

# Habitat-Modelle in der Wildökologie

**Diplomarbeit**  
eingereicht von  
**Massimo BASSI**

erstellt und betreut am  
Institut für Wildbiologie und Jagdwirtschaft  
Vorstand: o. Univ. Prof. Dr. Hartmut Gossow  
an der Universität für Bodenkultur, Wien

Wien, im August 2003

# Danksagung

"Wenn ich weiter gesehen habe, dann deswegen, weil ich auf den Schultern von Riesen stand." – Sir Isaac Newton

Mein Dank gilt zunächst Professor Hartmut Gossow, der mir diese Diplomarbeit angeboten, und somit eine zweite Chance gegeben hat. Ich hatte in ihm eine freundliche, hilfreiche und vor allem motivierende Unterstützung.

Der größte Dank gebührt natürlich meinen lieben Eltern, die mich während meines "ewigen Studiums" – nicht nur finanziell – tatkräftig unterstützt haben.

Bedanken möchte ich mich bei meiner Freundin Lu, die mich mühevoll – mit viel Geduld und Liebe – jeden Tag ein bißchen näher ans Ziel herangebracht hat.

Bleiben noch meine Geschwister und Freunde, die mir stets "zugesetzt" haben.

An dieser Stelle möchte ich auch Menschen danken, denen ich im alltäglichen Leben nie begegnet bin, die aber indirekt zum Gelingen dieser Arbeit beigetragen haben: Linus Torvalds für Linux, Donald E. Knuth für T<sub>E</sub>X, Leslie Lamport für L<sup>A</sup>T<sub>E</sub>X, Bram Moolenaar et al. für den 'vim' sowie den zahl- und namenlosen Mitstreitern der *open source community* für die vielen zusätzlich benutzten Programme.

Diesen Leuten ist es zu verdanken, daß ich meine Diplomarbeit ohne den Einsatz kommerzieller/proprietärer Software schreiben konnte.

Nicht zu vergessen sind natürlich all jene, auf deren Erkenntnisse ich mich stützen konnte.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung und Fragestellung</b>	<b>1</b>
<b>2</b>	<b>Material und Methoden</b>	<b>3</b>
<b>3</b>	<b>Realität, Modell und Simulation</b>	<b>5</b>
3.1	Realität bzw. reale Systeme . . . . .	5
3.2	Modell . . . . .	6
3.3	Simulation . . . . .	9
<b>4</b>	<b>Modellentwicklung</b>	<b>10</b>
4.1	Definitionen, Zielsetzungen und Anwendung . . . . .	10
4.1.1	Definitionen . . . . .	10
4.1.2	Zielsetzungen . . . . .	10
4.1.3	Anwendungsgebiete . . . . .	11
4.2	Modellbildungsprozess . . . . .	12
4.2.1	konzeptionelle Formulierung . . . . .	12
4.2.2	Statistische Formulierung . . . . .	12
4.2.3	Modellanpassung . . . . .	13
4.2.4	Vorhersage . . . . .	14
4.2.5	Evaluierung . . . . .	14
4.3	(Implizite) Annahmen und Fallstricke . . . . .	15
4.3.1	Pseudo-Equilibrium . . . . .	15
4.3.2	fundamental vs. realized niche . . . . .	15
4.3.3	Response-Verlauf . . . . .	16
4.3.4	Kausale vs. statistische Zusammenhänge . . . . .	17
4.4	Einfließende Variablen . . . . .	18
4.4.1	Zielgröße . . . . .	19
4.4.2	Habitatvariablen . . . . .	19
4.4.3	Einfluß der Tierarten auf ihr Habitat . . . . .	21
4.4.4	Datenquellen . . . . .	21
4.4.5	Auswahl der Variablen . . . . .	22
4.4.6	Sampling Design . . . . .	24

<b>5</b>	<b>Modellansätze</b>	<b>26</b>
5.1	Versuch einer Gliederung . . . . .	26
5.2	HSI-Modelle und HEP . . . . .	30
5.3	Verallgemeinertes lineares Modell (GLM) . . . . .	33
5.4	Verallgemeinertes additives Modell (GAM) . . . . .	36
5.5	Autologistisches Modell . . . . .	38
5.6	Ecological-Niche Factor Analysis (ENFA) . . . . .	40
5.7	FORSPACE . . . . .	44
5.8	Expertensysteme . . . . .	47
5.9	(Künstliche) Neuronale Netzwerke (ANN) . . . . .	49
5.10	Classification Trees . . . . .	51
5.11	GRASP . . . . .	53
5.12	GAP . . . . .	54
5.13	RSF . . . . .	55
5.14	Individual-based Models . . . . .	55
5.15	Populations Modelle . . . . .	56
<b>6</b>	<b>Modellanwendungen</b>	<b>57</b>
6.1	HSI . . . . .	57
6.2	GLM . . . . .	59
6.3	Autologistisches Modell . . . . .	60
6.4	ENFA . . . . .	61
6.5	FORSPACE . . . . .	63
6.6	Expertensystem . . . . .	65
6.7	ANN . . . . .	67
6.8	Classification trees . . . . .	70
6.9	Diskriminanzanalyse (DA) . . . . .	72
6.10	GAP . . . . .	74
6.11	RSF . . . . .	76
6.12	Logistische Regression . . . . .	77
<b>7</b>	<b>Modellvergleiche</b>	<b>79</b>
7.1	ENFA, GLM bzw. logistische Regression . . . . .	79
7.2	MR und ANN . . . . .	80
7.3	DA, ANN und logistische Regression . . . . .	81
7.4	ANUCLIM, DOMAIN, GAM, GLM und GARP . . . . .	81

7.5	DFA, CART, logistische Regression und Mahalanobis Distanz . . .	82
7.6	Hierarchical Set of Models und GAM . . . . .	83
<b>8</b>	<b>Schlußbetrachtung</b>	<b>84</b>
<b>9</b>	<b>Zusammenfassung</b>	<b>88</b>
<b>10</b>	<b>Glossar und Abkürzungsverzeichnis</b>	<b>90</b>
	<b>Literaturverzeichnis</b>	<b>103</b>

# Tabellenverzeichnis

4.1	Variablen für ein HSI-Modell für Auerwild . . . . .	20
5.1	Klassifizierung von Habitatmodellen . . . . .	27
5.2	<i>link</i> -Funktionen . . . . .	34
5.3	inverse <i>link</i> -Funktionen . . . . .	36
6.1	Formeln für verschiedene Datensätze . . . . .	60

# Abbildungsverzeichnis

3.1	Beziehungen zw. realem System und Modell . . . . .	5
3.2	Eigenschaften von Modellen . . . . .	8
4.1	5 Schritte zum Modell . . . . .	13
4.2	lineare und kurvilineare Response . . . . .	16
4.3	<i>Of Correlates and Causes</i> . . . . .	18
4.4	Vier Samplingstrategien . . . . .	24
5.1	Exemplarisches Baumdiagramm für terrestrisches Habitat . . . . .	31
5.2	Die 4 bzw. 8 nächsten Nachbarn einer Zelle . . . . .	39
5.3	Verteilung entlang einer EGV . . . . .	40
5.4	Eignungsindex einer Zelle . . . . .	43
5.5	Komponenten eines Expertensystems . . . . .	48
5.6	<i>BackPropagation Network</i> und Kohonen SOM . . . . .	50
5.7	<i>decision tree</i> . . . . .	52
6.1	Eignungsindex-Funktionen für den Winter . . . . .	57
6.2	Ausschnitt aus Habitateignungskarte für Steinbock . . . . .	62
6.3	Gewicht und Anzahl von Pflanzenfressern . . . . .	63
6.4	Rotwild im Jahresverlauf . . . . .	64
6.5	Merkmalsinterpretation im Expertensystem WIFES . . . . .	66
6.6	BPN für Rotschulterstärling und Sumpfzaunkönig . . . . .	68
6.7	Sensitivitätsanalyse der Habitatvariablen . . . . .	70
6.8	Ljubljana-Tier Sommerlebensraum . . . . .	71
6.9	Aktuelles Vorkommen und Kerngebiet . . . . .	73
6.10	<i>mimimum critical areas and functional connectivity</i> . . . . .	75

# 1. Einleitung und Fragestellung

Thema dieser Diplomarbeit sollen Habitat-Modelle in ihren verschiedensten Ausprägungen sein, wie sie in der Wildökologie Verwendung finden.

Zunächst sollen allgemeine Aspekte der Modellentwicklung behandelt werden. Hierbei wird auf Definitionen, Zielsetzungen und Anwendungsgebiete von Habitatmodellen eingegangen. Der Modellbildungsprozeß an sich soll beleuchtet werden. Wie ist er zu gestalten? Welche Bedeutung ist den einzelnen Schritten beizumessen? Fragen ergeben sich durch explizite und implizite Annahmen bezüglich grundlegender Konzepte. Wie vertragen sich einem Modell zugrunde liegende Annahmen mit den Postulaten der Ökologie? Werden kausale oder statistische Zusammenhänge beschrieben? Auch im Zusammenhang mit den Variablen, die zur Entwicklung eines Modells verwendet werden, gilt es einige Überlegungen anzustellen: Welche Faktoren beeinflussen Qualität, Eignung bzw. Verfügbarkeit eines Lebensraumes? Wie wird bestimmt, welche Variablen ins Modell einfließen sollen? Woher stammen die Daten? Wie sollen sie aufbereitet sein?

In einem eigenen Teil soll eine prinzipielle Darstellung der Ansätze erfolgen. Hier werden die verschiedenen Modelle besprochen und die ihnen zugrundeliegenden Annahmen diskutiert. Kann man Familien unterscheiden oder sind es lauter individuelle Ansätze ohne gemeinsamen theoretischen Hintergrund? Worin unterscheiden sie sich, was haben sie gemeinsam? Stehen diese in Einklang mit ökologischen Theorien? Geben Modelle tatsächlich kausale Zusammenhänge wieder oder sind es reine Hypothesen?

Ein Abschnitt soll der Besprechung konkreter Anwendungen der besprochenen Modelle gewidmet sein. Es wird versucht, anhand dieser Beispiele Schwächen und Stärken eines Ansatzes aufzuzeigen. Welche Einflüsse werden berücksichtigt? Welche nicht, oder welche sind entbehrlich? Fließen (Stör-)Faktoren wie Fütterung, Jagd, Freizeitaktivitäten, Landnutzung und Katastrophen (Waldbrand, Überflutung) ein? Sind Modelle auf andere Gebiete übertragbar? Mit welchen



Einschränkungen? Wo liegen die Schwerpunkte der Modelle (Naturschutz, Management, Wildschäden, Raubtiere)? Gibt es Unterschiede im Vergleich Europa als Kulturlandschaft und Nordamerika als – im *Pacific North West* weitgehend – Naturlandschaft?

Durch den Vergleich der verschiedenen Ansätze, besprochen anhand von Beispielen aus der Literatur, sollen folgende Aspekte erörtert werden: Gibt es das beste Modell? Oder sind alle gleichwertig? Welches Modell ist wofür besser geeignet?

Als Ausgangspunkt sollen die Habitat-Bewertungs- bzw. -Eignungsmodelle HEP (Habitat Evaluation Procedure) und HSI (Habitat Suitability Index) vom US Fish and Wildlife Service dienen (U.S. Fish and Wildlife Service, 1980a,b, 1981). Weitere Anhaltspunkte lieferten eine niederländische Arbeit (Kramer et al., 2001, Landscape forming processes and diversity of forested landscapes - description and application of the model FORSPACE), die Dissertation von Mike BOKALO (Bokalo, 2001, The Role of Forest Growth in Habitat Quality Dynamics) und jene von Ernst PARTL (Partl, 2001, Die Wechselwirkungen zwischen Wald als Habitat und Rehwild als Standortfaktor bei der Waldsanierung).

Speziell im Zusammenhang mit der letztgenannten Arbeit stellt sich die Frage: Gibt es einen "Technologietransfer" nur in Richtung Nordamerika -> Europa? Oder fließen Denkanstöße z.B. bezüglich Wildschädenprädisposition auch in die Gegenrichtung?

## 2. Material und Methoden

Der Begriff "Habitat-Modelle" ist ein – wie sich bald herausstellen sollte – sehr vages Konzept. Auf der Suche nach Ansätzen wird man mit einer Unzahl – teils individueller – Verfahren konfrontiert, deren Einordnung schwer fällt.

Bei der Suche nach Literatur bediente ich mich ausgiebig des Internets, im speziellen der "Elektronischen Zeitschriftenbibliothek" der Universität Regensburg (EZB, <http://rzblx1.uni-regensburg.de/ezeit/fl.phtml?bibid=UBWI>). Diese war mir beim Auffinden der einschlägigen wissenschaftlichen Zeitschriften behilflich. Zugang zu den in elektronischer Form verfügbaren Journals/Artikeln hatte ich über die Netze der Universität für Bodenkultur und der Hauptuniversität Wien. Weiters kamen die Online-Kataloge (ALEPH) der Haupt- (BOKU, UNIWIE, TU-WIEN) und Instituts-Bibliotheken (Inst. f. Wildbiologie u. Jagdwirtschaft, Inst. für Mathematik und Statistik, Inst. f. Zoologie, Inst. f. Botanik) zum Einsatz. Auch Suchmaschinen, in der Hauptsache Google (<http://www.google.com/>) wurden konsultiert.

Aufgrund der Tatsache, daß der Großteil der Literatur – wie zu erwarten war – in englischer Sprache vorliegt, wurden natürlich englische keywords benutzt. Schlagwörter, nach denen ich – zunächst – suchte, waren: *habitat, evaluation, suitability, model(l), model(l)ing, ecological, wildlife*.. Später, als sich die Richtung abzeichnete, waren es: *generalized linear model, generalized additive model, neural network, expert system, decision tree*, u.a.m.

Die meisten Beiträge wurden jedoch über Zitate bzw. Literaturverzeichnisse in den bereits durchgesehenen Artikeln gefunden (Schneeballsystem). Vereinzelt sind mir Bücher auch unverhofft untergekommen.

Im Text wird den englischen Fachausdrücken und Abkürzungen der Vorzug gegeben, v.a. in solchen Fällen, wo mir keine entsprechenden deutschen Fachausdrücke bekannt sind, bzw. akute Verballhornungsgefahr gegeben wäre.

Öfters verwendete Begriffe bzw. Ausdrücke, auf die im Text nicht näher eingegangen wird, sind in einem eigenen Verzeichnis angeführt (siehe S. 90), eventuell mit

kurzer Beschreibung und Literaturangabe. Die darin aufgenommenen Einträge sind im Text entsprechend gekennzeichnet (z.B. →ANN) und funktionieren in der pdf-Version als Link.

Im Literaturverzeichnis werden auch Einträge angeführt, die zwar hier im Text nicht zitiert werden, mir aber beim Erstellen der Arbeit hilfreich waren bzw. weitere Anhaltspunkte liefern können.

## 3. Realität, Modell und Simulation

### 3.1 Realität bzw. reale Systeme

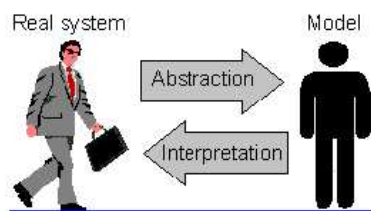
Ein System ist die Gesamtheit miteinander verknüpfter und sich gegenseitig beeinflussender Elemente, die entsprechend einem bestimmten Zweck organisiert sind. Das System hat eine gänzlich andere Qualität als die Summe seiner Elemente (Meadows et al., 1993).

**Klassen von Systemen** Systeme können anhand unterschiedlicher Kriterien in Klassen eingeordnet werden, die sich nicht gegenseitig ausschließen.

So kann man z.B. offene von (idealtypisch) geschlossenen Systemen unterscheiden: während beim ersten Wechselwirkungen mit der Umwelt möglich sind (z.B. Energiezufuhr Sonne -> Erde), ist das bei geschlossenen nicht der Fall (Aquarium).

In statischen Systemen gibt es keine Veränderung der Komponenten, während in dynamischen die Systemgrößen Änderungen unterworfen sind.

In diskreten Systemen ändern sich die Systemvariablen sprunghaft in endlichen Zeitabschnitten (z.B. Populationsgröße). In stetigen Systemen ändern sich die Parameter kontinuierlich in beliebig kleinen Zeitabschnitten (z.B. Körpergewicht). Stabile Systeme sind unempfindlich gegenüber Störungen. Änderungen halten sich in Grenzen, da sie vom System gepuffert werden. Instabile Systeme dagegen sind empfindlich gegenüber Störungen: Geringe Änderungen bewirken Unvorhergesehenes, das System *kippt*.



(Sharow, 1996)

Abbildung 3.1: Beziehungen zw. realem System und Modell

## 3.2 Modell

**Was ist ein Modell?** Ein Modell ist die formale Darstellung eines Teiles/Ausschnittes der realen Welt/Realität (Morrison et al., 1998). Nach Tipton (1980) ist ein Modell die (vereinfachte) Darstellung von (komplizierten) Systemen/Sachverhalten.

Man kann grundsätzlich physische bzw. gegenständliche Modelle (z.B. Modell-Flugzeug im Windkanal) von abstrakten "Abbildungen" eines Systems (z.B. Konzept, (Fluß-)Diagramm, mathematisches Modell) unterscheiden. Abstraktion bedeutet Beschränkung/Reduktion auf "wesentliche" Komponenten bzw. Wechselwirkungen der modellierten Realität. Diese Vereinfachung soll das System überschaubarer und damit nachvollziehbar machen.

Auch im alltäglichen Leben selektieren wir die uns wichtig erscheinende Information aus dem schier unüberschaubaren Informationsschwall, der andauernd über uns hereinbricht. Nur ein geringer Teil der Daten ist (über-)lebensnotwendig und wird ausgewertet. Der Rest kann – situationsabhängig – getrost ignoriert werden. Wir leben also (unbewußt) nach einem Modell, das in unserem Kopf aus vergangenen Erfahrungen entsteht und sind geübt in der vereinfachten Darstellung komplexer Zusammenhänge. Tatsächlich hätte es auch gar keinen Sinn, ein Modell in der selben Komplexität wie die Realität zu konstruieren, denn es würde den Vorteil der Übersichtlichkeit verlieren.

Realität und deren Modell sind über 2 Vorgänge miteinander verbunden: Abstraktion und Interpretation. Wie man durch Abstraktion von der Realität zum Modell gelangt, so führt der Weg zurück, also vom Modell zur Realität, über die Interpretation. Dabei werden Modell-Komponenten (Parameter, Variablen) und Modell-Verhalten als Komponenten, Eigenschaften und Verhalten des realen Systems ausgelegt und bewertet (Sharow, 1996).

**Zweck eines Modells** Im wesentlichen erfüllen Modelle folgende Funktionen:

- Beschreiben: Die Abläufe innerhalb eines realen Systems werden formal dargestellt. Dabei werden eventuelle Wissenslücken identifiziert.
- Erklären: Das Modell soll helfen, Einsicht in die Funktionsweise realer Systeme zu gewinnen.
- Vorhersage/Steuern: Das Modell ist Grundlage für Entscheidungen und Evaluation alternativer Managementstrategien.

**Typen und Eigenschaften von Modellen** Ähnlich wie Systeme können auch Modelle klassifiziert werden.

Bei qualitativen Modellen werden Elemente und ihre Beziehungen nur beschrieben. Bei quantitativen hingegen werden Komponenten und Wechselbeziehungen beziffert.

Einfache Modelle beruhen auf simplen Beschreibungen (exponentielles Wachstum einer Population  $N_t = N_0 * e^{rt}$ ). Ist das Modell zu einfach, kann es u.U. das reale Verhalten des Systems nicht mehr angemessen beschreiben. Beispiel für ein komplexes Modell ist das vom Club of Rome in Auftrag gegebene *World3*-Modell, (Meadows et al., 1993). Je komplexer, desto geringer ist die Einsicht in die Interaktion der Elemente.

Mechanistische (auch als *theoretical, functional, biological, heuristic* bezeichnete) Modelle bilden die tatsächliche Funktionsweise des realen Systems nach. Z.B. müßte ein mechanistisches Modell für den Temperaturgang im Laufe eines Jahres Faktoren wie Neigung der Erdachse, Breitengrad, Höhe, Exposition u.ä.m. berücksichtigen. Empirische Modelle beruhen auf einfachen, an experimentelle Daten angepassten Gleichungen. Sie beschreiben beispielsweise die Temperatur als Funktion des Kalendermonats ( $T_{[C^\circ]} = a + b * \cos(2 * \pi * \frac{(\text{month}+c)}{12})$ ).

Analytische Modelle haben, wie der Namen schon sagt, analytische Lösungen, numerische entsprechend nur numerische.

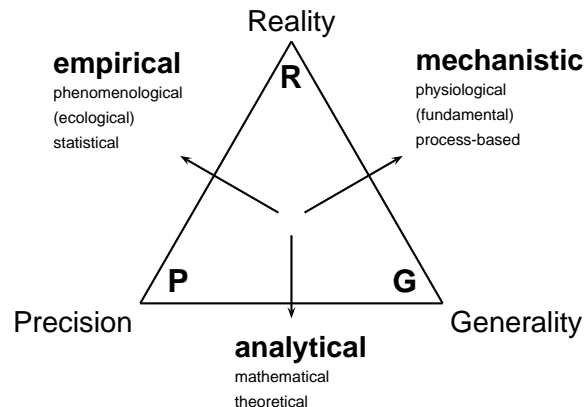
Statische (stationäre) Modelle entsprechen Momentaufnahmen (z.B. Verbreitung einer Tierart), dynamische hingegen beschreiben Veränderungen über Zeiträume (z.B. den Prozeß der Migration).

In diskreten Modellen ändern sich Systemvariablen (z.B. Populationsgröße) sprunghaft in endlichen Zeitabschnitten. In stetigen (*continuous*) Modellen ändern sich die Komponenten (z.B. Körpergewicht eines Individuums) kontinuierlich in beliebig kleinen Zeitabständen.

In deterministischen Modellen ist der Output bzw. das Verhalten allein durch die Ausgangslage (Input) bestimmt. Unter identischen Bedingungen sind identische Folgezustände reproduzierbar. Stochastische Modelle versuchen die Variabilität eines Systems durch Funktionen und Verteilungen zu berücksichtigen. Bei identischen Bedingungen sind Folgezustände nur durch Wahrscheinlichkeitsaussagen beschreibbar.

Im konzeptionellen Modell ist zwar die Struktur definiert, aber die Elemente und deren Beziehungen sind nicht quantifiziert.

Levins formulierte 1966 (zitiert in Guisan und Zimmermann (2000)) ein Prinzip,



Klassifizierung von Modellen auf der Grundlage ihrer Eigenschaften nach Levins (1966) bzw. Sharpe (1990).

(Guisan und Zimmermann, 2000, verändert)

Abbildung 3.2: Eigenschaften von Modellen

demzufolge ein Modell nur 2 der 3 wünschenswerten Eigenschaften (Allgemeingültigkeit, Realismus und Präzision) gleichzeitig aufweisen kann, während die dritte vernachlässigt werden muß (siehe Abb. 3.2).

**Modellentwicklung** Modellbildung (engl. *model(l)ing*) ist ein zyklischer, iterativer Prozess, bei dem nach dem "Versuch und Irrtum bzw. Erfolgs-"Prinzip vorgegangen wird. Ausgangspunkt ist im Idealfall nicht die Datensammlung, sondern die Modellentwicklung. Der Modellbildungsprozess selbst soll klären, welche Daten gebraucht werden. Allzu oft, so Tipton (1980), wird das Modell an bereits vorliegende Daten angepaßt.

Tipton (1980) gliedert die Modellentwicklung in 8 Schritte:

- verbale Beschreibung des Systems,
- Definition des Problems,
- grobe graphische Darstellung der kausalen Zusammenhänge,
- Erstellen eines detaillierten Flußdiagramms,
- Aufstellen der Gleichungen,
- Implementieren der Gleichungen als Software,
- Validieren des Modells (anhand anderer Daten),
- Evaluation, Kommunikation und Ausarbeitung von Empfehlungen zu Anwendung/Anwendbarkeit.

**Fehlerquellen** Fast jeder Schritt im Modellbildungsprozeß ist fehleranfällig. Im Zuge der Abstraktion, bedingt durch die Notwendigkeit der Reduktion auf das Wesentliche, kommt es zu Vereinfachungsfehlern. Durch die Anwendung ungeeigneter Rechenverfahren ergeben sich Verfahrensfehler (Diskretisierungsfehler, Abbrechfehler). Rechenfehler entstehen einerseits durch menschliches Versagen, andererseits in der "Maschine" beim Rundungsvorgang (Rundungsfehler bei Maschinenzahlen sind eigentlich Verfahrensfehler). Datenfehler ergeben sich sowohl während der Datenerhebung, als auch bei der Eingabe (Schnittstelle Mensch). Analog zur Abstraktion, kommt es im Zuge der Interpretation des Modelloutputs zu Auslegungsfehlern.

**Chancen und Gefahren** "*Models are always wrong ... but many of them are useful.*" behauptet zumindest Sharow (1996). Er fragt sich denn auch, wie es möglich ist, daß ein falsches Modell die richtige Antwort liefert: So wie alte Landkarten, die, obwohl sie eine flache Welt unterstellten und damit falsche Abstandsrelationen, für Reisende – in der Vergangenheit – trotzdem nützlich waren.

Prognostizierende Modelle werden von Managern als wertvolle Entscheidungsgrundlage zur Steuerung/Kontrolle realer Systeme genutzt. Allerdings stellen diese Modelle für Entscheidungsträger meist eine *black box* dar. Nicht-Wissen um die inneren Zusammenhänge und blindes Vertrauen bergen das Risiko von Fehlentscheidungen.

Beschreibende Modelle gewähren Einblick in die Funktionsweise realer Systeme. Dabei ist jedoch zu beachten, daß statistische Zusammenhänge nicht zwingend kausale Zusammenhänge widerspiegeln.

### 3.3 Simulation

Simulation wird im alltäglichen Sprachgebrauch oft mit Modell und Modellbildungsprozeß gleichgesetzt. In der mir vorliegenden Literatur steht Simulieren für "Spielen" mit dem (fertigen) Modell, z.B. im Sinne von (wiederholter) Anwendung des Modells, um verschiedene Szenarien durchzuspielen oder um aus den (stochastisch bedingten) zufälligen Schwankungen einen Mittelwert zu bilden.



## 4. Modellentwicklung

### 4.1 Definitionen, Zielsetzungen und Anwendung

#### 4.1.1 Definitionen

Definitionen sind im Zusammenhang mit Modellen insofern wesentlich, als daß sie die theoretischen Überlegungen zu Aufbau, Funktionsweise und Interpretation eines Modells beeinflussen. Es gibt in der Literatur immer wieder Aufrufe (Hall et al., 1997) und Versuche (Morrison und Hall, 2002) eine einheitliche Terminologie zu entwickeln und somit einen gemeinsamen Nenner zu finden. Tatsache ist, daß Begriffe wie Habitat, Biotop, Biozönose, Umwelt im anglo-amerikanischen Sprachgebrauch anders ausgelegt werden als z.B. im deutschsprachigen Raum. Die Konzepte werden nicht konsequent unterschieden, überlagert und teilweise als Synonym verwendet. Das mag wohl auch mit der Sprache an sich, der Auffassung der Worte selbst, zu tun haben.

Trotzdem müssen zwei Begriffe, auf die ständig Bezug genommen wird, definiert werden:

**Habitat:** Morrison und Hall (2002) liefern neben einer ausführlichen Diskussion des Habitat-Konzeptes (und anderer Begriffe wie Nische, Ökosystem, Population) diese knappe Definition: Habitat ist der physische Raum, in dem ein Tier lebt, samt der in diesem Raum vorkommenden biotischen und abiotischen Elemente (Ressourcen, Pflanzen, andere Tiere, ...).

**Nische:** Das Nischen-Konzept ist in diesem Sinne restriktiver: es beschränkt sich auf jene Elemente, die für die jeweilige Spezies (überlebens-)relevant sind. Der Raum, in dem sich ein Organismus aufhält/bewegt, wird sozusagen durch diese Faktoren überhaupt erst definiert.

#### 4.1.2 Zielsetzungen

Habitat-Modelle werden eingesetzt, um die Eignung eines Untersuchungsgebietes als Lebensraums für eine bestimmte Zieltierart oder auch mehrere -arten zu

bewerten bzw. um die Vorkommens-Wahrscheinlichkeit einer Spezies oder u.U. auch deren Abundanz zu schätzen.

Mit der Entwicklung von Modellen zur Wildtier-Habitat Beziehung werden laut Morrison et al. (1998) 5 Ziele verfolgt:

- *Formalisieren* – oder *Beschreiben* – des *aktuellen Wissensstandes* über eine Art oder ein Ökosystem.
- *Verstehen*, welche *Umwelt-Faktoren* die Verbreitung und Häufigkeit einer Spezies beeinflussen.
- *Vorhersage* der *Ausbreitung und Häufigkeit* einer Art.
- *Identifizieren* von *Schwächen* in unserem Verständnis und
- *Generieren* von *Hypothesen* über eine Art oder ein System.

Was die Vorhersage anbelangt, unterscheidet Morrison et al. (1998) zwischen *hindcasting* und *forecasting*. Bei *hindcasting* wird versucht, beobachtete Muster (Vorkommen, Abundanz) zu erklären. Es hat einen räumlich eingegrenzten, lokalen Bezug, beruht auf empirischen Ansätzen und arbeitet mit Korrelationen (statistischen Zusammenhängen). Bei *forecasting* handelt es sich um eine explizite Vorhersage, es kommt also auch eine zeitliche Dimension ins Spiel. Die in einem Untersuchungsgebiet gewonnenen Erkenntnisse werden auf andere Landstriche extrapoliert. Dabei kommen sogenannte theoretische bzw. mechanistische Ansätze zum Einsatz, die auf kausalen Zusammenhängen beruhen.

Für Guisan et al. (2002) sind die zwei wichtigsten Gründe für den Einsatz statistischer Modelle (1) deren Fähigkeit, das Verhalten ökologischer Systeme zu beschreiben und (2) die Möglichkeit Vorhersagen machen zu können.

Erklärende Modelle gewähren Einsicht in ökologische Prozesse ( $\rightarrow$ process) die ihrerseits bestimmte Muster ( $\rightarrow$ pattern) hervorbringen. Mit prognostizierenden Modellen sollen Erkenntnisse über Zusammenhänge zwischen Habitatvariablen und Vorkommen bzw. Abundanz einer Art auf andere, nicht in die ursprüngliche Studie einbezogene Gebiete übertragen werden.

### 4.1.3 Anwendungsgebiete

Modelle zur "Vorhersage der Habitatverteilung" (Guisan und Zimmermann, 2000) finden in vielen Bereichen Verwendung: von der Naturschutzbiologie (*conservation biology*), über die Klimawandelforschung (*climate change research*), Biogeographie (*biogeography*), bis hin zu Biotop- und Artenmanagement (*habitat*

*or species management*).

Die Naturschutzökologie bedient sich multivariater, räumlich expliziter Modelle "in allen Bereichen die einen gewissen Grad an ökologischem Realismus erfordern" (Hirzel et al., 2002): PVAs (*Population Viability Analyses*), Risikobewertung in Zusammenhang mit Biodiversitätsverlust (*biodiversity-loss risk assessment*), Biotopmanagement für gefährdete Tierarten (*landscape management for endangered species*), Sanierung von Ökosystemen (*ecosystem restoration*) und Ausbreitung von Exoten (*alien-invaders expansion*).

Guisan et al. (2000) beschreiben *predictive geographical modeling* als Werkzeug zur Bewertung der Auswirkung beschleunigter Landnutzung und anderer Umweltveränderungen (z.B. Klimaveränderung) auf die Verbreitung von Organismen, der Artenvielfalt oder einfach "nur" zum Testen biogeographischer Hypothesen.

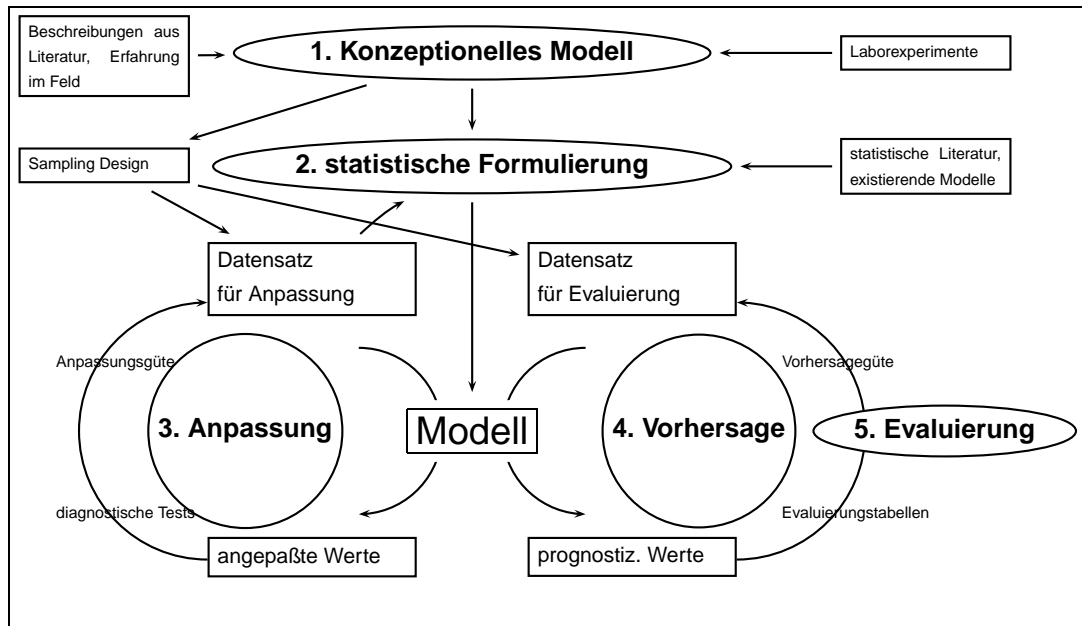
Nicht unerwähnt bleiben soll die Arbeit von Partl (2001), die sich mit der Habitatattraktivität für Rehwild in Zusammenhang mit der Verbißprädisposition von Jungwuchsflächen beschäftigt.

## 4.2 Modellbildungsprozess

Ein Modell entsteht zwar meistens in einem iterativen Prozeß, dennoch kann man 5 logische Schritte unterscheiden (Guisan und Zimmermann, 2000): konzeptionelle Formulierung (*conceptual model formulation*), statistische Formulierung (*statistical model formulation*), Anpassung des Modells (*model calibration*), Vorhersage (*model prediction*) und schließlich Evaluierung (*model evaluation* (siehe Abb. 4.1).

### 4.2.1 konzeptionelle Formulierung

Knackpunkt bei der Formulierung eines ökologischen Modells ist das Gefüge biologischer (und anderer) Konzepte, das den dabei getroffenen Annahmen zugrunde liegt. Dazu zählen Aspekte wie Pseudo-Equilibrium, fundamentale vs. realisierte Nische aber auch Überlegungen bezüglich der verwendeten Prädiktoren (direkt vs. indirekt). Auf einige dieser Punkte wird weiter unten eingegangen (siehe 4.3 und 4.4)



Die 5 Schritte im Modellbildungsprozess mit zwei Datensätzen - einer zum Anpassen und einer zum Evaluieren des Modells. Die Modellevaluierung erfolgt entweder mit dem *calibration*-Datensatz (und *bootstrap*, *cross validation* oder *jackknife* Techniken) oder mit einem unabhängigen Datensatz (und z.B. ROC-plot).

