

Räumliche Analyse von Linientransektstichproben mit Hilfe von Distance, ArcGIS und SAS

Felix Mader Joachim Saborowski
Institut für Forstliche Biometrie und Informatik,
Universität Göttingen
Büsgenweg 4
37077 Göttingen
fmader@gwdg.de jsaboro@gwdg.de

Zusammenfassung

Dieser Vortrag stellt eine Möglichkeit vor, räumliche, in einem GIS verwaltete Informationen in die Analyse von Linientransekteherhebungen einzubeziehen und zur Präzisierung der Dichte- und Abundanzschätzungen, sowie für deren räumliche Vorhersagen zu nutzen. Methodisch basiert das Vorgehen im wesentlichen auf dem von Hedley [9, 10] vorgestellten Count-Modell. Dieses verwendet ein log-lineares Modell, das die Anzahl der in einem Transektabschnitt entdeckten Objekte in Abhängigkeit dort präsenter geographischer Informationen modelliert und auch kleinräumige Vorhersagen für nicht beobachtete Regionen des Untersuchungsgebietes ermöglicht. Um die zu erwartende räumliche Autokorrelation zwischen den Transektabschnitten in die Parametrisierung der Modelle und die Berechnung von Vorhersagen einzubeziehen, wurde das Count-Modell von Hedley um Ansätze von Gotway u. Stroup [7] erweitert, die allgemein der Berücksichtigung räumlicher Abhängigkeit in verallgemeinerten linearen Modellen dienen. Zur Vorhersage neuer Beobachtungen werden dabei Ansätze des aus der Geostatistik bekannten Universellen Krigings (z. B. [4]) aufgegriffen. Neben einem Einblick in die zur Anwendung kommenden Modelle wird insbesondere auf deren software-technische Umsetzung mit Hilfe der Programme Distance (Linientransektanalyse), ArcGIS (Geodatenverwaltung, -bearbeitung) und SAS (zentrale Steuerung, statistische Analyse), sowie auf deren Interaktion mit Hilfe von SAS-Makros und sog. Geoprocessing-Skripten eingegangen.

Ein Anwendungsbeispiel befaßt sich mit den Wildzählungsdaten einer namibischen Antilopenart (*Oryx gazella*). Die in diesem Vortrag behandelte Thematik ist Teil eines aktuellen Promotionsvorhabens [12].

Schlüsselwörter: Log-lineares Modell, Verallgemeinerte Lineare Modelle, GLM, Räumliche Autokorrelation, PROC GLIMMIX, PROC GENMOD, Python, Geoprocessing-Skripte, Count-Modell, Distance Sampling, Geographische Informationssysteme, GIS, Wilddichteschätzung.

1 Einleitung

Das zur Familie des Distance Sampling [2] gehörende Linientransektverfahren findet weltweit Anwendung zur Dichte- und Abundanzschätzung unterschiedlichster biologischer Populationen. Gegenüber klassischen Stichprobenverfahren zeichnet es sich insbesondere dadurch aus, daß es nicht der Annahme bedarf, alle in der Stichprobenfläche

vorhandenen Untersuchungsobjekte zu entdecken. Vielmehr wird anhand der zur den entdeckten Objekten gemessenen Entfernungen ein Korrekturfaktor für nicht entdeckte Objekte geschätzt.

Im Zuge steigender Verfügbarkeit kostengünstiger und qualitativ hochwertiger geographischer Informationen besteht zunehmend der Bedarf, diese auch in die Analyse von Linientransektaufnahmen einzubeziehen, mit dem Ziel erstens Schätzungen zu präzisieren und zweitens kleinräumige Vorhersagen für beliebige Bereiche des Untersuchungsgebietes zu ermöglichen, auch für solche, die im Zuge der Aufnahmen nicht erfaßt wurden. Auf diese Weise generierte Verteilungskarten der untersuchten Art können wertvolles Hilfsmittel für das Management und die Untersuchung ökologischer Zusammenhänge sein. Die klassische Analyse des Linientransektverfahrens sieht jedoch den Einbezug zusätzlicher erklärender (räumlicher oder nicht-räumlicher) Variablen nicht vor und kleinräumige Dichteschätzungen lassen sich nur durch eine (geographische) Stratifizierung erreichen. Mit kleiner werdenden Straten gehen jedoch auch geringere Beobachtungsanzahlen (je Stratum) und damit letztlich weniger präzise Dichteschätzungen einher. Im wesentlichen basierend auf den Arbeiten von Hedley [9, 10] und Gotway u. Stroup [7] soll in Abschnitt 2 daher eine Möglichkeit vorgestellt werden, mit Hilfe derer herkömmliche Linientransekterhebungen mit zusätzlichen geographischen Informationen zusammengeführt und für die Steigerung der Präzision von Schätzungen und die lokale Vorhersage von Objektdichten und -anzahlen verwendet werden können. Die Darstellung der Methoden beschränkt sich dabei jedoch auf die wichtigsten grundlegenden Konzepte, für ausführlichere Betrachtungen sei auf die genannte Literatur verwiesen.

Der Schwerpunkt der Betrachtungen liegt auf der software-technischen Umsetzung der vorzustellenden Methoden, die mit Hilfe der Programme Distance, ArcGIS und SAS realisiert wurde. Sie wird in Abschnitt 3 beschrieben.

Abschnitt 4 enthält ein kurzes Anwendungsbeispiel, das sich mit Linientransekterhebungen die zur Dichteschätzung von Oryxantilopen in einem Untersuchungsgebiet im zentralen Osten Namibias durchgeführt wurden, befaßt.

2 Modelle und Methoden

2.1 Das Linientransektverfahren

Allen Verfahren des Distance Sampling, zu denen das Linientransektverfahren gehört, ist die Eigenschaft eigen, daß die wesentliche, zu erfassende Größe die Entfernung zu entdeckten Untersuchungsobjekten ist, um anhand dieser den Anteil übersehener Objekte schätzen zu können. Im Falle des Linientransektverfahrens erfolgt die Datenerhebung von k zufällig oder systematisch in das Untersuchungsgebiet gelegten Geraden (Transekten) der Gesamtlänge $L = \sum_{i=1}^k l_i$. Die Entfernungen zu entdeckten Objekten werden als senkrechte Distanz x zwischen Objekt und Transekt bestimmt. Anhand der erfaßten Distanzen läßt sich eine sog. Entdeckungs(wahrscheinlichkeits)funktion $g(x)$ parametrisieren, die die Wahrscheinlichkeit angibt, ein Objekt in Entfernung x vom Transekt zu entdecken. Unter einigen Annahmen, von denen die bedeutendste ist, alle

