

# 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 ausgewählter invasiver Neophyten im Kanton Zürich, Schweiz

vorgelegt von

**Dr. Reimo Lieske**

Teilnehmerkennzahl U1303, UNIGIS MSc Jahrgang 2007

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

Gutachter:

Ao. Univ. Prof. Dr. Josef Strobl  
Dr. Niklaus E. Zimmermann

Winterthur, 30.04.2010

Erklärung:

Ich versichere, diese Master Thesis ohne fremde Hilfe und ohne Verwendung anderer als der aufgeführten 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.

Winterthur, 30.04.2010

## **Danksagung**

Bei der Anfertigung dieser Arbeit konnte ich erfreulicherweise auf die Unterstützung der Eidgenössischen Forschungsanstalt für Wald, Schnee und Landschaft WSL zurückgreifen. Mein besonderer Dank gilt dabei Dr. Nicklaus E. Zimmermann, Leiter der Forschungseinheit Landnutzungsdynamik, der die Idee für dieses Thema aufbrachte und mit dem ich einige wertvolle Diskussionen führte. Darüber hinaus versorgte er mich zu Beginn meiner Arbeit mit den Daten und allerhand Funktionen für das Statistik-Programm R, um mir den Einstieg in diese Thematik zu erleichtern.

Mein Dank gilt auch Herrn Prof. Dr. Josef Strobl und dem gesamten UNIGIS-Team, die mich während meines Studiums hervorragend begleiteten.

## **Kurzfassung**

Neophyten sind Pflanzen, die in fremde Gebiete eingeführt wurden oder eingewandert sind. Manche dieser neuen Arten breiten sich sehr schnell aus und können verschiedenste Probleme verursachen. Man spricht dann von invasiven Neophyten. Auch in der Schweiz breiten sich invasive Neophyten zunehmend aus und müssen deshalb speziell überwacht werden. Für den Kanton Zürich wurde deshalb von der Baudirektion eine Datenbank aufgebaut, die eine Vielzahl von Fundortdaten enthält und ständig erweitert wird.

Da jedoch sowohl zwischen den einzelnen Arten als auch im geografischen Kontext große Differenzen hinsichtlich der Datengrundlage bestehen, ist der Bedarf an Prognosemodellen, die das Vorkommen einzelner Arten modellieren, sehr groß.

Deshalb wurden im Rahmen dieser Arbeit Habitatmodellierungen mit dem Ziel durchgeführt, das (potentielle) Vorkommen ausgewählter invasiver Neophyten hinreichend genau abbilden zu können. Zum Einsatz kamen Logistische Regressionen, wobei die bisher vorhandenen Fundortdaten sowie eine Vielzahl möglicher Umweltvariablen genutzt wurden. Besonderes Augenmerk lag auf der Verwendung von Variablen, die anthropogene Strukturen möglichst gut beschreiben.

Für insgesamt sieben Arten wurden Habitatmodelle erstellt. Es zeigte sich, dass vor allem Distanz-Variablen für die Modellgüte eine große Rolle spielen. Die Information über die Entfernung zu einer entsprechenden Struktur – beispielsweise einer Straße – beschreibt das Vorkommen bzw. Fehlen einer Art oft recht genau. In Kombination mit anderen Variablen konnten statistisch gut abgesicherte Modelle hergeleitet werden.

Die vorliegenden Ergebnisse dienen dazu, invasive Neophyten im Kanton Zürich in Zukunft noch besser zu überwachen. So kann der personelle Aufwand durch gezieltere Begehungen und gegebenenfalls Bekämpfungsmaßnahmen in Zukunft reduziert und das Verständnis für die ausschlaggebenden Umweltfaktoren, die das Vorkommen oder Fehlen einer Art bedingen, erhöht werden.

## **Abstract**

Invasive neophytes are non-indigenous plant-species, introduced since 1500, which get well established in nature. In the new areas they can cause a variety of damages. Therefore, the knowledge of their actual and potential habitats is very important. Even in Switzerland many non-native species have immigrated in the last centuries. Some species are monitored, and in the canton of Zurich, there is a database in which the central species are recorded.

Species distribution models can help to explain the distribution of species. The analysed and quantified species-environment relationships represent the core of the predictive geographical modelling.

The present study investigates the potential presence and absence of selected species in the canton of Zurich with the assistance of logistic regressions. Therefore presence-data of species – as the database mentioned above – were used in combination with a variety of predictor variables.

For a total of seven neophyte species a variety of models was created. Subsequent to model fitting, one model per species was described in detail and the distribution was visualised. The selected models show a high accuracy. It was found that especially distance variables (e.g. to buildings, streets, rivers) are well suited for modelling.

---

# Inhaltsverzeichnis

<b>I</b>	<b>Einleitung</b>	<b>1</b>
1	Motivation der Arbeit .....	1
2	Struktur der Arbeit .....	3
<b>II</b>	<b>Grundlagen</b>	<b>4</b>
1	Ökologische Grundlagen.....	4
1.1	Ökologische Nische.....	4
1.2	Invasive Neophyten .....	4
1.3	Konzeptionelle Überlegungen .....	5
2	Statistische Grundlagen und methodischer Ansatz.....	5
2.1	Statistische Modellbildung .....	6
2.1.1	Daten-Transformation .....	7
2.1.2	Kollinearität.....	7
2.2	Modellanpassung .....	8
2.2.1	Schrittweise Variablen-Selektion.....	8
2.2.2	Modellgüte .....	9
2.3	Validierung .....	11
2.4	Festlegung von Schwellenwerten .....	12
<b>III</b>	<b>Material und Methoden</b>	<b>13</b>
1	Software und Werkzeuge.....	14
2	Untersuchungsgebiet .....	14
3	Verbreitungsdaten .....	16
3.1	Vorkommen .....	20
3.2	Pseudo-Absenz.....	22

<b>4</b>	<b>Erklärende Variablen.....</b>	<b>25</b>
4.1	<b>Klima-Variablen.....</b>	<b>27</b>
4.2	<b>Topografische Variablen .....</b>	<b>27</b>
4.3	<b>Physikalische Boden-Variablen .....</b>	<b>28</b>
4.4	<b>Bodenbedeckungs-Variablen .....</b>	<b>28</b>
4.5	<b>Distanz-Variablen .....</b>	<b>28</b>
4.5.1	Distanzen zu Verkehrswegen.....	29
4.5.2	Distanzen zu Siedlungsflächen .....	29
4.5.3	Distanzen zu Gewässern.....	29
<b>5</b>	<b>Datentransformation .....</b>	<b>31</b>
<b>6</b>	<b>Kollinearität .....</b>	<b>32</b>
<b>7</b>	<b>Daten-Überlagerung .....</b>	<b>34</b>
<b>8</b>	<b>Modellierung im engeren Sinne.....</b>	<b>35</b>
8.1	<b>Statistische Modellbildung .....</b>	<b>36</b>
8.2	<b>Modellanpassung.....</b>	<b>36</b>
8.3	<b>Validierung .....</b>	<b>37</b>
8.3.1	Festlegung von Schwellenwerten.....	37
<b>9</b>	<b>Visualisierung.....</b>	<b>37</b>
<b>IV</b>	<b>Ergebnisse</b>	<b>38</b>
<b>1</b>	<b>Kollinearität .....</b>	<b>38</b>
<b>2</b>	<b>Modellierung im engeren Sinne und Visualisierung .....</b>	<b>39</b>
2.1	<i>Ambrosia artemisiifolia</i> .....	40
2.2	<i>Heracleum mantegazzianum</i> .....	44
2.3	<i>Reynoutria japonica</i> .....	48
2.4	<i>Impatiens glandulifera</i> .....	52
2.5	<i>Rhus typhina</i> .....	56
2.6	<i>Solidago canadensis</i> .....	60
2.7	<i>Solidago gigantea</i> .....	64
<b>V</b>	<b>Zusammenfassung, Diskussion und Ausblick</b>	<b>70</b>

---

<b>VI</b>	<b>Literatur</b>	<b>73</b>
<b>VII</b>	<b>Anhang</b>	<b>82</b>
1	Übersichtskarte des Untersuchungsgebietes .....	82
2	Verbreitungskarten (Vorkommen/Pseudo-Absenz) .....	83
3	Variablen-Wertebereiche.....	85
3.1	Klima-Variablen.....	85
3.2	Topografische Variablen .....	86
3.3	Physikalische Boden-Variablen .....	87
3.4	Bodenbedeckungs-Variablen .....	88
3.5	Distanzen zu Siedlungsflächen .....	89
3.6	Distanzen zu Verkehrswegen .....	90
3.7	Distanzen zu Gewässern .....	91
4	Variablen-Karten.....	92
4.1	Klima-Variablen.....	92
4.2	Topografische Variablen .....	93
4.3	Physikalische Boden-Variablen .....	94
4.4	Bodenbedeckungs-Variable.....	95
4.5	Distanzen zu Siedlungsflächen .....	96
4.6	Distanzen zu Verkehrswegen .....	97
4.7	Distanzen zu Gewässern .....	99
5	Variablensets der Modelle .....	100



## Abbildungsverzeichnis

Abbildung 1: Aktivitätsdiagramm: Modellierung ausführen. Grau dargestellte Aktivitäten werden in den folgenden Abschnitten ausführlicher dargestellt.....	13
Abbildung 2: Aktivitätsdiagramm I: Untersuchungsgebiet erstellen.....	15
Abbildung 3: Punkt und Flächen-Fundortdaten der Neophyten im Kanton Zürich.....	17
Abbildung 4: Aktivitätsdiagramm II: Fundortdaten aufbereiten. ....	18
Abbildung 5: Aktivitätsdiagramm III: Fundortdaten einfügen (kartierte Vorkommen). 20	
Abbildung 6: Aktivitätsdiagramm IV: Fundortdaten einfügen (Pseudo-Absenz). ....	24
Abbildung 7: Aktivitätsdiagramm V: Distanz-Variablen berechnen.....	30
Abbildung 8: Aktivitätsdiagramm VI: Erklärende Variablen aufbereiten. Nicht aufgeführt ist die Prüfung auf Multikollinearität. ....	31
Abbildung 9: Aktivitätsdiagramm VIII: Daten überlagern.....	34
Abbildung 10: Aktivitätsdiagramm VII: Modellierung im engeren Sinne durchführen. 35	
Abbildung 11: Scatterplot-Matrix der erklärenden Variablen mit Ausnahme der Distanz-Variablen (angegeben ist der absolute Regressionskoeffizient nach Pearson). ....	38
Abbildung 12: Scatterplot-Matrix der Distanz-Variablen (angegeben ist der absolute Regressionskoeffizient nach Pearson). ....	39
Abbildung 13: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von <i>Ambrosia artemisiifolia</i> . ....	41
Abbildung 14: Vorkommenswahrscheinlichkeit von <i>Ambrosia artemisiifolia</i> (Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). ....	43
Abbildung 15: Prognostiziertes Vorkommen von <i>Ambrosia artemisiifolia</i> entsprechend des Schwellenwertes (max. TSS; Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). ....	44
Abbildung 16: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von <i>Heracleum mantegazzianum</i> . ....	45
Abbildung 17: Vorkommenswahrscheinlichkeit von <i>Heracleum mantegazzianum</i> (Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). ....	47
Abbildung 18: Prognostiziertes Vorkommen von <i>Heracleum mantegazzianum</i> entsprechend des Schwellenwertes (max. TSS; Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). ....	48
Abbildung 19: Abbildung 20: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von <i>Reynoutria japonica</i> . ....	50
Abbildung 21: Vorkommenswahrscheinlichkeit von <i>Reynoutria japonica</i> (Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). ....	51

---

Abbildung 22: Prognostiziertes Vorkommen von <i>Reynoutria japonica</i> entsprechend des Schwellenwertes (max. TSS; Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). .....	52
Abbildung 23: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von <i>Impatiens glandulifera</i> . .....	53
Abbildung 24: Vorkommenswahrscheinlichkeit von <i>Impatiens glandulifera</i> (Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). .....	55
Abbildung 25: Prognostiziertes Vorkommen von <i>Impatiens glandulifera</i> entsprechend des Schwellenwertes (max. TSS; Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). .....	56
Abbildung 26: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von <i>Rhus typhina</i> . .....	57
Abbildung 27: Vorkommenswahrscheinlichkeit von <i>Rhus typhina</i> (Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). .....	59
Abbildung 28: Prognostiziertes Vorkommen von <i>Rhus typhina</i> entsprechend des Schwellenwertes (max. TSS; Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). .....	60
Abbildung 29: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von <i>Solidago canadensis</i> . .....	61
Abbildung 30: Vorkommenswahrscheinlichkeit von <i>Solidago canadensis</i> (Modell $\text{gesamt}^2$ , Variablenselektion mit BIC). .....	63
Abbildung 31: Prognostiziertes Vorkommen von <i>Solidago canadensis</i> entsprechend des Schwellenwertes (max. TSS; Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). .....	64
Abbildung 32: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von <i>Solidago gigantea</i> . .....	65
Abbildung 33: Vorkommenswahrscheinlichkeit von <i>Solidago gigantea</i> (Modell $\text{gesamt}^2$ , Variablenselektion mit BIC). .....	68
Abbildung 34: Prognostiziertes Vorkommen von <i>Solidago gigantea</i> entsprechend des Schwellenwertes (max. TSS; Modell $\text{gesamt}^2$ , Variablen-Selektion mit BIC). .....	69
Abbildung 35: Übersichtskarte des Untersuchungsgebietes. Siedlungen (Gebäude) sind grau dargestellt, Größere Städte und Flüsse beschriftet, Datengrundlage: BFL (1994). .....	82
Abbildung 36: Verbreitungskarten der modellierten Neophyten I, Datengrundlage: BFL (1994). .....	83
Abbildung 37: Verbreitungskarten der modellierten Neophyten II, Datengrundlage: BFL (1994). .....	84
Abbildung 38: Werteverteilung der Klima-Variablen vor (links) und nach einer eventuell vorgenommenen Transformation. .....	85
Abbildung 39: Werteverteilung der Topografischen Variablen vor (links) und nach einer eventuell vorgenommenen Transformation. .....	86

---

Abbildung 40: Werteverteilung der Physikalischen Boden-Variablen vor (links) und nach einer eventuell vorgenommenen Transformation.....	87
Abbildung 41: Werteverteilung der Bodenbedeckungs-Variablen Die Ursprungskategorien der reklassifizierten Kategorien von as97r (unten) können durch die Farbgebung zugeordnet werden.. .....	88
Abbildung 42: Werteverteilung der Distanz-Variablen Siedlungsflächen vor (links) und nach einer eventuell vorgenommenen Transformation.....	89
Abbildung 43: Werteverteilung der Distanz-Variablen Verkehrsflächen I vor (links) und nach einer eventuell vorgenommenen Transformation.....	90
Abbildung 44: Werteverteilung der Distanz-Variablen Verkehrsflächen II vor (links) und nach einer eventuell vorgenommenen Transformation.....	91
Abbildung 45: Werteverteilung der Distanz-Variablen Gewässern vor (links) und nach einer eventuell vorgenommenen Transformation. ....	91
Abbildung 46: Layer-Darstellung der Klima-Variablen.....	92
Abbildung 47: Layer-Darstellung der Topografischen-Variablen.....	93
Abbildung 48: Layer-Darstellung der physikalischen Boden-Variablen.....	94
Abbildung 49: Layer-Darstellung der Bodenbedeckungs-Variable.....	95
Abbildung 50: Layer-Darstellung der Distanz-Variablen zu Siedlungsflächen. ....	96
Abbildung 51: Layer-Darstellung der Distanz-Variablen zu Verkehrswegen I.....	97
Abbildung 52: Layer-Darstellung der Distanz-Variablen zu Verkehrswegen II. ....	98
Abbildung 53: Layer-Darstellung der Distanz-Variablen zu Gewässern. ....	99

## Tabellenverzeichnis

Tabelle 1: Klassifikationsmatrix zur Kontrolle der Modellgüte (nach ALLOUCHE et al. 2006). .....	9
Tabelle 2: Für die Definition des Untersuchungsgebietes verwendeter Datensatz mit weiteren Angaben. ....	14
Tabelle 3: Erläuterungen zum Aktivitätsdiagramm I: Untersuchungsgebiet erstellen (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).....	15
Tabelle 4: Verwendete Datensätze der Fundortdaten (Datums-Angaben und Anzahl der Einträge beziehen sich auf die Fläche des Kantons Zürich, siehe auch Schritt 2 in Abbildung 2). ....	16
Tabelle 5: Erläuterungen zum Aktivitätsdiagramm II: Fundortdaten aufbereiten (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).....	19
Tabelle 6: Erläuterungen zum Aktivitätsdiagramm III: Fundortdaten einfügen (kartierte Vorkommen) (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter). 21	
Tabelle 7: Übersicht der für die Modellierung aufbereiteten Neophyten-Arten mit Angaben zur Anzahl Datensätze und daraus resultierender Gitter-Punkte (im eigentlichen Modellierungsprozess verwendete Arten sind unterstrichen). ....	22
Tabelle 8: Erläuterungen zum Aktivitätsdiagramm IV: Fundortdaten einfügen (Pseudo-Absenz) (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter). .....	24
Tabelle 9: Aufbereitete erklärende Variablen. Letztendlich für die Modellierung im engeren Sinne verwendete Variablen sind unterstrichen (siehe auch Abschnitte 5 Datentransformation und 6 Kollinearität. ....	26
Fortsetzung Tabelle 10 .....	27
Tabelle 11: Erläuterungen zum Aktivitätsdiagramm V: Distanz-Variablen berechnen (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).....	30
Tabelle 12: Erläuterungen zum Aktivitätsdiagramm VI: Erklärende Variablen aufbereiten (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).....	31
Tabelle 13: Erläuterungen zum Aktivitätsdiagramm VIII: Daten überlagern (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).....	34
Tabelle 14: Erläuterungen zum Aktivitätsdiagramm VII: Modellierung im engeren Sinne durchführen. ....	35
Tabelle 15: Formulierten Teilmodelle für die Habitatmodellierung invasiver Neophyten. ....	36
Tabelle 16: Übersicht wichtiger Kennwerte der durchgeführten Modelle zur Vorkommensvorhersage von <i>Ambrosia artemisiifolia</i> . ....	41

---

Tabelle 17: Regressionskoeffizienten und Standardfehler der Variablen für das Modell $gesamt^2$ , Variablenselektion mit BIC von <i>Ambrosia artemisiifolia</i> . .....	42
Tabelle 18: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von <i>Heracleum mantegazzianum</i> . .....	45
Tabelle 19: Regressionskoeffizienten und Standardfehler der Variablen für das Modell $gesamt^2$ , Variablenselektion mit BIC von <i>Heracleum mantegazzianum</i> . .....	46
Tabelle 20: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von <i>Reynoutria japonica</i> . .....	49
Tabelle 21: Regressionskoeffizienten und Standardfehler der Variablen für das Modell $gesamt^2$ , Variablenselektion mit BIC von <i>Reynoutria japonica</i> . .....	50
Tabelle 22: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von <i>Impatiens glandulifera</i> . .....	53
Tabelle 23: Regressionskoeffizienten und Standardfehler der Variablen für das Modell $gesamt^2$ , Variablenselektion mit BIC von <i>Impatiens glandulifera</i> . .....	54
Tabelle 24: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von <i>Rhus typhina</i> . .....	57
Tabelle 25: Regressionskoeffizienten und Standardfehler der Variablen für das Modell $gesamt^2$ , Variablenselektion mit BIC von <i>Rhus typhina</i> . .....	58
Tabelle 26: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von <i>Solidago canadensis</i> . .....	61
Tabelle 27: Regressionskoeffizienten und Standardfehler der Variablen für das Modell $gesamt^2$ , Variablenselektion mit BIC von <i>Solidago canadensis</i> . .....	62
Tabelle 28: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von <i>Solidago gigantea</i> . .....	65
Tabelle 29: Regressionskoeffizienten und Standardfehler der Variablen für das Modell $gesamt^2$ , Variablenselektion mit BIC von <i>Solidago gigantea</i> . .....	67
Tabelle 30: Variablensets der Modelle für <i>Ambrosia artemisiifolia</i> . .....	100
Tabelle 31: Variablensets der Modelle für <i>Heracleum mantegazzianum</i> .....	100
Tabelle 32: Variablensets der Modelle für <i>Reynoutria japonica</i> . .....	101
Tabelle 33: Variablensets der Modelle für <i>Impatiens glandulifera</i> . .....	101
Tabelle 34: Variablensets der Modelle für <i>Rhus typhina</i> . .....	102
Tabelle 35: Variablensets der Modelle für <i>Solidago canadensis</i> . .....	102
Tabelle 36: Variablensets der Modelle für <i>Solidago gigantea</i> . .....	103

## Abkürzungs-, Variablen- und Symbolverzeichnis

	Dateinamen, Code, Befehle
	<i>Ambrosia artemisiifolia</i> - wissenschaftliche Artnamen
.	Dezimaltrennzeichen
'	Symbol für Zifferngruppierung
AIC	Akaikes Informationskriterium (siehe S. 8)
as97r	Bezeichnung für das Teilmodell „Arealstatistik 97, reklassifiziert“, siehe Tabelle 15 (S. 36)
AUC	Area under the ROC-Curve (siehe S. 9)
BIC	Bayessches Informationskriterium (siehe S. 8)
boden	Bezeichnung für das Teilmodell „Boden“, siehe Tabelle 15 (S. 36)
engl.	englisch
gesamt	Bezeichnung für das Teilmodell „gesamter Datensatz“, siehe Tabelle 15 (S. 36)
gesamt^2	Bezeichnung für das Teilmodell „gesamter Datensatz, mit quadratischen Termen“, siehe Tabelle 15 (S. 36)
gew	Bezeichnung für das Teilmodell „Gewässer“, siehe Tabelle 15 (S. 36)
GLM	generalised linear model (siehe S. 6)
indet.	indeterminabel, unbestimmbar
klima	Bezeichnung für das Teilmodell „Klima“, siehe Tabelle 15 (S. 36)
korr. D <sup>2</sup>	korrigierte Devianz (siehe S. 10)
KV	Kreuzvalidierung (siehe S. 11)
ln	natürlicher Logarithmus
m. $\kappa$ /TSS	maximales Cohen's Kappa bzw. TSS
max.	maximal
n	Anzahl Records/Stichprobengröße
prog.	prognostiziert
ROC	Receiver Operating Characteristic (siehe S. 9)
topo	Bezeichnung für das Teilmodell „Topografie“, siehe Tabelle 15 (S. 36)
trans	Bezeichnung für das Teilmodell Transportwege, siehe Tabelle 15 (S. 36)
TSS	True skill statistic (siehe S. 10)
urban	Bezeichnung für das Teilmodell „Gebäude“, siehe Tabelle 15 (S. 36)
WSL	Eidg. Forschungsanstalt für Wald, Schnee und Landschaft
$\kappa$	Cohen's Kappa (siehe S. 9)

Gebräuchliche Abkürzungen nach deutscher Rechtschreibung (DUDENREDAKTION 2001) sind nicht aufgeführt.

---

# I Einleitung

## 1 Motivation der Arbeit

In der jüngeren Geschichte ist es im Zuge menschlicher Aktivitäten – gewollt oder ungewollt – immer wieder zur Einführung oder Einwanderung von Pflanzen in fremde Gebiete gekommen. Gelingt es den Pflanzen dabei, in einem bestimmten geografischen Gebiet dauerhaft wildlebende Bestände auszubilden und liegt ihr Auftreten nach dem Jahr 1500 unserer Zeitrechnung, werden sie als Neophyten bezeichnet (BÄBLER et al. 1996). Einige dieser Neophyten können verschiedenste Probleme verursachen, man spricht dann von invasiven Neophyten. GIGON & WEBER (2005) verstehen darunter solche Arten, die „leicht verwildern, sich effizient ausbreiten und in den Bereichen Biodiversität, Gesundheit und/oder Ökonomie Schäden verursachen, sowie Neophyten, die ein solches Potential besitzen“. Um Schäden vorzubeugen oder Risikoabschätzungen und Bekämpfungsmaßnahmen zielgerichtet durchführen zu können, müssen die Vorkommen der Arten bekannt sein. Vielfach handelt es sich um sich noch ausbreitende Arten. Deshalb ist es nicht ausreichend, nur die aktuell bekannten Fundorte zu kennen. Vielmehr sollten auch die potentiellen Standorte in den Analysen berücksichtigt werden.

Habitatmodelle (engl. species distribution models) – Modelle zur räumlichen Vorhersage von Arten oder Artengemeinschaften (GUISAN & ZIMMERMANN 2000, SCHRÖDER 2004, HIRZEL et al. 2006) – können hierfür einen wichtigen Beitrag leisten. Sie verschneiden punkthafte Beobachtungsdaten der Arten mit flächenhaft vorliegenden Daten der Habitatausprägung (LANG & BLASCHKE 2007) und setzen so das Vorkommen von Arten in einem bestimmten Gebiet mit den dort herrschenden Umweltbedingungen (abiotische und biotische Lebensraumfaktoren) in Beziehung (PEARCE & BOYCE 2006). Durch den Begriff „Modell“ wird jedoch deutlich, dass immer nur eine unvollständige Abbildung der realen Beziehungen zwischen Arten und ihrer Umwelt möglich ist (KLEYER et al. 1999/2000). Die Quantifizierung von Art-Umwelt-Beziehungen stellt den Kern der prognostizierenden räumlichen Modellierung dar (GUISAN & ZIMMERMANN 2000). Aktuelle Beispiele für Habitatmodellierungen für Gebiete unterschiedlicher Ausdehnung finden sich in NIEDERBERGER (2008), GRAUTE (2009) und HERRMANN (2009).

Auch in der Schweiz breiten sich invasive Neophyten zunehmend aus. Sie werden deshalb in Listen geführt und speziell überwacht (vgl. GIGON & WEBER 2005). Für den Kanton Zürich liegen – dank einfacher Datenerfassung mit Hilfe eines Web GIS – zahlreiche Fundortdaten invasiver Neophyten vor (siehe GIS-ZENTRUM KANTON ZÜRICH 2006).

Ziel der vorliegenden Arbeit ist es, durch Verwendung der oben genannten Fundortdaten und mit Hilfe der Habitatmodellierung zu prüfen, ob potentielle Standorte invasiver Neophyten für den Kanton Zürich mit hinreichender Sicherheit beschrieben werden können. In diesem Zusammenhang stellen sich folgende Fragen:

- Gibt es Umweltvariablen, die einen besonders großen Einfluss auf das Vorkommen der Arten haben – und wenn ja, welche?
- Unterscheidet sich das Set der erklärenden Variablen zwischen den Arten?
- Wie gut sind die Modelle im Hinblick auf statistische Prüfgrößen?
- Können anhand der Ergebnisse Aussagen über das zukünftige Ausbreitungspotential gemacht werden?
- Können anhand der Umweltvariablen eventuell der Ursprung der Ausbreitung oder Ausbreitungswege erkannt werden?

Dabei sind die Ergebnisse dieser Arbeit für unterschiedliche Nutzergruppen relevant. Als Beispiele seien an dieser Stelle genannt:

- Naturschutzämter und -verbände angesichts der Problematik der Verdrängung (gefährdeter) einheimischer Arten
- Gesundheitsinstitutionen wegen möglicher Allergie-Auslöser
- Hochwasserschutz und Landwirtschaft infolge ökonomischer Schäden

Diesen Nutzergruppen sollen Informationen in Form von Kartenmaterial und Zusammenstellungen relevanter Umweltvariablen zur Verfügung gestellt werden, die eine bessere Risikoabschätzung, Kontrolle und gegebenenfalls gezielte Bekämpfungsmaßnahmen ermöglichen.

Eine detaillierte biologisch-ökologische Interpretation der Ergebnisse kann diese Arbeit jedoch nicht leisten. Das wirkt sich auch auf den entsprechenden Literaturüberblick und die abschließende Diskussion aus. Genauso wenig kann die Arbeit die ganze Thematik und Vielfalt der Habitatmodellierung darstellen. Die Ausführungen beschränken sich deshalb auf die im aktuellen Zusammenhang genutzten Techniken.



## **2 Struktur der Arbeit**

Nach diesem einführenden Abschnitt gliedert sich die Arbeit in sechs weitere Kapitel. Das zweite Kapitel stellt sowohl die ökologischen als auch statistischen Grundlagen vor und gibt einen Literaturüberblick. Das anschließende Kapitel bespricht die Methodik, die im Rahmen der Master Thesis verwendet wurde. Da das zweite Kapitel bereits viele Informationen gibt und um Redundanzen möglichst zu vermeiden, sind entsprechende Abschnitte kurz gehalten. Das vierte Kapitel stellt die Ergebnisse der Modellierung für jede Art separat dar. Kapitel fünf enthält Diskussion und Ausblick. Daran schließen sich das Literaturverzeichnis und der Anhang an.

Abschließend sei noch kurz auf die „Problematik“ englischen Fachvokabulars hingewiesen. Häufig haben sich englische Ausdrücke so eingebürgert und sind heute derart gebräuchlich, dass auf eine Übersetzung verzichtet wird.

---

## **II Grundlagen**

### **1 Ökologische Grundlagen**

#### **1.1 Ökologische Nische**

Bevor in diesem Abschnitt näher auf die Ökologie invasiver Neophyten eingegangen wird, sei einleitend ein im Zusammenhang mit Habitatmodellierungen wichtiges zugrunde liegendes Konzept angesprochen. Es handelt sich um das Konzept der ökologischen Nische (GUISAN & THULLER 2005). Dieses von HUTCHINSON (1959) spezifizierte Konzept beschreibt das Vorkommen einer Art als n-dimensionalen Hyperraum, wobei n die Anzahl der Umweltfaktoren bezeichnet (TOWNSEND et al. 2003). Die ökologische Nische einer Art wird durch die Grenzen definiert, innerhalb derer sie vorkommen kann (TOWNSEND et al. 2003). Dabei sind die Lebensbedingungen für eine gegebene Art innerhalb der einzelnen Toleranzbereiche aber nicht konstant, sondern können sich ändern. Beispielsweise kann eine Art in einem weiten Temperaturbereich vorkommen, ihr Reproduktionsoptimum liegt aber im mittleren Temperaturbereich. Diese Muster sollten in der Habitatmodellierung möglichst abgebildet werden können.

#### **1.2 Invasive Neophyten**

Die Ausbreitung von gebietsfremden Arten ist ein weltweites Phänomen (VITOUSEK et al. 1997, MACK et al. 2000, HULME 2006, 2009) und findet auch in Europa große Beachtung (z.B. PYSEK et al. 2002, z.B. CHYTRY et al. 2009, WILLIAMSON et al. 2009). Für die Schweiz liegen ebenfalls Studien über Neophyten mit unterschiedlichen Fragestellungen vor (GELPKE & WEBER 2005, BOHREN et al. 2006, NOBIS et al. 2009).

Die Herkunftsgebiete europäischer Neophyten liegen auf allen Kontinenten, mit Ausnahme der Antarktis (KOWARIK 2003). Viele Arten stammen aus Asien oder Nordamerika. Einigen Arten gelang und gelingt es, sich schnell auszubreiten. Diese Neophyten zeichnen sich durch eine starke Ausbreitungsfähigkeit über Samen, Wurzeln bzw. über das Rhizomsystem und rasches Höhenwachstum aus (HARTMANN et al. 1995). Die gleichen Autoren geben als örtliche Ausbreitungsmedien Wasser, Luft, Tiere sowie den Menschen an. Im letzteren Fall geschieht dies hauptsächlich indirekt über Transporte und Ablagerungen von Erdmaterial, Bauschutt und Deponiegut (HARTMANN et al. 1995). Allgemein kann festgestellt werden, dass überwiegend Standorte besiedelt werden, die von Menschen in der einen oder anderen Form geschaffen oder stark verändert wurden (HARTMANN et al. 1995), sich in einem (frühen) Sukzessionsstadium

befinden oder degenerierte Pflanzengesellschaften aufweisen (FALINSKI 1998). Häufig werden auch Bereiche entlang von Fließgewässern besiedelt (MASKELL et al. 2006), wobei diese in Mitteleuropa vielfach ebenfalls nur noch gestörte und degenerierte Pflanzengesellschaften aufweisen.

### **1.3 Konzeptionelle Überlegungen**

Die hier nur kurz beschriebenen Merkmale und Charakteristika der Neophyten gilt es in einem statistischen Ansatz sinnvoll zu integrieren. Wichtig ist hierbei vor allem die Auswahl geeigneter erklärender Variablen. Da viele invasive Neophyten auf in der einen oder anderen Weise anthropogen beeinflussten Standorten vorkommen, müssen entsprechende Variablen genutzt werden. Neben Variablen, die beispielsweise die Klima- und Boden-Verhältnisse wiedergeben, sollte der Einfluss von Distanz-Variablen geprüft werden. Diese geben über Kodierung der Entfernung des Neophyten-Fundortes von einer entsprechenden Struktur (z.B. Gebäude, Straße, Fluss) detailliert Auskunft über mögliche Standorte bzw. Standortbedingungen aber auch über Ausbreitungswege. Der Einsatz derartiger Variablen ist umso angebrachter, da durch die geringe räumliche Ausdehnung des Untersuchungsgebietes kaum größere Wertebereiche bei den klassischen erklärenden Variablen (z.B. Temperatur) zu erwarten sind.

## **2 Statistische Grundlagen und methodischer Ansatz**

Der folgende Abschnitt beschreibt kurz den prinzipiellen Ablauf einer Art-Habitat-Modellierung. Ausführliche Darstellungen zu dieser Thematik finden sich in GUIBAN & ZIMMERMANN (2000), GUIBAN et al. (2002), GUIBAN & THULLER (2005) sowie den Arbeiten von SCHRÖDER & REINEKING (2004a, 2004b) bzw. REINEKING & SCHRÖDER (2004a, 2004b).

Die hier besprochenen theoretischen Grundlagen greift Abschnitt III - Material und Methoden zu einem großen Teil auf. Lediglich speziellere Aspekte werden dort ausführlicher behandelt.

## 2.1 Statistische Modellbildung

Für die Habitatmodellierung sind heute eine Vielzahl unterschiedlicher Techniken verfügbar, auf deren Darstellung hier verzichtet werden soll. Detaillierte Übersichten und Vergleiche finden sich in GUI SAN & ZIMMERMANN (2000), MOISEN & FRESCINO (2002), SEGURADO & ARAUJO (2004), GUI SAN & THUILLER (2005), ELITH et al. (2006) oder AUSTIN (2007). Im Folgenden wird jedoch das für die Bearbeitung der Fragestellungen dieser Arbeit ausgewählte Verfahren der Logistischen Regression (engl. logistic regression model, LRM) beschrieben und die Gründe für diese Entscheidung kurz zusammengefasst.

- Mit Hilfe von LRMs ist es möglich, binomial verteilte Response-Variablen<sup>1</sup> zu modellieren. Bei Habitatmodellierungen entspricht dies den Vorkommen/Absenz-Daten (engl. presence/absence)<sup>2</sup>.
- Da LRMs sich von Verallgemeinerten Linearen Modellen (engl. generalised linear model, GLM) ableiten, erlauben sie Nichtlinearität sowie heterogene Varianzen (Hastie & Tibshirani 1990).
- LRMs sind auch bei Nichteinhaltung statistischer Voraussetzungen robust und stabil (ZARNETSKE et al. 2007).
- Es können erklärende Variablen unterschiedlichster Datentypen (kontinuierliche, kategoriale, binäre) gemeinsam modelliert werden (HILBE 2009).
- Es kann die Vorkommenswahrscheinlichkeit einer Art als eine (einfache) Funktion von erklärenden Variablen berechnet werden (PEETERS & GARDENIERS 1998), was die ökologische Interpretierbarkeit erhöht (SCHRÖDER 2000).

Die Logistische Regression ist die meist genutzte statistische Methode, um binäre Response-Variablen zu analysieren (HILBE 2009). Sie wird auch in der Habitatmodellierung am häufigsten verwendet (SCHRÖDER & REINEKING 2004a) und stellt heute den Standardansatz für derartige Modelle dar (KLAR et al. 2008).

---

<sup>1</sup> Der Begriff der Response-Variable steht in der Habitatmodellierung für die Variable der Verbreitungsdaten (abhängige Variable), Prädiktor-Variablen beschreiben die Habitateigenschaften (unabhängige Variablen).

<sup>2</sup> Echte Absenz-Daten sind aufgrund der Datenlage nicht vorhanden. Da jedoch Modelle, die Vorkommen/Absenz-Daten nutzen, das Vorkommen von Arten genauer vorhersagen können als Modelle, die nur Vorkommen-Daten verwenden (CHEFAOUI & LOBO 2008), werden sogenannte Pseudo-Absenzen generiert. Zur deren Bildung siehe Abschnitt III 3.2 - Pseudo-Absenz, S. 22.

### 2.1.1 Daten-Transformation

Für die beabsichtigte Art der Modellierung mittels Logistischer Regression sind die erklärenden Variablen idealerweise normalverteilt und die Beziehungen untereinander möglichst linear (QUINN & KEOUGH 2002, MAINDONALD & BRAUN 2007). Um diese Bedingungen zu erreichen, können die Original-Daten in eine andere Mess-Skala überführt werden. Hierfür kommen – je nach Art der Variablen – unterschiedliche Transformationen in Betracht. Geeignete Transformationen für die auch in dieser Arbeit verwendeten meist rechtsschief verteilten Variablen sind die Quadratwurzel- oder Logarithmus-Transformation, bei linksschief verteilten Variablen die Potenzierung (siehe beispielsweise SOKAL & ROHLF 1995, QUINN & KEOUGH 2002, KUTNER et al. 2004, SHEATHER 2009).

### 2.1.2 Kollinearität

Sind die erklärenden Variablen untereinander korreliert – was bei biologisch-ökologischen Daten fast immer der Fall ist – spricht man davon, dass die Daten durch Kollinearität beeinflusst sind (QUINN & KEOUGH 2002, GRAHAM 2003, KUTNER et al. 2004). Der Begriff Multikollinearität wird dabei häufig dann verwendet, wenn die Korrelation besonders hoch ist (KUTNER et al. 2004, SHEATHER 2009). Multikollinearität macht die Interpretation von Regressionskoeffizienten schwierig (FARAWAY 2006). Die Effekte der erklärenden Variablen auf die Response-Variable können nicht mehr eindeutig unterschieden werden. So kann es vorkommen, dass das gesamte Modell einen höchst signifikanten Einfluss der betrachteten Variablen ergibt, aber nur sehr wenige Regressionskoeffizienten statistisch signifikant sind (SHEATHER 2009). Im Fall der schrittweisen Variablen-Selektion (siehe Abschnitt 2.2.1, S. 8 ff.) können dadurch beispielsweise erklärende Variablen ausgeschlossen werden, weil andere Variablen, die einen geringeren Erklärungsgehalt besitzen, mit diesen korreliert sind und im Modell verbleiben (MAC NALLY 2000).

Um Multikollinearität aufzudecken und gegebenenfalls auszuschließen, gibt es verschiedene Ansätze (für eine Übersicht siehe beispielsweise GRAHAM 2003, KUTNER et al. 2004, DE FRUTOS et al. 2007, DORMANN et al. 2008). Die in dieser Arbeit verwendeten Ansätze werden – der Übersichtlichkeit halber – im Abschnitt III 6 - Kollinearität (S. 32) ausführlicher beschrieben.

## 2.2 Modellanpassung

In dieser Phase der Modellierung wird die Übereinstimmung zwischen den Beobachtungen und den prognostizierten Werten geprüft und das statistische Modell kalibriert (SCHRÖDER 2000, GUIBAN & ZIMMERMANN 2000). Durch Anpassung der Modellparameter soll die Güte des Modells optimiert werden.

### 2.2.1 Schrittweise Variablen-Selektion

Methoden der Variablen-Selektion haben das Ziel, die Teilmenge an erklärenden Variablen zu wählen, die für ein gegebenes Modell optimal ist (GUIBAN & ZIMMERMANN 2000, SHEATHER 2009). Generell gilt, je mehr erklärende Variablen in ein Modell einfließen, umso geringer ist der Fehler der Vorhersage, desto größer ist aber auch die Varianz. SHEATHER (2009) gibt zwei Gründe an, die für eine Variablen-Reduktion sprechen. Als erstes nennt er das Problem, welches als Überanpassung (engl. overfitting) bekannt ist. Das Modell ist so gut an die Trainingsdaten angepasst, dass zufällige Muster in diesen Daten mit abgebildet werden (HARRELL et al. 1996). Der zweite Punkt betrifft die Interpretierbarkeit des Modells. Mit zunehmender Zahl von Variablen fällt eine sinnvolle (ökologische) Interpretation schwerer.

Ausführliche Übersichten möglicher Kriterien zur Evaluierung von Subsets der erklärenden Variablen finden sich in REINEKING & SCHRÖDER (2004b), HILBE (2009) und SHEATHER (2009). Im Folgenden sollen lediglich die verwendeten Kriterien zur Variablen-Reduktion betrachtet werden.

#### *Akaike Informationskriterium*

Das Akaike Information Criterion (AIC) ist das in der Habitatmodellierung am häufigsten genutzte Kriterium, um Modelle zu vergleichen (RICHARDS 2005) und wird auch als Prüfkriterium zur schrittweisen Variablen-Selektion verwendet (z.B. GUIBAN et al. 2002, MAGGINI et al. 2006, ARAUJO & LUOTO 2007). Es stellt einen Kompromiss zwischen einerseits möglichst guter Modellanpassung und andererseits möglichst geringer Modellkomplexität dar (RICHARDS 2005, SHEATHER 2009). Dabei gilt, je kleiner der AIC-Wert innerhalb der gleichen Datengrundlage ist, desto besser ist das Modell.

#### *Bayessches Informationskriterium*

Das Bayesian Information Criterion (BIC) ist gegenüber AIC das strengere Selektionskriterium und führt typischerweise zu Modellen mit weniger erklärenden Variablen (REINEKING & SCHRÖDER 2006, MAINDONALD & BRAUN 2007). Unterschiede ergeben sich vor allem bei größeren Stichprobenumfängen, da beim BIC die Stichprobengröße in den „Strafterm“ einfließt, während dies bei AIC nicht der Fall ist<sup>3</sup> (gesamte Formeln siehe beispielsweise KUTNER et al. 2004).

<sup>3</sup> „Strafterm“: BIC:  $\ln(n) \cdot p$ ; AIC:  $2 \cdot p$  ( $p$  = Anzahl der verwendeten erklärender Variablen).

### 2.2.2 Modellgüte

Heute existiert in der Literatur eine Vielzahl an Gütemaßen zur Beurteilung von Habitatmodellen (Übersichten finden sich in FIELDING & BELL 1997, GUIBAN & ZIMMERMANN 2000, SCHRÖDER & REINEKING 2004b). Aus dieser großen Fülle sollen im Folgenden nur die in dieser Arbeit verwendeten Gütemaße beschrieben werden. Basis vieler Maße ist die in Tabelle 1 dargestellte Klassifikationsmatrix. Sie gibt die möglichen Kombinationen aus Vorkommen und Vorhersage bei binären Habitatmodellen wieder.

Tabelle 1: Klassifikationsmatrix zur Kontrolle der Modellgüte (nach ALLOUCHE et al. 2006).

