Habitatmodelle für Fledermäuse in Bayern

Bachelorarbeit

von Carola Kobayashi

Betreuung: Prof. Dr. Torsten Hothorn

Ludwig–Maximilians–Universität

Institut für Statistik

Besonderer Dank gilt Herrn Dr. Jörg Müller vom Nationalpark Bayerischer Wald, der die Daten für diese Arbeit zur Verfügung gestellt hat und mir bei Fragen stets hilfreich zur Seite stand. Ebenso möchte ich an dieser Stelle Herrn Dr. Claus Bässler danken, dessen Rat und Anregungen mir wichtige Impulse für den Teil der biologischen Auswertung dieser Arbeit gaben.

Bedanken möchte ich mich auch bei Prof. Dr. Torsten Hothorn sowie bei Esther Herberich und Nikolay Robinzonov für die engagierte Betreuung während Arbeitsphase und die zahlreichen hilfreichen Anregungen.

Für die kleine graphische Auffrischung der Titelseite danke ich meinem Großvater Dipl. Ing. Hans-Georg Weiß.

Inhaltsverzeichnis

1 Einführung

Viele der in Bayern vorkommenden Naturräume stellen einen Lebensraum für die insgesamt 23 dort nachgewiesenen Arten von Fledermäusen dar. Jedoch gelten nach der Roten Liste der gefährdeten Tiere und Gefäßpflanzen Bayerns (Schnappauf und Müller, 2005) sieben dieser 23 Arten als gefährdet, fünf als stark gefährdet, zwei Arten als vom Aussterben bedroht und die Alpenfledermaus (Hypsugo savii Bonaparte) gilt nach dieser Liste als ausgestorben oder verschollen.

Bis ins 18. Jahrhundert hinein wusste man nicht, dass Fledermäuse sich durch Ultraschall orientieren und konnte sich nicht erklären, wie Fledermäuse bei Nacht fliegen können, ohne dabei gegen Hindernisse zu stoßen. So entstand der Volksglaube, die Tiere seien vom Teufel besessen (Drunkenmölle, 2010). Ein weiterer Irrglaube, einheimische Fledermäuse seien Blutsauger, führte bis zum 19. Jahrhundert dazu, dass viele Fledermauskolonien vom Menschen vernichtet wurden. Aber auch heute hat es die Fledermaus nicht leicht. Einer der Hauptgründe für die Gefährdung der Tiere ist die Beeinträchtigung ihrer Quartiere, aber auch Umweltgifte oder die landschaftliche Veränderung der Jagdräume von Fledermäusen können Gründe für den Rückgang der Individuenzahl in Bayern sein. Das Hauptaugenmerk des Fledermausschutzes liegt daher auf der Identifizierung und dem Schutz der Quartiere der Fledermauskolonien (Meschede und Rudolph, 2004). Es gibt viele Faktoren, die die Quartierwahl von Fledermäusen beeinflussen könnten. Zum einen sind Klimafaktoren vorstellbar, zum anderen können auch die Landnutzung und damit eventuell zusammenhängende Nahrungsvorkommen eine wichtige Rolle spielen.

In dieser Arbeit werden Habitatmodelle für Fledermäuse in Bayern angepasst, um herauszufinden, auf welche Weise Klima- und Landbedeckungsvariablen die Artenzahl von Fledermäusen beeinflussen. Eines der Hauptprobleme hierbei ist die Vielzahl an Umweltvariablen, die für eine solche Analyse zur Verfügung stehen. Damit das resultierende Modell nicht unnötig komplex wird, besteht das Ziel in einer effektiven Variablenselektion. Bei der Anpassung von Modellen für die Artenverteilung gibt es zudem einige Pro-

bleme. Es ist unklar, in welcher Weise die Umweltvariablen auf die Zielvariable wirken. Beispielsweise kann es sich um einen linearen oder einen nicht–linearen Effekt handeln. Ein weiteres Problem kann durch räumliche Autokorrelation entstehen. Das Vorkommen von Arten kann sich auf die Nachbarumgebung sowohl positiv als auch negativ auswirken. Auch Nonstationarität, d.h. das Variieren einer Umweltvariable mit dem Raum, kann ein Problem für die Modellierung sein. Für die Modellierung in dieser Arbeit werden Artenverteilungsmodelle verwendet, wie sie von Hothorn *et al.* (2010b) beschrieben werden, deren Grundidee in einer Zerlegung des Prädiktors in eine globale und eine lokale Komponente besteht. Dadurch sollen die Probleme möglicher Nicht–Linearität sowie räumliche Autokorrelation und Nonstationarität berücksichtigt werden. Um eine effiziente Modellwahl und Variablenselektion durchzuführen, wird ein Boosting–Algorithmus verwendet.

2 Material

2.1 Untersuchungsraum

Mit einer Gesamtfläche von 70.547,8 km^2 ist Bayern das größte Bundesland Deutschlands. Mit einem Waldanteil, der etwa ein Drittel der gesamten Fläche ausmacht, handelt es sich um ein relativ waldreiches Land. Der größte Teil dieser Wälder liegt in den Bezirken Oberpfalz, Oberfranken und Unterfranken. Der Wasseranteil an der Gesamtfläche beträgt etwa 1,9%. Den Großteil der Wasserfläche machen die nacheiszeitlich entstandenen Seen in Südbayern aus. 51,6% der Fläche dienen der Primärwirtschaft, dabei werden etwa 36% als Dauergrünland und 63% als Ackerland genutzt. 9,8% ist der Anteil an überbauter Fläche in Bayern.
Das Klima in Bayern ist sehr differenziert, so dass es sowohl in Bezug auf die Temperatur als auch auf den Niederschlag sehr große Unterschiede zwischen den verschiedenen Re-

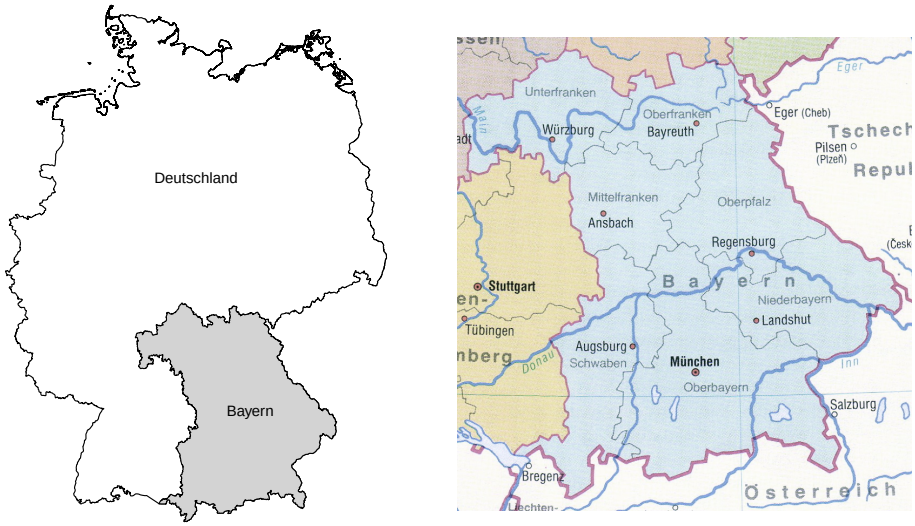


Abbildung 1: Untersuchungsraum Bayern (Bayernkarte aus Zahn (1992)).

gionen gibt. Diese Differenzen werden vor allem durch die Alpen und das Mittelgebirge bewirkt (Meschede und Rudolph, 2004).

2.2 Beschreibung des Datensatzes

Die Daten dieser Arbeit wurden vom Nationalpark Bayerischer Wald zur Verfügung gestellt und wurden aus drei unterschiedlichen Quellen bezogen:
Die Daten über die Fledermäuse stammen aus dem Atlaswerk „Fledermäuse in Bayern“ (Meschede und Rudolph, 2004), das vom Bayerischen Landesamt für Umwelt, dem Landesbund für Vogelschutz in Bayern e.V. und dem Bund Naturschutz Bayern e.V. herausgegeben wurde. Darin wird Bayern in 2.285 Quadranten mit einer durchschnittlichen Größe von 33,9 km^2 eingeteilt. Mit einer 0-1-Kodierung wurde festgehalten, welche

Arten jeweils in einem Quadranten beobachtet wurden. Seit 1985 werden diese Daten vom Bayerischen Landesamt für Umwelt gesammelt und umfassen den Zeitraum von 1985 bis 2008. Benutzt wurden nur die Beobachtungen von April bis Oktober, so dass es sich bei den gefundenen Quartieren um Sommerquartiere und Wochenstuben handelt. Die Sammlung der Daten findet im Rahmen des Artenhilfsprogramms für Fledermausschutz in Bayern statt. Eine wesentliche Einrichtung dieses Programms sind die beiden an den Universitäten München und Erlangen angesiedelten Koordinationsstellen für Fledermausschutz. Ziel der Einrichtung ist die Beratung in allen Fragen des Fledermausschutzes, die Bestandserfassung und Beobachtung der Bestandsentwicklung sowie die Ausführung gezielter Artenhilfsmaßnahmen. Von den 23 nachgewiesenen Arten kommen 22 im Datensatz vor. Für die Alpenfledermaus (Hypsugo savii Bonaparte) gibt es keine aktuellen Vorkommensnachweise. Die Zielvariable der Analyse in dieser Arbeit ist die Variable `Arten`, die zusammenfasst, wie viele verschiedene Arten in einem Quadranten nachgewiesen wurden.

Die Kovariablen setzen sich aus Klima– und Bodennutzungsfaktoren zusammen. Die bioklimatischen Variablen stammen aus dem Projekt WorldClim, bei dem die wichtigsten Klimadaten für alle Regionen der Erde, mit Ausnahme der Antarktis, erfasst wurden. Dazu sind Daten, die weltweit von Wetterstationen gemessen wurden, gesammelt worden. Gemessen wurde der monatliche Niederschlag sowie die mittlere, die minimale und die maximale monatliche Temperatur, die anschließend mit Hilfe eines Algorithmus interpoliert wurden. Daraus leitete man 19 bioklimatische Variablen ab. Der Großteil der Daten stammt aus den Jahren 1960 bis 2000, in einigen Fällen wurden die Daten ab 1950 verwendet. Die Daten liegen jeweils für Zellen von $0{,}93 \times 0{,}93 = 0{,}86$ km^2 vor, was für Klimadaten eine sehr hohe Auflösung ist. Für jeden Quadranten des Fledermausatlasses wurde der Mittelwert der darin liegenden Zellen berechnet. Nähere Informationen zu den WorldClim–Daten sind in Hijmans *et al.* (2005) nachzulesen.

Variablenname	Beschreibung
bio1	Jahresdurchschnittstemperatur
bio2	mittlere Tagesspannweite (Mittel der monatlichen Maximaltemperatur minus der monatlichen Minimaltemperatur)
bio3	Isothermie (gleichbleibende Temperaturverteilung)
bio4	Saisonabhängige Temperatur (Standardabweichung $*100$)
bio5	Maximale Temperatur im wärmsten Monat
bio6	Minimale Temperatur im kältesten Monat
bio7	Jahrestemperaturbereich (bio6−bio5)
bio8	Mittlere Temperatur im feuchtesten Quartal
bio9	Mittlere Temperatur im trockensten Quartal
bio10	Mittlere Temperatur im wärmsten Quartal
bio11	Mittlere Temperatur im kältesten Quartal
bio12	jährlicher Niederschlag
bio13	Niederschlag im feuchtesten Monat
bio14	Niederschlag im trockensten Monat
bio15	Saisonabhängiger Niederschlag (Variationskoeffizient)
bio16	Niederschlag im feuchtesten Quartal
bio17	Niederschlag im trockensten Quartal
bio18	Niederschlag im wärmsten Quartal
bio19	Niederschlag im kältesten Quartal

Tabelle 1: Überblick über die Variablen aus dem WorldClim-Datensatz.

Der andere Teil der Einflussvariablen, bestehend aus Landbedeckungsvariablen, stammt aus dem CORINE Landcover–Project CLC2000, das Teil des Programms CORINE (Coordination of Information on the Environment) der Europäischen Union ist. Ziel des Projektes ist es, einheitliche Daten der Bodenbedeckung zur Verfügung zu stellen, die zum räumlichen und zeitlichen Vergleich verwendet werden können. Die Sammlung der Daten basiert auf Satellitenbildern im Maßstab 1 : 100.000, die sowohl visuell als auch durch automatisierte Verfahren analysiert werden. Der erste derartige Datensatz bezieht sich auf das Jahr 1990. Hierbei wurden europaweit 44 verschiedene Landklassen festgestellt, 21 davon mit Relevanz für Bayern. Gemessen wird der Anteil der Landklasse an einem Quadranten. Der Datensatzt CLC2000 stellt eine Aktualisierung dieser Daten zum Jahr 2000 dar. Seit Februar 2006 steht auch der Datensatz CLC2006 zur Verfügung. Auf europäischer Ebene wird das Projekt vom European Topic Centre for

Terrestrial Environment (ETC–TE) koordiniert, den Auftrag des deutschen Teilprojekts, der flächendeckenden Kartierung Deutschlands, gab das Umweltbundesamt (UBA). Dieses Projekt wird vom Deutschen Fernerkundungsdatenzentrum (DFD) des Deutschen Zentrums für Luft- und Raumfahrt (DLR) geleitet. Weitere Informationen zu dem Projekt sind nachzulesen bei Deutsches Zentrum für Luft-und Raumfahrt e.V. (2005). Die Variablen der beiden Datensätze WorldClim und CLC2000 sowie deren Beschreibung sind in Tabelle 1 und Tabelle 2 aufgelistet.
Außerdem standen die Koordinaten der untersuchten Quadranten im Gauß–Krüger-System (`X` und `Y`), die Höhe über Normalnull (`GewHoehe`) und die Anzahl der Exkursionen in einer Zelle (`Naechte`) zur Verfügung.

2.3 Bearbeitung des Datensatzes

Um die Analyse durchzuführen, musste der Datensatz in geeigneter Weise aufbereitet werden. Die Variablen `Gletscher`, `Deponien` und `Verkehr` wurden aufgrund keiner oder zu weniger Ausprägungen entfernt. Außerdem wurden alle Beobachtungen entfernt, deren Ausprägung der Variable `Naechte` gleich Null war. Im Datensatz wurde diese Variable nur um eins erhöht, wenn bei einer Exkursion auch Arten gefunden wurden. Daher ist nicht eindeutig festzustellen, ob in einem solchen Quadranten keine Exkursionen stattfanden oder ob trotz Exkursionen keine Arten beobachtet wurden. Auf diese Weise gehen schließlich 1.760 der 1.927 Quadranten des ursprünglichen Datensatzes in die Analyse ein.
Die Variable `GewHoehe` wurde mit Hilfe der Formel

$$\texttt{GewHoehe\_neu} = \frac{(\texttt{GewHoehe} - \min(\texttt{GewHoehe}))}{\max(\texttt{GewHoehe})}$$

standardisiert. Diese standardisierte Höhe wird als Variationskoeffizient in der Berechnung der Nonstationarität verwendet. Da die Variablen `SAWasser`, `Abbauflaechen`, `Felsen`, `HeidenMoore`, `Industrie`, `Laubwald`, `Moore`, `Obst`, `Sumpf`, `WaldrandGeb`, `WasserFl`,

Variablenname	Beschreibung
`Abbauflaechen`	Anteil der jeweiligen Landbedeckung im Quadranten
`Acker`	
`Deponien`	
`Felsen`	
`Gletscher`	
`HeidenMoore`	
`Industrie`	
`Komplex`	
`Laubwald`	
`Mischwald`	
`Moore`	
`Nadelwald`	
`Obst`	
`Stadt`	
`Sumpf`	
`Verkehr`	
`WaldrandGeb`	
`WasserFl`	
`WasserSteh`	
`Weinbau`	
`Wiesen`	
`SAWald`	Zusammenfassung der Variablen `Laubwald`, `Mischwald`, `Nadelwald` und `WaldrandGeb`
`SAWasser`	Zusammenfassung der Variablen `WasserFl`, `WasserSteh` und `Sumpf`

Tabelle 2: Überblick über die Variablen aus dem Corine-Datensatz.

