

Master Thesis

im Rahmen des
Universitätslehrganges „Geographical Information Science & Systems“
(UNIGIS MSc) am Zentrum für Geoinformatik (Z_GIS)
der Paris Lodron-Universität Salzburg

zum Thema

„Habitatmodellierung der endemischen Landschneckenart *Trochulus biconicus*“

Der Nidwaldner Haarschnecke auf der Spur

vorgelegt von

dipl. Erdw. Beat Niederberger

u1279, UNIGIS MSc Jahrgang 2006

Zur Erlangung des Grades
„Master of Science (Geographical Information Science & Systems) – MSc(GIS)“

Gutachter:

Ao. Univ. Prof. Dr. Josef Strobl

Dr. Ariel Bergamini, WSL

Luzern, 21.4.08

Vorwort

Zum Glück habe ich vor einigen Jahren nicht allzu laut gelacht, als einer meiner Kollegen plötzlich anfang, in Geröllhalden Steine zu drehen und darunter nach Haarschnecken zu suchen und nicht wie wir anderen ganz normal nach neuen Höhleneingängen. In der Zwischenzeit drehe auch ich in dieser Gegend Steine – nur viel gezielter.

Es ist zu einem grossen Teil der abfärbenden Begeisterung des Biologen Markus Baggenstos zu verdanken, dass ich auf der Suche nach einem geeigneten Thema für meine Master Thesis auf die *Trochulus biconicus* gestossen bin. Ich hoffe, mit dieser Arbeit das Wissen zu dieser Art ein Stück vorwärts gebracht zu haben und mit deren weiteren Erforschung nicht nur emotional verbunden zu bleiben.

Es war mein Ziel, mit der Arbeit an der Master Thesis nicht nur meine Kenntnisse in geografischer Informationsverarbeitung zu zeigen, sondern auch meinen naturwissenschaftlichen Horizont etwas zu weiten. Ein biologisches Thema passt dazu hervorragend. Ich konnte mit dieser Arbeit wichtige Werkzeuge und Methoden im Umgang mit (geo-)statistischen Daten, ökologischen Zusammenhängen, geografischer Informationsverarbeitung und wissenschaftlichem Schreiben anwenden.

An der Eidgenössischen Forschungsanstalt für Wald, Schnee und Landschaft, WSL in Birmendorf fand ich schliesslich in Dr. Ariel Bergamini eine sehr gute Unterstützung. Mit seiner Hilfe konnte ich rasch das notwendige Wissen um die Methodik und Werkzeuge zur Habitatmodellierung aufbauen und auf „meinen Fall“ anwenden.

Mein grösster Dank geht daher an Dr. Ariel Bergamini für die intensive Betreuung und die spannenden Diskussionen und an Markus Baggenstos für alle Belange um die Nidwaldner Haarschnecke, das Projektumfeld und die 4 unvergesslichen Feldtage. Weiter danke ich herzlich Dr. Klaus Ecker für Umweltdaten und die Erklärungen zu deren Entstehung und meiner Lebenspartnerin Ruth Aregger für das erweiterte Korrekturlesen.

Erklärung über eigenständige Arbeit und Plagiarismus

Ich versichere, diese Master Thesis eigenständig und ohne Verwendung anderer als der angeführten Quellen angefertigt zu haben, und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat. Alle Ausführungen der Arbeit, die wörtlich oder sinngemäß übernommen wurden, sind entsprechend gekennzeichnet.

Luzern, 21.4.08

Beat Niederberger

Zusammenfassung

Trocholus biconicus ist eine kleine Landschnecke (Gehäusedurchmesser ca. 6 mm), die unter flachen Steinen lebt. Sie wurde bisher nur an wenigen, alpinen Fundstellen in der Zentralschweiz beidseits des *Engelbergertales* nachgewiesen.

Das Wissen über seltene oder schwierig zu findende Arten ist meist lückenhaft. Für deren Erforschung stehen kaum je ausreichend Mittel zur Verfügung. Neue Methoden erlauben, aus wenigen Funden zusammen mit Daten zu Topografie, Klima usw. auf das mögliche Verbreitungsgebiet einer Art zu schliessen. In dieser Arbeit wurden – ausgehend von einem Projekt zur Erforschung der *T. biconicus* – die Möglichkeiten der Habitatmodellierung getestet.

Den Modellierungsansätzen gemeinsam ist das Konzept der ökologischen Nische. Die Funktionsweise der Projektion vom geografischen Raum in den Umweltraum, aber auch die Gefahren für Fehlprognosen und Fehlinterpretationen wurden vor diesem Hintergrund diskutiert.

Der mechanistische Ansatz, über Arbeitshypothesen zu biologisch interpretierbaren Art-Umwelt-Interaktionen zu gelangen, scheiterte an der mangelnden Verfügbarkeit der dazu notwendigen Umweltvariablen. Als mögliche Alternative bot sich die statistische Untersuchung mit den vorhandenen Umweltvariablen an (korrelativer Ansatz). Qualität und Aussagekraft der Eingangsdaten spielen bei diesem Ansatz eine wichtige Rolle für die Güte der Prognosen. Die Daten zu den Beobachtungen wurden daher in einer Datenkritik untersucht. Die verwendeten Umweltvariablen wurden sorgfältig ausgewählt und deren Erstellung recherchiert.

Mit den neueren Werkzeugen zur Habitatmodellierung MAXENT und *Boosted Regression Trees* BRT und mit der etablierten Methode *Generalized Linear Models* GLM wurden in einem Untersuchungsgebiet Prognosen für die Verbreitung von *T. biconicus* erstellt.

Die verschiedenen Prognosen wurden bewertet und miteinander verglichen. Ein Mass für die Güte von Prognosemodellen ist der AUC-Wert (*Area Under the Curve*), der aus dem ROC-Plot (*Receiver Operator Characteristic*) abgeleitet ist. Ausserdem wurde eine Berechnungsart für einen geeigneten Schwellenwert gesucht, der die Wahrscheinlichkeiten aus den Prognosen in Präsenz-/Absenz-Informationen übersetzt.

MAXENT benötigt als Eingabe nur die positiven Beobachtungen (Präsenzen) und erstellt selbständig viele zufällige Punkte als sogenannte Pseudoabsenzen. GLM und BRT benötigen zusätzlich auch die Nicht-Funde (Absenzen) als Teil der Eingabe (Präsenz-/Absenzdatensatz). Als Variante wurden auch GLM und BRT mit dem Präsenz-/Pseudoabsenzdatensatz gerechnet.

Die Prognosen der verschiedenen Modelle wurden diskutiert und die Unterschiede in Unsicherheits- und Einigkeitskarten dargestellt. Diese Karten heben jene Gebiete hervor, wo sich die Prognosen, ob ein Vorkommen wahrscheinlich ist oder nicht, widersprechen. Damit lassen sich die künftigen Feldarbeiten gezielt planen: Einerseits sollen Gebiete besucht werden, wo sich die Prognosen stark widersprechen, um die Prognosen zu verbessern. Andererseits können in bisher nicht besuchten Gebieten mit hoher Prognosewahrscheinlichkeit neue Nachweise von *T. biconicus* versucht werden.

Die räumliche Extrapolation auf die umliegenden Regionen wies auf weitere potentielle Standorte hin. Die Möglichkeiten und Probleme beim projizieren in andere Gebiete wurden aufgezeigt. Die Jahresdurchschnittstemperatur erwies sich als wichtige Umweltvariable. Eine Projektion in andere klimatische Bedingungen wurde durch Veränderung dieser Umweltvariablen versucht und diskutiert.

Es konnte in dieser Arbeit nachgewiesen werden, dass die neueren Methoden MAXENT und BRT sich auch für gebirgiges Gelände sehr gut eignen.

Schlagerworte:

Endemische Landschnecke, Artverbreitungsmodellierung, Maxent, Unsicherheitskarte

Abstract

Trochulus biconicus is a tiny land snail (diameter of the shell about 6 mm), that lives underneath flat stones. Until now, it was observed only at a few alpine places in Central Switzerland on both sides of the *Valley of Engelberg*.

The knowledge about species that are rare or difficult to observe is mostly fragmentary. There are hardly enough resources available for a sufficient research of such species. Novel methods allow to predict the species distribution from a small sample size in combination with data to topography, climate et cetera. In this thesis, the possibilities of habitat suitability were investigated – based on a research project about *T. biconicus*.

The modelling approaches have the concept of the ecological niche in common. This thesis discussed the projection from the geographical space to the environmental space. Also, the risks of false prognosis and misinterpretations were mentioned.

The mechanistic approach, to obtain biologically interpretable species-environment-interactions through working hypothesis, failed because of the insufficient availability of needed environmental input variables. The statistical investigation with available environmental variables (correlative approach) was a possible alternative. Within this approach, quality and expressiveness of the input data play an important role for a good performance. Therefore, the observation data set was critically reviewed. The used environmental variables were chosen carefully and their compilation was investigated.

Predictions for the distribution of *T. biconicus* were drawn in an area of research, using the novel methods for habitat suitability modelling MAXENT and *Boosted Regression Trees* BRT and also the well-established method *Generalized Linear Models* GLM.

The different predictions were ranked and compared to each other. The AUC-value (*Area Under the Curve*) assesses the quality of a predictive performance. The AUC is derived from the ROC-Plot (*Receiver Operation Characteristic*). Furthermore, a mode of calculation for a threshold value was sought, that converts the continuous likelihood into binary presence-/absence-information.

MAXENT is a presence-only method. It autonomously generates many random points named pseudo-absences. In addition to presence data, GLM and BRT also need absence data (presence-/absence data set). As a variant, GLM and BRT were also calculated with the presence-/pseudo-absence data set.

The predictions of the different models were discussed and the differences were illustrated in uncertainty maps and consensus maps. These maps highlight areas where the predictions are not contradictory about the statement of occurrence. Therewith, future field investigations can be planned accurately: On one hand, areas should be visited where predictions are highly contradictory, in order to improve the predictions. On the other hand, new detections of *T. biconicus* can be tried in not yet visited areas with high likelihood of occurrence.

The spatial extrapolation in surrounding areas indicated further potential sites. The possibilities and problems of the projection to other regions were shown. The mean annual temperature proved to be an important environmental variable. A projection to diverging climatic conditions by shifting this variable was tried and discussed.

In this thesis it was shown that the novel methods MAXENT and BRT are highly suitable for mountainous terrain.

Keywords:

Endemic Landsnail, Maxent, species distribution modelling, uncertainty map

Inhaltsverzeichnis

1	Einführung	1
1.1	Motivation/Aufgabenstellung für die vorliegende Arbeit	2
1.2	Lösungsansatz	2
1.3	Erwartete Ergebnisse	3
1.4	Aufbau dieser Arbeit	4
2	Literaturüberblick	5
2.1	Der zu untersuchende Organismus: <i>Trochulus biconicus</i>	5
2.2	Ökologische Modellierung	6
2.3	Präsenzen, Absenzen und Pseudoabsenzen	8
2.4	Binäre und kontinuierliche Prognosen	9
2.5	Methodik biologischer Modellierung	10
2.6	Ansätze zur Habitatmodellierung	11
2.6.1	MAXENT	13
2.6.2	Generalized linear Models GLM	14
2.6.3	Boosted Regression Trees BRT	14
2.7	Die beste Methode?	16
2.8	Projektion in andere Umgebungen	17
3	Lösungsansatz	20
3.1	Modellieren natürlicher Vorgänge	20
3.2	Die ökologische Nische	21
3.3	Eingangsdaten	24
3.3.1	Beobachtungen zum Vorkommen der Art	24
3.3.2	Wahl der Umweltvariablen	24
3.4	Methodik der Modellierung	28
3.4.1	Wahl der Methode	28
3.4.2	Der AUC als Gütemass	29
3.4.3	Schwellenwerte	31
3.5	Unsicherheitskarten	33
3.5.1	Unsicherheitskarten aus den 'rohen' Prognosen: Standardabweichungskarten	34
3.5.2	Schwellenwertkarten	35
3.5.3	Methodenübergreifende Unsicherheitskarten	37

3.6	Projektion	37
3.7	Versuchsanordnung	37
4	Projektbeschreibung	39
4.1	Das Untersuchungsgebiet	40
4.2	Eingangsdaten in die Modellierung	41
4.2.1	Beobachtungen aus den Feldkampagnen	41
4.2.2	Vermessungsdaten	47
4.2.3	Umweltdaten der WSL	48
4.2.4	Geologische Karten	48
4.2.5	Satellitenbilder	48
4.3	Modellieransatz mit den verfügbaren Umweltvariablen	48
4.3.1	Reduktion aufgrund der Korrelation	49
4.3.2	Beschreibung ausgewählter Umweltvariablen	51
4.4	Modellieren mit MAXENT	57
4.4.1	Vereinfachen durch sukzessives Weglassen von Umweltvariablen	57
4.4.2	Prüfung auf Stabilität mit Teildatensätzen	58
4.5	Modellieren mit R	58
4.5.1	Vorbereitung der Beobachtungen	59
4.5.2	<i>Generalized Linear Models</i> GLM	60
4.5.3	<i>Boosted Regression Trees</i> BRT	63
4.5.4	Residuen	65
5	Ergebnisse	66
5.1	Untersuchungsgebiet	66
5.1.1	Prognosekarten	66
5.1.2	AUC	77
5.1.3	Unsicherheitskarten	78
5.2	Extrapolation	83
5.2.1	Extrapolation in andere Gebiete	83
5.2.2	Extrapolation in andere Klimate	89
6	Zusammenfassung, Diskussion, Ausblick	92
6.1	Zusammenfassung	92
6.2	Diskussion	93
6.2.1	Methodik	93
6.2.2	Werkzeuge zur Modellierung, Auswertung und Darstellung	98
6.3	Ausblick	99
A	Umweltvariablen	108

Abbildungsverzeichnis

1.1	Grafik zum Aufbau der Arbeit	4
2.1	Foto des Lebensraumes von <i>T. biconicus</i>	6
2.2	Nahaufnahme einer <i>T. biconicus</i>	6
2.3	Unterscheidung der Methoden nach benötigten Beobachtungs- daten	12
3.1	Darstellung des geografischen Raumes G und des Umwelt- Raumes U.	23
3.2	Mind Map zum mechanistischen Ansatz.	26
3.3	ROC-Plot mit Test- und Trainingskurve.	30
3.4	Die Konstruktion des Schwellenwertes nach Lobo et al. (2007).	34
4.1	Karte des Untersuchungsgebiets.	40
4.2	Karte der Beobachtungen: Präsenz, Absenz und Erstnachweis.	43
4.3	Mind Map zum korrelativen Ansatz.	49
4.4	Korrelationsmatrix mit Korrelationskoeffizient $r \leq 0.80$	52
4.5	Darstellung zur Aufteilung der Exposition in <i>Eastness</i> und <i>Northness</i>	54
4.6	MAXENT: das <i>User interface</i>	58
4.7	MAXENT: weitere Einstellungen.	58
5.1	Prognosekarte mit MAXENT Level A.	68
5.2	Prognosekarte mit MAXENT Level B.	69
5.3	Prognosekarte mit MAXENT Level C.	70
5.4	Prognosekarte mit MAXENT Level D.	71
5.5	Prognosekarte mit GLM Präsenz-/Absenzdaten.	72
5.6	Prognosekarte mit GLM Präsenz-/Pseudoabsenzdaten.	73
5.7	Prognosekarte mit BRT Präsenz-/Absenzdaten.	74
5.8	Prognosekarte mit BRT Präsenz-/Pseudoabsenzdaten.	75
5.9	Standardabweichungskarte von 10 Testläufen mit MAXENT Level A.	78
5.10	Standardabweichungskarte der Rohprognosen aus MAXENT Level A bis D.	80
5.11	Unsicherheitskarte MAXENT Level A bis D mit 'Sensitivität gleich Spezifität'.	81

5.12	Einigkeitskarte MAXENT Level A bis D mit 'Sensitivität gleich Spezifität'	81
5.13	Methodenübergreifende Unsicherheitskarte mit 'Sensitivität gleich Spezifität'	83
5.14	Methodenübergreifende Einigkeitskarte mit 'Sensitivität gleich Spezifität'	84
5.15	Die räumliche Extrapolation mit MAXENT Level B und Level C auf ein grösseres Gebiet.	85
5.16	Abhängigkeitscharakteristik von <code>Prec7</code> und <code>PDay</code>	86
5.17	Die Extrapolation in Gebiete mit deutlich anderen Umweltbedingungen verursacht 'Spannungen'.	87
5.18	Lokalnamen im umliegenden Gebiet.	88
5.19	Extrapolation der Habitateignung mit MAXENT (Level B) um plus 1°C.	90
5.20	Darstellung der Gebiete, welche bei einer Klimaänderung von 2°C ständig innerhalb der ökologischen Nische bleiben.	91
6.1	Charakteristik der Jahresdurchschnittstemperatur in MAXENT Level B und der Jack-Knife-Test dazu.	95
6.2	Charakteristik von <code>Wet</code> , <code>TWI</code> und <code>TWIs</code> aus MAXENT Level A.	97
A.1	Wertebereiche der Umweltvariablen für Fundstellen, Absenzen und Pseudoabsenzen	109

Tabellenverzeichnis

4.1	Tabelle der verwendeten Umweltvariablen.	50
4.2	Korrelationen mit Korrelationskoeffizienten $r \leq 0.85$	51
5.1	AUC-Werte der Prognosemodelle und den Validierungen.	77
5.2	Kennwerte der Standardabweichungskarten Level A bis D.	79

Abkürzungsverzeichnis

Abkürzungen im Text:

Abb.	Abbildung
bzw.	beziehungsweise
et al.	und andere
f.	folgende (Seite)
ff.	fortfolgende (Seiten)
BRT	Boosted Regression Trees / Generalized Boosted Model, siehe Abschnitt 2.6.3
GLM	Generalised Linear Model, siehe Abschnitt 2.6.2
LPT	Schwellenwert 'tiefster nachgewiesener Fund' (<i>lowest predicted threshold</i>)
Lobo <i>oder</i> Spz=Sns	Schwellenwert 'Spezifität gleich Sensivität' nach Lobo et al. (2007)
MaxKappa	Schwellenwert 'Maximales κ '
PsA	Pseudoabsenz(en)
10Proz <i>oder</i> 10%	Schwellenwert '10%-Quantil'

Abkürzungen im Literaturverzeichnis:

edn.	Edititor
pp.	Seiten
Tech.Rep	Technischer Report / Bericht
vol.	Volume

Kapitel 1

Einführung

Das Projekt **Nidwaldner Haarschnecke** bezweckt, Wissen über die Verbreitung und Biologie der endemischen Art *Trochulus biconicus* aufzubauen. Die Art wurde von Eder 1916 am Chaiserstuel entdeckt und danach von ihm wissenschaftlich beschrieben (Eder, 1917). Nach dieser vor allem systematisch geprägten Arbeit wurden keine weiteren publizierten Nachweise bekannt. Seit 2004 wurden beidseits des Engelbergertales und an Stellen, wo Eder vor 90 Jahren vergeblich gesucht hatte, Neufunde gemacht. Diese werfen Fragen auf und verlangen nach Interpretationen:

Im Rahmen der Arbeiten zur Aktualisierung der Roten Liste der Landschnecken der Schweiz widmet sich ein Projekt des Biologen Markus Baggenstos und der betroffenen Kantone Obwalden, Nidwalden, Uri und Bern der Erhebung und Verbreitung der Nidwaldner Haarschnecke (*T. biconicus*) (Baggenstos, 2006, 2007, in prep.). Es soll Auskunft über folgende Fragen geben:

- a) Wo kommt die Art heute vor und wie stark gefährdet ist sie?
- b) In welchen Gebieten (innerhalb und ausserhalb des Untersuchungsperimeters) soll zusätzliche Feldarbeit gemacht werden, um bisher unbekannte Populationen zu finden?
- c) Gibt es Hinweise auf die Ansprüche der Art an ihren Lebensraum und können daraus Schutzmassnahmen abgeleitet werden?

- d) Findet ein genetischer Austausch zwischen den heute bekannten Populationen statt oder handelt es sich um isolierte Teile einer voreiszeitlich zusammenhängenden Population?
- e) Welche Auswirkungen auf das Verbreitungsgebiet der Art sind durch die prognostizierten Klimaveränderungen zu erwarten?
- f) Können neue Erkenntnisse über die Lebensweise und die Verbreitungsgeschichte der Art gewonnen werden?

1.1 Motivation/Aufgabenstellung für die vorliegende Arbeit

Das Wissen über Biologie und Verbreitung seltener oder schwierig zu findender Arten ist meist lückenhaft. Zur Erforschung solcher Arten stehen kaum je ausreichend Mittel zur Verfügung. Neue Methoden erlauben, aus wenigen Funden in Kombination mit Daten zu Topografie, Klima usw. auf das mögliche Verbreitungsgebiet einer Art zu schliessen. Somit können die vielversprechendsten Gebiete für bisher unbekannte Populationen im Feld gezielt untersucht werden. Die Möglichkeiten und Grenzen solcher Habitatmodelle sollen im Rahmen der Untersuchungen zur Nidwaldner Haarschnecke auf ihre Gebirgstauglichkeit geprüft werden.

1.2 Lösungsansatz

Die oben unter a) bis f) gestellten Fragen müssen analysiert und auf ihre Umsetzbarkeit geprüft werden. Im Rahmen dieser Arbeit soll geprüft werden, ob das potentielle Habitat von *T. biconicus* modelliert werden kann. Aus dieser grundsätzlichen Aufgabe ergeben sich folgende Teilfragen, die in dieser Arbeit behandelt werden:

- 1) Kann ein Modell aufgebaut werden, welches auf empirisch erkannten Zusammenhängen zwischen Ausprägungen der Umwelt und dem Vorkommen der Art beruht? Können diese Ausprägungen der Umwelt

durch vorhandene Umweltvariablen abgebildet oder durch geeignete Verfahren von diesen abgeleitet werden?

- 2) Können mit den vorhandenen Umweltvariablen Modelle entwickelt werden, die plausible Prognosekarten liefern, ohne die empirischen Zusammenhänge zu berücksichtigen?
- 3) Eignen sich die Modelle auch für Extrapolationen in andere Bedingungen (räumliche Extrapolation, andere Klimate)?

1.3 Erwartete Ergebnisse

Das Ergebnis dieser Arbeit ist zum Einen eine kartografische Darstellung der prognostizierten Verbreitung. Darin sind die Gebiete mit potentiellen Vorkommen ersichtlich. Zusätzlich sind Gebiete mit geringer Prognosesicherheit markiert, denn diese sind für die Prüfung der Modelle wichtig. Diese Angaben erlauben, die weitere Feldarbeit gezielt zu planen. Zum Anderen ist der Nachweis zu erbringen, dass die angewandten Methoden den starken Gradienten im Gebirge und der Kleinräumigkeit der Habitate gewachsen sind. Dazu werden die Funddaten der Felderhebung vom Sommer 2007 und früherer Funde herangezogen und in einem Untersuchungsgebiet behandelt, das sich möglichst mit dem Interessengebiet des Forschungsprojektes deckt.

Durch die Resultate dieser Arbeit sollen die nächsten Feldarbeiten gezielt geplant und durchgeführt werden können. Die erarbeitete Methodik soll als eine Referenz für weitere Habitatmodellierungen unter ähnlichen Umständen dienen. Weitergehende biologische Interpretationen werden im übergeordneten oder in nachfolgenden Projekten behandelt.

1.4 Aufbau dieser Arbeit

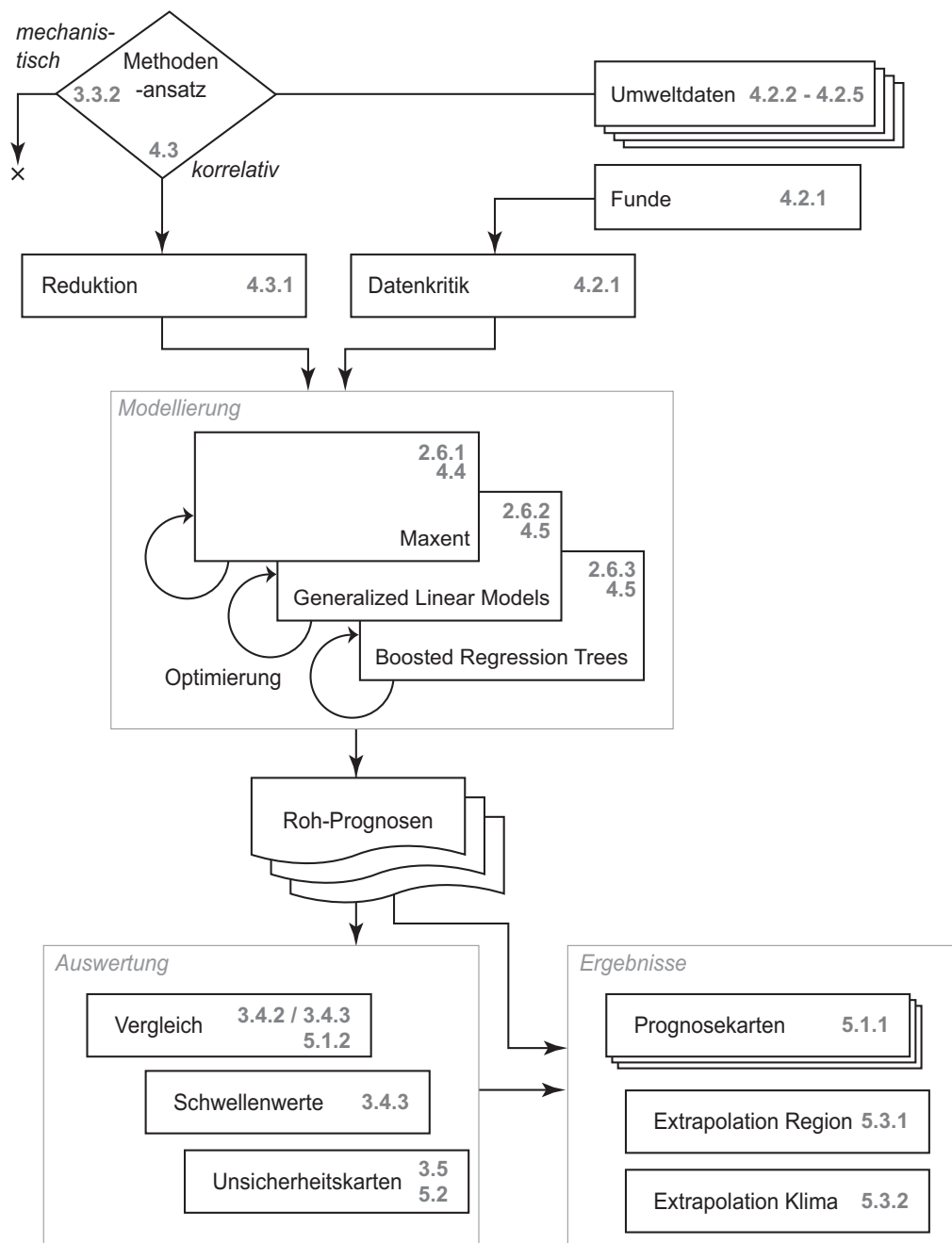


Abbildung 1.1: Grafisches Schema zum Aufbau der Arbeit. Die Nummerierung der entsprechenden Kapitel ist in grau angegeben.

Kapitel 2

Literaturüberblick

2.1 Der zu untersuchende Organismus:

Trochulus biconicus

T. biconicus wurde im Sommer 1916 von Leo Eder entdeckt und in Eder (1917) beschrieben. Eder und nach ihm Generationen von Schneckenforschern suchten die Art anschliessend vergeblich an vielen anderen Orten. Die Population am Chaiserstuel blieb die einzige und galt daher als stark gefährdet und als vom Aussterben bedroht (Abbildung 2.1). Erst Feldarbeiten zur Überprüfung der Roten Liste der Schweizer Landschnecken brachten 2004 neue Standorte zu Tage: Jörg Rüetschi gelang auf der Urner Seite der Bannalper Schonegg ein neuer Fund, an einer Stelle wo Eder vergeblich gesucht hatte. Markus Baggenstos entdeckte weitere Populationen auf dem Schwalmis (UR) und auf dem Widderfeld (OW).¹

T. biconicus ist eine lokalendemische Art. Das heisst, dass sie nur in einem sehr begrenzten Gebiet auftritt. Bei der Beurteilung der Schutzbedürftigkeit von Arten wird Endemismus und begrenztes Verbreitungsgebiet als wichtiges Kriterium aufgeführt (Baur et al., 2004)².

¹ Die Ausführungen zur *T. biconicus* sind aus einem Pressebericht von M. Baggenstos vom 19.8.2005, dem Projektbeschrieb (Baggenstos, 2007) und mündlichen Mitteilungen von M. Baggenstos zusammengestellt.

² Kapitel 11.6 "Prioritäten festlegen", Seite 204.



Abbildung 2.1: Der Lebensraum ist oft an kammnahen Stellen und gekennzeichnet durch ein Mosaik von Blaugras und flachen Steinen.



Abbildung 2.2: Nahaufnahme von *T. biconicus*. Das Gehäuse hat etwa 6 mm Durchmesser. Fotos: M. Baggenstos, 15.10.2007 am Chaiserstuel.

T. biconicus ist eine dunkle Landschnecke mit einem 5-7 mm grossen, dünnchaligen Gehäuse. Ihr Lebensraum sind Blaugrashalden oberhalb etwa 2100 m ü.M. mit flachen Steinen, unter welchen die Schnecke sich meist aufhält (Abbildungen 2.1 und 2.2).

Pfenniger et al. (2005) untersuchten anhand von DNA-Sequenzen die Stammesgeschichte der Gattung der Haarschnecken (*Trochulus*). Dabei zeigte sich, dass der gemeinsame Vorfahre Haare in Form von hornartigen Auswüchsen der äussersten Schalenschicht auf der Behausung trug. Diese Haare sind aber im Verlaufe der Evolution bei jenen Arten wieder verloren gegangen, die in weniger feuchten Habitaten leben - so auch bei *T. biconicus*.

Im Rahmen der laufenden Untersuchungen wurde in einem Teilprojekt auch die Aktivität im Tagesverlauf der Schnecken untersucht. Wigger (2007) beobachtete Fortbewegungs-Distanzen bis 30 cm innerhalb von 6 Stunden. Zudem zeigte sich, dass die Tiere vor allem nachtaktiv sind.

2.2 Ökologische Modellierung

In den letzten Jahren wurde die Modellierung der potentiellen Verteilung von Arten ein wichtiges Werkzeug zur Bearbeitung von Fragestellungen in der Ökologie, der Biogeografie, der Evolution und zuletzt auch in der Klima-

änderungsforschung (Guisan & Thuiller, 2005). Die theoretischen Ansätze indes sind meist schon einige Jahrzehnte alt, aber erst in neuerer Zeit durch geeignete Software der ökologischen Modellierung leichter zugänglich geworden.

Das Prinzip der ökologischen Modellierung ist, Funde aus dem geografischen Raum durch Umweltparameter in den *Umwelt-Raum* zu übersetzen, dort die Verteilungsmuster zu bestimmen und als Vorhersagekarte wieder zurück in den geografischen Raum zu projizieren. Das Prinzip wurde von Pearson (2007) anschaulich aufgearbeitet und dargestellt. Dieser theoretische Rahmen wurde in dieser Arbeit weitgehend übernommen und auf die speziellen Vorkommnisse angepasst (Abschnitt 3.2).

Mit der Debatte zur Klimaänderung hat sich das Interesse für Modellierungen für zukünftige Verbreitungen von Arten verstärkt. Die Zahl der Publikationen stieg in den letzten Jahren enorm. Trotz immer ausgefeilteren Methoden und verbesserten Eingangsdaten bleiben gewisse Einschränkungen der Aussagekraft aber bestehen. So ist die Anwendung von Modellen auf andere (meist zukünftige) Klimaten auf Grund der meist fehlenden Möglichkeiten zur Eichung der Modelle eine Herausforderung. Aufgrund der zahlreichen Unsicherheiten und Annahmen warnte Dormann (2007) davor, sich zu stark auf solche Prognosen zu verlassen. Neben statistischen Stolpersteinen kritisierte er unter anderem, dass keine genetischen Anpassungen berücksichtigt würden oder dass von heute als limitierend erkannten Faktoren angenommen werde, dass sie auch unter den veränderten Bedingungen limitierend sein sollten.

Weitere Prognoseunsicherheiten entstehen, weil die Modelle meist auf der aktuellen, besetzten ökologischen Nische basieren, bzw. auf der effektiven Verbreitung, und nicht auf der fundamentalen Nische (siehe Kapitel 3.2). Der Prognosefehler wird umso grösser sein, je geringer die Kenntnis über die fundamentale Nische ist. Die Veränderung biotischer Interaktionen unter den zukünftigen Bedingungen sind eine weitere Unbekannte. Es ist allerdings schwierig, das Ausmass dieser Unkenntnis abzuschätzen. Einen Hinweis über die Vorhersageleistung eines Modells geben Extrapolationen in andere Gegenden mit unabhängigen Funddaten. Allerdings sind solche Tests für Loka-

lendemiten nicht möglich. Nach Guisan & Thuiller (2005) und Davis et al. (1998) wurde die Ausbreitungskapazität einer Art meist zu wenig in die Berechnung einbezogen.

Das zweite wichtige Anwendungsfeld, welches zu einer Zunahme der Forschungsaktivitäten führte, ist das Ausarbeiten von Artenschutzmassnahmen. Modelle werden anstelle von kostenintensiver Feldarbeit eingesetzt. Allerdings ist auch hier ein Augenmerk auf die Unsicherheiten zu werfen. Guisan & Thuiller (2005) versprachen sich Verbesserungen durch die Berücksichtigung von Ausbreitungsmöglichkeiten von Individuen und Populationen sowie den Einbezug von Populationsdynamik und biotischen Interaktionen. Sie führten noch weitere Verbesserungsmöglichkeiten an, welche jedoch die Modelle zunehmend komplexer werden liessen. Damit wachsen die Ansprüche an die Datengrundlagen und es droht ein Verlust der Übersicht.

2.3 Präsenzen, Absenzen und Pseudoabsenzen

Im Idealfall liefert eine Felderhebung für jede Rasterzelle eine Beobachtung über das Vorkommen (Präsenz) oder Nicht-Vorkommen (Absenz) der Art. Diese Informationen sind so aber nur durch Rasterbeprobung in kleineren Untersuchungsgebieten zu erhalten. Bei einem Untersuchungsgebiet der Grösse wie es für diese Studie gewählt wurde und bei einem starken Relief, ist dieses Vorgehen jedoch nicht praktikabel.

Bei der Planung der Felderhebungen 2007 wurde ein üblicher Ausweg aus diesem Dilemma gewählt: Die Erhebungen wurden auf Potentialgebiete konzentriert, die aufgrund einer Expertenhypothese eine erhöhte Wahrscheinlichkeit für das Vorkommen versprachen (Baggenstos, 2007). Die Verteilung der Beobachtungen bleibt so jedoch auf das Potentialgebiet beschränkt. Dem Modell fehlt die Information, dass ausserhalb des Potentialgebiets das Vorkommen unwahrscheinlich ist (z.B. Talboden). Ausserdem befinden sich viele der Absenzen in einem Gebiet, das möglicherweise die Anforderungen der fundamentalen Nische erfüllt - das Nicht-Vorkommen jedoch anderweitig zu erklären ist (siehe Abschnitt 3.2). Die Methode muss also sowohl mit einer

Mangelinformation ausserhalb des Potentialgebietes wie auch mit möglicherweise widersprüchlicher Information innerhalb des Potentialgebiets auskommen.

Als Alternative bieten sich sogenannte Pseudoabsenzen an: Über das ganze Untersuchungsgebiet wird eine grosse Anzahl³ Punkte zufällig ausgewählt und als scheinbare Absenzen oder Pseudoabsenzen festgesetzt.

2.4 Binäre und kontinuierliche Prognosen

Meist sollen Habitat-Prognosen eine Karte mit Gebieten liefern, deren Umweltbedingungen für eine Art geeignet oder ungeeignet sind (binär). Die Resultate der Modellierungen sind jedoch oft (kontinuierliche) Wahrscheinlichkeiten zum Vorkommen (GLM und BRT), respektive zum Grad der Habitateignung, gemessen am modellierten Optimum (MAXENT)⁴. Um die gewünschten Binärkarten zu erhalten, müssen diese Roh-Prognosen durch Schwellenwerte⁵ in binäre Aussagen übersetzt werden (Abschnitt 3.4.3). Dabei geht allerdings ein gewisser Grad an Information verloren. Die Wahl des geeigneten Schwellenwertes wird vielfach diskutiert und unterschiedlich ausgelegt (Reineking & Schröder, 2004; Pearson, 2007; Lobo et al., 2007). Liu et al. (2005) stellte 12 Methoden zur Schwellenwertbestimmung einander gegenüber: insbesondere feste Schwellenwerte, d.h. solche, die nicht aus den modellierten Daten abgeleitet werden, versprechen wenig gute Resultate, aber auch die viel verwendete Maximierung des Kappa-Wertes ist wenig geeignet. Methoden, welche auf der Spezifität und der Sensitivität aufbauen, liefern brauchbare Schwellenwerte.

³ Als Voreinstellung werden in MAXENT 10'000 verwendet. Für die anderen Methoden wurden gleich viele Punkte verwendet. Das ist im Falle des in dieser Studie verwendeten Untersuchungsgebietes von 839×751 Zellen jeder 63. Punkt.

⁴ Solche Wahrscheinlichkeitskarten werden in dieser Arbeit als *rohe Prognosen* bezeichnet.

⁵ Ein Schwellenwert gibt an, ab welcher Wahrscheinlichkeit bzw. Habitateignung ein Vorkommen (für Werte grösser als der Schwellenwert) angenommen werden kann. Der einfachste Schwellenwert ist 50%.