		Daten	
		Vorkommen	Absenz
Modell	Vorkommen	korrekte Vorhersage des Vorkommens (a)	falsche Vorhersage des Vorkommens (b)
	Absenz	falsche Vorhersage der Absenz (c)	korrekte Vorhersage der Absenz (d)

$n = a + b + c + d$

#### *Fläche unter der Receiver Operating Characteristic Kurve*

In einem Receiver Operating Characteristic (ROC)-Diagramm wird der Anteil der korrekt klassifizierten Vorkommen (Sensitivität, engl. sensitivity) gegen den 1 - Anteil der korrekt vorhergesagten Absenzen (Spezifizität, engl. specificity) für alle möglichen Schwellenwerte aufgetragen (HANLEY & MCNEIL 1982, FIELDING & BELL 1997, für ausführliche Darstellungen siehe FAWCETT 2004, FAWCETT 2006, LASKO et al. 2005).<sup>4</sup>

Der Flächeninhalt unter der ROC-Kurve wird als Area under the ROC-Curve (AUC) bezeichnet. Er stellt ein wichtiges Maß dar, da er die Gesamtgüte unabhängig von einem Schwellenwert angibt (DELEO 1993, zitiert in FIELDING & BELL 1997). Der AUC kann Werte von 0.5 bis 1 aufweisen, wobei 0.5 keine bessere Modellprognose gegenüber einer Zufallsziehung bedeutet (FIELDING 2002). Nach MANEL et al. (2001) kennzeichnet ein AUC von 0.5 bis 0.7 geringe Modellgüte, ein AUC von 0.7 bis 0.9 ein brauchbares Modell und Werte über 0.9 indizieren ein Modell mit hoher Genauigkeit.

<sup>4</sup> Die Kodierung aus Tabelle 1 übernehmend, bedeutet dies:

$$\text{Sensitivität} = \frac{a}{(a + c)} \text{ und Spezifizität} = \frac{d}{(b + d)} .$$

### ***Cohen's Kappa***

Das gebräuchlichste Maß zur Abschätzung der Güte von Vorkommen/Absenz Vorhersagen ist Cohen's Kappa ( $\kappa$ ) (ALLOUCHE et al. 2006). Es gibt die normierte Differenz zwischen dem Anteil beobachteter korrekter Vorhersagen und dem Anteil erwarteter korrekter Vorhersagen an und kann Werte zwischen -1 und 1 annehmen (REINEKING & SCHRÖDER 2004a). Die Formel enthält die drei Parameter Prävalenz, Sensitivität und Spezifität<sup>5</sup>, die aus der Klassifikationsmatrix (Tabelle 1) berechnet werden können (für die genaue Formel siehe beispielsweise ALLOUCHE et al. 2006). MONSERUD & LEE-MANS (1992) schlagen folgende Bewertungen der Modellgüte für Cohen's Kappa vor:

<0.05	keine
0.05 - 0.20	sehr gering
0.20 - 0.40	gering
0.40 - 0.55	ausreichend
0.55 - 0.70	gut
0.70 - 0.85	sehr gut
0.85 - 0.99	exzellent
0.99 - 1.00	perfekt

Ein wichtiger Kritikpunkt an  $\kappa$  betrifft seine Abhängigkeit von der Prävalenz, also dem Verhältnis von Vorkommen zur Gesamtstichprobe (z.B. LANTZ & NEBENZAHL 1996, MCPHERSON et al. 2004). Kappa-Werte sind deshalb über verschiedene Studien hinweg selten vergleichbar (REINEKING & SCHRÖDER 2004a).

### ***True skill statistic***

Aufgrund der für Cohen's Kappa beschriebenen Einschränkungen schlugen ALLOUCHE et al. (2006) die sogenannte true skill statistic (TSS) als Gütemaß für die Habitatmodellierung vor. Sie wird nicht mehr von der Prävalenz beeinflusst, ist aber sonst mit Cohen's Kappa vergleichbar (ALLOUCHE et al. 2006).<sup>6</sup>

### ***Korrigierte Devianz***

Die korrigierte Devianz (korr.  $D^2$ ) ist ein ideales Maß, um Modelle, die unterschiedliche Variablen-Kombinationen enthalten, miteinander zu vergleichen (GUISAN & ZIMMERMANN 2000). In die korrigierte Devianz fließen neben der eigentlichen Devianz auch die Stichprobengröße und die Anzahl erklärender Variablen mit ein. Die Devianz als solche gibt an, wie nah ein Modell an ein perfektes Modell heranreicht. Bei einer perfekten Anpassung nimmt  $D^2$  den Wert 1 an. Der Wert von korr.  $D^2$  nimmt mit

<sup>5</sup> Die Kodierung aus Tabelle 1 übernehmend, bedeutet dies:

Prävalenz =  $\frac{(a + c)}{n}$ , für die beiden anderen Parameter siehe Fußnote 4.

<sup>6</sup> TSS = Sensitivität + Spezifität - 1



zunehmender Stichprobengröße oder mit abnehmender Anzahl erklärender Variablen zu (GUISAN & ZIMMERMANN 2000, KUTNER et al. 2004). Mehr Informationen und Formeln für Devianz und korrigierte Devianz finden sich beispielsweise in FARAWAY (2006), KUTNER et al. (2004) oder GUISAN & ZIMMERMANN (2000).

### *Akaikes und Bayessesches Informationskriterium*

AIC und BIC wurden schon im Abschnitt 2.2.1 Schrittweise Variablen-Selektion (S. 8 ff.) besprochen. Sie eignen sich auch, um Teilmodelle untereinander zu vergleichen. Modellen mit niedrigen AIC- bzw. BIC-Werten sollte der Vorzug gegeben werden.

## **2.3 Validierung**

Die Modellvalidierung dient dazu, eine unverzerrte, nicht optimistische Schätzung der Modellgüte zu erhalten und erlaubt somit Aussagen, ob und mit welcher Güte die Prognosen des Modells auf neue Fälle übertragen werden können (REINEKING & SCHRÖDER 2004a). Während der Validierung werden die Habitatmodelle auf Grundlage von Stichproben, die nicht zur Modellschätzung verwendet wurden, bewertet (SCHRÖDER & REINEKING 2004b). Die Modelle können sowohl extern als auch intern validiert werden (GUISAN & ZIMMERMANN 2000). Im ersten Fall stehen zwei unabhängige Datensätze zur Verfügung, die zur Modellanpassung (Trainingsdaten) bzw. Validierung (Testdaten) genutzt werden. Wenn kein ausreichend großer Datensatz vorhanden ist, um eine sinnvolle Auftrennung in Test- und Trainingsdatensatz zu gewährleisten, kann auch eine interne Validierung, beispielsweise in Form einer Kreuzvalidierung, durchgeführt werden (MAINDONALD & BRAUN 2007). In diesem Fall wird nur ein einziger Datensatz verwendet, der beispielsweise bei einer zehnfachen Kreuzvalidierung zufällig in zehn gleich große Teilmengen geteilt wird. Neun Teile dieser Daten dienen als Trainingsdaten während der verbleibende Teil als Testdatensatz verwendet wird. Diese Prozedur wird zehnmal wiederholt, so dass jede Beobachtung einmal zum Testen des Modells gedient hat (STEYERBERG et al. 2001).

## **2.4 Festlegung von Schwellenwerten**

Um die errechneten Vorkommenswahrscheinlichkeiten in eine binäre Verteilung (Vorkommen/Absenz) umzuwandeln und somit das prognostizierte Vorkommen der Arten zu klassifizieren, ist ein Schwellenwert erforderlich. Auch hier gibt es wieder eine Vielzahl von Berechnungsvarianten, um den für die jeweilige Fragestellung geeigneten Schwellenwert zu bestimmen (siehe beispielsweise LIU et al. 2005). In dieser Arbeit wurden die Schwellenwerte über den maximalen Cohen's Kappa-Wert bzw. über den maximalen TSS-Wert bestimmt (siehe auch entsprechende Ausführungen im Abschnitt 2.2.2, ab S. 10).

### III Material und Methoden

Der Arbeitsablauf wird mit Hilfe von Ablaufdiagrammen veranschaulicht. Diese bestehen aus einem Diagramm zur grafischen Darstellung der Abfolgeschritte sowie einer Tabelle mit einer kurzen Beschreibung der einzelnen Aktivitäten. Eine Übersicht des Gesamtablaufes findet sich in Abbildung 1. Auf die Wiedergabe des Programmiercodes wird aus Platzgründen dagegen verzichtet.

Einen Großteil der zugrunde liegenden Vorüberlegungen legt schon Abschnitt II Grundlagen (siehe S. 4 ff.) im Zuge der allgemeinen Einführung dar. Um Redundanzen zu vermeiden, beschreiben die folgenden Abschnitten den Ablauf der Modellierung und verzichten weitgehend auf theoretische Ausführungen und Literaturzitate.

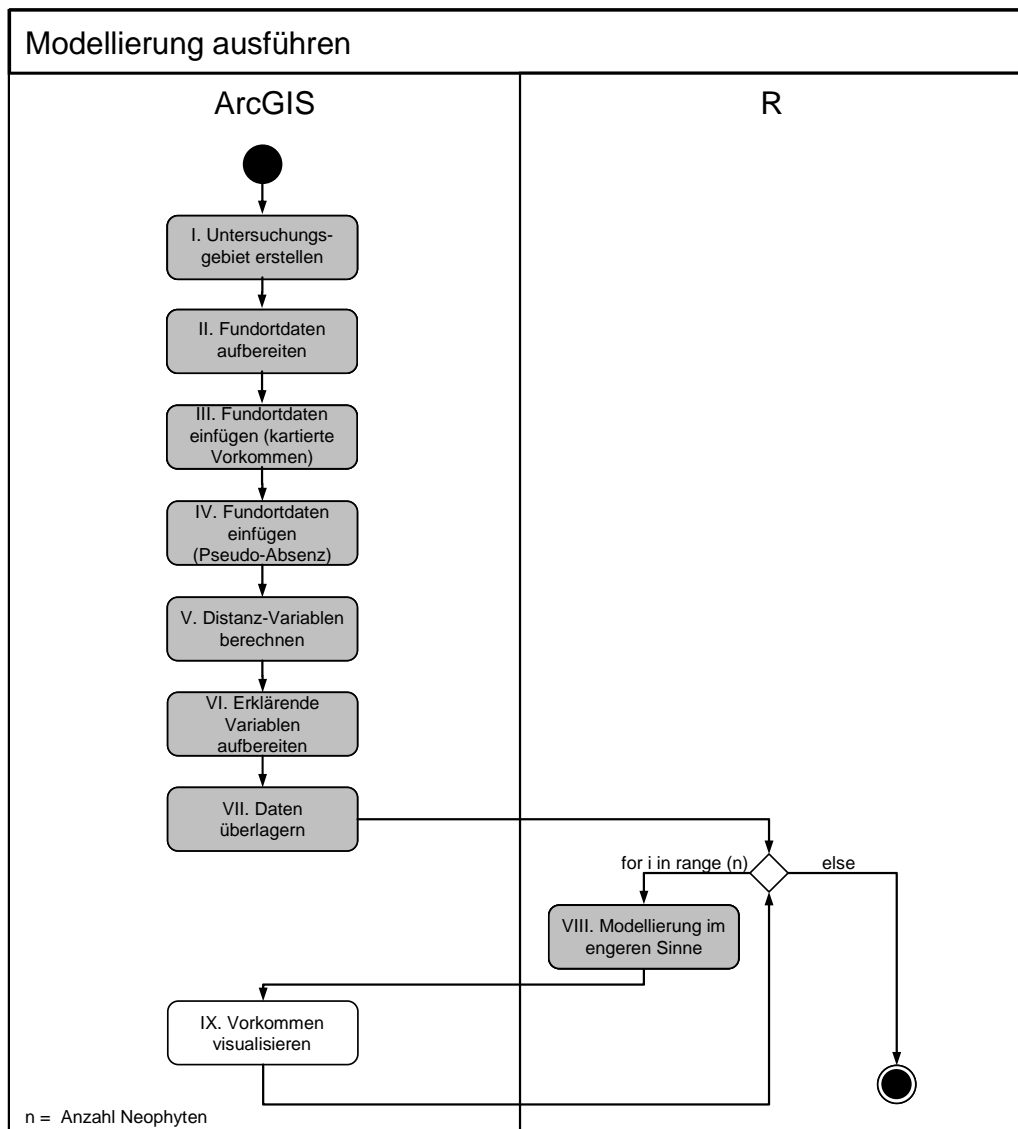


Abbildung 1: Aktivitätsdiagramm: Modellierung ausführen. Grau dargestellte Aktivitäten werden in den folgenden Abschnitten ausführlicher dargestellt.

## 1 Software und Werkzeuge

Sämtliche Arbeiten wurden mit Hilfe des GIS-Programms ArcGIS Desktop (Version 9.3, License Type: ArcView) (ESRI o.J.) und des Statistik-Programms R (Version 2.8.1) (THE R DEVELOPMENT CORE TEAM 2008) ausgeführt. Für beide Programme wurden neben den Basisinstallationen – je nach Bedarf – zusätzliche Erweiterungen (Extensions, Packages) verwendet. Diese werden in den jeweiligen Abschnitten erwähnt. Automatisierungen innerhalb von ArcGIS wurden mit Hilfe von Scripten in Python (Version 2.5.1) (PYTHON SOFTWARE FOUNDATION 1990-2010) realisiert. Die Farbgebung des Kartenmaterials orientiert sich an den Vorschlägen von BREWER & HARROWER (2010).

## 2 Untersuchungsgebiet

Das Untersuchungsgebiet umfasst den Kanton Zürich, Schweiz. Für diese Arbeit wurde auf eine generalisierte Version der Kantonsgrenze zurückgegriffen. Aufgrund der dadurch erzielten Vereinfachung der Geometrie (Vereinfachung der Grenzlinie) konnten für die Aufgabenstellung unnötige Informationen ausgeschlossen werden. Die verwendete räumliche Abgrenzung basiert auf den Angaben des Bundesamtes für Statistik (Tabelle 2).

Für die weitere Datenaufbereitung wurde innerhalb des Untersuchungsgebietes sowohl ein Punkte- als auch ein Polygon-Gitter generiert (Abbildung 2 und Tabelle 3)

*Tabelle 2: Für die Definition des Untersuchungsgebietes verwendeter Datensatz mit weiteren Angaben.*

Datensatz	Beschreibung	Fläche [km <sup>2</sup> ] <sup>a</sup>	Quelle
G1K01.shp	Kanton Zürich = ID 5, Generalisierungsstufe 1, Jahrgang 2001 (ausführliche Datenbeschreibung, siehe BFS 2007b)	1661.12	(BFS 2001)

<sup>a</sup> exklusive des Zürich- und Greifen-Sees, beachte auch die Generalisierungsstufe

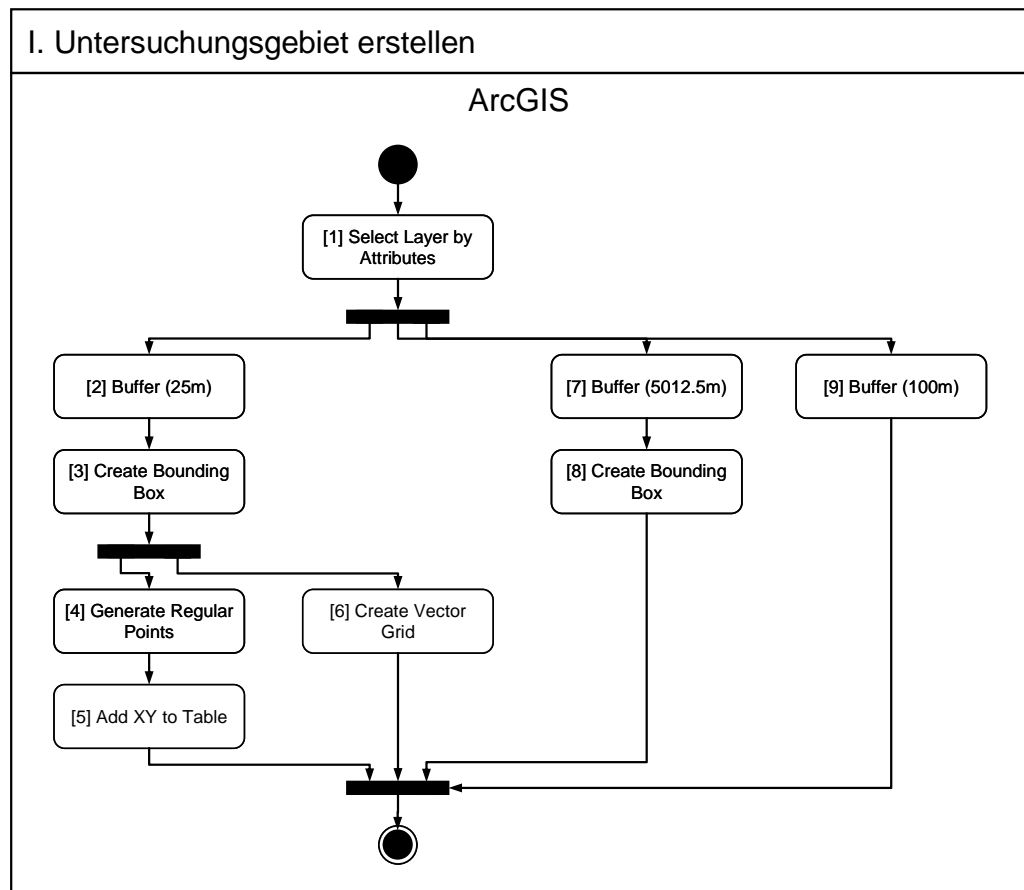


Abbildung 2: Aktivitätsdiagramm I: Untersuchungsgebiet erstellen.

Tabelle 3: Erläuterungen zum Aktivitätsdiagramm I: Untersuchungsgebiet erstellen (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).

Aktivität	Beschreibung
[1] Select Layer by Attributes	Kanton Zürich selektieren; Datenfile: G1K01.shp; Name = 'Zürich'; Data Management Tools
[2] Buffer (25m)	Puffer um Kanton Zürich erstellen, um nachfolgend vollständige Gitter-Abdeckung zu erreichen (angestrebte Gittergröße für Modellierung: 25m); linear unit: 25m, side type: full; Analysis Tools
[3] Create Bounding Box	Bounding-Box um den Puffer erstellen; Marine Geospatial Ecology Tools 0.7a15 (ROBERTS et al. in review)
[4] Generate Regular Points	Punkte-Gitter erstellen; point spacing X/Y: 25; Hawth's Tools 3.27 (BEYER 2006)
[5] Add XY to Table	Koordinaten der generierten Punkte einfügen; Hawth's Tools 3.27 (BEYER 2006)
[6] Create Vector Grid	Polygon-Gitter erzeugen; spacing between lines X/Y 25; Hawth's Tools 3.27 (BEYER 2006)
[7] Buffer (5000m)	Puffer um Kanton Zürich erstellen (angestrebte maximale Euclidische Distanz für Distanz-Daten: 5000 m (siehe Abschnitt III 4.5, S.28); linear unit: 5000m, side type: full; Analysis Tools
[8] Create Bounding Box	Bounding-Box um den Puffer erstellen; incl. Buffer von 12.5m; Marine Geospatial Ecology Tools 0.7a15 (ROBERTS et al. in review)
[9] Buffer (100m)	Puffer um Kanton Zürich erstellen; (zum Extrahieren der Raster-Daten der erklärenden Variablen); linear unit: 100m, side type: full; Analysis Tools

### 3 Verbreitungsdaten

Die für die Modellierung verwendeten Fundortdaten stammen aus der Neophytenkataster-Datenbank (WEBGIS) der Baudirektion des Kantons Zürich, Amt für Abfall, Wasser, Energie und Luft (AWEL), Abteilung Abfallwirtschaft und Betriebe (AW), Sektion Biosicherheit (BAUDIREKTION KANTON ZÜRICH; AMT FÜR RAUMORDNUNG UND VERMESSUNG 2009). Nähere Angaben zum Aufbau, Umfang und Nutzerkreis finden sich in AWEL (2005).

In der vorliegenden Version des Neophytenkatasters werden 26 Arten erfasst (Stand Oktober 2008), wobei der Zeitpunkt der ersten Erfassung in die Datenbank schwankt. Darüber hinaus ist der Beobachtungszeitpunkt nicht in allen Fällen klar verifizierbar, offenbar wurden auch Datensätze ohne Zeitangaben eingefügt. Die Mehrzahl der Fundortdaten weist ein Beobachtungsdatum ab 2002 auf. Die Datenbank erlaubt die Erfassung von Pflanzenvorkommen sowohl als Punkt- als auch als Flächen(Polygon)daten (BAUDIREKTION KANTON ZÜRICH; AMT FÜR RAUMORDNUNG UND VERMESSUNG 2009). Beide Datentypen wurden für diese Arbeit herangezogen, wobei auf zwei Sichten (engl. views) zurückgegriffen wurde, die jeweils alle jemals erfassten – aber um allfällige Fehler bereinigten – Daten enthielten (Tabelle 4). Ihre Verwendung macht insofern Sinn, als dass in der beabsichtigten Modellierung potentielle Standorte der Arten erkannt werden sollen, unabhängig davon, ob die Art inzwischen aufgrund durchgeführte Bekämpfungsmaßnahmen wieder ausgerottet werden konnte.

*Tabelle 4: Verwendete Datensätze der Fundortdaten (Datums-Angaben und Anzahl der Einträge beziehen sich auf die Fläche des Kantons Zürich, siehe auch Schritt 2 in Abbildung 2).*

Datensatz	Beschreibung <sup>a</sup>	min. Beobachtungsdatum <sup>b</sup>	max. Beobachtungsdatum <sup>b</sup>	max. Erstellungsdatum	Anzahl Einträge
NEOPHYTEN_ALLEJEMALS	Punkte View, enthält alle aktuellsten Eingaben an einer Koordinate, falls eliminiert, dann die vorgängig aktuellste Bestätigung oder Erfassung.	19.05.2002	02.02.2009	02.03.2009	6'442
NEOPHYTEN_F_V	Flächen View, enthält sämtliche je erfassten Bestände, ohne die gelöschten	31.05.2000	06.08.2009	02.03.2009	2'547

<sup>a</sup> siehe ANONYMOS (2008)

<sup>b</sup> teilweise auch Datensätze ohne Beobachtungsdatum oder Pseudo-Datum (z.B. 01.01.1900) vorhanden.

---

Alle Fundortangaben wurden auf ein Gitternetz mit einer Kantenlänge von  $25 \times 25$  m aufbereitet, indem auf die im Abschnitt III 2 - Untersuchungsgebiet (S. 14) beschriebenen generierten Punkte- und Polygon-Gitter zurückgegriffen wurde. Außerhalb der Kantonsgrenze Zürich liegende Fundorte blieben ausgeschlossen. Daraus ergeben sich 41'127 Gitter-Punkte (Abbildung 3). Zum genauen Ablauf der Aufbereitung der Verbreitungsdaten siehe Abbildung 4 und Tabelle 5.

## kartierte Neophyten-Vorkommen

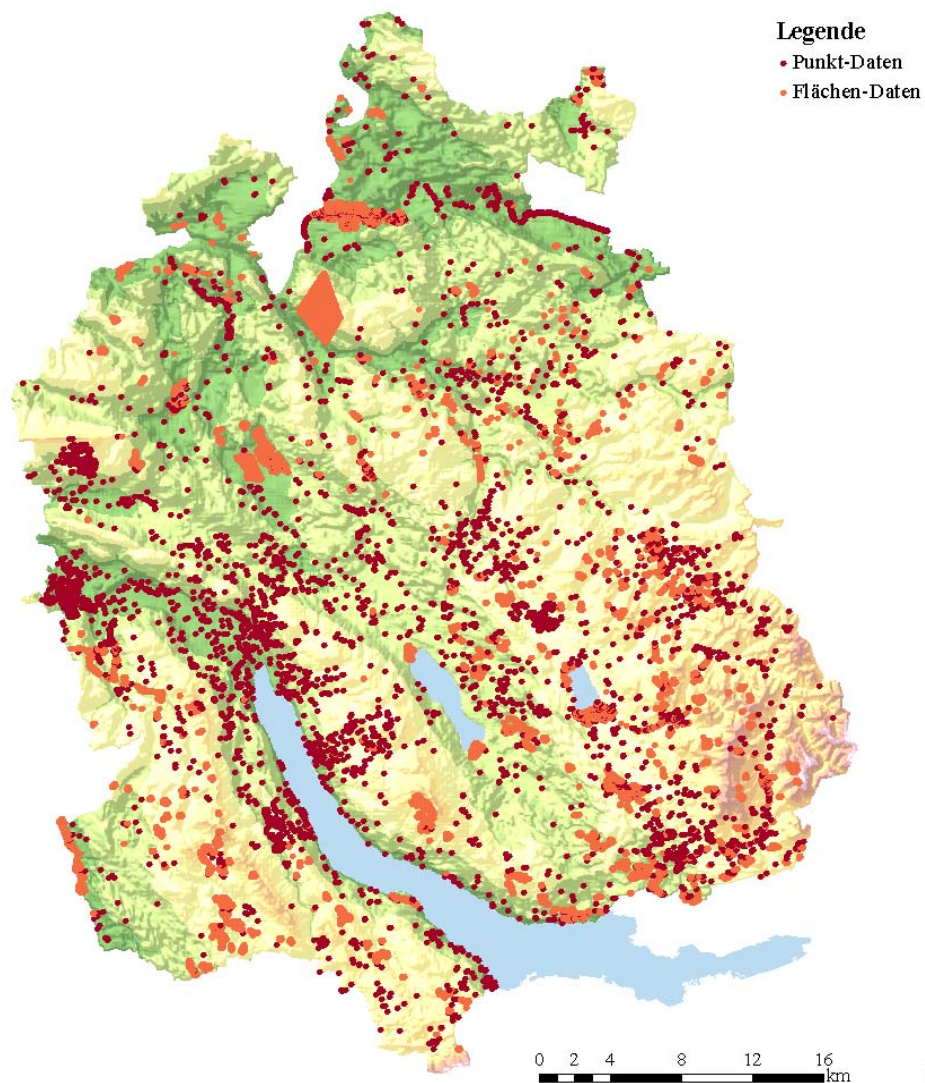


Abbildung 3: Punkt und Flächen-Fundortdaten der Neophyten im Kanton Zürich.

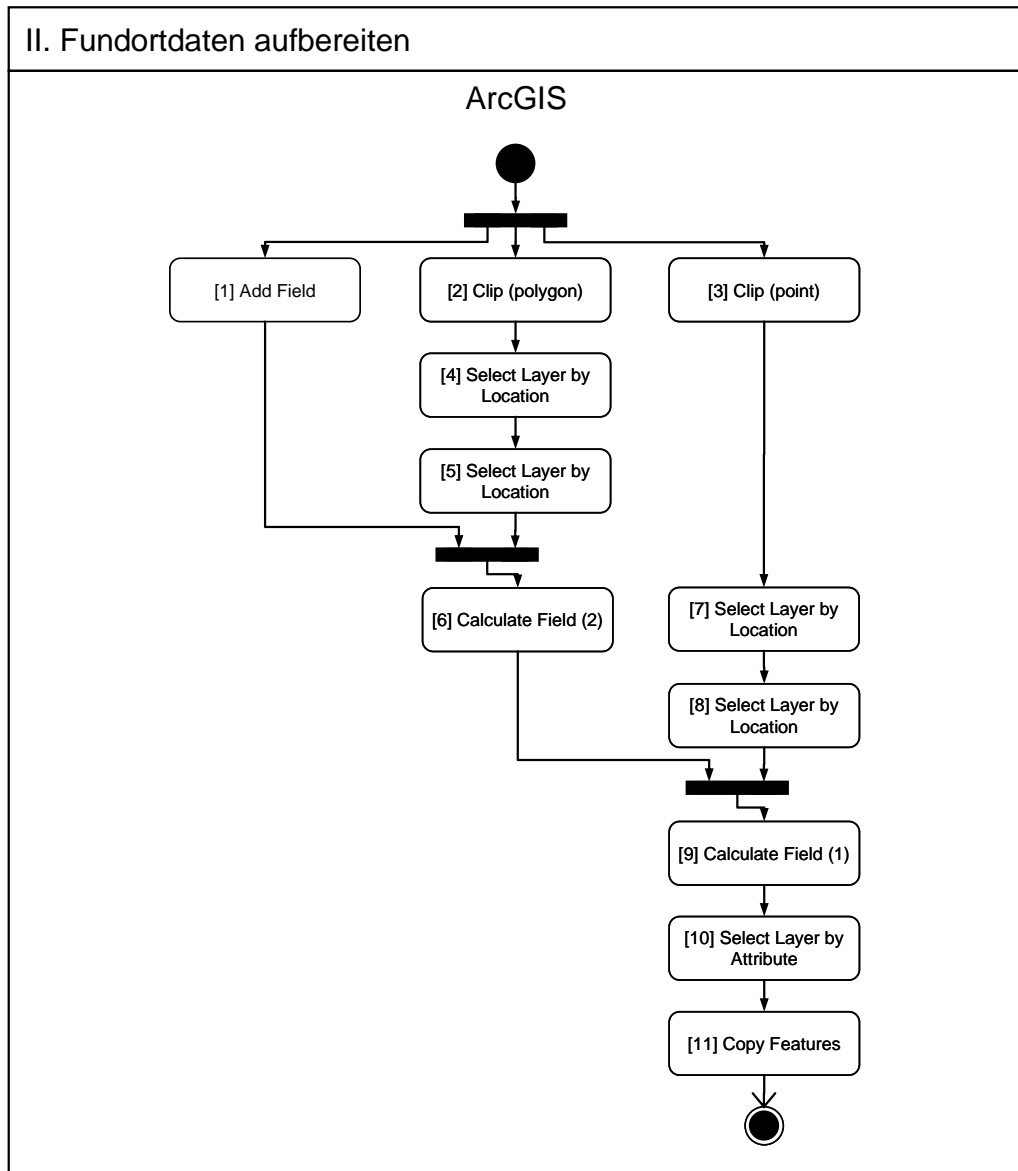


Abbildung 4: Aktivitätsdiagramm II: Fundortdaten aufbereiten.



*Tabelle 5: Erläuterungen zum Aktivitätsdiagramm II: Fundortdaten aufbereiten (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).*

Aktivität	Beschreibung
[1] Add Field	neues Feld für Punkte-Gitter-Datensatz definieren (siehe Schritt 4 in Abbildung 2); Field Name: neophyta_p, Default: 0; Data Management Tools
[2] Clip (polygon)	Fundort-Polygondatensatz NEOPHYTEN_F_V.shp mittels gepuffertem Untersuchungsgebiet ausschneiden (siehe Schritt 2 in Abbildung 2); Analysis Tools
[3] Clip (point)	Fundort-Punktensatz NEOPHYTEN_ALLEJEMALS.shp mittels gepuffertem Untersuchungsgebiet ausschneiden (siehe Schritt 2, in Abbildung 2); Analysis Tools
[4] Select Layer by Location	alle Polygon-Gitter selektieren, die Fundort-Polygone schneiden (siehe Schritt 6, in Abbildung 2); Relationship: Intersect; Data Management Tools
[5] Select Layer by Location	alle Punkte im Punkte-Gitter-Datensatz selektieren, die sich innerhalb der selektierten Polygon-Gitter befinden; Relationship: Intersect; Data Management Tools
[6] Calculate Field (2)	Feldwerte setzen; Field Name: neophyta_p, expression: 1; Data Management Tools
[7] Select Layer by Location	alle Polygon-Gitter selektieren, die Fundort-Punkte schneiden (siehe Schritt 6, in Abbildung 2); Relationship: Intersect; Data Management Tools
[8] Select Layer by Location	alle Punkte im Punkte-Gitter neobiota_fishnet.shp selektieren, die sich innerhalb der selektierten Polygon-Gitter befinden; Relationship: Intersect; Data Management Tools
[9] Calculate Field (1)	Feldwerte setzen; Field Name: neophyta_p, expression: 1 (überschreibt 2); Data Management Tools
[10] Select Layer by Attribute	nur Punkte des Punkte-Gitters selektieren, die Datensätze mit Fundorten enthalten; Expression: neophyta_p > 0; Data Management Tools
[11] Copy Features	in Datei neophyten_species.shp speichern; Data Management Tools

### 3.1 Vorkommen

Die Aufbereitung der Daten zu den kartierten Vorkommen erfolgte zunächst für elf Arten plus zwei Artengruppen. Verwendung fanden dabei diejenigen Arten, die beim Datenherrn auf besonderes Interesse hinsichtlich einer Modellierung des Verbreitungspotenzials stießen. Aufgrund der nicht ausreichenden Datenlage wurden im eigentlichen Modellierungsprozess jedoch lediglich sieben Arten verwendet. Die Aufbereitung der Punkt- und Polygondaten erfolgte analog des Vorgehens im vorangegangenen Abschnitt (Abbildung 5 und Tabelle 6). Eine Übersicht der untersuchten Arten mit Angaben zur Größe der Datensätze bzw. Anzahl der Gitter-Punkte findet sich in Tabelle 7.

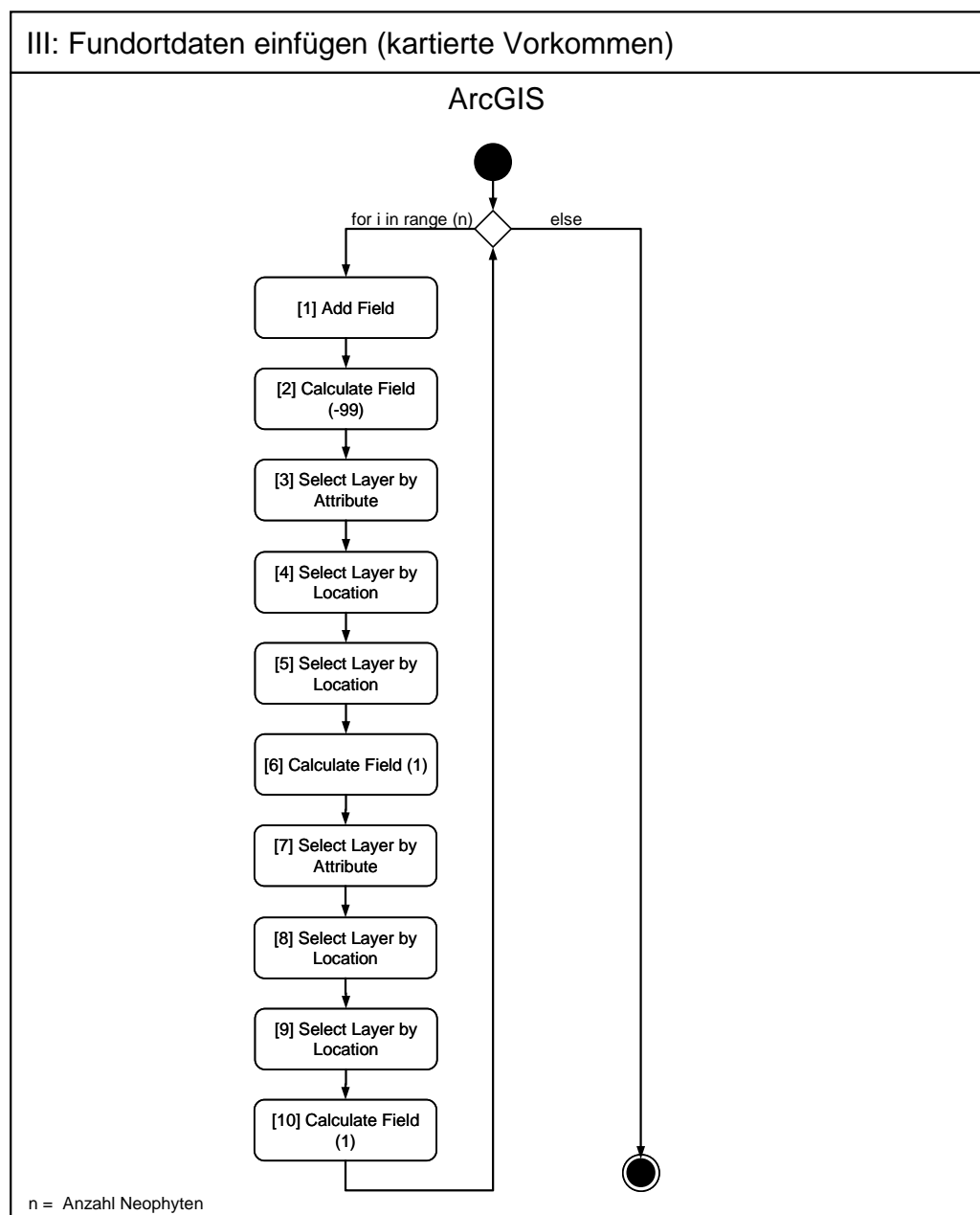


Abbildung 5: Aktivitätsdiagramm III: Fundortdaten einfügen (kartierte Vorkommen).

*Tabelle 6: Erläuterungen zum Aktivitätsdiagramm III: Fundortdaten einfügen (kartierte Vorkommen) (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).*

Aktivität	Beschreibung
[1] Add Field	Beginn der Schleife (pro Art für 13 Arten/Artengruppen ausführen): neues Feld in <code>neophyten_species.shp</code> anlegen; Field Type: LONG; Data Management Tools
[2] Calculate Field (-99)	initialen Wert setzen; Expression: -99; Data Management Tools
[3] Select Layer by Attribute	alle Fundort-Polygone von Art n selektieren; Data Management Tools
[4] Select Layer by Location	alle Polygon-Gitter selektieren, die die selektierten Fundort-Polygone schneiden (siehe Schritt 6 in Abbildung 2); Relationship: Intersect; Data Management Tools
[5] Select Layer by Location	alle Punktdaten in <code>neophyten_species.shp</code> selektieren, die sich innerhalb der selektierten Polygon-Gitter befinden; Relationship: Intersect; Data Management Tools
[6] Calculate Field (1)	Feldwerte setzen; Expression: 1; Data Management Tools
[7] Select Layer by Attribute	alle Fundort-Punkte von Art n selektieren; Data Management Tools
[8] Select Layer by Location	alle Polygon-Gitter selektieren, die die selektierten Fundort-Punkte schneiden (siehe Schritt 6 in Abbildung 2); Relationship: Intersect; Data Management Tools
[9] Select Layer by Location	alle Punktdaten in <code>neophyten_species.shp</code> selektieren, die sich innerhalb der selektierten Polygon-Gitter befinden; Relationship: Intersect; Data Management Tools
[10] Calculate Field (1)	Feldwerte setzen; Expression: 1; Data Management Tools Ende der Schleife

Tabelle 7: Übersicht der für die Modellierung aufbereiteten Neophyten-Arten mit Angaben zur Anzahl Datensätze und daraus resultierender Gitter-Punkte (im eigentlichen Modellierungsprozess verwendete Arten sind unterstrichen).

ID <sup>a</sup>	wissenschaftlicher Name <sup>b</sup>	deutscher Name <sup>b</sup>	Datensatz <sup>c</sup>	Gitter <sup>d</sup>
1	<u>Ambrosia artemisiifolia L.</u>	Beifuß-Ambrosie	756/1	779
2	<u>Heracleum mantegazzianum</u> <u>SOMMIER et LEVIER</u>	Riesen-Bärenklau	632/35	738
3	<i>Reynoutria</i> indet.	Knöterich (unbestimmt)	320/42	410
4	<u>Reynoutria japonica HOUTT.</u>	Japanischer Staudenknöterich	941/361	1'230
5	<i>Reynoutria sachalinensis</i> (F. SCHMIDT) NAKAI	Sachalin-Staudenknöterich	17/0	17
6	<i>Reynoutria X bohemica</i>	Bastard-Knöterich	9/0	10
7	<i>Polygonum polystachyum</i> Meisn.	Himalaya-Knöterich	12/5	20
8	<u>Impatiens glandulifera ROYLE</u>	Drüsiges Springkraut	886/190	1'846
9	<i>Senecio inaequidens</i> DC.	Schmalblättriges Greiskraut	44/0	45
10	<u>Rhus typhina L.</u>	Essigbaum	169/17	208
11	<i>Solidago</i> indet.	Goldrute (unbestimmt)	529/1033	24'254
12	<u>Solidago canadensis L.</u>	Kanadische Goldrute	267/118	1'177
13	<u>Solidago gigantea AIT.</u>	Riesen-Goldrute	671/622	10'044

<sup>a</sup> verwendete Arten ID gemäß Neophyten Datenmodell (BAUDIREKTION KANTON ZÜRICH; AMT FÜR RAUMORDNUNG UND VERMESSUNG 2009)

<sup>b</sup> Bezeichnung nach BÄBLER et al. (1996), MOSER et al. (2002)

<sup>c</sup> Anzahl Datensätze, angegeben als Anzahl Punkt/Anzahl Flächen-Datensätze

<sup>d</sup> Anzahl Gitter-Punkte

### 3.2 Pseudo-Absenz

Bedingt durch die Art der Datenhaltung und des Fehlens eines „Versuchsdesigns“ existieren lediglich Daten über das Vorkommen der Neophyten, Absenz-Daten fehlen. Deren Fehlen ist problematisch, da sie für die Modellierung essentiell sind. In der Praxis tritt diese Situation oft ein, wenn nicht explizit für derartige Modellierungen gesammelte Daten – z.B. aus Atlanten, Museen und Datenbanken (CHEFAOUI & LOBO 2008) – analysiert werden sollen. Man behilft sich häufig durch die Generierung sogenannter Pseudo-Absenzen. In der Literatur existieren verschiedene Ansätze für ihre Erzeugung. So können Pseudo-Absenzen:

- gänzlich zufällig über das Untersuchungsgebiet verteilt sein (STOCKWELL & PETERS 1999, STOCKWELL & PETERSON 2002)
- zufällig, aber gewichtet mit Umwelt-Faktoren sein (ZANIEWSKI et al. 2002)
- zufällig, aber geographisch gewichtet sein (HIRZEL et al. 2001)
- zufällig, aber gegebenenfalls geografisch und mit Umwelt-Faktoren gewichtet sein (FERRIER et al. 2002)
- Fundorte anderer Arten entsprechen, bei der die entsprechende Art nicht erfasst wurde (ELITH & LEATHWICK 2007)

Beeinflusst durch die Ausführungen von PEARCE & BOYCE (2006), die auf mögliche systematische Fehler im Zusammenhang mit „selektiven“ Kartierungen hinweisen<sup>7</sup> – die auch bei den vorliegenden Daten sehr wahrscheinlich sind –, wurde ein ähnlicher Ansatz wie von ELITH & LEATHWICK (2007) beschrieben, gewählt. Dazu wurde die Auswahllogik von Pseudo-Absenzen folgendermaßen implementiert: Potentielle Kandidaten für Absenz-Punkte waren alle Punkte, mit Ausnahme derjenigen, an der die Art selbst vorkam oder die sich in direkter Umgebung zu einem solchen Punkt befanden (im Einzelnen: die benachbarte acht Gitter-Punkte). Speziell wurden die Fundorte der Stauden-Knöterich- und Knöterich-Arten sowie der zwei Goldruten-Arten behandelt. Hier wurden bei den einzelnen Arten jeweils auch die nicht bestimmten (indet.) Knöterich- und Goldruten-Vorkommen als Präsenz-Punkte verwendet, um zufällige Zuordnungen von eigentlichen Präsenzen zu Pseudo-Absenzen zu vermeiden.

