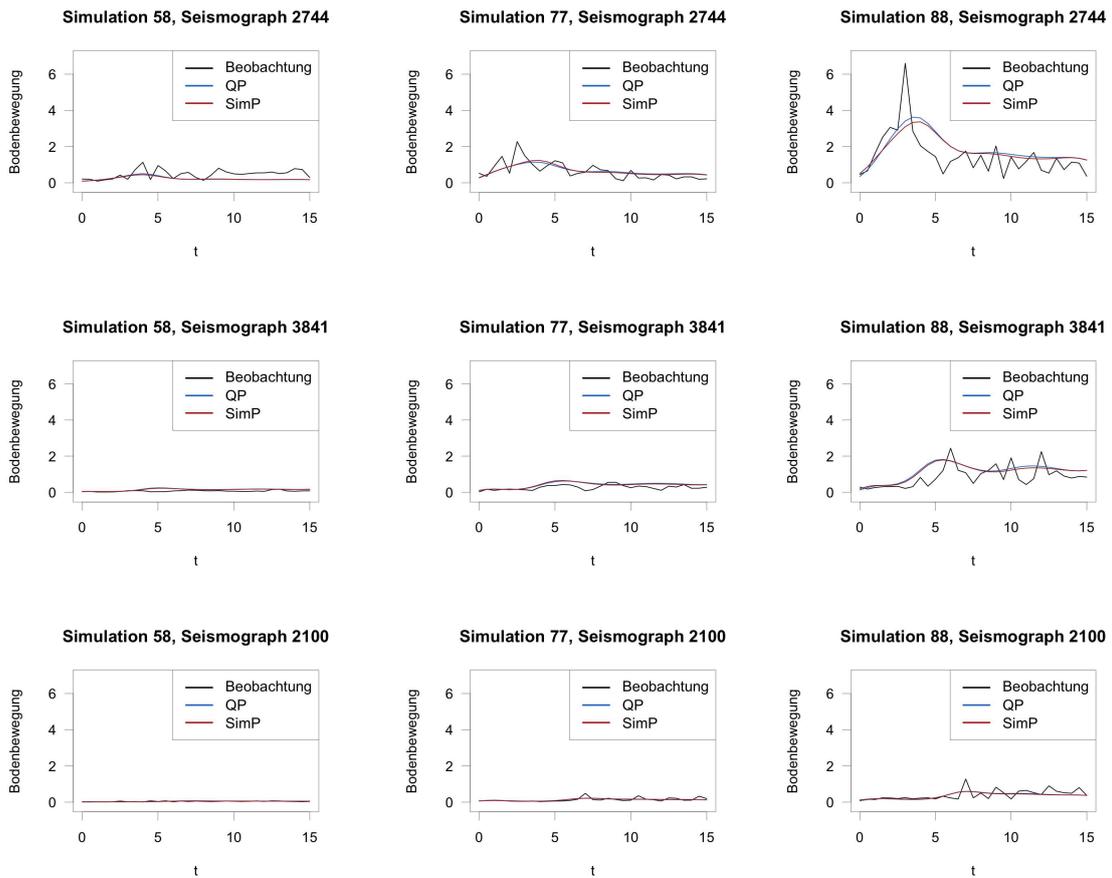
Bei Betrachtung der über den Raum geplotteten mittleren Residuen lassen sich deutliche Regionen erkennen, in welchen das Modell die wahren Bodenbewegungen strukturell über- beziehungsweise unterschätzt. Auffällig ist hierbei, dass das strukturell unterschätzte Gebiet komplett westlich des Hypozentrums liegt, das strukturell überschätzte Gebiet dagegen nicht westlich davon. Bezieht man die Topographie in die Betrachtung mit ein wie in Abbildung 21, so lassen sich darüber hinaus jedoch keine weiteren direkten Rückschlüsse über mögliche topographische Ursachen für die Über- und Unterschätzungen ziehen. Grundsätzlich könnte die Residualstruktur darauf hindeuten, dass weitere topographische Effekte bestehen, welche im Modell nicht auf adäquate Weise berücksichtigt wurden oder dass eventuell ein Richtungseffekt in der Ausbreitung der Erdbeben besteht.

Letztere Vermutung kann aufgrund der Ergebnisse in Abschnitt 5.5.2 auf Basis der vorgenommenen Analysen jedoch zumindest nicht bestätigt werden. Stattdessen wird vermutet, dass die räumliche Struktur der Residuen daher rührt, dass das als Punktquelle gewählte Hypozentrum gemittelt über alle Simulationen nicht der Region auf der Bruchfläche entspricht, von welcher die größten Bodenbewegungen ausgehen. Basierend auf dem teilweise neu geschätzten Modell unter Benutzung einer alternativen Punktquelle in Abschnitt 5.3.4 bestätigt sich diese Vermutung.

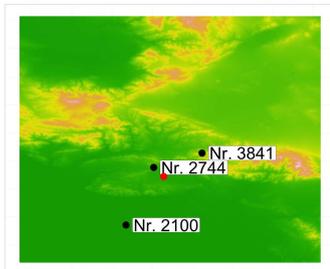
Evaluierung der Prognosegüte

Um eine generelle Einschätzung über die Prognosegüte des finalen Modells sowie des Modells inklusive des Quellparameters geben zu können werden die Prognosen in Abbildung 22 den wahren Beobachtungen gegenübergestellt. Grundlage für den Vergleich bilden jeweils drei ausgewählte Simulationen sowie Seismographen. Die Auswahl erfolgte dabei jeweils basierend auf der gemachten Beobachtung, dass die hypozentrale Distanz sowie der dynamische Reibungswiderstand μ_d die beiden Einflussgrößen mit der absolut größten Auswirkung auf die Bodenbewegungen darstellen. Für die Evaluation der Prognosegüte wurde entsprechend je eine Simulation mit einem niedrigen, mittleren und hohen Wert von μ_d ausgewählt, sowie je ein Seismograph mit einer niedrigen, mittleren und höheren hypozentralen Distanz. Die evaluierten Seismogramme können als repräsentativ für alle Seismogramme gesehen werden, die Prognosen der Modelle verhalten sich ähnlich im Vergleich mit den übrigen, hier nicht dargestellten Messungen.

Beide Modelle prognostizieren für die in der Abbildung betrachteten Seismogramme insgesamt sinnvolle Werte. Insbesondere die Seismogramme, welche sehr weit entfernt von der Punktquelle aufgenommen wurden, werden sehr gut von den Modellen nachvollzogen. Bei den etwas näher an der Punktquelle aufgenommenen Messungen werden die maximalen Ausschläge zu Beginn ebenfalls verhältnismäßig gut vorhergesagt, es fällt jedoch auf, dass, wenn ein mittlerer bis niedriger Reibungswiderstand vorliegt, nach Abklingen der Bodenbewegungen über die komplette Zeit etwas zu hohe Werte prognostiziert werden. Die Prognosen beider Modelle sind insgesamt gesehen nahezu identisch, was aufgrund der Ergebnisse des vorigen Unterkapitels durchaus zu erwarten war.



Lage der betrachteten Seismographen



Seismograph	<i>hypo.dis</i>	<i>z</i>	Landform
Nr. 2744	20799	374	Ebene
Nr. 3841	37031	1304	Steiler Gipfel
Nr. 2100	52347	-50	Ebene

Simulation	<i>magnitude_{sum}</i>	<i>struct</i>	<i>l</i>	μ_s	μ_d	<i>angle</i>
58	$1,17 \cdot 10^{22}$	<i>sediment</i>	0,8	0,61	0,48	33
77	$7,36 \cdot 10^{22}$	<i>rock</i>	1,1	0,63	0,27	28
88	$29,00 \cdot 10^{22}$	<i>rock</i>	0,8	0,75	0,10	27

Abbildung 22: Vergleich der Prognosen der Modelle inklusive des Quellparameters (QP) und ohne den Quellparameter (SimP) mit den wahren Beobachtungen anhand dreier Simulationen und Seismographen; Die Seismographen wurden ausgewählt nach variierendem Abstand zum Hypozentrum, Simulationen nach variierendem μ_d ; in rot markiert in der Karte ist das Epizentrum

5.3.4 Auswirkung der Wahl der Punktquelle

Zum Abschluss der Modellierung der zentralen Fragestellung wird nun noch darauf eingegangen, wie stark sich die Auswahl der Punktquelle auf die Ergebnisse auswirkt. Das bis hierhin vorgestellte finale Modell unter Ausschluss des Quellparameters basierend auf dem Hypozentrum wird dabei verglichen mit dem Modell auf Basis des senkrecht auf die Bruchfläche projizierten Punktes, für welchen die größte mittlere Bodenbewegung gemessen wurde (siehe Abbildung 6). Auf die in das Modell einfließenden Daten wirkt sich diese Änderung nur derart aus, dass die „hypozentralen Distanzen“ (Variable *hypo.dis*) auf Basis der alternativen Punktquelle neu berechnet werden und dass diese inhaltlich nun die Abstände zur alternativen Punktquelle angeben.

Grundlage der Verwendung des Hypozentrums als Punktquelle ist, dass aufgrund der Tatsache, dass das Hypozentrum über alle Simulationen den immer gleichen Startpunkt des Bruches auf der Bruchfläche angibt, vom Hypozentrum immer Bodenbewegungen ausgehen. Auch wenn bei einem schwachen Erdbeben der Bruch früh „stirbt“, sich also nicht weit ausbreitet, stellt das Hypozentrum dadurch eine nachvollziehbare Wahl für die Punktquelle dar. Insbesondere bei größeren Beben ist es jedoch möglich, dass die Region auf der Bruchfläche, welche Auslöser der größten Bodenbewegungen ist, weiter entfernt vom Hypozentrum liegt. Aufgrund dieser Motivation wird als alternative Punktquelle im Folgenden eben jener senkrecht auf die Bruchfläche projizierte Punkt verwendet, an welchem über alle Simulationen gemittelt die größte mittlere Bodenbewegung registriert wurde. Die beiden evaluierten Modelle werden primär bezüglich ihrer Prognosegüte miteinander verglichen, eine Übersicht über die geschätzten Effekte sowie ausgewählte Residualplots auf Basis der alternativen Punktquelle findet sich in Anhang F. Der Vollständigkeit halber werden dort auch die Ergebnisse des auf der alternativen Punktquelle basierenden Modells unter Einbezug des Quellparameters berichtet.

Für einen ersten Vergleich der Prognosegüte der Modelle findet sich in Abbildung 24 auf S.58 analog zum vorigen Abschnitt ein Vergleich der Prognosen mit den wahren Beobachtungen. Für eine Erläuterung, wieso eben jene Simulationen und Seismographen als Grundlage für den Vergleich ausgewählt wurden wird auf den vorigen Abschnitt verwiesen. Insgesamt zeigt sich beim Vergleich der Modelle, dass das Modell mit der alternativen Punktquelle die meisten Seismogramme etwas besser prognostiziert als dies im Modell basierend auf dem Hypozentrum der Fall ist. Lediglich fällt negativ auf, dass ersteres Modell bei nah an der Punktquelle aufgenommenen Messungen sowie einem niedrigen bis mittleren Reibungswiderstand nach Abklingen der maximalen Bodenbewegung die wahren Beobachtungen teilweise deutlich überschätzt. Im auf dem Hypozentrum basierenden Modell fällt diese Überschätzung zwar ebenfalls auf, jedoch fällt sie insgesamt weniger stark aus. Problematisch ist jedoch, dass das auf dem Hypozentrum basierende

Modell nicht nur nahe an der Punktquelle aufgenommene Seismogramme überschätzt, sondern auch Messungen bezüglich weiter entfernter Seismographen.

Dass die Verwendung der alternativen Punktquelle insgesamt zu besseren Prognosen der Bodenbewegungen führt bestätigt sich bei Betrachtung der Residuen. Insbesondere ein Vergleich der über den Raum geplotteten Residuen anhand Abbildung 23 zeigt, dass die räumliche Residualstruktur durch die Wahl einer adäquaten Punktquelle deutlich verringert werden kann.

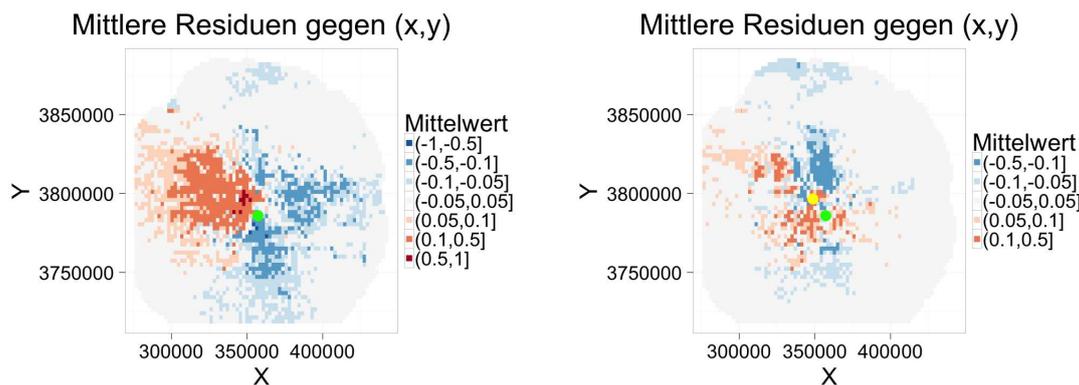


Abbildung 23: Gegenüberstellung der über den Raum geplotteten Residuen des Modells unter Benutzung des Hypozentrums als Punktquelle (linke Grafik) sowie des Modells unter Benutzung des senkrecht auf die Bruchfläche projizierten Punktes der maximalen mittleren Bodenbewegung (rechte Grafik); in grün eingezeichnet das Epizentrum, in gelb (nur rechte Grafik) der Punkt der maximalen mittleren Bodenbewegung

Punktquelle	Mittlere Bodenbewegung der Simulationen			
	(0;0, 1]	(0, 1; 0, 2]	(0, 2; 0, 4]	(0, 4; 0, 61]
Hypozentrum	0,0088	0,0271	0,0856	0,2416
Alternative Punktquelle	0,0093	0,0269	0,076	0,1998