Objekte direkt auf und sehr nahe dem Transekt mit Sicherheit zu entdecken ($g(0)=1$), ist ein erwartungstreuer Schätzer der Objektdichte D_s gegeben durch

$$\hat{D}_s = \frac{n}{2 \cdot L \cdot \hat{\nu}}, \quad (2.1)$$

wobei n die Anzahl der Beobachtungen und $\hat{\nu}$ die sog. effektive Streifenbreite (genauer: halbe effektive Streifenbreite) ist, die sich direkt aus der angepaßten Entdeckungsfunktion $\hat{g}(x)$ ergibt. Die effektive Streifenbreite wird interpretiert als die senkrechte Entfernung zum Transekt, bis zu der so viele Objekte übersehen, wie jenseits davon noch entdeckt wurden. Eine Linientransekthebung mit (halber) effektiver Streifenbreite $\hat{\nu}$ läßt sich daher auch als Streifenzählung mit $2 \cdot \hat{\nu}$ breiten Streifen betrachten. Treten die Untersuchungsobjekte in Gruppen (Herden, Rudel usw.) auf, so werden die Entfernungen nicht zu den einzelnen Individuen, sondern jeweils zum Schwerpunkt einer Gruppe gemessen. Die Objektdichte D_s entspricht dann der Gruppendichte, ein Schätzer der Individuendichte D ergibt sich durch

$$\hat{D} = \hat{D}_s \cdot \hat{E}(s), \quad (2.2)$$

wobei $\hat{E}(s)$ ein Schätzer der mittleren Gruppengröße - oftmals das arithmetische Mittel $\hat{E}(s) = \frac{1}{n} \sum_{i=1}^n s_i$ der Gruppengrößen s_i - ist.

Umfangreiche Details zur Methodik und Umsetzung der Verfahren des Distance Sampling finden sich in [2, 3].

2.2 Das Count-Modell

Wie leicht aus (2.1) ersichtlich, sieht der herkömmliche Dichteschätzer des Linientransektverfahrens den Einbezug zusätzlicher erklärender Kovariaten nicht vor. Des weiteren ist die Berechnung kleinräumiger Vorhersagen der lokalen Dichten nicht möglich; Dichteschätzungen beziehen sich stets auf das gesamte Untersuchungsgebiet, oder, bei einer stratifizierten Auswertung, auf das gesamte Stratum. Die Größe der Straten ist infolge geringerer Beobachtungsanzahlen (je Stratum) nach unten jedoch durch die erforderliche Präzision der Schätzer begrenzt, zudem ist eine verlässliche Schätzung der Entdeckungsfunktion oftmals erst ab einer Beobachtungsanzahl von 60-80 möglich. Beide Ziele - die Berechnung lokaler Vorhersagen und der Einbezug zusätzlicher erklärender Variablen - sind jedoch zunehmend in ökologischen 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 nützliches Hilfsmittel für das Management der untersuchten Population sein.

Eine Möglichkeit, den neuen Anforderungen gerecht zu werden, ist das von Hedley [9, 10] vorgestellte Count-Modell. Es sieht die Einbindung zusätzlicher, z. B. in einem geographischen Informationssystem (GIS) gespeicherter Informationen in die konventionelle Linientransektanalyse vor und ermöglicht sowohl die Ausweisung signifikanter Einflußgrößen, als auch räumliche Dichte- und Abundanzvorhersagen. Die Datenerhebung erfolgt wie zuvor beschrieben, zusätzlich zu den Entfernungen x , der Länge der Transekte L und ggf. den Gruppengrößen s sind die Positionen entdeckter Objekte und der Transekte nun aber z. B. per GPS (Global Positioning System) zu verorten, um entsprechende Geodatensätze in einem GIS erstellen zu können. Im Zuge der Analyse wird dann, z. B. mit Hilfe von GIS-Funktionen, jedes der k erfaßten Transekte in T_i Segmente (etwa) gleicher Länge c unterteilt ($l_i = \sum_{j=1}^{T_i} c_j$ und $L = \sum_{i=1}^k l_i$); 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 im Anschluß um jeden Transektabschnitt ein Streifen der Breite $2 \cdot \hat{v}$ erstellt und für jedes der so generierten Segmente die Anzahl der darin beobachteten Objekte n_i , sowie zusätzlich die an der Position des Segments gültigen Werte erklärender (räumlicher) Kovariaten bestimmt (s. Abschnitt 3.2). Ein log-lineares Modell zur Beschreibung der Objektanzahl eines Segments läßt sich dann durch

$$E(n_i) = \mu_i = \exp \left[\ln(2 \cdot c_i \cdot v) + \beta_o + \sum_{j=1}^{p-1} x_{ij} \beta_j \right] \quad (2.3)$$

formulieren, worin μ_i der Erwartungswert der Anzahl der im i -ten Segment beobachteten Objekte ist. Der natürliche Logarithmus der Fläche eines Segments $2 \cdot c_i \cdot v$ ist als sog. Offset-Term enthalten, um unterschiedliche Segmentgrößen zu berücksichtigen. x_{ij} sind die für das Segment bestimmten Werte der (räumlichen) Kovariaten, β_j , $j=0,1,\dots,p-1$ die zugehörigen, zu schätzenden Parameter. Die Maximum-Likelihood- oder Quasi-Likelihood-Schätzer [18, 13] des $p \times 1$ Parametervektors β ergeben sich durch iterative Lösung des Gleichungssystems

$$\mathbf{X}' \mathbf{W} \mathbf{X} \beta = \mathbf{X}' \mathbf{W} \mathbf{y}^* \quad (2.4)$$

Diese Darstellung ist für alle verallgemeinerten linearen Modelle (GLMs; [13]), zu denen auch das log-lineare Modell gehört, gültig, wobei \mathbf{X} die aus den Kovariaten zusammengesetzte $n \times p$ Designmatrix, $\mathbf{y}^* = \boldsymbol{\eta} + \Delta (\mathbf{y} - \boldsymbol{\mu})$, $\mathbf{W} = \Delta^{-1} \mathbf{V}^{-1} \Delta^{-1}$ Δ eine $n \times n$ Diagonalmatrix mit Elementen $\text{diag}[\partial \mu_i / \partial \eta_i]$ und \mathbf{V} die Varianzmatrix mit Elementen $\text{diag}[\phi \cdot v(\mu_i)]$ ist. Der lineare Prädiktor η und die Varianzfunktion $v(\mu)$ hängen von der Wahl der Link-Funktion des GLMs und der für y unterstellten Verteilung ab, ϕ ist ein ggf. zu schätzender Dispersionsparameter. Für das log-lineare Modell ist $\eta = \ln(\mu) = \mathbf{x}' \beta$ und $v(\mu) = \mu$, anstelle des Dispersionsparameters ϕ kommt hier oftmals ein Überdispersionsparameter σ_o^2 zur Anwendung, der die Berücksichtigung größerer als

unter der Poisson-Verteilung erwarteter Varianz ermöglicht; es ist dann $\mathbf{V} = \text{diag}[\sigma_o^2 \cdot \nu(\mu_i)]$. Die Parameterschätzungen $\hat{\boldsymbol{\beta}}$ sind konsistent, und approximativ normalverteilt mit Kovarianzmatrix $\text{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$.

