

**„Die Beziehung zwischen genetischem Polymorphismus
von Populationen und Umweltvariabilität:
Anwendung der Fitness-Set Theorie“**

D i s s e r t a t i o n

Z u r E r l a n g u n g d e s G r a d e s

„ D o k t o r

d e r N a t u r w i s s e n s c h a f t e n “

am Fachbereich Biologie
der Johannes Gutenberg-Universität
in Mainz

Oliver Lange
Geb. in Mainz

Mainz, 2002

Inhaltsverzeichnis

Abbildungsverzeichnis	6
Tabellenverzeichnis	7
1 EINFÜHRUNG	8
1.1 Bedeutung von Vielfalt	8
1.2 Mechanismen, die genetische Vielfalt aufrechterhalten	9
2 DAS QUANTITATIVE FITNESS-SET MODELL (QFM)	12
2.1 Einleitung	12
2.2 Material und Methode	13
2.2.1 Voraussetzungen und Begriffsdefinitionen	14
2.2.2 Herleitung des QFM	15
2.2.3 Das QFM Analyseprogramm	17
2.2.4 Anwendbarkeitsüberprüfung des Analyseprogramms	20
2.3 Ergebnisse	22
2.4 Diskussion	23
3 SENSITIVITÄTSANALYSE DES QFM ANALYSEPROGRAMMS	25
3.1 Einleitung	25
3.2 Material und Methode	25
3.3 Ergebnisse	28
3.3.1 Variation der Genotypanzahl	28
3.3.2 Variation der Populationsgröße	29
3.3.3 Stichprobeneffekt	31
3.4 Diskussion	32

4	FITNESSBÄNDER SCHÜTZEN VOR GENETISCHER VERARMUNG KLEINER POPULATIONEN	34
4.1	Einleitung	34
4.2	Material und Methode	34
4.3	Ergebnisse	36
4.4	Diskussion	38
5	ASYNCHRONE REPRODUKTION ERMÖGLICHT FEINKÖRNIGE INTERPRETATION GROBKÖRNIGER UMWELTVARIATION	41
5.1	Einleitung	41
5.2	Material und Methode	41
5.2.1	Herleitung des QFM bei asynchroner Reproduktion	41
5.2.2	Erhalt genetischer Vielfalt unter fein- und grobkörniger Sicht	43
5.3	Ergebnisse	44
5.4	Diskussion	46
6	AUSBLICK	49
6.1	Der Modellorganismus	49
6.2	Umweltschwankungen	50
7	ZUSAMMENFASSUNG	52
8	LITERATURVERZEICHNIS	53
9	ANHANG	56
9.1	Detaillierte Herleitung der Fitness-Set-Theorie nach Levins 1968	56
9.2	Motivation der Quantitativen Fitness-Set-Theorie	58
9.2.1	Notationsfehler im Originaltext	58

9.2.2	Extremfallbetrachtung	58
9.2.3	Nicht explizite Definition von Fitness	58
9.3	Beweis der Konstanz der Gesamtfitnessnormierung	58
9.4	Rückführung des QFM auf die Ergebnisse von Levins (1968)	60
9.4.1	Für grobkörnige Umwelt	60
9.4.2	Für feinkörnige Umwelt	61
9.5	Herleitung der Reihenfolgefunktion	63
9.6	Funktionen zur grafischen Darstellung der Fitness-Isoklinen in „R“	64
9.7	Module und Kernprozeduren aus dem QFM-Analyseprogramm	66
9.7.1	Datenstrukturen	66
9.7.2	Implementation des QFM	68
9.7.3	Ermitteln der Parameterkombination, „Fitting“	69
9.7.4	Routinen zur Auswertung (Anpassungsfehler E_{obs} und Schätzfehler E_{real})	71
9.7.5	Darstellung der Umweltvariation und Körnigkeit	71
9.7.6	Reduktion auf die Hauptlösungsmenge	74
9.7.7	Das Simulationsmodell zur Generierung von Testdaten	77

Abbildungsverzeichnis

Abb. 1: Fitnessisoklinenverläufe in feinkörniger und grobkörniger Umweltvariation	15
Abb. 2: Fitness eines Genotyps in mittelkörniger Umwelt	16
Abb. 3: Hauptlösung und 5 Nebenlösungen für Parameterkombination	18
Abb. 4: Körnigkeiten von Umwelten	19
Abb. 5: Ablauf der Sensitivitätsanalyse	19
Abb. 6: Streuung und Bias führen zu „Ungenauigkeiten“	20
Abb. 7: Einflüsse der Parameterwerte auf die Fehlermaße ME_{real} und ME_{obs}	23
Abb. 8: Ablauf einer Generation des Drift-Selektions-Modells für 6 Genotypen	26
Abb. 9: Normierter Schätzfehler ME_{real} , Anpassungsfehler ME_{obs} und die Nullverteilung der Mittelwerte in Abhängigkeit der Anzahl der Genotypen	28
Abb. 10: Zufallsverteilung (schwarz), sowie normierter Schätzfehler ME_{real} (weiß) und Anpassungsfehler ME_{obs} (schraffiert) in Abhängigkeit der Populationsgröße N	30
Abb. 11: Zuordnung eines Unschärfewertes zu einer Populationsgröße	36
Abb. 12: Dichteabhängigkeit der Wiederfunde im 0%-Unschärfebereich unter verschiedenen Populationsgrößen	37
Abb. 13: Die Häufigkeitsverteilung der Dichten im 1% Unschärfebereich in der Hauptlösungsmenge	38
Abb. 14: Polymorphieverlust in der Abhängigkeit der Umweltkörnigkeit	44
Abb. 15: Wechsel der Körnigkeit	45
Abb. 16: Erhöhung der Schwankungsamplitude (z.B. wärmere Sommer) führt zu konkaven Fitness-Sets und tendenziell zur disruptiven Selektion	51

Tabellenverzeichnis

Tabelle 1: Erläuterung der Parameter und Abkürzungen des Quantitativen Fitness-Set-Models.	14
Tabelle 2: Vergleich von Schätzfehler und Anpassungsfehler und ihre Interpretation. ME_{real} beschreibt die Qualität des Analyseprogramms im Erkennen der richtigen Parameterkombination, ME_{obs} beschreibt, wie gut die geschätzte Parameterkombination die simulierten Gesamtfitnesswerte erklärt.	21
Tabelle 3: Absolute Größe der Schätzfehler über alle 2056 Parameterkombination. Beide Fehlermaße sind als sehr klein zu werten. Damit ist die Fähigkeit des Analyseprogramms hoch einzuschätzen.	22
Tabelle 4: Einfluss der Parameterwerte auf den Schätzfehler des Modells. Angegeben sind Spearman's Rangkorrelationskoeffizienten. Werte kleiner $ 0.5 $ werden als „schwach“ gewertet.	23
Tabelle 5: Steigende Zahl der Hauptlösungen in Abhängigkeit vom Parameter Generationen“.	24
Tabelle 6: Werte der Sensitivitätsanalyse. Die Populationsgrößen sollen starke, mittlere und wenig genetische Drift erzeugen. Die Genotypen sind symmetrisch auf beide Umwelten verteilt, sodass sich die Fitness-Sets (f1,f2) als Gitter über die beiden Umwelten legen. Der Stichprobeneffekt wird exemplarisch an der Stichprobengröße „100“ demonstriert.	27
Tabelle 7: Einfluss der Anzahl der Genotypen auf die Fehlerverteilungen. Angegeben ist der Anteil der Elemente [%], bei den die Nullhypothese zutrifft.	29
Tabelle 8: Einfluss der Parameterwerte auf den Schätzfehler ME_{real} und Anpassungsfehler ME_{obs} des Modells. Spearman's Rangkorrelationen kleiner $ 0.5 $ werden als „schwach“ gewertet.	31
Tabelle 9: Experimentelle Fitness-Daten für konstante und oszillierende Umwelten (Kassen und Bell 1998) der Grünalge <i>Chlamydomonas reinhardtii</i> . Die Genotypen wurden in ihren Umwelten isoliert und dann den anderen Umwelten ausgesetzt.	47
Tabelle 10: Beispiele der Motivabfolge für $\tau_1=5$, $\tau_2=7$.	63

1 Einführung

Vielfalt, Biodiversität, genetische Unterschiede sind Schlagwörter des alltäglichen Gebrauchs geworden. Der Vielfalt wird in den Medien viel Raum eingeräumt – auf eine oftmals sehr beschreibende Art und Weise, die fast an die Erlebnisberichte von Forschungsreisenden des letzten Jahrhunderts erinnert. Schwieriger wird die Zuweisung kultureller, wissenschaftlicher oder wirtschaftlicher Bedeutung zu dieser Vielfalt, sodass diese sich in den heutigen argumentativen Kanon gesellschaftsrelevanter Ziele einfügt. Dabei wird der heutige Zustand und die jetzige Lebenserfahrung als Maßstab angesetzt – oft unter Missachtung historischer Wahrheit und mit dem Ziel des konservierten heutigen Zustandes. Über die Prozesse und die sie verursachenden Prinzipien hört man nur wenig, obwohl die Betrachtung der so gelobten Vielfalt nur ein einziger Schnappschuss aus einer ganzen Serie von Zuständen ist.

1.1 Bedeutung von Vielfalt

In der Medizin bedeutet Vielfalt bakterieller oder viraler Krankheitserreger die Notwendigkeit verschiedener Heilungsmethoden. Die genetische Vielfalt innerhalb einer Art eines Krankheitserregers bedeutet auch vielfältige Abwehrmechanismen gegenüber Medikamenten (Antibiotikaresistenzen, unwirksame Schutzimpfungen). Den volkswirtschaftlichen Schaden von „Grippe“ schätzt die Pharmafirma Roche 1999 in den USA mit durchschnittlich 14 Mrd. Dollar pro Jahr (Roche 1999).

In der Lebensmittelindustrie werden Reinzuchten von Mikroorganismen zur Produktion von gleichbleibender Qualität bei Wein, Milchprodukten oder auch Dauerwurst benötigt. Die Pharmaindustrie verwendet Mikroorganismen in biotechnischen Verfahren z.B. bei der Erzeugung von Insulin oder zur Umsetzung chemischer Produkte. Auch hier sind Prozesse gefürchtet, die Vielfalt erzeugen oder bewahren.

Andererseits werden auch viele Hoffnungen an noch unerkannte Biodiversität geknüpft: Von Mikroorganismen erhofft man sich u.a. noch unbekannte Antibiotika oder die Erzeugung von Polymeren aus nachwachsenden Rohstoffen. Die Bionik wirbt mit der Untersuchung der natürlichen Vielfalt zum Erkennen technischer Grundprinzipien. Auch erhofft man sich unerkannte Stoffgruppen (Arzneimittel, Genussmittel) aus seltenen, bzw. bisher unbekanntem Organismen (Meeresschwämme, tropische Spezies). Aus Sicht des Evolutionsbiologen ist klar, dass jede existierende Art, ob Mikroorganismus, Tier oder Pflanze, die derzeitige

bestmögliche Lösung eines zur jetzigen Zeit an diesem Ort existierenden Problems ist.

Die Vielfalt der Arten, aber auch die genetische Vielfalt innerhalb von Arten kann als Spiegel der vergangenen Umwelteinflüsse angesehen werden (Allendorf und Leary 1987). Sie gilt somit auch als „Rückversicherung“ für zukünftige Umweltänderungen (z.B. Global Change). Das gesellschaftliche Ziel der Arterhaltung führt damit auch zur Frage, welche wichtige Einflussfaktoren z.B. zum Erhalt genetischer Vielfalt in einem Auswilderungsprojekt sind.

Vielfalt bildet auch die Grundlage für Evolution. Sie ist die Voraussetzung für den Auswahlprozess. Diese Auswahl kann nun gerichtet und somit Selektion im eigentlichen Sinne sein, oder aber zufällig erfolgen (z.B. Gründereffekt). In beiden Fällen führt „Auswahl“ letztendlich zum Verlust genetischer Vielfalt und kann zur Artbildung führen. Aufrechterhaltung genetischer Vielfalt ist also auch als Antispeziationsmechanismus zu werten. Insbesondere für vorwiegend klonal reproduzierende Organismen besteht starke Gefahr der endemischen Artbildung - Mechanismen für eine „Flucht vor der Speziationsfalle“ sind also zu erwarten.

1.2 Mechanismen, die genetische Vielfalt aufrechterhalten

Vielfalt auf regionaler und großräumiger Ebene haben Wright und Fisher mit ihren populationsgenetischen Theorien erklären können. Sie zeigten, dass in einer einzelnen Population mit endlicher Populationsgröße in unendlicher Zeit immer nur ein Allel fixiert sein wird. Nimmt man unendlich viele solcher endlichen Populationen und startet sie mit einer identischen Allelverteilung, so findet man jedes Allel mit der ursprünglichen Häufigkeit wieder. Eine geringe, aber identische Migration zwischen allen Populationen führt zu einem Gleichgewicht zwischen dem Fixationsprozess (Genetische Drift) und der Migration (Hartl und Clark 1989). Dabei bleibt die Allelvielfalt in jeder Population erhalten. Für haploide Organismen bleibt dabei auch die gesamte genetische Vielfalt erhalten, in di- und polyploiden Arten verliert sich der Anteil an genetischer Vielfalt, der durch die Kombination von Allelen in heterozygoten Individuen auftritt (Fixierung von Allelen).

Vollständig isolierte und fixierte Populationen (Inselpopulationen) sind vor Ausleseprozessen geschützt, denn ohne Vielfalt gibt es keine Auswahl. Fixierte Inselpopulationen entstehen nicht nur als Ergebnis einer Zufallsauswahl (Buri 1956), sondern auch als Resultat von Selektion (Chappell et al. 1988). Die eventuell bestehende Übergangszone von einer zur

anderen Population ist dann genetisch als Cline sichtbar. Vielfalt bleibt, großräumig gesehen, erhalten.

Neben derartigen großräumigen, bzw. regionalen Mechanismen, können auch kleinräumige Strukturierungen zum Erhalt von Vielfalt führen. Trade-Offs bieten die Möglichkeit, verschiedenartige, aber als Gesamtfitness gleichgute Lösungen eines Problems koexistieren zu lassen. Einnischung einer Population in verschiedene räumlichen Zonen durch unterschiedliches Verhalten konnte z.B. bei der Vertikalwanderung verschiedener Daphnia-Arten beobachtet werden (Müller 1993).

Jenseits räumlicher sind auch zeitliche Mechanismen vorhanden. Eine sehr elegante Möglichkeit zum Erhalt von Vielfalt bietet sexuelle Reproduktion. (1) In dominanten Erbgängen liegen seltene, rezessive Allele hauptsächlich heterozygot vor und sind damit vor Selektion geschützt. (2) In intermediären Erbgängen können die Heterozygoten bevorteilt oder benachteiligt sein. In beiden Fällen kann es zu einem stabilen Gleichgewicht kommen, wobei die Gesamtfitness der Population unter der maximalen Fitness eines Genotyps liegt – die Population trägt eine „genetische Last“ (Dobzhanski 1957), unter der die Vielfalt erhalten bleibt. (3) Betrachtet man mehr als nur einen Locus, so wirkt Selektion auf den Phänotyp im Sinne der Gesamtheit aller Loci (Franklin und Lewontin 1970). Der zu erwartende Prozentsatz von Kombinationen vorteilhafter und benachteiligt ausgeprägter Loci maskieren das benachteiligte Allel vor der direkten Einwirkung von Selektion; dies führt zum Teil erst zu einer maximalen Populationsfitness (Lewontin 1971). Verstärkt tritt dies bei nicht-segregierenden Loci auf (linkage). Hier müssen benachteiligte Allele auf einem Locus, die mit vorteilhaften Allelen an einem anderen Locus gekoppelt sind, erst durch crossing-over entkoppelt werden (Gimelfarb 1998; Barton 2000). Vielfalt wird unter sexuellen Bedingungen längere Zeit aufrechterhalten als unter klonalen Bedingungen.

Die genetischen Grundlagen führen auch bei asexueller Fortpflanzung zu einem zeitlichen Mechanismus: Bei haploiden Chromosomensätzen prägt sich ein Locus nur dann direkt aus, wenn die codierende Sequenz einfach vorliegt. Wenn das Gen verdoppelt wurde, können beide Genorte auch unterschiedliche Allele tragen (Du et al. 2000). Dieser Zustand existiert auch bei di- und polyploiden Chromosomensätzen (z.B. Hämoglobin-Genfamilie, rDNA). Bei diesen Möglichkeiten können verschiedene Allele von gekoppelten Genorten ausgeprägt werden, sodass ein benachteiligtes Allel durch die vorteilhaften maskiert wird.

Bei verschiedenen Lebensabläufen ist auch unterschiedlicher Erhalt von Vielfalt zu erwarten. In extremen Fällen geht der Genotyp in eine Form der Diapause und kann mehrere Generationen darin verweilen. Von *Daphnia* ist bekannt, dass aus über 40 Jahre alte Proben aus Seesedimenten noch Individuen schlüpfen (Hairston et al. 2001). Moose können nach mehreren Jahrzehnten in Herbarien weiterwachsen (Aichele und Schwegler 1993), Bakteriensporen können extreme Temperaturen und lange Jahre überleben (Winkle 2002), auch Insekten verzögern das Schlüpfen ggf. um mehrere Jahrzehnte (Powell 2001). In allen Fällen können „alte“ Genotypen neu in der Population auftauchen. In der Botanik ist dafür der Begriff der Bodensamenbank etabliert.

Paarungsbevorzugung seltener Genotypen führen zum Erhalt gefährdeter Allele, da sie eine höhere Fitness aufgrund ihrer geringeren Frequenz erhalten. Auch Räuber produzieren derartige Mechanismen, sofern sie bekannte und häufige Beutetypen bevorzugen, und anders aussehende Genotypen seltener fressen. Im ersten Fall ist für den seltenen Genotyp die Anzahl an Nachkommen erhöht, im zweiten Fall ist seine Mortalität erniedrigt. Eine schöne Zusammenstellung und detailliertere Übersicht der Mechanismen über genetische Polymorphie findet sich im Buch „Ökologische Genetik“ (Stern und Tigerstedt 1974).

In allen bisherigen Beispielen wurde die jeweilige Umwelt als möglicherweise unterschiedlich, aber in sich als konstant angenommen. Dass korrelierte und unkorrelierte zufällige Fluktuationen die genetische Vielfalt erhalten können, lässt sich mathematisch herleiten (Cook und Hartl 1974; Hartl und Cook 1974). Dem entsprechen auch experimentelle Ergebnisse für die „Intermediate Disturbance Hypothesis“ (Connell 1978), die hohe Vielfalt bei „mittleren“ Störungsgraden postuliert (Weider 1992). Für zufällige Umweltschwankung sind schon für einen Locus eine beliebige Anzahl an Allelen möglich (Gillespie 1977).

