

Zelluläre Automaten für die Modellierung und Simulation von Stammzellen

Fabian Wanner & Manuel Widmer

Semesterarbeit SA-2006-28

Sommersemester 2006

Betreuer: **Simon Barkow-Oesterreicher**

Verantwortlicher Professor: **Eckart Zitzler**

25. Juli. 2006

Zusammenfassung

In der Entwicklungsbiologie stellen sich heutzutage immer noch viele Fragen, was die Entwicklung von Zellen in Organismen betrifft. So ist zum Beispiel noch weitgehend unbekannt, wie Gene untereinander wechselwirken, so dass sich im Zellverband bestimmte Verhältnisse von Genkonzentrationen ergeben. Interessant dabei ist die Tatsache, dass aufgrund dieser Verteilungen entschieden wird, welche Aufgabe eine Zelle übernehmen wird.

In dieser Arbeit wurde das Verhalten von Genen im Zellverband simuliert. Dazu wurden Wechselwirkung, Diffusion, Eigenproduktion und Selbstzerfall von Genen modelliert. Diese Modelle wurden simuliert und so die Entwicklung der Genkonzentrationen berechnet und grafisch dargestellt. Dies wurde mit einem zellulären Automaten bewerkstelligt, da dieser vom Aufbau her stark an einen biologischen Zellverband erinnert und deshalb für die vorliegende Aufgabe als geeignet erscheint.

In einem weiteren Schritt wurde versucht, für bekannte Gen-Muster aus der Literatur ein Gen-Wechselwirkungsmodell zu finden, welches dieses Muster generiert. Dazu wurde ein genetisches Optimierungsverfahren verwendet, bei welchem verschiedene Modelle erstellt und mittels zellulärem Automaten simuliert werden. Die dabei entstehenden Muster werden nach Ähnlichkeit zum vorgegebenem Muster bewertet und schliesslich die zugrundeliegenden Modelle verändert, um bessere Modelle zu finden.

Letztendlich ist das Ziel, aus gegebenen Genmustern auf mögliche zugrundeliegende Gen-Netzwerke schliessen zu können.

Abstract

In today's evolution biology there are still many unanswered questions, for example the cell development in organisms. An unsolved problem is how the genes interact among each other and as a result of this how they establish their typical gene concentration. The interesting fact is that this gene concentration determines what kind of cell it will be.

In this thesis, the authors modeled the most important gene parameters: Interaction between the genes, diffusion, self production and self reduction. Via simulation, the gene concentration of the found models was calculated and visualized. The simulation is based on cellular automata (CA). This kind of automata is predestinated for this usage because they have many similarities with the biological cell structure.

In a further step, we tried to find an interaction model which produces some well known gene pattern. To achieve this ambition, we used a genetic optimization algorithm. The algorithm generates a population of gene models and simulates them with the CA in order to get as close to a reference model as possible. Every model was tested on its affinity to the reference pattern and its interaction model was adapted accordingly to find better models.

To find an interaction model for a given gene pattern is considered as the main target.

Inhaltsverzeichnis

Abbildungsverzeichnis	iv
Tabellenverzeichnis	vi
Liste der Abkürzungen	vii
1 Einführung	1
1.1 Thematischer Hintergrund	1
1.2 Aufgabenstellung	2
2 Software	4
2.1 Übersicht	4
2.2 Benutzungsoberfläche	4
2.2.1 Interaktionsfenster (Main GUI)	4
2.2.2 Visualisierung der Zellstruktur	7
2.2.3 Konsolenausgabe	8
2.2.4 File I/O	10
2.3 Zellulärer Automat	10
2.3.1 CA Modelle	10
2.3.2 Funktionsweise	12
2.3.3 Validierung der Komponenten	13
2.4 Evolutionärer Algorithmus	18
2.4.1 Übersicht	19
2.4.2 Generierung der Startpopulation	21
2.4.3 "Mating Selection"	21
2.4.4 Rekombination	23
2.4.5 Mutation	25
2.4.6 Fitness-Evaluation	27
2.4.7 "Environmental Selection"	30
3 Ergebnisse	33
3.1 Experimente	33
3.1.1 Symmetrien	34
3.1.2 Skalierbarkeit	38
3.1.3 Optimierungsverlauf	43
3.1.4 Einfluss der verschiedenen Randbedingungs-Formen	44
3.1.5 Einfluss der Konzentrationsgrenze	45
3.2 Performance	46

3.2.1	Zeitmessung	46
3.2.2	Optimale Populationsgrösse	49
3.3	Anwendungen	52
3.3.1	Meristem-Muster nach Grandjean	52
3.3.2	Meristem-Muster nach Carles	60
4	Schlussfolgerungen	64
4.1	Erfolge	64
4.2	Probleme	64
4.3	Ausblick	65
A		68
A.1	Zeitplan	69
A.2	Optimale Popgrösse - Tabellen	71
A.3	Textausgabe zu Optimierung "Carles"	73
A.4	Verifizierung Simulator (Excel TM)	76
	Literaturverzeichnis	85

Abbildungsverzeichnis

2.1	UML-Diagramm der Programmklassen	5
2.2	GUI	6
2.3	Modellauswahl Menü	7
2.4	Leere Zellstruktur	8
2.5	Beispiel eines Genmusters	8
2.6	Beispiel für die Konzentrationsvisualisierung	9
2.7	Interaktionsmatrix	11
2.8	Beispiele spezieller Nachbarschaftsbeziehungen	13
2.9	Diffusionsbeispiel mit spezieller Randbedingung	14
2.10	Diffusionsbeispiel mit normaler Randbedingung	14
2.11	Selbstproduktion mit Diffusion	15
2.12	Verlauf der Genverdrängung	17
2.13	Konzentrationsverlauf des Gens A	17
2.14	Konzentrationsverlauf des Gens B	18
2.15	Aufbau der Modell-Population	19
2.16	Schritte des evolutionären Optimierungsverfahrens	20
2.17	Beispiel-Ablauf zum Tournament-Verfahren	22
2.18	Beispiel-Ablauf zur Rekombination	24
2.19	Arten von Parameter-Mutationen	26
2.20	Vorgehen bei der "Environmental Selection"	30
2.21	Vorgehen bei der "Environmental Selection" mit "fewBest"	32
3.1	Vertikal-achsensymmetrisches Muster, gerade Spaltenzahl	33
3.2	Vertikal-achsensymmetrisches Muster, gerade Spaltenzahl	35
3.3	Vertikal-achsensymmetrisches Muster, ungerade Spaltenzahl	35
3.4	Fitnessverläufe für ein vertikal-achsensymmetrisches Muster	36
3.5	Horizontal-achsensymmetrisches Muster, gerade Spaltenzahl	36
3.6	Horizontal-achsensymmetrisches Muster, ungerade Spaltenzahl	37
3.7	Fitnessverläufe für ein horizontal-achsensymmetrisches Muster	37
3.8	Punktsymmetrisches Muster, gerade Spaltenzahl	38
3.9	Punktsymmetrisches Muster, ungerade Spaltenzahl	38
3.10	Fitnessverläufe für ein punktsymmetrisches Muster	39
3.11	Oben: Referenzmuster mit Konzentration Gen 1, Gen 2. Unten: Optimiertes Muster mit Konzentration Gen 1 Gen 2	40
3.12	Oben: Referenzmuster mit Konzentration Gen 1, Gen 2. Unten: Optimiertes Muster mit Konzentration Gen 1 Gen 2	41
3.13	Abhängigkeit der Optimierungszeit von der Feldgrösse, blau: Messda- ten, rot: parametrisierte Kurve mit Formel	46

3.14	Abhängigkeit der Optimierungszeit von der Anzahl Gene	47
3.15	Abhängigkeit der Simulationszeit von der Anzahl Gene	48
3.16	Tabelle zu Diagramm aus Abbildung 3.15	48
3.17	Abhängigkeit der Optimierungszeit von der Populationsgrösse . .	49
3.18	Gegenüberstellung von Fitness-Verbesserungsfaktoren, Muster: Einfaches Modell, 1000 EA-Iterationen	50
3.19	Gegenüberstellung von Fitness-Verbesserungsfaktoren, Muster: Carles, 100 EA-Iterationen	50
3.20	Zeiten zum Erreichen einer Fitness-Grenze, Muster: Carles	52
3.21	Muster Grandjean1, Original und Nachbildung (Wuschel)	53
3.22	oben: Referenzmuster Grandjean1 mitte: Optimiertes Muster mit normaler Randbedingung unten: Optimiertes Muster mit spezieller Randbedingung	54
3.23	Textausgabe der Genkonzentrationen (normale Randbedingung)	54
3.24	Textausgabe der Genkonzentrationen (spezielle Randbedingung)	55
3.25	Muster Grandjean2, Original und Nachbildung (ATML1)	56
3.26	oben: Referenzmuster Grandjean2 mitte: Optimiertes Muster mit normaler Randbedingung unten: Optimiertes Muster mit spezieller Randbedingung	57
3.27	Textausgabe der Genkonzentrationen (normale Randbedingung)	58
3.28	Textausgabe der Genkonzentrationen (spezielle Randbedingung)	59
3.29	Links: Referenzmuster aus der Literatur, Rechts: Umsetzung in verwendetes Modell	60
3.30	Resultat der Optimierung mit 3 Genen	61
3.31	Fitnessverlauf der Optimierung mit 3 Genen	62

Tabellenverzeichnis

2.1	Konsolenoutput während der Optimierung	9
2.2	Parameterdefinitionen	12
2.3	Stabilitätskriterium	12
2.4	Validationsmodell	16
2.5	Validationsmodell	16
3.1	Optimiertes Modell mit normaler Randbedingung	40
3.2	Optimiertes Modell mit spezieller Randbedingung	41
3.3	Referenzmodell 6 Gene Teil 1	42
3.4	Referenzmodell 6 Gene Teil 2	43
3.5	Optimierungsergebnisse	44
3.6	Optimierungsergebnisse für konstantes Muster	44
3.7	Optimierungsergebnisse	45
3.8	Optimiertes Modell mit normaler Randbedingung	55
3.9	Optimiertes Modell mit spezieller Randbedingung	56
3.10	Optimiertes Modell mit normaler Randbedingung	58
3.11	Optimiertes Modell mit spezieller Randbedingung	59

Liste der Abkürzungen

- CA: Cellular Automaton / Zellulärer Automat
- EA: Evolutionärer/genetischer Algorithmus
- GUI: Graphical User Interface
- WUS: Wuschel (ein Gen)
- CLV: Clavata (ein Gen)
- ATML1: Arabidopsis thaliana meristem L1 layer (ein Gen)

Kapitel 1

Einführung

Im vorliegenden Artikel wird nicht zwischen Gen und Genprodukt (Protein) unterschieden. Mit "Genkonzentration" ist also die Konzentration des jeweiligen Genprodukts gemeint.

1.1 Thematischer Hintergrund

Ein wichtiges Teilgebiet der Biologie ist die Erforschung der Entwicklung von Organismen, das bis heute noch viele ungeklärte Fragen beinhaltet. Beispielsweise ist das Phänomen der Musterbildung in Zellformationen immer noch ungeklärt. Durch Musterbildung entstehen spezialisierte Zonen im Organismus, die z. B. für das ständige Wachstum von Pflanzen verantwortlich sind. Diese spezialisierten Zonen heissen bei Pflanzen Stammzellzonen oder Meristeme.

Die Zellen der Meristeme¹ sind undifferenziert/unspezialisiert, d.h. deren Aufgabe für die Pflanze an und für sich ist noch nicht festgelegt. Diese embryonalen Zellen sind theoretisch unbegrenzt teilungsfähig, d.h. aus ihnen entstehen neue, spezialisierte Zellen. Bei Pflanzen sitzen die besagten Zellen folglich am äussersten Ende der Sprossspitze und an den Wurzelenden, also überall dort, wo in der Pflanze Wachstum stattfindet.

Da die Kontrolle der Genverteilungen - und damit der Stammzellen - in diesen Zonen der Schlüssel dazu ist, beispielsweise bei Nutzpflanzen gewünschtes Verhalten zu erzeugen, ist das Verständnis der Meristeme von grossem Interesse. Damit könnte also die Bildung der verschiedenen Zelltypen "gesteuert" werden, also wo/wann/welche/wieviele bestimmte Zellen entstehen sollen. Die Funktion der einzelnen Zelltypen leiten sich aus dem Verhältnis der Konzentrationen der verschiedenen Gene ab. Diese Konzentrationsverteilung der Gene wird im weiteren Verlauf dieses Artikels als Genmuster oder einfach Muster bezeichnet. Diese Musterbildung von Genen ist Thema dieser Arbeit, wobei das ganze System vereinfacht als Simulation auf dem Computer nachgebildet werden soll.

In der Biologie spielen Modelle jeglicher Art eine bedeutende Rolle. Sie helfen dabei, Hypothesen aufzustellen, die Erklärungen für beobachtetes Verhalten liefern können. Durch Computermodellierung kann vorhandenes biologisches Wissen genutzt werden, um mit Hilfe von Simulationen neues Wissen abzuleiten. Auch für das Verständnis der Organisation der Stammzellzone von Pflanzen

¹Auch "embryonale Zellen" genannt.

können Computermodelle einen wichtigen Beitrag leisten. Für die Modellierung der Stammzellzone in Pflanzen wurden in der Vergangenheit bereits unterschiedliche Ansätze verfolgt [2]. Ein viel versprechender Ansatz stellt dabei die Modellierung mittels zellulärer Automaten (ZA) [4] dar. ZAs bilden eine Klasse von Computersimulationen, die gerade für die Simulation eines räumlichen Modells geschaffen sind. Da die räumliche Aufteilung der Stammzellzone und die Kommunikation zwischen ihren Teilzonen entscheidend für das Verständnis der biologischen Vorgänge ist, bieten sich ZAs also an.

1.2 Aufgabenstellung

In dieser Arbeit sollte die Musterbildung von Genen in biologischen Zellverbänden modelliert werden und schliesslich die Parameter dieses Modells optimiert werden, so dass es ein vorgegebenes Muster möglichst genau reproduziert.

Die Modellierung sollte folgende Punkte beinhalten: Die Zellstruktur (Geometrie des Zellverbandes), Genkonzentration pro Zelle und Gen, die Interaktion zwischen Genen und die Diffusion der einzelnen Gene. Nach dem Finden einer geeigneten Modellierung sollte ein Programm geschrieben werden, welches aufgrund der übergebenen Modellparameter die Entwicklung des Gen-Netzwerks simuliert. Das Resultat dieser Simulation sollte ein Muster sein, in welchem die Genkonzentrationen der einzelnen Felder und Gene des Zellverbands grafisch dargestellt wird. Als Vorlage und Anregung für einige Aspekte dieser Arbeit diene ein Artikel von [5], welcher sich mit ähnlichen Problemen auseinandersetzt.

Das Vorgehen zur Lösung dieser Aufgaben war das folgende: Die Simulation des Gennetzwerks sollte durch einen zellulären Automaten erfolgen, da dieser von Struktur her mit biologischen Zellen vergleichbar ist. Anschliessend sollte das Ergebnis geeignet visualisiert werden. Durch Variation der Modellparameter sollte nach der Implementierung dieses ersten Teils der Software ein Modellparameter-Satz gefunden werden, welcher in einem stabilen² Muster resultiert.

Zur Optimierung der zu einem gegebenen Zielmuster passenden Modellparameter sollte ein Evolutionärer Algorithmus [6] eingesetzt werden, da das Problem für herkömmliche Optimierungsverfahren wahrscheinlich zu komplex ist. Dieser Optimierer sollte eine Funktion enthalten, welche das durch Simulation eines Modells entstandene Endmuster hinsichtlich Ähnlichkeit zum vorgegebenen Zielmuster bewertet.

Die von uns bearbeiteten Teilschritte sind:

- **Zellulärer Automat:**

- Modellierung des Gen-Netzwerks: Geometrie des Zellverbands, Interaktion/Diffusion/Produktion/Abbau der Gene und Genkonzentration pro Feld und Gen.
- Entwurf und Implementierung eines multivariablen zellulären Automaten, welcher instande ist, mit kontinuierlichen Grössen (Genkonzentrationen) zu arbeiten. Die Transitionsfunktion zur Berechnung

²Ein Muster gilt als stabil, falls sich bei der Simulation irgendwann (bzw. nach gewisser maximaler Anzahl Simulationszyklen) ein Gleichgewicht einstellt und sich von einem Zeitschritt auf den nächsten nichts mehr ändert.

des nächsten Zeitschritts wurde in Abhängigkeit der Interaktion, Diffusion, Eigenproduktion und Abbau der Gene aufgestellt.

- Visualisierung der über den Zellverband verteilten Genkonzentrationen.
- Änderung des Modells bezüglich Einfluss von Interaktion, Produktion und Abbau der Gene.

• **Optimierung:**

- Ein klassischer evolutionärer Algorithmus zur Optimierung der Modellparameter wurde implementiert.
- Zur Bewertung der Modelle der Population wurden verschiedene Methoden umgesetzt, welche das Endmuster einer Simulation mit dem Zielmuster vergleicht. Jene Modelle, welche die kleinste Differenz zum Zielmuster ergeben, gelangen in die nächste Generation der Optimierung.
- Die Performance des Optimierungsverfahrens wurde bei Erhöhung der Populationsgrösse, Anzahl Gene und Felddimensionen analysiert.
- Es wurde versucht, jene Populationsgrösse zu finden, bei welcher der Optimierer am schnellsten arbeitet.
- Der Optimierer wurde mit Endmustern von Modellen getestet, welche von der Software selbst generiert wurden.
- Schliesslich wurde versucht, komplexere Muster sowie in der Literatur zu findende Muster durch den Optimierer anzunähern.

Kapitel 2

Software

2.1 Übersicht

Ein wesentlicher Teil der Semesterarbeit beinhaltet die Entwicklung eines Programms, das zu einem gegebenen Genmuster eine passende Topologie (Genmodell) findet. Die Hauptteile der Software sind Benutzungsoberfläche (GUI), Simulator (Zellulärer Automat) und Optimierer (Evolutionärer Algorithmus). Den Aufbau der Software zeigt das UML-Diagramm in Abbildung 2.1. Die wichtigsten Methoden und Variablen der einzelnen Klassen sind in der Grafik dargestellt. Des weiteren ist zu sehen, wie oft welche Klasse von anderen als Objekt instanziiert werden kann.

Im Artikel wird oft von Modellen gesprochen; damit sind Genmodelle gemeint. Das Genmodell ist im wesentlichen ein Parametersatz mit den Werten für Diffusion, Interaktion, Eigenproduktion und Abbau. Eine genaue Erklärung folgt in Abschnitt 2.3.1.

2.2 Benutzungsoberfläche

2.2.1 Interaktionsfenster (Main GUI)

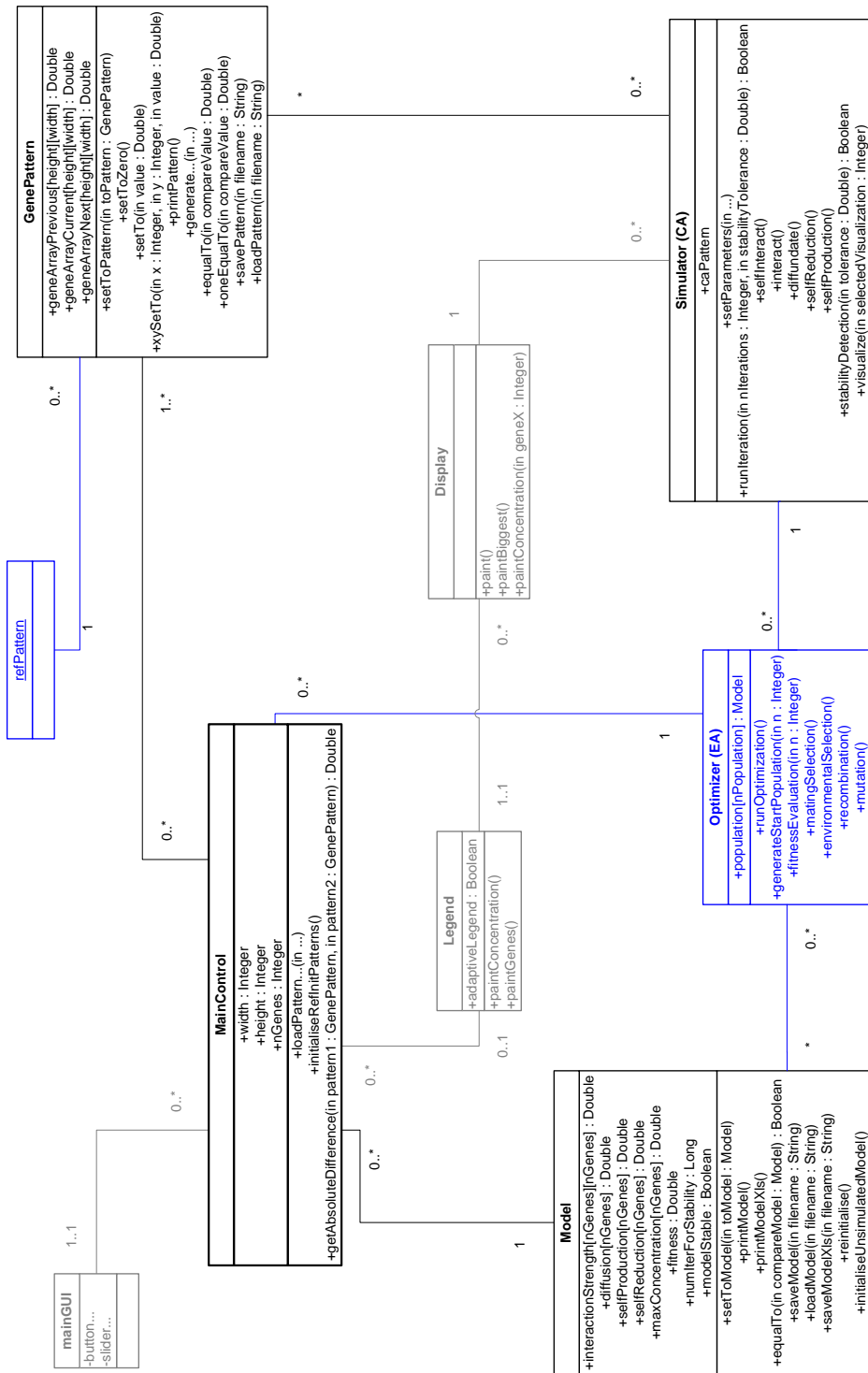
Das Interaktionsfenster erlaubt dem Benutzer, die wichtigsten Parameter für die Simulation von Hand einzustellen. In Figur 2.2 ist der Aufbau dargestellt.

CA Hoehe, CA Breite: Diese Parameter erlauben dem Benutzer die genaue Dimension der Zellstruktur vorzugeben.

Anzahl Gene: Mittels dieses Parameters lässt sich die gewünschte Anzahl Gene zuweisen, die im Modell verwendet werden sollen.

Anzahl Optimierungs Iterationen: Dieser Regler bestimmt die maximale Anzahl Optimierungsiterationen des Evolutionären Algorithmus, die durchlaufen werden sollen.

Populationsgrösse Dieser Parameter beeinflusst die Grösse der Modellpopulation. Er bestimmt, wie viele verschiedene Genmodelle man in jedem Iterationsschritt behält. Einzelheiten sind in Abschnitt 2.4 zu finden.



... **Nur für Optimierung** ... **Betrifft Benutzeroberfläche**

Abbildung 2.1: UML-Diagramm der Programmklassen

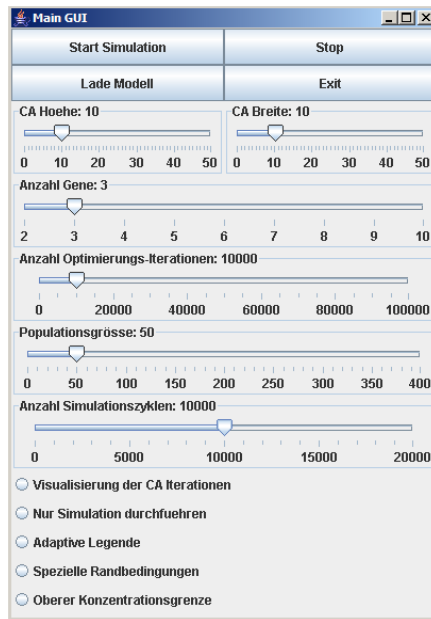


Abbildung 2.2: GUI

Anzahl Simulationszyklen: Der letzte Regler bestimmt die maximale Anzahl der Simulationsschritte für jedes Genmodell. Eine genauere Beschreibung findet man im Abschnitt 2.3.

Visualisierung der CA Iterationen: Ist diese Checkbox angeklickt, werden die Iterationsschritte des Zellulären Automaten graphisch dargestellt, so dass die Entwicklung mitverfolgt werden kann. Dieser Parameter macht nur dann Sinn, wenn man ein einzelnes Genmodell simulieren möchte, also im Zusammenhang mit der "Nur Simulation durchführen" - Checkbox. Bei einer Optimierung interessieren die Entwicklungen der "Zwischenmodelle" hingegen nicht.

Nur Simulation durchführen: Dieser Parameter bewirkt, dass die Optimierung umgangen wird und nur die Simulation eines Modells durchgeführt wird. Er wird verwendet, falls man sich zu einem Genmodell das zugehörige Zellmuster ansehen möchte.

Adaptive Legende: Die Verwendung dieses Parameters führt dazu, dass die Legende mit den Genkonzentrationen für jedes Gen separat an seine Genkonzentration angepasst wird.

Spezielle Randbedingung: Hier kann man zwischen zwei Arten von Randbedingungen wählen:

nicht angeklickt: Diese Einstellung wird im weiteren Text als "normale Randbedingung" bezeichnet. Sie bewirkt, dass die Zellstruktur als unendlich angesehen wird. Die vom Programm angezeigte Zellstruktur ist somit nur ein Ausschnitt daraus. Daher haben auch die

Randzellen sechs Nachbarn, und das Genprodukt kann in den nicht sichtbaren Bereich diffundieren.

angeklickt: Diese Art von Randbedingung wird im weiteren Verlauf des Artikels als "spezieller Randbedingung" betitelt. Hier wird die Zellstruktur als endlich angenommen und die Randzellen haben weniger Nachbarn als Zellen in der Mitte. Randzellen diffundieren Genprodukte nur in Zellen im Inneren der Struktur, es geht also nichts "verloren".

Obere Konzentrationsgrenze: Diese Eingabe bestimmt, ob eine obere Konzentrationsgrenze in jeder Zelle existiert. Die Konzentration des Genprodukts kann dadurch nie über diese Grenze ansteigen.

Über die folgenden Buttons kann der generelle Ablauf des Programms gesteuert werden.