(Guisan und Zimmermann, 2000, verändert)

Abbildung 4.1: 5 Schritte zum Modell

### 4.2.2 Statistische Formulierung

Dieser Schritt wird in der einschlägigen Literatur auch als *verification* bezeichnet. Unter diesen Punkt fällt die Wahl (1) eines geeigneten Algorithmus zur Vorhersage einer Response-Variable (Vorkommen/Nichtvorkommen, Abundanz) und zum Schätzen der entsprechenden Koeffizienten und (2) des optimalen statistischen Ansatzes.

Die meisten statistischen Modelle sind an einen spezifischen Zielgrößen-Typ gebunden und unterliegen damit bestimmten Wahrscheinlichkeitsverteilungen (bzw. Dichtefunktionen). Es handelt sich also um sogenannte  $\rightarrow$ parametrische Verfahren (siehe z.B. Annahmen bezüglich Regression in 5.3). Ob ein Ansatz angemessen ist oder nicht, kann mit einer Vielzahl von Tests und graphischen Methoden verifiziert werden.

### 4.2.3 Modellanpassung

In diesem Schritt erfolgt die Anpassung des Modells, d.h. die Schätzung und Justierung der Modellparameter und -konstanten um die Übereinstimmung zwischen Modell-Output und ursprünglichem Datensatz zu verbessern. Um die Genauigkeit und Vorhersagefähigkeit eines Modells zu erhöhen, ist es u.U. notwendig, die Anzahl der einfließenden Variablen auf ein vernünftiges Maß zu beschränken. Es gilt also zu entscheiden, welche erklärenden Variablen ins Modell aufgenommen werden sollen (siehe 4.4.5). Das Einstellen ihrer Koeffizienten ist gewöhnlich das kleinere Problem.

### 4.2.4 Vorhersage

Ist das Modell angepaßt, so ist der nächste Schritt die Vorhersage z.B. der potentiellen Verbreitung einer Art im modellierten Gebiet. Das Modellieren der Verbreitung einer Art (bzw. einer *community*) entspricht im wesentlichen dem Modellieren ihres Habitates. Habitatmodelle werden oft in (bestehende) geographische Informationssysteme integriert, als Ergebnis erhält man dementsprechend Karten die geeignetes bzw. ungeeignetes Habitat ausweisen.

### 4.2.5 Evaluierung

In diesem Stadium der Modellentwicklung soll geprüft werden, ob das Modell falsifizierbare Hypothesen und zutreffende Vorhersagen ökologischer Muster bereitstellt bzw. zu konkreten operationalen Parametern beiträgt. Guisan und Zimmermann (2000) ziehen den Begriff Evaluierung (*evaluation*) dem Begriff Validierung (*validation*) vor, weil ja nicht die Wahrhaftigkeit des Modells in Frage gestellt wird.

Zwei Haupt-Ansätze existieren zum Evaluieren der Vorhersagefähigkeit von Modellen: Der erste Ansatz besteht darin, einen einzigen Datensatz (*single data set*) sowohl zum Anpassen als auch zum Evaluieren heranzuziehen. Dazu muß der ursprüngliche Datensatz mittels Kreuzvalidierung (*cross-validation*,  $\rightarrow$ CV), *leave-one-out jackknife* ( $\rightarrow$ JK), oder  $\rightarrow$ bootstrap Techniken aufbereitet werden (Datenpartitionierung). Dieser Ansatz eignet sich für den Fall, daß der vorhandene Datensatz zu klein ist, um in separate Datensätze aufgeteilt zu werden, bzw. wenn es vorzuziehen ist, möglichst viele Beobachtungen in die Modellanpassung auf-

zunehmen.

Der zweite Ansatz ist die Verwendung von zwei unabhängigen Datensätzen, einen zum Anpassen und einen zum Evaluieren. Die Datensätze werden entsprechend als Trainings- bzw. als Evaluierungs-Datensatz bezeichnet. Entweder werden von vorne herein 2 getrennte Datensätze erhoben (z.B. Anpassungsdatensatz aus einer stratifizierten Stichprobe, Evaluierungsdatensatz aus vorangegangenen Beobachtungen) oder ein einheitlicher Datensatz wird, so er groß genug ist, zweigeteilt (*split sample approach*).

Die Vernachlässigung der Validierung stellt nach Guisan et al. (2002) ein ernsthaftes Problem in der ökologischen Modellierung dar. Nur wenige Studien, die sich einer Regressionsanalyse zu Vorhersagezwecken bedienen, setzen statistische Validierungstechniken ein, trotz der Vielzahl existierender Methoden. Noch weniger führen eine Validierung im Feld durch, und stellen damit Gültigkeit und Anwendbarkeit des Modells in Frage.

Van Horne (2002) spricht sich im Gegensatz dazu gegen eine Überbewertung der Modell-Validierung aus. Man könne nicht von jedem Modell eine Vorhersagegenauigkeit von  $P < 0.05$  erwarten. Man müsse die Tatsache akzeptieren, daß Habitat-Modelle (ein) Mittel sind, um das vorhandene Wissen über Tier-Habitat Beziehung zahlenmäßig auszudrücken. Auf dieser Grundlage sind die bestinformiertesten Entscheidungen zu treffen, sowie eventueller Forschungs-Bedarf aufzudecken.

### 4.3 (Implizite) Annahmen und Fallstricke

*"Any modeling implementation will include unrealistic biological assumptions, not meet statistical assumptions, omit causal relationships, and/or fail to meet modeling objectives"*, soweit Van Horne (2002).

#### 4.3.1 Pseudo-Equilibrium

Guisan et al. (2002) geben zu bedenken, daß der Einsatz von Regressions-Analysen zur Modellierung einer Spezies ein "Pseudo-Equilibrium" zwischen Organismen und deren Umwelt unterstellt. Die betrachtete Population sollte sich also weder auf Expansionskurs noch in einem Schrumpfungsprozeß befinden. Ansonsten kann der Einsatz dieser Techniken – wenn zum Identifizieren von Umwelt-

faktoren welche die Verbreitung dieser Art bestimmen eingesetzt – zu falschen Ergebnissen führen, wie z.B. "abgeschnittenen" Response-Graphen (Hirzel et al., 2001b).

### 4.3.2 fundamental vs. realized niche

GLMs und GAMs modellieren, aufgrund ihrer empirischen Natur, tatsächlich eher die ökologische – sozusagen die umgesetzte (*realized*) – als die "fundamentale" Nische. Sie beinhalten also implizit Interaktionen mit der Umwelt (Konkurrenz, Feinddruck), die sich von Region zu Region unterscheiden können.

Modelle die für ein und die selbe Art, aber in verschiedenen Gebieten, oder mit unterschiedlicher Auflösung entwickelt wurden, sind somit u.U. schwer vergleichbar. Die Vorhersage-Fähigkeit dieser (statischen!) Verbreitungs-Modelle ist daher häufig gering, da sie sich allein auf Habitatparametern als Prädiktoren beschränken und Prozesse (→process) wie eben Konkurrenz, Migration, Feuerdynamik u.a.m. vernachlässigen (Austin, 2002b). Dadurch erklärt sich auch der bescheidene Erfolg, wenn sie auf andere Gebiete übertragen werden (Guisan et al., 2002).

### 4.3.3 Response-Verlauf

Vor der Modellwahl sollte man sich Gedanken machen über die Gestalt der →Response einer Art auf eine bestimmte Umweltvariable (Gradient), z.B. Verteilung bezüglich Seehöhe. In der ökologischen Literatur werden mehrere theoretische Verläufe beschrieben: von der symmetrischen Gauß'schen Glockenkurve über schiefe, unimodale bis hin zu bimodaler Gestalt. Als gemeinsamen Nenner kann man jedenfalls die kurvilineare Gestalt erkennen, wie sie ja von der Nischentheorie postuliert wird (Austin und Meyers, 1996). Somit scheiden Modelle, die

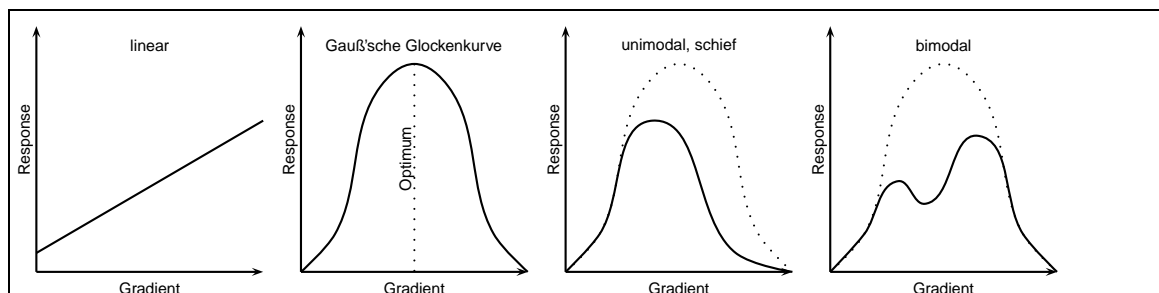


Abbildung 4.2: lineare und kurvilineare Response

nur lineare Zusammenhänge beschreiben können von vornherein aus.

Zum Beispiel ist die →PCA zur Analyse der meisten ökologisch relevanten Daten nicht geeignet. Wendet man dieses Verfahren auf kurvilineare Zusammenhänge an, führt dies zu starken Verzerrungen. Deshalb geht Austin (1999) hart ins Gericht mit diesem Ansatz: "*the linear assumption of PCA is incompatible with the niche theory.*"

Aber auch die Idealvorstellung einer Gauß'schen Glockenkurve muß in Frage gestellt werden. Rezentere Studien, so Austin (1999), zeigen, daß diese Annahme nur in wenigen Fällen angemessen ist. Der Einfluß anderer Arten (Konkurrenten, Räuber) kann nämlich eine Art aus ihrem Optimum verdrängen (siehe obige Diskussion, *fundamental vs. realized niche* 4.3.2). Dadurch entstehen schiefe, unimodale aber auch bimodale Verläufe (siehe Abb. 4.2).

#### 4.3.4 Kausale vs. statistische Zusammenhänge

Die Tatsache, daß zwischen zwei Variablen ein statistischer Zusammenhang nachweisbar gegeben ist, bedeutet noch lange nicht, daß zwischen den Beobachtungen auch ein kausaler Zusammenhang besteht.

Ein typisches Fettnäpfchen im Zusammenhang mit (statischen) Habitat (-Bewertungs-)Modellen ist Dichte als Indikator für Habitatqualität. Die Annahme, daß dort wo sich die meisten Individuen einer Art aufhalten, die bestmöglichen Bedingungen herrschen, ist durchaus nicht immer gültig. Van Horne (1983) beschreibt 3 solcher Situationen:

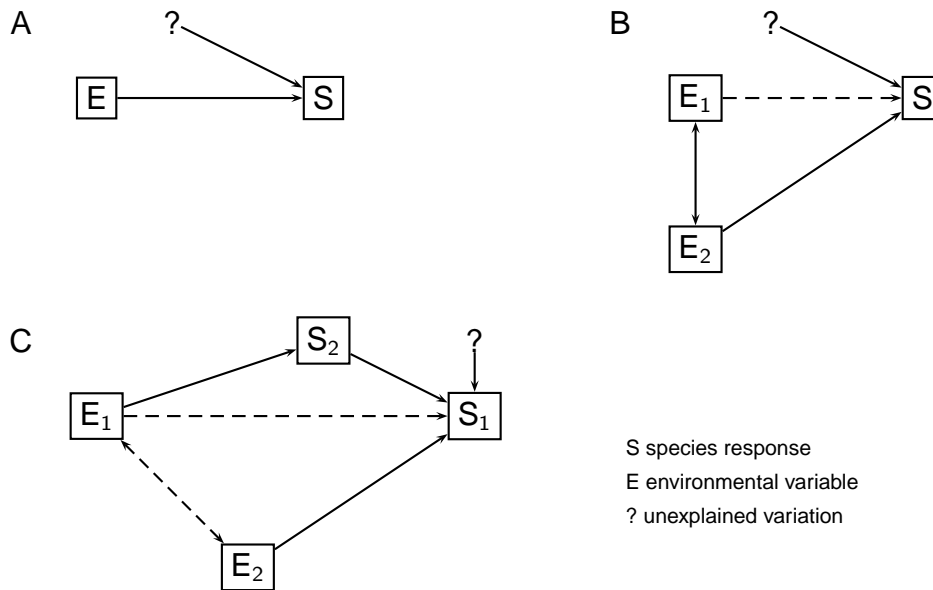
- jahreszeitlich unterschiedliche Verfügbarkeit von Lebensraum: Ist z.B. ausreichend Winterhabitat vorhanden, aber zu wenig Sommerhabitat, so spiegelt die beobachtete Dichte (Erhebungen werden meist in den wärmeren Monaten durchgeführt) nicht die Tragfähigkeit wider.
- über längere Zeiträume (Jahre) schwankendes Nahrungsangebot: Die beobachtete Dichte ist nicht das Ergebnis der aktuellen Bedingungen, sondern hat weiter zurückliegende Ursachen.
- *source-sink*-Situation: Während wenige dominante Tiere das geeignete Habitat besetzen, müssen subadulte Individuen in das umliegende ungünstigere Habitat ausweichen. Solche Situationen sind daran erkennbar, daß im *sink* zwar hohe Dichte beobachtbar ist, aber auch gesteigerte Mortalität und geringere Fertilität. Der *sink* muß also aus dem geeigneten Habitat, *source*,



versorgt werden.

Ähnlich kann Habitatverlust – sei es durch anthropogenen Einfluß oder eine Katastrophe – dazu führen, daß der verbleibende Lebensraum stärker beansprucht wird (Van Horne, 2002).

Morrison et al. (1998) illustrieren anhand einer Abbildung Annahmen und tatsäch-



(Morrison et al., 1998, verändert)

Abbildung 4.3: *Of Correlates and Causes*

liche Zusammenhänge zwischen Wildtieren und deren Habitat. Selbst bei einem einfachen, tatsächlich kausalen Zusammenhang (durchgehender Pfeil) zwischen einem Umweltfaktor (E) und einer Reaktion (S) einer Spezies bleibt ein Rest an unerklärter Varianz (?) (siehe Abb. 4.3A). Der Grund liegt in eventuellen Meß- und Verfahrensfehlern, aber auch in der Auswirkung unberücksichtigter Faktoren. Häufig mißt man z.B. eine Umweltvariable (E<sub>1</sub>) im Glauben, dadurch eine biologische Response (S) beschreiben zu können, während der eigentliche ursächliche Faktor (E<sub>2</sub>) unberücksichtigt bleibt (siehe Abb. 4.3B). Zwischen den beiden Faktoren (E<sub>1</sub> und E<sub>2</sub>) besteht zwar ein statistischer Zusammenhang, aber nicht unbedingt ein direkt kausaler. In solch einem Fall ist es irrig davon auszugehen, daß Beeinflussung der Variable E<sub>1</sub> eine Veränderung in der Reaktion S zur Folge haben muß.

Freilich sind noch komplexere Zusammenhänge möglich (siehe Abb. 4.3C). Zum

Beispiel könnte eine zweite Art (S2), etwa ein Konkurrent, Räuber oder Symbiont, Auswirkungen auf die untersuchte Spezies (S1) haben. Die beiden Arten "teilen" sich zwar den selben Umweltparameter E1, der in einer statistischen, aber keineswegs ursächlichen Beziehung zur Spezies S1 steht, aber die tatsächliche Kausalkette verläuft über S2.

## 4.4 Einfließende Variablen

Ob man nun von (ökogeographischen) Variablen, Faktoren, Parametern, Komponenten oder Gradienten, von Response oder Zielgröße spricht, einige Überlegungen sind – tunlichst bevor man sich auf ein bestimmtes Modell festlegt bzw. unabhängig davon – anzustellen.

### 4.4.1 Zielgröße

Beobachtungen für das Vorkommen bzw. die Abundanz der untersuchten Tierart stammen entweder aus direkten Nachweisen (Sichtung, Fang und/oder telemetrische Ortung, Gesang) oder aus indirektem Nachweis (Kot, Risse, Federn, Haare, Spuren, Huderstellen).

Die meisten statistischen Ansätze sind während ihrer Anpassung sowohl auf *presence* als auch auf *absence*-Daten angewiesen, also auf Angaben über nachweisliches Vorkommen bzw. gesichertes Nichtvorkommen der untersuchten Tierart an einem gegebenem Standort.

Die Aufnahme der *presence/absence*-Daten ist ein kritischer Schritt im Modellierungs-Prozeß (Hirzel et al., 2002). Insbesondere *absence*-Daten sind schwer mit der gebotenen Genauigkeit zu erheben.

Ein gegebener Stichprobenpunkt kann aus mehreren Gründen – eben auch fälschlicherweise – dem *absence*-Satz zugeordnet werden:

- (1) die Tierart konnte nicht beobachtet werden, obwohl sie vorkommt (wiederholtes Aufsuchen notwendig),
- (2) die Tierart kommt aus historischen Gründen nicht vor, obwohl das Habitat geeignet wäre (z.B. Ausrottung), oder
- (3) das Habitat ist tatsächlich ungeeignet.

Nur der letzte Fall ist für die Vorhersage relevant. Falsche *absences* können die Aussage eines Modells beträchtlich verzerren.

Einer der wenigen Ansätze, die nur mit nachgewiesenem Vorkommen, also reinen *presence*-Daten umgehen können, ist ENFA (siehe Abschnitt 5.6).

#### 4.4.2 Habitatvariablen

Als Variablen zur Beschreibung eines Lebensraumes kommen sämtliche abiotischen (Klima, Geologie, Gelände, Wasserhaushalt, ...) und biotischen (Vegetation, Fauna, ...) Faktoren in Frage. Typische, in der Literatur verwendete Habitatvariablen sind etwa verschiedene Kennzahlen zur Vegetation (Typ, Sukzessionsstadium, Kronenschluß, ...), Seehöhe, Neigung, Exposition. Aber es werden auch andere Parameter berücksichtigt, wie z.B. Beeinträchtigungen durch *Katastrophen* (Feuer) und anthropogene Einflüsse (Bevölkerungsdichte, Nähe zu Siedlung, Dichte des Straßennetzes, Erholungsaktivitäten). In manchen Modellen wird sogar der Einfluß anderer Arten miteinbezogen, z.B. Konkurrenz um Nistplätze. Bei Modellen für Fleischfresser ist natürlich die Verfügbarkeit von Beutetieren ein wichtiger Aspekt.

In Morrison et al. (1998, S. 361) findet man in Form einer Tabelle (*a hierachic classification of key environmental correlates for wildlife species*), die sich über mehrere Seiten erstreckt, eine ausführliche und strukturierte Übersicht.

Als Beispiel sind in Tabelle 4.1 die von Storch (2002) zur Erstellung eines HSI-Modells für das Auerwild aufgenommenen Parameter angeführt.

Tabelle 4.1: Variablen für ein HSI-Modell für Auerwild

Variable	Definition
Höhe über Tal bzw. Grenze Wald/ landwirtschaftl. genutzte Fläche	1 (<100m), 2 (100 bis <300m), 3 (300 bis <500), 4 (500 bis <700), ...
Hangneigung	Mittelwert [°]
Sukzessionsstadium	Schlag/Jungwuchs, Dickung, Stangenholz, Baumholz, Altholz, Plenterung, unbestockt
Kronenschluß	bedeckter Anteil des Waldbodens
Lücken im Kronendach	breiter als Bestandeshöhe; ja/nein
Bestandestyp	Nadelwald; Mischwald; Laubwald (Kodierung nach Forsteinrichtung)
Vorkommen Nahrungsbäume	Tanne oder Kiefer (excl. Legföhre); ja/nein

Fortsetzung nächste Seite

Variable	Definition
Bodendeckung Verjüngung	bedeckter Anteil des Waldbodens
Bodenbedeckung Heidelbeere	bedeckter Anteil des Waldbodens
Höhe Bodenvegetation	mittlere Höhe der dominanten Schicht [cm]
Hinweis auf Nutzung durch Auerw.	Spuren, Federn, Losung, Huderstellen; ja/nein

(Storch, 2002, verändert)

Austin (2002a) unterscheidet zwischen indirekten (komplexen) Gradienten, direkten (Regulator-)Gradienten und Ressourcen-Gradienten. Komplexe Gradienten, wie z.B. Seehöhe, haben keinen direkten Einfluß auf eine Spezies. Sie sind insofern von Bedeutung, als daß sie mit Faktoren, wie z.B. Niederschlag, korrelieren. Regulator-Gradienten, wie z.B. Temperatur, üben zwar direkten Einfluß aus, werden aber nicht konsumiert. Ressource-Gradienten, wie z.B. Vegetation, werden direkt konsumiert.

Das sind allerdings keine ausschließlichen Kategorien: Wasser kann in Form von Niederschlag als indirekter Gradient betrachtet werden, in Form einer Tränke stellt es einen Ressource-Gradienten dar.

Weiters kann zwischen proximalen und distalen Prädiktoren unterschieden werden: Proximal und distal bezieht sich auf die Position der Prädiktoren in der Kausalkette, die sie mit der untersuchten Spezies verbindet.

Modelle, die auf proximalen Ressource- und direkten Gradienten basieren, sind robuster und weiträumiger anwendbar.

Ein einschränkendes Kriterium bei der Wahl der Habitatvariablen die schließlich ins Modell aufgenommen werden sollen, ist natürlich die Verfügbarkeit dieser Daten in angemessener Form (digital) und Genauigkeit.

#### 4.4.3 Einfluß der Tierarten auf ihr Habitat

Üblicherweise wird in einem Modell nur der Einfluß der Umwelt auf eine Art/Gesellschaft nachgebildet. Eigentlich müßte auch der gegenläufige Prozess Eingang finden. Bestimmte Arten nehmen einen starken modifizierenden Einfluß auf ihr Habitat, beeinflussen die Dynamik der Energie- und Nährstoffflüsse und somit die Habitat-Verfügbarkeit (für sich selbst und/oder andere Arten). Als Beispiel nennt Van Horne (2002) den Biber. Ein Modell, in welchem dieses "feedback" Berücksichtigung findet, ist FORSPACE (Kramer et al., 2001).

#### 4.4.4 Datenquellen

An dieser Stelle sei noch einmal (Tipton, 1980) erwähnt: Das Modell soll sich nicht nach den Daten richten, sondern umgekehrt.

Als Quellen für Habitatvariablen kommen in Frage: eigens durchgeführte Erhebungen, GIS, DEM (*digital elevation model*) und DTM (*digital topographic model*), "traditionelle" Kartenwerke, Daten aus vorangegangenen Studien, Fernerkundungsdaten aus Satellitenbildern und Luftbildern (LANDSAT, CORINE), zusätzliche terrestrische Ergänzungserhebungen, Daten aus Forstinventuren u.a.m.

Bei der Wahl der Datenquellen stellt sich einerseits die Kostenfrage, andererseits besteht die Gefahr, es sich allzu "bequem" zu machen. Es ist natürlich wesentlich einfacher und bestechender, mit Karten und Computern zu arbeiten, als eigene (personal- und zeitaufwendige) Erhebungen im Feld durchzuführen. Allerdings muß man sich bei der Verwendung speziell digitaler Datenbestände u.U. gewissen Zwängen beugen (Auflösung, siehe unten). Zudem ist es ratsam, diese Daten vor Ort zu verifizieren bzw. Fernerkundungsdaten terrestrisch zu ergänzen.

Ein Problem bei der Verwendung thematischer Karten (erstellt aus Satellitenbildern und Flugzeugaufnahmen) ist, daß die Interpretation aus der Sicht der Landnutzung erfolgt, ökologische Aspekte bleiben außen vor (Guisan et al., 2000).

Van Horne (2002) bringt bezüglich der Verwendung von Daten aus bestehenden GIS einige Warnungen an. So ergibt sich aus der Vielfalt der sich anbietenden Variablen das Problem der Auswahl. Dabei ist es praktisch so gut wie unmöglich, durch eine "objektive" Analyse (beruhend auf Erfahrung, Literatur, Expertenurteil) die wichtigen Variablen auszuwählen. Eine a priori Selektion beeinträchtigt zudem den späteren analytischen Prozess.

Ein weiteres Problem ergibt sich aus der räumlichen Auflösung der GIS-Daten: damit diktiert das GIS praktisch den Maßstab. Die angemessene Pixel-Größe hängt davon ab, ob man auf Individuum- oder Populationsebene operiert. Im ersten Fall sollte sie in der Größenordnung der kleinsten Einheit sein, welche die Habitatwahl beeinflusst, im zweiten Fall in der Ordnung eines (jahreszeitlich angemessenen) Territoriums liegen.

#### 4.4.5 Auswahl der Variablen

In das endgültige Modell sollten nur jene Variablen bzw. Terme einfließen, die tatsächlich von Bedeutung sind. Zum einen aus Gründen der *parsimony*, zum an-

deren, um die Vorhersagefähigkeit des Modells nicht zu beeinträchtigen.

Bei Ansätzen wie z.B. HSI-Modellen ist man dabei auf eigene Erfahrungen, Plausibilitäts-Überlegungen oder Vergleich von Expertenmeinungen in der Literatur angewiesen.

Für rein statistische Ansätze gibt es Verfahren, anhand derer – mehr oder weniger automatisiert – ein "statistisch gesicherter" bzw. signifikanter Einfluß auf die Zielgröße bestimmt wird.

Guisan et al. (2002) erwähnen im Zusammenhang mit Variablen-Selektion den F-Test (für  $\rightarrow$ LS Regression),  $\chi^2$ -Test und Ansätze zur Reduktion des  $\rightarrow$ AIC (für GAM und GLM).

Bei Regressionsmodellen (einschließlich GAM und GLM) kommen u.a. sogenannte schrittweise-Verfahren in Betracht:

- schrittweise rückwärts: Man beginnt mit einem "vollen" Modell, d.h. alle in Frage kommenden Variablen werden ins Modell aufgenommen. Die jeweils "relativ unwichtigste" Variable wird mittels t-Test für den Koeffizienten (mit z.B.  $P < 0.05$ ) bestimmt und aus dem Modell entfernt.
- schrittweise vorwärts: Man beginnt mit einem "leeren" Modell, die jeweils "wichtigste" Variable wird ins Modell aufgenommen.
- Kombination der obigen Verfahren: Nachdem eine im Rückwärts-Verfahren als "unwichtig" erkannte Variable eliminiert worden ist, wird im Vorwärts-Verfahren kontrolliert, ob eine bereits ausgeschiedene Variable wieder ins Modell aufgenommen werden soll bzw. umgekehrt.

Guisan et al. (2002) sind dem Einsatz von *stepwise procedures* gegenüber skeptisch eingestellt. Einerseits kann es bei  $\rightarrow$ Kollinearität innerhalb der Prädiktoren dazu kommen, daß zwei korrelierende Variablen beide als nicht signifikant ausscheiden, obwohl jede für sich betrachtet einen bedeutenden Anteil der Varianz erklären könnte. Andererseits können bereits geringe "Störungen" der Response-Variable zu beträchtlich unterschiedlichen Subsets von Prädiktoren führen. Sie schlagen als mehrverheißende Alternativen sogenannte *shrinkage rules* vor, nämlich "*ridge regression*" und "*lasso*".

Bei erstgenanntem Verfahren werden alle Terme im Modell belassen, aber deren Koeffizienten werden – unter Verwendung eines quadratischen Strafterms – gegen 0 "geschrumpft". Dabei gilt es, einen Kompromiss zu finden zwischen der durch das Verfahren reduzierten Varianz einerseits und der bewirkten Verzerrung

andererseits.

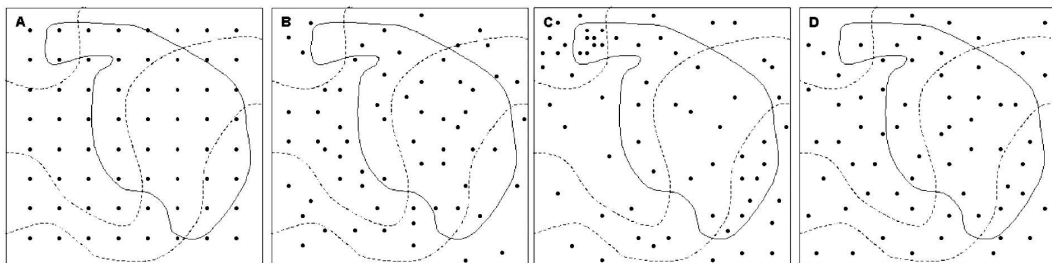
Bei *lasso* wird eine Grenze für die Summe der Absolutwerte der Koeffizienten festgelegt. Auch hierbei werden die Koeffizienten gegen 0 geschrumpft, mit dem Unterschied, daß einige exakt auf 0 gesetzt werden und damit die Variable aus dem Modell genommen wird. Hier muß sozusagen ein Kompromiss zwischen der Wahl eines Variablen-Subsets und einer *ridge regression* gefunden werden.

Van Horne (2002) erwähnt AIC sowohl als Verfahren zur Auswahl der wichtigen Variablen, als auch zur Auswahl des am besten geeigneten Modells aus einer Reihe möglicher Modelle, die nicht notwendigerweise auf dem selben Ansatz beruhen. Dadurch ermöglicht das AIC den Vergleich von wissens- und korrelationsbasierten Ansätzen und es werden – laut Autorin "künstliche" – Standards wie  $P < 0,05$  vermieden.

#### 4.4.6 Sampling Design

Hirzel und Guisan (2002) vergleichen die Auswirkung unterschiedlicher Strategien bei der Stichproben-"Nahme" sowie des Stichproben-Umfangs auf die Vorhersage-Fähigkeit von Habitataignungs-Modellen (hier ein GLM).

Die vier verglichenen Stichprobenverfahren sind:



(Hirzel und Guisan, 2002)

Abbildung 4.4: Vier Samplingstrategien

- *regular sampling*: Stichprobenpunkte liegen auf den Knoten eines regelmäßigen Gitternetzes (Abb. 4.4A);
- *random sampling*: Stichprobenpunkte werden zufällig im Erhebungsgebiet angeordnet (Abb. 4.4B);
- *equal-stratified sampling*: das Untersuchungsgebiet wird zunächst in verschiedene Strata aufgeteilt, danach wird in jedem Stratum die selbe Anzahl an Stichprobenpunkten in zufälliger Anordnung platziert (Abb. 4.4C);

- *proportional-stratified sampling*: wie oben, aber die Stichprobenpunkte werden im Verhältnis zum Flächenanteil des jeweiligen Stratum an der Gesamtfläche des Erhebungsgebietes zufällig angeordnet (Abb. 4.4D).

Der Stichprobenumfang hatte den stärksten Einfluß: größere Stichproben lieferten genauere Vorhersagen, sowohl was die Habitat-Eignung als auch das Vorkommen/Nicht-Vorkommen anbelangt.

Die Rangordnung der Verfahren von bester zu schlechtester Strategie war bei der Bewertung der Habitateignung *equal-stratified* gefolgt von *regular*, *random* und *proportional-stratified*. Bei der Vorhersage Vorkommen/Nicht-Vorkommen war die Reihenfolge *regular*, *equal-stratified*, *proportional-stratified* und *random*.

Da die Erhebung der Daten ein sowohl zeit- als auch geldaufwendiges Unterfangen darstellt, sollte man den jeweiligen Anforderungen Rechnung tragen. In diesem Sinne kommen die Autoren zum Schluß, daß das Stratifizieren der Stichproben – was ja zusätzlicher Information bedarf – die Modell-Genauigkeit nicht wesentlich steigert, und daß *regular samplings* einfacher (und kostengünstiger) zu implementieren sind.

Austin (2002a) beschreibt ein *gradsect* oder auch *SR<sup>3</sup>-strategy* (*Stratification Representation Replication Randomization*) genanntes Verfahren, das eine repräsentative Stichprobe des Untersuchungsgebietes liefern soll. Dabei werden Transekte so durch das Gebiet gelegt, daß sie die wesentlichen Umweltgradienten der Region kreuzen und zwar derart, daß die Zeit, die zum Aufsuchen der sich ergebenden Punkte notwendig ist, minimiert und Zugangs-Probleme vermieden werden. Ein kosteneffizientes Verfahren also, das den oben genannten *random-stratified*-Strategien sehr nahe kommt (Hirzel und Guisan, 2002).



## 5. Modellansätze

### 5.1 Versuch einer Gliederung

Ansätze gibt es ebensoviele wie Modelle selbst, es ist daher nicht einfach, sie in Klassen einzuordnen. Je nach angelegten Kriterien gelangen verschiedene Autoren zu unterschiedlichen Einteilungen. Ein und das selbe Modell scheint eventuell mehrmals bzw. in unterschiedlichen Kategorien auf.

Keine der gefundenen Aufstellungen enthält alle bekannten Modelle, es erhebt auch keiner der Autoren Anspruch auf Vollständigkeit.

Jeffers (1988) ordnet im *Practitioner's Handbook on the Modelling of Dynamic Change in Ecosystems* mathematische Modelle in folgende Familien ein: *functional relationship* (Populations- und Räuber-Beute-Modelle), *matrix models* (z.B. Leslie-Matrix, Markovs Modell), *statistical models* (→ANOVA, →MLR), *multivariate models* (→Ordination, *Discrimination* (→DA, →Mahalanobis distance), →Classification, →Clusteranalyse), *mathematical programming, game theory models*, und *catastrophe theory*.

Harvey (1996) spricht von "Techniken der Gradientenanalyse" und zählt mehrere (statistische) Ansätze auf (→MR, →IR, →PCA, →RDA, →COR, →WAE, →GLM, →ML, →WAI, →CA, →DCA, →CCA und →DCCA). Je nach →Response (linear oder unimodal) sowie Anzahl der abhängigen/unabhängigen Variablen ordnet er die Modelle in die Kategorien →Regression, →Calibration, →Ordination und →Canonical Ordination ein.

Lek und Guégan (1999) unterscheiden mathematische (analytische), numerische und statistische Methoden, sowie Techniken die auf künstlicher Intelligenz (→AI) beruhen: Expertensysteme (→XPS), genetische Algorithmen (→GA) und neuronale Netzwerke (→ANN).

Morrison et al. (1998) liefern eine Klassifizierung von *wildlife-habitat relationship*-Modellen, primär gegliedert nach den Schlüsseln Forschung- und Management-

Interessen einerseits und ausschließlich Management-Interessen andererseits (siehe Tabelle 5.1).

Dies dürfte eine der ausführlichsten Einteilungen sein. Allerdings werden Modelle angeführt, auf die im Rahmen dieser Arbeit nicht näher eingegangen wird und es fehlen rezente Ansätze.

Tabelle 5.1: Klassifizierung von Habitatmodellen

Models of Research and Management Interest
Models of vegetation structure
Stand growth models
Succession models
Models of species response to vegetation structure
Single-species models
Life history models
Habitat preference models
Optimal foraging models
Correlation models
Multivariate statistical models
Habitat suitability index (HSI) models
Habitat capability (HC) models
Habitat evaluation procedures (HEP)
Bayesian and pattern recognition (PATREC) models
Multiple species models
Coarse-filter and fine-filter models
Species-habitat matrices
Guild and lifeform models
Community structure models
Community and ecosystem simulation models
Gap analysis models
Biodiversity models
Hierarchy models
Models of population dynamics and viability
Models of landscapes
Vegetation disturbance models

Fragmentation models  
 Cumulative effect assessments  
 Models of insular biogeography  
 Models of ecosystems  
 Models of nutrient cycling and energy flow  
 Modeling historic range of natural conditions

---

Models of Management Interest

---

Models of indicator species  
 Models for monitoring species and habitats  
 Knowledge-based and decision-aiding models  
 Decision support models  
 Expert systems

---

(Morrison et al., 1998)

Guisan und Zimmermann (2000) gruppieren die wichtigsten (statistischen) Ansätze in 7 Kategorien: Multiple Regression und deren verallgemeinerte Form (klassische →LS →Regression, →GLM, →GAM) , →Classification *techniques* (→CART, *decision trees* und *rulebased classification*), *Environmental envelopes* (→BIOCLIM, →HABITAT, →DOMAIN), →Ordination *techniques* (→CCA, →RDA), →Bayesian approaches, Neuronale Netzwerke (→ANN) und "andere Ansätze" (GIS-basierte Modelle, →DFA, →ENFA).