Neben zufällig fluktuierenden Umweltänderungen existieren auch regelmäßige, oszillierende Umweltänderungen in verschiedenen zeitlichen Abstufungen: kurze Schwankungsperioden entstehen durch die Gezeiten oder Tages-Zyklen, längere sind die Jahreszeiten, der 11-jährige Zyklus der Sonnenaktivität bis hin zu den 41 000-Jahres-Milankovitch-Zyklen (Willis et al. 1999), auf die die eiszeitliche Vergletscherungen zurückgeführt werden. In diesen oszillierenden Umwelten existieren Lebewesen mit unterschiedlichen Generationsdauern, sodass die Begriffe „kurz“ oder „lang“ relativ zur betrachteten Art gesehen werden müssen. Zur Analyse und Untersuchung der zu erwartenden genetischen Muster entstand das Quantitative Fitness-Set Modell auf der Basis der Fitness-Set-Theorie (Levins 1968).

2 Das Quantitative Fitness-Set Modell (QFM)

2.1 Einleitung

Die Fitness-Set-Theorie (FT) erklärt genetische Vielfalt durch zeitlich aufeinanderfolgende Umwelten, in denen Arten unterschiedliche Fitness aufweisen (Levins 1968). Unter derartiger oszillierender Selektion existieren verschiedene Überlebensstrategien: (1) An eine der beiden Umwelten sehr gut angepasst sein und die andere gerade so überleben, (2) für beide Umwelten halbwegs angepasst sein, (3) eine Mischung aus beiden Strategien. Sofern diese Strategien genetisch bedingt sind, ergeben sich genetisch analysierbare Muster. Diese werden durch die FT beschrieben. Die Gesamtfitness eines Genotyps ergibt sich aus den Zeitdauern der Umwelten zueinander. Aus dem stabilisierten genetischen Muster lassen sich nun auch die formenden Kräfte der Vielfalt quantifizieren. Es lassen sich regulierende von nichtregulierenden Faktoren unterscheiden, und man kann damit bei der Aufrechterhaltung von Biodiversität helfen. Die FT bietet einen Ansatz, freie ökologische Nischen zuerkennen, z.B. um die Anfälligkeit eines Ökosystems für Invasionen abschätzen zu können. Die FT belegt oszillierende Selektion als einen Faktor, der genetische Diversität innerhalb von Arten und Artengemeinschaften stabilisiert. Sie ist damit auch für die Untersuchung von Evolution und Speziationsprozessen interessant.

Eine Schattenseite der Fitness-Set-Theorie (Levins 1968) liegt in der Herleitung als qualitative, erklärende Theorie. Folgende Kritikpunkte lassen sich formulieren: (1) Mathematisch sind die Extremfälle „grobkörnige“ und „feinkörnige Umwelt“ formuliert. Der allgemeinere Ansatz ist auf mathematisch-kontinuierlicher Ebene schwierig, wenn nicht unmöglich zu führen. (2) Als analytisches Werkzeug wird eine graphische Interpretation der Extreme eingeführt und es werden Wirkungsweisen und Prinzipien erklärt. Eine quantitative Erfassung der Modellparameter und der analytische Einsatz muss unterbleiben. (3) Es existiert kein quantitatives Maß für die „Körnigkeit“ der Umweltschwankung.

Zunächst stelle ich die Weiterentwicklung der Fitness-Set-Theorie zum Quantitativen Fitness-Set Modell (QFM) mit ihren Voraussetzungen vor. Hierbei wird die Gesamtfitness eines Genotyps im Zusammenspiel von zweistufiger Umweltszillation mit dem Lebenszyklus der Art mathematisch abgebildet. Die Herleitung des QFM beruht auf parthenogenetischen Organismen, und daher greift hier der Begriff „Fitness eines Genotyps“. Diese Sichtweise ist analog dem Verständnis von „Fitness von Allelen“ (Cook und Hartl 1974), lässt sich aber

auch durch die Vorstellung „Konkurrenzfähigkeit einer Art“ ersetzen. Analysen von Artengemeinschaften und ökologischen Nischen sind möglich, denn in allen drei Sichtweisen findet keine (oder nur selten) eine Vermischung der genetischen Basis (sexueller Zyklus parthenogenetischer Organismen, Crossing-Over, Hybridisierungen) statt.

Basierend auf dem QFM habe ich ein Computerprogramm als analytisches Werkzeug entwickelt. Zur Analyse der oszillierenden Umwelt benötigt man zu jedem Genotyp die Gesamtfitness, sowie die Einzelfitness in beiden konstanten Umwelten. Aufgrund der zu untersuchenden Parameterbereiche und Anpassungsgüte gibt das Programm die am besten passenden Parameterkombinationen aus.

Die generelle Einsatzfähigkeit wird an einer Modellumwelt belegt. Aus vorgegebenen Parameterkombinationen und Einzelfitnesswerten werden per Simulation Gesamtfitnesswerte erzeugt, aus denen das Analyseprogramm die ursprüngliche Parameterkombination wieder schätzt. Der Vergleich von ursprünglichen und geschätzten Parameterkombinationen ergibt ein Gütemaß für die Einsatzfähigkeit des Analyseprogramms.

2.2 Material und Methode

Die mathematische Herleitung des Modells basiert auf diskreten Parametern. Dies erleichterte die Umsetzung in ein computergestütztes Analysewerkzeug und ermöglichte die Reduktion vieler, analoger Parameterkombinationen nach dem Prinzip der einfachsten Lösung. Dadurch konnte die Anwendbarkeit auch für den gesamten Parameterraum bis zu einer Obergrenze von zehn Generationen überprüft werden. Die Auswertungen erfolgten mit Statistika (StatSoft 2001). Zum besseren Verständnis der Voraussetzungen, Herleitung und Grafiken sind die verwendeten Parameter und Variablen als Tabelle vorangestellt.

Tabelle 1 Erläuterung der Parameter und Abkürzungen des Quantitativen Fitness-Set-Modells. * Der Vergleich von τ_1 , τ_2 , l und x zueinander ist gültig, da alle in dem Zeitmaß Generationen „ $1/\tau_1$ “ bemessen sind.

Bezeichner	Bedeutung	Definitionsbereich	Einheit
n	Anzahl der Generationen	$n \in \mathbb{N} / \{0\}$	Generationen
τ_1	Zeitmaße pro Generation	$\tau_1 \in \mathbb{N} / \{0\}$	Zeitmaß (Anteile an Generation)*
τ_2	Dauer der Umweltperiode	$\tau_2 \in \mathbb{N} / \{0\}$	Zeitmaß (Anteile an Generation)*
λ	Dauer der Umwelt 1	$\lambda \in \{1, \dots, \tau_2 - 1\}$	Zeitmaß (Anteile an Generation)*
Messwerte			
f_1	Fitness in Umwelt 1	$f_1 \in [0..1]$	Relative normierte Fitness
f_2	Fitness in Umwelt 2	$f_2 \in [0..1]$	Relative normierte Fitness
Hilfsparameter in der Formulierung des Modells			
x	Zeitindex der aktuellen Fitness	$x \in \{t, \dots, t + \tau_1 - 1\}$	Zeitmaß (Anteile an Generation)*
i	aktuelle Generation	$i \in \{1, \dots, n\}$	Generation

2.2.1 Voraussetzungen und Begriffsdefinitionen

(1) Voraussetzung der Genotypkonstanz: In einer konstanten Umwelt hat ein Genotyp eine konstante Fitness. Er ist damit konstant über die Zeit, es finden im Beobachtungszeitraum weder a, plastische Veränderungen (z. B. räuberinduzierte Verteidigungsmechanismen, maternale Effekte) noch b, genetische Änderungen statt (sexueller Zyklus klonaler Organismen, Rekombination von Allelen, Hybridisierungen von Arten, Mutationen), noch darf c, kryptische, fitnessbeeinflussende Diversität innerhalb des (z.B. experimentell definierten) Genotyps existieren.

(2) Definition der relativen normierten Fitness: Fitness ist derjenige Faktor, der die Frequenzverschiebung eines Genotyps zwischen zwei Zeitpunkten beschreibt. Fitness ist damit analog zur Populationsgenetik definiert (Hartl und Clark 1989). Das QFM benötigt zudem eine normierte Fitness im Intervall $[0,1]$, die relativ auf den fittesten Genotyp bezogen ist.

(3) Definition der Umweltszillation: Eine Umweltszillation ist das regelmäßige Abwechseln zweier Umwelten nach unterschiedlich langen aber immer gleichbleibenden Zeitabständen.

(4) Definition des Fitness-Set: Ein Fitness-Set ist die Menge der Fitnesspaare der Genotypen, auf die eine spezifischen Umweltszillation einwirkt. Sie bilden in einem Fitness-Set-Diagramm die Stützstellen der eigentlich dreidimensionalen Gesamtfitnessfunktion.

(5) Definition der Fitnessisokline: Eine Fitnessisokline ist die Linie gleicher Gesamtfitness. Sie kann in einem Fitness-Set-Diagramm (s. Abb. 1) dargestellt werden. Ihre Form wird durch die Körnigkeit der Umwelt bestimmt: Sie sind linear bei feinkörniger Umwelt (Abb. 1A) und hyperbolisch bei grobkörniger Umwelt (Abb. 1B). Ihr Verlauf wird durch die Verhältnisse der jeweiligen Umweltdauern zueinander bestimmt (hier 1:1). Im Gleichgewichtszustand koexistieren nur Genotypen, die alle auf der maximalen Fitnessisokline liegen.

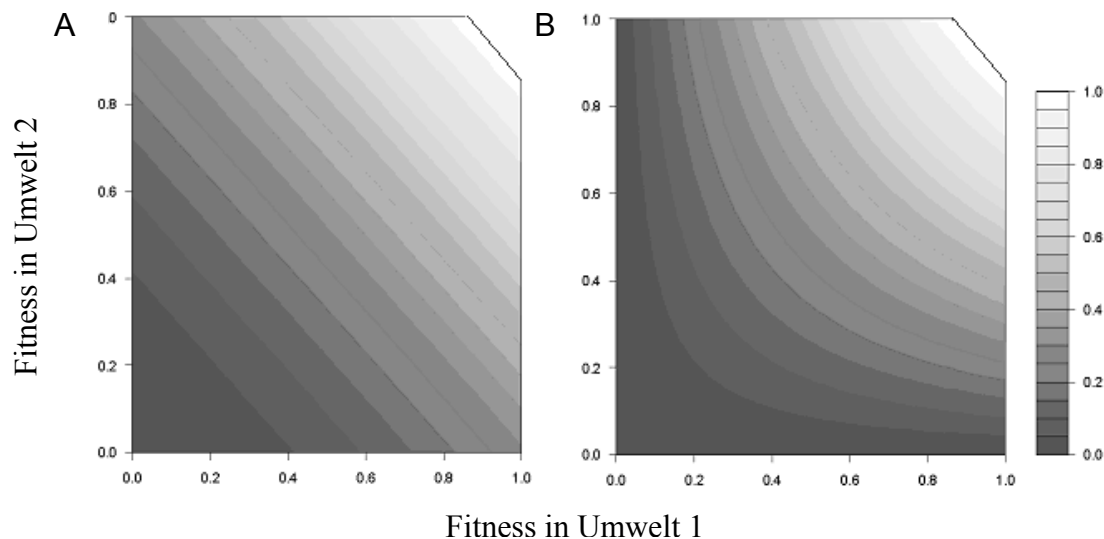


Abb. 1 Fitnessisoklinenverläufe in feinkörniger (A) und grobkörniger (B) Umweltvariation. Die Umwelten stehen im Verhältnis 1:1. Existiert kein Genotyp, der in beiden Umwelten maximal ist, so entsteht wie in der obigen Abbildung zwangsläufig „Ecken“ als Skalierungsartefakt. Parameterkombination für das QFM A: $(n, \tau_1, \tau_2, \lambda) = (2, 2, 2, 1)$, B: $(n, \tau_1, \tau_2, \lambda) = (2, 1, 2, 1)$.

(6) Voraussetzung der nichtüberlappenden Generationen: Als Einschränkung ist zunächst zu nennen, dass diese Formulierung nur für Arten mit nichtüberlappenden Generationen gilt. Asynchrone Entwicklungen können zu einer Mittelwertbildung der Fitness über Generationszeiträume führen; dieser Einfluss wird in einem späteren Kapitel untersucht.

2.2.2 Herleitung des QFM

Fitnessfunktion: Die Fitnessfunktion weist dem Genotyp zu jedem Zeitpunkt den Fitnesswert in der aktuellen Umwelt zu.

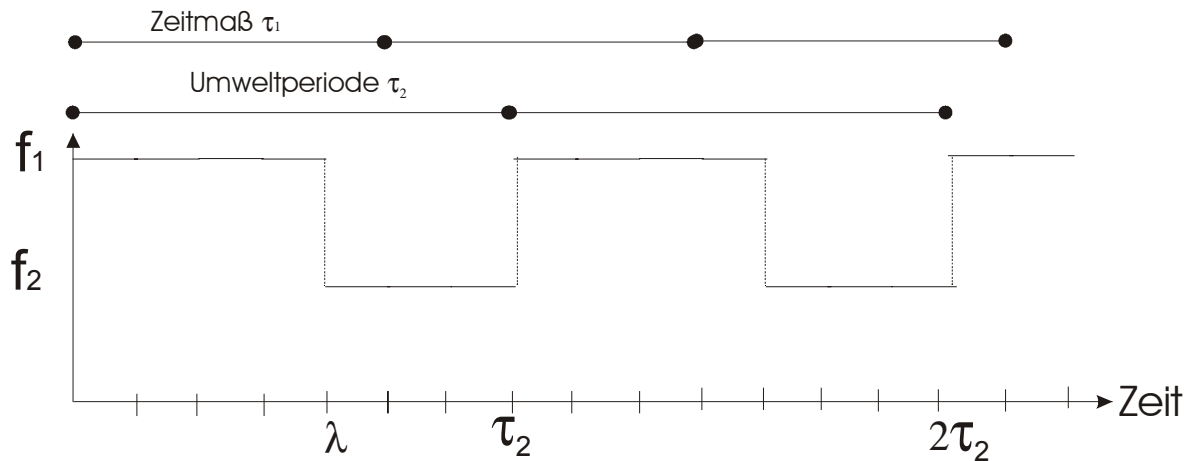


Abb. 2 Fitness eines Genotyps in mittelkörniger Umwelt. Die Umweltperiode τ_2 teilt sich im Verhältnis $\lambda: (\tau_2 - \lambda)$ auf. Zu jedem Zeitpunkt weist der Genotyp den entsprechenden aktuellen Fitnesswert (f_1 oder f_2) auf. Nach einer Generation wird die Fitness realisiert (z.B. Reproduktion, Sterben). Die realisierte Fitness entspricht dem Mittel der aufgetretenen aktuellen Fitnesswerte zwischen Beginn und Ende der Generation. Umweltperiode (τ_2 Zeiteinheiten) und Generationsdauer (τ_1 Zeiteinheiten) unterscheiden sich. Man sieht, dass beide Perioden erst nach einiger Zeit wieder gleichzeitig beginnen (s. a. Reihenfolgefunktion).

Die aktuelle Fitness w zum Zeitpunkt x lässt sich durch die Parameter $\tau_2, \lambda, f_1, f_2$ (s. Abb. 1, Tab. 1) mit folgendermaßen beschreiben:

$$w(x, \tau_2, \lambda, f_1, f_2) = \begin{cases} f_1 \forall (x \bmod \tau_2) \in [0.. \lambda[\\ f_2 \forall (x \bmod \tau_2) \in [\lambda.. \tau_2[\end{cases} \quad (2.1)$$

Durch den Einsatz des Modulo ist die Periodizität der Umweltvariation in das Intervall $[0..2\tau_2[$ projiziert. Wie man sieht, sind die Zeiten τ_1 und τ_2 von wesentlicher Interesse, denn nach dem Ablauf des Zeitmaßes für eine Generation τ_1 wird die Fitness realisiert (z.B. die Erzeugung von Nachkommenschaft, Sterben). Im Zusammenspiel mit der Umweltszillation τ_2 ergibt sich die realisierte Fitness W aus dem Mittelwert der auftretenden aktuellen Fitnesswerte im Zeitabschnitt t bis $t+\tau_1-1$:

$$W(t, \tau_1, \tau_2, \lambda, f_1, f_2) = \frac{1}{\tau_1} \sum_{x=t}^{t+\tau_1-1} w(x, \tau_2, \lambda, f_1, f_2) \quad (2.2)$$

Gemäß der Fitnessdefinition wirken die realisierten Fitnesswerte in aufeinanderfolgenden Generationen multiplikativ. So ergibt sich unter der Voraussetzung nicht-überlappender Generationen die Gesamtfitness aus:

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \prod_{i=1}^n W(r(i, \tau_1, \tau_2), \tau_1, \tau_2, \lambda, f_1, f_2) \quad (2.3)$$

Die noch unbekannte Reihenfolgefunktion $r(i, \tau_1, \tau_2)$ ergibt sich aus der Notwendigkeit, dass bei mittelkörniger Umwelt folgendes auffällt: Die Perioden von Umwelt und Fertilität kommen erst nach mehreren Wiederholungen zur Deckung. Somit ist auch die Abfolge der realisierten Fitnesswerte wichtig. Als Reihenfolgefunktion ergibt sich (s. Kapitel 9.5)

$$r(i, \tau_1, \tau_2) = [(i-1) \cdot \tau_1 \bmod \tau_2]; \quad r \in \{0, 1, \dots, \tau_2-1\} \quad (2.4)$$

Die Fitness-Set-Theorie ist damit auf eine explizite mathematische Formulierung zurückgeführt worden.

2.2.3 Das QFM Analyseprogramm

Die computergerechte Umsetzung des QFM in ein Analysewerkzeug wird durch die Diskretheit der Parameter vereinfacht: Die Elemente der Definitionsmenge sind alle möglichen, unterschiedlichen Parameterkombinationen. Damit ist diese abzählbar aber nicht unbeschränkt. Daher müssen oberen Schranken für n , τ_1 und τ_2 durch den Anwender vorgegeben werden. Die Definitionsmenge lässt sich dann mit den definierten Schranken vollständig erzeugen.

Erzeugt man alle möglichen Parameterkombinationen, so befinden sich zunächst redundante Elemente in der Definitionsmenge. Diese wurden nach dem Grundsatz der einfachsten Erklärung entfernt: Als Hauptlösung wurde diejenige Parameterkombination definiert, die das Ergebnis mit den kleinstmöglichen Parameterwerten erzeugt. Nebenlösungen erzeugen also für alle Genotypen aus den Einzelfitnesswerten f_1 und f_2 dieselbe Gesamtfitness. Zur Vereinfachung werden folgende Schreibweisen eingeführt:

(1) einen spezifischen Parameter $(n, \tau_1, \tau_2, \lambda)$: $(2, 2, 2, 1)$ für den gilt $n=2, \tau_1=2, \tau_2=2, \lambda=1$

(2) eine Hauptlösungsmenge $(n, \tau_1, \tau_2, \lambda) \leq (10, 10, 10, 9)$ ist die Menge der Parameter für die gilt: $n \in \{1, 2, \dots, 10\}, \tau_1 \in \{1, 2, \dots, 10\}, \tau_2 \in \{1, 2, \dots, 10\}$ und $\lambda \in \{1, 2, \dots, 9\}$

Typische Nebenlösungen sind Vielfache der Hauptlösungen (z.B. zur Hauptlösung $(n, \tau_1, \tau_2, \lambda) = (2, 2, 2, 1)$ die Nebenlösung $(4, 4, 4, 2)$). Andere, nicht triviale Nebenlösungen ergeben sich aus der Kommutativität der realisierten Fitness in der Gesamtfitness, denn

verschiedene Parameterkombinationen können dieselben Faktoren (in u.U. anderer Reihenfolge) erzeugen. Der Unterschied liegt dann in einem (noch) nicht erzeugten Umweltmotiv begründet (Abb. 3). Nebenlösungen erzeugen also identische Gesamtfitness und werden als solche erkannt und ausgeschlossen. Dies erfolgte hier experimentell durch Zuweisen eines Wertes (z.B. Blau=2,3; Rot=1,1) und Anwenden von Gleichung (2.3). Als Definitionsmenge des Programms wurden nur Hauptlösungen verwendet.

