

# BEWEGUNGSAKTIVITÄT VON REHWILD IM TAGES- UND JAHRESVERLAUF

## Bachelorarbeit

im Studienfach Statistik



vorgelegt von: Elizabeth Heller

Matrikelnummer: 37206431

Betreuer: Prof. Dr. Torsten Hothorn

Datum: 21. September 2009

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Voraussetzungen . . . . .	1
1.2	Ziel . . . . .	2
<b>2</b>	<b>Beschleunigungssensoren</b>	<b>3</b>
2.1	Triaxiale Beschleunigungssensoren für die Bestimmung der Bewegungsmuster von Wildtieren . . . . .	4
2.1.1	Statische Beschleunigung . . . . .	5
2.1.2	Dynamische Beschleunigung . . . . .	6
2.2	Die Anwendung von biaxialen Beschleunigungssensoren bei der Untersuchung von Wildrehen . . . . .	8
<b>3</b>	<b>Deskriptive Datenanalyse</b>	<b>11</b>
<b>4</b>	<b>Modelle</b>	<b>18</b>
4.1	Boosting . . . . .	19
4.2	Schätzung der zweidimensionalen Interaktion . . . . .	20
4.3	Additive Modelle für Männchen und Weibchen . . . . .	22
4.4	Additives Modell mit zufälligem Effekt . . . . .	29
4.4.1	Modell für die Weibchen . . . . .	29
4.4.2	Modell für die Männchen . . . . .	32
<b>5</b>	<b>Fazit</b>	<b>37</b>
	<b>Literaturverzeichnis</b>	<b>38</b>
	<b>Eidesstattliche Erklärung</b>	<b>40</b>
<b>A</b>	<b>Anhang</b>	<b>i</b>
A.1	Verwendete Software . . . . .	ii
A.2	Verwendete Hardware . . . . .	iii
A.3	Uhrzeiten . . . . .	iv
A.4	Kodierung der Tage . . . . .	vii
A.5	Vogelperspektive . . . . .	x

# 1 Einleitung

Seit Anfang der neunziger Jahre werden in der Biologie und Zoologie Beschleunigungssensoren für die Aufzeichnung von Aktivitätsdaten von Tieren immer häufiger eingesetzt.

Die ständigen Entwicklungen in der Elektrotechnik ermöglichen sowohl den Bau von immer kleineren und zuverlässigeren Beschleunigungssensoren, als auch die Verfügbarkeit von grösseren und preisgünstigeren Speicherkapazitäten. Dies erhöht die Attraktivität des Einsatzes der Sensoren für die Untersuchung der Lebensumstände von Tieren, die in der Wildnis leben - einerseits, weil es einfacher ist, die Geräte zweckmässig an dem Tier zu befestigen; andererseits, weil durch die vergrösserte Speicherkapazität die ununterbrochene Aufzeichnungszeit des Geräts beachtlich erweitert wird.

In Abhängigkeit sowohl von den Eigenschaften des jeweiligen Beschleunigungssensors und dessen Konfiguration, als auch von dem angestrebten Ziel der Untersuchung haben sich inzwischen verschiedene Techniken für eine sinnvolle Interpretation dieser riesigen Anzahl von Daten etabliert.

## 1.1 Voraussetzungen

Die vorliegende Arbeit befasst sich mit der Analyse der Aktivitätsdaten von freilebenden, mit Halsbandsendern ausgestatteten Rehen (*Capreolus capreolus*), die im Nationalpark Bayerischer Wald zwischen den Jahren 2006 und 2008 erhoben wurden. Dabei wird die Bearbeitung und Kodierung der Daten übernommen, die in einer früheren exploratorischen Arbeit im Rahmen des Statistischen Praktikums 2009 in der Ludwig-Maximilians-Universität München durchgeführt wurde. Die Ergebnisse der Praktikumsarbeit sind im daraus entstandenen Bericht [[Heller u. a. 2009](#)] festgehalten.

Die aus den Halsbändern ausgelesenen Daten lagen für die Auswertung im Praktikum als einzelne Dateien vor: insgesamt 44 Aktivitätsdateien von 41 Rehen. Die Menge an Rohdaten umfasste 3 182 582 Einträge im Zeitraum von Mai 2004 bis März 2009. Nach einer Datenbereinigung und -auswahl standen Daten von 32 Rehen im Zeitraum 2006-2008 zur Verfügung. Auf dieser Grundlage wurden mehrere Datensätze erstellt, von denen zwei auch in der vorliegenden Studie verwendet werden: der Datensatz REHE5 enthält Aktivitätsmessungen im 5 Minuten-Abstand, wie sie ursprünglich erhoben wurden. Der Datensatz REHE15 enthält die auf 15 Minuten aggregierten Daten.

Für unsere Studie wurden aus REHE5 bzw. REHE15 die Daten der Halsbänder vom Typ 2, d.h. mit den Seriennummern 4931, 4933 und 4931 entfernt, da deren Messungen sich

als zu den anderen Halsbandtypen wesentlich unterschiedlich erwiesen. Die Datenmenge beträgt nun 2 463 193 bzw. 832 102 Einträge. Es handelt sich um Aktivitätsdaten von 29 Rehen: 20 Männchen und 9 Weibchen.

## 1.2 Ziel

Primäres Ziel der statistischen Analyse ist die Ermittlung der Einflüsse verschiedener Variablen auf die vom Beschleunigungssensor aufgezeichnete Aktivität. Die Aktivität im Tages- und Jahresverlauf wurde im Praktikum mit Hilfe additiver Modelle getrennt analysiert. Hier wird die Interaktion der Tages- und Jahresaktivität, sowohl bei Weibchen als auch bei Männchen, unter besonderer Berücksichtigung der Eigenschaften der verwendeten Beschleunigungssensoren modelliert. Für die Schätzung der Modelle werden Methoden aus dem Gebiet des maschinellen Lernens herangezogen.

## 2 Beschleunigungssensoren

*Accelerometry* ist die englische Bezeichnung für ein Teilgebiet der Ingenieurwissenschaften, die sich mit der Berechnung der Geschwindigkeitszunahme oder -abnahme von Körpern oder Teilen von Körpern befasst.

Dabei wird die Änderung der Geschwindigkeit mit Hilfe von Beschleunigungssensoren gemessen. Die Einsatzbereiche der Beschleunigungssensoren ergeben sich auf Gebieten wie Untersuchungen zur Lebensdauer von Produkten durch Vibrations- und Schocktests, die Erfassung von Zuständen in der Prozess- und Automatisierungstechnik oder auch biomechanische Untersuchungen von Vibrations- und Stosseinflüssen auf den Menschen. Die genauesten Beschleunigungssensoren lassen sich zur Bestimmung der dynamischen Komponenten und der Richtung der Bewegung des Körpers einsetzen. Daher werden sie seit den neunziger Jahren des vorigen Jahrhunderts in der Biologie verwendet, um Kenntnisse über das Auftreten einer Vielzahl von Verhaltensweisen von Wildtieren zu gewinnen, die sonst sehr schwer zu beobachten sind [Shepard u. a. 2008]. Die Arbeit von Yoda über das Verhalten der Pinguine gilt als einer der ersten Versuche, Beschleunigungssensoren bei der Untersuchung des Verhaltens von Wildtieren zu verwenden (e.g. Yoda et al. 1999).

Die Grundprinzipien der Anwendung von Beschleunigungssensoren in der Biologie beruhen auf den gemeinsamen Eigenschaften, die durch die Beschleunigungsmuster von verschiedenen Tieren entstehen. Für die Biologie ist das Tierverhalten eine Reaktion auf die Umgebung und den physiologischen Zustand des Tieres (e.g. Wilson et al. 2007). Das Tierverhalten kann in Bewegung und Körperhaltung unterteilt werden. Beide Merkmale können als Parameter aufgefasst und durch die am Tier befestigten Beschleunigungssensoren gemessen werden.

Die Interpretation der vom Sensor gelieferten Beschleunigungen stellt eine schwierige Aufgabe dar, denn die statischen und dynamischen Komponenten müssen aus der gesamten Beschleunigung entnommen werden. Die statischen Komponenten beziehen sich auf die Körperhaltung, während die dynamischen Komponenten auf Bewegung hinweisen.

Die biomechanischen Prozesse, die den Tieraktivitäten wie etwa der Fortbewegung zugrunde liegen, sind eingehend untersucht worden [Alexander 1982] und werden von Biologen und Zoologen als Grundlage für die Zuordnung der Beschleunigungsmuster zu den Aktivitätsmustern der Tiere betrachtet.

Die Möglichkeit, solche Aktivitätsmuster zuverlässig zu bestimmen, hängt sowohl von der Häufigkeit der Beobachtungen ab, als auch von der Art und Weise wie das Gerät am Tier fixiert ist (e.g. Macdonald et al. 2008).

Je nach Bau sind die Beschleunigungssensoren in der Lage, die Beschleunigung von einer, zwei oder drei Achsen des Körpers gleichzeitig aufzuzeichnen. Entsprechend spricht man von mono-, bi- und triaxialen Beschleunigungssensoren.

## 2.1 Triaxiale Beschleunigungssensoren für die Bestimmung der Bewegungsmuster von Wildtieren

Der Einsatz von triaxialen Beschleunigungssensoren in der Untersuchung von Bewegungsmustern wird aus einer Reihe von Gründen bevorzugt: einerseits ermöglichen sie die Messung der Bewegung in allen Dimensionen, was sich besonders positiv auf die Einschätzung der gesamten Beschleunigung des Körpers auswirkt. Die Verfügbarkeit der drei Dimensionen der Bewegung eröffnet neue Möglichkeiten für die Erfassung und Identifikation verschiedener Tierverhalten. Andererseits ist die genaue Fixierung des Geräts für die Interpretation der Signale von triaxialen Beschleunigungssensoren nicht von zentraler Bedeutung, da die von den verschiedenen Kanälen gelieferten Informationen sich gegenseitig ausgleichen. Im Gegensatz dazu hängt die richtige Interpretation der Beschleunigung bei mono- und biaxialen Sensoren von der Befestigung des Gerätes und deren Übereinstimmung mit den Körperachsen des Tieres ab.

Da in der Literatur zu diesem Thema detaillierte Informationen zur Interpretation der gelieferten Signale zur Verfügung stehen, werden wir uns mit den Hauptbegriffen und Ergebnissen dieser Studien auseinandersetzen, um sie als Bezugsrahmen für unsere Untersuchungen verwenden zu können.

Bei der Betrachtung der wesentlichen Begriffe dieses Gegenstandsbereichs konzentrieren wir uns auf die Resultate aus drei Studien von Macdonald und Wilson [[Shepard u. a. 2008](#)],[[Wilson u. a. 2008](#)],[[Wilson u. a. 2007](#)].

In diesen Studien wurde ein Beschleunigungssensor benutzt, der in der Lage war, mit der relativ hohen Auflösung von 22 bit Aufzeichnungen zu machen.

Es muss zuerst betont werden, dass diese Wissenschaftler den Schwerpunkt ihrer Untersuchung auf die Gewinnung von erkennbaren Bewegungsmustern aus den Beschleunigungsdaten legen, und das bedeutet: sie bemühen sich um die Erfassung von Bewegungsmustern, die sich bestimmten Verhaltensweisen der Tiere zuordnen lassen.

Die drei Kanäle eines triaxialen Beschleunigungssensors werden als *Heave*, *Surge* und *Sway* bezeichnet. Abbildung 2.1 zeigt, wie hilfreich die Informationen der drei Kanäle zur genauen Bestimmung der Einzelheiten der Bewegung des Tieres sind.

Wie bereits erwähnt, wird die Beschleunigung in zwei Komponenten aufgeteilt: eine statische und eine dynamische.

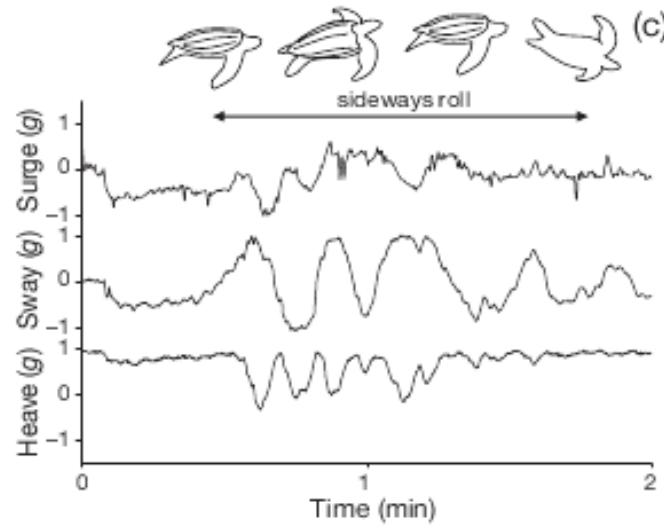


Abbildung 2.1: Die drei Kanäle eines triaxialen Beschleunigungssensors. Entnommen aus [Shepard u. a. 2008]

### 2.1.1 Statische Beschleunigung

Die statische Beschleunigung ist eine Grösse, welche die Neigung des Beschleunigungssensors bezüglich des Gravitationsfelds der Erde angibt. Diese Grösse ermöglicht die Berechnung des Winkels des besenderten Tieres. Die aus den drei Dimensionen berechnete statische Beschleunigung kann benutzt werden, um zwei wichtige Grössen zu erhalten: *pitch* oder Höhe und *roll* oder Rollen. Die Höhe wird aus dem Arkussinus von  $g$  gewonnen, wobei  $g$  die statische Beschleunigung auf dem *heave* Kanal des Beschleunigungssensors ist. Das Rollen wird genauso berechnet, aber mit der statischen Beschleunigung auf dem *sway* Kanal. Es ist auch möglich, *pitch* und *roll* einfach durch visuelle Inspektion aus den Werten der gesamten Beschleunigung einzuschätzen. Die Bestimmung von *pitch* und *roll* ist nach diesen Studien einer der wichtigsten Schritte in der Klassifizierung der Bewegungsmuster von Tieren, einerseits, weil jede Körperhaltung des Tieres den Umfang der möglichen Verhaltensweisen beschränkt und andererseits, weil je nach Tier es vorkommen kann, dass eine bestimmte Körperhaltung charakteristisch für ein bereits bekanntes Verhalten ist. Auch einige Arten der Fortbewegung ereignen sich mit einer minimalen dynamischen Beschleunigung, für ihre Erfassung ist vor allem die statische Beschleunigung von Interesse.

Abbildung 2.2 zeigt die Werte der statischen Beschleunigung bei zwei verschiedenen Körperhaltungen eines Magellan-Pinguins.