Ausgehend von der Habitatdefinition durch Morrison und Hall (2002) (siehe Kapitel 4.1.1) teilen Guisan et al. (2000) Habitat(-Verteilungs-)Modelle grob in 2 Ansätze ein:

- Art-spezifischer Ansatz (*species-specific approach*): Alle potentiellen Habitate einer Tierart werden aus biotischen (z.B. Vegetationstypen) und abiotischen (z.B. physikalischen) Deskriptoren als die gesamte potentielle Habitat-Verteilung abgeleitet. Dieser Ansatz erfordert im Allgemeinen gesicherte Angaben zur *absence* der untersuchten Tierart (siehe S. 19). Als Ergebnis erhält man eine einzelne Karte, auf der zwischen zwei (oder mehreren) für die untersuchte Tierart gleichermaßen geeigneten Habitaten nicht unterschieden werden kann.
- Habitat-spezifischer Ansatz (*habitat-specific approach*): Habitat-Typen werden gemäß einer vordefinierten Typologie, unabhängig von den ökologi-

schen Anforderungen einer bestimmten Art, ausgeschieden. Die Habitatklassifizierung selbst stellt einen integrierten Prädiktor für Vegetationsstruktur, Topographie, Geomorphologie, Landnutzungsform usw. dar. Zwei verschiedene Habitate, die für eine gegebene Tierart gleich geeignet sind, scheinen auf den Habitatkarten als unterschiedliche Einheiten auf. Für diesen Ansatz werden keine gesicherten *absence*-Daten benötigt.

Die meisten Beispiele für den artspezifischen Ansatz greifen auf statisch-komparative Modelle zurück. Dabei werden Beobachtungen an einer Tierart, durch Anpassen eines statistischen Zusammenhanges mit einer Kombination von Umweltfaktoren in Beziehung gesetzt. Dies unter der Annahme, daß sich die betreffende Tierart mit ihrer Umwelt im "Gleichgewicht" befindet (siehe 4.3.1). Die geographische Verbreitung einer Spezies soll also durch jene Umweltvariablen erklärt werden, welche die "realisierte" Nische am besten definieren (siehe 4.3.2). Dieser Ansatz impliziert, daß so viele Modelle entwickelt (und getestet) werden müssen, wie es Tierarten zu modellieren gilt. Viele statistische Techniken können zur Modellierung solcher Verbreitungen herangezogen werden, z.B. GLMs oder GAMs, *decision trees*, *Bayesian models*, *environmental envelopes* (→HABITAT und →BIOCLIM) oder Modelle auf der Grundlage von →CCA.

Während es für den artspezifischen Ansatz in der Literatur sehr viele Beispiele gibt, sind sie für den biotop-spezifischen Ansatz dünn gesät. Das hat mehrere Gründe:

- Es muß eine (eindeutige) Standard-Habitateinheit definiert werden.
- Für eine erfolgreiche Vorhersage müssen Habitateinheiten auf der Basis verschiedener räumlich expliziter (→spatially explicit) Umwelt-Deskriptoren (topographische, geologische, edaphische, hydrologische, klimatische, ... Daten) definiert sein.
- Die benötigte Information stammt aus verschiedensten Quellen (andere Modelle, Fernerkundung oder Erhebungen) und muß somit erst transformiert, klassifiziert und standardisiert werden, bevor sie zu einem GIS kombiniert werden kann.

Die wenigen Beispiele berücksichtigen nur eine begrenzte Anzahl von Habitattypen, die wiederum nur für eine geringe Anzahl von Spezies interessant sind. Guisan et al. (2000) können kein Beispiel anführen, das alle bekannten Habitatklassen im jeweiligen Untersuchungsgebiet berücksichtigt, und somit für

die gesamte darin vorkommende Fauna anwendbar wäre (z.B. GAP, siehe 5.12).

Die meisten (multivariaten) Ansätze beruhen auf ähnlichen Prinzipien (Hirzel et al., 2002):

- Das Untersuchungsgebiet wird – zumindest theoretisch – als eine Rasterkarte modelliert, die sich aus N anliegenden, isometrischen Zellen zusammensetzt.
- Die "abhängige" Variable liegt meistens in binärer Form – Vorkommen/Nichtvorkommen *presence/absence* der untersuchten Tierart – vor.
- Die "unabhängigen" ökogeographischen Variablen (*ecogeographical variables*, EGV) beschreiben Eigenschaften der Zellen.
- Eine Funktion der EGV wird angepaßt, um die diversen Zellen als geeignetes bzw. ungeeignetes Habitat zu klassifizieren.

Modelle bzw. Ansätze die hier erwähnt wurden, auf die aber in diesem Kapitel nicht weiter eingegangen wird, sind im Glossar (Seite 90) kurz beschrieben.

## 5.2 HSI-Modelle und HEP

In den USA gehören HSI-Modelle (*Habitat Suitability Index*) wie das HEP (*Habitat Evaluation Procedures*) als Entscheidungsinstrument in Landnutzungsfragen seit den '70er Jahren zum Repertoire des Naturschutzes. Die Vorgangsweise bei der Erstellung solcher Modelle ist in den Teilen 101, 102 und 103 des Ecological Services Manual des U.S. Fish and Wildlife Service beschrieben (U.S. Fish and Wildlife Service, 1980a,b, 1981).

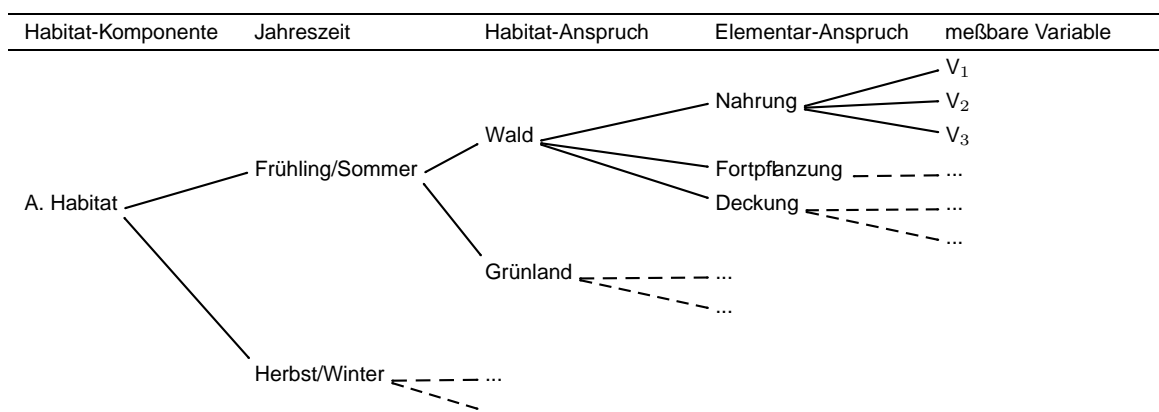
Van Horne (2002) bezeichnet HEP und HSI als wissensbasierte Ansätze. Sie werden also aus dem Erfahrungspool von Experten – die mit der jeweiligen Zielart vertraut sind – abgeleitet. Es wird von diesen Modellen nicht erwartet, daß sie – obwohl sie durchaus mit Korrelationen arbeiten – kausale Zusammenhänge herausarbeiten können.

**HSI-Modelle** Der HSI ist definiert als numerischer Index für das Vermögen eines gegebenen Lebensraums einer Zielart bestimmte Bedingungen zu bieten. Bei der Bewertung – auf Basis verschiedener Habitat-Variablen – vergleicht man die vorgefundenen Habitatbedingungen mit den für die jeweilige Tierart optimalen

Lebensbedingungen, bildet also – zumindest theoretisch – einen Quotienten. Bei optimalen Bedingungen entspricht der HSI dem Wert 1,0 und steht damit für die höchste potentielle Dichte bzw. die maximale Tragfähigkeit (*carrying capacity*) für sie Zieltierart. Es ist jedoch zu beachten, daß abgesehen von der Tragfähigkeit eines Habitats, auch noch andere Faktoren – wie Feinddruck, (zwischenartliche) Konkurrenz, anthropogene Störungen – die Populationsdichte beeinträchtigen. Bei ungeeigneten Bedingungen – die das Vorkommen einer Art nicht zulassen – wird er mit 0,0 festgelegt. Im Idealfall steht der HSI in direktem linearem Zusammenhang mit der Tragfähigkeit des Habitats (dies ist auch eine notwendige Bedingung für dessen Verwendung in HEP).

Ein HSI-Modell entsteht in 5 Phasen, die sich teilweise überlappen.

- 1.) **Zielformulierung:** Definieren von idealem und akzeptablem Output (Werte von 0 bis 1,0, direkter linearer Zusammenhang mit Tragfähigkeit); Festlegen der geographischen Anwendbarkeit; Bestimmen der Anwendbarkeit auf Jahreszeiten und Lebensabschnitte.
- 2.) **Identifizierung der Modell-Variablen:** Welche Umweltfaktoren würden, falls verändert, die Tragfähigkeit des Habitats beeinträchtigen?; Kandidaten sind alle meßbaren physikalischen, chemischen und biologischen Faktoren; Gliederung der Strukturvariablen als Baumdiagramm mit z.B den Ebenen Saison (*seasonal habitat*), Elementaransprüche (*life requisites*, Nahrung, Fortpflanzung, Deckung), Habitatanspruch (*cover type*), ev. Lebensabschnitt (*life stage*), meßbare Variable.



(U.S. Fish and Wildlife Service, 1981, verändert)

Abbildung 5.1: Exemplarisches Baumdiagramm für terrestrisches Habitat

- 3.) **Strukturierung des Modells:** die Variablen werden mittels verbalen (beschreibenden), mechanistischen, Muster-Erkennungs- oder Bayes-Wahrscheinlichkeits-Modellen ( $\rightarrow$  Satz von Bayes) miteinander verknüpft; bei mechanistischen Modellen unterscheidet man limitierende ( $I = \min(I_1, \dots, I_n)$ ), kumulative ( $I = \sum_{i=1}^n I_i$ ) und kompensatorische (arithmetisches  $I = \frac{\sum_{i=1}^n I_i}{n}$  oder geometrisches  $I = \sqrt[n]{\prod_{i=1}^n I_i}$  Mittel) Verknüpfung.
- 4.) **Dokumentation:** überlappt sich mit allen anderen Phasen; Modell soll auch für Außenstehende nachvollziehbar und anwendbar sein; wichtig zum Abschätzen der Übertragbarkeit; wesentlich für die Weiterentwicklung.
- 5.) **Verifizierung:** Überprüfung durch Autor (Ist Output akzeptabel?); Test mit Stichprobendaten; Überprüfung durch Experten; Verifizieren mit Geländedaten.

HSI-Modelle liefern also nicht nur den HSI als Index für die Habitat-Güte, sondern dokumentieren – zwangsweise – den gesamten Aufbau des Modells selbst.

Sie sind zwar dazu geeignet, die wesentlichen Umwelt-Faktoren – von denen angenommen wird, daß sie Vorkommen und Häufigkeit einer Wildtierart am stärksten beeinflussen – in einfacher und nachvollziehbarer Form darzustellen. Aber sie treffen keine Aussagen über kausale Zusammenhänge und bieten keine Information über Populations-Größe und -Entwicklung. Schamberger (zitiert in Morrison et al. (1998)) betrachtet diesen Ansatz deshalb eher als eine Hypothese zur Art-Habitat-Beziehung, denn als Beschreibung dem System zugrundeliegender kausaler Zusammenhänge.

**HEP** HEP ist ein Instrument zur Bewertung (abzusehender) Auswirkungen menschlicher Eingriffe auf die Umwelt, somit zur Abwägung von Alternativen und Projektplanung. Dieser Ansatz wurde aus dem Bedürfnis heraus entwickelt, den nicht-monetären Wert von Fisch- und Wildtier-Ressourcen zu dokumentieren. HEP gründet auf der Annahme, daß Habitat-Qualität und -Quantität numerisch beschrieben werden können. Diese numerische Darstellung soll es erlauben, Optionen und Alternativen in der Landnutzung zu vergleichen.

Die Habitatqualität einer Zielart wird in einem Index festgehalten, dem HSI. In der ersten Auflage des Modells war die Habitatqualität subjektiv begründet auf dem Urteil einer Expertengruppe. Der HSI wurde erst 1980 mit der neuen HEP-Version eingeführt.

Die HSI-Werte werden mit der für die jeweilige Ziel-Tierart verfügbaren Habitat-

Fläche multipliziert. Dadurch erhält man die sogenannten HUs (*Habitat Units*), das Kernstück der HEP. Diese Einheiten bilden die Basis für die Vergleiche, räumlicher oder zeitlicher Natur, und damit für das Abwägen von Alternativen bzw. Abschätzen von Auswirkungen anthropogener Eingriffe. Dabei ist es wichtig, daß der Verlust einer HU – sei es nun an Fläche oder Qualität – durch eine anderweitig gewonnene Einheit wieder wettgemacht werden kann. Damit erklärt sich die Bedingung, daß zwischen HSI und Tragfähigkeit des Habitats eine lineare Beziehung bestehen muß.

HEP ist ein ideales Werkzeug zur Dokumentation und Veranschaulichung der Auswirkungen geplanter Maßnahmen, auf dessen Grundlage Ressourcen-Manager fundierte Entscheidungen treffen können.

### 5.3 Verallgemeinertes lineares Modell (GLM)

GLMs (*Generalized Linear Models*) sind mathematische Erweiterungen bzw. Verallgemeinerungen linearer Modelle. Daten werden nicht in ein "unnatürliches Skalen-Korsett" gezwungen, Nichtlinearität sowie heterogene Varianzen sind erlaubt. Sie basieren auf einer angenommenen Beziehung – der sogenannten *link*-Funktion – zwischen dem Mittelwert der Response-Variable und der  $\rightarrow$ Linearkombination der erklärenden Variable(n).

Die Daten können verschiedensten Verteilungsfunktionen folgen, was der nicht normalen Fehlerverteilung der meisten ökologisch relevanten Daten eher entspricht. Dadurch sind GLMs flexibler und zur Analyse ökologischer Beziehungen, die durch die klassische Gauß-Verteilung schlecht dargestellt werden können, besser geeignet.

Um GLMs zu verstehen, ist es notwendig, auf die (multiple) lineare  $\rightarrow$ Regression ( $\rightarrow$ LR bzw.  $\rightarrow$ MLR) Bezug zu nehmen.

Die Formel der LR lautet:

$$Y = \alpha + \beta X + \varepsilon \quad \text{bzw.} \quad y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i,$$

wobei Y für die Zielgröße oder  $\rightarrow$ Response(-Variable) steht,  $\alpha$  (bzw.  $\beta_0$ ) für den Achsabschnitt (eng. *intercept*),  $\beta$  für den Vektor der sogenannten Regressionskoeffizienten (bei der einfachen LR Steigung, engl. *slope*) und X für den Vektor der Ausgangsgrößen oder  $\rightarrow$ Prädiktoren.  $\varepsilon$  stellt einen Fehlerterm dar, darin werden



sowohl Meßfehler, als auch der durch das Modell nicht erklärte Teil der Abweichung erfaßt.

Beim Anpassen eines Regressionsmodells wird durch die Anwendung von Schätzmethoden ( $\rightarrow$ LS,  $\rightarrow$ ML) versucht, die unerklärte Varianz zu minimieren.

Annahmen bei der LS Regression (Guisan et al., 2002; Stahel, 2002):

- die Fehler  $\varepsilon_i$  sind unabhängig und identisch verteilt (das schließt die Annahme ein, daß die Varianz von Y konstant über die Beobachtungen ist);
- die Fehler  $\varepsilon_i$  folgen der (gleichen) Normal-Verteilung;
- die Regressionsfunktion ist linear in den Prädiktoren.

Die Verletzung der ersten Annahme (z.B. folgen Zählungen (Anzahlen) einer Poisson-Verteilung, Varianz proportional zum Erwartungswert) schränkt die Anwendung der meisten  $\rightarrow$ parametrischen Verfahren ein. Abweichungen von den Annahmen 1 und 2 werden durch  $\rightarrow$ Transformation der Response-Variable gelöst, sodaß die Bedingungen der Normal-Verteilung und der konstanten Varianz wieder erfüllt sind. Verstöße gegen die dritte Annahme werden durch Erweiterung der Prädiktoren um polynomiale Terme oder nicht-lineare Transformationen der ursprünglichen Prädiktoren behoben.

In GLMs werden die Prädiktoren zu einem linearen Prädiktor (LP bzw.  $\eta_i$ ) zusammengefaßt und zum Erwartungswert  $\mu = E(Y)$  der Response-Variable Y durch eine *link*-Funktion  $g()$  in Beziehung gesetzt (Guisan et al., 2002; Schmid, 2002).

Die allgemeine Formel lautet:

$$g(E(Y)) = LP = \alpha + \beta X = \alpha + \sum_{j=1}^p \beta_j x_{ij} \quad \text{bzw.}$$

$$g(\mu(x_i)) = \eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$$

Im Gegensatz zum klassischen linearen Modell, das Normal-Verteilung und *identity link* ( $g(\mu) = \mu$ ) voraussetzt, kann in einem GLM Y einer beliebigen Verteilung aus der Exponential-Familie (Normal-, inverse Gauß-, Binomial-, negative Binomial-, Poisson-, Gamma-Verteilung) folgen, und die *link*-Funktion kann eine beliebige monotone differenzierbare Funktion (z.B. Logarithmus oder  $\text{logit}()$ , siehe Tab. 5.2) sein.

Tabelle 5.2: *link*-Funktionen

Name	Definition $\eta_i = g(\mu_i)$	kanonisch für die
identity	$\mu_i$	Normalverteilung ( $\mathcal{N}$ )
log	$\ln(\mu_i)$	Poissonverteilung ( $\mathcal{P}$ )
logit*	$\ln\left(\frac{\mu_i}{1-\mu_i}\right)$	Binomialverteilung ( $\mathcal{B}$ )
probit*	$\Phi^{-1}(\mu_i)$	
complementlog – log*	$\ln(-\ln(1 - \mu_i))$	
log – log	$-\ln(-\ln(\mu_i))$	
diverse Potenzen	$\mu_i^\lambda$ $\mu_i^{-1}$ $\mu_i^{-2}$	Gammaverteilung ( $\Gamma$ ) inverse Gaußverteilung

\* eignet sich nur für Binomialverteilung

(Schmid, 2002, verändert)

Die wesentlichen "Fortschritte" gegenüber der LS Regression sind also:

- Fähigkeit, mit einer größeren Anzahl von Verteilungen der Response-Variablen umzugehen. GLMs können auch qualitative und semi-quantitative (ordnationale) Response-Variablen anpassen.
- Beschreibung der Beziehung zwischen Response-Variable und linearem Prädiktor über die *link*-Funktion. Zusätzlich zur sichergestellten Linearität ist dies ein effizienter Weg, die "Vorhersage" in einen für die Response-Variablen möglichen Werte-Bereich zu "zwingen" (z.B. Werte zwischen 0 und 1 für Wahrscheinlichkeiten von Vorkommen/Nichtvorkommen).

Das Anpassen eines GLM entspricht im wesentlichen dem Anpassen einer multiplen LS Regression. In beiden Fällen können polynomiale Terme in den Satz der Prädiktoren aufgenommen werden, um nicht-linearen Zusammenhängen und multimodalen Responses Rechnung zu tragen. Polynome zweiter Ordnung erlauben die Nachbildung unimodaler Response, Polynome dritter oder höherer Ordnung "schiefer" und bimodaler Response. Andere Transformationen (sogenannte parametrische Glätter) ermöglichen spezifischere Response-Formen (Guisan und Zimmermann, 2000):  $\beta$ -Funktionen, *hierarchical set of models* (HOF) von Huisman et al. (1993), "n-transformierte" Funktionen.

Wie bei der Regression kann ein Streudiagramm der  $\rightarrow$ Residuen Hinweise auf

die geeignete Transformation liefern und Ausreißer (einflußreiche Beobachtungen) können mit Standard-Methoden identifiziert werden.

GLMs sind leicht in einem GIS zu implementieren, sofern die *link*-Funktion invertiert werden kann (Guisan und Zimmermann, 2000). Die Prädiktoren werden mit den jeweiligen – bereits geschätzten – Regressionskoeffizienten multipliziert, als Ergebnis erhält man den LP. Mit der invertierten *link*-Funktion erhält man Werte im Bereich der Skala der originalen Response-Variable.

Tabelle 5.3: inverse *link*-Funktionen

<i>link</i> $\eta_i = g(\mu_i)$	inverse <i>link</i> $\mu_i = g^{-1}(\eta_i)$
$\mu_i$	$\eta_i$
$\ln(\mu_i)$	$e^{\eta_i}$
$\ln \frac{\mu_i}{1-\mu_i}$	$\frac{e^{\eta_i}}{1+e^{\eta_i}}$
$\Phi^{-1}(\mu_i)$	$\Phi(\eta_i)$
$\ln(-\ln(1-\mu_i))$	$1 - e^{-e^{\eta_i}}$
$\mu_i^\lambda$	$\sqrt[\lambda]{\eta_i}$
$\mu_i^{-1}$	$\eta_i^{-1}$
$\mu_i^{-2}$	$\eta_i^{-\frac{1}{2}}$

Für ein binomiales GLM (*link*-Funktion ist  $\text{logit}()$ ), z.B., lautet die inverse logistische Transformation:  $p(y) = \frac{e^{\text{LP}}}{(1+e^{\text{LP}})}$ . Dadurch erhält man Werte zwischen 0 und 1, z.B. Wahrscheinlichkeits-Werte für das Vorkommen einer Art.

## 5.4 Verallgemeinertes additives Modell (GAM)

Das Finden geeigneter polynomischer Terme oder Transformationen der Prädiktoren zur besseren Anpassung eines linearen Modells kann ein langwieriges und fehleranfälliges Unterfangen sein (Guisan et al., 2002).

Einen "Quantensprung" in der Regressionsanalyse stellte deshalb die Entwicklung neuer Verfahren dar, welche in der Lage sind, eigenständig eine passende Transformation ausfindig zu machen. Dies war möglich durch eine weitere Verallgemeinerung der GLMs: die sogenannten GAMs (Generalized Additive Model(s)) stellen eine "semi-parametrische" (siehe unten) Erweiterung der GLMs dar.

Man kann die verschiedenen Regressions-Modelle als ineinander verschachtelt

betrachten, allgemein ausgedrückt:

$$\text{SLR} \subset \text{MLR} \subset \text{GLM} \subset \text{GAM}$$

Die Stärke der GAMs liegt in ihrer Fähigkeit, mit nicht-linearen und nicht-monotonen Beziehungen zwischen Response und Prädiktoren umgehen zu können. Die einzige zugrunde liegende Annahme ist, daß die Funktionen sich additiv verhalten und daß die Komponenten "glatt" sind.

GAMs benutzen – wie GLMs – eine *link*-Funktion  $g(\cdot)$  um die Beziehung zwischen dem Erwartungswert  $\mu = E(Y)$  der Response-Variable  $Y$  und einem linearen Prädiktor  $\eta_i$  festzulegen. Dieser wiederum ergibt sich aus der Summe der – nicht näher spezifizierten – glatten Funktionen der einzelnen Prädiktoren  $f(x_i)$ .

Die allgemeine Formel lautet (Yee und M., 2002; Yee und N.D., 1991):

$$g(E(Y)) = \alpha + \sum_{j=1}^p f_j(x_{ij}) \quad \text{bzw.}$$

$$g(\mu(x_i)) = \eta_i = \beta_0 + f_1(x_{i1}) + f_2(x_{i2}) + \dots + f_p(x_{ip})$$

Die glatten Funktionen  $f(x)$  werden aus den Daten geschätzt und zwar mit Hilfe von Verfahren, die zur Glättung von Streudiagrammen entwickelt wurden. Zur Auswahl stehen unter anderem: gleitender Mittelwert, gleitender Median, kubische Splines, B-Splines, →LOWESS (Yee und N.D., 1991).

Knackpunkt bei der Anpassung eines GAM ist somit die Wahl des angebrachten Glättungsverfahrens sowie die Stärke der Glättung (Guisan et al., 2002). Da (allein) die Daten die Art des Zusammenhanges zwischen Zielgröße und Deskriptoren bestimmen, werden GAMs manchmal als datengesteuert bezeichnet. Im Gegensatz dazu sind modellgesteuerte Verfahren – was die Form des Response-Graphen anbelangt – durch die unterstellte parametrische Beziehung an die dadurch vorgegebenen Formen gebunden.

Die Wahrscheinlichkeits-Verteilung der Response-Variable muß zwar weiterhin festgelegt werden, insofern sind GAMs wie GLMs →parametrische Verfahren. Prädiktoren können allerdings linear, polynomialisch und darüber hinaus "nicht-parametrisch" (→nichtparametrische Regressionsmethode) modelliert werden. Man spricht deshalb von GAMs als semiparametrischen Modellen.

Ein Problem bei statistischen Modellen bzw. Regressionsanalysen ist der Einfluß der *naughty noughts* – sozusagen der "böartigen Nullen" – (Austin und Meyers,

1996), ein Aspekt der also sowohl GAM als auch GLM betrifft. Dehnt man die Beobachtungen über den möglichen Lebensraum einer Spezies hinaus aus, fallen entsprechend viele Nullen als *absence*-Daten an. Je mehr solcher Nullen in die Regressionsanalyse aufgenommen werden, desto schwächer ist die Erklärungsfähigkeit des Modells und umso stärker wird die Responsefunktion verzerrt. Das Modell "versucht" also, das Vorkommen einer Spezies in Bereichen vorherzusagen, die jenseits der für die Art tragbaren Bedingungen liegen.

Folglich ist es unter Umständen notwendig, für verschiedene Regionen getrennte Modelle zu entwickeln, damit diese auch repräsentativ sind.

## 5.5 Autologistisches Modell

Augustin et al. (1996) stellen eine Ansatz zur Schätzung der geographischen Verbreitung tierischer (und pflanzlicher) Spezies vor, der auch mit unvollständigen Erhebungsdatensätzen (siehe unten) umgehen kann. Es handelt sich im wesentlichen um eine Erweiterung der logistischen Regression, also eines GLM mit *logit()*-link.

GLMs (und GAMs) modellieren die räumliche Verteilung, indem sie die Response-Variable (Abundanz, Vorkommen/Nichtvorkommen) mit räumlich referenzier-ten Variablen in Beziehung setzen (die Autoren verwenden hier den Ausdruck *covariates*, wohl um anzudeuten, worauf sie hinaus wollen). Derartige Modelle ignorieren schlichtweg (oder modellieren bestenfalls indirekt) die Abhängigkeit von ungemessenen Variablen, sowie die den Variablen intrinsische räumliche →Autokorrelation, die sich z.B. bei Herdentieren ergibt.

Benachbarte Zellen tendieren dazu, ähnliche Bedingungen aufzuweisen, und wenn die verfügbaren Variablen diese Bedingungen nicht vollständig wiedergeben (wie sie von der untersuchten Art wahrgenommen werden), dann zeigen die →Residuen des angepaßten Modells räumliche Autokorrelation. Zudem ist die Wahrscheinlichkeit, in einer beliebigen Zelle eine Art anzutreffen, ganz abgesehen von den Umweltbedingungen, nicht unabhängig davon, ob in einer benachbarten Zelle die selbe Art vorkommt oder nicht (Einzelgänger ↔ Rudel). Dies verursacht zusätzlich räumliche Autokorrelation, die durch Umweltvariablen schlecht beschrieben werden kann. Verwendet man Modelle, die räumliche Autokorrelation berücksichtigen, benötigt man folglich u.U. weniger Variablen und erhält bessere Hinweise darauf, welche Faktoren die Verteilung einer Art beeinflussen.

Im autologistischen Modell wird diesen Tatsachen explizit Rechnung getragen, indem ein eigener Term ( $\text{autocov}_i$ ) eingeführt wird.

$$\log\left(\frac{p_i}{1 - p_i}\right) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_n \text{autocov}_i \quad \text{mit}$$

$$\text{autocov}_i = \frac{\sum_{j=1}^{k_i} w_{ij} y_j}{\sum_{j=1}^{k_i} w_{ij}}, \quad \text{wobei} \quad w_{ij} = \frac{1}{h_{ij}}$$

Der Term  $\text{autocov}_i$  ist ein gewichtetes Mittel der (besetzten) benachbarten ( $k_i$ ) Zellen der Zelle  $i$ . Das Gewicht  $w_{ij}$  einer Zelle  $j$  ergibt sich aus dem Kehrwert der euklidischen Distanz  $h_{ij}$  zwischen den Zellen  $i$  und  $j$ . Das einfachste Schema für  $k_i = 4$  sowie  $k_i = 8$  Nachbarn ist in Abbildung 5.2 zu sehen.

Die Anzahl der Zellen einer solchen Gruppe (*clique*) kann natürlich auch größer



(Augustin et al., 1996)

Abbildung 5.2: Die 4 bzw. 8 nächsten Nachbarn einer Zelle

sein, abhängig davon, wie weit sich der (angenommene,) gegenseitige Einflußbereich erstreckt.

Das Anpassen eines autologistischen Modells stellt kein Problem dar, wenn für sämtliche Zellen *presence/absence*-Daten vorliegen. Wenn aber nur Daten aus stichprobenartigen Erhebungen vorliegen, kann die "Autokovariate" nicht berechnet werden, da ja das Muster der Besetzung der benachbarten Zellen nicht bekannt ist. Dieses Problem ergibt sich beim gewöhnlichen logistischen Modell nicht, da es derartige Informationen nicht auswertet.

In solchen Fällen bietet sich der sogenannte *Gibbs sampler* an. Mit dieser Technik ist es möglich, die *presence/absence*-Verteilung für Zellen, die nicht Teil der Stichprobe sind, zu schätzen. Dieser Mechanismus berechnet die Beobachtung für die Zelle  $i$  aus dem Muster der Besetzung in den benachbarten Zellen. Als Anhaltspunkt wird zunächst ein gewöhnliches logistisches Modell angepaßt. Dann wird eine beliebige (nicht erhobene) Zelle ausgewählt und aus der geschätzten umliegenden Verteilung mittels *Gibbs sampler* eine Beobachtung generiert. Diese "Maske" wird über alle Zellen gelegt, die nicht Teil der Stichprobe sind. In

einem iterativen Prozeß wird dieser Schritt so oft wiederholt, bis das Muster konvergiert. Als Ergebnis erhält man eine *presence/absence*-Karte, anhand der ein autologistisches Modell angepaßt werden kann.

## 5.6 Ecological-Niche Factor Analysis (ENFA)

Hirzel (2001) (siehe auch Hirzel et al. (2002)) stellt einen neuen multivariaten Ansatz vor, der nicht auf *absence*-Daten angewiesen ist und somit das auf Seite 19 dargestellte Problem umgeht. ENFA *Ecological-Niche Factor Analysis* berechnet Habitatsignungsfunktionen nur auf der Grundlage von *presence*-Daten, indem die Verteilung jener Zellen, wo die untersuchte Art beobachtet wurde, mit der Verteilung des gesamten Zellen-Satzes (= globale Verteilung) verglichen wird.

**Marginalität, Spezialisierung und ökologische Nische** Arten verteilen sich bezüglich einer ökogeographischen Variable (EGV), z.B. Temperatur, nicht zufällig. Eine Spezies wird bevorzugt in jenen Zellen vorkommen, die im optimalen (Temperatur-)Bereich liegen. Dies kann man quantifizieren, indem man die Temperatur-Verteilung jener Zellen, in denen die untersuchte Tierart beobachtet wurde, mit jener aller Zellen vergleicht. Die Verteilungen können sich in Bezug auf ihren Mittelwert und ihre Streuung unterscheiden (siehe Abbildung 5.3).

Eine Tierart kann sich sowohl durch *Marginalität* (Mittelwert der Spezies unterscheidet sich vom globalen Mittelwert) als auch durch *Spezialisierung* (Standardabweichung der Spezies ist kleiner als die globale) abheben.

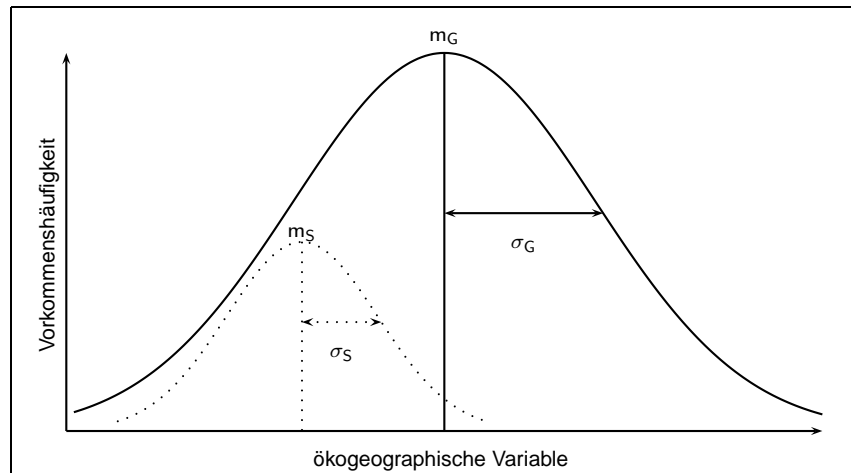
Marginalität (M) wird definiert als absolute Differenz aus globalem Mittelwert (Mittelwert der globalen Verteilung,  $m_G$ ) und Mittelwert der Spezies ( $m_S$ ), gebrochen durch 1,96 Standardabweichungen ( $\sigma_G$ ) der globalen Verteilung.

$$M = \frac{|m_G - m_S|}{1,96 \sigma_G}$$

Ein hoher Wert, d.h. nahe 1, deutet darauf hin, daß sich das Habitat der Tierart wesentlich vom Referenzgebiet unterscheidet. Analog dazu wird die Spezialisierung (S) definiert als Quotient aus der Standardabweichung der globalen Verteilung und jener der untersuchten Spezies.

$$S = \frac{\sigma_G}{\sigma_S}$$

Jeder Wert größer als 1 weist auf eine Spezialisierung der Tierart hin. Diese Gleichungen stellen nur das Prinzip des Ansatzes dar, die eigentlichen Berechnungen



Die Verteilung jener Zellen, in denen die untersuchte Spezies vorkommt (gestrichelte Linie, Spezies-Verteilung), entlang einer ökogeographischen Variable unterscheidet sich von der Verteilung aller Zellen (durchgehende Linie, globale Verteilung) bezüglich des Mittelwertes ( $m_S \neq m_G$ ), darüber wird die Marginalität definiert. Wenn sie sich zudem in der Standardabweichung unterscheidet ( $\sigma_S \neq \sigma_G$ ) liegt Spezialisierung vor.

(Hirzel, 2001, verändert)

Abbildung 5.3: Verteilung entlang einer EGV

im Modell werden durch multivariate Erweiterungen (siehe unten) bewerkstelligt. Es ist zu beachten, daß die ermittelten Werte (M und S) vom globalen Datensatz abhängen. Eine Spezies kann also auf der Ebene eines ganzen Nationalterritoriums extrem marginal und spezialisiert erscheinen, während sie sich auf ein Teilgebiet bezogen anders verhält.

Legt man diese Gleichungen auf einen größeren Satz von Variablen um, gelangt man zu Hutchinson's Konzept der ökologischen Nische: ein "Hyper-Volumen" im durch die ökologischen Variablen definierten multidimensionalen Raum. Die durch ENFA ermittelte multivariate Nische ist jene Teilmenge der Zellen im EGV-Raum, in denen die untersuchte Tierart vorkommen kann.

