Räumliche, GIS-gestützte Analyse von Linientransektstichproben

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

„Doctor rerum naturalium“

der Georg-August-Universität Göttingen

vorgelegt von

Felix Mader

aus Buxtehude

Göttingen, 2007

Referent: Prof. Dr. J. Saborowski, Institut für forstliche Biometrie und Informatik

Korreferent: Prof. Dr. W. May, Institut für Informatik

Tag der mündlichen Prüfung: 9. 3. 2007

Dies ist ein Ausdruck des Inhaltes unter
http://resolver.sub.uni-goettingen.de/purl/?webdoc-1535.

Für B. G.

Vorwort

Die vorliegende Arbeit entstand zwischen September 2001 und Januar 2007 am Institut für Forstliche Biometrie und Informatik der Universität Göttingen. Ideengeber der Themenstellung war eine vorhergehende Masterarbeit, die ich nach einem ersten Aufenthalt in Namibia im Jahr 2000 ebenfalls am Institut für Forstliche Biometrie und Informatik über die Wildzählungen der Farmer der Seeis Conservancy angefertigt habe. Zwischen September 2001 und Dezember 2003 wurde das Vorhaben durch ein Promotionsstipendium nach dem niedersächsischen Graduiertenförderungsgesetz, während der Aufnahmen in Namibia zusätzlich durch ein Aufstockungsstipendium des DAAD gefördert.

Die Liste derer, die zum Gelingen dieser Arbeit beigetragen, ist sehr lang, ich möchte aber einige Personen besonders hervorheben, ohne die sie gar nicht erst möglich gewesen wäre. Zu ihnen gehört zunächst mein Doktorvater, Prof. Dr. J. Saborowski, dem ich für die Freiheit bei der Themenwahl und ihrer Umsetzung und für Rat und Unterstützung in allen fachlichen, gelegentlich finanziellen und manchmal auch menschlichen Problemen danke. Birgit und Harald Förster von der Okatumba Wildlife Research danke ich, weil sie es waren, die mich erstmalig nach Namibia geholt und damit meine Faszination und Leidenschaft für dieses wunderschöne Land geweckt haben; für die Dauer meiner Aufnahmen in Namibia haben sie mich in ihrem Haus aufgenommen und logistisch und menschlich unterstützt. Den Farmern der Seeis Conservancy, auf deren Farmen ich Wild zählen und weitere Daten erfassen durfte, bin ich zutiefst dankbar; sie haben mich nicht nur „machen lassen“, sondern stets sehr herzlich aufgenommen und mit Fahrzeugen, Fahrern, Essen und einem Bett zum Schlafen versorgt. Hier möchte ich insbesondere die Familien Held, Pack und Wilckens nennen, ohne die ich die vielen kleinen und größeren Widrigkeiten nicht hätte überwinden können. Auch wenn letztlich nur ein Teil der erfaßten Daten in diese Arbeit eingegangen ist, so werden mir um so mehr die Freundlichkeit, mit der ich empfangen wurde, die Unterstützung, die ich erfahren habe, und die vielen schönen Momente als eine der schönen und großen Lebenserfahrungen in Erinnerung bleiben.

Prof. Dr. Dr. h. c. B. Sloboda, dem Leiter des Instituts für Forstliche Biometrie und Informatik, und den übrigen Angehörigen des Instituts danke ich für die Bereitstellung der räumlichen und technischen Möglichkeiten bzw. für ein ausgesprochen angenehmes Arbeitsklima und die stetige Unterstützung. Prof. Dr. W. May danke ich für die Übernahme des Korreferats. Schließlich möchte ich mich bei all den Freunden und Verwandten bedanken, die in den vergangenen Jahren oft auf mich verzichten mußten und mir dennoch treu geblieben sind.

Zum Schluß eine kleine Anmerkung zum Titel dieser Arbeit: Ich bin im Laufe der Zeit einige Male darauf angesprochen, ob es sich bei GIS-gestützten Analysen nicht immer auch um räumliche handelt. Grundsätzlich stimmt dies sicher-

lich. Die zusätzliche Betonung auf „räumlich“ soll jedoch hervorheben, daß nicht nur GIS-Technologie verwendet wurde, um einen räumlichen Bezug zwischen den erfaßten geographischen Informationen herzustellen, sondern daß zusätzlich in einem Teil der Analysen die räumliche Beziehung (in Form der räumlichen Autokorrelation) zwischen den untersuchten Größen auch in die anschließende statistische Modellierung einging.

Göttingen, Februar 2007 Felix Mader

Inhaltsverzeichnis

Abbildungsverzeichnis ix

Tabellenverzeichnis xi

1. Einleitung und Problemstellung 1

2. Modelle und Methoden (Teilprojekte 1 und 2) 5

2.1. Das Linientransektverfahren . . . 5
2.1.1. Grundlagen der Dichteschätzung . . . 6
2.1.2. Anpassungsgüte und Modellauswahl . . . 9
2.1.3. Annahmen und Voraussetzungen . . . 10
2.1.4. Varianzschätzung . . . 11
2.1.5. Zusätzliche Analyseoptionen . . . 13
2.1.5.1. Kurvige Transekte . . . 13
2.1.5.2. Datenbeschneidung . . . 15
2.1.5.3. Wiederholte Aufnahme permanenter Transekte . 16
2.1.5.4. Korrektur der Dichteschätzer bei abweichender Stichprobenfläche . . . 18
2.1.5.5. Vergleich zweier Dichteschätzungen . . . 19
2.2. Verallgemeinerte lineare Modelle (GLM) . . . 20
2.2.1. Die Exponentialfamilie . . . 22
2.2.1.1. Die Poisson-Verteilung . . . 23
2.2.2. Schätzung der Parameter . . . 24
2.2.2.1. Parameterschätzung im Poisson-Modell . . . 27
2.2.3. Anpassungsgüte und Residuenanalyse . . . 28
2.2.4. Varianzschätzung und Hypothesentests . . . 30
2.2.5. Überdispersion . . . 33
2.2.6. Quasi-Likelihood . . . 35
2.2.7. Räumliche Autokorrelation in GLMs . . . 37
2.2.7.1. Der GEE-Ansatz . . . 38
2.2.7.2. Der PL-Ansatz . . . 41
2.2.7.3. Varianzschätzung, Hypothesentests und Modellauswahl . . . 43
2.2.7.4. Räumliche Vorhersagen . . . 45
2.3. Das Count-Modell . . . 48
2.3.1. Modellformulierung . . . 48
2.3.2. Vorhersagen . . . 51
2.3.3. Varianzschätzung . . . 53

3. Anwendungssoftware (Teilprojekt 3) 57
3.1. Konventionelle Linientransektanalyse 60
3.2. Raumbezogene Analyse 65
3.2.1. Segmente und räumliche Verschneidungen 65
3.2.2. Parameterschätzungen und Vorhersagen 72

4. Untersuchungsgebiet und Wildarten 77
4.1. Untersuchungsgebiet 77
4.2. Untersuchte Arten 81

5. Datenerhebung 89
5.1. Wildzählungen 89
5.2. Raumbezogene Informationen 92

6. Ergebnisse 101
6.1. Validierung der Farmwegzählungen (Teilprojekt 1) 101
6.2. Räumliche Modellierung (Teilprojekt 2) 114

7. Diskussion und Ausblick 131
7.1. Validierung der Farmwegzählungen (Teilprojekt 1) 131
7.2. Räumliche Modellierung (Teilprojekt 2) 132

8. Zusammenfassung 141

A. Notation und verwendete Abkürzungen 145

B. Modelle und Methoden 149
B.1. Darstellungsformen der erwarteten Fisher Informationsmatrix . . 149
B.2. Fisher Scoring als IWLS 149
B.3. Darstellungsformen des Prädiktors im räumlichen GLM 150

C. Anwendungssoftware 153
C.1. Berechnung der Objektpositionen 153
C.2. Probleme der Segmenterstellung in ArcGIS 153

D. Ergebnisse 155
D.1. Validierung der Farmwegzählungen (Teilprojekt 1) 155

Literaturverzeichnis 157

Abbildungsverzeichnis

2.1. Datenerhebung auf gerade verlaufenden Transekten. 7
2.2. Entdeckungsfunktion und effektive Streifenbreite. 8
2.3. Datenerhebung auf kurvigen Transekten. 14
2.4. Unvollständige Erfassung der Stichprobenflächen. 19

3.1. Schematische Darstellung des Ablaufs der für die konventionelle Linientransektanalyse in Distance verwendeten SAS-Makros. . . . 63
3.2. Begriffe der Segmenterstellung. 65
3.3. Segmenterstellung. 66
3.4. Probleme der Segmenterstellung I. 67
3.5. Probleme der Segmenterstellung II. 68
3.6. Schematische Darstellung der für die Erstellung der Segmente verwendeten Makros und Skripte. 69
3.7. Räumliche Verschneidung der Segmente mit weiteren geographischen Informationen. 71

4.1. Lage des Untersuchungsgebietes. 78
4.2. Landschaft und Vegetation des Untersuchungsgebietes. 79
4.3. Die untersuchten Wildarten Hartebeest, Großer Kudu und Oryx und ihre Verbreitungsgebiete. 83
4.4. Die untersuchten Wildarten Steinbock und Warzenschwein und ihre Verbreitungsgebiete. 84

5.1. Lage der Transekte im Untersuchungsgebiet. 90
5.2. Erfassung der Beobachtungs- und Objektposition. 91
5.3. Ausschnitt aus den erfaßten Geodaten. 92
5.4. Brightness-, Greenness-, und Wetness-Kanal der Tasseled-Cap-Transformation am Beispiel eines in der Nähe des Untersuchungsgebietes liegenden Areals. 96

6.1. Geschätzte Objektdichten, Individuendichten, Begegnungsraten und Individuenanzahlen sowie zugehörige 95%ige Konfidenzintervalle je Aufnahmephase, -schema und Art. 108
6.2. Objektpositionen der Oryxbeobachtungen für das WegPferd- und das BuschPferd-Schema aus beiden Aufnahmephasen. 111
6.3. Objektpositionen der Hartebeestbeobachtungen und räumliche Vorhersagen der Individuenanzahl. 123
6.4. Objektpositionen der Kudubeobachtungen und räumliche Vorhersagen der Individuenanzahl. 124

6.5. Objektpositionen der Oryxbeobachtungen und räumliche Vorhersagen der Individuenanzahl. 125
6.6. Objektpositionen der Steinbockbeobachtungen und räumliche Vorhersagen der Individuenanzahl. 126
6.7. Objektpositionen der Warzenschweinbeobachtungen und räumliche Vorhersagen der Individuenanzahl. 127

C.1. Probleme der Segmentgenerierung bei Verwendung der ArcGIS Buffer-Funktion. 154

Tabellenverzeichnis

4.1. Eigenschaften und Habitatansprüche der fünf untersuchten Wildarten. 85

5.1. Koeffizienten der Tasseled-Cap-Transformation für die Kanäle des ETM+ Sensors. 97

6.1. Anzahl der je Aufnahmephase, -schema und Art beobachteten Gruppen und Individuen. 102
6.2. Geschätzte mittlere Gruppengrößen, zugehörige Standardfehler und Anzahl der zur Berechnung verwendeten Beobachtungen je Art und Aufnahmephase. 104
6.3. Verwendete Entdeckungsfunktionen und deren geschätzte Parameter je Art und Aufnahmeschema. 105
6.4. Vergleich der Individuendichteschätzungen ($\hat{D}$) beider Aufnahmephasen je Art und Schema. 112
6.5. p-Werte der T-Tests zum Vergleich der Individuendichteschätzungen der drei Aufnahmeschemata je Aufnahmephase und Art. . 113
6.6. Segmentlängen und Anzahl resultierender Segmente. 115
6.7. In den Modellen verwendete abhängige Variablen, Kovariaten und Segmentlängen sowie Parameterschätzungen und zugehörige Standardfehler und p-Werte des Modells mit unabhängigen Beobachtungen und mit Einbezug räumlicher Autokorrelation je Aufnahmephase und Art. 118
6.8. Erläuterung der in Tabelle 6.7 verwendeten Abkürzungen der Kovariatenbezeichnungen. 119
6.9. Geschätzte Kovarianzparameter der Modelle mit und ohne Berücksichtigung räumlicher Autokorrelation je Aufnahmephase und Art. 121
6.10. Schätzungen der Gesamtindividuenabundanz und deren Standardfehler für die Modelle mit und ohne Berücksichtigung räumlicher Autokorrelation sowie der konventionellen Linientransektanalyse je Aufnahmephase und Art. 128
6.11. Schätzungen der Gesamtindividuendichte und deren Standardfehler für die Modelle mit und ohne Berücksichtigung räumlicher Autokorrelation sowie der konventionellen Linientransektanalyse je Aufnahmephase und Art. 129

D.1. Nach dem Linientransektverfahren geschätzte Objektdichten und Individuendichten sowie zugehörige 95%ige Konfidenzintervalle je Aufnahmephase, Schema und Art. 155

D.2. Nach dem Linientransektverfahren geschätzte Begegnungsraten und Individuenanzahlen sowie zugehörige 95%ige Konfidenzintervalle je Aufnahmephase, -chema und Art. 156

1. Einleitung und Problemstellung

Als eines der ersten Länder weltweit hat Namibia bereits mit seiner Unabhängigkeit im Jahr 1990 den Schutz und Erhalt der Natur in die Verfassung aufgenommen. In Artikel 95 heißt es:

> The State shall actively promote and maintain the welfare of the people by adopting [...] policies aimed at [...] maintenance of ecosystems, essential ecological processes and biological diversity of Namibia and utilization of living natural resources on a sustainable basis for the benefit of all Namibians, both present and future [...]

Ein Ausdruck dieser Umweltschutzpolitik zum Wohle der Menschen ist das seit 1995 unter Leitung des *Ministry of Environment and Tourism* (MET) laufende Programm „*Community Based Natural Resource Management*“ (CBNRM). Wesentlicher Bestandteil dieses Programms ist die Bildung sog. *Conservancies* (MET, 1995). Dabei handelt es sich um Zusammenschlüsse *kommunaler*, d. h. in Staatsbesitz befindlicher Flächen, für die ein gemeinsamer Management- und Entwicklungsplan aufgestellt wird. Ziel ist es, die im wesentlichen aus Tourismus und Trophäenjagd erzielten Einnahmen einer Conservancy direkt an die lokale Bevölkerung zurückfließen zu lassen, um so die Lebensbedingungen zu verbessern und gleichzeitig den schonenden, nachhaltigen Umgang mit natürlichen Ressourcen zu fördern. Bis zur Gründung der Conservancies befand sich das Wild wie der Grund und Boden in Staatsbesitz, legale Bejagung erfolgte fast ausschließlich in Form staatlich regulierter Trophäenjagden ohne Nutzen für die lokale Bevölkerung; mit der Folge, daß die Bevölkerung etwaige Wildschäden zu akzeptieren hatte, illegale Jagd jedoch keineswegs verhindert wurde. Vielmehr waren die Bestände in weiten Teilen des kommunalen Farmlandes stark dezimiert (Ashley, 1995; Ashley u. Barnes, 1996). Mit der Übergabe des (eingeschränkten) Nutzungsrechtes an die im Gebiet der Conservancy lebenden Bevölkerung sollte – mit bisher guten Erfolgen (s. z. B. Barnes et al., 2001; Bandyopadyay et al., 2004) – das Bewußtsein und ein eigenes Interesse am Erhalt der natürlichen Ressourcen protegiert werden. Ein Ideengeber hierfür war die Entwicklung der Wildbestände auf privatem, kommerziellen Farmland. Kommerzielle Farmer können bereits seit 1967 zahlreiche der vorkommenden Wildarten nutzen, lediglich eingeschränkt durch ein Permit-System, das die Beantragung von Abschüssen beim MET und die Entrichtung von Gebühren erfordert (Jones u. Murphree, 2001; Barnes u. de Jager, 1996). Infolge des entstandenen wirtschaftlichen Nutzens ließen sich in den darauffolgenden Jahrzehnten rasant wachsende Wildbestände beobachten (Barnes u. de Jager, 1995).

Neben den derzeit 44 kommunalen Conservancies existieren mittlerweile auch ca. 26 *kommerzielle*, d. h. aus Privatfarmen bestehende Conservancies. Im Gegensatz zu den kommunalen Conservancies handelt es sich hierbei aber bisher

lediglich um freiwillige Zusammenschlüsse ohne rechtliche Grundlage und ohne besondere Vorteile für die einzelnen Conservancy-Mitglieder. Dies soll sich, nach mehrjährigen Planungen, 2007 in Form des *Parks and Wildlife Management Bill* ändern. Der Gesetzentwurf sieht u. a. die weitere Vereinfachung der Wildnutzung für (kommerzielle) Conservancy-Mitglieder und höhere Strafen für Wilderei in Conservancy-Gebieten, aber auch die Verpflichtung zur Erstellung von Managementplänen vor. Teil solcher Managementpläne ist, neben der Angabe getätigter und geplanter Abschüsse, die Angabe der aktuellen Wildbestände. Da es bislang keine gesetzlichen und standardisierten Vorgaben gibt, wie die Wildbestände zu erfassen sind, haben sich in den Conservancies verschiedene Zählverfahren etabliert. In den kommunalen Conservancies, die bereits zur Abgabe von Managementplänen verpflichtet sind, werden Wildzählungen im wesentlichen halbjährlich per Auto in Form sog. *Road Strip Counts* durchgeführt, ergänzt durch Wasserlochzählungen und – im Zwei-Jahres-Turnus – durch Luftzählungen. Unterstützung erhalten die kommunalen Farmer hierbei vom MET und von nicht-staatlichen Organisationen wie dem *World Wildlife Fund for Nature* (WWF).

Wenngleich sie noch nicht zur Abgabe von Managementplänen verpflichtet sind, führen viele kommerzielle Conservancies ebenfalls Wildzählungen durch, in denen ebenso unterschiedliche Verfahren wie Road Strip Counts und Wasserlochzählungen zum Einsatz kommen. Die Aufnahmen erfolgen hier jedoch vollständig in Eigenregie durch die Farmer.

Ein Verfahren, das in mehreren kommerziellen, und im Zuge der Luftzählungen auch in kommunalen Conservancies, Anwendung findet, ist das Linientransektverfahren. Allgemein werden in diesem Verfahren die zu zählenden Objekte von (im Idealfall) zufälligen Transekten aus erfaßt und anhand der zu den Objekten gemessenen Entfernungen deren „Entdeckbarkeit“ modelliert, um einen Korrekturfaktor für übersehene Objekte zu schätzen. Das Linientransektverfahren bietet sich für die Durchführung von Wildzählungen in den namibischen Conservancies aus mehreren Gründen an:

Die Datenerhebung ist vergleichsweise einfach und leicht zu vermitteln. Es bedarf nur einiger weniger Voraussetzungen, die zu erfüllen sind, um eine verläßliche Analyse zu gewährleisten. Da in beiden Conservancy-Formen Laien (Planung, Durchführung und Analyse von Stichprobenverfahren betreffend) maßgeblich an der Durchführung der Zählungen beteiligt sind, ist es von entscheidender Bedeutung, ein einfaches und eingängiges Verfahren zu verwenden, um die Motivation der Durchführenden zu erhalten und damit eine möglichst hohe Qualität der gesammelten Daten sicherzustellen. Lediglich die Analyse der Daten erfordert ein gewisses statistisches Grundverständnis.

Die Datenerhebung ist kostengünstig, da sie keine spezielle Ausrüstung erfordert, wenngleich sie durch technische Geräte vereinfacht und präzisiert werden kann. Teures Gerät würde die Bereitschaft kommerzieller Farmer, an den Zählungen überhaupt teilzunehmen, deutlich senken, da diese bislang keine finanziellen Vorteile aus der Mitgliedschaft in einer Conservancy ziehen. Ebenfalls aus Kostengründen kommen in kommerziellen Conservancies auch professionell im Auftrag durchgeführte Zählungen nur in Ausnahmefällen in Frage.

Schließlich ist das Verfahren für unterschiedlichste Habitattypen geeignet. In dem im Rahmen dieser Arbeit untersuchten Gebiet, wie in vielen Regionen Namibias, sind Vegetation und Landschaftsstruktur sehr heterogen und wechseln binnen kurzer Entfernungen zwischen dichtem Busch, Hügellandschaften und weiten offenen Ebenen. Das Linientransektverfahren verhält sich – ein geeignetes Stichprobendesign vorausgesetzt – robust gegenüber den unterschiedlichen „Entdeckbarkeiten“ in den verschiedenen Habitattypen.

Diesen Vorteilen steht jedoch ein nicht weniger gewichtiger Nachteil gegenüber: Vegetation, Bodenverhältnisse und kilometerlange Zäune auf kommerziellen Farmen machen ein Befahren der Farmen abseits der Wege nahezu unmöglich. Aus diesem Grund werden stets Farmwege oder öffentliche Straßen als Transekte verwendet. Hierdurch wird jedoch klar eine Grundannahme der meisten statistischen Verfahren, die zufällige Auswahl und Unabhängigkeit der Stichprobeneinheiten, verletzt. Die Repräsentativität der Ergebnisse für das gesamte Untersuchungsgebiet (die Conservancy oder ein Teil davon) wird damit in Frage gestellt. In der Regel kann Wegen ein abstoßender Effekt unterstellt werden, da es sich um Orte regelmäßiger oder gar dauernder Beunruhigung handelt. Entlang der Wege würde demnach zu wenig Wild gezählt und die Gesamtdichte unterschätzt. Im speziellen Fall der namibischen Farmwege läßt sich jedoch nicht von vornherein sagen, ob die Farmwege überhaupt einen Einfluß auf die Wildverteilungen haben, da es sich um Privatwege handelt, die zum überwiegenden Teil nur selten befahren werden. Neben dem genannten negativen direkten Effekt der Wege ist zudem auch ein positiver indirekter Effekt denkbar. Vermeintlich attraktive Orte wie Salzlecken und Wasserstellen liegen fast immer direkt an Farmwegen und könnten dazu führen, daß entlang der Wege „zuviel“ gezählt wird. **Das erste Ziel (Teilprojekt 1)**, das diese Arbeit verfolgte, war daher die Validierung von Dichte- und Abundanzschätzungen, die anhand von Linientransekterhebungen mit Wegtransekten ermittelt wurden. Zu diesem Zweck wurden über mehrere Monate Aufnahmen nach verschiedenen Aufnahmeschemata auf einem Farmkomplex durchgeführt, der Teil einer kommerziellen Conservancy im Zentrum Namibias ist. Die Schemata umfaßten dabei sowohl die Verwendung von Farmwegen als Transekte als auch ein den statistischen Ansprüchen genügendes Versuchsdesign mit zufällig angelegten Transekten.

Das zweite Teilprojekt widmete sich einem Ansatz der räumlichen Modellierung basierend auf Linientransekterhebungen. Im Zuge steigender Verfügbarkeit kostengünstiger und hochwertiger geographischer Informationen wächst in unzähligen Bereichen der Bedarf, diese in (räumlichen) Bezug zu erfaßten Daten zu stellen und in herkömmliche Analysen mit einzubeziehen. Zum einen mit dem Ziel, die in den Daten beobachtete Streuung (teilweise) zu erklären und damit Schätzungen zu präzisieren, zum anderen, um lokale Vorhersagen der untersuchten Größe zu ermöglichen. Die klassische Linientransektanalyse sieht jedoch den Einbezug zusätzlicher räumlicher wie nicht-räumlicher Informationen nicht vor. Um diesen dennoch zu ermöglichen, hat Hedley (2000) u. a. das sog. *Count-Modell* vorgestellt. Es basiert im wesentlichen auf der Verwendung eines log-linearen Modells, mit dessen Hilfe die Anzahl der in einem Transektabschnitt beobachteten Objekte in Abhängigkeit von zusätzlichen (räumlichen) Kovariaten modelliert wird. Basierend auf den Ansätzen von Albert u. McShane

(1995), Gotway u. Stroup (1997) und Gotway u. Wolfinger (2003) stellt diese Arbeit eine Möglichkeit vor, wie das Count-Modell von Hedley zusätzlich um den Einbezug räumlicher Autokorrelation[1] erweitert werden kann. Zur Umsetzung dieses Teilprojektes wurden die Linientransektdaten aus Teilprojekt 1 genutzt und zusätzlich eine umfangreiche Geodatenbank aufgebaut, die zahlreiche raumbezogene Informationen enthielt, denen ein Einfluß auf die lokale Verteilung des Wildes unterstellt wurde. Die zusätzlichen Geodaten wurden für das gesamte Untersuchungsgebiet anhand von Luft- und Satellitenbildern, topographischen Karten oder direkt per GPS (*Global Positioning System*) erfaßt.

Das dritte Ziel (Teilprojekt 3) dieser Arbeit bestand in der Entwicklung von Software-Modulen, die zum einen die Umsetzung der in beiden Teilprojekten angewandten Methoden erleichterte bzw. überhaupt ermöglichte und zum anderen eine weitestgehende Automatisierung der Arbeitsabläufe erlaubte. Zwar existiert Software für die konventionelle Analyse von Linientransekterhebungen und in begrenztem Umfang auch für die Modellierung räumlich autokorrelierter, nicht normalverteilter Daten, wie sie in der Anwendung des Count-Modells anfallen, gerade in der Umsetzung des Count-Modells fallen jedoch aufwendige und komplexe Arbeitsschritte an, die derzeit in keinem Programm direkt implementiert sind. Dies betrifft insbesondere die umfangreiche Erstellung und Manipulation von Geodaten. Für die Verarbeitung von Geodaten kam hier ESRIs ArcGIS, für das Datenmanagement und die wesentlichen statistischen Analysen das Software-Paket SAS zum Einsatz. Die Software-Module wurden entsprechend in Form sog. *Geoprocessing-Skripte* (ArcGIS) und SAS-Makros erstellt.

Wenngleich die nachfolgend vorgestellten Methoden gezielt für die Analyse der im Rahmen dieser Arbeit erfaßten namibischen Wildzählungsdaten zusammengetragen und die Software-Module entsprechend entwickelt wurden, so wurde hierbei dennoch das Ziel verfolgt, die Modelle so allgemein zu beschreiben und die Software-Module so zu entwickeln, daß ihre Anwendung ohne große Änderungen auch in anderen Projekten möglich ist. Die Kapitel 2 und 3 beschreiben die angewandten Modelle und Methoden sowie die entwickelten Software-Module daher weitestgehend unabhängig von den konkreten Daten. Letztere werden in Kapitel 4 und 5, die Ergebnisse in Kapitel 6 vorgestellt. Die Diskussion in Kapitel 7 bezieht sowohl besondere Aspekte der Durchführung von Linientransektaufnahmen auf kommerziellem namibischen Farmland als auch allgemeine Überlegungen und Ausblicke zu den vorgestellten Methoden ein.

[1] Der Begriff der (positiven) räumlichen Autokorrelation beschreibt, vereinfacht ausgedrückt, den Effekt, daß sich nahe gelegene Beobachtungen in der Regel ähnlicher sind als weiter entfernte. So ist beispielsweise die Höhe des Niederschlags in benachbarten Orten meist ähnlicher als in weit voneinander entfernten Orten.

2. Modelle und Methoden (Teilprojekte 1 und 2)

Der erste Teil dieses Kapitels befaßt sich mit der Methodik des Linientransektverfahrens. Er beschränkt sich hierbei auf eine kurze Darlegung der grundlegenden Ideen und Methoden sowie auf die zusätzlich zur Bearbeitung der beiden ersten Teilprojekte benötigten Ansätze. Umfangreicher, da im Zusammenhang mit Linientransektanalysen bisher nur von Hedley (2000) angewandt, fällt die Erläuterung der verallgemeinerten linearen Modelle (*Generalized Linear Models*, GLMs) im zweiten Teil aus. Die Ausführungen befassen sich, zunächst gänzlich unabhängig von Linientransekten, mit GLMs zur Auswertung unabhängiger Stichproben nicht normalverteilter Daten. Ein deutlicher Schwerpunkt wird hierbei auf die Poisson-Verteilung bzw. das zugehörige GLM, das log-lineare Modell (auch Poisson-Modell oder Poisson-Regression), gelegt, da es im Zuge von Wildzählungen vordergründig um die Modellierung von Anzahlen – nämlich die von Tieren bzw. Tiergruppen – gehen wird. Diese Grundlagen werden dann auf den Fall abhängiger und speziell räumlich abhängiger Beobachtungen erweitert und im letzten Abschnitt schließlich mit den Methoden des Linientransektverfahrens zusammengeführt.

Zunächst völlig losgelöst von der in den nachfolgenden Kapiteln beschriebenen und im Rahmen dieser Arbeit analysierten Datengrundlage erläutern die folgenden Abschnitte allgemein Methoden und Modelle für die räumliche Auswertung von Linientransekterhebungen unter Verwendung von verallgemeinerten linearen Modellen (*Generalized Linear Models*, GLMs). Sie zielen insofern insbesondere auf die Erarbeitung der in Kapitel 6 aufgeführten Ergebnisse ab, können aber für sich betrachtet ebenso als allgemeiner Leitfaden für eine Möglichkeit der räumlichen Analyse von Linientransektaufnahmen dienen.

2.1. Das Linientransektverfahren

Das Linientransektverfahren gehört neben dem Punkttransekt-, dem *Trapping-Web*- und dem *Cue-Counting*-Verfahren zur Familie des sog. *Distance Sampling*. Wesentliches Charakteristikum der Distance Sampling-Verfahren ist, daß sie entgegen herkömmlicher Stichprobenverfahren keiner Vollaufnahme der Stichprobenflächen bedürfen, sondern einen Korrekturfaktor für nicht entdeckte Objekte schätzen. Eine grundlegende Rolle spielen hierbei die zwischen den entdeckten Objekten und den Positionen des Beobachters gemessenen Entfernungen.

Die nachfolgend vorzustellenden Methoden beruhen im wesentlichen auf den Arbeiten von Buckland et al. (2001, 2004), den Standardreferenzen des Distance Sampling. Für tiefergehende Betrachtungen und Erläuterungen der übrigen Ver-

fahren des Distance Sampling sei auf diese und die darin genannte weiterführende Literatur verwiesen.

2.1.1. Grundlagen der Dichteschätzung

Am ehesten vergleichbar ist das Linientransektverfahren mit einer Streifenzählung, in der von einer Anzahl k zufällig in das Untersuchungsgebiet gelegter Geraden (Transekte) die Objekte des Interesses bis zu einer Entfernung w beiderseits der Geraden beobachtet und gezählt werden. Werden innerhalb der insgesamt so aufgenommenen Fläche $a = 2\,L\,w$ alle Objekte erfaßt, wobei L die Gesamtlänge der Geraden ($= \sum_{i=1}^{k} l_i$) ist, so läßt sich die Objektdichte D_s erwartungstreu durch das Verhältnis aus der Anzahl entdeckter Objekte n und der Stichprobenfläche schätzen.

$$\hat{D}_s = \frac{n}{2\,w\,L} = \frac{n}{a} \tag{2.1}$$

Gerade die Annahme, tatsächlich alle Objekte zu entdecken, ist jedoch oftmals nur schwer bis gar nicht zu erfüllen und im wesentlichen abhängig von der gewählten Streifenbreite w, den Eigenschaften der untersuchten Objekte (Größe, Färbung, Verhalten usw.), der Effizienz des Beobachters und der Art des Untersuchungsgebietes (Dichte und Höhe der Vegetation usw.). Genau an diesem Punkt setzen das Linientransektverfahren und im allgemeinen die Verfahren des Distance Sampling an. Das Linientransektverfahren setzt weder die Vollaufnahme der Fläche voraus, noch ist die aufzunehmende Fläche von vornherein durch Festsetzung der Streifenbreite w vorzugeben. Es können also auch solche Beobachtungen genutzt werden, die außerhalb des Streifens liegen, und damit der Stichprobenumfang und die Präzision der Schätzer erhöht werden. Statt dessen wird ein Korrekturfaktor geschätzt, der den Anteil der übersehenen Objekte bemißt. Hierfür wird zunächst neben der Anzahl der Objekte auch deren senkrechte Entfernung vom Transekt x erfaßt. Da sie im Gelände meist leichter zu erfassen sind, werden oftmals, wie in Abbildung 2.1 dargestellt, die Entfernung zwischen Beobachter und Objekt y und der Beobachtungswinkel φ zwischen Transekt und Objekt gemessen und die senkrechte Entfernung im Anschluß durch $x = y \cdot \sin(\varphi)$ berechnet.

Stellt man die Entfernungen x in Entfernungsintervallen klassiert in Form eines Histogramms dar ohne dabei nach Beobachtungen links- und rechtsseits des Transektes zu unterscheiden, so wird sich, wie Abbildung 2.2 a exemplarisch zeigt, meist eine Abnahme der Anzahl der entdeckten Objekte mit zunehmender Entfernung vom Transekt beobachten lassen. Ein solches Histogramm gibt einen ersten Einblick in den Entdeckungsprozeß, d. h. die Wahrscheinlichkeit mit der ein Objekt unter den aufgetretenen Bedingungen (Vegetation, Beobachter, Wetter usw.) in einer bestimmten Entfernung x vom Transekt entdeckt wird. Für das weitere Vorgehen bedarf es nun eines Modells, das diese Wahrscheinlichkeit in Abhängigkeit von der Entfernung darstellt, der sog. *Entdeckungswahrscheinlichkeitsfunktion* $g(x)$ (kurz: Entdeckungsfunktion), für die ebenfalls ein Beispiel in Abbildung 2.2 a dargestellt ist. Nicht jede beliebige Funktion ist eine zulässige Entdeckungsfunktion; die gängigsten sind in Buckland et al. (2001, Kap. 2

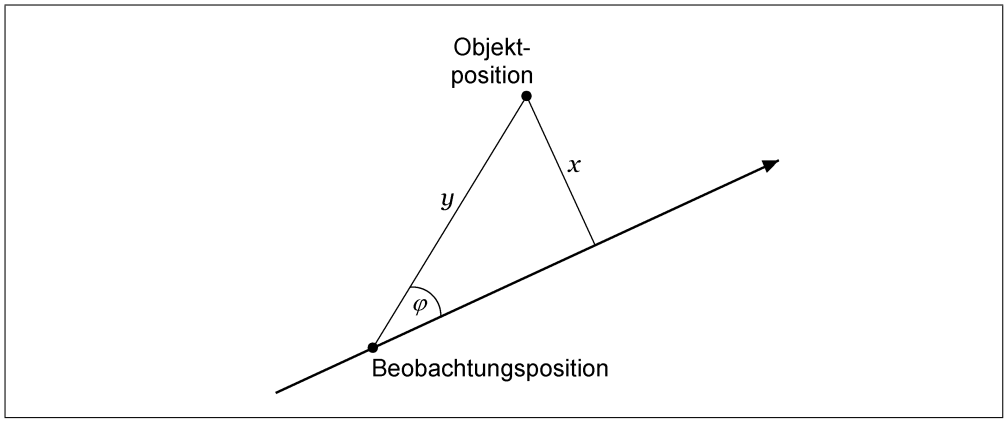

Abbildung 2.1.: Datenerhebung auf gerade verlaufenden Transekten. Sofern sie im Feld nicht direkt gemessen werden kann, wird die senkrechte Entfernung x anhand des Winkels φ und der Entfernung y durch $x = y \cdot \sin(\varphi)$ bestimmt.

u. 3) beschrieben. Sie setzen sich stets aus einer sog. Haupt- oder Schlüsselfunktion und – je nach Bedarf – zusätzlichen Erweiterungstermen zusammen, die die Anpassung an die beobachteten Daten weiter verbessern.

Unter der Annahme, daß zumindest auf und sehr nahe an dem Transekt alle Objekte entdeckt werden ($g(0) = 1$) und im Durchschnitt in jeder Entfernung vom Transekt gleich viele Objekte vorkommen (gewährleistet durch die zufällige Anordnung der Transekte), repräsentiert das graue Rechteck der Größe $1 \cdot w_r$ in Abbildung 2.2 b alle bis zu einer Entfernung w_r vorkommenden Objekte, die Fläche unterhalb der geschätzten Entdeckungsfunktion $\hat{g}(x)$ (schraffierte Fläche), also deren Integral $\int_0^{w_r} \hat{g}(x)\mathrm{d}x$, die hiervon entdeckten. w_r bezeichnet hierbei entweder die größte gemessene oder eine willkürlich festlegbare kleinere Entfernung (s. Kap. 2.1.5.2). $\hat{P}_a = \int_0^{w_r} \hat{g}(x)\mathrm{d}x / w_r$ ist demnach der Anteil der entdeckten an den insgesamt vorhandenen Objekten und beschreibt damit die mittlere (entfernungsunabhängige) Wahrscheinlichkeit, ein Objekt bis zu einer Entfernung w_r zu entdecken. Verwendet man $\hat{P}_a$ als Korrekturfaktor für die übersehenen Objekte, so läßt sich (2.1) zu

$$\hat{D}_s = \frac{n}{2\, w_r\, L\, \hat{P}_a} = \frac{n}{2\, L \int_0^{w_r} \hat{g}(x)\mathrm{d}x} \tag{2.2}$$

umformulieren.

Das Integral $\int_0^{w_r} \hat{g}(x)\mathrm{d}x$ besitzt neben der genannten eine weitere Interpretation als sog. *effektive Streifenbreite* $\hat{v}$ (genauer: halbe effektive Streifenbreite). Sie bezeichnet die Entfernung, bis zu der so viele Objekte übersehen, wie jenseits davon noch entdeckt wurden (die mit ① und ② markierten Flächen in Abb. 2.2 b sind gleich groß). Eine Streifenzählung mit $2 \cdot v$ breiten Streifen würde also (bei gleicher Gesamttransektlänge L) dieselbe Dichteschätzung wie eine Linientransektaufnahme mit einer effektiven Streifenbreite v ergeben. Die effektive Streifenbreite spielt eine entscheidende Rolle bei der in Kapitel 2.3 erläuterten räumlichen Linientransektanalyse.

Die Angabe des Dichteschätzers erfolgt in der Regel nicht in der in (2.2) dargestellten Form, sondern über die zur Entdeckungsfunktion gehörende Wahr-

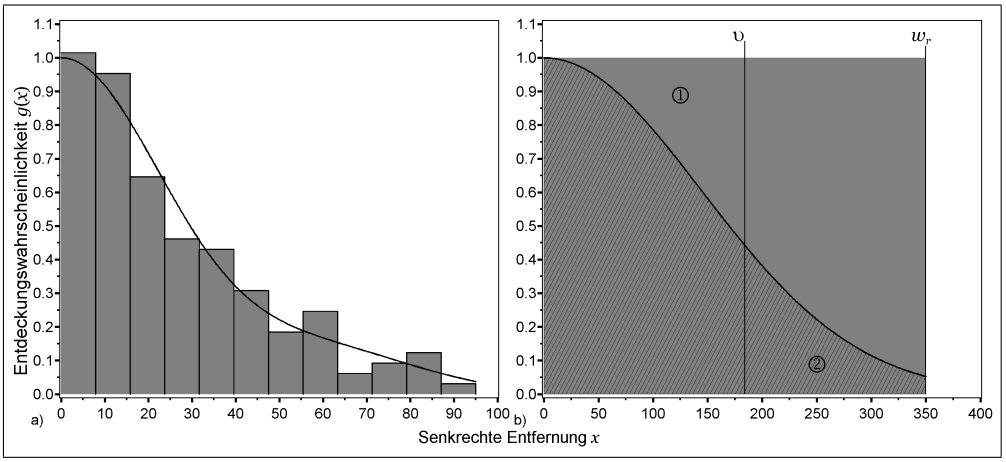

Abbildung 2.2.: a) Histogrammdarstellung der beobachteten Entfernungen und daran angepaßte Entdeckungsfunktion $g(x)$. b) Die effektive Streifenbreite υ bezeichnet die senkrechte Entfernung, bis zu der so viele Objekte übersehen, wie jenseits davon noch entdeckt wurden.

scheinlichkeitsdichtefunktion $f(x)$. Sie berechnet sich aus $g(x)$ durch

$$f(x) = \frac{g(x)}{\int_0^{w_r} g(x)\mathrm{d}x}, \tag{2.3}$$

besitzt damit dieselbe Form wie $g(x)$, aber zusätzlich die Eigenschaft, daß ihr Integral zwischen 0 und w_r gleich Eins ist ($\int_0^{w_r} f(x)\mathrm{d}x = 1$).

Unter der genannten Annahme, daß $g(0) = 1$, die Entdeckung von Objekten auf dem Transekt also sicher ist, folgt dann aus (2.3)

$$f(0) = \frac{1}{\int_0^{w_r} g(x)\mathrm{d}x}$$

und aus (2.2) die allgemeine Form der Darstellung der Objektdichteschätzung von Linientransektaufnahmen

$$\hat{D}_s = \frac{n \cdot \hat{f}(0)}{2\,L}. \tag{2.4}$$

Eine Erweiterung von (2.4) ist notwendig, wenn die Untersuchungsobjekte in Gruppen (in der Literatur oft als *Cluster* bezeichnet) auftreten, also beispielsweise bei Tieren in Form von Rudeln oder Herden. Die Entdeckung eines Individuums ist dann nicht unabhängig von der Entdeckung eines anderen in derselben Gruppe, und damit ist eine wesentliche Voraussetzung der verläßlichen Schätzung der Entdeckungsfunktion, die Unabhängigkeit der zugrundeliegenden Beobachtungen, verletzt. Umgangen wird dieses Problem, indem anstelle der Entfernungen zu den einzelnen Individuen nur eine Entfernung zum Schwerpunkt der Gruppe gemessen wird. $\hat{D}_s$ in (2.4) ist in diesem Falle also der Schätzer der Gruppen- oder Objektdichte.

Definition: Der Begriff *Objekt* bezeichnet allgemein die (z. B. soziale) Einheit der untersuchten Population, zu der Entfernungen gemessen werden. Treten die Individuen der untersuchten Population in Gruppen auf, so sind diese die Objekte. Treten die Individuen hingegen stets einzeln auf, so sind die Individuen selbst die Objekte.

Um auf die Individuendichte zu schließen, werden zusätzlich die beobachteten Gruppengrößen festgehalten, auf Basis derer die mittlere Gruppengröße $\mathrm{E}(s)$ geschätzt und um die (2.4) erweitert wird.

$$\hat{D} = \hat{D}_s \cdot \hat{\mathrm{E}}(s) \tag{2.5}$$

Einfachster Schätzer der mittleren Gruppengröße ist das arithmetische Mittel der beobachteten Gruppengrößen

$$\hat{\mathrm{E}}(s) = \bar{s} = \frac{1}{n} \sum_{i=1}^{n} s_i.$$

Unter bestimmten, aber häufig auftretenden Bedingungen sind jedoch alternative Schätzer vorzuziehen (s. hierzu Kap. 2.1.5.2 u. Buckland et al., 2001, Kap. 3.5 u. 4.8.2).

Ein Schätzer der Individuenabundanz M schließlich ergibt sich durch Multiplikation der Individuendichte mit der Fläche des Untersuchungsgebietes A.

$$\hat{M} = \hat{D} \cdot A$$

Der zugehörige Varianzschätzer ist gegeben durch

$$\mathrm{var}(\hat{M}) = \frac{1}{A^2} \cdot \mathrm{var}(\hat{D}).$$

Die Anpassung von $g(x)$ bzw. $f(x)$ erfolgt gruppiert oder ungruppiert über Maximum-Likelihood-Methoden, die in Kapitel 3 in Buckland et al. (2001) beschrieben sind. Die ungruppierte Analyse verwendet die tatsächlich gemessenen bzw. berechneten Entfernungen x, wohingegen in der gruppierten Auswertung die Anpassung des gewählten Modells an die zu Entfernungsklassen zusammengefaßten Daten erfolgt. Die Bildung dieser Klassen kann bereits vor der Datenerhebung oder im Anschluß daran erfolgen und hat zum Ziel, Auswirkungen verletzter Annahmen (s. nächster Abschnitt) abzuschwächen und ein Histogramm (genauer: eine Datengrundlage) zu erstellen, das sich durch ein möglichst einfaches Modell repräsentieren läßt.

2.1.2. Anpassungsgüte und Modellauswahl

Die Anpassungsgüte eines Modells läßt sich mit Hilfe des χ^2- (Buckland et al., 2001, Kap. 2.3.4 u. 3.4.4), des Kolmogorov-Smirnov- und des Cramér-von Mises-Tests sowie mit q-q-Plots (alle Buckland et al., 2004, Kap. 11.11) beurteilen. Die Bewertung erfolgt aber immer auch visuell anhand der Histogrammdarstellung der Daten und des angepaßten Modells. Die Anpassungsgüte unterschiedlicher

Modelle läßt sich über Likelihood-Ratio-Tests (Buckland et al., 2001, Kap. 3.4.3 u. 4.6.3), Akaikes Informationskriterium (AIC; Buckland et al., 2001, Kap. 3.4.2 u. 4.6.2; Akaike, 1973, 1985) oder dessen Erweiterung für kleine Stichprobenumfänge, AICc (Buckland et al., 2001, ebd.; Hurvich u. Tsai, 1989, 1995), objektiv vergleichen. Hier soll einzig Akaikes Informationskriterium kurz erläutert werden, da es das von Buckland et al. (2001, S. 70) empfohlene Maß zur Modellbeurteilung ist und unten in anderem Zusammenhang noch einmal Erwähnung findet. Es bemißt die Anpassung eines Modells in Abhängigkeit vom erreichten Maximum der log-Likelihood-Funktion ℓ und der Anzahl der dabei verwendeten Parameter p durch

$$\mathrm{AIC} = -2\,\ell + 2\,p. \tag{2.6}$$

Die Entscheidung für eines von mehreren an dieselbe Datenbasis angepaßten Modellen fällt für dasjenige mit dem kleinsten AIC-Wert.

2.1.3. Annahmen und Voraussetzungen

Neben der allgemein für Stichprobenverfahren erforderlichen Voraussetzung der zufälligen oder zumindest systematischen Anordnung der Stichprobeneinheiten (hier der Transekte) sind für das Linientransektverfahren drei wesentliche Annahmen zu erfüllen (Buckland et al., 2001, Kap. 2.1 u. 4.9):

1. Alle Objekte auf und nahe dem Transekt werden entdeckt ($g(0) = 1$).
2. Entfernungen und Winkel werden zur Ursprungsposition der Objekte gemessen.
3. Entfernungen und Winkel werden korrekt erfaßt.

Die erste Annahme ist zugleich die wichtigste. Ein schmaler Streifen um das Transekt und die Anzahl der darin beobachteten Objekte dienen als Referenz, um den Anteil der übersehenen Objekte in entfernteren Distanzklassen schätzen zu können. Streng genommen muß daher die Wahrscheinlichkeit einer Entdeckung auf und nahe dem Transekt nicht gleich 1 sein, es „genügt“, daß sie bekannt ist. Gerade dies kann jedoch oft nur für den Fall $\hat{g}(0) = 1$ behauptet werden. Ist die Wahrscheinlichkeit kleiner, so handelt es sich hierbei fast immer um eine ebenfalls zu schätzende aber nur schwer erfaßbare Größe, mit der sich fortgeschrittene Methoden des Distance Sampling beschäftigen (Buckland et al., 2001, Kap. 7.9.5 und insbesondere Buckland et al., 2004, Kap. 6). Graphisch läßt sich das Problem anhand von Abbildung 2.2 b verdeutlichen. Ist die Höhe des grauen Rechtecks nicht gleich Eins sondern unbekannt, so ist die Fläche $\hat{g}(0) \cdot w_r$ ebenfalls unbekannt und damit auch der Korrekturfaktor $\hat{P}_a$.

Gedanklich handelt es sich bei Linientransekterhebungen um eine Momentaufnahme. Bewegungen der Objekte führen daher zu Verzerrungen der Dichteschätzer, die jedoch ohne größere Bedeutung bleiben, wenn die Bewegungen zufällig, d. h. ungerichtet sind (Yapp, 1956) und die Geschwindigkeit der Objekte klein im Verhältnis zu der des Beobachters ist (Hiby, 1986). Gravierend dagegen können Bewegungen als Reaktion auf den Beobachter (Flucht oder Anziehung) sein. Fliehen viele Objekte unerkannt vor ihrer Entdeckung und wird irrtümlich

die Fluchtposition als Ursprungsposition angenommen, so läßt sich dies oftmals im Histogramm der senkrechten Entfernungen erkennen. Die meisten Beobachtungen werden dann nicht direkt auf oder sehr nahe dem Transekt gemacht, sondern ein Stück weit davon entfernt. Die Balken des Histogramms steigen also zunächst an, bevor sie, wie gewöhnlich, mit größer werdender Entfernung wieder abfallen. An derartige Daten lassen sich die zur Verfügung stehenden Modelle nur schwer anpassen, insbesondere ist aber auch die erste Annahme verletzt: Die Anzahl der tatsächlich nahe dem Transekt gewesenen Objekte ist unbekannt und damit auch $g(0)$.

Zufällige und nicht allzu große Fehler bei der Winkel- und Entfernungsmessung sind im allgemeinen unproblematisch (Gates et al., 1985). Werden Entfernungen aber systematisch zu groß/klein gemessen (bzw. berechnet), so führt dies zu negativ/positiv verzerrten Dichteschätzungen. Die Schätzung von Winkeln und Entfernungen ist zudem problematisch, da es hierbei oftmals zum sog. *Heaping*, dem Runden auf „beliebte“ Werte (z. B. 0°, 45°; 50 m, 100 m) kommt, das sich im Histogramm durch besonders hohe Werte für die entsprechenden Entfernungsklassen bemerkbar macht. Im Idealfall kommen daher Meßgeräte wie Laser-Entfernungsmesser, Maßbänder und Winkelmesser zum Einsatz.

2.1.4. Varianzschätzung

Die Varianz der Individuendichteschätzung läßt sich mit Hilfe der Delta-Methode (auch Gaußsches Fehlerfortpflanzungsgesetz) aus den Varianzen der daran beteiligten Komponenten schätzen (Buckland et al., 2001, Kap. 3.6.1).

$$\begin{aligned} \mathrm{var}(\hat{D}) &= \hat{D}^2 \left\{ \frac{\mathrm{var}(n)}{n^2} + \frac{\mathrm{var}[\hat{f}(0)]}{[\hat{f}(0)]^2} + \frac{\mathrm{var}[\hat{\mathrm{E}}(s)]}{[\hat{\mathrm{E}}(s)]^2} \right\} \\ &= \hat{D}^2 \left\{ [\mathrm{cv}(n)]^2 + [\mathrm{cv}\{\hat{f}(0)\}]^2 + [\mathrm{cv}\{\hat{\mathrm{E}}(s)\}]^2 \right\} \end{aligned} \tag{2.7}$$

und

$$\mathrm{se}(\hat{D}) = \sqrt{\mathrm{var}(\hat{D})} \qquad \mathrm{cv}(\hat{D}) = \frac{\mathrm{se}(\hat{D})}{\hat{D}}.$$

(2.7) verkürzt sich um den Term für die mittlere Gruppengröße, wenn es sich um Objekte handelt, die nicht in Gruppen auftreten.

Entstammen die entdeckten Objekte einer Stichprobe von Objekten, die zufällig, d. h. Poisson-verteilt sind, so ist $\mathrm{Var}(n) = \mathrm{E}(n)$ und $\mathrm{var}(n) = n$. Die meisten biologischen Populationen zeigen aber ein eher geklumptes Auftreten, so daß davon auszugehen ist, daß $\mathrm{Var}(n) > \mathrm{E}(n)$. Anstelle der Annahme einer Poisson-Verteilung sollte $\mathrm{Var}(n)$ daher nach Möglichkeit empirisch anhand der Streuung der Beobachtungsanzahlen je Transekt n_i geschätzt werden.

$$\mathrm{var}(n) = \frac{L \sum_{i=1}^{k} l_i \left[(n_i/l_i) - (n/L)\right]^2}{k-1} \tag{2.8}$$

$$\mathrm{se}(n) = \sqrt{\mathrm{var}(n)} \qquad \mathrm{cv}(n) = \frac{\mathrm{se}(n)}{n}$$

(Buckland et al., 2001, Kap. 3.6.2 u. 4.4). Die Schätzung erfolgt mit $k-1$ Freiheitsgraden.

Oftmals findet sich auch die Darstellung der Begegnungsrate n/L, der mittleren Anzahl entdeckter Objekte je (Transekt-)Längeneinheit. Ihre Varianz wird durch

$$\mathrm{var}(n/L) = \frac{\sum_{i=1}^{k} l_i/L\,(n_i/l_i - n/L)^2}{k-1} \tag{2.9}$$

$$\mathrm{se}(n/L) = \sqrt{\mathrm{var}(n/L)} \qquad \mathrm{cv}(n/L) = \frac{\mathrm{se}(n/L)}{n/L}$$

geschätzt und steht mit der von n im Zusammenhang $\mathrm{var}(n) = L^2 \cdot \mathrm{var}(n/L)$.

Die Parameter der Wahrscheinlichkeitsdichtefunktion $f(x)$, $\theta_1, \ldots, \theta_p$, werden, wie bereits erwähnt, über die Maximierung der zugehörigen Likelihood-Funktion $\mathcal{L}(\boldsymbol{\theta})$ oder der log-Likelihood-Funktion $\ell(\boldsymbol{\theta})$ geschätzt. Der Varianzschätzer einer Funktion $g(\hat{\boldsymbol{\theta}})$ der geschätzten Parameter $\hat{\boldsymbol{\theta}}$ ist dann

$$\mathrm{var}\left[g(\hat{\boldsymbol{\theta}})\right] = \frac{1}{n}\left[\frac{\partial g(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right]' \boldsymbol{F}(\hat{\boldsymbol{\theta}})^{-1}\left[\frac{\partial g(\hat{\boldsymbol{\theta}})}{\partial \hat{\boldsymbol{\theta}}}\right],$$

worin $\boldsymbol{F}(\hat{\boldsymbol{\theta}})$ die geschätzte Fisher-Informationsmatrix der Beobachtungen ist. Deren Einträge hängen von der gewählten Schlüsselfunktion und den Erweiterungstermen sowie von der Art der Auswertung (gruppiert/ungruppiert) ab. Näheres findet sich bei Buckland et al. (2001, Kap. 3.3). Die Schätzung von $f(0)$ erfolgt mit $n-p$ Freiheitsgraden.

Lassen die Daten eine Schätzung der mittleren Gruppengröße $\mathrm{E}(s)$ durch das arithmetische Mittel der beobachteten Gruppengrößen $\bar{s}$ zu, so ergibt sich die geschätzte Varianz mit $n-1$ Freiheitsgraden aus der bekannten Gleichung

$$\mathrm{var}(\bar{s}) = \frac{1}{n}\cdot \mathrm{var}(s) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(s_i - \bar{s})^2 \qquad \text{mit } n = \sum_{i=1}^{k} n_i.$$

$$\mathrm{se}(\bar{s}) = \sqrt{\mathrm{var}(\bar{s})} \qquad \mathrm{cv}(\bar{s}) = \frac{\mathrm{se}(\bar{s})}{\bar{s}}$$

Ein $100(1-\alpha)\%$iges Konfidenzintervall der Dichteschätzung läßt sich unter der Annahme normalverteilter Dichteschätzungen durch $\hat{D} \pm z_{(1-\alpha/2)} \cdot \mathrm{se}(\hat{D})$ berechnen, worin $z_{(1-\alpha/2)}$ das $100(1-\alpha)$-Quantil der Standardnormalverteilung ist. Burnham et al. (1987, S. 211 ff.) empfehlen jedoch wegen der meist rechtsschiefen Verteilung von D die Annahme einer log-Normalverteilung. Untere (LCL) und obere (UCL) Konfidenzgrenzen berechnen sich dann durch

$$\begin{aligned} \mathrm{LCL}(\hat{D}) &= \hat{D}/C \\ \mathrm{UCL}(\hat{D}) &= \hat{D}\cdot C, \end{aligned}$$

wobei

$$C = \exp\left\{z_{(1-\alpha/2)} \cdot \sqrt{\ln[1+\mathrm{cv}(\hat{D})]}\right\} \tag{2.10}$$

(Buckland et al., 2001, Kap. 3.6.1 u. 4.7.4). Dieser Ansatz ist ausreichend, wenn alle an $\hat{D}$ beteiligten Komponenten mit genügend vielen Freiheitsgraden (mindestens 30) geschätzt werden. Andernfalls empfehlen Buckland et al. (2001, Kap. 3.6.1) $z_{(1-\alpha/2)}$ in (2.10) durch einen auf einer t-Verteilung basierenden Wert $t_{(1-\alpha/2;\mathrm{df}_{\hat{D}})}$ zu ersetzen. Die hierfür benötigten Freiheitsgrade berechnen sich nach Satterthwaite (1946; s. auch Milliken u. Johnson, 1984) durch

$$\mathrm{df}_{\hat{D}} = \frac{[\mathrm{cv}(\hat{D})]^4}{[\mathrm{cv}(n)]^4/(k-1) + [\mathrm{cv}\{\hat{f}(0)\}]^4/(n-p) + [\mathrm{cv}\{\hat{\mathrm{E}}(s)\}]^4/(n-1)}. \quad (2.11)$$

2.1.5. Zusätzliche Analyseoptionen

Im folgenden werden einige Besonderheiten bzw. Erweiterungen der oben erläuterten Analyse von Linientransekterhebungen dargestellt, die für die Beantwortung der Fragestellungen dieser Arbeit bzw. aufgrund der besonderen Datengrundlage benötigt wurden.

2.1.5.1. Kurvige Transekte

Handelt es sich bei den Transekten – wie in den bisherigen Ausführungen angenommen – um gerade Linien, so lassen sich die senkrechten (und damit kürzesten) Entfernungen x zwischen Objekt und Transekt, wie erläutert, einfach aus der zwischen Beobachter und Objekt gemessenen Distanz y und dem Winkel zwischen Transekt und Objekt φ durch $x = y \cdot \sin(\varphi)$ berechnen (vergl. Abb. 2.1). Diese Berechnung ist insbesondere völlig unabhängig von der Position des Beobachters zum Zeitpunkt der Entdeckung, d. h. es ist unerheblich, wo entlang des Transektes die Entdeckung gemacht wurde. Für eine gewöhnliche Analyse gerader Transekte müssen lediglich besagte Entfernungen x, Anzahl k und Länge der Transekte l_i sowie die Anzahl der darauf gemachten Beobachtungen n_i bekannt sein. Geographische Positionen spielen in diesem Zusammenhang zunächst keine Rolle.

Definitionen: Die *Objektposition* bezeichnet, in geographischen oder kartesischen Koordinaten, die Stelle, an der das Objekt, d. h. z. B. ein Einzeltier oder eine Gruppe zusammengehöriger Tiere, zum ersten Mal entdeckt wurde, oder dessen vermutete Ursprungsposition vor einer Flucht infolge der Annäherung des Beobachters. Die *Beobachtungsposition* bezeichnet die Stelle (wiederum in geographischen oder kartesischen Koordinaten), von der aus Beobachtungswinkel (φ oder ω) und Entfernungen gemessen wurden und ist meist, aber nicht notwendigerweise[1], identisch mit der Position, von der aus das Objekt zuerst beobachtet wurde. Des weiteren von Bedeutung ist die *Transektposition*; sie gibt die Stelle auf dem Transekt an, zu der die kürzeste Entfernung gemessen wurde und ist insbesondere bei einer anschließenden räumlichen Analyse nach dem Count-Modell von Bedeutung (s. Kap. 2.3).

[1] Für die Bestimmung der kürzesten Entfernung x ist nicht die Beobachtungs- sondern lediglich die Objektposition von Bedeutung. Kann z. B. die Objektposition von der Position

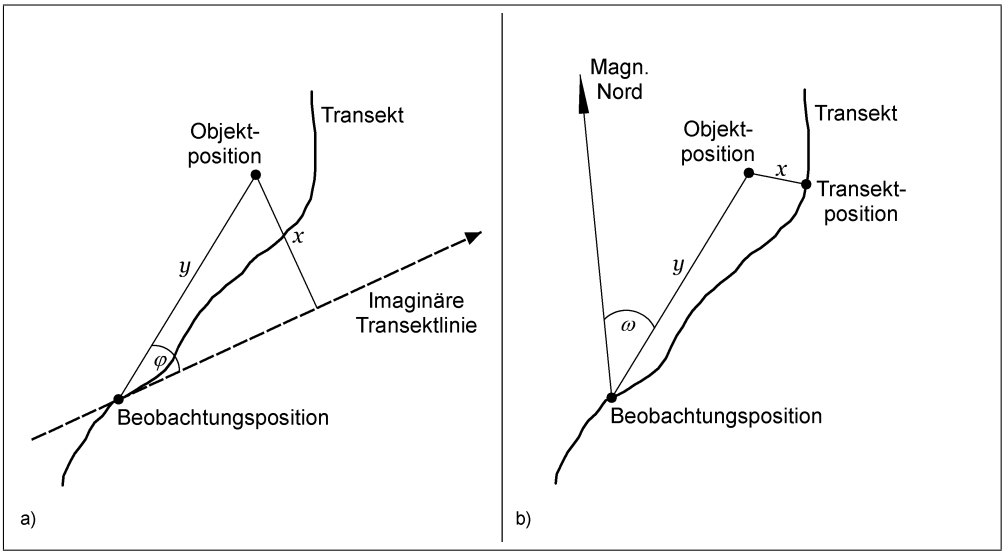

Abbildung 2.3.: a) Abweichung zwischen tatsächlichem Abstand eines Objektes vom Transekt und dem per $x = y \cdot \sin(\varphi)$ berechneten. b) Datenerfassung auf kurvigen Transekten: Erfaßt werden der Winkel ω, die Entfernung y und die Beobachtungsposition (GPS), auf Basis derer die Objektposition bestimmt wird. Deren kürzeste Entfernung zum Transekt x wird anschließend mit Hilfe von GIS-Funktionen bestimmt.

Etwas komplizierter gestaltet sich die Ermittlung der senkrechten Entfernungen bei nicht gerade verlaufenden Transekten. In Abbildung 2.3 a ist ein Beispiel für die Abweichung zwischen der per $y \cdot \sin(\varphi)$ berechneten und der tatsächlichen Entfernung dargestellt. Die Abweichung wird einzig durch den weiteren Verlauf des Transektes beeinflußt, so daß anhand der erfaßten Entfernung y nicht darauf geschlossen werden kann. Nicht zu unterschätzen ist in diesem Zusammenhang auch die Schwierigkeit, den Winkel φ auf kurvigen Transekten korrekt zu bestimmen.

Wenngleich Abbildung 2.3 a die offensichtliche Abweichung der berechneten Entfernung von der richtigen deutlich macht, so stellt sich doch die Frage, welche Entfernung denn die richtige ist. Zu welchem Punkt auf dem Transekt soll die Entfernung gemessen werden? Einem Ansatz von Hiby u. Krishna (2001) folgend, wurde hier als „richtig“ die insgesamt kürzeste Entfernung eines Objektes zu dem Transekt, von dem aus die Beobachtung gemacht wurde, definiert, unabhängig von der Beobachtungsposition. Da aber der weitere (und ggf. vorhergehende) Verlauf des Transektes oft nicht einsehbar oder die Objektposition vom vermuteten nächstgelegenen Punkt auf dem Transekt evtl. nicht wiederzuerkennen ist, ist diese Entfernung im Feld meist nur sehr schwer zu erfassen. Einfacher und zuverlässiger ist die Berechnung mit Hilfe eines GIS (Geographisches Informationssystem[2]). Hierfür ist dann jedoch die Kenntnis der Lage des

der ersten Entdeckung des Objektes aus nicht exakt eingemessen werden, so kann auch jede beliebige andere Position als Beobachtungsposition verwendet werden. Im Vordergrund steht hierbei stets die möglichst genaue Bestimmung der Objektposition.

[2]Ein geographisches Informationssystem besteht aus Computer-Software zur Erstellung, Edi-

Transektes und der Objektposition erforderlich. Der Transektverlauf kann z. B. per GPS (*Global Positioning System*) aufgezeichnet oder aus Luftbildern digitalisiert werden, die Bestimmung der Objektposition dagegen gestaltet sich etwas schwieriger. Das Anlaufen der Position ist meist zu zeitaufwendig und dürfte zudem durch die Beunruhigung in den meisten Fällen Einfluß auf den weiteren Entdeckungsprozeß nehmen. Der Ansatz, der hier gewählt wurde, berechnet die Objektposition anhand der zwischen ihr und der Beobachtungsposition gemessenen Entfernung y (identisch zur Aufnahme bei geraden Transekten) und dem zum geographischen Nordpol bestimmten Winkel ω^* (s. auch die Beschreibung der Datenaufnahme in Kap. 5.1). Letzterer ergibt sich aus dem im Feld erfaßten Winkel zwischen Objekt und magnetisch Nord ω und – sofern von nennenswerter Bedeutung – der aktuellen örtlichen Deklination (Nadelmißweisung). x- und y-Koordinaten der Objektpositionen lassen sich im Anschluß leicht über trigonometrische Funktionen berechnen, wobei, wie in Anhang C.1 erläutert, die Umrechnung von geographischen zu mathematischen Winkeln und von Grad ins Bogenmaß zu berücksichtigen ist. Ausgehend von der berechneten Objektposition und der geographischen Lage der Transekte, läßt sich im Anschluß mit Hilfe gängiger GIS-Funktionalität (s. Kap. 3.1) die kürzeste Entfernung x zwischen Objektposition und Transekt bestimmen. Diese geht dann in die weiteren Analysen ein, die ansonsten identisch zu denen gerader Transekte sind (eine einzelne, die Größe der Stichprobenfläche betreffende Ausnahme hiervon wird weiter unten erläutert).

2.1.5.2. Datenbeschneidung

Fast immer empfiehlt es sich, nicht alle erhobenen Daten (Entfernungen x) in die Analyse mit einzubeziehen. Für Objekte, die in Gruppen auftreten, betrifft dies zunächst die Schätzung der mittleren Gruppengröße: In großen Entfernungen werden oftmals auch zunehmend nur noch größere Gruppen beobachtet, kleine Gruppen und einzelne Individuen dagegen oftmals übersehen. Das arithmetische Mittel der Gruppengröße aus (2.1.1) ist dann mit einem positiven Fehler behaftet und damit kein geeigneter Schätzer mehr für $\mathrm{E}(s)$. Einfachste Lösungsmöglichkeit dieses Problems ist, so viele Beobachtungen, mit der größten Entfernung beginnend, von der Berechnung von $\bar{s}$ auszuschließen, daß sich ein Zusammenhang zwischen s und x (z. B. mittels Regression) nicht mehr nachweisen läßt. Aufwendigere, aber nicht notwendigerweise auf das Entfernen von Beobachtungen angewiesene Alternativen werden in Buckland et al. (2001, Kap. 3.5) erläutert.

Ein weiterer Grund, Daten am rechten Ende der Verteilung zu entfernen (*Right Truncation*), liegt in der Anpassung der Entdeckungsfunktion, die durch Ausreißer, einzelne Beobachtungen in sehr großer Entfernung, erschwert wird. Dies kann sich in einer erhöhten Anzahl benötigter Erweiterungsterme und damit letztlich in breiteren Konfidenzintervallen der Schätzungen bemerkbar machen. Buckland et al. (2001, S. 104) empfehlen daher pauschal die größten 5-10 %

tierung, Verwaltung, Analyse und (z. B. kartographischen) Ausgabe raumbezogener Daten und den zu bearbeitenden geographischen Daten (z. B. Bill u. Fritsch, 1991, S. 5; ESRI, 2006).

der Entfernungen oder besser noch diejenigen zu entfernen, für die $\hat{g}(x) < 0.15$, um Probleme der Modellanpassung durch Ausreißer zu umgehen. Es gehen dann nur alle bis zur (rechten) Kappungsentfernung w_r gemachten Beobachtungen in die Analyse ein.

Aber auch für Beobachtungen, die sehr nahe am Transekt liegen, kann es, wenn auch weit seltener, Beweggründe geben, diese vor einer Analyse zu entfernen (*Left Truncation*), wenngleich dies den oben aufgeführten Annahmen und Voraussetzungen zu widersprechen scheint. So kann beispielsweise beim Befliegen von Transekten der Bereich direkt unterhalb des Flugzeugs nur schwer oder gar nicht eingesehen werden, dafür läßt sich aber evtl. die These $g(x) \approx 1$ für ein kleines x aufrecht erhalten, wohingegen $g(0) < 1$. Ist diese Entfernung w_l nur gering, so kann die Anpassung der Entdeckungsfunktion an die Daten mit $x > w_l$ (und $x < w_r$) und die Extrapolation zurück zu $x = 0$ dennoch gültige Schätzungen liefern (Alldredge u. Gates, 1985). Alternativ hierzu läßt sich (gedanklich) auch die Transektlinie um die Entfernung w_l verschieben. Hierzu wird von allen gemessenen Entfernungen w_l subtrahiert und die Daten anschließend wie gewohnt ausgewertet. Diese Variante setzt allerdings voraus, daß $\hat{g}(w_l) = 1$.

Auch wenn übermäßig viele Beobachtungen auf oder nahe dem Transekt liegen, kann die linksseitige Beschneidung der Daten sinnvoll sein, wenn nämlich zu vermuten ist, daß ein schmaler Bereich um das Transekt herum wenig repräsentativ für das eigentliche Untersuchungsgebiet ist. Dies kann z. B. bei der Verwendung von Wegen oder Schneisen als Transekte der Fall sein, wo Objekte beim Wechsel darüber auch in sehr großer Entfernung entdeckt werden, wohingegen die Entdeckung von Objekten abseits des Weges nur bis zu deutlich kleineren Entfernungen möglich ist. Das Transekt selbst ist hier nicht repräsentativ für das Untersuchungsgebiet, und die Schätzung der Entdeckungsfunktion und der darauf beruhenden Parameter sind vermutlich verläßlicher und für das gesamte Untersuchungsgebiet aussagekräftiger, wenn die direkt auf dem Transekt gesammelten Daten von der Analyse ausgeschlossen werden.

Buckland et al. (2001, S. 108) weisen jedoch ausdrücklich darauf hin, in jedem Fall die möglichen Ursachen von Auffälligkeiten im Entfernungshistogramm zu ergründen, ehe die linksseitige Datenbeschneidung in Erwägung gezogen wird. So liefert die Methode der Extrapolation (Alldredge u. Gates, 1985) Dichteschätzungen, die sehr stark abhängig vom gewählten Modell der Entdeckungsfunktion sein können (Buckland et al., 2001, S. 16).

Weitere Einzelheiten zur Datenbeschneidung finden sich in Buckland et al. (2001, Kap. 4.3).

2.1.5.3. Wiederholte Aufnahme permanenter Transekte

Die wiederholte Aufnahme permanent angelegter Transekte läßt sich auch als Sonderfall einer stratifizierten Erhebung betrachten, in der die Straten durch die zeitlichen Wiederholungen gegeben sind und zudem in allen Straten (Wiederholungen) derselbe Aufwand (dieselbe Gesamttransektlänge) betrieben wurde. Mit der Durchführung derartiger Wiederholungsaufnahmen wird in der Regel wenigstens eines der drei folgenden Ziele verfolgt. Zunächst können die Dichteschätzungen der einzelnen Aufnahmen (Straten) der Untersuchung eines zeit-

lichen Trends, also der Darstellung von Zu- und Abgängen dienen. Zum anderen kann der Mittelwert $\hat{\bar{D}} = \sum_{h=1}^{H} \hat{D}_h / H$ der H Dichteschätzungen unter bestimmten Bedingungen eine genauere und präzisere Schätzung liefern. Die Verläßlichkeit der mittleren Dichteschätzung $\hat{\bar{D}}$ ist jedoch an die Bedingung gekoppelt, daß zwischen den einzelnen Aufnahmen keine Migration stattgefunden hat und sonstige Zu- oder Abgänge nicht vorhanden oder vom selben Betrag sind. Zum dritten kann es vorkommen, daß in einer einzelnen Aufnahme die Mindestanzahl der für eine Schätzung der Entdeckungsfunktion erforderlichen Beobachtungen von 60-80 (Buckland et al., 2001, S. 14) nicht erreicht wird. In diesem Fall läßt sich prüfen, ob die Entdeckungsfunktion über mehrere oder alle Wiederholungsaufnahmen (Straten) hinweg (*gepoolt*) geschätzt werden kann.

Die Untersuchung zeitlicher Trends wird hier nicht weiter verfolgt (s. hierfür Buckland et al., 2001, Kap. 8.4), einige Besonderheiten, die sich für die beiden anderen Ziele im Zuge der Linientransektanalyse ergeben, sollen jedoch im folgenden kurz erläutert werden.