Der andere Faktor, der für die Interpretation der Signale entscheidend ist, ist die Kenntnis der Zeitlänge, während der das Tier eine bestimmte Körperhaltung einnimmt. Die

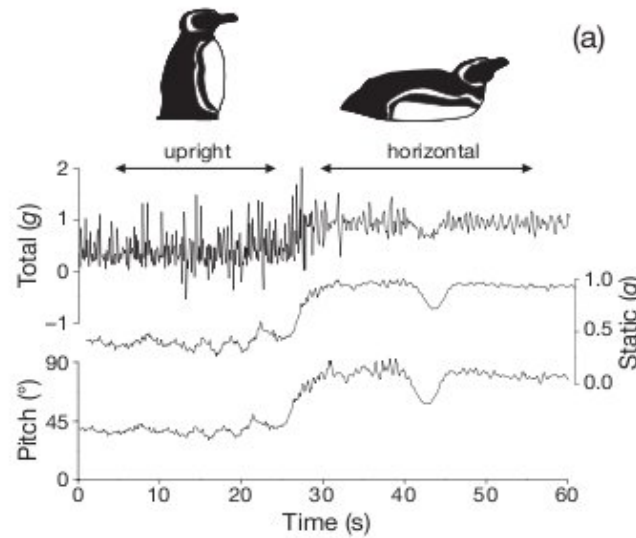


Abbildung 2.2: Statische Beschleunigung eines Magellan-Pinguins. Entnommen aus [Shepard u. a. 2008]

Reihenfolge verschiedener Körperhaltungen über eine relativ kurze Zeit kann für die Klassifizierung eines bestimmten Verhaltens ausschlaggebend sein.

### 2.1.2 Dynamische Beschleunigung

Die dynamische Komponente stellt die Änderung der Geschwindigkeit als Ergebnis der Körperbewegung des Tieres dar. Sie bezieht sich also auf fast alle Modalitäten der Fortbewegung. Sie kann als eine Subtraktion der statischen Beschleunigung aus der gesamten Beschleunigung aufgefasst werden. Die dynamische Beschleunigung wird ausgewertet nicht nur, um die Fortbewegungsmustern der Tiere zu klassifizieren, sondern auch um eine Schätzung der Geschwindigkeit der Fortbewegung zu erhalten. Beide Informationsquellen, d.h. die Art der Fortbewegung und die Geschwindigkeit können wiederum benutzt werden, um noch detaillierte Erkenntnisse über das Verhalten und den körperlichen Zustand der Tiere zu gewinnen.

Abbildung 2.3 zeigt die Änderung der dynamischen Beschleunigung während der Fortbewegung eines Amaisenbären.

Mit Hilfe dieser physikalischen Begriffe und der von ihnen beschriebenen Techniken sind diese Wissenschaftler in der Lage, die aus den Sensoren entstehenden Zeitreihen so zu interpretieren, dass bekannte Bewegungsmuster der Tiere sichtbar werden.



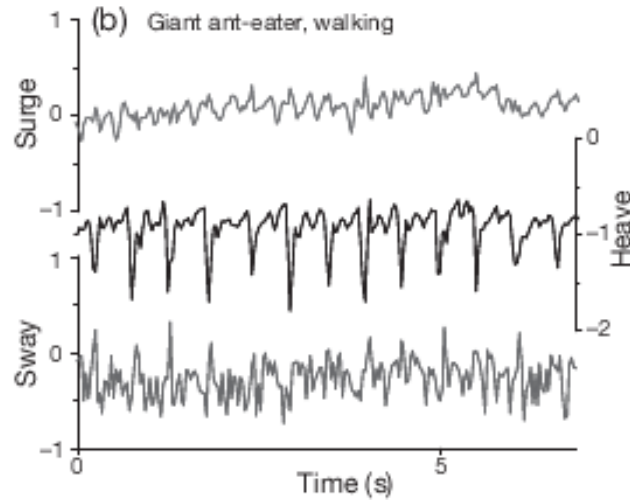


Abbildung 2.3: Dynamische Beschleunigung während der Fortbewegung. Entnommen aus [Shepard u. a. 2008]

Eine wichtige Voraussetzung der Analyse war in diesen Studien eine einleitende Beobachtung von in Zoos gehaltenen Tieren. Jedes Verhalten der Tiere wurde zusammen sowohl mit den Beschleunigungswerten als auch mit der Zeitdauer aufgezeichnet. Aus diesen Informationen wurden verschiedenen Statistiken erstellt, die für die Zuordnung der Signale zu bestimmten Verhaltensweisen der Wildtiere als Vorlage dienten.

Das Resultat dieser Arbeiten war, besonders bei [Wilson u. a. 2008], die Erstellung von einem Diagramm, das die automatische Klassifizierung von jedem Verhalten des Tieres in Abhängigkeit von den Werten der bereits besprochenen Parameter ermöglicht. Die vom Beschleunigungssensor gelieferten Werte sind dabei die Größen, welche die genaue Quantifizierung der Bewegung möglich machen.

Abbildung 2.4 zeigt eine schematische Karte für die Interpretation der Signale, die aus der Aktivität eines Kormorans entstehen [Wilson u. a. 2008].

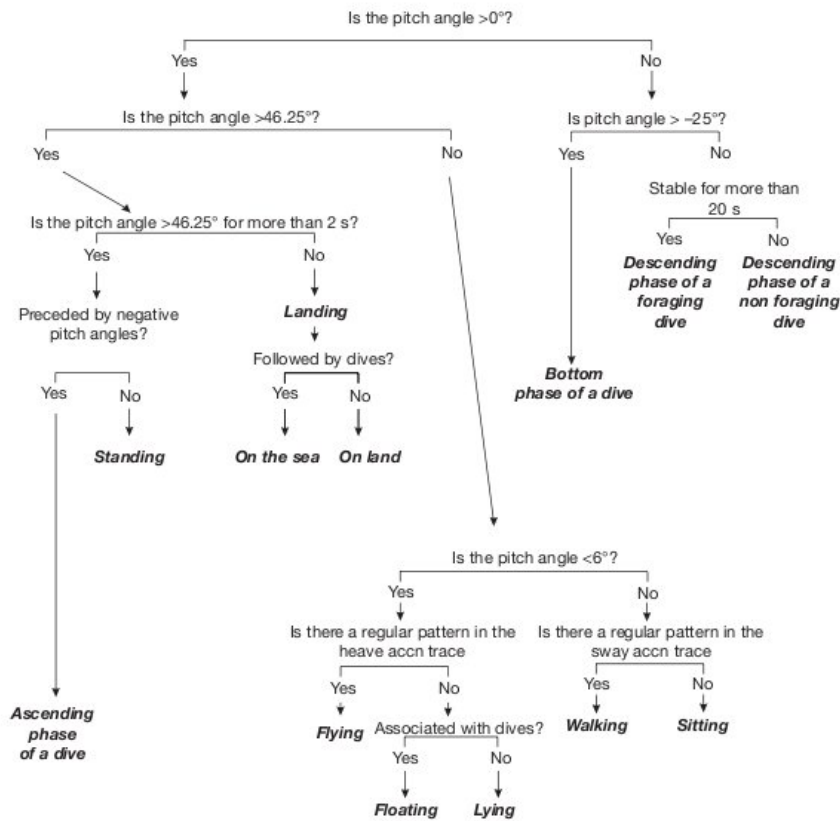


Abbildung 2.4: Diese Karte zeigt, wie die besprochenen Grössen benutzt werden können, um das Verhalten des Kormorans zu beschreiben.

Ein für unsere Arbeit wichtiges Ergebnis aus diesen Untersuchungen besteht in der Erkenntnis, dass die hohe Auflösung der Daten eine Einsicht in Vorgänge liefert, die nur Sekunden dauern, was sich wiederum sofort auf die Qualität der Bestimmung und Einschätzung der Muster auswirkt. Die Kriterien für die Klassifizierung bestimmter Verhaltensweisen werden also in Sekunden gemessen. Andererseits werden auch hier die Werte der ursprünglichen Beschleunigung benötigt, um die Berechnungen anzustellen, welche die Entdeckung der richtigen Muster ermöglichen.

## 2.2 Die Anwendung von biaxialen Beschleunigungssensoren bei der Untersuchung von Wildrehen

Die Eklärungen im vorangegangenen Abschnitt dienten dazu, einige Resultate und Methoden darzustellen, die einen Überblick über den gegenwärtigen Stand der Forschung

bieten und die für die Behandlung der uns zur Verfügung gestellten Daten nützlich sein werden. Im Folgenden beschreiben wir kurz die Eigenschaften der Beschleunigungssensoren, die bei der Messung der Aktivität der Rehe im Nationalpark Bayerischer Wald zum Einsatz kommen.

Die Beschleunigungssensoren sind Bestandteil der von der Firma Vectronic Aerospace gelieferten Halsbänder. Nach Angaben der Firma misst der eingebaute Aktivitätssensor „...kontinuierlich die Bewegung des Tieres. Diese Daten werden automatisch aufsummiert und alle 5 Minuten zusammen mit der Temperatur intern abgespeichert. Der Messzeitraum beträgt bis zu drei Jahren.“<sup>1</sup>.

Aus diesen Eigenschaften wird erstens ersichtlich, dass wir nicht über die Beschleunigung im eigentlichen Sinne verfügen, was die Aufteilung der Beschleunigung in Komponenten unmöglich macht, und zweitens, dass der minimale Abstand zwischen Beobachtungen fünf Minuten beträgt.

Dieser Beschleunigungssensor zeichnet die Aktivität des Tieres auf zwei Kanälen: X und Y auf. Im Bericht zum Statistischen Praktikum wurde durch graphische Analyse die Ähnlichkeit der Aktivitätsmuster in X- und Y-Richtung festgestellt. Der berechnete Korrelationskoeffizient betrug 0.95, so dass von einem starken Zusammenhang zwischen beiden Größen auszugehen war. Deswegen wurde aus der X- und Y-Aktivität durch Addition eine neue Variable gebildet: die Variable *akt*, mit einem Wertebereich von 0 bis 510. Diese Variable war die Zielgröße der Untersuchung und wird hier übernommen.

Abbildung 2.5 zeigt die Ähnlichkeit der beiden Aktivitätswerte für das Reh Stefan.

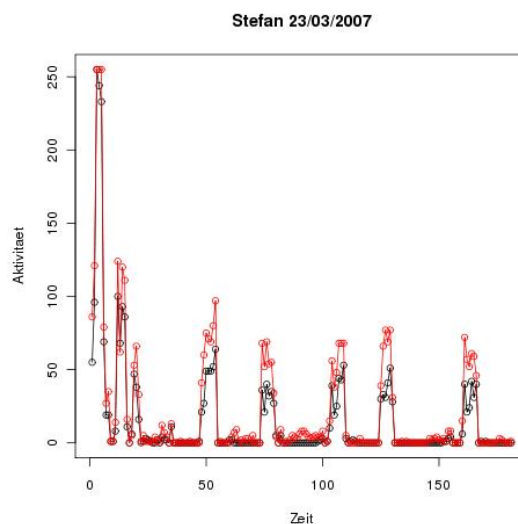


Abbildung 2.5: X- und Y-Aktivität von Reh Stefan am 23/03/2007

---

<sup>1</sup>Website: [http://www.vectronic-aerospace.com/html/dgps\\_plus.html](http://www.vectronic-aerospace.com/html/dgps_plus.html)

Wie man in Abbildung 2.3 für den Fall des Ameisenbären feststellen kann, sind die Signale für die *surge* und *sway* keineswegs gleich. Erst durch die Analyse der Unterschiede zwischen den verschiedenen Dimensionen und in Abhängigkeit davon, um welche Tier- und Bewegungsart es sich handelt, ergibt sich nach Alexander [Alexander 1982] die richtige biomechanische Interpretation des Verhaltens. Dies scheint bei den Aktivitätsdaten der Rehe nicht realisierbar zu sein, denn die Signale der zwei Kanäle unterscheiden sich kaum voneinander.

Darüber hinaus muss festgehalten werden, dass keine direkten Beobachtungen der Rehe mit den Geräten systematisch durchgeführt wurden, die uns eine Vorlage für die Bewegungsmuster liefern könnten.

Diese Tatsachen bestimmen entscheidend sowohl unseren Ansatz bei der Modellierung der Daten als auch die Grenzen unserer Untersuchung. Unsere Bemühungen müssen sich auf die Entdeckung von Regelmässigkeiten im Verlauf dieser unbestimmten Aktivität beschränken, auf die Bestimmung von genaueren Bewegungsmustern muss jedoch verzichtet werden.

### 3 Deskriptive Datenanalyse

Im Zusammenhang mit den Ausführungen und Ergebnissen des vorausgehenden Abschnittes konzentrieren wir uns in diesem Teil darauf, die Daten nach Gesetzmässigkeiten zu untersuchen. Da die Daten für die Aktivität der einzelnen Rehe nichts anderes als Zeitreihen sind, wenden wir einige Begriffe aus der Zeitreihentheorie auf sie an [[Cryer und Chan 2008](#)],[[Shumway und Stoffer 2006](#)].

In erster Linie wenden wir die Autokorrelationsfunktion auf einzelne Rehe an, um erste Anhaltspunkte für das Auffinden der möglichen sich wiederholenden Muster zu gewinnen.

Die sogenannten empirischen Autokorrelationskoeffizienten sind ein Mass für die Autokorrelation der Beobachtungen, die durch verschiedene Entfernungen oder Lags getrennt sind. Diese Koeffizienten ermöglichen eine erste Diagnose des probabilistischen Modells, das den Daten zugrunde liegen könnte, weil sie den Grad der Abhängigkeit zwischen den Beobachtungen erfassen.

Die Korrelation der Beobachtungen, die durch einen Lag  $k$  getrennt sind, kann durch

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}$$

berechnet werden [[Chatfield 2004](#)].

Die graphische Darstellung der empirischen Autokorrelationsfunktion heisst Korrelogramm. Das Korrelogramm ist ein hilfreiches Mittel für die Interpretation von einer Folge von Autokorrelationskoeffizienten, weil es die Abhängigkeiten zwischen Beobachtungen visualisiert.

Wenn in einer Zeitreihe saisonbedingte Schwankungen oder Saisonkomponenten enthalten sind, dann zeigt auch die graphische Darstellung der empirischen Autokorrelationsfunktion Schwankungen in genau derselben Frequenz der ursprünglichen Daten. Wir werden uns diese Eigenschaft bei der Suche nach periodischen Mustern zunutze machen.

Wir fangen mit der Analyse des männlichen Rehes Stefan an. Die Daten für das Reh Stefan erstrecken sich über zwei Jahre: 2007 und 2008. Abbildung [3.1](#) zeigt die Zeitreihe der Aktivität für die ersten fünf aufgezeichneten Tage des Jahres 2007 (23 März bis 27 März).

