



**Universität
Zürich** UZH

**Statistische Analyse und Vergleich ausgewählter Methoden
zur Segmentierung von Trajektorien
anhand von Simulationen**

GEO 511 Masterarbeit

Autorin

Myriam Braun

13-732-045

Betreut durch

Prof. Dr. Robert Weibel (Fakultätsverantwortlicher)

Oliver Burkhard

Dr. Chloe Bracis

Datum der Abgabe: 30. September 2018

Geographisches Institut Universität Zürich

Kontakt

Autorin

Myriam Braun

Breitibachstrasse 29
CH-8600 Dübendorf
myriam.braun@geo.uzh.ch

Betreuer & Betreuerin

Prof. Dr. Robert Weibel

Universität Zürich
Geografisches Institut
Winterthurerstrasse 190
CH-8057 Zürich
robert.weibel@geo.uzh.ch

Oliver Burkhard

Universität Zürich
Geografisches Institut
Winterthurerstrasse 190
CH-8057 Zürich
oliver.burkhard@geo.uzh.ch

Dr. Chloe Bracis

Senckenberg Gesellschaft für Naturforschung &
Institut Français de Recherche pour l'Exploitation de la Mer (Ifremer)
Centre Manche Mer du Nord
150, Quai Gambetta
FR-62200 Boulogne-sur-Mer
chloe.bracis@ifremer.fr

ZUSAMMENFASSUNG

Das Verhalten von Tieren interessiert Ökologinnen und Ökologen schon seit sehr langer Zeit. Mit der technischen Entwicklung ist es nun möglich, das Verhalten eines Tieres nicht nur durch Beobachtung zu erforschen, sondern auch mit technischen Hilfsmitteln, beispielsweise kleinen GPS Sensoren. Die Schwierigkeit besteht nun jedoch darin, das Verhalten eines Individuums aus den vorhandenen Bewegungsdaten abzuleiten. Es wurden bereits unterschiedlichste Methoden entwickelt, um anhand von Zeit und Koordinaten eine Trajektorie zu segmentieren. Ziel der Segmentierung ist es, möglichst homogene Segmente zu erhalten, die je ein Verhaltensmuster beschreiben. Die Methoden basieren auf unterschiedlichen mathematischen Grundlagen, die sowohl zu Stärken als auch zu Schwächen der jeweiligen Methode führen können. Um die am besten geeignete Methode für einen bestimmten Datensatz oder für eine spezifische Fragestellung auszuwählen, müssen die Methoden miteinander verglichen werden. Bisher wurden jedoch erst zwei Vergleichspublikationen veröffentlicht, in denen die Resultate verschiedener Methoden visuell verglichen wurden. Daher beschäftigt sich die vorliegende Arbeit mit dem statistischen Vergleich zweier Methoden (FPT und BCPA). Da Simulationen dazu eine gute Hilfestellung bieten, werden keine realen Datensätze analysiert, sondern die Eigenschaften der Methoden werden aufgrund von Simulationen identifiziert, evaluiert und verglichen. Dazu werden die Resultate der beiden bestehenden Publikationen mit den Resultaten dieses Projekts verglichen. Zudem werden 50 Realisierungen der Simulationen aufgrund eines Rauschens produziert, maximale Abweichungen in die Bewertung miteinbezogen, die Werte der Parameter der Methoden verglichen sowie die Kombination von zwei verschiedenen Simulationen analysiert.

Mit diesem Vorgehen konnte festgestellt werden, dass beide Methoden in der Lage sind, starke Geschwindigkeitsänderungen in den Bewegungsdaten zu erkennen. Während die BCPA auch schwächere Veränderungen meist in derselben Segmentierung identifizieren konnte, musste für die FPT eine erneute Segmentierung eines einzelnen Segmentes durchgeführt werden. Veränderungen in der Tortuosität sowie kontinuierliche Geschwindigkeitsänderungen waren für die FPT nicht erkennbar, da sie in ihre Analyse einzig die Geschwindigkeit miteinbezieht. Die BCPA hingegen konnte auch Veränderungen in der Tortuosität erkennen, nicht aber kontinuierliche Geschwindigkeitsänderungen. Des Weiteren kann gesagt werden, dass der Einbezug einer maximalen Abweichung die Resultate stark verbessern kann. Trotzdem wird es bei der Anwendung der Methoden auf reale Datensätze grosse Unterschiede geben und eine visuelle Sichtung der Resultate unter Einbezug des ökologischen Hintergrundwissens über die untersuchte Tierart wird weiterhin stets notwendig sein.

Schlagwörter

Movement Ecology, Movement Analysis, Bewegungsökologie, Bewegungsanalyse, FPT, BCPA, Simulation, Ökologie

DANKSAGUNG

Es waren sechs lehrreiche Jahre am GIUZ, mit vielen neuen Bekanntschaften und hoffentlich langjährigen tiefen Freundschaften, die nun zu Ende gehen. Diese Arbeit gibt mir die Gelegenheit, mich bei all den Personen zu bedanken, die mich auf dem Weg durch das Studium unterstützt haben.

Als erstes möchte ich mich bei meinen beiden Betreuern Röbi und Oliver sowie meiner externen Betreuerin Chloé für ihre wertvolle Unterstützung bedanken. Ich bin sehr dankbar, in diesem Jahr drei Experten an meiner Seite gehabt zu haben. Eure vielfältigen Ideen halfen mir während dem gesamten Arbeitsprozess sehr – nicht zuletzt auch die Hilfe bei den R-Problemen, die ich dank dir, Oliver, innert vernünftiger Frist lösen konnte.

Des Weiteren möchte ich mich bei meinen Eltern bedanken, ohne die ich bestimmt nicht an diesem Punkt in meinem Leben stehen würde. Eure Geduld, eure Liebe und Aufmerksamkeit haben mich bis hierhin getragen. Und wenn einmal mehr die Zeit des Lernens, der Prüfungen oder die Zeit des Arbeit-Schreibens anstand, habt ihr mich immer mit Köstlichkeiten gepflegt und mich unterstützt wo immer ihr konntet.

Eveline, mis liebe Schwöschterherzli, ich danke dir für deine Geduld. Wenn du dich mit mir hingesezt hast, um mein Thema zu verstehen. Wenn ich nicht wusste, wie ich meine Frage stellen, mein Problem beschreiben oder die Antwort auf eine deiner Fragen von dir formulieren sollte. Danke für deine umfangreiche Unterstützung, inklusive Korrekturen. Und natürlich auch ein grosses Dankeschön an dich, Damian, für deine Motivation und die Verköstigung während dieser Stunden. Ab nach Weggis, go Fischnusperli ässe!

Ein grosses Dankeschön geht auch an dich, Thomas. Ich weiss, dass ich diese Arbeit nicht vollendet hätte, wenn du mich nicht immer wieder aufgebaut und an mich geglaubt hättest. All die unzähligen Male, in denen du mich in deinen Armen gehalten und geduldig gewartet hast, bis meine Tränen versiegeten. Und all die vielen Male, in denen ich dich angeschnauzt habe, weil ich unzufrieden war. Es tut mir leid. Aber jetzt startet ein neuer Abschnitt, uf gaats zu neue Abentüür!!

Tja, dann möchte ich mich noch bei meinen beiden «GIUZ-Überbleibseln», Susanne und Vera bedanken. Fünf Jahre gemeinsames Studium, gemeinsame Ferien, Wanderungen und Erlebnisse. Und natürlich für das letzte Jahr, in dem wir viel Neues gelernt haben (z. B. dass Elefanten entweder Links- oder Rechtsrüssler sind oder dass Rehe farbenblind sind) und uns gegen Fehlermeldungen des R verbünden konnten. Ohne euch würde diese Arbeit jetzt nicht vor mir liegen.

Zu guter Letzt möchte ich mich auch bei allen weiteren Personen bedanken, die mich durch das Studium und die Masterarbeit getragen haben. Es war eine intensive, lehrreiche Zeit und ich hoffe, dass wir in Kontakt bleiben oder uns wieder einmal über den Weg laufen werden.

INHALT

ABBILDUNGSVERZEICHNIS	IX
TABELLENVERZEICHNIS	XII
ABKÜRZUNGEN	XIV
1 EINLEITUNG	1
1.1 Problemstellung.....	2
1.2 Zielsetzung und Umfang dieses Projekts.....	2
1.3 Aufbau dieser Arbeit.....	3
2 HINTERGRUND	5
2.1 Theoretischer Hintergrund.....	5
2.1.1 Movement Ecology – Bewegungsökologie.....	5
2.1.2 Movement Analysis – Bewegungsanalyse.....	8
2.1.3 Trajektorien.....	9
2.1.4 Path Characteristics.....	11
2.1.5 Simulationen.....	12
2.1.6 Segmentierung.....	15
2.2 Stand der Forschung.....	16
2.2.1 Methoden der Bewegungsanalyse.....	16
2.3 Bisherige Vergleichsstudien.....	20
2.4 Forschungslücke.....	22
2.5 Forschungsfragen.....	23
3 MATERIAL	25
3.1 Verwendete Simulationen.....	25
3.1.1 Definition der Parameter.....	25
3.2 Verwendete Methoden.....	27

INHALT

3.2.1	First Passage Time (FPT)	28
3.2.2	Behavioural Change Point Analysis (BCPA)	30
4	METHODIK	32
4.1	Methoden	32
4.1.1	Methodischer Rahmen	32
4.1.2	Beschreibung der Funktionen	32
4.1.3	Packages	36
4.1.4	Evaluationskriterien	36
4.2	Vorgehen	38
4.2.1	Repliziert	38
4.2.2	50 Realisierungen	39
4.2.3	Maximale Abweichung	39
4.2.4	FPT: mittlerer CP	40
4.2.5	BCPA: Wertevergleich der Parameter	40
4.2.6	Kombination der Speed & Time-scale switch Simulation	41
5	RESULTATE	42
5.1	Repliziert	42
5.2	50 Realisierungen	58
5.3	Maximale Abweichung	62
5.4	FPT: mittlerer CP	68
5.5	BCPA: Wertevergleich der Parameter	74
5.6	Kombination der Speed & Time-scale switch Simulation	83
6	DISKUSSION	87
6.1	Repliziert	87
6.2	50 Realisierungen	92
6.3	Maximale Abweichung	93

INHALT

6.4	FPT: mittlerer CP.....	95
6.5	BCPA: Wertevergleich der Parameter	96
6.6	Kombination der Speed & Time-scale switch Simulation	98
6.7	Forschungsfragen.....	100
6.8	Kritische Beurteilung des Projekts	103
7	FAZIT.....	105
7.1	Ausblick.....	107
8	LITERATUR.....	109
A	ANHANG	114
	EIGENSTÄNDIGKEITSERKLÄRUNG.....	115

ABBILDUNGSVERZEICHNIS

- Abbildung 1: Konzept der Bewegungsökologie von Nathan et al. (2008). Es zeigt drei Grundkomponenten (gelber Hintergrund) einer Bewegung eines Individuums (internal state, motion capacity und navigation capacity) und eine externe Komponente (blauer Hintergrund), welche die anderen Komponenten beeinflusst (external factors). Die Beziehungen zwischen den Komponenten werden mit Pfeilen dargestellt, um die Richtung der Beeinflussung anzugeben. Der daraus resultierende Bewegungspfad gibt Informationen an den internal state sowie die externen Faktoren zurück. 7
- Abbildung 2: Datensätze vorhanden auf Movebank. Grün: frei verfügbare Datensätze, grau: Anfrage bei Inhaber der Daten notwendig. (Max Planck Institute of Ornithology 2018, Zugriff: 1.9.2018)..... 10
- Abbildung 3: Aufbau des methodischen Rahmens. Die grosse Box zeigt den Inhalt der alles umfassenden Funktion (call Funktion). Im Schritt «Simulation generieren» sowie im Schritt «Resultate produzieren» wird jeweils ein csv-Dokument abgespeichert. Der Segmentierungsschritt beinhaltet die beiden Methoden FPT und BCPA. Ausserhalb der grossen Box werden dann statistische Auswertungen vorgenommen und Resultate visualisiert..... 32
- Abbildung 4: Beispiel zweier simulierten Trajektorien. Start ist jeweils bei Koordinate (0,0). Die Farben zeigen unterschiedliche Segmente. Durch das Rauschen in der Simulation sehen die Trajektorien mit jeder Realisierung anders aus..... 39
- Abbildung 5: Trajektorie der speed switch Simulation (oben). Orange zeigt die erste und letzte Phase, blau die zweite und pink die dritte. Start ist bei Koordinate (0,0). Dieselbe Trajektorie mit eingezeichneten CP (unten). Definierte CP sind als roter und grosser Punkt markiert. Durch FPT identifizierte CP in grün. Durch BCPA identifizierte CP in blau..... 43
- Abbildung 6: FPT der speed switch Simulation. X-Achse ist die Zeit [min], y-Achse die FPT. Definierte CP sind die roten vertikalen Linien, durch FPT identifizierte CP sind die grün gestrichelten vertikalen Linien..... 44
- Abbildung 7: Grafik der BCPA der speed switch Simulation. X-Achse ist die Zeit [min], y-Achse die persistence velocity [map unit/min]. Oben repräsentieren die violetten vertikalen Linien die durch die BCPA identifizierten CP, die horizontalen schwarze Linien der geschätzte Mittelwert, die horizontalen roten Linien das 95% Konfidenzintervall. Die farbigen Punkte repräsentiere die erwartete time-scale. Je kühler die Farbe desto tiefer die Werte und desto höher ist die Tortuosität (und umgekehrt). Unten repräsentieren die roten vertikalen Linien die korrekten CP und die blau gestrichelten die durch die BCPA identifizierten CP. 45
- Abbildung 8: Turning angle (oben) und geschätzte Geschwindigkeit [map unit/min] (unten) für die speed switch Simulation. X-Achse ist die Zeit [min]. Rote vertikale Linien sind die korrekten CP, grün gestrichelt die durch die FPT identifizierten CP und blau gestrichelt die durch die BCPA identifizierten CP. 47
- Abbildung 9: Trajektorie der time-scale switch Simulation (oben). Orange zeigt die erste und letzte Phase, blau die zweite und pink die dritte. Start ist bei Koordinate (0,0). Trajektorie (gleiche wie oben) mit eingezeichneten CP (unten). Definierte CP roter und grosser Punkt. Durch FPT identifizierte CP in grün und durch BCPA identifizierte CP in blau..... 48
- Abbildung 10: Grafik der FPT der time-scale switch Simulation. X-Achse ist die Zeit [min], y-Achse die FPT. Definierte CP sind die roten vertikalen Linien, durch die FPT identifizierte CP sind die grün gestrichelten, vertikalen Linien..... 49
- Abbildung 11: Grafik der BCPA der time-scale switch Simulation. X-Achse ist die Zeit [min], y-Achse die persistence velocity [map unit/min]. Oben repräsentieren die violetten vertikalen Linien die

ABBILDUNGSVERZEICHNIS

identifizierten CP, die horizontalen schwarze Linien der geschätzte Mittelwert, die horizontalen roten Linien das 95% Konfidenzintervall. Die farbigen Punkte repräsentieren die erwartete time-scale. Je kühler die Farbe desto tiefer die Werte und desto höher ist die Tortuosität (und umgekehrt). Unten repräsentieren die roten vertikalen Linien die korrekten CP und die blau gestrichelten die durch die BCPA identifizierten CP.....	50
Abbildung 12: Turning angle (oben) und geschätzte Geschwindigkeit [map unit/min] (unten) für die time-scale switch Simulation. X-Achse ist die Zeit [min]. Rote vertikale Linien sind die korrekten CP, grün gestrichelte Linien sind die durch die FPT identifizierten CP und blau gestrichelt die durch die BCPA identifizierten CP.....	52
Abbildung 13 Trajektorie der continuous speed switch Simulation (oben). Orange zeigt die erste und letzte Phase, blau die zweite und pink die dritte Phase. Start ist bei Koordinate (0,0). CP derselben Trajektorie (unten). Definierte CP als rote grosse Punkte. Durch FPT identifizierte CP in grün und durch BCPA identifizierte CP in blau.	53
Abbildung 14: Grafik der FPT der continuous speed switch Simulation. X-Achse ist die Zeit [h], y-Achse die FPT. Definierte CP sind die roten vertikalen Linien, durch die FPT identifizierten CP sind die grün gestrichelten vertikalen Linien.....	54
Abbildung 15: Grafik der BCPA der continuous speed switch Simulation. X-Achse ist die Zeit [h], y-Achse die persistence velocity [map unit/min]. Oben repräsentieren die violetten vertikalen Linien die identifizierten CP, die horizontalen schwarze Linien der geschätzte Mittelwert, die horizontalen roten Linien das 95% Konfidenzintervall. Die farbigen Punkte repräsentiere die erwartete time-scale. Je kühler die Farbe desto tiefer die Werte und desto höher die Tortuosität (und umgekehrt). Unten repräsentieren die roten vertikalen Linien die korrekten CP und die blau gestrichelten die durch die BCPA identifizierten CP.....	55
Abbildung 16: Turning angle (oben) und geschätzte Geschwindigkeit [map unit/min] (unten) für die continuous speed switch Simulation. X-Achse ist die Zeit [h]. Rote vertikale Linien sind die korrekten CP, grün gestrichelt die durch die FPT identifizierten CP und blau gestrichelt die durch die BCPA identifizierten CP.....	57
Abbildung 17: Boxplot der Anzahl Segmente und Anzahl CP für die speed switch Simulation jeweils für die Simulation, FPT und BCPA.....	58
Abbildung 18: Boxplot der Anzahl Segmente und Anzahl CP für die time-scale switch Simulation jeweils für die Simulation, FPT und BCPA.....	60
Abbildung 19: Boxplot der Anzahl Segmente und Anzahl CP für die continuous speed switch Simulation jeweils für die Simulation, FPT und BCPA.....	61
Abbildung 20:Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.....	63
Abbildung 21: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.....	65
Abbildung 22: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.....	67
Abbildung 23: speed switch Änderung 1: Anzahl korrekte CP basierend auf unterschiedlichen Radien der FPT. Links: exakt identifizierte CP, rechts: mit maximaler Abweichung von ± 3 identifizierte CP. ...	69

ABBILDUNGSVERZEICHNIS

Abbildung 24: speed switch Änderung 2: Anzahl korrekte CP basierend auf unterschiedlichen Radien der FPT. Links: exakt identifizierte CP, rechts: mit maximaler Abweichung von ± 3 identifizierte CP. ...	70
Abbildung 25: speed switch Änderung 3: Anzahl korrekte CP basierend auf unterschiedlichen Radien der FPT. Links: exakt identifizierte CP, rechts: mit maximaler Abweichung von ± 3 identifizierte CP. ...	72
Abbildung 26: Boxplot zu Precision (links) und Recall (rechts) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 segmentiert durch die FPT.....	74
Abbildung 27: Wertevergleich der Parameter der BCPA für die speed switch Simulation. X-Achse oben: Sensitivitätsparameter K, X-Achse unten: Clusterwidth, Y-Achse links: Anzahl exakt korrekt identifizierte CP, y-Achse rechts: window size.	76
Abbildung 28: Wertevergleich der Parameter der BCPA für die speed switch Simulation. X-Achse oben: Sensitivitätsparameter K, X-Achse unten: Clusterwidth, Y-Achse links: Anzahl korrekt identifizierte CP mit Abweichung ± 3 , y-Achse rechts: window size.	77
Abbildung 29: Wertevergleich der Parameter der BCPA für die time-scale switch Simulation. X-Achse oben: Sensitivitätsparameter K, X-Achse unten: Clusterwidth, Y-Achse links: Anzahl exakt korrekt identifizierter CP, y-Achse rechts: window size.	79
Abbildung 30: Wertevergleich der Parameter der BCPA für die time-scale switch Simulation. X-Achse oben: Sensitivitätsparameter K, X-Achse unten: Clusterwidth, Y-Achse links: Anzahl korrekt identifizierter CP mit Abweichung ± 3 , y-Achse rechts: window size.....	80
Abbildung 31: Kombination der speed switch und der time-scale switch Simulation: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.....	85
Abbildung 32: Kombination der speed switch und der time-scale switch Simulation: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.....	85
Abbildung 33: Abgeänderte Darstellung aus Gurarie et al. (2016a). Sie zeigt die simulierten Trajektorien (oben) und die Resultate der beiden Methoden (FPT mittig, BCPA unten). In den Trajektorien kennzeichnet blau die erste und letzte Phase, grün die zweite und rot die dritte. Die vertikalen Linien (unten) zeigen die CP, die horizontalen schwarzen Linien den geschätzten Mittelwert, die horizontalen roten Linien repräsentieren das 95% Konfidenzintervall. Hellere Farben der Punkte bedeuten tiefere Werte der time-scale und damit eine höhere Tortuosität. Die Hintergrundfarben präsentieren die vier Segmente der Simulation.....	90
Abbildung 34: Auswirkungen der Parameter der BCPA (clusterwidth, Sensitivitätsparameter K und window size ws) auf die Anzahl Segmente dreier Trajektorien von subadulten, männlichen Wildschweinen (Morelle et al. 2017a).	97

TABELLENVERZEICHNIS

Tabelle 1: Zusammenfassende Tabelle von in vier Kategorien eingeteilte Methoden der Bewegungsanalyse aus Gurarie et al. (2016a). Fett geschrieben sind die vier Methoden, welche Gurarie et al. (2016a) in ihrer Publikation verwendeten. Die Methoden der letzten Gruppe sind multistate random walks, hidden Markov models oder state space models.	17
Tabelle 2: Parameter der speed switch, time-scale switch (Gurarie et al. 2016a) und der continuous speed switch (Soleymani et al. 2017) Simulation.	26
Tabelle 3: Parameter für die drei veränderten speed switch Simulationen.	27
Tabelle 4: Parameter für die speed und time-scale switch Simulation. Diese kombiniert die jeweiligen Veränderungen der Parameter aus der speed switch und der time-scale switch Simulation von Gurarie et al. (2016a).	27
Tabelle 5: In diesem Projekt verwendete R-Packages.	36
Tabelle 6: Daten der speed switch Simulation und der durch die FPT und BCPA segmentierten Trajektorie der speed switch Simulation.	46
Tabelle 7: Daten der time-scale switch Simulation und der durch die FPT und BCPA segmentierten Trajektorie der time-scale switch Simulation.	51
Tabelle 8: Daten der continuous speed switch Simulation und der durch die FPT und BCPA segmentierten Trajektorie der continuous speed switch Simulation.	56
Tabelle 9: Zusammenfassende Statistik der speed switch Simulation segmentiert durch die FPT (oben) und BCPA (unten).	59
Tabelle 10: Zusammenfassende Statistik der time-scale switch Simulation segmentiert durch die FPT (oben) und BCPA (unten).	60
Tabelle 11: Zusammenfassende Statistik der continuous speed switch Simulation segmentiert durch die FPT (oben) und BCPA (unten).	62
Tabelle 12: Anzahl Realisierungen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Simulationen, die zwischen den korrekten CP weitere CP identifiziert haben.	64
Tabelle 13: Anzahl Simulationen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Simulationen, die zwischen den korrekten CP weitere CP identifiziert haben.	65
Tabelle 14: Anzahl Realisierungen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Realisierungen, die zwischen den korrekten CP weitere CP identifiziert haben.	66
Tabelle 15: speed switch Änderung 1: Anzahl Realisierungen bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert wurden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert wurden. Radius der FPT gleich 10.	68
Tabelle 16: speed switch change 1: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. Radius der FPT gleich 10.	69
Tabelle 17: speed switch Änderung 2: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert wurden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert wurden. Radius der FPT gleich 10.	70

TABELLENVERZEICHNIS

Tabelle 18: speed switch Änderung 2: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert wurden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert wurden. Radius der FPT gleich 5.....	71
Tabelle 19: speed switch change 3: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. Radius der FPT gleich 10.....	71
Tabelle 20: Zusammenfassende Statistik der speed switch Simulation mit Radius 4 segmentiert durch die FPT: exakt (oben) und mit maximaler Abweichung von ± 3 (unten).....	73
Tabelle 21: speed switch: Anzahl Simulationen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Simulationen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. Radius der FPT beträgt 4.....	74
Tabelle 22: Zusammenfassende Statistik der speed switch Simulation mit $K = 2$ segmentiert durch die BCPA: exakt (oben) und mit maximaler Abweichung von ± 3 (unten).	81
Tabelle 23: speed switch: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. K der BCPA gleich 2.	82
Tabelle 24: Zusammenfassende Statistik der time-scale switch Simulation mit $K = 2$ segmentiert durch die BCPA: exakt (oben) und mit maximaler Abweichung von ± 3 (unten).	82
Tabelle 25: time-scale switch: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. K der BCPA gleich 2.	83
Tabelle 26: Zusammenfassende Statistik der speed und time-scale switch Simulation segmentiert durch die FPT (oben) und die BCPA (unten). Anzahl korrekt identifizierter CP ohne Abweichung (exakt).	83
Tabelle 27: Anzahl Realisierungen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Realisierungen, die zwischen den korrekten CP weitere CP identifiziert haben.....	86

ABKÜRZUNGEN

ARS	<i>Area restricted search</i>
BCPA	<i>Behavioural Change Point Analysis</i>
BIC	<i>Bayesian Information Criterion</i>
BPMM	<i>Bayesian partitioning of Markov models</i>
CP	<i>Change Point</i>
CRW	<i>Correlated random walks</i>
CVM	<i>Correlated velocity movement</i>
D	<i>Fractal dimension</i>
DWT	<i>Discrete wavelet transform</i>
FPT	<i>First Passage Time</i>
GPS	<i>Global positioning systems</i>
HMM	<i>Hidden Markov model</i>
MRW	<i>Multistate random walk</i>
NSD	<i>Net squared displacement</i>
PCM	<i>Penalized contrast method</i>
RTG	<i>Random trajectory generator</i>
RT	<i>Residence Time</i>
RW	<i>Random walks</i>
SSM	<i>State-space models</i>
VHF	<i>Very high frequency</i>

1 EINLEITUNG

Bewegung ist eine wichtige Grundlage des Lebens. Nicht nur von uns Menschen – auch das Leben von Pflanzen und Tieren ist durch die Bewegung geprägt. Unterschiedlichste Formen von Bewegung ermöglichen eine Nutzung des gesamten vorhandenen Raums, sei es in der Luft, am oder im Boden, im Wasser. Denn erst die Bewegung ermöglicht die Interaktion zwischen Individuen, innerhalb oder zwischen unterschiedlichen Spezies sowie mit der Umwelt. Daher ist auch nicht verwunderlich, dass unterschiedlichste Disziplinen der Forschung Interesse am Thema «Bewegung» zeigen. Dazu gehören zum Beispiel der Transport, die menschliche Mobilität, verschiedene Verhaltens- und Umweltstudien, die Informationswissenschaften oder die Bewegungsökologie (*movement ecology*) (Dodge et al. 2016). Letztere ist die Grundlage dieser Arbeit und wird in Kapitel 2.1.1 genauer erläutert.

Insbesondere das Interesse am Verhalten von Tieren hat in den letzten Jahren zugenommen. Immer tiefer und vollständiger beschäftigt sich die Forschung mit Fragen zu «warum», «wie», «wann» und «wohin» sich Tiere bewegen und wie dieses Verhalten mit externen Faktoren zusammenhängt (Dray et al. 2010, Nathan et al. 2008). Durch welche Prozesse werden Bewegungen ausgelöst und welchen Einfluss haben die unterschiedlichen Grössen? Wie und weshalb nutzen Tiere gewisse Ressourcen? Wie verwerten und reproduzieren sie diese? Wie interagieren die Tiere untereinander oder wie verhalten sie sich in Kontakt mit anderen Spezies? (Demšar et al. 2015) Aber auch Fragen zur Verteilung von Populationen und ihren Dynamiken sind für Ökologen von grossem Interesse (Mueller & Fagan 2008).

Solche Fragen werden schon seit langer Zeit von Forscherinnen und Forschern untersucht. Die Analyse von Bewegung hilft dem Menschen, diese Fragen zu beantworten. Beispielsweise lassen sich Entscheidungen bezüglich einer Bewegung beziehungsweise deren beeinflussenden Faktoren erkennen und verstehen. Doch erst mit der technischen Entwicklung wurde es für Forscher möglich, auch Bewegungen kleinerer oder scheuer Tiere sowie Bewegungen in unwegsamen Lebensräumen zu erforschen. Sensoren, wie zum Beispiel GPS-Geräte, wurden an Tieren befestigt um ihre Bewegungen räumlich und zeitlich aufzuzeichnen. Die Vorteile der technologischen Entwicklung führten zu einer Zunahme am Interesse der Bewegungsdaten und deren Analyse. Mit Hilfe solcher Daten lassen sich Prozesse der Ökologie und der Evolution besser und detaillierter verstehen und erklären.

Um Bewegungsdaten zu analysieren, gibt es unterschiedliche Vorgehensweisen. Eine davon ist die Segmentierung. Die Methoden zur Segmentierung eines Bewegungspfades (Trajektorie) basieren auf verschiedenen Bewegungscharakteristiken. Anhand dieser Charakteristiken unterteilen die Segmentierungsmethoden eine Trajektorie in einzelne Segmente. Diese Segmente zeigen jeweils eine spezifische Bewegungsart und erlauben so, ökologische Fragen betreffend Zeit und Ort zu beantworten.

1.1 Problemstellung

Bis zum heutigen Zeitpunkt besteht die Literatur hauptsächlich aus Publikationen aus drei verschiedenen Richtungen. Die erste Richtung beinhaltet Publikationen zu spezifischen ökologischen Fragen, welche mit Methoden der Bewegungsanalyse beantwortet werden sollen. Die zweite Richtung besteht aus Publikationen, welche neue Methoden entwickeln, um dem eigenen Bedürfnis der Fragestellung zu genügen. Diese werden an einem realen Datensatz oder allenfalls an passenden Simulationen präsentiert. Während der letzten Jahre führte die Entwicklung neuer Methoden zu einer immer länger werdenden Liste an möglichen Methoden. Dadurch wurde die Wahl einer Methode zu einem langwierigen Prozess und oftmals erfolgte die schnelle Wahl auf eine wenig nützliche Methode. Ein Vergleich der unterschiedlichen Methoden wäre also sehr hilfreich. Vergleichende Studien machen die dritte Richtung aus und sind noch immer rar, denn dazu wurde erst wenig publiziert.

Bisher wurden zwei Publikationen veröffentlicht, in welchen unterschiedliche Segmentierungsmethoden visuell miteinander verglichen werden. Diese beiden Vergleichspublikationen wurden von Gurarie et al. (2016a) und Edelhoff et al. (2016a) durchgeführt (siehe Kapitel 2.3). Ein Grund für solch wenige vergleichende Publikationen ist vermutlich die Menge sowie die Komplexität der Methoden, aber auch die vielfältigen Eigenschaften der Bewegungen und die grosse Breite an möglichen Forschungsfragen. Weitere Vergleiche würden jedoch die Wahl der Methode für eine Ökologin oder einen Ökologen vereinfachen und die Resultate der Segmentierung verbessern. Damit könnten Forschungsfragen schneller, genauer und informationsreicher beantwortet werden.

1.2 Zielsetzung und Umfang dieses Projekts

Das Hauptziel dieses Projekts ist der Vergleich zweier Methoden zur Analyse von Bewegungsdaten, in Bezug auf zwei verschiedene Bewegungscharakteristiken. Der Schwerpunkt liegt dabei auf den Veränderungen der Geschwindigkeit und der Tortuosität. Dabei werden die Vor- und Nachteile jeder Methode dargelegt. Dieser Vergleich soll den Ökologinnen und Ökologen einerseits eine Hilfestellung geben, um geeignete Werte für die Parameter der Methoden auszuwählen und andererseits zeigen, wie robust sich die Methoden bezüglich des Rauschens in einer Trajektorie verhalten. Um dieses Ziel zu erreichen, werden mehrere aufeinanderfolgende Schritte durchgeführt. Diese haben sich während des ganzen Projekts laufend entwickelt. In einem ersten Schritt werden Simulationen aus zwei Publikationen repliziert. Aufgrund des in den Simulationen enthaltenen Rauschens werden jeweils 50 Realisierungen der Simulationen durchgeführt. Anschliessend wird der Einfluss von maximalen Abweichungen analysiert. Um allfällig bessere Wertekombinationen für die Parameter der Methoden zu finden, werden diese für die beiden Methoden grafisch verglichen. Dann werden zwei der drei (speed switch und time-scale switch)

EINLEITUNG

Simulationen kombiniert und diese Resultate ebenfalls analysiert. Die Resultate all dieser Schritte werden in einen Kontext gesetzt und Schlussfolgerungen zur Leistung der beiden Methoden gezogen.

In diesem Projekt werden die Simulationen aus den Publikationen von Gurarie et al. (2016a) und Soleymani et al. (2017) verwendet. Bei den gewählten Simulationen kann zum einen genau gesagt werden, ob die Methoden robust gegenüber einem Rauschen sind und zum anderen ist definiert, inwiefern sie die Geschwindigkeit oder Tortuosität miteinbeziehen. Weitere Simulationsarten und Publikationen sind in diesem Projekt nicht berücksichtigt. Ebenso werden keine neuen Methoden oder andere Arten von Simulationen entwickelt, sondern bestehende Methoden werden an bestehenden Simulationen ausgeführt. Diese werden evaluiert, um bei späteren ökologischen Analysen informativere Resultate mit höherer Qualität generieren zu können. Dazu soll das bestehende Wissen über die Leistung der beiden Methoden sowie deren Robustheit gegenüber einem Rauschen erweitert werden. Auf der einen Seite kann dies helfen, die Wahl einer Methode für eine Ökologin oder einen Ökologen zu erleichtern oder eine bereits getroffene Wahl zu überprüfen. Auf der anderen Seite können zukünftige Resultate in der Bewegungsanalyse verbessert und damit informativer werden.

Nicht zuletzt soll der Ablauf der Analysen dieses Projekts helfen, weitere Methoden auf dieselbe Weise zu analysieren und zu vergleichen. Ein einheitliches Vorgehen ist daher unumgänglich. Aus diesen Gründen wird der Programmierung des gesamten R-Codes ein grosses Gewicht gegeben. Die einzelnen Funktionen werden detailliert beschrieben und der R-Code ist so aufgebaut, dass er problemlos auf andere Ausgangslagen angewendet werden kann.

Dieses Projekt soll eine Grundlage liefern, welche zur Analyse von Methoden beliebig erweitert und vertieft werden kann und dadurch die aktuelle Forschung im Bereich der Bewegungsanalyse ein Stück weiterbringt.

1.3 Aufbau dieser Arbeit

Der Hauptteil dieser Arbeit ist in sechs Kapitel gegliedert. In Kapitel 2 wird zuerst der theoretische Hintergrund erläutert und anschliessend die für dieses Projekt relevante Literatur präsentiert. Im ersten Teil geht es um die Bewegungsökologie und um die Bewegungsanalyse. Der Begriff der Trajektorien und ihre Eigenschaften (*path characteristics*) werden erklärt. Im zweiten Teil werden dann bestehende Publikationen zur Segmentierung beschrieben. Daraus ergeben sich Forschungslücken und die Forschungsfragen werden gestellt. Das Kapitel 3 stellt die Simulationen vor, welche die Grundlage für dieses Projekt sind und es wird beschrieben, wie sie generiert werden. In Kapitel 4 wird der methodische Rahmen beschrieben (Ablauf im R-Code) und im Anschluss wird das Vorgehen dieses Projekts detailliert erläutert. Das Kapitel beinhaltet alle Methoden, welche auf die in Kapitel 3 beschriebenen Simulationen

EINLEITUNG

angewendet werden, die einzelnen Schritte des Vorgehens sowie die Evaluationskriterien. In Kapitel 5 werden dann die Resultate präsentiert. Kapitel 6 beinhaltet die Diskussion der Resultate. Sie soll helfen, die Forschungsfragen aus Kapitel 2 zu beantworten. Des Weiteren werden allfällige Limitierungen vorgestellt und das Projekt wird kritisch beurteilt. Am Schluss wird in Kapitel 7 ein Fazit gezogen und weitere Forschungsrichtungen werden vorgeschlagen.

2 HINTERGRUND

In diesem Kapitel wird zuerst theoretisches Hintergrundwissen vermittelt, welches wichtig ist für das Verständnis dieses Projekts. Im zweiten Teil dieses Kapitels wird dann der Stand der Forschung präsentiert. Auf dieser Grundlage werden Forschungslücken extrahiert, welche durch die aktuelle und zukünftige Forschung noch zu beantworten sind. Daraus leiten sich die Forschungsfragen dieses Projekts ab.

2.1 Theoretischer Hintergrund

Bewegung ist ein fundamentaler Teil jeglichen Lebens. Sie spielt nicht nur für jedes einzelne Individuum eine wichtige Rolle, sondern auch für Gemeinschaften, Populationen und ganze Ökosysteme, denn sie ermöglicht die Evolution und somit die Vielfalt an Leben (Nathan et al. 2008).

Nathan et al. (2008: 19052) beschreiben die Bewegung eines Organismus als «change in the spatial location of the whole individual in time». Um den Ortswechsel eines Organismus als Ganzes zu beschreiben, erläutert Trepl (2007) den allgemeinen Begriff der «Mobilität». Die Mobilität beinhaltet sowohl die Fähigkeit zum Ortswechsel (Vagilität) als auch die einzelnen Prozesse, welche diesen Ortswechsel auslösen. Ein Organismus kann sich aktiv von einem Ort an einen anderen bewegen (Lokomotion) oder er kann es passiv geschehen lassen (Transport). Wird die Mobilität nicht nur auf individueller Ebene betrachtet, sondern als Kollektiv, so spielt neben der Fortbewegung auch die Ausbreitung eine wichtige Rolle (Trepl 2007).

Durch Bewegung können Schlüsselprozesse und Muster in verschiedensten Disziplinen identifiziert werden: in der Ökologie genauso wie in der Epidemiologie, Geografie, Soziologie oder im Transportwesen (Holloway und Miller 2018). In den folgenden Kapiteln wird tiefer auf die Bewegungsökologie und die Bewegungsanalyse sowie deren Bestandteile eingegangen.

2.1.1 Movement Ecology – Bewegungsökologie

Unterschiedliche Disziplinen der Wissenschaft und Technologie haben Interesse an Bewegungsdaten. Dazu gehören zum Beispiel der Transport, die menschliche Mobilität, verschiedene Verhaltens- und Umweltstudien, die Informationswissenschaft oder die Bewegungsökologie (*Movement ecology*; Dodge et al. 2016). Die eher neuere Disziplin der Bewegungsökologie haben insbesondere Nathan et al. (2008) geprägt (Holden 2006).

Nathan et al. (2008) weisen darauf hin, dass ein vernetztes Verständnis der Ursachen, Mechanismen und Konsequenzen von Bewegungen eines Organismus zentral sind, um einerseits geschädigte Landschaften

HINTERGRUND

zu verwalten und wiederherzustellen, andererseits aber auch um invasive Arten, Allergene, Gifte oder ansteckende Krankheiten zu kontrollieren. Sie prägen den Begriff der «Movement Ecology» entscheidend und schlagen ein Paradigma vor, welches konzeptuelle, theoretische, methodische und empirische Faktoren verbindet, damit die Bewegungen aller Organismen untersucht werden können. Sie führen ein Konzept ein, welches die Bewegung eines individuellen Organismus als Zusammenspiel von vier verschiedenen Mechanismen beschreibt. Dazu gehören der *internal state* (Wieso bewegt er sich?), die *motion capacity* (Wie bewegt er sich?), die *navigation capacity* (Wann und wohin bewegt er sich?) und externe Faktoren (*external factors*). Der *internal state* beinhaltet sowohl die physiologische Ausstattung als auch den psychologischen Zustand eines Individuums, die es dazu antreiben gewisse Ziele anzustreben. Um beispielsweise Energie aufnehmen zu können, Schutz zu finden oder sich vermehren zu können, müssen die Organismen Nahrung suchen, fliehen oder Partner finden. Die *motion capacity* beinhaltet die Fähigkeit eines Individuums, sich selbst zu bewegen (*motile locomotion*) oder sich bewegen zu lassen (*externally vectored transport*). Dazu gehört, dass Vögel fliegen, Fische schwimmen oder Gazellen springen können. Einige Tierarten können sich auch auf mehrere Arten bewegen: Beispielsweise können gewisse Vogelarten sowohl fliegen als auch gehen oder schwimmen. Die *navigation capacity* verhilft Organismen, sich im Raum und/oder in der Zeit zu orientieren. Dazu wählen sie wohin (Richtung und Position) sie sich bewegen und/oder wann (Einleitung und Abschluss) ihre Bewegung stattfindet. Häufig ist dies direkt mit den Zielen verknüpft, die ein Organismus erreichen möchte. Das heisst, ein Ort mit optimaler Nahrung, ein Ort von genügender Sicherheit oder ein anderes Individuum von Interesse kann die Bewegung steuern. Manchmal beeinträchtigen aber auch externe Faktoren die Erreichung eines Ziels (Nathan et al. 2008).

Das Konzept von Nathan et al. (2008) kombiniert und verknüpft diese vier Mechanismen um eine fünfte Frage zu beantworten: Wie spielen diese vier Komponenten zusammen um eine bestimmte Bewegung zu erzeugen? Das Konzept soll Wissenschaftlern erlauben, die fundamentalen Mechanismen herauszufinden, die zu einem bestimmten Bewegungspfad führen. Dies beinhaltet jegliche Massstäbe; von einem einzelnen Schritt über verschiedene Bewegungsphasen bis hin zum ganzen Lebenspfad. (Nathan et al. 2008).

Bewegungen erfolgen also über mehrere räumliche und zeitliche Massstabsebenen und werden dadurch selbst ein Prozess in Raum und Zeit (Holloway & Miller 2018, Nathan et al. 2008). Die Bewegungsanalyse versucht diese Fragen, basierend auf raum-zeitlichen Informationen, zu beantworten.

HINTERGRUND

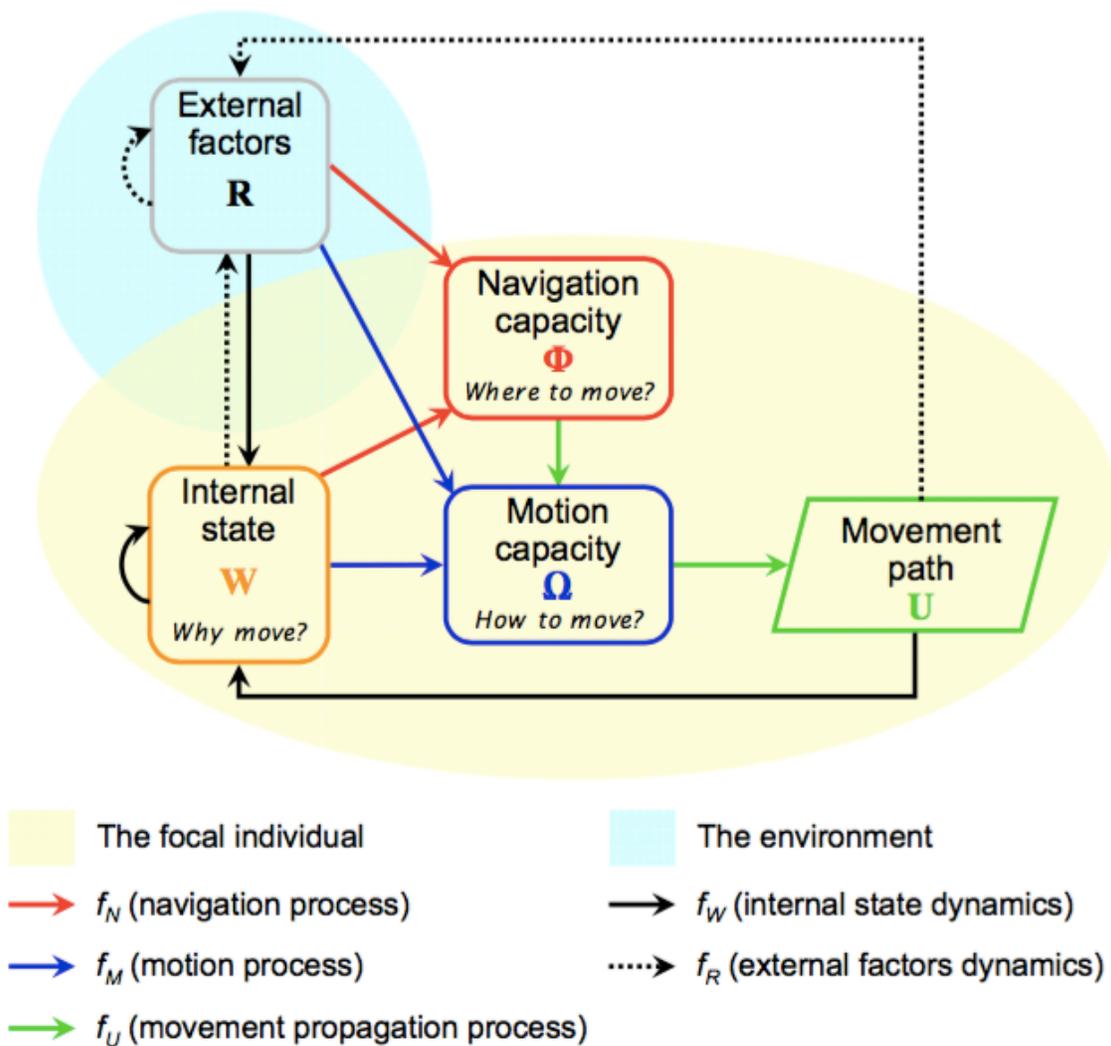


Abbildung 1: Konzept der Bewegungsökologie von Nathan et al. (2008). Es zeigt drei Grundkomponenten (gelber Hintergrund) einer Bewegung eines Individuums (internal state, motion capacity und navigation capacity) und eine externe Komponente (blauer Hintergrund), welche die anderen Komponenten beeinflusst (external factors). Die Beziehungen zwischen den Komponenten werden mit Pfeilen dargestellt, um die Richtung der Beeinflussung anzugeben. Der daraus resultierende Bewegungspfad gibt Informationen an den internal state sowie die externen Faktoren zurück.