Angestrebt wurde eine gleich große Anzahl von Vorkommen und Pseudo-Absenzen, da deren Verhältnis zueinander Einfluss auf einige Gütemaße hat (siehe Abschnitt II 2.2.2 - Modellgüte, S. 9 ff.). Hierzu wurde eine entsprechende Anzahl Punkte aus dem „Pool“ der potentiellen Pseudo-Absenz-Punkte pro Art zufällig ausgewählt. Lediglich bei *Solidago gigantea* konnte dieses Verhältnis nicht realisiert werden. Der genaue Ablauf der Generierung der Pseudo-Absenzen ist aus Abbildung 6 und Tabelle 8 ersichtlich.

---

<sup>7</sup> Beispielsweise können straßennahe Gebiete in den Daten überrepräsentiert sein, da Kartierungen mit Hilfe von Autos durchgeführt werden. Eine andere Fehlerquelle könnte die besonders gründliche Erfassung der Flora in großen Städten sein.

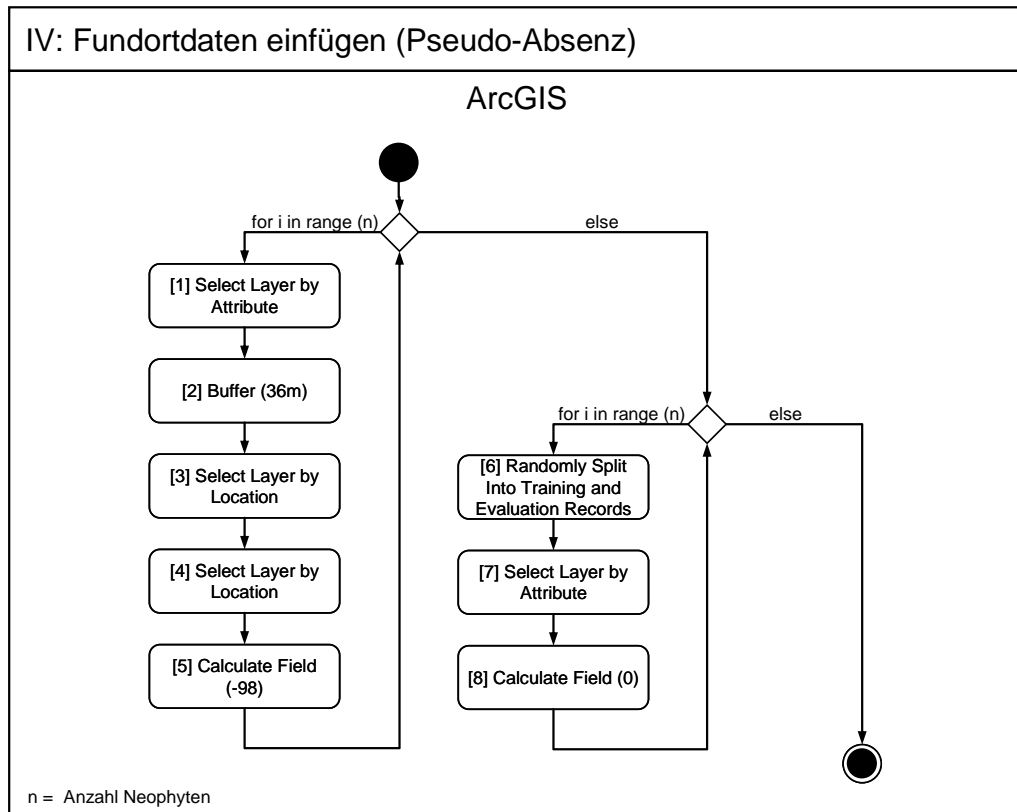


Abbildung 6: Aktivitätsdiagramm IV: Fundortdaten einfügen (Pseudo-Absenz).

Tabelle 8: Erläuterungen zum Aktivitätsdiagramm IV: Fundortdaten einfügen (Pseudo-Absenz) (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).

Aktivität	Beschreibung
[1] Select Layer by Attribute	Beginn der Schleife (pro Art für 11 Arten ausführen): alle Datensätze selektieren, bei der die Art vorkommt; Data Management Tools
[2] Buffer	Puffer um selektierte Punkte im Punkte-Gitter erstellen, um alle benachbarten Punkte zu markieren; linear unit: 36 m, side type: full; Analysis Tools
[3] Select Layer by Location	alle Punktdaten in neophyten_species.shp selektieren, die sich innerhalb der Puffer-Polygone befinden; Relationship: Intersect; Data Management Tools
[4] Select Layer by Location	Auswahl umkehren, selection type: Switch_Selection; Data Management Tools
[5] Calculate Field	Feldwerte setzen; Expression: -98 (potentielle Pseudo-Absenz); Data Management Tools
	Ende der Schleife
[6] Randomly Split Table into Training and Evaluation Records	Beginn der Schleife (pro Art für 11 Arten ausführen): Percentage of evaluation records: $\sum \text{presence} * 100 / \sum(\text{potentielle pseudo-absence})$ ; Where clause: "Art n" = -98; Marine Geospatial Ecology Tools 0.7a15 (ROBERTS et al. in review)
[7] Select Layer by Attribute	markierte evaluation records selektieren; Data Management Tools
[8] Calculate Field	Feldwerte setzen; Expression: 0 (Pseudo-Absenz); Data Management Tools
	Ende der Schleife

## 4 Erklärende Variablen

Als potentiell erklärende Variablen wurde auf eine Vielzahl von Daten zurückgegriffen. Wie schon im Abschnitt II 1 (S. 4) besprochen, lag das Hauptaugenmerk bei der Auswahl dabei auf dem möglichen ökologischen Aussagegehalt für die Verbreitung der Neophyten. Insgesamt verwendet diese Arbeit 24 Variablen, wobei teilweise auf bereits aufbereitete Daten von Mitarbeitern der Eidgenössischen Forschungsanstalt für Wald, Schnee und Landschaft (WSL) zurückgegriffen werden konnte. Alle Variablen konnten in einer Auflösung von  $25 \times 25$  m genutzt werden. Lediglich die Variablen zur Arealstatistik lagen in einer räumlichen Auflösung von  $100 \times 100$  m vor. Tabelle 9 gibt einen Überblick der genutzten erklärenden Variablen. Hier ist auch die Datenquelle vermerkt sowie gekennzeichnet, welche Variablen verändert oder neu generiert wurden. Der Ablauf der Aufbereitung bzw. die Generierung der erklärenden Variablen ist in Abbildung 7 bzw. Abbildung 8 und in den dazugehörigen Tabellen ersichtlich.

Großer Wert wurde auch auf eine gründliche Analyse der statistischen Verteilungen (Abschnitt 5 Datentransformation) und möglicher Korrelationen (Abschnitt 6 Kollinearität) der erklärenden Variablen gelegt (siehe auch Ausführungen auf S. 6 ff.).

Tabelle 9: Aufbereitete erklärende Variablen. Letztendlich für die Modellierung im engeren Sinne verwendete Variablen sind unterstrichen (siehe auch Abschnitte 5 Datentransformation und 6 Kollinearität).

Variable	Abk.	Typ <sup>a</sup>	Separation/Reklassifizierung	G <sup>b</sup>	Quelle
<b>Klima</b>					
<u>Jahresmitteltemperatur</u>	TAVE	C			(ZIMMERMANN & KIENAST 1999)
<u>Potentielle Globalstrahlung im Dezember</u>	SRAD12	C			(ZIMMERMANN & KIENAST 1999)
<b>Topografie</b>					
<u>Exposition</u> (aspect)	ASPC	C	cos(45° - aspect) + 1	×	(ZIMMERMANN, unpubliziert)
<u>Topografische Exponiertheit</u>	TOPS	C			(ZIMMERMANN & ROBERTS 2001)
<u>Neigung</u> (slope)	SLP	C			(ZIMMERMANN & ROBERTS 2001)
<b>Boden (physikalisch)</b>					
<u>Skelettgehalt</u>	SKELE	F	0(0), 1(1-4), 2(5-10), 3(11-20), 4(21-48)	×	(ZIMMERMANN, unpubliziert)
Vernässungsgrad	VNASS	C	exclusive 999 (NA)	×	(ZIMMERMANN, unpubliziert)
<u>Wasserspeichervermögen</u>	WSPEI	F	0(0), 1(>0 <70), 2(≥70 < 100), 3 (≥100)	×	(ZIMMERMANN, unpubliziert)
<b>Bodenbedeckung</b>					
Arealstatistik 1985	AS85	F			(BFS 1979/85)
<u>Arealstatistik 1997</u>	AS97	F			(BFS 1992/97)
<u>Arealstatistik 1997, reklassifiziert</u>	AS97R	F			(BFS 1992/97)
Wald/Gehölze	1	F	Klasse: 09, 11, 14, 17	×	
Landwirtschaft, int.	2	F	Klasse: 75, 76, 78, 81	×	
Landwirtschaft, ext.	3	F	Klasse: 82,83	×	
Garten, Park, Umschwung	4	F	Klasse: 41,45, 46, 47, 48, 49, 52, 59	×	
Nass-Standort	5	F	Klasse: 95	×	
sonstiges	6	F		×	
<b>Distanz</b>					
<u>Autobahn</u>	STR01	C	Straßennetz; ObjectVal: A_Zufahrt, Autob_Ri, Autobahn, Autostr, Ein_Ausf	×	(BFL 2007a)
<u>asphaltierte Straße</u>	STR02	C	Straßennetz; ObjectVal: 1_Klass, 2_Klass, 3_Klass, Bundesst, Nebenstr, Q_Klass	×	(BFL 2007a)
<u>nicht asphaltierter Feld/Waldweg</u>	STR03	C	Straßennetz; ObjectVal: 4_Klass, 5_Klass, Fahrweg, Wirtweg	×	(BFL 2007a)
<u>Fußweg</u>	STR04	C	Straßennetz; ObjectVal: 5_Klass_Spur, 6_Klass, 6_Klass_Spur, Fussweg, Parkweg	×	(BFL 2007a)



Fortsetzung Tabelle 10

Variable	Abk.	Typ <sup>a</sup>	Separation/Reklassifizierung	G <sup>b</sup>	Quelle
<u>Eisenbahn</u>	EIS01	C	Eisenbahnnetz; ObjectVal: Gt-Bahn, I-Geleis, MS-Bahn, NS_Bahn1, NS_Bahn2, SS_Bahn1, SS_Bahn2, Str_Bhof; TunnelType <> Tunnel	×	(BFL 2007a)
<u>Bach</u>	GWN01	C	Gewässernetz; ObjectVal: Bach	×	(BFL 2007a)
<u>Fluss/Kanal</u>	GWN02	C	Gewässernetz; ObjectVal: Fluss, Kanal	×	(BFL 2007a)
<u>See</u>	GWN03	C	Gewässernetz; ObjectVal: See	×	(BFL 2007a)
<u>Gebäude</u>	GEB01	C	Gebäude	×	(BFL 2007a)
Bahnhofareal	ANL01	C	Anlagen; ObjectVal: Z-BhArea	×	(BFL 2007a)
Flughafenareal	ANL02	C	Anlagen; ObjectVal: Z_FlArea	×	(BFL 2007a)
Flughafenbahnhofareal	ANL03	C	Anlagen; ObjectVal: Z_FlugBh	×	(BFL 2007a)

<sup>a</sup> C: kontinuierliche (continuous) variable, F: kategoriale (factor) variable

<sup>b</sup> G(generiert): x = im Rahmen dieser Arbeit veränderte oder neu generierte Variablen

## 4.1 Klima-Variablen

Als Klima-Variablen dienten die Jahresmitteltemperatur und die potentielle Globalstrahlung im Dezember (nähere Informationen zu Berechnung der Variablen finden sich in ZIMMERMANN 2000a, 2000b, ZIMMERMANN & KIENAST 1999). Andere Klima-Variablen fanden nicht Verwendung, da aufgrund der geringen räumlichen Ausdehnung des Untersuchungsgebietes, verbunden mit einer vergleichsweise geringen Höhenstufung, durch andere Variablen kein Informationsgewinn zu erwarten war (siehe Problematik Kollinearität S.7).

## 4.2 Topografische Variablen

Topografische Variablen – abgeleitet aus dem Digitalen Höhenmodell des Bundesamtes für Landestopografie (BFL 1994) – standen in Form von Neigung (Hangneigung), Exposition (erfasst als Azimut in Grad im Uhrzeigersinn von Nord) und der topografischen Exponiertheit (Hügel/Senke) zu Verfügung. Im Falle der Exposition erfolgte eine Transformation der Werte. Hierfür fand die Kosinusfunktion Verwendung, wobei gleichzeitig eine Phasenverschiebung um 45° – entsprechend der oftmals günstigsten Vegetationsbedingungen in Nordnordost-Richtung – durchgeführt wurde. Durch Addition von eins wurden negative Werte vermieden (detailliertere Informationen zu dieser Thematik finden sich in TRIMBLE & WEITZMAN 1956, BEERS et al. 1966, STAGE 1976).

### **4.3 Physikalische Boden-Variablen**

Physikalische Parameter des Bodens konnten durch die Variablen Vernässung (Lage des porengesättigten Horizontes), Skelettgehalt (Anteil von Kiesen und Steinen) (BLUM 2007) und Wasserspeichervermögen in der Modellierung berücksichtigt werden.

### **4.4 Bodenbedeckungs-Variablen**

Aufbereitete Informationen zur Bodenbedeckung bzw. Bodennutzung sind in der Schweiz in Form der Arealstatistik verfügbar (BFS 2010). In dieser Arbeit fanden die Arealstatistik der Jahre 1979/85 (kurz Arealstatistik 85) und 1992/1997 (kurz Arealstatistik 97) Verwendung. In ihnen werden Merkmale in 72 (Arealstatistik 85) bzw. 74 Kategorien (Arealstatistik 97) klassifiziert (BFS 2007a). Zusätzlich wurde die Arealstatistik 1992/1997 reklassifiziert (siehe Tabelle 9), um für die Fragestellung wichtige Kategorien zu selektieren und zusammenzufassen. Hierbei wurden zum einen diejenigen Kategorien ausgewählt und teilweise zusammengefasst, die aufgrund ökologischer Überlegungen einen Einfluss auf die Artverbreitung erwarten lassen (z.B. Nass-Standort). Zum anderen wurde jedoch nur auf solche Kategorien zurückgegriffen, bei denen die Information nicht schon in anderen Variablen enthalten war (Distanz-Variablen, z.B. Gebäude).

### **4.5 Distanz-Variablen**

Die Distanz-Variablen – die über die Entfernung des Neophyten-Fundortes von einer entsprechenden Struktur Auskunft gibt – wurden aus Teilen des digitalen Landschaftsmodells VECTOR25 des Bundesamtes für Landestopographie generiert (BFL 2007a). Neben den bestehenden thematischen Ebenen (z.B. Straßen- und Wegnetz) erfolgte eine feinere Aufteilung mit Hilfe des Feldes Objektart (ObjectVal, z.B. Autobahn, Fußweg) und im Falle der Berechnung der Distanzen zum Eisenbahnnetz auch des Feldes Tunneltyp. Dem lag die Überlegung zugrunde, dass die einzelnen Objektarten einen unterschiedlich starken Erklärungsgehalt der räumlichen Verbreitung der Neophyten besitzen können. Beispielsweise verfügt eine Autobahn als lineare Ausbreitungsstruktur über andere Standortbedingungen als ein Fußweg. Die genaue Definition der Variablen ist in Tabelle 9 ersichtlich, der Arbeitsablauf in Abbildung 7 und Tabelle 11. Nähere Informationen zum Attribut ObjectVal finden sich in BFL (2007b).

Als maximale Distanz wurde eine Entfernung von 5000 m definiert. Hintergrund dieser Definition waren vor allem Limitationen in Bezug auf die räumliche Ausdehnung der Daten des digitalen Landschaftsmodells (Datenverfügbarkeit für Gebiete außerhalb der Schweiz). Darüber hinaus wären größere Distanzen in den Modellen auch schwer interpretierbar gewesen.

Nachfolgend sind die zwölf generierten Variablen kurz beschrieben.

#### **4.5.1 Distanzen zu Verkehrswegen**

Verkehrswege werden im VECTOR25-Modell als thematische Ebenen Straßen- und Eisenbahn-Netz erfasst. Im Falle des Straßennetzes erfolgte eine feingliedrigere Unterscheidung in vier Straßenklassen: Autobahnen, asphaltierte Straßen, nicht asphaltierte Feld- und Waldwege sowie Fußwege.

Die thematische Ebene Eisenbahnnetz wurde nicht feiner separiert. Von der Distanzberechnung wurden jedoch Tunnelabschnitte sowie Straßenbahngleise ausgeschlossen.

#### **4.5.2 Distanzen zu Siedlungsflächen**

Distanzinformationen zu Siedlungsflächen wurden aus den thematischen Ebenen Gebäude und Anlagen extrahiert. Aus ersterer Ebene wurde die Variable Gebäude generiert. Sie enthält alle Objektarten. Aus der Ebene Anlagen wurde von jeder Objektart eine Variable gebildet (Bahnhofareal, Flughafenareal und Bahnhofflughafenareal).

#### **4.5.3 Distanzen zu Gewässern**

Es wurden die Distanz-Variablen Bäche, Flüsse/Kanäle und Seen gebildet, indem einige Objektarten der thematischen Ebene Gewässernetz verwendet wurden.

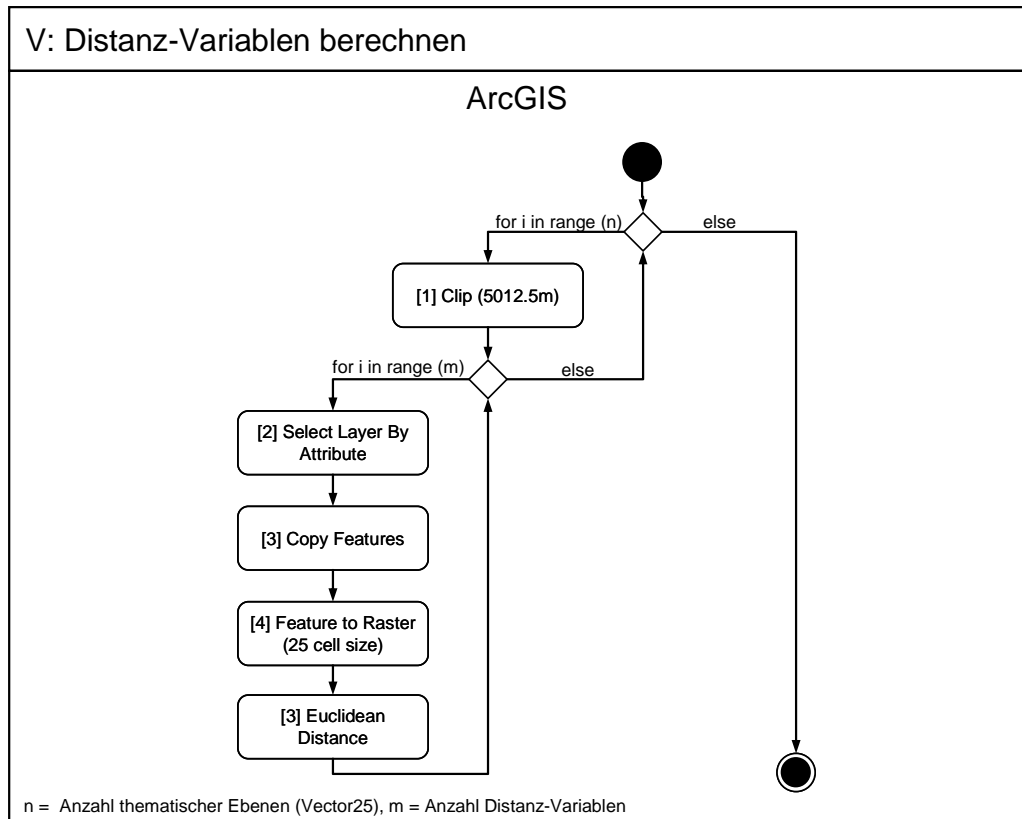


Abbildung 7: Aktivitätsdiagramm V: Distanz-Variablen berechnen.

Tabelle 11: Erläuterungen zum Aktivitätsdiagramm V: Distanz-Variablen berechnen (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).

Aktivität	Beschreibung
[1] Clip	VECTOR25-Daten mittels Bounding-Box (siehe Schritt 8 in Abbildung 2) ausschneiden, <i>Analysis Tools</i>
[2] Select Layer by Attribute	Beginn der Schleife (pro thematischer Ebene für 5 thematische Ebenen ausführen): alle Polygone/Linien selektieren, die Objectval-Auswahl entsprechen (siehe Tabelle 9); <i>Data Management Tools</i>
[3] Copy Features	in Datei „variable“.shp speichern; <i>Data Management Tools</i>
[4] Feature to Raster	als Raster-Datensatz abspeichern; Field: Objectval; Output cell size: 25; <i>Conversion Tool</i>
[5] Euclidian Distance	Euclidische Distanz berechnen; Maximum distance: 5000, Output cell size: 25; <i>Spatial Analyst Tools</i> Ende der Schleife

## 5 Datentransformation

Alle verwendeten erklärenden Variablen wurden – wenn nötig und durchführbar – transformiert (siehe II 2.1.1. Daten-Transformation, S.7). Für jede Variable wurde das anhand von Dichte- und Boxplot-Diagrammen geprüft. Bei auffallend mehrgipfligen Verteilungen wurden auch Umwandlungen von kontinuierlichen zu kategorialen Variablen durchgeführt. Die Werteverteilung der Original-Variablen und eventuell vorgenommener Transformationen bzw. Klassifikationen sind im Anhang VII 3 (S. 85 ff.) dargestellt.

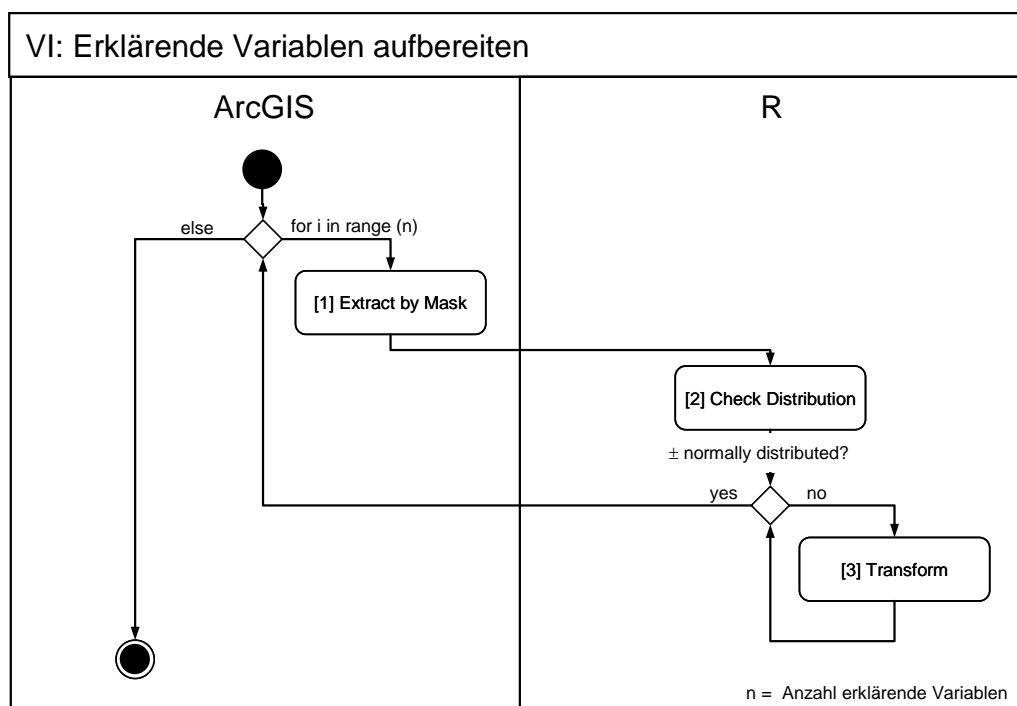


Abbildung 8: Aktivitätsdiagramm VI: Erklärende Variablen aufbereiten. Nicht aufgeführt ist die Prüfung auf Multikollinearität.

Tabelle 12: Erläuterungen zum Aktivitätsdiagramm VI: Erklärende Variablen aufbereiten (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).

Aktivität	Beschreibung
[1] Extract by Mask	Beginn der Schleife (für jede erklärende Variable ausführen): Rasterdaten mittels gepuffertem Untersuchungsgebiet (100 m) ausschneiden (siehe Schritt 9, in Abbildung 2); <i>Spatial Analyst Tools</i>
[2] Check Distribution	Verteilung prüfen
[3] Transform	transformieren
	Ende der Schleife

## 6 Kollinearität

Zur Aufdeckung möglicher Kollinearität verfolgte diese Arbeit zwei Ansätze. Zum einen wurden zur visuellen Interpretation möglicher Korrelationen Streudiagramme erstellt sowie die Regressionskoeffizienten nach Pearson aller Variablenpaare berechnet. Der Regressionskoeffizient gibt aber nur einen Hinweis auf die Stärke der Kollinearität, feste Schwellwerte können nicht angegeben werden (KUTNER et al. 2004). Lediglich bei GREEN (1979, zitiert in FIELDING & HAWORTH 1995) findet sich ein Vorschlag, bei Korrelationskoeffizienten über 0.7 eine der beiden Variablen auszuschließen. Diese Aussage deckt sich jedoch gut mit den Erfahrungen, die an der WSL mit ähnlichen Modellierungstechniken gemacht wurden (ZIMMERMANN, persönliche Mitteilung). Beim Auftreten von Korrelationen  $\geq 0.7$  wurden deshalb diejenigen Variablen ausgeschlossen, die für die ökologische Interpretation weniger gut geeignet schienen.

In einem zweiten Schritt wurde während des eigentlichen Modellierungsprozesses für jedes initiale Modell pro Neophyt (siehe Abschnitt II 2.1

Statistische Modellbildung, S. 36) der sogenannte variance inflation factor (VIF) berechnet. Er gibt die inverse Toleranz (1 - Regressionskoeffizient zwischen der Variablen und allen restlichen Variablen) der erklärenden Variablen an (siehe z.B. SHEATHER 2009). Hohe Werte zeigen dabei starke Kollinearität an. Werte größer zehn weisen auf eine Multikollinearität hin (QUINN & KEOUGH 2002, KUTNER et al. 2004). Die VIF-Werte wurden mit Hilfe des Paketes `car` (FOX 2009) berechnet und geprüft. Analog zum obigen Vorgehen wurde bei VIF-Werten größer zehn diejenigen Variablen ausgeschlossen, die für die ökologische Interpretation weniger gut geeignet schienen.

## 7 Daten-Überlagerung

Die Verbreitungsdaten der Neophyten (Vorkommen und Pseudo-Absenz) wurden mit den Werten der erklärenden Raster-Variablen überlagert (siehe Abbildung 9 und Tabelle 13).

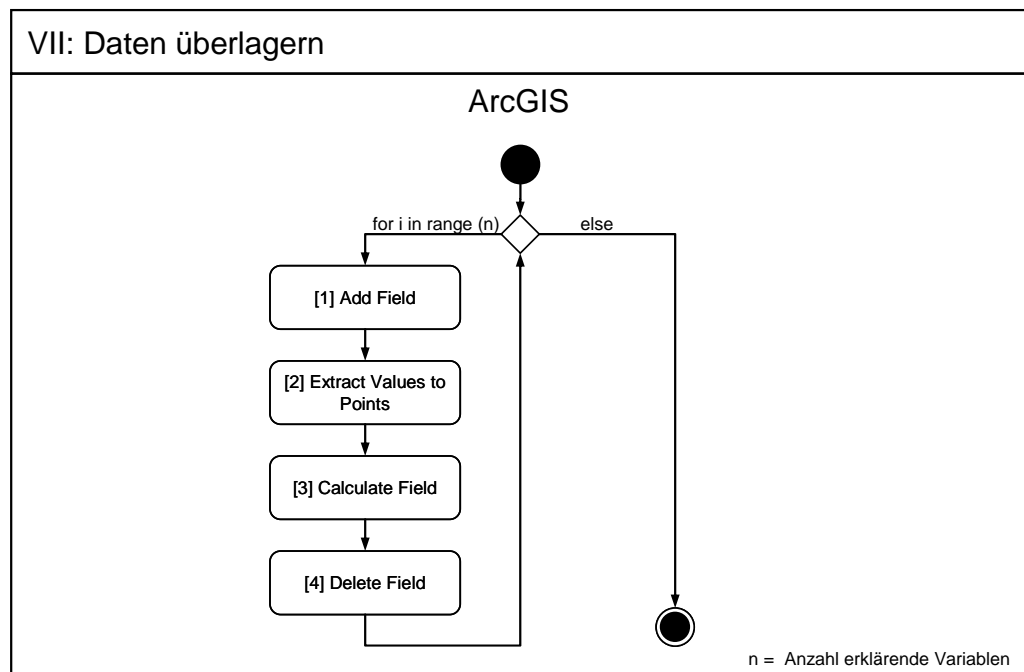


Abbildung 9: Aktivitätsdiagramm VIII: Daten überlagern.

Tabelle 13: Erläuterungen zum Aktivitätsdiagramm VIII: Daten überlagern (inklusive Angabe der verwendeten Toolbox und wichtiger Parameter).

Aktivität	Beschreibung
[1] Add Field	Beginn der Schleife (pro erklärender Variable ausführen): neues Feld für Datensatz <code>neobiota_zh.shp</code> definieren; Field Name: „Abkürzung des Variablennamens“, Default: 0; Data Management Tools
[2] Extract Values To Points	Zellwerte der Raster in den Punktelayer extrahieren, Default-Field: <code>RASTERVALU</code> ; Spatial Analyst Tools
[3] Calculate Field	Werte aus dem Feld <code>RASTERVALU</code> in das Variablenfeld kopieren; Data Management Tools <sup>a</sup>
[4] Delete Field	Feld <code>RASTERVALU</code> löschen; Data Management Tools Ende der Schleife

<sup>a</sup> Der Ablauf der Reklassifizierung der Variable Arealstatistik 97 ist hier nicht dargestellt (siehe Tabelle 9, S.26).



## 8 Modellierung im engeren Sinne

Unter diesem Titel sollen alle den eigentlichen Modellierungsprozess betreffende Schritte verstanden werden, exklusive der Daten-Aufbereitung und Visualisierung der Ergebnisse. Die Abfolge ist aus Abbildung 10 und Tabelle 14 ersichtlich.

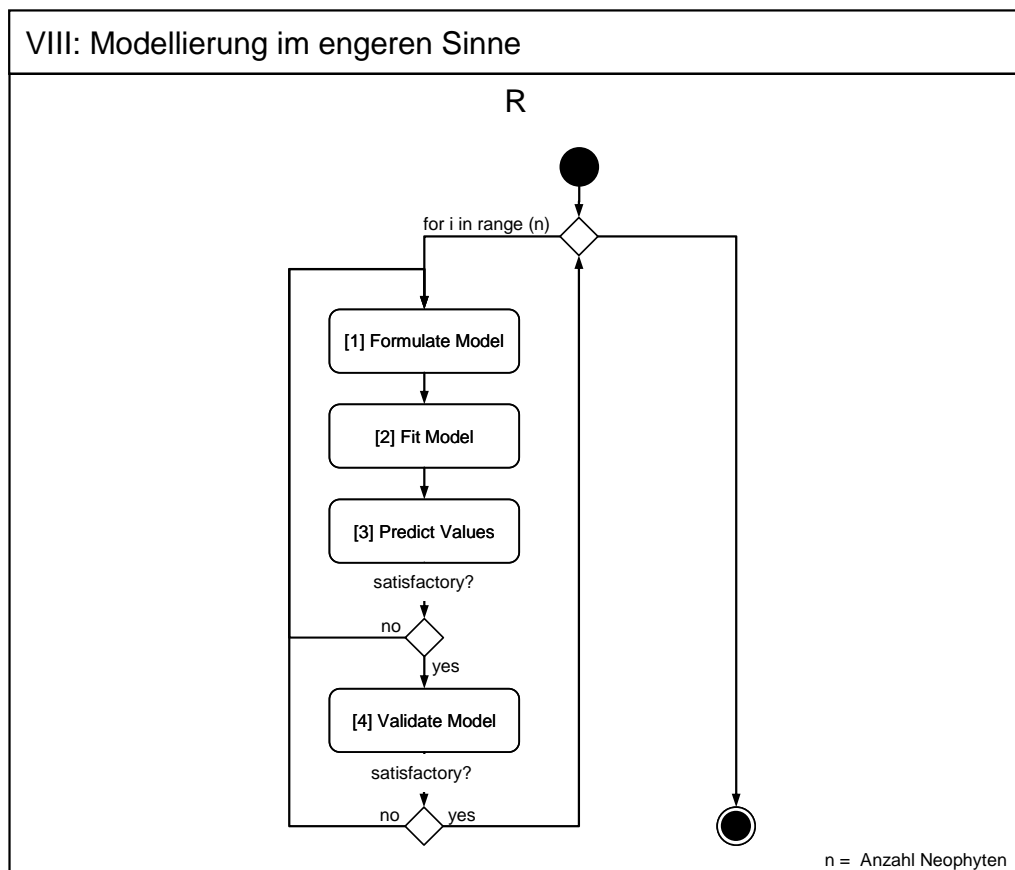


Abbildung 10: Aktivitätsdiagramm VII: Modellierung im engeren Sinne durchführen.

Tabelle 14: Erläuterungen zum Aktivitätsdiagramm VII: Modellierung im engeren Sinne durchführen.

Aktivität	Beschreibung
[1] Formulate Model	Beginn der Schleife (pro Art für 7 Arten ausführen): Modelle formulieren und berechnen
[2] Fit Model	Modelle anpassen, Schrittweise Variablen-Selektion,
[3] Predict Values	Modellvoraussagen machen, Gütemaße berechnen
[4] Validate Model	Kreuzvalidierung durchführen
	Ende der Schleife

## 8.1 Statistische Modellbildung

Für jeden der sieben Neophyten wurde eine Reihe von Teilmodellen gebildet. Grundsätzlich wurden je ein Modell pro Variablensatz, ein Modell mit allen Variablen sowie ein Modell mit quadratischen Termen formuliert (siehe Tabelle 15). Als Variablensatz wird hier eine Gruppe von erklärenden Variablen verstanden, die aufgrund theoretischer Überlegungen zusammengefasst wurden. So enthält beispielsweise der Variablensatz „trans“ alle Variablen mit den linearen Strukturen der Verkehrswege. Dem lag die Absicht zugrunde, die Relevanz dieser Variablensätze für das Vorkommen bzw. Fehlen der Art vergleichend abschätzen zu können. Das Modell mit quadratischen Termen wurde gebildet, um dem Konzept der ökologischen Nische (siehe S. 4) besser gerecht werden zu können.

Tabelle 15: Formuliert Teilmodelle für die Habitatmodellierung invasiver Neophyten.

Modell	Variablensatz	
trans	Verkehrswege	$STR01 + \log(STR02 + 25) + \log(STR03 + 25) + \log(STR04 + 25) + \sqrt{EIS01}$
gew	Gewässer	$\log(GWN01 + 25) + \sqrt{GWN02} + \sqrt{GWN03}$
urban	Gebäude	$\sqrt{GEB01}$
klima	Klima	$TAVE^6 + SRAD12$
topo	Topografie	$ASPC + TOPS + \log(SLP + 1)$
boden	Boden	$SKELE + WSPEI$
as97r	Arealstatistik 1997, reklassifiziert	AS97R
gesamt	gesamter Variablensatz	$\sqrt{GEB01} + STR01 + \log(STR02 + 25) + \log(STR03 + 25) + \log(STR04 + 25) + \sqrt{EIS01} + \log(GWN01 + 25) + \sqrt{GWN02} + \sqrt{GWN03} + TAVE^6 + SRAD12 + ASPC + TOPS + \log(SLP + 1) + SKELE + WSPEI + AS97R$
gesamt <sup>2</sup>	gesamter Variablensatz, mit quadratischen Termen	$\sqrt{GEB01} + I((\sqrt{GEB01})^2) + STR01 + I(STR01^2) + \log(STR02+25) + I((\log(STR02+25))^2) + \log(STR03+25) + I((\log(STR03+25))^2) + \log(STR04+25) + I((\log(STR04+25))^2) + \sqrt{EIS01} + I((\sqrt{EIS01})^2) + \log(GWN01+25) + I((\log(GWN01+25))^2) + \sqrt{GWN02} + I((\sqrt{GWN02})^2) + \sqrt{GWN03} + I((\sqrt{GWN03})^2) + TAVE^6 + I((TAVE^6)^2) + SRAD12 + I(SRAD12^2) + ASPC + I(ASPC^2) + TOPS + I(TOPS^2) + \log(SLP+1) + I((\log(SLP+1))^2) + SKELE + WSPEI + AS97R$

## 8.2 Modellanpassung

Die schrittweise Variablen-Selektion erfolge mittels Vorwärts- und Rückwärts-Algorithmus und AIC- bzw. BIC-Minimierung. Der Vorteil dieses Algorithmus gegenüber einer alleinigen Vorwärts- oder Rückwärts-Prozedur besteht darin, dass Variablen, die erst durch das Vorhandensein oder Fehlen einer anderen Variable signifikant werden, wieder in das Modell aufgenommen werden können (HILBE 2009).

Die Modelgüte wurde über die Gütemaße AIC, BIC, korr.  $D^2$ , AUC und maximales  $\kappa$  bzw. maximale TSS abgeschätzt. Da sich die Werte für das maximale  $\kappa$  bzw. TSS mit Ausnahme der Modelle für die Art *Solidago gigantea* aufgrund des Verhältnisses Vorkommen/Pseudo-Absenz von 1:1 (siehe Erläuterungen zur Generierung von Pseudo-Absenzen auf S.22 ff.) nicht unterscheiden, werden sie gemeinsam aufgeführt. Visuell erfolgte die Ausgabe von ROC-Plot mit Hilfe des Packetes `lroc` (CHONGSUVIVATWONG 2010).

### 8.3 Validierung

Die Validierung erfolgte über eine zehnfache Kreuzvalidierung und der Funktion `CVbinary()` aus dem Paket DAAG (MAINDONALD & BRAUN 2009), die die Vorhersagegüte der Modelle schätzt. Ausgegeben wird das geschätzte Verhältnis korrekter Vorhersagen (MAINDONALD & BRAUN 2007). Um die Vorhersagegüte innerhalb der Teilmodelle einer Art besser vergleichen zu können, wurde die zufällige Zuordnung der Daten in die Teilmengen vor der eigentlichen Kreuzvalidierung durchgeführt, galt also für alle Teilmodelle. Da die Schätzung von Zuordnung zu Zuordnung schwanken kann, wurde die Funktion darüber hinaus zehnmal ausgeführt und der Mittelwert der Ausgabewerte berechnet.

#### 8.3.1 Festlegung von Schwellenwerten

Schwellenwerte für die Überführung der Vorkommenswahrscheinlichkeiten in eine binäre Vorkommen/Absenz-Struktur wurden über die Berechnung der Gütemaße Cohen's Kappa und TSS bestimmt. Die Berechnung erfolgte in Stufen von 0.02.

## 9 Visualisierung

Die Visualisierung der Vorkommenswahrscheinlichkeiten und prognostizierten Vorkommen erfolgte in ArcGIS mit dem Spatial Analyst Raster Calculator. Dazu wurde mit Hilfe der errechneten Koeffizienten und der relevanten erklärenden Variablen die Regressionsgleichung formuliert und ausgeführt.

---

## IV Ergebnisse

### 1 Kollinearität

Die folgenden beiden Abbildungen zeigen die Verteilungen der erklärenden Variablen und ihre Korrelationen untereinander an. Nicht mit aufgeführt sind dabei diejenigen Variablen, die aufgrund zu hoher Korrelationswerte von der weiteren Modellierung ausgeschlossen wurden. Alle Variablenpaare wurden auf mögliche Korrelation überprüft. Aus Gründen der besseren Lesbarkeit sind die Variablen jedoch auf zwei Abbildungen aufgetrennt.

Nach ersten Modellierungen wurde darüber hinaus auch die Variable `an102` (Flughafenareal) ausgeschlossen. Sie weist aufgrund der äußerst geringen Flächen überwiegend die maximale Distanz von 5000 m auf (siehe Anlage VII 3.6, S. 90), was in der Modellierung zu Verfälschungen und Artefakten führte (beispielsweise Ringe hoher Vorkommenswahrscheinlichkeit um die Flughafenareale).

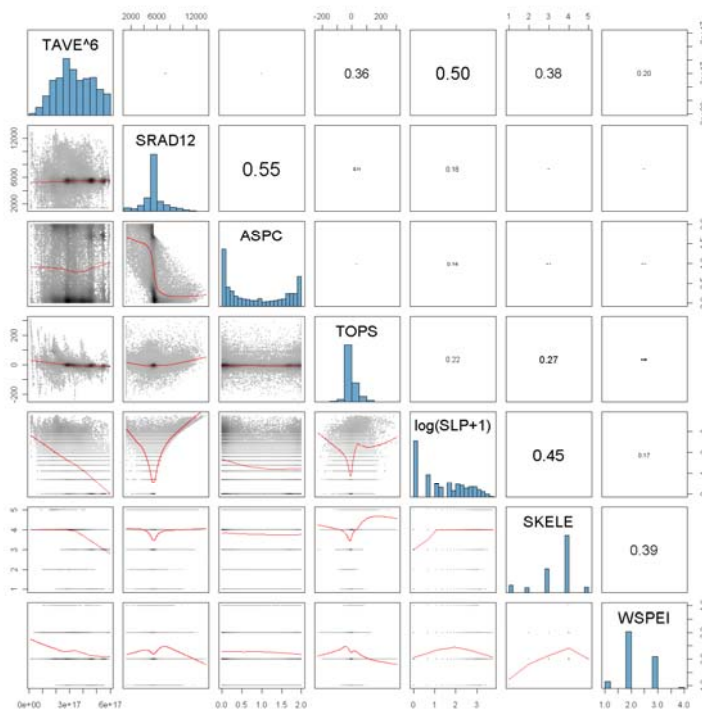


Abbildung 11: Scatterplot-Matrix der erklärenden Variablen mit Ausnahme der Distanz-Variablen (angegeben ist der absolute Regressionskoeffizient nach Pearson).