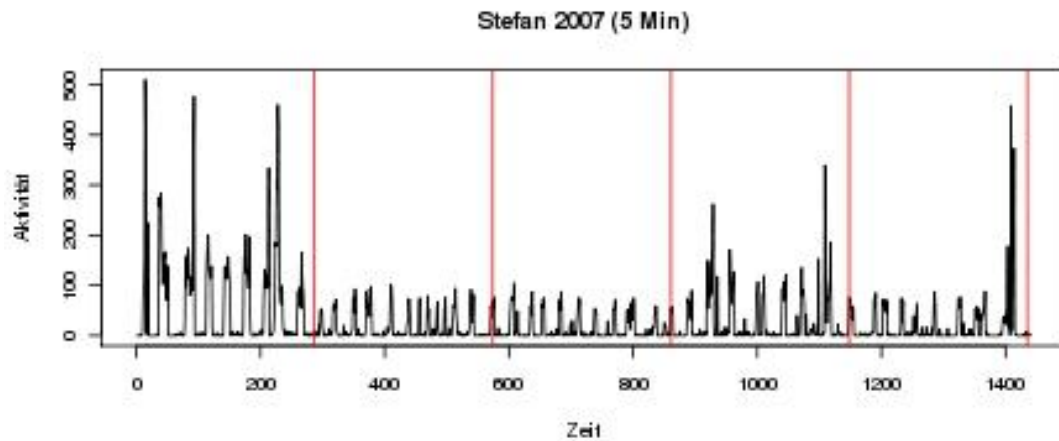


Abbildung 3.1: Tage vom 23. März bis zum 27. März 2007

Die roten Linien teilen die Beobachtungen in Tage auf. Auch wenn man in dieser Zeitreihe keine ausgeprägten Muster erkennen kann, die Anhaltspunkte für die Bestimmung von präzisen Bewegungsmustern liefern könnten, ist es nicht zu übersehen, dass jeder Tag eine Reihe von kleinen Erhebungen in mehr oder weniger regelmässigen Abständen enthält.

Abbildung 3.2 stellt das Korrelogramm für diesen Zeitabschnitt dar.

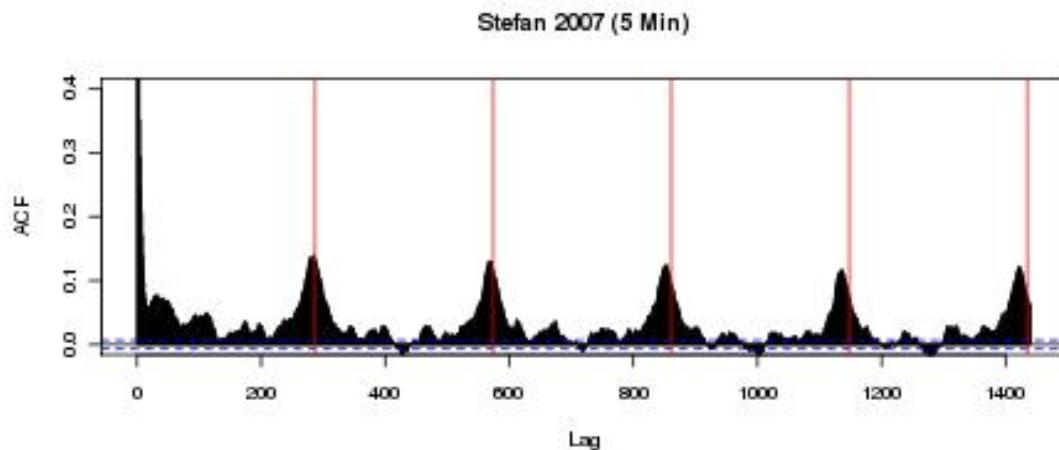


Abbildung 3.2: Korrelogramm für die Aktivität von Stefan mit Lag = 1440

Das starke regelmässige Muster, das in diesem Korrelogramm zum Vorschein kommt, wird saisonbedingte Schwankung genannt. Eine nähere Betrachtung zeigt, dass das Korrelogramm für diese Zeitreihe eine Regelmässigkeit zutage bringt, welche der Korrelation

zwischen den Beobachtungen entspricht, die durch einen Lag von 288 Zeiteinheiten entsteht. Diese 288 Zeiteinheiten entsprechen genau einem Tag. Man kann leicht feststellen, dass auch die Vielfachen von 288, d.h. 576, 864, usw. eine sehr starke positive Korrelation haben. Solche Schwankungen für tägliche Werte treten auf, wenn den durch einen Tag getrennten Beobachtungen eine sehr starke periodische Gesetzmässigkeit zugrunde liegt.

Die erhaltenen Werte für die Autokorrelationsfunktion bringen uns zu der Überzeugung, dass es angebracht ist, einen Modellierungsansatz zu wählen, der diese Periodizität von Tagen streng berücksichtigt. Ein Ansatz, der sie ausser Acht lässt, könnte zwar plausibel erscheinen, ist jedoch nicht geeignet, um die wesentlichen Züge der Aktivitätswerte zu erfassen.

Abbildung 3.3 zeigt das Ergebnis der Anwendung verschiedener Glättungsverfahren auf die Daten von Reh Stefan. Diese Verfahren wurden ohne Beachtung der lokalen Struktur der Zeitreihe eingesetzt. Die Nichtbeachtung der zugrundeliegenden repetitiven Muster führt in diesen Fällen zur Darstellung einer den Daten nicht innewohnenden Tendenz, die die Auffindung von Bewegungsmustern unmöglich macht.

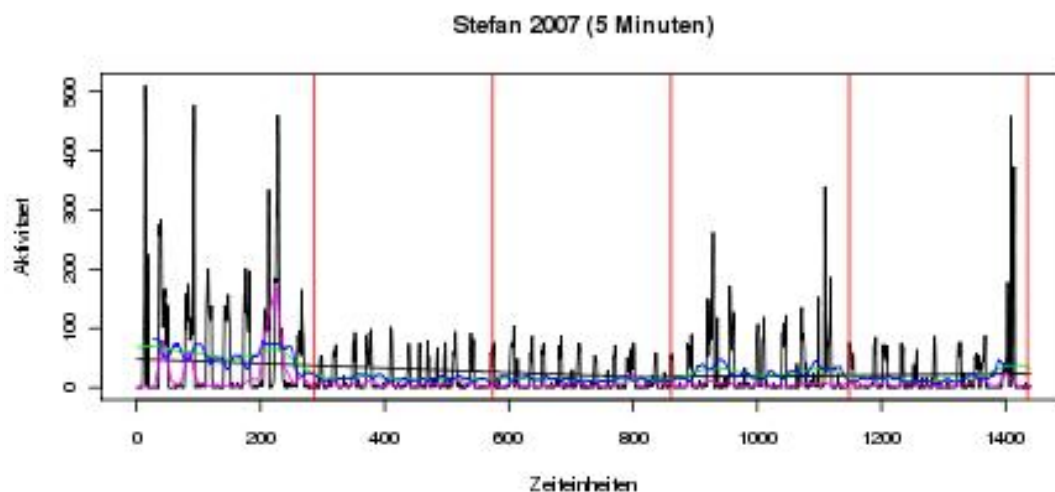


Abbildung 3.3: Glättungsverfahren

Bei diesen Untersuchungen stellt sich sofort die Frage, ob die Autokorrelationskoeffizienten über das ganze Jahr die gleichen Werte zeigen, oder anders formuliert, ob die sich wiederholenden Muster über die verschiedenen Jahreszeiten keine Veränderungen aufweisen.

Die graphische Darstellung der Daten aus einer anderen Periode des Jahres 2007 (20. Juli bis 24. Juli) zeigt in Abbildung 3.4 im Allgemeinen höhere Aktivitätswerte als die vorher für Ende März gezeigten Daten. Auch hier kann man nur sehr vage eine Regelmässigkeit

der Aktivität bei jedem Tag erkennen. Für diese Tage sind die höchsten Aktivitätswerte am Ende und am Anfang des Tages zu beobachten, wobei diese Werte jedesmal eine relativ lange Zeitspanne in Anspruch nehmen. Im Gegensatz dazu zeigt sich im mittleren Teil des Tages eine relative Abwesenheit von Bewegung.

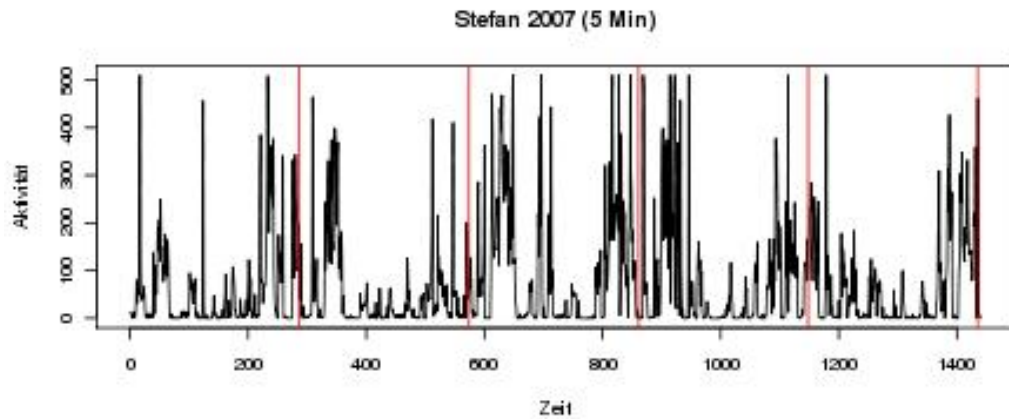


Abbildung 3.4: Tage vom 20. Juli bis zum 24. Juli

Abbildung 3.5 zeigt das Autokorrelogramm für diese Sommertage mit besonders hohen Aktivitätswerten. Dabei kann man sofort einen repetitiven Vorgang erkennen, der aber eine andere Gestalt aufweist, als die für die Aktivität Anfang März (vgl. Abbildung 3.2).

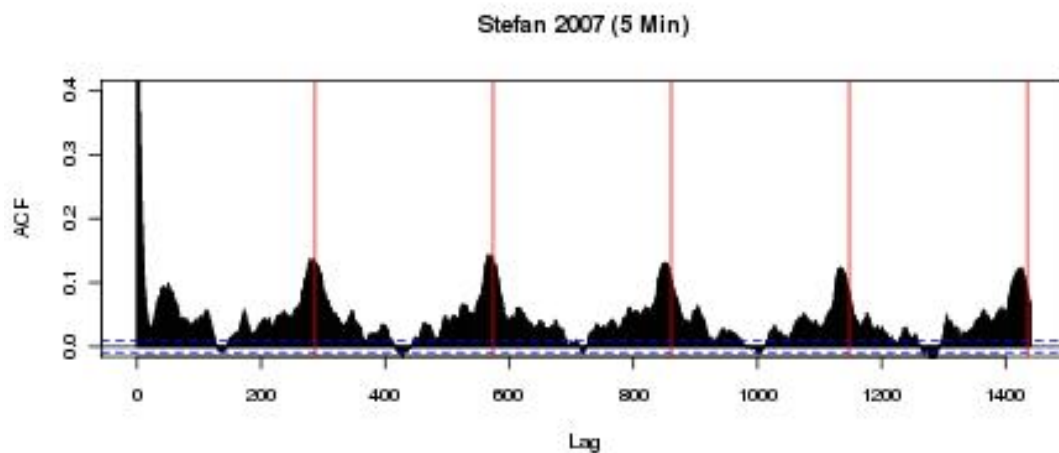


Abbildung 3.5: Korrelogramm für die Aktivität von Stefan mit  $\text{Lag} = 1440$

Wir wollen nun eine interessante systematische Änderung der Aktivität von Reh Stefan



in einem späteren Jahresabschnitt betrachten.

Abbildung 3.6 zeigt die Zeitreihe für die Tage vom 28. Oktober bis zum 1. November 2007. Wir finden hier auch Anhebungen, die durch mehr oder weniger regelmässige Abstände getrennt sind. Die Anhebungen haben aber fast immer die gleiche Höhe, die im mittleren Bereich der Aktivitätswerte liegt.

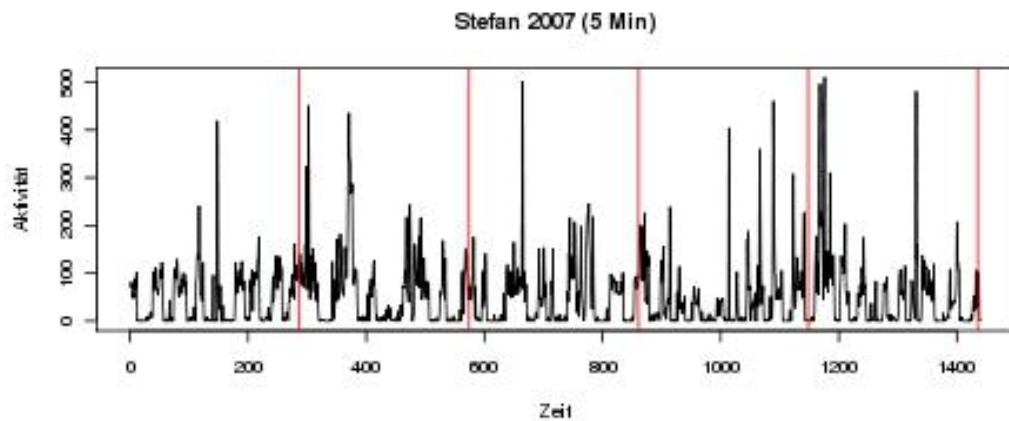


Abbildung 3.6: Tage vom 28. Oktober bis zum 1. November 2007

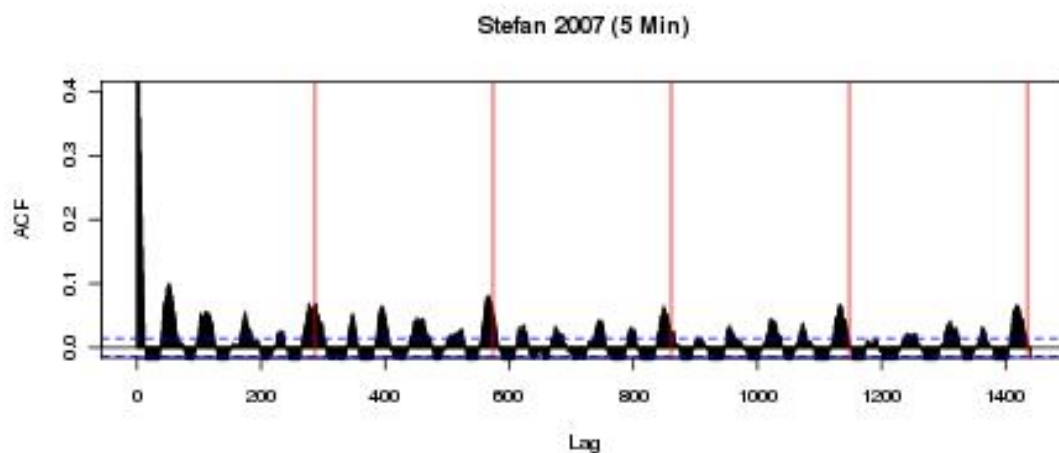


Abbildung 3.7: Korrelogramm für die Aktivität von Stefan mit Lag = 1440

Im Korrelogramm 3.7 kann man sofort ein regelmässiges Muster erkennen, dieses unterscheidet sich jedoch erheblich von denen die den Tagen im Frühling (Abbildung 3.2) und im Sommer (Abbildung 3.5) entsprechen. Es fällt auf, dass negative Korrelationen sich in einem gleichmässigen Rythmus mit positiven abwechseln. Innerhalb der Tage sind mehrere Erhebungen zu sehen, die fast alle gleich hoch sind.

