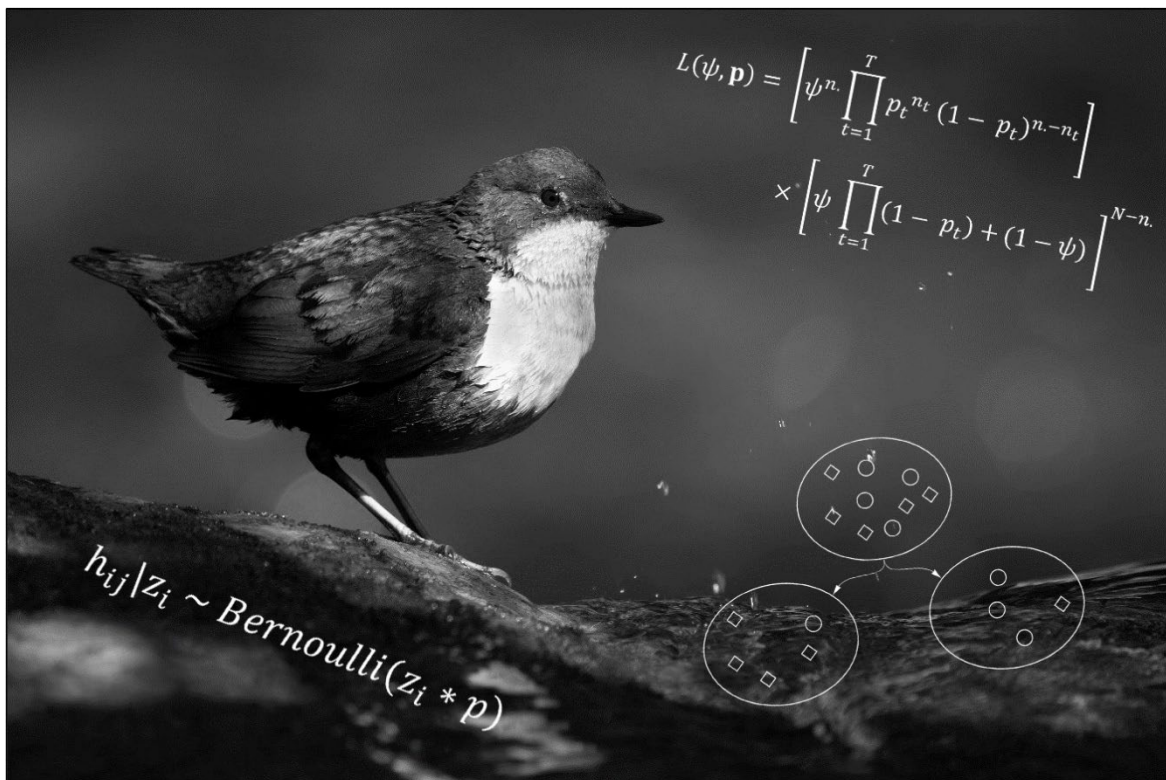


Vorhersage von Nachweiswahrscheinlichkeiten und Vorkommen der Wasseramsel anhand mathematischer Modelle



Florian Hatt, C6c

Betreut von Hanspeter Siegfried, Informatik

Maturitätsarbeit 2023

Kantonsschule Zürcher Oberland Wetzikon

Abstract

In dieser Arbeit wurde das Vorkommen der Wasseramsel an elf Kilometern Bachlauf im Zürcher Oberland erfasst. Damit die Datenaufnahme auch an schlecht einsehbaren Bereichen durchgeführt werden konnte, wurde die Methode der Punktkartierung gewählt, bei welcher an definierten Beobachtungsstandorten Präsenz-/Absenzdaten während mehreren Kontrollgängen erhoben wurden. Daraus liessen sich Site-Occupancy-Modelle erstellen, welche die Nachweis- und Vorkommenswahrscheinlichkeit anhand von äusseren Faktoren berechnen können. Es zeigte sich, dass frühmorgens die besten Chancen bestehen, Wasseramseln nachzuweisen, diese Chancen aber zwischen verschiedenen Phasen der Brutzeit schwanken. Zusätzlich wurde versucht, das Vorkommen anhand von Bachstrukturen an den Beobachtungsstandorten vorherzusagen. Es gelang allerdings kaum, verlässliche Aussagen über den Einfluss dieser Elemente zu treffen. Ein Decision-Tree-Modell als Vergleich lieferte noch unzuverlässigere Resultate. Die Gründe für all diese Befunde werden ausführlich diskutiert.

Inhaltsverzeichnis

ABSTRACT	2
1. EINLEITUNG	5
2. GRUNDLAGEN	6
2.1 WASSERAMSEL	6
2.1.1 LEBENSRAUMANSPRÜCHE	6
2.1.2 BRUTGESCHÄFT UND TERRITORIALITÄT	6
2.2 SITE-OCCUPANCY-MODELLE	7
2.2.1 MAXIMUM-LIKELIHOOD-METHODE.....	8
2.2.2 GRUNDLEGENDE ANNAHMEN UND VORAUSSETZUNGEN	10
2.2.3 MODELLVERGLEICH	10
2.2.4 SENSITIVITÄTSANALYSE.....	11
2.2.5 ANZAHL KONTROLLGÄNGE UND WAHRSCHEINLICHER BESETZUNGSSTATUS.....	12
2.3 DECISION-TREE-MODELLE	12
2.3.1 BILDUNG EINES DECISION-TREES	13
2.3.2 OVERFITTING.....	15
3. METHODIK	16
3.1 ÜBERBLICK.....	16
3.2 DATENAUFNAHME	16
3.3 SITE-OCCUPANCY-MODELLE	17
3.3.1 GRUNDLEGENDES MODELL	17
3.3.2 MODELLIERUNG MIT COVARIABLEN	18
3.3.3 SENSITIVITÄTSANALYSE.....	20
3.3.4 ANZAHL KONTROLLGÄNGE UND WAHRSCHEINLICHER BESETZUNGSSTATUS.....	21
3.4 DECISION-TREE-MODELL	21
4. RESULTATE	22
4.1 ÜBERBLICK ÜBER DIE GESAMMELTEN PRÄSENZ-/ABSENZDATEN.....	22
4.2 SITE-OCCUPANCY-MODELLE	22
4.2.1 GRUNDLEGENDES MODELL	22
4.2.2 MODELLIERUNG MIT ZEITLICHEN COVARIABLEN	23
4.2.3 MODELLIERUNG MIT BESETZUNGSSTATUS DER NACHBARSTANDORTE ALS COVARIABLE.....	24
4.2.4 MODELLIERUNG MIT BACHSTRUKTUREN ALS COVARIABLEN	24
4.2.5 SENSITIVITÄTSANALYSE.....	25
4.2.6 ANZAHL KONTROLLGÄNGE UND WAHRSCHEINLICHER BESETZUNGSSTATUS.....	26
4.3 DECISION-TREE-MODELL	26
5. DISKUSSION	27
5.1 ABSCHÄTZUNG DER VORKOMMEN MITTELS PRÄSENZ-/ABSENZDATEN	27
5.2 EINFLUSS VON ZEITLICHEN COVARIABLEN AUF DIE DETECTION PROBABILITY	27
5.3 EINFLUSS DES BESETZUNGSSTATUS DER NACHBARSTANDORTE AUF DIE OCCUPANCY	29
5.4 VORHERSAGEN ZUM VORKOMMEN ANHAND DER BACHSTRUKTUREN.....	29

5.4.1	MANGELNDE AUSSAGEKRAFT DER ERHOBENEN BACHSTRUKTUREN.....	29
5.4.2	GRÖSSE DES STICHPROBENUMFANGS.....	30
5.4.3	EIGNUNG DER VERWENDETEN MODELIERUNGSANSÄTZE.....	30
5.5	EINHALTUNG GRUNDLEGENDER ANNAHMEN UND VORAUSSETZUNGEN DER SINGLE-SEASON SITE-OCCUPANCY-MODELLE.....	31
6.	SCHLUSSWORT.....	33
7.	LITERATURVERZEICHNIS.....	34
8.	ABBILDUNGSVERZEICHNIS.....	36
9.	TABELLENVERZEICHNIS.....	38
10.	DANK.....	39
11.	ANHANG.....	40

1. Einleitung

Vögel faszinieren mich schon seit Jahren. Besonders der pandemiebedingte Lockdown ermöglichte es mir, viel Zeit draussen zu verbringen. Durch zahlreiche Beobachtungen von Wasseramseln in dieser Zeit und auch durch die Maturitätsarbeit meines Bruders (Hatt, 2022) wuchs mein Bedürfnis, mich auch in meiner Maturitätsarbeit vertieft mit dieser Art auseinanderzusetzen.

Da ich unterdessen beim kantonalen Projekt «Avimonitoring» mitarbeite, wurde mir bewusst, dass der Nachweis von Wasseramseln verglichen mit anderen Arten schwierig sein kann. Normalerweise werden bei ornithologischen Kartierungen nämlich kontinuierlich Daten gesammelt, während man eine bestimmte Strecke abläuft. Bei Wasseramseln kann die übliche Vorgehensweise bei Kartierungen aber an ihre Grenzen stossen. Einerseits führen Wasseramseln ein recht diskretes Leben. Reviergesänge vermögen kaum das Bachrauschen zu übertönen (Glutz von Blotzheim & Bauer, 1985). Andererseits sind die langen, linearen Reviere teilweise schlecht zugänglich oder einsehbar. Ohne ein Wissen über Nachweis- oder Vorkommenswahrscheinlichkeiten kann es in solchen Fällen schwierig werden, korrekte Aussagen über Wasseramselvorkommen zu machen. Dies ist z.B. auch im Aathal zwischen Wetzikon und Uster der Fall, da der Aabach hier nur an wenigen Stellen eingesehen werden kann. Aus diesem Grund war unklar, ob Wasseramseln hier regelmässig vorkommen. Es stellte sich mir daher die Frage, ob sich diese Situation mit einer angepassten Kartierungsmethode verbessern liesse. Dabei wurde ich inspiriert von Steffen (2020). Er versuchte, das Vorkommen einer ganz anderen, schwer nachweisbaren Art, nämlich dem Hermelin, zu bestimmen. Dabei verwendete er nicht kontinuierlich entlang einer Wegstrecke erhobene Daten, sondern Präsenz-/Absenzdaten von ausgewählten Geländepunkten.

Die vorliegende Arbeit greift diese Idee auf und versucht, die Nachweis- und Vorkommenswahrscheinlichkeit von Wasseramseln mittels Punktbeobachtungen zu quantifizieren, weil diese auch an schlecht einsehbaren Bachabschnitten möglich sind. Im Zentrum stehen folgende Fragestellungen:

- Gelingen mit Punktbeobachtungen valide Abschätzungen von Wasseramsel-Vorkommen?
- Erlauben es mathematische Modelle, den Einfluss von Tages- und Jahreszeit auf die Nachweiswahrscheinlichkeit abzuschätzen? Kann mit den Erkenntnissen aus diesen Modellen die in dieser Arbeit verwendete Kartierungsmethode optimiert werden?
- Welche Bachstrukturen müssen in einem Modell berücksichtigt werden, sodass dieses Vorhersagen zum Vorkommen von Wasseramseln an einem bestimmten Beobachtungspunkt ermöglicht? Welchen Einfluss auf die Vorkommenswahrscheinlichkeit hat das Vorkommen von Wasseramseln in der Nähe?

2. Grundlagen

2.1 Wasseramsel

Die Wasseramsel (*Cinclus cinclus*) ist ein Vertreter der Familie der Wasseramseln innerhalb der Ordnung der Sperlingsvögel. Wasseramseln besitzen als einzige Singvögel die Fähigkeit, tauchend und schwimmend nach Nahrung zu jagen (Zopfi, 1988). Ihr Beuteschema umfasst dabei hauptsächlich diverse Arten von Wirbellosen. In Europa haben Wasseramseln ein ausgedehntes Verbreitungsgebiet. Aufgrund ihrer Lebensraumansprüche konzentriert sich das Vorkommen allerdings auf mittlere Höhenlagen und gebirgige Regionen (Glutz von



Blotzheim & Bauer, 1985). In der Schweiz sind die Bestände leicht zunehmend, weisen allerdings hohe Schwankungen auf. Zwischen 1988 und 2008 ist die Zahl der Individuen im Kanton Zürich sogar um den Faktor zwei gestiegen (Knaus et al., 2018). Wasseramseln gelten nicht als gefährdet und lassen sich sogar mit passenden Nisthilfen relativ leicht fördern (Rockenbauch, 1985).

Abbildung 1: Adulte Wasseramsel am Chämptnerbach

2.1.1 Lebensraumansprüche

Ein optimaler Lebensraum für Wasseramseln weist gemäss Literatur folgende Merkmale auf: Von Vorteil für das Vorkommen ist ein Bachabschnitt mit ständiger Wasserführung, eine Bachsohle aus Kies oder Schotter sowie die Existenz starker Turbulenzen (Glutz von Blotzheim & Bauer, 1985) und möglicher Verstecke im Uferbereich (Bosch & Lurz, 2019; Martinez et al., 2020). Genutzt werden auch Ansitz-Steine sowie Nebenbäche, die bei höherem Wasserstand des Hauptbaches zur Nahrungssuche aufgesucht werden können. Wasseramseln besitzen somit entlang von Bächen und Flüssen recht eindimensionale, langgestreckte Reviere (Glutz von Blotzheim & Bauer, 1985).

2.1.2 Brutgeschäft und Territorialität

Wasseramseln sind streng territorial und schützen ihr Revier vor Eindringlingen. Sogar im Winter werden die Reviere verteidigt und nur bei Nahrungsknappheit können Artgenossen an verbleibenden Futterquellen toleriert werden (Zopfi, 1988). Die Paarbildung der Wasseramsel beginnt bereits im August und dauert etwa bis März. Die Balz im engeren Sinne findet allerdings erst ab Januar statt, erreicht ihren Höhepunkt dann während der Nestbau-Phase und nimmt nach der Eiablage deutlich ab. Vor möglichen Zweitbruten ist sie wieder vermehrt zu beobachten (Glutz von Blotzheim & Bauer, 1985). Der Nachwuchs, welcher frühestens im März und spätestens im Juli schlüpft (Hegelbach, 2004), hält sich nicht an die Reviergrenzen. Bei der Futterbeschaffung für die Jungen beschränken sich die Altvögel

meist auf die günstigsten Revierabschnitte, folgen ihrem Nachwuchs aber auch problemlos aus dem Revier hinaus (Glutz von Blotzheim & Bauer, 1985). Die Revierdichte schwankt zwischen 0.5 bis 2.5 Brutpaaren pro Kilometer (Hegelbach, 2004; Glutz von Blotzheim & Bauer, 1985).

2.2 Site-Occupancy-Modelle

Unter einem Modell ganz allgemein versteht man eine vereinfachte Abbildung der Wirklichkeit, die benutzt werden kann, um Beobachtungen zu erklären und mögliche Veränderungen zu prognostizieren (MacKenzie et al., 2018). Unter anderem ist es damit auch möglich, Verbreitung und Vorkommen von Tieren oder anderen Lebewesen aufgrund gesammelter Daten vorherzusagen. Dafür ist das Zählen oder Kartieren dieser Arten zwingend notwendig. Dies wird schon lange praktiziert, sodass verschiedenste statistische Methoden entwickelt wurden, mit denen sich diverse Aspekte zum Vorkommen untersuchen lassen (MacKenzie et al., 2018). Doch egal welche Untersuchungsmethode zum Sammeln der Daten verwendet wird, bei einer Absenz der untersuchten Art wird man nie ganz sicher sein können, ob die Spezies in der Realität tatsächlich abwesend ist, oder ob sie schlicht und einfach nicht entdeckt werden konnte (Dunham & Rieman, 1999; Moilanen, 2002; Tyre et al., 2003; Engler et al., 2004). Solche «False Zeros» oder auch «False Negatives» sollten auf keinen Fall ignoriert werden, da die Anzahl besetzter Standorte ansonsten unterschätzt werden kann (Fiske & Chandler, 2011). Für viele heimlich lebende oder seltene Arten können oftmals nur Präsenz-/Absenzdaten generiert werden (Gu & Swihart, 2004), nicht aber Abundanzdaten, also Angaben zur Anzahl der Individuen. Um auf der Basis von Präsenz-/Absenzdaten trotzdem Angaben zu Verbreitung und Vorkommen zu erhalten, wurden sogenannte Site-Occupancy-Modelle entwickelt (Powell & Gale, 2015; Kéry & Royle, 2016). Wird lediglich eine einzelne Art in einer spezifischen Jahreszeit untersucht, so verwendet man eine spezielle Form dieser Modelle, nämlich ein Single-Species, Single-Season Site-Occupancy-Modell (MacKenzie et al., 2018). Dies ist auch in der vorliegenden Arbeit der Fall. Der sprachlichen Einfachkeit halber ist im folgenden lediglich von Site-Occupancy-Modell(en) die Rede.

Ein Site-Occupancy-Modell basiert auf der Annahme, dass die beobachteten Daten einer Untersuchung Ergebnisse von Zufallsvariablen sind (Kéry & Royle, 2016; MacKenzie et al., 2018). Eine Zufallsvariable weist jedem Ergebnis eines Zufallsversuchs eine reelle Zahl zu. Die Zufallsvariable X könnte beispielsweise die gewürfelte Augenzahl sein, mit den Zahlen 1 bis 6 als mögliche Werte. Im vorliegenden Fall des Site-Occupancy-Modelles wird die Zufallsvariable z_i eingeführt, welche den Vorkommensstatus der Spezies an einem Standort i beschreibt. Da z_i nur zwei Werte annehmen kann (0 für Standort unbesetzt und 1 für besetzt), handelt es sich um eine sogenannte Bernoulli-Zufallsvariable. Es lässt sich aber nur zu einer gewissen Wahrscheinlichkeit der tatsächliche Wert von z_i vorhersagen. Etwas formeller schreibt man:

$$z_i \sim \text{Bernoulli}(\psi)$$

Diese Formulierung besagt nichts anderes, als dass die Bernoulli-Zufallsvariable z_i mit der Wahrscheinlichkeit ψ (psi) den Wert 1 annimmt. Unter ψ versteht man die Occupancy (Vorkommenswahrscheinlichkeit), also die Wahrscheinlichkeit, mit der ein ausgewählter Ort innerhalb des untersuchten Gebiets von der entsprechenden Art besetzt ist, sprich mindestens ein Individuum dort vorkommt. Beim finalen Resultat des Untersuchungsdurchgangs j an Standort i handelt es sich auch um eine Bernoulli-Zufallsvariable, welche mit h_{ij} bezeichnet wird. Um diese formell beschreiben zu können, wird die Detection Probability (Nachweiswahrscheinlichkeit) p benötigt. Diese stellt die Wahrscheinlichkeit dar, dass die Art bei einer

Begehung entdeckt wird, *unter der Bedingung, dass sie anwesend ist.* (Dies ist sehr wichtig zum weiteren Verständnis der Grundlagen und im Endeffekt der ganzen Arbeit.) h_{ij} wird nun folgendermassen definiert:

$$h_{ij}|z_i \sim \text{Bernoulli}(z_i * p)$$

Man merke, dass an einem unbesetzten Standort $z_i = 0$ und somit auch $z_i * p = 0$ gilt, was den Fall $h_{ij} = 0$ erzwingt. Sprich, es ist unmöglich, an einem unbesetzten Standort die gesuchte Art nachzuweisen. Das Ergebnis von h_{ij} ist schliesslich auch im Feld beobachtbar. Mit genug Daten lassen sich dann aus den gesammelten Werten von h_{ij} die Wahrscheinlichkeiten ψ und p ableiten (MacKenzie et al., 2018). Dies geschieht mit einer geeigneten Software, welche allerdings nur in englischer Sprache verfügbar ist, weshalb im Folgenden der Einfachheit halber für ψ und p die Begriffe Occupancy und Detection Probability bevorzugt werden. Auch bei einigen weiteren Fachbegriffen wird der englische Begriff verwendet, weil keine gebräuchliche deutsche Übersetzung existiert.