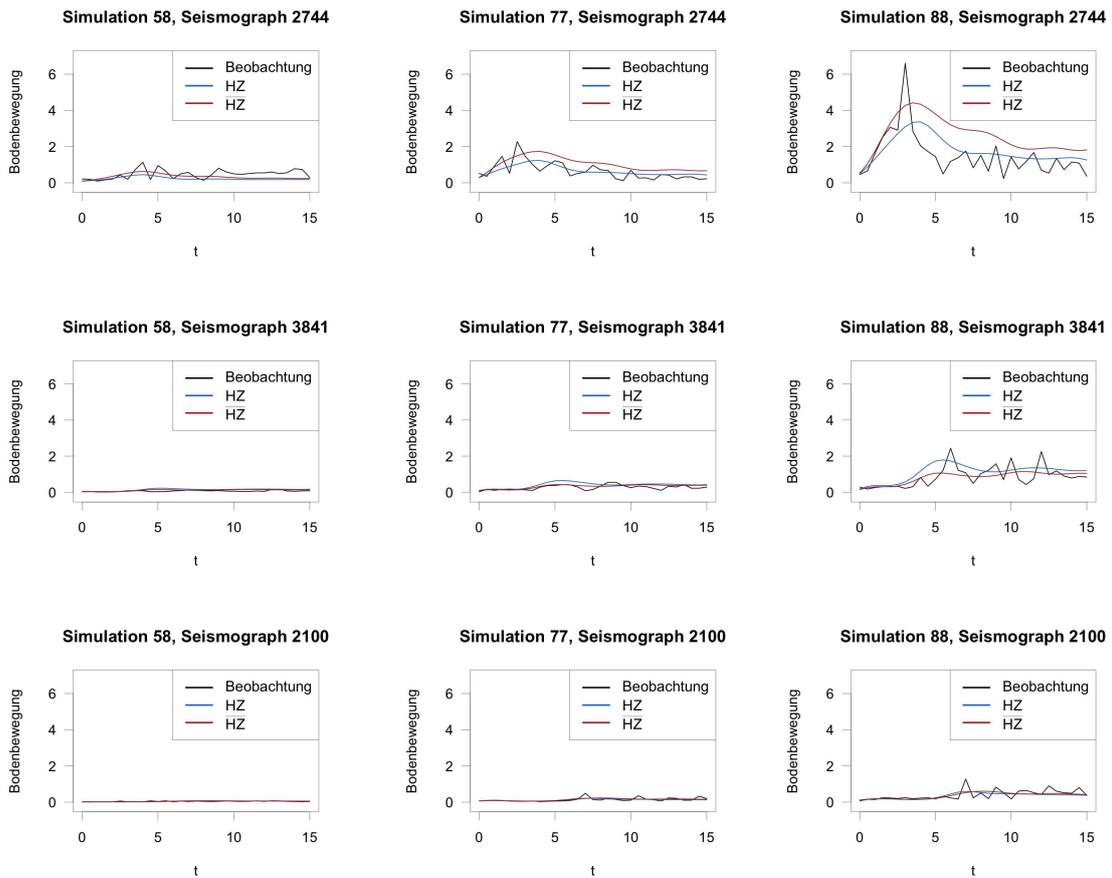
Tabelle 10: Vergleich der auf dem Hypozentrum sowie auf der alternativen Punktquelle basierenden Modellen bezüglich des MSE auf den Testdaten. In die Betrachtung fließen jeweils nur Simulationen ein, deren über die Zeit sowie über alle Seismogramme gemittelte Bodenbewegung im angegebenen Intervall liegt; in grün eingezeichnet ist das Epizentrum, in gelb (nur rechte Grafik) der Punkt der größten mittleren Bodenbewegung

Auch ein Vergleich der Modelle anhand des MSE auf den Testdaten in Tabelle 10 bestätigt die insgesamt bessere Prognosegüte des auf der alternativen Punktquelle basierenden Modells. Auf Simulationen, welche in insgesamt stärkeren Erdbeben resultieren, fällt die Verbesserung der Prognosen dabei recht deutlich aus. Die MSE-Werte bestätigen darüber hinaus jedoch auch die eingangs angebrachte Vermutung, dass das Hypozentrum bei sehr kleinen Beben im Vergleich zur alternativen Punktquelle wohl näher an der für die Model-

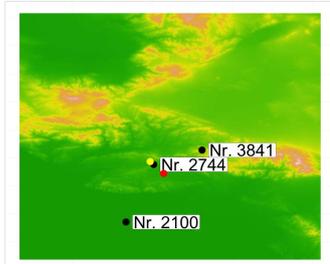
lierung informativsten Punktquelle liegt, an welcher auf der Bruchfläche die größte Bewegung stattfindet. Entsprechend hat das Modell basierend auf dem Hypozentrum für Simulationen, welche in eben solchen kleineren Erdbeben resultieren, durchaus einen kleineren MSE auf den Testdaten.

Zusammenfassend wird durch Vergleich der Ergebnisse basierend auf den beiden evaluierten Punktquellen deutlich, dass die Punktquelle durchaus einen relevanten Einfluss auf die Prognosen und damit auch auf die Modellschätzung aufweist, wobei letzteres hier nicht betrachtet wurde. Die optimale Punktquelle für die Modellierung stellt pro Simulation grundsätzlich jener Punkt auf der Bruchfläche dar, an welchem die größte Bewegung registriert wird, da dieser im Allgemeinen am informativsten für die Ausbreitung des Bebens ist. Da dieser Punkt vor Durchführung der Simulationen nicht ohne Weiteres lokalisierbar ist, sollte das Ziel sein eine möglichst nahe alternative Punktquelle zu wählen. Je weiter entfernt die Punktquelle gewählt wird, desto „falschere“ Abstände zur Punktquelle fließen (durch die Variable *hypo.dis*) in die Modellierung ein und desto schlechter kann erwartungsgemäß der zugehörige Effekt geschätzt werden. Da die hypozentrale Distanz in den Modellen zur Erklärung der Bodenbewegungen einen recht starken Einfluss aufweist kann nicht ausgeschlossen werden, dass dadurch auch die Schätzung der übrigen Effekte negativ beeinflusst wird. Wie in Anhang F zu sehen ist unterscheiden sich die basierend auf der alternativen Punktquelle geschätzten Effekte zumindest für die vorliegende Datensituation jedoch nicht stärker von den Effekten des Modells, welches das Hypozentrum als Punktquelle annimmt.

Aufgrund der Tatsache, dass der sich auf der Bruchfläche befindliche Punkt, an welchem die größte Bewegung stattfindet, wie erwähnt je nach Simulation variiert ist eine allgemeine Empfehlung, auf welche Punktquelle sich der Anwender letztendlich stützen sollte schwierig. Grundsätzlich zeigte sich jedoch, dass das Hypozentrum als Punktquelle besser zur Modellierung der Bodenbewegungen kleinerer Beben geeignet ist, die alternative Punktquelle dagegen besser für die Modellierung der Bodenbewegungen, welche durch größere Erdbeben verursacht werden.



Lage der betrachteten Seismographen



Seismograph	<i>hypo.dis</i>	<i>hypo.dis_{neu}</i>	<i>z</i>	Landform
Nr. 2744	20799	17648	374	Ebene
Nr. 3841	37031	38495	1304	Steiler Gipfel
Nr. 2100	52347	58509	-50	Ebene

Simulation	<i>magnitude_{sum}</i>	<i>struct</i>	<i>l</i>	μ_s	μ_d	<i>angle</i>
58	$1,17 \cdot 10^{22}$	<i>sediment</i>	0,8	0,61	0,48	33
77	$7,36 \cdot 10^{22}$	<i>rock</i>	1,1	0,63	0,27	28
88	$29,00 \cdot 10^{22}$	<i>rock</i>	0,8	0,75	0,10	27

Abbildung 24: Vergleich der Prognosen des finalen Modells ohne den Quellparameter basierend auf dem Hypozentrum (HZ) und dem Modell basierend auf der alternativen Punktquelle (HZ) mit den wahren Beobachtungen anhand dreier Simulationen und Seismographen; Die Seismographen wurden ausgewählt nach variierendem Abstand zum Hypozentrum, Simulationen nach variierendem μ_d ; in rot markiert in der Karte ist das Epizentrum, in gelb der Punkt der größten mittleren Bodenbewegung, welcher senkrecht auf die Bruchfläche projiziert die alternative Punktquelle darstellt; *hypo.dis_{neu}* gibt die Abstände der Seismographen zur alternativen Punktquelle an

5.4 Erklärung der Momentenmagnitude durch Simulationsparameter

Um ein besseres Verständnis der auf der Bruchfläche stattfindenden geophysikalischen Prozesse zu erlangen wird nun untersucht, auf welche Art die Simulationsparameter den in der Modellierung betrachteten Quellparameter $magnitude_{sum}$ beeinflussen. Grundlage hierfür waren - anders als in der übrigen Modellierung - alle 150 Simulationen, für die Daten zum Quellparameter vorlagen. Da sowohl die Simulations- als auch die Quellparameter innerhalb einer Simulation jeweils konstant gehalten wurden betrug die Stichprobengröße für die folgende Analyse gleich 150. Folglich musste bei der Berechnung der evaluierten Modelle - erneut im Gegensatz zu den Hauptmodell dieses Kapitels - nicht auf die besonders effizienten Schätzalgorithmen zurückgegriffen werden. Die evaluierten (Generalisierten) Additiven Modelle wurden mit der Funktion `gam` aus dem `mgcv`-Paket geschätzt. Darüber hinaus war auch die in Kapitel 5.2 beschriebene Modellwahl zur Selektion der Knotenzahlen nicht notwendig, da die Penalisierung der Splines nicht durch die nun deutlich geringere Stichprobengröße negativ beeinflusst wurde.

In der Modellierung wurden alle metrischen Simulationsparameter mittels auf kubischen P-Splines und 5 Knotenpunkten basierenden glatten Effekten berücksichtigt. Die Zielgröße $magnitude_{sum}$ wurde wegen des extremen Wertebereichs wieder mit dem Faktor $\frac{1}{10^{22}}$ skaliert. Aufgrund der Rechtsschiefe des Quellparameters erwies sich ein Normalverteilungsmodell mit log-transformierter Zielgröße bei Betrachtung der Residuen als deutlich adäquater im Vergleich zu demselbigen unter Benutzung des untransformierten Quellparameters. Als weitere Alternativen zu ersterem Modell wurden weiterhin (1) ein Normalverteilungsmodell mit log-Link, (2) ein Gamma-Modell mit log-Link sowie (3) eine t-Verteilung für den Response bei erneuter Verwendung der log-transformierten Zielgröße evaluiert. All diese Modell erwiesen sich bei erneuter Betrachtung der Residuen als schlechter bzw. nicht besser im Vergleich zum Normalverteilungsmodell mit log-Transformation.

Basierend auf dem oben ausgewählten Modell wurden weiterhin drei Interaktionseffekte jeweils einzeln testweise in das Modell aufgenommen. Grundlage für die Beschränkung auf diese drei Interaktionen waren jeweils auffällige Strukturen in der deskriptiven dreidimensionalen Betrachtung der Kovariablen und der Zielgröße. Die Interaktionen $ms : struct$, $ms : I_{l \leq 0,5}$ sowie $angle : struct$ wurden jeweils testweise in das Modell aufgenommen, anhand des AIC erwies sich lediglich die Aufnahme von letzterer Interaktion als sinnvoll. Eine Übersicht über die Ergebnisse des finalen Modells findet sich in Tabelle 11. Die glatten Effekte sind zusätzlich in Abbildung 25 dargestellt.

Variable	glatter Effekt	edf/Schätzer*	p-Wert
Intercept	✗	1,79	-
$struct_{sediment}$	✗	0,05	0,1390
l	✓	3,82	< 0,0001
μ_s	✓	1,94	0,0013
μ_d	✓	3,12	< 0,0001
$angle : struct_{rock}$	✓	3,19	< 0,0001
$angle : struct_{sediment}$	✓	3,75	< 0,0001

*: edf bei glatten Effekten, Schätzer bei skalaren Effekten

Tabelle 11: Ergebnisse des Normalverteilungs-Modells mit log-transformierter Zielgröße

Der Intercept lässt sich hier als die mittlere logarithmierte Magnitude interpretieren, falls die Simulation einen Gesteinsboden umfasst sowie die übrigen Kovariablen ungefähre Werte von $l = 0,7$, $\mu_s = 0,68$, $\mu_d = 0,3$, $angle = 30$. In diesem Fall wird eine mittlere Magnitude von $\exp(1,79) = 5,99$ durch das Modell erwartet.

Für $struct_{sediment}$ wurde ein linearer Effekt von 0,05 geschätzt. In Simulationen mit einem aus Sediment bestehenden Untergrund wurde also im Mittel eine um den Faktor $\exp(0,05)=1,05$ - also um 5% - erhöhte Magnitude beobachtet als in Simulationen mit Gesteinsböden. Dass bei den weicheren Sedimentböden eine größere aggregierte Verschiebung der Messpunkte auf der Bruchfläche (Definition der skalaren Momentenmagnitude) erwartet wird erscheint aus physikalischer Sicht sinnvoll. Im Gegensatz zu den übrigen Simulationsparametern ist die Bodenbeschaffenheit die einzige Variable, welche keinen signifikanten Effekt aufweist.

Das lineare slip-weakening l beeinflusst die Magnitude auf nichtlineare Art. Die Magnitude nimmt dabei immer weiter ab, je höher diese Distanz ausfällt, über welche der Reibungswiderstand vom statischen auf den dynamischen Wert abfällt. Lediglich im unteren Wertebereich bis ca. 0,3 lässt sich ein gegenläufiger Effekt beobachten. Verglichen mit dem Wert $l = 1,5$, für welchen die geringste Magnitude erwartet wird, ist die größte erwartete Magnitude bei $l = 0,3$ um ca. 45% erhöht.

Der statische Reibungskoeffizient μ_s weist einen ungefähr quadratischen Effekt auf. Je höher der anfängliche Reibungswiderstand ist, desto höher fällt dabei die mittlere Magnitude aus, wobei dieser Effekt ab ca. $\mu_s = 0,7$ stagniert. Vergleicht man $\mu_s = 0,75$ - den Wert des größten geschätzten Effektes - mit dem Wert des geringsten geschätzten Effektes $\mu_s = 0,6$, so werden für Simulationen mit ersterem Wert im Mittel um 22% erhöhte Werte der skalaren Momentenmagnitude erwartet.

Unter allen Simulationsparametern wirkt sich der dynamische Reibungskoeffizient μ_d am stärksten auf die skalare Momentenmagnitude aus. Die größten Magnituden ergeben sich bei kleinen μ_d -Werten, bei steigenden Reibungskoeffizienten fallen die Magnituden deutlich ab. Die mittleren Werte der Magnitude sind bei $\mu_d = 0,1$ im Vergleich zu $\mu_d = 0,5$ um den Faktor 13,9 bzw. um 1390% erhöht.

