

Diplomarbeit

Simulation einer Multikapillarsäule

Elisabeth Böhmer

16. Juli 2015

Betreuer:

Prof. Dr. Sven Rahmann

Prof. Dr. Jörg Rahnenführer

Fakultät für Informatik

Algorithm Engineering (Ls11)

Technische Universität Dortmund

<http://ls11-www.cs.tu-dortmund.de>

Inhaltsverzeichnis

1	Einleitung	3
1.1	Motivation und Hintergrund	3
1.2	Aufbau der Arbeit	3
1.3	Ziel der Arbeit	4
2	Grundlagen	5
2.1	Chromatographie	5
2.1.1	Der chromatographische Prozess	5
2.1.2	Detektion	6
2.1.3	Technische Daten	6
2.2	Peakcharakteristika	7
2.2.1	Lage	7
2.2.2	Form	7
2.2.3	Breite	7
2.3	Probabilistische Arithmetische Automaten	7
2.3.1	Definition eines PAA	8
2.3.2	Algorithmus zur Berechnung der Wartezeit	8
3	Modell	11
3.1	2-Parameter Modell	11
3.1.1	PAA zu diesem Modell	12
3.1.2	Grenzen des Modells	12
3.2	3-Zustände Modell	15
3.2.1	PAA	16
3.3	Weitere Modelle	16
4	Methoden	17
4.1	Simulation	17
4.1.1	Simulation des 2-Parameter Modells	17
4.1.2	Simulation des 3-Zustände Modells	17
4.2	PAA	17
4.2.1	PAA für das 2-Parameter Modell	17

4.2.2	PAA für das 3-Zustände Modell	17
5	Implementierung	19
5.1	Simulation der Modelle	19
5.1.1	Festlegung der Einheiten	19
5.1.2	Step-by-step	19
5.1.3	By-Event	20
5.1.4	Vergleich von Step-by-Step und By-Event	20
5.2	Simulation des PAA	20
6	Evaluation	21
7	Diskussion	23
7.1	Ausblick	23
A	Weitere Informationen	25
	Abbildungsverzeichnis	26
	Algorithmenverzeichnis	27
	Literaturverzeichnis	29

ToDo

	P.
1. (Wo) Soll die Aufgabenstellung noch mal rein? Also Entsprechung zwischen Peak- parametern und Sim-Parametern	1
2. Ziel und Vorgehensweise	4
3. Umformulierung auf hypothetische Peaks	4
4. Vorgehensweise graphisch darstellen!	4
5. Muss englische Worterklärung kursiv oä?	5
6. Quelle Anzahl Kapillaren einer MCC	5
7. Beide Möglichkeiten der Wechselwirkungen beschreiben	5
8. Ausführlich den chroma Proz beschreiben	5
9. Inklusive Gleichgewicht der Phasen im Inneren des Pulks	5
10. Bild Chromatogramm	6
11. MS und IMS näher beleuchten	6
12. Wo kann ich sonst Peaks beobachten?	6
13. Technische Daten belegen	6
14. Minimalzeitpunkt angeben, sowie Ende der Simulation	7

15. Zitat, dass einige Stoffe gererell tailen	7
16. Interquantilskoeffizient, Quelle, Erklären, warum	7
17. Zitat IQR, diesen erkläen, warum dieses Mass	7
18. Muss ich dann Markov erklären?	8
19. Algos für die Verteilung und Wartezeit-Berechnung vorstellen	8
20. Referenzdaten raus, Hintergründe an anderer Stelle geeignet erklären	9
21. Formale Beschreibung meines Modells	11
22. PAA für das Modell	12
23. Umformulierung auf hypothetische Peaks	12
24. ?die beiden unterkapitel modell und modell3s ganz am ende wieder in eine datei packen, wegen seitenumbruch	14
25. Erweiterung des Modells	15
26. In Intro auf Gleichgewicht bei Adsoption eingehen, -> Modellerweiterung	15
27. genaue Beschreibung der möglichen Modelle	16
28. 3sPAA aufschreiben	16
29. Bild 3sPAA?	16
30. kommt diese berechnung nicht eher in die methoden?	16
31. Berechnungen bei den verschiedenen Modellen, Text und Bilder	17
32. Kapitel Implementierung schreiben	19
33. Stey by step-Simulation beschreiben -> Methoden	19
34. by event-simulation beschreiben -> Methoden	20
35. Implementierung des PAA beschreiben; erst wenn evtl nachprogrammierung fertig	20
36. Kapitel Evaluation schreiben	21
37. Kapitel Diskussion schreiben	23

1.(Wo) Soll die Aufgabenstellung noch mal rein? Also Entsprechung zwischen Peakparametern und Sim-Parametern

Kapitel 1

Einleitung

1.1 Motivation und Hintergrund

“Multikapillarsäule”, MCC (engl. Multi Capillary Column)

Trennsäule in der Gaschromatographie

“Simulation”

Keine physikalische Simulation der Moleküle

Keine Interpolation vorhandener Messungen

Kein Überlagern verschiedener Kurven, um Gesamt-Spektrum zu erhalten
sondern: Modell für chromatographischen Prozess

Im Rahmen der Diplomarbeit soll eine Multikapillarsäule simuliert werden.