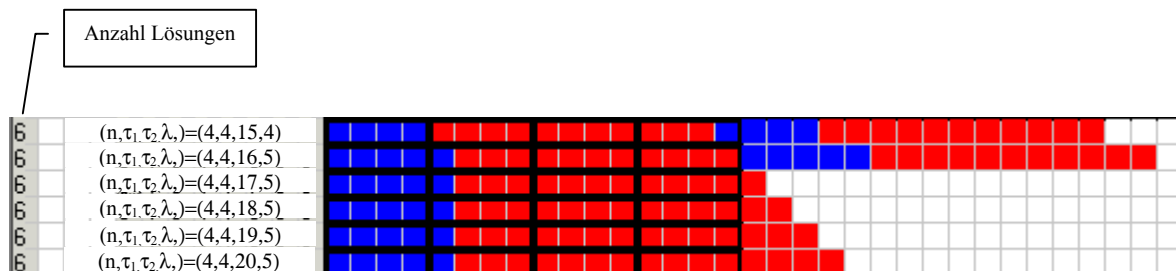


Abb. 3 Hauptlösung und 5 Nebenlösungen für die Parameterkombination $n=4$, $\tau_1 = 4$, $\tau_2 = 15$, $\lambda=4$. Die erste Ziffer bezeichnet die Anzahl der Parameterkombinationen mit identischer Gesamtfitnessfunktion. Die Zeiteinheiten in der Umwelt 1 wirkt, sind blau eingezeichnet, rot diejenigen in denen Umwelt 2 wirkt. Die vier schwarz umrandeten Zeilen entsprechen je einer der vier Generationen. Die realisierte Fitness ist der Mittelwert innerhalb einer Generation. Es existiert also für $(4,4,15,4)$ in dieser Reihenfolge die realisierten Fitnesses „Blau“, „Rot“, „Rot“ und „¼ blau ¾ rot“. Nur die Hauptlösung $(4,4,15,4)$ geht in die Definitionsmenge des Modells ein. Man erkennt, dass identische Lösungen durch die Kommutativität der Addition bei der Mittelwertbildung der realisierten Fitness ebenso entstehen, wie durch die Multiplikation der realisierten Fitnesswerte zur Gesamtfitness (Zeile 1 und 2). Auch Unterschiede im nicht beobachteten Zeitraum führen zur identischen Gesamtfitness (Zeilen 3,4,5,6).

Mit demselben experimentellen Ansatz lässt sich auch die Körnigkeit der erzeugten Umwelt beschreiben. Im QFM stellt sich eine grobkörnige Umwelt durch stark unterschiedliche realisierte Fitness dar. In Abb. 3 ist eine grobkörnige Umwelt dargestellt, denn die Faktoren sind fast nur entweder rot oder blau, aber es treten beide auf. Feinkörnige Umwelten sind gleichförmiger, d.h. a, „bunter“, b, nur blau oder c, nur rot (s. Abb. 4). Quantitativ lässt sich die Körnigkeit als die am Maximum normierten Varianz der realisierten Fitness W darstellen, also gemäß Gleichung (2.3)

$$\text{Körnigkeitsmaß} = \frac{\text{Var}(W)}{\text{Var}(\{\text{blau}, \text{rot}\})}, \quad W := \text{realisierten Fitnesswerte eines Parameters} \quad (2.5)$$

Feinkörnige Umweltschwankungen besitzen einen niedrigen, grobkörnige einen hohen Körnigkeitswert.

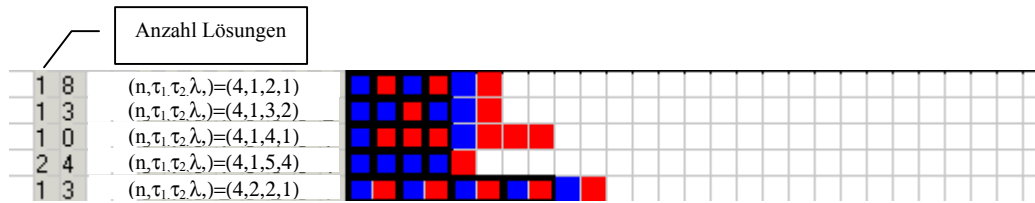


Abb. 4 Körnigkeiten von Umwelten: Grobkörnige Umwelten $(4, 1, 2, 1)$ mit Körnigkeitswert=1; $(4, 1, 3, 2)$ mit Körnigkeitswert=0.75). Feinkörnige Umwelten haben wenig Variation $(4, 1, 5, 4) \rightarrow$ (Körnigkeitswert=0) von völliger Einheitlichkeit (keine Umweltvariation, die Fitnessisoklinen stehen hier senkrecht auf eine der Umweltachsen), oder in der Form $(4, 2, 2, 1) \rightarrow$ Körnigkeitswert=0, in der dauerhaft eine „mittlere Fitness“ wirkt.

Die Optimierungsroutine ist aufgrund der Diskretheit des Parameterraumes denkbar einfach: Jeder mögliche Parametersatz wird verwendet, um die Fitnessfunktion zu parametrisieren. Nun kann man für jeden Genotyp des Fitness-Sets die Gesamtfitness berechnen (G_{est}). Als Güte der Anpassung gilt nun der absolute Abstand $|G_{est}(GT_1) - G_{obs}(GT_1)|$ für jeden Genotypen. Als Anpassungsfehler gilt das arithmetische Mittel über alle Abstände zwischen beobachtete Daten (E_{obs}) und der Fitnessfunktion. Der Parametersatz mit dem kleinsten Anpassungsfehler E_{obs} gilt als Lösung.

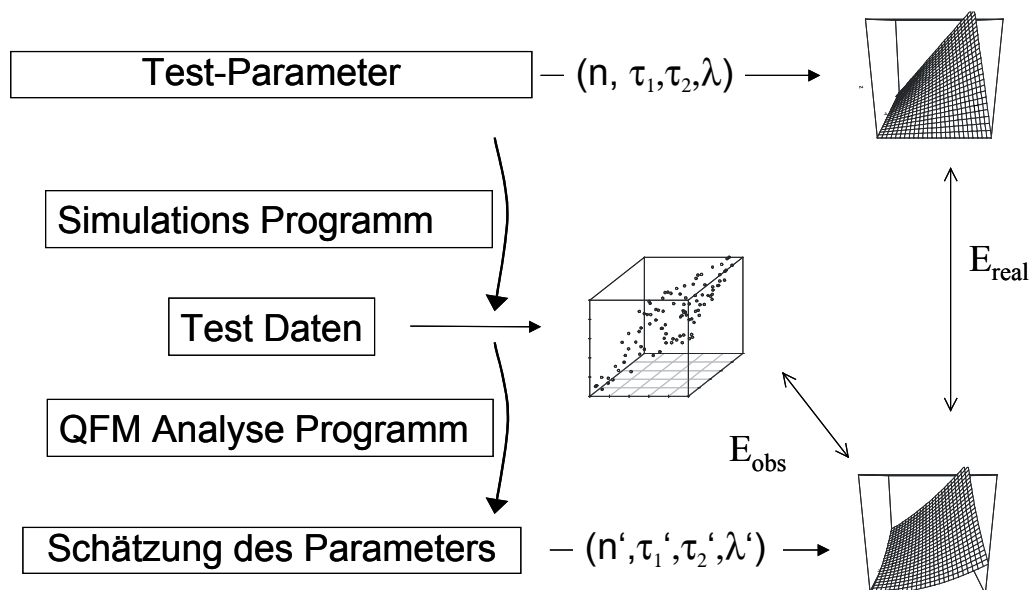


Abb. 5 Ablauf der Sensitivitätsanalyse. Mittels des Testparameters werden Testdaten erzeugt. Mittels des QFM-Analyseprogramms werden die Parameter geschätzt. Zwei Fehlermaße werden berechnet: Der Anpassungsfehler E_{obs} berechnet sich als mittlerer Abstand zwischen Test-Daten und Gesamtfitnessfunktion des geschätzten Parameters. Der Anpassungsfehler ist also klein, wenn sich die Gesamtfitnessfunktion durch den gewählten Parametersatz gut an die beobachteten Daten anschmiegt. Der Schätzfehler E_{real} berechnet sich als mittlerer Abstand zwischen den beiden Gesamtfitnessfunktionen, parametrisiert durch den echten und den geschätzten Parametern. Der Schätzfehler ist also klein, wenn geschätzter Parametersatz und echter Parametersatz zueinander ähnliche Fitnessfunktionen parametrisieren.

2.2.4 Anwendbarkeitsüberprüfung des Analyseprogramms

Die Einsatzfähigkeit eines Werkzeuges lässt sich durch Simulationen validieren. Das Analyseprogramm wird dabei Testdatensätzen ausgesetzt, die mit Hilfe einer Drift-Selektions-Simulation aus bekannten Parameterkombinationen erzeugt wurden. Werden diese Parameterkombinationen durch das Analyseprogramm wieder erkannt, arbeitet es gut. Wenn nicht, gibt es zwei Möglichkeiten: Die 25 Replikate der Schätzungen können streuen oder konsequent abweichen (Abb. 6).

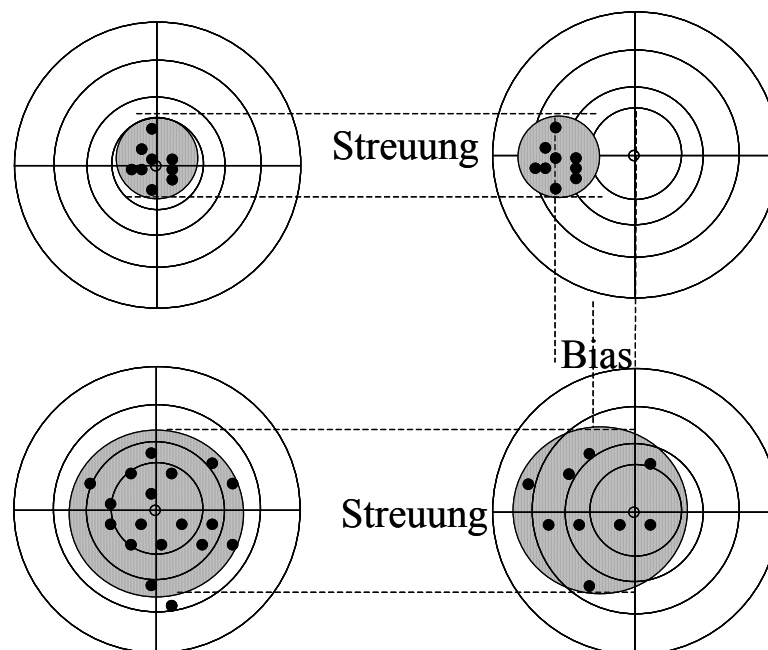


Abb. 6 Streuung und Bias führen zu „Ungenauigkeiten“. Die Summe aus beiden ist damit ein Maß zur Abschätzung der Schätzgenauigkeit des Modells. Dazu wird der ME gemäß Gleichung (2.6) für den Schätzfehler und den Anpassungsfehler eingeführt (siehe auch Abb. 5).

Als Schätzfehler eines Testdatensatzes bezeichne ich das arithmetische Mittel E_{real} der Abstände zwischen der wahren Gesamtfitness G_{real} und der geschätzten Gesamtfitness G_{est} . Damit ist E_{real} analog zu E_{obs} definiert. Zur Normierung wurden beide Parametersätze mit einem Standard-Fitness-Set aus 100 Genotypen (Gitter von 10x10 Genotypen mit gleichverteilten Einzelfitnesses in den Umwelten) berechnet. Über 25 Replikate jedes erzeugten Testszenarios werden Mittelwert und Standardabweichung (SD) von E_{real} berechnet. Das Fehlermaß für den Schätzfehler ergibt sich aus

$$ME_i = |\text{Bias}(E_i)| + |\text{SD}(E_i)|, \quad (2.6)$$

E_i entspricht der mittleren Abweichung (Bias), da der Erwartungswert für den Schätzfehler 0 ist. In ME gehen statistischer und stochastischer Fehler gleichgewichtet ein und ME quantifiziert somit gleichzeitig systematische Abweichung und die Streuung. Es wurde nicht

der Mean Squared Error (MSE) verwendet (Rice 1995), da das Quadrieren von Werten zwischen 0 und 1 zu zu kleinen Werten im Computer führen kann. Analog lässt sich der Anpassungsfehler ME_{obs} aus E_{obs} berechnen. Bei kleinen ME_{real} ist das Analyseprogramm für die Parameterkombination gut und bei großen ME_{real} schlecht einsetzbar. Der Grund lässt sich im Vergleich ME_{real} zu ME_{obs} genauer klassifizieren (s. Tab. 2).

Tabelle 2 Vergleich von Schätzfehler und Anpassungsfehler und ihre Interpretation. ME_{real} beschreibt die Qualität des Analyseprogramms im Erkennen der richtigen Parameterkombination, ME_{obs} beschreibt, wie gut die geschätzte Parameterkombination die simulierten Gesamtfitnesswerte erklärt. ($ME \in [0,2]$)

ME_{real}	ME_{obs}	Interpretation
Groß	Groß	Fehlerhafte Schätzung wegen fehlerbehafteter Testdaten, da das Analysemodell keine Parameterkombination gut Anpassen konnte. Möglicherweise sind die Parameter Obergrenzen zu niedrig gewählt oder die Realität konnte durch die Parameter nicht abgebildet werden.
Groß	Klein	Fehlerhafte Schätzung, weil eine andere Parameterkombination den Testdaten ähnelt (z.B. zu starke genetische Drift, Stichprobengröße zu klein).
Klein	Groß	Gute Schätzung trotz fehlerbehafteter Testdaten, Vorteil eines diskreten Parameterraumes
Klein	Klein	Gute Schätzungen

Die Testdaten wurden durch folgendes einfaches Drift-Selektions-Modell erzeugt. Die Startpopulation besteht aus $N=100\ 000$ Individuen mit $GT=16$ Genotypen. Nun berechnet sich die erwartete Häufigkeit eines Genotyps in der nächsten Generation als das Produkt aus Genotyphäufigkeit und der realisierte Fitness in diesem Zeitraum (Gleichung 2.2). Aus der so berechneten Verteilung werden nun zufällig wieder N Individuen gezogen und bilden die nächste Generation. Zum Schluss berechnet sich die Gesamtfitness G_{obs} nach der Fitnessdefinition für die normierte Gesamtfitness. Die 16 Genotypen wurden in den Einzelumwelten uniformverteilt, sodass die Genotypen ein regelmäßiges Gitter (4×4) als Fitness-Set bilden.

Als Grundmenge für die Simulationen wurde eine Parameterdatei $(n, \tau_1, \tau_2, \lambda) \leq (10, 10, 10, 9)$ mit 2056 Hauptlösungen erzeugt. Die Fehlermaße ME_{obs} und ME_{real} wurden über 25 Replikate und jeden Parametersatz berechnet. Zufallseffekte können Abweichungen nach oben wie nach unten verursachen. Bei hohen Test-Parametersätzen bedeutet dies, dass Abweichungen nach oben nicht möglich wären, wenn als Grundmenge zum Schätzen ebenfalls $(n, \tau_1, \tau_2, \lambda) \leq (10, 10, 10, 9)$ verwendet werden würde. Im Analyseprogramm wurde daher eine größere Grundmenge $(n, \tau_1, \tau_2, \lambda) \leq (12, 12, 12, 11)$ eingesetzt (4529 mögliche Lösungen), was auch ein Überbestimmen hoher Test-Parameter ermöglicht.

Kleine Werte für die Fehlermaße ME belegen eine gute Anwendbarkeit, große Werte eine schlechte Anwendbarkeit. Korreliert man nun die nach Anzahl der Nebenlösungen

gewichteten Parameterkombinationen (n , τ_1 , τ_2 , λ) gegen die Fehlermaße ME_{obs} und ME_{real} , so zeigt sich der Einfluss der Parameter auf die Schätzgenauigkeit.

2.3 Ergebnisse

Ich untersuche die Fähigkeiten des Analyseprogramms anhand simulierter Testdaten. Zwei Fehlermaße werden berechnet: ME_{real} beschreibt die Fähigkeit des Analysemodells, die ursprüngliche Parameterkombination wiederzuentdecken: ME_{obs} beschreibt die Anpassungsfehler der gefundenen Parameterkombination an die simulierten Daten.

Tabelle 3 Absolute Größe der Schätzfehler über alle 2056 Parameterkombinationen. Beide Fehlermaße sind als sehr klein zu werten. Damit ist die Fähigkeit des Analyseprogramms hoch einzuschätzen. SD = Standardabweichung, Anzahl der Parameterkombinationen nach Gewichtung mit Nebenlösungen = 4500. Wertebereich von ME = [0;2]; ME_{real} ist signifikant kleiner als ME_{obs} (Vorzeichen-Test, Anzahl unverbundene Ränge 2056, $ME_{real} < ME_{obs}$ in ~99.85% der Fälle, $Z=45.18$, $p<0.01$; signifikant auch nach Gewichtung mit Nebenlösungen).

	ME_{real}	ME_{obs}
Mittelwert \pm SD	0.001078 \pm 0.000941	0.004188 \pm 0.002394
Median	0.000897	0.003636
10 - 90 Perzentil	0 - 0.002387	0.001647 - 0.007664
Min - Max	0 - 0.007713	0.000908 - 0.013009

Tabelle 3 zeigt Standardstatistiken über alle Testdatensätze. Der Wertebereich von ME ist [0;2]. Daher sind die Mittelwerte als sehr klein und die Fähigkeiten des Analyseprogramms als hoch zu bewerten. Auffälligerweise ist $ME_{obs} > ME_{real}$ für alle Standardstatistiken. Dies belegt, dass die Schätzungen des Analysemodells im Mittel näher an den wahren Parameterkombinationen sind als die simulierten (und Zufallseffekten unterworfenen) Testdaten.

Trotz der geringen Abweichungen können die Parameter unterschiedlichen Einfluss auf die Schätzgenauigkeiten haben, die durch die Standardstatistiken nicht aufgedeckt werden können. Tabelle 4 und Abb. 7 zeigen daher die Einflüsse der Parameter auf die Fehlermaße. Die für ME_{real} signifikanten und einflussreichen Korrelationen für τ_1 (und auch τ_2) erscheinen im Scatterplot ohne deutlichen Einfluss. Hohe Werte für den Zeiteinheitenparameter τ_1 ermöglichen erst hohe Werte für τ_2 (Hauptlösungsproblem). Hohe Werte für beide Parameter sind ein Zeichen für sehr filigrane Umweltverhältnisse, die schwer abbildbar sind und dadurch zu Fehlern führen (s. auch Kapitel 9.5). Der Fehler ME_{obs} hingegen sinkt stark mit steigender Generationenzahl, ist aber von den anderen Parametern unbeeinflusst.