Der Winkel der Hintergrundspannung *angle* weist bei Simulationen mit Gesteinsböden und Simulationen mit Sedimentböden einen etwas unterschiedlichen Einfluss auf die Magnitude auf. Bei beiden Untergründen lässt sich die Struktur erkennen, dass größere Winkel zu einer Abnahme der Magnitude führen. Bei Gesteinsböden verläuft diese Abnahme dabei relativ gleichmäßig und stagniert ab einem Winkel von ca. 33° . In Simulationen mit einem aus Sediment bestehenden Untergrund dagegen fällt die Magnitude bis zu einem Winkel von ca. 27° schneller ab als im übrigen Wertebereich und die weitere Abnahme verläuft weniger linear als bei Gesteinsböden. Bei einem Gesteinsuntergrund wird bei einem Winkel von 25° eine um 70% erhöhte Magnitude erwartet im Vergleich zu einem Winkel von 34° . Bei Sedimentböden beträgt der maximale Unterschied 84% und ergibt sich beim Vergleich der Winkel 25° und $34,5^\circ$.

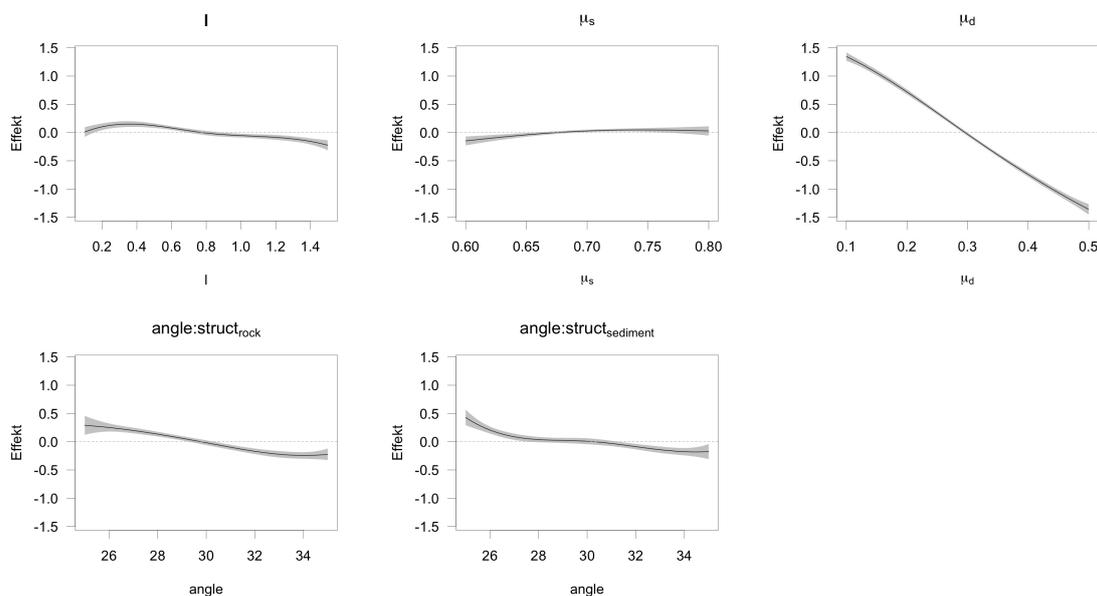


Abbildung 25: Glatte Effekte der Simulationsparameter; grau eingezeichnet sind bayesianische Konfidenzbänder (Wood 2006)

Der Anteil der erklärten Devianz des geschätzten Modells beträgt 98,2%. Die aggregierte Momentenmagnitude lässt sich folglich fast komplett durch die fünf betrachteten Simulationsparameter erklären. Ein QQ-Plot sowie eine Darstellung der Residuen findet sich in Abbildung 26. Insgesamt lässt sich das verwendete Modell als für die Datensituation passabel beschreiben.

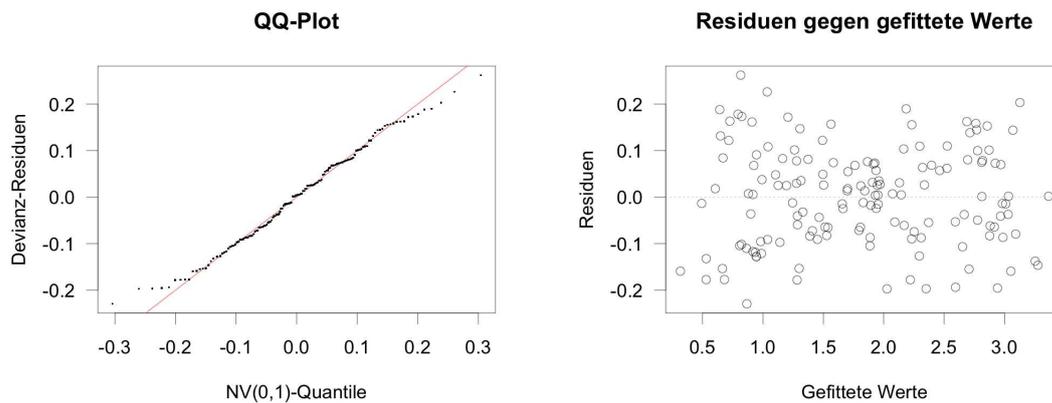


Abbildung 26: *QQ-Plot und Darstellung der Residuen gegen die gefitteten Werte*

5.5 Nebenfragestellungen

Nach der Untersuchung der beiden Hauptfragestellungen werden nun die beiden Nebenfragestellungen bearbeitet. Konkret wird dabei auf die Fragestellungen eingegangen, ob sich die Verläufe der Bodenbewegungen in einem ebenen Gebiet anders verhalten als in einer gebirgigen Region (Abschnitt 5.5.1) und ob sich die simulierten Beben möglicherweise auf nicht-isotrope Weise ausbreiten (5.5.2).

5.5.1 Unterschiedliche Ausbreitung nach Geländeform

Grundlage für die Hypothese, dass sich die ausbreitenden Druckwellen der Erdbeben unterschiedlich auswirken, wenn die Messung in einer weitläufigen Ebene oder einem gebirgigen Gebiet vorgenommen wurde, bildet die Überlegung, dass die Wellen vor Erreichen einer Messstation in einer gebirgigen Region durch die raue Erdoberfläche öfter reflektiert und dadurch abgeschwächt werden. Bei der Ausbreitung der Erdbebenwellen hin zu einer Messstation auf einer Ebene findet eine solche Reflektierung dagegen in deutlich geringerem Umfang statt. Dass diese Struktur grundsätzlich in den Simulationsdaten wiederzufinden ist, ist in Abbildung 10 der Deskription ersichtlich, welche sich auf die Zeitpunkte bezieht, zu welchen pro Seismograph die maximale Bodenbewegung registriert wurde. In den Ebenen ist darin eine deutlich gleichmäßigere Struktur zu erkennen als in einem Großteil der gebirgigen Gebiete.

Da sich die durch Abbildung 10 aufgekommene Hypothese insbesondere darauf bezieht, dass bei Messungen in Gebirgen der in einem vergleichbaren ebenen Gebiet gemessene maximale Ausschlag oft ausbleibt oder später stattfindet, lässt sich die Fragestellung im finalen Modell dadurch überprüfen, ob es einen unterschiedlichen Effekt der hypozentralen Distanz gibt, je nachdem ob die Messung in einer Ebene aufgenommen wurde oder

nicht. Ein Vergleich des finalen Modells der Hauptfragestellung mit demselbigen Modell inklusive des eben erörterten Interaktionseffektes findet sich in Tabelle 12.

	Gamma ohne Interaktion	Gamma mit Interaktion
MSE auf Testdaten	0,0757	0,0873
Anteil erklärter Devianz	70,7%	71,0%

Tabelle 12: Vergleich des finalen Modells der Hauptfragestellung mit und ohne Interaktionseffekt bezüglich des MSE auf den Testdaten sowie dem Anteil der erklärten Devianz

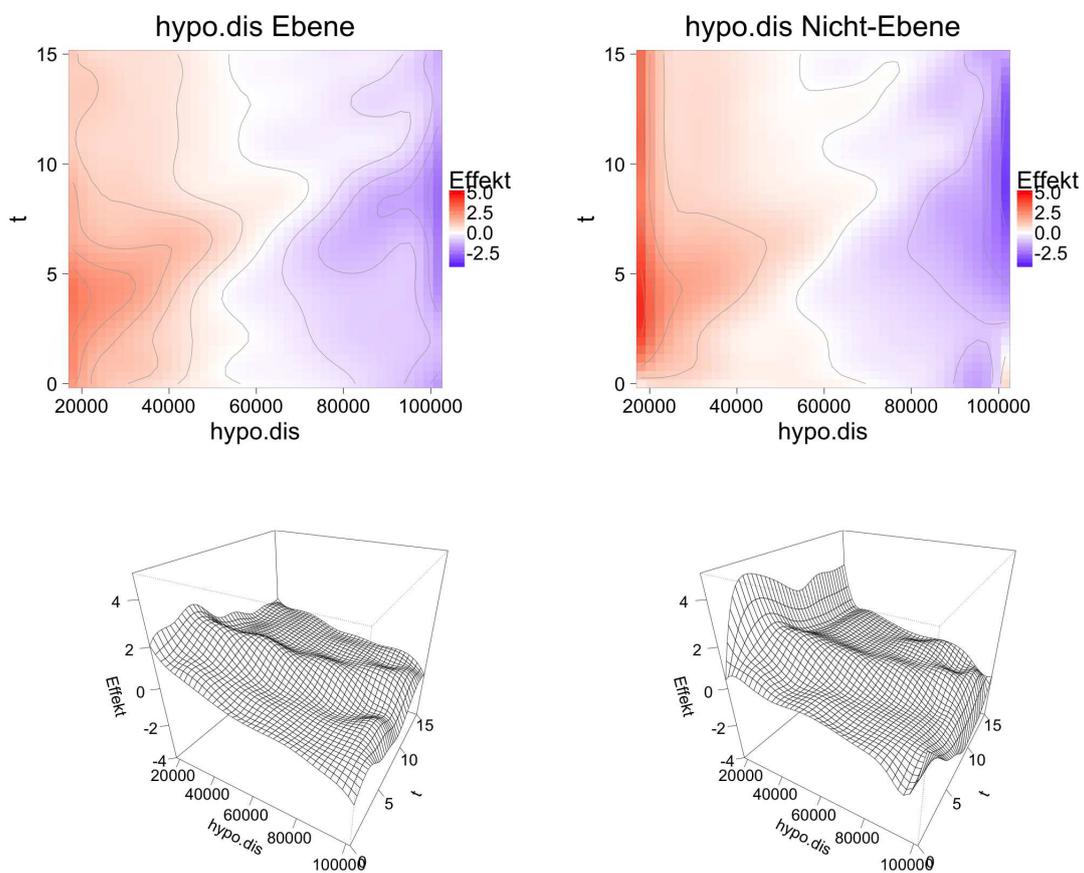


Abbildung 27: Darstellung der beiden durch Einbezug der Interaktion geschätzten Effekte der hypozentralen Distanz für Messstationen in einer Ebene (linke Spalte) und Messstationen, welche sich nicht in einer Ebene befinden (rechte Spalte)

Basierend auf dem auf den Testdaten berechneten MSE der beiden Modelle schneidet das Modell inklusive der Interaktion deutlich schlechter ab als das Modell ohne die Interaktion. Da der Anteil der erklärten Devianz außerdem steigt, kann gefolgert werden, dass die Aufnahme des Interaktionseffektes hier lediglich zu Overfitting auf den Trainingsdaten

führt. Betrachtet man die in Abbildung 27 dargestellten glatten Effekte ist darüber hinaus bis auf die Ränder kein größerer Unterschied zwischen den beiden geschätzten Effekten zu erkennen. Die Hypothese kann also abgelehnt werden.

5.5.2 Nicht-isotroper Richtungseffekt

Eine möglicherweise bestehende nicht-isotrope Ausbreitung der Erdbeben wird durch Einbezug eines zeitkonstanten, zweidimensionalen glatten Effektes der Koordinaten der Seismographen basierend auf dem finalen Modell der Hauptfragestellung untersucht. Um eine sinnvolle Interpretation dieses räumlichen Effektes sicherzustellen wurde die hypozentrale Distanz als Einflussvariable hierzu aus dem Modell genommen. Der räumliche Effekt gibt somit das mittlere Niveau der über die Zeit gemessenen Bodenbewegungen pro Ort an, wenn die übrigen Kovariablen konstant gehalten werden. Dargestellt ist der geschätzte Effekt in Abbildung 28.

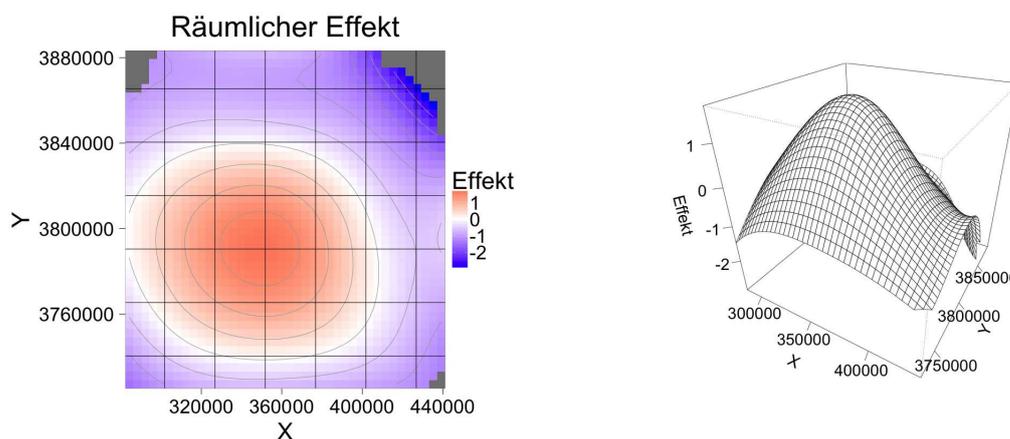


Abbildung 28: Darstellung des glatten räumlichen Effektes der Koordinaten; Mittelpunkt des Rasters in der linken Grafik ist der Punkt des größten geschätzten Effektes; das Raster hat eine Auflösung von jeweils 25km in X- und Y-Richtung

Anhand des Konturplots lässt sich keine richtungsabhängige Struktur erkennen. Bei Konstanthalten der übrigen Einflussgrößen variiert das allgemeine Niveau der Bodenbewegungen über den Raum sehr gleichmäßig. Angemerkt werden muss jedoch, dass die Hypothese durch Einbezug des räumlichen Effektes in das Modell im Rahmen dieser Arbeit eben nur diesbezüglich untersucht wurde, wie sich das Niveau der über die Zeit gemessenen Bodenbewegungen über den Raum unterscheidet. Insbesondere die aufgrund von Abbildung 9 vermuteten Strukturen, dass die durch die Erdbeben verursachten Bodenbewegungen je nach Ausbreitungsrichtung früher oder später einsetzen, lassen sich auf diese Weise nicht überprüfen.

