2

Modellierung der Art-Habitat-Beziehung – ein Überblick über die Verfahren der Habitatmodellierung

Boris Schröder[1] & Björn Reineking[2]

[1] Institut für Geoökologie, Universität Potsdam, Postfach 601553, D-14415 Potsdam, Email: boschroe@rz.uni-potsdam.de

[2] UFZ Umweltforschungszentrum Leipzig-Halle GmbH, Sektion Ökosystemanalyse, Postfach 500136, D-04301 Leipzig und ETH Zürich, UNS, Haldenbachstr. 44, ETH-Zentrum HAD, CH-8092 Zürich, Email: bjoern.reineking@ufz.de

2.1 Grundsätzliches - Voraussetzungen, theoretischer Hintergrund, Ziele & Prinzip, Design zur Datenerhebung

2.1.1 Was sind Habitatmodelle?

Habitatmodelle beschreiben funktionale Zusammenhänge der Beziehung zwischen Organismen und ihrem Lebensraum und quantifizieren die Qualität des Habitats aus der Sicht dieser Organismen (Morrison et al. 1998; Schröder 2000). Statistische Habitatmodelle schätzen aus Verbreitungsdaten (Responsevariable) und Habitateigenschaften (Prädiktorvariablen) für jeweils abgegrenzte homogene Untersuchungseinheiten die Vorkommenswahrscheinlichkeit bzw. prognostizieren die Inzidenz, d.h. Vorkommen oder Nichtvorkommen der Art (Scott et al. 2002). Zudem erlauben sie, die Wichtigkeit einzelner Habitatparameter für die Prognose zu analysieren und auf dieser Grundlage Habitatpräferenzen abzuleiten (z.B. Lindenmayer et al. 1991; Peeters & Gardeniers 1998).

2.1.2 Mögliche Fragestellungen der Habitatmodellierung

In der Habitatmodellierung werden zwei verschiedene, aber miteinander verbundene Fragestellungen verfolgt (Fielding & Haworth 1995; MacNally 2000; Morrison et al. 1998). Zum einen das Verständnis der Art-Habitat-Beziehungen, d.h. die Analyse und Quantifizierung der Habitatansprüche sowie die Charakterisierung von Aspekten der realisierten Nische (z.B. Austin et al. 1990; De Swart et al. 1994; Eyre et al. 1992; Peeters & Gardeniers 1998). Mögliche Fragestellungen aus den Bereichen Autökologie, Biogeographie und Naturschutzbiologie sind beispielsweise: Welche Umwelteigenschaften bestimmen die räumliche Verteilung von Art *x* oder Artengemeinschaft *y*? - Lässt sich der Lebensraum von Art *y* im regionalen Maßstab durch die Umwelteigenschaften *a*, *b* und *c* mit hinreichender Genauigkeit beschreiben? - Wie soll ein Reservat, das dem Schutz von Art *x* dient „gemanagt" werden, bzw. welches wäre die optimale Erweiterung des Reservats, um den Schutzeffekt zu vergrößern? - Wo verspricht eine Wiederansiedlung gefährdeter Arten Erfolg?

Zum anderen geht es in der Habitatmodellierung um die Prognose der räumlichen Verteilung von Organismen. Hier sind mögliche Fragestellungen: - Welche Verteilung ist in einem nicht untersuchten Gebiet zu erwarten? - Welche Veränderung der Zusammensetzung von Artengemeinschaft x ist durch den Klimawandel zu erwarten? - Wie kann ein sinnvolles Monitoringsystem zur Beurteilung des Erfolgs von Managementmaßnahmen für Art *y* gestaltet werden?

Als dritter, allerdings innerhalb der Habitatmodellierung im Gegensatz zur medizinischen Statistik nicht

Dormann CF, Blaschke T, Lausch A, Schröder B, Söndgerath D (Hrsg.)(2004) Habitatmodelle – Methodik, Anwendung, Nutzen. Tagungsband zum Workshop vom 8.-10. Oktober 2003 am UFZ Leipzig. UFZ-Berichte 9/2004.

so relevanter Punkt sollte noch das Testen einzelner Hypothesen angeführt werden; etwa wenn allein der Frage nachgegangen wird, ob ein bestimmter Umweltfaktor einen signifikanten Einfluss auf die Verteilung der Organismen hat. Die Fragestellung, die mittels der Habitatmodellierung beantwortet werden soll, bestimmt die Auswahl des statistischen Verfahrens und der anvisierten Modellkomplexität sowie die Strategie bei der Modellbildung (Harrell 2001).

2.1.3 Theoretischer Hintergrund

Der theoretische Hintergrund der Habitatmodellierung umfasst die Theorie der realisierten Nische (Austin et al. 1990; Franklin 1995; Hutchinson 1957; Leibold 1995) sowie der Mechanismen der skalenabhängigen Habitatselektion (Johnson 1980; Mackey & Lindenmayer 2001). Poff (1997) und Schröder (2001) definieren Konzepte von Filterkaskaden, um zu erklären, welche Arten aus dem regionalen Artenpool tatsächlich in den lokalen Gemeinschaften vorkommen. Die Habitateigenschaften wirken wie selektive Filter, die hierarchisch auf verschiedenen räumlichen und zeitlichen Skalen definiert werden können (s. Beispiele bei Oppel et al. 2004; Rolstad et al. 2000). Um einen Filter zu passieren, muss die Art funktionale Eigenschaften (*traits*) aufweisen, die der selektiven Charakteristik des Filters entsprechen (Poff 1997; Townsend et al. 2003; Weiher & Keddy 1999). In der von Schröder (2001) definierten Filterkaskade (Abb. 2.1) findet zuerst ein Abgleich der abiotischen Verhältnisse mit den Ressourcenansprüchen der Arten statt, wodurch Aspekte der fundamentalen Nische beschrieben werden. Hierfür ließen sich nach Poff (1997) oder Mackey & Lindenmayer (2001) mehrere abiotische Filter auf verschiedenen räumlichen und zeitlichen Skalen definieren. Danach wird geklärt, ob die räumliche Verteilung der Habitatpatches in der Landschaft mit der Ausbreitungscharakteristik der Art harmoniert oder ob die Fragmentierung der Landschaft durch Isolation die effektive Besiedlung verhindert (Johst et al. 2002). Selbst wenn beide Bedingungen erfüllt sind, so ist möglich, dass der dritte Filter durch biotische Interaktionen wie interspezifische Konkurrenz, Parasitismus oder Prädation die Persistenz einer Population verhindert (realisierte Nische, z.B. Malanson et al. 1992). Da die Habitatmodelle aus Verbreitungsdaten geschätzt werden, wird normalerweise die realisierte und nicht die fundamentale Nische modelliert; implizit sind dabei biotische Interaktionen und negative stochastische Effekte mit berücksichtigt (Guisan et al. 2002).

2.1.4 Annahmen, Möglichkeiten und Einschränkungen

Habitatmodelle liefern Prognosen für Verteilungsmuster von Arten im Raum. Der räumliche Bezug entsteht über die Habitateigenschaften von Raumeinheiten, üblicherweise auf der Ebene von Kartierungseinheiten oder Rasterquadraten. Sie sind nicht dynamisch, da sie aus einzelnen „schlaglichtartigen" Erhebungen abgeleitet werden, für die implizit eine Quasi-Gleichgewichtssituation angenommen wird (Austin 2002; Guisan & Zimmermann 2000; Kleyer et al. 2000; O'Connor 2002), d.h. dass die Umwelt sich im Vergleich zur Lebenserwartung des Organismus nur langsam verändert. Die Dynamik eines Habitats ist dann eine Habitateigenschaft, die - beispielsweise als Frequenz eines Störungsereignisses - ebenso wie andere, statische Habitateigenschaften in die Analyse eingehen kann (Schröder & Reineking 2004a). Aus Habitatmodellen abgeleitete Prognosen für Szenarien beschreiben lediglich Potentiale der möglichen Verbreitung; ob und wie die prognostizierten Zustände erreicht werden, kann nur durch dynamische Modelle simuliert werden; denn Habitatmodelle können keine Abbildung der Populationsdynamik verbunden mit Aussagen zu Populationsgrößen leisten (Schamberger & O'Neil 1986). Diese wird erst durch die Verknüpfung mit Modellen zur räumlichen (Meta-)Populationsdynamik möglich (z.B. Collingham et al. 2000; Söndgerath & Schröder 2002; Wadsworth et al. 2000; Wahlberg et al. 1996, oder Arbeiten zur *population viability analysis*: PVA).

Eine grundsätzliche Annahme ist, dass die Tiere die Biotope derart nutzen und ihre Habitate so auswählen, dass ihre Fitness optimiert wird (Southwood 1977). Habitate höherer Qualität werden proportional häufiger genutzt, so dass bei einer überproportionalen Nutzung von einer Habitatpräferenz, bei einer subproportionalen von einer Meidung ausgegangen werden kann (Aebischer et al. 1993; Manly et al. 1993). Für alle statistischen Verfahren, die nicht explizit räumliche oder zeitliche Autokorrelation berücksichtigen (s.u.), muss die Grundannahme der Unabhängigkeit der einzelnen Stichprobenwerte in der Stichprobe erfüllt sein.

2.1.5 Probenahme-/Sampling Design zur Datenerhebung

Grundlage des Modellierungsprozesses ist die Formulierung eines konzeptionellen Modells, das die Auswahl der adäquaten Untersuchungsskala, eines geeigneten Probenahmedesigns und der zu erhebenden erklärenden Variablen umfasst (Guisan & Zimmermann 2000). Eine sorgfältige Planung, welche Variablen zu welchen Zeitpunkten und an welchen Orten erhoben werden sollen (*sampling design*), ist die Grundlage einer Erfolg versprechenden Habitatmodellierung. Nach Hirzel & Guisan (2002) und Wessels et al. (1998) sowie eigenen Erfahrungen (Bonn & Schröder 2001; Oppel et al. 2004) erhält man die besten und robustesten Ergebnisse bei Anwendung eines stratifizierten Zufalls-

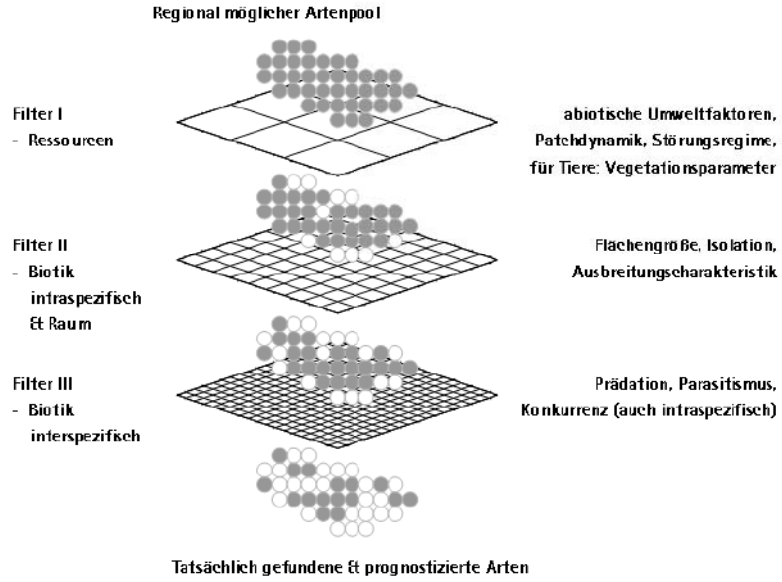


Abb. 2.1. Filterkaskade zur Erklärung der Zusammensetzung von Gemeinschaften durch Ausschluss regional möglicher Arten auf Grund von Landschaftsfiltern (nach Schröder 2001).

designs (*randomly stratified sampling*). Die Stratifikation erfolgt nach den wichtigsten Gradienten, die aufgrund von Vorstudien oder Literaturauswertungen ausgewählt und in wenige Klassen aufgeteilt werden. Für die sich ergebenden Klassenkombinationen, die Straten, erfolgt dann eine zufällige Probeflächenauswahl mit ausreichendem Mindestabstand (räumliche Autokorrelation!) und in ausreichender Wiederholung. Als Faustregel für die Datensatzgröße geben Steyerberg et al. (2001b) einen Wert von zehn Präsenzen pro Variable an. Zu bedenken sind auch die Gradientenlängen, welche die Form der Responsekurven bestimmen (Oksanen & Minchin 2002) und Einfluss auf den Gültigkeitsbereich der Modelle haben.

2.1.6 Erklärende Variablen (Prädiktorvariablen, Kovariaten)

Eine Vielzahl von Variablen ist potentiell dazu geeignet, die räumlichen Verteilungsmuster von Arten zu erklären (vgl. Abb. 2.1). Hierzu zählen neben den geologischen, topographischen und edaphischen Habitatfaktoren (Austin et al. 1996; Lamouroux & Capra 2002), klimatische Faktoren (s. *climatic envelopes* Davis et al. 1998; Pearson & Dawson 2003), die Landnutzung und ihre Entwicklung (Pearson et al. 1999; Verboom et al. 1991), biotische Habitatfaktoren wie z.B. Prädation (Reading et al. 1996), Parasiten (Balcom & Yahner 1996) oder Konkurrenz (Massolo & Meriggi 1998), die Landschaftsstruktur und -heterogenität (Fahrig & Johnson 1998) sowie Flächengrößen, Konnektivität und Verinselung der Landschaft (Adler & Wilson 1985; Biedermann 2003; Kuhn & Kleyer 1999). Die erklärenden Variablen können als Messungen und Beobachtungen vorliegen oder aus Karten oder per GIS-Analysen abgeleitet werden (Guisan & Zimmermann 2000). Ihre Auswahl erfolgt grundsätzlich hypothesengesteuert (Morrison et al. 1998). Die Modelle sind um so robuster und allgemeingültiger, je enger die erklärenden Variablen mit den zugrunde liegenden Mechanismen und physiologischen Prozessen zusammenhängen (Poff 1997). Austin (1985) unterscheidet in diesem Zusammenhang a) Gradienten von Ressourcen, die konsumiert werden können, b) direkte Gradienten mit physiologischer Bedeutung und c) indirekte Gradienten, die zwar keine direkte physiologische Bedeutung haben, aber leichter messbar sind und Kombinationen von Ressourcen- und direkten Gradienten ersetzen können (Guisan et al. 1999).

Die verwendeten statistischen Methoden sind stets korrelativ; aus ihnen lässt sich prinzipiell keine Kausalität ableiten, wohl aber die Beschreibung funktioneller Art-Habitat-Beziehungen (Austin 2002). Je mehr direkt wirksame erklärende Variablen im Modell Berücksichtigung finden, desto größer ist die Wahrscheinlichkeit, dass diese die dahinter liegenden Mechanismen und Prozesse gut beschreiben. Die Modelle liefern damit Hypothesen, die durch Experimente, theoretische Analysen und wiederholte Untersuchungen geprüft werden müssen (Austin 2002; Morrison et al. 1998).