Tabelle 4 Einfluss der Parameterwerte auf den Schätzfehler des Modells. Angegeben sind Spearman's Rangkorrelationskoeffizienten. Werte kleiner |0.5| werden als „schwach“ gewertet. Die Parameter zeigen demnach keinen starken Einfluss auf die allgemeine Schätzgenauigkeit (ME_{real}), aber hohe Abhängigkeit der Anpassungsfehler ME_{obs} von der Anzahl der Generationen. Alle Korrelationen sind signifikant ($p < 0.01$ nach Bonferoni Korrektur).

Parameter	ME_{real}	ME_{obs}
n	0.10	-0.82***
τ_1	0.51***	-0.20***
τ_2	0.41***	-0.19***
λ	0.22***	-0.11***

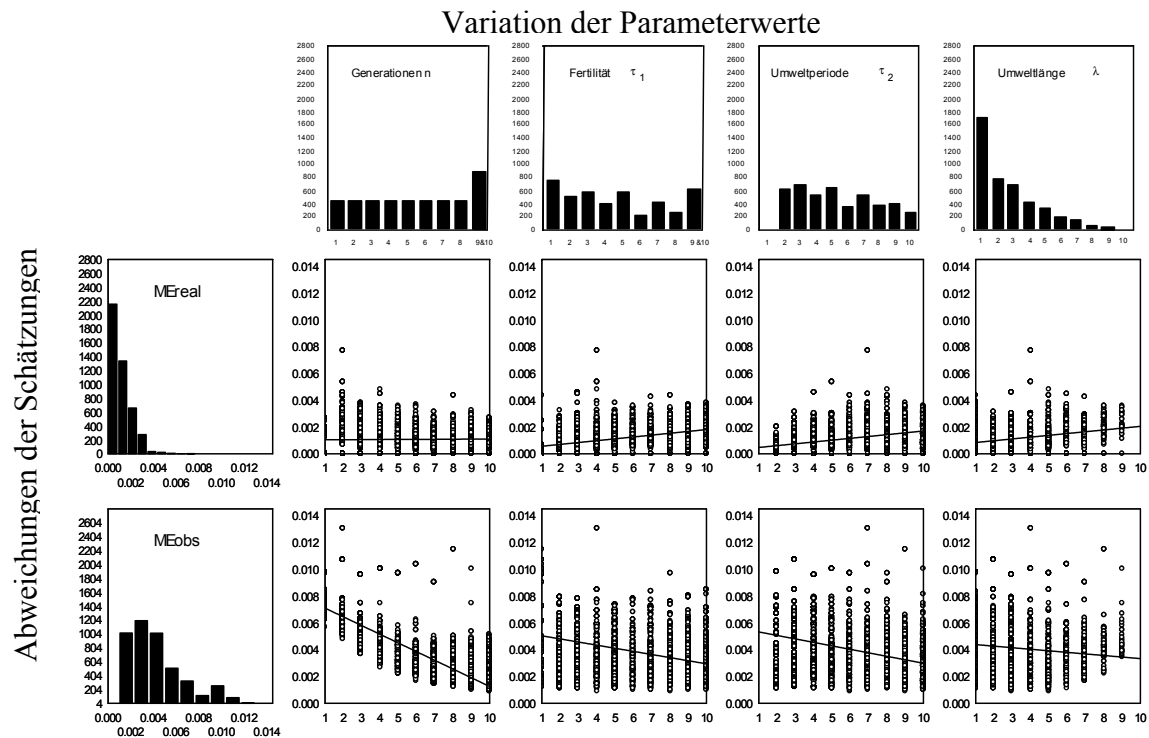


Abb. 7 Einflüsse der Parameterwerte auf die Fehlermaße ME_{real} und ME_{obs} . Als Trendlinien sind die Regressionsgeraden eingezeichnet. Die Abhängigkeiten sind nur für den Anpassungsfehler ME_{obs} von der Anzahl der Generationen offensichtlich.

2.4 Diskussion

Das Quantitative Fitness-Set-Modell ist eine Erweiterung der Fitness-Set-Theorie (Levins 1968). Mit vier Parametern (Generationenanzahl n , Zeiteinheiten τ_1 , Umweltperiode τ_2 , Umweltanteil λ) wird oszillierende Selektion in zwei sich abwechselnden Umwelten beschrieben. Damit ermöglicht das QFM die Erklärung und Analyse genetisch polymorpher Populationen. Die Verhältnisse der Umweltschwankungen aber auch die Körnigkeit der Umweltschwankung werden aus der Sicht des Tieres quantifiziert. Das Modell wurde am Beispiel klonaler Arten hergeleitet; sie ist jedoch für die Analyse von Artengemeinschaften und die Analyse der Vielfalt von Haplotypen oder Allelen in sexuell reproduzierenden

Populationen erweiterbar.

Die Genauigkeit des auf dem QFM basierende Analyseprogramms wurde an Testdaten durch Simulationen untersucht. Die berechneten Fehlermaße zeigen im Mittel nur geringe Abweichungen von den Testparametern (s. Tabelle 3). Die Korrelationsanalyse fand keine starken Beeinflussung der Arbeitsgenauigkeit des Analysemodells unter den untersuchten Bedingungen durch den gewählten Parameterbereich (s. Tabelle 4).

Die Korrelationen sind auch nach Bonferoni-Korrektur noch signifikant, aber als schwach zu werten, sodass ein Einfluss dieser Parameter auf die Fehleranfälligkeit unter natürlichen Bedingungen fragwürdig ist. Diese Aussage unterstützt auch der Scatterplot (Abb. 7).

Nur die Anzahl der Generationen hat einen deutlichen Einfluss auf den Anpassungsfehler, der auch durch $R=-0.82^{***}$ abgesichert ist. Im Gegensatz dazu weist ME_{real} keine Abhängigkeit auf ($R=0.1^{***}$). Wäre die geringere Abweichung durch eine erhöhte Informationsmenge begründet, so würde man auch eine Verbesserung des Schätzfehlers ME_{real} erwarten. Dies ist aber nicht der Fall. Der Anpassungsfehler ME_{obs} misst, wie genau der Parametersatz die beobachteten Daten beschreibt. Dieser Fehler ist erwartungsgemäß kleiner, wenn die Dichte des (diskreten) Parameterraumes in Lösungsnähe größer ist. Dies ist hier der Fall (Tabelle 5). Da bei vielen Generationen mehr Hauptlösungen existieren, ist der Parameterraum dichter, die Gesamtfitnessfunktion kann den beobachteten Daten näher angepasst werden und die Abnahme (und Korrelation) von Anpassungsfehler ME_{obs} bei gleichbleibenden Schätzfehler ME_{real} mit steigenden Parameterwerten von n ist verständlich.

Tabelle 5 Steigende Zahl der Hauptlösungen in Abhängigkeit vom Parameter „Generationen“.

Generationen	1	2	3	4	5	6	7	8	9	10	Σ
Anzahl Hauptlösungen	32	100	159	193	231	245	258	268	276	294	2056

Die Robustheit des Modells wird zudem durch den Vergleich von Anpassungsfehler ME_{obs} und Schätzfehler ME_{real} gezeigt. Aus dem Anpassungsfehler ME_{obs} sollte eigentlich ein erhöhter Schätzfehler des Analysemodells resultieren. Die umgekehrte Aussage wird jedoch von allen Standardstatistiken getroffen ($ME_{real} < ME_{obs}$ in Tabelle 3). Offensichtlich wird der stochastische Effekt der Simulation durch die hochdiskrete Lösungsmenge kompensiert. Somit schätzt das Analyseprogramm unter den Testbedingungen im Mittel sogar besser als die Simulationsdaten den Parameterraum beschreiben.

3 Sensitivitätsanalyse des QFM Analyseprogramms

3.1 Einleitung

Das QFM Analyseprogramm hilft dabei, Biodiversität aufgrund schwankender Umweltbedingungen zu analysieren. Es quantifiziert und qualifiziert oszillierende Selektion, sodass einerseits Aussagen über die Häufigkeit von Umweltänderungen aber auch über die Körnigkeit der Umweltänderungen aus der Sicht der Art getroffen werden können.

Das QFM dient dabei als Werkzeug, um aus experimentellen und Beobachtungsdaten Parameter zu schätzen, die dann in biologischen Begriffen interpretierbar sind. In einer Sensitivitätsanalyse werden Einflussfaktoren auf die Modelleigenschaften, z.B. die Schätzgenauigkeit untersucht. Dazu werden die Einflussfaktoren in einer künstlichen Umwelt variiert. Auf die so gewonnenen Daten wird dann das Werkzeug angewendet.

Für das QFM Analyseprogramm benötigt ein Anwender Aussagen über die Genauigkeit der Parameterschätzungen und Richtlinien zur Anwendung: (1) Existiert ein Einfluss der Diversität auf das Modellverhalten? (2) Welchen Einfluss hat genetische Drift und welche Populationen sind den Effekten der Fitness-Set Theorie ausgesetzt? (3) Welchen Einfluss hat das Ziehen einer Stichprobe auf die Schätzgenauigkeit des QFM?

3.2 Material und Methode

Die Testdaten wurden durch folgendes Drift-Selektions-Modell bei konstanten Populationsgrößen erzeugt (Abb. 8): Die Startpopulation besteht aus N Individuen mit GT Genotypen zum Startzeitpunkt. Nun berechnet sich die erwartete Häufigkeit eines Genotyps (gem. der Fitnessdefinition) in der nächsten Generation als das Produkt aus Genotyphäufigkeit und der realisierte Fitness in diesem Zeitraum (Gleichung 2.2). Aus dieser Verteilung werden nun zufällig N Individuen mit Zurücklegen gezogen und bilden die nächste Generation.

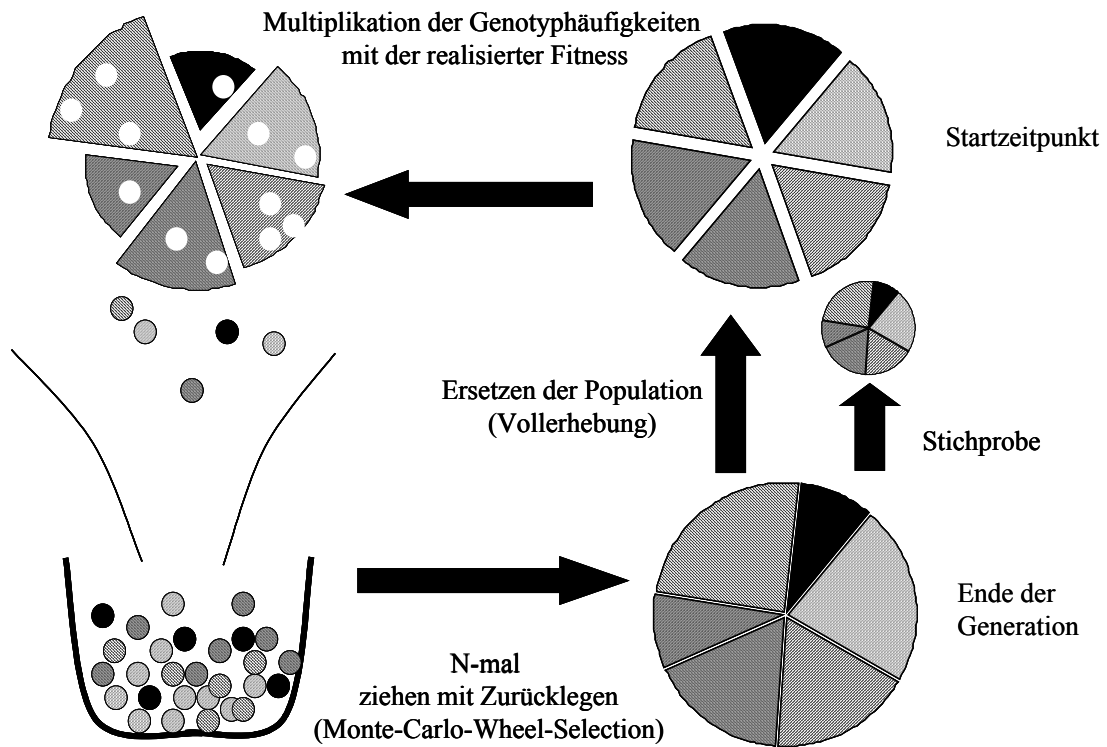


Abb. 8 Ablauf einer Generation des Drift-Selektions-Modells für 6 Genotypen. Jedes Segment stellt die absolute Häufigkeit eines Genotyps dar. Nach der Selektion sind die Häufigkeiten verändert, was durch eine unterschiedlichen Segmentradius dargestellt ist. Die konstante Populationsgröße N wird durch N -faches „ziehen mit Zurücklegen“ im Urnenmodell erhalten. Zum Ende der Generation ist die unterschiedliche Häufigkeit der Genotypen wieder auf die Populationsgröße normiert und wird somit wieder als Winkelunterschied dargestellt.

Die beobachtete Gesamtfitness G_{obs} berechnet sich aus der Genotypfrequenzverschiebung zwischen Startzeitpunkt und Endzeitpunkt (Abb. 5). Die Genotypen wurden in den Einzelumwelten gleichmäßig verteilt, sodass sich ein Gitter als Fitness-Set bildet. Tabelle 6 führt die untersuchten Wertebereiche auf.

Tabelle 6 Werte der Sensitivitätsanalyse. Die Populationsgrößen sollen starke, mittlere und wenig genetische Drift erzeugen. Die Genotypen sind symmetrisch auf beide Umwelten verteilt, sodass sich die Fitness-Sets (f_1, f_2) als Gitter über die beiden Umwelten legen. Der Stichprobeneffekt wird exemplarisch an der Stichprobengröße „100“ demonstriert.

Einflussgröße	Werte
Test-Parameterkombinationen	2056 Hauptlösungen $(n, \tau_1, \tau_2, \lambda) \leq (10, 10, 10, 9)$
Schätz-Parameterkombinationen	4529 Hauptlösungen $(n, \tau_1, \tau_2, \lambda) \leq (12, 12, 12, 11)$
Populationsgröße N	100 000, 1000, 100
Anzahl Genotypen GT	16 (für N=1000 exemplarisch 4, 9, 16, 25, 49, 100)
Stichprobengröße	Vollerhebung, 100
Replikate	25

Über die Replikate wurden nun Schätzfehler ME_{real} und Anpassungsfehler ME_{obs} berechnet (Gleichung 2.6). Der relative Vergleich von ME_{real} zu ME_{obs} gibt darüber Aufschluss, welcher Typ von Fehler zu Abweichungen führt (s. Tabelle 2).

Zu jedem Parametersatz existiert ein maximal abweichender Parametersatz. Um vergleichbare Werte über alle Test-Parameterkombinationen zu erhalten, wurden Schätz- und Anpassungsfehler am maximalen Parametersatz normiert. Als ein Versuchsergebnis erhält man die Verteilung von ME_{obs} und ME_{real} über alle 2056 Test-Parameterkombinationen. Nun lässt sich deren Verteilung z.B. im Balkendiagramm in Abhängigkeit der variierten Einflussfaktoren diskutieren. Mittels eines Vorzeichentests lässt sich auch die Abweichung signifikant absichern (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Als Vergleichsmaß habe ich zudem die Verteilung von zufällig ausgewählten Ergebnissen erzeugt. Diese würde entstehen, wenn das Analysemodell zufällig eine Lösung gleichverteilt aus der Schätz-Parameterkombinationen wählen würde.

Zur Beantwortung der Frage nach dem Einfluss der Diversität auf das Modellverhalten wurde die Anzahl der Genotypen zum Startzeitpunkt variiert. Zum einen erwartet man dort eine Verbesserung der Schätzung mit steigender Genotypenzahl, da sich das QFM auf mehr Informationen stützt. Andererseits existiert eine höhere Diskretheit der Fitnessabbildung, da pro Genotypklasse weniger Individuen vorhanden sind.

Für eine geeignete Genotypenanzahl kann danach der Einfluss der Populationsgröße und genetischer Drift untersucht werden. Auch hier wirkt die Diskretheit der genauen Fitnessabbildung entgegen. Alle Diskretisierungsereignisse führen zu Abweichungen der simulierten Daten von den realen Parameterwerten. Die entscheidenden Fragen sind, wann die

dadurch erzeugten Testdaten zu fehlerbehaftet sind, um den Selektionsprozess abzubilden, und/oder ob das Modell anfällig für systematische Fehler durch „Rundungsprozesse“ ist.

Das Ziehen einer Stichprobe entspricht einem Zufallsexperiment, in dem die Populationsgröße auf die Stichprobengröße ohne Selektionseinwirkung reduziert wird. Solange die Stichprobe die Grundgesamtheit ausreichend gut abbildet, erwartet man auch eine gute Schätzung.

3.3 Ergebnisse

3.3.1 Variation der Genotypanzahl

Die Anzahl der Genotypen weist in Abb. 9 einen starken Einfluss auf die Schätzgenauigkeit ME_{real} auf. Erwartungsgemäß sind Parameter bei sehr niedriger Diversität nur ungenau zu schätzen, denn im Extremfall bleibt nur ein einzelner Genotyp übrig. Dem kann jeder Parametersatz entsprechen, was durch den kleinen Anpassungsfehler ME_{obs} gezeigt ist. Bei niedriger Diversität ist demnach eine genaue Schätzung durch zu geringer Informationsdichte (wenige Genotypen) erschwert.

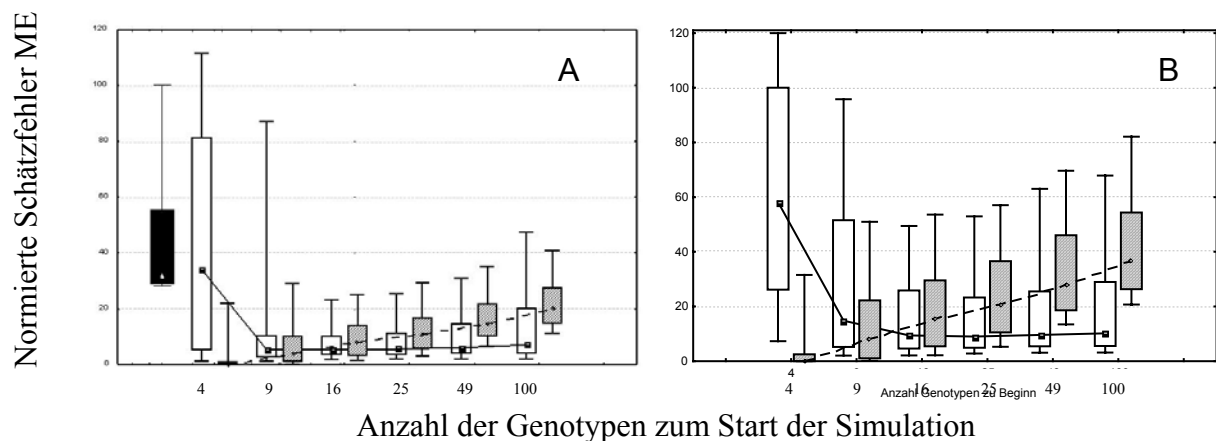


Abb. 9 Normierter Schätzfehler ME_{real} (weiß), Anpassungsfehler ME_{obs} (schraffiert) und die Nullverteilung der Mittelwerte (schwarz) in Abhängigkeit der Anzahl der Genotypen zu Simulationsbeginn. $N=1000$, A) Vollerhebung, B) Stichprobengröße 100. Der Punkt bezeichnet den Median, die Box die 10%-90% Perzentile, die Whiskers die Minima und Maxima.