Wird die Entdeckungsfunktion über alle Wiederholungsaufnahmen (Straten) hinweg geschätzt ($\hat{g}_{pool}(x)$), so ergibt sich die Dichteschätzung eines einzelnen Stratums h durch

$$\hat{D}_h = \frac{\hat{f}(0)_{pool}\, n_h\, \bar{s}_h}{2\, L_h}. \tag{2.12}$$

n_h, $\bar{s}_h$ und L_h werden hier stratenweise angegeben, da sich Begegnungsraten und mittlere Gruppengrößen in der Regel auch schon für kleinere Stichprobenumfänge recht präzise schätzen lassen. Voraussetzung für die Verwendung einer gemeinsamen $f(0)$-Schätzung ist jedoch, daß die Entdeckungsfunktionen der Straten tatsächlich gleich sind, oder, genauer ausgedrückt, die Beobachtungen der unterschiedlichen Straten demselben Entdeckungsprozeß entstammen. Buckland et al. (2001, S. 111) und Buckland et al. (1993, S. 374) nennen zur Überprüfung dieser Annahme zwei Verfahren. In beiden wird zunächst ein beliebiges Modell an den gepoolten Datensatz angepaßt und dasselbe Modell (d. h. Art und Anzahl der Erweiterungsterme sind fix) anschließend für die Anpassung an die Stratendaten verwendet. Die Maxima der log-Likelihood-Funktionen des gemeinsamen (*gepoolten*) Datensatzes ℓ_{pool} und der einzelnen Straten ℓ_h lassen sich dann über einen Likelihood-Ratio-Test

$$\chi^2 = 2\left[\left(\sum_{h=1}^{H} \ell_h\right) - \ell_{pool}\right]$$

auswerten. Unter der Nullhypothese eines einheitlichen Entdeckungsprozesses ist die Teststatistik χ^2-verteilt mit $p(H-1)$ Freiheitsgraden, wobei p die Anzahl der Modellparameter ist. Im zweiten und meist bevorzugten Ansatz werden die durch (2.6) bestimmten AIC-Werte der stratenweise angepaßten Modelle aufsummiert und die Summe mit dem AIC-Wert des gepoolten Datensatzes verglichen. Ist die Summe kleiner als der AIC-Wert der gepoolten Analyse, so sollte von unterschiedlichen Entdeckungsprozessen ausgegangen und die Entdeckungsfunktion stratenweise geschätzt werden.

Für eine allgemeine Abhandlung der (Post-)Stratifizierung, insbesondere auch

für die Auswertung geographischer Straten sowie die Auswirkung auf die Varianzschätzung s. Buckland et al. (2001, Kap. 3.6.2 u. 3.7.1).

Im Falle der Berechnung einer mittleren Dichteschätzung anhand der einzelnen Aufnahmen, ist die Tatsache, daß es sich um wiederholte Aufnahmen derselben Transekte handelt auch in der Varianzschätzung von $\hat{D}_s$ bzw. $\hat{D}$ zu berücksichtigen. Die Transekte und damit die Ergebnisse einer Aufnahme sind dann nicht unabhängig von denen einer anderen. Im vereinfachten Fall, in dem alle Transekte gleich oft, H mal, erfaßt wurden, ist der Varianzschätzer für n gegeben durch

$$\mathrm{var}_{pt}(n) = \frac{L \sum_{i=1}^{k} l_i \, (n_i/l_i - n/L)^2}{k-1}. \tag{2.13}$$

Auf den ersten Blick ist dieser identisch zu (2.8), jedoch entspricht n_i in (2.8) der Anzahl der Beobachtungen auf dem i-ten Transekt *einer* Aufnahme, in (2.13) dagegen der Anzahl aller in den H Aufnahmen vom i-ten Transekt gemachten Beobachtungen. Entsprechend handelt es sich bei l_i und L um die H-fache Länge der ursprünglichen Länge eines bzw. aller Transekte. k besitzt in beiden Gleichungen dieselbe Bedeutung der Anzahl der Transekte. Die Anzahl der Freiheitsgrade $(k-1)$ ist ebenfalls in beiden Fällen identisch. Die Varianzschätzung für n bzw. n/L erfolgt also, als wären nicht $k \cdot H$ Transekte aufgenommen worden, sondern k Transekte der Länge $H \cdot l_i$. Im Falle der mittleren Dichteschätzung anhand wiederholter Aufnahmen ist also $\mathrm{var}(n)$ in (2.7) durch $\mathrm{var}_{pt}(n)$ zu ersetzen.

2.1.5.4. Korrektur der Dichteschätzer bei abweichender Stichprobenfläche

Liegen Transekte sehr nahe an der Grenze des Untersuchungsgebietes oder verlaufen diese von einem Ende des Gebietes zum anderen, so können Teilflächen existieren, deren Entfernung zum Transekt kleiner als w_r ist, die aber dennoch nicht aufgenommen werden, weil sie außerhalb des Untersuchungsgebietes liegen (s. Abb. 2.4 a). Die tatsächlich erfaßte Fläche ist also kleiner als die in der Gleichung des Dichteschätzers (2.4) verwendete, $a = 2\,w_r\,L$. Ähnlich verhält es sich, wie in Abbildung 2.4 b dargestellt, bei kurvigen Transekten. Die Stichprobenfläche, die sich durch die Erstellung eines Streifens der Breite $2 \cdot w_r$ um die Transektlinien ergibt, ist nicht notwendigerweise gleich der eines geraden Transektes derselben Länge. Ist der Anteil von Flächen außerhalb des Untersuchungsgebietes bzw. der Anteil gekrümmter Transektabschnitte groß, so sollte dies in der Berechnung der Dichteschätzer und deren Streuungsmaße berücksichtigt werden.

Sei $\tilde{a}$ die tatsächlich aufgenommene (aufnehmbare) Stichprobenfläche, so ergibt sich ein einfacher, korrigierter Schätzer der Individuendichte D durch

$$\hat{D}_{\tilde{a}} = \hat{D}\,\frac{a}{\tilde{a}} = \hat{D}\,\frac{2\,L\,w_r}{\tilde{a}}$$

und die zugehörige Varianzschätzung durch

$$\mathrm{var}(\hat{D}_{\tilde{a}}) = \mathrm{var}(\hat{D}) \cdot \left(\frac{a}{\tilde{a}}\right)^2.$$

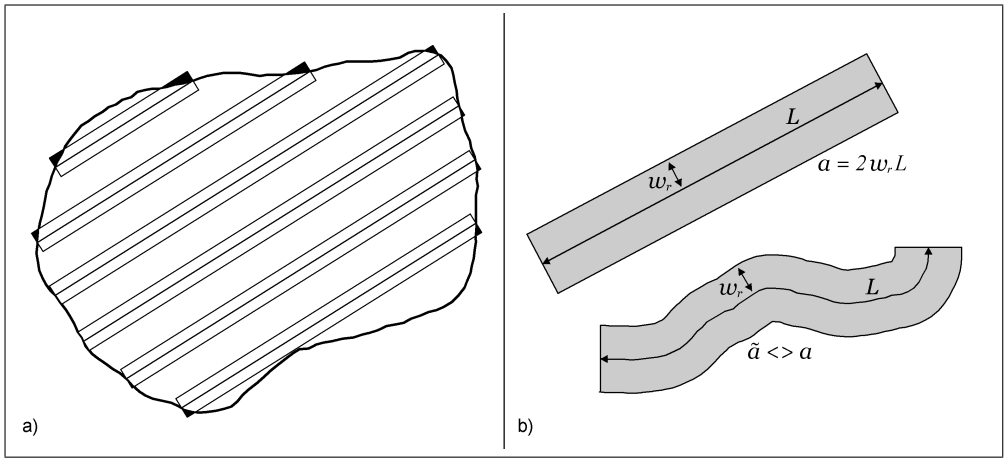

Abbildung 2.4.: Unvollständige Erfassung der Stichprobenflächen: Verlaufen die Transekte von einem Ende des Untersuchungsgebietes zum anderen, so kann es Bereiche geben, die nicht erfaßt werden, obwohl sie innerhalb der Kappungsentfernung w_r liegen (a). Ähnliches gilt bei kurvigen Transekten; hier ist die tatsächlich aufgenommene Fläche nicht notwendigerweise gleich der eines geraden Transektes gleicher Länge (b).

$\tilde{a}$ läßt sich z. B. mit geeigneter GIS-Software bestimmen, indem Streifen (sog. *Buffer*) der Breite $2 \cdot w_r$ um die Transekte gelegt werden und die Teile außerhalb des Untersuchungsgebietes „abgeschnitten“ werden.

Eine komplexere Korrekturmöglichkeit, die die Schätzung von $f(x)$ beeinflußt, wird in Buckland et al. (2001, Kap. 6.7) vorgestellt.

2.1.5.5. Vergleich zweier Dichteschätzungen

Ein vermuteter Unterschied zwischen zwei voneinander unabhängigen Dichteschätzungen, $\hat{D}_1$ und $\hat{D}_2$, läßt sich mit Hilfe eines einfachen T-Tests überprüfen. Die geschätzte Varianz der Differenz beider Schätzungen ist

$$\mathrm{var}(\hat{D}_1 - \hat{D}_2) = \mathrm{var}(\hat{D}_1) + \mathrm{var}(\hat{D}_2)$$

und die Teststatistik

$$t = \frac{\hat{D}_1 - \hat{D}_2}{\sqrt{\mathrm{var}(\hat{D}_1 - \hat{D}_2)}}. \tag{2.14}$$

Die zugehörigen Freiheitsgrade lassen sich durch

$$\mathrm{df}_{(\hat{D}_1 - \hat{D}_2)} = \frac{[\mathrm{var}(\hat{D}_1) + \mathrm{var}(\hat{D}_2)]^2}{[\mathrm{var}(\hat{D}_1)]^2/\mathrm{df}_{\hat{D}_1} + [\mathrm{var}(\hat{D}_2)]^2/\mathrm{df}_{\hat{D}_2}} \tag{2.15}$$

bestimmen, die der beiden Dichteschätzungen, $\mathrm{df}_{\hat{D}_1}$ und $\mathrm{df}_{\hat{D}_2}$, über die Satterthwaite-Approximation aus (2.11). Die Nullhypothese gleicher Dichten ($H_0 : D_1 = D_2$) wird abgelehnt, wenn der Betrag der berechneten Teststatistik größer als das $100(1-\alpha)$%-Perzentil einer entsprechenden t-Verteilung ist.

$$|t| > t_{(1-\alpha/2;\mathrm{df}_{(\hat{D}_1 - \hat{D}_2)})}$$

Alternativ kann auch das $100(1-\alpha)\%$ige Konfidenzintervall der Differenz der Dichteschätzungen

$$(\hat{D}_1 - \hat{D}_2) \pm t_{(1-\alpha/2;\mathrm{df}_{(\hat{D}_1-\hat{D}_2)})} \cdot \sqrt{\mathrm{var}(\hat{D}_1 - \hat{D}_2)}$$

interpretiert werden.

Auch diese Gleichungen sind anzupassen, wenn für beide Dichteschätzungen Parameter gemeinsam geschätzt werden (z. B. $\hat{f}_{pool}(0)$); s. hierzu Buckland et al. (2001, Kap. 3.6.5).

2.2. Verallgemeinerte lineare Modelle (GLM)

In vielen Untersuchungen werden lineare Modelle der Form

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i$$

verwendet, um Vorhersagen zu ermöglichen oder Schlußfolgerungen über die Bedeutung potentiell erklärender Variablen zu erlangen. Hierbei ist y_i der Wert der Untersuchungsvariablen für die i-te Beobachtung und $\boldsymbol{x}_i$ ein Spaltenvektor potentiell erklärender, bekannter Variablen für Beobachtung i. Basierend auf den Beobachtungen y_i wird der Vektor unbekannter Parameter $\boldsymbol{\beta}$ über die Methode der kleinsten Quadrate geschätzt. Als Abweichung zwischen beobachtetem und vom Modell vorhergesagtem Wert ergibt sich das Residuum ϵ_i.

$\boldsymbol{x}_i'\boldsymbol{\beta}$ modelliert den Erwartungswert $\mathrm{E}(Y_i) = \mu_i$ der Zufallsvariablen Y_i. In Matrixschreibweise ausgedrückt ist also

$$\boldsymbol{\mu}_{(n\times 1)} = \boldsymbol{X}_{(n\times p)}\boldsymbol{\beta}_{(p\times 1)}.$$

$\boldsymbol{X}$ ist hierin die aus den $\boldsymbol{x}_i'$ zusammengesetzte *Designmatrix*. Zu den gängigsten Verfahren, die diese Art von Modellen verwenden, gehören Regressions-, Varianz- (ANOVA) und Kovarianzanalyse (ANCOVA).

Schlußfolgerungen in Form von Hypothesentests bezüglich der geschätzten Parameter eines solchen Modells bedürfen jedoch der Annahme der Unabhängigkeit der Beobachtungen sowie der Normalverteilung und Varianzhomogenität der Untersuchungsvariablen, d. h. $Y_i \sim N(\mu_i, \sigma^2)$ bzw. der Residuen $\epsilon_i \sim N(0, \sigma^2)$. Insbesondere die Annahmen der Normalverteilung und der Varianzhomogenität der Untersuchungsvariablen sind aber vielfach nicht erfüllt. Vielmehr zeigen Beobachtungen größerer Dimension oftmals eine stärkere Streuung als die kleinerer Dimension, die Varianz ist hier also nicht konstant sondern eine Funktion des Mittelwertes. Ebenso schließen Art und Skalierung der Untersuchungsvariablen häufig bereits die Annahme der Normalverteilung aus. So sind beispielsweise Anzahlen stets auf ganzzahlige Werte größer oder gleich Null beschränkt und folgen oft einer Poisson-Verteilung. Erfolgswahrscheinlichkeiten, bemessen als Quotient aus Anzahl der Erfolge und Anzahl der Versuche sind in der Regel binomialverteilt und können nur Werte im Intervall 0 bis 1 annehmen. Das lineare Modell $\boldsymbol{X\beta}$ ist jedoch nicht auf bestimmte Wertebereiche begrenzt, sondern nimmt (theoretisch) den gesamten Bereich der reellen Zahlen ein, so daß

für Daten mit begrenztem Wertespektrum unsinnige Vorhersagen außerhalb des gültigen Bereiches möglich sind. Eine Möglichkeit, dieses Problem zu beheben, sind Transformationen, die jedoch meist vordergründig auf die Homogenisierung der Varianzen abzielen, das Ziel, die Schiefe der Verteilung der Daten in Symmetrie zu ändern, jedoch oft verfehlen.

Eine zweite Möglichkeit, lineare Modelle für die Analyse nicht normalverteilter Daten, die aus einer zur Exponentialfamilie gehörenden Verteilung stammen, zu verwenden, bieten *verallgemeinerte lineare Modelle* (*Generalized Linear Models*, GLMs; Nelder u. Wedderburn, 1972; McCullagh u. Nelder, 1989). Sie folgen der allgemeinen Form

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta}.$$

Die Exponentialfamilie umfaßt viele bekannte Verteilungen wie die Normal-, die Poisson-, die Binomial- und die Gamma-Verteilung.

GLMs bestehen aus den drei folgenden, wesentlichen Bestandteilen:

1. Die lineare Komponente

   $$\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}.$$

   Diese ist identisch zum gewöhnlichen linearen Modell, d. h. der (lineare) Prädiktor η_i ergibt sich aus einer Linearkombination der erklärenden Variablen $\boldsymbol{x}_i'$.

2. Die Link-Funktion $g(\cdot)$

   $$g(\mu_i) = \eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}.$$

   Da sich die meisten nicht normalen Verteilungen nicht linear bezüglich ihres Erwartungswertes verhalten, ist die Grundidee der GLMs, das lineare Modell nicht direkt an den Mittelwert μ anzupassen, sondern an eine (nichtlineare) Funktion davon – die Link-Funktion. Sie stellt die Beziehung zwischen dem Erwartungswert der Zufallsvariable und dem linearen Prädiktor η her. Es handelt sich dabei gewöhnlich um ein einfaches mathematisches Modell (eine monotone differenzierbare Funktion). Verallgemeinerte additive Modelle (*Generalized Additive Model*, GAM; Hastie u. Tibshirani, 1990) stellen eine sehr flexible Erweiterung dar, in der anstelle des linearen Prädiktors $\boldsymbol{x}_i'\boldsymbol{\beta}$ die Summe allgemeiner Funktionen der unabhängigen Variablen verwendetet werden ($\sum_{j=0}^{p-1} f_j(x_j)$ statt $\sum_{j=0}^{p-1} \beta_j x_j$).

   Für $g(\mu) = \mu$ (sog. *Identity Link*) reduziert sich das GLM zum gewöhnlichen linearen Modell.

   Die inverse Link-Funktion $g^{-1}(\cdot)$ ist die Umkehrfunktion der Link-Funktion, über die die Vorhersagen auf der Link-Skala auf die ursprüngliche Skala zurückgerechnet werden, also $g^{-1}(\eta) = \mu$.

3. Die Varianzfunktion $v(\cdot)$. Die Varianz einer Beobachtung ist gegeben durch

   $$\mathrm{var}(y_i) = v(\mu_i) \cdot a_i(\phi) = \frac{\phi \cdot v(\mu_i)}{w_i},$$

wobei ϕ ein konstanter bekannter oder zu schätzender Skalierungs- oder Dispersionsparameter und w_i ein beobachtungsspezifisches, bekanntes Gewicht ist. Durch die Varianzfunktion $v(\mu)$ läßt sich die verteilungsspezifische Abhängigkeit der Varianz vom Mittelwert in das Modell integrieren und so inhomogene Varianzen berücksichtigen.

In den folgenden Abschnitten sollen die Eigenschaften und Erweiterungen von GLMs dahingehend näher erläutert werden, wie sie für die Analyse und Vorhersage räumlich korrelierter Anzahldaten genutzt werden können. Zunächst soll hierzu ein Blick auf die Eigenschaften der Verteilungen der Exponentialfamilie sowie auf die Link-Funktion geworfen werden. Darauf aufbauend wird die Schätzung des unbekannten Parametervektors $\boldsymbol{\beta}$ über die Maximum-Likelihood-Methode, die Beurteilung der Anpassungsgüte des Modells und die Berechnung zugehöriger Streuungsmaße und Hypothesentests erläutert. Über *Quasi-Likelihood* wird dann die Anwendung von GLMs für Verteilungen, die nicht zur Exponentialfamilie gehören, erweitert. Quasi-Likelihood ist die Grundlage u. a. für die Einbeziehung sog. *Überdispersion* und für die Berücksichtigung abhängiger Beobachtungen. Letztere lassen sich mit Hilfe von *Generalized Estimating Equations* (GEEs) analysieren, die zunächst allgemein und abschließend für die Berücksichtung räumlich korrelierter Daten eingeführt werden. Des weiteren werden ein äquivalenter Ansatz, das sog. Pseudo-Likelihood, und Möglichkeiten der räumlichen Vorhersage vorgestellt.

Umfangreiche allgemeine Einführungen in die Theorie und Anwendung verallgemeinerter linearer Modelle liefern u. a. Aitkin et al. (1989), Dobson (2001) und Firth (1991); auf Literatur zu den speziellen Anwendungen und insbesondere der räumlichen Autokorrelation wird in den entsprechenden Abschnitten verwiesen.

2.2.1. Die Exponentialfamilie

Läßt sich die Wahrscheinlichkeitsverteilung einer Zufallsvariablen Y in der Form

$$f(y;\theta,\phi) = \exp\left\{\frac{y\,\theta - b(\theta)}{a(\phi)} + c(y;a(\phi))\right\} \tag{2.16}$$

darstellen, so ist die Verteilung Mitglied der *Exponentialfamilie*. a, b und c sind verteilungsspezifische Funktionen, wobei $a(\phi) = \phi/w$ und $c = c(y;\phi/w)$. ϕ wiederum ist der bereits erwähnte Skalierungs- oder Dispersionsparameter und w ein beobachtungsspezifisches Gewicht.

Es läßt sich zeigen, daß

$$\mathrm{E}(Y) = b'(\theta) \tag{2.17}$$

und

$$\mathrm{var}(Y) = b''(\theta)\cdot a(\phi) = v(\mu)\cdot a(\phi).$$

Die Varianz setzt sich also aus zwei Funktionen zusammen. Die erste, oft als Varianzfunktion $v(\mu)$ bezeichnet, hängt vom natürlichen Parameter θ der Verteilung und damit vom Mittelwert μ ab, da $\theta = \theta(\mu)$, die zweite Funktion lediglich vom Dispersionsparameter ϕ.

So wie sich normalverteilte Daten linear bezüglich der unabhängigen Variablen verhalten, gilt dies allgemein für Beobachtungen aus Verteilungen der Exponentialfamilie für den natürlichen Parameter θ. Der sog. *kanonische Link* ist daher die Link-Funktion, für die $\theta = \eta$ gilt. Obwohl andere Link-Funktionen verwendet werden können, bietet der kanonische Link insbesondere für kleine Stichprobenumfänge günstige statistische Eigenschaften (McCullagh u. Nelder, 1989, S. 32).

2.2.1.1. Die Poisson-Verteilung

Da sich der weitere Verlauf der Arbeit maßgeblich mit der Modellierung von Anzahldaten beschäftigen wird, soll die Poisson-Verteilung bzw. das zugehörige GLM, das log-lineare Modell, neben der allgemeinen Betrachtung der Exponentialfamilie und der GLMs gesondert hervorgehoben werden. Die Poisson-Verteilung (Poisson, 1837) beschreibt die Wahrscheinlichkeit, daß in einem gegebenen Intervall eine vorgegebene Anzahl voneinander unabhängiger Ereignisse eintritt.

$$P(Y = y; \mu) = P(y) = \frac{\mu^y \cdot \exp(-\mu)}{y!}, \quad \mu > 0, \quad y = 0, 1, 2, \ldots \tag{2.18}$$

Das auf der Poisson-Verteilung basierende GLM, als Poisson-Modell, log-lineares Modell oder Poisson-Regression bezeichnet, wird in der Regel für die Modellierung der Anzahl von Ereignissen mit geringer Erfolgswahrscheinlichkeit verwendet.

Der Erwartungswert und die Varianz einer Poisson-verteilten Zufallsvariable sind gleich.

$$\mathrm{E}(Y) = \mathrm{var}(Y) = \mu$$

Geschätzt werden Erwartungswert und Varianz durch das arithmetische Mittel $\bar{y} = 1/n \sum_{i=1}^{n} y_i$ der beobachteten Anzahlen y_i. Mit größer werdendem μ wird die rechtsschiefe Poisson-Verteilung zunehmend symmetrischer und geht für $\mu \to \infty$ in eine Normalverteilung mit Erwartungswert und Varianz μ über.

Formt man (2.18) zu

$$P(y) = \exp\{y \cdot \ln(\mu) - \mu - \ln(y!)\} \tag{2.19}$$

um, so ist mit $\theta = \ln(\mu)$, $b(\theta) = \mu$, $c(y; \phi) = -\ln(y!)$ und $a(\phi) = 1$ die Zugehörigkeit zur Exponentialfamilie zu erkennen (vergl. (2.16)). Der kanonische Link ($\eta = \theta$) ist für das Poisson-Modell durch den natürlichen Logarithmus $g(\mu) = \ln(\mu)$ gegeben. Er streckt den positiven Wertebereich Poisson-verteilter (Anzahl)Daten auf den gesamten reellen Wertebereich. Die additiven Effekte in η werden durch die inverse Link-Funktion $g^{-1}(\eta) = \exp(\eta) = \mu$ zu multiplikativen in μ, und μ ist notwendigerweise positiv.

Eine sog. *Offset*-Variable t ist eine in der Designmatrix $\boldsymbol{X}$ enthaltene Kovariate[3], deren Parameter in $\boldsymbol{\beta}$ jedoch konstant und gleich 1 ist, also nicht geschätzt wird. Sie kann verwendet werden, um unterschiedliche Größen des zeitlichen

[3] Genauer ausgedrückt enthält $\boldsymbol{X}$ den natürlichen Logarithmus der Offset-Variablen ($\ln(t)$).

oder räumlichen Intervalls, in dem Ereignisse gezählt wurden, zu berücksichtigen. (2.19) wird dann zu

$$P(y_i) = \exp\{y_i \cdot \underbrace{\ln(t_i\,\mu_i)}_{\theta_i} - \underbrace{t_i\,\mu_i}_{b(\theta_i)} \underbrace{- \ln(y!)}_{c(y_i;\phi)}\}, \tag{2.20}$$

worin t_i der Wert der Offset-Variablen für Beobachtung i ist. Mit $\mu_i = \exp(\eta_i) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})$ und (2.17) folgt dann

$$E(Y_i) = b'(\theta_i) = \exp[\ln(t_i\,\mu_i)] = t_i\,\mu_i = t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) = \tilde{\mu}_i = b(\theta_i).$$

t_i wirkt demnach multiplikativ auf die Vorhersage des linearen Modells ($\tilde{\mu}_i = t_i \cdot \exp(\eta_i)$). Aus $g(\tilde{\mu}_i) = \ln(t_i\,\mu_i) = \ln(t_i) + g(\mu_i)$ läßt sich dann die erweiterte Link-Funktion

$$g(\mu_i) = \eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} = \ln(\tilde{\mu}_i) - \ln(t_i)$$

ableiten. Ist das Beobachtungsintervall für alle Beobachtungen gleich groß, so ist $t_i = 1\ \forall\ i$ und es ergibt sich wiederum der kanonische Link des Poisson-Modells $\theta = \ln(\mu)$.

2.2.2. Schätzung der Parameter

Seien $y_1, y_2, \ldots, y_n$ Realisierungen der stochastisch unabhängigen Zufallsvariablen $Y_1, Y_2, \ldots, Y_n$ mit einer Verteilung aus der Exponentialfamilie. Ziel ist es, die Parameter $\beta_0, \beta_1, \ldots, \beta_{p-1}$ zu schätzen, die mit den Y_i über $\mathrm{E}(Y_i) = \mu$ und $g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$ in Beziehung stehen. Die Anpassung des Modells erfolgt über die *Maximum-Likelihood*-Methode (ML), d. h. es werden diejenigen β_j bestimmt, die den Wert der Likelihood-Funktion $\mathcal{L}(\boldsymbol{y};\boldsymbol{\theta},\phi)$ oder äquivalent der log-Likelihood-Funktion $\ell(\boldsymbol{y};\boldsymbol{\theta},\phi)$ maximieren. Der Beitrag einer Beobachtung zur Likelihood-Funktion entspricht dabei gerade dem Wert der Wahrscheinlichkeitsfunktion der Beobachtung.

$$\mathcal{L}_i = f(y_i;\theta_i,\phi)$$

Die gemeinsame Likelihood-Funktion aller Beobachtungen ist dann das Produkt der einzelnen Beiträge,

$$\mathcal{L}(\boldsymbol{y};\boldsymbol{\theta},\phi) = \prod_{i=1}^{n} \mathcal{L}_i,$$

die gemeinsame log-Likelihood-Funktion entsprechend deren Summe,

$$\begin{aligned} \ell(\boldsymbol{y};\boldsymbol{\theta},\phi) &= \sum_{i=1}^{n} \ln f(y_i,\theta_i,\phi) \\ &= \boldsymbol{y}'\boldsymbol{A}^{-1}\boldsymbol{\theta} - [\boldsymbol{b}^{1/2}]'\boldsymbol{A}^{-1}[\boldsymbol{b}^{1/2}]' + \boldsymbol{1}_n'\boldsymbol{c}, \end{aligned} \tag{2.21}$$

worin $\boldsymbol{y}$ der $n \times 1$ Beobachtungsvektor, $\boldsymbol{\theta}$ der Vektor der natürlichen Parameter $[\theta_1, \ldots, \theta_n]'$, $\boldsymbol{b}^{1/2} = \left[\sqrt{b(\theta_1)}, \ldots, \sqrt{b(\theta_n)}\right]'$, $\boldsymbol{A}$ die $n \times n$ Diagonalmatrix mit Elementen diag $[a_i(\phi)]$, $\boldsymbol{c}$ der $n \times 1$ Vektor mit den Einträgen $c(y_i;\phi)$ und $\boldsymbol{1}_n$ ein $n \times 1$ Vektor aus Einsen ist.

Die Schätzung des Parametervektors $\boldsymbol{\beta}$ ergibt sich durch Berechnung der partiellen Ableitungen der log-Likelihood-Funktion und Nullsetzen der resultierenden Gleichungen. Für Parameter β_j berechnet sich die partielle Ableitung durch

$$s_j(\boldsymbol{\beta}) = \frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} \frac{y_i - \mu_i}{\mathrm{var}(y_i)} \frac{\partial \mu_i}{\partial \eta_i} x_{ij}.$$

$\boldsymbol{s}(\boldsymbol{\beta})$ bezeichnet die, auch Gradientenvektor genannte, *Score*-Funktion. In Matrixschreibweise ergibt sich für $\boldsymbol{\beta}$

$$\boldsymbol{s}(\boldsymbol{\beta}) = \boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}), \tag{2.22}$$

wobei $\boldsymbol{D}$ die $n \times p$ Matrix der partiellen Ableitungen $\partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$ und $\boldsymbol{V}$ die $n \times n$ Varianzmatrix $\mathrm{diag}\,[\phi \cdot v(\mu_i)]$ ist. Das zur Schätzung des Parametervektors $\boldsymbol{\beta}$ zu lösende Gleichungssystem ist also

$$\boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = 0. \tag{2.23}$$

Wird der kanonische Link des GLMs verwendet, so vereinfacht sich (2.23) zu

$$\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{\mu}) = 0, \tag{2.24}$$

da die Elemente von $\boldsymbol{D}$ dann durch

$$\boldsymbol{D}_{ij} = \frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij} = \frac{\partial \mu_i}{\partial \theta_i} x_{ij} = \frac{\partial b'(\theta_i)}{\partial \theta_i} x_{ij} = b''(\theta_i) \cdot x_{ij}$$

gegeben sind. Die erste Umformung erfolgt, da $\boldsymbol{x}_i'\boldsymbol{\beta} = \eta_i$, die zweite, da für den kanonischen Link $\eta_i = \theta_i$ und die dritte, da nach (2.17) $\mu_i = b'(\theta_i)$. Die Inverse der Varianzmatrix enthält die Elemente

$$V_{ii}^{-1} = \frac{1}{\mathrm{var}(y_i)} = \frac{1}{\phi \cdot v(\mu_i)} = \frac{1}{\phi \cdot b''(\theta_i)},$$

so daß das Produkt $\boldsymbol{D}'\boldsymbol{V}^{-1}$ die Elemente $1/\phi \cdot x_{ij}$ enthält. Beim Nullsetzen der resultierenden Score-Gleichungen entfällt der Term $1/\phi$, so daß sich die in (2.24) angegebene Form ergibt.

Zur Lösung der Score-Gleichungen $\boldsymbol{s}(\boldsymbol{\beta}) = 0$ wird des weiteren die *Hesse-Matrix* $\boldsymbol{H}$, die Matrix der zweiten partiellen Ableitungen $\partial\ell/(\partial\beta_s\partial\beta_r)$, benötigt.

$$\boldsymbol{H} = -\boldsymbol{X}'\boldsymbol{W}_{obs}\boldsymbol{X},$$

wobei $\boldsymbol{X}$ die Designmatrix ist und $\boldsymbol{W}_{obs}$ eine $n \times n$ Diagonalmatrix mit i-tem Diagonalelement

$$w_{obs_{ii}} = w_{exp_i} + (y_i - \mu_i)\,\frac{v(\mu_i) \cdot g''(\mu_i) + v'(\mu_i) \cdot g'(\mu_i)}{[v(\mu_i)]^2 \cdot [g(\mu_i)]^3},$$

$g'(\mu_i) = \partial\eta_i/\partial\mu_i$ und

$$w_{exp_{ii}} = \frac{1}{\phi \cdot v(\mu_i) \cdot [g'(\mu_i)]^2} = \frac{1}{\phi \cdot v(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2. \tag{2.25}$$

$\boldsymbol{H}$ enthält also die Elemente

$$\boldsymbol{H}_{sr} = -\sum_{i=1}^{n} w_{obs_{ii}}\, x_{is}\, x_{ir}.$$

Die negative Hesse-Matrix ist die *beobachtete (Fisher) Informationsmatrix* $\boldsymbol{F}_{obs}$.

Da $\partial\ell/(\partial\beta_s\,\partial\beta_r)$ von $\boldsymbol{y}$ abhängt, wird statt dessen meist der Erwartungswert $\mathrm{E}\left[\partial\ell/(\partial\beta_s\,\partial\beta_r)\right]$ verwendet. Der negative Erwartungswert der Ableitungen ist die *erwartete (Fisher) Informationsmatrix*

$$\begin{aligned} -\mathrm{E}\left(\frac{\partial\ell}{\partial\beta_s\,\partial\beta_r}\right) \quad = \quad \boldsymbol{F}_{exp} \quad &= \quad \boldsymbol{X}'\boldsymbol{W}_{exp}\boldsymbol{X} \\ &= \quad \boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D} \end{aligned} \tag{2.26}$$

(s. Anhang B.1) mit den Elementen

$$\boldsymbol{F}_{exp_{sr}} = \sum_{i=1}^{n} w_{exp_i}\, x_{is}\, x_{ir}.$$

$\boldsymbol{W}_{exp}$ ist eine $n \times n$ Diagonalmatrix mit den in (2.25) definierten Elementen. Wie schnell zu ersehen ist, ist $\boldsymbol{F}_{exp}$ leichter zu berechnen als $\boldsymbol{F}_{obs}$ und zudem stets positiv semidefinit (Fahrmeir u. Tutz, 2001, S. 42). Für kanonische Link-Funktionen ist $\boldsymbol{H} = \mathrm{E}(\boldsymbol{H})$ und damit $\boldsymbol{F}_{exp} = \boldsymbol{F}_{obs}$.

Da die Score-Gleichungen nichtlinear in β_j sind, sind die sie iterativ zu lösen. Basierend auf einer aktuellen Schätzung des Parametervektors $\hat{\boldsymbol{\beta}}_{(m)}$ berechnet sich die nächste durch

$$\hat{\boldsymbol{\beta}}_{(m+1)} = \hat{\boldsymbol{\beta}}_{(m)} + \boldsymbol{F}_{(m)}^{-1}\boldsymbol{s}_{(m)}. \tag{2.27}$$

Die Berechnungen werden so lange wiederholt, bis ein vorgegebenes Konvergenzkriterium erreicht ist. Wird für $\boldsymbol{F}$ die beobachtete Informationsmatrix $\boldsymbol{F}_{obs}$ verwendet, so handelt es sich um den *Newton-Raphson*-Anpassungsalgorithmus, für $\boldsymbol{F} = \boldsymbol{F}_{exp}$ um das sog. *Fisher Scoring*.

(2.27) läßt sich für die Fisher-Scoring-Methode zu

$$\boldsymbol{X}'\boldsymbol{W}_{exp_{(m)}}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{m+1} = \boldsymbol{X}'\boldsymbol{W}_{exp_{(m)}}\boldsymbol{y}^*_{(m)} \tag{2.28}$$

umformen (s. Anhang B.2), wobei $\boldsymbol{y}^*_{(m)}$ die Elemente

$$y^*_{i_{(m)}} = \boldsymbol{x}_i'\hat{\boldsymbol{\beta}}_{(m)} + (y_i - \hat{\mu}_{i_{(m)}})\left(\frac{\partial\hat{\eta}_i}{\partial\hat{\mu}_i}\right)_{(m)} \tag{2.29}$$

besitzt. Die Form der Darstellung in (2.28) wird oft in der Literatur verwendet. Sie entspricht den Normalgleichungen zur Anpassung eines gewöhnlichen linearen Modells an die Untersuchungsvariable $\boldsymbol{y}^*$ über gewichtete kleinste Quadrate. Im Unterschied dazu erfolgt die Anpassung hier jedoch iterativ, da sowohl $\boldsymbol{y}^*$, als auch $\boldsymbol{W}_{exp}$ von $\hat{\boldsymbol{\beta}}$ abhängt. Die Anpassung des GLMs kann daher auch als Anpassung über iterativ gewichtete kleinste Quadrate (*Iteratively Weighted*

Least Squares, IWLS) betrachtet werden (Charnes et al., 1976). Für den Identity Link ($\eta = \mu$) vereinfacht sich $\boldsymbol{W}_{exp} = \boldsymbol{V}^{-1}(\partial\boldsymbol{\mu}/\partial\boldsymbol{\eta})^2$ zu $\boldsymbol{W}_{exp} = \boldsymbol{V}^{-1}$ und $\boldsymbol{y}^* = \boldsymbol{\eta} + (\boldsymbol{y} - \boldsymbol{\mu})\,(\partial\boldsymbol{\eta}/\partial\boldsymbol{\mu})$ zu $\boldsymbol{y}$ und damit (2.28) zu

$$\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y},$$

dem Gleichungssystem verallgemeinerter kleinster Quadrate. Unter der Annahme konstanter Varianzen, also $\boldsymbol{V}^{-1} = \operatorname{diag}\left[1/\sigma^2\right]$, kann auch $\boldsymbol{V}^{-1}$ auf beiden Seiten der Gleichung entfallen, so daß sich die Normalgleichungen des klassischen linearen Modells ergeben.

2.2.2.1. Parameterschätzung im Poisson-Modell

Die log-Likelihood-Funktion einer Poisson-verteilten Zufallsvariable ist, einen Offset-Term einbeziehend (vergl. (2.20)),

$$\begin{aligned}\ell(\boldsymbol{y};\boldsymbol{\beta}) &= \sum_{i=1}^{n}\{y_i \cdot \ln(t_i\,\mu_i)\} - \sum_{i=1}^{n}\{t_i\,\mu_i\} - \sum_{i=1}^{n}\ln(y_i!) \\ &= \boldsymbol{y}'\boldsymbol{\theta} - \mathbf{1}_n'\boldsymbol{b} + \mathbf{1}_n'\boldsymbol{c}, \end{aligned} \tag{2.30}$$

worin $\mathbf{1}_n$ ein $n \times 1$ Vektor aus Einsen ist, $\boldsymbol{\theta} = [\ln(t_1\,\mu_1), \ldots, \ln(t_n\,\mu_n)]'$, $\boldsymbol{c} = [-\ln(y_1!), \ldots, -\ln(y_n!)]'$ und $\boldsymbol{b} = [t_1\,\mu_1, \ldots, t_n\,\mu_n]'$. (2.21) vereinfacht sich zu (2.30), da für die Poisson-Verteilung $a(\phi) = 1$ und damit $\boldsymbol{A} = \boldsymbol{I}$. Die Score-Funktion eines Parameters β_j ist dann

$$\begin{aligned} s_j(\boldsymbol{\beta}) &= \frac{\partial\ell}{\partial\beta_j} = \sum_{i=1}^{n} y_i\,x_{ij} - \sum_{i=1}^{n} t_i\,\mu_i\,x_{ij} \\ &= \sum_{i=1}^{n}\left\{y_i - t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right\} x_{ij}, \end{aligned} \tag{2.31}$$

da

$$\frac{\partial\mu_i}{\partial\beta_j} = \frac{\partial[\exp(\boldsymbol{x}_i'\boldsymbol{\beta})]}{\partial\beta_j} = x_{ij}$$

und

$$\frac{\partial[\ln(t_i\,\mu_i)]}{\partial\beta_j} = \frac{\partial[\ln(t_i) + \ln\{\exp(\boldsymbol{x}_i'\boldsymbol{\beta})\}]}{\partial\beta_j} = x_{ij}.$$

Da es sich um den kanonischen Link ($\eta = \theta = \ln(\tilde{\mu})$) handelt und damit $\boldsymbol{D}'\boldsymbol{V}^{-1} = \boldsymbol{X}'$, ergibt sich für den gesamten Parametervektor

$$\boldsymbol{s}(\boldsymbol{\beta}) = \boldsymbol{X}'(\boldsymbol{y} - \tilde{\boldsymbol{\mu}}),$$

mit $\tilde{\boldsymbol{\mu}} = [t_1 \cdot \exp(\boldsymbol{x}_1'\boldsymbol{\beta}), \ldots, t_n \cdot \exp(\boldsymbol{x}_n'\boldsymbol{\beta})]'$.

Die Hesse-Matrix beinhaltet bei Verwendung des kanonischen Links die Elemente

$$\begin{aligned} \boldsymbol{H}_{sr} &= \frac{\partial^2\ell}{\partial\beta_s\,\partial\beta_r} = \frac{\partial}{\partial\beta_s}\left\{\sum_{i=1}^{n}\left(y_i - t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\right)\,x_{ir}\right\} \\ &= -\sum_{i=1}^{n} t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\,x_{is}\,x_{ir}. \end{aligned}$$

Die beobachtete Hesse-Matrix und deren Erwartungswert sind identisch und damit auch $\boldsymbol{F}_{obs}$ und $\boldsymbol{F}_{exp}$. Sie besitzen die Elemente

$$\begin{aligned} \boldsymbol{F}_{sr} &= -\mathrm{E}\left(-\sum_{i=1}^{n} t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\, x_{is}\, x_{ir}\right) \\ &= \sum_{i=1}^{n} t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\, x_{is}\, x_{ir}. \end{aligned}$$

In Matrixform ergibt sich

$$\boldsymbol{F} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X},$$

mit $\boldsymbol{W} = \mathrm{diag}\,[t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta})]$, da für das Poisson-Modell $\mathrm{var}(y_i) = \tilde{\mu}_i$ und für den kanonischen Link $\partial\tilde{\mu}_i/\partial\eta_i = t_i \cdot \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) = \tilde{\mu}$ und damit $w_{exp_{ii}} = 1/\tilde{\mu}_i \cdot \tilde{\mu}_i^2 = \tilde{\mu}_i$ (vergl. (2.25)).

2.2.3. Anpassungsgüte und Residuenanalyse

Zur Beurteilung der Anpassungsgüte von GLMs werden oftmals die sog. skalierte Devianz D^* und die skalierte Pearson-Statistik χ^{2*} herangezogen. Die skalierte Devianz ist definiert als zweifache Differenz zwischen maximal erreichbarer log-Likelihood $\ell(\boldsymbol{y}, \boldsymbol{y})$ – dies entspricht der log-Likelihood eines gesättigten Modells mit $p = n$ Parametern – und unter dem gegebenen Modell mit $p < n$ Parametern realisierter $\ell(\boldsymbol{y}, \hat{\boldsymbol{\mu}})$.

$$D^* = 2\,\{\ell(\boldsymbol{y}; \boldsymbol{y}) - \ell(\boldsymbol{y}; \hat{\boldsymbol{\mu}})\}$$

Die maximal erreichbare log-Likelihood ergibt sich durch Einsetzen der beobachteten y_i anstelle der vorhergesagten $\hat{\mu}_i$ in die log-Likelihood-Funktion.

Die Devianz, gegeben durch $D = \phi\, D^*$, ist die Verallgemeinerung der Fehlerquadratsumme (SS_{error}) im gewöhnlichen linearen Modell.

Die Pearson-Statistik ist definiert als

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)}.$$

Ihre skalierte Version ist, der Berechnung der skalierten Devianz entsprechend, $\chi^{2*} = \chi^2/\phi$. Große Werte beider Teststatistiken sprechen gegen die Nullhypothese, daß das gewählte Modell das wahre ist. Unter der Nullhypothese folgen die skalierten Versionen der Teststatistiken asymptotisch einer χ^2-Verteilung mit $n - p$ Freiheitsgraden, über die sich zugehörige p-Werte berechnen lassen.

$$D^*,\, \chi^{2*} \overset{a}{\sim} \chi^2_{n-p}$$

Für das Poisson-Modell (mit $a(\phi) = 1$) sind skalierte und nicht-skalierte Devianz und Pearson-Statistik gleich und gegeben durch

$$D^* = D = 2\sum_{i=1}^{n}\left[y_i \cdot \ln\left(\frac{y_i}{\hat{\mu}_i}\right) - (y_i - \hat{\mu}_i)\right] \tag{2.32}$$

und

$$\chi^{2*} = \chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}.$$

Für Poisson-Modelle, die den kanonischen Link ($\eta = \ln(\mu)$) verwenden, ist nach (2.24) $\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\hat{\boldsymbol{\mu}}$. Enthält das Modell zudem einen Interzeptterm (d. h. die Elemente einer Spalte in der Designmatrix sind alle gleich Eins), so ist die Summe der beobachteten Werte gleich der Summe der vorhergesagten Werte ($\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{\mu}_i$). Die Devianz aus (2.32) vereinfacht sich dann zu

$$D^* = D = 2\sum_{i=1}^{n} \left[y_i \cdot \ln\left(\frac{y_i}{\hat{\mu}_i}\right)\right].$$

Beide Statistiken sind zur Beurteilung der Anpassungsgüte eines Modells jedoch mit Vorsicht zu verwenden, wenn die Stichprobenumfänge je Kovariatenkombination (unterschiedliche Zeilen in $\boldsymbol{X}$) klein sind. Im Extremfall, in dem jede Kombination nur einmal vorkommt (ungruppierte Daten), gilt die obige approximative Verteilung nicht und beide Statistiken sind unbrauchbar (Fahrmeir u. Tutz, 2001, S. 51; McCullagh u. Nelder, 1989, S. 118 f.). Ebenso können die skalierten Versionen der Devianz und Pearson-Statistik nicht als Maß der Anpassungsgüte gedeutet werden, wenn Über- oder Unterdispersion (s. u.) vorliegt.

Wesentliche Bedeutung besitzt die skalierte Devianz bei den im nachfolgenden Abschnitt erläuterten Hypothesentests zur Beurteilung der im Modell enthaltenen Effekte. Sie gelten auch im Fall ungruppierter Daten.

In klassischen linearen Modellen mit normalverteilten und homoskedastischen Fehlern besitzen die rohen Residuen ($y_i - \hat{\mu}_i$) für große Stichprobenumfänge die gewünschten Eigenschaften symmetrischer Verteilung um Null und konstanter Varianz. Entstammen die Daten jedoch nicht normalen Verteilungen, sind beide Eigenschaften meist nicht gegeben. So sind beispielsweise Varianz und Schiefe der rohen Residuen einer Poisson-verteilten Variablen y gleich μ, so daß die rohen Residuen auch für große Stichprobenumfänge heteroskedastisch sind und eine asymmetrische Verteilung besitzen. Zur (insbesondere graphischen) Beurteilung eines Modells sollten die rohen Residuen daher dahingehend transformiert werden, eine möglichst symmetrische Verteilung und konstante Varianz zu erreichen. Insbesondere für das Poisson-Modell werden hierfür die Anscombe-Residuen (Anscombe, 1953) empfohlen (McCullagh u. Nelder, 1989, S. 38) und sind für dieses gegeben durch

$$r_i^{(A)} = \frac{\frac{3}{2}(y_i^{2/3}) - \hat{\mu}_i^{2/3}}{\hat{\mu}_i^{1/6}}. \tag{2.33}$$

Insbesondere in der Anwendung der unten erläuterten *Generalized Estimating Equations* sind die Pearson-Residuen

$$r_i^{(P)} = \frac{y_i - \hat{\mu}_i}{\sqrt{v(\hat{\mu}_i)}} \tag{2.34}$$

von Bedeutung. Eine speziell für kleine Stichprobenumfänge sinnvolle Student-Standardisierung, die Residuen, deren Varianz konstant und asymptotisch gleich Eins ist ($\mathrm{var}(r^{(P)^*}) \stackrel{a}{=} 1$), hervorbringt, kann über das Pendant der aus linearen Modellen bekannten Projektionsmatrix (*Hat Matrix*) $\boldsymbol{P}$ erfolgen. Für GLMs ist diese definiert durch

$$\boldsymbol{P} = \boldsymbol{W}_{exp}^{1/2}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{W}_{exp}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}_{exp}^{1/2}, \tag{2.35}$$

mit der Diagonalmatrix $\boldsymbol{W}_{exp}$ und den in (2.25) definierten Elementen. Die Student-Standardisierung der Pearson-Residuen erfolgt dann durch

$$r_i^{(P)^*} = \frac{y_i - \hat{\mu}_i}{\sqrt{\psi \cdot v(\hat{\mu}_i)\,(1 - p_{ii})}}, \tag{2.36}$$

wobei p_{ii} der i-te Diagonaleintrag aus $\boldsymbol{P}$ ist. Der Divisor $\sqrt{(1 - p_{ii})}$ erzielt die Standardisierung, so daß $\mathrm{var}(r_i^{(P)^*}) = \mathrm{const}$, $\sqrt{\psi \cdot v(\hat{\mu}_i)}$ die Student-Standardisierung (McCullagh u. Nelder, 1989, S. 396). ψ ist hierin entweder der bereits eingeführte Skalierungs- oder Dispersionsparameter ϕ oder ein Über- oder Unterdispersionsparameter σ_o^2 (s. u.).

Neben den genannten existieren weitere Residuen, wie z. B. Devianz-, Score- und Likelihood-Residuen sowie zusätzliche Anpassungen und Standardisierungen (s. z. B. McCullagh u. Nelder, 1989, 2.4; Lindsey, 1997, Anh. B.2; Cameron u. Trivedi, 1998, Kap. 5.2).

Zur Überprüfung der Wahl der Varianzfunktion, der Link-Funktion, der Varianzhomogenität, der Skalierung der einbezogenen Kovariaten und zur Aufdeckung von Ausreißern und einflußreichen Beobachtungen existieren zahlreiche graphische Möglichkeiten, insbesondere in Form von Residuenplots. Auflistungen und Hinweise zur Interpretationen derartiger Plots finden sich u. a. in McCullagh u. Nelder (1989, Kap. 12.6), Cameron u. Trivedi (1998, Kap. 5.2), Fahrmeir u. Tutz (2001, Kap. 4.2) und Lindsey (1997, S. 227 f.). Trotz umfangreicher Möglichkeiten der Residuenanalyse, ist diese jedoch nicht unproblematisch, da jede aufgedeckte Abweichung vom Idealzustand meist mehrere mögliche Ursachen haben kann. Entsprechend aufwendig und langwierig kann sich die Suche nach der tatsächlichen Ursache gestalten.

2.2.4. Varianzschätzung und Hypothesentests

Zu den asymptotischen Eigenschaften des Maximum-Likelihood-Schätzers gehört, daß $\hat{\boldsymbol{\beta}}$ für $n \to \infty$ gegen den wahren Parametervektor $\boldsymbol{\beta}$ strebt (Konsistenz) und approximativ normalverteilt ist.

$$\hat{\boldsymbol{\beta}} \stackrel{a}{\sim} N(\boldsymbol{\beta}, \boldsymbol{F}_{exp}^{-1})$$

mit asymptotischer Kovarianzmatrix

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \boldsymbol{F}_{exp}^{-1} = (\boldsymbol{X}'\boldsymbol{W}_{exp}\boldsymbol{X})^{-1},$$

worin $\boldsymbol{F}_{exp}^{-1}$ die Inverse der oben definierten Fisher-Informationsmatrix und $\boldsymbol{\beta}$ der wahre Parametervektor ist. Konfidenzintervalle der Schätzer lassen sich somit durch

$$\hat{\beta}_j \pm z_{1-\alpha/2} \cdot \sigma(\hat{\beta}_j)$$

bestimmen. $z_{1-\alpha/2}$ ist hierin das $100(1-\alpha/2)$-Quantil der Standardnormalverteilung und $\sigma(\hat{\beta}_j)$ der Standardfehler der Parameterschätzung $\hat{\beta}_j$, also die Wurzel des j-ten Diagonaleintrags aus $\boldsymbol{F}_{exp}^{-1}$.

Die Varianz des linearen Prädiktors $\hat{\eta}_i$ ist

$$\mathrm{var}(\hat{\eta}_i) = \boldsymbol{x}_i' \boldsymbol{F}_{exp}^{-1} \boldsymbol{x}_i.$$

Über die inverse Link-Funktion lassen sich Konfidenzintervalle für die vorhergesagten Werte $\hat{\mu}_i$ berechnen.

$$g^{-1}(\boldsymbol{x}_i' \hat{\boldsymbol{\beta}} \pm z_{1-\alpha/2} \cdot \mathrm{se}(\hat{\eta}_i))$$

Die Konfidenzintervalle werden so – außer im Falle des Normalverteilungsmodells – asymmetrisch und spiegeln damit die nichtlineare Eigenschaft der Link-Funktion wider.

Standardfehler der vorhergesagten Werte können mit Hilfe der Delta-Regel (z. B. Bishop et al., 1975) bestimmt werden. Danach ist

$$\mathrm{var}(\hat{\mu}_i) \approx \left(\frac{\partial \hat{\mu}_i}{\partial \hat{\eta}_i} \right)^2 \cdot \mathrm{var}(\hat{\eta}_i). \tag{2.37}$$

Im Poisson-Modell ist also $\mathrm{var}(\hat{\mu}_i) \approx [\exp(\hat{\eta}_i)]^2 \cdot \mathrm{var}(\hat{\eta}_i)$.

Viele Testprobleme lassen sich mit Hilfe der Hypothesen

$$H_0 : \boldsymbol{L}'\boldsymbol{\beta} = \xi \quad \text{gegen} \quad H_1 : \boldsymbol{L}'\boldsymbol{\beta} \neq \xi \tag{2.38}$$

formulieren, wobei $\boldsymbol{L}$ die sog. Kontrastmatrix ist. Von besonderer Bedeutung ist hierbei oftmals der Sonderfall

$$H_0 : \boldsymbol{\beta}_r = \boldsymbol{0} \quad \text{gegen} \quad H_1 : \boldsymbol{\beta}_r \neq \boldsymbol{0}, \tag{2.39}$$

worin $\boldsymbol{\beta}_r$ ein r-dimensionaler Subvektor von $\boldsymbol{\beta}$ ist. Dies entspricht einem Test des Submodells mit $p-r$ Parametern gegen das volle Modell mit p Parametern. Beide Modelle verwenden aber dieselbe Wahrscheinlichkeitsverteilung und Link-Funktion. Nur auf die Hypothesenformulierung aus (2.39) wird hier näher eingegangen, für Tests der allgemeinen Hypothesenformulierung aus (2.38) s. z. B. Fahrmeir u. Tutz (2001, Kap. 2.2.2), Littel et al. (2002, Kap. 10.6.3) oder SAS (2004, Kap. 29) und speziell für Anzahldaten Cameron u. Trivedi (1998, Kap. 2.6).

Zur Überprüfung der Nullhypothese aus (2.39) erfolgt die Anpassung des Modells einmal uneingeschränkt (wie in den vorhergehenden Abschnitten beschrieben) und einmal mit der Einschränkung, daß $\boldsymbol{\beta}_r = 0$. Sei $\hat{\boldsymbol{\beta}}$ die uneingeschränkte und $\tilde{\boldsymbol{\beta}}$ die unter der Einschränkung erhaltene ML-Schätzung, dann ist die *Likelihood-Ratio-Teststatistik* definiert als

$$S = 2\left\{ \ell(\hat{\boldsymbol{\beta}}) - \ell(\tilde{\boldsymbol{\beta}}) \right\}.$$

Ist s groß, so spricht dies für die Ablehnung der Nullhypothese und damit für die Beibehaltung der Effekte im Modell. Oftmals wird die Likelihood-Ratio-Teststatistik auch als Test der Devianz-Differenz des Teil- und vollen Modells $s = D_r - D_p$ dargestellt, die äquivalent zur oben dargestellten Form ist.

Da für die Likelihood-Ratio-Teststatistik die iterative Anpassung des Modells für jede zu testende Hypothese wiederholt werden muß, steht als weniger rechenintensive Alternative die Wald-Statistik zur Verfügung, die für die Nullhypothese aus (2.39) durch

$$W = \boldsymbol{\beta}_r' \boldsymbol{A}_r^{-1} \boldsymbol{\beta}_r \tag{2.40}$$

definiert ist, worin $\boldsymbol{A}_r$ die zu den Parametern aus $\boldsymbol{\beta}_r$ gehörende Submatrix von $\boldsymbol{A} = \boldsymbol{F}_{exp}^{-1}$ ist. Ist $\boldsymbol{\beta}_r$ lediglich ein skalare Komponente aus $\boldsymbol{\beta}$, so ist die Wald-Statistik durch den t-Wert

$$t_r = \frac{\hat{\beta}_r}{\sqrt{\hat{a}_{rr}}},$$

gegeben, mit $\hat{a}_{rr} = \mathrm{var}(\hat{\beta}_r)$, dem r-ten Diagonaleintrag der geschätzten asymptotischen Kovarianzmatrix $\boldsymbol{F}_{exp}^{-1}$.

Unter der Nullhypothese, daß $\boldsymbol{\beta}_r = 0$, sind Wald- und Likelihood-Ratio-Statistik asymptotisch χ^2-verteilt mit r Freiheitsgraden.

$$W, S \overset{a}{\sim} \chi^2_{1-\alpha;r}$$

Wald-Statistiken sind zwar leichter zu berechnen, es wird jedoch angenommen, daß die angegebenen Signifikanzniveaus hier weiter von den tatsächlichen abweichen als dies für Likelihood-Ratio-Teststatistiken der Fall ist (SAS, 2004, S. 1414).

Enthält das Modell einen unbekannten Skalierungsparameter[4] ψ, wie die Varianz ($\psi = \phi = \sigma^2$) für normalverteilte Daten oder den Überdispersionsparameter ($\psi = \sigma_o^2$) im überstreuten Poisson-Modell (s. nächster Abschnitt), wird aber durch einen konsistenten Schätzer $\hat{\psi}$ ersetzt, so ergeben sich angepaßte Likelihood-Ratio- und Wald-Teststatistik durch Dividieren der oben genannten durch $\hat{\psi}$. Für Hypothesentests wird dann anstelle der χ^2-Verteilung die F-Verteilung verwendet, um zu berücksichtigen, daß es sich bei $\hat{\psi}$ um eine Schätzung handelt. Im gewöhnlichen linearen Modell mit normalverteilten Fehlern sind Wald- und Likelihood-Ratio-Teststatistik identisch zur Quadratsumme der zu prüfenden Hypothese (SS_h). Die genannte Anpassung entspricht dort also dem Verhältnis SS_h/MS_{error}. Teilen dieses Ausdrucks durch die Anzahl der Freiheitsgrade der zu testenden Hypothese resultiert in der bekannten F-Statistik MS_h/MS_{error}. Für GLMs ist die F-Statistik äquivalent durch das Verhältnis aus angepaßter Likelihood-Ratio- oder Wald-Statistik und der Anzahl der Hypothesenfreiheitsgrade r definiert.

$$\tilde{S} = \frac{s}{r\,\hat{\psi}} \qquad \text{oder} \qquad \tilde{W} = \frac{w}{r\,\hat{\psi}} \tag{2.41}$$

Wenngleich die Likelihood-Ratio-Statistik eigentlich für Überdispersionsmodelle nicht geeignet definiert ist, da in diesem Fall die Likelihood-Funktion nicht voll spezifiziert werden kann, wird dieses Vorgehen dennoch oft als Annäherung verwendet. Die Wald-Statistik dagegen bezieht lediglich das erste und zweite Moment der Verteilung ein und ist damit auch für Überdispersionsmodelle

[4] Der Skalierungsparameter $\hat{\psi}$ wird hier wiederum als Schätzer des Dispersionsparameters ϕ, als auch eines etwaigen Überdispersionsparameters σ_o^2 verwendet, wenngleich beide ausdrücklich nicht identisch sind.

geeignet definiert (Fahrmeir u. Tutz, 2001, S. 48). Die Teststatistik besitzt r Zähler- und $p_{\hat{\psi}}$ Nennerfreiheitsgrade, wobei $p_{\hat{\psi}}$ die Anzahl der Parameter des Modells ist, das für die Schätzung von ψ verwendet wurde.

2.2.5. Überdispersion

Hohe Werte der Devianz- und der Pearson-Statistik können, wie oben erläutert, für ein schlecht angepaßtes Modell bzw. für Probleme innerhalb des Modells sprechen. Ist das gewählte Modell aber das richtige, so wird das Phänomen einer deutlich größeren als erwarteten Varianz im Falle binomial- oder Poisson-verteilter Variablen als *Überdispersion* bezeichnet. Die Beobachtungen treten dann geklumpter auf als vom Modell unterstellt. Die beiden häufigsten Ursachen hierfür sind unbeobachtete Heterogenität, die durch die Kovariaten des linearen Prädiktors nicht erfaßt wird, und positive (zeitliche, räumliche oder sonstige) Korrelation zwischen den Beobachtungen. Der seltenere Fall der *Unterdispersion* beschreibt den umgekehrten Fall kleinerer als erwarteter Varianz. Wird Überdispersion (Unterdispersion) nicht im Modell berücksichtigt, so resultiert dies in ggf. deutlich unterschätzten (überschätzten) Standardfehlern und damit in zu großen (kleinen) Teststatistiken; zu testende Nullhypothesen werden also zu oft abgelehnt (beibehalten). Um Über- oder Unterdispersion in der Modellierung der Varianz in Binomial- oder Poisson-Modellen zu berücksichtigen, wird oftmals ein zusätzlicher Dispersionsparameter σ_o^2 in das Modell aufgenommen. σ_o^2 wird in vielen Aspekten behandelt wie der Dispersionsparameter der Exponentialfamilie ϕ, ist jedoch nicht identisch mit diesem, so daß auch im über-/unterstreuten Poisson-Modell und Binomial-Modell $\phi = 1$ gilt. Der besseren Lesbarkeit wegen wird im nachfolgenden nur der Begriff der weit häufiger auftretenden Überdispersion verwendet, alle Aussagen gelten aber entsprechend auch für den Fall der Unterdispersion.

Im Falle der Schätzung eines Überdispersionsparameters, sind die log-Likelihood-Funktionen der Binomial- und Poisson-Verteilung keine gültigen log-Likelihood-Funktionen mehr, da diese mit $a(\phi) = 1$ standardmäßig (s. Kap. 2.2.1.1) einen konstanten Dispersionsparameter vorsehen. Es handelt sich aber um gültige *Quasi-Likelihood*-Funktionen (s. nächster Abschnitt), für die ebenfalls die meisten der asymptotischen Eigenschaften der Likelihood-Funktionen gelten, so daß die oben erläuterten Methoden der Parameterschätzung und Hypothesentests weitestgehend unverändert angewendet werden können. Im einfachsten Fall bleibt die Schätzung des Parametervektors $\boldsymbol{\beta}$ von der des zusätzlichen Überdispersionsparameters unbeeinflußt, lediglich die Streuungsmaße werden entsprechend korrigiert, so daß z. B. im überstreuten Poisson-Modell $\mathrm{var}(y) = \sigma_o^2\,\mu$ anstelle von $\mathrm{var}(y) = \mu$ angenommen wird. Zur Schätzung von σ_o^2 wird oft die skalierte Devianz $D^* = D/\sigma_o^2$ herangezogen. Bei $n - p$ Freiheitsgraden besitzt sie den Erwartungswert $n - p$. Einsetzen dieses Erwartungswertes für D^* und Auflösen nach σ_o^2 ergibt dann den Schätzer

$$\hat{\sigma}_o^2 = \frac{D}{n-p} = \frac{2\left\{\ell(\boldsymbol{y};\boldsymbol{y}) - \ell(\boldsymbol{y};\hat{\boldsymbol{\mu}})\right\}}{n-p}. \tag{2.42}$$

Gleichermaßen kann mit der Pearson-Statistik verfahren werden.

$$\hat{\sigma}_o^2 = \frac{\chi^2}{n-p} \tag{2.43}$$

(2.43) liefert auch dann einen approximativ ($n >> p$) verzerrungsfreien und von $\hat{\boldsymbol{\beta}}$ unabhängigen Schätzer, wenn nur wenige oder keine Wiederholungen je Kovariatenkombination vorliegen. (2.42) dagegen ist nur dann konsistent, wenn viele Wiederholungen für alle Kovariatenkombinationen vorliegen (McCullagh u. Nelder, 1989, S. 127 f.). Die Deutung der skalierten Devianz- oder Pearson-Statistik als Maß der Anpassungsgüte ist bei Schätzung eines zusätzlichen Dispersionsparameters natürlich nicht mehr möglich.

Angepaßte Standardfehler der Parameterschätzungen im überstreuten Modell ergeben sich dann aus denen des gewöhnlichen Poisson- oder Binomial-Modells, multipliziert mit σ_o (die geschätzte Kovarianzmatrix wird mit $\hat{\sigma}_o^2$ multipliziert). Angepaßte Wald- und Likelihood-Ratio-Statistiken lassen sich aus denen des gewöhnlichen Modells durch Dividieren durch $\hat{\sigma}_o^2$ berechnen (s. oben).

Alternativ zur Berechnung per Devianz- oder Pearson-Statistik läßt sich σ_o^2 auch als weiterer unbekannter Parameter per Maximum-Likelihood schätzen. McCullagh u. Nelder (1989, S. 126) merken jedoch an, daß es sich dabei um eine zwar mathematisch einfache, jedoch evtl. wissenschaftlich nicht plausible Variante handelt, da die Schätzung von σ_o^2 hier auch die der übrigen Parameter aus $\boldsymbol{\beta}$ und damit $\hat{\boldsymbol{\mu}}$ beeinflußt.

Wenngleich die Schätzung eines Dispersionsparameters per Devianz- oder Pearson-Statistik sehr einfach ist, so wird dem Vorgehen jedoch vielfach vorgeworfen zu einfach zu sein, und Kritiker weisen darauf hin, daß das tatsächliche Problem in der Wahl der falschen Varianzfunktion liegt. Als Alternative zum überstreuten log-linearen Modell wird dann oft die Negativ-Binomial-Verteilung (Greenwood u. Yule, 1920) genannt, die hier kurz vorgestellt, im folgenden jedoch nicht weiter betrachtet werden soll. Die Verteilungsfunktion der Negativ-Binomial-Verteilung ist gegeben durch

$$P(Y = y; k, p) = \frac{\Gamma(y+k)}{\Gamma(k) \cdot \Gamma(y+1)}\, p^k\, (1-p)^y.$$

Sie beschreibt die Wahrscheinlichkeit, daß $k + y$ Versuche benötigt werden, um bei Erfolgswahrscheinlichkeit p genau y Erfolge zu erzielen. k ist also die Anzahl der Mißerfolge vor dem y-ten Erfolg. Für ein festes k gehört die Negativ-Binomial-Verteilung zur Exponentialfamilie mit kanonischem Link $\eta = \theta = \ln[\mu/(\mu + k)]$. Erwartungswert und Varianz der Negativ-Binomial-Verteilung sind $\mathrm{E}(Y) = \mu$ und $\mathrm{var}(Y) = \mu + \mu^2/k$, wobei k der sog. Aggregationsparameter ist, welcher nicht identisch zum Dispersionsparameter ϕ ist. Da die Varianz hier eine quadratische Funktion des Erwartungswertes ist, bietet die Negativ-Binomial-Verteilung eine weit größere Flexibilität in der Modellierung der Varianz als dies im Poisson-Modell der Fall ist. Je größer k wird, desto stärker nähert sich die Varianz dem Mittelwert an. Der Grenzfall der Negativ-Binomial-Verteilung für $k \to \infty$ ist also die Poisson-Verteilung. Für kleine k dagegen kann die Varianz beträchtlich größer als μ werden. Kleine k sprechen also für eine

starke Aggregation der Beobachtung, d. h. die Daten treten geklumpter auf, als in einer Poisson-verteilten Population zu erwarten wäre. Für $k = 1$ ergibt sich die geometrische Verteilung.

Für eine nähere Betrachtung der Negativ-Binomial-Verteilung s. z. B. McCullagh u. Nelder (1989, Kap. 11.2), Hilbe (1994), Lawless (1987) und Cameron u. Trivedi (1998).

Diskussionen und Fallstudien zum Problem der Überdispersion und über den Umgang mit ihr finden sich u. a. in (Collett, 1991, Kap. 6), Hinde u. Démetrio (1998), Poortema (1999) und Liang u. McCullagh (1993). Formelle Test auf Überdispersion beschreiben Dean (1992) und Cameron u. Trivedi (1998).

2.2.6. Quasi-Likelihood

Die auf den vorhergehenden Seiten beschriebene Modellanpassung und Parameterschätzung per Maximum-Likelihood (ML) war bis hierhin beschränkt auf Untersuchungsvariablen, die einer Verteilung aus der Exponentialfamilie folgen, deren Likelihood-Funktion also vollständig spezifiziert werden kann. Vielfach lassen die erhobenen Daten jedoch keinen definitiven Rückschluß auf die zugrundeliegende Verteilung zu, wohl aber auf bestimmte charakteristische Eigenschaften, wie den Einfluß externer Größen (z. B. Behandlungen) auf die Untersuchungsvariable, die Änderung der Streuung mit dem Mittelwert und die Abhängigkeit oder Unabhängigkeit der Beobachtungen voneinander. Wedderburn (1974) hat mit der *Quasi-Likelihood*-Theorie (QL) die Grundlage geschaffen, Methoden der Modellanpassung und Inferenz der gewöhnlichem ML-Theorie auch auf solche Untersuchungsvariablen zu erweitern, für die lediglich besagte Charakteristika, nicht aber die gesamte Verteilungsfunktion angegeben werden kann. Voraussetzung hierfür sind lediglich die beiden folgenden Bedingungen.

1. Es kann eine Funktion angegeben werden, die das lineare Modell $\boldsymbol{X\beta}$ mit dem Erwartungswert $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\mu}$ in Beziehung setzt.

2. Die Kovarianzmatrix der Untersuchungsvariablen läßt sich in der Form $\mathrm{cov}(\boldsymbol{y}) = \psi \boldsymbol{V}(\boldsymbol{\mu})$ spezifizieren, wobei ψ ein für alle Beobachtungen konstanter aber evtl. unbekannter Dispersions- oder Überdispersionsparameter und $\boldsymbol{V}(\boldsymbol{\mu})$ eine Matrix bekannter Funktionen $v(\mu)$ ist.

Die Funktion

$$\frac{\partial Q(\mu, y)}{\partial \mu} = \frac{y - \mu}{\psi \cdot v(\mu)} \tag{2.44}$$

besitzt wesentliche Eigenschaften der ersten Ableitung gewöhnlicher log-Likelihood-Funktionen (s. McCullagh u. Nelder, 1989, S. 325), entsprechend sollte das Integral, die Quasi-Likelihood-Funktion

$$Q(\mu, y) = \int_y^\mu \frac{y - t}{\psi \cdot v(t)} \, \mathrm{d}t,$$

die Charakteristika gewöhnlicher log-Likelihood-Funktionen besitzen, wodurch die Verwendung derselben Methoden gerechtfertigt wird. Die allgemein durch

$$\ell(y; \theta, \phi) = \left[\frac{y\,\phi - b(\theta)}{a(\phi)} \right] + c(y; \phi)$$

gegebene log-Likelihood-Funktion einer zur Exponentialfamilie gehörenden Verteilung stellt einen Sonderfall der Quasi-Likelihood-Funktionen dar (die Abhängigkeit der Funktion von μ ist durch den funktionalen Zusammenhang $\theta(\mu)$ gegeben), da die für die Schätzung des Parametervektors $\boldsymbol{\beta}$ zu lösenden Gleichungen lediglich auf Basis von

$$Q(\mu, y) = \left[\frac{y\,\phi - b(\theta)}{a(\phi)}\right]$$

bestimmt werden. Selbst für Verteilungen der Exponentialfamilie muß also nicht die volle log-Likelihood-Funktion angegeben werden.

Die Spezifikation der Quasi-Likelihood-Funktion für das Poisson-Modell mit Offset-Term und ohne Überdispersion läßt sich nach (2.30) leicht ersehen aus

$$Q(\tilde{\mu}, y) = y \cdot \ln(\tilde{\mu}) - \tilde{\mu}$$

und

$$\frac{\partial Q(\tilde{\mu}, y)}{\partial \tilde{\mu}} = \frac{y}{\tilde{\mu}} - 1 = \frac{y - \tilde{\mu}}{\tilde{\mu}}.$$

Die Gleichung in (2.44) ist also mit $v(\tilde{\mu}) = \tilde{\mu}$ und $\psi = \phi = 1$ erfüllt. Im überstreuten Poisson-Modell mit Offset-Term, $v(\tilde{\mu}) = \tilde{\mu}$ und $\mathrm{var}(y) = \sigma_o^2 \cdot v(\tilde{\mu})$ ist

$$Q(\tilde{\mu}, y) = \frac{y \cdot \ln(\tilde{\mu}) - \tilde{\mu}}{\sigma_o^2}$$

und

$$\frac{\partial Q(\tilde{\mu}, y)}{\partial \tilde{\mu}} = \frac{y - \tilde{\mu}}{\sigma_o^2\,\tilde{\mu}}.$$

Der Skalierungs- oder Dispersionsparameter ϕ ist hier weiterhin gleich Eins, hinzu kommt jedoch der Überdispersionsparameter σ_o^2, der in der Formulierung der Exponentialfamilie nicht vorgesehen ist.

Die Schätzung des Parametervektors $\boldsymbol{\beta}$ erfolgt wie im Falle der gewöhnlichen ML-Methode (vergl. Kap. 2.2.2) durch Ableiten (nach $\boldsymbol{\beta}$) und Nullsetzen der resultierenden *Quasi-Score-Funktionen*

$$\boldsymbol{s}(\boldsymbol{\beta}) = \boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})/\psi. \tag{2.45}$$

wobei $\boldsymbol{D}$ wiederum die $n \times p$ Matrix der partiellen Ableitungen $\partial\boldsymbol{\mu}/\partial\boldsymbol{\beta}$ ist.