Start Simulation: Startet einen Run und es erscheint ein Auswahlmenu (Abbildung 2.3) für bestimmte vorgegebene Muster.

Lade Modell: Über diesen Knopf lässt sich ein beliebiges vorher gespeichertes/ingetipptes Genmodell aus einer Datei laden und anschliessend simulieren. Dieser Button ist also nur relevant, wenn man ein einzelnes Modell simulieren will.

Stop, Exit: Die beiden letzten Buttons bewirken einerseits den Abbruch des aktuellen Runs (*Stop*) bzw. das Schliessen des Programms (*Exit*).

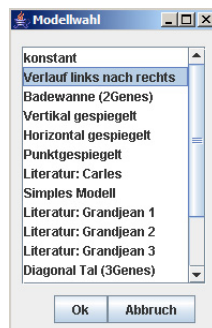


Abbildung 2.3: Modellauswahl Menü

2.2.2 Visualisierung der Zellstruktur

Es wurden zwei Arten von Visualisierung implementiert, das Grundgerüst sieht dabei immer gleich aus. Ein Beispiel ist in Abbildung 2.4 ersichtlich.

Verteilungsvisualisierung

Diese Methode visualisiert die Verteilung der Gene als Ganzes, d.h. in einem Bild sind alle Gene vorhanden. Jedes Gen bekommt eine bestimmte, eindeutige Farbe zugeteilt. Die Zelle wird mit der entsprechenden Farbe eingefärbt, falls folgende zwei Punkte erfüllt sind:

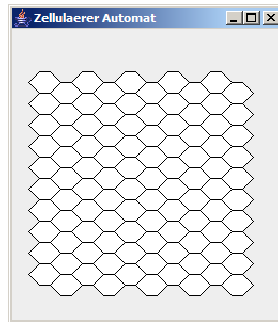


Abbildung 2.4: Leere Zellstruktur

1. Das Gen hat die höchste Konzentration in dieser Zelle gegenüber den anderen Genen
2. Die Genkonzentration liegt über einem Schwellenwert

In der Abbildung 2.5 sieht man ein Beispiel für diese Darstellungsart. An der Farbe der Zelle lässt sich feststellen welches Gen jeweils die höchste Konzentration besitzt.

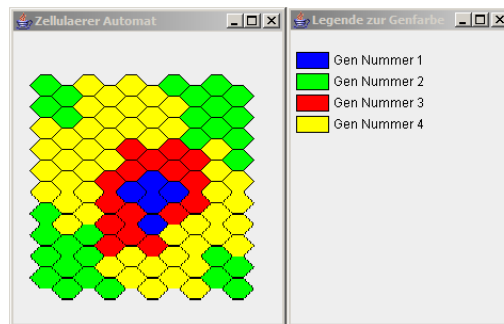


Abbildung 2.5: Beispiel eines Genmusters

Konzentrationsvisualisierung

Hier wird das ganze Zellmuster für ein einzelnes Gen verwendet. Es braucht also gleich viele Fenster wie Gene, um alles darstellen zu können. Diese Darstellung ist immer im Zusammenhang mit einer Legende zu verstehen, die angibt, was die einzelnen Farben bedeuten, v.a. im Zusammenhang mit der adaptiven Legende. Ein Beispiel für diese Visualisierungsart findet man in Abbildung 2.6.

2.2.3 Konsolenausgabe

Auf der Konsole kann man jeden Optimierungsschritt mitverfolgen. Es wird die Fitness für jedes Populationselement¹ nach dem Sortieren² und nach der Envi-

¹Entspricht einem Genmodell

²Damit ist gemeint, dass man alle Genmodelle nach der Höhe ihrer Fitness in absteigender Reihenfolge sortiert

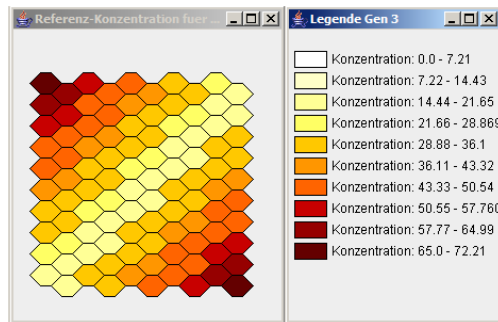


Abbildung 2.6: Beispiel für die Konzentrationsvisualisierung

romental Selection angegeben (Siehe Abschnitt 2.4). Zudem sieht man, wieviele Simulationsiterationen (Siehe Abschnitt 2.3) für jedes Modell erforderlich waren, bis es stabil wurde. Diese Werte beziehen sich auf die Modelle nach der Enviromental Selection. Der letzte Parameter, *Mean Fault*, gibt den Mittelwert der Konzentrations-Differenz³ über alle Gene und Zellen zum Zielmuster an. Ein Beispiel steht in Tabelle 2.1.

```
[1.] Optimierungs-Iteration
fitnessen nach sortiervorgang:  2.0  0.18  0.1  0.03
fitnessen nach env.selektion:   2.0  0.18  0.1  0.03
Anzahl Iterationen bis stabil: 221  184  410  176
Beste Fitness: 2.0085005175845425
Mean Fault: 47.59474428213822
```

Tabelle 2.1: Konsolenoutput während der Optimierung

Am Ende einer Optimierung oder nach dem Abbruch durch den Benutzer wird eine genaue Übersicht über die Simulationsergebnisse ausgegeben. Es werden alle Genkonzentrationswerte, also sowohl Referenzkonzentration als auch optimierte Konzentration, ausgegeben. Interessant sind u.a. folgende Punkte:

Min Diff: Die kleinste Differenz zwischen Zielmuster und optimiertem Muster, die über alle Gene in einer Zelle vorkommt.

Max Diff: Analog zur kleinsten Differenz, wird hier zusätzlich die Zelle angegeben, die die grösste Differenz hervorruft.

Diff: Die gesamte, über alle Zellen aufsummierte Differenz zum Referenzmuster.

Maximal je erreichte Konzentration: Unter diesem Punkt steht für jedes Gen die maximale Konzentration, die während der Modellsimulation je aufgetreten ist.

Bestes Modell: In diesem Abschnitt stehen die Parameter für das Modell, das die grösste Fitness erzielt hat. Dieses Modell wird unter dem Namen "bestModel.txt" im aktuellen Verzeichnis abgespeichert. Somit kann man es später laden, um sich z.B. die Simulation graphisch darstellen zu lassen.

³Mit "Differenz" ist im folgenden stets der Unterschied in der Genkonzentration gemeint.

2.2.4 File I/O

Das Programm erlaubt dem Benutzer, die besten, vom EA gefundenen Modelle abzuspeichern und wieder zu laden. Dadurch wird es auch möglich, vom Benutzer erzeugte Genmodelle zu laden. In der aktuellen Implementation ist das zwar möglich, aber mit einigen Schwierigkeiten verbunden. Man muss die Formatierung strengstens befolgen. Eine genauere Beschreibung zur Editierung solcher Files findet man in jedem *bestModel.txt*-File das vom Programm generiert wurde. Zu beachten ist, dass man nach einer Optimierung das File umbenennen oder verschieben sollte, weil es sonst im nächsten Optimierungsrun einfach überschrieben wird.

2.3 Zellulärer Automat

2.3.1 CA Modelle

Es wurde ein Automat mit einer sechseckigen Zellstruktur verwendet. Jede Zelle interagiert nach einer definierten Transitionsfunktion mit jeder seiner Nachbarzellen. Diese Interaktion wurde in zwei verschiedenen CA Modellen realisiert, die sich in ihrer Transitionsfunktion unterscheiden. Mit CA Modellen sind hier verschiedene Softwareimplementationen des Zellulären Automaten gemeint. Die Transitionsfunktion besteht aus drei Teilen.

1. Diffusion
2. Eigenproduktion / Abbau
3. Interaktion

Die Diffusion beschreibt, wie ein Gen sich in andere Zellen verteilt. Die Eigenproduktion und der Abbau bestimmen die Konzentrationsveränderung des Gens ohne Einwirkung anderer Gene. Die eigentliche Interaktion bestimmt die Topologie des Netzwerks. Durch sie wird die gegenseitige Beeinflussung und die damit verbundene Veränderung der Genkonzentrationen realisiert.

CA Modell A

Jedes Gen hat einen festen Diffusionsparameter, der besagt, wieviel Prozent der aktuellen Genkonzentration in der Zelle bei jedem Schritt auf die Nachbarzellen verteilt wird. Der Parameter ist dabei unabhängig von der Position der Zelle. Aus einer Randzelle mit wenig Nachbarn diffundiert also der gleiche prozentuale Anteil wie aus Zellen im Zentrum. Der Wertebereich für den Diffusionsparameter geht von 0 bis 100 und ist als Prozentwert zu verstehen. Der Genanteil, der in jede Nachbarzelle diffundiert, ist durch folgende Formel gegeben:

$$\Delta Diff = \frac{Diffusionskonstante * Genkonzentration}{\#Nachbarn} \quad (2.1)$$

Die Eigenproduktion und der Abbau wurde in diesem CA Modell in einem Parameter zusammengefasst, er ist charakteristisch für jedes Gen. Dieser Parameter ist eine feste, absolute Grösse, die zwischen 0 und 100 variiert, und die

Abnahme bzw. Zunahme der Genkonzentration beeinflusst. In jedem Zeitschritt verändert sich die Genkonzentration um diesen Wert.

Die Topologie wird in diesem CA Modell, wie auch im zweiten CA Modell (CA Modell B), durch eine quadratische Matrix realisiert. In der Tabelle 2.7 ist ein Beispiel mit den drei Genen A, B und C gegeben. Jede Zeile der Matrix steht für ein Gen und welchen Einfluss es auf die anderen ausübt. Die Diagonalelemente dieser Matrix müssen alle den Wert Null haben, da die Auto-Interaktion bereits durch den Eigenproduktionsparameter modelliert ist. Die Matrixwerte

$$\begin{array}{c}
 \text{Beeinflusste} \\
 \text{Gene} \\
 \text{A} \quad \text{B} \quad \text{C} \\
 \text{Beeinflussende} \\
 \text{Gene} \\
 \text{A} \\
 \text{B} \\
 \text{C}
 \end{array}
 \begin{pmatrix}
 0 & T_{BA} & T_{CA} \\
 T_{AB} & 0 & T_{CB} \\
 T_{AC} & T_{BC} & 0
 \end{pmatrix}$$

Abbildung 2.7: Interaktionsmatrix

repräsentieren einen prozentualen Anteil von 0 bis 100. Dieser Prozentsatz von der Konzentration des beeinflussenden Gens wird berechnet und je nach Vorzeichen von der Konzentration des zu beeinflussenden Gens hinzu addiert bzw. subtrahiert.

CA Modell B

Die Diffusion ist gegenüber dem CA Modell A unverändert.

Eigenproduktion und Abbau sind in diesem Modell hingegen getrennte Parameter. Jedes Gen hat einen festen Parameter für den Abbau. Der Wertebereich für diesen Parameter geht von 0 bis 100, wobei die Zahlen als Prozentwerte zu verstehen sind. In jedem Iterationsschritt wird der entsprechende prozentuale Anteil von der aktuellen Konzentration abgezogen.

Die Eigenproduktion und die Interaktion sind in diesem CA Modell eng verknüpft. Die Interaktion beschreibt hier, wie stark die Eigenproduktion eines Gens von den anderen Genen beeinflusst wird in Abhängigkeit derer Genkonzentrationen. Wie schon beim CA Modell A wurde die Interaktion durch eine quadratische Matrix realisiert. Jedes Gen hat wiederum einen festen Wert für die Eigenproduktion, der sich im Bereich von 0 bis 100 bewegt. Diese Angaben sind als absolute Konzentrationswerte zu verstehen. Die maximale Selbstproduktion ohne Interaktion kann demzufolge maximal 100 betragen, was einer Erhöhung der Konzentration um 100 entspricht. Die Interaktionsformel wurde, leicht verändert, aus dem Artikel von Jönsson [2] übernommen. Es handelt sich dabei um die Formel (5) aus diesem Artikel. Die Änderung bezieht sich dabei ausschliesslich auf die Funktion $g(x)$.

$$\Delta A = g(S_A + T_{AB}B + T_{AC}C) - d_A A \quad (2.2)$$

wobei

$$g(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2.3)$$

Erklärungen zu den verwendeten Variablen können aus Tabelle 2.2 entnommen werden.

Parameter	Definition
X	Genkonzentration von X
S_X	Eigenproduktion von X
T_{XY}	Interaktionsparameter (wie wirkt Y auf X)
d_X	Abbau von X

Tabelle 2.2: Parameterdefinitionen

Vergleich der CA Modelle A und B

Für die durchgeführten Experimente wurde nur das CA Modell B verwendet. Es liefert bessere Resultate als CA Modell A und stellt die Wirklichkeit realistischer dar. Daher gibt es keine Möglichkeit, das CA Modell A im GUI zu wählen.

2.3.2 Funktionsweise

Der Zelluläre Automat hat jeweils drei Zeitschritte gespeichert: die Genkonzentration aus dem vorhergehenden Zeitschritt, die aktuelle Genkonzentration und jene für den nächsten Schritt. In jedem Iterationsschritt werden die Genkonzentrationen für den nächsten Zeitschritt berechnet und am Ende als aktueller Wert gespeichert. Am Ende jedes Iterationsschrittes wird zudem das Muster auf seine Stabilität hin geprüft. Als Bedingung hierbei dient, dass alle drei Zeitschritte sich in ihrer Differenz unterhalb einer gewünschten Toleranzgrenze bewegen müssen um als stabil zu gelten. Das genaue Stabilitätskriterium ist aus Tabelle 2.3 ersichtlich.

Stabilitätskriterium	
$ Genkonzentration[t] - Genkonzentration[t - 1] $	$\leq Toleranz$
$ Genkonzentration[t + 1] - Genkonzentration[t] $	$\leq Toleranz$
$ Genkonzentration[t + 1] - Genkonzentration[t - 1] $	$\leq Toleranz$

Tabelle 2.3: Stabilitätskriterium

Abhängig von den Benutzereingaben wird in jedem Iterationsschritt geprüft, ob die Genkonzentration eine obere Schranke erreicht hat, und allfällige Überschwinger werden korrigiert.

Aufgrund der verwendeten begrenzten Zellstruktur⁴ haben nicht alle Zellen die gleiche Anzahl Nachbarn. Daher kann man zwischen zwei verschiedenen Ansätzen wählen. Auf der einen Seite kann man mit normaler Randbedingung rechnen und so tun, als ob es ausserhalb des dargestellten Zellmusters noch weitere Zellen gibt. Auf der anderen Seite wird bei der speziellen Randbedingung die Zellstruktur als endlich angenommen; es kommen also echte Randzellen vor. In diesem Ansatz geht keine Genkonzentration "verloren", da nichts in den nicht sichtbaren Bereich diffundieren kann.

⁴Es wäre z.B. auch eine torusförmige Zellstruktur denkbar, durch welche die Zellen links und rechts am Rand miteinander verbunden würden.

2.3.3 Validierung der Komponenten

Für die Gewährleistung der Korrektheit aller Programmteile wurde eine Validierung vorgenommen. Im Einzelnen wurde geprüft: die Diffusion (mitsamt ihrer Kernfunktion *getNeighbours()*), die Selbstproduktion / Abbau und die Interaktion. Zu diesem Zweck wurde auch eine Validierung mit ExcelTM durchgeführt, welche Anhang 4 zu entnehmen ist.

Diffusion

Das Kernstück der Diffusion ist die Funktion *getNeighbours()*, sie bekommt die Koordinaten (x,y) einer Zelle und gibt ihre Nachbarn zurück. Die Schwierigkeit bei dieser Funktion lag bei den Randzellen und falls z.B. die Höhe und / oder die Breite der Zellstruktur nur Eins beträgt. Für die Validierung wurde ein separates Testprogramm geschrieben. Als ersten Schritt stellt man die gewünschte Zellstruktur ein, dann kann man einzelne Koordinaten übergeben und das Programm zeichnet die jeweiligen Nachbarn ein. In Abbildung 2.8 sind einige Beispiele gegeben. Die eigentliche Zelle wurde dabei rot markiert, ihre Nachbarn mit grau.

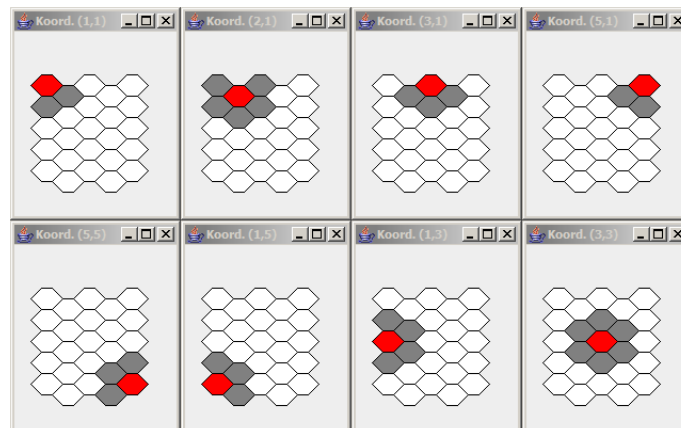


Abbildung 2.8: Beispiele spezieller Nachbarschaftsbeziehungen

Bei der Überprüfung der Diffusion wurde zuerst der genaue Ablauf von Hand berechnet. Es wurden einige wenige Zellen mit einer bestimmten Konzentration gefüllt und bestimmt, wie es im nächsten Zeitschritt aussehen sollte. Das Gleiche wurde dann im Programm simuliert und die Ergebnisse überprüft. Auch wurden qualitative Überprüfungen gemacht ob das bildlich dargestellte Resultat intuitiv logisch erscheint. Teilweise mit der speziellen Randbedingung und teilweise mit der normalen Randbedingung. Je ein Beispiel einer Diffusion mit spezieller und normaler Randbedingung ist in Abbildung 2.9 und 2.10 dargestellt. In diesen Bildern ist ersichtlich, wie sich die Diffusion verhält, und was die spezielle Randbedingung ausmacht. Alle 50 Iterationsschritte wurde die Konzentration graphisch dargestellt. In beiden Beispielen wurde ein Diffusionsparameter von 10 eingestellt, und die Genkonzentration der Zellen auf der linken Seite wurde auf 100 gesetzt.

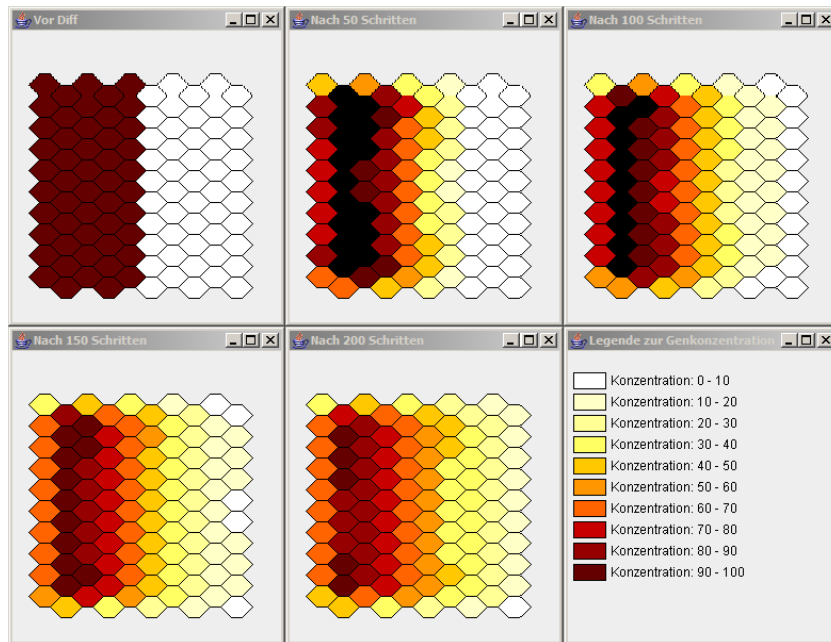


Abbildung 2.9: Diffusionsbeispiel mit spezieller Randbedingung

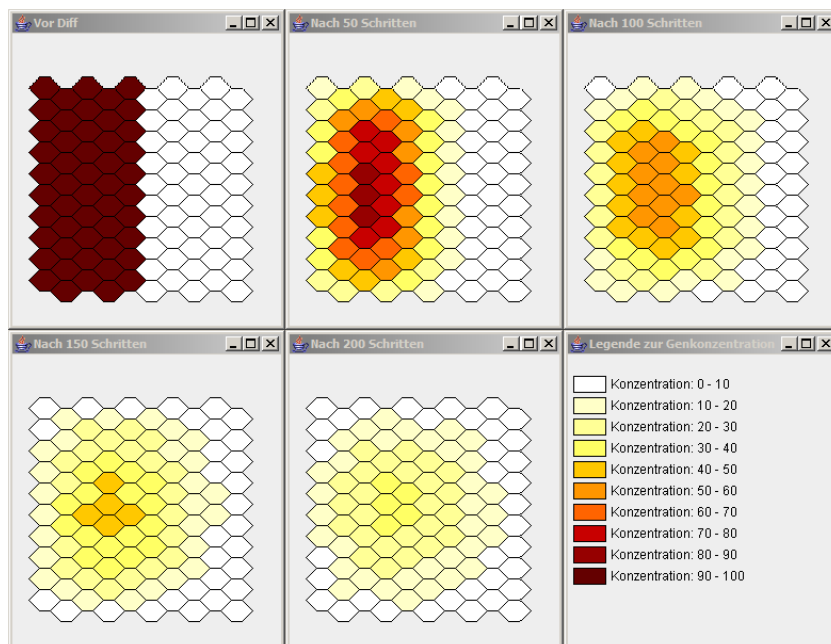


Abbildung 2.10: Diffusionsbeispiel mit normaler Randbedingung

Eigenproduktion und Abbau

Die Validierung dieser Komponenten war vergleichsweise einfach. Wiederum wurde hier das Szenario zuerst von Hand berechnet, und anschliessend anhand einer Simulation überprüft. Dabei war zu beachten, dass nach CA Modell B, das in Abschnitt 2.3.1 beschrieben ist, die Selbstproduktion mit der Interaktion verknüpft ist. Für die Validierung wurden alle Interaktionsparameter auf Null gesetzt. Um den Effekt sichtbar zu machen, wurden die mittleren vier Zellen mit einer Konzentration von 100 gefüllt. Der Selbstproduktionsparameter wurde auf 1 gesetzt, zusätzlich wurde eine Diffusion von 10 eingestellt. Die Konzentrationsverteilung wurde, beginnend nach 30 Iterationsschritten, alle 15 Schritte neu gezeichnet. Der ganze Verlauf ist in Abbildung 2.11 ersichtlich.

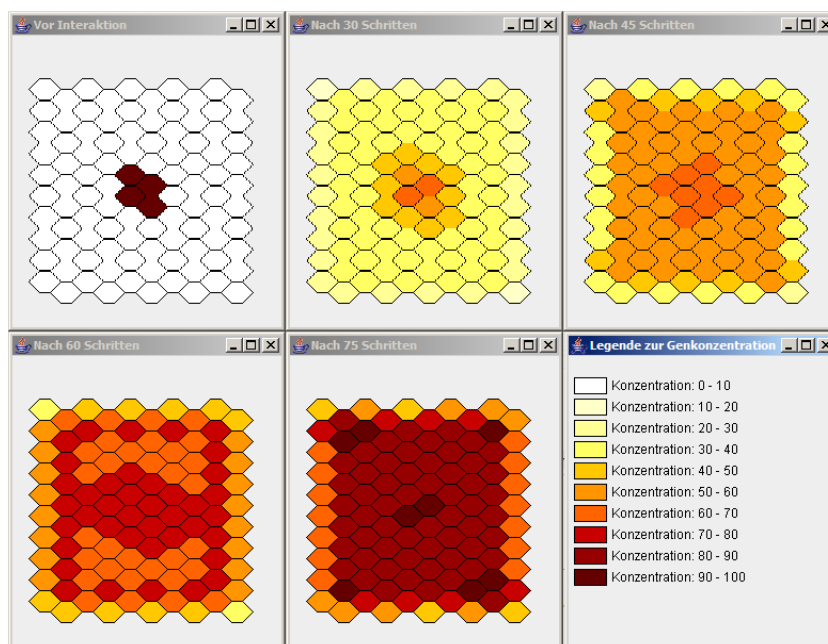


Abbildung 2.11: Selbstproduktion mit Diffusion

Interaktion

Die Validierung dieses Teils war weitaus aufwändiger. Schon bei einer kleinen Zellstruktur mit nur zwei Genen wächst das Problem rasant. Es wurde eine quadratische Zellstruktur der Breite drei mit zwei Genen verwendet. Alle Parameter wurden einfach von Hand gewählt. Das genaue Modell ist aus Tabelle 2.4 zu entnehmen. Überprüft wurden die Ergebnisse auf drei Arten: Einerseits wurden die ersten paar Schritte von Hand ausgeführt. Die zweite Methode bestand aus einer ExcelTM-Tabelle⁵, in der das ganze Modell genau durchgerechnet wurde. Drittens wurde unsere Software verwendet. Alle Ergebnisse wurden miteinander verglichen und auf ihre Übereinstimmung überprüft.

⁵Siehe Anhang 4

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion	
A	10	2	0.5	0	2
B	10	1	0.5	-1	0

Tabelle 2.4: Validationsmodell

Um die Vorgänge zu visualisieren, wurde ausserdem ein einfaches Szenario mit einer quadratischen Zellstruktur der Länge zehn mit zwei Genen simuliert. Die Parameter wurden hier so gewählt, dass sich das eine Gen auf Kosten der Konzentration des anderen ausbreitet. Das genaue Modell ist der Tabelle 2.5 zu entnehmen. In Abbildung 2.12 wird dargestellt, wie Gen A das Gen B verdrängt. Das Gen A ist blau eingefärbt, das Gen B grün. Zu Beginn gibt es keine Durchmischung der Gene, wir haben auf der linken Seite also eine Konzentration von Null für Gen B, genauso wie auf der rechten Seite eine Konzentration von Null für Gen A. Die Konzentrationswerte zu Beginn sind konstant, Gen A wurde in allen Zellen auf 50 gesetzt, Gen B in allen Zellen auf 100. Wie sich die Genkonzentrationen der beiden Gene konkret ändern, ist aus den Abbildungen 2.13 und 2.14 zu entnehmen. Da es zu Beginn ein Weilchen in Anspruch nimmt bis beide Gene ausreichend diffundiert sind, braucht es Zeit bis man etwas erkennen kann. Daher wurde der zweite Screenshot nach 50 Iterationsschritten gemacht, der Dritte nach 68, und weiter im Abstand von zwei Schritten.