6 Fazit und Ausblick

Zusammenfassend kann gesagt werden, dass der gewählte Modellierungsansatz eines funktionalen Regressionsmodells unter Annahme einer Gammaverteilung für die Zielgröße insgesamt eine gute Modellierung der über die Zeit gemessenen Bodenbewegungen ermöglicht. Der verwendete, in der Funktion bam implementierte effiziente Schätzalgorithmus war dabei eine zentrale Grundlage für die Anwendung. Insbesondere aufgrund des großen Umfangs der Daten und der darauf basierend notwendig gewordenen Modellwahl, um eine möglichst gute Schätzung der wahren Effekte sicherzustellen, wäre die durchgeführte Analyse auf Basis alternativer, weniger effizienter Algorithmen, deutlich aufwendiger gewesen. Der verwendete Ansatz zur Klassifizierung der Topographie in die Landformen ermöglichte eine adäquate Repräsentierung der verschiedenen Landschaftsformen in der Modellierung und kann aufgrund der Flexibilität durch die individuell wählbaren Radien für zukünftige Anwendungen weiterempfohlen werden.

Neben dem Befund, dass der dynamische Reibungswiderstand μ_d neben dem Abstand zur angenommenen Punktquelle den stärksten Effekt auf die Bodenbewegungen aufweist, wurde weiterhin festgestellt, dass eine zusätzliche Aufnahme der skalaren Momentenmagnitude als zentralem Quellparameter in ein Modell, welches bereits die Simulationsparameter als Einflussgrößen enthält, keine Zusatzinformation in Bezug auf die Vorhersage der Bodenbewegungen beinhaltet. Letzteres Ergebnis stimmt damit überein, dass die Magnitude anhand der Simulationsparameter sehr gut vorhergesagt werden kann. Bei Untersuchung der in die Modelle einbezogenen Topographie war zu beobachten, dass in Ebenen insgesamt etwas geringere Bodenbewegungen gemessen werden als in gebirgigeren Gebieten. Bezüglich der Punktquelle zeigte sich, dass die Annahme der konkret verwendeten Punktquelle einen durchaus relevanten Einfluss auf die Modellgüte aufweist.

Für eine mögliche zukünftige Weiterentwicklung des vorgestellten Modellierungsansatzes bestehen unterschiedliche Ansatzpunkte. Grundsätzlich stellt eine über alle Simulationen konstante Punktquelle eine unrealistische Annahme dar und kann sich - wie gezeigt - negativ auf die Modellgüte auswirken. Eine genauere Bestimmung der pro Simulation optimalen Punktquelle ist basierend darauf zu empfehlen. Für eine optimale Prognose der Bodenbewegungen für neue physikalische Bedingungen könnte - auch unter Einbezug der genaueren Beschaffenheit der Bruchfläche - weiterhin evaluiert werden, ob beziehungsweise wie gut eben jene optimalen Punktquellen vor Durchführung der Simulation vorhergesagt werden können. Darüber hinaus könnten auch weitere Eigenschaften der Bruchfläche beziehungsweise des genauen Bruchverlaufs bezüglich eines Einbezugs in die Modellierung evaluiert werden.

7 Literatur

- Biek, Verena und Bumsun Park (2015). “*Statistische Analyse erdbebenbedingter Bodenbewegungen und Erdbebenquellodynamik*”. Unveröffentlichter Abschlussbericht des Kurses „Statistisches Consulting“ im Master Statistik der LMU München.
- Breslow, Norman E und David G Clayton (1993). “Approximate inference in generalized linear mixed models”. In: *Journal of the American statistical Association* 88.421, S. 9–25.
- Dumbser, Michael und Martin Käser (2006). “An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes—II. The three-dimensional isotropic case”. In: *Geophysical Journal International* 167.1, S. 319–336.
- Gelman, Andrew et al. (2014). *Bayesian data analysis*. Bd. 2. Chapman & Hall/CRC Boca Raton, FL, USA.
- Goldsmith, Jeff et al. (2016). *refund: Regression with Functional Data*. R package version 0.1-15. URL: <https://CRAN.R-project.org/package=refund>.
- Gu, Chong (1992). “Cross-validating non-Gaussian data”. In: *Journal of Computational and Graphical Statistics* 1.2, S. 169–179.
- Heinecke, Alexander et al. (2014). “Petascale high order dynamic rupture earthquake simulations on heterogeneous supercomputers”. In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE Press, S. 3–14.
- Jenness, Jeff (2006). “Topographic Position Index (tpi_jen. avx) extension for ArcView 3. x, v. 1.3 a. Jenness Enterprises”. In: URL: <http://www.jennessent.com/arcview/tpi.htm>.
- Loehlin, John C (2004). *Latent variable models: An introduction to factor, path, and structural equation analysis*. Psychology Press.
- McKay, MD und RJ Beckman (1979). “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”. In: *TECHNOMETRICS* 21.2.
- Morris, Jeffrey S (2014). “Functional regression”. In: *arXiv preprint arXiv:1406.4068*.
- NatCatSERVICE (2016). *Münchener Rückversicherungs-Gesellschaft NatCatSERVICE*. https://www.munichre.com/site/corporate/get/documents_E2109916638/mr/assetpool.shared/Documents/5_Touch/_NatCatService/Significant-Natural-Catastrophes/2015/1980_2015_Welt_all_eco_d.pdf. Zugegriffen am 21.11.2016.

- Pelties, Christian, AA Gabriel und JP Ampuero (2013). “Verification of an ADER-DG method for complex dynamic rupture problems”. In: *Geoscientific Model Development Discussion* 6, S. 5981–6034.
- Pelties, Christian et al. (2012). “Three-dimensional dynamic rupture simulation with a high-order discontinuous Galerkin method on unstructured tetrahedral meshes”. In: *Journal of Geophysical Research: Solid Earth* 117.B2.
- R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Ramsay, James O und Bernard W Silverman (2005). *Functional Data Analysis*. Springer.
- Reiss, Philip T und R T Ogden (2009). “Smoothing parameter selection for a class of semiparametric linear models”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2, S. 505–523.
- Scheipl, Fabian, Jan Gertheiss, Sonja Greven et al. (2016). “Generalized functional additive mixed models”. In: *Electronic Journal of Statistics* 10.1, S. 1455–1492.
- Scheipl, Fabian, Ana-Maria Staicu und Sonja Greven (2015). “Functional additive mixed models”. In: *Journal of Computational and Graphical Statistics* 24.2, S. 477–501.
- Stöcker, Almond und Tobias Kühn (2015). “*Statistische Analyse erdbebenbedingter Bodenbewegungen und Erdbebenquellendynamik*”. Unveröffentlichter Abschlussbericht des Kurses „Statistisches Consulting“ im Master Statistik der LMU München.
- Weiss, Andrew (2001). “Topographic position and landforms analysis”. In: *Poster presentation, ESRI user conference, San Diego, CA*. Bd. 200.
- Wood, Simon N (2006). *Generalized additive models: an introduction with R*. CRC press.
- Wood, Simon N (2011). “Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models”. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73.1, S. 3–36.
- Wood, Simon N (2013). “On p-values for smooth components of an extended generalized additive model”. In: *Biometrika* 100, S. 221–228.
- Wood, Simon N, Yannig Goude und Simon Shaw (2015). “Generalized additive models for large data sets”. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 64.1, S. 139–155.
- Wood, Simon N et al. (2016). “Generalized additive models for gigadata: modelling the UK black smoke network daily data”. In: *Journal of the American Statistical Association* just-accepted, S. 1–40.

8 Anhang

A Daten (digital)

B R-Codes (digital)

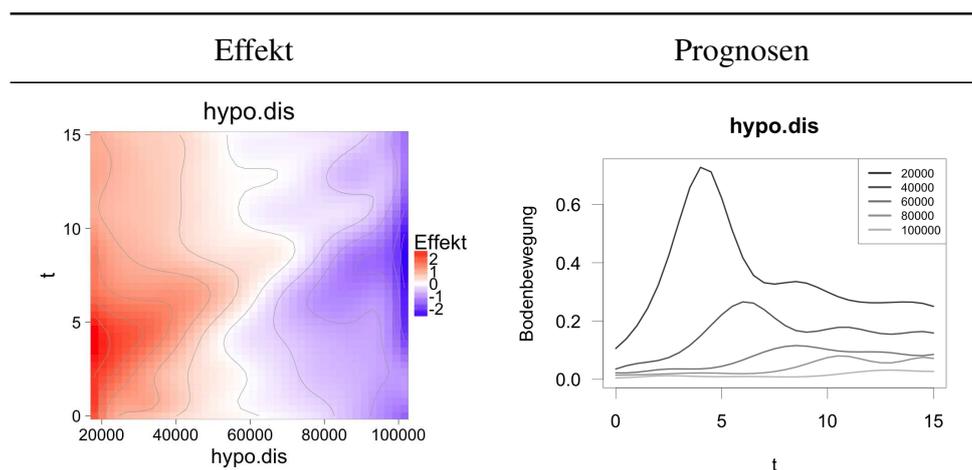
C Finale Modelle zu Haupt- und Nebenfragestellungen (digital)

D Evaluation der Kovariablen-Effekte

Im Folgenden werden die Effekte der aus der Modellwahl resultierten Modelle graphisch evaluiert. Grundlage sind die physikalischen Gegebenheiten von Simulation Nr. 56 sowie Seismograph Nr. 1. Die Werte aller auf diese Weise festgesetzten Einflussvariablen finden sich nach diesem Absatz tabellarisch dargestellt. Pro Kovariable wurde darauf basierend eine Grafik erstellt, in welcher die prognostizierten Bodenbewegungen für unterschiedliche Werte der Kovariable abgetragen sind. Die übrigen Einflussgrößen werden dabei konstant gehalten. Für das finale Modell ohne Quellparameter wird darüber hinaus eine Übersicht über alle geschätzten Effekte gegeben.

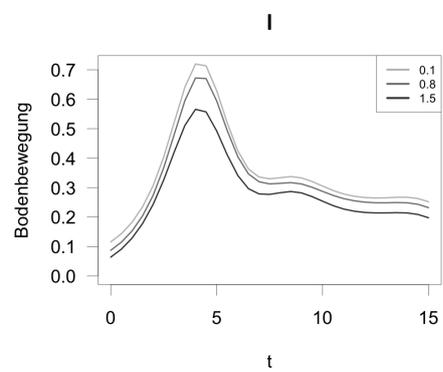
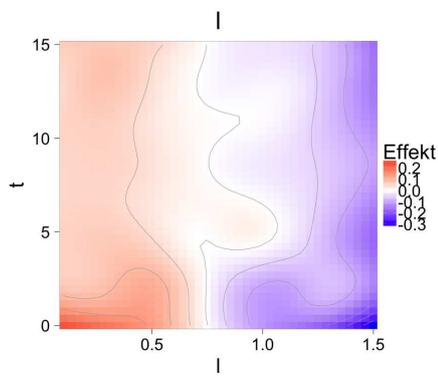
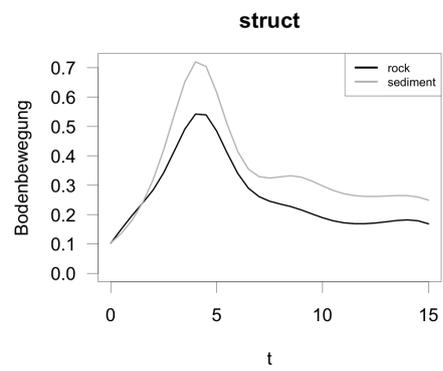
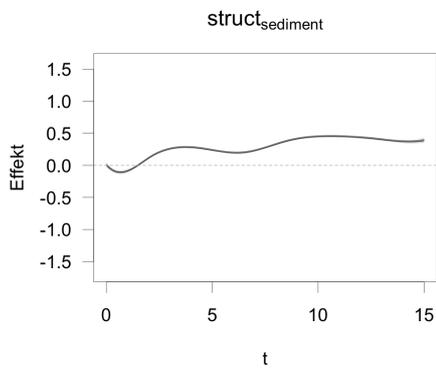
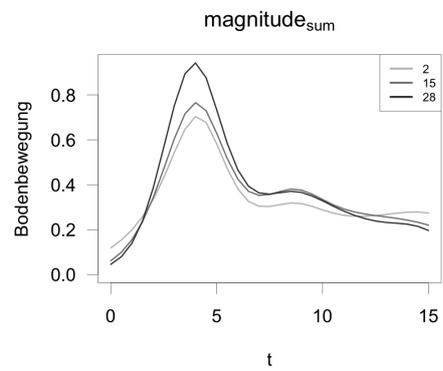
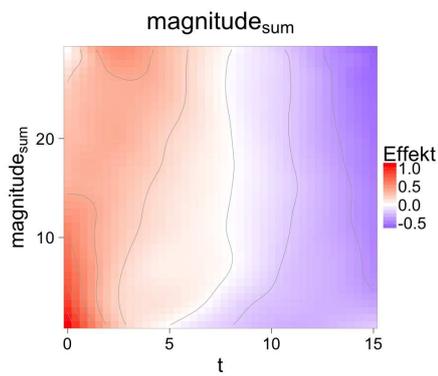
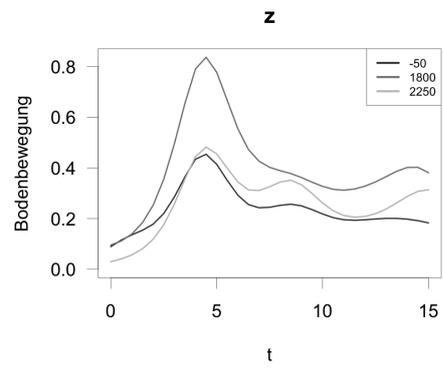
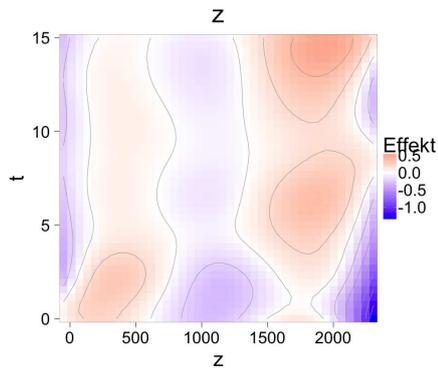
<i>hypo.dis</i>	<i>z</i>	Landform	<i>magnitude_{sum}</i>	<i>struct</i>	<i>l</i>	μ_s	μ_d	<i>angle</i>
20098	367	Ebene	3,53	<i>sediment</i>	0,5	0,75	0,43	31