2.1.2 Movement Analysis – Bewegungsanalyse

Die Bewegungsanalyse (*movement analysis*) hat zum Ziel, Bewegung zu verstehen, zu modellieren, vorherzusagen und zu visualisieren (Dodge et al. 2016). Sie ist ein Teilbereich der Bewegungsökologie und basiert auf der Information über den Ort und die Zeit eines Aufnahmepunkts (Demšar et al. 2015). Die Bewegungsanalyse ist sehr komplex und bringt unterschiedliche Schwierigkeiten mit sich. Nathan et al. (2008) legen vier Hauptschwierigkeiten der Bewegungsanalyse dar. Die erste Schwierigkeit ist es, einen durch die Bewegung entstandenen Pfad in eine Abfolge von homogenen Teilen zu zerlegen. Eine zweite Schwierigkeit ist die Klassierung der Segmente anhand ihrer Grundfunktionen beziehungsweise des Verhaltens eines Individuums. Drittens erschweren die unterschiedlichen Massstäbe, auf denen die Prozesse ein Verhalten steuern, die Analyse. Und die vierte Schwierigkeit ist die Verlinkung der Bewegungsdaten mit dem ökologischen Kontext. Diese von Nathan et al. (2008) dargelegten Schwierigkeiten beinhalten zwei Analysebereiche: die Klassierung und die Segmentierung (Soleymani 2016, Dissertation).

In der Klassierung wird ein Klassierungsmodell gebildet, welches auf Trajektorien mit bereits vorhandener Bezeichnung zum jeweiligen Verhalten basiert. Das Ziel besteht darin, Trajektorien ohne Verhaltenszuweisung den zugehörigen Klassen zuzuordnen. Dies geschieht aufgrund der Beziehungen zwischen den der Trajektorien zugrundeliegenden Eigenschaften und nicht aufgrund der Geometrie des Pfades (Soleymani 2016, Dissertation).

Die Segmentierung hingegen hat zum Ziel, einen Bewegungspfad in möglichst homogene Segmente zu zerlegen. Dies kann aufgrund der geometrischen Eigenschaften (*path characteristics*), dem semantischen Kontext oder den physiologischen Eigenschaften geschehen, die mittels anderen Methoden gemessen werden können (Soleymani 2016, Dissertation).

Die Bewegungsanalyse versucht also, anhand von aufgenommenen Daten (z. B. GPS-Daten), das Verhalten eines Individuums zu rekonstruieren. Die Verbindung dieser Messungen mit raum-zeitlichen geografischen Informationen ermöglicht es, die geografischen Prozesse zu verstehen, welche die Ökologen studieren wollen (Holloway & Miller 2018). Dies sind Fragen zum Wie und Warum der Bewegung von Tieren. Dazu gehören beispielsweise: «Wie und warum nutzen Tiere spezifische Ressourcen?», «Wie und warum interagieren sie mit Artgenossen?» oder «Wie und warum konkurrieren und reproduzieren sie sich?». Die Antworten zu diesen Fragen geben Informationen zur Leistung eines Tieres (Cagnacci et al. 2010). Diese Fragen können oftmals nicht mit reinen raum-zeitlichen Daten beantwortet werden, sondern müssen mit weiteren Informationen ergänzt werden. Diese stammen zum Beispiel von Beschleunigungsmessern oder aus Fernerkundungsdaten (Hebblewhite & Haydon 2010).

2.1.3 Trajektorien

Demšar et al. (2015: 2) definieren Trajektorien als «a discrete time series of measured locations». Laut Buchin et al. (2011) repräsentieren sie die Standorte eines sich bewegenden Objekts während eines gewissen Zeitintervalls. Trajektorien sind also diskrete Bewegungspfade bestehend aus X- und Y-Koordinaten und geordnet nach der jeweilig zugehörigen Zeit. Sie werden durch die Aufnahme von Standorten zu gewissen Zeitpunkten generiert (Buchin et al. 2011).

Aufgenommen werden diese Daten sowohl mit Hilfe von *global positioning systems* (GPS), *very high frequency* (VHF) als auch mit weiteren satellitengestützten Aufnahmesystemen (Demšar et al. 2015). Durch den technischen Fortschritt der letzten Jahre nimmt die Menge an Bewegungsdaten immer weiter zu. Einerseits ist es nun möglich, nebst grossen auch sehr kleine oder scheue Tiere mit Sensoren auszustatten, da sich die Grösse von Aufnahmesensoren verkleinert hat und in Zukunft noch weiter verkleinern wird (Demšar et al. 2015, Dray et al. 2010). Andererseits können Daten mit immer kürzeren Aufnahmeintervallen und über längere Zeiträume aufgezeichnet werden. Diese Entwicklung führt zu Bewegungsdaten mit sehr hoher räumlicher und zeitlicher Auflösung (Dodge et al. 2016). Holloway & Miller (2018) schreiben sogar von der Möglichkeit zur nahezu kontinuierlichen Aufnahme von Daten und den zugehörigen Informationen.

Die Aufnahme von Standorten kann in regelmässigen oder unregelmässigen Zeitintervallen gemacht werden. Oftmals werden regelmässige Intervalle bevorzugt. Unregelmässige Zeitintervalle werden teilweise dann verwendet, wenn die Aufnahme an die Geschwindigkeit eines Tieres gekoppelt sein soll (Buchin et al. 2011). Dabei werden höhere Geschwindigkeiten mit kleineren Zeitintervallen aufgenommen als tiefere Geschwindigkeiten. Die Zeitintervalle zur Aufnahme der Standorte können unterschiedlich lang sein. Kürzere Zeitintervalle führen zu einer höheren zeitlichen Auflösung und damit zu einer grösseren Menge an Datenpunkten. Sie erlauben eine detailreiche Sicht auf kurzzeitiges Verhalten, führen aber auch zu einem grösseren Anteil an Aufnahme Fehlern (z. B. räumliche Unsicherheiten) (Barraquand & Benhamou 2008; Edelhoff et al. 2016a). Im Gegensatz zur hohen Aufnahmerate kann bei einer tieferen Aufnahmerate nur Verhalten auf grösseren Massstäben interpretiert werden (Barraquand & Benhamou 2008; Edelhoff et al. 2016a).

GPS-Daten haben sich zu einer äusserst wichtigen Forschungsgrundlage entwickelt, dennoch bergen sie einige Schwierigkeiten. Nicht immer sind sie die beste Möglichkeit zur Datenerfassung, denn sie können durch Umwelteinflüsse wie klimatische Faktoren, die Bodenbeschaffenheit oder das Habitat sowie durch das tierische Verhalten (z.B. Bewegung, Orientierung des Aufnahme geräts) beeinflusst werden und zu Aufnahme Fehlern führen (Cagnacci et al. 2010). Bei Meerestieren ist dies beispielsweise oft der Fall. Standorte werden nur aufgenommen, wenn sich das Individuum genügend nahe an der Wasseroberfläche

HINTERGRUND

oder an Land befindet (Gurarie et al. 2009a). Zudem sind die Lebensdauer der Batterie sowie der Speicherplatz entscheidende Limitierungen (Hebblewhite & Haydon 2010; Tomkiewicz et al. 2010). Sobald zudem Informationen zur Umwelt gebraucht werden, müssen GPS-Daten mit weiteren Daten von externen Aufnahmegegeräten kombiniert werden (Cagnacci et al. 2010).

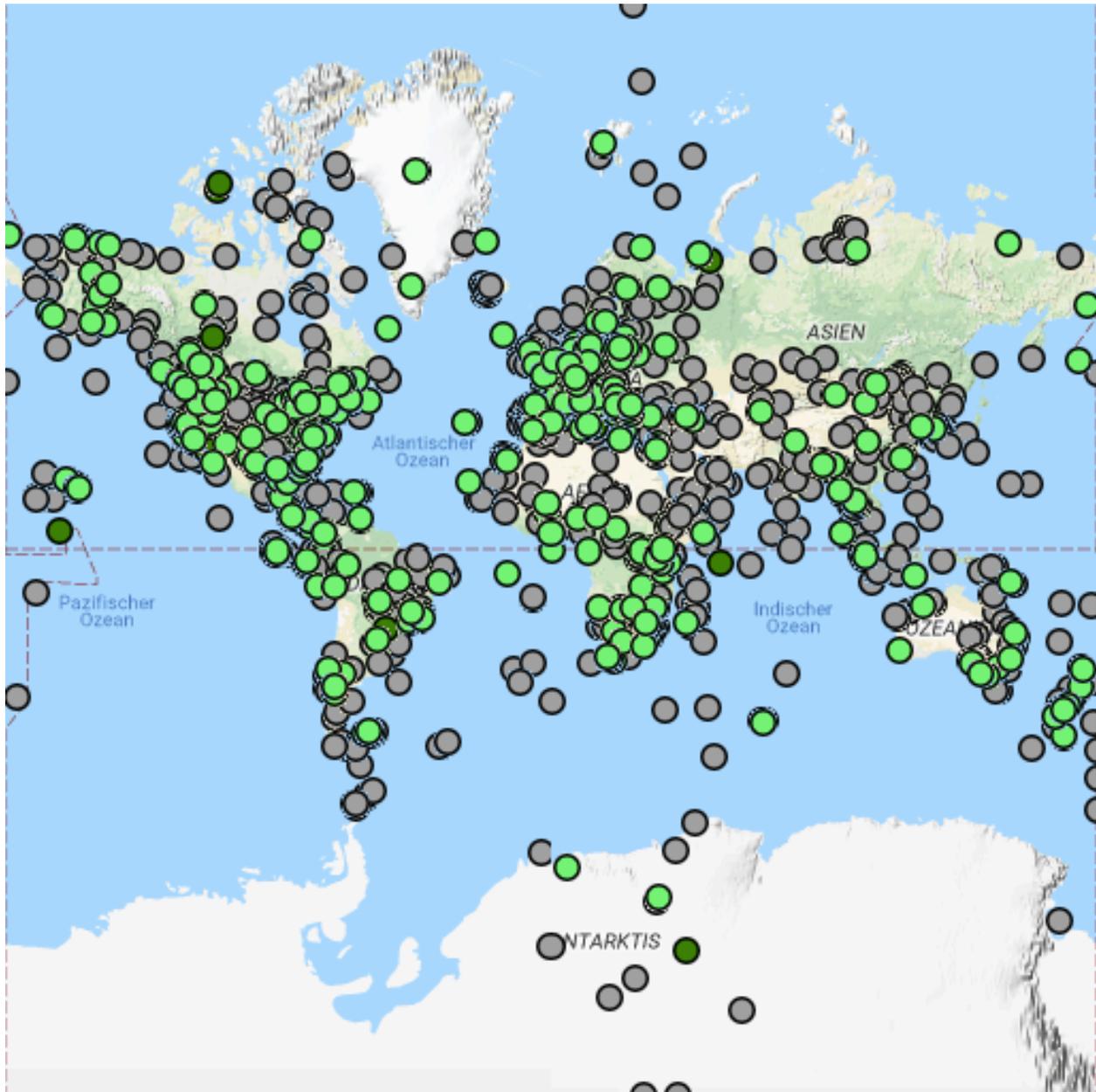


Abbildung 2: Datensätze vorhanden auf Movebank. Grün: frei verfügbare Datensätze, grau: Anfrage bei Inhaber der Daten notwendig. (Max Planck Institute of Ornithology 2018, Zugriff: 1.9.2018)

HINTERGRUND

Die Aufnahme von Bewegungsdaten hat in der Ökologie eine enorm hohe Wichtigkeit erlangt. Was früher mit sehr wenigen und schlechten Daten begann, ist nun eine äusserst datenreiche Disziplin geworden (Demšar et al. 2015). Diese Datensätze werden immer öfters in Datenbanken der Allgemeinheit zur Verfügung gestellt. Eine solche Datenbank ist zum Beispiel Movebank (www.movebank.org, Max Planck Institute for Ornithology 2018). Nebst mehr als 4000 Studien mit über 650 Mio. Standorten befinden sich auf dem *movebank data repository* über 100 Datenpakete und beinahe 300 Datendokumente, welche bereits in Publikationen verwendet wurden (Max Planck Institute for Ornithology 2018).

2.1.4 Path Characteristics

Aus Trajektorien lassen sich unterschiedliche Informationen, sogenannte *path characteristics*, ableiten. Aufgrund von X- und Y-Koordinaten in Kombination mit der zugehörigen Zeit, können mathematische Werte berechnet werden. Diese geben aufschlussreiche Informationen über die dazugehörige Bewegung. Genau an solchen Informationen ist die Bewegungsanalyse interessiert.

Eine Vereinheitlichung der unterschiedlichen Synonyme schufen Dodge et al. (2008) und teilten die *path characteristics* in drei Hauptgruppen ein: die *primitive parameters* und davon abgeleitet die *primary* und *secondary derivatives*. Zu den *primitive parameters* gehören die absolute räumliche Position (X- & Y-Koordinaten), die zugehörige Zeit sowie das Zeitintervall. Dies sind die rohen Daten, welche man beispielsweise mit einem GPS generiert. Daraus können weitere Parameter abgeleitet werden, welche die *primary* und *secondary derivatives* bilden. Als *primary derivative* zählt zum Beispiel die Distanz, welche aus zwei aufeinanderfolgenden Datenpunkten berechnet werden kann. Als *secondary derivative* gilt die räumliche Verteilung als Funktion der Distanz (Dodge et al. 2008).

Die *primary* und *secondary derivatives* können zwischen einzelnen Punkten berechnet werden, oder zwischen Sequenzen, die aus mehreren Punkten bestehen (Edelhoff et al. 2016a). Ein Beispiel für ersteres ist die Berechnung der Geschwindigkeit, für letzteres der *straightness index*. Der *straightness index* ist das Verhältnis zwischen der Euklidischen Distanz vom Startpunkt zum Endpunkt einer Trajektorie und die Summe der *step lengths* (Edelhoff et al. 2016a). Die *step length* beschreibt die euklidische Distanz zwischen zwei aufeinanderfolgenden Datenpunkten (Morelle et al. 2017a). Der *turning angle* beschreibt den Winkel zwischen drei Datenpunkten und ist somit eine Messgrösse für die Richtungsänderung (Morelle et al. 2017a). Die *orthogonal persistence velocity* beschreibt die Magnitude einer Bewegung und die *turning velocity* ist die Tendenz hin zu einer bestimmten Richtung (Morelle et al. 2017a, Gurarie et al. 2009b). Diese beiden zuletzt genannten Signale sind robust gegenüber den *sampling intervals*, da sie die Geschwindigkeit mit dem *turning angle* kombinieren (Edelhoff et al. 2016a). Die Tortuosität ist ein Indiz für die Orientierung einer Trajektorie (Edelhoff et al. 2016a) und beschreibt «the amount of turning in a given space or time» (Codling et al. 2008, 824). Die Tortuosität wird in der BCPA (Gurarie et al. 2009a)

HINTERGRUND

als Produkt der geschätzten Geschwindigkeit und dem Kosinus der *turning angles* berechnet. Eine Tabelle mit weiteren *path characteristics* präsentiert Edelhoff et al. (2016a).

Räumlich-zeitliche Autokorrelation ist in Bewegungsdaten allgegenwärtig. Je näher die Datenpunkte zeitlich liegen, desto höher wird die Abhängigkeit aufeinanderfolgender Punkte (Dray et al. 2010). Das heisst, je kleiner das *sampling interval*, desto grösser ist die räumlich-zeitliche Autokorrelation (Cagnacci et al. 2010). Autokorrelation ist eine Eigenschaft von zufälligen Variablen. Das heisst, dass Werte von nahegelegenen Stichproben sich bezüglich einer räumlich zufälligen Annahme entweder sehr (positive Autokorrelation) oder überhaupt nicht (negative Autokorrelation) ähnlich sind. Die Autokorrelation kann sowohl räumlich als auch zeitlich sein. Der Hauptunterschied liegt darin, dass bei einer zeitlichen Autokorrelation nur die vergangenen Werte beeinflusst werden können, während die räumliche Abhängigkeit in alle Richtungen beeinflussen kann (Dray et al. 2010).

Die Bewegung eines Individuums wird durch interne und externe Faktoren gesteuert, welche auf unterschiedlichen räumlichen und zeitlichen Massstäben stattfinden (Dodge et al. 2008). Daher sind auch die *path characteristics* abhängig vom räumlichen und zeitlichen Massstab (Soleymani et al. 2014). Der räumliche Massstab reicht von einer sehr lokalen Ebene im Zentimeter-Bereich bis hin zur globalen Ebene. Der zeitliche Massstab variiert zwischen sehr kurzem Verhalten und langzeitigem Verhalten (Dodge et al. 2008). Um diesen Unterschieden gerecht zu werden, muss der räumliche und zeitliche Massstab in einer Analyse bekannt sein, da unterschiedliches Verhalten auf unterschiedlichen Massstäben zu unterschiedlichen Werten der *path characteristics* führen kann (Nathan et al. 2008, Thiebault & Tremblay 2013). Die Beachtung der unterschiedlichen Massstäbe ist eine der vier Schwierigkeiten, welche Nathan et al. (2008) präsentieren (siehe Kapitel 2.1.2).

2.1.5 Simulationen

Simulationsdaten werden immer wieder verwendet, um tierische Daten nachzuahmen. Sie haben den grossen Vorteil, dass keine wilden Tiere gefangen werden müssen. Es müssen keine Aufnahmegeräte angebracht und abgenommen werden, wodurch Kontakte mit Menschen vermieden werden können. Dies minimiert zum einen ethische Konflikte, zum anderen kann die logistische Organisation stark vereinfacht werden (Technitis et al. 2015). Des Weiteren können die einzelnen Parameter individuell bestimmt werden. Simulationen werden einerseits direkt in Methoden eingegliedert und als statistisches Modell mit realen Daten verglichen (z.B. SSM) oder sie werden als Input für Methoden verwendet. Ersteres kann beispielsweise sinnvoll sein, wenn nur grosse Zeitintervalle in den Daten vorliegen, man aber wissen möchte, wo sich das Tier dazwischen hätte aufhalten können. Mit Hilfe eines auf statistischen Annahmen basierenden simulierten Bewegungsmodells kann dies herausgefunden werden (Technitis et al. 2015).

HINTERGRUND

Simulationen können aber auch als Ersatz für nicht vorhandene Datensätze verwendet werden. Damit dienen sie als Input für Methoden.

Es gibt zwei Gruppen von Simulationsmodellen: Ausbreitungsmodelle und sogenannte *correlated random walks* (CRW). Im ersten Fall werden Bewegungen als Verbreitung in einem bestimmten Raum betrachtet, während im zweiten Fall individuelle Pfade, basierend auf vorherigen Standorten und Parametern aus empirischen oder theoretischen Verteilungen, gebildet werden (Holloway & Miller 2018). Fauchald & Tveraa (2003) haben beispielsweise zwei unterschiedliche Simulationen generiert, um die Eigenschaften der Analysemethode FPT (siehe Kapitel 3.2) zu untersuchen.

Ähnliche Simulationen hat auch Pinaud (2008) generiert. Er hat zum Beispiel Simulationen verwendet, um mittels der FPT-Methode einen *area restricted search* (ARS) innerhalb der Simulationen zu identifizieren. Dazu simulierte er die Trajektorie eines Tieres, welches sich in einer Umgebung mit runden Flächen von gegebener Grösse bewegte. Ausserhalb einer solchen Fläche nahm die Bewegung des Tieres einen relativ direkten CRW mit konstanter Geschwindigkeit an. Die *turning angles* beschreiben eine Von Mises Verteilung. Innerhalb einer Fläche nahm das Tier vier unterschiedliche Suchverhalten an. Diese variierten in der Geschwindigkeit und der Sinuosität sowie dem Grenzverhalten (Pinaud 2008). Des Weiteren hat Pinaud (2008) Trajektorien simuliert, welche auf realen tierischen Parametern (*Albatross Diomedea exulans*) basieren. Diesen hat er ein Rauschen ähnlich dem eines Argos Aufnahmesystems hinzugefügt, um dessen Einfluss zu testen.

Mit ähnlichen Simulationen und hinzugefügtem Rauschen haben auch Barraquand & Benhamou (2008) Simulationen generiert. Ihren Simulationen liegen ebenfalls CRW zugrunde. Sie präsentierten eine Methode (Residence time in Kombination mit Lavielle (2005)), mit welcher möglichst profitable Orte innerhalb einer Trajektorie gefunden werden können.

Auch Gurarie und Ovaskainen (2011) verwendeten den CRW. Einerseits nutzten sie einen *discrete-step CRW* mit konstanten Intervallen und andererseits brauchten sie einen CRW mit zufälligen Intervallen. Zusätzlich verwendeten sie einen *continuous-time correlated velocity movement* (CVM). Ihr Ziel war es, die charakteristischen zeitlichen und räumlichen Massstäbe so zu definieren, dass sie das Bewegungsverhalten auf kleinen genau so wie auf grossen Massstäben beschreiben können.

Thiebault und Tremblay (2013) hingegen verwendeten eine gerade Linie, einen Kreis und vier RWs (*random walks*). Diese Simulationen beinhalten kein zusätzliches Rauschen. Die gerade Linie sowie der Kreis beschrieben nur eine Art von Bewegung. Bei beiden bleibt die Geschwindigkeit konstant, die *turning angles* unterscheiden sich jedoch.

HINTERGRUND

Le Corre et al. (2014) generierten Simulationen basierend auf dem Verhalten von Karibus. In den Simulationen wurden zwei saisonale Gebiete und zwei Zwischenstationen eingebaut. Weitläufige Bewegungen wurden während den Migrationsphasen simuliert und ARS während den saisonalen Gebieten und Zwischenstationen. Die ARS Gebiete haben eine bestimmte Grösse und die simulierten Tiere können, ausser während der Migrationszeit, die Flächen nicht verlassen (Le Corre et al. 2014).

Auch Edelhoff et al. (2016a) haben eine einfache, auf das Individuum bezogene Simulation verwendet. Dabei wird ein jährlicher Pfad eines Individuums beschrieben. Die Tageszeit, das Habitat sowie der Bezug zum Mittelpunkt des Lebensraums beeinflussen das Verhalten des Tieres. Die *turning angles* sind uniform verteilt zwischen $-\pi$ und π und die *step lengths* entsprechen einer Gamma-Verteilung. In der Nacht liegt die Wahrscheinlichkeit sich zu bewegen bei 0.3, tagsüber bei 1. Zu jedem Zeitpunkt kann das Individuum zwischen 20 Standorten wählen (Edelhoff et al. 2016c). Edelhoff et al. (2016a) haben diese Simulation verwendet, um unterschiedliche Methoden zu vergleichen, ähnlich wie Gurarie et al. (2016a).

Solche auf das Individuum basierte Simulationen sind für Technitis et al. (2015) von hoher Wichtigkeit. Sie betonen, dass Simulationen von individuellen Trajektorien weit wichtiger sind als statistische Verteilungen wie beispielsweise Brownian oder Lévy-Prozesse, weil letztere selber bereits Mängel mit in die Analyse bringen. Um ihrer Forderung nachzukommen, haben sie in ihrer Publikation einen Algorithmus («random trajectory generator» RTG) entwickelt, der die Vorteile dreier führenden Bewegungsmodelle kombiniert. Die drei Bewegungsmodelle sind das Random Walk Modell, das *space-time prisms* und das Brownian Bridge Bewegungsmodell. Der Algorithmus kann effizient zufällige Trajektorien generieren, welche mit dem kleinsten *directionalen* Bias zwischen zwei Punkten verlaufen. Trotz dieser Vorteile betonen Technitis et al. (2015) jedoch, dass ihr Algorithmus bezüglich des Realitätsgrades noch immer limitiert ist. Daher wollen sie in Zukunft auch mehr nutzerspezifische Parameter in den Algorithmus einbauen (z. B. Energie Budget, Anziehungs-/Abstossungspunkte) (Technitis et al. 2015).

In einem weiteren Anwendungsfeld können Simulationen auch genutzt werden, um Eigenschaften von Methoden zu entdecken und zu testen. Gurarie et al. (2016a) empfehlen Simulationen im Allgemeinen, um die Sensitivität sowie die Robustheit von Methoden zu beurteilen. In ihrer Publikation haben sie drei unterschiedliche Simulationen verwendet, um vier verschiedene Segmentierungsmethoden zu vergleichen. Die Simulationen variierten in der mittleren Geschwindigkeit, der Tortuosität sowie dem räumlichen Bias. Die in diesem Projekt verwendete Grundsimulation ist dieselbe wie Gurarie et al. (2016a) in ihrer Publikation verwendeten und als R Code zur Verfügung stellten.

2.1.6 Segmentierung

Innerhalb der Bewegungsanalyse gibt es zwei verschiedene Bereiche: die Klassifikation von Bewegungen anhand ihrer Muster sowie die Segmentierung von Trajektorien. Unter der Klassifikation von Bewegungen versteht man die Einteilung von Bewegungsmustern in Kategorien. Diese Kategorien sind vordefiniert und die Einteilung geschieht aufgrund spezifischer Eigenschaften der Bewegungsmuster.

Während man in der Klassifikation die Bewegungsmuster einteilt, werden in der Segmentierung die Trajektorien in homogene Bewegungsmuster aufgeteilt. Die Segmentierung ist Hauptbestandteil dieses Projekts und wird in den folgenden Kapiteln tiefer erklärt.

Edelhoff et al. (2016a) verwenden den Begriff «Segmentierung» um Änderungen im Bewegungsverhalten eines Tieres basierend auf dessen Trajektorie zu bestimmen. Dabei wird eine Trajektorie in mehrere Untertrajektorien aufgeteilt – in die sogenannten Segmente. Die Punkte, an denen Änderungen stattfinden und so die Segmente definieren, werden *change-* oder *breakingpoints* (CP) genannt (Edelhoff et al. 2016a). Die Segmente beschreiben jeweils unterschiedliche Bewegungsmuster, welche innerhalb eines Segments möglichst homogen sind. Die Bewegungsmuster lassen sich durch die *path characteristics* beschreiben. (Buchin et al. 2011).

Buchin et al. (2011) unterscheiden zwischen diskreter und stetiger Segmentierung. Während erstere nur an bestehenden Datenpunkten segmentieren kann, kann die stetige Segmentierung auch zwischen den Datenpunkten einen CP finden. Oft werden dadurch weniger Segmente gefunden (Buchin et al. 2011). Ob eine diskrete oder stetige Segmentierung verwendet wird, hängt unter anderem von der vorhandenen Datenmenge ab (Buchin et al. 2013). Ist eine Trajektorie bereits hoch aufgelöst, besteht nur ein geringer Unterschied zwischen den beiden Möglichkeiten. Sind die Intervalle zwischen den Datenpunkten jedoch gross, kann die Verwendung der stetigen Segmentierung von Vorteil sein. Es muss jedoch beachtet werden, dass der stetigen Segmentierung eine Interpolation vorausgeht (Buchin et al. 2013).

Ökologinnen und Ökologen verwenden Segmentierungsmethoden immer häufiger. Garstang et al. (2014) haben beispielsweise die BCPA verwendet, um einen möglichen Zusammenhang zwischen den Bewegungsarten von Afrikanischen Elefanten (*Loxodonta africana*) und den saisonalen Änderungen während den Regenzeiten zu erkennen. Le Corre et al. (2014) wollten unter der Verwendung der FPT den Zeitpunkt der Migration von migrierenden Karibus im nördlichen Québec und Labrador identifizieren. Solche Analysen von migrierenden Tierarten werden im Zusammenhang mit der menschlichen Ausbreitung sowie dem Klimawandel immer wichtiger, denn Habitatseinschränkungen oder -verluste sowie Veränderungen in der Phänologie haben einschneidende Einflüsse auf das Migrationsverhalten von Tieren (Le Corre et al. 2014).

HINTERGRUND

Manchmal interessiert es auch, wann ein Individuum zum ersten Mal einen bestimmten Punkt im Raum erreicht. Beispielsweise, wenn Tiere ihr Revier kontrollieren oder nur limitierte Ressourcen vorhanden sind. Dies haben McKenzie et al. (2009) mithilfe einer Form der FPT und einem Datensatz eines Rotfuchses (*Vulpes vulpes*) gezeigt.

Laut Zheng (2015) dient die Segmentierung oft auch als erster Schritt, wenn anschliessend ein Clustering oder ein Klassifikationsprozess stattfindet. Diesbezüglich empfehlen Barraquand und Benhamou (2008), eine Clusteranalyse unabhängig von der Methode durchzuführen. Denn wenn die Clusteranalyse direkt in die Methode integriert wird, können zusätzlich zu den Segmentierungsfehlern noch Klassierungsfehler dazukommen. Werden sie jedoch separat voneinander behandelt, kann jeweils die beste Methode ausgewählt und angewendet werden.

2.2 Stand der Forschung

Das zunehmende Interesse an Bewegungsdaten und die enorme Fülle an verfügbaren Datensätzen brachte und bringt noch immer viele Methoden zu deren Analyse hervor. Oftmals werden neue, den zu analysierenden Daten entsprechende Methoden entwickelt. Ein Vergleich unter Verwendung unterschiedlicher Datensätze ist hingegen rar.

2.2.1 Methoden der Bewegungsanalyse

Die Methoden unterscheiden sich in ihrem mathematischen und statistischen Aufbau, der Verwendung von Variablen und den der Methode zugrundeliegenden Fragen (Gurarie et al. 2016a). Die Methoden können deshalb sehr unterschiedlich gruppiert werden. Gurarie et al. (2016a) und Edelhoff et al. (2016a) sind bisher die einzigen, die eine Publikation mit einer Zusammenstellung unterschiedlicher Methoden veröffentlicht haben.

Gurarie et al. (2016a) haben die Gruppierung der Methoden aufgrund des analytischen Hintergrundes jeder Methode gebildet. Dabei haben sie die Gruppen aufgeteilt in «metric-based methods», «classification and segmentation», «phenomenological methods» und «mechanistic movement models». Die Gruppe der «metric-based methods» ist durch die geometrische Analyse von eindimensionalen Kurven inspiriert. Die zweite Gruppe bildet sich aus Klassierungs- und Segmentierungsmethoden. Die dritte Gruppe sind die «phenomenological methods», welche die Struktur oder Periodizität von Bewegungspfaden analysieren. Die letzte Gruppe besteht aus Methoden, welche die Veränderungen des Verhaltens durch ein statistisches Verfahren analysieren. In diesen Methoden wird ein Bewegungsmodell verwendet. Dazu gehören beispielsweise *Multistate Random Walks* (MRW), *hidden Markov models* (HMM) oder *state space models* (SSM). Diese Gruppe nennen Gurarie et al. (2016a) «mechanistic movement models»

HINTERGRUND

Tabelle 1: Zusammenfassende Tabelle von in vier Kategorien eingeteilte Methoden der Bewegungsanalyse aus Gurarie et al. (2016a). Fett geschrieben sind die vier Methoden, welche Gurarie et al. (2016a) in ihrer Publikation verwendeten. Die Methoden der letzten Gruppe sind *multistate random walks*, *hidden Markov models* oder *state space models*.

Category	Method	References
Metric-based	Fractal analysis	Fritz, Said & Weimerskirch (2003), Laidre <i>et al.</i> (2004) Nams & Bourgeois (2004); Tremblay, Roberts & Costa (2007)
	Tortuosity measures	Bovet & Benhamou (1988); Benhamou (2004)
	First passage time (FPT)	Fauchald & Tveraa (2003)
Classification and segmentation	Residence time (RT)	Barraquand & Benhamou (2008)
	Penalized contrasts	Lavielle (2005), Calenge (2006)
	Bayesian partitioning (BPMM)	Calenge (2006)
	k-clustering	van Moorster <i>et al.</i> (2010)
Phenomenological time-series analysis	RT (segmentation step)	Barraquand & Benhamou (2008)
	Autocorrelation functions	Boyce <i>et al.</i> (2010)
	Change point analysis (BCPA)	Gurarie, Andrews & Laidre (2009), Gurarie (2013) Kranstauber <i>et al.</i> (2012)
Mechanistic movement modelling	Wavelet	Polansky <i>et al.</i> (2010)
	Multistate random walk (MRW)	Morales <i>et al.</i> (2004)
	Ignoring location error	Forester <i>et al.</i> (2007), Langrock <i>et al.</i> (2012) Patterson <i>et al.</i> (2008), McClintock <i>et al.</i> (2012)
	Accounting for error	Jonsen <i>et al.</i> (2013), Breed <i>et al.</i> (2012)

Auch Edelhoff et al. (2016a) haben eine Gruppierung von Methoden vorgenommen (siehe Anhang A). Ihre Gruppierung bezieht sich auf drei Kategorien von Forschungsfragen. Die erste Gruppe hilft Fragen bezüglich der Beschreibung von Bewegungsmustern zu beantworten. Daher nennen sie diese «movement pattern description». Die zweite Gruppe hilft bei der Suche nach den CP, daher trägt sie die Bezeichnung «change point detection». Und die dritte Gruppe möchte grundlegende Prozesse der Bewegung identifizieren und wird daher «process identification» genannt (Edelhoff et al. 2016a). Nebst der Literatur, wie sie Gurarie et al. (2016a) ebenfalls angegeben haben, haben Edelhoff et al. (2016a) neben einer kurzen Beschreibung auch die analytische Kategorie sowie die Eigenschaften der Trajektorien (siehe Kapitel 2.1.3) als Input für die Methoden aufgelistet. Die analytischen Kategorien beschreiben, ob die Methode eine Trajektorie basierend auf der Topologie oder als Zeitreihe (*time series*) analysiert, oder ob sie ein Zustandsraummodell (*state-space model*) zur Analyse beinhalten. Die Eigenschaften der Trajektorien unterscheiden sie in primitiv, primär und sekundär. Diese Einteilung entspricht der Deklaration von Dodge et al. (2008). Können die Methoden nebst diesen Eigenschaften noch zusätzliche Informationen, wie beispielsweise Aktivitätsdaten, verwenden, geben Edelhoff et al. (2016a) dies ebenfalls an. Zusätzlich stellen sie zu ihrer Publikation eine zusammenfassende Zusammenstellung der Methoden zur Verfügung (Supplement S1). Ergänzend zu dieser Zusammenstellung werden im Folgenden einige Methoden präsentiert, die weder Gurarie et al. (2016a) noch Edelhoff et al. (2016a) genauer beschrieben haben, die jedoch in der Tabelle von Gurarie et al. (2016a) aufgelistet sind. Dies sind die *Fractal analysis*, die RT (*Residence Time*, inkl. Segmentierungsschritt) und die DWT (*discrete wavelet transform*) von Soleymani

HINTERGRUND

et al. (2017). Letztere wird als aktuelles Beispiel für *wavelet* Methoden genannt. Die beiden Methoden, welche in dieser Arbeit verwendet wurden (FPT und BCPA), werden in Kapitel 3.4 vorgestellt.

Fractal analysis

Fractal analysis ist eine Methode, um Verhaltensänderungen über verschiedene Massstäbe zu identifizieren. Die Methode misst die Tortuosität von Trajektorien (*Fractal dimension* (D)). Die Tortuosität ist ein guter Indikator für die Qualität eines Habitats bezüglich Nahrung, Schutz und weiteren Faktoren, welche die Bewegungen eines Individuums beeinflussen (Nams & Burgeois 2004). Sie variiert zwischen 1 (eine gerade Linie) und 2 (maximale Tortuosität, welche eine Fläche komplett ausfüllt) (Fritz et al. 2003, Nams & Burgeois 2004). Ein höheres D ist somit ein Hinweis auf eine erhöhte Nahrungssuche. D wird durch eine Methode berechnet, die in Nams und Burgeois (2004) als *FractalMean* definiert ist. Um variierende Werte der Länge der Trajektorie zu vermeiden, berechnet der *FractalMean* den Durchschnitt unter Verwendung von 50 unterschiedlichen Startpunkten. Diese werden zufällig gewählt und jeweils in beide Richtungen (vor und zurück) berechnet.

Von Laidre et al. (2004) wurde die *Fractal analysis* verwendet, um Unterschiede im saisonalen Verhalten von Narwalen (*Monodon monoceros*) zu analysieren. Fritz et al. (2003) haben mit Trajektorien von Wanderalbatrossen (*Diomedea exulans*) die Massstäbe identifiziert, auf welchen die Tortuosität zu unterschiedlichen Umwelt- und Verhaltensweisen führt. Nams und Burgeois (2004) wollten herausfinden, welchen Einfluss unterschiedliche Habitateigenschaften in der Winterzeit auf Fichtenmarder (*Martes americana*) haben und welche Rolle unterschiedliche Massstäbe dabei spielen.

Residence time (RT, segmentation step)

Diese Methode von Barraquand & Benhamou (2008) hat zum Ziel, die Dauer des Aufenthalts eines Individuums an unterschiedlichen Orten zu bestimmen. Damit lässt sich herausfinden, wie das Individuum seine Zeit auf unterschiedliche Orte aufteilt. Diese Methode hat im Gegensatz zu anderen Methoden (z. B. FPT) den Vorteil, dass sie eine Trajektorie automatisch segmentiert. Die Segmentierung basiert auf der Zeit, welche das Individuum in der Umgebung der Datenpunkte verbringt. Zusätzlich unterliegt ihr ein klarer statistischer Prozess. Die Methode kann als Erweiterung der FPT angesehen werden (Barraquand & Benhamou 2008). Hier wird, zusätzlich zur vom Kreismittelpunkt ausgehend vorwärts und rückwärts gemessenen Zeit, die Zeit dazugerechnet, welche das Individuum im Kreis verbringt. Die Zeit zwischen dem Verlassen des Kreises und dem erneuten Zurückkehren sowie die dabei zurückgelegte Distanz, ist limitiert. Die Wahl des Radius sollte basierend auf ökologischen Faktoren des Individuums geschehen. Da dies oftmals schwierig ist, empfehlen Barraquand & Benhamou (2008) den Radius anhand der Körpergröße zu bestimmen. Wird eine Analyse auf einer detaillierten, kleinen Ebene gemacht, soll die *step length* für ein langsames Tier in dichter Umgebung einigen Körperlängen entsprechen, diejenige für

HINTERGRUND

ein schnelles Tier in offenem Gelände einigen Dutzend Körperlängen. Der Radius soll dann ein paar *step lengths* betragen. Grundsätzlich werden lokalere Bewegungen mit kleinerem Radius identifiziert und grössere Bewegungen mit grösseren Radien. Vorteile der RT sind, dass sie robust ist gegenüber Daten mit einem Rauschen und dass sie sich auf unterschiedliche Situationen anwenden lässt (Barraquand & Benhamou 2008).

Discrete wavelet transform (DWT)

Diese Segmentierungsmethode von Soleymani et al. (2017) ist relativ neu. Das Ziel der Methode ist es ebenfalls, CPs von kontinuierlichen Veränderungen des Verhaltens zu identifizieren, welche auf unterschiedlichen Massstäben stattfinden. Dazu wird die Zeitreihe einer Trajektorie in *low-frequency approximation* und *high-frequency detail sub-bands* zerteilt. Erstere erkennt kontinuierliche Verhaltensänderungen über grössere Massstäbe (z. B. Migration mit Überwintern und Brutzeit) während letztere abrupte, kleinere CP (tägliche Bewegung mit Nahrungssuche und Ruhephase) erkennt. Veränderungen erkennt die Methode dadurch, dass sie Übereinstimmungen zwischen der Trajektorie und einer *Wavelet*-Funktion erkennt und diese als *Wavelet*-Koeffizient quantifiziert. Mit der Skalierung der *Wavelet*-Funktion können unterschiedliche Massstäbe betrachtet werden. Mittels einer *Peak analysis* werden die einzelnen Teilbänder dann in Segmente aufgeteilt. Die Höhe der *Peaks* wird durch einen Grenzwert definiert. Alles was unter dem Grenzwert liegt, zählt als gleiches Segment, ansonsten wird ein neues Segment gebildet. Ein Vorteil dieser Methode ist, dass sie auf unterschiedliche *path characteristics* angewendet werden kann. Ein weiterer Vorteil sind die unterschiedlichen Massstäbe, auf welchen das Verhalten analysiert werden kann. Eine Schwierigkeit ist das Grenzwerte für die Höhe des *Peaks* müssen visuell bestimmt werden. Dies bedingt ein gewisses Vorwissen bezüglich des möglichen Verhaltens der Individuen (Soleymani et al. 2017).

Bisher wurde diese Methode erst auf Simulationen von Gurarie et al. (2016a) und Truthahngeier angewendet (Soleymani et al. 2017). Die Autoren haben dabei einige Vorteile gegenüber anderen Methoden festgestellt. Die DWT bietet die Möglichkeit, sehr feine Verhaltensänderungen in der Trajektorie zu erkennen, unterschiedliche *path characteristics* als Input zu verwenden und auch unregelmässige Daten sowie Daten mit einem Rauschen zu akzeptieren. Zusätzlich ist sie rechnerisch effizient und biologisch intuitiv. Als Nachteile sehen Soleymani et al., dass einerseits Randeffekte auftreten können und andererseits eine Trajektorie mit gewisser Länge notwendig ist. Weitere Informationen zur Methode liefern Soleymani et al. (2017).

Kombinationen von Methoden

Des Weiteren gibt es Analysen, welche eine Segmentierungsmethode mit weiteren Methoden kombinieren. Die Segmentierung bildet dabei den ersten Schritt, während unterschiedliche Clustering- oder Klassierungsmethoden die nächsten Schritte bilden.

Beispielsweise haben Zhang et al. (2015) eine BCPA in Kombination mit einer Clusteranalyse verwendet. Dabei wird im ersten Schritt die Trajektorie durch die BCPA segmentiert, anschliessend wird eine hierarchische multivariate Clusteranalyse durchgeführt, um die Anzahl von Verhaltensmustern zu identifizieren. Zum Schluss wird noch ein k-means Clustering angewendet um die Segmente den jeweiligen Verhaltensmustern zuzuordnen.

Einen ähnlichen Ansatz verwendeten Morelle et al. (2017a). Sie hängten drei Schritte aneinander, wobei die Segmentierung mittels BCPA der erste Schritt ist. Anschliessend definieren sie *candidate movement models* und beschreiben diese als nicht-lineare und lineare Formeln zwischen *net squared displacement* (NSD) und der Zeit. Dann wird jede *candidate* Formel auf die NSD Segmente gepasst und mit diversen statistischen Kriterien werden die am besten passenden Muster ausgesucht.

Auch Teimouri et al. (2018) haben eine mehrteilige Vorgehensweise gewählt. Diese basiert auf zwei Schritten. Der erste ist ebenfalls die Anwendung der BCPA. Im zweiten Schritt wird eine Kolmogorov-Smirnov-Distanzmetrik verwendet, um ein agglomeratives hierarchisches Clustermodell der Segmente herzustellen. Laut Teimouri et al. (2018) ist ihr Vorteil gegenüber von Zhang et al. (2015), dass sie die Anzahl Cluster nicht vorgängig definieren mussten und dass sie die Verteilung der Parameterwerte als Ganzes und nicht nur einzelne statistische Werte wie beispielsweise der Median beachten können.

2.3 Bisherige Vergleichsstudien

Ein zunehmend wichtiger Bestandteil in der Bewegungsanalyse ist der Vergleich von unterschiedlichen Methoden. Bisher gibt es jedoch erst zwei Publikation, welche mehrere Methoden vergleichen. Diese werden in den folgenden Abschnitten kurz zusammengefasst.

Gurarie et al. (2016a): What is the animal doing? Tools for exploring behavioural structure in animal movements.

Gurarie et al. (2016a) waren die ersten, die einen Methodenvergleich anhand von Simulationen und Datensätzen publiziert haben. Dieser soll Forschern helfen, eine für ihre Fragestellung geeignete Methode auszuwählen. In dieser Publikation verglichen sie vier Methoden. Die vier Methoden wählten sie basierend auf einer von ihnen vorgenommenen Gruppierung unterschiedlichster Methoden (siehe Kapitel 2.2.3). Der Vergleich der Methoden ist rein visuell.

HINTERGRUND

Die vier gewählten Methoden – FPT, BPMM, BCPA und MRW – wurden mit einem eigens dafür produzierten Package namens «WADDLE» (Gurarie et al. 2016c) analysiert. WADDLE basiert auf den Packages AdehabitatLT (Calenge 2009b), BCPA (Gurarie et al. 2014) und MRW (Gurarie et al. 2016c). Es beinhaltet sowohl Methoden zur Bildung von Simulationen, zur Segmentierung, Analyse und Visualisierung sowie auch reale Datensätze.

In der Verwendung von Simulationen, in welchen die mittlere Geschwindigkeit, die Tortuosität und der räumliche Bias variieren, haben die Autoren Schwierigkeiten festgestellt: Methoden, welche die Autokorrelation in den Bewegungsparametern nicht berücksichtigen, führen zu falschen CP. Methoden, welche den räumlichen Bias nicht miteinbeziehen, führen zu fehlenden CP aufgrund missachteter Orientierungswechsel. Dieses Problem nennen die Autoren «model misspecification». Es beinhaltet die Unfähigkeit einer Methode, einen CP zu erkennen der auf einer *path characteristic* basiert, welche die Methode nicht in die Analyse miteinbezieht. Sie betonen die Wichtigkeit zur Beachtung dieses Problems bei der Auswahl sowie der Anwendung einer Methode. Des Weiteren haben sie beim Methodenvergleich herausgefunden, dass zudem weitere Konflikte bestehen und teilweise Kompromisse eingegangen werden müssen. Es muss eine Menge an mutmasslichen Annahmen getroffen werden und der Output weist eine hohe Komplexität auf, aber es besteht das Potenzial, möglichst viel Information zu generieren.

Aus ihrer Analyse haben Gurarie et al. (2016a) Grundsätze für die explorative Analyse gebildet. Sie empfehlen, bei der Wahl von Input-Parametern der eigenen Intuition zu vertrauen, die Struktur der Daten zu analysieren, die Machbarkeit zu überprüfen, bei mehreren Tieren zuerst nur wenige zu analysieren, Annahmen zu treffen und zu testen, Simulationen zu verwenden wenn das Potenzial für Fehler hoch ist sowie verschiedene Methoden zu kombinieren, um möglichst viele nützliche Informationen zu bekommen. Im Allgemeinen empfehlen sie, möglichst genaue explorative Analysen durchzuführen bevor komplizierte Modelle angewendet werden (Gurarie et al. 2016a).