Es geht dabei nicht um eine physikalische Simulation auf molekularer Ebene, sondern um die Entwicklung eines abstrakten, probabilistischen Modells, welches mit möglichst wenig Parametern auskommt.

In der Simulation soll jeweils nur ein Stoff, charakterisiert durch diese Parameter, simuliert werden. Das Resultat einer Simulation ist jeweils ein einzelner Peak, welcher durch seine Lage und Form beschrieben werden kann. Die Lage entspricht dabei der Retentionszeit am Maximum des Peaks. Die Form ist durch seine Breite an einer bestimmten Höhe (zum Beispiel Halbwertsbreite) sowie Schiefe gekennzeichnet.

1.2 Aufbau der Arbeit

In Kapitel 2 wird zu Einem die Chromatographie, insbesondere die Gaschromatographie vorgestellt und der chromatographische Prozess erklärt. Zum Anderen PAA

1.3 Ziel der Arbeit

2.Ziel und Vorgehensweise Umgekehrt soll es möglich sein, für eine gegebene Peaklage und Peakform die nötigen Simulationsparameter zu ermitteln, mit denen ein solcher Peak simuliert werden kann.

3.Umformulierung auf hypothetische Peaks

Gesucht ist letztendlich eine allgemeine Entsprechung der Simulationsparameter zu den Parametern mit denen ein Peak beschrieben werden kann, falls diese existiert. Eine Berechnung des Peaks nur durch die Simulationsparameter und umgekehrt eine Vorhersage der Simulationsparameter zu bekannten Peakdaten soll damit realisiert werden.

Ziel dieser Diplomarbeit ist es, eine Entsprechung von Peakcharakteristika, wie sie in Kapitel 2 beschrieben werden, zu den Simulationsparametern, die für das endgültige Modell verwendet werden, zu finden, falls es eine solche Entsprechung gibt. Das heißt es soll sowohl möglich sein, zu einem gegebenen Peak die Parameter bestimmen, mit denen eben jener Peak simuliert werden kann. Umgekehrt soll zu gegebenen Simulationsparametern, der dadurch erzeugte Peak vorhergesagt werden können.

Um dieses Ziel zu erreichen, wird zunächst mit einem sehr einfachen Modell begonnen, welches nach und nach angepasst werden muss, bis die durch die Referenzdatensätze vorgegebenen Peaks ausreichend angenähert werden können.

4.Vorgehensweise graphisch darstellen!

1. Start mit einfachem 2-Parameter-Modell
2. Experimentelle Arbeit: Simulation mit verschiedenen Parametern
3. Überprüfung, ob Referenzpeaks angenähert werden können
4. Verfeinerung/Erweiterung des Modells
5. Wiederholung von 2-4 bis gegebene Peaks ausreichend angenähert

Kapitel 2

Grundlagen

2.1 Chromatographie

Die Chromatographie ist ein Verfahren zur Auftrennung von Stoffgemischen. Die Auftrennung erfolgt dabei zwischen zwei sogenannten Phasen, der stationären und der mobilen Phase, welche sich in unterschiedlichen Aggregatzuständen befinden und untereinander nicht mischen.

Es existieren zum einen die Flüssigchromatographie (LC, engl. Liquid Chromatography), **5.Muss englische Worterklärung kursiv öä?** bei der die mobile Phase eine Flüssigkeit ist und die stationäre Phase ein Feststoff. Bei der Gaschromatographie (GC) ist die mobile Phase ein Gas und es wird zusätzlich nach der stationären Phase unterschieden. Ist diese ein Feststoff, so spricht man von gepackten Säulen. Bei der Kapillartechnik hingegen werden die Trennsäulen innen mit einem Flüssigkeitsfilm als stationäre Phase beschichtet.

Beispielsweise kann die GC in einer Multikapillarsäule (MCC, engl. Multi Capillary Colum) stattfinden. Sie besteht aus ca. 1000 **6.Quelle Anzahl Kapillaren einer MCC** einzelnen Kapillaren. Jede davon ist innen mit der sog. stationären Phase beschichtet. Außerdem kommt ein Trägergas, die sog. mobile Phase, zum Einsatz, welches die Analyte durch die Säule transportiert.