Prädiktorvariablen können auf unterschiedlichen Skalen gemessen werden; metrisch skalierte Variablen „verbrauchen“ im linearen Modell (s.u.) jeweils nur einen Freiheitsgrad (*degree of freedom*, abgekürzt mit df). Sie sind deshalb kategorialen Variablen, die bei k Kategorien $(k-1)$ Freiheitsgrade beanspruchen, wenn immer möglich vorzuziehen (Harrell 2001). Je mehr Freiheitsgrade verbraucht werden, desto höher ist die Modellkomplexität und damit auch die Gefahr der Überanpassung des Modells (*overfitting*). Diese führt zu einem instabilen, unzuverlässigen Modell, das vielleicht für den Datensatz, auf dessen Grundlage es geschätzt wurde (Trainingsdaten), gute Prognosen ergibt, aber dafür umso schlechtere Ergebnisse für unabhängige Testdaten liefert (Harrell 2001). Überanpassung ist ein häufig anzutreffendes Phänomen, weshalb Fragen der Modellvereinfachung und Modellvalidierung eine große Bedeutung im Modellbildungsprozess haben.

Sparsamere, weniger komplexe Modelle mit reduzierter Anzahl an Variablen und/oder geringerer Flexibilität in den Responsekurven (s.u.) haben normalerweise eine geringere Varianz (*variance*), die aber durch eine erhöhte Verzerrung (*bias*) „erkauft“ wird. Ihre Vorhersagen sind zwar sicherer, zeigen dafür aber systematische Abweichungen vom „wahren“ Wert. Dieser Zielkonflikt zwischen Ungenauigkeit und Verzerrung der Modellschätzungen wird als Bias-variance trade-off beschrieben (s. Abb. 2.2 Hastie et al. 2001). Da die Varianz häufig über den Bias dominiert, liefern verzerrte Schätzverfahren häufig bessere Prognosen (Harrell 2001). Dieser Aspekt wird im Beitrag von Reineking und Schröder (2004, dieser Band) ausführlicher diskutiert.

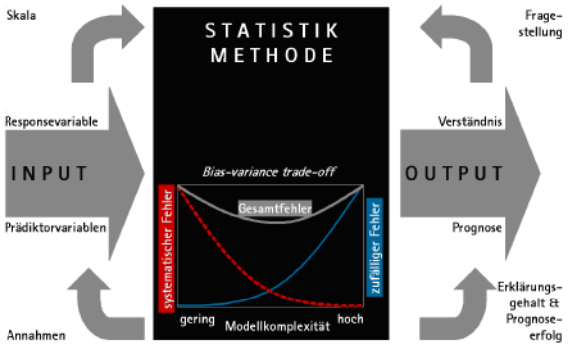


Abb. 2.2. Skala der Responsevariable und Fragestellung bestimmen die Auswahl des statistischen Verfahrens. Die Flexibilität des Verfahrens bestimmt, welche Annahmen hinsichtlich der Prädiktorvariablen und der Beziehung zwischen Prädiktoren und Response getroffen werden müssen. Je höher die Flexibilität, desto höher häufig der Prognoseerfolg, aber desto geringer der Erklärungsgehalt.

2.2 Methodenüberblick

Die vorzustellenden Verfahren passen alle in das in Abb. 2.2 dargestellte Schema. Sie unterscheiden sich hinsichtlich des mit der Habitatmodellierung verfolgten Ziels, d.h. des gewünschten Outputs und natürlich auch hinsichtlich des verfügbaren Inputs. Für verschiedene skalierte Responsevariablen eignen sich unterschiedliche Methoden.

Parametrische Verfahren sind immer weniger flexibel als semi-parametrische/nicht-parametrische, die hinsichtlich der Prädiktorvariablen und ihrer Beziehung zur Responsevariablen weniger Annahmen verlangen. Diesen Verfahren liegen z.B. keine Verteilungsannahmen zugrunde. Je flexibler die Methoden sind, desto höher ist potentiell der Prognoseerfolg, aber desto geringer ist dann oft auch der zum Verständnis der Art-Habitat-Beziehung gelieferte Erklärungsgehalt. Andererseits gilt: je flexibler die Methoden sind, desto komplexer können die Modelle sein, d.h. desto mehr Freiheitsgrade können sie verbrauchen und umso vorsichtiger muss der Gefahr des *overfitting* begegnet werden. Im Fall von Datenarmut, also bei sehr kleinen Datensätzen, können hingegen oftmals mit Modellen, von dem wir im Grunde wissen, dass sie zu einfach sind, bessere Ergebnisse erzielt werden.

Das grundsätzliche Prinzip der Habitatmodellierung, das bei allen unten aufgeführten Verfahren in vergleichbarer Weise Anwendung findet, verdeutlicht Abb. 2.3. Hierbei steht die dargestellte logistische Regressionsfunktion (visualisiert durch eine Responsekurve oben rechts, s.u.) beispielhaft für die Vielzahl anwendbarer statistischer Verfahren. In 2.2.2 wird die logistische Regression vergleichsweise umfangreich dargestellt.

2.2.1 Historische Entwicklung

Die Analyse der Beziehung zwischen den Arten und ihrer Umwelt wurde vom U.S. Fish & Wildlife Service (1981) mit der Entwicklung von Habitateignungsindex-Modellen (*habitat suitability index-models*), die Teil eines Verfahrens zur Habitatbewertung waren (Pearsall et al. 1986; U.S. Fish & Wildlife Service 1980) erstmals institutionalisiert. Die Habitateignungsindices wurden als geometrische Mittelwerte aus einer Menge auf das Intervall [0,1] skalierter Umweltvariablen berechnet, von denen basierend auf Expertenwissen der größte Einfluss auf Verteilung und Abundanz der Arten erwartet wurde (z.B. Reading et al. 1996).

Im Zuge der allgemeinen Verfügbarkeit geeigneter Software wurden dann verstärkt statistische Verfahren zur quantitativen Analyse empirischer Daten und zur Modellbildung eingesetzt (Brennan et al. 1986; Morrison et al. 1998). Anfangs war dies vor allem die Diskriminanzanalyse, die dann durch Regressionsanalysen

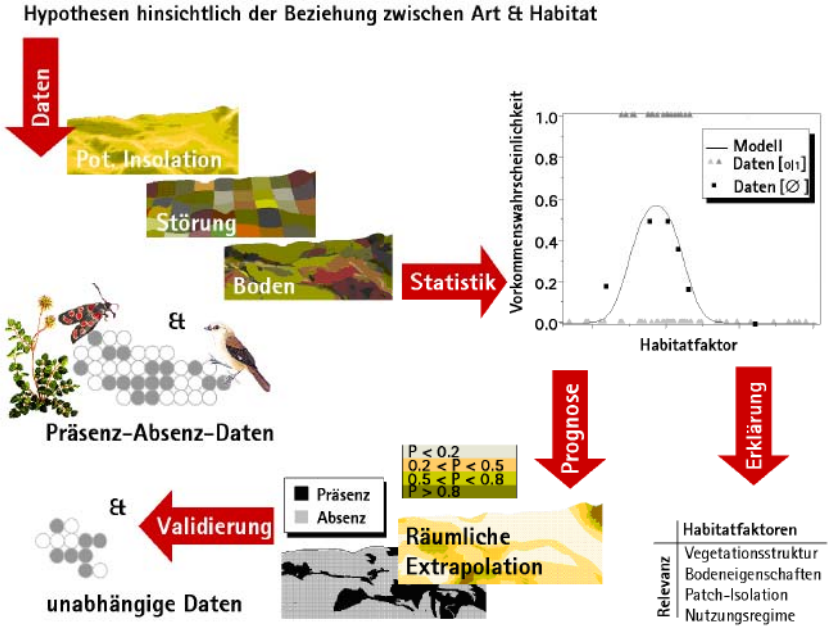


Abb. 2.3. Grundprinzip der Habitatanalyse und -modellierung: Erhebung von Präsenz-Absenz-Daten der zu modellierenden Arten und ausgewählter erklärender Variablen, Schätzung eines statistischen Modells und Modellbewertung, Analyse der relevanten Habitatfaktoren, Prognose und räumliche Extrapolation der Vorkommenswahrscheinlichkeiten, externe oder interne Modellvalidierung.

mehr oder weniger abgelöst wurde - in der Reihenfolge ihrer Entwicklung: allgemeine lineare Modelle (*general linear models*), verallgemeinerte lineare Modelle (*generalised linear models*: GLMs, McCullough & Nelder 1989) mit ihrem wohl wichtigsten Vertreter, der logistischen Regression, verallgemeinerte gemischte lineare Modelle (*generalised linear mixed models*: GLMMs, Wolfinger & OConnell 1993) sowie verallgemeinerte additive Modelle (*generalised additive models*: GAMs, Hastie & Tibshirani 1990). Aktuell werden auch Verfahren aus der Bayes-Statistik und aus dem Bereich des Datamining wie Klassifikations- und Regressionsbäume (*classification and regression trees*: CART) sowie neuronale Netze (*artificial neural networks*: ANN) angewendet. Abb. 2.4 zeigt die Ergebnisse einer Web-Recherche über die Anwendung dieser Verfahren in Veröffentlichungen zur Habitatmodellierung aus dem letzten Jahrzehnt, die im folgenden kursorisch beschrieben werden.

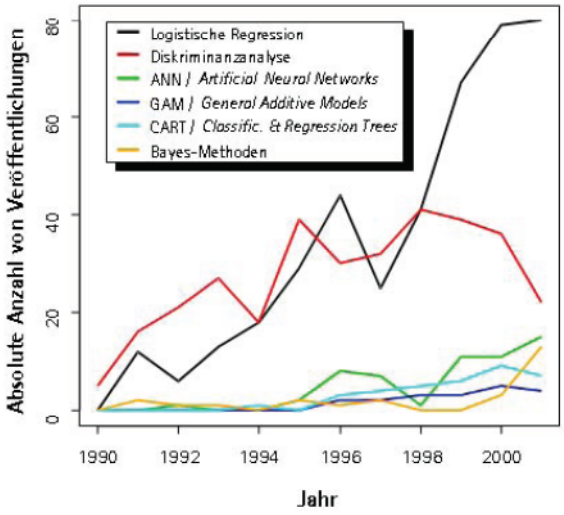


Abb. 2.4. Ergebnisse einer Recherche im Web of Science 1990-2001: Absolute Anzahl der Veröffentlichungen, die eines der angeführten Verfahren zur Habitatanalyse verwenden.

2.2.2 Regressionsanalyse

Regressionsverfahren modellieren die funktionelle Abhängigkeit einer Responsevariable - z.B. das Vorkommen einer Art - von einer oder mehreren erklärenden Variablen (Prädiktorvariablen). Crawley (2002); Quinn & Keough (2002); Guisan et al. (2002) liefern exzellente Überblicke zur Entwicklung der Regressionsverfahren, die hier kurz zusammengefasst werden.

Lineare Regression

Als eines der ältesten Verfahren der Statistik stellt die lineare Regression die Grundlage der Regressionsverfahren dar. Das einfache lineare Regressionsmodell folgt Gl. 2.1.

$$Y = \alpha + \sum_{j=1}^{p} \beta_j X_j + \varepsilon \qquad (2.1)$$

Hierbei ist Y die Responsevariable (abhängige Variable), α eine Konstante (y-Achsenabschnitt, Interzept), X_j, $j = 1, \ldots, p$ die Prädiktorvariablen (erklärende oder unabhängige Variablen, Kovariaten), β_j die zugehörigen Regressionskoeffizienten (Effektgrößen) und ε ist eine $N(0, \sigma^2)$-verteilte Zufallsvariable die nicht durch das Modell erklärte Variabilität z.B.aufgrund von Messfehlern repräsentiert. Wird nur eine Kovariate modelliert ($p = 1$), so spricht man von einfacher, sonst von multipler linearer Regression.

Durch Anwendung des Verfahrens der Minimierung der Abweichungsquadrate (*ordinary least-squares*: OLS) wird bei der Modellschätzung der Anteil der unerklärten Variabilität minimiert. Dem Modell zugrunde liegenden Annahmen sind konstante Varianz der abhängigen Variable über den gesamten Bereich der Beobachtungen (Homoskedastizität), normalverteilte Fehlerverteilung und die Linearität der Regressionsfunktion. Verletzungen der ersten und dritten Annahme können häufig durch Anwendung geeigneter Transformationen der Responsevariablen begegnet werden. Durch Verwendung von Polynomen der Prädiktorvariablen und anderen nichtlinearen Transformationen sowie von Interaktionen kann die Verletzung der dritten Annahme umgangen werden; diese Verfahren führen zu Modellen, die zwar nicht mehr linear in den Prädiktorvariablen, wohl aber in den Parametern sind (Allgemeines Lineares Modell, *general linear model*, Guisan et al. 2002).

Verallgemeinerte lineare Modelle (*Generalised Linear Models*: GLM)

Flexibler als die lineare Regression sind die verallgemeinerten linearen Modelle, welche die Modellierung von Art-Habitat-Beziehungen für alle denkbaren Skalen der Umweltvariablen (diskret oder stetig; nominal-, ordinal- oder intervallskaliert) in einem einzigen theoretischen Rahmen erlaubt (Yee & Mitchell 1991). Hier bildet ebenfalls eine Linearkombination der Prädiktorvariablen den linearen Prädiktor *LP*. *LP* fällt allerdings nicht mit dem Erwartungswert zusammen, sondern der Erwartungswert $E(Y_i|\mathbf{x}_i) = \mu_i$ wird durch eine Linkfunktion - $g(\mu_i)$ - mit *LP* verbunden, so dass der transformierte Wert auf der Ebene im p-dimensionalen Raum liegt, den die rechte Seite von Gl. 2.2 beschreibt.

$$g(\mu(\mathbf{x})) = LP = \alpha + \sum_{j=1}^{p} \beta_j x_j \qquad (2.2)$$

Während im einfachen linearen Modell als Linkfunktion die Identität gewählt wird, d.h. $g(z) = z$ gilt, die Fehler normalverteilt sind und die Varianz unabhängig von den Prädiktorvariablen ist, können durch GLMs auch Modelle behandelt werden, bei denen die Varianz von der Summer der jeweiligen Prädiktorvariablen abhängt. Für unterschiedliche nichtlineare Fehlerverteilungen, wie z.B. Poissonfehler bei Zählungsdaten oder Binomialfehler bei Verhältnisdaten stehen kanonische Linkfunktionen zur Verfügung. Für die beiden oben genannten Fälle sind dies beispielsweise log- und logit-Funktion (Crawley 2002). Neben dem logit-Link stehen für binäre abhängige Variablen auch andere Linkfunktionen zur Verfügung (z.B. probit, log-log, s. Fox 2002). Diese finden aber in der Habitatmodellierung selten Anwendung.