`WasserSteh` und `Weinbau` nur wenige Ausprägungen aufweisen, wurden sie kategorisiert mit den Ausprägungen $= 0$ und > 0, d.h. es gibt nur die Information, ob die entsprechende Landbedeckung im betroffenen Quadranten vorhanden ist oder nicht.
Die Variable `Stadt` wurde in drei Gruppen eingeteilt, die erste mit einem Stadtanteil von 0% im untersuchten Quadranten, die zweite mit einem Anteil bis zu 10% und die dritte mit allen Quadranten, die einen höheren Stadtanteil als 10% haben.

2.4 Deskriptive Analyse

2.4.1 Zielvariable: Artenzahl

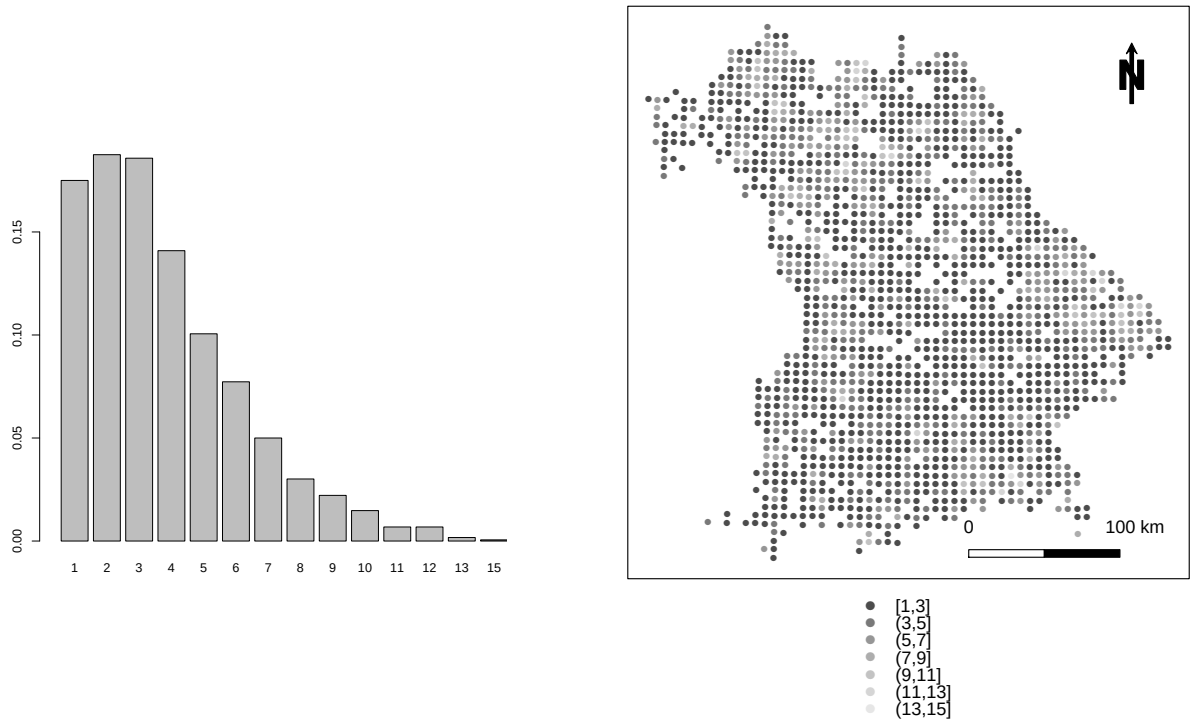


Abbildung 2: Verteilung der Variable `Arten` und Artendichte in Bayern.

Die maximale Anzahl an Fledermausarten, die in einem Quadranten beobachtet wurde, beträgt 15, die minimale Anzahl liegt bei einer beobachteten Art. Hierbei muss beachtet werden, dass Quadranten, in denen keine Fledermausart beobachtet wurde, im Vorfeld gelöscht wurden (vgl. Kapitel 2.3). Der Median liegt bei drei Arten im Quadranten. Für den Großteil der Daten liegt die Artenzahl zwischen zwei und fünf. Eine Übersicht über die Artenzahlen von Fledermäusen in Bayern gibt Abbildung 2. Insgesamt handelt es sich um eine unsymmetrische, deutlich linkssteile Verteilung. Es lässt sich erkennen, dass im Osten Bayerns im Gebiet des Bayerischen Wald tendenziell mehr Fledermausarten in einem Quadranten gefunden wurden. Auch im Nordwesten sowie im Süden in der Gegend des Chiemsees scheint es vermehrt Quadranten zu geben, in denen relativ viele

Fledermausarten beobachtet wurden. Außerdem ist gut erkennbar, dass die Zahl der gefundenen Arten in den meisten Quadranten zwischen eins und drei liegt.

2.4.2 Kovariablen

Einige der Kovariablen sollen an dieser Stelle vorgestellt werden. Nähere Information zu den hier nicht erwähnten Variablen finden sich im Anhang.
In Abbildung 3 sieht man die Dichteschätzer der Variablen `bio1` (Jahresdurchschnittstemperatur), `bio2` (mittlere Tagesspannweite) und `bio4` (Saisonabhängige Temperatur). Auf der x–Achse der Variablen `bio2` und `bio4` ist hierbei die Temperatur in $°C \cdot 10$ abgetragen. $-0{,}2°C$ ist die Temperatur des Quadranten mit der niedrigsten Jahresdurchschnittstemperatur und etwa $10°C$ die des Quadranten mit der höchsten Temperatur. Im Mittel beträgt die Jahresdurchschnittstemperatur in Bayern ca. $7{,}8°C$. Die Verteilung ist deutlich rechtssteil. Ein Großteil der Daten liegt zwischen $7{,}5°C$ und $8{,}3°C$. Im Datensatz befinden sich außerdem einige Variablen, die auf verschiedene Weise die Schwankungen der Temperatur messen. Eine dieser Variablen ist `bio2`, die mittlere Tagesspannweite. Die Spannweite im Datensatz liegt etwa zwischen $7{,}0°C$ und $10{,}3°C$. Ein anderes Beispiel für ein Schwankungsmaß der Temperatur ist die Variable `bio4` (Saisonabhängige Temperatur), die mit Hilfe der Standardabweichung berechnet wird. Die Saisonabhängige Temperatur schwankt zwischen $5{,}7°C$ und $7{,}2°C$, das arithmetische Mittel liegt bei etwa $6{,}8°C$. Es gibt nur vereinzelt Quadranten, in denen ein kleinerer Wert als $6{,}35°C$ gemessen wurde, so dass in der späteren Analyse der Effekte hier eventuell keine Aussagen getroffen werden können.

Neben Variablen, die die Temperatur beschreiben, gibt es im Datensatz auch Variablen, die Informationen über den Niederschlag in Bayern enthalten. Zwei dieser Variablen sollen an dieser Stelle beschrieben werden. Ihre Verteilung ist in Abbildung 4 zu sehen. Der jährliche Niederschlag wird durch die Variable `bio12` beschrieben und nimmt in den

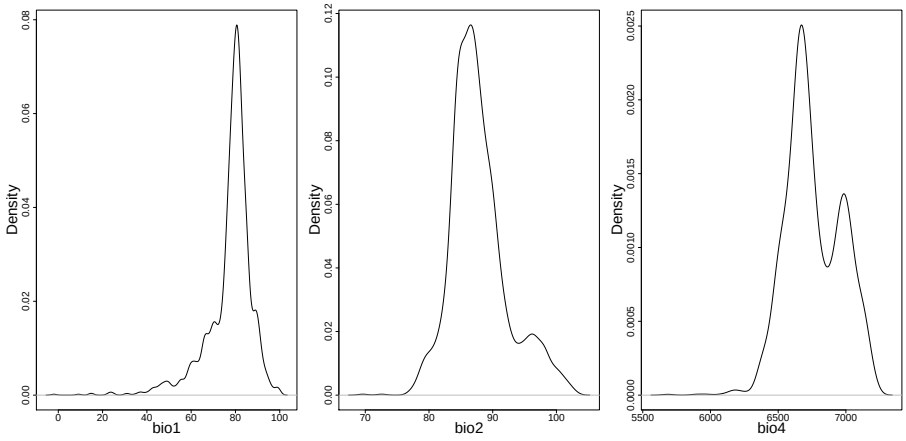


Abbildung 3: Verteilung der Variablen `bio1` (Jahresdurchschnittstemperatur), `bio2` (mittlere Tagesspannweite) und `bio4` (Saisonabhängige Temperatur).

vorliegenden Daten eine linkssteile Verteilung an. Im Quadranten mit dem niedrigsten jährlichen Niederschlag beträgt dieser 590,4 mm, im Quadranten mit dem höchsten Niederschlag 1.464 mm. Der Durchschnitt der jährlichen Niederschläge in Bayern liegt bei 820,4 mm pro Jahr. Betrachtet man die Variable `bio13` (Niederschlag im feuchtesten Monat), so kann man Werte von 70,8 mm bis 170,7 mm beobachten. Auch hierbei handelt es sich um eine deutlich linkssteile Verteilung.

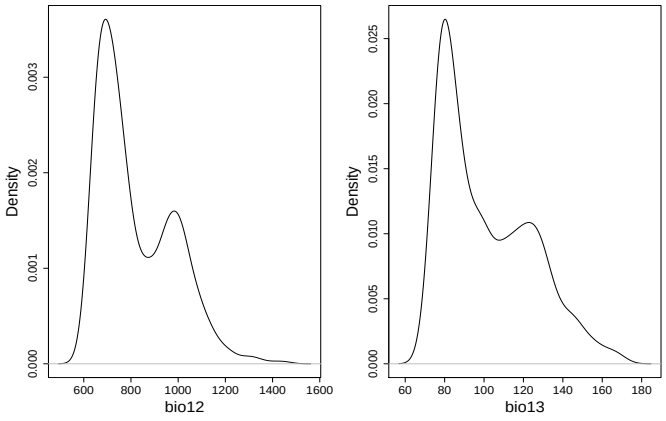


Abbildung 4: Verteilung der Variablen `bio12` (jährlicher Niederschlag) und `bio13` (Niederschlag im feuchtesten Monat).

Im Folgenden wird die Verteilung einiger Landbedeckungsvariablen analysiert. Die Auswahl fiel auf Variablen, die von Meschede und Rudolph (2004) als bedeutsame Landschaftskategorien bezeichnet wurden. Informationen zur Verteilung der anderen Bedeckungsvariablen finden sich im Anhang.

Schon in vielen Studien über Fledermäuse wurde die Landbedeckung „Wald“ als eine einflussreiche Variable erkannt, so z.B. von Mehr *et al.* (2010). Wie schon erwähnt, ist eine große Fläche Bayerns bewaldet. Die Variable `SAWald` fasst die drei Waldtypen Mischwald, Nadelwald und Laubwald und die Waldrandgebiete zusammen. Die Verteilung dieser fünf Variablen ist in Abbildung 5 gezeigt.

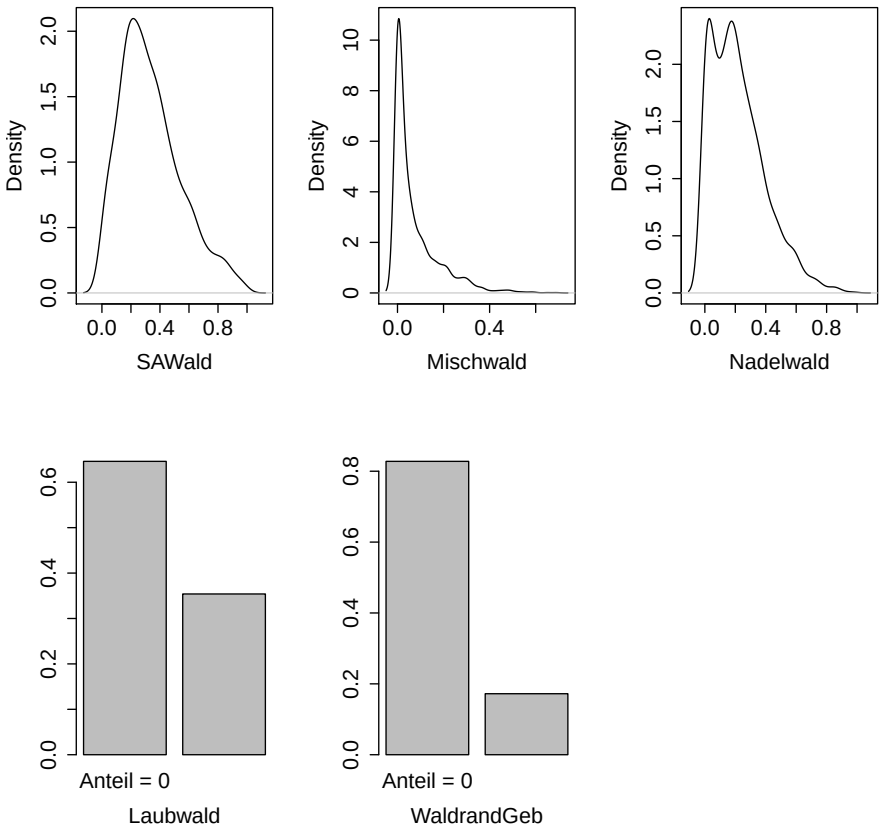


Abbildung 5: Verteilung der Variablen `SAWald`, `Mischwald`, `Nadelwald` und `Laubwald` und `WaldrandGeb`.

Durchschnittlich beträgt der Waldanteil in einem Quadranten 33,65%. Für die Mehrzahl der Quadranten liegt der Waldanteil zwischen 18% und 46%. Insgesamt handelt es sich um eine linkssteile Verteilung, bei der sich der Anteil der Bedeckung zwischen 0% und 100% befindet. Mit einem Mittelwert von 21,88% ist der Nadelwald stärker vertreten als der Mischwald mit 7,36%. In etwa einem Drittel der Quadranten gibt es Laubwald, jedoch handelt es sich hier meist um sehr geringe Anteile. Waldrandgebiet gibt es in etwa 17% der Quadranten.