Bei der Erstellung von Modellen für die Aktivität (vgl. 4.1) werden wir auf diese Änderung der Aktivität im letzten Jahresabschnitt zurückkommen.

Um dieses Kapitel abzuschliessen sei noch erwähnt, dass die Visualisierung der auf 15 Minuten aggregierten Daten fast kein Unterschied zu denen im 5 Minuten-Abstand erkennen lässt.

Abbildung 3.8 zeigt die Aktivitätswerte aus dem Datensatz REHE15 für den gleichen Zeitraum wie Abbildung 3.1 aus dem Datensatz REHE5. Man kann hier sowohl die gleiche Anzahl von Erhebungen erkennen als auch deren Positionierung. Diese Tatsache lässt die Vermutung aufkommen, dass die Arbeit mit beiden Datensätzen zu sehr ähnlichen Ergebnissen führen wird.

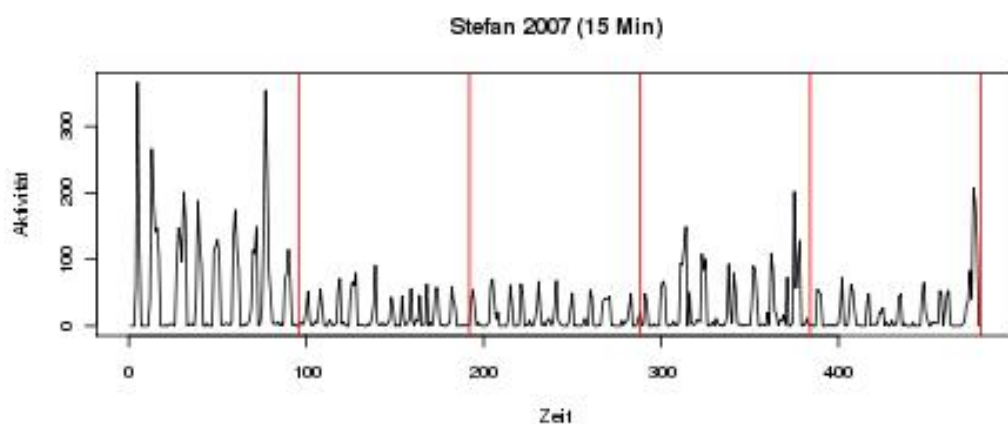


Abbildung 3.8: Tage vom 23. März bis zum 27. März 2007

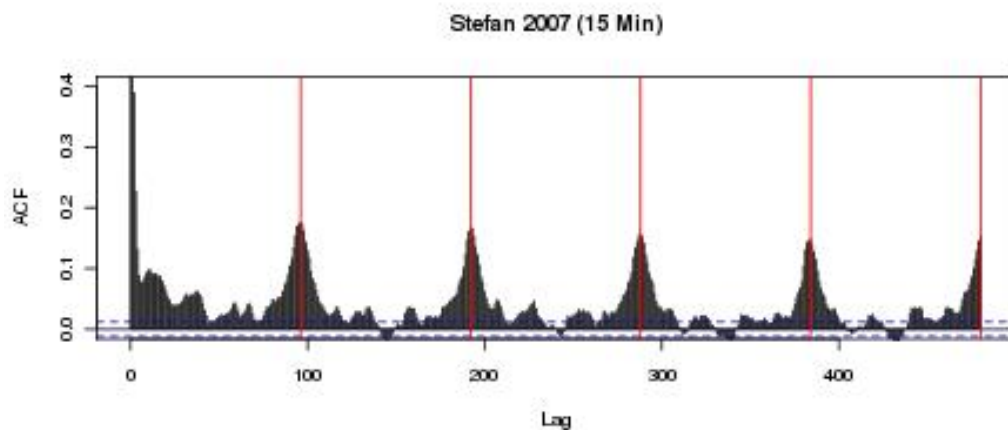


Abbildung 3.9: Korrelogramm für die Aktivität von Stefan mit Lag = 480

Das Korrelogramm in Abbildung 3.9 ist mit dem Korrelogramm in Abbildung 3.2 fast identisch.

Bei dieser deskriptiven Analyse haben wir uns ausschliesslich auf die Daten eines einzelnen Rehes konzentriert. Eine Anwendung der gleichen Techniken auf die anderen Rehe zeigt ähnliche Ergebnisse. Der Hauptunterschied besteht darin, dass die Muster der Autokorrelationsfunktionen je nach Reh andere Gestalten annehmen können. Die Periodizität der Tage bleibt aber immer erhalten.

Die Verwendung der empirischen Autokorrelationsfunktion erwies sich als sehr nützlich bei der Entdeckung der täglichen Periodizität der Zeitreihen der Aktivitätswerte. Auch die Änderung dieser Periodizität in Abhängigkeit von der Jahreszeit wurde hervorgehoben. Wir ziehen aus diesen Kenntnissen den Schluss, dass es notwendig ist, die Wechselwirkung zwischen dem Tages- und dem Jahresverlauf zu modellieren.

Der Tagesverlauf wird im Datensatz REHE5 durch die Variable *minuten* erfasst, im Datensatz REHE15 durch die Variable *m15*. Dem Jahresverlauf entspricht in beiden Datensätzen die Variable *tag*.

## 4 Modelle

In diesem Teil werden wir die Ergebnisse der vorangegangenen Untersuchungen einbeziehen, um ein Modell für die Aktivität von Rehen zu erhalten. Aus dem Praktikumsbericht entnehmen wir die Erkenntnis, dass sich die Aktivität zwischen Geschlechtern signifikant unterscheidet. Deshalb gruppieren wir hier die Rehe nach Geschlecht und erstellen zwei getrennte Modelle.

Wir sind davon ausgegangen, dass sowohl der Tagesablauf als auch der Jahresablauf einen mit Sicherheit nicht linearen Einfluss auf die Aktivität ausüben. Die Untersuchung der empirischen Autokorrelationsfunktion der Aktivitätswerte für verschiedene Rehe in beiden Geschlechtskategorien legte uns nahe, dass zwischen den beiden Einflussgrößen, nämlich *minuten* bzw. *m15* und *tag*, Interaktionseffekte auftreten müssen.

Ein erstes Modell zur Beschreibung solcher Interaktionseffekte kann folgendermassen aufgestellt werden [Fahrmeir u. a. 2007]:

Für die 5 Minuten-Auflösung:

$$akt = f(minuten, tag) + \epsilon \quad (4.1)$$

und für die 15 Minuten-Auflösung:

$$akt = f(m15, tag) + \epsilon \quad (4.2)$$

mit  $\epsilon \sim N(0, \sigma^2)$ .

Die Wechselwirkung der beiden metrischen Kovariablen wird in nichtparametrischer Form durch additives Einbeziehen einer nicht näher spezifizierten glatten zweidimensionalen Funktion  $f(minuten, tag)$  berücksichtigt. Der Effekt zwischen den Kovariablen wird dabei ohne die Aufnahme von eindimensionalen Haupteffekten modelliert.

Für die Schätzung der zweidimensionalen Interaktion verwenden wir einen Gradient Boosting- Algorithmus [Bühlmann und Hothorn 2007]. Einige Eigenschaften dieses Algorithmus werden im nächsten Abschnitt betrachtet.

## 4.1 Boosting

Boosting ist ein Algorithmus, der ursprünglich für die Lösung von Klassifizierungsproblemen entworfen wurde, aber immer häufiger als Regressionsmethode eingesetzt wird. Im Folgenden geben wir eine schematische Beschreibung des Boosting-Algorithmus *Ada-Boost.M1*, der aus den Arbeiten von Freund und Schapire stammt [Hastie u. a. 2009]. Die Hauptidee dieses Algorithmus besteht in der Verschmelzung von mehreren „schwachen“ Klassifikatoren zu einem einzigen „guten“. Shapire zeigte, dass die Leistung eines schwachen Lernalgorithmus immer verbessert werden kann, und zwar durch die Anwendung zweier zusätzlicher Klassifikatoren auf modifizierte Versionen der Eingabedaten. Ein schwacher Lerner ist ein Algorithmus, der einen Klassifikator für zwei Klassen erzeugt, dessen Leistung viel besser ist, als das Werfen einer Münze. Der Algorithmus wendet den „schwachen“ Klassifikator auf modifizierte Versionen der Daten an, um eine Reihe  $G_m(x), m = 1, 2, \dots, M$  von „schwachen“ Klassifikatoren zu erzeugen. Die Vorhersagen aus diesen Klassifikatoren werden zusammengefasst, um die endgültige Optimierung zu erhalten.

$$G(x) = \text{sign}\left(\sum_{m=1}^M \alpha_m G_m(x)\right) \quad (4.3)$$

Der Algorithmus berechnet die  $\alpha_1, \alpha_2, \dots, \alpha_M$  und erwägt deren Einfluss bezüglich  $G_m$ . Ihre Auswirkung besteht darin, den genauesten Klassifikatoren einen besseren Platz in der Reihenfolge einzuräumen.

Der Algorithmus wird in [Hastie u. a. 2009] folgendermassen beschrieben:

1. Initialize the observation weights  $w_i = 1/N, i = 1, 2, \dots, N$
2. For  $m = 1$  to  $M$  :
  - a) Fit a classifier  $G_m(x)$  to the training data using weights  $w_i$
  - b) Compute

$$err_m = \frac{\sum_{i=1}^N w_i I(y_i \neq G_m(x_i))}{\sum_{i=1}^N w_i}$$

- c) Compute  $\alpha_m = \log((1 - err_m)/err_m)$
  - d) Set  $w_i \leftarrow w_i \exp \cdot (\alpha_m \cdot I(y_i \neq G_m(x_i)))$ ,  $i = 1, 2, \dots, N$
3. Output  $G_x = \text{sign}(\sum_{m=1}^M \alpha_m G_m(x))$

Spätere Untersuchungen dehnten den Einsatzbereich der Boosting-Algorithmen auf andere Gebiete aus, die nichts mit der Klassifizierung zu tun haben, wie die Regression.

Für unsere Modelle verwenden wir eine Implementierung von Gradient Boosting, die im R-Package *mboost* zur Verfügung steht.

Nach [Hothorn u. a. 2009] kann Boosting als ein Gradientenverfahren interpretiert werden, welches die Lösung des Optimierungsproblems

$$\eta^*(\mathbf{z}) = \underset{\eta(\mathbf{z})}{\operatorname{argmin}} \mathbb{E}(\rho(\mathbf{y}, \eta(\mathbf{z})))$$

sucht, wobei  $\rho(., .)$  eine angemessene Loss-Funktion, wie z.B.  $\rho(y, \eta) = 0.5|y - \eta|^2$  oder die (log) -likelihood Funktion, ist. Nach der Initialisierung der Funktion mit einem angemessenen Startwert  $\hat{\eta}^{[0]}$  berechnet der Boosting-Algorithmus den negativen Gradienten:

$$u_i = -\frac{\partial}{\partial \eta} \rho(y_i, \eta) \Big|_{\eta = \hat{\eta}^{[m-1]}(z_i)}, i = 1, \dots, n$$

und fittet einen Base-Lerner  $g$  zu  $u = (u_1, \dots, u_n)'$

Für die Modellierung der Interaktionsoberfläche werden Base-Lerner, die auf der Basis von penalisierten Splines funktionieren, auf zwei Dimensionen übertragen.

Die für die Modellierung verwendete Funktion *mboost* implementiert ein Gradientenverfahren für die Optimierung beliebiger Loss-Funktionen, wobei komponentenweise Modelle als Base-Lerner benutzt werden.

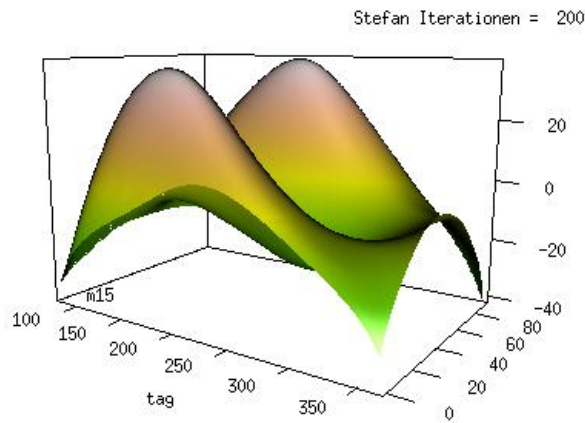
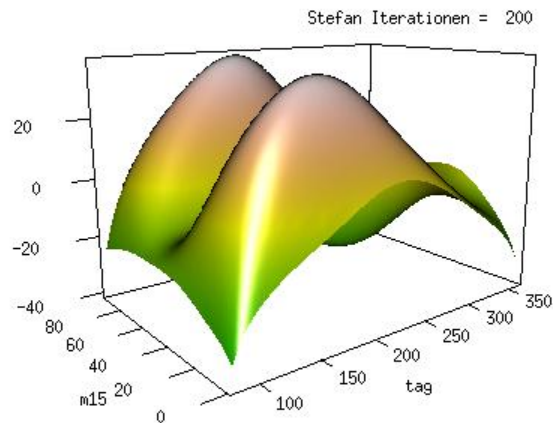
Die Interaktionsoberflächen wurden mit Hilfe der Funktion *bspatial* geschätzt. Diese Funktion implementiert bivariate Tensorprodukt P-Splines.

Das Modell mit zufälligem Effekt verwendet zusätzlich die Funktion *brandom*, die einen Intercept für jedes Reh fittet.

Die optimale Anzahl von Iterationen wird mit Hilfe von *AIC* und *mstop* ermittelt. Aufgrund der grossen Datenmenge war es nicht möglich, diese Funktionen bei unserer Untersuchung zu verwenden.

## 4.2 Schätzung der zweidimensionalen Interaktion

Um an die Überlegungen des letzten Abschnittes anzuknüpfen, zeigen wir die geschätzte zweidimensionale Funktion des Tages- und Jahresablaufs für das Reh Stefan im Jahr 2007.