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion	
A	5	1	1	0	-10
B	5	1	1	0	0

Tabelle 2.5: Validationsmodell

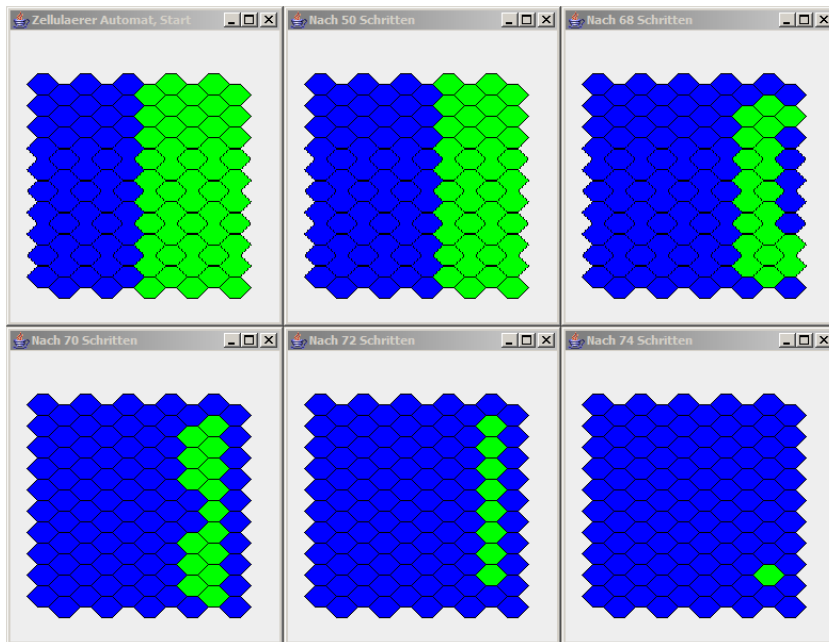


Abbildung 2.12: Verlauf der Genverdrängung

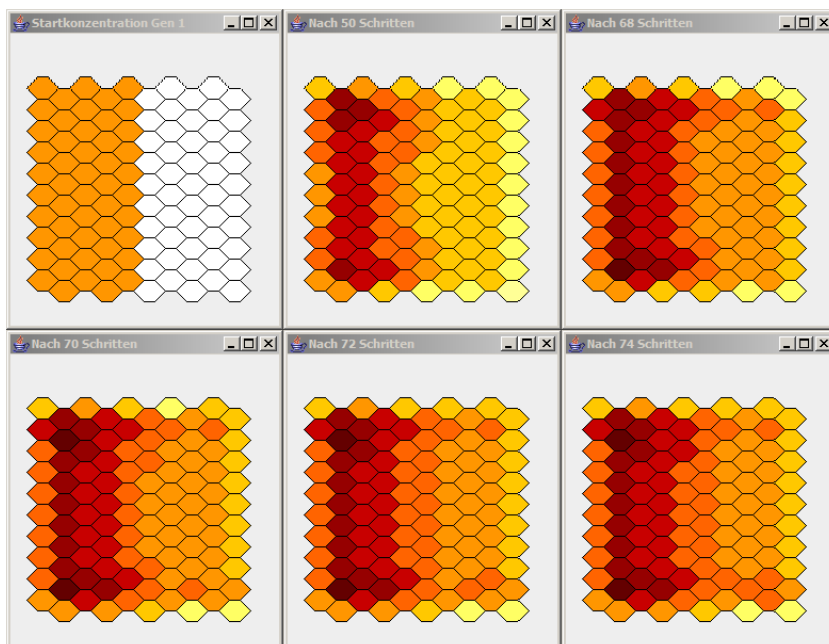


Abbildung 2.13: Konzentrationsverlauf des Gens A

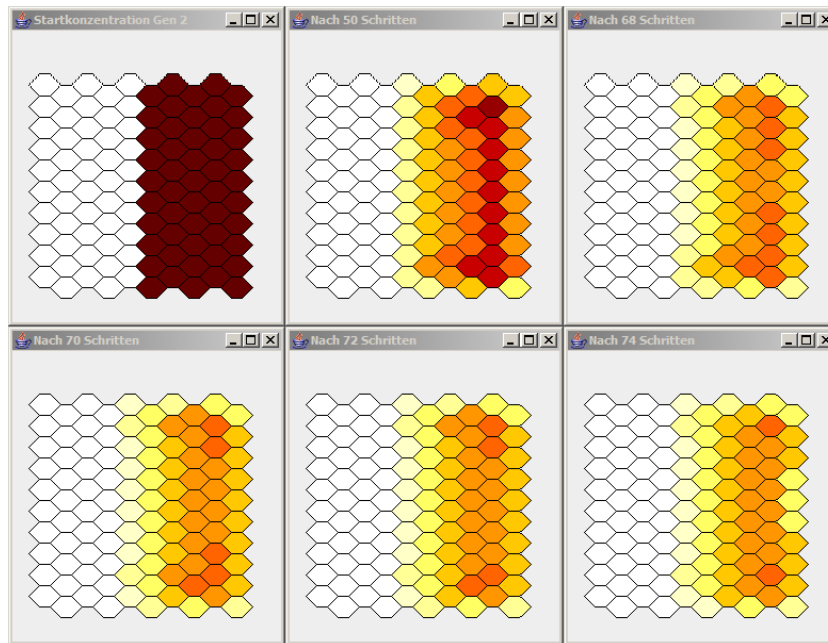


Abbildung 2.14: Konzentrationsverlauf des Gens B

2.4 Evolutionärer Algorithmus

Evolutionäre Algorithmen⁶ sind Optimierungsverfahren, welche der Evolution in der Natur nachempfunden sind. Konkret werden die Fortpflanzung (also Rekombination von Erbgut) sowie die Mutation von Erbgut nachgebildet. Wie in der Evolution wird ebenfalls eine Selektion der stärksten Kandidaten vorgenommen, wobei die schwächeren ausscheiden.

Diese Art von Optimierung wird vor allem dort angewandt, wo eine grosse Anzahl von Parametern angepasst werden müssen und/oder wo nicht im vornherein klar ist, welche Auswirkung die Änderung eines einzelnen Parameter auf das ganze System hat. Bei der hier behandelten Optimierungsaufgabe treffen diese beiden Eigenschaften zu.

In unserem Fall sollen die folgenden Parameter eines Kandidaten-Modells optimiert werden:

- Interaktionsstärken zwischen den Genen
- Diffusion
- Eigenproduktion der Gene
- Abbau der Gene

Für eine genaue Beschreibung dieser Parameter siehe Abschnitt 2.3.1.

⁶Auch genetische Algorithmen/Verfahren genannt.

Die Population des EA besteht in unserer Software aus Modellen, welche oben beschriebene Parameter enthalten. Diese Population wird in einem Array der dreifachen Populationsgrösse abgelegt, da stets doppelt so viele Kinder wie Eltern erschaffen werden. Desweiteren können die Elemente des Kinder-Array-Teils als "leer" und jene des Eltern-Array-Teils als "für Rekombination ausgewählt" markiert werden. Diese Unterteilung wird anhand einer Beispiel-Population der Grösse 4 in Abbildung 2.15 dargestellt.

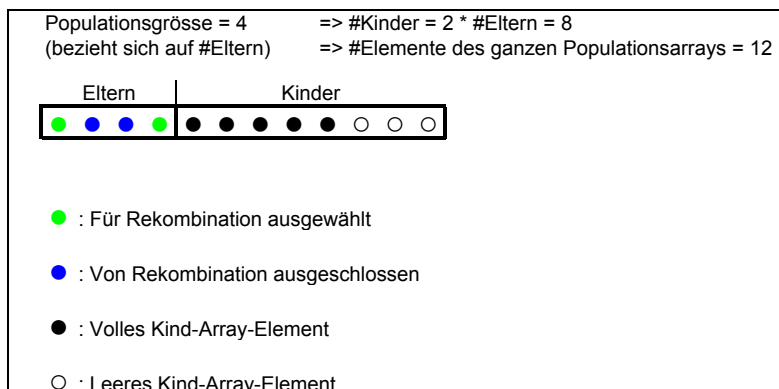


Abbildung 2.15: Aufbau der Modell-Population

Das Ziel der Optimierung ist, zu einem gegebenen Genmuster (Zielmuster) ein Gennetzwerk zu finden, welches möglichst genau dieses Muster generiert. Dazu werden zuerst mehrere Zufallsmodelle simuliert und deren Endmuster bewertet, indem der jeweilige Abstand zum Zielmuster berechnet wird. Dann werden diese Modelle mutiert und untereinander rekombiniert, um neue Modelle zu schaffen. Nach diesem Schritt werden nur die besten Modelle⁷ in die nächste Optimierungsiteration übernommen und schliesslich wiederholen sich die Schritte Fortpflanzungsauswahl (Mating Selection), Rekombination/Mutation, Bewertung (Fitness-Evaluation) und Selektion nach Fitness.

2.4.1 Übersicht

In Abbildung 2.16 ist der Ablauf des Optimierungsverfahrens dargestellt.

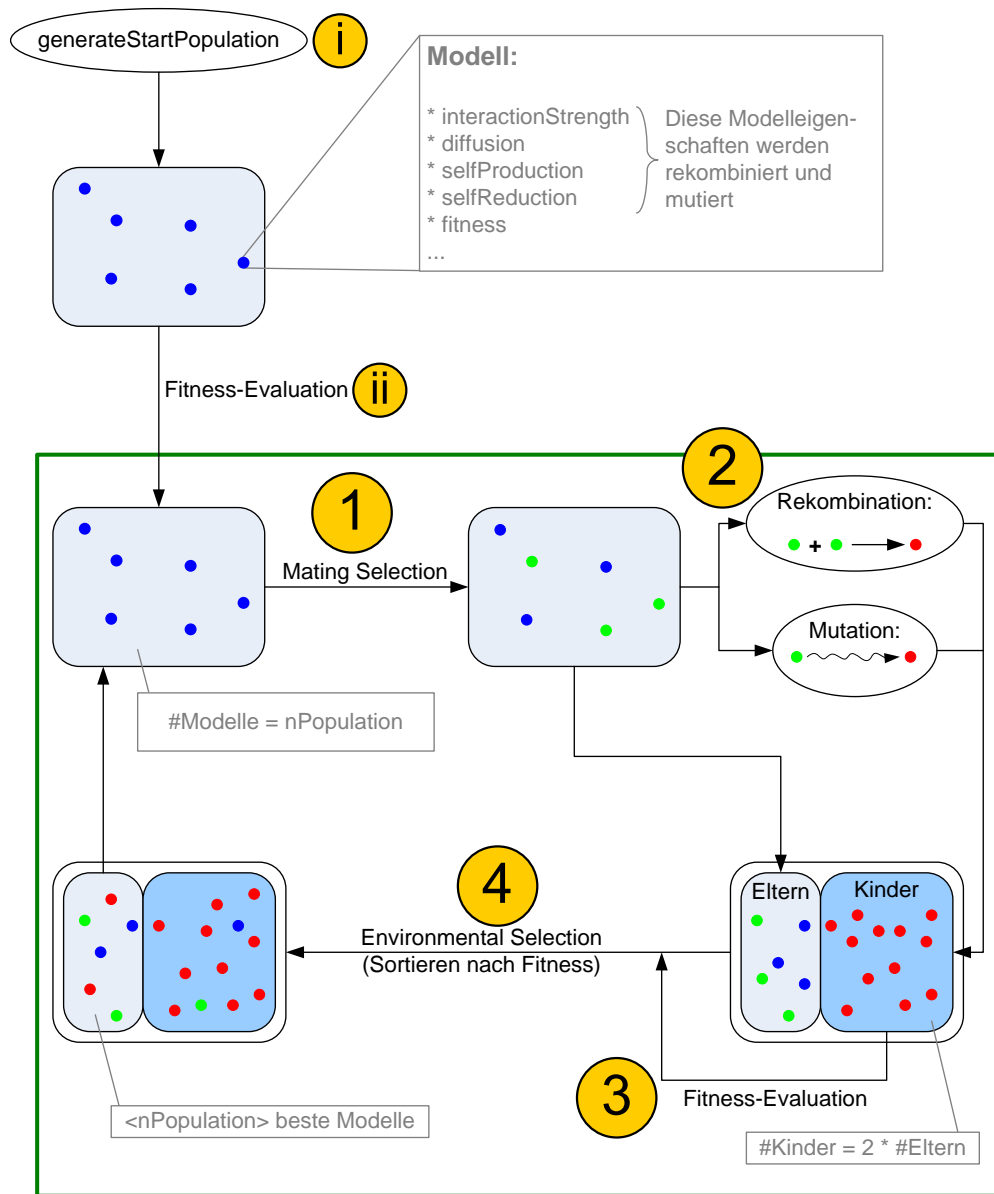
Die Schritte **i** und **ii** werden lediglich einmal beim Starten des Optimierers durchgeführt, um eine Anfangspopulation zu schaffen und deren Elementen eine Fitness zuzuweisen. Hingegen die Schritte **1** bis **4** werden wiederholt ausgeführt und stellen die eigentliche Optimierung dar.

In Schritt **2** wird wiederholt zufällig entweder eine Rekombinationen bzw. eine Mutation durchgeführt bis genügend⁸ "Kind-Modelle" erstellt worden sind.

Die einzelnen Schritte werden in den folgenden Abschnitten detailliert erläutert.

⁷Ausnahme: Siehe Abschnitt 2.4.7, "fewBest"

⁸Genügend bedeutet hier, dass doppelt so viele Kinder wie Eltern vorhanden sind.



Dieser Block wird wiederholt ausgeführt (→ #Optimierungsiterationen)

Abbildung 2.16: Schritte des evolutionären Optimierungsverfahrens

2.4.2 Generierung der Startpopulation

Um eine Startpopulation zu schaffen, wird eine der Populationsgrösse entsprechende Anzahl von Zufallsmodellen erstellt. Die Zufallswerte (Prozent-Zahlen) der einzelnen Variablen liegen dabei in folgenden Intervallen:

- $interactionStrength[i][j] \begin{cases} \in]-100,100[& \text{für } i \neq j \\ = 0, & \text{sonst} \end{cases}$
- $diffusion[i] \in [0,100[$
- $selfProduction[i] \in [0,100[$
- $selfReduction[i] \in [0,100[$

Aufgrund der Tatsache, dass es sehr unwahrscheinlich ist, dass der Zufallsgenerator gerade 0 als Wert zurückgibt, wird ein Interaktionswert mit einer Wahrscheinlichkeit von 10% auf 0 gesetzt. Ansonsten wird der Wert per Zufall bestimmt. Diese Korrektur wurde deshalb eingebaut, da in Wirklichkeit viele Gene nicht in Interaktion miteinander stehen.

2.4.3 "Mating Selection"

Während der "Mating Selection" wird entschieden, welche Kandidaten untereinander kombiniert werden dürfen, um Kinder zu schaffen. Diese Auswahl geschieht im sogenannten Tournament-Verfahren, welches wie folgt abläuft:

1. Zwei Eltern-Kandidaten werden zufällig ausgewählt.
2. Der bessere dieser zwei Kandidaten wird für Rekombination ausgewählt, der andere (für die aktuelle Optimierungsiteration) davon ausgeschlossen.
3. Die Schritte 1 und 2 werden wiederholt bis alle Eltern-Kandidaten entweder gewählt oder abgewählt wurden bzw. bis nur ein Kandidat übrig bleibt. Das allenfalls übrigbleibende Modell wird sowieso selektiert.

Handelt es sich um eine ungerade Populationsgrösse, kann das Tournamentverfahren nicht korrekt angewandt werden, da es auf Pärchenbildung beruht. In diesem Fall wird der übriggebliebene, nicht markierte Kandidat mitselektiert.

In Abbildung 2.17 ist das Tournamentverfahren für ganzzahlige Fitness-Werte und einer ungeraden Anzahl Kandidaten gezeigt.

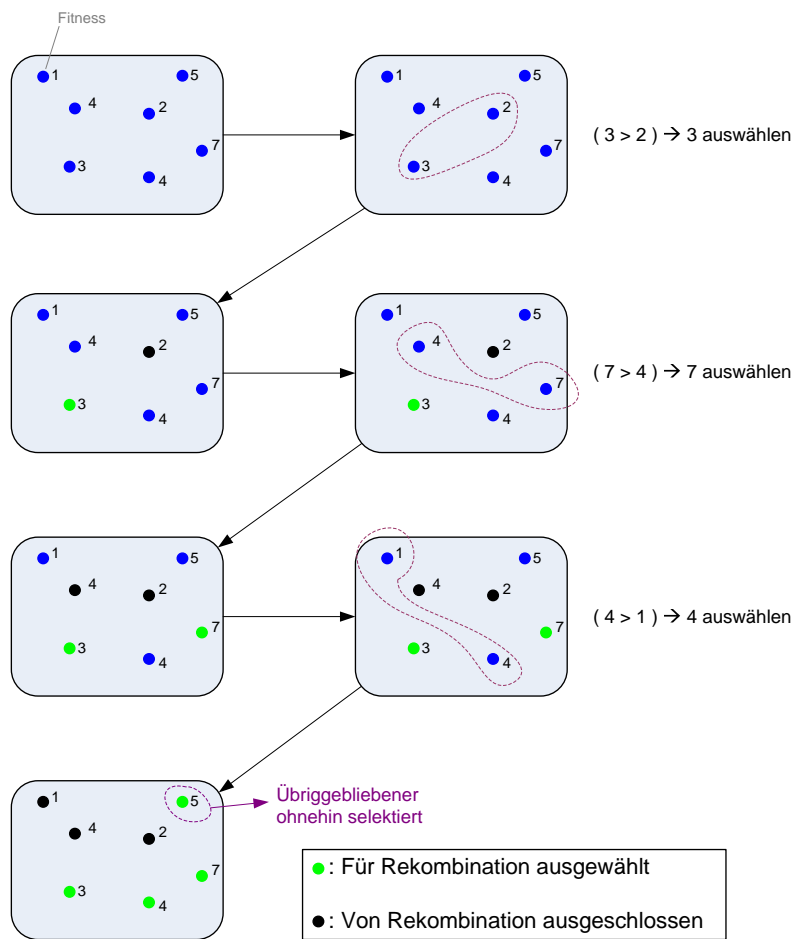


Abbildung 2.17: Beispiel-Ablauf zum Tournament-Verfahren

2.4.4 Rekombination

Bei der Rekombination von Modellen werden die Parameter - also Interaktionsstärke/Diffusion/Eigenproduktion/Abbau der Gene - zweier Modelle miteinander vermischt. Dabei entscheidet der Zufall, ob einer, zwei, drei oder alle vier Parameter kombiniert werden. Da die Parameter selbst Vektoren bzw. Matrizen sind, wird innerhalb dieser wieder eine Zufallsauswahl der ausgetauschten Werte getroffen.

Das Vorgehen sieht folgendermassen aus:

1. Zwei Modelle der Eltern-Population, welche während der "Mating Selection" (siehe Abschnitt 2.4.3) für Rekombination ausgewählt wurden, werden zufällig ausgewählt.
2. Im Kinder-Teil des Populationsarrays wird ein leerer Platz gesucht (siehe Abbildung 2.15).
3. Es wird eine Zufallsauswahl der zu durchmischenden Parameter getroffen.
4. Innerhalb der im vorhergehenden Schritt ausgewählten Parameter werden die Werte zufällig ausgewählt, die getauscht werden sollen.
5. Das Kind-Modell übernimmt die Werte des ersten Eltern-Modells. Dann wird schrittweise durch die auszutauschenden Parameter durchgegangen und jene Parameterwerte des 2.Elternteils ins Kind übernommen, welche zum Austausch markiert wurden.
6. Falls das entstandene Modell bereits existiert, wird bei Schritt 1 neu gestartet. Nach 1000 misslungenen Rekombinationsversuchen wird die Rekombination aus Effizienzgründen für die aktuelle Optimierungsiteration deaktiviert und nur noch Mutation zur Generierung von Kinder-Modellen verwendet.
7. Das anfangs gewählte Kind wird als "besetzt" markiert (siehe Abbildung 2.15)

Schritt 5 wird in Abbildung 2.18 veranschaulicht.

Legende:

- : Für Rekombination (zufällig) ausgewählter Parameter
- xx : Für Rekombination (zufällig) ausgewählter Parameter-Wert
- : Im aktuellen Schritt geänderter Parameter
- xx : Im aktuellen Schritt geänderter Parameter-Wert

	1.Eltern-Kandidat	2.Eltern-Kandidat																												
Interaktion:	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">0</td><td style="width: 50%; text-align: center;">30</td></tr> <tr><td style="width: 50%; text-align: center;">7</td><td style="width: 50%; text-align: center;">0</td></tr> </table>	0	30	7	0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">0</td><td style="width: 50%; text-align: center;">36</td></tr> <tr><td style="width: 50%; text-align: center;">19</td><td style="width: 50%; text-align: center;">0</td></tr> </table>	0	36	19	0																				
0	30																													
7	0																													
0	36																													
19	0																													
Diffusion:	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">25</td></tr> <tr><td style="width: 50%; text-align: center;">13</td></tr> </table>	25	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">43</td></tr> <tr><td style="width: 50%; text-align: center;">27</td></tr> </table>	43	27																								
25																														
13																														
43																														
27																														
Eigenproduktion:	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">34</td></tr> <tr><td style="width: 50%; text-align: center;">65</td></tr> </table>	34	65	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">86</td></tr> <tr><td style="width: 50%; text-align: center;">12</td></tr> </table>	86	12																								
34																														
65																														
86																														
12																														
Abbau:	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">23</td></tr> <tr><td style="width: 50%; text-align: center;">81</td></tr> </table>	23	81	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">12</td></tr> <tr><td style="width: 50%; text-align: center;">7</td></tr> </table>	12	7																								
23																														
81																														
12																														
7																														
Kind-Entwicklung																														
	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">0</td><td style="width: 50%; text-align: center;">30</td></tr> <tr><td style="width: 50%; text-align: center;">7</td><td style="width: 50%; text-align: center;">0</td></tr> </table>	0	30	7	0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">25</td></tr> <tr><td style="width: 50%; text-align: center;">13</td></tr> </table>	25	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">34</td></tr> <tr><td style="width: 50%; text-align: center;">65</td></tr> </table>	34	65	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">23</td></tr> <tr><td style="width: 50%; text-align: center;">81</td></tr> </table>	23	81	(1 Elternteil übernommen) Schritt 1	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">0</td><td style="width: 50%; text-align: center;">36</td></tr> <tr><td style="width: 50%; text-align: center;">7</td><td style="width: 50%; text-align: center;">0</td></tr> </table>	0	36	7	0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">25</td></tr> <tr><td style="width: 50%; text-align: center;">13</td></tr> </table>	25	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">34</td></tr> <tr><td style="width: 50%; text-align: center;">65</td></tr> </table>	34	65	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">23</td></tr> <tr><td style="width: 50%; text-align: center;">81</td></tr> </table>	23	81	Schritt 2
0	30																													
7	0																													
25																														
13																														
34																														
65																														
23																														
81																														
0	36																													
7	0																													
25																														
13																														
34																														
65																														
23																														
81																														
	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">0</td><td style="width: 50%; text-align: center;">36</td></tr> <tr><td style="width: 50%; text-align: center;">7</td><td style="width: 50%; text-align: center;">0</td></tr> </table>	0	36	7	0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">43</td></tr> <tr><td style="width: 50%; text-align: center;">13</td></tr> </table>	43	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">34</td></tr> <tr><td style="width: 50%; text-align: center;">65</td></tr> </table>	34	65	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">23</td></tr> <tr><td style="width: 50%; text-align: center;">81</td></tr> </table>	23	81	Schritt 3	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">0</td><td style="width: 50%; text-align: center;">36</td></tr> <tr><td style="width: 50%; text-align: center;">7</td><td style="width: 50%; text-align: center;">0</td></tr> </table>	0	36	7	0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">43</td></tr> <tr><td style="width: 50%; text-align: center;">13</td></tr> </table>	43	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">34</td></tr> <tr><td style="width: 50%; text-align: center;">65</td></tr> </table>	34	65	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">12</td></tr> <tr><td style="width: 50%; text-align: center;">81</td></tr> </table>	12	81	Schritt 4
0	36																													
7	0																													
43																														
13																														
34																														
65																														
23																														
81																														
0	36																													
7	0																													
43																														
13																														
34																														
65																														
12																														
81																														
	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">0</td><td style="width: 50%; text-align: center;">36</td></tr> <tr><td style="width: 50%; text-align: center;">7</td><td style="width: 50%; text-align: center;">0</td></tr> </table>	0	36	7	0	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">43</td></tr> <tr><td style="width: 50%; text-align: center;">13</td></tr> </table>	43	13	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">34</td></tr> <tr><td style="width: 50%; text-align: center;">65</td></tr> </table>	34	65	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="width: 50%; text-align: center;">12</td></tr> <tr><td style="width: 50%; text-align: center;">7</td></tr> </table>	12	7	Schritt 5															
0	36																													
7	0																													
43																														
13																														
34																														
65																														
12																														
7																														

Abbildung 2.18: Beispiel-Ablauf zur Rekombination

2.4.5 Mutation

Das Vorgehen bei der Mutation von Modellen ist jenem der Rekombination ähnlich. Der Unterschied liegt darin, dass bloss ein(!) Eltern-Modell mit Zufallszahlen kombiniert wird. Das Ganze läuft wie folgt ab:

1. Ein beliebiges Modell der Eltern-Population wird zufällig ausgewählt. Es kommen auch jene in Frage, welche nicht für Rekombination selektiert wurden.
2. Im Kinder-Teil des Populationsarrays wird ein leerer Platz gesucht (siehe Abbildung 2.15).
3. Es wird eine Zufallsauswahl der zu mutierenden Parameter getroffen.
4. Innerhalb der im vorhergehenden Schritt ausgewählten Parameter werden die Werte zufällig ausgewählt, die mutiert werden sollen.
5. Das Kind-Modell übernimmt die Werte des Eltern-Modells. Dann wird schrittweise durch die zu mutierenden Parameter durchgegangen und jene Parameterwerte geändert, welche für Mutation markiert wurden. Die einzelnen Parameterwerte mutieren - wie auch in Abbildung 2.19 veranschaulicht - auf folgende Weise:
 - Interaktionsmatrix: Ein Interaktionswert wird mit einer Wahrscheinlichkeit von 10% auf 0 gesetzt, falls er nicht schon vorher gleich Null war. Andernfalls wird zum aktuellen Wert eine Zufallszahl zwischen -40 und +40 dazu addiert. Wenn nach dieser Addition ein Wert kleiner als -100^9 resultiert, wird die entsprechende Variable auf eine zufällige Zahl zwischen -100 und -90 gesetzt. Sollte die vorher beschriebene Addition einen Wert grösser als 100 ergeben, wird ein Wert zwischen 90 und 100 eingesetzt
 - Diffusion: Ein zufälliger Wert zwischen -40 und +40 wird addiert. Ist das Resultat kleiner als 0, wird eine Zufallszahl zwischen 0 und 10 eingetragen. Falls eine Zahl grösser als 100 resultiert, wird die Diffusion auf einen zufälligen Wert zwischen 90 und 100 gesetzt.
 - Eigenproduktion: Wie Diffusion
 - Abbau: Wie Diffusion
6. Das anfangs gewählte Kind wird als "besetzt" markiert (siehe Abbildung 2.15)

⁹Die Interaktionswerte müssen zwischen -100 und +100 liegen

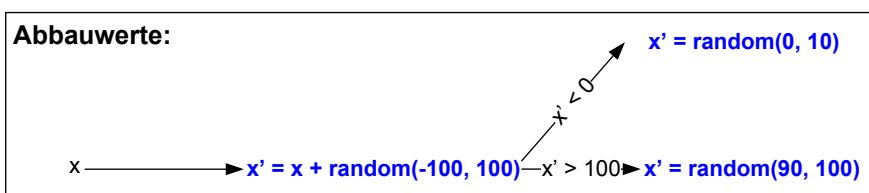
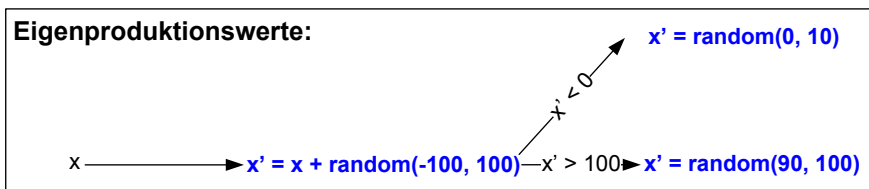
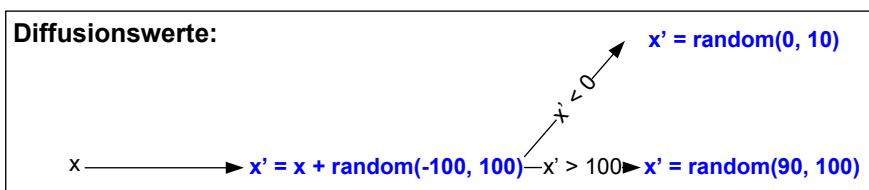
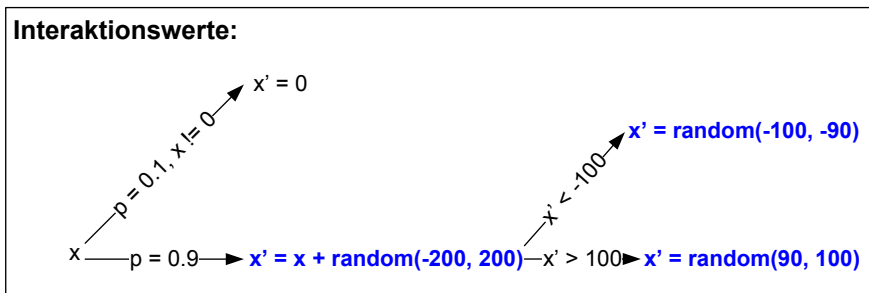


Abbildung 2.19: Arten von Parameter-Mutationen

2.4.6 Fitness-Evaluation

Um ein Modell bewerten zu können, wird der zelluläre Automat (siehe Kapitel 2.3) damit gestartet und bis zur Stabilität des Musters bzw. bis zur angegebenen maximalen Anzahl Simulationsschritte (siehe Abschnitt 2.2) laufen gelassen. Ein Modell, bei welchem die maximale Anzahl Simulationsiterationen überschritten wird, wird als instabil eingestuft und erhält folglich in der Bewertung einen Straffaktor von 0.1. Die Eigentliche Bewertung beruht auf dem Vergleich des bei der Simulation erhaltenen Endmusters und des vorgegebenen Zielmusters.