2.5 Methodik biologischer Modellierung

Aus der Erfahrung von bisherigen Fundstellen auf mögliche weitere zu schließen, kann als der klassische, biologische Ansatz bezeichnet werden, Arten im Feld zu suchen. Die Erfahrung entsteht durch eine Kombination von wahrgenommenen Umweltbedingungen und den hypothetischen Wirkungen dieser Umweltbedingungen auf die Art. Leider lässt sich dieser subjektive 'Expertenansatz' weder exakt beschreiben, noch flächig auf das ganze Untersuchungsgebiet übertragen und schon gar nicht automatisieren. Die rechnergestützte Modellierung funktioniert jedoch ganz ähnlich: Aus den Umweltbedingungen an den Fundstellen - dargestellt durch die Umweltvariablen - wird durch statistische Methoden auf die Anforderungen der Art an die Umweltbedingungen geschlossen.

Als Kritik an den verwendeten Modellen ist anzufügen, dass jeweils von stationären Bedingungen ausgegangen wird. So wird angenommen, dass die heutige Verbreitung im Gleichgewicht mit den herrschenden Bedingungen ist (Dormann, 2007)⁶.

Insofern sind die verwendeten Modellansätze für die Prognostik durchwegs unvollständig. Zusätzlich muss angenommen werden, dass der derzeit beobachtete Zustand möglicherweise ebenfalls transient ist (d.h. in einem Übergang befindlich). Ein Veränderungsprozess kann kaum aus der Beobachtung eines einzelnen Zustandes abgeleitet werden.

Wahl des Ansatzes

Die Wahl des Ansatzes wird durch das Hauptinteresse beeinflusst: Sind die relativen Einflüsse der Wirkungsvariablen wichtig und somit eine interpretierbare Lösung (Ansatz in Abschnitt 3.3.2) oder liegt das Hauptinteresse bei einer möglichst optimalen Prognose, welche nicht unbedingt eine direkte, biologische Interpretation der Einflussgrößen zulässt (Ansatz in Abschnitt 4.3)?

⁶ Siehe auch Pearson (2007) und die Ausführungen im Abschnitt 3.2

Nach Pearson (2007) können diese Ansätze mit 'mechanistisch' bzw. 'korrelativ' bezeichnet werden⁷. Der mechanistische Ansatz findet vorzugsweise Verwendung, wenn aus dem Resultat biologische Erkenntnisse bzw. kausale Zusammenhänge und daraus Schutz- oder Nutzungsmöglichkeiten abgeleitet werden sollen⁸. Der korrelative Ansatz hingegen findet Verwendung für Fragestellungen nach dem potentiellen Vorkommen und der Habitateignung der untersuchten Art(en). Bei den korrelativen Ansätzen sind solche zu bevorzugen, die nicht nur jede Variable einzeln nach der Korrelation mit den Fundstellen untersuchen, sondern auch die Interaktionen unter den Variablen berücksichtigen (Elith et al., 2006). Das heisst, es werden alle n Variablen in einem n -dimensionalen Raum untersucht. Die meisten Ansätze sind jedoch als Kombination dieser beiden Extreme 'mechanistisch' oder 'korrelativ' anzusehen.

Die Fragestellung beeinflusst also die Wahl des Ansatzes. Eine Zusammenstellung einiger Ziele und publizierte Beispiele dazu sind in Pearson (2007), Table 1 aufgeführt. Ein Entscheidungsbaum zur Wahl des geeigneten Ansatzes ist in Guisan & Zimmermann (2000)⁹ aufgezeichnet.

2.6 Ansätze zur Habitatmodellierung

Eine ausführliche und fundierte Übersicht zu den in der Ökologie verwendeten Modellen für die Prognose von Artverteilungen gaben Guisan & Zimmermann (2000). Allerdings ist diese schon etwas älter und somit sind die neuesten Entwicklungen darin nicht berücksichtigt. Eine aktuellere Zusammenstellung mit Angaben zu den verwendeten Methoden findet sich in Guisan & Thuiller (2005)¹⁰. Anhand der verwendeten Beobachtungsdaten können 3 Gruppen

⁷ Andere Bezeichnungen sind 'mechanisch-empirischer' bzw. 'statistischer (probabilistischer)' Ansatz (Hengl, 2007).

⁸ Guisan & Zimmermann (2000) unterschieden die Wirkungsvariablen gemäss der Art ihrer Wirkung in *direkte* und *indirekte*. Indirekte Wirkungsvariablen sind beispielsweise die Topografie oder das Mesoklima, welche die direkten Wirkungsvariablen wie Boden oder verfügbare Feuchtigkeit beeinflussen. Je nach Fokus der Modellierung sind eher direkte oder eher indirekte Wirkungsvariablen zu verwenden.

⁹ Fig. 4, Seite 154: Kriterien für die Modellauswahl.

¹⁰ Tabelle 3, Seite 1001: Publizierte Methoden zur Erstellung von Prognosen mit Speziesverteilungsmodellen (Referenzen, Methoden und Links).

von Methoden zur Modellierung unterschieden werden (Elith et al., 2006):

- In die erste Gruppe fallen die Methoden, welche nur Funddaten (Präsenzen) verwenden, wie z.B. BIOCLIM. Diese Methoden versuchen eine Charakterisierung der Umweltbedingungen an den Fundstellen und prognostizieren aufgrund der Ähnlichkeit weitere Vorkommen¹¹.
- Die zweite Gruppe umfasst die verschiedenen Ansätze, welche auf Regressionen aufbauen. Zu diesen gehören die häufig angewendeten *Generalized linear models* GLM und die *Generalized additive models* GAM. Sie suchen nach Zusammenhängen von Auftreten und Umweltvariablen und benötigen dazu Präsenzen und Absenzen. Die Absenzen können auch durch Pseudoabsenzen ersetzt werden.
- Zur dritten Gruppe gehören die neuesten Entwicklungen aus dem Forschungsgebiet des *machine learning*, wie MAXENT. Diese Methoden untersuchen, wie sich das Vorkommen (Präsenzen) vom Hintergrund (Pseudoabsenzen) abhebt.

Die drei Gruppen unterscheiden sich signifikant im Modellierungserfolg¹². Eine weitere Unterteilung kann anhand der notwendigen Beobachtungen vorgenommen werden (Abb. 2.6).

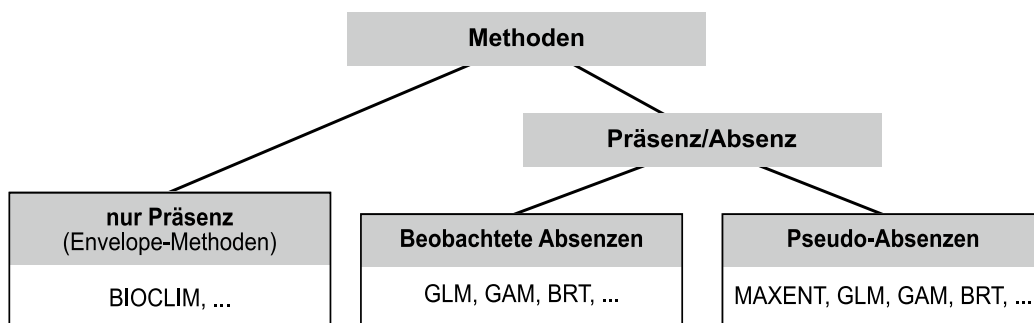


Abbildung 2.3: Die Methoden können gemäss der benötigten Beobachtungen des Vorkommens unterschieden werden.

¹¹ Diese sind auch unter dem Begriff *Envelope-Methoden* bekannt.

¹² Siehe Elith et al. (2006), Fig. 3 und Appendix, Fig. S2.

2.6.1 MAXENT

MAXENT ist eine universell einsetzbare Methode zur Charakterisierung von Wahrscheinlichkeitsverteilungen aus unvollständiger Information. Die Ursprünge von MAXENT liegen in der Sprachforschung und wurden durch Phillips et al. (2004) in die Ökologie eingeführt. Der theoretische Ansatz beruht auf dem Prinzip, wonach die geschätzte Wahrscheinlichkeitsverteilung von den Umweltbedingungen an den Fundstellen abgeleitet werden kann. Gleichzeitig sollen Annahmen vermieden werden, welche von den Daten nicht gestützt werden. MAXENT versucht, die Wahrscheinlichkeitsverteilung der maximalen Entropie zu finden, das heisst diejenige Verteilung, welche am gleichmässigsten über das ganze Gebiet verteilt ist, eingeschränkt durch die an den Fundstellen angetroffenen Umweltbedingungen (Phillips et al., 2006; Pearson, 2007; Gibson et al., 2007)¹³. Elith et al. (2006) zeigten auf, dass MAXENT im Vergleich zu anderen Methoden sehr gute Resultate erzielt.

Wie unter 2.6 beschrieben, benötigt MAXENT keine Absenz-Beobachtungen für die Modellierung des Verbreitungsgebietes einer Art. Die Methode ist nicht auf die Verwendung von kontinuierlichen Umweltvariablen beschränkt. Sie kann auch kategorische Umweltvariablen, wie Informationen zur Bodenbedeckung, berücksichtigen. Das Resultat ist eine Karte mit Werten, die eine relative Habitataignung anzeigen¹⁴.

Für die präzise, mathematische Definition von MAXENT und für eine ausführlichere Diskussion der Anwendung auf Artverteilungsmodellierung sei auf Phillips et al. (2004), Dudík et al. (2004) und Phillips et al. (2006) verwiesen. MAXENT berechnet auch eine Anzahl von Schwellenwerten, eine Modellvalidierungs-Statistik und ermöglicht mittels einer Jack-Knife-Prozedur, den Beitrag jeder Umweltvariablen zum Modell einzeln abzulesen. Die Implementierung ist betriebssystemunabhängig in Java und stellt ein benutzerfreundliches, grafisches Interface zur Verfügung. Das Programm lässt sich

¹³ Die Berechnungen mit Maxent wurden mit Version 3.1 durchgeführt. Für die Extrapolationen wurde jedoch auf die neuere Version 3.2.1 zurückgegriffen, da sie eine verbesserte Darstellung der Spannungen ('clamping') aufweist. Die Software kann unter <http://www.cs.princeton.edu/~schapire/maxent/> bezogen werden.

¹⁴ Verschiedene Ausgabevarianten stehen zur Auswahl: rohe Wahrscheinlichkeitswerte, kumulative Werte, welche in der Summe über das Untersuchungsgebiet 1 ergeben oder logistische Werte, welche auf einen Wertebereich von 0 bis 1 normiert sind.

auch im Batch-Modus ansprechen.¹⁵

2.6.2 Generalized linear Models GLM

Unter dem Begriff *Generalisierte Lineare Modelle* GLM werden logistische Regressionen zusammen mit Poisson-Regressionen, 'normalen' Regressionen sowie weiteren Spezialfällen zusammengefasst.

Bei der Untersuchung von Regressionen wird die Beziehung von unabhängigen Variablen (hier die Umweltbedingungen) zu einer abhängigen (hier die Beobachtungen der Art) gesucht. Ein Spezialfall von Regressionen stellen die sogenannten logistischen Regressionen dar. Logistische Regressionen sind immer dann nützlich, wenn die abhängige Variable nur die Werte 0 und 1 annehmen kann. Mit normalen Regressionen sind solche Variablen nicht geeignet zu modellieren, da sie unsinnige Werte unterhalb von 0 und oberhalb von 1 liefern. Bei logistischen Regressionen werden die abhängigen Variablen einer sogenannten 'Logit Transformation' unterzogen und liegen danach als Wahrscheinlichkeiten mit Werten zwischen 0 und 1 vor (Hastie et al., 2001).

Durch Parametrisieren wird jene Linearkombination der erklärenden Variablen gesucht, welche die Beobachtungen möglichst optimal annähert. Es können auch die quadrierten Umweltvariablen mit einbezogen werden, um allfällige quadratische Abhängigkeiten aufzulösen. Weiter kann die Signifikanz für jede erklärende Variable berechnet werden. So können Rückschlüsse auf deren Aussagekraft gewonnen werden und nicht-signifikante Variablen können weggelassen werden. In das finale Modell werden schliesslich nur die signifikanten Umweltvariablen aufgenommen. Das Ergebnis ist ein Satz von Regressionskoeffizienten (Hengl, 2007).

2.6.3 Boosted Regression Trees BRT

Boosted Regression Trees sind eine Erweiterung von Regressionsbäumen (*Regression Trees*). Regressionsbäume beschreiben den Variablenraum durch

¹⁵ Eine kurze Einführung in die Handhabung von Maxent und die Interpretationsmöglichkeiten der Resultate ist in Phillips (2006) gegeben.

(Knoten-)Kriterien, welche eine Menge Objekte sukzessive in zwei immer homogener werdende Teile zerlegen (Hengl, 2007). An den Ästen einer solchen Gabelung setzt das jeweils nächste Kriterium an (binäre Bäume). Dabei wird immer jenes Kriterium durch Regression gesucht, welches die Beobachtungen am besten in zwei Gruppen aufteilt¹⁶. Das Ende eines Baumes ist erreicht, wenn am Ende jedes Astes ein einzelnes Objekt (maximales überfitten) bzw. identische Objekte sind oder das zu definierende Mass an Aufteilungsschärfe zwischen den verbleibenden Objekten unterschritten wird. Ein übliches Mass für die Aufteilungsschärfe ist der quadratische Fehler.

Regressionsbäume sind wie die multiple logistische Regression geeignet, Modelle für binäre Beobachtungen zu finden.

Ein Nachteil von Bäumen ist, dass das Weglassen oder Hinzufügen von wenigen Beobachtungen zu ganz anderen Baumstrukturen führen kann. Dies kann sich z.B. bei Kreuzvalidierungen deutlich auswirken. Entsprechend vorsichtig sind die Bäume zu interpretieren. Ebenso als Nachteil kann die binäre Natur empfunden werden, bei der jeweils nur ein einziger Wert geprüft werden kann (z.B. untere Begrenzung des Vorkommens in der Höhenlage). Bei genügender Grösse eines Baumes wiegt sich dieser Nachteil allerdings ziemlich auf.

Bei der Modellierung von Habitaten sind die Kriterien an den Astgabelungen aus dem Satz an Umweltvariablen entnommen. Dabei werden die Werte an den Fundstellen auf den Wertebereich des ganzen Untersuchungsgebiets angewendet¹⁷. Für detailliertere Erklärungen und ein illustratives Beispiel anhand eines Spam-Filters siehe Hastie et al. (2001).

Die Boosting-Methode

Kriterien, die nur wenig besser unterscheiden als zufällig, sind - wie oben angesprochen - ein Problem. Freund & Schapire (1997) haben eine Methode

¹⁶ Die zwei Gruppen sind hier 'Vorkommen wahrscheinlich' bzw. 'Vorkommen nicht wahrscheinlich'.

¹⁷ Zum besseren Verständnis siehe auch die Verteilungsplots der Umweltvariablen im Anhang A. Dabei sind die Präsenzen die Werte aus den Fundstellen, und die Pseudoabsenzen stellen eine Annäherung an den ganzen Wertebereich des Untersuchungsgebiets dar.

eingeführt, die durch Iterationen mit jeweils stärker gewichteten Fehlern eine Verbesserung anstrebt. Beim ersten Durchgang sind noch alle Beobachtungen mit $\frac{1}{N}$ gewichtet. Die Anlage entspricht so dem klassischen Vorgehen. Beim nächsten Durchgang werden diejenigen Beobachtungen stärker gewichtet, welche beim vorherigen Durchgang falsch vorhergesagt wurden und jene mit richtiger Vorhersage schwächer. Dies wird m -Mal wiederholt. Die in den Iterationsstufen ermittelten Kriterien werden nun entsprechend ihrem Erfolg in der Fehlerminimierung gewichtet und zur resultierenden Struktur kombiniert.

2.7 Die beste Methode?

Die Frage nach der besten Methode lässt sich nicht *a priori* beantworten: Die Funddaten, die Verwendung von Absenzen oder Pseudoabsenzen, das Verhalten der Art sowie die Güte und Fülle an Umweltinformationen haben starke Auswirkungen und können zu recht unterschiedlichen Prognoseleistungen führen (Pearson et al., 2006; Brotons et al., 2004)¹⁸. Insbesondere Guisan et al. (2007) zeigten, dass die Qualität der Modelle stark von den modellierten Arten selbst abhängt und weniger von den verwendeten Methoden.

Verschiedene Ringversuche¹⁹ haben sich mit der Frage nach der Rangierung der Methoden befasst (Elith et al., 2006; Guisan et al., 2007) und viele der publizierten Anwendungen von MAXENT beschäftigen sich mit Vergleichen zu etablierteren Methoden. Beispiele gibt es für verschiedenste Arten:

- Gibson et al. (2007): Anwendung von MAXENT im Vergleich zu GLM an seltenen Papageien in Australien. Das GLM wurde mit Pseudoabsenzen gerechnet, weil mit Präsenz/Absenz kein befriedigendes Modell erzeugt werden konnte. Bei 96 Präsenzen und Anwendung des Schwellenwertes

¹⁸ Die beiden Artikel stehen stellvertretend für weitere. Sie schliessen allerdings in ihren Vergleichen MAXENT und BRT nicht mit ein. Die Erkenntnisse sind aber in dieser Frage trotzdem gültig.

¹⁹ Verschiedene Methoden zur ökologischen Modellierung wurden von entsprechenden Spezialisten an denselben Datensätzen durchgeführt und miteinander verglichen.

'10%-Quantil'²⁰ lieferten beide Methoden gute aber nicht deckungsgleiche Prognosen. Mit dem GLM wurden AUC-Werte von 0.867 erreicht, mit MAXENT 0.889 (28 Präsenzen wurden als Testdaten verwendet).

- Yesson & Culham (2006): Vergleich von BIOCLIM und MAXENT unter Verwendung von 20 Cyclamen-Arten und dem Versuch, die Auswirkungen geänderter Klimabedingungen auf die einzelnen Arten vorher zu sagen. Die Studie wurde mit Umweltparametern in einem relativ groben 1/4 Grad-Raster für ein grosses Untersuchungsgebiet durchgeführt. MAXENT zeigte sich in der Modellierung der rezenten Verbreitung überlegen²¹. Die Toleranz für klimatische Änderungen wurde durch die beiden Methoden unterschiedlich beurteilt. Diese Beurteilungen können aber nicht bewertet werden, da sie nicht überprüfbar sind.
- Pearson et al. (2007): Vergleich MAXENT und GARP²² bei sehr wenig Präsenzen (<25) von seltenen Geckos. Zum Testen der Modelle wurde ein Jack-Knife Verfahren mit den Beobachtungen an zwei Schwellenwerten angewandt: der 'tiefste nachgewiesene Fund'²³ und der '10%-Quantil Schwellenwert'. Der Schwellenwert nach dem 'tiefsten nachgewiesenen Fund' erwies sich jedoch als nicht geeignet. Maxent zeigte sich vor allem bei sehr kleiner Anzahl Präsenzen als geeigneter.

Es gibt also trotz allem eine gewisse Rangordnung in der Güte der Methoden, die sich aus den erwähnten Arbeit ableiten lässt. Die neueren Methoden MAXENT und BRT schneiden gegenüber den etablierten wie GLM fast durchwegs besser ab.

2.8 Projektion in andere Umgebungen

Modelle können nicht nur für die Prognosen innerhalb eines Untersuchungsgebietes verwendet werden, sondern auch für die Prognose von Vorkommen

²⁰ Siehe Abschnitt 3.4.3.

²¹ Gemessen an der κ -Statistik.

²² GARP *Genetic Algorithm for Rule-Set Prediction* ist eine Methode aus der Gruppe des 'machine learning' (www.lifemapper.org/desktopgarp).

²³ Siehe Abschnitt 3.4.3.

in andere Gebiete. Ebenso können Prognosen in andere klimatische Bedingungen gemacht werden. Dazu werden normalerweise die Umweltvariablen des anderen Gebietes bzw. die neuen Ausprägungen als erklärende Variablen eingesetzt. Prognosen in andere Gebiete oder andere Klimate werden unter dem Begriff 'Projektion' zusammengefasst. Solche Projektionen stützen sich auf die im Untersuchungsgebiet beobachteten Umweltbedingungen. Sind die Bedingungen in der anderen Umgebung ausserhalb dieses Bereiches oder in der Kombination der Variablen neu, muss extrapoliert werden. Dies führt aber zu erheblichen Unsicherheiten.

Bei allen zeitlichen Prognosen, stellt sich das Problem der Güteabschätzung in besonderem Masse. Modelle, die in die Vergangenheit projizieren, können an geeigneten Indizien getestet werden. Garzón et al. (2006) verwendet dazu die Pollenanalyse²⁴.

Dormann (2007) erarbeitete einige Fehlerquellen für Extrapolationen (Auswahl): Die Annahme des Gleichgewichtes der Artverteilung mit den Umweltbedingungen ist oft nicht erfüllt. Die Frage, wie sich eine Art auch ohne Änderung des Klimas in der Verbreitung ändern würde, wird meist nicht betrachtet. Weiter sind die Umweltvariablen meist stark voneinander abhängig. Eine Änderung einer Variablen führt zu Reaktionen der anderen: Beispielsweise beeinflusst im Gebirge die Temperatur den Niederschlag und dieser wiederum die Vegetation. Geänderte Umweltbedingungen führen zu genetischen Antworten oder Verhaltensänderungen, welche mit den verfügbaren Parametern nicht vorhergesehen werden können. Zudem sind grundsätzlich die klimatischen Voraussagen mit einer sehr hohen Unsicherheit verbunden, insbesondere wenn sie von einer globalen Skala auf kleinräumige Gebiete angewendet werden.

Thuiller (2004) stellte verschiedene Modellansätze einander gegenüber und untersuchte deren Eignung für Voraussagen zur Klimaänderung. Er wies darauf hin, dass nicht zwingend das Modell mit der besten räumlichen Vorhersage der jetzigen Verbreitung zugleich die beste Prognose in die Zukunft darstellt.

²⁴ Für *T. biconicus* liegen jedoch bisher keine fossilen Funde vor. Untersuchungen in datierbaren Höhlensedimenten im Verbreitungsgebiet könnten dazu weiterhelfen, wenn auch Funde in genügender Zahl kaum zu erwarten sind.

Die mit den Umweltvariablen repräsentierte Skala (räumlicher Extent) und Rasterauflösung (räumlicher *Grain*) ist nicht unbedingt auch die entscheidende für die Verbreitung einer Art und mindert deshalb den Erklärungswert des Modells. Darum können Modelle in einem Untersuchungsgebiet sehr wohl gut angepasst sein, bei Projektionen in entferntere Gebiete oder andere Bedingungen aber versagen (Lausch, 2004; Dormann, 2007).

Guisan & Theurillat (2005) untersuchten mittels GLM, wie sich die Verbreitung verschiedener Pflanzenarten in den Schweizer Alpen unter veränderten Klimabedingungen entwickelt. Die Frage, wie die Pflanzen in höhere Stufen wandern, ist darin zentral: Eine Temperatur-Höhenbeziehung wurde entwickelt. Trotz den vielen Vorbehalten scheint diese Studie eine solide Vorhersage zu leisten.

Kapitel 3

Lösungsansatz

3.1 Modellieren natürlicher Vorgänge

Beim Modellieren natürlicher Vorgänge besteht das Interesse darin, aus Beobachtungen die Zusammenhänge und Gesetzmässigkeiten der Natur so zu verstehen, dass diese Vorgänge, in einem geeigneten Gesamtmodell verarbeitet, nachgebildet werden können. Da das 'Gesamtmodell Natur' in all seinen räumlichen, zeitlichen und kausalen Abhängigkeiten jedoch zu komplex ist, wird versucht, vorerst Teile davon zu modellieren.

Zum Einen gibt es die abiotischen Prozesse, welche sich auf die physikalischen Gesetzmässigkeiten stützen und durch diese annähernd vollständig beschrieben werden können. Dies sind beispielsweise Gebirgshebung, Gletscherfluss, Schneedecke, Wasserabfluss, Sedimenttransport, Sonneneinstrahlung usw. Erschwerend kommt aber hinzu, dass die Prozesse ungleich schnell ablaufen und daher viele labile Gleichgewichte und Fliessgleichgewichte bzw. Ungleichgewichtszustände entstehen. Insbesondere in alpinen Gebieten sind die Gradienten in allen Richtungen sehr stark und häufig auch kleinräumig wechselnd.

Zum Anderen sind am System Natur zahlreiche biologische Prozesse beteiligt. Diese zeichnen sich durch nochmals deutlich komplexere Abhängigkeiten von einander und von den abiotischen Prozessen aus. Zu den biologischen

Prozessen gehören beispielsweise biologische Interaktionen in all ihren Ausprägungen (Konkurrenz, Symbiose, Parasitismus etc.), Wachstum, Reproduktion und Ausbreitung. Auch hier ist die Prozessgeschwindigkeit von Belang: Reproduktions- und Ausbreitungsgeschwindigkeit von Arten, genetische Anpassung, Wachstum von Individuen usw. Viele Prozesse laufen zudem auf unterschiedlichen Skalen ab (Lausch, 2004).

Um diese unüberschaubare Komplexität zu bewältigen, kann mit statistischen Methoden versucht werden, Abhängigkeiten auf mess- oder interpolierbare Werte zu reduzieren. Mit der Anwendung drei solcher Methoden befasst sich diese Arbeit (siehe auch Abschnitt 2.7). Es wird versucht, aus Beobachtungen (d.h. Präsenz oder Absenz der Art) auf die Ansprüche der Art an die Umwelt und daraus auf das Verbreitungsgebiet dieser Art (hier *T. biconicus*) zu schließen.

3.2 Die ökologische Nische

Als Eingangsdaten in die Modellierung werden die Beobachtungen mit Informationen zu geografischer Lage und Vorkommen¹ benötigt. Eine Suchstrategie gibt vor, wie diese zu erheben sind. Dabei werden Lokalitäten auf das Vorkommen der Art untersucht und im geografischen Raum \mathbf{G} (Abbildung 3.1, links) als Karte dargestellt. Ein Untersuchungsgebiet kann jedoch gerade in gebirgigem Raum nie vollständig abgesucht werden. Die Informationen beschränken sich also auf den abgesuchten Weg (\mathbf{d}). Die Prognose soll hingegen über das gesamte Untersuchungsgebiet Auskunft geben.²

Der Umwelt-Raum \mathbf{U} (Abb. 3.1, rechts) kann als n -dimensionaler Raum beschrieben werden, entsprechend der Anzahl zu untersuchender Umweltvariablen³, in dem jeder Punkt eine Kombination von Umweltwerten an einer Fund- oder Nicht-Fundstelle darstellt. Die Modellierung versucht, aus

¹ Es werden nur die Informationen zu Präsenz oder Absenz verwendet. Die Individuendichte wird hier für die Modellierung nicht berücksichtigt.

² Die Ausführungen in diesem Kapitel folgen grob Pearson (2007), S.8-11 und 13 ff., sowie den Abbildungen dazu.

³ Das in Abb. 3.1 gezeigte Beispiel besteht - stellvertretend für viele - nur aus den zwei Umweltvariablen E1 und E2.

den Funden im Umwelt-Raum eine Umhüllende⁴ zu bilden. Absenzen werden möglichst ausgeschlossen. Die verschiedenen Modellansätze unterscheiden sich vor allem in der Art, wie die Umhüllende gelegt wird. Diese Umhüllende beschreibt im Idealfall die *fundamentale Nische* (f) einer Art. Gemäss der Definition von Hutchinson (1957) entspricht dies jenem Satz von Umweltbedingungen, in welchem die Art längerfristig überleben und gedeihen kann. Allerdings wird die fundamentale Nische nur in der Masse beschrieben, wie sie durch Funde an Grenzbedingungen nachgewiesen ist. Übersetzt in den geografischen Raum, entspricht die fundamentale Nische der *potentiellen Verbreitung*. Die *effektive Verbreitung* oder die *besiedelte Nische* kann sich auf einen Teil der fundamentalen Nische beschränken (e): Beispielsweise können höher gelegene Gebiete, welche aufgrund fehlender Ausbreitungsmöglichkeiten nach oben bisher nicht erreicht wurden, trotzdem zur fundamentalen Nische gehören. So liefern Gebiete wie (c) Absenzdaten, obwohl die Bedingungen der fundamentalen Nische dort erfüllt sind. Die Begründungen dazu sind jeweils aus Sicht der zu beschreibenden Umweltvariable zu suchen.

Angenommen, die Ausbreitung einer Art ist von einem Entstehungszentrum (auch 'Hot Spot' genannt) ausgegangen, ist es eine Frage der Zeit, der Ausbreitungskapazität und allfälliger Barrieren, ob alle Gebiete der fundamentalen Nische besiedelt werden. So konnte das Gebiet bei (a) innerhalb der Besiedlungsgeschichte von einer genügenden Anzahl Individuen erreicht werden und sich eine stabile Population erhalten. Hingegen ist das Gebiet bei (b) noch nicht oder nicht mehr besiedelt. Die entsprechende Absenz-Beobachtung (k) liegt deswegen innerhalb der Umhüllenden. Die Talententwicklung kann ein Grund für eine solche Barriere sein (physische Barriere)⁵.

Die tatsächlich besiedelten Nischen im Umwelt-Raum (e) entsprechen der effektiven Verteilung im geografischen Raum (h). Hingegen können auch Funde an Orten gemacht werden, an denen die Art nicht längerfristig überlebensfähig ist. Solche *verschleppten Funde* können entweder durch das Wandern von Exemplaren von einem Habitat zu einem anderen begründet werden (aktive Wanderung) oder via passiven Transport durch z.B. gravitative Prozesse

⁴ Die Umhüllende bezeichnet hier die kleinste Fläche, die alle Punkte einer Punkteschar einschliesst.

⁵ Weitere Möglichkeiten sind Brand-Katastrophen oder Krankheiten.

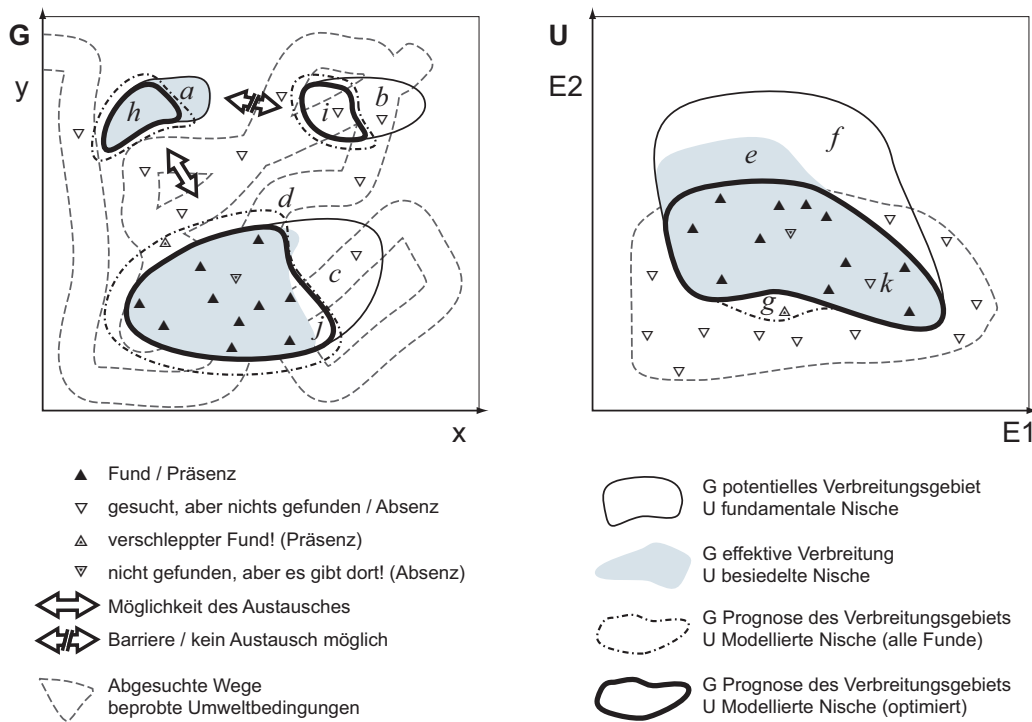


Abbildung 3.1: Die Beobachtungen im geografischen Raum G (links) können an Hand der Umweltbedingungen am Fundort in den Umwelt-Raum U (rechts) übersetzt werden. Die Umhüllende der Präsenzpunkte im Umwelt-Raum stellt geeignete Umweltbedingungen (Kombinationen von E1 und E2) für die Art dar.

verursacht werden⁶. Sie geben aber einen scheinbar grösseren Bereich für die fundamentale Nische vor (g), was zu einer zu optimistischen Prognose führt⁷.

Es kann aber auch vorkommen, dass die Art an einer Lokalität nicht beobachtet wird, obwohl sie dort vorkommt⁸. Oft wird es besiedelte Gebiete geben, die nicht besucht wurden und somit nicht entdeckt werden (h). Das Finden solcher Gebiete ist eine der Hauptaufgaben und Stärken der Modellierung.

Problematisch für den Modellierungsprozess sind jedoch vor allem die

⁶ Das Phänomen wird auch als *source-sink dynamics* besprochen, wobei *source* das Auswanderungsgebiet meint und *sink* den Fundort mit den für den langfristigen Erhalt der Art nicht geeigneten Umweltbedingungen (Pulliam, 2000; Pearson, 2007).

⁷ Siehe dazu die Ausführungen zum Schwellenwert 'tiefster nachgewiesener Wert' im Unterabschnitt 3.4.3.

⁸ Bei schwer zu findenden Arten wie der *T. biconicus*, ist dieser Fall durchaus wahrscheinlich.

verschleppten Funde, welche die Umhüllende fälschlicherweise erweitern (g). Eine falsche Absenzbestimmung hingegen wird bei genügender Präsenzzahl nur einen marginalen Einfluss haben.

Durch die Rückprojektion der im Umweltraum modellierten Nische auf den geografischen Raum, ergibt sich eine Verbreitungsprognose. Die nicht besiedelten Teile der fundamentalen Nische (b und c) werden dabei nicht als mögliches Verbreitungsgebiet ausgeschieden. Die Abgrenzung kann aber wie bei (j) gelegentlich für das Modell schwierig sein. Gebiete innerhalb der effektiven Verbreitung, welche Umweltbedingungen aufweisen, die nicht in den Präsenzdaten vertreten sind (a), werden vom Modell nicht erkannt. Hingegen können Hypothesen über das Fehlen der Art bei (i) Hinweise zur Besiedlungsgeschichte liefern.

3.3 Eingangsdaten

Die Eingangsdaten zur Modellierung von Artverteilungen bestehen zum Einen aus Lokalitäten mit Beobachtungen zum Vorkommen der Art in einem Gebiet, zum Andern aus einem Satz aus Umweltvariablen zu diesem Gebiet.

3.3.1 Beobachtungen zum Vorkommen der Art

Die Angaben zum Vorkommen der zu untersuchenden Art stammen aus Beobachtungen der Felderhebungen von 2004 bis 2007. Bevor sie zur Modellierung herangezogen werden, müssen sie einer Datenkritik unterzogen werden (siehe Abschnitt 4.2.1).

3.3.2 Wahl der Umweltvariablen

Unter Umweltvariablen versteht man quantitative oder qualitative Messungen von Umweltbedingungen, welche kontinuierlich über das Untersuchungsgebiet vorhanden sind⁹ oder durch geeignete Interpolationsverfahren ergänzt

⁹ Für jede Rasterzelle gibt es einen Wert.

werden können. Nach Hengl (2007) stellt deren Verfügbarkeit und Qualität ein zentraler Input in die Modellierung dar. Dudík et al. (2004) ergänzten, dass die Umweltvariablen aus verschiedenen Quellen stammen können:

- a) Aus einem Höhenmodell abgeleitete Umweltvariablen, wie Höhe, Hangneigung, Hangorientierung, topografische Indices.
- b) Aus meteorologisch-klimatischen Werten, die basierend auf Messungen bzw. Messreihen von Meteostationen mit Hilfe des Höhenmodells auf das Untersuchungsgebiet interpoliert werden, wie Niederschlag, Temperatur, Evapotranspiration.
- c) Aus Felderhebungen basierende Umweltvariablen, wie Bodenbedeckung, Geologie, Vegetation.
- d) Aus Fernerkundungsquellen (Satelliten- und Luftbilder) aufbereitete Umweltvariablen.

Mit steigendem Verständnis der physikalischen Zusammenhänge können zunehmend klimatische Umweltvariablen direkt aus Höhenmodellen gerechnet werden. Sie können in Zukunft mehr und mehr die unter b) interpolierten Werte ersetzen, welche mit stark wechselnden, meist grösseren Fehlern behaftet sind.

In Abschnitt 2.5 wurden die beiden (Extrem-)Ansätze 'mechanistisch' und 'korrelativ' eingeführt.

Mechanistischer Ansatz

Im mechanistischen Ansatz werden Umweltvariablen mit ökologisch direkt interpretierbarer Wirkung gewählt (direkte Wirkungsvariablen). Daraus sollen ein Prozessverständnis¹⁰ abgeleitet und die limitierenden Umweltvariablen herausgefunden werden. Ausgangslage sind subjektiv empirisch aufgestellte Hypothesen aus der Feldbeobachtung oder dem Prozessverständnis des 'Experten'. Da verschiedene Einflüsse jedoch nicht direkt aus Messwerten oder

¹⁰ Das heisst das Erkennen statistisch nachweisbarer Zusammenhänge.

Berechnungen abgeleitet werden können und die Einflüsse der Realwelt nie vollständig verstanden und nachgebildet werden können, bleibt ein solches Modell immer beschränkt. Das geforderte Mass an Verstehen der Interaktion der Art mit der Umwelt kann nur in seltenen Fällen gefunden werden. Das Aufstellen und Prüfen von Hypothesen kann für konkrete Fragen jedoch ein gutes Vorgehen sein. Solche Modelle sind - mit Blick auf die in Abschnitt 3.1 gemachten Überlegungen - in der Regel sehr schwierig und nur mit grossem Aufwand zu erstellen: Abbildung 3.2 zeigt diesen Ansatz.