Unter der Annahme einer korrekt angegebenen Mittelwert- und Varianzstruktur, wie in Wedderburn (1974), McCullagh (1983) und McCullagh u. Nelder (1989, Kap. 9.2) angenommen, berechnet sich die Kovarianzmatrix der Quasi-Score-Funktionen analog zur erwarteten Fisher Informationsmatrix aus (2.26) durch

$$\boldsymbol{F} = \boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D}/\psi = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}/\psi, \tag{2.46}$$

die asymptotische Kovarianzmatrix von $\boldsymbol{\beta}$ durch

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = \psi(\boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D})^{-1} = \psi(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} = \boldsymbol{F}^{-1},$$

wobei $\boldsymbol{W} = \boldsymbol{\Delta}^{-1'}\boldsymbol{V}^{-1}\boldsymbol{\Delta}^{-1}$ und $\boldsymbol{\Delta}$ die $n \times n$ Diagonalmatrix mit den Elementen $\mathrm{diag}\,[\partial\eta_i/\partial\mu_i]$ ist. In dieser Darstellung ist leicht zu erkennen, welchen Einfluß

ein etwaiger Überdispersionsparameter ($\psi = \sigma_o^2$) auf die Kovarianzmatrix und damit die Präzision der Modellparameter nimmt.

$\hat{\boldsymbol{\beta}}$ ist wiederum konsistent und asymptotisch normal.

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} \mathrm{N}(\boldsymbol{\beta}, \boldsymbol{F}^{-1})$$

Häufigste Anwendungsgebiete für Quasi-Likelihood sind die beschriebene zusätzliche Schätzung eines Überdispersionsparameters und der nachfolgend näher erläuterte Einbezug der Korrelation zwischen Beobachtungen. Wedderburn (1974), McCullagh (1983) und McCullagh u. Nelder (1989, Kap. 9) beschreiben die Quasi-Likelihood-Theorie im Detail.

2.2.7. Räumliche Autokorrelation in GLMs

Oft kann – z. B. bedingt durch das Versuchsdesign – nicht von der bis hierhin für die Inferenz geforderten Unabhängigkeit der Beobachtungen ausgegangen werden. Als gängigstes Beispiel sind hier die auch in gewöhnlichen linearen Modellen häufig angewendeten Meßwiederholungen zu nennen. Die untersuchte Größe wird hierbei mehrfach am selben Objekt erhoben, neuere Messungen werden daher stark mit vorhergehenden Messungen am selben Objekt korrelieren (Autokorrelation). Die Korrelation ist hier also zeitlich bedingt. Wird diese Korrelation nicht im Modell berücksichtigt, so können daraus deutlich unterschätzte Standardfehler der Parameterschätzungen und letztlich zu große Teststatistiken und zu enge Konfidenzintervalle resultieren. Bei Daten mit geographischem Bezug läßt sich oftmals, wie Toblers Gesetz der Geographie „*Everything is related to everything else, but near things are more related than distant things*“ (Tobler, 1970) formuliert, oftmals (positive) Autokorrelation beobachten, d. h. nahe gelegene Beobachtungen sind sich in der Regel ähnlicher als weiter entfernte. Zur bisher beschriebenen Erklärung oder Vorhersage von Beobachtungen durch Kovariaten kommt hier also deren räumlicher Bezug zu anderen Beobachtungen ins Spiel. Um dies auch in der Notation der nachfolgend erläuterten Modelle zu verdeutlichen, werden die betreffenden Größen mit einer Positionsangabe $\boldsymbol{s}$ versehen. So ist $y(\boldsymbol{s}_i)$ die Realisation der Zufallsvariablen Y an der Stelle $\boldsymbol{s}_i$, wobei der Vektor $\boldsymbol{s}$ in der Regel zweidimensionale kartesische Koordinaten enthält.

Der Einbezug und die Analyse räumlicher Autokorrelation in gewöhnlichen linearen Modellen ist – insbesondere unter dem Schlagwort Geostatistik – seit mehreren Jahrzehnten Forschungsthema, dem sich zahlreiche Lehrbücher widmen (s. z. B. Matheron, 1971; Journel u. Huijbregts, 1978; Ripley, 1981; Cressie, 1991; Goovaerts, 1997; Chilès u. Delfiner, 1999; Olea, 1999; Schabenberger u. Gotway, 2004). Die Entwicklung von Modellen und Methoden, welche die Anwendung geostatistischer und weiterer Verfahren der räumlichen Statistik auch im Rahmen von GLMs, also auch für nicht normalverteilte Daten, ermöglichen, hat – teilweise zusätzlich beschränkt durch die bis dahin unzureichende Rechenleistung von Computern – erst Mitte der 1990er Jahre begonnen und ihre Umsetzung gehört bis heute nur sehr bedingt zum Standardumfang statistischer Software. Eine wesentliche theoretische Hürde war, daß die üblichen Annahmen klassischer geostatistischer Verfahren wie gewöhnliches und universelles *Kriging*

(Krige, 1951) im Falle nicht normalverteilter Daten von vornherein nicht erfüllt sind. Hierbei ist insbesondere die Annahme der Stationarität 2. Ordnung zu nennen. Dazu gehört die konstante Varianz der untersuchten Größe an jeder beliebigen Position im Untersuchungsgebiet. Werden Anteile, Anzahlen oder andere nicht normalverteilte Größen *gekrigt*, so ist diese Bedingung oftmals nicht erfüllt, da die Varianz hier meist nicht konstant, sondern eine Funktion des Erwartungswertes ist.[5] Wie bei der Anpassung gewöhnlicher linearer Modelle an nicht normalverteilte Daten kommt auch hier das Problem hinzu, daß die linearen Modelle der Geostatistik (theoretisch) nicht auf das mögliche Wertespektrum der Untersuchungsvariablen beschränkt sind und damit unzulässige Vorhersagen möglich werden.

Inzwischen existieren aber Lösungsansätze, die den Einbezug räumlicher Kovarianzstrukturen in GLMs zu ermöglichen (z. B. Albert u. McShane, 1995; Gotway u. Stroup, 1997; Diggle et al., 1998; Gotway u. Wolfinger, 2003). Zwei grundlegende und gängige Ansätze sollen hier nachfolgend erläutert werden. In diesem Zusammenhang werden Begriffe und Methoden aus dem Bereich der Geostatistik erwähnt, auf die jedoch nicht detailliert eingegangen werden kann. Für allgemeine Einführungen und Detailbetrachtungen der Methoden und Verfahren der Geostatistik und weiterer räumlicher Analysemethoden sei daher auf die oben genannte Literatur verwiesen.

2.2.7.1. Der GEE-Ansatz

Wie im oben beschriebenen Fall des überstreuten Poisson-Modells kann auch für den Fall (räumlich, zeitlich oder andersartig) abhängiger Beobachtungen die gemeinsame Likelihood-Funktion nicht vollständig spezifiziert werden. Bei Kenntnis der Mittelwert- und Kovarianzstruktur können aber auch hier die Quasi-Score-Gleichungen – in diesem Zusammenhang *Generalized Estimating Equations* (GEE) genannt –

$$\boldsymbol{s}(\boldsymbol{\beta}) = \boldsymbol{D}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}(\boldsymbol{s}) - \boldsymbol{\mu}(\boldsymbol{s})) = 0 \tag{2.47}$$

zur Schätzung des Parametervektors $\boldsymbol{\beta}$ verwendet werden, wobei $\boldsymbol{D} = \partial\boldsymbol{\mu}(\boldsymbol{s})/\partial\boldsymbol{\beta}$ und $\boldsymbol{\Sigma}$ die Kovarianzmatrix der Beobachtungen ist. Sie ersetzt die Matrix $\boldsymbol{V}$ aus (2.45), deren Diagonalform die Unabhängigkeit der Beobachtungen implizierte. GEEs verwenden also dieselbe Link-Funktion und systematische Komponente ($\boldsymbol{X}\boldsymbol{\beta}$) wie gewöhnliche GLMs, beziehen zusätzlich aber die Abhängigkeit zwischen den Beobachtungen in Form der Kovarianzmatrix $\boldsymbol{\Sigma}$ ein. McCullagh u. Nelder (1989, S. 333 ff.) merken zwar an, daß die Inverse der Kovarianzmatrix $\boldsymbol{\Sigma}$ bestimmte Bedingungen erfüllen muß, um eine Lösung der GEEs aus (2.47) zu gewährleisten, Liang u. Zeger (1986) und Zeger u. Liang (1986) konnten für Meßwiederholungsdaten (McShane et al. (1997) für räumlich korrelierte Daten) aber zeigen, daß $\boldsymbol{\beta}$ konsistent geschätzt werden kann, wenn sich die Kovarianz-

[5] Auch im Fall normalverteilter Daten ist diese Bedingung nicht automatisch erfüllt. Da die Varianz hier aber im Gegensatz zu den übrigen Verteilungen der Exponentialfamilie keine Funktion des Erwartungswertes ist, kann oftmals vereinfachend von annähernd konstanter Varianz ausgegangen werden.

matrix in der Form

$$\boldsymbol{\Sigma}(\boldsymbol{\mu}, \boldsymbol{\alpha}) = \boldsymbol{V}^{1/2}\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{V}^{1/2} \tag{2.48}$$

darstellen läßt, wobei $\boldsymbol{R}(\boldsymbol{\alpha})$ eine $n \times n$ Matrix, die sog. Arbeitskorrelationsmatrix (*Working Correlation Matrix*) und $\boldsymbol{V}^{1/2}$ eine Diagonalmatrix mit den Elementen diag $\left[\sqrt{v(\mu_i)}\right]$ ist.

Einen sog. *Nugget*-Effekt σ_n^2 und einen Überdispersionsparameter σ_o^2 einbeziehend, läßt sich (2.48) erweitern zu

$$\boldsymbol{\Sigma}(\boldsymbol{\mu}, \boldsymbol{\alpha}) = \sigma_n^2\boldsymbol{V} + \sigma_o^2\boldsymbol{V}^{1/2}\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{V}^{1/2} \tag{2.49}$$

(Gotway u. Wolfinger, 2003). Die Elemente von $\boldsymbol{R}(\boldsymbol{\alpha})$ werden in der Praxis anhand der entfernungs- und richtungsabhängigen Korrelationsfunktion des räumlichen Prozesses $\rho(\boldsymbol{s}_i - \boldsymbol{s}_j; \boldsymbol{\alpha})$ bestimmt, deren unbekannte Parameter $\boldsymbol{\alpha}$ (Schwellenwert und effektive Reichweite) die räumliche Abhängigkeit vollständig spezifizieren und anhand des empirischen Semivariogramms[6] (Matheron, 1962, 1963) bestimmt werden. Die Varianz einer einzelnen Beobachtung ist dann (einen Nugget-Effekt einbeziehend)

$$\mathrm{var}[y(\boldsymbol{s}_i)] = (\sigma_n^2 + \sigma_o^2) \cdot v[\mu(\boldsymbol{s}_i)]$$

und die Kovarianz zwischen zwei beliebigen Beobachtungen

$$\mathrm{cov}[y(\boldsymbol{s}_i), y(\boldsymbol{s}_j)] = \sigma_o^2\sqrt{v\left[\mu(\boldsymbol{s}_i)\right] \cdot v\left[\mu(\boldsymbol{s}_j)\right]} \cdot \rho(s_i - s_j; \boldsymbol{\alpha}).$$

Die durch $\rho(s_i - s_j; \boldsymbol{\alpha})$ spezifizierte Korrelationsstruktur sollte der wahren aus Effizienzgründen und insbesondere für räumliche Vorhersagen (s. u.) möglichst ähnlich sein.

Für voneinander unabhängige Beobachtungen und damit $\boldsymbol{R}(\boldsymbol{\alpha}) = \boldsymbol{I}$ ist leicht zu ersehen, daß die GEEs aus (2.47) zu den Score-Funktionen gewöhnlicher GLMs aus (2.23) werden.

Von Zeger u. Liang (1986) und Liang u. Zeger (1986) ursprünglich für die Auswertung von Meßwiederholungsdaten mit nicht normaler Verteilung entwickelt, wurde die Darstellung aus (2.48) erstmals von Albert u. McShane (1995) auch zur Beschreibung räumlicher Kovarianzstrukturen verwendet. Gotway u. Stroup (1997) haben kurze Zeit später allgemeine Modelle für die Parameterschätzung und insbesondere auch für die räumliche Vorhersage auf Basis von GEEs vorgestellt. Zur Lösung der GEEs verwenden sie die bereits in Kapitel 2.2.2 für gewöhnliche GLMs vorgestellte Darstellung des Fisher Scorings durch iterativ gewichtete Kleinste-Quadrate-Schätzungen

$$\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}^*(\boldsymbol{s}), \tag{2.50}$$

mit

$$\boldsymbol{W} = \boldsymbol{\Delta}^{-1'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}^{-1},$$

[6] Eine Auflistung und Beschreibung möglicher theoretischer Semivariogramm-Modelle findet sich z. B. in Cressie (1991, S. 61 f.) und Journel u. Huijbregts (1978, S. 163-171).

und

$$\boldsymbol{y}^*(\boldsymbol{s}) = \boldsymbol{\eta}(\boldsymbol{s}) + \boldsymbol{\Delta}(\boldsymbol{y}(\boldsymbol{s}) - \boldsymbol{\mu}(\boldsymbol{s})),$$

wobei $\boldsymbol{\Delta} = \operatorname{diag}[\partial\eta(\boldsymbol{s}_i)/\partial\mu(\boldsymbol{s}_i)]$. Zunächst unter der Annahme bekannter Korrelationsparameter $\boldsymbol{\alpha}$ ergibt sich die Parameterschätzung für die linearisierte Variable $\boldsymbol{y}^*(\boldsymbol{s})$ nach der Methode der verallgemeinerten kleinsten Quadrate (*Generalized Least Squares, GLS*) durch iterative Lösung der Gleichungen

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{y}^*(\boldsymbol{s}). \tag{2.51}$$

Da $\boldsymbol{\alpha}$ in der Regel aber ebenfalls unbekannt sein wird, erfolgt die Schätzung von $\boldsymbol{\beta}$ und $\boldsymbol{\alpha}$ iterativ abwechselnd. Begonnen wird zumeist mit der Annahme der Unabhängigkeit der Beobachtungen, also $\boldsymbol{R}(\boldsymbol{\alpha}) = \boldsymbol{I}$, und einer ersten Parameterschätzung $\hat{\boldsymbol{\beta}}_{(1)}$ als Lösung der Gleichungen aus (2.51). Dem folgt eine Schätzung der Korrelationsparameter $\boldsymbol{\alpha}$ – meist auf Basis der mit $\boldsymbol{\beta}_{(1)}$ erzielten Residuen $(\boldsymbol{y}(\boldsymbol{s}) - \hat{\boldsymbol{\mu}}_{(1)}(\boldsymbol{s}))$ – und der Kovarianzmatrix aus (2.48) oder (2.49). Im einfachsten Fall (z. B. Gotway u. Stroup, 1997) wird im Anschluß daran die Schätzung des Parametervektors $\boldsymbol{\beta}$ sowie der zugehörigen Streuungsmaße nur einmal aktualisiert ($\hat{\boldsymbol{\beta}}_{(2)}$), davon ausgehend, daß die so ermittelten Kovarianzparameter die wahren sind. Idealerweise wird jedoch ein doppelt iterativer Algorithmus verwendet, in dem sich der Aktualisierung von $\hat{\boldsymbol{\beta}}$ so lange eine Aktualisierung von $\hat{\boldsymbol{\alpha}}$ anschließt, bis ein gemeinsames Konvergenzkriterium erfüllt ist.

Zur Schätzung von $\boldsymbol{\alpha}$ gibt es gleich mehrere Ansätze. Ein häufig angewandter ist, die Pearson-Residuen

$$r^P(\boldsymbol{s}_i) = \frac{y(\boldsymbol{s}_i) - \hat{\mu}(\boldsymbol{s}_i)}{\sqrt{v\left[\hat{\mu}(\boldsymbol{s}_i)\right]}}$$

für eine gutachterliche Anpassung einer theoretischen Semivariogrammfunktion an das empirische Semivariogramm zu verwenden, wobei empfohlen wird (z. B. Albert u. McShane, 1995; Schabenberger u. Gotway, 2004, S. 362), die Residuen zuvor wie in (2.36) zu studentisieren, um deren Heteroskedastizität zu reduzieren. Für die Pearson-Residuen r^P gilt im Gegensatz zu den beobachteten Werten y und den rohen Residuen $(y - \mu)$ approximativ die Annahme eines konstanten Mittelwertes und einer konstanten Varianz (Stationarität 2. Ordnung), die Interpretation der anhand der Pearson-Residuen erstellten empirischen Semivariogramme ist daher zulässig.

Mit Hilfe der Pearson-Residuen läßt sich ggf. auch der Überdispersionsparameter

$$\sigma_o^2 = \frac{1}{n-p}\sum_{i=1}^{n}\left(r^P(\boldsymbol{s}_i)\right)^2$$

konsistent schätzen.

Die gutachterliche Schätzung der Korrelationsparameter bietet sich jedoch wegen des damit verbundenen Aufwandes nur bei einer sehr geringen Anzahl an Aktualisierungen der Schätzungen von $\boldsymbol{\beta}$ und $\boldsymbol{\alpha}$ an. Im Zuge des doppelt iterativen Algorithmus werden daher vordergründig automatisierte Anpassungen der Semivariogrammfunktion erfolgen. Dies kann entweder mit Hilfe der

Methode der kleinsten Quadrate oder wegen der Heteroskedastizität der Beobachtungen des empirischen Semivariogramms besser über gewichtete kleinste Quadrate (Cressie, 1985) geschehen.

Alternativ läßt sich $\boldsymbol{\alpha}$ auch über die Formulierung weiterer GEEs oder über die sog. *Composite-Likelihood*-Methode (CL) schätzen. Siehe Prentice (1988) und Zhao u. Prentice (1990) für die Herleitung zusätzlicher GEEs und für die Anwendung auf Meßwiederholungsdaten sowie Schabenberger u. Gotway (2004, Kap. 4.5.3.1) und Albert u. McShane (1995) für die Anwendung auf räumliche Daten. Lindsay (1988) beschreibt die allgemeine Herleitung der CL-Methode, Lele (1997), Heagerty u. Lele (1998) und Schabenberger u. Gotway (2004, Kap. 4.5.3.2) wenden sie zur Schätzung räumlicher Kovarianzparameter an.

Des weiteren läßt sich der GEE-Ansatz dahingehend erweitern, $\boldsymbol{\alpha}$ und $\boldsymbol{\beta}$ simultan zu schätzen (Zhao u. Prentice, 1990; Liang et al., 1993), was auf der einen Seite mit Effizienzgewinnen (klein für $\boldsymbol{\beta}$, größer für $\boldsymbol{\alpha}$) einhergeht, auf der anderen aber zusätzlicher Annahmen (insbesondere der der korrekten Spezifizierung der Korrelationsstruktur) bedarf.

2.2.7.2. Der PL-Ansatz

Als Alternative zu Quasi-Likelihood und maßgeblich, um die Aufnahme zufälliger Effekte in GLMs (dann *Generalized Linear Mixed Models*, GLMM) zu ermöglichen, haben Wolfinger u. O'Connel (1993) das *Pseudo-Likelihood-* (PL) bzw. *Restricted-Pseudo-Likelihood*-Verfahren (REPL) vorgeschlagen. Beide Ansätze unterscheiden sich im wesentlichen dadurch vom QL-Verfahren des vorhergehenden Abschnitts, daß hier in jedem Iterationsschritt der Modellanpassung eine echte Likelihood-Funktion verwendet wird, um die unbekannten Parameter zu schätzen. Die Formulierung von GLMMs erlaubt zwei Herangehensweisen an die Berücksichtigung räumlicher Autokorrelation: Im *marginalen* Modell werden räumlich korrelierte Residuen angenommen, wohingegen im *bedingten* Modell die räumliche Autokorrelation durch einen latenten, stationären Gauß-Prozeß $\boldsymbol{S}$ (in Form zufälliger Effekte) hervorgerufen wird. Die Beobachtungen (und Residuen) sind dann bedingt (gegeben $\boldsymbol{S}$) unabhängig. Hier soll lediglich das marginale Modell (als Spezialfall des bedingten) näher betrachtet werden; eine Betrachtung und der Vergleich beider Ansätze finden sich in Gotway u. Wolfinger (2003) und Schabenberger u. Gotway (2004, Kap. 6.3.5.2). Ebenfalls nicht berücksichtigt werden soll hier der *Penalized-Quasi-Likelihood*-Ansatz von Breslow u. Clayton (1993) der ebenfalls die Formulierung von GLMMs ermöglicht.

Grundidee des PL-Ansatzes[7] ist, den Erwartungswert über eine Taylor-Entwicklung um einen Vektor $\tilde{\boldsymbol{\beta}}$ zu linearisieren. Aus der inversen Link-Funktion wird dann

$$\mathrm{E}(\boldsymbol{y}) = g^{-1}(\boldsymbol{X}\boldsymbol{\beta}) = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu} \doteq \tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{\Delta}}^{-1}\boldsymbol{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}),$$

worin $\tilde{\boldsymbol{\Delta}} = \mathrm{diag}\,[\partial\eta_i/\partial\mu_i]_{\tilde{\boldsymbol{\beta}}}$ und $\tilde{\mu} = g^{-1}(\boldsymbol{X}\tilde{\boldsymbol{\beta}})$. Dieser Ausdruck läßt sich um-

[7] Es wird hier nur der Teil der GLMM-Theorie aufgeführt, der für die marginale Modellformulierung benötigt wird.

formulieren zu

$$\boldsymbol{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\Delta}}(\boldsymbol{\mu} - \tilde{\boldsymbol{\mu}}) \doteq \boldsymbol{X}\boldsymbol{\beta}.$$

Die linke Seite dieses Ausdrucks ist der Erwartungswert von

$$\boldsymbol{X}\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{\Delta}}(\boldsymbol{Y} - \tilde{\boldsymbol{\mu}}) \equiv \boldsymbol{\nu}.$$

Das Modell

$$\boldsymbol{\nu} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu}) \tag{2.52}$$

ist dann ein lineares Modell mit $\boldsymbol{\nu}$ als abhängiger Variablen, den Parametern der festen Effekte $\boldsymbol{\beta}$ und $\mathrm{var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma}_\nu$. $\boldsymbol{\nu}$, von Wolfinger u. O'Connel (1993) als *Pseudo-Daten* bezeichnet, ist identisch zur linearisierten Variablen $\boldsymbol{y}^*$ aus dem GEE-Ansatz. (Marginaler) Erwartungswert und Varianz dieser Pseudo-Daten sind nach Wolfinger u. O'Connel (1993)

$$\begin{aligned} \mathrm{E}(\boldsymbol{\nu}) &= \boldsymbol{X}\boldsymbol{\beta} \qquad \text{und} \\ \mathrm{cov}(\boldsymbol{\nu}) &= \boldsymbol{\Sigma}_\nu(\boldsymbol{\mu}, \boldsymbol{\alpha}), \end{aligned} \tag{2.53}$$

mit

$$\boldsymbol{\Sigma}_\nu = \psi \boldsymbol{\Delta} \boldsymbol{V}^{1/2} \boldsymbol{R}(\boldsymbol{\alpha}) \boldsymbol{V}^{1/2} \boldsymbol{\Delta} = \boldsymbol{\Delta}\boldsymbol{\Sigma}(\boldsymbol{\mu}, \boldsymbol{\alpha})\boldsymbol{\Delta}, \tag{2.54}$$

worin ψ wiederum ein Skalierungsparameter (Dispersions- oder Überdispersionsparameter) ist. Die Formulierung verallgemeinerter linearer gemischter Modelle ergibt sich aus der obigen durch eine zusätzliche Taylor-Entwicklung über den Vektor der zufälligen Effekte.

Da der Mittelwert der Pseudo-Daten linear bezüglich $\boldsymbol{\beta}$ ist, lassen sich diese mit Hilfe gewöhnlicher linearer Modelle mit räumlich korrelierten Fehlern betrachten und analysieren. Unter der Annahme, daß $\boldsymbol{\nu}$ approximativ normal verteilt ist mit Mittelwert $\boldsymbol{X}\boldsymbol{\beta}$ und Kovarianz $\boldsymbol{\Sigma}_\nu$, lassen sich für die Schätzung der Parameter der fixen Effekte $\boldsymbol{\beta}$ und der Korrelationsparameter $\boldsymbol{\alpha}$ Maximum-Likelihood- oder *Restricted-Maximum-Likelihood*-Methoden (REML) verwenden. Schabenberger u. Gotway (2004, S. 366) merken hierzu an, daß es der Annahme der Normalverteilung genau genommen gar nicht bedarf; vielmehr heißt es dort, die unter der Normalverteilungsannahme zu maximierende Likelihood- oder Restricted-Likelihood-Funktion sei – unabhängig von der wahren Verteilung der ν – geeignet und „nur ein Fahrzeug“ zum Erreichen der Parameterschätzungen und Vorhersagen. Je nachdem, ob für die Modellanpassung ML oder REML zum Einsatz kommt, handelt es sich um den PL- bzw. um den REPL-Ansatz von Wolfinger u. O'Connel (1993). REML- werden im allgemeinen ML-Methoden vorgezogen, da letztere größeren negativen Bias der Kovarianzparameter zeigen. Ursache hierfür ist – vereinfacht ausgedrückt – die Unfähigkeit der ML-Methode, die Anzahl der zu schätzenden Parameter der festen Effekte in der Schätzung der Kovarianzparameter zu berücksichtigen. Näheres zur Schätzung der Kovarianzparameter per ML oder REML findet sich u. a. in Mardia u. Marshall (1984), Dietrich u. Osborne (1991) und Pardo-Iguzquisa (1998).

Da sich auch hier die jeweils aktuellen Schätzungen von $\boldsymbol{\Sigma}_\nu$ und $\boldsymbol{\beta}$ beeinflussen, ist erneut doppelt iterativ vorzugehen. Erleichtert wird das Procedere dadurch, daß sich für $\hat{\boldsymbol{\beta}}$ und $\hat{\psi}$ geschlossene Formen aus der zu maximierenden Likelihood-Funktion auskoppeln lassen (sog. *profiling*, für eine Erläuterung

s. z. B. Schabenberger u. Gotway, 2004, Kap. 5.5.2, 5.5.3 u. 6.3.5.2). So ergibt sich für die Schätzung des Parametervektors $\boldsymbol{\beta}$ wiederum – wie im GEE-Ansatz – ein Gleichungssystem verallgemeinerter kleinster Quadrate (GLS)

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}_\nu^{-1}\boldsymbol{\nu}. \tag{2.55}$$

Enthält das Modell einen zu schätzenden Skalierungsparameter ($\psi \neq 1$), so ergibt sich für diesen im PL-Ansatz

$$\hat{\psi} = \frac{(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}_\nu^*)^{-1}(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n}, \tag{2.56}$$

und im REPL-Ansatz

$$\hat{\psi} = \frac{(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{\Sigma}_\nu^*)^{-1}(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}{n - \operatorname{rang}(\boldsymbol{X})}, \tag{2.57}$$

worin sich $\boldsymbol{\Sigma}_\nu^*$ durch Teilen von $\boldsymbol{\Sigma}_\nu$ durch die Residualvarianz ergibt. Die iterative Schätzung der Parameter per PL oder REPL läuft dann in den Schritten:

1. Initialisierung der Pseudo-Daten aus (2.52) durch ein gewöhnliches (auf Unabhängigkeit der Beobachtungen basierendes) GLM.
2. Schätzung der räumlichen Autokorrelationsparameter $\boldsymbol{\alpha}$ per ML (PL) oder REML (REPL) durch Minimierung der jeweiligen Likelihood-Funktion und Berechnung von $\hat{\boldsymbol{\Sigma}}_\nu$.
3. Berechnung von $\hat{\boldsymbol{\beta}}$ und ggf. $\hat{\psi}$ per (2.55) und (2.56) bzw. (2.57).
4. Aktualisierung der Pseudo-Daten.
5. Wiederholung der Schritte 2-4, bis ein vorgegebenes Konvergenzkriterium erreicht ist.

Es wird empfohlen (Albert u. McShane, 1995; Littel et al., 1996, S. 330), im Anschluß an die Modellanpassung eine Überprüfung der berechneten Kovarianzparameter vorzunehmen, indem die geschätzte Semivariogrammfunktion gemeinsam mit dem empirischen Semivariogramm der (studentisierten) Pearson-Residuen des finalen Modells dargestellt werden.

Detailinformationen zu GLMMs finden sich u. a. in Wolfinger u. O'Connel (1993), Breslow u. Clayton (1993), McCulloch u. Searle (2001) und SAS (2006) sowie speziell für den Fall räumlicher Autokorrelation in Diggle et al. (1998), Gotway u. Wolfinger (2003), Schabenberger u. Gotway (2004, Kap. 6.3), Christensen et al. (2000), Christensen u. Waagepetersen (2002) und Zhang (2002).

2.2.7.3. Varianzschätzung, Hypothesentests und Modellauswahl

Bei den Gleichungen zur Berechnung des Parametervektors $\boldsymbol{\beta}$ handelt es sich sowohl beim GEE- wie auch beim PL-Ansatz um solche der Methode der verallgemeinerten kleinsten Quadrate (GLS), so daß für die der Modellanpassung folgende Inferenz dieselben Verfahren verwendet werden können.

Die Parameterschätzung $\hat{\boldsymbol{\beta}}$ ist konsistent und asymptotisch normalverteilt mit (modellbasierter) Kovarianzmatrix

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1} = (\boldsymbol{D}'\boldsymbol{\Sigma}^{-1}\boldsymbol{D})^{-1}, \tag{2.58}$$

sofern es sich bei $\boldsymbol{R}(\boldsymbol{\alpha})$ um die wahre Korrelationsmatrix handelt. Die Schätzung von $\boldsymbol{\beta}$ ist auch dann konsistent, wenn die Korrelationsstruktur in Form der Korrelationsmatrix $\boldsymbol{R}(\boldsymbol{\alpha})$ falsch spezifiziert wurde (Liang u. Zeger, 1986); es ist lediglich die korrekte Spezifikation der Mittelwertstruktur notwendig. Für die verläßliche Schätzung der zugehörigen Streuungsmaße und für räumliche Vorhersagen (s. u.) dagegen ist die Angabe der korrekten Korrelationsstruktur von entscheidender Bedeutung. So haben Albert u. McShane (1995) in einer Simulationsstudie gezeigt, daß die Annahme der Unabhängigkeit in einem tatsächlich räumlich autokorrelierten Prozeß nur geringe Auswirkungen auf die Parameterschätzungen $\hat{\boldsymbol{\beta}}$ hat, Varianzen jedoch deutlich unterschätzt werden. Im Falle unbekannter und nur unsicher zu bestimmender räumlicher Korrelation empfehlen sie, die Kovarianz aus (2.58) durch eine robuste Alternative, einen sog. *Sandwich-Schätzer*, zu ersetzen. In SAS (2006, S. 19 ff., 128 f.) sind zahlreiche Erweiterungen und Anpassungen dieses Schätzers aufgeführt. Sie alle setzen jedoch voraus, daß für die Auswertung mindestens zwei voneinander unabhängige Subjekte mit gleicher interner Korrelationsstruktur zur Verfügung stehen. Da dies im räumlichen Kontext nur selten gegeben ist, soll auf die robuste Varianzschätzung hier nicht weiter eingegangen werden; für ein Beispiel, in dem diese angewendet wird, s. Albert u. McShane (1995). In der Deutung von Streuungsmaßen und Hypothesentests ist die etwaige Ungewißheit der Korrelationsstruktur jedoch entsprechend zu berücksichtigen.

Die Schätzung von $\boldsymbol{\alpha}$ ist nur dann konsistent, wenn die Mittelwertstruktur und die Kovarianzfunktion korrekt spezifiziert wurden.

Hypothesen der Form $\boldsymbol{L}'\boldsymbol{\beta} = 0$ lassen sich mit Hilfe der angepaßten Wald-Statistik

$$\tilde{W} = \frac{(\boldsymbol{L}\hat{\boldsymbol{\beta}})'(\boldsymbol{L}[\boldsymbol{X}'\boldsymbol{\Sigma}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\alpha}})^{-1}\boldsymbol{X}]^{-1}\boldsymbol{L}')^{-1}(\boldsymbol{L}\hat{\boldsymbol{\beta}})}{\operatorname{rang}(\boldsymbol{L})} \tag{2.59}$$

testen. Zugehörige p-Werte können anhand einer F-Verteilung mit $\operatorname{rang}(\boldsymbol{L})$ Zähler- und $n - \operatorname{rang}[\boldsymbol{X}(\boldsymbol{s})]$ Nennerfreiheitsgraden bestimmt werden. Soll nur ein einzelner Parameter auf Gleichheit mit Null getestet werden ($H_0 : \beta_j = 0$), ist $\boldsymbol{L}$ also ein $1 \times p$ Vektor gefüllt mit Nullen außer einer Eins für den j-ten Eintrag, so vereinfacht sich der Test zur bekannten T-Statistik

$$t = \frac{\hat{\boldsymbol{\beta}}_j}{\sqrt{\hat{s}^2_{jj}}},$$

worin $\hat{s}^2_{jj}$ der j-te Diagonaleintrag der Kovarianzmatrix aus (2.58) ist. Die Statistik besitzt unter der Nullhypothese $k = n - \operatorname{rang}[\boldsymbol{X}(\boldsymbol{s})]$ Freiheitsgrade, so daß sich neben p-Werten zur obigen Teststatistik Konfidenzintervalle der Parameterschätzungen über

$$\hat{\boldsymbol{\beta}}_j \pm t_{1-\alpha/2;k} \cdot \sqrt{\hat{s}^2_{jj}}$$

berechnen lassen. Alternativ können diese wegen der asymptotischen Normalverteilung der Parameterschätzungen auch über das entsprechende Quantil der Standardnormalverteilung bestimmt werden.

Die von den gewöhnlichen GLMs bekannten Maße zur Beurteilung der Anpassungsgüte, Devianz- und Pearson-Statistik sowie klassische, in Likelihood-Verfahren zur Anwendung kommende Maße zur Modellauswahl wie AIC und AICc dürfen für korrelierte Daten nicht mehr interpretiert werden, da sie die Kenntnis der vollständigen Likelihood-Funktion voraussetzen. Hinzu kommt, daß das Hinzufügen oder Weglassen von Kovariaten den Aufbau von $\boldsymbol{X}$ und damit die Linearisierung der Daten und letztlich die Pseudo-Daten selbst verändert. Die Likelihood-Funktionen einzelner Modelle sind damit nicht vergleichbar. Einzige Ausnahme hiervon ist – bei gleicher Linearisierung – der Vergleich bezüglich $\boldsymbol{R}(\boldsymbol{\alpha})$ geschachtelter Modelle, also beispielsweise der Vergleich eines Modells mit Nugget-Effekt mit einem Modell ohne Nugget-Effekt.

Eine Adaption des AIC an GEEs ist das QIC_u (Hardin u. Hilbe, 2003, S. 142), das jedoch bislang nicht in statistischer Software verfügbar ist. Wie das AIC eignet es sich zur objektiven Auswahl eines aus mehreren zur Verfügung stehenden genesteten oder nichtgenesteten Modellen und bezieht dabei, ähnlich dem AIC, den von einem Modell erzielten Wert der Quasi-Likelihood-Funktion und die Anzahl der zu schätzenden Parameter mit ein. Ebenfalls nicht zum Standardumfang statistischer Software gehört das QIC (Pan, 2001; s. auch Hardin u. Hilbe, 2003, S. 139 ff.), das unterstützend zur Auswahl eines geeigneten Kovarianzmodells herangezogen werden kann. Hardin u. Hilbe (2003, Kap. 4) geben eine verständliche und allgemeine Einführung in mögliche Verfahren und Ansätze zur Modellauswahl und -evaluation in GEEs zur Analyse von (zeitlichen) Meßwiederholungsdesigns.

2.2.7.4. Räumliche Vorhersagen

Gerade im Fall raumbezogener Daten wird mit der Berücksichtigung räumlicher Autokorrelation oftmals nicht oder nicht nur auf die verbesserte Schätzung von Effekten (im Sinne der erwartungstreuen und präzisen Schätzung von Parametern) abgezielt, sondern auch auf die Vorhersage von Beobachtungen $y(\boldsymbol{s}_0)$ an nicht erfaßten Positionen $\boldsymbol{s}_0$. Im Bereich der linearen Modelle ist Kriging ein zu diesem Zweck oftmals eingesetztes Verfahren. Die Übertragung der Modelle auf GLMs ist jedoch nicht ohne weiteres möglich (s. u.); ihr haben sich u. a. Gotway u. Stroup (1997) und Gotway u. Wolfinger (2003) gewidmet (s. auch Schabenberger u. Gotway, 2004, Kap. 6.3.6), deren Ansätze hier teilweise vorgestellt werden sollen. Begonnen wird jedoch zunächst mit einer kurzen Darstellung der Grundzüge des Krigings für lineare Modelle.

Die optimale Vorhersage für eine nicht beobachtete Position ergibt in der klassischen Geostatistik als gewichtete Summe der beobachteten Daten $\boldsymbol{y}(\boldsymbol{s})$

$$y(\boldsymbol{s}_0) = \sum_{i=1}^{n} w_i \, y(\boldsymbol{s}_i),$$

wobei die Gewichte w_i so zu wählen sind, daß der mittlere quadratische Vorhersagefehler (*Mean-Squared Prediction Error*, MSPE) minimal wird. Um dies in

der Praxis umzusetzen, sind bestimmte Annahmen bezüglich des Mittelwertes und der Kovarianz des räumlichen Prozesses zu treffen. Meist sind dies die eines stationären Prozesses 2. Ordnung mit

$$\begin{aligned} \mathrm{E}[\boldsymbol{y}(\boldsymbol{s})] &= \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\mu} \quad \text{und} \\ \mathrm{cov}[y(\boldsymbol{s}_i), y(\boldsymbol{s}_j)] &= \sigma_n^2 + \sigma^2 \cdot \rho(s_i - s_j), \end{aligned} \tag{2.60}$$

worin $\sigma_n^2 + \sigma^2$ die Varianz der Zufallsvariablen (Schwellenwert oder *Sill*), σ_n^2 die Residual- oder Nugget-Varianz und $\rho(s_i - s_j)$ die Korrelationsfunktion des Prozesses ist. Es ist dann

$$\begin{aligned} \mathrm{E}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_0 \end{bmatrix} &= \begin{bmatrix} \boldsymbol{X}\boldsymbol{\beta} \\ \boldsymbol{X}_0\boldsymbol{\beta} \end{bmatrix} \quad \text{und} \\ \mathrm{cov}\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}_0 \end{bmatrix} &= \begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yy_0} \\ \boldsymbol{\Sigma}_{y_0y} & \boldsymbol{\Sigma}_{y_0y_0} \end{bmatrix}, \end{aligned}$$

worin $\boldsymbol{y}$ die an den Positionen $\boldsymbol{s}$ beobachteten Daten und $\boldsymbol{y}_0$ die vorherzusagenden Werte an den Positionen $\boldsymbol{s}_0$ sind. Ferner ist $\boldsymbol{\Sigma}_{yy}$ die $n \times n$ die Kovarianzmatrix der beobachteten Daten, $\boldsymbol{\Sigma}_{y_0y_0}$ die $k \times k$ Kovarianzmatrix der Vorhersagen und $\boldsymbol{\Sigma}_{y_0y}$ die $k \times n$ Kovarianzmatrix zwischen Vorhersagen und beobachteten Daten.

Die beste lineare, verzerrungsfreie Vorhersage (*Best Linear Unbiased Predictor*, BLUP; Goldberger, 1962), in der Geostatistik als *Universeller-Kriging-Prädiktor* bekannt, ist dann

$$\hat{\boldsymbol{y}}_0 = \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{y} + (\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})\hat{\boldsymbol{\beta}}, \tag{2.61}$$

worin $\hat{\boldsymbol{\beta}}$ die Verallgemeinerte-Kleinste-Quadrate-Schätzung (GLS)

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{y}$$

ist. Der zugehörige Vorhersagefehler $\mathrm{E}[(\hat{\boldsymbol{y}}_0 - \boldsymbol{y}_0)^2]$ ist

$$\begin{aligned} \mathrm{MSPE}(\hat{\boldsymbol{y}}_0) &= \boldsymbol{\Sigma}_{y_0y_0} - \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yy_0} + (\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X}) \\ &\quad \times (\boldsymbol{X}'\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})'. \end{aligned} \tag{2.62}$$

Mehr zur Herleitung und den Eigenschaften dieses Prädiktors findet sich u. a. in Cressie (1991, Kap. 3.4) und Olea (1999, Kap. 6)

Da der Prädiktor aus (2.61) linear in $\hat{\boldsymbol{\beta}}$ ist, ist seine Anwendung für GLMs nicht direkt möglich. Im Ansatz von Gotway u. Stroup (1997) wird (2.61) daher zunächst umgestellt zu

$$\hat{\boldsymbol{y}}_0 = \boldsymbol{X}_0\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). \tag{2.63}$$

Gemäß der Formulierung des Erwartungswertes aus (2.60) läßt sich dies auch darstellen durch

$$\hat{\boldsymbol{y}}_0 = \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}). \tag{2.64}$$

Anstelle des Erwartungswertes von $\boldsymbol{y}$ im linearen Modell ($\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$) setzten Gotway u. Stroup (1997) hier den Erwartungswert von $\boldsymbol{y}$ im GLM ($\boldsymbol{\mu} = g^{-1}(\boldsymbol{X}\boldsymbol{\beta})$)

ein. Zur Schätzung von $\hat{\boldsymbol{\beta}}$ wurde von ihnen der GEE-Ansatz und der IWLS-Algorithmus aus (2.50) verwendet.

Die Matrix der mittleren quadratischen Vorhersagefehler für diesen Prädiktor geben Gotway u. Wolfinger (2003) (nach Vjipurkar u. Gotway, 2001) mit

$$\begin{aligned}\mathrm{MSPE}(\hat{\boldsymbol{y}}_0) &= \boldsymbol{\Sigma}_{y_0y_0} - \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yy_0} + (\boldsymbol{D}_0 - \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{D}) \\ &\quad \times(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X})^{-1}(\boldsymbol{D}_0 - \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{D})' \end{aligned} \tag{2.65}$$

an, wobei $\boldsymbol{D}$ die $n\times p$ Matrix der partiellen Ableitungen $\partial\boldsymbol{\mu}/\partial\boldsymbol{\beta} = \boldsymbol{\Delta}^{-1}\boldsymbol{X}$ für die beobachteten Positionen $\boldsymbol{s}$ ist und $\boldsymbol{D}_0$ entsprechend für die vorherzusagenden Positionen formuliert wird.

Um, basierend auf den Pseudo-Daten des PL-Ansatzes, Vorhersagen für ein GLMM zu erhalten, gehen Gotway u. Wolfinger (2003) wie folgt vor: Da die (marginalen) Momente der Pseudo-Daten aus (2.52) dieselbe Form besitzen wie die Daten gewöhnlicher linearer Modelle, läßt sich – einen stationären Prozeß 2. Ordnung unterstellend – für sie der Universelle-Kriging-Prädiktor aus (2.61) direkt nutzen. Der BLUP einer neuen Pseudo-Beobachtung ist dann (nach Umformung von (2.61) in (2.63))

$$\boldsymbol{\nu}_0 = \boldsymbol{X}_0\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}}), \tag{2.66}$$

wobei die Kovarianzmatrizen $\boldsymbol{\Sigma}_{\nu_0\nu}$ und $\boldsymbol{\Sigma}_{\nu\nu}$ analog zu denen in (2.61) definiert sind, jedoch für die Pseudo-Daten. Analog zu (2.62) ist die MSPE-Matrix dieser Vorhersagen

$$\begin{aligned}\mathrm{MSPE}(\hat{\boldsymbol{\nu}}_0) &= \boldsymbol{\Sigma}_{\nu_0\nu_0} - \boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Sigma}_{\nu\nu_0} + (\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{X}) \\ &\quad \times(\boldsymbol{X}'\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{X})^{-1}(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{X})'. \end{aligned} \tag{2.67}$$

Diese Matrix enthält die mittleren quadratischen Vorhersagefehler auf der Pseudo-Skala $\mathrm{E}[(\hat{\nu}_{0_i} - \nu_{0_i})^2]$ auf der Hauptdiagonalen. Um die Daten nun auf der Originalskala und nicht auf der Skala der Pseudo-Daten vorherzusagen, wird angenommen, daß $\boldsymbol{\nu}_0$ und die vorherzusagenden Beobachtungen $\boldsymbol{y}_0$ sowie ihre Prädiktoren $\hat{\boldsymbol{\nu}}_0$ bzw. $\hat{\boldsymbol{y}}_0$ ebenfalls Gleichung (2.52) erfüllen, so daß

$$\hat{\boldsymbol{\nu}}_0 = \hat{\boldsymbol{\eta}}_0 + \hat{\boldsymbol{\Delta}}_0(\hat{\boldsymbol{y}}_0 - \hat{\boldsymbol{\mu}}_0),$$

mit $\hat{\boldsymbol{\mu}}_0 = g^{-1}(\hat{\boldsymbol{\eta}}_0)$. Aufgelöst nach $\hat{\boldsymbol{y}}_0$ ergibt sich dann

$$\hat{\boldsymbol{y}}_0 = \hat{\boldsymbol{\mu}}_0 + \hat{\boldsymbol{\Delta}}_0^{-1}(\hat{\boldsymbol{\nu}}_0 - \hat{\boldsymbol{\eta}}_0). \tag{2.68}$$

Im Fall der hier ausschließlich besprochenen marginalen GLMMs ist dieser Prädiktor äquivalent zu dem von Gotway u. Stroup (1997) entwickelten des GEE-Ansatzes aus (2.64) (s. Anhang B.3). Der Vorhersagefehler dieses Prädiktors ist nach Gotway u. Wolfinger (2003) (s. auch Schabenberger u. Gotway, 2004, S. 369 f.)

$$\mathrm{MSPE}(\hat{\boldsymbol{y}}_0) = [\hat{\boldsymbol{\Delta}}_0^{-1}]^2 \cdot \mathrm{MSPE}(\hat{\boldsymbol{\nu}}_0). \tag{2.69}$$

Er ist äquivalent zu dem in (2.65) gezeigten (Gotway u. Wolfinger, 2003). Konfidenzintervalle der Vorhersagen lassen sich unter der Annahme einer approximativen Normalverteilung der Vorhersagefehler $(\hat{\boldsymbol{y}}_0 - \boldsymbol{y}_0)$ mit Mittelwert 0 und Varianz $\mathrm{MSPE}(\hat{\boldsymbol{y}}_0)$ berechnen durch

$$\hat{y}_0 \pm t_{1-\alpha/2,q} \cdot \sqrt{\mathrm{MSPE}(\hat{\boldsymbol{y}}_0)},$$

worin $t_{1-\alpha/2,q}$ das $100(1-\alpha)$%ige Perzentil einer t-Verteilung mit q Freiheitsgraden ist. q läßt sich mit Hilfe der Satterthwaite-Approximation bestimmen; näheres hierzu findet sich z. B. in SAS (2006, S. 126 f.).

Letztlich unterscheiden sich die beiden hier vorgestellten Modelle des GEE- und des marginalen PL-Ansatzes lediglich in der Art der Schätzung der Kovarianzparameter $\boldsymbol{\alpha}$. Während diese im GEE-Ansatz über eine gutachterliche oder automatisierte Anpassung einer Semivariogrammfunktion an das empirische Semivariogramm der (student-standardisierten) Modellresiduen erfolgt, wird im PL-Ansatz nicht das empirische Semivariogramm bestimmt, sondern $\hat{\boldsymbol{\alpha}}$ über die Maximierung der Likelihood-Funktion bestimmt. Letzteres bringt zweierlei Vorteile mit sich: Zum einen kann die Schätzung der Parameter in $\boldsymbol{\beta}$ und $\boldsymbol{\alpha}$ vollständig automatisiert erfolgen, ohne wiederholte gutachterliche oder extern automatisierte Schätzungen der Kovarianzparameter anhand der jeweils aktuellen Modellresiduen durchführen zu müssen. Zum anderen werden Aussagen über die Streuung der Kovarianzparameter ermöglicht (näheres hierzu s. SAS, 2006, S. 123 ff. und Wolfinger et al., 1994), die jedoch mit Vorsicht zu betrachten sind, da sie nur unter der Annahme der Normalverteilung der Pseudo-Daten gelten. Die Schätzung der Kovarianzparameter per ML kann sich jedoch auch nachteilig auswirken, da sie stets alle beobachteten Punkte mit einbezieht und damit ggf. der Kovarianz zwischen weit entfernten Punkten, die nur durch wenige Beobachtungen repräsentiert werden, zuviel Gewicht verleiht. Für gutachterliche Anpassungen einer Semivariogrammfunktion dagegen wird empfohlen, nur Beobachtungspaare einzubeziehen, deren Entfernung nicht über die Hälfte bis Zweidrittel der maximalen Entfernung hinausgeht (Journel u. Huijbregts, 1978, S. 194). Die ML-Methode kann also überangepaßte Modelle liefern; für ein Beispiel s. Littel et al. (1996, Kap. 9.7). In keinem der beiden Ansätze wird jedoch, wie auch in den klassischen Kriging-Verfahren, die Unsicherheit der Schätzung der Kovarianzparameter mit in die Schätzung der Vorhersagevarianz einbezogen.

2.3. Das Count-Modell

Sind die Methoden und Modelle des Linientransektverfahrens und der GLMs bisher gänzlich unabhängig voneinander betrachtet worden, so soll im folgenden eine Kombinationsmöglichkeit der Analyse von Linientransekten und GLMs – im wesentlichen basierend auf der Arbeit von Hedley (2000) (s. alternativ Buckland et al., 2004, Kap. 4; Hedley u. Buckland, 2004) – vorgestellt werden.

2.3.1. Modellformulierung

Ziel des von Hedley (2000) entwickelten Verfahrens war es, basierend auf herkömmlichen Linientransekterhebungen, die in Walzählungen zu beobachtende

räumliche Heterogenität geeignet in einem Modell zu berücksichtigen. Darüber hinaus bestand Bedarf an lokalen Vorhersagen der Walabundanz, um Gebiete hoher und niedriger Dichten auszuweisen und somit Entscheidungshilfen für das Management von Walpopulationen zu liefern und letztlich auch Aussagen über zugrundeliegende ökologische Prozesse formulieren zu können. Eine Stratifizierung der Analyse nach Gebieten hoher und niedriger Dichten war zwar möglich und wurde auch durchgeführt, führte aber in den Straten niedriger Dichten aufgrund entsprechend geringer Beobachtungsanzahlen zu sehr hohen Streuungen. Anstelle der Stratifizierung galt es daher ein Verfahren zu entwickeln, das den Einbezug geographischer Faktoren wie in diesem Fall der Entfernung zur Packeisgrenze ermöglicht, um so auch in Gebieten mit nur wenigen Beobachtungen verläßlichere Ergebnisse zu erreichen. Der herkömmliche Dichteschätzer des Linientransektverfahrens aus (2.4) sieht den Einbezug zusätzlicher räumlicher oder nicht-räumlicher Kovariaten jedoch nicht vor. Ebenso ermöglicht er keine lokalen Vorhersagen; Dichteschätzungen beziehen sich auf das gesamte Untersuchungsgebiet, oder bei einer stratifizierten Auswertung auf das gesamte Stratum.

Beide von Hedley verfolgten Ziele, lokale Vorhersagen und Einbezug zusätzlicher erklärender Variablen, sind auch zunehmend in anderen Projekten von Bedeutung. Nicht zuletzt wegen der stetig steigenden Verfügbarkeit geographischer Informationen und der rasanten Weiterentwicklung der Software-Produkte zur Verarbeitung und Analyse solcher Daten besteht immer öfter auch der Bedarf, diese in die Analyse herkömmlicher Linientransektaufnahmen einzubeziehen. Die Ausweisung signifikant auf die Verteilung der untersuchten biologischen Population einflußnehmender Größen läßt Rückschlüsse auf ökologische Zusammenhänge zu, lokale Vorhersagen der Dichte und anhand dieser erstellte Karten können ein nützliches Hilfsmittel zum Management der untersuchten Population sein.

Hedley (2000) hat gleich mehrere Verfahren entwickelt, die diesen neuen Anforderungen gerecht werden, ohne die Vorteile des Linientransektverfahrens aufzugeben. Von ihnen soll hier lediglich das sog. *Count-Modell* näher betrachtet werden. Es sieht die Einbindung zusätzlicher, z. B. in einem GIS gespeicherter Informationen in die konventionelle Linientransektanalyse vor und ermöglicht sowohl die Ausweisung signifikanter Einflußgrößen wie auch die räumliche Vorhersage von Abundanzen.

Zunächst wiederum (wie bei der Erläuterung des Linientransektverfahrens) von einer gewöhnlichen Streifenzählung mit $2 \cdot w$ breiten Streifen ausgehend, wird jeder der k erfaßten Streifen in T_i Segmente (etwa) gleicher Länge c unterteilt ($l_i = \sum_{j=1}^{T_i} c_j$ und $\sum_{i=1}^{k} l_i = L$); das Segment am Ende eines Transektes fällt entsprechend kürzer aus. Die Länge c ist dabei generell frei wählbar, sollte aber so gesetzt werden, daß die geographischen Eigenschaften innerhalb eines Segments annähernd konstant sind. Erst durch diese Diskretisierung der eigentlich kontinuierlich verlaufenden Transekte wird die kleinräumige Anbindung zusätzlicher (geographischer) Informationen an Transektabschnitte ermöglicht. So wird anschließend jedem einzelnen Segment die Anzahl der darin beobachteten Objekte n_i sowie zusätzlich die an der Position des Segments gültigen Werte erklärender (räumlicher) Kovariaten zugewiesen. Ein log-lineares Modell zur

Beschreibung der Objektanzahl eines Segments läßt sich dann durch

$$\mathrm{E}(n_i) = \exp\left[\ln(2\,c_i\,w) + \beta_0 + \sum_{j=1}^{p-1} x_{ij}\,\beta_j\right] \tag{2.70}$$

formulieren. Die Fläche des i-ten Segments $2 \cdot c_i \cdot w$ ist hierin als Offset-Term enthalten, so daß unterschiedliche Segmentgrößen berücksichtigt werden können. $\boldsymbol{x}_i$ ist ein Vektor (räumlicher) Kovariaten des i-ten Segments und β_j, $j = 0, 1, \ldots, p-1$ sind die zu schätzenden Modellparameter. Die Modellformulierung entspricht also der eines gewöhnlichen GLMs[8] mit Offset-Term und Poisson-verteilten n_i (vergl. Kap. 2.2.1.1). Die Objektdichte für ein Segment ist im Fall der Streifenzählung identisch zur Individuendichte (s. Kap. 2.1.1) und ergibt sich aus (2.70) einfach durch $D_i = D_{s_i} = n_i/(2\,c_i\,w)$.

Die Anpassung dieses Modells an Linientransekterhebungen läßt sich nun über die effektive Streifenbreite υ erreichen. Diese ist, wie in Kapitel 2.1.1 erläutert, definiert als die Entfernung beiderseits des Transektes, bis zu der so viele Objekte übersehen wurden, wie jenseits davon bis zur rechten Kappungsgrenze w_r noch entdeckt wurden. D. h. eine Linientransekterhebung läßt sich auch als Streifenzählung mit $2 \cdot \upsilon$ breiten Streifen betrachten. Entsprechend läßt sich (2.70) für Linientransektaufnahmen umformulieren zu

$$\mathrm{E}(n_i) = \exp\left[\ln(2\,c_i\,\upsilon) + \beta_0 + \sum_{j=1}^{p-1} x_{ij}\,\beta_j\right], \tag{2.71}$$

wobei zunächst vernachlässigt werden soll, daß es sich beim Offset-Term (genauer bei υ) um eine ebenfalls zu schätzende Größe handelt. Die Unterteilung der Transekte in Segmente erfolgt hier gleichermaßen wie oben beschrieben.

Sollen nicht die Objekt-, sondern die Individuenanzahlen modelliert werden, so kann n_i in (2.71) durch die Anzahl der im i-ten Segment beobachteten Individuen $m_i = \sum_{j=1}^{n_i} s_{ij}$ ersetzt werden, worin s_{ij} die Gruppengröße des j-ten in Segment i entdeckten Objektes ist. Ist die Anzahl der beobachteten Individuen je Segment deutlich größer als die Anzahl der beobachteten Objekte, so ist dies auch bei der Wahl der Fehlerverteilung zu berücksichtigen. Während für (2.71) ein evtl. überstreutes Poisson-Modell in Frage kommt, können bei direkter Modellierung der Individuenanzahl und großen Gruppen andere Varianzfunktionen wie die der Negativ-Binomial-Verteilung (s. Kap. 2.2.5) besser geeignet sein. Die direkte Modellierung der Individuenanzahlen setzt jedoch voraus, daß die Gruppengrößen s_{ij} fehlerfrei erfaßt wurden. Alternativ kann die mittlere Gruppengröße – evtl. geographisch stratifiziert oder ebenfalls in Abhängigkeit (räumlicher) Kovariaten – über eine verläßliche Methode geschätzt werden (s. Buckland et al., 2001, Kap. 3.5), um die Schätzung(en) anschließend mit denen aus (2.71) zu multiplizieren.

Die Formulierung in (2.71) impliziert eine konstante Entdeckungswahrscheinlichkeit entlang aller Transekte bzw. analog eine konstante effektive Streifenbreite. Hedley (2000, Kap. 4.1) hat die Ansätze daher zusätzlich dahingehend

[8] Hedley (2000) beschreibt darüber hinaus auch die Möglichkeit der Verwendung von verallgemeinerten additiven Modellen (GAMs; Hastie u. Tibshirani, 1990) anstelle von GLMs.

erweitert, segment- oder gar objektspezifische Entdeckungswahrscheinlichkeiten p_i bzw. p_{ij} zu berücksichtigen, auf die hier ebenfalls kurz eingegangen werden soll.

Die effektive Streifenbreite υ ergibt sich auch als Produkt der Kappungsentfernung w_r und der mittleren Entdeckungswahrscheinlichkeit P_a ($\upsilon = w_r\, P_a$). (2.71) läßt sich also auch darstellen als

$$\mathrm{E}(n_i) = \exp\left[\ln(2\, c_i\, w_r\, P_a) + \beta_0 + \sum_{j=1}^{p-1} x_{ij}\, \beta_j\right].$$

Aus dieser Form der Darstellung ist leicht die Möglichkeit des Einbezugs segmentspezifischer Entdeckungswahrscheinlichkeiten p_i zu ersehen.

$$\mathrm{E}(n_i) = \exp\left[\ln(2\, c_i\, w_r\, p_i) + \beta_0 + \sum_{j=1}^{p-1} x_{ij}\, \beta_j\right] \tag{2.72}$$

Äquivalent läßt sich dies auch durch

$$\mathrm{E}(\tilde{n}_i) = \exp\left[\ln(2\, c_i\, w_r) + \beta_0 + \sum_{j=1}^{p-1} x_{ij}\, \beta_j\right], \quad \tilde{n}_i = \frac{n_i}{p_i} \tag{2.73}$$

formulieren, wobei anstelle der zwischen 0 und w_r beobachteten Objektanzahl n_i die (geschätzte) Anzahl der vorhandenen Objekte $\tilde{n}_i$ als abhängige Variable verwendet wird. $\tilde{n}_i$ ergibt sich durch den Horvitz-Thompson-Schätzer (Horvitz u. Thompson, 1952; Thompson, 1992) $\tilde{n}_i = n_i/p_i$, ist diesem in der Regel jedoch nur ähnlich, da die Entdeckungswahrscheinlichkeit p_i meist durch ihren Schätzer $\hat{p}_i$ zu ersetzen ist. Wenngleich beide Gleichungen äquivalent sind, werden die Fehler beider Modelle evtl. unterschiedliche Verteilungen aufweisen. Während für (2.72) eine ggf. überstreute Poisson-Verteilung angenommen werden kann, dürfte die Überdispersion der $\tilde{n}_i$ in (2.73) wegen der Anpassung der beobachteten Anzahlen durch ihre Entdeckungswahrscheinlichkeit meist größer sein. Die Notwendigkeit der Darstellung in (2.73) ergibt sich jedoch, wenn objektspezifische Entdeckungswahrscheinlichkeiten zu berücksichtigen sind. Dies läßt sich in (2.72) nicht umsetzen, wohl aber in (2.73) mit $\tilde{n}_i = \sum_{j=1}^{n_i} 1/p_{ij}$.

Auch für den Fall der direkten Modellierung der Individuenanzahlen m_i lassen sich segment- und objektspezifische Entdeckungswahrscheinlichkeiten einbeziehen. Für die Erstgenannte wird n_i in (2.72) durch m_i ersetzt, für letztere $\tilde{n}_i$ in (2.73) durch $\tilde{m}_i = \sum_{j=1}^{n_i} s_{ij}/p_{ij}$. Auch hier ist zu bedenken, daß in beiden Fällen die fehlerfreie Erfassung der Gruppengrößen Voraussetzung für eine verläßliche Schätzung ist.

2.3.2. Vorhersagen

Um auf Grundlage der vorgestellten Modelle lokale Vorhersagen zu berechnen, empfiehlt Hedley (2000, S. 35), ein möglichst engmaschiges Gitter mit quadratischen Zellen der Größe f_0 über das Gebiet zu legen, für die Zellen die Kovariaten

zu bestimmen und mit Hilfe der geschätzten Parameter $\hat{\boldsymbol{\beta}}$ über eines der oben aufgeführten Modelle die zu erwartenden Objekt- ($\mathrm{E}[\boldsymbol{n}_0]$) oder Individuenanzahlen ($\mathrm{E}[\boldsymbol{m}_0]$) zu schätzen. Als Offset-Term wird dann die Zellgröße f_{0_i} verwendet. Der Schätzer der Objektdichte für Zelle i ergibt sich dann durch

$$\hat{D}_{s_{0_i}} = \hat{n}_{0_i}/f_{0_i},$$

der Schätzer der Individuendichte durch

$$\hat{D}_{0_i} = \hat{n}_{0_i}/f_{0_i} \cdot \hat{\mathrm{E}}(s_{0_i}) \quad \text{oder} \quad \hat{D}_{0_i} = \hat{m}_{0_i}/f_{0_i}.$$

Die Ausdrücke vereinfachen sich, wenn alle Zellen dieselbe Größe besitzen oder eine globale Schätzung der mittleren Gruppengröße $\mathrm{E}(s)$ verwendet wird.[9]

Wie aus dem Absatz zuvor bereits hervorgeht, wird im Ansatz von Hedley die zu vermutende räumliche Autokorrelation der Objekt- bzw. Individuenanzahlen nahe gelegener Segmente und Zellen in der Vorhersage neuer Beobachtungen nicht berücksichtigt. Dasselbe gilt für die Schätzung der festen Effekte des Modells, die ebenfalls unter der Annahme der Unabhängigkeit der Beobachtungen erfolgt. Während die Berücksichtigung der Autokorrelation auf die Parameterschätzungen selbst in der Regel nur geringen Einfluß nehmen wird, so gilt dies – in Abhängigkeit von der Stärke der räumlichen Abhängigkeit – umso weniger für Varianzschätzungen der Parameter und für neue vorherzusagende Beobachtungen. Anstelle der von Hedley verwendeten, auf Unabhängigkeit der Beobachtungen basierenden Score-Funktionen aus (2.22),

$$\boldsymbol{s}(\boldsymbol{\beta}) = \boldsymbol{D}'\boldsymbol{V}^{-1}(\boldsymbol{y}(\boldsymbol{s}) - \tilde{\boldsymbol{\mu}}(\boldsymbol{s})),$$

dürfte daher die Verwendung der GEEs aus (2.47),

$$\boldsymbol{s}(\boldsymbol{\beta}) = \boldsymbol{D}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}(\boldsymbol{s}) - \tilde{\boldsymbol{\mu}}(\boldsymbol{s})),$$

besser geeignet sein, wobei die abhängige Variable $\boldsymbol{y}(\boldsymbol{s})$ hier den Objekt- oder Individuenanzahlen entspricht, also $y_i(\boldsymbol{s}) = n_i$, $y_i(\boldsymbol{s}) = m_i$, $y_i(\boldsymbol{s}) = \tilde{n}_i$ oder $y_i(\boldsymbol{s}) = \tilde{m}_i$. Im Fall des hier besprochenen (überstreuten) Poisson-Modells ist

$$\boldsymbol{D} = \begin{bmatrix} x_{11}\,\tilde{\mu}_1 & \cdots & x_{1p}\,\tilde{\mu}_1 \\ \vdots & & \vdots \\ x_{n1}\,\tilde{\mu}_n & \cdots & x_{np}\,\tilde{\mu}_n \end{bmatrix},$$

$\boldsymbol{V} = \mathrm{diag}\left[\sigma_o^2\,\tilde{\mu}_i\right]$ und $\tilde{\boldsymbol{\mu}} = [t_1 \cdot \exp(\boldsymbol{x}_1'\boldsymbol{\beta}), \ldots, t_n \cdot \exp(\boldsymbol{x}_n'\boldsymbol{\beta})]$, worin t_i der Wert des Offset-Terms für die jeweilige Beobachtung/Vorhersage und σ_o^2 der Überdispersionsparameter ist. Die Kovarianzmatrix der GEEs läßt sich gemäß (2.49) durch

$$\boldsymbol{\Sigma}(\tilde{\boldsymbol{\mu}}, \boldsymbol{\alpha}) = \sigma_n^2\boldsymbol{V} + \sigma_o^2\boldsymbol{V}^{1/2}\boldsymbol{R}(\boldsymbol{\alpha})\boldsymbol{V}^{1/2}$$

oder einer vereinfachten Form hiervon definieren. Hierin ist σ_n^2 ein Nugget-Effekt und $\boldsymbol{R}(\boldsymbol{\alpha})$ eine Korrelationsmatrix mit distanzabhängigen Elementen $\rho(s_i - s_j; \boldsymbol{\alpha})$.

[9] Wie es zu unterschiedlich großen Vorhersagezellen kommen kann, wird in Kapitel 3.2.1 erläutert.

Die Auswertung der Modelle kann dann mit Hilfe der in Kapitel 2.2.7 beschrieben GEE- oder PL-Ansätze erfolgen. Für die Vorhersage neuer Beobachtungen kann der Ansatz von Hedley, ein Gitter über das Untersuchungsgebiet zu legen, übernommen werden, die Vorhersage für eine Zelle erfolgt dann jedoch nicht über eines der Modelle (2.71), (2.72) oder (2.73), sondern über einen der Prädiktoren aus (2.64),

$$\hat{y}_0 = \hat{\tilde{\mu}}(\boldsymbol{s}_0) + \boldsymbol{\Sigma}_{y_0 y}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y}(\boldsymbol{s}) - \hat{\tilde{\boldsymbol{\mu}}}(\boldsymbol{s})),$$

oder (2.68),

$$\hat{y}_0 = \hat{\tilde{\mu}}(\boldsymbol{s}_0) + \hat{\boldsymbol{\Delta}}_0^{-1}(\hat{\boldsymbol{\nu}}(\boldsymbol{s}_0) - \hat{\boldsymbol{\eta}}(\boldsymbol{s}_0)),$$

wobei $\hat{\tilde{\mu}}(\boldsymbol{s}_0)$ den geschätzten Erwartungswerten der zellweisen Objekt- $\hat{n}_0$ bzw. Individuenanzahl $\hat{m}_0$ entspricht, hier aber basierend auf den unter Einbezug räumlicher Autokorrelation geschätzten Parametern $\hat{\boldsymbol{\beta}}$ bestimmt wird.

Schätzer der Objekt- ($\hat{N}$) und Individuenabundanz ($\hat{M}$) für das gesamte Untersuchungsgebiet ergeben sich durch Aufsummieren der zellweisen Vorhersagen $\hat{n}_{0_i}$ oder $\hat{m}_{0_i}$. Die Gesamtobjekt- ($\hat{D}_s$) oder -individuendichte ($\hat{D}$) schließlich kann durch

$$\begin{aligned} \hat{D}_s &= \hat{N}/A \quad \text{bzw.} \\ \hat{D} &= \hat{M}/A \end{aligned}$$

geschätzt werden, mit $A = \sum_{i=1}^{n_Z} f_{0_i}$, wobei A die Größe der Untersuchungsgebietes und n_Z die Anzahl der Zellen innerhalb des Untersuchungsgebietes ist. Sind alle Zellen des Vorhersagegitters gleich groß, so vereinfacht sich $A = \sum_{i=1}^{n_Z} f_{0_i}$ zu $n_Z \cdot f_0$.

2.3.3. Varianzschätzung

Ein möglicher Ansatz zur Berechnung eines Präzisionsmaßes für die Gesamtabundanz der Objekte ($\hat{N} = \sum_{i=1}^{n_Z} n_{0_i}$) oder Individuen ($\hat{M} = \sum_{i=1}^{n_Z} m_{0_i}$) orientiert sich am sog. (universellen) Block-Kriging. Es handelt sich hierbei um ein Vorhersageverfahren, in dem das Ziel weniger in der Vorhersage der Zielgröße an bestimmten Punkten $\boldsymbol{s}_0$ als vielmehr in der Vorhersage des Mittelwertes für eine Fläche B, den Block, liegt.[10] Das Verfahren gleicht in weiten Teilen dem des bereits erläuterten Universellen Krigings.[11] Da im Rahmen dieser Arbeit das Interesse nicht am Block-Kriging-Prädiktor selbst, sondern lediglich an dessen MSPE bestand, beschränkt sich die nachfolgende Erläuterung auf diesen Teil. Für detailliertere Betrachtungen des Block-Krigings s. z. B. Cressie (1991, Kap. 3.4), Goovaerts (1997, Kap. 5.5), Olea (1999, Kap. 12) und Journel u. Huijbregts (1978, Kap. 5).

[10] Genau genommen handelt es sich bereits bei $\hat{n}_0$ bzw. $\hat{m}_0$ um Vorhersagen für Flächen, nämlich die Gitterzellen, und nicht für einzelne Punkte. Der Flächenbezug wurde in den bisher erläuterten Ansätzen jedoch nicht im Zuge des Krigings, sondern in Form des Offset-Terms des GLMs hergestellt. Die eigentliche Vorhersage bezog sich jedoch stets auf einen Punkt, den Zellschwerpunkt, dessen Repräsentativität für die gesamte Zellfläche unterstellt wurde.

[11] Genauer: Das Punkt-Kriging ist ein Sonderfall des Block-Krigings.

Unter der Annahme der Stationarität zweiter Ordnung ist der (universelle) Block-Kriging-MSPE einer mittleren Vorhersage $\hat{\bar{z}}_0$ gegeben durch

$$\begin{aligned}\mathrm{MSPE}(\hat{\bar{y}}_0) &= \bar{\boldsymbol{\Sigma}}_{BB} - \bar{\boldsymbol{\Sigma}}_{By}\boldsymbol{\Sigma}_{yy}^{-1}\bar{\boldsymbol{\Sigma}}_{yB} + (\bar{\boldsymbol{x}}_B - \bar{\boldsymbol{\Sigma}}_{yB}'\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X}) \\ &\quad \times \mathrm{cov}(\hat{\boldsymbol{\beta}})(\bar{\boldsymbol{x}}_B - \bar{\boldsymbol{\Sigma}}_{yB}'\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{X})'. \end{aligned} \tag{2.74}$$

und ähnelt den Darstellungen aus (2.65) und (2.67). Es ist dabei $\bar{\boldsymbol{\Sigma}}_{yB}$ der Punkt-zu-Block-Kovarianzvektor, $\bar{\boldsymbol{\Sigma}}_{BB}$ die Block-zu-Block-Kovarianz, $\boldsymbol{\Sigma}_{yy}$ die Matrix der Kovarianzen zwischen den beobachteten Punkten, $\boldsymbol{X}$ die Designmatrix der beobachteten Punkte und $\bar{\boldsymbol{x}}_B$ der Designvektor des Blocks. In der Praxis werden $\bar{\boldsymbol{\Sigma}}_{yB}$, $\bar{\boldsymbol{\Sigma}}_{BB}$ und $\bar{\boldsymbol{x}}_B$ zumeist approximiert, indem endlich viele Diskretisierungspunkte in den Block gelegt und über diese mittlere Kovarianzen und Kovariaten berechnet werden. So ergibt sich für die Block-zu-Block-Kovarianz

$$\bar{\boldsymbol{\Sigma}}_{BB} = \frac{1}{n_B^2}\sum_{i=1}^{n_B}\sum_{j=1}^{n_B}\mathrm{cov}(\boldsymbol{s}_{Bi}, \boldsymbol{s}_{Bj}).$$

Das j-te Element des $n \times 1$ Punkt-zu-Block-Kovarianzvektors berechnet sich durch

$$\left(\bar{\boldsymbol{\Sigma}}_{yB}\right)_j = \frac{1}{n_B}\sum_{i=1}^{n_B}\mathrm{cov}(\boldsymbol{s}_{Bi}, \boldsymbol{s}_j).$$

Es handelt sich bei $\bar{\boldsymbol{\Sigma}}_{BB}$ also um die mittlere Kovarianz zwischen allen n_B Diskretisierungspunkten, bei $\left(\bar{\boldsymbol{\Sigma}}_{yB}\right)_j$ um die mittlere Kovarianz zwischen allen Diskretisierungspunkten und dem j-ten beobachteten Punkt. Hierbei ist n die Anzahl der beobachteten Punkte, $\boldsymbol{s}_{Bi}$ die Position eines Diskretisierungspunktes und $\boldsymbol{s}_j$ die eines beobachteten Punktes.

Das j-te Element des mittleren $1 \times p$ Kovariatenvektors wird durch

$$(\bar{x}_B)_j = \frac{1}{n_B}\sum_{i=1}^{n_B} x_j(\boldsymbol{s}_{Bi})$$

approximiert.