Die Ähnlichkeit zweier Muster wird dabei auf einen einzigen Wert, den sogenannten Fitness-Wert abgebildet. Zur Bildung dieser Bewertung wurden verschiedene Verfahren¹⁰ entwickelt und evaluiert. Bei den meisten Verfahren wurde zuerst der Abstand zum Zielmuster berechnet und dann der Kehrwert davon als Fitness-Wert angenommen, da man schliesslich eine möglichst kleine Differenz zum Zielmuster erhalten möchte.

Die verwendeten Algorithmen zur Differenzbildung seien hier kurz erläutert. Es wird dabei lediglich erklärt, welchen Einfluss die Genkonzentrationsdifferenz einer einzigen Zelle eines Gens auf den ganzen Differenzwert hat. Der totale Differenzwert schliesslich besteht aus der Summe all dieser Einzeldifferenzen:

1. Falls Differenz in Genkonzentration > 1 :
$$diff_{total} += \frac{10000}{\#gene \cdot caBreite \cdot caHöhe}$$
2. $diff_{total} += Differenz$
3. Falls $Differenz < 10$:
$$diff_{total} -= \frac{100}{\#gene \cdot caBreite \cdot caHöhe},$$

sonst: $diff_{total} += \frac{Differenz}{\#gene \cdot caBreite \cdot caHöhe}.$
4. Falls $Differenz < 5$ und Zielkonzentration $\neq 0$:
 $diff_{total} -= 2,$

falls $5 \leq Differenz < 10$ und Zielkonzentration $\neq 0$:
 $diff_{total} -= 1,$

falls $10 \leq Differenz < 50$:
 $diff_{total} += 1,$

falls Differenz > 50 :
 $diff_{total} += 2.$

¹⁰Im Programm-Code mit "getDifference1()" bis "getDifference9()" sowie mit "getFitness()" bezeichnet.

5. Zuerst wird von allen Feldern mit Zielkonzentration $\neq 0$ die Minimum-Differenz ($diff_{min}$) gesucht.

Falls $Differenz \leq diff_{min}$ und Zielkonzentration $\neq 0$:

$diff_{total} -= 1$,

sonst: $diff_{total} += 1$

6. Falls $Differenz < 5$ und Zielkonzentration $\neq 0$:

$diff_{total} -= 1$,

falls $Differenz < 5$ und Zielkonzentration = 0:

$diff_{total} -= 0.5$,

falls $Differenz \geq 5$:

$diff_{total} += 1$.

7. Die grösste gefundene Differenz ($diff_{max}$) wird zurückgegeben.

8. Falls Modellkonzentration = 0 und Zielkonzentration $\neq 0$:

$diff_{total} += 1$,

falls Modellkonzentration $\neq 0$ und Zielkonzentration $\neq 0$:

$diff_{total} += \frac{Differenz-10}{5}$,

sonst:

$diff_{total} += \frac{Differenz}{50}$.

Dann:

$$diff_{total} = \frac{diff_{total} + 2 \cdot \#gene \cdot caBreite \cdot caHöhe}{2 \cdot \#gene \cdot caBreite \cdot caHöhe} \cdot 100 + 1$$

9. $diff_{total} += Differenz$

Zusätzlich wird die kleinste und grösste Differenz gesucht. ($diff_{min}, diff_{max}$)

Die zurückgegebene Differenz berechnet sich nach folgender Formel:

$$diff_{total} = \frac{diff_{total} \cdot diff_{min} \cdot diff_{max}}{caHöhe \cdot caBreite \cdot \#Gene}$$

Ausserdem wird bei allen Verfahren zur Differenzbildung überprüft, ob ungültige Differenzwerte auftreten. Dies kann vorkommen, wenn die Genkonzentration zu hoch ist, um als Double-Variable¹¹ gespeichert zu werden. Für solche unerlaubten Werte wird eine Fitness von 0¹² zurückgegeben.

¹¹Maximaler Double-Wert = $1.7976931348623157 \cdot 10^{308}$

¹²bzw. $diff_{total}$ = maximaler Double-Wert

Es hat sich herausgestellt, dass es sehr wichtig ist, kleine Unterschiede in der Genkonzentration zu belohnen, indem zum Beispiel ein bestimmter Wert von der Gesamtdifferenz $diff_{total}$ abgezogen wird. Auf diese Weise werden Muster bevorzugt, welche stellenweise eine gute Übereinstimmung mit dem Zielmuster aufweisen. Hingegen bei einfachen Verfahren wie 1 oder 2 kann es vorkommen, dass gute Übereinstimmungen von schlechten "überdeckt" werden und so nicht mehr am Fitnesswert erkennbar sind.

Aus diesen Gründen wurden schliesslich nur die Verfahren 8 und 9 verwendet. Es ist anzunehmen, dass diese weiter verbessert werden können, wodurch die Güte des gesamten Optimierers gesteigert wird; denn der Vergleich von Zielmuster und optimiertem Muster stellt sozusagen das Herzstück der Optimierung dar.

Für Modelle, welche in instabilen Mustern resultieren, wurde ein Fitness-Straffaktor von 0.1 eingebaut, da deren Endmuster von eher zufälliger Natur ist und nichts über die eigentliche Güte des Modells aussagt. Instabile Modelle wurden aus dem Grund nicht vollständig ignoriert, da es möglich ist, ausgehend von diesen durch Mutation oder Rekombination neue stabile Modelle zu erschaffen.

2.4.7 "Environmental Selection"

Die Aufgabe dieser Selektion ist es, die besten Modelle auszuwählen und in die nächste Eltern-Generation zu übernehmen. Dies geschieht, indem die gesamte Population von Modellen nach Fitness sortiert wird und schliesslich der bessere Teil der resultierenden Population als neue Eltern-Population gesetzt wird. Dies geschieht in einem Schritt durch Sortieren mit dem "Quick-Sort"-Algorithmus, wie Abbildung 2.20 verdeutlicht. Stellvertretend für die Modelle sind nur deren Fitness-Werte in die Grafik eingetragen.

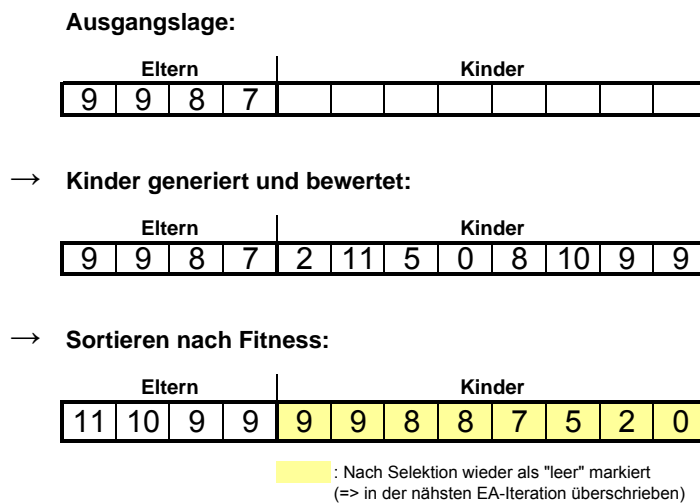


Abbildung 2.20: Vorgehen bei der "Environmental Selection"

Erweiterung

Nach dem bisher beschriebenen Vorgehen des EA werden nach vielen Iterationen kaum mehr bessere Modelle erschaffen und alle Eltern-Modelle scheinen von einem einzigen - besonders gut bewerteten - Modell abzustammen. Ausserdem ist es mit der Zeit immer schwieriger¹³, durch Rekombination neue Modelle zu finden. Aus diesen Gründen wurde ein Feature namens "fewBest" eingebaut, welches dafür sorgt, dass ein Teil der Eltern-Population während der "Environmental Selection" für eine Optimierungsiteration zufällig gesetzt wird. In den folgenden (mindestens 19) EA-Iterationen wird dieser Teil der Eltern per Tournamentverfahren ausgewählt.

Diese Erweiterung wird aktiviert, wenn sich der Fitness-Wert des besten Modells über 20 Iterationen hinweg nicht ändert. In Abbildung 2.21 ist ein Beispiel für das Vorgehen bei Aktivierung des Features gezeigt. Deaktiviert wird das Feature, wenn sich wieder eine Verbesserung/Änderung des besten Fitness-Werts zeigt.

¹³D.h., viele Rekombinationsversuche werden benötigt, um ein neues Modell hervorzubringen siehe Abschnitt 2.4.4

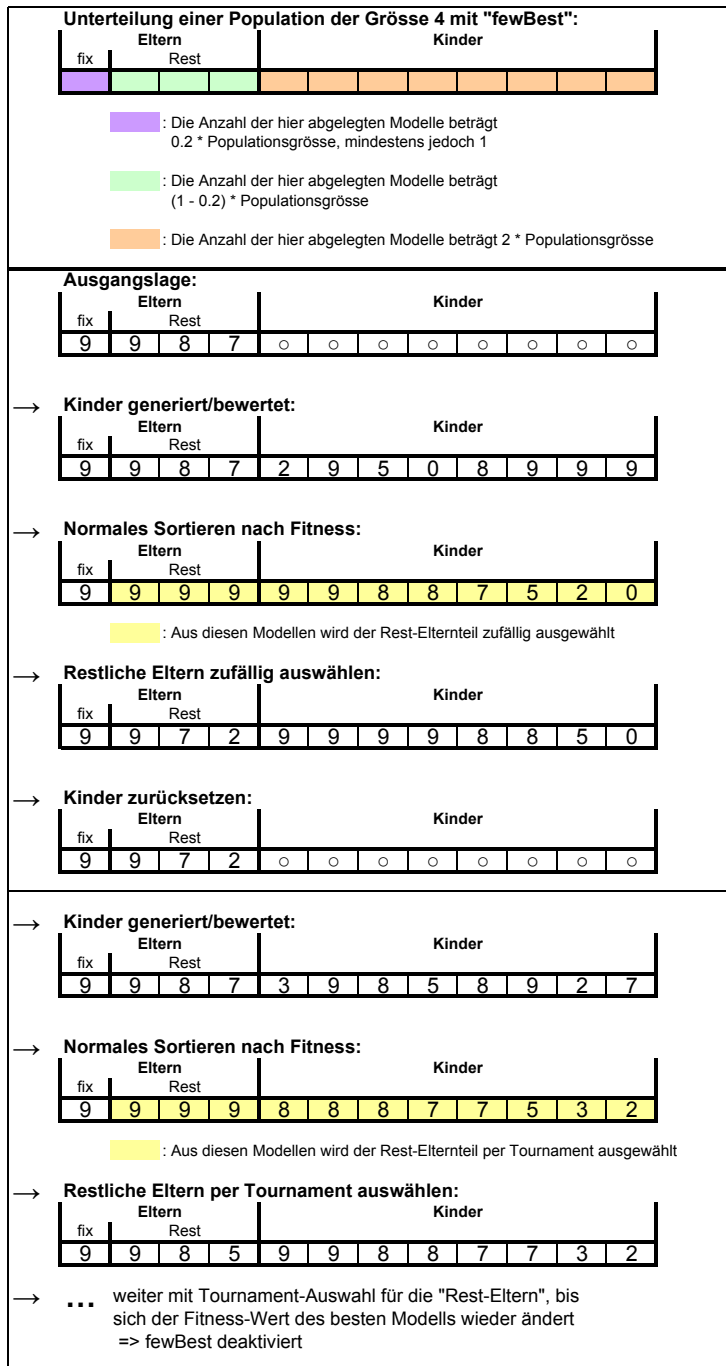


Abbildung 2.21: Vorgehen bei der "Environmental Selection" mit "fewBest"

Kapitel 3

Ergebnisse

3.1 Experimente

Zu Beginn der Arbeit stand im Vordergrund, die Limitationen unseres Programmes zu erkennen und aufzuzeigen. Daher wurde zu Beginn versucht, einfache Muster nachzubilden um anschliessend zu erkennen, womit das Programm Mühe bekundet und womit es problemlos fertig wird.

Während der Simulation diverser Muster kam die Theorie auf, dass man eventuell mit mehr Genen aufwändigere Muster erreichen könnte, da dann viel mehr Interaktionsmöglichkeiten existieren. Auf diese Theorie wird in Abschnitt 3.3.2 näher eingegangen.

Die Ergebnisse der Tests können wie folgt gelesen werden:

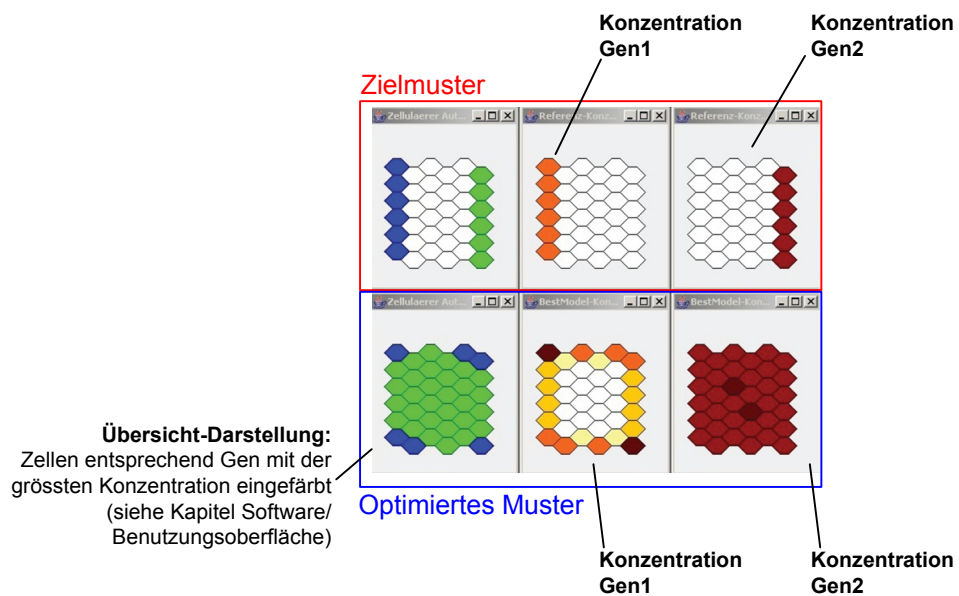


Abbildung 3.1: Vertikal-achsensymmetrisches Muster, gerade Spaltenzahl

3.1.1 Symmetrien

Während der Evaluation des Optimierungsalgorithmus haben wir festgestellt, dass für Zielmuster mit gerader Spaltenzahl, bzw. mit einer geraden Zahl für die Breite, stets punktsymmetrische optimierte Muster gefunden werden. Hingegen werden für Zielmuster mit ungerader Spaltenanzahl immer vertikal-achsensymmetrische¹ Muster generiert. Daher wurde systematisch untersucht, welchen Einfluss die Symmetrieform der gegebenen Zellstruktur auf das Resultat der Optimierung hat.

Es wurden nie optimierte Muster gefunden, welche horizontal-achsensymmetrisch² oder gar unsymmetrisch waren. Eine Erklärung für erstere Beobachtung könnte sein, dass die verwendete Zellstruktur an und für sich nicht diese Art von Symmetrie aufweist³. Folglich können sich die Gene auch nicht gleichermassen nach oben und unten ausbreiten. Für die Tatsache, dass unsymmetrische Muster anscheinend unmöglich erreicht werden können, kann argumentiert werden, dass für keinen der verwendeten Modell-Parameter eine Richtungsabhängigkeit besteht und sich so alle Gene gleichermassen - sprich: symmetrisch - in alle Richtungen ausbreiten.

Zu den verschiedenen Symmetrie-Arten wurden Tests mit gerader und ungerader Spaltenzahl durchgeführt, deren Resultate nachfolgend aufgeführt sind. Sämtliche Versuche wurden mit folgenden Einstellungen durchgeführt:

- Breite x Höhe = $\begin{cases} 6 \times 6, & \text{für gerade Spaltenzahl} \\ 5 \times 5, & \text{für ungerade Spaltenzahl} \end{cases}$
- Anzahl Gene = 2
- Anzahl Optimierungsiterationen = 200
- Populationsgrösse = 50
- Anzahl Simulationszyklen = maximal 5000
- Normale Randbedingung
- Ohne obere Konzentrationsgrenze

vertikal-achsensymmetrisch

¹Von links nach rechts gespiegelt.

²Von oben nach unten gespiegelt.

³Unten "überlappen" die geraden Spalten, oben jedoch nicht.

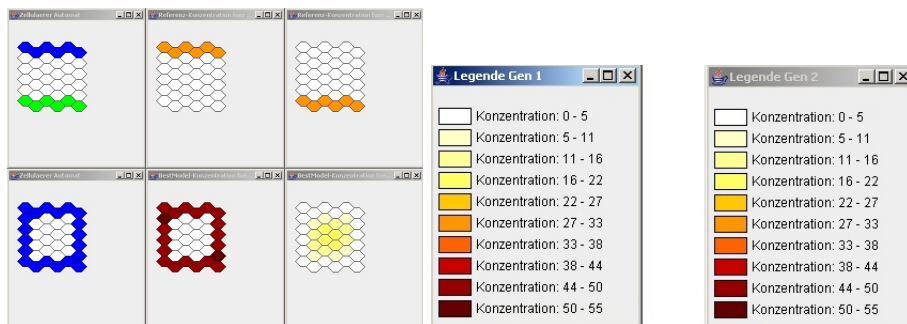


Abbildung 3.2: Vertikal-achsensymmetrisches Muster, gerade Spaltenzahl

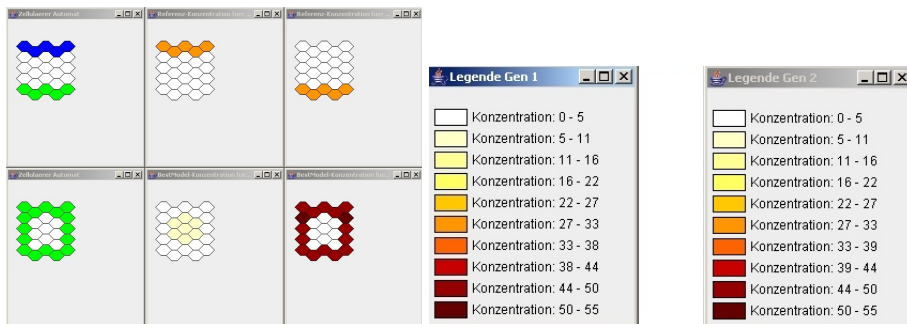


Abbildung 3.3: Vertikal-achsensymmetrisches Muster, ungerade Spaltenzahl

Wie erwartet steigt in Abbildung 3.4 die Fitness für eine ungerade Spaltenzahl schneller an und kommt auf einen höheren Wert, da für diesen Fall die Zellstruktur vertikal-achsensymmetrisch ist. Es ist anzunehmen, dass die gegebene Zellstruktur einen entscheidenden Einfluss darauf hat, ob ein punktsymmetrisches oder achsensymmetrisches Muster leichter generiert werden kann. Dies zeigt sich auch im Verlauf der Fitness.

Es ist erkennbar, dass für eine ungerade Spaltenzahl ein vertikal-achsensymmetrisches und für gerade Spaltenzahl ein punktsymmetrisches Muster gefunden wird.

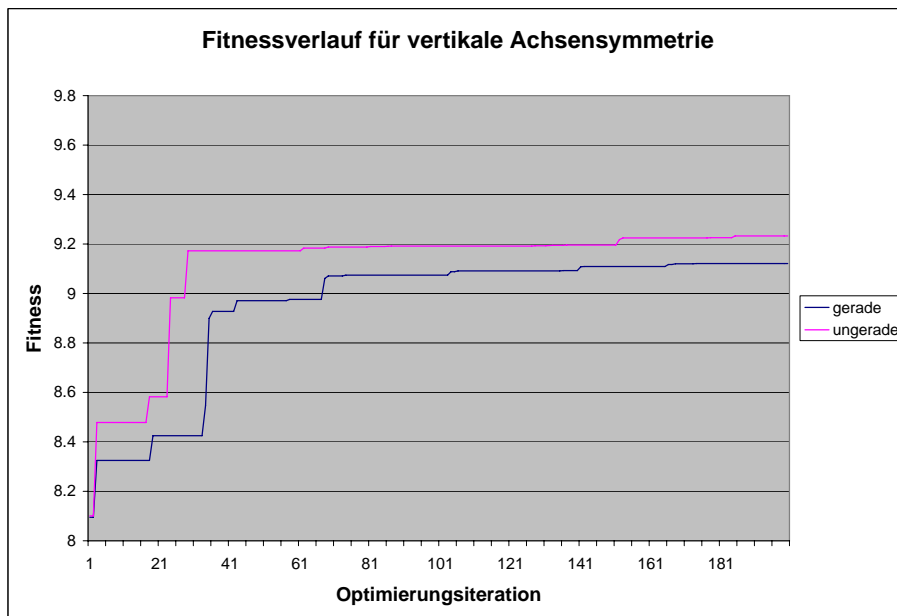


Abbildung 3.4: Fitnessverläufe für ein vertikal-achsensymmetrisches Muster

horizontal-achsensymmetrisch

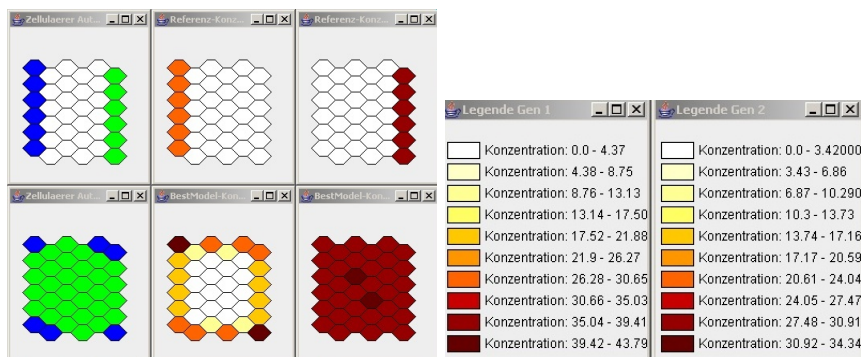


Abbildung 3.5: Horizontal-achsensymmetrisches Muster, gerade Spaltenzahl

Die Resultate in Abbildung 3.6 und 3.5 bestätigen, dass es nicht möglich ist, horizontal-achsensymmetrische Muster zu erzeugen, welche nicht auch gleichzeitig vertikal-achsensymmetrisch sind. Folglich werden derartige Muster durch ähnliche punktsymmetrische bzw. vertikalachsensymmetrische Muster angenähert.