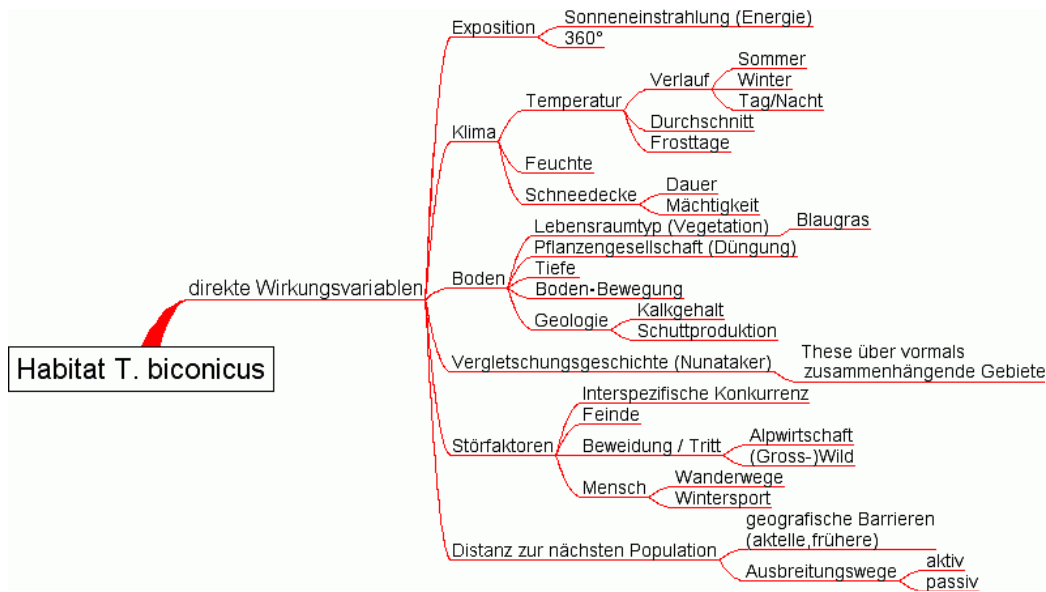


Abbildung 3.2: Im mechanistischen Ansatz werden Hypothesen zur Abhängigkeit des Vorkommens von den Umweltbedingungen aufgestellt und durch statistische Methoden anhand von Umweltvariablen zu belegen versucht.

Im folgenden wird gezeigt und begründet, warum der mechanistische Ansatz auf die Habitatmodellierung von *T. biconicus* nicht angewendet werden kann:

Die **Höhenlage** und die **Expositionsangaben** können mit ausreichender Genauigkeit aus einem Höhenmodell hergeleitet werden. Hingegen sind die **Klimadaten** auf Grund der Interpolation bereits mit grösseren Fehlern behaftet. Zudem spielt vermutlich das Mikroklima unter den Steinen für *T. biconicus* eine wichtigere Rolle als das modellierbare Mesoklima.

Die **Boden**parameter können einerseits durch die Bodenbedeckung der amtlichen Vermessung und andererseits durch die Geologie (Gesteinsart, Lage, Verwitterungsverhalten, chemische Zusammensetzung, ...) abgedeckt werden. Die Bodenbedeckung ist zwar sehr präzise ausgeschieden, hat aber auf Grund der von der Fragestellung differierenden Erfassungsoptik eine recht grobe Unterteilung und Abgrenzung der Klassen. Die Informationen zur Geologie stehen leider nicht für das ganze Untersuchungsgebiet und in genügender Tiefe zur Verfügung.

Eine Hypothese besagt, dass die einzelnen rezenten Populationen ein Überbleibsel einer einst zusammenhängenden, grösseren Population darstellen. Die einzelnen Populationen haben nach dieser Hypothese die quartären **Vergletscherungen** auf Nunatakern überdauert, wurden als Folge der Talentwicklung jedoch getrennt. Die Quartärgeschichte und eine genaue Abschätzung des Gletscherhöchststandes sind im Untersuchungsgebiet jedoch noch zu wenig bekannt. Zudem sind Beobachtungen zu Gletscherschliffgrenzen oft schwierig. Die Talentwicklung des Engelberger- und Sarneraatales ist noch über weite Strecken unbekannt. Die Landschaft dürfte aber auch im Einzugsgebiet der Gletscher ziemlich stark umgestaltet worden sein.

Für das Verständnis von **Störfaktoren** auf die Art reichen die Beobachtungen und Kenntnisse noch nicht aus. Die Bestossung der Alpen wird nicht systematisch und mit genügender Genauigkeit erfasst - Wildbewegungen schon gar nicht. Wanderwege sind im Verbreitungsgebiet nicht allzu zahlreich und der Wintersport ist in den meisten Gebieten (noch) nicht stark ausgeprägt.

Die **Distanz zur nächsten Population** ist abhängig von der Ausbreitungsgeschwindigkeit und -kapazität der Art. Bislang ist aber noch wenig zu diesen beiden Faktoren bekannt und daher bleiben diese Vorgänge vorerst nicht quantifizierbar¹¹.

Der mechanistische Ansatz scheitert hier also daran, dass (a) viele Größen nicht oder nur schwer bzw. mit grossem Aufwand messbar sind, (b) das

¹¹ Erste Erkenntnisse zur Lebensweise von *T. biconicus* sind aus den weiteren Studien im Zusammenhang mit dem Projekt zur Erforschung der *T. biconicus* von Baggenstos (in prep.) zu erwarten.

Modell recht komplex wird und daher schwer zu überschauen und aufwändig zu berechnen ist sowie (c) die Abhängigkeiten komplex sind. Zudem ist (d) die Liste der Einflussgrößen sicher nicht vollständig. Die vorliegende Arbeit befasst sich deshalb mit der Modellierung basierend auf dem korrelativen Ansatz.

Korrelativer Ansatz

Bei der Anwendung des korrelativen Ansatzes sucht man zuerst möglichst viele Umweltvariablen und prüft diese dann auf ihren Effekt auf die Wirklichkeit des Vorkommens von *T. biconicus* (z.B. in MAXENT). Zu Grunde liegt die Annahme, dass aus dem Vorkommen (bzw. Nicht-Vorkommen) einer Art auf deren Anforderungen an die Umwelt geschlossen werden kann. Die wichtigsten Umweltvariablen werden durch den angewendeten Algorithmus gefunden und parametrisiert. Somit kann das Modell auf eine gute Voraussage optimiert werden. Verschiedene Einflüsse werden in mess- oder berechenbare Größen zusammengefasst (substituiert).

Das korrelative Vorgehen führt nicht zwingend zu biologisch interpretierbaren Resultaten. Dennoch ist es erwünscht, solche daraus ableiten zu können. Damit dieses Ziel erreicht wird, ist ein möglichst einfaches Modell mit ökologisch interpretierbaren Variablen anzustreben. Weisen Variablen eine hohe Korrelation der Werte auf, erklären sie einen ähnlichen Sachverhalt und bringen somit keine zusätzliche Information in das Modell. Es ist also anzustreben, solche Variablen auf eine zusammenzufassen - das Kriterium für die Auswahl ist die erwähnte einfach ökologisch interpretierbare Wirkung der Variable auf das Vorkommen der Art.

3.4 Methodik der Modellierung

3.4.1 Wahl der Methode

Die Suche nach der besten Methode wurde in Abschnitt 2.7 beschrieben. Der Fokus dieser Arbeit soll mehr darauf liegen, zu zeigen, ob die Methoden

für wenig mobile Arten und Gebiete mit starkem Relief geeignet sind. Dazu sollen die beiden neueren Ansätze BRT und MAXENT mit einer klassischen Methode GLM verglichen werden.

3.4.2 Der AUC als Gütemass

Zum Vergleich der Modelle bzw. deren Prognoseleistungen werden Gütemasse benötigt. Zur Bewertung gibt es eine Vielzahl an Gütemassen, jedoch kann kein Gütemass alle Aspekte eines Modells wiedergeben (Reineking & Schröder, 2004). Darum verwenden werden hier verschiedene Gütemasse, welche zusammen ein verlässlicheres Gesamtbild ergeben.

Ein in der Signalverarbeitung entwickeltes Verfahren, um die Güte von Modellen bei der Unterscheidung von tatsächlichen Signalen vom zufälligen Rauschen zu unterscheiden, ist die *Receiver Operating Characteristics* ROC-Kurve (Abbildung 3.3). In einem Diagramm wird die Rate der falschen Prognosen für das Vorkommen (*'1-specificity'*) gegen die Rate der richtigen Prognosen für Vorkommen (*'sensitivity'*) aufgetragen. Konkret wird für jeden Schwellenwert zwischen 0 und 1 eine *confusion Matrix* oder Wahrheitsmatrix erstellt und die Erfolgsrate als Punkt der ROC-Kurve dargestellt. Bei einem Zufallsgenerator-Modell würde die Kurve die Diagonale von 0/0 zu 1/1 zeigen (grüne Linie in Abb. 3.3). Modelle, die besser als zufällig voraussagen, zeigen eine Kurve, die oberhalb dieser Diagonalen liegen.

Als Gütemass wird die Fläche unter der Kurve gemessen und als AUC-Wert (*Area Under the Curve*) bezeichnet. Je besser ein Prognosemodell die Daten trifft, desto näher liegt dieser Wert bei 1. Der AUC-Wert dient der Abschätzung der Prognosegenauigkeit und ist somit ein Kriterium für die Wahl des besten Modells (Fielding & Bell, 1997).

Kritikpunkte

Der Hauptkritikpunkt am AUC ist seine Abhängigkeit von der Grösse des Untersuchungsgebietes. Da bei Vergrößerung des Untersuchungsgebietes in der Regel das Flächenverhältnis zu Gunsten von ungeeigneten Flächen wächst,

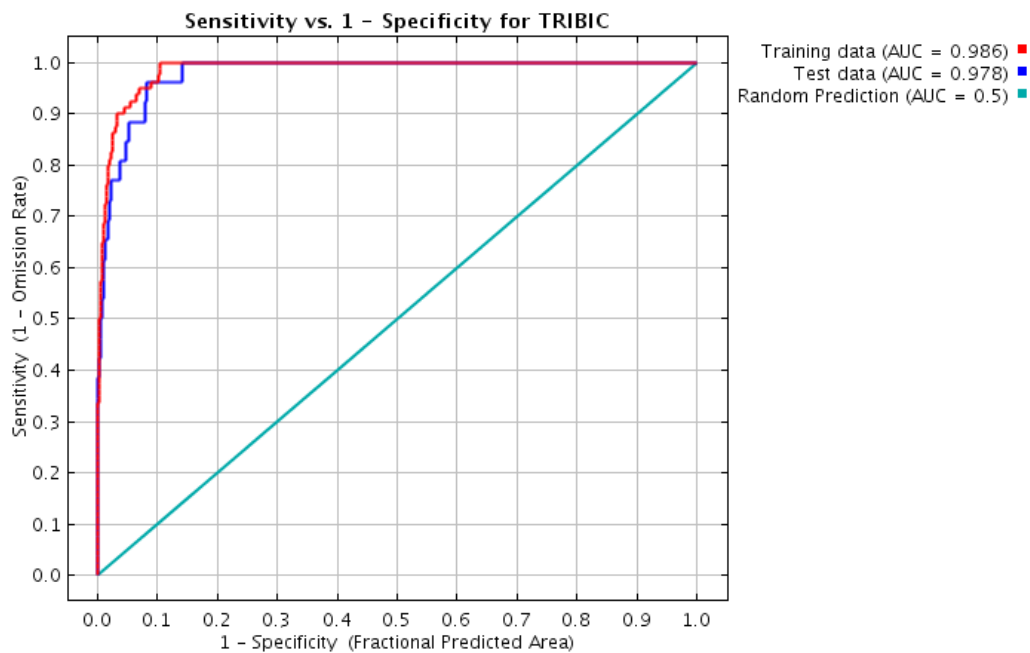


Abbildung 3.3: Im ROC-Plot wird die Prognoseleistung eines Modells dargestellt: Liegt die Kurve auf der Diagonalen (grün), ist das Modell nicht besser als eine zufällige Schätzung; führt die ROC-Kurve nahe an der linken oberen Ecke vorbei, ist die Übereinstimmung der vorhergesagten und der beobachteten Werte nahezu perfekt. Der AUC-Wert gibt die Fläche unter der ROC-Kurve an: Je näher er gegen 1 geht, desto besser ist die Prognoseleistung. Zur Erstellung des Modells wurden hier $3/4$ der Beobachtungen als Trainingsdaten benutzt (rote Kurve); das andere Viertel diente zum Testen des Modells (blaue Kurve): In einem robusten Modell liegt die Kurve der Testdaten sehr nahe bei der Kurve der Trainingsdaten.

nimmt die Modell-Güte scheinbar zu (Lobo et al., 2007). Die Grenze des Untersuchungsgebietes wurde darum für alle Methoden identisch und die Ausdehnung nicht allzuweit über das erwartete Verbreitungsgebiet hinaus gewählt (siehe Abschnitt 4.1).

Ein weiterer Schwachpunkt des AUC-Wertes ist, dass er keine Angaben über die räumliche Verteilung der Fehler (Residuen) gibt. Dieser Schwachpunkt kann durch die Untersuchung der Residuen nach räumlicher Autokorrelation behoben werden (siehe Abschnitt 4.5.4). Des Weiteren ist beim AUC zu beachten, dass Fehler in der Absenzvorhersage und solche in der Präsenzvorsage gleich gewichtet werden. Dies ist bei der Verwendung von Pseudoabsenzen, welche hier 100-mal häufiger sind, problematisch und muss berücksichtigt werden.

Trotz diesen Kritikpunkten scheint der AUC im Moment ein geeignetes Mass zur Abschätzung der Modellgüte zu sein und wird häufig eingesetzt.

3.4.3 Schwellenwerte

Die Antwort auf die in der Einführung (Abschnitt 1) gestellten Frage a) nach dem Vorkommen, ist nicht eine Habitatsignungs- oder Wahrscheinlichkeitskarte mit kontinuierlichen Werten von 0 bis 1, sondern eine binäre Aussage über das Vorkommen oder Nicht-Vorkommen. Daraus folgt die Frage nach dem Schwellenwert, an welchem die Habitatsignung gerade so gross ist, dass die Art auftreten kann. Der beste Schwellenwert minimiert die Fehlprognosen für Präsenzen wie für Absenzen¹². Für weitere Referenzen und eine ausführliche Diskussion sei auf Pearson (2007),¹³ verwiesen, für eine kritische Betrachtung auf Lobo et al. (2007).

¹² Für die theoretischen und mathematischen Hintergründe dazu sei auf Fielding & Bell (1997) verwiesen.

¹³ Table 4, Seite 35: Auswahl publizierter Methoden zum Setzen von Schwellenwerten für Vorkommen.

Feste Schwellenwerte

Ein erster Ansatz ist, den Schwellenwert *a priori* vorzugeben. Das heisst, ihn unabhängig von den modellierten Werten festzusetzen. Oft verwendete Werte sind 0.01, 0.05 oder 0.10 (bei einer Werteverteilung von 0 bis 1). Da der Schwellenwert jedoch artspezifisch und von der verwendeten Methode abhängig ist, sind so kaum verlässliche oder biologisch sinnvolle Aussagen zu erwarten (Liu et al., 2005).

Tiefster nachgewiesener Fund

Besser ist es, den tiefsten prognostizierten Wert der Fundstellen zu verwenden ('*lowest predicted threshold*' LPT oder '*Minimum training presence*'). Er kann so verstanden werden, dass damit alle Gebiete eingeschlossen werden, an denen die Bedingungen gerade noch gut genug sind um ein Vorkommen zu ermöglichen. Somit kann er als der 'konservativste' der (abhängigen) Schwellenwerte bezeichnet werden. Die in Abschnitt 3.2 als 'verschleppte Funde' bezeichneten Beobachtungen können aber nicht durch die Umweltbedingungen am Fundort erklärt werden. Sie stellen also Populationsteile dar, die entweder zu einer nicht längere Zeit überlebensfähigen *Sink*-Population gehören oder sich nur vorübergehend an diesem Ort aufhalten. In einem solchen Fall wird mit diesem Schwellenwert der modellierte Lebensraum gegenüber dem tatsächlichen zu gross vorhergesagt. Dieser Ansatz ist auch sehr anfällig auf Fehler in der Lageangabe der Fundstellen – insbesondere bei starken Gradienten. Daher stellt dieser Schwellenwert kein robustes Mass dar.

Quantil

Eine verbesserte Möglichkeit ist, nur einen bestimmten Prozentsatz der Fundstellen innerhalb der prognostizierten Fläche zu haben und die restlichen als Ausreisser (gegen unten) anzusehen. Das heisst, man gibt eine minimale Sensitivität vor. Diese Konstruktion weist Parallelen zum 'Konfidenzintervall' - einem verbreiteten statistisches Verfahren - auf. Üblich sind 99%, 95% oder 90% der Funde innerhalb der Vorhersagefläche zu haben. Dieser Ansatz hat

den Vorteil, dass nicht ein einzelner (Ausreisser-)Wert den Schwellenwert definiert. Er nimmt jedoch einen willkürlich festgelegten Anteil an Ausreißern an (siehe auch die Ausführungen zur ökologischen Nische in Abschnitt 3.2).

Im Ausgabedokument von MAXENT erscheint nur der 90%-Wert unter der Bezeichnung '*10 percentile training presence*'.

Maximales Kappa

Das von Cohen (1960) eingeführte κ -Kriterium ist ein Mass für die Treffgenauigkeit eines Schätzverfahrens oder eines Prognosemodells, das auch die Zufallstreffer berücksichtigt; es sagt also aus, um wie viel besser als zufällig es ist (in einem Bereich zwischen 0 und 1). Den 'MaxKappa'-Schwellenwert erhält man, indem das Maximum des κ gesucht wird. Da es aber alle Werte aus der Konfusionsmatrix benötigt, ist seine Anwendung eigentlich auf die Präsenz/Absenz-Modelle beschränkt. Das Anwenden auf die Pseudoabsenzen führt zu sehr hohen Schwellenwerten, welche entsprechend sehr kleine Flächen ergeben.

Sensitivität gleich Spezifität

Schlussendlich ist die Wahl des Schwellenwertes auch eine Frage des Fokus einer Untersuchung: für den umfassenden Schutz eines Habitats wird man eher einen tiefen Wert wählen – für eine sichere Suchstrategie eher einen hohen Schwellenwert. Daher kommt auch der Vorwurf der Subjektivität von Schwellenwerten.

3.5 Unsicherheitskarten

Bei mehreren Modellen oder bei mehreren Testläufen des gleichen Modells interessieren neben zusammenfassenden Gütemassen auch lokalisierbare Informationen. Zellen mit hoher Konstanz über alle zu vergleichenden Modelle, weisen eine entsprechend hohe Prognosesicherheit auf. Umgekehrt sind Zellen mit stark wechselnder Bewertung, mit grösserer Unsicherheit behaftet

Im Abschnitt 3.4.2 über die Verwendung des AUC wurde die Kritik von Lobo et al. (2007) an dessen Verwendung angeführt. Lobo et al. schlugen als Alternative vor, einen Schwellenwert zu verwenden, der sich aus dem 'nordwestlichsten' Punkt der ROC ableitet (Pfeil in Abbildung 3.4). Damit wird eine Egalität der Fehlerrate für Präsenzen wie auch für Absenzen erreicht. Dieser Schwellenwert entspricht dem *Equal training sensitivity and specificity* aus MAXENT.

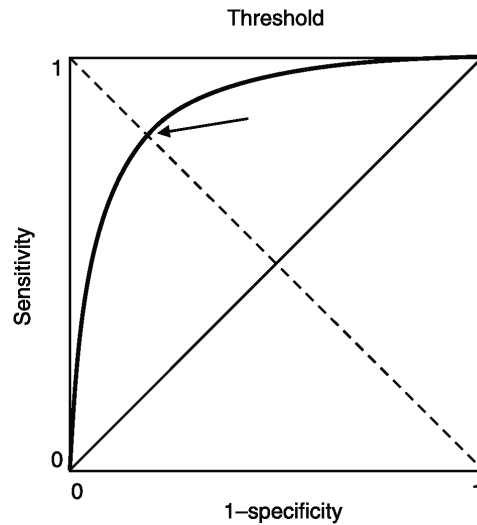


Abbildung 3.4: Die Konstruktion des Schwellenwertes nach einem Vorschlag von Lobo et al. (2007), dargestellt in einem ROC-Plot.

und sollten für die Verbesserung der Modelle bei künftigen Feldbegehungen prioritär besucht werden.

3.5.1 Unsicherheitskarten aus den 'rohen' Prognosen: Standardabweichungskarten

Die Variabilität von Prognosen verschiedener Modelle kann durch die Differenz vom maximal zum minimal erreichten Wert dargestellt werden. Dieses Mass ist jedoch recht anfällig auf 'missglückte' Durchläufe mit extremen Werten. Als Alternative bietet sich die Standardabweichung an. Sie reagiert stabiler auf Ausreisser, wenn auch die absoluten Werte daraus schwieriger zu interpretieren sind. Die qualitative Aussage bleibt jedoch gleich. Für die Darstellung der Standardabweichung soll ein Bereich von 0 bis zum maximal möglichen Wert der Standardabweichung gewählt werden¹⁴. Solche Karten werden im Folgenden 'Standardabweichungskarten' genannt.

Obschon auch hier die Vorbehalte für den Vergleich von rohen Prognosen

¹⁴ Der maximale Wert der Standardabweichung ist abhängig von der Anzahl zu vergleichenden Prognosekarten. Für 3 und 4: 0.5774; für 5 und 6: 0.5477; ...; für 9 und 10: 0.5270; ...

gelten, ist der Vergleich innerhalb der selben Methode noch einigermaßen zulässig, hingegen darf nicht mit rohen Prognosen aus anderen Methoden verglichen werden.

3.5.2 Schwellenwertkarten

Um die Prognosekarten verschiedener Methoden miteinander zu vergleichen, werden die rohen Prognosen durch einen Schwellenwert in binäre Präsenz/Absenz-Karten überführt. Zwei Darstellungsvarianten werden gewählt: Die Einigkeitskarte, die aussagt, wie hoch der Anteil der Modelle ist, die sich über das Vorkommen der Art einig sind, und die Unsicherheitskarte, welche aussagt, wo und wie stark sich die Modelle widersprechen.

Die **Einigkeitskarte** zu einem Schwellenwert (Formel 3.1) rechnet den Durchschnitt aller Prognosen und zeigt somit den höchsten Wert in jenen Gebieten mit einhelligem Vorkommen an. Bei vier verglichenen Prognosen können folgende Fälle auftreten: 'alle Prognosen sind sich über das Vorkommen einig' (100%), 'eine Prognose sagt kein Vorkommen voraus' (75%), 'zwei Prognosen sind gegensätzlich zu den anderen zwei' (50%) und 'nur eine Prognose sagt ein Vorkommen voraus' (25%)¹⁵.

$$Einigkeitskarte_{th} = \frac{\sum_{i=1}^n Prognosekarte_n(th)}{n} \quad (3.1)$$

Der Wert n ist die Anzahl zu vergleichender Prognosen
und th der angewendete Schwellenwert ist.

Es besteht die Gefahr, dass schlechte Prognosen ungebührlich stark zum Ergebnis beitragen. Um dieser Gefahr zu begegnen, kann zusätzlich eine Gewichtung der einzelnen Prognosen durch Multiplikation mit ihrem AUC-Wert - 0.5, d.h. alles, was besser als zufällig ist, eingeführt werden. Die Formel 3.2 zeigt die entsprechende Adaption. Sind die AUC-Werte der Prognosemodelle jedoch sehr nahe beieinander und ist die Anzahl verglichener Modelle klein, wird das Resultat kaum von einem nicht gewichteten abweichen.

¹⁵ Siehe Abbildung 5.12 auf Seite 81.

$$\text{Einigkeitskarte}_{th} = \frac{\sum_{i=1}^n \text{Prognosekarte}_n(th) \cdot (AUC_n - 0.5)}{\sum_{i=1}^n (AUC_n - 0.5)} \quad (3.2)$$

Die **Unsicherheitskarte** (siehe Abbildung 5.11 Seite 81) zeigt die Gebiete mit abweichenden Prognosen an. Bei vier Prognosen gibt es drei mögliche Fälle: Die Unsicherheit ist am geringsten, wenn alle Prognosen das Gleiche voraussagen (Präsenz oder Absenz); eine mittlere Unsicherheit liegt vor, wenn eine Prognose von den drei anderen abweicht; wenn je zwei Prognosen identisch sind, ist die Unsicherheit am grössten.

Formel 3.3 wurde verwendet, um für jede Rasterzelle eine Prognose-Unsicherheit anzugeben. Zuerst wird die durchschnittliche Prognose berechnet: Ist diese 0 oder 1, ist die Unsicherheit minimal; beträgt sie 0.5, ist sie maximal. Durch die Subtraktion von 0.5 wird der Wertebereich "nach unten verschoben", d.h. er wird symmetrisch zu Null gelegt (Bereich neu -0.5 bis +0.5). Anschliessend werden die negativen Teile durch den Betrag $\text{abs}(x)$ umgeklappt (Wertebereich: 0 bis +0.5). Die Skala wird anschliessend durch Multiplikation mit 2 wieder auf den Bereich 0 bis 1 gespreizt und so die Aussagen 0 (maximale Unsicherheit) bis 1 (grosse Sicherheit, alle Prognosen sagen dasselbe) erreicht. Zur Darstellung der Unsicherheit werden die Werte getauscht (1-x), so dass schliesslich 1 die grösste Unsicherheit darstellt.

$$\text{Unsicherheitskarte}_{th} = 1 - \left(\left| \frac{\sum_{i=1}^n \text{Prognosekarte}_n(th)}{n} - 0.5 \right| \cdot 2 \right) \quad (3.3)$$

Die Korrektur mit der Gewichtung mittels des AUC-Wertes ist analog zur Formel 3.2 zu erstellen.

Welche der zwei Darstellungsarten bevorzugt wird, hängt von deren Verwendung ab: Die Unsicherheitskarte hilft beim Identifizieren von Gebieten mit Schwierigkeiten für die Modellierung - die Einigkeitskarte bei der Einschätzung der Vorkommenswahrscheinlichkeit ("sichere Fundgebiete"). Die Genauigkeit der Karten darf jedoch nicht überschätzt werden: Sinnvolle Aussagen sind in eher grösseren Gebieten mit einheitlicher Einstufung zu erwarten und weniger aus der Interpretation von einzelnen Rasterzellen.

3.5.3 Methodenübergreifende Unsicherheitskarten

Die Schwellenwerte 'tiefster nachgewiesener Wert' und 'Maximales Kappa' sind anfällig auf Ausreisser, beziehungsweise sehr restriktiv. Für den abschliessenden Quervergleich über die Methodengrenzen hinweg sind deshalb die Unsicherheits- und die Einigkeitskarten für die Schwellenwerte 'Sensitivität gleich Spezifität' und das '10%-Quantil' zu erstellen.

3.6 Projektion

Die Projektion¹⁶ in andere Gebiete als das Untersuchungsgebiet liefert die Hinweise auf weitere, bisher unbekanntes Vorkommen. Dabei ist zu beachten, dass die Genauigkeit der Extrapolation tendenziell schwächer wird, je weiter entfernt die Werte der verwendeten Umweltbedingungen von den im Untersuchungsgebiet vorhandenen sind. MAXENT liefert dazu ein Bild¹⁷, welches, auf Grund von Überschreitungen des aus dem Untersuchungsgebiet bekannten Wertebereiches, Auskunft über die Stärke der Unsicherheiten gibt. Zudem bildet die zunehmende räumliche Distanz eine zusätzliche Fehlerquelle: Einflüsse, welche im relativ begrenzten Untersuchungsgebiet gleichmässig wirken, können über grössere Distanzen wichtiger werden, wie beispielsweise die durch Gebirgsrücken beeinflusste Grosswetterlage.

3.7 Versuchsanordnung

Die etablierte Methode der GLM wurde den neueren *Maximum Entropy Model* (MAXENT) und BRT gegenüber gestellt. Dazu wurden mit diesen Ansätzen an Hand der *T. biconicus*-Funde die Modelle berechnet, das Optimierungspotential ausgeschöpft und auf das Untersuchungsgebiet angewandt. Die *Präsenz/Absenz*-Methoden GLM und BRT wurden zuerst wie üblich

¹⁶ Der Begriff Projektion umfasst die Extrapolation des Modelles in andere Gebiete und in andere klimatische Bedingungen.

¹⁷ In MAXENT unter dem Begriff *clamping* zu finden. Die Darstellung wurde von Version 3.1 zu 3.2.1 geändert.

durchgeführt. Anschliessend wurden sie zum Vergleich mit Maxent nochmals mit extern erzeugten Pseudoabsenzen durchgeführt und geprüft. Hierzu wurden drei zufällige Verteilungen mit 10'000 Punkten generiert. Der Ersatz von Absenzen durch Pseudoabsenzen wurde auch in Guisan et al. (2007) angewendet.

Zum Vergleich der Resultate wurden verschiedene Tests angewendet. Weil kein unabhängiger Testdatensatz zur Verfügung stand, wurden Versuche zur Stabilität der Prognose mit zufälligen Teildatensätzen angestellt und die Resultate untereinander verglichen (bei Maxent mit 10 Durchläufen mit jeweils 75% der Funde als Trainingsdaten und 25% als Testdaten, bei den R-basierten Methoden mit 10-facher Kreuzvalidierung). Je geringer der Einfluss der Fundanzahl auf das Testergebnis und je ähnlicher die Tests einer Methode zueinander sind, desto stabiler und damit vertrauenswürdiger ist das Modell.

Durch die Überlagerung der Prognosen wurden jene Gebiete als unsicher in der Prognose identifiziert, wo sich die Vorhersagen am stärksten widersprechen. Zwei Darstellungsvarianten dieser Unsicherheit wurden in Kapitel 3.5 vorgestellt.

MAXENT bietet eine einfache Möglichkeit, in andere Gebiete zu extrapolieren. Mit verschiedenen Umweltdatensätzen wurden solche Extrapolationen gemacht und diskutiert. Es sollte geprüft werden, ob der einfache Versuch, die Jahresdurchschnittstemperatur zu verändern und auf diese geänderten Umweltbedingungen zu extrapolieren, plausible Resultate liefert.

Es war vorgesehen, die Frage nach dem Einfluss der Rastergrösse auf die Prognosegenauigkeit zu untersuchen. Dazu wurde in einem Versuchsgebiet ein detaillierteres Höhenmodell erstellt, das eine Verbesserung der Auflösung von 25×25 m auf gut 5×5 m zulässt. Die in den vorangegangenen Untersuchungen als wichtig erkannten Umweltparameter mussten damit nachgerechnet und die Modelle darauf angewendet werden. Mit der in Abschnitt 6.2.1 dargelegten Begründung wurde diese Teilaufgabe jedoch nicht weiter verfolgt.

Kapitel 4

Projektbeschreibung

Ein grosser Teil der Arbeit eines Modellierungsprojektes fällt mit der Abgrenzung des Untersuchungsgebietes, der Suche und Aufbereitung der Eingangsdaten und der Wahl der Methodik an. Nach der Durchführung der Modellierungen liegen die Resultate zunächst als rohe Werteverteilung für die modellierte Habitategnung vor. Zudem geben Kennwerte Auskunft über die Güte der Modelle. Um die Modelle untereinander vergleichen zu können und um die Modelle in eine binäre Aussage zur Verbreitung der Art zu übersetzen, werden Schwellenwerte verwendet und auf die rohen Prognosekarten angewendet. Daraus lassen sich auch Arealstatistiken ableiten. Weitere Tests zum Vergleich der Prognosen geben Auskunft über die Unterschiede.

Die Planung künftiger Feldarbeit war ein Kernanliegen dieser Arbeit. Bei der zukünftigen Feldarbeit geht es nicht nur darum, neue Populationen zu finden, sondern auch darum, Daten zu sammeln um die Modelle gezielt zu verbessern. Mit den Unsicherheitskarten wurden die Gebiete mit grosser Unsicherheit hervorgehoben und für die Überprüfung priorisiert.

Schliesslich wurde versucht, die Modelle nicht nur im Untersuchungsgebiet anzuwenden, sondern auch darüber hinaus zu extrapolieren, um entferntere Gebiete zu beurteilen oder Hinweise auf Arten mit ähnlicher ökologischer Nische zu finden. Ein weiterer Anwendungsversuch war der Projektion in andere Klimate gewidmet.

4.1 Das Untersuchungsgebiet

Die Begrenzung des Untersuchungsgebietes wurde so gewählt, dass möglichst viele der zu erwartenden Verbreitungsgebiete¹ abgedeckt und alle Fundstellen umfasst sind. Gleichzeitig sollte aber der Anteil von ungeeigneten Gebieten nicht allzu gross werden. Aus technischen Gründen musste das Gebiet zudem eine rechteckige Form aufweisen. Für diese Untersuchung wurde daher das Gebiet zwischen Melchsee-Frutt (x= 663'868) und Uri-Rotstock (x= 684'855) bzw. Stanserhorn (y= 198'709) und Titlis (y= 179'944) gewählt. Das

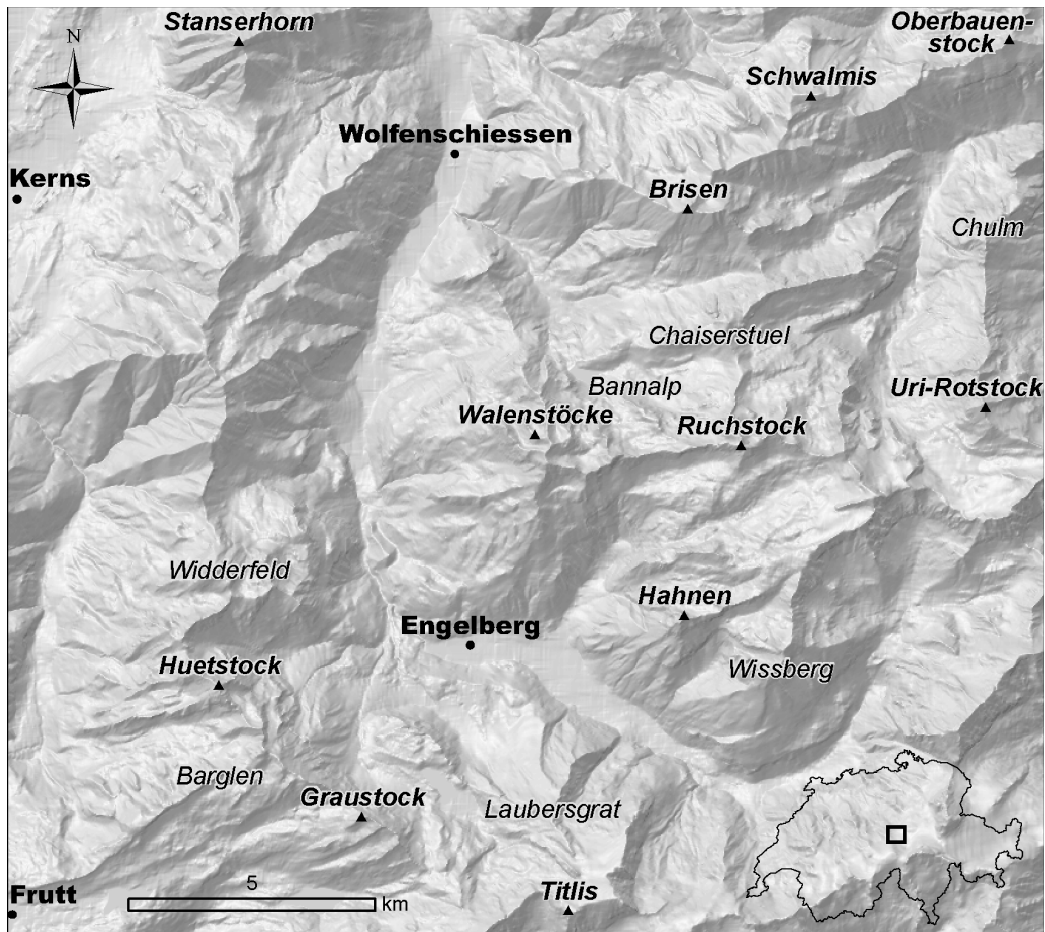


Abbildung 4.1: Karte des Untersuchungsgebietes mit den verwendeten Lokalnamen. Unten rechts: Lage des Untersuchungsgebietes in der Schweiz.

¹ Die erwarteten Verbreitungsgebiete wurden aus den Potentialgebieten in Baggenstos (2007) abgeleitet.

Untersuchungsgebiet umfasst die beiden Bergketten, die das Engelbergertal begrenzen, sowie die oberen Teile des Melchtals und des Grosstals. Am nördlichen Alpenrand gelegen, erstreckt es sich von 450 m ü.M. am Talausgang bis auf 3238 m ü.M. auf dem Titlis, bei einer durchschnittlichen Höhe von 1660. Mit Ausnahme der südlichsten Gebiete und der quartären Lockergesteine der Talungen ist die Geologie hauptsächlich durch die mesozoischen Sedimente der helvetischen Decken geprägt. Weite Gebiete sind verkarstet und weisen eine ausgeprägt unterirdische Entwässerung auf. In den höchsten Lagen halten sich einige kleinere Gletscher und Firnfelder. In Engelberg (1035 m ü.M.) betreibt MeteoSchweiz eine Messstation, deren Messungen der langjährigen² Monatsmittel der Temperatur von -2.7°C im kältesten Monat (Januar) bis 14.1°C im wärmsten (Juli) reichen. Niederschläge sind meist durch Westwindlagen oder durch Gewitter bedingt. Im Winter sind Inversionslagen, im Sommerhalbjahr Föhnlagen häufig.

4.2 Eingangsdaten in die Modellierung

4.2.1 Beobachtungen aus den Feldkampagnen

Die Funde von 2004 bis 2006 sowie die Beobachtungen der intensiven Kampagne im Sommer 2007 liegen als Datensätze vor: Zusammen sind das 113 Präsenzen und 169 Absenzen (davon liegt eine Absenz ausserhalb des Untersuchungsgebiets). Für die Planung der Feldkampagne 2007 wurde aufgrund von Expertenmeinung ein Potentialgebiet ausgeschieden, in welchem das Vorkommen der Art möglich schien. In mehreren Tages- und Mehrtagestouren wurde versucht, das Potentialgebiet einigermaßen gesamthaft zu besuchen. Wo es die Zugänglichkeit erlaubte, wurde auch neben den Pfaden gesucht. Im Zuge der Aufnahmen wurden folgende Parameter erfasst:

Einzelbeobachtungen

- Vorkommen von *T. biconicus* (Ja/Nein, leere Gehäuse)

² Die Messperiode ist 1961 bis 1990.