Edelhoff et al. (2016a): Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns

Die Publikation von Edelhoff et al. (2016a) ist auf drei Forschungsfragen aufgebaut, welche durch die Segmentierung zu beantworten versucht werden. Erstens die quantitative Beschreibung von Verhaltensmustern, zweitens die Identifizierung von signifikanten CP und drittens die Identifizierung von zugrundeliegenden Prozessen. Anschliessend diskutieren die Autoren Vorteile und Einschränkungen verschiedener Methoden und präsentieren Vorgehensweisen, um eine geeignete Methode auszuwählen.

Im ersten Schritt soll eine generelle Analyse der Daten stattfinden. Im zweiten Schritt werden die Datenstruktur und ihre Eigenschaften analysiert. Dazu gehört, ob das Zeitintervall zwischen den

HINTERGRUND

Aufnahmepunkten regelmässig ist und ob eine zeitliche Autokorrelation vorhanden ist. Im dritten Schritt wird, basierend auf dem Ziel der Analyse, eine Methode ausgewählt.

Im Anschluss testen Edelhoff et al. (2016a) drei Methoden an einer Individuum-basierten Simulation. Dies sind die *thresholding*-Methode, die BCPA und ein *hidden Markov model* (HMM). Die erste Methode ist ein auf der Topologie basierender Ansatz, die BCPA basiert auf Zeitreihen-Analysen und das HMM gehört zu den Zustandsraummodellen (*state-space model*). Diese drei analytischen Kategorien beantworten die drei oben genannten Forschungsfragen.

Die auf der Topologie basierten Methoden sind nützlich, wenn spezifische Hypothesen bezüglich den Verhaltensmustern schon im Vorherein gestellt werden können (Van Moorter et al. 2010). Zusätzlich sind dies die Methoden, welche die geringsten Forderungen bezüglich Datenaufbau haben und keine spezifischen Annahmen über Dateneigenschaften oder *path characteristics* machen. Sie werden oftmals als erstes Analyseinstrument verwendet. Die Zeitreihen-Analysen liefern tiefere Einblicke in Verhaltensänderungen und Korrelationen in den Daten (Gurarie et al. 2009), sind aber auch anspruchsvoller bezüglich dem Datenaufbau. Die Zustandsraummodelle sind am anspruchsvollsten in der Anwendung, bieten aber auch den besten Einblick in die Verhaltensdaten. Sie beinhalten zusätzlich am meisten Annahmen bezüglich Bewegungseigenschaften und erlauben dadurch eine gewisse Unsicherheit (Edelhoff et al. 2016a).

Edelhoff et al. (2016a) betonen die aktuellen Schwierigkeiten für eine Wissenschaftlerin oder einen Wissenschaftler, einen Überblick über die unterschiedlichen Methoden zu bekommen und eine für die jeweiligen Daten und Forschungsfragen geeignete Methode auszuwählen. Es ist daher von enormer Wichtigkeit, verschiedene Datensätze mit unterschiedlichen Charakteristiken zu analysieren und vergleichen. Zusätzlich müssen unterschiedliche Methoden untereinander verglichen werden und Hilfestellungen zur Wahl einer geeigneten Methode bereitgestellt werden.

2.4 Forschungslücke

Wie aus dem letzten Kapitel erkennbar ist, sind die beiden vergleichenden Publikationen von Gurarie et al. (2016a) und Edelhoff et al. (2016a) erst ein Startpunkt für Vergleiche von Methoden der Bewegungsanalyse. Bisher mangelt es an weiteren Vergleichen, aber es wird in verschiedensten Publikationen gefordert, Methoden aufgrund unterschiedlicher Forschungsfragen zu analysieren, damit einerseits ein breites Wissen aufgebaut werden kann und andererseits möglichst viele Erfahrungen mit unterschiedlichen Datensätzen gemacht werden können. Zu wissen, wie die Segmentierungsmethoden auf unterschiedliche *path characteristics* reagieren oder wie sie mit Unregelmässigkeiten (z. B. einem Rauschen) in Daten umgehen, ist für eine Ökologin oder einen Ökologen von enormer Wichtigkeit.

Des Weiteren wurden solche Vergleiche von Methoden noch nie statistisch ausgewertet. Obwohl die Verwendung von Simulationen sich als Grundlage für eine solche Analyse anbietet, wurden die Resultate basierend auf Grafiken bisher nur visuell verglichen und analysiert. Mit der ständig ansteigenden Anzahl an Datensätzen, der Zunahme an hoch aufgelösten Daten sowie dem zunehmenden Interesse an solchen Daten, können statistische Resultate für die Analyse von grosser Wichtigkeit sein. Beispielsweise bei Datensätzen bestehend aus mehreren Individuen können statistische Resultate von Vorteil sein, da ein visuelles Begutachten sehr viel Zeit in Anspruch nehmen kann.

2.5 Forschungsfragen

Basierend auf dem Stand der Forschung und den Forschungslücken lassen sich die Forschungsfragen bilden. Diese sollen dazu führen, dass anhand der Resultate dieses Projekts die Ziele erreicht werden können und einen aktuellen Beitrag zur Forschung liefern.

Forschungsfrage I Inwiefern sind die First Passage Time und die Behavioural Change Point Analysis für eine Ökologin oder einen Ökologen von Nutzen? Welche Arten von Forschungsfragen lassen sich durch die beiden Segmentierungsmethoden beantworten?

Hypothese I Segmentierungsmethoden wie die First Passage Time oder die Behavioural Change Point Analysis werden oftmals angewandt, ohne zu wissen welche Leistung sie erbringen können. Jede Methode hat ihre Eigenschaften, die sie in die Analyse miteinbringt und die für gewisse Forschungsfragen besser geeignet sind als andere. Simulationen und ihre statistische Auswertung werden helfen, den Nutzen der beiden Methoden herauszufinden.

Forschungsfrage II Wie robust gestalten sich die Methoden gegenüber einem Rauschen in der Simulation? Wie genau können die Methoden einen Change Point identifizieren? Welchen Einfluss haben die Werte der Parameter der jeweiligen Methode sowie die zu identifizierenden *path characteristics*?

Hypothese II Werden unterschiedliche Individuen innerhalb einer Gruppierung (beispielsweise einer Herde) analysiert, zeigen sie ihr eigenes Verhalten. Dieses unterscheidet sich von den anderen Individuen in unterschiedlicher Masse. Um solche Unterschiede in ähnlichen Trajektorien zu berücksichtigen, beinhalten die Simulationen ein zufälliges Rauschen. Dieses Rauschen hilft, die Methoden auf ihre Genauigkeit in der Identifizierung von Change Points zu untersuchen. Doch auch die der Methode zugrundeliegenden Eigenschaften können diese Genauigkeit der Methode beeinflussen.

Forschungsfrage III Welche Stärken und Schwächen lassen sich aus den Resultaten der Segmentierung für die jeweilige Methode ableiten?

HINTERGRUND

Hypothese III Anhand von Informationen über die Stärken und Schwächen der Methoden lässt sich für eine Ökologin oder einen Ökologen gezielter eine Auswahl treffen, um eine für den Datensatz geeignete Methode zu verwenden. Unterschiedliche Werte für die Parameter der Methoden sowie der Simulationen können helfen, die Vorzüge oder Nachteile einer Methode zu identifizieren.

3 MATERIAL

Dieses Kapitel beschreibt die in diesem Projekt verwendeten Simulationen. Sie basieren auf der Simulation aus Gurarie et al. (2016a). Im ersten Teil wird die Simulation in ihrer zugrundeliegenden Struktur erklärt. Anschliessend werden die verwendeten Werte für die Parameter der Simulationen präsentiert. Ebenso werden die beiden Methoden vorgestellt, mit denen die aus den Simulationen generierten Daten untersucht und verglichen werden. Es sind die beiden Methoden FPT (First Passage Time) und BCPA (Behavioural Change Point Analysis), die ebenfalls von Gurarie et al. (2016a) verwendet wurden.

3.1 Verwendete Simulationen

Der Simulationsalgorithmus für die Geschwindigkeits- und *time-scale switching* Modelle basieren auf einem CVM (*continuous time correlated velocity movement*) (Johnson et al. 2008, Gurarie & Ovaskainen 2011, Gurarie & Ovaskainen 2013). In diesem Modell wird die Geschwindigkeit $v(t)$ als ein 2D Ornstein-Uhlenbeck Prozess ausgedrückt. Ein Ornstein-Uhlenbeck-Prozess wird auch Gaussian-Markov-Prozess genannt. Dieser beschreibt ein stochastischer Prozess, welcher stationär und kontinuierlich in der Wahrscheinlichkeit ist sowie dem Gaussian- und Markov-Prozess entspricht (Finch 2004).

$$dv = \frac{1}{\tau} v dt + \frac{2v}{\sqrt{\pi\tau}} dW_t$$

Die mittlere tangentielle Geschwindigkeit des Prozesses v sowie die *time-scale* der Autokorrelation τ beschreiben den Zeitpunkt, an welchem die Autokorrelation der Geschwindigkeit um den Faktor $1/e$ fällt. Das Modell beinhaltet ein zweidimensionales nicht-korreliertes weisses Rauschen dW_t (Wiener-Prozess) mit der Varianz dt (Gurarie et al. 2016b). Der momentane Wert eines weissen Rauschens ist statistisch unabhängig von allen anderen Werten, denn es gibt keine inneren Zusammenhänge in einem weissen Rauschen (Isermann, R. 1974).

3.1.1 Definition der Parameter

Die in Kapitel 3.1.1 beschriebene Simulation hat drei Parameter, welche die Nutzerin oder der Nutzer bestimmen können. Dies sind die *characteristic time-scale*, die mittlere Geschwindigkeit sowie die Länge der jeweiligen Phase (Anzahl Datenpunkte pro Segment). Zusätzlich zu diesen drei Parametern, ist in diesem Projekt auch die *sampling frequency* zu bestimmen. Diese muss in Sekunden angegeben werden (z. B. alle 2 Minuten einen Punkt ergibt eine *sampling frequency* von 7200). Die Koordinaten sind keine geographischen Koordinaten, sondern kartesische Koordinaten.

Speed switch & time-scale switch (Gurarie et al. 2016a) und continuous speed switch (Soleymani et al. 2017)

In der Analyse von Gurarie et al. (2016a) wechselt in jeder einzelnen Simulationen jeweils nur ein Parameter. In der ersten Simulation ist es die mittlere Geschwindigkeit, in der zweiten Simulation ist es die Tortuosität. In beiden Simulationen ist das Verhalten jeweils bekannt: die erste und vierte Phase bestehen aus je 100 Punkten und repräsentieren eine intensive Bewegung (tiefe Geschwindigkeit und hohe Tortuosität), die mittleren beiden Phasen bestehen aus jeweils 50 Punkten und repräsentieren eine Bewegung mit höherer Geschwindigkeit und tieferer Tortuosität. Dabei ist die Geschwindigkeit in der dritten Phase am höchsten, die Tortuosität hingegen am tiefsten. Die *sampling frequency* liegt bei beiden Simulationen bei 1 (1s).

In der dritten Simulation (continuous speed switch model) nimmt die Geschwindigkeit mit jeder Phase stufenweise um 1 zu. Die Geschwindigkeit der ersten Phase beträgt 1, die der letzten 5. Die Phasenlänge beträgt jeweils 1000 Punkte und die *sampling frequency* ist 2 (2s).

In diesem Projekt werden die Parameter in einem ersten Schritt gleich gewählt wie in Gurarie et al. (2016a) und Soleymani et al. (2017). Dadurch lassen sich die Resultate mit denjenigen von Gurarie et al. (2016a) und Soleymani et al. (2017) vergleichen.

Tabelle 2: Parameter der speed switch, time-scale switch (Gurarie et al. 2016a) und der continuous speed switch (Soleymani et al. 2017) Simulation.

	Speed switch	Time-scale switch	Continous speed switch
Geschwindigkeit	1, 5, 10, 1	1	1, 2, 3, 4, 5
Tortuosität	2	2, 20, 40, 2	2
Phasenlänge	100, 50, 50, 100	100, 50, 50, 100	1000
Sampling freq.	1	1	2

Speed switch model mit Veränderungen

Um dann herauszufinden, wie sensitiv die Erkennung der FPT im mittleren CP ist, wird die Speed switch mit weiteren Varianten untersucht. Dabei werden sowohl die Geschwindigkeit wie auch die Phasenlänge separat variiert und dann kombiniert. In der speed switch 1 wird die Geschwindigkeit von 5 auf 10 zurück auf 5 und wieder auf 10 geändert. Die Phasenlänge beträgt 50, wie in der ursprünglichen Simulation. In der speed switch 2 gibt es nur den Wechsel von 5 auf 10 in der Geschwindigkeit, dafür wird die Phase auf 100 verlängert. In der speed switch 3 werden beide kombiniert. Das heisst, die Geschwindigkeit steigt zweimal von 5 auf 10 und die Phasenlänge beträgt 100. Um weitere Faktoren ausschliessen zu können, wird die Tortuosität und die *Sampling frequency* konstant belassen.

MATERIAL

Tabelle 3: Parameter für die drei veränderten speed switch Simulationen.

	Speed switch 1	Speed switch 2	Speed switch 3
Geschwindigkeit	5, 10, 5, 10	5, 10	5, 10, 5, 10
Tortuosität	2	2	2
Phasenlänge	50	100	100
Sampling freq.	1	1	1

Speed und time-scale switch

In einem weiteren Schritt werden die speed switch und die time-scale switch Simulationen kombiniert. Das heisst, die Geschwindigkeitsänderungen werden aus der speed switch übernommen, die Tortuosität wird aus der time-scale switch übernommen. Diesmal bleiben die Phasenlänge sowie die *Sampling frequency* unverändert.

Tabelle 4: Parameter für die speed und time-scale switch Simulation. Diese kombiniert die jeweiligen Veränderungen der Parameter aus der speed switch und der time-scale switch Simulation von Gurarie et al. (2016a).

	Speed und time-scale switch
Geschwindigkeit	1, 5, 10, 1
Tortuosität	2, 20, 40, 2
Phasenlänge	100, 50, 50, 100
Sampling frequency	1

3.2 Verwendete Methoden

Zwei der von Gurarie et al. (2016a) verwendeten Methoden werden in dieser Arbeit angewendet. Dies sind die First Passage Time (FPT, siehe Kapitel 3.2.1) und die Behavioural Change Point Analysis (BCPA siehe Kapitel 3.2.2). Diese beiden Methoden wurden einerseits gewählt, da sie bereits in Gurarie et al. (2016a) und Soleymani et al. (2017) verwendet wurden und dadurch verglichen werden können. Andererseits sind beide Methoden in Edelhoff et al. (2016a) als Zeitreihen-Analysen deklariert (für die FPT ist es die Penalized Contrast Method (PCM) von Lavielle (2005)). Die BPMM (*Bayesian partitioning of Markov models*) wurde nicht gewählt, da es sich um eine Klassierungsmethode handelt. Die MRW (*Multistate random walk*) wurde aufgrund der hohen Komplexität der *mechanistic movement models* nicht gewählt – sie müsste in einem eigenen Projekt weiter untersucht werden. Ebenfalls wurden keine weiteren Methoden verglichen, da dieses Projekt auch einen Rahmen schaffen will, mit welchem weitere Methoden auf relativ einfache Weise analysiert werden können.

3.2.1 First Passage Time (FPT)

Die *First Passage Time* (FPT) wurde erstmals analysiert durch Johnson et al. (1992) und Fauchald und Tveraa (2003) richtig analysiert. Johnson et al. (1992) gaben die FPT als Alternative für das *mean squared displacement* an, um die Ausbreitung von Tieren zu analysieren. Die FPT kann als Dauer verstanden werden, wie lange ein Individuum braucht, um einen Kreis von bestimmter Grösse einmal zu durchqueren. Es misst also die Zeit, die ein Individuum innerhalb einer bestimmten Fläche verbringt. Diese ist abhängig vom Radius sowie der Geschwindigkeit und der Tortuosität der Bewegung des Individuums (Fauchald & Tveraa 2003). Das heisst, sie unterscheidet hauptsächlich zwischen Streckenbewältigung und Nahrungssuche. Für grosse Massstäbe heisst das, dass hohe FPT-Werte in Aufenthaltsgebieten vorkommen während tiefe FPT-Werte Migration beschreiben (Le Corre et al. 2014). Das Ausbreitungsverhalten kann dann durch die Skalierung der *mean FPT* herausgefunden werden, welche durch den Radius des Kreises beeinflusst wird. Die *mean FPT* ist laut Johnson et al. (1992) jedoch eine Funktion der Landschaft und nicht des tierischen Verhaltens. Fauchald und Tveraa (2003) untersuchten die Eigenschaften der FPT als Messmethode für *search effort* entlang einer Trajektorie. Der Ort beziehungsweise der Massstab, an dem ein Individuum seinen *search effort* verstärkt, kann durch die Varianz der FPT entgegen unterschiedlichen räumlichen Massstäben, herausgefunden werden. Der zum *search effort* zugehörige Massstab kann verfälscht werden durch lange *sampling intervals*. Wenn die *sampling intervals* gross sind, dann werden tortuose Bewegungen verpasst und damit das Maximum der Varianz tiefer erreicht als der tatsächliche Massstab eigentlich wäre (Fauchald & Tveraa 2003).

Fauchald und Tveraa (2003) berechnen die FPT entlang eines Pfades. Um jeden Punkt der Trajektorie liegt ein gewisser Radius. Innerhalb dieses Radius wird die Zeit zwischen dem erstmaligen Durchlaufen des Radius vorwärts und rückwärts gemessen. Da die FPT aufgrund eines Radius berechnet wird, müssen die ersten und letzten Punkte ausgeschlossen werden. Das heisst, alle die Punkte, an denen die FPT vorwärts beziehungsweise rückwärts unbekannt sind (Fauchald & Tveraa 2003). Barraquand und Benhamou (2008) bemängeln, dass die FPT, welche von Fauchald und Tveraa (2003) verwendet wurden, nur eine rein visuelle Analyse zulassen. Um die Trajektorie mit Einbezug der FPT zu segmentieren, wird oft Lavielle (2005) verwendet (Barraquand & Benhamou 2008).

Um die FPT an Umwelteinflüssen oder an Suchverhalten anderer Individuen zu koppeln, kann man die FPT über ein geografisches Netz mitteln (Fauchald & Tveraa 2003).

Die FPT ist bisher hauptsächlich angewendet worden, um tierisches Verhalten bei der Nahrungssuche zu analysieren. Oftmals wurden marine Säugetiere und Meeresvögel (Pinaud & Weimerskirch 2007, Bailey & Thompson 2006) und einzelne terrestrische Säugetiere (Frair et al. 2010, Le Corre et al. 2008) analysiert.

MATERIAL

Pinaud (2008) hat in seiner Publikation die FPT verwendet, um ARS zu identifizieren. Die FPT hat CP erkennen können, wenn die Geschwindigkeit und die *sinuosity* sich veränderten. Zusätzlich konnte sie diese auch auf unterschiedlichen Massstäben erkennen. Er empfiehlt jedoch, die FPT in zwei Schritten anzuwenden. Im ersten Schritt wird die ganze Trajektorie analysiert, während im zweiten Schritt ein zuvor identifiziertes Segment analysiert wird, um allfällige Verhaltensmuster innerhalb des Segments zu erkennen (Pinaud 2008). Zum Beispiel können Zwischenstopps innerhalb der Migration zu weiterführenden Informationen führen. Einerseits können reguläre Pausen erkannt werden, jedoch auch nicht zu erwartende Pausen aufgrund von menschlichen Störungen oder anderen Einflüssen (Le Corre et al. 2014). Le Corre et al. (2014) haben die FPT ebenfalls in zwei Schritten angewendet. Sie nutzten die FPT um die Winterpause sowie den Zeitpunkt der Kalbung zu identifizieren um damit Start und Ende der Frühjahresmigration von migrierenden Karibus zu datieren. Auch sie verwendeten Lavielle's Methode um zu segmentieren. McKenzie et al. (2009) kombinieren die *mean FPT* mit einer Fokker-Planck Formel, um speziell die Verbreitung von Tieren im Raum zu erkennen. Sie haben territoriales Verhalten anhand des Rotfuchses (*Vulpes vulpes*) gezeigt. Dabei haben sie geschaut, wie lange der Rotfuchs hat, um seine Nahrung im Heimatgebiet zu lokalisieren. Le Corre et al. (2014) haben einen Radius gewählt basierend auf dem Maximum der mittleren Varianz der FPT. Diese war die Hälfte der Grösse der mittleren Winterpause ($100 \pm 3 \text{ km}$) und ähnlich zur Grösse der Fläche zum Kalberungszeitpunkt ($55 \pm 1 \text{ km}$). Die Grösse des Kreises wurde aber kleiner gewählt, als dass sie eine ganze Migrationszeitspanne beinhaltet hätte.

Lavielle's (2005) Segmentation: Penalized Contrast Method (PCM)

Die Methode von Lavielle basiert auf einer *minimized penalized contrast function* und die CP werden anhand von homogenem Mittelwert oder der Varianz gefunden (Barraquand & Benhamou 2008, Le Corre et al. 2014). Die Nutzer müssen Entscheidungen bezüglich drei Parametern machen. Erstens, die Variable, mit welcher die Methode segmentieren soll (Mittelwert, Varianz oder beide), zweitens, die minimale Länge eines Segments (*l.min*) und drittens, die maximale Anzahl Segmente, welche identifiziert werden darf (*Kmax*). Barraquand und Benhamou (2008) haben bei der Verwendung der FPT die gemittelten FPT Werte empfohlen, daher haben auch Le Corre et al. (2014) sich für den Mittelwert entschieden. Barraquand und Benhamou (2008) haben eine minimale Segmentlänge definiert (10, Kreidurchmesser/*step length*), eine maximale Segmentanzahl (30) und den *tracking mode* für Mittelwert und Varianz gesetzt. Le Corre et al. (2014) haben den Mittelwert ausgewählt, die minimale Länge der Segmente als 1 definiert und die maximale Anzahl Segmente bei 30 angesetzt. Lavielle (2005) empfiehlt für *Kmax* einen Wert zu nehmen, der höher ist als für die Daten angenommen wird. Le Corre et al. (2014) haben auch untersucht, wie sich die Methode verhält, wenn sie die genaue Anzahl Segmente (vorhandenes

Wissen aufgrund der Simulation) verwendeten. Sie kamen zum Schluss, dass mehr Fehler entstanden als wenn sie keine genaue Anzahl vorgegeben haben.

3.2.2 Behavioural Change Point Analysis (BCPA)

Behavioural Change Point Analysis ist eine Segmentierungsmethode, welche 2009 in der Publikation von Gurarie et al. (2009) vorgestellt wurde. Seither wird BCPA oft verwendet (Kranstauber et al. 2011, Garstang et al. 2014) und auch mit weiteren Methoden ergänzt (Garstang et al. 2014, Morelle et al. 2017a, Zhang et al. 2015). Diverse Datensätze von nördlichen Seebären (Gurarie et al. 2009), Zwergpinguinen (Zhang et al. 2015), Neunaugen (Gurarie et al. 2016a), Wölfen (Gurarie et al. 2016a), Wildschweinen (Morelle et al. 2017a), Elefanten (Garstang et al. 2014) und weiteren Tierarten, wurden schon durch die BCPA untersucht und segmentiert.

BCPA basiert auf Techniken, welche ihren Ursprung in der Zeitreihenanalyse und der Signalverarbeitung haben (Gurarie et al. 2016a).

BCPA ist eine Methode, welche auf der Wahrscheinlichkeit beruht, dass ein *Change Point* (CP) an einer gewissen Stelle liegt. Dazu werden zuerst allgemeine Informationen zu Distanz (*step length*), Geschwindigkeit (*approximate speed*) und Orientierung (*turning angle*) sowie weiteren Werten berechnet. Diese Informationen werden im nächsten Schritt weiterverwendet. In diesem Schritt wird ein *Window* auf der Trajektorie verschoben und innerhalb von jedem *Window* der wahrscheinlichste CP berechnet. Dabei wird jeder Punkt innerhalb des *Windows* einmal als CP betrachtet und aufgrund der zuvor generierten Informationen berechnet, ob sich der Mittelwert, die Varianz und/oder die *characteristic time* oder Autokorrelation vor beziehungsweise nach dem Punkt unterscheiden. Dann wird jedem Punkt das passendste Modell zugeordnet. Diese acht Modelle (M0 – M7) beschreiben keine Parameterveränderung (M0: Mittelwert, Varianz und *characteristic timescale*/Autokorrelation) zu einem Parameter, der sich verändert bis zu allen Parameter verändern sich (Gurarie et al. 2014). Die Werte der drei statistischen Masse beschreiben die *persistence* und *turning velocity*, welche durch die Transformierung der *step length* und des *turning angles* berechnet werden. Das heisst, eine Zunahme des Mittelwerts der *persistence velocity* oder der Autokorrelation weist auf eine schneller werdende und direkte Bewegung hin (z.B. Migration) und ein Anstieg in der Varianz auf Anhalten und Weitergehen (z.B. Nahrungssuche) (Garstang et al. 2014). Ob dieser Punkt ein CP ist, wird schlussendlich aufgrund des *Bayesian Information Criterion* (BIC) entschieden. Dieses BIC vergleicht die acht verschiedenen Modellannahmen; ob es keine Änderungen gibt, oder ob sich ein, zwei oder alle drei Parameter ändern (Edelhoff et al. 2016a). BCPA wählt also die Stelle des wahrscheinlichsten CP im Fenster (*window*) anhand des tiefsten BIC.

MATERIAL

Teimouri et al. (2018) haben, inspiriert durch Thiebault und Tremblay (2013), die Trajektorie zweimal segmentieren lassen, einmal unter Verwendung der *persistence velocity* und einmal mit der *turning velocity*. Dies soll die Zuverlässigkeit der Wahl der CP verbessern.

BCPA ist robust gegenüber unregelmässigen Zeitintervallen zwischen den Punkten (Gurarie et al. 2016a) sowie fehlenden Daten und Ausreissern (Gurarie et al. 2009). Sie ist rechnerisch effizient implementiert und einfach anzuwenden (Gurarie et al. 2009). Die BCPA beinhaltet die Autokorrelation explizit in der Wahrscheinlichkeitsfunktion (*likelihood function*) (Teimouri et al. 2018). Fleming et al. (2014) weisen darauf hin, dass bei einer Methode wie der BCPA, welche Trajektorien Schritt für Schritt basierend auf der *sampling rate* analysiert, die zeitliche Skala eines Bewegungsprozesses ähnlich zum *sampling interval* sein sollte, um vernünftige Resultate zu generieren.

Teimouri et al. (2018) weisen auf die Wichtigkeit der Wahl der *window size* hin. Um diese auszuwählen variierten sie die *window size* und untersuchten die *diagnostic plots*. Sie wählten eine *window size* von 30 Datenpunkten (in ihrem Fall ist dies ein Zeitintervall von 30 Minuten). Die Ortsinformationen wurden alle 10s aufgenommen und zur Analyse auf 1min gemittelt. Zhang et al. (2015) wählten ebenfalls eine *window size* von 30 Datenpunkten, um die von Gurarie et al. (2009) empfohlene Minimalanzahl nicht zu unterschreiten, jedoch aber die Wechsel im Verhalten der Tiere auf kleinstmöglichem Massstab zu identifizieren. Das *sampling interval* ihrer Daten liegt bei 1min.

Morelle et al. (2017a) verwendeten eine *window size* von 30, der Sensitivitätsparameter K wurde 2 gesetzt und die Clustergrösse auch 2. In diesen Daten wurde eine *sampling frequency* von 15min verwendet.

Ein Nachteil der BCPA ist laut Teimouri et al. (2018), dass sie nur CP identifiziert, aber kein weiterer ökologischer Kontext dazu liefert. Um also tieferen Einblick in die *behavioural states* zu bekommen und die Segmente noch zu klassieren, müssen der Segmentierung durch die BCPA anschliessend noch weitere Methoden folgen (Teimouri et al. 2018).

4 METHODIK

4.1 Methoden

Dieses Kapitel beschreibt den methodischen Rahmen, in welchem die Simulationen segmentiert und analysiert werden. Es werden die verwendeten Funktionen sowie die Evaluationskriterien beschrieben. Anschliessend wird das Vorgehen präsentiert, welches mit unterschiedlichen Parametern und Zusätzen den methodischen Rahmen nutzt, um vergleichbare Resultate zu generieren.

4.1.1 Methodischer Rahmen

Der Ausgangspunkt dieses Projekts ist stark an die Publikation von Gurarie et al. (2016a) angelehnt. Der im R verwendete Code wurde von ihnen übernommen und an dieses Projekt angepasst. Dabei wurden alle Schritte aneinandergehängt und mit einer alles umfassenden Funktion in einen sinnvollen Ablauf gebracht. Im Folgenden werden die einzelnen Schritte und Teilfunktionen kurz erläutert.

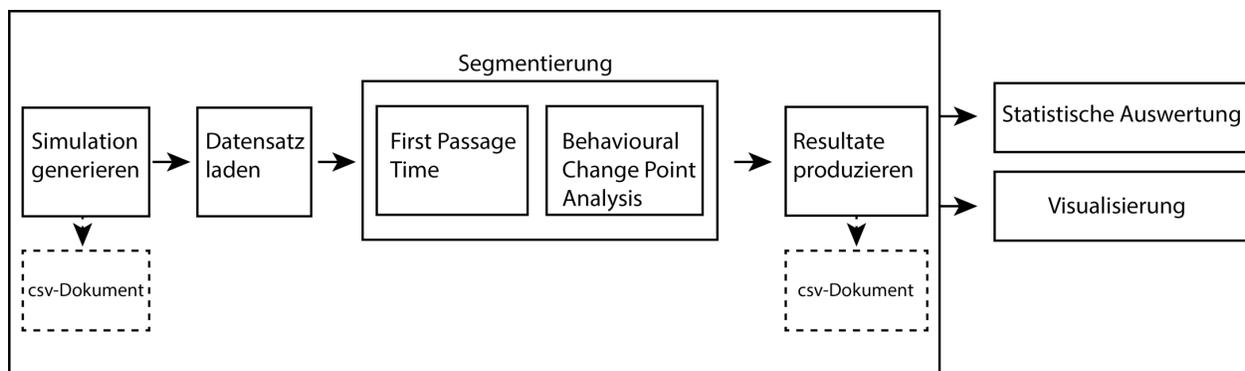


Abbildung 3: Aufbau des methodischen Rahmens. Die grosse Box zeigt den Inhalt der alles umfassenden Funktion (call Funktion). Im Schritt «Simulation generieren» sowie im Schritt «Resultate produzieren» wird jeweils ein csv-Dokument abgespeichert. Der Segmentierungsschritt beinhaltet die beiden Methoden FPT und BCPA. Ausserhalb der grossen Box werden dann statistische Auswertungen vorgenommen und Resultate visualisiert.

4.1.2 Beschreibung der Funktionen

Simulation – `Get.multiCVM.simulation()`

Im ersten Schritt werden die Simulationen erzeugt. Die dafür verwendete Funktion ist die `multiCVM()`-Funktion. Diese ist Bestandteil des `WADDLE` packages von Gurarie et al. (2016a) und simuliert ein *correlated velocity movement continuous time model* (CVM). Die `multi-CVM()`-Funktion ist so gestaltet, dass die Nutzerin oder der Nutzer die mittlere Geschwindigkeit «`nu`» und die *characteristic time scale* «`tau`» bestimmen kann um ein Model mit verschiedenen Phasen zu generieren. Diese Simulation ermöglicht es, laut Gurarie et al. (2016c), die Robustheit unterschiedlichster Segmentierungsmethoden zu

METHODIK

testen. Die multiCVM-Funktion wird in diesem Projekt innerhalb der `Get.multiCVM.simulation`-Funktion erweitert. Das Ziel dieser Funktion ist es, ein csv-Dokument zu erstellen, welches die Koordinaten, eine beliebige Zeit als POSIXct (Start bei 2000-01-01) sowie den CP und die Segmentnummer (basierend auf der durch die Nutzerin oder den Nutzer definierten Phasenlänge) beinhaltet. Dieses Dokument soll für weitere Analysen (allenfalls auch ausserhalb von R) zur Verfügung stehen.

- Input: *characteristic timescale*, mittlere Geschwindigkeit, Phasenlänge, *sampling frequency*, Name des Dokuments welches generiert wird
- Output: csv-Dokument mit Koordinaten, Zeit als POSIXct, CP, Segmentnummer, *characteristic timescale*

Vorbereitung der Daten – Load.datasets()

Um auch reale Datensätze verwenden zu können, kann in der `load.datasets`-Funktion ein Dokument in csv-Format, `csv.gz`-Format oder `.rda` eingelesen werden. Das zuvor durch die `get.multiCVM.simulation`-Funktion generierte Dokument wird hier eingelesen. Zu Beginn werden die Spaltennamen vereinheitlicht. Die X-Koordinate wird zu «x», die Y-Koordinate zu «y» und der Spaltenname «date» wird für das Datum und die Zeit verwendet. Teilweise muss die Zeit in das POSIXct-Format umgewandelt werden. Datensätze, welche aus mehreren Individuen bestehen, werden zu einzelnen *dataframes* in eine Liste gespalten. Innerhalb eines *dataframes* eines Individuums werden doppelte Daten entfernt. Zum Schluss wird jedem Datenpunkt eine eigene ID angefügt.

- Input: der Name des Dokuments
- Output: Liste mit einem oder mehreren *dataframes* mit einheitlichen Spaltennamen und IDs

Segmentierungsmethode 1: FPT – Get.FPT.Gurarie()

In dieser Funktion wird die First Passage Time (FPT) durchgeführt. Dazu muss der Datensatz in ein `ltraj`-Objekt umgewandelt werden. Dann werden zuerst die FPT und die Varianz der FPT für mehrere Radien (5 bis zur Hälfte der Anzahl Datenpunkte) berechnet. Die Varianz entgegen der Radien zu visualisieren hilft, Flächen mit erhöhter Aktivität (*search effort*) zu identifizieren. Diese werden durch die Maxima in der Varianz bestimmt (Fauchald & Tveraa 2003). Anschliessend wird die FPT mit dem von den Nutzern gewählten Radius erneut berechnet. Um die FPT nicht nur visuell interpretieren zu können, wird sie durch die Funktionen von Lavielle (`lavielle()`, `chooseseg()` und `findpath()`; Lavielle, 2005) segmentiert. Das Resultat dieser `get.FPT.Gurarie`-Funktion sind die CP-Zuweisung zum jeweiligen Datenpunkt, die Segmentnummer und die Anzahl Datenpunkte pro Segment sowie die durch die `fpt`-Funktion generierten Werte.

- Input: *dataframe* (Output von `load.datasets()`), Radius
- Output: *dataframe* mit zusätzlichen CP von FPT, Segmentnummer, Anzahl Datenpunkte pro Segment und FPT-Werte

Segmentierungsmethode 2: BCPA – Get.BCPA()

Da die `GetVT()`-Funktion aus dem `bcpa`-package (Gurarie et al. 2014) bestimmte Spaltennamen braucht, werden die Spalten in einem ersten Schritt wieder umbenannt: «x» zu «X», «y» zu «Y» und «date» zu «Time». Die `GetVT`-Funktion berechnet dann die Werte gewisser Eigenschaften (z.B. *turning angle*, geschätzte Geschwindigkeit etc.) über das durch die Nutzerin oder den Nutzer gewählte Zeitintervall (*sec*, *min*, *hour*, *day*). Anschliessend verwendet die `WindowSweep`-Funktion, die durch die `GetVT`-Funktion generierten Werte, um die eigentliche BCPA-Analyse durchzuführen. Diese kann als Parameter die *persistence velocity* ($V \cdot \cos(\Theta)$) oder die logarithmierte Geschwindigkeit ($\log(v)$) benutzen. Zusätzlich zu diesem Parameter müssen andere Parameter von der Nutzerin oder dem Nutzer definiert werden. Dies sind die *window size*, *window step*, der Sensitivitätsparameter K und die Autokorrelation oder *characteristic timescale* (Gurarie et al. 2009 `bcpa`). In einem weiteren Schritt werden die CP durch die `ChangePointSummary()`-Funktion extrahiert. Hier muss die Nutzerin oder der Nutzer die *cluster width* bestimmen, mit welcher nahe zusammenliegende CP zu einem CP zusammengefasst werden. Und auch hier werden am Schluss den Datenpunkten des *dataframes* die CP und die Segmentnummer hinzugefügt und mit der Segmentanzahl und der Anzahl Datenpunkte pro Segment zu einer Liste zusammengefügt. Ebenfalls dieser Liste hinzugefügt werden die Werte aus der `GetVT`-Funktion und der `WindowSweep`-Funktion.

- Input: *dataframe* (Output von `load.datasets()`), Zeiteinheit für die Analyse, *window size*, *window step*, Sensitivitätsparameter, *cluster width*
- Output: Liste von *dataframes* mit zusätzlichen CP von BCPA, Segmentnummer, Anzahl Datenpunkte pro Segment, Anzahl Segmente, Werte aus `GetVT`-Funktion, Werte aus `WindowSweep`-Funktion

Produktion von Resultaten – Get.results()

Diese Funktion wird nur ausgeführt, wenn der Ausgangsdatensatz eine Simulation war. Hier werden Resultate generiert, welche auf dem Vergleich von identifizierten CP und den durch die Simulation vorgegebenen CP basieren. Dazu gehören die Anzahl Segmente, die Anzahl Datenpunkte pro Segment (Segmentlänge), die Anzahl korrekt identifizierter Segmente (mit maximaler Abweichung ± 3 vom CP), die Anzahl CP, die korrekt identifizierten CP welche exakt, ± 1 , ± 2 oder ± 3 vom realen CP entfernt liegen,

METHODIK

sowie diejenigen mit einer maximalen Abweichung von ± 1 , ± 2 und ± 3 , die Precision und der Recall der CP.

- Input: Kombiniertes Output von Simulation, FPT und BCPA als Liste
- Output: Liste mit allen Evaluationskriterien (Anzahl Segmente, Anzahl Datenpunkte pro Segment (Segmentlänge), Anzahl korrekter Segmente (mit maximaler Abweichung ± 3 vom CP), Anzahl CP, korrekt identifizierte CP welche exakt, ± 1 , ± 2 oder ± 3 vom realen CP entfernt liegen sowie diejenigen mit einer maximalen Abweichung von ± 1 , ± 2 und ± 3 , die Precision der CP und Segmente und der Recall der CP)

Alles in einem Ablauf – Call_fun()

Diese fünf Funktionen (`get.multiCVM.simulation`, `load.datasets`, `get.FPT.Guarie`, `get.BCPA` und `get.results`) werden in einer einzigen, grossen Funktion – der `call_fun`-Funktion – kombiniert. Der Nutzer kann also, anstatt jede Funktion einzeln laufen zu lassen, alle Parameter, welche die fünf Funktionen brauchen, auf ein Mal definieren und erhält dann alle automatisch berechneten Outputs gleichzeitig. Zudem beinhaltet die `call_fun`-Funktion einige weitere Parameter, die durch den Nutzer definiert werden müssen. Dazu gehören, ob eine Simulation verwendet wird oder nicht, der Simulations- oder Datensatzname (aufgeteilt in zwei Parameter) sowie die Anzahl Simulationen mit den gleichen Parametern, die generiert werden sollen. Anschliessend werden alle Outputs der fünf Funktionen zu einer verschachtelten Liste zusammengefügt und als Endresultat ausgegeben. Zusätzlich wird ein neues csv-Dokument abgespeichert, welches neben der ID, den Koordinaten, den Daten, der Zeit, den CP und der Segmentnummer der Simulation auch die CP und Segmentnummern der FPT und BCPA beinhalten sowie die *persistence velocity*, die *turning angle*, die geschätzte Geschwindigkeit und die *step length*.

- Input: Parameter aller fünf Funktionen sowie ob eine Simulation vorhanden ist oder nicht, Name der Simulation, Name des Datensatzes, Anzahl Simulationen
- Output: verschachtelte Liste mit den Outputs der fünf Funktionen

Statistische Auswertungen – Get.statistics()

Um die generierte Datenmenge statistisch auswerten zu können, wird die `get.statistics`-Funktion verwendet.

- Input: Name der verschachtelten Liste (Output von `call_fun`-Funktion), eine Liste der CP (mit/ohne Abweichung)
- Output: Diverse statistische Ergebnisse: Boxplot zu Anzahl Segmente, Anzahl CP, Precision und Recall der FPT mit Abweichungen, Precision und Recall der BCPA mit Abweichungen, Tabelle

mit Anzahl und Indizes der identifizierten CP und der korrekt identifizierten CP für FPT und BCPA; Liste mit den statistischen Werten für jede der 50 Realisierungen, Liste mit Simulationsname und identifizierten CP sowie allfälligen identifizierten CP zwischen den eigentlich korrekten CP

Visualisierungen – Lets.plot()

Diese Funktion beinhaltet die Visualisierung einer Realisierung.

- Input: Name eines *dataframes* innerhalb der verschachtelten Liste, ein Farbvektor, der Radius der FPT
- Output: Diverse Grafiken: FPT-Plot, BCPA-Plot Original, BCPA-Plot, Diagnostic Plot Original, geschätzte Geschwindigkeit-Zeit-Plot, Turning angle-Zeit-Plot, Step length-Plot

4.1.3 Packages

Um die Funktionen wunschgemäss in R abzuspielen, sind folgende Packages notwendig:

Tabelle 5: In diesem Projekt verwendete R-Packages.

R-package	Verwendung	Literatur
adehabitatLT	FPT	Calenge et al. 2009
bcpa	BCPA Segmentierung	Gurarie et al. 2014
base	Diverse Datenbearbeitung	R-Core-Team 2013
gplots	Textplots	Warnes, G.R. 2016
graphics	Grafische Darstellungen	R-Core-Team 2013
plyr	Datenverarbeitung	Wickham 2016
rio	Datensatz laden	Becker et al. 2018
utils	Datensatz laden / csv abspeichern	R-Core-Team 2013
waddle	Simulation, BCPA Grafik	Gurarie et al. 2016c

4.1.4 Evaluationskriterien

Damit die generierten Resultate der Methoden nicht nur visuell verglichen werden können, werden zusätzlich einige statistische Grössen verwendet. Für einige ökologische Fragen ist es wichtig, dass die Segmente, das heisst Start- und Endpunkt eines Segments korrekt identifiziert werden. Für andere ökologische Fragestellungen sind die CP als einzelne Punkte oder als Gruppe wichtiger. Zum Schluss empfiehlt sich jedoch in jedem Fall, bestehendes Wissen zum Verhalten der untersuchten Tierart bei der

nachträglichen Analyse der Segmentierung miteinzubeziehen. Die Segmente lassen sich dann einfacher dem Verhalten zuweisen, welches die Tiere in diesem Moment vermutlich hatten (Le Corre et al. 2014).

Segmente

Bei verschiedenen Fragestellungen besteht ein Interesse an der Länge der Segmente. Wie lange dauert die Migrationsphase dieses Tieres? Unterscheidet sie sich von anderen Individuen derselben Kolonie? Gibt es Unterschiede zu letztjährigen oder noch älteren Migrationsphasen? Kann man diese mit dem Klimawandel oder menschlichen Einflüssen in einen Zusammenhang bringen? Aber auch Fragen zu veränderten Routen oder Nahrungsplätzen können von Interesse sein. Wie verändert sich die Migrationsroute? Wohin wird sich die Route möglicherweise verschieben? An welchen Plätzen machen die Tiere Rast? Wie sieht es mit den Bedingungen dort aus? Oder es interessieren Fragen zur Anzahl Segmente. Stimmt die Anzahl der unterschiedlichen Verhaltensphasen mit den erwarteten Phasen überein? Gibt es Individuen mit unterschiedlichen Verhaltensphasen? Wie ausgeprägt sind diese?

- *Anzahl Segmente*: Dieses Kriterium gibt die Anzahl der Segmente wieder. Einerseits diejenigen der Simulation und andererseits die jeweilige Anzahl identifizierter Segmente durch die beiden Methoden.
- *Anzahl Datenpunkte pro Segment*: Dieses Kriterium liefert die Anzahl Datenpunkte für jedes Segment, wiederum für die Simulation und die beiden Methoden. Aus diesem Kriterium lässt sich die zeitliche Dauer des jeweiligen Segments berechnen.

Change Points

Besteht ein Interesse an den CP, kann zum Beispiel die zeitliche Information eine Rolle spielen. Mit Bezug zu obigem Beispiel (Migration) könnten folgende Fragen auftreten: Wann ist der Beginn der Migrationsphase? Über welchen Zeitraum starten die unterschiedlichen Individuen bzw. wann sind sie wieder an ihrem Ausgangsort zurück? Wie hat sich der Zeitpunkt des Beginns oder Endes mit den Jahren verändert? Gibt es Zwischenstopps, die es zuvor nicht gegeben hat? Kann man dies bestimmten Umwelteinflüssen zuweisen?

- *Anzahl CP*: Dieses Kriterium liefert die Anzahl CP, einerseits von der Simulation und andererseits die Anzahl der von den beiden Methoden identifizierten CP.
- *Anzahl korrekt identifizierten CP*: Hier werden die identifizierten CP mit denjenigen aus der Simulation verglichen und die Anzahl übereinstimmender CP gezählt. Dabei werden unterschiedliche Abweichungen miteinbezogen. Es werden die exakt identifizierten, die um 1 Datenpunkt verschoben, die um 2 Datenpunkte verschoben und die um 3 Datenpunkte

verschobenen CPs betrachtet. Zudem werden sie mit einer maximalen Abweichung von ± 1 , ± 2 und ± 3 vom korrekten CP angeschaut.