2.1.1 Der chromatographische Prozess

Die Substanzen unterscheiden sich vor allem durch ihre Wechselwirkungen mit der stationären Phase. Während dieser Wechselwirkungen haften die Teilchen an der stationären Phase, bewegen sich also nicht fort. Finden wenig Wechselwirkungen statt, passieren die Teilchen die Säule schneller, als wenn viele Wechselwirkungen stattfinden. Dies beeinflusst die Retentionszeit, also die Zeit, die zum Durchlaufen der Säule gebraucht wird. **7.Bei-
de Möglichkeiten der Wechselwirkungen beschreiben 8.Ausführlich den chroma Proz beschreiben
9.Inklusive Gleichgewicht der Phasen im Inneren des Pulks**

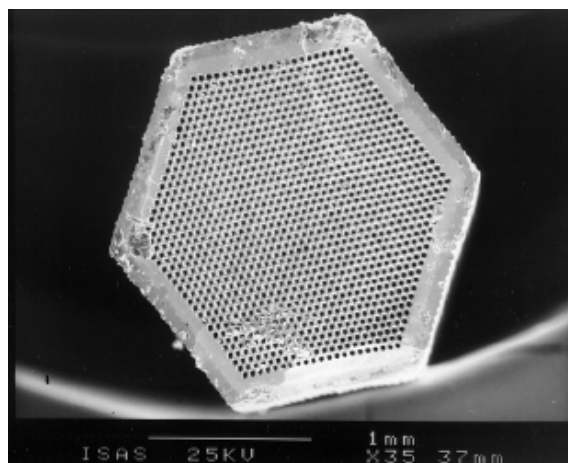


Abbildung 2.1: Querschnitt einer MCC ¹

2.1.2 Detektion

Nach Durchlaufen der Säule wird detektiert, welche Menge an Substanzen austreten. Dabei wird nicht die Art des Analyts festgestellt, sondern nur die Menge der zum jeweiligen Zeitpunkt austretenden Stoffe. Dabei zeichnet der Detektor ein Chromatogramm auf, welches beispielhaft in **10.Bild Chromatogramm** zu sehen ist

Alternativ kann die Gaschromatographie auch als Vorverarbeitung für Verfahren wie Massenspektrometrie (MS) oder Ionen-Mobilitäts-Spektrometrie (IMS) dienen. In diesen Fällen werden die aus der MCC austretenden Moleküle direkt ionisiert und in den entsprechenden Geräten weiter analysiert.

11.MS und IMS näher beleuchten

12.Wo kann ich sonst Peaks beobachten? Zu beobachten ist, dass schnelle Teilchen Peaks zu frühen Zeitpunkten erzeugen, die eine relativ geringe Varianz aufweisen, hingegen spätere Peaks tendenziell breiter werden. Ideale Peaks haben die Form einer Gaußkurve.

2.1.3 Technische Daten

Eine MCC besteht aus ca. 1000 Kapillaren mit je

- 20 – 80 μm Durchmesser
 - Stationäre Phase ist Flüssigkeitsfilm, ca. 0,1 – 0,8 μm dick
- MCC etwa 2 – 6 mm dick und 20 cm lang **13.Technische Daten belegen**

¹<http://yas.yanaco.co.jp/products/import-gc-ims.html>

2.2 Peakcharakteristika

Um später die simulierten Peaks mit denen aus den Referenzdatensätzen gut vergleichen zu können, sollen nun einige Eigenschaften aufgeführt werden, die einen Peak beschreiben.

2.2.1 Lage

Die offensichtlichste Eigenschaft ist die Lage des Peaks, genauer gesagt, die Lage des Maximums des Peaks. **14.Minimalzeitpunkt angeben, sowie Ende der Simulation**

2.2.2 Form

Außerdem können die Peaks verschiedene Formen haben. Im theoretischen Idealfall haben sie die Form einer Gaußkurve, jedoch tritt oft ein Tailing auf, welches verschiedene Ursachen haben kann. Unter Anderen seien hier genannt: zusätzliche Adsorptionseffekte, die beim Altern einer Säule auftreten [1], technische Ursachen wie kleinste Hohlräume zwischen der Säule und dem Gaseinlass bzw. -Austritt, sowie einige Stoffe, die generell zu Tailing neigen **15.Zitat, dass einige Stoffe generell tailen**. Beim Tailing steigt die Kurve zunächst stark an, sinkt jedoch nach Erreichen des Maximalwerts deutlich langsamer ab, es entsteht ein Schwanz (engl. Tail). Auch der umgekehrte Fall, Fronting genannt, kann auftreten, ist jedoch in den vorliegenden Datensätzen nicht zu beobachten. Yule-Bowley-Koeffizient **16.Interquanti skoeffizient, Quelle, Erklären, warum**