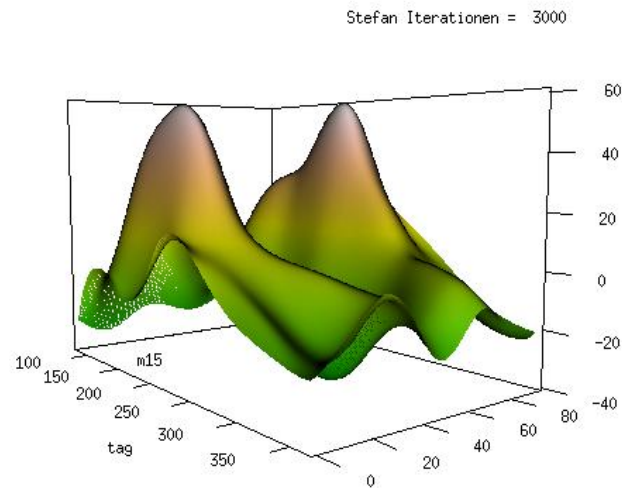


Bei der Betrachtung der drei Zeitabschnitte (Tage 82 bis 86, 200 bis 204, 300 bis 305)<sup>1</sup>, mit denen wir uns bei der deskriptiven Datenanalyse beschäftigt haben, kann man feststellen, dass die Interaktionsoberfläche in diesen Umgebungen eine gleichförmige Gestalt annimmt. Dies bedeutet, dass in Übereinstimmung mit den sich wiederholenden Mustern der Autokorrelationsfunktion die Aktivität eine Zeit lang ähnliche Werte aufweist, bis sich ein anderes Muster manchmal plötzlich, manchmal allmählich abzeichnet. In den letzten Tagen des Jahres, ungefähr ab dem Tag 300, kann man eine plötzliche Änderung der Form der Oberfläche sehen. Diese entspricht der Veränderung der Muster in der Autokorrelationsfunktion von Abbildung 3.7. Es handelt sich dabei um eine bemerkenswerte Veränderung des Aktivitätsverlaufs.

Das Korrelogramm für die Tage vom 28. Oktober bis zum 1. November 2007 3.7 legte die

<sup>1</sup>Die Kodierung der Tage ist im Anhang A.4 zu finden.

Wahl einer sinusförmigen Funktion für die Beschreibung des Tagesverlaufs nahe. Genau diese Gestalt ist nach 3000 Iterationen in der Interaktionsoberfläche für den genannten Zeitraum zu erkennen.



Einer der grössten Vorzüge bei der Verwendung einer Interaktionsoberfläche für die Visualisierung der Aktivität besteht in der Möglichkeit, die Eigentümlichkeiten jedes einzelnen Tieres in einem bestimmten Zeitraum anzusehen. Das Erscheinen von Oberflächen mit stark abweichenden Gestalten könnte auf einen Krankheitszustand oder auf starke Umwelteinflüsse hinweisen [Shepard u. a. 2008].

### 4.3 Additive Modelle für Männchen und Weibchen

Als nächster Schritt wollen wir das additive Modell 4.2 auf alle Rehe anwenden. Zuerst erfolgt die Schätzung der Interaktion für die Weibchen. Wir verwenden den Datensatz REHE15. Dabei setzen wir die Anzahl der Iterationen auf 200 und die Anzahl der Knoten auf 10. Abbildung 4.1a zeigt das Konturdiagramm für alle Weibchen.



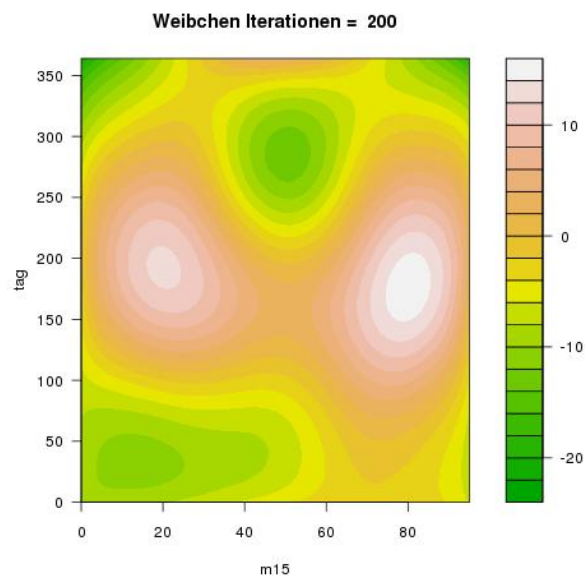


Abbildung 4.1a: Konturdiagramm für alle Weibchen.

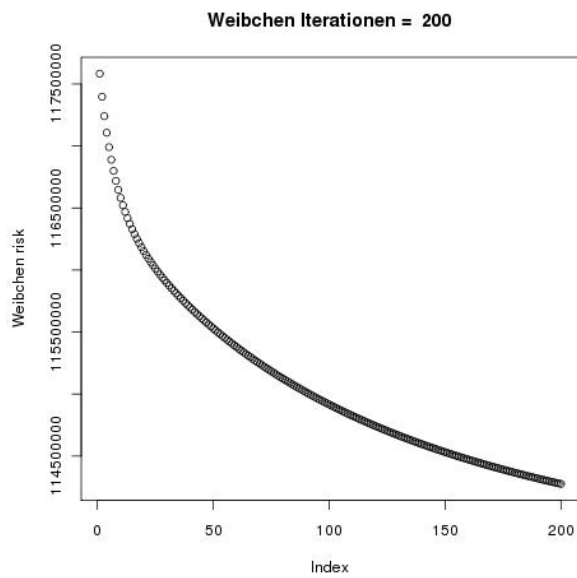


Abbildung 4.1b: Risk

Die dazu gehörige Interaktionsoberfläche wird in den folgenden Abbildungen gezeigt. Die im Konturdiagramm sich abzeichnenden Aktivitätsmaxima treten auch hier deutlich zum Vorschein.

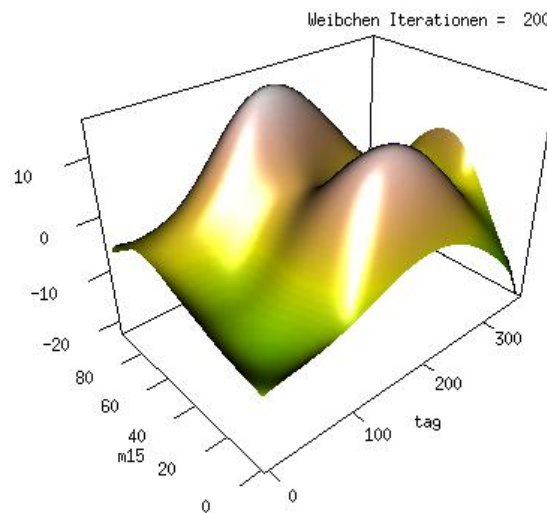


Abbildung 4.2: Weibchen 15 Minuten

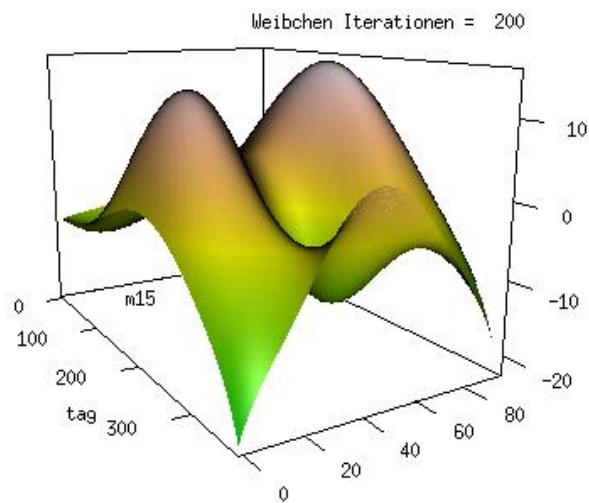


Abbildung 4.3: Weibchen 15 Minuten

Das Modell wurde auch mit 1000 und 3000 Iterationen gerechnet. Nach 1000 Iterationen bleibt die Grundgestalt der Oberfläche erhalten, der Tagesverlauf erscheint jedoch, vor allem am Jahresanfang, nicht so flach wie in [Abbildung 4.2](#). Ungefähr bei Tag 50 ist

eine bemerkenswerte Vertiefung in den ersten Stunden des Tages zu sehen. Dies lässt sich auch im Konturdiagramm erkennen. Die Risikofunktion verläuft weiterhin fallend, so dass man nicht sagen kann, die optimale Anzahl von Iterationen sei erreicht worden.

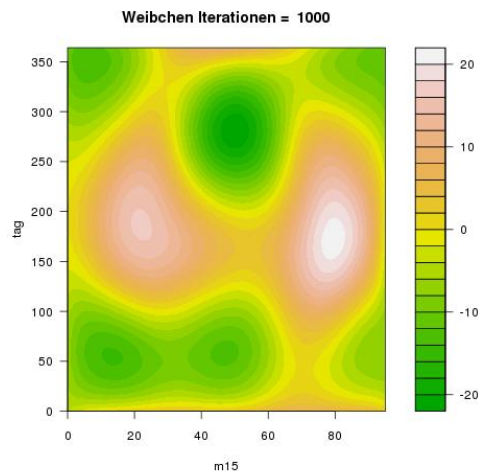


Abbildung 4.4: Weibchen 15 Minuten

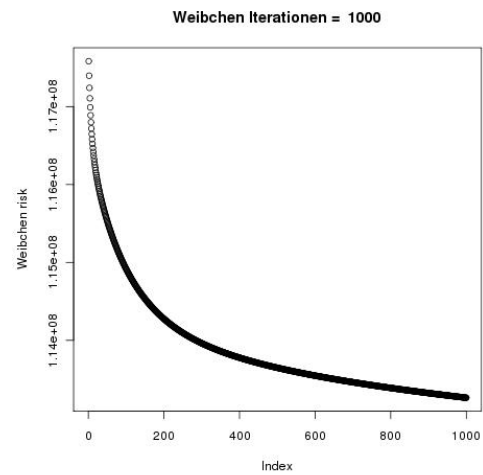


Abbildung 4.5: Risk

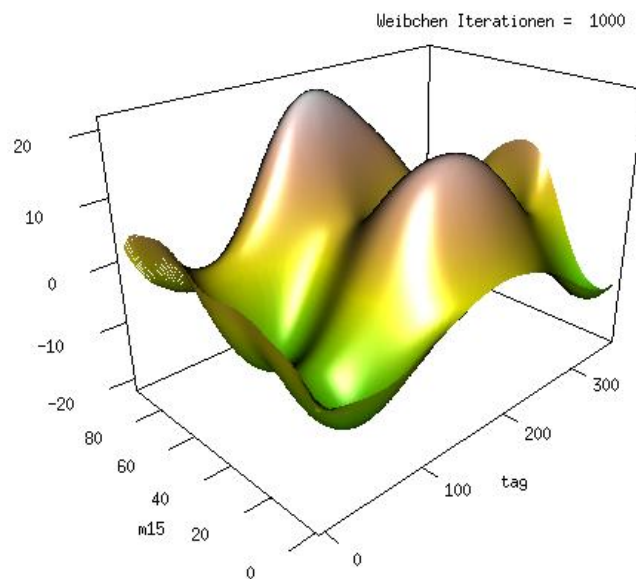


Abbildung 4.6: Weibchen 15 Minuten

Nach 3000 Iterationen erscheint der Tagesverlauf welliger, wobei zwei ausgeprägte Erhebungen zu sehen sind, die der Morgen- bzw. Abenddämmerung entsprechen. Im Bereich zwischen den Tagen 150 und 250 und zwischen den Viertelstunden 30 und 70 kann man

einen Aktivitätszuwachs in der Interaktionsoberfläche deutlich erkennen.

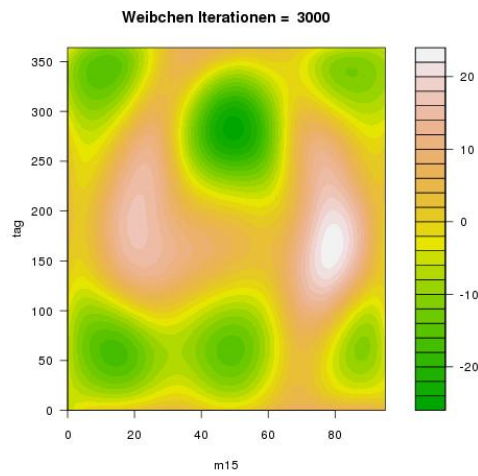


Abbildung 4.7: Weibchen 15 Minuten

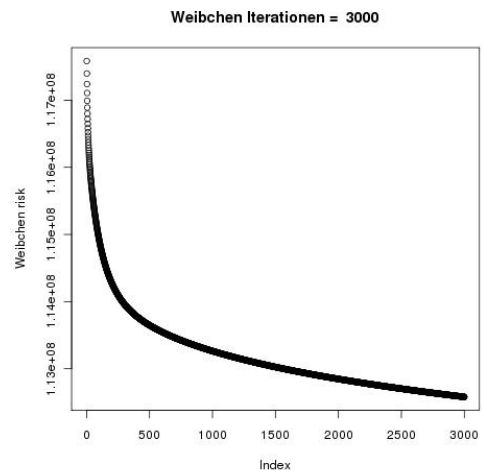


Abbildung 4.8: Risk

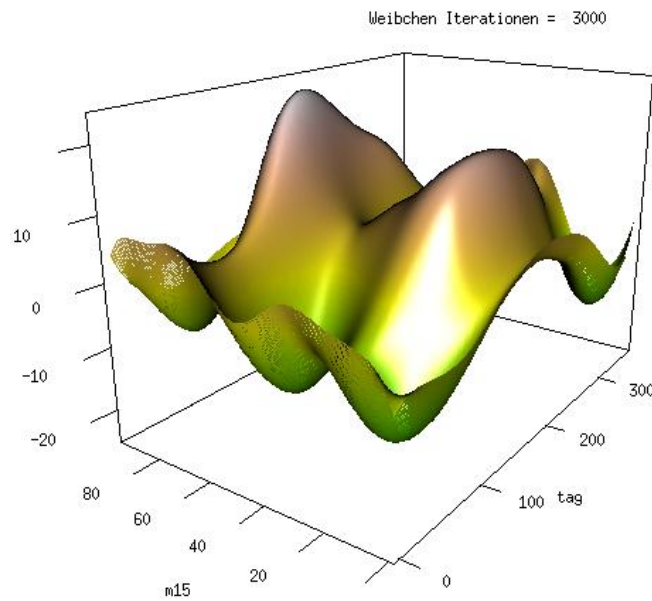


Abbildung 4.9: Weibchen 15 Minuten

Im Folgenden zeigen wir die Ergebnisse der additiven Modellierung mit 200 Iterationen für alle Männchen in 5 bzw. 15 Minuten-Auflösung. Beide Konturdiagramme lassen eine Symmetrie erkennen, die bei den Weibchen nicht vorhanden war.