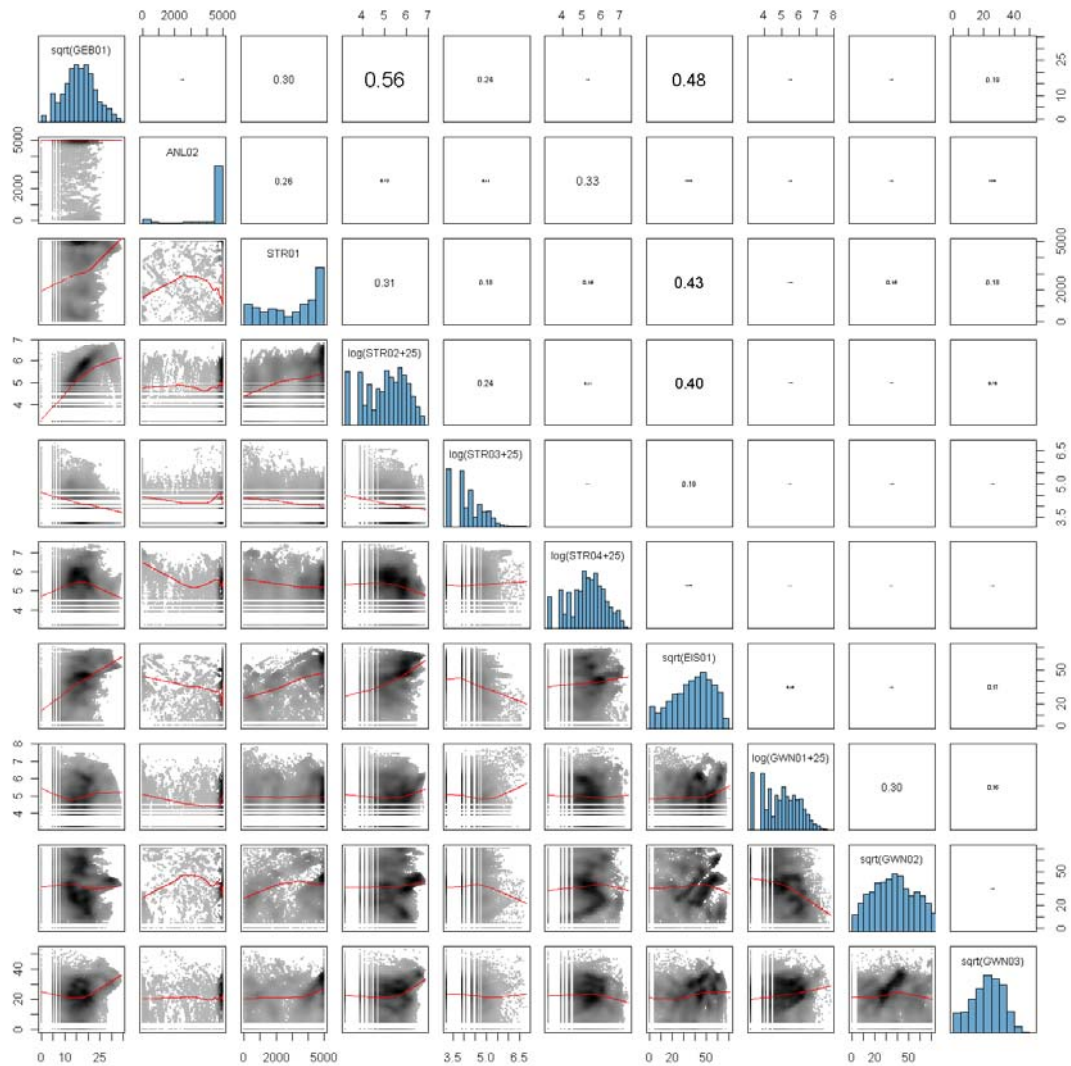


Abbildung 12: Scatterplot-Matrix der Distanz-Variablen (angegeben ist der absolute Regressionskoeffizient nach Pearson).

## 2 Modellierung im engeren Sinne und Visualisierung

Dieser Abschnitt erläutert die Ergebnisse der Modellbildungen für jede Neophytenart getrennt. In einer Übersichtstabelle sind jeweils für alle Teilmodelle – initial und nach durchgeführter schrittweiser Variablen-Selektion (nachfolgend kurz reduzierte Modelle bzw. AIC-Modelle oder BIC-Modell) – die diversen Gütemaße und der Schwellenwert, bei dem  $\kappa$ - bzw. TSS maximale Werte annehmen – aufgeführt. Dabei werden die reduzierten Modelle nur dargestellt, wenn sie sich von den initialen Modellen unterscheiden. Deren Variablensätze finden sich im Anhang VII 5 (S. 97ff.). Darüber hinaus sind ROC-Diagramme für alle Teilmodelle nach BIC-Reduktion dargestellt. Für das ausgewählte Modell werden tabellarisch die Regressionskoeffizienten der Variablen, deren Standardfehler und Signifikanzniveaus aufgeführt. Die beiden letzteren

werden im Text jedoch nicht näher besprochen. Aus Gründen der Modellgüte, Modellkomplexität und ökologischer Interpretierbarkeit (Stichwort ökologische Nische) wurde letztendlich immer auf das Gesamtmodell mit quadratischen Termen, angepasst durch eine schrittweise Variablen-Selektion und Prüfkriterium BIC, zurückgegriffen.

Am Ende jedes Abschnittes stellen zwei Karten die Vorkommenswahrscheinlichkeit der Art sowie das prognostizierte Vorkommen (binäre Darstellung von Vorkommen/Absenz) für das ausgewählte Modell dar<sup>8</sup>.

## **2.1 *Ambrosia artemisiifolia***

Betrachtet man die Teilmodelle, wird deutlich, dass die Distanz-Variablen, die direkte anthropogene Veränderungen der Landschaft indizieren (Verkehrswege, Gebäude), einen hohen Einfluss auf das Vorkommen oder die Absenz der Beifuß-Ambrosie besitzen. Erkennbar ist dies an den in den Teilmodellen trans und urban im Vergleich zu anderen Modellen niedrigen AIC- bzw. BIC-Werten, den hohen korr. D<sup>2</sup>- und AUC- sowie max.  $\kappa$ - bzw. TSS-Werten (Tabelle 16). Grafisch wird dieser Sachverhalt in den ROC-Plots in Abbildung 13 deutlich (steiler Anstieg des roten Graphen in die linke obere Ecke).

---

<sup>8</sup> Zur geografischen Einordnung urbaner Gebiete, aufgeführter Orts- und Flussnamen siehe Anhang VII 1 Übersichtskarte des Untersuchungsgebietes auf S. 82.

Tabelle 16: Übersicht wichtiger Kennwerte der durchgeführten Modelle zur Vorkommensvorhersage von *Ambrosia artemisiifolia*.

GLM	AIC/BIC	KV	korr. D <sup>2</sup>	AUC	Schw.	m. $\kappa$ /TSS
<b>trans</b>						
gesamt	1325.8	0.828	0.776	0.882	0.50	0.657
reduziert (BIC)	1330.4	0.822	0.387	0.880	0.40	0.656
<b>gew</b>						
gesamt	1961.8	0.634	0.094	0.701	0.54	0.286
<b>urban</b>						
gesamt	1110.2	0.876	0.488	0.883	0.50	0.752
<b>klima</b>						
gesamt	2034.7	0.650	0.060	0.656	0.52	0.312
reduziert (AIC)	2032.8	0.650	0.060	0.658	0.52	0.313
<b>topo</b>						
gesamt	2115.8	0.536	0.022	0.569	0.44	0.176
reduziert (AIC)	2114.4	0.536	0.023	0.565	0.44	0.169
<b>boden</b>						
gesamt	1818.7	0.677	0.161	0.630	0.36	0.365
<b>as97r</b>						
gesamt	1445.6	0.776	0.334	0.761	0.58	0.552
<b>gesamt</b>						
gesamt	923.0	0.897	0.590	0.949	0.52	0.814
reduziert (AIC)	911.3	0.896	0.590	0.948	0.58	0.809
reduziert (BIC)	922.4	0.898	0.581	0.946	0.48	0.802
<b>gesamt ^2</b>						
gesamt	934.8	0.895	0.594	0.951	0.54	0.822
reduziert (AIC)	907.6	0.900	0.593	0.950	0.50	0.813
reduziert (BIC)	911.9	0.899	0.587	0.947	0.54	0.809

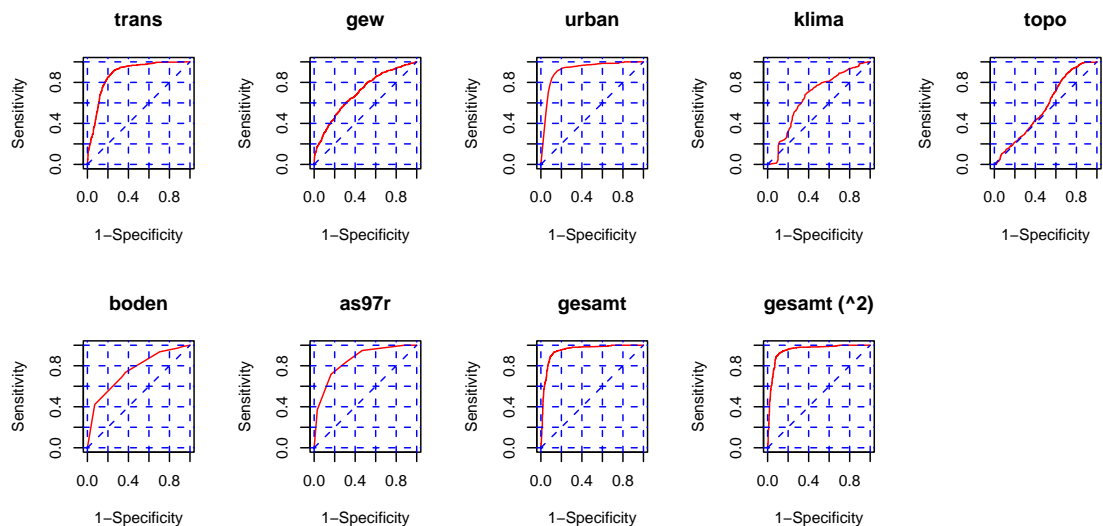


Abbildung 13: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von *Ambrosia artemisiifolia*.

Wurden alle Variablen in das Modell aufgenommen, steigt die Modellgüte nochmals an, wobei das Modell mit quadratischen Termen im Vergleich zum Modell ohne quadratische Terme leicht besser abschneidet. Betrachtet man die reduzierten Modelle des erstgenannten Modells, indizieren die Gütemaße für das AIC-Modell die höchste Güte. Dennoch wurde das BIC-Modell gewählt, da es bei nur wesentlich geringerer Modellgüte den Vorteil einer geringeren Modellkomplexität besitzt (weniger Variablen).

Dieses Modell (gesamt<sup>2</sup>, Variablen-Selektion mit BIC) wird im Folgenden besprochen. Der AUC-Wert von über 0.94 zeigt ein Modell mit hoher Prognosegüte an. Beim Schwellenwert von 0.54 erreichen  $\kappa$ /TSS mit Werten von über 0.8 eine exzellente Modellgüte, die sich auch im hohen korr. D<sup>2</sup>-Wert widerspiegelt. Die durchgeführte Kreuzvalidierung (KV) weist mit einem Wert von ca. 0.9 auf eine hohe Vorhersagegüte hin.

Tabelle 17: Regressionskoeffizienten und Standardfehler der Variablen für das Modell gesamt<sup>2</sup>, Variablenselektion mit BIC von *Ambrosia artemisiifolia*.

Variable	Koeffizient	Standardfehler	
(Intercept)	1.70E-01	7.42E-01	
sqrt(GEB01)	-2.46E-01	2.20E-02	***
STR01	-2.03E-04	5.51E-05	***
I((log(STR02 + 25))^2)	-6.36E-02	1.48E-02	***
log(GWN01 + 25)	2.95E-01	9.18E-02	**
sqrt(GWN02)	1.52E-02	5.27E-03	**
sqrt(GWN03)	5.06E-02	9.56E-03	***
log(SLP + 1)	8.46E-01	3.09E-01	**
I((log(SLP + 1))^2)	-3.75E-01	1.11E-01	***
AS97R2 (Landwirtschaft, int.)	7.70E-01	3.04E-01	*
AS97R3 (Landwirtschaft, ext.)	3.60E-01	3.07E-01	
AS97R4 (Garten/Park/Umschwung)	1.57E+00	3.44E-01	***
AS97R5 (Nass-Standort)	-1.59E+01	6.04E+02	
AS97R6 (sonstiges)	9.18E-01	2.99E-01	**

Signifikanzniveau: \* = 0.05, \*\* = 0.01, \*\*\* = 0.001

Die in Tabelle 17 aufgeführten Regressionskoeffizienten zeigen, dass mit zunehmender Entfernung von Gebäuden (GEB01) und Autobahnen (STR01) die Vorkommenswahrscheinlichkeit von *Ambrosia artemisiifolia* abnimmt. Im Fall von kleineren asphaltierten Straßen (STR02) hingegen, scheint es ein Vorkommensoptimum im mittleren Entfernungsbereich zu geben. Deutlich für einen Verbreitungsschwerpunkt in Siedlungsflächen spricht auch der hohe Regressionskoeffizient für die Faktorstufe Garten/Park/Umschwung der Variablen der Arealstatistik 97 (AS97). Dagegen wird *Ambrosia*



*artemisiifolia* mit zunehmender Entfernung zu Gewässern (GWN01 - GWN03) seltener. Von den topografischen Variablen weist die Hangneigung (SLP) einen signifikanten Einfluß auf das Vorkommen auf. Eine höhere Vorkommenswahrscheinlichkeit tritt im mittleren Wertebereich auf.

Abbildung 14 und Abbildung 15 geben die Vorkommenswahrscheinlichkeit bzw. das prognostizierte Vorkommen von *Ambrosia artemisiifolia* mit der deutlichen Betonung urbaner Regionen um Zürich, Zürichsee und Winterthur kartografisch wieder.

## Vorkommenswahrscheinlichkeit *Ambrosia artemisiifolia*

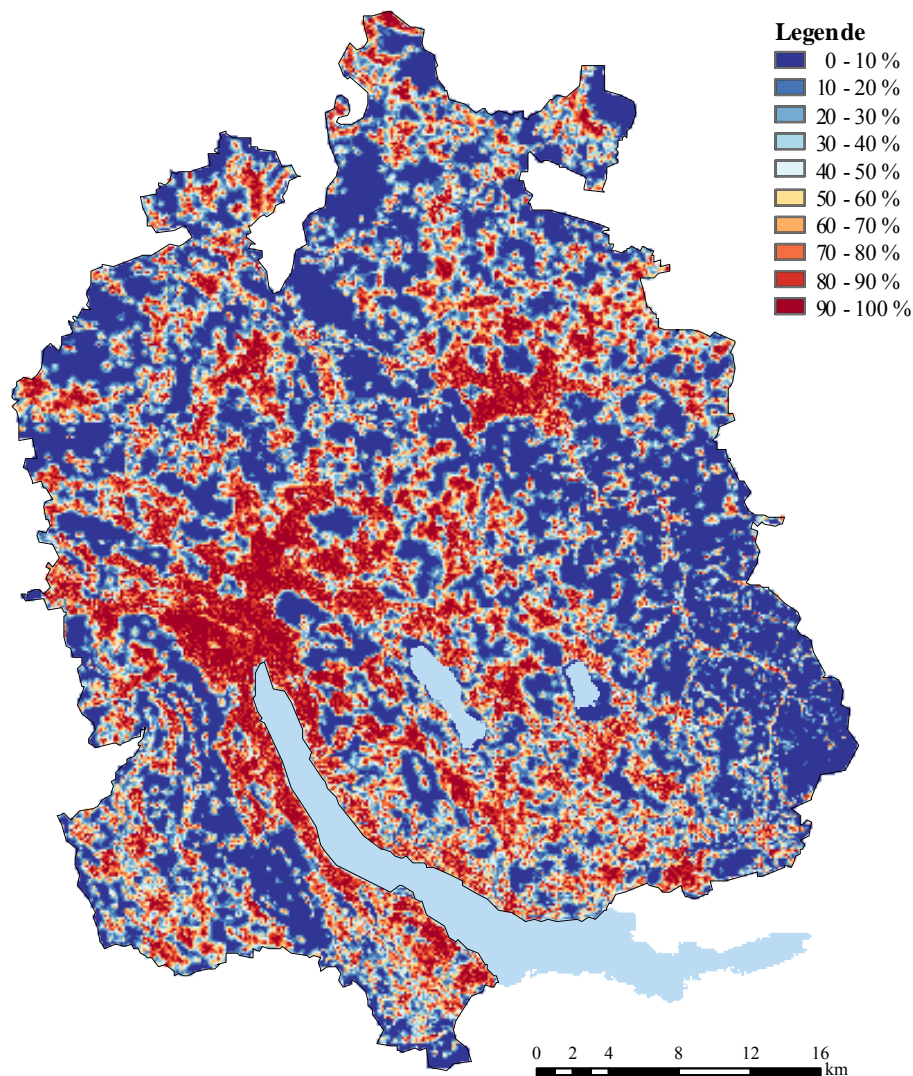


Abbildung 14: Vorkommenswahrscheinlichkeit von *Ambrosia artemisiifolia* (Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

## prognostiziertes Vorkommen *Ambrosia artemisiifolia*

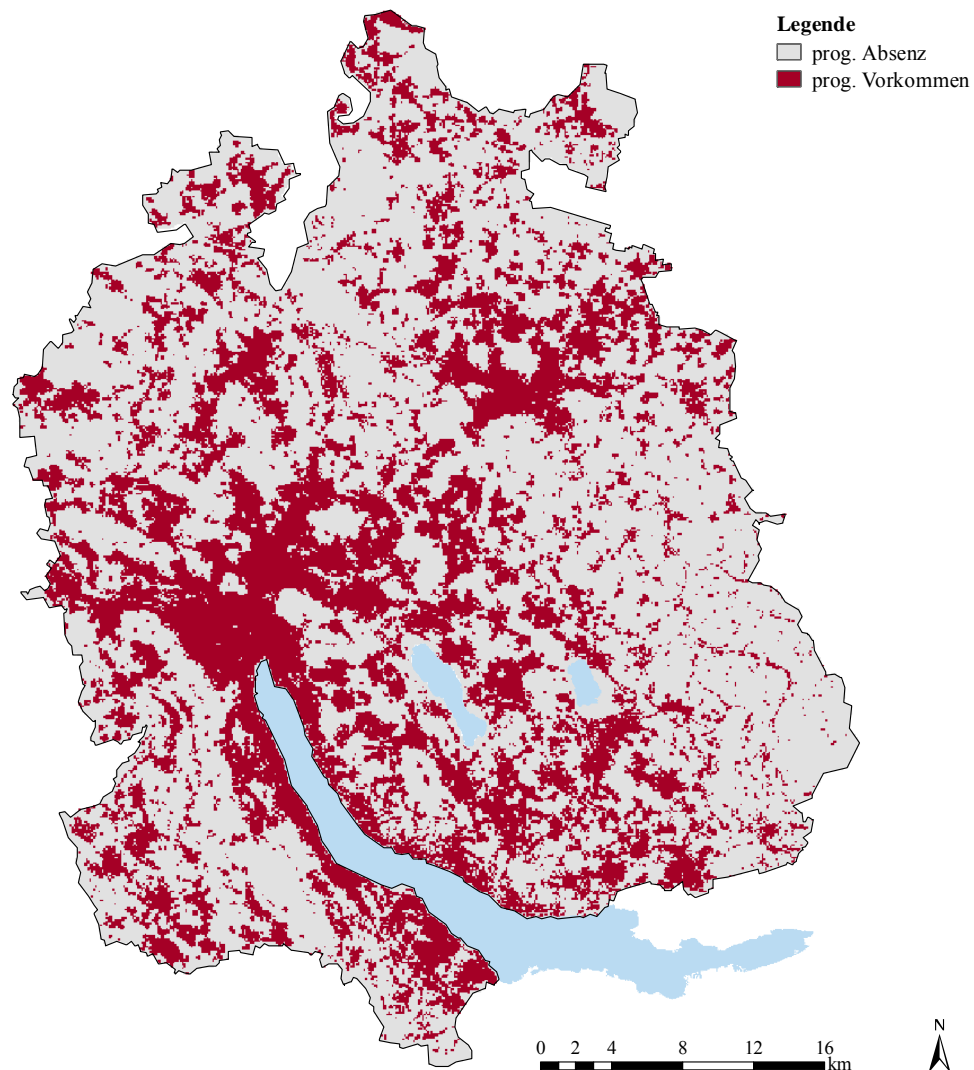


Abbildung 15: Prognostiziertes Vorkommen von *Ambrosia artemisiifolia* entsprechend des Schwellenwertes (max. TSS; Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

### 2.2 *Heracleum mantegazzianum*

Die Vorkommenswahrscheinlichkeit des Riesen-Bärenklaus kann vor allem auf die Entfernung zu Verkehrswegen zurückgeführt werden. Das Teilmodell *trans* weist bereits vergleichsweise niedrige AIC- bzw. hohe AUC-Werte auf (Tabelle 18, siehe auch Abbildung 16). Geringeren Erklärungsgehalt besitzen die Teilmodelle *urban* und *as97r*, wohingegen die Variablen der restlichen Teilmodelle einen nur sehr geringen Einfluss auf das Vorkommen von *Heracleum mantegazzianum* ausüben.

Table 18: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von *Heracleum mantegazzianum*.

GLM	AIC/BIC	KV	korr. D <sup>2</sup>	AUC	Schw.	m. $\kappa$ /TSS
<b>trans</b>						
gesamt	1514.9	0.764	0.263	0.822	0.46	0.542
<b>gew</b>						
gesamt	2043.1	0.539	0.003	0.554	0.48	0.122
reduziert (AIC)	2041.1	0.540	0.004	0.551	0.48	0.114
reduziert (BIC)	2041.2	0.549	0.004	0.537	0.50	0.108
<b>urban</b>						
gesamt	1617.3	0.728	0.211	0.772	0.54	0.470
<b>klima</b>						
gesamt	2021.5	0.575	0.014	0.562	0.50	0.153
reduziert (BIC)	2023.9	0.577	0.012	0.559	0.50	0.161
<b>topo</b>						
gesamt	1966.2	0.548	0.041	0.603	0.38	0.153
reduziert (AIC)	1964.3	0.549	0.042	0.602	0.40	0.154
reduziert (BIC)	1968.3	0.575	0.039	0.605	0.50	0.160
<b>boden</b>						
gesamt	1923.7	0.603	0.062	0.566	0.50	0.206
reduziert (BIC)	1930.2	0.569	0.059	0.551	0.44	0.178
<b>as97r</b>						
gesamt	1806.7	0.670	0.120	0.618	0.50	0.340
<b>gesamt</b>						
gesamt	1397.4	0.784	0.331	0.865	0.52	0.595
reduziert (AIC)	1392.4	0.785	0.331	0.865	0.50	0.595
reduziert (BIC)	1419.1	0.782	0.312	0.853	0.42	0.577
<b>gesamt ^2</b>						
gesamt	1379.3	0.787	0.347	0.874	0.50	0.611
reduziert (AIC)	1368.2	0.793	0.347	0.872	0.48	0.606
reduziert (BIC)	1406.9	0.788	0.318	0.853	0.50	0.589

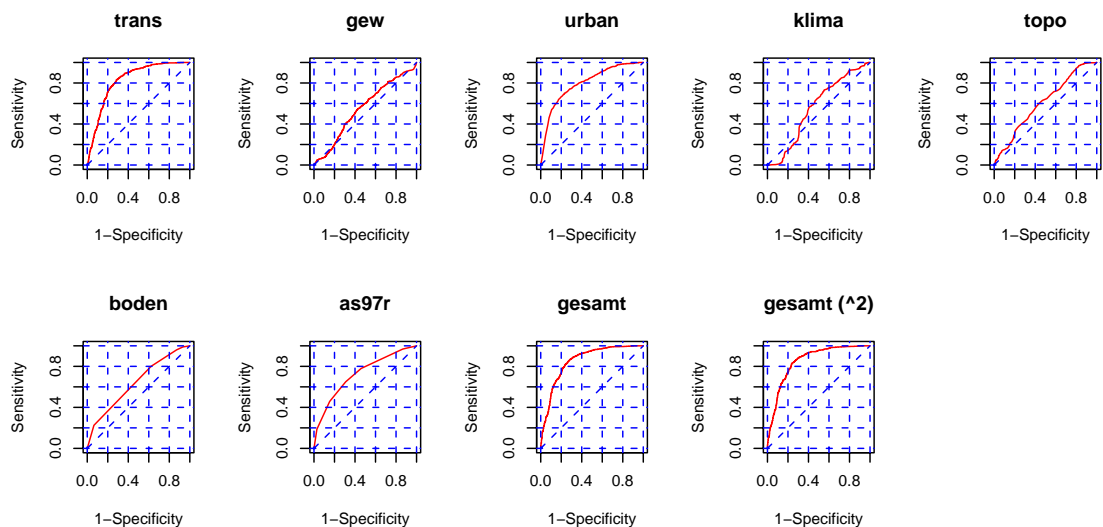


Abbildung 16: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von *Heracleum mantegazzianum*.

Beim Vergleich der beiden Gesamt-Modelle sind nur geringfügige Unterschiede in den Gütemaßen festzustellen. Die AUC-Werte von jeweils über 0.85 indizieren ein brauchbares Modell, was auch durch die max.  $\kappa$ /TSS-Werte von knapp 0.6 bestätigt wird. Die Kreuzvalidierung weist in beiden Fällen auf eine hohe Vorhersagegüte hin.

Für die weitere Analyse wird das Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC verwendet, da es die Theorie der ökologischen Nische besser widerspiegelt, verbunden mit einer geringen Anzahl erklärender Variablen. Wie bereits beim Teilmodell trans sichtbar, zeigt auch Tabelle 19, dass vor allem Entfernungen zu Verkehrswegen einen Einfluss auf das Vorkommen von *Heracleum mantegazzianum* besitzen. Je geringer die Entfernung zu nicht asphaltierten Feld- und Waldwegen (STR03) sowie Fußwegen (STR04) ist, desto größer ist die Wahrscheinlichkeit für ein Vorkommen. Darüber hinaus tragen auch asphaltierte Straßen (STR02) und zu einem geringeren Teil Autobahnen (STR01) sowie Eisenbahnlinien (EIS01) zur Vorhersagegüte des Modells bei, wobei bei ihnen die quadratischen Terme nach der Variablen-Selektion im Modell verbleiben (Optimum im mittleren Entfernungsbereich). Des Weiteren ist das Vorkommen negativ mit der Entfernung zu Gebäuden (GEB01) korreliert, in der Nähe von Gebäuden kommt *Heracleum mantegazzianum* verstärkt vor. Dagegen nimmt mit zunehmender Entfernung zu Flüssen bzw. Kanälen (GWN02) seine Vorkommenswahrscheinlichkeit zu. Als einzige Nichtdistanz-Variable verbleibt die potentiell Globalstrahlung im Dezember (SRAD12) im Modell, allerdings mit nur sehr geringen negativen Regressionskoeffizienten.

In Abbildung 17 und Abbildung 18 ist die Vorkommenswahrscheinlichkeit bzw. das prognostizierte Vorkommen von *Heracleum mantegazzianum* für das oben näher besprochene Modell dargestellt.

Tabelle 19: Regressionskoeffizienten und Standardfehler der Variablen für das Modell  $\text{gesamt}^2$ , Variablenselektion mit BIC von *Heracleum mantegazzianum*.

Variable	Koeffizient	Standardfehler
(Intercept)	8.05E+00	7.29E-01 ***
sqrt(GEB01)	-1.03E-01	1.34E-02 ***
I(STR01 <sup>2</sup> )	-5.35E-08	8.22E-09 ***
I((log(STR02 + 25)) <sup>2</sup> )	-7.88E-02	1.01E-02 ***
log(STR03 + 25)	-6.81E-01	9.45E-02 ***
log(STR04 + 25)	-2.35E-01	7.43E-02 **
I((sqrt(EIS01)) <sup>2</sup> )	-3.00E-04	7.30E-05 ***
sqrt(GWN02)	1.24E-02	3.60E-03 ***
I(SRAD12 <sup>2</sup> )	-1.69E-08	4.31E-09 ***

Signifikanzniveau: \*\* = 0.01, \*\*\* = 0.001

# Vorkommenswahrscheinlichkeit

## *Heracleum mantegazzianum*

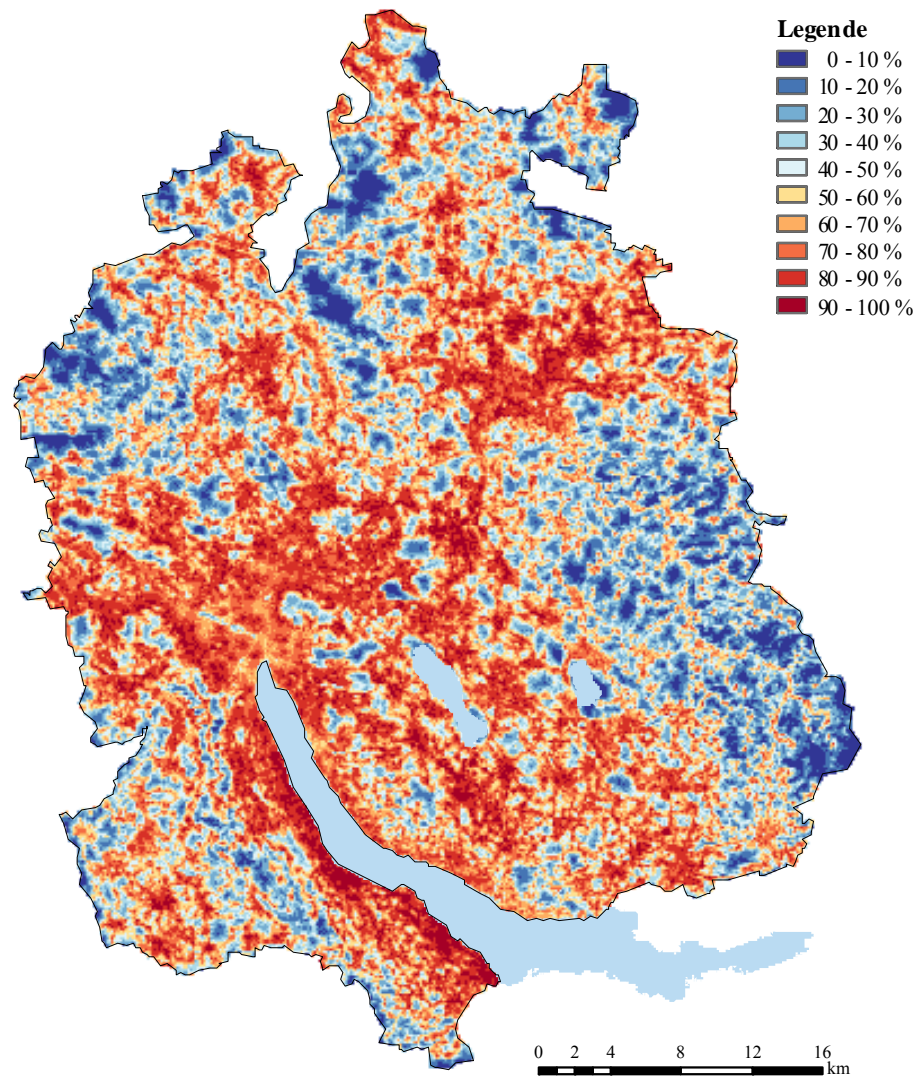


Abbildung 17: Vorkommenswahrscheinlichkeit von *Heracleum mantegazzianum* (Modell  $gesamt^2$ , Variablen-Selektion mit BIC).

## prognostiziertes Vorkommen *Heracleum mantegazzianum*

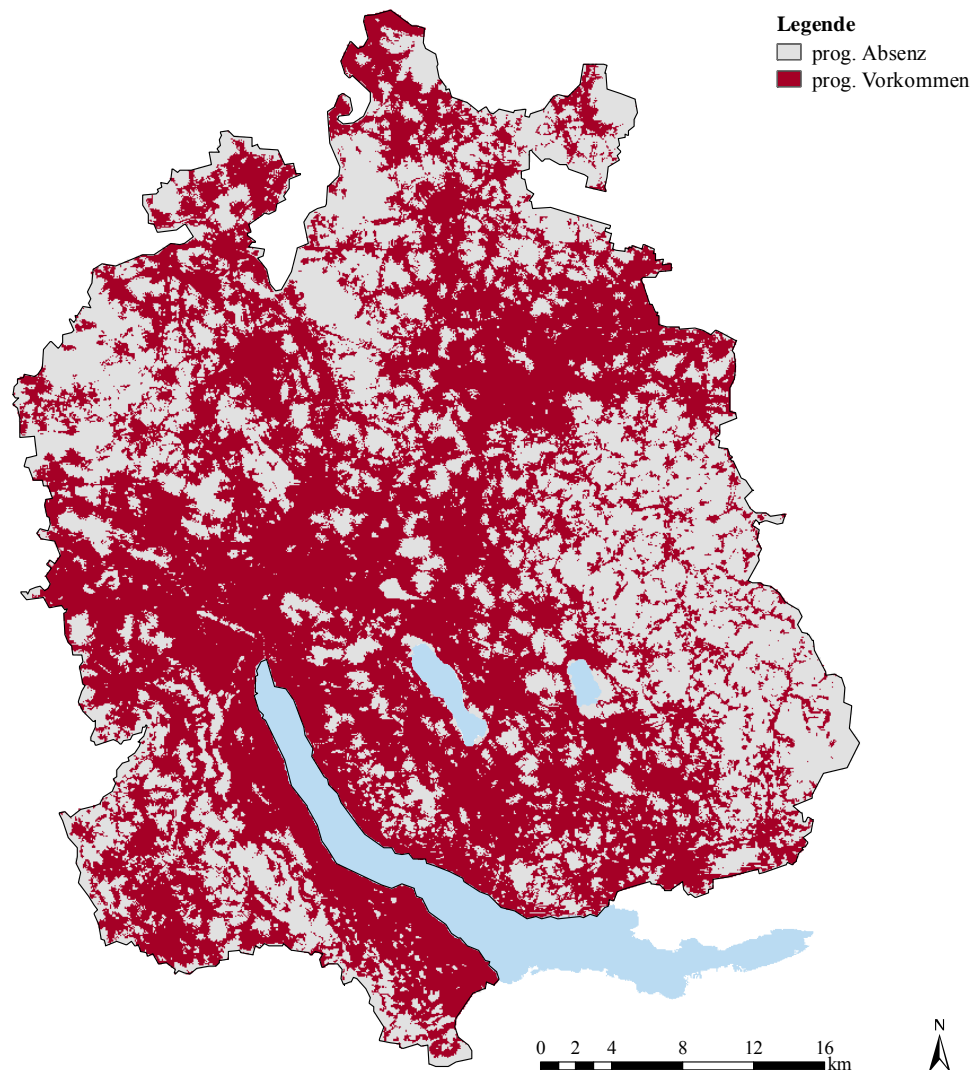


Abbildung 18: Prognostiziertes Vorkommen von *Heracleum mantegazzianum* entsprechend des Schwellenwertes (max. TSS; Modell  $gesamt^2$ , Variablen-Selektion mit BIC).

### 2.3 *Reynoutria japonica*

Das Vorkommen des Japanischen Staudenknöterichs kann zu einem großen Teil bereits durch die Teilmodelle trans und urban (Entfernung zu Verkehrswegen und Gebäuden) erklärt werden. Diese beiden Modelle weisen bei vergleichsweise niedrigen AIC-Werten hohe AUC-Werte auf (Tabelle 20, grafisch in Abbildung 19).

In den beiden Gesamtmodellen kann der Erklärungsgehalt nochmals gesteigert werden und ist beim Modell mit quadratischen Termen am höchsten (AUC: 0.918). Bei letzterem wird durch die schrittweise Variablen-Selektion (red. BIC) ein bezüglich

Modellgüte, Vorhersagegüte und Variablenzahl sehr gut angepasstes Modell gebildet (vergleiche Kennwerte in Tabelle 20 und Tabelle 21), auf welches nachfolgend näher eingegangen wird.

Tabelle 20: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von *Reynoutria japonica*.

GLM		AIC/BIC	KV	korr. D <sup>2</sup>	AUC	Schw.	m. $\kappa$ /TSS
<b>trans</b>	gesamt	2354.2	0.799	0.312	0.851	0.48	0.602
<b>gew</b>	gesamt	3418.0	0.465	0.000	0.500	0.50	0.020
<b>urban</b>	gesamt	2575.3	0.758	0.246	0.789	0.56	0.536
<b>klima</b>	gesamt	3323.8	0.576	0.026	0.603	0.52	0.163
<b>topo</b>	gesamt	3072.1	0.632	0.100	0.694	0.48	0.276
	reduziert (BIC)	3073.5	0.629	0.100	0.697	0.52	0.520
<b>boden</b>	gesamt	2946.2	0.669	0.138	0.514	0.52	0.337
	reduziert (BIC)	2952.2	0.629	0.136	0.382	0.50	0.337
<b>as97r</b>	gesamt	2982.9	0.686	0.127	0.602	0.56	0.372
<b>gesamt</b>	gesamt	2003.3	0.825	0.422	0.902	0.46	0.664
	reduziert (AIC)	1998.8	0.825	0.422	0.902	0.46	0.662
	reduziert (BIC)	2037.4	0.825	0.406	0.895	0.44	0.663
<b>gesamt ^2</b>	gesamt	1879.8	0.831	0.464	0.918	0.46	0.683
	reduziert (AIC)	1870.9	0.833	0.464	0.917	0.50	0.681
	reduziert (BIC)	1893.1	0.832	0.453	0.913	0.42	0.681

Die Regressionskoeffizienten, Standardfehler und Signifikanzniveaus des Modells mit quadratischen Termen (BIC-Modell) stellt Tabelle 21 dar. Davon sollen nur einige besonders relevant erscheinende Variablen besprochen werden. Die Variablen *STR02* und *STR03* (asphaltierte Straße und nicht asphaltierter Feld/Waldweg) sind sowohl mit linearem als auch quadratischem Termen im Modell vertreten. Entsprechend der Regressionskoeffizienten kommt *Reynoutria japonica* überwiegen in unmittelbarer Umgebung dieser Verkehrswege vor, mit zunehmender Entfernung davon sinkt die Vorkommenswahrscheinlichkeit und nimmt bei größeren Entfernungen wieder leicht zu. Ähnliche Muster der Vorkommenswahrscheinlichkeit ergeben sich auch für *GWN02* sowie *GWN03* (Fluss/Kanal und See).

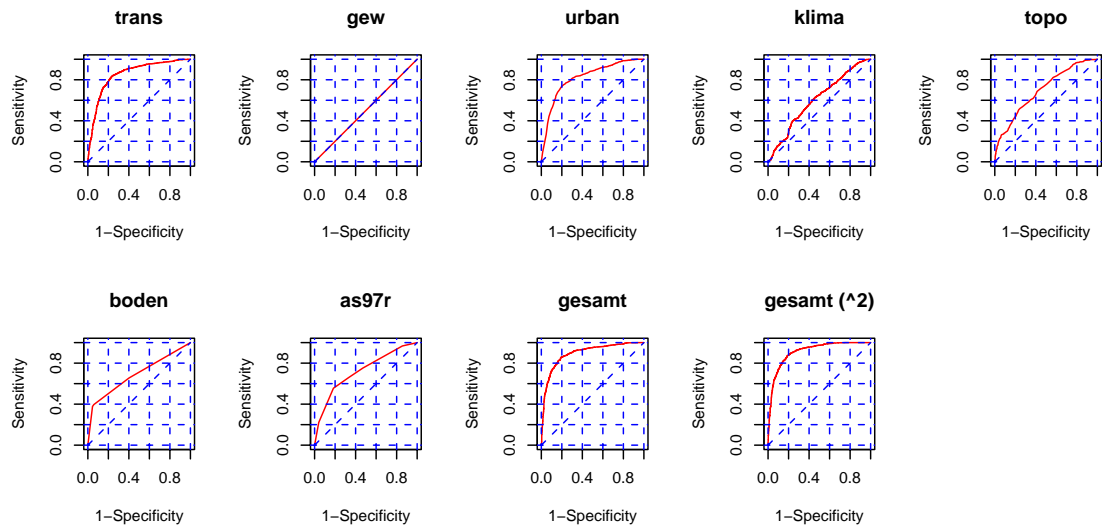


Abbildung 19: Abbildung 20: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von *Reynoutria japonica*.

Tabelle 21: Regressionskoeffizienten und Standardfehler der Variablen für das Modell *gesamt*<sup>2</sup>, Variablenselektion mit BIC von *Reynoutria japonica*.

Variable	Koeffizient	Standardfehler	
(Intercept)	3.23E+01	3.13E+00	***
sqrt(GEB01)	-1.16E-01	1.32E-02	***
log(STR02 + 25)	-2.51E+00	6.91E-01	***
I((log(STR02 + 25)) <sup>2</sup> )	2.26E-01	7.56E-02	**
log(STR03 + 25)	-2.89E+00	8.08E-01	***
I((log(STR03 + 25)) <sup>2</sup> )	3.03E-01	9.35E-02	**
log(STR04 + 25)	-4.15E-01	6.67E-02	***
sqrt(EIS01)	-2.83E-02	4.00E-03	***
log(GWN01 + 25)	-3.14E+00	5.01E-01	***
I((log(GWN01 + 25)) <sup>2</sup> )	3.12E-01	5.09E-02	***
sqrt(GWN02)	-6.76E-02	1.50E-02	***
I((sqrt(GWN02)) <sup>2</sup> )	9.68E-04	1.90E-04	***
TAVE	-7.94E-03	2.32E-03	***
I((TAVE <sup>6</sup> ) <sup>2</sup> )	8.23E-36	1.71E-36	***
I(SRAD12 <sup>2</sup> )	-1.43E-08	4.23E-09	***
TOPS	-1.19E-02	1.98E-03	***
log(SLP + 1)	9.21E-01	2.06E-01	***
I((log(SLP + 1)) <sup>2</sup> )	-2.94E-01	7.18E-02	***
WSPEI1	-1.44E+00	2.17E-01	***
WSPEI2	-1.40E+00	2.29E-01	***
WSPEI3	2,78E-01	4,67E-01	

Signifikanzniveau: \*\* = 0.01, \*\*\* = 0.001



Des Weiteren nimmt die Vorkommenswahrscheinlichkeit mit zunehmender Entfernung zu Gebäuden (GEB01) ab. Von den Topografischen Variablen übt die Hangneigung einen größeren Einfluss auf das Vorkommen von *Reynoutria japonica* aus. Mit zunehmender Hangneigung steigt die Vorkommenswahrscheinlichkeit zunächst an und sinkt, nachdem sie ihren Höhepunkt erreicht hat, wieder leicht ab. Darüber hinaus findet der Japanische Knöterich bei Böden mit einem Wasserspeichervermögen, dass in einem Bereich  $\geq 101$  mm liegt (Kategorie 1 und 2) gute Wachstumsbedingungen.

Abbildung 21 und Abbildung 22 stellen die Ergebnisse des oben besprochenen Habitatmodells kartografisch dar.