- *Precision*: der Anteil korrekt identifizierter CP an allen identifizierten CP

$$\frac{\# \text{ korrekt identifizierte CP}}{\# \text{ identifizierte CP}}$$

- *Recall*: der Anteil korrekt identifizierter CP an den tatsächlichen CP

$$\frac{\# \text{ korrekt identifizierte CP}}{\# \text{ tatsächliche CP}}$$

Sowohl für die Precision wie auch für den den Recall werden jeweils die korrekt identifizierten CPs sowie diejenigen Datenpunkten mit einer maximalen Abweichung von ± 1 , ± 2 und ± 3 verwendet.

4.2 Vorgehen

4.2.1 Repliziert

Ausgangspunkt dieses Projekts ist die Replizierung der Simulationen von Gurarie et al. (2016a) und Soleymani et al. (2017). Dies sind die speed switch, time-scale switch und continous speed switch Simulation. Gurarie et al. (2016a) ermutigen interessierte Leserinnen und Leser in ihrer Publikation direkt, die Analyse selbst zu replizieren. Dazu stellen sie den R-Code und die Beispiele als zusätzliches Material zur Verfügung (WADDLE package, Gurarie et al. 2016c). Dies wird in diesem Projekt in einem ersten Schritt gemacht. Im Gegensatz zu Gurarie et al. (2016a) werden die durch die FPT und BCPA segmentierten Trajektorien aber nicht nur visuell dargestellt, sondern die Anzahl CP, deren Indizes, die Anzahl Datenpunkte pro Segment sowie die Anzahl korrekt identifizierten CP werden zusätzlich berechnet.

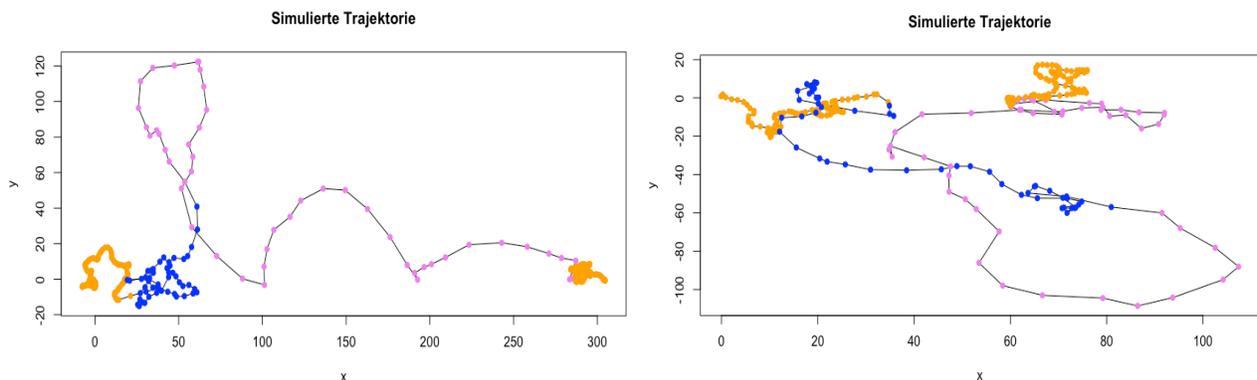
Die Parameter werden wie im Grundlagepaper von Gurarie et al. (2016a) bei allen drei Simulationen wie folgt gesetzt: Der Radius für die FPT beträgt 10 Einheiten. Die *window size* der BCPA beträgt 30 Datenpunkte, der *window step* 1, K ebenfalls 1 und die *cluster width* liegt bei einer zeitlichen Spanne von 3.

Auffällig dabei ist, dass die Simulationen mit jeder Realisierung anders aussehen und die Methoden zu anderen Resultaten kommen. Diese Unterschiede entstehen aufgrund des in der Simulation integrierten Rauschens. Aus diesem Grund werden im nächsten Schritt mehrere Realisierungen einer Simulation durchgeführt.

4.2.2 50 Realisierungen

Da die Simulation ein Rauschen beinhaltet, sieht jede Realisierung, obwohl sie aus denselben Parametereinstellungen besteht, auf den ersten Blick sehr unterschiedlich aus. Gurarie et al. (2016a) haben eine einzige, komplett willkürliche Realisierung jedes Simulationstyps gewählt und an diesen als illustrierendes Beispiel ihre explorative Analyse durchgeführt. Um ihre Aussagen zu den Methoden zu bestätigen oder allenfalls zu bemängeln, werden in diesem Projekt mehrere Realisierungen des jeweiligen Simulationstyps durchgeführt. Es werden jeweils 50 Realisierungen der *speed switch*, der *tortuosity switch* sowie der *continuous-speed switch* Simulation durchgeführt. So kann die Robustheit der Methoden getestet werden.

Auch im ökologischen Sinne ist die Verwendung von mehreren, leicht unterschiedlichen Realisierungen sinnvoll. Angenommen, eine Ökologin oder ein Ökologe möchte die Routen einer Herde von Tieren analysieren, so werden die Routen der einzelnen Tiere in der Herde sowohl zeitlich wie auch räumlich leichte Differenzen aufweisen. Ist die Methode aber robust gegenüber kleiner räumlicher und zeitlicher Variabilität, kann sie nahezu bedenkenlos auf die Individuen der Herde angewendet werden. Deshalb sollte sie möglichst zuverlässige Resultate liefern.



4.2.3 Maximale Abweichung

Während den ersten beiden Schritten ist aufgefallen, dass die durch die Methoden identifizierten CPs zwar selten genau auf den CPs der Simulation zu liegen kommen, oftmals aber nur wenige Datenpunkte vom korrekten CP entfernt bestimmt wurden. Um dieser Beobachtung nachzukommen, werden in einem dritten Schritt neben dem exakten CP jeweils auch drei unterschiedliche maximale Abweichungen berücksichtigt. Diese maximalen Abweichungen werden definiert als ± 1 , ± 2 und ± 3 . Damit werden von den Methoden

gefundene CP, welche jeweils bis maximal einen, zwei oder drei Datenpunkte vor oder nach dem exakten CP identifiziert werden, ebenfalls als korrekt gezählt.

Ökologisch gesehen ist die Beachtung der Abweichungen auf jeden Fall von Vorteil. Je nach Zeitintervall der Datenaufnahme verläuft der Wechsel im Verhalten der Tiere nämlich über mehrere Datenpunkte ab. Dadurch sind sie für die Methoden schwieriger zu erkennen und können Verzögerungen hervorrufen. Diese Problematik kann mit der Berücksichtigung von maximalen Abweichungen aufgefangen werden.

4.2.4 FPT: mittlerer CP

Dieser Schritt bezieht sich auf die Aussagen von Gurarie et al. (2016a) und Soleymani et al. (2017) sowie auf Beobachtungen der ersten drei Schritte. Die Änderung der Geschwindigkeit von 5 auf 10 hat kleinere Auswirkungen auf das Resultat der Methoden ergeben als die Änderungen von 1 auf 5 oder von 10 auf 1 zurück. Dies könnte auch an einer Unregelmässigkeit in den Daten liegen, die jedoch keine Verhaltensänderung beschreibt. Es muss also untersucht werden, wie neben klaren Verhaltensänderungen auch subtilere identifiziert werden können. Dazu werden die Parameter der Simulation wie folgt geändert:

Die erste Simulation verfügt über Geschwindigkeitsänderungen von 5 auf 10 zurück auf 5 und wieder auf 10. Jede Phase dauert 50 Datenpunkte (und damit 50s). Die zweite Simulation verfügt nur über zwei Phasen à je 100 Datenpunkte und die Geschwindigkeit springt von 5 auf 10. Die dritte Simulation kombiniert die ersten beiden Simulationen, indem die Geschwindigkeit sich wie in der ersten Simulation ändert (5, 10, 5, 10), die Phasen jedoch 100 Datenpunkte lang sind. (vgl. Kapitel 3.1.1 Definition der Parameter).

Diese Simulationen werden nun durch die FPT mit unterschiedlich grossen Radien segmentiert. Folgende Radien werden verwendet: 2, 3, 4, 5, 10, 15, 20, 25, 30, 40 und 50. Der Radius, mit welchem am meisten korrekte CP identifiziert werden, wird anschliessend erneut auf die *speed switch* Simulation angewendet.

4.2.5 BCPA: Wertevergleich der Parameter

Die BCPA verfügt über mehrere Parameter, für welche die Nutzerin oder der Nutzer die Werte selbst bestimmen kann. Doch welche Werte ergeben die beste Zusammensetzung bzw. das beste Resultat? Gibt es überhaupt eine beste Zusammensetzung? Und wofür ist dies die beste Zusammenstellung?

Dieser Schritt dient dazu, Antworten auf diese Fragen zu finden. Dazu werden die *speed switch* und die *time-scale switch* Simulation mit unterschiedlichen Parameterzusammensetzungen segmentiert. Die zu analysierenden Parameter sind die *window size*, der Sensitivitätsparameter K und die *cluster width*. Für die *window size* werden Werte in 10er-Schritten von 10 bis 50 verwendet. Für K die Werte 1, 2, und 4. Und für die *cluster width* die Werte 2 bis 8. Anhand der korrekt identifizierten CP soll schlussendlich die beste

Kombination gefunden werden. Diese Kombination wird dann sowohl auf die speed switch wie auch auf die time-scale switch Simulation erneut angewendet.

4.2.6 Kombination der Speed & Time-scale switch Simulation

In diesem letzten Schritt werden die Erkenntnisse aus den beiden vorherigen Schritten (Kapitel 4.3.4 und 4.3.5) angewendet. Dabei werden für die Geschwindigkeit dieselben Werte verwendet wie bei der *speed switch* Simulation und für die Tortuosität dieselben Werte wie bei der *time-scale switch* Simulation. Dies ergibt die Geschwindigkeitswerte 1, 5, 10, 1 und die Tortuositätswerte 2, 20, 40, 2. Die Phasenlänge dauert jeweils 100s (100 Datenpunkte pro Phase).

5 RESULTATE

In diesem Kapitel werden nun die Resultate aus den einzelnen Vorgehensschritten aufgezeigt. Die Unterkapitel beziehen sich eins zu eins auf diejenigen im Kapitel 4.2 Vorgehen. Im nächsten Kapitel werden die Resultate dann diskutiert und in einen grösseren Zusammenhang gesetzt.

5.1 Repliziert

Speed switch Simulation (Gurarie et al. 2016a)

In der Abbildung 6 (oben) ist zu erkennen, dass die erste und letzte Phase eine höhere Tortuosität und eine kleinere *step length* aufweisen. In der dritten Phase wird die grösste Distanz zurückgelegt, da die Geschwindigkeit am höchsten ist. In der Abbildung 6 (unten) lässt sich erkennen, dass die BCPA zusätzliche CP im dritten und im letzten Segment identifiziert hat. Die FPT hingegen hat nur jene zwei identifiziert, welche nahe am ersten und am letzten korrekten CP liegen.

RESULTATE

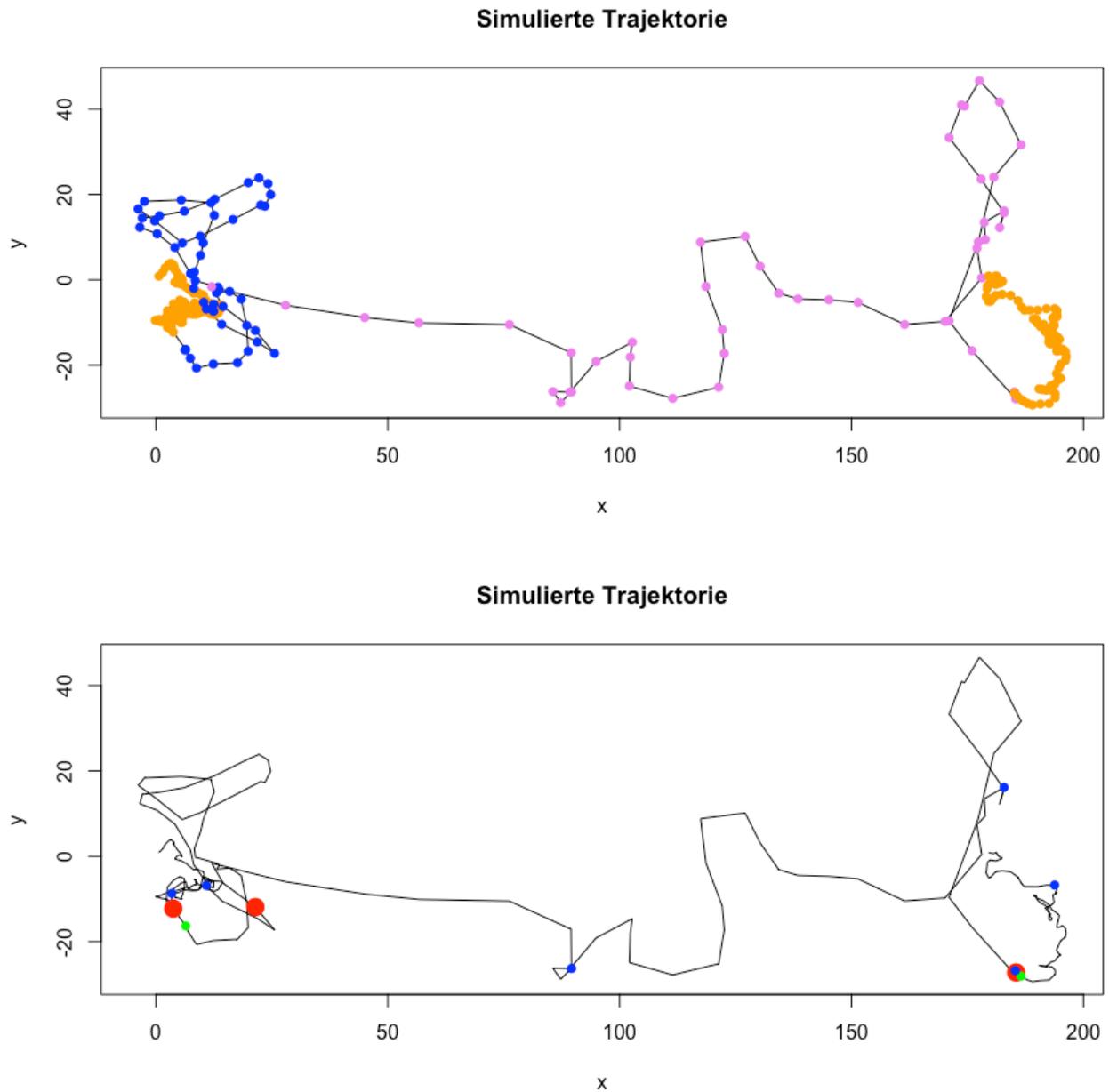


Abbildung 5: Trajektorie der speed switch Simulation (oben). Orange zeigt die erste und letzte Phase, blau die zweite und pink die dritte. Start ist bei Koordinate (0,0). Dieselbe Trajektorie mit eingezeichneten CP (unten). Definierte CP sind als roter und grosser Punkt markiert. Durch FPT identifizierte CP in grün. Durch BCPA identifizierte CP in blau.

In Abbildung 7 ist zu erkennen, dass die FPT in der Nähe des ersten und des letzten CP jeweils einen CP erkannt hat, den mittleren CP jedoch nicht. Zudem hat sie keine zusätzlichen CP identifiziert. Da die von der Methode gefundenen CP dem ersten und letzten CP nur nahe, nicht aber exakt darauf liegen, hat sie keinen CP korrekt identifiziert (Tabelle 6).

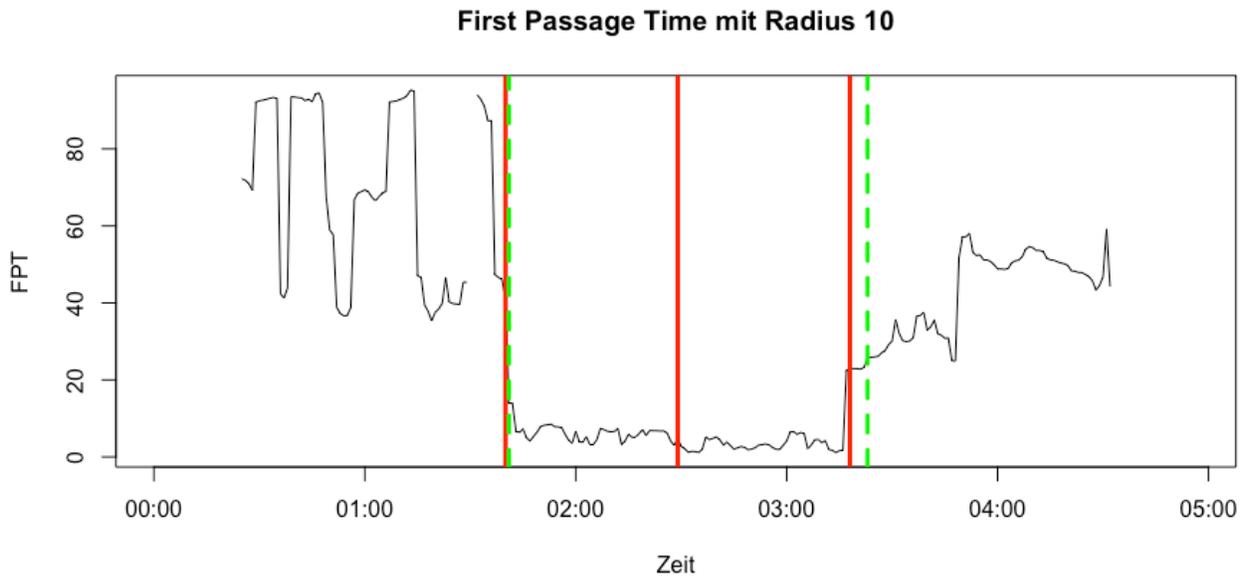


Abbildung 6: FPT der speed switch Simulation. X-Achse ist die Zeit [min], y-Achse die FPT. Definierte CP sind die roten vertikalen Linien, durch FPT identifizierte CP sind die grün gestrichelten vertikalen Linien.

Die BCPA hat ebenfalls keinen CP korrekt identifiziert (Abbildung 8, Tabelle 6). Beim zweiten und letzten CP liegt sie jedoch sehr nahe. Zusätzlich hat die Methode noch drei weitere CP identifiziert. Zwei davon liegen im dritten Segment, einer liegt im letzten Segment. In der Abbildung 8 (oben) liegt die time-scale meist bei 0, im fünften und im sechsten Segment (identifiziert durch die BCPA) liegt die time-scale bei 0.85, im letzten Segment bei 3.4.

RESULTATE

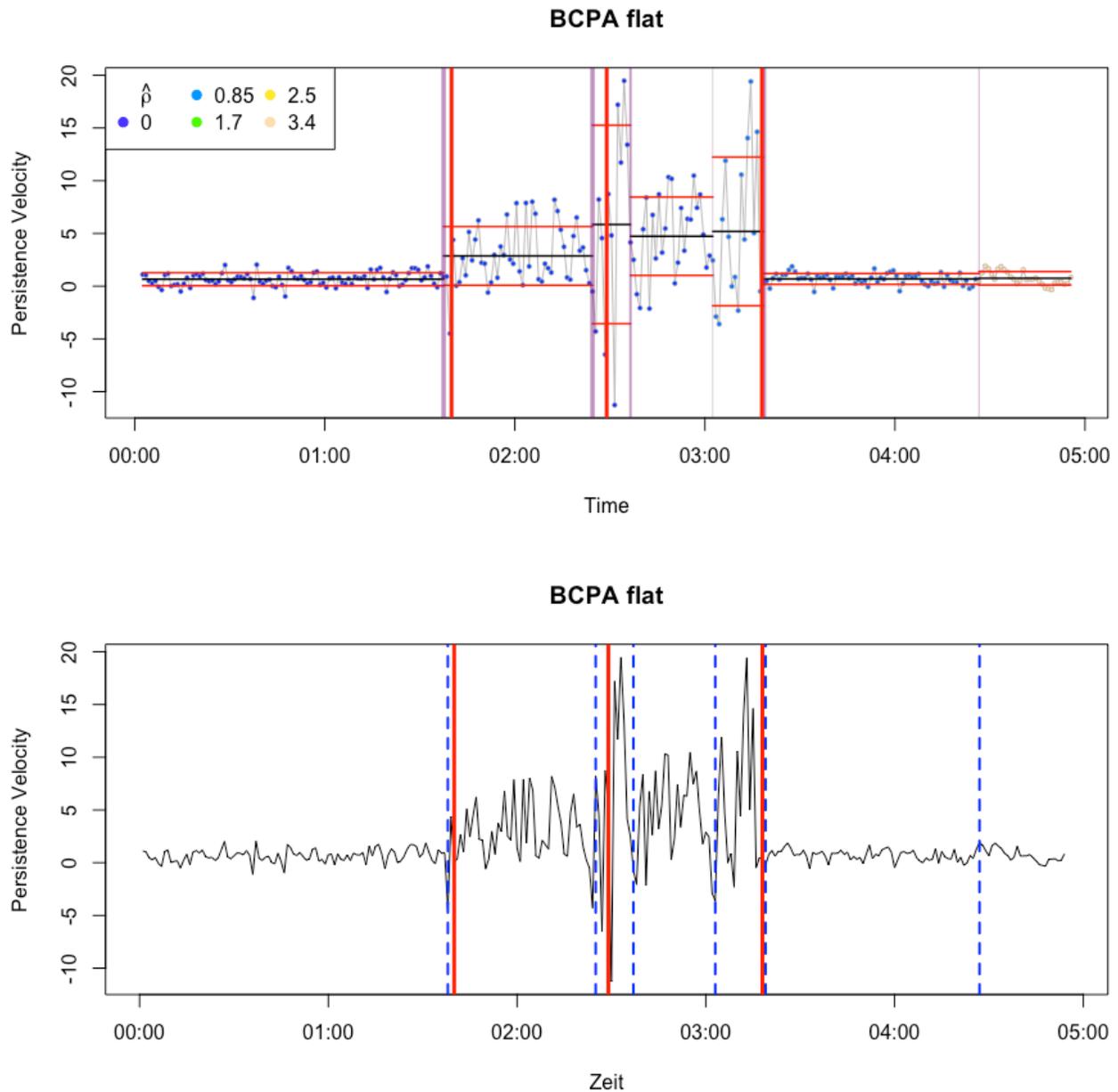


Abbildung 7: Grafik der BCPA der speed switch Simulation. X-Achse ist die Zeit [min], y-Achse die persistence velocity [map unit/min]. Oben repräsentieren die violetten vertikalen Linien die durch die BCPA identifizierten CP, die horizontalen schwarze Linien der geschätzte Mittelwert, die horizontalen roten Linien das 95% Konfidenzintervall. Die farbigen Punkte repräsentiere die erwartete time-scale. Je kühler die Farbe desto tiefer die Werte und desto höher ist die Tortuosität (und umgekehrt). Unten repräsentieren die roten vertikalen Linien die korrekten CP und die blau gestrichelten die durch die BCPA identifizierten CP.

RESULTATE

Tabelle 6: Daten der speed switch Simulation und der durch die FPT und BCPA segmentierten Trajektorie der speed switch Simulation.

	Simulation	FPT	BCPA
# CP	3	2	6
# Segmente	4	3	7
Indizes CP	100, 149, 198	101, 203	98, 145, 157, 183, 199, 267
# Datenpkt. pro Segm.	99, 49, 49, 99	100, 102, 94	97, 47, 12, 26, 16, 68, 30
# korrekt ident. CP		0	0

In der Abbildung 9 der *turning angles* (oben) erkennt man keinen Zusammenhang zwischen einer Winkeländerung und den CP. In der Abbildung zur geschätzten Geschwindigkeit (unten) lässt sich erkennen, dass bei einer grösseren Änderung der Geschwindigkeit zwischen zwei Datenpunkten ein CP erkannt wird. Die BCPA hat zusätzlich noch weitere CP erkannt, bei denen sich die Geschwindigkeit jedoch nicht stark verändert hat.

RESULTATE

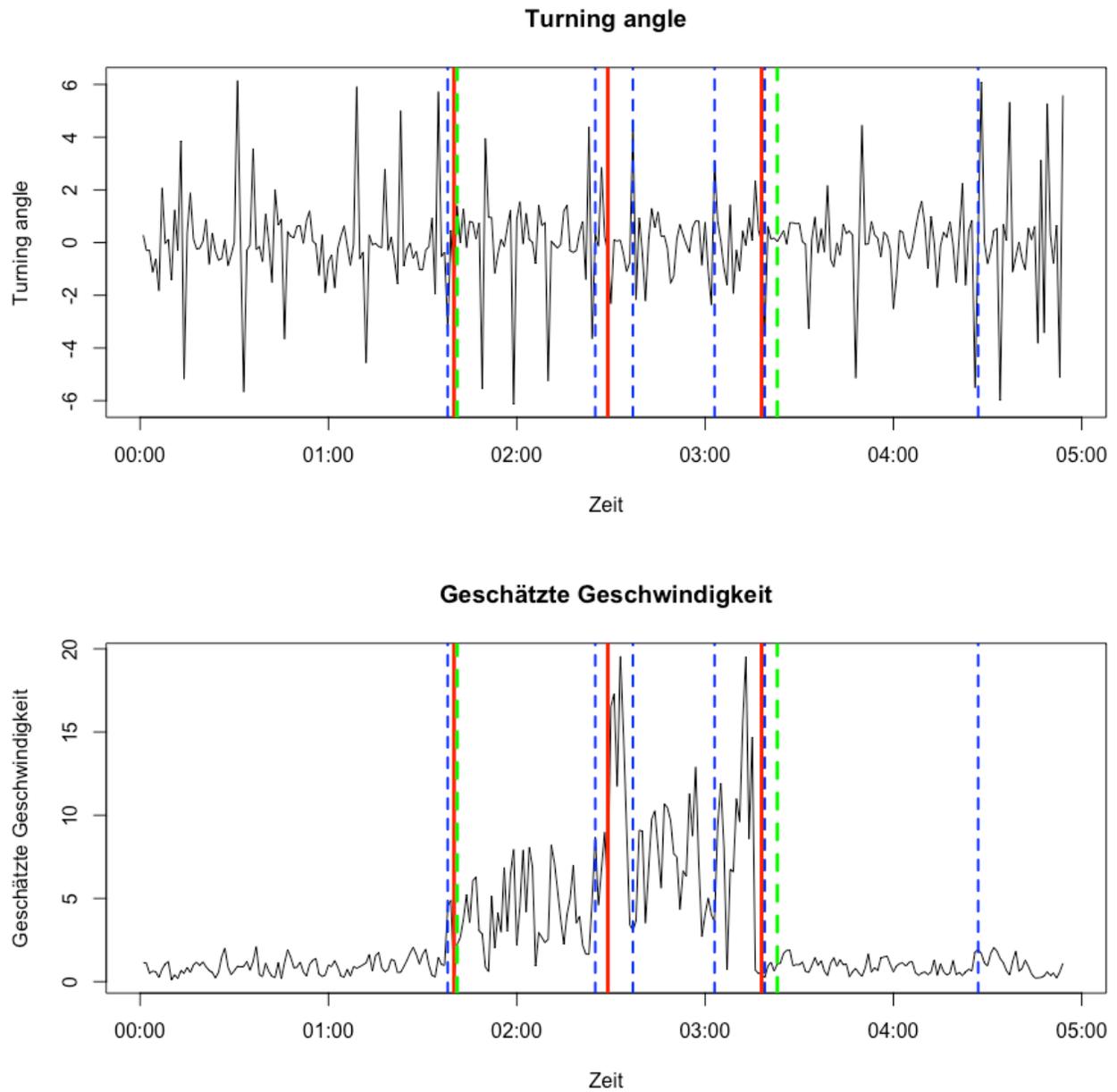


Abbildung 8: Turning angle (oben) und geschätzte Geschwindigkeit [map unit/min] (unten) für die speed switch Simulation. X-Achse ist die Zeit [min]. Rote vertikale Linien sind die korrekten CP, grün gestrichelt die durch die FPT identifizierten CP und blau gestrichelt die durch die BCPA identifizierten CP.

Time-scale switch Simulation (Gurarie et al. 2016a)

In Abbildung 10 ist erkennbar, dass das erste und letzte Segment die höchste Tortuosität aufweisen, das dritte Segment hingegen die niedrigste. Keine der beiden Segmentierungsmethoden hat einen CP korrekt identifiziert.

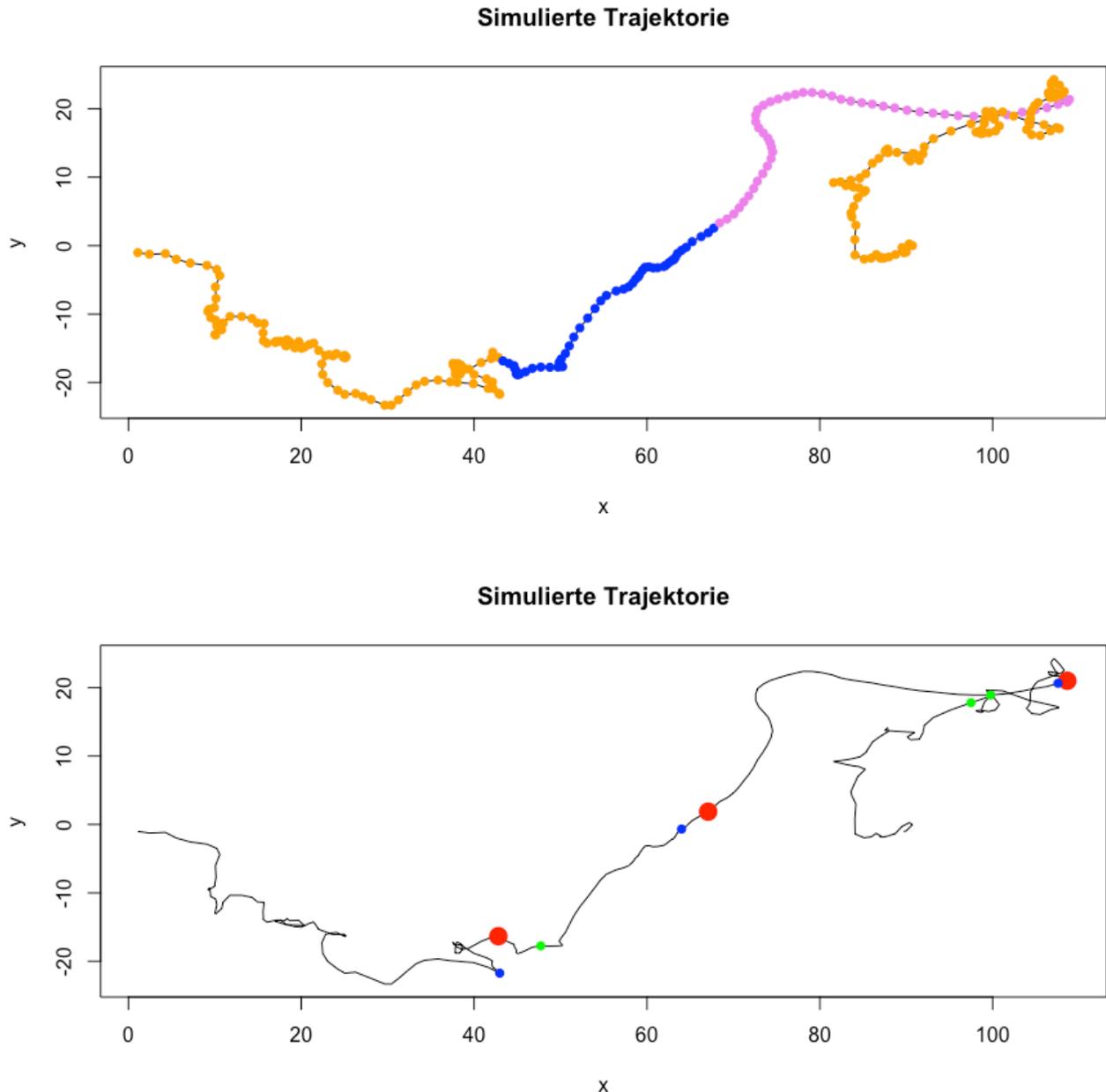


Abbildung 9: Trajektorie der time-scale switch Simulation (oben). Orange zeigt die erste und letzte Phase, blau die zweite und pink die dritte. Start ist bei Koordinate (0,0). Trajektorie (gleiche wie oben) mit eingezeichneten CP (unten). Definierte CP roter und grosser Punkt. Durch FPT identifizierte CP in grün und durch BCPA identifizierte CP in blau.

RESULTATE

In dieser Simulation hat die FPT die korrekte Anzahl CP gefunden, jedoch keinen CP am richtigen Ort identifiziert. Der erste und der letzte CP wurden zwar in der Nähe, aber nicht so nahe wie in der speed switch Simulation lokalisiert. Der dritte identifizierte CP liegt im letzten Segment (Abbildung 11).

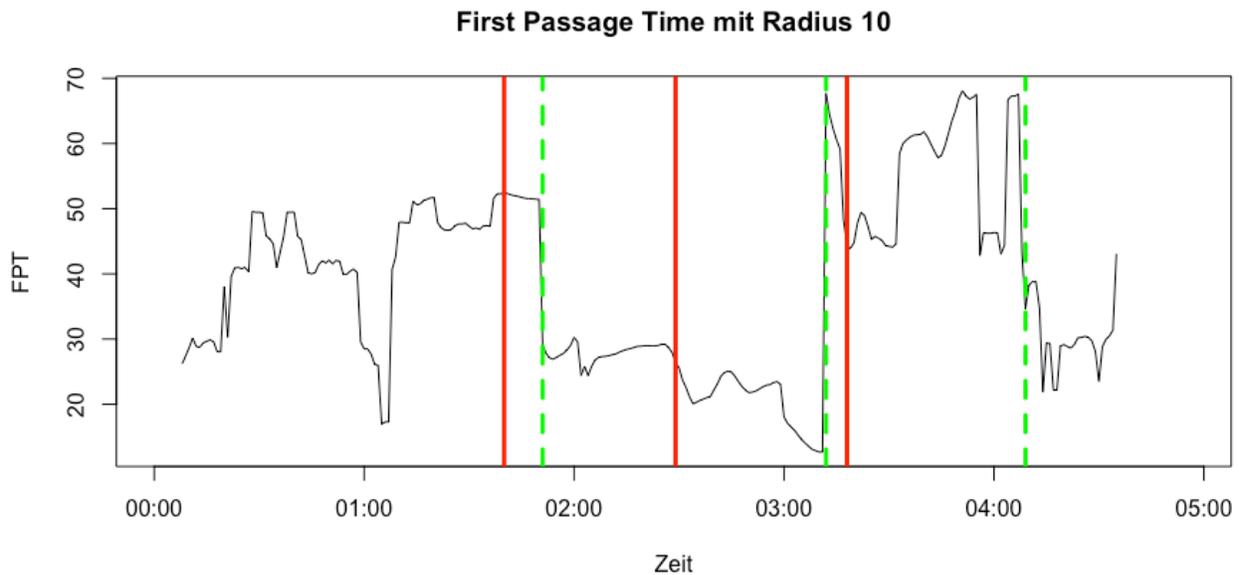


Abbildung 10: Grafik der FPT der time-scale switch Simulation. X-Achse ist die Zeit [min], y-Achse die FPT. Definierte CP sind die roten vertikalen Linien, durch die FPT identifizierte CP sind die grün gestrichelten, vertikalen Linien.

Die BCPA hat in dieser Simulation ebenfalls die korrekte Anzahl CP gefunden (Abbildung 12). Sie liegt mit dem letzten identifizierten CP nur einen Datenpunkt vom eigentlichen CP entfernt (vgl. Tabelle 7). Der zweite gefundene Punkt liegt nahe dem zweiten CP. Der erste identifizierte CP liegt relativ weit vom eigentlichen ersten CP entfernt. In dieser Simulation haben also beide Methoden zwar die richtige Anzahl CP erkannt, jedoch keinen CP tatsächlich korrekt identifiziert. In der Abbildung 12 (oben) erkennt man auch, dass das dritte Segment eine höhere Tortuosität (11) aufweist als die anderen Segmente (0).

RESULTATE

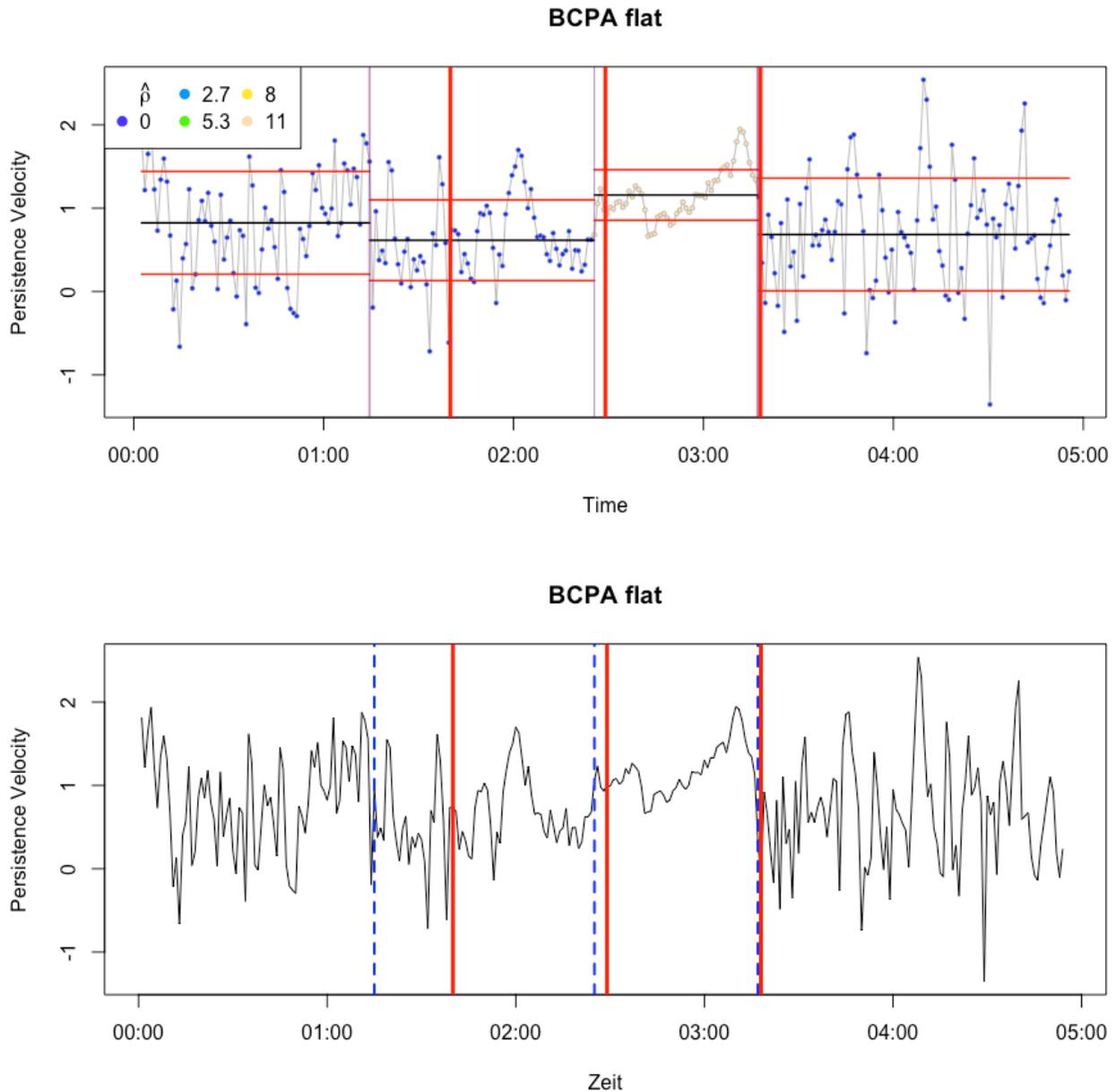


Abbildung 11: Grafik der BCPA der time-scale switch Simulation. X-Achse ist die Zeit [min], y-Achse die persistence velocity [map unit/min]. Oben repräsentieren die violetten vertikalen Linien die identifizierten CP, die horizontalen schwarze Linien der geschätzte Mittelwert, die horizontalen roten Linien das 95% Konfidenzintervall. Die farbigen Punkte repräsentieren die erwartete time-scale. Je kühler die Farbe desto tiefer die Werte und desto höher ist die Tortuosität (und umgekehrt). Unten repräsentieren die roten vertikalen Linien die korrekten CP und die blau gestrichelten die durch die BCPA identifizierten CP.

RESULTATE

Tabelle 7: Daten der time-scale switch Simulation und der durch die FPT und BCPA segmentierten Trajektorie der time-scale switch Simulation.

	Simulation	FPT	BCPA
# CP	3	3	3
# Segmente	4	4	4
Indizes CP	100, 149, 198	111, 192, 249	75, 145, 197
# Datenpkt. pro Segm.	99, 49, 49, 99	110, 81, 57, 48	74, 70, 52, 100
# korrekt ident. CP		0	0

In der Abbildung 13 (oben) lässt sich erkennen, dass der *turning angle* im dritten Segment kaum, im ersten und letzten Segment jedoch sehr stark variiert. Dies führt zu einer time-scale von 11 im dritten Segment (Abbildung 12). Vor allem der Übergang vom dritten ins vierte Segment lässt sich visuell eindeutig erkennen (Abbildung 13). Die Geschwindigkeit (unten) schwankt im dritten Segment ebenfalls nur wenig, im ersten und im letzten jedoch sehr stark.

RESULTATE

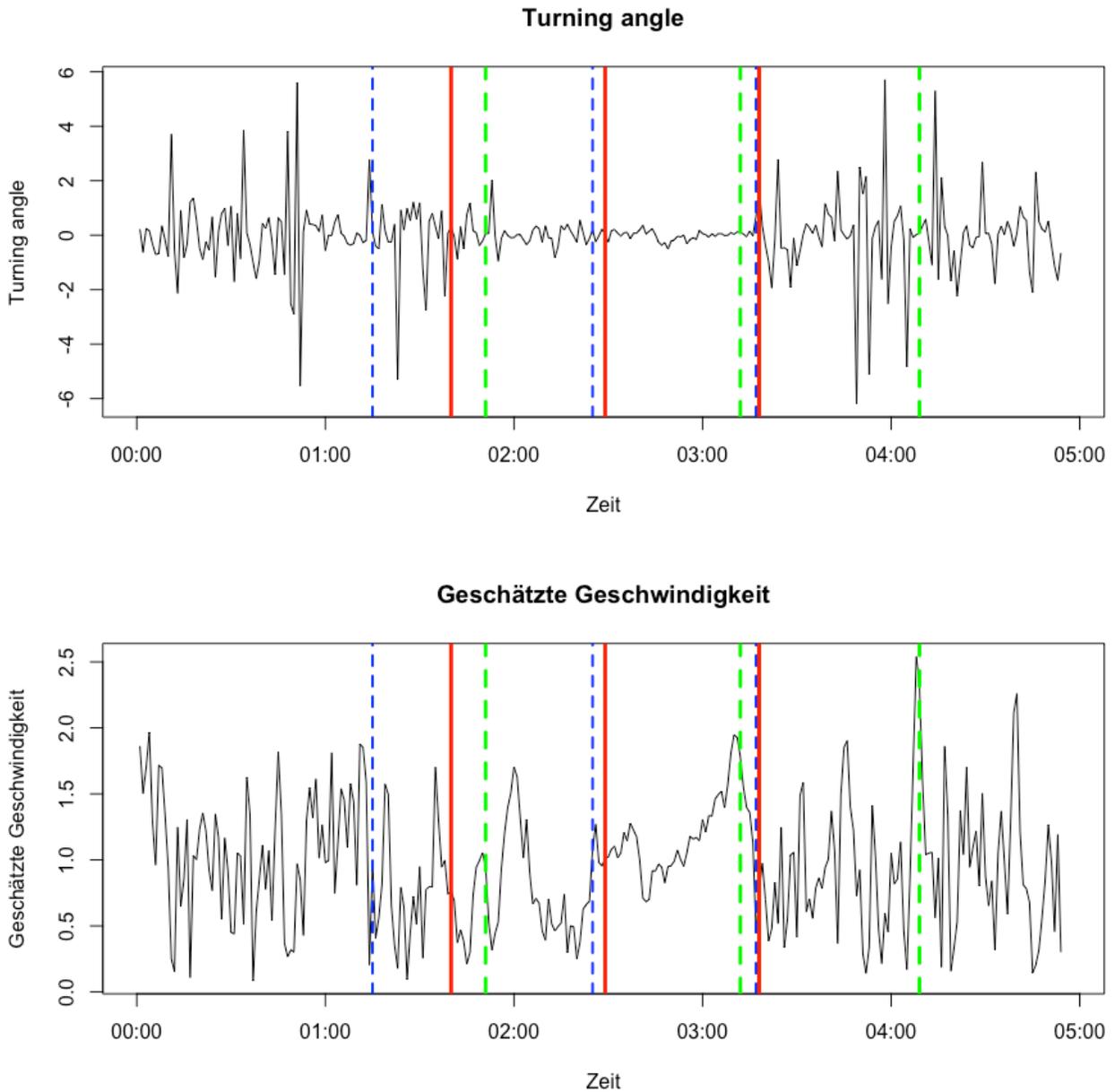


Abbildung 12: Turning angle (oben) und geschätzte Geschwindigkeit [map unit/min] (unten) für die time-scale switch Simulation. X-Achse ist die Zeit [min]. Rote vertikale Linien sind die korrekten CP, grün gestrichelte Linien sind die durch die FPT identifizierten CP und blau gestrichelt die durch die BCPA identifizierten CP.

RESULTATE

Continuous speed switch Simulation (Soleymani et al 2017)

In dieser Simulation scheinen alle Segmente eine sehr hohe Tortuosität zu haben (Abbildung 14). Erkennbar ist, dass die Länge der zurückgelegten Strecke mit jedem Segment etwas zunimmt. Die Lage der CP in Bezug zu den korrekten CP lässt sich aus dieser Abbildung kaum identifizieren.