- Standort/Lebensraumtyp und Charakterisierung des Substrates
- Kleinrelief
- Bemerkungen zur Populationsgrösse und weitere Beobachtungen

Aufnahmeflächen

An 32 sogenannten Aufnahmeflächen vom 10×10 m wurden detaillierte Daten zur Populationsgrösse und zum Habitat gesammelt. Auf 20 Aufnahmeflächen wurde *T. biconicus* gefunden (Präsenz), auf 12 nicht (Absenz). Die Platzierung der Aufnahmeflächen wurde möglichst über das Potentialgebiet verteilt zufällig vorgenommen und es wurde nach einer standartisierten Aufnahmemethode³ vorgegangen. Zusätzlich zu den Parametern der Einzelbeobachtungen wurden weitere Daten erhoben:

- Dichteschätzung: Die Anzahl Funde von *T. biconicus* (lebend; nur Gehäuse) während einer Suche von 5 Minuten (dreifache Durchführung mit Mittelwertbildung)
- weitere Mollusken Arten
- Exposition und Neigung
- Angaben zur Fels- und Schuttbedeckung
- Kalkgehalt und Kornverteilung der Steine
- Mikroklima unter einem besiedelten Stein, das heisst im potentiellen Lebensraum: Temperatur und Luftfeuchte
- weitere Beobachtungen wie Hinweise zur Vergletscherungsgeschichte

Die Feldarbeit wurde durch Biologen und eingeführte Laien bestritten. Die Verortung der Beobachtungen und ein Teil der Datenerfassung wurde mit

³ Das verwendete Erfassungsformular und die Beschreibung der Aufnahmemethode ist in (Baggenstos, in prep.) erläutert.

GPS-Geräten durchgeführt⁴. Die Bestimmungen wurden durch die Molus-
kensenzialisten M. Baggenstos und J. Rüetschi vorgenommen.

Um Widersprüche zu vermeiden, wurden die Rasterzellen, welche gleich-
zeitig eine Präsenz- wie auch Absenzinformation enthielten, weggelassen und
bei denjenigen mit mehreren Präsenzpunkten (oder Absenzzpunkten) die über-
zähligen Punkte ausgeschlossen. Die verwendete Anzahl Präsenzen reduzierte
sich so auf 106 (siehe Abb. 4.2).

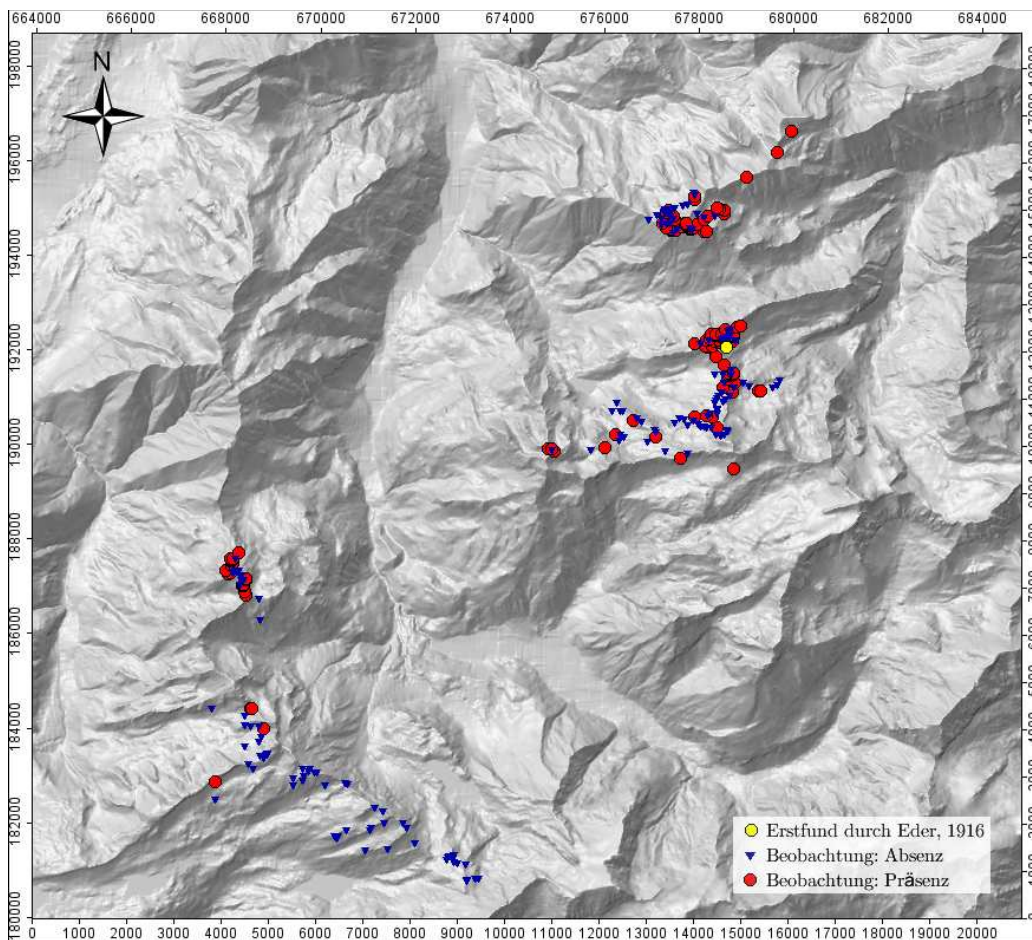


Abbildung 4.2: Die Verteilung der Präsenzen (rot) und der Absenzen (blau) im Untersu-
chungsgebiet und Lage des Erstfundes am Chaiserstuel (gelb). Rechts und unten ist die
Ausdehnung des Untersuchungsgebietes in Metern, links und oben sind die Koordinaten
nach dem Schweizer Koordinatenreferenz System angegeben.

⁴ Es standen zwei Geräte im Einsatz: ein MobileMapper von Thales und ein GeoXH
von Trimble.

Datenkritik

An Beobachtungsdaten werden mancherlei Qualitätsansprüche gestellt. Angelehnt an die Kriterien von Hengl (2007), wird hier eine Datenkritik durchgeführt (die Kriterien sind im folgenden kursiv):

Ausreichender Umfang der Daten *Der Datensatz muss einen genügenden Umfang aufweisen, damit die statistischen Aussagen fundiert sind.* Verschiedene Versuche mit MAXENT haben gezeigt⁵, dass auch mit weniger als 25 Funden eine gute Prognose gemacht werden kann. Mit etwas über 100 Fundstellen ist diese Studie auf der sicheren Seite, insbesondere weil sie sich damit im Bereich der für solche Studien üblichen Nachweiszahlen bewegt⁶. Allerdings gelten noch die im nächsten Punkt und unter 3.2 gemachten Vorbehalte.

Repräsentative Verteilung *Die Daten müssen das ganze Interessengebiet abdecken und zwar sowohl in geografischer Hinsicht, wie auch in der Diversität der Umweltvariablen⁷.* Der erste Punkt ist – im speziellen bei den Absenzdaten – nicht erfüllt, bildet aber eine Grundmotivation, die Modellierung überhaupt durchzuführen. Der zweite Punkt ist für die hier gewählten Modellierungsverfahren deutlich wichtiger, hingegen weit schwieriger abzuschätzen: Steile Felswände sind beispielsweise nur sehr unzureichend vertreten. Die Art ist durch die Funddaten jedoch an mindestens vier wichtigen Lokalitäten⁸ jeweils mehrfach nachgewiesen, was ausreichend sein dürfte.

Unabhängigkeit *Die Suchstrategie muss möglichst objektiv und ohne Beeinflussung durch andere Gegebenheiten sein.* Die Suchstrategie wird in Baggenstos (in prep.) detailliert beschrieben. Die oft als Lösung angepriesene zufällige Beprobung oder eine Rasterbeprobung sind im gebirgigen Umfeld fast nicht zu beherrschen, bzw. nicht durchführbar: Die

⁵ Siehe Abschnitt 2.7 und die Beschreibung der Strategien von MAXENT bei kleinen Beobachtungsdatenmengen in Phillips et al. (2004).

⁶ Vergleiche dazu auch die Anmerkungen in der Einleitung und den Untersuchungen zum Umfang an Funddaten in Guisan et al. (2007).

⁷ Vergleiche dazu Abschnitt 3.2.

⁸ Namentlich sind dies: Chaiserstuel, Brisen, Walenstöcke - Ruchstock und Widderfeld.

Zugänglichkeit ist oft stark eingeschränkt und das subjektive Empfinden für Distanzen ist stark von der Wegsamkeit und der Verfassung des Suchers abhängig. Einer gewissen Beeinflussung der Suchintensität durch Wege und Pfade ist kaum auszuweichen, sollte sich hier aber in Grenzen halten, weil nach Möglichkeit sofort auf eine flächige Suchstrategie⁹ gewechselt wurde. Je grösser jedoch die beeinflussenden Gradienten sind, desto näher sollten die Suchpunkte liegen. Dies führt dazu, dass in steilen Felswänden sehr ausgiebig gesucht werden müsste.

Lagefehler *Unzureichende Genauigkeit in der Lokalisierung, Fehler im Erfassen (z.B. Tippfehler) oder Fehlmanipulationen in der Datenaufbereitung verursachen Fehler, welche den Zusammenhang von Beobachtung und Umweltvariablen falsch wiedergeben. Die Auswirkungen solcher Fehler auf die Modellierung können zu Prognosen führen, die mit der tatsächlichen Verbreitung sehr schlecht übereinstimmen.* Auffällige Lagefehler können durch optische Kontrolle auf einer Karte eliminiert werden. Unentdeckte Fehler werden die Modellbildung erschweren oder verfälschen. Wenn die Modellkennzahlen sehr schlecht werden und die verschiedenen Ansätze sich stark widersprechen, sollte der Prozessablauf der Datenerhebung und -verarbeitung genauer untersucht werden. Die meisten der während der Kampagne im Sommer 2007 erhobenen Fundstellen wurden mit GPS-Geräten erfasst und deren Fehler ausgewertet¹⁰. Sie erfüllen die notwendige Genauigkeit. Einige Fundstellen mussten jedoch mit Hilfe von Luftbildern verortet werden. Sie sind aber klar in der Minderzahl. Bei der Diskussion der Prognosen muss auf die Ausgangsdaten zurückgegriffen werden können, im Speziellen mit Hinblick auf die Verwendung des tiefsten prognostizierten Wertes als Schwellenwert.

Konsistente Methodik *Alle Prozesse müssen reproduzierbar sein und den*

⁹ In flacheren Gebieten kann leichter von den Wegspuren abgewichen werden als bei steilen Passagen. Der Lebensraum der Art kann aber auch auf kleinen Grasbändern in den Felswänden gegeben sein. Ein Hinweis dazu ist der sehr exponiert gelegene Fund auf der Barglen von M. Trüssel (2.10.2006).

¹⁰ Die Lagefehler der GPS-Messungen wurden entweder online oder nachträglich mit der Bodenstation ausgeglichen. Die Vorgabe in Baggenstos (2007) für Lagefehler <3m wird somit erfüllt.

Standardmethoden entsprechen. Die vorliegende Arbeit versuchte, die Teilaufgabe der Modellierung zu dokumentieren und somit reproduzierbar zu machen.

Ausreichende Genauigkeit *Die Angaben zu den Umweltvariablen müssen in signifikanter Genauigkeit vorliegen.* Die Stabilität der Umweltvariablen gegenüber zufälligen Schwankungen und Artefakten ist im Speziellen bei den durch Interpolation aus Messstationen und Hochrechnung auf das Höhenmodell gewonnenen klimatischen Daten etwas in Frage gestellt. Die direkt aus dem Höhenmodell abgeleiteten Variablen weisen bei starken topografischen Wechsels nur eine knapp ausreichende Genauigkeit auf¹¹ und stellen durch die Mittelwertbildung über eine Rasterzelle nicht unbedingt den wirksamen Wert dar. Die Ausscheidung der Bodenbedeckung liegt zwar in teilweise sehr kleinen Flächen und präzisen Begrenzungen vor, die inhaltliche Aussage wurde aber mit einem anderen Fokus (eher wirtschaftliche Nutzung) und von verschiedenen Bearbeitern vorgenommen (siehe auch Abschnitt 3.3.2).

Räumliche Auflösung *Die Angaben zu den Umweltvariablen müssen von aussagekräftiger räumlicher Auflösung sein.* Grundsätzlich scheint eine Rastergröße von 25×25 m für den Lebensraum der Art und das dafür entscheidende Mikroklima sehr grob zu sein. Für eine Näherung ist dies jedoch ausreichend und nachvollziehbar. Die prognostizierten Gebiete sagen daher eher etwas über die generellen Bedingungen aus, als über das Mikroklima. Die Frage nach verbesserten Aussagen durch Höhenmodelle mit feinerer Auflösung wird in Abschnitt 6.2.1 nochmals aufgegriffen.

Vergleichbare Zeitschnitte *Der Erfassungszeitraum der Umweltvariablen und der Funde müssen sich entsprechen.* Die Veränderung der Umweltbedingungen mit der Zeit beeinflusst das Verbreitungsgebiet von Arten. Dem ist bei der Wahl der Funddaten und der Umweltvariablen Rechnung zu tragen. Es wurde hier angenommen, dass sich kurzfristige

¹¹ Beispielsweise werden Grasbänder in steilen Felswänden nicht wiedergegeben und spitze Kämme und Gipfel werden verflacht.

jährliche Schwankungen zwar auf die Häufigkeit, nicht aber grundsätzlich auf das Vorkommen oder Nicht-Vorkommen der Art auswirken. Daher wurden für die Umweltparameter Durchschnittswerte über eine Periode von 30 Jahren gewählt. Die Funde hingegen stammen alle aus 3 Jahren. Die Diskrepanz zwischen dem Zeitintervall 1961 bis 1990 der Umweltdaten und den Fundjahren 2005 bis 2007 - begründet durch die Verfügbarkeit der Daten - ist knapp vertretbar. Die Funde von Eder (1917) ohne Nachprüfung im Felde in die Modellierung zu übernehmen, wäre aber nicht zulässig¹².

Durch den derzeitigen Kenntnisstand konnte erst schwer abgeschätzt werden, wie sich die Tagestemperatur und das Wetter auf die Beobachtungswahrscheinlichkeit auswirken. Die Annahme war jedoch, dass das Mikroklima unter den Steinen sehr konstant ist und das Tageswetter nur Auswirkungen auf die Aktivität der Tiere hat. Die grösste Aktivität war von Wigger (2007) in den Nachtstunden gefunden worden. Daher sah die Suchstrategie vor, die Tiere unter den Steinen zu suchen.

4.2.2 Vermessungsdaten

Eine grössere Datensammlung wurde von der kantonalen Vermessungsstelle LIS Nidwalden AG auf einer Harddisk für das Gebiet der Kantone Obwalden und Nidwalden zur Verfügung gestellt. Die Sammlung umfasst topografisches Kartenmaterial, Luftbilder, die Daten der amtlichen Vermessung inklusive Klassierung der Bodenbedeckung, Vermessungspunkte und andere topografische Angaben. Die Karten und Luftbilder wurden nur zu Orientierungszwecken benutzt. Die Klassierung der Bodenbedeckung konnte jedoch in die Modellierung miteinbezogen werden.

¹² Die Art wurde am *Locus classicus* in den neueren Beobachtungen bestätigt. Eder hat im Sommer 1916 an weiteren Orten vergeblich gesucht, wo in den letzten Jahren jedoch Funde nachgewiesen wurden.

4.2.3 Umweltdaten der WSL

Umfangreiches Datenmaterial der WSL mit Klimainterpolationen¹³ und aus dem Höhenmodell abgeleiteten Informationen standen zur Verfügung. Die Recherchen und Berechnungen wurden von Klaus Ecker, WSL, durchgeführt und zur Verfügung gestellt¹⁴. Die Umweltvariablen liegen landesweit in einem aus dem DHM25¹⁵ abgeleiteten 25 × 25 m-Raster vor.

4.2.4 Geologische Karten

Es wurde versucht, innerhalb dieses Projektes das vorhandene geologische Kartenmaterial zusammenzutragen und aufzubereiten. Schwierigkeiten mit der Georeferenzierung und Harmonisierung der recht unterschiedlichen Quellen, sowie der sehr grosse, manuelle Digitalisierungsaufwand haben zum Abbruch der Aufbereitung dieser Informationsquelle geführt.

4.2.5 Satellitenbilder

Auf die Verwendung von Satellitenbildern wurde verzichtet, da sie in der verfügbaren Auflösung keinen Informationsgewinn versprachen. Am ehesten wären Rückschlüsse zur Vegetation zu gebrauchen gewesen. Diese waren aber durch die 'Bodenbedeckung' der kantonalen Vermessungsstellen bereits in besserer Qualität vorhanden.

4.3 Modellieransatz mit den verfügbaren Umweltvariablen

Der in Abschnitt 3.3.2 aufgezeigte Ansatz geht von den verfügbaren Umweltdaten aus und sucht nach statistischen Abhängigkeiten zwischen den Umwelt-

¹³ Die hier verwendeten Klimalayer wurden von Niklaus Zimmermann, WSL, gerechnet und zur Verfügung gestellt. Siehe Zimmermann & Kienast (1999).

¹⁴ Aus den monatlich vorliegenden Variablen wurden nur März, Mai, Juli und August verwendet.

¹⁵ DHM25 ©1994, Bundesamt für Landestopographie.

variablen und dem Vorkommen der Art. Die 'Mind Map' (Abbildung 4.3) zeigt diesen Ansatz. Nicht alle dort aufgeführten Datenquellen standen in

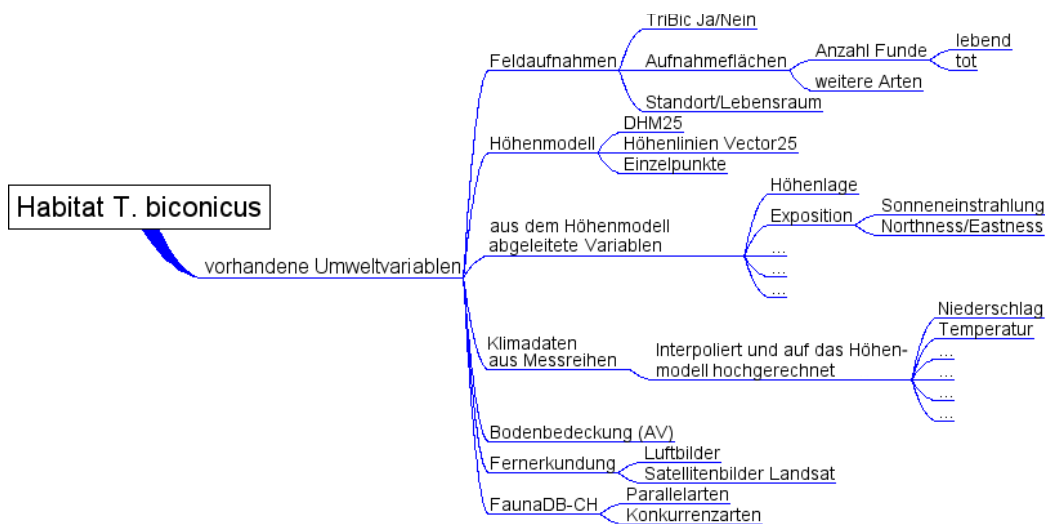


Abbildung 4.3: Der korrelative Ansatz zur Modellierung orientiert sich an den vorhandenen Daten.

genügender Abdeckung oder Ergiebigkeit zur Verfügung. Dennoch verfügte diese Studie über eine grosse Zahl an Umweltvariablen.

Alle Umweltvariablen mussten im selben 25 × 25 m-Raster vorliegen und genau auf das Untersuchungsgebiet zugeschnitten werden.

4.3.1 Reduktion aufgrund der Korrelation

Die über 100 Umweltvariablen wurden auf Korrelation untersucht: Eine hohe Korrelation zwischen zwei Variablen bedeutet, dass die erste einen grossen Anteil des Informationsgehaltes der zweiten abdeckt und die zweite kaum mehr neue Information zum Modell beitragen kann. In einem ersten Schritt werden Variablen mit Korrelationskoeffizient¹⁶ $r \geq 0.85$ eliminiert. Dabei

¹⁶ Hier wird der Absolutwert der Korrelationen nach Pearson verwendet, welcher den linearen Zusammenhang zweier Variablen beschreibt. 0 bedeutet kein Zusammenhang und 1 ein vollständig linearer Zusammenhang. Andere, nicht lineare Zusammenhänge werden nicht berücksichtigt; vgl. dazu den durch die Winkelfunktionen Cosinus bzw. Sinus gegebenen Zusammenhang von *Northness* und *Eastness* in Abb. 4.4.

Abkürzung	Name	Originalbezeichnung	Wertebereich [Einheit]	Herkunft	Level			
					A	B	C	D
Bod	Bodenbedeckung	Bodenbedeckung bzw. CodeReclass	3 Kategorien (von urspr. 60)	LIS Nidwalden AG / LISAG Uri	*	*	*	
Cur_pln	Kurvatur quer zur Fallrichtung	cur25_plan	-42.55 bis 48.04	DHM	*			
Cur_prf	Kurvatur in der Fallrichtung	cur25_prfl	-57.11 bis 43.93	DHM	*			
Eness	Exposition: Ostanteil	eness25	-1 bis +1	DHM	*	*	*	
Nness	Exposition: Nordanteil	nness25	-1 bis +1	DHM	*	*	*	
Etpt7	Evapotranspiration im Juli	etpt7	5 bis 57 [1/10 mm / Tag]	Interpolation und DHM	*			
PDay	Regentage während der Wachstumszeit	pday	41 bis 49 [Tage]	Interpolation und DHM	*	*		
Prec7	Niederschlag im Juli	prec7	1490 bis 1952 [1/10 mm / Monat]	Interpolation und DHM	*	*		
Slope	Hangneigung	slp25f_di	0° - 12.6° [90° Grad]	DHM	*			
SolRad	Potentielle, direkte Sonneneinstrahlung	shwc25ss60	0 bis 29'565 [kJ/m ² /Tag]	DHM	*	*	*	
SWB	Wasserbilanz	swb	-557 bis 995 [1/10 mm / Jahr]	Interpolation und DHM	*			
TAve	Jahresdurchschnittstemperatur	taveyy	-498 bis 894 [1/100 °C]	Interpolation und DHM	*	*	*	*
TOPO / TOPOs	Topografischer Index	topo / topos	-∞ (-2566 bzw. -968) bis +∞ (4778 bzw. 1562)	DHM	*			
tp2000s	Topografische Position Fokus: 2 km	tp2000s	-381.7 bis 497.9 [m]	DHM	*	*	*	*
TWI / TWIs	Topografischer Wetness-Index	twi25 / twi25s	-654 bis 1062 bzw. -551 bis 687	DHM	*			
Wet	Wetnessindex	wet25f_di	0 bis 0.506	DHM	*	*	*	*

Tabelle 4.1: Zusammenstellung der für die Modellierung verwendeten Umweltvariablen und der im Untersuchungsgebiet angetroffenen Wertebereiche dieser Variablen. Die 'Originalbezeichnung' bezieht sich auf den Namen in der Datensammlung von K. Ecker, WSL (ausser bei Bod). Die 'Level' beziehen sich auf die sukzessive Verringerung der Anzahl Umweltvariablen bei der Modellierung (Abschnitt 4.4).

	Cur_prf	SolRad	TWI	TWIs	TOPO	TOPOS	TAve	PDay	Wet	tp2000s	Slope	Eness	Nness	SWB	Prec7	Eptp7	Bod
Cur_pln	0.51	0.17	0.33	0.20	0.60	0.38	0.07	0.03	0.39	0.21	0.13	0.03	0.07	0.08	0.01	0.02	0.04
Cur_prf	.	0.05	0.30	0.10	0.77	0.43	0.21	0.11	0.73	0.22	0.50	0.17	0.07	0.08	0.12	0.01	0.06
SolRad	.	.	0.19	0.37	0.10	0.25	0.24	0.28	0.20	0.22	0.53	0.14	0.60	0.16	0.17	0.68	0.22
TWI	.	.	.	0.75	0.38	0.36	0.17	0.06	0.57	0.21	0.47	0.21	0.10	0.10	0.03	0.29	0.07
TWIs	0.40	0.51	0.30	0.28	0.53	0.46	0.59	0.07	0.06	0.30	0.31	0.40	0.19
TOPO	0.76	0.30	0.12	0.67	0.42	0.42	0.23	0.09	0.03	0.18	0.17	0.04
TOPOS	0.37	0.17	0.51	0.59	0.24	0.21	0.21	0.04	0.20	0.24	0.12
TAve	0.78	0.28	0.44	0.18	0.12	0.35	0.12	0.77	0.79	0.15
PDay	0.19	0.20	0.10	0.05	0.36	0.22	0.67	0.76	0.15
Wet	0.25	0.82	0.32	0.07	0.00	0.13	0.30	0.04
tp2000s	0.09	0.08	0.26	0.05	0.14	0.46	0.04
Slope	0.29	0.01	0.05	0.09	0.47	0.08
Eness	0.05	0.04	0.00	0.03	0.23
Nness	0.18	0.18	0.36	0.21
SWB	0.15	0.61	0.21
Prec7	0.66	0.22
Etp7	0.55

Tabelle 4.2: Korrelationen mit Korrelationskoeffizienten $r \leq 0.85$. Mit diesen Umweltvariablen wurden die Modellierungen begonnen (entspricht Level A).

wird jeweils die besser interpretierbare der weniger gut interpretierbaren Variablen bzw. der weniger gut zu verallgemeinernden vorgezogen. So ist z.B. im hoch korrelierten Paar 'Höhe über Meer' und 'Jahresdurchschnittstemperatur' der zweiten Variablen den Vorrang zu geben. Falls das Modell in die Südalpen extrapoliert werden soll, ist die Temperatur biologisch sinnvoller als die Höhe. Durch die Reduktion auf $r \leq 0.85$ entsteht ein reduzierter Satz von 18 Variablen (Tabelle 4.1). Im Anhang A sind die Wertebereiche dieser Umweltvariablen grafisch dargestellt. Dieser *Level A* genannte Satz an Umweltvariablen ist der Ausgangspunkt für alle Modell-Varianten. Durch weitere Einschränkung auf $r \leq 0.80$ entstand der Satz von 9 Umweltvariablen (*Level B*) beziehungsweise durch Einschränkung auf $r \leq 0.75$ *Level C* mit 7 Variablen.

4.3.2 Beschreibung ausgewählter Umweltvariablen

Die Herleitung und Aussage der in GLM, BRT und MAXENT verwendeten Umweltvariablen mit $r \leq 0.8$ (entspricht Level B), werden nachfolgend beschrieben.

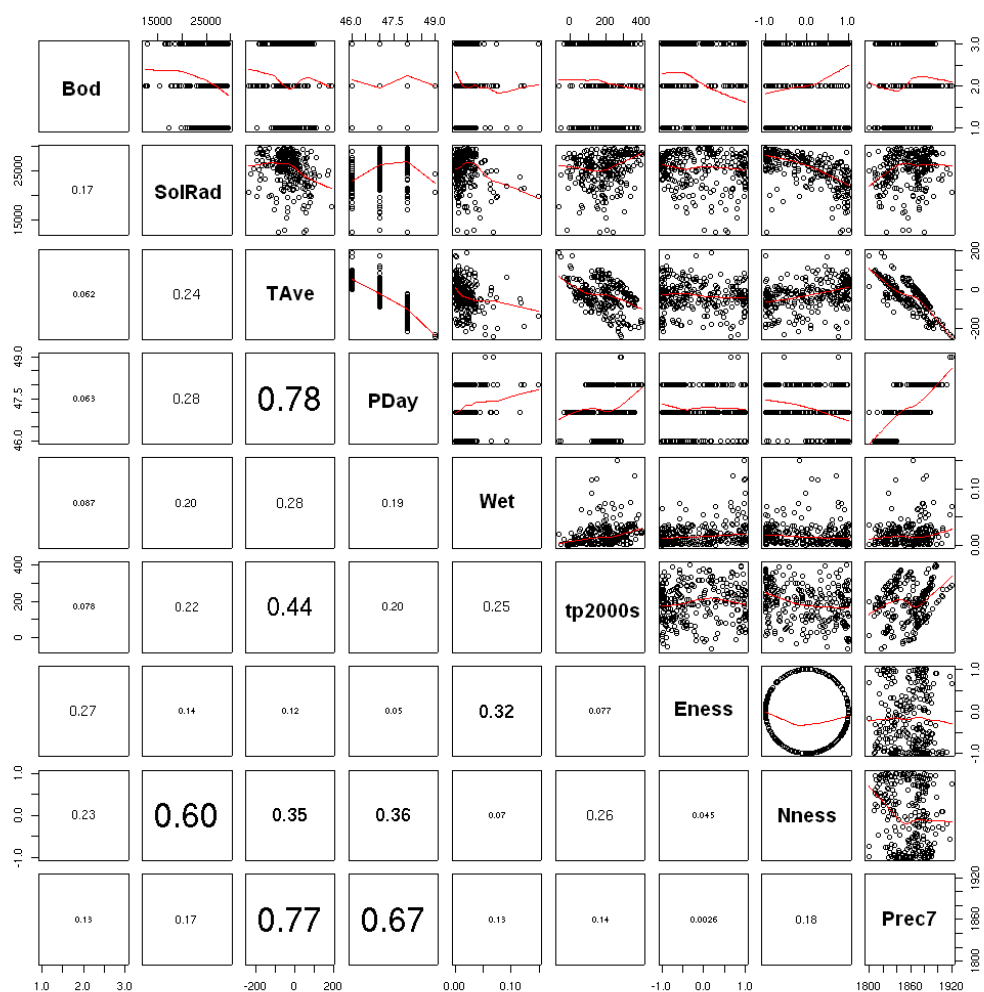


Abbildung 4.4: Korrelationsmatrix mit Korrelationskoeffizient $r \leq 0.80$: Jedes Streudiagramm in der oberen Hälfte der Matrix ist ein Plot der Umweltvariable der Zeile gegen diejenige der Spalte. In der unteren Hälfte der Matrix sind die Korrelationskoeffizienten r nach Pearson angegeben, skaliert gemäss ihrer Grösse. Die Abkürzungen der Umweltvariablen sind in Tabelle 4.1 aufgelistet.

Aus dem DHM25 abgeleitete Umweltvariablen

Hangneigung Slope: In einem quadratischen Raster kann die Hangneigung nicht wie in einer Dreiecksvermaschung eindeutig aus den Nachbarzellen abgeleitet werden. Verschiedene Ansätze zur Lösung dieses Problems sind bekannt. Hier wurde die '*Deterministic Infinity*'-Methode angewendet (Wilkin et al., 2007), welche eine optimale Bestimmung der Hangneigung in jeder Zelle aus ihren Nachbarn verspricht. Durch die begrenzte Genauigkeit eines digitalen Höhenmodelles entstehen kleinere Senken, die meist als Artefakte angesehen werden können. Durch das Auffüllen dieser Senken, werden solche Modelle - vor allem aus hydrologischen Überlegungen - vorbehandelt, um nachfolgend Analysen zu ermöglichen.

Potentielle Sonneneinstrahlung SolRad: Die Berechnung der Sonneneinstrahlung auf eine Rasterzelle in Abhängigkeit von der geografischen Lage sowie von Position und Neigung im Höhenmodell ist in verschiedenen Applikationen mit unterschiedlichen Ansätzen umgesetzt. Die Berechnungsarten unterscheiden sich unter anderem im Einbezug oder Nicht-Einbezug von Parametern wie Atmosphärendurchlässigkeit, Bewölkungsdichte, indirekte Abstrahlung, Schattenwurf und so weiter. Die Berechnung auf einen Rasterdatensatz kann von längeren, summierten Abschnitten (z.B. ein Jahr oder eine Saison) bis zu kurzen Zeitpunkten (Datum, Stunde) ausgelegt werden oder aber auf spezielle Konstrukte, wie beispielsweise die Summe der Sonneneinstrahlung aller Tage um 12 Uhr.

Hier wurde der auf die Nordhemisphäre korrigierte Code von L. Kumar für die *potential direct clear sky insolation (shortwave)* verwendet und damit die Tagessumme für den 21. Juni (Sommersonnenwende) im Stundenintervall gerechnet und summiert¹⁷.

Exposition *Eastness*: *Eness* bzw. *Northness*: *Nness*: Nordanteil = $\cos(\text{aspect})$ bzw. Ostanteil = $\sin(\text{aspect})$. Die Exposition (*aspect*) wird in Grad (360°) angegeben. Daher ergibt sich ein nicht kontinuierlicher Übergang

¹⁷ Details unter http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml1_1.html.

von 360° zu 0° . Dieser Sprung hat jedoch keine topologische Bedeutung und muss deshalb verhindert werden. Dazu wird die Exposition in zwei Datensätze aufgeteilt, welche die Orientierung der *Fallrichtung* in Nord-Südrichtung bzw. in Ost-Westrichtung zwischen 1 und -1 als Kontinuum wiedergeben.

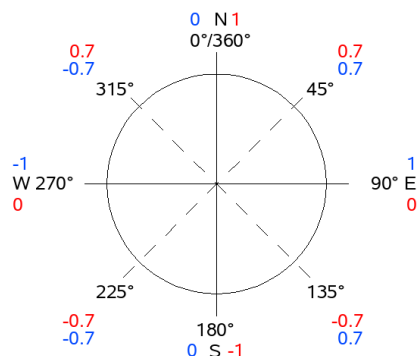


Abbildung 4.5: Die Aufteilung der Exposition in *Northness* (rot) und in *Eastness* (blau) verhindert die Diskontinuität der Exposition beim Übergang von 360° zu 0° .

Topografische Position *tp2000s*: Dieser Wert zeigt die relative Position eines Ortes in einem topografischen Gradienten (hohe Werte zeigen Gipfel- oder Gratlage, tiefe Mulden- oder Tallage). Die Aussage wird durch den Vergleich des Höhenwertes einer Zelle mit dem Mittelwert dessen Umgebung in einem Umkreis von 2 km Radius errechnet¹⁸ und mit einem Filter¹⁹ gedämpft (*smooth*).

Topografische Indices *TOP0* und *TOP0s*: Diese Indices zeigen die topografische Position an. Auf ein Höhenmodell angewendet, wird von jedem Punkt in einem kreisförmigen Suchfenster mit wachsendem Radius die durchschnittliche Höhe bestimmt. Diese wird mit dem Höhenwert des Mittelpunktes verglichen. Die Aussagen der resultierenden Karten können als relative topografische Exposition der Zellen zu ihrem Umfeld im jeweiligen Suchradius bezeichnet werden. Durch Standardisierung dieser Karten und Integration zu einer kombinierten Karte wird erreicht,

¹⁸ Siehe auch <http://esapubs.org/archive/mono/M077/018/appendix-A.htm> .

¹⁹ Hier kommt ein 3×3 -Zellen *mowing window* zur Anwendung.

dass in einer einzigen Karte kleinräumige bis grossflächige Expositionen dargestellt werden können (Zimmermann & Roberts, 2001). Positive Werte zeigen Rücken, Gipfel und exponierte Stellen an - negative Werte zeigen Senken, Rinnen und Täler. Werte nahe bei Null weisen auf eine gleichmässige Fläche (horizontal oder geneigt) oder einen Wendepunkt hin. TOPOS ist dieselbe Karte wie TOPO, nach Anwendung eines Dämpfungsfilters²⁰. Interessanterweise ist die Korrelation von TOPO zu TOPOs mit $r = 0.76$ erstaunlich klein (siehe Tabelle 4.2).

Wetnessindex Wet: Basierend auf der Hangneigung (Slope) und dem hydrologischen Einzugsgebiet einer Zelle (*Specific Catchment area*) wird auf deren Feuchtigkeitsangebot geschlossen. Das hydrologische Einzugsgebiet wird über die Rückverfolgung von Zuflüssen der Nachbarzellen bestimmt. Üblich ist die Verwendung von:

$$TWI = \ln \left(\frac{\text{hydrol. Einzugsgebiet}}{\tan(\text{Hangneigung})} \right)$$

Dabei tritt allerdings bei einer Hangneigung von 0° ein Fehler auf. Die hier verwendete Formulierung von Tarboton (2005)²¹ verhindert dies. Auf das Füllen der Senken in einem Höhenmodell und auf den DI-Algorithmus wurde unter 'Hangneigung' bereits hingewiesen.

Das Untersuchungsgebiet weist einen grossen Anteil an unterirdischer Entwässerung auf. Einer direkten Übersetzung in eine ökologische Aussage der Wirkungsvariablen sind daher Grenzen gesetzt. Die für das Habitat von *T. biconicus* wichtigen Kretenlagen dürften aber recht gut wiedergegeben sein.

Aus Messstationen interpolierte Umweltvariablen

Jahresdurchschnittstemperatur Tave: Ausgehend von Temperaturmessreihen von 1961 bis 1990 der Messstationen von MeteoSchweiz, bestimmten Zimmermann & Kienast (1999) die langjährigen Monatsmittel und projizierten diese mittels der Höhenabhängigkeit auf Meeresni-

²⁰ Es wird ein *Focal Mean* mit 2 Zellen Radius verwendet.

²¹ Umgesetzt im TauDEM-Zusatzpaket für ArcGIS.

veau. Die Höhenabhängigkeit wurde durch Regression der Temperaturen mit der topografischen Höhe der Messstationen bestimmt. Auf dem Referenzniveau wurden die Messwerte interpoliert²² und mit der Höhenabhängigkeit auf die Höhe der Zellen umgerechnet. Für die Jahresdurchschnittstemperatur (T_{Ave}) wurden schliesslich die Monateswerte summiert und durch 12 dividiert.