## Vorkommenswahrscheinlichkeit *Reynoutria japonica*

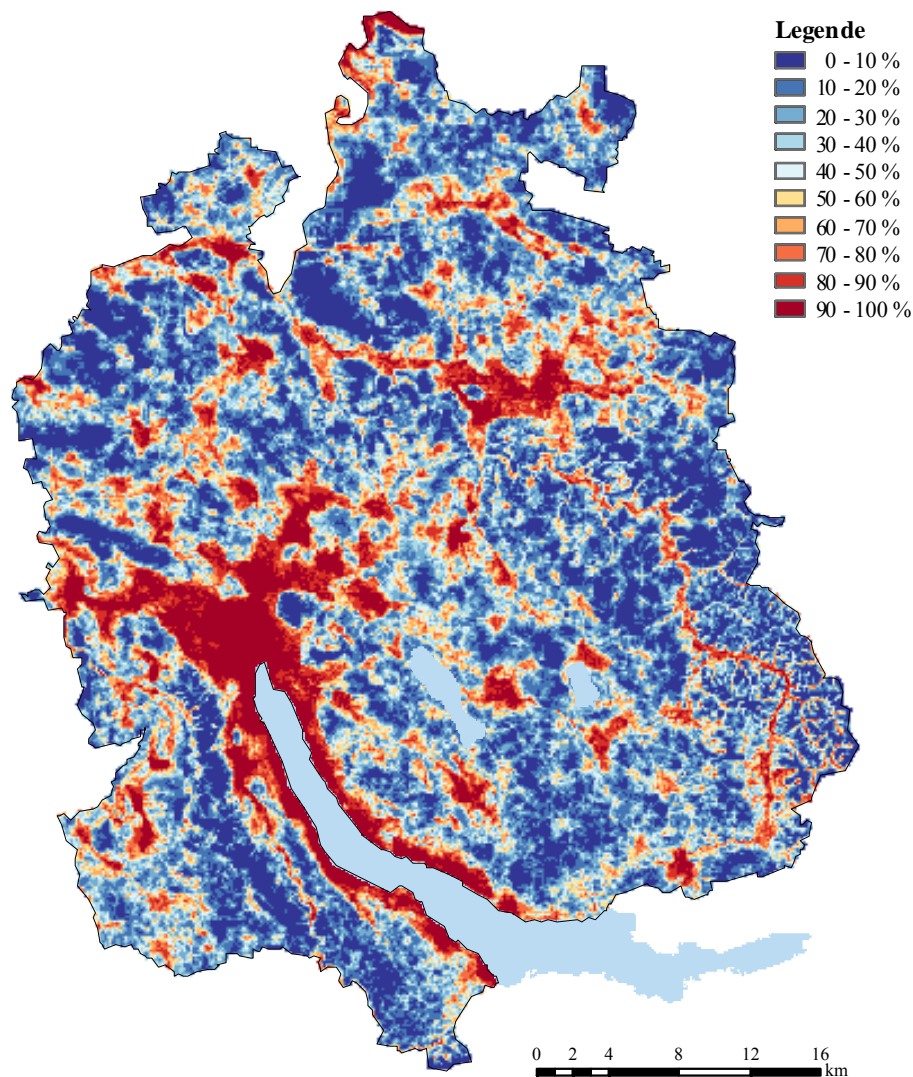


Abbildung 21: Vorkommenswahrscheinlichkeit von *Reynoutria japonica* (Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

## prognostiziertes Vorkommen

*Reynoutria japonica*

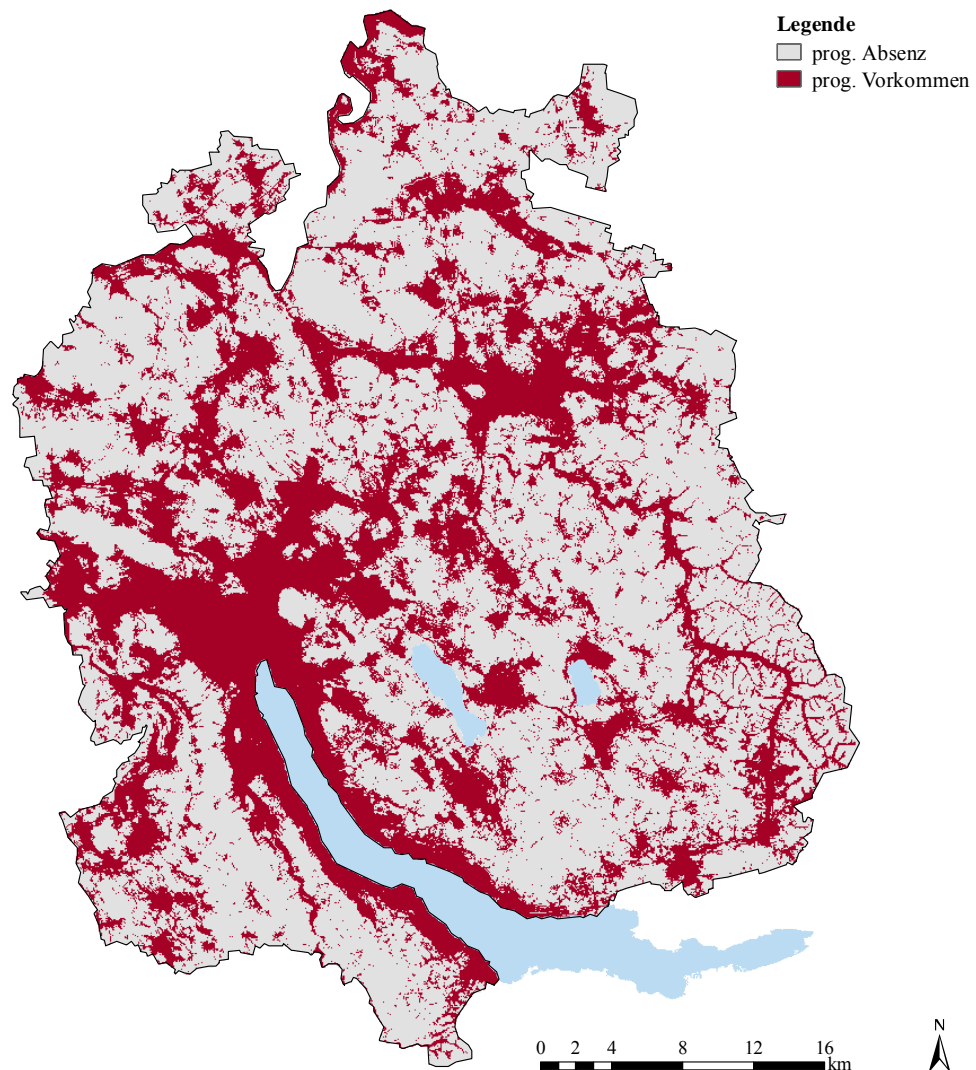


Abbildung 22: Prognostiziertes Vorkommen von *Reynoutria japonica* entsprechend des Schwellenwertes ( $\max. TSS$ ; Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

### 2.4 *Impatiens glandulifera*

Betrachtet man die Teilmodelle, die zur Habitatmodellierung des Drüsigen Springkrauts erstellt wurden, scheint kein Variablenset allein, einen Großteil des Vorkommens/Fehlens der Art erklären zu können. Vielmehr ist es eine Kombination von erklärenden Variablen, die sich auf die einzelnen Teilmodelle verteilen (vergleiche Teil- und Gesamtmodelle in Tabelle 22, Abbildung 23). Beide Gesamtmodelle zeigen eine deutliche Zunahme der Modellgüte gegenüber den Teilmodellen (ablesbar an den vergleichsweise tiefen AIC-, BIC-Werten und den hohen AUC-Werten) und eine jeweils akzeptable Vorhersagegüte (KV: > 0.75).

Tabelle 22: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von *Impatiens glandulifera*.

GLM	AIC/BIC	KV	korr. D <sup>2</sup>	AUC	Schw.	m. $\kappa$ /TSS
<b>trans</b>						
gesamt	4917.5	0.591	0.040	0.630	0.50	0.194
reduziert (AIC)	4916.3	0.781	0.040	0.631	0.48	0.197
<b>gew</b>						
gesamt	4887.7	0.608	0.046	0.641	0.54	0.260
reduziert (AIC)	4885.8	0.607	0.046	0.639	0.52	0.258
<b>urban</b>						
gesamt	5117.1	0.502	0.001	0.512	0.48	0.037
<b>klima</b>						
gesamt	4997.9	0.559	0.024	0.599	0.54	0.197
<b>topo</b>						
gesamt	4686.3	0.639	0.085	0.691	0.52	0.282
<b>boden</b>						
gesamt	5000.4	0.594	0.024	0.584	0.50	0.208
reduziert (BIC)	5008.1	0.567	0.022	0.476	0.50	0.134
<b>as97r</b>						
gesamt	4906.9	0.591	0.042	0.415	0.48	0.185
<b>gesamt</b>						
gesamt	3912.0	0.755	0.240	0.822	0.48	0.526
reduziert (AIC)	3910.9	0.754	0.241	0.822	0.46	0.527
reduziert (BIC)	3929.1	0.751	0.236	0.819	0.50	0.513
<b>gesamt ^2</b>						
gesamt	3634.1	0.781	0.298	0.852	0.46	0.581
reduziert (AIC)	3632.2	0.783	0.298	0.851	0.50	0.578
reduziert (BIC)	3654.2	0.779	0.292	0.847	0.48	0.567

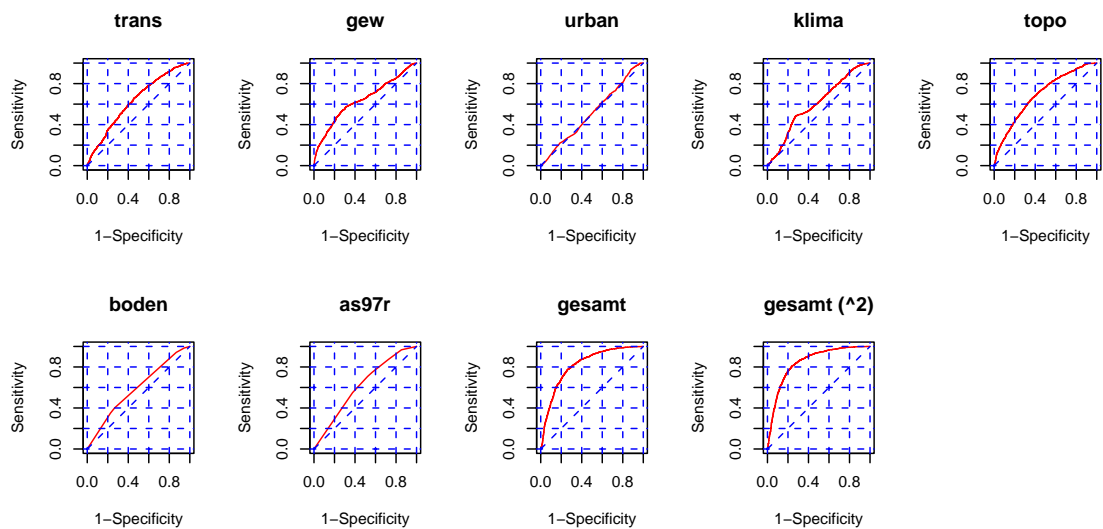


Abbildung 23: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von *Impatiens glandulifera*.

Tabelle 23: Regressionskoeffizienten und Standardfehler der Variablen für das Modell gesamt<sup>2</sup>, Variablenselektion mit BIC von *Impatiens glandulifera*.

Variable	Koeffizient	Standardfehler
(Intercept)	1.75E+01	1.67E+00 ***
sqrt(GEB01)	1.59E-01	3.37E-02 ***
I((sqrt(GEB01))^2)	-5.34E-03	9.99E-04 ***
I(STR01^2)	3.01E-08	5.84E-09 ***
I((log(STR02 + 25))^2)	-5.36E-02	6.37E-03 ***
I((log(STR03 + 25))^2)	-7.67E-02	8.40E-03 ***
I((log(STR04 + 25))^2)	-2.63E-02	5.02E-03 ***
sqrt(EIS01)	-4.28E-02	1.13E-02 ***
I((sqrt(EIS01))^2)	7.75E-04	1.59E-04 ***
log(GWN01 + 25)	-1.85E+00	3.69E-01 ***
I((log(GWN01 + 25))^2)	1.79E-01	3.73E-02 ***
sqrt(GWN02)	-1.66E-01	1.09E-02 ***
I((sqrt(GWN02))^2)	1.64E-03	1.34E-04 ***
TAVE	-6.43E-03	1.76E-03 ***
I((TAVE^6)^2)	-6.48E-36	1.41E-36 ***
I(SRAD12^2)	-1.56E-08	2.74E-09 ***
TOPS	-1.53E-02	1.89E-03 ***
I(TOPS^2)	-5.45E-05	1.95E-05 **
I((log(SLP + 1))^2)	-1.36E-01	1.81E-02 ***
SKELE1	2.55E-01	2.88E-01
SKELE2	1.54E-01	2.25E-01
SKELE3	-7.92E-02	2.20E-01
SKELE4	-2.28E+00	4.23E-01 ***
AS97R2 (Landwirtschaft, int.)	-1.01E+00	1.32E-01 ***
AS97R3 (Landwirtschaft, ext.)	-8.77E-01	1.44E-01 ***
AS97R4 (Garten/Park/Umschwung)	-9.97E-01	2.46E-01 ***
AS97R5 (Nass-Standort)	-1.88E+00	1.81E-01 ***
AS97R6 (sonstiges)	-8.84E-01	1.35E-01 ***
AS97R6 (sonstiges)	-8.84E-01	1.35E-01 ***

Signifikanzniveau: \*\* = 0.01, \*\*\* = 0.001

Wie schon bei den anderen Neophytenarten soll nachfolgend das BIC-Modell des Modells mit quadratischen Termen vorgestellt werden. Auch in diesem Fall erfolgt keine vollständige Nennung aller im Modell verbliebenen Variablen. Vielmehr werden nur die vermeintlich ökologisch relevantesten Variablen besprochen. Neben der Nähe zu Gebäuden (Zunahme der Vorkommenswahrscheinlichkeit mit der Entfernung) sind es vor allem die Variablen, die die Distanz zu Bächen sowie Flüssen und Kanälen kodieren, mit denen das Vorkommen von *Impatiens glandulifera* gut modelliert werden kann (geringe Entfernungen erhöhen die Vorkommenswahrscheinlichkeit, aber beachte auch die signifikanten quadratischen Terme). Habitate, deren Böden geringe Volumenprozent an Kiesen und Steinen aufweisen (Kategorien eins und zwei der Variable

Skelettgehalt, *SKELE*), scheinen gute Standortbedingungen für die Art zu bieten.

Abbildung 24 und Abbildung 25 geben die Vorkommenswahrscheinlichkeit bzw. das prognostizierte Vorkommen von *Impatiens glandulifera* kartografisch wieder. Deutlich erkennbar sind die prognostizierten Vorkommen entlang der Flüsse.

## Vorkommenswahrscheinlichkeit *Impatiens glandulifera*

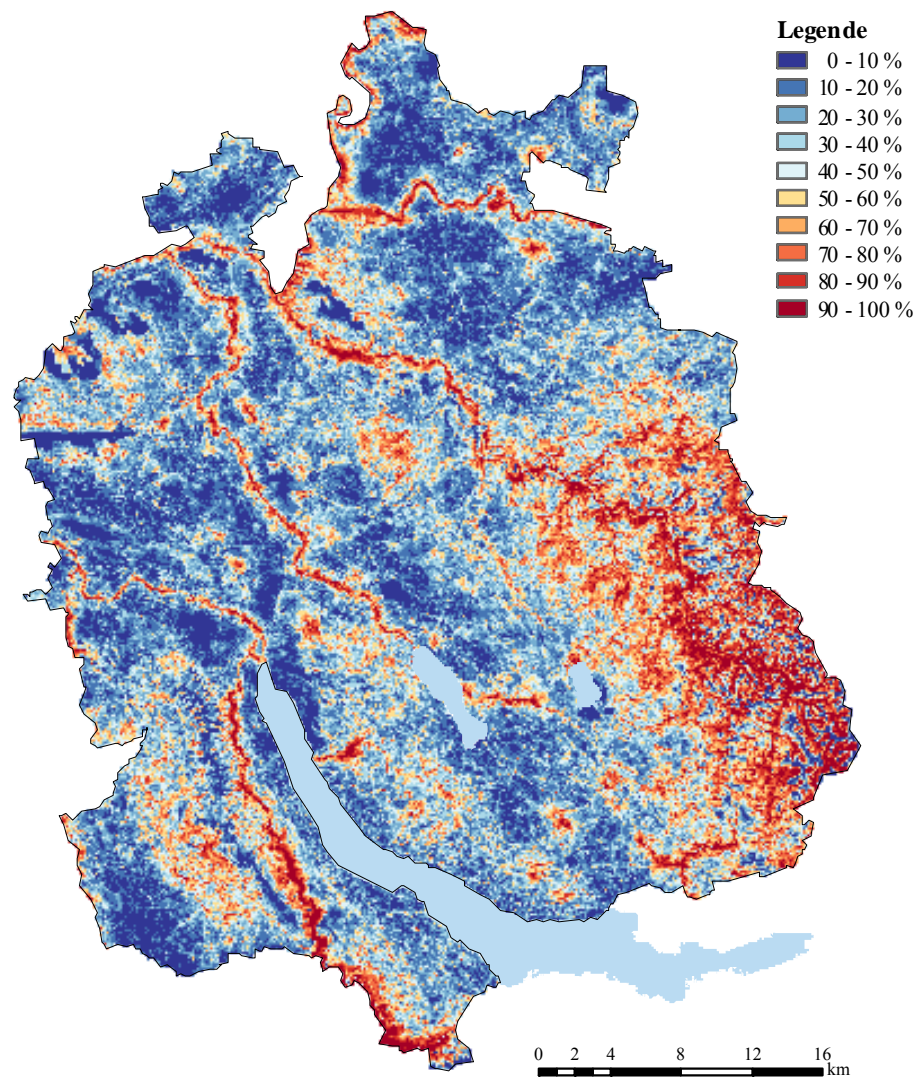


Abbildung 24: Vorkommenswahrscheinlichkeit von *Impatiens glandulifera* (Modell  $gesamt^2$ , Variablen-Selektion mit BIC).

## prognostiziertes Vorkommen *Impatiens glandulifera*

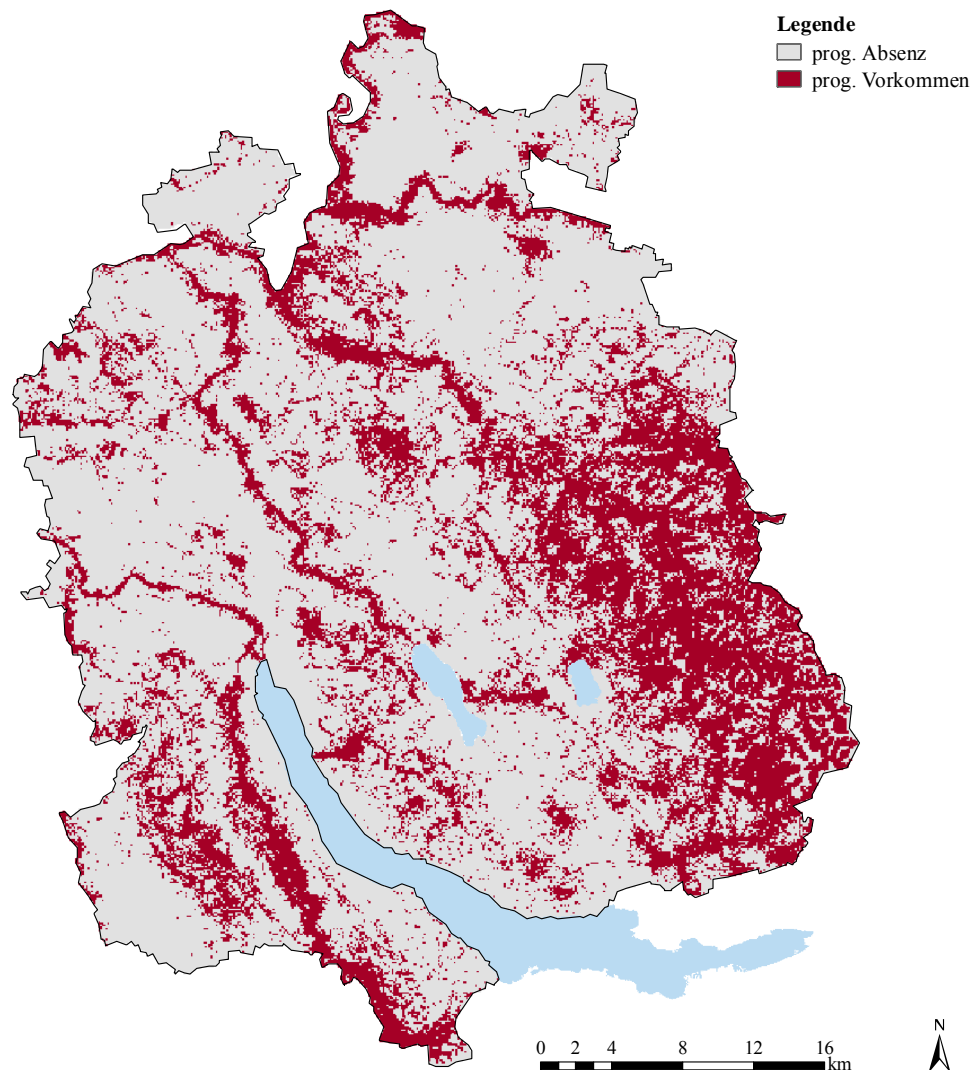


Abbildung 25: Prognostiziertes Vorkommen von *Impatiens glandulifera* entsprechend des Schwellenwertes (max. TSS; Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

### 2.5 *Rhus typhina*

Auch das Vorkommen des Essigbaums ist klar an urbane Strukturen gebunden. Die Teilmodelle *trans* und *urban* erklären schon einen großen Teil des Vorkommensmusters (AUC-Werte von knapp unter 0.9 bzw. über 0.9, deutlich erkennbar auch in den entsprechenden ROC-Plots in Abbildung 26).

Tabelle 24: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von *Rhus typhina*.

GLM	AIC/BIC	KV	korr. D <sup>2</sup>	AUC	Schw.	m. $\kappa$ /TSS
<b>trans</b>						
gesamt	366.1	0.832	0.378	0.879	0.42	0.649
reduziert (AIC)	362.4	0.835	0.381	0.878	0.48	362.4
<b>gew</b>						
gesamt	561.0	0.593	0.034	0.627	0.50	0.221
reduziert (BIC)	561.4	0.589	0.031	0.606	0.54	0.221
<b>urban</b>						
gesamt	272.9	0.870	0.533	0.938	0.58	0.769
<b>klima</b>						
gesamt	559.4	0.583	0.036	0.599	0.44	0.279
<b>topo</b>						
gesamt	551.3	0.624	0.051	551.3	0.48	0.322
reduziert (AIC)	549.6	0.637	0.053	549.6	0.46	0.317
<b>boden</b>						
gesamt	535.3	0.589	0.081	0.500	0.66	0.240
reduziert (AIC)	531.3	0.608	0.087	0.499	0.66	0.240
<b>as97r</b>						
gesamt	414.7	0.761	0.293	0.756	0.60	0.529
<b>gesamt</b>						
gesamt	257.9	0.858	0.619	0.961	0.48	0.803
reduziert (AIC)	242.3	0.865	0.619	0.959	0.36	0.793
reduziert (BIC)	255.7	0.858	0.567	0.947	0.38	0.740
<b>gesamt ^2</b>						
gesamt	259.7	0.867	0.655	0.972	0.44	0.822
reduziert (AIC)	237.3	0.880	0.639	0.965	0.40	0.822
reduziert (BIC)	248.2	0.863	0.588	0.952	0.38	0.764

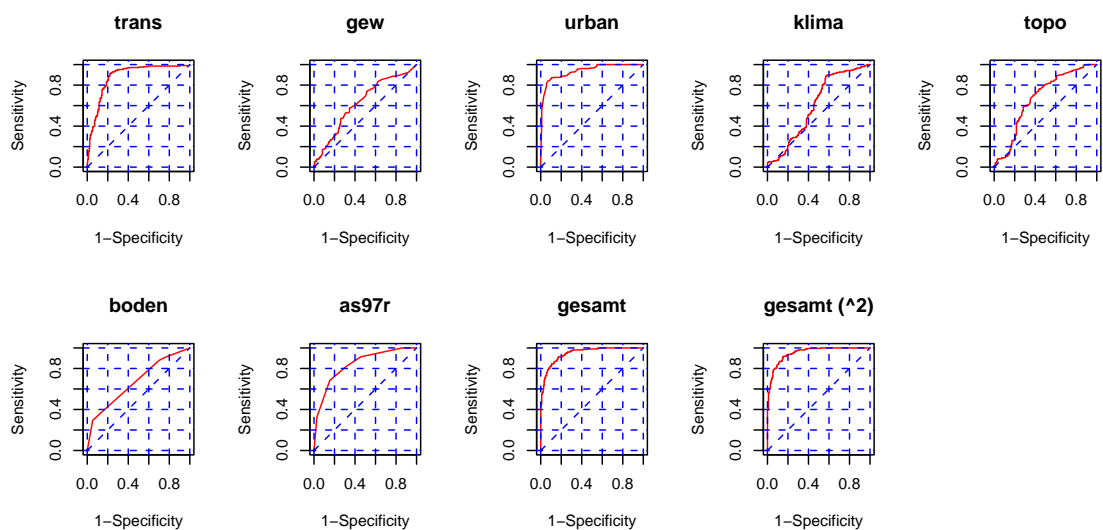


Abbildung 26: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von *Rhus typhina*.

Auch im Modellierungsfall von *Rhus typhina* wurde das nach BIC reduzierte Modell mit quadratischen Termen (gesamt<sup>2</sup>, red. BIC) gewählt. Es weist mit seinem sehr hohen AUC-Wert von über 0.95 auf eine exzellente Modellgüte hin. Darüber hinaus ist der BIC-Wert vergleichsweise niedrig und auch die korr. D<sup>2</sup> und der max. κ bzw. TSS-Wert indizieren gute Modellgüte. Die Resultate der Kreuzvalidierung zeugen von einer sehr gute Vorhersagegüte des Modells.

In Tabelle 25 finden sich mit Ausnahme der Hangneigung nur Variablen, die die Distanz zu Gebäuden, (GEB01), Autobahnen und anderen asphaltierten Straßen (STR01, STR02) sowie Flüssen und Kanälen (GWN02) beschreiben. Dabei nimmt die Vorkommenswahrscheinlichkeit mit zunehmender Entfernung ab (GEB01), oder weist aufgrund eines quadratischen Terms einen mittleren Entfernungsbereich mit verringerter Vorkommenswahrscheinlichkeit auf (STR01, STR02, GWN02). Im Falle der Hangneigung wurde ein statistisch signifikanter positiver linearer Zusammenhang nachgewiesen, zunehmende Hangneigung wirkt sich positiv auf die Standorteigenschaften von *Rhus typhina* aus.

Tabelle 25: Regressionskoeffizienten und Standardfehler der Variablen für das Modell gesamt<sup>2</sup>, Variablenselektion mit BIC von *Rhus typhina*.

Variable	Koeffizient	Standardfehler
(Intercept)	7.14E+00	9.71E-01 ***
sqrt(GEB01)	-4.14E-01	4.69E-02 ***
I(STR01 <sup>2</sup> )	-5.01E-08	1.96E-08 *
I((log(STR02 + 25)) <sup>2</sup> )	-7.33E-02	2.55E-02 **
sqrt(GWN02)	-1.08E-01	4.01E-02 **
I((sqrt(GWN02)) <sup>2</sup> )	1.43E-03	5.18E-04 **
log(SLP + 1)	6.17E-01	1.64E-01 ***

Signifikanzniveau: \* = 0.05, \*\* = 0.01, \*\*\* = 0.001

Abbildung 27 und Abbildung 28 geben einen Überblick über die Vorkommenswahrscheinlichkeit und das prognostizierte Vorkommen von *Rhus typhina*. In der ersten Abbildung ist die scharfe Begrenzung hoher Vorkommenswahrscheinlichkeiten auf die urbanen Gebiete, wie beispielsweise das Gebiet rund um Zürich, den Zürichsee und Winterthur, deutlich erkennbar.



# Vorkommenswahrscheinlichkeit

## *Rhus typhina*

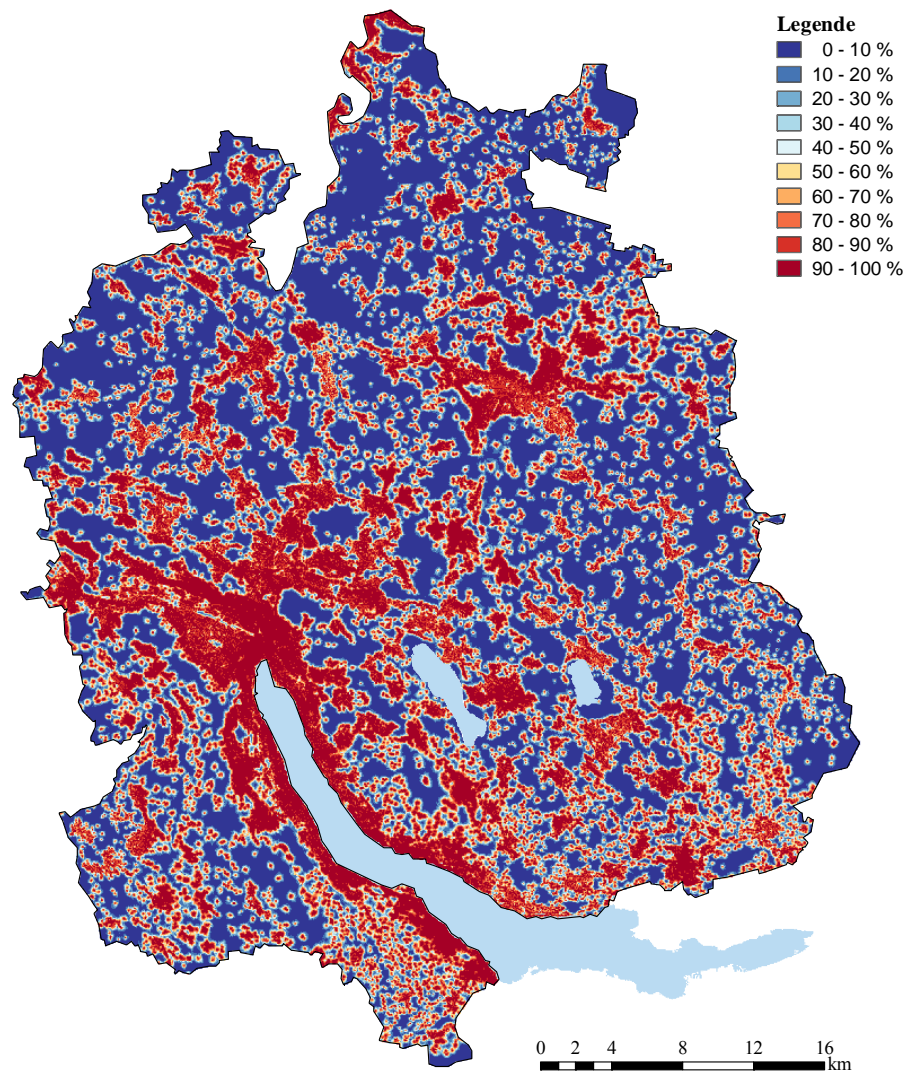


Abbildung 27: Vorkommenswahrscheinlichkeit von *Rhus typhina* (Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

## prognostiziertes Vorkommen

### *Rhus typhina*

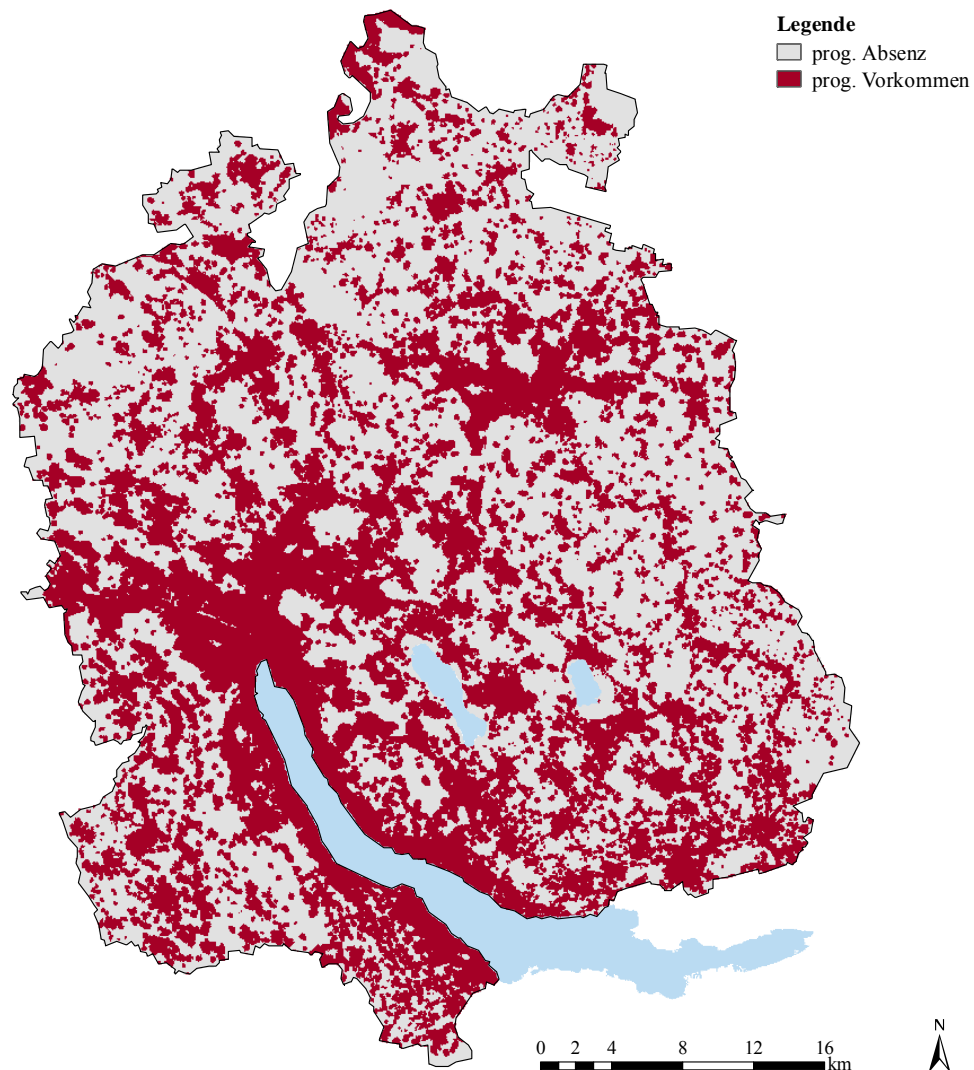


Abbildung 28: Prognostiziertes Vorkommen von *Rhus typhina* entsprechend des Schwellenwertes (max. TSS; Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

## 2.6 *Solidago canadensis*

Betrachtet man die Teilmodelle für die Kanadische Goldrute zeigt lediglich das Teilmodell boden eine gewisse Modellgüte (Tabelle 26 und Abbildung 29). Erst die Integration aller Variablen in das Gesamtmodell verbessert diese substantiell. Auch bei dieser Art unterscheiden sich die beiden Gesamtmodelle (mit und ohne quadratische Terme) hinsichtlich der Kenngrößen nicht wesentlich voneinander, das quadratische Modell modelliert das Vorkommen/Fehlen von *Solidago canadensis* etwas besser (AUC: > 0.8).

Table 26: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von *Solidago canadensis*.

GLM	AIC/BIC	KV	korr. D <sup>2</sup>	AUC	Schw.	m. $\kappa$ /TSS
<b>trans</b>						
gesamt	3099,5	0,590	0,052	0,644	0,46	0,192
<b>gew</b>						
gesamt	3245,9	0,551	0,007	0,564	0,52	0,121
reduziert (BIC)	3249,1	0,540	0,005	0,557	0,54	0,109
<b>urban</b>						
gesamt	3215,6	0,525	0,016	0,574	0,44	0,108
<b>klima</b>						
gesamt	3215,6	0,525	0,016	0,574	0,44	0,108
reduziert (BIC)	3216,6	0,539	0,015	0,569	0,42	0,097
<b>topo</b>						
gesamt	3260,1	0,542	0,002	0,544	0,50	0,099
reduziert (AIC)	3258,4	0,543	0,003	0,523	0,50	0,101
reduziert (BIC)	3259,1	0,522	0,002	0,416	0,48	0,073
<b>boden</b>						
gesamt	2911,3	0,687	0,110	0,683	0,52	0,405
<b>as97r</b>						
gesamt	3230,8	0,563	0,012	0,449	0,48	0,127
<b>gesamt</b>						
gesamt	2645,1	0,714	0,197	0,794	0,62	0,477
reduziert (AIC)	2638,7	0,714	0,197	0,795	0,62	0,472
reduziert (BIC)	2676,2	0,701	0,184	0,782	0,58	0,433
<b>gesamt ^2</b>						
gesamt	2552,3	0,724	0,229	0,817	0,58	0,489
reduziert (AIC)	2540,7	0,728	0,230	0,816	0,58	0,495
reduziert (BIC)	2559,9	0,719	0,221	0,807	0,60	0,491

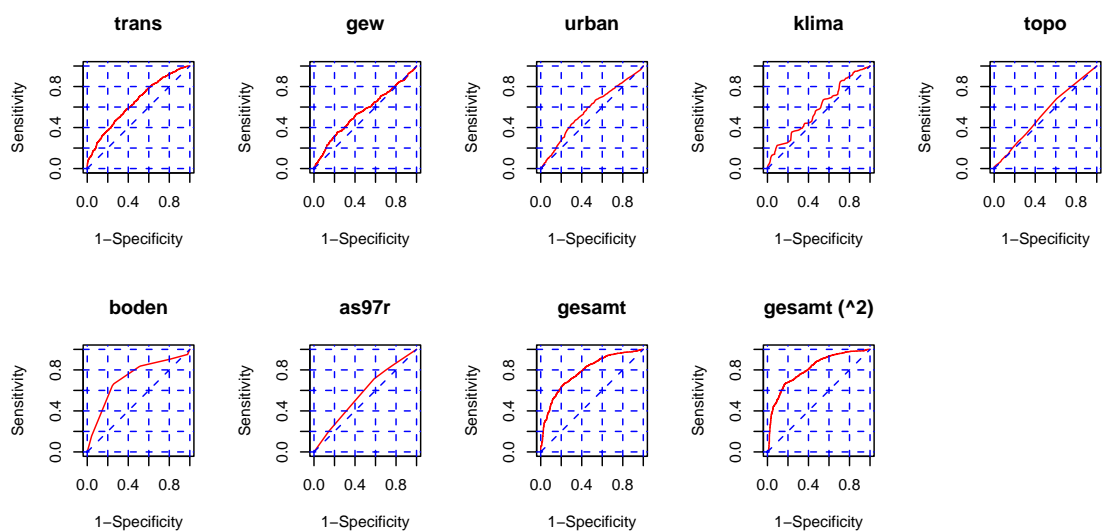


Abbildung 29: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von *Solidago canadensis*.

Wiederum fiel die Wahl auf die BIC-Variante des  $\text{gesamt}^2$ -Modells. Es weist eine akzeptable Vorhersagegüte (KV:  $> 0.7$ ) und Modellgüte (AUC: 0.807) auf, verbunden mit einer überschaubaren Anzahl Termen (Tabelle 27). Wiederum sollen nur ausgewählte Variablen näher vorgestellt werden. Die Entfernungen zu Eisenbahngleisen (EIS01) und Flüssen bzw. Kanälen (GWN02) ist für das Vorkommen von *Solidago canadensis* relevant. Dabei ist die Wahrscheinlichkeit des Vorkommens in unmittelbarer Nähe zu diesen Strukturen besonders hoch, sinkt dann ab um in größerer Entfernung wieder leicht zuzunehmen. Weiterhin bilden die Distanzen zu asphaltierten Straßen, nicht asphaltierten Feld/Waldwegen und Fußwegen (STR02, STR03, STR04 - weitgehend fallende Regressionsgeraden) und die Distanzen zu Autobahnen (STR01 - weitgehend parabelförmiger Verlauf mit Optimum im mittleren Entfernungsbereich) wichtige Komponenten des Modells. Darüber hinaus scheinen sowohl geringe (SKELE, 1-4 Volumenprozent) als auch deutlich höhere Skelettanteile des Bodens ( $>11$  Volumenprozent) die Wahrscheinlichkeit des Vorkommens von *Solidago canadensis* positiv zu beeinflussen.

Tabelle 27: Regressionskoeffizienten und Standardfehler der Variablen für das Modell  $\text{gesamt}^2$ , Variablenselektion mit BIC von *Solidago canadensis*.

Variable	Koeffizient	Standardfehler	
(Intercept)	-1.26E+01	2.00E+00	***
STR01	7.16E-04	1.36E-04	***
I(STR01 <sup>2</sup> )	-1.23E-07	2.53E-08	***
log(STR02 + 25)	1.86E+00	5.99E-01	**
I((log(STR02 + 25)) <sup>2</sup> )	-1.85E-01	6.38E-02	**
I((log(STR03 + 25)) <sup>2</sup> )	-2.94E-02	9.71E-03	**
log(STR04 + 25)	2.06E+00	5.20E-01	***
I((log(STR04 + 25)) <sup>2</sup> )	-1.76E-01	5.02E-02	***
sqrt(EIS01)	-5.64E-02	1.22E-02	***
I((sqrt(EIS01)) <sup>2</sup> )	6.50E-04	2.02E-04	**
sqrt(GWN02)	-4.52E-02	1.25E-02	***
I((sqrt(GWN02)) <sup>2</sup> )	7.42E-04	1.57E-04	***
I((sqrt(GWN03)) <sup>2</sup> )	5.16E-04	1.21E-04	***
I((TAVE <sup>6</sup> ) <sup>2</sup> )	1.07E-35	1.01E-36	***
TOPS	6.26E-03	1.95E-03	**
SKELE1	3.22E+00	4.83E-01	***
SKELE2	1.81E-01	4.22E-01	
SKELE3	2.17E+00	4.08E-01	***
SKELE4	1.80E+00	6.40E-01	**
WSPEI1	1.25E-01	3.62E-01	
WSPEI2	-6.48E-01	3.72E-01	.
WSPEI3	NA	NA	

Signifikanzniveau: . = 0.1, \*\* = 0.01, \*\*\* = 0.001

Abbildung 30 und Abbildung 31 zeigen die Vorkommenswahrscheinlichkeit und das prognostizierte Vorkommen von *Solidago canadensis* für das Modell mit quadratischen Termen (BIC-Modell).