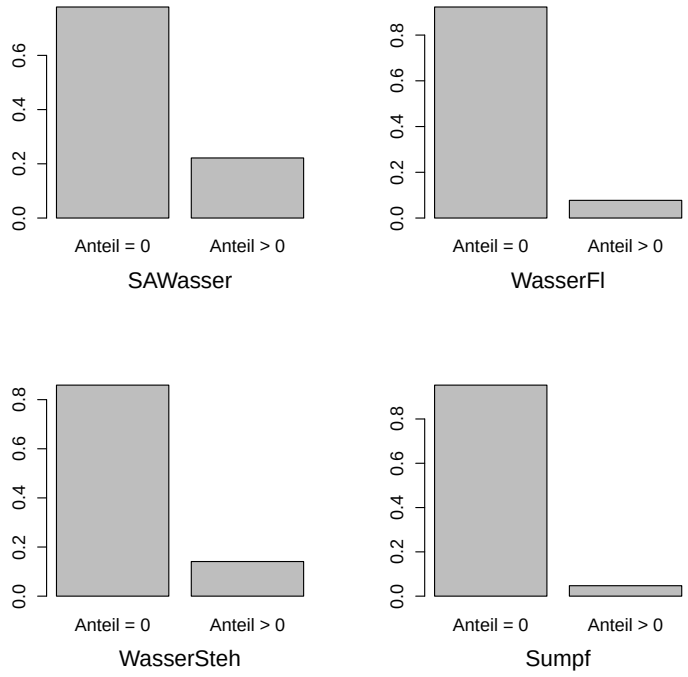


Abbildung 6: Verteilung der Variablen (SAWasser), `WasserFl`, `WasserSteh` und `Sumpf`.

Die Variablen, die den Wasseranteil in den Quadranten beschreiben, liegen nach der Datenaufbereitung alle in kategorisierter Form vor. Sie sind in Abbildung 6 dargestellt. Die Variable `SAWasser` fasst die Gewässer und somit die Variablen für stehendes Gewässer (`WasserSteh`), fließendes Gewässer (`WasserFl`) und Sümpfe (`Sumpf`) zusammen. In etwa

22% der Quadranten ist Gewässer vorhanden. Hierbei ist stehendes Gewässer, das in ca. 14% aller Quadranten vorkommt, am häufigsten vertreten, gefolgt von fließendem Gewässer und Sümpfen.

Der Anteil von `Acker` in den Quadranten reicht von 0% bis 98%. Das arithmetische Mittel der linkssteilen Verteilung liegt bei 30%. Der Großteil der Beobachtungspunkte weist einen Ackeranteil zwischen 5% und 50% auf. Der durchschnittliche Anteil von der ebenfalls stark linkssteil verteilten Variable `Wiesen` ist mit 17,23% im Quadranten deutlich geringer. Die meisten Daten liegen hier zwischen 4% und 23%.

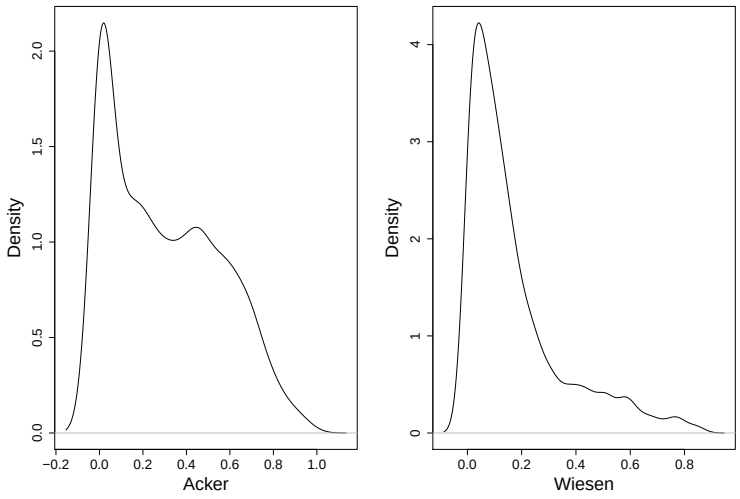


Abbildung 7: Verteilung der Variablen `Acker` und `Wiesen`.

Als letztes soll noch die Verteilung der Höhe über Normalnull analysiert werden. Bei dieser Größe handelt es sich erfahrungsgemäß um eine einflussreiche geographische Komponente. Die Höhe der Quadranten im Datensatz reicht von 151,9 m bis 1.925 m. Die höchsten Gebiete liegen im Bereich der Alpen, die niedrigsten in Talauen. Es handelt sich um eine linkssteile Verteilung mit einem Mittelwert von 492,8 m.

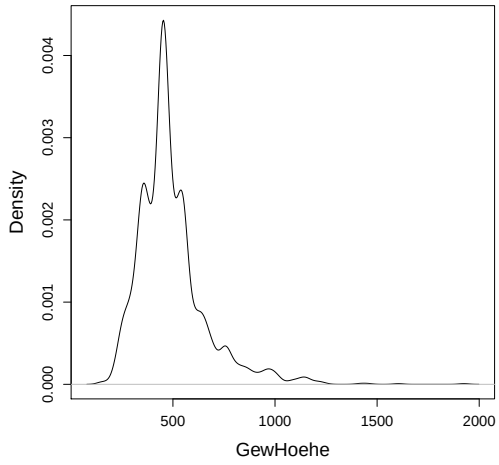


Abbildung 8: Verteilung der Variable `GewHoehe` (Höhe über dem Meeresspiegel).

2.5 Besonderheiten bei Modellen über die Artenverteilung

Ziel dieser Arbeit ist die Erstellung eines Habitatmodells, das die Artenzahl von Fledermäusen erklärt. Bei der Modellanpassung der vorliegenden Daten gibt es einige Punkte, die beachtet werden sollten. Zum einen beinhaltet der Datensatz eine große Menge an Kovariablen. Ein Hauptziel der Anpassung ist es also, herauszufinden, welche Kovariablen von Bedeutung sind, und auf diese Weise die Modellkomplexität so weit wie möglich zu reduzieren. Zum anderen sollte man darauf achten, dass es aufgrund des Raumes Abhängigkeiten zwischen den einzelnen Beobachtungen geben kann. Diese können sich positiv auswirken, was bedeuten würde, dass die Entdeckung einer Fledermausart in einem Quadranten die Wahrscheinlichkeit dafür, dass sich diese Art auch im Nachbarquadranten befindet, erhöht, obwohl dies durch die beobachteten Umweltvariablen nicht vorhergesagt werden würde. Dieses Phänomen bezeichnet man als positive räumliche Autokorrelation. Aber auch negative räumliche Autokorrelation, zum Beispiel aufgrund zwischenartlicher Konkurrenz, ist möglich (Legendre, 1993). Für die Modellierung wurde daher ein additives Modell mit der Poisson-verteilten Zielgröße „Anzahl der Fledermausarten“ aufgestellt. Die Modellanpassung erfolgte mit der so genannte Methode „Spatial

Boosting“, die im Folgenden erläutert werden soll. Die genauen Details sind nachzulesen in Hothorn *et al.* (2010b).

3 Methoden

3.1 Generalisiertes additives Modell

Bei den vorliegenden Daten handelt es sich um eine Zählvariable als Response, daher gilt die Annahme, dass $Y_i|\mathbf{x}_i \sim Po(\lambda_i)$, wobei der Parameter λ_i zugleich dem Erwartungswert und der Varianz entspricht. Im generalisierten additiven Modell geht man davon aus, dass sich der Prädiktor η_i aus glatten eindimensionalen Funktionen der einzelnen Kovariablen zusammensetzt:

$$\eta_i = f(\mathbf{x}_i, s_i) = \sum_j f_{(j)}(x_{ij}, s_i)$$

Über die Exponentialfunktion wird der Prädiktor mit der erwarteten Artenzahl λ_i verknüpft:

$$\lambda_i = \mathbb{E}(\text{Artenzahl}_i|\mathbf{x}_i, s_i) = \exp(f(\mathbf{x}_i, s_i)) \tag{1}$$

Das bedeutet, dass die mittlere erwartete Artenzahl an einem Punkt s (gegeben durch die Koordinaten im Gauß–Krüger–System), abhängig von den Umweltvariablen $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})$, dem Wert der Exponentialfunktion ausgewertet an der Stelle der Regressionsgleichung entspricht.

Mehr *et al.* (2010) stellen in ihrer Arbeit fest, dass zum Beispiel der positive Einfluss der Variablen `Stadt` und `Verkehr` verschwindet, sobald man die Anzahl der Exkursionen im Quadranten i ($\#\text{Exkursionen}_i$) mit einbezieht. Dass die Artenzahl in der Stadt höher ist, liegt also nicht daran, dass es hier tatsächlich mehr Arten gibt, sondern wird vor allem dadurch erklärt, dass dort zum einen öfter nach Fledermäusen gesucht wurde und zum anderen die Tiere dort auch leichter zu finden sind als beispielsweise im Wald.

Deshalb wird die Anzahl von Exkursionen als so genannter Offset in den Prädiktor aufgenommen, und ihr Effekt damit auf 1 gezwungen. Das heißt, dass die mittlere erwartete Artenzahl in einem Feld mit #Exkursionen $= k, k > 1$ sich bei gleicher Ausprägung aller übrigen Kovariablen multiplikativ um den Faktor k von einem Feld mit #Exkursionen $=$ 1 unterscheidet. Die erwartete Artenzahl λ_i ist dann $\lambda_i = \#\text{Exkursionen}_i \cdot \exp(f(\mathbf{x}_i, s_i))$ und es ergibt sich die strukturelle Komponente

$$\mathbb{E}(\text{Anzahl}_i|\mathbf{x}_i, s_i) = \lambda_i = \#\text{Exkursionen}_i \cdot \exp(f(\mathbf{x}_i, s_i)) =$$

$$= \exp(\underbrace{\log(\#\text{Exkursionen}_i)}_{\text{Offset}} + f(\mathbf{x}_i, s_i)).$$

3.2 Spatial Boosting

Bei hochdimensionalen Datensätzen sind übliche Schätzverfahren, wie z.B. penalisierte Schätzung nicht mehr anwendbar. Es kommt zu numerischen Rechenproblemen. Boosting ist ein möglicher Algorithmus zur Schätzung hochdimensionaler Regressionsmodelle für additive Prädiktoren. Das iterative Anpassen einzelner schwacher Schätzer führt zu einem insgesamt numerisch guten Schätzergebnis und überzeugt durch seine effektive Variablenselektion. Beim Spatial Boosting werden die Kovariablen in eine globale und eine lokale Komponente aufgeteilt. Die globale Komponente beachtet hierbei ausschließlich die Umweltvariablen sowie mögliche lineare oder nicht–lineare Effekte und Interaktionsterme. Ein rein globales Modell würde annehmen, dass die Effekte der Umweltvariablen fest und universal sind. Bei Auftreten von Nonstationarität variieren diese Effekte jedoch mit dem Raum. Die lokale Komponente beschreibt daher die räumliche Autokorrelation als Funktion $f_s(s)$ nur abhängig vom Raum. Die Nonstationarität wird als Funktion $f_{ns}(\mathbf{x}, s)$ in Abhängigkeit vom Raum und den Umweltvariablen modelliert. Durch die lokale Komponente erhält man eine Schätzung der unbeobachteten Heterogenität, die durch räumliche Autokorrelation oder nonstationäre Effekte verursacht wird. Dies ist deshalb von Bedeutung, da man davon ausgehen muss, nicht alle tatsächlichen Einflussvariablen erfasst zu haben. Die Annahme der Unabhängigkeit von $Y_i|\mathbf{x}_i$, $i = 1, \ldots, n$

kann aber nur getroffen werden, wenn alle Kovariablen gegeben sind. Deswegen werden die restlichen nicht erfassten Kovariablen sozusagen zu einem räumlichen Effekt der unbeobachteten Heterogenität zusammengefasst. Dies ist bei den meisten der bisher verwendeten Verfahren nicht der Fall.

Durch die Zerlegung hat die Regressionsfunktion, die in die Modellgleichung einfließt, folgende Form:

$$f(\mathbf{x}, s) = \underbrace{f_{env}(\mathbf{x})}_{global} + \underbrace{f_{ns}(\mathbf{x}, s) + f_s(s)}_{lokal} \tag{2}$$

Mit dieser Modellzerlegung wird auch die Variabilität in drei Komponenten zerlegt: die Variabilität erklärt durch die Umweltvariablen ($f_{env}(\mathbf{x})$), Variabilität, die von räumlicher Autokorrelation verursacht wird ($f_s(s)$) und die Variabiliät verursacht durch nonstationäre Umwelteffekte, d.h. zusätzlich räumlich variierende Effekte der Umweltvariablen ($f_{ns}(\mathbf{x}, s)$).

3.3 Beschreibung der Modellkomponenten

Da das Modell vom Raum abhängig ist, ist es nur auf das betreffende Untersuchungsgebiet anwendbar. f_{env} kann hingegen für Prognosen außerhalb Bayerns genutzt werden, da in diesem Term die räumlichen Effekte herausgerechnet wurden und somit die Prädiktionen nicht verzerrt werden. Der Term kann auf zwei Arten modelliert werden: Die einfachste Möglichkeit ist ein parametrischer Ansatz mit dem linearen Prädiktor $f_{env}(\mathbf{x}) = \mathbf{x}^T\beta$, wobei β der zu schätzende Vektor der Regressionskoeffizienten ist. Eine bisher genutzte Möglichkeit hier die Autokorrelation miteinzubeziehen, ist z.B. die Spezifizierung einer Arbeitskovarianz in Generalized Estimating Equations (GEE) (Dormann *et al.*, 2007). Eine andere Möglichkeit der Modellierung ist ein nonparametrischer Ansatz mit additiven glatten Funktionen, also $f_{env}(\mathbf{x}) = \sum_{j=1}^{p} f_j(x_j)$, wobei $\mathbf{x} = (x_1, \ldots, x_p)$. In jeder einzelnen Kovariable kann so ein möglicher nicht–linearer Effekt auf flexible Weise

geschätzt werden. Komplexere Modelle erlauben zusätzlich Interaktionen, wie z.B. Random Forests oder Boosted Regression Trees. $f_s(s)$ stellt eine glatte zweidimensionale Oberflächenfunktion dar, die die unbeobachtete Heterogenität, eingeführt durch lokale Einflüsse, modelliert. So werden räumliche Autokorrelationsmuster erkannt. $f_{ns}(\mathbf{x}, s)$ repräsentiert die räumliche Nicht–Stationarität.

3.4 Modellanpassung durch Spatial Boosting

Die Modellanpassung wird durch die Minimierung der negativen Log–Likelihood der zugrunde liegenden Verteilung durchgeführt. Die Artenzahl folgt einer $Po(\lambda_i)$ Poissonverteilung mit $\lambda_i = \mathbb{E}(y_i|x_i, s_i)$ und $\lambda_i(f) = \#\text{Exkursionen}_i \cdot \exp(f(\mathbf{x}_i, s_i))$. Damit ist die negative Log–Likelihood–Funktion

$$\hat{f} = \underset{f}{\operatorname{argmin}} \sum_{i=1}^{n} \rho(y_i, \lambda_i(f))$$

mit

$$\rho(y_i, \lambda_i(f)) = \lambda_i - y_i \log \lambda_i$$

als Beitrag einer Beobachtung zur Gesamt–Log-Likelihood.