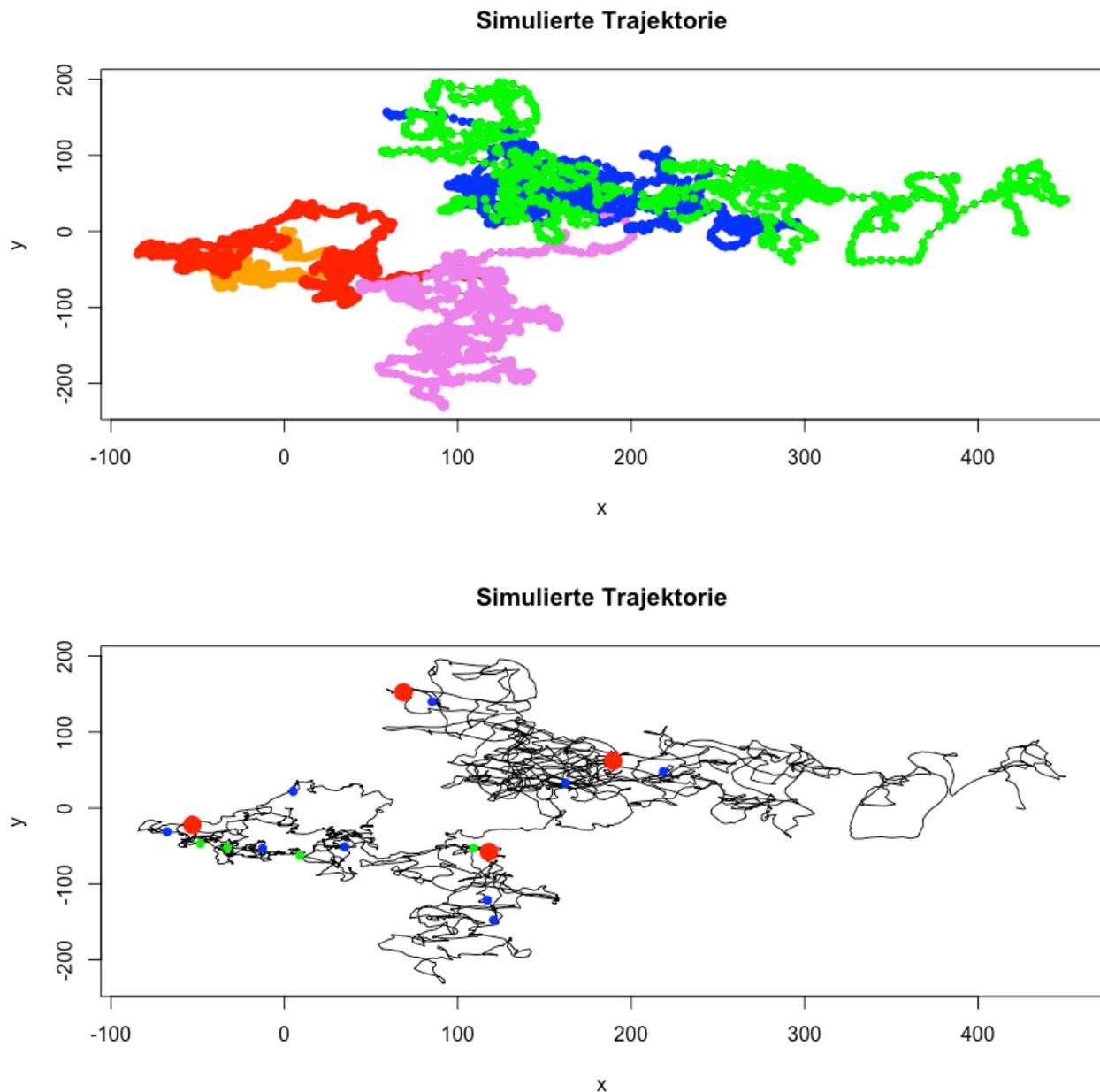


Abbildung 13 Trajektorie der continuous speed switch Simulation (oben). Orange zeigt die erste und letzte Phase, blau die zweite und pink die dritte Phase. Start ist bei Koordinate (0,0). CP derselben Trajektorie (unten). Definierte CP als rote grosse Punkte. Durch FPT identifizierte CP in grün und durch BCPA identifizierte CP in blau.

RESULTATE

Die FPT hat zwar die richtige Anzahl CP erkannt (4), jedoch keinen CP richtig identifiziert (Tabelle 8). Den zweiten CP hat sie jedoch nur um zwei Datenpunkte verfehlt. Auch in der Nähe des ersten CP hat sie einen CP identifiziert. Die identifizierten CP liegen alle innerhalb der ersten beiden Segmente, drei davon im ersten Segment. In diesem Segment schwanken die Werte am stärksten (Abbildung 15).

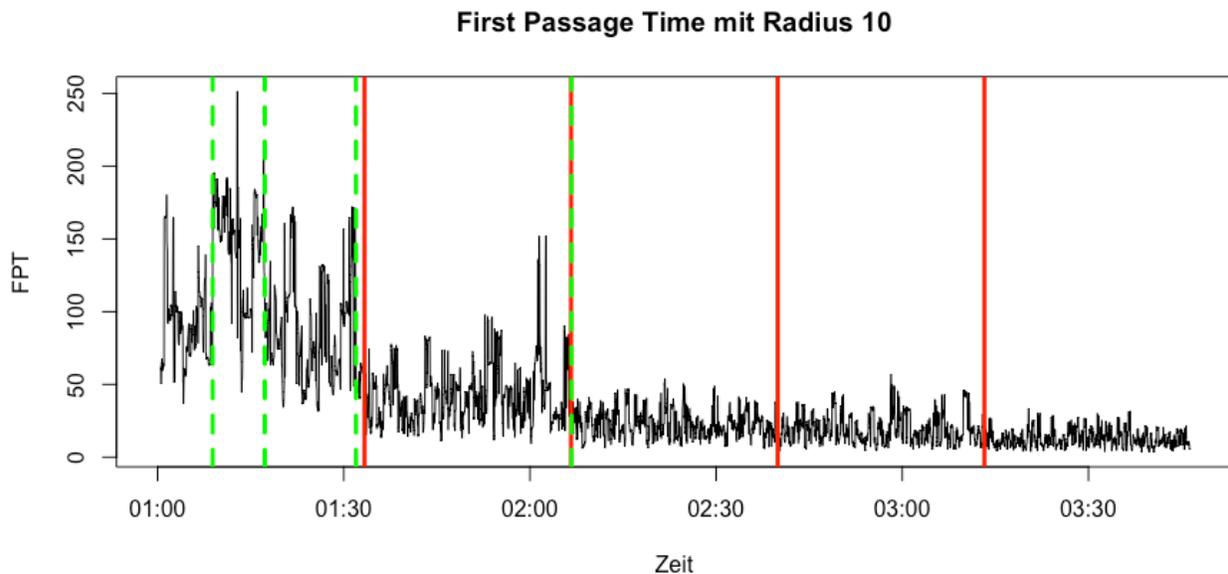


Abbildung 14: Grafik der FPT der continuous speed switch Simulation. X-Achse ist die Zeit [h], y-Achse die FPT. Definierte CP sind die roten vertikalen Linien, durch die FPT identifizierte CP sind die grün gestrichelten vertikalen Linien.

Die BCPA hat wie die FPT keinen CP richtig identifiziert (Tabelle 8). Neun Datenpunkte entfernt liegt der am nächsten an einen korrekten CP identifizierte CP. Dies ist der letzte CP. Aufgrund der visuellen Interpretation könnte man meinen, dass auch der erste CP relativ nahe scheint. Dieser liegt tatsächlich aber 48 Datenpunkte vom korrekten CP entfernt. Es liegen immer zwei CP in einem Segment, ausser im dritten, in dem nur ein identifizierter CP liegt. Die BCPA hat fast doppelt so viele CP (9) identifiziert wie korrekt wären. Die time-scale (Abbildung 16 oben) liegt immer in den höchsten beiden Klassen (0.032 und 0.042).

RESULTATE

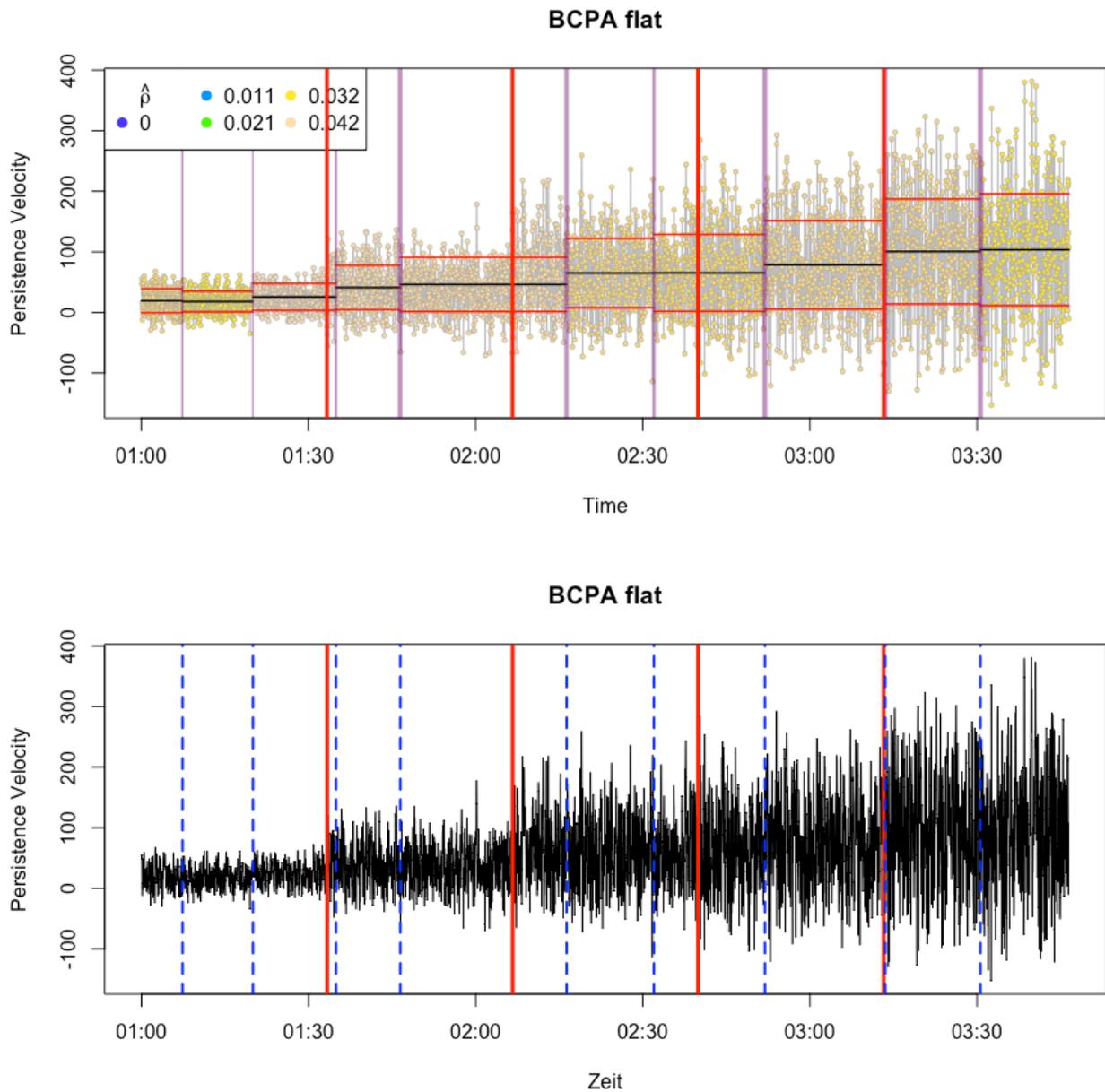


Abbildung 15: Grafik der BCPA der continuous speed switch Simulation. X-Achse ist die Zeit [h], y-Achse die persistence velocity [map unit/min]. Oben repräsentieren die violetten vertikalen Linien die identifizierten CP, die horizontalen schwarzen Linien der geschätzte Mittelwert, die horizontalen roten Linien das 95% Konfidenzintervall. Die farbigen Punkte repräsentieren die erwartete time-scale. Je kühler die Farbe desto tiefer die Werte und desto höher die Tortuosität (und umgekehrt). Unten repräsentieren die roten vertikalen Linien die korrekten CP und die blau gestrichelten die durch die BCPA identifizierten CP.

RESULTATE

Tabelle 8: Daten der continuous speed switch Simulation und der durch die FPT und BCPA segmentierten Trajektorie der continuous speed switch Simulation.

	Simulation	FPT	BCPA
# CP	4	4	9
# Segmente	5	5	10
Indizes CP	1000, 1999, 2998, 3997	266, 518, 959, 2001	221, 600, 1048, 1394, 2289, 2760, 3358, 4006, 4518
# Datenpkt. pro Segm.	999, 999, 999, 999, 999	265, 252, 441, 1042, 2995	220, 379, 448, 346, 895, 471, 598, 648, 512, 478
# korrekt ident. CP		0	0

In beiden Grafiken der Abbildung 17 lässt sich kaum etwas erkennen. Visuell lassen sich das erste vom zweiten, das zweite vom dritten und das vierte vom fünften Segment ungefähr anhand der Geschwindigkeit erkennen.

RESULTATE

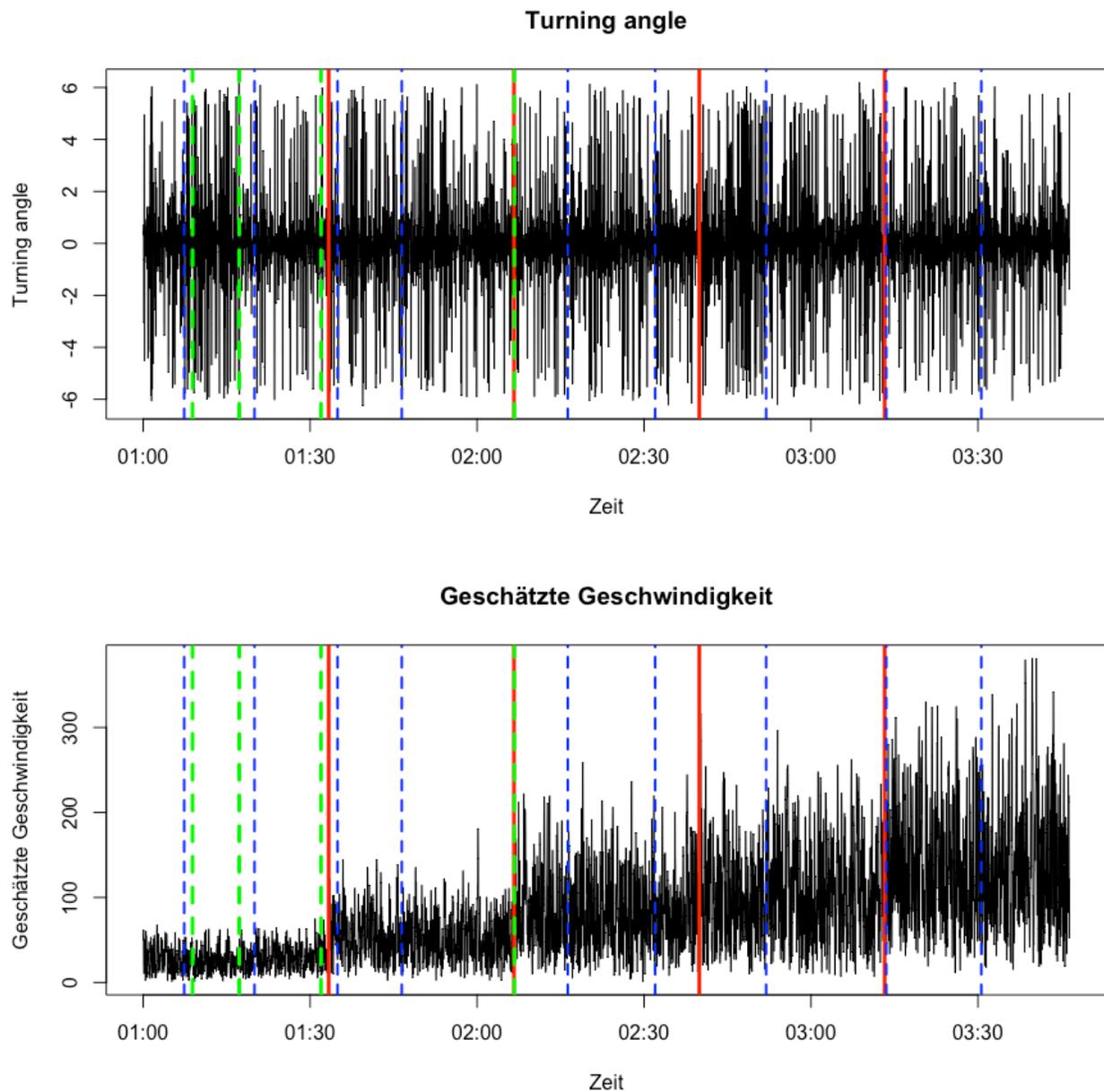


Abbildung 16: Turning angle (oben) und geschätzte Geschwindigkeit [map unit/min] (unten) für die continuous speed switch Simulation. X-Achse ist die Zeit [h]. Rote vertikale Linien sind die korrekten CP, grün gestrichelt die durch die FPT identifizierten CP und blau gestrichelt die durch die BCPA identifizierten CP.

Wie in Gurarie et al. (2016a) betont wird, ist diese eine Realisierung der jeweiligen Simulation nur ein einziges Beispiel. Aufgrund des Rauschens, welches der Simulation jedoch hinterlegt ist, sieht jede Simulation trotz identischen Werten der Parameter wieder anders aus. Um herauszufinden, ob die Methode gegenüber diesem Rauschen robust ist, werden im nächsten Schritt jeweils 50 Realisierungen jeder Simulation ausgeführt und beurteilt.

5.2 50 Realisierungen

Speed switch Simulation (Gurarie et al. 2016a)

Abbildung 17 zeigt die Anzahl Segmente und die Anzahl CP, sowohl in den ursprünglichen 50 Realisierungen wie auch in den Analysen von den beiden Methoden identifiziert. Logischerweise ist die Anzahl CP immer um 1 kleiner als die Anzahl Segmente.

Die Anzahl Segmente variiert bei der FPT zwischen 3 und 6 (2 Ausreisser). Der Median liegt bei 3 und ist somit um 1 tiefer als korrekt wäre (4 Segmente). Bei 75% der Realisierungen hat die FPT 3 oder 4 Segmente gefunden. Die BCPA erreicht einen Median von 7 für die Anzahl Segmente. Dieser ist fast doppelt so hoch als die eigentlich korrekte Anzahl Segmente ist. Die Anzahl Segmente variiert innerhalb der 50 Realisierungen jedoch zwischen 4 (1 Ausreisser) und 11 (5 Ausreisser). In 50% der Realisierungen hat die BCPA die Realisierung in 6 oder 7 Segmente geteilt.

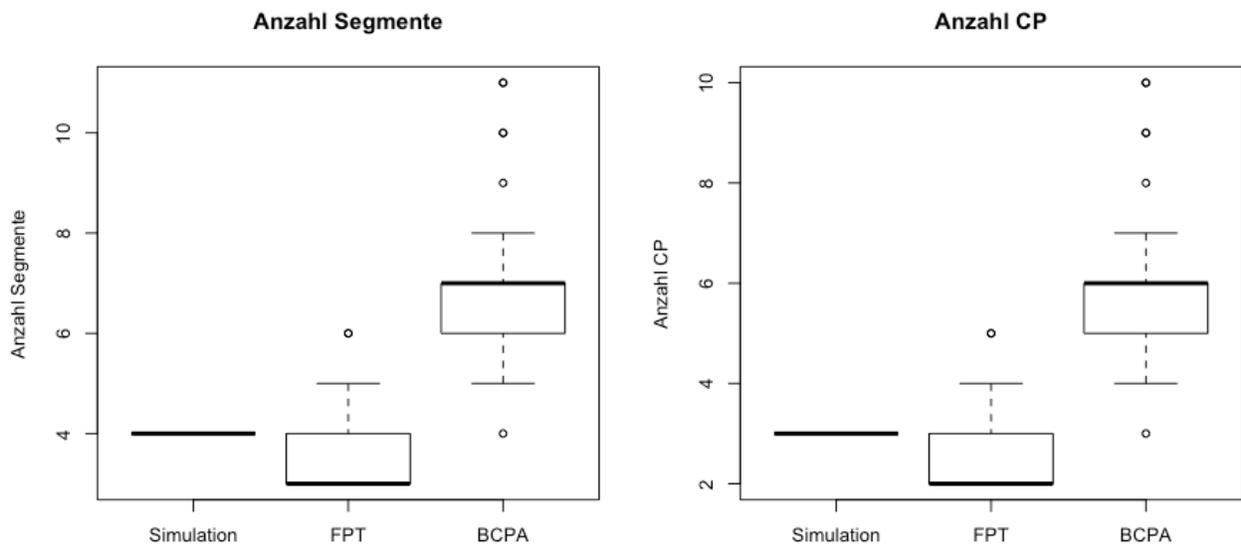


Abbildung 17: Boxplot der Anzahl Segmente und Anzahl CP für die speed switch Simulation jeweils für die Simulation, FPT und BCPA.

RESULTATE

Tabelle 9: Zusammenfassende Statistik der speed switch Simulation segmentiert durch die FPT (oben) und BCPA (unten).

FPT				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	2	0	0	0
1. Quartil	2	0	0	0
Median	2	0	0	0
Mittelwert	2.52	0.14	0.067	0.047
3. Quartil	3	0	0	0
Max	5	1	0.5	0.333

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	3	0	0	0
1. Quartil	5	0	0	0
Median	6	1	0.106	0.333
Mittelwert	5.7	0.6	0.111	0.2
3. Quartil	6	1	0.2	0.333
Max	10	2	0.4	0.667

Nur in 7 von 50 Fällen hat die FPT einen CP korrekt identifiziert (14%). Die BCPA hat für die Hälfte (26) der Realisierungen mindestens einen CP korrekt identifiziert (52%). In 4 Realisierungen hat sie 2 CP korrekt identifiziert (8%). Der Median der Precision und des Recalls liegt sowohl für die FPT wie auch für die BCPA bei 0.

Time-scale switch Simulation (Gurarie et al. 2016a)

In dieser Simulation hat die FPT für die Hälfte der Realisierungen mindestens 4 Segmente gefunden, für die andere Hälfte der Realisierungen höchstens gleich viele. Die Werte variieren zwischen 2 und 7 Segmenten. Der Median für die 50 Realisierungen segmentiert durch die BCPA liegt bei 4.5, und ist damit nur wenig höher als derjenige der FPT. Die Anzahl Segmente variiert hier zwischen 2 und 9 Segmenten. Bei 50% der Realisierungen werden jedoch zwischen 4 und 6 Segmenten erkannt.

RESULTATE

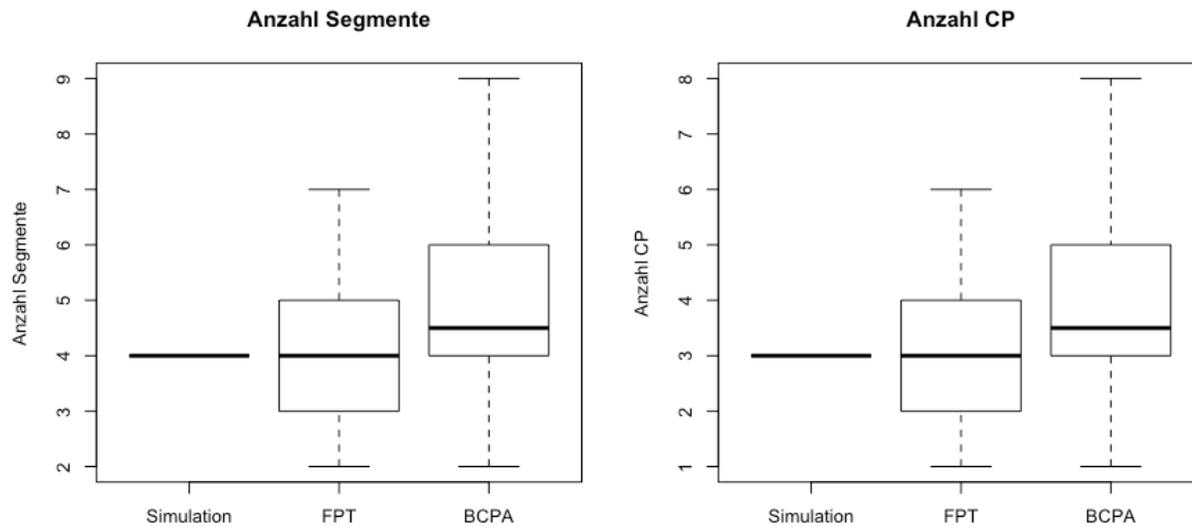


Abbildung 18: Boxplot der Anzahl Segmente und Anzahl CP für die time-scale switch Simulation jeweils für die Simulation, FPT und BCPA.

Tabelle 10: Zusammenfassende Statistik der time-scale switch Simulation segmentiert durch die FPT (oben) und BCPA (unten).

FPT				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	1	0	0	0
1. Quartil	2	0	0	0
Median	3	0	0	0
Mittelwert	3.24	0.04	0.013	0.013
3. Quartil	4	0	0	0
Max	6	1	0.333	0.333

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	1	0	0	0
1. Quartil	3	0	0	0
Median	3.5	0	0	0
Mittelwert	3.8	0.34	0.087	0.113
3. Quartil	5	1	0.192	0.333
Max	8	2	0.667	0.667

RESULTATE

Da der Median der Anzahl korrekt identifizierter CP durch die FPT bereits bei 0 liegt, liegt er auch bei der Precision und beim Recall bei 0. Bei nur gerade 4% der Realisierungen wird 1 CP richtig identifiziert. Für die BCPA liegt der Median beider Kriterien ebenfalls bei 0. Bei 32% der Realisierungen wird mindestens 1 CP korrekt identifiziert.

Continuous speed switch Simulation (Soleymani et al. 2017)

In dieser Simulation liegt die Anzahl Segmente bei 5 (Abbildung 19 links). Die FPT hat, ausser mit wenigen Ausreissern (4), zwischen 2 und 4 Segmente gefunden. 75% der Realisierungen hat sie in 2 oder 3 Segmente eingeteilt. Der Median der BCPA liegt bei 12 für die Anzahl Segmente und variiert für alle Realisierungen zwischen 9 und 15 (Oberer Whisker).

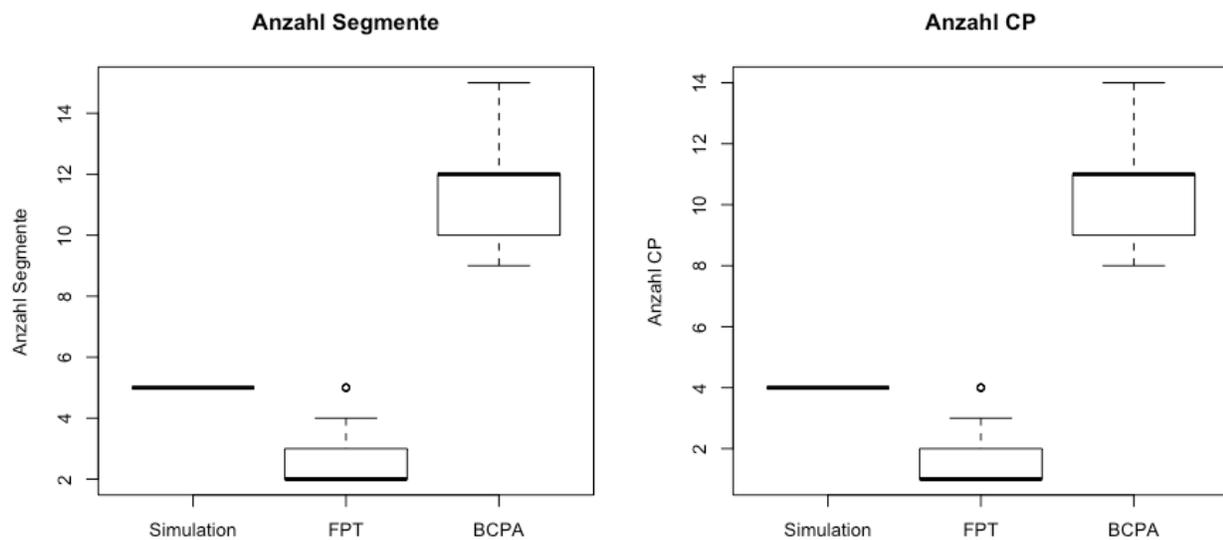


Abbildung 19: Boxplot der Anzahl Segmente und Anzahl CP für die continuous speed switch Simulation jeweils für die Simulation, FPT und BCPA.

RESULTATE

Tabelle 11: Zusammenfassende Statistik der continuous speed switch Simulation segmentiert durch die FPT (oben) und BCPA (unten).

FPT				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	1	0	0	0
1. Quartil	1	0	0	0
Median	1	0	0	0
Mittelwert	1.64	0.02	0.2	0.005
3. Quartil	2	0	0	0
Max	4	1	1	0.25

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	8	0	0	0
1. Quartil	9.25	0	0	0
Median	11	0	0	0
Mittelwert	10.72	0	0	0
3. Quartil	11	0	0	0
Max	14	0	0	0

Die FPT hat in einer Realisierung 1 CP korrekt identifiziert. Der Median der Precision und des Recalls liegen dennoch bei 0. Die BCPA hat keinen CP korrekt identifiziert. Damit liegen sowohl die Precision wie auch der Recall bei 0.

5.3 Maximale Abweichung

Speed switch Simulation (Gurarie et al. 2016a)

In der Abbildung 20 lässt sich erkennen, dass bereits die Berücksichtigung einer Abweichung von ± 1 bei der FPT den Median der Precision sowie des Recalls auf 0.333 ansteigen lässt. Mit der Berücksichtigung einer Abweichung von ± 2 und ± 3 steigt der Median bei der Precision auf 0.5 an. Das heisst, mindestens 50% der identifizierten CP werden mit einer maximalen Abweichung von ± 3 korrekt identifiziert. Der Median des Recalls bleibt bei 0.333 ab einer Abweichung von ± 1 . Mit einer Abweichung von ± 2 sind es jedoch 75% auf 0.333 oder mehr. Dies bedeutet, dass 33.3% der korrekten CP bereits mit der Berücksichtigung einer Abweichung von ± 1 als korrekt identifiziert werden.

RESULTATE

Bei der BCPA steigt der Median bei einer Abweichung von ± 1 auf 0.333 an. Mit der Berücksichtigung einer Abweichung von ± 1 wird also bereits bei mindestens 50% der Realisierungen ein CP korrekt identifiziert. Wird eine maximale Abweichung von ± 2 und ± 3 berücksichtigt, liegt der Median bei 0.5. Wird eine Abweichung von ± 3 berücksichtigt, beträgt die Spannweite aller Werte nur ungefähr 0.4, während sie bei der FPT von 0 bis 1 reicht. Auch der Median des Recalls steigt an, bis er bei einer maximalen Abweichung von ± 2 auf 1 liegt. 76% der Realisierungen haben drei CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert (12 Ausreisser).

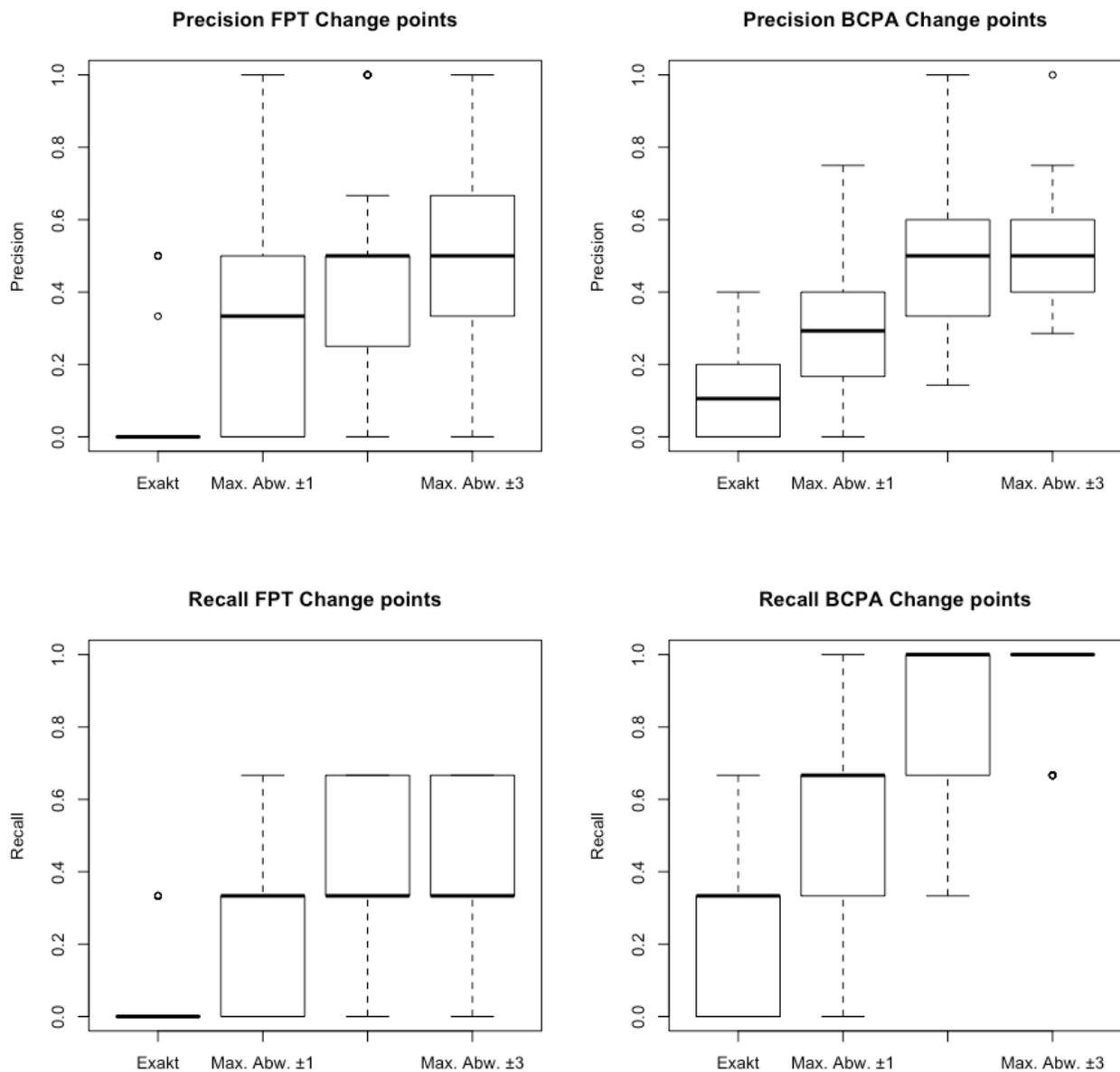


Abbildung 20: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.

RESULTATE

Klar ersichtlich aus der Tabelle 12 ist, dass die FPT bei keiner Realisierung den mittleren CP erkannt hat. Der erste und dritte CP werden jeweils bei 30 Realisierungen korrekt identifiziert.

Die BCPA hat bei allen Realisierungen jeweils den ersten und den dritten CP erkannt. Auch der mittlere CP wird bei 38 Realisierungen korrekt identifiziert. Im Gegensatz zur FPT hat die BCPA aber bei mehr Realisierungen zusätzlich zwischen den korrekt identifizierten CP einen oder mehrere CP identifiziert.

Tabelle 12: Anzahl Realisierungen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Simulationen, die zwischen den korrekten CP weitere CP identifiziert haben.

	< I	I	I < II	II	II < III	III	III <
FPT	24	32	5	0	3	30	24
BCPA	23	50	29	38	33	50	23

Time-scale switch Simulation (Gurarie et al. 2016a)

Die Precision in der Abbildung 21 zeigt, dass die FPT in der time-scale switch Simulation, trotz Einbezug der maximalen Abweichungen, keinen CP korrekt identifiziert hat (mit wenigen Ausnahmen der Ausreisser). Aus diesem Grund ist auch der Median des Recalls gleich 0.

Die BCPA hat besser abgeschnitten. Sie erreicht bereits mit der Berücksichtigung einer Abweichung von ± 1 einen Median von 0.333. Mit einer Abweichung von ± 2 erreicht sie einen Median von 0.5 und mit einer maximalen Abweichung von ± 3 einen Median von 0.6. 60% aller identifizierten CP von 50% der Realisierungen sind damit korrekt identifiziert. Die Streuung ist jedoch relativ gross. Der Recall steigt mit Berücksichtigung einer maximalen Abweichung von ± 1 auf 0.333 und mit ± 2 auf 0.667. Bei der Berücksichtigung einer maximalen Abweichung von ± 3 hat die BCPA in 75% der Realisierungen 66.7% oder mehr der korrekten CP korrekt identifiziert.

Die Werte in der Tabelle 13 bestätigen die Precision und den Recall von 0 in Abbildung 21. Es wird ersichtlich, dass die FPT häufiger CP zwischen den korrekten CP identifiziert hat als tatsächlich korrekte CP. Bei der BCPA ist es ziemlich ausgeglichen zwischen den korrekt identifizierten und den dazwischen liegenden CP. Grundsätzlich werden jedoch mehr korrekte CP identifiziert als dazwischen liegende. Der dritte CP wird am häufigsten korrekt identifiziert während es bei der FPT der erste CP ist.

RESULTATE

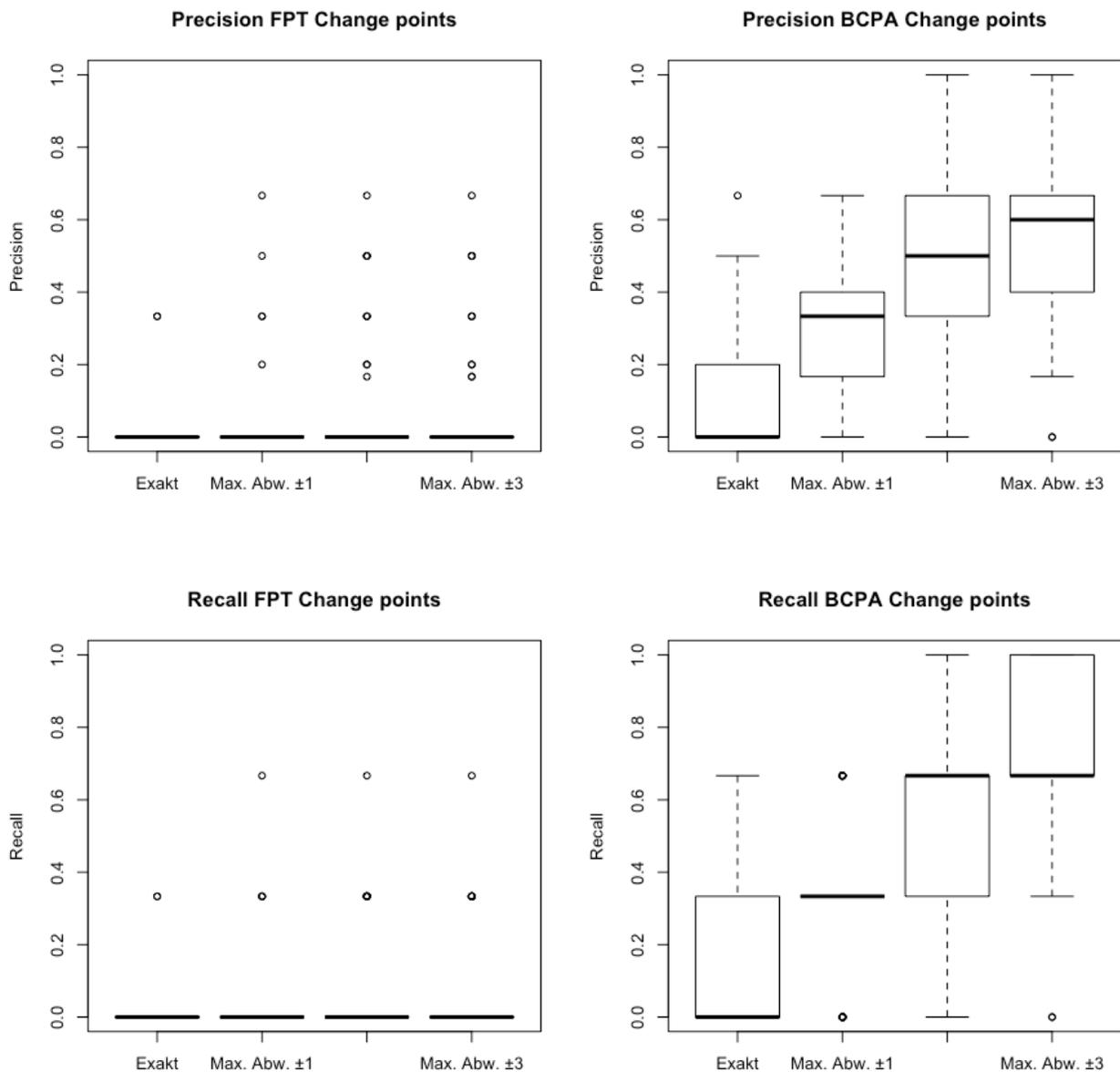


Abbildung 21: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.

Tabelle 13: Anzahl Simulationen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Simulationen, die zwischen den korrekten CP weitere CP identifiziert haben.

	< I	I	I < II	II	II < III	III	III <
FPT	24	8	33	1	32	4	30
BCPA	20	38	16	22	22	43	14

RESULTATE

Continuous speed switch Simulation (Soleymani et al. 2017)

In dieser Simulation liefern weder die FPT noch die BCPA brauchbare Resultate. Ausser wenigen Ausreissern (9 bei der FPT und 1 bei der BCPA bei Abweichung ± 3) werden nahezu in 100% der Realisierungen keine der identifizierten CP korrekt identifiziert und daher werden auch mehrheitlich keine korrekten CP als korrekt identifiziert.

In der Tabelle 14 fällt auf, dass die FPT ab dem dritten CP bei keiner Realisierung einen CP findet. Die BCPA hingegen identifiziert (ausser mit einer Ausnahme des zweiten CP) keinen CP korrekt. Jedoch findet sie bei sämtlichen Realisierungen immer zwischen den korrekten CP mindestens 1 CP.

Tabelle 14: Anzahl Realisierungen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Realisierungen, die zwischen den korrekten CP weitere CP identifiziert haben.

	< I	I	I < II	II	II < III	III	III < IV	IV	IV <
FPT	31	6	19	3	4	0	0	0	0
BCPA	50	0	50	1	50	0	50	0	50

RESULTATE

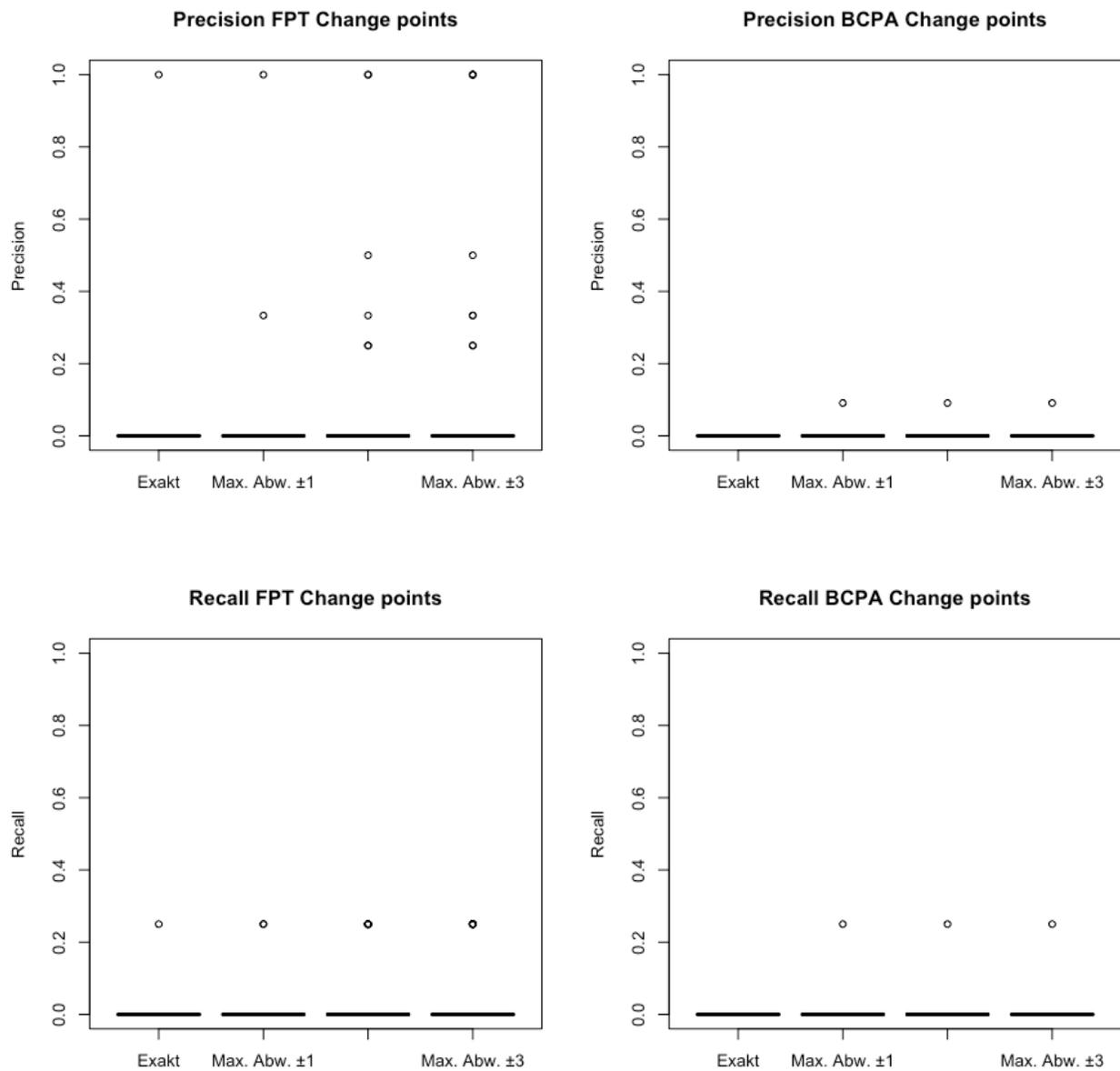


Abbildung 22: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.