Die Schätzung der Regressionskoeffizienten erfolgt mittels des *Maximum likelihood*-Verfahrens (ML). Dabei wird eine Likelihoodfunktion aufgestellt, welche die Wahrscheinlichkeit der beobachteten Daten als eine Funktion der zu schätzenden Regressionskoeffizienten ausdrückt. Die Koeffizienten werden dabei so geschätzt, dass die Wahrscheinlichkeit, bei gegebenem Modell die empirischen Daten zu beobachten, maximiert wird. ML-Parameterschätzung ist verglichen mit dem OLS-Algorithmus unverzerrt (*unbiased*). In Fällen, in denen die Voraussetzungen der linearen Regression erfüllt sind, entspricht der OLS-Schätzer dem ML-Schätzer mit Identitätslink und normalverteilten Fehlern (Crawley 2002). Ein weiterer Vorteil der GLMs ist, dass sie ermöglichen, die Vorhersagen adäquat zu begrenzen (Guisan & Zimmermann 2000). Dies kommt z.B. bei der logistischen Regression zum Tragen, mit der Modelle auf der Grundlage von Präsenz (1)-Absenz (0)-Daten und Umweltvariablen geschätzt werden können. Die geschätzten Werte für die Responsevariable liegen zwischen 0 und 1 und sind als Vorkommenswahrscheinlichkeiten zu interpretieren.

Spezialfall: Logistische Regression (*Logistic Regression Model*: LRM)

Die logistische Regression ist das am häufigsten verwendete Verfahren der Habitatmodellierung zur Modellierung von Präsenz(1)-Absenz(0)-Daten, d.h. für Fälle, in denen die Responsevariable nur Einsen und Nullen enthält. Instruktive Beispiele finden sich u.a. bei Akçakaya et al. (1995); Bonn & Schröder (2001); Fielding & Haworth (1995); Lindenmayer et al. (1991); Manel et al. (2001); Özesmi & Mitsch (1997); Pearce

et al. (1994); Peeters & Gardeniers (1998); Schadt et al. (2002); Scholten et al. (2003); Lehrbücher zum Thema sind u.a. Agresti (1996); Harrell (2001); Hosmer & Lemeshow (2000).

Die Modellierung mit logistischer Regression erfolgt unter der Annahme, dass die Responsevariable bernoulliverteilt (binomialverteilt mit Binomialkoeffizient 1) ist. Dafür wird ein spezielles GLM verwendet, das für einen binomialverteilten Fehler als Linkfunktion den logit verwendet. Damit ergibt sich aus Gl. 2.2 die in Gl. 2.3 aufgeführte Grundgleichung.

$$\begin{aligned} E(Y|\mathbf{x}) &= \pi(\mathbf{x}) = P(Y=1|\mathbf{x}) \\ g(\mathbf{x}) = logit(\pi(\mathbf{x})) &= \ln\left(\frac{\pi(\mathbf{x})}{1-\pi(\mathbf{x})}\right) \\ &= LP = \alpha + \sum_{j=1}^{p} \beta_j x_j \end{aligned} \tag{2.3}$$

Die sich daraus ergebende Responsekurve zeigt Gl. 2.4.

$$\pi(LP) = \frac{e^{LP}}{1+e^{LP}} = \frac{1}{1+e^{-LP}} = (1+e^{-LP})^{-1} \tag{2.4}$$

Im Falle einer linearen Funktion für den linearen Prädiktor, d.h. wenn $LP = \alpha + \beta x$, ist die Responsekurve sigmoidal (S-förmig bei positivem Regressionskoeffizienten). Für eine quadratische Funktion $LP = \alpha + \beta_1 x + \beta_2 x^2$ ist sie unimodal (sog. Gauss-Logit-Modell, s. Jongman et al. 1995). Das Gauss-Logit-Modell lässt sich auch in einer äquivalenten Form formulieren: $LP = \alpha - (x-u)^2/(2t^2)$, in der u das Optimum und t die Toleranz bezeichnet. So lassen sich beide artabhängigen Parameter aus der Responsekurve ableiten (Yee & Mitchell 1991).

Möchte man mit dem LRM Vorkommenswahrscheinlichkeiten berechnen, so müssen die Werte der erklärenden Variablen in Gl. 2.4 eingesetzt werden. Stehen in einem Geoinformationssystem (GIS) Karten der Prädiktorvariablen zur Verfügung, so kann die Berechnung für jede räumliche Einheit durchgeführt werden, was einer räumlichen Extrapolation entspricht.

Die Residuen, d.h. die Differenz zwischen beobachtetem Wert y_i und geschätzten Wert $\hat{y}_i$ können für jede Beobachtung i nur einen der zwei möglichen Werte annehmen: entweder $(1-\hat{y}_i)$, wenn $y_i = 1$, oder $-\hat{y}_i$, wenn $y_i = 0$. Sie entstammen damit also nicht - wie im Falle des linearen Modells - einer Normalverteilung.

Zur *Maximum-likelihood*-Parameterschätzung der Regressionskoeffizienten eines logistischen Regressionsmodells wird die in Gl. 2.5 aufgeführte Likelihoodfunktion aufgestellt. Dabei bezeichnet der Term $\pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$ die Wahrscheinlichkeit einer einzelnen Beobachtung i (s.o., Hosmer & Lemeshow 2000).

$$\begin{aligned} L(\hat{\beta}) &= Pr(y_1,\ldots,y_n) = \prod_{i=1}^{n} Pr(y_i) \\ &= \prod_{i=1}^{n} \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i} \end{aligned} \tag{2.5}$$

Da angenommen wird, dass alle n Beobachtungen voneinander unabhängig sind, ist die Likelihood aller Beobachtungen das Produkt über alle n Wahrscheinlichkeiten, s. Gl. 2.5. Weil das Rechnen mit Summen leichter ist als das Rechnen mit Produkten, wird für gewöhnlich mit dem Logarithmus der Funktion, der log *Likelihood* LL (Gl. 2.6), gerechnet; dadurch werden die Exponenten Koeffizienten.

$$LL(\hat{\beta}) = \sum_{i=1}^{n} [y_i \ln(\pi(x_i)) + (1-y_i)\ln(1-\pi(x_i))] \tag{2.6}$$

Aus dieser Funktion lässt sich mit der residualen Devianz $D = -2LL$ eine Teststatistik herleiten, die zur Beurteilung der Güte der Anpassung des spezifizierten Modells eingesetzt wird (Fielding & Bell 1997; Reineking & Schröder 2004b). Um zu testen, ob eine erklärende Variable signifikant zur Modellverbesserung beiträgt, wird das Verhältnis der Devianzen für ein Modell ohne bzw. mit Berücksichtigung dieser Variable betrachtet (*Likelihood-Ratio*-Test, Gl. 2.7, Hosmer & Lemeshow 2000). Unter der Nullhypothese („Kein Unterschied zwischen den beiden Modellen, d.h. die Devianzen sind gleich.“) ist die Teststatistik LR bei ausreichend großem Stichprobenumfang näherungsweise χ^2-verteilt mit df Freiheitsgraden (df = Differenz der Anzahl der Parameter im vollen und reduzierten Modell).

$$\begin{aligned} LR &= D(\text{Modell ohne } x_j) - D(\text{Modell mit } x_j) \\ &= -2\ln\left(\frac{\hat{L}(\text{Modell ohne } x_j)}{\hat{L}(\text{Modell mit } x_j)}\right) \end{aligned} \tag{2.7}$$

Der Quotient aus der Devianz und der Anzahl der Freiheitsgrade, also der Term D/df wird als Dispersionsparameter (oder Skalierungsparameter) bezeichnet. Er beschreibt das Verhältnis von beobachteter und erwarteter Varianz. Bei Verwendung des logistischen Regressionsmodells hat der Dispersionsparameter den Wert 1. Mittels eines F-Tests kann geprüft werden, ob die zugrunde liegende Annahme der Binomialverteilung erfüllt ist (Crawley 2002). Wenn die Unterschiede zwischen beobachteter Inzidenz und vorhergesagten Vorkommenswahrscheinlichkeiten größer sind, als auf der Basis des zugrunde liegenden Binomialmodells zu erwarten, ist der Dispersionsparameter größer als 1 und damit zusätzliche Variabilität/*overdispersion* zu beobachten. In einem solchen Fall werden die Standardfehler der Regressionskoeffizienten zu optimistisch geschätzt. *Overdispersion* kann z.B. dann vorkommen, wenn die Beobachtungen nicht voneinander unabhängig

sind, keine zufällige Stichprobe vorliegt oder die Beobachtungen nicht aus einer Binomialverteilung stammen, was einer Verletzung der Annahmen entspricht. Bei moderater *overdispersion* sollten korrigierte Standardfehler der Regressionskoeffizienten SE_{adj} berechnet werden, Gl. 2.8, (Crawley 2002).

$$SE_{adj}(\hat{\beta}_j) = SE(\hat{\beta}_j)\sqrt{\frac{D}{df}} \qquad (2.8)$$

Alternativ können statt der ML-Parameterschätzung *Quasi-likelihood*-Verfahren verwendet werden, die neben den Regressionskoeffizienten auch den Dispersionsparameter aus den Daten schätzen (Quinn & Keough 2002). Bei starker *Overdispersion* empfiehlt es sich, das Modell neu zu spezifizieren, da es im Extremfall seine Aussagekraft völlig verliert (Crawley 2002).

LRMs für nominale und ordinale Responsevariablen

Ist die Responsevariable nicht binär, sondern nominalskaliert - beispielsweise wenn es um das Vorkommen von Männchen und Weibchen auf eine Fläche geht - oder ordinal - z.B. bei der Modellierung von Häufigkeitsklassen - so ist das Instrumentarium der GLMs bzw. LRMs auch zur Modellierung dieser Fälle geeignet. Im ersten Fall können multinomiale Logit-Modelle geschätzt werden (Fox 2002; Trexler & Travis 1993). Beispielhafte Anwendungen multinomialer GLMs finden sich bei Ramsey & Usner (2003) sowie Augustin et al. (2001). Bei ordinalen Responsevariablen bieten sich zwei unterschiedliche Modelltypen an (Guisan & Zimmermann 2000). Sind die Klassen durch Kategorisierung einer kontinuierlichen Responsevariablen entstanden, so ist die Verwendung des sog. *proportional-odds*-logistischen Regressionsmodells (Fox 2002) angeraten. Wenn dagegen die Klassen auf einer diskreten, geordneten Responsevariablen, etwa bei Entwicklungsstadien, beruhen, so empfiehlt es sich, das sog. *continuation-ratio*-Modell anzuwenden (Bender & Benner 2000; Harrell et al. 1998).

GLMs für Zählungen - Poissonregression

Werden statt Präsenz-Absenz- oder nominalskalierten Daten metrisch skalierte Abundanzdaten oder Zählungen verwendet, so können Poissonregressionen (auch *log-linear models*) durchgeführt werden, die analog zur logistischen Regression mit poissonverteiltem Fehler und log-Linkfunktion arbeiten. Sollte statt der Poisson- die neg-binomial-Verteilung adäquat sein, so ist auch hier eine Modellierung mit GLMs möglich (Pearce & Ferrier 2001; White & Bennetts 1996). Beispiele für Poissonregressionen in der ökologischen Literatur finden sich bei Laurance (1997); Lindenmayer et al. (1991); MacNally et al. (2003) sowie Vincent & Haworth (1983). Pearce & Ferrier (2001) zeigen allerdings, dass sich der Mehraufwand einer abundanzbasierten Erhebung und Modellierung in dem von ihnen untersuchten Fall nicht gelohnt hat. Dabei ist auch zu bedenken, dass die Erhebung von Abundanz im Vergleich zu Inzidenzdaten stärker fehlerbehaftet ist (Mühlenberg 1993).

GLM-Erweiterungen - Umgang mit Autokorrelation I: Autologistische GLMs

Eine wichtige Annahme, die für das Modell getroffen wird, ist die Unabhängigkeit der erhobenen Daten (Hosmer & Lemeshow 2000). Diese Annahme kann u.a. durch räumliche Autokorrelation untergraben werden (Fielding & Haworth 1995; Legendre 1993). Autokorrelation findet sich überall, wo Variablen in Zeitreihen (zeitliche Autokorrelation) bzw. entlang von Umweltgradienten (räumliche Autokorrelation) aufgenommen werden (Koenig 1999). In der Ökologie ist die räumliche Autokorrelation für kurze Distanzen zumeist positiv, d.h. an nahe benachbarten Orten gemessene oder beobachtete Variablen beeinflussen sich über den Raum; ihre Werte sind positiv korreliert (Legendre & Fortin 1989). Entlang eines Gradienten ist diese positive Autokorrelation für kurze Distanzen gekoppelt mit einer negativen Autokorrelation für große Distanzen (Legendre & Fortin 1989).

Bei positiver räumlicher Autokorrelation kann ein Wert, der an einem bestimmten Ort erhoben wurde, zu einem gewissen Anteil durch die Werte in der Nachbarschaft vorhergesagt werden (Legendre 1993). Dieser Wert ist dann aber stochastisch nicht unabhängig, was eine Verletzung der Modellannahmen bedeutet: jede Beobachtung resultiert in diesem Fall nicht im Gewinn eines weiteren ganzen Freiheitsgrades für die Modellbildung. Diese Verringerung der Freiheitsgrade macht eine „naive" Anwendung „klassischer" Testverfahren unzuverlässig (Legendre & Fortin 1989).

Eine Erweiterung der logistischen Regression, in denen die räumliche Autokorrelation der abhängigen und unabhängigen Variablen zur Verbesserung der Prognose in das Modell einbezogen wird, sind autologistische GLMs. Dem Modell wird eine weiteren Kovariate hinzugefügt, die entweder aus den tatsächlichen Vorkommen oder aus den geschätzten Vorkommenswahrscheinlichkeiten in der Nachbarschaft abgeleitet wird (Augustin et al. 1996, 1998). Der lineare Prädiktor kann dann z.B. die in Gl. 2.9 gezeigte Form annehmen.

$$LP = \alpha + \sum_{j=1}^{p} \beta_j X_j + \sum_{k \neq l} \delta_{kl} Y_l \qquad (2.9)$$

Dabei beschreibt der Parameter δ_{kl} die (wegen $\delta_{kl} = \delta_{lk}$ symmetrisch angenommene) Interaktionen zwischen den räumlichen Einheiten (z.B. Rasterzellen)

k und l. Um für unbeobachtete räumliche Einheiten die Vorkommenswahrscheinlichkeit zu schätzen und damit die Autokovariate schätzen zu können, verwenden Augustin et al. (1996, 1998) den sog. *Gibbs sampler* mit der Markovketten-Monte-Carlo-Methode (*Markov Chain Monte Carlo*: MCMC) (Besag & Creen 1993; Khaemba & Stein 2001). Beispielhafte Anwendungen mit z.T. weniger aufwendigen, heuristischen Verfahren finden sich bei Lichstein et al. (2002); Osborne & Alonso (2000); Osborne et al. (2001); Schröder (2000); Smith (1994) sowie Wu & Huffer (1997).