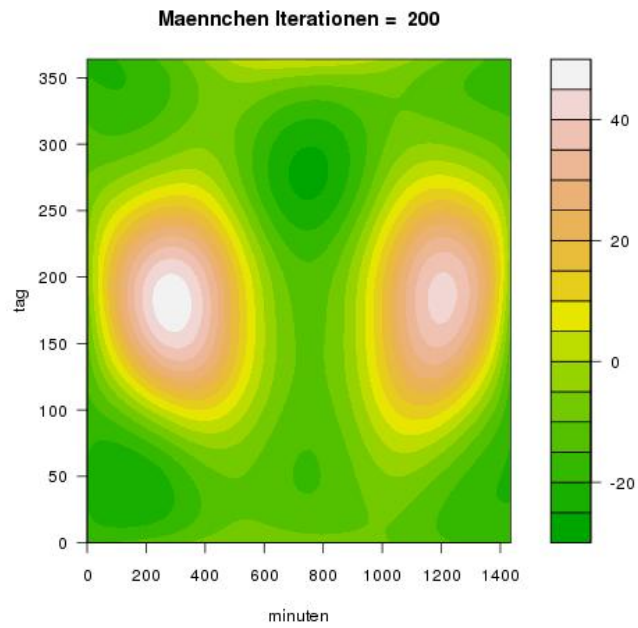


Abbildung 4.10a: Männchen 5 Minuten

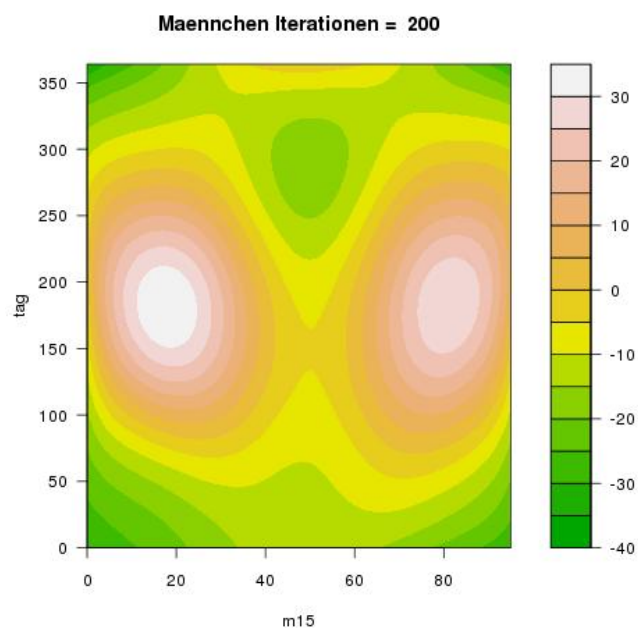


Abbildung 4.10b: Männchen 15 Minuten

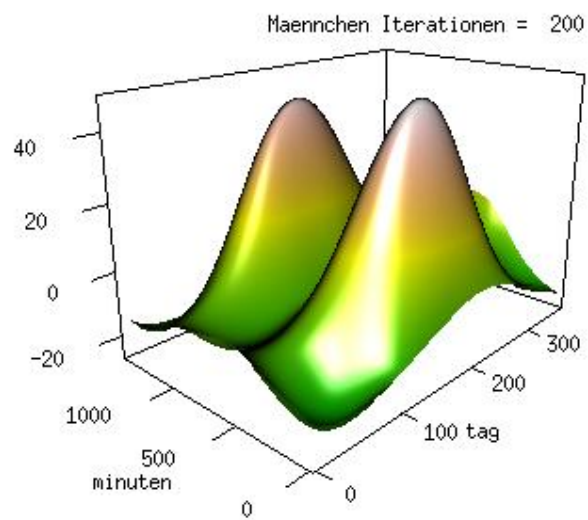


Abbildung 4.11a: Männchen 5 Minuten

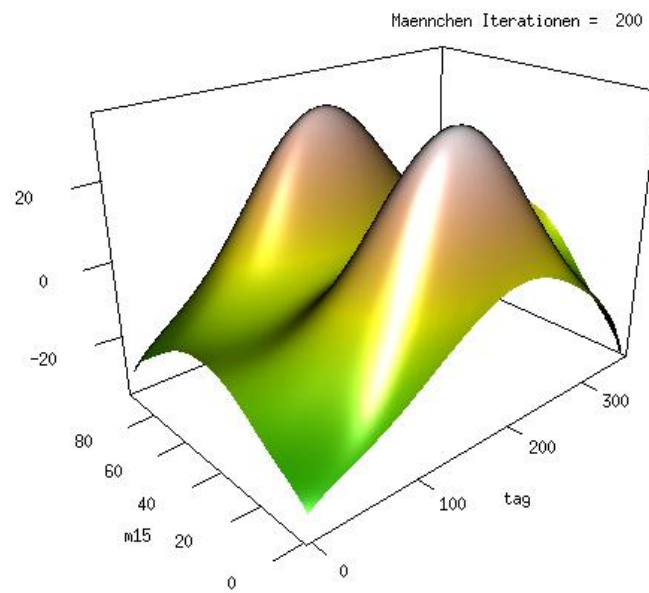


Abbildung 4.11b: Männchen 15 Minuten

Die Unterschiede, die durch die Wahl der Auflösung bei gleicher Anzahl von Iterationen entstehen, waren bei den Weibchen nicht so ausgeprägt. Der Betrachtung der Interaktionsoberflächen kann man entnehmen, dass die Aktivitätswerte für die Männchen insgesamt höher als die für die Weibchen sind. Darüber hinaus ist der Verlauf hin zu den Maxima steiler. Dies ergibt eine für die Männchen charakteristische Gestalt, die sich im Gegensatz zu der der Weibchen auch mit zunehmender Anzahl von Iterationen nicht wesentlich ändert.

## 4.4 Additives Modell mit zufälligem Effekt

Bei den Modellen 4.1 und 4.2 wurden die Aktivitätswerte in den aufgezeichneten Jahren so zusammengefasst, dass keine individuellen Effekte auf die Zielgrösse betrachtet wurden. Da gerade diese Betrachtung für uns von Interesse ist, erweitern wir das additive Modell um einen Term, der diese Effekte mit einbezieht. Dabei wird angenommen, dass die zufälligen Effekte unabhängig und identisch verteilt sind.

Damit erhalten wir folgendes additives Modell mit zufälligem Effekt [Hothorn u. a. 2009]:

$$akt = f(minuten, tag) + random(name) \quad (4.4)$$

Bei diesem Modellierungsansatz wird davon ausgegangen, dass die Form der Interaktionsoberfläche für alle Individuen die gleiche ist. Die individuellen Unterschiede werden durch einen Intercept erfasst, d.h. durch eine Verschiebung der Oberfläche nach oben bzw. nach unten. Wir schätzen wieder getrennte Modelle für Männchen und Weibchen.

### 4.4.1 Modell für die Weibchen

Wir betrachten zuerst das Modell mit 200 Iterationen im 5 Minuten-Abstand. Die Schätzungen der Intercepte für Elfriede, Judith, Olympia und Resi sind negativ, d.h. diese Rehe zeigen ein niedrigeres Aktivitätsniveau im Vergleich zur Oberfläche für alle Weibchen.

Name	Intercept
Elfriede	-0.1356880
Helene	1.6586635
Herta	2.9523744
Judith	-12.5356056
Lara	13.5560280
Olga	0.8781106
Olympia	-4.2047456
Oxana	2.3886081
Resi	-4.3017170

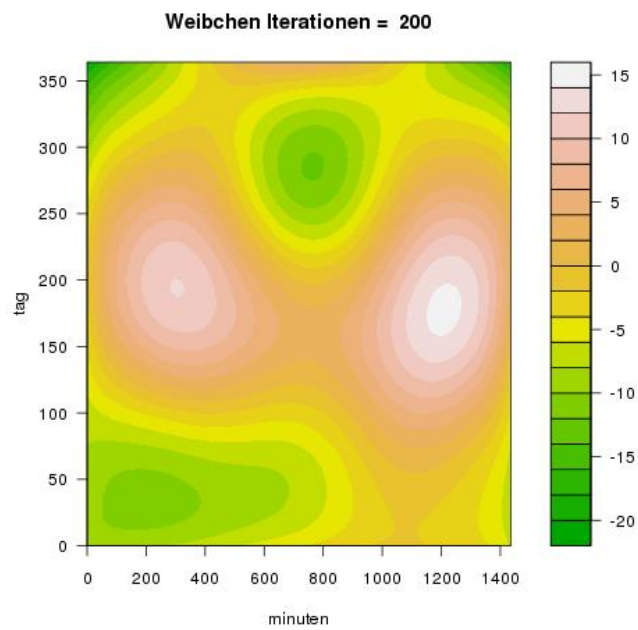


Abbildung 4.12a: Weibchen 5 Minuten mit zufälligem Effekt

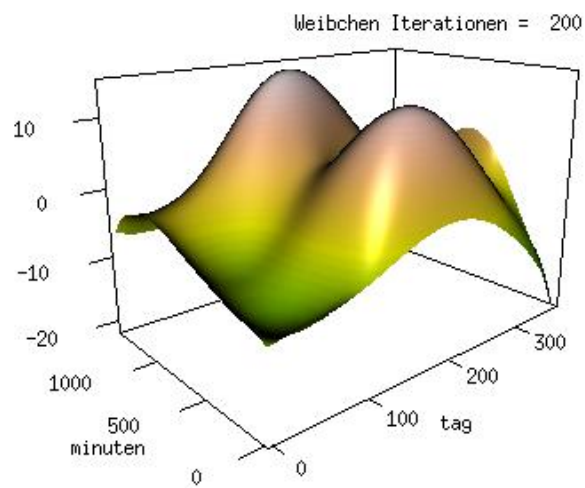


Abbildung 4.12b: Weibchen 5 Minuten mit zufälligem Effekt



Judith mit einen Intercept von -12.5356056 und Lara mit einem von 13.5560280 zeigen die grössten Abweichungen vom mittleren Aktivitätsniveau.

Sowohl das Konturdiagramm als auch die Interaktionsoberfläche unterscheiden sich kaum von denen aus dem Modell ohne zufälligen Effekt. Das Modell mit 2000 Iterationen ergibt:

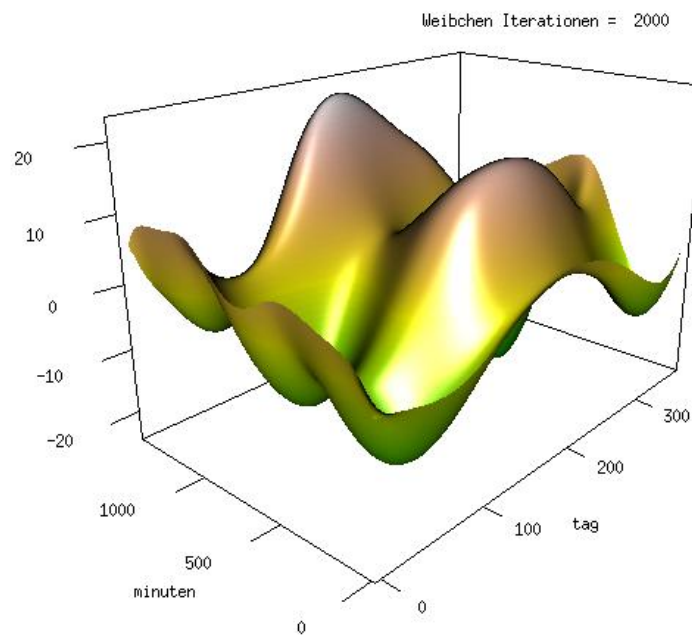


Abbildung 4.13: Weibchen 5 Minuten mit zufälligem Effekt

Schätzungen	
Name	Intercept
Elfriede	0.08074343
Helene	1.94301262
Herta	3.60609170
Judith	-13.61430895
Lara	14.99066869
Olga	0.47370383
Olympia	-4.54544734
Oxana	2.83497960
Resi	-5.70519469

Mit Ausnahme von Elfriede bleiben die Vorzeichen der Intercepte erhalten, so dass man von stabilen Schätzungen für die individuellen Aktivitätsniveaus ausgehen kann.

#### 4.4.2 Modell für die Männchen

Wir betrachten in diesem Abschnitt das Modell mit zufälligem Effekt für die Männchen. Wie bei den Ausführungen zum additive Modell ohne zufälligen Effekt zeigen wir zuerst die Ergebnisse für die 5 Minuten-Auflösung. Die grössten Abweichungen, beide in positiver Richtung, sind bei Erich und Oleg zu sehen. Unter den Rehen mit negativem Intercept zeigt Heiner das niedrigste Aktivitätsniveau.

Schätzungen für 5 Minuten	
Name	Intercept
Eddy	-6.4066063
Erich	13.6089303
Georg	-1.0043275
Hansi	-0.6306767
Heiner	-8.5105134
Heribert	-2.7753414
Holger	4.3495831
Jakob	-6.5378737
Jasper	-2.1428956
Joe	6.4420015
Leo	-5.6711652
Norbert	0.5679298
Odilo	4.6822730
Oleg	13.6385723
Ralph	-2.8531231
Richard	3.0024483
Rokko	-6.4485779
Roland	-4.8560705
Rudi	-3.4400555
Stefan	2.7311967

Folgende Abbildungen zeigen das Konturdiagramm und die Interaktionsoberfläche für dieses Modell. Beide sind denen sehr ähnlich, die zum Modell ohne zufälligen Effekt gehören.