Finales Modell inklusive Quellparameter



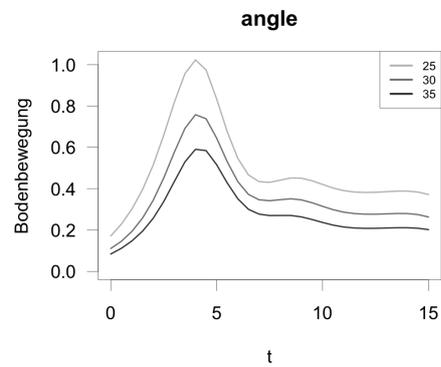
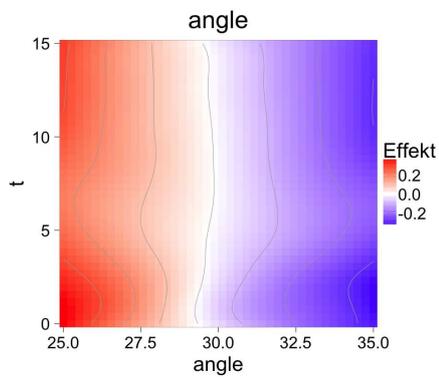
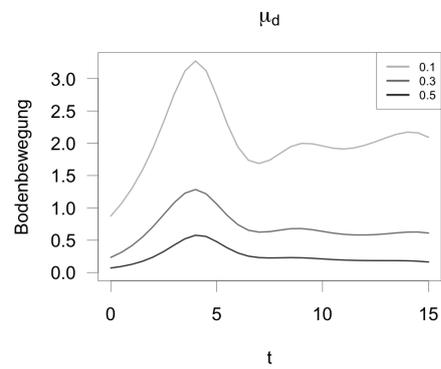
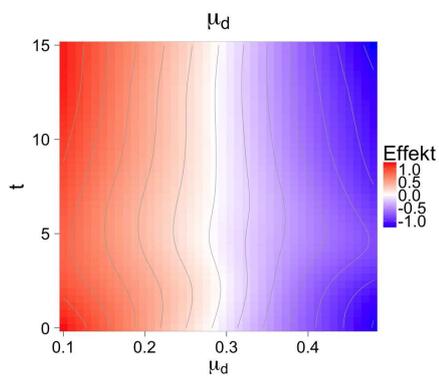
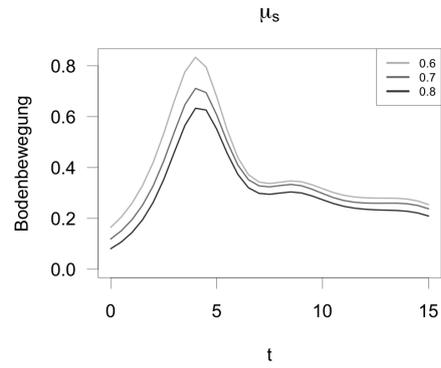
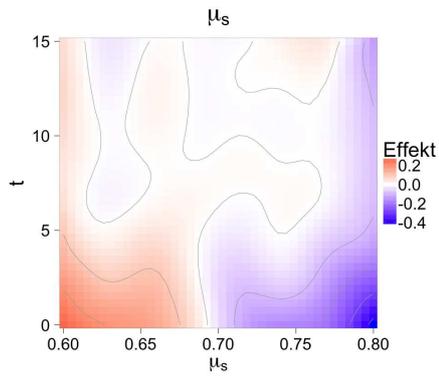
Effekt

Prognosen



Effekt

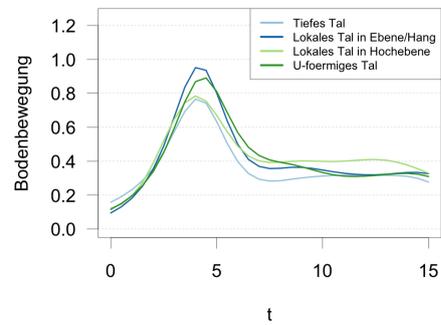
Prognosen

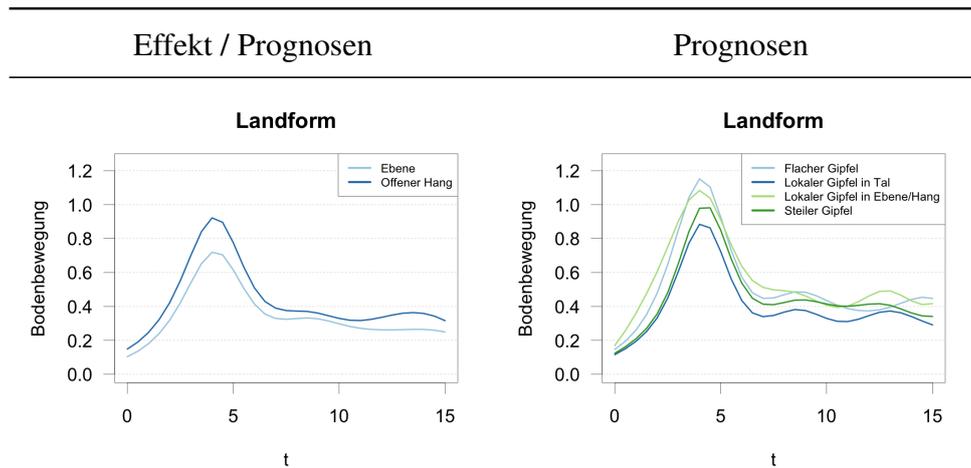


Landform

Landform-Effekte:
siehe Ergebnis-Kapitel

Landform





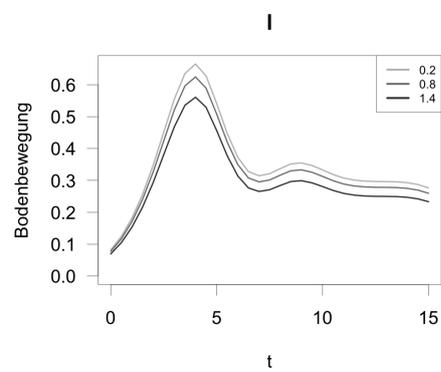
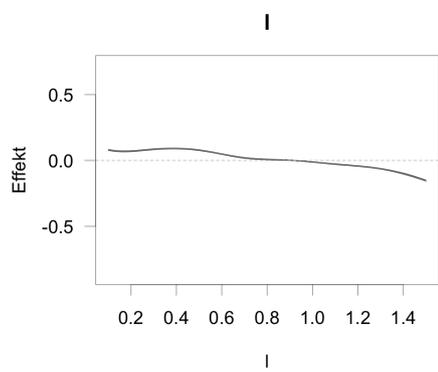
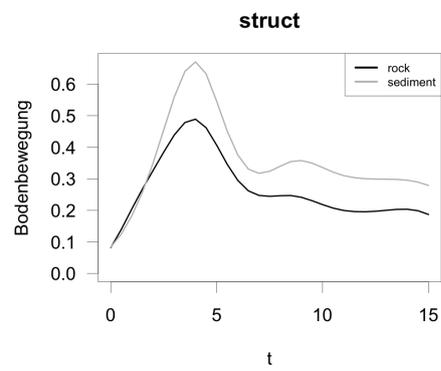
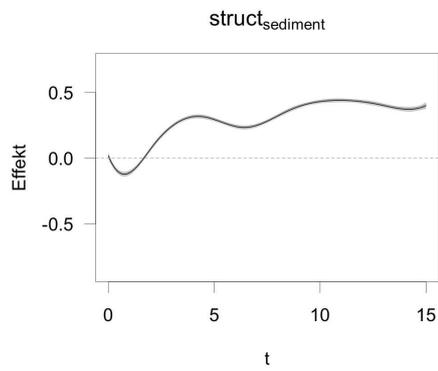
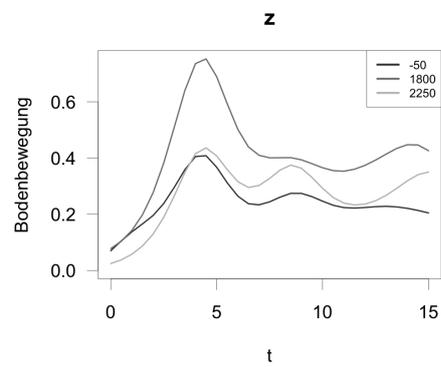
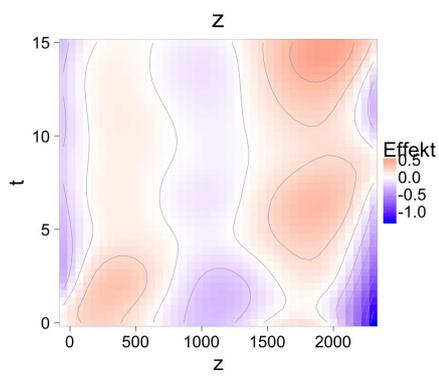
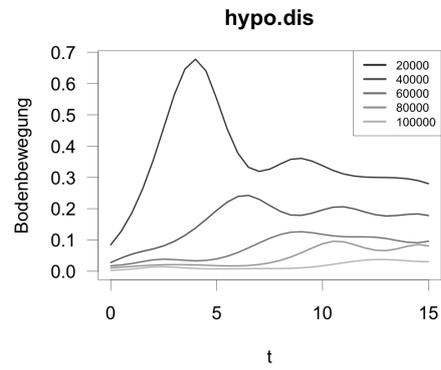
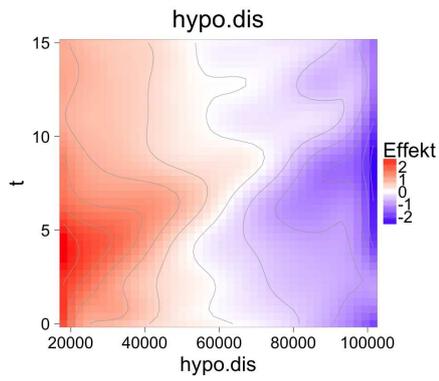
Finales Modell ohne Quellparameter

Variable	glatter Effekt	edf/Schätzer*	p-Wert	
Intercept (skalar)	✗	0,92	-	
Intercept	✓	8,82	-	
<i>hypo.dis</i>	✓	85,39	< 0,0001	
<i>z</i>	✓	36,73	< 0,0001	
Simulation	<i>struct</i> _{sediment}	✓	9,90	< 0,0001
	<i>l</i>	✓	7,56	< 0,0001
	μ_s	✓	69,88	< 0,0001
	μ_d	✗	-5,48	< 0,0001
	<i>angle</i>	✓	6,64	< 0,0001
Landform	$LF_{\text{Tiefes Tal}}$	✓	8,34	< 0,0001
	$LF_{\text{Lokales Tal in Ebene/Hang}}$	✓	8,05	< 0,0001
	$LF_{\text{Lokales Tal in Hochebene}}$	✓	9,48	< 0,0001
	$LF_{\text{U-förmiges Tal}}$	✓	8,21	< 0,0001
	$LF_{\text{Offener Hang}}$	✓	7,25	< 0,0001
	$LF_{\text{Flacher Gipfel}}$	✓	8,32	< 0,0001
	$LF_{\text{Lokaler Gipfel in Tal}}$	✓	9,15	< 0,0001
	$LF_{\text{Lokaler Gipfel in Ebene/Hang}}$	✓	9,59	< 0,0001
	$LF_{\text{Steiler Gipfel}}$	✓	9,39	< 0,0001

*: edf bei glatten Effekten, Schätzer bei skalaren Effekten

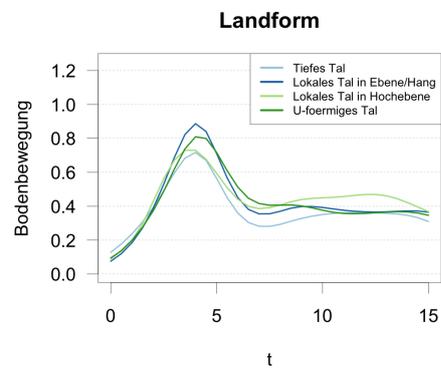
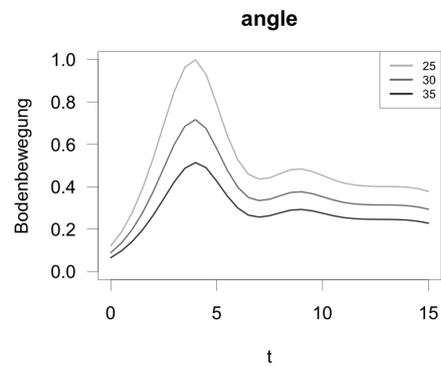
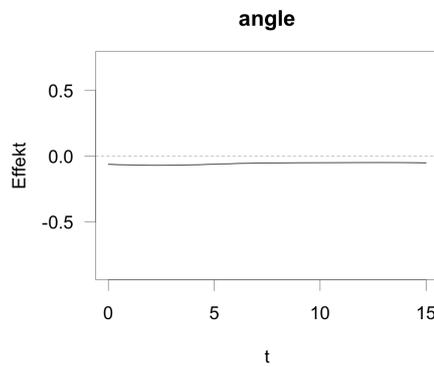
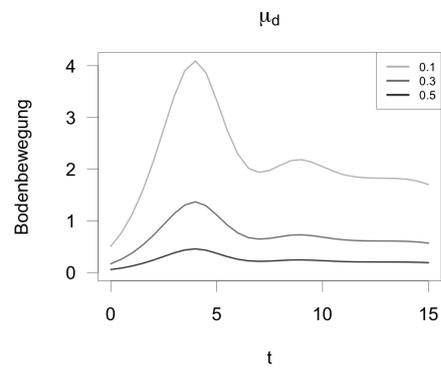
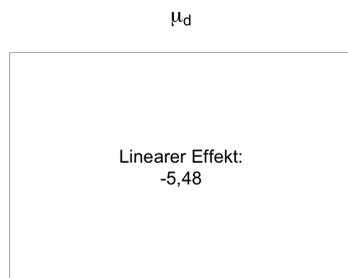
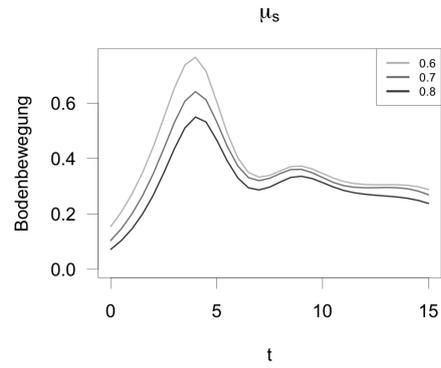
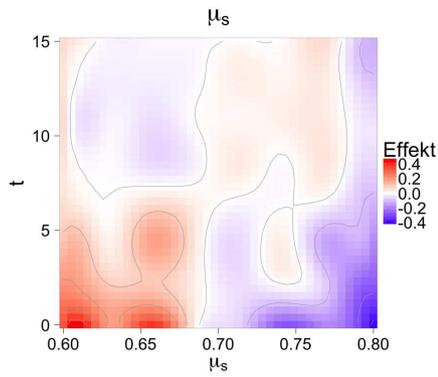
Effekt