2.2.3 Breite

Als drittes können Peaks unterschiedlich breit sein. Da jedoch ein Peak mit größerer Intensität, also möglicherweise größerer Stoffmenge, automatisch auch einen breiteren Peak erzeugt, kann die Halbwertsbreite als Maß für die Peakbreite herangezogen werden. Die Halbwertsbreite wird auf halber Maximalhöhe des Peaks gemessen. Im Fall von auftreten dem Tailing oder Fronting ist es sinnvoll, je einen Wert für rechts und links des Maximalwerts zu berechnen. Interquartilsabstand **17.Zitat IQR, diesen erklären, warum dieses Mass**

2.3 Probabilistische Arithmetische Automaten

Ein Probabilistischer Arithmetischer Automat (PAA) nach [2] ist ein Modell, mit dem eine Folge zufälliger Operationen beschrieben werden kann. Für PAA existieren Algorithmen, welche eine gemeinsame Verteilung von Zuständen und Werten oder auch die Verteilung der Wartezeit für einen Wert berechnen. Wie in Kapitel 3 beschrieben wird, kann das Modell zur Simulation einer Multikapillarsäule auch als PAA formuliert werden. Mit dieser Formulierung ist die Zeit, die zum Durchlaufen einer Säule gebraucht wird, dann die Wartezeit für den Wert, welcher der Länge der Säule entspricht. Deshalb kann ein PAA

nützlich sein, um neben der eigentlichen Simulation auch noch eine erwartete Verteilung der Ankunftszeiten der Teilchen zu berechnen.

Zunächst sei hier eine Definition für den PAA gegeben, anschließend wird der Algorithmus zur Berechnung der Wartezeit beschrieben.

2.3.1 Definition eines PAA

2.3.1 Definition (PAA). Ein Probabilistischer Arithmetischer Automat (PAA) ist ein Tupel $\mathcal{P} = (\mathcal{Q}, q_0, T, \mathcal{V}, v_0, \mathcal{E}, (e_q)_{q \in \mathcal{Q}}, (\theta_q)_{q \in \mathcal{Q}})$, dabei ist:

- \mathcal{Q} eine endliche Menge von Zuständen
- $q_0 \in \mathcal{Q}$ der Startzustand
- $T : \mathcal{Q} \times \mathcal{Q} \rightarrow [0,1]$ eine Übergangsfunktion mit $\sum_{q' \in \mathcal{Q}} T(q, q') = 1$ das heißt $(T(q, q'))_{q, q' \in \mathcal{Q}}$ ist eine stochastische Matrix
- \mathcal{E} eine endliche Menge von Emissionen
- $e_q : \mathcal{E} \rightarrow [0,1]$ eine Wahrscheinlichkeitsverteilung der Emissionen für jeden Zustand
- \mathcal{V} eine Menge von Werten
- v_0 der Startwert
- $\theta_q : \mathcal{V} \times \mathcal{E} \rightarrow \mathcal{V}$ eine Operation für jeden Zustand

Dabei entspricht $\mathcal{P} = (\mathcal{Q}, q_0, T)$ einer Markovkette und $\mathcal{P} = (\mathcal{Q}, q_0, T, \mathcal{E}, (e_q)_{q \in \mathcal{Q}})$ einem Hidden Markov Model. **18.Muss ich dann Markov erklären?**

2.3.2 Algorithmus zur Berechnung der Wartezeit

19.Algos für die Verteilung und Wartezeit-Berechnung vorstellen

20. Referenzdaten raus, Hintergründe an anderer Stelle geeignet erklären

Kapitel 3

Modell

3.1 2-Parameter Modell

Kapitelübersicht: Man kann unterscheiden zwischen verschiedenen Modellen (2p/3s) und den Arten, wie diese umgesetzt werden (Teilchensim oder PAA)

Als Grundlage für die Entwicklung eines Simulationsmodells dient die Beobachtung des chromatographischen Prozesses. Dieser ist wie in Kapitel 2 beschrieben gekennzeichnet durch viele Teilchen, welche häufig zwischen der stationären und mobilen Phase wechseln. Es liegt also nahe, dass im Modell ebenfalls Teilchen simuliert werden, die zwischen zwei Zuständen, welche die beiden Phasen repräsentieren, wechseln. Der Phasenwechsel geschieht jeweils mit einer bestimmten Wahrscheinlichkeit. Dabei kann es möglich sein, dass die Wahrscheinlichkeiten für den Wechsel in die eine oder andere Richtung voneinander abhängig sind. Da dies jedoch nicht sein muss, wird für das Modell der allgemeinere Fall von unabhängigen Wahrscheinlichkeiten angenommen. Es sei also p_s die Wahrscheinlichkeit, dass ein Teilchen, welches sich bereits in der stationären Phase befindet, auch stationär bleibt und $1 - p_s$ die Wahrscheinlichkeit, dass es in die mobile Phase übergeht. Analog seien p_m und $1 - p_m$ die Wahrscheinlichkeiten, dass ein mobiles Teilchen in der mobilen Phase bleibt bzw. zur stationären Phase wechselt.