2.2.1 Maximum-Likelihood-Methode

Die oben vorgestellten Wahrscheinlichkeitswerte können mittels der Maximum-Likelihood Methode ermittelt werden (MacKenzie et al., 2002). Diese liefert diejenigen Werte für p und ψ , welche aufgrund der gesammelten Daten am wahrscheinlichsten sind. Benötigt wird dafür eine Likelihood-Funktion, welche eine Wahrscheinlichkeitsverteilung der Parameter beschreibt. An der Stelle der gesuchten Werte für die Parameter muss die Funktion ihr Maximum aufweisen. Dieses wird, der Komplexität der Funktion geschuldet, meist numerisch ermittelt (MacKenzie et al., 2018).

Im Folgenden wird gezeigt, wie die Likelihood-Funktion für den gesamten Datensatz aufgestellt werden kann. Für ein Site-Occupancy-Modell werden positive Nachweise mit «1» und negative Nachweise mit «0» dokumentiert. Daraus resultiert für jeden Standort eine Detection-History, welche die Abfolge von beobachteter Präsenz und Absenz darstellt. Diese könnte beispielsweise für Standort i nach fünf Kontrollgängen 01101 sein. Die Wahrscheinlichkeit, exakt diese Detection-History zu erhalten, beträgt

$$\psi_i(1 - p_{i1})p_{i2}p_{i3}(1 - p_{i4})p_{i5}.$$

Diverse Gründe können dazu führen, dass zu manchen Zeiten an gewissen Orten keine Daten erhoben werden können. In diesem Modell führen fehlende Daten allerdings zu keinerlei Problemen, da z.B. die Wahrscheinlichkeit für Detection-History 0–10– genauso berechnet werden kann. Die fehlenden Durchgänge werden im Term einfach ignoriert, sodass

$$\psi_j(1 - p_{j1})p_{j3}(1 - p_{j4})$$

als Wahrscheinlichkeit resultiert. Wie müsste diese aber aussehen, wenn kein einziger positiver Nachweis erbracht werden konnte? Für dieses Resultat gibt es zwei mögliche Erklärungen. Entweder war die Art anwesend, wurde aber nicht entdeckt *oder* der Standort ist unbesetzt. Wie das «oder» bereits signalisiert, müssen die Wahrscheinlichkeiten der beiden Ereignisse addiert werden. Die Wahrscheinlichkeit, die Detection-History 00000 an Ort k zu erhalten, beträgt dementsprechend

$$\begin{aligned} & \psi_k(1 - p_{k1})(1 - p_{k2})(1 - p_{k3})(1 - p_{k4})(1 - p_{k5}) + (1 - \psi_k) \\ & = \psi_k \prod_{t=1}^5 (1 - p_{kt}) + (1 - \psi_k). \end{aligned}$$

Man beachte, dass Occupancy und Detection Probability hier als standort- und zeitspezifisch betrachtet werden. Daraus könnte man noch kein Modell anfertigen, da zu viele Parameter enthalten sind. Die Wahrscheinlichkeiten können aber abhängig von Covariablen modelliert werden (mehr dazu später). Deshalb werden die Formeln hier in der allgemeinen Form präsentiert. Unter der Annahme, dass Occupancy und Detection Probability zwischen den Standorten konstant und sämtliche Standorte unabhängig voneinander sind (vgl. Abschnitt 2.2.2) kann die Gesamtwahrscheinlichkeit und somit die Likelihood-Funktion für den Datensatz als Produkt aller Terme (für jeden Ort einen) ausgedrückt werden als

$$L(\psi, \mathbf{p}) = \left[\psi^{n \cdot} \prod_{t=1}^T p_t^{n_t} (1 - p_t)^{n - n_t} \right] \times \left[\psi \prod_{t=1}^T (1 - p_t) + (1 - \psi) \right]^{N - n \cdot}.$$

Dabei ist T die Anzahl Kontrollgänge, p_t die Detection Probability zu Zeit t , n_t die Anzahl der Standorte, an denen die Art zu Zeitpunkt t entdeckt wurde, $n \cdot$ die Anzahl Standorte mit mindestens einem positiven Nachweis, N die gesamte Anzahl untersuchter Standorte und \mathbf{p} steht für die Parameter $p_1 - p_T$ (MacKenzie et al., 2002).

Wie bereits angeschnitten, können ψ und p als Funktion von Covariablen (Habitatseigenschaften, Wetter, Tageszeit etc.) aufgefasst werden, welche zwischen Standorten und Kontrollgängen variieren können. Hierzu verwendet man eine *logit*-Funktion, welche den Vorteil hat, dass sie die Linearkombination aus den Werten der Covariablen, welche im Bereich $\pm\infty$ zu liegen kommen können, mit der Wahrscheinlichkeitsskala von 0 bis 1 verknüpft. Für einen beliebigen Wahrscheinlichkeitswert (z.B. ψ) wäre die *logit*-Funktion definiert als

$$\text{logit}(\psi_i) = \ln\left(\frac{\psi_i}{1 - \psi_i}\right) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_U x_{iU}.$$

Die Variablen $x_{i1} - x_{iU}$ stehen für die Werte der U verschiedenen Covariablen, die Variablen $\beta_1 - \beta_U$ für die dazugehörigen Koeffizienten (sind natürlich unbekannt), welche für die Gewichtung der entsprechenden Werte verantwortlich sind. β_0 ist der Achsenabschnittsterm, welcher den Funktionsgraphen nach links und rechts verschieben kann. Die obige Gleichung kann nach ψ_i umgeformt werden:

$$\psi_i = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_U x_{iU})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_U x_{iU})}$$

Dieser Ausdruck kann nun in die Likelihood-Funktion eingebaut werden. Die Koeffizienten lassen sich somit auch mittels der Maximum-Likelihood-Methode berechnen.

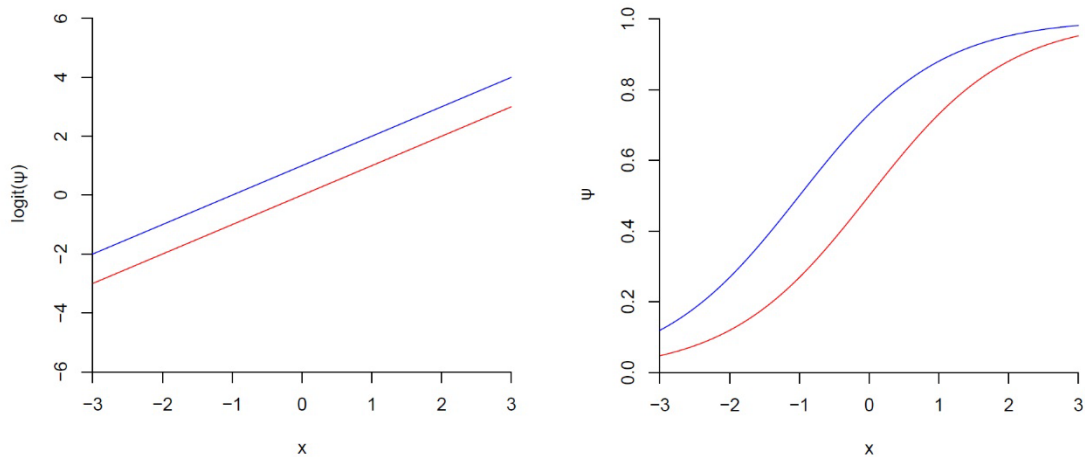


Abbildung 2: Veranschaulichung der logit-Funktion. Beziehungen, welche auf der logit-Skala linear sind (links), verlaufen auf der Wahrscheinlichkeits-Skala (rechts) nicht mehr linear und nehmen nur Werte zwischen 0 und 1 an.

2.2.2 Grundlegende Annahmen und Voraussetzungen

Ein Single-Season, Single-Species Site-Occupancy-Modell beruht auf einer Handvoll grundlegender Annahmen:

1. Der Besetzungsstatus eines Standortes verändert sich während der Untersuchungsdauer nicht. Ein besetzter Standort bleibt also besetzt und ein unbesetzter bleibt unbesetzt.
2. Occupancy und Detection Probability sind konstant oder jegliche Heterogenität ist als Funktion von Covariablen modelliert.
3. Sämtliche Datenpunkte sind unabhängig voneinander. Es gibt also sowohl örtlich als auch zeitlich keinerlei Abhängigkeiten.
4. Die Art wird stets korrekt bestimmt. Positive Nachweise beruhen nie auf einer Fehlbestimmung.

Die Verletzung dieser Grundannahmen kann zu verzerrten und fehlerhaften Werten führen. Solche Fehler sind aber je nach Art der Verletzung nicht unbedingt gravierend (MacKenzie et al., 2018). Die Einhaltung der Annahmen wird im Diskussionsteil genauer betrachtet.

2.2.3 Modellvergleich

Ein wichtiger Aspekt im Prozess des Modellierens ist die Suche nach dem aussagekräftigsten Modell. Eines, welches unsere Daten perfekt vorhersagt, mag zwar vielleicht auf den ersten Blick als geeignet erscheinen, produziert aber auf andere Daten angewandt erhebliche Fehler. Dieses Phänomen, welches man als «Overfitting» bezeichnet, gilt es zu verhindern. Mit steigender Präzision nimmt gleichzeitig auch die Varianz zu. Diese zwei Werte müssen also gegeneinander abgewogen und ein geeignetes Mittelmaß gefunden werden. Hierzu wird das Akaike-Informationskriterium (AIC) verwendet, welches definiert ist als

$$AIC = -2 \ln(L) + 2K.$$

L steht für den Wert der Likelihood-Funktion an der Stelle der errechneten Parameter und K für die Anzahl verwendeter Parameter im Modell (Powell & Gale, 2015). Wie man der Definition entnehmen kann, ist es erstrebenswert, Modelle mit möglichst niedrigem AIC zu kreieren. Der absolute Wert des AIC spielt dabei aber keine Rolle, da er mit wachsendem

Datensatz ebenfalls zunimmt. Relevant hingegen ist die Differenz eines Modells zum bestplatzierten aus dem Kandidaten-Set (MacKenzie et al., 2018). Für ein Modell k entspricht diese Differenz

$$\Delta AIC_k = AIC_k - AIC_{min}.$$

Als grobe Faustregel gilt, dass Modelle mit $\Delta AIC_k < 2$ annähernd gleiche Aussagekraft bieten, bei $4 < \Delta AIC_k < 7$ «deutlich weniger» und bei $\Delta AIC_k > 10$ «im Wesentlichen keine» (Burnham & Anderson, 2002). Es gibt aber Situationen, in denen das AIC-Gewicht w_k eine fairere Bewertung ermöglicht (MacKenzie et al., 2018). Dieses wird aus einem Set von R Modellen berechnet als

$$w_k = \frac{\exp(-0.5\Delta AIC_k)}{\sum_r^R \exp(-0.5\Delta AIC_r)}$$

und kann interpretiert werden als die Wahrscheinlichkeit, dass das Modell k das «beste» innerhalb des Sets ist (Burnham & Anderson, 2004). Dementsprechend gilt für die Summe der AIC-Gewichte $\sum_r^R w_r = 1$.

Doch auch mit den beschriebenen Klassifikationsmöglichkeiten ist es nicht immer sinnvoll, ein einziges Modell als «das beste» zu erküren, wenn mehrere ungefähr gleichwertige vorliegen. Um dieses Problem zu lösen, können die errechneten Parameterwerte mehrerer Modelle unter Berücksichtigung des AIC-Gewichts gemittelt werden. Als Beispiel wäre der gemittelte Wert für $\psi = \sum_{r=1}^R w_r \psi_r$ (MacKenzie et al., 2018).

2.2.4 Sensitivitätsanalyse

Wie im vorhergehenden Abschnitt erwähnt, ist das AIC kein absoluter Indikator, wie aussagekräftig ein Modell tatsächlich ist. Daher sollte jedes Modell stets auch auf mangelnde Anpasstheit, also ein «lack-of-fit», getestet werden (McCullagh & Nelder, 1989). Nur so kann gezeigt werden, dass ein Modell realistisch ist und die wichtigsten Merkmale des untersuchten Systems erfasst wurden. Erhebliche Mängel führen zu ungenauen Schlussfolgerungen oder können zu grosse Standardfehler begünstigen (MacKenzie & Bailey, 2004).

Für die Analyse werden die Modelle dem Goodness-of-Fit-Test von MacKenzie & Bailey (2004) unterzogen, welcher ein sogenanntes resampling-Verfahren verwendet (Dormann & Kühn, 2009). Dabei werden aus den Parameterwerten des Modells künstlich neue Datensätze generiert («bootstrap-Daten»), für welche ein Chi-Quadrat-Test durchgeführt wird. X^2 beschreibt die Abweichung der erwarteten Daten gegenüber den tatsächlich beobachteten und wird berechnet als

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}.$$

Dabei steht O_i für die Anzahl tatsächlicher Beobachtungen von Klasse i (hier unterschiedliche Detection-Histories) und E_i für die erwartete Anzahl davon. X^2 wird nun für eine möglichst grosse Anzahl von bootstrap-Daten errechnet und dann mit dem X^2 -Wert der Felddaten verglichen. Daraus resultiert der \hat{c} -Wert, welcher als das Verhältnis von X^2 der Felddaten zum Durchschnitt von X^2 der bootstrap-Daten definiert ist:

$$\hat{c} = \frac{X_{Obs}^2}{\bar{X}_B^2}$$

Der \hat{c} -Wert sollte etwa 1 betragen, weil für $\hat{c} > 1$ die Chi-Quadrat-Statistik überverteilt ist, was einen «lack-of-fit» anzeigen kann. Werte von 2-3 werden teilweise noch als genügend betrachtet (M. Kéry, schriftliche Mitteilung vom 27.09.2022). Zusätzlich wird der p -Value berechnet, welcher die Wahrscheinlichkeit angibt, dass $X_{Obs}^2 \geq \bar{X}_B^2$ (MacKenzie & Bailey, 2004). Für Modelle mit einem p -Value > 0.05 kann man davon ausgehen, dass sie aussagekräftig sind (Strimas-Mackey et al., 2020).

2.2.5 Anzahl Kontrollgänge und wahrscheinlicher Besetzungsstatus

Die Ansprüche an die Genauigkeit einer Untersuchung zum Vorkommen einer Art können stark variieren. Um mit einer Sicherheit von 95% ausschliessen zu können, dass anwesende Wasseramseln unentdeckt geblieben sind, ist selbstverständlich eine andere Anzahl Kontrollgänge nötig, als wenn man sich bereits mit 80% begnügen würde. Somit kann es von Interesse sein, für eine Untersuchung die notwendige Anzahl der Besuche pro Standort zu kennen. Mit gegebener Detection Probability p lässt sich leicht ausrechnen, wie viele Kontrollgänge tatsächlich erfordert werden, um eine gewünschte Gesamt-Nachweiswahrscheinlichkeit p_t zu erhalten (Powell & Gale, 2015). Dieser Wert ist gleichbedeutend mit der Sicherheit, anwesende Individuen mindestens einmal nachgewiesen zu haben. Er stellt somit die Gegenwahrscheinlichkeit zum Fall dar, in k Kontrollgängen keinen Nachweis trotz Präsenz erhalten zu haben.

$$p_t = 1 - (1 - p)^k$$

Diese Gleichung kann umgeformt werden, um k in Abhängigkeit von p und p_t zu berechnen.

$$k = \log_{1-p}(1 - p_t)$$

Des Weiteren lässt sich die Frage beantworten, wie wahrscheinlich es ist, dass ein Standort ohne Nachweise dennoch besetzt ist. Für die bedingte Wahrscheinlichkeit von Ereignis A unter der Voraussetzung, dass B eingetreten ist, gilt nach dem Satz von Bayes:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

Übertragen auf die obige Frage gilt mit Besetzungsstatus z eines Standorts (unbesetzt=0, besetzt=1) und Detection-History 00000:

$$\begin{aligned} P(z = 1 | h = 00000) &= \frac{P(h = 00000 | z = 1) * P(z = 1)}{P(h = 00000)} \\ &= \frac{(1 - p)^5 * \psi}{\psi(1 - p)^5 + (1 - \psi)} \end{aligned}$$

2.3 Decision-Tree-Modelle

Decision-Tree- oder Entscheidungsbaum-Algorithmen gehören zur Familie der Algorithmen des sogenannten «überwachten Lernens» und sind somit ein Teilgebiet des «maschinellen Lernens». Im Gegensatz zu anderen Algorithmen des «überwachten Lernens» können Decision-Trees auch zum Lösen von Klassifikationsproblemen genutzt werden (Chauhan, 2020). Der Algorithmus soll anhand von gesammelten Trainingsdaten einen Zusammenhang zwischen Daten und deren Klassifikation finden, um so ein Modell zu erstellen, welches möglichst präzise Vorhersagen zur Klassifikation von neuen Daten treffen kann (Anuradha & Gupta, 2014).

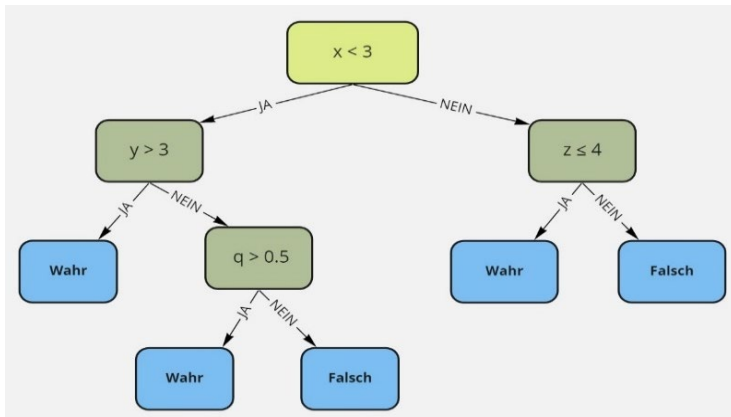


Abbildung 3: Simplex Beispiel eines Decision-Trees: Geprüft wird, ob eine Aussage auf Objekte mit den Attributen x, y, z, q zutrifft oder nicht. Hellgrün = Wurzelknoten, dunkelgrün = Entscheidungsknoten, blau = Blattknoten

die am Wurzelknoten in den Baum gespeist werden, folgen dem Pfad mit den zutreffenden Antworten auf die Entscheidungsfragen. Sobald der Pfad in einem Blattknoten endet, wird die Eingabe klassifiziert (Yang, 2019). Ein Nachteil der Decision-Tree-Methode ist allerdings, dass dem Input nur kategorische Klassifikationen zugeordnet werden können (Charbuty & Abdulazeez, 2021).