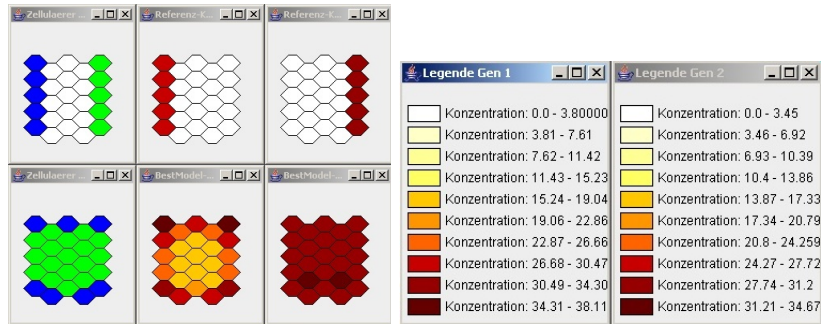


Abbildung 3.6: Horizontal-achsensymmetrisches Muster, ungerade Spaltenzahl

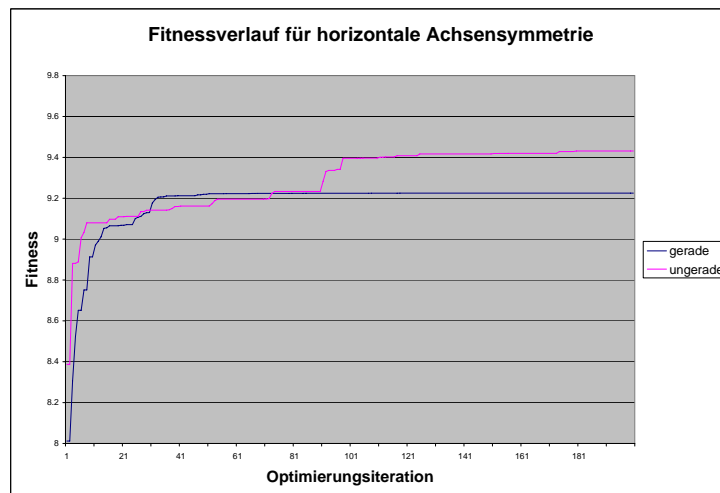


Abbildung 3.7: Fitnessverläufe für ein horizontal-achsensymmetrisches Muster

Punktsymmetrie

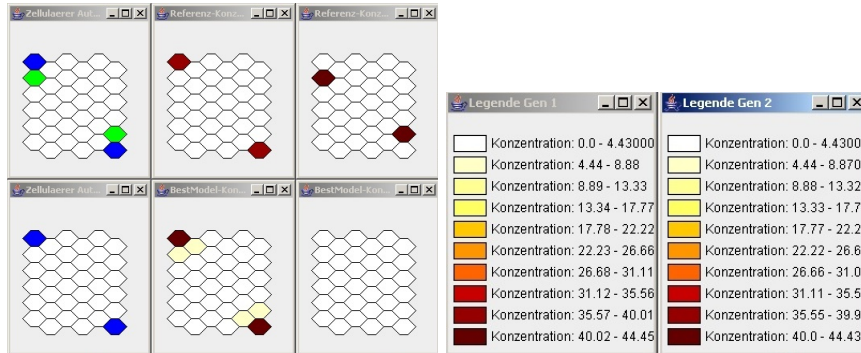


Abbildung 3.8: Punktsymmetrisches Muster, gerade Spaltenzahl

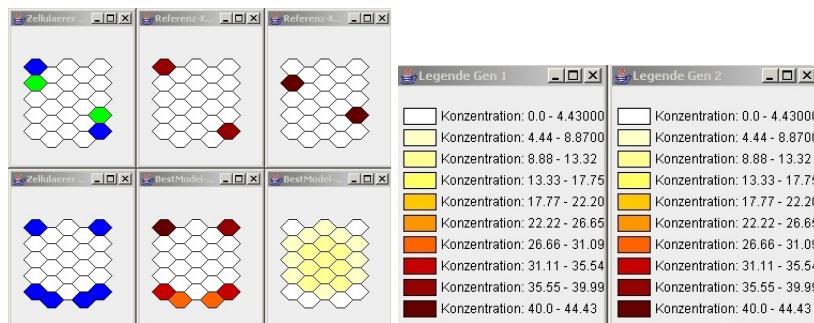


Abbildung 3.9: Punktsymmetrisches Muster, ungerade Spaltenzahl

In Abbildung 3.10 ist der Unterschied im Fitness-Verlauf zwischen gerader und ungerader Spaltenzahl am deutlichsten erkennbar, was daher rührt, dass es (für ungerade Spaltenzahl) schlecht möglich ist, punktsymmetrische Muster durch vertikalachsensymmetrische anzunähern. Spiegelt man z.B. das Feld ganz links oben von links nach rechts, ergibt das beim gegebenen Muster bereits einen Fehler. Daher resultiert jedes Feld mit gut angenäherter Genkonzentration gleich in einem Fehler an der vertikal gespiegelten Stelle.

Fazit

Nach obigen Erkenntnissen ist es empfehlenswert, für punktsymmetrische Zielmuster eine gerade Anzahl Spalten und für vertikalachsensymmetrische Zielmuster eine ungerade Spaltenzahl zu verwenden. Es ist zu erwarten, dass dadurch bessere Fitness-Werte resultieren.

3.1.2 Skalierbarkeit

Eine wichtige Frage im Bezug auf das Programm ist die Skalierbarkeit. Zentral dabei ist der Punkt, ob das Programm auch in der Lage wäre, Modelle zu finden, von denen man weiss, dass sie theoretisch erreichbar sind, auch wenn die

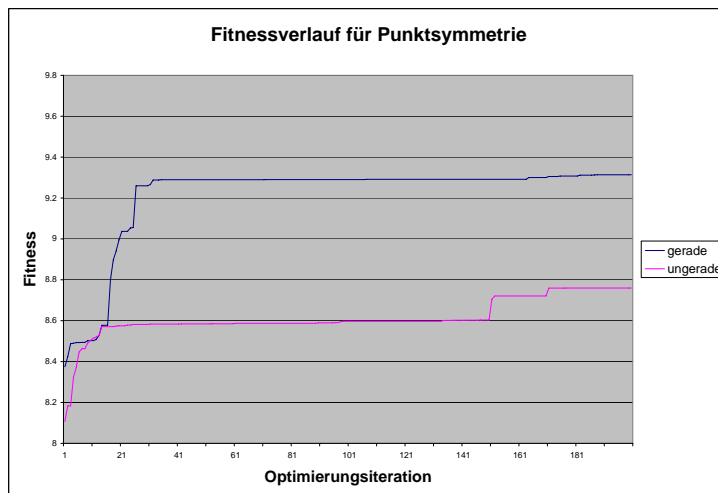


Abbildung 3.10: Fitnessverläufe für ein punktsymmetrisches Muster

Zellstruktur sehr gross gewählt wird und / oder mit einer grossen Anzahl Gene gerechnet wird. Die Komplexität des Optimierungsproblems berechnet sich nach Formel 3.1. Mit Komplexität ist hier die Anzahl variierbarer Parameter gemeint, n bezieht sich auf die Anzahl Gene.

$$\mathcal{O}(n) = n^2 + 2n \quad (3.1)$$

Es wurden vier verschiedene Szenarien durchprobiert. Im ersten Szenario wurde mit zwei Genen und einer Zellstruktur von 8x8 Zellen gerechnet. In einem weiteren Szenario wurde die Zellstruktur auf 25x25 Zellen vergrössert, die Anzahl Gene gleich gelassen. Beim dritten Szenario wurde die Anzahl Gene auf sechs erhöht, die Zellstruktur wurde auf 3x3 Zellen eingestellt. Das letzte Szenario bestand aus sechs Genen mit einer Zellstruktur von 20x20 Zellen. Diese vier Experimente findet man in dieser Reihenfolge in den folgenden Unterabschnitten.

Wenig Gene, kleine Zellstruktur

Zu Beginn wurden einfache Modelle aufgestellt und deren Muster mittels der CA-Simulation bestimmt. Für Modelle mit zwei bis drei Genen und einer Höhe und Breite der Zellstruktur kleiner als zehn bekundet das Programm relativ wenig Mühe, ein Modell zu finden, das das Muster sehr genau annähern kann. In der Regel werden solche Genmodelle innerhalb weniger Minuten gefunden. In Tabelle 2.4 ist ein einfaches Referenzmodell mit zwei Genen beschrieben. Es handelt sich um das Modell, das man im Programm unter Modellauswahl als *simples Modell* wählen kann.

Rechnet man mit normaler Randbedingung, dann ist das Referenzmuster in 1047 CA-Iterationen stabil. Bei dem Versuch, dieses Muster durch das Programm anzunähern, bekommt man schon nach ca. 200 EA-Iterationen ein Muster das sehr nahe (mittlerer Fehler ≤ 1) am Referenzmuster liegt. Interessant dabei ist, dass dieses durch das Programm gefundene Muster viel schneller stabil

wird als das Referenzmuster, es braucht also weniger CA-Iterationen. Das optimierte Muster braucht dafür nur 157 CA-Iterationsschritte. In Tabelle 3.1 ist das beste durch das Programm gefundene Modell abgebildet. Wie man aus der Abbildung 3.11 sieht, wird das Referenzmuster nahezu perfekt nachgebildet. Der maximale Fehler, der in einer Zelle vorkommt, beträgt 2.7, der durchschnittliche Fehler 0.62.

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion
A	40.9	9.44	6.56	0 10.7
B	75.7	9.09	3.51	-3.67 0

Tabelle 3.1: Optimiertes Modell mit normaler Randbedingung

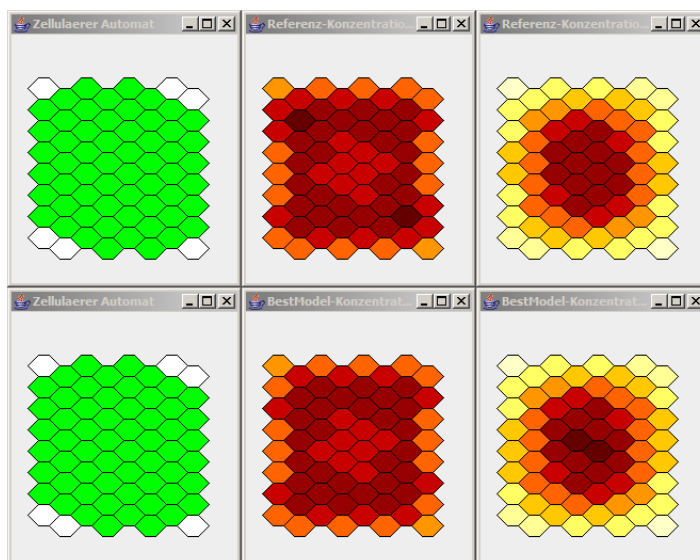


Abbildung 3.11: Oben: Referenzmuster mit Konzentration Gen 1, Gen 2. Unten: Optimiertes Muster mit Konzentration Gen 1 Gen 2

Für den Fall mit spezieller Randbedingung hat das Programm wesentlich mehr Mühe. Die Simulation des Referenzmodells (Tabelle 2.4) braucht in diesem Fall 2524 CA-Iterationsschritte, das sind mehr als doppelt so viele wie unter normaler Randbedingung. Eine Erklärung hierfür kann sein, dass es durch das unterschiedliche Verhalten an den Rändern länger dauert, bis sich ein Gleichgewicht eingestellt hat. Des weiteren bekundet das Programm mehr Mühe, ein passendes Modell zu finden. Mit spezieller Randbedingung muss man viel länger optimieren lassen, bis ein äquivalent gutes Modell gefunden wird. Dieses Phänomen konnte bis jetzt noch nicht erklärt werden. Eigentlich sollte die Art der Randbedingung keinen Einfluss auf die Geschwindigkeit der Optimierung haben. Diese Aussage über die Optimierungsdauer ist auch insofern fragwürdig, da es manchmal auch Optimierungsläufe gibt, in denen das Muster sehr schnell gefunden wird. Jedoch kommen diese wesentlich seltener vor als bei dem Modell mit normaler Randbedingung. Die Dauer der Optimierung ist aber v.a. auch

vom Zufall abhängig, da viele Werte zufällig gesetzt werden. Es kann bei einfachen Modellen teilweise unter 200 EA-Iterationen brauchen, bis eine nahezu perfekte Lösung gefunden wird, manchmal aber auch über 1000. Eine genauere Untersuchung über dieses Problem findet man im Abschnitt 3.1.3.

Für das Referenzmuster mit spezieller Randbedingung brauchte es c.a. 1000 EA-Iterationen, bis ein vernünftiges (mittlerer Fehler ≤ 1) Modell gefunden wurde. Auch hier ist das durch den Optimierer gefundene Modell schneller stabil, als das eigentliche Referenzmodell (nach 649 CA-Iterationen). Der maximale Zellfehler⁴ gegenüber dem Referenzmuster beträgt hier 1.46, der mittlere Fehler 0.69. Das optimierte Modell findet man in Tabelle 3.2. Auch hier wurde das Zielmuster fast perfekt erreicht, wie aus der Abbildung 3.12 ersichtlich ist.

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion
A	38.4	9.56	1.7	0 10.8
B	56.7	5.94	2.92	-5.2 0

Tabelle 3.2: Optimiertes Modell mit spezieller Randbedingung

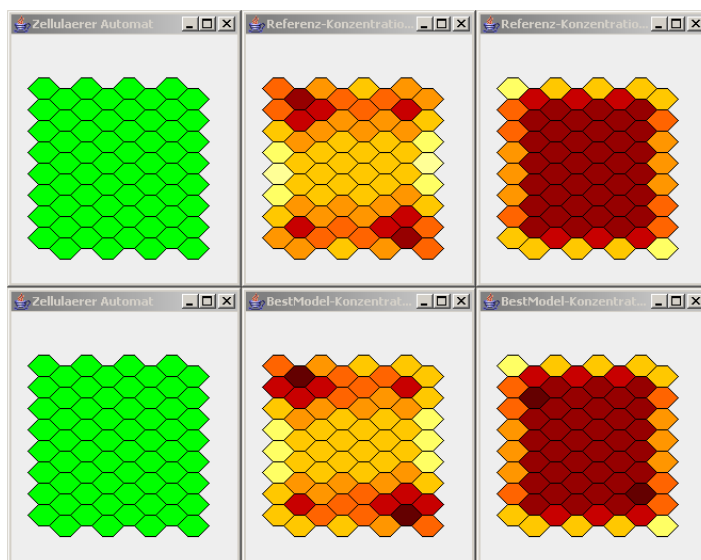


Abbildung 3.12: Oben: Referenzmuster mit Konzentration Gen 1, Gen 2. Unten: Optimiertes Muster mit Konzentration Gen 1 Gen 2

Wenig Gene, grosse Zellstruktur

In diesem Unterabschnitt wird weiterhin das gleiche Modell verwendet wie bisher (Tabelle 2.4). Die Zellstruktur wurde auf 25×25 Zellen vergrössert. Das Optimieren dauert hier wesentlich länger als bei Modellen mit kleinen Dimensionen, jedoch können die Muster immer noch genau (mittlerer Fehler ≤ 1) angenähert

⁴Damit ist die grösste vorkommende Differenz über alle Zellen zwischen dem Referenzmuster und dem optimierten Muster gemeint

werden. Die Skalierbarkeit bezüglich der Zellgrösse erscheint gut, die Optimierungsdauer steigt proportional zur Zelldimension. Die Anzahl der Optimierungsschritte nimmt dabei zu, obwohl sich die Parameterzahl nicht vergrössert hat. Das liegt daran dass die Modelle genauer⁵ sein müssen um den gewünschten Fehler zu erreichen.

Viele Gene, kleine Zellstruktur

Um diese Konstellation mit vielen Genen zu testen, wurde ein "Modelfinder"-Programm⁶ geschrieben, das Genmodelle mit sechs Genen und einer Zellgrösse von 8x8 generiert und anschliessend simuliert. Nur die Modelle, die stabile Muster hervorbringen wurden abgespeichert. Aufgefallen ist dabei, dass nie ein Modell gefunden wurde, bei dem die Genkonzentration nach der Simulation für alle Gene grösser Null war. Meistens hatten drei Gene eine Konzentration, die sich wesentlich von Null unterschied, die restlichen waren praktisch gleich Null.

Um dem entgegenzuwirken wurde ein neues Genmodell definiert, wiederum ein sehr einfaches, bei dem keines der Gene eine Konzentration von überall Null hat. Es ist aus Tabelle 3.3 und 3.4 zu entnehmen. Nun bekundet das Programm Mühe, überhaupt ein stabiles Modell zu finden. Das könnte daran liegen, dass es für dieses vermeintlich einfache Muster viel weniger Modelle gibt, die ähnliche Muster erzeugen können als für das durchschnittlich gefundene Muster, das die Modelle des "Modelfinder"-Programms erzeugen.

Um Zeit zu sparen, wurde die Zellgrösse auf 3x3 Zellen verkleinert. Wiederum zeigte sich das selbe Bild. Es wurden zu Beginn der Optimierung nur instabile Modelle gefunden. Wählt man hingegen ein solches automatisch generiertes Modell als Referenzmodell, dann wird sehr schnell ein stabiles optimiertes Modell wiedergefunden. Um dem Problem der instabilen Muster Herr zu werden, wurde ein neues Feature im Programm eingebaut. Die Startpopulation kann nun solange generiert werden, bis mindestens ein stabiles Modell darin vorkommt, des weiteren wurde der *Fitness-Straffaktor*, den instabile Modelle bei der Fitnessbewertung bekommen, verschärft. Nach diesen Veränderungen war das Programm auch in der Lage, vermeintlich einfache 6 Gen Modelle, wie das aus den Tabellen 3.3 und 3.4, zu optimieren.

Gen	Diffusion	Eigenprod.	Eigenred.
A	5	1	1.5
B	5	1	2
C	5	1	2
D	5	1	1.5
E	5	1	2
F	5	1	1.5

Tabelle 3.3: Referenzmodell 6 Gene Teil 1

⁵Damit ist gemeint, dass identische Modelle bei kleiner Zellstruktur schon gute Muster liefern, bei grosser Zellstruktur hingegen nicht.

⁶Es handelt sich hier um ein reines Hilfsprogramm

Gen	Interaktionsmatrix					
A	0	1	-1	1	1	-1
B	1	0	1	-1	-1	-1
C	-1	-1	0	-1	1	1
D	-1	-1	1	0	1	1
E	-1	1	1	1	0	-1
F	1	1	-1	-1	-1	0

Tabelle 3.4: Referenzmodell 6 Gene Teil 2

Viele Gene, grosse Zellstruktur

Solche Modelle sind sehr zeitintensiv auf einem normalen Desktop Rechner. Das Programm rechnete mehrere Tage, bis befriedigende (mittlerer Fehler ≤ 2) Genmodelle gefunden wurden. Die hohe Parameterkomplexität zusammen mit der grossen Zellstruktur benötigt einen grossen Rechenaufwand. Durch das quadratische Ansteigen des Rechenaufwandes ist eine Optimierungsdauer von mehreren Tagen nicht weiter verwunderlich.

3.1.3 Optimierungsverlauf

Die Optimierungsdauer für einfache Modelle, bis gute Muster gefunden wurden, war sehr unterschiedlich. Teilweise wurden gute Modelle in wenigen Iterationsschritten gefunden. Dann kam es aber auch vor, dass nach über 2000 Schritten noch kein vernünftiges Modell generiert wurde. Um das statistisch genauer zu betrachten, wurde ein 3 Gen Modell mit quadratischen Zelldimensionen von 2x2 bis 10x10 durchprobiert. Jede Dimension wurde zwanzig mal optimiert, das macht bei 9 verschiedenen Dimensionen insgesamt 180 Runs. Die Populationsgrösse lag bei 20 Modellen. Das Abbruchkriterium für die Optimierung war ein durchschnittlicher Fehler von kleiner als 1.5. Falls das nicht erreicht wurde, brach der Optimierer nach 1000 Iterationen ab.

Sehr auffallend sind die teilweise extrem grossen Differenzen innerhalb einer Zelldimension. Bei der 3x3-Zellstruktur wurde im schnellsten Fall ein gutes Modell nach 7 Optimierungsiterationen gefunden, zweimal hingegen wurde auch nach über 1000 EA-Iterationen kein genügend gutes Modell gefunden. Die 10x10-Zellstruktur konnte im besten Fall bereits nach 33 EA-Iterationen reproduziert werden. In der Tabelle 3.5 ist eine Auflistung zu finden mit der Minimalen (*min Iter*), der Maximalen (*max Iter*) und der durchschnittlichen (E_{normal}) Optimierungsdauer. Die durchschnittliche Optimierungsdauer wurde nur über die tatsächliche Anzahl beendeter Runs (*#beendete Runs*) berechnet. Da man bei den nicht beendeten Runs nicht weiss, wie lange es noch gegangen wäre, bis die gewünschte Toleranz erreicht worden wäre, wurden sie nicht in die Berechnung miteinbezogen. Um diesem Umstand der unbeendeten Runs trotzdem Rechnung zu tragen, wurde eine weitere Kennzahl eingeführt: Die bereinigte, durchschnittliche Optimierungsdauer ($E_{bereinigt}$). Die Berechnung dieses Parameters erfolgte nach folgender Gleichung.

$$E_{bereinigt} = \frac{1}{\frac{\#beendeteRuns}{\#Runs}} \cdot E_{normal} \quad (3.2)$$

Dimension	min Iter	max Iter	E_{normal}	#beendete Runs	$E_{bereinigt}$
2x2	5	119	24.75	20	24.75
3x3	7	867	116.73	18	129.7
4x4	12	676	151.85	13	233.6
5x5	17	979	199.63	11	363
6x6	16	957	322.75	8	806.9
7x7	43	189	151.71	7	433.5
8x8	44	819	218.8	9	486.2
9x9	34	855	230.64	11	419.34
10x10	33	811	203	12	338.2

Tabelle 3.5: Optimierungsergebnisse

Die Ergebnisse lassen Folgendes erkennen: Im Durchschnitt dauert eine Optimierung mit grösseren Zelldimensionen länger als eine mit kleinen Dimensionen. Aber da vieles zufällig gebildet wird, wie z.B. die Startpopulation, kann es durchaus vorkommen, dass das Programm für ein 10x10-Muster schneller eine optimale Lösung findet als für ein 5x5-Muster. Die Wahl von 20 Runs pro Dimension erwies sich im Nachhinein als zu wenig, um verlässliche Aussagen zu machen.

3.1.4 Einfluss der verschiedenen Randbedingungs-Formen

Der Einfluss der verschiedenen Arten von Randbedingungen auf den Optimierungserfolg wurde im folgenden Abschnitt untersucht. Erste Ergebnisse haben gezeigt, dass je nach Zielmuster die ein oder andere Art von Randbedingung von Vorteil sein kann.

Um den Einfluss der Randbedingungen zu prüfen, wurde in einem ersten Schritt versucht, ein Muster mit zwei Genen, die über alle Zellen je einen konstanten Konzentrationswert besitzen, einmal mit spezieller und einmal mit normaler Randbedingung anzunähern. Es wurde eine Zelldimension von 4x4 Zellen gewählt. Das Referenzmuster besitzt eine Konzentration von 55 für das erste Gen und 60 für das zweite. Es wurden je 30 Runs mit spezieller Randbedingung und je 30 mit normaler Randbedingung durchgeführt. Als Abbruchbedingung bei der Optimierung galt ein mittlerer Fehler pro Zelle von kleiner eins oder falls nach 10000 EA-Iterationen kein Modell gefunden wurde. Durch die relativ kleine Dimension haben alle Optimierungsläufe innerhalb von 10000 Schritten ein Modell gefunden, dessen mittlerer Fehler kleiner eins ist. Die Resultate sind eindeutig ausgefallen. Sobald mit spezieller Randbedingung gerechnet wurde, hat das Programm deutlich schneller eine Lösung gefunden. Eine Übersicht über die wichtigsten Resultate findet man in Tabelle 3.6, die mittlere Iterationsdauer ist in der Spalte E angegeben. Die spezielle Randbedingung ist für konstante

Randbedingung	min Iter	max Iter	E
normal	44	8191	2060.9
speziell	23	242	80.2

Tabelle 3.6: Optimierungsergebnisse für konstantes Muster

Muster klar besser, das Programm findet 25.7 mal schneller passende Modelle.

Fazit

Die Frage, ob mit spezieller oder normaler Randbedingung gerechnet werden soll, kann nicht alleine aufgrund des Optimierungserfolges beantwortet werden. Eine wichtige Rolle spielen auch die in der Wirklichkeit gegebenen Bedingungen. Die Wirklichkeit, die man abzubilden versucht, hat vielleicht keine spezielle Randbedingung (wenn man z.B. nur einen Ausschnitt aus dem Zellverband betrachtet) oder sie sind unbedingt notwendig (z.B. bei Randzellen).

3.1.5 Einfluss der Konzentrationsgrenze

Zu Beginn der Arbeit wurde immer mit einer Konzentrationsgrenze gerechnet. Es wurde davon ausgegangen, dass die Genkonzentration in einer Zelle zwischen 0 und 100 liegen kann. Sobald sich ein Wert grösser 100 in einer Zelle gebildet hat, wurde er automatisch wieder auf 100 gesetzt. Da das Programm immer stabile Muster ausspuckte, kam der Verdacht auf, es läge eventuell nur an dieser Konzentrationsgrenze.

Um den Einfluss dieser Grenze zu testen, wurde ein einfaches Referenzmuster gewählt mit zwei Genen und einer quadratischen Dimension von 3x3 Zellen. Ein Gen wurde auf den konstanten Wert von 85 gesetzt, das andere auf 90. Nun wurde dieses Referenzmuster je 40 mal mit dieser Konzentrationsgrenze und 40 mal ohne diese Grenze durchprobiert. Die Abbruchbedingung für die Optimierung war ein durchschnittlicher Fehler pro Zelle von kleiner eins oder falls innerhalb von 10000 Optimierungsschritten kein genügend gutes Modell gefunden wurde.

Wenn man sich die Resultate anschaut, präsentiert sich ein anderes Bild als man zu Beginn annehmen würde. Das Programm tut sich schwerer mit der Modellfindung, wenn die Konzentrationsgrenze eingeschaltet ist. Es schafft es dann genau acht mal, ein genügend gutes Modell zu finden, das entspricht einer Erfolgsquote von lediglich 20%. Ohne Konzentrationsgrenze geschah dies immerhin 37 mal, was einer Erfolgsquote von 92.5% entspricht. Die Resultate finden sich in der Tabelle 3.7. Die Variable *min Iter* gibt die # Iterationen für den schnellsten Run an, die Variable *max Iter* die für den langsamsten, aber erfolgreichen. Unter dem Punkt E_{normal} stehen die Mittelwerte für die # Iterationsschritte der beendeten Optimierungsläufe. Unter $E_{bereinigt}$ wurde der grossen Anzahl nicht beendeter Optimierungsläufe Rechnung getragen, es wurde Formel 3.2 verwendet.

Konz. Grenze	min Iter	max Iter	E_{normal}	#beendete Runs	$E_{bereinigt}$
100	791	9261	3476.9	8	17384.4
Max.Double ⁷	93	8845	3255.5	37	3519.5

Tabelle 3.7: Optimierungsergebnisse

⁷Hier wird die obere Grenze durch den maximalen double Wert bestimmt. Bei 64 Bit liegt der bei 18446744073709551616 (= 2^{64})

Fazit

Die Beschränkung durch die Konzentrationsgrenze erweist sich als nicht notwendig, ja sogar hinderlich für den Optimierungserfolg.