Damit ergibt sich als Modell ein einfacher Automat mit zwei Zuständen $\mathcal{Q} = \{s, m\}$ und den oben beschriebenen Transitionen. Eine graphische Darstellung des Modells zeigt

3.1 **21. Formale Beschreibung meines Modells**

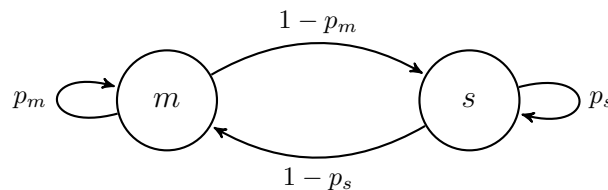


Abbildung 3.1: Graphische Darstellung des 2-Parameter Modells

Für die Simulation müssen viele Teilchen, ausgehend vom mobilen Zustand, den Automaten durchlaufen. Dabei wird zusätzlich zum Zustand der Teilchen auch der Ort, an dem sie sich befinden, verwaltet. Wenn sich ein Teilchen im mobilen Zustand befindet, wird dieser Ortszähler erhöht. Außerdem werden die insgesamt benötigten Schritte gezählt. Dadurch leiten sich dann die Ankunftszeiten der Teilchen bei Erreichen eines bestimmten Wertes des Ortszählers ab.

Der genaue Ablauf der Simulation wird in Kapitel 4 beschrieben.

3.1.1 PAA zu diesem Modell

Blabla...

Der PAA ist formal definiert durch: $\mathcal{Q} = \{s, m\}$, $q_0 = m$, $T = \begin{pmatrix} p_m & 1 - p_m \\ 1 - p_s & p_s \end{pmatrix}$, $\mathcal{E} = \{0, 1\}$, $e_s(0) = 1, e_s(1) = 0, e_m(0) = 0, e_m(1) = 1$, $\mathcal{V} = [0, \dots, l]$, $v_0 = 0$, $\theta_s = \theta_m = +$

22.PAA für das Modell

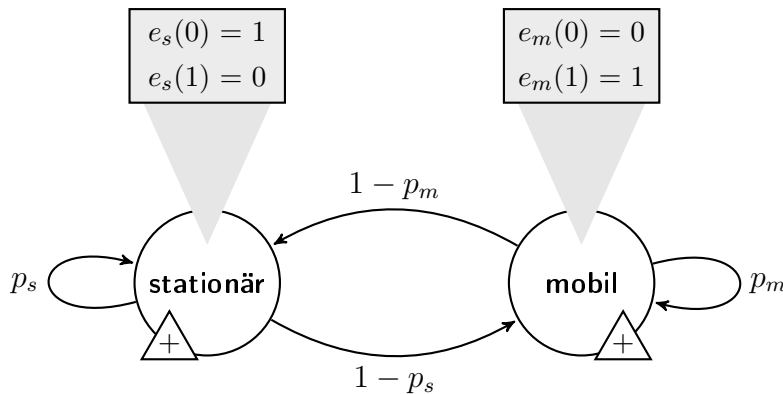


Abbildung 3.2: PAA für das 2-Parameter Modell

Und noch mehr Text, der sich auf 3.2 bezieht

3.1.2 Grenzen des Modells

Eine genaue Analyse der Peaks, die mit dem 2-Parameter Modell erzeugt werden können, findet sich in Kapitel 6. An dieser Stelle sei nur vorweg genommen, dass es zwei Hauptprobleme mit dem Modell zu geben scheint.

23.Umformulierung auf hypothetische Peaks Einfach darstellen, wie es ist (Zu jedem Zeitpunkt scheint es eine Minimalbreite zu geben, die mit diesem Modell nicht unterschritten werden kann) , und dass es evtl wünschenswert wäre, schmalere Peaks zu manchen Zeitpunkten zu haben

** Die Sache mit den Peaks die zu schmal für meine Sim waren fällt ja quasi weg, es gibt ohne Daten keine Hinweise darauf, dass es nötig ist, schmalere Peaks zu erzeugen. Daher einfach aufschreiben, DASS da eine Grenze ist (nach oben aber fast beliebig breit)

Zum Anderen weisen fast keine der simulierten Peaks ein Tailing auf. Lediglich ein sehr stark eingeschränkter Parameterbereich erzeugt ein Tailing. Das führt dazu, dass Tailing nur in einem sehr kleinen Retentionszeitbereich auftritt, allerdings nicht, wie es in tatsächlichen Daten zu beobachten ist, zu den eher späten Zeiten.