Die Funktion $\hat{f}$, die die Verlustfunktion minimiert, wird mit einem Componentwise Functional Gradient Descent Boosting–Algorithmus geschätzt. Für Modelle der Form (2) können auch Methoden wie Markov–Chain–Monte–Carlo– (MCMC–) Algorithmen (Fahrmeir *et al.* (2004), Kneib *et al.* (2008)) oder penalisierte Schätzung von generalisierten additiven Modellen verwendet werden. Diese Methoden sind jedoch rechenaufwändig bzw. auf Daten mit einer geringen Zahl an Einflussvariablen oder einer kleinen bis mittleren Beobachtungszahl ausgelegt und es gibt keine effizienten Verfahren der Variablenselektion. Auf diese Weise würden unwichtige Parameter das finale Modell unnötig komplex machen. Die Modellinferenz hat hier aber vor allem die Selektion von informativen

Parametern zum Ziel. Sollte keine räumliche Autokorrelation vorliegen, dann sollte auch die Modellkomponente $f_s(s)$ nicht in das Modell aufgenommen werden, d.h. $f_s(s) \equiv 0$ und genauso $f_{env}(\mathbf{x}) \equiv 0$, falls keine der Umweltvariablen einen Einfluss hat. Hier ist man allerdings mehr an den Effekten der einzelnen Umweltvariablen, also an dem Ergebnis $f_j(x_j) \equiv 0$ interessiert, was bedeutet, dass die Variable x_j keinen Einfluss auf die Artenzahl von Fledermäusen hat. Der Idealfall wäre ein globales Modell, in das nur wenige Umweltkomponenten aufgenommen werden.

Componentwise Functional Gradient Descent Boosting–Algorithmus

Für den Componentwise Functional Gradient Descent Boosting–Algorithmus wird $\hat{f} \equiv 0$ als konstantes Modell initialisiert. Im nächsten Schritt werden die Residuen für das aktuelle Modell berechnet. Unter dem Residuum versteht man hier den negativen Gradienten u_i der Verlustfunktion ρ berechnet an jeder Beobachtung y_i.

$$u_i = -\frac{\partial}{\partial f}\rho(y_i, f)|_{f=\hat{f}^{[m-1]}(x_i)}, i = 1, \ldots, n$$

Nun wird diejenige Basisprozedur g_{j^*} ($f_j(x_j)$, f_{ns} oder f_s) ausgewählt, welche die Residuen am besten beschreibt, d.h. die Summe der quadrierten Differenz der Residuen und der Modellkomponente minimiert:

$$j^* = \underset{1 \leq j \leq p}{\operatorname{argmin}} \sum_{i=1}^{n} (u_i - \hat{g}_j(x_i))^2$$

Nur diese Komponente wird aktualisiert mit z.B. 10% der Prädiktionen (Schrittweite ν) und zum aktuellen Modellfit hinzugefügt.

$$\hat{f}_{j^*}^{[m]}(\cdot) = \hat{f}_{j^*}^{[m-1]}(\cdot) + \nu g_{j^*}^{[m]}(\cdot)$$

Für alle anderen Komponenten gilt:

$$\hat{f}_j^{[m]}(\cdot) = \hat{f}_j^{[m-1]}(\cdot), \forall j \neq j^*$$

Anschließend werden die Residuen wieder neu berechnet und die entsprechende Modellkomponente aktualisiert. Diese Schritte werden wiederholt, bis eine vorher festgelegte Anzahl von Iterationen durchgeführt wurde. Das finale Modell $\hat{f}$ setzt sich zusammen aus der Summe aller gefitteten Modelle der einzelnen Komponenten $\hat{f}_{env}$, $\hat{f}_{ns}$ und $\hat{f}_s$. Die mathematischen Details werden von Bühlmann und Hothorn (2007) und Kneib *et al.* (2007) beschrieben.

Basisprozedur

Die so genannte Basisprozedur, die auch als Baselearner bezeichnet wird, bestimmt, wie die Residuen gefittet werden. Die Wahl der Baselearner ist entscheidend, da sie bestimmen, in welcher Form die einzelnen Modellkomponenten in das finale Modell eingehen. Für f_{env} kommen lineare Modelle, Smoothing–Splines, univariate penalisierte Splines (P–Splines) oder Regressionsbäume in Frage. Wobei letztere Methode genau mit den Boosted Regression Trees übereinstimmt. Für f_s werden die Baselearner als bivariater Tensorprodukt P–Spline gewählt, was einer glatten zweidimensionalen Oberflächenfunktion entspricht. Für die nicht–stationäre Komponente f_{ns} bietet sich ein Produkt eines Tensorprodukt P–Splines mit einer Umweltvariable x_j an. Interaktionen können z.B. über lineare Terme von Produkten berücksichtigt werden oder, wenn man noch flexibler sein möchte, über zwei- oder dreidimensionale glatte Funktionen.

Wie bereits erwähnt, wurden Umweltvariablen mit nur wenigen Ausprägungen kategorisiert, so dass man nun zwei unterschiedliche Variablentypen vorliegen hat. Für die stetigen Variablen wurden als Baselearner penalisierte Regressionssplines (mit sechs Freiheitsgraden) verwendet und für die faktorisierten Variablen einfache lineare Modelle, die über Ridge–Regression (Parameter λ bestimmt durch sechs Freiheitsgrade) geschätzt wurden.

3.5 Modellwahl und Variablenselektion

Es gibt sechs verschiedene Grundmodelle, die alle möglichen Einflussszenarien beschreiben, indem sie verschiedene Restriktionen an die einzelnen Modellkomponenten anlegen (Tabelle 3).

Modell	$f_{env}(\mathbf{x})$	$f_{ns}(\mathbf{x}, s)$	$f_s(s)$
(Spatial)	$\equiv 0$	$\equiv 0$	
(Additive)	$\sum_{j=1}^{p} f_j(x_j)$	$\equiv 0$	$\equiv 0$
(Add/Spatial)	$\sum_{j=1}^{p} f_j(x_j)$	$\equiv 0$	
(Tree/Spatial)		$\equiv 0$	
(Add/Vary)	$\sum_{j=1}^{p} f_j(x_j)$		
(Tree/Vary)			

Tabelle 3: Modellrestriktionen

Das Modell (Spatial), das nur den lokalen Einfluss misst und alle anderen Komponenten auf Null setzt, wäre das beste Modell, wenn keine der erhobenen Umweltvariablen einen Einfluss auf den Response hat. Wenn dagegen nur diese Umweltvariablen Auswirkungen haben ohne räumliche Variation und dabei die einzelnen Variablen additiv und ohne Interaktionen auf den Response wirken, wäre das Modell (Additive) das Richtige. (Add/Spatial) modelliert einen additiven Effekt der Umweltvariablen sowie einen zusätzlichen räumlichen Effekt ohne Nonstationarität oder Interaktionen zu berücksichtigen. Mit Regressionsbäumen als Baselearner für f_{env} können Interaktionen besser modelliert werden, ansonsten ist das Modell (Tree/Spatial) gleich wie das vorherige. Am komplexesten sind die letzten beiden Modelle, die damit auch die größte Flexibilität bieten: (Add/Vary) modelliert wieder additive Effekte für f_{env} und erlaubt gleichzeitig räumliche Autokorrelation und Nicht–Stationarität. Dies ist auch bei (Tree/Vary)

der Fall, zusätzlich sind Interaktionen bei den Umweltvariablen erlaubt, was insgesamt heißt, dass überhaupt keine Restriktionen an die Modellkomponenten angelegt werden. Aus diesen sechs Grundmodellen wird für die vorliegenden Daten in Kapitel 4 das beste Modell ausgewählt.

Die eigentliche Modellwahl wird in zwei Schritten gemacht. Für jedes der sechs oben genannten Modelle wird die ideale Iterationszahl bestimmt. Diese ergibt sich als m_{stop} mit dem minimalen empirischen Risiko, berechnet mit Bootstrap– und Kreuzvalidierungsverfahren. Eine andere Möglichkeit wäre, m_{stop} durch das Informationskriterium nach Akaike (AIC), das korrigierte AIC oder das Bayesianische Informationskriterium (BIC) zu bestimmen. Da es sich aber um einen hochdimensionalen Datensatz handelt, ist die Berechnung über Bootstrap und Kreuzvalidierung am geeignetsten. Die Wahl des idealen Stoppkriteriums hat den Zweck, Overfitting zu vermeiden. Im zweiten Schritt wird mit der neu bestimmten optimalen Anzahl an Boosting–Schritten die Modellanpassung wiederholt. Die sechs Modelle werden untereinander anhand der negativen Log–Likelihood verglichen. Die beste Modellanpassung hat das Modell, das in wiederholten Bootstrapstichproben die kleinste negative Log–Likelihood hat.

Auch die Schrittweite ν muss festgelegt werden. Für bisherige Probleme schien die Wahl dieser Schrittweite von eher geringer Bedeutung zu sein, solange sie klein genug gewählt wurde, um den Effekt des aktuellen Fits zu dämpfen. Eine kleinere Schrittgröße bedeutet typischerweise eine größere Anzahl an Iterationsschritten und somit mehr Berechnungszeit, wobei sich die Prädiktionsgenauigkeit im Allgemeinen nicht verschlechtert. Aus diesem Grund genügt es meist, den Parameter ν „ausreichend klein“ zu wählen (Bühlmann und Hothorn, 2007). Bei bisherigen Problemen wurde die Schrittweite daher oft auf den Wert $\nu = 0{,}1$ festgelegt. In der Auswertung dieser Arbeit stellte sich jedoch heraus, dass ein weiteres Verringern der Schrittgröße die Ergebnisse für die vorliegenden Daten weiter verbessern kann (vgl. Kapitel 4).

Da immer nur eine Modellkomponente pro Iterationsschritt angepasst wird, führt eine kleine Anzahl an Iterationen zu einem sparsamen Modell. Somit ist diese Methode eine sehr gute Möglichkeit der Variablenselektion. Zusätzlich wird für das beste Modell eine Stability Selection, wie sie von Meinshausen und Bühlmann (2010) beschrieben wird, angewandt, um sicher zu stellen, dass tatsächlich nur einflussreiche Variablen und Komponenten aufgenommen werden und man keine Effekte interpretiert, die in Wirklichkeit gar nicht bestehen. Dazu wird die empirische Wahrscheinlichkeit berechnet, wie oft die Variable in Teildaten ausgewählt wird. Variablen, deren Wahrscheinlichkeit größer einem festgelegten Grenzwert sind, gelten als einflussreich. Die Wahrscheinlichkeit mindestens eine einflusslose Variable oder Komponente auszuwählen ist dabei kleiner einem festgesetzten Wert α. Auf diese Weise erhält man ein Modell, das so komplex wie nötig, aber so einfach wie möglich ist.

4 Auswertung

Alle Auswertungen wurden mit Hilfe der statistischen Software `R` durchgefürt (R Development Core Team, 2009). Die sechs bereits erläuterten Modelle (Tabelle 3) wurden für die drei verschiedene Schrittweiten $\nu = 0{,}1$, $\nu = 0{,}05$ und $\nu = 0{,}01$ aufgestellt. Wie bereits in Kapitel beschrieben, spielt die Wahl der Schrittgröße ν eine untergeordnete Rolle und basiert auf Erfahrungswerten. Eine Möglichkeit diesen Parameter ähnlich wie das Stopkriterium m_{stop} oder den Glättungsparameter λ zu berechnen ist bisher noch nicht gegeben. Während zum Beispiel bei einer Binomialverteilung eine Schrittweite von $\nu = 0{,}1$ ausreicht, ist bei der Poissonverteilung meist eine kleinere Schrittweite nötig.
Als Variationskoeffizient für die Modelle (Add/Vary) und (Tree/Vary) wurde die standardisierte Höhe über dem Meeresspiegel gewählt, da aus anderen Studien bekannt ist, dass dies eine einflussreiche geographische Variable ist.
Als erstes wurde für jedes Modell und jede Schrittweite mittels Kreuzvalidierung ein

individuelles Stopkriterium bestimmt.
In den Abildungen 9, 10 und 11 wird der Vergleich der sechs verschiedenen Modelle anhand der Out–of–Bootstrap–Verteilung der negativen Log–Likelihood für die drei verschiedenen Schrittgrößen gezeigt. Für die Schrittweiten $\nu = 0{,}1$ und $\nu = 0{,}05$ ist neben den Boxplots für alle sechs Modelle eine Vergrößerung des unteren Bereichs abgetragen, um die Modelle besser miteinander vergleichen zu können.
Kleine Werte der negativen Log–Likelihood stehen für eine bessere Modellanpassung. Die Buchstaben a, b, und c, die über der Grafik abgetragen sind, teilen Modelle der gleichen Modellgüte in eine Gruppe ein. Als Vergleichsmethode wurde ein multipler Vergleich nach Tukey verwendet, der untersucht, welche Modelle sich signifikant im Mittelwert der negativen Log–Likelihood unterscheiden.
Insgesamt lässt sich erkennen, dass für die Schrittweite $\nu = 0{,}01$ die negative Log-Likelihood die kleinsten Werte und somit die besten Modellanpassungen liefert. Für alle drei Schrittweiten lassen sich jedoch die gleichen Tendenzen feststellen. Die Modelle (Tree/Spatial) und (Tree/Vary) fallen für alle drei Schrittweiten am schlechtesten aus. Die übrigen vier Modelle liegen jeweils sehr dicht beieinander.

Vergleicht man die Modelle, die mit der Schrittweite $\nu = 0{,}1$ angepasst wurden (Abbildung 9), so fällt auf, dass das Modell (Spatial) am besten abschneidet. Dennoch wurde für diese Schrittweite auch das Modell (Add/Spatial) in der weiteren Analyse betrachtet, dessen Werte der negativen Log-Likelihood, wie auch die von Modell (Add/Vary) nur etwas größer sind als die des Modells (Spatial). Da das Modell (Add/Spatial) aber einfacher als das Modell (Add/Vary) aufgebaut und daher auch leichter zu interpretieren ist, wurde dieses Modell für die Analyse ausgewählt.