Evapotranspiration und Niederschlag im Juli Et_{pt7} , $Prec7$: Die Evapotranspiration wurde nach der Formel von Turc (1963)²³ abgeleitet. Darin sind die prozentuale Bewölkung und die potentielle direkte Sonneneinstrahlung enthalten²⁴. Der Niederschlag wurde analog zur Durchschnittstemperatur aus einer lokal aus Messreihen zu bestimmenden Relation zwischen Höhe und Messwert und der Interpolation der Messwerte zwischen den Stationen bestimmt. Die Durchschnittswerte wurden aus der Messreihe von 1961 bis 1990 abgeleitet.

Regentage $PDay$: Die Anzahl Regentage in der Vegetationszeit (*Number of precipitation days per growing season*) wurde ebenfalls aus den Messreihen 1961 – 1990 interpoliert und auf das Höhenmodell hochgerechnet.

Wasserbilanz SWB : Aus Bodeneigenschaften, wie der Wasserspeicherkapazität und klimatischen Parametern, kann nach einem Verfahren von Grier & Running (1977) auf die den Pflanzen zur Verfügung stehende jährliche Wassermenge geschlossen werden (Roberts et al., 1993). Dabei wird die Differenz der monatlichen Niederschläge und der potentiellen Evapotranspiration für die Erdscholle über das ganze Jahr aufsummiert. Die hier verwendete Methode unterscheidet sich zu jener von Grier & Running (1977) in der Definition des hydrologischen Jahres²⁵. Die Bezeichnung leitet sich aus *site water balance* ab.

²² Die Details zur verwendeten thin-plate SPLINE-Interpolation sind in Zimmermann & Kienast (1999) beschrieben.

²³ Zitiert in Zimmermann & Kienast (1999).

²⁴ Die Bewölkung *Cloudiness* wurde aus den Messreihen der Meteostationen berechnet, die Sonneneinstrahlung aus dem Höhenmodell und der geografischen Lage (Details dazu in Zimmermann & Kienast (1999)).

²⁵ Für die Details zu den Parametern, zur Beschreibung der Erdscholle und des Wasserhaushaltes über das Jahr wird auf Roberts et al. (1993), Guisan et al. (2006) und <http://www.wsl.ch/staff/niklaus.zimmermann/biophys.html#SWB> verwiesen.

Umweltvariablen aus Erhebungen (Vermessung)

Bodenbedeckung Bod: Diese Angaben wurden aus der Bodenbedeckung (Datenmodell DM01) der Kantone NW, OW und UR zusammengeführt und auf Grund der Spalte 'Kulturart' neu eingeteilt. Im betroffenen Gebiet des Kantons Bern ist die Bodenbedeckung noch nicht erfasst. Die so entstandenen 60 Klassen wurden mit Rücksicht auf die Klassen mit den Präsenz- und den Absenpunkten stark vereinheitlicht: Das Attribut Bod enthält nur mehr die Klassen

500: "humusierete und übrige Flächen",

603: "vegetationslos: Fels",

604: "vegetationslos: Geroell, Sand", was als Lockergestein interpretiert wird.

4.4 Modellieren mit MAXENT

Die vorgegebenen Einstellungen in MAXENT wurden grundsätzlich übernommen. MAXENT entscheidet aufgrund der Anzahl Präsenzen, ob die Umweltvariablen als lineare Terme, als quadratische Terme oder als Produktterme, beziehungsweise als Kombinationen davon berücksichtigt werden (siehe Phillips et al. (2004) und die Hilfe zu MAXENT). Als Ausgabeform wurde 'logistisch' gewählt (Abbildung 4.6).

Die Präsenzdaten und die Umweltvariablen können in MAXENT separat eingelesen werden. Das Verschneiden der Beobachtungspunkte mit den Umweltvariablen wird von MAXENT übernommen.

4.4.1 Vereinfachen durch sukzessives Weglassen von Umweltvariablen

Es wurden mit Variablensätzen für *Level A* (18 Variablen), *Level B* (9 Variablen) und *Level C* (7 Variablen) gearbeitet, welche nach dem in 4.3.1 beschriebenen Verfahren zum Eliminieren zu hoher Korrelationen zwischen den

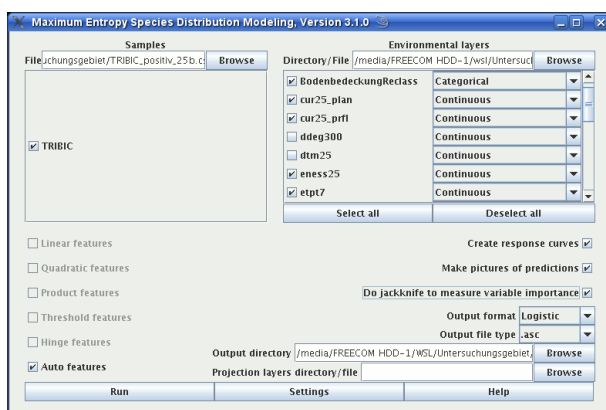


Abbildung 4.6: Das Hauptfenster von MAXENT: Die Beobachtungen (links oben), Umweltvariablen (rechts oben) und die Ausgabeoptionen (unten).

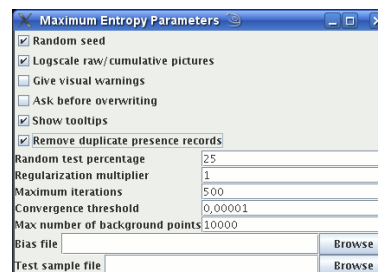


Abbildung 4.7: Speziellere Einstellungen: 25% der Präsenzen als Testdaten verwenden, 10'000 Pseudoabsenzen im Hintergrund erzeugen, usw.

Umweltvariablen erstellt wurden. Um die Grenzen von MAXENT auszuloten, wurde noch ein *Level D* eingeführt, der nur noch die wichtigsten 3 Variablen aus *Level C* verwendet; d.h. nur noch diejenigen Variablen, welche mehr als 5% zum Modell beitragen (*Percent contribution* > 5%).

4.4.2 Prüfung auf Stabilität mit Teildatensätzen

Die Berechnungen wurden auf jedem Level einmal mit dem vollen Präsenzdatensatz (Bezeichnung z.B. 'LevelA100') und zehn mal mit einer zufälligen Auswahl von 75% der Präsenzdaten als Trainingspunkten und den anderen 25% als Testpunkte durchgeführt (Bezeichnung z.B. 'LevelB025_10'). Die Kennwerte und die Rangfolge der Variablen aus den zehn Testmodellen wurden hinsichtlich Stabilität miteinander verglichen: Bleiben die Werte in einem engen Rahmen, ist es ein Hinweis auf ein stabiles Modell.

4.5 Modellieren mit R

Die Modellierungsmethoden GLM und BRT wurden in der statistischen Programmierumgebung R durchgeführt²⁶. Für die Modellierungen und Auswer-

²⁶ R Development Core Team (2007), <http://www.R-project.org>.

tungen wurden die von A. Bergamini, WSL, zusammengestellten R-Codes verwendet und angepasst²⁷.

4.5.1 Vorbereitung der Beobachtungen

Den R-basierten Modellierungen wurden die Beobachtungen als Punktdatensatz mit entsprechenden Werten der Umweltvariablen als Attribute übergeben. Bei Verwendung von Pseudoabsenzen wurde analog verfahren.

- a) Die Punktdaten (Beobachtungen) wurden mit den Umweltvariablen verschnitten und deren Werte als Attribute zu den Beobachtungen hinzugefügt²⁸ (inklusive Koordinaten). Das Attribut für die An- oder Abwesenheit der zu modellierenden Art wurde entsprechend mit 1 bzw. 0 angegeben.
- b) Datenreduktion: Der Datensatz wurde so reduziert, dass jede 25×25 m-Zelle höchstens einfach belegt war. Zellen mit sowohl Präsenz-, wie auch Absenzpunkten wurden weggelassen (`TRIBIC_25.txt`).
- c) Einlesen und Formatieren²⁹ des Beobachtungsdatensatzes in R: Die Jahresdurchschnittstemperatur (`taveyy`) wurde nach Kelvin `tavkel` umgerechnet, weil bei den linearen Modellen die negativen Werte nicht sinngemäss interpretiert werden.

So entstand der reduzierte Datensatz `d.red` für die Modellierung.

- d) Einlesen der Umweltvariablen für das ganze Untersuchungsgebiet: Mit dem Modell sollte eine Prognosekarte erstellt werden. Dazu mussten die Werte aller verwendeten Umweltvariablen für jede Zelle des Rasters

²⁷ Es wurden folgende R-Bibliotheken verwendet: Generalized Boosted Regression Models (`gbm`), Visualizing the performance of scoring classifiers (`ROCR`), Main Package of Venables and Ripley's MASS (`MASS`), Analysis of geostatistical data (`geoR`), Recursive Partitioning (`rpart`), Companion to Applied Regression (`car`), SAGA Geoprocessing and Terrain Analysis in R (`RSAGA`) und ODBC Database Access (`RODBC`).

²⁸ SAGA: `Shapes > Grid > Add Grid Values to Points`; GRASS: `v.what.rast`; ArcView mit Hilfe des Plug-Ins „Hawth's tool“ (<http://www.spatial ecology.com/htools>) > `Intersect Point Tool`.

²⁹ Die kategorischen Variablen müssen als solche gekennzeichnet werden (`as.factor`).

verfügbar sein. Analog zur Erstellung des Datensatzes unter a) mit den Beobachtungen, wurde ein Datensatz mit allen Rasterzellen des Untersuchungsgebietes und den zugehörigen Werten der Umweltvariablen als Attribute erstellt und eingelesen (`d.pred`).

4.5.2 *Generalized Linear Models* GLM

Beispielhaft werden die R-Code Fragmente und deren Ausgaben angegeben^{30,31}.

- a) Erstes Aufrufen der Funktion 'glm' (Umweltvariablen teilweise auch als quadratische Terme in der Formel):

```
g.mod <- glm(formula = TRIBIC ~ eness25 + tp2000s + prec7
             + etpt7+ shwc_ss + slp25f_ + twi25s + swb + tavkel
             + I(etpt7^2) + I(shwc_ss^2) + I(slp25f_^2) +
             I(topos^2) + I(twi25s^2) + I(swb^2) + I(tavkel^2),
             family = "binomial", data = d.red)
```

Für die Berechnung mit dem Pseudoabsenzdatensatz muss wegen dem zahlenmässigen Übergewicht der Pseudoabsenzen gegenüber den Präsenzen zusätzlich eine Gewichtung³² in die Formel einfließen (Maggini et al., 2006).

- b) Das Modell vereinfachen: Der Funktion `Step.AIC` wird das Objekt `g.mod` übergeben. Danach wird aufgrund des AIC³³ das Modell automatisch vereinfacht, d.h. es werden so lange Variablen aus dem Modell entfernt, bis das AIC ein Minimum erreicht. Die so erhaltenen Modelle sind allerdings immer noch sehr gross und viele Variablen sind

³⁰ Die dargestellten Werte stammen aus den Berechnungen mit dem Präsenz-/Absenz-Datensatz.

³¹ Die für diese Studie unter der Bezeichnung GLM verwendete Methodik ist in einem Tutorial des *Laboratory for Dynamic Synthetic Vegetation Phenomenology* der Universität Montana ausführlich beschrieben: <http://ecology.msu.montana.edu/labdsv/R/labs/lab4/lab4.html>.

³² Die Gewichtung wird so festgelegt, dass jede Präsenz das 'Gewicht' $W = 1$ bekommt und die Gewichtsummen gleich sind: $\sum W_{Präsenz} = \sum W_{Pseudoabsenz}$.

³³ Das *Akaike Information Criterion* AIC sucht die Güte der Anpassung eines Modells an die Beobachtungen zu optimieren und die Anzahl erklärender Variablen (Komplexität) zu minimieren.

nicht signifikant. Über die aktuellen Resultate orientiert der Aufruf von `summary(red.mod)`.

- c) Von Hand weiter reduzieren: Schrittweise werden nicht signifikante Variablen ausgeschlossen und das Modell neu gerechnet. Durch jedes Weglassen einer Variablen steigt allerdings die Residual-Devianz leicht an. Wie weit reduziert werden kann, muss fallweise entschieden werden.

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6112  -0.6819  -0.1893   0.6555   2.4622

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -6.944e+04  2.474e+04  -2.807 0.005000 **
eness25     -7.715e-01  2.914e-01  -2.647 0.008117 **
tp2000s      2.254e-02  3.803e-03   5.928 3.07e-09 ***
prec7        1.274e-01  2.284e-02   5.577 2.44e-08 ***
etpt7        2.027e+00  9.324e-01   2.174 0.029690 *
shwc_ss     -2.193e-03  1.055e-03  -2.080 0.037563 *
slp25f_      4.169e+00  1.373e+00   3.038 0.002385 **
twi25s       1.644e-02  6.170e-03   2.665 0.007700 **
swb         -2.224e-02  7.425e-03  -2.995 0.002746 **
tavkel       5.013e+02  1.811e+02   2.768 0.005636 **
I(etpt7^2)  -3.540e-02  1.436e-02  -2.466 0.013657 *
I(shwc_ss^2) 5.073e-08  2.065e-08   2.456 0.014041 *
I(slp25f_^2) -9.530e-01  3.193e-01  -2.985 0.002837 **
I(topos^2)   -1.808e-05  6.939e-06  -2.606 0.009162 **
I(twi25s^2)  5.178e-05  1.546e-05   3.349 0.000813 ***
I(swb^2)     7.932e-05  2.429e-05   3.265 0.001093 **
I(tavkel^2) -9.078e-01  3.314e-01  -2.739 0.006160 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

```

(Dispersion parameter for binomial family taken to be 1)

```

Null deviance: 353.01 on 262 degrees of freedom
Residual deviance: 223.86 on 246 degrees of freedom
AIC: 257.86

```

Number of Fisher Scoring iterations: 6

Erklärung: In der ersten Spalte sind die Variablen (oben) bzw. deren

Quadrate (unten) aufgelistet. **Estimate** zeigt die Steigung der Regressionsgeraden, **Std. Error** steht für die Standardabweichung der Fehler dazu. Der **z value** ist eine Testgrösse, mit der überprüft wird, ob die Steigung der Regressionsgerade signifikant von 0 verschieden ist oder nicht. Je grösser der absolute Wert von **z**, desto wahrscheinlicher ist die Steigung unterschiedlich von 0. **Pr** zeigt diese Wahrscheinlichkeit an. Ist die Steigung der Regressionsgeraden 0, trägt die Variable nicht zur Verbesserung des Modells bei.

d) Wenn alle verbleibenden Variablen eine hohe Signifikanz aufweisen, ist die Modelloptimierung abgeschlossen. Die Modellformel kann auf den Datensatz mit allen Rasterzellen (**d.pred**) angewendet und als Prognose in eine Datei geschrieben werden (**glm_prediction.txt**). Diese kann anschliessend in einer Prognosekarte dargestellt werden³⁴.

e) Modell Evaluation: AUC berechnen

```
> auc <- auc.rank(c.actual,c.predicted)35
> auc
[1] 0.8837083
```

f) Kreuzvalidierung: Die Beobachtungen werden in 10 Gruppen aufgeteilt. Innerhalb dieser Gruppen ist das Verhältnis von Präsenz- zu Absenzdaten gleich wie im Gesamtdatensatz. Auf jede dieser Gruppen wird das optimierte Modell aus d) angewendet, aber neu parametrisiert. Für die Evaluation wird für jede Gruppe der AUC-Wert berechnet. Ist der Durchschnitt der AUC-Werte der Läufe mit den Teildatensätzen (**apply(, ,mean)**) im Vergleich zum Lauf mit dem gesamten Datensatz nur wenig kleiner, ist dies ein Hinweis auf ein stabiles Modell.

```
> d.cross <- x.eval(red.mod,cr,d.red,c.actual,10)36
```

³⁴ In SAGA Import Shapes from XYZ und anschliessend Shapes to Grid; in ArcView darstellen mit xy anschliessend Features to Raster; durch die Verwendung des Paketes RSAGA können auch direkt Ascii-Raster Dateien geschrieben werden.

³⁵ **auc.rank**: Funktion zur Berechnung des AUC (geschrieben von A. Bergamini, WSL).

³⁶ **x.eval**: Funktion zur automatischen Durchführung einer Kreuzvalidierung (geschrieben von A. Bergamini, WSL).


```

> d.cross
      AUC
1  0.8312500
2  1.0000000
3  0.7937500
4  0.8187500
5  0.7625000
6  0.8562500
7  0.8375000
8  0.8562500
9  0.8312500
10 0.8190476
> apply(d.cross,2,mean)
      AUC
0.84065476

```

- g) Zum Schluss müssen die Residuen auf eine eventuelle räumliche Autokorrelation untersucht werden. Dazu werden Korrelogramme (Funktion `correlog()` aus Paket `ncf`) und Variogramme (Funktionen `variog()` und `variofit()` des Pakets `geoR`) erzeugt.

4.5.3 *Boosted Regression Trees* BRT

Die vorbereitenden Schritte a) bis c) aus Abschnitt 4.5.1 bleiben sich gleich.

- a) Laden der Bibliothek `gbm` und das Modell erstellen:

```

t.gbm <- gbm(d.red[,1] ~ CodeReclas + eness25 + nness25 +
  tp2000s + wet25f_ + prec7 + etpt7 + pday + shwc_ss +
  slp25f_ + topos + twi25 + twi25s + cur25_pln +
  cur25_prf + swb + topo + tavkel,
  distribution="bernoulli", data=d.red[,], n.trees=10000,
  interaction.depth=5, shrinkage=0.005)

```

Durch Aufruf der Funktion `gbm()` wird die Modellbildung gestartet. Der Zeitbedarf für die Modellbildung ist im Vergleich mit den anderen Modellierungsmethoden eher gross.

Mit `summary(t.gbm)` kann eine Liste der Variablen mit der relativen Wichtigkeit in Prozent ausgegeben werden:

```

      var rel.inf
1     tavkel 12.726
2     tp2000s 11.188
3     cur25_pln 7.995
4     prec7 7.864
5     nness25 6.786
6     twi25s 6.503
7     twi25 6.489
...

```

Für die Berechnung mit Pseudoabsenzen wurde dieselbe Gewichtung herangezogen, wie sie für die Berechnung in GLM hergestellt worden war.

- b) Die optimale Anzahl an Iterationen zur Verhinderung der Überanpassung (*overfit*) wurde nach der vorgeschlagenen Methode `cv` in der Funktion `gbm.perf` berechnet. Die Prognose wurde mit dieser Anzahl erstellt.

```

best.iter <- gbm.perf(t.gbm, method="cv", plot.it=F)
2462

```

- c) Herstellen der Prognosekarte³⁷.
- d) AUC berechnen³⁸. AUC = 0.965046
- e) 10-fache Kreuzvalidierung³⁹ (V1 bis V10) mit Einzel- und Durchschnittswerten.

	V1	V2	V3	V4	V5
best.iter	3314	2573	2376	2275	2014
AUC	0.6875	0.6812	0.7312	0.7750	0.8500

	V6	V7	V8	V9	V10
best.iter	3819	2026	3113	3332	2317
AUC	0.6062	0.8313	0.8562	0.8062	0.7000