2.3.1 Bildung eines Decision-Trees

Decision-Trees sind keine vorgegebenen Strukturen. Sie müssen spezifisch aufgebaut werden. Wie bereits erwähnt, wird für die Bildung des Baumes ein Datensatz benötigt, welcher für alle Daten die beobachtete und somit korrekte Klassifikation beinhaltet. Die Problematik von «False Zeros», wie sie im Abschnitt 2.2 bereits erläutert wurde, wird bei Decision-Trees also ignoriert. Diese Methode sollte also nur angewendet werden, wenn «False Zeros» sehr unwahrscheinlich oder unmöglich sind. Mit dieser Annahme wird nun anhand der gesammelten Daten der Decision-Tree «trainiert», um die passendsten Kriterien für das Splitting an den Knoten festzulegen. Doch wie erfolgt dieser Trainingsprozess? Hierzu fokussiert der Algorithmus immer nur darauf, welcher Split der Daten zum jetzigen Zeitpunkt gerade als

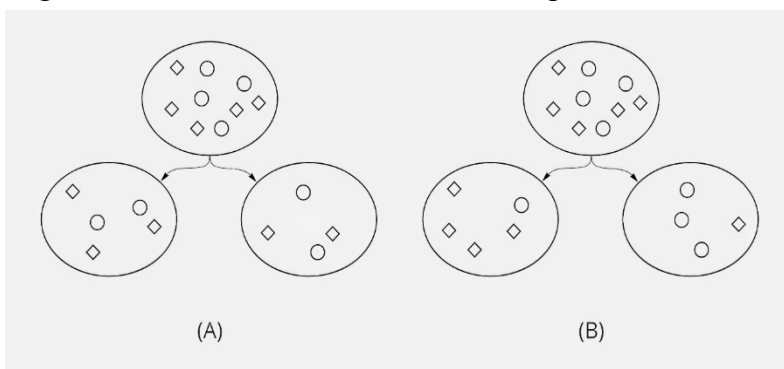


Abbildung 4: Beispiel für mögliche Teilungen des Datensatzes: Split B liefert eine höhere Homogenität der Klassifikationen (Kreis und Quadrat) in den Subsets als Split A.

der beste erscheint. Ein Split wird immer dadurch erstellt, dass der Datensatz aufgrund von Eigenschaften der Daten aufgeteilt wird. Es werden sämtliche mögliche Splits getestet, wobei derjenige schlussendlich ausgewählt wird, welcher in den entstandenen Daten-Untergruppen, den Subsets, eine möglichst grosse Homogenität der Klassifizierungen hervorbringt. Für alle Subsets wird dieses Verfahren wiederholt, bis der Decision-Tree nicht mehr weiterwachsen kann oder mögliche Kriterien zum Anhalten erreicht sind (siehe Abschnitt 2.3.1) (Chauhan, 2020).

Ein Decision-Tree ist so aufgebaut, dass vom Wurzelknoten, dem Startpunkt des Baumes, Pfade wegführen. Anhand von Entscheidungskriterien werden diese Pfade an Entscheidungsknoten aufgeteilt. Dieser Vorgang wird als Splitting bezeichnet. Jede Abzweigung entspricht dabei einer möglichen Antwort auf das abgefragte Kriterium. Diese Antworten führen wiederum zu neuen Knoten, sodass ein rekursiver Prozess entsteht (Chauhan, 2020). Eingabedaten,

der beste erscheint. Ein Split wird immer dadurch erstellt, dass der Datensatz aufgrund von Eigenschaften der Daten aufgeteilt wird. Es werden sämtliche mögliche Splits getestet, wobei derjenige schlussendlich ausgewählt wird, welcher in den entstandenen Daten-Untergruppen, den Subsets, eine möglichst grosse Homogenität der Klassifizierungen hervorbringt.

Es existieren diverse Algorithmen zur Erstellung von Decision-Trees, welche sich verschiedener Kriterien bedienen, mit welchen die Homogenität quantifiziert werden kann (Charbuty & Abdulazeez, 2021):

1. Die *Entropie* ist ein Mass für die Unordnung oder Zufälligkeit eines Datensatzes. Ihr Wert liegt immer zwischen 0 und 1, wobei maximale Ordnung die Entropie 0 aufweist, während der Wert bei perfekter Unordnung 1 beträgt. Zur Berechnung der Entropie werden die Wahrscheinlichkeiten benötigt, mit welchen ein Datenpunkt zur Klasse i gehörig ist. Diese Werte werden mit $\log_2()$ von sich selbst multipliziert und schlussendlich über alle Eigenschaften aufsummiert.

$$E(X) = - \sum p_i \log_2(p_i)$$

Damit keine negativen Werte für die Entropie entstehen, wird die ganze Summe noch mit -1 multipliziert (Rekha, 2020).

2. Der *Information Gain* («Informationsgewinn») berechnet die Differenz der Entropie vor dem Split mit derjenigen nach dem Split (Chauhan, 2020). Somit könnte er also auch als Reduktion der Entropie betitelt werden (Silipo, 2019). Daraus folgt, dass der beste Split logischerweise den höchsten Information Gain hervorbringt. Wenn ein Datenset X in K verschiedene, kleinere Subsets X_k geteilt wird, beträgt der Informationsgewinn

$$IG = E(X) - \sum_{k=1}^K E(X_k).$$

Der Information Gain bevorzugt tendenziell kleinere Datenmengen innerhalb der Subsets und neigt dazu, Attribute mit einer grossen Anzahl Werte zu priorisieren. Ein weiterer Nachteil ist, dass der Test nur für kategorische Variablen funktioniert (Chauhan, 2020).

3. Um die Verzerrung des Information Gains zu korrigieren, wird der *Gain Ratio* («Gewinnverhältnis») eingeführt. Dieser berücksichtigt zusätzlich eine sogenannte Split-Information, durch welche der Informationsgewinn geteilt wird.

$$Gain\ Ratio = \frac{IG}{\sum_{k=1}^K w_k \log_2(w_k)}$$

Um die Split-Information zu generieren, wird das Verhältnis w_k der Anzahl Daten im Subset k zur Anzahl Daten im vorherigen Datenset verwendet (Silipo, 2019).

4. Der *Gini-Index* favorisiert im Gegensatz zum Informationsgewinn eher grössere Subsets. Für K verschiedene davon werden verrechnete Wahrscheinlichkeitswerte aufsummiert:

$$Gini\ Index = \sum_{k=1}^K w_k (1 - \sum p_i)$$

Hier taucht der Ausdruck w_k wieder auf, welcher verantwortlich ist für die Gewichtung des darauffolgenden Terms. Dieser bildet die Differenz aus 1 minus der Summe der Wahrscheinlichkeiten, dass ein Objekt zur Klasse i gehört. Für geeignete Splits sollte der Gini-Index kleine Werte erreichen (Silipo, 2019). Diese Testvariante ist auf kontinuierliche Variablen ausgelegt (Chauhan, 2020).

5. Der *Chi-Quadrat-Test* wurde in dieser Arbeit bereits bei der Sensitivitätsanalyse von Site-Occupancy-Modellen eingeführt. Zur Erinnerung: Die Formel zur Berechnung von X^2 lautet

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i}.$$

Getestet wird die Signifikanz des Zusammenhangs zwischen einem Kriterium und der Klassifikation. Für jede mögliche Antwort wird die Anzahl gleicher Klassifikationen ausgezählt, was dem beobachteten Wert entspricht. Der erwartete Wert hingegen kommt aus dem Durchschnitt der Anzahlen jeder Klassifikationsmöglichkeit zustande.

Tabelle 1: Beispielswerte für ein zu testendes Kriterium mit Antworten «Hoch», «Mittel» und «Tief» bei möglichen Klassifikationen in «Wahr» und «Falsch»

	Anzahl «Wahr»	Anzahl «Falsch»	Erwarteter Wert
Hoch	2	3	$(2+3)/2 = 2.5$
Mittel	5	2	$(5+2)/2 = 3.5$
Tief	1	5	$(1+5)/2 = 3$

Für jeden der sechs beobachteten Werte wird nun X^2 berechnet und schlussendlich aufsummiert. Je höher der Wert von X^2 , desto mehr steigt die Signifikanz des Zusammenhangs zwischen Kriterium und Klassifikation. Die Beschränkung auf kategoriale Variablen stellt allerdings einen Nachteil dieses Tests dar (Serengil, 2020).

2.3.2 Overfitting

Eine Überanpassung oder ein Overfitting von Decision-Trees kann relativ leicht erreicht werden, wenn der Baum zu lange wächst und ihm keine Limiten gesetzt werden. Theoretisch könnte das Modell perfekt die Trainingsdaten vorhersagen, wenn für alle Daten ein eigener Blattknoten existiert (Chauhan, 2020). Zu grosse Bäume sind allerdings zu detaillierte Modelle, die nicht auf neue Daten angewendet werden können. Es gibt zwei Methoden, welche eine Überanpassung verhindern sollen (Silipo, 2019):

1. Das *Pruning* ermöglicht es, bereits kreierte Decision-Trees wieder zu verallgemeinern. Dazu werden allerdings zusätzlich zu den Trainingsdaten auch Testdaten benötigt. Diese können erhalten werden, wenn der gesamte Datensatz in zwei kleinere Partitionen aufgeteilt wird. Der Decision-Tree, welcher anhand der Trainingsdaten erstellt wurde, wird nun auf die Testdaten angewendet. Nun werden Entscheidungsknoten darauf getestet, ob sie durch Blattknoten ersetzt werden können, ohne dabei die Anzahl von Missklassifikationen zu erhöhen (Chauhan, 2020).
2. Ein *Früh-Stopp* oder *Pre-Pruning* sorgt dafür, dass ein Overfitting gar nicht erst erreicht wird. Ein mögliches Kriterium für das Anhalten des Baumwachstums wäre eine minimale Anzahl von Datenpunkten innerhalb eines Subsets, welches beim Erstellen des Decision-Trees generiert wird. Wird diese Grenze erreicht, werden keine weiteren Splittings mehr vorgenommen (Silipo, 2019). Eine maximale Anzahl an Entscheidungsknoten kann ebenfalls als ein Kriterium verwendet werden, welches die Grösse des Decision-Trees beschränkt (Kumar, 2021).

3. Methodik

3.1 Überblick

Mit einer systematischen Kartierung wurden an definierten Beobachtungsstandorten Präsenz-/Absenzdaten von Wasseramseln erfasst. Für die Fragestellungen relevante Umweltdaten wurden zeitgleich aufgenommen. Aus den gesammelten Daten konnten danach verschiedene Site-Occupancy-Modelle kreiert werden, welche die beobachteten Felddaten möglichst gut erklären sollen. Diese wurden verglichen und auf ihre Tauglichkeit geprüft, sodass schlussendlich von den untersuchten Umweltfaktoren, welche Occupancy und Detection Probability möglicherweise beeinflussen, die relevantesten ermittelt werden können. Anhand der Daten zu Bachstrukturen wurde zudem versucht, einen Decision-Tree zu erstellen.

3.2 Datenaufnahme

Auf einer Strecke von rund elf Kilometern wurden zwischen Oberuster und Bäretswil im Zeitraum 8. Februar bis 8. Juni 2022 am Aa- und Chämptnerbach systematisch Wasseramseln gesucht und kartiert. Im Vorfeld wurden in regelmässigen Abständen von 200 Metern Beobachtungsstandorte definiert. Weil vier dieser "auf dem Papier" festgelegten Standorte nicht zugänglich waren, wurden sie aus der Kartierung ausgeschlossen, sodass schlussendlich 58 Beobachtungsorte resultierten. Aus ausschliesslich organisatorischen Gründen wurden diese in fünf Sektoren A – E aufgeteilt.

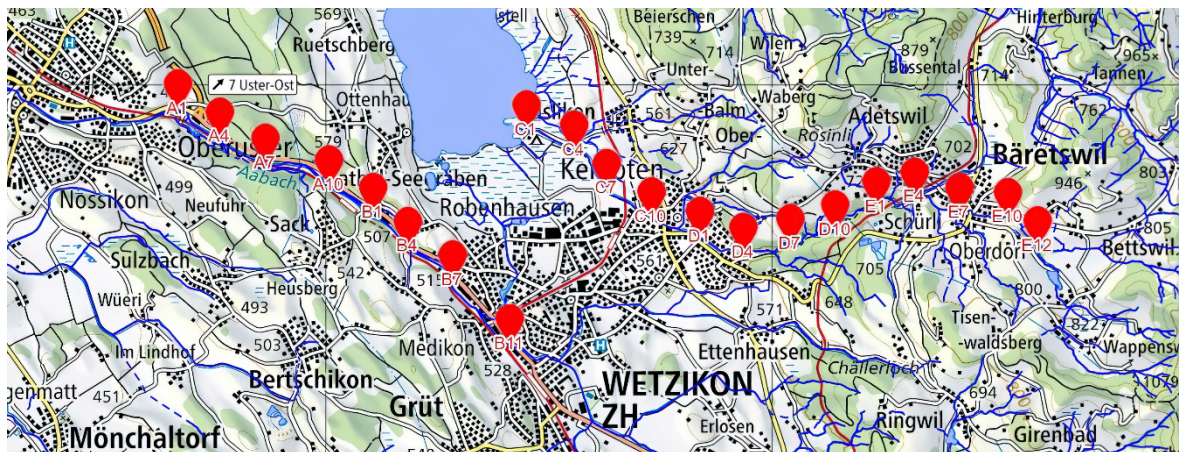


Abbildung 5: Karte der Beobachtungsstandorte. Der Übersichtlichkeit geschuldet sind nicht alle Standorte markiert. Ein Link zu einer Karte mit allen Standorten ist im Anhang zu finden.

Jeder Sektor und somit auch jeder Standort wurde fünfmal begangen, je im Abstand von drei bis vier Wochen und abwechslungsweise bachauf- und abwärts. Diese Anzahl Beobachtungen pro Standort war einerseits gerade noch zeitlich bewältigbar, andererseits versprach sie eine ausreichende Datenbasis, um solide Modellierungen durchzuführen (MacKenzie et al., 2018). Das zeitliche Beobachtungsfenster pro Kontrollgang orientierte sich an der Aktivitätszeit der Wasseramsel. Die grösste Aktivität ist in den Stunden nach Sonnenaufgang und vor Sonnenuntergang zu verzeichnen (Glutz von Blotzheim & Bauer, 1985). Dabei scheinen die Morgenstunden für Kartierungen am ergiebigsten zu sein (Hatt, 2022). Somit galt als Faustregel, dass der Beobachtungsstart ungefähr zur Zeit des Sonnenaufgangs sein sollte.

Im Folgenden wird kurz der Ablauf einer Begehung beschrieben. An jedem Standort wurden während exakt zehn Minuten sämtliche Beobachtungen von Wasseramseln erfasst. Als Nachweis zählten sicher bestimmte visuelle und akustische Beobachtungen. Sehr häufig

entstand eine Kombination, da sich fliegende Wasseramseln oftmals mit ihrem Ruf ankündigten und dadurch leichter gesehen werden konnten. Ein permanenter Blick auf den Bach war unerlässlich, um stumme, schnell vorbeifliegende Vögel nicht zu verpassen. Die Strecken zwischen den Standorten wurden entweder zu Fuss oder per Fahrrad zurückgelegt. Die typische Ausrüstung setzte sich zusammen aus einem Feldstecher des Typs Nikon Monarch und einem Klemmbrett mit Protokollblatt und Karte sowie einem Stift. Für jeden Standort wurde zusätzlich auch der Beobachtungszeitpunkt festgehalten. Bei der letzten Begehung wurden bei jedem Standort die Bachstrukturen mittels eines Kriterienrasters erfasst (siehe Abbildung 6). Dieses wurde von Hatt (2022) übernommen und leicht angepasst, sodass nur noch zehn anstatt elf Kriterien beachtet wurden (der Aspekt "Freizeitnutzung" wurde nicht erfasst).

1	Variabilität der Bachbreite	Stark abwechselnd schmal und breiter	3	Begradigt, abwechselnd etwas schmaler/breiter werdend	2	Kanalisierte Einheitsbreite	1		
2	Absolute Wasserbreite	>2m	2	<2m			1		
3	Nebenbäche?	Ja	2	Nein			1		
4	Variabilität der Wassertiefe	Stark wechselnd	3	Etwas variierend	2	Einheitlich	1		
5	Stärke der Turbulenzen	Starke Turbulenzen, ebenso schnell fließende wie fast stehende Abschnitte	3	Turbulenzen (z.B. an Schwellen), unterschiedliche Durchflussschwindigkeiten	2	Keine Turbulenzen, einheitliche Fließgeschwindigkeit	1		
6	Steine als Sitzwarten	Viele, >5 im Sichtfeld	3	Einzelne, 1-5 im Sichtfeld	2	Keine	1		
7	Art der Bachsohle	Vor allem Kies & Steine	4	Vor allem Sand & Schlack	3	Vor allem organisch (Laub & Pflanzen)	2	Versiegelt, verbaut oder versintert	1
8	Naturnähe des Ufers	Natürlich, reich strukturiert, oft unterspült	3	Mit Blöcken verbaut, begradigt, Strukturen vorhanden, manchmal unterspült	2	Völlig einheitlich & gleichförmig, z.B. monotone Grasböschung	1		
9	Stärke des Uferbewuchses	Durchgehende Deckung durch Bäume & Sträucher gegeben	3	Einzelne Sträucher o.ä. als Deckung vorhanden	2	Keine Deckung	1		
10	Entfernung der Uferlinie zum Weg	Größerer Abstand $\geq 4m$	3	Abstand von wenigen Metern, ca. 2-3m	2	Weg direkt neben Bach	1		

Abbildung 6: Kriterienraster zum Erfassen der Bachstrukturen

3.3 Site-Occupancy-Modelle

Für die Modellierungen mit Site-Occupancy-Modellen wurde das Programm *R* (Version 4.1.2) verwendet, welches das Programmieren in der gleichnamigen Programmiersprache ermöglicht. Die notwendigen R-Skripts liessen sich mit *RStudio* (Version 2022.07.1) erstellen und ausführen. (Ein Link zu allen R-Skripts ist im Anhang zu finden.) Um die passenden Funktionen nutzen zu können, wurde das R-Package *unmarked* (Version 1.1.1) benötigt. Dieses bietet eine ganze Palette an Funktionen für statistische Analysen von Daten, welche aus Untersuchungen mit unmarkierten Tieren stammen. So können für Spezies, welche nicht mit perfekter Sicherheit nachgewiesen werden können, Occupancy, Abundanz und Dichte geschätzt werden (Fiske & Chandler, 2011). In *unmarked* sind verschiedene Modelle implementiert, wie beispielsweise das in dieser Arbeit verwendete Single-Species, Single-Season Site-Occupancy-Modell nach MacKenzie et al. (2002).