GLM-Erweiterungen - Umgang mit (Auto-)Korrelation II: *Generalised Estimating Equations*: GEEs

Die im Zuge der Diskussion zur *overdispersion* erwähnten *quasi-likelihood*-Verfahren sind die Grundlage der sog. *Generalised Estimating Equation regression models* (GEE), die explizit als robuste Methode zur Modellierung korrelierter Kovariaten entwickelt wurden (Quinn & Keough 2002). Die meisten Beispiele in der Literatur beziehen sich auf die Analyse von wiederholten Messungen bzw. Zeitreihen (Horton & Lipsitz 1999), doch bieten sich GEEs auch für die Modellierung räumlich autokorrelierter Daten an (Gotway & Stroup 1997). Ein ausgezeichnetes Beispiel für die Verwendung von GEEs in der Habitatmodellierung findet sich bei Gumpertz et al. (2000). Eine Anwendung eines verallgemeinerten gemischten linearen Modells (GLMM) in der Habitatmodellierung, das sich ebenfalls zur Modellierung korrelierter Kovariaten eignet zeigen Milsom et al. (2000).

GLM-Erweiterungen - Flexibilisierung I: Verallgemeinerte additive Modelle (*Generalised Additive Models*: GAM)

In einigen Fällen ist das parametrische Verfahren der logistischen Regression nicht flexibel genug, um die Form der Responsekurve abbilden zu können, z.B. im Fall asymmetrischer unimodaler oder bimodaler Responsekurven, oder man daran interessiert ist, welche Form der Responsekurve die Daten besonders gut beschreibt (Lehmann et al. 2002b; Yee & Mitchell 1991). Verallgemeinerte additive Modelle (*generalised additive models*: GAMs) ermöglichen eine flexiblere, daten-, nicht modellgeleitete Anpassung, indem sie neben der polynomialen auch die nicht-parametrische Modellierung der Prädiktoren zulassen (Hastie & Tibshirani 1990). Gl. 2.10 zeigt die allgemeine Form der GAMs.

$$g(\pi) = \alpha + \sum_{j=1}^{p} f_i(x_j) \qquad (2.10)$$

Dabei sind die f_j, $j = 1, \ldots p$, nicht-parametrische Glättungsfunktionen (*smooth functions* z.B. *smoothing splines*). Sie werden aus den Daten abgeleitet, können je nach Verfahren eine unterschiedliche Anzahl von Freiheitsgraden verbrauchen und sind allein an die Bedingung gebunden, um 0 symmetrisch zu sein ($E(f_j(x)) = 0$). Wenn für alle Prädiktoren $f(x) = \beta x$ gilt, dann entspricht das GAM einem „klassischen" GLM. Insofern stellt also das GLM einen Spezialfall des GAM dar, so wie die multiple lineare Regression ein Spezialfall des GLM ist. Im Zuge einer Modellierung mit GAMs kann getestet werden, ob die Glättungsfunktionen durch parametrische, lineare Funktionen und damit das GAM durch ein - wenn immer möglich zu bevorzugendes, weil einfacheres - GLM ersetzt werden kann (Yee & Mitchell 1991). Während GAMs aufgrund der größeren Flexibilität häufig bessere Prognosen liefern als GLMs, haben diese den Vorteil der besseren Interpretierbarkeit und Vergleichbarkeit der Modelle z.B. anhand der Optimums- und Toleranzparameter. Beispielhafte Anwendungen von GAMs in der Habitatmodellierung finden sich bei Austin (2002); Leathwick et al. (1996); Lehmann et al. (2002a,b) sowie Pearce & Ferrier (2000).

GLM-Erweiterungen - Flexibilisierung II: Künstliche neuronale Netzwerke (*Artificial neural networks*: ANN)

Ein künstliches neuronales Netzwerk ist ein sog. *black box*-Verfahren, das sich sehr gut zur prädiktiven Modellierung eignet (Lek & Guégan 1999). Die im Zuge der Habitatmodellierung am häufigsten verwendeten künstlichen neuronalen Netzwerke sind sogenannte *backpropagation networks* (BPN: Rumelhart et al. 1986). Ein zweiter, in der ökologischen Modellierung zum Einsatz kommender Typ von neuronalen Netzen sind selbst organisierende Kohonenkarten (Céréghino et al. 2001; Foody 1999), auf die hier aber nicht näher eingegangen wird. Die einfachste Form eines BPN ist ein *one-layer feed forward neural network*, in dem die „Neuronen" (Knoten, *nodes*, *perceptrons*) in drei aufeinander folgenden Lagen - einem Input-, einem versteckten/*hidden* und einem Outputlayer - angeordnet sind (Abb. 2.5). Im Fall der Habitatmodellierung besitzt der Outputlayer nur einen einzigen Knoten (Özesmi & Özesmi 1999).

Der Informationsfluss verläuft nur in einer Richtung, vom Inputlayer, in dem die erklärenden Variablen eingelesen werden, zum Outputlayer, der in der Trainingsphase die Werte der Responsevariable erhält, bei der Modellanwendung hingegen die prognostizierten Vorkommenswahrscheinlichkeiten ausgibt. Die Knoten eines jeden Layers sind mit allen Knoten des dahinter liegenden Layers verknüpft, nicht aber lateral innerhalb eines Layers. Jede Verknüpfung bekommt ein Gewicht ω zugeordnet, vom dem die eingehende Information abhängt. Diese wird durch jeden Knoten durch eine sigmoidale Transferfunktion ϕ, Gl. 2.11 unten, verarbeitet,

die ihrerseits einen zwischen 0 und 1 liegenden Output generiert (Manel et al. 1999a).

Das ANN-Modell lernt anhand eines Trainingsdatensatzes, der aus Prädiktorvariablen und Responsevariable besteht, indem es iterativ die Fehler der eigenen Prognosen durch Veränderung der Gewichte minimiert. Abb. 2.5 (verändert nach Goodman 1996) veranschaulicht diesen Prozess, indem sie den Informationsfluss sowie den konzeptuellen Fluss des Fehlergradienten für ein logistisches Regressionsmodell (Abb. 2.5 oben) sowie ein einfaches neuronales Netz (Abb. 2.5 unten) als gerichteten Graphen darstellt. Die Abbildung veranschaulicht auch die Flexibilität von ANNs, die als Verallgemeinerung von Regressionsfunktionen interpretiert werden können (s. Gl. 2.11, Ripley 1996).

$$Y = \phi_o \left(\alpha + \sum_h \omega_h \phi_h \left(\alpha_h + \sum_i \omega_i X_i \right) \right)$$

$$\phi(z) = \frac{e^z}{1+e^z} \quad (2.11)$$

Hierbei bezeichnen wie in Abb. 2.5 die Indizes *i*, *h* und *o* die Elemente des Input-, Hidden bzw. Outputlayers, α eine Konstante, ω die Gewichte jeder Verknüpfung zwischen den Neuronen und ϕ eine sigmoidale Transferfunktion, die der Linkfunktion bei der logistischen Regression entspricht.

Goodman & Harrell (1999) stellen ANNs als eine serielle Verknüpfung von logistischen Regressionsmodellen dar, in denen die Werte der erklärenden Variablen am Inputlayer simultan an einige parallele GLMs übermittelt werden.

Wird ein ANN zu intensiv trainiert, so besteht die große Gefahr des *overtrainings* bzw. *overfittings*, d.h. dass neben den realen Strukturen in den Daten auch reines Umweltrauschen modelliert wird. Deshalb spielt bei dem Anpassen von ANNs die Regularisierung, d.h. die Einschränkung der Modellkomplexität, eine große Rolle (Ripley 1996; Reineking & Schröder 2004b). Zudem wird bei der Modellierung mit neuronalen Netzen neben dem repräsentativen Trainingsdatensatz auch ein möglichst unabhängiger Testdatensatz zur Abschätzung der Prognosegüte benötigt. Ist die Verwendung zweier unabhängiger Datensätze nicht möglich, so kann der Datensatz auch durch Kreuzvalidierung oder andere Resamplingverfahren (Schröder & Reineking 2004b; Verbyla & Litvaitis 1989) aufgeteilt werden.

Instruktive Beispiele der Anwendung von ANNs in der Habitatmodellierung oder verwandten Bereichen finden sich bei Bradshaw et al. (2002); Cairns (2001); Gevrey et al. (2003); Hilbert & Ostendorf (2001); Hoang et al. (2001); Mastrorillo et al. (1997); Olden (2003); Özesmi & Özesmi (1999); Recknagel (2000) und Reyjol et al. (2001). Vergleiche mit statistischen Verfahren wie Diskriminanzanalyse (s.u.) oder GLMs zeigen Manel et al. (1999a,b); Moisen & Frescino (2002) sowie Olden & Jackson (2002).

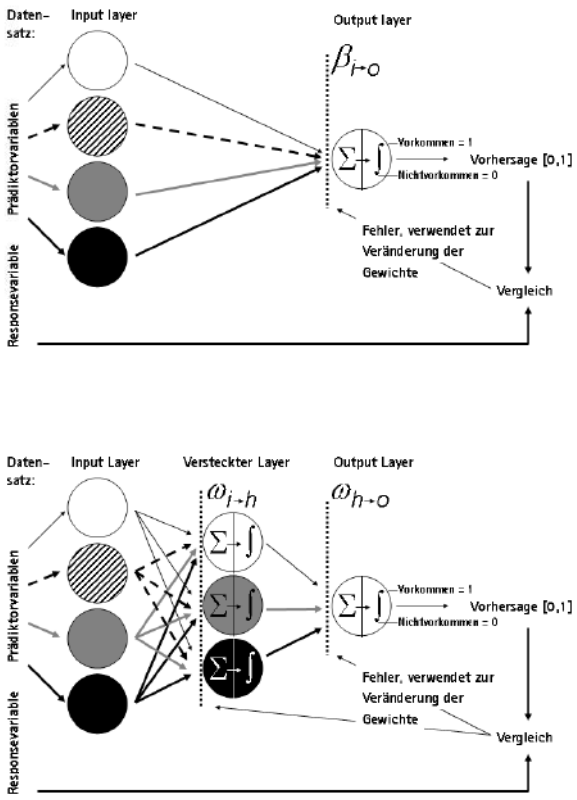


Abb. 2.5. Repräsentation eines LRMs (oben) und eines *One-hidden-layer feed forward* ANNs (unten) als gerichteter Graph (verändert nach Goodman 1996): der Informationsfluss erfolgt von links nach rechts, der konzeptuelle Fluss des Fehlergradienten von rechts nach links.

2.2.3 Klassifikationsverfahren

Neben den regressionsanalytischen Verfahren finden in der statistischen Habitatmodellierung auch Klassifikationsverfahren Verwendung, die mit zunehmender Flexibilisierung aber auch als erweiterte Regressionsverfahren betrachtet werden können (s.u.).

Diskriminanzanalyse (*Discriminant function analysis*:DFA)

Die Diskriminanzanalyse (nach Fisher 1936) ist ein parametrisches Verfahren, mit dem analysiert werden kann, ob sich Gruppen - z.B. besiedelte und unbesiedelte Patches - hinsichtlich ihrer Umwelteigenschaften signifikant voneinander unterscheiden bzw. welche Umweltvariablen zur Unterscheidung der Gruppen geeignet sind (Kleyer et al. 2000). Hierbei müssen

die untersuchten Merkmalsvariablen metrisch skaliert sein (Ausnahme: Diskriminanzanalyse mit Multinomialregel nach Deichsel & Trampisch 1985), die Gruppierungsvariable nominal (Backhaus et al. 2000). Morrison et al. (1998) konstatieren eine Ablösung der Diskriminanzanalyse durch die logistische Regression (vgl. Abb. 2.3). Beispielhafte Anwendungen finden sich bei Clarke et al. (2003); Corsi et al. (1999); Fielding & Haworth (1995); Green (1971) und Tappeiner et al. (1998).

Klassifikations- und Regressionsbäume (*Classification and regression trees*: CART)

In Fällen, in denen die Vielzahl verfügbarer Prädiktorvariablen eine intensive Analyse erschwert und in denen keine überprüfbaren Hypothesen zur Art-Habitat-Beziehung vorliegen, werden mitunter Verfahren verwendet, die der Suche nach Zusammenhängen mit dem Ziel der Klassifikation dienen und damit eher aus dem Bereich des Datamining stammen. Hierzu gehören neben regelbasierten Klassifikationsverfahren die nicht-parametrischen Klassifikations- und Regressionsbäume (auch *recursive partitioning regression*, Breiman & Friedman 1984; Venables & Ripley 1997). Nach Crawley (2002) eignen sich diese als einfach bewertete Verfahren sehr gut zur Dateninspektion, da sie einen guten Überblick über die den Daten inhärenten Strukturen vermitteln.

CARTs unterteilen den durch die p Prädiktorvariablen aufgespannten (p-dimensionalen) Raum in Regionen, in denen die Responsevariable annähernd konstante Werte annimmt. Diese Konstante wird als Wert der Responsevariablen für die jeweilige Region - hier als Vorkommenswahrscheinlichkeit - geschätzt. Durch binäre rekursive Partitionierung werden die Daten sukzessive entlang der Achsen der Prädiktorvariablen an den Stellen - Knoten (*nodes*) - aufgeteilt, die einen maximalen Unterschied der Responsevariablen auf dem entstehenden linken und rechten Ast (*branch*) erzeugen (Crawley 2002). Dies wird solange wiederholt, bis alle Responsewerte eines Knoten identisch sind oder die Datenmenge für eine weitere Aufteilung zu gering ist (terminale Knoten). Auch andere Regeln der Teilung wie beispielsweise die Maximierung der Reduktion der Abweichungsquadrate sind möglich (Moisen & Frescino 2002). An allen terminalen Knoten werden dann die durchschnittlichen (bei kontinuierlichen Variablen) bzw. häufigsten Werte (bei kategorialen Variablen) der Responsevariablen als vorhergesagte Werte bestimmt. Das entstehende Modell ist ein Klassifikationsbaum, wenn die Responsevariable diskret ist und ein Regressionsbaum im Fall einer stetigen Responsevariable.