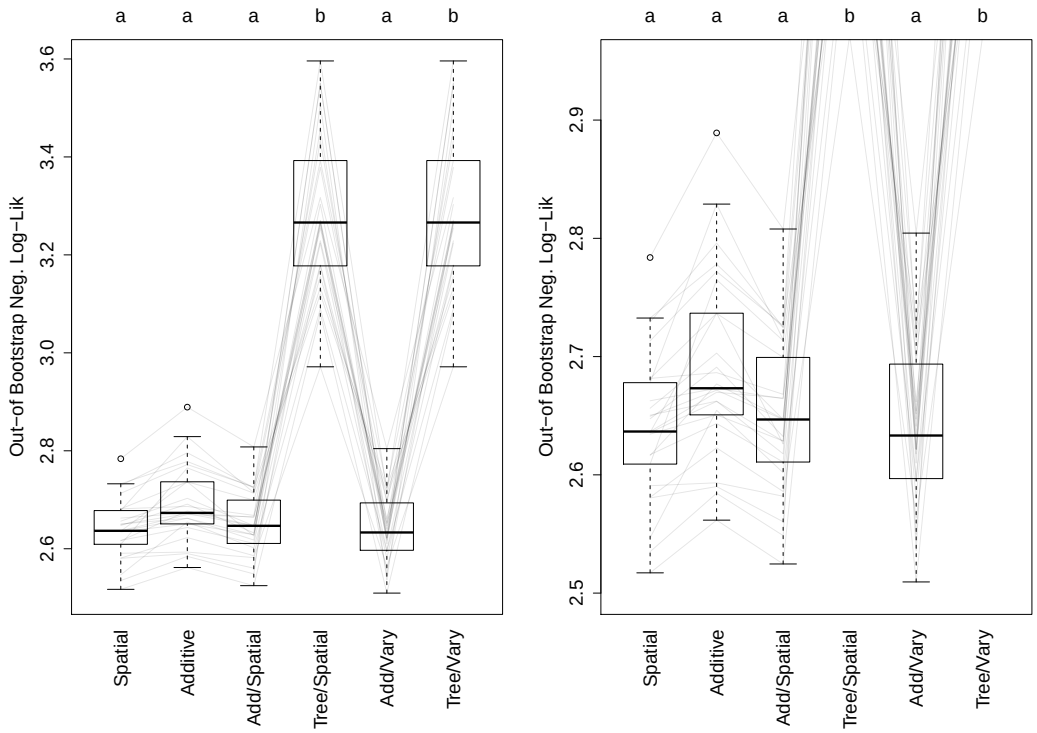


Abbildung 9: Modellvergleich anhand der negativen Log–Likelihood der Out-of-Bootstrap–Verteilung für die Schrittweite $\nu = 0{,}1$.

Auf der rechten Seite von Abbildung 10 kann man erkennen, dass sich die Werte der negativen Log-Likelihood für die Modelle (Spatial), (Add/Spatial) und (Add/Vary) kaum unterscheiden. Betrachtet man nur die Modelle (Spatial), (Additive), (Add/Spatial) und (Add/Vary), so fällt für die beiden bisher betrachteten Schrittgrößen auf, dass das rein globale Modell (Additive) jeweils am schlechtesten abschneidet und das Hinzunehmen der räumlichen Information das Modell verbessert. Das Aufnehmen des Variationkoeffizienten trägt allerdings nicht wesentlich zur Verbesserung der Modellgüte bei. Das Weglassen der globalen Komponente führt tendenziell zu einer Verbesserung des Modells.

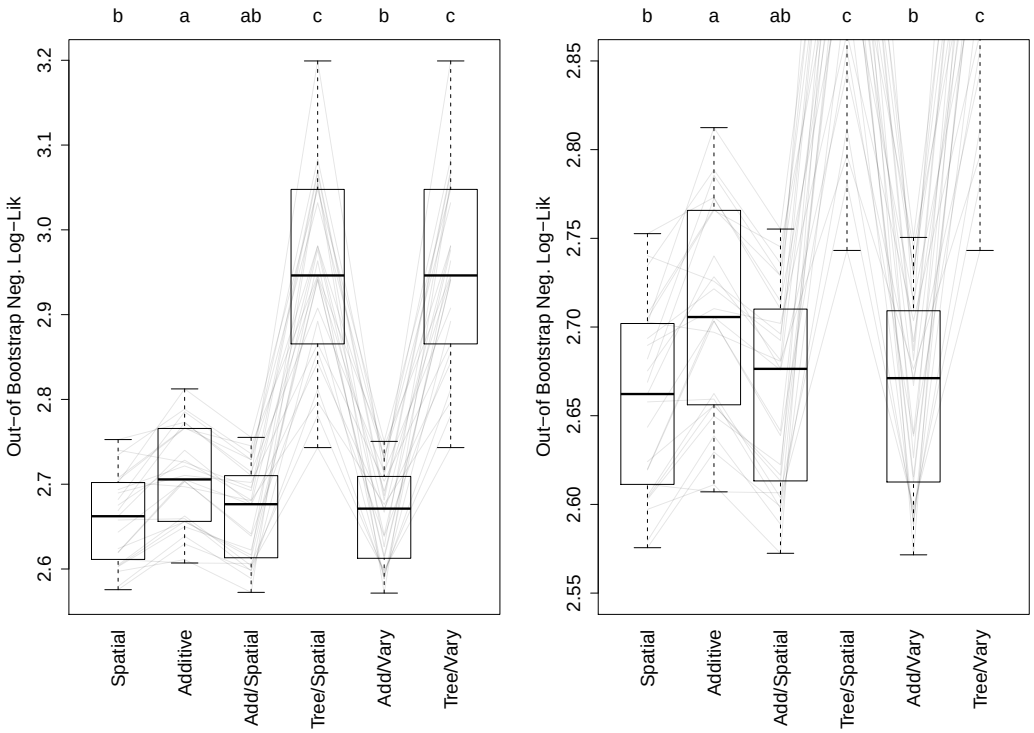


Abbildung 10: Modellvergleich anhand der negativen Log–Likelihood der Out–of–Bootstrap–Verteilung für die Schrittweite $\nu = 0{,}05$.

Für das Modell mit der Schrittweite $\nu = 0{,}01$ kann man erkennen, dass das Modell (Spatial), in das nur die räumliche Komponente eingeht, etwas besser ist als das Modell (Additive) nur mit den Kovariablen. Dies weist darauf hin, dass der Einfluss der räumlichen Autokorrelation größer ist als der der Umweltvariablen. Von Fledermäusen ist beispielsweise bekannt, dass es so genannte Quartierverbände gibt, d.h. man kann oft davon ausgehen, dass der Fund eines Quartiers auf das Vorhandensein weiterer Quartiere in der Umgebung hinweist (Meschede und Rudolph, 2004). Die Hinzunahme von glatten additiven Funktionen (Add/Spatial) oder das Einfließen der standardisierten Höhe über dem Meeresspiegel (Add/Vary) verbessern die Modellanpassung nur geringfügig. Die beiden Modelle, die Regressionsbäume als Baselearner verwenden, schneiden wie schon erwähnt am schlechtesten ab (Tree/Spatial, Tree/Vary), während die Modelle

(Add/Spatial) und (Add/Vary) am besten sind. Diese beiden Modelle sind sich sehr ähnlich, was darauf hinweist, dass es keine starken Interaktionseffekte in den Daten gibt. Wieder beziehen sich die weiteren Analysen auf Grund der besseren Interpretierbarkeit auf das Modell (Add/Spatial).

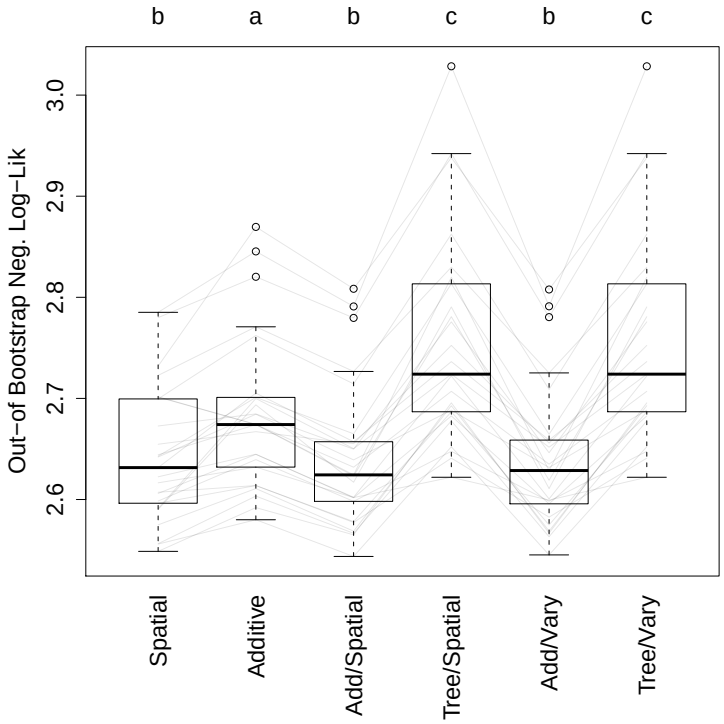


Abbildung 11: Modellvergleich anhand der negativen Log–Likelihood der Out–of–Bootstrap–Verteilung für die Schrittweite $\nu = 0{,}01$.

Die erklärte Variabilität des Modells (Add/Vary) ist in Abbildung 12 dargestellt. Diese Variabilität ist für alle drei Schrittweiten sehr ähnlich verteilt, weshalb hier nur die Boxplots der Variabilität für das Modell mit der Schrittweite $\nu = 0{,}01$ dargestellt ist.
Die größte Variabilität wird durch den Offset erklärt ($\log(\#\text{Exkursionen}_i)$), was bedeutet, dass die erwartete Artenzahl eines Quadranten in den Daten vor allem dadurch beeinflusst wird, wie oft dort nach Fledermäusen gesucht wurde. Die Variabilität, die

durch die räumliche Komponente und die Umweltvariablen erklärt wird, ist vergleichsweise gering.

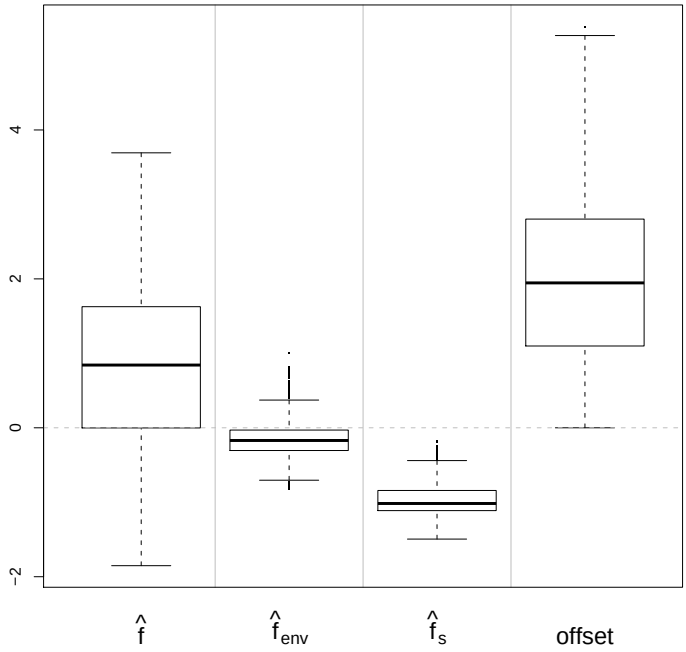


Abbildung 12: Aufspaltung der erklärten Variabilität für das Modell (Add/Spatial) mit der Schrittweite $\nu = 0{,}01$.

Es soll nun verglichen werden, welche Variablen jeweils für die verschiedenen Schrittweiten in das Modell aufgenommen wurden.
Auf das Modell (Add/Spatial) wurde für alle drei Schrittweiten die bereits in Kapitel 3.5 erläuterte Methode Stability Selection angewendet, um zu überprüfen, ob die ausgewählten Parameter tatsächlich einen Einfluss auf die Responsevariable haben. Der dabei verwendete Grenzwert beträgt 0,9.
Für die Schrittweite $\nu = 0{,}1$ wurde nur die räumliche Komponente $f_s(s)$ ausgewählt, was genau dem Modell (Spatial) entspricht. Es wird also angenommen, dass keine der

Umweltvariablen einen Effekt auf den Response hat, d.h. $f_{env}(\mathbf{x}) \equiv 0$.
Mit einer Schrittweite von $\nu = 0{,}05$ wurden neben der räumlichen Komponente $f_s(s)$ noch drei der Umweltvariablen für das finale Modell (Add/Spatial) ausgewählt. Der geschätzte Effekt $f_{partial}$ dieser drei Variablen ist in Abbildung 13 dargestellt. Interpretiert werden die Effekte folgendermaßen: Bei Festhalten aller anderen Variablen ändert sich die erwartete mittlere Artenzahl multiplikativ um den Faktor $\exp(f_{partial})$. Ein geschätzter Effekt größer Null bedeutet also einen positiven, ein geschätzter Effekt kleiner Null hingegen einen negativen Einfluss.

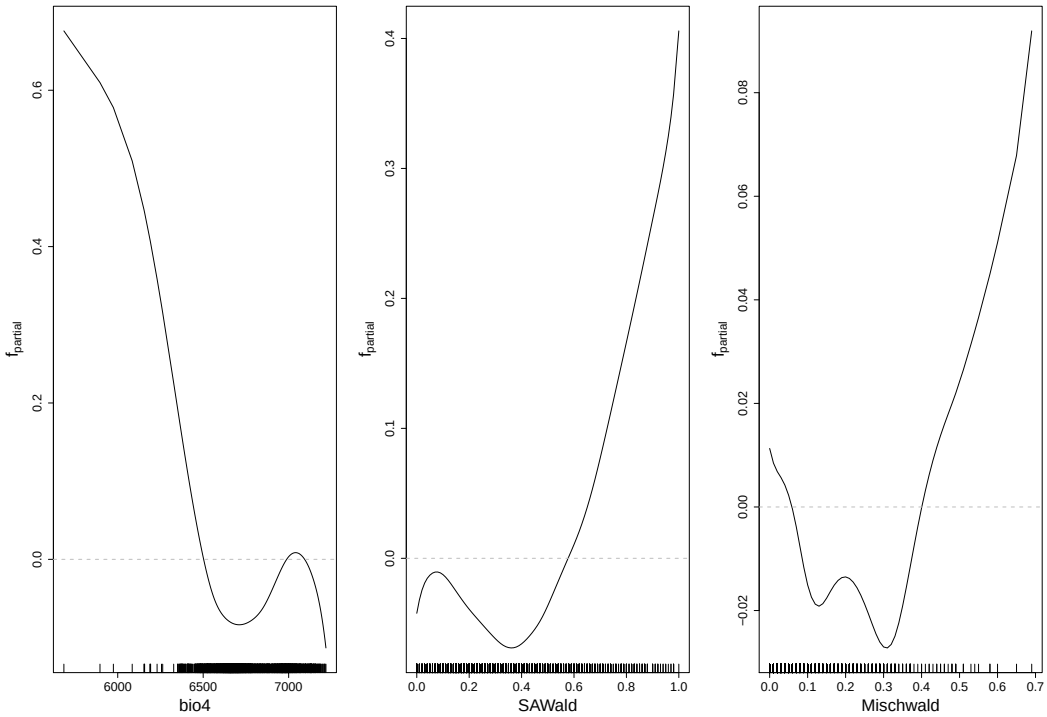


Abbildung 13: Schätzung der Effekte des Modells (Add/Spatial) für die Schrittweite $\nu = 0{,}05$.