Vorhersagen des Erwartungswertes $\hat{\mu}_0$ werden ermöglicht, indem in (2.3) die Parameterschätzungen $\hat{\boldsymbol{\beta}}$, sowie Offset-Term und Kovariaten der vorherzusagenden Beobachtungen eingesetzt werden. In der Anwendung des Count-Modells wird hierfür ein Gitter mit Zellen der Größe f_0 über das Untersuchungsgebiet gelegt, für dessen Zellen die räumlichen Kovariaten und damit \mathbf{x}_0 bestimmt werden. $\ln(f_0)$ wird anstelle der Segmentfläche als Offset-Term verwendet. Die Varianz einer solchen Vorhersage ist für das log-lineare Modell gegeben durch $\text{var}(\hat{\mu}_0) \approx [\exp(\hat{\eta}_0)]^2 \cdot \text{var}(\hat{\eta}_0)$, wobei $\text{var}(\hat{\eta}_0) = \mathbf{x}'_0 (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} \mathbf{x}'_0$.

2.2.1 Einbezug räumlicher Autokorrelation

Nachteil der Verwendung eines gewöhnlichen log-linearen Modells im Count-Modell ist, daß es die zu erwartende räumliche Abhängigkeit zwischen den Segmenten nicht berücksichtigt. So werden sich die Beobachtungsanzahlen benachbarter Segmente in der Regel ähnlicher sein als die weiter entfernter Segmente. Ein früher Ansatz des Einbezugs räumlicher Korrelationen in GLMs stammt, basierend auf den Generalized Estimating Equations [11, 20], von Albert und McShane [1]. Gotway und Stroup [7] haben diesen Ansatz aufgegriffen, in einen allgemeineren Kontext gestellt und zudem eine Möglichkeit für räumliche Vorhersagen formuliert, die sich an das aus der Geostatistik bekannte Universelle Kriging [z. B. 4] anlehnt. Die Modellformulierung erfolgt für das log-lineare Modell wie für den Fall unkorrelierter Beobachtungen durch (2.3), die Parameterschätzung über Quasi-Likelihood-Methoden oder äquivalent durch iterative Lösung von (2.4). Anstelle der diagonalen Varianzmatrix $\mathbf{V} = \text{diag}[\sigma_o^2 \cdot \nu(\mu_i)]$ ($\sigma_o^2 = 1$, wenn keine Über- oder Unterdispersion vorliegt) wird hier jedoch eine Kovarianzmatrix $\boldsymbol{\Sigma}_{yy}(\boldsymbol{\mu}, \boldsymbol{\alpha})$ verwendet, die wie folgt formulierbar sein muß [8].

$$\boldsymbol{\Sigma}_{yy}(\boldsymbol{\mu}, \boldsymbol{\alpha}) = \boldsymbol{\sigma}_n^2 \mathbf{V} + \mathbf{V}^{1/2} \mathbf{R}(\boldsymbol{\alpha}) \mathbf{V}^{1/2} \tag{2.5}$$

Hierin ist $\boldsymbol{\sigma}_n^2$ ein optionaler Nugget-Effekt, \mathbf{V} die bereits zuvor verwendete Varianzmatrix und $\mathbf{R}(\boldsymbol{\alpha})$ eine $n \times n$ Korrelationsmatrix. Die Elemente von $\mathbf{R}(\boldsymbol{\alpha})$ werden in der Praxis anhand der entfernungs- und richtungsabhängigen Korrelationsfunktion des räumlichen Prozesses bestimmt, deren unbekannt Parameter $\mathbf{R}(\boldsymbol{\alpha})$ (Schwellenwert und effektive Reichweite) die räumliche Abhängigkeit vollständig spezifizieren und z. B. anhand des empirischen Semivariogramms der Pearson-Residuen

$$r^P = \frac{y - \hat{\mu}}{\sqrt{\hat{\mu}}} \tag{2.6}$$

bestimmt werden können.

Zur Vorhersage neuer Beobachtungen leiten Gotway und Stroup [7] aus dem klassischen Universellen-Kriging-Prädiktor die Formulierung

$$\hat{y}_0 = \hat{\mu}_0 + \Sigma_{0y} \Sigma_{yy}^{-1} (y - \hat{\mu}) \quad (2.7)$$

her, wobei $\hat{\mu}_0$ der $k \times 1$ Vektor mit Vorhersagen gemäß (2.3), Σ_{0y} die $k \times n$ Kovarianzmatrix zwischen vorherzusagenden und beobachteten Positionen und Σ_{yy} die $n \times n$ Kovarianzmatrix zwischen beobachteten Positionen ist. Die Vorhersage für eine neue Beobachtung \hat{y}_0 setzt sich hier also zusammen aus der Schätzung des Erwartungswertes $\hat{\mu}_0$ an der Position und einer „Korrektur“ infolge der Residuen für beobachtete Positionen $y - \hat{\mu}$. Infolge der räumlichen Kovarianzstruktur bekommen dabei nahe gelegene Residuen stärkeres Gewicht als weiter entfernte. Die Vorhersage für eine neue Beobachtung wird also gewissermaßen anhand der benachbarten Beobachtungen angepaßt. Den mittleren quadratischen Vorhersagefehler (MSPE) des Prädiktors aus (2.7) geben Gotway und Wolfinger [8] mit

$$\begin{aligned} \text{MSPE}(\hat{y}_0) = & \Sigma_{00} - \Sigma_{0y} \Sigma_{yy}^{-1} \Sigma_{y0} + (\mathbf{D}_0 - \Sigma_{0y} \Sigma_{yy}^{-1} \mathbf{D}) \\ & \times (\mathbf{X}' \mathbf{W} \mathbf{X})^{-1} (\mathbf{D}_0 - \Sigma_{0y} \Sigma_{yy}^{-1} \mathbf{D}) \end{aligned} \quad (2.8)$$

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

3 Software-technische Umsetzung

Gemäß den vorhergehenden Erläuterungen läßt sich der Analyseablauf nach einer erfolgten Aufnahme wie folgt unterteilen:

- Durchführung einer klassischen Linientransektanalyse mit dem Ergebnis (u. a.) einer Dichteschätzung \hat{D}_{LT} und der effektiven Streifenbreite \hat{v} .
- Erstellen der Segmente und Bestimmen der Anzahl der Beobachtungen je Segment n_i und der zugehörigen räumlichen Kovariaten \mathbf{x}_i .
- Modellformulierung und Schätzung des Parametervektors β nach (2.4), mit $\mathbf{W} = \Delta^{-1} \mathbf{V}^{-1} \Delta^{-1}$ oder $\mathbf{W} = \Delta^{-1} \Sigma_{yy}^{-1} \Delta^{-1}$ und ggf. der räumlichen Kovarianzparameter α .
- Erstellen des Vorhersagegitters, Bestimmen der räumlichen Kovariaten, Berechnung der Vorhersagen nach (2.3) oder (2.7).

Viele der in diesen Punkten enthaltenen Arbeitsschritte sind bereits in verfügbarer Software implementiert. So bietet die frei erhältliche Windows-Software Distance [17] zahlreiche Möglichkeiten der Planung und Analyse von Linientransekterhebungen und anderer zur Familie des Distance Sampling gehörenden Verfahren.