```

> apply(xeval10, 1, mean)
best.iter      AUC
2715.9      0.7525

```

³⁷ In Analogie zu Punkt d) in Abschnitt 4.5.2.

³⁸ Siehe Punkt e) in Abschnitt 4.5.2.

³⁹ Die Kreuzvalidierung wurde mit derselben Funktion wie in Punkt f) in Abschnitt 4.5.2 gerechnet.

Der durchschnittliche AUC der Kreuzvalidierung ist in diesem Fall, im Vergleich zum AUC des Durchganges mit allen Daten, deutlich gesunken. Dies ist kein gutes Qualitätszeichen und weist auf eine Überanpassung (*'Overfitting'*) hin .

4.5.4 Residuen

Die Signifikanzen der Variablen werden in Regressionsmodellen nur dann richtig gerechnet, wenn die Residuen unabhängig voneinander sind. Ob dies der Fall ist, kann beispielsweise mit einem Test auf räumliche Autokorrelation überprüft werden. Nach Dormann et al. (2007) gibt es zahlreiche Gründe für räumliche Autokorrelationen, jedoch sind vier davon häufig:⁴⁰ 1) Biologische Prozesse wie Artbildung, Aussterben, Ausbreiten oder Artinteraktionen sind distanzabhängig; 2) Nicht lineare Beziehungen zwischen Art und Umweltvariablen wurden fälschlicherweise als lineare Beziehungen modelliert; 3) Dem statistischen Modell fehlt eine wichtige Wirkungsvariable, welche eine räumliche Struktur des Auftretens erklärt und deshalb eine räumliche Struktur im Modell bewirken würde; 4) Räumliche Autokorrelation kann durch die räumliche Auflösung des Modells (bzw. der Variablen des Modells) bedingt sein, weil größere Raster eine Glättung verursachen (*'spatial smoothing'*).

Der Nachweis von räumlichen Autokorrelationen in den Residuen ist relativ einfach durch Korrelogramme zu erbringen. Das Beheben ist jedoch schwieriger: Die Methodik zur Korrektur ist bei GLM sehr aufwändig und bei BRT und MAXENT nicht bekannt.

⁴⁰ Weitere Ausführungen und die Verweise auf die Originalzitate siehe in Dormann et al. (2007).

Kapitel 5

Ergebnisse

5.1 Untersuchungsgebiet

5.1.1 Prognosekarten

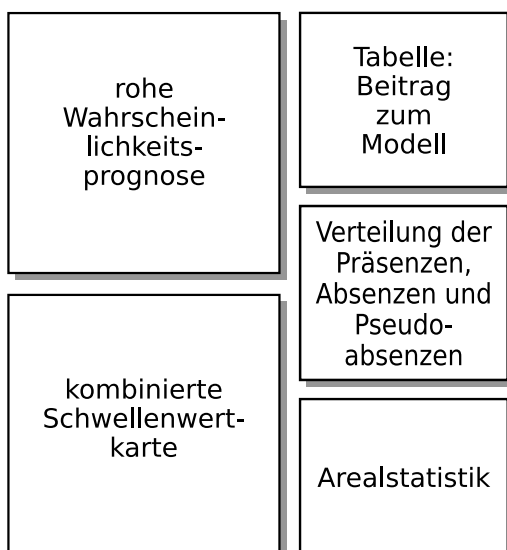
In den Darstellungen der rohen Wahrscheinlichkeitsprognosen wurde willkürlich der feste Schwellenwert 0.01 für die Unterscheidung zwischen Hintergrund (keine Farbe) und Vorhersage gewählt, mit Ausnahme der Prognosen mit BRT resultierten damit anschauliche Karten. Für jede Prognose wurden die Schwellenwerte 'Tiefster nachgewiesener Wert' (LPT), 'Spezifität gleich Sensitivität', '10%-Quantil' und 'Maximales κ ' (maxKappa) gerechnet und in einer kombinierten Schwellenwertkarte dargestellt. In der kombinierten Schwellenwertkarte wurden die Flächen der vier Schwellenwerte mit einem einheitlichen Farbcode belegt und überlagert. Dies ist deshalb möglich, weil die Flächen zu den Schwellenwerten sukzessive kleiner werden - also Teilmengen der Fläche A des jeweils tieferen Schwellenwertes sind: Beispielsweise

$$A_{LPT} \supseteq A_{Spez=Sens} \supseteq A_{10\%Quantil} \supseteq A_{MaxKappa} \quad (5.1)$$

Für jeden dieser Schwellenwerte wurden die Summen der mit Vorkommen von *T. biconicus* prognostizierten Areale in Hektaren berechnet. Zudem zeigen Boxplots die Werteverteilung der Prognose für Präsenzen, Absenzen und

Pseudoabsenzen. Bei einer optimalen Prognose ist die Verteilung der Präsenzen klar von derjenigen der Absenzen getrennt (Präsenzen mit höheren Werten und Absenzen mit tieferen). Die Pseudoabsenzen stellen den "Hintergrund" dar und sind daher in der Verteilung auf die Werte nahe 0 konzentriert. Die Schwellenwerte sind in diesen Boxplots als Linien dargestellt.

Die nachfolgenden Abbildungen 5.1 bis 5.8 sind nach folgendem Schema aufgebaut:



Anmerkungen

- Der Farbcode für die verschiedenen Schwellenwerte ist in allen Darstellungen gleich gehalten: 'tiefster nachgewiesener Fund' in türkis; '10%Quantil' in violett; 'Spezifität gleich Sensitivität' in orange und 'Maximales κ ' in blau.
- Die Lage der verwendeten Lokalnamen ist aus Abbildung 4.1 auf Seite 40 zu entnehmen.
- Die Abkürzungen der Variablen in 'Beitrag zum Modell' entsprechen den Definitionen in der Tabelle 4.1 auf Seite 50. Die Abkürzungen in der Arealstatistik stehen für die Schwellenwerte 'tiefster nachgewiesener Fund' (LPT), 'Spezifität gleich Sensitivität' (Spz=Sns), '10%-Quantil'(10%) und 'Maximales κ ' (MaxKappa).

Prognose mit MAXENT Level A

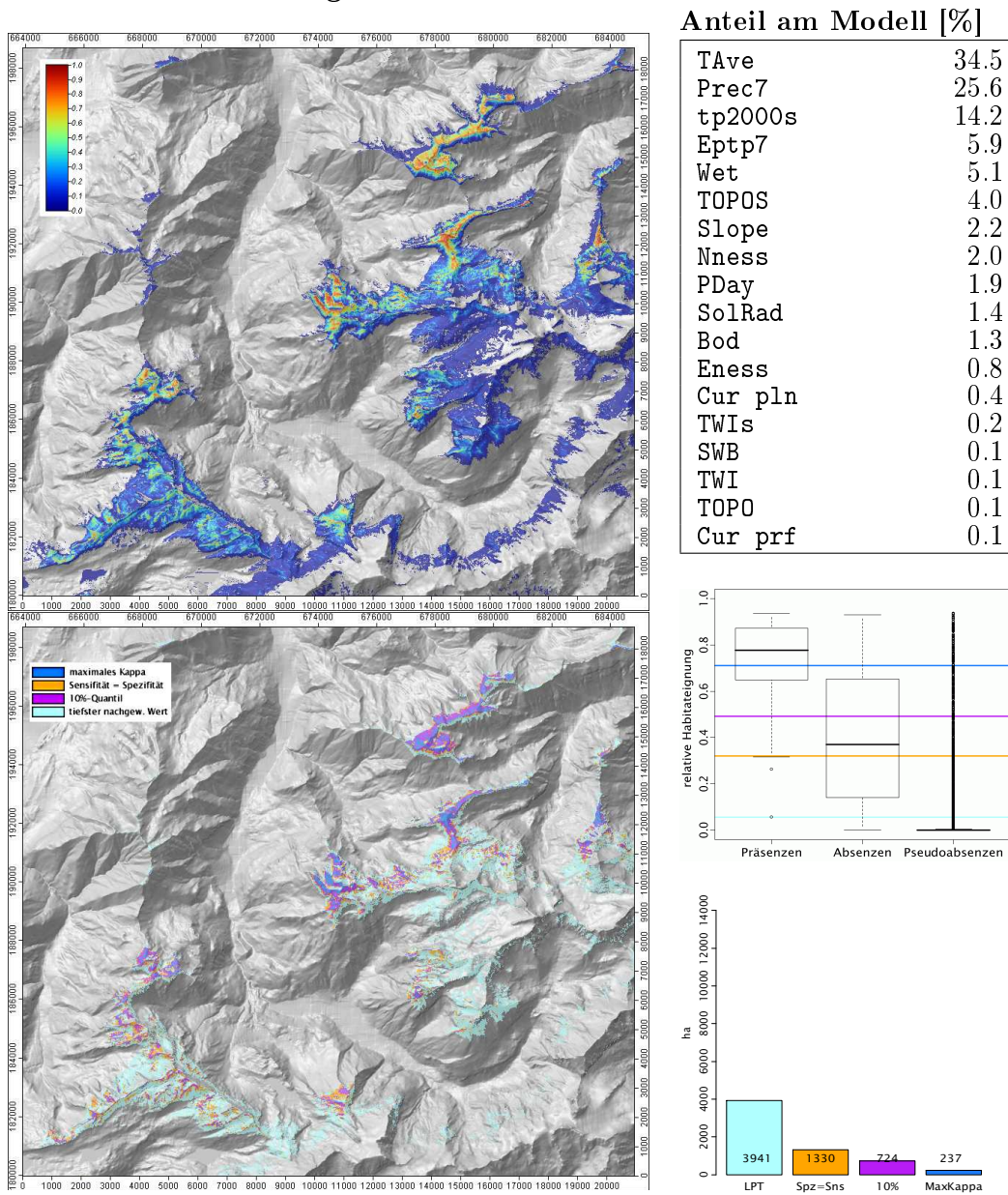


Abbildung 5.1: Prognosekarte mit MAXENT Level A: Die wichtigsten Gebiete *Chaiserstuel*, *Brisen*, *Walenstöcke*, *Schwalmis* und *Widderfeld* sowie bisher unbesuchte Gebiete am Nordgrat des *Uri-Rotstock* und *Hahnen* werden klar hervorgehoben. Etwas undeutlicher sind Gebiete auf dem *Chulm* und am *Laubersgrat* bezeichnet.

Prognose mit MAXENT Level B

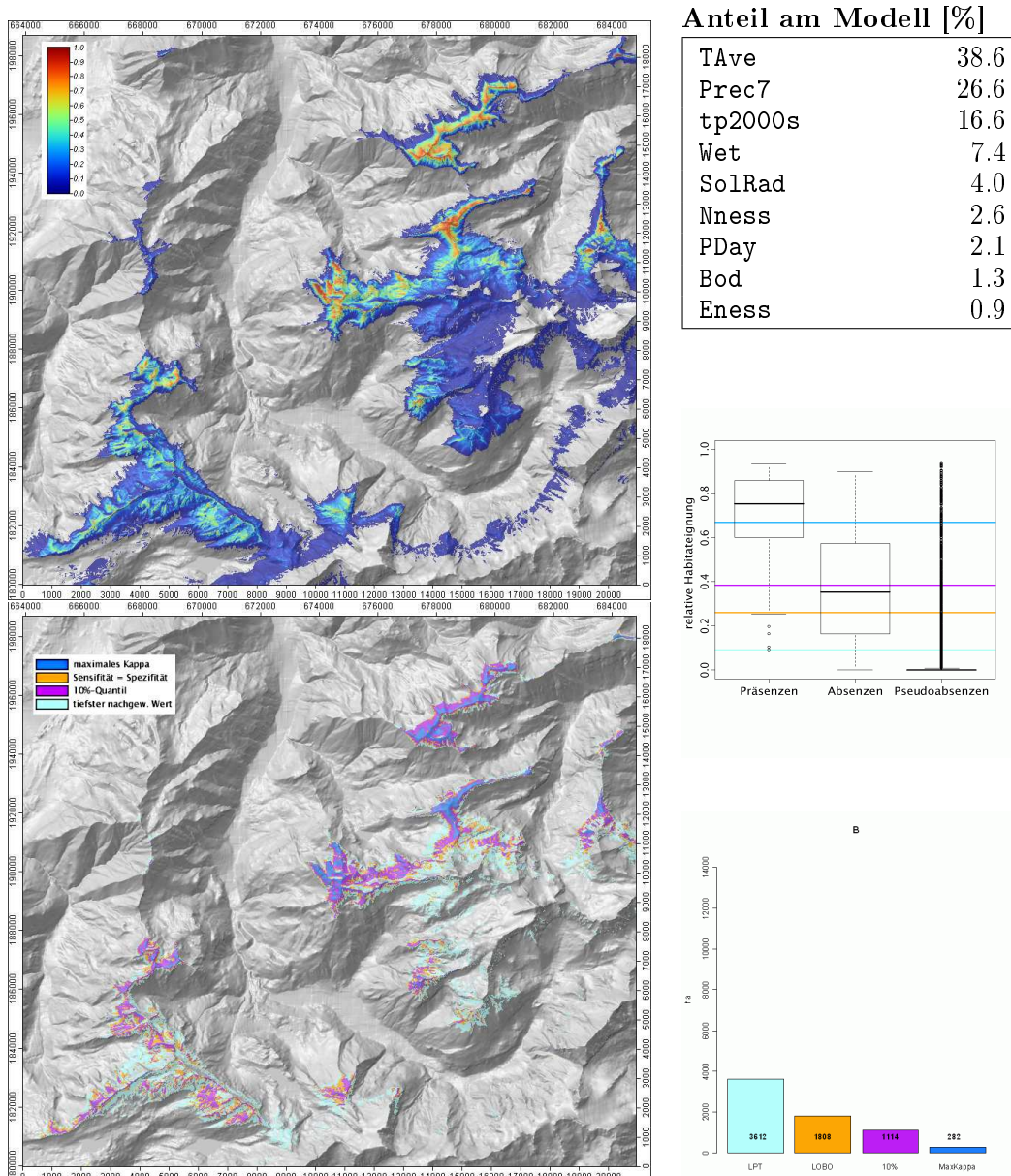


Abbildung 5.2: Prognosekarte mit MAXENT Level B: Im Vergleich zu Level A ergibt verschieben sich die prognostizierten Gebiete leicht nach Süden und nach oben (höhere topografische Höhe). Die Präsenzen und Absenzen sind auch nach der Reduktion auf 9 Umweltvariablen noch gut unterschieden (siehe Boxplot). Die Flächen mit den Schwellenwerten 'tiefster nachgewiesener Fund' und '10%-Quantil' sind etwas kleiner als bei Level A, hingegen ist die Fläche mit dem Schwellenwert 'Sensitivität = Spezifität' grösser.

Prognose mit MAXENT Level C

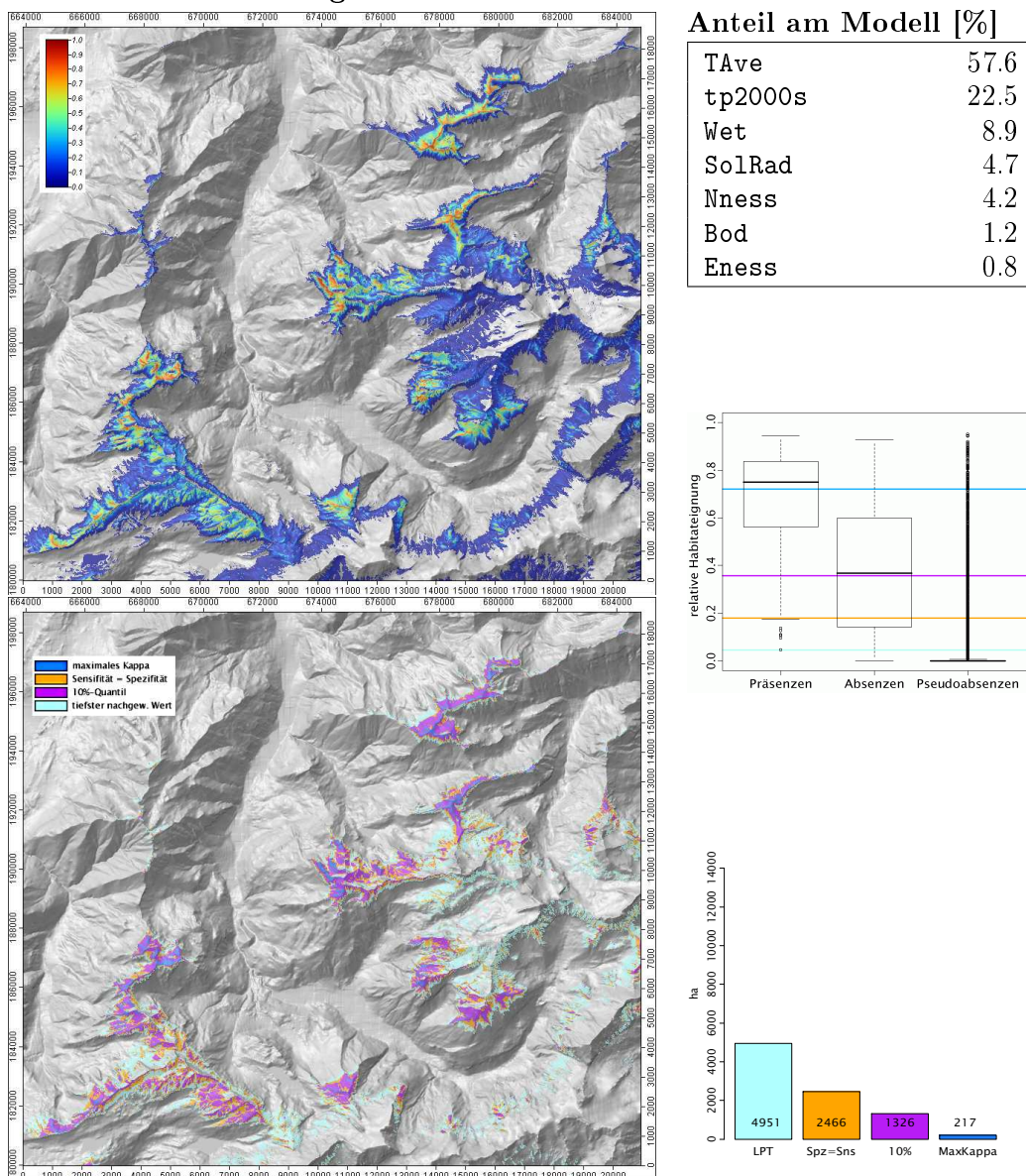


Abbildung 5.3: Prognosekarte mit MAXENT Level C: Insbesondere mit dem Schwellenwert 'tiefster nachgewiesener Fund' werden viele kleine Fragmente geeigneter Gebiete prognostiziert. Als ökologisch relevant sind aber nur grössere Gebiete anzusehen. Die Unterscheidung zwischen Präsenzen und Absenzen ist im Vergleich zu Level B etwas weniger deutlich. In der Arealstatistik wurden mit Ausnahme von 'Maximales κ ' alle Eignungsgebiete etwas grösser (siehe Boxplot).

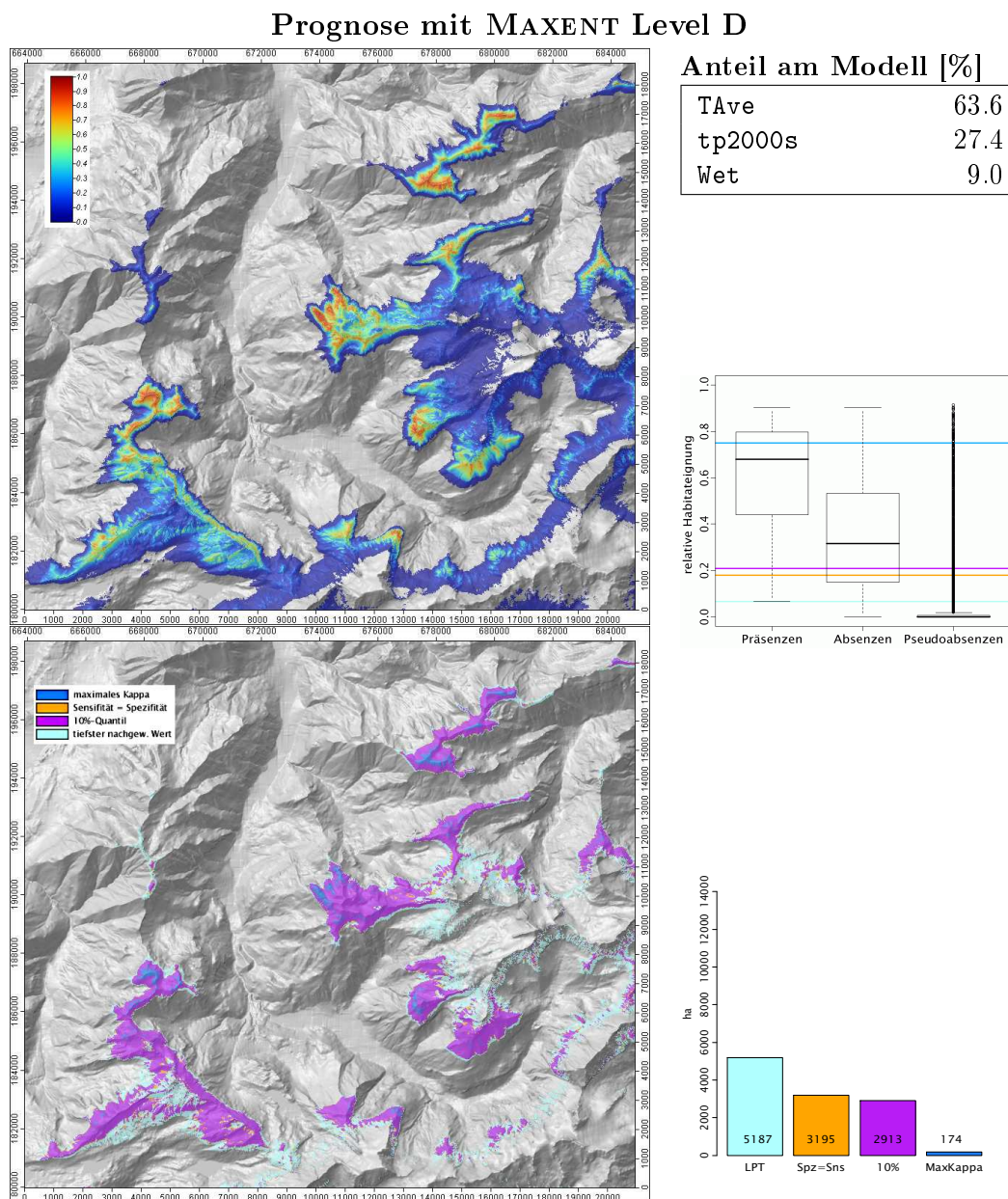
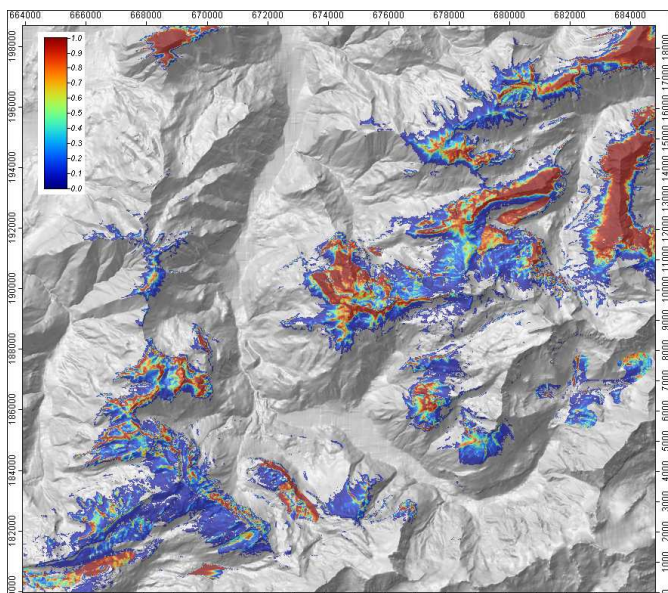


Abbildung 5.4: Prognosekarte mit MAXENT Level D: Die Reduktion der Anzahl Umweltvariablen ist ausgereizt: Die Prognoseschärfe und der Detailreichtum nehmen deutlich ab. Alle Schwellenwerte führen zu unglaublich grossen Verbreitungsgebieten, ausser 'Maximales κ '. Dieser Schwellenwert führt zu einem extrem kleinen Verbreitungsgebiet. Hier liegen sogar mehr als die Hälfte der Präsenzen ausserhalb des prognostizierten Verbreitungsgebietes. Dennoch wird das Verbreitungsgebiet durch MAXENT erstaunlich gut gefunden.

Prognose mit GLM Präsenz-/Absenzdaten

z-Wert und Signifikanz



(Intercept)	-2.81	**
Eness	-2.65	**
tp2000s	5.93	***
Prec7	5.58	***
Etpt7	2.17	*
SolRad	-2.08	*
Slope	3.04	**
TWIs	2.67	**
SWB	-3.00	**
TAve	2.77	**
(Etpt7) ²	-2.47	*
(SolRad) ²	2.46	*
(Slope) ²	-2.99	**
(TOP0s) ²	-2.61	**
(TWIs) ²	3.35	***
(SWB) ²	3.27	**
(TAve) ²	-2.74	**

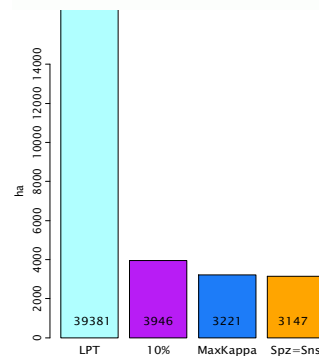
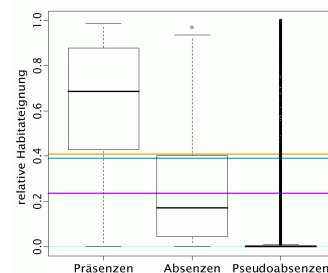
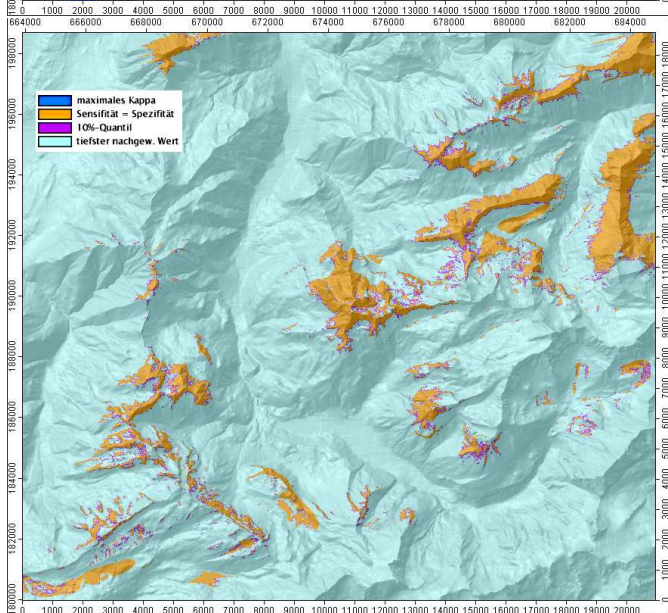


Abbildung 5.5: Prognosekarte mit GLM Präsenz-/Absenzdaten: Das Modell wird stark durch *Prec7* und *tp2000s* gesteuert, es sind aber viele weitere Variablen als signifikant eingestuft. Die Schwellenwerte 'Spezifität = Sensitivität' und 'Maximales κ ' unterscheiden gut zwischen Präsenzen und Absenzen (siehe Boxplot). Für grosse Gebiete im Norden und Osten (*Stanserhorn*, *Oberbauenstock* und *Uri-Rotstock* bis *Chulm*) ist eine hohe Fundwahrscheinlichkeit prognostiziert.

Prognose mit GLM Präsenz-/Pseudoabsenzdaten

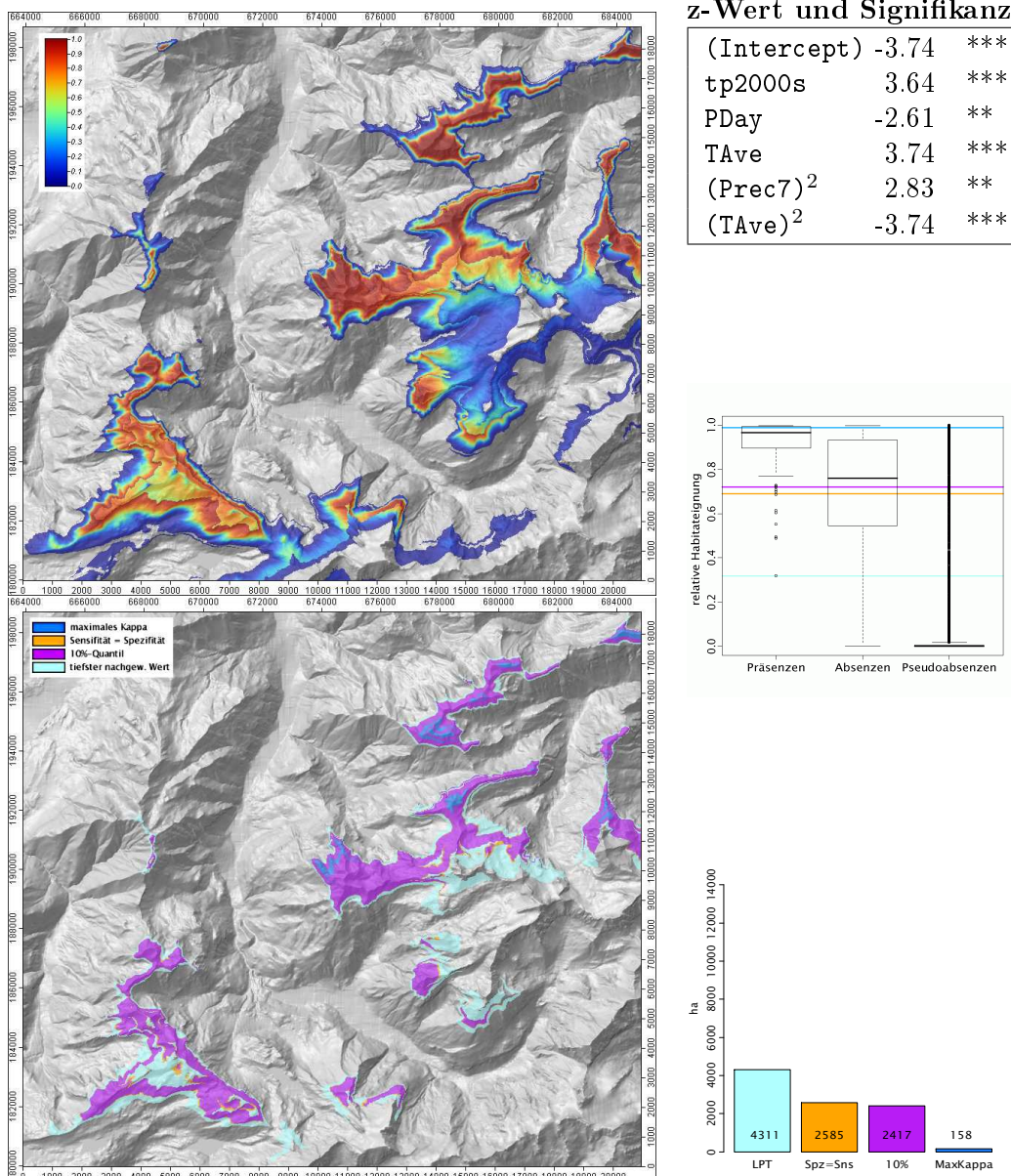


Abbildung 5.6: Prognosekarte mit GLM Präsenz-/Pseudoabsenzdaten: Nur wenige Variablen sind signifikant (TAve und tp2000s liefern wie in den MAXENT-Modellen einen wichtigen Beitrag). Auffallend ist das Gebiet um den *Oberbauenstock*, das bei den MAXENT-Modellen nur schwach prognostiziert wurde. Die intensiv abgesuchten Gebiete ohne Funde um den *Graustock* werden erstaunlich stark als wahrscheinliche Fundgebiete prognostiziert.

Prognose mit BRT Präsenz-/Absenzdaten

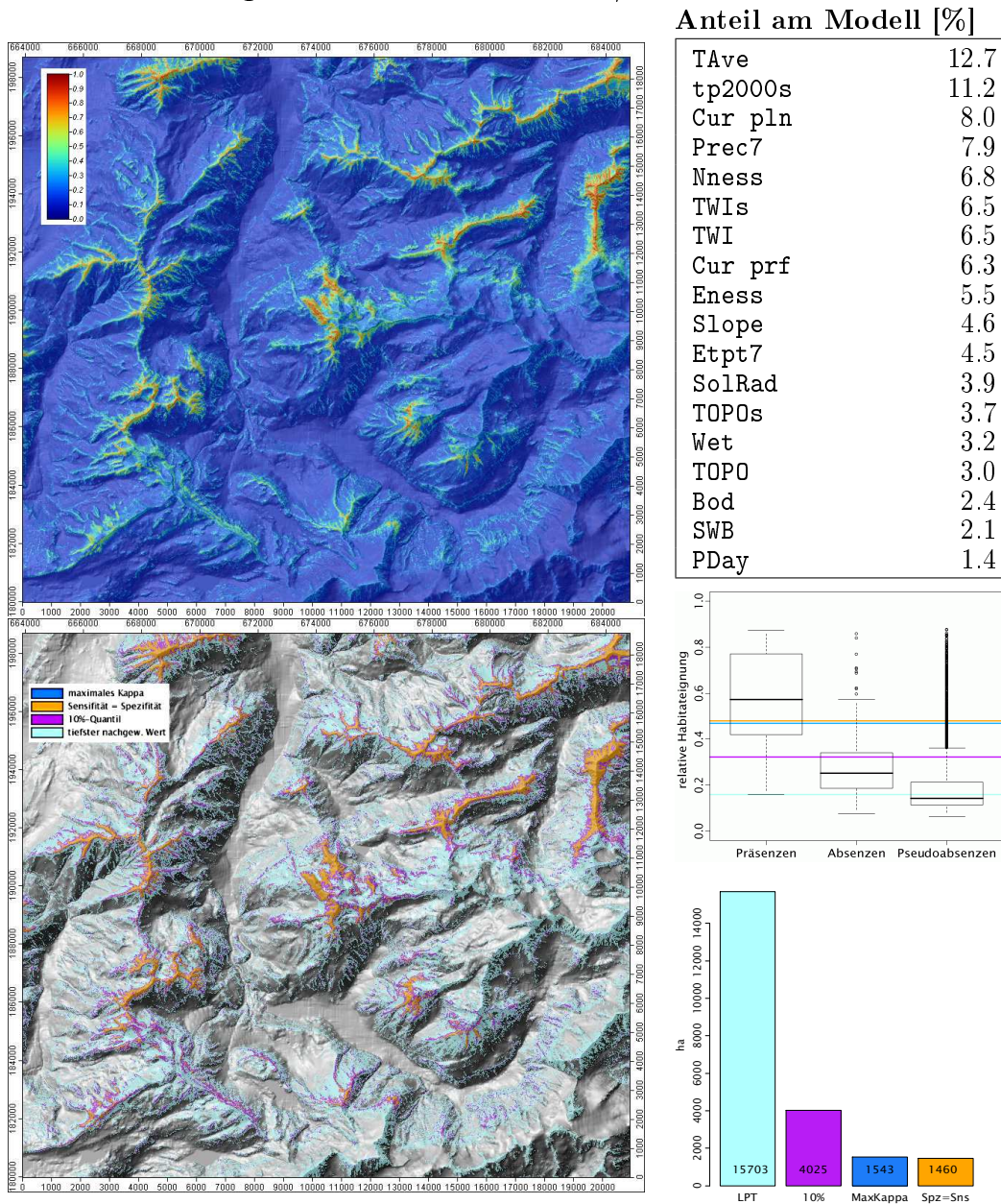


Abbildung 5.7: Prognosekarte mit BRT Präsenz-/Absenzdaten: Über das ganze Untersuchungsgebiet wird eine relativ hohe Rest-Fundwahrscheinlichkeit prognostiziert, bei gleichzeitig recht guter Trennung von Präsenzen und Absenzen. Viele Variablen tragen ähnlich stark zum Modell bei. Die Prognosegebiete wirken sehr fragmentiert (fast alle Grate und Gipfel der mittleren Höhen) und der Schwellenwert mit dem 'tiefsten nachgewiesenen Fund' verursacht ein viel zu grosses Gebiet.

Prognose mit BRT Präsenz-/Pseudoabsenzdaten

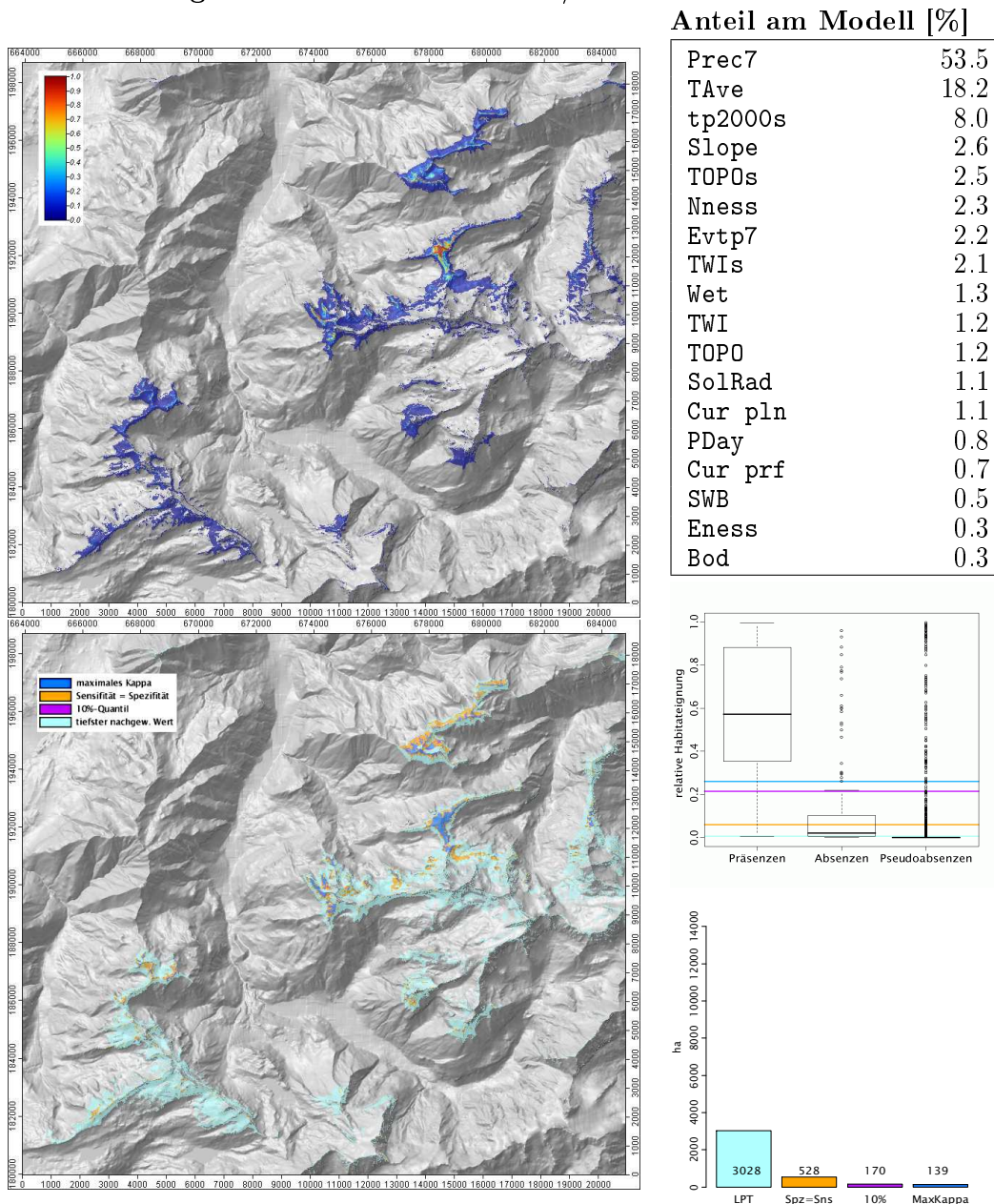


Abbildung 5.8: Prognosekarte mit BRT Präsenz-/Pseudoabsenzdaten: Hauptsächlich durch Prec7 und untergeordnet durch TAve und tp2000s gesteuert, werden nur sehr kleine Gebiete als Habitate prognostiziert. Möglicherweise ist das Modell überangepasst. Der Chaisertuel als wichtigstes Fundgebiet (bezüglich Funddichte) wird stark betont. Die Unterscheidung in Präsenzen und Absenzen ist sehr deutlich, obwohl der Schwellenwert 'tiefster nachgewiesener Fund' einen sehr tiefen Wert aufweist.

Die Prognosen der verschiedenen Levels von MAXENT sehen sich grundsätzlich sehr ähnlich (Abbildungen 5.1 bis 5.4). Erst bei Level D - dem Modell, das nur mit den drei wichtigsten Umweltvariablen gerechnet wurde - wird der mangelnde Informationsgehalt infolge der geringen Anzahl Umweltparameter allmählich sichtbar.

Auffällig an der Prognose von GLM mit Präsenz-/Absenzdaten (Abbildung 5.5) ist die starke Betonung der Grate im Nordosten und Osten des Untersuchungsgebietes. Der Schwellenwert 'tiefster nachgewiesener Fund' versagt bei dieser Prognose deutlich, da einer der Funde in einer Rasterzelle mit Vorkommenswahrscheinlichkeit 0 lag. Dadurch wurde das ganze Gebiet als mögliches Verbreitungsgebiet klassiert. Im GLM-Modell mit den Präsenz-/Pseudoabsenzdaten (Abbildung 5.6) zeigen sich auffällige Sprünge oder Diskontinuitäten in ansonsten sehr kontinuierlichen Gradienten. Die Verteilung der rohen Wahrscheinlichkeiten weist allgemein zu hohe Werte auf. Die Unterscheidung zwischen Präsenzen und Absenzen ist nur mässig, so dass die Schwellenwerte kaum eine gute Trennung finden. Dies kann als Hinweis auf die beschränkte Verwendbarkeit der Absenzen gesehen werden.

Die Prognose mit BRT anhand der Präsenz-/Absenzdaten (Abbildung 5.7) kann den Hintergrundwert für das Nicht-Vorkommen nicht richtig finden, d.h. es wird eine geringe Vorkommenswahrscheinlichkeit über das ganze Untersuchungsgebiet prognostiziert. Selbst nach Anwendung der Schwellenwerte wirkt die Prognose etwas wild, mit vielen kleinräumigen Wechseln. Die Rohprognose von BRT mit dem Präsenz-/Pseudoabsenzdatensatz (Abbildung 5.8) hat hingegen ein sehr stark gegen tiefe Werte verschobenes Mittelfeld der Werteverteilung (vgl. die Boxplots zu dieser Prognose: Der Durchschnitt der Absenzen liegt nur wenig über 0). Die Schwellenwerte sind zwar ebenfalls recht tief, führen aber im Vergleich zu den andern Prognosen dennoch zu kleinen Werten in der Arealstatistik.

Auffallend ist der Wechsel in der Rangfolge der Schwellenwerte bei den GLM- und BRT-Prognosen mit Präsenz-/Absenzdaten, wo der Schwellenwert 'Sensitivität gleich Spezifität' zur kleinsten prognostizierten Fläche führt (Abbildungen 5.5 und 5.7).

AUC	Maxent				GLM		BRT	
	A	B	C	D	PA	PPsA	PA	PPsA
Prognose	0.994	0.990	0.987	0.985	0.884	0.979	0.965	0.999
Validierung	0.990	0.984	0.979	0.976	0.841	0.977	0.752	0.999
	± 0.015	± 0.004	± 0.009	± 0.008	± 0.063	± 0.006	± 0.084	± 0.001

Tabelle 5.1: AUC-Werte für die Prognosemodelle mit allen Funden und die durchschnittlichen AUC-Werte der Validierungen (MAXENT: 10× mit 75% des Präsenzdatensatzes; GLM und BRT mit 10× Kreuzvalidierung).

PA: Präsenz-/Absenzdatensatz; PPsA: Präsenz-/Pseudoabsenzdatensatz.

5.1.2 AUC

In Tabelle 5.1 sind die AUC-Werte aller Modelle angegeben. Auffallend sind die allgemein hohen Werte von meist deutlich über 0.9. Gemäss der Einteilung von Swets (1988) sind AUC-Werte über 0.9 als „*excellent*“, Werte zwischen 0.8 und 0.9 als „*good*“, solche über 0.7 noch als „*useful*“ und solche darunter als „*poor*“ zu bezeichnen¹. Somit fällt lediglich das GLM-Modell mit den Präsenz-/Absenzdaten etwas ab. Das stark tiefere Mittel der Kreuzvalidierungen des BRT-Präsenz-/Absenzdaten-Modells weist auf eine mögliche Überanpassung hin, sicher aber auf eine sehr schlechte Fähigkeit, neue Funde zu integrieren. Die MAXENT-Werte fallen beim sukzessiven Weglassen von erklärenden Variablen nur leicht ab. Selbst in Level D mit nur mehr 3 Umweltvariablen erreicht MAXENT ein sehr gutes Resultat. Der höchste Wert wird mit dem BRT-Präsenz-/Pseudoabsenzdaten-Modell erreicht. Wenn die Überlegungen zur Prognosekarte (Abschnitt 5.1.1 und Abbildung 5.8) mit einbezogen werden, ist hier der Verdacht auf eine Überanpassung angebracht. Mit zukünftigen Funddaten wird diese Vermutung überprüft werden können.

Die oft sehr kleinen Unterschiede über etwa 0.97 dürfen aber nicht zu stark gewichtet und daraus vorschnell eine Rangfolge der Modelle abgeleitet werden. Aufgrund der AUC-Werte kann also kein verlässlicher Qualitätsunterschied zwischen den erstellten Prognosemodellen festgestellt werden. Elith

¹ Eine vergleichbare Einteilung schlugen auch Hosmer & Lemeshow (2000) vor. Sie bezeichneten AUC-Werte ab 0.7 als „*acceptable*“, solche ab 0.8 als „*excellent*“ und solche ab 0.9 als „*outstanding*“.

et al. (2006) verglichen die verschiedenen Methoden bezüglich ihrer Neigung zu besseren AUC-Werten. Dabei schnitt MAXENT erfolgreicher als GLM und BRT etwas besser als MAXENT ab. Zwar waren in dieser Untersuchung die durchschnittlichen AUC-Werte allgemein tiefer als jene in Tabelle 5.1, bei einzelnen spezialisierten Arten wurden aber ebenfalls ähnlich hohe Werte erzielt. Somit können die Ergebnisse von Elith et al. (2006) - bei aller Vorsicht solcher Vergleiche - nachvollzogen bzw. bestätigt werden.

5.1.3 Unsicherheitskarten

Standardabweichungskarten

Für den Vergleich von rohen Prognosekarten wurde die Standardabweichungskarte eingeführt (Abschnitt 3.5.1). Beispielhaft werden hier zwei Karten angeführt: Der Vergleich von 10 Testläufen (Teildatensätze) mit Level A und der Vergleich der vier Prognosekarten aller Levels, welche mit dem vollen Beobachtungsdatensatz erstellt wurden. Abbildung 5.9 zeigt die Standard-

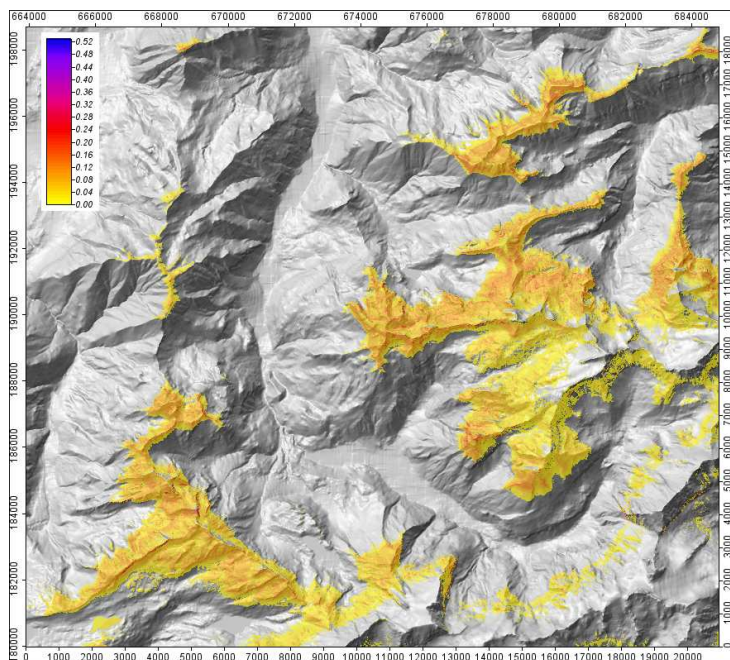


Abbildung 5.9: Standardabweichungskarte aus 10 Testläufen des Level A mit je 75% der Präsenzdaten. Die Resultate sind sehr ähnlich: Die durchschnittliche Standardabweichung aller Zellen beträgt nur 0.0092 oder 1.8% der maximal möglichen Abweichung. Die grössten Abweichungen treten entlang von Graten und Felswänden auf. Werte unter 0.01 sind transparent.

abweichungskarte für die 10 Testläufe von Level A mit MAXENT unter Verwendung von 75% der Präsenzen. Die Unterschiede sind erwartungsgemäss eher in den Gebieten mit höheren Prognosewerten zu finden. Sie sind jedoch allgemein sehr klein. Die Werte der Standardabweichungskarten sind, wie in Abschnitt 3.5.1 ausgeführt, vor allem qualitativ interpretierbar: Die höchste erreichte Standardabweichung mit einem Wert von 0.407 beträgt etwa 77% der theoretisch maximal möglichen Abweichung.

Die räumliche Verteilung der Unsicherheiten erweist sich als recht uneinheitlich. Dies kann als Indiz für eine konsistente Prognose gewertet werden.

Die Standardabweichungskarten für die Testläufe der anderen drei Levels sehen ähnlich aus. Ein Trend lässt sich bei den von A nach D sinkenden Durchschnittswerten und Standardabweichungen ablesen. Aus den maximalen Werten, welche bei Level B ein Minimum von 0.352 (oder 67% des theoretischen Maximums) erreichen, ist jedoch kein Trend ersichtlich (Tabelle 5.2).

Testläufe für Level	Durchschnitt	Standardabweichung	höchster Wert	% des theor. Maximums
10 × A	0.00926	0.0242	0.4067	77%
10 × B	0.00807	0.0217	0.3516	67%
10 × C	0.00832	0.0209	0.3660	69%
10 × D	0.00662	0.0166	0.4764	90%
Theoretischer Maximalwert der Standardabweichung bei 10 Vergleichen: 0.5270				

Tabelle 5.2: Kennwerte der Standardabweichungskarten zu den Testläufen für die Level A bis D mit MAXENT. 'Durchschnitt', 'Standardabweichung' und 'Maximum' beziehen sich auf alle Rasterzellen der Standardabweichungskarte zum angegebenen Level.

Beim Vergleich der vier Prognosen aus MAXENT für Level A bis D ergibt sich ein etwas deutlicheres Bild (Abbildung 5.10). Es lassen sich auch Bereiche identifizieren in denen die Unsicherheit höher ist. Dort soll bei einem nächsten Feldbesuch prioritär nach *T. biconicus* gesucht werden. Unglücklicherweise – aber nicht überraschend – liegen diese Bereiche oft in den steilen Felswänden oder an schwer zugänglichen Stellen! Solche Gebiete wurden zum Beispiel

am Nordost-Ausläufer des *Titlis*, im Bereich *Huetstock* und beim *Hahnen* identifiziert.

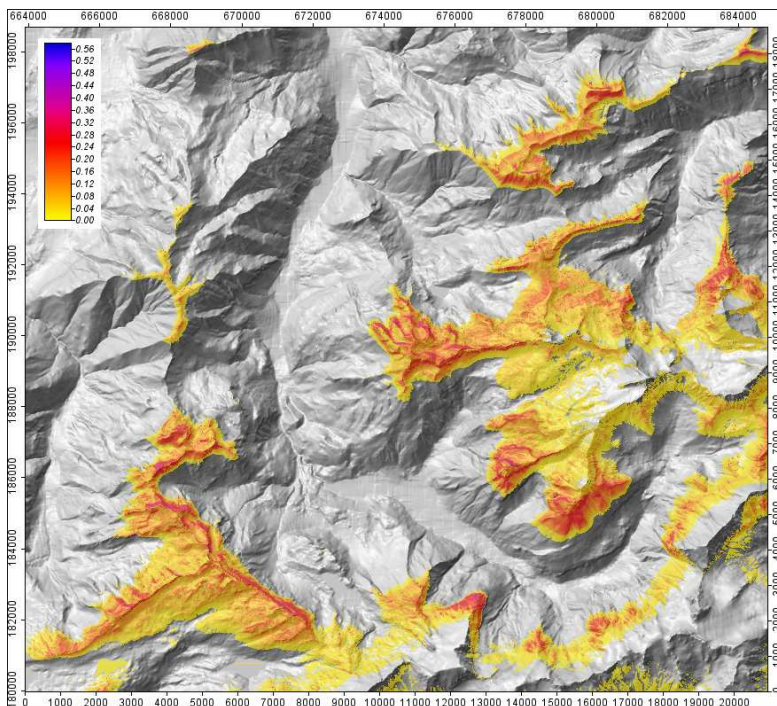


Abbildung 5.10: Die Standardabweichung der vier mit MAXENT produzierten, rohen Habitatprognosen der Level A bis Level D. Die Unsicherheiten sind grösser als in Abbildung 5.9 und vor allem in steilen Felswänden am höchsten. Werte unter 0.01 sind transparent.

Schwellenwertkarten: Unsicherheits- und Einigkeitskarten

Die Umwandlung von rohen Wahrscheinlichkeitsprognosen zu Schwellenwertkarten verspricht einen konsistenteren Vergleich (Abschnitt 3.5.2 Seite 35). Es wurden zwei Darstellungsvarianten eingeführt: Die Unsicherheitskarte mit der Aussage, wo Unterschiede zwischen den Modellen vorliegen, und die Einigkeitskarte, die aufzeigt, wo sich die Prognosen einig sind und wo sie sich widersprechen. Als Beispiel werden in Abbildung 5.11 und 5.12 die Unsicherheits- und Einigkeitskarte gezeigt, welche den Vergleich der vier MAXENT-Level mit dem Schwellenwert 'Sensitivität gleich Spezifität' darstellen. Dieser Schwellenwert wurde gewählt, weil er für die MAXENT-Prognosen einen guten Kompromiss zwischen der Anzahl Fehler beim Erklären der Beobachtungen und dem Ausschluss von 'Ausreissern' bildet.

Unsicherheiten, die sich nur in einzelnen, isolierten Rasterzellen äussern, dürfen nicht zu stark gewichtet werden. Hingegen geben grössere Gebiete mit

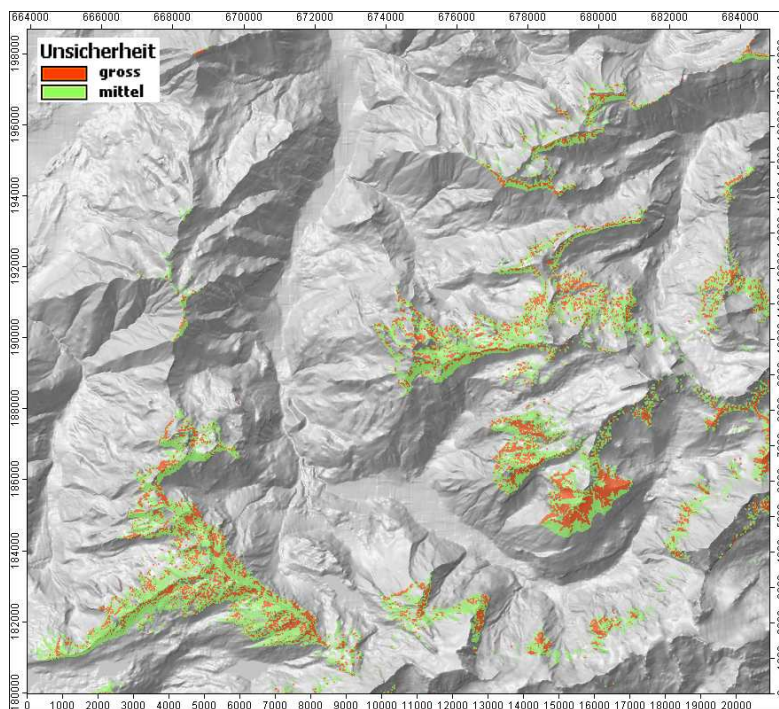


Abbildung 5.11: Die Unsicherheitskarte für den Schwellenwert 'Sensitivität gleich Spezifität', gerechnet mit MAXENT für die Level A bis D. Die auffälligste Unsicherheit befindet sich am *Wissberg*. Dieses Gebiet sollte prioritär besucht werden.

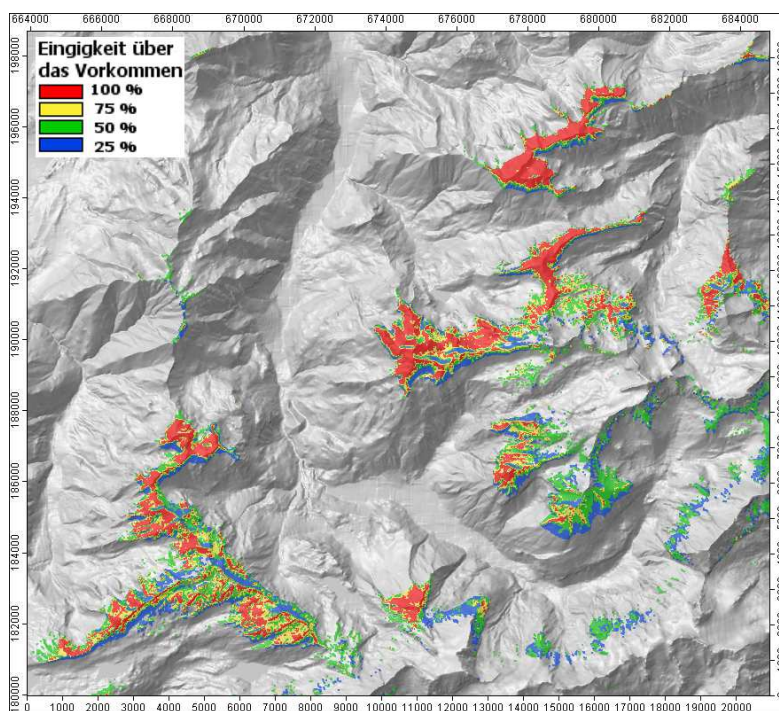


Abbildung 5.12: Eignigkeitskarte für den Schwellenwert 'Sensitivität gleich Spezifität', gerechnet mit MAXENT für die Level A bis D. Auffallend ist das Gebiet beim *Laubersgrat*, wo sich alle Prognosen einig sind.

erhöhter Unsicherheit einen Hinweis auf zusätzlich notwendige Beobachtungen. Die Unsicherheitskarte in Abbildung 5.11 zeigt um alle prognostizierten Gebiete zahlreiche, stark fragmentierte grössere Unsicherheiten. Zusammenhängende Gebiete grösserer Unsicherheit sind fast ausschliesslich auf den bisher nicht abgesuchten *Wissberg* beschränkt.

Die Einigkeitskarte in Abbildung 5.12 weist um die wichtigsten Fundstellen *Brisen*, *Chaiserstuel*, *Walenstöcke* und *Widderfeld* eine grosse Einigkeit auf. Für andere Gebiete hingegen zeigt die Karte eine oft kleinräumig wechselnde Einigkeit: Zum Beispiel in den Gebieten *Graustock - Barglen - Frutt*, *Hahnen*, *Wissberg* und die Gebiete nördlich des *Ruchstocks*. Auffallend ist die grosse Einigkeit beim *Laubersgrat* und am Nordgrat des *Uri-Rotstocks*.

Der Vergleich von binären Schwellenwertkarten ist auch über verschiedene Methoden hinweg zulässig. Es wurde eine Kombination gewählt, die aus jeder Methode je zwei Vertreter verwendet: Von MAXENT Level B und Level C, von GLM und BRT die jeweiligen Berechnungen mit dem Präsenz-/Absenzdatensatz und dem Präsenz-/Pseudoabsenzdatensatz. Diese Kombination vereint die Informationen aller Methoden bzw. deren Prognosen und kann daher als Grundlage für die Empfehlungen zur weiteren Planung der Feldarbeiten dienen.

Verbesserung der Prognosen Zur Verbesserung der Prognosen sind Beobachtungen aus Gebieten mit grossen Unterschieden zwischen den Prognosen interessant. Allerdings sind Absenzbeobachtungen aus den Modellen, die sich auf den Präsenz-/Pseudoabsenzdatensatz stützen, nicht direkt verwendbar.

Gebiete mit hoher Priorität sind (siehe Abbildung 5.13): *Laubersgrat* und *Hohmad* bis *Bonistock* (zwischen *Barglen* und *Frutt*). In mittlerer Priorität sind: Das Gebiet nördlich des *Hahnen* (bis *Gemsispiel*), *Wissberg*, *Huetstock*, *Uri-Rotstock* bis *Chulm* sowie *Schwalmis* bis *Oberbauenstock*.

Neue Fundgebiete Ein weiteres Ziel der Modellierung war das Aufzeigen neuer, potentieller Fundgebiete. Es zeigte sich aber auch, dass einige bekannte Fundgebiete weiter untersucht werden sollten (siehe Ab-