Beziehen sich die erfaßten Größen nicht tatsächlich auf Punkte, sondern auf Flächen und unterscheiden sich die Diskretisierungsflächen des Blocks in ihrer Größe, so lassen sich der Block-Kriging-Prädiktor und dessen MSPE berechnen, indem die Kovarianzen und Kovariaten flächengewichtet gemittelt werden (Goovaerts, 1997, S. 156 f.). Es ergibt sich:

$$\begin{aligned}\bar{\boldsymbol{\Sigma}}_{BB} &= \frac{1}{B^2}\sum_{i=1}^{n_B}\sum_{j=1}^{n_B} b_i\, b_j \cdot \mathrm{cov}(\boldsymbol{s}_{Bi}, \boldsymbol{s}_{Bj}), \\ \left(\bar{\boldsymbol{\Sigma}}_{yB}\right)_j &= \frac{1}{B}\sum_{i=1}^{n_B} b_i \cdot \mathrm{cov}(\boldsymbol{s}_{Bi}, \boldsymbol{s}_j) \quad \text{und} \\ (\bar{x}_B)_j &= \frac{1}{B}\sum_{i=1}^{n_B} b_i\, x_j(\boldsymbol{s}_{B_i}),\end{aligned}$$

wobei $B = \sum_{i=1}^{n_B} b_i$.

Auf das Count-Modell mit räumlich autokorrelierten Beobachtungen läßt sich das beschriebene Vorgehen nun wie folgt anwenden: Das gesamte Untersuchungsgebiet entspricht dem Block, für den der MSPE zu berechnen ist, die n_Z Zellen entsprechen den n_B Diskretisierungsflächen des Blocks, die Flächen der Vorhersagezellen f_{0_i} den Gewichtungsfaktoren b_i.

Wie oben erwähnt, handelt es sich beim Block-Kriging-Prädiktor um die Vorhersage des Mittelwertes der untersuchten Größe innerhalb des Blocks. In der Anwendung auf das Count-Modell ist dies die Vorhersage der mittleren Objekt- oder Individuenanzahl bezogen auf eine mittlere Zellfläche $\bar{f}_0 = 1/n_Z \sum_{i=1}^{n_Z} f_{0_i}$. Um eine Vorhersage der Gesamtabundanz und einen Schätzer ihrer Varianz zu erhalten, ist die Vorhersage daher mit n_Z, der MSPE aus (2.74) mit n_Z^2 zu multiplizieren.

3. Anwendungssoftware (Teilprojekt 3)

Die nachfolgenden Seiten erläutern die Ergebnisse des dritten Teilprojektes, also die zur Umsetzung der im vorhergehenden Kapitel beschriebenen Methoden entwickelten Software-Module. Es beschreibt insofern eher technische Aspekte, die für das Verständnis der nachfolgenden Kapitel nicht unbedingt erforderlich sind.

Für die meisten der in Kapitel 2.1 und 2.2 beschriebenen Analysemethoden existiert bereits mehr oder minder umfassende Software. U. a. für die Auswertung von Linientransekterhebungen stellt das frei erhältliche Programm Distance 4.1[1] (Thomas et al., 2003), angebunden an die Standardreferenzen des Distance Sampling (Buckland et al., 2001, 2004), zahlreiche Möglichkeiten zur Verfügung. Die Software wird an der *Research Unit for Wildlife Population Assessment* (RUWPA) der Universität St. Andrews, Schottland speziell für die Auswertung von Distance-Sampling-Verfahren entwickelt. Neben den meisten der in dieser Arbeit benötigten und in Kapitel 2.1 erläuterten Methoden umfaßt das Programm etliche weitere Funktionen wie z. B. die Planung des Versuchsdesigns[2] und den Einbezug von Kovariaten in die Schätzung der Entdeckungsfunktion (sog. *Multicovariate Distance Sampling*, MCDS). Sie stellt jedoch nicht alle der in Kapitel 2.1.5 vorgestellten zusätzlichen Analyseoptionen zur Verfügung, so daß hierfür neue Module zu entwickeln waren. Auch die Anwendung des Count-Modells von Hedley (2000) gehört nicht zum Funktionsumfang. Ihre Implementierung ist zwar für die nächste Programmversion geplant, wird aber noch starken Einschränkungen unterliegen, da mit ihr umfangreiche Arbeitsschritte verbunden sind, die außerhalb geographischer Informationssysteme nur schwer umzusetzen sind.

Auch zur Auswertung unabhängiger Beobachtungen nicht normalverteilter Daten existieren bereits seit geraumer Zeit umfangreiche Analysewerkzeuge in zahlreichen freien und kommerziellen statistischen Programmen. Dies gilt jedoch nur eingeschränkt für den Fall korrelierter und insbesondere räumlich autokorrelierter Daten. So ist für die hier verwendete Analysesoftware SAS 9.1.3 erst seit Ende 2005 eine Prozedur verfügbar, die über den entsprechenden Funktionsumfang verfügt.

Insbesondere für die Umsetzung des Count-Modells wurde darüber hinaus eine GIS-Software benötigt, mit der die zahlreichen Geodaten erstellt, verwaltet, manipuliert und räumlich miteinander in Bezug gesetzt werden konnten.

[1] http://www.ruwpa.st-and.ac.uk/distance/

[2] Die in Abbildung 5.1 dargestellten Buschtransekte sind mit Hilfe dieser Funktion generiert worden.

Die Wahl fiel hier auf das Programmpaket ArcGIS 9.1[3]. Die Software des im Bereich der geographischen Informationssysteme weltweit führenden Herstellers ESRI bot (in der ArcInfo-Lizenzierung) bereits fast die gesamte in diesem Projekt benötigte Funktionalität. Zudem bietet es ab der Version 9.0 die Möglichkeit, mit Hilfe sog. *Geoprocessing-Skripte* (ESRI, 2004, 2005) Arbeitsabläufe zu automatisieren und eigene Module zur Umsetzung nicht enthaltener Funktionen zu entwickeln. Als Skriptsprache wurde Python[4] verwendet.

Es zeigte sich, daß, nicht zuletzt wegen der notwendigen Interaktion zwischen den drei beteiligten Software-Produkten, sehr viele und damit insgesamt langwierige und fehleranfällige Arbeitsschritte zu vollführen waren, um die Rohdaten für die eigentliche Analyse vorzubereiten. Zusätzliche Prämissen bei der Entwicklung neuer Software-Module waren daher die Zusammenfassung von Arbeitsschritten in klar definierte Blöcke, die einen bestimmten Aufgabenbereich umfassen sowie die zentrale Steuerung aller erforderlichen Programmaufrufe aus einem der beteiligten Programme heraus. Für die zentrale Steuerung bot sich SAS an, da es von den drei Programmen über die mit Abstand umfangreichsten Möglichkeiten des Datenim- und -exports zum Austausch von Informationen zwischen den Programmen verfügt sowie über umfangreiche Optionen zur Interaktion mit dem Betriebssystem, die die Steuerung der beiden übrigen Programme ermöglichen. Mit einer eigenen Makrosprache verfügt SAS darüber hinaus über die Möglichkeit der Automatisierung von Arbeitsabläufen und insbesondere der Entwicklung eigener Prozeduren bis hin zu komplexen statistischen Analysewerkzeugen. Es kam daher für die Entwicklung aller nicht anderweitig zur Verfügung gestellter Modelle zum Einsatz. Genutzt wurden dabei die SAS-Module BASE, STAT, GRAPH und IML. Das Datenmodell von SAS legt Daten zwar in Form von Tabellen in voneinander unabhängigen Dateien ab, stellt aber u. a. in Form von SAS-eigenen Programmierelementen (sog. `DATA Steps`) und einer in weiten Teilen mit der Standard-SQL-Syntax identischen SQL-Syntax (`PROC SQL`) umfangreiche Möglichkeiten zur Verfügung, um z. B. Tabellen miteinander in Beziehung zu setzen – ähnlich dem Datenmodell relationaler Datenbanken –, die in einer Tabelle enthaltenen Informationen zu filtern, zu bearbeiten und neue Daten (in Form neuer Spalten/Variablen) hinzuzufügen.

Die Zusammenfassung von Arbeitsschritten und ebenso die Entwicklung zusätzlicher, neuer Funktionalität erfolgte mit Hilfe von SAS-Makros und in Python programmierten Geoprocessing-Skripte. Insgesamt wurden 27 Makros und 11 Skripte programmiert, von den in den folgenden Abschnitten die wichtigsten kurz beschrieben werden. Nicht erläutert werden Makros und Skripte, die ausschließlich projektspezifische Aufgaben übernehmen und nur schwer eine Anwendung in anderen Projekten zulassen sowie solche, die kleine wiederkehrende Aufgaben durchführen. Hierzu gehört z. B. der Import der namibischen Beobachtungsdaten aus der verwendeten Datenbankanwendung und die Verknüpfung mit den per GPS erfaßten Koordinaten der Beobachtungspositionen sowie das Löschen, Kopieren oder Konvertieren von Dateien.

Wenngleich die nachfolgend beschriebenen Makros und Skripte speziell an-

[3] http://esri.com/software/arcgis/index.html
[4] http://www.python.org/

hand und zur Auswertung der namibischen Wildzählungsdaten, die in Kapitel 5 beschrieben werden, entwickelt wurden, so sind sie dabei gezielt so gehalten worden, daß sie sich vollständig und ohne großen Aufwand aus diesem Kontext heraustrennen und auch in anderen Projekten, die die räumliche und in weiten Teilen auch nicht-räumliche Analyse von Linientransekten zum Ziel haben, einsetzen lassen. Die Erläuterung der Makros und Skripte erfolgt daher allgemein und ohne konkreten Bezug zu den Daten aus Kapitel 5 und kann als Einführung[5] in die Anwendung der vorgestellten Makros und Skripte oder zumindest als Ideengeber für die technischen Aspekte räumlicher Linientransektprojekte betrachtet werden. Wie und an welcher Stelle sie für die namibischen Daten zum Einsatz kamen, wird in Form entsprechender Verweise in Kapitel 6 aufgeführt.

Nomenklatur: In SAS erstellte Makros werden durch ein führendes %-Zeichen gekennzeichnet (z. B. `%CalcPerpDist`), in Python programmierte Geoprocessing-Skripte durch die Endung „.py“ (z. B. `CreatePoints.py`). Bei den übrigen hervorgehobenen Bezeichnungen (z. B. `MakeXYEventLayer`) handelt es sich um sog. *Geoprocessing Tools*. Dies sind ArcGIS-eigene Datenmanagement- oder Analysefunktionen, auf die in den Geoprocessing-Skripten zugegriffen wird.

Es lag bis zu dieser Stelle mehrfach die Betonung auf der Automatisierung der entwickelten Module und der damit verbundenen Arbeitsabläufe. Es sei jedoch ausdrücklich darauf hingewiesen, daß das Ziel hiervon keineswegs ein vollständig automatisierter Programmablauf von der Berechnung der kürzesten Entfernungen bis zur Vorhersage lokaler Dichten war – wenngleich dies möglich wäre –; vielmehr sollten die einzelnen Module Funktionalitäten bereitstellen, die so bisher nicht oder nur verbunden mit erheblich größerem Aufwand zur Verfügung standen. Die Automatisierung ersetzt jedoch nicht die sachkundige Entscheidung über zu verwendende Modelle (z. B. Entdeckungsfunktion oder räumliche Kovarianzstruktur) und Parameter (z. B. Segmentlänge).

Neben den genannten Software-Produkten Distance, ArcGIS und SAS kamen im Rahmen dieses Projektes MS Access zur Eingabe und Speicherung der im Zuge der Wildzählungen erfaßten Beobachtungsdaten sowie G7ToWin[6] zur Speicherung der GPS-Informationen zum Einsatz. Als Geodatenformat für alle permanent zu erstellenden Geometrien wurde das Format der sog. *Personal Geodatabase Feature Class*, für temporäre Geodaten das gängige *Shapefile*-Format verwendet.

[5] Eine detailliertere Beschreibung der Makros und Skripte findet sich in diesen selbst, die unter `http://www.uni-forst.gwdg.de/~fmader/promotion/smalts/` erhältlich sind.
[6] Frei erhältlich unter `http://www.gpsinformation.org/ronh/`.

ArcGIS-Terminologie: In den nachfolgenden Erläuterungen werden einige ArcGIS-spezifische Begriffe verwendet, die hier kurz eingeführt werden sollen. In *Attributtabellen* werden Sachinformationen gespeichert, die ein Objekt der realen Welt, das im GIS durch eine Geometrie repräsentiert wird, näher beschreiben (z. B. Höhe, Baumart, Durchmesser für Bäume repräsentierende Punktgeometrien). Zwischen Geometrien und den Datensätzen der zugehörigen Attributtabelle besteht eine 1:1-Beziehung, d. h. für jede Geometrie findet sich genau ein Eintrag (ein Zeile/ein Datensatz) in der zugehörigen Attributtabelle und umgekehrt. Als sog. *Feature Class* wird in ArcGIS eine Einheit zur Speicherung von Geometrien desselben Typs (im wesentlichen Punkte, Linien oder Polygone) und der zugehörigen Attribute (in der Attributtabelle) bezeichnet. Zumeist umfaßt eine Feature Class thematisch zusammengehörige Objekte (Wasserstellen, Grundstücke, Straßen usw.). Eine *Personal Geodatabase* (pGDB) ist ein neueres Format (in Form einer MS Access Datenbank) zur Speicherung von Feature Classes. Als sog. *M-Wert* wird die (auf der Linie) gemessene reale Entfernung einer Position auf einer Linie zum Startpunkt der Linie bezeichnet. *Routengeometrien* sind spezielle Liniengeometrien, die die Verwendung von M-Werten ermöglichen.

3.1. Konventionelle Linientransektanalyse

Dieser Abschnitt umfaßt im wesentlichen den software-technischen Part des ersten Teilprojektes, der Validierung der Farmwegzählungen. Mit Hilfe von SAS-Makros und Python-Skripten waren hierbei – allgemein formuliert – folgende Arbeitsschritte umzusetzen:

1. Berechnung der kürzesten Entfernungen zwischen Objekten und zugehörigen Transekten. Zuvor ggf. Erstellung von Feature Classes, die die Beobachtungs- und Objektpositionen repräsentieren.

2. Durchführung einer herkömmlichen Linientransektanalyse (in Distance), ggf. unter Berücksichtigung abweichender realer Stichprobenflächen und externer Schätzungen der mittleren Gruppengröße.

3. Vergleich von Dichteschätzungen mit Hilfe des T-Tests aus Kapitel 2.1.5.5.

Die einzelnen Punkte zielen also nicht explizit auf die Überprüfung der Verläßlichkeit von auf Wegtransekten basierenden Schätzungen ab, sondern ganz allgemein auf den Vergleich von Dichteschätzungen und die hierfür zu schaffenden Grundlagen.

Berechnung der kürzesten Entfernungen zwischen Objekten und zugehörigen Transekten Basierend auf einer SAS-Tabelle, die Beobachtungsinformationen (Winkel, Entfernung, Art, Gruppengröße usw.) und die (geographischen oder projizierten) Koordinaten der Beobachtungspositionen enthält und einem

Geodatensatz, der die Transekte in Form von Routengeometrien (s. ArcGIS-Terminologie oben) repräsentiert, bestimmt das Makro `%CalcPerpDist` die kürzeste Entfernung zwischen Objektpositionen und beliebig geformten Transekten wie in Kapitel 2.1.5.1 beschrieben. Die Makro ruft nacheinander die Skripte `CreatePoints.py`, `CalcObjPos.py` und `CalcPerpDist.py` auf. Anhand der exportierten Beobachtungstabelle werden in `CreatePoints.py` zunächst Punktgeometrien erstellt (`MakeFeatureLayer`, `FeatureClassToFeatureClass`), welche die Beobachtungspositionen repräsentieren. Optional lassen sich die Koordinaten hierbei projizieren oder in eine andere Projektion (Räumliches Referenzsystem) überführen (`Project`). `CalcObjPos.py` berechnet anhand der ebenfalls in der Beobachtungstabelle gespeicherten Winkel und Entfernungen sowie einer optional anzugebenden lokalen Deklination (Nadelmißweisung) die Objektpositionen (s. Kap. 2.1.5.1 und Anh. C.1) und erstellt einen entsprechenden Geodatensatz. Anschließend werden, wiederum optional, die Beobachtungen gelöscht, die außerhalb des Untersuchungsgebietes oder vor dem Startpunkt bzw. jenseits des Endpunktes des zugehörigen Transektes liegen. Hierfür ist ggf. die Angabe einer zusätzlichen Polygon Feature Class erforderlich, die das Untersuchungsgebiet repräsentiert. Diese Funktion ist nützlich, wenn während der Aufnahme nicht sicher bestimmt werden kann, ob eine Beobachtung noch innerhalb des Untersuchungsgebietes bzw. der zu beobachtenden Fläche liegt. Im Anschluß daran bestimmt `CalcPerpDist.py` die kürzeste Entfernung zwischen jedem Objekt und dem Transekt, von dem aus die Beobachtung gemacht wurde (`LocateFeaturesAlongRoutes`). Zusätzlich werden für jede Beobachtung sog. M-Werte gespeichert. Dabei handelt es sich um Entfernungsmessungen vom Startpunkt des Transektes bis zu dem Punkt, an dem die kürzeste Entfernung zum Transekt bestimmt wurde (Transektposition). Diese Information ist im Zuge einer nachfolgenden räumlichen Analyse für die Zuordnung einer Beobachtung zu einem Segment erforderlich. Liegen bereits Geometrien der Beobachtungs- oder Objektpositionen vor, so kann alternativ auch auf die Erstellung dieser Feature Classes verzichtet und direkt die Berechnung der kürzesten Entfernungen durchgeführt werden. Wurden ausschließlich gerade Transekte aufgenommen, ist keine weiterführende räumliche Analyse der Daten geplant und enthält die Beobachtungstabelle Angaben des Winkels φ anstelle von ω (vergl. Abb. 2.3), so berechnet `%CalcPerpDist` die senkrechte Entfernung durch $x = y \cdot \sin(\varphi)$ und gibt diese in die Ergebnistabelle aus. GIS-Funktionalitäten werden in diesem Fall also nicht benötigt. Das Ergebnis des `%CalcPerpDist`-Makros ist eine SAS-Tabelle, die neben den ursprünglichen Beobachtungsdaten für jedes Objekt die kürzeste Entfernung zum zugehörigen Transekt und ggf. den M-Wert der zugehörigen Transektposition enthält.

Durchführung der herkömmlichen Linientransektanalyse Für eine herkömmliche Analyse von Linientransektstichproben bietet sich in erster Linie die oben erwähnte Software Distance an. Um dem Ziel der zentralen Steuerung aller benötigten Arbeitsabläufe aus einem Programm heraus gerecht zu werden, wurde jedoch nicht die graphische Oberfläche (GUI) von Distance sondern dessen Batch-Betrieb (s. hierzu Thomas et al., 2003, S. 80 f. u. Laake et al., 1996) genutzt,

um aus SAS heraus die Analysen starten zu können. Hierfür wurden mehrere Makros entwickelt, die von Distance in diesem Fall benötigte Konfigurations- und Datendateien schreiben und den Import der Ergebnisse aus Distance sowie deren Darstellung in SAS übernehmen. Eine Übersicht über den Ablauf der beteiligten Makros zeigt Abbildung 3.1. `%DistAnaly` ist das einzige vom Nutzer aufzurufende Makro, ihm werden alle benötigten Parameter, wie die zu verwendenden SAS-Tabellen, zu erstellende Ausgabedateien und eine Vielzahl an Auswertungsoptionen übergeben, auf die die nachfolgenden Makros zugreifen. Mit über 50 Optionen kann Einfluß auf die auszuwertenden Daten und die Analyse genommen werden. Als in der Regel bedeutsamste Optionen sind hier die zu verwendende Beobachtungs-, Transekt- und ggf. Stratumstabelle, das für die Schätzung der Entdeckungsfunktion zu verwendende Modell sowie linke und rechte Kappungsentfernung zu nennen. Als Beobachtungstabelle kann z. B. direkt die von `%CalcPerpDist` erstellte Tabelle verwendet werden. Alle von Distance berechneten Schätzungen und Kennzahlen werden in eine SAS-Tabelle importiert und stehen damit im Anschluß direkt zur Verfügung. Hierzu gehören insbesondere die Objekt- und Individuendichteschätzung, die Individuenabundanzschätzung für das gesamte Untersuchungsgebiet und die für die spätere Erstellung der Segmente benötigte Schätzung der effektiven Streifenbreite v sowie die Variationskoeffizienten und Konfidenzintervalle aller geschätzten Parameter.

Die SAS-Makros übernehmen hier ausschließlich Aufgaben des Datenmanagements und der Steuerung der Analyse, die eigentliche Auswertung dagegen findet komplett in Distance statt. Auch stellen sie nicht die gesamte von Distance angebotene Funktionalität zu Verfügung; insbesondere die Analyse von Punkttransekten und das sog. *Multicovariate Distance Sampling* (MCDS) sind nicht möglich, lassen sich bei Bedarf aber relativ leicht implementieren. Die Makros ermöglichen jedoch, wie gewünscht, den Aufruf von Distance aus SAS heraus und insbesondere die anschließende direkte Verwendung der Distance-Ergebnisse in SAS. Des weiteren lassen sich so weitergehende, nicht in Distance verfügbare Analysen in SAS umsetzen, dabei aber der von Distance angebotenen Funktionsumfang nutzen. Eine Möglichkeit ist beispielsweise, Distance-Analysen innerhalb von SAS-Makros in Schleifen zu implementieren, um z. B. im Rahmen einer Bootstrap-Analyse ein bestimmtes Modell an sich ändernde Daten anzupassen und die Ergebnisse der einzelnen Schätzungen in SAS weiterzuverarbeiten. Ein weiteres Beispiel stellt das Makro `%PoolPdfTest` dar. Es führt die in Kapitel 2.1.5.3 aufgeführten Tests auf Gleichheit der Entdeckungsfunktionen verschiedener Straten durch. Das Makro ruft hierfür `%DistAnaly` zunächst für den gesamten, alle Straten umfassenden Datensatz und dann für die Daten der einzelnen Straten auf. Die in den Modellanpassungen erzielten AIC-Werte und Maxima der log-Likelihood-Funktion werden gespeichert und am Ende mit der berechneten AIC-Summe, der χ^2-Teststatistik und dem zugehörigen p-Wert ausgegeben. Entsprechend der Ergebnisse kann dann die eigentliche Auswertung der Daten durchgeführt bzw. abgeändert werden.

Die ebenfalls im Zuge wiederholter Aufnahmen permanent angelegter Transekte in Kapitel 2.1.5.3 erläuterte abweichende Varianzschätzung läßt sich direkt in Distance umsetzen.

`%SmplFrac` ermöglicht die in Kapitel 2.1.5.4 beschriebene Berechnung der rea-

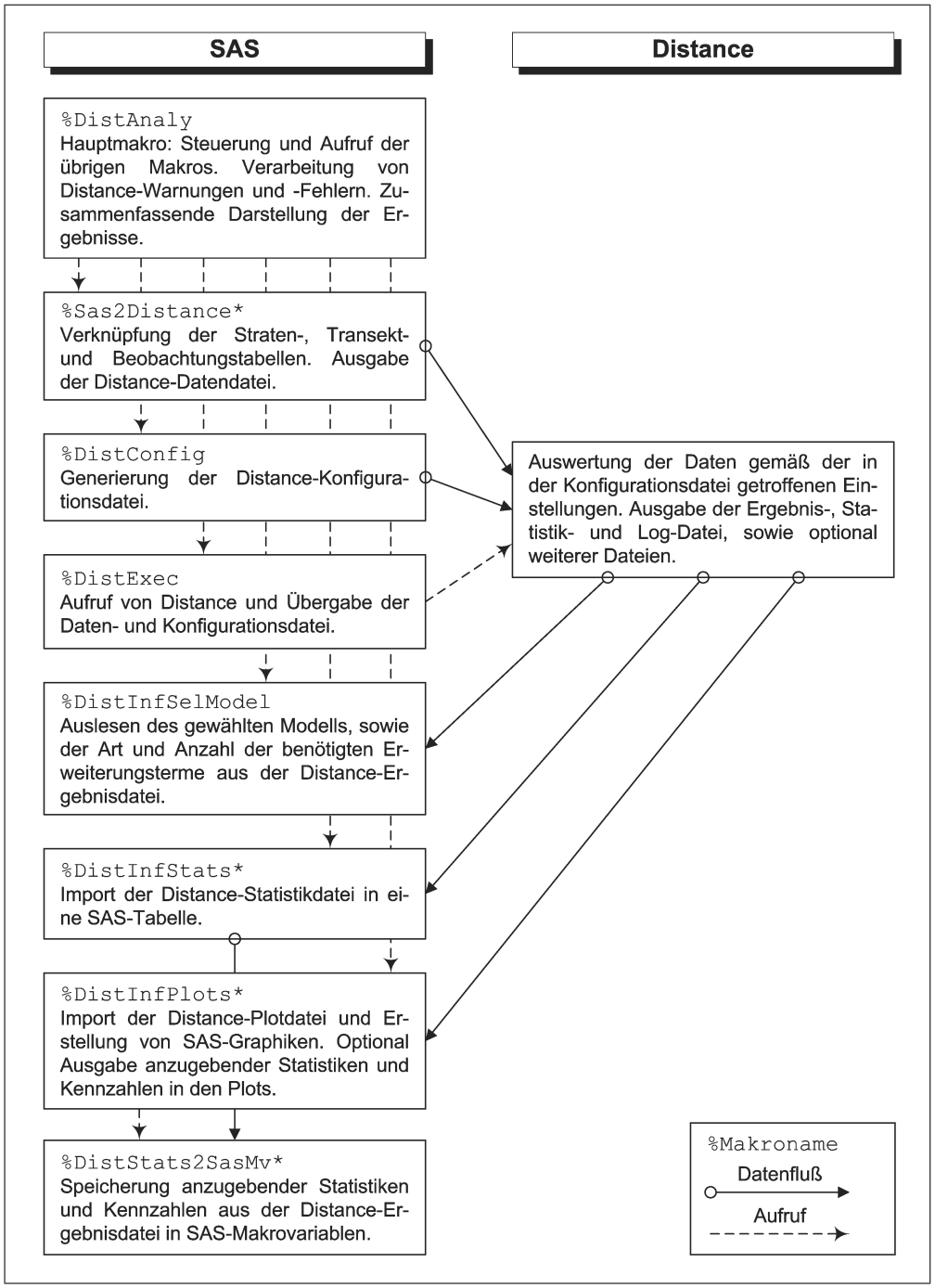

Abbildung 3.1.: Schematische Darstellung des Ablaufs der für die konventionelle Linientransektanalyse in Distance verwendeten SAS-Makros. Die mit * gekennzeichneten Makros lassen sich auch separat aufrufen, um Daten aus SAS zur Verwendung im Distance-GUI zu exportieren oder Distance-Ergebnisdateien in SAS zu importieren.

len Stichprobenfläche im Falle kurviger Transekte oder wenn diese sehr nahe an der Grenze des Untersuchungsgebietes liegen. Das Makro liest zunächst aus der Ergebnistabelle eines vorhergehenden `%DistAnaly`-Aufrufs die geschätzte effektive Streifenbreite $\hat{v}$ ein und übergibt diese gemeinsam mit anzugebenden Geodatensätzen der Transekte (Liniengeometrien) und des Untersuchungsgebietes (Polygongeometrien) an das Python-Skript `SmplFrac.py`. Dieses erstellt temporäre Streifengeometrien (`Buffer`) der Breite $2 \cdot \hat{v}$ um die Transekte und entfernt außerhalb des Untersuchungsgebietes liegende Streifenteile (`Clip`). Die Größe der sich daraus ergebenden Flächen ($\tilde{a}$) wird von `%SmplFrac` importiert und ins Verhältnis zur angenommenen Stichprobenfläche ($a = 2\,\hat{v}\,L$) gesetzt. Der resultierende Faktor kann bei einem erneuten `%DistAnaly`-Aufruf angegeben werden. Alle Ergebnisse werden dann entsprechend der realen Stichprobenfläche angepaßt.

Ein T-Test zum Vergleich zweier Dichteschätzungen läßt sich mit Hilfe des Makros `%DensTTest` durchführen. Das Makro liest aus zwei `%DistAnaly`-Ergebnistabellen die Individuen- oder Objektdichteschätzungen, die zugehörigen Freiheitsgrade und Variationskoeffizienten ein und berechnet anhand dieser die in (2.14) aufgeführte Teststatistik und den zugehörigen p-Wert.

Distance verwendet zur Schätzung der mittleren Gruppengröße standardmäßig die Beobachtungen, die auch für die Anpassung der Entdeckungsfunktion verwendet werden. Diese Auswahl läßt sich zwar weiter einschränken – z. B., um eine Korrelation zwischen Gruppengröße s und Entfernung x auszuschließen –, aber nicht erweitern. Auch die Angabe einer externen Schätzung der mittleren Gruppengröße ist nicht möglich. Dies ist jedoch beispielsweise dann sinnvoll, wenn neben den Daten zur Schätzung der Entdeckungsfunktion zusätzliche Daten für die Schätzung der mittlere Gruppengröße zur Verfügung stehen (s. z. B. Kap. 6.1). Es ließen sich so präzisere Schätzungen der mittleren Gruppengröße und damit auch der Individuendichte und -abundanz erzielen. Das SAS-Makro `%ExtCsEst` verwendet eine anzugebende mittlere Gruppengröße und deren Variationskoeffizienten oder Standardfehler, um die Einträge der Individuendichte- und Abundanzschätzung sowie deren Variationskoeffizienten in der Ergebnistabelle eines `%DistAnaly`-Aufrufs anzupassen.

Als letztes hier vorzustellendes Makro für die gewöhnliche Linientransektanalyse kann `%PreDistPlots` vor der eigentlichen Auswertung dazu verwendet werden, Übersichtgraphiken u. a. zur Auswahl eines geeigneten Modells für die Entdeckungsfunktion, der linken und rechten Kappungsentfernung sowie zur Beurteilung eines Zusammenhangs zwischen Entfernung und Gruppengröße zu erstellen. Das Makro erstellt Histogramme, ähnlich den von Distance erzeugten, welche die Anzahl der Beobachtungen in frei wählbaren Entfernungsintervallen zeigen und Scatterplots, die die beobachteten Gruppengrößen gegen die zugehörigen Entfernungen abtragen. In den Scatterplots wird zusätzlich eine einfache Regressionslinie, die zugrundeliegende Modellformulierung und ein p-Wert zur Beurteilung des linearen Zusammenhangs zwischen Gruppengröße und Entfernung dargestellt.

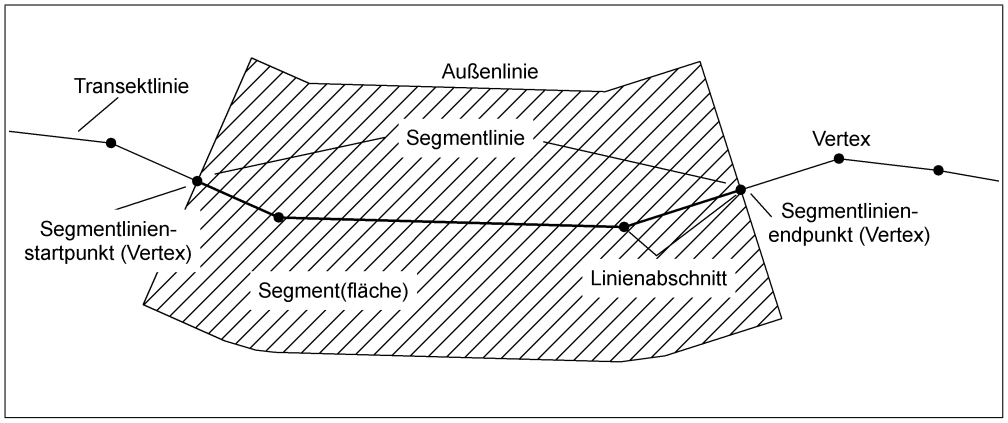

Abbildung 3.2.: Begriffe der Segmenterstellung: Die gesamte Transektlinie, deren Verlauf durch in Vertices miteinander verbundene Geraden (Linienabschnitte) definiert ist, wird in gleich lange Teilstücke (Segmentlinien) unterteilt, um die herum die Segmentflächen erstellt werden.

3.2. Raumbezogene Analyse

Soll es nicht bei einer gewöhnlichen Linientransektanalyse bleiben, sondern zusätzlich der räumliche Zusammenhang zwischen der lokalen Wilddichte und weiteren vorhandenen geographischen Informationen mit Hilfe der Modelle des Count-Modells aus Kapitel 2.3 untersucht werden, so folgt als nächster Arbeitsschritt die Erstellung der Segmente entlang der Transekte, die Ermittlung der Anzahl der hierin beobachteten Objekte und der Werte der zusätzlichen geographischen Informationen für jedes Segment. Im Anschluß daran läßt sich die Anzahl der in einem Segment entdeckten Objekte oder Individuen in Abhängigkeit der ermittelten räumlichen Kovariaten modellieren.

3.2.1. Segmente und räumliche Verschneidungen

Segmenterstellung Zur besseren Verständlichkeit der Erläuterung der Segmenterstellung verdeutlicht Abbildung 3.2 die hierbei verwendeten Begriffe.

Für die Segmenterstellung sind, wie in Kapitel 2.3 beschrieben, zunächst die Transektlinien in Teile gleicher, zuvor festgelegter Länge zu „zerschneiden“ (die Länge des letzten Teilstücks eines Transektes wird entsprechend angepaßt). Liegen die Transektgeometrien in Form sog. Routen (s. ArcGIS-Terminologie oben) vor, so läßt sich diese Unterteilung leicht mit Hilfe sog. Linienereignisse umsetzten. Zur Definition solcher Linienereignisse, der späteren Segmentlinien, wird neben der Routengeometrie lediglich eine Tabelle mit Start- und End-M-Werten (s. o.) der einzelnen Segmente benötigt, also die Angabe von wo bis wo – gemessen von Startpunkt der Route – ein Segment entlang der Route reicht. Bei Kenntnis der Gesamtlänge eines Transektes ist eine solche Tabelle leicht automatisch generierbar. Aus den so definierten Linienereignissen (`MakeRouteEventLayer`) lassen sich dann neue, die Segmentlinien repräsentierende Liniengeometrien erstellen und anhand dieser wiederum die Segmentflächen.

Besonders deutlich wird der Vorteil der Verwendung von Routengeometrien

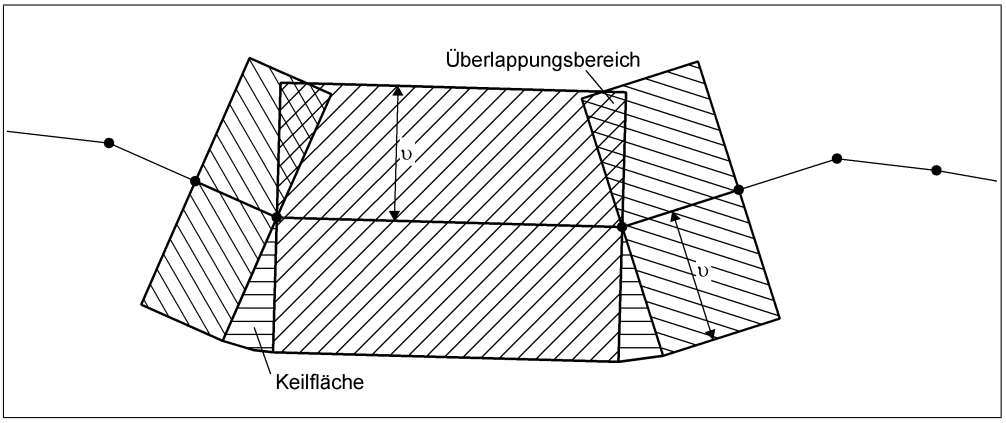

Abbildung 3.3.: Segmenterstellung: Um die (geraden) Linienabschnitte werden Rechtecke der Breite $\hat{v}$ und in Kurven zusätzliche Keilflächen erstellt. Anschließend werden die Grenzen zwischen den Teilflächen eines Segments und Überlappungsbereiche aufgelöst.

und Linienereignissen im Fall kurvig verlaufender Transekte. Während für gerade verlaufende Transekte die Positionen, an denen die Linien zu trennen sind auch leicht außerhalb eines GIS berechnet werden könnten, ist die Berechnung dieser Positionen für kurvige Transekte weit aufwendiger.

Um anschließend die Segmentflächen zu generieren, werden Streifen (in GIS-Programmen als *Flat Buffer* bezeichnet) der Breite $2 \cdot \hat{v}$ um die Segmentlinien erstellt. Für gerade oder nur leicht kurvig verlaufende Transekte ist dies in ArcGIS leicht mit Hilfe der Standardfunktion `Buffer` umzusetzen. Für deutlich gekrümmte Transekte sollten dagegen andere Lösungen ins Auge gefaßt werden (z. B. das unten näher erläuterte Geoprocessing-Skript `CreateComplex-Buffers.py`), da die von ArcGIS generierten Flächen dann teilweise deutlich von den eigentlich zu erstellenden Segmenten abweichen. Ursache hierfür ist die Art und Weise, in der die von ArcGIS berechneten (temporären) Außenpunkte der zu erstellenden Segmentgeometrie miteinander verbunden werden (näheres hierzu s. Anh. C.2).

Der Verlauf einer kurvigen Transektlinie (oder -route) ist technisch durch zahlreiche miteinander verbundene Vertices definiert, d. h. die Linie besteht aus aneinandergefügten Geraden unterschiedlicher Länge. Als einfache, in Abbildung 3.3 dargestellte Möglichkeit der Berechnung der Segmentgeometrien bietet es sich daher an, um die (geraden) Linienabschnitte einer Segmentlinie Rechtecke der Breite $\hat{v}$ zu erstellen und diese zu einer Gesamtfläche (der Segmentfläche) zu verschmelzen, wobei Überlappungsbereiche aufgelöst werden. Zusätzlich sind in Kurven auf der „offenen“ Seite Keilflächen zu erstellen, die nicht durch die Rechtecke erfaßt werden.

Für gerade Transekte sind die Ergebnisse des ArcGIS `Buffer`-Befehls und des alternativen Vorgehens identisch, wenngleich die `Buffer`-Funktion einen deutlichen Geschwindigkeitsvorteil aufweist. Für kurvige Transekte empfiehlt sich, je nach Stärke, Anzahl der Kurven und Größe der geschätzten effektiven Streifenbreite $\hat{v}$ ggf. der alternative Ansatz. Im Falle kurviger Transekte können zudem

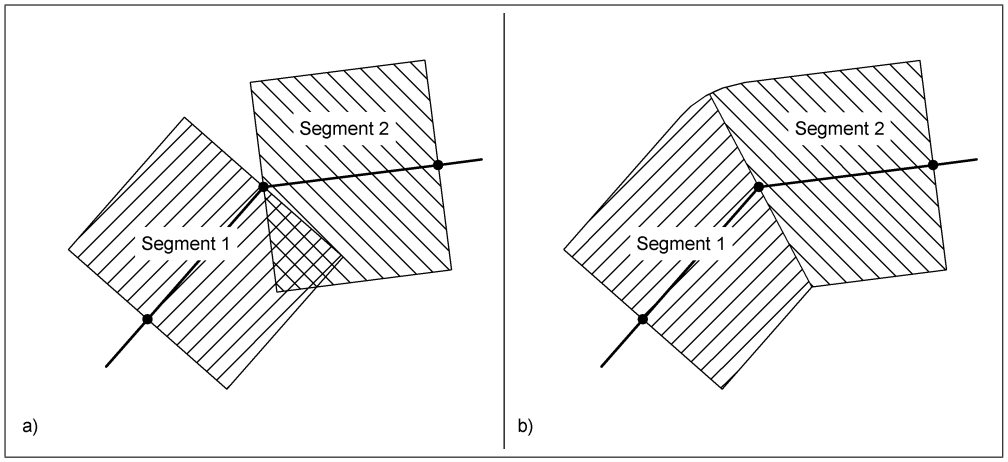

Abbildung 3.4.: Probleme der Segmenterstellung I: Fällt ein Segmentlinienendpunkt auf einen Vertex, in dem das Transekt seine Richtung ändert, so führt die Segmenterstellung zu falschen Ergebnissen (a), sofern die Richtungsänderung dabei nicht berücksichtigt wird (b).

zwei weitere Probleme entstehen, die jedoch in beiden Ansätzen nicht berücksichtigt werden. Das erste, dargestellt in Abbildung 3.4 a entsteht, wenn ein Segmentlinienendpunkt zufällig auf einen Vertex fällt, in dem die Transektlinie ihre Richtung ändert. Es kommt dann auf der „geschlossenen“ Transektseite zu einer Überlappung mit dem nachfolgenden Segment. Auf der „offenen“ Seite dagegen wird eine zu berücksichtigende Teilfläche in keinem der beiden Segmente erfaßt. Abbildung 3.4 a zeigt einen simplen Lösungsansatz für dieses Problem: Ähnlich dem Vorgehen für ein einzelnes Segment wird auf der „offenen“ Transektseite eine Keilfläche generiert, die aber zu gleichen Teilen unter den beiden Segmenten aufgeteilt wird. Auf der „geschlossenen“ Transektseite werden die für die betroffenen Linienabschnitte zu erstellenden Rechtecke ebenfalls so angepaßt, daß es nicht zu einem Überlappungsbereich kommt.

Dieses Problem dürfte in den meisten Anwendungen ein eher theoretisches bleiben, da ein Segmentlinienendpunkt nur sehr selten exakt auf einen Transektlinienvertex fallen wird, selbst dann, wenn die Transektlinie sehr präzise und damit verbunden mit sehr vielen Vertices digitalisiert wurde.

Das zweite, in Abbildung 3.5 verdeutlichte Problem dürfte von weit größerer Praxisrelevanz sein und kann sich ergeben, wenn eine Segmentlinie im Bereich einer Kurve endet. Die Segmentfläche müßte rechtwinklig zum letzten Linienabschnitt abschließen (gestrichelte Linie in Abb. 3.5 a), es existiert jedoch eine Teilfläche jenseits des Linienendes, deren Entfernung zu einem vorhergehenden Linienabschnitt kleiner ist als $\hat{v}$. Im Prinzip stellt dieser Effekt keinen Fehler dar, da die Fläche ja tatsächlich erfaßt[7] wurde, er führt aber wiederum zu einer Überlappung mit dem nachfolgenden Segment (Abb. 3.5 b). Dieses Problem läßt sich jedoch, wie der Abbildung leicht zu entnehmen ist, nicht durch eine einfa-

[7] Mit „erfaßt“ ist an dieser Stelle nicht gemeint, daß die Fläche bei der Aufnahme tatsächlich einsehbar gewesen sein muß und Objekte darin aufgenommen wurden, sondern lediglich, daß sie sich innerhalb des Abstandes $\hat{v}$ von der Segmentlinie befindet.

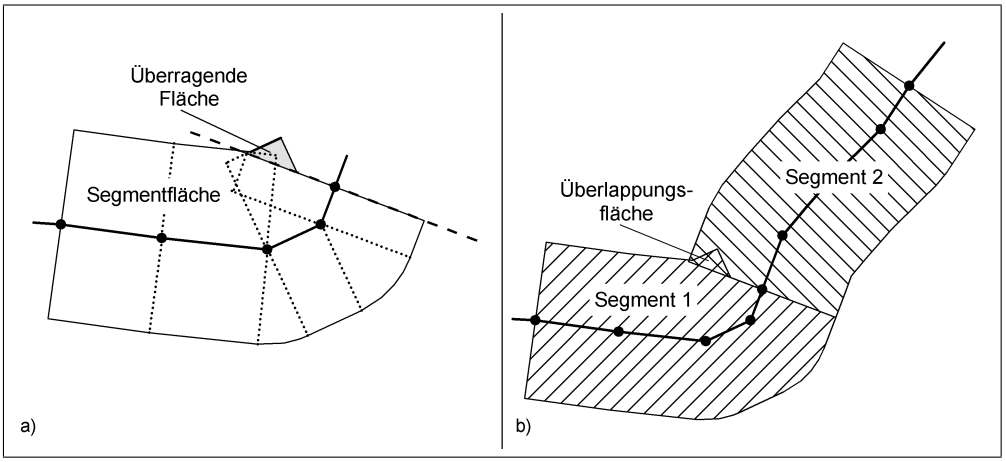

Abbildung 3.5.: Probleme der Segmenterstellung II: Endet eine Segmentlinie im Bereich einer Kurve, ragt – in Abhängigkeit von der Kurvenkrümmung und Streifenbreite $\hat{v}$ – die Segmentfläche über das eigentliche Ende des Segments hinaus (a), so daß es zu einer Überlappung mit dem nachfolgenden Segment kommt (b).

che Anpassung der generierten temporären Rechtecke und Teilflächen beheben. Eine vergleichsweise einfache Lösung wäre, dem ersten Segment den Vorrang einzuräumen, da dies ja zuerst aufgenommen wurde. Ein Segment würde also nur aus der Fläche bestehen, die nicht zuvor durch andere erfaßt wurde. Berücksichtigt man aber das Verfahren, nach dem eine Beobachtung einem Segment (genauer einer Segmentlinie) zugewiesen wird, nämlich dem nächstgelegenen, so wird klar, daß dieses Vorgehen dazu führen kann, daß ein Objekt einem Segment zugewiesen wird, die Fläche um das Objekt herum aber einem anderen. Die übrigen geographischen Informationen dieser Fläche gehen im späteren Modell also nicht für das Segment ein, in dem das Objekt beobachtet wurde. Sinnvoller wäre es demnach, die Überlappungsfläche so zwischen den Segmenten aufzuteilen, daß die Teilflächen dem jeweils näheren Segment zugeordnet werden. Die Umsetzung dieses an sich simplen Ansatzes ist für beliebige Transektverläufe, Streifenbreiten und Segmentlängen jedoch so komplex, daß im Rahmen dieser Arbeit darauf verzichtet wurde. Die Ergebnisse der im folgenden vorzustellenden Makros und Skripte können demnach sich überlappende Segmente enthalten. Größe und Anzahl der betroffenen Flächen hängen von der Anzahl der Kurven, der Stärke der Kurvenkrümmungen, der Positionen der Segmentlinienendpunkte und insbesondere von der Streifenbreite ab.

Der Ablauf der Makros und Skripte ist schematisch in Abbildung 3.6 dargestellt. Die wesentliche Funktionalität wird von den Geoprocessing-Skripten bzw. den darin angesprochen ArcGIS-Funktionen bereitgestellt, SAS dient im wesentlichen der Steuerung der Abläufe.

In `%CreateSegments` wird zunächst die zu verwendende Segmentlänge und (halbe) Segmentbreite angegeben. Alternativ zur Angabe einer Segmentbreite kann auch die geschätzte effektive Streifenbreite $\hat{v}$ aus einer `%DistAnaly`-Ergebnistabelle eingelesen werden. Das Makro importiert die Transektlängen aus der Attributtabelle der entsprechenden Routen Feature Class und erstellt

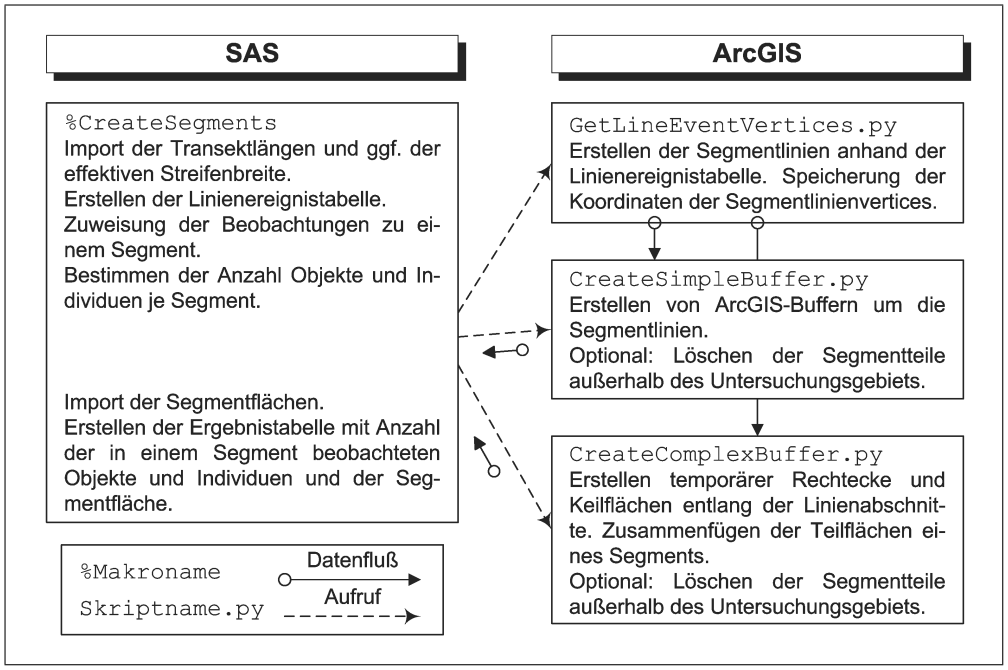

Abbildung 3.6.: Schematische Darstellung der für die Erstellung der Segmente verwendeten Makros und Skripte.

anhand dieser eine Segmenttabelle, die insbesondere die Start- und End-M-Werte (Linienereignisse) der zu erstellenden Segmentlinien enthält (ein Datensatz je Segmentlinie, Start- und End-M-Wert werden in Form von Spalten/Variablen gespeichert). Darüber hinaus erfolgt an dieser Stelle bereits die Zuordnung der Beobachtungen zu einem Segment. Eine Beobachtung wird dabei dem Segment zugewiesen, innerhalb dessen Routenabschnitts (M-Wert-Intervall zwischen Segmentlinienstart- und -endpunkt) der M-Wert der zugehörigen Transektposition (berechnet in `%CalcPerpDist` oben) liegt. Im Anschluß hieran wird die Anzahl der Objekte und Individuen je Segment bestimmt und die Segmenttabelle entsprechend erweitert. Anhand der in der Segmenttabelle gespeicherten Linienereignisse zerteilt `GetLineEventVertices.py` die Transektrouten in die einzelnen Segmentlinien (`MakeRouteEventLayer`). Des weiteren werden die (projizierten, kartesischen) Koordinaten der Segmentlinienvertices gespeichert (`FeatureVerticesToPoints`, `AddXY`). Liegen gerade Transekte vor, kann mit `CreateSimpleBuffer.py` die ArcGIS-eigene Funktion `Buffer` genutzt werden, um für jede Segmentlinie die entsprechende Segmentfläche zu generieren. Die gespeicherten Koordinaten der Segmentlinienvertices werden in diesem Fall nicht weiter verwendet. Für kurvige Transekte erstellt `CreateComplexBuffer.py` die Segmentflächen. Hierfür werden, wie oben beschrieben, anhand der Koordinaten der Segmentlinienvertices temporäre Rechtecke und Keilflächen entlang der Linienabschnitte erstellt, die dann zu den Segmentflächen verbunden werden (`Dissolve`). Beide Skripte entfernen im Anschluß optional die Segmentteile, die außerhalb des Untersuchungsgebietes liegen (`Clip`). Die resultierenden Segmentgeometrien werden in Form von Polygonen in einem Geodatensatz abgelegt.

`%CreateSegments` importiert aus dessen Attributtabelle die Flächenangaben der Segmente und fügt diese der Segmenttabelle hinzu. Ergebnisse sind demnach die Segmentgeometrien einerseits und die Segmenttabelle (im SAS-Format) mit Anzahl der Objekte und Individuen je Segment sowie der Segmentfläche andererseits (ein Datensatz je Segment).

Es sei an dieser Stelle noch einmal darauf hingewiesen, daß die Makros und Skripte derzeit nicht in der Lage sind, die Überlappungsbereiche zweier Segmente, die sich in Kurven ergeben können, zu entfernen. Eine Beobachtung wird aber korrekt nur einmal und zwar dem Segment mit der nächstgelegenen Segmentlinie zugewiesen. Zu beachten ist in diesem Zusammenhang auch, daß diese Art der Zuordnung es nicht erfordert, daß sich das Objekt tatsächlich in der betreffenden Segmentfläche befindet. Die Definition der effektiven Streifenbreite v erfordert es vielmehr, daß es Beobachtungen gibt, für die $x > \hat{v}$ gilt, die also außerhalb des Segments liegen, dem sie zugewiesen werden.

Räumliche Verschneidungen Unterschiedliche geographische Informationen räumlich miteinander in Bezug zu setzen gehört seit jeher zu den grundlegendsten Aufgaben von GIS-Software-Produkten. Im Rahmen dieses Projektes erwiesen sich drei Formen des „Räumlich-In-Bezug-Setzens“ als sinnvoll zur Bestimmung räumlicher Kovariaten für die Segmente. Sie werden graphisch in Abbildung 3.7 verdeutlicht.

1. Entfernungsmessung: Für jedes Segment wird die Entfernung zwischen seinem Schwerpunkt und dem nächstgelegenen Objekt einer weiteren Informationsebene, z. B. zur nächsten Siedlung, berechnet.

2. Punktverschneidung: An der Position des Segmentschwerpunkts wird der Zellwert eines Rasterdatensatzes abgefragt. Diese Verschneidungsart eignet sich beispielsweise zur Bestimmung der Geländehöhe anhand eines digitalen Geländemodells (DGM) oder zur Abfrage von aus Satellitenbildern generierten Informationen.

3. Flächenverschneidung: Hierbei wird die gesamte Segmentfläche mit einer ebenfalls flächenhaften Informationsebene räumlich verschnitten. Dabei wird das Segment in Teilflächen unterteilt, die den jeweils darin gültigen Wert der anderen Informationsebene enthalten. Um einen für das gesamte Segment gültigen Wert zu erhalten, wird entweder der (über die Teilflächen) flächengewichtete Mittelwert berechnet oder der Wert der größten Teilfläche verwendet. Flächenhaft vorliegende Informationen wie Bodentypen oder Vegetationsklassen lassen sich so einem Segment zuordnen.

Während also die Flächenverschneidung die gesamte Segmentfläche in die Berechnung mit einbezieht, gehen Punktverschneidung und Entfernungsmessung von der Repräsentativität des Segmentschwerpunkts für das ganze Segment aus. Mit zunehmender Segmentgröße ist diese aber immer weniger gegeben.

Zur software-technischen Umsetzung wurden ein SAS-Makro (`%SpatialQueryPolys`), das erneut im wesentlichen die Steuerung und Zusammenstellung der Ergebnisse übernimmt, und ein Python-Skript (`SpatialQueryPolys.py`),

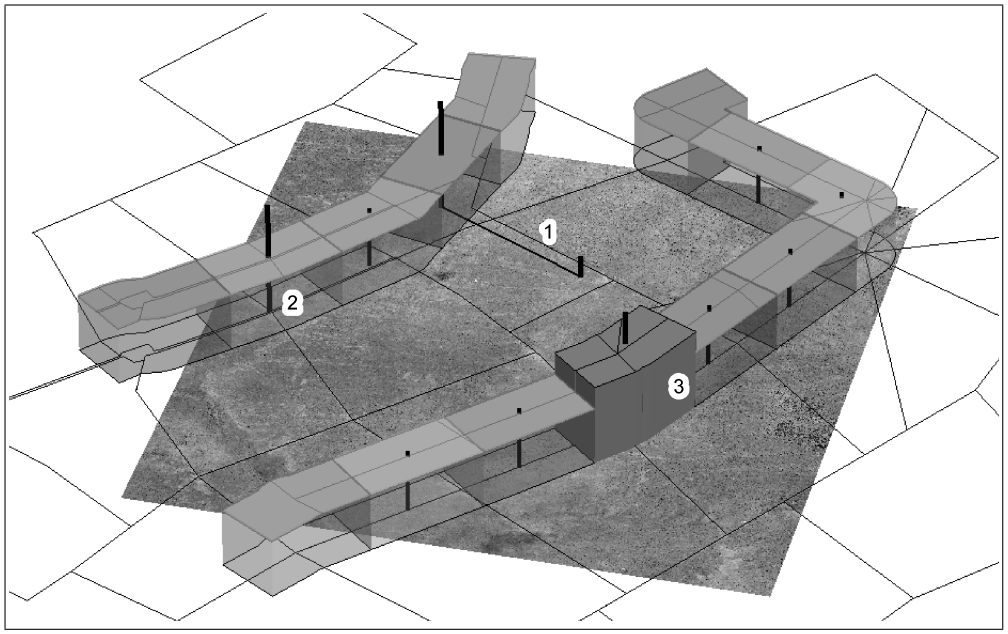

Abbildung 3.7.: Räumliche Verschneidung der Segmente mit weiteren geographischen Informationen. 1) Entfernungsmessung: Berechnung der Distanz zwischen Segmentschwerpunkt und nächstgelegenem Objekt (z. B. Salzlecken). 2) Punktverschneidung: Bestimmung der an der Stelle des Segmentschwerpunkts gültigen Information (z. B. Geländehöhe). 3) Flächenverschneidung: Berechnung der Größe der eingehenden Teilflächen und eines darüber gewichteten Mittelwertes der Informationswerte der Teilflächen (z. B. Rinderanzahl) oder Verwendung des Informationswertes der größten Teilfläche.

das die eigentlichen Verschneidungen durchführt, programmiert. Dem Skript werden der Segmentgeodatensatz und für die drei Verschneidungsarten jeweils (fast) beliebig viele weitere Vektor- bzw. Rastergeodatensätze übergeben.

`SpatialQueryPolys.py` beginnt mit der Berechnung der Segmentschwerpunkte (`FeatureToPoint`) und deren (projizierten) Koordinaten. Im Anschluß werden die kürzesten Entfernungen (`Near`) zwischen den Segmentschwerpunkten und Objekten der hierfür angegeben Geodatensätze und die Werte der an der Position des Schwerpunktes liegenden Rasterzellen bestimmt (`SurfaceSpot`). Flächenverschneidungen führt das Skript mit Hilfe der ArcGIS-Funktion `Identity` durch. Die Berechnung des flächengewichteten Mittelwertes oder die Bestimmung des Wertes der größten Teilfläche findet in `%SpatialQueryPolys` statt.

Alle ermittelten Informationen werden in Form neuer Variablen an die von `%CreateSegments` erstellte Segmenttabelle angefügt (zwischen der Tabelle der räumlichen Kovariaten und der in `%CreateSegments` erstellten Segmenttabelle besteht eine 1:1-Beziehung). Im Anschluß liegt eine Ergebnistabelle vor, die die Anzahl der Objekte n und Individuen m je Segment einerseits, und die Koordinaten des Segmentschwerpunkts sowie Werte zusätzlicher räumlicher Kovariaten andererseits enthält (Werte des Kovariatenvektors $\boldsymbol{x}'$). Die Tabelle eignet sich damit als Datengrundlage der log-linearen Regressionen des Count-Modells und ebenfalls für deren Erweiterung um die Berücksichtigung räumlicher Autokor-

relation.

Um Vorhersagen zu ermöglichen, sind darüber hinaus die räumlichen Kovariaten gleichermaßen für die Zellen eines über das Untersuchungsgebiet gelegten Gitters zu bestimmen. `CreatePolyGrid.py`, das über `%CreatePolyGrid` aus SAS heraus aufgerufen werden kann, erstellt ein solches Gitter mit beliebig wählbarer Zellbreite in Form eines Polygongeodatensatzes. Zellbereiche, die außerhalb des Untersuchungsgebietes liegen, werden dabei wahlweise entfernt (`Clip`), so daß ein Teil der Zellen von der Vorgabe der Zellgröße abweichende Flächen besitzt. Für die resultierenden Flächen (Zellen) lassen sich im Anschluß die räumlichen Kovariaten in der gleichen Weise wie für die Segmente mit Hilfe von `%SpatialQueryPolys` und `SpatialQueryPolys.py` bestimmen. Die Ergebnistabelle enthält, mit Ausnahme der Angabe über die Anzahl der beobachteten Objekte und Individuen, die gleichen Informationen wie die Segmenttabelle.

3.2.2. Parameterschätzungen und Vorhersagen

Zur Umsetzung des (statistischen) Modellierungsteils des Count-Modells steht in SAS bereits seit geraumer Zeit die Prozedur `GENMOD` zur Verfügung, die allgemein der Formulierung und Lösung verallgemeinerter linearer Modelle über ML- oder QL-Methoden dient. Auch die Berücksichtigung von korrelierten Beobachtungen ist begrenzt in Form von GEEs möglich, dient jedoch vorwiegend der Analyse von Meßwiederholungen. Der Einbezug räumlicher Autokorrelation ist nicht implementiert. Vorhersagen neuer Beobachtungen berechnet die Prozedur für das log-lineare Modell durch

$$\hat{y}_0 = \hat{\mu}_0 = \exp(\boldsymbol{x}_0'\hat{\boldsymbol{\beta}}). \tag{3.1}$$

Dies entspricht dem von Hedley (2000) vorgeschlagenen Vorgehen für das Count-Modell (vergl. Kap. 2.3). $\hat{\mu}_0$ ist hierin der geschätzte Erwartungswert der Anzahl der in einer Vorhersagezelle vorhandenen Objekte n_0 oder Individuen m_0. $\boldsymbol{x}_0$ ist der Designvektor der Vorhersage, zusammengesetzt aus den räumlichen Kovariaten der Vorhersagezelle. Die Modellformulierung gleicht also der zur Schätzung des Parametervektors $\boldsymbol{\beta}$, wobei hier bereits dessen Schätzung $\hat{\boldsymbol{\beta}}$ und anstelle der räumlichen Kovariaten der Segmente die der Vorhersagezellen verwendet werden. Die tatsächliche Position der vorherzusagenden Zelle spielt in diesem Modell keine Rolle.

Für eine umfangreiche Erläuterung der Möglichkeiten und Syntax von `PROC GENMOD` sowie exemplarische Anwendungen der Prozedur s. SAS (2004, Kap. 31) und Littel et al. (2002, Kap. 10).

Um das log-lineare Modell um den Einbezug räumlicher Autokorrelation zu erweitern, steht in SAS, wie eingangs erwähnt, seit kurzer Zeit eine neue Prozedur (`PROC GLIMMIX`) zur Verfügung. Sie dient allgemein der Formulierung und Lösung von verallgemeinerten linearen gemischten Modellen (GLMMs) mit unabhängigen oder korrelierten Fehlern. Zur Schätzung der Parameter der festen und gemischten Effekte sowie ggf. der Kovarianzparameter verwendet `PROC GLIMMIX` standardmäßig den teilweise in Kapitel 2.2.7.2 beschriebenen REPL-Ansatz von Wolfinger u. O'Connel (1993). Auf die Modellformulierung und den Anpassungsalgorithmus läßt sich in sehr umfangreichem Maße Einfluß nehmen;

eine detaillierte Beschreibung der Möglichkeiten und Syntax der Prozedur liefert SAS (2006).

Zur Angabe der räumlichen Kovarianzstruktur unterstützt `PROC GLIMMIX` bekannte Modelle wie das sphärische, Exponential- oder Gauß-Modell. Die Prozedur führt, bei Vorliegen räumlicher Autokorrelation, jedoch kein Universelles Kriging der Pseudo-Daten durch, sondern berechnet die Vorhersagen ebenfalls, wie in Gleichung (3.1) dargestellt, über die inverse Link-Funktion des linearen Prädiktors, wobei $\hat{\boldsymbol{\beta}}$ hier unter Berücksichtigung der Autokorrelation bestimmt wird. Um Vorhersagen nach den in Kapitel 2.2.7.4 beschriebenen Ansätzen zu ermöglichen, läßt sich das bereits seit 1992 verfügbare gleichnamige Makro `%GLIMMIX`[8] nutzen, das ebenfalls die Formulierung und Lösung von GLMMs zum Ziel hat, dabei aber wiederholte Aufrufe der Prozedur `MIXED` nutzt, die zur Vorhersage räumlich autokorrelierter Beobachtungen Universelles Kriging durchführt. Das Makro ist der Prozedur jedoch in zwei wesentlichen Aspekten deutlich unterlegen. So zeigte sich im Zuge der Modellierungen der namibischen Wildzählungsdaten, daß das Makro im Vergleich zu `PROC GLIMMIX` zum einen wesentlich mehr Rechenzeit benötigte und zum anderen oftmals nicht konvergierte, also gar keine Ergebnisse oder deutlich von denen der `GLIMMIX`-Prozedur abweichende lieferte. Dies galt selbst dann, wenn die in `PROC GLIMMIX` geschätzten Kovarianzparameter in `%GLIMMIX` als fest und bekannt angegeben wurden. Die Berechnung räumlicher Vorhersagen war entsprechend vielfach nicht möglich. Um dennoch ein Universelles Kriging der Pseudo-Daten durchzuführen, wurde ein zusätzliches Makro programmiert, das diesen Arbeitsschritt für den log-Link und die ggf. überstreute Varianz des Poisson-Modell übernahm. Dem Makro `%MargGlmSpatPred` werden eine Tabelle mit den räumlichen Kovariaten und Positionen der beobachteten (Segmentschwerpunkte) und vorherzusagenden Positionen (Schwerpunkte der Vorhersagezellen) sowie bestimmte Ergebnistabellen der `GLIMMIX`-Prozedur übergeben. Letztere enthalten die Parameterschätzungen der festen Effekte, deren Streuungsmaße und die geschätzten Kovarianzparameter. Das Makro berechnet dann die benötigten Kovarianzmatrizen sowie Vorhersagen und Vorhersagefehler anhand der in Kapitel 2.2.7.4 beschriebenen Modelle. Optional wird auch die Summe der Vorhersagen und deren Vorhersagefehler, wie in Kapitel 2.3.3 beschrieben, berechnet.

Außer den Vorhersagen unter Einbezug räumlicher Autokorrelation ließen sich also alle erforderlichen Analyseschritte wie die Schätzung der Parameter und die Durchführung von Hypothesentests mit Hilfe der beiden genannten Prozeduren durchführen. Um die Aufstellung und Überprüfung geeigneter Modelle zu erleichtern, wurden zusätzlich einige weitere, im folgenden kurz erläuterte Makros entwickelt:

Für eine explorative Datenanalyse log-linearer Modelle läßt sich das Makro `%GlmEda` verwenden. Da der Erwartungswert $\tilde{\mu}_i$ im log-linearen Modell mit Offset-Term t_i ($\ln(\tilde{\mu}_i) = \ln(t_i) + \boldsymbol{x}_i'\boldsymbol{\beta}$) durch $\ln(y_i) = \ln(t_i) + \boldsymbol{x}_i'\hat{\boldsymbol{\beta}} + \epsilon_i$ geschätzt wird, sollte $\ln(y_i) - \ln(t_i)$ linear bezüglich $\boldsymbol{x}_i$ sein. Das Makro plottet $\ln(1/n_i \sum_{j=1}^{n_i} y_{ij}) - \ln(1/n_i \sum_{j=1}^{n_i} t_{ij})$ gegen $\boldsymbol{x}_i$, wobei n_i die Anzahl identischer

[8] http://support.sas.com/ctx/samples/index.jsp?sid=536. Für ein (nicht-räumliches) Anwendungsbeispiel s. Littel et al. (1996, Kap. 11.6).

$\boldsymbol{x}_i$ ist. Liegt nur eine Beobachtung je Kovariatenkombination vor, vereinfacht sich der Ausdruck zu $\ln(y_i) - \ln(t_i)$. Die Graphiken eignen sich für die visuelle Überprüfung der Link-Funktion und geben bereits Hinweise auf die Vorzeichen und Dimension der zu schätzenden Parameter. Für $y_i = 0$ verwendet das Makro $\ln(y_i + 0.01)$. Liegen mehrere Beobachtungen je Designvektor $\boldsymbol{x}_i$ vor, so ist zudem die Überprüfung möglich, ob $\mathrm{E}(y_i) = \mathrm{var}(y_i)$. Zu diesem Zweck erstellt das Makro eine Graphik, in der der geschätzte Erwartungswert $\hat{\mu}_i^0 = 1/n_i \sum_{j=1}^{n_i} y_{ij}$ gegen die empirische Varianz $1/(n_i - 1) \sum_{j=1}^{n_i} (y_{ij} - \hat{\mu}_i^0)$ abgetragen wird. Gilt die Annahme der Gleichheit von Varianz und Erwartungswert, verläuft der Scatterpot entlang der Winkelhalbierenden.

`%GenmodAutoSelect` eignet sich zur automatisierten Auswahl von Kovariaten in `PROC GENMOD`. Das Vorgehen hierbei entspricht wahlweise dem aus klassischen linearen Modellen bekannten der schrittweisen oder Vorwärts-Selektion. Beginnend mit dem Modell, das lediglich einen Interzeptterm enthält, werden diesem sukzessive die Effekte hinzugefügt, die zusätzlich zu den bereits vorhandenen den größten erklärenden und signifikanten Beitrag liefern. Bemessen wird der erklärende Beitrag an der Wald-Teststatistik des sog. Typ-3-Kontrastes eines Effektes aus (2.40) oder (2.41). Im Falle der schrittweisen Selektion wird zusätzlich nach dem Hinzufügen eines Effektes überprüft, ob alle dann im Modell enthaltenen Effekte noch signifikant sind. Zur Steuerung der Aufnahme und Entfernung von Effekten können im Makro zwei Signifikanzgrenzen angegeben werden. Die erste gibt den p-Wert an, den ein nicht im Modell enthaltener Effekt höchstens erreichen darf, um noch in das Modell aufgenommen zu werden, die zweite den p-Wert, den ein im Modell enthaltener Effekt mindestens erreichen muß, um im Modell zu bleiben. Wie in gewöhnlichen linearen Modellen auch, ist dieses Vorgehen nicht unproblematisch, da es Gefahr läuft, zwar das Modell auszuwählen, das sich am besten zur Modellierung der konkreten vorliegenden Daten, evtl. aber nicht zur allgemeinen Beschreibung derartiger Daten eignet. Es führt keinen robusten Vergleich mehrerer sinnvoller Modelle durch und bemißt auch die Differenz zwischen den erzielten p-Werten verschiedener Effekte nicht, sondern wählt stets nur den Effekt mit dem kleinsten p-Wert, selbst wenn die p-Werte anderer Effekte nur marginal größer sind. Zudem enthält das Vorgehen keine sachlogische Überprüfung der geschätzten Parameter.

`PROC GENMOD` ermöglicht u. a. die Ausgabe einer Tabelle, die vorhergesagte Werte und unterschiedliche Residuen eines Modells enthält. Hierzu gehören u. a. die Pearson-Residuen aus (2.34) und ihre student-standardisierte Version aus (2.36). Das Makro `%GlmResPlot` liest diese Tabelle ein, fügt ihr die Anscombe-Residuen aus (2.33) hinzu und erstellt anhand der Tabellenwerte bis zu 20 einzeln auswählbare Graphiken. Bei den Graphiken handelt es sich um von verschiedenen Autoren (McCullagh u. Nelder, 1989, Kap. 12.6; Cameron u. Trivedi, 1998, Kap. 5.2; Fahrmeir u. Tutz, 2001, Kap. 4.2 und Lindsey, 1997, S. 227f.) vorgeschlagene Möglichlichkeiten zur Überprüfung der Wahl der Link- und Varianzfunktion, der Varianzhomogenität der Residuen, zur Ausweisung von einflußreichen Beobachtungen und Ausreißern sowie einer etwaigen Korrelation zwischen den Beobachtungen. Mit jeder Graphik werden ein kurzer Interpretationshinweis und eventuelle Ursachen für Abweichungen vom Sollzustand

ausgegeben.

`%GlimmixAutoSelect` dient wie `%GenmodAutoSelect` der automatisierten Variablenauswahl, verwendet für die Modellanpassungen jedoch die `GLIMMIX`-Prozedur und damit den PL- oder REPL-Algorithmus. Hier werden die Waldstatistiken aus (2.59) und die zugehörigen p-Werte verwendet, um aus den zur Verfügung stehenden Kovariaten diejenigen auszuwählen, welche die Streuung der Daten am besten erklären. Die Kovarianzparameter können entweder fest vorgegeben oder in jedem Selektionsschritt per PL oder REPL geschätzt werden. Es gelten hier dieselben Vorbehalte wie oben. Zudem konnte beobachtet werden, daß Modellanpassungen oftmals nicht konvergierten, der Beitrag, den ein zusätzlicher Effekt in einem Modell erzielte, also nicht bemessen und damit auch nicht mit dem anderer Effekte verglichen werden konnte.

`%EmpSV` schließlich verwendet die SAS-Prozedur `VARIOGRAM`, um das klassische empirische Semivariogramm (Matheron, 1962) und dessen robuste Variante (Cressie u. Hawkins, 1980) für eine Zielvariable und frei wählbare Entfernungsklassen zu berechnen. Beide Seimivariogramme werden für die gewählten Entfernungsklassen graphisch dargestellt und können dann zur Auswahl einer geeigneten theoretischen Semivariogrammfunktion und zur gutachterlichen Bestimmung der zugehörigen Parameter Schwellenwert (Sill), Nugget und Reichweite (Range) genutzt werden. Im Kontext räumlicher GLMs lassen sich die empirischen Semivariogramme der Pearson-Residuen aus (2.34) interpretieren, um die Kovarianzparameter zu bestimmen oder die per PL- oder REPL geschätzten zu überprüfen.