Für die Bearbeitung und Analyse von Geodaten stehen zahlreiche kommerzielle und frei verfügbare Software-Produkte zu Verfügung; hier kam ESRI's ArcGIS Desktop 9.1 in der ArcInfo-Lizenzierung zum Einsatz.

U. a. mit der Prozedur GENMOD bietet SAS die Möglichkeit der Analyse log-linearer und weiterer verallgemeinerter linearer Modelle, die Berücksichtigung korrelierter Be-

obachtungen beschränkt sich jedoch auf Meßwiederholungen, die Angabe räumlicher Kovarianzstrukturen ist nicht möglich. Diese bietet jedoch die seit einiger Zeit verfügbare und auf dem GLIMMIX-Makro basierende GLIMMIX-Prozedur, die in der Version 9.1.3 noch nicht zum Standardumfang des SAS/STAT-Moduls gehört, aber auf der SAS-Support-Website¹ verfügbar ist.

Zentrales Ziel war die weitestgehende Automatisierung der Arbeitsabläufe und die zentrale Steuerung aus einem der beteiligten Programme heraus zu ermöglichen. Die Automatisierung wurde mit Hilfe von SAS-Makros und sog. Geoprocessing-Skripten (s. u.) erreicht, als zentrale Steuerungseinheit bot sich aufgrund umfangreicher vorgefertigter Möglichkeiten der Datenmanipulation, sowie des Datenim- und -exports SAS an. Insgesamt wurden 27 Makros und 11 Skripte entwickelt, von denen nachfolgend einige erläutert werden.² Ausführlichere Beschreibungen finden sich in [12].

3.1 Klassische Linientransektanalyse

Da SAS selbst keinerlei Funktionen zur Analyse von Linientransektmessungen anbietet, wurde der Batch-Modus von Distance genutzt, um (derzeit begrenzt auf konventionelle Linientransektanalysen) auf dessen Funktionsumfang zugreifen zu können. Die Software wird in der Regel über eine graphische Oberfläche bedient, stellt aber auch die Möglichkeit des Batch-Betriebs zur Verfügung. Die Steuerung erfolgt in diesem Fall über eine Daten- und eine Konfigurationsdatei (im ASCII-Format), Ergebnisse werden ebenfalls in Form von ASCII-Dateien ausgegeben. Die entwickelten Makros exportieren in SAS-Tabellen vorliegende Linientransektdaten, schreiben die Daten- und Konfigurationsdatei, importieren die Ergebnisdateien und legen die Ergebnisse in Form von SAS-Tabellen ab oder stellen sie graphisch dar. Zentrales, als einziges vom Nutzer aufzurufendes Makro ist %DistAnaly. Mit Hilfe von über 50, größtenteils optionalen Makroparametern kann Einfluß auf die zu exportierenden Daten, die durchzuführende Analyse und die zu erstellenden Ergebnisdateien genommen werden. Optional können ausgewählte Ergebnisse, wie z. B. die geschätzte effektive Streifenbreite $\hat{\nu}$, zur weiteren Verwendung direkt in Makrovariablen gespeichert werden.

3.2 Segmenterstellung und Bestimmung der Kovariaten

Das Erstellen der Segmentflächen um die entstehenden Abschnitte, die Ermittlung der Anzahl der in ein Segment fallenden Beobachtungen und die räumliche Verschneidung von Segmenten mit weiteren geographischen Informationen zur Bestimmung der Kovariaten erfordern Funktionalität die in das Spektrum geographischer Informationssysteme fällt. ArcGIS Desktop bot sich hier an, da es (in der ArcInfo-Lizenzierung) bereits die meisten der erforderlichen Teilschritte in Form von bereits implementierten Befehlen anbot. Darüber hinaus bietet ArcGIS seit der Version 9 die Möglichkeit der Entwicklung sog. Geoprocessing-Skripte [5, 6], die, ähnlich SAS-Makros, die Entwicklung zu-

¹ <http://support.sas.com/ctx/samples/index.jsp?sid=536>

² Alle Makros und Skripte sind in Kürze unter <http://www.uni-forst.gwdg.de/~fmader/smalts/> verfügbar.

sätzlicher Funktionalität und die Automatisierung umfangreicher Arbeitsabläufe ermöglichen. Bei den Geoprocessing-Skripten handelt es sich um in einer Skriptsprache (z. B. JScript, VBScript oder Python) geschriebene Programme, für die ArcGIS eine Schnittstelle zum Zugriff auf sämtliche GIS-Funktionalitäten bereitstellt. Sie greifen damit das Konzept der aus ArcInfo workstation bekannten AMLs (ArcMacro Language) auf, im Gegensatz zur ArcInfo-eigenen Makrosprache AML steht hier nun aber über die GIS-Funktionalität hinaus der gesamte Umfang der von der jeweiligen Skriptsprache angebotenen Möglichkeiten zur Verfügung. Die von ESRI favorisierte Skriptsprache ist Python³; Python 2.1 und PythonWin (u. a. Editor, Debugger) sind in der ArcGIS-Desktop-Distribution enthalten.

Ein einfaches Beispiel eines Python Geoscripts BufferBsp.py ist

```
# Import der Systemmodule
import sys, string, win32com.client

# Definition der Übergabeparameter
LineFcP = sys.argv[1] # 1. Übergabeparameter: Linien-Feature-Class
                    # (z. B. C:\temp\linien.shp)
BufferWidth = sys.argv[2] # Buffer-Breite (z. B. 50)
BufferFcP= sys.argv[3] # Ausgabe-Feature-Class (z. B. C:\temp\buffer.shp)

try:
    # Erstellen des Geoprocessor Objekts
    try:
        gp = win32com.client.Dispatch("esriGeoprocessing.GpDispatch.1")
    except:
        raise Exception, "Geoprocessor-Objekt-Fehler "

    # Erstellen von Buffer-Geometrien und Speicherung in einer
    # neuen Feature Class
    try:
        gp.toolbox = "analysis" # Zu verwendende ArcInfo-Toolbox
        gp.Buffer_analysis(LineFcP, BufferFcP, str(BufferWidth)
                           + " Meters", "FULL", "FLAT", "NONE", "#")
        # Buffer_analysis(InputFeatureClassPfad, OutputFeatureClassPfad,
        #                 Bufferbreite, weitere Optionen)
    except:
        raise Exception, "Buffer-Fehler"

# Fehlerbehandlung
except Exception, ErrorDesc:
    print ErrorDesc
```