3.2 Performance

3.2.1 Zeitmessung

Die Entwicklung der Optimierungsgeschwindigkeit (bzw. der pro Optimierungsiteration benötigten Zeit) wurde in Abhängigkeit der Variablen Feldgrösse, Anzahl Gene und Populationsgrösse aufgezeichnet. Die Resultate sind in den drei folgenden Unterkapiteln nachzulesen.

Feldgrösse

Die pro Optimierungsiteration benötigte Zeit ist gemäss einer Fittingkurve (in der Grafik rot gezeichnet) ungefähr quadratisch abhängig von der Feldgrösse. Das rührt daher, dass die Feldfläche bei einer Vergrösserung in die horizontale und vertikale Richtung quadratisch zunimmt.

Da der Berechnungsaufwand des Simulators (siehe Abschnitt 2.3) proportional zur Feldfläche zunimmt, ergibt sich der beschriebene Zusammenhang.

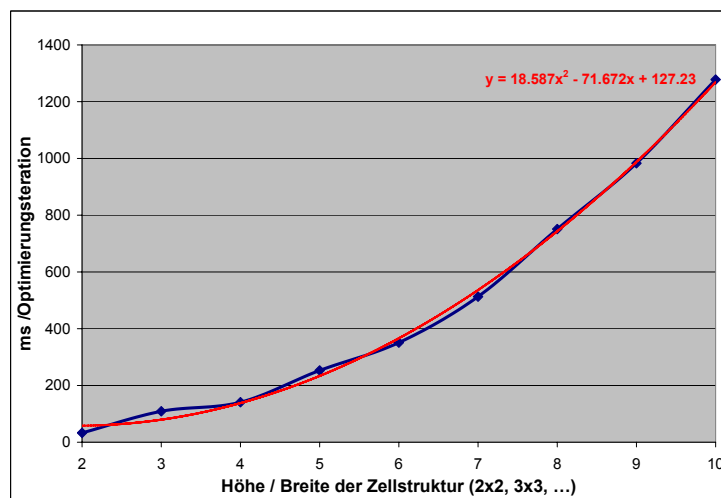


Abbildung 3.13: Abhängigkeit der Optimierungszeit von der Feldgrösse, blau: Messdaten, rot: parametrisierte Kurve mit Formel

Anzahl Gene

Die durchschnittliche Optimierungszeit scheint hier - gemäss Abbildung 3.14 - linear von der Anzahl Gene abhängig zu sein.

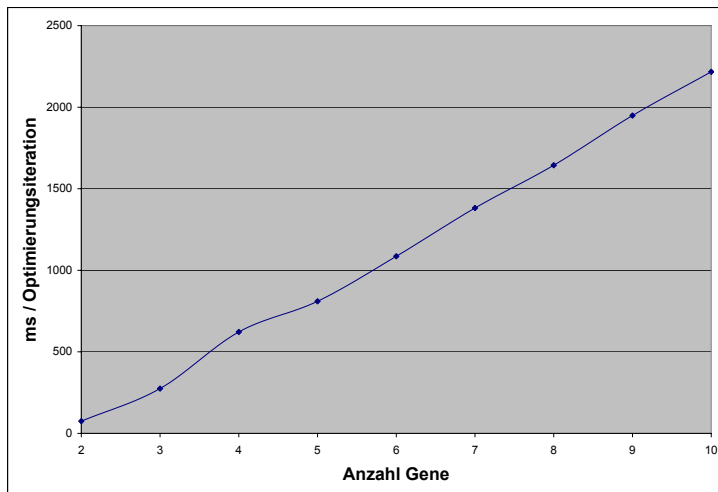


Abbildung 3.14: Abhängigkeit der Optimierungszeit von der Anzahl Gene

Betrachtet man dies von Seiten der Fitness-Bewertung (Simulation nicht eingeschlossen) ist das verständlich, da für jedes zusätzliche Gen nochmals der selbe Aufwand zur Berechnung der Differenz zum Zielmuster nötig ist.

Allerdings für Mutation und Rekombination kann sich ein grösserer Aufwand ergeben, da die Anzahl der mutier-/rekombinierbaren Modellparameter (schneller als) quadratisch mit der Anzahl der Gene wächst. Es muss dabei beachtet werden, dass der zeitliche Aufwand für die Mutation bzw. Rekombination eines Modells verglichen mit jenem für die Simulation und Bewertung eines Modells sehr klein ist. Deshalb zeigt sich dieser überproportionale Anstieg aufgrund Mutation/Rekombination nicht in der Grafik.

Es bleibt also noch ein Faktor zur Diskussion übrig: Die Simulationszeit. Zur Bestimmung der Abhängigkeit der Simulationszeit von der Anzahl Gene wurden für 2 bis 10 Gene je 1000 zufällig generierte Modelle simuliert und der dabei gefundene arithmetische Mittelwert für Simulationszeit und Anzahl Simulationszyklen in eine Tabelle (Abb. 3.16) eingetragen. Anhand der blossen Simulationszeit ohne Einbezug der Anzahl Simulationsiterationen wurde schliesslich die blaue Kurve in Abbildung 3.15 erzeugt. Da diese Kurve allerdings stark von der Anzahl Simulationszyklen abhängt, wurde die grüne Kurve generiert, indem die durchschnittliche Zeit pro Simulationsiteration⁸ für jede Genanzahl wie folgt berechnet wurde: $zeitProSimulationsiteration = \frac{zeitProSimulation}{anzahlSimulationsiterationen}$. Erst diese Normierung macht eine Interpretation der untersuchten Abhängigkeit möglich.

Für Diffusion, Selbstproduktion und Selbstzerfall würde man bei zusätzlichen Genen lineare Zunahme der Simulationszeit erwarten, was Abbildung 3.15 bestätigt. Bei genauerem Hinsehen lässt sich jedoch eine leichte Beugung der Gerade

⁸In Abbildung 3.15 in 10 μ s-Schritten angegeben.

erkennen. Dies rührt wahrscheinlich daher, dass die Anzahl Elemente der Interaktionsmatrix - und damit der Simulationsaufwand für die Interaktion der Gene - nahezu quadratisch⁹ ansteigt. Folglich ergibt sich die schwache quadratische Komponente der Fitting-Kurve in der Grafik.

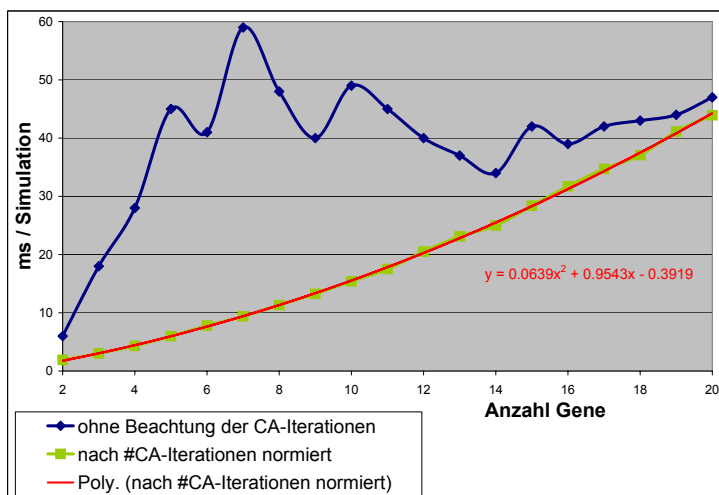


Abbildung 3.15: Abhängigkeit der Simulationszeit von der Anzahl Gene

Anzahl Gene	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Breite des Feldes	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Höhe des Feldes	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Simulationssiterationen (maximal)	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000	5000
Ø Zeit pro Simulation [ms]	6	18	28	45	41	59	48	40	49	45	40	37	34	42	39	42	43	44	47
Ø #CA-Iterationen:	313	598	646	752	526	628	424	302	318	257	195	160	136	148	123	121	116	107	107
Ø Zeit pro Simulationssiteration [*10µs]	2	3	4	6	8	9	11	13	15	18	21	23	25	28	32	35	37	41	44

Abbildung 3.16: Tabelle zu Diagramm aus Abbildung 3.15

⁹Die Anzahl Elemente ungleich Null der Interaktionsmatrix entspricht $nGenes^2 - nGenes$.

Populationsgrösse

Wie erwartet steigt die Optimierungszeit linear zu Populationsgrösse an, da im Schnitt für jedes Element der Population - also für jedes Modell - der selbe Aufwand nötig ist.

Es können kleine Abweichungen aufgrund unterschiedlicher Anzahl benötigter Simulationsiterationen auftreten. Über eine grössere Population hinweg wird dieser Effekt jedoch vernachlässigbar.

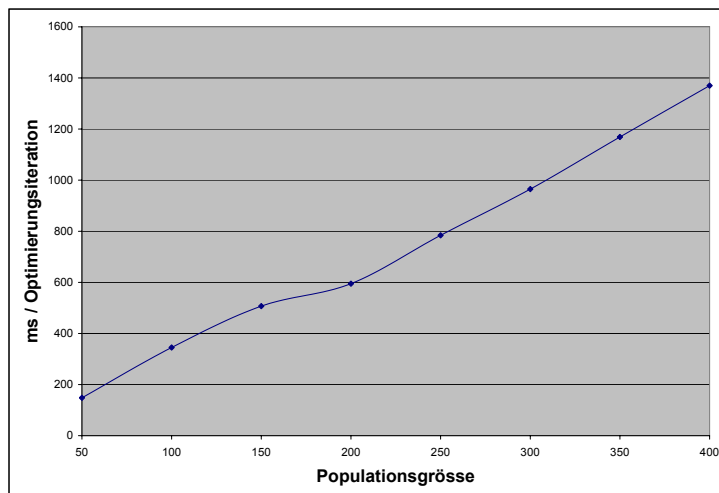


Abbildung 3.17: Abhängigkeit der Optimierungszeit von der Populationsgrösse

3.2.2 Optimale Populationsgrösse

Um eine möglichst schnelle Optimierung zu gewährleisten, wurden Tests durchgeführt um herauszufinden, welches die optimale Populationsgrösse sein könnte. Dazu wurden zwei Methoden angewandt, welche nachfolgend beschrieben sind.

Methode 1

In einem ersten Versuch wurden dafür nacheinander mehrere Optimierungen mit derselben Anzahl Iterationen und verschiedenen Populationsgrössen gestartet. Jede dieser Einstellungen wurde fünf mal durchgespielt, um so einen Mittelwert bilden zu können, welcher das Problem einer allfälligen schlechten Startpopulation¹⁰ etwas entschärfen soll. Nach jeder Optimierung wurde berechnet, wie gross der durchschnittliche Fitness-Verbesserungsfaktor von der besten Fitness einer Iteration auf jene der nächsten Iteration ist¹¹.

Dieses Vorgehen wurde für zwei verschiedene Zielmuster angewandt; die Resultate sind in den Abbildungen 3.18 und 3.19 dargestellt.

¹⁰Siehe Abschnitt 4.2

¹¹ $\text{verbesserungsfaktor} = \text{currentBestFitness} / \text{previousBestFitness} \rightarrow$ durchschnittlicher Verbesserungsfaktor entspricht dem arithmetischen Mittel dieser Faktoren

Populationsgrösse	Ø Fitness-verbesserungsfaktor:										Durchschnittswerte über alle Runs		
	Run 1	Zeit pro EA-Iteration	Run 2	Zeit pro EA-Iteration	Run 3	Zeit pro EA-Iteration	Run 4	Zeit pro EA-Iteration	Run 5	Zeit pro EA-Iteration	Verbesserungsfaktor	zeit	Ø Verbesserungsfaktor pro Sekunde
10	1.004744	25	1.00325	28	1.003642	24	1.003874	23	1.004263	22	1.003954654	24.4	0.44945
20	1.004343	52.406	1.001817	47.281	1.004896	48.563	1.003411	53.562	1.005213	51.484	1.003935977	50.659	0.21568
50	1.004854	131	1.003579	137.125	1.004272	122.718	1.004791	143.391	1.003208	136.031	1.004140889	134.05	0.08488
100	1.003448	259.703	1.003601	279.64	1.003838	284.25	1.003087	257.61	1.004446	272.609	1.003683829	270.76	0.03839

Zeitwerte: Ø Zeit pro Optimierungsiteration in [ms]

Als Zielmuster wurde das Endmuster des folgenden einfachen Modells mit 2 Genen verwendet:				# EA-Iterationen: 1000										
Interaktion:	Diffusion:	Eigenproduktion:	Eigenreduktion:	Breite x Höhe = 3 x 3										
<table border="1"> <tr><td>0</td><td>2</td></tr> <tr><td>-1</td><td>0</td></tr> </table>	0	2	-1	0	<table border="1"> <tr><td>10</td></tr> <tr><td>10</td></tr> </table>	10	10	<table border="1"> <tr><td>20</td></tr> <tr><td>10</td></tr> </table>	20	10	<table border="1"> <tr><td>5</td></tr> <tr><td>5</td></tr> </table>	5	5	# Gene: 2
0	2													
-1	0													
10														
10														
20														
10														
5														
5														

Abbildung 3.18: Gegenüberstellung von Fitness-Verbesserungsfaktoren, Muster: Einfaches Modell, 1000 EA-Iterationen

Populationsgrösse	Ø Verbesserungsfaktor	Endfitness	Ø Zeit / EA-Iteration	Ø Verbesserungsfaktor pro Minute
10	1.002683764	9.636412164	1572.03	0.49765
20	1.002257696	8.83423406	4364.84	0.17185
50	1.001652536	10.20607899	7566.57	0.09338
100	1.003624243	10.42924679	7712.19	0.11131
200	1.002884375	10.41311365	45251.56	0.01763
300	1.002787721	10.55546465	38392.5	0.02059
400	1.002680592	10.45948927	124436.9	0.00628

Zeitwerte: Ø Zeit pro Optimierungsiteration in [ms]


<p>Zielmuster mit 3 Genen:</p> 	<p># EA-Iterationen: 100</p> <p>Breite x Höhe = 9 x 7</p> <p># Gene: 3</p> <p># Runs: 1</p>
---	---

Abbildung 3.19: Gegenüberstellung von Fitness-Verbesserungsfaktoren, Muster: Carles, 100 EA-Iterationen

Die Resultate der Abbildungen 3.19 und 3.18 in der Spalte "ØVerbesserungsfaktor pro Zeit"¹² deuten darauf hin, dass eine Population von 10 optimal ist. D.h., dass mit einer Population von bloss 10 in der selben Zeit grössere Verbesserungen der Fitnesswerte zu erwarten sind als mit einer grösseren Population.

Method 2

Das Resultat von Methode 1 ist jedoch mit Vorsicht zu geniessen, da es bei einer kleinen Population wahrscheinlich eher vorkommt, dass der Algorithmus in einem lokalen Optimum stecken bleibt. Denn dann haben neue Modelle¹³ kleine Chancen, in die nächste Generation der Optimierungspopulation zu gelangen, weil sie in der ersten Entwicklungsstufe¹⁴ oft schlechter bewertet werden als jene, die unter den bestbewerteten sind. Ausserdem wäre es möglich, dass die Performance-Unterschiede in den Resultaten von Methode 1 daher kommen, dass bei grossen Populationen schon zu Beginn gute Fitness-Werte vorhanden sind und so nicht mehr so starke Anstiege zu beobachten sind wie bei kleinen Populationen.

Aus diesen Gründen wurde ein weiteres Vorgehen zum Leistungsvergleich verschiedener Populationsgrössen entwickelt: Die Optimierung wird unabhängig von der Populationsgrösse erst dann abgebrochen¹⁵, wenn eine bestimmte Schranke für die Fitness überschritten wurde. So lässt sich erkennen, mit welcher Populationsgrösse wie lange optimiert werden muss, um zu einem gewissen Resultat zu kommen. Es ist denkbar, dass kleine Populationen bei diesen Tests schlechter abschneiden, da die Optimierung dann evtl. nur langsam aus lokalen Optima hinausfindet.

Abbildung 3.20 zeigt die Auswertung der Testreihe nach dem zweiten Verfahren. Es wurden fünf bis zehn¹⁶ Runs pro Populationsgrösse durchgeführt und schliesslich die Mittelwerte der einzelnen Auswertungen für die Grafik verwendet. Konkret wurde zur Berechnung der Zeit, die benötigt wurde, um die Fitnessgrenze zu erreichen, die durchschnittliche Zeit pro Optimierungsiteration mit der Anzahl der nötigen Iterationen multipliziert. Die genauen Werte der einzelnen Runs sind Anhang 2 zu entnehmen.

Die gelbe Kurve in Abbildung 3.20 bestätigt den Verdacht, dass die Optimierung mit kleiner Population eher in lokalen Optima stecken bleibt, denn hier wurde für eine Population von 10 extrem viel Zeit benötigt, um zur vorgegebenen Fitness zu gelangen. In einem Run wurde sogar die maximale Anzahl EA-Iterationen von 1000 gesprengt.

¹²Dieser Wert wurde folgendermassen berechnet:

$$\frac{\text{ØVerbesserungsfaktor} \cdot \#Iterationen - 1}{\#Iterationen \cdot \text{ØZeitproEA-Iteration}} \cdot 1000 \text{ für Verbesserungsfaktor pro Sekunde}$$

bzw. $\cdot 1000 \cdot 60$ für Verbesserungsfaktor pro Minute

¹³Damit sind Modelle gemeint, welche sich stark von den aktuell am besten bewerteten unterscheiden und somit aus lokalen Optima heraus helfen können.

¹⁴Mit der ersten Entwicklungsstufe ist die Optimierungsiteration gemeint, in der das neue Modell zum ersten Mal auftritt und somit noch nicht verändert bzw. verbessert werden konnte.

¹⁵Es wurde eine Grenze von 1000 für die Anzahl Optimierungsiterationen gesetzt. Beim Überschreiten dieser Grenze wurde in jedem Fall abgebrochen, da dann nicht mehr von einer effizienten Optimierung gesprochen werden kann.

¹⁶Für die Fitnessgrenze von 9 wurden zehn, für jene von 9.5 fünf Runs durchgeführt.

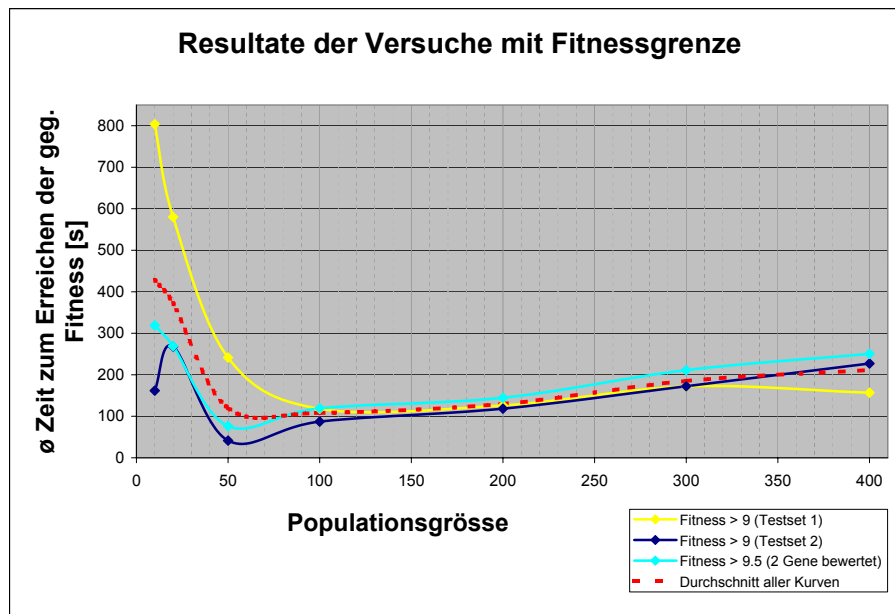


Abbildung 3.20: Zeiten zum Erreichen einer Fitness-Grenze, Muster: Carles

Auffällig ist, dass in allen Versuchen eine Population zwischen 50 und 100 am schnellsten war. Der Grund dafür könnte sein, dass mit dieser Einstellung genügend Kandidaten vorhanden sind, um aus lokalen Optima herauszufinden, wobei gleichzeitig nicht zu viele unnütze Modelle¹⁷ simuliert werden müssen. Dies würde den Geschwindigkeitsvorteil gegenüber grösseren Populationen erklären. Das Resultat sollte jedoch als Tendenz angesehen werden, d.h. in einzelnen Fällen kann es schliesslich doch vorkommen, dass eine andere Populationsgrösse in einer schnelleren Optimierung resultiert.

3.3 Anwendungen

Zur Evaluation unserer Software haben wir versucht, verschiedene Meristem-Muster aus der Literatur nachzubilden. Dazu haben wir drei Muster aus Artikeln von Cristel C. Carles [1] und Olivier Grandjean [3] übernommen und in unsere Darstellungsform übertragen. Anschliessend wurden Optimierungen mit jenen Zielmustern durchgeführt. Die Resultate dieser Simulationen werden in den folgenden Abschnitten 3.3.2 und 3.3.1 diskutiert.

3.3.1 Meristem-Muster nach Grandjean

Die Muster in diesem Abschnitt basieren auf dem Artikel von Grandjean [3]. Im Programm kann man diese zwei Muster als Grandjean1 und Grandjean2 laden. Die Originale findet man im Artikel auf Seite 4, es handelt sich dabei um die Muster H (Grandjean2) und I (Grandjean1). Die Nachbildung der Muster

¹⁷Modelle, welche für das Fortschreiten der Fitness während der Optimierung nicht entscheidend sind.

ist rein qualitativ zu verstehen, die Konzentrationen wurden auf einen beliebigen Wert gesetzt, die einzige Bedingung war, dass es wie das Literaturmuster auszusehen hat. Daher kann es durchaus sein, dass man mittels geschickterer Konzentrationswahl bessere Ergebnisse erzielen würde.

Muster Grandjean1

In Abbildung 3.21 sieht man das Original und seine Umsetzung in unserem Programm, mit leicht unterschiedlichen Farben. Bei den grün eingefärbten Zellen im Originalmuster handelt es sich um Zellen, in denen das Gen Wuschel überwiegt.

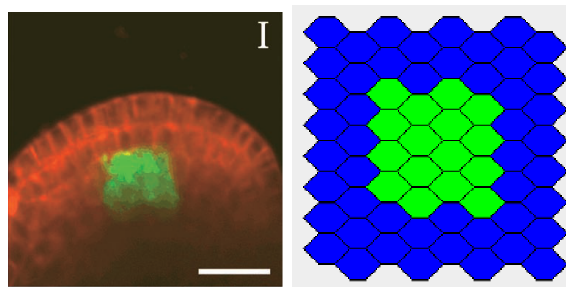


Abbildung 3.21: Muster Grandjean1, Original und Nachbildung (Wuschel)

Die genaue Nachbildung dieses Musters ist aus Abbildung 3.22 oben zu entnehmen. Das blaue Gen hat eine konstante Konzentration von 35 in allen Zellen, was aus der Abbildung ersichtlich ist. Das grüne Gen hat in der Mitte eine Konzentration von 40 und am Rand von 30.

Es wurde mit folgenden Einstellungen optimiert:

- Anzahl Gene: 2
- Anzahl Optimierungsiterationen: 10000
- Populationsgrösse: 50
- Anzahl Simulationszyklen (maximal): 10000
- Normale Randbedingung
- Ohne obere Konzentrationsgrenze

Bei einer Simulation mit normaler Randbedingung wird das Zielmuster perfekt nachgebildet (Abbildung 3.22 mitte). Die Konzentrationsverteilung hingegen variiert stärker gegenüber den Referenzkonzentrationen. Das ist hingegen nicht weiter schlimm, da es ja nur eine Art Richtwert für das Programm darstellt. Es ist auch rein intuitiv klar, dass die vorgegebene Konzentrationsverteilung so nicht erreicht werden kann. Eine Genkonzentration für das erste Gen, die über alle Zellen konstant ist und eine für das zweite Gen, die variiert, ist unmöglich zu erreichen. Desweiteren ist der Konzentrationsprung zwischen zwei Nachbarzellen, die am Übergang liegen, mit einem Betrag von 10 sehr hoch.

Wie man aus der Textausgabe (Abbildung 3.23) sieht wurde das blaue Gen (Gen 1) sehr genau angenähert, beim grünen Gen (Gen 2) gibt es hingegen viel

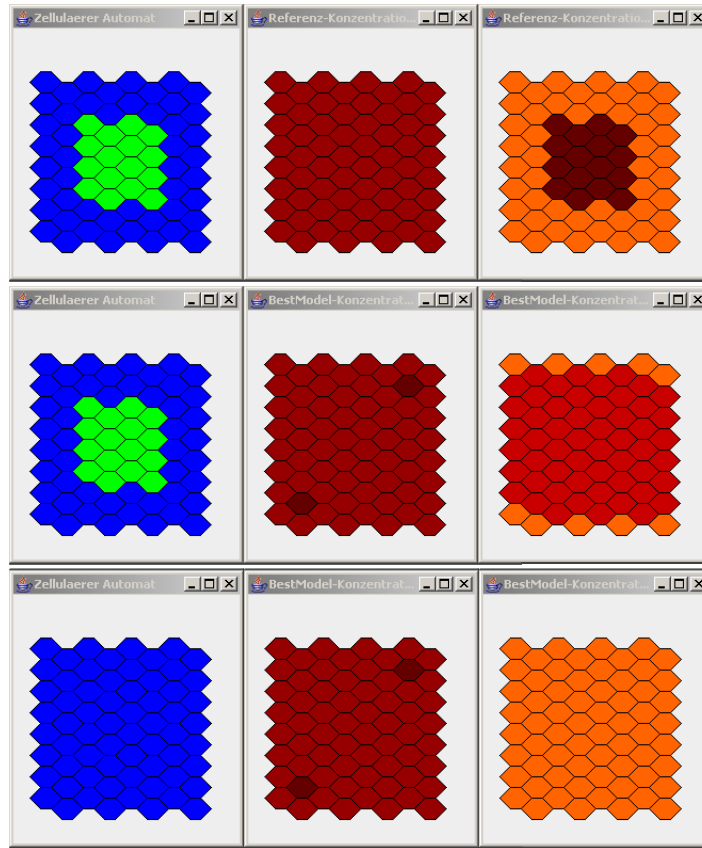


Abbildung 3.22: **oben**: Referenzmuster Grandjean1 **mitte**: Optimiertes Muster mit normaler Randbedingung **unten**: Optimiertes Muster mit spezieller Randbedingung

GEN 1

Optimiertes Muster:

34.73	35.38	34.8	35.2	34.79	35.22	34.9	34.95
35.17	35.31	35.29	35.05	35.26	35.08	35.5	35.07
35.01	35.21	35	34.95	34.97	34.98	35.24	35
34.99	35.21	34.97	34.94	34.94	34.97	35.21	34.99
34.99	35.21	34.97	34.94	34.94	34.97	35.21	34.99
35	35.24	34.98	34.97	34.95	35	35.21	35.01
35.07	35.5	35.08	35.26	35.05	35.29	35.31	35.17
34.95	34.9	35.22	34.79	35.2	34.8	35.38	34.73

Zielmuster:

35	35	35	35	35	35	35	35
35	35	35	35	35	35	35	35
35	35	35	35	35	35	35	35
35	35	35	35	35	35	35	35
35	35	35	35	35	35	35	35
35	35	35	35	35	35	35	35
35	35	35	35	35	35	35	35
35	35	35	35	35	35	35	35

GEN 2

Optimiertes Muster:

28.3	32.98	29.99	33.11	29.99	33.11	29.91	29.86
31.42	34.8	34.8	34.97	34.8	34.98	34.63	31.51
31.55	34.84	35.02	35.02	35.02	35.02	34.85	31.55
31.55	34.84	35.02	35.01	35.01	35.02	34.84	31.55
31.55	34.84	35.02	35.01	35.01	35.02	34.84	31.55
31.55	34.85	35.02	35.02	35.02	35.02	34.84	31.55
31.51	34.63	34.98	34.8	34.97	34.8	34.8	31.42
29.86	29.91	33.11	29.99	33.11	29.99	32.98	28.3

Zielmuster:

30	30	30	30	30	30	30	30
30	30	30	30	30	30	30	30
30	30	40	40	40	40	30	30
30	30	40	40	40	40	30	30
30	30	40	40	40	40	30	30
30	30	40	40	40	40	30	30
30	30	30	30	30	30	30	30
30	30	30	30	30	30	30	30

Abbildung 3.23: Textausgabe der Genkonzentrationen (normale Randbedingung)

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion	
Blau	5.07	8.75	7.89	0	6.62
Grün	3.04	0.55	8.17	-17.1	0

Tabelle 3.8: Optimiertes Modell mit normaler Randbedingung

grössere Differenzen. Wie man sieht konnte im Zentrum des Musters das grüne Gen nur knapp die Oberhand gewinnen, was auf das nicht ganz optimale Referenzmuster zurückzuführen ist. Aber trotzdem kann man sagen, das Programm hat sein Ziel erreicht und ein passendes Modell zum gesuchten Muster geliefert. Das gefundene Modell steht in der Tabelle 3.8.