bildung 5.14). Dabei handelt es sich um die Gebiete vom *Widderfeld* gegen Osten (*Bocki-Stock*), am *Schwalmis* gegen Norden und auf den *Walenstöcken*.

Innerhalb des Untersuchungsgebietes sind einige bisher nicht abgesuchte Gebiete lohnend: Am *Hahnen*, der Grat vom *Uri-Rotstock* bis zum *Chulm* und der *Oberbauenstock*.

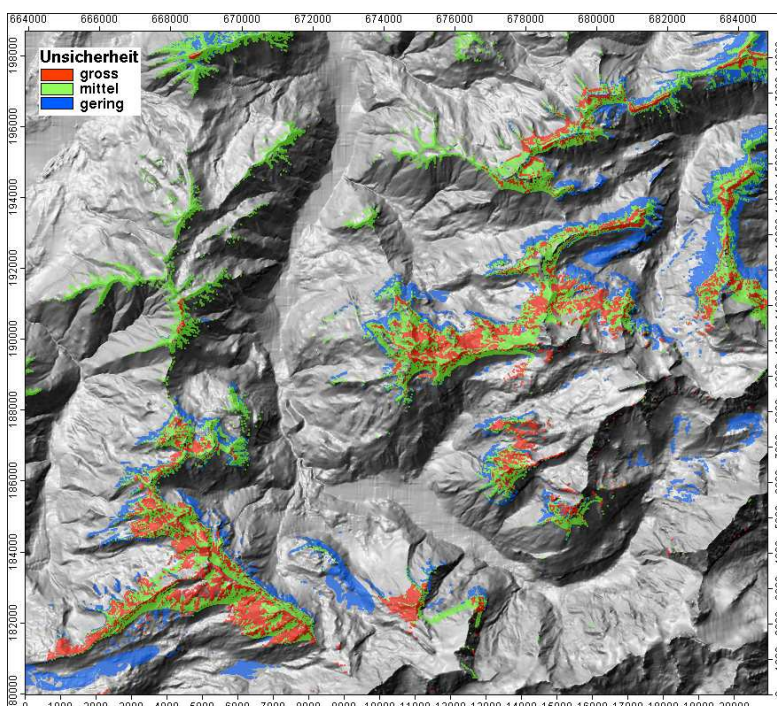


Abbildung 5.13: Zusammenstellung der Resultate aus den Methoden MAXENT, GLM und BRT als Unsicherheitskarte mit dem Schwellenwert 'Sensitivität gleich Spezifität'.

5.2 Extrapolation

5.2.1 Extrapolation in andere Gebiete

Die Prognosen aus MAXENT für die Level B und C (Abbildung 5.15) wurden auf ein grösseres Gebiet hochgerechnet, das fast die ganze Schweiz umfasst. In Abbildung 5.15 ist aus Übersichtlichkeitsgründen nur die weitere Region des Untersuchungsgebietes dargestellt. Die Bodenbedeckung musste allerdings als Parameter weggelassen werden, da sie nicht über das ganze Gebiet zur Verfügung stand.

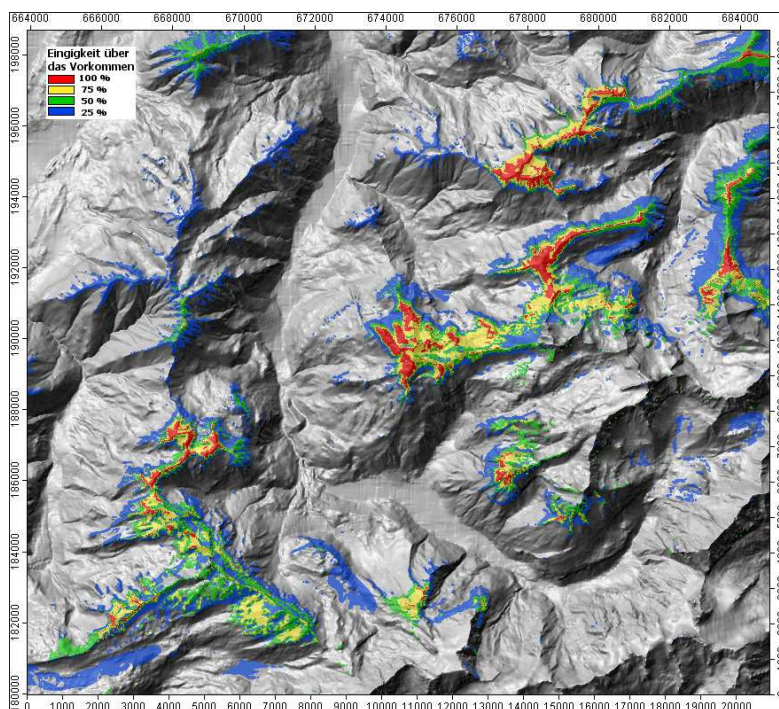


Abbildung 5.14:
Zusammenstellung
der Resultate aus
den Methoden
MAXENT, GLM
und BRT als
Einigkeitskarte mit
dem Schwellenwert
'Sensitivität gleich
Spezifität'.

Die grossen Unterschiede der Extrapolation aus Level B zu der aus Level C erstaunen, denn die Prognosen innerhalb des Untersuchungsgebietes sind sich sehr ähnlich. Im Gegensatz zu Level C wurden in Level B der Juli-Niederschlag und die Regentage während der Wachstumsperiode als Umweltparameter berücksichtigt. Diese zwei inhaltlich sehr ähnlichen Umweltvariablen zeigen vor allem im Gebiet des hinteren *Muotatals* und in den *Berner Alpen*² eine ausgeprägte Anomalie von zu hohen Werten (im hinteren *Muotatal* bis über 230 mm Niederschlag im Juli und bis 51 Regentage). Solch hohe Werte werden im Untersuchungsgebiet nicht erreicht und deren Auswirkungen auf die Habitateignung muss vom Modell geschätzt werden. Im Gebiet des hinteren *Muotatals* und des *Klausenpasses* wird die Habitateignung durch den exponentiellen Charakter des Modells stark überzeichnet. Die *Berner Alpen* sind topografisch deutlich höher und fallen wohl deshalb für die Verbreitung ausser Betracht.

Die durch MAXENT ermittelten Abhängigkeit des Vorkommens von *T. biconicus* von diesen beiden Umweltvariablen ist in Abbildung 5.16 aufge-

² Die *Berner Alpen* sind nur teilweise am unteren Rand der Abbildung 5.18 sichtbar.

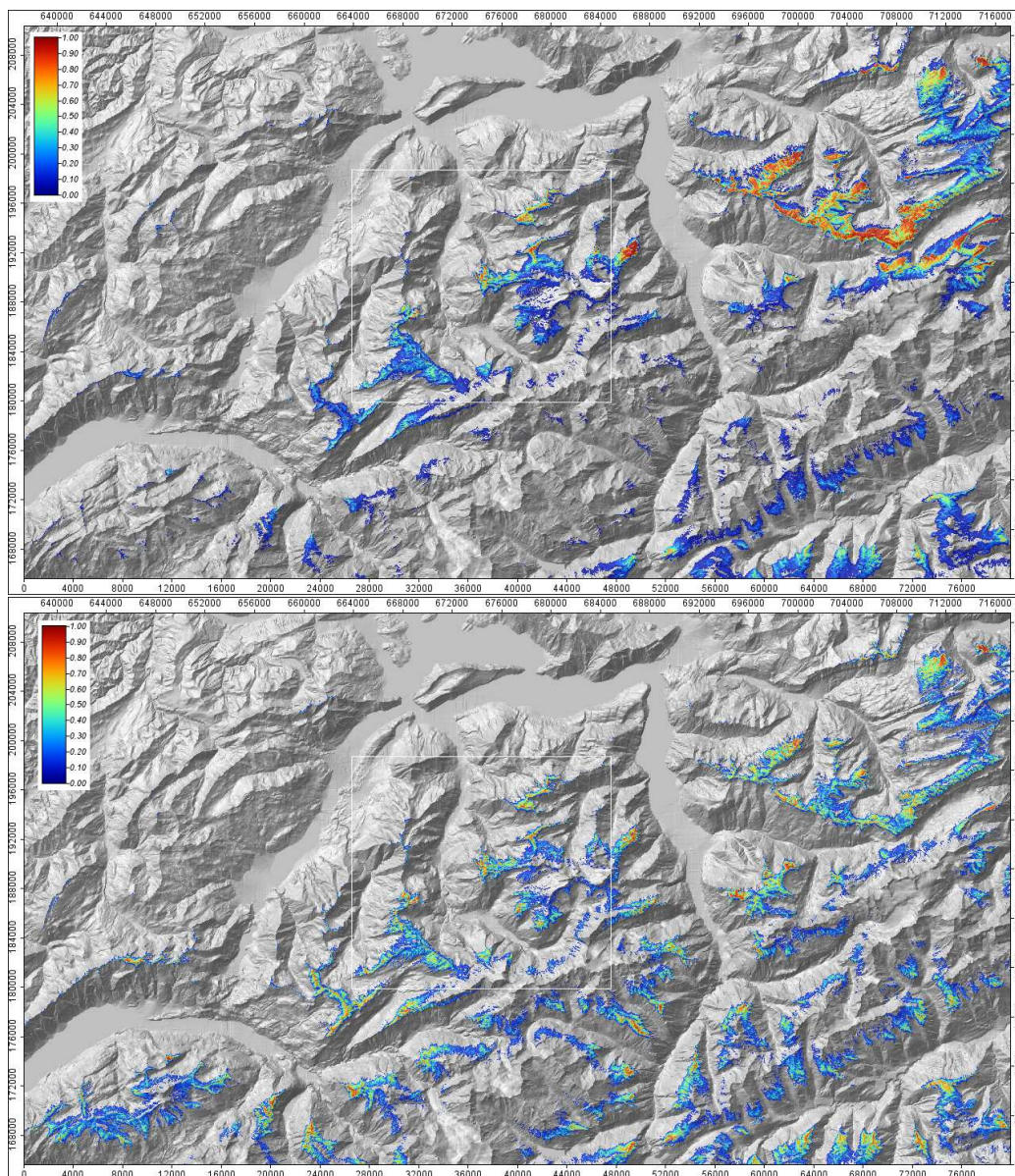


Abbildung 5.15: Die räumliche Extrapolation der Modelle aus MAXENT Level B (oben) und Level C (unten) auf ein grösseres Gebiet. Das Untersuchungsgebiet ist mit einem weissen Rahmen gekennzeichnet. Ausdehnung und Koordinaten sind am Rand angegeben. Obwohl sich die Prognosen im Untersuchungsgebiet nur geringfügig unterscheiden, weisen die räumlichen Extrapolationen in die umliegenden Gebiete deutliche Unterschiede auf.

zeigt: Interessanterweise korreliert das Vorkommen von *T. biconicus* positiv mit dem Juli-Niederschlag, zu viele Regentage während der ganzen Wachstumsperiode wirken sich aber negativ auf ihr Vorkommen aus. Die durch die Anomalien der Umweltvariablen erzeugten 'Spannungen' sind in Abbildung 5.17 dargestellt. In der Extrapolation für Level C sind offensichtlich

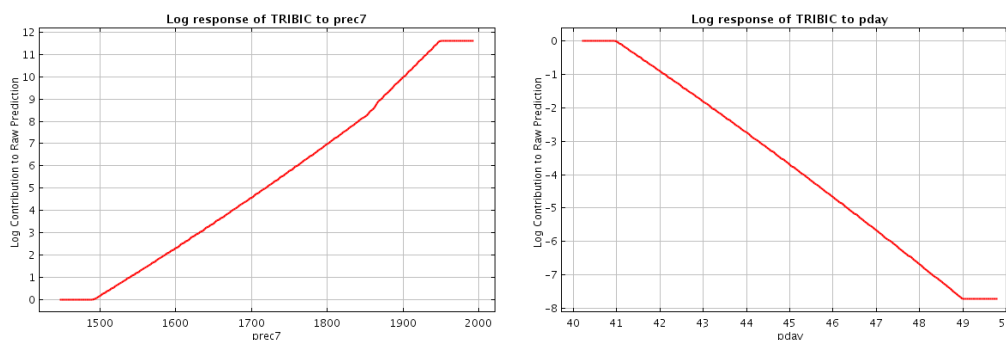


Abbildung 5.16: Die in MAXENT im Level B ermittelten Abhängigkeiten vom Juli-Niederschlag (links) und von der Anzahl Regentage in der Wachstumsperiode (rechts). Bei Extrapolationen mit Werten ausserhalb des im Untersuchungsgebiet angetroffenen Bereiches wird der am nächsten liegende Punkt der Kurve verwendet (konstante Linienenden beim Maximum und beim Minimum der Kurve).

im Untersuchungsgebiet besser abgedeckte Umweltvariablen enthalten, beziehungsweise nur solche, die keine grossen Anomalien aufweisen.

Das Verteilungsmuster für das Vorkommen von *T. biconicus* in der Extrapolation mit Level C zeigt eine sehr starke Abhängigkeit von einem Höhenband (repräsentiert durch die Jahresdurchschnittstemperatur) und von der topografischen Position (Umweltvariable `tp2000s`). Ob dies auch tatsächlich den ökologisch geeigneten Gebieten entspricht, ist zumindest mit Vorsicht zu interpretieren. Die durch 'Spannungen' verursachten Unsicherheiten sind jedoch hier deutlich geringer als diejenigen in der Extrapolation aus Level B.

Zu den in Abschnitt 5.1.3 aufgezeigten Gebieten, welche in zukünftigen Feldarbeiten prioritär bearbeitet werden sollten, kommen durch die räumliche Extrapolation einige dazu (die Lage der genannten Gebiete ist aus Abbildungen 5.18 ersichtlich):

- Östlich angrenzend an das Untersuchungsgebiet sind sehr hoch prognostizierte Habitateignungen am *Gitschen* und den *Sunnigen Stöck*

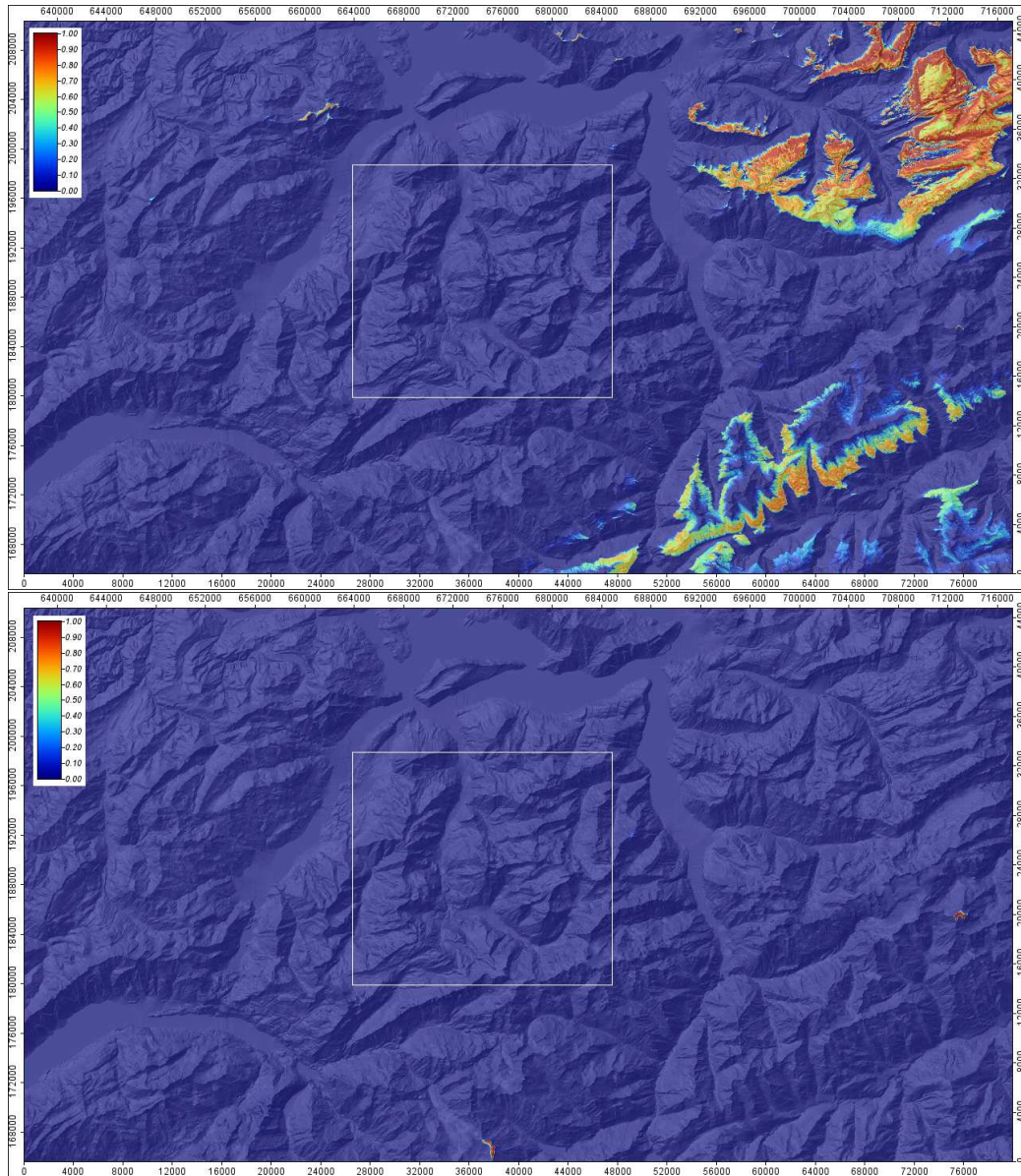


Abbildung 5.17: Die Extrapolation in Gebiete mit deutlich anderen Umweltbedingungen ist problematisch und fehleranfällig. Je weiter ausserhalb des aus dem Untersuchungsgebiet bekannten Wertebereiches eine Prognose gemacht werden muss, desto stärker wird ein Modell strapaziert. Diese als 'Spannungen' bezeichneten Effekte sind für Level B (oben) und Level C (unten) dargestellt.

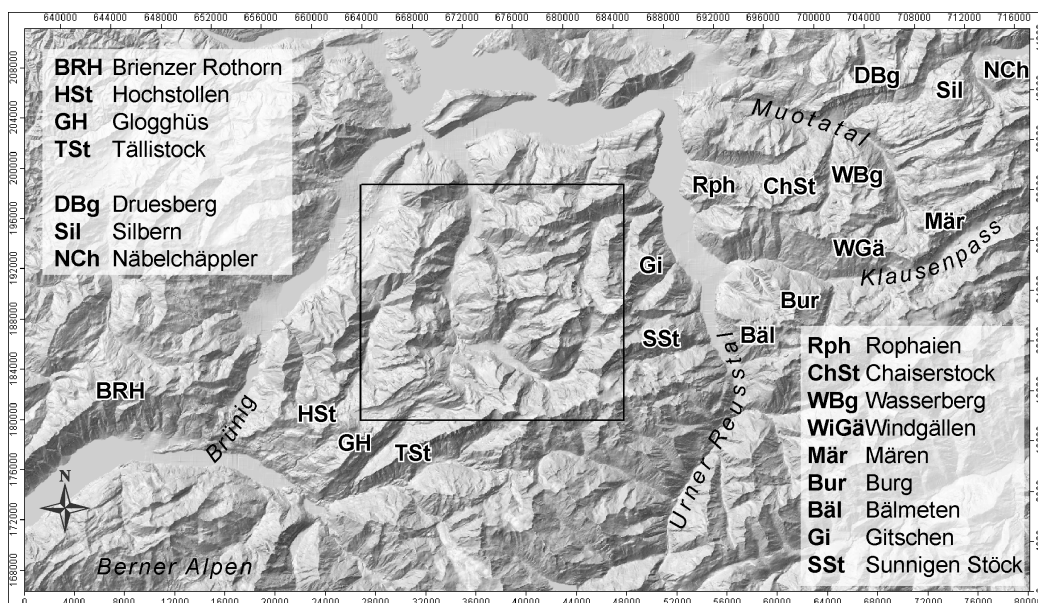


Abbildung 5.18: Karte mit den Lokalnamen im umliegenden Gebiet, auf die im Text verwiesen wird.

ausgewiesen. Diese Gebiete sind ebenfalls prioritär zu besuchen.

- Weiter gegen Osten liegt das *Urner Reusstal*. Auf der anderen Talseite dieses Tals sind einige Gebiete mit hoher Habitateignung hervorgehoben. Falls dort kein Nachweis von *T. biconicus* gelingen sollte, wäre die Suche nach Arten, welche dieselbe oder eine ähnliche ökologische Nische besetzen, nicht minder interessant.

Die interessanten Stellen liegen am *Bälmeten*, bei der *Burg* und beidseits des *Klausenpasses*. Die Gebiete im hinteren *Muotatal* und gegen den *Rophaien* mit hohen 'Spannungen' in der Extrapolation mit MAXENT Level B, weisen meist auch mit Level C hohe Prognosewerte auf, jedoch ohne grössere 'Spannungen'. Dazu gehören die Gebiete um die *Schächentaler Windgällen* bis zum *Chaiserstock*, eventuell auch am *Wasserberg*, weiter die stark verkarsteten Gebiete von der *Mären* bis zur *Silbern*, sowie auf dem *Näbelchäppler*. Die letztgenannten Gebiete sind jedoch während den Hauptvergletscherungen vollständig überflossen worden.

- Westlich und südlich anschliessend an das Untersuchungsgebiet treten nur wenige hohe Prognosewerte auf: Die Gebiete vom *Hochstollen* zum *Glogghüs* und am *Tällistock* sind in die hohe Priorität aufzunehmen, das *Brienzer Rothorn* eher in eine tiefere.
- Vom Untersuchungsgebiet Richtung Süden ändert der geologische Untergrund zu kristallinen Gesteinen. Auf kristallinem Untergrund sind keine Nachweise für *T. biconicus* bekannt.

5.2.2 Extrapolation in andere Klimate

Die Extrapolation in andere Klimate wurde durch eine generelle Erhöhung, beziehungsweise Erniedrigung der Jahresdurchschnittstemperatur versucht (Abbildung 5.19). Der Vorgang ist derselbe, wie die Extrapolation in andere Regionen. Allerdings können andere Wirkungsvariablen, die mit grosser Sicherheit auch einen Zusammenhang mit einer Änderung des Klimas aufweisen, nicht angepasst werden. Insbesondere trifft dies auf die Bodenbedeckung zu. In Level B sind zusätzlich die Variablen `PDay` und `Prec7` von diesem Vorbehalt betroffen. Zudem kann nicht überprüft werden, ob bei einer Klimaänderung zusätzliche Parameter wichtig werden, die unter den heutigen Umweltbedingungen keine oder nur eine untergeordnete Rolle für das Vorkommen von *T. biconicus* spielen, beispielsweise die Niederschlagsmenge und deren Verteilung über das Jahr.

Die Unterschiede der Prognosekarte in Abbildung 5.2, mit einer um 1°C erhöhten Jahresdurchschnittstemperatur, zur Prognosekarte in Abbildung 5.2, mit den unveränderten Temperaturwerten, sind nicht auf den ersten Blick ersichtlich. Einige Gebiete werden durch die Erwärmung in der Habitateignung scheinbar besser bewertet, andere schlechter. Es ist fraglich, ob dies eine ökologisch sinnvolle Extrapolation in ein anderes Klima darstellt.

Test auf die ökologische Relevanz

Um einen Hinweis auf die Glaubwürdigkeit der extrapolierten Prognosen zu erhalten, kann die Projektion in kältere Klimate (Eiszeit) versucht werden.

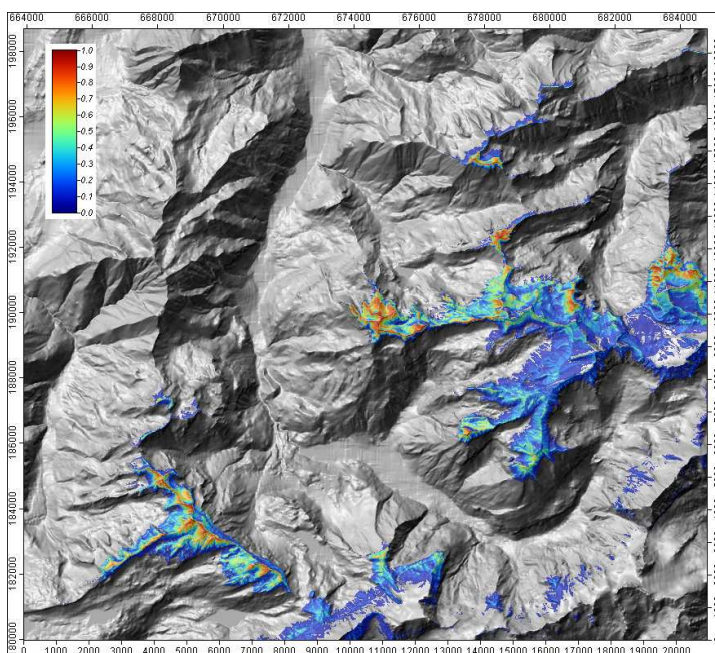


Abbildung 5.19: Die Habitataignung nach der Extrapolation mit MAXENT (Level B) in ein Klima mit um 1°C erhöhter Jahresdurchschnittstemperatur. Im Vergleich mit der Prognose in Abbildung 5.2 "verlieren" die Gebiete am *Chaiserstuel*, *Brisen* und *Widderfeld* an Bedeutung und die Gebiete beim *Uri-Rotstock*, den *Walenstöcken* und *Graustock* bis *Huetstock* werden "aufgewertet". Werte unter 0.046 (Prognosewert des tiefsten nachgewiesenen Fundes) sind transparent gesetzt.

Die Temperaturvariable weist eine starke Korrelation mit der Höhe auf³. Daher ist mit ansteigender Temperatur auch ein 'höher wandern' des modellierten Eignungsgebietes zu erwarten, beziehungsweise das Absinken bei tieferen Temperaturen.

Sinkt das Verbreitungsgebiet jedoch vollständig in Gebiete, die während Eiszeiten von Gletschern bedeckt waren, wäre ein Überleben von *T. biconicus* an diesem Ort nicht möglich gewesen. Die Annahmen zur Extrapolation der Modellierung müssten in diesem Fall als falsch bezeichnet werden.

Die Extrapolation wird zunehmend fehleranfällig, je weiter weg die zu prognostizierenden Bedingungen von den zugrunde liegenden Umweltbedingungen sind. Diesem Umstand wurde hier Rechnung getragen, indem in einem ersten Schritt die Projektionen möglichst nahe an den Trainingsbedingungen gehalten wurden. In Abbildung 5.20 wurde die Prognose für die Bedingungen einer Klimaerwärmung von $+1^{\circ}\text{C}$ mit jener eines kühleren Klimas von -1°C verglichen. Die Gebiete mit gleicher Habitataignung sind (gelb) hervorgehoben. Die Tendenz, mit zunehmender Erwärmung gegen höhere Gebiete zu wandern, ist sehr ausgeprägt. Die Unterschiede für $\Delta T 2^{\circ}\text{C}$ sind bereits

³ Pearson Korrelationskoeffizient $r = -0.998$ zwischen T_{Ave} und dem Höhenwert aus dem Höhenmodell.

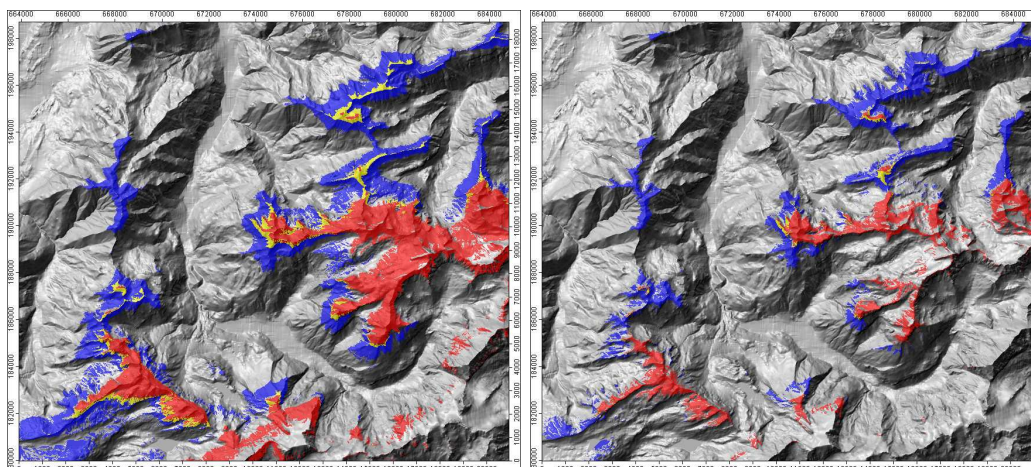


Abbildung 5.20: Die gelben Gebiete umfassen jene Stellen, wo die Habitateignung bei $+1^{\circ}\text{C}$ und bei -1°C gegenüber den Klimaten von 1961-1990 durchgehend gegeben ist. Die roten Gebiete kommen bei einer Erwärmung neu dazu. Die blauen Gebiete sind nur bei -1°C als geeignete Gebiete prognostiziert; links mit dem Schwellenwert 'tiefster nachgewiesener Fund' und rechts mit 'Sensitivität gleich Spezifität'. Da die gelben Gebiete schon bei $\Delta T 2^{\circ}\text{C}$ sehr klein sind, muss der Ansatz, durch Veränderung der Jahresdurchschnittstemperatur in andere Klimate zu projizieren, als falsch betrachtet werden.

enorm. Nur noch kleine Flächen sind unter beiden Klimaten gleichzeitig besiedelt.

Bereits eine um wenige Grad kühlere Jahresdurchschnittstemperatur führt in der Projektion zum Verschwinden von geeigneten Habitaten im Untersuchungsgebiet (ohne Abbildung). Das Temperaturminimum für die letzte Hauptvergletscherung ist nach Schreiner (1992) jedoch schätzungsweise 15° bis 16°C unter den heutigen Temperaturen. Ohne Abwanderung in tiefere Lagen im Vorfeld der Gletscher und anschließende Rückwanderung, hätte *T. biconicus* keine Überlebenschance gehabt. Die Auswanderungstheorie wird durch die heutige Verteilung allerdings nicht gestützt. Daher muss der Ansatz, durch Veränderung der Jahresdurchschnittstemperatur in andere Klimate zu projizieren, als falsch betrachtet werden.

Kapitel 6

Zusammenfassung, Diskussion, Ausblick

6.1 Zusammenfassung

Die vorliegende Arbeit untersuchte die Anwendbarkeit von Habitatmodellen anhand der Funddaten von *T. biconicus*. Durch die Auswahl der Modellierungsansätze und die Ausschöpfung deren Optimierungspotentials konnten sehr gute Prognosen erstellt werden. Damit wurden aber nicht alle Ansprüche abgedeckt. Insbesondere waren kausale Aussagen über Wirkungszusammenhänge zwischen Umweltvariablen und dem Auftreten von *T. biconicus* nur mit grossen Vorbehalten möglich. Als Folge davon blieb die Extrapolation in benachbarte Gebiete und andere Klimate unsicher.

Es konnte jedoch gezeigt werden, dass die neuen Ansätze mit MAXENT und *Boosted Regression Trees* BRT mit den Gegebenheiten im Untersuchungsgebiet sehr gut zu recht kommen. Durch die Verwendung einer grossen Zahl von zufällig über das ganze Untersuchungsgebiet verteilten Pseudoabsenzen konnte die Prognoseleistung der BRT sogar deutlich gesteigert werden. In gleicher Weise konnte die Prognosegüte der GLM ebenfalls deutlich gesteigert werden.

Durch die Habitatmodellierung wurden weitere Gebiete erkannt, in denen die Bedingungen der ökologischen Nische für *T. biconicus* erfüllt, aber bisher

keine Funde bekannt sind. Es wurde ein Verfahren hergeleitet, das die Unsicherheiten der Prognosen darstellt und somit zu priorisierende Lokalitäten für zukünftige Felduntersuchungen aufzeigt. Mit Beobachtungen aus diesen Lokalitäten soll eine Verbesserung der Prognosesicherheit erzielt werden können.

6.2 Diskussion

6.2.1 Methodik

Beantwortung der Eingangsfragen

Der mechanistische Ansatz, über Arbeitshypothesen zu aussagekräftigen Wirkungsvariablen zu kommen, scheiterte an der Verfügbarkeit der Umweltdaten, welche auch indirekt nicht oder nur unzureichend hergeleitet werden konnten (Abschnitt 3.3.2). Aus den vorhandenen Umweltdaten waren jedoch gute Prognosen herzustellen (Abschnitt 4.3). Die Umweltvariablen wurden darauf untersucht, wie gut sie das Vorkommen von *T. biconicus* erklären (vgl. die Tabellen zu den Prognosekarten in 5.1.1). Die Prognosen konnten jedoch nur eingeschränkt als tatsächliche Wirkungsvariablen in die ökologische Interpretation einfließen (Abschnitt 6.2.1). Somit war das Ausformulieren von konkreten Schutzmassnahmen nicht direkt auf in der Modellierung bestimmte Wirkungszusammenhänge abzustützen. Jedoch lieferte das potentielle Verbreitungsgebiet eine gute Grundlage dafür. Ebenso lieferten die Prognosen Hinweise zur Priorisierung zukünftiger Felduntersuchungen in bisher nicht untersuchten Gebieten. Prioritär untersucht werden sollen Gebiete mit prognostizierten Populationen und solche mit grossen Differenzen in den Prognosemodellen (Abschnitt 3.5).

Projektionen in Gebiete, die sich in den Bedingungen vom Untersuchungsgebiet unterscheiden, sind problematisch. Dies gilt sowohl für Prognosen in geografisch andere Regionen wie auch in andere Klimate (Abschnitt 5.2). Das Beispiel der räumlichen Extrapolation mit MAXENT Level B mit den auftretenden Anomalien im Juli-Niederschlag und in der Anzahl Regentage in

Abschnitt 5.2.1 zeigte anschaulich, wie sich Modelle im Untersuchungsgebiet ähnlich sein können (siehe Prognosekarten für Level B und Level C in den Abbildungen 5.2 und 5.3) und die Extrapolationen trotzdem unterschiedlich ausfallen können (Abbildung 5.15).

Der simple Ansatz, eine Klimaänderung nur durch Veränderung der Jahresdurchschnittstemperatur zu simulieren (Abschnitt 5.2.2), stützte sich auf die Annahme, dass die Veränderung der Jahresdurchschnittstemperatur ausschliesslich das Temperatursignal betreffe. Die Interpretation der Darstellung mit $\Delta T 2^\circ\text{C}$ (Abbildung 5.20) legte aber den Schluss nahe, dass die Jahresdurchschnittstemperatur viel mehr die damit hoch korrelierte Höhe erklärt. Deshalb stellt sie das ökologische Temperatursignal nur unzureichend dar. Die Veränderung der Umweltvariablen 'Jahresdurchschnittstemperatur' entspricht also eher einer Hebung oder einem Absenken der Topografie. Das Vorgehen, durch simples Verändern der Jahresdurchschnittstemperatur in andere Klimate zu projizieren, führte deshalb zu nicht plausiblen Resultaten.

Die Ausbreitungsgeschichte von *T. biconicus* ist ein sehr spannendes Forschungsgebiet. Das Wissen um Gebiete mit geeigneten Umweltbedingungen, die aber nicht besiedelt sind, kann wichtige Hinweise auf die Ausbreitungskapazität der Art und die Veränderungen der Landschaft über längere Zeiträume geben. Dazu dürfte die Untersuchung des *Laubersgrat* eine Schlüsselfunktion haben. Dort befindet sich das Anrissgebiet des oder der späteiszeitlichen Bergstürze von *Engelberg* (Hess, 1986). Diese müssten eine bestehende Population zerstört haben. Eine Neubesiedlung ist aufgrund der geografischen Barrieren nicht wahrscheinlich.

Die ausserhalb des Untersuchungsgebietes gefundenen Gebiete mit hoher Habitateignung werden entweder zu neuen Nachweisen von *T. biconicus* oder zur räumlich besseren Eingrenzung ihres Vorkommens führen. Die Gebiete auf der orografisch rechten Seite des *Urner Reusstales* werden womöglich Hinweise auf Arten mit ähnlicher ökologischer Nische liefern.

Die Frage, ob die hier angewandten Modellierungstechniken mit der verfügbaren Datengrundlage in einem Gebiet mit sehr starken topografischen Gradienten für die Modellierung einer sehr immobilen Art tauglich sind,

konnte bejaht werden. Die starken Gradienten führten sogar zu besonders guten Prognosen.

Die ökologische Aussagekraft der Umweltvariablen

Die **Jahresdurchschnittstemperatur** wurde bei allen Modellansätzen als wichtige Variable ermittelt (siehe die Tabellen über den Beitrag zum Modell in den Abbildungen 5.1 bis 5.8). Diese Variable war jedoch, bedingt durch die Interpolation über grosse Distanzen zwischen den Messstationen, mit starken Fehlern behaftet (siehe auch Abschnitt 4.3.2) und berücksichtigte zudem die Exposition nicht. Vermutlich bildete daher die **Exposition** (vor allem *Northness*) eine wichtige Interaktion mit der Jahresdurchschnittstemperatur. Denn die Exposition alleine lieferte keinen markanten Erklärungsbeitrag zum Modell (Abbildung 6.1): Der sehr kurze blaue Balken von *nness* im

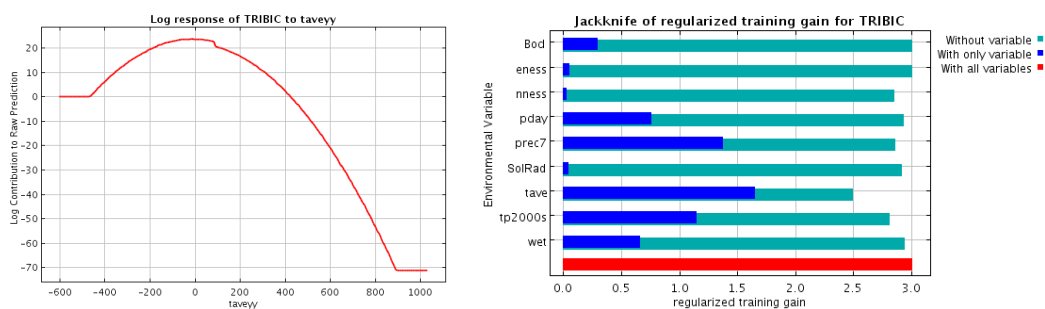


Abbildung 6.1: Die Charakteristikkurve (links) der Umweltvariablen 'Jahresdurchschnittstemperatur' (*taveyy*) zeigt deren Einfluss auf das Vorkommen von *T. biconicus* in MAXENT Level B mit einem deutlichen Maximum im Bereich von 0°C. Der Jack-Knife-Test (rechts) zeigt, wie gut das Modell mit allen Variablen ist (roter Balken), wieviel davon beim Weglassen einzelner Umweltvariablen übrig bleibt (grüne Balken) und wie viel die jeweilige Umweltvariable alleine erklären würde (blaue Balken).

Jack-Knife-Test steht für den geringen Erklärungsbeitrag dieser Variablen zum Modell. Das Zurückweichen des grünen Balkens zeigt jedoch, dass beim Weglassen dieser Variable das Modell eine deutliche Qualitätseinbusse erleidet. Ein ähnliches Verhalten zeigt auch *SolRad*. Die Wichtigkeit der Umweltvariablen *TAVE* zeigt sich im Jack-Knife-Test durch das starke Zurückweichen des grünen Balkens.

Der hohe Beitrag der Variablen 'Jahresdurchschnittstemperatur' zum Modell in den MAXENT-Modellen und ihre gleichzeitig hohe Signifikanz im linearen und im quadratischen Term im GLM zeigt das Vorhandensein eines optimalen Temperatur-, beziehungsweise Höhenbereichs für das Vorkommen von *T. biconicus*.

Allerdings ist die Aussage der Jahresdurchschnittstemperatur mit sehr vielen anderen Umweltvariablen hoch korreliert. Deshalb ist die Jahresdurchschnittstemperatur als Vertreter all dieser Variablen zu betrachten. Eine Reduktion auf das ökologisch interpretierbare Temperatursignal ist daher nicht zulässig.

Eine Arbeitshypothese besagt, dass *T. biconicus* auf Nunatakern die Vergletscherungen überdauert hat. Zusammen mit der geringen Ausbreitungskapazität ergibt sich so die Annahme, dass die Art nur an Stellen zu finden sei, wo keine Gletscherüberbedeckung vorhanden war. Diese Bedingung ist hauptsächlich an Gipfellagen und exponierten Gebirgskämmen erfüllt, was durch die Umweltvariablen für die topografische Position `tp2000s` vermutlich sehr gut nachgezeichnet wird. Auch diese Variable ist bei den Modellen durchwegs von grosser Bedeutung bzw. Signifikanz.

Auf die Rolle des Juli-Niederschlages wurde im Zusammenhang mit der räumlichen Extrapolation bereits hingewiesen. Vermutlich treffen die Aussagen zum Juli in die aktivste Zeit von *T. biconicus* und der Vegetation in Allgemeinen an den Fundorten. Die Relevanz ist daher nachvollziehbar. Der Zusammenhang scheint sinnvoll, da Schnecken auf eine gewisse Feuchtigkeit angewiesen sind (`Prec7` ist positiv mit dem Vorkommen von *T. biconicus* korreliert). In der ökologischen Aussage ähnlich, jedoch in der Interpretation der Auswirkung auf das Vorkommen von *T. biconicus* komplizierter, sind die Umweltvariablen `Wet`, `TWI` und `TWIs` (Abbildung 6.2). Die Zuordnung der Bodenbedeckungsart (`Bod`) war im Vergleich mit dem Luftbild und den Feldbeobachtungen oft nicht ganz nachvollziehbar. Dies ist möglicherweise damit erklärbar, dass die Bodenbedeckungsarten unter anderen Gesichtspunkten als den hier benötigten zugeordnet wurden. Die ökologisch interpretierbare Aussage dieser Umweltvariablen dürfte wohl das Vorhandensein von flachen Steinen in den Schuttfeldern oder schuttfeldnahen Vegetationsflächen sein.