³ <http://python.org>

Mit dem Erstellen des Geoprocessor-Objekts steht die gesamte Bandbreite der von ArcGIS zur Verfügung gestellten Befehle zur Bearbeitung und Analyse von Geodaten zur Verfügung. In dem Beispiel werden um die in einer sog. Feature Class⁴ gespeicherten Liniengeometrien Buffer(Streifen)geometrien erstellt und diese in einer weiteren Feature Class gespeichert. Zur Steuerung des Ablaufs eines Skriptes lassen sich u. a. Übergabeparameter definieren, deren Werte beim Aufruf des Skriptes mit übergeben werden. Im Beispiel sind dies die Eingabe- und Ausgabe-Feature-Class, sowie die Breite der Streifen. Geoprocessing-Skripte lassen sich entweder als neue Werkzeuge (Tools) in den ArcGIS-Desktop-Programmen (z. B. ArcMap, ArcCatalog) verwenden oder über die Eingabeaufforderung („DOS-Box“) starten. Ein möglicher Aufruf des Beispielskriptes oben wäre:

```
BufferBsp.py "C:\temp\linien.shp" "50" "C:\temp\buffer.shp"
```

Über die Funktion CALL SYSTEM() bzw. das X Statement ist es somit auch möglich, die von ArcGIS bereitgestellte Funktionalität aus SAS heraus zu nutzen ohne manuell zwischen den Programmen wechseln und Daten „per Hand“ im- bzw. exportieren zu müssen. Mit Hilfe entsprechend formulierter Makros lassen sich dabei auch die Werte der Übergabeparameter flexibel halten.

```
%MACRO buffer(
    LineFcP=,
    BufferW=,
    BufferFcP=
);
DATA _null_;
command = quote("BufferBsp.py") || " " ||
    quote("&LineFcP") || " " ||
    quote("&BufferW") || " " ||
    quote("&BufferFcP");
CALL SYSTEM(command);
RUN;

%MEND buffer;
```

⁴ Eine Feature Class bezeichnet in ArcGIS eine Einheit zur Speicherung von Objekten desselben Typs (z.B. Punkte, Linien oder Flächen). Bekanntestes Feature-Class-Format ist das der Shapefiles.

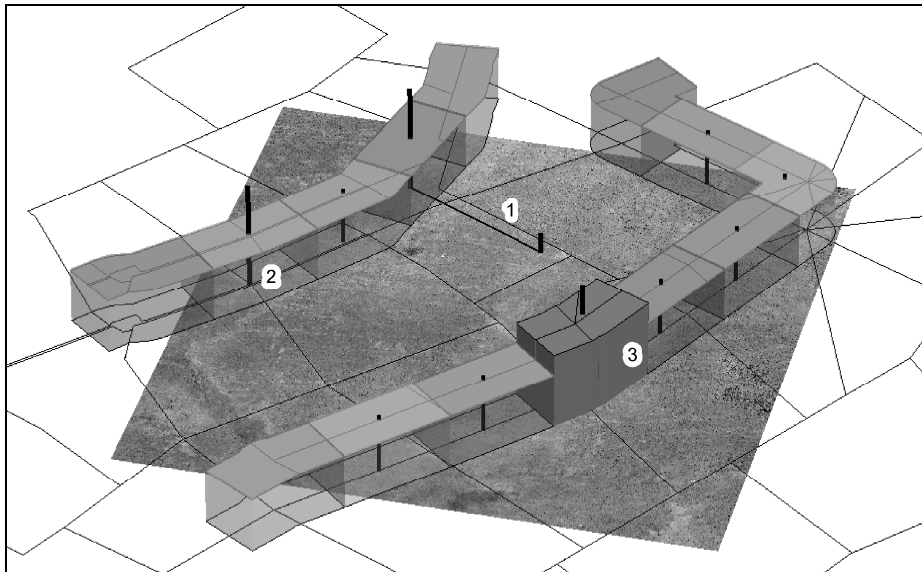


Abbildung 1: Ermittlung der Werte der räumlichen Kovariaten je Segment: 1) Entfernungsmessung, 2) Punktverschneidung, 3) Flächenverschneidung.

Das Zusammenspiel aus SAS-Makros und Geoprocessing-Skripten wurde genutzt, um die erforderlichen Bearbeitungsschritte zu realisieren (%CreateSegments). Die Makros und Skripte ermöglichen dabei u. a. die Angabe der zu verwendenden Segmentbreite (= effektive Streifenbreite \hat{v}), der Segmentlänge c , und (fast) beliebig vieler räumlich mit den generierten Segmenten in Bezug zu setzenden geographischen Informationen. Drei Arten des räumlichen In-Bezug-Setzens, die in Abbildung 1 verdeutlicht sind, werden derzeit unterstützt: Entfernungsmessungen, Punkt- und Flächenverschneidungen. Im ersten Fall wird die kürzeste Entfernung des Segmentsschwerpunktes zum nächstgelegenen Objekt einer anderen Feature Class berechnet, z. B. zur nächsten Tränke oder zur nächsten öffentlichen Straße. Für Punktverschneidungen wird der Zellwert eines Rasterdatensatzes (z. B. ESRI Grid, GeoTiff) an der Stelle des Segmentsschwerpunktes abgefragt (z. B. Geländehöhe, Niederschlagsmenge), für Flächenverschneidungen werden die Segmentgeometrien (Polygone) mit den Flächen einer weiteren Feature Class geometrisch verschnitten. Anhand der für jedes Segment resultierenden Teilflächen wird entweder ein (über diese Teilflächen) flächengewichteter Mittelwert bestimmt oder der Wert der größten Teilfläche verwendet (z. B. Weidequalität). Die so bestimmten Kovariatenwerte werden gemeinsam mit der ebenfalls bestimmten Anzahl der Beobachtungen je Segment und der Segmentfläche in einer SAS-Tabelle abgelegt, die dann direkt für die weiteren Analysen verwendet werden kann. Die Makros und Skripte ermöglichen hierbei, wie in Abbildung 1 dargestellt, nicht nur die Verwendung gerade verlaufender sondern auch gekrümmter Transekte.

3.3 Parameterschätzung

Für die Schätzung der unbekannt Parameter der Kovariaten bieten sich in SAS die Prozeduren GENMOD [14] (ohne räumliche Autokorrelation) und GLIMMIX [15, 16] (mit räumlicher Autokorrelation) an. PROC GENMOD dient allgemein der Lösung verallgemeinerter linearer Modelle (GLM; [13]), zu denen auch das log-lineare Modell gehört, und bestimmt die Parameter über die ML- bzw. QL-Methode. Über

```
PROC GENMOD data = segments;
CLASS VegType;
MODEL n = DistWater VegType /
      offset = LnArea dist = poisson
      link = log scale = pearson
;
RUN;
```

läßt sich ein log-lineares Modell (`link = log`) mit der Poisson- als Wahrscheinlichkeitsverteilung (`dist = poisson`) und Offset-Variable `LnArea` berechnen. Die Werte der Offset-Variablen entsprechen im Fall der Transektsegmente dem natürlichen Logarithmus der Segmentfläche ($= \ln(2 \cdot c_i \cdot \hat{v})$). Ein Überdispersionsparameter σ_o^2 wird anhand der Pearson- χ^2 -Statistik berechnet (`scale = pearson`; s. [14, S. 1408]).