Bereits neun Genotypen führen zu einer starken Verbesserung der Schätzeigenschaften. Die genaueste Schätzung (kleinste Minima/Maxima und kleinste Perzentile) bei $N=1000$ erfolgt mit 16 Genotypen. Eine weitere Erhöhung der Diversität führt zu erhöhten Schätzfehlern für ME_{obs} und ME_{real} (Tabelle 7, paarweise Vergleiche, Werte $> 50\%$).

Tabelle 7 zeigt, dass ab 16 Genotypen $ME_{\text{real}} < ME_{\text{obs}}$ ist. Nach Tabelle 2 ist dies ein Indiz für

die Modellrobustheit, da das „Rauschen“, das in ME_{obs} abgebildet ist, nicht groß genug ist, um die Schätzgenauigkeit stark zu beeinflussen. Diese Verbesserung der Schätzung bei höherer Diversität (mehr Genotypen) hat Grenzen: Für 49 und vermehrt für 100 Genotypen ist dieses Rauschen schon wieder stärker als noch für 25 Genotypen. Zu geringe und zu hohe Vielfalt führen zu zufallsbedingten Fehlern.

Tabelle 7 Einfluss der Anzahl der Genotypen auf die Fehlerverteilungen. Angegeben ist der Anteil der Parameterkombinationen [%], bei denen die Nullhypothese zutrifft. Vorzeichen-Test mit 2056 unverbundenen Rängen, Vollerhebung, N=1000. Alle Tests sind signifikant mit $p < 0.01$. Beim Vorzeichentest ist der Umkehrschluss bei signifikanter Ablehnung der Nullhypothese gültig.

Genotypenanzahl	4	9	16	25	49	100
$H_0: ME_{real} < ME_{obs}$	2.63%	25.44%	73.54%	99.27%	98.74%	95.91%
Paarweise Vergleiche $H_0:$	4 < 9	9 < 16	16 < 25	25 < 49	49 < 100	
ME_{real}	13.86%	67.89%	62.40%	69.79%	76.79%	
ME_{obs}	99.85%	98.63%	96.74%	99.51%	99.90%	

3.3.2 Variation der Populationsgröße

Abb. 10 zeigt den Einfluss der Diskretisierung aufgrund endlicher Populationsgrößen sehr deutlich: In großen Populationen und bei Vollerhebung funktioniert das Analysewerkzeug fast perfekt. In kleinen Populationen sind Streuung und Median stark erhöht. Trotz der starken Zufallseffekte bei N=100 weicht die Verteilung der Schätzfehler von der Zufallsverteilung ab.

Der Anpassungsfehler ME_{obs} ist immer größer als ME_{real} . Wie im vorgehenden Abschnitt über den Diversitätseinfluss ist auch hier dies als Robustheitsindiz zu verstehen. Für N=100000 und 16 Genotypen ist bei Vollerhebung der Diskretisierungseinfluss schon zu gering, um die Schätzung stark zu beeinträchtigen.

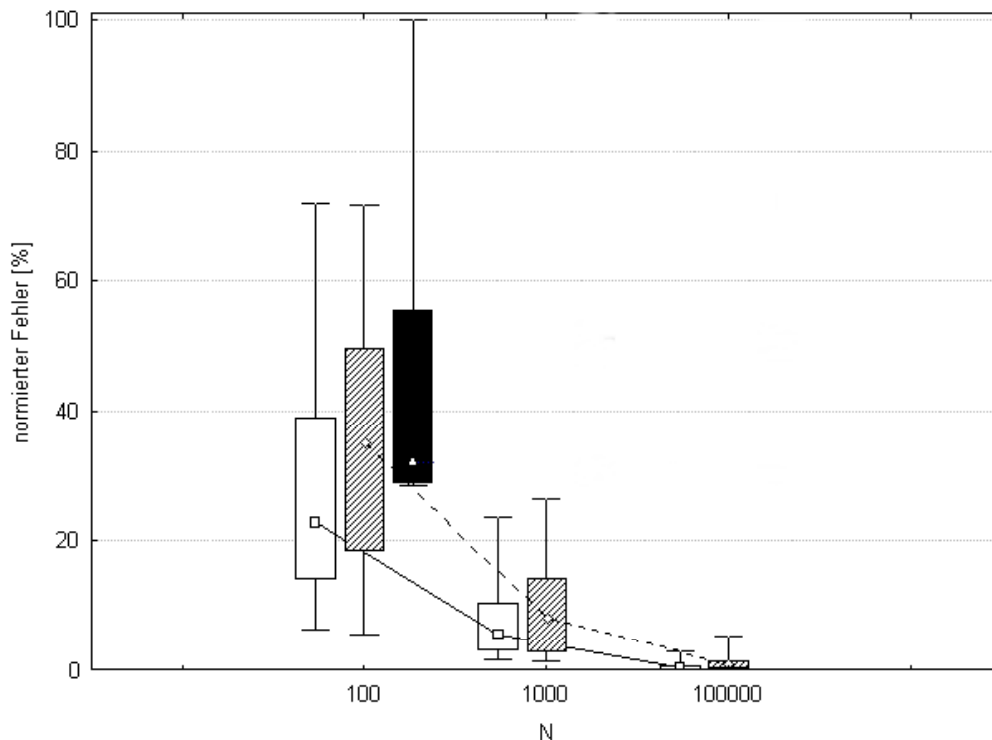


Abb. 10 Zufallsverteilung (schwarz), sowie normierter Schätzfehler ME_{real} (weiß) und Anpassungsfehler ME_{obs} (schraffiert) in Abhängigkeit der Populationsgröße N . Der Punkt bezeichnet den Median, die Box die 10%-90% Perzentile, die Whiskers die Minima und Maxima. Der Fall $N=100000$ als Vollerhebung wurde im vorhergehenden Kapitel untersucht. Der Boxplot „Zufall“ entspricht einer Zufallsverteilung, in der sich aus dem mittleren Fehler und Standardabweichung der paarweisen Vergleiche aller Parameterkombinationen zusammensetzt. Die Verteilungen sind in der hier dargestellten Relation (von links nach rechts) innerhalb und zwischen den Populationsgrößenkategorien hoch signifikant unterschiedlich (Vorzeichen-Test, 2056 unverbundene Ränge, $p < 0.001$)

Die Korrelationsanalyse in Tabelle 8 (Vollerhebung) zeigt, dass der wesentliche Einfluss auf den Schätzfehler durch die Populationsgröße ausgeübt wird ($R=0.93^{***}$). Innerhalb der Populationsgrößen jedoch hat die Anzahl der Generationen für $N=100$ einen schwachen bis für $N=1000$ zu einen stark zu nennenden Einfluss. Je mehr Generationen beobachtet werden, umso genauer wird die Schätzung.

Tabelle 8 Einfluss der Parameterwerte auf den Schätzfehler ME_{real} und Anpassungsfehler ME_{obs} des Modells. Spearman's Rangkorrelationen kleiner $|0.5|$ werden als „schwach“ gewertet. Die Anzahl an Generationen (n) und die Populationsgröße (N) zeigen den stärksten Einfluss. Alle Korrelationen mit *** sind signifikant ($p < 0.01$ nach Bonferoni Korrektur), unmarkierte sind nicht signifikant $p > 0.05$. Gewichtung nach Anzahl der Nebenlösungen, da der Parameterraum auf die Hauptlösungen reduziert wurde.

Vollerhebung	N=100		N=1000		N=100 000	
Parameter	ME_{real}	ME_{obs}	ME_{real}	ME_{obs}	ME_{real}	ME_{obs}
n	-0.52***	-0.44***	-0.70***	-0.77***	0.10***	-0.82***
τ_1	-0.05***	0.03	-0.09***	-0.10***	0.51***	-0.20***
τ_2	-0.19***	0.06***	-0.11***	-0.15***	0.41***	-0.19***
λ	-0.11***	0.05	-0.02	-0.09***	0.22***	-0.11***
Korrelation mit N			-0.93***	-0.92***		
Stichprobe 100						
n			-0.86***	-0.84***	-0.90***	-0.89***
τ_1			-0.05	-0.05***	-0.06	-0.06***
τ_2			-0.25***	-0.19***	-0.25***	-0.21***
λ			-0.15***	-0.12***	-0.16***	-0.14***
Korrelation mit N			-0.14***	-0.14***		

3.3.3 Stichprobeneffekt

Die Untersuchung des Stichprobeneinflusses wurde anhand der Daten der Genotypvariation vorgenommen und in Abb. 9B dargestellt (Anzahl der Testdatensätze: 24672). Der Stichprobeneinfluss bei Populationsgrößenvariation geht aus Tabelle 8 hervor.

Das Ziehen einer Stichprobe verstärkt Schätzfehler und Anpassungsfehler ($ME_i^{Vollerhebung} < ME_i^{Stichprobe}$: 97.40%*** für $i=ME_{real}$ und 90.51%*** für $i=ME_{obs}$, Vorzeichentest, Anzahl unverbundener Ränge = 12175). Die Verteilung der Schätzfehler der Stichproben ist von der Zufallsverteilung der Schätzfehler signifikant verschieden ($ME_{real}^{Stichprobe} < ME_{Zufallsverteilung}$: 83.51%, Anzahl unverbundener Ränge = 12336). Stichprobe verursachen im Vergleich mit den Vollerhebungen schlechtere Schätzungen. Aber selbst die geringe Stichprobe von 100 Individuen ist noch klar von einem Zufallsprozess entfernt.

Höhere Generationsanzahlen verbessern die Schätzung (Tabelle 8). Der Einfluss der Populationsgröße ist gering. Damit wirkt in diesem Wertebereich die Stichprobengröße nicht negativ mit der Populationsgröße zusammen.

3.4 Diskussion

Für die Einflussfaktoren untersuche ich hier zunächst die Anwendbarkeit des QFM als Analysewerkzeug. Danach beschreibe ich jeweils die Auswirkung von oszillierender Selektion auf natürliche Populationen. Diese Untersuchung erfolgt durch die Anwendung des QFM auf eine simulierte, endliche Population parthenogenetischer Individuen, die oszillierender Selektion unterworfen sind.

Wie beeinflusst Diversität das Modellverhalten? Durch die Variation der Genotypenzahl wurde untersucht, wie Unterschiede in der Diversität auf die Schätzgenauigkeit wirkt (Abb. 9). Selbst bei einer Vollerhebung (Abb. 9A) führt wenig Diversität zu starken Schwankungen in der Schätzung, 16 Genotypen lieferten bei einer Populationsgröße von $N=1000$ die besten Schätzungen, die sich mit zunehmender Genotypenzahl wieder verschlechtert. Bei zu großer Vielfalt werden die erwarteten Prozesse der FT ungenau abgebildet, denn steigende Vielfalt bei gleichbleibender Populationsgröße bedeutet, dass jede Genotypklasse mit weniger Individuen besetzt ist. Betrachtet man die Genotypen als „Subpopulationen“, so sind diese klein und es herrscht in ihnen höhere genetische Drift.

Wann sind die durch genetische Drift erzeugten Testdaten zu fehlerbehaftet, um den Selektionsprozess abzubilden? Besonders deutlich wird der Diskretisierungseffekt bei Variation der Populationsgröße in Abb. 10: Kann die Population mit $N=100\ 000$ die gesamte Komplexität der Parameterkombinationen nachvollziehen, so führt Diskretisierung zu Fehlern im Schätzverhalten. Dabei erweist sich das QFM als robust, denn die Schätzungen sind im Mittel näher an den wahren Parameterkombinationen als an den Testdaten ($ME_{\text{real}} < ME_{\text{obs}}$ in Abb. 10). Damit lässt sich das Analyseprogramm durchaus auch für kleine Populationen einsetzen, wenn auch mit größerer Ungenauigkeit als in großen Populationen.

Dies bedeutet für natürliche Populationen, dass die Populationsdynamik auch einen Einfluss auf den Fitness-Set ausüben kann. Diese Ergebnisse sind demnach auf stark fluktuierende Populationen nicht direkt übertragbar und mit Vorsicht anzuwenden.

In der Natur lassen sich die Zusammenhänge zwischen Schätzgüte und Populationsgröße bzw. Diversität nicht beeinflussen, da oft weder die Diversität noch die Populationsgröße vom Untersuchenden verändert werden kann. Jedoch werden folgende Eigenschaften natürlicher Populationen offenbar: Selbst mittelgroße Populationen ($N=1000$) können unterschiedlichen Drifteffekten je nach Biodiversität unterworfen sein. Die Abbildbarkeit der oszillierenden

Selektion ist also nicht nur eine Eigenschaft der Populationsgröße, sondern auch der in der Population enthaltenen Anzahl an Genotypen und ihrer relativen Fitnessunterschiede. Geht man von antagonistischen Verhältnissen zwischen hoher Diskretisierung (kleine Population oder viele Genotypen) und Biodiversität aus, so müsste eine obere Grenze für Biodiversität im Kontext der Fitness-Set Theorie existieren.

Ist das Modell anfällig für systematische Fehler durch Rundungsprozesse? Für die Anwendung lassen sich zur Zeit folgende Schlüsse ziehen: Das QFM verfällt selbst unter ungünstigen Driftbedingungen noch nicht in ein zufälliges Verhalten. In kleinen und mittleren Populationen sinkt der Schätzfehler mit steigender Anzahl Generationen. Stichproben bewirken diesen Effekt auch bei großen Populationen. Mit einer verlängerten Beobachtungsdauer lässt sich also die Genauigkeit des Verfahrens steigern. Weiterhin ist der Einfluss nicht kontrollierbarer Faktoren zu bedenken: Bei einer Stichprobe von 100 Individuen ist die Schätzung in mittleren Populationen ($N=1000$) fast genauso gut wie in einer großen Population ($N=100\ 000$).

Der diskrete Parameterraum wirkt dem Zufall entgegen, sodass die Schätzung im Mittel genauer ist als die Testdaten. Der Einsatz molekularer Marker ermöglicht das Erkennen vieler verschiedener Genotypen, sodass man ein Übersteigen der hier erwünschten Zahl leicht erwarten könnte. Dabei muss man sich allerdings ins Gedächtnis rufen, dass es hierbei nicht um die Betrachtung neutraler Mutationen, sondern um selektiv unterscheidbare Genotypen geht. Träger neutraler Mutationen wirken im Kontext der Fitness-Set Theorie identisch, da sie denselben Punkt im Fitness-Set einnehmen.

Hier ergibt sich ein interessanter Aspekt zur Neutralisten-Selektionisten Debatte (Kimura 1983), in der „Neutralisten“ die genetische Vielfalt auf neutraler Mutationen zurückführen. Polymorphie wird dann z.B. durch genetische Drift in geographischen Strukturen erhalten. Selektionisten berufen sich auf genetische Systeme (z.B. Überdominanz) und verschiedenartiger Selektion in geographischen Strukturen, die Polymorphie erhalten. Dazu sind fitnessbeeinflussende Mutationen nötig. In der Fitness-Set Theorie sind beide Sichtweisen integriert: Neutrale Mutationen nehmen identische Plätze im Fitness-Set ein, fitnessveränderte Genotypen können besser, gleichgut oder schlechter als die vorhandenen Genotypen sein. Zudem erklärt die Fitness-Set Theorie, wie oszillierende Fitness „quasi-neutrale Fitnessbereiche“ schafft: Mutationen entlang der Fitnessisokline sind fitness-neutral.

4 Fitnessbänder schützen vor genetischer Verarmung kleiner Populationen

4.1 Einleitung

Der große Einfluss der genetischen Drift ist in der Populationsgenetik Grundlage vieler erfolgreich eingesetzter Modelle (Hartl und Clark 1989). Für die Diskussion von Selektion wird der Einfluss der Populationsgröße als zufälliger Prozess zumeist als störend empfunden und durch verschiedene mathematische Annahmen ausgeschlossen. In Kapitel 3 konnte ich jedoch zeigen, dass die Schätzgenauigkeit des QFM als Analysewerkzeug abhängig von Populationsgröße und Stichprobengröße ist.

Damit das Analysewerkzeug einsetzbar ist, müssen Aussagen und Empfehlungen erarbeitet werden, die einem Anwender Richtlinien zum Einsatz zur Hand geben. Nur eine derartige Qualitätskontrolle des Werkzeugs ermöglicht den vertrauenswürdigen Einsatz.

Doch sind diese Abweichungen in der Genauigkeit nicht nur experimentelle Artefakte unter Simulationsbedingungen. Kleine Populationsgrößen sind als Bottleneck- oder Foundereffekt in evolutionären Situationen bekannt. Da bereits die Vollerhebung derartige Abweichungen zeigt, stellt sich die Frage nach der Bedeutung ungenauer Abbildbarkeit von Fitnessisoklinen für die genetische Vielfalt. Welche Typen von Parametersätzen werden dann als „Alternativen“ erkannt?

Weiterhin lassen sich in der Schwankung über die Parametersätze auch Anzeichen von Inhomogenität sehen. Sind also verschiedene Parametersätze unterschiedlich anfällig für Schwankungen und bedeutet das eine Möglichkeit, evolutionäre Zustände (Generalisten und Spezialisten) zu erklären ?

4.2 Material und Methode

Der Drifteffekt wurde für vier Populationsgrößen untersucht ($N=10^2, 10^3, 10^4, 10^5$). Dabei kam das einfache Drift-Selektions-Modell aus Kap. 2 mit sechzehn gleichverteilten Genotypen als Fitness-Set zum Einsatz. Überprüft wurden alle Hauptlösungen $(n, \tau_1, \tau_2, \lambda) \leq (10, 10, 10, 9)$, d.h. 2056 Parameterkombinationen (Eingabemenge). Als Definitionsmenge des Schätzvorganges kam dieselbe Hauptlösungsmenge zum Einsatz.

Für jedes der 25 Replikate pro Parametersatz wurden die geschätzten Parameterkombinationen aufgezeichnet. Für jede Populationsgröße existiert eine Wiederfundmatrix als quadratische Matrix aus Eingabemenge und Definitionsmenge (2056x2056). Die Matrixeinträge ergeben sich aus der absoluten Häufigkeit der Schätzung eines spezifischen Parametersatzes.

Analog zur Wiederfundmatrix ist die Abstandsmatrix aufgebaut. Diese beinhaltet als Abstandsmaß zweier Parameterkombinationen den mittleren absoluten Abstand zweier Parameterkombinationen bei einem gitterförmigen äquidistanten 10x10 Fitness-Set. Zur Wahrung der Vergleichbarkeit der einzelnen Parameterkombinationen wurden die maximalen Abstände jeder Parameterkombination auf 100% normiert. Legt man nun um den gesuchten Parameterbereich einen „Unschärfbereich“ von 0..x%, so lässt sich

- (1) die Anzahl der Wiederfunde im x%-Unschärfbereich berechnen als die Summe der Wiederfundmatrixeinträge, deren Abstandsmatrixeintrag kleiner x% ist,
- (2) die Dichte des x% Unschärfbereichs definieren als die Anzahl der Parameter, die im x%-Unschärfbereich liegen.