Die Variable `bio4` (Saisonabhängige Temperatur) beschreibt Temperaturschwankungen. Wie bereits in Kapitel 2.4.2 beschrieben, liegen bis zu einer Schwankung von etwa $6{,}35°C$

nur wenige Daten vor, so dass in diesem Bereich nur sehr ungenaue Aussagen getroffen werden können. Temperaturschwankungen bis zu einem Wert von etwa 6,5°C wirken sich positiv, alle größeren Schwankungen negativ auf die Zielvariable aus. Das Phänomen, dass Artenzahlen bei größeren Temperaturschwankungen sinken, wird häufig beobachtet. Dies kommt daher, dass Temperaturen physiologisch wirksam und häufig ein strenger ökologischer Filter sind.
Außerdem wurden die Variablen `SAWald` und `Mischwald` ausgewählt. Die Variable `SAWald` hat ab einem Bedeckungsanteil von 58% im Quadranten einen positiven Einfluss auf die Artenzahl der Fledermäuse, die erwartete Zahl der Arten steigt an, je höher der Waldanteil einer Zelle über diesem Wert liegt. In Quadranten, deren Waldanteil geringer ist, sinkt der mittlere Erwartungswert der Artenzahl. Generell kann man sagen, dass die Anzahl der Arten steigt, je mehr Nischen es gibt. Der Anteil des Waldes könnte möglicherweise ein Indikator dafür sein. Von Fledermäusen ist bekannt, dass Bäume im Sommer die wichtigsten natürlichen Verstecke für die Tiere darstellen. Sie nutzen Baumhöhlen, Baumspalten und abstehende Rinde als Lebensraum. Außerdem hängen im bayerischen Staatswald 150.000 Nistkästen, der Großteil davon in Nadelwäldern. Alle Fledermäuse in Bayern nutzen den Lebensraum Wald und gehölzreiche Habitate, allerdings in unterschiedlich intensiver Weise (Meschede und Rudolph, 2004). Für die Variable `Mischwald` ergibt sich ab einem Bedeckungsanteil von 40% ein positiver Effekt auf die erwartete Artenzahl.

Für das Modell mit der Schrittweite $\nu = 0{,}01$ wurden schließlich die neun in den Abbildungen 14, 15 und 16 dargestellten Umweltvariablen und die räumliche Komponente $f_s(s)$ ausgewählt.

Für die Variable `bio2` (mittlere Tagesspannweite) liegen im Bereich bis etwa 7,7°C nur wenige Daten vor, so dass hier keine Aussage getroffen werden kann. Ab einer mittleren Tagesspannweite von 7,7°C gibt es einen positiven Einfluss auf die Artenzahl, der jedoch

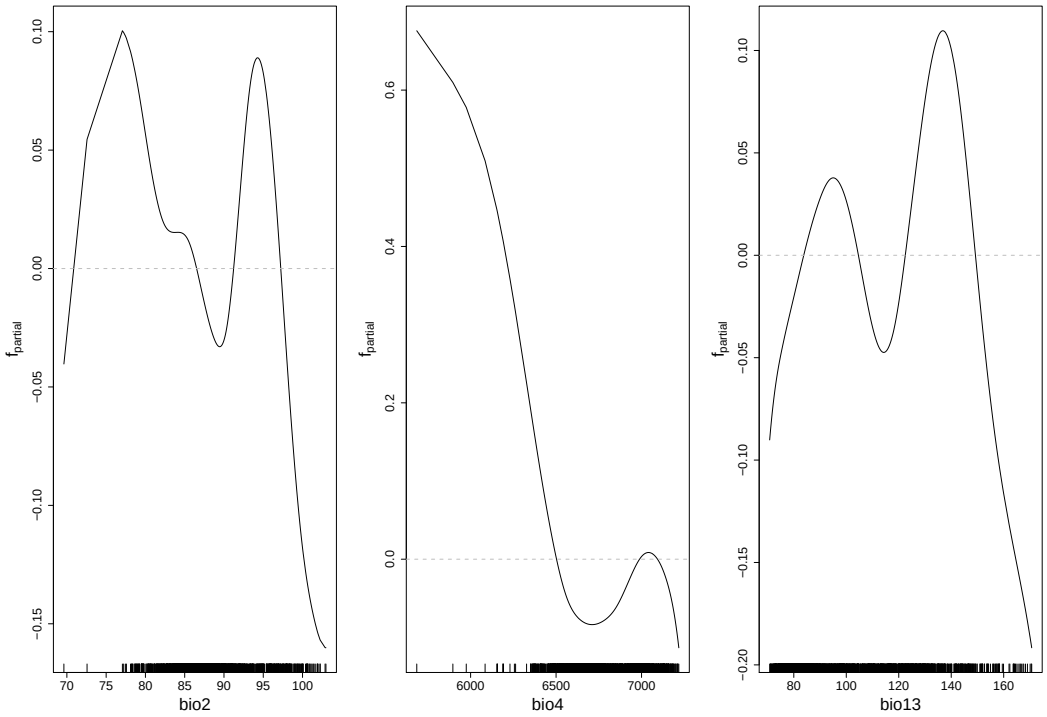


Abbildung 14: Schätzung der Effekte der Variablen `bio2`, `bio4` und `bio13`.

mit zunehmener Tagesspannweite geringer wird, so dass sich zwischen einer Spannweite von etwa 8,7°C und 9,1°C ein negativer Einfluss ergibt. Zwischen 9,1°C und etwa 9,7°C gibt es wieder einen positiven Effekt, überschreitet die Tagesspannweite jedoch den Wert von 9,7°C, so hat dies eine negative Wirkung auf die Zahl der Fledermausarten. Da es sich hier wie bei `bio4` um ein Schwankungsmaß der Temperatur handelt, ist die Abnahme der erwarteten Fledermausarten für hohe Werte durchaus nachvollziehbar. Die Interpretation der Variable `bio4` entspricht der des Modells mit $\nu = 0{,}05$.

Für eine Niederschlagsmenge von etwa 123 mm bis 150 mm ist der Effekt der Variable `bio13` (Niederschlag im feuchtesten Monat) positiv, für Niederschlagsmengen, die außerhalb dieses Intervalls liegen, ergibt sich ein negativer Effekt, wobei für hohe Niederschlagsmengen nur wenige Daten vorliegen, so dass diesbezügliche Angaben keine zuverlässige Aussagekraft haben. Im Allgemeinen kann die Niederschlagsmenge ein Maß

für die Produktivität in einem System sein, was auch eine erhöhte Artenanzahl zur Folge haben kann. Gebiete mit hoher Niederschlagsmenge sind jedoch oft montane und alpine Regionen, in denen die Temperaturbedingungen die Produktivität hemmen können. Dies wiederum könnte sich negativ auf die erwartete Artenzahl auswirken.

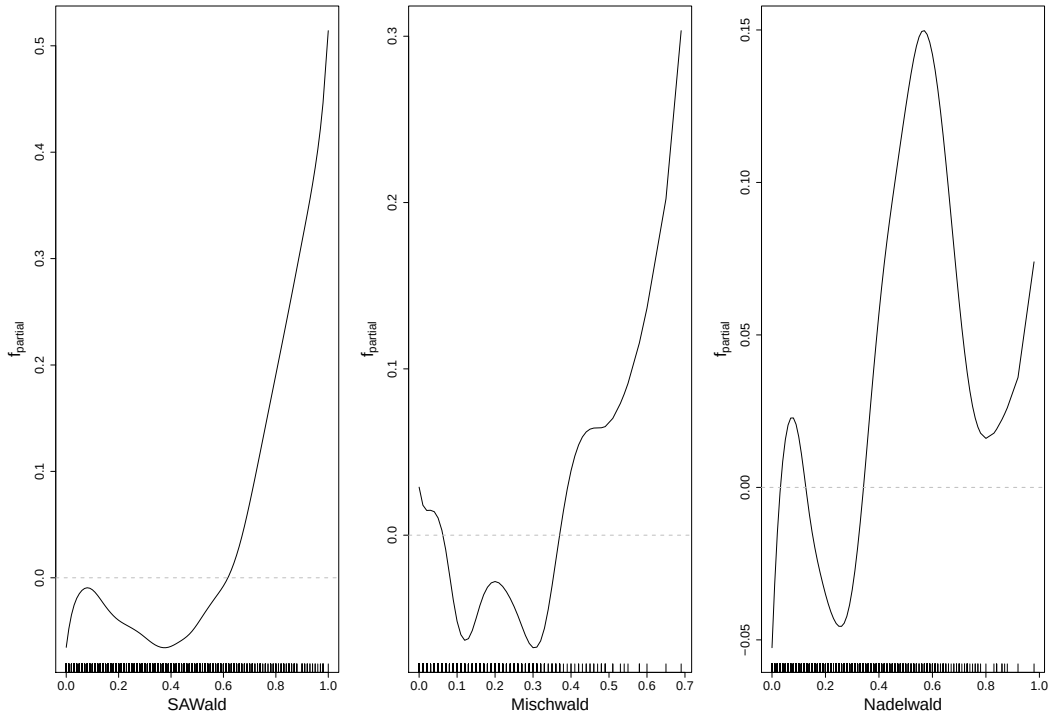


Abbildung 15: Schätzung der Effekte der Variablen `SAWald`, `Mischwald` und `Nadelwald`.

Die Interpretation der Variable `SAWald` entspricht wieder der des Modells mit der Schrittweite $\nu = 0{,}05$ mit dem Unterschied, dass hier die Schwelle vom negativen zum positiven Einfluss nicht bei 58%, sondern bei etwa 62% liegt. Auch die Variablen `Nadelwald` und `Mischwald` wurden in das Modell aufgenommen. Unterschiedliche Fledermausarten bevorzugen unterschiedliche Waldtypen. Nadelwälder bieten allerdings nicht so viele natürliche Quartiermöglichkeiten wie Laub- oder Mischwälder. Die Tiere sind hier viel stärker auf Nistkästen als Quartierhilfen angewiesen (Meschede und Rudolph, 2004). Für

das Modell gibt es ab etwa 37% Nadelwaldbedeckung eines Quadranten einen positiven Effekt auf die Artenzahl, bei Mischwald ab etwa 34%.

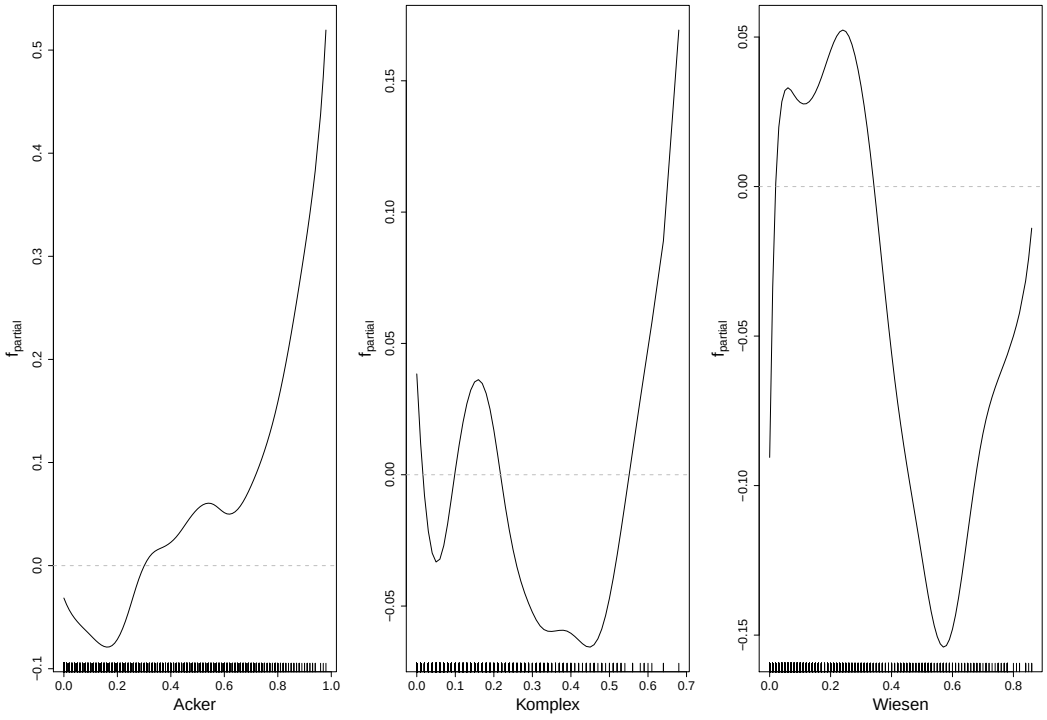


Abbildung 16: Schätzung der Effekte der Variablen `Acker`, `Komplex` und `Wiesen`.

Ebenso weist die Variable `Acker` ab einem Bedeckungsanteil von ca. 30% einen positiven Effekt auf. Der Grund dafür könnte sein, dass fast alle Fledermausarten primärwirtschaftlich genutzte Flächen, zu denen unter anderem Ackerland gehört, als Jagdraum nutzen. Je nach Nahrungsvorliebe geschieht dies in unterschiedlich intensiver Weise. So jagt zum Beispiel der Abendsegler im Herbst häufig auf abgeernteten Feldern (Meschede und Rudolph, 2004).

Im Bereich zwischen 22% und 55% hat die Variable `Komplex` einen negativen Einfluss. `Komplex` steht für Gebiete, in denen Waldrandflächen und Streuobstwiesen ineinander übergehen. Für niedrigere bzw. höhere Werte ergibt sich ein positiver Effekt.

Wiesen zeigen ab einem Anteil von etwa 35% im Quadranten eine negative Wirkung auf die Anzahl der Fledermausarten.
Insgesamt könnte es bei Fledermäusen problematisch sein, den Einfluss jeder Bedeckungsvariable einzelnd zu interpretieren, da die Tiere mobil sind. Sie sind also nicht an isolierte Habitate, d.h. an einzelne Biotope gebunden. Auch wenn Fledermäuse einige Landschaftsformen mehr nutzen als andere, bevorzugen sie in der Regel eine vielfältige, komplexe und relativ großräumige Landschaft (Meschede und Rudolph, 2004). Auch die Interpretation der einzelnen Klimavariablen ist problematisch, da Klimavariablen in der Regel sehr stark miteinander korrelieren.
Vergleicht man die Modelle der drei Schrittweiten anhand der ausgewählten Variablen, so ist auffällig, dass die Komponente $f_s(s)$ für jedes der Modelle ausgewählt wurde. Für die Schrittweite $\nu = 0{,}1$ wurde keine der Umweltvariablen ausgewählt. Die Variablen, die in das finale Modell der Schrittweite $\nu = 0{,}05$ eingingen, wurden alle auch für $\nu = 0{,}01$ ausgewählt.

Als letztes soll noch der räumliche Effekt des Modells (Add/Spatial) der drei verschiedenen Schrittweiten miteinander verglichen werden. Dieser ist in den Abbildungen 17 und 18 dargestellt. Zu sehen ist der Einfluss der Modellkomponente $f_s(s)$ und somit die erklärte Variabilität der unbeobachteten Heterogenität. Bei der Interpretation der Grafiken gilt: Je heller die Fläche, desto größer ist der positive Einfluss auf die Zielvariable. Der räumliche Effekt der drei Schrittweiten ist strukturell relativ ähnlich. So ist im Raum Unterfranken eine geringere Artenzahl zu erwarten als in den anderen Bezirken. Im Bereich von Mittelfranken und Oberbayern ist die erwartete Artenzahl von Fledermäusen wiederum niedriger als in den Bezirken Oberfranken, Oberpfalz, Niederbayern und Schwaben. Für das Modell mit der Schrittweite $\nu = 0{,}01$ ist insbesondere noch zu beobachten, dass der Effekt der räumlichen Komponente auf den Response im Raum München geringer ist als im Rest Oberbayerns.