PROC GLIMMIX wurde für die Lösung verallgemeinerter linearer gemischter Modelle (GLMM) entwickelt und berechnet Parameterschätzungen über die Pseudo-Quasi-Likelihood-Methode (PQL; [19]), die für die hier behandelten marginalen Modelle (ohne zufällige Effekte) aber äquivalent zu den ML- bzw. QL-Methoden sind. Zusätzlich lassen sich in PROC GLIMMIX aber räumliche Kovarianzstrukturen (z. B. sphärisch, gauß'sch) angeben und deren Parameter ebenfalls schätzen. Alternativ lassen sich, z. B. nach einem PROC-GENMOD-Aufruf (mit unabhängigen Beobachtungen), die Kovarianzparameter auch gutachterlich anhand eines empirischen Semivariogramms der sog. Pearson-Residuen [Option `reschi = VarName` im GENMOD-OUTPUT-Statement, s. u.] schätzen und als `fix` in PROC GLIMMIX angeben. Für die Berechnung empirischer Semivariogramme steht in SAS die Prozedur VARIOGRAM zur Verfügung.

Abgesehen von der Spezifikation der räumlichen Kovarianzparameter gleichen viele Statements und Optionen der GLIMMIX-Prozedur denen von PROC GENMOD.

```
PROC GLIMMIX data = segments;
CLASS VegType;
MODEL n = DistWater VegType /
      dist = poisson
      link = log
      offset = LnArea
;
RANDOM _residual_ / subject = intercept
      type = sp(sph) (xCoord yCoord);
```

```
RANDOM _residual_; /* Zusätzlicher Nugget-Effekt */  
PARMS (1300) (1) (2); /* Range, Nugget, Überdispersion/Sill */  
RUN;
```

Das erste RANDOM-Statement enthält die Angaben, daß das Modell räumlich autokorrelierte Fehler enthält (`type = sp(·)`), wobei alle Beobachtungen korreliert sind (`subject = intercept`) und es sich um eine sphärische Kovarianzstruktur handelt (`sp(sph)`) und die (kartesischen) x- und y-Koordinaten in den Variablen, `xCoord` und `yCoord` gespeichert sind. Durch ein zweites (optionales) RANDOM-Statement kann ggf. ein zusätzlicher Nugget-Effekt σ_n^2 hinzugefügt werden. Das PARMS-Statement ermöglicht wie in anderen Prozeduren die hier sehr zu empfehlende Angabe von Startwerten für die Kovarianzparameterschätzungen. Hier ist auch die Festsetzung oder Begrenzung von Kovarianzparametern möglich (Optionen `hold =`, `lowerb =` und `upperb =`).

3.4 Vorhersagen

Zur Berechnung kleinräumiger Vorhersagen für das gesamte Untersuchungsgebiet wird zunächst mit Hilfe weiterer, über SAS-Makros gesteuerter Geoprocessing-Skripte ein Gitter aus quadratischen, gleich großen Flächen über das Untersuchungsgebiet gelegt (`%CreatePolyGrid`). Im Anschluß werden für die Flächen, wie zuvor für die Segmente, die Werte der räumlichen Kovariaten bestimmt (`%SpatialQueryPolys`). Eine Ergebnistabelle, die diese Werte und die Flächengröße der Zellen des Vorhersagegitters enthält wird wiederum im SAS-Format abgelegt. Da die Strukturen der Segmenttabelle und der Vorhersagetabelle mit Ausnahme der Information über die Anzahl der beobachteten Objekte gleichen, lassen sich Vorhersagen des Count-Modells ohne räumliche Korrelation sehr leicht in PROC GENMOD berechnen. Wie viele anderen statistische Prozeduren führt PROC GENMOD Parameterschätzungen nur anhand von Datensätzen durch, in denen die abhängige Variable nicht leer (`missing`) ist, Vorhersagen werden jedoch für alle Datensätze berechnet. Es sind also lediglich Beobachtungs- und Vorhersagetabelle (`per SET`) zusammenzufügen und die Ergebnistabelle als Analysetabelle in PROC GENMOD zu verwenden. Vorhersagen (und weitere statistische Kenngrößen) können auch hier über das OUTPUT-Statement in eine neue Tabelle ausgeschrieben werden

```
OUTPUT out = predictions pred = mu;
```

Die Variable `mu` enthält dann den in Abhängigkeit von den für eine Vorhersagezelle bestimmten räumlichen Kovariaten \mathbf{x}_0 berechneten Erwartungswert $\mu_o = \exp(\mathbf{x}_0' \hat{\boldsymbol{\beta}})$. Die oben angesprochenen Pearson-Residuen können über die Option `reschi = VariablenName` in die Ausgabetablelle geschrieben werden.

Die Berechnung von Vorhersagen unter Einbezug räumlicher Autokorrelation gestaltet sich dagegen deutlich schwieriger. Im Gegensatz zu PROC MIXED (und damit auch zum GLIMMIX-Makro(!), das PROC MIXED verwendet) führt PROC GLIMMIX kein

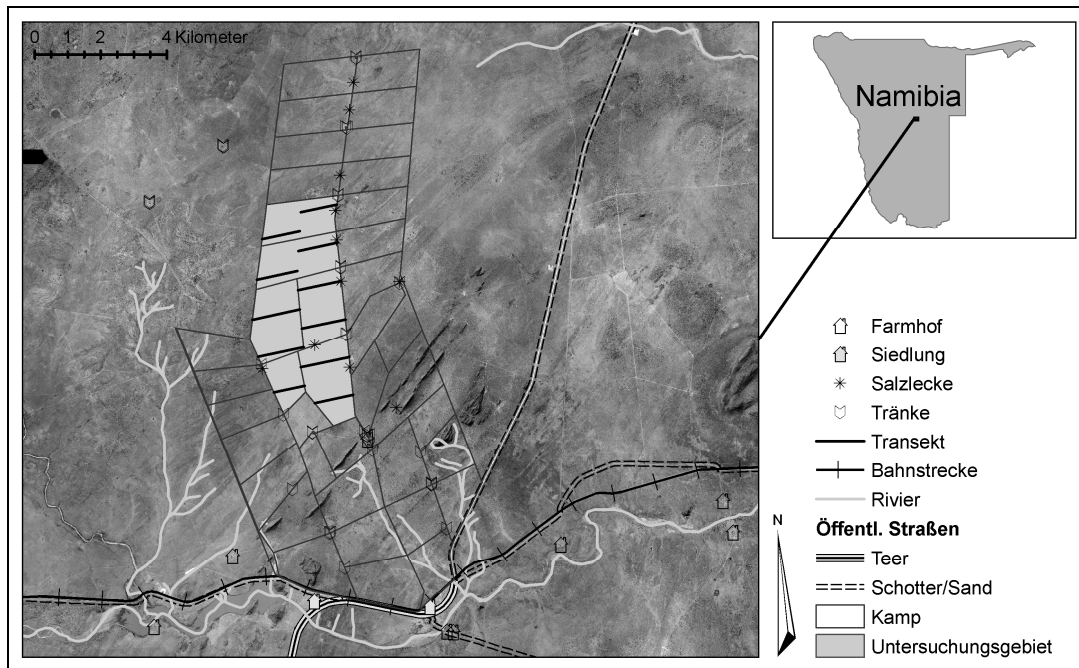


Abbildung 2: Untersuchungsgebiet, Transekte und ein Teil der erfaßten Geodaten.