4. Untersuchungsgebiet und Wildarten

Dieses und das nachfolgende Kapitel erläutern die eingangs erwähnte praktische Umsetzung der in Kapitel 2 vorgestellten Methoden anhand von Wildzählungen auf kommerziellem, d. h. in Privatbesitz befindlichem Farmland in Zentralnamibia. Wie beschrieben, galt es, im Zuge der Erhebungen Daten für das Erreichen zweier Ziele zu sammeln. Zum einen sollte die Verläßlichkeit der Verwendung von Farmwegen als Transekte überprüft und zum anderen Möglichkeiten aufgezeigt werden, zusätzliche räumliche Informationen in die Analyse einzubeziehen.

Dieses Kapitel stellt zunächst das Untersuchungsgebiet und die untersuchten Wildarten vor, Kapitel 5 die Durchführung der Linientransektaufnahmen und die zusätzlich erfaßten geographischen Informationen.

4.1. Untersuchungsgebiet

Der als Untersuchungsgebiet ausgewählte Farmkomplex, dargestellt in Abbildung 4.1, umfaßt drei im zentralen Osten Namibias gelegene Farmen von insgesamt gut 9 000 ha Größe, von denen jedoch nur ein Teil tatsächlich für die Aufnahmen verwendet wurde (s. Kap. 5.1). Die drei Farmen gehören, wie einige Nachbarfarmen, zur kommerziellen Omitara Conservancy, andere Farmen in der Region sind Mitglieder der ebenfalls kommerziellen Seeis Conservancy. Der Farmkomplex liegt genau auf der Grenze der Verwaltungsgebiete Khomas und Omaheke und jeweils etwa 100 km ostnordöstlich bzw. westnordwestlich von deren Verwaltungszentren, Windhoek und Gobabis, entfernt.

Die nachfolgende Übersicht soll einen kleinen Eindruck über die örtlichen Gegebenheiten und Habitate der Region vermitteln.

Landschaft: In der gesamten Region vorherrschender Vegetationstyp ist die Kameldornsavanne, wobei die lokal sehr heterogene Vegetation von dichtesten Büschen bis zu weiten, offenen Grasflächen reicht. Weiteres charakteristisches Geländemerkmal sind viele, teils langgezogene, steil ansteigende und felsige Hügel. Die mittlere Geländehöhe über NN liegt im Farmkomplex bei etwa 1 600 m, fällt jedoch allgemein von Nordwesten nach Südosten deutlich ab; im Bereich des Farmkomplexes bereits von 1 650 auf 1 550 m. Beispiele für die Heterogenität der Landschaftsstrukturen sind in Abbildung 4.2 dargestellt.

Klima: Das Untersuchungsgebiet liegt, wie der Großteil Namibias, im tropischen Sommerregengebiet; Niederschläge fallen fast ausschließlich in der Zeit von November bis April, mit Schwerpunkt in der sog. Großen Regenzeit von Februar bis April. Der mittlere Niederschlag in dieser Region Namibias liegt

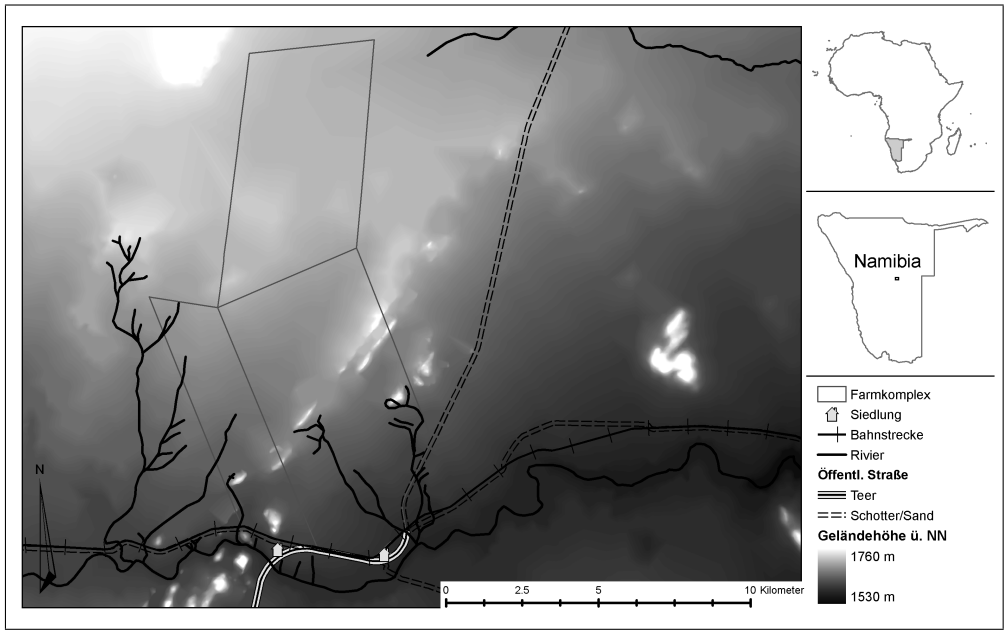

Abbildung 4.1.: Lage des Untersuchungsgebietes.

bei etwa 350 mm, jedoch mit starken zeitlichen und räumlichen Schwankungen (Mendelsohn et al., 2002). Dürreperioden über mehrere Jahre, in denen nur ein Bruchteil des mittleren Niederschlags fällt, sind keine Seltenheit. Die Tagestemperaturen liegen ganzjährig bei mindestens 20-25 °C, in den heißesten Sommermonaten Dezember und Januar auch deutlich über 30 °C. Während in den Sommermonaten auch nachts Temperaturen von 20 °C und mehr erreicht werden, kühlt es in den Wintermonaten Juli und August nachts deutlich bis teilweise unter den Gefrierpunkt ab.

Wasser: Permanent wasserführende Flüsse gibt es in der Region nicht, lediglich das Rivier[1] des White Nossob liegt an der südlichen Grenze des Untersuchungsgebietes. Von ihm gehen kleinere Ausläufer nach Norden ab. Menschen und Tiere sind daher auf Staudämme und Grundwasser angewiesen, das auf allen Farmen an mehreren Stellen über Windräder oder Motoren gefördert wird.

Erschließung und Infrastruktur: Etwas südlich des Farmkomplexes liegt die in Ost-West-Richtung verlaufende Teerstraße zwischen Windhoek und Gobabis. Neben dieser existieren nur einige weitere von ihr abgehende, meist geschotterte Straßen, von denen nur eine in direkter Nähe des Untersuchungsgebietes verläuft. Neben den wenigen öffentlichen Straßen gibt es auf allen Farmen ein enges, meist entlang der Farmzäune verlaufendes Wegenetz, das jedoch nur von Farmern und deren Angestellten genutzt wird.

In weiten Teilen parallel zur Teerstraße verläuft die einmal täglich genutzte

[1] Riviere sind trockene Flußläufe, die nur nach starken Regenfällen Wasser führen.

Abbildung 4.2.: Die Dichte der Vegetation in der Region des Untersuchungsgebietes reicht von weiten, offenen Grasebenen (a) bis zu fast völlig verbuschten Flächen mit nur wenigen Metern Sichtweite (b). Weitere charakteristische Landschaftsmerkmale sind die steil aus der ansonsten ebenen Fläche aufragenden Berge und die kilometerlangen, schnurgeraden Zäune (c).

Eisenbahnstrecke Gobabis-Windhoek.

Ein weiteres, sehr auffälliges Landschaftsmerkmal sind die oft kilometerlangen Zäune, die die Farmen in sog. Kamps von Größen zwischen 40 und 1 200 ha (durchschnittlich ca. 200 ha) unterteilen. In ihnen werden vorwiegend Rinder gehalten, die, um die Weide zu schonen, in regelmäßigen Abständen von Kamp zu Kamp umgestellt werden, so daß zur selben Zeit immer nur etwa 20 % der Kamps besetzt sind. Die Zäune der Rinderkamps haben eine Höhe von etwa 1.2 m und bestehen aus drei bis fünf einfachen, horizontal gespannten Drähten. Manche Farmen oder Farmteile sind darüber hinaus kriech- oder wildsicher eingezäunt. Kriechsichere Zäune verhindern durch zusätzlich eng über dem Boden gezogene Drähte den Wechsel von kriechendem Wild wie Oryx und Hartebeest. Wildsichere Zäune, mit bis zu 2.5 m Höhe können noch dazu auch von springendem Wild wie Kudu und Springbock nicht überquert werden. Auch der gesamte Farmkomplex, in dem das Untersuchungsgebiet liegt, ist kriechsicher eingezäunt, so daß zumindest für die Arten Oryx, Hartebeest und Steinbock die Zu- oder Abwanderung von Tieren im Laufe der Aufnahmen weitestgehend ausgeschlossen werden kann.

Tränken zur Versorgung der Rinder, die auch vom Wild genutzt werden, finden sich meist an den Kreuzungspunkten mehrerer Kamps. Sie sind stets gefüllt, stehen dem Wild also auch dann zur Verfügung, wenn sich keine Rinder in den anliegenden Kamps befinden.

Besiedelung: Die größten Städte der weiteren Umgebung sind Windhoek, die Hauptstadt Namibias, im Westen mit ca. 250 000 Einwohnern und Gobabis mit ca. 12 000 Einwohnern im Osten jeweils etwa 100 km entfernt. In direkter Umgebung des Untersuchungsgebietes existieren lediglich das Dorf Omitara im Süden mit unter 100 Einwohnern sowie eine provisorische Siedlung am südlichen bzw. südwestlichen Rand mit etwa 200 Bewohnern. Die Bevölkerungsdichte insgesamt liegt in dieser Gegend bei unter einem Einwohner je Quadratkilometer (Mendelsohn et al., 2002).[2] Auf den Farmen selbst konzentriert sich die Besiedlung auf ein vergleichsweise kleines Areal um das Farmgehöft, wo Farmer und Angestellte sowie deren Familien leben. Wenngleich also die Erschließung der Region durch Zäune und Wege mit einer Dichte von etwa 1.5 km/km^2 recht hoch erschient, so konzentriert sich die Besiedelung doch auf sehr wenige und kleine Areale. Weite Teile des Untersuchungsgebietes werden nur ein- bis zweimal wöchentlich auf Kontrollfahrten von Menschen betreten; Bereiche außerhalb der Farmwege noch weit seltener.

Wirtschaft und Jagd: Haupteinkommensquelle der Farmer dieser Region ist, wie oben angedeutet, die Rinderzucht. Die tragbare Rinderdichte der Weide und damit das wirtschaftliche Wohlergehen eines Farmers hängt jedoch sehr stark von der Menge, Verteilung und Dauer der Niederschläge ab. Immer mehr Farmer bauen sich daher ein zweites wirtschaftliches Standbein

[2] Zum Vergleich: Die Bevölkerungsdichte in Deutschland beträgt etwa 229 Einwohner je Quadratkilometer (Brockhaus, 1997).

über den Jagdtourismus auf. Das Wild ist besser an die klimatischen Verhältnisse und Dürreperioden angepaßt und bedarf kaum zusätzlicher Investitionen, stellt damit also, bezogen auf die Niederschlagsabhängigkeit, eine deutlich risikoärmere Alternative dar. Noch deutlich vor der Bejagung durch Gastjäger steht aber auf allen Farmen die Jagd zur Eigenversorgung der Farmer und ihrer Angestellten. Da kostenfrei und von bester Qualität, wird auf fast allen Farmen der gesamte Eigenbedarf an Fleisch über die Jagd gedeckt.

Die zum Untersuchungsgebiet gehörenden Farmen werden bislang nicht für Jagdtourismus oder ähnliche Aktivitäten genutzt, jedoch wurden in der Aufnahmepause von Anfang April bis Mitte Juni 2002 Hartebeest und Oryx gefangen und aus dem Gebiet entfernt. Da die genauen Anzahlen der gefangenen Tiere aber bekannt waren, ließ sich die Entnahme in der Auswertung der durchgeführten Aufnahmen berücksichtigen. Der Wildfang brachte zudem den Vorteil mit sich, daß zuvor der Oryx- und Hartebeestbestand des Farmkomplexes vollständig per Hubschrauberzählung bestimmt wurde und so (mit Einschränkungen, s. Kap. 6.1) mit den Ergebnissen der Linientransektanalysen verglichen werden konnten.

Mancherorts ebenfalls von größerer Bedeutung für Rinder- und Wildbestände sind Wilderer. Insbesondere in der Nähe der öffentlichen Straßen und Siedlungen werden regelmäßig Spuren von Wilderei entdeckt. Verläßliche Angaben über die Größenordnung der durch Wilderei entstandenen Schäden sind jedoch nicht vorhanden.

4.2. Untersuchte Arten

Bei den Wildarten, für die im Zuge der Aufnahmen Daten erhoben wurden, handelt es sich um die drei größten in der Region endemischen Antilopenarten Großer Kudu (*Tragelaphus strepsiceros*), Oryx (*Oryx gazella*) und Hartebeest (*Alcelaphus buselaphus caama*) sowie um den Steinbock (*Rhaphicerus campestris*) und das Warzenschwein (*Phacochoerus aethiopicus*). Alle fünf Arten werden jagdlich genutzt, wobei die ersten drei sowohl zur Deckung des Eigenbedarfs als auch für Trophäenjagden genutzt werden, die letzten beiden dagegen vorwiegend für die Trophäenjagd, die im Untersuchungsgebiet aber nahezu keine Bedeutung besitzt. Neben den genannten Arten kommen Springbock (*Antidorcas marsupialis*) und Kronenducker (*Sylvicapra grimmia*) auf der Herbivoren- sowie Gepard (*Acinonyx jubatus*), Leopard (*Panthera pardus*), Schabrackenschakal (*Canis mesomelas*) und Wüstenluchs (*Felis caracal*) auf der Carnivorenseite vor. Die Dichte dieser Arten ist im Untersuchungsgebiet jedoch so gering, daß nicht genügend Daten für eine Auswertung gesammelt werden konnten.

Die nachfolgenden Beschreibungen der untersuchten Arten konzentrieren sich im wesentlichen bereits auf die für die späteren Analysen potentiell bedeutsamen phänotypischen sowie verhaltensbiologischen Eigenschaften der Tiere. Die Angaben wurden zusammengetragen aus Estes (1997); Grzimek (1968, 1988); Walker (1968, 1983); Kingdon (1997); Batty (2002); Newell (1999a,b); Sanders (2005); Dorst u. Dandelot (1973); Huffman (2004a,b, 2005) und Creel (2005).

Für darüber hinausgehende Informationen sei auf diese Autoren verwiesen. Abbildungen der fünf Arten sowie Übersichten über deren Verbreitung sind in Abbildung 4.3 und 4.4 dargestellt, eine kurze Zusammenfassung der wesentlichen Eigenschaften und Habitatansprüche enthält Tabelle 4.1.

Hartebeest/Südafrikanische Kuhantilope (*Alcelaphus buselaphus caama*) Das Hartebeest war früher in verschiedenen Unterarten über den gesamten afrikanischen Kontinent sehr zahlreich verbreitet, ist heute jedoch nur noch in wenigen Gebieten vertreten. Dies gilt insbesondere für die in Namibia heimische Unterart *A. b. caama*, die hier unter besonderem Schutz steht, gerade in der Region um das Untersuchungsgebiet aber (wieder) sehr zahlreich vertreten ist. Für den Rückgang maßgeblich verantwortlich sind die starke Bejagung, die zunehmende Zerstörung des bevorzugten Habitattyps – weite offene Flächen –, der Zaunbau, der die Wanderung in bessere Weidegebiete verhindert, und die Haltung von Rindern, mit denen sie als fast ausschließliche Grasfresser in Nahrungskonkurrenz stehen. Die südafrikanische Kuhantilope ist etwa rothirschgroß, mit langem schmalen Kopf und hakenförmig nach hinten gebogenen Hörnern bei beiden Geschlechtern. Ihre soziale Organisation richtet sich maßgeblich nach der Verfügbarkeit von Wasser und dem Weideangebot. In der Trockenzeit und in generell trockeneren Gebieten schließen sich bis zu 300 Tiere (früher bis zu 10 000) zu Gruppen zusammen, um in in bessere Weidegebiete zu ziehen. Soziale Hierarchien und das sonst vorhandene Territorialverhalten älterer Bullen werden dann ggf. gänzlich aufgegeben. Während der Regenzeit oder in sehr guten Weidegebieten dagegen bilden sich deutlich kleinere seßhafte Gruppen, und das Territorialverhalten ist sehr ausgeprägt. Mehrere Kühe leben dann mit ihrem Nachwuchs in Gruppen von zwei bis zehn Tieren, junge und sehr alte Bullen leben in meist etwas kleineren Junggesellengruppen. Geschlechtsreife, starke Bullen besetzen Reviere, die sie gegen andere Bullen verteidigen. Abhängig von der saisonalen Wasserverfügbarkeit und guten Weidegründen lassen sich in derselben Population alle Zwischenformen zwischen vollständig standortstreu und rein nomadisch lebend sowie die beschriebenen sozialen Gefüge beobachten. Kuhantilopen bevorzugen weite offene Flächen mit kurzem Gras und offenem Wasser, dringen aber auch in trockene Gebiete und verbuschte Flächen vor. Den Großteil ihres Nahrungsbedarfs decken sie mit Gras. Um Wasser aufzunehmen, fressen sie aber auch Melonen, Wurzeln, Knollen etc. Sie gelten als ausgesprochen wachsame Tiere mit sehr gutem Gesichts- und Hörsinn und als sehr ausdauernde Läufer, die bei drohender Gefahr sehr früh und schnell fliehen. Ihre Hauptaktivitätszeit liegt in den frühen Morgen- und Abendstunden.

Großer Kudu (*Tragelaphus strepsiceros*) Der große Kudu ist nach dem Eland (*Taurotragus oryx*) die zweitgrößte Antilope Afrikas und wegen der langen korkenzieherartig gedrehten Hörner der Bullen sicher auch eine der eindrucksvollsten. An den hornlosen und deutlich kleineren Kühen fallen besonders die überdimensioniert wirkenden großen Ohren auf, die bereits auf den besonders gut ausgeprägten Hörsinn hindeuten. Ebenfalls gut

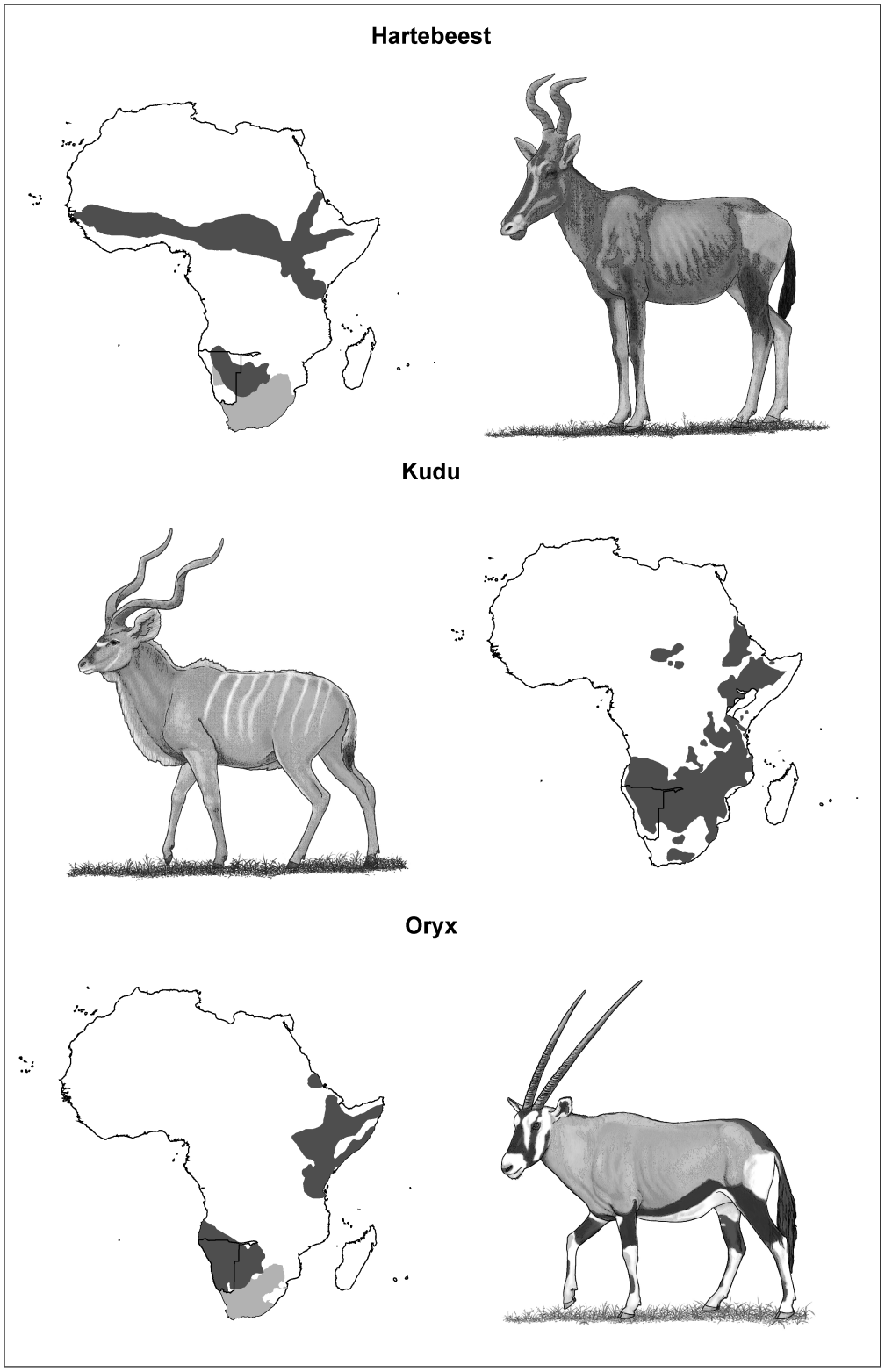

Abbildung 4.3.: Die untersuchten Wildarten Hartebeest, Großer Kudu und Oryx und ihre Verbreitungsgebiete (Verbreitungsgebiete nach AMD, 1998). Legende s. Abbildung 4.4.

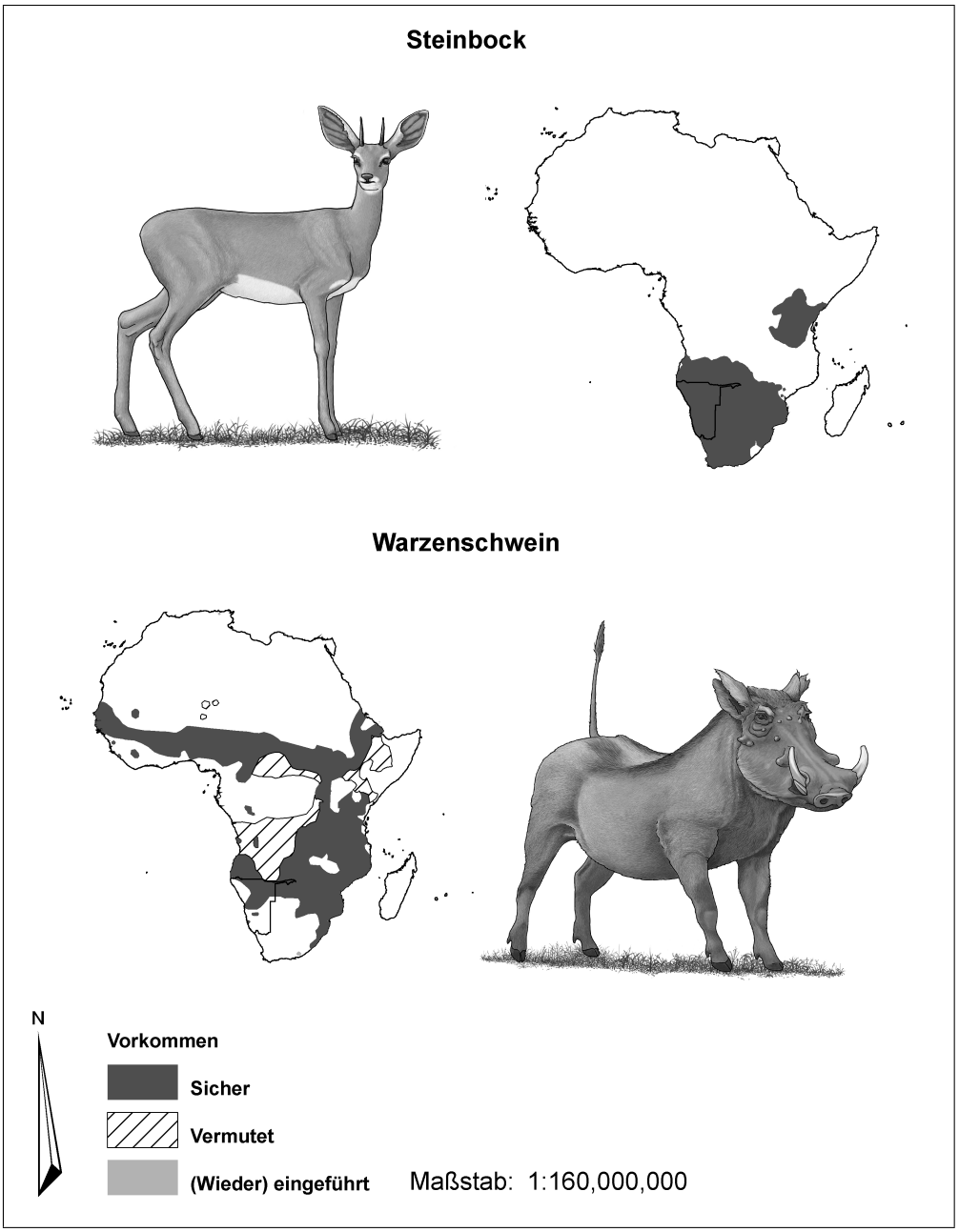

Abbildung 4.4.: Die untersuchten Wildarten Steinbock und Warzenschwein und ihre Verbreitungsgebiete (Verbreitungsgebiete nach AMD, 1998).

Tabelle 4.1.: Eigenschaften und Habitatansprüche der fünf untersuchten Wildarten.

	Kudu	**Hartebeest**	**Oryx**	**Steinbock**	**Warzenschwein**
Schulterhöhe	♂ 135 cm (122-150) ♀ 121 cm (100-140)	♂, ♀ 120-145 cm	♂, ♀ 120 cm (115-125)	♂, ♀ 50 cm (45-60)	♂, ♀ 75 cm (65-84)
Gewicht	♂ 257 kg (190-315) ♀ 170 kg (120-215)	♂ 135-200 kg ♀ 120-180 kg	♂ 176 kg (167-209) ♀ 162 kg (116-188)	♂, ♀ 11 kg (9-13)	♂ 82 kg (60-135) ♀ 65 kg (45-71)
Trophäe	nur ♂; lange, korkenzieherartige Hörner; 120 cm (max. 180)	♂ u. ♀; hakenförmig nach hinten gebogene Hörner; 48 cm (45-70)	♂ u. ♀; lange, lanzenartige Hörner; 105 cm (90-120)	nur ♂; kurze, gerade und spitz zulaufende Hörner; 13 cm (7-19)	♂ u. ♀; halbkreisförmig nach oben gebogene obere Eckzähne; ♂ 26-64 cm, ♀ 15-26 cm
Nahrung	Laub, daneben Gräser und Kräuter	Gräser und Kräuter, lehmige und salzige Erde, selten Laub	Gräser, Kräuter, Saftwurzeln, -früchte, Melonen, Blätter, Knospen	Vorwiegend Laub, daneben Gräser, Kräuter, Knollen und Wurzeln	Kurzes Gras, Sämereien, unterirdische Pflanzenteile, selten tierische Kost
Lebensraum	Steiniges, locker bis dicht bewaldetes oder bebuschtes Flach-, Hügel- und Bergland	Freie Steppen bis Trockensavannen sowie Buschgebiete in Ebenen und im Hügelland	Trockensteppen, Busch- und Baumsavannen in Ebenen und im Hügelland, Halbwüsten und Wüsten	Offenes, buschdurchsetztes Flachland, auch Grasland und Baumsteppen	Offenes Gelände, Savannenlandschaften, lichte Busch- und Grassteppen, vor allem in Wassernähe
Paarungszeit	Mai-Juli	Ganzjährig mit Höhepunkt im Feb./März	Ganzjährig mit Höhepunkt zw. Nov. und Jan.	Ganzjährig mit Höhepunkt im Juni/Juli	Zum Ende der Regenzeit/Anfang der Trockenzeit (Juni/Juli)
Setzzeit	Zu Beginn der Regenzeit (Feb./März)	Ganzjährig mit Höhepunkt gegen Ende der Trockenzeit (Okt./Nov.)	Ganzjährig mit Höhepunkt im Aug./Sep. Oft synchron innerhalb einer Gruppe	Ganzjährig mit Höhepunkt zum Ende der Trockenzeit (Nov./Dez.)	Zum Ende der Trockenzeit (Nov./Dez.)
Anz. Junge	1	1	1, selten 2	1, selten 2	3 (1-4, selten bis 8)

ausgeprägt ist der Geruchssinn, der Gesichtssinn dagegen weniger. Kudus bevorzugen Hügel- und Berglandschaften mit lockerem und dichtem Busch, sind aber auch im buschbestandenen Flachland zu beobachten; offene Landschaften werden dagegen eher gemieden. Deckung ist die wesentliche Habitateigenschaft. Genügend und dichtes Buschwerk vorausgesetzt, leben sie auch in dicht von Menschen besiedelten Gebieten. In Namibia gelten sie infolge der künstlich angelegten Brunnen und Tränken sogar als Kulturfolger. Bei starker Bejagung ziehen sich Kudus jedoch in weniger zugängliche Hügel- und Berggebiete zurück und sind fast ausschließlich nachts aktiv, so daß sie, selbst wenn zahlreich vorhanden, ggf. nur selten beobachtet werden. Ihr rötlich-braunes bis blau-graues Fell mit neun bis zehn senkrechten weißen Streifen macht sie zwischen dichten Büschen fast unsichtbar; so stehen sie bei etwaiger Gefahr auch meist völlig still und lassen diese sehr nahe an sich herankommen, bevor sie fliehen. Kudus sind nicht territorial, genügend Wasser vorausgesetzt, aber sehr standortstreu. Während der Trockenzeit sind sie oft an Flußbetten und am Fuß von Hügeln und Bergen zu sehen, wo die Vegetation dann noch reichhaltiger ist. Sie ernähren sich fast ausschließlich von Laub, fressen, wenn kein offenes Wasser verfügbar ist, aber auch Knollen und Melonen. Die Tiere bilden geschlechtergetrennte Gruppen mit meist nur wenigen Tieren (2-15, selten mehr); alte Bullen wandern gelegentlich allein umher. Die Gruppen sind während der Regenzeit und auf Farmland im allgemeinen kleiner, in der Trockenzeit finden sich auch deutlich mehr Tiere in den verbliebenen Weidegründen zusammen. Kudus sind, abhängig vom Jagddruck, mehr oder minder nachtaktiv, äsen nachts oder in den frühen Morgen- und späten Abendstunden und nehmen morgens und abends Wasser auf. Tagsüber ruhen sie meist im dichten Busch. Sie sind hervorragende Springer, die leicht Hindernisse bis 2.5 m Höhe überwinden.

Oryx/Gemsbock/Spießbock (*Oryx gazella*) Gemsböcke sind – von den hier beschriebenen Arten – die bei weitem am besten an trockene Bedingungen angepaßten Tiere. Sie bevorzugen zwar steinige, offene Flächen mit zumindest begrenztem Zugang zu Wasser, dringen aber auch tief in Halbwüsten und Wüsten vor, in denen nur wenige andere Huftiere überlebensfähig sind. Sie sind u. a. in der Lage, ihre Körpertemperatur um fast 10 °C zu erhöhen, um so die Transpiration zu reduzieren und können, wenn nötig den gesamten Wasserbedarf über Pflanzen (Melonen, Wurzeln, Knollen usw.) decken. Sie verlegen die Äsungszeit dann in die Nachtstunden, wenn die Vegetation mehr Feuchtigkeit enthält, und sind verstärkt am Fuß von Hügeln und Bergen sowie entlang von Flußläufen zu finden, wo die Vegetation noch üppiger ist. Sie ernähren sich vorwiegend von Gras, das auch bereits sehr trocken sein kann. In weniger trockenen Gebieten liegt die Hauptaktivitätszeit in den frühen Morgen- und Vormittagsstunden sowie in den Nachmittags- und Abendstunden. In der Hitze der Mittagssonne ziehen sich die Tiere in den Schatten zurück. Spießböcke sind in der Lage, Wasservorkommen und Niederschläge über weite Entfernungen auszumachen, und legen auf der Suche nach Wasser und Salzlecken, an denen sie

vermehrt anzutreffen sind, sehr weite Strecken zurück. Diese stark nomadische Lebensweise, in der geschlechtergemischte Herden mit 50 bis 200 Tieren umherziehen, geht während der Regenzeit und in sehr guten, dauerhaften Weidegebieten über in eine eher seßhafte mit geschlechtergetrennten und deutlich kleineren Gruppen, bestehend aus durchschnittlich 14 Tieren (2-40). Vermutet wird, daß in dieser Phase einzelne geschlechtsreife Bullen auch Reviere besetzten, sich anderen Bullen gegenüber aber weit toleranter verhalten, als dies bei anderen Arten der Fall ist. Das Territorialverhalten ist demnach, wenn überhaupt, nur gering ausgeprägt. Oryxkühe und -bullen sind aus der Entfernung nur sehr schwer zu unterscheiden; beide Geschlechter tragen die markanten spitz zulaufenden Hörner und sind sich auch darüber hinaus äußerlich sehr ähnlich. Die Kühe sind lediglich 10-20 % kleiner und besitzen schlankere, aber etwa gleich lange Hörner. Das Fell der Spießböcke ist weißgrau bis bräunlichgrau mit weißen und schwarzen Markierungen. Besonders auffällig ist die schwarze halfterartige Gesichtszeichnung. Oryxantilopen gelten als wachsame Tiere mit hervorragendem Gesichts-, Riech- und Hörsinn, die bei Gefahr sehr früh fliehen. Sie sind schnelle und sehr ausdauernde Läufer, jedoch auch bevorzugtes Opfer von Wilderern, da sie sich durch Hunde stellen lassen.

Steinbock (*Rhaphicerus campestris*) Der Steinbock ist, mit einer Schulterhöhe von lediglich etwa 50 cm, die mit Abstand kleinste der untersuchten Antilopenarten. Beide Geschlechter besitzen ein einheitlich gelb-bräunliches Fell, die Männchen tragen kurze, spitz zulaufende Hörner, die Weibchen sind gelegentlich etwas schwerer. Steinböcke sind fast ausschließlich einzeln oder paarweise anzutreffen. Es wird vermutet, daß sie in dauerhaften Beziehungen leben, die Paare wesentliche Teile des Tages und auch viel längere Zeiträume aber getrennt verbringen. Sie besetzen gemeinsame Reviere zwischen 5 und 100 ha Größe. Bevorzugtes Habitat sind weite und offene Landschaften, die jedoch genügend Deckung in Form von Büschen, hohem Gras u. ä. aufweisen müssen. Sie gelten als Kulturfolger, die in vom Menschen durch Kultivierung und Wegebau für sie geöffnete Gebiete vordringen. Sie ernähren sich vornehmlich von Laub, aber auch von Gräsern und Sprößlingen sowie Wurzeln und Knollen und sind nicht auf offenes Wasser angewiesen, nehmen es aber an, wenn vorhanden. Abhängig von der Besiedlungsdichte und dem Jagddruck liegt die Hauptaktivitätszeit fast vollständig in der Nacht oder vorwiegend tagsüber mit den üblichen Höhepunkten frühmorgens und am späten Nachmittag. Bei Gefahr verbergen sich Steinböcke zunächst im Gebüsch oder hohem Gras, wo sie aufgrund ihrer geringen Größe und der Fellfärbung nur sehr schwer auszumachen sind. Erst beim unmittelbaren Herannahen der Gefahr fliehen sie mit hoher Geschwindigkeit und oft im Zickzacklauf, um sich dann erneut in geeigneter Deckung abzudrücken.

Warzenschwein (*Phacochoerus aethiopicus*) Das Warzenschwein ist die bei weitem wasserabhängigste der hier vorgestellten Arten. Es nimmt täglich Wasser auf und suhlt in heißen Perioden auch täglich. Dennoch handelt

es sich um die an offene und trockene Lebensräume am besten angepaßte Schweineart und zudem um die einzige grasende. Warzenschweine ernähren sich während der Regenzeit vorwiegend von Gras und buddeln in den trockenen Monaten unterirdische Pflanzenteile aus; tierische Kost wird nur selten angenommen. Warzenschweinbachen und -keiler lassen sich im wesentlichen an der Größe – Bachen sind 15-20 % kleiner – und anhand der mächtigen, bis 60 cm langen, nach oben gebogenen oberen Eckzähne der Keiler unterscheiden. Durch das Suhlen nehmen die Tiere die Farbe des Bodens an und sind dann nur schwer auszumachen. Warzenschweine sind, sofern der Jagddruck durch den Menschen nicht zu groß ist, ausgesprochen tagaktive Tiere, die lediglich in der Nacht und selten auch kurze Zeit in den heißen Mittagsstunden ihre Erdhöhlen aufsuchen. Ihre Hauptaktivität zur Nahrungssuche liegt, wie bei den meisten der übrigen Steppen- und Savannenbewohner, in den frühen Vormittags- und späten Nachmittagsstunden. Sie leben meist in kleinen Gruppen von zwei bis sechs Tieren (sehr selten bis 18), bestehend aus einer oder zwei Bachen und deren letzten beiden Würfen. Keiler schließen sich während der Rauschzeit einer solchen Gruppe für einige Zeit an und ziehen sonst allein oder in Junggesellengruppen von bis zu fünf Tieren umher. In der Regel leben die Gruppen in einem Gebiet von 0.4 bis 600 ha um ihren Geburtsort herum. Die Gebiete verschiedener Gruppen überlappen sich stark, feste Punkte wie Suhlen, Wasserstellen und Bauten werden friedlich gemeinsam genutzt, meist aber nicht zur selben Zeit. Die Haut der Warzenschweine ist, abgesehen von der schwarzen, braunen, gelblichen oder rötlichen Hals- und Rückenmähne, fast nackt. Zudem besitzen sie kein Unterhautfett, so daß sie relativ empfindlich gegenüber extremen Temperaturen sind. Bei besonders kalten oder heißen Bedingungen verlassen sie ihren Bau später bzw. ziehen sich über die Mittagshitze dortin zurück. Droht Gefahr, so fliehen Warzenschweine in hohem Tempo auf der Suche nach einem Erdloch, aus dem heraus sie sich ggf. verteidigen können.

5. Datenerhebung

Wie eingangs erwähnt, kommt in mehreren kommunalen und kommerziellen Conservancies das Linientransektverfahren zur Erfassung der aktuellen Wildbestände zum Einsatz. In dem Farmkomplex, in dem das Untersuchungsgebiet liegt, werden seit Ende der 1990er Jahre mehrmals jährlich Zählungen nach diesem Verfahren durchgeführt, wobei, wie in fast allen namibischen Anwendungen, Wege als Transekte verwendet werden, da Zäune, Bodenverhältnisse und dichte Vegetation ein Befahren abseits der Wege vielfach unmöglich machen. Damit einher geht die ebenfalls angesprochene kritische Frage nach der Verlässlichkeit der darauf basierenden Dichteschätzungen. Um dieser Frage nachzugehen, wurden im Untersuchungsgebiet auf permanent angelegten Transekten Linientransektaufnahmen nach drei verschiedenen Schemata durchgeführt. Zum einen wurden den Farmeraufnahmen entsprechende Zählungen durchgeführt, in denen Farmwege als Transekte verwendet wurden und von der Ladefläche eines Geländewagens beobachtet wurde. Zum zweiten wurden zufällig in das Untersuchungsgebiet gelegte Transekte per Pferd abgeritten. Zum dritten wurden die Wegtransekte ebenfalls per Pferd abgeritten, um einen möglichen Einfluß der Erfassung per Auto von dem der Verwendung von Farmwegen trennen zu können.

Die Datenerhebungen der drei Schemata erfolgte so, daß sie auch für das Ziel des zweiten Teilprojektes, die Erweiterung klassischer Linientransektaufnahmen um eine räumliche Komponente, verwendet werden konnten. D. h. es war nicht nur die Länge der Transekte und die Entfernungen zu beobachteten Objekten, sondern darüber hinaus die Position der Transekte und jeder einzelnen Beobachtung festzuhalten, um so einen räumlichen Bezug zu anderen geographischen Informationen wie z. B. der Weidequalität oder der nächstgelegenen Wasserstelle herstellen zu können und mit Hilfe dieser zusätzlichen Informationen Dichteschätzungen zu präzisieren und kleinräumige Vorhersagen zu ermöglichen. Benötigt wurden also verortete Linientransektdaten einerseits und eine Bandbreite zusätzlicher geographischer Informationen andererseits. Letztere wurden entweder direkt während der Aufnahmen erfaßt oder im Nachhinein mit Hilfe von Luft- und Satellitenbildern, topographischen Karten und von Farmern zur Verfügung gestellten Informationen erstellt. Der für eine räumliche Analyse unabdingbare geographische Bezug der Daten wurde dabei entweder durch die Aufnahme per GPS-Gerät oder im Zuge der Bearbeitung im GIS sichergestellt.

5.1. Wildzählungen

Abbildung 5.1 zeigt das Untersuchungsgebiet und den Verlauf der insgesamt zwölf Transekte. Das Wegtransekt mit einer Länge von 15.6 km entspricht der

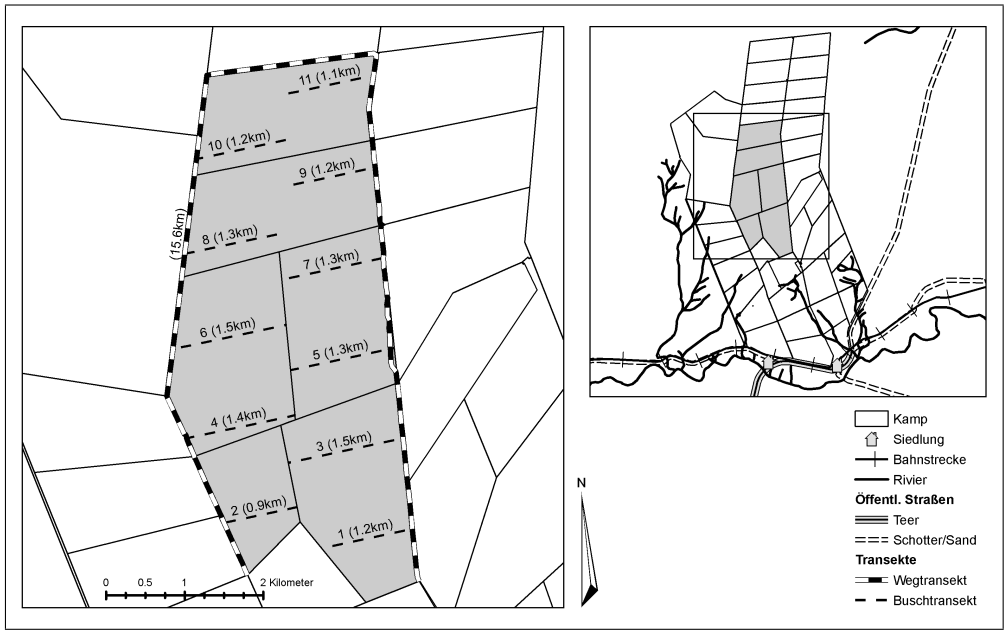

Abbildung 5.1.: Lage der Transekte im Untersuchungsgebiet.

Aufnahmepraxis der in der Region durchgeführten Farmerzählungen, die elf sog. Buschtransekte entstammen einer systematischen Anordnung mit zufälligem Startpunkt – mit der Einschränkung, daß ein Transekt stets nur durch einen Kamp verläuft, um zeitintensive Zaunüberquerungen zu vermeiden. Der Abstand zwischen den Transekten wurde mit 1.2 km so gewählt, daß sich die einsehbaren Bereiche benachbarter Transekte nicht überlappen, um Doppelsichtungen weitestgehend auszuschließen. Die elf Buschtransekte mit Längen zwischen 0.9 und 1.5 km (Gesamtlänge: 14.0 km) wurden in zwei Etappen an zwei aufeinanderfolgenden Tagen abgeritten. Das Wegtransekt wurde an einem Aufnahmetag per Auto abgefahren, an einem weiteren per Pferd abgeritten. Der für die Buschtransekte notwendige Pferdeeinsatz ging mit einer deutlichen Absenkung der Aufnahmegeschwindigkeit einher, so daß nicht der gesamte Farmkomplex als Untersuchungsgebiet verwendet werden konnte, sondern nur ein kleineres Areal von etwa 1 600 ha Größe (grau markierte Fläche in Abb. 5.1). Die Navigation auf den Buschtransekten erfolgte zunächst über einen GPS-Empfänger. Nach einigen Probeläufen zeigte sich jedoch, daß ein Großteil der Aufmerksamkeit der Navigation und dem Bemühen, auf dem Transekt zu bleiben, zum Opfer fiel. Die Transekte wurden daher permanent markiert und das GPS-Gerät nur noch unterstützend sowie zur Aufzeichnung der Beobachtungspositionen verwendet. Die Zählungen wurden in zwei Aufnahmephasen zwischen dem 28. November 2001 und 10. April 2002 sowie zwischen dem 8. Juni und 2. September 2002 insgesamt 20 mal durchgeführt (zwölf Aufnahmen je Aufnahmeschema in Phase 1, acht in Phase 2), um genügend viele Daten für eine Auswertung zu erfassen.

Die eigentliche Datenerfassung verlief für alle drei Aufnahmeschemata nach demselben Prinzip: Der Zählbeginn lag etwa eine Stunde nach Sonnenaufgang, gezählt wurde stets vom selben Beobachter von der Ladefläche eines Geländewa-

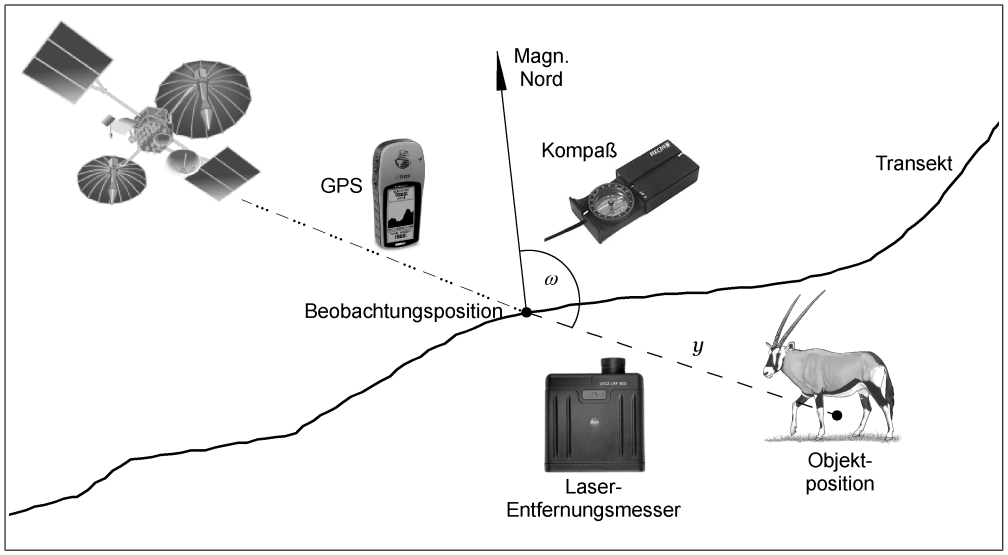

Abbildung 5.2.: Erfassung der Beobachtungs- und Objektposition: Bei Entdeckung eines Objektes wurden die Position des Beobachters per GPS, die Entfernung zum Objekt y per Laser-Entfernungsmesser und der Winkel ω zu magnetisch Nord per Kompaß erfaßt.

gens oder vom Rücken eines Pferdes aus. Bei Entdeckung eines Objektes wurde gestoppt und, wie in Abbildung 5.2 dargestellt, die aktuelle Position per GPS-Empfänger gespeichert, der Winkel zwischen magnetisch Nord und Objekt ω sowie die Entfernung zwischen Objekt und Beobachter y gemessen und notiert. Darüber hinaus wurden Uhrzeit, Art und Anzahl der Tiere sowie ggf. weitere beobachtungsspezifische Bemerkungen festgehalten.

Die Positionsermittlung erfolgte mit einem gewöhnlichen 12-Kanal-GPS-Empfänger (Garmin ETrex Summit). Aufgrund der sehr guten Empfangsqualität ließ sich auch ohne zusätzliche Korrektursignale oder höherwertige Geräte eine mittlere Abweichung von weniger als 4 m erreichen, die für die weitere Bearbeitung durchaus ausreichend war. Zusätzlich zu den in Form sog. Wegpunkte gespeicherten Beobachtungspositionen wurden die automatisch gespeicherten *Tracks*[1] genutzt, um im Anschluß den exakten Transektverlauf im GIS nachvollziehen zu können. Für die Winkelmessung kam der Kompaß DP 65 der Firma Recta zum Einsatz, die Entfernungsmessung erfolgte mit dem Leica Laser-Entfernungsmesser LRF 800. Trotz einer angegebenen Reichweite des Entfernungsmessers von 800 m, bei einer Genauigkeit von ± 1 m, lieferte dieser aufgrund kleiner anzuvisierender Ziele und störender Äste und Büsche nur bis zu einer Entfernung vom maximal 400 m verwertbare Angaben. Da der weitaus größte Teil der Beobachtungen jedoch deutlich unter dieser Entfernung lag (99.6 % bis 400 m, 97.9 % bis 300 m), ergaben sich hieraus keine nachhaltigen Probleme.

[1] Während der gesamten Laufzeit speichert der GPS-Empfänger in durch Geschwindigkeit sowie Stärke und Häufigkeit der Richtungsänderung bedingten Intervallen die aktuelle Position. Hieraus ergeben sich die sog. Tracks.

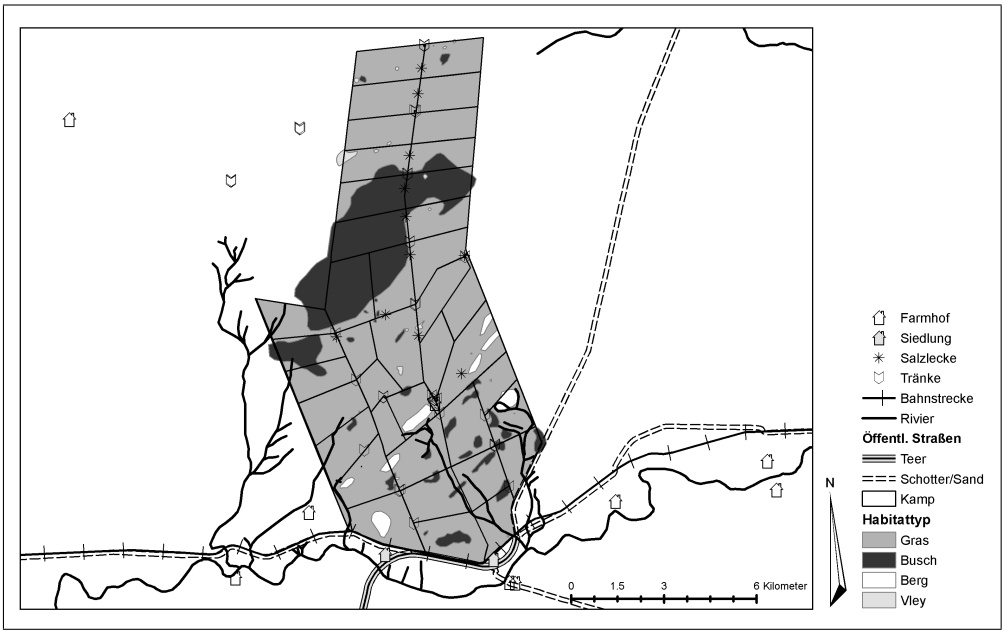

Abbildung 5.3.: Ausschnitt aus den erfaßten Geodaten.

5.2. Raumbezogene Informationen

Da für den Bereich Zentralnamibias kaum verläßliche geographische Informationen zur Verfügung standen – den verfügbaren Geodaten mangelte es an Angaben über das zugrundeliegende räumliche Referenzsystem, der erforderlichen räumlichen Auflösung oder Genauigkeit –, waren alle raumbezogenen Daten, von denen angenommen wurde, daß sie einen Einfluß auf die Verteilung des Wildes haben könnten, zunächst auf Grundlage von Luft- und Satellitenbildern, topographischen Karten und mit Hilfe der Angaben von Farmern digital zu erstellen. Die nachfolgende Auflistung enthält jeweils eine kurze Beschreibung des Inhalts und der Bedeutung der generierten Geodaten sowie eine Hypothese bezüglich der vermuteten Auswirkung auf die lokale Wilddichte. Abbildung 5.3 zeigt einen Teil der erstellten Geodaten.

Während für Wasserstellen und Salzlecken die zu diesen Objekten gemessene Entfernung von Bedeutung sein könnte, ist für den Vegetationstyp oder die Geländehöhe der Wert direkt an der Stelle eines Segments von Interesse. Es wurden daher drei Formen des räumlichen „In-Bezug-Setzens“ ausgewiesen, die zusammen mit ihrer software-technischen Umsetzung bereits in Kapitel 3.2.1 beschrieben wurden. Für *Entfernungsinformationen* geht die Distanz zwischen einem Segment und dem nächstgelegenen Objekt, z. B. der nächsten Wasserstelle, als Wert der Kovariaten in die spätere Analyse ein. Für *Punktinformationen* ist dies der Wert direkt an der Stelle des Segmentschwerpunkts (z. B. Geländehöhe), für *Flächeninformationen* wird ein flächengewichteter Mittelwert bestimmt. In diesem Fall liegen mehrere Werte der hinzuzufügenden Information im Bereich der Segmentfläche (z. B. Rinderanzahl).

Entfernungsinformationen

Tränken: Von ihnen wurde vermutet, daß sie gerade außerhalb der Regenzeit (im wesentlichen der Zeitraum der zweiten Aufnahmephase) einen großen (anziehenden) Einfluß auf das Wild und hierbei insbesondere auf Warzenschweine haben würden. 35 Tränken wurden innerhalb und im Umkreis des Farmkomplexes per GPS erfaßt, wobei meist mehrere Tränken auf kleinem Raum angelegt waren (z. B. in zwei Kamps beiderseits des Zauns), so daß 18 räumlich deutlich getrennte Positionen auszumachen waren. Vier der Tränken wurden als (weitestgehend) unerreichbar für Wild eingestuft, da sie so umzäunt waren, daß ihre Nutzung durch Wild als sehr unwahrscheinlich angesehen werden konnte. Es wurde ferner danach getrennt, ob eine Tränke für alle untersuchten Arten erreichbar war, also innerhalb des Farmkomplexes lag, oder nur für Kudu und Warzenschwein. Insgesamt fünf Tränkenpositionen lagen direkt an den untersuchten Flächen.

Salzlecken: Um das Wild mit benötigten Mineralien zu versorgen, bringen viele Farmer an verschiedenen Stellen der Farm – oft direkt an den Tränken – Salzlecksteine aus. Diese haben laut Auskunft der Farmer eine extrem anziehende Wirkung, zumindest auf die größeren Wildarten. Elf solcher Plätze wurden, wie die Tränken, per GPS erfaßt; sechs von ihnen lagen direkt an oder im Untersuchungsgebiet.

Farmhöfe: Farmhöfe stellen neben den Siedlungen die einzigen Orte dauernder menschlicher Aktivität dar. Es ließ sich daher vermuten, daß das Wild in der Regel in größerem Abstand zu diesen bleibt, wenngleich vereinzelt Tiere an Tränken in direkter Hofnähe beobachtet werden konnten. Innerhalb des Farmkomplex liegt lediglich ein einzelner Farmhof im Süden des untersuchten Gebietes, einige weitere liegen südlich des Farmkomplexes.

Siedlungen: Für die beiden im Süden bzw. Südwesten des Farmkomplexes liegenden Siedlungen wurde, stärker noch als für die Farmhöfe, ein abweisender Einfluß vermutet. Neben dauernder menschlicher Aktivität kann nicht ausgeschlossen werden, daß die Siedlungen Ausgangspunkt für Wilderer sind. Dies beträfe insbesondere Oryxantilopen, da sie leicht zu jagen und daher bevorzugte Opfer der Wilderer sind.

Besiedelte Orte: Diese Variable enthält für jeden Segmentschwerpunkt das Minimum der Entfernungen zum nächsten Farmhof und zur nächsten Siedlung, um einen generellen Einfluß dauerhaft von Menschen bewohnter Plätze zu berücksichtigen, ohne diesen in den Effekt von Siedlungen und Farmhöfen zu trennen.

Öffentliche Straßen: Die in der Nähe des Farmkomplexes verlaufenden Straßen sind zwar nicht stark frequentiert, aber dennoch Orte regelmäßiger Beunruhigung und zudem, wie vereinzelt von Farmern berichtet wurde, ebenfalls Ausgangspunkt für Wilderei, so daß ihnen ein negativer Einfluß auf die lokale Dichte aller untersuchten Arten unterstellt wurde.

Eisenbahn: Eine weitere, wenngleich nur einmal täglich genutzte Beunruhigungslinie ist die im Süden des Farmkomplexes verlaufenden Bahntrasse. Auch sie ist evtl. Ausgangsort für Wilderer und wirkt sich dann negativ auf die Wilddichten aus.

Wilderei: Ähnlich dem Vorgehen bei besiedelten Orten wurden unter dieser Variablen alle potentiellen Ausgangsorte für Wilderei zusammengefaßt. Sie enthält den für ein Segment jeweils kleinsten Wert der Variablen „Siedlungen“, „Öffentliche Straßen“ und „Eisenbahn“.

Riviere: Ein anziehender Effekt wurde dagegen den Rivieren und deren stärkeren Ausläufern unterstellt. Da sich hier nach starken Regenfällen das Wasser sammelt und das Grundwasser höher ansteht, ist die Vegetation am Rand des Flußbettes und der Rivierausläufer deutlich üppiger und länger grün als auf den übrigen Flächen.

Vleys: Auch den sog. Vleys[2] wurde ein positiver Effekt zugeschrieben. Da sich nach Niederschlägen auch hier das Wasser sammelt und geraume Zeit hält, ist die Vegetation um Vleys herum ebenfalls oft grüner und dichter. Oryxantilopen konnten zudem mehrfach beobachtet werden, wie sie die Vleys als Ruheplätze nutzen. Insgesamt 21 Vleys mit Größen zwischen 0.04 und 18 ha wurden auf dem Farmkomplex anhand von Luftbildern und digital vorliegenden topographischen Karten digitalisiert. Um die Entfernung zum nächstgelegen Vley auch für die Arten, die nicht durch die kriechsichere Umzäunung des Farmkomplexes beeinflußt werden (Kudu und Warzenschwein), korrekt zu bestimmen, wurden darüber hinaus weitere Vleys auf den benachbarten Farmen erfaßt.

Berge: Um die Hypothese zu überprüfen, daß insbesondere Kudus sich bevorzugt in der Nähe von oder direkt auf den felsigen Hügeln und Bergen aufhalten, wurden aus Luftbildern und topographischen Karten alle Berge und Hügel im und um den Farmkomplex herum digitalisiert.

Wege: Um – sofern überhaupt vorhanden – einen Einfluß der Wege zu quantifizieren, der nicht durch daran gelegene Tränken und Salzlecken erklärt werden kann, wurde für die Segmente auch die Entfernung zum nächstgelegenen Weg bestimmt.

Punktinformationen

Koordinaten: Zur Aufdeckung eines eventuell vorhandenen räumlichen Trends, der nicht durch andere Kovariaten erklärt werden kann, wurden (projizierte) x- und y-Koordinaten des Segmentschwerpunkts sowie x^2, y^2 und $x \cdot y$ als mögliche erklärende Variablen in das Modell aufgenommen. Die Koordinaten der Segmentschwerpunkte gingen darüber hinaus in die Schätzung der Parameter der räumlichen Autokorrelation ein.

[2] Vleys sind mit tonigen Sedimenten überzogene, meist vegetationsfreie Pfannen, um die herum die Vegetation jedoch oft üppiger ist.

Geländehöhe: Um einen Einfluß der Geländehöhe zu überprüfen, wurde auf Basis der in topographischen Karten (Maßstab 1:50 000) enthaltenen Höhenlinien ein digitales Geländemodell (DGM) in Form eines Rasterdatensatzes mit einer räumlichen Auflösung von 10 m × 10 m erstellt. Wie die zuvor genannten Koordinaten dürfte die Geländehöhe selbst keinen direkten Einfluß auf die Verteilung der Wildarten besitzen, es ist jedoch denkbar, daß sie indirekt weitere Informationen wie sich mit der Geländehöhe ändernde Habitattypen oder Vegetationszusammensetzungen enthält.

Niederschlag: Von wesentlicher Bedeutung für die lokale Wilddichte dürften die Niederschläge der vergangenen Wochen/Monate/Saison sein. Auch diese dürften wiederum keinen direkten eigenen Einfluß besitzen, jedoch stellvertretend für das verfügbare Weideangebot stehen. Je nach Intensität und Dauer der Niederschläge ließen sich vermeintlich attraktive Weideflächen ausweisen. Es gelang jedoch nicht, für den Zeitraum vor und während der Aufnahmen genügend und verläßliche Niederschlagsdaten zu sammeln, die eine flächendeckende Interpolation zugelassen hätten. Eine Verwendung der frei verfügbaren afrikaweiten 10-Tages-Niederschlagsschätzungen[3] (CPC/FEWSa) des Climate Prediction Center (CPC) der National Oceanic & Atmospheric Administration (NOAA) kam wegen der zu groben Auflösung von etwa 10 km × 10 km nicht in Betracht. Statt dessen wurde versucht, die Weidequalität anhand von aus Satellitenbildern ableitbaren Vegetationsindizes zu bestimmen (s. u.).

Brightness: Hierbei handelt es sich um den ersten Kanal der sog. *Tasseled-Cap-Transformation* (Kauth u. Thomas, 1976; Crist u. Cicone, 1984), die, ähnlich einer Hauptkomponentenanalyse, die Informationen der verschiedenen Kanäle eines Satellitenbildes, in wenigen, neu berechneten Kanälen verdichtet, ohne dabei allzuviel der Ursprungsinformation (d. h. -variabilität) zu verlieren. Auch hier stehen die Achsen des neu berechneten Koordinatensystems annähernd senkrecht aufeinander, im Gegensatz zur Hauptkomponentenanalyse sind die auf empirischen Erhebungen beruhenden Parameter der Tasseled-Cap-Transformation (sie entsprechen den Eigenvektoren der Hauptkomponentenanalyse) jedoch vorgegeben und die Interpretation der resultierenden Kanäle klar definiert. So werden die Werte des Brightness-Kanals als Maß für den Anteil offenen Bodens auf dem vom Bildpixel repräsentierten Teil der Erdoberfläche gewertet.

Im Falle der Berechnung aus Landsat-7-Szenen (ETM+ Sensor) bezieht die Tasseled-Cap-Transformation die Kanäle 1, 2, 3, 4, 5 und 7 ein, wobei der Brightness-Kanal als weitestgehend gleichgewichtete (mit stärkerem Gewicht auf Kanal 4) Linearkombination der sechs eingehenden Kanäle berechnet wird. Tabelle 5.1 zeigt die Koeffizienten zur Berechnung der Tasseled-Cap-Transformation anhand der Daten des ETM+ Sensors und Abbildung 5.4 ein Beispiel der so berechneten Kanäle für ein in der Nähe des Untersuchungsgebietes liegendes Areal. Mehr zur Tasseled-Cap-Transformation findet sich u. a. in Jensen (1996) und Lillesand u. Kiefer (1994),

[3] Erhältlich unter http://www.cpc.ncep.noaa.gov/products/fews/data.html.

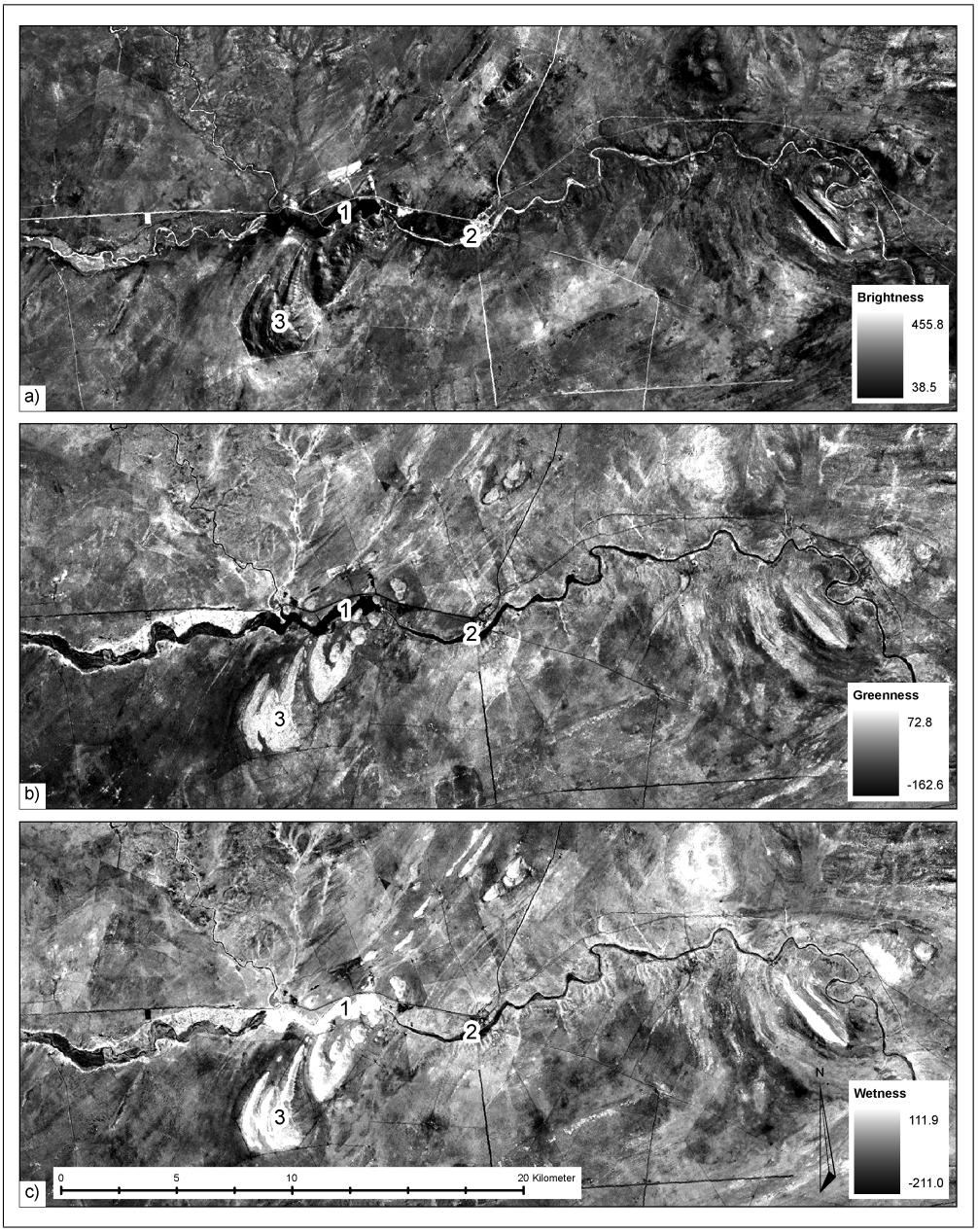

Abbildung 5.4.: Brightness- (a), Greenness- (b), und Wetness-Kanal (c) der Tasseled-Cap-Transformation am Beispiel eines in der Nähe des Untersuchungsgebietes liegenden Areals. 1 markiert ein zum Zeitpunkt der Aufnahme gefülltes Staubecken, 2 einen Teil des darin mündenden, ausgetrockneten Riviers, 3 eine Berggruppe. Gut erkennbar sind hohe Brightness-Werte für Wege, Straßen und das ausgetrocknete Rivier, niedrigere für die Wasserfläche und Berge. Die Greenness dagegen zeigt sehr niedrige Werte im Bereich des Wassers und der Wege, dafür hohe entlang der Rivierausläufer und auf den Bergen. Für die Wetness können erwartungsgemäß sehr hohe Werte im Bereich des Staubeckens und ebenfalls sehr hohe Werte im Bereich der Berge beobachtet werden.

Tabelle 5.1.: Koeffizienten der Tasseled-Cap-Transformation für die Kanäle des ETM+ Sensors (aus Huang et al., 2002).

Kanal	1 Blau	2 Grün	3 Rot	4 Nahes IR	5 Mittleres IR	7 Mittleres IR
Brightness	0.3561	0.3972	0.3904	0.6966	0.2286	0.1596
Greenness	-0.3344	-0.3544	-0.4556	0.6966	-0.0242	-0.2630
Wetness	0.2626	0.2141	0.0926	0.0656	-0.7629	-0.5388

zur Herleitung der Koeffizienten für den ETM+ Sensor in Huang et al. (2002).

Hohe Brightness-Werte sprechen für geringen Bewuchs, also für offene Flächen und damit das bevorzugt von Hartebeestantilopen genutzte Habitat.

Berechnet wurde der Brightness-Kanal und die nachfolgend erläuterten Greenness- und Wetness-Kanäle auf Grundlage einer Landsat-7-Szene der U.S. Geological Survey (USGS, 2004) mit einer räumlichen Auflösung von 30 m × 30 m vom 24. April 2004.[4] Zeitnah zu den Aufnahmephasen entstandene Satellitenbilder standen nicht zur Verfügung. Daß für beide Aufnahmephasen die Informationen desselben Satellitenbildes verwendet wurden und dieses zudem deutlich nach Beendigung der Aufnahmen entstanden ist, ist daher in den später folgenden Auswertungen zu berücksichtigen.

Greenness: Sie berechnet sich, wie Tabelle 5.1 zu entnehmen ist, für den ETM+ Sensor maßgeblich als Differenz aus der Summe der drei den Bereich des sichtbaren Lichts erfassenden Kanäle (1, 2 und 3) und dem des nahen Infrarots (4). Interpretiert wird die neue, zweite Achse der Tasseled-Cap-Transformation als Maß der Vegetationsdichte. Diese Herleitung beruht auf der Eigenschaft gesunder, grüner Vegetation ca. 40-50 % der im nahen Infrarotbereich (0.7-1.1 μm) einfallenden Energie zu reflektieren, jedoch 80-90 % des sichtbaren Lichts (0.4-0.7 μm) für die Photosynthese zu absorbieren (Jensen, 1983). Im Gegensatz dazu reflektiert tote Vegetation einen Großteil des sichtbaren Wellenspektrums und absorbiert mehr im Bereich des nahen Infrarots. Trockene Böden schließlich absorbieren einen größeren Teil des sichtbaren Lichts als tote und weniger als lebende Vegetation, reflektieren aber einen geringeren Teil des nahen Infrarotlichts als grüne und tote Vegetation (Jensen, 1996). Diese charakteristischen Unterschiede machen sich neben der Greenness zahlreiche weitere Vegetationsindizes wie der unten beschriebene NDVI zunutze.

Der Vegetationsdichte wurde ein nicht unerheblicher Einfluß auf die artspezifische Verteilung des Wildes unterstellt. Zum einen betraf diese das Nahrungsangebot, zum anderen die Habitatpräferenz für eher offene, weit einsehbare oder eher dichte, verbuschte Flächen. Eine eventuell über die Vegetationsdichte hinaus aussagekräftige Klassifizierung der Vegetation

[4] Frei erhältlich bei der *Global Landcover Facility* unter http://www.landcover.org.

bezüglich der Artenzusammensetzung wurde aufgrund mangelnder Kontrollinformationen nicht versucht.

Wetness: Der dritte Kanal der Tasseled-Cap-Transformation bemißt die Menge der in Boden und Vegetation enthaltenen Feuchtigkeit. Die kleinsten Werte werden für wasserfreie, versiegelte Flächen, die größten für trübe, etwas kleinere für klare Wasseroberflächen erreicht. Da Wasser ganzjährig der nahezu alles limitierende Faktor ist, dürften Gebiete hohen Wassergehalts stets besonders attraktiv für Tiere aller Arten sein, gleich ob es sich um offene Wasserflächen oder üppige, wasserhaltige Vegetation handelt. Jedoch dürfte insbesondere die Verläßlichkeit der Wetness-Information aufgrund der großen zeitlichen Differenz zwischen Datenerhebung und Aufnahme des Satellitenbildes als fragwürdig angesehen werden.