## Vorkommenswahrscheinlichkeit *Solidago canadensis*

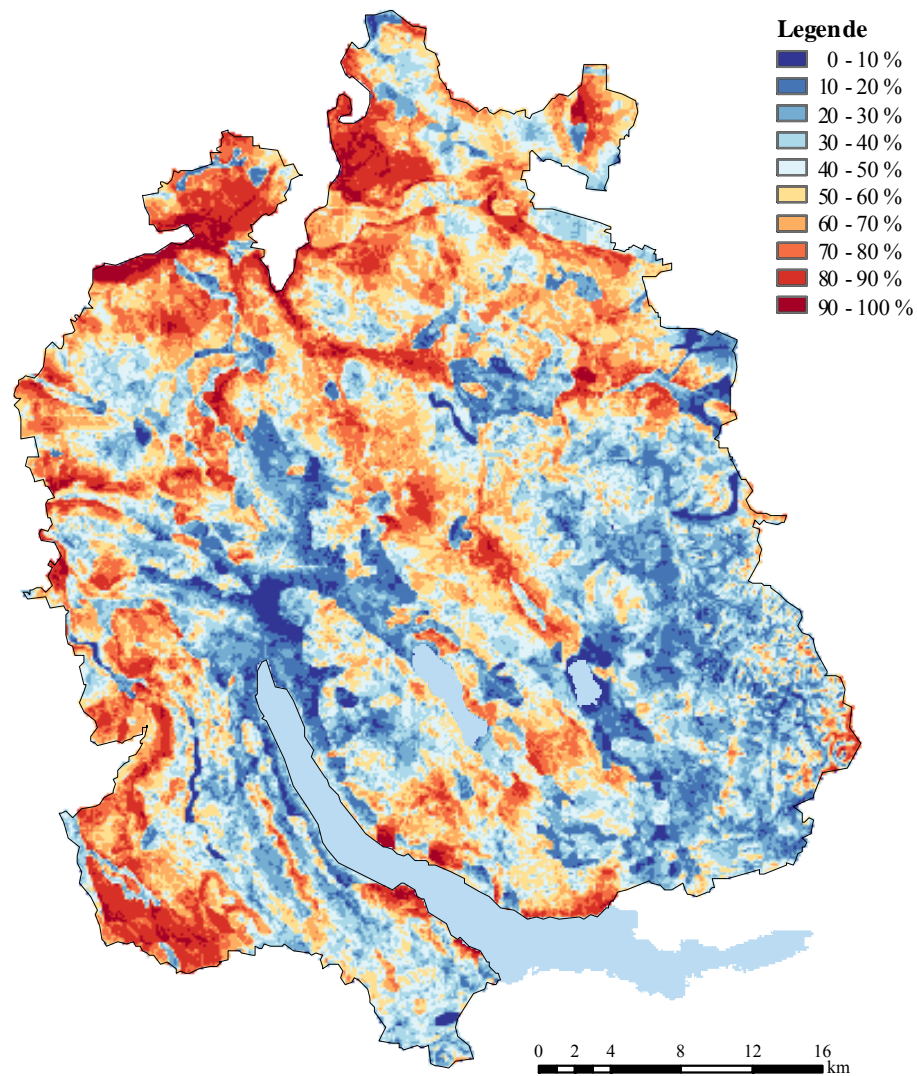


Abbildung 30: Vorkommenswahrscheinlichkeit von *Solidago canadensis* (Modell  $\text{gesamt}^2$ , Variablen-selektion mit BIC).

## prognostiziertes Vorkommen *Solidago canadensis*

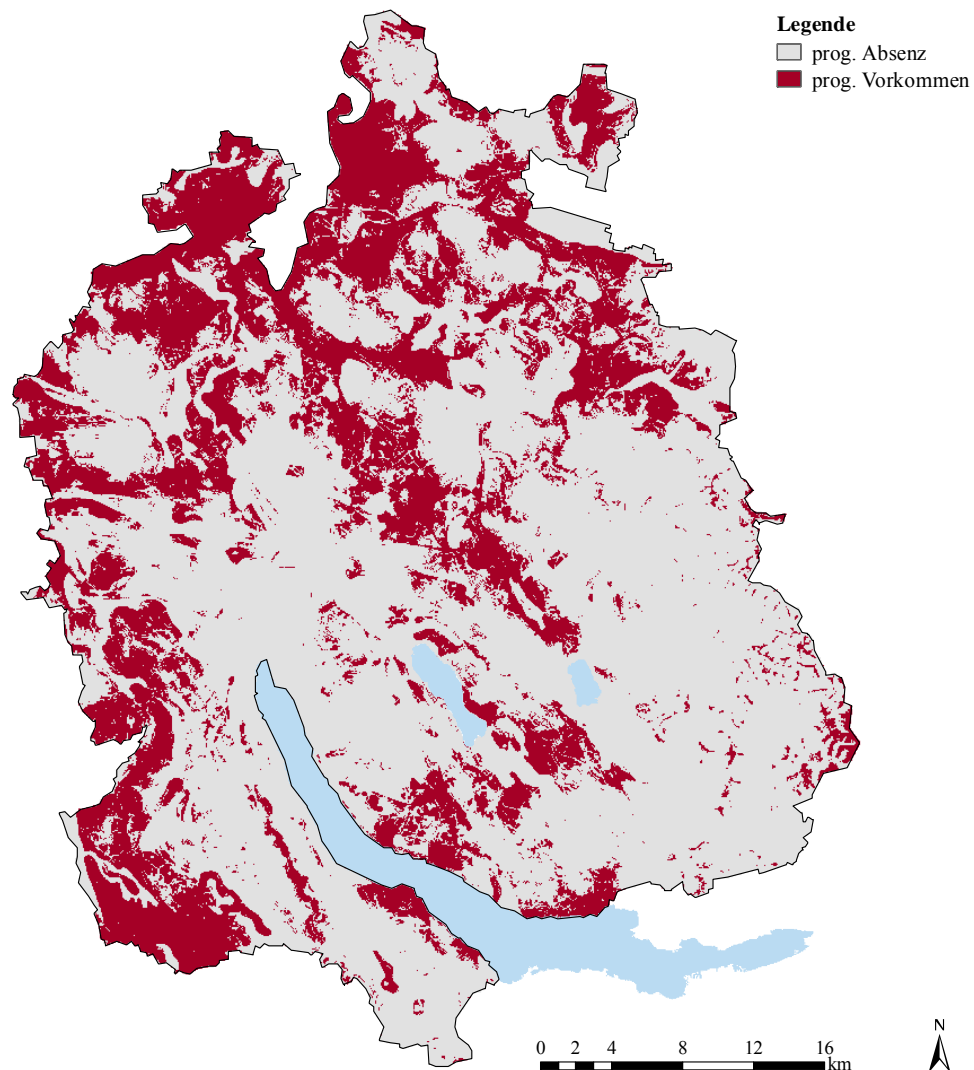


Abbildung 31: Prognostiziertes Vorkommen von *Solidago canadensis* entsprechend des Schwellenwertes ( $\max. TSS$ ; Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

### 2.7 *Solidago gigantea*

Für die Riesen-Goldrute war als einzige modellierte Art kein ausgeglichenes Verhältnis von Vorkommen und Pseudo-Absenz realisierbar. Deshalb ergeben sich unterschiedliche Werte für maximales  $\kappa$  und max. TSS (Tabelle 28). Keines der Teilmodelle allein kann das Vorkommen von *Solidago gigantea* hinreichend gut erklären. Lediglich die beiden Modelle trans und urban zeigen eine gewisse Prognosegüte (Tabelle 28 und Abbildung 32).

Table 28: Übersicht wichtiger Kennwerte der berechneten Modelle zur Vorkommensvorhersage von *Solidago gigantea*.

GLM	AIC/BIC	KV	korr. D <sup>2</sup>	AUC	Schw.	Kappa	Schw.	TSS
<b>trans</b>								
gesamt	17790.0	0.742	0.181	0.778	0.56	0.450	0.60	0.452
<b>gew</b>								
gesamt	19875.0	0.673	0.085	0.695	0.58	0.318	0.60	0.319
<b>urban</b>								
gesamt	18003.0	0.758	0.171	0.727	0.44	0.485	0.48	0.459
<b>klima</b>								
gesamt	21713.0	0.616	0.001	0.527	0.60	0.070	0.60	0.060
reduziert (AIC)	21714.0	0.616	0.000	0.492	0.60	0.060	0.60	0.051
<b>topo</b>								
gesamt	20490.0	0.641	0.057	0.671	0.64	0.264	0.66	0.281
<b>boden</b>								
gesamt	20567.0	0.618	0.053	0.516	0.80	0.205	0.80	0.242
reduziert (BIC)	20570.0	0.618	0.053	0.507	0.68	0.202	0.70	0.235
<b>as97r</b>								
gesamt	17729.0	0.712	0.184	0.565	0.50	0.360	0.50	0.342
<b>gesamt</b>								
gesamt	14014.0	0.808	0.356	0.875	0.52	0.589	0.52	0.581
reduziert (AIC)	14011.0	0.809	0.356	0.875	0.50	0.589	0.52	0.581
<b>gesamt ^2</b>								
gesamt	12827.0	0.822	0.412	0.895	0.50	0.620	0.66	0.614
reduziert (AIC)	12825.0	0.822	0.412	0.895	0.50	0.620	0.62	0.614
reduziert (BIC)	12850.0	0.822	0.410	0.895	0.52	0.620	0.66	0.611

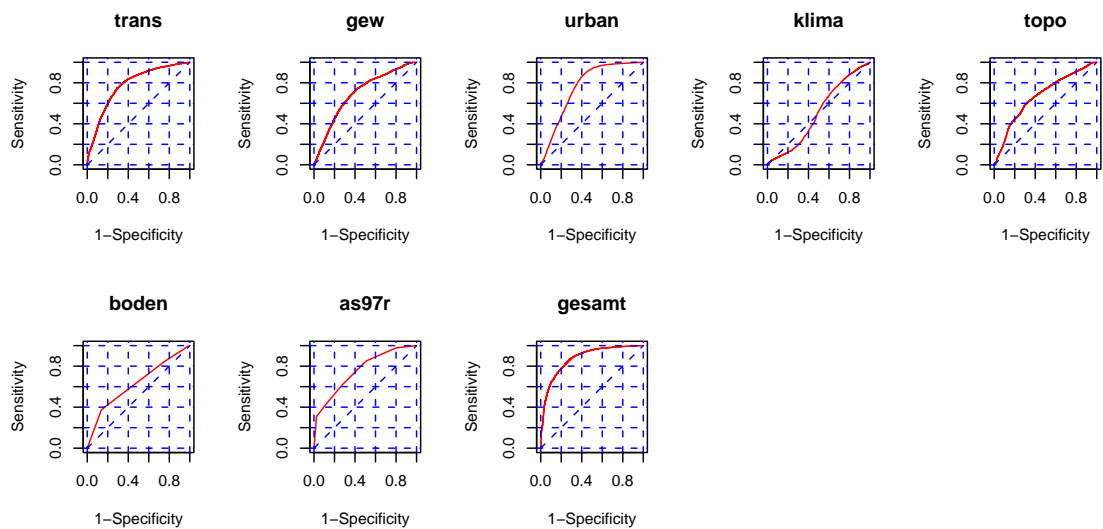


Abbildung 32: ROC-Plots der berechneten reduzierten Modelle (Methode BIC) zur Vorkommensvorhersage von *Solidago gigantea*.

Für die weitere Analyse und Visualisierung wurde das Modell gesamt<sup>2</sup>, reduziert (BIC) ausgewählt. Es besitzt eine hohe Vorhersagegüte (KV: 0.822), verbunden mit einem niedrigen BIC- und einem hohen AUC-Wert sowie vergleichsweise hohen Werten für korr. D<sup>2</sup> und max. TSS. Tabelle 29 enthält nähere Informationen zu diesem Modell. Einige der dort aufgeführten Variablen sollen näher besprochen werden. Entsprechend des Regressionskoeffizienten für GEB01 ist die Vorkommenswahrscheinlichkeit in unmittelbarer Gebäudenähe gering, steigt dann schnell an und nimmt mit steigender Entfernung wieder ab. Ein ähnliches Muster zeigt sich auch bei den Variablen STR03 (Distanz zu nicht asphaltierter Feld/Waldweg), GWN01 (Distanz zu Bach) sowie GWN02 (Distanz zu Fluss/Kanal). Die Regressionskurven mit den Variablen EIS01 (Distanz zu Eisenbahngleis) und GWN03 (Distanz zu See) zeigen dahingegen einen umgekehrten Verlauf. Nach hohen Vorkommenswahrscheinlichkeiten von *Solidago gigantea* im unmittelbaren Bereich der Strukturen nimmt die Vorkommenswahrscheinlichkeit mit zunehmender Entfernung ab um schließlich wieder anzusteigen. Erwähnt seien noch die Variablen TAVE (Jahresmitteltemperatur) und TOPS (Topografische Exponiertheit) mit weitgehend steigender Regressionsgeraden. *Solidago gigantea* ist zudem in Gärten, Parks und Umschwungflächen (Kategorie 4 der Variablen AS97R) seltener, während sie auf extensiven landwirtschaftliche Flächen häufiger vorkommt (Kategorie 3).



Tabelle 29: Regressionskoeffizienten und Standardfehler der Variablen für das Modell  $\text{gesamt}^2$ , Variablenselektion mit BIC von *Solidago gigantea*.

Variable	Koeffizient	Standardfehler	
(Intercept)	-2.56E+01	1.27E+00	***
sqrt(GEB01)	3.71E-01	2.01E-02	***
I((sqrt(GEB01))^2)	-9.14E-03	6.62E-04	***
log(STR02 + 25)	4.16E-01	3.11E-02	***
log(STR03 + 25)	2.60E+00	3.54E-01	***
I((log(STR03 + 25))^2)	-2.65E-01	4.22E-02	***
I((log(STR04 + 25))^2)	3.34E-02	2.68E-03	***
sqrt(EIS01)	-6.45E-02	5.73E-03	***
I((sqrt(EIS01))^2)	1.12E-03	9.55E-05	***
log(GWN01 + 25)	2.42E+00	2.06E-01	***
I((log(GWN01 + 25))^2)	-2.60E-01	2.13E-02	***
sqrt(GWN02)	8.13E-02	5.77E-03	***
I((sqrt(GWN02))^2)	-1.01E-03	7.18E-05	***
sqrt(GWN03)	-1.11E-01	7.96E-03	***
I((sqrt(GWN03))^2)	1.54E-03	1.75E-04	***
TAVE	1.32E-02	1.28E-03	***
I((TAVE^6)^2)	-1.32E-35	9.32E-37	***
I(SRAD12^2)	1.17E-08	1.69E-09	***
ASPC	1.40E-01	3.52E-02	***
TOPS	1.05E-02	8.84E-04	***
I(TOPS^2)	-3.31E-05	8.28E-06	***
log(SLP + 1)	-1.63E-01	3.02E-02	***
SKELE1	-1.22E+00	1.99E-01	***
SKELE2	2.33E-01	1.69E-01	
SKELE3	-7.97E-01	1.61E-01	***
SKELE4	-2.83E-01	2.50E-01	
WSPEI1	3.16E-01	1.44E-01	*
WSPEI2	6.18E-01	1.46E-01	***
WSPEI3	NA	NA	
AS97R2 (Landwirtschaft, int.)	1.12E-01	6.99E-02	
AS97R3 (Landwirtschaft, ext.)	7.10E-01	7.31E-02	***
AS97R4 (Garten/Park/Umschwung)	-4.473e-01	1.09E-01	***
AS97R5 (Nass-Standort)	1.63E+00	1.03E-01	***
AS97R6 (sonstiges)	-1.89E-01	7.22E-02	**

Signifikanzniveau: \* = 0.05, \*\* = 0.01, \*\*\* = 0.001

Abbildung 33 und Abbildung 34 geben die Vorkommenswahrscheinlichkeit bzw. das prognostizierte Vorkommen von *Solidago gigantea* kartografisch wieder. Deutlich erkennbar ist das Fehlen der Art in den Stadtgebieten von Zürich und Winterthur sowie in den Gemeinden entlang des Zürichsees.

# Vorkommenswahrscheinlichkeit

## *Solidago gigantea*

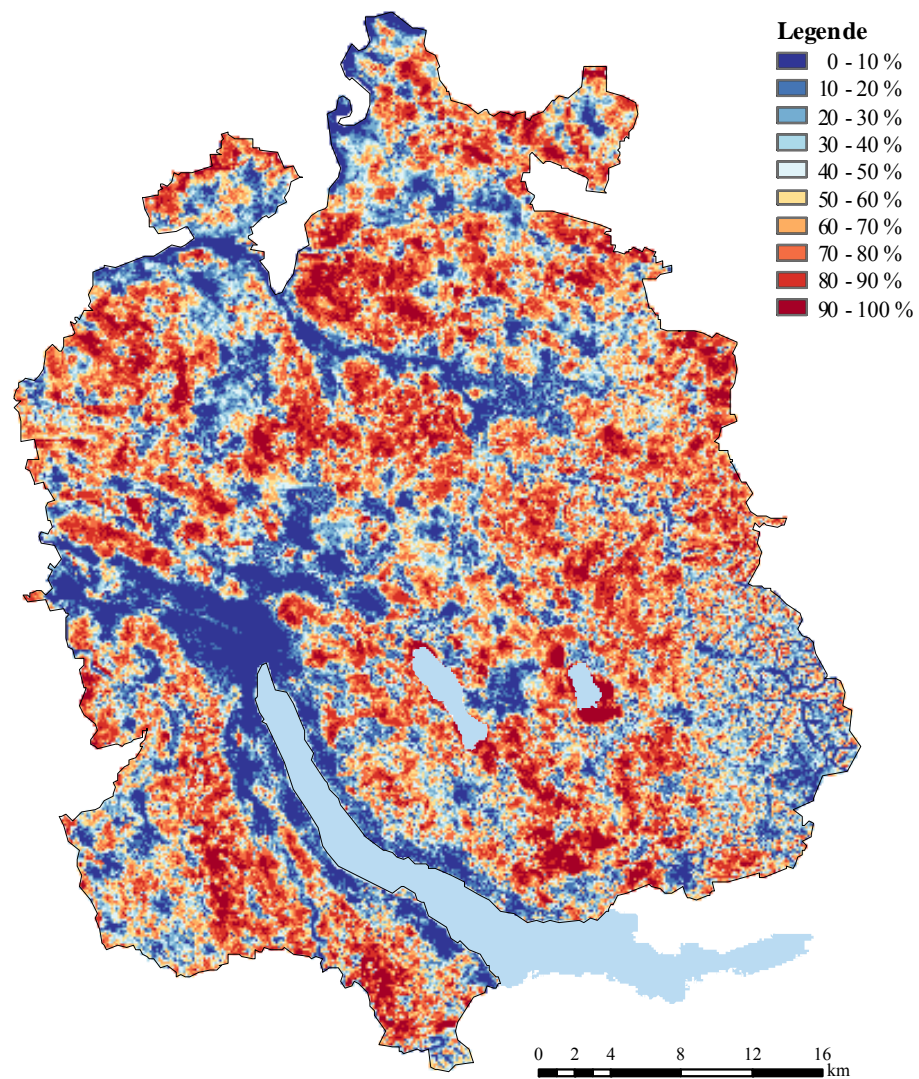


Abbildung 33: Vorkommenswahrscheinlichkeit von *Solidago gigantea* (Modell  $gesamt^2$ , Variablen-selektion mit BIC).

# prognostiziertes Vorkommen

## *Solidago gigantea*

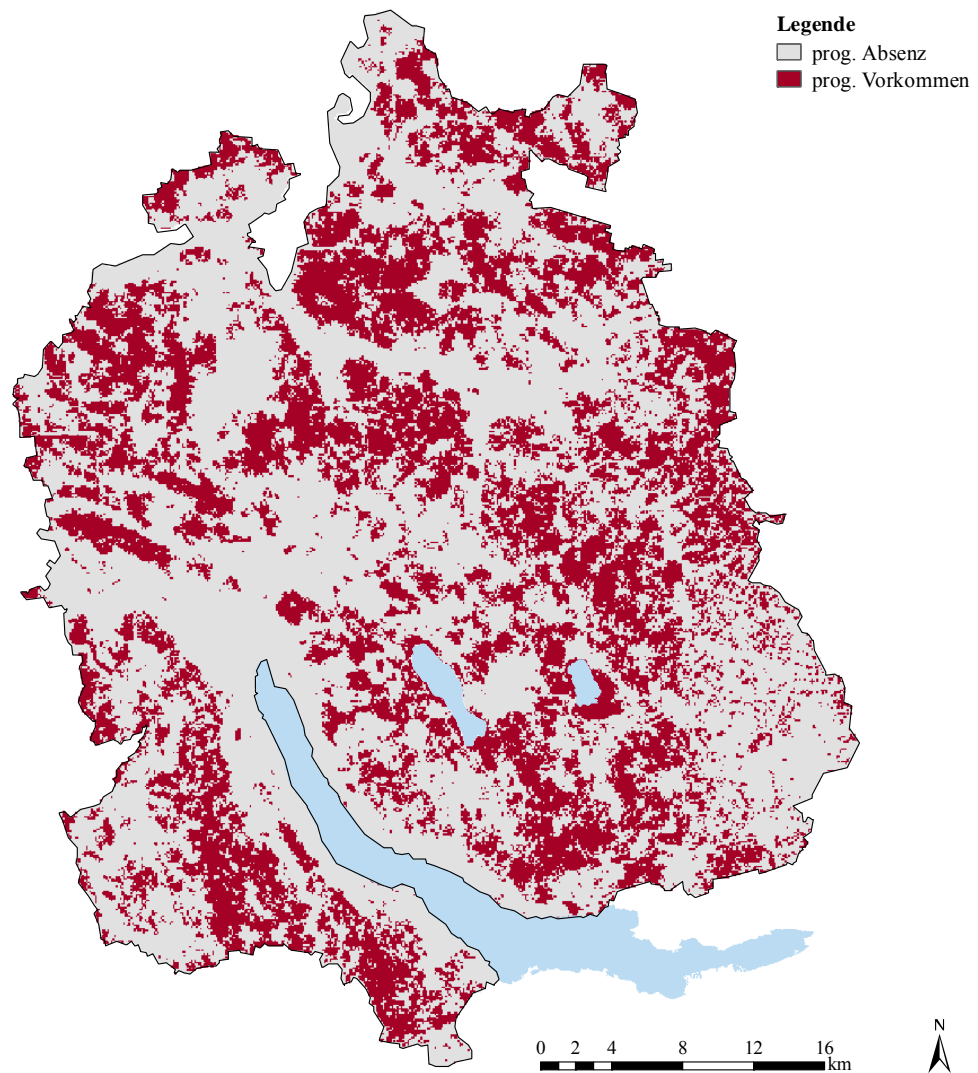


Abbildung 34: Prognostiziertes Vorkommen von *Solidago gigantea* entsprechend des Schwellenwertes (max. TSS; Modell  $\text{gesamt}^2$ , Variablen-Selektion mit BIC).

---

## V Zusammenfassung, Diskussion und Ausblick

Das Ziel der vorliegenden Arbeit war es, mit Hilfe geeigneter Habitatmodellierungstechniken das Vorkommen invasiver Neophyten im Kanton Zürich zu beschreiben. Zu diesem Zweck wurden zahlreiche Logistische Regressionen für sieben besonders relevante Neophyten durchgeführt.

Von allen verwendeten Arten konnten statistisch gut abgesicherte Modelle erstellt werden. Als eines von einer Vielzahl verwendeter Gütemaße sei hier der AUC-Wert genannt. Er liegt bei den ausgewählten Modellen zwischen 0.807 (*Solidago canadensis*) und 0.952 (*Rhus typhina*), was von einer jeweils hohen Modellgüte zeugt. Durchgeführte Kreuzvalidierungen zur Validierung der Modelle weisen ebenfalls auf eine gute Prognosegüte hin. Besonders gute Habitatmodelle konnten für *Ambrosia artemisiifolia* und *Rhus typhina* formuliert werden. Verbunden mit einer geringen Zahl erklärender Variablen ermöglichen sie sehr gute Prognosen des Vorkommens.

Für die Habitatmodellierung erwiesen sich die Distanzen zu bestimmten anthropogenen oder natürlichen Strukturen als besonders wichtige erklärende Variablen. Sie besitzen einen großen Erklärungsgehalt hinsichtlich des Vorkommens oder Fehlens der Arten. Über alle Arten hinweg waren das die Distanz-Variablen, die die Entfernung zu Gebäuden, großen Straßen (Autobahn, sonstige asphaltierte Straßen) oder Gewässern (Bäche oder Flüsse) beschreiben. Sie sind in fast jedem Modell enthalten und besitzen meist einen höchst signifikanten Einfluss auf das Vorkommen/Fehlen der Art. Bei allen Modellen mit Ausnahme von *Ambrosia artemisiifolia* und *Rhus typhina* ist auch die Entfernung zu Eisenbahngleisen im Modell enthalten. Als hilfreich hat sich die Reklassifizierung der Arealstatistik 97 in sechs Kategorien erwiesen. Diese Variable trägt häufig ebenfalls zu einem großen Teil zum Erklärungsgehalt der Modelle bei. Dahingegen spielen Klima- und Topografie-Variablen nur eine untergeordnete Rolle, was aber aufgrund der geringen räumlichen Ausdehnung des Untersuchungsgebietes plausibel ist. Eigenschaften des Bodens werden in Form des Skelettgehaltes und der Wasserspeicherkapazität gelegentlich in das Modell aufgenommen (z.B. beim Modell von *Impatiens glandulifera*).

Neben den in fast allen Modellen enthaltenen Variablen unterscheiden sich die einzelnen Art-Habitat-Modelle jedoch im Hinblick auf die Variablenanzahl und der enthaltenen Variablen. So wird beispielsweise die Potentielle Globalstrahlung im Dezember bei *Heracleum mantegazzianum* oder die Jahresmitteltemperatur bei *Reynoutria japonica* im Modell einbezogen, bei anderen Modellen tragen sie nicht zur Prognosegüte bei.

Wie bereits in der Einleitung formuliert, wird auf eine ökologische Interpretation der Ergebnisse in dieser Arbeit verzichtet. So muss beispielsweise offen bleiben, ob im Habitatmodell von *Rhus typhina* die Hangneigung einen ökologisch begründbaren Einfluss auf das Vorkommen der Art besitzt. Existieren möglicherweise an Hängen durch die Unterlassung bestimmter Nutzungsarten (Mahd) gute Etablierungsmöglichkeiten für den Essigbaum?

Abschätzungen über das zukünftige Ausbreitungspotential können ansatzweise lediglich für *Ambrosia artemisiifolia* gemacht werden. Nur bei dieser Art ist die Erfassung von Fundorten obligatorisch. Deshalb kann davon ausgegangen werden, dass ein großer Teil der aktuellen oder historischen Vorkommen erfasst wurde. Vergleicht man die Prognosekarte mit den bisherigen Funddaten, kann von einer weiteren Ausbreitung in den Stadtgebieten von Zürich und Winterthur inklusive angrenzender Ortschaften sowie den Gemeinden entlang des Zürichsees ausgegangen werden. Darüber hinaus wird ein vermehrtes Auftreten östlich des Zürichsees prognostiziert.

Im Fall von *Rhus typhina* ist klar der Ursprung der Ausbreitung auszumachen. Diese Art, ursprünglich als Ziergehölz gepflanzt, verwildert ausgehend von den Siedlungen. An diesem Beispiel werden auch die Grenzen der Habitatmodellierung sichtbar: vom aktuellen Vorkommen werden potentielle Habitate modelliert. Wenn eine Art jedoch in starker Ausbreitung begriffen ist, können Vorhersagen verfälscht werden (siehe beispielsweise auch die Ausführungen in GUIBAN & ZIMMERMANN 2000, GUIBAN & THUILLER 2005). Klare lineare Vorkommen und damit auch potentielle Ausbreitungswege können bei *Impatiens glandulifera* erkannt werden. Hier weisen Fluss- und Kanalufer hohe Vorkommenswahrscheinlichkeiten auf.

Auffallend ist das fast vollständige Fehlen von *Solidago gigantea* in den Stadtgebieten von Zürich und Winterthur. Diese unwahrscheinliche Prognose wird möglicherweise durch das Fehlen von Fundorten in den Stadtgebieten, verbunden mit der Zuweisung von Pseudo-Absenzen in diesen Gebieten verursacht. Deshalb soll in diesem Zusammenhang noch einmal auf die Voraussetzungen für derartige Modellierungen

hingewiesen werden. Im Idealfall geht einer derartigen Analyse eine Datenerhebung mit systematischem Probenahmedesign voraus. Standorte werden aufgesucht und das Vorkommen, aber auch das Fehlen einer Art wird dokumentiert. Diese für die Habitatmodellierung geeigneteren „echten“ Absenzen sind aufgrund der Konzeption des Web GIS bzw. der Datenbank der Baudirektion des Kantons Zürich jedoch auch in Zukunft nicht möglich. Die sich eigentlich als sinnvoll erwiesene Logik der Generierung der Pseudo-Absenzen könnte im Falle der Goldruten- aber auch der Knöterich-Arten zu Verzerrungen geführt haben. Die Ursachen liegen in einer Vielzahl von Vorkommen, die als nicht bestimmbar deklariert wurden. Sie wurden aus der Analyse der jeweiligen Arten-Gruppen ausgeschlossen. So wäre es zukünftig durchaus vorstellbar, diese Arten-Gruppen gemeinsam zu modellieren. Das würde den Fehler des unerkannten Vorkommens und den Ausschluss aus dem Modellierungsprozess verhindern. Dies wäre jedoch mit einer verringerten Flexibilität der Modellierung der unterschiedlichen ökologischen Ansprüche der Arten verbunden. Interessant wäre eine Validierung der Modelle in räumlicher Hinsicht, indem die Güte der Modelle beispielsweise anhand von Daten anderer Kantone überprüft wird.

Abschließend kann festgestellt werden, dass diese Arbeit für einige besonders relevante invasive Neophyten statistisch gut abgesicherte Modelle bereitstellt, verbunden mit entsprechenden Karten zur Vorkommenswahrscheinlichkeit und prognostiziertem Vorkommen. Die Daten geben interessierten Nutzergruppen die Möglichkeit, durch die hier (statistisch) festgestellten relevanten Umweltfaktoren für das Vorkommen der Arten neue Erkenntnisse zu gewinnen. Dabei ist es aber ausdrücklich erwünscht, die vorliegenden Daten kritisch ökologisch zu hinterfragen. Des Weiteren kann das vorliegende Kartenmaterial dazu dienen, die Verbreitung der Neophyten in Zukunft noch besser zu überwachen. Gebiete mit hohen Vorkommenswahrscheinlichkeiten bestimmter Arten können so z.B. gezielt begangen und Bekämpfungsmaßnahmen gegebenenfalls durchgeführt werden.

---

## VI Literatur

- ALLOUCHE, O.; TSOAR, A. & KADMON, R. (2006): Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43: 1223-1232.
- ANONYMOS. (2008): Neophytenkataster: Punkt und Flächenerfassung, Version: 10.01.2008 (ed. BAUDIREKTION KANTON ZÜRICH; AMT FÜR RAUMORDNUNG UND VERMESSUNG; ABTEILUNG GIS-ZENTRUM).
- ARAUJO, M. B. & LUOTO, M. (2007): The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology And Biogeography* 16(6): 743-753.
- AUSTIN, M. P. (2007): Species distribution models and ecological theory: A critical assessment and some possible new approaches. *Ecological Modelling* 200(1-2): 1-19.
- AWEL [Amt für Abfall, Wasser, Energie und Luft des Kantons Zürich] (2005): WEB GIS gestützte Erfassung von Neophyten. URL: <http://www.biosicherheit.zh.ch/internet/bd/awel/awb/bs/de/neophyten/gisErfassung.html>, Letzter Zugriff: 30.03.2009
- BÄBLER, M.; JÄGER, E. J. & WERNER, K. (1996): Rothmaler, Exkursionsflora von Deutschland, Gefäßpflanzen: Grundband. Gustav Fischer Verlag, Jena.
- BAUDIREKTION KANTON ZÜRICH; AMT FÜR RAUMORDNUNG UND VERMESSUNG, A. G.-Z. (2009): Neophytenkataster, Amt für Abfall, Wasser, Energie und Luft (AWEL), Abteilung Abfallwirtschaft und Betriebe (AW), Sektion Biosicherheit (ed. W. AMT FÜR ABFALL, ENERGIE UND LUFT (AWEL), ABTEILUNG ABFALLWIRTSCHAFT UND BETRIEBE (AW), SEKTION BIOSICHERHEIT), Zürich.
- BEERS, T. W.; DRESS, P. E. & WENSEL, L. C. (1966): Aspect transformation in site productivity research. *Journal of Forestry* 64(10): 691-692.
- BEYER, H. L. (2006): Hawth's Analysis Tools for ArcGIS. URL: <http://www.spatial ecology.com/htools/index.php>, Letzter Zugriff: 29.07.2009
- BFL (1994): Digital elevation model, DHM25, Bundesamt für Landestopografie. Wabern.
- BFL (2007a): Digital landscape model, VECTOR25, Bundesamt für Landestopografie. Wabern.
- BFL (2007b): VECTOR25: Das digitale Landschaftsmodell der Schweiz. Bundesamt für Landestopografie. <http://www.swisstopo.admin.ch/internet/swisstopo/de/home/products/landscape/vector25.parsysrelated1.47641.downloadList.50692.DownloadFile.tmp/vector25infode.pdf>, Letzter Zugriff: 10.03.2010
- BFS (1979/85): Arealstatistik. Bundesamt für Statistik. Neuchâtel.
- BFS (1992/97): Arealstatistik. Bundesamt für Statistik. Neuchâtel.

- BFS (2001): Generalisierte Gemeindegrenzen der Schweiz, Kantone, Bundesamt für Statistik. Neuchâtel.
- BFS (2007a): Arealstatistik nach Nomenklatur 1992: GEOSTAT-Datenbeschreibung. Bundesamt für Statistik. [http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/arealstatistik\\_1979.Document.97374.pdf](http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/arealstatistik_1979.Document.97374.pdf), Letzter Zugriff: 10.03.2010
- BFS (2007b): Generalisierte Gemeindegrenzen der Schweiz: GEOSTAT-Datenbeschreibung, Version 1.3, Bundesamt für Statistik. [http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/generalisierte\\_gemeindegrenzen.Document.97376.pdf](http://www.bfs.admin.ch/bfs/portal/de/index/dienstleistungen/geostat/datenbeschreibung/generalisierte_gemeindegrenzen.Document.97376.pdf), Letzter Zugriff: 31.01.2010
- BFS (2010): Arealstatistik 2004/09 (NOAS04): Steckbrief. Bundesamt für Statistik. <http://www.bfs.admin.ch/bfs/portal/de/index/infothek/nomenklaturen/blank/blank/noas04/01.html>, Letzter Zugriff: 12.03.2010
- BLUM, W. E. H. (2007): Bodenkunde in Stichworten. 6. völlig neu bearbeitete Auflage, Gebr. Borntraeger Berlagsbuchgesellschaft, Berlin.
- BOHREN, C.; MERMILLOD, G. & DELABAYS, N. (2006): Common ragweed (*Ambrosia artemisiifolia* L.) in Switzerland: development of a nationwide concerted action. *Journal Of Plant Diseases And Protection*: 497-503.
- BREWER, C. & HARROWER, M. (2010): Colorbrewer 2.0: color advice for cartography. <http://colorbrewer2.org/>, Letzter Zugriff: 09.04.2010
- CHEFAOUI, R. M. & LOBO, J. M. (2008): Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecological Modelling* 210(4): 478-486.
- CHONGSUVIVATWONG, V. (2010): Package 'epicalc': Epidemiological calculator. <http://cran.r-project.org/web/packages/epicalc/epicalc.pdf>, Letzter Zugriff: 11.04.2010
- CHYTRY, M.; PYSEK, P.; WILD, J.; PINO, J.; MASKELL, L. C. & VILA, M. (2009): European map of alien plant invasions based on the quantitative assessment across habitats. *Diversity and Distributions* 15(1): 98-107.
- DE FRUTOS, A.; OLEA, P. P. & VERA, R. (2007): Analyzing and modelling spatial distribution of summering lesser kestrel: The role of spatial autocorrelation. *Ecological Modelling* 200(1-2): 33-44.
- DORMANN, C. F.; PURSCHKE, O.; MARQUEZ, J. R. G.; LAUTENBACH, S. & SCHRODER, B. (2008): Components of uncertainty in species distribution analysis: A case study of the great grey shrike. *Ecology* 89(12): 3371-3386.
- DUDENREDAKTION (2001): Duden: Die deutsche Rechtschreibung, Band 1. Bibliographisches Institut & F.A. Brockhaus AG, Mannheim.



- ELITH, J.; GRAHAM, C. H.; ANDERSON, R. P.; DUDÍK, M.; FERRIER, S.; GUISAN, A.; HIJMANS, R. J.; HUETTMANN, F.; LEATHWICK, J. R.; LEHMANN, A.; LI, J.; LOHMANN, L. G.; LOISELLE, B. A.; MANION, G.; MORITZ, C.; NAKAMURA, M.; NAKAZAWA, Y.; OVERTON, J. M. M.; PETERSON, A. T.; PHILLIPS, S. J.; RICHARDSON, K.; SCACHETTI-PEREIRA, R.; SCHAPIRE, R. E.; SOBERÓN, J.; WILLIAMS, S.; WISZ, M. S. & ZIMMERMANN, N. E. (2006): Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2): 129-151.
- ELITH, J. & LEATHWICK, J. R. (2007): Predicting species' distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions* 13: 165-175.
- ESRI (o.J.): ArcView, Create, View and Edit Your Data: Overview. <http://www.esri.com/software/arcgis/arcview/index.html>, Letzter Zugriff: 08.03.2010
- FALINSKI, J. B. (1998): Invasive alien plants and vegetation dynamics. *Plant Invasions: Ecological Mechanisms And Human Responses*: 3-21.
- FARAWAY, J. J. (2006): Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. Chapman & Hall/CRC.
- FAWCETT, T. (2004): ROC Graphs: Notes and Practical Considerations for Researchers. *HP Labs Technical Report HPL-2003-4*.
- FAWCETT, T. (2006): An introduction to ROC analysis. *Pattern Recognition Letters* 27: 861-874.
- FERRIER, S.; WATSON, G.; PEARCE, J. & DRIELSMA, M. (2002): Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. I. Species-level modelling. *Biodiversity and Conservation*, 11: 2275-2307.
- FIELDING, A. H. (2002): What Are the Appropriate Characteristics of an Accuracy Measure? In: J. M. SCOTT, P. HEGLUND, M. L. MORRISON, J. B. HAUFLER, M. G. RAPHAEL, W. A. WALL & F. B. SAMSON (Hrsg.): Predicting Species Occurrences: Issues of Accuracy and Scale. Island Press, Washington: 271-280.
- FIELDING, A. H. & BELL, J. F. (1997): A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24(1): 38-49.
- FIELDING, A. H. & HAWORTH, P. F. (1995): Testing the generality of bird-habitat models. *Conservation Biology* 9(6): 1466-1481.
- FOX, J. (2009): Package 'car': Companion to Applied Regression. <http://cran.r-project.org/web/packages/car/car.pdf>, Letzter Zugriff: 04.04.2010
- GELPKKE, G. & WEBER, E. (2005): Situation und Handlungsbedarf bezüglich invasiver Neophyten im Kanton Zürich. Studie im Auftrag der Sektion Biosicherheit (SBS), Amt für Abfall, Wasser, Energie und Luft (AWEL), Baudirektion Kanton Zürich.

- GIGON, A. & WEBER, E. (2005): Invasive Neophyten in der Schweiz: Lagebericht und Handlungsbedarf. Bericht zu Händen des Bundesamtes für Umwelt, Wald und Landschaft, Bern.
- GIS-ZENTRUM KANTON ZÜRICH (2006): Neophytenkataster: Punkt und Detailerfassung. <http://www.gis.zh.ch/dokus/awel/abfallwirtschaft/neophyten.html>, Letzter Zugriff: 03.03.2010
- GRAHAM, M. H. (2003): Confronting multicollinearity in ecological multiple regression. *Ecology* 84(11): 2809-2815.
- GRAUTE, S. (2009): Einfluss der Klimaveränderung auf die Biodiversität der Tagfalter im Alpenraum. Master, Zentrum für Geoinformatik (Z\_GIS) der Paris-Lodron Universität Salzburg.
- GREEN, R. H. (1979): Sampling design and statistical methods for environmental biologists. Wiley, New York.
- GUISAN, A.; EDWARDS, T. C. & HASTIE, T. (2002): Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling* 157(2-3): 89-100.
- GUISAN, A. & THUILLER, W. (2005): Predicting species distribution: offering more than simple habitat models. *Ecology Letters* 8: 993-1009.
- GUISAN, A. & ZIMMERMANN, N. E. (2000): Predictive habitat distribution models in ecology. *Ecological Modelling* 135(2-3): 147-186.
- HANLEY, J. A. & MCNEIL, B. J. (1982): The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve. *Radiology* 143(1): 29-36.
- HARRELL, F. E.; LEE, K. L. & MARK, D. B. (1996): Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* (15): 361-387.
- HARTMANN, E.; SCHULDES, H.; KÜBLER, R. & KONOLD, W. (1995): Neophyten: Biologie, Verbreitung und Kontrolle ausgewählter Arten. ecomed, Landsberg.
- HASTIE, T. J. & TIBSHIRANI, R. J. (1990): Generalized Additive Models. Chapman & Hall, London.
- HERRMANN, D. (2009): Modellierung der Auswirkungen des Klimawandels auf die landlebenden europäischen Säugetiere. Master, Zentrum für GeoInformatik (Z\_GIS) der Paris Lodron-Universität Salzburg.
- HILBE, J. M. (2009): Logistic Regression Models. Chapman & Hall/CRC, Boca Raton.
- HIRZEL, A. H.; HELFER, V. & METRAL, F. (2001): Assessing habitat-suitability models with a virtual species. *Ecological Modelling* 145: 111-121.
- HIRZEL, A. H.; LE LAY, G.; HELFER, V.; RANDIN, C. & GUISAN, A. (2006): Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling* 199(2): 142-152.

- HULME, P. E. (2006): Beyond control: wider implications for the management of biological invasions. *Journal of Applied Ecology* 43(5): 835-847.
- HULME, P. E. (2009): Trade, transport and trouble: managing invasive species pathways in an era of globalization. *Journal of Applied Ecology* 46(1): 10-18.
- HUTCHINSON, G. E. (1959): Homage to Santa Rosalia or Why Are There So Many Kinds of Animals? *American Naturalist* 93(870): 145-159.
- KLAR, N.; FERNÁNDEZ, N.; KRAMER-SCHADT, S.; HERRMANN, M.; TRINZEN, M.; BÜTTNER, I. & NIEMITZ, C. (2008): Habitat selection models for European wildcat conservation. *Biological Conservation* 141: 308-319.
- KLEYER, M.; KRATZ, R.; LUTZE, G. & SCHRÖDER, B. (1999/2000): Habitatmodelle für Tierarten: Entwicklung, Methoden und Perspektiven für die Anwendung. *Zeitschrift für Ökologie und Naturschutz* 8: 177-194.
- KOWARIK, I. (2003): Biologische Invasionen: Neophyten und Neozoen in Mitteleuropa. Eugen Ulmer, Stuttgart.
- KUTNER, M. H.; NETER, J.; NACHTSHEIM, C. J. & LI, W. (2004): Applied Linear Statistical Models. 5. Auflage, McGraw-Hill/Irwin, New York.
- LANG, S. & BLASCHKE, T. (2007): Landschaftsanalyse mit GIS. Eugen Ulmer, Stuttgart (Hohenheim).
- LANTZ, C. A. & NEBENZAHL, E. (1996): Behavior and interpretation of the kappa statistic: Resolution of the two paradoxes. *Journal Of Clinical Epidemiology* 49(4): 431-434.
- LASKO, T. A.; BHAGWAT, J. G.; ZOU, K. H. & OHNO-MACHADO, L. (2005): The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics* 38(5): 404-415.
- LIU, C. R.; BERRY, P. M.; DAWSON, T. P. & PEARSON, R. G. (2005): Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28(3): 385-393.
- MAC NALLY, R. (2000): Regression and model-building in conservation biology, biogeography and ecology: The distinction between and reconciliation of 'predictive' and 'explanatory' models. *Biodiversity And Conservation* 9(5): 655-671.
- MACK, R. N.; SIMBERLOFF, D.; LONSDALE, W. M.; EVANS, H.; CLOUT, M. & BAZZAZ, F. A. (2000): Biotic invasions: Causes, epidemiology, global consequences, and control. *Ecological Applications* 10(3): 689-710.
- MAGGINI, R.; LEHMANN, A.; ZIMMERMANN, N. E. & GUISAN, A. (2006): Improving generalized regression analysis for the spatial prediction of forest communities. *Journal of Biogeography*.