Die wesentlichen Achsen dieses n-dimensionalen Raumes werden mittels Faktoren-Analyse ( $\rightarrow$ FA) ermittelt. Dies aus zwei Gründen: Erstens sind ökologische Variablen nicht unabhängig voneinander, im Gegenteil sie weisen  $\rightarrow$ Multikollinearität und Redundanz auf. Ein Ziel der FA ist es, aus  $V$  *korelierten Variablen* die selbe Anzahl *unkorrelierter Faktoren* zu extrahieren. Die nachfolgenden Berechnungen werden – ohne größeren Informationsverlust – auf der Grundlage weniger Faktoren durchgeführt, sozusagen im F-dimensionalen Raum.

Zweitens hängt die Spezialisierung von Interaktionen zwischen den Variablen ab.



Z.B. variiert die von einer Spezies bevorzugte Temperatur mit der herrschenden Feuchtigkeit. Eine Tierart spezialisiert sich also eher auf eine Kombination von Variablen als auf mehrere einzelne – voneinander unabhängige – Variablen. Die FA erlaubt eine Extraktion einer Linearkombination ursprünglicher Variablen, die Marginalität und Spezialisierung der untersuchten Spezies erklären soll.

**Faktorextraktion und -Interpretation** Die ökogeographischen Variablen werden zunächst normalisiert (Box-Cox Transformation), weil das Verfahren zur Faktorextraktion mittels → Eigenanalyse theoretisch Multi-Normalität voraussetzt. FA scheint jedoch recht robust gegenüber Abweichungen zu sein, und so messen Hirzel et al. (2001a) diesem Schritt wenig Bedeutung zu. In einem weiteren Schritt werden die EGV standardisiert (→ Standardisieren). Nach Ermitteln der Kovarianz-Matrizen der globalen Verteilung und der "Spezies-Verteilung", einigen Substitutionen, Eigenanalyse und Rücksubstitution gelangt man schließlich zu den Gleichungen:

$$M = \frac{\sqrt{\sum_{i=1}^V m_i^2}}{1,96} \quad \text{und} \quad S = \frac{\sqrt{\sum_{i=1}^V \lambda_i}}{V}$$

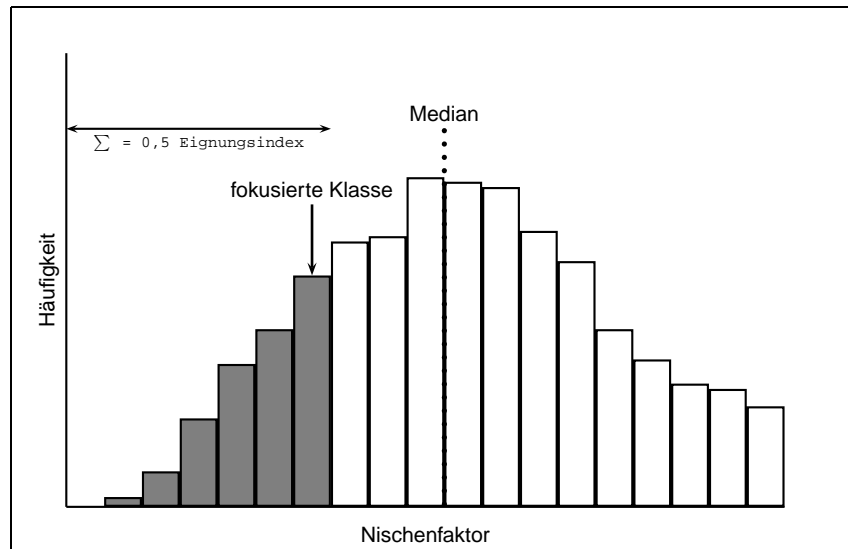
All diese Prozeduren sind in der Software Biomapper (Hirzel et al., 2001a) implementiert. Das Programm steht im Internet zum Download frei<sup>1</sup>.

$m_i$  sind die Koeffizienten des Marginalitätsfaktors. Je höher der Wert eines Koeffizienten, desto stärker unterscheidet sich das Habitat einer Spezies vom durchschnittlich verfügbaren Lebensraum. Negative Koeffizienten deuten darauf hin, daß die untersuchte Art unterdurchschnittliche Werte im Verhältnis zum gesamten Untersuchungsgebiet bevorzugt und umgekehrt. Anhand der Gesamt-Marginalität  $M$  kann man unterschiedliche Spezies im selben Gebiet direkt vergleichen. Die weiteren Faktoren sind anders zu interpretieren: Je höher ihr Wert, um so eingeschränkter ist der Lebensraum der Spezies. Hier spielen Vorzeichen keine Rolle. Die Eigenvektoren  $\lambda_i$  drücken also den Grad der Spezialisierung aus. Anhand der Gesamt-Spezialisierung  $S$  kann man unterschiedliche Arten vergleichen, sofern man die Werte auf das selbe Gebiet bezieht.

**Habitat-Eignungs-Karten** Zur Berechnung der Habitategnung für eine bestimmte Art sind verschiedene Verfahren vorstellbar. Der in Biomapper implementierte Ansatz funktioniert folgendermaßen: Der Bereich eines Faktors der von

<sup>1</sup><http://www.unil.ch/biomapper/>

der untersuchten Art in Anspruch genommen wird, wird in Klassen unterteilt, und zwar derart, daß der Median exakt zwischen zwei Klassen zu liegen kommt (siehe Abbildung 5.4). Die Eignung jeder Zelle der globalen Verteilung ergibt sich



Der Eignungsindex jeder Zelle der globalen Verteilung, für den jeweiligen Nischenfaktor, ergibt sich aus ihrer Lage (Pfeil) in der Spezies-Verteilung (Histogramm). Er wird berechnet aus der zweifachen gerasterten Fläche (Summe der Zellen der Spezies-Verteilung, die in den dem Median "abgewandten" Klassen liegen) dividiert durch die gesamte Anzahl der Zellen der Spezies-Verteilung (Fläche des Histogramms).

(Hirzel, 2001, verändert)

Abbildung 5.4: Eignungsindex einer Zelle

aus ihrer Lage (Pfeil) im Verhältnis zur Verteilung der Spezies (Histogramm). Sie wird berechnet aus der verdoppelten gerasterten Fläche (Summe aller Zellen der Spezies-Verteilung die in dieser Klasse und darunter, oder ev. darüber, liegen) dividiert durch die gesamte Anzahl der Zellen der Spezies-Verteilung (gesamte Fläche des Histogramms). Dadurch werden die Werte normiert ( $\rightarrow$  Transformation), das heißt sie liegen zwischen 0 und 1. Zellen, die in einer der zwei am Median anliegenden Klassen liegen, haben demnach einen Wert von 1, Zellen die außerhalb der Spezies-Verteilung liegen, den Wert 0.

Der Gesamtindex ergibt sich aus der Kombination der Werte für jeden Faktor. Marginalität und Spezialisierung werden gleich gewichtet, innerhalb der Spezialisierung werden die Gewichte proportional zu den Eigenwerten der Faktoren aufgeteilt.

Wiederholt man diese Prozedur für jede Zelle, erhält man eine Habitateignungskarte. Um diese (semi-)quantitative Karte in eine *presence/absence*-Karte um-

zuwandeln, reicht es einen Schwellenwert festzulegen, oberhalb welchem eine Zelle als geeignet erachtet wird. Dieser Schwellenwert wird mit einem unabhängigen Datensatz und einem  $\rightarrow$ ROC plot bestimmt.

**Vorbehalte und Einschränkungen** Ein Vorbehalt, der wohl auch für andere Ansätze gilt, ist, daß es sich bei ENFA um eine rein beschreibende Methode handelt, es können keine kausalen Zusammenhänge abgeleitet werden.

Eine Einschränkung stellt die Tatsache dar, daß ENFA nur mit linearen Abhängigkeiten umgehen kann. Nichtlineare Interaktionen können – derzeit noch – nicht angepaßt werden, es sei denn durch Transformation oder nichtlineare Kombinationen der ursprünglichen ökogeographischen Variablen.

## 5.7 FORSPACE

Aus dem Wunsch heraus, die Rolle landschaftsformender Prozesse in bewaldeten Gebieten zu analysieren und deren Einfluß auf die räumliche Verteilung von Vegetationstypen sowie auf die Metapopulations-Dynamik von Indikator-Arten abzuschätzen, wurde in den Niederlanden FORSPACE (*FORest dynamics in SPAtially Changing Environments*) entwickelt (Kramer et al., 2001).

Dieses Modell ist wohl am ehesten in die Klasse *Models of nutrient cycling and energy flow* einzuordnen (siehe Tabelle 5.1 auf Seite 27).

Dabei handelt es sich um ein prozeßbasiertes ( $\rightarrow$ process) Modell mit explizitem räumlichen Bezug ( $\rightarrow$ spatial explicit), und sozusagen nur implizit um ein Habitat-Modell. FORSPACE beschreibt die grundlegenden Prozesse, welche Ausbreitung, Keimung, Etablierung, Wachstum und Mortalität von Pflanzen (Gräser, Sträucher und Bäume) einerseits und Fertilität, Mortalität und Dynamik des Körpergewichts – in Abhängigkeit von Verbrauch und Aufwand – von Pflanzenfressern andererseits bestimmen.

Die untersuchten landschaftsformenden Prozesse sind der Weideeinfluß von großen Pflanzenfressern (Rinder, Pferde, Hirsch, Reh und Wildschwein) sowie die Auswirkung von (Wald-)Bränden. Das Modell beschränkt sich (derzeit noch) auf die vorherrschenden Pflanzenarten der nährstoffarmen sandigen Böden in den Niederlanden (Kiefer, Birke, Eiche und Buche sowie verschiedene Kräuter und Gräser).

Ziel von FORSPACE ist es, anhand von realistischen – d.h. (im Feld) meßba-

ren – Parametern besagte Prozesse möglichst realistisch und prägnant nachzubilden. Die Initialisierung des Modells erfordert keine detaillierte Feld-Erhebung. Dazu benötigte Werte werden bestehenden Tabellenwerken und Datenbanken entnommen.

FORSPACE kann auf verschiedene Fragestellungen angewandt werden, z.B.:

- Wie entwickelt sich die Artenvielfalt unter dem Einfluß interagierender landschaftsformender Prozesse?
- Welches Ausmaß bzw. welche Häufigkeit müssen landschaftsformende Prozesse annehmen damit die Heterogenität maximiert wird?
- Wie groß muß ein Schutzgebiet sein, damit es eine funktionsfähige Metapopulation einer Zielart beherbergen kann?

Dieses Modell eignet sich also für das Durchspielen verschiedener Szenarien und als Hilfsmittel bei der Entscheidungsfindung.

FORSPACE arbeitet mit einer räumlichen "Auflösung" von 20x20 m und mit Zeitintervallen von einem Monat. Es werden kontinuierliche und diskrete Prozesse sowohl für Tier- (Gewichtsänderung, Laktation,...  $\leftrightarrow$  Geburt, Tod, ...) als auch für Pflanzenarten (Höhenwachstum, Änderung der Biomasse, ...  $\leftrightarrow$  Aufgehen der Samen, Sterben alter Kohorten) modelliert.

Das Modell ist sehr detailliert, basiert aber auf einfachen, leicht nachvollziehbaren Gleichungen. So wird zum Beispiel der monatliche Energiebedarf ( $R_{qr,b,Ad}$ ) der Population (ausgedrückt in MJ) als Summe der für Trächtigkeit ( $P_{rg,b}$ ), Laktation ( $L_{ct,b}$ ) und Erhaltung ( $MntPerGroup_{b,Ad}$ ) erforderlichen Energie berechnet.

$$R_{qr,b,Ad} = P_{rg,b} + L_{ct,b} + MntPerGroup_{b,Ad}$$

Der Energiebedarf für die Trächtigkeit ( $P_{rg,b}$ ) wiederum ergibt sich aus dem Anteil der trächtigen ( $F_{Prg,b}$ ) an der Reproduktion beteiligten Tiere ( $NR_{pr,b,Ad}$ ), den monatlichen "Kosten" der Trächtigkeit ( $C_{Prg,b}$ ) und dem (metabolischen) Körpergewicht des Tieres ( $WghInd_{b,Ad}^{0,71}$ ).

$$P_{rg,b} = F_{Prg,b} * NR_{pr,b,Ad} * C_{Prg,b} * WghInd_{b,Ad}^{0,71}$$

Der Verbiß durch Reh-, Rot- und Schwarzwild wird als "Weiden an Blättern und kleinen Ästen von Bäumen und Sträuchern" modelliert. Der Grund liegt wohl darin, daß mit den selben Gleichungen das Weideverhalten von Rindern und Pfer-

den<sup>2</sup> nachgebildet wird. Ob eine Fläche von einem Pflanzenfresser aufgesucht wird, hängt von ihrer "Qualität" ab. Diese wiederum hängt ausschließlich von der Verdaulichkeit (Werte werden für die jeweilige Tierart einer Tabelle entnommen) der darauf wachsenden Pflanzen ab, nicht von der Menge an verdaulichem Pflanzenmaterial. Erst wenn auf der höchstwertigen Fläche sämtliche verfügbare "Stoffwechsel-Energie" aufgebraucht ist, suchen die Tiere die nächste auf. FORSPACE berücksichtigt sogar Viehtritt als Mortalitätsfaktor bei Pflanzen. Allerdings wird dessen Wirkung – für alle Tierart gleich – allein über die aufgenommene Biomasse bestimmt, was gegenüber den Wildarten nicht unbedingt plausibel ist (Verhältnis Körpergewicht zu Aufstandsfläche nicht bei allen gleich).

**Räumliche Analyse** Diese wird im wesentlichen mittels dreier Komponenten realisiert: einem Datenbank(-Management-)-System, einem (rasterbasierten) GIS-Viewer und einer eigens programmierten Benutzer-Schnittstelle. Diese ist sowohl für die Bedienung als auch für die Berechnung und den Datenaustausch zwischen Datenbank und GIS-Viewer zuständig.

Die räumliche Analyse kann um benutzerdefinierte Prozeduren erweitert werden. Diese beruhen auf Datenbank-Abfragen, Tabellenkalkulation und (script-basierten) Modulen.

Eine typische Fragestellung bei der räumlichen Analyse wäre z.B.: Wieviel geeignetes Habitat steht über welchen Zeitraum, innerhalb eines bestimmten Radius, vorgegeben durch die spezifischen Eigenschaften der Flächeneinheiten, zur Verfügung?

**Metapopulations-Dynamik** Das in FORSPACE integrierte Metapopulationsmodell kann nur dann sinnvoll auf eine Spezies angewandt werden, wenn diese sich im gewählten Betrachtungsmaßstab (→scale) auch tatsächlich wie eine →Metapopulation verhält. Das Besondere an diesem Modul ist, daß die Dynamik der modellierten Metapopulation nicht auf eine statische, unveränderliche Umwelt bezogen wird, sondern auf eine ständigen Änderungen unterworfenen Landschaft – eben die in FORSPACE modellierte.

Im Modell kann geeignetes Habitat von lokalen Populationen besiedelt werden.

---

<sup>2</sup>Waldweide wird von den Autoren als in den Niederlanden relativ neue Praxis beschrieben. Das Vieh hält sich dabei – soweit man aus dem Modell schließen kann – das ganze Jahr über (ohne Zusatzfütterung) im Freien auf.

Diese können mit der Zeit – aufgrund der eigenen Dynamik oder jener der Umwelt – erlöschen. Unbesiedelter und neu entstandener Lebensraum kann neu besiedelt werden. Das (ausgewogene) Zusammenspiel aus Besiedlung geeigneten Habitats, Erlöschen lokaler Populationen und (Wieder-)Besiedlung ev. neu entstandener Lebensräume bestimmt Vorkommenshäufigkeit und Aussterberisiko in einer Metapopulation.

**Feuermodell** Das eingesetzte Modell ist einfach gehalten. Es beschreibt die Anhäufung von brennbarem Material, das Auftreten eines Brandes in Abhängigkeit vom angehäuften Material (und einiger anderer Bedingungen) sowie die Ausbreitung des Feuers als Maß für dessen Intensität. Die Intensität bestimmt ihrerseits die Höhe der Flammen in der Krautschicht. Das Feuer kann sich – abhängig von der Bestandesstruktur – vertikal in die Strauch- und Baumschicht ausbreiten.

Die Auswirkungen eines Brandes schlagen sich im Modell als erhöhte Mortalität und Minderung der Biomasse nieder.

Eine Aussage des Feuermodells könnte z.B. sein, wie man durch kleine kontrollierte Brände (*small-scale prescribed burning*) räumliche Heterogenität herstellen und größere Flächenbrände vermeiden kann.

In Zukunft soll FORSPACE auf weitere Arten ausgedehnt werden. Weiters sollen unterschiedliche forstliche Bewirtschaftungsarten einbezogen und andere (Wohlfahrts-)Wirkungen (CO<sub>2</sub>-Fixierung, Erholung) berücksichtigt werden. Auch soll die Verwundbarkeit durch Windstürme in Abhängigkeit von der Bestandesstruktur modelliert werden.

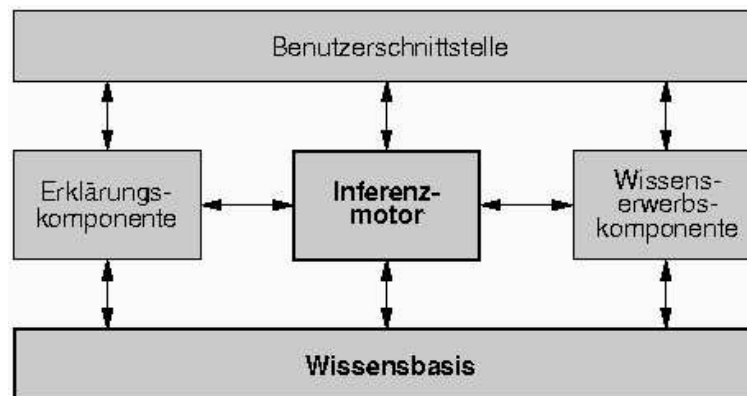
## 5.8 Expertensysteme

Expertensysteme – auch als wissensbasierte Systeme bezeichnet – sind eine Teildisziplin bzw. ein Anwendungsgebiet der künstlichen Intelligenz. Der Begriff Expertensystem bringt den externen Aspekt – ein Verhalten analog zu menschlichen Experten – zum Ausdruck. Mit dem Begriff wissensbasiert wird eher die interne Systemstruktur angesprochen: Wissensbasierte Systeme verarbeiten Symbole, die Informationen über die Welt repräsentieren, um intelligentes Verhalten zu erreichen.

Ziel eines Expertensystems ist es, das Spezialwissen und die Schlußfolgerungsfähigkeiten qualifizierter Fachleute in einem Softwaresystem nachzubilden und so zu vergleichbaren oder besseren Problemlösungen zu kommen (Abecker, 2000). Dadurch sollen einerseits Anwender profitieren, und andererseits Experten entlastet werden.

Zunächst muß das Fachwissen in eine Wissensbasis (siehe Abb. 5.5) überführt werden (Greif, 2000). Bindeglied zwischen Expertensystem und Experten ist dabei ein sogenannter *knowledge engineer*. Dessen Aufgabe ist es, das Wissen zu strukturieren, formalisieren und in maschinengerechter Form darzustellen. Dies geschieht durch sogenannte logische Wissensrepräsentation, das sind Regeln der Form: WENN Vorbedingung DANN Aktion (z.B. WENN X und/oder Y DANN Z). Damit werden Fakten, Zusammenhänge, Schlußfolgerungsmechanismen und strategisches Wissen (wie wird ein Problem angegangen) gespeichert.

Zudem muß ein Expertensystem in der Lage sein, das gespeicherte Wissen (Fak-



(Greif, 2000)

Abbildung 5.5: Komponenten eines Expertensystems

ten und Regeln) zu interpretieren, d.h. Schlußfolgerungen zu ziehen. Dies wird mittels einer Problemlösungskomponente, der Inferenzmaschine gelöst. Damit werden die aus einer Problemstellung extrahierten Fakten nach den entsprechenden Regeln verknüpft, und auf neue Fakten geschlossen. In der Wissensbasis wird also nur festgelegt, was bei bestimmten Bedingungen getan werden soll, über die Reihenfolge der einzelnen Problemlösungsschritte entscheidet allein die Inferenzmaschine.

Die Benutzerschnittstelle übernimmt die Kommunikation mit dem Anwender des Systems. Sie nimmt Problemstellungen entgegen, reicht sie an die Inferenzma-

schine weiter und präsentiert die Lösungsvorschläge.

Die Erklärungskomponente soll Lösungswege analysieren und Entscheidungen begründen (wie wurde eine Lösung gefunden). Die Wissenserwerbekomponente schließlich, bietet die Möglichkeit, neues Wissen in die Wissensbasis aufzunehmen und diese auf Konsistenz und Vollständigkeit zu Überprüfen.

## 5.9 (Künstliche) Neuronale Netzwerke (ANN)

ANNs (Artificial Neural Networks) sind intelligente Maschinen: Sie bilden im Wesentlichen den tierischen bzw. menschlichen Gehirnkomplex und damit "Denkvorgänge" nach. In einem Lernprozeß werden die nötigen "Erfahrungen" gesammelt, die später auf neue Probleme angewandt werden.

In Anlehnung an Anatomie und Physiologie des menschlichen Gehirns setzt sich ein ANN aus einer mehr oder minder großen Anzahl von Neuronen zusammen (Sauer, 1999). Neuronen, auch Verarbeitungseinheiten genannt, sind (einfache) Prozessoren, die entweder als Hardware (auf Beschleunigerkarten) oder als Softwaresimulationen implementiert sind. Diese Einheiten sind untereinander (mehrfach) verbunden und in Schichten organisiert: Eingabeschicht, (k)eine oder mehrere (verborgene) Zwischenschichten (*hidden layer* bestehen aus Einheiten, die von außen nicht beeinflusst werden können) und Ausgabeschicht. Die einzelnen Schichten bestehen aus einer unterschiedlichen Anzahl meist gleichartiger Verarbeitungseinheiten. Jedes Element der Eingabeschicht ist (indirekt) mit jedem Element der Ausgabeschicht verbunden.

Die Informationsverarbeitung erfolgt durch Ignorieren, Verstärken oder Abschwächen von Signalen. Ein Neuron nimmt über "Dendriten" Signale auf (siehe Abb. 5.6 *A processing element*). Falls die Summe der eingehenden Signale einen bestimmten Wert überschreitet, wird das Signal weiterverarbeitet (Aktivierungsfunktion) und (an den "Synapsen") entsprechend verstärkt oder gedämpft weitergeleitet.

Neuronale Netzwerke machen nichts anderes, als verschiedene Eingabemuster bestimmten Ausgabemustern zuzuordnen. Zusammenhänge zwischen Ein- und Ausgabe werden als Gewichte dargestellt.

Ein ANN kennt 2 Modi: einen Lernmodus (Trainingsphase) und einen Ausführungsmodus (Recall Modus). Während der Trainingsphase werden die Gewichte – bestimmten Lernregeln folgend – so lange verändert, bis ein bestimmtes Ein-

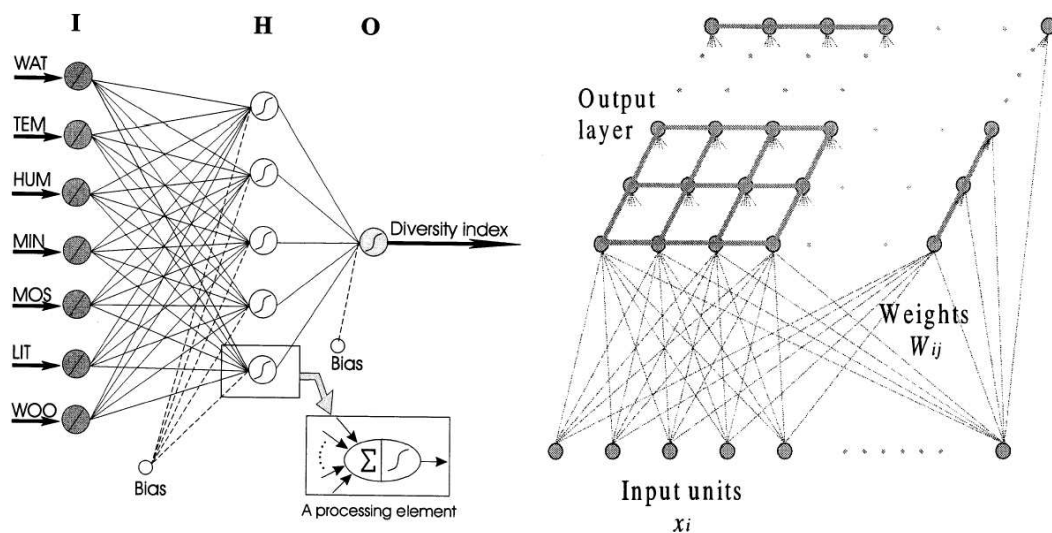


gabemuster dem gewünschten Ausgabemuster zugeordnet wird. Im Recall Modus werden die trainierten ANN mit Mustern gefüttert, die sie noch nicht gelernt haben. Die Qualität der Ausgabe wird durch Dauer und Intensität der vorangegangenen Lernphase bestimmt.

Es gibt nicht das ANN schlechthin, sondern verschiedenste Typen neuronaler Netzwerke, die für unterschiedliche Problemstellungen geeignet sind (Steuerung, Optimierung, Verallgemeinerung, Vorhersage u.a.m.). Sie unterscheiden sich in mehrfacher Hinsicht: Anzahl der Schichten, Art des Lernens (Lernregeln), *feed forward* (Signale werden nur "vorwärts" weitergereicht) bzw. *feed back*, Verarbeitung binärer oder stetiger Daten. Die Wahl des Netzwerk-Types hängt also von der Natur des Problems ab.

ANNs werden als leistungsstarkes Modellierungs-Werkzeug beschrieben, speziell wenn die den Ein- und Ausgabedaten zugrunde liegende Beziehung unbekannt ist (Lek und Guégan, 1999). Sie sind also bestens für die Analyse komplexer, nichtlinearer Zusammenhänge zwischen ungenauen und *noisy* Daten geeignet, was in der Ökologie meistens der Fall ist.

Momentan erfreuen sich 2 ANNs großer Beliebtheit: *Backpropagation*-Netzwerke



(Lek und Guégan, 1999; Lek-Ang et al., 1999)

Abbildung 5.6: *BackPropagation Network* und Kohonen SOM

ke (BPN) und (Kohonen-)Selbstorganisierende 'Karten' (SOM) (siehe Abb. 5.6 links bzw. rechts). BPN sind drei- oder mehrschichtige *feed-forward* Netze, die inneren Schichten sind versteckt. Information fließt unidirektional von der Eingangs-

beschicht über die versteckten Schichten zur Ausgabeschicht, ohne seit- oder rückwärts gerichtete Verbindungen. Backpropagation ist der Namen der Lernregel: Von der Ausgabeebene wird der Einfluß der Ebenen auf alle Fehler zurückgerechnet (Sauer, 1999). Das BPN erstellt das Modell allein auf der Grundlage der eingespeisten Daten. Praktisch wird angenommen, daß die Information die notwendig ist, um die Beziehung zwischen Ein- und Ausgabe herzuleiten, implizit in den Daten enthalten ist.

Kohonen SOM fallen in die Kategorie der unüberwacht lernenden ANN. Sie ahmen die Fähigkeit des Gehirns zur Selbstorganisation nach: SOM organisieren ihre Verbindungsstruktur gemäß einfachen Regeln selbständig. Sie bestehen nur aus einem Input- sowie einem Output-Layer. Die Eingabeschicht fungiert nur als *flow-through* Layer (keine Veränderung des Eingabe-Musters). Die Ausgabeschicht besteht aus einem 2-dimensionalen, quadratisch angeordneten Neuronen-Netz. Jedes Neuron ist mit den  $n$  nächsten Nachbarn verbunden. Die "Neuronen" speichern einen Gewichtungs-Vektor, einen für jeden entsprechenden Eingabewert. Mittels multivariater Algorithmen wird nach Mustern in den Daten gesucht.

ANN sind ein 'black box'-Ansatz: Die gefundenen Regeln entziehen sich weitgehend einer fachlichen Interpretation, da sie in der Vielzahl von Parametern mit kompliziertem Zusammenspiel versteckt sind (Stahel, 2002).

## 5.10 Classification Trees

*Classification trees*, manchmal auch als *decision trees* bezeichnet, sind im Bereich  $\rightarrow$ AI, genauer gesagt in der Disziplin des  $\rightarrow$ machine learning anzusiedeln. Ein *Entscheidungsbaum* ist eine hierarchische Struktur. In jedem *Knoten* (*node*, Verzweigung) wird bestimmt, welcher tieferliegende *Knoten* durchsucht wird. Jedes *Blatt* (*leaf*, Endpunkt) stellt eine Entscheidung dar. Das Absuchen eines Entscheidungsbaumes ermöglicht die Generierung neuen Wissens durch Schlussfolgerung (Xie et al., 2001).

Dabei wird der Wert einer diskreten abhängigen Variable aus einem endlichen Satz von Werten (*Klassen* genannt) von den Werten eines Satzes unabhängiger (diskreter oder kontinuierlicher) Variablen, den sogenannten *Attributen*, abgeleitet (Debeljak et al., 2001).

Ein reales System wird in Form einer Tabelle dargestellt. Die darin enthaltenen

Daten werden zum Erlernen bzw. automatischen Konstruieren eines Entscheidungsbaumes benutzt. Jede Zeile (*example, Beispiel*) in der Tabelle hat die Form  $(x_1, x_2, \dots, x_N, y)$ , wobei  $x_i$  die Werte der  $N$  *Attribute* (z.B. Seehöhe eines Standortes, Nähe zur nächsten Siedlung, Vegetationsdecke) darstellen und  $y$  den Wert der *Klasse* (z.B. Punkt ist als Lebensraum ungeeignet/geeignetes Winterhabitat/geeignetes Sommerhabitat).

Der *induzierte* (erlernte) Entscheidungsbaum hat in jedem *Knoten* einen Test (*IF-THEN rules*, einfache WENN-DANN-Regeln, wie z.B. WENN *Attribut* größer/kleiner/gleich DANN ...) zum Überprüfen des Wertes eines bestimmten *Attributes*, und in jedem *Blatt* den Wert einer *Klasse* (siehe Abb. 5.7).

Für jedes neue *Beispiel* (Zeile in der Tabelle, Stichprobenpunkt), für das der Wert

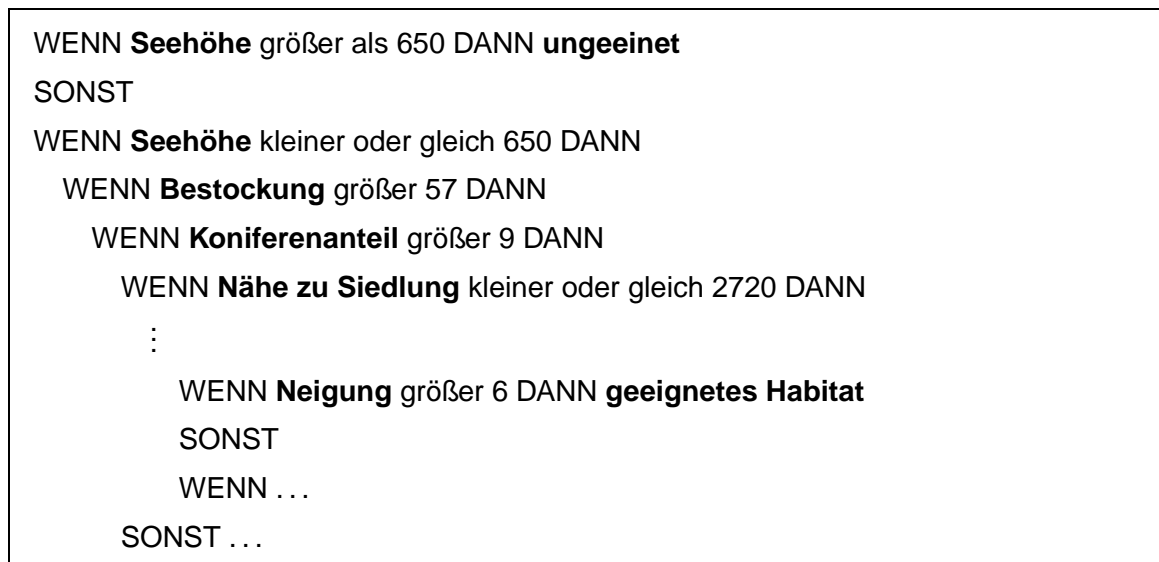


Abbildung 5.7: *decision tree*

der *Klasse* vorhergesagt werden soll, wird der Baum von der Wurzel ausgehend interpretiert. In jedem *Knoten* wird der vorgeschriebene Test durchgeführt, und entsprechend zum Ergebnis des Tests wird der entsprechende (Unter-)Baum ausgewählt. Handelt es sich beim ausgewählten *Knoten* um ein *Blatt*, dann wird der Wert der *Klasse* bestimmt.

Der allgemeine Lösungsweg zum *Induzieren* von Entscheidungsbäumen wird als *Top-Down Induction of Decision Trees* (TDIDT) bezeichnet. Die Konstruktion des Baumes erfolgt rekursiv, beginnend mit dem gesamten Satz an Trainingsbeispielen (gesamte Tabelle). Bei jedem Schritt wird das informativste *Attribut* als Wurzel für den (Unter-)Baum ausgewählt, und das aktuelle Trainingsset wird entspre-

chend dem Wert der ausgewählten Variable in Subsets aufgegliedert. Bei diskreten *Attributen* wird für jeden Wert, den die Variable annehmen kann, ein entsprechender Ast generiert. Bei kontinuierlichen *Attributen* wird ein Schwellenwert festgelegt, an dem sich der Baum, abhängig vom Wert der Variable, verzweigt.

Zur Validierung eines Entscheidungsbaumes wird idealerweise ein zweiter, unabhängiger Datensatz herangezogen. Steht kein solcher zur Verfügung, muß man sich mit  $\rightarrow$ resampling Methoden, z.B. Kreuzvalidierung, behelfen.

Algorithmen zum Implementieren von *classification trees* sind CART<sup>3</sup>, ASSISTANT<sup>4</sup> und C4.5<sup>5</sup> bzw. C5<sup>6</sup>.

## 5.11 GRASP

GRASP (*Generalized Regression Analysis and Spatial Prediction*) stellt eigentlich keinen selbständigen Ansatz dar (Lehmann et al., 2003).

GRASP ist vielmehr eine Art Benutzerschnittstelle zu einer Sammlung von in "S-Plus" (einem Statistikpaket) implementierten Funktionen. Damit steht ein Werkzeugkasten zur Verfügung, der eine einfache und rasche Möglichkeit bietet, Daten zu analysieren, Modelle zu erstellen und zu evaluieren, sowie auf der Grundlage des entwickelten Modells Vorhersagen zu treffen.

Die aktuelle Version von GRASP beruht auf GAM, es sollen jedoch auch GLM und ANN integriert werden. GLM, weil es sich dabei um eine Methode mit robustem theoretischen Hintergrund und einer Vielzahl von analytischen Werkzeugen handelt. ANN, weil sie sich besser zur Implementierung von Interaktionen zwischen den Prädiktoren eignen und damit die darin enthaltene Information besser auswerten als andere Ansätze.

Die Integration von Methoden zur Residuenanalyse, für den Umgang mit reinen *presence*-Datensätzen, zur Analyse von zwischenartlicher Konkurrenz, Kreuzva-

---

<sup>3</sup>Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., (1984): Classification and Regression Trees. Wadsworth, Belmont.