NDVI: Der *Normalized Difference Vegetation Index* (Kriegler et al., 1969; Rouse et al., 1973) ist der wohl gängigste, aus Satellitenbildinformationen abgeleitete Vegetationsindex. Wie der Greenness-Kanal der Tasseled-Cap-Transformation nutzt auch er die charakteristischen Absorptions- und Reflexionseigenschaften toter und lebender Vegetation im Bereich des sichtbaren und nahen infraroten Lichts, bezieht in die Berechnung aber lediglich den roten und den nahen infraroten Kanal mit ein. Für den ETM+ Sensor berechnet er sich durch

$$\mathrm{NDVI} = \frac{\mathrm{Kanal4} - \mathrm{Kanal3}}{\mathrm{Kanal4} + \mathrm{Kanal3}}$$

und liefert Werte zwischen -1 und 1. Werte von 0.1 und kleiner deuten auf blanken Boden, Sand, Wasserflächen oder Schnee hin, Werte zwischen 0.2 und 0.3 auf Büsche oder Grasland, sehr hohe Werte (0.6-0.8) auf sehr dichte Vegetation (Weier u. Herring, 2001). Die durch den Nenner ausgedrückte Normalisierung soll den Vergleich räumlich oder zeitlich verschiedener NDVI-Berechnungen erleichtern, da Einflußfaktoren wie Neigung, Neigungsrichtung und Sonnenstand kompensiert werden (Lillesand u. Kiefer, 1994).

Flächeninformationen

Rinder: Wirkt sich die (dauerhafte) Anwesenheit von Rindern negativ auf die lokale Wilddichte aus? Zur Klärung dieser Frage wurden für jeden Kamp des Untersuchungsgebietes zwei weitere Variablen berechnet, jeweils gültig für den gesamten Zeitraum einer Aufnahmephase. Die erste beziffert den Anteil der Tage jeder Aufnahmephase, in dem der betreffende Kamp mit Rindern besetzt war, unabhängig von der Anzahl der Rinder. Die zweite Variable enthält die mittlere Anzahl Rinder, die an jedem Tag der Aufnahmephase in dem jeweiligen Kamp standen. Der Wert für ein Segment wurde als flächengewichteter Mittelwert der für die Kamps berechneten Werte bestimmt, durch die das Segment verläuft. Vermutet wurde, daß infolge dauernder Beunruhigung und Nahrungskonkurrenz – letzteres gilt

im wesentlichen für Hartebeest – das Wild rinderbesetzte Flächen meidet, eine hohe Rinderdichte und dauerhafte Besetzung eines Kamps sich also negativ auf die lokale Wilddichte auswirken. Die berechneten Größen sagen jedoch nichts darüber aus, ob sich in dem betreffenden Segment tatsächlich Rinder befanden. Vielmehr ließ sich während der Aufnahmen eine deutliche Konzentration der Rinder (meist um die Wasserstellen) beobachten.

Habitattyp: Als letzte räumliche Informationen wurden, basierend auf Luftbildern und Farmerinformationen, vier Habitattypen (weitestgehend offenes Grasland, dichter Busch, Berge und Vleys) ausgewiesen und für jedes Segment die Klasse bestimmt, die den größten Flächenteil des Segments ausmachte. Mit Hilfe dieser Variablen sollten die oben erläuterten Habitatpräferenzen der fünf Arten getestet werden.

6. Ergebnisse

Die nachfolgende Darstellung der Ergebnisse erfolgt getrennt nach den Teilprojekten 1 und 2 dieser Arbeit. Mit Hilfe der herkömmlichen Linientransektanalyse und einiger zusätzlicher, in Kapitel 2.1.5 erläuterter Analyseoptionen soll zunächst die Frage des ersten Teilprojektes beantwortet werden, ob bei Verwendung der Farmwege als Transekte verläßliche Dichteschätzungen erzielt werden können. Der zweite Abschnitt befaßt sich mit den Ergebnissen der in Kapitel 2.2 und 2.3 beschriebenen räumlichen Analysen und Vorhersagen, in die die zahlreichen im vorhergehenden Kapitel erläuterten geographischen Informationen eingingen.

In den einzelnen zu erläuternden Arbeitsschritten wird auf die zur Umsetzung der jeweiligen Aufgabe zusätzlich entwickelten und bereits in Kapitel 3 beschriebenen Software-Module Bezug genommen. Wie dort werden SAS-Makros durch ein führendes %-Zeichen, in Python programmierte Geoprocessing-Skripte durch die Endung `.py` gekennzeichnet.

6.1. Validierung der Farmwegzählungen (Teilprojekt 1)

Tabelle 6.1 zeigt die Anzahl der je Aufnahmephase, -schema und Art beobachteten Gruppen und Individuen. Ohne an dieser Stelle bereits die unten näher betrachteten zugrundeliegenden Entdeckungsprozesse und damit die klassische Linientransektanalyse vorwegzunehmen, geben die auf 100 km Transektlänge normierten Anzahlen bereits Hinweise auf Dichteunterschiede zwischen den beiden Aufnahmephasen und zwischen den drei Aufnahmeschemata. Außer für Kudus und Warzenschweine sind die Beobachtungsanzahlen für die Weg-Schemata geringer als die der Buschzählungen. Die ungleichen Beobachtungsanzahlen der drei übrigen Arten sprechen damit bereits für die Hypothese, daß diese Arten die Wege tatsächlich meiden und entsprechend weniger Objekte (Tiergruppen) beobachtet werden können. Eine alternative, aber vermutlich nicht zutreffende, Deutung wäre die unterschiedlicher Entdeckungsprozesse zwischen Weg- und Buschtransekten. Dies würde bedeuten, daß zwar weniger Objekte von den Wegen aus beobachtet werden, gleichzeitig aber die zugehörigen Entdeckungsfunktionen deutlich stärker abfallen, so daß Dichteschätzungen via Distance Sampling letztlich doch ähnliche Werte für Busch- und Wegtransekte annehmen könnten. Dies ist jedoch eher unwahrscheinlich, da sich die Vegetation und sonstigen Umweltbedingungen entlang der Wegtransekte nicht von denen entlang der Buschtransekte unterscheiden. Lediglich die weit einsehbaren Wege selbst könnten deutlich erhöhte Entdeckungswahrscheinlichkeiten aufweisen, die sich dann aber in erhöhten Beobachtungsanzahlen der Wegtransekte niederschlagen

Tabelle 6.1.: Anzahl der je Aufnahmephase, -schema und Art beobachteten Gruppen und Individuen. Angaben in Klammern sind normierte Anzahlen je 100 km Transektlänge.

Phase	Art	Schema	Anzahl Gruppen	Anzahl Individuen
1	Hartebeest	BuschPferd	57 (34)	336 (201)
		WegAuto	24 (13)	161 (86)
		WegPferd	29 (16)	337 (180)
	Kudu	BuschPferd	39 (23)	88 (53)
		WegAuto	33 (18)	104 (56)
		WegPferd	54 (29)	146 (78)
	Oryx	BuschPferd	109 (65)	469 (280)
		WegAuto	37 (20)	196 (105)
		WegPferd	51 (27)	171 (92)
	Steinbock	BuschPferd	36 (22)	38 (23)
		WegAuto	24 (13)	26 (14)
		WegPferd	15 (8)	18 (10)
	Warzenschwein	BuschPferd	35 (21)	70 (42)
		WegAuto	15 (8)	33 (18)
		WegPferd	44 (24)	94 (50)
2	Hartebeest	BuschPferd	31 (28)	99 (89)
		WegAuto	9 (7)	17 (14)
		WegPferd	14 (11)	49 (39)
	Kudu	BuschPferd	24 (22)	45 (40)
		WegAuto	31 (25)	82 (66)
		WegPferd	31 (25)	76 (61)
	Oryx	BuschPferd	51 (46)	158 (142)
		WegAuto	15 (12)	33 (26)
		WegPferd	17 (14)	30 (24)
	Steinbock	BuschPferd	24 (22)	27 (24)
		WegAuto	12 (10)	13 (10)
		WegPferd	10 (8)	10 (8)
	Warzenschwein	BuschPferd	24 (22)	46 (41)
		WegAuto	11 (9)	29 (23)
		WegPferd	27 (22)	60 (48)

müßten – dies ist jedoch deutlich erkennbar nicht der Fall. Auch zählplattformabhängige (Auto/Pferd) Unterschiede lassen sich für die Wegtransekte nur in geringem Umfang erkennen. So könnte vermutet werden, daß das Wild nicht die Wege meidet, sondern infolge des weit hörbaren Autos flieht, bevor es überhaupt entdeckt werden kann. Die Beobachtungsanzahlen des WegPferd-Schemas liegen zwar in fast allen Fällen etwas über denen des WegAuto-Schemas, aber meist (außer Kudu und Warzenschwein) deutlich unter denen des BuschPferd-Schemas. Ein klarer Unterschied zwischen den Zählplattformen ist lediglich für Warzenschweine zu erkennen. Hier liegen die Anzahlen für BuschPferd und WegPferd relativ nahe beieinander, die des WegAuto-Schemas dagegen fallen deutlich ab. Hier könnte demnach die Zählplattform selbst einen Einfluß haben. Die im Vergleich zum WegAuto-Schema erhöhten Beobachtungsanzahlen des WegPferd-Schemas könnten Ihre Ursache aber auch (oder zusätzlich) in der deutlich geringeren Geschwindigkeit des Pferdes und den dadurch länge-

ren Beobachtungszeiten haben. Dies dürfte insbesondere für die farblich sehr gut an den Untergrund angepaßten Warzenschweine gelten. Die gegenüber dem WegPferd-Schema leicht höheren Beobachtungsanzahlen des WegAuto-Schemas für Steinböcke hingegen könnten durch deren Eigenschaft, bei drohender Gefahr aufzuspringen und im Zickzacklauf zu fliehen, begründet sein. Die kleinen und ebenfalls farblich sehr gut angepaßten Steinböcke sind oftmals überhaupt nur so zu entdecken und der „Aufschreckradius“ des Autos dürfte größer als der des Pferdes sein. Die Objektanzahlen der Kudus schließlich liegen für alle drei Schemata etwa im gleichen Bereich, die Individuenanzahlen für die Wegtransekte sind hier sogar etwas größer als die der Buschtransekte. Für Kudus ist also weder ein Einfluß der Wege noch der Zählplattform zu erkennen.

Auffällig sind des weiteren die von der ersten zur zweiten Aufnahmephase deutlich abfallenden Beobachtungsanzahlen aller drei Aufnahmeschemata für Oryx und Hartebeest. Diese sind begründet durch den bereits erwähnten Wildfang, der in der Aufnahmepause im April 2002 durchgeführt wurde und bei dem 60 Oryx und 90 Hartebeest aus dem Farmkomplex entfernt wurden. Es fällt auf, daß – relativ betrachtet – der Rückgang der Individuenanzahlen in fast allen Fällen deutlich stärker als der Rückgang der Gruppenanzahlen ist. Die verbliebenen Tiere haben sich also offensichtlich in der zweiten Aufnahmephase in kleineren Gruppen zusammengefunden. Vor dem Wildfang war der Bestand des gesamten Farmkomplexes in einer Vollzählung per Helikopter mit 210 Oryx und 190 Hartebeest bestimmt worden.

Die weitere Auswertung der Daten mit Hilfe der konventionellen Linientransektanalyse untergliedert sich in die nachfolgend näher erläuterten Schritte:

1. Zusammenführung und Aufbereitung der Daten
2. Berechnung der mittleren Gruppengröße je Aufnahmephase und Art
3. Schätzung der Entdeckungsfunktion je Aufnahmephase und Art und Überprüfung der Möglichkeit einer globalen Schätzung
4. Berechnung der Dichteschätzung je Aufnahmephase, Art und Aufnahmeschema
5. Vergleich der Dichteschätzungen beider Aufnahmephasen je Art und Aufnahmeschema
6. Vergleich der Dichteschätzungen der drei Aufnahmeschemata je Art und Aufnahmephase

Es handelt sich insofern um eine „konventionelle“ Analyse der Linientransekterhebungen, als an dieser Stelle noch keinerlei zusätzliche räumliche Informationen in die Modellierung mit eingingen, wenngleich zur Bestimmung der kürzesten Entfernungen zwischen Objekt und Transekt natürlich GIS-Funktionalität zum Einsatz kam.

Zusammenführung und Aufbereitung der Daten Zunächst waren die Beobachtungs- und Objektpositionen repräsentierende Geometrien innerhalb der

Tabelle 6.2.: Geschätzte mittlere Gruppengrößen ($\bar{s}$) zugehörige Standardfehler (in Klammern) und Anzahl der zur Berechnung verwendeten Beobachtungen ($n_{\bar{s}}$) je Art und Aufnahmephase.

Art	$\bar{s}_1$	$\bar{s}_2$	$n_{\bar{s}_1}$	$n_{\bar{s}_2}$
Hartebeest	5.14 (0.72)	3.06 (0.43)	93	54
Kudu	2.68 (0.18)	2.36 (0.18)	126	86
Oryx	4.24 (0.44)	2.66 (0.25)	197	83
Steinbock	1.09 (0.03)		121	
Warzenschwein	1.98 (0.11)		121	

Geodatenbank zu erstellen, um anschließend GIS-Funktionalität zur Berechnung der kürzesten (senkrechten) Entfernungen zwischen entdeckten Objekten und zugehörigem Transekt zu nutzen. Die Verwendung von GIS-Funktionen war erforderlich, da zumindest das Wegtransekt abschnittsweise nicht exakt gerade verlief und damit senkrechte Entfernungen nicht einfach über $x = y \cdot \sin(\varphi)$ berechnet werden konnten (s. Kap. 2.1.5.1). Darüber hinaus wurden in diesem Arbeitsschritt bereits die im Rahmen der räumlichen Analyse für die Zuordnung einer Beobachtung zu einem Segment erforderlichen M-Werte[1] bestimmt.

Die beschriebenen Aufgaben wurden mit Hilfe des SAS-Makros `%CalcPerpDist` und der daraus gestarteten Python-Skripte `CreatePoints.py`, `CalcObjPos.py` und `CalcPerpDist.py` durchgeführt.

Berechnung der mittleren Gruppengröße je Aufnahmephase und Art Wie bereits aus Tabelle 6.1 erkennbar ist, waren die Beobachtungsanzahlen insbesondere für die beiden Wegschemata teilweise sehr gering. Während die unterschiedlichen Schemata natürlich mit unterschiedlichen Beobachtungsanzahlen und unterschiedlichen Entdeckungsfunktionen einhergehen können, so erscheinen unterschiedliche mittlere Gruppengrößen je Schema jedoch nicht sinnvoll. Gerade infolge der geringen Stichprobenumfänge können diese Unterschiede bei schemaweiser Schätzung jedoch bereits durch einzelne Ausreißer erheblich sein. Folge hiervon könnten – selbst bei gleicher Objektdichteschätzung – stark unterschiedliche Individuendichteschätzungen sein. Um dies auszuschließen, wurde die mittlere Gruppengröße $\bar{s}$ je Aufnahmephase und Art über alle Schemata hinweg als gewöhnliches arithmetisches Mittel bestimmt.

Um hierbei eine mögliche Überschätzung der mittleren Gruppengröße infolge eines zu hohen Anteils großer Gruppen in großen Entfernungen weitestgehend auszuschließen, wurden, mit der größten Entfernung beginnend, so viele Beobachtungen entfernt, bis sich ein linearer Zusammenhang zwischen senkrechter Entfernung und Gruppengröße bei einem Signifikanzniveau von $\alpha = 0.15$ nicht mehr zeigen ließ. Hierbei kam das Makro `%PreDistPlots` unterstützend zum Einsatz. Tabelle 6.2 enthält die so bestimmten mittleren Gruppengrößen, deren Standardfehler sowie die Anzahl der zur Schätzung verwendeten Beobachtungen. Für Steinbock und Warzenschwein waren die Unterschiede zwischen den Gruppengrößen in Aufnahmephase 1 und 2 darüber hinaus so gering, daß diese

[1] Ein sog. M-Wert bezeichnet die (auf der Linie) gemessene reale Entfernung einer Position auf einer Linie zum Startpunkt der Linie (s. Kap. 3.1).

Tabelle 6.3.: Verwendete Entdeckungsfunktionen (Hauptfunktionen: HN: Halfnormal, UN: Uniform, HA: Hazard-rate; Erweiterungsterme: CO: Cosine, PO: Polynomial), Anzahl der geschätzten Modellparameter, Anzahl der zur Schätzung verwendeten Beobachtungen ($n_{\hat{v}}$), $\hat{f}(0)$, geschätzte effektive (halbe) Streifenbreiten ($\hat{v}$, in Metern) und zugehörige Standardfehler (in Klammern) je Art und Aufnahmeschema.

Art	Schema	Model	# Par.	$n_{\hat{v}}$	$\hat{f}(0)$	$\hat{v}$
Hartebeest	BuschPferd	HN	1	84	0.0061	163.83 (13.91)
	WegAuto	HN	1	25	0.0065	153.98 (22.88)
	WegPferd	HN	1	27	0.006	166.35 (26.75)
Kudu	BuschPferd	HN	1	59	0.0079	126.56 (14.34)
	WegAuto	UN+CO	2	58	0.0087	114.71 (19.50)
	WegPferd	HN	1	56	0.0103	97.41 (12.05)
Oryx	BuschPferd	HN	1	149	0.0089	112.44 (6.60)
	WegAuto	HN	1	40	0.0068	146.80 (19.79)
	WegPferd	HN	1	61	0.0056	178.49 (15.71)
Steinbock	BuschPferd	HN	1	56	0.0379	26.38 (2.85)
	WegAuto	HN	1	33	0.0228	43.86 (7.39)
	WegPferd	UN+PO	1	24	0.0327	30.57 (3.28)
Warzenschwein	BuschPferd	HA	2	52	0.0188	53.27 (5.78)
	WegAuto	HA	2	25	0.0145	69.19 (15.64)
	WegPferd	HA	2	51	0.0181	55.36 (7.31)

als über das Jahr hinweg konstant betrachtet und aus den Daten beider Phasen geschätzt wurden. Bekräftigt wird diese Entscheidung durch die Charakteristika der Arten, die keine jahreszeitlichen Schwankungen der Gruppengrößen zeigen (s. Kap. 4.2). Anders verhält es sich bei den übrigen drei Arten; insbesondere bei Oryx und Hartebeest ist ein deutlicher Abfall der mittleren Gruppengröße von Phase 1 zu 2 zu beobachten. Dieser könnte sowohl durch saisonale Schwankungen begründet sein als auch durch den zwischen den Phasen durchgeführten Wildfang, der diese beiden Arten betraf. Für Kudus fällt der Unterschied wiederum deutlich geringer aus. Um jedoch auch kleine saisonale Schwankungen zu berücksichtigen, wurde auch hier nach Phasen getrennt geschätzt. Um etwaige Unterschiede zwischen den mittleren Gruppengrößen statistisch abzusichern, wurde ein Zweistichproben-T-Test durchgeführt. Aufgrund der stark rechtsschiefen Verteilung der Daten wurde zudem eine logarithmische Transformation versucht, die jedoch auch keine genügende Annäherung an eine Normalverteilung lieferte. Sowohl für Original- als auch für transformierte Daten wies der T-Test signifikante ($\alpha = 0.05$) Unterschiede zwischen beiden Aufnahmephasen für Oryx und Hartebeest aus. Die nichtparametrische Alternative, der Wilcoxon-Rangsummentest, wies für keine Art signifikante Unterschiede aus, einhergehend mit sehr ähnlichen Medianen in allen Fällen.

Schätzung der Entdeckungsfunktion je Aufnahmephase und Art und Überprüfung der Möglichkeit einer globalen Schätzung Tabelle 6.3 zeigt die für die weitere Analyse wesentlichen Kennwerte der per `%DistAnaly` und Distance angepaßten Entdeckungsfunktionen. Aufgrund der geringen Stichprobenum-

fänge wurden alle Entdeckungsfunktionen anhand der Entfernungsdaten beider Aufnahmephasen geschätzt, wenngleich lediglich die Buschzählungen für Oryx einen formellen Test auf Gleichheit der Entdeckungsfunktionen zuließen (`%PoolPdfTest`). Sowohl AIC-Werte wie auch der Likelihood-Ratio-Test aus Kapitel 2.1.5.3 sprachen hier knapp für die Schätzung einer globalen Entdeckungsfunktion. Für die übrigen Arten wurde eine konstante Entdeckungsfunktion über alle Aufnahmen eines Schemas hinweg unterstellt, da formelle Tests hier wegen der geringen Beobachtungsanzahlen wenig sinnvoll erschienen. Diese Annahme ist jedoch gewagt, da insbesondere saisonale Unterschiede im Verlauf der Entdeckungsfunktion infolge unterschiedlicher Vegetationsdichten nicht auszuschließen sind. Um die dennoch weiterhin geringe Datenbasis weiter zu vergrößern, hätte eine Zusammenfassung der Daten z. B. für die drei großen Arten Kudu, Oryx und Hartebeest in Betracht gezogen werden können, die ebenfalls in Tabelle 6.3 aufgeführten effektiven Streifenbreiten zeigen jedoch, daß diese drei Arten trotz vergleichbarer Körpergrößen teils deutlich unterschiedliche Entdeckungswahrscheinlichkeiten besitzen und somit eine getrennte Schätzung der Entdeckungsfunktionen vorzuziehen ist. Die in Kapitel 4.2 dargestellten unterschiedlichen Eigenschaften der Arten, insbesondere bezüglich ihres Fluchtverhaltens und des bevorzugten Habitats sprechen ebenfalls für eine getrennte Schätzung der Entdeckungsfunktionen.

Bereits gemessen an der ebenfalls in Tabelle 6.3 dargestellten Anzahl benötigter Modellparameter, verlief die Anpassung der Modelle in fast allen Fällen sehr gut. Dies ist insbesondere bemerkenswert, da die empfohlene Mindestanzahl von 60 bis 80 Beobachtungen für viele Arten und Schemata bei weitem nicht erreicht wurde. Im Falle der Wegzählungen mußte jedoch in fast allen Fällen eine linke Kappungsgrenze w_l gesetzt werden, um den Einfluß von Beobachtungen auszuschalten, die weit voraus direkt auf dem Transekt gemacht wurden. Für fast alle Arten und Schemata zeigte sich das Half-normal-Modell ohne zusätzliche Erweiterungsterme als das am besten geeignete. Lediglich für die Warzenschweine wurde für alle drei Schemata das Hazard-Rate-Modell gewählt, dessen Eigenschaft einer zunächst (nahe $x = 0$) parallel zur x-Achse verlaufenden, dann aber stark abfallenden Entdeckungsfunktion sich mit den beobachteten Entfernungsdaten dieser Art deckte. Der Vergleich der geschätzten effektiven Streifenbreiten zwischen den Arten zeigt erwartungsgemäß kleine Werte für Steinbock, gefolgt von Warzenschweinen und, mit deutlichem Abstand, den drei großen Antilopenarten. Weitere Aussagen lassen sich jedoch nicht gesichert treffen. So könnten infolge der höheren Beobachtungsgeschwindigkeit bei Zählungen per Auto zwar geringe effektive Streifenbreiten vermutet werden, evtl. aufgrund der teils sehr geringen Beobachtungsanzahlen läßt sich dies in den Schätzungen jedoch nicht erkennen, geschweige denn bei den insgesamt großen Streuungen statistisch absichern.

Berechnung der Dichteschätzung je Aufnahmephase, Art und Aufnahmeschema und Vergleich der Dichteschätzungen zwischen Aufnahmephasen und Aufnahmeschemata Gegenüber der gewöhnlichen Schätzung der Objekt- und Individuendichte über (2.4) bzw. (2.5) waren hier zwei Erweiterungen zu

berücksichtigen: Da ein Teil des Wegtransektes direkt an der Farmgrenze verlief (s. Abb. 5.1), die Nachbarfarm jedoch nicht zum Untersuchungsgebiet gehörte und somit dort keine Daten erfaßt wurden, war zu berücksichtigen, daß die reale Stichprobenfläche $\tilde{a}$ kleiner als die in den Gleichungen unterstellte $a = 2 \cdot \hat{v} \cdot L$ war. Mit Hilfe von `%SmplFrac` und `SmplFrac.py` wurden anhand der geschätzten Entdeckungsfunktionen Streifen der Breite $\hat{v}$ um alle Transekte gelegt und der Flächenanteil bestimmt, der innerhalb des tatsächlichen Untersuchungsgebietes lag. Für die Buschtransekte lag dieser Anteil bei über 99 %, für die Wegtransekte deutlich niedriger bei etwa 86 %. Würde dieser Anteil nicht in der Analyse berücksichtigt, so wären Objekt- und Individuendichtschätzungen der Wegschemata systematisch und deutlich zu gering. Das Progamm Distance stellt für solche Fälle die *Sampling-Fraction*-Option zur Verfügung, die alle Dichteschätzungen entsprechend korrigiert, so daß die unterschiedlichen Anteile der tatsächlichen Stichprobenflächen leicht durch einen erneuten Aufruf von `%DistAnaly`, lediglich erweitert um den entsprechenden Parameter, berücksichtigt werden konnten.

Während hiermit die Ergebnisse der Objektdichteschätzungen erreicht waren, waren in den Schätzungen der Individuendichten zusätzlich die oben erwähnten Schätzungen der mittleren Gruppengrößen zu berücksichtigen. Distance bezieht stets nur die Beobachtungen in die Schätzung von $\mathrm{E}(s)$ mit ein, die zum auszuwertenden Stratum (hier Aufnahmephasen) gehören und für die Schätzung der Entdeckungsfunktion verwendet werden, ermöglicht jedoch nicht die Angabe einer externen Schätzung. Wie oben erläutert, bestand daher die Gefahr, daß die Individuendichteschätzungen durch einzelne Ausreißer in den beobachteten Gruppengrößen maßgeblich beeinflußt würden. Dies galt insbesondere für die beiden Arten mit ohnehin sehr variablen Gruppengrößen, Oryx und Hartebeest. Es wurden daher die in Tabelle 6.2 dargestellten, über die Aufnahmeschemata und für Steinbock und Warzenschwein zusätzlich über Aufnahmephasen gepoolten Schätzungen der mittleren Gruppengrößen verwendet, um anhand der von Distance berechneten Objektdichteschätzungen aktualisierte Individuendichteschätzungen zu bestimmen. Die zugehörigen Streuungsmaße und Freiheitsgrade der Individiduendichteschätzungen wurden ebenfalls angepaßt. Durchgeführt wurde die Anpassung der Individuendichteschätzungen und Streuungsmaße mit dem Makro `%ExtCsEst`. Die resultierenden Gruppen- und Individuendichteschätzungen, Schätzungen der Begegnungsraten und der Individuenanzahl je Aufnahmephase, Art und Schema sowie zugehörige 95%ige Konfidenzintervalle sind in Abbildung 6.1 dargestellt. Tabellarische Darstellungen der Ergebnisse enthalten die Tabellen D.1 und D.2 im Anhang.

Zunächst ohne Berücksichtigung der Streuungsmaße der Schätzungen fällt auf, daß sowohl Objekt- wie auch Individuendichteschätzungen der Wegschemata für die Arten Hartebeest, Oryx und Steinbock deutlich (50-80 %) unter denen der Buschtransekte liegen. Für Kudus liegen die Schätzungen aller drei Schemata etwa im selben Bereich, für Warzenschweine gilt dies lediglich für die beiden Pferdeschemata, die Schätzungen des WegAuto-Schemas fallen dagegen weit geringer aus. Eine Erklärung hierfür könnte eine sehr frühe und unbeobachtete Fluchtreaktion der Warzenschweine auf das von weitem hörbare Auto sein. Kudus hingegen verharren sehr lange in Deckung, bevor sie fliehen. Dies und evtl. die Tatsache, daß sie weit seltener gejagt werden als Oryx und Harte-

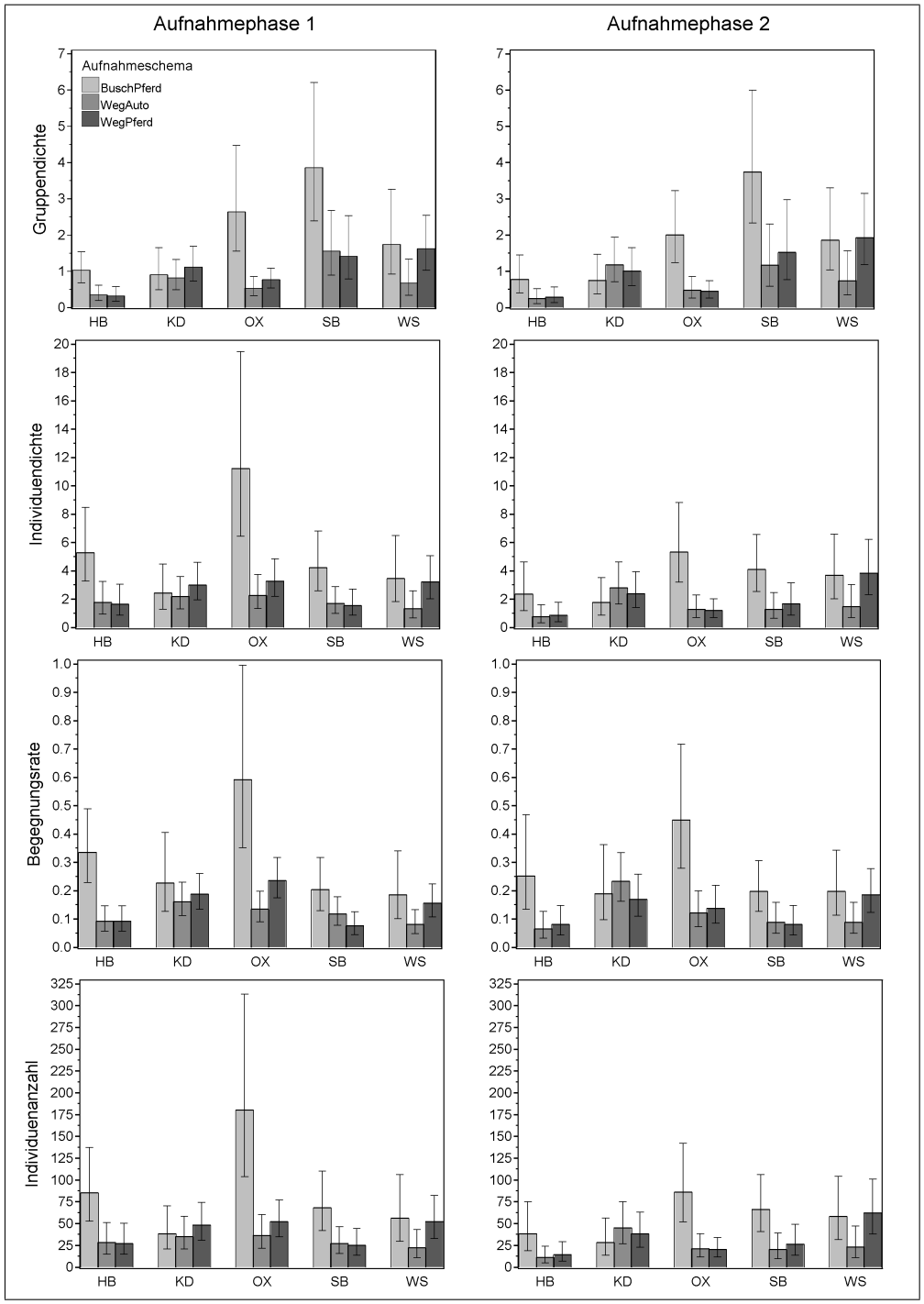

Abbildung 6.1.: Geschätzte Objektdichten, Individuendichten (Anzahl Gruppen/Tiere je Quadratkilometer), Begegnungsraten (Anzahl Gruppen je Kilometer) und Individuenanzahlen sowie zugehörige 95%ige Konfidenzintervalle je Aufnahmephase, -schema und Art (HB = Hartebeest, KD = Kudu, OX = Oryx, SB = Steinbock, WS = Warzenschwein). Die Individuenanzahlen beziehen sich auf das Untersuchungsgebiet von etwa 16.1 km^2 Größe.

beest, könnte Ursache für die sehr ähnlichen Ergebnisse der drei Schemata sein. Oryx und Hartebeest hingegen gelten als sehr wachsame und früh fliehende Tiere (s. Kap. 4.2). Da Jagd zumeist direkt von den Wegen aus stattfindet, ist es möglich, daß diese beiden Arten die Wege ausdrücklich meiden, wodurch die im Vergleich zu den Schätzungen des PferdBusch-Schemas geringen Schätzungen der Wegtransekte erklärt werden könnten. Dies betrifft insbesondere Oryx, die mit Abstand am stärksten jagdlich genutzte Art.

Unklar ist jedoch der insbesondere bei Betrachtung der Gruppendichteschätzungen deutliche Unterschied zwischen dem BuschPferd-Schema und den beiden Wegschemata für die jagdlich eher unbedeutenden Steinböcke. Eine mögliche, aber anhand der vorliegenden Daten nicht validierbare Ursache hierfür könnte die bevorzugte Nutzung von Wegen durch Prädatoren wie Geparden und Leoparden sein.

Die Ergebnisse stützen demnach insgesamt die auf Basis der Beobachtungsanzahlen in Tabelle 6.1 geäußerten Vermutungen, daß auf Wegzählungen basierende Dichteschätzungen der Arten Hartebeest, Oryx und Steinbock trotz der nur geringen Nutzung der Wege negativ verzerrt sind. Für Warzenschweine gilt dies offensichtlich explizit für Zählungen per Auto.

Der Vergleich der Dichteschätzungen beider Aufnahmephasen desselben Schemas zeigt für alle fünf Arten durchaus klar zu erwartende Ergebnisse. Die Dichteschätzungen der Arten Kudu, Steinbock und Warzenschwein liegen für alle drei Schemata in beiden Phasen nahe beieinander, für Oryx und Hartebeest dagegen verzeichnen alle drei einen sehr deutlichen Rückgang der Gruppen- und Individuendichte. Da es sich gerade bei diesen beiden Arten um diejenigen handelt, die die kriechsichere Umzäunung des Farmkomplexes nicht überqueren können, dürfte dieser Rückgang maßgeblich durch den zwischen beiden Aufnahmephasen durchgeführten Wildfang begründet sein. Auffällig ist hierbei, daß, wie oben im Zusammenhang mit Tabelle 6.1 bereits angedeutet, Gruppen- und Individuendichten unterschiedlich stark zurückgehen. Während die Individuendichteschätzungen für beide Arten in der zweiten Aufnahmephase auf etwa die Hälfte der Schätzungen aus der ersten Phase sanken, fielen die Gruppendichteschätzungen um maximal ein Drittel, jedoch einhergehend mit ebenfalls deutlich sinkenden mittleren Gruppengrößen (s. Tab. 6.2). Die im Gebiet verbliebenen Tiere bildeten demnach während der zweiten Aufnahmephase weniger, aber vor allem deutlich kleinere Gruppen. Ob dies jedoch gänzlich auf die Entnahme von Tieren zurückzuführen ist oder ob auch saisonale Trends eine Rolle spielen, läßt sich anhand der Daten nicht ergründen.

Im Zuge des Wildfangs wurden 60 Oryxantilopen und 90 Hartebeest gefangen. Zuvor war der Gesamtbestand des Farmkomplexes per Helikopterzählung mit 210 Oryx und 190 Hartebeest bestimmt worden. Wenngleich die Dichteschätzungen klar einen Rückgang infolge des Wildfangs zeigen, so lassen sich diese Zahlen nur sehr schwer mit den in Abbildung 6.1 dargestellten Ergebnissen vergleichen. Das eigentliche Untersuchungsgebiet macht nur etwa 18 % des gesamten Farmkomplexes aus, gehört jedoch zu dessen wildreicheren Gebieten; dies gilt insbesondere für Oryx. Der südliche Teil des Farmkomplexes dagegen ist nahezu wildfrei. Insofern verwundert es nicht, daß die geschätzte Oryx-Individuenanzahl für das Untersuchungsgebiet in Phase 1 ($\hat{M} = 180$) be-

reits sehr nahe an dem Wert der Vollzählung des gesamten Farmkomplexes liegt, wenngleich das zugehörige, breite Konfidenzintervall auch deutlich geringe Werte enthält. Da während der Helikopterzählung keine Positionen der entdeckten Objekte erfaßt wurden, lassen sich keine präziseren und nur auf das Untersuchungsgebiet bezogene Aussagen treffen.

Die hohen Variationskoeffizienten der Begegnungsraten n/L sind maßgeblich für die Streuung der Gruppen- bzw. Individuendichteschätzungen verantwortlich. Der Anteil der Varianz der Begegnungsrate an der Gesamtvarianz der Individuendichteschätzungen liegt meist über 50 % und reicht bis über 70 %. Die zweite Position nimmt die Varianz von $\hat{f}(0)$ ein, der Anteil der mittleren Gruppengröße dagegen liegt meist lediglich bei etwa 20 %.

Es fällt auf, daß die Varianzen aller in Abbildung 6.1 dargestellten Parameter des WegPferd- und des WegAuto-Schemas denen des BuschPferd-Schemas meist sehr ähnlich und vielfach sogar geringer sind; und dies trotz der geringen Beobachtungsanzahlen der Wegschemata. Wesentliche Ursache hierfür sind die Schätzungen der Varianz der Begegnungsrate. Während die Variationskoeffizienten der $f(0)$-Schätzungen (nicht dargestellt) für die Wegtransekte tatsächlich größer sind als die der Buschtransekte, verhält es sich bei den Variationskoeffizienten der Begegnungsraten umgekehrt. Ursache hierfür ist eine notwendigerweise abweichende Berechnung von $\mathrm{var}(n/L)$ für die Wegschemata. Da es sich bei den zu einer Aufnahmephase gehörenden Zählungen um zeitlich nahe beieinander liegende wiederholte Aufnahmen derselben Transekte handelt, sollten diese in der Analyse nicht als voneinander unabhängig betrachtet werden. In der Folge werden die t Aufnahmen desselben Transektes analysiert, als handele es sich um die Aufnahme eines einzelnen Transektes mit insgesamt n_i Beobachtungen und $t \cdot L$ Gesamtlänge (s. Kap. 2.1.5.3). Dies wirkt sich jedoch einzig auf die Berechnung der Varianz der Begegnungsrate aus, die übrigen Bestandteile der Dichteschätzer und deren Varianzschätzungen bleiben unbeeinflußt. Für die beiden Wegschemata und das verwendete Erhebungsdesign bedeutet dies, daß je Aufnahmephase nur ein (entsprechend längeres) Transekt analysiert werden kann, für das BuschPferd-Schema hingegen liegen je Aufnahmephase elf (wiederholt aufgenommene) Transekte vor. Eine Berechnung der empirischen Varianz der Begegnungsrate nach (2.13) ist für ein einzelnes Transekt jedoch nicht möglich. Das Programm Distance unterstellt daher in diesem Fall eine Poisson-Verteilung der Beobachtungsanzahl, so daß $\mathrm{var}(n) = n$ und $\mathrm{var}(n/L) = n/L^2$. Erfahrungsgemäß wird die wahre Varianz so jedoch (oft deutlich) unterschätzt (Buckland et al., 2001, S. 108), darauf basierende Konfidenzintervalle sind dementsprechend zu eng. Bei der Berechnung der zugehörigen Konfidenzintervalle kommt erschwerend hinzu, daß die Verwendung von Quantilen der t-Verteilung nicht möglich ist, da der Varianzschätzer für n/L keine Freiheitsgrade besitzt. Statt dessen werden Quantile der Standardnormalverteilung verwendet, was zu noch engeren Konfidenzintervallen führt.

Das Problem der Varianzschätzung der Begegnungsrate wirkt sich auch auf die Konfidenzintervalle der Gruppen- und Individuendichte aus. Da die Varianz der Begegnungsrate über die Delta-Regel aus (2.7) in die Varianz der Gruppen- und Individuendichte eingeht, fallen letztere ebenfalls zu gering aus. Deren Freiheitsgrade lassen sich infolge der fehlenden Freiheitsgrade von $\mathrm{var}(n/L)$ darüber

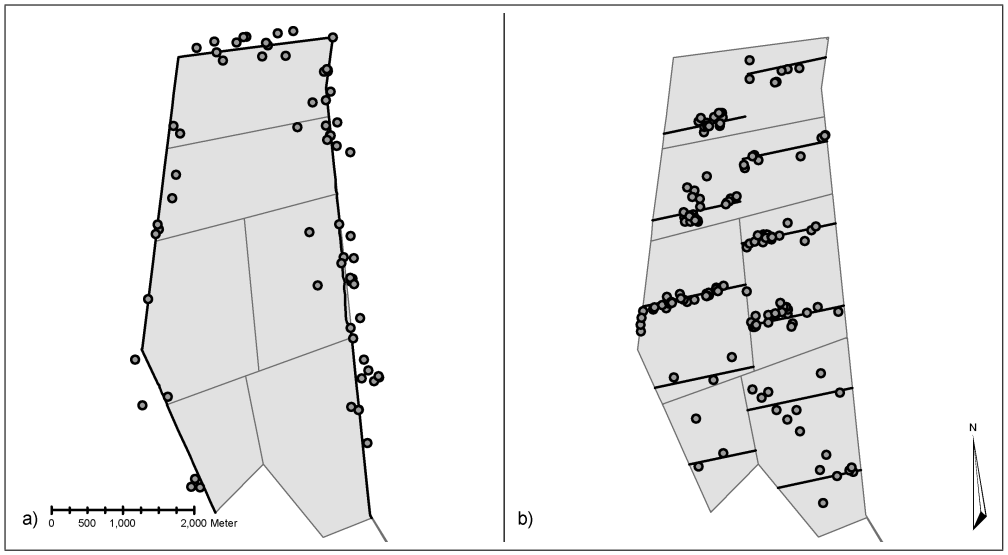

Abbildung 6.2.: Objektpositionen der Oryxbeobachtungen für das WegPferd- (a) und das BuschPferd-Schema (b) aus beiden Aufnahmephasen.

hinaus nicht über die Satterthwaite-Approximation aus (2.11) bestimmen. Bisher (bis Programmversion 4.1) wurde in Distance dennoch die Satterthwaite-Approximation verwendet, jedoch unter Ausschluß der zur Begegnungsrate gehörenden Terme. Auf Basis der resultierenden Freiheitsgrade und der entsprechenden Quantile einer t-Verteilung wurden dann Konfidenzintervalle für $\hat{D}$ und $\hat{D}_s$ bestimmt. In der kommenden Programmversion von Distance – und dies ist in den Abbildung 6.1 zugrundeliegenden Berechnungen bereits umgesetzt (`%ExtCsEst`) – soll für $k = 1$ keine Approximation der Anzahl der Freiheitsgrade von $\hat{D}$ und $\hat{D}_s$ stattfinden. Statt dessen erfolgt die Berechnung ihrer Konfidenzintervalle ebenfalls anhand der Quantile der Standardnormalverteilung (Thomas, pers. Mitlg., 7. 12. 2006).

Um die oben getroffenen Aussagen bezüglich unterschiedlicher Dichteschätzungen der Schemata und Aufnahmephasen statistisch abzusichern, wurde der in Kapitel 2.1.5 vorgestellte und in `%DensTTest` umgesetzte T-Test paarweise für die zu vergleichenden Dichteschätzungen durchgeführt. Um die Anzahl der Freiheitsgrade der Teststatistiken nach (2.15) berechnen zu können, war es erforderlich, Freiheitsgrade für die eingehenden Dichteschätzungen zu berechnen. Wie zuvor erläutert, war dies jedoch für die Wegtransekte nicht möglich. Um die T-Tests dennoch als Anhaltspunkte berechnen zu können, wurde für die Wegtransekte hier die zeitliche Unabhängigkeit unterstellt und auch für sie der empirische Varianzschätzer der Begegnungsrate aus (2.9) verwendet. Abgesehen von der zu vermutenden Unterschätzung der wahren Varianz infolge der Nichtbeachtung der zeitlichen Korrelation, ist hierbei auch zu bedenken, daß die Wegtransekte im Gegensatz zu den Buschtransekten die räumliche Heterogenität der Wilddichte nicht erfassen. Abbildung 6.2 verdeutlicht dieses Problem. In beiden Teilen der Abbildung ist zu erkennen, daß es wildreichere und wildärmere Bereiche innerhalb des Untersuchungsgebietes gibt. In der empirischen Vari-

Tabelle 6.4.: Vergleich der Individuendichteschätzungen beider Aufnahmephasen je Art und Schema sowie p-Werte der T-Tests. p-Werte kleiner 0.05 sind fett hervorgehoben.

Art	Schema	$\hat{D}_1$	$\hat{D}_2$	$\hat{D}_1 - \hat{D}_2$	p-Wert
Hartebeest	BuschPferd	5.27	2.35	2.92	0.180
	WegAuto	1.77	0.74	1.02	**0.035**
	WegPferd	1.64	0.86	0.78	0.070
Kudu	BuschPferd	2.41	1.76	0.65	0.445
	WegAuto	2.18	2.78	-0.60	0.538
	WegPferd	2.99	2.36	0.62	0.567
Oryx	BuschPferd	11.20	5.32	5.87	0.578
	WegAuto	2.25	1.27	0.98	0.124
	WegPferd	3.26	1.19	2.07	0.081
Steinbock	BuschPferd	4.20	4.08	0.12	0.945
	WegAuto	1.69	1.27	0.42	0.229
	WegPferd	1.54	1.65	-0.11	0.848
Warzenschwein	BuschPferd	3.44	3.66	-0.22	0.918
	WegAuto	1.33	1.46	-0.13	0.804
	WegPferd	3.20	3.81	-0.61	0.810

anzschätzung der Buschtransekte wird dies durch Transekte mit niedrigen und solche mit hohen Begegnungsraten berücksichtigt. Bei einem einzelnen Transekt ist dies jedoch nicht möglich. Anhand der Begegnungsrate dieses einzelnen Transektes läßt sich nicht erkennen, ob es Bereiche hoher oder niedriger Wilddichten gibt; die Beobachtungen könnten alle direkt am Anfang des Transektes gemacht worden sein oder sich gleichmäßig darüber verteilen. Die offenkundig starke räumliche Heterogenität ist jedoch ein Argument für den Einbezug zusätzlicher raumbezogener Informationen in die Dichteschätzungen (Teilprojekt 2). Gelingt es, lokal oder auch großflächig schwankende Wilddichten durch zusätzliche geographische Informationen zu erklären, so sollten sich auch die Streuungen der Dichteschätzungen deutlich reduzieren lassen.

Unter dem genannten und einem weiteren, unten erläuterten Vorbehalt sind die in Tabelle 6.4 aufgeführten Ergebnisse der T-Tests zum Vergleich der Dichteschätzungen innerhalb eines Schemas zwischen den Aufnahmephasen zu interpretieren. Die Dichteschätzungen für Kudu, Steinbock und Warzenschwein in Abbildung 6.1 ließen vermuten, daß es keine großen Veränderungen zwischen beiden Aufnahmephasen gegeben hat; dies wird durch die Ergebnisse der T-Tests bestätigt. Für Oryx und Hartebeest hingegen waren infolge des Wildfangs deutliche Unterschiede meßbar, die jedoch nur in einem einzigen Fall signifikant (zu einem Fehlerniveau von $\alpha = 0.05$) sind. Bei der Interpretation der p-Werte ist jedoch zu bedenken, daß diese keine Korrektur für multiple Vergleiche enthalten, also die versuchsbezogene Irrtumswahrscheinlichkeit nicht einhalten. Zudem unterstellt der T-Test eine Normalverteilung der Dichteschätzungen, die jedoch im allgemeinen nicht gegeben ist (s. Kap. 2.1.4). Die Tests führen hier also tendenziell zu anti-konservativen Entscheidungen. Ursache der dennoch hohen p-Werte sind die hohen Varianzen der Dichteschätzungen, wobei nach den obigen Anmerkungen durchaus davon auszugehen ist, daß die Varianzen der Schätzungen

Tabelle 6.5.: p-Werte der T-Tests zum Vergleich der Individuendichteschätzungen der drei Aufnahmeschemata je Aufnahmephase und Art. p-Werte kleiner 0.05 sind fett hervorgehoben.

Phase	Art	Schema	BuschPferd	WegAuto	WegPferd
1	Hartebeest	BuschPferd	.	0.069	0.052
		WegAuto	0.069	.	0.829
		WegPferd	0.052	0.829	.
	Kudu	BuschPferd	.	0.800	0.645
		WegAuto	0.800	.	0.486
		WegPferd	0.645	0.486	.
	Oryx	BuschPferd	.	0.346	0.427
		WegAuto	0.346	.	0.500
		WegPferd	0.427	0.500	.
	Steinbock	BuschPferd	.	**0.037**	**0.032**
		WegAuto	**0.037**	.	0.742
		WegPferd	**0.032**	0.742	.
	Warzenschwein	BuschPferd	.	0.120	0.899
		WegAuto	0.120	.	0.070
		WegPferd	0.899	0.070	.
2	Hartebeest	BuschPferd	.	**0.039**	0.055
		WegAuto	**0.039**	.	0.688
		WegPferd	0.055	0.688	.
	Kudu	BuschPferd	.	0.253	0.385
		WegAuto	0.253	.	0.646
		WegPferd	0.385	0.646	.
	Oryx	BuschPferd	.	**0.036**	**0.030**
		WegAuto	**0.036**	.	0.755
		WegPferd	**0.030**	0.755	.
	Steinbock	BuschPferd	.	**0.010**	0.053
		WegAuto	**0.010**	.	0.422
		WegPferd	0.053	0.422	.
	Warzenschwein	BuschPferd	.	0.120	0.958
		WegAuto	0.120	.	0.264
		WegPferd	0.958	0.264	.

beider Wegschemata noch unterschätzt werden.

Die gleichen Vorbehalte gelten auch für die in Tabelle 6.5 aufgeführten p-Werte. Die Tabelle enthält die Ergebnisse der T-Tests für die Vergleiche zwischen den Aufnahmeschemata. Auch hier lassen sich nur wenige der Aussagen von oben absichern. So weisen die Daten der umfangreicheren ersten Aufnahmephase nur für Steinböcke einen signifikanten Unterschied zwischen Weg- und Buschzählungen aus. Für die zweite Aufnahmephase lassen sich zumindest für Hartebeest, Oryx und Steinbock Unterschiede zwischen den Schemata BuschPferd und WegAuto nachweisen (für Oryx zusätzlich auch zu WegPferd). Den Ergebnissen von oben (s. Abb. 6.1) nach war für Kudus in beiden Aufnahmephasen kein wesentlicher Unterschied zwischen den drei Aufnahmeschemata auszumachen. Ebenso konnte kein auffälliger Unterschied der Warzenschweindichteschätzungen zwischen dem BuschPferd- und dem WegPferd-Schema festgestellt werden. Für beide Arten bestätigen die Ergebnisse der T-Tests diese

Beobachtungen. Der vermutete autospezifische Effekt auf die Dichteschätzungen der Warzenschweine läßt sich jedoch ebenfalls nicht absichern.

Fazit Wenngleich nur sehr begrenzt statistisch abgesichert, lassen die oben aufgeführten Ergebnisse doch die Schlußfolgerung zu, daß die Dichteschätzungen der Zählungen per Auto – außer für Kudu – vermutlich deutlich negativ verzerrt sind. Die nach gleichem Schema von Farmern durchgeführten Zählungen liefern für die übrigen vier Arten demnach zu geringe Dichteschätzungen. Werden anhand dieser Zahlen Management- und Jagdpläne erstellt, so dürfte eine Unterschätzung zwar weit weniger gravierende Folgen als eine Überschätzung der Bestände haben, dennoch geben die (vermuteten) Dimensionen der Verzerrung Anlaß dazu, das Stichprobendesign und damit die Wahl der Transekte noch einmal zu überdenken, trotz der mit anderen (zufälligen) Designs verbundenen Schwierigkeiten.

6.2. Räumliche Modellierung (Teilprojekt 2)

Die zur Umsetzung der Ziele des zweiten Teilprojektes, des Einbezugs räumlicher Kovariaten und der Berechnung räumlicher Vorhersagen, erforderlichen Arbeitsschritte lassen sich wie folgt zusammenfassen:

1. Erstellung der Segmente und Bestimmung der Objekt- und Individuenanzahl je Segment sowie der räumlichen Kovariaten.

2. Auswahl der abhängigen Variablen (Anzahl der Objekte oder Anzahl der Individuen je Segment), der zu verwendenden Segmentlänge und geeigneter Kovariaten.

3. Parametrisierung der gewählten Modelle unter der Annahme unabhängiger und räumlich autokorrelierter Segmente.

4. Berechnung räumlicher Vorhersagen und der Gesamtabundanz.

Die Ergebnisse der Validierung der Farmwegzählungen oben ließen eine Verzerrung der Dichteschätzungen durch die Verwendung von Wegtransekten vermuten. Die vier Arbeitsschritte sollen daher nur für die Daten des BuschPferd-Schemas etwas detaillierter betrachtet werden. Sollte sich hier ein Einfluß von Wegen, Tränken oder Salzlecken zeigen lassen, so spräche dies ebenfalls gegen die Verwendung von Wegen als Transekte. Besonderes Augenmerk soll jedoch nicht auf den Einzelheiten der Analysen der fünf untersuchten Arten liegen, sondern auf allgemeinen Aspekten, Problemen und Auffälligkeiten, die im Zuge der Auswertung beobachtet werden konnten.

Erstellung der Segmente und Bestimmung der Objekt- und Individuenanzahl je Segment sowie der räumlichen Kovariaten: Für die Breite der zu erstellenden Segmente wurden die im ersten Teilprojekt bestimmten und in Tabelle 6.3 aufgeführten effektiven Streifenbreiten $\hat{v}$ verwendet. Da es für die Wahl

Tabelle 6.6.: Segmentlängen (in Metern) und Anzahl resultierender Segmente.

Segmentlänge c	Anzahl
100	145
200	75
300	51
400	40
500	32
600	29
700	25
800	22
900	22
1000	21

der Segmentlänge c keinerlei Erfahrungswerte gab, sondern lediglich die Richtlinie, daß die geographischen Informationen innerhalb eines Segments möglichst homogen sein sollten, wurden alle Berechnungen zunächst für die Segmentlängen 100 m bis 1 000 m in 100 m-Intervallen durchgeführt. Tabelle 6.6 enthält die Anzahl der für die zehn Segmentlängen resultierenden Segmente. Da in beiden Aufnahmephasen dasselbe Stichprobendesign verwendet wurde, ist die Anzahl der Segmente für beide Phasen identisch.

Die Generierung der Segmentgeometrien erfolgte mit Hilfe des `%CreateSegments`-Makros, die anschließende räumliche Verschneidung mit den in Kapitel 5.2 aufgeführten zusätzlichen geographischen Informationen mit `%SpatialQueryPolys`. Beide Makros und die daraus aufgerufenen Geoprocessing-Skripte sind in Kapitel 3.2.1 näher erläutert.

Auswahl der abhängigen Variablen, der zu verwendenden Segmentlänge und geeigneter Kovariaten: Der umfangreichste Arbeitsschritt bestand in der Auswahl der für die Berechnung der Vorhersagen zu verwendenden Kovariaten und zudem in der Entscheidung darüber, ob die Anzahl der Objekte (Tiergruppen) n oder der Individuen m als zu modellierende Größe verwendet werden sollten. Da die Auswahl der Kovariaten sehr groß war, wurde unter der Annahme unabhängiger Beobachtungen für beide abhängigen Variablen und alle Segmentlängen zunächst eine automatische, schrittweise Variablenselektion durchgeführt. Zur Beurteilung des Beitrags, den ein Effekt zur Erklärung der beobachteten Streuung liefert, wurde die Wald-Statistik des zugehörigen Typ-3-Kontrastes aus (2.41) verwendet. Schrittweise wurden dem Modell die Kovariaten hinzugefügt, die zusätzlich zu den bereits enthaltenen als signifikant ausgewiesen wurden. Das hierbei angesetzte Signifikanzniveau wurde gezielt höher als üblich bei $\alpha = 0.15$ angesetzt, um für jede Segmentlänge einen Überblick über die wichtigsten erklärenden Variablen und nicht nur die jeweils beste oder wenigen besten zu erhalten. Für die weitere Modellierung wurden dann maßgeblich die Kovariaten verwendet, die für die zehn Segmentlängen am häufigsten ausgewählt wurden. Da, wie unten erläutert werden wird, für die abschließende Parametrisierung recht große Segmentlängen von 600 m oder 700 m verwendet werden mußten, und mit diesen nur geringe Segmentanzahlen einhergingen (s. Tab. 6.6), sollten

nicht mehr als zwei oder maximal drei Kovariaten ausgewählt werden. In fast allen Fällen ergab die automatische Variablenselektion ohnehin höchstens zwei klar präferierte Kovariaten, so daß diese Beschränkung leicht einzuhalten war.

In der Regel waren die Kovariaten, die sich für die Modellierung der Objektanzahl n ergaben, denen für die Individuenanzahl m sehr ähnlich, so daß eine Entscheidung über die zu verwendende Abhängige meist anhand der unten erläuterten Residuenplots fiel. Eine deutliche Ausnahme hiervon zeigten die Hartebeest-Daten der ersten Aufnahmephase. Während die automatische Variablenselektion für die Objektanzahl keine Kovariate klar präferierte, wurden für die Individuenanzahl gleich drei Variablen für die Mehrzahl der Segmentlängen ausgewählt. Ursache hierfür dürfte die zwar recht homogene Verteilung der Objekte, aber offensichtliche Konzentration größerer Gruppen in bestimmten Gebieten sein (s. Abb. 6.3). In Phase 2 waren die beobachteten Gruppen vermutlich infolge des Wildfangs deutlich kleiner, so daß sich hier kein so klarer Unterschied zeigte.

Wechselwirkungen zwischen den Kovariaten wurden bewußt nicht getestet, da dies die Modellformulierungen deutlich komplexer hätte werden lassen, verläßliche Aussagen aufgrund der geringen Beobachtungsanzahlen jedoch nicht zu erwarten waren. Neben der automatischen Selektion wurde der Einfluß von Kovariaten auch mit Hilfe von Graphiken überprüft, in denen der Wert der Kovariaten der (transformierten) Individuen- oder Objektanzahl gegenübergestellt wurde (näheres s. Kap. 3.2.2).

Der Kovariatenauswahl folgte die Wahl einer Segmentlänge, anhand deren konkreter Parameterschätzungen dann die Vorhersagen berechnet wurden. Prinzipiell sind zwar möglichst kurze Segmentlängen und damit eine hohe räumliche Auflösung zu bevorzugen, die Wahl der Segmentlänge ist jedoch dadurch nach unten begrenzt, daß ein großer Anteil der Segmente ohne Beobachtungen ($n = m = 0$) die Modellanpassung negativ beeinträchtigt. Je kürzer die Segmentlänge gewählt wird, desto größer ist dieser Anteil und desto stärker werden die Individuen- oder Objektanzahlen in Segmenten mit Beobachtungen unterschätzt. Die Ursache hierfür liegt zum einen darin, daß im gewöhnlichen log-linearen Modell (mit unabhängigen Beobachtungen) keine negativen Vorhersagen möglich sind. Die Vorhersage für ein Segment ohne Beobachtungen wird daher immer größer als Null sein. Zum anderen ist die Summe der Residuen in log-linearen Modellen, wie in klassischen linearen Modellen, gleich Null. Gibt es also viele Segmente ohne Beobachtungen, deren Objekt- oder Individuenanzahl nur überschätzt werden kann, so müssen die Objekt- oder Individuenanzahlen der Segmente mit Beobachtungen zwangsweise sehr oft unterschätzt werden. Dieser Effekt ist umso stärker, je größer der Anteil der Segmente ohne Beobachtungen ist. Einige methodische Möglichkeiten, mit diesem allgemein unter dem Begriff *Zero Inflated Poisson Process* (ZIP) bekannten Problem umzugehen, werden im Zuge der Diskussion in Kapitel 7 angesprochen. Eine einfache Möglichkeit ist, die Segmentlänge soweit zu erhöhen, bis der Anteil der Segmente ohne Beobachtungen genügend klein ist. Zur Beurteilung, welche Segmentlänge ausreichend ist, kann u. a. ein Plot herangezogen werden, der beobachtete Anzahlen den vom Modell vorhergesagten gegenüberstellt. Die Graphik sollte möglichst gleichmäßig um die Winkelhalbierende streuende Punkte zeigen.

Abweichungen hiervon können aber auch Anzeichen für fehlende oder falsch skalierte Kovariaten oder für die Wahl der falschen Link-Funktion sein.

Neben dieser Graphik wurden zahlreiche weitere als Entscheidungshilfen für die endgültige Wahl der Kovariaten, einer Segmentlänge und ggf. der abhängigen Variablen herangezogen. Hierzu gehörten insbesondere solche, die die Streuung und Verteilung der verschiedenen Residuenarten darstellen (s. Kap. 3.2.2).

Die jeweils verwendete abhängige Variable, die ausgewählten Kovariaten und die verwendete Segmentlänge je Aufnahmephase und Wildart sind in Tabelle 6.7 aufgeführt. Eine kurze Beschreibung der abgekürzten Kovariatenbezeichnungen enthält Tabelle 6.8. Mit einer Ausnahme (Kudu, Phase 2) fiel die Wahl auf die Daten der 600 m langen Segmente. Die Auswahl der Kovariaten erfolgte ausschließlich anhand von Ergebnissen der Modelle unter der Annahme räumlicher Unabhängigkeit. Für Modelle, die die räumliche Autokorrelation zwischen den Segmenten berücksichtigen, wurde ebenfalls eine automatische Variablenselektion versucht, die jedoch zu oft an Konvergenzproblemen scheiterte. In den geglückten Versuchen konnte jedoch beobachtet werden, daß die Ergebnisse denen der entsprechenden Modelle mit unabhängigen Beobachtungen stets sehr ähnlich waren.

Wie aus Tabelle 6.7 zu ersehen ist, kam nur eine sehr begrenzte Auswahl der zur Verfügung stehenden geographischen Informationen zum Einsatz. Wesentliche Ursache hierfür dürften die geringen Beobachtungsanzahlen und damit verbunden die geringe Anzahl zu verwendender Segmente sein, die oft nur die Ausweisung einer einzelnen einflußreichen Kovariate zuließ. Denkbar ist jedoch auch, daß wesentliche erklärende geographischen Informationen nicht oder – im Falle der aus Satellitenbildern generierten Informationen – für den falschen Zeitpunkt erfaßt wurden. Da die Daten beider Aufnahmephasen über einen längeren Zeitraum erhoben wurden, ist zudem fraglich, ob die unterstellte Konstanz der Bedingungen und der Wildverteilung über den jeweiligen Zeitraum gerechtfertigt ist. Eine Analyse der Daten der 20 einzelnen Zählungen kam wegen der geringen Beobachtungsanzahlen und der damit verbundenen Probleme jedoch nicht in Betracht.

Für die Umsetzung der beschriebenen Abläufe wurden mehrere SAS-Makros verwendet: Mit `%GenmodAutoSelect` wurde die automatische Variablenselektion unter Annahme der Unabhängigkeit der Segmente durchgeführt, `%GlimmixAutoSelect` ist das Pendant für räumlich autokorrelierte Beobachtungen. Die visuelle Beurteilung der Modelle erfolgte mit den Makros `%GlmEda` und `%GlmResPlot`. `%GlmEda` stellt die (transformierte) abhängige Variable den Werten potentiell erklärender Kovariaten graphisch gegenüber und ermöglicht damit die Überprüfung eines (linearen) Zusammenhangs. `%GlmResPlot` erstellt zahlreiche Graphiken, vornehmlich Residuenplots, die zur Beurteilung der Modellanpassung sowie der Wahl der Varianz- und Link-Funktion herangezogen werden können (näheres s. Kap. 3.2.2).

Parametrisierung der gewählten Modelle für unabhängige und räumlich autokorrelierte Segmente: Tabelle 6.7 enthält auch die anhand der Daten der angegebenen Segmentlänge geschätzten Parameter der ausgewählten Kovaria-

Tabelle 6.7.: In den Modellen verwendete abhängige Variablen (NIndiv: Anzahl der Individuen m, NObj: Anzahl der Objekte/Tiergruppen n), Kovariaten und Segmentlängen (c) sowie Parameterschätzungen ($\hat{\boldsymbol{\beta}}$) und zugehörige Standardfehler (in Klammern) und p-Werte des Modells mit unabhängigen Beobachtungen ($IGLM$) und mit Einbezug räumlicher Autokorrelation ($SpGLM$) je Aufnahmephase und Art. p-Werte kleiner 0.05 sind fett hervorgehoben.

Phase	Art	Abh. Var.	c	Kovariate	$\hat{\boldsymbol{\beta}}_{IGLM}$	p-Wert$_{IGLM}$	$\hat{\boldsymbol{\beta}}_{SpGLM}$	p-Wert$_{SpGLM}$
1	Hartebeest	NIndiv	600	DistVleys	-0.45 (0.35)	0.206	-0.41 (0.34)	0.247
				bright	0.18 (0.05)	**<0.001**	0.17 (0.04)	**0.001**
				green	0.15 (0.07)	**0.025**	0.15 (0.06)	**0.026**
	Kudu	NObj	600	CatDayAnt	-3.46 (1.60)	**0.030**	-3.55 (2.13)	0.109
				DistVleys	0.48 (0.27)	0.072	0.37 (0.33)	0.275
	Oryx	NIndiv	600	DistVleys	1.20 (0.22)	**<0.001**	1.18 (0.22)	**<0.001**
	Steinbock	NObj	600	DistFarmy	0.57 (0.31)	0.065	0.55 (0.31)	0.086
				alt	-0.11 (0.04)	**0.009**	-0.10 (0.04)	**0.018**
	Warzenschwein	NIndiv	600	DistPath	-5.62 (2.05)	**0.006**	-5.05 (1.76)	**0.008**
				x*y	0.02 (0.01)	**0.011**	0.02 (0.01)	**0.016**
2	Hartebeest	NObj	600	DistPath	4.10 (2.07)	**0.048**	3.67 (2.11)	0.095
	Kudu	NIndiv	700	bright	0.13 (0.04)	**0.002**	0.12 (0.04)	**0.004**
	Oryx	NObj	600	DistRivier	0.56 (0.36)	0.12	0.57 (0.36)	0.123
				DistWater	0.91 (0.35)	**0.010**	0.92 (0.36)	**0.017**
	Steinbock	NObj	600	DistPath	2.16 (1.47)	0.143	2.12 (1.47)	0.16
	Warzenschwein	NObj	600	DistVleys	-1.20 (0.31)	**<0.001**	-1.21 (0.34)	**0.002**

Tabelle 6.8.: Erläuterung der in Tabelle 6.7 verwendeten Abkürzungen der Kovariatenbezeichnungen. Für Detailinformationen s. Kapitel 5.2.

Abkürzung	Kurzbeschreibung
alt	Geländehöhe über N. N. an der Stelle des Zellschwerpunktes
bright	Brightness-Wert an der Stelle des Zellschwerpunktes
CatDayAnt	Mittlerer Anteil der Tage an dem die Kamps, in denen die Zelle liegt, mit Rindern besetzt waren
DistFarmy	Entfernung zum nächstgelegenen Farmhof
DistPath	Entfernung zum nächstgelegenen Farmweg
DistRivier	Entfernung zum nächstgelegenen Rivier
DistVleys	Entfernung zum nächstgelegenen Vley
DistWater	Entfernung zur nächstgelegenen Wasserstelle
green	Greenness-Wert an der Stelle des Zellschwerpunktes
x*y	Produkt aus x- und y-Koordinate des Zellschwerpunktes

ten für das Modell mit unabhängigen Beobachtungen ($IGLM$) und das unter Berücksichtigung räumlicher Autokorrelation ($SpGLM$) sowie die zugehörigen Standardfehler. Bei den aufgeführten p-Werten handelt es sich um die der zu den Parametern gehörenden Typ-3-Kontraste. Da die Auswahl der Kovariaten nicht anhand der konkreten Segmentlänge sondern anhand der Modelle für alle Segmentlängen getroffen wurde, kommen p-Werte vor, die deutlich größer als das gängige Signifikanzniveau von $\alpha = 0.05$ sind.

Die Schätzung der Parameter der festen Effekte erfolgte für $IGLM$ per Quasi-Likelihood, da in alle Modelle ein zusätzlicher Überdispersionsparameter aufgenommen wurde, für $SpGLM$ per Restricted-Pseudo-Likelihood. Wie in der Fachliteratur beschrieben, sind sich auch hier die Parameterschätzungen beider Modellansätze durchweg sehr ähnlich, wenngleich im Laufe der Modellselektion in einigen wenigen Fällen deutliche Schwankungen der Parameterschätzungen in Abhängigkeit von den Werten der Kovarianzparameter beobachtet werden konnten. Ebenfalls erwartungsgemäß sind die Standardfehler der Parameterschätzungen aus $SpGLM$ in der Regel etwas größer, insgesamt aber im Vergleich zu den Parameterschätzungen selbst recht groß – ein Ausdruck der Modellunsicherheit. Zu bedenken ist zudem, daß die (modellbasierten) Standardfehler der $SpGLM$-Modelle die Kenntnis der wahren Kovarianzfunktion und -parameter unterstellen. Da diese ebenfalls gutachterlich oder per Maximum-Likelihood der Pseudo-Daten geschätzt wurden, kann hiervon sicherlich nicht ausgegangen werden. Die angegebenen Standardfehler für $SpGLM$ dürften also die wahren unterschätzen. Dies gilt entsprechend umso mehr für die Standardfehler der $IGLM$-Schätzungen, die keinerlei Korrelation zwischen den Segmenten berücksichtigen.

Wie ein Blick auf Tabelle 6.7 schnell zeigt, ließen sich die im Zuge der Auflistung der zusätzlichen geographischen Informationen formulierten Hypothesen nur sehr begrenzt bestätigen oder widerlegen. Gleiches gilt für artspezifische Hypothesen, die sich aus den Eigenschaften und Habitatpräferenzen der fünf Untersuchungsarten herleiten ließen. Die Begründung hierfür liegt in erster Linie in den geringen Beobachtungs- und Segmentanzahlen und den damit verbundenen großen Standardfehlern sowie evtl. in wichtigen aber fehlenden Kovariaten. Es

soll hier lediglich auf drei auffällige Aspekte gesondert eingegangen werden.

Die Ergebnisse für Oryx zeigen, daß zwar vermutete einflußreiche Kovariaten (DistRivier, DistVleys) ausgewählt wurden, die Parameterschätzungen jedoch das falsche Vorzeichen besitzen. Die Oryxdichten sind demnach in der Nähe von Rivieren und Vleys nicht besonders hoch sondern besonders niedrig. Abgesehen von der Möglichkeit, daß diese Variablen einen zufällig beobachteten, aber nicht real existenten Einfluß auf die Oryxdichte besitzen – dies gilt gleichermaßen für alle ausgewählten Kovariaten und wird gestützt durch die geringen Stichprobenumfänge größerer Segmentlängen –, könnten die Variablen auch Ersatz für eine tatsächlich vorhandene, jedoch nicht erfaßte Einflußgröße wie Jagddruck sein. Vleys und Riviere sind gut einsehbar, Wasserstellen sehr oft von Menschen frequentiert. Zudem konnte beobachtet werden, daß sich Oryxantilopen auch sehr oft in stark verbuschten Gebieten aufhalten. Dies alles spricht dafür, daß sich zumindest in diesem konkreten Untersuchungsgebiet Oryxantilopen eher in verbuschte Bereiche zurückziehen. Die entsprechende Kovariate (Vegetationstyp) wurde jedoch nicht ausgewählt.

Insgesamt werden für alle Arten außer Kudu in wenigstens einer der Phasen Kovariaten ausgewählt, die in enger Verbindung zu menschlicher Aktivität stehen. Dies ist entweder die Entfernung zum Farmhof, zur nächsten Wasserstelle oder – am häufigsten – zu den Farmwegen. Die Auswahl der (an den Farmwegen gelegenen) Wasserstellen und insbesondere der Farmwege selbst stützt die Aussagen aus dem ersten Teilprojekt und spricht gegen die Repräsentativität der Farmwege und damit gegen deren Verwendung als Transekte. Lediglich für Kudu lassen die ausgewählten Kovariaten keine solche Schlußfolgerung zu; auch dies deckt sich mit den Ergebnissen aus Teilprojekt 1.

Der dritte und letzte Aspekt betrifft die Parameterschätzung des Effektes DistPath für Warzenschwein in Aufnahmephase 1. Der im Vergleich zu den übrigen Parameterschätzungen mit Abstand kleinste und daher auffällige Wert spricht für einen starken anziehenden Effekt der Farmwege. Dieser ließe sich durch die bevorzugte Verwendung der Farmwege erklären, da hier keine störende Vegetation vorhanden ist. Gerade für die robusten Warzenschweine erscheint diese Erklärung jedoch etwas fragwürdig. Wahrscheinlicher ist daher ein verdeckter Effekt der tatsächlichen Ursache, evtl. der Wasserstellen.

Tabelle 6.9 enthält die geschätzten Kovarianzparameter der einzelnen Modelle aus Tabelle 6.7. Der Überdispersionsparameter σ_o^2 wurde für *IGLM* und *SpGLM* anhand der Pearson-Statistik geschätzt. Wie für die Parameterschätzungen der festen Effekte liefern beide Ansätze auch ähnliche Ergebnisse für $\hat{\sigma}_o^2$. Darüber hinaus zeigt sich – mit einer Ausnahme (Hartebeest, Phase 2) – erwartungsgemäß, daß die geschätzten Überdispersionsparameter der Analysen, in denen die Individuenanzahl als abhängige Variable verwendet wurde, deutlich größer ausfallen als in den Analysen der Objektanzahl. Ursache hierfür ist die teils erhebliche Streuung der Gruppengrößen (insbesondere Hartebeest, Phase 1). Während die meisten Analysen eine Überdispersion der untersuchten Größe ergaben, zeigte sich für Steinbock (Phase 1 und 2) und Warzenschwein (Phase 2) sogar eine leichte Unterdispersion. Die Klumpung der Gruppen bzw. Individuen dieser beiden Arten war also geringfügig kleiner als unter dem Poisson-Modell erwartet.