Um eine zu starke Anpassung des Modells (*overfitting*) zu vermeiden, muss der Baum analog zur Variablenselektion (Reineking und Schröder 2004b, dieser Band) in der Regressionsanalyse „gestutzt" werden (sog. pruning). Häufig geschieht dies durch Kreuzvalidierung (Moisen & Frescino 2002). CARTs zeichnen sich dadurch aus, dass sie auch nichtlineare, nichtadditive und hierarchische Beziehungen zwischen den Prädiktorvariablen berücksichtigen können (Miller & Franklin 2002). Instruktive Beispiele finden sich bei Bell (1996); Cairns (2001); De'ath & Fabricius (2000); De'ath (2002); Franklin et al. (2000); Miller & Franklin (2002); Moisen & Frescino (2002) und Vayssières et al. (2000).

Multivariate adaptive regression splines: MARS

Eine Erweiterung der CARTs sind multivariate adaptive Regressionssplines (MARS, Friedman 1991). Sie stellen ein flexibles, nicht-parametrisches Regressionsverfahren dar, in dem anstelle der stückweise konstanten Funktionen der CARTs multivariate *splines* in den einzelnen Regionen angepasst werden. Durch Anpassung der entsprechenden Funktionswerte an den Grenzen der Regionen ergeben sich dann kontinuierliche Funktionen. Vergleiche zwischen Regressionsbäumen und MARS-Modellen im Habitatmodellkontext finden sich bei Moisen & Frescino (2002) sowie Prasad & Iverson (2000).

Regelbasierte Verfahren basierend auf genetischen Algorithmen (*Genetic Algorithm Rule-set Prediction*: GARP)

GARP, genetische Algorithmen zur Ableitung von Regelsätzen zur Prognose der räumlichen Verteilung von Organismen, ist ein Expertensystemansatz, bei dem Verfahren des künstliches Lernens (*machine learning*) verwendet werden (Stockwell 1992; Stockwell & Peters 1999). Diese Methoden der künstlichen Intelligenz umfassen Entscheidungsbäume, neuronale Netze und genetische Algorithmen. Letztere werden in GARP zur Ableitung der Regeln, d.h. „Wenn-dann" Beziehungen, unter gleichzeitiger Maximierung der Signifikanz und des Prognoseerfolgs verwendet. Simultan werden von diesem System verschiedene Modelle bzw. Regeltypen generiert und getestet. Dazu gehören i) Regeln, welche die gesamten klimatischen Bedingungen umfassen, unter denen eine Art existieren kann (*envelope rules*), ii) sog. *GARP-rules*, die sich von den vorgenannten dahingehend unterscheiden, dass einzelne Variablen nicht berücksichtigt werden, iii) Regeln, die sich auf einzelne Kategorien oder Werte einzelner Variablen beziehen (*atomic rules*) und iv) Regeln, die logistischen Regressionsmodellen entsprechen, in denen für bestimmte Werte der im linearen Prädiktor enthaltenen Variablen Vorkommen oder Nichtvorkommen vorhergesagt werden (*logit rules*). Die Regeln werden durch iterative und schrittweise Verbesserung durch den genetischen

Algorithmus spezifiziert, der nach evolutionären Prinzipien - Mutation, *crossing-over*, Reproduktion, Selektion - arbeitet und stochastische Elemente aufweist. Initiale Modelle bzw. Regelsätze, von denen der iterative Prozess der Regelableitung ausgeht, werden durch parametrische statistische Verfahren wie beispielsweise logistische Regression erstellt. Ausgehend von dieser Start-„population" wird dann für einen zufällig gewählten Trainingsdatensatz eine Bewertung hinsichtlich Prognosegüte und Signifikanz vorgenommen. Zufällig, aber proportional zu ihrer relativen Güte werden aus diesen Regeln die neuen Regeln der nächsten Generation entnommen und stochastischen Variationen unterworfen. Bei Mutationen werden nur kleine Änderungen der Regeln durchgeführt, während beim *crossing-over* ganze Regelelemente zwischen Regeln ausgetauscht werden (Stockwell & Peters 1999). *Overfitting* wird dabei durch *data-splitting* oder *resampling* verhindert (vgl. Schröder und Reineking 2004, in diesem Band).

GARP produziert aufgrund der stochastischen Elemente des Algorithmus keine eindeutigen Lösungen (Anderson 2003), d.h. mit jedem Modelllauf erhält man Vorhersagen, die sich leicht voneinander unterscheiden. Beispiele der Anwendung von GARP - z.T. unter alleiniger Verwendung von Präsenzdaten - finden sich bei Anderson et al. (2002); Anderson (2003); Anderson & Martinez-Meyer (2004); Lim et al. (2002); Peterson et al. (2002) und Rojas-Soto et al. (2003).

2.2.4 Weitere Verfahren

Zur Erläuterung weiterer Verfahren, wie der sog. *environmental niche factor analysis* (ENFA Hirzel & Guisan 2002; Hirzel et al. 2001; Reutter et al. 2003; Zaniewski et al. 2002), sog. *climate envelopes* (Davis et al. 1998; Pearson & Dawson 2003), regelbasierten Habitatmodellen auf der Basis der Fuzzy Logik (Schröder 1997) sowie Verfahren aus der Bayes-Statistik (Aspinall 1992; Aspinall & Veitch 1993; Fleishman et al. 2003, 2001; Högemander & Møller 1995; Hooten et al. 2003; MacNally et al. 2003; Ter Braak et al. 2003) sei an dieser Stelle auf die angeführte Literatur verwiesen.

2.3 Auswahl eines geeigneten Verfahrens

Die Wahl des geeigneten statistischen Verfahrens und der adäquaten Modellierungsstrategie muss immer angesichts der Fragestellung bzw. des angestrebten Outputs erfolgen. Abb. 2.6 verdeutlicht plakativ das Verhältnis von Responsekurven, welche mit Hilfe eines parametrischen GLMs, eines semi-parametrischen GAMs und eines regelbasierten ANNs erhalten werden. Die datengeleiteten Verfahren erlauben viel flexiblere Anpassungen an die Daten und damit eine größere Formvielfalt der Responsekurven. Dadurch sind sie in der Prognosegüte den parametrischen GLMs zumeist überlegen (Manel et al. 1999a; Moisen & Frescino 2002; Olden & Jackson 2002). Andererseits wird häufig die bessere Interpretierbarkeit und höhere Vergleichbarkeit der parametrischen Modelle hervorgehoben, die auch durchweg dem Prinzip der geringstmöglichen Modellkomplexität/*principle of parsimony* eher gerecht werden (Austin 2002; Özesmi & Özesmi 1999). Letztlich geht es also um zwei zusammenhängende Zielkonflikte (s. Abb. 2.2) und darum, Kompromisse hinsichtlich des *bias-variance trade-offs* sowie hinsichtlich Prognoseerfolg vs. Erklärungsgehalt zu treffen.

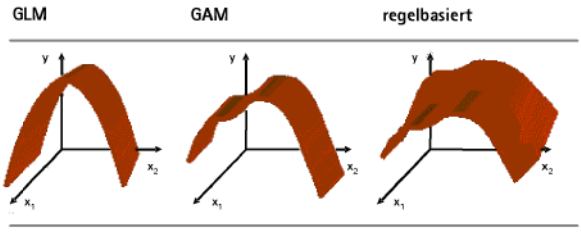


Abb. 2.6. Beispielhafte Responsekurven der Vorkommenswahrscheinlichkeit y in Abhängigkeit der erklärenden Variablen x_1 und x_2, erzeugt mittels eines GLM (modellgeleitete parametrische Statistik), eines GAM (datengeleitete semi-parametrische Statistik) sowie eines datengeleiteten regelbasierten Ansatzes. Von links nach rechts nimmt die Vorhersagegüte zu aber der Erklärungsgehalt zur Art-Habitat-Beziehung ab (verändert nach Lehmann in http://www.cscf.ch/grasp/grasp-s/show/grasp_files/frame.htm).

2.4 Modellierungsstrategie

Die Strategie, die im Zuge der Modellbildung und -evaluation verfolgt wird, hängt davon ab, welchen Zweck das Habitatmodell erfüllen soll. Einige grundsätzliche Hinweise sollten in jedem Fall berücksichtigt werden. Datenerhebung ist meistens teurer als Datenanalyse; deshalb sollten möglichst effiziente und genaue Modellierungsmethoden verwendet werden. Prädiktive Modellierung mit dem Ziel „Prognose, Vorhersage" und die genaue Bestimmung der Effektgrößen mit dem Ziel „Erklärung" sowie das Testen einzelner Hypothesen gehen Hand in Hand (Harrell 2001). Die Unzuverlässigkeit der geschätzten Koeffizienten entspricht dem Ausmaß, in dem das Modell falsche Vorhersagen für einen unabhängigen Datensatz liefert. Dennoch bekommen je nach Verwendungszweck des Modells einzelne Aspekte des Modellierungsprozesses unterschiedliches Gewicht. Wir folgen hier Harrell (2001) sowie Hosmer & Le-

meshow (2000), deren wichtigste Aussagen zu diesen Punkten der Abwägung wir hier zusammenfassen.

2.4.1 Modellierungsstrategie für ein Prognosemodell

Ein Modell, das gute Prognosen liefern soll, sollte auf der Grundlage möglichst vieler Daten hoher Qualität erfolgen, die lange Gradienten abdecken. Eine intensive Dateninspektion mit Hilfe grafischer und deskriptiver Verfahren erleichtert dann die Formulierung guter Hypothesen zur Spezifikation der relevanten Prädiktorvariablen und möglicher Interaktionen. (Hosmer & Lemeshow 2000) schlagen in diesem Zusammenhang vor, für alle Prädiktorvariablen univariate LRMs zu schätzen, und nur Variablen für die anschließende Modellselektion zu berücksichtigen, die im univariaten Fall mindestens auf dem Niveau $p < 0,25$ signifikant sind.

Dabei sollte man sich auch Gedanken darüber machen, welchen Grad an Komplexität bzw. Nichtlinearität für die einzelnen Prädiktorvariablen erlaubt, also wie viele Freiheitsgrade man ihnen „zugesteht". Wenn die Anzahl zu schätzender Parameter im Verhältnis zum Stichprobenumfang groß ist, können Techniken der Datenreduktion verwendet werden (Harrell 2001). Eine adäquate Anzahl erklärender Variablen wird z.B. bei Steyerberg et al. (2001b) für einen ausgeglichen Datensatz (Prävalenz, d.h. Anteil der Vorkommen ungefähr 50%) mit ca. 10 Präsenzen pro Variable angegeben. Das entspricht der $p/10$ Faustregel mit $p = \min[\sum(\text{Präsenzen}), \sum(\text{Absenzen})]$ (vgl. Guisan & Zimmermann 2000).

Stärkere bivariate Korrelationen zwischen einzelnen Prädiktorvariablen sollten nicht vorhanden sein, da Multikollinearität dazu führt, dass die Standardfehler des geschätzten Modells nicht mehr korrekt bestimmt werden, die darauf beruhenden Tests nicht mehr aussagekräftig sind und z.B. die Variablenselektion unzuverlässig wird (Harrell 2001). Fielding & Haworth (1995) empfehlen, bei Korrelationen, geschätzt mit dem Spearman-Rangkorrelationskoeffizienten ρ_S, von 0,7 nur eine der korrelierten Prädiktorvariablen für die Modellbildung zu berücksichtigen. Um diesen Informationsverlust zu vermeiden, können alternativ auch Verfahren der Variablenaggregation, wie z.B. die Hauptkomponentenanalyse (*principal components analysis*: PCA) eingesetzt werden und (Li et al. 1997; Quinn & Keough 2002). Die PCA aggregiert die Prädiktoren zu Linearkombinationen, die dann linear unabhängig sind. Allerdings geht dies häufig auf Kosten der Interpretierbarkeit.

Wenn möglich, sollte der gesamte Datensatz zur Modellbildung hinzugezogen werden und die Trennung in Trainings- und Testdatensätze mittels Resamplingverfahren erfolgen - *Bootstrapping* oder Kreuzvalidierung, (s. Schröder & Reineking 2004b, dieser Band). Nur wenn die folgenden Schritte aufgrund der Komplexität der Analyse oder fehlender Operationalisierbarkeit nicht für jede Resampling-Stichprobe wiederholt werden können, sollte ein Testdatensatz vor der Analyse zurückgehalten werden. Ökologische Datensätze sind zu wertvoll, als dass sie verschwendet werden dürften (Harrell 2001). Wichtig ist dann die Überprüfung der verschiedenen zugrunde liegenden Annahmen:

- Linearität der Beziehung zwischen Prädiktorvariablen und logit der Responsevariablen; u.U. müssen einige Prädiktorvariablen transformiert werden. Um beispielsweise unimodale Responsekurven zu erzeugen, müssen auch quadratische Terme der Prädiktorvariablen ins Modell gelangen.
- Additivität der Prädiktorvariablen: wenn sich der Einfluss von einzelnen Prädiktoren auf die abhängige Variable in Abhängigkeit anderer Prädiktoren verändert, dann sollten Interaktionsterme berücksichtigt werden, die als Produkte der in Frage kommenden Variablen berechnet werden.
- Einzelne Beobachtungen können einen besonders starken Einfluss auf das Modell haben (*overly influential observations*). Wie für die lineare Analyse steht auch für GLMs/LRMs eine ganze Reihe von regressionsdiagnostischen Verfahren zur Verfügung (Landwehr et al. 1984; Pregibon 1981). Mit ihrer Hilfe können i)untypische Beobachtungen (durch Residuenplots) und ii) Beobachtungen mit besonders starkem Einfluss auf die Modellbildung (durch Berechnung von Hebelwerten (*leverages*) und Maßzahlen für den Einfluss wie z.B. Cook's Distanz) gefunden werden (vgl. umfangreiche Darstellungen bei Fox 2002; Nicholls 1989).
- Verteilungsannahmen können durch Testen des Dispersionsparameters (s.o.) überprüft werden. Falls notwendig sollte ein anderes Modell gewählt werden. Unter Umständen bietet sich aber auch die Verwendung verzerrter Schätzverfahren, wie der Quasilikelihood-Schätzung an. Hierbei ist anzumerken, dass das vorsichtige Anpassen eines nicht ganz adäquaten Modells weniger „gefährlich" ist, als die schlechte Anpassung und das Overfitting eines adäquaten Modells (Harrell 2001).

Um die Modellkomplexität zu verringern, sollte eine rückwärts schrittweise Variablenselektion durchgeführt werden (Harrell 2001; Reineking & Schröder 2004b). Abhängig vom Datensatz kann dies zur Verringerung aber auch zur Verbesserung der Prognosegüte führen (vgl. unterschiedliche Ergebnisse in Reineking & Schröder 2003; Steyerberg et al. 1999). Auch die Variablenselektion kann durch Bootstrapping unterstützt werden (s. z.B. Wisnowski et al. 2003).