<sup>4</sup>Cestnik, B., Kononenko, I., Bratko, I., (1986): ASSISTANT 86: a knowledge elicitation tool for sophisticated users. In: Bratko, I.; Lavrač, N. (Eds.): Progress in Machine Learning. Sigma, Wilmslow, pp. 31-45.

<sup>5</sup>Quinlan, J.R., (1993): Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.

<sup>6</sup>Quinlan, J.R., (1998): C5/See5 Software. <http://www.rulequest.com/>. Eine auf 400 Einträge pro Datensatz beschränkte Demoversion von C5 steht unter <http://www.rulequest.com/download.html> für verschiedene UNIX-Varianten zum Download sowie als See5 für Windows-Betriebssysteme.

lidierung (→resampling-Methoden) und Variablenselektion ist ebenfalls angekündigt.

## 5.12 GAP

GAP (*Gap Analysis Program*) ist ein vom *USGS National Biological Service* koordiniertes Projekt. Es handelt sich dabei um ein Instrument zur Inventarisierung von Habitat-Typen und Wildtierarten. Es bedient sich eines GIS, um Gebiete zu identifizieren, die für die Erhaltung der Biodiversität vorrangig sind (Van Horne, 2002).

Eine Gap Analyse entsteht in 5 definierten Schritten, beschrieben in *A handbook for gap analysis* (USGS Gap Analysis Program, 1998). Im wesentlichen werden Vegetationsdecke, Arten-Verbreitung und administrative Grenzen kartiert, verschnitten und statistisch ausgewertet.

Endprodukt ist eine thematische, digitale Karte (im Maßstab 1:100.000) der aktuellen Habitat-Typen und der Verbreitung jeder Wirbeltierart. Diese sollen einen Überblick über den Status Quo gewähren, Tierarten und Vegetations-Gesellschaften identifizieren, die nicht angemessen repräsentiert sind (eben die Namen gebenden *gaps*) und als Grundlage für Entscheidungen in Fragen der Landschaftsnutzung dienen. Damit soll der Schwerpunkt in der Planung von der Reaktion auf die Initiative verlegt werden. Hauptziel ist es, "weitere Arten vor der Aufnahme in die Liste der als bedroht oder gefährdet eingestuftten Spezies zu bewahren" (Scott, 2000).

GAP versteht sich als erweiterter "Grob-Filter"-Ansatz (*coarse filter approach*) zur Erhaltung der Biodiversität: Die Vegetationstypen dienen als Filter für die Planung der Landschaftsnutzung.

Werte wie Habitat-Qualität und Abundanz einer Art finden nicht Eingang bzw. sie sind für die meisten Spezies/Pflanzengesellschaften unbekannt. Somit kann GAP keine Vorhersagen über die Existenzfähigkeit einer Population treffen. Bedarfsweise wäre zu diesem Zweck eine PVA (siehe 5.15) angebracht.

Zudem unterscheidet sich der räumliche Maßstab, in dem Organismen ihre Umwelt nutzen, abhängig von Körpergröße, Freß-Verhalten, Mobilität u.a.m. erheblich. Endemische Populationen oder weit verbreitete Arten können ev. durch den "Rost" fallen. Aufgrund der grobkörnigen Auflösung ist GAP also kein "Allheilmittel" für die Erhaltung der Biodiversität, einzelne Arten bedürfen eines individuellen

Schutzes (Crist, 2000).

## 5.13 RSF

*Resource Selection Functions* sind ein Nebenprodukt von Modellen zur Nachbildung der natürlichen Selektion. Sie beschreiben die Habitat-Nutzung durch (Wild-)Tiere.

"Ökologie ist das wissenschaftliche Studium der Verteilung und Abundanz von Organismen. Wenn die Verteilung der für eine Tierart relevanten Ressourcen, bekannt ist, kann die Verteilung bzw. Abundanz dieser Tierart durch RSF beschrieben werden. Damit kann jede Funktion die proportional zur Wahrscheinlichkeit der Inanspruchnahme einer Ressource-Einheit ist, als RSF bezeichnet werden" (Boyce und McDonald, 1999).

RSF bedienen sich unter anderem log-linearer und logistischer Regressionsmethoden, also GLMs (Guisan et al., 2002).

Eine einfache RFS könnte lauten (Boyce und McDonald, 1999):

$$w_i = \frac{o_i}{\pi_i} \quad \text{mit} \quad \pi_i = \frac{a_i}{a_+}$$

$a_+$  steht für die Gesamtheit der verfügbaren Ressource-Einheiten,  $a_i$  für jene der Kategorie  $i$ .  $o_i$  ist der Anteil beanspruchter Ressource-Einheiten in der Kategorie  $i$ .  $w_i$  – auch als *selection ratio* bezeichnet – steht also für den beanspruchten Anteil von Ressource-Einheiten einer Kategorie  $i$  an den insgesamt verfügbaren. Man kann diesen Ansatz weiterentwickeln, z.B. zum Modellieren der Nutzung von Habitat-Typen im Verhältnis zur deren Verfügbarkeit. Etwa indem man die RSF durch ein log-lineares Modell beschreibt:

$$w(x) = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k},$$

wobei die  $x_i$  für die unabhängigen Habitat-Variablen stehen und die  $\beta_i$  als "Selektions-Koeffizienten" bezeichnet werden. Ein derartiges Modell wird durch ein GLM angepaßt.

## 5.14 Individual-based Models

Bei diesem Ansatz werden einzelne Individuen (und deren Umwelt) als Objekte mit entsprechenden Eigenschaften (z.B. Alter, Geschlecht, sozialer Rang) nach-

gebildet. Das Verhalten einer gesamten Population wird durch das Zusammenspiel dieser Individuen bzw. deren Interaktion mit der Umwelt abgebildet. Derartige Modelle berücksichtigen – im Gegensatz zu den meisten oben angeführten – kausale Zusammenhänge ( $\rightarrow$ process), sind hoch spezifisch und sehr komplex. Z.B. beeinflussen Verfügbarkeit von Futter und Wahrscheinlichkeit von einem Beutegreifer geschlagen zu werden (Eigenschaften der *patches* bzw. Individuen anderer Spezies) die Wahl zwischen Habitattypen (Nahrung  $\Leftrightarrow$  Deckung). Die Entscheidung wird in Abhängigkeit vom physiologischen Status (ausgedrückt als Wahrscheinlichkeit, einen bestimmten Schwellenwert an Körperfett erhalten zu können) des Individuums getroffen (Van Horne, 2002).

Aufgrund ihrer Komplexität sind IBMs schwer zu entwickeln, schwer zu kommunizieren und schwer zu verstehen (Grimm et al., 1999).

## 5.15 Populations Modelle

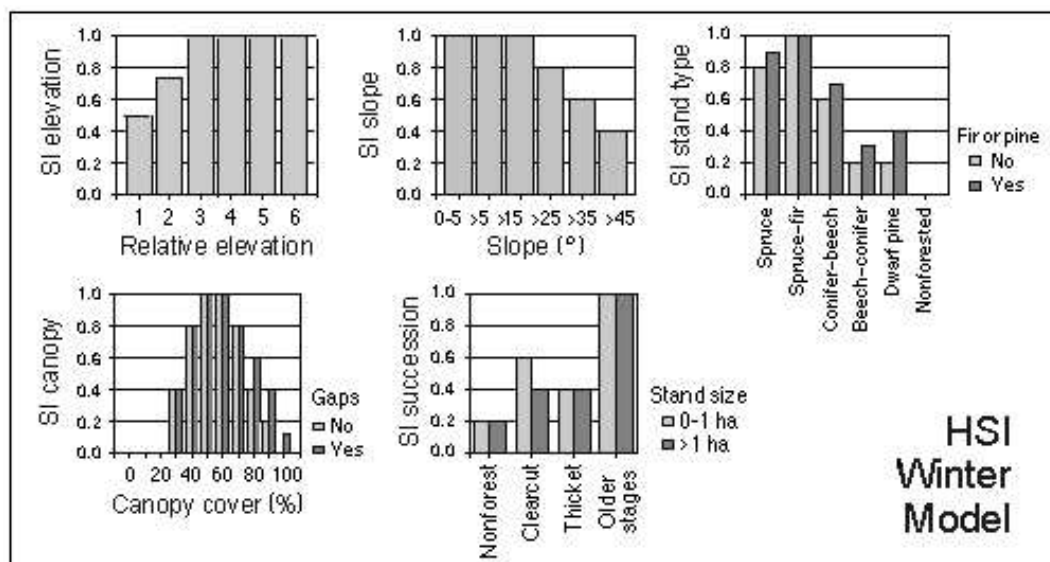
Modelle zur Populations-Dynamik, *population viability models* (Morrison et al., 1998, S. 50ff) sowie *life-table approaches* (Van Horne, 2002) werden im Zusammenhang mit Habitatmodellen oft erwähnt und verwendet, um das Aussterberisiko einer Population zu bewerten. Dabei wird z.B. die Auswirkung der Fragmentierung eines Lebensraums auf Dynamik, Überlebensfähigkeit und Migrationsverhalten einer (Meta-)Population untersucht. Realistischerweise sollten darin auch Faktoren wie dichteabhängige demographische Mechanismen, Konkurrenten und Räuber mit berücksichtigt werden.

## 6. Modellanwendungen

### 6.1 HSI

In den bayerischen Alpen wurde ein Habitatmodell zur Bewertung von Auerwildlebensräumen entwickelt (Storch, 2002). Als HSI-Modell beruht es auf einer Kombination von Forschungsdaten und Expertenurteil. Der Datensatz für die Habitatvariablen wurde aus einer früheren telemetrischen Studie abgeleitet, aus der folgende Faktoren als signifikant hervorgegangen waren: relative Höhe über Talboden bzw. Grenze landwirtschaftlich genutzte Fläche - Wald (ele), Hangneigung (slo), Sukzessionsstadium/Betriebsklasse (suc), Kronenschluß (can), Lücken im Kronendach (yes/no), Bestandestyp (typ), Bäume die als Winternahrung dienen (yes/no), Bodendeckung Verjüngung (reg), Bodendeckung Heidelbeere (bil) und Höhe der Bodenvegetation (veg). Da das Modell auf den ganzen bayerischen Alpenraum anwendbar sein sollte, wurden auch Variablen aufgenommen, die im eigentlichen Untersuchungsgebiet nicht relevant waren. Ins Modell flossen 8 Variablen ein.

Für jede Modellvariable wurde eine Eignungsindex-Funktion (*SI function*, bzw.



(Storch, 2002)

Abbildung 6.1: Eignungsindex-Funktionen für den Winter

SI) mit Werten zwischen 0 (für ungeeignet) und 1 (für optimal) erstellt, welche



den angenommenen Zusammenhang (siehe Abb. 6.1) zwischen der Variable und der Nutzung durch das Auerhuhn beschreibt. Bei Variablen, welche die Habitategnung verringern, wurde der Mindestwert auf  $> 0$  gesetzt (um zu verhindern, daß der Gesamtindex nicht 0 wird). Die SIs wurden anschließend zu einfachen Gleichungen zusammengefaßt, um den Gesamtwert der Habitategnung HSI zu berechnen.

Die mathematische Verknüpfung der SI-Werte wurde ihrer angenommenen Rolle für die Habitategnung entsprechend gewählt: Multiplikation bei großem Einfluß (limitierende Wirkung durch kleine Werte), arithmetisches Mittel bei kompensatorischer Wirkung und geometrisches Mittel bei teilweise kompensatorischer Wirkung (kleine SIs fallen stärker ins Gewicht).

Aufgrund der jahreszeitlich unterschiedlichen Habitatansprüche des Auerhuhns wurden für Winter (Schneelage) und Sommer (schneefreie Zeit) getrennte Eignungsindizes berechnet und abschließend in einem Index für die ganzjährige Habitategnung zusammengefaßt.

In den Index für die Wintereignung flossen die Variablen Sukzessionsstadium, Kronenschluß, Bestandestyp, Hangneigung und Seehöhe ein ( $HSI_{wi} = (SI_{suc} \times SI_{can}) \times \sqrt[2]{SI_{typ} \times SI_{slo}} \times SI_{ele}$ ).

Der Index für die Sommereignung schloß die Variablen Sukzessionsstadium, Kronenschluß, Bodendeckung Heidelbeere, Bodendeckung Verjüngung, Höhe der Bodenvegetation, Hangneigung und Seehöhe ein ( $HSI_{su} = 0.25 \times \{(SI_{suc} \times SI_{can}) + (2 SI_{bil} \times SI_{reg}) + SI_{veg}\} \times SI_{slo} \times SI_{ele}$ ).

Der Index für die ganzjährige Habitategnung wurde als geometrisches Mittel aus den beiden saisonalen Indizes gebildet ( $HSI_{year} = \sqrt[2]{HSI_{wi} \times HSI_{su}}$ ).

Das fertige Modell wurde anhand zweier Datensätze "verifiziert": jenem der zur Anpassung des Modells verwendet worden war (telemetrische Ortungen) und einem zweiten unabhängigen (indirekte Nachweise, nur Sommer). Für beide Datensätze wurden die 403 bzw. 169 Bestände in 5 HSI-Klassen (sehr gut, gut, mittel, mäßig, schlecht) eingeteilt. Aus den (telemetrischen und indirekten) Beobachtungen wurden Indizes für die Habitatnutzung durch das Auerwild abgeleitet. Schließlich wurde eine einfache  $\rightarrow$ Spearmansche Rangkorrelation zwischen den Mittelwerten der HSI-Klassen und der Habitatnutzung berechnet. Die Mittelwerte aller 5 HSI-Klassen wiesen einen signifikanten Zusammenhang mit den Indizes für Habitatnutzung auf. Je höher der HSI-Wert, desto stärker war die Nutzung durch das Auerwild.

## 6.2 GLM

Zimmermann und Breitenmoser (2002) haben Habitatvariablen und deren Bedeutung für die Wiederbesiedlung des Schweizer Jura durch den Luchs untersucht, um den verfügbaren Lebensraum abschätzen zu können. Das Modell wurde mit Daten aus dem südlichen Jura entwickelt (ständig vom Luchs besetzt) und anschließend auf den gesamten Gebirgszug übertragen (die nördliche Hälfte wird nicht ständig besiedelt).

Es wurden circa 6.300 telemetrische Ortungen von 11 adulten, territorialen Luchsen verwendet, unter der Annahme, daß diese Tiere das beste Habitat besetzen. Das gesamte Untersuchungsgebiet erstreckte sich über 680km<sup>2</sup>, als Stichprobeneinheit wurden Flächen von 1x1km<sup>2</sup> Ausdehnung festgelegt. Für das Modell wurde die Stichprobe auf jene Flächeneinheiten beschränkt, die innerhalb des MCP (*minimum convex polygon*) zu liegen kamen. Dabei blieben 1.085 Quadrate übrig.

Aus diesen wurden 400 zufällig so ausgewählt, daß ein Mindestabstand gewährleistet war, um die räumliche →Autokorrelation zu reduzieren. Mit diesem Satz wurde das Modell entwickelt, die übrigen 685 Quadrate wurden zur Evaluierung des Modells herangezogen.

Response-Variable war das Vorkommen/Nichtvorkommen (*presence/absence*) von Luchs. Der Datensatz wurde nochmals aufbereitet, sodaß 3 neue entstanden: (1) Luchs (beide Geschlechter), (2) (nur) Weibchen und (3) (nur) Männchen. Entscheidungsgrundlage bei der Auswahl der Prädiktoren war einerseits die Erfahrung der Autoren, was die ökologischen Ansprüche des Luches anbelangt und andererseits die Verfügbarkeit der Daten in digitaler Form. So gingen 18 Variablen ins Modell ein: bewaldete Fläche (ha/km<sup>2</sup>), andere bestockte Flächen (ha/km<sup>2</sup>), Obst- und Weingüter (ha/km<sup>2</sup>), Seehöhe (m), Neigung (°), Straßen und Eisenbahnen (ha/km<sup>2</sup>), ....

Die Prädiktoren wurden nicht orthogonalisiert, um die spätere ökologische Interpretation des Modells zu erleichtern. Ins endgültige Modell wurden nur jene Variablen aufgenommen, die signifikant zur Varianzreduktion beitragen konnten. Weiters wurden Variablen ausgeschlossen, die weniger als 1% der Streuung erklärten, um nicht Prädiktoren mit geringer bzw. keiner biologischer Bedeutung zu erhalten.

Tabelle 6.1: Formeln für verschiedene Datensätze

Response variables	GLM formulas
Presence/absence of lynx	$\text{elev}^2 + \text{slo} + \text{forest}$
Presence/absence of females	$\text{elev}^2 + \text{slo} + \text{forest}$
Presence/absence of males	$\text{elev} + \text{slo} + \text{forest} + \text{roads}$

(Zimmermann und Breitenmoser, 2002, verändert)

Ins Modell für den Datensatz "Luchs" fanden 3 Prädiktoren Eingang: Seehöhe, Neigung und Waldanteil (siehe Tab. 6.1). Das selbe gilt für den Datensatz "Weibchen", wenn auch mit leicht unterschiedlichem Erklärungswert. Beim Modell für den Datensatz "Männchen" wurde eine weitere Variable aufgenommen, nämlich die Straßendichte.

Die Karte für die potentielle Verteilung der Männchen weist die kleinste Fläche auf. Weibchen und beide Geschlechter miteinander haben eine größere potentielle Verbreitung.

Neigung und Seehöhe sind die Prädiktoren mit dem höchsten Erklärungswert. Das ist nicht sehr typisch für den Luchs, der durchaus in tiefgelegenen Wäldern lebt. Aber im Untersuchungsgebiet war Wald, als Ergebnis anthropogener Eingriffe, an hochgelegene Hanglagen gebunden. Dieses Faktum unterstreicht den lokalen Charakter des Modells und demonstriert, daß Variablen nicht notwendigerweise eine biologische Bedeutung für die untersuchte Tierart haben müssen. Zum Evaluieren des Modells wurde ein  $\rightarrow$ ROC-plot durchgeführt. Die dadurch erhaltenen AUC-Werte bestätigen die Gültigkeit des Modells.

Außerdem wurde –sozusagen als zusätzliche Evaluierung - das Überleben junger Luchse (Daten aus telemetrischen Ortungen 7 anderer, nicht territorialer Tiere) in Abhängigkeit von deren Habitatwahl untersucht. Jene subadulten Tiere die über – die vom Modell ausgewiesenen – Korridore in geeignetes Habitat auswanderten, überlebten. Einige beerbten ihre Mutter, und konnten so im Inneren des Untersuchungsgebietes ihr Leben fristen. Tiere die kein Territorium besetzen konnten oder sich in suboptimalem Lebensraum aufhielten, gingen ein.

### 6.3 Autologistisches Modell

Augustin et al. (1996) haben in Schottland ein autologistisches Modell zur räumlichen Verteilung von Rothirschen angepaßt. Als Ausgangspunkt wurde ein bereits

für diese Region entwickeltes logistisches Modell benutzt, das um einen "autokovariaten" Term erweitert wurde.

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 \text{altitude}_i^2 + \beta_2 \text{northing}_i + \beta_3 \text{mires}_i + \beta_4 \text{easting}_i + \beta_5 \text{pine}_i + \beta_6 \text{autocov}_i$$

Einfließende Variablen waren demnach Seehöhe, Lage im Koordinatensystem, Sümpfe und Kiefer. Nach dem Anpassen des autologistischen Modells, leisteten die Prädiktoren *northing*, *easting* und *mires* keinen signifikanten Beitrag zur Erklärung der Vorkommenswahrscheinlichkeit mehr. Diese Variablen haben also ins ursprüngliche logistische Modell als Stellvertreter für die räumliche Autokorrelation der *presence/absence*-Daten Eingang gefunden (schrittweise vorwärts Verfahren).

Interessant ist, daß das neu angepaßte autologistische Modell besser in der Vorhersage der Verteilung (*clustering*) der Hirsche war, das ursprüngliche logistische Modell die Anzahl der besetzten Zellen aber besser vorhersagen konnte. Dies läßt sich damit erklären, daß die Schätzung der logistischen Regression die durchschnittliche Reaktion auf die Habitatbedingungen über die ganze Region wiedergibt, während der autologistische Ansatz lokale Variationen berücksichtigt, die sich aus den jeweiligen Nachbarzellen ableiten.

## 6.4 ENFA

Hirzel et al. (2002) wenden ENFA auf die gesamte Steinbock-Population der Schweiz an. Der Steinbock wird dort seit seiner Wiederansiedlung sorgfältig überwacht, sodaß zuverlässige *presence*-Daten vorliegen. Die weitere Ausbreitung der Art wird durch die starke Fragmentierung ihres Lebensraums behindert, sodaß vermutlich noch nicht alle geeigneten Habitate besiedelt sind. Das Nichtvorkommen der Spezies deutet also nicht zwangsweise auf gering- bzw. ungeeignetes Habitat hin. Diese Tatsache legt die Verwendung eines Verfahrens nahe, das nicht auf *absence*-Daten angewiesen ist.

Die gesamte Fläche der Schweiz wurde in 1ha große Zellen aufgeteilt. Es wurden 34 ökogeographische Variablen (Seehöhe, Neigung, Exposition, Anteil der Fläche mit einer Neigung > 30° in einem Kreis mit 1200 m Radius, Abstand zur nächsten Siedlung, ...) verwendet. Topographische Daten flossen direkt ins Modell ein, Häufigkeits- und Abstandsdaten mußten erst für jede Zelle abgeleitet

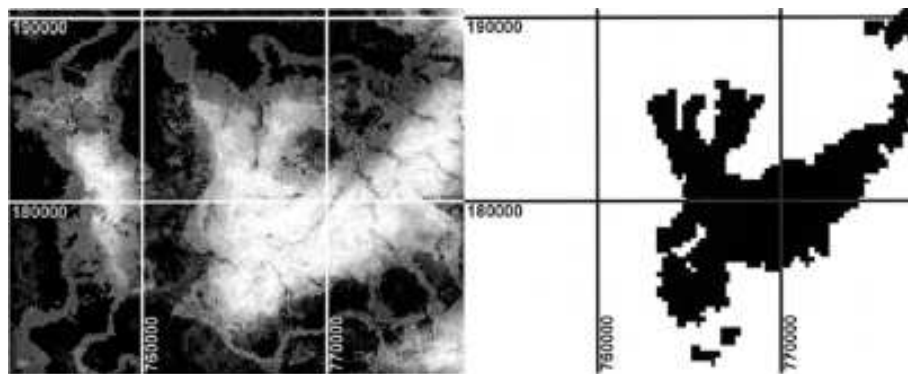
werden.

Der Datensatz wurde vor der Modellanpassung nach dem Zufallsprinzip in 2 annähernd gleich große Datensätze aufgeteilt. Einer wurde zur Modellentwicklung herangezogen, der zweite zur Validierung.

Die ENFA ergab eine Gesamt-Marginalität von  $M = 1,1$  und für die Gesamt-Spezialisierung einen Wert von  $S = 2,2$ . Das führt die Autoren zu dem Schluß, daß sich das Habitat des Steinbocks drastisch von den durchschnittlich in der Schweiz vorherrschenden Verhältnissen unterscheidet und daß der Steinbock unter ziemlich eingeschränkten Bedingungen vorkommt.

Die Marginalitätskoeffizienten zeigen, daß das Vorkommen des Steinbocks an steile, felsige Hochlagen mit reichlich Weideangebot gebunden ist. Der Steinbock meidet bewaldete Flächen und anthropogene Einflüsse. Exposition, Schnee und Wasser haben nur marginale Auswirkungen. Die Interpretation der Faktoren bzw. ihrer Koeffizienten steht somit in Einklang mit den Erfahrungen und Beobachtungen der Fachleute und den in der Literatur zitierten relevanten ökogeographischen Variablen.

Die auf der Grundlage von 5 Faktoren erstellte Habitateignungskarte weist Flä-



Links die vom Modell generierte Karte (geeignetes Habitat in weiß). Rechts die tatsächlichen Beobachtungen (in schwarz). Für die kleine weiße Fläche ganz links gibt es keine Entsprechung in der rechten Karte. Entweder die Fläche ist aufgrund ihrer isolierten Lage (noch) unbesiedelt oder sie ist für eine *viable* Population zu klein.

(Hirzel et al., 2002)

Abbildung 6.2: Ausschnitt aus Habitateignungskarte für Steinbock

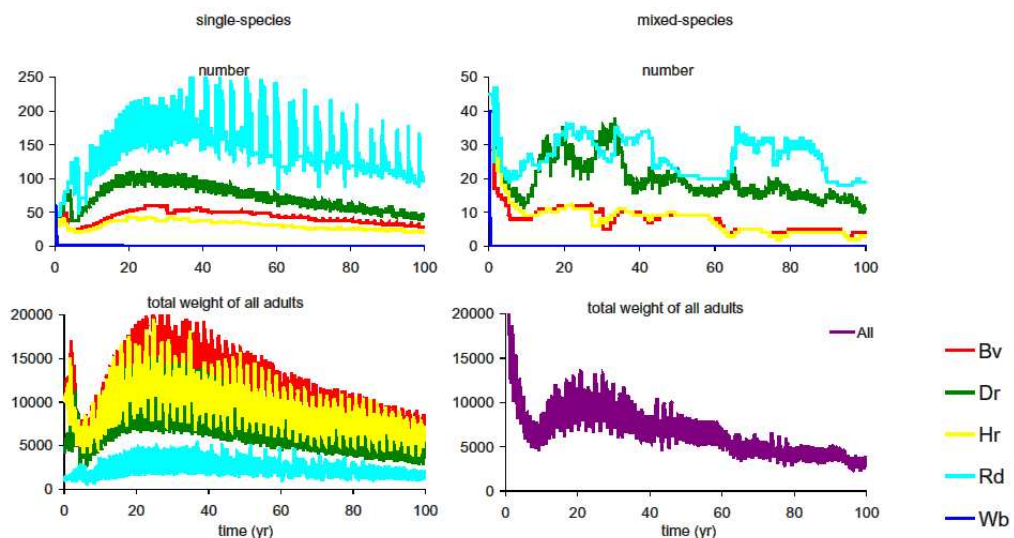
chen als geeignetes Habitat aus, auf denen der Steinbock nicht beobachtet worden war (siehe Abb. 6.2). Diese *false positives* geben keinen Aufschluß über die Modellgüte. Das sind Flächen, die wahrscheinlich aufgrund ihrer isolierten Lage (noch) nicht besiedelt worden sind oder zu klein sind, um langfristig eine überle-

bensfähige Population zu tragen.

## 6.5 FORSPACE

Das Modell FORSPACE wurde in einem Naturschutzgebiet (Imbos) der Niederlanden auf eine 200ha große Fläche angewandt, um die Auswirkungen von Beweidung (sowie Verbiß) und Feuer auf die Vegetation zu analysieren. Dazu wurden 4 Szenarien durchgespielt:

- ungestörte Entwicklung der Vegetation
- Beweidung ohne Feuereinfluß: alle Pflanzenfresserarten (Reh, Rot- und Schwarzwild, Rind, Pferd) gemeinsam, bzw. jede einzeln
- Feuereinfluß ohne Beweidung (unterschiedliche Frequenzen)
- Beweidung und Feuereinfluß: alle Pflanzenfresserarten gemeinsam bei unterschiedlichen Frequenzen.



Anzahl der adulten Tiere und Summe ihrer Körpergewichte für die verschiedenen Szenarien: links ausschließliches Vorkommen (jeweils nur eine Art), rechts gleichzeitiges Vorkommen der Arten Rind Bv, Pferd Hr, Rothirsch Dr und Rehwild Rd (Wildschwein Wb).

(Kramer et al., 2001)

Abbildung 6.3: Gewicht und Anzahl von Pflanzenfressern

Schwarzwild wurde nicht weiter berücksichtigt, da sich bald herausstellte, daß das untersuchte Gebiet keine andauernde Präsenz dieser Tierart garantieren konnte (zu wenig offenes Gelände, zu seltene Eichel-Mast).

Mit steigendem Körpergewicht (Reh-, Rotwild, Pferd, Rind) wird die Fluktuationen

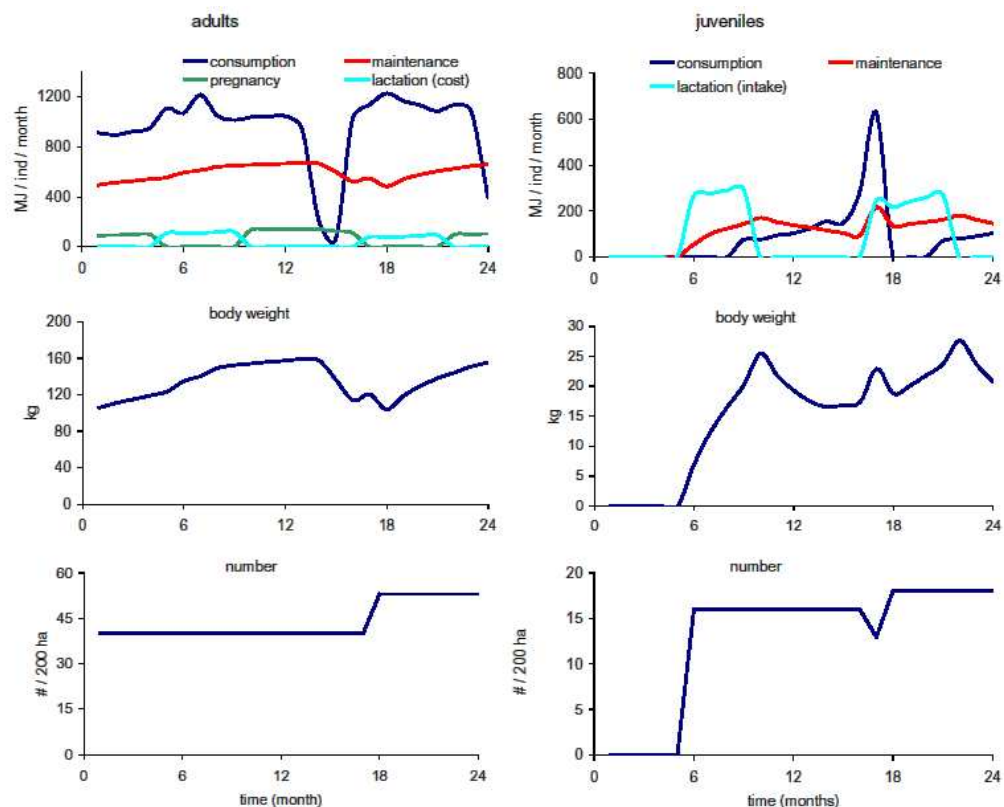
gedämpft und die Dichte geringer (siehe Abb. 6.3).

Summiert man allerdings das Körpergewicht der erwachsenen Tiere, so ergeben sich für die schwereren Arten die höchsten Werte.

Bei gemeinsamem Vorkommen aller Pflanzenfresser fluktuiert auch die Population der schwereren Arten stärker. Das Gesamtgewicht aller erwachsenen Tiere ist bei "alle Pflanzenfresser" nur im Vergleich zu "nur Rehwild" höher. Alle anderen Arten erreichen, so sie alleine vorkommen, höhere Körpergewichtssummen.

Rehwild erreicht im Modell – wenn keine der anderen Arten vorkommt – extrem hohe Dichten (durchschnittlich 75/100ha), aber es fluktuiert entsprechend stark (von unter 50 bis 250 auf 200ha).

Abbildung 6.4 zeigt die Energiebilanz eines adulten weiblichen Rotwildtieres über



Entwicklung von Energiebilanz, Körpergewicht sowie Anzahl der Individuen adulter und juveniler Rothirsche über einen Zeitraum von 24 Monaten aus der FORSPACE-Simulation für das Naturschutzgebiet Imbos.

(Kramer et al., 2001)

Abbildung 6.4: Rotwild im Jahresverlauf

einen Zeitraum von 24 Monaten.

Die Energieaufnahme zeigt einen deutlichen Abfall in den Wintermonaten. Dieser stimmt zeitlich mit vermindertem Energiebedarf für die Erhaltung überein, was

wiederum mit dem geringeren Körpergewicht zusammenhängt. Die Kosten für Laktation und Trächtigkeit fluktuieren natürlich genau entgegengesetzt und liegen das ganze Jahr über unter dem Bedarf an Erhaltungsenergie.

Bei Jungtieren wird während der ersten Lebensmonate der Energiebedarf aus der Muttermilch gedeckt, später durch Beweidung. Die Umstellung von Muttermilch auf Grünfutter schlägt sich auf das Körpergewicht der juvenilen Tiere nieder. Der starke Abfall an Energieverbrauch ist modellbedingt: mit dem Erreichen des zweiten Lebensjahres wechseln die Jungtiere in die Kohorte der 2-jährigen Adulten. Auffällig ist auch, daß sich der Ausfall einiger Jungtiere sofort auf das durchschnittliche Körpergewicht auswirkt, wahrscheinlich weil dadurch mehr Futter zur Verfügung steht.

Feuer beeinflusst die Dynamik dahingehend, daß bei gemeinsamem Vorkommen aller Pflanzenfresserarten Rind und Pferd, selbst bei niedriger Frequenz, keine nachhaltig lebensfähige Population aufrecht erhalten können. Bei niedriger Frequenz (Ansammlung von brennbarem Material) brennen größere Flächen ab, was zur Folge hat, daß nicht genügend Nahrung zur Verfügung steht.

## 6.6 Expertensystem

Partl (2001) hat ein Expertensystem erstellt ("Wildökologie - Waldverjüngung", WIFES), das in der forstlichen Praxis zur Vermeidung von Wildschäden als Planungs- und Entscheidungsgrundlage eingesetzt werden soll. Das System wurde auf der Grundlage wissenschaftlich fundierter Erkenntnisse aus der Fachliteratur, Daten aus Erhebungen und Expertenurteilen entwickelt und mit einem GIS kombiniert. Es beinhaltet sowohl ein Habitatqualitätsmodell für Rehwild als auch ein Wildschadensprädispositionsmodell.

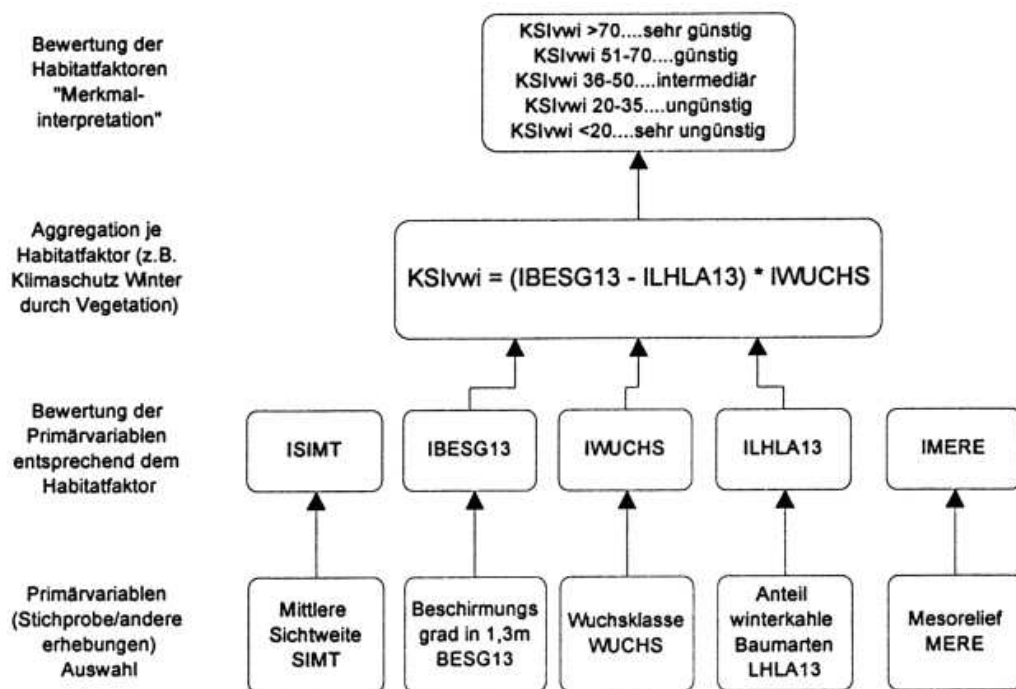
Die dabei verwendeten Daten stammen aus 4 Fallstudien, durchgeführt im Rahmen der Forschungsinitiative gegen das Waldsterben (FIW) und vom Spezialforschungsbereich Waldökosystemsanie rung (SFB). Damit standen Parameter aus permanenten (forstlichen) Stichprobeninventuren (Geländemorphologie, Bodenzustand, Waldstruktur, Verbißeinfluß) zur Verfügung. Auf jedem Stichprobenpunkt wurde zudem der wildökologische Bestandestyp angesprochen (→WÖBT) sowie indirekte Wildnachweise (Losung, Wildwechsel, . . .) erhoben. Weiters flossen Daten aus Panseninhaltsanalysen (zum späteren Vergleich mit der errechneten Verbißpräferenz) und Protokolle zum Bejagungsaufwand und -erfolg ein.