5.4 FPT: mittlerer CP

Speed switch: Änderung 1

In dieser Simulation wechselt die Geschwindigkeit von 5 auf 10 zurück auf 5 und wieder auf 10. Die Segmentlänge beträgt jeweils 50 Datenpunkte.

Tabelle 15: speed switch Änderung 1: Anzahl Realisierungen bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert wurden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert wurden. Radius der FPT gleich 10.

< I	I	I < II	II	II < III	III	III <
34	14	3	20	39	15	5

Wie in der Tabelle 16 ersichtlich ist, werden 28% bis 40% der CP jeweils korrekt identifiziert. Im Vergleich zu der normalen speed switch Simulation (0%) ist dies eine starke Verbesserung.

Wird diese Simulation mit unterschiedlichen Radien segmentiert (Abbildung 23) wird erkennbar, dass ohne Einbezug der maximalen Abweichung weniger als 50% der CP korrekt identifiziert werden. Mit Einbezug der maximalen Abweichung von ± 3 werden für alle Radien (ausser 1 und 50) über 50% der CP korrekt identifiziert. Bei einem Radius von 4 liegen der Mittelwert und der Median bei 1. Ansonsten liegt der Mittelwert immer tiefer.

RESULTATE

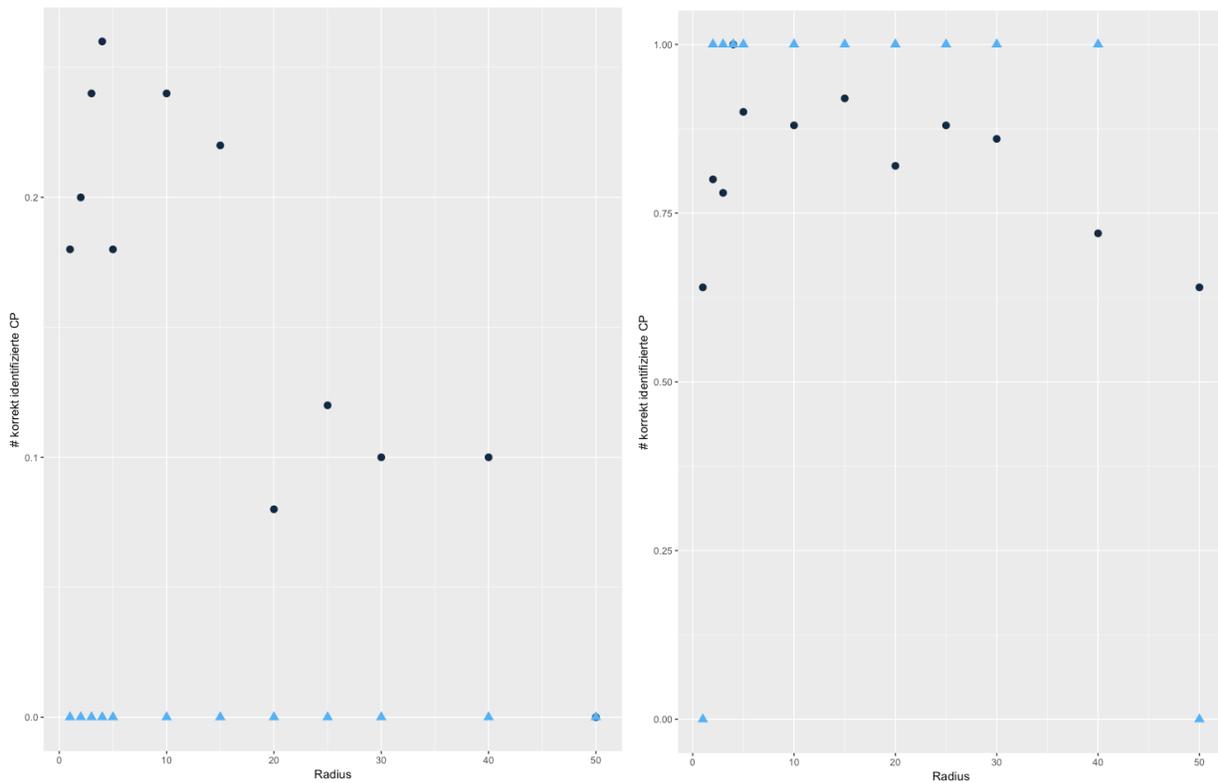


Abbildung 23: speed switch Änderung 1: Anzahl korrekte CP basierend auf unterschiedlichen Radien der FPT. Links: exakt identifizierte CP, rechts: mit maximaler Abweichung von ± 3 identifizierte CP.

Mit einem Radius von 4 wird die Simulation erneut segmentiert. Ausser dem ersten CP bleiben alle Werte gleich. Der erste CP erreicht 15 statt 14 CP.

Tabelle 16: speed switch change 1: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. Radius der FPT gleich 10.

< I	I	I < II	II	II < III	III	III <
34	15	3	20	39	15	5

Speed switch: Änderung 2

In dieser Simulation wechselt die Geschwindigkeit von 5 auf 10, die Segmentlänge beträgt hier neu 100 Datenpunkte.

RESULTATE

Tabelle 17: speed switch Änderung 2: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert wurden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert wurden. Radius der FPT gleich 10.

$< I$	I	$I <$
46	23	11

Die Tabelle 17 zeigt, dass in der speed switch Änderung 2 Simulation bei 46% der Realisierungen der CP korrekt gefunden wird. Im ersten Segment wird jedoch oft mindestens ein CP identifiziert.

Durch den Vergleich der unterschiedlichen Radien (Abbildung 24) wird ersichtlich, dass auch in dieser Simulation ohne Berücksichtigung einer Abweichung weniger als 50% der CP korrekt identifiziert werden. Mit Berücksichtigung einer Abweichung von ± 3 , wird nur mit dem Radius 5 bei der Hälfte der Realisierungen ein CP korrekt identifiziert.

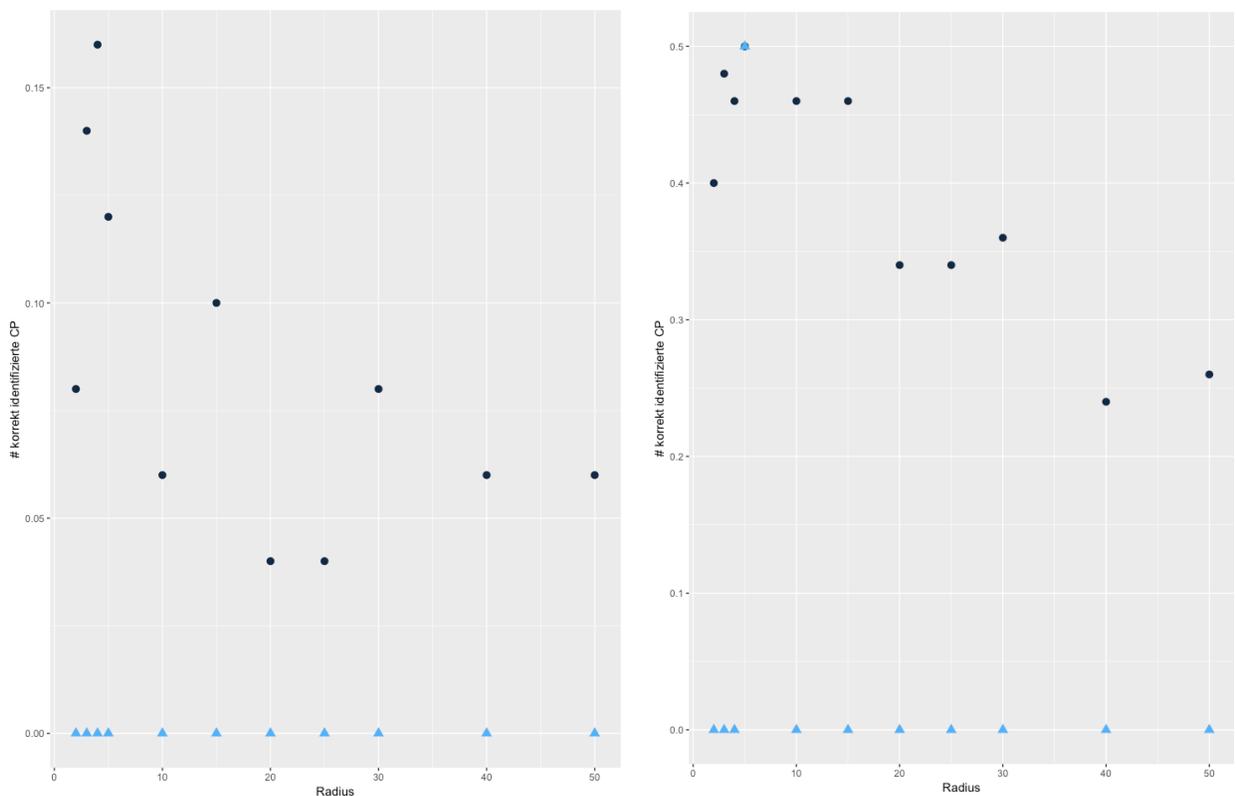


Abbildung 24: speed switch Änderung 2: Anzahl korrekte CP basierend auf unterschiedlichen Radien der FPT. Links: exakt identifizierte CP, rechts: mit maximaler Abweichung von ± 3 identifizierte CP.

RESULTATE

Mit Radius 5 wird erneut segmentiert. Damit erhöht sich die Anzahl korrekt gefundener CP auf 25, was 50% der Realisierungen entspricht. Allgemein haben sich die Werte aber auch hier nicht stark verändert.

Tabelle 18: speed switch Änderung 2: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert wurden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert wurden. Radius der FPT gleich 5.

< I	I	I <
47	25	10

Speed switch: Änderung 3

Die dritte Runde kombiniert die ersten beiden Änderungen. Die Geschwindigkeit wechselt von 5 auf 10 zurück auf 5 und wieder auf 10. Die Segmentlänge beträgt jetzt aber 100 Datenpunkte.

Tabelle 19: speed switch change 3: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. Radius der FPT gleich 10.

< I	I	I < II	II	II < III	III	III <
32	16	8	16	35	18	0

Bei 32% bis 36% der Realisierungen werden die CP korrekt identifiziert. Im ersten und dritten Segment werden jedoch viele zusätzliche CP identifiziert (64% und 70%). Im letzten Segment werden keine CP erkannt (Tabelle 19).

Die Abbildung 25 zeigt, dass ohne Berücksichtigung einer maximalen Abweichung weniger als die Hälfte der CP korrekt identifiziert werden. Mit Berücksichtigung einer maximalen Abweichung von ± 3 wird bei über 50% der Realisierungen ein CP korrekt identifiziert, ausser mit Radius 40 und 50. Mit Radius 4 und 5 liegt der Mittelwert etwas über dem Median.

Die Segmentierung wird mit einem Radius von 4 erneut durchgeführt.

RESULTATE

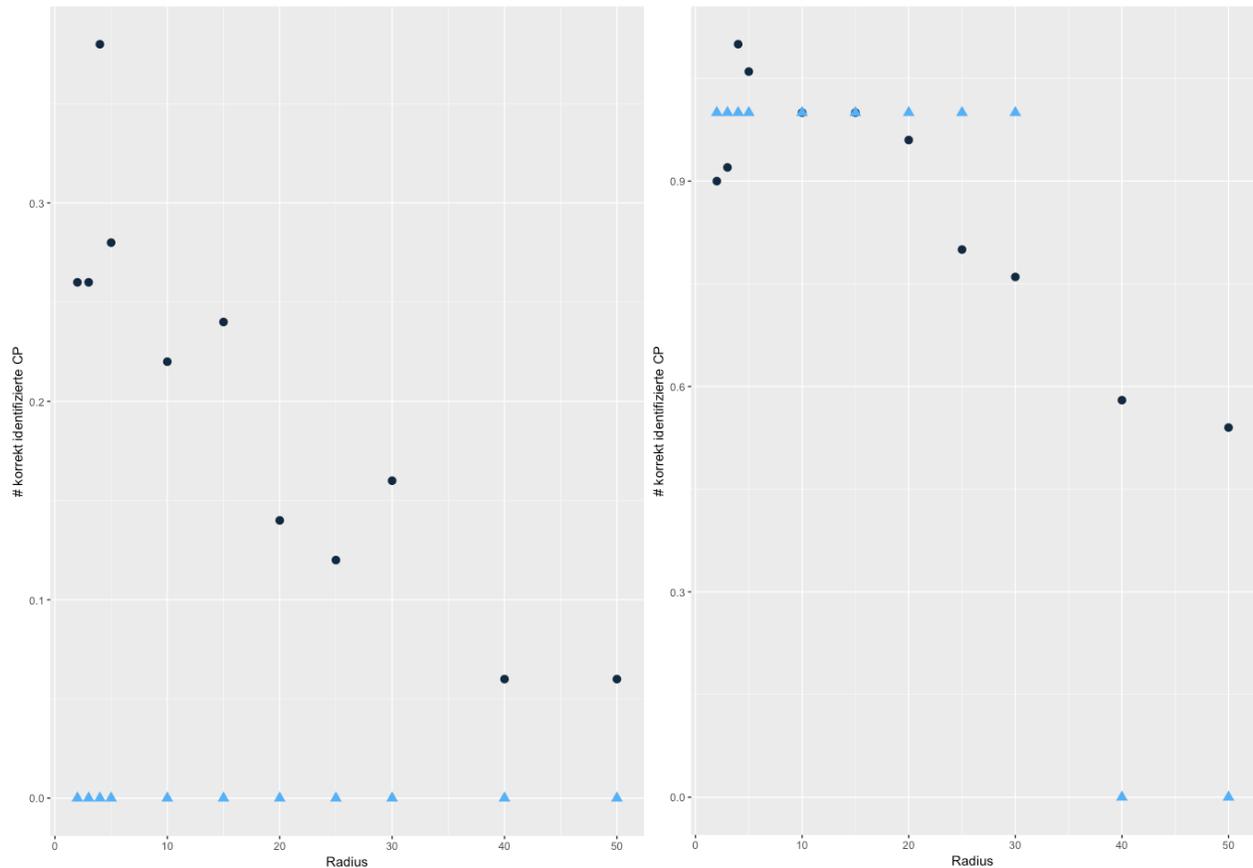


Abbildung 25: speed switch Änderung 3: Anzahl korrekte CP basierend auf unterschiedlichen Radien der FPT. Links: exakt identifizierte CP, rechts: mit maximaler Abweichung von ± 3 identifizierte CP.

Damit erhöht sich die Anzahl jedes CP um ein bis 2 weitere korrekt identifizierte CP. Der kleinere Radius hat somit einen leicht positiven Effekt auf die Anzahl der korrekt identifizierten CP.

Originale speed switch mit Radius 4

In dieser Simulation mit Radius 4 erhöht sich der Median der Anzahl identifizierten CP um 1 (von 2 auf 3). Die weiteren Mittelwerte (Anzahl korrekt identifizierte CP, Precision und Recall) steigen etwas an, auf den Median hat dies jedoch keinen Einfluss. Dieser bleibt sich gleich. Auch die Maximum-Werte bleiben dieselben. Das 3. Quartil erhöht sich für alle Elemente. Die Anzahl CP steigt von 3 auf 4, die Anzahl korrekt identifizierter CP von 0 auf 1, die Precision und der Recall je von 0 auf 0.333. Durch die Berücksichtigung einer maximalen Abweichung von ± 3 erhöht sich der Median der Anzahl korrekt identifizierter CP von 0 auf 1, der Median der Precision von 0 auf 0.5 und der Median des Recalls von 0 auf 0.333.

RESULTATE

Tabelle 20: Zusammenfassende Statistik der speed switch Simulation mit Radius 4 segmentiert durch die FPT: exakt (oben) und mit maximaler Abweichung von ± 3 (unten).

FPT				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	1	0	0	0
1. Quartil	2	0	0	0
Median	3	0	0	0
Mittelwert	2.94	0.36	0.145	0.12
3. Quartil	4	1	0.333	0.333
Max	5	1	0.5	0.333

FPT				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	1	0	0	0
1. Quartil	2	1	0.25	0.333
Median	3	1	0.5	0.333
Mittelwert	2.94	1.32	0.488	0.44
3. Quartil	4	2	0.667	0.667
Max	5	2	1	0.667

Vergleicht man die Boxplots zu Precision und Recall der ursprünglichen Simulation und derjenigen mit Radius 4, zeigt sich, dass mit Radius 4 mehr CP exakt identifiziert werden als mit Radius 10. Der Median liegt dennoch immer bei 0. Werden die unterschiedlichen Abweichungen berücksichtigt, ist die Streuung bei beiden Radien gross, der Median liegt beim Radius 4 eher tiefer. Mit der Berücksichtigung einer Abweichung von ± 3 liegt der Median sowohl bei der Precision als auch beim Recall bei 0.5, die Box des Radius 4 ist jedoch etwas breiter und damit etwas schlechter als bei Radius 10.

Auch beim Recall ist der Radius 4 bei den exakt identifizierten CP etwas besser, wie auch wenn eine Abweichung von ± 1 berücksichtigt wird. Der Median liegt aber bei allen Abweichungen von beiden Radien auf 0.333. Wird eine maximale Abweichung von ± 3 berücksichtigt, ist kein Unterschied mehr ersichtlich.

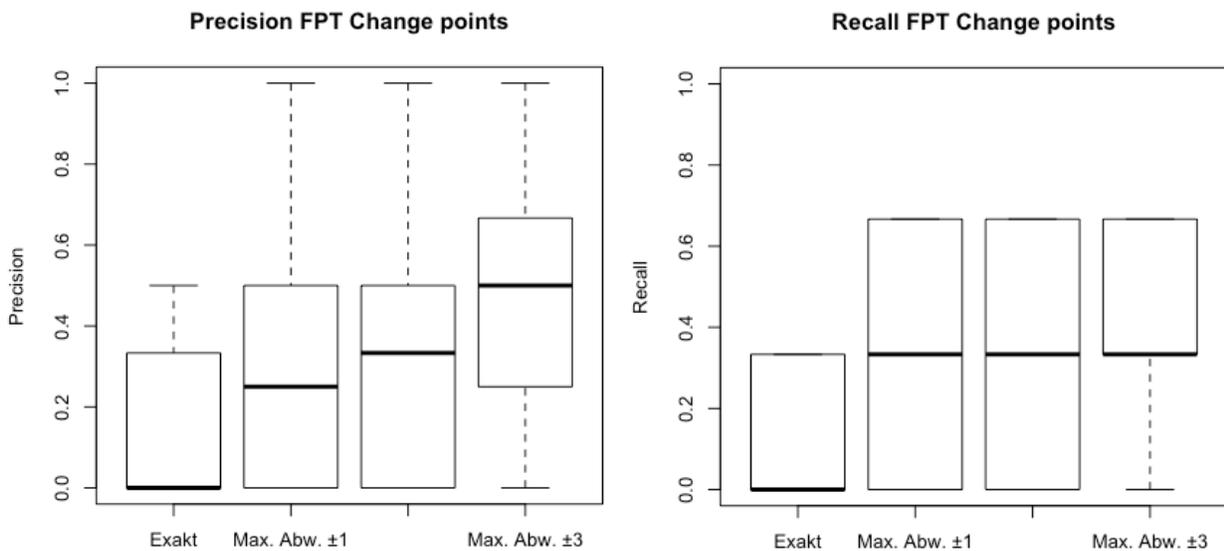


Abbildung 26: Boxplot zu Precision (links) und Recall (rechts) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 segmentiert durch die FPT.

Tabelle 21: speed switch: Anzahl Simulationen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Simulationen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. Radius der FPT beträgt 4.

< I	I	I < II	II	II < III	III	III <
29	32	0	0	0	34	25

Die Änderung des Radius von 10 auf 4 hat auf den mittleren CP keinen Einfluss. Zwischen dem zweiten und dritten CP sowie zwischen dem dritten und vierten CP werden keine CP mehr identifiziert, dafür sind es im ersten und im letzten Segment etwas mehr.

5.5 BCPA: Wertevergleich der Parameter

Speed switch Simulation

Der Wertevergleich der speed switch Simulation zeigt, dass der höchste Wert bei exakt identifizierten CP ein korrekt identifizierter CP ist. Aus Abbildung 27 lässt sich erkennen, dass eine *window size* von 10 eher nicht zu empfehlen ist, da meist keine CP gefunden werden. Mit *window size* 20 werden bereits mehr CP korrekt identifiziert, aber noch immer weniger als mit *window size* 30, 40 und 50. Die *window size* 40 scheint ziemlich robust zu sein bezüglich den anderen beiden Parametern. Der Sensitivitätsparameter K in Kombination mit tiefer *cluster width* (2-5) liefert allgemein gute Resultate. Ansonsten scheint der Einfluss von K eher klein zu sein. Die *cluster width* scheint einen Einfluss zu haben. Ausser bei *window size* 40 führt eine niedrige *cluster width* (zwischen 2 und 6) tendenziell zu besseren Resultaten. Die einzige

RESULTATE

Kombination, bei welcher der Median und der Mittelwert gleich sind, ist mit $K = 2$, $window\ size = 30$ und $cluster\ width = 3$.

Mit einer maximalen Abweichung von ± 3 zeigt die Abbildung 28 ein stärkeres Bild. Es werden mit vielen Wertekombinationen drei korrekte CP identifiziert. Ausnahmen bilden ebenfalls die $window\ size$ 10 und die höheren Werte der $cluster\ width$ (6-8). Eine $cluster\ width$ von 2 ist in diesem Fall (unter Berücksichtigung der maximalen Abweichung von ± 3) mit Vorsicht zu behandeln. Es kann vorkommen, dass die BCPA direkt aufeinanderfolgende Datenpunkte als CP erkennt. Dies geschieht, weil die Abweichung mehr Datenpunkte umfasst als die $cluster\ width$. Dadurch kann ein CP doppelt gezählt werden, wodurch ein falsches Bild entsteht. Aus diesem Grund wird die $cluster\ width$ von 2 in diesem Projekt nicht weiter berücksichtigt.

RESULTATE

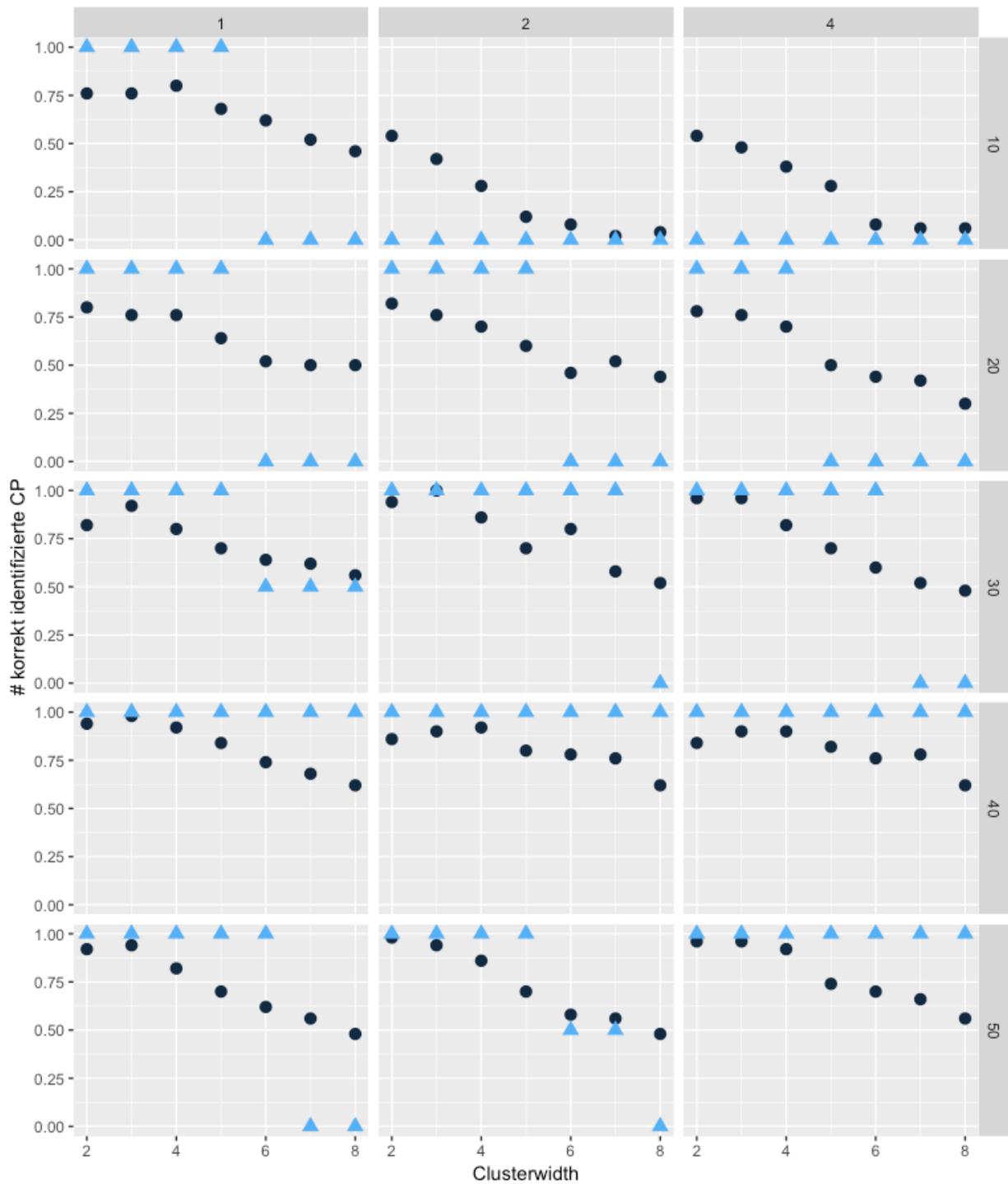


Abbildung 27: Wertevergleich der Parameter der BCPA für die speed switch Simulation. X-Achse oben: Sensitivitätsparameter K , X-Achse unten: Clusterwidth, Y-Achse links: Anzahl exakt korrekt identifizierte CP, y-Achse rechts: window size.

RESULTATE

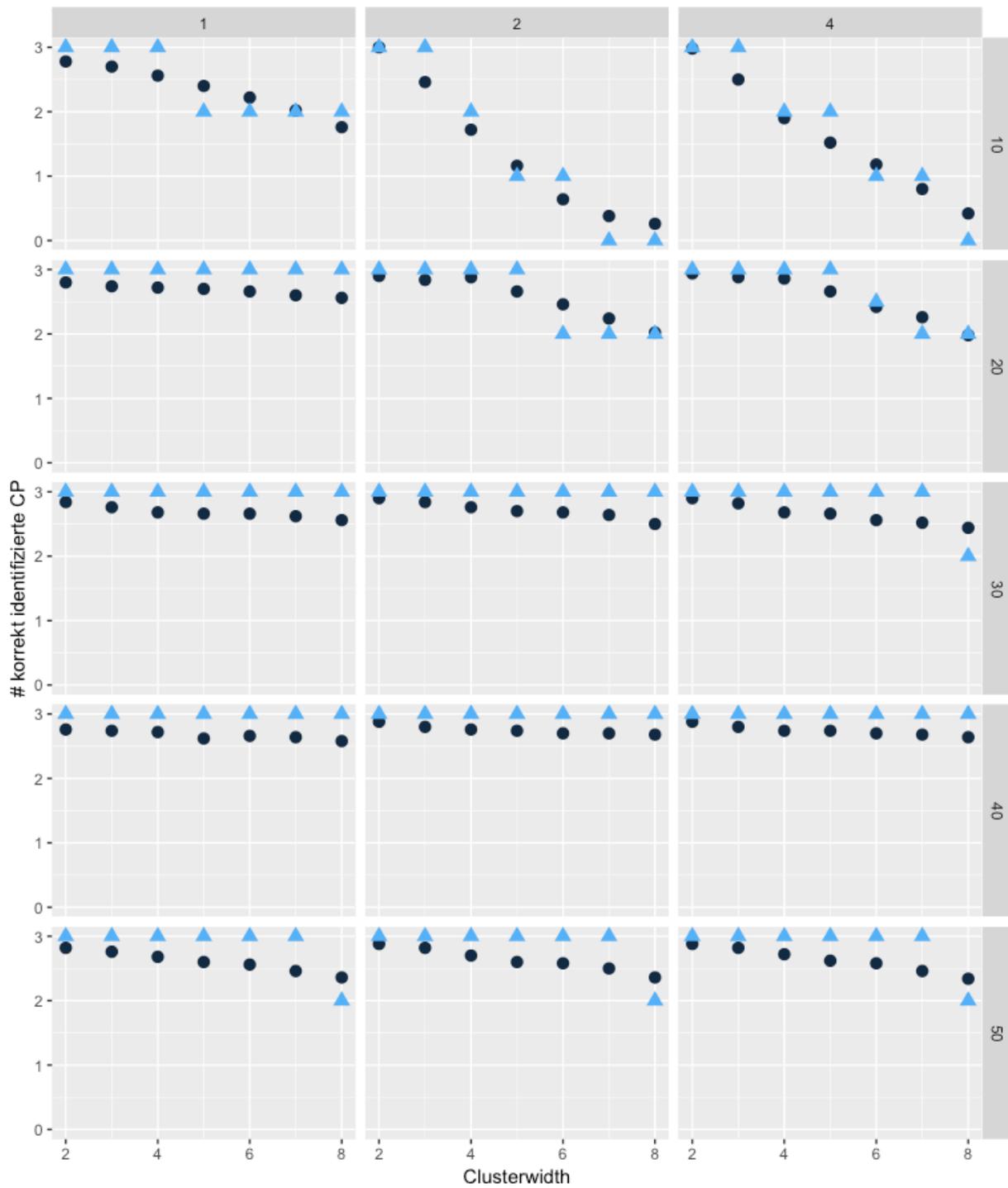


Abbildung 28: Wertevergleich der Parameter der BCPA für die speed switch Simulation. X-Achse oben: Sensitivitätsparameter K , X-Achse unten: Clusterwidth, Y-Achse links: Anzahl korrekt identifizierte CP mit Abweichung ± 3 , y-Achse rechts: window size.

Time-scale switch Simulation

In der time-scale switch Simulation wird in weniger Fällen ein CP korrekt identifiziert (Abbildung 29). Mit einer *clusterwidth* von 2 identifiziert die BCPA meist 1 korrekter CP. Häufig wird auch mit einer *clusterwidth* von 3 und einem K von 4 ein CP korrekt identifiziert. Oder mit einem K von 2 und einer *window size* von 30 oder 10.

Mit einer maximalen Abweichung von ± 3 erkennt das zuvor genannte Problem der *clusterwidth* sehr gut. Die BCPA hat in einigen Fällen direkt aufeinanderfolgende CP identifiziert, welche in einer *clusterwidth* = 2 nicht miteinbezogen werden, bei der Abweichung ± 3 jedoch trotzdem gezählt werden. In der Abbildung 30 lässt sich erkennen, dass mit einer *clusterwidth* von 2 bei K gleich 2 oder 4 immer 3 korrekte CP identifiziert werden. Dies geschieht, da die *clusterwidth* kleiner ist als die maximale Abweichung und damit ein CP doppelt gezählt werden kann.

Ansonsten werden oftmals 2 CP korrekt identifiziert. Ein gutes Resultat wird mit einer *window size* von 20 oder 30 erreicht. Vor allem die *window size* 20 in Kombination mit K gleich 4 liefert ein gutes Resultat.

RESULTATE

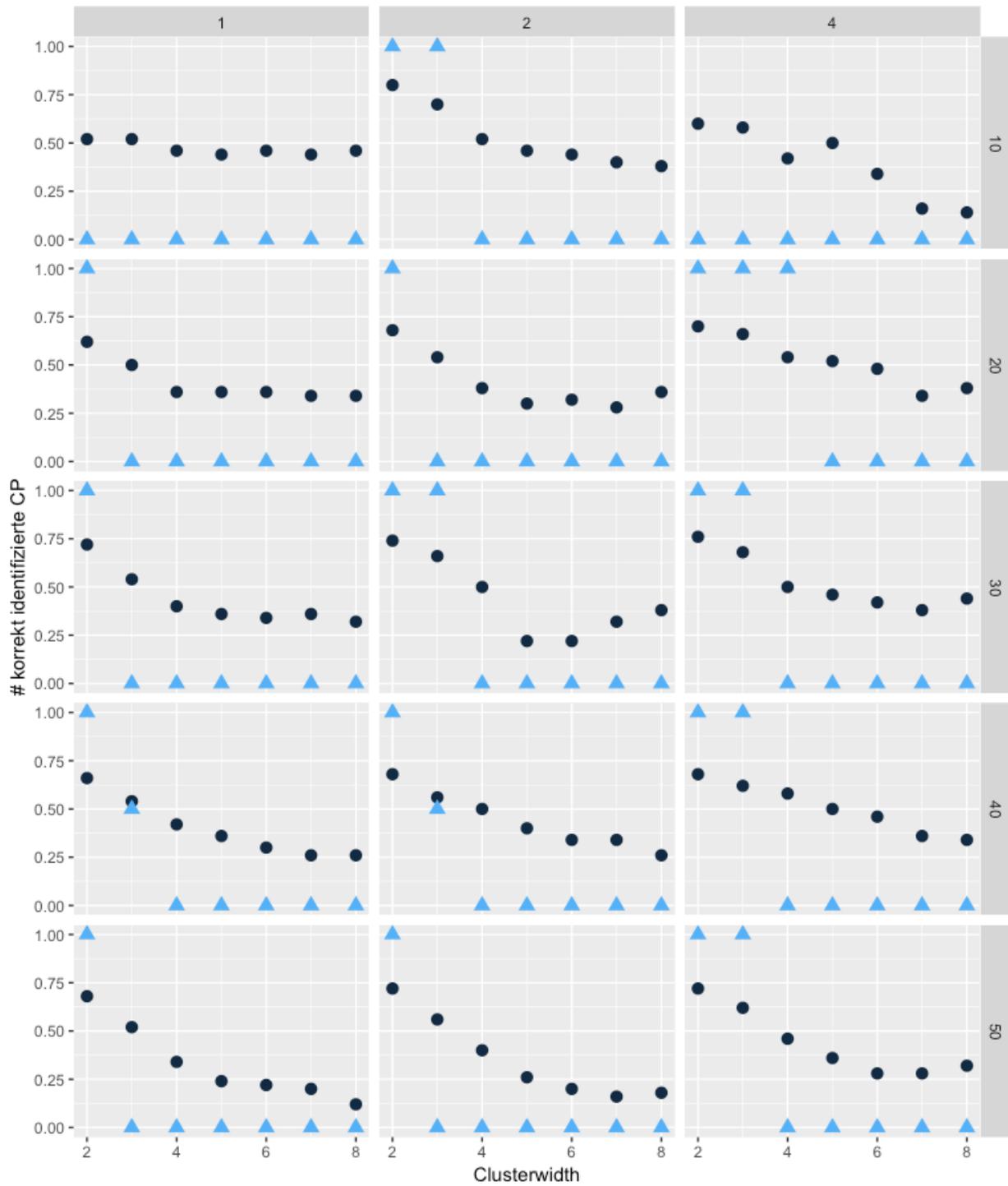


Abbildung 29: Wertevergleich der Parameter der BCPA für die time-scale switch Simulation. X-Achse oben: Sensitivitätsparameter K , X-Achse unten: Clusterwidth, Y-Achse links: Anzahl exakt korrekt identifizierter CP, y-Achse rechts: window size.

RESULTATE

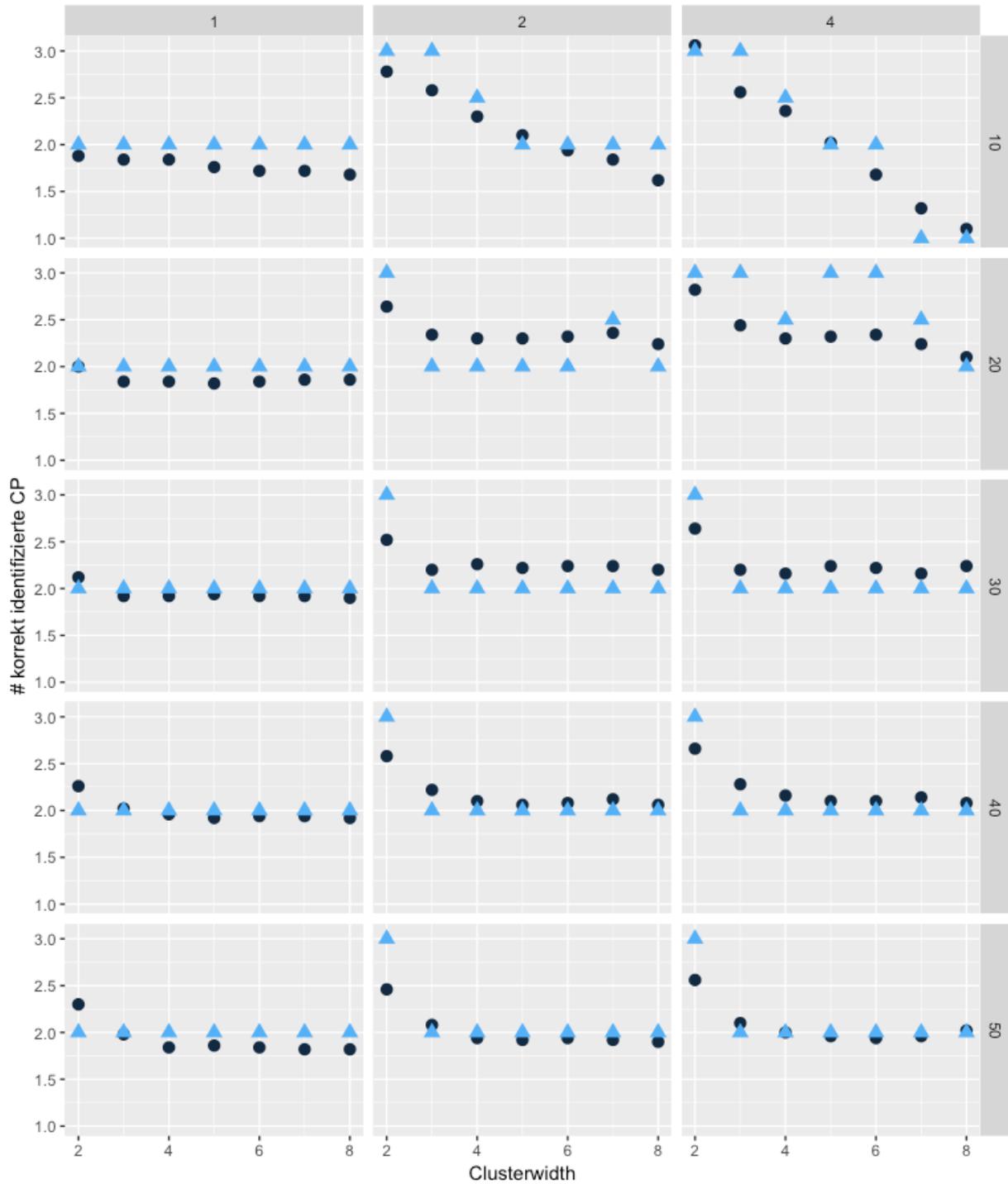


Abbildung 30: Wertevergleich der Parameter der BCPA für die time-scale switch Simulation. X-Achse oben: Sensitivitätsparameter K , X-Achse unten: Clusterwidth, Y-Achse links: Anzahl korrekt identifizierter CP mit Abweichung ± 3 , y-Achse rechts: window size.

RESULTATE

Kombiniert man die vier Zusammenstellungen, muss ein Kompromiss gefunden werden. Dabei müssen sowohl die speed change wie auch die time-scale switch mit und ohne Abweichung gut abschneiden. Daher wird für die Kombination der speed switch und time-scale switch die Parameter-Kombination $K = 2$, $window\ size = 30$ und $cluster\ width = 3$ gewählt.

Speed switch mit $K = 2$

Die Tabelle 22 (oben) zeigt, dass die BCPA mit K gleich 2 mehr als doppelt so viele CP identifiziert hat. Die Anzahl der korrekt identifizierten CP hat sich nur minimal verändert. Die Precision hat sich stark verschlechtert, da mehr CP identifiziert worden waren. Auch der Recall ist fast identisch.

Die Abweichung ± 1 hat einen tieferen Median als die originale speed switch Simulation mit K gleich 1. Ansonsten hat sich nichts verändert (Tabelle 22 unten).

Tabelle 22: Zusammenfassende Statistik der speed switch Simulation mit $K = 2$ segmentiert durch die BCPA: exakt (oben) und mit maximaler Abweichung von ± 3 (unten).

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	11	0	0	0
1. Quartil	13	0	0	0
Median	14.5	1	0.063	0.333
Mittelwert	14.5	0.64	0.045	0.213
3. Quartil	16	1	0.071	0.333
Max	18	2	0.167	0.667

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	11	2	0.111	0.667
1. Quartil	13	3	0.179	1
Median	14.5	3	0.2	1
Mittelwert	14.5	2.82	0.198	0.94
3. Quartil	16	3	0.214	1
Max	18	3	0.273	1

In der Tabelle 23 wird ersichtlich, dass durch die Veränderung des Parameters K von 1 auf 2 nur die Anzahl zwischen den CP zugenommen hat, nicht aber die Anzahl korrekt identifizierten CP.

RESULTATE

Tabelle 23: speed switch: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. K der BCPA gleich 2.

< I	I	I < II	II	II < III	III	III <
50	50	50	41	50	50	49

Time-scale switch mit K = 2

Auch in dieser Simulation hat sich der Median der Anzahl CP stark erhöht, nämlich mehr als verdreifacht (Tabelle 24 oben). Der Median der Anzahl korrekten CP liegt noch immer bei 0. Die Precision hat sich verschlechtert und der Recall ist in etwa identisch zur originalen time-scale switch Simulation mit K gleich 1.

Der Recall hat sich für alle Abweichungen verbessert. Mit Abweichung ± 3 liegt er sogar bei 1, im Gegensatz zur originalen time-scale switch mit K gleich 1 (Tabelle 24 unten).

Tabelle 24: Zusammenfassende Statistik der time-scale switch Simulation mit K = 2 segmentiert durch die BCPA: exakt (oben) und mit maximaler Abweichung von ± 3 (unten).

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	6	0	0	0
1. Quartil	11	0	0	0
Median	12	0	0	0
Mittelwert	11	0.44	0.038	0.147
3. Quartil	13	1	0.077	0.333
Max	16	2	0.2	0.667

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	6	1	0.077	0.333
1. Quartil	11	2	0.167	0.667
Median	12	3	0.2	1
Mittelwert	11	2.48	0.213	0.827
3. Quartil	13	3	0.25	1
Max	16	3	0.333	1

RESULTATE

In der Tabelle 25 wird ersichtlich, dass sich alle Werte stark erhöht haben, ausser die Anzahl Realisierungen, welchen den dritten CP korrekt identifiziert haben.

Tabelle 25: time-scale switch: Anzahl Realisierungen, bei welchen die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert werden und Anzahl Realisierungen, bei welchen zwischen den korrekten CP weitere CP identifiziert werden. K der BCPA gleich 2.

< I	I	I < II	II	II < III	III	III <
48	48	37	34	41	42	50

5.6 Kombination der Speed & Time-scale switch Simulation

Die FPT hat bei der Anzahl CP durchschnittlich einen CP mehr als bei der speed switch Simulation und gleich viele wie in der time-scale switch. Der Mittelwert der Anzahl korrekt identifizierter CP ist in dieser Simulation ebenfalls etwas höher, auf den Median hat dies aber keinen Einfluss.

Die Precision der BCPA hat sich stark verschlechtert, da sie grundsätzlich mehr CP gefunden hat, diese aber nicht unbedingt korrekt sind. Bei der Anzahl korrekt identifizierten CP sind die Werte ähnlich wie bei der speed switch Simulation, aber besser als bei der time-scale switch Simulation. Der Median der korrekt identifizierten CP liegt hier bei 1, bei der time-scale switch liegt er bei 0. Der Median des Recalls hat sich im Vergleich zur time-scale switch verbessert (0.333 statt 0) und ist vergleichbar mit der speed switch Simulation.

Tabelle 26: Zusammenfassende Statistik der speed und time-scale switch Simulation segmentiert durch die FPT (oben) und die BCPA (unten). Anzahl korrekt identifizierter CP ohne Abweichung (exakt).

FPT				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	2	0	0	0
1. Quartil	2	0	0	0
Median	3	0	0	0
Mittelwert	3.04	0.52	0.183	0.173
3. Quartil	4	1	0.333	0.333
Max	6	2	1	0.667

BCPA				
	# CP	# korrekt ident. CP	Precision CP	Recall CP
Min	9	0	0	0
1. Quartil	12	0.25	0.015	0.083
Median	13	1	0.071	0.333
Mittelwert	13.12	0.88	0.07	0.293
3. Quartil	14	1	0.089	0.333
Max	17	2	0.222	0.667

Vergleicht man die Precision und den Recall der FPT unter Einbezug der Abweichungen, kann man erkennen, dass der Median der Precision ähnlich liegt, derjenige des Recalls aber nicht mehr bei 0.333 sondern bei 0.667. Es werden also bei der Hälfte der Simulationen nicht nur ein CP, sondern zwei CP als korrekt identifiziert. Im Vergleich zur time-scale switch Simulation hat die FPT in dieser Simulation viel besser abgeschnitten.

Die Precision der BCPA hat sich stark verschlechtert, da die BCPA mehr CP identifiziert hat. Der Recall ist mit Ausnahme der Abweichung von ± 1 sehr ähnlich zur speed switch Simulation. Mit Berücksichtigung einer Abweichung von ± 1 liegen (mit wenigen Ausnahmen) alle Simulationen bei 2 von 3 korrekt identifizierten CP. Im Vergleich zur time-scale switch Simulation hat die BCPA nur betreffend Recall besser abgeschnitten. Bei den exakt identifizierten CP liegt der Median bei 0.333 statt 0. Bei einer Abweichung von ± 1 liegt er bei 0.667 statt bei 0.333 und bei Abweichung ± 2 und ± 3 sogar bei 1 statt bei 0.667. Bei der Abweichung von ± 1 haben (mit wenigen Ausnahmen) alle Simulationen 2 von 3 CP korrekt identifiziert. Mit Berücksichtigung der Abweichung ± 3 haben alle Simulationen (mit wenigen Ausnahmen) alle 3 CP korrekt identifiziert.