24. die beiden unterkapitel modell und modell3s ganz am ende wieder in eine datei packen, wegen
seitenumbruch

3.2 3-Zustände Modell

25. Erweiterung des Modells 26. In Intro auf Gleichgewicht bei Adsorption eingehen, -> Modellerweiterung

Der letzte Absatz zu weiteren möglichen Erweiterungen kann dann als kleiner Absatz ganz ans Ende des Modellkapitels und wird im Ausblick noch mal aufgegriffen

Bisher fand keine Unterscheidung zwischen Adsorption und Lösung der Teilchen an bzw. in der stationären Phase statt. In der Realität können beide Wechselwirkungen parallel zueinander statt finden. Außerdem ist es möglich, dass sich die Wahrscheinlichkeiten, in einen der beiden Zustände überzugehen oder in die mobile Phase zurückzukehren, unterscheiden. Daher liegt es nahe, einen neuen Zustand einzuführen, sodass die Adsorption und die Lösung voneinander getrennt behandelt werden. Die Tatsache, dass Tailing, wie bereits anfangs erwähnt, unter anderem durch zusätzlich Adsorptionseffekte verursacht sein kann, lässt vermuten, dass durch eine Anpassung der Parameter ein Tailing in der Simulation verursacht werden kann.

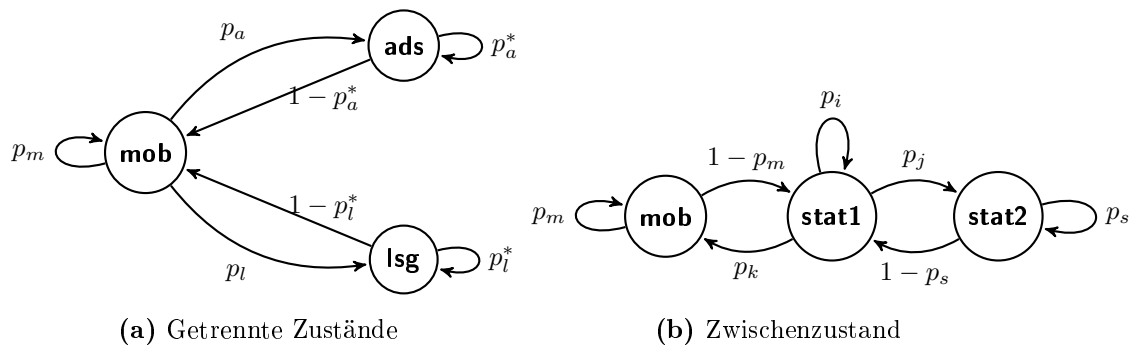


Abbildung 3.3: Mögliche weitere Modelle

In 3.3 sind zwei Möglichkeiten gezeigt, dem bisherigen Modell einen neuen Zustand hinzuzufügen. In beiden Fällen gibt es je zwei stationäre Zustände, einen für die Adsorption, einen für die Lösung.

In 3.3a sind die beiden stationären Zustände getrennt voneinander, dazwischen finden keine direkten Übergänge statt. Anschaulich kann man sich vorstellen, dass bei Übertritt in die stationäre Phase schon festgelegt wird, welcher Art der Übergang sein wird. Keine Ahnung, ob das hier rein kommt: Es scheint intuitiv sinnvoll zu sein, die Übergangswahrscheinlichkeiten von und zu den beiden Zuständen sehr unterschiedlich zu wählen, damit sie sich nicht einfach wieder zu einer Gesamtwahrscheinlichkeit aufaddieren. In einen der Zustände sollen sie also nur seltener kommen, dafür dort lange verweilen, denn ein solches Verhalten würde zu Tailing führen.

In 3.3b dient der erste stationäre Zustand als Übergangszustand zum zweiten stationären Zustand. Dabei ist der Übergangszustand die Adsorption als Hinweis darauf, dass die Teilchen, die sich in der stationären Phase lösen, zunächst mit deren Oberfläche in

Kontakt treten und zumindest kurzzeitig adsorbiert sind, gleiches gilt für die andere Richtung. Nachdem ein Teilchen adsorbiert wurde, besteht die Möglichkeit, dass es sich noch in der stationären Phase löst oder auch adsorbiert bleibt. Auch hier liegt es nahe, die Wahrscheinlichkeiten für die verschiedenen Übergänge so zu gestalten, dass die Teilchen in einem der stationären Zustände seltener aber länger bleiben.