Das gleiche Muster (Abbildung 3.22 oben) wurde ein zweites Mal optimiert, dieses Mal mit der speziellen Randbedingung. Aufgrund des konstanten Gens 1 könnten diese Einstellungen schneller zu einem besseren Ergebnis kommen, wie wir in Abschnitt 3.1.4 gesehen haben. Rein aufgrund des Referenzbildes im Paper würde man allerdings eine Optimierung mit normaler Randbedingung vorziehen, da man sich hier in der Mitte einer Zellstruktur befindet. Es wurde mit folgenden Einstellungen optimiert:

- Anzahl Gene: 2
- Anzahl Optimierungsiterationen: 2365
- Populationsgrösse: 50
- Anzahl Simulationszyklen (maximal): 10000
- Spezielle Randbedingung
- Ohne obere Konzentrationsgrenze

GEN 1								GEN 2							
Optimiertes Muster:								Optimiertes Muster:							
34.99	35.02	34.99	35.01	34.99	35.01	35	35	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35.01	35.01	35.01	35.01	35.01	35.01	35.02	35	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35	35.01	35	35	35	35	35.01	35	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35	35.01	35	35	35	35	35.01	35	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35	35.01	35	35	35	35	35.01	35	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35	35.01	35	35	35	35	35.01	35	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35	35.01	35	35	35	35	35.01	35	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35	35.02	35.01	35.01	35.01	35.01	35.01	35.01	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
35	35	35.01	34.99	35.01	34.99	35.02	34.99	30.03	30.03	30.03	30.03	30.03	30.03	30.03	30.03
Zielmuster:								Zielmuster:							
35	35	35	35	35	35	35	35	30	30	30	30	30	30	30	30
35	35	35	35	35	35	35	35	30	30	30	30	30	30	30	30
35	35	35	35	35	35	35	35	30	30	40	40	40	40	30	30
35	35	35	35	35	35	35	35	30	30	40	40	40	40	30	30
35	35	35	35	35	35	35	35	30	30	40	40	40	40	30	30
35	35	35	35	35	35	35	35	30	30	40	40	40	40	30	30
35	35	35	35	35	35	35	35	30	30	30	30	30	30	30	30
35	35	35	35	35	35	35	35	30	30	30	30	30	30	30	30

Abbildung 3.24: Textausgabe der Genkonzentrationen (spezielle Randbedingung)

Eine Optimierung mit spezieller Randbedingung liefert entgegen den Erwartungen keine guten Ergebnisse (Abbildung 3.22 unten). Das gesuchte Muster konnte nicht nachgebildet werden. Allerdings wurde hier, wie aus der Textausgabe aus Abbildung 3.24 zu entnehmen ist, das konstante Gen 1 noch genauer angenähert. Bei Gen Nummer 2 tat sich das Programm hingegen schwerer,

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion	
Blau	0.008	0.53	9.90	0	-14.5
Grün	0.01	7.72	5.32	9.8	0

Tabelle 3.9: Optimiertes Modell mit spezieller Randbedingung

es wurde als konstante Verteilung auf dem Niveau der tieferen Konzentration angenähert. Das lässt darauf schliessen das konstante Konzentrationen im Zusammenhang mit Variablen nicht mit der speziellen Randbedingung möglich sind. Das beste vom Optimierer gefundene Modell steht in Tabelle 3.9. Auffällig ist hier, dass beide Gene praktisch nicht diffundieren, was für eine konstante Konzentration spricht.

Muster Grandjean2

Aus der Abbildung 3.25 kann man das Originalmuster und seine Umsetzung in unserem Programm betrachten. Hier stimmen die Farben nicht überein, es sollte hingegen klar sein auf was man hinaus will. Die gelb eingefärbten Randzellen des Originalbildes enthalten eine hohe Konzentration des ATML1 Gens.

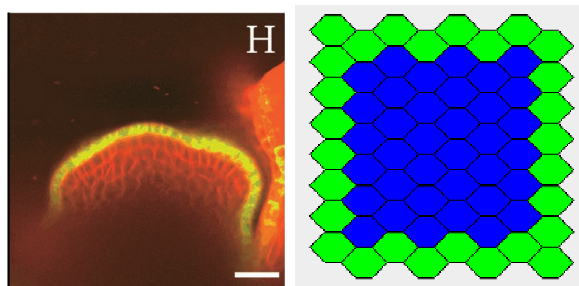


Abbildung 3.25: Muster Grandjean2, Original und Nachbildung (ATML1)

Es wurde wiederum eine qualitative Nachbildung des Musters durchgeführt. Graphisch aufgezeichnet wurde die Referenzkonzentration in Abbildung 3.26 oben.

In diesem Muster wurde die Konzentration von Gen 1 (blau) in allen Zellen auf 25 gesetzt. Das grüne Gen 2 wurde am Rand mit der Konzentration 30 versehen, im Inneren mit der Konzentration 20.

In einem ersten Run wurde mit normaler Randbedingung gerechnet mit folgenden Einstellungen:

- Anzahl Gene: 2
- Anzahl Optimierungsiterationen: 10000
- Populationsgrösse: 50
- Anzahl Simulationszyklen (maximal): 10000
- Normale Randbedingung

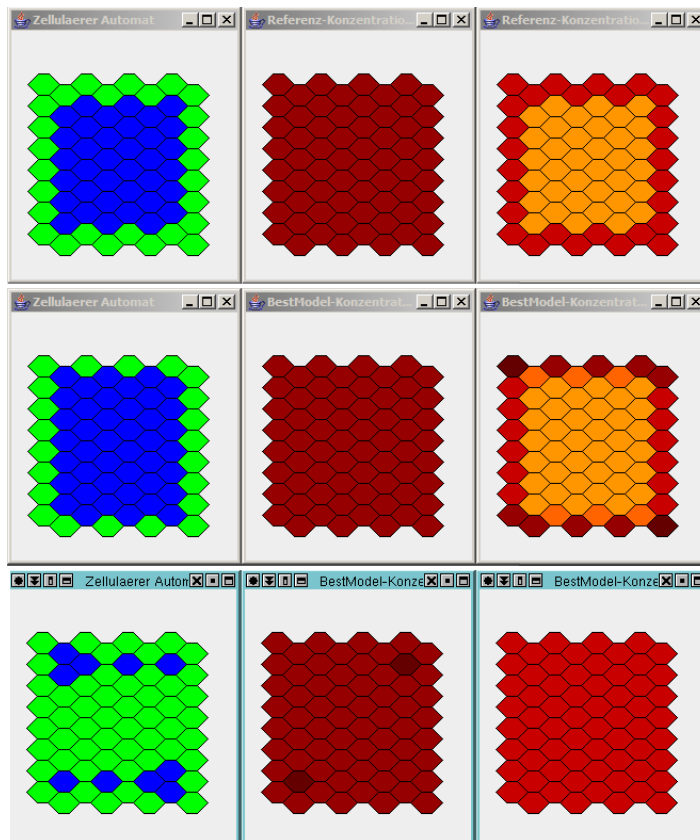


Abbildung 3.26: **oben**: Referenzmuster Grandjean2 **mitte**: Optimiertes Muster mit normaler Randbedingung **unten**: Optimiertes Muster mit spezieller Randbedingung

- Ohne obere Konzentrationsgrenze

Die Optimierung mit normaler Randbedingung liefert hier kein perfektes Ergebnis, hingegen ist die Tendenz des Zielmusters klar ersichtlich (Abbildung 3.26 mitte). Lediglich sechs Zellen unterscheiden sich vom Referenzmuster. Das konstante Gen, hier Gen 1, wurde wiederum sehr genau angenähert. Nur in den Ecken ist es etwas schlechter, aber auch nur um c.a. 0.4 (siehe Abbildung 3.27). Auch beim Gen mit der "variablen" Konzentration zeigt sich das gleiche Bild wie im vorhergehenden Muster Grandjean1. Die Zellen in der Mitte haben fast die optimale Konzentration, die Randzellen sind hingegen etwas ungenau. Dieses Modell hat aber den Konzentrationsprung zwischen Randzellen und mittleren Zellen gut angenähert. Die Randzellen, die die falsche Farbe haben, haben eine Konzentration des grünen Gens (Gen 2) von ungefähr 24, die mit der richtigen Farbe von ungefähr 28 bis 31. Es stellt sich die Frage, wie können solch grosse Unterschiede entstehen, da zum einen das Gen 1 fast eine konstante Konzentration aufweist. Es hilft ein Blick auf das Modell in Tabelle 3.10. Auffällig sind dabei die hohe Selbstproduktionsrate von Gen 2 und die starke Interaktion von Gen 1 auf Gen 2. Wie man aus Abschnitt 2.3.1 weiss, sind diese Faktoren eng gekoppelt. Die Konzentration von Gen 1 entscheidet also im wesentlichen wie stark die tatsächliche Eigenproduktion des Genes 2 ausfällt. Da diese Interaktion und Eigenproduktion so stark ist, genügt schon eine kleine Konzentrationsänderung von Gen1, um die Konzentration von Gen 2 stark zu beeinflussen. Das Ganze wird unterstützt von der Tatsache, dass Gen 2 fast nicht diffundiert und so diese Konzentrationsdifferenz nicht ausgleichen kann.

GEN 1								GEN 2							
Optimiertes Muster:								Optimiertes Muster:							
24.63	24.93	24.73	24.93	24.73	24.93	24.73	24.73	35.92	24.11	31.93	24.06	31.93	24.06	31.97	31.99
24.83	25.03	25.03	25.03	25.03	25.03	25.03	24.83	28.05	20.07	20.07	20	20.07	20	20.14	28.02
24.83	25.03	25.03	25.03	25.03	25.03	25.03	24.83	28	20.06	19.99	19.99	19.99	19.99	20.06	28
24.83	25.03	25.03	25.03	25.03	25.03	25.03	24.83	28	20.06	19.99	19.99	19.99	19.99	20.06	28
24.83	25.03	25.03	25.03	25.03	25.03	25.03	24.83	28	20.06	19.99	19.99	19.99	19.99	20.06	28
24.83	25.03	25.03	25.03	25.03	25.03	25.03	24.83	28	20.06	19.99	19.99	19.99	19.99	20.06	28
24.83	25.03	25.03	25.03	25.03	25.03	25.03	24.83	28.02	20.14	20	20.07	20	20.07	20.07	28.05
24.73	24.73	24.93	24.73	24.93	24.73	24.93	24.63	31.99	31.97	24.06	31.93	24.06	31.93	24.11	35.92
Zielmuster:								Zielmuster:							
25	25	25	25	25	25	25	25	30	30	30	30	30	30	30	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	30	30	30	30	30	30	30

Abbildung 3.27: Textausgabe der Genkonzentrationen (normale Randbedingung)

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion
Blau	7.99	0.4	8	0 -35.4
Grün	0.04	9.03	0.83	7.98 0

Tabelle 3.10: Optimiertes Modell mit normaler Randbedingung

Auch hier wurde das gleiche Muster noch auf den Einfluss durch die spezielle Randbedingung getestet. Auf die Wirklichkeit bezogen macht die spezielle Randbedingung durchaus Sinn, wie man in Abbildung 3.25 sehen kann. Rein

aus den Erfahrungen, die mit dem Muster Grandjean1 gemacht wurden, sind aber keine guten Ergebnisse zu erwarten. Es wurde mit folgenden Einstellungen optimiert:

- Anzahl Gene: 2
- Anzahl Optimierungsiterationen: 10000
- Populationsgrösse: 50
- Anzahl Simulationszyklen (maximal): 10000
- Spezielle Randbedingung
- Ohne obere Konzentrationsgrenze

GEN 1								GEN 2							
Optimiertes Muster:								Optimiertes Muster:							
24.85	25.04	24.87	24.99	24.87	24.99	24.9	24.91	24.99	24.99	24.99	24.99	24.99	24.99	24.99	24.99
24.96	25	25	24.96	25	24.96	25.04	24.94	24.99	24.99	24.99	24.99	24.99	24.99	24.99	24.99
24.93	24.99	24.96	24.96	24.96	24.96	24.99	24.93	24.99	24.99	24.99	24.99	24.99	24.99	24.99	24.99
24.93	24.99	24.96	24.96	24.96	24.96	24.99	24.93	24.99	24.99	24.99	24.99	24.99	24.99	24.99	24.99
24.93	24.99	24.96	24.96	24.96	24.96	24.99	24.93	24.99	24.99	24.99	24.99	24.99	24.99	24.99	24.99
24.94	25.04	24.96	25	24.96	25	25	24.98	24.99	24.99	24.99	24.99	24.99	24.99	24.99	24.99
24.91	24.9	24.99	24.87	24.99	24.87	25.04	24.85	24.99	24.99	24.99	24.99	24.99	24.99	24.99	24.99
Zielmuster:								Zielmuster:							
25	25	25	25	25	25	25	25	30	30	30	30	30	30	30	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	20	20	20	20	20	20	30
25	25	25	25	25	25	25	25	30	30	30	30	30	30	30	30

Abbildung 3.28: Textausgabe der Genkonzentrationen (spezielle Randbedingung)

Gen	Diffusion	Eigenprod.	Eigenred.	Interaktion
Blau	0.06	0.74	7.90	0 -15.6
Grün	0.12	6.0	8.51	4.9 0

Tabelle 3.11: Optimiertes Modell mit spezieller Randbedingung

Wie erwartet zeigt sich hier ein äquivalentes Bild wie bei Grandjean1 mit spezieller Randbedingung. Das blaue Gen mit der Nummer 1 wurde hingegen noch etwas genauer optimiert als im Modell mit normaler Randbedingung. Das grüne Gen (Gen 2) wurde, analog zum Gen 2 bei Grandjean1, mit spezieller Randbedingung, konstant. Der Konzentrationswert liegt exakt zwischen den beiden vorkommenden Werten (30 und 20) dieses Gens, also genau bei 25 (Abbildung 3.28). Für diese Tatsache spricht auch hier die praktisch inexistente Diffusion (Tabelle 3.11). Es ist auch zu beachten, dass die Genkonzentrationen beider Gene nahezu identisch sind, und es ist unklar, welches Gen jetzt dominieren¹⁸ würde (Siehe dazu Abbildung 3.26 unten).

¹⁸dominieren bedeutet hier, dass die Genkonzentration eines Gens in einer Zelle die Konzentrationen aller anderer Gene deutlich übersteigt und somit die Funktion der Zelle bestimmt.

Nach den bisherigen Erkenntnissen kann man sagen, dass die spezielle Randbedingung die Muster bzw. Modellfindung behindert. Nur für den Fall, dass man konstante Genkonzentrationen im ganzen Muster hat, bringt sie Vorteile. Nur ist ein solches Szenario in der Wirklichkeit nicht sehr wahrscheinlich.

3.3.2 Meristem-Muster nach Carles

Hier wurde versucht, ein Muster zu generieren, welches in Arbeiten von Cristel C. Carles über das Meristem von *Arabidopsis thaliana* [1] zu finden ist. Abbildung 3.29 zeigt die Umsetzung des Musters aus der Literatur auf unsere Darstellungsweise, wobei für einen besseren Vergleich ähnliche Farben verwendet wurden.

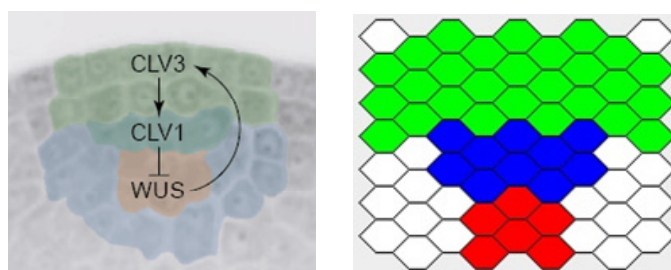


Abbildung 3.29: Links: Referenzmuster aus der Literatur, Rechts: Umsetzung in verwendetes Modell

Es wurden Optimierungen mit unterschiedlicher Anzahl Genen durchgeführt, wobei jedoch nur immer die drei Gene aus dem Referenzmuster in die Bewertung eines optimierten Musters einfließen. Durch eine höhere Anzahl Genen müssten solche, eher komplexeren Muster wie das vorliegende eher erreichbar sein.

In der Abbildung 3.30 ist eine Gegenüberstellung der Resultate¹⁹ zu sehen. Diese werden anschliessend diskutiert. In Anhang 3 ist für jede betrachtete Einstellung die Textausgabe mit genauen Genkonzentrationen gegeben

Es wurde mit folgenden Einstellungen von Seiten der Benutzeroberfläche optimiert. Die Anzahl Gene und die Populationsgrösse sind für jeden Optimierungsversuch durch Komma getrennt aufgeführt:

- Anzahl Gene: 3, 4, 6, 8, 10
- Anzahl Optimierungsiterationen: 2176
- Populationsgrösse: 100, 20, 50, 100, 150
- Anzahl Simulationszyklen (maximal): 5000
- Normale Randbedingung (für 4 Gene mit spezieller Randbedingung)
- Ohne obere Konzentrationsgrenze

Zur genauen Höhe der Genkonzentrationen sollten stets die Textausgaben in Anhang 3 konsultiert werden.

¹⁹Die Resultate sind nicht quantitativ - also bezüglich Farbgebung - vergleichbar, da es sich um verschiedene Optimierungen handelt, welche in unterschiedlichen Farbkodierungen resultieren.

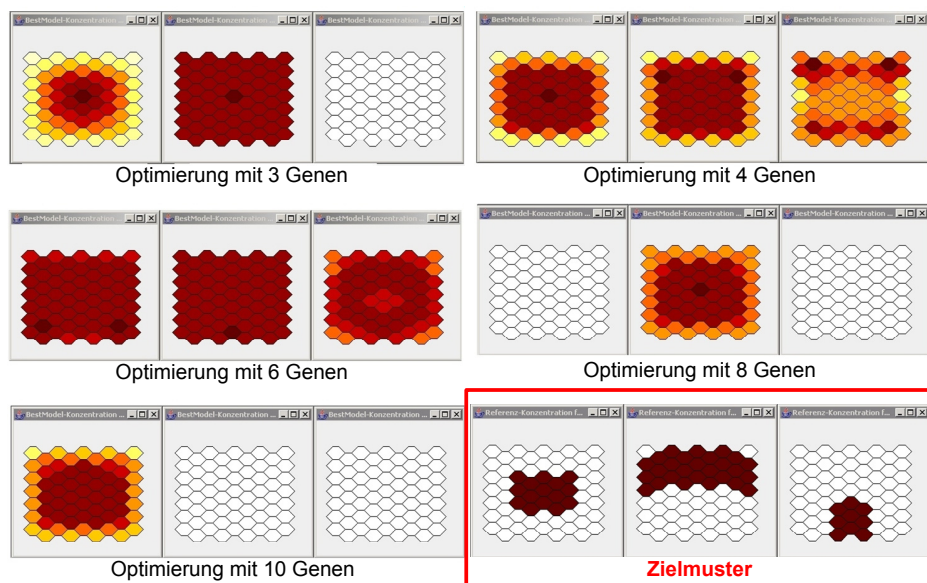


Abbildung 3.30: Resultat der Optimierung mit 3 Genen

Mit 3 Genen

Die Form des Vorkommens von Gen 1 wurde hier ansatzweise angenähert, für Gen 2 stimmt bloss die Höhe der Genkonzentration mit jener des Zielmuster überein. Gen 3 ist im optimierten Muster gar nicht vorhanden.

Betrachtet man den Verlauf der Fitness in Abbildung 3.31, lässt sich nicht erwarten, dass z.B. nach doppelt so vielen Optimierungsiterationen ein viel passenderes Muster resultieren würde, da nach 1000 Iterationen keine Verbesserung mehr erkennbar ist.

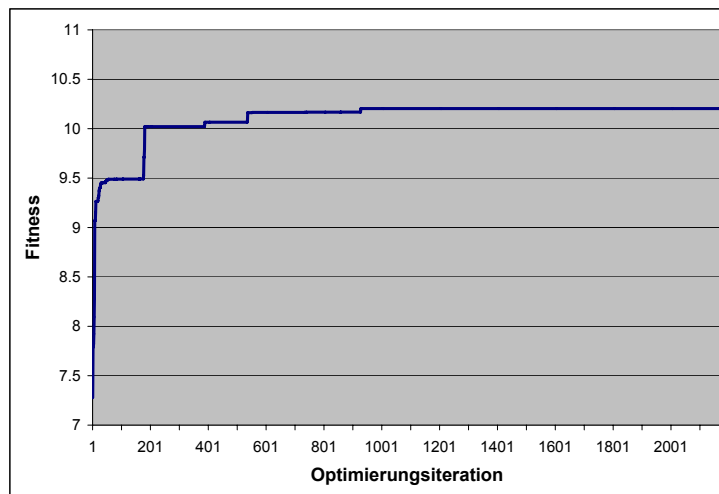


Abbildung 3.31: Fitnessverlauf der Optimierung mit 3 Genen

Mit 4 bis 10 Genen

Für mehr als 3 Gene lässt sich das vorgegebene Muster kaum im Resultat der Optimierung wiedererkennen. Für die Optimierung mit 4 bzw. 10 Genen ist dies für Gen 1 noch ansatzweise der Fall und das Gen konzentriert wie im Referenzmuster sich in der Mitte.

Die Höhe der Genkonzentration wurde generell mit nur geringen Abweichungen erreicht. Also an jenen Stellen, wo im Zielmuster eine Genkonzentration grösser als Null vorhanden ist, wurde diese stets - falls im optimierten Muster kein leeres Genfeld²⁰ generiert wurde - mit Abweichungen von weniger als 10 angenähert.

Da sich die Komplexität des Problems²¹ mit 4 Genen bereits stark erhöht (Anzahl Parameter steigt von 18 auf 28 an), ist anzunehmen, dass für diesen Fall schon bedeutend mehr Optimierungssiterationen nötig sind. Ausserdem ist fragwürdig, ob der Optimierer bei immens längerer Rechenzeit nicht ohnehin in einem lokalen Fitness-Maximum hängen bleiben würde, denn dieses Problem trat teilweise bereits nach wenigen Iterationen auf.

Auffällig ist bei der Optimierung mit 8 und 10 Genen, dass jeweils 2 leere Genfelder generiert wurden. Bei den dazugehörigen Modellen scheint immer noch eine kleinere Differenz zum Zielmuster zu resultieren als bei allen anderen Modellen. Ausserdem wird die Simulation für derart viele Gene noch rechenintensiver und es wäre daher angebracht, die Optimierung mehrere Tage lang laufen zu lassen, um zu einem guten Ergebnis zu kommen. Leider stand gegen Ende dieser Arbeit nicht mehr genügend Zeit hierzu zur Verfügung.

²⁰Leeres Genfeld: Für betreffendes Gen Konzentration in allen Feldern gleich Null.

²¹Die Komplexität, also die Anzahl Parameter des Modells steigt in etwa quadratisch zu der Anzahl Gene an: $O(n) = n^2 + 2n$. Siehe dazu Abschnitt 3.1.2.

Fazit

Dem Optimierungsverfahren fällt es sehr schwer, das Muster von Carles [1] zu approximieren. Dies liegt wohl vor allem daran, dass es sich um ein Muster handelt, welches nicht von einem unserer Modell²² generiert wurde und schon darum nur näherungsweise erreicht werden kann. Dazu kommt, dass das hier untersuchte Muster eine komplizierte Geometrie aufweist und es deshalb innerhalb aller möglichen Modelle wahrscheinlich verschwindend wenige gibt, welche ein ähnliches Muster hervorbringen.

Auffällig ist, dass die Resultate der Optimierung für mehr Gene als die drei eigentlich zu optimierenden immer schlechter ausfallen, als wenn bloss jene drei Gene verwendet werden. Entweder handelt es sich dabei um einen Zufall - also dass durch Mutation und Rekombination einfach schlechte Modelle entstanden - oder die Komplexität ist zu gross, als dass man mit einer Optimierungs-Laufzeit von einigen Stunden eine akzeptable Lösung finden könnte. Die zweite Erklärung scheint jedoch plausibler.

²²Siehe Abschnitt 2.3.1

Kapitel 4

Schlussfolgerungen

4.1 Erfolge

Es wurde prinzipiell gezeigt, dass es mit dem verwendeten Ansatz möglich ist, Gennetzwerke zu simulieren und bis zu einem gewissen Komplexitätsgrad für vorgegebene Muster zu optimieren.