Die Stichprobenpunkte wurden mittels hierarchischer → Clusteranalyse anhand der Habitatfaktoren Nahrung, Sicht- und Wetterschutz 4 Habitattypen zugeordnet. Dadurch wurde sowohl nahrungsunabhängigem Besiedlungsanreiz als auch dem Nahrungsangebot Rechnung getragen und die Grundlage für das spätere Wildschadensprädispositionsmodell geschaffen.

Die Habitattypen wurden anhand der indirekten Nachweise bezüglich ihrer Nutzung durch Rehwild analysiert. Dabei wurden folgende Aspekte untersucht:

- Habitatattraktivität als Anzahl der Wildnachweise in Abhängigkeit von Habitattyp, mittlerer Sichtweite und Randliniensituation,
- Verbißpräferenz und Verbißeinfluß von Rehwild auf die Waldvegetation,
- Wechselwirkungen zwischen räumlicher Verteilung der Jungwuchsflächen und Verbißeinfluß,
- Beobachtbarkeit des Rehwilds im jahreszeitlichen Verlauf (Bejagung).



(Partl, 2001)

Abbildung 6.5: Merkmalsinterpretation im Expertensystem WIFES

Das im Expertensystem enthaltene Habitatqualitätsmodell entspricht einem HSI-Modell. Als Eingangsgrößen für das Modell werden die Parameter aus den Fallstudien herangezogen. Diese Primärvariablen werden zunächst bezüglich der

wesentlichen Habitatfaktoren (Feind-, Klimaschutz, Wohnraum, Nahrung, Beunruhigung) bewertet und über mathematische Funktionen, entsprechend der unterstellten Beziehung zwischen den Parametern aggregiert (siehe Abb. 6.5). Als Ergebnis erhält man eine Kennziffer, die entsprechend dem aktuellen Wissensstand interpretiert wird. Diese Merkmalsinterpretationen für Feindschutz, Klimaschutz, Wohnraum und Nahrung werden über logische Regeln zu einem Habitatqualitätswert kombiniert.

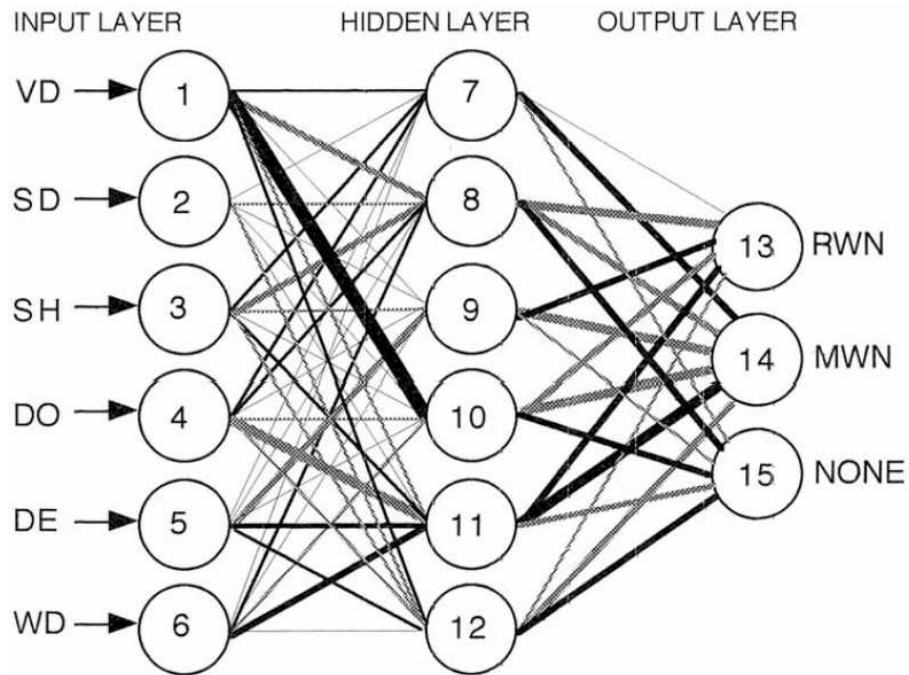
Das Wildschadensprädispositionsmodell entspricht im wesentlichen dem von Reimoser und Gossow (1996) vorgeschlagenen Ansatz: Prädisposition = nahrungsunabhängiger Besiedlungsanreiz / Nahrungsangebot. Der nahrungsunabhängige Besiedlungsanreiz wird als Kombination der Faktoren Einstand, Randlinien und Beunruhigung dargestellt. Maßnahmen zur Erhöhung des Nahrungsangebotes bzw. Herabsetzung der Habitatattraktivität können folglich die Wildschadensanfälligkeit herabsetzen. Völk (1999), z.B., konnte in Zusammenhang mit Schälschäden durch Rotwild aufzeigen, daß waldbauliche Maßnahmen wie Erhöhung des Anteils an mehrschichtigen Waldbeständen (Herabsetzen der Schälattraktivität, Erhöhung des schadensfrei nutzbaren Nahrungsangebotes) die Schadenshäufigkeit effizienter herabsetzen können, als jagdwirtschaftliche Maßnahmen dazu im stande wären.

Die Modellevaluierung des Expertensystems erfolgte in einem der Fallstudiengebiete. Tatsächliche Beobachtungen bestätigten die Simulationsergebnisse der integrierten Modelle. Der Autor stellt fest, daß die Anwendung des Expertensystems zu einer erfolgreichen Schadensprophylaxe beitragen kann.

## 6.7 ANN

Özesmi und Özesmi (1999) haben in den USA das Habitat von zwei Sumpfbrütern (Rotschulterstärling (*Agelaius phoeniceus*) und Sumpfsaunkönig (*Cistothorus palustris*) modelliert, und dabei zwischenartliche Interaktionen analysiert. Als *presence*-Daten wurden Nistplätze (RWN und MWN) verwendet, als Habitatvariablen (6) wurden Wassertiefe (WD), Abstand zu offenem Wasser (DO) respektive Ufer (DE) und Höhe (SH), Dichte (SD) sowie Beständigkeit (VD) der Ufervegetation herangezogen.

Das Modell wurde als  $\rightarrow$ BPN mit einem *hidden layer* konzipiert. Die Eingangswerte wurden  $\rightarrow$ standardisiert, die Initialwerte der Gewichte wurden zufällig



(Özesmi und Özesmi, 1999)

Abbildung 6.6: BPN für Rotschulterstärling und Sumpfzaunkönig

auf Werte  $\pm 0,1$  gesetzt. Im Inneren wurde als Aktivierungsfunktion eine symmetrische logistische Funktion mit dem Bereich  $-0,5$  bis  $+0,5$  benutzt, an den Outputeinheiten eine asymmetrische logistische Funktion mit einem Bereich von  $0$  bis  $1$  ( $\equiv$  Wahrscheinlichkeiten). Dabei kam NetProp3<sup>1</sup> zum Einsatz. Dieses Programm benutzt einen Teil des Datensatzes zum Trainieren des ANN, den anderen zum Validieren. Validieren hat hier einen anderen Sinn: Die Prozedur *autotrain* soll verhindern, daß das Netzwerk zwar den Trainingsdatensatz gut beherrscht, das "Erlernte" aber nicht verallgemeinern, also auf neue Daten anwenden, kann. Das Validieren im eigentlichen Sinne erfolgte mit einem Datensatz, der an in einem anderen Uferabschnitt aufgenommen worden war.

Die Autoren der Studie gehen auch auf die Kritik ein, ANNs wären reine *black box* Modelle, die zwar gute Vorhersagen liefern, aber keine weitere Interpretation zu-

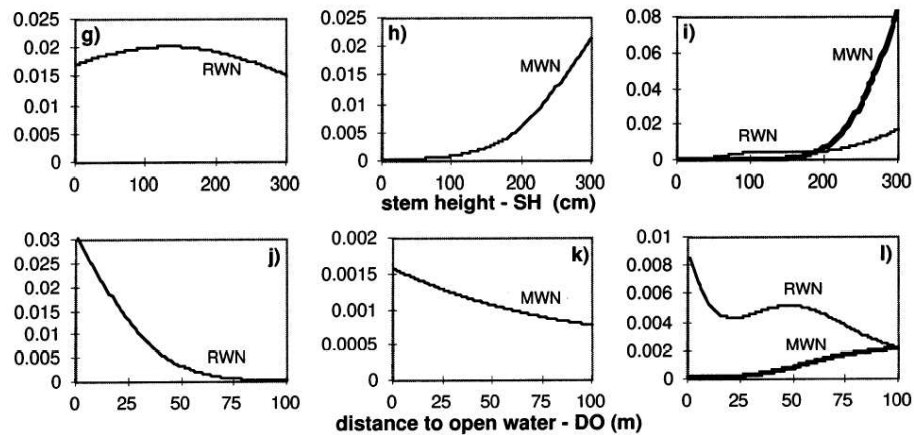
<sup>1</sup>Nevada backPropagation, OpenSource-Software (GPL), ist ein "*feedforward backpropagation multilayer perceptron simulator*", oder, um dem ganzen den Schrecken zu nehmen, ein "*multivariate nonlinear regression program*". Es kann unter <http://brain.cs.unr.edu/publications/nevprop.zip> im Quellcode für UNIX heruntergeladen werden. Besagte zip-Datei enthält auch eine Windowsversion im Binärformat. Auf <ftp://ftp.scs.unr.edu/pub/cmbr/private/wormhole/npmac.sea> gibt es eine (ältere) Mac-Version.

lassen und die dem System zugrundeliegenden Mechanismen blieben unverstanden. Sie finden, daß bei hinlänglichem Verständnis der Ökologie der untersuchten Spezies es durchaus möglich ist, die Ergebnisse eines ANN zu interpretieren. Zu diesem Zweck führen sie drei Argumente ins Feld: NIDs (siehe unten), Relevanz- und Sensitivitätsanalyse.

NIDs (*Neural Interpretation Diagram*) sind nichts anderes als eine leicht veränderte Art, ein ANN darzustellen (siehe Abb. 6.6). In einem NID werden die Verbindungen zwischen den Einheiten entsprechend ihrer Gewichtung mit mehr oder weniger starken Linien dargestellt. Das Vorzeichen des Gewichts spiegelt sich in der "Farbgebung" wieder: schwarze Linien stehen für positive Signale (Exzitation), graue Linien für negative Signale (Inhibition). Betrachtet man den *hidden layer*, kann man das Zusammenspiel der Variablen und ihren Beitrag zum Endergebnis (hier Nist-Wahrscheinlichkeit) interpretieren. Man kann z.B. erkennen, daß durchschnittliche Vegetationsbeständigkeit, Halmhöhe und Entfernung zu offenem Wasser die Wahrscheinlichkeit für den Sumpfzaunkönig positiv beeinflussen (Einheit 7 in Abb. 6.6). Hohe Entfernung vom Ufer, Nähe zu offenem Wasser und tiefes Wasser begünstigen vor allem das Auftreten des Sumpfzaunkönigs, aber auch des Rotschulterstärlings (siehe Einheit 11 in Abb. 6.6)

Um den Beitrag der einzelnen Variablen an der Attraktivität eines Lebensraumes abschätzen zu können, werden sie auf ihre Relevanz hin untersucht. Die Relevanz einer Variable ergibt sich aus der Quadratsumme ihrer Gewichte, dividiert durch die Quadratsumme der Gewichte aller Variablen. Variablen mit höheren Gewichten entlang ihrer Querverbindungen haben demnach einen stärkeren Einfluß. Für den Rotschulterstärling war dies der Abstand zum offenen Wasser, gefolgt von Dauerhaftigkeit der Vegetation und Dichte der Ufervegetation (DO, VD, SD, DE, WD, SH). Beim Sumpfzaunkönig waren es Dauerhaftigkeit der Vegetation, Abstand zum Ufer und Abstand zum offenen Wasser (VD, DE, DO, WD, SH, SD). Wenn sich beide Arten den Lebensraum teilen, sieht die Sache etwas anders aus: am ausschlaggebendsten ist Dauerhaftigkeit der Vegetation, gefolgt von Abstand zu offenem Wasser respektive Ufer (VD, DO, DE, SH, WD, SD).

Schließlich wurde noch eine Sensitivitätsanalyse durchgeführt: Alle Variablen werden auf ihren Mittelwert gesetzt, nur die analysierte wird variiert, nämlich von ihrem Minimum bis zu ihrem Maximum, und gegen die Nist-Wahrscheinlichkeit aufgetragen. So ist für den Rotschulterstärling eine Vegetationshöhe von 130cm und eine Vegetationsdichte von  $\approx 160$  Halmen/m<sup>2</sup> optimal. Mit zunehmender Ufe-



(Özesmi und Özesmi, 1999)

Abbildung 6.7: Sensitivitätsanalyse der Habitatvariablen

rentfernung, abnehmender Entfernung zu offenem Wasser und zunehmender Wassertiefe nimmt die Brut-Wahrscheinlichkeit zu. Der Sumpfzaunkönig bevorzugt höhere und dichtere, weniger beständige Vegetation. Für die anderen Variablen gilt mehr oder weniger das selbe wie für den Rotschulterstärling. Wenn die beiden Arten sich den Lebensraum teilen, verschieben sich die Präferenzen: Optima verschieben sich, flachen ab, neue zeichnen sich ab, Maxima und Minima werden ausgeprägter usw. (siehe Abb. 6.7). Und doch ist es nicht so, daß sich die beiden Spezies aus dem Wege gingen.

## 6.8 Classification trees

Debeljak et al. (2001) haben in Slowenien (Karstgebiet) für den Rothirsch ein Habitateignungsmodell entwickelt. Den Anstoß dazu lieferte unter anderem die – durch übermäßigen Verbiss – fast zum Stillstand gekommene Naturverjüngung. Die dazu verwendeten Daten stammen aus telemetrischen Beobachtungen (3 adulte Tiere, mehr als 70 Ortungen pro Tier) einerseits und aus einem bereits bestehenden GIS andererseits.

Das gesamte Untersuchungsgebiet erstreckt sich über 114.800ha und ist somit größer als die von den Tieren beanspruchten Territorien (bestimmt mittels *adaptive kernel method*). Das ist notwendig, weil das auf →machine learning basierende Modell sowohl auf positive als auch auf negative *Beispiele* angewiesen ist. Das GIS hat eine Auflösung von 200m. Jedes Pixel wird durch 9 Attribute (Seehöhe, Exposition, Neigung, Nähe zu Siedlung, Koniferenanteil, Bestockung, ...)

beschrieben.

Da sich bei einem der Tiere Winter- und Sommerlebensraum wesentlich unterscheiden, wurden insgesamt 4 "Habitate" bzw. Datensätze modelliert: Sommerhabitat des Ljubljana-Tieres, Winterhabitat des Ljubljana-Tieres, Habitat des Kočevje Tieres #1 und Habitat des Kočevje Tieres #2.

Für jeden Datensatz wurden *decision trees* induziert, die "vorhersagen" sollten, ob ein Punkt zum *home range* eines Tieres gehört (und damit implizit geeignetes Habitat darstellt) oder nicht. In Abbildung 6.8 ist der *main branch* (enthält die meisten positiven Fälle) für das Sommerhabitat des Ljubljana-Tieres zu sehen. Im

```
Elevation > 594:
: ...Elevation <= 712:
:   ...PercConifers > 59:
:     ...Wood <= 190:
:       ...ProxSet > 600:
:         ...ProxSet <= 1780:
:           ...Wood <= 110: InHomeRange
```

Hauptzweig (*main branch*) des für den Sommerlebensraum des Ljubljana-Tieres induzierten Entscheidungsbaumes (Output von C5, siehe Kapitel 5.10). Damit werden 49 der insgesamt 53 Punkte erfaßt.

(Debeljak et al., 2001, verändert)

Abbildung 6.8: Ljubljana-Tier Sommerlebensraum

Sommer hält sich das Tier in Seehöhen von 600 bis 700m auf, bevorzugt in bewaldeten Gebieten mit einem Koniferenanteil größer 60% und einer "Bestockung" von 110m<sup>3</sup>/ha in einer Entfernung größer 600m von Siedlungen.

Das Ergebnis sind demnach 4 individuelle(!) Habitatmodelle, nicht ein allgemeines Habitatmodell für Rotwild. Doch die Autoren postulieren – auf der Grundlage der gemeinsamen Eigenschaften der "maschinell" erstellten Modelle – ein allgemeines Modell, sozusagen aus der Sicht eines Wildtierexperten. Sie sind sich dabei durchaus bewußt, daß das Extrapolieren ausgehend von 3 Tieren auf eine ganze Population nur ein provisorisches Modell liefern kann.

Das allgemeine Modell kann im wesentlichen durch 5 *Attribute* beschrieben werden: Seehöhe, Nähe zu Siedlung, Waldvorkommen, Koniferenanteil und Bestockung. Für die beobachteten Querverbindungen zwischen diesen Attributen lassen sich durchaus plausible Erklärungen finden.

Die Seehöhe lag in den individuellen Modellen zwischen 413 (Ljubljana-Tier Winter) und 712m (Ljubljana-Tier Sommer). Das kann durch geeignete Nahrung und klimatische Bedingungen erklärt werden: Unter 400m wären die Tiere starkem

anthropogenem Druck (Siedlungen, landw. Nutzung) ausgesetzt. Oberhalb von 800m Seehöhe verändern sich die klimatischen Bedingungen drastisch, was unter anderem kürzere Vegetationsperioden zur Folge hat.

Der Abstand zur nächsten Siedlung lag zwischen 600 und 3102m. Das deutet darauf hin, daß sich die Tiere auch auf nicht bewaldeten Flächen aufhalten. Dies ist wahrscheinlich dem Aufsuchen alternativer Futterquellen (Felder, Bestandesränder) zuzuschreiben.

Aus den individuellen Modellen gehen zwei bevorzugte Bestandestypen hervor: ein gering bestockter Typ mit hohem Koniferenanteil und ein höher bestockter Typ mit geringerem Koniferenanteil. Der erste Typ (junger Nadelwald) wird bei strengen Wetterbedingungen aufgesucht (Winterhabitat), denn er bietet bessere mikroklimatische Bedingungen, Deckung (was Laubbäumen nicht der Fall wäre) und eine spärliche Nahrungsgrundlage (Verbiß und Schäle). Der zweite Typ (ältere Laubbaumbestände) ist typisches Sommerhabitat, speziell für führende Tiere. Er bietet eine gut entwickelte Grasschicht und Gestrüpp, insgesamt also bessere Futterbedingungen als der Nadelwald.

Das abgeleitete allgemeine Modell ist demnach aus wildbiologischer Sicht konsistent und liefert gute Hinweise auf Faktoren, die für die Habitateignung wesentlich sind. Trotzdem betonen die Autoren, daß das Modell durch die Qualität der Ausgangsdaten beschränkt ist. Für genauere Aussagen wären Angaben über Bestandesstruktur, -zusammensetzung, Gras- und Strauchschicht, Lage von Futterstellen und forstliche Maßnahmen (Zeitpunkt, Art, ...) notwendig. Auch müßten die Telemetrieprotokolle vollständiger sein: Geschlecht, jahres- und tageszeitliche Habitateigenschaften sollten berücksichtigt werden.

Auf jeden Fall, so die Autoren, lassen sich aus den gewonnenen Erkenntnissen Richtlinien für die Planung forstlicher Maßnahmen ableiten.

## 6.9 Diskriminanzanalyse (DA)

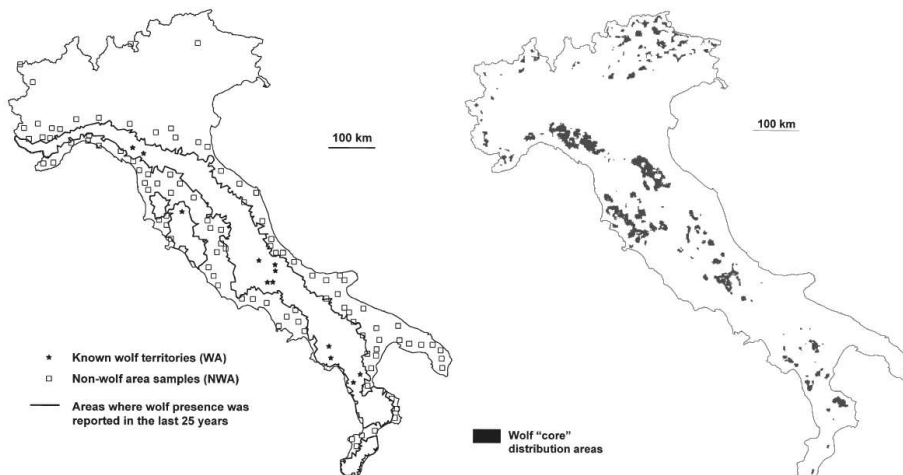
Corsi et al. (1999) haben für Italien im Rahmen des nationalen Schutzprogrammes für den Wolf ein Modell entwickelt, das auf Diskriminanzanalyse (→DA) und →Mahalanobis distance beruht. Das Modell zielt mehr darauf ab, räumliche Verbreitungsmuster zu analysieren, als die Habitateignung zu untersuchen.

Grundlage des Modells bildeten *presence/absence*-Daten, 13 Habitatvariablen und ein Datensatz mit Angaben über tot aufgefundene Wölfe. Die Ortungen toter

Wölfe wurde mit der Verteilung des (vom angepaßten Modell) als ungeeignet ausgewiesenen Habitats verglichen, sozusagen als zusätzliches Validierungswerkzeug herangezogen.

Die *presence*-Daten stammen aus 12 von Wölfen besetzten Streifgebieten (WAs) (telemetrische Ortungen, Schneespuren). Die *absence*-Daten hingegen stammen aus Gebieten, für die es zwar über die vorangegangenen 25 Jahre keinen Wolfsnachweis gibt, die sich aber in "Reichweite" eines migrierenden Wolfes befinden (Abstand <100km von bekannten Wolfsterritorien). Die Autoren gehen davon aus, daß es sich folglich um ungeeignetes Habitat handelt. In dieses Gebiet wurden 100 Kreisflächen (NWAs) in der Größenordnung eines durchschnittlichen Wolfsterritoriums platziert (siehe Abb. 6.9 links).

Bei der Wahl der Habitatvariablen griffen die Autoren auf ihre Erfahrungen betref-



Links: — aktuelles Wolfs-Vorkommen (Nachweise über die letzten 25 Jahre), \* Wolfs-Streifgebiete (WA), □ Nicht-Wolfgebiete (NWA). Rechts: schwarz eingefärbt das Wolfs-"Kern"-Gebiet (Ergebnis der Diskriminanzanalyse)

(Corsi et al., 1999)

Abbildung 6.9: Aktuelles Vorkommen und Kerngebiet

fend die grundlegenden Bedürfnisse des Wolfes zurück, mußten aber auch die Verfügbarkeit der Daten berücksichtigen. Von den ursprünglich 13 Habitatvariablen flossen ins endgültige Modell 9 ein (landwirtschaftl. genutzte Flächen, Wald, Siedlungen, Besiedlungsdichte, Straßendichte, Dichte von Mülldeponien, Anzahl der Huftierarten und 2 Indices zur Beschreibung räumlicher Gegebenheiten).

Die Variablen wurden normalisiert (verschiedene Transformationen) und anschließend einer schrittweise vorwärts Diskriminanzanalyse unterzogen. Das fertig angepaßte Modell wurde schließlich auf die ganze italienische Halbinsel über-



tragen, dabei wurden 5,7% der Fläche als für den Wolf relevant ausgewiesen (siehe Abb. 6.9 rechts).

Unabhängig von der DA wurde versucht, mittels Mahalanobis Distanz potentielle Verbindungen zwischen den Wolfsgebieten zu beschreiben. Dabei mußte – sozusagen – die Entfernung eines gegebenen Standortes zu den besten Umweltbedingungen (Flächenschwerpunkt einer WA) als Index für die Qualität der Umwelt herhalten.

Die Autoren betonen, daß der gewählte Ansatz eher zum Identifizieren räumlicher Muster geeignet ist, als zum Identifizieren kritischer Habitatfaktoren. Auch sind die auf das Gebiet nördlich des Flusses Po bzw. auf den Alpenraum bezogenen Ergebnisse mit Vorsicht zu genießen.

Weiters geben sie zu bedenken, daß sich einer der Faktoren, der die Verbreitung des Wolfes wesentlich beeinflusst, nämlich die Einstellung des Menschen zum Wolf, nicht einfach in eine Variable verpacken läßt. Man könnte natürlich annehmen, sie verberge sich in anderen Variablen, wie z.B. Straßendichte oder Landnutzungsform. Das impliziert aber Dichteabhängigkeit: anthropogene Störungen nehmen mit der Siedlungsdichte zu. Die Annahme steht damit auf schwachen Beinen, denn die Einstellung des Menschen zum Wolf folgt nicht unbedingt diesem Zusammenhang, sondern hängt eher mit einschlägigen (Koexistenz-)Erfahrungen oder hartnäckigen Vorurteilen zusammen.

## 6.10 GAP

Erhaltung von Biodiversität – GAPs eigentliche Zielsetzung – bedeutet Erhaltung von Populationen, das wiederum impliziert Erhaltung von genügend großem Lebensraum um eine *minimum viable Population* (MVP) tragen zu können. Das durch GAP ausgewiesene Habitat läßt keine dahingende Interpretation zu: Es können weder Aussagen darüber gemacht werden, ob die einzelnen *patches* genügend groß sind, noch darüber ob sie in genügendem Maße vernetzt sind. Allen et al. (2002) haben GAP (für 10 Tierarten in 3 counties, South Carolina) deshalb um die Aspekte *minimum critical area* (MCA) und *functional connectivity* erweitert.

Eine MCA ist eine (Schätzungen der) Mindestfläche, die notwendig ist, um eine MVP zu erhalten. MCAs werden auf der Grundlage einer einfachen Gleichung ( $MCA = \frac{[(\text{home range area}) \times N_e]}{2}$ ) geschätzt. Die Werte für die Größe des Streifgebiete-

tes einer Art stammen aus von Fachleuten überprüfter Literatur.

$N_e$  steht für die effektive Populationsgröße, also die Anzahl der an der Reproduktion beteiligten Tiere, bzw. stellt einen Schätzwert für die MVP dar. Das Modell wurde mit 2 verschiedenen Werten für  $N_e$  durchgerechnet:

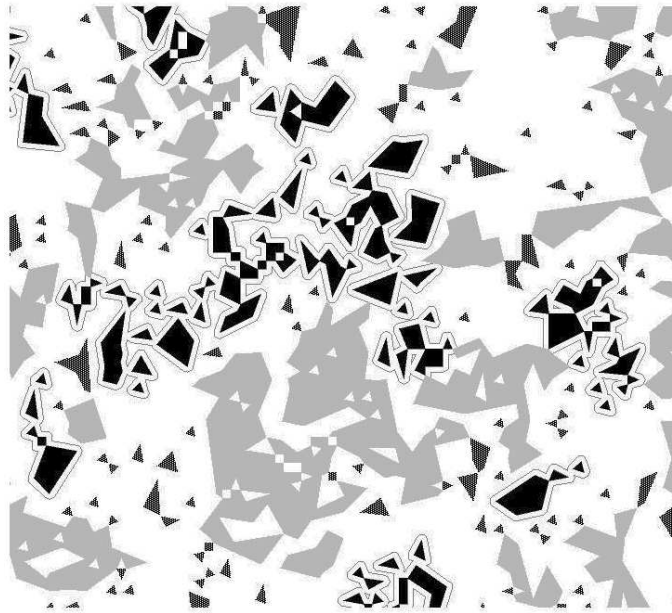
- mit  $N_e = 50$  um auszuschließen, daß stochastische demographische Schwankungen zum Austerben einer lokalen Population führen können.
- mit  $N_e = 500$  um auszuschließen, daß Inzucht und genetische Drift zu Genkonvergenz, also Diversitätsverlust führen.

Die Zahl 2 unter dem Bruch trägt dem *intersexual overlap* Rechnung, also der Tatsache, daß sich die Streifgebiete fortpflanzungsaktiver Tiere überlappen.

Ist der Lebensraum einer Spezies fragmentiert, kann durch Migration zwischen den lokalen Populationen der funktionale Zusammenhang (*functional connectivity*) aufrechterhalten bleiben, sofern die Individuen in der Lage sind, zwischen den *patches* zu wechseln. Die *functional connectivity* ändert sich sprunghaft, sobald die Entfernung zwischen den *patches* bzw. die Ausbreitungsfähigkeit einer Art einen kritischen Schwellenwert erreicht. Werte für die Ausbreitungsdistanzen (*dispersal distances*) der verschiedenen Tierarten wurden entweder der Literatur entnommen oder auf der Grundlage der Körpermasse geschätzt.

Um MCA und *functional connectivity* ins Modell zu integrieren wurden zunächst alle *patches*  $< MCA$  ausgeschlossen (gerasterte Flächen in Abb. 6.10). Im zweiten Schritt wurden alle *patches*  $\geq MCA$  mit einem Puffer in der Größenordnung der Ausbreitungsdistanz versehen und *patches*  $< MCA$ , die dadurch in den Ausbreitungsbereich eines größeren *patches* zu liegen kamen, wieder aufgenommen (graue Flächen). Um die restlichen *patches* wurde ebenfalls ein Puffer gelegt: Netzwerke aus *patches* die zwar individuell  $< MCA$  aber in Summe  $\geq MCA$  waren, wurden ebenfalls wieder aufgenommen (schwarze Flächen mit weißer Pufferzone).

Für jede der berücksichtigten 10 Säugerarten wurde je ein reines GAP, ein GAP mit MCA und ein GAP mit MCA und *functional connectivity* entwickelt. Das MCA-Modell wies die geringste Habitatfläche aus. Bei Berücksichtigung der *functional connectivity* waren die Flächen fast durchwegs gleich groß wie im reinen GAP-Modell. Dies aber nur bei größeren Arten (Schwarzbär, Rotluchs, ...). Für kleinere Arten (Sternmull, östliche Erntemaus) – mit entsprechend geringer Ausbreitungsfähigkeit – lagen die Werte der beiden modifizierten Modelle deutlich unter dem



Habitat der östlichen Erntemaus: *patches*  $\geq$  MCA (grau), *cluster* zusammenhängender *patches*  $<$  MCA und "Ausbreitungspuffer" (schwarz mit weißer Umrandung), restliche *patches*  $<$  MCA (gerastert).

(Allen et al., 2002, verändert)

Abbildung 6.10: *minimum critical areas and functional connectivity*

einfachen GAP-Modell. Wie man auch Abbildung 6.10 entnehmen kann, wird das verfügbare Habitat durch ein reines GAP-Modell u.U. also überschätzt.

Daraus läßt sich auch die unverhältnismäßig wichtige Rolle einzelner *patches* im Verbund erkennen. Sie garantieren den funktionalen Zusammenhang und damit die Erhaltung von (Meta-)Populationen.

## 6.11 RSF

Mace et al. (1999) haben im Nordosten der USA (Montana) ein Habitatmodell für den Grizzly entwickelt. Das besondere an diesem Modell ist, daß es als eines der wenigen die Auswirkungen der Jahreszeit auf die Habitatwahl berücksichtigt. Die Daten der telemetrischen Ortungen (8 adulte Weibchen) wurden auf 3 Jahreszeiten aufgeteilt. Kriterium für die Einteilung der Jahreszeiten (Frühling, Sommer, Herbst) waren Veränderungen im Freißverhalten des Bären.

Als Habitatvariablen wurden Straßendichte, Verkehrsaufkommen, Siedlungen, Lagerplätze, Seehöhe und *greenness* (Grünintensität, LANDSAT Thematic Mapper) berücksichtigt.

Für jede Jahreszeit wurde eine RSF ( $w(x) = \exp(\beta_1 x_1 + \dots + \beta_p x_p)$ , also eine logistische Regression) angepaßt. Da man an der kumulativen Wirkung aller Variablen auf die Habitatwahl interessiert war, wurden sämtliche ins Modell aufgenommen. (Damit nimmt man allerdings eine reduzierte Erklärungsfähigkeit des Modells in Kauf.) Die erhaltenen RSF-Werte wurden normiert (auf den höchsten Wert skaliert) und damit eine Karte erstellt. Aus dieser Karte ("*realized*" map) kann man praktisch die relative Wahrscheinlichkeit für das Vorkommen eines (weiblichen) Bären ablesen. Es wurde auch eine Karte ("*potential*" map) generiert, bei der die Werte der Variablen für anthropogene Einflüsse (Siedlungen, Straßen, ...) auf Null gesetzt worden waren. Die Differenz aus den beiden RSF-Werten wurde als Reduktion des potentiellen Habitats durch menschlichen Störungseinfluß interpretiert.

Aus den entstandenen Modellen läßt sich folgendes ableiten: Im Frühling halten sich Bären in tieferen Lagen auf, wo der Schnee früher schmilzt und die Vegetation freigibt. Die Tieflagen fallen auch mit den meisten anthropogenen Einflüssen zusammen, sodaß sich hier die stärkste Einschränkung des potentiellen Habitats ergibt. Im Sommer wechseln die Weibchen auch in mittelhohe Lagen, nutzen aber weiterhin die Tallagen und sind damit ebenso starken menschlichen Störungen ausgesetzt. Im Herbst ziehen sich die Bären in mittlere und Hoch-Lagen zurück und meiden dadurch starke anthropogene Einflüsse.

Da die telemetrischen Ortungen untertags erfolgten, kann allerdings nicht ausgeschlossen werden, daß die Tiere nachts (störungsarme Situation) doch auch tiefere Lagen nutzen. So konnte z.B. Fischer (2003) für Rothirsche nachweisen, daß sich die Gesamtaktivität, über den Jahresverlauf betrachtet, im Sommer stark zugunsten der Nachtaktivität verschiebt.

Bei Tieren, die sich bei der Wahl ihres Habitats nach den jahreszeitlichen Bedingungen richten, macht es also durchaus Sinn, für jede Jahreszeit ein eigenes Modell zu entwickeln.

## 6.12 Logistische Regression

Mladenoff et al. (1995) haben im Norden der USA (Northern Great Lakes) ein Habitatmodell für den Timberwolf angepaßt. Neben den sonst üblichen, zur Habitatbeschreibung herangezogenen Variablen, wurden hierbei auch Angaben zur Bevölkerungsdichte, Beutetierdichte, Straßendichte und Besitzverhältnisse be-

rücksichtigt bzw. verglichen. Ferner wurden Kennzahlen zur Beschreibung der Landschaftsstruktur herangezogen: *patch*-Größe, Randlinien, *fractal dimension* (Maß für die Unregelmäßigkeit/Rauheit einer Fläche), *landscape diversity*, *landscape dominance* und *contagion* (Maß für den Verklumpungsgrad).