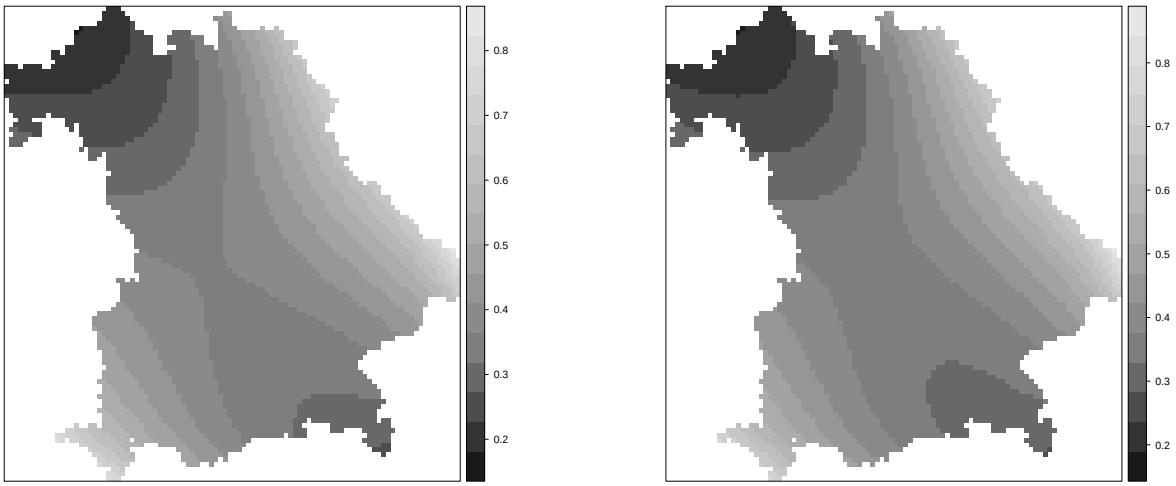


Abbildung 17: räumlicher Effekt des Modells (Add/Spatial) für $\nu = 0{,}1$ (links) und $\nu = 0{,}05$ (rechts)

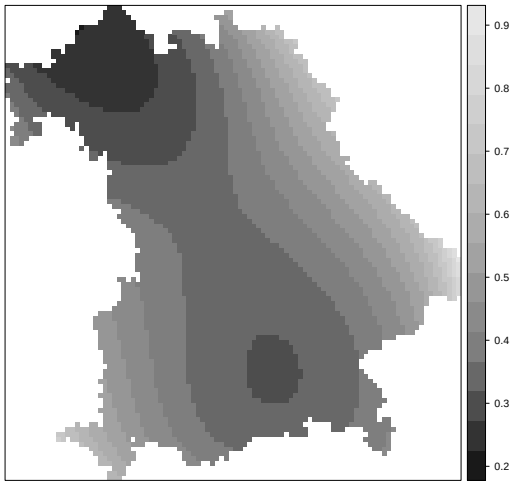


Abbildung 18: räumlicher Effekt des Modells (Add/Spatial) für $\nu = 0{,}01$

5 Zusammenfassung

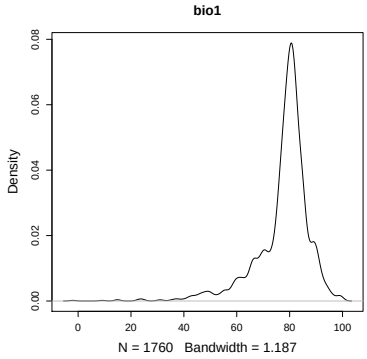
In dieser Arbeit wurde versucht, ein Habitatmodell für Fledermäuse in Abhängigkeit von Raum und Umweltvariablen aufzustellen und dabei diejenigen Umweltvariablen zu identifizieren, die die Artenvielfalt der Fledermäuse beeinflussen. Dafür wurde ein generalisiertes additives Modell mit Poisson-verteiltem Response aufgestellt, dessen Prädiktor sich in eine globale und eine lokale Komponente unterteilt. Mit Hilfe der Methode Spatial Boosting konnten sowohl Modellwahl als auch Variablenselektion durchgeführt und die Effekte der ausgewählten Variablen flexibel geschätzt werden. Aufgrund der Modellaufteilung kann auch die Verteilung der erklärten Variabilität untersucht werden.
Besonders auffallend war, dass die Variabilität der Artenzahl im Datensatz in erster Linie von der Anzahl der Exkursionen abhängig war. Zwar konnte dies als Offset herausgerechnet werden, so dass die Effekte der Kovariablen dennoch geschätzt werden konnten, problematisch wird es jedoch, wenn man auf Basis dieser Daten Vorhersagen für ununtersuchte Gebiete treffen möchte. Um dies zu verbessern, müsste man versuchen, die Diskrepanz zwischen tatsächlicher und beobachteter Artenzahl zu reduzieren, was trotz einer Vielzahl von Studien über Fledermäuse sehr schwierig und aufwändig ist.
Nur ein kleiner Teil der Variabilität kann durch die Umweltvariablen erklärt werden, dennoch konnten einige Zusammenhänge, wie der positive Einfluss der Waldvariablen auf die Artenzahl, erkannt werden.

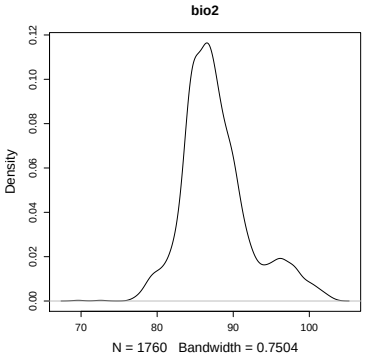
Ein noch nicht gelöstes Problem der ausgewählten Methodik ist die Wahl des Hyperparameters ν. Anders als die Parameter m_{stop} und der Glättungsparameter λ wurde dieser Parameter nicht berechnet, sondern basierend auf Erfahrungswerten festgesetzt. Eine Möglichkeit, diesen Parameter auf andere Weise zu bestimmen, ohne den Algorithmus zu kompliziert und rechenintensiv zu gestalten, ist bisher noch nicht gefunden.

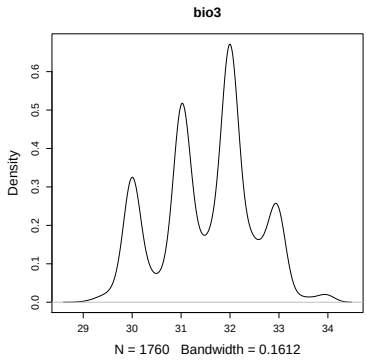
Um mehr über den Einfluss von Umweltvariablen auf die Artenvielfalt von Fledermäusen

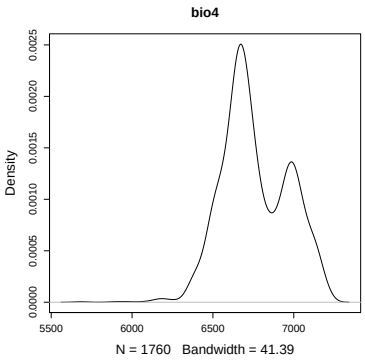
zu erfahren, könnte es auch sinnvoll sein, einzelne Arten gezielt zu untersuchen, da die verschiedenen Arten teilweise sehr unterschiedliche Vorlieben bei der Wahl von Quartieren und Jagdgebieten haben.

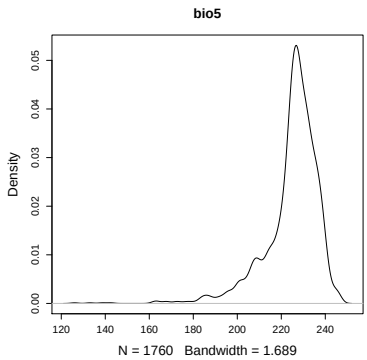
A Verteilung der Kovariablen

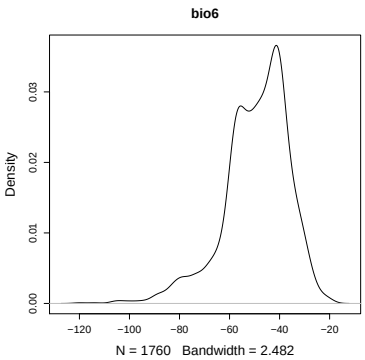


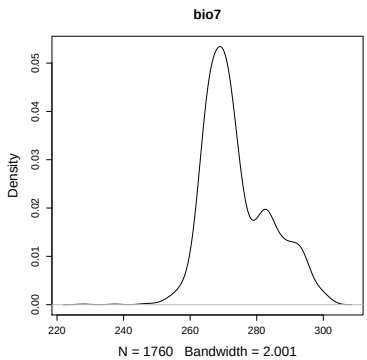


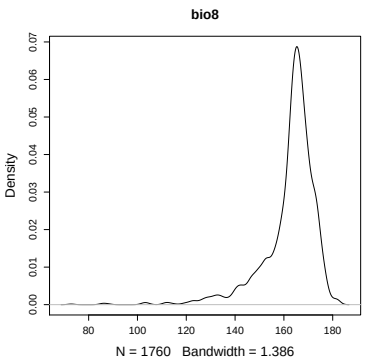


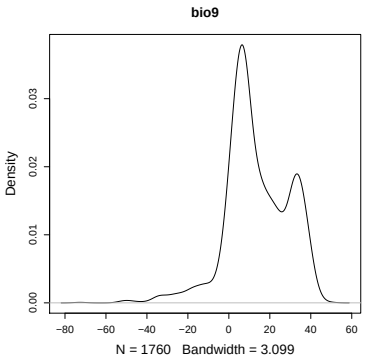


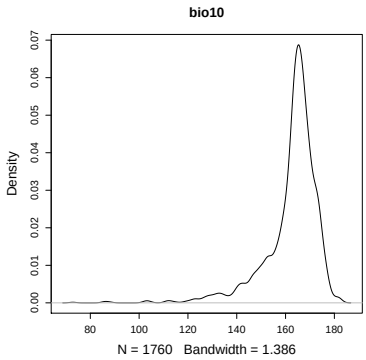


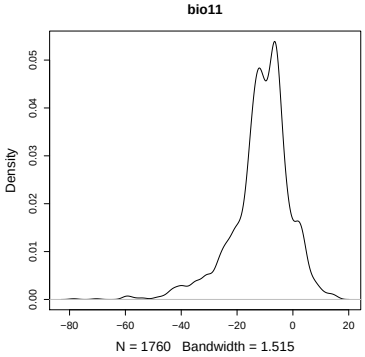


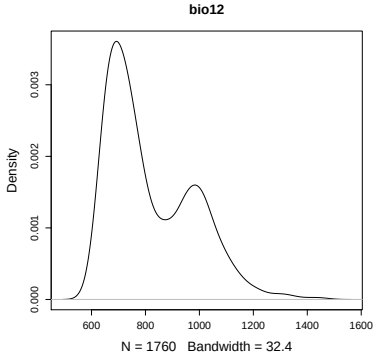


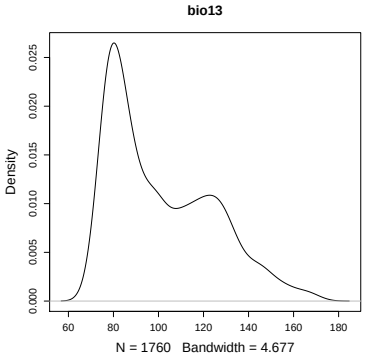


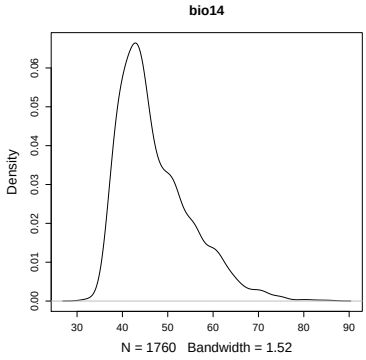

bio9
Density
N = 1760 Bandwidth = 3.099

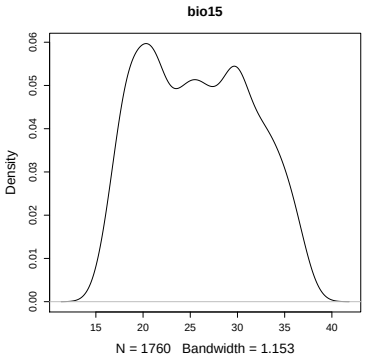

bio10
Density
N = 1760 Bandwidth = 1.386

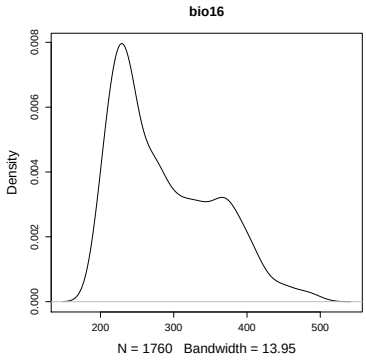

bio11
Density
N = 1760 Bandwidth = 1.515

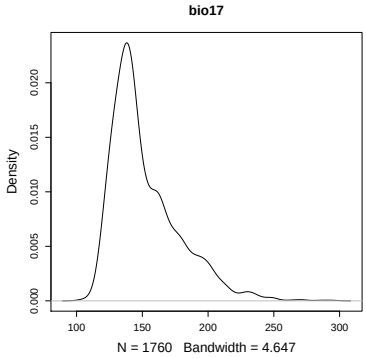

bio12
Density
N = 1760 Bandwidth = 32.4

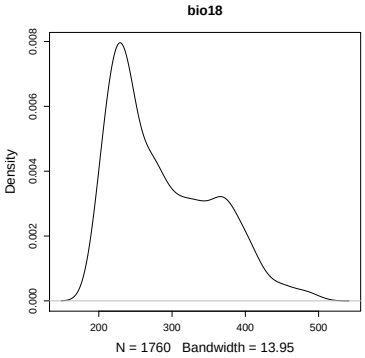

bio13
Density
N = 1760 Bandwidth = 4.677

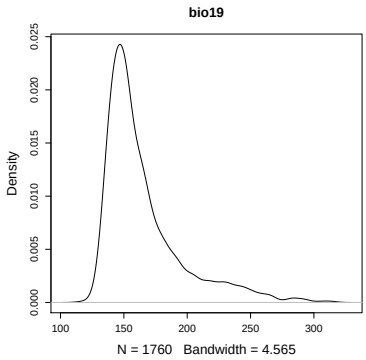

bio14
Density
N = 1760 Bandwidth = 1.52


bio15
Density
N = 1760 Bandwidth = 1.153

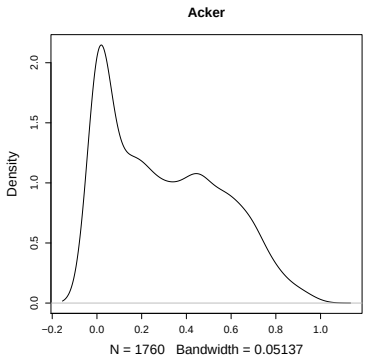

bio16
Density
N = 1760 Bandwidth = 13.95

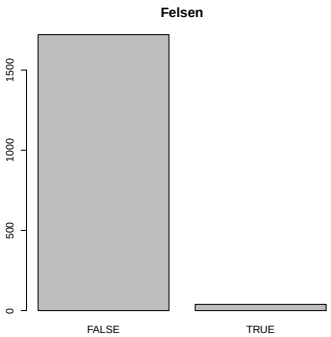

bio17
Density
0.000
0.005
0.010
0.015
0.020
100
150
200
250
300
N = 1760 Bandwidth = 4.647