Der Erwartungswert der Anzahl der Wiederfunde ist die Anzahl der Replikate (25), da im optimalen Fall jeder Parameter jedes Mal wiedergeschätzt wird. Die beobachtete Verteilung der Wiederfunde über die Eingabemenge lässt sich mittels χ^2 -Statistik auf signifikante Unterschiedlichkeit testen. Dies habe ich für alle Populationsgrößen und elf verschiedenen Unschärfewerte getan. Die Abhängigkeiten lassen sich dann grafisch durch Zuordnen von Populationsgröße zum Unschärfewert zu der entsprechenden Wert der χ^2 -Statistik (FG 2055, $\alpha \leq 5\%$) ermitteln.

Analog der Wiederfundmatrix ergibt sich aus den Simulationen die Fehlfundmatrix. Hierbei sind die Matrixeinträge jedoch die Häufigkeit, mit der ein Parametersatz fehlerhafterweise für den echten Parametersatz gefunden wurde.

4.3 Ergebnisse

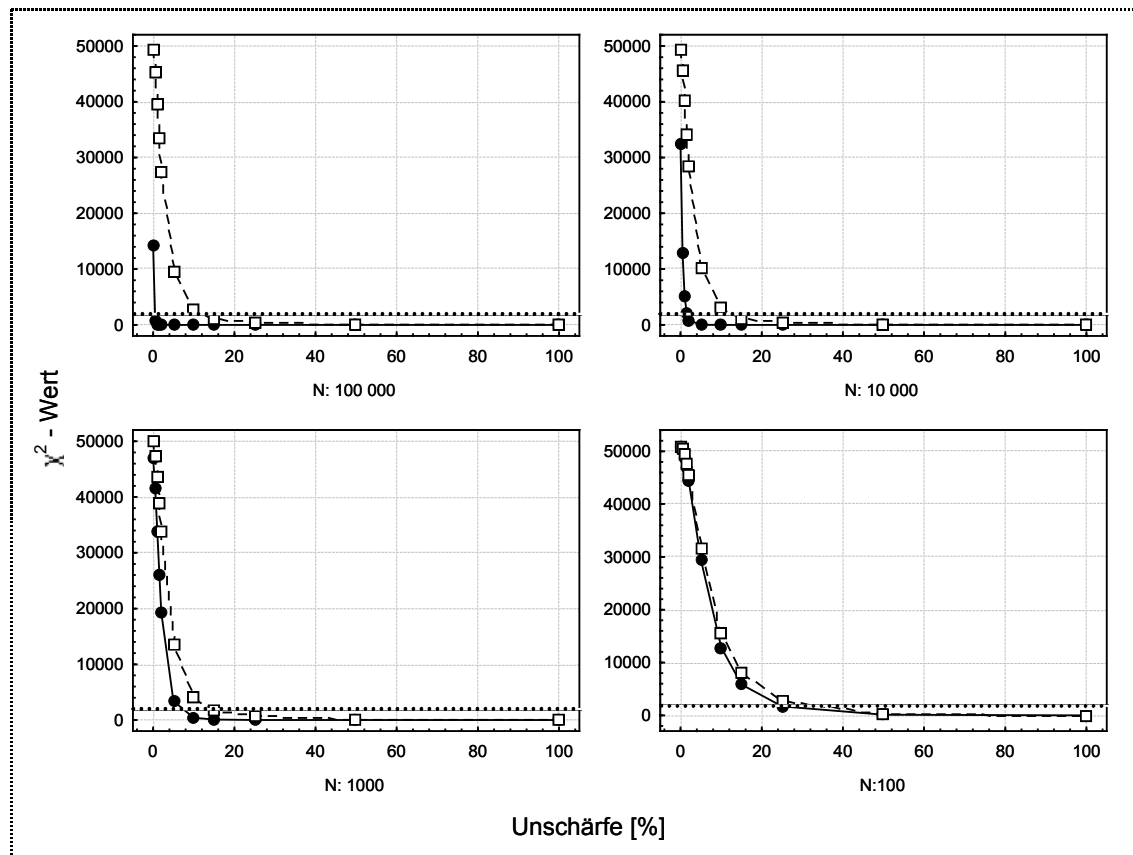


Abb. 11 Zuordnung eines Unschärfewertes zu einer Populationsgröße. Der χ^2 -Test quantifiziert die Abweichung der Wiederfunde für einen vorgegebenen Unschärfebereich. 5% Signifikanzniveau (χ^2 -Wert 1950, FG 2055) ist als gepunktete Linie dargestellt. Die ausgefüllten Punkte repräsentieren die Daten der Vollerhebung, die leeren bei einer Stichprobengröße 100.

Abb. 11 zeigt die Ungenauigkeit der Schätzung in Abhängigkeit vom Unschärfewert und der Populationsgröße. Kleine Populationen führen durch genetische Drift zu einer unscharfen Abbildbarkeit der oszillierenden Fitness. Selbst sehr große Populationen ($N=100\,000$) führen zu signifikant vom Erwartungswert unterscheidbaren Wiederfinden unter exakten Bedingungen (0%-Unschärfe).

Für eine vergleichbare Mächtigkeit des QFM als analytisches Tool empfiehlt sich die Verwendung eines populationsgrößenabhängigen Unschärfewertes, z.B. auf dem 5% Signifikanzniveaus des χ^2 -Tests. Es zeigt sich zudem, dass die Stichprobenerhebung für Populationsgrößen $N=1000 \dots 100\,000$ eine vergleichbare Unschärfeabhängigkeit aufweisen; für $N=100$ und Stichprobe=100 ergeben sich durchgängig höhere Werte. Damit lässt sich die Unschärfe über die bekannte Stichprobengröße und nicht über die zunächst unbekante (oder geschätzte Populationsgröße) abbilden.

Auch lässt sich in Abb. 11 erkennen, dass bereits kleine Unschärfewerte die Schätzgenauigkeit stark erhöhen. Daraus lässt sich als Hypothese für den Vorgang der Fehlbestimmung annehmen, dass ein Parametersatz umso wahrscheinlicher fehlerhaft geschätzt wird, je ähnlicher er dem echten Parametersatz ist.

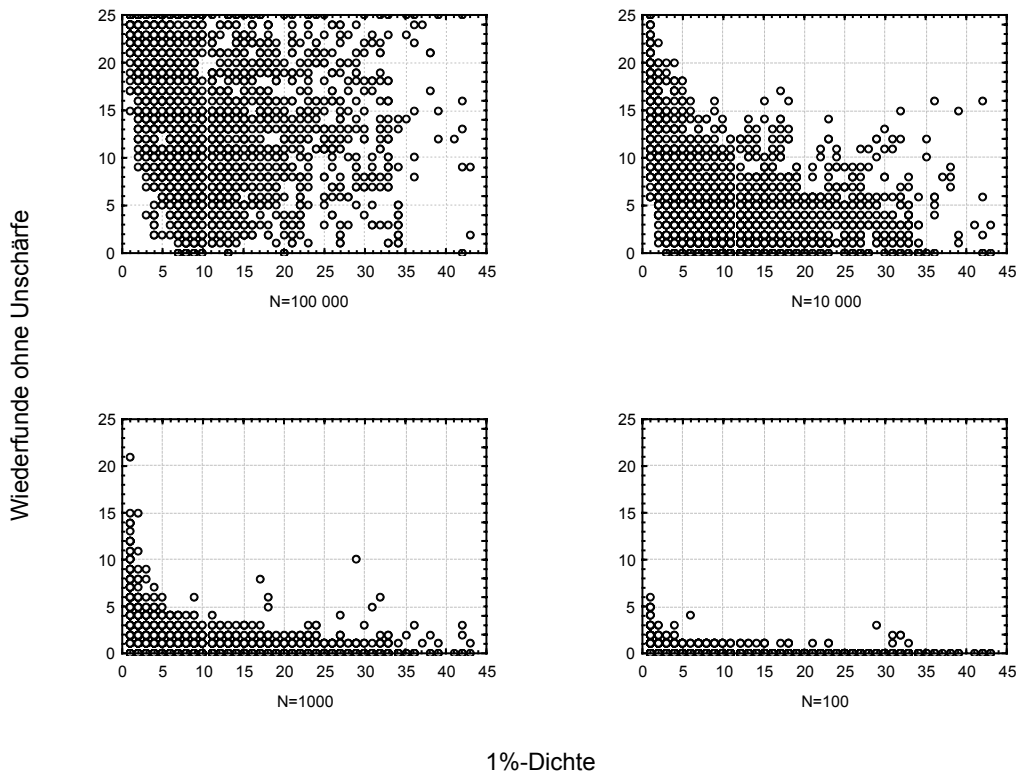


Abb. 12 Dichteabhängigkeit der Wiederfunde im 0%-Unschärfebereich unter verschiedenen Populationsgrößen. Die Dichte ist als die Anzahl der Parameterkombinationen im 1%-Unschärfebereich definiert. Als Wiederfunde werden hier nur die exakten Wiederfunde gewertet.

Die Wiederfundwerte in Abb. 12 zeigen zunächst die erwartete Abhängigkeit von der Populationsgröße, d.h. die Wiederfunde sind in großen Populationen wahrscheinlicher. Weiterhin zeigt sich bei kleineren Populationsgrößen, dass Parameter in diskreten Bereichen, d.h. weniger dichten Bereichen des Parameterraumes, eher wiedergefunden werden, als solche in hohen Dichten. Dies unterstützt die Hypothese, dass vornehmlich die Ähnlichkeit von Parameterkombinationen zu einer Fehlschätzung führt. In hohen Dichten kann bereits eine kleine Abweichung einer anderen Parameterkombination entsprechen und es somit zu einer Fehlschätzung kommen. Niedrige Dichten hingegen sind per Definition diskret und bilden auch ungenaue „Fitnessdaten“ auf die ursprüngliche Parameterkombination ab.

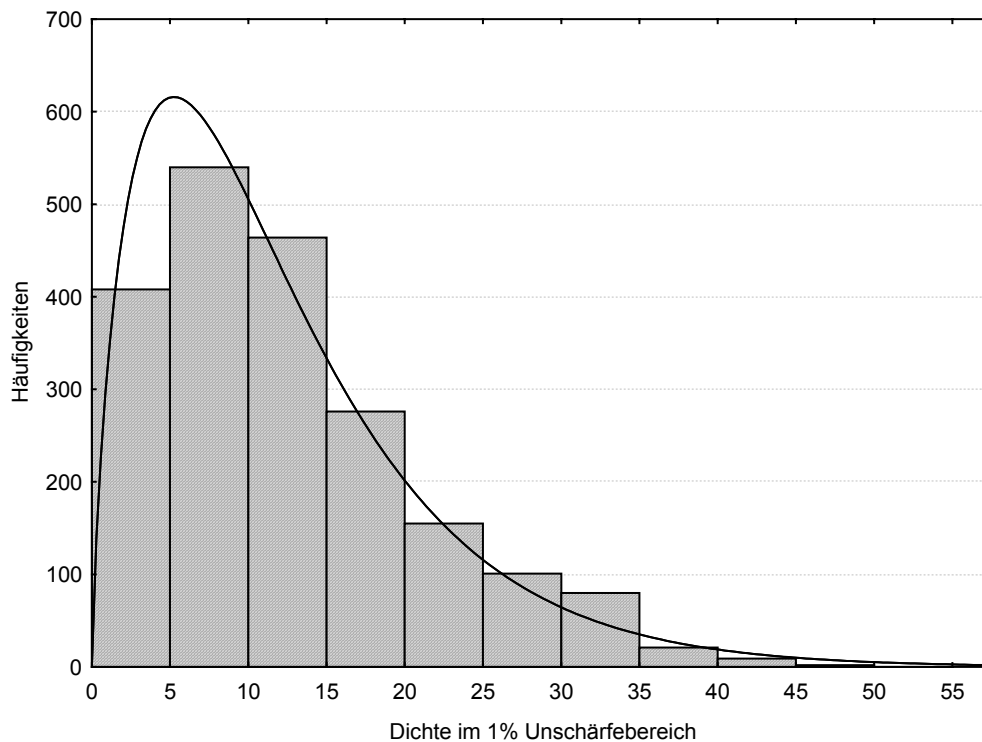


Abb. 13 Die Häufigkeitsverteilung der Dichten im 1% Unschärfebereich in der Hauptlösungsmenge $(n, \tau_1, \tau_2, \lambda) \leq (10, 10, 10, 9)$. Die Dichte ist angegeben als die Anzahl alternativer Parameterkombinationen im 1% Unschärfebereich aus der Hauptlösungsmenge. Die Verteilung scheint einer Gammaverteilung zu folgen, weicht jedoch signifikant von ihr ab ($\chi^2 = 48.75119^{***}$, FG = 8).

Die Dichte um die Parameterkombinationen ist unheimlich (Abb. 13). Es existieren also dichte und nicht dichte Parameterbereiche. Die Ähnlichkeit zu einer Gammaverteilung ist grafisch offensichtlich, eine Übereinstimmung lässt sich jedoch signifikant ablehnen ($\chi^2 = 48.75119^{***}$, FG = 8). Nichtsdestotrotz kann man zumindest auf einen graduellen Übergang von dichten zu nicht-dichten Parameterbereichen spekulieren. Es scheint eine spezifischen räumliche Anordnung im Parameterraum vorzuliegen, die zu dieser Verteilung der 1%-Dichten führt.

4.4 Diskussion

Diskretisierung durch endliche Populationsgrößen erzeugt zufällige Prozesse, die zu einer ungenauen Abbildung von Fitness führen. Es wirken daher Parameterkombinationen, die ähnlich zu der angenommenen Parameterkombination sind. Durch die Einführung des Unschärfewertes und die Erweiterung von einer Einzellösung auf eine Lösungsmenge erweitern sich die Fitnessisoklinen zu Fitnessbändern, in denen sich die Dynamik der Fitness - Set Theorie bewegen kann. Die unscharfe Abbildbarkeit nimmt mit steigender Populationsgröße ab. Es besteht also auch ein direkter Zusammenhang zwischen

Populationsgröße und einem Unschärfebereich, der sich z.B. über ein 5% Signifikanzniveau quantifizieren lässt.

Es ergeben sich daraus Konsequenzen für die Anwendung des QFM als Analysewerkzeug. Zunächst kann man nicht erwarten, dass eine einzige Lösung die Wirklichkeit repräsentiert. Da finite Populationen Fitnessbänder erzeugen, werden auch die experimentellen Ergebnisse von Freilanduntersuchungen notgedrungen eine Lösungsmenge darstellen. Die Dominanz des Stichprobeneffektes über den Einfluss der Populationsgröße (Abb. 11) führt zu einer Robustheit des Analysemodells gegenüber fehlerhaften Populationsgrößenschätzungen und Populationsgrößenschwankungen. Die Nutzung der Stichprobengröße zur Kalibrierung ist ausreichend.

Noch interessanter sind die aus dem Unschärfeeffekt ableitbaren evolutionären Konsequenzen. Die Unschärfe fasst die Umgebung der Parameterkombination in eine Menge zusammen. Damit wirkt auf die genetische Vielfalt der Population nicht nur genau eine Fitnessisokline ein, sondern ein Fitnessband aus ähnlichen Fitnessisoklinen. Steigende Unschärfe mit sinkender Populationsgröße bedeutet eine Verbreiterung des Fitnessbandes (bis auf 30% bei einem 5% Signifikanzniveau), und damit eine Sicherung der genetischen Vielfalt kleiner Populationen unter oszillierender Selektion. Gründerpopulationen und Bottlenecksituationen entziehen sich damit weitgehendst der genetischen Verarmung aufgrund (oszillierender) Selektion und sind damit „nur“ den Verlustprozessen der genetischen Drift ausgesetzt. Für die Neutralisten-Selektionisten Debatte (Kimura 1983) ergibt sich hier eine Teillösung für den Fall der oszillierenden Umwelten: Innerhalb des Fitnessbandes gelten die Argumente der Neutralisten, außerhalb des Fitnessbandes gelten die der Selektionisten.

Zudem lassen sich Prognosen über Generalisten und Spezialisten formulieren. Die uneinheitliche Dichte des Parameterraumes führt zu unterschiedlichen Ausprägungen der Fitnessbänder: Dichte Bereiche resultieren in quasi-kontinuierlichen Fitnessbändern, während diskretere Bereiche aus nur wenigen alternativen Fitnessisoklinen bestehen. Letzteres müsste als evolutionäre Konsequenz in Spezialisten münden, die strikt an ihr enges, stets wirkendes Fitnessband angepasst sind. Die Entwicklung (kostenintensiver) Plastizität ist hier von Nachteil. Diese Anpassung entspricht auch dem Kanalisierungs-Effekt (Kawecki 2000), der als evolutionäre Anpassung an „unerwartete“ Umweltänderungen eine Unterdrückung von Spezialisierung postuliert. Spezies, die dichte Parameterräume repräsentieren sind demnach

evolutionär als Generalisten bevorzugt, da ihre Populationen selbst bei mittlere Größe stets einem Fitnessband (einer „plastischen Fitnessisokline“) ausgesetzt sind.

Auch ergeben sich ökologische Konsequenzen und Konsequenzen zur Anwendbarkeit des QFM als erklärendes Modell. Umweltschwankungen sind nicht immer identisch – durch den Unschärfefaktor ist dies aber auch nicht zwingend notwendig, solange die Veränderungen, z.B. der Periodizität oder der Intensität der wirkenden Einzelfitnesswerte, zu Schwankungen innerhalb des Unschärfebereichs führen. Der Unschärfeefferkt führt somit zur Modellrobustheit und ermöglicht das Wirken der Fitness-Set Theorie auch bei nicht strikten, d.h. „verrauschten“ Umweltschwankungen.

5 Asynchrone Reproduktion ermöglicht feinkörnige Interpretation grobkörniger Umweltvariation

5.1 Einleitung

Die Fitness-Set Theorie geht in der Herleitung von einem gleichzeitigen Ablauf des Reproduktionszyklusses in allen Individuen der beobachteten Art aus. Für viele Tier- und Pflanzenarten in temperaten Breiten, die sich jährlich einmal reproduzieren und eine Winter-Diapause eingehen, trifft diese Annahme gut zu. Andere Arten in temperaten Breiten (z.B. Aphiden), viele Arten in den Tropen, aber auch die meisten Mikroorganismen, reproduzieren sehr oft innerhalb eines Jahres oder sogar kontinuierlich. Dieses Verhalten ist in der Herleitung der Fitness-Set Theorie unberücksichtigt, obwohl die meisten experimentellen Daten (Kassen 2002) sich auf Algen, Bakterien, Viren und nur einige auf Insekten beziehen.