RESULTATE

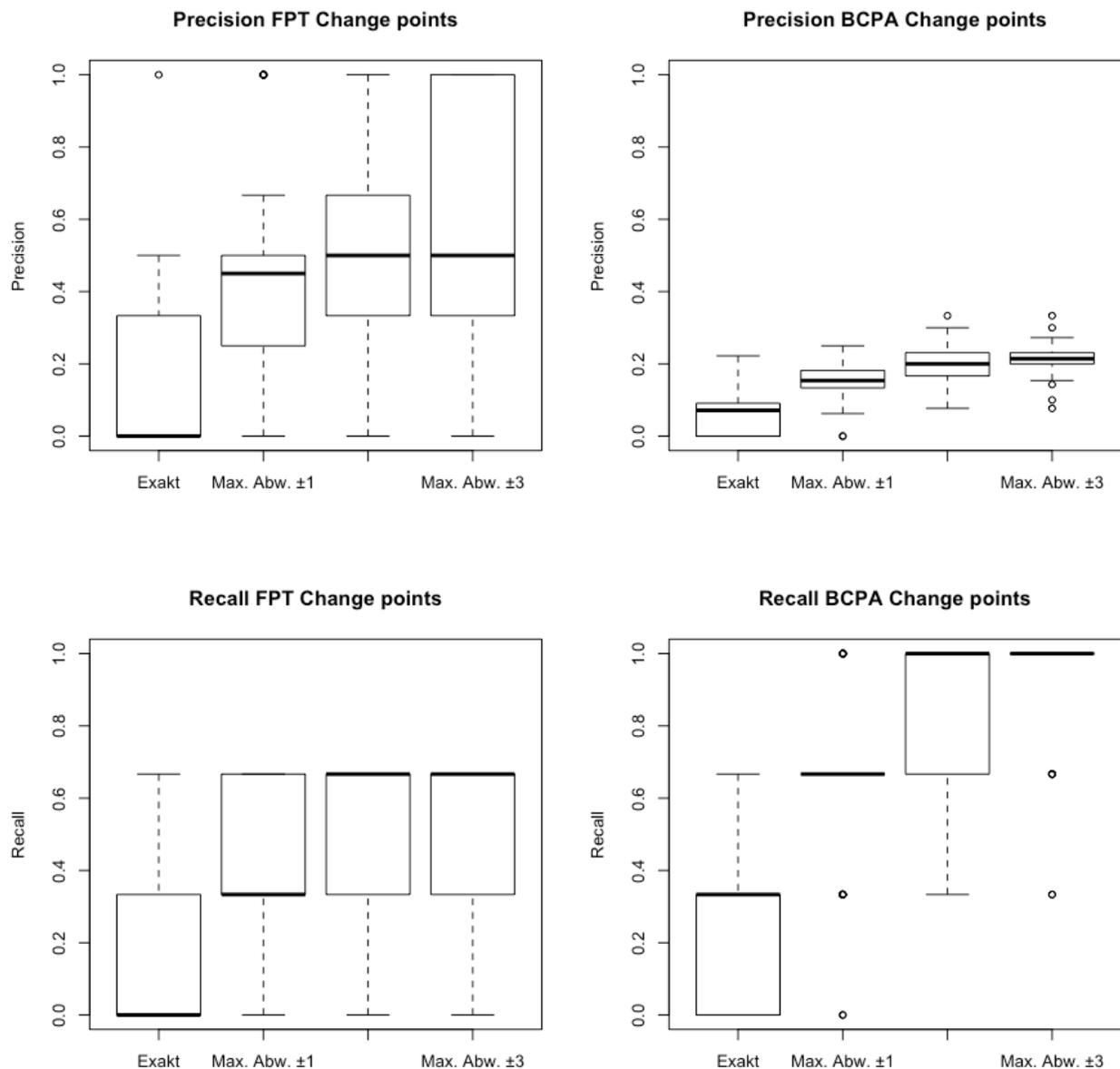


Abbildung 31: Kombination der speed switch und der time-scale switch Simulation: Boxplot zu Precision (oben) und Recall (unten) mit exakter Identifizierung sowie maximalen Abweichungen von ± 1 , ± 2 und ± 3 . Links: durch FPT segmentiert, rechts: durch BCPA segmentiert.

Die FPT sowie die BCPA haben, was die Anzahl korrekt identifizierter CP betrifft, besser abgeschnitten in der kombinierten Simulation als in den anderen beiden Simulationen. Beziehungsweise wurden bereits mit kleinerer Abweichung mehr CP korrekt identifiziert.

RESULTATE

Tabelle 27: Anzahl Realisierungen, welche die korrekten CP mit einer maximalen Abweichung von ± 3 korrekt identifiziert haben und Anzahl Realisierungen, die zwischen den korrekten CP weitere CP identifiziert haben.

	< I	I	I < II	II	II < III	III	III <
FPT	26	37	0	0	0	42	22
BCPA	50	47	50	44	50	49	50

6 DISKUSSION

In diesem Kapitel werden die in Kapitel 5 generierten Daten in einen Kontext gesetzt. Dies soll zeigen, dass die Resultate dieses Projekts einen aktuellen Beitrag zur Forschung liefern. In den ersten sechs Kapiteln (Kapitel 6.1 bis 6.6) werden die Resultate wenn möglich mit bestehender Literatur verglichen. Das Ziel ist es, ihre Erkenntnisse zu bestätigen oder zu widerlegen und neues Wissen zu ergänzen. Diese Kapitel bilden die Grundlage, um die Forschungsfragen im darauffolgenden Kapitel (Kapitel 6.7) zu beantworten. Zum Schluss wird dieses Projekt kritisch hinterfragt und Einschränkungen werden aufgezeigt.

6.1 Repliziert

Das Ziel dieses Kapitels ist es, einige der Resultate von Gurarie et al. (2016a) und Soleymani et al. (2017) bestätigen zu können oder allenfalls abweichende Resultate zu begründen. Dazu werden im Folgenden die Aussagen von Gurarie et al. (2016a) und Soleymani et al. (2017) in Kombination mit den Resultaten aus diesem Projekt genauer diskutiert.

Speed switch Simulation

Gurarie et al. (2016a) fanden heraus, dass beide Methoden die CP zwischen tiefen und hohen Geschwindigkeiten ausgezeichnet identifizierten. Die BCPA hatte etwas mehr Mühe, den schwächeren mittleren CP zu identifizieren. Sie begründen dies mit dem Einfluss der *persistence velocity* bei hoher Tortuosität auf die Geschwindigkeitsänderungen. Letztere werden abgeschwächt und sind nicht mehr so klar erkennbar. Auch Soleymani et al. (2017) betonen die Schwäche beider Methoden bei der Erkennung der mittleren Phasen. Zusätzlich hat die BCPA einige kleinere CP innerhalb der langsameren Phasen identifiziert (Gurarie et al. 2016a).

Aussage 1: «Each of the methods did an excellent job of identifying the change between low and high velocities in the first simulation» (Gurarie et al. 2016a: 74)

Die CP zwischen dem ersten und zweiten Segment sowie zwischen dem dritten und vierten Segment lassen sich sowohl in der Realisierung von Gurarie et al. (2016a) wie auch in denjenigen dieses Projekts visuell gut erkennen. Auch Lavielle (2005) hat diese beiden CP ungefähr korrekt identifiziert. Bei der Segmentierung durch die BCPA wurden der erste und der letzte CP gut erkannt, sowohl bei Gurarie et al. (2016a) wie auch in der Realisierung dieses Projekts. Starke Veränderungen in der mittleren Geschwindigkeit sind damit sowohl für die FPT wie auch für die BCPA identifizierbar. Damit kann diese erste Aussage bestätigt werden.

DISKUSSION

Aussage 2: «The BCPA did a poorer job of capturing the more subtle middle transition [...]» (Gurarie et al. 2016a: 74) und «[...] FPT (three segments) and BCPA (six segments) fail to capture the intermediate transitions.» (Soleymani et al. 2017: 1117)

Nicht nur bei Gurarie et al. (2016a) hatte die BCPA etwas Mühe mit der korrekten Identifizierung des mittleren CP. Bei Gurarie et al. (2016a) liegt der identifizierte CP vor dem korrekten CP, in diesem Projekt wurden je ein CP vor und hinter dem korrekten CP identifiziert.

Der mittlere CP lässt sich im Resultat der FPT allenfalls visuell erahnen, durch die Methode von Lavielle (2005) wurde er jedoch nicht erkannt. Gurarie et al. (2016a) haben sich zum mittleren CP in ihrer Publikation nicht geäußert. Auch in diesem Projekt haben beide Methoden Schwierigkeiten diese Veränderung in der Geschwindigkeit zu erkennen.

Die Anzahl der Segmente unterscheidet sich um eins. In der Realisierung von Gurarie et al. (2016a) identifizierte die BCPA sechs Segmente, in dieser Realisierung sieben. Dies zeigt, dass die Anzahl der Segmente leicht schwanken kann, sie sich aber um diese Anzahl herum manifestiert und tendenziell mehr Segmente identifiziert werden als tatsächlich korrekt wären. Damit können beide Teile dieser Aussage bestätigt werden. Als Ergänzung lässt sich sagen, dass die FPT mit Lavielle (2005) den mittleren CP nicht erkennen kann.

Aussage 3: «[The BCPA] picked up some more subtle changes within the slower phases.» (Gurarie et al. 2016a: 74)

Während die BCPA in Gurarie et al. (2016a) im ersten und im letzten Segment zusätzlich je einen CP identifizierte, hat sie in der Realisierung dieses Projekts im dritten und vierten Segment je einen zusätzlichen CP gefunden. Der in diesem Projekt identifizierte CP im letzten Segment bestätigt, dass die BCPA zusätzlich schwächere CP bei tiefen Geschwindigkeiten und dadurch hoher Tortuosität identifiziert. Jedoch hat sie auch einen zusätzlichen CP im dritten Segment, welches durch die höchste Geschwindigkeit und tiefste Tortuosität in dieser Simulation charakterisiert ist, gefunden. Dies zeigt, dass fälschlicherweise identifizierte CP auch bei höheren Geschwindigkeiten auftreten können. Diese Aussage von Gurarie et al. (2016a) kann somit bestätigt werden. Hinzuzufügen ist jedoch, dass fälschlicherweise identifizierte CP nicht nur in der langsamen Phase auftreten können.

Time-scale switch

Gemäss Gurarie et al. (2016a) liefert die FPT in der time-scale switch Simulation keine brauchbaren Informationen. Soleymani et al. (2017) ergänzen, dass die FPT auch für die mittleren beiden Phasen unbrauchbar war. Die BCPA hingegen hat alle drei CP genau identifiziert (Gurarie et al. 2016a) und damit korrekt in vier Phasen geteilt (Soleymani et al. 2017). Gurarie et al. (2016a) begründen die Stärke der

DISKUSSION

BCPA in dieser Simulation damit, dass die BCPA die Autokorrelation explizit in die Segmentierung miteinbezieht. Im Gegensatz dazu macht die FPT nur Gebrauch von der Geschwindigkeit und schneidet deshalb bei dieser Simulation schlechter ab.

Aussage 1: «[...] the FPT is fairly uninformative [...]» (Gurarie et al. 2016: 75) und «[...] FPT remains uninformative about the intermediate transitions [...]» (Soleymani et al. 2017: 1117)

Ändert sich die mittlere Geschwindigkeit nicht, sondern nur die time-scale, lieferte die FPT in Gurarie et al. (2016a) keine brauchbaren Informationen (Gurarie et al. 2016a). In der Realisierung dieses Projekts liegen die durch Lavielle (2005) identifizierten CP hingegen nicht weit weg von den korrekten CP. Der mittlere CP wird zwar wieder nicht erkannt und ein zusätzlicher CP im letzten Segment wird identifiziert, aber die Segmentierung gibt ein erstes brauchbares Resultat. Die Aussage kann mit dieser Realisierung zwar bestätigt werden, es kann aber auch Fälle geben, in der die FPT brauchbare Informationen liefert.

Aussage 2: «The BCPA accurately detected the change points for all three major transitions [...]» (Gurarie et al. 2016a: 75) und «[...] BCPA detects all the four phases accurately [...]» (Soleymani et al. 2017: 1117)

Laut Gurarie et al. (2016a) hat die BCPA in ihrer Realisierung alle CP korrekt identifiziert. Zusätzlich wurden im letzten Segment noch zwei weitere CP gefunden. Die Autoren präsentieren die BCPA für diese Simulation als geeignete Methode. Die Realisierung in diesem Projekt zeigt ein etwas anderes Bild. Der letzte und zweitletzte CP werden zwar nahe der korrekten CP identifiziert, der erste CP wird jedoch klar verfehlt. Der Wechsel von 2 auf 20 in der time-scale hat der BCPA hier Mühe bereitet. Das dritte Segment, in welchem die time-scale sehr hoch ist (11) wird in beiden Realisierungen relativ genau identifiziert. Damit kann die Aussage nur teilweise bestätigt werden. Nicht in jeder Realisierung kann die BCPA alle CP korrekt identifizieren.

DISKUSSION

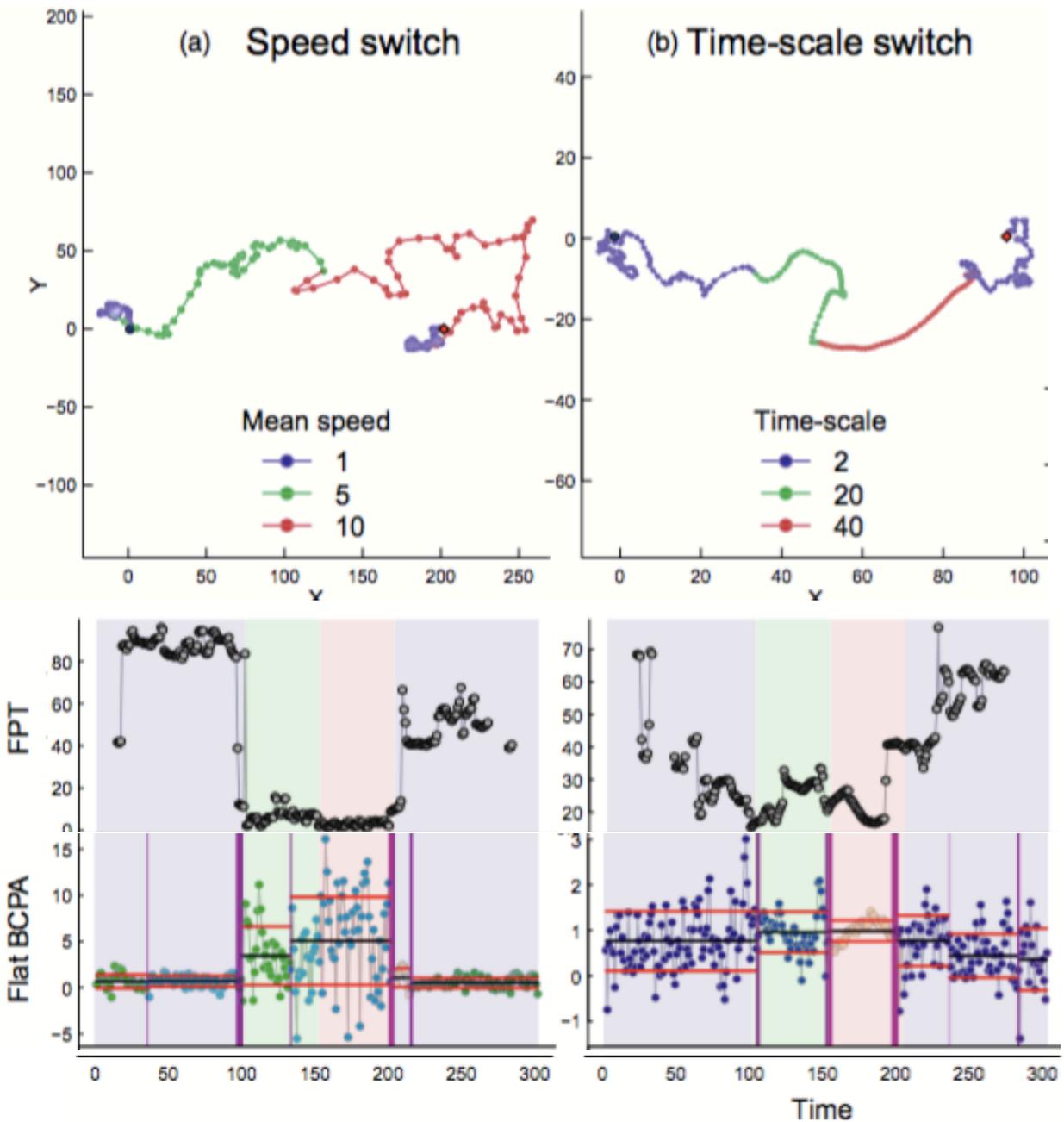


Abbildung 33: Abgeänderte Darstellung aus Gurarie et al. (2016a). Sie zeigt die simulierten Trajektorien (oben) und die Resultate der beiden Methoden (FPT mittig, BCPA unten). In den Trajektorien kennzeichnet blau die erste und letzte Phase, grün die zweite und rot die dritte. Die vertikalen Linien (unten) zeigen die CP, die horizontalen schwarzen Linien den geschätzten Mittelwert, die horizontalen roten Linien repräsentieren das 95% Konfidenzintervall. Hellere Farben der Punkte bedeuten tiefere Werte der time-scale und damit eine höhere Tortuosität. Die Hintergrundfarben präsentieren die vier Segmente der Simulation.

Continuous speed switch

Soleymani et al. (2017) haben für diese Simulation festgestellt, dass sowohl die FPT wie auch die BCPA keine nützlichen Resultate liefern. Beide Methoden teilen die Trajektorie in zu viele Segmente und die Lage der CP stimmt nicht mit den korrekten CP überein (Soleymani et al. 2017).

Aussage 1: «[...] identified excessive numbers of segments» (Soleymani et al. 2017: 1117).

Unter der Verwendung der FPT liefert die Realisierung dieses Projekts etwas andere Resultate. Im Gegensatz zu der Realisierung von Soleymani et al. (2017) stimmt die Segmentanzahl mit der korrekten Anzahl überein. Die BCPA hingegen hat die Trajektorie dieses Projekts in sehr viele Segmente unterteilt. Es sind, wie auch in der speed switch dieses Projekts, etwa doppelt so viele CP wie tatsächlich korrekt wären. Obige Aussage kann für die BCPA bestätigt werden, für die FPT jedoch nicht. Diese kann durchaus auch weniger Segmente identifizieren.

Aussage 2: «[...] the positions of the extracted segments [identified by the BCPA] are not representative» (Soleymani et al. 2017, Appendix S4: 4)

Die Lage der CP sieht in diesem Projekt sehr zufällig aus, da die CP ziemlich gleichmässig über die Trajektorie verteilt sind. Nur zwei der identifizierten CP liegen nahe eines CP. Die Aussage von Soleymani et al. (2017) kann damit bestätigt werden.

Aussage 3: «[...] FPT did not show any flexibility and resulted in completely missing the segments» (Soleymani et al. 2017, Appendix S4: 2)

In der Realisierung dieses Projekts liegen zwei der vier CP nahe an dem ersten und zweiten korrekten CP. Ab der Geschwindigkeit 3 werden keine CP mehr identifiziert, dafür im Segment mit Geschwindigkeit 1 zwei zusätzliche. Die FPT-Werte schwanken bei den Geschwindigkeiten 1 und 2 viel stärker als bei den Geschwindigkeiten 3, 4 und 5. Die Zeit, in der das Individuum innerhalb des Radius bleibt, ist vor allem für die Geschwindigkeiten 3, 4 und 5 sehr ähnlich. Dadurch werden von der FPT keine CP mehr identifiziert. Das Resultat der BCPA ähnelt also stark demjenigen von Soleymani et al. (2017) im Gegensatz zum Resultat der FPT. Die Aussage 3 kann somit nur teilweise bestätigt werden, denn die FPT hat zwei Segmente ziemlich genau identifiziert, die anderen 3 hingegen eindeutig verfehlt.

Zusammenfassen lässt sich also sagen, dass die Aussagen von Gurarie et al. (2016a) grösstenteils bestätigt werden können. Insbesondere die Aussage, dass beide Methoden die Veränderung der tiefen und hohen Geschwindigkeiten (erster und letzter CP) der speed switch Simulation exakt identifizieren können, hat sich als zutreffend erwiesen. Gewisse Aussagen von Gurarie et al. (2016a) können zwar bestätigt werden, brauchen jedoch gewisse Erweiterungen. Beispielsweise erkennt die BCPA nicht nur in der langsamen

Phase weitere CP, sondern auch in anderen Phasen können fälschlicherweise identifizierte CP auftreten. Eine Aussage von Gurarie et al. (2016a) muss widerlegt werden. Die BCPA erkennt nicht immer alle drei CP der time-scale switch Simulation. Mit Ausnahme von dieser letzten Aussage können alle Aussagen von Gurarie et al. (2016a), mit teilweise entsprechenden Ergänzungen, betätigt werden. Bei Soleymani et al. (2017) hingegen sind es zwei Aussagen, die widerlegt werden müssen. Erstens, ebenfalls wie bei Gurarie et al. (2016a), kann die BCPA nicht unbedingt alle Segmente der time-scale switch Simulation korrekt erkennen. Zweitens ist die FPT durchaus in der Lage, die ersten beiden CP in der continuous speed switch Simulation korrekt zu erkennen und die Menge an Segmenten in Grenzen zu halten. Bestätigt werden kann jedoch, dass die FPT nicht in der Lage war, die mittleren Segmente in der speed switch Simulation zu erkennen. Des Weiteren können die Aussagen bestätigt werden, dass die BCPA sehr viele Segmente in der continuous speed switch Simulation erkannt hat und deren CP nicht repräsentativ lokalisiert wurden.

6.2 50 Realisierungen

Wie aus den Resultaten von Kapitel 5.2 mit 50 Realisierungen ersichtlich wird, bestätigen sie die Resultate einer einzigen zufälligen Realisierung aus Kapitel 5.1 sowie derjenigen von Gurarie et al. (2016a) und Soleymani et al. (2017) zu einem grossen Teil. Die FPT sowie die BCPA sind in der Lage, starke Veränderungen in der Geschwindigkeit zu identifizieren und den CP unweit des korrekten CP zu positionieren. Mit schwächeren Veränderungen haben sowohl die FPT wie auch die BCPA Schwierigkeiten. Während die BCPA den CP in der Nähe des korrekten CP identifiziert oder mehrere in dessen Nähe findet, kann die FPT den mittleren CP überhaupt nicht erkennen. Ebenfalls grosse Mühe hat die FPT mit der time-scale switch und mit der continuous speed switch Simulation. In der time-scale switch Simulation spielt die von Gurarie et al. (2016a) definierte «model misspecification» auf jeden Fall eine Rolle. Diese weist darauf hin, dass die Nutzerin oder der Nutzer sich bewusst sein soll, auf welchen *path characteristics* die Methode basiert. Da die FPT die Zeit misst, welche ein Individuum zur Durchquerung des Radius benötigt und damit auf der Geschwindigkeit basiert (Gurarie et al. 2016a), lässt sich auch ihr Versagen in der time-scale switch erklären. Die BCPA hingegen basiert auf der *persistence velocity* und damit sowohl auf der Geschwindigkeit wie auch auf dem *turning angle*. Aus diesem Grund konnte die BCPA auch die CP in der time-scale switch Simulation nahezu korrekt identifizieren. In der continuous speed switch Simulation versagen jedoch beide Methoden und konnten, ausser wenigen Ausnahmen, keine CP korrekt identifizieren. Eine solche Ausnahme war zufälligerweise die Realisierung aus Kapitel 6.1 dieses Projekts. Die durch diese Realisierung widerlegte Aussage 3 von Soleymani et al. (2016) zur FPT kann mit den 50 Realisierungen wieder bestätigt werden. Die FPT hat in der ersten Phase mehrere CP identifiziert, da bei der Geschwindigkeit 1 grosse Schwankungen stattfanden, jedoch nur sehr

wenige davon korrekt lokalisiert. Die BCPA hingegen hat über die ganze Trajektorie verschiedene CP identifiziert. Schwankungen in der Geschwindigkeit und in den *turning angles* hatten jedoch den grösseren Einfluss, als dies die Zunahme der Geschwindigkeit hatte. Die Variabilität innerhalb der *turning angles* wie auch innerhalb der mittleren Geschwindigkeit ist so stark, dass die Zunahme der mittleren Geschwindigkeit um jeweils 1 nicht genügend ins Gewicht fällt. Um kontinuierliche Verhaltensänderungen zu erkennen, sind deshalb sowohl die FPT wie auch die BCPA ungeeignet. Eine alternative Methode ist die DWT von Soleymani et al. (2016, siehe Kapitel 2.2.3). Bei dieser Methode wurde gezeigt, dass sie kontinuierliche Verhaltensänderungen erkennen kann. Solche kontinuierlichen Verhaltensänderungen kommen hauptsächlich bei hoch aufgelösten Daten zum Vorschein, die beispielsweise die Beschleunigung eines Individuums durch schnell aufeinanderfolgende Messungen aufnehmen kann. Sind die Daten weniger hoch aufgelöst, werden eher Sprünge in der Geschwindigkeit und/oder in der Tortuosität festgestellt und damit auch Methoden wie die FPT oder BCPA wieder brauchbarer. Dies zeigen die Resultate der speed switch und der continuous speed switch Simulation. Sowohl die FPT wie auch die BCPA haben den Verlauf der mittleren Geschwindigkeit von 1 bis 5 in der continuous speed switch Simulation nicht erkannt. Würden die Datenpunkte zwischen der Geschwindigkeit 1 und 5 ausgelassen (aufgrund tieferer *sampling frequency*), würde ein abrupter Sprung von 1 direkt auf 5 stattfinden (wie in der speed switch Simulation). In diesem Fall haben die Resultate gezeigt, dass sowohl die FPT als auch die BCPA in der Lage sind, einen solchen Sprung zu identifizieren. Bei der Auswahl der Methode muss also immer auch die *sampling frequency* berücksichtigt werden. Allenfalls können in einem Datensatz die Datenpunkte ausgedünnt werden, um bessere Resultate zu erzielen. Eine Methode, die testet, ob eine Reduktion der Datenpunkte einen signifikanten Einfluss auf die Topologie der Trajektorie hat, ist die «Line Simplification» (Edelhoff et al. 2016b). Eine Ausdünnung der Daten sollte jedoch immer mit Vorsicht durchgeführt werden, da durch diesen Prozess wertvolle Informationen verloren gehen können.

Zudem bietet sich an, den CP immer mit einer gewissen zeitlichen Flexibilität zu betrachten. Da ein Individuum sein Verhalten meist nicht von einer Sekunde auf die andere wechselt, sondern eine gewisse Zeit dafür benötigt, ist oft auch kein klarer CP definierbar. Aus diesem Grund muss eine gewisse Abweichung auch in einer quantitativen Analyse miteinbezogen werden. Daher wird in Kapitel 6.3 mit unterschiedlichen maximalen Abweichungen gearbeitet.

6.3 Maximale Abweichung

Wie die Resultate der speed switch Simulation zeigen, wird bereits bei der Berücksichtigung einer Abweichung von ± 1 das Resultat der korrekt identifizierten CP gesteigert. Bei der FPT werden bei 50% der Realisierungen ein CP von 3 korrekt identifiziert, bei der BCPA sind es 2 von 3 CP. Mit einer

DISKUSSION

Abweichung von ± 3 werden ca. 50% der durch die FPT identifizierten CP korrekt identifiziert. Die Resultate streuen sich jedoch noch immer von 0 bis 1 und der zweite CP wird in keiner Realisierung erkannt. Die BCPA hat mit einer Abweichung von ± 3 ebenfalls ca. 50% der identifizierten CP korrekt identifiziert. Die Variabilität ist jedoch deutlich kleiner als bei der FPT. Die BCPA hat auch den ersten und den letzten CP in jeder der 50 Realisierung erkannt. Dies führt dazu, dass die BCPA nahezu 100% der CP korrekt identifiziert hat. Bei der time-scale switch und der continuous speed switch Simulation hat die FPT auch mit einer Abweichung von ± 3 , ausser bei wenigen Ausnahmen, keine CP korrekt identifiziert. Die BCPA hingegen hat bei der time-scale switch Simulation ca. 60% der identifizierten CP korrekt identifiziert. Dies entspricht ca. 2/3 der korrekten CP. Bei der continuous speed switch Simulation hat auch die Abweichung von ± 3 bei der BCPA keinen positiven Einfluss auf das Resultat (ausser bei wenigen Ausnahmen). Da es bis anhin keine Literatur gibt, die eine vergleichbare Auswertung gemacht hat, lassen sich diese Resultate mit keiner Publikation vergleichen. Wenn in einer zukünftigen Studie ebenfalls statistische Auswertungen gemacht werden sollten, empfiehlt sich resultierend aus diesem Projekt auf jeden Fall, eine gewisse maximale Abweichung um die identifizierten CP zuzulassen. Die Grösse der Abweichung hängt von der *sampling frequency* und dem zu analysierenden Massstab beziehungsweise den möglichen Verhaltensweisen in den Daten ab. Wird die Abweichung zu klein gewählt, hat sie einen zu geringen Effekt, wird sie jedoch zu gross gewählt, können kürzere Verhaltensänderungen übersehen werden oder falsche CP einander zugeordnet. Eine Möglichkeit, die in der Bewegungsanalyse allenfalls verwendet werden könnte, stammt aus der Segmentierung von Trajektorien unterschiedlicher Transportmitteln. Prelipcean et al. (2016) präsentieren das «interval alignment», welches die Verwendung von maximalen Abweichungen überflüssig machen könnte. In der Ökologie wurde diese Methode jedoch noch nie angewendet. Werden aber maximale Abweichungen verwendet, kann auf die Erfahrungen aus diesem Projekt zurückgegriffen werden. Diese zeigen nämlich, dass es sich lohnt, mit einer geringen Abweichung zu beginnen und diese schrittweise zu erhöhen. Die Resultate zeigen relativ schnell, ob die Berücksichtigung der Abweichung einen Mehrwert bringt und hilft, die Fragestellung zu beantworten, oder ob sie nur eine weitere Ungenauigkeit mit sich bringt und die Aussagekraft verwischt. Werden die Methoden auf reale Datensätze angewandt kommt man nicht umhin, das Wissen um die relevanten Zeitspannen eines Verhaltensmusters einer spezifischen Tierart in die Analyse einfließen zu lassen. So haben Le Corre et al. (2014) zum Beispiel definiert, dass der im Juni identifizierte CP den Beginn der Kalbungszeit bei einer Karibuherde markiert. Die Autoren wussten bereits im Vorherein, wann die Kalbungszeit ungefähr beginnt und konnten so, basierend auf ihrem bestehenden Vorwissen, für diese Abweichung eine gewisse sinnvolle Zeitspanne bestimmen.

6.4 FPT: mittlerer CP

Da die FPT trotz Abweichung den mittleren CP in der speed switch Simulation nicht identifizieren konnte, wurde der Wechsel von der Geschwindigkeit 5 auf 10 separat und detaillierter betrachtet. Dies entspricht einem Ansatz von Pinaud (2008, inspiriert durch Fauchald & Tveraa 2003), der eine Segmentierung in mehreren Schritten vorschlägt. Auch Le Corre et al. (2014) haben die Karibudaten in zwei Schritten segmentiert. Die erste Segmentierung diente dazu, in der ganzen Trajektorie die Winterpausen zu identifizieren. Da die FPT-Werte innerhalb der Kalbungsphasen viel tiefer sind als diejenigen der Winterpausen, segmentierten Le Corre et al. (2014) in einem zweiten Schritt die Zeitspanne zwischen den Winterpausen separat. So konnte den Daten gerecht werden, weil der Kontrast zwischen den Winterpausen und den anderen Perioden zu hoch war, als dass in demselben Schritt auch die Kalbungsphasen hätten identifiziert werden können. Die Analyse zeigte anschliessend, dass die Kalbungsphasen auch viel kürzer waren und tiefere FPT-Werte aufwiesen, wodurch sie im Gesamtbild schlechter identifiziert werden konnten. Dies zeigt einmal mehr, dass die Bewegung eines Individuums durch unterschiedliche ökologische Faktoren auf verschiedenen räumlichen Skalen beeinflusst wird (Fauchald & Tveraa 2003) und so zu unterschiedlichen Verhaltensmustern auf verschiedenen Massstäben führt (Nathan et al. 2008, Thiebault & Tremblay 2013). Beispielsweise kann ein Meeresvogel zwischen fliegen und auf dem Wasser landen hin und her wechseln. Während er fliegt, kann er während einer längeren Zeit in eine bestimmte Richtung fliegen, er kann aber auch möglichst schnell vor einem Feind fliehen, oder in kurzen Zeitabständen nach Futter suchen usw. (Thiebault & Tremblay 2013). Da die FPT massstabsabhängig ist, identifiziert sie je nach Massstab möglicherweise auch andere Verhaltensmuster (Fauchald & Tveraa 2003).

Die Resultate aus dem Kapitel 5.4 zeigen, dass die FPT in der Lage ist, den Wechsel von 5 auf 10 zu erkennen, wenn dieser separat betrachtet wird. Die Unterschiede zwischen der mittleren Geschwindigkeit 5 und 10 waren jedoch zu klein im Vergleich zu der Variabilität der mittleren Geschwindigkeit innerhalb der Phasen, sodass nie mehr als die Hälfte der CP korrekt identifiziert wurden. Die Resultate zeigen aber auch, dass die Chance für ein treffendes Resultat erhöht wird, wenn ein Segment separat betrachtet wird und nicht nur innerhalb einer ganzen Trajektorie. Le Corre et al. (2014) weisen zudem darauf hin, dass fälschlicherweise identifizierte CP oder fälschlicherweise nicht identifizierte CP auch eine biologische Bedeutung haben können. Beispielsweise können fehlende CP zur Kalbungszeit darauf hinweisen, dass eine Karibukuh kein Kalb in sich trägt und ihre Bewegung daher nur analog zur Herde verlangsamt, jedoch kein typisches Kalbungsverhalten zeigt (Le Corre et al. 2014). Die Methoden können Hinweise auf solche Muster geben, die Resultate müssen aber auch hier im Nachhinein mit Hilfe von Hintergrundwissen richtig interpretiert werden (Le Corre et al. 2014).

Eine andere Möglichkeit um eine Trajektorie auf unterschiedlichen Massstabsebenen zu untersuchen kann auch darin bestehen, den Radius auf die Situation und den Massstab des zu analysierenden Verhaltensmusters anzupassen. In diesem Projekt wurden unterschiedliche Radien zur Segmentierung verwendet. Ein kleinerer Radius konnte zwar den Mittelwert der korrekt identifizierten CP etwas anheben, hatte jedoch keinen Einfluss auf den Median. Die FPT wird oft verwendet, um die Massstäbe, auf denen «area-restricted search» (ARS) stattfindet, zu identifizieren. Die unterschiedlichen Massstäbe des ARS können aufgrund verschiedener Radien und der darauf basierenden FPT-Werte herausgefunden werden. Wird anschliessend die relative Varianz der FPT-Werte in Relation zu den verschiedenen Radien dargestellt, präsentieren die Maxima der Kurve den Massstab, auf welchem ein Individuum seinen Suchaufwand bezüglich Nahrung oder Ruheplatz intensiviert (Fauchald & Tveraa 2003, siehe auch Kapitel 3.2). Anhand dieser Radien kann dann das Verhalten auf verschiedenen Massstäben identifiziert werden. Da der Radius in diesem Projekt kaum einen Einfluss auf die Resultate hatte, ist in diesen Simulationen vermutlich auch kein mehrskaliges Verhalten vorhanden.

6.5 BCPA: Wertevergleich der Parameter

Die Resultate in Kapitel 5.5 zeigen, dass Veränderungen in den Werten der Parameter im konkreten Beispiel dieses Projekts darüber entscheiden, ob der Median bei keinem CP oder bei einem korrekt identifizierten CP liegt. Wird eine Abweichung von ± 3 berücksichtigt, erhöht die Veränderung den Median auf zwei oder drei korrekt identifizierte CP. Ausnahme bildet die *window size* 10, bei welcher der Median zwischen 0 und 3 liegt. Bei der time-scale switch Simulation liegt er ebenfalls bei 0 oder 1 beziehungsweise zwei oder drei CP mit Abweichung ± 3 . Der Median liegt hingegen seltener bei 1 bzw. 3 CP als dies bei der speed switch Simulation der Fall ist. Der Wertevergleich dieses Projekts kann nicht direkt mit demjenigen von Morelle et al. (2017a) verglichen werden. In diesem Projekt wird der Einfluss auf die Anzahl korrekt identifizierte CP analysiert, während bei Morelle et al. (2017a) die Anzahl identifizierter Segmente verwendet wird. Morelle et al. (2017b: Fig. S1.2) haben ebenfalls die *window size*, den Sensitivitätsparameter K und die *cluster width* miteinander verglichen (Abbildung 34). Die Abbildung zeigt, dass in ihrem Beispiel die Anzahl Segmente mit zunehmender *cluster width* abnimmt. Bei der *cluster width* 4 gibt es eine klare Änderung der Steigung. Mit tieferer *cluster width* und höherer *window size* (20 bis 50) nimmt die Anzahl Segmente stärker ab als mit höherer *cluster width*. Dieses Verhalten ist sehr ähnlich zu der Anzahl exakt korrekt identifizierten CP in diesem Projekt. Morelle et al. (2017a) sind zum Schluss gekommen, dass die Variation der Werte für die drei Parameter in ihrem Fall keine grossen Auswirkungen auf die Anzahl der Segmente hatte.

DISKUSSION

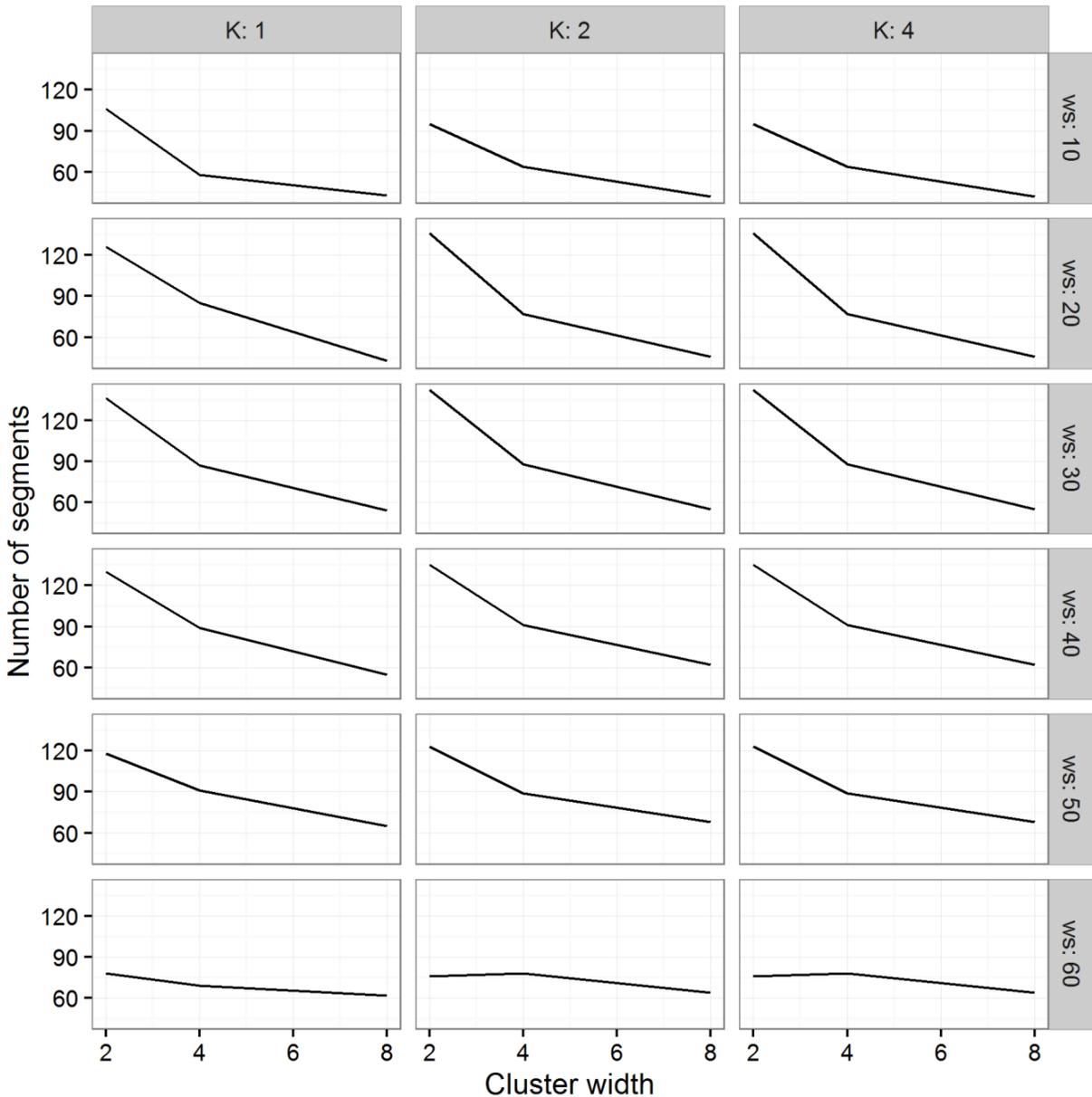


Abbildung 34: Auswirkungen der Parameter der BCPA (clusterwidth, Sensitivitätsparameter K und window size ws) auf die Anzahl Segmente dreier Trajektorien von subadulten, männlichen Wildschweinen (Morelle et al. 2017a).

Der Wertevergleich der Parameter in diesem Projekt zeigt jedoch (beide Simulationen mit und ohne Abweichung ± 3 kombiniert), dass gewisse Konstellationen nicht verwendet werden sollen. Beispielsweise ist eine *window size* von 10 sowie eine *clusterwidth* grösser als 6 eher nicht zu empfehlen. Zudem hat der Sensitivitätsparameter K einen Einfluss auf die Menge der identifizierten CP. Mit K gleich 1 werden weniger CP identifiziert als wenn K grösser ist. Der Einfluss einer Werteveränderung macht selten mehr als einen CP aus. Diese Resultate zeigen, dass es sich auf jeden Fall lohnt, mit unterschiedlichen

Wertekombinationen zu experimentieren und eine Analyse, ähnlich zu dieser in diesem Kapitel, durchzuführen.

6.6 Kombination der Speed & Time-scale switch Simulation

Die FPT hat in der Kombination von speed und time-scale switch Simulation minimal besser abgeschnitten als in Kapitel 5.4 (beide mit Radius 4), denn der Mittelwert der Anzahl korrekten CP (ohne Abw.) liegt etwas höher (Tabelle 27 oben). Der Median liegt trotzdem noch bei 0, wenn keine Abweichung miteinbezogen wird. Die Precision sowie der Recall haben sich durch die Kombination kaum verändert (Abbildung 31). Wird aber eine Abweichung von ± 1 berücksichtigt, liegt der Median der Precision bereits höher und die Streuung innerhalb der 50 Simulationen ist kleiner. Wird eine Abweichung von ± 3 berücksichtigt, liegt der Median des Recalls nicht mehr bei 0.333, sondern bei 0.667. Es werden also bei 50% der Realisierungen nicht mehr nur einer der korrekten CP identifiziert, sondern 2. Die Anzahl der korrekt identifizierten CP hat für den ersten und letzten CP mit Beachtung einer Abweichung von ± 3 etwas zugenommen (Tabelle 28). Der zweite CP wird noch immer in keiner Realisierung identifiziert, die Anzahl der identifizierten CP zwischen den korrekten CP kann jedoch minimiert werden.

Aus dem Vergleich der Resultate wird ersichtlich, dass der Einbezug einer maximalen Abweichung einen grösseren Einfluss auf die Anzahl korrekt identifizierten CP beziehungsweise auf den Recall hat als die Veränderung des Radius. Denn andere Radien hätten in dieser Simulation vermutlich keinen grossen Einfluss gehabt, ähnlich wie dies in den Resultaten von Kapitel 5.4 zu erkennen ist. In realen Datensätzen müssten unterschiedliche Radien in Betracht gezogen werden, da jedes Individuum auf eine Reihe unterschiedlicher ökologischen Faktoren reagiert, welche auf unterschiedlichen räumlichen Massstäben zusammenspielen (Fauchald & Tveraa 2003). ARS (*area restricted search*) kann beispielsweise durch das Maximum in der Varianz der FPT-Werte identifiziert werden (Fauchald & Tveraa 2003, siehe Kapitel 3.2).

Des Weiteren spielt die Grösse des Radius eine Rolle, wenn kleinräumiges Verhalten identifiziert werden möchte. In diesem Fall muss der Radius kleiner sein als wenn grossräumiges Verhalten (z. B. Migration) identifiziert werden soll. Eine weitere Möglichkeit wäre ein Vorgehen in mehreren Schritten, wie Le Corre et al. (2014) verwendet haben, um schwächere Veränderungen zu identifizieren (siehe Kapitel 3.2). Dies zeigen auch die Resultate aus Kapitel 4.2.3, in welchem die Veränderung in der Geschwindigkeit des zweiten CP analysiert wurde.