Prognosen



Effekt

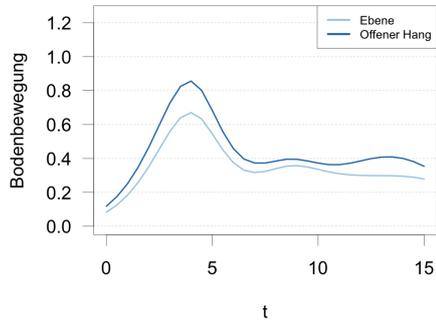
Prognosen



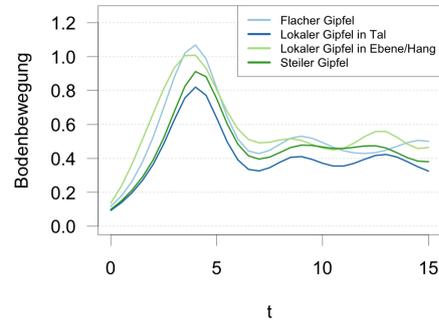
Effekt / Prognosen

Prognosen

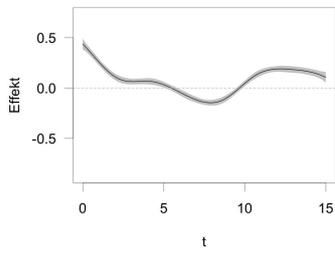
Landform



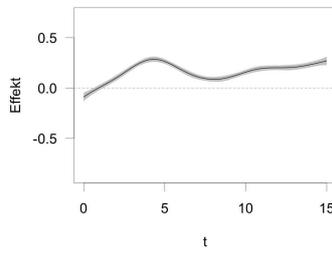
Landform



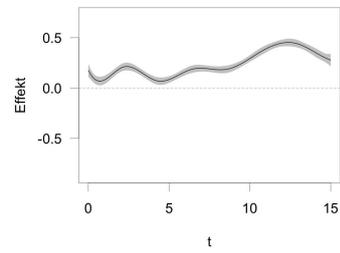
LF_{Tiefes Tal}



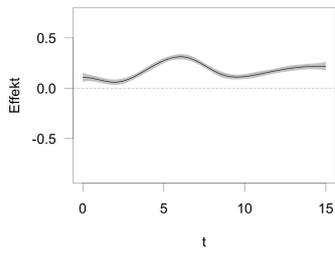
LF_{Lokales Tal in Ebene/Hang}



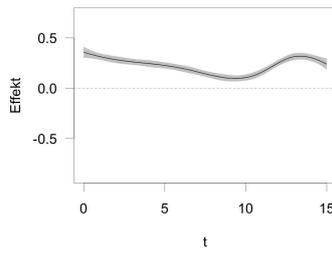
LF_{Lokales Tal in Hochebene}



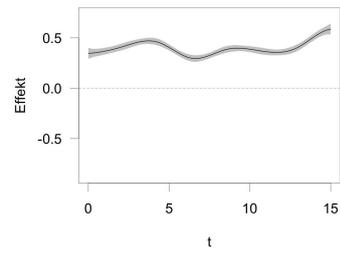
LF_{U-förmiges Tal}



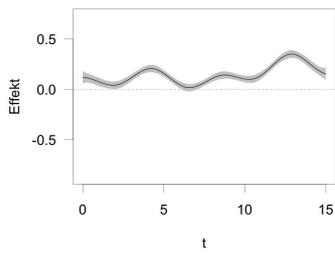
LF_{Offener Hang}



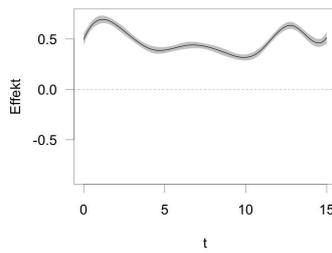
LF_{Flacher Gipfel}



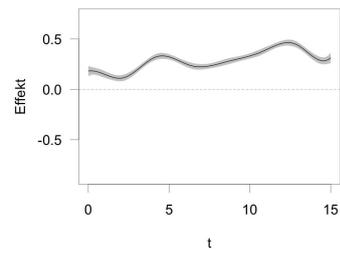
LF_{Lokaler Gipfel in Tal}



LF_{Lokaler Gipfel in Ebene/Hang}

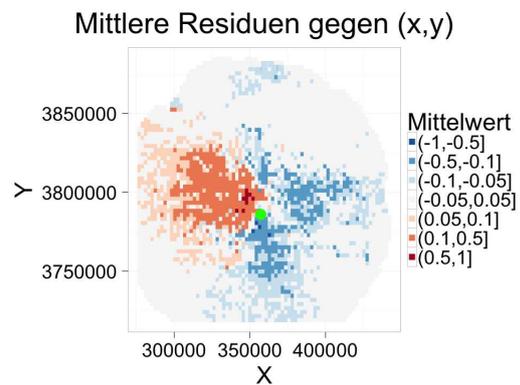
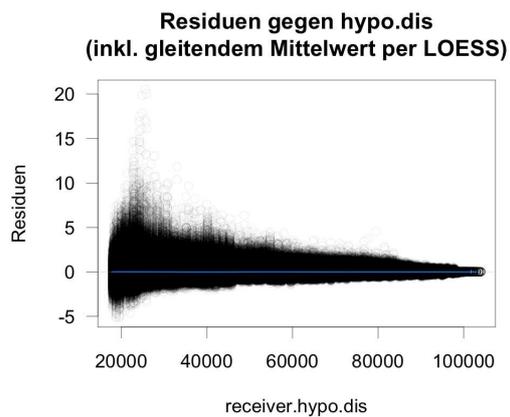
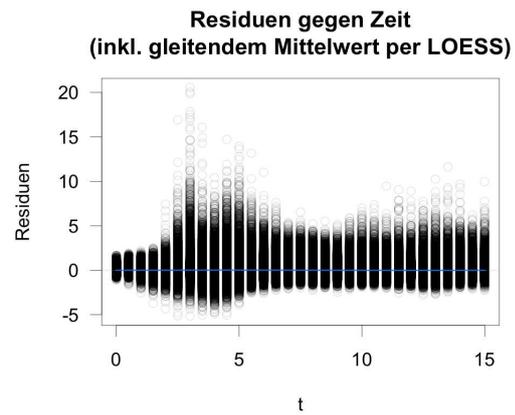
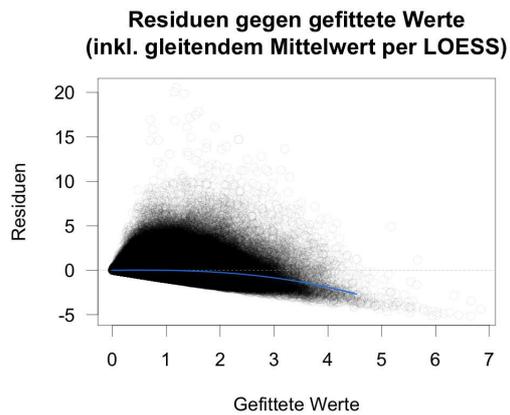


LF_{Steiler Gipfel}



E Residualplots Finales Modell inklusive Quellparameter

Zum durch die Modellwahl erhaltenen Modell inklusive des Quellparameters finden sich hier ausgewählte Residualplots. Grundlage hierfür sind wie im Hauptteil der Arbeit die aus den Trainings- und Testdaten zusammengefassten Daten, die Residualplots nur bezüglich der Trainingsdaten weisen eine vergleichbare Form auf.



F Modellergebnisse alternative Punktquelle

Im Folgenden sind die Ergebnisse basierend auf der alternativen Punktquelle, des senkrecht auf die Bruchfläche projizierten Punktes der größten mittleren Bodenbewegung, dargestellt. Es werden sowohl die Ergebnisse für das volle Modell inklusive des Quellparameters gezeigt als auch die Ergebnisse des Modells ohne Quellparameter.

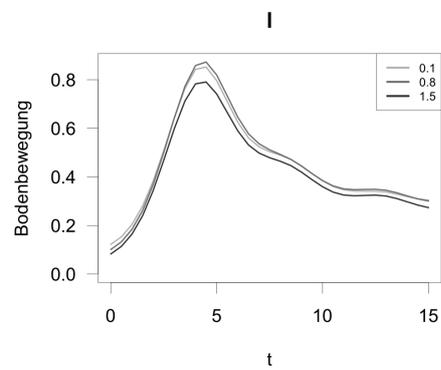
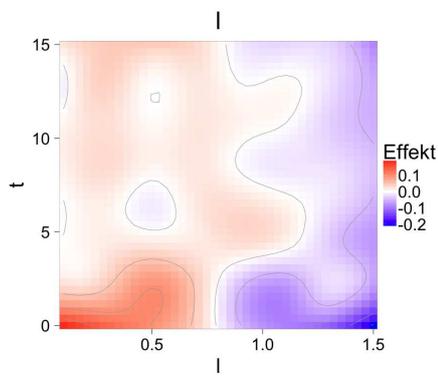
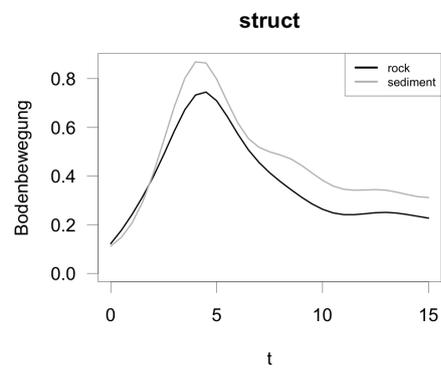
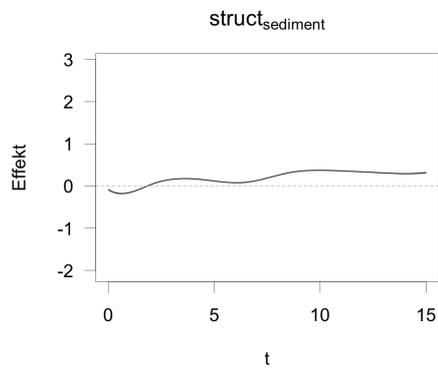
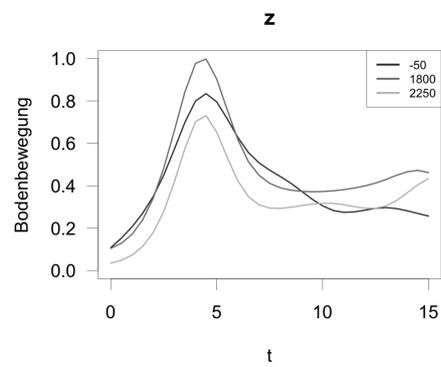
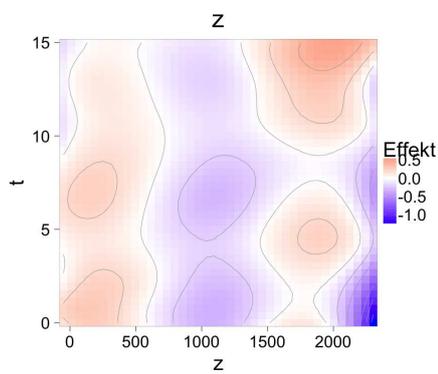
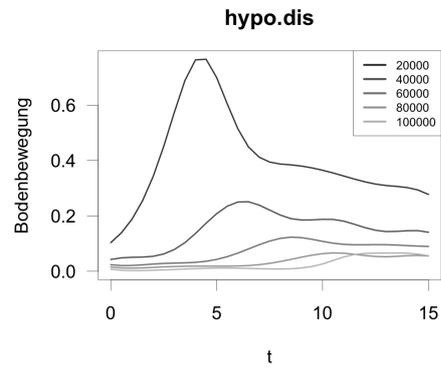
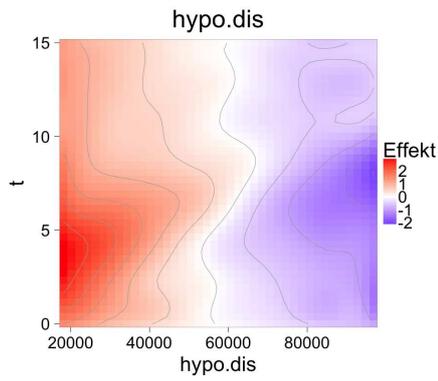
Modell inklusive Quellparameter

Variable	glatter Effekt	edf/Schätzer*	p-Wert	
Intercept (skalar)	✗	-2,35	-	
Intercept	✓	8,37	-	
<i>hypo.dis</i>	✓	84,94	< 0,0001	
<i>z</i>	✓	37,07	< 0,0001	
<i>magnitude_{sum}</i>	✓	63,61	< 0,0001	
Simulation	<i>struct_{sediment}</i>	✓	9,87	< 0,0001
	<i>l</i>	✓	44,77	< 0,0001
	μ_s	✓	14,24	< 0,0001
	μ_d	✓	32,14	< 0,0001
	<i>angle</i>	✓	32,00	< 0,0001
Landform	<i>LF_{Tiefes Tal}</i>	✓	7,58	< 0,0001
	<i>LF_{Lokales Tal in Ebene/Hang}</i>	✓	8,40	< 0,0001
	<i>LF_{Lokales Tal in Hochebene}</i>	✓	9,74	< 0,0001
	<i>LF_{U-förmiges Tal}</i>	✓	6,53	< 0,0001
	<i>LF_{Offener Hang}</i>	✓	8,79	< 0,0001
	<i>LF_{Flacher Gipfel}</i>	✓	9,26	< 0,0001
	<i>LF_{Lokaler Gipfel in Tal}</i>	✓	9,45	< 0,0001
	<i>LF_{Lokaler Gipfel in Ebene/Hang}</i>	✓	9,35	< 0,0001
	<i>LF_{Steiler Gipfel}</i>	✓	9,28	< 0,0001