Vor der Anpassung der logistischen Regression wurden die Variablen der Wolfsgebiete mit jenen der Nicht-Wolfsgebiete verglichen (t-Test). Dabei zeigte sich, daß in von Wölfen besiedelten Gegenden der Anteil an Mischwäldern und bewaldeten Feuchtgebieten höher war, geringer war der Anteil an landwirtschaftlich genutzten Flächen, Laubwäldern und größeren Seen. Straßen- und Bevölkerungsdichte waren im Wolfsgebiet niedriger, die Dichte der Beutetiere wies hingegen keinen signifikanten Unterschied auf. Das Wolfsgebiet unterschied sich vom Nicht-Wolfsgebiet auch bezüglich der Landschaftsparameter: Die *patches* wiesen einfachere Formen auf (geringere *fractal dimension*), waren weniger verklumpt (geringere *contagion*) und vielfältiger (höhere *landscape diversity* und niedrigere *landscape dominance*).

Von den ursprünglich 12 Variablen fanden, nach einer schrittweisen Variablenselektion, allerdings nur zwei Eingang ins Regressionsmodell.

Das Anpassen eines Modells für den Wolf wird durch sein komplexes Sozialverhalten und die weiträumigen Interaktionen zwischen den Populationen nicht gerade begünstigt. Zudem ist der Wolf als "Großräuber", an der Spitze der Nahrungspyramide angesiedelt, nicht an spezifische Vegetationstypen oder Ökosysteme gebunden. Diese Tatsache spiegelt sich im Modell, das nur Straßendichte und *fractal dimension* berücksichtigt, wider.

## 7. Modellvergleiche

Nur wenige Autoren wagen den Vergleich verschiedener Ansätze, meistens eher um die Leistungsfähigkeit des persönlich bevorzugten Ansatzes hervorzuheben als diesen in Frage zu stellen.

Einige der hier angeführten Vergleiche beruhen auf effektiv durchgeführten Studien, d.h. die diversen Ansätze werden anhand konkreter Datensätze verglichen, andere sind eher theoretische Auseinandersetzungen.

### 7.1 ENFA, GLM bzw. logistische Regression

Der Vergleich (statistischer) Modellierungs-Techniken ist eigentlich nur dann möglich, wenn die "Realität" bekannt ist und erfaßt werden kann (*where 'truth' is known*) (Austin, 2002a). Das Generieren derartiger "wahrer" Daten setzt voraus, daß die Reaktion einer Art auf die diversen Umweltgradienten sowie die Prozesse, welche Multikollinearität zwischen diesen Prädiktoren bedingen, bekannt sind. In diesem Sinne ist auch Austin (2002b) zu verstehen, wenn er davon spricht, daß das Ignorieren ökologischer Zusammenhänge die Anwendung statistischer Modelle in der Ökologie einschränkt.

Hirzel et al. (2001b) erzeugen tatsächlich künstliche Daten aufgrund expliziter ökologischer Theorien: Sie schaffen eine virtuelle Spezies, die vollkommen durch ihre ökologische Nische definiert ist. Damit ist es möglich, eine bzw. die "wahre" Habitateignungskarte zu generieren. Diese bildet die Grundlage für den Vergleich von zwei Habitateignungs-Modellen, nämlich ENFA und GLM.

Es werden drei Szenarien – mit jeweils zwei Stichprobengrößen – simuliert: *spreading* (Population besiedelt unbesetzten, geeigneten Lebensraum), *equilibrium* (Populationsdichte ist so hoch, daß sämtliches geeignetes Habitat besetzt ist) und *over-abundance* (Populationsdichte derart hoch, daß auch ungeeignetes Habitat besiedelt wird).

Im *spreading*-Szenario war ENFA effizienter als GLM (höheres  $\rightarrow$  Bestimmtheitsmaß  $R^2$ ). Beim *equilibrium*-Szenario war kein signifikanter Unterschied erkennbar. Beim *over-abundance*-Szenario mit kleiner Stichprobe lieferte GLM wesentlich bessere Ergebnisse, bei größerer Stichprobe verschwanden

den die Unterschiede.

ENFA erwies sich als sehr robust, was Quantität (Stichprobengröße) und Qualität (Zuverlässigkeit der *presence/absence* Daten) der Eingangsvariablen anbelangt. Die Autoren kommen zu Schluß, daß ENFA speziell dann geeignet ist, wenn die Qualität der Daten dürftig bzw. unbekannt ist. GLM liefert geringfügig bessere Ergebnisse, wenn die verfügbaren *presence/absence* Daten ausreichend zuverlässig sind.

Hirzel et al. (2002) vergleichen ENFA und logistische Regression. ENFA hat den bereits angesprochenen Vorteil, nicht auf *absence*-Daten angewiesen zu sein. Allerdings impliziert dies, daß die *presence*-Daten unverfälschte Stichproben der aktuellen Verteilung darstellen. Und das, so vermuten die Autoren, ist nicht immer der Fall.

Für stabile Populationen gut untersuchter Arten ist den klassischeren Ansätzen – wie eben der logistischen Regression – der Vorzug zu geben, da diese in der Lage sind, aus *absence*- bzw. Häufigkeitsdaten relevante Information zu extrahieren. Es bedarf aber einer Abwägung der Vorteile, die durch die Einbeziehung von *absence*-Daten erwachsen, einerseits und des dadurch entstehenden Mehraufwandes, bzw. der geringen Qualität, andererseits.

Weiters geben die Autoren zu bedenken, daß die Verfahren zur Variablenselektion (siehe 4.4.5) sehr sensibel sind, was den gewählten Algorithmus und die Reihenfolge der Eingabe angeht. Im Gegensatz dazu werden bei der FA keine Variablen verworfen, sondern unterschiedlich gewichtet.

## 7.2 MR und ANN

Lek et al. (1996) vergleichen die Vorhersage-Fähigkeit von multipler Regression (MR) und neuronalen Netzwerken (ANN) für die Schätzung der Dichte von Forellen-Laichplätzen.

Wegen der unterschiedlichen Größenordnung der Eingangsvariablen wurden diese standardisiert ( $\rightarrow$ Standardisieren). Beide Methoden wurden sowohl mit transformierten (Potenzierungs-Funktionen) als auch untransformierten Werten durchgerechnet. Die Notwendigkeit der  $\rightarrow$ Transformation ergibt sich aus der nicht-linearen Beziehung zwischen den Variablen. Damit können zwar ANNs aber nicht MR umgehen. Als Indikatoren für die Leistungsfähigkeit der Modelle wer-

den →Korrelationskoeffizient ( $r$ ) und →Bestimmtheitsmaß ( $R^2$ ) zw. beobachteten und geschätzten Werten, der Anstieg (*slope*  $b$ ) der Regression(-sgeraden) sowie Verteilung und Korrelation der →Residuen herangezogen.

In die MR fanden (nach einem schrittweisen Ausleseverfahren) nur 4 erklärende Variablen Eingang, da sich nur deren Regressions-Koeffizienten signifikant von 0 unterschieden. Bei den ANNs flossen hingegen alle zur Verfügung stehenden Variablen ein. Die Variante ANN mit untransformierten Variablen lieferte die besten Ergebnisse.

Höheres  $R^2$ , bessere Vorhersage von kleinen Werten, Normalverteilung der Residuen und deren Unabhängigkeit von der (Größe der) vorherzusagenden Variable lassen die Autoren zum Schluß kommen, daß ANNs "performanter" sind.

### 7.3 DA, ANN und logistische Regression

Manel et al. (1999) untersuchen das Vorkommen eines Flußvogels anhand eines großen Datensatzes (Indisches und Nepalesisches Himalaya-Gebiet) und vergleichen dabei *multiple discriminant analysis* (→DA), neuronale Netzwerke (ANN) und →logistische Regression.

ANN brauchen zwar mehr Rechenzeit, können aber mit nicht-linearen Zusammenhängen umgehen. Allerdings ist es ohne zusätzlichen Aufwand nicht möglich, kausale Zusammenhänge zu identifizieren. Logistische Regression und Diskriminanz-Analyse beanspruchen weniger Rechenzeit und lassen Rückschlüsse auf kausale Zusammenhänge zu. Dadurch ist es möglich, überprüfbare Hypothesen zu entwickeln.

Die Autoren kommen zum Schluß, daß ANNs "konventionelle" Ansätze nicht "out-performen", sofern diese richtig angewandt werden.

### 7.4 ANUCLIM, DOMAIN, GAM, GLM und GARP

Elith (2000) vergleicht 5 Ansätze, nämlich →ANUCLIM (Synonym für →BIOCLIM), →DOMAIN, →GAM, →GLM und →GARP. Während GLMs und GAMs sowohl auf →presence bzw. Abundanz-Daten als auch auf *absence*-Daten für Art angewiesen sind, können die anderen Ansätze auch nur mit dem *presence*-Datensatz umgehen. GLMs und GAMs sind, unter den verglichenen Ansätzen, die einzigen mit integrierten anerkannten Prozeduren zur Selektion



der eingehenden Variablen; die anderen Ansätze greifen auf Experten-Urteile, →PCA und univariate Regression zurück. GLMs und GAMs integrieren Methoden zur Berücksichtigung von Wechselwirkungen zwischen den Variablen und liefern statistische Schätzungen zu Fehler und Abweichung. ANUCLIM und DOMAIN basieren auf einfachen Konzepten, doch mangelt es an umfassender Dokumentation. Die komplexen statistischen Konzepte bei GLMs und GAMs werden durch umfangreiche Dokumentation und Literatur ausgeglichen. GARP hat zwar eine benutzerfreundliche Schnittstelle, ist aber schwer zu interpretieren, und es gibt wenige veröffentlichte Anwendungen.

Die Modelle wurden auf 7 Spezies angewandt. Es ergaben sich keine wesentlichen Unterschiede im Modellierungs-Erfolg zwischen den verschiedenen Modellen. Nur ANUCLIM war etwas schwächer, wohl einfach aufgrund der Tatsache, daß weniger Variablen einfließen. Die Autorin macht u.a. die Qualität der Daten dafür verantwortlich und schließt daraus, daß es sich unter Umständen auszahlt, mehr Aufwand in bessere Daten zu investieren, als in hochentwickelte Modelle. Sie erklärt sich die bei Forschern vorherrschende Präferenz für GLM und GAM damit, daß diese leichter zu interpretieren und somit auf ökologische Zusammenhänge umgelegt werden können.

## **7.5 DFA, CART, logistische Regression und Mahalanobis Distanz**

Dettmers et al. (2002) vergleichen →logistische Regression, →DFA, →Mahalanobis distance, →CART und →PCA. Modelliert wurde das jeweilige Habitat (definiert durch 26 Variablen, jeweils 4 Modelle) von 6 Vogelarten. Um die Treffsicherheit der verschiedenen Ansätze zu vergleichen, wurden die für Tennessee angepaßten Modelle auf zwei unabhängige Datensätze (Georgia und Virginia) angewandt.

Das DFA-Modell konnte, auf den Trainingsdatensatz angewandt, mehr als 90% der Beobachtungen richtig einordnen. Auf die beiden unabhängigen Datensätze bezogen konnte es nur mehr weniger als 70%, für die meisten Arten weniger als 50% der Beobachtungen richtig vorhersagen. CART lieferte für die Datensätze Georgia und Virginia bei zwei Vogelarten über 90% bzw. 85% korrekte Vorhersagen, für die anderen Arten lag die Treffsicherheit unter 65%. Logistische

Regression und Mahalanobis Distanz lieferten für alle Vogelarten weniger als 66% korrekte Klassifizierungen.

Die Autoren geben eine Treffsicherheit von 75-80% als sowohl für Forscher als auch Modellanwender (z.B. Manager) akzeptabel an. Sie berufen sich dabei u.a. auf Morrison et al. (1998, allerdings ältere Ausgabe), demnach erklären selbst die besten Modelle weniger als 50% der Schwankungen des Vorkommens einer Art.

Die schlechte Performance der Modelle, was die Datensätze Georgia und Virginia anbelangt, war zu erwarten, da sich die geographischen Gegebenheiten – und damit die Reaktion einer Art auf die Habitatbedingungen – doch unterscheiden. Mit einer Stratifizierung der Datensätze und einem Modell für die gesamte Region hätte man dem Rechnung tragen können.

Zwei Schlußfolgerungen der Autoren seien noch angeführt:

CART ist der beste Ansatz zum Modellieren von Arten mit annähernd binomialer Verbreitung bezüglich eines Habitatfaktors (z.B. wenn ihr Vorkommen an eine Mindest-Seehöhe gebunden ist).

Das Anpassen von Modellen für Generalisten gestaltet sich schwierig, da diese sich bei der Wahl ihres Lebensraumes naturgemäß nicht festlegen und so keine Anhaltspunkte liefern.

## **7.6 Hierarchical Set of Models und GAM**

Huisman et al. (1993) führen zwar keinen direkten Vergleich durch, würden aber einfachen parametrischen und "statistisch zulässigen" Modellen den Vorzug gegenüber den aufwendigeren GAMs geben. Und zwar aus Gründen der "Sparsamkeit", ganz nach dem Prinzip: "Halte es möglichst einfach".

## 8. Schlußbetrachtung

**Das beste Modell** Gleich vorweg: "*Just as there is no perfect automobile, there is no single "best" modeling approach.*" (Van Horne, 2002).

Alle Autoren die sich ernsthaft mit dieser Frage beschäftigt haben, kommen mehr oder weniger zum selben Schluß: Die verschiedenen Ansätze stehen nicht in Konkurrenz zueinander, vielmehr ergänzen sie sich. Speziell in Fällen, wo es um hohe Einsätze geht (z.B. Risiko, daß eine Population ausstirbt), ist es ratsam mehrere Modelle parallel zu entwickeln, sie anzuwenden und zu testen. Oder, um es mit Morrison et al. (1998) zu sagen: "*Models are like politicians: support them, use them, but don't unquestioningly trust them. Compare them. Find out if several models of the same environment or species tell you the same story, and if they don't, find out why; are you learning more about the structure of the models – their state variables and relations – then the real- world biological entity of interest? ... The lesson here is: Know thy models and thy species.*"

Van Horne (2002) weist mit Nachdruck darauf hin, daß es zielführender ist, Modellwahl und Zielsetzungen aufeinander abzustimmen, anstatt gedankenlos den "letzten Stand der Technik" in der Habitatanalyse zu übernehmen.

**Ausblick** Habitatmodelle operieren in unterschiedlichen Maßstäben und mit unterschiedlicher Auflösung. Es kann durchaus sinnvoll sein, die verschiedenen Ansätze zu verbinden, beispielsweise um im Biotopmanagement mehrere Arten bzw. unterschiedliche Regionen berücksichtigen zu können. Einen möglichen Ansatz besprechen Gross und DeAngelis (2002) am Beispiel ATLSS (*Across Trophic Level System Simulation*).

Das Versagen aktueller Ansätze, was das Aufdecken und Modellieren biotischer Interaktionen – im speziellen von (interspezifischer) Konkurrenz – anbelangt, stellt eine der größten Beschränkungen und gleichzeitig eine Herausforderung für die nächste Zukunft dar.

Ein Ansatz, derartige Interaktionen in statische Verbreitungs-Modelle zu integrieren, könnten (Gleichungs-)Systeme simultaner Regressionen oder GLMs bzw. GAMs (*System of Simultaneous Regressions SSR*) sein, wie sie in der Ökonometrie bereits Verwendung finden (Guisan und Zimmermann, 2000). Hierbei fließt das Vorkommen bzw. die Abundanz einer Art als zusätzlicher Prädiktor in

die Gleichungen aller anderen modellierten Arten ein.

$$\begin{aligned} Y_1 &= \alpha_1 + \beta_1 X + Y_{(-1)} \gamma_1, \\ Y_2 &= \alpha_2 + \beta_2 X + Y_{(-2)} \gamma_2, \\ &\quad \vdots \\ Y_n &= \alpha_n + \beta_n X + Y_{(-n)} \gamma_n \end{aligned}$$

Dabei stellt der zusätzliche Term  $Y_{(-i)} \gamma_i$  das Matrizenprodukt aller anderen (ausschließlich der  $i$ -ten) Zielgrößen  $Y$  – also die Response z.B. anderer Tierarten – dar.

In einem iterativen Prozeß wird das Gleichungssystem so lange durchlaufen, bis sich ein Gleichgewicht einstellt.

Ein anderer vielversprechender Ansatz ist die Verwendung von  $\rightarrow$ CART-Techniken in Ergänzung zu GLMs oder GAMs (Austin, 2002b; Guisan et al., 2002).

**Softwarepakete** Auffälligerweise gehen die Autoren der im Rahmen dieser Arbeit zitierten Literatur kaum auf die Softwarepakete ein, mit denen die besprochenen Modelle realisiert wurden. Die explizit erwähnten und gängigen Statistikpakete sind **S-PLUS**, **SPSS** und **SAS**.

Im Zusammenhang mit der Erstellung von Habitat-Modellen sind natürlich solche Anwendungen interessant, die man um eigene Module erweitern kann und die über eine Schnittstelle zu einem geographischen Informations-System verfügen. So lassen sich mit dem *open source* Software-Paket **R**<sup>1</sup> (Yee und M., 2002), z.B., einige der hier vorgestellten Ansätze (GLM, GAM, ...) ohne größeren Aufwand realisieren. Ein GLM z.B. wird, nach erfolgter Eingabe der Daten, einfach mit dem Aufruf:

```
> fm <- glm(formula=y ~ x, family=binomial(link=logit), data=df)
```

berechnet. Anschließend kann man sich mit `summary(fm)` einen Bericht über Koeffizienten, Residuen, unerklärte Varianz, AIC u.a.m. ausgeben lassen.

**R** läuft auf allen gängigen Plattformen (UNIX, Macintosh, Windows) und verfügt auch über eine Schnittstelle zu GRASS, einem – ebenfalls freien – rasterbasierten GIS<sup>2</sup>.

Das ursprünglich in **S-PLUS** implementierte GRASP, nutzt eben diese Tatsache

<sup>1</sup><http://www.r-project.org/>

<sup>2</sup>[http://www.geog.uni\\_hannover.de/grass/](http://www.geog.uni_hannover.de/grass/)

und steht mittlerweile ebenfalls in einer **R**-Version zum Download<sup>3</sup> zur Verfügung. Denkbar sind natürlich eigenständige Implementierungen in geeigneten Programmiersprachen, wie z.B. FORTRAN. Biomapper (Hirzel et al., 2001a), die Software zu ENFA, ist in Borland Delphi 4 realisiert.

**Caveat** Grimm et al. (1999) warnen davor, ökologischen Modellen bzw. ihren Aussagen blindes Vertrauen zu schenken. Ein Mindestmaß an Verständnis dafür, wie diese Ergebnisse zustandekommen ist erforderlich. Auch sollten die dem Modell zugrundeliegenden Annahmen bekannt sein.

**Von Forschern und Paradigmen** Austin (1999) betont wie wichtig es ist, die grundlegenden Annahmen eines Paradigmas ständig in Frage zu stellen, um Inkonsistenzen zwischen angewandten Methoden und Hypothesen zu vermeiden. Fälle in denen derartige Überlegungen sträflich unterlassen werden, stehen an der Tagesordnung. Der Autor scheut nicht davor zurück, konkrete Beispiele zu nennen. Aber er deckt nicht nur auf. Er ruft zu verstärkter Kommunikation auf, wenn auch mit seltsamen Worten: "*The clash between paradigms is not often marked by controversy but more by a conspiracy of silence, 'don't knock my paradigm and I won't knock yours'. ... We know our current paradigms are inadequate, therefore we should help by trying to subvert each others' current paradigm.*"

**Feedback Europa -> Nordamerika** Auf die Frage ob, es denn ein Feedback von Europa in die USA bezüglich der Wildschadensprädispositionsmodelle gäbe, ist die Antwort prinzipiell nein.

Wie Völk (1999) bemerkt, spielen Wildschäden ausserhalb der dichtbesiedelten europäischen Kulturlandschaft kaum eine Rolle. Aber mit zunehmender Umstellung von großflächiger Bewirtschaftung auf kleine, schachbrettartig angeordnete *patches* und der dadurch veränderten Randliniensituation könnte dieser Aspekt bald auch in anderen Regionen eine Rolle spielen.

So beschreiben etwa Ziegler und Nolte (2001) durch Schwarzbären verursachte Schäden im Bundestaat Washington. Im Frühjahr decken die Tiere durch Rindenschäle an Douglasie anscheinend einen wesentlichen Teil ihres Bedarfes an Kohlehydraten. Die *timber companies* versuchen durch zuckerreiches Ablenkfut-

---

<sup>3</sup><http://www.cscf.ch/grasp/grasper/>

ter die Bären vom Schälen abzuhalten. Während der kritischen Zeit werden in den schadensanfälligen Beständen Futterstellen mit Pellets beschickt (Fersterer et al., 2001).

Fersterer (2000) schlägt – mit Hinblick auf europäische Erfahrungen im Umgang mit Wildschäden – waldbauliche Maßnahmen vor, welche die Schadensanfälligkeit der Bestände herabsetzen sollen. Der Logik des von Reimoser und Gossow (1996) propagierten Prädispositionsmodells folgend, sind das einerseits Maßnahmen, welche die Habitatattraktivität beeinflussen (höhere Stammzahl durch Naturverjüngung, Mischung mit Laubbäumen, Schlagabraum als mechanischer Schutz) und andererseits Maßnahmen zur Erhöhung des – schadensfrei nutzbaren – Nahrungsangebotes (Gestaltung von Schlägen, Ansäen von Futterpflanzen zur Überbrückung der kritischen Zeit, Totholz).

Die Voraussetzungen für einen "Export" der europäischen Erfahrungen und Erkenntnisse im Zusammenhang mit Wildschäden scheinen also gegeben zu sein.

## 9. Zusammenfassung

Habitatmodelle finden in den verschiedensten Bereichen Verwendung: vom Biotopmanagement, über Artenschutz und Landnutzungsplanung bis hin zur Wildschadensprävention. Vorhersage, z.B. von geeignetem Habitat, stellt dabei nur ein Ziel dar. Habitatmodelle werden auch "nur" zur Beschreibung von Wildtier-Habitat-Beziehungen entwickelt, bzw. um diese Zusammenhänge überhaupt erst nachvollziehen zu können, also zum Gewinnen neuer Erkenntnisse.

Die Entwicklung eines Modells läuft in einem zyklischen, interaktiven Prozeß ab. Doch kann man 5 Schritte unterscheiden: konzeptionelles Modell, statistische Formulierung, Anpassung, Vorhersage und Evaluierung. Während aller Phasen des Modellbildungsprozesses, speziell aber bei der Entscheidung für einen bestimmten Ansatz, ist es notwendig zu überprüfen, ob Methoden und grundlegende Konzepte kompatibel sind. Sonst kommt es zu Inkonsistenzen bzw. verzerrten Aussagen. Aber auch bei der Wahl der Variablen, die zur Beschreibung eines Habitas herangezogen werden, sind einige Überlegungen (Kosten, Verfügbarkeit, Genauigkeit) anzustellen.

Habitatmodelle beruhen auf einer Vielzahl mathematisch-statistischer aber auch anderer Ansätze, z.B. auf einfachen Annahmen über Zusammenhänge. Im Rahmen dieser Arbeit wird auf HSI-Modelle (*Habitat Suitability Index*), *Habitat Evaluation Procedures* (HEP), Regressionsmodelle (GAM, GLM, autologistisches Modell), *Ecological Niche Factor Analysis* (ENFA), neuronale Netzwerke (ANN), *classification trees*, Expertensysteme (XPS), *FORest dynamics in SPAtially Changing Environments* (FORSPACE), *Gap Analysis Program* (GAP), *Resource Selection Functions* (RSF) und individuenbasierte Modelle (IBM) eingegangen.

Für einzelne der angeführten Ansätze (HSI-Modell, GLM, autologistisches Modell, ENFA, FORSPACE, Expertensystem, ANN, *classification trees*, Diskriminanzanalyse, GAP, logistische Regression, RSF) wird anhand konkreter Beispiele aus der Literatur ihre Anwendung diskutiert. Hauptaugenmerk liegt hier auf den eingeflossenen Habitatvariablen und den Aussagen der Modelle.

Wieder anhand von Beispielen aus der Literatur werden direkte Vergleiche verschiedener Ansätze besprochen. Dadurch ist es möglich, Stärken bzw. Schwä-

chen der Modelle zu identifizieren. Allerdings spiegeln faßt alle dieser Beiträge die Präferenzen der jeweiligen Autoren wieder.

Sämtliche Autoren, die sich eingehend mit den verschiedenen Modell-Ansätzen auseinandergesetzt haben, kommen zum selben Schluß: Das "beste" Modell gibt es nicht. Man kann höchstens davon sprechen, daß ein bestimmtes Modell für eine bestimmte Fragestellung besser geeignet ist als andere.



## 10. Glossar und Abkürzungsverzeichnis

**AI** *Artificial Intelligence*, Künstliche Intelligenz (KI). Die Erforschung und möglichst exakte Simulation von natürlichem intelligentem Verhalten. Teilgebiete sind z.B. →XPS, →ANN und →machine learning.

**AIC** *Akaike Information Criterion*, Informations-Kriterium von Akaike. Das AIC zeichnet beim Vergleich mehrerer (Regressions-)Modelle – die sich in der Anzahl der Terme unterscheiden – das "beste" aus (kleinstes AIC). Ein "größeres" Modell (Anzahl der Terme) ist nicht unbedingt besser als ein "kleineres" (Stahel, 2002).

**ANN** *Artificial Neural Network*, künstliches neuronales Netzwerk. Forschungsgebiet der →AI.

**ANOVA** *ANalysis Of VAriance*, Varianzanalyse.

**ANUCLIM** siehe →BIOCLIM. ANUCLIM ist zwar speziell für die Modellierung von Vegetations-Habitaten entwickelt worden, kann aber ohne weiteres – durch Berücksichtigen zusätzlicher Parameter – auch auf das Tierreich übertragen werden (Elith, 2000).

**Autokorrelation** besteht, wenn die einzelnen Beobachtungen nicht unabhängig voneinander sind, z.B. entlang eines Gradienten oder in Zeitreihen. Dies verletzt die den meisten statistischen Ansätzen zugrundeliegende Annahme, daß die Beobachtungen voneinander unabhängig sind.

**Bayesian approaches** Auf dem →Satz von Bayes beruhende Ansätze verknüpfen die *a priori* Wahrscheinlichkeiten für Beobachtung einer Art mit deren Wahrscheinlichkeiten für Vorkommen in Abhängigkeit vom Wert der verschiedenen Prädiktoren: z.B.  $p(y|x) = \frac{p(y) p(x|y)}{p(y) p(x|y) + p(-y) p(x|-y)}$ . Bedingte *a priori* Wahrscheinlichkeiten ( $p(y|x)$ ) können z.B. relative Häufigkeiten des Art-Vorkommens sein. *A priori* Wahrscheinlichkeiten ( $p(x|y)$ ) werden aus der Literatur oder aus früheren Studien abgeleitet (Guisan und Zimmermann, 2000).

Kritikpunkt: Der Wert für die *a priori* Wahrscheinlichkeit ist schwer zu bestimmen, gründet letzten Endes auf Experten-Urteil. Da die *a posteriori*

Wahrscheinlichkeit wesentlich von der *a priori* Wahrscheinlichkeit beeinflusst wird, ist sie entsprechend fehlerbehaftet (Morrison et al., 1998).

**Bestimmtheitsmaß**  $R^2$ , jener Anteil der Varianz von Y, die durch X erklärt wird.  $R^2$  wird um so größer, je mehr Terme hinzugefügt werden. Der Wertebereich von  $R^2$  ist abhängig von der Variabilität des Regressors (je größer X, desto größer  $R^2$ ). Ein hoher Wert von  $R^2$  sagt nichts über die Modellgüte aus (Kumke, 2001).

**BIOCLIM** *BIOCLIMatic Prediction System*. Durch Kartierung von Klimadaten (Langzeit Aufzeichnungen von Niederschlag, Temperatur, Sonneneinstrahlung), topographischer Höhe und Vorkommen der Zieltierart entsteht ein artspezifisches "Klima-Profil". Mit Hilfe der Parameter berechnet man eine (minimale,) geradlinige, rechtwinklige "Hülle" im mehrdimensionalen "klimatischen Raum", die als *bounding box* oder *climatic envelope* bezeichnet wird. Es wird also die klimatische Eignung eines Gebietes "vorhergesagt" (Elith, 2000).

**bootstrap** Methode zum Bestimmen von Vertrauensintervallen, zum Schätzen von Parametern und Testen von Hypothesen. Auch als →resampling-Methode benutzt (verwandt mit →jackknife).

**BPN** *BackPropagation Network*, auch als *multi-layer perceptron* bekannt, →ANN.

**Calibration** Eichung. Siehe →IR.

**CA** *Correspondence Analysis*, Korrespondenz-Analyse. Auch bekannt als *reciprocal averaging*. Zur Quantifizierung der Beziehung zwischen den Objekten wird ein Assoziationsmaß ( $\chi^2$ -Abstand) herangezogen (Kumke, 2001). Zählt zu den →Ordination-Methoden.

**Canonical Ordination** Kanonische Verfahren umfassen alle Methoden zur simultanen Analyse von mindestens zwei Datenmatrizen. Voraussetzung ist dabei, daß beide Datensätze denselben Objekten entstammen. Kanonische Verfahren basieren auf →Eigenanalyse und, in vielen Fällen, auf multipler Regression (Kumke, 2001). Methoden: →CCA, →RDA.

**CART** *Classification and Regression Tree analysis*, System zum Induktion von *classification trees*.

**CCA** *Canonical Correspondence Analysis*, Kanonische Korrespondenzanalyse. Erweiterung der  $\rightarrow$ CA. Die CCA hat das Ziel, die Resonanz eines multivariaten Ensembles von Spezies (Matrix Y) durch sie beeinflussende Variablen (Matrix X) zu modellieren (Gradienten-Analyse). Die Speziesresonanz entlang eines Gradienten ist dabei nicht linear (Kumke, 2001), sondern folgt einer unimodalen Verteilung (Harvey, 1996). siehe  $\rightarrow$ Canonical Ordination.

**Classification** Klassifikation. Aus den Merkmalen eines Individuums (Eingabegrößen) wird dessen Gruppenzugehörigkeit (Ausgabegröße, zweiwertig oder nominal) bestimmt, Gruppen a priori festgelegt (im Gegensatz zur  $\rightarrow$ Ordination). Methoden:  $\rightarrow$ DA.

**Clusteranalyse** *Cluster Analysis*. Clusteranalysen haben das Ziel, Objekte zu gruppieren. Objekte mit relativ ähnlichen Eigenschaften werden dabei derselben Kategorie zugeordnet. Basis der Clusteranalysen sind Assoziationsmaße (Ähnlichkeits- bzw. Abstandskoeffizienten) (Kumke, 2001).

**COR** *CORrelation analysis*, Korrelations-Analyse. Es werden die Beziehungen der Deskriptoren untereinander analysiert, z.B. Bestimmen der Kovarianz(-) (gemeinsame Dispersion der Variablen um deren Mittelwert) und Korrelation(s-Matrix) (Maß für die Abhängigkeit zwischen Zufallsvariablen) oder des multiplen Korrelations-Koeffizienten (Maß für die Stärke des Zusammenhanges zwischen einer Resonanzvariablen und einer Linearkombination von erklärenden Variablen).

**CV** *cross-validation*, Kreuzvalidierung,  $\rightarrow$ resampling-Methode.

**DA** *Discriminant Analysis*, Diskriminanz-Analyse; Anwendungsgebiet: Klassifikations-Probleme, Muster-Erkennung, sind die gegebenen Variablen dazu geeignet zwischen vordefinierten Gruppen zu unterscheiden; für größere Datensätze zu wenig flexibel (Alternative  $\rightarrow$ ANN) (Corsi et al., 1999; Stahel, 2002).

**DCA** *Detrended Correspondence Analysis*. Durch Entfernen des Trends werden Verzerrungen, die bei der  $\rightarrow$ CA auftreten unterbunden.

**DCCA** *Detrended Canonical Correspondence Analysis*. Durch Entfernen des Trends werden Verzerrungen die bei der  $\rightarrow$ CCA auftreten unterbunden.

**DFA** *Discriminant Function Analysis*, siehe  $\rightarrow$ DA.

**DOMAIN** Anwendung multivariater Distanz-Messung (*Gower metric*) auf Kartierung. Es werden – im Gegensatz zu  $\rightarrow$ BIOCLIM – keine Grenzen definiert, sondern es wird ein Maß für die Gleichartigkeit von Habitaten bestimmt (Elith, 2000).

**Eigenanalyse** Ermitteln von Eigenvektoren  $u$  und Eigenwerten  $\lambda$  einer Matrix, z.B. der  $n \times p$ - Matrix der Deskriptoren. Lineare Algebra ist ein unverzichtbares Hilfsmittel in der multidimensionalen Statistik.

**ENFA** *Ecological Niche Factor Analysis* (Hirzel, 2001; Hirzel et al., 2002).

**FA** *factor analysis*, Faktoren-Analyse. Die Zielgröße wird als Ergebnis einer Überlagerung von Faktoren interpretiert, die nicht direkt beobachtbar sind (latente Variablen), sondern erst aus den beobachteten Größen  $X$  erschlossen werden müssen.  $\rightarrow$ Ordinations-Methode, verwandt mit  $\rightarrow$ PCA (Stahel, 2002).

**Fehler 1. Art** Ablehnung der Nullhypothese obwohl diese korrekt ist.

**Fehler 2. Art** Beibehalten der Nullhypothese obwohl diese falsch ist.

**Fehler 3. Art** Falsches Problem gelöst ;-)

**GA** *Genetic Algorithms*, genetische Algorithmen. Diese Optimierungsmethode ist an die Evolutionstheorie angelehnt: charakteristische Beschreibungen von Problemlösungen werden codiert und bilden sogenannte Individuen, die mit Hilfe einer Fitnessfunktion bewertet werden. Durch wiederholte Auswahl, Rekombination und Eliminierung werden immer neue Individuen, also Lösungen erzeugt. Ziel ist es, daß diese im Laufe der Generationen bzgl. ihrer Fitness immer besser werden.

**GAM** *General(ized) Additive Model*, allgemeines additives Modell.  $\rightarrow$ nichtparametrische Regressionsmethode für mehrere Variablen. Annahme: Effekte auf Zielgröße verhalten sich additiv (Stahel, 2002).

**GAP** *Gap Analysis Program*. GAP<sup>1</sup> ist ein vom *USGS Biological Resources Division* entwickelter, proaktiver, GIS-basierter Ansatz zur Erhaltung der Biodiversität. Dabei wird untersucht, in welchem Ausmaß einheimische Tierarten (zunächst Wirbeltiere, später auch Insekten, Mollusken und Crustaceen) und natürliche Pflanzengemeinschaften im (bestehenden) – vegetationskundlich ausgewiesenen – Schutzgebiete-Netz vertreten sind. Unterrepräsentierte Tierarten bzw. Gemeinschaften stellen *gaps* (Lücken) dar, die durch Schutzmaßnahmen geschlossen werden sollen. Motto: *keep common species common*.

**GARP** *Genetic Algorithm for Rule set Prediction*. In ein GIS eingebundenes Programm zur Vorhersage der Verbreitung von Arten aus Umwelt-Variablen. Es werden genetische Algorithmen (→GA) verwendet, um Regeln für die Vorhersage zu generieren, testen und modifizieren. Das Modell entsteht aus Test- und Trainings-Datensätzen (resampelte originale Vorkommen- oder Vorkommen/Nicht-Vorkommen-Datensätze). Vorhersagen können als relative Wahrscheinlichkeit für das Vorkommen einer Art oder auch als ein Index für die Güte eines Habitats interpretiert werden (Elith, 2000).