Ergebnis dieses Schrittes ist dann das „finale" Modell. Dies sollte in zweierlei Hinsicht analysiert werden:

- Grafische Interpretation der geschätzten Responsekurven, am besten durch dreidimensionale Responsekurven für je zwei Prädiktorvariablen gleichzeitig (vgl. Rudner et al. 2004, dieser Band).
- Überprüfung der vorhergesagten Werte. Im Falle räumlicher Extrapolation bedeutet dieser Punkt auch, nach räumlichen Mustern in den Residuen zu schauen bzw. zu testen, ob räumliche Autokorrelation der Residuen vorliegt. Nach Austin (2002) sollte diese Überprüfung zum Standardrepertoire in der Habitatmodellierung gehören. Räumliche Autokorrelation kann durch Maßzahlen wie Moran's I quantifiziert werden (Anselin 1993; Cliff & Ord 1981). Um den Einflussbereich der räumlichen Autokorrelation zu quantifizieren, können experimentelle Variogramme für die Residuen erstellt werden (Goovaerts 1998; Wallace et al. 2000). Bio et al. (2002) führen dies explizit für Habitatmodelle durch.

Im Anschluss an diese Überprüfungen nimmt die interne Validierung des finalen Modells hinsichtlich seiner Kalibrierung und Diskriminierung einen wichtigen Platz ein (Reineking & Schröder 2003; Schröder & Reineking 2004b). Sie erlaubt eine adäquate, nicht-optimistische Einschätzung der Anpassungs- und Prognosegüte. Mittels externer Validierung durch Tests auf Übertragbarkeit des Modells, d.h. Überprüfung des Modells anhand von Daten aus anderen Untersuchungsgebieten und/oder Untersuchungszeiträumen kann zudem der Geltungsbereich eines Modells abgeschätzt werden (Beispiele bei Bonn & Schröder 2001; Dennis & Eales 1999; Freeman et al. 1997; Glozier et al. 1997; Lamouroux et al. 1999; Schröder 2000; Schröder & Richter 2000; Thomas & Bovee 1993). Dies ist von großer Bedeutung, wenn das Modell zur Erstellung von Prognosen in anderen Gebieten eingesetzt werden soll.

Sollte im Zuge der internen Validierung Überanpassung festgestellt werden, so empfiehlt Harrell (2001), die Schätzwerte der Parameter durch Anwendung von *Shrinkage*-Verfahren zu verkleinern. Beispiele dafür finden sich bei Reineking & Schröder (2003) sowie - allerdings nicht aus dem Kontext der Habitatmodellierung - bei Steyerberg et al. (2001a) und Tibshirani (1995).

Modellierungsstrategie für ein Modell zur Abschätzung der Effektgrößen oder zum Hypothesentest

Geht es bei der Habitatmodellierung eher um die Schätzung der Effektgrößen oder um das Testen einzelner spezifizierter Hypothesen, so ist nach Harrell (2001) das Verfolgen der geringstmöglichen Modellkomplexität/*principle of parsimony* von nicht so großer Bedeutung wie bei der Erstellung prädiktiver Modelle. Vielmehr ist hier besonderer Wert auf die zugelassene Komplexität der interessierenden Prädiktoren und eventuelle Interaktionsterme zu legen. Der Kompromiss bezüglich des *Bias-variance trade-offs* schwenkt in diesem Fall also von der leichten Bevorzugung des *bias* bei prädiktiver Modellierung auf die Seite der *variance*. Die interne Validierung ist vor allem hinsichtlich der Quantifizierung der Überanpassung relevant, während die externe Validierung darauf zielt, zu analysieren, ob in unterschiedlichen Gebieten dieselben Umwelteigenschaften die Verteilung der Organismen erklären oder ob beispielsweise lokale Adaptationen oder saisonale Unterschiede festzustellen sind.

2.5 Zusammenfassung - *state-of-the-art* in der Habitatmodellierung

Zusammenfassend heben Lindenmayer et al. (1999) hervor, dass der Prozess der Modellformulierung ein langwieriges Unterfangen iterativer Modellschätzungen ist. Sie betonen zudem, dass die abschließende Modellauswahl ein großes Maß an Erfahrung, Verständnis der zugrunde liegenden Theorie und die ökologische und empirische Rechtfertigung der ausgewählten Variablen voraussetzt. Dieser Überblick und die weiteren Beiträge im Block „Statistische Habitatmodelle - Status quo & aktuelle Entwicklungen" sollen dazu eine Hilfestellung leisten.

Der tabellarische Überblick (Tab. 2.1) soll die „Minimalanforderungen" an eine gute Habitatmodellierung als auch einige zusätzliche Aspekte des *state-of-the-art* zusammenstellen (vgl. auch Reineking & Schröder 2004b,a; Schröder & Reineking 2004b, in diesem Band):

2.6 Danksagung

Die Autoren bedanken sich bei Hans-Peter Bäumer und Michael Rudner, beide Universität Oldenburg, für hilfreiche Kommentare und Diskussionen zum Manuskript.

Literaturverzeichnis

Adler, G. & Wilson, M. 1985. Small mammals on massachusetts islands: the use of probability functions in clarifying biogeographic relationships. *Oecologia*, 66:178–186.

Aebischer, N. J., Robertson, P. A. & Kenward, R. E. 1993. Compositional analysis of habitat use from animal radio-tracking data. *Ecology*, 74(5):1313–1325.

Agresti, A. 1996. *An Introduction to Categorical Data Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, New York.

Tabelle 2.1. *State-of-the-art* und Empfehlungen für Minimalanforderungen an kommunizierbare Habitatmodelle.

Modelldarstellung	Regressionskoeffizienten mit Standardfehler und Signifikanz nach LR-Test sowie Gütekriterien s.u.
Gütekriterien	
Diskriminierung	schwellenwertunabhängig: AUC ± CI
schwellenwertabhängig:	Cohen's κ
Kalibrierung & Refinement	R2 nach Nagelkerke, Interzept und Steigung der Kalibrierungsgerade
Alle Aspekte	Visualisierung mittels Attributgrafik
Modellbildung	*keep it simple*!
Komplexitätsbegrenzung	*Penalized maximum likelihood* und Variablenselektion oder Lasso (vgl. Reineking & Schröder 2004b, in diesem Band)
Variablenselektion	sinnvolle, hypothesengesteuerte Vorauswahl, rückwärts schrittweise unter Verwendung von Informationsmaßen (AIC) oder Lasso (vgl. Reineking & Schröder 2004b, in diesem Band)
Modellevaluation	
Regressionsdiagnostik	Tests auf untypische oder besonders einflussreiche Beobachtungen (*overly influential observations*) und Verletzung der Modellannahmen
Visualisierung	Visualisierung der Responsekurven als große Hilfe bei Plausibilitätscheck und Interpretation
interne Validierung	durch Resampling, d.h. *bootstrapping* oder zehnfache Kreuzvalidierung
externe Validierung	bei gleicher Ausgangssituation in den Datensätzen: Übertragbarkeitstest nach Schröder (2000), Angabe von Konfidenzintervallen für alle Gütekriterien
weitere Tests	Tests auf räumliche Autokorrelation der Residuen Test auf räumliche und/oder zeitliche Übertragbarkeit zur Festlegung des Gültigkeitsbereichs des Modells

Akçakaya, H. R., McCarthy, M. A. & Pearce, J. L. 1995. Linking landscape data with population viability analysis: management options for the helmeted honeyeater *Lichenostomus melanops cassidix*. *Biological Conservation*, 73:169–173.

Anderson, R. P. 2003. Real vs. artefactual absences in species distributions: tests for *Oryzomys albigularis* (Rodentia: Muridae) in Venezuela. *Journal of Biogeography*, 30(4):591–606.

Anderson, R. P. & Martinez-Meyer, E. 2004. Modeling species' geographic distributions for preliminary conservation assessments: an implementation with the spiny pocket mice (Heteromys) of Ecuador. *Biological Conservation*, 116(2):167–179.

Anderson, R. P., Peterson, A. T. & Gómez-Laverde, M. 2002. Using niche-based GIS modeling to test geographic predictions of competitive exclusion and competitive release in South American pocket mice. *Oikos*, 98(1):3–16.

Anselin, L. 1993. Discrete space autoregressive models. In Goodchild, M. F., Parks, B. O. & Steyart, L. T., editors, *Environmental Modeling with GIS*, pages 455–469. Oxford Univ. Press, New York.

Aspinall, R. 1992. An inductive modelling procedure based on bayes theorem for analysis of pattern in spatial data. *International Journal of Geographical Information Systems*, 6:105–121.

Aspinall, R. & Veitch, N. 1993. Habitat mapping from satellite imagery and wildlife survey data using a bayesian modeling procedure in a gis. *Photogrammetric engineering & remote sensing*, 59:537–549.

Augustin, N., Mugglestone, M. & Buckland, S. 1996. An autologistic model for the spatial distribution of wildlife. *Journal of applied ecology*, 33(2):339–347.

Augustin, N. H., Cummins, R. P. & French, D. D. 2001. Exploring spatial vegetation dynamics using logistic regression and a multinomial logit model. *Journal of Applied Ecology*, 38(5):991–1006.

Augustin, N. H., Mugglestone, M. A. & Buckland, S. T. 1998. The role of simulation in modelling spatially correlated data. *Environmetrics*, 9(2):175–196.

Austin, G. E., Thomas, C. J., Houston, D. C. & Thompson, D. B. 1996. Predicting the spatial distribution of buzzard *Buteo buteo* nesting areas using a GIS and remote sensing. *Journal of Applied Ecology*, 33:1541–1550.

Austin, M. P. 1985. Continuum concept, ordination methods and niche theory. *Annual review of Ecology and Systematics*, 16:39–61.

Austin, M. P. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157:101–118.

Austin, M. P., Nicholls, A. O. & Margules, C. R. 1990. Measurement of the realised qualitative niche: environmental niche of five eucalyptus species. *Ecological Monographs*, 60(2):161–178.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. 2000. *Multivariate Analysemethoden - Eine Anwendungsorientierte Einführung*. Springer, Berlin.

Balcom, B. J. & Yahner, R. H. 1996. Microhabitat and landscape characteristics associated with the threatened Allegheny woodrat. *Conservation biology*, 10(2):515–525.

Bell, J. F. 1996. Application of classification trees to the habitat preference of upland birds. *Journal of Applied Statistics*, 23(2-3):349–360.

Bender, R. & Benner, A. 2000. Calculating ordinal regression models in sas and s-plus. *Biometrical Journal*, 42(6):677–699.

Besag, J. & Creen, P. J. 1993. Spatial statistics and bayesian computation. *Journal of the Royal Statistical Society: Series B*, 55(1):25–38.

Biedermann, R. 2003. Body size and area-incidence relationships: is there a general pattern? *Global Ecology and Biogeography*, 12(5):381–388.

Bio, A. M. F., De Becker, P., De Bie, E., Huybrechts, W. & Wassen, M. 2002. Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biodiversity and Conservation*, 11(12):2189–2216.

Bonn, A. & Schröder, B. 2001. Habitat models and their transfer for single and multi species groups: a case study of carabids in an alluvial forest. *Ecography*, 24:483–496.

Bradshaw, C. J. A., Davis, L. S., Purvis, M., Zhou, Q. & Benwell, G. L. 2002. Using artificial neural networks to model the suitability of coastline for breeding by new zealand fur seals (arctocephalus forsteri). *Ecological Modelling*, 147(2):111–131.

Breiman, L. & Friedman, L. H. 1984. *Classification and Regression Trees*. Wadsworth, Belmont.

Brennan, L. A., Block, W. M. & Guttiérrez, R. 1986. The use of multivariate statistics for developing habitat suitability index models. In Verner, J., Morrison, M. L. & Ralph, C. J., editors, *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates.*, pages 177–182. University of Wisconsin Press, Madison.

Cairns, D. M. 2001. A comparison of methods for predicting vegetation type. *Plant Ecology*, 156(1):3–18.

Céréghino, R., Giraudel, J. L. & Compin, A. 2001. Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps. *Ecological Modelling*, 146(1-3):167–180.

Clarke, R. T., Wright, J. F. & Furse, M. T. 2003. RIVPACS models for predicting the expected macroinvertebrate fauna and assessing the ecological quality of rivers. *Ecological Modelling*, 160(3):219–233.

Cliff, A. D. & Ord, J. K. 1981. *Spatial processes: models and applications*. Pion Limited, London.

Collingham, Y. C., Wadsworth, R. A., Huntley, B. & Hulme, P. E. 2000. Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology*, 37:13–27.

Corsi, F., Dupre, E. & Boitani, L. 1999. A large-scale model of wolf distribution in italy for conservation planning. *Conservation Biology*, 13(1):150–159.

Crawley, M. J. 2002. *Statistical Computing: An Introduction to Data Analysis using S-Plus*. John Wiley & Sons, New York.

Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B. & Wood, S. 1998. Making mistakes when predicting shifts in species range in response to global warming. *Nature*, 391:783–786.

De Swart, E. O. A. M., Van der Valk, A. G., Koehler, K. J. & Barendegt, A. 1994. Experimental evaluation of realized niche models for predicting responses of plant species to change in environmental conditions. *Journal of Vegetation Science*, 5:541–442.

De'ath, G. 2002. Multivariate regression trees: A new technique for modeling species-environment relationships. *Ecology*, 83(4):1105–1117.

De'ath, G. & Fabricius, K. E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology*, 81(11):31783192.

Deichsel, G. & Trampisch, H. J. 1985. *Clusteranalyse und Diskriminanzanalyse*. biometrie. Gustav Fischer Verlag, Stuttgart.

Dennis, R. L. & Eales, H. T. 1999. Probability of site occupancy in the large heath butterfly *Coenonympha tullia* determined from geographical and ecological data. *Biological Conservation*, 87:295–302.

Eyre, M. D., Carr, R., McBlane, R. P. & Foster, G. N. 1992. The effects of varying side-water duration on the distribution of water beetle assemblages, adults and larvae (coleoptera: Haliplidae, dytiscidae, hydrophilidae). *Archiv für Hydrobiologie*, 124:281–291.

Fahrig, L. & Johnson, I. 1998. Effect of patch characteristics on abundance and diversity of insects in an agricultural landscape. *Ecosystems*, 1:197–205.

Fielding, A. H. & Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24:38–49.

Fielding, A. H. & Haworth, P. F. 1995. Testing the generality of bird-habitat models. *Conservation biology*, 9(6):1466–1481.

Fisher, R. A. 1936. The use of multiple measurements in taxonomic problems. *Annuals Eugenics*, 7:179–188.

Fleishman, E., MacNally, R. & Fay, J. P. 2003. Validation tests of predictive models of butterfly occurrence based on environmental variables. *Conservation biology*, 17(3):806–817.