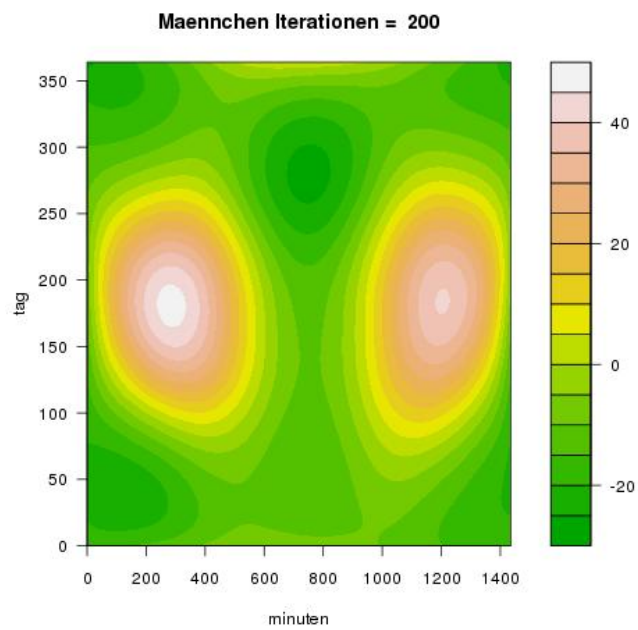


Abbildung 4.14a: Männchen 5 Minuten mit zufälligem Effekt

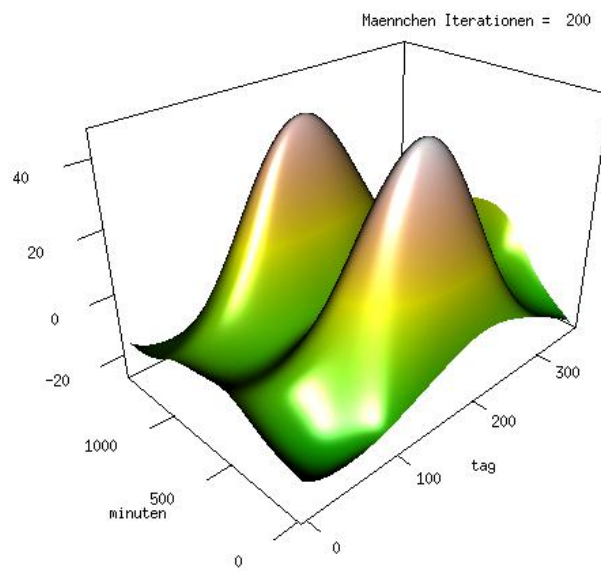


Abbildung 4.14b: Männchen 5 Minuten mit zufälligem Effekt

Die Schätzungen für die 15 Minuten-Auflösung lauten:

Schätzungen für 15 Minuten	
Name	Intercept
Eddy	-5.3595033
Erich	12.0037022
Georg	-1.2180024
Hansi	-1.2367285
Heiner	-7.8323939
Heribert	-2.9676344
Holger	4.8363592
Jakob	-6.5158948
Jasper	-1.9370775
Joe	5.5245916
Leo	-3.6997871
Norbert	-0.1353533
Odilo	4.1294171
Oleg	12.4962427
Ralph	-2.7576937
Richard	3.0725692
Rokko	-5.2510255
Roland	-3.9961271
Rudi	-3.2793552
Stefan	2.4635982

Norbert ist das einzige Reh, für welches sich das Vorzeichen des Intercepts ändert. Der Betrag des Schätzers ist aber im Vergleich zu den anderen klein. Aus dem Modell mit 1000 Iterationen ergeben sich keine wesentlichen Unterschiede bei den Schätzern.

Zum Schluss zeigen wir die graphischen Ergebnisse für 200 und für 1000 Iterationen.

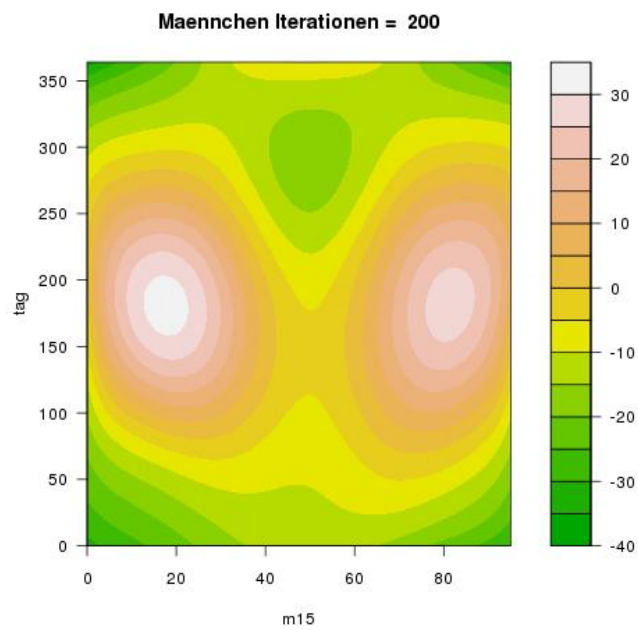


Abbildung 4.15a: Männchen 15 Minuten mit zufälligem Effekt

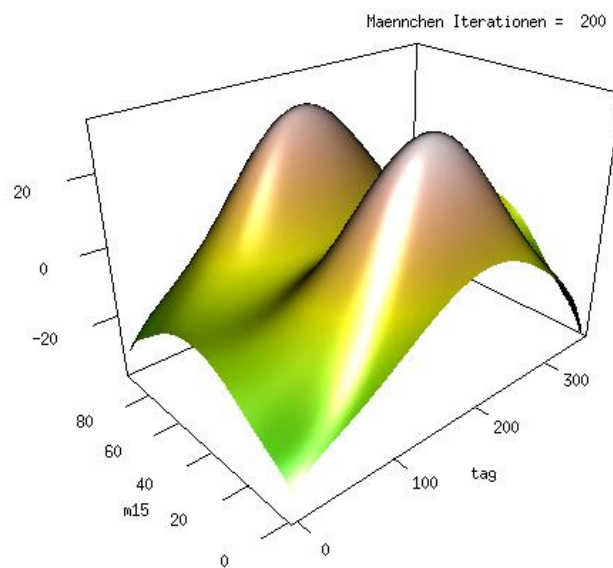


Abbildung 4.15b: Männchen 15 Minuten mit zufälligem Effekt

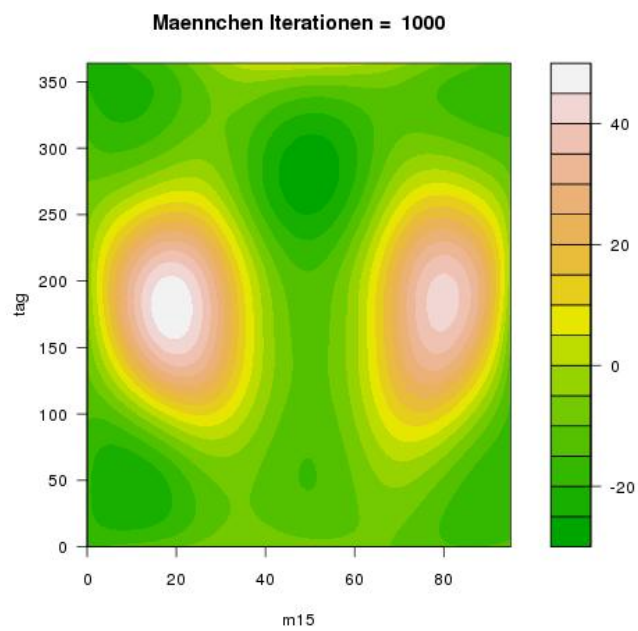


Abbildung 4.16a: Männchen 15 Minuten mit zufälligem Effekt

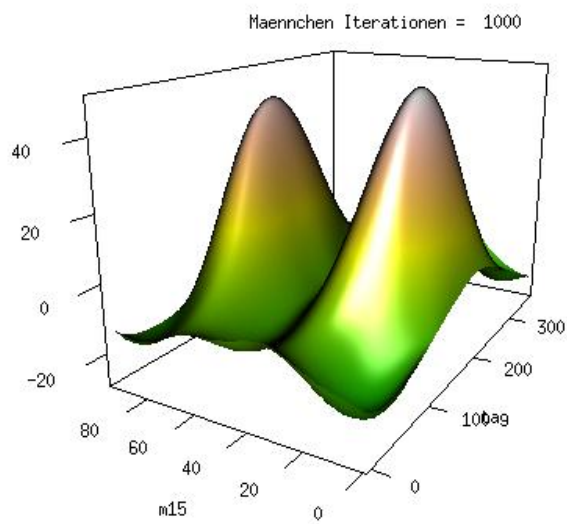


Abbildung 4.16b: Männchen 15 Minuten mit zufälligem Effekt

## 5 Fazit

Unser primäres Ziel war die Auffindung von Bewegungsmustern aus den Aktivitätsdaten der Wildrehe. Zu diesem Zweck gruppieren wir die Tiere nach Geschlecht und konzentrierten uns auf die Ausarbeitung zweier verschiedener Modelle, welche die wesentlichen Merkmale von jedem Geschlecht enthalten sollten. Unter Berücksichtigung der Interaktion zwischen dem Tages- und Jahresablauf haben wir additive Modelle für die Aktivitätsdaten erstellt. Die erhaltenen Interaktionsoberflächen zeigten verschiedene charakteristische Gestalten für Weibchen und Männchen. Dies eröffnet die Möglichkeit einer differenzierteren Analyse der Aktivität im Zusammenhang mit Kenntnissen im Bereich der Biologie des Rehes und vielleicht auch mit anderen Kovariablen. Mit zunehmender Anzahl von Iterationen zeichneten sich bei den Weibchen verschiedene Gestalten der Oberfläche ab. Dies war bei den Männchen nicht der Fall. Darüber hinaus konnte festgestellt werden, dass zwischen den Interaktionsoberflächen, die aus den zwei verschiedenen Datensätzen REHE15 und REHE5 entstehen, zwar Unterschiede existieren, die Tendenz der Aktivität jedoch stabil bleibt.

# Literaturverzeichnis

## Alexander 1982

ALEXANDER, McNeill: *Locomotion of animals*. First Edition. Glasgow : Blackie et Son Limited, 1982 [2](#), [2.2](#)

## Bühlmann und Hothorn 2007

BÜHLMANN, Peter ; HOTHORN, Torsten: Boosting Algorithms: Regularization, Prediction and Model Fitting. In: *Statistical Science* (2007) [4](#)

## Chatfield 2004

CHATFIELD, Chris: *The Analysis of Time Series. An Introduction*. Sixth Edition. Boca Raton : CRC Press, 2004 [3](#)

## Cryer und Chan 2008

CRYER, Jonathan D. ; CHAN, Kung-Sik: *Time Series Analysis with Applications in R*. Second Edition. New York : Springer, 2008 [3](#)

## Fahrmeir u. a. 2007

FAHRMEIR, Ludwig ; KNEIB, Thomas ; LANG, Stefan: *Regression: Modelle, Methoden und Anwendungen*. 1. Auflage. Berlin : Springer, 2007 [4](#)

## Hastie u. a. 2009

HASTIE, Trevor ; TIBSHIRANI, Robert ; FRIEDMAN, Jerome: *The Elements of Statistical Learning*. Second Edition. New York : Springer, 2009 [4.1](#), [4.1](#)

## Heller u. a. 2009

HELLER, Elizabeth ; MAIERHOFER, Sarah ; MANUILOVA, Ekaterina: *Aktivitätsanalyse von Rehen im Nationalpark Bayerischer Wald*. 2009. – Bericht zum statistischen Praktikum, Semesterferien WiSe 2008-2009 [1.1](#)

## Hothorn u. a. 2009

HOTHORN, Torsten ; KNEIB, Thomas ; TUTZ, Gerhard: Variable Selection and Model Choice in Geosadditive Regression Models. In: *Biometrics* (2009) [4.1](#), [4.4](#)

## Shepard u. a. 2008

SHEPARD, Emily L. C. ; WILSON, Rory P. ; QUINTANA, Flavio ; MACDONALD, David W.: Identification of animal movement patterns using tri-axial accelerometry. In: *Inter Research* (2008) [2](#), [2.1](#), [2.1](#), [2.2](#), [2.3](#), [4.2](#)

## Shumway und Stoffer 2006

SHUMWAY, Robert H. ; STOFFER, David S.: *Time Series Analysis and Its Applications With R Examples*. Second Edition. New York : Springer, 2006 [3](#)



**Wilson u. a. 2008**

WILSON, Rory P. ; QUINTANA, Flavio ; SHEPARD, Emily L. C.: Identification of imperial cormorant *Phalacrocorax atriceps* behaviour using accelerometers. In: *Inter Research* (2008) [2.1](#), [2.1.2](#)

**Wilson u. a. 2007**

WILSON, Rory P. ; SHEPARD, Emily L. C. ; LIEBSCH, N.: Prying into the intimate details of animal lives: use of a daily diary on animals. In: *Inter Research* (2007) [2.1](#)

## Eidesstattliche Erklärung

Ich versichere hiermit, dass ich meine Bachelorarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

München, den 21. September 2009

---

ELIZABETH HELLER

## A Anhang

## A.1 Verwendete Software

Im Folgenden werden die Programme vorgestellt, die zum Erstellen dieser Arbeit verwendet wurden. Es wurden nur Open-Source-Programme eingesetzt.

R is a language and environment for statistical computing and graphics.

- **Das R Package “mboost”**

Functional gradient descent algorithms (boosting) for optimizing general loss functions utilizing componentwise least squares, either of parametric linear form or smoothing splines, or regression trees as base learners for fitting generalized linear, additive and interaction models to potentially high-dimensional data.

Website: <http://cran.r-project.org/web/packages/mboost/index.html>

- **Die Programmiersprache “ R” version 9.2.1**

R is a language and environment for statistical computing and graphics. R is available as Free Software under the terms of the Free Software Foundation’s GNU General Public License in source code form

Website: <http://www.r-project.org>

- **Das R Package “rgl”**

The rgl package is a visualization device system for R, using OpenGL as the rendering backend. An rgl device at its core is a real-time 3D engine written in C++. It provides an interactive viewpoint navigation facility (mouse + wheel support) and an R programming interface. Website: <http://rgl.neoscientists.org/about.shtml>

- **L<sup>A</sup>T<sub>E</sub>X**

Diese Arbeit wurde mit L<sup>A</sup>T<sub>E</sub>X geschrieben. Als Distribution wurde das Package Tex Live als Editoren Emacs 22 und Texmaker verwendet.

- **GNU/Linux**

Es wurde die Debian Distribution unter der Version Lenny amd64 verwendet.