Es konnte zu Referenzmustern, die vom Programm selber erzeugt wurden, problemlos passende optimierte Modelle wiedergefunden werden. Auch zu einfacheren Mustern aus der Literatur erzeugte das Programm passende Modelle. Das Muster Grandjean1 konnte perfekt nachgebildet werden, bei Grandjean2 konnte zumindest die Tendenz ganz klar gezeigt werden.

Bei komplizierten Mustern wurden die Limitationen des Programmes allerdings klar aufgezeigt. Das aufwändige Meristem-Muster aus dem Artikel von Carles [1] konnte nicht gefunden werden. Problematisch war auch die Tatsache, dass die Optimierung oft in lokalen Maxima gefangen war, was dazu führte, dass keine gute Annäherung gemacht werden konnte. Die Gründe für diese Tatsache konnten nicht eruiert werden, sind aber bei derart komplexen Optimierungsproblemen oft zu beobachten. Desweiteren verlangsamt sich die Optimierung extrem bei grossen Zellstrukturen (ab Feldgrösse 10x10) und vielen Genen (für Anzahl Gene grösser als 3). Dann ist die Software nicht mehr vernünftig auf einem Desktop PC einsetzbar.

Trotz dieser Schwierigkeiten kann man von einem Erfolg reden, da auch zu Beginn der Arbeit nicht klar war ob so etwas überhaupt realisierbar ist. Mit diversen Erweiterungen und Verbesserungen (siehe auch Abschnitt 4.3) liessen sich noch bessere Resultate erzielen.

4.2 Probleme

- **Performance:**

Die Software im Allgemeinen lässt von der Geschwindigkeit her zu wünschen übrig. Daher sind die durchgeführten Optimierungen sehr zeitaufwendig. Dies liegt wohl auch daran, dass sie in JavaTM implementiert wurde und so auf einer virtuellen Maschine läuft, was gewisse Leistungs Nachteile bringt.

Eine Implementierung in C++ könnte bereits einige Geschwindigkeitsvor-

teile bringen und auch eine Optimierung des Codes bezüglich Ausführungsgeschwindigkeit würde wahrscheinlich schon kleine Verbesserungen bringen.

- **Innovative Modelle haben's schwer...**

Neu generierte Modelle, die stark vom zurzeit besten Modell abweichen, erhalten sehr selten eine bessere Bewertung als das beste Modell. Somit besteht teilweise bereits nach kurzer Zeit die ganze Eltern-Population aus Abkömmlingen eines einzigen Modells, welche alle nahezu die gleiche Fitness sowie die selbe benötigte Anzahl Simulationsiterationen aufweisen. Es ergeben sich so zwar nach wie vor winzige Fortschritte hinsichtlich der Fitness, jedoch sind die Verbesserung minim. Auch Versuche mit höherer Mutationsrate zur Generierung von stark abgeänderten Modellen führten nicht zum gewünschten Erfolg. Das "fewBest"-Feature (siehe Abschnitt 2.4.7 hilft auch nicht weiter, denn die dabei gewonnenen "normalen" Eltern-Modelle vermögen nicht gegen die fix gesetzten besten Eltern zu bestehen. Ein Lösungsansatz wäre, die normalen Eltern durch eine neue zufällige Startpopulation zu ersetzen, um neue Inputs zu kreieren. Allerdings entscheidet auch bei dieser Lösung der Zufall, ob neue wirklich gute Modelle entstehen.

Womöglich handelt es sich hierbei um ein Problem bei evolutionären Algorithmen im Allgemeinen.

4.3 Ausblick

Zum Erreichen besserer Optimierungsergebnisse sind noch verschiedene Änderungen denkbar. Es folgt eine Liste von möglichen Erweiterungen bzw. Verbesserungen des Programms:

1. Optimierer

- **Schlechte Startpopulation verhindern:**

Beim aktuellen Optimierer kommt es häufig vor, dass man einen "unglücklichen Start" erwischt. D.h., die Startpopulation besteht aus relativ schlechten Modellen, so dass es vorkommen kann, dass ein weiterer Optimiervorgang bereits nach erheblich weniger Optimierungsiterationen das bessere Resultat erzielt. Daher auch die breite Streuung an EA-Iterationen in Abschnitt 3.1.3.

Bei der aktuellen Implementierung lässt sich dieses Problem auch nicht mit grösseren Populationen lösen, da sich das System auch dann schnell auf ein erfolgreiches Modell und dessen ähnlich gut bewertet Nachkommen konzentriert.

Eine Möglichkeit dem beschriebenen Verhalten entgegenzuwirken, wäre, mehrere Populationen parallel durch das Optimierungsverfahren laufen zu lassen. Das Resultat davon ist das selbe, wie wenn mehrere Male hintereinander eine Optimierung gestartet wird, wobei anzunehmen ist, dass eine der Optimierungen eine akzeptable Startpopulation schafft. Denkbar ist auch, Rekombination von Modellen verschiedener Populationen zu erlauben, um so die "guten Eigenschaften" der besten Modelle unterschiedlicher Populationen zu kombinieren.

- **Neue Fitness-Bewertungsmethoden:**

Die Bewertungsmethode kann dahingehend verbessert werden, dass sie dem Zielmuster ähnliche Muster besser erkennt. Einige Vorschläge dazu sind:

- Komplexere Methoden zur Bewertung von Endmustern könnten z.B. miteinbeziehen, ob an einer gewissen Stelle in der Zellstruktur tatsächlich das Gen konzentrationsmässig überwiegt, welches dies nach dem Zielmuster tun sollte.
- Die Bewertungen verschiedener Funktionen könnten zusammen in die endgültig abgegebene Wertung mit einfließen. So liessen sich die guten Ansätze verschiedener Methoden kombinieren. Eine Möglichkeit wäre, auch einen qualitativen Vergleich der Muster zu berücksichtigen, welcher nur das Vorhandensein von Genen überhaupt beurteilt und dabei nicht auf die exakte Konzentration eingeht.

- **Wie der Vater so der Sohn...**

Um eine übermässige Fortpflanzung des besten Modells zu vermeiden und so mehr andere, neuartigen Modellen in die Population miteinzubeziehen, könnte das generierte Kind-Modell zuerst mit den Eltern-Modellen verglichen werden und schliesslich nur dann akzeptiert werden, wenn sich eine minimale Änderung zeigt.

Auf diese Weise würde verhindert, dass die ganze Population aus Kandidaten besteht, welche sich voneinander kaum unterscheiden. Dadurch würde erreicht, dass eher aus lokalen Fitness-Maxima herausgefunden wird

- **Vorgegebene Topologie:**

Um die Komplexität der Optimierung zu verringern, könnte die Interaktionsmatrix oder zumindest die Vorzeichen deren Elemente bereits vor der Optimierung festgesetzt werden. So müssen die Parameter der Interaktionsmatrix nicht mehr in der Optimierung berücksichtigt werden, sondern bleiben fix. Auf diese Weise könnten auch bereits bekannte Topologien aus der Literatur in der Software verwendet werden.

- **Mehr Rechenleistung!**

Eine Frage ist, ob sich mit mehr Rechenleistung und -Zeit deutlich bessere Modelle finden liessen. Die Durchführung von Optimierungen auf leistungsstarken Grossrechnern wäre zu diesem Zweck ideal, da auf diesen eine parallele Abarbeitung der Optimierung auf mehreren Prozessoren möglich ist.

2. Benutzungsoberfläche

- **Bessere Visualisierung der Genkonzentrationen:**

Sämtliche Genkonzentrationen liessen sich in einem Bild darstellen, indem z.B. die Konzentrationsverhältnisse zwischen den Genen auf eine einzige Farbe kodiert werden¹. Bei mehr als zwei Genen würde jedoch auch diese Darstellung schnell unübersichtlich und so liessen

¹Als einfaches Beispiel: Falls Gen "blau" und Gen "gelb" in einer Zelle vorhanden → Zelle grün färben.

sich die Konzentrationen nicht mehr so leicht anhand der Farben ablesen, da viele ähnliche Farbkodierungen nötig wären.

Weitere Darstellungsmöglichkeiten, wie z.B. jede Zelle nach Anzahl Genen zu unterteilen und schliesslich die einzelnen Teile je nach Konzentration des entsprechenden Gens zu färben, sind auch nicht unbedingt geeignet, um die Konzentrationen sämtlicher Gene auf einen Blick zu erfassen. Bei einer solchen Darstellung muss der Benutzer zuerst jede Zelle einzeln betrachten, um sich ein Bild des Ganzen machen zu können.

3. Simulator

- **Oszillationsdetektion:**

Es kann in Erwägung gezogen werden, Modelle, welche in oszillierenden Mustern resultieren ähnlich gut zu bewerten, wie solche, die stabile Muster generieren. Voraussetzung dazu ist, dass solche Oszillationen auch erkannt werden können. Ein Verfahren hierzu müsste noch implementiert werden, könnte jedoch - je nachdem über wieviele Zeitschritte hinweg die Erkennung einer Schwingung möglich sein soll - sehr rechenintensiv werden.

Ausserdem muss entschieden werden, wie die einzelnen Schritte einer Schwingung - also die Muster der einzelnen Schwingungszustände - mit in die Bewertung des Modells einfließen. Dies dürfte bei starken Schwingungen nicht allzu einfach sein.

Da schwingende Muster in der aktuellen Implementation als nicht stabil klassifiziert werden und deshalb eine schlechte Bewertung erhalten, könnte eine mildere Bewertung derartiger Modelle in zusätzlichen Kandidaten mit guter Fitness resultieren.

Anhang A

A.1 Zeitplan

A.2 Optimale Popgrösse - Tabellen

Fitnessgrenze 9.5

Populationsgrösse	Lauf 1		Lauf 2		Lauf 3		Lauf 4		Lauf 5		Durchschnitt über alle Runs		Durchschnittliche Zeit [s] zum Erreichen der Fitnessgrenze	
	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen		
10	1556.39	233	2052.92	13	1415.85	13	2641.15	118	1805.58	465	1894	168.4	319	5.3
20	2106.67	6	4921.96	185	5407.35	156	1833.33	9	3119.45	31	3478	77.4	269	4.5
50	7474.24	17	4513.43	7	7950.5	12	6636.7	20	4756.4	5	6266	12.2	76	1.3
100	13046.8	18	9593.77	13	10400	5	8960.93	14	8526	9	10106	11.8	119	2.0
200	27427.1	12	19966.2	6	19820.3	6	22171	1	23625	7	22602	6.4	145	2.4
300	29192.7	6	27921.8	5	31253.9	8	31420	8	31099.6	8	30178	7	211	3.5
400	43460.1	9	42068.6	5	39665.6	5	38901.7	7	37925	5	40404	6.2	251	4.2

Fitnessgrenze 9 (Versuch 1)

Populationsgrösse	Lauf 1		Lauf 2		Lauf 3		Lauf 4		Lauf 5		Lauf 6		Lauf 7		Lauf 8		Lauf 9		Lauf 10		Durchschnittliche Zeit [s] zum Erreichen der Fitnessgrenze		
	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen			
10	2144	45	1355	33	1965	201	2334	45	2839	1000	994	5	3534	1000	4043	825	2583	113	1713	153	2350	342	803.79667
20	3273	174	2883	121	3236	85	1973	16	3016	21	6122	267	2713	27	6759	511	6570	147	3583	77	4013	144.6	580.25766
50	5198	12	8800	18	7776	21	4924	7	8332	18	11678	74	4896	12	6153	17	17453	66	7721	46	8293	29.1	241.32919
100	9820	11	10594	5	9380	6	10406	5	34866	17	10195	8	10368	9	11943	8	9753	6	13876	15	13119	9	118.07144
200	19472	9	20223	7	19016	5	20620	3	19744	5	21242	6	16263	10	20058	10	19523	6	21010	3	19717	6.4	126.18903
300	28292	7	29427	3	28435	6	27388	7	30055	4	30134	7	30620	6	28476	7	31837	7	26866	5	29153	5.9	172.00252
400	39275	5	42734	1	42781	6	40946	4	40623	6	47245	3	40191	4	44750	2	37789	4	36115	3	41245	3.8	156.73037

Fitnessgrenze 9 (Versuch 2)

Populationsgrösse	Lauf 1		Lauf 2		Lauf 3		Lauf 4		Lauf 5		Lauf 6		Lauf 7		Lauf 8		Lauf 9		Lauf 10		Durchschnittliche Zeit [s] zum Erreichen der Fitnessgrenze		
	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen	Zeit pro Iteration	#Iterationen			
10	896	79	1178	97	763	43	953	3	786	7	1377	16	1542	83	1211	27	1287	1000	1149	97	1114	145	161.78769
20	992	8	3052	9	2054	13	1901	28	2830	7	2241	9	3752	1000	2468	18	2076	15	2688	7	2405	111.4	267.96900
50	5221	7	5883	11	5803	5	5688	3	6813	4	5840	8	4966	6	7123	6	5434	12	4747	10	5752	7.2	41.41203
100	12049	9	12493	7	15630	3	12086	10	11212	7	12059	5	10947	11	12333	6	11518	7	10743	7	12107	7.2	87.17137
200	25503	6	29164	2	26766	1	28294	5	23766	5	24989	7	21281	6	28669	6	25656	4	26010	4	26010	5	118.48884
300	36002	7	39555	2	40859	6	33643	6	44453	1	50964	3	36669	5	34075	5	41949	4	41863	4	38785	4	172.37972
400	59531	1	49164	6	50699	7	50272	5	54177	3	50964	3	51744	5	52693	3	45402	7	51590	4	51623	4.4	227.14315

A.3 Textausgabe zu Optimierung "Carles"

A.4 Verifizierung Simulator (Excel™)

20.0300565036	2.0030055036	1.9450035020	0.8379614260	1.8675415703	1.8572932475	1.8518502795	1.4006011007	0.0662292149	0.1001502752	0.1303822777	0.1001502752
2.4201230031	2.6237639924	2.3486311502	1.6089231673	1.8372932475	1.8060549704	1.8518502795	1.4006011007	0.0662292149	0.1210061502	0.1411881999	0.1210061502
2.2191239488	2.2056474444	2.2103353160	1.2415890743	1.8518502795	1.8060549704	1.8518502795	1.4006011007	0.0662292149	0.1109561974	0.1102823750	0.1109561974
1.3245842972	1.7625373886	1.3037923236	0.5649341524	1.3037923236	1.5647527996	1.4411294889	1.4006011007	0.0662292149	0.0662292149	0.0881268694	0.0662292149
1.6270675255	1.9194502964	1.5751772662	1.0813448645	1.0813448645	1.5647527996	1.4411294889	1.4006011007	0.0662292149	0.0813533763	0.0959725148	0.0813533763
1.4814972049	1.4713967476	0.8137407844	0.8363190949	0.8363190949	1.4411294889	1.4411294889	1.4006011007	0.0662292149	0.0740748602	0.0735698374	0.0740748602
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2.0632402254	2.7107177463	2.0151135500	0.8702249126	0.8702249126	1.8617943529	1.8454604561	1.4006011007	0.0662292149	0.1031620113	0.1355358873	0.1031620113
2.5106317293	2.9429373905	2.5106317293	1.6687258719	1.6687258719	1.8617943529	1.8454604561	1.4006011007	0.0662292149	0.1255315865	0.1471468895	0.1255315865
2.2954598695	2.2803740268	1.2556428549	1.2889905244	1.2889905244	1.8617943529	1.8454604561	1.4006011007	0.0662292149	0.1147729935	0.1140187013	0.1147729935
1.3820564713	1.8600310623	1.3820564713	0.5954609147	0.5954609147	1.8617943529	1.8454604561	1.4006011007	0.0662292149	0.0891028236	0.0930001553	0.0891028236
1.7127623818	2.0319010219	1.7127623818	1.1379858311	1.1379858311	1.8617943529	1.8454604561	1.4006011007	0.0662292149	0.0656381191	0.1015953811	0.0656381191
1.5539543869	1.5423871165	1.5539543869	0.8811750867	0.8811750867	1.8617943529	1.8454604561	1.4006011007	0.0662292149	0.0776977193	0.0771193568	0.0776977193
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2.1198019282	2.8090037069	2.1198019282	0.9009606780	0.9009606780	1.8562248562	1.8044057219	1.4006011007	0.0662292149	0.1059900964	0.1404501853	0.1059900964
2.5967603611	3.0569978704	2.5967603611	1.7255999329	1.7255999329	1.8562248562	1.8044057219	1.4006011007	0.0662292149	0.1298380181	0.1528498935	0.1298380181
2.3677960918	2.3510751684	1.2987650000	1.3341389000	1.3341389000	1.8562248562	1.8044057219	1.4006011007	0.0662292149	0.1183898046	0.1175537584	0.1183898046
1.4377514378	1.9559427805	1.4377514378	0.6254793883	0.6254793883	1.8562248562	1.8044057219	1.4006011007	0.0662292149	0.0718875719	0.0977971390	0.0718875719
1.7969335494	2.1430072251	1.7969335494	1.1935862210	1.1935862210	1.8562248562	1.8044057219	1.4006011007	0.0662292149	0.0898466775	0.1071503613	0.0898466775
1.6248158824	1.6117074464	0.8987731650	0.9252747035	0.9252747035	1.8562248562	1.8044057219	1.4006011007	0.0662292149	0.0812407941	0.0805653723	0.0812407941
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2.1729412792	2.9026675973	2.1729412792	0.9302264630	0.9302264630	1.8508268486	1.7949711517	1.4006011007	0.0662292149	0.1086470640	0.1451333799	0.1086470640
2.6786911810	3.1660955515	2.6786911810	1.7796702769	1.7796702769	1.8508268486	1.7949711517	1.4006011007	0.0662292149	0.1339345590	0.1583034776	0.1339345590
2.4363432333	2.4179717021	1.4363432333	1.3771220724	1.3771220724	1.8508268486	1.7949711517	1.4006011007	0.0662292149	0.1218171617	0.1208988851	0.1218171617
1.4917315142	2.0520288487	1.4917315142	0.6548729731	0.6548729731	1.8508268486	1.7949711517	1.4006011007	0.0662292149	0.0745866575	0.1025144241	0.0745866575
1.8795493561	2.2526505753	1.8795493561	1.2481273632	1.2481273632	1.8508268486	1.7949711517	1.4006011007	0.0662292149	0.0939774678	0.1126325288	0.0939774678
1.6940936069	1.6793769844	0.9401396315	0.9685961526	0.9685961526	1.8508268486	1.7949711517	1.4006011007	0.0662292149	0.0847046803	0.0839688492	0.0847046803
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2.2288777605	2.9918735226	2.2288777605	0.9580789440	0.9580789440	1.8455944172	1.7857018251	1.4006011007	0.0662292149	0.1111443888	0.1495936761	0.1111443888
2.7566001411	3.2702842116	2.7566001411	1.8310573292	1.8310573292	1.8455944172	1.7857018251	1.4006011007	0.0662292149	0.1378300071	0.1635142106	0.1378300071
2.5012984650	2.4812107105	1.4812107105	1.4180257539	1.4180257539	1.8455944172	1.7857018251	1.4006011007	0.0662292149	0.1250649233	0.1240635085	0.1250649233
1.5440582811	2.1429817492	1.5440582811	0.6839274971	0.6839274971	1.8455944172	1.7857018251	1.4006011007	0.0662292149	0.0772027914	0.1071490875	0.0772027914
1.9605832336	2.3607261792	1.9605832336	1.3015940024	1.3015940024	1.8455944172	1.7857018251	1.4006011007	0.0662292149	0.0980291617	0.1180363090	0.0980291617
1.7618002581	1.7454157982	0.9807211159	1.0111208685	1.0111208685	1.8455944172	1.7857018251	1.4006011007	0.0662292149	0.0880900129	0.0872707899	0.0880900129
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000	0.0000000000
2.2698518957	3.0767846527	2.2698518957	0.9845736133	0.9845736133	1.8405219529	1.7766027126	1.4006011007	0.0662292149	0.1134925948	0.1538392326	0.1134925948
2.8306570269	3.9697945603	2.8306570269	1.8798771172	1.8798771172	1.8405219529	1.7766027126	1.4006011007	0.0662292149	0.1415328513	0.1694889728	0.1415328513
2.5628466990	2.5411639970	1.5628466990	1.4569334133	1.4569334133	1.8405219529	1.7766027126	1.4006011007	0.0662292149	0.1270581998	0.1270581998	0.1270581998
1.5947804714	2.2397287447	1.5947804714	0.7123310453	0.7123310453	1.8405219529	1.7766027126	1.4006011007	0.0662292149	0.0797390236	0.1116988647	0.0797390236
2.0400133971	2.4671416901	2.0400133971	1.3539740538	1.3539740538	1.8405219529	1.7766027126	1.4006011007	0.0662292149	0.1020066699	0.1233570845	0.1020066699
1.8279492871	1.8098446544	1.0250567107	1.0528332903	1.0							

Textausgabe des Programms

# Schritt 0				# Schritt 5:				# Schritt 10:				# Schritt 15:			
=====				=====				=====				=====			
GEN 1:				GEN 1:				GEN 1:				GEN 1:			
0	0	0		8.6	9.44	8.6		14.51	17.38	14.51		18.7	23.86	18.7	
0	0	0		9.16	9.72	9.16		16.47	18.38	16.47		22.25	25.69	22.25	
0	0	0		8.88	8.88	8.88		15.51	15.47	15.51		20.53	20.43	20.53	
GEN 2:				GEN 2:				GEN 2:				GEN 2:			
0	0	0		4.68	5.15	4.68		8.66	10.45	8.66		12.04	15.63	12.04	
0	0	0		4.99	5.3	4.99		9.88	11.07	9.88		14.51	16.91	14.51	
0	0	0		4.84	4.84	4.84		9.28	9.26	9.28		13.31	13.24	13.31	
# Schritt 1:				# Schritt 6:				# Schritt 11:				# Schritt 16:			
=====				=====				=====				=====			
GEN 1:				GEN 1:				GEN 1:				GEN 1:			
2	2	2		9.96	11.15	9.96		15.47	18.79	15.47		19.38	24.99	19.38	
2	2	2		10.76	11.55	10.76		17.74	19.95	17.74		23.25	26.99	23.25	
2	2	2		10.36	10.35	10.36		16.63	16.58	16.63		21.38	21.26	21.38	
GEN 2:				GEN 2:				GEN 2:				GEN 2:			
1	1	1		5.53	6.2	5.53		9.38	11.5	9.38		12.65	16.63	12.65	
1	1	1		5.99	6.43	5.99		10.83	12.25	10.83		15.39	18.05	15.39	
1	1	1		5.76	5.76	5.76		10.12	10.09	10.12		14.07	13.98	14.07	
# Schritt 2:				# Schritt 7:				# Schritt 12:				# Schritt 17:			
=====				=====				=====				=====			
GEN 1:				GEN 1:				GEN 1:				GEN 1:			
3.84	3.94	3.84		11.22	12.8	11.22		16.36	20.14	16.36		20.03	26.07	20.03	
3.91	3.97	3.91		12.29	13.34	12.29		18.95	21.47	18.95		24.2	28.23	24.2	
3.88	3.88	3.88		11.76	11.75	11.76		17.68	17.62	17.68		22.19	22.05	22.19	
GEN 2:				GEN 2:				GEN 2:				GEN 2:			
1.96	2.01	1.96		6.35	7.27	6.35		10.08	12.55	10.08		13.24	17.62	13.24	
2	2.03	2		6.97	7.58	6.97		11.77	13.42	11.77		16.27	19.19	16.27	
1.98	1.98	1.98		6.67	6.66	6.67		10.94	10.9	10.94		14.81	14.71	14.81	
# Schritt 3:				# Schritt 8:				# Schritt 13:				# Schritt 18:			
=====				=====				=====				=====			
GEN 1:				GEN 1:				GEN 1:				GEN 1:			
5.55	5.83	5.55		12.4	14.39	12.4		17.19	21.44	17.19		20.63	27.1	20.63	
5.74	5.93	5.74		13.75	15.07	13.75		20.1	22.93	20.1		25.1	29.42	25.1	
5.64	5.64	5.64		13.08	13.06	13.08		18.69	18.61	18.69		22.95	22.8	22.95	
GEN 2:				GEN 2:				GEN 2:				GEN 2:			
2.9	3.05	2.9		7.15	8.33	7.15		10.75	13.59	10.75		13.82	18.6	13.82	
3	3.1	3		7.95	8.74	7.95		12.69	14.59	12.69		17.12	20.31	17.12	
2.95	2.95	2.95		7.56	7.54	7.56		11.75	11.7	11.75		15.53	15.42	15.53	
# Schritt 4:				# Schritt 9:				# Schritt 14:				# Schritt 19:			
=====				=====				=====				=====			
GEN 1:				GEN 1:				GEN 1:				GEN 1:			
7.13	7.66	7.13		13.49	15.92	13.49		17.97	22.68	17.97		21.19	28.09	21.19	
7.49	7.84	7.49		15.14	16.75	15.14		21.2	24.34	21.2		25.96	30.56	25.96	
7.31	7.31	7.31		14.33	14.3	14.33		19.63	19.54	19.63		23.67	23.51	23.67	
GEN 2:				GEN 2:				GEN 2:				GEN 2:			
3.81	4.09	3.81		7.92	9.39	7.92		11.4	14.61	11.4		14.37	19.55	14.37	
4	4.19	4		8.92	9.9	8.92		13.61	15.75	13.61		17.96	21.43	17.96	
3.9	3.9	3.9		8.43	8.41	8.43		12.54	12.48	12.54		16.24	16.11	16.24	

Literaturverzeichnis

- [1] *Shoot apical meristem maintenance: the art of a dynamic balance*, 2003
Cristel C. Carles and Jennifer C. Fletcher
- [2] *Modeling the organization of the Wuschel expression domain in the shoot apical meristem*, 2005
H. Jönsson, M. Heisler, G. Venugopala Reddy, V. Agrawal, V. Gor, B. E. Shapiro, E. Mjolsness and E. M. Meyerowitz
- [3] *In vivo analysis of cell division, cell growth, and differentiation at the shoot apical meristem in Arabidopsis*, 2004
Grandjean, O. and Vernoux, T. and Laufs, P. and Belcram, K. and Mizukami, Y. and Traas, J.
- [4] *Cellular Automaton Modeling and Biological Pattern Formation. Characterization, Applications, and Analysis*. A. Deutsch and S. Dorfman. Springer - Birkhuser, Basel, Boston, 2004.
- [5] *Ingeneue: a versatile tool for reconstituting genetic networks, with examples from the segment polarity network*. E. Meir, E. M. Munro, G. M. Odell, and G. Von Dassow. J Exp Zool, 294(3):21651, 2002.
- [6] *Einfuehrung in Evolutionaere Algorithmen*. V. Nissen. Vieweg, Braunschweig, Wiesbaden, Germany, 1997.