*: edf bei glatten Effekten, Schätzer bei skalaren Effekten

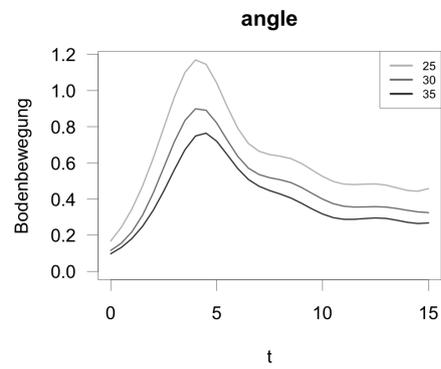
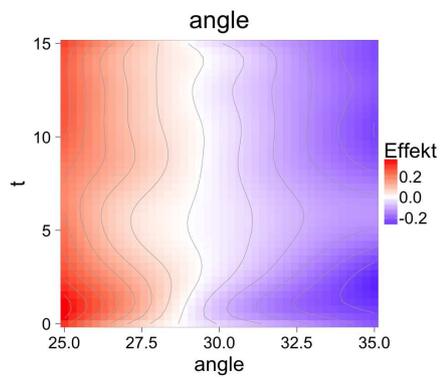
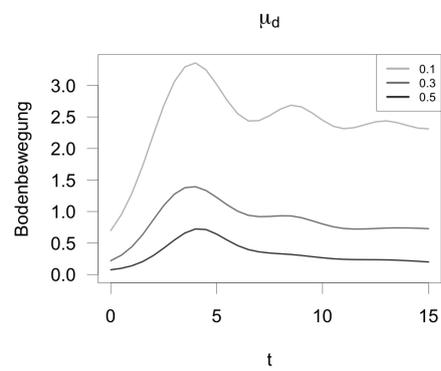
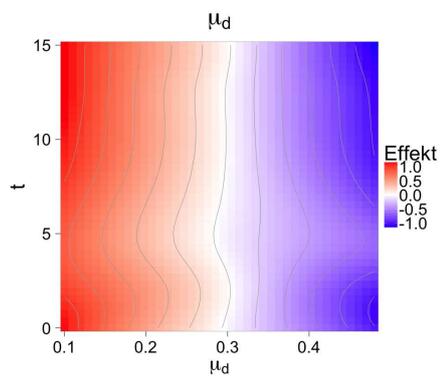
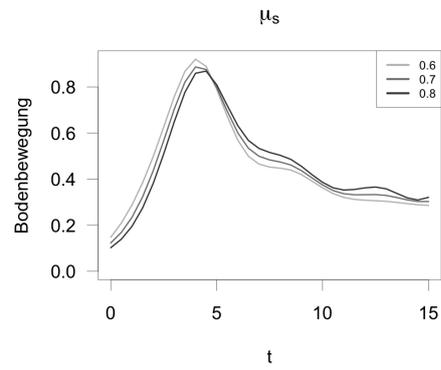
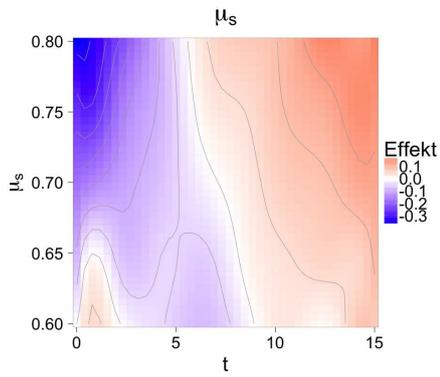
Effekt

Prognosen



Effekt

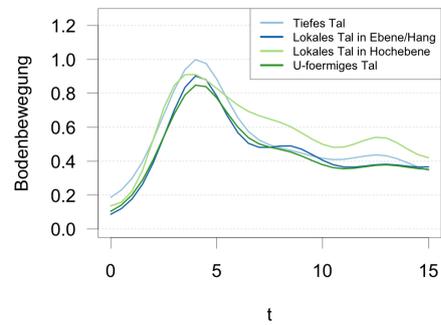
Prognosen



Landform

Landform-Effekte:
siehe nächste Seite

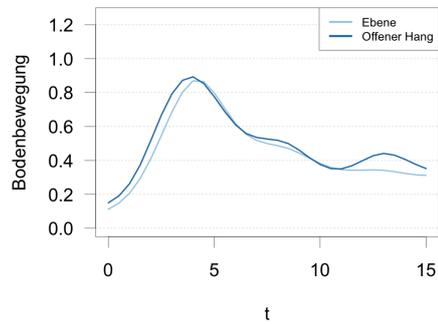
Landform



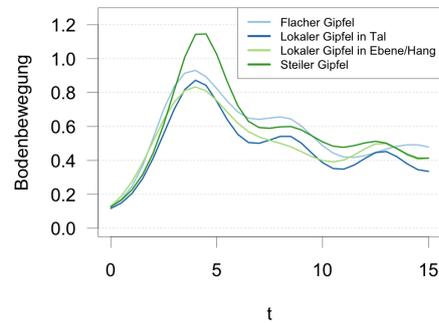
Effekt / Prognosen

Prognosen

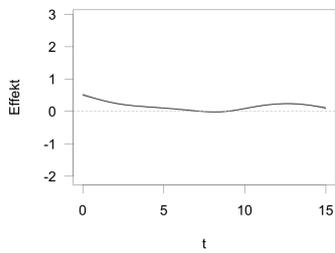
Landform



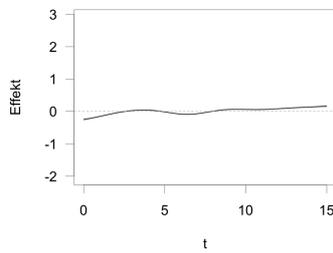
Landform



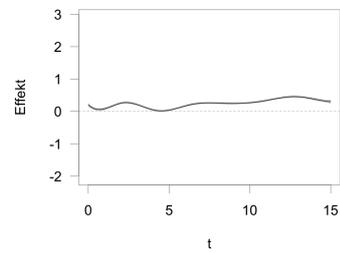
LF_{Tiefes Tal}



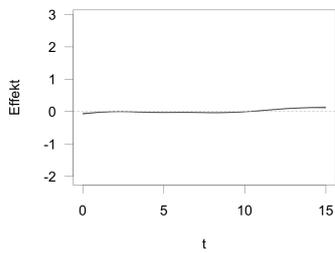
LF_{Lokales Tal in Ebene/Hang}



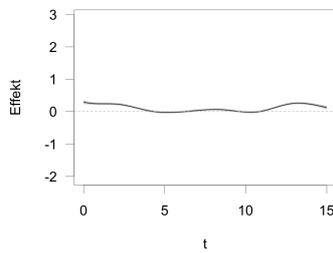
LF_{Lokales Tal in Hochebene}



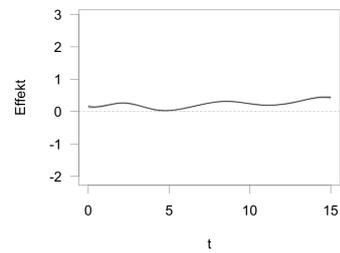
LF_{U-förmiges Tal}



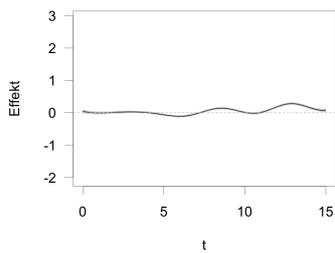
LF_{Offener Hang}



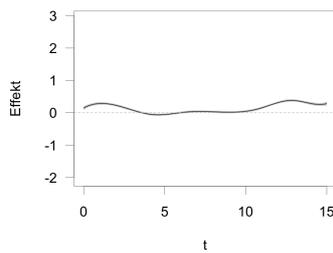
LF_{Flacher Gipfel}



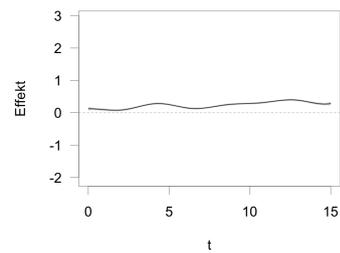
LF_{Lokaler Gipfel in Tal}

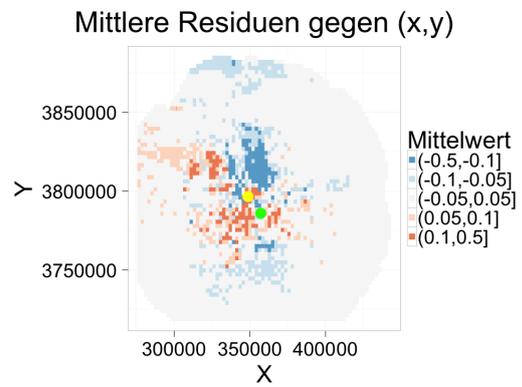
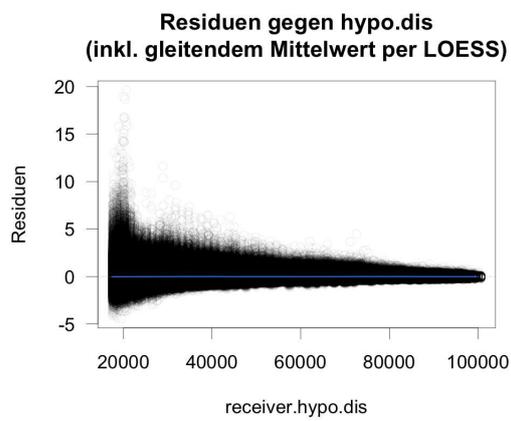
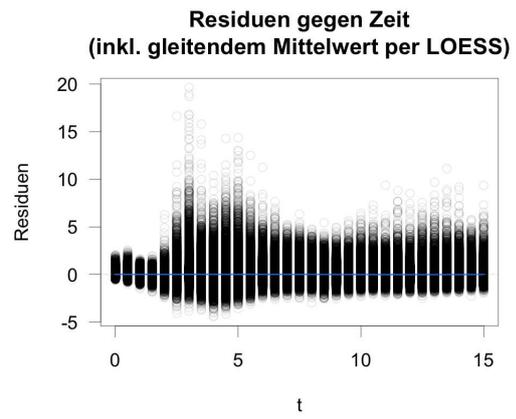
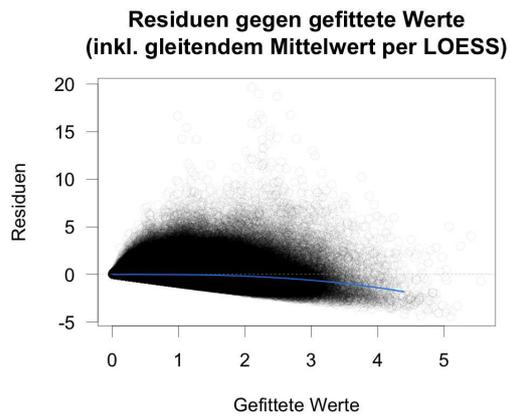


LF_{Lokaler Gipfel in Ebene/Hang}



LF_{Steiler Gipfel}





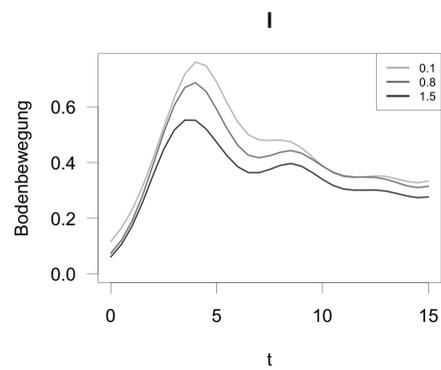
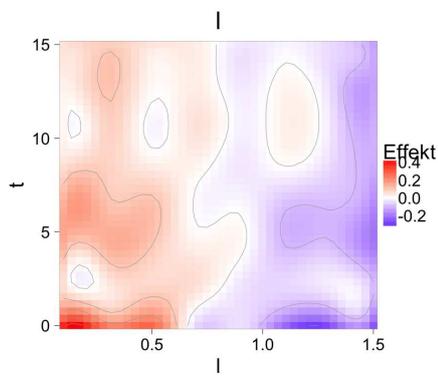
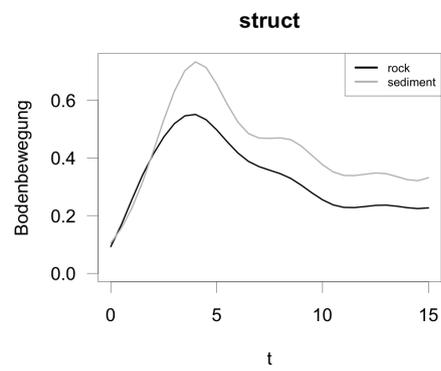
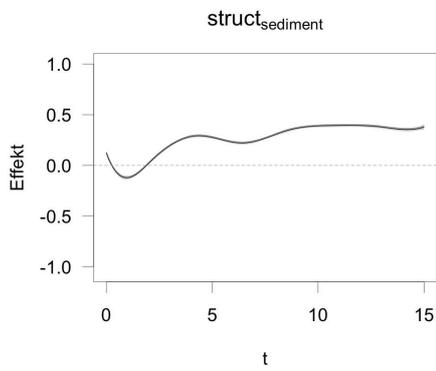
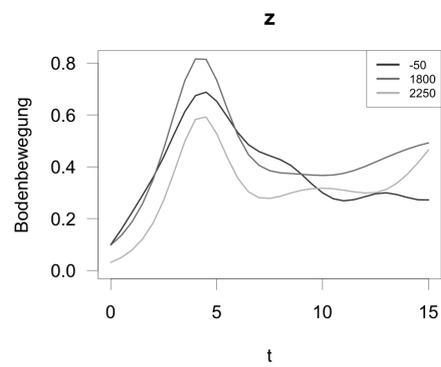
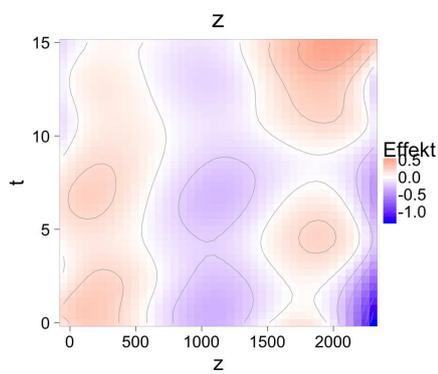
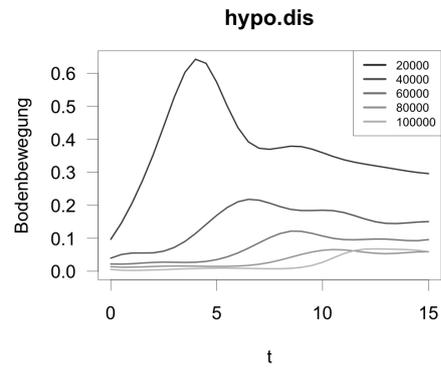
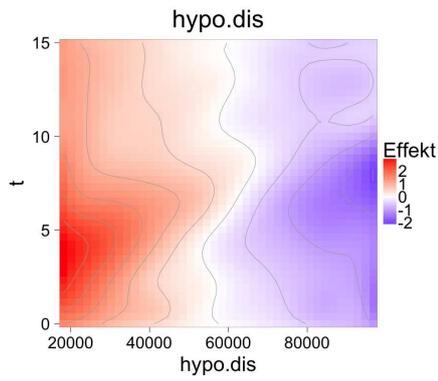
Modell ohne Quellparameter