Universelles Kriging für Beobachtungen durch, in denen die abhängige Variable leer ist. Vorhersagen des (marginalen) log-linearen Modells ohne zufällige Effekte werden wie in PROC GENMOD über $\mu_o = \exp(\mathbf{x}_o' \hat{\boldsymbol{\beta}})$ berechnet. Im wesentlichen unter Verwendung von SAS/IML wurde daher das SAS-Makro %MargGlmSpatPred entwickelt, das die Berechnung räumlicher Vorhersagen nach (2.7) und deren MSPEs nach (2.8) durchführt. Das Makro wird im Anschluß an PROC GLIMMIX aufgerufen und benötigt neben der zusammengesetzten Beobachtungs- und Vorhersagetabelle die GLIMMIX-ODS-Tabellen CovParms, ParameterEstimates und CovB, die die geschätzten (oder festgesetzten) Kovarianzparameter $\hat{\boldsymbol{\alpha}}$, die geschätzten Parameter der festen Effekte $\hat{\boldsymbol{\beta}}$ und deren Kovarianzmatrix $\text{cov}(\hat{\boldsymbol{\beta}})$ enthalten.

Die berechneten Vorhersagen beider Ansätze können im Anschluß, wiederum unter Verwendung von Geoprocessing-Skripten, in die Attributtabelle der Feature Class des Vorhersagegitters geschrieben (%JoinSasFcAt) und dann, z. B. in ArcMap, kartographisch dargestellt werden (s. Abb. 3).

4 Anwendungsbeispiel

Die Daten, für die die Anwendung der vorstehend beschriebenen Daten und Software-Umsetzungen exemplarisch vorgestellt werden sollen, entstammen Wildzählungen, die zwischen Juni und September 2002 in einem Untersuchungsgebiet im zentralen Osten Namibias durchgeführt worden sind. Untersucht wurden fünf endemische Wildarten, von denen hier nur eine, die Oryxantilope (*Oryx gazella*) näher betrachtet werden soll. Das Untersuchungsgebiet von ca. 16.1 km² Größe ist Teil eines insgesamt ca. 90 km² großen Farmkomplexes. Insgesamt achtmal wurden die elf, in Abbildung 2 dargestellten Transekte mit einer Gesamtlänge L von 14.0 km abgeritten und dabei die Entfernungen

Tabelle 1: Parameterschätzungen der Kovariaten „Kürzeste Entfernung zum nächsten Vley“ ($\hat{\beta}$), Standardfehler (in Klammern) und p-Werte, Schätzungen des Überdispersionsparameters σ_o^2 und der Reichweite des räumlichen Prozesses.

	$\hat{\beta}$	p-Wert ($\hat{\beta}$)	$\hat{\sigma}_o^2$	Reichweite
Modell ohne räuml. Autokorr.	1.18 (0.22)	<0.001	5.81	/
Modell mit räuml. Autokorr.	1.20 (0.22)	<0.001	5.68	636.5 m

zu den entdeckten Tiergruppen und deren Gruppengröße erfaßt. Die Wiederholung der Aufnahmen war erforderlich, da in den einzelnen Erhebungen nicht genügend große Beobachtungsanzahlen erreicht wurden. Neben der Verortung der Objektpositionen und der Transekte wurde eine umfangreiche Geodatenbank aufgebaut, die zusätzliche, die lokale Wilddichte potentiell erklärende geographische Informationen enthielt, von denen einige ebenfalls in Abbildung 2 dargestellt sind. Hierzu gehörten beispielsweise die Entfernungen zur nächsten Tränke, Salzlecke, Siedlung und öffentlichen Straße (Entfernungsmessung), Geländehöhe und Vegetationsindizes (Punktverschneidung), sowie der Habitattyp (Flächenverschneidung).

In den acht Aufnahmen wurden insgesamt 109 Objekte (Tiergruppen) entdeckt. Die mittlere Gruppengröße wurde mit 4.24 bestimmt, die effektive Streifenbreite anhand der Entfernungsdaten mit 112.44 m. Der klassische Linientransektdichteschätzer aus (2.1) ergab damit eine Individuendichte von 11.20 Tieren/km² mit einem Standardfehler von 2.96.

Die effektive Streifenbreite ging im Anschluß als Streifenbreite in die Segmentgenerierung ein; als Segmentlänge c wurden 600 m verwendet, so daß sich eine Segment- und damit „Beobachtungs“anzahl für die nachfolgenden Analysen von 29 ergab. Nicht zuletzt infolge der geringen Segmentanzahl konnte von den zahlreichen zur Verfügung stehenden räumlichen Kovariaten nur eine, die Entfernung zu sog. Vleys⁵, als signifikant ausgewiesen werden; die Parameterschätzungen für das Modell mit (PROC GLMIIX) und ohne (PROC GENMOD) räumliche Autokorrelation und die zugehörigen Standardfehler sind in Tabelle 1 aufgeführt. Zu erkennen sind sehr ähnliche Parameterschätzungen mit identischen Standardfehlern beider Modellansätze. Ersteres läßt sich oft beobachten und stimmt auch mit Simulationsergebnissen [1] überein; die Parameterschätzungen werden also nur gering durch die Verwendung von Σ_{yy} anstelle von V beeinflußt. Die identischen Standardfehler hingegen überraschen ein wenig; hier liefert das Modell mit Autokorrelation in der Regel größere Werte. Ebenfalls in Tabelle 1 aufgeführt sind die Schätzungen der Überdispersionsparameter und der Reichweite des räumlichen Prozesses, die ebenfalls in PROC GLIMMIX bestimmt wurde. Ein Nugget-Effekt wurde nicht verwendet.

Abbildung 3 zeigt die nach (2.3) und (2.7) berechneten Vorhersagen der Individuenabundanz für beide Modellansätze, sowie die Positionen der entdeckten Objekte.

⁵ Vleys sind mit tonigen Sedimenten überzogene, meist vegetationsfreie Pfannen, um die herum die Vegetation jedoch oft üppiger ist.