27.genaue Beschreibung der möglichen Modelle

3.2.1 PAA

Analog zum PAA für nur zwei Zustände kann für drei Zustände ein PAA definiert werden.

28.3sPAA aufschreiben

29.Bild 3sPAA?

30.kommt diese berechnung nicht eher in die methoden? Berechnung für nächsten Schritt '

Habe Matrix mit m, a, l und Orten 0...Länge mit den jeweiligen W'keiten Neue Matrix:

$$s'_x(n) = s_{x-1}(1) * p_{1,n} + \sum_{m=2}^{|N|} s_x(m) * p_{m,n}$$

3.3 Weitere Modelle

Eine weitere Möglichkeit, das Modell zu verändern, besteht darin, das Gleichgewicht, welches sich zwischen den beiden Phasen aufbaut, zu berücksichtigen. Damit müssten die Wahrscheinlichkeiten, den Zustand zu wechseln, nicht mehr fest vorgegeben sein, sondern sich dynamisch während der Simulation aus der aktuellen Verteilung der Teilchen an einem bestimmten Ort auf die Zustände berechnen.

Außerdem existieren Idee, noch eine Sättigung der freien Plätze zur Adsorption einzubauen

Kapitel 4

Methoden

4.1 Simulation

31.Berechnungen bei den verschiedenen Modellen, Text und Bilder

4.1.1 Simulation des 2-Parameter Modells

4.1.2 Simulation des 3-Zustände Modells

4.2 PAA

Wie werden die neuen Einträge berechnet? Graphik, Text dazu

Wie werden die Arrays gekürzt

4.2.1 PAA für das 2-Parameter Modell

4.2.2 PAA für das 3-Zustände Modell

Kapitel 5

Implementierung

32.Kapitel Implementierung schreiben

5.1 Simulation der Modelle

Zur Simulation der beiden Modelle wurde Python (Python3) verwendet.

Abhängigkeiten von Numpy (Arrays zur schnelleren Verarbeitung), Scipy (Berechnung von Peakdaten) und PyPlot (graphische Darstellung)

Mit der 3-s kann auch 2p simuliert werden, indem nicht benötigte Übergangswahrscheinlichkeiten auf 0 gesetzt werden. Die reine 2p-Simulation ist aber deutlich schneller, da nur auf boolschen Werten gearbeitet werden muss.

5.1.1 Festlegung der Einheiten

20 cm entspricht 200000 Schritten, in die MCC unterteilt wird. 1 Schritt ist damit 1 μm Trägergas hat nach etwa 0,1 Sekunden die Säule durchlaufen. Eine Zeiteinheit entspricht demnach ca. $0.1/200000$ Sekunden 0.0000005s 0.0005ms 0,5 μs

5.1.2 Step-by-step

Es kann entweder jeder Zeitschritt der Chromatographie simuliert und dabei für jedes Teilchen entschieden werden, ob es sich bewegt. Dabei angenommen, dass die Wahrscheinlichkeit der Wechselwirkung unabhängig vom aktuellen Ort oder dem Zeitpunkt des letzten Phasenwechsels ist. Es wird für jedes Teilchen festgehalten, wo es sich befindet und ähnlich wie bei der echten Chromatographie beobachtet, wann wie viele Teilchen die Säule durchquert haben. **33.Step by step-Simulation beschreiben -> Methoden**

5.1.3 By-Event

Bei Übergangswahrscheinlichkeiten, die sehr groß oder klein sind, kommt es selten zu Zustandsänderungen, sodass es in diesem Fall effizienter ist, eine Wartezeitmethode zu nutzen: Es wird für jedes Teilchen entschieden, wann es den Zustand wechselt und, falls es mobil ist, wie weit es bis dahin weiter wandert. Es werden also nur die tatsächlichen Ereignisse simuliert, Schritte für Zeitpunkte, zu denen nichts passiert, entfallen. **34.by event-simulation beschreiben -> Methoden**

5.1.4 Vergleich von Step-by-Step und By-Event

Resultierende Peaks sind (natürlich) gleich. Der Laufzeitvergleich hängt stark von den gewählten Parametern, aber auch Länge der Säule, ab. Bei Parametern nah an 1 lohnt meist die By-Event, da dann nur seltener Ereignisse erzeugt werden und viele Schritte in denen nichts (also kein Zustandswechsel) passiert, übersprungen werden können. Bei sehr vielen Teilchen lohnt sich eher die Step-by-Step Variante, da die Teilchen, die in Numpy-Arrays als True und False - Werte vorliegen sehr viel schneller verarbeitet werden können, als die Listen, die bei der Event -Variante verwendet werden. Eine weitere Parallelisierung scheint hier aber nicht zu lohnen. Zumindest bei Parametern extrem nah bei 1 kann sich die Laufzeit der By-Event stark vergrößern, wenn die Säule zu kurz ist, da die Teilchen “viel zu weit über das Ziel hinausschiessen”, aber das erst später festgestellt wird, wenn das Teilchen wieder betrachtet wird und schon entsprechend viele Zeitschritte vergangen sind.