Tabelle 6.9.: Geschätzte Kovarianzparameter der Modelle mit und ohne Berücksichtigung räumlicher Autokorrelation je Aufnahmephase und Art. $\hat{\sigma}_o^2(IGLM)$ ist die Schätzung des Überdispersionsparameters des Modells für unabhängige Beobachtungen, $\hat{\sigma}_o^2(SpGLM)$ und $\widehat{\text{range}}(SpGLM)$ sind die Schätzungen für den Überdispersionsparameter und die Reichweite der Kovarianzfunktion des Modells mit räumlicher Autokorrelation. Die mit * markierten Werte sind aus den empirischen Semivariogrammen der Pearson-Residuen gutachterlich bestimmte Reichweiten.

Phase	Art	$\hat{\sigma}_o^2(IGLM)$	$\hat{\sigma}_o^2(SpGLM)$	$\widehat{\text{range}}(SpGLM)$
1	Hartebeest	8.47	7.99	741.8
	Kudu	1.25	1.39	1023.5
	Oryx	5.81	5.68	636.5
	Steinbock	0.86	0.81	732.3
	Warzenschwein	2.89	2.81	*800.0
2	Hartebeest	2.62	2.76	*700.0
	Kudu	2.13	2.17	1110.5
	Oryx	1.42	1.38	588.6
	Steinbock	0.85	0.84	479.5
	Warzenschwein	0.83	0.87	777.9

Eine Abschätzung der räumlichen Kovarianzparameter Schwellenwert (für $SpGLM$ gleichbedeutend mit $\sigma_o^2(SpGLM)$) und Reichweite (range($SpGLM$)) erfolgte zunächst gutachterlich anhand empirischer Semivariogramme der Pearson-Residuen der zugehörigen Modelle ohne Autokorrelation. Bedingt durch die relativ hohen Segmentlängen und geringen Stichprobenumfänge, konnten die Semivariogramme jedoch meist nur sehr grobe Anhaltspunkte liefern, so daß die Angabe einer theoretischen Semivariogramm-/Kovarianzfunktion ausnahmslos nicht möglich war. Wegen fehlender empirischer Semivariogrammwerte für kurze Entfernungsklassen war auch die Ansprache eines etwaigen Nugget-Effektes nur schwer möglich. Unter Annahme der Normalverteilung der Pseudo-Daten erfolgte die Schätzung der Reichweiten daher ebenfalls im Zuge des Pseudo-Likelihood-Algorithmus. Zudem wurde, ebenfalls aufgrund des geringen Stichprobenumfangs, Isotropie unterstellt. Da die Modelle für unterschiedliche Kovarianzfunktionen oftmals nicht konvergierten, eine klare Präferenz für eine bestimmte Funktion anhand der empirischen Semivariogramme jedoch nicht möglich war, wurde für alle Analysen das sphärische Kovarianzmodell unterstellt, für das die Modelle durchweg am sichersten und besten konvergierten. Lediglich in zwei Fällen, in Tabelle 6.9 durch * gekennzeichnet, konvergierte der Algorithmus auch für das sphärische Modell nicht, so daß die gutachterlich aus den empirischen Semivariogrammen bestimmten Reichweiten verwendet wurden. In allen anderen Fällen lieferte die Pseudo-Likelihood-Schätzung der Kovarianzparameter Ergebnisse, die den gutachterlich bestimmten beachtlich nahe waren. Für vier der fünf Arten (Hartebeest, Kudu, Oryx, Warzenschwein) ergaben sich zudem in beiden Aufnahmephasen recht ähnliche Reichweiten. Dies läßt sich auch als Anzeichen dafür deuten, daß sich das soziale Verhalten und die sozialen Strukturen dieser vier Arten zwischen den Aufnahmephasen nicht wesentlich geändert haben. Für Steinböcke hingegen ergab sich eine deutlich kleinere Reichweite in Aufnahmephase 2, möglicherweise begründet durch die in diesen Zeitraum fallende

Paarungszeit der Steinböcke.

Ein Nugget-Effekt wurde versuchshalber in alle Modelle einbezogen, führte jedoch ausnahmslos zu unplausiblen Schätzungen aller Kovarianzparameter. Dies deckt sich mit der Interpretation der empirischen Semivariogramme, die, wie oben angedeutet, ebenfalls keinen eindeutigen Schluß auf die Notwendigkeit von Nugget-Effekten zuließen.

Die Schätzung der Parameter der einzelnen Modelle erfolgte nicht mit Hilfe von Makros, sondern direkt mit den von SAS zur Verfügung gestellten Prozeduren `PROC GENMOD` (unabhängige Beobachtungen) und `PROC GLIMMIX` (räumlich autokorrelierte Beobachtungen). Lediglich zur gutachterlichen Abschätzung der räumlichen Kovarianzparameter anhand empirischer Semivariogramme der Pearson-Residuen wurde das Makro `%EmpSV` verwendet.

Berechnung räumlicher Vorhersagen und der Gesamtabundanz: Die räumlichen Vorhersagen der Individuenanzahl, die die beiden Modellansätze mit den zuvor beschriebenen Parameterschätzungen für das gesamte Untersuchungsgebiet lieferten, sind in den Abbildungen 6.3 bis 6.7 dargestellt. Für die Modelle, in denen nicht die Anzahl der Individuen, sondern die der Tiergruppen (Objekte) modelliert wurde, wurden die Vorhersagen mit der in Teilprojekt 1 bestimmten, in Tabelle 6.2 dargestellten mittleren Gruppengröße multipliziert. Bereits auf den ersten Blick ist zu erkennen, daß die Vorhersagen der Modelle mit räumlicher Autokorrelation deutlich realistischere Wildverteilungen liefern. Die Verteilungskarten der Modelle ohne räumliche Autokorrelation lassen in der Mehrzahl der Fälle die zugrundeliegende Kovariate klar erkennen oder erlauben, aufgrund der hohen räumlichen Heterogenität der aus Satellitenbildern generierten Informationen, kaum eine Aussage über Gebiete niedriger und hoher Wilddichte. Unabhängig von den verwendeten Kovariaten gelingt es dem Modell mit räumlicher Autokorrelation hingegen in allen Fällen bemerkenswert gut, die Vorhersagen an den tatsächlichen Beobachtungen auszurichten. So sind die Vorhersagen in den Bereichen, in denen viele Beobachtungen gemacht wurden, deutlich höher, in den Bereichen mit weniger oder ohne Beobachtungen deutlich niedriger. Für die Vorhersagen des Modells ohne räumliche Autokorrelation läßt sich dies nur sehr begrenzt feststellen. Der Einbezug der räumlichen Autokorrelation nimmt demnach maßgeblichen Einfluß auf die Vorhersagen, die Bedeutung der Kovariaten dagegen verblaßt. Dies wird umso deutlicher erkennbar, als die *SpGLM*-Vorhersagen für alle fünf Arten zum großen Teil jeweils dieselben Gebiete hoher und niedriger Wilddichten für beide Aufnahmephasen ausweisen, obwohl für keine Art in beiden Aufnahmephasen dieselbe Kovariate verwendet wurde. Für das Modell ohne Autokorrelation ist dies lediglich ansatzweise für Oryx zu beobachten. Die in beiden Aufnahmephasen ähnlichen räumlichen Verteilungen der Vorhersagen des *SpGLM*-Modells legen demnach auch den Schluß nahe, daß die Verteilung der Wildarten keinen übermäßigen saisonalen Schwankungen unterlag. Unterstützt wird diese Aussage durch die Tatsache, daß die Beobachtungen beider Aufnahmephasen über einen mehrmonatigen Zeitraum gesammelt wurden. Die Daten beider Phasen beinhalten also die Habitatpräferenzen der Arten und stellen nicht bloß ein einmaliges und evtl.

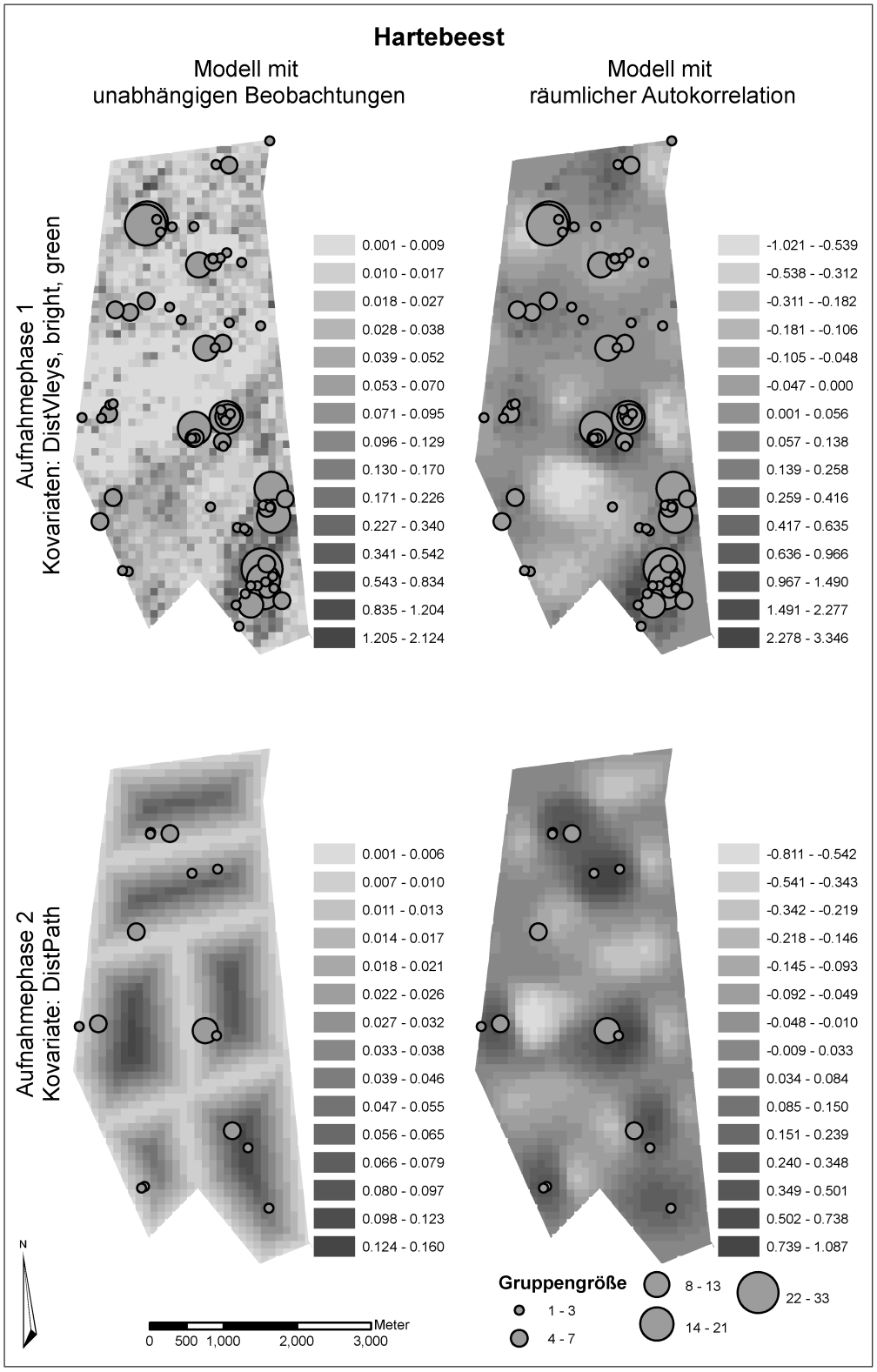

Abbildung 6.3.: Objektpositionen der Hartebeestbeobachtungen und räumliche Vorhersagen der Individuenanzahl. Größe der Zellen des zugrundeliegenden Gitters: $100\,\mathrm{m} \times 100\,\mathrm{m}$.

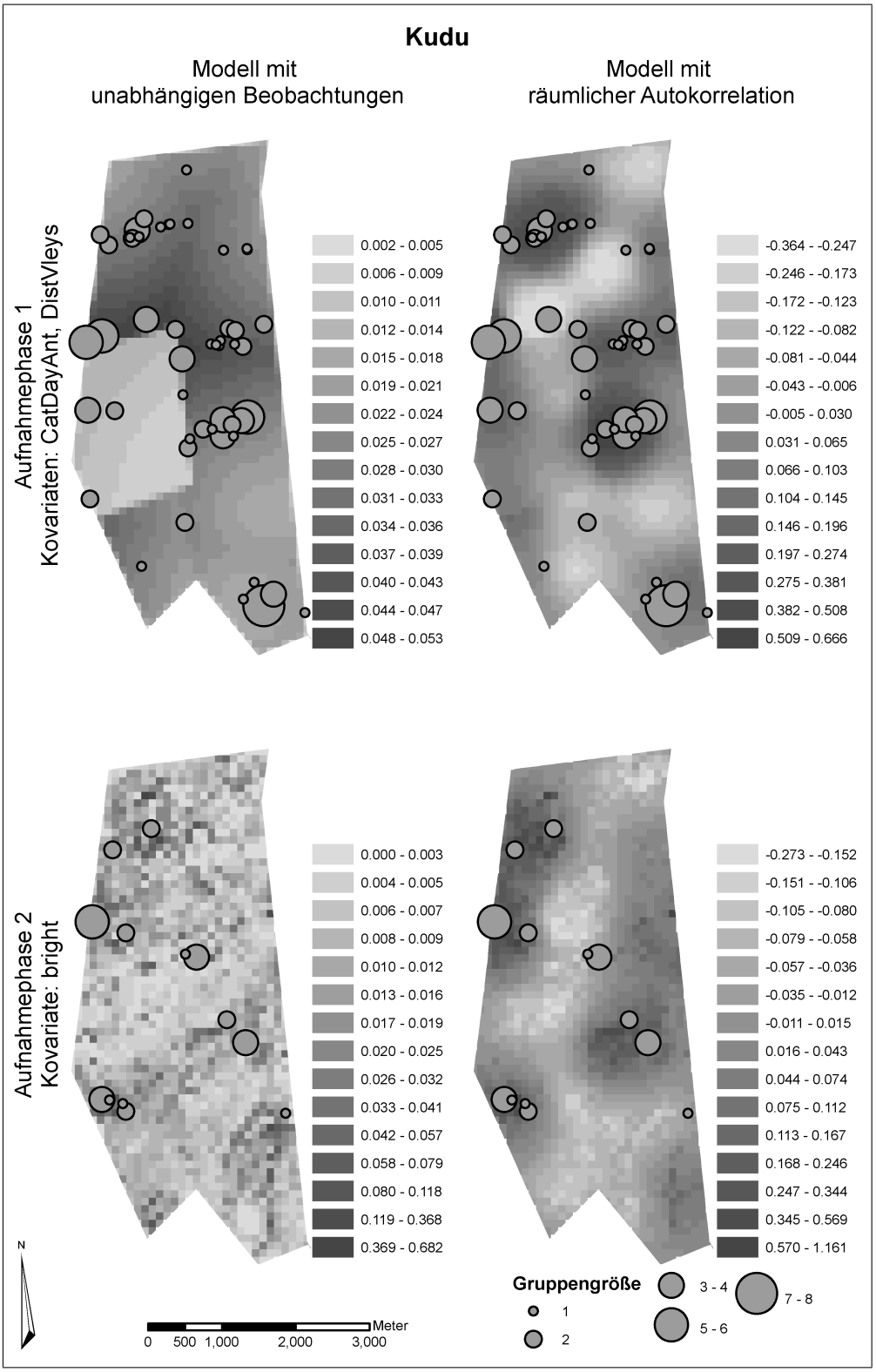

Abbildung 6.4.: Objektpositionen der Kudubeobachtungen und räumliche Vorhersagen der Individuenanzahl. Größe der Zellen des zugrundeliegenden Gitters: $100\,\mathrm{m} \times 100\,\mathrm{m}$.

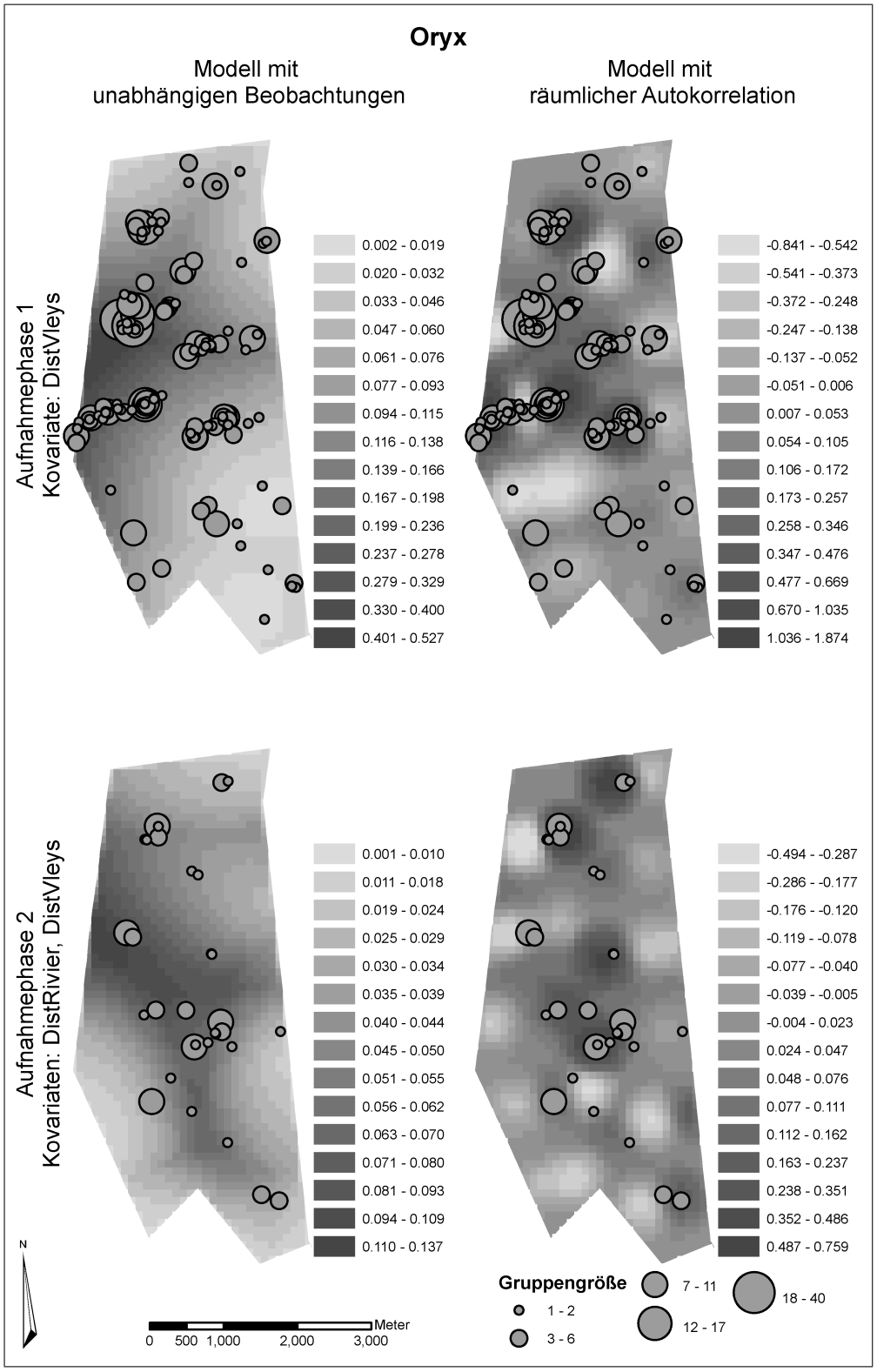

Abbildung 6.5.: Objektpositionen der Oryxbeobachtungen und räumliche Vorhersagen der Individuenanzahl. Größe der Zellen des zugrundeliegenden Gitters: $100\,\mathrm{m} \times 100\,\mathrm{m}$.

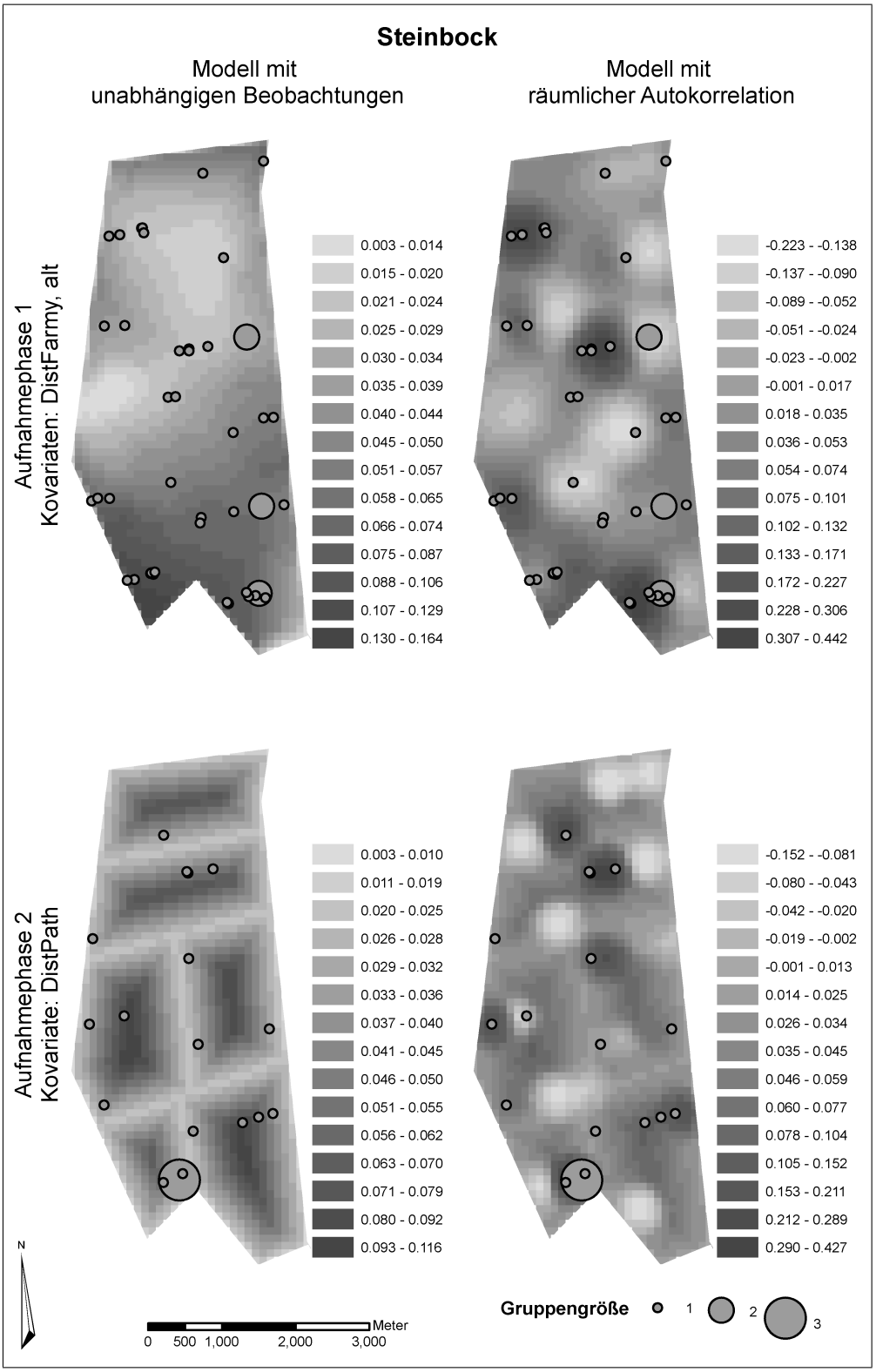

Abbildung 6.6.: Objektpositionen der Steinbockbeobachtungen und räumliche Vorhersagen der Individuenanzahl. Größe der Zellen des zugrundeliegenden Gitters: $100\,\mathrm{m} \times 100\,\mathrm{m}$.

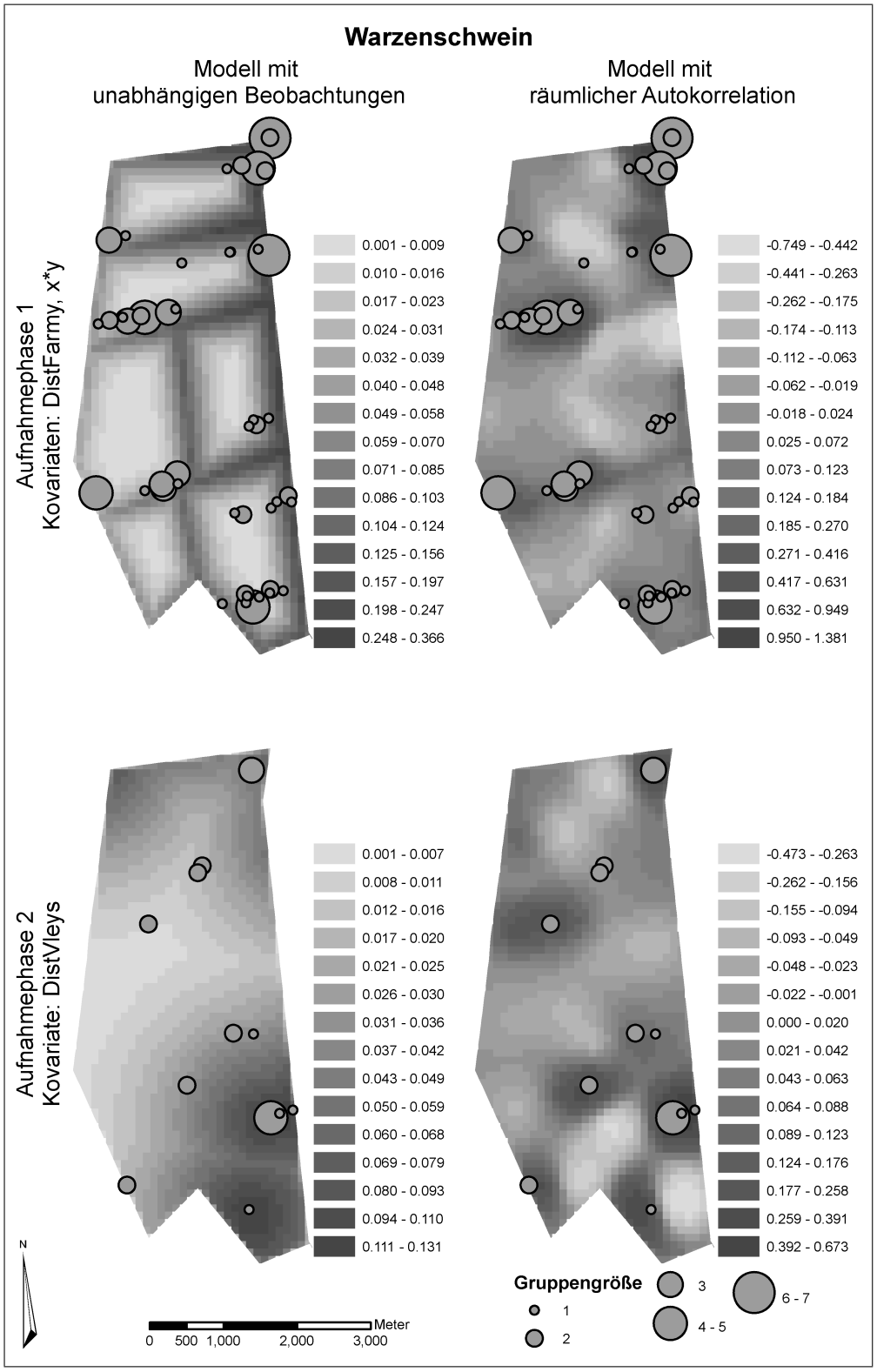

Abbildung 6.7.: Objektpositionen der Warzenschweinbeobachtungen und räumliche Vorhersagen der Individuenanzahl. Größe der Zellen des zugrundeliegenden Gitters: 100 m × 100 m.

Tabelle 6.10.: Schätzungen der Gesamtindividuenabundanz ($\hat{M}$) und deren Standardfehler (in Klammern) für die Modelle mit ($SpGLM$) und ohne ($IGLM$) Berücksichtigung räumlicher Autokorrelation sowie der konventionellen Linientransektanalyse (DS) je Aufnahmephase und Art.

Phase	Art	$\hat{M}_{IGLM}$	$\hat{M}_{SpGLM}$	$\hat{M}_{DS}$
1	Hartebeest	88 (5.21)	114 (20.07)	85 (20.11)
	Kudu	39 (2.67)	66 (6.96)	38 (11.21)
	Oryx	170 (1.17)	161 (18.12)	180 (47.60)
	Steinbock	72 (1.98)	68 (4.10)	68 (15.61)
	Warzenschwein	100 (1.61)	86 (10.79)	56 (16.96)
2	Hartebeest	42 (5.95)	56 (12.15)	38 (12.48)
	Kudu	20 (1.21)	29 (3.65)	28 (9.22)
	Oryx	74 (7.08)	77 (10.72)	86 (20.71)
	Steinbock	67 (1.88)	67 (3.88)	66 (15.02)
	Warzenschwein	54 (3.19)	54 (6.23)	58 (16.24)

weitestgehend zufälliges Ereignis dar.

Ein mit den Vorhersagen der Modelle unter Einbezug räumlicher Autokorrelation verbundener Nachteil ist, wie die Legenden der Karten zeigen, die Möglichkeit negativer Vorhersagen. In einem extremen Fall, der während der Modellselektion auftrat, war der Anteil negativer Vorhersagen sogar so groß, daß auch deren Summe, also die Schätzung der Gesamtabundanz, negativ war. Negative Vorhersagen treten auf, wenn die inverse Link-Funktion des linearen Prädiktors, für das log-lineare Modell also $\exp(\boldsymbol{X}\hat{\boldsymbol{\beta}})$, bereits sehr kleine Werte annimmt. Der Universelle-Kriging-Ansatz (vergl. (2.64) und (2.68)) „korrigiert“ diese Werte über ein gewichtetes Mittel der an den beobachteten Stellen erzielten Residuen $(\boldsymbol{y} - \hat{\boldsymbol{\mu}})$, wobei die Gewichte negative Werte annehmen können, die Korrektur also „nach unten“ durchgeführt wird. Somit werden auch eigentlich unsinnige negative Vorhersagen möglich. Ein Reduktion des Anteils negativer Vorhersagen kann durch die Erhöhung der Werte des Offset-Terms erreicht werden, hier also durch die Vergrößerung der Zellen des Vorhersagegitters. Da der Offset-Term (auf der Originalskala) multiplikativ auf $\exp(\boldsymbol{X}\hat{\boldsymbol{\beta}})$ wirkt, ergeben sich hierfür automatisch ebenfalls größere Werte. Negative „Korrekturen“ müssen also entsprechend klein ausfallen, um insgesamt negative Vorhersagen zu erhalten. Mit Sicherheit ausschließen lassen sich negative Vorhersagen durch dieses Vorgehen jedoch nicht. Zudem führt die Erhöhung der Zellgröße ggf. dazu, daß die kartographische Darstellung der Vorhersagen nur noch sehr grobe Aussagen über die Wildverteilung zuläßt. Das Modell mit Einbezug räumlicher Autokorrelation ist demnach gut für die ordinale Darstellung der Wildverteilung geeignet, liefert für die Vorhersagen an bestimmten Position aber ggf. unsinnige Werte.

Die letzte zu bestimmende Kenngröße, die Anzahl der je Art insgesamt im Untersuchungsgebiet vorhandenen Individuen, wurde für beide Modellansätze über die Summe der Vorhersagen für die einzelnen Zellen ermittelt und ist in Tabelle 6.10 aufgeführt. Ebenfalls in der Tabelle enthalten sind die jeweiligen Standardfehler und – zum Vergleich – die Ergebnisse der konventionellen Linientransektanalyse aus dem ersten Teilprojekt. Tabelle 6.11 enthält darüber

Tabelle 6.11.: Schätzungen der Gesamtindividuendichte ($\hat{D}$) und deren Standardfehler (in Klammern) für die Modelle mit ($SpGLM$) und ohne ($IGLM$) Berücksichtigung räumlicher Autokorrelation sowie der konventionellen Linientransektanalyse (DS) je Aufnahmephase und Art.

Phase	Art	$\hat{D}_{IGLM}$	$\hat{D}_{SpGLM}$	$\hat{D}_{DS}$
1	Hartebeest	5.50 (0.32)	7.07 (1.25)	5.27 (1.25)
	Kudu	2.45 (0.17)	4.12 (0.43)	2.41 (0.71)
	Oryx	10.55 (0.07)	10.03 (1.13)	11.20 (2.96)
	Steinbock	4.49 (0.12)	4.22 (0.25)	4.20 (0.96)
	Warzenschwein	6.23 (0.10)	5.33 (0.67)	3.44 (1.04)
2	Hartebeest	2.62 (0.37)	3.47 (0.76)	2.35 (0.77)
	Kudu	1.24 (0.08)	1.78 (0.23)	1.76 (0.58)
	Oryx	4.62 (0.44)	4.82 (0.67)	5.32 (1.28)
	Steinbock	4.17 (0.12)	4.20 (0.24)	4.08 (0.93)
	Warzenschwein	3.38 (0.20)	3.35 (0.39)	3.66 (1.03)

hinaus die entsprechenden Dichteschätzungen. Die in den Tabellen dargestellten Standardfehler des $IGLM$-Ansatzes wurden, wie in Kapitel 2.3.3 erläutert, auf Basis der Varianzen der zellweisen Vorhersagen, die des $SpGLM$-Ansatzes mit Hilfe des Block-Kriging-MSPEs bestimmt. Mit nur zwei Ausnahmen (Hartebeest und Kudu, Phase 1) liegen die Schätzungen der Gesamtabundanz der beiden Count-Modell-Ansätze trotz der sehr unterschiedlichen Verteilungskarten beachtlich nahe beieinander. Darüber hinaus liegen die Schätzungen in der Mehrzahl der Fälle auch sehr nahe an denen der herkömmlichen Linientransektanalyse. Eine Ausnahme, in der die Abundanzschätzung beider log-linearen Modelle recht deutlich über der konventionellen Linientransektschätzung liegt, sind die Warzenschweindaten aus Aufnahmephase 1. Die auffällige Parameterschätzung für den DistPath-Effekt für diese Daten ist bereits oben besprochen worden. Die Ergebnisse in Tabelle 6.10 und 6.11 legen, in Verbindung mit dem hohen Standardfehler der Parameterschätzung, die Vermutung nahe, daß der tatsächliche Einfluß der Entfernung zum nächsten Weg anhand der beobachteten Daten überschätzt wurde, die Parameterschätzung also zu klein ist. Dafür spricht auch, daß das $SpGLM$-Modell die Vorhersage des $IGLM$-Modells unter Einbezug der räumlichen Korrelation nach unten „korrigiert“.

Die Standardfehler der drei Modellansätze sind erwartungsgemäß für das Count-Modell unter der Annahme unabhängiger Beobachtungen am kleinsten, werden infolge der tatsächlich vorhandenen Autokorrelation jedoch deutlich unterschätzt. Entsprechend fallen die Standardfehler des Modells mit räumlicher Autokorrelation sehr viel größer aber in fast allen Fällen noch deutlich kleiner als die der gewöhnlichen Linientransektanalyse aus. Die Verwendung zusätzlicher räumlicher Kovariaten ermöglicht demnach auch unter Berücksichtigung räumlicher Autokorrelation die Reduktion der Streuung für Dichte- und Abundanzschätzungen aus Linientransektstichproben.

Die Generierung des Vorhersagegitters wurde mit Hilfe des SAS-Makros `%CreatePolyGrid` umgesetzt, die Ermittlung der räumlichen Kovariaten wiederum mit `%SpatialQueryPolys`. Beide Makros rufen zusätzliche Python-Skripte auf und werden gemeinsam mit diesen in Kapitel 3.2 detaillierter erläutert. Die Be-

rechnung der Vorhersagen des Modells ohne räumliche Autokorrelation konnte direkt mit `PROC GENMOD` durchgeführt werden, für die des Modells mit räumlicher Autokorrelation hingegen wurde die Prozedur `GLIMMIX` in Kombination mit dem `%MargGlmSpatPred`-Makro verwendet.

7. Diskussion und Ausblick

Wie im Ergebnisteil erfolgt die Diskussion und der Ausblick auf weitere Entwicklungen und Möglichkeiten getrennt nach den beiden ersten Teilprojekten dieser Arbeit. Dabei wird gleichermaßen auf allgemeine methodische Aspekte und auf die konkreten namibischen Daten eingegangen.

7.1. Validierung der Farmwegzählungen (Teilprojekt 1)

Die Ergebnisse des ersten Teilprojektes stellen die Brauchbarkeit der Farmwege als Transekte sehr deutlich in Frage. Zwar lassen sich Dichteunterschiede zwischen den Aufnahmeschemata aufgrund deren großen Streuungen nur in wenigen Fällen statistisch absichern und für robuste Verfahren der Varianzschätzungen (z. B. nichtparametrisches Bootstrapping) wäre dies voraussichtlich noch seltener der Fall, die Dichteschätzungen selbst geben jedoch sehr deutlich Anlaß zur Vermutung, daß die Wahl des Stichprobendesigns durchaus erheblichen Einfluß besitzt. Außer für Kudu sollte daher bei Verwendung von Wegtransekten von einer deutlichen Unterschätzung der wahren Wilddichte ausgegangen werden. Die Ergebnisse des zweiten Teilprojektes stützen diese Aussage. Hier wurden – außer für Kudu – mehrfach die Farmwege selbst oder daran gelegene Wasserstellen als auf die lokale Wildverteilung Einfluß nehmende Faktoren ausgewählt.

Ein weiterer problematischer Aspekt der Verwendung von Farmwegen ist die Schwierigkeit der Anpassung einer Entdeckungsfunktion an die Daten der Wegzählungen. Zahlreiche Beobachtungen direkt auf dem weit einsehbaren Transekt/Weg machen die Anpassung einer Entdeckungsfunktion sehr schwierig und erfordern oftmals die Wahl einer (in gewissen Grenzen) willkürlichen linken Kappungsentfernung w_l. Wie allgemein angeraten (Buckland et al., 2001, Kap. 7.8.5; Norton-Griffith, 1978, Kap. 10.5), sollte daher trotz deren vergleichsweise seltener Nutzung von der Verwendung von Wegen als Transekte abgesehen werden. Da verläßliche Stichprobendesigns in kommerziellen namibischen Conservancies jedoch automatisch mit deutlich erhöhtem technischen, logistischen und letztlich finanziellen Aufwand verbunden sind, wird sich ihre Anwendung hier vermutlich frühestens mit einer entsprechenden Änderung der gesetzlichen Vorgaben in der Praxis durchsetzen.

Für das Erreichen des eigentlichen Ziels der Aufnahmen, die Überwachung der nachhaltigen Nutzung der Wildbestände, könnte das derzeitige Procedere dennoch hilfreich sein. Hierbei sind weniger die konkreten Dichteschätzungen von Bedeutung als vielmehr deren Entwicklung über die Zeit. Da die wahren Dichten für alle Arten (außer Kudu) zudem unterschätzt werden, ist die Gefahr, basierend auf diesen Zahlen zu starke Eingriffe zu planen, relativ gering.

Die Interpretation zeitlicher Trends ist jedoch an die Bedingung gekoppelt, daß der Faktor, um den die wahre Wilddichte unterschätzt wird, für alle Aufnahmen konstant ist. Die in Abbildung 6.1 dargestellten Dichteschätzungen deuten durchaus darauf hin, daß die Faktoren, um die die Ergebnisse der Wegtransekte für die vier unterschätzten Arten unter denen der Buschtransekte liegen, für beide Aufnahmephasen von etwa gleicher (aber eben nicht notwendig identischer) Größenordnung sind. Wenngleich die im Rahmen dieser Arbeit erhobenen Daten keine sichere Aussage darüber zulassen, so dürfte das Maß der Unterschätzung der wahren Wilddichte jedoch im allgemeinen räumlich, zeitlich und farmspezifisch variieren. So wird die räumlich und zeitlich sehr heterogene Verteilung von Niederschlägen und die (temporäre) Existenz natürlicher Wasserstellen (Vleys, Riviere) die vermeintlich anziehende Wirkung von künstlichen Wasserstellen deutlich beeinflussen. Unterstellt man zudem, wie die Ergebnisse des zweiten Teilprojektes vermuten lassen, den Wegen selbst einen negativen Einfluß, so wird dessen Stärke maßgeblich davon abhängen, wie stark die Beunruhigung und insbesondere der Jagddruck von den als Transekte dienenden Wegen ist. Diese Aspekte deuten bereits darauf hin, daß es sich nicht um einen zeitlich und räumlich weitestgehend konstanten Faktor sondern um einen ggf. ebenfalls aufwendig zu bestimmenden Parameter handeln dürfte.

7.2. Räumliche Modellierung (Teilprojekt 2)

Die Ansätze für Diskussionen und Weiterentwicklungen der für die räumliche Analyse von Linientransektdaten vorgestellten Modelle und Methoden lassen sich unter den folgenden vier Punkten zusammenfassen:

- (Räumliche) Kovariaten
- Modellselektion
- Varianzschätzung
- Andere Modellansätze

(Räumliche) Kovariaten Die Ergebnisse des zweiten Teilprojektes haben gezeigt, daß sich mit Hilfe der räumlichen Analyse und dem Einbezug zusätzlicher Kovariaten die Streuung der Dichteschätzungen durchaus reduzieren und zudem lokale Vorhersagen der Abundanz berechnen lassen. Mit der geänderten Modellformulierung gehen jedoch auch vier nicht unwesentliche Vorteile des herkömmlichen Linientransektverfahrens verloren: Zum einen vervielfachen sich die Kosten der Datenerhebung. Objektpositionen sind (z. B. per GPS) zu verorten, benötigte zusätzliche geographische Informationen zu erstellen oder zu beschaffen. Zum zweiten werden Methoden, Modelle und software-technische Umsetzung ungleich komplexer und damit weniger leicht nachvollziehbar. Zum dritten gelten die klassischen Linientransektdichteschätzungen, ein verläßliches Stichprobendesign vorausgesetzt, als robust, für die lokalen Vorhersagen der räumlichen Modelle dürfte dies jedoch nur sehr eingeschränkt gelten, wie insbesondere die Möglichkeit (unsinniger) negativer Vorhersagen verdeutlicht. Zum

vierten eignen sich die Verfahren des Distance Sampling im allgemeinen insbesondere für die Dichteschätzungen seltener Objekte, für die räumliche Analyse hingegen werden möglichst viele Beobachtungen benötigt, um die Segmentlänge möglichst klein wählen und damit eine möglichst hohe räumliche Auflösung erreichen zu können.

Trotz der großen Anzahl zur Verfügung stehender Kovariaten gelang es, wie Tabelle 6.7 zeigt, nur in einem Fall, mehr als eine Kovariate als einflußreich auszuweisen. Wesentlich hierfür waren geringe Beobachtungsanzahlen, die eine, gemessen an der Ausdehnung des Untersuchungsgebietes, große Segmentlänge erforderten, so daß wiederum nur Informationen weniger Segmente zur Beurteilung der Kovariaten zur Verfügung standen. Daneben stellt sich jedoch auch die Frage nach wesentlichen, aber nicht verfügbaren geographischen Informationen und nach der Qualität der vorhandenen. So fehlten z. B. Informationen über aktuelle oder zurückliegende Niederschläge und Vegetationsindizes waren nur für einen Zeitpunkt weit nach den Aufnahmen verfügbar. Des weiteren ist es denkbar, daß einige Kovariaten, wie die Entfernung zu einer potentiellen Wildereiquelle, einen signifikanten Einfluß besitzen, sich dieser aber erst über größere Entfernungen, als sie in dem vergleichsweise kleinen Untersuchungsgebiet gemessen werden konnten, nachweisen läßt. Es ist also möglich, daß, bedingt durch die geringe Größe des Untersuchungsgebietes, die Spannweite der Werte einiger Kovariaten zu gering ist, um einen signifikanten Beitrag zu zeigen.

Eine im Rahmen dieser Arbeit ebenfalls nicht in Betracht gezogene Einflußgröße auf die Verteilung einer Wildart ist die der übrigen Arten. Zumindest für die drei großen Antilopenarten läßt sich vermuten, daß neben der artspezifischen Habitatpräferenz auch die An- bzw. Abwesenheit der beiden übrigen Arten von Bedeutung ist. So lassen sich anhand der Abbildungen 6.3 bis 6.5 zumindest grob Ballungsgebiete ausweisen, die sich für die drei großen Arten jedoch nur wenig überlappen. Um einen etwaigen wechselseitigen Einfluß in die Modelle aufzunehmen, wäre ein zusätzlicher Iterationsschritt erforderlich, in dem die Berechnung der lokalen Vorhersagen einer Art (u. a.) in Abhängigkeit von den aktuellen Vorhersagen einer oder mehrerer der übrigen Arten erfolgt.

Der räumlichen Autokorrelation kommt ggf. nicht nur die Rolle des Ersatzes für fehlende Kovariaten zu. Selbst wenn ein Modell signifikante und plausible Kovariaten enthält und realistische und sinnvolle Vorhersagen liefert, kann es sinnvoll sein, die räumliche Korrelation zusätzlich zu berücksichtigen. Dann nämlich, wenn „echte“ Autokorrelation beobachtet werden kann, wie es gerade für biologische Populationen oftmals der Fall ist. Während sich beispielsweise Niederschläge, wie die meisten Größen, die mit Hilfe geostatistischer Methoden untersucht werden, an zwei benachbarten Orten nur deshalb ähnlich sind, weil die Bedingungen (z. B. Geländehöhe) an beiden Orten sehr ähnlich sind, konzentrieren sich Tiere (nicht nur) in einem bestimmten Gebiet, weil die Habitatbedingungen dort besonders günstig sind, sondern (evtl. maßgeblich), weil sich dort weitere Tiere derselben Art aufhalten. Ursache ist also die untersuchte Größe selbst, kein externer Effekt.

Ein letzter Punkt betrifft die Berechnung der Kovariatenwerte. Es ist bereits angesprochen worden, daß mit zunehmender Segmentlänge der Segmentschwerpunkt und die anhand dessen bestimmten Kovariatenwerte immer weniger re-

präsentativ für das gesamte Segment werden. Beginnt beispielsweise ein langes Transekt in der Nähe einer Tränke, so liegt der Segmentschwerpunkt bereits ein Stück von dieser Tränke entfernt, die zugehörige Kovariate wird entsprechend einen relativ hohen Wert annehmen, auch dann, wenn direkt am Segmentanfang mehrere Beobachtungen in Tränkennähe gemacht wurden. Zudem existieren Beobachtungen (Objektpositionen), die zwar einem Segment zugewiesen wurden, aber nicht innerhalb dessen Fläche liegen (gemäß der Definition der effektiven Streifenbreite existieren Beobachtungen mit $x > \hat{v}$); die geographischen Eigenschaften an der Stelle der Objektposition werden also evtl. gar nicht erfaßt. Ein alternative Berechnung der Kovariatenwerte könnte daher darin bestehen, die zusätzlichen Informationen direkt an den Objektpositionen zu bestimmen und je Segment den Mittelwert, das Maximum oder den am häufigsten vorkommenden Wert zu bestimmen. Im Falle des genannten Tränkenbeispiels wäre der Mittelwert der tatsächlichen Entfernungen zwischen Tränke und Objekten deutlich kleiner als die zwischen Segmentschwerpunkt und Tränke, so daß die Attraktivität der Tränke besser dargestellt würde.

Modellselektion Für die Auswahl der die in einem Segment beobachtete Individuen- oder Objektanzahl erklärenden Kovariaten wurde in dieser Arbeit neben weiteren Kriterien ein einfacher aber nicht unproblematischer Ansatz der automatisierten Modellauswahl über die Wald-Statistiken der Typ-3-Kontraste gewählt. Da speziell die Verfügbarkeit geographischer Informationen und damit die potentieller Kovariaten rasant zunimmt, gilt es robuste Selektionsverfahren zu entwickeln, die auch für GLMs mit (räumlich) korrelierten Beobachtungen anwendbar sind. Das in Kapitel 2.2.4 angesprochene QIC und QIC_u sind erste Ansätze hierfür, die auf dem für unabhängige Daten weit verbreiteten AIC basieren. Ihre Anwendung hat sich jedoch offensichtlich bislang nicht durchgesetzt, so daß es an ihrer Umsetzung in statistischer Software fehlt. Zudem verwendet das QIC eine modifizierte Form des robusten Sandwich-Schätzers der Varianz, der jedoch im Kontext räumlicher Kovarianzstrukturen nur selten herleitbar ist, da hier in der Regel nur ein unabhängiges Subjekt (eine Realisation des räumlichen Prozesses) vorliegt. Die Anwendbarkeit der beiden Kennzahlen für räumlich korrelierte Daten ist daher fraglich.

Nach Möglichkeit ebenfalls automatisierte und robuste Verfahren werden auch für die Auswahl der Segmentlänge benötigt. Hedley u. Buckland (2004) haben in ihren Analysen zwar festgestellt, daß die Wahl der Segmentlänge wenig Einfluß auf die Schätzung der Gesamtabundanz nimmt, diese Beobachtung dürfte jedoch an eine umfangreiche Datenbasis gekoppelt sein. Liegen dagegen, wie in dieser Arbeit, nur wenige Beobachtungen vor, so kann die Wahl der Segmentlänge durchaus erheblichen Einfluß auf die Parameterschätzungen nehmen. Wesentlich ist dabei der Anteil der Segmente ohne Beobachtungen. Je kürzer die Segmentlänge gewählt wird, desto höher ist dieser Anteil und desto stärker werden die Parameterschätzungen hiervon dominiert.

Varianzschätzung Der kritischste und zugleich dringlichste Aspekt zukünftiger Entwicklungen dürfte die verläßliche Schätzung der Varianz der Vorhersagen

und Schätzungen des Count-Modells und allgemein räumlicher GLMs sein. Die Ergebnisse aus Kapitel 6.2 zeigen, daß eine Reduktion der Varianz der Abundanzschätzungen durch Einbezug von Kovariaten möglich, die Berücksichtigung räumlicher Autokorrelation aber insbesondere dann von Bedeutung ist, wenn durch die Kovariaten nur ein geringer Teil der beobachteten Streuung erklärt werden kann, so daß weiter von einer starken räumlichen Abhängigkeit auszugehen ist. Dennoch kann die Verläßlichkeit der Varianzschätzungen der lokalen Vorhersagen und der Gesamtabundanz aus mehreren Gründen angezweifelt werden: Zunächst geht die effektive Streifenbreite in beiden Varianten des Count-Modells (mit und ohne Berücksichtigung räumlicher Autokorrelation) als Teil des Offset-Terms in das log-lineare Modell ein, ohne daß in den anschließenden Varianzschätzungen berücksichtigt wird, daß es sich hierbei nicht um einen bekannten, sondern ebenfalls geschätzten Parameter handelt. Ein allgemeines, nicht auf die Anwendung in GLMs begrenztes Problem der Krigingvarianzen (MSPE) besteht darüber hinaus darin, daß in sie zwar die Streuung der geschätzten Parameter der festen Effekte $\hat{\boldsymbol{\beta}}$, nicht jedoch die Streuung der in aller Regel ebenfalls geschätzten Kovarianzparameter $\boldsymbol{\alpha}$ eingehen; diese werden als bekannt vorausgesetzt. Darüber hinaus handelt es sich um unbedingte Fehler, die nicht von der Streuung der für die Vorhersage verwendeten beobachteten Werte, sondern von der gemeinsamen (räumlichen) Anordnung der beobachteten und vorherzusagenden Punkte abhängen.

Um das Ziel robuster Varianzschätzungen für das Count-Modell ohne räumliche Autokorrelation zu erreichen, hat Hedley (2000, Kap. 5) drei klassische Verfahren der robusten Varianzschätzung an die Bedingungen des Count-Modells angepaßt. Die Verfahren sollen die Tatsache, daß es sich bei der effektiven Streifenbreite um eine Schätzung handelt, und insbesondere die räumliche Abhängigkeit zwischen den Segmenten berücksichtigen. Es handelt sich um das *Jackknife*-, das parametrische und das nichtparametrische *Bootstrap*-Verfahren. Das nichtparametrische Bootstrapping hat neben der Annahme unabhängiger und identisch verteilter Beobachtungen (iid) den wesentlichen Nachteil, daß beim Ziehen mit Zurücklegen (ZmZ) vollständiger Transekte zur Erstellung der einzelnen Bootstrap-Stichproben evtl. eine nur unzureichende räumliche Abdeckung erreicht wird, sofern die Anzahl der Transekte nicht sehr groß ist. Es soll daher nur auf die beiden verbleibenden Verfahren kurz eingegangen werden. Weitere Möglichkeiten der Varianzschätzung bieten darüber hinaus die unten angesprochenen Ansätze der Bayes-Statistik.

Für das Jackknife-Verfahren empfiehlt Hedley (2000), nicht die einzelnen Beobachtungen, sondern vollständige Transekte als Stichprobeneinheiten zu definieren. Sukzessive wird jeweils eine Stichprobeneinheit samt der zugehörigen Beobachtungen aus dem Originaldatensatz entfernt und alle Analyseschritte mit Hilfe der Daten der verbliebenen $k-1$ Transekte durchgeführt. Dies beginnt mit der (ggf. segment- oder beobachtungsweisen) Schätzung der Entdeckungswahrscheinlichkeit und endet mit der Berechnung der Summe der lokalen Vorhersagen zur Schätzung der Gesamtabundanz. Die Jackknife-Schätzung der Gesamtindividuenanzahl ohne Verwendung des i-ten Transektes $\hat{M}^{(-i)}$ berechnet sich wie in Kapitel 2.3.2 beschrieben. Im nächsten Schritt werden dann sog. Pseudowerte $\tilde{M}^{(-i)}$ (nicht zu verwechseln mit den Pseudo-Daten ν aus dem PL-Ansatz)

berechnet.

$$\tilde{M}^{(-i)} = k \cdot \hat{M} - (k-1)\hat{M}^{(-i)}, \quad i = 1, \ldots, k$$

Infolge der so erhöhten Variabilität der Pseudowerte, werden diese behandelt, als würde es sich um k unabhängige Werte handeln (Efron u. Tibshirani, 1993). Der Jackknife-Varianzschätzer der anhand des Originaldatensatzes geschätzten Gesamtabundanz $\hat{M}$ ist dann

$$\mathrm{var}(\hat{M}) = \frac{\sum_{i=1}^{k}(\tilde{M}^{(-i)} - \tilde{\bar{M}})^2}{k(k-1)} = \frac{\mathrm{var}(\tilde{M}^{(-i)})}{k},$$

wobei $\tilde{\bar{M}}$ der Mittelwert der Pseudowerte

$$\tilde{\bar{M}} = \frac{1}{k}\sum_{i=1}^{k}\tilde{M}^{(-i)}$$

ist.

Im Gegensatz zum nichtparametrischen Bootstrapping werden die Bootstrap-Stichproben im parametrischen Bootstrapping nicht durch Ziehen einer Stichprobe (ZmZ) aus den Originaldaten, sondern aus dem an die Originaldaten angepaßten Modell gebildet. Vorgegangen wird dabei wie folgt:

1. Anpassung eines Modells an die Originaldaten und darauf basierend Ermittlung der Dichteschätzung entlang der Transekte. In der Praxis werden hierfür Dichteschätzungen für diskrete, nahe aufeinanderfolgende Punkte auf dem Transekt berechnet.

2. Erzeugen einer Poisson-verteilten Zufallszahl $\tilde{n}$ mit Erwartungswert $\mathrm{E}(n)$, der in den Originaldaten zu erwartenden Gesamtbeobachtungsanzahl, geschätzt durch $n = \sum_{i=1}^{k} n_i$.

3. Erzeugen einer im Intervall $(0, \hat{D}_{max})$ gleichverteilten Zufallszahl, wobei $\hat{D}_{max}$ die größte geschätzte Dichte entlang dem Transekt ist.

4. Erzeugen einer im Intervall $(0, L)$ gleichverteilten Zufallszahl, wobei $L = \sum_{i=1}^{T} c_i = \sum_{i=1}^{k} l_i$ die Gesamtlänge aller Transekte ist.

5. Die im 3. und 4. Schritt erzeugten Zufallszahlen definieren einen Punkt in einem zweidimensionalen kartesischen Koordinatensystem. Liegt dieser Punkt unterhalb der in Schritt 1 modellierten Dichtekurve, wird er angenommen, andernfalls abgelehnt.

6. Wiederholung der Schritte 3 bis 5 bis $\tilde{n}$ Punkte angenommen wurden.

7. Berechnung der auf jedes Segment entfallenden Punkte (dies ist die simulierte Beobachtungsanzahl), Anpassen des Ausgangsmodells an die generierten Daten und Berechnung der zugehörigen Abundanz- oder Dichteschätzungen für das gesamte Untersuchungsgebiet.

8. B-fache Wiederholung der Schritte 2 bis 7. Für verläßliche Varianzschätzungen sollten mindestens 200 Bootstrap-Stichproben gezogen werden, d. h. $B \geq 200$.

Die Stichprobenvarianz der Bootstrap-Individuenabundanzschätzung

$$\mathrm{var}_B(\hat{M}) = \frac{1}{B-1} \sum_{i=1}^{B} \left(\hat{M}_{B_i} - \bar{M}_B \right)^2$$

mit

$$\bar{M}_B = \frac{1}{B} \sum_{i=1}^{B} \hat{M}_{B_i}$$

ist dann eine Schätzung des Anteils der Varianz von $\hat{M}$ infolge der räumlichen Modellierung.

Auf Grundlage der Ergebnisse eines realen Anwendungsbeispiels äußert Hedley (2000, S. 54, 61 f.) die Vermutung, daß es dem parametrischen Bootstrapping nicht gelingt, die räumliche Korrelation zwischen den Segmenten vollständig zu reproduzieren und die wahren Varianzen daher unterschätzt werden. Wird die räumliche Autokorrelation in den Modellen (in Schritt 7 oben) direkt berücksichtigt, so könnte das parametrische Bootstrapping umso mehr als robuste Alternative zu analytischen Varianzschätzern eingesetzt werden. Zudem läßt sich auf diese Weise die Unsicherheit der Kovarianzparameterschätzungen und – geeignete Selektionsmechanismen vorausgesetzt – ggf. auch die Unsicherheit der Wahl der Kovarianzstruktur berücksichtigen.

Besonders das Jackknife-Verfahren bietet sich aus mehreren Gründen für eine praktische Anwendung an. Da der Ablauf der Analyse jeder einzelnen Jackknife-Stichprobe der Analyse des Originaldatensatzes entspricht, läßt sich das Verfahren relativ leicht mit Hilfe der vorhandenen Arbeitsabläufe umsetzen. Zudem beginnt jede der Analysen mit der Schätzung der Entdeckungswahrscheinlichkeit (P_a, p_i oder p_{ij}); so geht deren Unsicherheit direkt in die Jackknife-Varianz ein. Darüber hinaus läßt sich auf diese Weise sogar die Unsicherheit der Modellselektion der Entdeckungsfunktion berücksichtigen, indem die Auswahl für jede der Jackknife-Stichproben getrennt erfolgt (z. B. per AIC). Eine Schwachstelle des Jackknife-Verfahrens ist jedoch die Begrenzung der Anzahl der Jackknife-Stichproben auf die Anzahl der Stichprobeneinheiten (Transekte) der Originaldaten. Im parametrischen Bootstrapping hingegen wird die Anzahl der Bootstrap-Stichproben nicht durch die Originaldaten, sondern lediglich durch Rechenzeit und -leistung limitiert. Zudem unterstellt es als einziges der drei genannten Verfahren nicht die Unabhängigkeit und identische Verteilung der Beobachtungen.

Bereits die Ergebnisse von Hedley (2000) zeigen, daß es weiteren Aufwandes bedarf, um die Verläßlichkeit sowie Möglichkeiten und Grenzen der Anwendbarkeit vorhandener Varianzschätzer aufzuzeigen und ggf. weitere robuste Schätzer zu entwickeln. Wesentliche Bedeutung wird hierbei simulierten Daten zukommen (s. z. B. Augustin et al., 1998), in denen die wahren Parameter bekannt sind. Speziell für die Simulation der Verteilung von Wildtieren bietet WISP[1] (Zucchini et al., 2002), eine Programmbibliothek der freien Statistiksoftware R[2] (Ihaka u. Gentleman, 1996), umfangreiche Möglichkeiten.

[1] http://www.ruwpa.st-and.ac.uk/estimating.abundance/WiSP/index.html
[2] http://www.r-project.org/

Andere Modellansätze Die Möglichkeiten, die in Kapitel 2 vorgestellten Modelle zur räumlichen Analyse von Linientransektdaten zu erweitern oder zu modifizieren, sind zahlreich: So ist z. B. die Verwendung anderer Link- oder Varianzfunktionen oder auch anderer Kriging-Ansätze zur Vorhersage nicht stationärer Daten (*Median Polish Kriging*, *Trans Gaussian Kriging*, Disjunktives Kriging) denkbar. Die nachfolgende Auflistung beschreibt daher nur einen Ausschnitt möglicher Ansätze, der sich maßgeblich an dem Beibehalt der Verwendung verallgemeinerter linearer Modelle orientiert.

Neben dem hier vorgestellten und verwendeten Count-Modell beschreibt Hedley (2000, Kap. 4) zwei Ansätze, in denen anstelle der Anzahl der in einem Segment beobachteten Objekte/Individuen die (angepaßte) Entfernung zwischen zwei aufeinanderfolgenden Beobachtungen $\tilde{o}_i$ als abhängige Variable verwendet wird. Das *waiting-distance*-Modell verwendet eine für alle Transekte einheitliche (geschätzte) Entdeckungsfunktion, das *waiting-area*-Modell ermöglicht darüber hinaus die Verwendung unterschiedlicher effektiver Streifenbreiten; die abhängige Variable ist in diesem Fall die *waiting area* $\tilde{r} = 2 \cdot \tilde{o}_i \cdot \hat{v}_i$. Die Modellformulierung erfolgt für beide Ansätze wiederum mit Hilfe verallgemeinerter linearer oder additiver Modelle. Näheres zu beiden Ansätzen findet sich auch bei Buckland et al. (2004, Kap. 4.3.1), Hedley u. Buckland (2004) und Borchers et al. (2002, Kap. 12.2.2).

Bereits im Abschnitt 2.2.7.2 kurz angesprochen wurde die Möglichkeit, räumliche Abhängigkeit in verallgemeinerten linearen Modellen in Form zufälliger Effekte zu berücksichtigen. Der lineare Prädiktor des (bedingten) Erwartungswertes $\boldsymbol{\mu}(\boldsymbol{s}) = \mathrm{E}[\boldsymbol{Y}(\boldsymbol{s})|\boldsymbol{S}]$ ergibt sich dann in Abhängigkeit der Kovariaten und eines latenten Gaußschen Zufallsprozesses $\boldsymbol{S}$ mit Mittelwert 0 und Kovarianzfunktion $\mathrm{cov}(\sigma^2_{\boldsymbol{S}}, \rho(\boldsymbol{s}_i - \boldsymbol{s}_j))$

$$g[\boldsymbol{\mu}(\boldsymbol{s})] = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{S}(\boldsymbol{s}).$$

Zusätzlich werden bedingt (gegeben $\boldsymbol{S}$) unabhängige Beobachtungen mit Varianz

$$\mathrm{var}[\boldsymbol{Y}(\boldsymbol{s})|\boldsymbol{S}] = \phi\, v(\boldsymbol{\mu}).$$

unterstellt, wobei ϕ ein ggf. zu schätzender Dispersionsparameter und $v(\cdot)$ die Varianzfunktion der unterstellten (bedingten) Verteilung ist. Parameterschätzungen und räumliche Vorhersagen lassen sich, ähnlich wie im marginalen Modell, mit Hilfe des PL-Ansatzes von Wolfinger u. O'Connel (1993) und einem universellen Kriging der Pseudo-Daten berechnen. Gotway u. Wolfinger (2003) vergleichen anhand eines realen und simulierter Datensätze das marginale mit dem bedingten Modell, kommen aber zum Ergebnis, daß nicht prinzipiell einer der Ansätze dem anderen überlegen ist. Insbesondere fehlt es bisher an objektiven Bewertungskriterien, welches Modell für einen konkreten Datensatz geeigneter ist. Zusätzliche Informationen über GLMMs im Kontext raumbezogener Daten liefern z. B. Schabenberger u. Gotway (2004, Kap. 6.3) sowie die im nächsten Absatz genannten Autoren.

Wachsender Beliebtheit und Anwendung erfreuen sich im Zusammenhang mit GLMMs Methoden der Bayes-Statistik (z. B. Lee, 2004; Gelman et al., 2003). Im Gegensatz zu Ansätzen der klassischen Statistik (zu denen auch

der PL-Ansatz gehört) werden hier die zu schätzenden Parameter (Parameter der festen und zufälligen Effekte sowie Kovarianzparameter) nicht als konstant, sondern, wie die Zufallsvariablen, deren Realisierungen beobachtet wurden, ebenfalls als zufällig mit einer sog. a-priori-Verteilung angenommen. Zusammen mit der gemeinsamen Dichtefunktion der beobachteten Daten wird sie zur Berechnung der a-posteriori-Verteilung verwendet. Im Gegensatz zu klassischen Verfahren steht damit nicht nur der Punktschätzer, sondern dessen gesamte a-posteriori-Verteilung für die Inferenz zur Verfügung, so daß z. B. direkte Aussagen über Konfidenzintervalle der Parameter möglich sind. Die Parameterschätzungen selbst ergeben sich durch den Erwartungs- oder Modalwert der a-posteriori-Verteilung. Vielfach, und dies gilt u. a für räumliche Abhängigkeitsstrukturen, werden die Modelle jedoch so komplex, daß a-posteriori-Verteilungen analytisch oder numerisch nicht mehr bestimmbar sind. In diesem Fall läßt sich die a-posteriori-Verteilung mit Hilfe von *Markov-Chain-Monte-Carlo*-Verfahren (MCMC; Robert u. Casella, 1999) approximieren. Der Begriff MCMC beschreibt eine Sammlung von Methoden für die Simulation aus komplexen multivariaten Verteilungen, insbesondere aus solchen, deren Wahrscheinlichkeitsdichtefunktion analytisch nicht herzuleiten ist. Kernidee der MCMC-Algorithmen (z. B. Gibbs Sampler, Metropolis Hastings) ist, eine Markov-Kette zu erstellen, deren Übergangswahrscheinlichkeiten analytisch beschreibbar sind und die die gewünschte multivariate Verteilung als Grenzverteilung besitzt. Durch eine genügend große Anzahl an Wiederholungen dieser Kette läßt sich dann die benötigte Verteilung simulieren. Die Anwendung von MCMC im Zusammenhang mit GLMMs mit räumlicher Kovarianzstruktur wird z. B. in Christensen et al. (2000), Christensen u. Waagepetersen (2002), Diggle et al. (1998) und Zhang (2002) beschrieben. Die Autoren verwenden und diskutieren unterschiedliche MCMC-Algorithmen und a-priori-Verteilungen und geben zahlreiche Beispiele.

In Diggle et al. (1998) und Christensen u. Waagepetersen (2002) werden Anzahldaten mit einem hohen Anteil von Nullen analysiert. Ohne daß dies explizit von den Autoren erwähnt wird, lassen die Ergebnisse vermuten, daß die Parameterschätzungen der festen Effekte in GLMMs (in Verbindung mit Bayes-Methoden) weniger stark von den Nullen dominiert werden als dies, wie in dieser Arbeit beobachtet, für das marginale GLM der Fall ist.

Zur Umsetzung der rechenintensiven MCMC-Verfahren haben sich im wesentlichen zwei frei erhältliche Software-Produkte durchgesetzt: BayesX[3] des Instituts für Statistik der LMU München und das u. a. von der MRC Biostatistics Unit, Cambridge entwickelte WinBUGS[4].

Ein weiterer Ansatz, der gezielt auf die Analyse von Anzahldaten mit vielen Nullen abzielt ist der der sog. *Zero-Inflated-Poisson-Process*-Modelle (ZIP). Die erste Formulierung von Lambert (1992) modelliert dabei die Verteilung einer Zufallsvariablen Y_i durch

$$Y_i \sim \begin{cases} 0, & \text{mit Wahrscheinlichkeit } p_i \\ \text{Poisson}(\mu_i), & \text{mit Wahrscheinlichkeit } 1 - p_i, \end{cases}$$

[3] http://www.stat.uni-muenchen.de/~bayesx/bayesx.html
[4] http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml

wobei p über das logit- und μ über das log-lineare Modell geschätzt wird. Es handelt sich also um eine Mischung aus Binomial- und Poisson-Modell, in denen p und μ in Abhängigkeit identischer oder auch unterschiedlicher Kovariaten modelliert werden. In einer aktuellen Arbeit modellieren Rathbun u. Fei (2006) die Anzahl von Eichensämlingen mit Hilfe eines räumlichen ZIP-Modells. Die Autoren verwenden hierfür ein räumliches Probit-Modell in Verbindung mit einem gewöhnlichen log-linearen Modell. Die Schätzung der Parameter der festen Effekte des Probit- und log-linearen Modells sowie der Kovarianzparameter erfolgt über MCMC-Verfahren.

Insbesondere bei den angesprochenen Methoden der Bayes-Statistik und des räumlichen ZIP-Modells handelt es sich um vielversprechende Entwicklungen, die sich gut auf das Count-Modell von Hedley (2000) übertragen lassen. Insbesondere dürften so verläßlichere Aussagen über die Varianz der Parameterschätzungen und Vorhersagen ermöglicht werden.

8. Zusammenfassung

Motiviert durch Wildzählungen, die mit Hilfe des Linientransektverfahrens in einigen namibischen Conservancies durchgeführt werden, verfolgt diese Arbeit drei Ziele:

1. Allgemeine Formulierung und Anwendung der Methoden und Modelle des Linientransektverfahren sowie darauf aufbauend zusätzlicher Analysemöglichkeiten zur Überprüfung der Verläßlichkeit von Dichteschätzungen, die sich bei Verwendung von Farmwegen als Transekte ergeben (Teilprojekt 1).

2. Allgemeine Formulierung und Anwendung von Methoden und Modellen, die den Einbezug zusätzlicher räumlicher Informationen in die Linientransektanalyse sowie räumliche Dichte- oder Abundanzvorhersagen ermöglichen (Teilprojekt 2).

3. Entwicklung von Software-Modulen innerhalb von Standard-Software, die eine weitestgehend automatisierte und zentral gesteuerte computer-technische Umsetzung der ersten beiden Teilprojekte ermöglichen (Teilprojekt 3).

Teilprojekt 1 Conservancies sind Zusammenschlüsse benachbarter Farmen mit dem Ziel einer gemeinsamen und nachhaltigen Nutzung der natürlichen Ressourcen und hierbei insbesondere der endemischen Wildbestände. In mehreren solcher Conservancies werden seit mehreren Jahren Wildzählungen mit Hilfe des Linientransektverfahrens zur Erfassung der aktuellen Wildbestände endemischer Arten durchgeführt. Bei diesem zur Familie des Distance Sampling gehörenden Verfahren werden von in das Untersuchungsgebiet gelegten Linien (Transekten) die senkrechten Entfernungen zwischen Transekt und allen hiervon entdeckten Untersuchungsobjekten erfaßt. Anhand dieser Entfernungen läßt sich die Wahrscheinlichkeit, ein Objekt in einer bestimmten Entfernung vom Transekt zu entdecken, modellieren (Entdeckungsfunktion) und damit auch ein Korrekturfaktor für nicht entdeckte Objekte schätzen. Wie in vielen Stichprobenverfahren erfordert die Verläßlichkeit der resultierenden Dichte- und Abundanzschätzungen auch hier die Unabhängigkeit und zufällige Auswahl der Stichprobeneinheiten (Transekte). Im Fall der namibischen Wildzählungen werden jedoch oftmals Farmwege als Transekte verwendet, da das Befahren der Flächen abseits nur selten möglich ist. Das erste Teilprojekt sollte daher die Frage klären, inwiefern diese Zählungen dennoch verläßliche Ergebnisse liefern. Hierfür wurde ein Farmkomplex im zentralen Osten Namibias als Untersuchungsgebiet ausgewiesen und auf diesem wiederholte Aufnahmen nach drei Schemata durchgeführt. Das erste – Zählung per Auto von Farmwegen aus (WegAuto) – entsprach dem herkömmlichen Zählverfahren, für das zweite wurde ein Stichprobendesign mit

systematisch angeordneten Transekten verwendet, die per Pferd abgeritten wurden (BuschPferd). Im dritten Aufnahmeschema wurden – ebenfalls per Pferd – die Wegtransekte abgeritten (WegPferd), um zwischen Effekten des Weges und der Zählplattform (Auto/Pferd) unterscheiden zu können. Im Zuge der Aufnahmen wurden die fünf Wildarten Großer Kudu (*Tragelaphus strepsiceros*), Oryx (*Oryx gazella*), Hartebeest (*Alcelaphus buselaphus caama*), Steinbock (*Rhaphicerus campestris*) und Warzenschwein (*Phacochoerus aethiopicus*) erfaßt. Wenngleich nur sehr begrenzt statistisch abgesichert, zeigten die Vergleiche der Dichteschätzungen recht deutliche Unterschiede. So lagen die Dichteschätzungen der drei Arten Hartebeest, Oryx und Steinbock für die beiden Weg-Schemata deutlich unter denen des BuschPferd-Schemas; ein Hinweis darauf, daß auf Wegtransekten basierende Dichteschätzungen für diese Arten negativ verzerrt sind, evtl. begründet durch Beunruhigung und Jagddruck. Für Kudus hingegen ließen sich kaum Unterschiede zwischen den drei Schemata zeigen, für Warzenschweine lediglich zählplattformabhängige. Hier wiesen beide Pferd-Schemata vergleichbare Dichteschätzungen auf, die des Auto-Schemas hingegen fielen deutlich ab. Dies läßt sich als Anzeichen für eine frühe, unbeobachtete Fluchtreaktion der Warzenschweine auf das Auto interpretieren. Die Dichteschätzungen sind in diesem Fall ebenfalls negativ verzerrt.