	Variable	glatter Effekt	edf/Schätzer*	p-Wert
	Intercept (skalar)	✗	-2,30	-
	Intercept	✓	8,97	-
	<i>hypo.dis</i>	✓	84,83	< 0,0001
	<i>z</i>	✓	37,05	< 0,0001
Simulation	<i>struct</i> _{sediment}	✓	9,93	< 0,0001
	<i>l</i>	✓	75,31	< 0,0001
	μ_s	✓	32,95	< 0,0001
	μ_d	✓	3,91	< 0,0001
	<i>angle</i>	✓	3,69	< 0,0001
Landform	<i>LF</i> _{Tiefes Tal}	✓	7,63	< 0,0001
	<i>LF</i> _{Lokales Tal in Ebene/Hang}	✓	8,41	< 0,0001
	<i>LF</i> _{Lokales Tal in Hochebene}	✓	9,73	< 0,0001
	<i>LF</i> _{U-förmiges Tal}	✓	6,42	< 0,0001
	<i>LF</i> _{Offener Hang}	✓	8,76	< 0,0001
	<i>LF</i> _{Flacher Gipfel}	✓	9,25	< 0,0001
	<i>LF</i> _{Lokaler Gipfel in Tal}	✓	9,45	< 0,0001
	<i>LF</i> _{Lokaler Gipfel in Ebene/Hang}	✓	9,35	< 0,0001
	<i>LF</i> _{Steiler Gipfel}	✓	89,28	< 0,0001

*: edf bei glatten Effekten, Schätzer bei skalaren Effekten

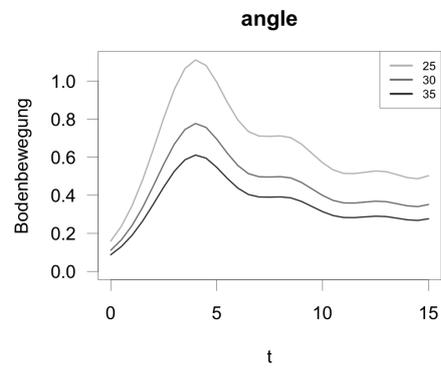
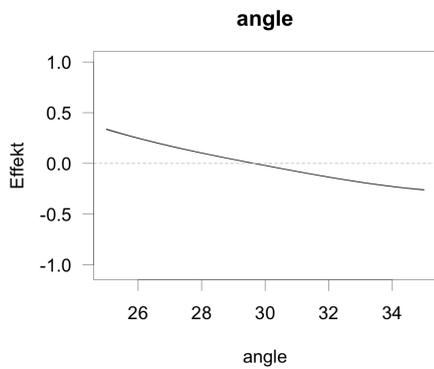
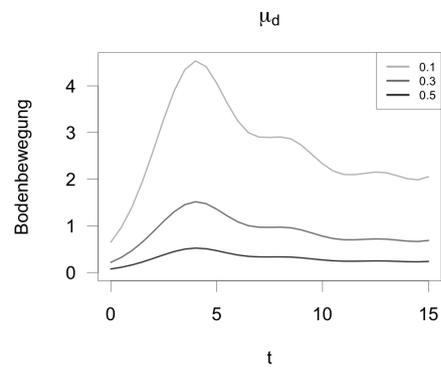
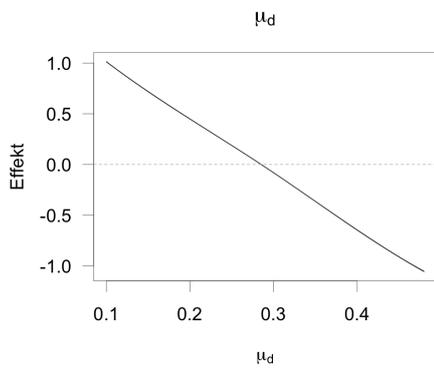
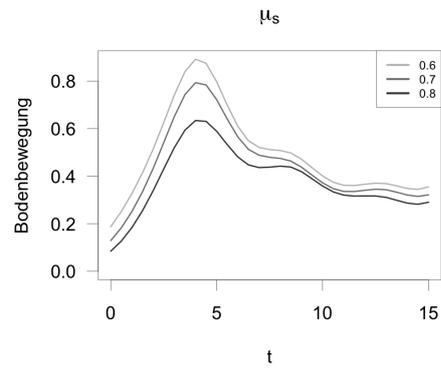
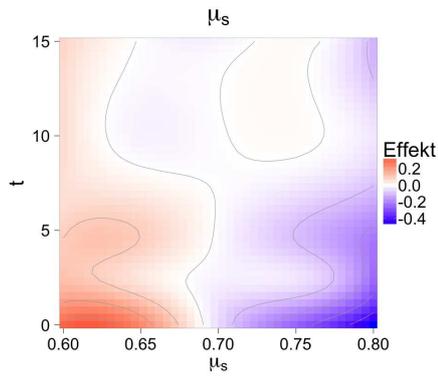
Effekt

Prognosen



Effekt

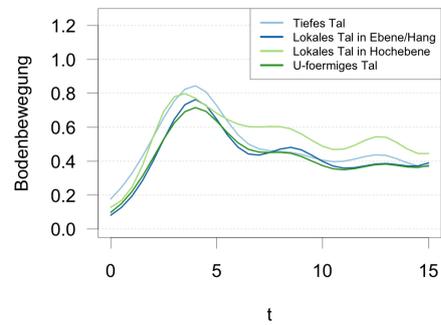
Prognosen



Landform

Landform-Effekte:
siehe nächste Seite

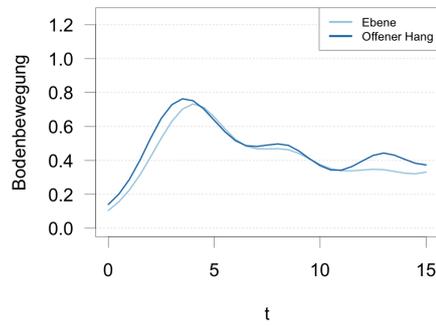
Landform



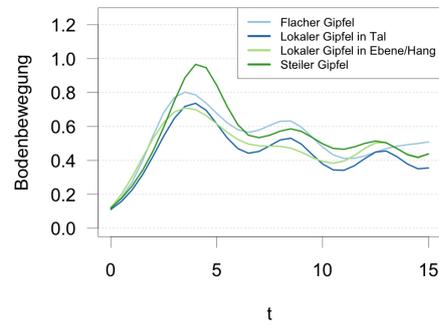
Effekt / Prognosen

Prognosen

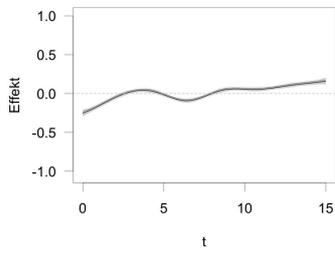
Landform



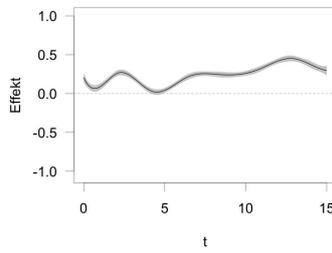
Landform



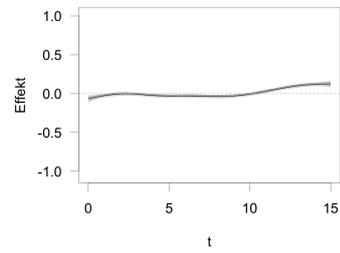
LF_{Tiefes Tal}



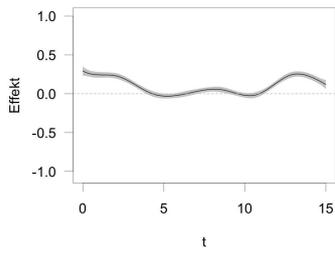
LF_{Lokales Tal in Ebene/Hang}



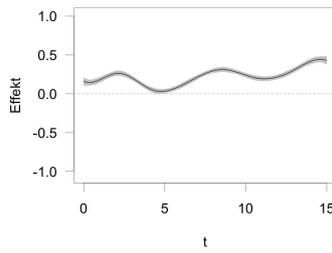
LF_{Lokales Tal in Hochebene}



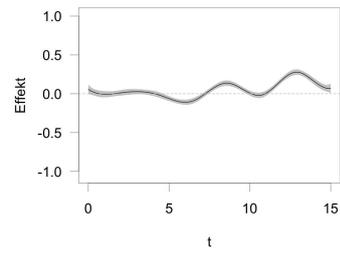
LF_{U-förmiges Tal}



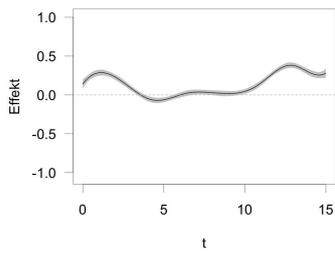
LF_{Offener Hang}



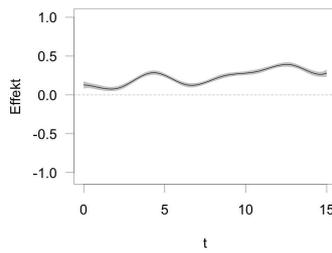
LF_{Flacher Gipfel}



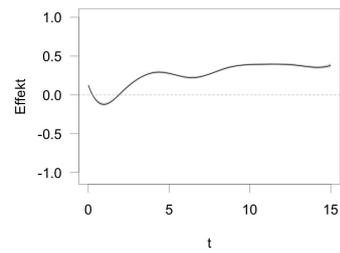
LF_{Lokaler Gipfel in Tal}

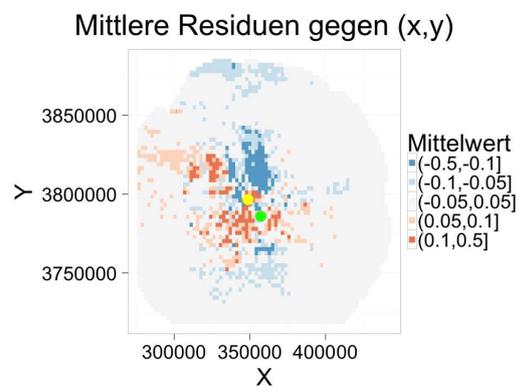
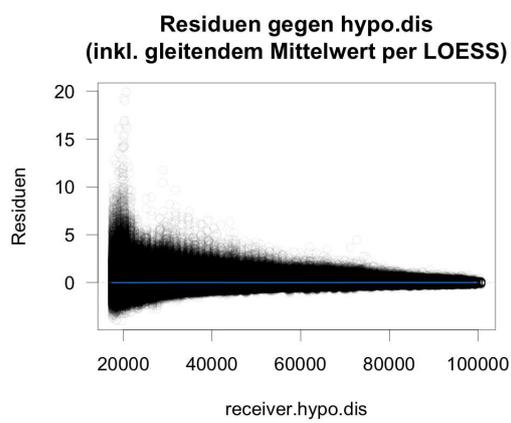
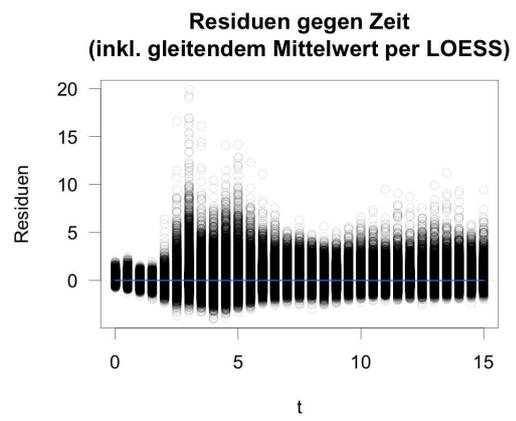
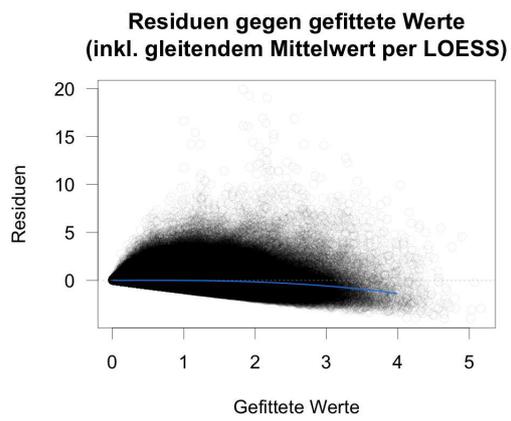


LF_{Lokaler Gipfel in Ebene/Hang}



LF_{Steiler Gipfel}





9 Eigenständigkeitserklärung

Hiermit bestätige ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken (dazu zählen auch Internetquellen) entnommen sind, wurden unter Angabe der Quelle kenntlich gemacht.

München, 28.11.16

Bauer A.

Alexander Bauer