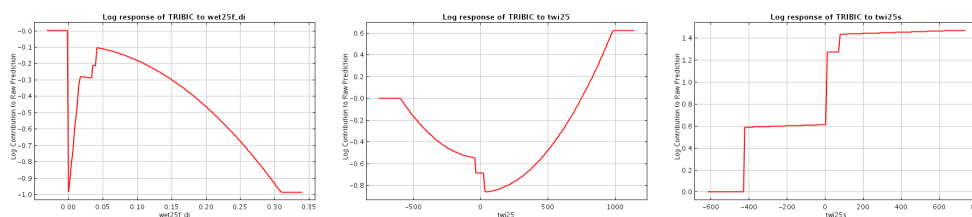


Abbildung 6.2: Die Charakteristikkurven der Umweltvariablen *Wet*, *TWI* und *TWIs* aus MAXENT Level A zeigen den ermittelten Effekt auf das Auftreten von *T. biconicus*. Sie lassen keine einfache Interpretation zu.

Gemäss der Feldbeobachtungen ist es jedoch entscheidend, ob eine Schuttbedeckung einigermaßen stabil oder in Bewegung ist. In bewegten Schuttfeldern konnte bisher nie ein Nachweis von *T. biconicus* gemacht werden. Die Aussagekraft der Bodenbedeckungsart ist dementsprechend gering. Sie spielte deshalb in keinem der Modelle eine wichtige Rolle.

Die Abhängigkeit von der Rasterauflösung

Alle Daten, die sich auf das DHM25 stützten, hatten entsprechend eine räumliche Auflösung von 25×25 m. In Teilen des Untersuchungsgebietes waren jedoch ausreichend Höhenpunkte vorhanden, um ein feineres Höhenmodell zu erstellen¹. Neben der Frage des geeigneten Interpolationsverfahrens mussten auch grundsätzliche Fragen erörtert werden. Fragen wie, ob die Auflösung von 25×25 m genügend sei oder nicht und welche Umweltvariablen überhaupt von einer Verfeinerung profitieren würden, haben sich Lassueur et al. (2006) in sehr ähnlicher Form auch gestellt. Alle Daten, die aus meteorologischen Messstationen interpoliert werden, sind sowieso nur für ein Meso- oder Makroklima relevant und daher für mikroklimatische Vorhersagen nur sehr bedingt tauglich (Guisan & Zimmermann, 2000). Hingegen sind die Werte von Variablen, welche direkt aus dem Höhenmodell abgeleitet werden, durchaus relevant. Lassueur et al. (2006) untersuchten daher die Hangneigung, die Exposition und die Krümmung näher, mit dem Resultat, dass die Aussagen aus einem 25×25 m-Höhenmodell durch solche aus einem sehr hochauflösenden

¹ Im Gebiet *Bannalp* konnte ein Höhenmodell aus fotogrammetrisch und trigonometrisch ermittelten Höhenpunkten erstellt werden, das einer Rasterauflösung von 5×5 m gerecht wurde.

Höhenmodell² im Allgemeinen nicht deutlich überboten werden. Einzig die aus der Exposition abgeleiteten Variablen werden in einem sehr hochauflösenden Höhenmodell besser wiedergegeben. Möglicherweise werden in höher gelegenen Gebieten (über 2000 m ü.M) infolge der Akzentuierung der Topografie die Krümmungswerte durch sehr hochauflösende Höhenmodelle besser wiedergegeben und daher für die Modellierung wichtiger.

In den Modellierungen für *T. biconicus* war der Beitrag von *Northness* und *Eastness* zum Modell nie besonders hoch und die Variablen für die Krümmung fielen sehr schnell aus den Berechnungen weg. Daher wurde an dieser Stelle auf die eingehende Diskussion und auf die Durchführung von Tests mit diesen Variablen verzichtet.

6.2.2 Werkzeuge zur Modellierung, Auswertung und Darstellung

Im Folgenden wird betrachtet, inwiefern die in den Applikationen umgesetzten Funktionalitäten den Arbeitsfluss einer Habitatmodellierung unterstützen.

Die Erstellung von Modellen mit MAXENT ist sehr komfortabel und kann im 'batch modus' zu einem guten Teil automatisiert werden. Das Resultat ist eine gut strukturierte HTML-Seite. Allerdings ist dieses Werkzeug auch etwas starr und in sich geschlossen. Es kann so kaum in einen übergeordneten Prozess integriert werden und das Einbinden der Resultate in eine Auswertungsumgebung ist kaum möglich. Die ausgegebenen Prognosebilder sind optisch ansprechend, aber als rohe Wahrscheinlichkeitsprognosen auch etwas trügerisch (siehe auch Abschnitt 3.4.3 Schwellenwerte). Ausserdem ist die MAXENT-User Group³ schnell und sehr aktiv.

Die R-basierten Methoden bieten kein mit MAXENT vergleichbar attraktives *Frontend* und die Einstiegshürde in die R-Umgebung ist gross. Die

² Unter sehr hoch auflösenden Höhenmodellen werden solche verstanden, die eine Rasterbreite von 1 × 1 m oder genauer aufweisen.

³ <http://groups.google.com/group/Maxent?hl=en>

immense Funktionalität und die Transparenz der Arbeitsschritte rechtfertigen jedoch den Einarbeitungsaufwand. Anpassungen und spezielle Wünsche lassen sich fast beliebig umsetzen. Durch die Schnittstellen zu Datenbank-⁴ und GIS-Applikationen⁵ wird diese Umgebung für Geostatistik noch interessanter. Die Möglichkeiten wurden in dieser Arbeit bei Weitem nicht ausgeschöpft. Bei der Berechnung von BRT-Modellen sind längere Rechenzeiten in Kauf zu nehmen.

Für den letzten Teil der Arbeit - die Analyse und den Vergleich der verschiedenen Prognosen - bot MAXENT keine Funktionalitäten an. Die Kombination von R und SAGA hatte sich in dieser Arbeit bewährt.

Die Applikationen zur Generierung, Sammlung und Aufbereitung der Umweltdaten waren nicht im Fokus dieser Arbeit und wurden deshalb nicht behandelt.

6.3 Ausblick

Die Nagelprobe für diese Arbeit werden die Resultate des nächsten Feldsommers sein. Es wird sich zeigen, ob durch die neuen Daten aus den Gebieten mit widersprüchlichen Prognosen, die Modelle verbessert werden. Ebenso wird es sehr spannend sein, die zusätzlichen, als geeignet eingestuften Gebiete nach *T. biconicus* abzusuchen. Die Arbeiten um diese Art sollten jedoch auch auf andere Forschungsgebiete ausgedehnt werden. So sind Untersuchungen zur Genetik der nächste entscheidende Schritt zum Verständnis der Ausbreitungsgeschichte von *T. biconicus*.

Die Fragen auf Seite 1 nach der Populationsstruktur d) und nach der Lebensweise f) können nicht direkt durch die geografische Analyse beantwortet werden. Die Populationsgeschichte kann jedoch in einem weiterführenden Projekt mit Hilfe von genetischen Untersuchungen behandelt werden. Dazu liefern die Erkenntnisse dieser Untersuchung gute Grundlagen. Ebenso

⁴ Beispielsweise Postgres oder andere via ODBC.

⁵ Beispielsweise SAGA 2 oder GRASS; ArcGIS kann R via die Python-Schnittstelle nutzen.

verhält es sich mit den Erkenntnissen zum „Lebenswandel“ und zur Ausbreitungsgeschwindigkeit der Art. Diese Merkmale können nicht direkt aus der Geoanalyse abgeleitet werden - die Resultate dieser Arbeit bieten jedoch eine Diskussionsbasis.

Das Feld für mögliche Verbesserungen im Bereich der Modellierungsalgorithmen ist im Rahmen dieser Arbeit schwer abzuschätzen. Jedoch ist die Forderung von Guisan & Thuiller (2005) nach Entwicklung von standardisierten und robusten Modellierungsumgebungen mit Nachdruck zu unterstützen. Mit der immer grösseren Zahl von Modellierungen zu einzelnen Arten könnte durch übergeordnete "Meta-Modellierungen" allmählich ein grösseres Verständnis für die Wechselwirkung der modellierten Arten untereinander und auf die Umweltbedingungen aufgebaut werden. Dazu würden insbesondere auch Regeln gehören, wie die Metadaten für Modellierungen zu erfassen sind.

Ein Versuch in diese Richtung ist mit OpenModeller⁶ weit fortgeschritten. OpenModeller bietet eine gemeinsame Plattform für verschiedene Modellierungsansätze und Datenquellen.

Im Bereich der Umweltvariablen eröffnen hochauflösende Höhenmodelle im Zusammenhang mit der Formulierung abiotischer Zusammenhänge (siehe auch Abschnitt 3.1) grössere Experimentierfelder, die zur Verbesserung der Verbreitungsprognosen beitragen. Ein grosser Schritt wäre, wenn Umweltdaten und daraus abgeleitete Variablen in einer soliden Geodateninfrastruktur zugänglich gemacht würden und somit die allgemeine Basis für Modellierungen bildeten. Daraus entstünde die Chance, dass weniger Arbeiten mit ungenügenden oder unvergleichbaren Datengrundlagen entstehen würden.

Das Interesse am Einfluss der Geologie und der Talentwicklungsgeschichte des *Engelbergertales* war stark persönlich motiviert. Leider war die Aufarbeitung und Zusammenführung der verschiedenen geologischen Kartierungen sehr aufwändig und musste daher im Rahmen dieser Arbeit zurückgestellt werden (Abschnitt 4.2.4). Zur Vergletscherungsgeschichte sind in jüngster Zeit einige Studien verfasst und Modellierungsansätze ausgearbeitet worden⁷.

⁶ <http://openmodeller.sourceforge.net/>

⁷ Beispielsweise die Dissertation von Florineth (1998) mit Modellen zur letzten Hauptvergletscherungsphase in den Schweizer Alpen.

Eine aktuelle Ausgabe der Vergletscherungskarte ist in Vorbereitung ⁸. Allerdings sind all diese Informationsquellen noch in einem zu groben Massstab, als dass sie für die Modellierung lokaler Prozesse nutzbar sein könnten. Umgekehrt können aber Erkenntnisse über die Ausbreitungsgeschichte von *T. biconicus* in die Überlegungen der Quartären Landschaftsentwicklung einfließen und bisher ungenutzte Hinweise liefern.

Ich hoffe, mit dieser Arbeit einen Mosaikstein zum Verständnis der *T. biconicus* und ihrem Lebensraum beigetragen zu haben und das hier erworbene Wissen auch anderweitig einfließen lassen zu können.

⁸ Mündliche Mitteilung von M. Maisch, 22.9.07 .

Literaturverzeichnis

- Baggenstos, Markus**, 2006. Erhebungen zur Verbreitung und Biologie der Nidwalder Haarschnecke (*Trochulus biconicus*) als Grundlage für die Formulierung von Schutzmassnahmen. unpubliziert. Projektantrag
- Baggenstos, Markus**, 2007. Erhebungen zur Verbreitung und Biologie der Nidwalder Haarschnecke (*Trochulus biconicus*) als Grundlage für die Formulierung von Schutzmassnahmen. unpubliziert. URL <http://www.oekoberatung.ch/pdf/haarschneckenprojekt.pdf>, projektbeschrieb
- Baggenstos, Markus**, in prep. Erhebungen zur Verbreitung und Biologie der Nidwaldner Haarschnecke (*Trochulus biconicus*) als Grundlage für die Formulierung von Schutzmassnahmen. Tech. Rep., Oekologische Beratung Markus Baggenstos
- Baur, Bruno, Duelli, Peter, Edwards, Peter J., Jenny, Markus, Klaus, Gregor, Klaus, Gregor, Künzle, Irene, Martinez, Sylvia, Pauli, Daniela, Peter, Kathrin, Schmid, Bernhard, Seidl, Irmi & Suter, Werner**, 2004. *Biodiversität in der Schweiz - Zustand, Erhaltung, Perspektiven*. Forum Biodiversität Schweiz
- Brotons, Llíus, Thuiller, Wilfried, Araújo, Miguel B. & Hirzel, Alexandre H.**, 2004. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. In *Ecography*, vol. 27, pp. 437 – 448
- Cohen, Jacob**, 1960. A coefficient of agreement for nominal scales. In *Educational and Physiological Measurement*, vol. 20, pp. 37 – 46
- Davis, A. J., Jenkinson, L. S., Lawton, J. H., Shorrocks, B. & Wood, S.**, 1998. Making mistakes when predicting shifts in species range in response to global warming. In *Nature*, vol. 391(6669), pp. 783 – 786
- Dormann, Carsten F.**, 2007. Promising the future? Global change projections of species distributions. In *Basic and applied Ecology*, vol. 8(5), pp. 387–397
- Dormann, Carsten F., McPherson, Jana M., Araújo, Miguel B., Bivand, Roger, Bolliger, Janine, Carl, Gudrun, Davies, Richard G.**,

- Hirzel, Alexandre, Jetz, Walter, Kissling, W. Daniel, Kühn, Ingolf, Ohlemüller, Ralf, Peres-Neto, Pedro R., Reineking, Björn, Schröder, Boris, Schurr, Frank M. & Wilson, Robert, 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. In *Ecography*, vol. 30(5), pp. 609–628
- Dudík, Miroslav, Phillips, Steven J. & Schapire, Robert E., 2004. Performance Guarantees for Regularized Maximum Entropy Density Estimation. In *Proceedings of the 17th Annual Conference on Computational Learning Theory*. (pp. 472–486)
- Eder, Leo, 1917. Eine neue Schweizer Helicide. In *Revue Suisse de Zoologie*, vol. 25, pp. 441–452
- Elith, Jane, Graham, Catherine H., Anderson, Robert P., Dudík, Miroslav, Ferrier, Simon, Hijmans, Robert J., Huettmann, Falk, Leathwick, John R., Lehmann, Anthony, Li, Jin, Lohmann, Lucia G., Loiselle, Bette A., Manion, Glenn, Moritz, Craig, Nakamura, Miguel, Nakazawa, Yoshinori, Overton, Jacob McC., Peterson, A. Townsend, Phillips, Steven J., Richardson, Karen, Scachetti-Pereira, Ricardo, Schapire, Robert E., Soberón, Jorge, Williams, Stephen, Wisz, Mary S. & Zimmermann, Niklaus E., 2006. Novel methods improve prediction of species' distribution from occurrence data. In *Ecography*, vol. 29, pp. 129–151
- Fielding, Alain H. & Bell, John F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. In *Environmental Conservation*, vol. 24, pp. 38 – 49
- Florineth, Duri, 1998. Geometry of the Last Glacial Maximum (LGM) in the Swiss Alps and its paleoclimatological significance. Ph.D. thesis, Geologisches Institut der Universität Bern
- Freund, Yoav & Schapire, Robert E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. In *Journal of Computer and System Science*, vol. 55, pp. 119–139
- Garzón, Marta Benito, Blazek, Radim, Neteler, Markus, Sánchez de Dios, Rut, Sainz Ollero, Helios & Furlanello, Cesare, 2006. Predicting habitat suitability with machine learning models: The potential area of *Pinus sylvestris* L. in the Iberian Peninsula. In *Ecological Modelling*, vol. 197, pp. 383–393
- Gibson, Lesli, Barrett, Brent & Burbidge, Allan, 2007. Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. In *Diversity and Distributions*, vol. 13, pp. 704–713

- Grier, C. G. & Running, S. W.**, 1977. Leaf area of mature northwestern coniferous forests: relation to site water balance. In *Ecology*, vol. 58, pp. 893–899
- Guisan, Antoine, Broennimann, Olivier, Engler, Robin, Vust, Matthias, Yoccoz, Nigel G., Lehmann, Anthony & Zimmermann, Niklaus E.**, 2006. Using Niche-Based Models to Improve the Sampling of Rare Species. In *Conservation Biology*, vol. 20(2), pp. 501–511
- Guisan, Antoine & Theurillat, Jean-Paul**, 2005. *Appropriate Monitoring Networks are required for testing Model-based Scenarios of Climate Change Impact on Mountain Plant Distribution*, vol. 23 Advances in Global Change Research. Kluwer Academic Publishers, (pp. 467–476)
- Guisan, Antoine & Thuiller, Wilfried**, 2005. Predicting species distribution: offering more than simple habitat models. In *Ecology Letters*, vol. 8, pp. 993 – 1009
- Guisan, Antoine & Zimmermann, Niklaus E.**, 2000. Predictive habitat distribution models in ecology. In *Ecological Modelling*, vol. 135, pp. 147 – 186
- Guisan, Antoine, Zimmermann, Niklaus E., Elith, Jane, Graham, C.H., Phillips, Steven & Peterson, A.T.**, 2007. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? In *Ecological Monographs*, vol. 77, pp. 615–630
- Hastie, Trevor, Tibshirani, Robert & Friedman, Jerome**, 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York
- Hengl, Tomislav**, 2007. A Practical Guide to Geostatistical Mapping of Environmental Variables. Tech. Rep., European Commission, JRC, Institute for Environment and Sustainability
- Hess, Paul**, 1986. Der prähistorische Bergsturz von Engelberg. No. 19 in Engelberger Dokumente
- Hosmer, David W. & Lemeshow, Stanley**, 2000. *Applied logistic regression*. Wiley Series in probability and statistics. John Wiley & Sons, New York, 2nd edition edn.
- Hutchinson, G.E.**, 1957. Concluding remarks. In *Cold Spring Harbor Symposium on Quantitative Biology 22*. (pp. 415 – 427)
- Lassueur, Thierry, Joost, Stéphane & Randin, Christophe F.**, 2006. Very high resolution digital elevation models: Do they improve models of plant species distribution? In *Ecological Modelling*, vol. 198, pp. 139 – 153

- Lausch, Angela**, 2004. Raum, Zeit, Struktur und Skala in Habitatmodellen - Eine Einführung. In *Habitatmodelle – Methodik, Anwendung, Nutzen; Tagungsband zum Workshop vom 8.-10. Oktober 2003 am UFZ Leipzig* (editors **Dormann, CF, Blaschke, Thomas, Lausch, Angela, Schröder, Boris & Söndgerath, D.**), UFZ-Berichte. (pp. 69–82)
- Liu, Canran, Berry, Pam M., Dawson, Terence P. & Pearson, Richard G.**, 2005. Selecting thresholds of occurrence in the prediction of species distributions. In *Ecography*, vol. 28(3), pp. 385 – 393
- Lobo, Jorge M., Jimenez-Valverde, Alberto & Real, Raimundo**, 2007. AUC: a misleading measure of the performance of predictive distribution models. In *Global Ecology and Biogeography*, vol. 17(2), pp. 145 – 151
- Maggini, Ramona, Lehmann, Anthony, Zimmermann, Niklaus E. & Guisan, Antoine**, 2006. Improving generalized regression analysis for the spatial prediction of forest communities. In *Journal of Biogeography*, vol. 33, pp. 1729–1749
- Pearson, Richard G.**, 2007. Species' Distribution Modeling for Conservation Educators and Practitioners. Handout. URL <http://biodiversityinformatics.amnh.org/index.php?section=sdm&mid=40>
- Pearson, Richard G., Raxworthy, Christopher J., Miguel, Nakamura & Peterson, A. Townsend**, 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. In *Journal of Biogeography*, vol. 34, pp. 102–117
- Pearson, Richard G., Thuiller, Wilfried, Araújo, Miguel B., Martinez-Meyer, Enrique, Brotons, Llíus, McClean, Colin, Miles, Lera, Segurado, Pedro, Dawson, Terence P. & Lees, David C.**, 2006. Model-based uncertainty in species range prediction. In *Journal of Biogeography*, vol. 33, pp. 1704–1711
- Pfenninger, Markus, Hrabáková, Magda, Steinke, Dirk & Dèpraz, Aline**, 2005. Why do snails have hairs? A Bayesian inference of character evolution. In *BMC Evolutionary Biology*, vol. 5(59)
- Phillips, Steven J.**, 2006. *A Brief Tutorial on Maxent*. AT&T Research. URL <http://www.cs.princeton.edu/~schapire/maxent/tutorial/tutorial.doc>
- Phillips, Steven J., Anderson, Robert P. & Schapire, Robert E.**, 2006. Maximum entropy modeling of species geographic distributions. In *Ecological Modelling*, vol. 190, pp. 231–259

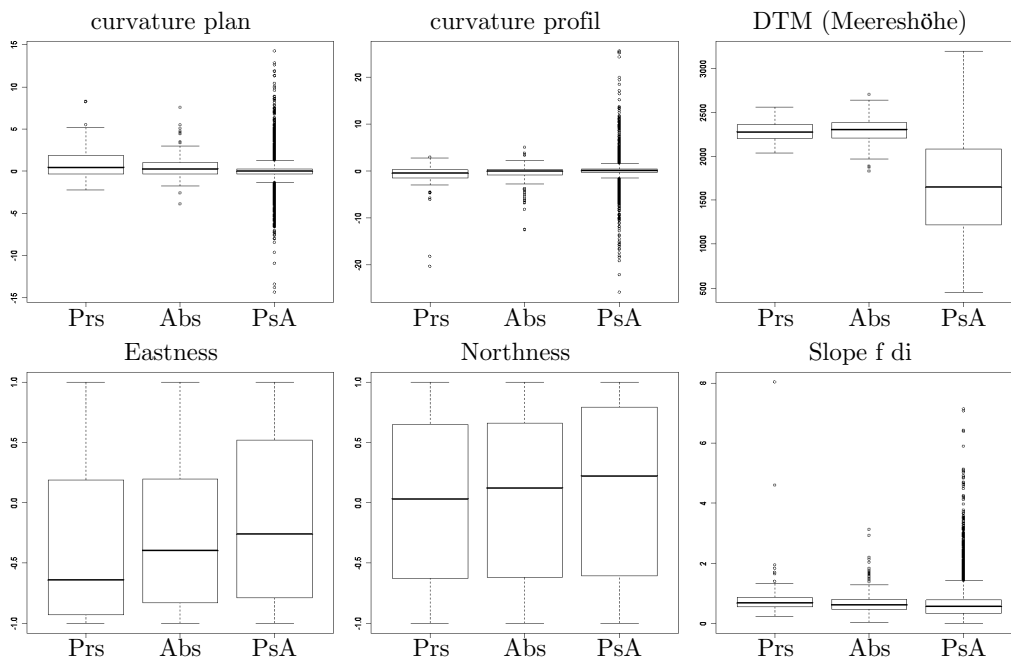
- Phillips, Steven J., Dudík, Miroslav & Schapire, Robert E.**, 2004. A Maximum Entropy Approach to Species Distribution Modeling. In *Proceedings of the Twenty-First International Conference on Machine Learning*. (pp. 655–662)
- Pulliam, H. Ronald**, 2000. On the relationship between niche and distribution. In *Ecology Letters*, vol. 3, pp. 349–361
- R Development Core Team**, 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>
- Reineking, Biörn & Schröder, Boris**, 2004. Gütemasse für Habitatmodelle. In *Habitatmodelle – Methodik, Anwendung, Nutzen; Tagungsband zum Workshop vom 8.-10. Oktober 2003 am UFZ Leipzig* (editors **Dorrmann, Carsten F., Blaschke, Thomas, Lausch, Angela, Schröder, Boris & Söndgerath, Dagmar**). (pp. 27–37)
- Roberts, D. W., Fisher, R. F., Long, J. M. & Jack, S. N.**, 1993. The Leaf Area Allocation Model. Final Report. Tech. Rep., Environmental Protection Agency (EPA)
- Schreiner, Albert**, 1992. *Einführung in die Quartärgeologie*. Schweizerbart'sche Verlagsbuchhandlung
- Swets, John A.**, 1988. Measuring the accuracy of diagnostic systems. In *Science*, vol. 240, pp. 1285–1293
- Tarboton, David G.**, 2005. *Terrain Analysis Using Digital Elevation Models (TauDEM)*. Utah State University. URL <http://hydrology.neng.usu.edu/taudem/>
- Thuiller, Wilfried**, 2004. Patterns and uncertainties of species' range shifts under climate change. In *Global Change Biology*, vol. 10, pp. 2020–2027
- Wigger, Fabian**, 2007. Der mikroklimatische und zeitabhängige Aktivitätsrhythmus von *Trochulus biconicus*. Projektarbeit, NLU Biogeographie, Universität Basel
- Wilkin, Teddy A., Perrins, Christopher M. & Sheldon, Ben C.**, 2007. The use of GIS in estimating spatial variation in habitat quality: a case study of lay-date in the Great Tit *Parus major*. In *Ibis*, vol. 149 (Suppl. 2), pp. 110–118
- Yesson, Chris & Culham, Alastair**, 2006. A phyloclimatic study of *Cyclamen*. In *BMC Evolutionary Biology*, vol. 6. URL <http://www.biomedcentral.com/1471-2148/6/72>

- Zimmermann, Niklaus E. & Kienast, Felix**, 1999. Predictive mapping of alpine grasslands in Switzerland: Species versus community approach. In *Journal of Vegetation Science*, vol. 10, pp. 469–482
- Zimmermann, Niklaus E. & Roberts, David W.**, 2001. Final Report of the MLP climate and biophysical mapping project. Tech. Rep., Swiss Federal Research Institute WSL, Birmensdorf

Anhang A

Umweltvariablen

Die Kistendiagramme (*Boxplots*) zeigen für die in Abschnitt 4.3.1 selektierten kontinuierlichen Umweltvariablen die Verteilung der Werte für die Fundstellen (Präsenzen: Prs), die abgesuchten Stellen ohne Nachweis (Absenzen: Abs) und die Werteverteilung an den 10'000 zufällig über das Untersuchungsgebiet verteilten Pseudoabsenzen (PsA):



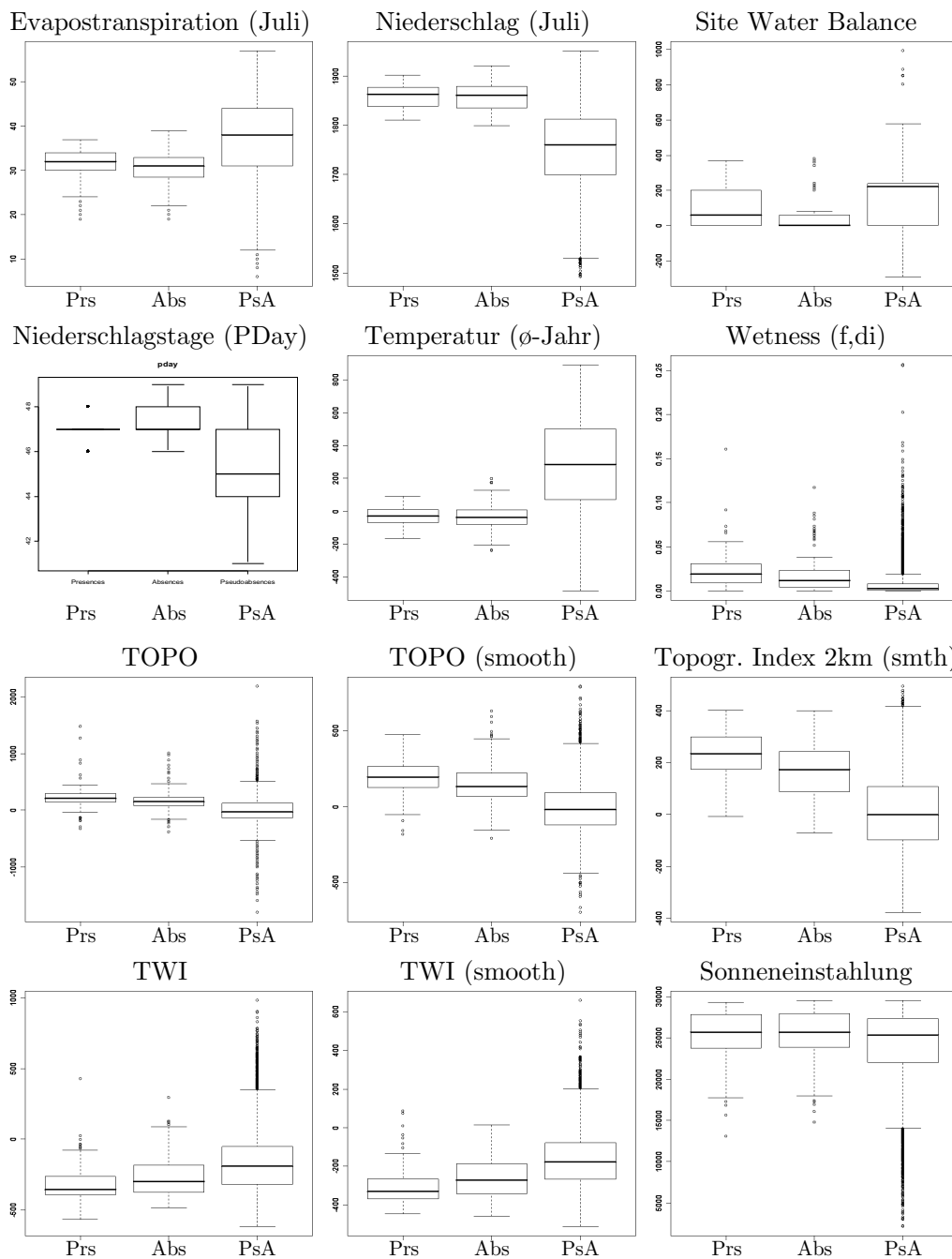


Abbildung A.1: Wertebereiche der Umweltvariablen für Fundstellen, Absenzen und Pseudoabsenzen