5.2 Simulation des PAA

Die Berechnung des PAA existiert in Java, dabei Referenz auf mosdi (Zitiere Paper, in dem das vorgestellt wird und sage, dass ich das verwende) **35.Implementierung des PAA beschreiben; erst wenn evtl nachprogrammierung fertig** da allerdings zu langsam: abgespeckte version nachprogrammiert in julia Größte Unterschiede dabei: start eingeschränkte funktionalität, da addition (von 1) unterstützt, das macht teilweise das arbeiten auf true/false möglich, außerdem auch nur jeweils feste ausgabe von 0 oder 1, festgelegt nach zustand etc. außerdem arbeiten mit arrays variabler länge, nur der relevante bereich (wkeiten != 0) wird betrachtet und zusätzlich zu kleine Werte zu beginn und ende der teilchenmasse abgeschnitten, um die länge weiter zu beschränken und dadurch zeit zu sparen (laufzeit linear/quadratisch oder so in irgendwas? noch mal bei mosdi nachgucken) länge sowieso beschränkt durch länge der säule (?????) und nicht durch maximale wartezeit

Warum Julia?

Was kann meine PAA-Version und was nicht?

Laufzeitvergleich bei gleichen Ergebnissen

Nachbearbeitung der Ergebnisse wieder in Python

Kapitel 6

Evaluation

36.Kapitel Evaluation schreiben

Was kann ich mit meinem Modell?

Simulation von Peaks die ideal-gaußförmig sind oder tailing aufweisen

Grenzen des Modells: Minimalbreite kann nicht unterschritten werden

Formel zur Umrechnung von Peakdaten in Parameter, oder falls das nicht drin ist, zumindest Zusammenhänge, dazu zb Plots von drei festen Parametern Evtl Tabelle mit exemplarischen Daten? Sinnvolle Parameterbereiche

3a modell liefert tailing, welches bei 3b nicht gefunden wurde (vielleicht einfach nur die falschen parameter ausprobiert?)

Kapitel 7

Diskussion

37.Kapitel Diskussion schreiben

7.1 Ausblick

Das ganze mit existierenden Daten vergleichen, um zu gucken, ob schmalere Peaks notwendig sind und alle realen Peakdatenkombinationen erreicht werden können, dann eventuell noch mal das Modell anpassen:

Aufgreifen der nicht umgesetzten Modellerweiterungen: Berücksichtigung des Gleichgewichts der Phasen und beschränkte Anzahl freier Plätze bei der Adsorption.

Eine weitere Möglichkeit, das Modell zu verändern, besteht darin, das Gleichgewicht, welches sich zwischen den beiden Phasen aufbaut, zu berücksichtigen. Damit müssten die Wahrscheinlichkeiten, den Zustand zu wechseln, nicht mehr fest vorgegeben sein, sondern sich dynamisch während der Simulation aus der aktuellen Verteilung der Teilchen an einem bestimmten Ort auf die Zustände berechnen.

Außerdem existieren Idee, noch eine Sättigung der freien Plätze zur Adsorption einzubauen

Das ganze auch für spätere Zeitpunkte überprüfen (echte Messlänge, dauert nur halt lange mit der Sim, evtl dafür die eventsim nutzbar?)

Falls Formel für entsprecung nicht gefunden: Das Sekantenverfahren erwähnen?

Anhang A

Weitere Informationen

Abbildungsverzeichnis

2.1	Querschnitt einer MCC	6
3.1	Graphische Darstellung des 2-Parameter Modells	11
3.2	PAA für das 2-Parameter Modell	12
3.3	Mögliche weitere Modelle	15

Algorithmenverzeichnis

Literaturverzeichnis

- [1] KOLB, BRUNO: *Gaschromatographie in Bildern [eine Einführung]*. Wiley-VCH, Weinheim, 2., überarbeitete und erweiterte Auflage, 2003.
- [2] MARSCHALL, TOBIAS, INKE HERMS, HANS-MICHAEL KALTENBACH und SVEN RAHMANN: *Probabilistic Arithmetic Automata and Their Applications*. IEEE/ACM Trans. Comput. Biol. Bioinformatics, 9(6):1737–1750, November 2012.