- MAINDONALD, J. & BRAUN, W. J. (2007): Data analysis and graphics using R: an example-based approach. R. GILL, B. D. RIPLEY, S. ROSS, B. W. SILVERMAN & M. STEIN (Hrsg.): Cambridge Series in Statistical and Probabilistic Mathematics, 2. Auflage Cambridge University Press, Cambridge: 502.
- MAINDONALD, J. & BRAUN, W. J. (2009): Package 'DAAG': Data Analysis And Graphics data and functions. <http://cran.r-project.org/web/packages/DAAG/DAAG.pdf>, Letzter Zugriff: 04.04.2010
- MANEL, S.; WILLIAMS, H. C. & ORMEROD, S. J. (2001): Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38(5): 921-931.
- MASKELL, L. C.; BULLOCK, J. M.; SMART, S. M.; THOMPSON, K. & HULME, P. E. (2006): The distribution and habitat associations of non-native plant species in urban riparian habitats. *Journal Of Vegetation Science* 17(4): 499-508.
- MCPHERSON, J. M.; JETZ, W. & ROGERS, D. J. (2004): The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact? *Journal of Applied Ecology* 41(5): 811-823.
- MOISEN, G. G. & FRESCINO, T. S. (2002): Comparing five modelling techniques for predicting forest characteristics. *Ecological Modelling* 157(2-3): 209-225.
- MONSERUD, R. A. & LEEMANS, R. (1992): Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling* 62(4): 275-293.
- MOSER, D. M.; GYGAX, A.; B., B.; WYLER, N. & PALESE, R. (2002): Rote Liste der gefährdeten Farn- und Blütenpflanzen der Schweiz. BUNDESAMT FÜR UMWELT, WALD UND LANDSCHAFT, BERN; ZENTRUM DES DATENVERBUNDNETZES DER SCHWEIZER FLORA, CHAMBÉSY; CONSERVATOIRE ET JARDIN BOTANIQUES DE LA VILLE DE GENÈVE, CHAMBÉSY, (Hrsg.): BUWAL-Reihe „Vollzug Umwelt“, Bern: 118.
- NIEDERBERGER, B. (2008): Habitatmodellierung der endemischen Schneckenart *Trochulus biconicus*. Master, Zentrum für GeoInformatik (Z\_GIS) der Paris Lodron-Universität Salzburg.
- NOBIS, M. P.; JAEGER, J. A. G. & ZIMMERMANN, N. E. (2009): Neophyte species richness at the landscape scale under urban sprawl and climate warming. *Diversity And Distributions* 15(6): 928-939.
- PEARCE, J. L. & BOYCE, M. S. (2006): Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* 43(3): 405-412.
- PEETERS, E. & GARDENIERS, J. J. P. (1998): Logistic regression as a tool for defining habitat requirements of two common gammarids. *Freshwater Biology* 39(4): 605-615.
- PYSEK, P.; JAROSIK, V. & KUCERA, T. (2002): Patterns of invasion in temperate nature reserves. *Biological Conservation* 104(1): 13-24.

- PYTHON SOFTWARE FOUNDATION (1990-2010): Python 2.5.1. <http://www.python.org/download/releases/2.5.1/>, Letzter Zugriff: 08.03.2010
- QUINN, G. P. & KEOUGH, M. J. (2002): *Experimental Design and Data Analysis for Biologists*. Cambridge University Press, Cambridge.
- REINEKING, B. & SCHRÖDER, B. (2004a): Gütemaße für Habitatmodelle. In: C. F. DORMANN, T. BLASCHKE, A. LAUSCH, B. SCHRÖDER & D. SÖNDGERATH (Hrsg.): *Habitatmodelle - Methodik, Anwendung, Nutzen. Teil I: Statistische Habitatmodelle - Status quo & aktuelle Entwicklungen Band 9/2004, Tagungsband zum Workshop vom 5.-10. Oktober 2003 am UFZ Leipzig, Leipzig: 27-38.*
- REINEKING, B. & SCHRÖDER, B. (2004b): Variablenselektion. In: C. F. DORMANN, T. BLASCHKE, A. LAUSCH, B. SCHRÖDER & D. SÖNDGERATH (Hrsg.): *Habitatmodelle - Methodik, Anwendung, Nutzen. Teil I: Statistische Habitatmodelle - Status quo & aktuelle Entwicklungen Band 9/2004, Tagungsband zum Workshop vom 5.-10. Oktober 2003 am UFZ Leipzig, Leipzig: 39-45.*
- REINEKING, B. & SCHRÖDER, B. (2006): Constrain to perform: Regularization of habitat models. *Ecological Modelling* 193(3-4): 675-690.
- RICHARDS, S. A. (2005): Testing ecological theory using the information-theoretic approach: Examples and cautionary results. *Ecology* 86(10): 2805-2814.
- ROBERTS, J. J.; BEST, B. D.; DUNN, D. C.; TREML, E. A. & HALPIN, P. N. (in review): Marine Geospatial Ecology Tools: An integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. *Environmental Modelling & Software*.
- SCHRÖDER, B. (2000): *Zwischen Naturschutz und theoretischer Ökologie: Modelle zur Habitateignung und räumlichen Populationsdynamik für Heuschrecken im Niedermoor*. Dissertation, Institut für Geographie und Geoökologie der Technische Universität Braunschweig.
- SCHRÖDER, B. (2004): Überblick zum Block "Statistische Habitatmodelle - Status quo & aktuelle Entwicklungen". In: C. F. DORMANN, T. BLASCHKE, A. LAUSCH, B. SCHRÖDER & D. SÖNDGERATH (Hrsg.): *Habitatmodelle - Methodik, Anwendung, Nutzen. Teil I: Statistische Habitatmodelle - Status quo & aktuelle Entwicklungen Band 9/2004, Tagungsband zum Workshop vom 5.-10. Oktober 2003 am UFZ Leipzig, Leipzig: 3-4.*
- SCHRÖDER, B. & REINEKING, B. (2004a): Modellierung der Art-Habitat-Beziehung - ein Überblick über die Verfahren der Habitatmodellierung. In: C. F. DORMANN, T. BLASCHKE, A. LAUSCH, B. SCHRÖDER & D. SÖNDGERATH (Hrsg.): *Habitatmodelle - Methodik, Anwendung, Nutzen. Teil I: Statistische Habitatmodelle - Status quo & aktuelle Entwicklungen Band 9/2004, Tagungsband zum Workshop vom 5.-10. Oktober 2003 am UFZ Leipzig, Leipzig: 5-25.*

- SCHRÖDER, B. & REINEKING, B. (2004b): Validierung von Habitatmodellen. In: C. F. DORMANN, T. BLASCHKE, A. LAUSCH, B. SCHRÖDER & D. SÖNDGERATH (Hrsg.): Habitatmodelle - Methodik, Anwendung, Nutzen. Teil I: Statistische Habitatmodelle - Status quo & aktuelle Entwicklungen Band 9/2004, Tagungsband zum Workshop vom 5.-10. Oktober 2003 am UFZ Leipzig, Leipzig: 47-56.
- SEGURADO, P. & ARAUJO, M. B. (2004): An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31(10): 1555-1568.
- SHEATHER, S. J. (2009): A Modern Approach to Regression with R. G. CASELLA, S. FIENBERG & I. OLKIN (Hrsg.): Springer Texts in Statistics, 393, 1. Auflage Springer, Berlin.
- SOKAL, R. R. & ROHLF, F. J. (1995): Biometry: The principles and practice of statistics in biological research. 3. Auflage, W. H. Freeman and Company, New York.
- STAGE, A. R. (1976): An Expression for the Effect of Aspect, Slope, and Habitat Type on Tree Growth. *Forest Science* 22(4): 457-460.
- STEYERBERG, E. W.; HARRELL, F. E.; BORSBOOM, G.; EIJKEMANS, M. J. C.; VERGOUWE, Y. & HABBEMA, J. D. F. (2001): Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology* 54(8): 774-781.
- STOCKWELL, D. R. B. & PETERS, D. (1999): The GARP modeling system: problems and solutions to automated spatial prediction. *International Journal of Geographic Information Science* 13(2): 143-158.
- STOCKWELL, D. R. B. & PETERSON, A. T. (2002): Controlling bias in biodiversity data. Predicting Species Occurrences: Issues of Accuracy and Scale. In: J. M. SCOTT, P. J. HEGLUND, M. L. MORRISON, J. B. HAUFLER, M. G. RAPHAEL, W. A. WALL & F. B. SAMSON (Hrsg.): Predicting species occurrence: Issues of accuracy and scale. Island Press, Washington, DC: 537-546.
- THE R DEVELOPMENT CORE TEAM (2008): R: A Language and Environment for Statistical Computing, Reference Index, Version 2.8.1 (2008-12-22). R FOUNDATION FOR STATISTICAL COMPUTING (Hrsg.), Wien.
- TOWNSEND, G. D.; HARPER, J. L. & BEGON, M. E. (2003): Ökologie. Springer, Berlin.
- TRIMBLE, G. R. & WEITZMAN, S. (1956): Site Index Studies of Upland Oaks in the Northern Appalachians. *Forest Science* 2(3): 162-173.
- VITOUSEK, P. M.; DANTONIO, C. M.; LOOPE, L. L.; REJMANEK, M. & WESTBROOKS, R. (1997): Introduced species: A significant component of human-caused global change. *New Zealand Journal of Ecology* 21(1): 1-16.
- WILLIAMSON, M.; DEHNEN-SCHMUTZ, K.; KUHN, I.; HILL, M.; KLOTZ, S.; MILBAU, A.; STOUT, J. & PYSEK, P. (2009): The distribution of range sizes of native and alien plants in four European countries and the effects of residence time. *Diversity And Distributions* 15(1): 158-166.

- ZANIEWSKI, A. E.; LEHMANN, A. & OVERTON, J. M. (2002): Predicting species spatial distributions using presence-only data: a case study of native New Zealand ferns. *Ecological Modelling* 157: 261-280.
- ZARNETSKE, P. L.; EDWARDS, T. C., JR. & MOISEN, G. G. (2007): Modeling forest bird species' likelihood of occurrence in Utah with Forest Inventory and Analysis and Landfire map products and ecologically based pseudo-absence points. In: R. E. MCROBERTS, G. A. REAMS, P. C. VAN DEUSEN & W. H. MCWILLIAMS (Hrsg.): Proceedings of the seventh annual forest inventory and analysis symposium; October 3-6, 2005; Portland, ME. Gen. Tech. Rep. WO-77. Washington, DC: U.S. Department of Agriculture, Forest Service: 291-305.
- ZIMMERMANN, N. E. (2000a): 1.1 shortwave.aml. [http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml1\\_1.html](http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml1_1.html), Letzter Zugriff: 17.04.2010
- ZIMMERMANN, N. E. (2000b): 2.1 tave.aml. [http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml2\\_1.html](http://www.wsl.ch/staff/niklaus.zimmermann/programs/aml2_1.html), Letzter Zugriff: 17.04.2010
- ZIMMERMANN, N. E. & KIENAST, F. (1999): Predictive mapping of alpine grasslands in Switzerland: species versus community approach. *Journal of Vegetation Science* 10(4): 469-482.
- ZIMMERMANN, N. E. & ROBERTS, D. W. (2001): Final report of the MLP climate and biophysical mapping project. *Scientific project report, WSL*: 1-18.

---

## VII Anhang

### 1 Übersichtskarte des Untersuchungsgebietes

#### Kanton Zürich

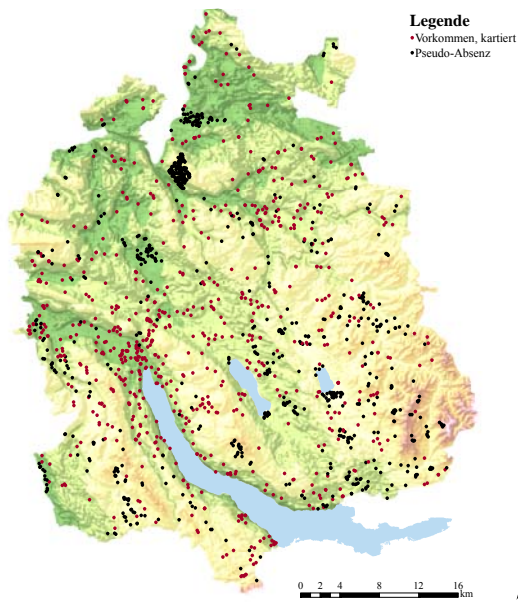


Abbildung 35: Übersichtskarte des Untersuchungsgebietes. Siedlungen (Gebäude) sind grau dargestellt, Größere Städte und Flüsse beschriftet, Datengrundlage: BFL (1994).

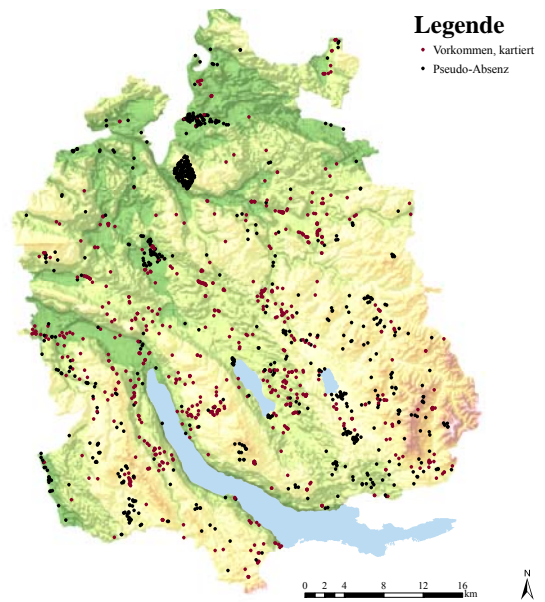


## 2 Verbreitungskarten (Vorkommen/Pseudo-Absenz)

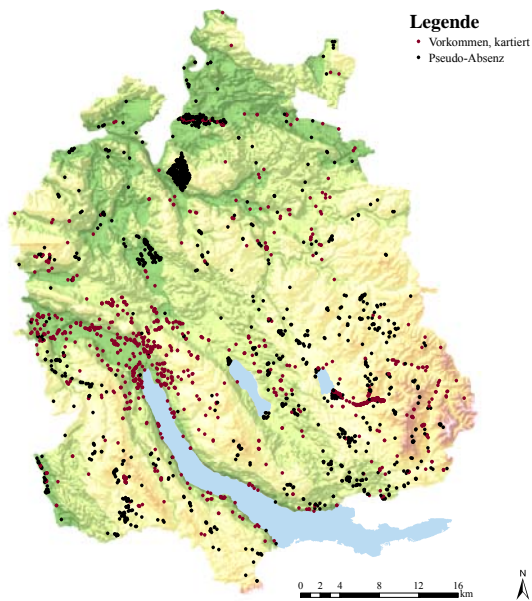
kartiertes Vorkommen und Pseudo-Absenz  
*Ambrosia artemisiifolia*



kartiertes Vorkommen und Pseudo-Absenz  
*Heracleum mantegazzianum*



kartiertes Vorkommen und Pseudo-Absenz  
*Reynoutria japonica*



kartiertes Vorkommen und Pseudo-Absenz  
*Impatiens glandulifera*

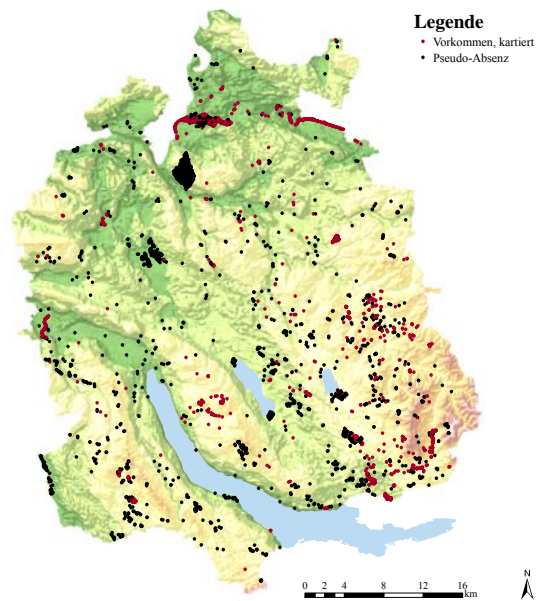
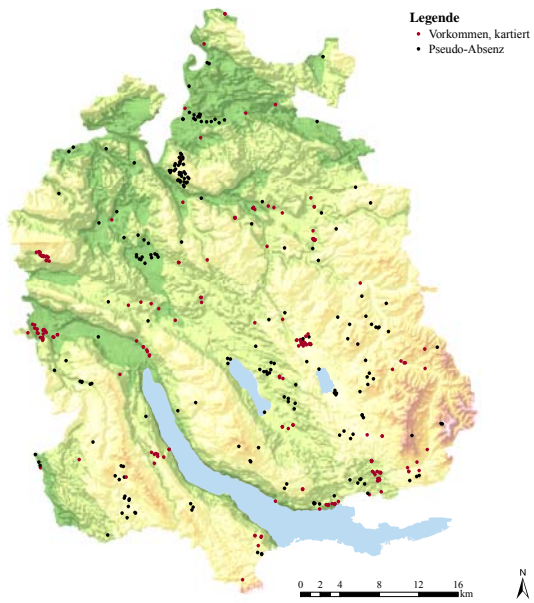
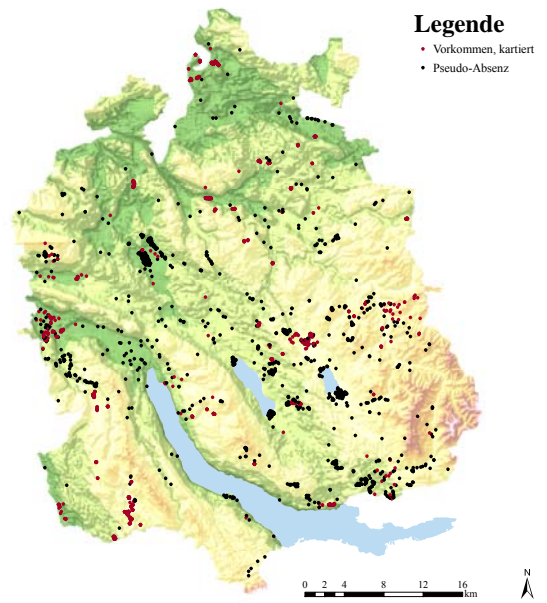


Abbildung 36: Verbreitungskarten der modellierten Neophyten I, Datengrundlage: BFL (1994).

kartiertes Vorkommen und Pseudo-Absenz  
*Rhus typhina*



kartiertes Vorkommen und Pseudo-Absenz  
*Solidago canadensis*



kartiertes Vorkommen und Pseudo-Absenz  
*Solidago gigantea*

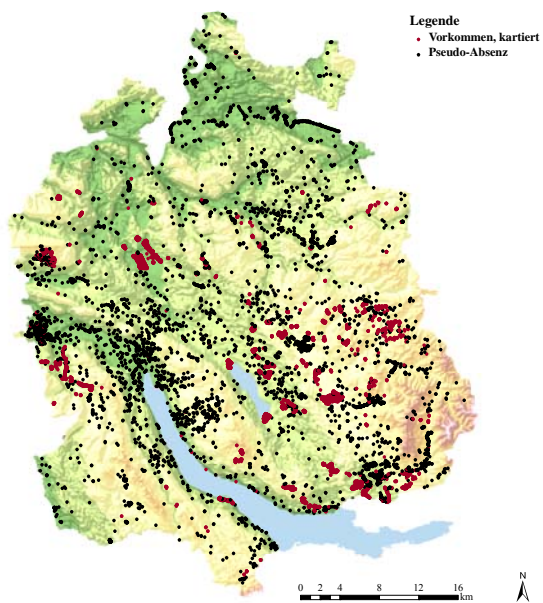


Abbildung 37: Verbreitungskarten der modellierten Neophyten II, Datengrundlage: BFL (1994).

### 3 Variablen-Wertebereiche

#### 3.1 Klima-Variablen

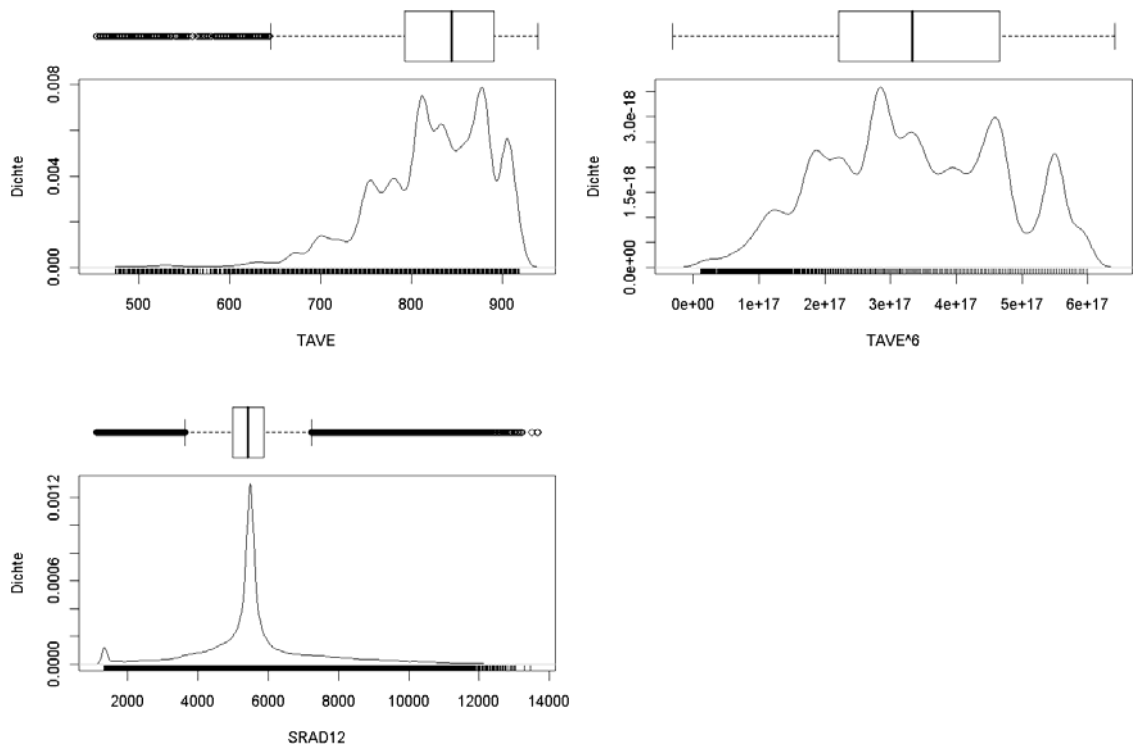


Abbildung 38: Werteverteilung der Klima-Variablen vor (links) und nach einer eventuell vorgenommenen Transformation.

### 3.2 Topografische Variablen

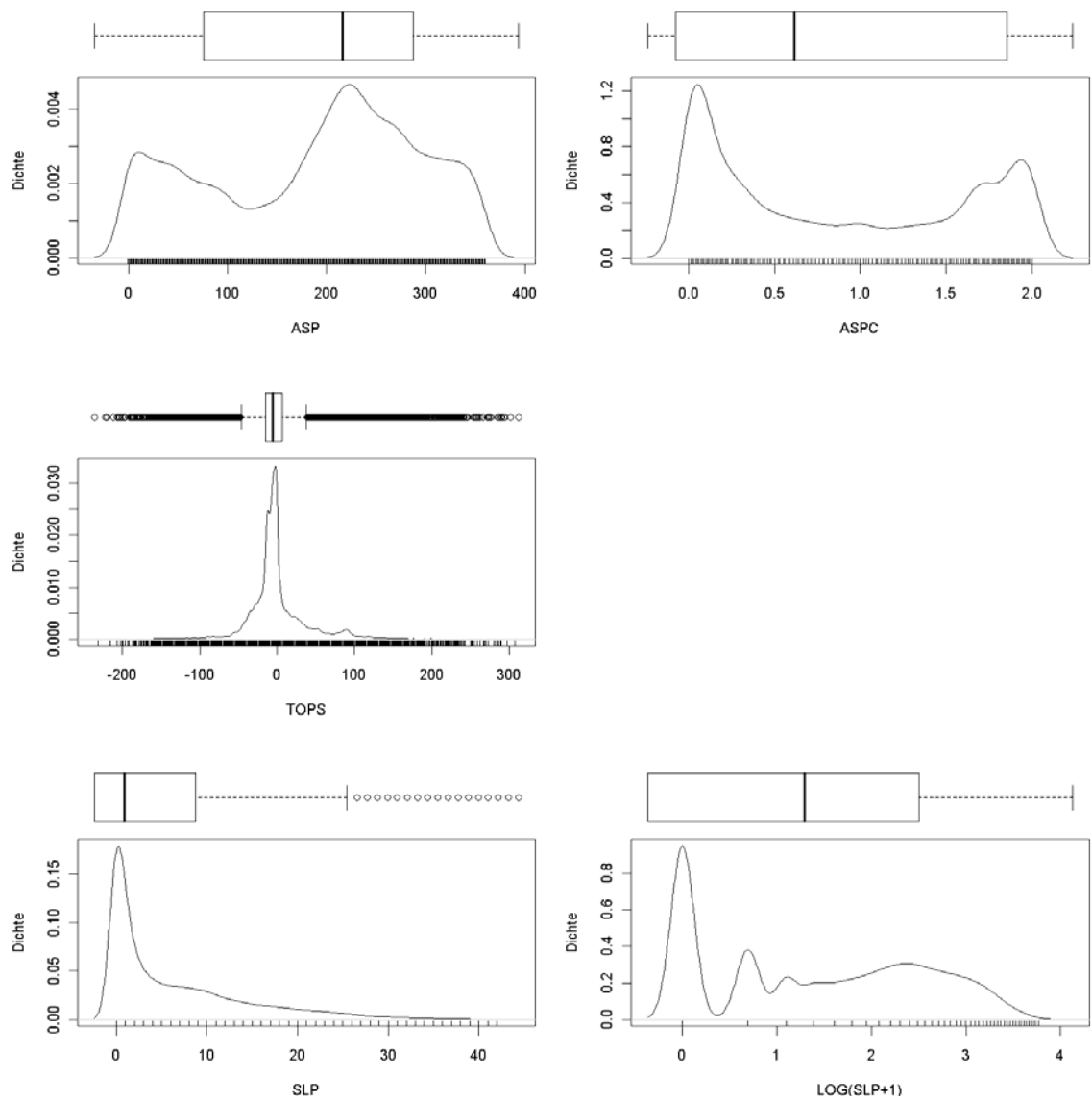


Abbildung 39: Werteverteilung der Topografischen Variablen vor (links) und nach einer eventuell vorgenommenen Transformation.

### 3.3 Physikalische Boden-Variablen

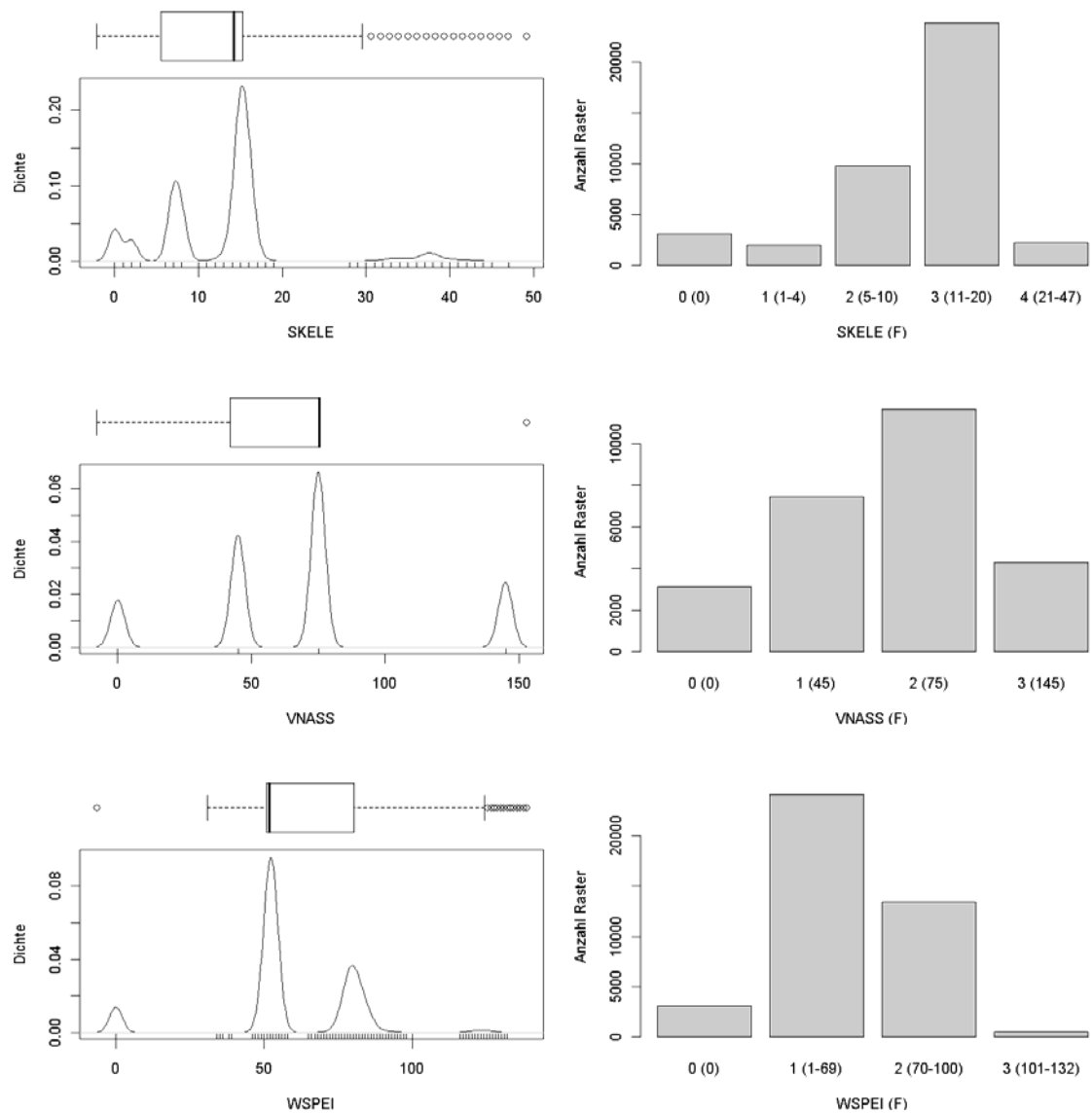


Abbildung 40: Werteverteilung der Physikalischen Boden-Variablen vor (links) und nach einer eventuell vorgenommenen Transformation.

### 3.4 Bodenbedeckungs-Variablen

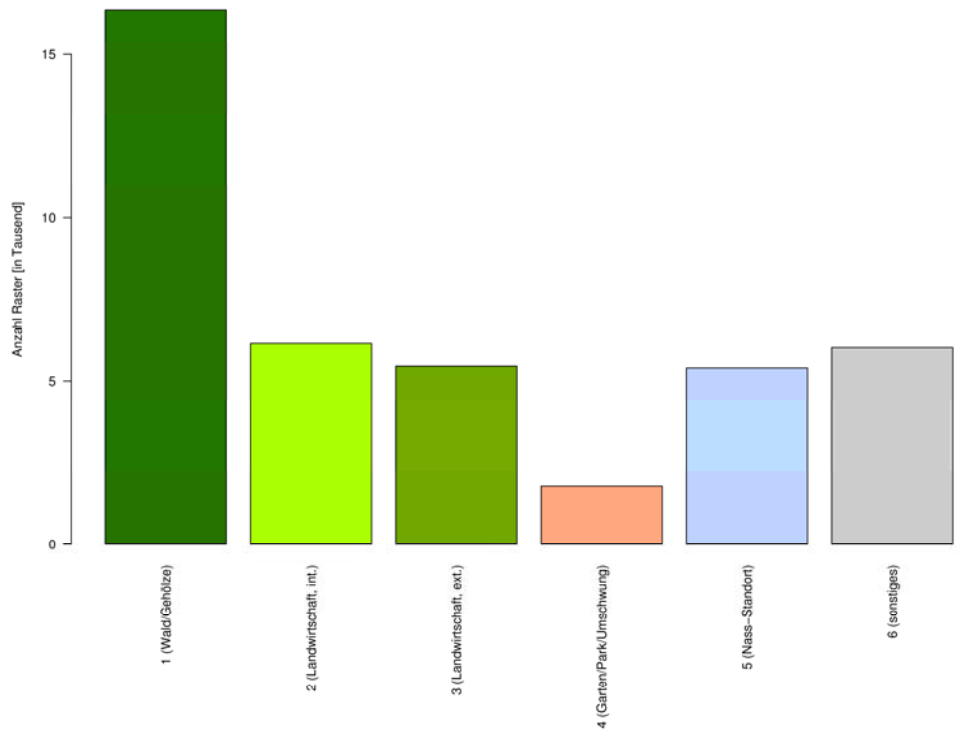
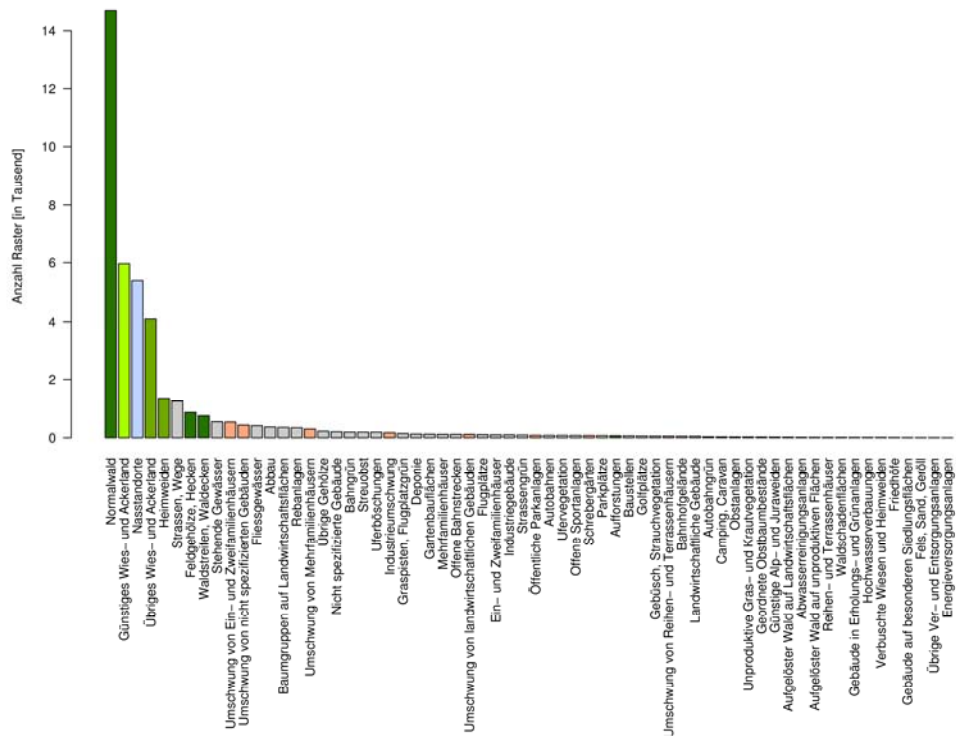


Abbildung 41: Werteverteilung der Bodenbedeckungs-Variablen Die Ursprungskategorien der reklassifizierten Kategorien von as97r (unten) können durch die Farbgebung zugeordnet werden..

### 3.5 Distanzen zu Siedlungsflächen

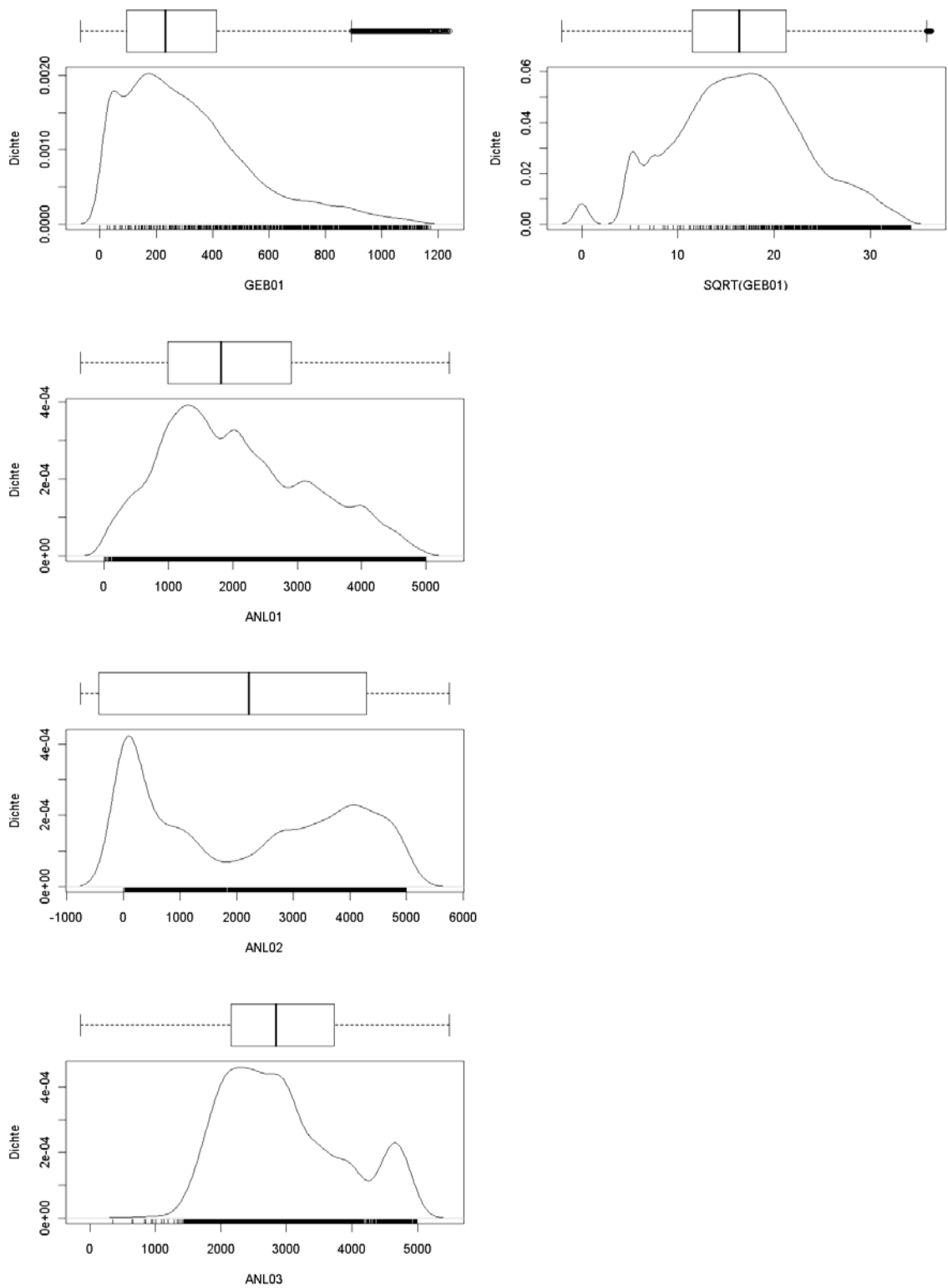


Abbildung 42: Werteverteilung der Distanz-Variablen Siedlungsflächen vor (links) und nach einer eventuell vorgenommenen Transformation.

### 3.6 Distanzen zu Verkehrswegen

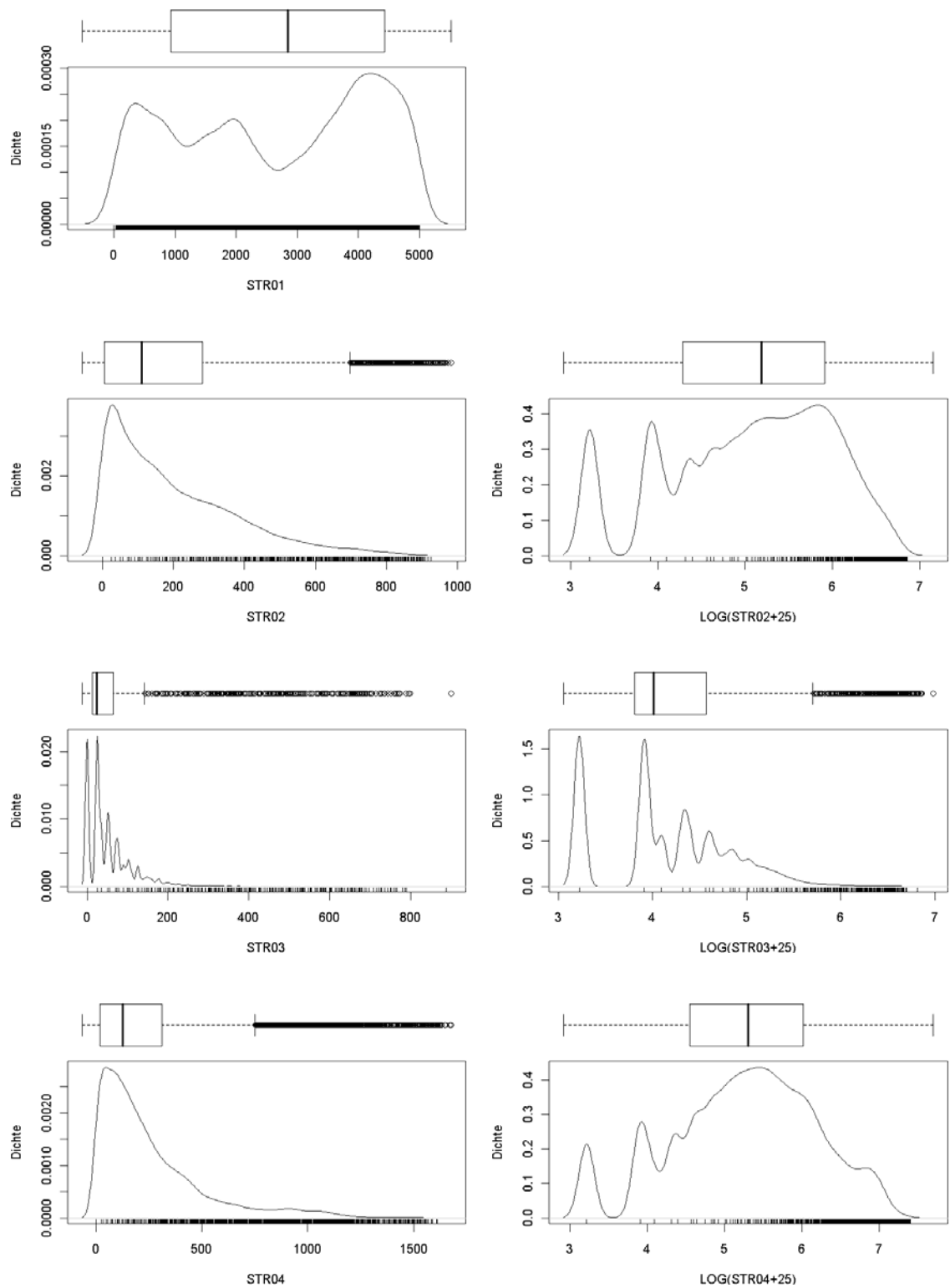


Abbildung 43: Werteverteilung der Distanz-Variablen Verkehrsflächen I vor (links) und nach einer eventuell vorgenommenen Transformation.