3.3.1 Grundlegendes Modell

In einem ersten Schritt wurde ein Modell erstellt, welches für Occupancy und Detection Probability je nur einen Wert errechnet. Somit wird angenommen, dass die entsprechenden Wahrscheinlichkeiten an allen Standorten und bei allen Begehungen gleich sind. In *R* müssen als erstes die benötigten Packages geladen werden. Am wichtigsten für die Modellierung ist das Package *unmarked*. Um Excel-Dateien laden zu können, wird aber auch *readxl*

gebraucht. Als Nächstes werden die Daten geladen. Diese müssen in der richtigen Struktur vorliegen, sodass jede Zeile einem Standort und jede Spalte einem Kontrollgang entspricht. Die Daten aus den Feldunterlagen wurden bereits in genau dieser Form in eine Excel-Datei eingetragen, um eine spätere Umsortierung zu umgehen.

```
library(unmarked)
library(readxl)

detections <- read_excel("C:/.../DetectionHistory.xlsx")
```

Abbildung 7: Startsequenz im R-Skript für das grundlegende Modell.

Nun wird ein *unmarkedFrame*-Objekt erstellt, welches von *unmarked* genutzt wird, um alle Daten zu organisieren. Für die gewählte Modellierung spezifisch wird dazu die Funktion *unmarkedFrameOccu()* verwendet. Um das Modell anschliessend zu generieren, wird die Funktion *occu()* aufgerufen. Da hier weder für Detection Probability noch für Occupancy Covariablen verwendet werden, muss zweimal das Argument «~1» angegeben werden.

```
umf <- unmarkedFrameOccu(y=detections)
basic <- occu(~1 ~1, data = umf)
```

Abbildung 8: Erstellen des *unmarkedFrame*-Objektes und Aufrufen der *occu*-Funktion.

Aus diesen Schritten resultieren die Werte für p und ψ auf der *logit*-Skala. Da es sich um Wahrscheinlichkeiten handelt, müssen diese Werte aber auf eine Skala von 0 bis 1 transformiert werden. Dies geschieht mit der Funktion *predict()*. Das Resultat ist der Schätzwert mit dazugehörigem Standardfehler und 95%-Vertrauensintervall.

```
basic.psi <- predict(basic, newdata=data.frame(site=1), type="state")
basic.p <- predict(basic, newdata=data.frame(site=1), type="det")
```

Abbildung 9: Rücktransformierung der errechneten p - und ψ -Werte mit der *predict*-Funktion.

Die Ergebnisse wurden mit der «Naïve Occupancy» beziehungsweise der «Naïve Detection Probability» verglichen. Diese Werte beschreiben Occupancy und Detection Probability, würden «False Zeros» ignoriert werden.

3.3.2 Modellierung mit Covariablen

Man muss davon ausgehen, dass die Occupancy zwischen Standorten und die Detection Probability zusätzlich auch zwischen Kontrollgängen variieren können (MacKenzie et al., 2018). Für ein Modell, welches die Daten möglichst gut erklären soll, muss dies miteinbezogen werden. Zuerst wurden dabei die Covariablen für die Detection Probability berücksichtigt und erst später diejenigen für die Occupancy.

Es wurde angenommen, dass die Nachweiswahrscheinlichkeit mit der Anzahl Minuten nach Sonnenaufgang abnimmt (vgl. Abschnitt 3.2). Zudem wurde vermutet, dass auch die Jahreszeit einen Einfluss auf die Nachweiswahrscheinlichkeit haben könnte. Deshalb wurde jedem Kontrollgang eine kategorische Covariable zugewiesen, je nachdem, in welchem Monat die Begehung stattfand. Als Vergleich wurde zudem eine kontinuierliche Covariable verwendet, mit welcher es in R allerdings nicht möglich ist, Schwankungen je nach Monat darstellen zu

können, sondern nur Anstiege oder Abfälle über den gesamten Beobachtungszeitraum hinweg.

Da diese Covariablen für jeden Standort und Kontrollgang verschieden sein können, handelt es sich um Observation-Level Covariablen, was in *R* mit «obsCovs» abgekürzt wird. Um diese ins *unmarkedFrameOccu*-Objekt laden zu können, müssen sie Teil eines *list*-Objekts sein, welches theoretisch auch nur eine einzige Observation-Level-Covariable beinhalten könnte. Erst in der Funktion *occu()* wird angegeben, welche Covariablen für das Modell genutzt werden. Da diese für die Detection Probability verwendet werden, müssen sie in der Formel an erster Stelle stehen, während an zweiter Stelle mit «~1» festgelegt wird, dass die Occupancy konstant bleiben soll.

```
minuten_cov <- read_excel("C:/.../MinutenCovs.xlsx")
monat_cov <- read_excel("C:/.../MonatCovs.xlsx")
datum_cov <- read_excel("C:/.../DatumCovs.xlsx")

obs_covs <- list(minuten=minuten_cov, monat=monat_cov, datum=datum_cov)
umf <- unmarkedFrameOccu(y = detections, obsCovs = obs_covs)

det_minuten <- occu(~minuten ~1, data = umf)
det_monat <- occu(~monat ~1, data = umf)
det_datum <- occu(~datum ~1, data = umf)
det_min_dat <- occu(~minuten+datum ~1, data = umf)
det_min_mon <- occu(~minuten+monat ~1, data = umf)
```

Abbildung 10: Erstellung verschiedener Modelle mit variierender Detection Probability

Es wurden Modelle mit je einer einzelnen Covariable sowie Kombinationen aus mehreren erstellt und mit den Funktionen *fitList()* und *modSel()* nach ihrem AIC geordnet.

```
fitList <- fitList("p(.)psi(.)" = basic,
                 "p(minuten)" = det_minuten,
                 "p(monat)" = det_monat,
                 "p(datum)" = det_datum,
                 "p(minuten+datum)" = det_min_dat,
                 "p(minuten+monat)" = det_min_mon)
modSel <- modSel(fitList)
```

Abbildung 11: Vergleich der Modelle untereinander anhand des AIC.

Aus allen Modellen mit $\Delta AIC < 2$ werden nun Durchschnittswerte für die Parameter berechnet. Dazu wird die Funktion *modavgPred()* des Packages *AICcmodavg* genutzt. Als Argumente dürfen eine Liste der besten Modelle und die verwendeten Covariablen nicht fehlen.

```
library(AICcmodavg)
det_best_list <- list(det1 = det_first,
                    det2 = det_second,
                    ...
                    )
det_p_predict <- modavgPred(det_best_list,
                          parm.type = "detect",
                          newdata = data.frame(predict_cov))
```

Abbildung 12: Berechnung der Durchschnittswerte der Parameter derjenigen Modelle, bei denen $\Delta AIC < 2$.

Für die Occupancy pro Standort wurde angenommen, dass sie von der Anzahl besetzter Nachbarstandorte abhängig sein könnte. Jedem Beobachtungsort wurde also ein Wert zwischen 0 und 2 zugeordnet. Da diese Covariablen nur zwischen Standorten variieren können, handelt es sich um «siteCovs». In der Formel der Funktion `occu()` wird nun die Detection Probability mit «~1» konstant gehalten, während für die Occupancy die Nachbarstandort-Daten verwendet werden. Die Bezeichnung in der Formel muss identisch mit derjenigen aus der Excel-Datei sein.

```
umf <- unmarkedFrameOccu(y = detections, siteCovs = daten_nachbar)
occu_nachbar <- occu(~1 ~Nachbar, data = umf)
```

Abbildung 13: Berücksichtigung der Anzahl besetzter Nachbarstandorte für die Berechnung der Occupancy der Standorte.

Als Covariablen für die Occupancy wurden aber auch die Bachstrukturen herbeigezogen. Bachstrukturen, welche an (fast) allen Standorten gleich bewertet wurden, flossen nicht in die Modellierung ein. Wie bei der Detection Probability wurde für jede Kombination der Covariablen ein Modell erstellt. Da mit steigender Anzahl Covariablen aber sehr viele verschiedene Kombinationsmöglichkeiten entstehen, erspart man sich die manuelle Arbeit. Viel einfacher ist es, die Funktion `dredge()` des Pakets *MuMIn* zu verwenden. Diese errechnet für alle möglichen Modelle das AIC und sortiert sie entsprechend diesem, sodass man auf den ersten Blick erkennen kann, welches die aussagekräftigsten Modelle sind.

```
umf <- unmarkedFrameOccu(y=detections, siteCovs = daten_bachstruktur)
occu_habitat <- occu(~1 ~Bachsohle + Steine + ... + Ufer, data = umf)
occu_habitat_dredge <- dredge(global.model = occu_habitat,
                             rank = "AIC")
```

Abbildung 14: Prüfung der Relevanz verschiedener Bachstrukturen für die Occupancy, unter Verwendung der Funktion `dredge()`.

Die nun erhaltenen Resultate wurden auf ihre biologische Sinnhaftigkeit überprüft. Biologisch sinnlose Befunde, z.B. ein negativer Einfluss von Ansitz-Steinen auf das Vorkommen der Wasseramseln, führten zum manuellen Ausschluss der entsprechenden Covariablen. Mit den restlichen wurde im Anschluss daran der ganze Prozess wiederholt, was eine Selektion des aussagekräftigsten Modells, welches gleichzeitig biologische Sinnhaftigkeit gewährt, ermöglichte.

3.3.3 Sensitivitätsanalyse

Als Sensitivitätsanalyse wurde der Goodness-of-Fit-Test von MacKenzie & Bailey (2004) verwendet, um die Aussagekraft der Modelle zu überprüfen. Die Funktion `mb.gof.test()` aus dem Package *AICcmodavg* führt diesen in R aus. Als Anzahl der generierten bootstrap-Daten wurde der empfohlene Wert von 1000 (MacKenzie & Bailey, 2004) verwendet. Der Test wurde auf alle Modelle angewendet, welche in ihrem Kandidatenset am besten eingestuft wurden. Modelle, welche zu einem Durchschnitts-Modell zusammengefasst wurden, mussten einzeln getestet werden, da die Funktion `mb.gof.test()` nicht auf gemittelte Modelle angewendet werden kann.

```
model_test <- mb.gof.test(model, nsim = 1000)
```

Abbildung 15: Aufrufen der für die Sensitivitätsanalyse notwendigen Funktion `mb.gof.test()`.

3.3.4 Anzahl Kontrollgänge und wahrscheinlicher Besetzungsstatus

Mittels der im Grundlagenteil eingeführten Formel wurde analysiert, wie gross die Gesamtnachweiswahrscheinlichkeit nach den fünf vollbrachten Kontrollgängen tatsächlich ist. Zudem liess sich errechnen, wie viele Beobachtungsgänge für gewünschte Gesamtnachweiswahrscheinlichkeiten nötig sind.

Aus den Werten von Occupancy und Detection Probability konnte durch den Satz von Bayes auch die Wahrscheinlichkeit ermittelt werden, dass ein Standort besetzt ist, obwohl zu keiner Zeit eine Wasseramsel dort nachgewiesen werden konnte.

3.4 Decision-Tree-Modell

Zum Erstellen eines Decision-Trees wurde das Programm *RapidMiner Studio* (Version 9.10.011) verwendet. Diese Applikation kommt ganz ohne programmierbaren Code aus, sondern beruht auf dem Konzept, dass die verwendeten Daten und Operatoren wie Bausteine ins Prozessfenster gesetzt und dann mit Linien passende Verknüpfungen gezogen werden.

Der Decision-Tree-Operator wurde allerdings nicht sofort auf die Daten angewendet. Mit einer «Cross-Validation» wurde nämlich der Datensatz zuerst in Trainings- und Testdaten aufgeteilt. Der Trainingsteil der Daten konnte dann mit dem Decision-Tree-Operator verknüpft werden. Als Einstellungen für den Decision-Tree-Operator wurde das Splitting-Kriterium auf *Information Gain* gesetzt und sowohl Pruning als auch Pre-Pruning aktiviert. Die Testdaten wurden mit dem Apply-Model-Operator verbunden, sodass der erstellte Decision-Tree direkt getestet werden kann. Schliesslich wurde noch eine Performance-Messung angehängt, um eine Quantifizierung der Aussagekraft des Entscheidungsbaumes mittels dem «accuracy»-Wert vorzunehmen. Dieser steht für den prozentualen Anteil korrekter Klassifikationen (Chugh, 2022).

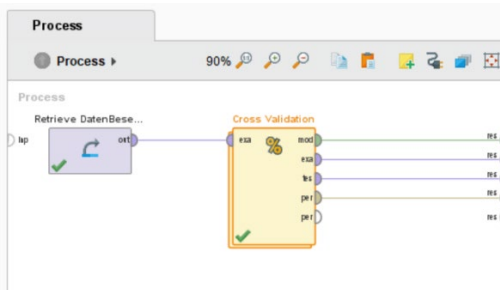


Abbildung 16: Prozessfenster von RapidMiner Studio zur Verarbeitung der Daten mit dem Cross-Validation-Operator (siehe Abbildung 17)

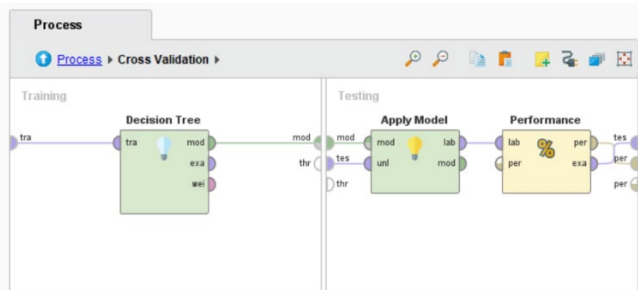


Abbildung 17: Operatoren innerhalb des Cross-Validation-Operators zum Erstellen eines Decision-Trees

Analog zur Modellierung der Site-Occupancy-Modelle wurden biologisch unsinnige Bachstrukturen ausgeschlossen und die Bildung des Decision-Trees mit den übrig gebliebenen Daten wiederholt. Nur der so entstandene Entscheidungsbaum wurde weiter verwendet.

4. Resultate

4.1 Überblick über die gesammelten Präsenz-/Absenzdaten

Nach fünf Kontrollgängen konnte von den 58 Standorten an 37 ein Wasseramselvorkommen verzeichnet werden. Meist gelang nur bei 1-3 Beobachtungsgängen auch tatsächlich ein positiver Nachweis. An sechs Orten wurden viermal Wasseramseln beobachtet, während eine Präsenz bei jedem Besuch des Standorts nirgends festgestellt werden konnte. Im untersten Teil des Aabachs in Oberuster gelangen einige Nachweise von Wasseramseln. Bachaufwärts wurden auf einem Abschnitt von etwa einem Kilometer keine Beobachtungen gemacht. Hinter dem Saurier-Museum konnten allerdings bei vier Gelegenheiten Wasseramseln dokumentiert werden. Bis nach Wetzikon folgt eine längere Strecke mit fast durchgehenden Nachweisen. In diesen wird aber kaum ein Muster ersichtlich, welches klare Reviergrenzen erkennen liesse. Entlang des teils kaum zugänglichen Aabachs gelang also doch eine grosse Zahl an Beobachtungen. Im gesamten Ried und weit nach Kempten hinein konnten bis auf wenige Ausnahmen Wasseramseln dokumentiert werden. Im Tobel existieren einige Beobachtungspunkte, an welchen recht zuverlässig Nachweise gelangen. Da der Bach hier gut einsehbar war, konnten während der Datenaufnahme sogar an vier Stellen Nistplätze gefunden werden. Insgesamt war fast das ganze Chämptnertobel von Wasseramseln besetzt. Erst an den obersten Standorten des Untersuchungsgebiets wurden Beobachtungen der Tiere wieder spärlicher. Alle gesammelten Präsenz-/Absenzdaten sind im Anhang angefügt.



Abbildung 18: Karte der Beobachtungsstandorte mit Anzahl positiver Nachweise pro Standort

4.2 Site-Occupancy-Modelle

4.2.1 Grundlegendes Modell

Mit dem grundlegenden Modell wurde für Occupancy und Detection Probability je nur ein einziger Wert errechnet. Dabei handelt es sich um Durchschnittswerte aller Standorte beziehungsweise Kontrollgänge. Die Occupancy ψ beträgt 69%, wobei die «Naïve Occupancy» bei 64% liegt. Der Wert der Detection Probability p liegt bei 43% mit «Naïve»-Wert 46%.

Dieses grundlegende Modell wurde in der Folge verfeinert, indem es mit Covariablen ergänzt wurde. Deren Einfluss auf die Detection Probability beziehungsweise die Occupancy konnte durch einen Modellvergleich mittels AIC getestet werden.

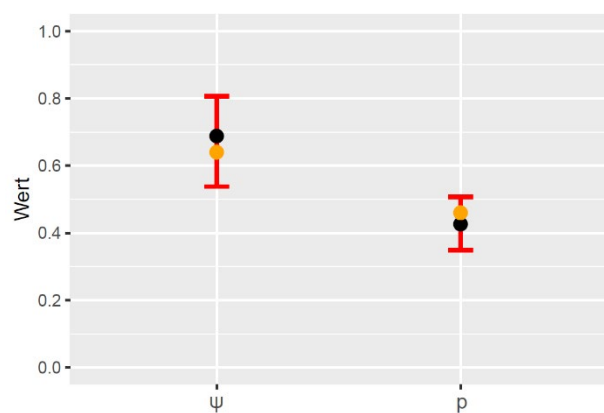


Abbildung 19: Errechnete Werte für die Occupancy ψ und die Detection Probability p gemäss dem grundlegenden Modell (schwarz). Das 95%-Vertrauensintervall wird durch die roten Balken gekennzeichnet. Die orangen Punkte markieren die «Naïve»-Werte.

4.2.2 Modellierung mit zeitlichen Covariablen

Sowohl die Tageszeit als auch die Jahreszeit können einen Einfluss darauf haben, wie aktiv Wasseramseln sind und wie gut sie dementsprechend entdeckt werden können. Der Effekt dieser Covariablen wurde in mehreren Modellen aufgeschlüsselt. Das Modell, welches die Minuten nach Sonnenaufgang sowie den Monat der Begehung berücksichtigt, schnitt mit einem AIC von 318.5 im Kandidatenset am besten ab. Mit nur 0.75 AIC-Punkten mehr folgt in der Tabelle das Modell, welches nur die Covariable der Minuten enthält. Deutlich schlechter schneiden die beiden Modelle ab, welche nur die Jahreszeit als Covariable verwenden. Das grundlegende Modell bildet (als Vergleich) das Schlusslicht.