## A.2 Verwendete Hardware

Für die Erstellung der Modelle wurde ein Rechner mit folgenden Eigenschaften eingesetzt:

- **Acer Aspire X3810**
  - CPU Intel Core 2 Quad Processor
  - RAM 8 GB DDR3 memory

## A.3 Uhrzeiten

Tabelle A.1: Kodierung der Uhrzeit in 5 und 15 Minuten-Abstand

Uhrzeit	minuten	m15	Uhrzeit	minuten	m15	Uhrzeit	minuten	m15
00:00	0	0	08:00	480	32	16:00	960	64
00:05	5	0	08:05	485	32	16:05	965	64
00:10	10	0	08:10	490	32	16:10	970	64
00:15	15	1	08:15	495	33	16:15	975	65
00:20	20	1	08:20	500	33	16:20	980	65
00:25	25	1	08:25	505	33	16:25	985	65
00:30	30	2	08:30	510	34	16:30	990	66
00:35	35	2	08:35	515	34	16:35	995	66
00:40	40	2	08:40	520	34	16:40	1000	66
00:45	45	3	08:45	525	35	16:45	1005	67
00:50	50	3	08:50	530	35	16:50	1010	67
00:55	55	3	08:55	535	35	16:55	1015	67
01:00	60	4	09:00	540	36	17:00	1020	68
01:05	65	4	09:05	545	36	17:05	1025	68
01:10	70	4	09:10	550	36	17:10	1030	68
01:15	75	5	09:15	555	37	17:15	1035	69
01:20	80	5	09:20	560	37	17:20	1040	69
01:25	85	5	09:25	565	37	17:25	1045	69
01:30	90	6	09:30	570	38	17:30	1050	70
01:35	95	6	09:35	575	38	17:35	1055	70
01:40	100	6	09:40	580	38	17:40	1060	70
01:45	105	7	09:45	585	39	17:45	1065	71
01:50	110	7	09:50	590	39	17:50	1070	71
01:55	115	7	09:55	595	39	17:55	1075	71
02:00	120	8	10:00	600	40	18:00	1080	72
02:05	125	8	10:05	605	40	18:05	1085	72
02:10	130	8	10:10	610	40	18:10	1090	72
02:15	135	9	10:15	615	41	18:15	1095	73
02:20	140	9	10:20	620	41	18:20	1100	73
02:25	145	9	10:25	625	41	18:25	1105	73
02:30	150	10	10:30	630	42	18:30	1110	74
02:35	155	10	10:35	635	42	18:35	1115	74
02:40	160	10	10:40	640	42	18:40	1120	74
02:45	165	11	10:45	645	43	18:45	1125	75
02:50	170	11	10:50	650	43	18:50	1130	75

*Fortsetzung auf der nächsten Seite*

<b>Uhrzeit</b>	<b>minuten</b>	<b>m15</b>	<b>Uhrzeit</b>	<b>minuten</b>	<b>m15</b>	<b>Uhrzeit</b>	<b>minuten</b>	<b>m15</b>
02:55	175	11	10:55	655	43	18:55	1135	75
03:00	180	12	11:00	660	44	19:00	1140	76
03:05	185	12	11:05	665	44	19:05	1145	76
03:10	190	12	11:10	670	44	19:10	1150	76
03:15	195	13	11:15	675	45	19:15	1155	77
03:20	200	13	11:20	680	45	19:20	1160	77
03:25	205	13	11:25	685	45	19:25	1165	77
03:30	210	14	11:30	690	46	19:30	1170	78
03:35	215	14	11:35	695	46	19:35	1175	78
03:40	220	14	11:40	700	46	19:40	1180	78
03:45	225	15	11:45	705	47	19:45	1185	79
03:50	230	15	11:50	710	47	19:50	1190	79
03:55	235	15	11:55	715	47	19:55	1195	79
04:00	240	16	12:00	720	48	20:00	1200	80
04:05	245	16	12:05	725	48	20:05	1205	80
04:10	250	16	12:10	730	48	20:10	1210	80
04:15	255	17	12:15	735	49	20:15	1215	81
04:20	260	17	12:20	740	49	20:20	1220	81
04:25	265	17	12:25	745	49	20:25	1225	81
04:30	270	18	12:30	750	50	20:30	1230	82
04:35	275	18	12:35	755	50	20:35	1235	82
04:40	280	18	12:40	760	50	20:40	1240	82
04:45	285	19	12:45	765	51	20:45	1245	83
04:50	290	19	12:50	770	51	20:50	1250	83
04:55	295	19	12:55	775	51	20:55	1255	83
05:00	300	20	13:00	780	52	21:00	1260	84
05:05	305	20	13:05	785	52	21:05	1265	84
05:10	310	20	13:10	790	52	21:10	1270	84
05:15	315	21	13:15	795	53	21:15	1275	85
05:20	320	21	13:20	800	53	21:20	1280	85
05:25	325	21	13:25	805	53	21:25	1285	85
05:30	330	22	13:30	810	54	21:30	1290	86
05:35	335	22	13:35	815	54	21:35	1295	86
05:40	340	22	13:40	820	54	21:40	1300	86
05:45	345	23	13:45	825	55	21:45	1305	87
05:50	350	23	13:50	830	55	21:50	1310	87
05:55	355	23	13:55	835	55	21:55	1315	87
06:00	360	24	14:00	840	56	22:00	1320	88

*Fortsetzung auf der nächsten Seite*

<b>Uhrzeit</b>	<b>minuten</b>	<b>m15</b>	<b>Uhrzeit</b>	<b>minuten</b>	<b>m15</b>	<b>Uhrzeit</b>	<b>minuten</b>	<b>m15</b>
06:05	365	24	14:05	845	56	22:05	1325	88
06:10	370	24	14:10	850	56	22:10	1330	88
06:15	375	25	14:15	855	57	22:15	1335	89
06:20	380	25	14:20	860	57	22:20	1340	89
06:25	385	25	14:25	865	57	22:25	1345	89
06:30	390	26	14:30	870	58	22:30	1350	90
06:35	395	26	14:35	875	58	22:35	1355	90
06:40	400	26	14:40	880	58	22:40	1360	90
06:45	405	27	14:45	885	59	22:45	1365	91
06:50	410	27	14:50	890	59	22:50	1370	91
06:55	415	27	14:55	895	59	22:55	1375	91
07:00	420	28	15:00	900	60	23:00	1380	92
07:05	425	28	15:05	905	60	23:05	1385	92
07:10	430	28	15:10	910	60	23:10	1390	92
07:15	435	29	15:15	915	61	23:15	1395	93
07:20	440	29	15:20	920	61	23:20	1400	93
07:25	445	29	15:25	925	61	23:25	1405	93
07:30	450	30	15:30	930	62	23:30	1410	94
07:35	455	30	15:35	935	62	23:35	1415	94
07:40	460	30	15:40	940	62	23:40	1420	94
07:45	465	31	15:45	945	63	23:45	1425	95
07:50	470	31	15:50	950	63	23:50	1430	95
07:55	475	31	15:55	955	63	23:55	1435	95



## A.4 Kodierung der Tage

Kodierung der Tage Januar - April							
Datum	tag	Datum	tag	Datum	tag	Datum	tag
Jan 01	0	Feb 01	31	Mar 01	59	Apr 01	90
Jan 02	1	Feb 02	32	Mar 02	60	Apr 02	91
Jan 03	2	Feb 03	33	Mar 03	61	Apr 03	92
Jan 04	3	Feb 04	34	Mar 04	62	Apr 04	93
Jan 05	4	Feb 05	35	Mar 05	63	Apr 05	94
Jan 06	5	Feb 06	36	Mar 06	64	Apr 06	95
Jan 07	6	Feb 07	37	Mar 07	65	Apr 07	96
Jan 08	7	Feb 08	38	Mar 08	66	Apr 08	97
Jan 09	8	Feb 09	39	Mar 09	67	Apr 09	98
Jan 10	9	Feb 10	40	Mar 10	68	Apr 10	99
Jan 11	10	Feb 11	41	Mar 11	69	Apr 11	100
Jan 12	11	Feb 12	42	Mar 12	70	Apr 12	101
Jan 13	12	Feb 13	43	Mar 13	71	Apr 13	102
Jan 14	13	Feb 14	44	Mar 14	72	Apr 14	103
Jan 15	14	Feb 15	45	Mar 15	73	Apr 15	104
Jan 16	15	Feb 16	46	Mar 16	74	Apr 16	105
Jan 17	16	Feb 17	47	Mar 17	75	Apr 17	106
Jan 18	17	Feb 18	48	Mar 18	76	Apr 18	107
Jan 19	18	Feb 19	49	Mar 19	77	Apr 19	108
Jan 20	19	Feb 20	50	Mar 20	78	Apr 20	109
Jan 21	20	Feb 21	51	Mar 21	79	Apr 21	110
Jan 22	21	Feb 22	52	Mar 22	80	Apr 22	111
Jan 23	22	Feb 23	53	Mar 23	81	Apr 23	112
Jan 24	23	Feb 24	54	Mar 24	82	Apr 24	113
Jan 25	24	Feb 25	55	Mar 25	83	Apr 25	114
Jan 26	25	Feb 26	56	Mar 26	84	Apr 26	115
Jan 27	26	Feb 27	57	Mar 27	85	Apr 27	116
Jan 28	27	Feb 28	58	Mar 28	86	Apr 28	117
Jan 29	28			Mar 29	87	Apr 29	118
Jan 30	29			Mar 30	88	Apr 30	119
Jan 31	30			Mar 31	89		

Kodierung der Tage Mai - August							
<b>Datum</b>	<b>tag</b>	<b>Datum</b>	<b>tag</b>	<b>Datum</b>	<b>tag</b>	<b>Datum</b>	<b>tag</b>
May 01	120	Jun 01	151	Jul 01	181	Aug 01	212
May 02	121	Jun 02	152	Jul 02	182	Aug 02	213
May 03	122	Jun 03	153	Jul 03	183	Aug 03	214
May 04	123	Jun 04	154	Jul 04	184	Aug 04	215
May 05	124	Jun 05	155	Jul 05	185	Aug 05	216
May 06	125	Jun 06	156	Jul 06	186	Aug 06	217
May 07	126	Jun 07	157	Jul 07	187	Aug 07	218
May 08	127	Jun 08	158	Jul 08	188	Aug 08	219
May 09	128	Jun 09	159	Jul 09	189	Aug 09	220
May 10	129	Jun 10	160	Jul 10	190	Aug 10	221
May 11	130	Jun 11	161	Jul 11	191	Aug 11	222
May 12	131	Jun 12	162	Jul 12	192	Aug 12	223
May 13	132	Jun 13	163	Jul 13	193	Aug 13	224
May 14	133	Jun 14	164	Jul 14	194	Aug 14	225
May 15	134	Jun 15	165	Jul 15	195	Aug 15	226
May 16	135	Jun 16	166	Jul 16	196	Aug 16	227
May 17	136	Jun 17	167	Jul 17	197	Aug 17	228
May 18	137	Jun 18	168	Jul 18	198	Aug 18	229
May 19	138	Jun 19	169	Jul 19	199	Aug 19	230
May 20	139	Jun 20	170	Jul 20	200	Aug 20	231
May 21	140	Jun 21	171	Jul 21	201	Aug 21	232
May 22	141	Jun 22	172	Jul 22	202	Aug 22	233
May 23	142	Jun 23	173	Jul 23	203	Aug 23	234
May 24	143	Jun 24	174	Jul 24	204	Aug 24	235
May 25	144	Jun 25	175	Jul 25	205	Aug 25	236
May 26	145	Jun 26	176	Jul 26	206	Aug 26	237
May 27	146	Jun 27	177	Jul 27	207	Aug 27	238
May 28	147	Jun 28	178	Jul 28	208	Aug 28	239
May 29	148	Jun 29	179	Jul 29	209	Aug 29	240
May 30	149	Jun 30	180	Jul 30	210	Aug 30	241
May 31	150			Jul 31	211	Aug 31	242

Kodierung der Tage September - Dezember							
<b>Datum</b>	<b>tag</b>	<b>Datum</b>	<b>tag</b>	<b>Datum</b>	<b>tag</b>	<b>Datum</b>	<b>tag</b>
Sep 01	243	Oct 01	273	Nov 01	304	Dec 01	334
Sep 02	244	Oct 02	274	Nov 02	305	Dec 02	335
Sep 03	245	Oct 03	275	Nov 03	306	Dec 03	336
Sep 04	246	Oct 04	276	Nov 04	307	Dec 04	337
Sep 05	247	Oct 05	277	Nov 05	308	Dec 05	338
Sep 06	248	Oct 06	278	Nov 06	309	Dec 06	339
Sep 07	249	Oct 07	279	Nov 07	310	Dec 07	340
Sep 08	250	Oct 08	280	Nov 08	311	Dec 08	341
Sep 09	251	Oct 09	281	Nov 09	312	Dec 09	342
Sep 10	252	Oct 10	282	Nov 10	313	Dec 10	343
Sep 11	253	Oct 11	283	Nov 11	314	Dec 11	344
Sep 12	254	Oct 12	284	Nov 12	315	Dec 12	345
Sep 13	255	Oct 13	285	Nov 13	316	Dec 13	346
Sep 14	256	Oct 14	286	Nov 14	317	Dec 14	347
Sep 15	257	Oct 15	287	Nov 15	318	Dec 15	348
Sep 16	258	Oct 16	288	Nov 16	319	Dec 16	349
Sep 17	259	Oct 17	289	Nov 17	320	Dec 17	350
Sep 18	260	Oct 18	290	Nov 18	321	Dec 18	351
Sep 19	261	Oct 19	291	Nov 19	322	Dec 19	352
Sep 20	262	Oct 20	292	Nov 20	323	Dec 20	353
Sep 21	263	Oct 21	293	Nov 21	324	Dec 21	354
Sep 22	264	Oct 22	294	Nov 22	325	Dec 22	355
Sep 23	265	Oct 23	295	Nov 23	326	Dec 23	356
Sep 24	266	Oct 24	296	Nov 24	327	Dec 24	357
Sep 25	267	Oct 25	297	Nov 25	328	Dec 25	358
Sep 26	268	Oct 26	298	Nov 26	329	Dec 26	359
Sep 27	269	Oct 27	299	Nov 27	330	Dec 27	360
Sep 28	270	Oct 28	300	Nov 28	331	Dec 28	361
Sep 29	271	Oct 29	301	Nov 29	332	Dec 29	362
Sep 30	272	Oct 30	302	Nov 30	333	Dec 30	363
		Oct 31	303			Dec 31	364

## A.5 Vogelperspektive

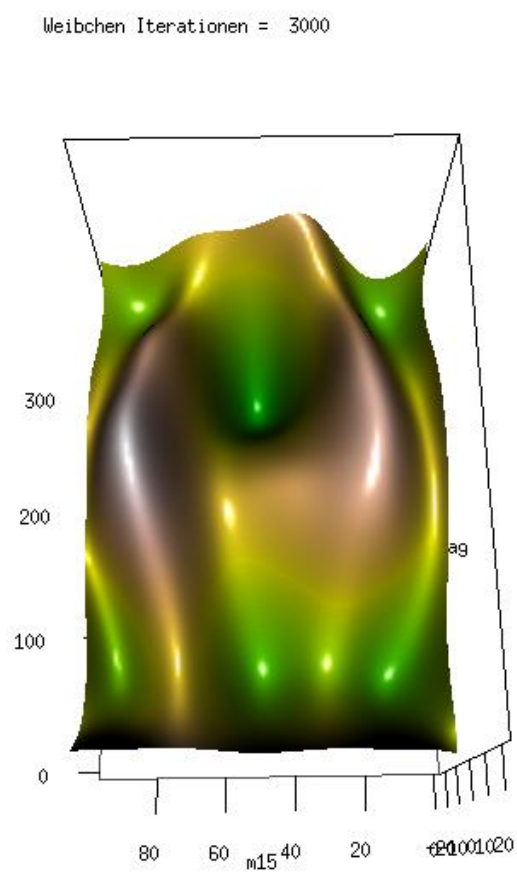


Abbildung A.1: Weibchen 15 Minuten