**GIS** Geographisches Informations-System.

**GLM** *Generalized Linear Model*, verallgemeinertes lineares Modell. Klasse von (Regressions-)Modellen, dazu zählen Logistische Regression, Poisson-Regression und log-lineare Modelle. Erweiterung der linearen Regression um nichtlinearen Zusammenhang zw. erklärenden Variablen und Response sowie nicht-normale Fehlerverteilung.

**Gradient** ökologischer Gradient: räumlich variierender Aspekt der Umwelt.

**HABITAT** Modell, ähnlich →BIOCLIM, zur Berechnung konvexer Hüllen (*convex polytope envelopes, convex hulls*) (Guisan und Zimmermann, 2000).

**HEP** *Habitat Evaluation Procedures* (U.S. Fish and Wildlife Service, 1980a). Modell zur Bewertung von Auswirkungen anthropogener Störungen auf die Umwelt, Vergleich von Alternativen auf der Basis von →HUs.

**HOF** *Hierarchical Set of Models* (Huisman et al., 1993).

---

<sup>1</sup><http://www.gap.uidaho.edu/>

**HSI** *Habitat Suitability Index*, dimensionsloser Wert zwischen 0,0 und 1,0. 0,0 bezeichnet ungeeignetes Habitat und 1,0 optimales Habitat (U.S. Fish and Wildlife Service, 1981).

**HU** *Habitat Unit*, ergibt sich aus der Multiplikation der verfügbaren (Habitat-)Fläche mit ihrem HSI. Grundlage für Vergleich von Habitaten (räumlich) bzw. Habitats-Veränderungen (zeitlich) (U.S. Fish and Wildlife Service, 1980a, 1981).

**IBM** *Individual Based Model*, Individuenbasiertes Modell. Populationen werden durch einzelne Individuen – mit eigenständigem Verhalten – nachgebildet.

**IR** *Inverse Regression*, auch  $\rightarrow$ Calibration, inverse Regression. Regression in die "falsche" Richtung, der wahre Wert X wird aus dem gemessenen Wert Y geschätzt, Messmethode wird mittels einer Ausgleichsgeraden geeicht.

**JK** *jackknife*. Methode zum Schätzen von Parametern und  $\rightarrow$ resampling von Datensätzen (verwandt mit  $\rightarrow$ bootstrap).

**Kollinearität** (Multi-)Kollinearität besteht, wenn zwei (oder mehrere) Regressoren stark (linear) korrelieren. Folge: große Varianzen bei Schätzung der Regressionskoeffizienten, zu große Werte für  $\beta_i$  und  $R^2$ , Interpretation erschwert. Ursachen: Methoden der Datenerhebung oder überdefiniertes Modell ( $p > n$ , mehr Regressoren als Beobachtungen). Erkennung (mittels Softwarepaketen): Toleranzwert, *Variance Inflation Factor* (VIF). Behebung: Erhebung weiterer Daten, Eliminierung zusammenhängender Regressoren.

**Korrelationskoeffizient** Maß für die Stärke eines *statistischen* Zusammenhangs, die Ableitung eines *kausalen* Zusammenhangs ist nicht zulässig.

**Linearkombination** *linear combination*.  $\sum_{j=0}^p \beta_j x_{ji}$  oder auch  $\beta X$ , wobei  $\beta$  und X Vektoren sind.

**logistische Regression** *logistic regression*. Zielgröße ist nicht eine kontinuierliche, sondern eine *binäre* Variable (0-1 Variable, *dummy variable*), die das Auftreten eines bestimmten Ereignisses angibt (z.B. Vorkommen einer Art). Gesucht ist die Abhängigkeit dieses Ereignisses von einer oder mehreren erklärenden Variablen. Gehört zur Klasse der  $\rightarrow$ GLM.

**logit** *logit*-Transformation. Wird meist zur  $\rightarrow$ Transformation von Wahrscheinlichkeiten (Intervall von 0 bis 1) verwendet, um Werte im Intervall von  $-\infty$  bis  $\infty$  zu erhalten:  $\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$ .

**LOWESS** *Locally Weighted Scatterplot Smoother, Locally Weighted Sum of Squares*. Glättungsverfahren (*smoother*).

**LR** *Linear Regression*, lineare Regression. Dazu zählen auch quadratische und polynomiale Regression (Stahel, 2002).

**LS** *Least Squares*, (Methode der) Kleinste(n) Quadrate. Gebräuchlichste Methode zur Bestimmung der Steigung (b) und des Achsabschnittes (a) bei der (linearen) Regression, siehe auch  $\rightarrow$ ML.

**machine learning** maschinelles Lernen. Die Forschung zu maschinellem Lernen beschäftigt sich mit der Entwicklung von Computersystemen, die in der Lage sind, aus den zugeführten Eingabeinformationen neues Wissen zu generieren bzw. vorhandenes Wissen zu verbessern. Teildisziplin der  $\rightarrow$ AI.

**Mahalanobis distance** Mahalanobis Distanz (M. Abstand, M. Radius). Abstandskoeffizient (Assoziationsmaß), charakterisiert die Unähnlichkeit zwischen zwei Objekten (Kumke, 2001).

**MDA** *Multiple Discriminant Analysis*, siehe  $\rightarrow$ DA.

**Metapopulation** Eine Metapopulation setzt sich aus mehreren, sich unabhängig voneinander fortpflanzenden, lokalen Populationen zusammen. Migration zwischen diesen Subpopulationen beeinflusst die lokale Dynamik, z.B. durch Wiederbesiedlung nach lokalem Aussterben. Starker Bezug zur  $\rightarrow$ spatial explicit *landscape ecology*.

**ML** *Maximum Likelihood* (größte Wahrscheinlichkeit). Methode zum Schätzen der Regressionskoeffizienten, siehe auch  $\rightarrow$ LS.

**MLR** *Multiple Linear Regression*, multiple lineare Regression; erfasst den Zusammenhang zwischen einer Zielgröße und mehreren Ausgangsgrößen. Es werden keine Annahmen über erklärende Variablen (Datentyp, Verteilung) gemacht und es besteht die Möglichkeit der Variablen-Transformation, daraus folgt deren hohe Flexibilität (Stahel, 2002).

**MR** *Multiple Regression*, multiple Regression. Eine abhängige Variable wird als Funktion mehrerer unabhängiger Variablen dargestellt.

**Multikollinearität** → Kollinearität

**multivariate Regression** Modelle, in denen gleichzeitig mehrere Zielgrößen ( $y_i$ ) in Abhängigkeit von erklärenden Variablen ( $x_i$ ) beschrieben werden, Problemstellung der multivariaten Statistik.

**MVP** *Minimum Viable Population*. Schätzung für die kleinste überlebensfähige Population die isoliert bestehen kann, bezogen auf einen bestimmten Zeitraum (z.B. 100 Jahre), ausgedrückt mit einer gewissen Wahrscheinlichkeit (z.B. 95%).

**nichtlineare Regression** → Regression mit nichtlinearem Zusammenhang zwischen Prädiktor(-en) und zu erklärender Variable. Nichtlineare Regressions-Modelle haben die allgemeine Form  $y_i = f(x_i, \theta) + \varepsilon_i$ , die Parameter ( $\theta$ ) werden mittels →LS-Methode geschätzt. Z.B. das logistische Wachstumsmodell:  $y = \frac{\theta_1}{1 + \theta_2 e^{-\theta_3 x}} + \varepsilon$  mit  $\theta_1 = K, \theta_2 = \frac{K}{N_0} - 1, \theta_3 = r$  das entspricht  $N_t = \frac{K}{1 + \frac{K - N_0}{N_0} e^{-rt}}$  (Kumke, 2001).

**nichtlinearer Zusammenhang** Bei Veränderung der Ursache ergibt sich keine gleichbleibend proportionale Veränderung der Wirkung.

**nichtparametrische Regressionsmethode** Zum Schätzen der Regressionsfunktion mit wenigen Konstanten (→ nichtlineare Regression) werden sog. Glättungsverfahren (*smoother*) eingesetzt. Der Begriff hat nichts mit der unterstellten – bzw. nicht unterstellten – Verteilung (→ nichtparametrische Verfahren) zu tun, sondern mit der (geringen) Anzahl an Parametern.

**nichtparametrische Verfahren** Es wird nicht vorausgesetzt, daß die betrachteten Zufallsvariablen bestimmten Verteilungsfunktionen folgen.

**Ordination** Verfahren zur "Dimensions-Reduktion". Ziel von Ordinationsverfahren ist es, die Struktur eines multivariaten Datensatzes zu vereinfachen und auf die wesentlichen Informationen zu reduzieren. Bei der Vereinfachung wird versucht, die Variationsstruktur des Datensatzes beizubehalten. Man versucht bei der Ordination alle im Datensatz vorhandenen Deskriptoren in einem reduzierten Raum – meistens nicht mehr als zwei Dimensionen –



darzustellen. Als Ergebnis einer Ordination erhält man eine graphische Repräsentation der vereinfachten Datenstruktur in zwei Dimensionen (Kumke, 2001). Methoden: →PCA, →CA, →PCoA, →FA.

**parametrische Verfahren** Es wird angenommen, daß die beobachteten Variablen einer bestimmten Verteilungsfunktion folgen. Z.B. legt die einfache lineare Regression (Annahme über die Verteilung der Fehler) mit den Parametern  $\alpha$ ,  $\beta$  und  $\sigma$  die parametrische Familie fest. Diese drei Größen bestimmen die Verteilung der Variable  $Y \sim \mathcal{N}(\alpha + \beta X, \sigma^2)$  (Stahel, 2002, S. 260).

**PATREC** *PATtern RECOgnition*, Muster-Erkennung. Modell zur Vorhersage der Auswirkungen veränderter Umweltbedingungen auf Wildtiere in Form einer Risiko-Analyse. PATREC beruht auf dem →Satz von Bayes, fällt also in die Kategorie →Bayesian approaches. Man berechnet damit die Wahrscheinlichkeit einer Reaktion (z.B. Populationsdichte) bei gegebener Wahrscheinlichkeit des Vorhandenseins verschiedener Umweltbedingungen (Morrison et al., 1998).

**pattern ecological p.** Anordnung (gleichmäßig, zufällig, gruppiert, regelmäßig) von Beobachtungen, z.B. die räumliche Verteilung einer Tierart im Untersuchungsgebiet (Smallwood, 1993; Van Horne, 2002).

**PCA** *Principal Components Analysis*, Hauptkomponenten-Analyse. Die Streuung der Daten soll durch Projektion der Datenpunkte auf eine geeignete Richtung so gut wie möglich wiedergegeben werden. Varianz der projizierten Punkte soll möglichst groß sein. →Ordinations-Methode, Anwendungsgebiet: Dimensions-Reduktion (Stahel, 2002).

**PCoA** *Principal Coordinates Analysis*, Hauptkoordinatenanalyse (metrische multidimensionale Skalierung). →Ordinations-Methode. Darstellung von Variablen im kartesischen Koordinatensystem und Bestimmung ihrer Ähnlichkeit mittels Abstands-Koeffizienten (Kumke, 2001).

**Performance** Leistungsfähigkeit, Vorhersagegüte.

**Prävalenz** *prevalence*, "(Vorkommens-)Häufigkeit". Verhältniszahl aus Punkten mit positiver Beobachtung zu allen aufgenommenen Punkten. Wirkt sich auf

→Sensitivität und →Spezifität eines Modells aus. Bei hoher Prävalenz hohe Sensitivität, bei niedriger Prävalenz hohe Spezifität (Manel et al., 1999).

**Prädiktoren** Synonym für erklärende Variablen, Regressoren, Deskriptoren, manchmal wird auch der Ausdruck unabhängige Variablen verwendet.

**process ecological p.** Mechanismus, ausgelöst durch Systemkomponenten, der eine Veränderung/Entstehung eines Musters (→pattern) bewirkt (z.B. Abwanderung bedingt durch innerartliche Konkurrenz) (Smallwood, 1993; Van Horne, 2002).

**PVA** *Population Viability Analysis*, nicht-räumliche demographische Gefährdungsanalyse einer Population.

**RDA** *ReDundancy Analysis*, Redundanz-Analyse. Methode (Gradienten-Analyse) zur Modellierung der Resonanz eines multivariaten Datensatzes. RDA hat das Ziel, die Varianz von  $Y$  ( $n * p$ ) durch eine zweite Datenmatrix  $X$  ( $n * m$ ) zu erklären. Es wird eine lineare Beziehung innerhalb von  $Y$  unterstellt (Kumke, 2001).

**Regression** Regressions-Analyse. Technik zur Modellierung der Beziehung zwischen mindestens 2 Variablen (einer oder mehrerer unabhängigen und einer abhängiger), um das Verhalten der abhängigen Variable zu prognostizieren. Kann nur statistische Zusammenhänge nachweisen, keinen ursächlichen Einfluß. Man unterscheidet lineare Regression (→LR) und →nichtlineare Regression. Werden mehr als 2 Variablen benutzt, spricht man von multipler Regression (→MR) bzw. multipler linearer Regression (→MLR). Annahmen:  $X$  ohne Fehler gemessen, alle Fälle sind unabhängig voneinander, Fehler in  $Y$  sind normalverteilt und unabhängig von  $X$ .

**resampling** Methoden (z.B. →bootstrap, →CV und →JK) zur Aufbereitung einer Stichprobe: aus einer gegebenen Stichprobe werden viele "Pseudo-Stichproben" erzeugt.

**Residuen** *residuals*, Restwerte, Abweichung zwischen (beobachteten) Daten und (geschätzter) Anpassung  $y_i - \hat{y}_i$ . Maß für jenen Anteil der Variabilität, der nicht durch das Modell erklärt werden kann.

**Response** *response*, Syn. Resonanz: (Graph der Funktion der) zu erklärende(n) Variable(n).

**ROC** *Relative (Receiver) Operating Characteristics*, ein schwellenwertunabhängiges Maß für die Genauigkeit der Vorhersage eines Modells, wird zur Validierung/Evaluierung herangezogen. Ein *ROC-plot* wird erstellt, indem man den Anteil der *true positives* ( $\rightarrow$ Sensitivität) auf der y-Achse gegen den Anteil der *false positives* auf der x-Achse aufträgt. Die Fläche unter der ROC-Kurve (AUC) kann als Index für die Leistungsfähigkeit eines Modells betrachtet werden, weil sie unabhängig von einem im Trainingsdatensatz willkürlich festgelegten Schwellenwert (und der  $\rightarrow$ Prävalenz einer Spezies) ist (Pearce und Ferrier, 2000).

**Satz von Bayes** *Bayes's theorem*.  $P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{\sum_{j=1}^k P(A_j)P(B|A_j)}$ .

**scale** auch *landscape level*, Betrachtungsmaßstab. Bezeichnet die raumzeitliche Erstreckung von Phänomenen, nicht das Abbildungsverhältnis auf (Land-)Karten. Charakterisiert durch *extent* (Ausdehnung des Untersuchungsgebietes) und *grain (coarsness, Größe der patches)*. Der Betrachtungsmaßstab muß auf den beobachteten Prozeß ( $\rightarrow$ process) abgestimmt sein: erzeugt der Prozeß ein grobkörnigeres Muster ( $\rightarrow$ pattern) ist nur ein Trend erkennbar, bei zu feinkörnigem Muster ergibt sich nur ein Rauschen (Lang, 1999).

**Sensitivität** *sensitivity* (Empfindlichkeit), *true positive fraction*, z.B. Prozentsatz durch ein Modell richtig erkannter "Vorkommen" (*presence*).

**Spearmanische Rangkorrelation** Maß für den Zusammenhang zwischen zwei Variablen, das nicht nur den linearen Anteil mißt. Gründet sich auf Rängen. Ist robuster als der gewöhnliche  $\rightarrow$ Korrelationskoeffizient (Stahel, 2002).

**SEPM** *Spatially Explicit Population Models* (Dunning et al., 1995).

**SLR** *Simple Linear Regression*, einfache lineare Regression, siehe  $\rightarrow$ LR.

**SOM** (Kohonen) *Self-Organizing Mapping*,  $\rightarrow$ ANN.

**spatial explicit** explizit räumliche Betrachtungsweise/Beschreibung der Struktur einer Landschaft. Die spezifische räumliche Anordnung von *patches* wird als *configuration* bezeichnet. Diese bezieht sich nicht nur auf

das Verteilungsmuster, sondern zieht auch die räumlichen Attribute (z.B. Verklumpungs- (*contagion*) und den Vernetzungsgrad (*connectivity*)) der einzelnen *patches* in Betracht. Ergänzung zur "traditionellen" Quantifizierung der *composition* ("Informationsgehalt" einer Landschaft, *patches* als Informationsträger, Eigenschaften: z.B. Anzahl (*abundance*), Verschiedenartigkeit (*variability*)). Zu weiteren Begriffen der *landscape metrics* siehe Lang (1999).

**Spezifität** *specificity* (Genauigkeit), *true negative fraction*, z.B. Prozentsatz durch ein Modell richtig erkannter "Nicht-Vorkommen" (*absence*).

**Standardisieren** → Transformation von Variablen, sodaß der Mittelwert ("Standard-Lage") der transformierten Variable 0 und deren Standardabweichung ("Standard-Streuung") 1 wird. Die entsprechende Funktion lautet:  $g(x) = \frac{x_i - \bar{x}}{\sigma}$ . Dient dem Vergleich verschieden dimensionierter Variablen (Dimension geht "verloren").

**Transformation** die gebräuchlichsten Verfahren sind: → Standardisieren, Symmetrisieren (Ausgebügeln der "Schiefe", "Dämpfen" von Ausreißern, Entzerrung erfolgt z.B. durch Logarithmus-<sup>2</sup>, Wurzel-, Potenz- oder Winkel-Transformation), Normieren (Daten auf Maximalwert skaliert, sodaß alle Werte zw. 0 und 1 liegen) und Zentrieren (beziehen der Daten auf ihren Mittelwert  $g(x) = x_i - \bar{x}$ ). Lineare Transformation (z.B.  $g(x) = a * (x_i - b)$ ) zum Umrechnen von Fahrenheit in Celsius).

Zweck der Transformation: Erleichtern der Interpretation, Linearisieren von Beziehungen, Homogenisieren von Streuungen, Anpassen an eine bestimmte Verteilungsfunktion.

**Validierung** *validation*. Prüfung der Stabilität sowie der Vorhersagegüte eines Modells (bei Inter- und Extrapolation).

**WAE** *Weighted Averaging of Environmental variables*.

**WAI** *Weighted Averaging of Indicator values*.

<sup>2</sup> $\log(x_i + c)$  Konstante wird addiert, um  $\log(0)$  zu verhindern

**WÖBT** Wildökologischer Bestandestyp<sup>3</sup>. Vegetationsstrukturtyp mit weitgehend identer Eignung für eine Wildtierart hinsichtlich seiner primären Lebensbedürfnisse (Nahrungsangebot, Feind- und Klimaschutz, Wohnraum). Im WÖBT nicht berücksichtigt werden die Habitatfaktoren Beunruhigung, Gelände und Klima. Die Strukturtypen werden nach Typ und Höhe des Bewuchses, ihrer Schichtigkeit und ihrem Bodenbegrünungsgrad klassifiziert. Die Einteilung orientiert sich zudem an den in der Forstwirtschaft verbreiteten Bestandesentwicklungsphasen (Wuchsklassen), was eine leichtere Zuordnung von forstwirtschaftlichen Maßnahmen gewährleistet (Partl, 2001).

**XPS** ES, *expert system*, Expertensystem. In Expertensystemen werden Spezialwissen und Schlußfolgerungsfähigkeit qualifizierter Fachleute auf einem eng begrenzten Anwendungsgebiet im Computer nachgebildet. Die so entstandenen Systeme sollen Fachleute in ihrer Entscheidungsfindung unterstützen. Teildisziplin der →AI.

---

<sup>3</sup>Reimoser, F. (1986): Wechselwirkung zwischen Waldstruktur, Rewildverteilung und Rewildbejagbarkeit in Abhängigkeit der waldbaulichen Betriebsform. Dissertation, Universität für Bodenkultur, Wien.

## Literaturverzeichnis

- Abecker, A. (2000): *Einstieg in Expertensysteme*. <http://www.dfki.uni-kl.de/~aa-becker/Mosbach/xps.pdf>.
- Allen, C., Simpson, K. und Johnson, A. (2002): *Improving Vertebrate Modeling in Gap Analyses: Incorporating Minimum Viable Populations and Functional Connectivity in Patchy Environments*. Gap Analysis Program Annual Bulletins 11: 24–30, [http://www.gap.uidaho.edu/Bulletins/11/Improving\\_vertibrate\\_Modeling.htm](http://www.gap.uidaho.edu/Bulletins/11/Improving_vertibrate_Modeling.htm).
- Augustin, N., Muggleston, M. und Buckland, S. (1996): *An autologistic model for the spatial distribution of wildlife*. Journal of Applied Ecology 33: 339–347.
- Austin, M. (1999): *A silent clash of paradigms: Some inconsistencies in community ecology*. Oikos 86: 170–178.
- Austin, M. (2002a): *Case Studies of the Use of Environmental Gradients in Vegetation and Fauna Modeling: Theory and Practice in Australia and New Zealand*. In: *Predicting Species Occurrence: Issues of Accuracy and Scale* (herausgegeben von Scott, J., Heglund, P., Morrison, M., Haufler, J., Raphael, M., Wall, W. und Samson, F.), S. 73–82, Island Press, Covelo, CA.
- Austin, M. (2002b): *Spatial prediction of species distribution: an interface between ecological theory and spatial modelling*. Ecological Modelling 157: 101–118.
- Austin, M. und Meyers, J. (1996): *Current approaches to modelling the environmental niche of eucalypts: implication for management of forest biodiversity*. Forest Ecology and Management 85: 95–106.
- Bokalo, M. (2001): *The Role of Forest Growth in Habitat Quality Dynamics*. Dissertation, Universität für Bodenkultur, Wien, Institut für Waldwachstumsforschung.
- Boyce, M. und McDonald, L. (1999): *Relating populations to habitats using resource selection functions*. Trends in Ecology and Evolution 14: 268–272.
- Corsi, F., Duprè, E. und Boitani, L. (1999): *A large scale model of wolf distribution in Italy for conservation planning*. Conservation Biology 13: 150–159.

- Crist, P. (2000): *A Handbook for Conducting Gap Analysis, Gap Analysis*. <http://www.gap.uidaho.edu/handbook/pdf/GapAnalysis.pdf>.
- Debeljak, M., Džeroski, S., Jerina, K., Kobler, A. und Adamič, M. (2001): *Habitat suitability modelling for red deer (Cervus elaphus L.) in South-central Slovenia with classification trees*. *Ecological Modelling* 138: 321–330.
- Dettmers, R., Buehler, D. und J.B., B. (2002): *A Test and Comparison of Wildlife-habitat Modeling Techniques for Predicting Bird Occurrence at a Regional Scale*. In: *Predicting Species Occurrence: Issues of Accuracy and Scale* (herausgegeben von Scott, J., Heglund, P., Morrison, M., Haufler, J., Raphael, M., Wall, W. und Samson, F.), S. 607–615, Island Press, Covelo, CA.
- Dunning, J. J., Steward, D., Danielson, B., Noon, B., Root, T., Lamberson, R. und Stevens, E. (1995): *Spatially Explicit Population Models: Current Forms and Future Uses*. *Ecological Applications* 5(1): 3–11.
- Elith, J. (2000): *Quantitative Methods for Modeling Species Habitat: Comparative Performance and an Application to Australian Plants*, Kap. 4, S. 39–58. Springer-Verlag, Berlin.
- Fersterer, P. (2000): *Spring Black Bear Timber Damage and its Management in Western Washington*. Diplomarbeit, Universität für Bodenkultur, Wien, Institut für Wildbiologie und Jagdwirtschaft.
- Fersterer, P., Nolte, D., Ziegltrum, G. und Gossow, H. (2001): *Effect of feeding stations on the home ranges of American black bears in western Washington*. *Ursus* 12: 51–54.
- Fischer, A. (2003): *Untersuchungen zur Raum-Zeit-Nutzung und zur Habitatstrukturwahl von Rotwild (Cervus elaphus L.) im Arlberggebiet unter besonderer Berücksichtigung von anthropogenen Störeinflüssen*. Dissertation, Universität für Bodenkultur, Wien, Institut für Wildbiologie und Jagdwirtschaft.
- Greif, G. (2000): *Moderne Aspekte der Wissensverarbeitung*. Diplomarbeit, Technische Universität Graz, Institut für Informationsverarbeitung und Computergestützte Medien (IICM), <http://www.iicm.edu/greif/thesis.html>.

- Grimm, V., Wyszomirski, T., Aikman, D. und Uchmański, J. (1999): *Individual-based modelling and ecological theory: synthesis of a workshop*. Ecological Modelling 115: 275–282.
- Gross, L. und DeAngelis, D. (2002): *Multimodeling: New Approaches for Linking Ecological Models*. In: *Predicting Species Occurrence: Issues of Accuracy and Scale* (herausgegeben von Scott, J., Heglund, P., Morrison, M., Haufler, J., Raphael, M., Wall, W. und Samson, F.), S. 467–474, Island Press, Covelo, CA.
- Guisan, A., Edwards, T.C., J. und Hastie, T. (2002): *Generalized linear and generalized additive models in studies of species distribution: setting the scene*. Ecological Modelling 157: 89–100.
- Guisan, A., Gonthier, Y., Cherix, D., Zimmermann, N., Kienast, F., Edwards, T., Walters, T. und Weiss, A., D. (2000): *Landscape Potential for Animal Species Survival, Colonization and Dispersal: A Spatial Simulation Study*. CSCF S. 1–35.
- Guisan, A. und Zimmermann, N. (2000): *Predictive habitat distribution models in ecology*. Ecological Modelling 135: 147–186.
- Hall, L., Krausman, P. und M.L., M. (1997): *The habitat concept and a plea for standard terminology*. Wildlife Society Bulletin 25(1): 173–182.
- Harvey, L. (1996): *Macroecological studies of species composition, habitat and biodiversity using GIS and canonical correspondence analysis*. In: *Third International Conference/Workshop on Integrating GIS and Environmental Modelling, Santa Fe, NM, January 21-26, 1996*, Santa Fe, NM.
- Hirzel, A. (2001): *When GIS come to life. Linking landscape- and population ecology for large population management modelling: the case of Ibex (Capra ibex) in Switzerland*. Dissertation, Universität de Lausanne, Faculté des Sciences, Institut d'Ecologie.
- Hirzel, A. und Guisan, A. (2002): *Which is the optimal sampling strategy for habitat suitability modelling?* Ecological Modelling 157: 331–341.
- Hirzel, A., Hauser, J., Chessel, D. und Perrin, N. (2002): *Ecological-Niche Factor Analysis: How to compute habitat-suitability maps without absence data?* Ecology 83(7): 2027–2036.



- Hirzel, A., Hausser, J. und Perrin, N. (2001a): *Biomapper*. <http://www.unil.ch/biomapper/>.
- Hirzel, A., Helfer, V. und Metral, F. (2001b): *Assessing habitat-suitability models with virtual species*. *Ecological Modelling* 145: 111–121.
- Huisman, J., Olff, H. und Resco, L. (1993): *A hierarchical set of models for species response analysis*. *Journal of Vegetation Science* 4: 37–46.
- Jeffers, J. (1988): *SCOPE34 Practitioner's Handbook on the Modelling of Dynamic Change in Ecosystems*. <http://www.icsu-scope.org/downloadpubs/scope34/index.html>.
- Kramer, K., Baveco, H., Bijlsma, R., Clerkx, A., Dam, J., van Goethem, J., Groen, T., Groot Bruinderink, G., Jorritsma, I., Kalkhoven, J., Kuiters, A., Lammertsma, D., Prins, R., Sanders, M., Wegman, R., van Wieren, S., Wijdeven, S. und van der Wijngaart, R. (2001): *Landscape forming processes and diversity of forested landscapes - description and application of the model FORSPACE*, Bd. 216 von *Alterra-rapport*. Alterra, Green World Research, Wageningen.
- Kumke, T. (2001): *Multivariate Statistik in der Ökologie*. [http://www.bio.uni\\_potsdam.de/oeksys/vstatoek.pdf](http://www.bio.uni_potsdam.de/oeksys/vstatoek.pdf).
- Lang, S. (1999): *Aspekte und Spezifika der nordamerikanischen landscape metrics innerhalb der Landschaftsökologie und experimentelle Untersuchungen zum Proximity Index*. Magisterarbeit, Universität Salzburg, Naturwissenschaftliche Fakultät.
- Lehmann, A., Overton, J. und Leathwick, J. (2003): *GRASP: generalized regression analysis and spatial prediction*. *Ecological Modelling* 160: 165–183.
- Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J. und Aulagnier, S. (1996): *Application of neural networks to modelling nonlinear relationships in ecology*. *Ecological Modelling* 90: 39–52.
- Lek, S. und Guégan, J. (1999): *Artificial neural networks as a tool in ecological modelling, an introduction*. *Ecological Modelling* 120: 65–73.
- Lek-Ang, S., Deharveng, L. und Lek, S. (1999): *Predictive models of collembolan diversity and abundance in a riparian habitat*. *Ecological Modelling* 120: 247–260.

- Mace, R., Waller, J., Manley, T., Ake, K. und Wittinger, W. (1999): *Landscape Evaluation of Grizzly bear habitat in Western Montana*. Conservation Biology 13: 367–377.
- Manel, S., J.-M., D. und Ormerod, S. (1999): *Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a Himalayan river bird*. Ecological Modelling 120: 337–347.
- Meadows, D., Meadows, D. und Randers, J. (1993): *Die neuen Grenzen des Wachstums*. Rowohlt Taschenbuch Verlag GmbH, Reinbeck bei Hamburg.
- Mladenoff, D., Sickley, T., Haight, R. und Wydeven, A. (1995): *A regional landscape analysis and prediction of favorable Gray Wolf habitat in northern Great Lakes region*. Conservation Biology 9: 279–294.
- Morrison, M., Bruce, G. und Mannan, R. (1998): *Wildlife-habitat relationship: concepts and application*. 2 Aufl., The University of Wisconsin Press.
- Morrison, M. und Hall, L. (2002): *Standard Terminology: Toward a Common Language to Advance Ecological Understanding and Application*. In: *Predicting Species Occurrence: Issues of Accuracy and Scale* (herausgegeben von Scott, J., Heglund, P., Morrison, M., Hauffer, J., Raphael, M., Wall, W. und Samson, F.), S. 43–52, Island Press, Covelo, CA.
- Özesmi, S. und Özesmi, U. (1999): *An artificial neural network approach to spatial habitat modelling with interspecific interaction*. Ecological Modelling 116: 15–31.
- Partl, E. (2001): *Die Wechselwirkungen zwischen Wald als Habitat und Rehwild als Standortfaktor bei der Waldsanierung*. Dissertation, Universität für Bodenkultur, Wien, Institut für Wildbiologie und Jagdwirtschaft.
- Pearce, J. und Ferrier, S. (2000): *Evaluating the predictive performance of habitat models developed using logistic regression*. Ecological Modelling 133: 225–245.
- Reimoser, F. und Gossow, H. (1996): *Impact of ungulates on forest vegetation and its dependence on silvicultural system*. Forest Ecology and Management 88: 107–119.

- Sauer, J. (1999): *Neuronale Netze, Fuzzy Control-Systeme und Genetische Algorithmen, Skriptum zur Vorlesung im SS 1999*. <http://rfhs8012.fh-regensburg.de/~saj39122/NN/INHALT.pdf>.
- Schmid, C. (2002): *Verallgemeinerte Lineare Modelle*. [http://www.uni-ulm.de/~cschmid/oldstat/se4\\_2.htm](http://www.uni-ulm.de/~cschmid/oldstat/se4_2.htm).
- Scott, J. (2000): *A Handbook for Conducting Gap Analysis, Preface*. <http://www.gap.uidaho.edu/handbook/pdf/preface.pdf>.
- Sharow, A. (1996): *Quantitative Population Ecology*. <http://www.ento.vt.edu/>.
- Smallwood, K. (1993): *Understanding ecological pattern and process by association and order*. *Acta Oecologica* 14: 443–462.
- Stahel, W. (2002): *Statistische Datenanalyse*. 4. Aufl., Vieweg, Braunschweig/Wiesbaden.
- Storch, I. (2002): *On Spatial Resolution in Habitat Models: Can Small-scale Forest Structure Explain Capercaillie Numbers?* *Conservation Ecology* 6(1): 6, [online] URL: <http://www.consecol.org/vol6/iss1/art6>.
- Tipton, A. (1980): *Mathematical Modeling in Wildlife Management*. In: *Wildlife management technical manual* (herausgegeben von Schemnitz, S.), Kap. 13, S. 211–220, Wildlife Soc., Washington.
- U.S. Fish and Wildlife Service (Hg.) (1980a): *Habitat Evaluation Procedures*, Bd. 102 von *Ecological Services Manual*. U.S. Fish and Wildlife Service, Washington, D.C.
- U.S. Fish and Wildlife Service (Hg.) (1980b): *Habitat as a Basis for Environmental Assessment*, Bd. 102 von *Ecological Services Manual*. U.S. Fish and Wildlife Service, Washington, D.C.
- U.S. Fish and Wildlife Service (Hg.) (1981): *Standards for the Development of Habitat Suitability Index Models*, Bd. 103 von *Ecological Services Manual*. U.S. Fish and Wildlife Service, Washington, D.C.
- USGS Gap Analysis Program (1998): *A handbook for gap analysis*. Moscow, USGS Gap Analysis Program, <http://www.gap.uidaho.edu/handbook/default.html>.

- Van Horne, B. (1983): *Density as a misleading indicator of habitat quality*. Journal of Wildlife Management 47(4): 893–901.
- Van Horne, B. (2002): *Approaches to Habitat Modeling: The Tensions between Pattern and Process and between Specificity and Generality*. In: *Predicting Species Occurrence: Issues of Accuracy and Scale* (herausgegeben von Scott, J., Heglund, P., Morrison, M., Haufler, J., Raphael, M., Wall, W. und Samson, F.), S. 63–72, Island Press, Covelo, CA.
- Völk, F. (1999): *Barkpeeling in relation to red deer density and forest structure in Austria*. In: *Wildlife, Land, and People: Priorities in the 21st Century. Second International Wildlife Management Congress Proceedings* (herausgegeben von Field, R., Warren, R., Okarma, H. und Sievert, P.), S. 79–83, The Wildlife Society, Bethesda, Maryland, USA.
- Xie, J., Liu, J. und Doepker, R. (2001): *DeerKBS: a knowledge-based system for white-tailed deer management*. Ecological Modelling 140: 177–192.
- Yee, T. und M., M. (2002): *Vector generalized additive models in plant ecology*. Ecological Modelling 157: 141–156.
- Yee, T. und N.D., M. (1991): *Generalized additive models in plant ecology*. Journal of Vegetation Science 2: 587–602.
- Ziegltrum, G. und Nolte, D. (2001): *Black bear forest damage in Washington state*. Ursus 12: 169–172.
- Zimmermann, F. und Breitenmoser, U. (2002): *A Distribution Model for the Eurasian Lynx (Lynx lynx) in the Jura Mountains, Switzerland*. In: *Predicting Species Occurrence: Issues of Accuracy and Scale* (herausgegeben von Scott, J., Heglund, P., Morrison, M., Haufler, J., Raphael, M., Wall, W. und Samson, F.), S. 653–659, Island Press, Covelo, CA.