Fleishman, E., Nally, R. M., Fay, J. P. & Murphy, D. D. 2001. Modeling and predicting species occurrence using broad-scale environmental variables: an example with butterflies of the great basin. *Conservation biology*, 15(6):1674–1685.

Foody, G. M. 1999. Applications of the self-organising feature map neural network in community data analysis. *Ecological Modelling*, 120(2-3):97–107.

Fox, J. 2002. *An R and S-Plus Companion to Applied Regression*. Sage, Thousand Oaks, 312 edition.

Franklin, J. 1995. Predictive vegetation mapping: geographic modelling of biospatial patterns in relation to environmental gradients. *Progress in physical geography*, 19(4):391–409.

Franklin, J., McCullough, P. & Gray, C. 2000. Terrain variables used for predictive mapping of vegetation communities in Southern California. In Wilson, J. P. & Gallant, J. C., editors, *Terrain analysis: principles and applications*, pages 331–354. Wiley, New York.

Freeman, M. C., Bowen, Z. H. & Crance, J. H. 1997. Transferability of habitat suitability criteria for fishes in warmwater streams. *North American Journal of Fisheries Management*, 17(1):20–31.

Friedman, J. H. 1991. Multivariate adaptive regression splines. *Annual Statistics*, 19:1–141.

Gevrey, M., Dimopoulos, I. & Lek, S. 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, 160(3):249–264.

Glozier, N. E., Culp, J. M. & Scrimgeour, G. J. 1997. Transferability of habitat suitability curves for a benthic minnow, rhinichthys cataractae. *Journal of freshwater ecology*, 12(3):379–394.

Goodman, P. 1996. Nevprop software version 3 - manual.

Goodman, P. H. & Harrell, F. E. 1999. Neural networks: advantages and limitations for biostatistical modeling. *Journal of the American Statistical Association*.

Goovaerts, P. 1998. Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties. *Biology and Fertility of Soils*, 27(4):315–334.

Gotway, C. A. & Stroup, W. W. 1997. A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 2:157–178.

Green, R. H. 1971. A multivariate statistical approach to the Hutchinsonian niche: bivalve moluscs of Central Canada. *Ecology*, 52(4):543–556.

Guisan, A., Edwards, T. C. & Hastie, T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157(2-3):89–100.

Guisan, A., Weiss, S. B. & Weiss, A. D. 1999. GLM versus CCA spatial modeling of plant species distribution. *Plant Ecology*, 143(1):107–122.

Guisan, A. & Zimmermann, N. 2000. Predictive habitat distribution models in ecology. *Ecological Modelling*, 135:147–186.

Gumpertz, M. L., Wu, C.-T. & Pye, J. M. 2000. Logistic regression for southern pine beetle outbreaks with spatial and temporal autocorrelation. *Forest Science*, 46:95–107.

Harrell, F. E., J., Margolis, P. A., Gove, S., Mason, K. E., Mulholland, E. K., Lehmann, D., Muhe, L., Gatchalian, S. & Eichenwald, H. F. 1998. Development of a clinical prediction model for an ordinal outcome: the world health organization multicentre study of clinical signs and etiological agents of pneumonia, sepsis and meningitis in young infants. *Statistics in Medicine*, 17(8):909–944.

Harrell, Frank E., J. 2001. *Regression Modeling Strategies - with Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer Series in Statistics. Springer, New York.

Hastie, T., Tibshirani, R. & Friedman, J. H. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Berlin.

Hastie, T. J. & Tibshirani, R. J. 1990. *Generalized additive models*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, Glasgow, Weinheim.

Hilbert, D. W. & Ostendorf, B. 2001. The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates. *Ecological Modelling*, 146(1-3):311–327.

Hirzel, A. & Guisan, A. 2002. Which is the optimal sampling strategy for habitat suitability modelling? *Ecological Modelling*, 157(2-3):331–341.

Hirzel, A. H., Helfer, V. & Metral, F. 2001. Assessing habitat-suitability models with a virtual species. *Ecological modelling*, 145(2-3):111–122.

Hoang, H., Recknagel, F., Marshall, J. & Choy, S. 2001. Predictive modelling of macroinvertebrate assemblages for stream habitat assessments in queensland (australia). *Ecological Modelling*, 146(1-3):195–206.

Högemander, H. & Møller, J. 1995. Estimating distribution maps from atlas data using methods of statistical image analysis. *Biometrics*, 51:393–404.

Hooten, M. B., Larsen, D. R. & Wikle, C. K. 2003. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, 18(5):487–502.

Horton, N. J. & Lipsitz, S. R. 1999. Review of software to fit generalized estimating equation regression models. *American Statistician*, 53:160–169.

Hosmer, D. W. & Lemeshow, S. 2000. *Applied Logistic Regression*. John Wiley & Sons, New York, 2nd edition.

Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposium on Quantitative Biology*, 22:415–427.

Johnson, D. H. 1980. The comparison of usage and availability measurements for evaluating resource preference. *Ecology*, 61:65–71.

Johst, K., Brandl, R. & Eber, S. 2002. Metapopulation persistence in dynamic landscapes: the role of dispersal distance. *Oikos*, 98(2):263–270.

Jongman, R. H. G., Ter Braak, C. J. F. & van Tongeren, O. F. R., editors 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press, Cambridge.

Khaemba, W. M. & Stein, A. 2001. Spatial statistics for modeling of abundance and distribution of wildlife species in the Masai Mara ecosystem, Kenya. *Environmental and Ecological Statistics*, 8(4):345–360.

Kleyer, M., Kratz, R., Lutze, G. & Schröder, B. 1999/2000. Habitatmodelle für Tierarten: Entwicklung, Methoden und Perspektiven für die Anwendung. *Zeitschrift für Ökologie und Naturschutz*, 8:177–194.

Koenig, W. D. 1999. Spatial autocorrelation of ecological phenomena. *Trends in Ecology and Evolution*, 14(1):22–26.

Kuhn, W. & Kleyer, M. 1999. A statistical habitat model for the blue winged grasshopper (*Oedipoda caerulescens*) considering the habitat connectivity. *Zeitschrift für Ökologie und Naturschutz*, 8(4):207–218.

Lamouroux, N. & Capra, H. 2002. Simple predictions of instream habitat model outputs for target fish populations. *Freshwater Biology*, 47(8):1543–1556.

Lamouroux, N., Capra, H., Pouilly, M. & Souchon, Y. 1999. Fish habitat preferences in large streams of southern France. *Freshwater Biology*, 42(4):673–687.

Landwehr, J. M., Pregibon, D. & Shoemaker, A. C. 1984. Graphical methods for assessing logistic regression models. *Journal of the American Statistical Association*, 79(385):61–71.

Laurance, W. F. 1997. A distributional survey and habitat model for the endangered northern Bettong *Bettongia tropica* in tropical queensland. *Biological Conservation*, 82:47–60.

Leathwick, J. R., Whitehead, D. & McLeod, M. 1996. Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: a modelling approach. *Environmental Software*, 11(1-3):81–90.

Legendre, P. 1993. Spatial autocorrelation: trouble of new paradigm? *Ecology*, 74:1659–1673.

Legendre, P. & Fortin, M.-J. 1989. Spatial pattern and ecological analysis. *Vegetatio*, 80:107–138.

Lehmann, A., Leathwick, J. R. & Overton, J. M. 2002a. Assessing New Zealand fern diversity from spatial predictions of species assemblages. *Biodiversity and Conservation*, 11(12):2217–2238.

Lehmann, A., Overton, J. M. & Austin, M. P. 2002b. Regression models for spatial prediction: their role for biodiversity and conservation. *Biodiversity and Conservation*, 11(12):2085–2092.

Leibold, M. A. 1995. The niche concept revisited: mechanistic models and community context. *Ecology*, 76:1371–1382.

Lek, S. & Guégan, J. F. 1999. Artificial neural networks as a tool in ecological modelling, an introduction. *Ecological Modelling*, 120(2-3):65–73.

Li, W., Wang, Z., Ma, Z. & Tang, H. 1997. A regression model for the spatial distribution of red-crown crane in Yancheng Biosphere Reserve, China. *Ecological Modelling*, 103:115–121.

Lichstein, J. W., Simons, T. R., Shriner, S. A. & Franzreb, K. E. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs*, 72:445–463.

Lim, B. K., Peterson, A. T. & Engstrom, M. D. 2002. Robustness of ecological niche modeling algorithms for mammals in guyana. *Biodiversity and Conservation*, 11(6):1237–1246.

Lindenmayer, D. B., Cunningham, R. B. & McCarthy, M. A. 1999. The conservation of arboreal marsupials in the montane ash forests of the central highlands of Victoria, South East Australia: VIII. landscape analysis of the occurence of arboreal marsupials. *Biological Conservation*, 89:83–92.

Lindenmayer, D. B., Cunningham, R. B., Tanton, M. T., Nix, H. A. & Smith, A. P. 1991. The conservation of arboreal marsupials in the montane ash forests of the central highlands of Victoria, South East Australia: III. the habitat requirements of Leadbeater's Possum *Gymnobelideus leadbeateri* and models of the diversity and abundance of arboreal marsupials. *Biological Conservation*, 56:295–315.

Mackey, B. G. & Lindenmayer, D. B. 2001. Towards a hierarchical framework for modelling the spatial distribution of animals. *Journal of Biogeography*, 28:1147–1166.

MacNally, R. 2000. Regression and model-building in conservation biology, biogeography and ecology: The distinction between and reconciliation of ‚predictive' and ‚explanatory' models. *Biodiversity and Conservation*, 9(5):655–671.

MacNally, R., Fleishman, E., Fay, J. P. & Murphy, D. D. 2003. Modelling butterfly species richness using mesoscale environmental variables: model construction and validation for mountain ranges in the great basin of western north america. *Biological Conservation*, 110(1):21–31.

Malanson, G. P., Westman, W. E. & Yan, Y.-L. 1992. Realised versus fundamental niche functions in a model of chaparral response to climatic change. *Ecological Modelling*, 64:261–277.

Manel, S., Dias, J. M., Buckton, S. T. & Ormerod, S. J. 1999a. Alternative methods for prediciting species distribution: an illustration with himalayan river birds. *Journal of Applied Ecology*, 36(5):734–747.

Manel, S., Dias, J. M. & Ormerod, S. J. 1999b. Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a himalayan river bird. *Ecological Modelling*, 120:337–348.

Manel, S., Williams, H. C. & Ormerod, S. J. 2001. Evaluating presenceabsence models in ecology: the need to account for prevalence. *Journal of Applied Ecology*, 38(5):921–931.

Manly, B. F. J., McDonald, L. L. & Thomas, D. L. 1993. *Resource Selection by Animals - Statistical Design and Analysis for Field Studies*. Chapman & Hall, London, Glasgow, New York, 1st edition.

Massolo, A. & Meriggi, A. 1998. Factors affecting habitat occupancy by wolves in northern Apennines (northern Italy): a model of habitat suitability. *Ecography*, 21(2):97–107.

Mastrorillo, S., Lek, S., Dauba, F. & Belaud, A. 1997. The use of artificial neural networks to predict the presence of small-bodied fish in a river. *Freshwater Biology*, 38(2):237–246.

McCullough, P. & Nelder, J. A. 1989. *Generalized Linear Models*. Chapman & Hall, London, 2nd edition.

Miller, J. & Franklin, J. 2002. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling*, 157(2-3):227–247.

Milsom, T. P., Langton, S. D., Parkin, W. K., Peel, S., Bishop, J. D., Hart, J. D. & Moore, N. P. 2000. Habitat models of bird species' distribution: an aid to the management of coastal grazing marshes. *Journal of Applied Ecology*, 37(5):706–727.

Moisen, G. G. & Frescino, T. S. 2002. Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling*, 157(2-3):209–225.

Morrison, M. L., Marcot, B. G. & Mannan, R. W. 1998. *Wildlife-Habitat Relationships - Concepts and Applications*. University of Wisconsin Press, Madison, 2nd edition.

Mühlenberg, M. 1993. *Freilandökologie*. UTB für Wissenschaft. Quelle & Meyer, Heidelberg, Wiesbaden, 3te edition.

Nicholls, A. O. 1989. How to make biological surveys go further with generalised linear models. *Biological Conservation*, 50:51–75.

O'Connor, R. J. 2002. The conceptual basis of species distribution modelling: time for paradigm shift. In Scott, J. M., Heglund, P. J., Morrison, M., Haufler, J. B. & Wall, W. A., editors, *Predicting Species Occurrences: Issues of Accuracy and Scale*, pages 25–33. Island Press.

Oksanen, J. & Minchin, P. R. 2002. Continuum theory revisited: what shape are species responses along ecological gradients? *Ecological Modelling*, 157(2-3):119–129.

Olden, J. D. 2003. A species-specific approach to modeling biological communities and its potential for conservation. *Conservation biology*, 17(3):854–863.

Olden, J. D. & Jackson, D. A. 2002. A comparison of statistical approaches for modelling fish species distributions. *Freshwater Biology*, 47(10):1976–1995.

Oppel, S., Schaefer, H. M., Schmidt, V. & Schröder, B. 2004. Habitat selection by the Pale-headed brush-finch, *Atlapetes pallidiceps*, in southern Ecuador: implications for conservation. *Biological Conservation*, in press.

Osborne, P. E. & Alonso, J. C. 2000. Building models of avian habitat use at large spatial scales using GIS and remote sensing. In *4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs*, Banff, Alberta, Canada, September 2 - 8, 2000.

Osborne, P. E., Alonso, J. C. & Bryant, R. G. 2001. Modeling landscape-scale habitat use using GIS and remote sensing: a case study with great bustards. *Journal of Applied Ecology*, 38:458–471.

Özesmi, S. L. & Özesmi, U. 1999. An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, 116(1):15–31.

Özesmi, U. & Mitsch, W. J. 1997. A spatial habitat model for the marsh-breeding red-winged blackbird (*Agelaius phoeniceus* l.) in coastal Lake Erie wetlands. *Ecological Modelling*, 101(2,3):139–152.

Pearce, J. & Ferrier, S. 2000. An evaluation of alternative algorithms for fitting species distribution models using logistic regression. *Ecological Modelling*, 128(2-3):127–147.

Pearce, J. & Ferrier, S. 2001. The practical value of modelling relative abundance of species for regional conservation planning: a case study. *Biological Conservation*, 98(1):33–43.

Pearce, J. L., Burgman, M. A. & Franklin, D. C. 1994. Habitat selection by helmeted honeyeater. *Wildlife Research*, 21:53–63.

Pearsall, S. H., Durham, D. & Eagar, D. C. 1986. Evaluation methods in the United States. In Usher, M., editor, *Wildlife Conservation Evaluation*, pages 111–133. Chapman and Hall, London / New York.

Pearson, R. G. & Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12:361–372.