Tabelle 2: Vergleich verschiedener Modelle mit Covariablen zur Detection Probability. Sie sind geordnet nach ihrem AIC. Als Referenzgrösse wird auch das grundlegende Modell $p(\cdot)$ aufgeführt. w steht für das AIC-Gewicht

	AIC	Δ AIC	w
$p(\text{Minuten}+\text{Monat})$	318.5	0	0.49
$p(\text{Minuten})$	319.2	0.75	0.34
$p(\text{Minuten}+\text{Datum})$	320.8	2.35	0.15
$p(\text{Monat})$	326.6	8.15	0.01
$p(\text{Datum})$	328.2	9.71	0.004
$p(\cdot)$	328.6	10.08	0.003

In diesem Fall gibt es nur zwei Modelle, deren Δ AIC < 2 ist. Diese beiden wurden unter Berücksichtigung ihres AIC-Gewichtes (w) gemittelt und daraus die Detection Probability für jeden einzelnen Kontrollgang pro Standort berechnet. Der maximale Wert von 69% wurde bei einem Beobachtungsgang sechs Minuten vor Sonnenaufgang in Februar erreicht. Das Minimum von 13% stammt aus dem Juni bei 202 Minuten nach Sonnenaufgang.

Wenn im gemittelten Modell die Covariable der Monate auf einem Durchschnittswert konstant gehalten wird, kann der alleinige Einfluss der Minuten nach Sonnenaufgang ermittelt werden. Die Detection Probability liegt bei Sonnenaufgang durchschnittlich bei 62%. Nach einer Stunde sinkt sie auf 44%, beträgt nach zwei Stunden nur noch 29% und erreicht nach 180 Minuten den Wert von 17% (siehe Abbildung 20). Wird hingegen die Covariable der Minuten nach Sonnenaufgang auf dem Durchschnittswert von 72 Minuten gehalten, so wird ersichtlich, dass im Februar und April die Detection Probability bei 47% zu liegen kommt. Im Monat Mai werden 41% erreicht, der Juni liegt mit 40% nur knapp darunter und mit 37% ist die Detection Probability im März am niedrigsten (siehe Abbildung 21).

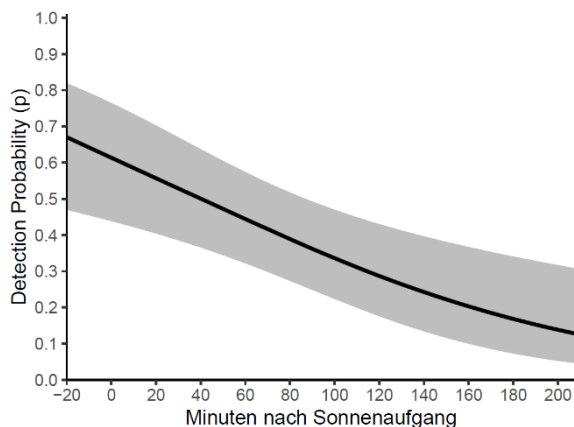


Abbildung 20: Einfluss der Tageszeit auf die Detection Probability (p) mit grauem 95%-Vertrauensintervall.

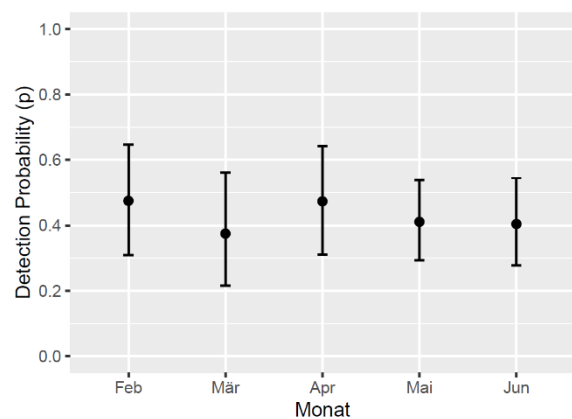


Abbildung 21: Detection Probability (p) in den Monaten Februar bis Juni mit 95%-Vertrauensintervall.

4.2.3 Modellierung mit Besetzungsstatus der Nachbarstandorte als Covariable

Ob der Besetzungsstatus der Nachbarstandorte einen Einfluss auf die Occupancy hat, konnte in einem weiteren Modell aufgeschlüsselt werden. Dieses wurde mit einem AIC von 312.9 bewertet und liegt somit 15.7 AIC-Punkte unter dem Wert von 328.6 des grundlegenden Modells. Sind beide angrenzenden Standorte unbesetzt, so ergibt sich eine Occupancy von gerade einmal 24%. Weiss man, dass an einem der beiden Nachbarorte Wasseramseln vorkommen, steigt die Occupancy bereits auf 72% und erreicht 95%, wenn der Standort beidseitig an besetzte Standorte grenzt (siehe Abbildung 22).

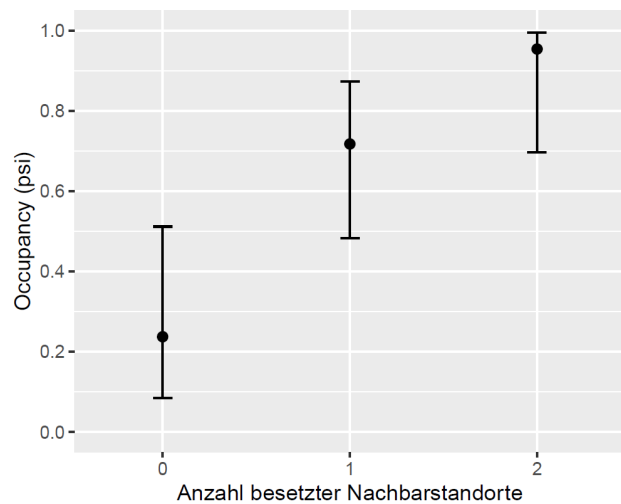


Abbildung 22: Werte der Occupancy bei variierender Anzahl besetzter Nachbarstandorte mit 95%-Vertrauensintervall

4.2.4 Modellierung mit Bachstrukturen als Covariablen

Auf der Suche nach geeigneten Modellen, welche die Occupancy in Abhängigkeit von Bachstrukturen modellieren, wurden alle theoretisch denkbaren Modelle nach ihrem AIC geordnet. Da es aufgrund der erhobenen Bachstrukturen (siehe Abbildung 6 sowie Anhang) sehr viele Kombinationsmöglichkeiten gibt, sind im Prinzip über 1000 verschiedene Modelle möglich. Es zeigte sich, dass einige dieser Modelle Aussagen machen, welche biologisch keinen Sinn ergeben und mit der Literatur im Widerspruch stehen. So prognostizieren die Modelle beispielsweise, dass ein Wasseramselvorkommen umso unwahrscheinlicher wird, je mehr Ansatzsteine vorhanden sind. Ebenso haben gemäss Modell eine grosse Variabilität der Wassertiefe oder ein dichter Uferbewuchs als Deckungsmöglichkeit schlechte Auswirkungen auf das Vorkommen. Diese Bachstrukturen wurden folgerichtig aus der Modellierung ausgeschlossen. Von den verbleibenden Modellen wiesen nur vier die Eigenschaft $\Delta AIC < 2$ auf. In all diesen tritt die Covariable der Naturnähe des Ufers auf, im bestplatzierten Modell sogar alleinig. Bei der jeweils zweiten Covariable in den übrigen drei Modellen handelt es sich um die Qualität des Substrats der Bachsohle, die Entfernung der Uferlinie zum Weg sowie die Stärke der Turbulenzen. Bei allen diesen Covariablen handelt es sich um kategorisierte Werte (vgl. Abbildung 6).

Tabelle 3: Vergleich der vier besten Modelle (nach Ausschluss von Covariablen, welche zu biologisch sinnlosen Modellen führten). «Ufer» steht für die Naturnähe des Ufers, «Bachsohle» für die Qualität des Substrats, «Weg» für dessen Abstand zum Bach und «Turb.» für die Stärke der Turbulenzen. Zusätzlich dargestellt ist das Gewicht der Covariablen innerhalb der einzelnen Modelle. Dieses ist nicht absolut zu verstehen, sondern bildet eine relative Grösse.

	Ufer	Bachsohle	Weg	Turb.	AIC	ΔAIC	w
$\psi(\text{Ufer})$	1.6				325.3	0	0.44
$\psi(\text{Ufer+Bachsohle})$	1.6	0.3			325.3	1.37	0.22
$\psi(\text{Ufer+Weg})$	1.5		0.2		327.2	1.85	0.17
$\psi(\text{Ufer+Turbulenzen})$	1.5			0.1	327.3	1.97	0.16

Analog zur Vorgehensweise bei den Covariablen der Detection Probability hätte man auch hier aus den hoch rangierten Modellen ($\Delta AIC < 2$) ein Durchschnitts-Modell erstellen können. Da aber selbst in den Modellen mit zwei Covariablen die Naturnähe des Ufers einen markant grösseren Einfluss hat und das AIC-Gewicht w des bestplatzierten Modells mit 0.44 deutlich über dem der übrigen liegt, kommt der Naturnähe des Ufers ein so grosser Einfluss auf die Occupancy zu, dass der Verzicht auf das Mitteln der Modelle gerechtfertigt war. Durch den Prozess des Ausschlusses der allermeisten Modelle blieb nur ein einziges übrig. Dieses berücksichtigt allein den Einfluss der Naturnähe des Ufers auf die Occupancy. Mit höher werdender Kategorie dieser Bachstruktur steigt die Occupancy von 22% über 59% bis auf 88% (siehe Abbildung 23). Das Vertrauensintervall ist allerdings gerade bei der Kategorie 1 sehr gross und reicht dort von 4% bis 68%.

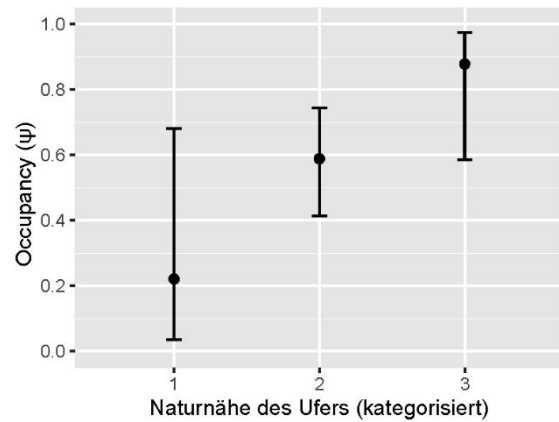


Abbildung 23: Occupancy in Abhängigkeit der kategorisierten Naturnähe des Ufers mit 95%-Vertrauensintervall.



Abbildung 24: Karte mit allen Beobachtungsstandorten. Die Farbe visualisiert den Wert der Occupancy in Abhängigkeit der Naturnähe des Ufers. Ein schwarzer Punkt markiert einen tatsächlich besetzten Standort.

4.2.5 Sensitivitätsanalyse

Mit einer Sensitivitätsanalyse kann die absolute Aussagekraft eines Modells ermittelt werden und nicht wie beim AIC nur die relative. Dies ist also für die Einordnung der Resultate ein wichtiger Schritt, da nur so ermittelt werden kann, ob die Befunde der Modelle überhaupt valide Aussagen über die Realität zulassen. Beim Goodness-of-Fit-Test von MacKenzie & Bailey (2004) erhält man für jedes getestete Modell zwei Werte (\hat{c} und der p -Value), welche schliesslich dessen Aussagekraft quantifizieren (vgl. Abschnitt 2.2.4). Der p -Value aller getesteten Modelle liegt deutlich über den erforderlichen 0.05 und ist somit im grünen Bereich. Der \hat{c} -Wert überschreitet allerdings bei allen Modellen 1, was bedeutet, dass eine Unteranpassung vorliegt. Doch selbst der grösste hier erreichte Wert von 1.23 liegt nahe genug bei 1, um die Aussagekraft des Modells als genügend betrachten zu können (vgl. Abschnitt 2.2.4).

Tabelle 4: p -Value und \hat{c} für jedes der getesteten Modelle. In Klammern sind die verwendeten Covariablen für die Occupancy (ψ) und die Detection Probability (p) angegeben. Ein Fehlen von Covariablen wird durch (.) signalisiert.

	p -Value	\hat{c}
$\psi(\cdot)p(\cdot)$	0.278	1.16
$\psi(\text{Ufer})p(\cdot)$	0.244	1.21
$\psi(\text{Nachbar})p(\cdot)$	0.277	1.11
$\psi(\cdot)p(\text{Minuten}+\text{Monat})$	0.275	1.17
$\psi(\cdot)p(\text{Minuten})$	0.211	1.23

4.2.6 Anzahl Kontrollgänge und wahrscheinlicher Besetzungsstatus

Bei einer durchschnittlichen Detection Probability von 43% (Wert gemäss dem grundlegenden Modell, vgl. Abschnitt 4.2.1) beträgt die gesamte Nachweiswahrscheinlichkeit nach fünf Kontrollgängen 96%. Möchte man mit einer Sicherheit von 80% ausschliessen können, anwesende Wasseramseln verpasst zu haben, so werden nur drei Beobachtungsgänge benötigt. Für 99% sind hingegen neun Durchgänge vonnöten (siehe Tabelle 5 und Abbildung 25).

Tabelle 5: *W'keit, anwesende Wasseramseln nach bestimmter Anzahl Kontrollgänge nachzuweisen*

Anzahl	Wahrscheinlichkeit
1	43%
2	67%
3	81%
4	89%
5	94%
6	96%
7	98%
8	99%
9	99%
10	99.6%
11	99.8%
12	99.9%

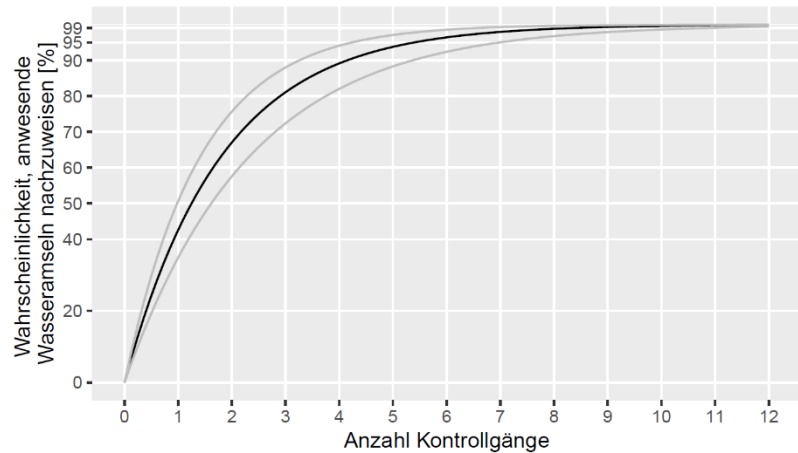


Abbildung 25: *Gesamt-Nachweiswahrscheinlichkeit in Abhängigkeit der Anzahl Kontrollgänge. Die grauen Kurven basieren auf Werten der Detection Probability bei unterer und oberer Grenze des Vertrauensintervalls.*

Ausgehend von den errechneten Werten des grundlegenden Modells lässt sich mit dem Satz von Bayes (vgl. Grundlagen) bestimmen, dass ein Standort nach fünf negativen Nachweisen mit einer 12-prozentigen Wahrscheinlichkeit dennoch besetzt sein könnte.

4.3 Decision-Tree-Modell

Die Vorhersage von Wasseramselvorkommen anhand aller gesammelten Bachstrukturen mittels eines Decision-Trees führte analog zur Site-Occupancy-Modellierung zu mehreren Befunden, welche gegen die biologische Sinnhaftigkeit verstossen. Nach Ausschluss der entsprechenden Bachstrukturen verblieben nur noch die Naturnähe des Ufers, der Abstand des Weges zum Bach, die Qualität des Substrats der Bachsohle sowie die Variabilität der Bachbreite in der Modellierung. Im Pruning fiel jedoch auch die letztgenannte Bachstruktur aus dem finalen Decision-Tree (Abbildung 26), sodass dieser nur drei Bachstrukturen beinhaltet.

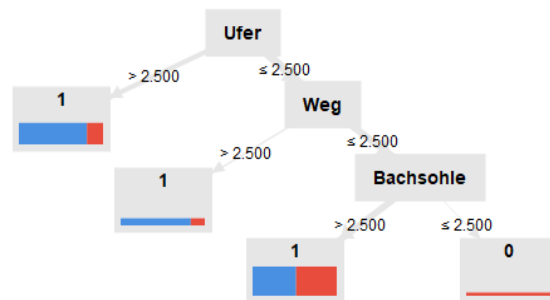


Abbildung 26: *Finaler Decision-Tree. In den Blattknoten ist der prognostizierte Besetzungsstatus (0 oder 1) angegeben. Die Dicke der Linien widerspiegelt die Anzahl Daten, welche die entsprechende Bedingung erfüllen. Der blaue Teil der Balken steht für den Anteil der Daten, welcher in Realität einen Besetzungsstatus 1 aufweist, während der rote Anteil für den Teil der Daten mit realem Status 0 steht.*

Der erste Entscheidungsknoten beinhaltet als Split-Kriterium die kategorisierte Naturnähe des Ufers, sodass diese den grössten Einfluss auf den Besetzungsstatus hat. Auffällig ist die tiefe Trefferquote korrekter Klassifikationen beim letzten Split. Dies äussert sich auch im «accuracy»-Wert des gesamten Decision-Trees, welcher nur 52% beträgt, wobei Werte unter 60% eine ungenügende Aussagekraft signalisieren (Allwright, 2022).

5. Diskussion

5.1 Abschätzung der Vorkommen mittels Präsenz-/Absenzdaten

Die erste Fragestellung dieser Arbeit lautet, ob mit Punktbeobachtungen valide Abschätzungen von Wasseramsel-Vorkommen gelingen. Angesichts der vielen Nachweise und der hohen Detection Probability von 43% pro Durchgang erweist sich die gewählte Methode als überraschend brauchbar. Die geringe Differenz der «Naïve»-Werte zur Occupancy beziehungsweise Detection Probability belegt ebenfalls die Tauglichkeit der Vorgehensweise, da die Kartierung offenbar nur sehr wenige besetzte Standorte fälschlicherweise als unbesetzt einordnete. Besonders die Vermutung, dass auf die gewählte Weise auch an kaum zugänglichen Bachabschnitten zuverlässig Sichtungen gelingen, konnte bestätigt werden. Mit den positiven Nachweisen lässt sich zwar nicht die exakte Anzahl Reviere ermitteln, aber eine Schätzung ist durchaus möglich. Die Revierlänge von Wasseramseln beträgt je nach Qualität des Reviers 0.4 bis 2 Kilometer (Hegelbach, 2004; Glutz von Blotzheim & Bauer, 1985). Durchschnittliche Reviere umfassen also einen Gewässerabschnitt von gut einem Kilometer. Ein Blick auf Abbildung 18 lässt aufgrund der Verteilung der Nachweise vermuten, dass man es im Aathal zwischen Oberuster und Wetzikon mit solchen durchschnittlichen Revieren zu tun haben könnte. Somit wären hier mindestens drei bis maximal fünf Reviere vorhanden. Im Abschnitt von Auslikon bis Oberkempten gibt es höchstwahrscheinlich kein Brutpaar. Zu diesem Schluss kam im Vorjahr auch Hatt (2022), welcher seine Daten mit einer klassischen Revierkartierung erhob. Im anschließenden Chämptnertobel bis Bäretswil fällt auf, dass bei mehr als der Hälfte der Standorte entweder drei oder vier von maximal fünf möglichen Nachweisen gelangen. Diese Quote ist deutlich höher als im Aathal und liesse sich erklären durch den Befund von Hatt (2022), dass das Chämptnertobel für die Wasseramsel perfekte Bedingungen bietet, weswegen die Revierlänge hier nahe am bekannten Minimum liegt. Kürzere Reviere führen zu einer höheren Präsenz der Individuen an einzelnen Beobachtungspunkten und damit zu einer höheren Detection Probability. Somit scheint sich sogar die Revierqualität in den vorliegenden Präsenz-/Absenzdaten zu widerspiegeln. Wie zuverlässig dies der Fall ist, müsste in einer Folgestudie untersucht werden – mir sind keine Arbeiten zu diesem Thema bekannt. Eine solche Studie müsste auch den Einfluss der Einsehbarkeit der Bachabschnitte berücksichtigen, welche sicher eine Rolle spielt. Aber die Modellierung des Einflusses der Naturnähe des Ufers auf die Occupancy diagnostiziert für das Chämptnertobel tatsächlich eine höhere Lebensraumqualität als für das Aathal (siehe Abbildung 24; mehr zu dieser Modellierung in Abschnitt 5.4)

Als Fazit lässt sich sagen, dass man durch Präsenz-/Absenzdaten auch an schlecht einsehbaren Bachabschnitten einen guten Überblick über das Vorkommen im Untersuchungsgebiet erhält. Die Punktbeobachtungs-Methode liefert zwar keine exakte Anzahl Reviere, erlaubt aber recht präzise Schätzungen und gibt vermutlich sogar Hinweise auf die Qualität des Habitats.