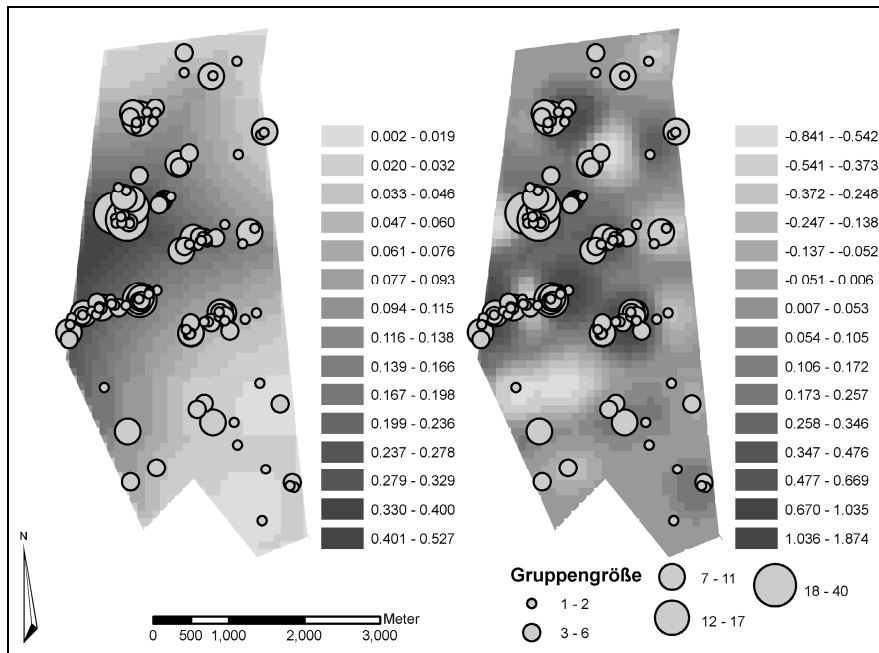


Abbildung 3: Räumliche Vorhersagen der Individuenabundanz von Oryxantilopen des Modells ohne (links) und mit (rechts) Berücksichtigung räumlicher Autokorrelation.

Tabelle 2: Schätzungen der Oryx-Gesamtabundanz \hat{M} und Standardfehler (in Klammern) des klassischen Linientransektverfahrens (LT) des Count-Modells ohne (IGLM) und des Count-Modells mit Berücksichtigung räumlicher Autokorrelation (SpGLM).

\hat{M}_{IGLM}	\hat{M}_{SpGLM}	\hat{M}_{LT}
170 (1.17)	161 (18.12)	180 (47.60)

Es ist deutlich zu erkennen, daß das Modell mit Autokorrelation sich deutlich stärker an den Objektpositionen ausrichtet und damit letztlich auch ein realistischeres Abbild der Dichteverteilung liefert. Nachteil des Modells mit Autokorrelation ist jedoch, wie die Legende zeigt, die Möglichkeit negativer und damit unsinniger Vorhersagen. Diese können sich als Folge der im Zusammenhang mit (2.7) erläuterten Anpassung des Erwartungswertes für eine neue Position ergeben.

Tabelle 2 zeigt die anhand der insgesamt drei Ansätze (klassisches Linientransektverfahren, Count-Modell mit und ohne Berücksichtigung räumlicher Autokorrelation) ermittelten Oryx-Gesamtabundanzschätzungen \hat{M} und zugehörige Standardfehler für das gesamte Untersuchungsgebiet. Zu erkennen sind durchaus vergleichbare (insbesondere gemessen an den hohen Standardfehlern für \hat{M}_{SpGLM} und \hat{M}_{LT}) Abundanzschätzungen der drei Ansätze. Die Standardfehler hingegen fallen sehr unterschiedlich aus. Der Einbezug einer Kovariaten kann offensichtlich zur Reduktion der Varianz der Abundanzschätzung beitragen, der sehr kleine Standardfehler des einfachen Count-Modell dürfte jedoch für die deutliche Unterschätzung der wahren Varianz infolge der Nichtbeachtung der räumlichen Autokorrelation sprechen.

Literatur

- [1] Albert P.S., McShane L.M. (1995) A Generalized Estimating Equations Approach for Spatially Correlated Binary Data: Applications to the Analysis of Neuroimaging Data. *Biometrics*, 51, 627–638.
- [2] Buckland S.T., Anderson D.R., Burnham K.P., Laake J.L., Borchers D.L., Thomas L. (2001) *Introduction to Distance Sampling - Estimating abundance of biological populations*. Oxford University Press, Oxford.
- [3] Buckland S.T., Anderson D.R., Burnham K.P., Laake J.L., Borchers D.L., Thomas L. (2004) *Advanced Distance Sampling – Estimating abundance of biological populations*. Oxford University Press, Oxford.
- [4] Cressie N.A.C. (1991) *Statistics for spatial data*. John Wiley & Sons, New York.
- [5] ESRI (2004) Environmental Systems Research Institute. *Geoprocessing in ArcGIS*. ESRI, 380 New York Street, Redlands, California.
- [6] ESRI (2005) Environmental Systems Research Institute. *Writing Geoprocessing Scripts With ArcGIS*. ESRI, 380 New York Street, Redlands, California.
- [7] Gotway C.A., Stroup W.W. (1997) A Generalized Linear Model Approach to Spatial Data Analysis and Prediction. *Journal of the American Statistical Association*, 2, 157–178.
- [8] Gotway C.A., Wolfinger R.D. (2003) Spatial Prediction of Counts and Rates, *Statistics in Medicine*, 22, 1415–1532.
- [9] Hedley S.L. (2000) *Modelling Heterogeneity in Cetacean Surveys*. PhD thesis, University of St Andrews, Scotland.
- [10] Hedley S.L.; Buckland S.T. (2004) Spatial Models for Line Transect Sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 9, 181–199
- [11] Liang K.Y., Zeger S.L. (1986) Longitudinal Data Analysis Using Generalized Linear Models. *Biometrika*, 73, 13–22.
- [12] Mader F. (2007) *Räumliche, GIS-gestützte Analyse von Linientransektstichproben*. Dissertation, Universität Göttingen – in Vorbereitung.
- [13] McCullagh P., Nelder J.A. (1989) *Monographs on statistics and applied probability*. Vol. 37: *Generalized linear models*. 2nd ed., Chapman & Hall, London.
- [14] SAS Institute Inc. (2004) *SAS/STAT 9.1 User's Guide*. SAS Institute Inc., Cary, NC.
- [15] SAS Institute Inc. (2006) *The GLIMMIX Procedure*. Version: June, 2006. http://www.sas.com/apps/demosdownloads/sasstatglimmix_PROD__sysdep.jsp?packageID=000353, Abruf: 21.02.2007
- [16] Schabenberger O (2005) Introducing the GLIMMIX Procedure for Generalized Linear Mixed Models. *Proceedings of the 18th North East SAS Users Group Inc.*

(NESUG) Conference 18th Annual Conference Proceedings. Portland, Maine, paper AN4

- [17] 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. (2003) *Distance 4.1. Release 2*. St Andrews, Scotland, – <http://www.ruwpa.stand.ac.uk/distance/>
- [18] Wedderburn R.W.M. (1974) Quasi-Likelihood Functions, Generalized Linear Models and the Gauss-Newton Method. *Biometrika*, 61, 439–447
- [19] Wolfinger R.D., O’Connell M. (1993) Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. *Journal of Statistical Computation and Simulation*, 48, 233–243
- [20] Zeger S.L., Liang K.Y. (1986) Longitudinal Data Analysis for Discrete and Continuous Outcomes. *Biometrics*, 42, 121–130.