Synchrone und asynchrone Reproduktion haben sich im Laufe der Evolution mehrfach entwickelt. Auch die Bildung von Jahrzehnte überdauernden Dauerstadien sind von Bakterien, z.B. *Bacillus anthracis* (Winkle 2002), ebenso belegt wie von Crustaceen z.B. *Daphnia galeata* (Hairston et al. 2001). Diese Strategie ist aus der Botanik als das Prinzip der Bodensamenbank bekannt und lässt sich als eine Strategie ansprechen, mit der genetische Vielfalt über lange Zeiträume hinweg bewahrt werden kann. In diesem Kapitel zeige ich, dass ähnlich zur Bodensamenbank asynchrone Reproduktion als Strategie zur (kurzfristigen) Bewahrung genetischer Polymorphie in grobkörniger Umweltvariation angesehen werden kann. Weiterhin ergeben sich daraus Hypothesen zur Radiation von Arten und Tendenzen zur Entwicklung niedriger Artenvielfalt in grobkörnigen Umwelten, z.B. den temperaten Breiten im Gegensatz zu hoher Artenvielfalt in feinkörnigen Umweltvariation, z.B. unter tropischen Umweltbedingungen.

5.2 Material und Methode

5.2.1 Herleitung des QFM bei asynchroner Reproduktion

Im QFM werden nicht-überlappende Generationen mit einer synchronisierten Reproduktion vorausgesetzt. Bei überlappenden Generationen existieren synchronisierte (z.B. Bäume) und asynchrone Reproduktion (Wasserflöhe, Bakterien). Synchronisierte Abläufe sind oftmals mittels diskreter Altersklassenmodelle abgebildet. Asynchrone Reproduktion kann hingegen

durch die Betrachtung z.B. der mittleren Reproduktionsrate, d.h. als Mittelwertsbetrachtung modelliert werden. Im Folgenden ist das QFM für „überlappende Generationen bei vollständiger asynchroner Reproduktion“ hergeleitet, womit die komplizierte Altersklassenmodellierung vermieden werden kann. Dabei muss man Annahmen (6) aus Kapitel 2 des QFM folgendermaßen modifizieren:

(6b) Überlappende Generationen: Die Reproduktion der Individuen erfolgt nicht gleichzeitig oder korreliert, und alle Altersstadien treten gemeinsam auf. Aus diese Annahme resultiert auch die Genotypasynchronität, d.h. jeder Genotyp reproduziert für sich asynchron. Damit schließe ich Strategien der zeitlichen Einnischung aus.

Im QFM stellt sich die Gesamtfitness g eines Genotyps als das Produkt von arithmetischen Mitteln W dar. Diese Mittelwerte sind über einen festen Zeitraum τ_1 ermittelt, der zu einem spezifischen Zeitpunkt t startet. Dieser wird durch die Reihenfolgefunktion r festgelegt. Es treten maximal τ_2 verschiedene Umweltmotive auf und damit existieren maximal τ_2 verschiedene Mittelwerte.

Im synchronen Modell treten diese in einer spezifischen Reihenfolge hintereinander auf (s. Gleichung 2.3). Im asynchronen Modell treten alle Umweltmotive gleichzeitig auf, da zu jedem Zeitpunkt alle Altersstadien existieren. Die asynchrone Gesamtfitness ist also

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \prod_{i=1}^n \frac{\sum_{j=0}^{\tau_2-1} [W(r(j, \tau_1, \tau_2), \tau_1, \tau_2, \lambda, f_1, f_2)]}{\tau_2} \quad (5.1)$$

Da hier alle möglichen Umweltmotive gleichzeitig auftreten verliert die Reihenfolgefunktion ihre Bedeutung; damit vereinfacht sich Gleichung (5.1) zu

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \prod_{i=1}^n \frac{\sum_{j=0}^{\tau_2-1} [W(j, \tau_1, \tau_2, \lambda, f_1, f_2)]}{\tau_2} \quad (5.2)$$

Damit ist der Term innerhalb des Produktes von der Produktlaufvariablen i entkoppelt und das Produkt lässt sich als Potenz schreiben

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \left(\frac{\sum_{j=0}^{\tau_2-1} [W(j, \tau_1, \tau_2, \lambda, f_1, f_2)]}{\tau_2} \right)^n = \left(\frac{1}{\tau_2} \sum_{j=0}^{\tau_2-1} [W(j, \tau_1, \tau_2, \lambda, f_1, f_2)] \right)^n \quad (5.3)$$

Die realisierte Fitness W entspricht nun ebenfalls dem arithmetischen Mittel der aufgetretenen Einzelfitnesswerte (Gleichung 2.2):

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \left(\frac{1}{\tau_2} \sum_{j=0}^{\tau_2-1} \left[\frac{1}{\tau_1} \sum_{x=j}^{j+\tau_1-1} w(j, \tau_2, \lambda, f_1, f_2) \right] \right)^n = \left(\frac{1}{\tau_2 \tau_1} \sum_{j=0}^{\tau_2-1} \sum_{x=j}^{j+\tau_1-1} w(j, \tau_2, \lambda, f_1, f_2) \right)^n \quad (5.4)$$

Wir sehen nun, dass j jeden möglichen Anfangszeitpunkt innerhalb der Umweltperiode einnimmt ($0.. \tau_2-1$) und dass x von diesem Anfangszeitpunkten ausgehend alle Umweltmotive exploriert. Die Umformung zur rechten Seite der Gleichung (5.4) zeigt, dass der Mittelwert über die gemeinsame Periode von Umwelt- und Reproduktionsperiode gebildet wird ($\tau_1 * \tau_2$). Dies bedeutet, dass hier ein ganzzahliges Vielfaches (τ_1 -mal) der Umweltperiode auftaucht. Der Mittelwert über ein Vielfaches der Umweltperiode entspricht aber dem über eine Umweltperiode. Es ergibt sich also

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \left(\frac{\lambda}{\tau_2} \cdot f_1 + \frac{(\tau_2 - \lambda)}{\tau_2} f_2 \right)^n \quad (5.5)$$

Diese Formel ist identisch mit der für feinkörnige Umweltvariation. Aus Asynchronität der Reproduktion resultiert eine feinkörnige Sichtweise von grob- oder mittelkörniger Umweltvariation.

5.2.2 Erhalt genetischer Vielfalt unter fein- und grobkörniger Sicht

Der Erhalt genetischer Vielfalt wird als Verlust von Genotypen in dem einfachen Drift-Selektions-Modell aus Kapitel 2 quantifiziert ($N=100\ 000$ Individuen, Mittelwerte über 25 Replikate). Als Fitness-Set wurde eine regelmäßiges Gitter aus 16 Genotypen gewählt.

5.3 Ergebnisse

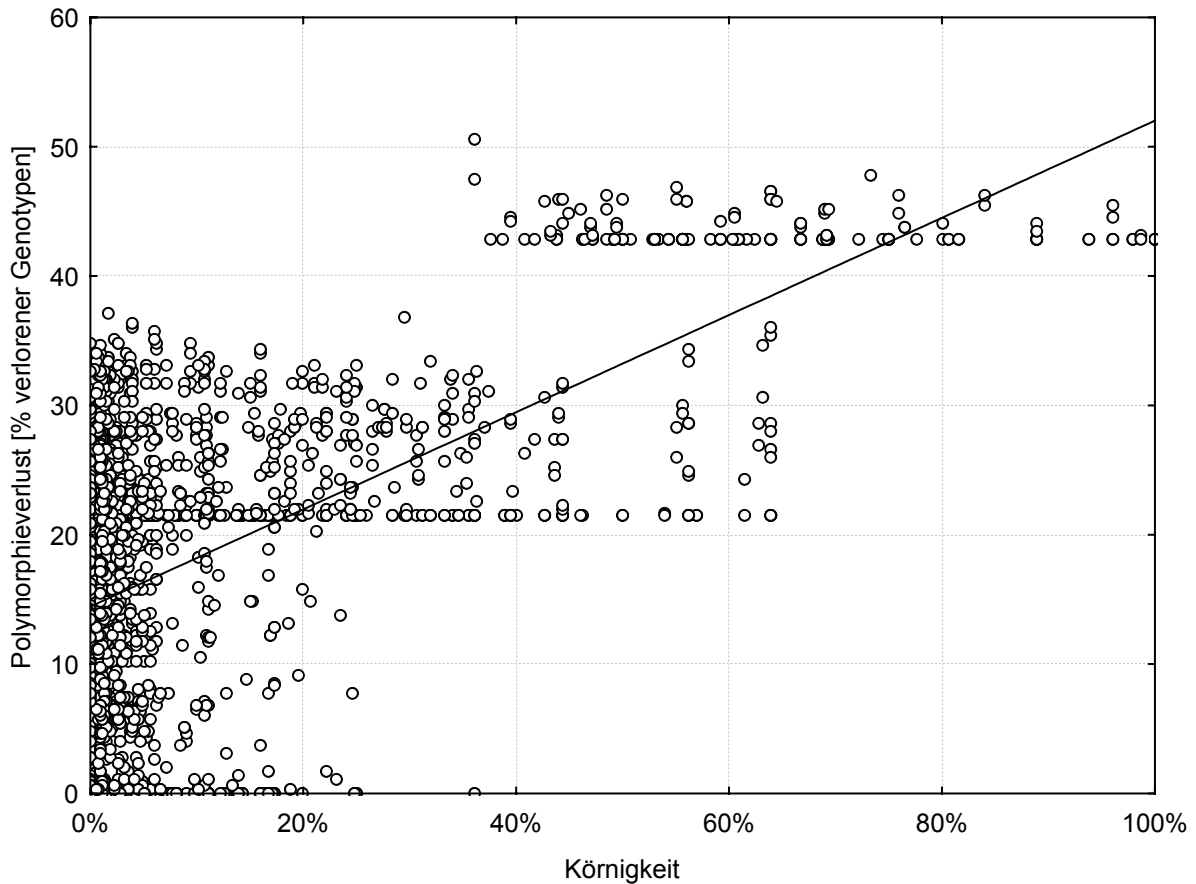


Abb. 14 Polymorphieverlust in der Abhängigkeit der Umweltkörnigkeit. Die Körnigkeit wurde nach Gleichung (2.5) berechnet. Grobkörnige Umwelten führen zu stärkeren Verlust von Genotypen als feinkörnige Umwelten. Der Fitness-Set aus 16 Genotypen im gleichmäßigen Gitter führt zu gleichzeitigen Verlusten von Genotypen und damit zur Stufenbildung. Eingezeichnet wurden alle 2056 Hauptlösungen aus $(n, \tau_1, \tau_2, \lambda) < (10, 10, 10, 9)$ im Drift-Selektions-Modell mit $N=100000$ parthenogenetischen Individuen.

Unter grobkörniger Umweltvariation ergibt sich stärkerer Polymorphieverlust als unter feinkörnigen Bedingungen. Mathematisch formuliert zeigt sich dieses Problem als Vergleich von geometrischen (g) und arithmetischen Mittelwert (a), denn es gilt:

$$\begin{aligned}
 G_{\text{feinkörnig}} &= \left(\frac{\lambda}{\tau_2} \cdot f_1 + \frac{(\tau_2 - \lambda)}{\tau_2} f_2 \right)^n \geq \left(f_1^{\frac{\lambda}{\tau_2}} \cdot f_2^{\frac{(\tau_2 - \lambda)}{\tau_2}} \right)^n = G_{\text{grobkörnig}} \\
 \Leftrightarrow \quad \frac{\lambda}{\tau_2} \cdot f_1 + \frac{(\tau_2 - \lambda)}{\tau_2} f_2 &\geq f_1^{\frac{\lambda}{\tau_2}} \cdot f_2^{\frac{(\tau_2 - \lambda)}{\tau_2}} \\
 \Leftrightarrow \quad a = \frac{1}{\tau_2} \cdot \sum_{i=1}^{\tau_2} f_i &\geq \sqrt[\tau_2]{\prod_{i=1}^{\tau_2} f_i} = g
 \end{aligned} \tag{5.6}$$

Von diesen Mittelwerten wissen wir, dass die Gleichheit nur für $f_1=f_2=\dots=f_{\tau_2}$ gilt; ansonsten gilt $a > g$ (Bronstein et al. 1997). Daraus folgt, dass die Gesamtfitness in

feinkörniger Umwelt immer größer oder gleich der Gesamtfitness in grobkörniger Umwelt ist und somit die Fitnessunterschiede im Fitness-Set geringer sind. Somit ist der Polymorphieverlust in feinkörniger Umweltvariation kleiner als in grobkörniger. Unter der Annahme, dass ein rascher Polymorphieverlust insbesondere für (zyklisch) parthenogenetische Arten in eine (Spezialisierungs- und) Speziierungsfalle führt, lässt sich eine Tendenz für das Leben in feinkörnigen Umweltvariation bzw. die Evolution einer feinkörnigen Sichtweise grobkörniger Umweltvariation postulieren.

Wechsel in der Körnigkeit

Die Körnigkeit der Umweltvariation, d.h. schnellere oder langsamere Fluktuation der Umwelt, kann sich durch äußere Wechsel (z.B. Eindeichen von Grünland und dem Verhindern regelmäßiger Überflutung) oder innere Änderungen, d.h. durch Evolution von Fähigkeiten der Art (z.B. längere Lebensdauer, Mehrjährigkeit, asynchrone Reproduktion) verändern.

Hierbei erweist sich der „generalistische“ Genotyp (f_i , f_i) als wesentlich: Er liegt an dem Berührungspunkt von grobkörniger und feinkörniger Fitnessisokline. Einen Wechsel von feinkörniger Umweltvariation zu grobkörniger Umweltvariation überlebt er zumindest unbeeinträchtigt. Die entstehende grobkörnige maximale Fitnessisokline liegt zudem höher als die entsprechende feinkörnige und damit belegt hier Abb. 15 den Generalisten als fitter.

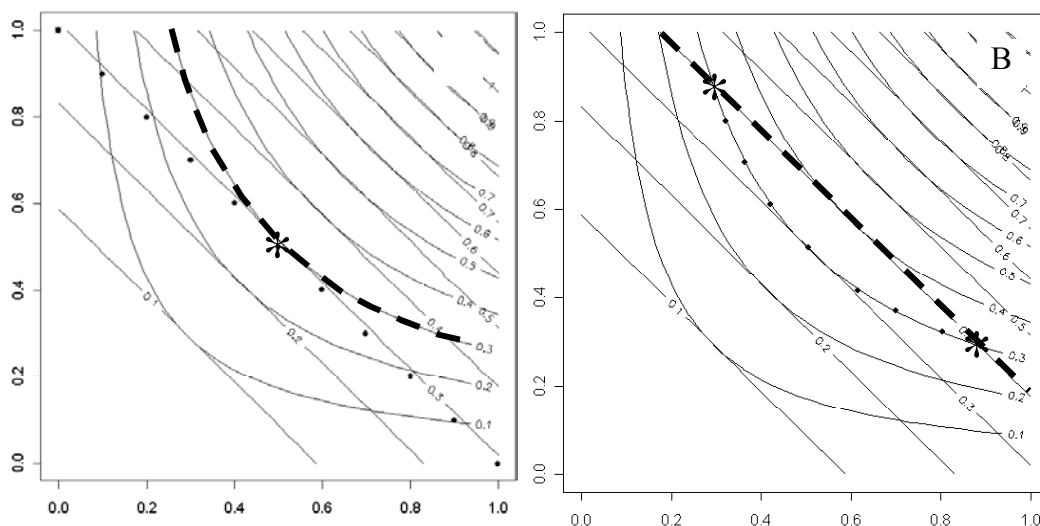


Abb. 15 Wechsel der Körnigkeit. Überlagerte Fitnessisoklinen - feinkörnig zu grobkörnig (A), grobkörnig zu feinkörnig (B). In A überlebt (*) der Generalist, in B die Spezialisten. • Genotypen auf der maximalen Fitnessisokline vor dem Wechsel, ---- : Fitnessisokline nach dem Wechsel. Beispiel mit (2,1,2,1) für grobkörnig; (2,2,2,1) für feinkörnige Umweltvariation.

Bei einem Wechsel von grobkörniger zu feinkörniger Umweltvariation (Abb. 15B) bestimmen die Spezialisten die maximale Fitnessisokline, und der Generalist ist unterlegen. Spezialisten könnten sich also in grobkörniger Umweltvariation durch asynchrone Reproduktion und der damit verbundenen „feinkörnigen Sichtweise“ einen Fitnessvorteil gegenüber Generalisten verschaffen.

5.4 Diskussion

Asynchrone Reproduktion führt zu einer feinkörnigen Sichtweise auch bei grobkörniger Umweltvariation. Der Verlust genetischer Vielfalt ist bei feinkörniger Variation kleiner als bei grobkörniger Umweltvariation (Abb. 14). Damit lässt sich eine feinkörnige Sichtweise der Umwelt, also auch asynchrone Reproduktion, als Strategie zur Aufrechterhaltung genetischer Polymorphie postulieren.

Andere vorbeugende Mechanismen der Aufrechterhaltung bestehender genetischer Polymorphie sind zum einen sexueller Art, z.B. (a) Heterozygotenvorteil führt zu balanciertem Polymorphismus (Dobzhanski 1957), (b) dominante Vererbung schützt seltene Allele in heterozygoten Individuen vor Selektion. (c) Linkage positiv und negativ selektierter Gene verlangsamt Polymorphieverlust dramatisch (Gimelfarb 1998). Zum anderen gibt es räumliche Schutzmechanismen z.B. kann in (d) räumlich getrennten Populationen, in denen durch genetische Drift oder Selektion unterschiedliche Allele fixiert wurden (Buri 1956; Chappell et al. 1988), aufgrund fehlender genetischer Vielfalt keine weitere Selektion wirken. (e) Clines in Allelfrequenzen sind die Ergebnisse räumlicher Übergangsstrukturen, in denen entweder graduell ändernde Selektion wirkt oder sich eine Kontaktzone zweier Extremhabitats beschreibt (Chappell et al. 1988). Die räumlichen und sexuellen Mechanismen führen bei lokaler Verarmung zu einem globalen Diversitätsschutz, wobei auf Artniveau die genetische Vielfalt erhalten bleibt. Ein dritter Ansatz ist die zeitliche Strategie, z.B. das Prinzip der Bodensamenbank (Turelli et al. 2001), das sich auf Sporen haploider Organismen ebenso erweitern lässt wie auf sexuell oder asexuell erzeugten Diapausestadien von Tieren (Caceres 1997).

Viele der sexuell realisierten Schutzmechanismen sind für vorwiegend parthenogenetischen Arten und im verschärften Fall bei parthenogenetischen haploiden Arten nicht wirksam. Die Nutzung räumliche Strukturen führt einerseits in die „Gefahr“ der allopatrischen Speziation, bzw. ist innerhalb von Populationen oder in homogenen Lebensräumen nicht möglich. Dies

ist jedoch bei Bakterien der Regelfall und auch bei Eukaryoten, z.B. Pilzen, Algen aber auch zyklisch parthenogenetischer Arten wie Daphnien nicht selten. In derartigen Gruppen ist die asynchrone Reproduktion ein wichtiger Verlustvermeidungsmechanismus.