Obwohl die FPT die Tortuosität nicht in ihre Analyse miteinbeziehen kann, hat die Kombination von Geschwindigkeit und time-scale einen positiven Effekt auf die FPT. Dass die FPT in der Kombination besser abgeschnitten hat ist erfreulich, denn in realen Datensätzen sind Wechsel in nur einer *path*

DISKUSSION

characteristic selten. Beispielsweise nimmt die Tortuosität zu und die Geschwindigkeit ab, wenn sich ein Individuum in einem Bereich befindet, in dem es nach Nahrung oder Ruheplätzen sucht (Fauchald & Tveraa 2003). Dies ist nur ein Beispiel unter vielen. Denn die Verhaltensmuster eines Individuums haben unterschiedliche Eigenschaften, die durch verschiedene *path characteristics* beschrieben werden (Barraquand & Benhamou 2008). Doch um zu wissen, welche *path characteristics* für eine bestimmte Analyse sinnvoll sind, braucht die Nutzerin oder der Nutzer ein gutes Verständnis über das Individuum sowie klare Forschungsfragen (Edelhoff et al. 2016a).

Die BCPA hat sich in der Anzahl korrekt identifizierten CP sowie im Recall gegenüber der time-scale switch Simulation mit $K = 2$ verbessert (ohne Abweichung). Der Recall der speed-switch Simulation mit $K = 2$ ist derselbe. Die Precision ist ebenfalls ähnlich zu den beiden Simulationen mit $K = 2$. Auch die Precision sowie der Recall mit Abweichung ± 3 sind sehr ähnlich. Wird die Precision hingegen mit den beiden ursprünglichen Simulationen ($K = 1$) verglichen (Kapitel 5.3), hat sie sich stark verschlechtert. Das heisst, dass die Methode zwar mehr CP identifiziert hat, diese aber nicht korrekt identifiziert wurden. Dies zeigen auch die Resultate der Tabelle 23 (oben) und Tabelle 25 (oben). Mit der Berücksichtigung einer Abweichung von ± 3 werden zwar in ähnlich vielen Realisierungen die CP korrekt erkannt, jedoch werden auch zwischen den CP viel mehr identifiziert als mit $K = 1$, sowohl bei der speed switch wie auch bei der time-scale switch Simulation. Die Resultate aus Kapitel 5.3 zeigen, dass auch die BCPA den Recall gegenüber der time-scale switch deutlich verbessern konnte. Die speed-switch Simulation hat die Verwendung von $K = 2$ kaum beeinflusst. Aus diesen Resultaten wird ersichtlich, dass der Einfluss verschiedener Wertekombinationen für die BCPA relativ klein ist, wenn gewisse Werte ausgeschlossen werden (siehe Kapitel 5.5). Trotzdem sollen die Auswirkungen unterschiedlicher Wertekombinationen der Parameter immer analysiert werden.

Die BCPA schneidet auch in dieser Simulation besser ab als die FPT. Da die BCPA mit unterschiedlichen *path characteristics* verwendet werden kann (*persistence velocity*, *log der Geschwindigkeit*), hat sie grundsätzlich ein grösseres Potenzial als die FPT. Edelhoff et al. (2016a) haben beispielsweise das NSD (*net squared displacement*) als *path characteristic* verwendet. Dass die BCPA einerseits mehr Parameter hat, die vor der Analyse zu bestimmen sind und andererseits unterschiedliche *path characteristics* verwenden kann, hat zwar seine Vorteile, lässt die Segmentierung jedoch auch komplexer werden als bei einer FPT (Edelhoff et al. 2016a).

Solche Vergleiche einzelner Werte für die Parameter der Methoden sind in der Literatur bisher rar. Morelle et al. (2014, siehe Kapitel 6.5) sind bisher die einzigen, die einen Vergleich für die Parameterwerte der BCPA erstellten. Sie haben die Auswirkungen der unterschiedlichen Kombinationen auf die Anzahl Segmente analysiert. Die Anzahl Segmente wird in der Literatur oft verwendet, um

Resultate zu vergleichen (z. B. Soleymani et al. 2017). Dieses Kriterium kann jedoch zu falschen Interpretationen führen, da die korrekte Anzahl Segmente nicht unbedingt heissen muss, dass die CP auch am korrekten Punkt in der Trajektorie identifiziert wurden. Dies zeigen beispielsweise die Resultate in Kapitel 5.1. Sowohl in der time-scale switch Realisierung wie auch in der Realisierung der continuous speed switch Simulation hat die FPT die korrekte Anzahl Segmente identifiziert, die CP jedoch teilweise weit entfernt von den korrekten CP gesetzt.

Eine Schwierigkeit ist daher, ein geeignetes Kriterium zu finden, um die Korrektheit der Segmentierung an realen Datensätzen zu beurteilen. Diesbezüglich können die Evaluationskriterien dieses Projekts (siehe Kapitel 4.1.4) für weitere Analysen mit Simulationen oder Datensätzen mit bekannten Verhaltensmustern verwendet werden. Für reale Datensätze ohne bekannte Verhaltensmuster (ein grosser Prozentsatz der verfügbaren Daten) sind diese Evaluationskriterien eher ungeeignet. Es wurden jedoch bisher noch keine Publikation veröffentlicht, in der die Resultate des Segmentierungsprozesses statistisch ausgewertet wurden. Einen Ansatz dazu liefern einzig Le Corre et al. (2014), indem sie die Resultate mit den zu erwartenden Resultaten verglichen. Der Segmentierungsprozess wird meist visuell analysiert (wie in Gurarie et al. 2016a) oder es wird in einem zweiten Schritt eine Klassierung vorgenommen, welche dann mit Evaluationskriterien wie der Precision oder dem Recall evaluiert wird. Beispielsweise haben Zhang et al. (2015) die BCPA mit einer k-Means Clusteranalyse kombiniert und die Resultate der Clusteranalyse statistisch ausgewertet. Schlussendlich, dies haben sowohl dieses Projekt als auch Gurarie et al (2016) gezeigt, kommt man nicht darum herum jede Segmentierung auch visuell zu betrachten und unter Einbezug des ökologischen Hintergrundwissens zum untersuchten Individuum zu evaluieren.

6.7 Forschungsfragen

Forschungsfrage I Inwiefern sind die First Passage Time und die Behavioural Change Point Analysis für eine Ökologin oder einen Ökologen von Nutzen? Welche Arten von Forschungsfragen lassen sich durch die beiden Segmentierungsmethoden beantworten?

Die Resultate dieses Projekts zeigen, dass sowohl die FPT wie auch die BCPA einen ersten Überblick verschaffen können bezüglich der Zeitpunkte von Veränderung in den Verhaltensmustern eines Individuums. Mithilfe beider Methoden ist es möglich, anhand von spezifischen Eigenschaften der Trajektorie die Wechsel im Verhalten zu identifizieren und sie räumlich und zeitlich zu verorten. Dieses Wissen aus der Trajektorie zu extrahieren kann für eine Ökologin oder einen Ökologen wertvolle Informationen enthalten, die ihnen helfen, ihre Forschungsfragen zu beantworten. Wie in Kapitel 2.1.1 beschrieben wird, interessiert sich eine Ökologin oder ein Ökologe in der Bewegungsökologie für Fragen zum wieso, wie, wann und wohin einer Bewegung eines Individuums sowie den beeinflussenden externen

DISKUSSION

Faktoren (Nathan et al. 2008). Durch die beiden Segmentierungsmethoden lassen sich Fragen beantworten, deren hauptsächliches Interesse der Zeitpunkt (wann?) und der Ort (wohin?) ist. Beispielsweise können Fragen zum Beginn oder zum Ende einer Migration, zur Dauer derselben oder zu Unterschieden zwischen einzelnen Individuen oder zwischen einzelnen Jahren beantwortet werden. Durch die Verortung im Raum können Migrationsrouten nachvollzogen werden, inklusive Zwischenstopps. Um die Fragen zum wieso, wie und den externen Faktoren zu beantworten, müssten weitere Informationen vorhanden sein beziehungsweise Methoden verwendet werden, die der Bewegung zugrundeliegende Prozesse erkennen und analysieren können (z. B. SSM).

Die FPT und die BCPA bieten einer Ökologin oder einem Ökologen eine Hilfestellung, Fragen bezüglich des Zeitpunktes und des Ortes einer Bewegung eines Individuums zu beantworten.

Forschungsfrage II Wie robust gestalten sich die Methoden gegenüber einem Rauschen in der Simulation? Wie genau können die Methoden einen Change Point identifizieren? Welchen Einfluss haben die Werte der Parameter der jeweiligen Methode sowie die zu identifizierenden *path characteristics*?

Aus den Resultaten der Evaluationskriterien (siehe Kapitel 4.1.4) lässt sich sagen, dass beide Methoden oftmals starke Schwankungen in der Anzahl der identifizierten CP, der Anzahl korrekt identifizierten CP, der Precision und dem Recall aufweisen. Es lässt sich aus den Resultaten jeweils eine Tendenz ablesen, die jedoch sowohl besser als auch schlechter ausfallen kann. Grundsätzlich kann gesagt werden, dass die FPT starke Veränderungen in der Geschwindigkeit in nahezu allen Fällen erkennen kann, während sie in schwächeren oder kontinuierlichen Veränderungen der Geschwindigkeit sowie in Veränderungen der Tortuosität meist versagt. Die BCPA hingegen kann sowohl starke als auch schwächere Veränderungen in demselben Durchgang identifizieren, ebenso Veränderungen in der Tortuosität erkennen. Bei kontinuierlichen Veränderungen versagt jedoch auch die BCPA. Unter Berücksichtigung dieser Punkte kann gesagt werden, dass beide Methoden selten einen CP exakt identifizieren können. Wird jedoch bereits eine geringe Abweichung berücksichtigt, lassen sich die meisten Resultate stark verbessern. Mit einer noch grösseren Abweichung können die Resultate erneut gesteigert werden. Während die FPT in der speed switch Simulation nie mehr als zwei Drittel der korrekten CP pro Realisierung korrekt identifizieren konnte, sind es bei der BCPA mit nur wenigen Ausnahmen alle CP. Nur in der kombinierten Simulation (mit veränderter Geschwindigkeit und Tortuosität) konnten diese Resultate noch etwas verbessert werden, indem ein grösserer Anteil der Realisierungen gute Resultate erzielte.

Es hat sich gezeigt, dass die unterschiedlichen Kombinationen in den Parametern der Methoden Einfluss darauf haben können, ob 50% der Realisierungen einen CP mehr oder weniger korrekt identifizieren. Dies sind jedoch Resultate, die für jeden Datensatz erneut getestet werden sollten, da die Eigenschaften einer Trajektorie die Wertekombinationen der Parameter beeinflussen. Die Verwendung der

DISKUSSION

unterschiedlichen *path characteristics* in der Simulation (mittlere Geschwindigkeit und Tortuosität) wurden durch die der Methode zugrundeliegenden *path characteristics* (Geschwindigkeit und *persistence velocity*) unterschiedlich gut erkannt. Wie bereits Gurarie et al. (2016a) festgestellt haben, versagt die FPT in der time-scale switch Simulation während die BCPA hier passable Resultate lieferte.

Zusammenfassend kann also gesagt werden, dass beide Methoden eine gewisse Tendenz zu den jeweiligen Resultaten aufweisen. Diese kann mit Berücksichtigung einer maximalen Abweichung verstärkt werden. Eine Ökologin oder ein Ökologe müssen sich bewusst sein, welche *path characteristic* einer Methode zugrunde liegt. Wird dies in der Analyse beachtet, ist es durchaus möglich, gute Resultate zu generieren.

Forschungsfrage III Welche Stärken und Schwächen lassen sich aus den Resultaten der Segmentierung für die jeweilige Methode ableiten?

Wie bereits in den anderen beiden Antworten zu den Forschungsfragen ausgeführt, liegt eine Stärke der FPT darin, starke Veränderungen in der Geschwindigkeit zu identifizieren. Des Weiteren bietet sie die Möglichkeit, Segmente in einem zweiten Schritt erneut zu segmentieren, um auch schwächere Veränderungen zu erkennen. Die FPT findet tendenziell weniger oder gleich viele CP wie korrekt wären. Zusätzlich braucht sie mindestens eine geringe Abweichung, um einen CP korrekt zu identifizieren. Mit nur einem zu bestimmenden Parameter (Radius) ist sie zudem sehr schnell und einfach anzuwenden. Jedoch braucht die FPT einen zusätzlichen Schritt, bis die segmentierte Trajektorie zur Sichtung bereit ist (z. B. Segmentierung mit Lavielle 2005). Der Output der FPT ist dann wiederum relativ einfach zu interpretieren. Eine Schwäche der FPT ist, dass sie einzig Veränderungen in der Geschwindigkeit erkennen kann. Falls diese unterschiedlich stark sind, erkennt sie nur diejenigen mit dem grössten Kontrast.

Im Gegensatz zur FPT kann die BCPA auch Veränderungen in der Tortuosität erkennen. Sie ist tendenziell sensitiver gegenüber Veränderungen. Dies führt jedoch meist zu mehr als doppelt so vielen CP als korrekt wären. Wird eine eher grössere Abweichung zugelassen, identifiziert die BCPA nahezu 100% der korrekten CP. Eine Ausnahme bilden die kontinuierlichen Veränderungen in der Geschwindigkeit. Mit diesen hat die BCPA ebenfalls Mühe. Um die BCPA anzuwenden, braucht es ein gewisses Vorwissen bezüglich der Wertezusammensetzung der einzelnen Parameter, ebenso bei der Interpretation des Outputs. Aus dem Output der BCPA lassen sich aber grundsätzlich mehr Informationen ableiten als aus demjenigen der FPT.

Zusammenfassend kann gesagt werden, dass die FPT die simplere Methode ist um Änderungen in der Geschwindigkeit zu erkennen, jedoch weniger Informationen in ihrem Output beinhaltet. Die BCPA

hingegen ist komplizierter, liefert dementsprechend aber mehr Informationen, auch bezüglich Veränderungen in der Tortuosität. Bei einer Fragestellung mit realen Bewegungsdaten lohnt es sich allenfalls, erst die FTP auf die Daten anzuwenden, um mit relativ geringem Aufwand ein Gefühl für mögliche Resultate zu erhalten. Bleiben dann noch weiterführende Fragen offen, kann zusätzlich die BCPA angewendet werden um so weitere und detailliertere Informationen zu erhalten.

6.8 Kritische Beurteilung des Projekts

Dieses Projekt basiert zu einem grossen Teil auf der vergleichenden Literatur von Gurarie et al. (2016a). Da der erste Teil des Projekts die Replizierung der Resultate von Gurarie et al. (2016a) beinhaltet, wurden ihre Erkenntnisse aus der Publikation extrahiert und mit den Resultaten dieses Projekts verglichen. Viele der Erkenntnisse basieren auf den mathematischen Grundlagen der Simulationen sowie der Methoden und sind daher unveränderlich. Trotzdem konnte eine weitere Realisierung zusätzliche Informationen bereitstellen.

Da die Simulation ein Rauschen beinhaltete, wurde die Analyse mit jeweils 50 Realisierungen durchgeführt. Ob diese Zahl jedoch genügend gross ist um tatsächlich aussagekräftig zu sein, wurde nicht überprüft. Diese Zahl wurde auch aus zeitlichen Gründen gewählt, da eine erhöhte Anzahl automatisch rechnerisch zu längeren Zeiten führen würde. Es müsste also überprüft werden, ob eine grössere Anzahl an Realisierungen die Resultate stark verändern würde oder ob es keinen grossen Einfluss hätte.

Des Weiteren wurde in diesem Projekt nur ein einziger Simulationstyp mit unterschiedlichen *path characteristics* verwendet. Wie die Resultate mit anderen Simulationen aussehen würden, wurde in diesem Projekt nicht geprüft. Andere Simulationen könnten beispielsweise unterschiedliche *sampling frequencies*, Datenlücken, kein zufälliges Rauschen oder Veränderungen in anderen *path characteristics* beinhalten. Solche Eigenschaften von Datensätzen müssen unbedingt ebenfalls analysiert werden, um zusätzliches Wissen zu generieren. Zudem müssen die gewonnenen Erkenntnisse müssten nun auch auf reale Datensätze angewendet werden. Genau dieser Punkt bereitet jedoch die allergrössten Schwierigkeiten, wird durch unzählige Faktoren beeinflusst und konnte innerhalb des zeitlichen Rahmens dieses Projekts nicht weiter behandelt werden. Empfohlen würde, mit einem Datensatz mit bekannten Verhaltensmustern zu beginnen und diese Erkenntnisse auf Datensätze ohne bekannte Verhaltensmuster anzuwenden.

Die Resultate der Evaluationskriterien wurden zwar untereinander verglichen, jedoch wurden keine statistischen Tests durchgeführt, um beispielsweise die Signifikanz einer Werteveränderung eines Parameters zu prüfen. Diese Informationen hätten allenfalls einen noch tieferen Einblick in die Leistung der Methoden gegeben. Die hauptsächliche Schwierigkeit liegt jedoch darin, solche statistischen Analysen

DISKUSSION

auch an realen Datensätzen durchzuführen. Bei diesen gibt es keine bekannten und korrekten Informationen anhand welcher die Analyse hätte überprüft werden können. Aus diesem Grund wurden auch die Methoden in den Publikationen von Gurarie et al. (2016a) und Edelhoff et al. (2016a) nur visuell verglichen.

Eine weitere Einschränkung dieses Projekts ist die fehlende Projektion in den Daten. Anstatt eines geografischen Koordinatensystems wird hier nur ein kartesisches Koordinatensystem verwendet. Dieses lässt sich natürlich nicht eins zu eins auf reale Daten übertragen. Dadurch sind nur schwammige Formulierungen wie «kleine Abweichung» möglich. Was dies bei realen Daten schliesslich heisst, muss je nach Datensatz und Fragestellung erneut geprüft werden.

Des Weiteren konnte im Umfang dieses Projekts die Parameterwahl aus den Kapiteln 6.4, 6.5 und 6.6 nur exemplarisch durchgeführt werden. Insbesondere bezüglich Anwendung auf reale Datensätze sollten hier noch weitere und differenziertere Analysen durchgeführt werden. Hierzu müsste in einem ersten Schritt das gleiche Vorgehen auf gut bekannte Datensätze angewendet werden und ein Evaluationskriterium gefunden werden, anhand dessen weitere Wertevergleiche der Parameter durchgeführt werden könnten. Allenfalls würden auch weitere statistische Analysen zusätzliche Informationen liefern, welche dann besser verallgemeinert werden könnten.

Das Thema der räumlichen und zeitlichen Autokorrelation spielt in der Bewegungsanalyse eine wichtige Rolle. Da dieses Thema jedoch schon in Gurarie et al. (2016a) behandelt wurde und in diesem Projekt dieselben Methoden wie auch Simulationen verwendet wurden, wurden hier keine weiteren Analysen zur Autokorrelation durchgeführt. Es ist jedoch ein Aspekt, der in der Bewegungsanalyse immer beachtet werden sollte.

7 FAZIT

Die Segmentierung einer Trajektorie durch eine geeignete Methode stellt eine grosse Herausforderung dar. Aus diesem Grund war das Ziel dieser Arbeit, zwei Segmentierungsmethoden in Bezug auf zwei unterschiedliche Bewegungscharakteristiken zu vergleichen. Diese beiden Bewegungscharakteristiken sind die mittlere Geschwindigkeit und die Tortuosität. Gurarie et al. (2016a) betonen wie wichtig es sei, verschiedene Methoden zu vergleichen und ihre Eigenschaften zu identifizieren. Da bisher nur zwei vergleichende Publikationen (Gurarie et al. 2016a und Edelhoff et al. 2016a) veröffentlicht und die Resultate darin nur visuell verglichen wurden, nimmt dieses Projekt bezüglich der statistischen Vergleiche eine Vorreiterrolle ein, auf welcher die zukünftige Forschung aufbauen kann. Somit widmet sich dieses Projekt dem statistischen Vergleich zweier Methoden, um ihre Stärken und Schwächen zu erkennen, geeignete Werte für die Parameter dieser Methoden auszuwählen und ihre Leistung bezüglich des zufälligen Rauschens in der Trajektorie zu analysieren.

Der erste Schritt in diesem Vergleich bestand aus der Replizierung der Resultate von Gurarie et al. (2016a) und Soleymani et al. (2017). Dabei wurden die speed switch, time-scale switch und die continuous speed switch Simulationen durch die beiden Methoden FPT und BCPA segmentiert. Anhand dieser Resultate konnten die Aussagen der beiden Publikationen hauptsächlich bestätigt, oft ergänzt und selten widerlegt werden. Dies zeigt, dass eine Methode vergleichbare Resultate liefert, obwohl die Trajektorie etwas anders aussah. Aus den Erkenntnissen dieses ersten Schritts konnten die nächsten Schritte entwickelt werden.

Da die Simulation ein zufälliges Rauschen beinhaltet, sieht jede einzelne Realisierung der Simulation wieder anders aus. Aus diesem Grund wurde jede der drei Simulationen jeweils 50 Mal produziert und zusammenfassende statistische Masse (Anzahl identifizierte CP, Anzahl korrekt identifizierte CP, Precision und Recall) wurden berechnet. Die Resultate zeigten, dass beide Methoden bezüglich der Anzahl identifizierten CP starke Schwankungen aufwiesen. Die FPT hatte tendenziell weniger, die BCPA dafür fast doppelt so viele CP identifiziert als eigentlich korrekt wären. Mit Ausnahme der durch die BCPA segmentierten speed switch Simulation, konnte keines der statistischen Masse (Anzahl korrekt identifizierte CP, Precision und Recall) einen Median höher als 0 aufweisen. Daraus folgte, dass den Methoden eine gewisse Abweichung zugestanden werden muss, damit sinnvolle Resultate erreicht werden konnten.

Da die identifizierten CP selten exakt auf dem korrekten CP lokalisiert wurden, wurden drei verschiedene maximale Abweichungen beachtet. Die drei Abweichungen betrug ± 1 , ± 2 und ± 3 Datenpunkte Entfernung vom korrekten CP. So konnte gezeigt werden, dass oftmals bereits eine Abweichung von ± 1

FAZIT

einen grossen Einfluss auf die Resultate hat und mit der Abweichung von ± 3 bereits 100% der korrekten CP identifiziert werden konnten. Die BCPA erreichte dabei höhere Werte als die FPT, da letztere den mittleren CP innerhalb der speed switch Simulation nicht erkennen konnte und sowohl in der time-scale switch Simulation wie auch in der continuous speed switch Simulation unbrauchbare Resultate lieferte. Der Einbezug einer maximalen Abweichung ist bei der Anwendung der beiden Methoden daher unerlässlich. Wie gross dieser tatsächlich sein soll, konnte im Rahmen dieses Projekts nicht abschliessend analysiert werden. Jedoch wurde gezeigt, dass bereits die Berücksichtigung dreier Datenpunkte vor- bzw. nach dem korrekten CP ein sehr gutes Resultat liefern.

Da, wie bereits erwähnt, die FPT in der speed switch Simulation den mittleren CP (schwächster CP) nicht identifizieren konnte, wurde in einem nächsten Schritt diese Veränderung in der Geschwindigkeit separat analysiert und einen Vergleich aufgrund unterschiedlicher Radien gezogen. Dies hat gezeigt, dass die FPT in der Lage ist, auch schwächere Veränderungen zu identifizieren, sofern diese separat analysiert werden. Pinaud (2008) empfiehlt dieses Vorgehen und Le Corre et al. (2014) haben es in ihrer Publikation ebenfalls angewendet. Die Grösse des Radius hatte in dieser Simulation nur einen geringen Einfluss. In einem realen Datensatz kann der Radius aufgrund der Massstabsabhängigkeit der Bewegungsdaten jedoch eine entscheidende Rolle dafür spielen, ob die Resultate brauchbar sind oder nicht.

Da die BCPA einige Parameter hat, welche die Nutzerin oder der Nutzer im Voraus bestimmen müssen, wurden unterschiedliche Wertezusammensetzungen getestet und analysiert. Dies hat gezeigt, dass es gewisse Kombinationen gibt, die für diese Simulationen nicht empfehlenswert sind. Es gibt jedoch auch mehrere Kombinationen, bei denen sich die Resultate kaum unterscheiden lassen. Dies hat gezeigt, dass es wichtig ist, eine solche Analyse vor jeder Anwendung der BCPA durchzuführen, da aufgrund verschiedener Eigenschaften der Trajektorien andere Wertekombinationen sinnvoll sein können.

Aus den beiden Analysen zu den Werten der Parameter (Radius bei der FPT und *window size*, *cluster width* und Sensitivitätsparameter *K* bei der BCPA) wurde jeweils ein Wert bzw. eine Kombination der Werte ausgewählt und auf eine Simulation angewendet, die sowohl in der Geschwindigkeit wie auch in der Tortuosität (speed & time-scale switch Simulation) variierte. Die Werte der Evaluationskriterien haben dann gezeigt, dass sich die FPT gegenüber beiden separaten Simulationen verbessern konnte. Die BCPA hat sich gegenüber der time-scale switch Simulation verbessert und lieferte ähnliche Resultate wie bei der speed switch Simulation. Diese Simulation liegt am nächsten an der Realität.

Zusammenfassend kann gesagt werden, dass beide Methoden starke Geschwindigkeitsänderungen relativ genau identifizieren können, während die BCPA auch schwächere Veränderungen innerhalb derselben Segmentierung erkennen kann. Die FPT kann diese ebenfalls erkennen, braucht jedoch einen separaten Schritt dazu, in dem nur ein einzelnes Segment erneut segmentiert wird. Variiert die Tortuosität in einer

Trajektorie, liefert die FPT keine brauchbaren Informationen, die BCPA kann die CP hingegen korrekt identifizieren. Wird nach kontinuierlichen Verhaltensänderungen gesucht, versagen beide Methoden. Allgemein können die Resultate beider Methoden verbessert werden, wenn eine maximale Abweichung mit einbezogen wird. Welche Werte dabei Sinn machen, muss jeweils aufgrund der Grundlagedaten und der spezifischen Fragestellung entschieden werden.

7.1 Ausblick

Diese Arbeit stellt einen Ausgangspunkt für die weitere Forschung bezüglich statistischer Auswertungen in der Bewegungsanalyse dar. Des Weiteren bietet sie Informationen betreffend Vorgehen und Auswertung beim Vergleich von Segmentierungsmethoden. Aufgrund der Komplexität der Bewegungsanalyse ergeben sich unzählige Möglichkeiten zur weiteren Forschung, die Ökologinnen und Ökologen helfen können, geeignete Methoden für ihren Datensatz und ihre Forschungsfragen auszuwählen.

Aus der kritischen Beurteilung dieses Projekts wird klar, dass verschiedene Aspekte verbessert und insbesondere auch erweitert werden können. Die erneute Durchführung der Analyse unter Verwendung von anderen Segmentierungsmethoden, Simulationen und vor allem von realen Datensätzen ist notwendig, um der Breite und Komplexität der Bewegungsanalyse gerecht zu werden.

Die Analyse könnte beispielsweise erneut durchgeführt werden, jedoch mit einer anderen Anzahl von Realisierungen. Dies gäbe einen Anhaltspunkt dazu, wie zuverlässig die Resultate tatsächlich sind und ob die in diesem Projekt ausgearbeiteten Schlussfolgerungen bestätigt werden können.

Um die Aussagen dieses Projekts zudem zu verfeinern, wäre es interessant, wenn die Simulationen unterschiedliche *sampling frequencies* beinhalten würden, geprägt wären durch unregelmässige Zeitintervalle mit grösseren Lücken oder keinem bzw. grösserem Rauschen. Denn reale Daten werden nie frei von solchen Unregelmässigkeiten sein. Optimalerweise sollten die Simulationen auch ein Individuum repräsentieren, wie dies beispielsweise bei den Daten in Edelhoff et al. (2016a) der Fall war. Dabei soll das Verhalten auf den Tag-Nacht-Rhythmus abgestimmt sein und sich auf unterschiedlichen Massstäben abspielen. In diesem Zusammenhang sollte man sich auch detaillierter mit der räumlichen und zeitlichen Autokorrelation auseinandersetzen und deren Einfluss genauer untersuchen. Dazu würde auch gehören, dass die Koordinaten der Trajektorie nicht mehr in einem kartesischen Koordinatensystem analysiert werden, sondern in einem geographischen. Alle diese Eigenschaften unterschiedlich zu kombinieren, und zu analysieren, wie sich die Resultate der Segmentierung dabei verändern, wäre in der weiteren Forschung sinnvoll.

FAZIT

Ein Schwerpunkt in diesem Projekt war die statistische Analyse. Sie bietet grosses Potenzial, ist jedoch erst ein Anfang und braucht noch viel Forschung. Der wichtigste Punkt, der zurzeit erforscht werden sollte, ist die Anwendung der Evaluation auf reale Daten, in denen man keine «korrekten» Change Points kennt. Es sollte also analysiert werden, welche Evaluationskriterien sinnvoll sind und wie diese zuverlässig auf reale Daten angewandt werden können. Um diesen Punkt zu erarbeiten, ist eine intensive Zusammenarbeit von Statistikerinnen/Statistikern und Ökologinnen/Ökologen zwingend.

Diese Vielfalt an Simulationen und der Einbezug unterschiedlicher Evaluationskriterien wird dabei helfen, auch weitere Methoden besser zu analysieren, zu vergleichen und zu verstehen, welche Stärken und Schwächen jede einzelne Methode mit sich bringt. Dadurch kann erforscht werden, welche ökologischen Forschungsfragen am besten mit welcher Methode analysiert werden können. So wird auch die Komplexität dieses Fachgebietes Schritt für Schritt abnehmen und es können schneller bessere Resultate erzielt werden. Dies kommt dem gesamten Verständnis der Ökologie zu Gute und wird eine grosse Menge an neuen Erkenntnissen betreffend dem Verhalten unterschiedlicher Tierarten innerhalb ihrer vielfältigen und sich stetig verändernden Lebensräume hervorbringen.

8 LITERATUR

- Bailey, H. & Thompson, P. (2006): Quantitative analysis of bottlenose dolphin movement patterns and the relationship with foraging. *Journal of Animal Ecology*, 75(2), 456–465.
- Barraquand, F. & Benhamou, S. (2008): Animal movements in heterogeneous landscapes: identifying profitable places and homogeneous movement bouts. *Ecology*, 89(12), 3336–3348.
- Becker, J., Chan, C., Chan, G., Leeper, T.J., Gandrud, C., MacDonald, A., Zahn, I., Stadlmann, S. & Williamson, R. (2018): rio: A swiss-army knife for data I/O. R package version 0.5.10.
- Buchin, M., Driemel, A., van Kreveld, M. & Sacristán (2011): Segmenting trajectories: a framework and algorithms using spatiotemporal criteria. *Journal of Spatial Information Science*, 3, 33–63.
- Buchin, M., Kruckenberg, H. & Kölzsch, A. (2013): Segmenting trajectories by movement states. In: Timpf, S. & Laube, P. *Advances in spatial data handling. Advances in Geographic Information Science*, 15–25.
- Cagnacci, F., Boitani, L., Powell, R.A., Boyce, M.S. (2010): Animal ecology meets GPS-based radiotelemetry: a perfect storm of opportunities and challenges. *Philosophical Transactions of the Royal Society*, 365, 2157–2162.
- Calenge, C. (2006): The package “adehabitat” for R software: a tool for the analysis of space and habitat use by animals.
- Calenge, C., Dray, S. & Royer-Carenzi, M. (2009a): The concept of animals’ trajectories from a data analysis perspective. *Ecological Informatics*, 4, 34–41.
- Calenge, C. (2009b): adehabitatLT: Analysis of animal movements. R package version 0.3.23.
- Codling, E.A., Plank M.J. & Benhamou, S. (2008): Random walk models in biology. *Journal of the Royal Society Interface* 5, 813–834.
- Demšar, U., Buchin, K., Cagnacci, F., Sai, K., Speckmann, B., Van de Weghe, N., Weiskopf, D. & Weibel, R. (2015): Analysis and visualisation of movement: an interdisciplinary review. *Movement Ecology*, 3(5), 1–24.
- Dodge, S., Weibel, R. & Lautenschütz, A-K. (2008): Towards a taxonomy of movement patterns. *Information Visualisation*, 1–13.
- Dodge, S., Weibel, R., Ahearn, S.C., Buchin, M. & Miller, J.A. (2016): Analysis of movement data. *International Journal of Geographical Information Science*, 30(5), 825–834.
- Dray, S., Royer-Carenzi, M. & Calenge, C. (2010): The exploratory analysis of autocorrelation in animal-movement studies. *Ecological Research*, 25, 673–681.
- Edelhoff, H., Signer, J. & Balkenhol, N. (2016a): Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. *Movement Ecology*, 4(21), 1–21.

LITERATUR

- Edelhoff, H., Signer, J. & Balkenhol, N. (2016b): Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. Supplement 1. *Movement Ecology*, 4(21), 1–21.
- Edelhoff, H., Signer, J. & Balkenhol, N. (2016c): Path segmentation for beginners: an overview of current methods for detecting changes in animal movement patterns. Supplement 2. *Movement Ecology*, 4(21), 1–21.
- Fauchald, P. & Tveraa, T. (2003): Using First-Passage-Time in the analysis of area-restricted search and habitat selection. *Ecology*, 84(2), 282–288.
- Finch, S. (2004): Ornstein-Uhlenbeck Process. 1–14.
- Fleming, C.H., Calabrese, J.M., Mueller, T., Olson, K.A., Leimgruber, P. & Fagan, W.F. (2014): From fine-scale foraging to home ranges: a semivariance approach to identifying movement modes across spatiotemporal scales. *The American Naturalist*, 183(5), E154–E167.
- Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J. & Pedrotti, L. (2010): Resolving issues of imprecise and habitat-biased locations in ecological analyses using GPS telemetry data. *Philosophical transactions of the royal society*, 365, 2187–2200.
- Frair, J.L., Merrill, E.H., Visscher, D.R., Fortin, D., Beyer, H.L. & Morales, J.M. (2005): Scales of movement by elk (*Cervus elaphus*) in response to heterogeneity in forage resources and predation risk. *Landscape Ecology*. 20, 273–287.
- Fritz, H., Said, S. & Weimerskirch, H. (2003): Scale-dependent hierarchical adjustments of movement patterns in a long-range foraging seabird. *The royal society*, 270, 1143–1148.
- Garstang, M., Davis, R.E., Leggett, K., Frauenfeld, O.W., Greco, S., Zipser, E. & Peterson, M. (2014): Response of African elephants (*Loxodonta africana*) to seasonal changes in rainfall. *PLoS ONE*, 9(10), 1–13.
- Gurarie E., Russel, D., Andrews, R.D., Laidre, K.L. (2009a): A novel method for identifying behavioural changes in animal movement data. *Ecology Letters*, 12, 395–408.
- Gurarie E., Russel, D., Andrews, R.D., Laidre, K.L. (2009b): A novel method for identifying behavioural changes in animal movement data. Supplementary Information S1. *Ecology Letters*, 12, 395–408.
- Gurarie, E. & Ovaskainen, O. (2011): Characteristic spatial and temporal scales unify models of animal movement. *The americal naturalist*, 1778(1), 113–123.
- Gurarie, E. & Ovaskainen, O. (2013): Towards a general formalization of encounter rates in ecology. *Theoretical Ecology*, 6(2), 189–202.
- Guarie, E. (2014): bcpa: Behavioural Change Point Analysis in R: The bcpa package. R package version 1.1.
- Gurarie, E., Bracis, C., Delgado, M., Meckley, T.D., Kojola, I. & Wagner, C.M. (2016a): What is the animal doing? Tools for exploring behavioural structure in animal movements. *Journal of animal ecology*, 85, 69–84.

LITERATUR

- Gurarie, E., Bracis, C., Delgado, M., Meckley, T.D., Kojola, I. & Wagner, C.M. (2016b): What is the animal doing? Tools for exploring behavioural structure in animal movements. Supplementary Materials 1. *Journal of animal ecology*, 85, 69–84.
- Gurarie, E., Bracis, C., Delgado, M., Meckley, T.D., Kojola, I. & Wagner, C.M. (2016c): What is the animal doing? Tools for exploring behavioural structure in animal movements. Waddle. R package. *Journal of animal ecology*, 85, 69–84.
- Hebblewhite, M. & Haydon, D.T. (2010): Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. *Philosophical Transactions of the Royal Society*, 356, 2303–2312.
- Holden, C. (2006): Inching toward movement ecology. *Science*, 313(5788), 779–782.
- Holloway, P. & Miller, J.A. (2018): Analysis and modeling of movement. In: Huang, B: *Comprehensive geographic information systems. GIS methods and techniques*, Bd. 1, 162–178.
- Isermann, R. (1974): *Prozessidentifikation. Identifikation und Parameterschätzung dynamischer Prozesse mit diskreten Signalen*. Springer-Verlag.
- Johnson, A.R., Wiens, J.A., Milne, B.T. & Crist, T.O. (1992): Animal movements and population dynamics in heterogeneous landscapes. *Landscape Ecology*, 7(1), 63–75.
- Johnson, D.S., London, J.M., Lea, M-A. & Durban, J.W. (2008): Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5), 1208–1215.
- Kranstauber, B., Cameron, A., Weinzerl, R., Fountain, T., Tilak, S., Wikelski, M. & Kays, R. (2011): The movebank data model for animal tracking. *Environmental Modelling & Software*, 26, 834–835.
- Laidre, K.L., Heide-Lørgensen, M.P., Logsdon, M.L., Hobbs, R.C., Dietz, R. & VanBlaricom, G.R. (2004): Fractal analysis of narwhal space use patterns. *Zoology* 107, 3–11.
- Lavielle, M. (2005): Using penalized contrasts for the change-point problem. *Signal Processing*, 85, 1501–1510.
- Le Corre, M., Dussault, C. & Côté, S.D. (2014): Detecting changes in the annual movements of terrestrial migratory species: using the first-passage time to document the spring migration of caribou. *Movement Ecology* 2(19), 1–11.
- Max Planck Institute for Ornithology (2018): About Movebank. Movebank.org, Zugriff: 1.9.2018
- McKenzie, H.W., Lewis, M.A. & Merrill, A.H. (2009): First Passage Time analysis of animal movement and insights into the functional response. *Bulletin of mathematical biology*, 71, 107–129.
- Morelle, K., Bunnefeld, N., Lejeune, P. & Oswald, S.A. (2017a): From animal tracks to fine-scale movement modes: a straightforward approach for identifying multiple spatial movement patterns. *Methods of Ecology and Evolution*, 8, 1488–1498.
- Morelle, K., Bunnefeld, N., Lejeune, P. & Oswald, S.A. (2017b): From animal tracks to fine-scale movement modes: a straightforward approach for identifying multiple spatial movement patterns. Appendix S1. *Methods of Ecology and Evolution*, 8, 1488–1498.

LITERATUR

Mueller, T. & Fagan, W.F. (2008): Search and navigation in dynamic environments – from individual behaviors to population distributions. *Oikos*, 117, 654–664.

Nams, V.O. & Bourgeois, M. (2004): Fractal analysis measures habitat use at different spatial scales: an example with American marten. *Canadian Journal of Zoology*, 82, 1738–1747.

Nathan, R., Getz, W.M., Revilla, E., Holyoak, M., Kadmon, R., Saltz, D. & Smouse, P.E. (2008): A movement ecology paradigm for unifying organismal movement research. *PNAS*, 105(49), 19052–19059.

Pinaud, D. (2008): Quantifying search effort of moving animals at several spatial scales using first-passage time analysis: effect of the structure of environment and tracking systems. *Journal of Applied Ecology*, 45, 91–99.

Pinaud, D. & Weimerskirch, H. (2007): At-sea distribution and scale-dependent foraging behavior of petrels and albatrosses: a comparative study. *Journal of Animal Ecology*, 76, 9–19.

Prepicean, A.C., Gidofalvi, G. & Susilo, Y.O. (2016): Measures of transportation mode segmentation of trajectories. *International Journal of Geographical Information Science*, 30(9): 1763–1784.

R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.

Soleymani, A., Pennekamp, F., Dodge, S. & Weibel, R. (2017): Characterizing change points and continuous transitions in movebank behaviours using wavelet decomposition. *Methods in Ecology and Evolution*, 8(9), 1–11.

Soleymani, A. (2016): Cross-scale analysis in classification and segmentation of movement. Dissertation. MNF UZH.

Soleymani, A., Cachat, J., Robinson, K., Dodge, S., Kalueff, A.V. & Weibel, R. (2014): Integrating cross-scale analysis in the spatial and temporal domains for classification of behavioral movement. *Journal of Spatial Information Science*, 8, 1–25.

Technitis, G., Othman, W., Safi, K. & Weibel, R. (2015): From A to B, randomly: a point-to-point random trajectory generator for animal movement. *International Journal of Geographical Information Science*, 29(6), 912–934.

Teimouri M., Indahl, U.G., Sickel, H. & Tveite, H. (2018): Deriving animal movement behaviour using movement parameters extracted from location data. *International Journal of Geo-Information*, 7(2), 1–13.

Thiebault, A. & Tremblay, Y. (2013): Splitting animal trajectories into fine-scale behaviourally consistent movement units: breaking points relate to external stimuli in a foraging seabird. *Behav Ecol Sociobiol*, 67, 1013–1026.

Tomkiewicz, S.M., Fuller, M.R., Kie, J.G. & Bates, K.K. (2010): *Philosophical Transactions of the Royal Society*, 365, 2163–2176.

Trepl, L. (2007): *Allgemeine Ökologie. Band 2. Population*. Lang, Peter GmbH.

van Moorter, B., Visscher, D.R., Jerde, C.L., Frair, J.L. & Merrill, E.H. (2010): Identifying movement states from location data using cluster analysis. *Journal of Wildlife Management*, 74(3), 588–594.

LITERATUR

Warnes, G., Bolker, B., Bonebakker, L., Gentleman, R. Huber, W., Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M. & Venables, B. (2016): gplot: Various R programming tools for plotting data. R package version 3.0.1.

Wickham, H. (2016): plyr: Tools for splitting, applying and combining data. R package version 1.8.4.

Zhang, J., O'Reilly, K.M., Perry, G.L.W., Taylor, G.A. & Dennis, T.E. (2015): Extending the functionality of behavioural change-point analysis with k-means clustering: a case study with the little penguin (*Eudyptula minor*). PLoS ONE, 10(4), 1–9.

Zheng, Y. (2015): Trajectory data mining: An overview. ACM Transactions on Intelligent Systems and Technology, 6(3), 29:1–29:41.

Table 2 Characteristics of the methodological approaches for the three different categories of research questions. Different methods for answering the three type of broad research questions (study aims) are listed together with the analytical category they stem from, a short description of each method as well as the considered categories of input path-signals and important references

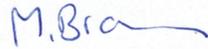
Study aim	Method	Analytical category	Description	Input signal	References
Movement pattern description	Thresholding	Topology-based	Applies thresholding schemes (cut-off values) to separate relocations into different groups based on single or multiple path parameters (e.g., short- vs. Long-range movements)	Primary and secondary signals	[45, 80, 84, 127]
	Supervised Classification	Topology-based	Relocations (steps) of a trajectory are assigned to certain classes of movement behavior based on a classification scheme fitted with a training dataset	Primary and secondary signals, additional information like activity data	[129-131]
	Clustering	Topology-based	Unsupervised classification for identifying distinctive groups within a multivariate set of path-signals	Primary and secondary signals, additional information like activity data	[21, 132]
	Bayesian Partitioning of Markov Models (BPMM)	Topology- and time-series based	Classification algorithm for determining the number and sequence of homogenous classes within a sequential path-signal (time series)	Primary and secondary signals	[35, 91, 92]
Change-point detection	Line Simplification	Topology- or time-series based	Tests whether reducing the number of vertices in a trajectory significantly impacts path topology to determine change points (can also be applied with graphs of sequential path-signals)	Primitive signals (spatial position)	[12, 133]
	Change Point Test	Topology-based	Detects significant changes in the observed movement direction (orientation) between the starting point and an attraction point of a trajectory	Primitive signals (spatial position)	[86, 134]
	Spatio-Temporal Criteria Segmentation	Topology-based	Special type of thresholding seeking optimal segmentation of a trajectory based on monotone criteria: relocations are included in a segment as long as they fulfill certain predefined requirements	Primitive, primary and secondary signals	[32, 87]
	Piecewise Regression	Time-series analysis	Splits time-series model into representative segments based on a significant change-point (fits a polynomial model for each segment)	Primary and secondary signals	[86, 87]
	Penalized Contrast Method (PCM)	Time-series analysis	Non-parametric segmentation of a path-signal: the unknown number of segments is estimated by minimizing a penalized contrast function	Mostly secondary signals	[31, 40, 135]
	Behavioral Change Point Analysis (BCPA)	Time-series analysis	Likelihood-based method for detecting significant change points; applies moving window over continuous autocorrelated time series of a path-signal	Mostly secondary signals	[28, 35]
	Pruned Exact Linear Time (PELT) Algorithm	Time-series analysis	Search method for detecting optimal number and locations of change points minimizing different cost and penalty functions	primary and secondary signals	[42, 136, 137]
	Behavioral Movement Segmentation (BMS)	Time-series analysis	Combined search algorithm which optimizes segmentation based on parsimony and subsequent clustering for assigning segments to similar behaviors	primary and secondary signals, additional information like activity data	[43]
Process identification	Hidden-Markov Models (HMM)	State-space models	Estimate the sequence and composition of a predefined number of discrete states (e.g., movement behaviors) as well as the switching-probabilities between these states	Primary signals, additional information like activity data	[33, 49, 53-55]
	State-Space Models with Location Filtering	State-space models	More complex models which can model hidden movement states and also correct for errors in the observation process (e.g., GPS errors)	Primitive (spatial position) and primary signals, additional information like activity data	[51, 52, 65, 88, 90, 138]
	Hierarchical State-Space Models	State-space models	Hierarchical models accounting for variability of number and composition of movement states between individuals (further making inferences at population level)	Primary signals	[48, 52, 89]
	Bayesian Partitioning of Markov Models (BPMM)	Topology- and time-series based	Can also be used as partitioning algorithm determining the number and sequence of homogenous models ("states") within a sequential path-signal	primary and secondary signals	[35, 91, 92]

Für die Referenzen, siehe Edelhoff et al. (2016a)

EIGENSTÄNDIGKEITSERKLÄRUNG

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und die den verwendeten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Dübendorf, 28.09.2018



Ort und Datum

Unterschrift