5.2 Einfluss von zeitlichen Covariablen auf die Detection Probability

Durch den Vergleich mehrerer Site-Occupancy-Modelle gelang es zu belegen, dass bei den Begehungen die Tageszeit und der Monat einen klaren Einfluss auf die Detection Probability haben. Dass die erstgenannte Covariable von Bedeutung ist, findet man auch in der Fachliteratur. So sprechen Glutz von Blotzheim & Bauer (1985) von der höchsten Aktivität der Wasseramsel in den ersten Stunden nach Sonnenaufgang. Die Stärke der Aktivitätsreduktion wird dabei allerdings nicht quantifiziert. Mit der entsprechenden Modellierung gelang es

nun sichtbar zu machen, in welchem Masse diese Reduktion stattfindet (siehe Abbildung 20). Die Detection Probability nimmt von Sonnenaufgang an kontinuierlich und stark ab, sodass sie nach zwei Stunden um mehr als die Hälfte reduziert ist und nach drei Stunden sogar unter einem Drittel des ursprünglichen Wertes liegt. Ohne zusätzliche Vergleichswerte lässt sich natürlich nicht beurteilen, ob das Modell hier aussergewöhnliche Werte ermittelt hat. Da die vorliegenden Resultate für künftige Kartierungen aber von grosser Bedeutung sind, wären ergänzende Studien sehr wünschenswert.

Die Jahreszeit hat ebenfalls Auswirkungen auf die Detection Probability, allerdings deutlich weniger ausgeprägt als die Tageszeit (siehe Abbildung 21). Zudem liegt kein kontinuierlicher Trend im Jahresverlauf vor, weshalb die Covariable «Datum» verworfen und lediglich die Covariable «Monate» beibehalten wurde. Diese hat den Vorteil, dass sie in *R* Schwankungen im Untersuchungsverlauf darstellen kann (vgl. Abschnitt 3.3.2). Das Auftreten dieser (leichten) Schwankungen passt zu den verschiedenen Phasen des Brutgeschäftes (vgl. Abschnitt 2.1.2; siehe auch Glutz von Blotzheim & Bauer [1985] sowie Hegelbach [2004]). Der hohe Wert im Februar könnte dadurch zustande kommen, dass während dieser Zeit die Balz in vollem Gange ist. Wasseramseln sind also besonders aktiv und machen oft durch Rufe und Gesänge auf sich aufmerksam. Interaktionen zwischen Wasseramseln sind die Regel und nicht die Ausnahme. Da die Reviergrenzen möglicherweise noch nicht restlos geklärt sind, kommt es des Öfteren zu lautstarken Auseinandersetzungen und Verfolgungsjagden zwischen Rivalen, welche auch dafür sorgen, dass die Art leichter entdeckt werden kann. Stark im Kontrast dazu steht die Situation im März. Dieser Monat weist die niedrigste Detection Probability im Untersuchungsverlauf auf. Die Eiablage gehört wahrscheinlich bereits der Vergangenheit an. Der Verdacht liegt nahe, dass ein gewisser Anteil der Individuen mit dem Bebrüten der Eier beschäftigt ist und auf dem Nest sitzt. Wenn sich weniger Tiere frei bewegen, so führt das natürlich zu einer tieferen Chance, die Spezies zu entdecken. Erst wenn im April der Nachwuchs bereits geschlüpft ist, steigt die Nachweiswahrscheinlichkeit wieder. Die adulten Tiere sind damit beschäftigt, möglichst viel Nahrung in kürzester Zeit für die Jungen zu beschaffen. Dies äussert sich natürlich in einer erhöhten Aktivität. Durch das Verlassen des Nestes erster immaturer Wasseramseln findet zudem eine Erhöhung der Anzahl beobachtbarer Individuen statt, was wohl auch zum Anstieg der Detection Probability im April beitrug. Vermutlich durch die Verselbstständigung der überlebenden Jungtiere wurde die Lebensweise der im Revier verbleibenden Wasseramseln gegen Ende der Brutzeit hin wieder etwas unauffälliger. Zusätzlich erschwerte die zunehmende Vegetationsdichte vermehrt die Sicht auf den Bach. Dies könnten Gründe dafür sein, weshalb in den Monaten Mai und Juli die Detection Probability wieder unter dem Niveau des Monats April liegt.

Doch lässt sich aus den Erkenntnissen der Modelle die gewählte Kartierungsmethode verbessern und effizienter gestalten? Dazu lohnt sich ein erneuter Blick auf die Abhängigkeit der Detection Probability von der Tageszeit (siehe Abbildung 20). Offensichtlich liessen sich die Beobachtungschancen von Wasseramseln deutlich erhöhen, fände ein durchschnittlicher Kontrollgang früher am Morgen statt. Würden an einem Tag statt zwischen elf und zwölf nur sieben Standorte (in einem von dementsprechend acht Sektoren) besucht und noch konsequenter darauf geachtet, bei Sonnenaufgang mit der Datenaufnahme zu beginnen, so liesse sich die Detection Probability wohl um etwa sieben Prozentpunkte auf 50% erhöhen. Damit wäre bereits nach drei Kontrollgängen in acht Sektoren eine Gesamt-Nachweiswahrscheinlichkeit von 88% erreicht, was etwa dem bisherigen Wert nach vier Kontrollgängen in fünf Sektoren entspricht. Gibt man sich mit diesem Wert zufrieden, so kann je nach Kapazitäten und Dauer der Anreise eine Reduktion auf drei Beobachtungsgänge pro Standort eine sinnvolle Massnahme sein. Denn mit 3 Kontrollgängen mal 8 Sektoren müssten nur vier zusätzliche Begehungen aufgewendet werden. Allerdings würde dabei die Länge des

morgendlichen Einsatzes deutlich kürzer, sodass sich in der Bilanz die Dauer der Datenaufnahme um einen Viertel reduzieren liesse. Mit deutlich geringerem Aufwand könnte also bereits ein gleich gutes Resultat erzielt werden. Dies würde auch für einen Ansatz mit höheren Ansprüchen an die Gesamt-Nachweiswahrscheinlichkeit gelten. Der zeitliche Gewinn wäre etwas geringer, aber immer noch in einem vergleichbaren Rahmen.

Jetzt stellt sich die Frage, auf welche Beobachtungsgänge man verzichten sollte. Angesichts der Tatsache, dass im Februar die Brutzeit gerade erst begonnen hat und die Reviergrenzen womöglich noch nicht ganz gefestigt sind, könnte es trotz der hohen Detection Probability sinnvoll sein, diesen Kontrollgang zu streichen, um eine Verzerrung der Resultate zu vermindern. Durch die grössere Mobilität der Jungtiere gegen Ende der Brutzeit hin könnte ebenfalls vermehrt ein falsches Bild tatsächlich besetzter Standorte entstehen (siehe auch Abschnitt 5.5). Die Juni-Begehung hat den weiteren Nachteil, dass es durch die lange Dauer des Tageslichtes sehr schwierig ist, bereits bei Sonnenaufgang am Beobachtungspunkt zu stehen, was zu einer verminderten Detection Probability führt. Besonders wenn man auf öffentliche Verkehrsmittel angewiesen ist, lässt sich dieses Problem kaum umgehen. Ein Verzicht auf diesen Kontrollgang kann also zur Erhöhung der Detection Probability beitragen.

Der Vorschlag dieser Arbeit lautet dementsprechend, sich auf die Monate März bis Mai zu beschränken und in diesen dank kürzeren Begehungen mit geringerem Aufwand die gleiche Gesamt-Nachweiswahrscheinlichkeit zu erreichen. Ob sich die Effizienz durch eine kürzere Verweildauer an den Standorten (z.B. acht anstatt bisher zehn Minuten) weiter steigern liesse, müsste in einer Folgestudie überprüft werden.

5.3 Einfluss des Besetzungsstatus der Nachbarstandorte auf die Occupancy

Wenig überraschend sind die Erkenntnisse aus der Modellierung mit der Covariable «Anzahl besetzter Nachbarstandorte». Da die meisten Beobachtungspunkte in einem Abstand von rund 200 Metern zueinander angeordnet sind, kommen zwangsläufig mehrere in einem Revier zu liegen, da deren Länge zwischen 0.4 und 2 Kilometern beträgt (Hegelbach, 2004; Glutz von Blotzheim & Bauer, 1985). Findet man einen scheinbar besetzten Standort vor, so kann man in den meisten Fällen annehmen, dass er sich innerhalb der Reviergrenzen eines Paares befindet. Daraus folgt, dass unter den benachbarten Standorten höchstwahrscheinlich ebenso besetzte Stellen existieren. Zu genau diesem Schluss kommt auch das entsprechende Modell. Findet man also eine Wasseramsel, so kann man mit einer Sicherheit von 72% bis 95% (siehe Abbildung 22) davon ausgehen, dass in der näheren Umgebung ebenso Wasseramseln anzutreffen sind.

5.4 Vorhersagen zum Vorkommen anhand der Bachstrukturen

Sowohl das Site-Occupancy-Modell als auch der Decision-Tree diagnostizieren einen starken Einfluss der Naturnähe des Ufers auf das Vorkommen der Wasseramsel, während andere Bachstrukturen gemäss den Modellierungen entweder einen schwächeren Einfluss zeigen oder aber zu biologisch sinnlosen Modell-Aussagen führen. Eine kritische Analyse der gewählten Vorgehensweise drängt sich also auf.

5.4.1 Mangelnde Aussagekraft der erhobenen Bachstrukturen

Ganz grundsätzlich ist es nur bedingt sinnvoll, Bachstrukturen an einzelnen Punkten zu erfassen und daraus auf das Vorkommen der Wasseramsel schliessen zu wollen. In einem besetzten Wasseramsel-Revier kommen zwar offensichtlich alle notwendigen Elemente in einer ausreichenden Anzahl vor. Sie müssen aber keinesfalls gleichmässig über den

Bachabschnitt verteilt, sondern in der Summe vorhanden sein. Durch die gewählte Vorgehensweise der Punktbeobachtung wurden indessen immer nur sehr kurze Bachabschnitte beurteilt. Je nach Einsehbarkeit waren diese Bachabschnitte ca. 10 bis 30 Meter lang. Der grösste Teil des Gewässers wurde damit nicht erfasst. Die erhobenen Daten bilden somit nur einen sehr kleinen Teil des ganzen Brutgewässers ab. Ob die untersuchten Abschnitte repräsentativ für das Gewässer sind, hängt somit ein Stück weit vom Zufall ab. Dies gilt allerdings nicht für alle Strukturen in gleichem Masse. Grossräumig recht homogene Bacheigenschaften wie die Naturnähe des Ufers dürften auch bei Punktaufnahmen repräsentativ beurteilt werden. So lässt sich erklären, weshalb genau diese Bachstruktur sowohl im Site-Occupancy-Modell als auch im Decision-Tree ein biologisch sinnvolles Resultat lieferte. Sehr heterogen verteilte Strukturen wie beispielsweise Ansitzsteine können hingegen durch die gewählte Vorgehensweise sehr leicht übergewichtet oder unterschätzt werden. Der Einfluss des Zufalls ist hier also besonders gross, was das Auftreten biologisch sinnloser Befunde begünstigt, welche im Bezug auf die Ansitzsteine in beiden Modellierungen aufgetreten sind. Eine weitere Gruppe von Bachstrukturen ist weder ausgeprägt homogen noch sehr heterogen verteilt. Dazu gehören beispielsweise die Variabilität der Wassertiefe, die Art der Bachsohle oder die Stärke der Turbulenzen. Die Wahrscheinlichkeit, dass mittels Punktbeobachtungen repräsentative Daten entstehen, sollte in einem durchschnittlichen Bereich zu liegen kommen. Die Vorhersagen der Modelle zu dieser Gruppe von Bachstrukturen waren sehr uneinheitlich. Teilweise gab es biologisch sinnvolle Aussagen, allerdings nicht sehr stark ausgeprägt. Zudem entstanden teils Widersprüche zwischen den beiden Modellen und auch innerhalb der Modellauswahl der Site-Occupancy-Modelle.

Aus diesen Beobachtungen und Befunden lässt sich für Projekte mit einem Stichprobenumfang wie in dieser Arbeit folgender Schluss ziehen: Heterogen verteilte Strukturen sollten bei Punktkartierungen nicht verwendet werden, da sie keine verlässlichen Aussagen zum Vorkommen der Wasseramsel erlauben.

5.4.2 Grösse des Stichprobenumfangs

Liessen sich die oben beschriebenen Probleme mit einem deutlich grösseren Stichprobenumfang lösen? Zu einem gewissen Grad lässt sich diese Frage bejahen. Einige Kategorien von Bachstrukturen tauchten nämlich nur an sehr wenigen Standorten auf. So wurde z.B. die schlechteste Kategorie «1» der Naturnähe des Ufers nur an einer einzigen Stelle gefunden, was sich auch im grossen Vertrauensintervall der Occupancy an diesem Ort widerspiegelt (siehe Abbildung 23). Je grösser die Anzahl der Beobachtungspunkte, desto wahrscheinlicher wird es, dass sich zufallsbedingte Fehleinschätzungen ausmitteln und die Modelle repräsentative Aussagen liefern. Es bräuchte allerdings eine riesige Anzahl an Datenpunkten, damit tatsächlich quantitativ verlässliche Vorhersagen entstehen können. In der Praxis wären solche grundsätzlich präzisen Aussagen aber trotzdem nicht ohne Weiteres anwendbar. Steht man nämlich an einem Standort, so ist weiterhin unbekannt, ob die sichtbaren Bachstrukturen repräsentativ sind. Die durch die Modelle gewonnenen Erkenntnisse haben somit lediglich einen grundsätzlichen Charakter, es sei denn, man erfasst die Bachelemente ganzer Abschnitte vollständig, was dann allerdings nicht mehr mit Punktbeobachtungen vereinbar ist.

5.4.3 Eignung der verwendeten Modellierungsansätze

Sowohl der Decision-Tree als auch das Site-Occupancy-Modell liefern zu bestimmten Bachstrukturen biologisch sinnvolle Resultate. Allerdings ist die Aussagekraft des Decision-Trees deutlich ungenügend, da der «accuracy»-Wert nur bei 52% liegt. Gemäss Allwright (2022) müsste der Wert mindestens 60% betragen, damit die Aussagekraft des Modells

akzeptabel ist. Mehrere Gründe tragen zum schlechten Abschneiden des Decision-Trees bei. Einerseits ist sicher die Stichprobenzahl viel zu klein (siehe oben). Da Entscheidungsbäume im Gegensatz zu Site-Occupancy-Modellen absolute Aussagen zum Besetzungstatus machen (besetzt oder unbesetzt), führen kleine Stichproben schnell zu einem schwachen «accuracy»-Wert. Das Site-Occupancy-Modell hat in dieser Hinsicht den Vorteil, dass es Wahrscheinlichkeitswerte generiert. Das Wesen von Wahrscheinlichkeiten ist ja gerade, dass auch der Gegenfall eintreten könnte, ohne dass die errechneten Werte deswegen in Frage gestellt werden. Beim Decision-Tree allerdings existieren nur «richtig» oder «falsch». Hier zeigt sich ein weiteres Problem der Entscheidungsbäume. Sie bauen auf einem strikten «Wenn-Dann-Algorithmus» auf. In natürlichen Systemen gibt es aber kaum je derart klare Ursache-Wirkungs-Gefüge. Zusätzlich zu unberücksichtigten Nebenfaktoren spielt immer auch ein Stück weit der Zufall mit. Bloss weil eine bestimmte Struktur vorhanden ist, führt dies nicht automatisch zur Anwesenheit einer Wasseramsel. Hinzu kommt eine weitere Schwäche der Decision-Tree-Modellierung. Sie ignorieren die Möglichkeit von «False-Zeros», also von fälschlicherweise als unbesetzt taxierten Standorten. Dieser Aspekt war ja gerade ausschlaggebend für die Entwicklung der Site-Occupancy-Modelle (MacKenzie et al., 2018). In dieser Arbeit wird dieser Mangel des Decision-Trees allerdings nicht allzu stark ins Gewicht fallen, da mit der erreichten Gesamt-Nachweiswahrscheinlichkeit von 94% höchstens vereinzelt «False-Zeros» auftreten dürften. Als Fazit lässt sich trotzdem sagen, dass Decision-Trees für Fragestellungen, wie sie in dieser Arbeit behandelt wurden, ungeeignet sind. Der ungenügende «accuracy»-Wert ist somit keine Überraschung.

Aber auch die Aussagen von Site-Occupancy-Modellen können nicht als zwingend korrekt betrachtet werden. Selbst wenn eine Bachabschnitt perfekte Strukturen aufweist, kann er aus Gründen, welche nicht auf Zufall beruhen, unbesetzt bleiben. Wie Hatt (2022) zeigen konnte, blieben grundsätzlich geeignete Gewässerstrecken unbesetzt, da eine als Neststandort nutzbare Stelle fehlte. Dieser Faktor ist nicht unbedingt als klassische Bachstruktur zu betrachten, da er theoretisch nur an einer einzigen Stelle im ganzen Revier benötigt wird. Sein Fehlen kann mit einer Punktbeobachtungsmethode unmöglich erkannt werden. Zudem können aus populationsdynamischen Gründen in einzelnen Jahren schlicht zu wenig Individuen vorhanden sein, um alle geeigneten Bachabschnitte zu besetzen. Beide beschriebenen Situationen können zu einer falschen Gewichtung der untersuchten Bachelemente führen. Daher bleibt unklar, wie verlässlich die Aussagen von Site-Occupancy-Modellen bezüglich des Einflusses von Bachstrukturen sind. Als Trend sind die biologisch sinnvollen Modell-Aussagen aber sicher brauchbar. Sie decken sich ja auch mit Aussagen der Literatur (vgl. Glutz von Blotzheim & Bauer [1985], Bosch & Lurz [2019] und Martinez et al. [2020]), die diesen Strukturen einen positiven Einfluss zuschreibt.