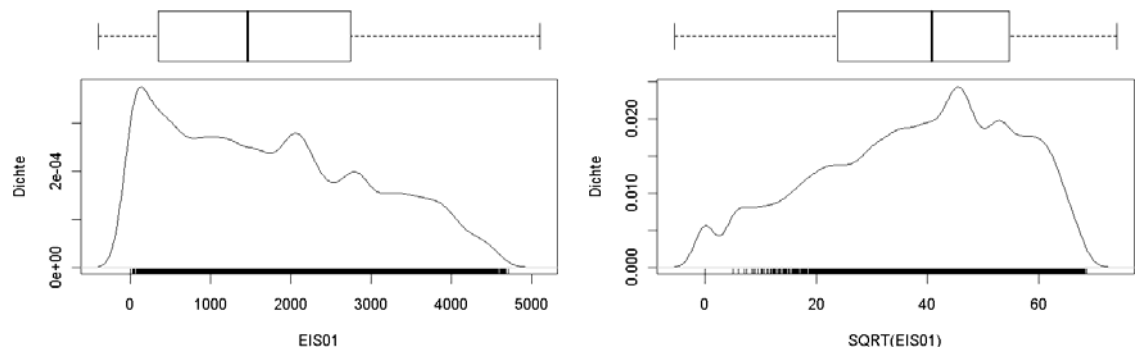


Abbildung 44: Werteverteilung der Distanz-Variablen Verkehrsflächen II vor (links) und nach einer eventuell vorgenommenen Transformation.

### 3.7 Distanzen zu Gewässern

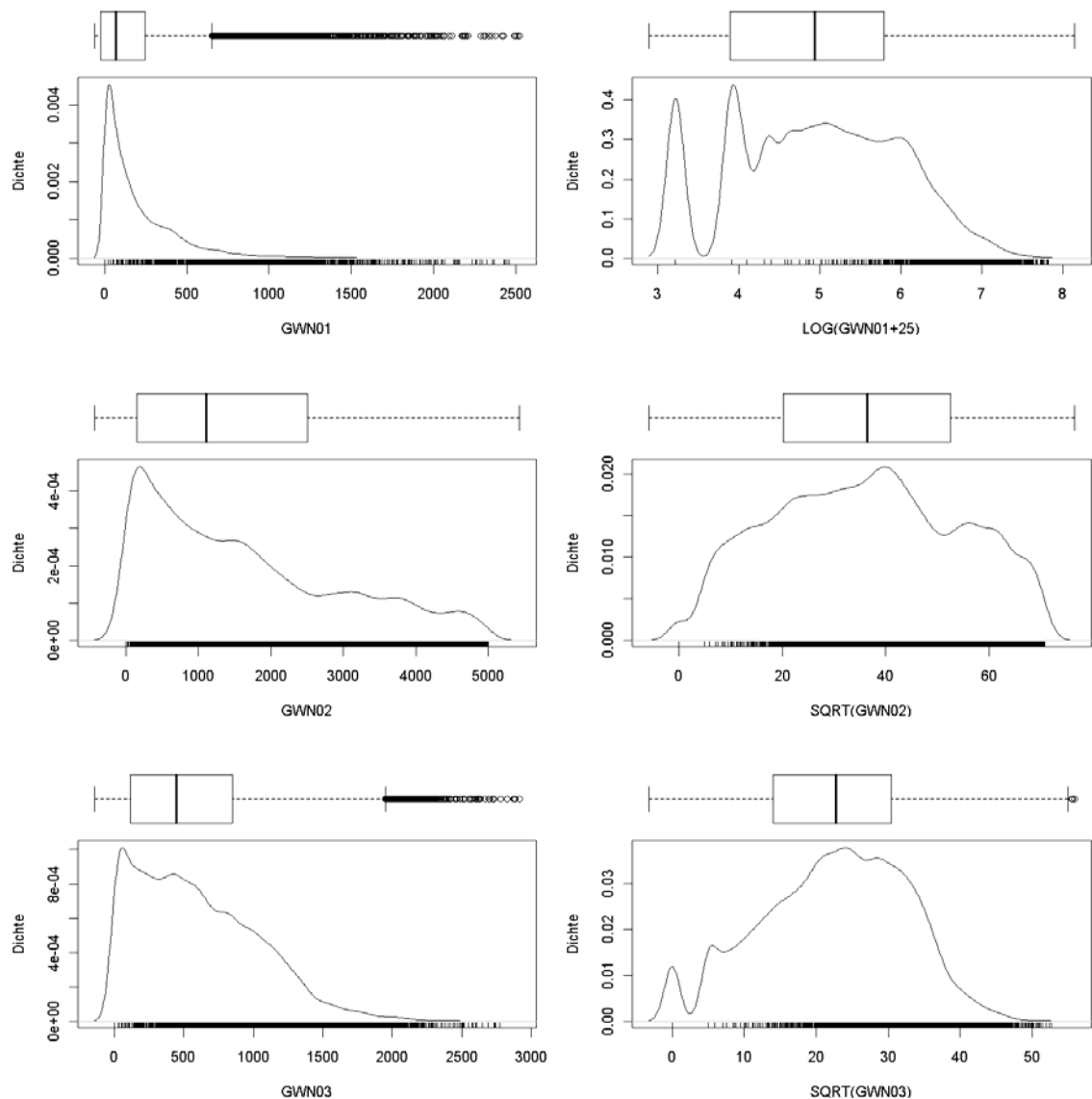


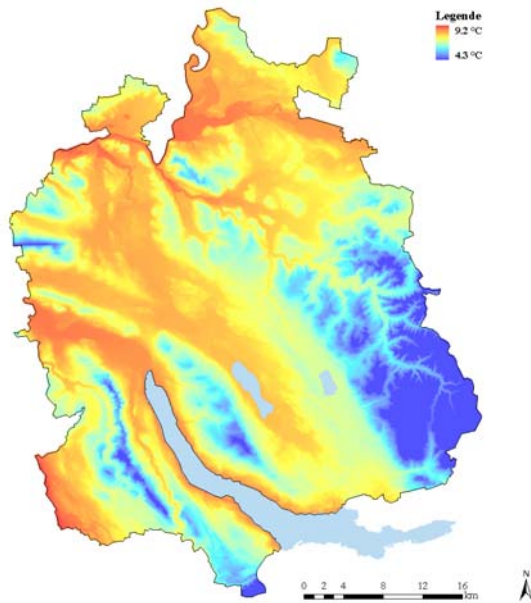
Abbildung 45: Werteverteilung der Distanz-Variablen Gewässern vor (links) und nach einer eventuell vorgenommenen Transformation.

## 4 Variablen-Karten

### 4.1 Klima-Variablen

Klima-Variable

Jahresmitteltemperatur (TAVE)



Klima-Variable

pot. Globalstrahlung im Dezember (SRAD12)

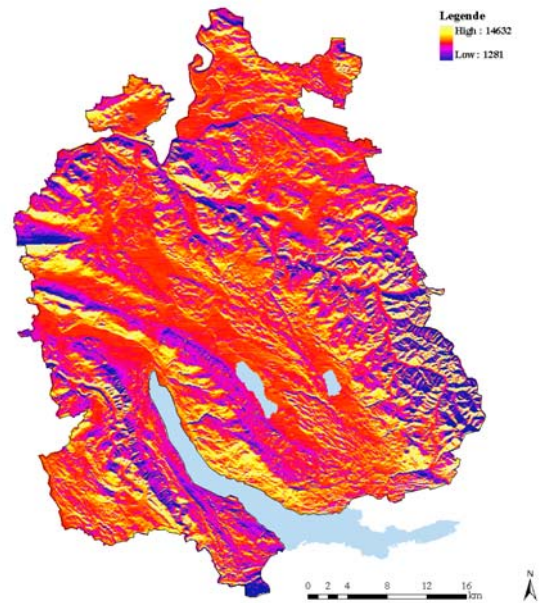
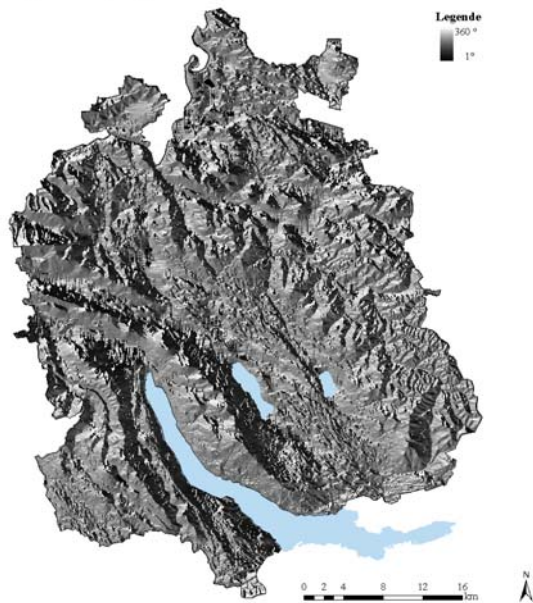


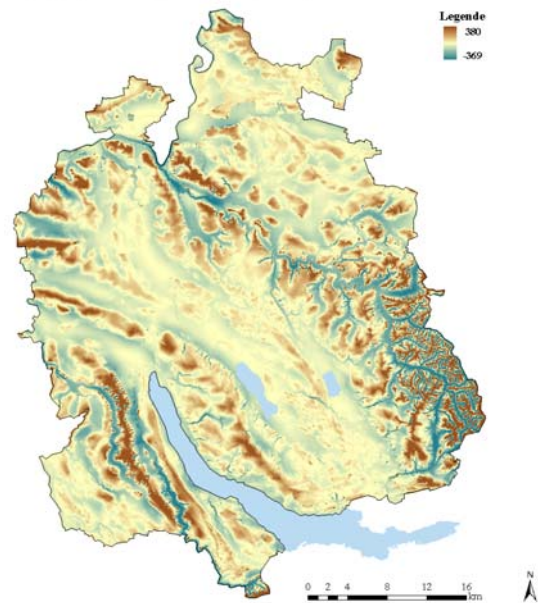
Abbildung 46: Layer-Darstellung der Klima-Variablen.

## 4.2 Topografische Variablen

Topografische-Variablen  
*Exposition (ASP)*



Topografische-Variablen  
*Topografische Exponiertheit (TOPS)*



Topografische-Variablen  
*Hangneigung (SLP)*

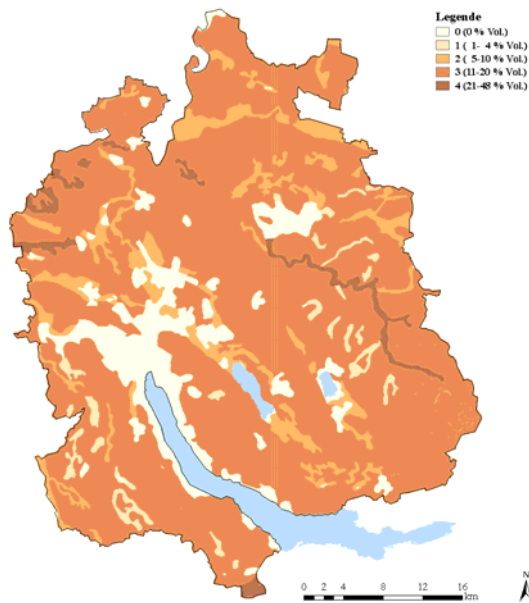


Abbildung 47: Layer-Darstellung der Topografischen-Variablen.

### 4.3 Physikalische Boden-Variablen

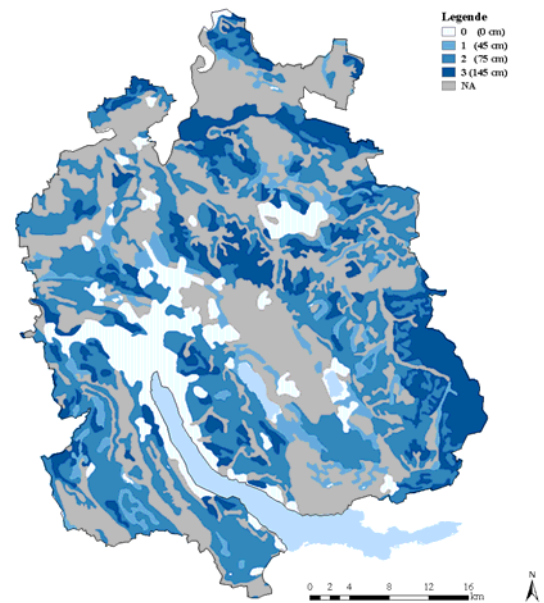
#### Physikalische Boden-Variablen

*Skelettgehalt (SKELE)*



#### Physikalische Boden-Variablen

*Vernässungsgrad (VNASS) - Lage d. porengesättigt. Horizontes*



#### Physikalische Boden-Variablen

*Wasserspeichervermögen (WSPEI)*

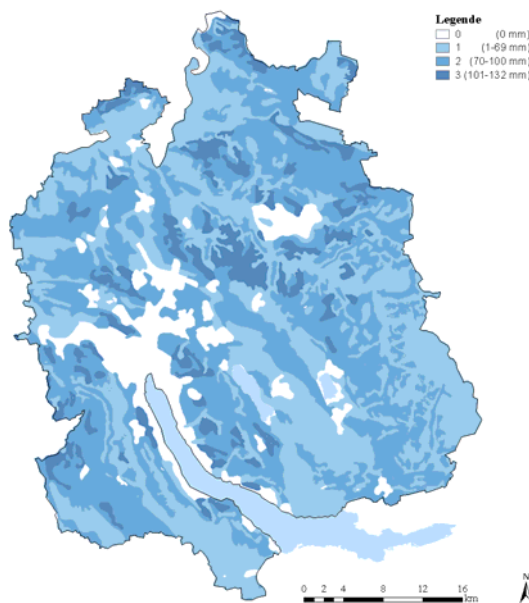


Abbildung 48: Layer-Darstellung der physikalischen Boden-Variablen.

## 4.4 Bodenbedeckungs-Variable

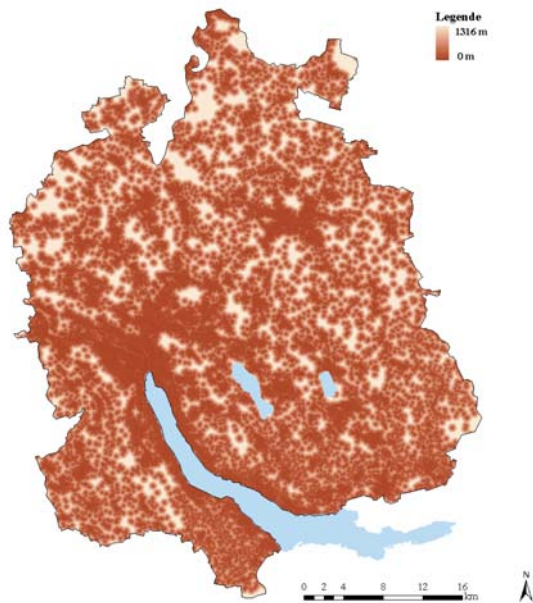


Abbildung 49: Layer-Darstellung der Bodenbedeckungs-Variable.

## 4.5 Distanzen zu Siedlungsflächen

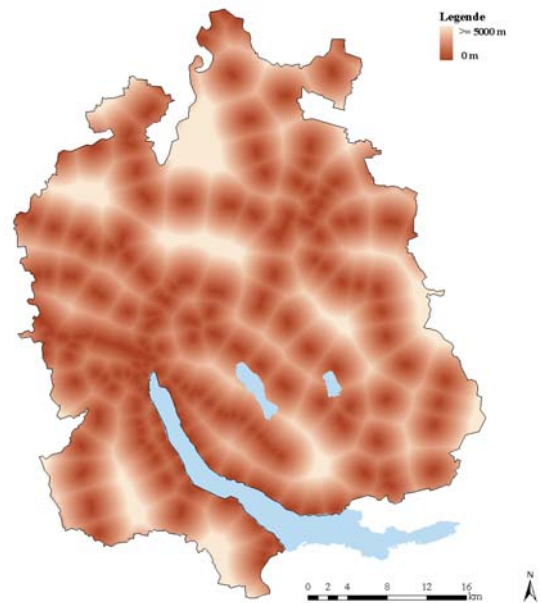
Distanz-Variable

*Gebäude (GEB01)*



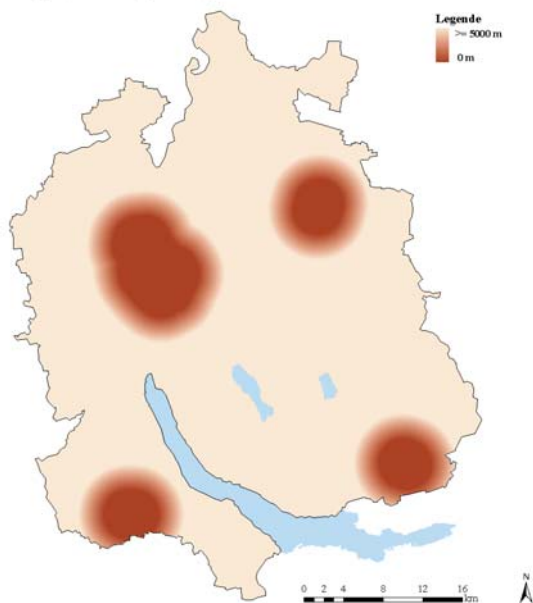
Distanz-Variable

*Bahnhofareal (ANL01)*



Distanz-Variable

*Flughafenareal (ANL02)*



Distanz-Variable

*Flughafenbahnhofareal (ANL03)*

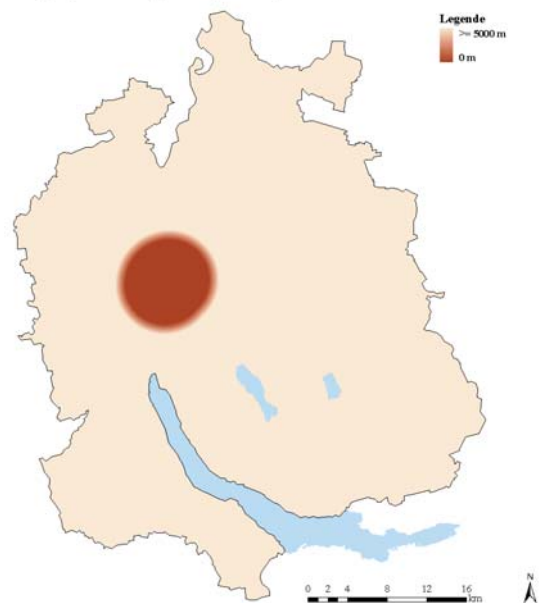
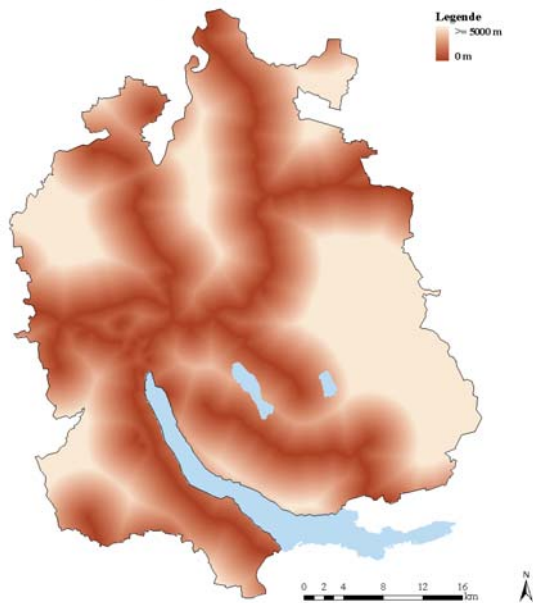


Abbildung 50: Layer-Darstellung der Distanz-Variablen zu Siedlungsflächen.

## 4.6 Distanzen zu Verkehrswegen

Distanz-Variable

Autobahn (STR01)



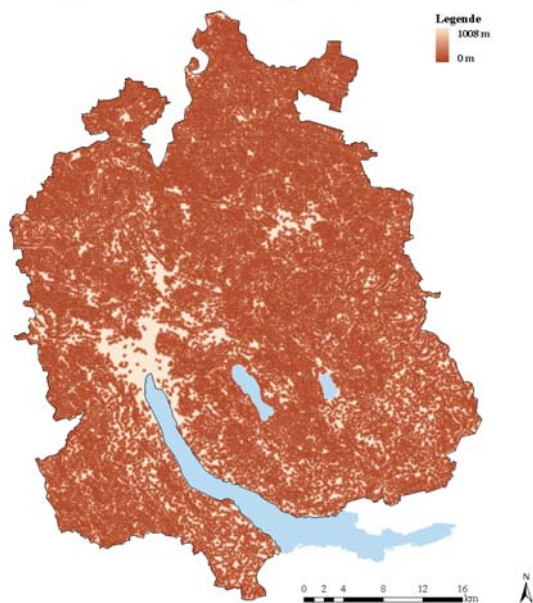
Distanz-Variable

asphaltierte Strasse (STR02)



Distanz-Variable

nicht asphaltierter Feld/Waldweg (STR03)



Distanz-Variable

Fussweg (STR04)

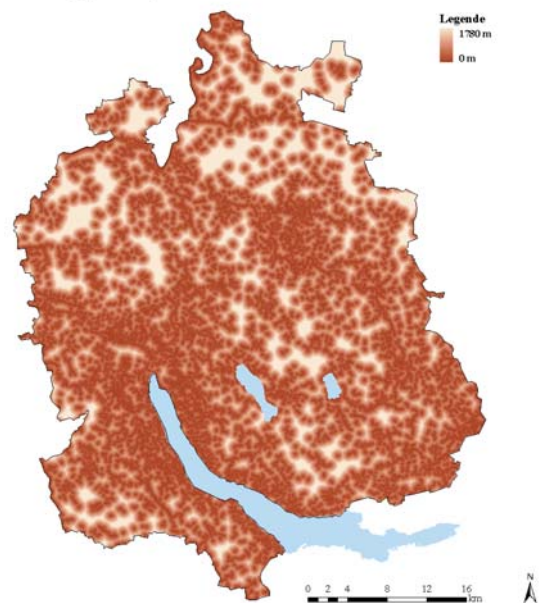


Abbildung 51: Layer-Darstellung der Distanz-Variablen zu Verkehrswegen I.

### Distanz-Variable

*Eisenbahn (EIS01)*

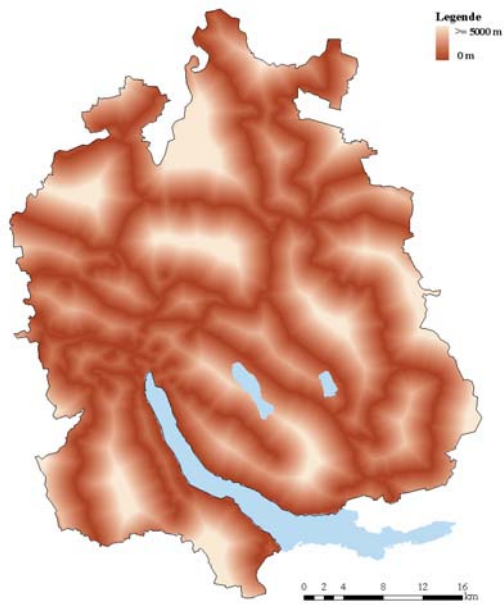


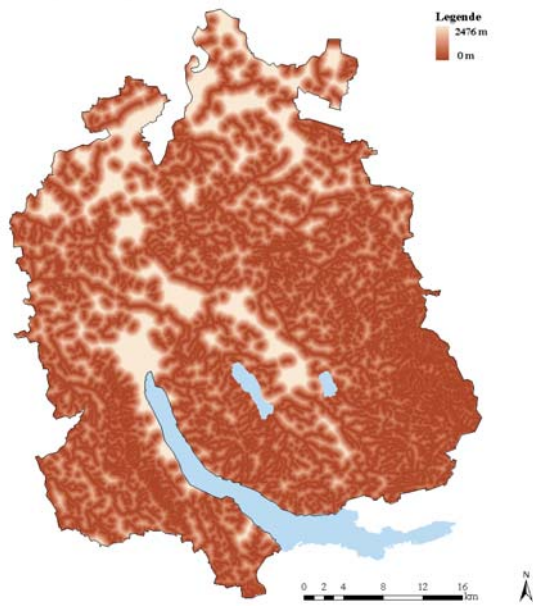
Abbildung 52: Layer-Darstellung der Distanz-Variablen zu Verkehrswegen II.



## 4.7 Distanzen zu Gewässern

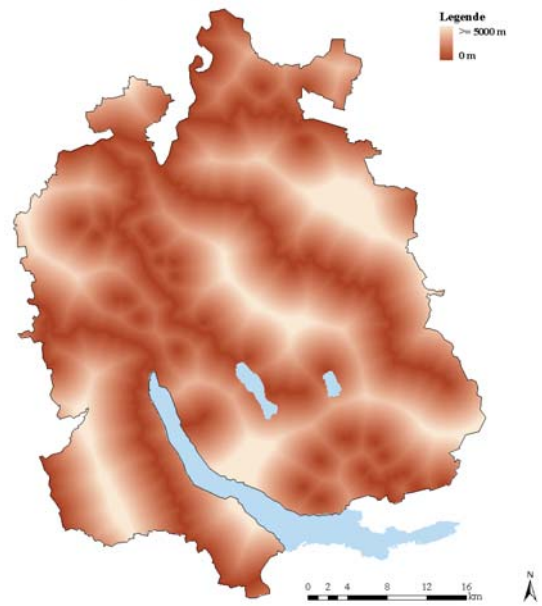
### Distanz-Variable

Bach (GWN01)



### Distanz-Variable

Fluss/Kanal (GWN02)



### Distanz-Variable

See (GWN03)

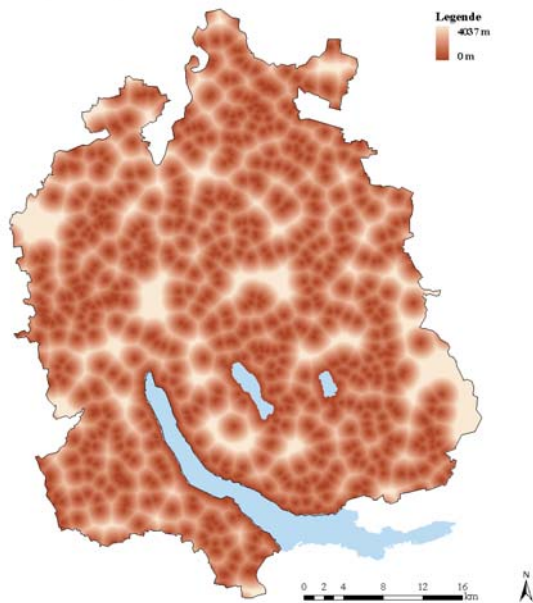


Abbildung 53: Layer-Darstellung der Distanz-Variablen zu Gewässern.

## 5 Variablensets der Modelle

Variablensets der Modelle nach schrittweiser Variablen-Selektion.

Tabelle 30: Variablensets der Modelle für *Ambrosia artemisiifolia*

GLM	Variablenset
<b>trans</b>	
red. BIC	$\text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR04} + 25) + \text{sqrt}(\text{EIS01})$
<b>klima</b>	
red. AIC	$\text{TAVE}^6$
<b>topo</b>	
red. AIC	$\text{TOPS} + \log(\text{SLP} + 1)$
<b>gesamt</b>	
red. AIC	$\text{sqrt}(\text{GEB01}) + \text{STR01} + \log(\text{STR02} + 25) + \log(\text{GWN01} + 25) + \text{sqrt}(\text{GWN02}) + \text{sqrt}(\text{GWN03}) + \text{TAVE} + \text{ASPC} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\text{sqrt}(\text{GEB01}) + \text{STR01} + \log(\text{STR02} + 25) + \text{sqrt}(\text{GWN02}) + \text{sqrt}(\text{GWN03}) + \text{TAVE} + \text{AS97}$
<b>gesamt ^2</b>	
red. AIC	$\text{sqrt}(\text{GEB01}) + \text{STR01} + \text{I}((\log(\text{STR02} + 25))^2) + \log(\text{STR03} + 25) + \text{I}((\log(\text{STR03} + 25))^2) + \log(\text{GWN01} + 25) + \text{sqrt}(\text{GWN02}) + \text{sqrt}(\text{GWN03}) + \text{I}((\text{TAVE}^6)^2) + \log(\text{SLP} + 1) + \text{I}((\log(\text{SLP} + 1))^2) + \text{WSPEI} + \text{AS97R}$
red. BIC	$\text{sqrt}(\text{GEB01}) + \text{STR01} + \text{I}((\log(\text{STR02} + 25))^2) + \log(\text{GWN01} + 25) + \text{sqrt}(\text{GWN02}) + \text{sqrt}(\text{GWN03}) + \log(\text{SLP} + 1) + \text{I}((\log(\text{SLP} + 1))^2) + \text{AS97R}$

Tabelle 31: Variablensets der Modelle für *Heracleum mantegazzianum*

GLM	Variablenset
<b>trans</b>	
red. AIC	$(\text{STR01}) + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + + \log(\text{STR04} + 25) + \text{sqrt}(\text{EIS01})$
red. BIC	$(\text{STR01}) + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + + \log(\text{STR04} + 25) + \text{sqrt}(\text{EIS01})$
<b>gew</b>	
red. AIC	$\text{sqrt}(\text{GWN02}) + \text{sqrt}(\text{GWN03})$
red. BIC	$\text{sqrt}(\text{GWN02})$
<b>klima</b>	
red. BIC	$\text{TAVE}^6$
<b>topo</b>	
red. AIC	$\text{TOPS} + \log(\text{SLP} + 1)$
red. BIC	$\text{TOPS}$
<b>boden</b>	
red. BIC	$\text{SKELE}$
<b>gesamt</b>	
red. AIC	$\text{sqrt}(\text{GEB01}) + \text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \text{sqrt}(\text{EIS01}) + \text{sqrt}(\text{GWN02}) + \text{SRAD12} + \text{TOPS} + \log(\text{SLP} + 1) + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\text{sqrt}(\text{GEB01}) + \text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \text{sqrt}(\text{EIS01}) + \text{sqrt}(\text{GWN02}) + \text{SRAD12} + \log(\text{SLP} + 1)$
<b>gesamt ^2</b>	
red. AIC	$\text{sqrt}(\text{GEB01}) + \text{I}((\text{sqrt}(\text{GEB01}))^2) + \text{STR01} + \text{I}(\text{STR01}^2) + \text{I}((\log(\text{STR02} + 25))^2) + \log(\text{STR03} + 25) + \text{I}((\log(\text{STR03} + 25))^2) + \log(\text{STR04} + 25) + \text{I}((\log(\text{STR04} + 25))^2) + \text{I}((\text{sqrt}(\text{EIS01}))^2) + \text{sqrt}(\text{GWN02}) + \text{TAVE} + \text{I}((\text{TAVE}^6)^2) + \text{SRAD12} + \text{I}(\text{SRAD12}^2) + \text{I}(\text{ASPC}^2) + \text{TOPS} + \log(\text{SLP} + 1) + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\text{sqrt}(\text{GEB01}) + \text{I}(\text{STR01}^2) + \text{I}((\log(\text{STR02} + 25))^2) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \text{I}((\text{sqrt}(\text{EIS01}))^2) + \text{sqrt}(\text{GWN02}) + \text{I}(\text{SRAD12}^2)$

Tabelle 32: Variablensets der Modelle für *Reynoutria japonica*.

GLM	Variablenset
<b>trans</b>	
red. BIC	$\text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR04} + 25) + \sqrt{\text{EIS01}}$
<b>topo</b>	
red. BIC	TOPS
<b>boden</b>	
red. BIC	WSPEI
<b>gesamt</b>	
red. AIC	$\sqrt{\text{GEB01}} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \sqrt{\text{EIS01}} + \sqrt{\text{GWN02}} + \sqrt{\text{GWN03}} + \text{SRAD12} + \text{TOPS} + \log(\text{SLP} + 1) + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\sqrt{\text{GEB01}} + \log(\text{STR02} + 25) + \log(\text{STR04} + 25) + \sqrt{\text{EIS01}} + \text{SRAD12} + \text{TOPS} + \text{WSPEI}$
<b>gesamt ^2</b>	
red. AIC	$I((\sqrt{\text{GEB01}})^2) + I(\text{STR01}^2) + \log(\text{STR02} + 25) + I((\log(\text{STR02} + 25))^2) + \log(\text{STR03} + 25) + I((\log(\text{STR03} + 25))^2) + \log(\text{STR04} + 25) + I((\sqrt{\text{EIS01}})^2) + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + 25 + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + \text{TAVE} + I((\text{TAVE}^6)^2) + I(\text{SRAD12}^2) + \text{ASPC} + I(\text{ASPC}^2) + \text{TOPS} + I(\text{TOPS}^2) + \log(\text{SLP} + 1) + I((\log(\text{SLP} + 1))^2) + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\sqrt{\text{GEB01}} + \log(\text{STR02} + 25) + I((\log(\text{STR02} + 25))^2) + \log(\text{STR03} + 25) + I((\log(\text{STR03} + 25))^2) + \log(\text{STR04} + 25) + \sqrt{\text{EIS01}} + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + \text{TAVE} + I((\text{TAVE}^6)^2) + I(\text{SRAD12}^2) + \text{TOPS} + \log(\text{SLP} + 1) + I((\log(\text{SLP} + 1))^2) + \text{WSPEI}$

Tabelle 33: Variablensets der Modelle für *Impatiens glandulifera*.

GLM	Variablenset
<b>trans</b>	
red. AIC	$\text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25)$
<b>gew</b>	
red. AIC	$\log(\text{GWN01} + 25) + \sqrt{\text{GWN02}}$
<b>boden</b>	
red. BIC	SKELE
<b>gesamt</b>	
red. AIC	$\sqrt{\text{GEB01}} + \text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \log(\text{GWN01} + 25) + \sqrt{\text{GWN02}} + \sqrt{\text{GWN03}} + \text{TAVE} + \text{SRAD12} + \text{ASPC} + \text{TOPS} + \log(\text{SLP} + 1) + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \sqrt{\text{GWN02}} + \text{TAVE} + \text{SRAD12} + \text{TOPS} + \log(\text{SLP} + 1) + \text{SKELE} + \text{AS97R}$
<b>gesamt ^2</b>	
red. AIC	$\sqrt{\text{GEB01}} + I((\sqrt{\text{GEB01}})^2) + \text{STR01} + I(\text{STR01}^2) + \log(\text{STR02} + 25) + I((\log(\text{STR02} + 25))^2) + I((\log(\text{STR03} + 25))^2) + \log(\text{STR04} + 25) + I((\log(\text{STR04} + 25))^2) + \sqrt{\text{EIS01}} + I((\sqrt{\text{EIS01}})^2) + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + \sqrt{\text{GWN03}} + I((\sqrt{\text{GWN03}})^2) + \text{TAVE} + I((\text{TAVE}^6)^2) + \text{SRAD12} + I(\text{SRAD12}^2) + \text{ASPC} + I(\text{ASPC}^2) + \text{TOPS} + I(\text{TOPS}^2) + I((\log(\text{SLP} + 1))^2) + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\sqrt{\text{GEB01}} + I((\sqrt{\text{GEB01}})^2) + I(\text{STR01}^2) + I((\log(\text{STR02} + 25))^2) + I((\log(\text{STR03} + 25))^2) + I((\log(\text{STR04} + 25))^2) + \sqrt{\text{EIS01}} + I((\sqrt{\text{EIS01}})^2) + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + \text{TAVE} + I((\text{TAVE}^6)^2) + I(\text{SRAD12}^2) + \text{TOPS} + I(\text{TOPS}^2) + I((\log(\text{SLP} + 1))^2) + \text{SKELE} + \text{AS97R}$

Tabelle 34: Variablensets der Modelle für *Rhus typhina*.

GLM	Variablenset
<b>trans</b>	
red. AIC	$\log(\text{STR02} + 25) + \log(\text{STR04} + 25) + \sqrt{\text{EIS01}}$
<b>gew</b>	
red. BIC	$\log(\text{GWN01} + 25)$
<b>topo</b>	
red. AIC	$\text{TOPS} + \log(\text{SLP} + 1)$
<b>boden</b>	
red. AIC	$\text{SKELE}$
<b>gesamt</b>	
red. AIC	$\sqrt{\text{GEB01}} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \text{ASPC} + \log(\text{SLP} + 1) + \text{WSPEI} + \text{AS97R}$
red. BIC	$\sqrt{\text{GEB01}} + \log(\text{STR02} + 25) + \log(\text{SLP} + 1)$
<b>gesamt ^2</b>	
red. AIC	$\sqrt{\text{GEB01}} + I((\sqrt{\text{GEB01}})^2) + I((\log(\text{STR02} + 25))^2) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + I((\log(\text{STR04} + 25))^2) + I((\sqrt{\text{EIS01}})^2) + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + I((\text{TAVE}^6)^2) + \text{SRAD12} + \text{ASPC} + \log(\text{SLP} + 1) + \text{WSPEI}$
red. BIC	$\sqrt{\text{GEB01}} + I(\text{STR01}^2) + I((\log(\text{STR02} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + \log(\text{SLP} + 1)$

Tabelle 35: Variablensets der Modelle für *Solidago canadensis*.

GLM	Variablenset
<b>gew</b>	
red. BIC	$\log(\text{GWN01} + 25) + \sqrt{\text{GWN02}}$
<b>klima</b>	
red. AIC	$(\text{TAVE}^6)$
<b>topo</b>	
red. AIC	$\text{ASPC} + \log(\text{SLP} + 1)$
red. BIC	$\log(\text{SLP} + 1)$
<b>gesamt</b>	
red. AIC	$\text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \sqrt{\text{EIS01}} + \sqrt{\text{GWN02}} + \sqrt{\text{GWN03}} + \text{TAVE} + \text{ASPC} + \text{TOPS} + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \sqrt{\text{EIS01}} + \sqrt{\text{GWN03}} + \text{TAVE} + \text{TOPS} + \text{SKELE}$
<b>gesamt ^2</b>	
red. AIC	$\text{STR01} + I(\text{STR01}^2) + \log(\text{STR02} + 25) + I((\log(\text{STR02} + 25))^2) + I((\log(\text{STR03} + 25))^2) + \log(\text{STR04} + 25) + I((\log(\text{STR04} + 25))^2) + \sqrt{\text{EIS01}} + I((\sqrt{\text{EIS01}})^2) + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + I((\sqrt{\text{GWN03}})^2) + I((\text{TAVE}^6)^2) + \text{SRAD12} + \text{ASPC} + I(\text{ASPC}^2) + \text{TOPS} + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\text{STR01} + I(\text{STR01}^2) + \log(\text{STR02} + 25) + I((\log(\text{STR02} + 25))^2) + I((\log(\text{STR03} + 25))^2) + \log(\text{STR04} + 25) + I((\log(\text{STR04} + 25))^2) + \sqrt{\text{EIS01}} + I((\sqrt{\text{EIS01}})^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + I((\sqrt{\text{GWN03}})^2) + I((\text{TAVE}^6)^2) + \text{TOPS} + \text{SKELE} + \text{WSPEI}$

Tabelle 36: Variablensets der Modelle für *Solidago gigantea*.

GLM	Variablenset
<b>klima</b>	
red. BIC	SRAD12
<b>boden</b>	
red. BIC	SKELE
<b>gesamt</b>	
red. AIC	$\sqrt{\text{GEB01}} + \text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \log(\text{GWN01} + 25) + \sqrt{\text{GWN02}} + \sqrt{\text{GWN03}} + \text{TAVE} + \text{SRAD12} + \text{ASPC} + \text{TOPS} + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\sqrt{\text{GEB01}} + \text{STR01} + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + \log(\text{STR04} + 25) + \log(\text{GWN01} + 25) + \sqrt{\text{GWN02}} + \sqrt{\text{GWN03}} + \text{TAVE} + \text{SRAD12} + \text{ASPC} + \text{TOPS} + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
<b>gesamt ^2</b>	
red. AIC	$\sqrt{\text{GEB01}} + I((\sqrt{\text{GEB01}})^2) + \text{STR01} + \log(\text{STR02} + 25) + I((\log(\text{STR02} + 25))^2) + \log(\text{STR03} + 25) + I((\log(\text{STR03} + 25))^2) + \log(\text{STR04} + 25) + I((\log(\text{STR04} + 25))^2) + \sqrt{\text{EIS01}} + I((\sqrt{\text{EIS01}})^2) + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + \sqrt{\text{GWN03}} + I((\sqrt{\text{GWN03}})^2) + \text{TAVE} + I((\text{TAVE}^6)^2) + \text{SRAD12} + I(\text{SRAD12}^2) + \text{ASPC} + I(\text{ASPC}^2) + \text{TOPS} + I(\text{TOPS}^2) + \log(\text{SLP} + 1) + I((\log(\text{SLP} + 1))^2) + \text{SKELE} + \text{WSPEI} + \text{AS97R}$
red. BIC	$\sqrt{\text{GEB01}} + I((\sqrt{\text{GEB01}})^2) + \log(\text{STR02} + 25) + \log(\text{STR03} + 25) + I((\log(\text{STR03} + 25))^2) + I((\log(\text{STR04} + 25))^2) + \sqrt{\text{EIS01}} + I((\sqrt{\text{EIS01}})^2) + \log(\text{GWN01} + 25) + I((\log(\text{GWN01} + 25))^2) + \sqrt{\text{GWN02}} + I((\sqrt{\text{GWN02}})^2) + \sqrt{\text{GWN03}} + I((\sqrt{\text{GWN03}})^2) + \text{TAVE} + I((\text{TAVE}^6)^2) + I(\text{SRAD12}^2) + \text{ASPC} + \text{TOPS} + I(\text{TOPS}^2) + \log(\text{SLP} + 1) + \text{SKELE} +$