Der Bearbeitung der konkreten namibischen Wildzählungsdaten geht eine allgemeine Beschreibung der Modelle des Linientransektverfahrens vorher. Des weiteren werden zusätzliche Analyseoptionen wie z. B. die Auswertung wiederholter Aufnahmen permanent angelegter Transekte, die Berechnung senkrechter Entfernungen für nicht gerade verlaufende Transekte und ein statistischer Test zum Vergleich von Dichteschätzungen erläutert.

Teilprojekt 2 Das zweite Teilprojekt widmete sich der auf Linientransektaufnahmen basierenden räumlichen Analyse und lokalen Vorhersagen von Wilddichten. Dabei wurde ein Ansatz von Hedley (2000) aufgegriffen, in dem Streifen, deren Breite abhängig von der geschätzten Entdeckungsfunktion bestimmt wird, um die Transekte gelegt und diese im Anschluß in Segmente etwa gleicher Länge unterteilt werden. Für die einzelnen Segmente werden dann die Anzahl der darin gemachten Beobachtungen einerseits und die Werte zusätzlicher (räumlicher), die lokale Wilddichte potentiell erklärender Kovariaten andererseits zugewiesen. Über ein log-lineares Modell läßt sich dann die Anzahl der Beobachtungen je Segment in Abhängigkeit von den ermittelten Kovariaten modellieren. Die resultierenden Parameterschätzungen lassen sich im Anschluß für die Vorhersage der Dichte oder Abundanz in nicht erfaßten Bereichen des Untersuchungsgebietes nutzen. Das Verfahren kann so z. B. dazu dienen, ökologische Zusammenhänge zwischen lokaler Wilddichte und Habitateigenschaften aufzuzeigen.

Der Ansatz von Hedley (2000) wurde zudem dahingehend erweitert, die zu erwartende räumliche Autokorrelation zwischen den Segmenten sowohl in der Parameterschätzung als auch in der Vorhersage neuer Beobachtungen zu berücksichtigen. Hierbei kamen Ansätze von Albert u. McShane (1995), Gotway u. Stroup (1997) und Gotway u. Wolfinger (2003) zum Einsatz, die sich dem Einbezug räumlicher Autokorrelation in verallgemeinerte lineare Modelle (GLMs)

und der Anwendung des aus der Geostatistik bekannten Krigings widmen.

Die Anwendung der Methoden erfolgte auf die Daten des BuschPferd-Schemas aus Teilprojekt 1. Hierfür wurde zusätzlich eine umfassende Geodatenbank aufgebaut, die als Grundlage für die zu bestimmenden räumlichen Kovariaten diente. Berechnungen wurden sowohl für das ursprüngliche Modell von Hedley (2000) als auch für das mit Berücksichtigung räumlicher Autokorrelation durchgeführt. Anhand lokaler Vorhersagen für das gesamte Untersuchungsgebiet wurden zudem die Gesamtdichten der fünf Wildarten bestimmt und mit den Schätzungen des klassischen Linientransektverfahrens verglichen. Für alle fünf Arten lieferte das Modell ohne Autokorrelation wenig realistische Wilddichteverteilungen, da jeweils nur wenige Kovariaten als signifikant ausgewiesen werden konnten. Ursache hierfür waren im wesentlichen geringe Beobachtungsanzahlen und ein evtl. zu kleines Untersuchungsgebiet, in dem die Werte der Kovariaten nur kleine Spannweiten besitzen. Dem Modell mit Autokorrelation hingegen gelang es, lokale Vorhersagen zu erstellen, die sich stark an den tatsächlichen Beobachtungen orientierten und damit sehr realistische Wilddichteverteilungen ergaben. Der Einfluß der Kovariaten konnte daher im Vergleich zu dem der räumlichen Autokorrelation als zu vernachlässigen betrachtet werden. Die Schätzungen der Gesamtdichte waren bis auf wenige Ausnahmen für beide Ansätze ähnlich und auch vergleichbar mit den Schätzungen der klassischen Linientransektanalyse. Die zugehörigen Standardfehler konnten gegenüber der klassischen Analyse durch den Einbezug zusätzlicher erklärender Kovariaten reduziert werden.

Die Darstellung der verwendeten Methoden umfaßt, wiederum gänzlich unabhängig von den konkreten namibischen Daten, u. a. die allgemeine Beschreibung verallgemeinerter lineare Modelle, mit besonderer Betrachtung des log-linearen Modells, Möglichkeiten der Berücksichtigung korrelierter und speziell räumlich korrelierter Beobachtungen sowie Ansätze zur räumlichen Vorhersage neuer Beobachtungen über Kriging-ähnliche Verfahren.

Teilprojekt 3 Die Umsetzung der erforderlichen Modelle und Methoden erfolgte mit Hilfe der Software-Produkte Distance (klassische Linientransektanalyse), ArcGIS (Erstellung, Verwaltung und Bearbeitung der Geodaten) und SAS (zentrale Steuerung, Datenmanagement, statistische Auswertung und Modellierung). In Form von SAS-Makros und in Python programmierten sog. Geoprocessing-Skripten (ArcGIS) wurden zahlreiche zusätzliche Module entwickelt, die einen weitestgehend automatisierten Ablauf der Arbeitsschritte und die zentrale Steuerung aller erforderlichen Operationen aus SAS heraus ermöglichten. Mit Hilfe der Geoprocessing-Skripte war beispielsweise die Berechnung der kürzesten Entfernung einer Beobachtung zum zugehörigen Transekt, die Generierung der Segmente, deren räumliche Verschneidung mit beliebigen anderen räumlichen Informationen zur Bestimmung der Kovariaten und die Speicherung der lokalen Vorhersagen in einem Geodatenformat zur anschließenden kartographischen Darstellung möglich. SAS-Makros dienten vorwiegend der zentralen Steuerung der beiden anderen Programme und der Verarbeitung und Zusammenführung von deren Ergebnissen. Darüber hinaus fand hier die gesamte statistische Mo-

dellierung statt. Hierfür wurden zum Teil SAS-eigene Prozeduren verwendet aber auch eigene Makros programmiert; so z. B. für den T-Test zum Vergleich der Dichteschätzungen der Aufnahmeschemata in Teilprojekt 1, für die automatisierte Variablenselektion zur Ausweisung signifikanter Kovariaten in Teilprojekt 2 und für die Berechnung lokaler Vorhersagen unter Berücksichtigung räumlicher Autokorrelation.

Alle Makros und Skripte wurden gezielt so allgemein entwickelt, daß ihre Anwendung auch in anderen Projekten mit dem Ziel der räumlichen oder auch nicht-räumlichen Analyse von Linientransektaufnahmen ohne weiteres möglich ist. Die insgesamt 27 Makros und 11 Skripte sind unter http://www.uni-forst.gwdg.de/~fmader/promotion/smalts/ erhältlich.

A. Notation und verwendete Abkürzungen

Zufallsvariablen werden durch große, kursiv gedruckte lateinische Buchstaben (z. B. $Y_1, \ldots, Y_n$), ihre Realisierungen durch die entsprechenden Kleinbuchstaben ($y_1, \ldots, y_n$) gekennzeichnet. Griechische Buchstaben wie β geben in der Regel Parameter, mit einem Dach versehen ($\hat{\beta}$) Parameterschätzungen an. Fettgedruckte Klein- bzw. Großbuchstaben sind Vektoren (z. B. $\boldsymbol{y}$) bzw. Matrizen (z. B. $\boldsymbol{X}$) zufälliger oder nicht-zufälliger Variablen.

Abkürzungen

AIC	Akaikes Informationskriterium (Akaike's Information Criterion)
GAM	Verallgemeinertes additives Modell (Generalized Additive Model)
GEE	Verallgemeinerte Schätzfunktion (Generalized Estimating Equation)
GLM	Verallgemeinertes lineares Modell (Generalized Linear Model)
GLMM	Verallgemeinertes lineares gemischtes Modell (Generalized Linear Mixed Model)
GLS	Verallgemeinerte kleinste Quadrate (Generalized Least Squares)
GPS	Global Positioning System
IWLS	Iterativ gewichtete kleinste Quadrate (Iteratively Weighted Least Squares)
LCL	Untere Konfidenzgrenze (Lower Confidence Limit)
MCDS	Multi-kovariates Distance Sampling (Multicovariate Distance Sampling)
ML	Maximum-Likelihood
MSPE	Mittlerer quadratischer Vorhersagefehler (Mean Squared Prediction Error)
NDVI	Normalized Difference Vegetation Index
pGDB	Personal Geodatabase
PL	Pseudo-Likelihood
QL	Quasi-Likelihood
REPL	Restricted-Pseudo-Likelihood
RMSPE	Standardfehler der Vorhersage (Root Mean Squared Prediction Error)
UCL	Obere Konfidenzgrenze (Upper Confidence Limit)
ZIP	Zero Inflated Poisson Process

Kapitel 2.1 Das Linientransektverfahren

φ	Beobachtungswinkel zwischen Transekt und Objektposition
θ	Zu schätzender Modellparameter der Entdeckungsfunktion

υ	(Halbe) effektive Streifenbreite ($= \int_0^{w_r} g(x)\mathrm{d}x$)
ω	Beobachtungswinkel zwischen magnetisch Nord und Objektposition
a	Stichprobenfläche bei geraden Transekten ($= 2 \cdot L \cdot \upsilon$)
$\tilde{a}$	Tatsächliche Stichprobenfläche
A	Größe des Untersuchungsgebietes
D	Individuendichte
D_s	Objektdichte
$\mathrm{E}(s)$	Mittlere Gruppengröße
$f(x)$	Wahrscheinlichkeitsdichtefunktion von $g(x)$ ($= g(x)/\int_0^{w_r} g(x)\mathrm{d}x$)
$g(x)$	Entdeckungs(wahrscheinlichkeits)funktion. Gibt die Wahrscheinlichkeit an, ein Objekt in Entfernung x vom Transekt zu entdecken
H	Anzahl der Straten
k	Anzahl der Transekte
ℓ	log-Likelihood-Funktion
l	Länge eines Transektes
L	Gesamtlänge der Transekte ($= \sum_{i=1}^{k} l_i$)
$\mathcal{L}$	Likelihood-Funktion
M	Individuenanzahl/-abundanz innerhalb des Untersuchungsgebietes ($= D \cdot A = D_s \cdot \mathrm{E}(s) \cdot A$)
n	Anzahl der Beobachtungen. Mit Index i Anzahl der Beobachtungen auf dem i-ten Transekt, $i = 1, \ldots, k$
N	Objektanzahl/-abundanz innerhalb des Untersuchungsgebietes ($= D_s \cdot A$)
p	Anzahl der zu schätzenden Modellparameter
P_a	Mittlere Entdeckungswahrscheinlichkeit ein Objekt bis zur Entfernung w_r zu entdecken ($= \int_0^{w_r} g(x)\mathrm{d}x/w_r$)
s	Gruppengröße. Anzahl der Individuen eines beobachteten Objektes
w	Halbe Streifenbreite
w_l	Linke Kappungsentfernung
w_r	Rechte Kappungsentfernung
x	Senkrechte, kürzeste Entfernung zwischen Objektposition und Transekt
y	Entfernung zwischen Beobachtungs- und Objektposition

Kapitel 2.2 Verallgemeinerte lineare Modelle (GLM)

α	Kovarianzparameter des räumlichen Prozesses
β	Modellparameter
χ^2	Pearson-Statistik
χ^{2*}	Skalierte Pearson-Statistik
ϵ	Residuum ($= y - \hat{\mu}$)
ϕ	Skalierungs- oder Dispersionsparameter
η	Linearer Prädiktor ($= g(\mu) = \boldsymbol{x}'\boldsymbol{\beta}$)
μ	Erwartungswert von y
$\tilde{\mu}$	Erwartungswert von y unter Einbezug einer Offset-Variablen t
ν	Pseudo-Variable ($= y^*$)
θ	Natürlicher Parameter

$\rho(\cdot)$	Korrelationsfunktion des räumlichen Prozesses
σ_n^2	Nugget-Effekt
σ_o^2	Überdispersionsparameter
ψ	Skalierungs-, Dispersions- oder Überdispersionsparameter, d. h. ϕ oder σ_o^2
D	Devianz
D^*	Skalierte Devianz
$g(\cdot)$	Link-Funktion
$g^{-1}(\cdot)$	Inverse Link-Funktion ($= \mu$)
ℓ	log-Likelihood-Funktion
$\mathcal{L}$	Likelihood-Funktion
n	Anzahl der Beobachtungen
$r^{(A)}$	Anscombe-Residuum
$r^{(P)}$	Pearson-Residuum
$r^{(P*)}$	Student-standardisiertes Pearson-Residuum
t	Offset-Variable
$v(\cdot)$	Varianzfunktion
x	Kovariate
y	Abhängige, zu modellierende Variable
y^*	Linearisierte abhängige Variable ($= \nu$)

Kapitel 2.3 Das Count-Modell

β	Modellparameter
υ	(Halbe) effektive Streifenbreite ($= \int_0^{w_r} g(x)\mathrm{d}x$)
ν	Pseudo-Variable ($= y^*$)
$\rho(\cdot)$	Korrelationsfunktion des räumlichen Prozesses
σ_n^2	Nugget-Effekt
σ_o^2	Überdispersionsparameter
a	Stichprobenfläche bei geraden Transekten ($= 2 \cdot L \cdot \upsilon$)
$\tilde{a}$	Tatsächliche Stichprobenfläche
A	Größe des Untersuchungsgebietes
b	Größe einer Block-Kriging-Diskretisierungsfläche
B	Größe eines Block-Kriging-Blocks
c	Länge eines Segments
D	Individuendichte
D_s	Objektdichte
$\mathrm{E}(s)$	Mittlere Gruppengröße
f	Größe einer Vorhersagezelle
k	Anzahl der Transekte
ℓ	log-Likelihood-Funktion
l	Länge eines Transektes
L	Gesamtlänge der Transekte ($= \sum_{i=1}^{k} l_i$) und der Segmente ($= \sum_{i=1}^{T} c_i$)
$\mathcal{L}$	Likelihood-Funktion
m	Anzahl der Individuen innerhalb eines Segments oder einer Vorhersagezelle

$\tilde{m}$	Horvitz-Thompson(-ähnlicher) Schätzer der Anzahl der in einem Segment vorhandenen Individuen
M	Individuenanzahl/-abundanz innerhalb des Untersuchungsgebietes ($= D \cdot A = D_s \cdot \mathrm{E}(s) \cdot A$)
n	Anzahl der Objekte (Gruppen) innerhalb eines Segments oder einer Vorhersagezelle
$\tilde{n}$	Horvitz-Thompson(-ähnlicher) Schätzer der Anzahl der in einem Segment vorhandenen Objekte (Gruppen)
n_B	Anzahl der Block-Kriging-Diskretisierungsflächen
n_Z	Anzahl der Vorhersagezellen
N	Objektanzahl/-abundanz innerhalb des Untersuchungsgebietes ($= D_s \cdot A$)
p	Segment- (p_i) oder objektspezifische (p_{ij}) Entdeckungswahrscheinlichkeit
P_a	Mittlere Entdeckungswahrscheinlichkeit ein Objekt bis zur Entfernung w_r zu entdecken ($= \int_0^{w_r} g(x)\mathrm{d}x/w_r$)
s	Gruppengröße. Anzahl der Individuen eines beobachteten Objektes
T	Anzahl der Segmente
w	Halbe Streifenbreite
w_r	Rechte Kappungsentfernung
x	(Räumliche) Kovariate

B. Modelle und Methoden

B.1. Darstellungsformen der erwarteten Fisher Informationsmatrix

Nach (2.26) ist die erwartete Fisher Informationsmatrix durch

$$\boldsymbol{F}_{exp} = \boldsymbol{X}'\boldsymbol{W}_{exp}\boldsymbol{X} \tag{B.1}$$

gegeben, wobei $\boldsymbol{X}$ die $n \times p$ Designmatrix und $\boldsymbol{W}_{exp}$ die $n \times n$ Diagonalmatrix mit Elementen $\operatorname{diag}[w_{exp_{ii}}]$ aus (2.25) ist. $\boldsymbol{W}_{exp}$ läßt sich zerlegen in

$$\boldsymbol{W}_{exp} = \boldsymbol{B}'\boldsymbol{V}^{-1}\boldsymbol{B}, \tag{B.2}$$

mit den Diagonalmatrizen

$$\boldsymbol{B} = \begin{bmatrix} \frac{\partial \mu_1}{\partial \eta_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial \mu_n}{\partial \eta_n} \end{bmatrix} \quad \text{und} \quad \boldsymbol{V} = \begin{bmatrix} \operatorname{var}(y_i) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \operatorname{var}(y_n) \end{bmatrix}.$$

(B.2) in (B.1) eingesetzt, vereinfacht sich $\boldsymbol{X}'\boldsymbol{B}'$ zu

$$\boldsymbol{X}'\boldsymbol{B}' = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{p1} & \cdots & x_{pn} \end{bmatrix} \begin{bmatrix} \frac{\partial \mu_1}{\partial \eta_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\partial \mu_n}{\partial \eta_n} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mu_1}{\partial \beta_1} & \cdots & \frac{\partial \mu_n}{\partial \beta_1} \\ \vdots & & \vdots \\ \frac{\partial \mu_1}{\partial \beta_p} & \cdots & \frac{\partial \mu_n}{\partial \beta_p} \end{bmatrix} = \boldsymbol{D}'.$$

Entsprechend vereinfacht sich $\boldsymbol{B}\boldsymbol{X}$ zu $\boldsymbol{D}$. (B.1) läßt sich also auch darstellen als

$$\boldsymbol{F}_{exp} = \boldsymbol{D}'\boldsymbol{V}^{-1}\boldsymbol{D}.$$

B.2. Fisher Scoring als IWLS

Der Lösungsalgorithmus des Fisher Scoring zur Schätzung des Parametervektors $\boldsymbol{\beta}$ ist

$$\hat{\boldsymbol{\beta}}_{(m+1)} = \hat{\boldsymbol{\beta}}_{(m)} + \boldsymbol{F}_{exp_{(m)}}^{-1}\boldsymbol{s}_{(m)}.$$

Multiplizieren beider Seiten mit $\boldsymbol{F}_{exp_{(m)}}$ ergibt

$$\boldsymbol{F}_{exp_{(m)}}\hat{\boldsymbol{\beta}}_{(m+1)} = \boldsymbol{F}_{exp_{(m)}}\hat{\boldsymbol{\beta}}_{(m)} + \boldsymbol{s}_{(m)}.$$

Die linke Seite dieser Gleichung läßt sich darstellen als $\boldsymbol{X}'\boldsymbol{W}_{exp_{(m)}}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{(m+1)}$ (vergl. (2.26)), die rechte Seite als $\boldsymbol{X}'\boldsymbol{W}_{exp_{(m)}}\boldsymbol{X}\hat{\boldsymbol{\beta}}_{(m)} + \boldsymbol{s}_{(m)}$. Hierbei handelt

es sich um einen $1 \times p$ Vektor mit k-tem Element

$$\begin{aligned}
& \sum_{j=1}^{p}\sum_{i=1}^{n} \frac{x_{ij}\, x_{ik}}{\mathrm{var}(y_i)} \left(\frac{\partial \hat{\mu}_i}{\partial \hat{\eta}_i}\right)^2_{(m)} \hat{\boldsymbol{\beta}}_{j_{(m)}} + \frac{x_{ij}}{\mathrm{var}(y_i)} \left(\frac{\partial \hat{\mu}_i}{\partial \hat{\eta}_i}\right)_{(m)} (y_i - \hat{\mu}_{i_{(m)}}) \\
= & \sum_{j=1}^{p}\sum_{i=1}^{n} w_{exp_{i_{(m)}}} x_{ij}\, x_{ik} + \sum_{i=1}^{n} \hat{w}_{exp_{i_{(m)}}} (y_i - \hat{\mu}_{i_{(m)}}) \left(\frac{\partial \hat{\eta}_i}{\partial \hat{\mu}_i}\right)_{(m)} x_{ij}.
\end{aligned}$$

Dies läßt sich auch darstellen als

$$\boldsymbol{X}'\boldsymbol{W}_{exp_{(m)}}\boldsymbol{y}^*,$$

worin $\boldsymbol{W}_{exp_{(m)}}$ eine Diagonalmatrix mit den in (2.25) definierten Elementen ist und

$$\boldsymbol{y}^* = \hat{\boldsymbol{\eta}}_{(m)} + (\boldsymbol{y} - \hat{\boldsymbol{\mu}}_{(m)}) \left(\frac{\partial \hat{\boldsymbol{\eta}}}{\partial \hat{\boldsymbol{\mu}}}\right)_{(m)}.$$

B.3. Darstellungsformen des Prädiktors im räumlichen GLM

Die beiden auf Basis eines verallgemeinerten linearen Modells mit räumlich autokorrelierten Fehlern vorgestellten Prädiktoren aus (2.64) und (2.68) sind

$$\begin{aligned}
\hat{\boldsymbol{y}}_0 &= \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Sigma}_{y_0 y}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}) \qquad \text{und} && \text{(B.3)}\\
\hat{\boldsymbol{y}}_0 &= \hat{\boldsymbol{\mu}}_0 + \hat{\boldsymbol{\Delta}}_0^{-1}(\hat{\boldsymbol{\nu}}_0 - \hat{\boldsymbol{\eta}}_0), && \text{(B.4)}
\end{aligned}$$

mit $\hat{\boldsymbol{\mu}} = g^{-1}(\boldsymbol{x}'\hat{\boldsymbol{\beta}})$. $\hat{\boldsymbol{\nu}}_0$ ist nach (2.66) gegeben durch

$$\boldsymbol{\nu}_0 = \boldsymbol{X}_0\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}}).$$

Eingesetzt in (B.4) ergibt sich

$$\begin{aligned}
\hat{\boldsymbol{y}}_0 &= \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Delta}_0^{-1}\left[\boldsymbol{X}_0\hat{\boldsymbol{\beta}} + \boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) - \hat{\boldsymbol{\eta}}_0\right]\\
&= \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Delta}_0^{-1}\boldsymbol{\Sigma}_{\nu\nu_0}\boldsymbol{\Sigma}_{\nu\nu}^{-1}(\boldsymbol{\nu} - \boldsymbol{X}\hat{\boldsymbol{\beta}}). && \text{(B.5)}
\end{aligned}$$

Laut Definition der Pseudo-Variablen aus (2.52) ist

$$\boldsymbol{\nu} = \hat{\boldsymbol{\eta}} + \boldsymbol{\Delta}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}).$$

Dies eingesetzt in (B.5) ergibt

$$\begin{aligned}
\hat{\boldsymbol{y}}_0 &= \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Delta}_0^{-1}\boldsymbol{\Sigma}_{\nu\nu_0}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\left[\hat{\boldsymbol{\eta}} + \boldsymbol{\Delta}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}) - \boldsymbol{X}\hat{\boldsymbol{\beta}}\right]\\
&= \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Delta}_0^{-1}\boldsymbol{\Sigma}_{\nu\nu_0}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Delta}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}). && \text{(B.6)}
\end{aligned}$$

$\boldsymbol{\Sigma}_{y_0 y}\boldsymbol{\Sigma}_{yy}^{-1}$ aus (B.3) ist identisch zu $\boldsymbol{\Delta}_0^{-1}\boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Delta}$ aus (B.6), da

$$\boldsymbol{\Sigma}_{\nu_0\nu} = \boldsymbol{\Delta}_0\boldsymbol{\Sigma}_{y_0 y}\boldsymbol{\Delta} \qquad \text{und} \qquad \boldsymbol{\Sigma}_{\nu\nu} = \boldsymbol{\Delta}\boldsymbol{\Sigma}_{yy}\boldsymbol{\Delta},$$

(vergl. (2.54)) und damit

$$\begin{aligned}
\boldsymbol{\Delta}_0^{-1}\boldsymbol{\Sigma}_{\nu_0\nu}\boldsymbol{\Sigma}_{\nu\nu}^{-1}\boldsymbol{\Delta} &= \boldsymbol{\Delta}_0^{-1}\boldsymbol{\Delta}_0\boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Delta}\left(\boldsymbol{\Delta}\boldsymbol{\Sigma}_{yy}\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta} \\
&= \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Delta}\boldsymbol{\Delta}^{-1}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Delta}^{-1}\boldsymbol{\Delta} \\
&= \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}.
\end{aligned}$$

Insgesamt ergibt sich also

$$\begin{aligned}
\hat{\boldsymbol{y}}_0 &= \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Sigma}_{y_0y}\boldsymbol{\Sigma}_{yy}^{-1}(\boldsymbol{y} - \hat{\boldsymbol{\mu}}) \\
&= \hat{\boldsymbol{\mu}}_0 + \boldsymbol{\Delta}_0^{-1}\left[\hat{\boldsymbol{\nu}}_0 - \hat{\boldsymbol{\eta}}_0\right].
\end{aligned}$$

C. Anwendungssoftware

C.1. Berechnung der Objektpositionen

Werden im Feld die Entfernung zwischen Objekt- und Beobachtungsposition y und der Winkel zwischen Objekt und magnetisch Nord ω erfaßt, so ergibt sich die Korrektur der aktuellen örtlichen Deklination (Nadelmißweisung) d durch

$$\omega_2 = (\omega + d)\%360,$$

wobei % den Modulo-Operator bezeichnet. Negative Werte für d stellen westliche, positive östliche Deklinationen dar (geographisch Nord liegt westlich/östlich des magnetischen Nordpols).

Die Umrechnung von geographischen Winkeln zu mathematischen erfolgt über

$$\omega_3 = (360 - \omega_2 + 90)\%360.$$

Die Umrechnung ist erforderlich, da die im Feld per Kompaß gemessenen geographischen Winkel von Norden (0°) aus mit dem Uhrzeigersinn, mathematische Winkel hingegen von Osten (0°) aus gegen den Uhrzeigersinn gemessen werden.

Da die meisten Software-Produkte zur Berechnung trigonometrischer Funktionen Winkelangaben im Bogenmaß anstelle von Grad erfordern, ist ggf. eine weitere Umrechnung erforderlich.

$$\omega_4 = \omega_3 \frac{\pi}{180}$$

x- und y-Koordinate der Objektposition sind dann schließlich gegeben durch

$$\begin{aligned} ObjX &= ObsX + y \cdot \cos(\omega_4) \text{ und} \\ ObjY &= ObsY + y \cdot \sin(\omega_4), \end{aligned}$$

wobei $ObsX$ und $ObsY$ die kartesischen, d. h. projizierten, x- und y-Koordinaten der Beobachtungsposition sind.

C.2. Probleme der Segmenterstellung in ArcGIS

Wird die ArcGIS-Funktion `Buffer` zur Erstellung von Segmenten von kurvigen Transekten verwendet, so können hierbei inkorrekte Segmentgeometrien entstehen. Ursache hierfür ist die Art und Weise, in der die berechneten temporären Segmentvertices miteinander verbunden werden, um die Flächen zu erstellen. Abbildung C.1 verdeutlicht das Vorgehen von ArcGIS am Beispiel zweier Segmente. Die Fläche eines Segments (oder allgemein eines *Buffers*) wird nicht

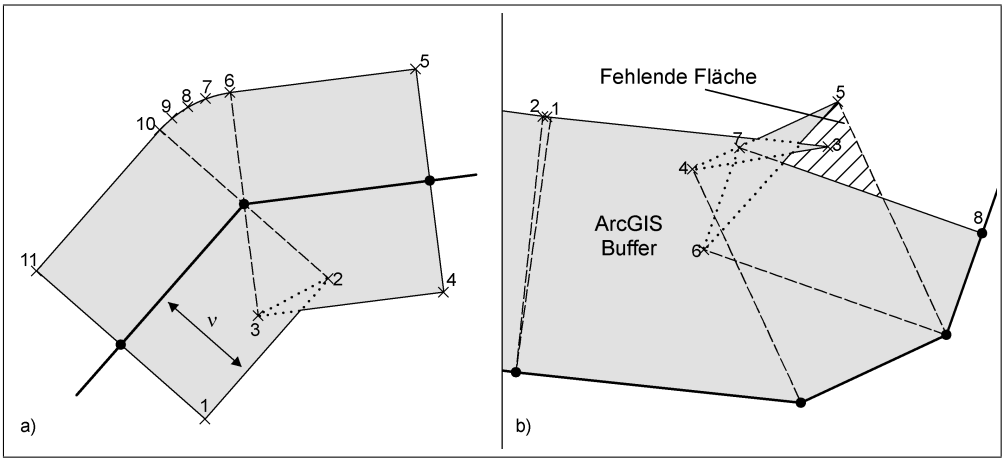

Abbildung C.1.: Probleme der Segmentgenerierung bei Verwendung der ArcGIS `Buffer`-Funktion. Die Segmentgeometrie wird durch die Außenlinie definiert, die sich durch Verbinden der temporären, in Entfernung $\hat{v}$ zur Segmentlinie liegenden Vertices ergibt (a). Bei in Kurven endenden Segmenten kann dies dazu führen, daß Teilflächen, deren Entfernung zur Linie kleiner ist als $\hat{v}$, dennoch nicht in die Buffer-Geometrie eingehen (gestreifte Fläche in b).

durch das Zusammenfügen von Teilflächen erreicht, sondern durch einen einzelnen Polygonzug definiert. Hierfür werden, dem Verlauf der Segmentlinie folgend, für jeden Linienvertex links- und rechtsseitig in der (senkrechten) Entfernung $\hat{v}$ temporäre Punkte erstellt (Punkte 1-6, 10 und 11 in Abb. C.1 a), die dann reihum miteinander verbunden werden. Ändert die Transektlinie in einem Vertex ihre Richtung, so werden auf der „offenen“ Seite zusätzliche temporäre Punkte in Entfernung $\hat{v}$ von diesem Vertex erstellt (Punkte 7-9), die den Verlauf der Außenlinie abrunden. Die von diesem Polygonzug eingeschlossene Fläche wird anschließend als Segment- oder allgemein als Buffer-Geometrie erzeugt. Sich schneidende Linien innerhalb der Fläche werden dabei aufgelöst, so daß es nicht zu Überlappungsbereichen kommt. Wie Abbildungen C.1 a und 3.2 zeigen, führt dies in der Regel auch zum gewünschten und richtigen Ergebnis. Abbildung C.1 b zeigt jedoch am Beispiel nur der linken Streifenseite eines einzelnen Segments, daß dieses Vorgehen bei in Kurven endenden Segmenten dazu führen kann, daß Teilflächen mit einer Entfernung kleiner $\hat{v}$ dennoch nicht in die Segmentgeometrie eingehen.[1] Die Größe der fehlenden Fläche hängt von der Stärke der Kurvenkrümmung und insbesondere von der Streifenbreite $\hat{v}$ ab.

[1] Die in der Abbildung dargestellte ArcGIS Streifengeometrie wurde als beidseitiger Streifen (Option `Side Type = both` der Funktion `Buffer`) erstellt und der Teil rechts des Transektes manuell entfernt. Die Option `Side Type = left` des Befehls zur Erstellung nur des linksseitigen Streifenteils liefert Ergebnisse, die noch stärker vom Soll abweichen!

D. Ergebnisse

D.1. Validierung der Farmwegzählungen (Teilprojekt 1)

Tabelle D.1.: Nach dem Linientransektverfahren geschätzte Objektdichten ($\hat{D}_s$) und Individuendichten ($\hat{D}$) (Anzahl Gruppen/Tiere je Quadratkilometer) sowie zugehörige 95%ige Konfidenzintervalle (in Klammern) je Aufnahmephase, Schema und Art (HB = Hartebeest, KD = Kudu, OX = Oryx, SB = Steinbock, WS = Warzenschwein).

Phase	Art	Schema	$\hat{D}_s$	$\hat{D}$
1	Hartebeest	BuschPferd	1.03 (0.69, 1.53)	5.27 (3.28, 8.47)
		WegAuto	0.34 (0.20, 0.59)	1.77 (0.96, 3.24)
		WegPferd	0.32 (0.18, 0.56)	1.64 (0.88, 3.03)
	Kudu	BuschPferd	0.90 (0.49, 1.65)	2.41 (1.31, 4.46)
		WegAuto	0.81 (0.50, 1.31)	2.18 (1.32, 3.58)
		WegPferd	1.11 (0.74, 1.67)	2.99 (1.95, 4.58)
	Oryx	BuschPferd	2.64 (1.56, 4.47)	11.20 (6.44, 19.45)
		WegAuto	0.53 (0.33, 0.84)	2.25 (1.36, 3.73)
		WegPferd	0.77 (0.55, 1.08)	3.26 (2.20, 4.83)
	Steinbock	BuschPferd	3.85 (2.39, 6.20)	4.20 (2.60, 6.78)
		WegAuto	1.55 (0.92, 2.62)	1.69 (1.00, 2.86)
		WegPferd	1.41 (0.81, 2.46)	1.54 (0.89, 2.69)
	Warzenschwein	BuschPferd	1.74 (0.93, 3.26)	3.44 (1.82, 6.48)
		WegAuto	0.67 (0.35, 1.29)	1.33 (0.68, 2.57)
		WegPferd	1.62 (1.04, 2.52)	3.20 (2.03, 5.04)
2	Hartebeest	BuschPferd	0.77 (0.41, 1.45)	2.35 (1.20, 4.61)
		WegAuto	0.24 (0.12, 0.50)	0.74 (0.34, 1.60)
		WegPferd	0.28 (0.14, 0.55)	0.86 (0.42, 1.77)
	Kudu	BuschPferd	0.75 (0.38, 1.46)	1.76 (0.89, 3.49)
		WegAuto	1.18 (0.72, 1.91)	2.78 (1.67, 4.61)
		WegPferd	1.00 (0.62, 1.63)	2.37 (1.43, 3.92)
	Oryx	BuschPferd	2.00 (1.24, 3.22)	5.32 (3.22, 8.80)
		WegAuto	0.48 (0.27, 0.83)	1.27 (0.71, 2.28)
		WegPferd	0.45 (0.27, 0.73)	1.19 (0.70, 2.01)
	Steinbock	BuschPferd	3.74 (2.33, 5.99)	4.08 (2.54, 6.55)
		WegAuto	1.16 (0.60, 2.25)	1.27 (0.66, 2.45)
		WegPferd	1.52 (0.80, 2.87)	1.65 (0.87, 3.13)
	Warzenschwein	BuschPferd	1.85 (1.04, 3.30)	3.66 (2.04, 6.57)
		WegAuto	0.74 (0.36, 1.51)	1.46 (0.71, 3.00)
		WegPferd	1.93 (1.20, 3.10)	3.81 (2.33, 6.21)

Tabelle D.2.: Nach dem Linientransektverfahren geschätzte Begegnungsraten (n/L) (Anzahl Gruppen je Kilometer) und Individuenanzahlen ($\hat{M}$) sowie zugehörige 95%ige Konfidenzintervalle (in Klammern) je Aufnahmephase, -schema und Art (HB = Hartebeest, KD = Kudu, OX = Oryx, SB = Steinbock, WS = Warzenschwein). Die Individuenanzahlen beziehen sich auf das Untersuchungsgebiet von etwa 16.1 km^2 Größe.

Phase	Art	Schema	n/L	$\hat{M}$
1	Hartebeest	BuschPferd	0.33 (0.23, 0.49)	85 (53, 137)
		WegAuto	0.09 (0.06, 0.15)	28 (15, 51)
		WegPferd	0.09 (0.06, 0.15)	27 (15, 50)
	Kudu	BuschPferd	0.23 (0.13, 0.40)	38 (21, 70)
		WegAuto	0.16 (0.11, 0.23)	35 (21, 58)
		WegPferd	0.19 (0.13, 0.26)	48 (31, 74)
	Oryx	BuschPferd	0.59 (0.35, 0.99)	180 (104, 313)
		WegAuto	0.13 (0.09, 0.20)	36 (22, 60)
		WegPferd	0.24 (0.18, 0.32)	52 (35, 77)
	Steinbock	BuschPferd	0.20 (0.13, 0.32)	68 (42, 110)
		WegAuto	0.12 (0.08, 0.18)	27 (16, 46)
		WegPferd	0.07 (0.04, 0.13)	25 (14, 44)
	Warzenschwein	BuschPferd	0.19 (0.10, 0.34)	56 (30, 106)
		WegAuto	0.08 (0.05, 0.13)	22 (11, 43)
		WegPferd	0.16 (0.11, 0.22)	52 (33, 82)
2	Hartebeest	BuschPferd	0.25 (0.13, 0.47)	38 (19, 75)
		WegAuto	0.06 (0.03, 0.13)	11 (5, 24)
		WegPferd	0.08 (0.04, 0.15)	14 (7, 29)
	Kudu	BuschPferd	0.19 (0.10, 0.36)	28 (14, 56)
		WegAuto	0.23 (0.16, 0.33)	45 (27, 75)
		WegPferd	0.17 (0.11, 0.26)	38 (23, 63)
	Oryx	BuschPferd	0.45 (0.28, 0.72)	86 (52, 142)
		WegAuto	0.12 (0.07, 0.20)	21 (12, 38)
		WegPferd	0.14 (0.09, 0.22)	20 (12, 34)
	Steinbock	BuschPferd	0.20 (0.13, 0.31)	66 (41, 106)
		WegAuto	0.09 (0.05, 0.16)	20 (10, 39)
		WegPferd	0.08 (0.04, 0.15)	26 (14, 49)
	Warzenschwein	BuschPferd	0.20 (0.11, 0.34)	58 (32, 104)
		WegAuto	0.09 (0.05, 0.16)	23 (11, 47)
		WegPferd	0.18 (0.12, 0.28)	62 (38, 101)

Literaturverzeichnis

Aitkin et al. 1989 Aitkin, M. ; Anderson, D. ; Francis, B. ; Hinde, J.: *Statistical modelling in GLIM.* Oxford : Oxford Science Publications, 1989

Akaike 1973 Akaike, H.: Information theory and an extension of the maximum likelihood principle. In: Petran, B.N. (Hrsg.) ; Csàaki, F. (Hrsg.): *International Symposium on Information Theory.* 2nd ed. Budapest, Hungary : Akadèemiai Kiadi, 1973, S. 267–281

Akaike 1985 Akaike, H.: Prediction and entropy. In: Atkinson, A.C. (Hrsg.) ; Fienberg, S.E. (Hrsg.): *A Celebration of Statistics.* Berlin : Springer, 1985, S. 1–24

Albert u. McShane 1995 Albert, P.S. ; McShane, L.M.: A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data. In: *Biometrics* 51 (1995), Nr. 2, S. 627–638

Alldredge u. Gates 1985 Alldredge, J.R. ; Gates, C.E.: Line transect estimators for left-truncated distributions. In: *Biometrics* 41 (1985), S. 273–280

AMD 1998 University of Rome "La Sapienza" (Hrsg.): *African Mammals Databank – A databank for the conservation and management of the African mammals.* 1998. – http://www.gisbau.uniroma1.it/amd.php. – Zugriffsdatum: 12 .03 .2006

Anscombe 1953 Anscombe, F.J.: Contribution to the Discussion of H. Hotelling's Paper. In: *Journal of The Royal Statistical Society Series B – Statistical Methodology* 15 (1953), S. 229–230

Ashley 1995 Ashley, C.: Tourism, communities, and the potential impacts on local incomes and conservation. / Department of Environmental Affairs (DEA), Ministry of Environment and Tourism (MET). Private Bag 13306, Windhoek, Namibia, 1995. – DEA Research discussion paper 10. http://www.met.gov.na/publications/research/rdp_0010.pdf

Ashley u. Barnes 1996 Ashley, C. ; Barnes, J.: Wildlife Use for Economic Gain: the potential for wildlife to contribute development in Namibia / Department of Environmental Affairs (DEA), Ministry of Environment and Tourism (MET). Private Bag 13306, Windhoek, Namibia, 1996. – DEA Research discussion paper 12. http://www.met.gov.na/publications/research/rdp_0012.pdf

Augustin et al. 1998 Augustin, N.H. ; Mugglestone, M.A. ; Buckland, S.T.: The role of simulation in modelling spatially correlated data. In: *Environmetrics* 9 (1998), S. 175–196

Bandyopadyay et al. 2004 Bandyopadyay, S. ; Shyamsundar, P. ; Wang, L. ; Humavindu, M.N.: Do households gain from community-based natural resource management? An evaluation of community conservancies in Namibia / Department of Environmental Affairs (DEA), Ministry of Environment and Tourism (MET). Private Bag 13306, Windhoek, Namibia, 2004. – DEA Research discussion paper 68. http://www.met.gov.na/publications/research/RDP68.pdf

Barnes u. de Jager 1995 Barnes, J.I. ; Jager, J.L.V. de: Economic and financial incentives for wildlife use on private land in Namibia and the implications for policy / Department of Environmental Affairs (DEA), Ministry of Environment and Tourism (MET). Private Bag 13306, Windhoek, Namibia, 1995. – DEA Research discussion paper 8. http://www.met.gov.na/publications/research/rdp_0008.pdf

Barnes u. de Jager 1996 Barnes, J.I. ; Jager, J.L.V. de: Economic and financial incentives for wildlife use on private land in Namibia and the implications for policy. In: *South African Journal of Wildlife Research* 26 (1996), Nr. 2, S. 37–46

Barnes et al. 2001 Barnes, J.I. ; MacGregor, J. ; Weaver, C.: Economic analysis of community wildlife use initiatives in Namibia / Department of Environmental Affairs (DEA), Ministry of Environment and Tourism (MET). Private Bag 13306, Windhoek, Namibia, 2001. – DEA Research discussion paper 42. http://www.met.gov.na/publications/research/RDP42.pdf

Batty 2002 Batty, K. ; Animal Diversity Web (Hrsg.): *Alcelaphus buselaphus*. 2002. – http://animaldiversity.ummz.umich.edu/site/accounts/information/Alcelaphus_buselaphus.html. – Zugriffsdatum: 03 .03 .2006

Bill u. Fritsch 1991 Bill, R. ; Fritsch, D.: *Grundlagen der Geo-Informationssysteme*. Bd. 1 – Hardware, Software und Daten. Karlsruhe : Wichmann, 1991

Bishop et al. 1975 Bishop, Y.M.M. ; Fienberg, S.E. ; Holland, P.W.: *Discrete Multivariate Analysis*. Cambridge, MA : MIT Press, 1975

Borchers et al. 2002 Borchers, D.L. ; Zucchini, W. ; Buckland, S.T.: *Estimating Animal Abundance – Closed Populations*. London : Springer, 2002

Breslow u. Clayton 1993 Breslow, N.R. ; Clayton, D.G.: Approximate Inference in Generalized Linear Mixed Models. In: *Journal of The American Statistical Association* 88 (1993), S. 9–25

Brockhaus 1997 *Der Brockhaus: in 15 Bänden*. Leipzig-Mannheim : F. A. Brockhaus GmbH, 1997

Buckland et al. 1993 Buckland, S.T. ; Anderson, D.R. ; Burnham, K.P. ; Laake, J.L.: *Distance Sampling: Estimating Abundance of Biological Populations.* London : Chapman & Hall, 1993

Buckland et al. 2001 Buckland, S.T. ; Anderson, D.R. ; Burnham, K.P. ; Laake, J.L. ; Borchers, D.L. ; Thomas, L.: *Introduction to Distance Sampling – Estimating abundance of biological populations.* Oxford : Oxford University Press, 2001

Buckland et al. 2004 Buckland, S.T. ; Anderson, D.R. ; Burnham, K.P. ; Laake, J.L. ; Borchers, D.L. ; Thomas, L.: *Advanced Distance Sampling – Estimating abundance of biological populations.* Oxford : Oxford University Press, 2004

Burnham et al. 1987 Burnham, K.P. ; Anderson, D.R. ; White, G.C. ; Brownie, C. ; Pollock, K.H.: *American Fisheries Society monographs.* Bd. 5: *Design and analysis methods for fish survival experiments based on release-recapture.* Bethesda, MD : American Fisheries Society, 1987

Cameron u. Trivedi 1998 Cameron, A.C. ; Trivedi, P.K.: *Econometric Society monographs.* Bd. 30: *Regression analysis of count data.* Cambridge : Cambridge University Press, 1998

Charnes et al. 1976 Charnes, A. ; Frome, E.L. ; Yu, P.L.: The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family. In: *Journal of The American Statistical Association* 71 (1976), S. 169–171

Chilès u. Delfiner 1999 Chilès, J.-P. ; Delfiner, P.: *Geostatistics.* New York : John Wiley & Sons, 1999

Christensen et al. 2000 Christensen, O.F. ; Møller, J. ; Waagepetersen, R.: Analysis of spatial data using generalized linear mixed models and Langevin-type Markov chain Monte Carlo / Department of Mathematical Sciences, Aalborg University. Fredrik Bajers Vej 7 G, 9220 Aalborg East, Denmark, 2000. – Research Report R-00-2009. http://citeseer.ist.psu.edu/christensen00analysis.html

Christensen u. Waagepetersen 2002 Christensen, O.F. ; Waagepetersen, R.: Bayesian Prediction of Spatial Count Data Using Generalised Linear Mixed Models. In: *Biometrics* 58 (2002), S. 280–286

Collett 1991 Collett, D.: *Modelling Binary Data.* London : Chapman & Hall, 1991

CPC/FEWSa Climate Prediction Center / Famine Early Warning System (Hrsg.): *RFE 2.0 Dekadal African Rainfall Estimates.* – http://www.cpc.ncep.noaa.gov/products/fews/data.html. – Zugriffsdatum: 13 .07 .2005

Creel 2005 Creel, E. ; Animal Diversity Web (Hrsg.): *Phacochoerus africanus*. 2005. – http://animaldiversity.ummz.umich.edu/site/accounts/information/Phacochoerus_africanus.html. – Zugriffsdatum: 13 .04 .2006

Cressie 1985 Cressie, N.A.C.: Fitting Variogram Models by Weighted Least Squares. In: *Journal of the International Association for Mathematical Geology* 17 (1985), S. 563–586

Cressie 1991 Cressie, N.A.C.: *Statistics for spatial data*. New York : John Wiley & Sons, 1991

Cressie u. Hawkins 1980 Cressie, N.A.C. ; Hawkins, D.M.: Robust Estimation of the Variogram. In: *Journal of the International Association for Mathematical Geology* 12 (1980), S. 115–125

Crist u. Cicone 1984 Crist, E.P. ; Cicone, R.C.: A physically-based transformation of Thematic Mapper data – the TM Tasseled Cap. In: *IEEE Transactions On Geoscience and Remote Sensing* GE-22 (1984), S. 256–263

Dean 1992 Dean, C.B.: Testing for Overdispersion in Binomial and Poisson Regression Models. In: *Journal of The American Statistical Association* 84 (1992), S. 1085–1088

Dietrich u. Osborne 1991 Dietrich, C.R. ; Osborne, M.R.: Estimation of covariance parameters in kriging via restricted maximum likelihood. In: *Mathematical Geology* 23 (1991), Nr. 1, S. 119–135

Diggle et al. 1998 Diggle, P.J. ; Tawn, J.A. ; Moyeed, R.A.: Model-based Geostatistics. In: *Applied Statistics* 47 (1998), Nr. 3, S. 299–350

Dobson 2001 Dobson, A.J.: *An Introduction to Generalized Linear Models*. 2nd ed. Boca Raton : Chapman & Hall, 2001

Dorst u. Dandelot 1973 Dorst, J. ; Dandelot, P.: *Säugetiere Afrikas: Ein Taschenbuch für Zoologen und Naturfreunde*. Hamburg : Parey, 1973

Efron u. Tibshirani 1993 Efron, B. ; Tibshirani, R.J.: *Monographs on statistics and applied probability*. Bd. 57: *An introduction to the bootstrap*. New York : Chapman & Hall, 1993

ESRI 2004 Environmental Systems Research Institute: *Geoprocessing in ArcGIS*. 380 New York Street, Redlands, California : Environmental Systems Research Institute, 2004. – http://support.esri.com/index.cfm?fa=knowledgebase.documentation.viewDoc\&{}PID=43\&{}MetaID=1003

ESRI 2005 Environmental Systems Research Institute: *Writing Geoprocessing Scripts With ArcGIS*. 380 New York Street, Redlands, California : Environmental Systems Research Institute, 2005. – http://support.esri.com/index.cfm?fa=knowledgebase.documentation.viewDoc\&{}PID=43\&{}MetaID=1003

ESRI 2006 Environmental Systems Research Institute (Hrsg.): *GIS Dictionary*. 2006. – http://support.esri.com/index.cfm?fa=knowledgebase.gisDictionary.browse. – Zugriffsdatum: 27.11.2006

Estes 1997 Estes, R.D.: *The Behavior Guide to African Mammals*. 9th ed. South Africa : Russel Friedman Books, 1997

Fahrmeir u. Tutz 2001 Fahrmeir, L. ; Tutz, G.: *Multivariate Statistical Modeling Based on Generalized Linear Models*. 2nd ed. New York : Springer, 2001

Firth 1991 Firth, D.: Generalized Linear Models. In: Hinkley, D.V. (Hrsg.) ; Reid, N. (Hrsg.) ; Snell, E.J. (Hrsg.): *Statistical Theory and Modelling*. London : Chapman & Hall, 1991

Gates et al. 1985 Gates, C.E. ; Evans, W. ; Gober, D.R. ; Guthery, F.S. ; Grant, W.E.: Line transect estimation of animal densities from large data sets. In: Beasom, S.L. (Hrsg.) ; Roberson, S.F. (Hrsg.): *Game Harvest Management*. Kingsville, TX, USA : Ceasar Kleberg Wildlife Research Institute, Texas A&I University, 1985, S. 37–50

Gelman et al. 2003 Gelman, A. ; Carlin, J.B. ; Stern, H.S. ; Rubin, D.B.: *Bayesian Data Analysis*. 2nd ed. London : Chapman & Hall/CRC, 2003

Goldberger 1962 Goldberger, A.S.: Best Linear Unbiased Prediction in the Generalized Linear Regression Model. In: *Journal of The American Statistical Association* 57 (1962), S. 369–375

Goovaerts 1997 Goovaerts, P.: *Geostatistics for Natural Resources Evaluation (Applied Geostatistics)*. NY, USA : Oxford University Press, 1997

Gotway u. Stroup 1997 Gotway, C.A. ; Stroup, W.W.: A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. In: *Journal of Agricultural, Biological, and Environmental Statistics* 2 (1997), Nr. 2, S. 157–178

Gotway u. Wolfinger 2003 Gotway, C.A. ; Wolfinger, R.D.: Spatial Prediction of Counts and Rates. In: *Statistics in Medicine* 22 (2003), S. 1415–1532

Greenwood u. Yule 1920 Greenwood, M. ; Yule, G.U.: An Inquiry into the Nature of Frequency Distributions of Multiple Happenings, with Particular Reference to the Occurence of Multiple Attacks of Disease or Repeated Accidents. In: *Journal of The Royal Statistical Society Series A – Statistics In Society* 83 (1920), S. 255–279

Grzimek 1968 Grzimek, B. (Hrsg.): *Grzimeks Tierleben – Enzyklopädie des Tierreichs*. Zürich : Kindler, 1968

Grzimek 1988 Grzimek, B. (Hrsg.): *Grzimeks Enzyklopädie*. Bd. 5 – Säugetiere. Zürich : Kindler, 1988

Hardin u. Hilbe 2003 Hardin, J.W. ; Hilbe, J.M.: *Generalized Estimating Equations.* London : Chapman & Hall, 2003

Hastie u. Tibshirani 1990 Hastie, T.J. ; Tibshirani, R.J.: *Monographs on statistics and applied probability.* Bd. 43: *Generalized additive models.* London : Chapman & Hall, 1990

Heagerty u. Lele 1998 Heagerty, P.J. ; Lele, S.R.: A composite likelihood approach to binary spatial data. In: *Journal of The American Statistical Association* 93 (1998), S. 1099–1111

Hedley 2000 Hedley, S.L.: *Modelling Heterogeneity in Cetacean Surveys.* St Andrews, Scotland, University of St Andrews, PhD thesis, August 2000. – http://www.creem.st-and.ac.uk/sharon/public_html/thesis.pdf

Hedley u. Buckland 2004 Hedley, S.L. ; Buckland, S.T.: Spatial Models for Line Transect Sampling. In: *Journal of Agricultural, Biological, and Environmental Statistics* 9 (2004), Nr. 2, S. 181–199

Hiby 1986 Hiby, A.R.: Results of a hazard rate model relevant to experiments on the 1984/85 IDCR minke whale assessment cruise. In: *Report of the International Whaling Comission* 36 (1986), S. 497–498

Hiby u. Krishna 2001 Hiby, L. ; Krishna, M.B.: Line Transect Sampling from a Curving Path. In: *Biometrics* 57 (2001), S. 727–731

Hilbe 1994 Hilbe, J.: Log Negative Binomial Regression Using the GENMOD Procedure. In: *Proceedings of the Nineteenth Annual SAS User's Group International Conference* Bd. 14, 1994, S. 1199–1204

Hinde u. Démetrio 1998 Hinde, J. ; Démetrio, C.: Overdispersion: Models and Estimation. In: *Computational Statistics & Data Analysis* 27 (1998), S. 151–170

Horvitz u. Thompson 1952 Horvitz, D.G. ; Thompson, D.J.: A Generalization of Sampling without Replacement from a Finite Universe. In: *Journal of The American Statistical Association* 47 (1952), S. 663–685

Huang et al. 2002 Huang, C. ; Wylie, B. ; Yang, L. ; Homer, C. ; Zylstra, G.: Derivation of a tasselled cap transformation based on Landsat 7 at-satellite reflectance. In: *International Journal of Remote Sensing* 23 (2002), Nr. 8, S. 1741–1748

Huffman 2004a Huffman, B. ; The Ultimate Ungulate Page (Hrsg.): *Alcelaphus buselaphus.* 2004. – http://www.ultimateungulate.com/Artiodactyla/Alcelaphus_buselaphus.html. – Zugriffsdatum: 13 .04 .2006

Huffman 2004b Huffman, B. ; The Ultimate Ungulate Page (Hrsg.): *Tragelaphus strepsiceros.* 2004. – http://www.ultimateungulate.com/Artiodactyla/Tragelaphus_strepsiceros.html. – Zugriffsdatum: 13 .04 .2006

Huffman 2005 Huffman, B. ; The Ultimate Ungulate Page (Hrsg.): *Raphicerus campestris*. 2005. – http://www.ultimateungulate.com/Artiodactyla/Raphicerus_campestris.html. – Zugriffsdatum: 13 .04 .2006

Hurvich u. Tsai 1989 Hurvich, C.M. ; Tsai, C.L.: Regression and time series model selection in small samples. In: *Biometrika* 76 (1989), S. 297–307

Hurvich u. Tsai 1995 Hurvich, C.M. ; Tsai, C.L.: Model selection for extended quasi-likelihood models in small samples. In: *Biometrics* 51 (1995), S. 1077–1084

Ihaka u. Gentleman 1996 Ihaka, R. ; Gentleman, R.: R: A Language for Data Analysis and Graphics. In: *Journal of Computational and Graphical Statistics* 5 (1996), Nr. 3, S. 299–314. – http://www.r-project.org/

Jensen 1983 Jensen, J.R.: Biophysical remote sensing. In: *Annals of the Association of American Geographers* 73 (1983), Nr. 1, S. 111–132

Jensen 1996 Jensen, J.R.: *Introductory digital image processing: a remote sensing perspective*. 2nd ed. Upper Saddle River, NJ : Prentice Hall, 1996

Jones u. Murphree 2001 Jones, B. ; Murphree, M.: The evolution of policy and community conservation in Namibia and Zimbabwe. In: Hulme, D. and Murphree, M. (Hrsg.): *African Wildlife and Livelihoods*. Oxford, UK : James Currey Ltd., 2001

Journel u. Huijbregts 1978 Journel, A.G. ; Huijbregts, C.J.: *Mining Geostatistics*. London : Academic Press, 1978

Kauth u. Thomas 1976 Kauth, R.J. ; Thomas, G.S.: The tasseled cap – a graphic description of the spectral-temporal development of agricultural crops as seen in Landsat. In: *Proceedings on the Symposium on Machine Processing of Remotely Sensed Data*. West Lafayette, Indiana: LARS, Purdue University : Purdue University, 1976, S. 4B41–4B51

Kingdon 1997 Kingdon, J.: *The Kingdon field guide to African mammals*. San Diego : Academic Press, 1997

Kriegler et al. 1969 Kriegler, F.J. ; Malila, W.A. ; Nalepka, R.F. ; Richardson, W.: Preprocessing Transformations and Their Effects on Multispectral Recognition. In: *Proceedings of the Sixth International Symposium on Remote Sensing of Environment* Bd. 2. Ann Arbor, Michigan, USA : Infrared and Optics Laboratory, Willow Run Laboratories, Environmental Research Institute of Michigan, 1969, S. 97–131

Krige 1951 Krige, D.G.: A Statistical Approach to Some Basic Mine Valuation Problems on the Witwaterrand. In: *Journal of the Chemical, Metallurgical, and Mining Society of South Africa* 52 (1951), S. 119–139

Laake et al. 1996 Laake, J.L. ; Buckland, S.T. ; Anderson, D.R. ; Burnham, K.P.: *Distance User's Guide V2.2*. Colorado Cooperative Fish &

Wildlife Research Unit Colorado State University, Fort Collins, CO. 1996. – http://www.ruwpa.st-and.ac.uk/distance/distance22download.html

Lambert 1992 LAMBERT, D.: Zero-inflated Poisson regression, with an application to defects in manufacturing. In: *Technometrics* 34 (1992), Nr. 1, S. 1–14

Lawless 1987 LAWLESS, J.E.: Negative Binomial and Mixed Poisson Regression. In: *The Canadian Journal of Statistics – Revue Canadienne De Statistique* 15 (1987), S. 209–225

Lee 2004 LEE, P.M.: *Bayesian Statistics: An Introduction.* 3rd ed. London : Arnold, 2004

Lele 1997 LELE, S.: Estimating functions for semivariogram estimation. In: BASAWA, I.V. (Hrsg.) ; GODAMBE, V.P. (Hrsg.) ; TAYLOR, R.L. (Hrsg.): *Selected Proceedings of the Symposium in Estimating Functions.* Hayward, CA : Institute of Mathematical Statistics, 1997, S. 381–396

Liang u. McCullagh 1993 LIANG, K.Y. ; MCCULLAGH, P.: Case Studies in Binary Dispersion. In: *Biometrics* 49 (1993), S. 623–630

Liang u. Zeger 1986 LIANG, K.Y. ; ZEGER, S.L.: Longitudinal Data Analysis Using Generalized Linear Models. In: *Biometrika* 73 (1986), S. 13–22

Liang et al. 1993 LIANG, K.Y. ; ZEGER, S.L. ; QAQISH, B.: Multivariate regression analysis for categorical data (with discussion). In: *Journal of The Royal Statistical Society Series B – Statistical Methodology* 54 (1993), S. 3–40

Lillesand u. Kiefer 1994 LILLESAND, T.M. ; KIEFER, R.W.: *Remote sensing and image interpretation.* 3rd ed. New York : Wiley, 1994

Lindsay 1988 LINDSAY, B.G.: Composite likelihood methods. In: *Contemporary Mathematics* 80 (1988), S. 221–239

Lindsey 1997 LINDSEY, J.K.: *Applying Generalized Linear Models.* New York : Springer, 1997

Littel et al. 1996 LITTEL, R.C. ; MILLIKEN, G.A. ; STROUP, W.W. ; WOLFINGER, R.D.: *SAS System for Mixed Models.* Cary, NC : SAS Institute Inc., 1996

Littel et al. 2002 LITTEL, R.C. ; STROUP, W.W. ; FREUND, R.J.: *SAS for Linear Models.* 4th ed. Cary, NC : SAS Institute Inc., 2002

Mardia u. Marshall 1984 MARDIA, K.V. ; MARSHALL, R.J.: Maximum likelihood estimation of models for residual covariance in spatial regression. In: *Biometrika* 71 (1984), Nr. 1, S. 135–146

Matheron 1962 MATHERON, G.: *Memoires du Bureau de Recherches Geologiques et Minieres.* Bd. 14: *Traite de Geostatistique Appliquee, Tome I.* Paris : Editions Technip, 1962

Matheron 1963 MATHERON, G.: Principles of Geostatistics. In: *Economic Geology and The Bulletin of The Society of Economic Geologists* 58 (1963), S. 1246–1266

Matheron 1971 MATHERON, G.: The theory of regionalized variables and its applications / Ecole des Mines de Paris, Centre de Geosciences/Geostatistique. 35 rue Saint Honoré, 77305 Fonainebleau, France, 1971. – Les cahiers du CMM de Fontainebleau, Fasc. 5. http://cg.ensmp.fr/bibliotheque/1971/MATHERON/Ouvrage/DOC_00167/MATHERON_Ouvrage_00167.pdf

McCullagh 1983 MCCULLAGH, P.: Quasi-Likelihood Functions. In: *Annals of Statistics* 11 (1983), S. 59–67

McCullagh u. Nelder 1989 MCCULLAGH, P. ; NELDER, J.A.: *Monographs on statistics and applied probability.* Bd. 37: *Generalized linear models.* 2nd ed. London : Chapman & Hall, 1989

McCulloch u. Searle 2001 MCCULLOCH, C.E. ; SEARLE, S.R.: *Generalized, Linear, and Mixed Models.* New York : Wiley, 2001 (Wiley Series in Probability and Statistics)

McShane et al. 1997 MCSHANE, L.M. ; ALBERT, P.S. ; PALMATIER, M.: A Latent Process Regression Model for Spatially Correlated Count Data. In: *Biometrics* 53 (1997), Nr. 2, S. 698–706

Mendelsohn et al. 2002 MENDELSOHN, J. ; JARVIS, A. ; ROBERTS, C. ; ROBERTSON, T.: *Atlas of Namibia: A Portrait of the Land and its People.* Cape Town : David Philip Publishers, 2002

MET 1995 MINISTRY OF ENVIRONMENT & TOURISM (Hrsg.): *Namibia's environmental assessment policy for sustainable development and environmental conservation.* Ministry of Environment & Tourism, Private Bag 13306, Windhoek, Namibia. 1995

Milliken u. Johnson 1984 MILLIKEN, G.A. ; JOHNSON, D.E.: *Analysis of Messy Data.* Belmont, California : Lifetime Learning Publications, 1984

Nelder u. Wedderburn 1972 NELDER, J.A. ; WEDDERBURN, R.W.M.: Generalized Linear Models. In: *Journal of The Royal Statistical Society Series A – Statistics In Society* 135 (1972), S. 370–384

Newell 1999a NEWELL, T.L. ; ANIMAL DIVERSITY WEB (Hrsg.): *Raphicerus campestris.* 1999. – http://animaldiversity.ummz.umich.edu/site/accounts/information/Raphicerus_campestris.html. – Zugriffsdatum: 03 .03 .2006

Newell 1999b NEWELL, T.L. ; ANIMAL DIVERSITY WEB (Hrsg.): *Tragelaphus strepsiceros.* 1999. – http://animaldiversity.ummz.umich.edu/site/accounts/information/Tragelaphus_strepsiceros.html. – Zugriffsdatum: 03 .03 .2006

Norton-Griffith 1978 Norton-Griffith, M. ; Grimsdell, J.J.J. (Hrsg.): *Counting animals.* 2nd ed. P.O. Box 48177, Nairobi, Kenya : African Wildlife Leadership Foundation, 1978 (Handbooks on techniques in African wildlife ecology)

Olea 1999 Olea, R.A.: *Geostatistics for earth engineers and earth scientists.* Boston : Kluwer Academic Publishers, 1999

Pan 2001 Pan, W.: Akaike's information criterion in generalized estimating equations. In: *Biometrics* 57 (2001), S. 120–125

Pardo-Iguzquisa 1998 Pardo-Iguzquisa, E.: Maximum likelihood estimation of spatial covariance parameters. In: *Mathematical Geology* 30 (1998), Nr. 1, S. 95–108

Poisson 1837 Poisson, S.-D.: *Recherches sur la Probabilité des Jugements en Matière Criminelle et en Matière Civile.* Paris : Bachelier, 1837

Poortema 1999 Poortema, K.L.: On Modelling Overdispersion of Counts. In: *Statistica Nederlandica* 53 (1999), S. 5–20

Prentice 1988 Prentice, R.L.: Correlated Binary Regression with Covariates Specific to Each Binary Observation. In: *Biometrics* 44 (1988), S. 1033–1084

Rathbun u. Fei 2006 Rathbun, S.L. ; Fei, S.: A spatial zero-inflated poisson regression model for oak regeneration. In: *Environmental and Ecological Statistics* 13 (2006), Nr. 4, S. 409–426

Ripley 1981 Ripley, B.D.: *Spatial statistics.* New York : John Wiley & Sons, 1981

Robert u. Casella 1999 Robert, C.P. ; Casella, G.: *Monte Carlo Statistical Methods.* New York : Springer, 1999

Rouse et al. 1973 Rouse, J.W. ; Haas, R.H. ; Schell, J.A. ; Deering, D.W.: Monitoring vegetation systems in the great plains with ERST. In: *Proceedings of the 3rd ERTS Symposium*, 1973, S. 309–317. – NASA SP-351 (1)

Sanders 2005 Sanders, S. ; Animal Diversity Web (Hrsg.): *Oryx gazella.* 2005. – http://animaldiversity.ummz.umich.edu/site/accounts/information/Oryx_gazella.html. – Zugriffsdatum: 03 .03 .2006

SAS 2004 SAS Institute Inc.: *SAS/STAT 9.1 User's Guide.* Cary, NC : SAS Institute Inc., 2004

SAS 2006 SAS Institute Inc.: *The GLIMMIX Procedure.* 2006. – http://support.sas.com/rnd/app/da/glimmix.html. – Zugriffsdatum: 21 .02 .2007

Satterthwaite 1946 Satterthwaite, F.E.: An approximate distribution of estimates of variance components. In: *Biometric Bulletin* 2 (1946), S. 110–114

Schabenberger u. Gotway 2004 Schabenberger, O. ; Gotway, C.A.: *Statistical Methods for Spatial Data Analysis.* Boca Raton : Chapman & Hall/CRC, 2004

Thomas et al. 2003 Thomas, L. ; Laake, J.L. ; Strindberg, S. ; Marques, F.F.C. ; Buckland, S.T. ; Borchers, D.L. ; Anderson, D.R. ; Burnham, K.P. ; Hedley, S.L. ; Pollard, J.H. ; Bishop, J.R.B.: *Distance 4.1. Release 2.* Research Unit for Wildlife Population Assessment, University of St Andrews, Scotland. 2003. – http://www.ruwpa.st-and.ac.uk/distance/distance41download.html

Thompson 1992 Thompson, S.K.: *Sampling.* New York : John Wiley & Sons, 1992

Tobler 1970 Tobler, W.: A computer movie simulating urban growth in the Detroit region. In: *Economic Geography* 46 (1970), S. 234–240

USGS 2004 U.S. Geological Survey: *Landsat ETM+ Scene.* 2004. – WRS-2, path 177, row 75, orthorectified, acquisition date: 2004-02-12. Sioux Falls, South Dakota: USGS

Vjipurkar u. Gotway 2001 Vjipurkar, U.P. ; Gotway, C.A.: Assessment of forecasts and forecast uncertainty using generalized linear models for time series count data. In: *Journal of Statistical Computation and Simulation* 68 (2001), S. 321–349

Walker 1968 Walker, E.P. (Hrsg.): *Mammals of the world.* Bd. 2. 2nd ed. Baltimore : The Johns Hopkins Press, 1968

Walker 1983 Walker, E.P. (Hrsg.): *Walker's mammals of the world.* 4th ed. Baltimore : The Johns Hopkins University Press, 1983

Wedderburn 1974 Wedderburn, R.W.M.: Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method. In: *Biometrika* 61 (1974), S. 439–447

Weier u. Herring 2001 Weier, J. ; Herring, D. ; Earth Observatory, NASA (Hrsg.): *Measuring Vegetation (NDVI & EVI).* 2001. – http://earthobservatory.nasa.gov/Library/MeasuringVegetation/. – Zugriffsdatum: 30 .05 .2005

Wolfinger u. O'Connel 1993 Wolfinger, R.D. ; O'Connel, M.: Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. In: *Journal of Statistical Computation and Simulation* 48 (1993), S. 233–243

Wolfinger et al. 1994 Wolfinger, R.D. ; Tobias, R. ; Sall, J.: Computing Gaussian Likelihoods and Their Derivatives for General Linear Mixed Models. In: *SIAM Journal on Scientific Computing* 15 (1994), Nr. 6, S. 1294–1310

Yapp 1956 Yapp, W.B.: The theory of line transects. In: *Bird Study* 3 (1956), S. 93–104

Zeger u. Liang 1986 ZEGER, S.L. ; LIANG, K.Y.: Longitudinal Data Analysis for Discrete and Continous Outcomes. In: *Biometrics* 42 (1986), S. 121–130

Zhang 2002 ZHANG, H.: On Estimation and Prediction for Spatial Generalized Linear Mixed Models. In: *Biometrics* 58 (2002), S. 129–136

Zhao u. Prentice 1990 ZHAO, L.P. ; PRENTICE, R.L.: Correlated binary regression using a quadratic exponential model. In: *Biometrika* 77 (1990), S. 642–648

Zucchini et al. 2002 ZUCCHINI, W. ; BORCHERS, D.L. ; ERDELMEIER, M. ; BISHOP, J.: *WISP – A Wildlife Simulation Package*. Institut fur Statistik und Ökonometrie, Georg-August-Universität, Göttingen. 2002. – http://www.ruwpa.st-and.ac.uk/estimating.abundance/WiSP/index.html

Lebenslauf

Name:	Felix Mader
Geburtsdatum:	14. 05. 1976
Geburtsort:	Buxtehude
Staatsangehörigkeit:	deutsch
Familienstand:	ledig

Schulabschluß

05. 1995	Abitur, Gymnasium Osterholz-Scharmbeck

Wissenschaftlicher Werdegang

seit 01. 2004	Wissenschaftlicher Angestellter am Institut für Forstliche Biometrie und Informatik der Georg-August-Universität Göttingen
09. 2001-02. 2007	Promotionsarbeit zum Thema „Räumliche, GIS-gestützte Analyse von Linientransektstichproben“ am Institut für Forstliche Biometrie und Informatik der Georg-August-Universität Göttingen, gefördert durch ein Promotionsstipendium nach dem niedersächsischen Graduiertenförderungsgesetz (09. 2001-12. 2003) und ein Aufstockungsstipendium des DAAD (10. 2001-04. 2002 und 06.-09. 2002)
10. 1999-04. 2001	Fortsetzung des Studiums der „Forstwissenschaften und Waldökologie“ im Studienschwerpunkt „Waldökosystemanalyse und Informationsverarbeitung“ (Master-Studiengang) an der Fakultät für Forstwissenschaften und Waldökologie der Georg-August-Universität Göttingen Masterarbeit zum Thema „Wilddichteschätzungen mit Hilfe des „line-transect“-Verfahrens – Schätzung von Kudu-, Oryx- und Hartebeestdichten im Gebiet der Seeis Conservancy, Namibia“ Abschluß des Studiums mit dem Master of Science (M. Sc.) in Forstwissenschaften am 26. 03. 2001
10. 1996-06. 1999	Studium der „Forstwissenschaften und Waldökologie“ an der Fakultät für Forstwissenschaften und Waldökologie der Georg-August-Universität Göttingen (Bachelor-Studiengang)

	Abschluß mit dem Bachelor of Science (B. Sc.) in Forstwissenschaften am 15. 06. 1999

Studienbegleitendende Tätigkeiten

10. 1998-12. 2003	Wissenschaftliche Hilfskraft am Institut für Forstliche Biometrie und Informatik der Georg-August-Universität Göttingen

Praktika und Auslandsaufenthalte

09. 2001-09. 2002	Datenerhebung im Gebiet der Seeis Conservancy, Namibia
02. 2000-05. 2000	Praktikum bei Okatumba Wildlife Research, Omitara, Namibia
10. 1996-03. 1997	Praktikum im staatlichen Forstamt Osterholz-Scharmbeck