5.5 Einhaltung grundlegender Annahmen und Voraussetzungen der Single-Season Site-Occupancy-Modelle

1. Es ist davon auszugehen, dass sich an einzelnen Standorten der Besetzungsstatus während der Untersuchungsdauer verändert hat. Solch eine Verletzung der Voraussetzungen ist beispielsweise an Standort C1 (siehe Abbildung 5) sehr wahrscheinlich. An diesem Punkt konnte nur an einem einzigen Tag eine Wasseramsel beobachtet werden. Diese Tatsache an sich wäre noch kein eindeutiger Hinweis auf einen Wechsel des Besetzungsstatus. Betrachtet man allerdings alle Wasseramsel-Nachweise entlang des Chämptnerbachs im Ried, so merkt man, dass der oben genannte der einzige ist. Würde es sich bei C1 um einen durchgehend besetzten Standort und somit auch um einen Teil eines Reviers handeln, so wären weitere positive Nachweise im Bereich des unteren Chämptnerbachs zu erwarten. Da dem aber nicht so ist, kann man mit sehr grosser Sicherheit davon

ausgehen, dass der Nachweis auf ein umherziehendes Individuum zurückzuführen ist. Gleichentags wie die Wasseramsel konnten an der Stelle auch Kotflecken dokumentiert werden, welche aufgrund ihrer Grösse und Lokalisierung auf Steinen am Bach mit an Sicherheit grenzender Wahrscheinlichkeit einer Wasseramsel zugeordnet werden können. Im weiteren Verlaufe der Untersuchung konnten keine solchen Anzeichen auf ein Vorkommen der Art festgestellt werden. Somit ist es unpassend, an diesem Ort von einem besetzten Standort zu sprechen. Es ist anzunehmen, dass das gleiche für weitere Beobachtungspunkte gilt. Da in R aber auch an solchen Orten mit Besetzungsstatus 1 gerechnet wurde, liegt die tatsächliche Occupancy wohl unter dem berechneten Anteil besetzter Standorte. Es ist anzunehmen, dass die Detection Probability durch die Annahmen-Verletzung hingegen eher negativ verzerrt wurde, denn Standorte mit nur einem positiven Nachweis in fünf Kontrollgängen sorgen dafür, dass die Detection Probability tiefer eingeschätzt wird, als sie in Realität ist. Umherziehende Individuen können diesem Effekt aber auch zumindest teilweise entgegenwirken. So ist es denkbar, dass diese an besetzten Orten auftauchen und dort die Anzahl Sichtungen erhöhen. Die Nachweiswahrscheinlichkeit wird aber durch den letztgenannten Effekt schwächer beeinflusst als durch den erstgenannten. Eine Erhöhung der Anzahl positiver Nachweise eines ohnehin schon besetzten Standorts fällt weniger ins Gewicht als ein Standort mit niedriger Detection Probability, der ansonsten gar keinen Einfluss auf die Nachweiswahrscheinlichkeit gehabt hätte, da keine Beobachtungen gelangen. Somit kann man davon ausgehen, dass durch umherziehende Wasseramseln in der Bilanz eine Unterschätzung von p ausgelöst wird. Individuen, welche sich nicht an ihre Reviergrenzen halten, könnten einerseits unverpaarte Tiere sein oder natürlich Jungtiere, welche sich ohne Probleme uneingeschränkt fortbewegen können. Die Eltern sind allerdings auch in der Lage, ihrem Nachwuchs aus dem Revier hinaus zu folgen (Glutz von Blotzheim & Bauer, 1985).

2. Occupancy und Detection Probability sind offensichtlich nicht konstant. Zwar wurde versucht, mögliche Heterogenität als Funktion von Covariablen zu modellieren. Eine Garantie, jegliche Ungleichheiten dadurch erklären zu können, existiert allerdings nicht. Da beim Goodness-of-Fit-Test sämtlichen Modellen eine Unteranpassung diagnostiziert wurde, kann man annehmen, dass sich tatsächlich weitere, unbekannte Umweltfaktoren auf Occupancy und Detection Probability auswirken. Eine Abschätzung, in welche Richtung eine Verzerrung dadurch ausfallen könnte, ist aber kaum möglich.
3. Die Frage, ob sämtliche Datenpunkte unabhängig voneinander sind, muss wohl verneint werden. Zeitlich sollten zwar keine Abhängigkeiten vorhanden sein, örtlich hingegen schon. Da die Standorte nahe genug beieinander liegen, dass Wasseramseln mehrere Standorte besuchen können, ist die Unabhängigkeitsbedingung nicht mehr erfüllt. Wird beispielsweise ein vorbeifliegendes Individuum registriert, so sinkt die Wahrscheinlichkeit, wenige Minuten später am Nachbarort, aus dessen Richtung die Wasseramsel geflogen kam, einen positiven Nachweis zu vollbringen, da sich hier wahrscheinlich gerade kein Vogel aufhält. Dieser Effekt kann aber selbstverständlich auch umgekehrt auftreten. Die gegenseitige Beeinflussung der Nachweise sollte sich also im Durchschnitt aufheben. Auch gemäss MacKenzie et al. (2018) ist es absolut unproblematisch, wenn Individuen an mehreren Standorten auftreten können. Somit sollte durch die Unabhängigkeitsverletzung keine Verzerrung der Resultate zustande gekommen sein.
4. Fehlbestimmungen, welche fälschlicherweise zu positiven Nachweisen führten, können ausgeschlossen werden, da unsichere Nachweise nicht auf dem Protokollblatt erfasst wurden. Diese Modellvoraussetzung wurde also perfekt eingehalten.

6. Schlusswort

Aufgrund meiner Begeisterung sowohl für Biologie als auch für Mathematik und Informatik bildet diese Arbeit für mich eine perfekte Vereinigung dieser Fachbereiche. Während ich für die Datenaufnahme auf viel eigenes Wissen und mehrjährige Erfahrungen mit ornithologischen Kartierungen zurückgreifen konnte, begann ich beim Erstellen der Modellierungen bei Null. Als Ausgangspunkt diente mir die Bachelorarbeit von Steffen (2020). Sie gab mir Anhaltspunkte für ein mögliches Vorgehen und lieferte auch wertvolle Literaturtipps. Das meiste Wissen eignete ich mir schliesslich über Fachliteratur und ausgewählte Internetseiten an. Motivierend war, wie rasch ich die zu Beginn recht kompliziert erscheinenden Grundlagen zu den Site-Occupancy-Modellen verstand. Eine besonders grosse Hilfe stellte das umfassende Buch von MacKenzie et al. (2018) dar. Für die konkrete Umsetzung der Modellierungen im Programm *R* waren mehrere Internetseiten (z.B. Wilkinson [2020] und Paterson [2020]) eine unschätzbare Hilfe. Diese ermöglichten es mir, die R-Skripts tatsächlich auch zu verstehen und nicht einfach nur anzuwenden, ohne exakt zu wissen, wie sie funktionieren. Es hat mir Freude bereitet, mich in diese neuen Themenbereiche einzuarbeiten und ein gutes Verständnis dafür aufzubauen.

Etwas unsicher war ich bei der Aufnahme der Decision-Tree-Modellierungen in die Arbeit, da diese die Problematik der «False-Zeros» nicht berücksichtigen. Da diese Modellierung aber nicht wie die Site-Occupancy-Modelle den Anteil tatsächlich besetzter Standorte berechnen soll, ging ich davon aus, dass die mögliche Fehlerquelle nicht allzu stark ins Gewicht fallen sollte. Dies auch, weil ich bereits nach den ersten Beobachtungsgängen merkte, dass eine recht hohe Gesamt-Nachweiswahrscheinlichkeit resultieren könnte. Schlussendlich war es herausfordernd, aber auch äusserst interessant, für die Diskussion zu überlegen, wie die mangelnde Tauglichkeit des Decision-Trees erklärt werden könnte.

Eine erfreuliche nachträgliche Bestätigung, dass ich bei meinen Modellierungen nichts übersehen hatte, lieferte mir der schriftliche Austausch mit M. Kéry, einem Mitarbeiter der Schweizerischen Vogelwarte Sempach, der im Bereich der Modellierung von populationsbiologischen Aspekten über ein ausserordentliches Fachwissen verfügt und diverse Fachpublikationen und -bücher (mit-)verfasst hat. Zu hören, dass ein Profi auf diesem Gebiet sich von meiner Maturitätsarbeit beeindruckt zeigte, gab mir zusätzliche Motivation beim fundierten Weiterarbeiten.

Es freut mich natürlich, durch die Arbeit klare Erkenntnisse besonders zur Nachweiswahrscheinlichkeit gewonnen zu haben und dadurch fähig zu sein, Vorschläge zu machen, wie die Kartierungsmethode effizienter gestaltet werden kann. Ebenso finde ich die gerade nicht vorhandene Tauglichkeit der Modellierungen mit den Bachstrukturen äusserst interessant. Dies bildet für mich einen sehr zufriedenstellenden Abschluss des Prozesses meiner Maturitätsarbeit, welcher mir ohnehin schon weitestgehend Freude bereitet hat.

Es wäre spannend, mit der Punktkartierungsmethode ein grösseres Monitoring von Wasseramseln im Zürcher Oberland aufzuziehen und somit mehr über ihr Vorkommen und ihre Verbreitung erfahren zu können. Die Einfachheit der Vorgehensweise liesse auch zu, dass die Wasseramseln erfasst würden von Personen, welche bisher über keine Erfahrungen mit ornithologischen Kartierungen verfügen. Mit noch mehr Daten liessen sich auch die Aussagen zur zeitlichen Variabilität der Detection Probability weiter präzisieren. Ich bin gespannt, ob sich so ein Projekt irgendwann realisieren lässt und bin motiviert, mich hierfür zu engagieren.

7. Literaturverzeichnis

- Allwright, S. (2022). *What is a good accuracy score in machine learning?* Abgerufen am 18. Oktober 2022 von Stephen Allwright: <https://stephenallwright.com/good-accuracy-score/>
- Anuradha, D., & Gupta, G. (2014). A self explanatory review of decision tree classifiers. *International conference on recent advances and innovations in engineering (ICRAIE-2014)*. IEEE.
- Bosch, P., & Lurz, P. (2019). *Die Wasseramsel, Die neue Brehmbücherei* (Bd. 489). Magdeburg: VerlagsKG Wolf.
- Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multi-Model Inference, 2nd Edition*. New York: Springer.
- Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: understanding AIC and BIC in model selection. *Sociological Methods & Research*, 33(2), S. 261-304.
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), S. 20-28.
- Chauhan, N. S. (2020). *Decision Tree Algorithm, Explained*. Abgerufen am 7. August 2022 von KDnuggets: <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>
- Chugh, V. (2022). *Which Metric Should I Use? Accuracy vs. AUC*. Abgerufen am 18. Oktober 2022 von KDnuggets: <https://www.kdnuggets.com/2022/10/metric-accuracy-auc.html>
- Dénes, F. V., Silveira, L. F., & Beissinger, S. B. (2015). Estimating abundance of unmarked animal populations: accounting for imperfect detection and other sources of zero inflation. *Methods in Ecology and Evolution*(6), S. 543–556.
- Dormann, C. F., & Kühn, I. (2009). *Angewandte Statistik für die biologischen Wissenschaften*. Helmholtz Zentrum für Umweltforschung-UFZ.
- Dunham, J. B., & Rieman, B. E. (1999). Metapopulation structure of bull trout: influences of physical, biotic, and geometrical landscape characteristics. *Ecological Applications*, 9(2), S. 642-655.
- Engler, R., Guisan, A., & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of applied ecology*, 41(2), S. 263-274.
- Fiske, I. J., & Chandler, R. B. (2011). unmarked: An R Package for Fitting Hierarchical Models of Wildlife Occurrence and Abundance. *Journal of Statistical Software*(43), S. 1-23.
- Glutz von Blotzheim, U., & Bauer, K. (1985). *Handbuch der Vögel Mitteleuropas* (Bd. 10). Wiebelsheim: Aula Verlag.
- Gu, W., & Swihart, R. K. (2004). Absent or undetected? Effects of non-detection of species occurrence on wildlife–habitat models. *Biological Conservation*, 116(2), S. 195-203.

- Hatt, N. (2022). *Eignung des Chämptnerbachs als Lebensraum für die Wasseramsel (Cinclus cinclus)*. Maturitätsarbeit.
- Hegelbach, J. (2004). Zunahme des Brutbestands der Wasseramsel *Cinclus cinclus* am Küsnachter Bach von 1987 bis 2002. *Ornithologischer Beobachter*(101), S. 99-108.
- Kéry, M., & Royle, J. A. (2016). *Applied Hierarchical Modeling in Ecology*. Academic Press.
- Kéry, M., & Schaub, M. (2012). *Bayesian Population Analysis Using WinBUGS: a Hierarchical Perspective*. Amsterdam: Elsevier.
- Knaus, P., Antoniazza, S., Wechsler, S., Guélat, J., Kéry, M., Strebel, N., & Sattler, T. (2018). *Schweizer Brutvogelatlas von 2013-2016. Verbreitung und Bestandesentwicklung der Vögel in der Schweiz und im Fürstentum Liechtenstein*. Sempach: Schweizerische Vogelwarte.
- Kumar, S. (2021). *3 Techniques to Avoid Overfitting of Decision Trees*. Abgerufen am 11. August 2022 von Towards Data Science: <https://towardsdatascience.com/3-techniques-to-avoid-overfitting-of-decision-trees-1e7d3d985a09>
- MacKenzie, D. I., & Bailey, L. L. (2004). Assessing the fit of site-occupancy models. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(3), S. 300-318.
- MacKenzie, D. I., Nicholas, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8), S. 2248-2255.
- MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L. L., & Hines, J. E. (2018). *Occupancy Estimation and Modeling (Second Edition)*. London: Academic Press.
- Martinez, N., Stickelberger, C., Fässler, F., Strebel, N., & Roth, T. (2020). Occurrence of White-throated Dipper *Cinclus cinclus* and Grey Wagtail *Motacilla cinerea* in relation to the biological quality of rivers. *Ornithologischer Beobachter*(117), S. 164-176.
- McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models*. New York: Springer.
- Moilanen, A. (2002). Implications of empirical data quality to metapopulation model parameter estimation and application. *Oikos*, 96(3), S. 516-530.
- Paterson, J. E. (2020). *Introduction to occupancy models in R*. Abgerufen am 18. 8 2022 von JAMES E PATERSON, PH.D: https://jamesepaterson.github.io/jamespatersonblog/2020-09-01_occupancyintroduction.html
- Powell, L. A., & Gale, G. A. (2015). *Estimation of Parameters for Animal Populations*. Lincoln: Caught Napping Publications.
- Rekha, M. (2020). *Entropy, Information gain and Gini Index; the crux of a Decision Tree*. Abgerufen am 10. August 2022 von The Clairvoyant Vision: <https://blog.clairvoyantsoft.com/entropy-information-gain-and-gini-index-the-crux-of-a-decision-tree-99d0cdc699f4>
- Rockenbauch, D. (1985). Wasseramsel (*Cinclus c. aquaticus*) und Zivilisation - am Beispiel des Fluss-Systems der Fils (Schwäbische Alp). *Ökologie der Vögel*, 7, S. 171-184.

- Serengil, S. I. (2020). *The Math Behind CHAID Decision Tree Algorithm*. Von Sefik Ilkin Serengil: <https://sefiks.com/2020/03/18/a-step-by-step-chaid-decision-tree-example/> abgerufen
- Silipo, R. (2019). *From a Single Decision Tree to a Random Forest*. Abgerufen am 10. August 2022 von Towards Data Science: <https://towardsdatascience.com/from-a-single-decision-tree-to-a-random-forest-b9523be65147>
- Steffen, F. (2020). *Nachweiswahrscheinlichkeit von Hermelinen (*Mustela erminea*) mittels Spurentunnel*. Zürcher Hochschule für Angewandte Wissenschaften ZHAW.
- Strimas-Mackey, M., Hochachka, W. M., Ruiz-Gutierrez, V., Robinson, O. J., Miller, E. T., Auer, T., . . . Johnston, A. (2020). Best Practices for Using eBird Data v1.0. Zenodo. Abgerufen am 7. August 2022 von <https://zenodo.org/record/3620739#.Yu-T-vjP1nI>
- Tyre, A. J., Tenhumberg, B., Field, S. A., Niejalke, D., Parris, K., & Possingham, H. P. (2003). Improving precision and reducing bias in biological surveys: estimating false-negative error rates. *Ecological Applications*, 13(6), S. 1790-1801.
- Wilkinson, D. (2020). *Fitting occupancy models with unmarked*. Abgerufen am 18. 8 2022 von Lodestar: <https://doi90.github.io/lodestar/fitting-occupancy-models-with-unmarked.html>
- Yang, F.-J. (2019). An Extended Idea about Decision Trees. *2019 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE.
- Zopfi, C. (1988). *Territorialität der Wasseramsel (*Cinclus cinclus*) am Küsnachter Bach unter Berücksichtigung der Wasserqualität*. Diplomarbeit, Universität Zürich.