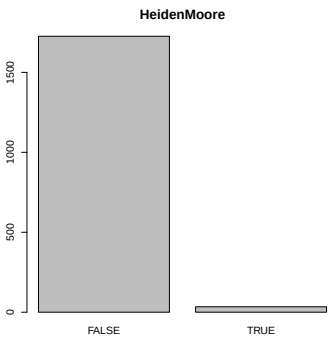

bio18
Density
0.000
0.002
0.004
0.006
0.008
200
300
400
500
N = 1760 Bandwidth = 13.95

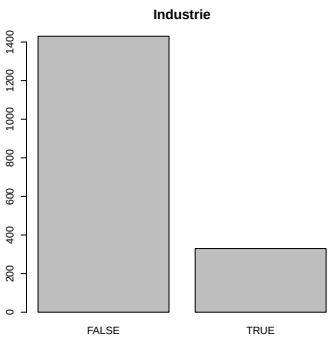

bio19
Density
0.000
0.005
0.010
0.015
0.020
0.025
100
150
200
250
300
N = 1760 Bandwidth = 4.565

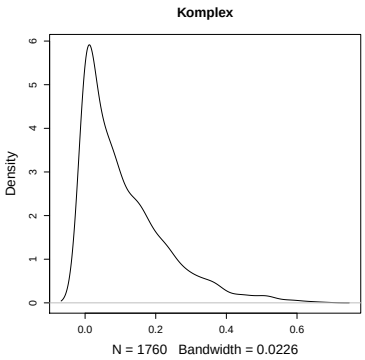

Abbauflaechen
0
500
1000
1500
FALSE
TRUE

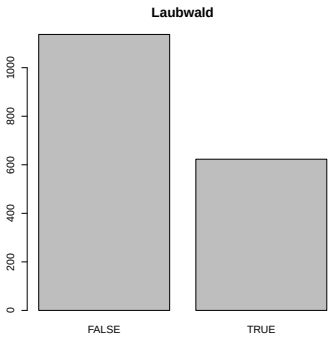

Acker
Density
0.0
0.5
1.0
1.5
2.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
N = 1760 Bandwidth = 0.05137

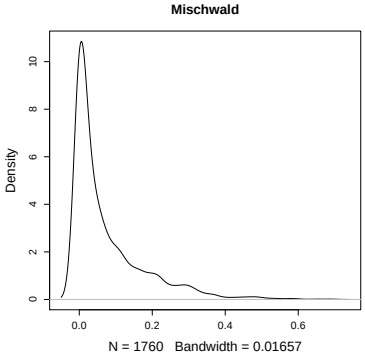

Felsen
0
500
1000
1500
FALSE
TRUE

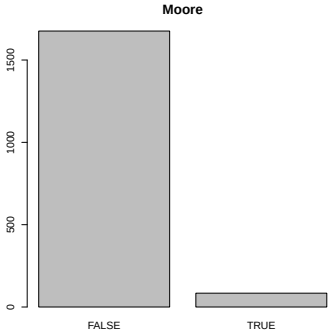

HeidenMoore
0
500
1000
1500
FALSE
TRUE

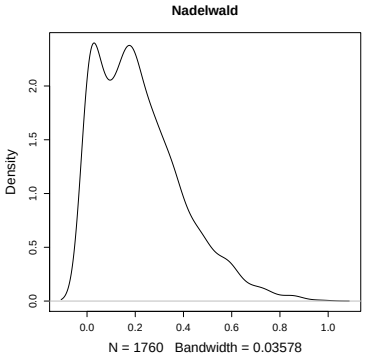

Industrie
0
200
400
600
800
1000
1200
1400
FALSE
TRUE

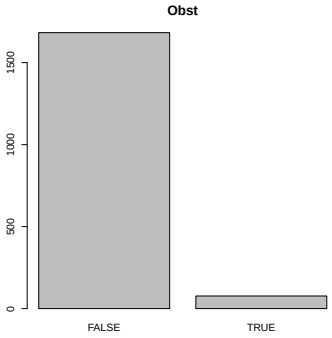

Komplex
Density
N = 1760 Bandwidth = 0.0226

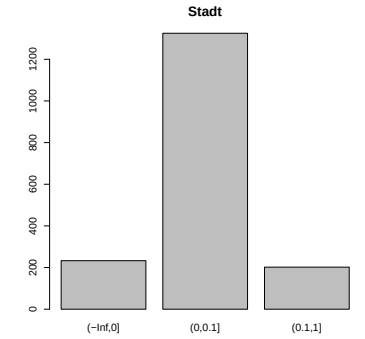

Laubwald
FALSE
TRUE

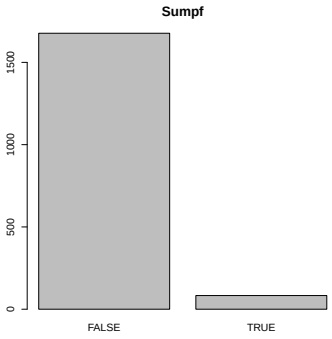

Mischwald
Density
N = 1760 Bandwidth = 0.01657


Moore
FALSE
TRUE

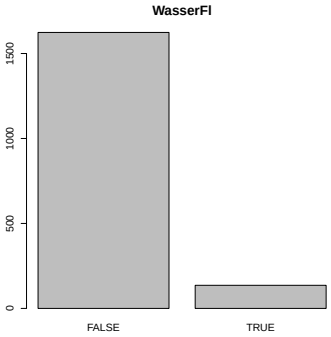

Nadelwald
Density
N = 1760 Bandwidth = 0.03578


Obst
FALSE
TRUE


Stadt
(−Inf,0]
(0,0.1]
(0.1,1]

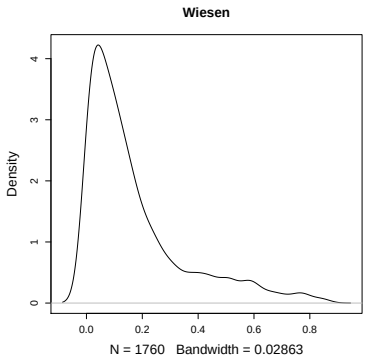

Sumpf
FALSE
TRUE

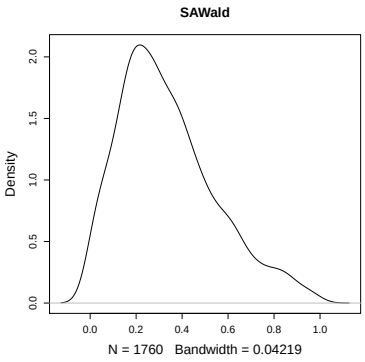

WaldrandGeb
0
200
400
600
800
1000
1200
1400
FALSE
TRUE


WasserFl
0
500
1000
1500
FALSE
TRUE

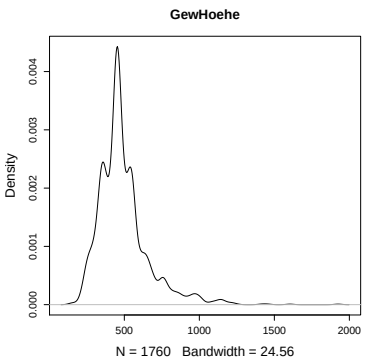

WasserSteh
0
200
400
600
800
1000
1200
1400
FALSE
TRUE


Weinbau
0
500
1000
1500
FALSE
TRUE


Wiesen
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
N = 1760 Bandwidth = 0.02863


SAWald
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
N = 1760 Bandwidth = 0.04219


SAWasser
0
200
400
600
800
1000
1200
FALSE
TRUE


GewHoehe
Density
0.000
0.001
0.002
0.003
0.004
500
1000
1500
2000
N = 1760 Bandwidth = 24.56

B Elektronischer Anhang

Der elektronische Anhang enthält die vier Ordner `data`, `preproc`, `analysis` und `report`. In `data` befinden sich die zur Verfügung gestellten Datensätze `AACorineRaumClima.csv` mit den Umweltvariablen, `Fledermaeuse1927mitNacht.csv` mit den Daten der Fledermäuse sowie die Datensätze `DEU_adm0.Rda`, `DEU_adm1.Rda`, `variables.csv` und `XY.Rda`, die für die Auswertungen verwendet wurden. Außerdem befinden sich in diesem Ordner noch die Dateien `flederm.Rda`, mit dem bearbeiteten Datensatz, `Artenzahl_0_1.Rda`, `Artenzahl_0_05.Rda`, `Artenzahl_0_01.Rda`, `Artenzahl_Modelle_0_1.Rda`, `Artenzahl_Modelle_0_05.Rda` und `Artenzahl_Modelle_0_01.Rda` mit den Modellen für die verschiedenen Schrittweiten.

Der Ordner `preproc` enthält die R-Datei `einlesen_(1).R`. Dieser Code wurde zur Aufbereitung der Daten verwendet.
Im Ordner `analysis` befinden sich die R-Dateien
`Grafik_deskriptiv_(2).R`
`Modellfit_(3).R`
`Inferenz_(4).R`
`Grafik_0_1_(5).R`
`Grafik_0_05_(5).R`
`Grafik_0_01_(5).R`
`Grafik_deskriptiv_(2).R` beinhaltet den Code der deskriptiven Analyse. Der Code für Modellfit und Modellinferenz befindet sich in `Modellfit_(3).R` und `Inferenz_(4).R`. Die Graphiken zu den Modellen der verschiedenen Schrittweiten wurden mittels `Grafik_0_1_(5).R`, `Grafik_0_05_(5).R` und `Grafik_0_01_(5).R` erstellt. Die Zahlen in Klammer zeigen an, in welcher Reihenfolge die R-Dateien auszuführen sind.

Der Ordner `report` enthält die elektronische Form dieses Berichts.

Literatur

Bates D, Maechler M (2010). *lme4: Linear Mixed-Effects Models Using S4 Classes*. R package version 0.999375-33, URL `http://CRAN.R-project.org/package=lme4`.

Bühlmann P, Hothorn T (2007). "Boosting Algorithms: Regularization, Prediction and Model Fitting." *Statistical Science*, **22**(4), 477–505. doi:10.1214/07-STS242. With discussion.

Deutsches Zentrum für Luft-und Raumfahrt eV DF (ed.) (2005). *CORINE Land Cover 2000 – Europaweit harmonisierte Aktualisierung der Landnutzungsdaten für Deutschland*, volume UBA-FB000826. URL `http://www.corine.dfd.dlr.de/`.

Dormann CF, MMcPherson J, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Kissling WD, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schnurr FM, Wilson R (2007). "Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: A Review." *Ecography*, **30**, 609–628. doi:10.1111/j.2007.0906-7590.05171.x.

Drunkenmölle H (2010). "Großer Lauschangriff auf wendige Ultraschall–Jäger." *Nordwest Zeitung*. 16.08.2010.

Fahrmeir L, Kneib T, Lang S (2004). "Penalized Structured Additive Regression for Space–Time Data: A Bayesian Perspective." *Statistica Sinica*, **14**, 715–745.

Graves S, with help from Sundar Dorai-Raj HPP (2006). *multcompView: Visualizations of Paired Comparisons*. R package version 0.1-0.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005). "Very High Resolution Interpolated Climate Surfaces for Global Land Areas." *International Journal of Climatology*, **25**, 1965–1978. URL `http://www.worldclim.org/`.

Hothorn T, Bretz F, Westfall P (2008). "Simultaneous Inference in General Parametric Models." *Biometrical Journal*, **50**(3), 346–363.

Hothorn T, Bühlmann P, Kneib T, Schmid M, Hofner B (2010a). *mboost: Model-Based Boosting*. R package version 2.0-6, URL `http://CRAN.R-project.org/package=mboost`.

Hothorn T, Hornik K, Zeileis A (2006). "Unbiased Recursive Partitioning : A Conditional Inference Framework." *Journal of Computational and Graphical Statistics*, **15**(3), 651–674.

Hothorn T, Müller J, Schröder B, Kneib T, Brandl R (2010b). "Decomposing Environmental, Spatial, and Spatiotemporal Components of Species Distributions." *Ecological Monographs*. Accepted 2010-07-15.

Kneib T, Hothorn T, Tutz G (2007). "Variable Selection and Model Choice in Geoadditive Regression Models." *Technical Report 3*, Institut für Statistik, Ludwig-Maximilians-Universität München. URL `http://epub.ub.uni-muenchen.de/2063/`.

Kneib T, Müller J, Hothorn T (2008). "Spatial Smoothing Techniques for the Assessment of Habitat Suitability." *Environmental and Ecological Statistics*, **15**(3), 343–364.

Legendre P (1993). "Spatial Autocorrelation: Trouble or new Paradigm?" *Ecology*, **74**(6), 1659–1673. By the Ecological Society of America.

Mehr M, Brandl R, Hothorn T, Dziock F, Förster B, Müller J (2010). "Land use is more important than climate for bat richness and community on a state wide regional scale." Unpublished.

Meinshausen N, Bühlmann P (2010). "Stability Selection." *Journal of the Royal Statistical Society, Series B*, **72**(4), 1–32.

Meschede A, Rudolph BU (2004). *Fledermäuse in Bayern*. Verlag Eugen Ulmer. ISBN 3-8001-3884-0.

Neuwirth E (2007). *RColorBrewer: ColorBrewer palettes*. R package version 1.0-2, URL `http://CRAN.R-project.org/package=RColorBrewer`.

Pebesma EJ, Bivand RS (2005). "Classes and Methods for Spatial Data in R." *R News*, **5**(2), 9–13. URL `http://cran.r-project.org/doc/Rnews/`.

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`.

Sarkar D (2010). *lattice: Lattice Graphics*. R package version 0.18-3, URL `http://CRAN.R-project.org/package=lattice`.

Schnappauf W, Müller E (2005). "Rote Liste der gefährdeten Tiere und Gefäßpflanzen Bayerns." Bayerisches Staatsministerium für Umwelt, Gesundheit und Verbraucherschutz. Kurzfassung, URL `http://www.starnberg.bund-naturschutz.de`.

Zahn U (1992). *Diercke Weltatlas*. Westermann.

Erklärung zur Urheberschaft

Hiermit versichere ich, die vorliegende Bachelorarbeit selbständig verfasst, und keine anderen als die angegebenen Quellen verwendet zu haben. Die Arbeit wurde weder in dieser noch in ähnlicher Form als Prüfungsleistung für eine andere Prüfung eingereicht.

München, den 26. August 2010

_____

(Carola Kobayashi)