Pearson, S. M., Turner, M. G. & Drake, J. B. 1999. Landscape change and habitat availability in the southern Appalachian highlands and Olympic peninsula. *Ecological Applications*, 9(4):1288–1304.

Peeters, E. T. H. M. & Gardeniers, J. J. P. 1998. Logistic regression as a tool for defining habitat requirements of two common gammarids. *Freshwater Biology*, 39(4):605–615.

Peterson, A. T., Ball, L. G. & Cohoon, K. P. 2002. Predicting distributions of Mexican birds using ecological niche modelling methods. *Ibis*, 144(1):E27–E32.

Poff, N. L. 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. *Journal of the North American Benthological Society*, 16:391–409.

Prasad, A. M. & Iverson, L. R. 2000. Predictive vegetation mapping using a custom built model-chooser: comparison of regression tree analysis and multivariate adaptive regression splines. In *4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4): Problems, Prospects and Research Needs*, Banff, Alberta, Canada, September 2 - 8, 2000.

Pregibon, D. 1981. Logistic regression diagnostics. *Annals of Statistics*, 9(4):705–724.

Quinn, G. P. & Keough, M. J. 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge.

Ramsey, F. L. & Usner, D. 2003. Persistence and heterogeneity in habitat selection studies using radio telemetry. *Biometrics*, 59(2):332–340.

Reading, R. P., Clark, T. W., Seebeck, J. H. & Pearce, J. 1996. Habitat suitability index model for the eastern barred bandicoot, *Perameles gunnii*. *Wildlife research*, 23(2):221–236.

Recknagel, F. 2000. ANNA Artificial Neural Network model for predicting species abundance and succession of blue-green Algae. *Hydrobiologia*, 410:47–57.

Reineking, B. & Schröder, B. 2003. Computer-intensive methods in the analysis of species-habitat relationships. In Breckling, B., Reuter, H. & Mitwollen, A., editors, *Gene, Bits und Ökosysteme - Implikationen neuer Technologien für die ökologische Theorie*, pages 165–182. Peter Lang.

Reineking, B. & Schröder, B. 2004a. Gütemaße für Habitatmodelle. *UFZ-Bericht*, 9/2004:27–38.

Reineking, B. & Schröder, B. 2004b. Variablenselektion. *UFZ-Bericht*, 9/2004:39–46.

Reutter, B. A., Helfer, V., Hirzel, A. H. & Vogel, P. 2003. Modelling habitat-suitability using museum collections: an example with three sympatric *Apodemus* species from the Alps. *Journal of Biogeography*, 30(4):581–590.

Reyjol, Y., Lim, P., Belaud, A. & Lek, S. 2001. Modelling of microhabitat used by fish in natural and regulated flows in the river Garonne (France). *Ecological Modelling*, 146(1-3):131–142.

Ripley, B. D. 1996. *Pattern recognition and neural networks*. Cambridge University Press, Cambridge.

Rojas-Soto, O. R., Alcántara-Ayala, O. & Navarro, A. G. 2003. Regionalization of the avifauna of the Baja California Peninsula, Mexico: a parsimony analysis of endemicity and distributional modelling approach. *Journal of Biogeography*, 30(3):449–462.

Rolstad, J., Løken, B. & Rolstad, E. 2000. Habitat selection as a hierarchical spatial process: the green woodpecker at the northern edge of its distribution range. *Oecologia*, 124(1):116–129.

Rudner, M., Schröder, B. & Biedermann, R. 2004. Habitatmodellierung in GIMOLUS: e-Learning Module zur Verwendung der logistischen Regression zur Analyse der Art-Umwelt-Beziehung. *UFZ-Bericht*, 9/2004:57–65.

Rumelhart, D. E., Hinton, G. E. & Wiilliams, R. J. 1986. Learning representations by back-propagating errors. *Nature*, 323:533–536.

Schadt, S., Revilla, E., Wiegand, T., Knauer, F., Kaczensky, P., Breitenmoser, U., Bufka, L., Cervený, J., Koubek, P., Huber, T., Staniša, C. & Trepl, L. 2002. Assessing the suitability of central european landscapes for the reintroduction of eurasian lynx. *Journal of Applied Ecology*, 39(2):189–203.

Schamberger, M. L. & O'Neil, L. J. 1986. Concepts and constraints of habitat-model testing. In Verner, J., Morrison, M. L. & Ralph, C. J., editors, *Wildlife 2000: Modeling Habitat Relationships of Terrestrial Vertebrates*, pages 5–10. University of Wisconsin Press, Madison.

Scholten, M., Wirtz, C., Fladung, E. & Thiel, R. 2003. The modular habitat model (MHM) for the ide, *Leuciscus idus* (L.) - a new method to predict the suitability of inshore habitats for fish. *Journal of Applied Ichthyology*, 19(5):315–329.

Schröder, B. 1997. Fuzzy Logik und klassische Statistik - ein kombiniertes Habitateignungsmodell für *Conocephalus dorsalis* (Latreille 1804) (Orthoptera: Tettigoniidae). *Verhandlungen der Gesellschaft für Ökologie*, 27:219–226.

Schröder, B. 2000. *Zwischen Naturschutz und Theoretischer Ökologie: Modelle zur Habitateignung und räumlichen Populationsdynamik für Heuschrecken im Niedermoor*. Doktorarbeit, TU Braunschweig.

Schröder, B. 2001. Habitatmodelle für ein modernes Naturschutzmanagement. In Gnauck, A., editor, *Theorie und Modellierung von Ökosystemen - Workshop Kölpinsee 2000*,

pages 201–224. Shaker, Aachen.

Schröder, B. & Reineking, B. 2004a. Modellierung der Art-Habitat-Beziehung - ein Überblick über die Verfahren der Habitatmodellierung. *UFZ-Bericht*, 9/2004:5–26.

Schröder, B. & Reineking, B. 2004b. Validierung von Habitatmodellen. *UFZ-Bericht*, 9/2004:47–55.

Schröder, B. & Richter, O. 1999/2000. Are habitat models transferable in space and time? *Zeitschrift für Ökologie und Naturschutz*, 8:195–205.

Scott, J. M., Heglund, P. J., Morrison, M., Haufler, J. B. & Wall, W. A., editors 2002. *Predicting Species Occurrences: Issues of Accuracy and Scale*. Island Press.

Smith, P. A. 1994. Autocorrelation in logistic regression modelling of species' distributions. *Global Ecology and Biogeography Letters*, 4(2):47–61.

Söndgerath, D. & Schröder, B. 2002. Population dynamics and habitat connectivity affecting the spatial spread of populations - a simulation study. *Landscape Ecology*, 17:57–70.

Southwood, T. R. E. 1977. Habitat, the templet for ecological strategies? *Journal of Animal Ecology*, 46:337–365.

Steyerberg, E. W., Eijkemans, M. & Habbema, J. 2001a. Application of shrinkage techniques in logistic regression analysis: a case study. *Statistica Neerlandica*, 55(1):76–88.

Steyerberg, E. W., Eijkemans, M. J. C. & Habbema, J. D. F. 1999. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology*, 52(10):935–942.

Steyerberg, E. W., Harrell, Frank E., J., Borsboom, G. J. J. M., Eijkemans, M. J. C., Vergouwe, Y. & Habbema, J. D. F. 2001b. Internal validation of predictive models - efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology*, 54(8):774–781.

Stockwell, D. 1992. *Machine learning and the problem of prediction and explanation in ecological modelling*. PhD-thesis, Australian National University.

Stockwell, D. R. B. & Peters, D. 1999. The GARP Modeling System: problems and solutions to automated spatial prediction. *International Journal of Geographical Information Science*, 13(2):143–158.

Tappeiner, U., Tasser, E. & Tappeiner, G. 1998. Modelling vegetation patterns using natural and anthropogenic influence factors: preliminary experience with a GIS based model applied to an Alpine area. *Ecological Modelling*, 113(1-3):225–237.

Ter Braak, C. J. F., Hoijtink, H., Akkermans, W. & Verdonschot, P. F. M. 2003. Bayesian model-based cluster analysis for predicting macrofaunal communities. *Ecological Modelling*, 160(3):235–248.

Thomas, J. A. & Bovee, K. D. 1993. Application and testing of a procedure to evaluate transferability of habitat suitability criteria. *Regulated Rivers: Research and Management*, 8(3):285–294.

Tibshirani, R. 1995. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1):267–288.

Townsend, C. R., Dolédec, S., Norris, R., Peacock, K. & Arbuckle, C. 2003. The influence of scale and geography on relationships between stream community composition and landscape variables: description and prediction. *Freshwater Biology*, 48(5):768–785.

Trexler, J. C. & Travis, J. 1993. Nontraditional regression analyses. *Ecology*, 74:1629–1637.

U.S. Fish & Wildlife Service 1980. *Habitat Evaluation Procedures (HEP)*. USDI Fish and Wildlife Services, Division of Ecological Serveices, Washington DC.

U.S. Fish & Wildlife Service 1981. *Standards for the Development of Habitat Suitability Index Models*, volume 103 of *Ecological Services Manual*. U.S. Fish and Wildlife Services, Washington DC.

Vayssières, M. P., Plant, R. E. & Allen-Diaz, B. H. 2000. Classification trees: an alternative non-parametric approach for predicting species distributions. *Journal of Vegetation Science*, 11(6):679–694.

Venables, W. & Ripley, B. 1997. *Modern applied statistics with S-Plus*. Springer, New York, 2te edition.

Verboom, J., Schotman, A., Opdam, P. & Metz, J. A. J. 1991. European nuthatch metapopulations in a fragmented agricultural landscape. *Oikos*, 61:149–156.

Verbyla, D. L. & Litvaitis, J. A. 1989. Resampling methods for evaluation of classification accuracy of wildlife habitat models. *Environmental Management*, 13(6):783–787.

Vincent, P. J. & Haworth, J. M. 1983. Poisson regression models of species abundance. *Journal of Biogeography*, 10:153–160.

Wadsworth, R. A., Collingham, Y. C., Willis, S. G., Huntley, B. & Hulme, P. E. 2000. Simulating the spread and management of alien riparian weeds: are they out of control? *Journal of Applied Ecology*, 37:28–38.

Wahlberg, N., Moilanen, A. & Hanski, I. 1996. Predicting the occurrence of endangered species in fragmented landscapes. *Science*, 273(5281):1536–1538.

Wallace, C. S. A., Watts, J. M. & Yool, S. R. 2000. Characterizing the spatial structure of vegetation communities in the Mojave Desert using geostatistical techniques. *Computers And Geosciences*, 26(4):397–410.

Weiher, E. & Keddy, P., editors 1999. *Ecological Assembly Rules - Perspectives, Advances, Retreats*. Cambridge University Press, Cambridge, 1st edition.

Wessels, K. J., Van Jaarsveld, A. S., Grimbeek, J. D. & Van Der Linde, M. J. 1998. An evaluation of the gradsect biological survey method. *Biodiversity and Conservation*, 7(8):1093–1121.

White, G. C. & Bennetts, R. E. 1996. Analysis of frequency count data using the negative binomial distribution. *Ecology*, 77(8):2549–2557.

Wisnowski, J. W., Simpson, J. R., Montgomery, D. C. & Runger, G. C. 2003. Resampling methods for variable selection in robust regression. *Computational Statistics & Data Analysis*, 43(3):341–355.

Wolfinger, R. & OConnell, M. 1993. Generalized linear mixed models: a pseudolikelihood approach. *Journal of Statistical Computation and Simulation*, 48:233–243.

Wu, H. & Huffer, F. W. 1997. Modelling the distribution of plant species using the autologistic regression model. *Environmental and ecological statistics*, 4(1):31–48.

Yee, T. W. & Mitchell, N. D. 1991. Generalized additive models in plant ecology. *Journal of Vegetation Science*, 2:587–602.

Zaniewski, A. E., Lehmann, A. & Overton, J. M. 2002. Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Mo-*

delling, 157(2-3):261–280.

2.7 Datenblatt

2.7.1 Software

Die Autoren empfehlen die Verwendung von R (free software) oder S-Plus®(Insightful); bestimmte Verfahren lassen sich auch mit SPSS®und vielen anderen Statistikprogrammen umsetzen.

2.7.2 Webresources

R unter www.r-project.org, wichtige Bibliotheken für LRMs unter S-Plus bei F. Harrell: http://hesweb1.med.virginia.edu/biostat/s/splus.html, GARP unter http://biodi.sdsc.edu und http://beta.lifemapper.org/desktopgarp

2.7.3 Kommentierte Literatur

Siehe umfangreiche Angaben zu beispielhaften Veröffentlichungen in den einzelnen Kapiteln. Empfehlenswerte Bücher sind:

Fox (2002) An R and S-Plus companion to applied regression. - Sage.

Harrell (2001) Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. - Springer.

Hosmer & Lemeshow (2000) Applied logistic regression. - Wiley.

Quinn & Keough (2002) Experimental design and data analysis for biologists. - Cambridge University Press.

Crawley (2002) Statistical computing - an introduction to data analysis using S-Plus. - Wiley.

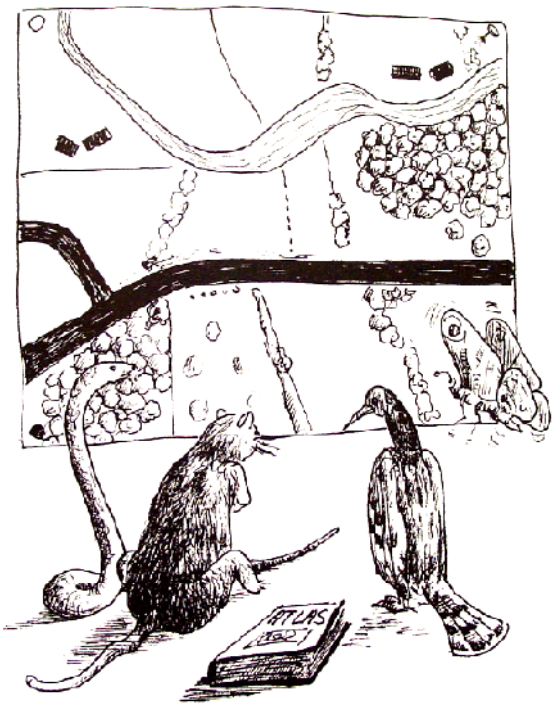
HABITATMODELLE

Methodik, Anwendung, Nutzen

Herausgeber

Carsten F. Dormann
Thomas Blaschke
Angela Lausch
Boris Schröder
Dagmar Söndgerath

Tagungsband
zum Workshop
8.-10. Oktober 2003,
UFZ Leipzig



UFZ - UMWELTFORSCHUNGSZENTRUM
LEIPZIG-HALLE GMBH IN DER HELMHOLTZ-GEMEINSCHAFT