8. Abbildungsverzeichnis

Abbildung 1: Adulte Wasseramsel am Chämpfnerbach (eigenes Bild)	6
Abbildung 2: Veranschaulichung der logit-Funktion. Beziehungen, welche auf der logit-Skala linear sind (links), verlaufen auf der Wahrscheinlichkeits-Skala (rechts) nicht mehr linear und nehmen nur Werte zwischen 0 und 1 an.	10
Abbildung 3: Simples Beispiel eines Decision-Trees: Geprüft wird, ob eine Aussage auf Objekte mit den Attributen x, y, z, q zutrifft oder nicht. Hellgrün = Wurzelknoten, dunkelgrün = Entscheidungsknoten, blau = Blattknoten.....	13
Abbildung 4: Beispiel für mögliche Teilungen des Datensatzes: Split B liefert eine höhere Homogenität der Klassifikationen (Kreis und Quadrat) in den Subsets als Split A.....	13
Abbildung 5: Karte der Beobachtungsstandorte. Der Übersichtlichkeit geschuldet sind nicht alle Standorte markiert. Ein Link zu einer Karte mit allen Standorten ist im Anhang zu finden. Karte von Swisstopo (map.geo.admin.ch (18.08.2022)).....	16
Abbildung 6: Kriterienraster zum Erfassen der Bachstrukturen	17
Abbildung 7: Startsequenz im R-Skript für das grundlegende Modell.	18
Abbildung 8: Erstellen des unmarkedFrame-Objektes und Aufrufen der occu-Funktion.....	18

Abbildung 9: Rücktransformierung der errechneten p- und ψ -Werte mit der predict-Funktion.	18
Abbildung 10: Erstellung verschiedener Modelle mit variierender Detection Probability	19
Abbildung 11: Vergleich der Modelle untereinander anhand des AIC.	19
Abbildung 12: Berechnung der Durchschnittswerte der Parameter derjenigen Modelle, bei denen $\Delta AIC < 2$	19
Abbildung 13: Berücksichtigung der Anzahl besetzter Nachbarstandorte für die Berechnung der Occupancy der Standorte.	20
Abbildung 14: Prüfung der Relevanz verschiedener Bachstrukturen für die Occupancy, unter Verwendung der Funktion dredge().	20
Abbildung 15: Aufrufen der für die Sensitivitätsanalyse notwendigen Funktion mb.gof.test().	20
Abbildung 16: Prozessfenster von RapidMiner Studio zur Verarbeitung der Daten mit dem Cross-Validation-Operator (siehe Abbildung 17).	21
Abbildung 17: Operatoren innerhalb des Cross-Validation-Operators zum Erstellen eines Decision-Trees.	21
Abbildung 18: Karte der Beobachtungsstandorte mit Anzahl positiver Nachweise pro Standort. Karte von Swisstopo (map.geo.admin.ch (21.10.2022)).	22
Abbildung 19: Errechnete Werte für die Occupancy ψ und die Detection Probability p gemäss dem grundlegenden Modell (schwarz). Das 95%-Vertrauensintervall wird durch die roten Balken gekennzeichnet. Die orangen Punkte markieren die «Naïve»-Werte.	22
Abbildung 20: Einfluss der Tageszeit auf die Detection Probability (p) mit grauem 95%-Vertrauensintervall.	23
Abbildung 21: Detection Probability (p) in den Monaten Februar bis Juni mit 95%-Vertrauensintervall.	23
Abbildung 22: Werte der Occupancy bei variierender Anzahl besetzter Nachbarstandorte mit 95%-Vertrauensintervall.	24
Abbildung 23: Occupancy in Abhängigkeit der kategorisierten Naturnähe des Ufers mit 95%-Vertrauensintervall.	25
Abbildung 24: Karte mit allen Beobachtungsstandorten. Die Farbe visualisiert den Wert der Occupancy in Abhängigkeit der Naturnähe des Ufers. Ein schwarzer Punkt markiert einen tatsächlich besetzten Standort. Karte von Swisstopo (map.geo.admin.ch (21.10.2022)).	25
Abbildung 25: Gesamt-Nachweiswahrscheinlichkeit in Abhängigkeit der Anzahl Kontrollgänge. Die grauen Kurven basieren auf Werten der Detection Probability bei unterer und oberer Grenze des Vertrauensintervalls.	26
Abbildung 26: Finaler Decision-Tree. In den Blattknoten ist der prognostizierte Besetzungsstatus (0 oder 1) angegeben. Die Dicke der Linien widerspiegelt die Anzahl Daten, welche die entsprechende Bedingung erfüllen. Der blaue Teil der Balken steht für den Anteil der Daten, welcher in Realität einen Besetzungsstatus 1 aufweist, während der rote Anteil für den Teil der Daten mit realem Status 0 steht.	26

9. Tabellenverzeichnis

Tabelle 1: Beispielswerte für ein zu testendes Kriterium mit Antworten «Hoch», «Mittel» und «Tief» bei möglichen Klassifikationen in «Wahr» und «Falsch».....	15
Tabelle 2: Vergleich verschiedener Modelle mit Covariablen zur Detection Probability. Sie sind geordnet nach ihrem AIC. Als Referenzgrösse wird auch das grundlegende Modell $p(\cdot)$ aufgeführt. w steht für das AIC-Gewicht.....	23
Tabelle 3: Vergleich der vier besten Modelle (nach Ausschluss von Covariablen, welche zu biologisch sinnlosen Modellen führten). «Ufer» steht für die Naturnähe des Ufers, «Bachsohle» für die Qualität des Substrats, «Weg» für dessen Abstand zum Bach und «Turb.» für die Stärke der Turbulenzen. Zusätzlich dargestellt ist das Gewicht der Covariablen innerhalb der einzelnen Modelle. Dieses ist nicht absolut zu verstehen, sondern bildet eine relative Grösse.	24
Tabelle 4: p-Value und c für jedes der getesteten Modelle. In Klammern sind die verwendeten Covariablen für die Occupancy (ψ) und die Detection Probability (p) angegeben. Ein Fehlen von Covariablen wird durch (\cdot) signalisiert.....	25
Tabelle 5: Sicherheit, anwesende Wasseramseln nach bestimmter Anzahl Kontrollgänge nachzuweisen	26

10. Dank

Meinen Eltern und meinem Bruder, die mir bei der Feldarbeit eine echte Unterstützung waren und mich durch wertvolle Hinweise und anregende Gespräche während der ganzen Arbeit weitergebracht haben.

Meinem Betreuer Hanspeter Siegfried, der während dem ganzen Prozess durch wertvolles Feedback eine grosse Hilfe war, mich mit nützlichen Quellen unterstützt hat und als aufmerksamer und kritischer Leser wichtige Rückmeldungen gegeben hat.

Ebenso Marc Kéry von der Schweizerischen Vogelwarte, der mir eine nachträgliche Bestätigung lieferte, dass ich bei meinen Modellierungen nichts übersehen hatte und der mir stets zeitnah auf Fachfragen antwortete.

Und natürlich auch allen anderen, die mich bei meiner Arbeit unterstützt haben, wie beispielsweise Franz Steffen, der mir seine Bachelorarbeit zur Verfügung gestellt hat, welche mir als Grundlage für meine ganze Arbeit diente.

Euch allen danke ich von Herzen!

11. Anhang

Übersicht über alle gesammelten Präsenz-/Absenzdaten

Standort	Begehung 1	Begehung 2	Begehung 3	Begehung 4	Begehung 5	Total
A1	0	0	1	0	0	1
A2	1	1	1	1	0	1
A3	0	0	0	1	1	1
A4	1	0	0	0	1	1
A5	0	0	0	0	0	0
A6	0	0				0
A7	0	0		0	0	0
A8	0	0	0	0	0	0
A9	0	0	0	0	0	0
A9x						
A10	1	1	1	1	0	1
A10x						
A11	1	0	1	1	0	1
B1	0	1	0	1	0	1
B2	1	0	0	1	0	1
B3	1	0	0	1	0	1
B4	0	0	0	0	0	0
B5	1	0	0		1	1
B6	1	0	1	0	0	1
B7	1	0	1	0	0	1
B7x						
B8	1	1	1	0	0	1
B9	0	0	1	0	1	1
B10	1	0	0	0	0	1
B11	1	0	1	0	1	1
C1	0	1	0	0	0	1
C2	0	0	0	0	0	0
C3	0	0	0	0	0	0
C4	0	0	0	0	0	0
C5	0	0	0	0	0	0
C6	0	0	0	0	0	0
C7	0	0	0	0	0	0
C8	0	0	0	0	0	0
C9	0	0	0	0	0	0
C10	0	0	0	0	0	0
C11	0	0	1	0	0	1
C12	0	0	1	0	0	1
D1	0	0	0	0	0	0
D2	0	0	0	0	0	0
D3	0	1	0	0	0	1
D4	1	0	1	1	1	1
D5	0	0	0	1	1	1
D6	0	1	1	0	1	1
D7	0	1	0	0	0	1
D8	1	1	1	0	0	1
D9	1	1	1	1	0	1
D10	0	0	0	0	0	0
D11	1	0	0	1	0	1
D12	1	0	1	1	0	1
E1	0	1	0	0	0	1
E2	1	0	1	1	0	1
E3	1	0	0	1	1	1
E4	1	1	1	1	0	1
E5	1	1	1	0	1	1
E6	1	0	1	0	0	1
E7	0	0	0	0	1	1
E8	1	0	0	0	1	1
E9	0	0	0	0	0	0
E10	0	0	1	0	0	1
E11	0	0	0	0	0	0
E11x						
E12	0	0	0	0	0	0

Übersicht über alle gesammelten Bachstruktur-Daten

Hinweis:

VB = Variabilität der Bachbreite
 Bb = Absolute Bachbreite
 Nb = Nebenbäche
 VW = Variabilität der Wassertiefe
 Tu = Turbulenzen
 St = Steine
 Bs = Bachsohle
 U = Ufer
 Ub = Uferbewuchs
 W = Weg
 (vgl. Abbildung 6)

Ort	VB	Bb	Nb	VW	Tu	St	Bs	U	Ub	W	Ort	VB	Bb	Nb	VW	Tu	St	Bs	U	Ub	W
A1	2	2	1	2	2	2	4	2	3	2	C8	1	2	1	1	1	2	4	2	3	1
A2	2	2	1	3	3	3	4	2	2	2	C9	1	2	1	1	1	3	4	2	3	1
A3	1	2	1	2	2	2	4	2	3	3	C10	1	2	1	2	2	3	4	2	3	2
A4	2	2	1	2	2	3	4	2	1	1	C11	1	2	1	1	2	3	4	2	3	1
A5	2	2	1	2	2	2	4	2	2	2	C12	1	2	1	1	2	3	4	2	3	1
A6	2	2	1	2	2	3	4	2	3	3	D1	2	2	1	3	2	3	4	2	1	2
A7	2	2	1	2	2	3	4	2	3	2	D2	2	2	1	3	3	3	4	3	3	3
A8	2	2	1	2	2	3	4	2	2	2	D3	2	2	1	2	2	2	1	2	3	3
A9	2	2	1	2	2	2	1	2	2	2	D4	2	2	1	3	3	3	4	3	3	3
A10	1	2	1	1	1	2	4	2	2	2	D5	2	2	1	3	2	3	4	3	2	3
A11	2	2	1	2	2	2	4	2	1	1	D6	3	2	1	3	3	3	4	3	2	3
B1	2	2	1	2	2	3	4	2	3	1	D7	3	2	1	3	3	3	4	3	2	2
B2	2	2	1	2	2	3	4	2	3	1	D8	3	2	1	2	3	3	4	3	1	2
B3	2	2	1	1	2	2	4	2	3	3	D9	2	2	1	3	3	3	4	3	2	2
B4	2	2	1	3	2	3	4	2	2	2	D10	3	2	1	3	3	3	4	3	2	2
B5	2	2	1	2	2	3	4	2	3	1	D11	3	2	1	3	3	3	4	3	1	2
B6	2	2	1	2	2	2	4	2	3	2	D12	3	2	1	3	3	3	4	3	2	2
B7	2	2	1	2	1	2	4	2	3	3	E1	3	2	1	3	3	3	4	3	2	2
B8	2	2	1	2	2	3	4	2	3	3	E2	3	2	1	3	2	3	4	3	2	2
B9	1	2	2	2	1	1	4	2	3	2	E3	3	2	1	3	3	3	4	3	2	2
B10	1	2	1	2	1	1	4	2	3	2	E4	3	2	1	3	3	3	4	3	2	3
B11	1	2	1	2	2	1	4	2	3	2	E5	3	2	1	3	3	3	4	3	2	2
C1	2	2	1	2	3	3	4	2	3	1	E6	3	2	2	3	3	3	4	3	2	2
C2	1	2	1	2	2	3	4	2	3	1	E7	2	1	1	3	3	3	4	3	2	2
C3	1	2	1	2	2	3	4	2	3	1	E8	2	2	1	3	3	3	4	3	2	3
C4	2	2	1	2	2	3	4	2	3	2	E9	3	2	1	3	2	2	1	2	3	2
C5	2	2	1	2	2	3	4	2	3	1	E10	3	2	1	3	3	3	4	3	3	3
C6	1	2	1	1	2	2	4	1	2	1	E11	3	2	1	3	3	3	4	3	3	3
C7	1	2	1	1	1	1	4	2	2	1	E12	3	2	1	3	3	3	4	3	3	3

Links

Link zu Karte mit allen Beobachtungsstandorten:

<https://s.geo.admin.ch/9b2c27f23a>

oder

bit.ly/karte-standorte

Link zu allen R-Skripts:

<https://github.com/FlorianH27/Wasseramsel-Site-Occupancy-Modell>

oder

bit.ly/r-skripts

Arbeitsprotokoll

Besprechung 17. Januar Erstes «Sondierungsgespräch»

Vorstellung der ersten Ideen zur Arbeit

Fragestellung:

- Wie gut widerspiegeln Site-Occupancy Modelle das Vorkommen der Wasseramsel?

Ausgangslage:

- Vorkommen der Wasseramsel am Chämptnerbach bekannt und gut nachweisbar
- An anderen Bächen u.a. aufgrund schlechterer Beobachtungsbedingungen weniger untersucht/bekannt (z.B. Aabach)

Ziel:

- Effiziente und standardisierte Methode entwickeln, die Daten zu Wasseramsel Vorkommen generieren soll
- Site-Occupancy-Modellierungen sollen anhand dieser Daten möglichst gutes Abbild der realen Vorkommen ermöglichen
- Überprüfung der Modellierung anhand von Bachabschnitten mit Brutnachweisen (bekanntem Vorkommen)

Feldarbeit 8. Februar Erste Begehung in Sektor B, bachabwärts

Begehung auch als Probebegehung gedacht, falls etwas dringend überdacht werden müsste. Die Methode funktioniert aber gut. Es gelangen überraschend viele Nachweise in schlecht einsehbaren Teilen des Aabachs.

Feldarbeit 12. Februar Sektor A bachabwärts

Feldarbeit 16. Februar Sektor C bachabwärts

Feldarbeit 18. Februar Sektor D bachabwärts (Helfer) und E bachabwärts

Feldarbeit 5. März Sektor B bachaufwärts

Feldarbeit 8. März Sektor A bachaufwärts

Feldarbeit 10. März Sektor C bachaufwärts (Helfer)

Besprechung 11. März zum Ausfüllen des Formulars

Vorschlag H. Siegfried zu Modellierungen mit AI, Decision-Trees

Feldarbeit 13. März Sektor D bachaufwärts (Helfer) und E bachaufwärts

Feldarbeit 11. April Sektor C bachabwärts

Besprechung 14. April Fragen und Abklärungen zu Modellen

Insbesondere zu den Modellierungen mit AI, Decision-Trees

Feldarbeit 18. April Sektor B bachabwärts

Feldarbeit 27. April Sektor D bachabwärts und E bachabwärts (Helfer)

Feldarbeit 28. April Sektor A bachabwärts

Feldarbeit 29. April Sektor C bachaufwärts

Feldarbeit 7. Mai Sektor B bachaufwärts

Feldarbeit 15. Mai Sektor A bachaufwärts und D bachabwärts (Helfer)

Feldarbeit 26. Mai Sektor E bachaufwärts

Eintrag 29. Mai Mit dem Schreiben der Arbeit etwas vorwärts gekommen im Abschnitt Datenaufnahme.

Feldarbeit 4. Juni Sektor D bachaufwärts

Feldarbeit 5. Juni Sektor B bachabwärts (Helfer) und C bachabwärts (Helfer)
Erhebung der Bachstrukturen nach dem Kriterienraster von Hatt (2022)

Feldarbeit 6. Juni Sektor E bachabwärts + Erhebung Bachstrukturen (auch in Sektor D)

Feldarbeit 8. Juni Sektor A bachabwärts + Erhebung Bachstrukturen

Eintrag 12. Juni Schreiben in Kapitel Methode, Abschnitt Datenaufnahme

Besprechung 16. Juni zu weiteren Schritten

Geplante Einarbeitung in die Modellierungen, Programm RapidMiner für das Erstellen des Decision-Trees

Eintrag 23. Juni Keine Reaktion von der Person der Uni auf Anfragen von H. Siegfried

Eintrag 30. Juni Schreiben in Kapitel Methodik, Abschnitt Datenaufnahme

Eintrag 7. Juli Schreiben in Kapitel Grundlagen

Eintrag 11. Juli In der Zwischenzeit recht vorangekommen mit den Site-Occupancy-Modellen.

- Wertvolle Websites gefunden für die Modellierung in R, was das Erstellen der Skripts enorm erleichtert hat.
- Anzahl Kontrollgänge berechnet, «Nachbar-Standort-Modellierung» zeigt klare Resultate
- Bachstrukturen korrelieren aber nicht wirklich mit Wasseramsel-Vorkommen. Es entstehen viele biologisch sinnlose Resultate.
- Decision-Trees scheinen weniger geeignet für Modellierung (tiefe Accuracy). Ebenso viele biologisch sinnlose Befunde.

Besprechung 12. Juli zu Modellierungen

Die Decision-Trees werden nicht weiter vertieft, aber selbstverständlich in der bisherigen Form in der Arbeit belassen, um den Vergleich aufzeigen zu können und die schlechte Eignung diskutieren zu können.

Eintrag 17. Juli Schreiben in Kapitel Grundlagen

Eintrag 21. Juli Mail an Marc Kéry von der Schweizerischen Vogelwarte. Frage, ob bei den gewählten Modellierungen etwas Entscheidendes übersehen wurde oder weitere Modelle zu einem Vergleich hinzugezogen werden sollten

Eintrag 22. Juli Antwort von Marc Kéry: Keine weiteren Modellierungsansätze, Tipps für grafische Darstellung der Resultate

Eintrag 27. Juli In der Zwischenzeit mit Modellierungen weitergekommen, Zeit nach Sonnenaufgang als Covariable für Detection Probability liefert brauchbare Resultate.

Schreiben in Kapitel Methodik, Abschnitt Site-Occupancy-Modell

Eintrag 29. Juli Modell mit Zeit nach Sonnenaufgang als negative Wurzelfunktion ausprobiert, da Wurzelfunktionen aber für negative Werte undefiniert sind, wurde die Idee wieder verworfen.

Schreiben in Kapitel Methodik, Abschnitt Site-Occupancy-Modell

Eintrag 30. Juli Datum als Covariable für Detection Probability angewendet. Mehr Verständnis von ggplot zum Erstellen der Grafiken gewonnen.

Eintrag 2. August In der Zwischenzeit Weiterführung von Datums-Covariablen, Schreiben in Kapitel Methodik, Abschnitt Site-Occupancy-Modell, Schreiben im Kapitel Grundlagen

Eintrag 3 - 5. August Schreiben im Kapitel Grundlagen

Eintrag 7. August Grundlagen Site-Occupancy-Modelle als erste Version fertig

Eintrag 8 - 11. August Schreiben im Kapitel Grundlagen im Abschnitt Decision-Tree

Eintrag 12. August Grundlagen Decision-Tree überarbeiten

Schreiben in Kapitel Methodik im Abschnitt Site-Occupancy-Modelle

Eintrag 23. August In der Zwischenzeit diverse Überarbeitungen

Eintrag 27. August Schreiben im Kapitel Grundlagen in Abschnitten Wasseramsel

Eintrag 31. August Korrekturabgabe Grundlagen

Besprechung 15. September Korrektur

In Zwischenzeit im Kapitel Resultate weitergeschrieben

Eintrag 24. September In der Zwischenzeit im Kapitel Resultate weitergeschrieben, Karte für Occupancy erstellt

Eintrag 19. Oktober In der Zwischenzeit Resultate überarbeitet und abgeschlossen, Karte für Nachweise erstellt. Mit der Diskussion bereits sehr weit, erste Teile auch schon überarbeitet

Eintrag 20. Oktober Diskussion fertig und überarbeitet

Eintrag 21. Oktober Abstract, Schlusswort, kleinere Überarbeitungen (Layout, Bereinigung von Fehler durch automatische Word-Zitierfunktion, etc.)