Experimentelle Belege unterstützen diese Sichtweise (Kassen und Bell 1998). Die Autoren haben mit der haploiden, parthenogenetischen Grünalge *Chlamydomonas reinhardtii* Selektionsversuche in konstanten und variierenden Umwelten (Licht, Dunkel) durchgeführt. Sie belegen eine klare Anpassung an konstant helle, konstant dunkle und variierenden Umwelten. Jedoch konnten sie keine Unterschiede für feinkörnige und grobkörnige Umweltvariation feststellen. Mit den publizierten Daten ließ sich eine Fitness-Set Analyse durchführen:

Tabelle 9 Experimentelle Fitness-Daten für konstante und oszillierende Umwelten (Kassen und Bell 1998) der Grünalge *Chlamydomonas reinhardtii*. Die Genotypen wurden in ihren Umwelten isoliert und dann den anderen Umwelten ausgesetzt. Die Berechnung der Körnigkeit erfolgte nach Gleichung (2.5). Der Anpassungsfehler ist der Mittelwert der Abstände von der parametrisierten Gesamtfitnessfunktion zu den beobachteten Gesamtfitnesswerten.

Genotypen	Fitness 1 (hell)	Fitness 2 (dunkel)	Gesamtfitness („Feinkörnig“)	Gesamtfitness („Grobkörnig“)
Lichspezialist 1	1.195	0.408	0.771	1.370
Lichspezialist 2	1.038	0.463	0.831	1.757
Lichspezialist 3	1.378	0.088	0.648	1.124
Dunkelspezialist 1	0.205	0.878	0.381	0.370
Dunkelspezialist 2	0.988	1.003	0.561	0.267
Dunkelspezialist 3	0.728	0.938	0.498	0.364
Finegrained1	1.385	0.628	0.751	1.640
Finegrained2	1.278	0.733	0.681	1.567
Finegrained3	1.008	0.538	1.078	1.834
Coarsegrained1	1.435	0.478	0.901	0.68
Coarsegrained2	0.918	0.193	0.731	0.467
Coarsegrained3	1.108	0.828	0.578	0.734
Parameter der QFM-Analyse			(1,1,2,1)	(3,3,5,4)
Körnigkeit des Parameters			0%	10%
Anpassungsfehler			0.18431	0.24197

Eine Fitness-Set Analyse ergab, dass beide variierenden Ansätze eine feinkörnige Sichtweise repräsentieren. Für das QFM Parameter $(n, \tau_1, \tau_2, \lambda)$ ergaben sich (1,1,2,1) im feinkörnigen Experiment und (3,3,5,4) im grobkörnigen Experiment. Beide Parameterkombinationen sind als feinkörnig einzustufen. Dies kann ein Grund sein, warum weder Evolution von Vielseitigkeit (feinkörnig) noch von Plastizität (grobkörnig) belegt werden konnte (Kassen und Bell 1998). Zudem erscheint der Fitnessanteil im Licht als der einzig wirklich bedeutende Faktor (senkrechte Fitnessisoklinen im feinkörnigen, fast senkrechte im grobkörnigen Fall).

Neben der Anwendbarkeit des QFM auf experimentelle Daten sind auch allgemeinere Situationen erklärbar: Generalisten sind bei Änderungen von fein- zu grobkörnigen Umweltvariation im Vorteil. Diesen Vorgang würde man bei tropischen Taxa erwarten, die sich dann in temperaten Breiten ausgebreitet haben. Hierbei wäre die Gründerpopulation der plötzlichen Grobkörnigkeit temperater Sommer-Winter Fluktuationen ausgesetzt und nur ein Genotyp wäre, im Extremfall, „der Fitteste“. Spezialisten würden bei Änderungen von grob zu feinkörniger Variation überleben. Aus den Taxa temperater Breiten, die sich in tropische Feinkörnigkeit ausbreiten können, überleben diejenigen der Gründerpopulation, die am besten an die Einzelumwelten angepasst waren. Im Extremfall können hier zwei spezialisierte Genotypen überleben und damit Paradebeispiele für die Möglichkeit disruptiver Selektion darstellen. Man müsste in diesen Phylogenien eine Tendenz zur Radiationen bei Einwanderung in die Tropen bzw. beim Driften der Kontinente von temperaten in tropische Breiten finden. Die Prozesse der Fitness-Set Theorie wären somit evolutionstechnische Gründe für eine größere Artenvielfalt tropischer und den relativen Artenarmut temperater Zonen.

6 Ausblick

Die Möglichkeiten der Fitness-Set Theorie sind mit der vorliegenden Arbeit noch nicht ausgeschöpft. Auch das QFM ist sicherlich nicht „das eine“ Modell zur vollständigen quantitativen Erfassung aller Details. Von den einschränkenden Annahmen (s. Kapitel 2.2.1) wurde bisher nur die der synchronisierten Reproduktion untersucht. Im Folgenden sollen die Erweiterungen des Modells und mögliche Folgeuntersuchungen formuliert werden.

6.1 Der Modellorganismus

Der unterstellte Modellorganismus lebt nur eine Generation und daher weist die Population keine Altersstruktur auf. Er reproduziert sich synchronisiert und damit leben alle Genotypen in identischen Umweltabfolgen. In den Untersuchungen war die finite Populationsgröße über die Zeit konstant.

Die Annahme der Synchronität bewies sich in Kapitel 5 als wesentlich für eine grobkörnige Umweltsichtweise. In Altersklassenmodellen können Mischformen asynchroner Phasen und synchroner Phasen auftreten: durch Sterben aller Altersklassen bis auf ein Lebensstadium, z.B. der Diapauseeier von Cladoceren oder Rotatorien in austrocknenden Gewässern, wird die vorhergehende Asynchronität wieder synchronisiert. Dies wird auch als Moran-Effekt bezeichnet (Gurney et al. 1991; Gurney et al. 1994)).

Unterschiedliche „Fitnessbeiträge verschiedener Alterklassen“ ergeben möglicherweise eine Überlagerung verschiedener Fitness-Sets. Das Prinzip der Bodensamenbank produziert zeitlich überlappende Generationen und gleicht Genotypfrequenzverschiebungen aus. Über diesen Dämpfungseffekt (Gottlieb 1974) ist als Polymorphieerhaltung zu werten und kann über das Prinzip der Bodensamenbanken erfolgen (Turelli et al. 2001). Daher ist für die Anwendung oder experimentelle Validerung der Fitness-Set Theorie wichtig, dass die Life-History einer untersuchten Art mit dem unterstellten Modellorganismus übereinstimmt.

Die Populationsgröße skaliert über das Unschärfepinzipp den Einfluss der Fitnessisoklinen oder -bänder auf den Fitness-Set. Starke Populationsgrößenschwankungen sind daher ein nicht zu unterschätzender Einflussfaktor bei der Anwendung und Analyse. Insbesondere sind diese Schwankungen dann zu erwarten, wenn eine Umweltschwankung eintritt: Die schlecht angepassten Individuen können eher sterben, als eine neue Population mit den nun bevorzugten Genotypen aufgebaut ist – ein Zusammenbruch der Populationsgröße ist die

Folge. Das Zusammenspiel regelmäßiger und zufälliger Umweltschwankungen können über den Resonanz-Effekt auch zu starken Populationsausbrüchen führen (Blarer und Doebeli 1999). Variierende Populationsgröße kann jedoch relativ einfach in das untersuchte Selektions-Drift-Simulationsmodell integriert werden (s. Abb. 8). Solange bleibt jedoch die Wirkungsweise der Fitness-Set Theorie unter diesen Bedingungen noch unklar.

In der Untersuchung führt Fortpflanzung zu keiner Fitnessveränderung. Das Muster des Fitness-Sets bleibt konstant, bis auf das Verschwinden einiger Genotypen aufgrund von Selektion und genetischer Drift. Sexuelle Reproduktion führt jedoch zu derartigen Veränderungen und somit zu einer Dynamisierung des Fitness-Sets: Bisher nicht vorhandene Genotypen können auftreten oder verschwinden, ohne dass Drift oder Selektion dies verursacht haben. Die Interaktion von Rekombination und Selektion führt zu polymorphieerhaltenden Systemen (z.B. Überdominanz). Daraus ergibt sich eine zeitliche Dynamik des Fitness-Sets und Muster, die nicht auf der Fitness-Set-Theorie basieren. Dieses Zusammenspiel von Populationsgenetik und Fitness-Set Theorie wurde in den bisherigen Untersuchungen nicht berücksichtigt.

Jenseits der Veränderung des Genotyps ist die Fitness auch vom physiologischen und morphologischen Zustand des Individuums abhängig. Plastizität, die eine Fitnessveränderung in der Zeit bewirkt, erzeugt aus den einzelnen Punkten der Genotypen Flächen von Phänotypen. Da Fitnesswerte als Mittelwerte über mehrere Individuen oder Stichproben hinweg ermittelt werden, ließe sich durch Einbeziehen der Streuung der Werte, bzw. der einzelnen Werte ein erweitertes Fitness-Set-Theorie-Modell für Plastizität entwickeln.

6.2 Umweltschwankungen

Aber auch die Umweltabbildung ist abstrahiert und nicht vollständig analysiert: Das QFM basiert in der hier präsentierten Form auf regelmäßigem Abwechseln zweier Umwelten. Das QFM bietet aufgrund der Definition verschiedener Fitnesssebenen Möglichkeiten der Erweiterung. Insbesondere die Formulierung der aktuellen Fitness (Gleichung 2.1) lässt eine Anpassungen auch an mehrere Umwelten zu. Es unterscheidet sich in dieser expliziten Trennung zweier Umwelten von vielen vorliegenden theoretischen Arbeiten, die fluktuierende Umwelt als gauss-verteilte Zufallsvariable um einen Mittelwert betrachten (Tuljapurkar 1990; Ellner und Sasaki 1996; Turelli et al. 2001). Ob diese Ergebnisse auch für bimodale Verteilungen gültig sind, wie man sie für oszillierende Fluktuationen annehmen muss, bleibt

fraglich.

Auch Veränderungen der Amplitude der Umweltschwankung führen zu Veränderungen am Fitness-Set. Eine Verkleinerung der Amplituden führt zu konvexen bis hin zu linearen Fitness-Sets, eine Amplitudenvergrößerung zu konkaven Fitness-Sets (Levins 1968). Dies setzt jedoch eine lineare Umsetzung der Umweltamplitudenveränderung in Fitnessverschiebung für alle Genotypen voraus. Ein derartiger Fall ist in Abb. 16 dargestellt.

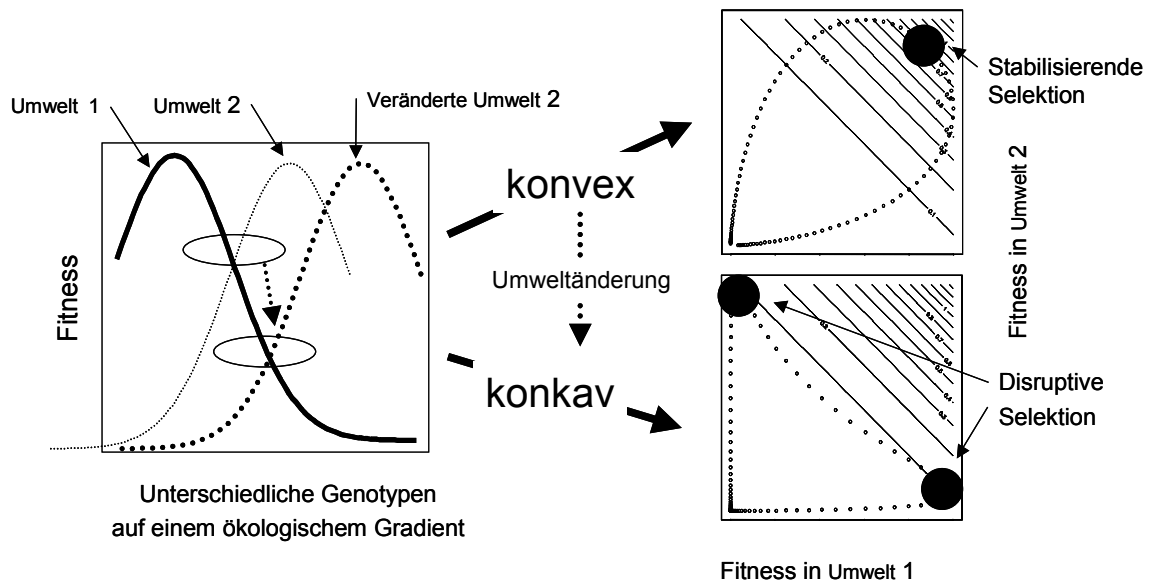


Abb. 16 Erhöhung der Schwankungsamplitude (z.B. wärmere Sommer) führt zu konkaven Fitness-Sets und tendenziell zur disruptiven Selektion (Levins 1968).

Nun lassen sich in geologischen Zeitmaßstäben regelmäßige Warm- und Kaltzeiten aufgrund planetarischer Umlaufzeiten, z.B. die sog. Milankovitch-Zyklen, nachweisen (Willis et al. 1999). Unter diesen Veränderungen der Amplitude müssten beim Übergang von kleinen zu großen Amplituden Artaufspaltung durch disruptive Selektion wahrscheinlicher gewesen sein als im umgekehrten Fall.

Die Regelmäßigkeit der Umweltschwankungen erscheint zunächst als sehr rigoreose Einschränkung. Betrachtet man jedoch die induzierte Unschärfen z.B. durch die Populationsgröße, so kann man erhoffen, dass die Genauigkeit der Schwankungsperiodizität die Abbildbarkeitsgenauigkeit nicht überschreitet. Auch zufällige Ereignisse sind durch die Kommutativität von Multiplikation und Addition z.T. abbildbar. Wann jedoch die Modellvoraussetzungen durch Zufälligkeit oder Ungenauigkeit der Umweltschwankungen zu stark verletzt sind, wurde hier nicht untersucht.

7 Zusammenfassung

Das Quantitative Fitness-Set Modell (QFM) ist eine Erweiterung der Fitness-Set Theorie (Levins 1968). Das QFM kann Abstufungen zwischen grob- und feinkörnigen regelmäßigen Schwankungen zweier Umwelten darstellen. Umwelt- und artspezifische Parameter, sowie die bewirkte Körnigkeit, sind quantifizierbar (Tabelle 1).

Experimentelle Daten lassen sich analysieren, und das QFM erweist sich in großen Populationen als sehr genau (Kapitel 2), was durch den diskreten Parameterraum unterstützt wird. Kleine Populationen und/oder hohe genetische Diversität (Kapitel 3) führen zu Schätzungenungenauigkeiten, die auch in natürlichen Populationen zu erwarten sind. Ein populationsgrößenabhängiger Unschärfewert erweitert die Punktschätzung eines Parametersatzes auf eine Intervallschätzung.

Diese Intervalle wirken in finiten Populationen als Fitnessbänder, d.h. Bereiche ähnlicher Fitness. Daraus ergibt sich die Hypothese, dass bei Arten, die in dichten kontinuierlichen Fitnessbändern leben, Generalisten und in diskreten Fitnessbändern Spezialisten evolvieren .

Die Evolution asynchroner Reproduktionsstrategien führt zur Bewahrung genetischer Diversität (Kapitel 5). Zudem ergibt sich aus dem Wechsel von grobkörniger zu feinkörniger Umweltvariation eine Bevorzugung der spezialisierter Genotypen. Aus diesem Angriffspunkt für disruptive Selektion lässt sich die Hypothese der vermehrten Artbildung in Übergangsszenarien von grobkörniger zu feinkörniger Umweltvariation formulieren. Dies ist somit eine prozessorientierte Erklärung für den Artenreichtum der (feinkörnigen) Tropen.

8 Literaturverzeichnis

- Aichele und Schwegler (1993). *Unsere Moos- und Farnpflanzen*, Kosmos Naturführer.
- Allendorf, F. W. und R. F. Leary (1987). Heterozygosity and fitness in natural populations of animals. *Conservation Biology*. M. E. Soulé, Sinauer, Sunderland, MA: 57-76.
- Barton, N. H. (2000). "Genetic hitchhiking." *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 355(1403): 1553-1562.
- Blarer, A. und M. Doebeli (1999). "Resonance effects and outbreaks in ecological time series." *Ecology Letters* 2(3): 167-177.
- Bronstein, I. N., K. A. Semendjajew, G. Musiol und H. Mühlig (1997). *Taschenbuch der Mathematik*. Frankfurt am Main, Verlag Harry Deutsch.
- Buri, P. (1956). "Gene frequency in small populations of mutant *Drosophila*." *Evolution* 10: 367-402.
- Caceres, C. E. (1997). "Temporal variation, dormancy, and coexistence: A field test of the storage effect." *Proceedings of the National Academy of Sciences of the United States of America* 94(17): 9171-9175.
- Chappell, M. A., J. P. Hayes und L. R. G. Snyder (1988). "Hemoglobin polymorphisms in deer mice (*Peromyscus maniculatus*): Physiology of beta-globin variants and alpha-globin recombinants." *Evolution* 42: 681-688.
- Connell, J. H. (1978). "Diversity in tropical rain forests and coral reefs." *Science* 199: 1302-1310.
- Cook, R. D. und D. L. Hartl (1974). "Uncorrelated random environments and their effects on gene frequency." *Evolution* 28: 265-274.
- Dobzhanski, T. (1957). "Genetic loads in natural populations." *Science* 126: 191-194.
- Du, W., J. Brown, D. Sylvester, J. Huang, A. Chalker, C. So, D. Holmes, D. Payne und N. Wallis (2000). "Two active forms of UDP-N-acetylglucosamine enolpyruvyl transferase in gram-positive bacteria." *Journal of Bacteriology* 182(15): 4146-4152.
- Ellner, S. und A. Sasaki (1996). "Patterns of genetic polymorphism maintained by fluctuating selection with overlapping generations." *Theoretical Population Biology* 50(1): 31-65.
- Franklin und Lewontin (1970). "Is the gene the unit of selection?" *Genetics* 65: 707-734.
- Gillespie, J. H. (1977). "A general model to account for enzyme variation in natural populations. III Multiple alleles." *Evolution* 31: 85-90.
- Gimelfarb, A. (1998). "Stable Equilibria in Multilocus genetic systems: A statistical investigation." *Theoretical Population Biology* 54: 133-145.

- Gottlieb, L. D. (1974). "Genetic stability in a peripheral isolate of *Stephanomeria exigua* ssp. *coronaria* that fluctuates in population size." *Genetics* 76: 551-556.
- Gurney, W. S. C., P. H. Crowley und R. M. Nisbet (1991). "Locking life-cycles on to seasons: circlemap models of population dynamics and local adaptation." *Journal of Mathematical Biology* 30: 251-279.
- Gurney, W. S. C., P. H. Crowley und R. M. Nisbet (1994). "Stage-specific quiescent as a mechanism for synchronizing life cycles to seasons." *Theoretical Population Biology* 46: 319-343.
- Hairston, N. G., C. L. Holtmeier, W. Lampert, L. J. Weider, D. M. Post, J. M. Fischer, C. E. Cáceres, J. A. Fox und U. Gaedke (2001). "Natural selection for grazer resistance to toxic cyanobacteria: Evolution of phenotypic plasticity?" *Evolution* 55(11): 2203-2214.
- Hartl, D. L. und A. G. Clark, Eds. (1989). *Principles of population genetics*. Sunderland, Sinauer Associates, Inc.
- Hartl, D. L. und R. D. Cook (1974). "Autocorrelated Random Environments and their effects on gene frequency." *Evolution* 28: 275-280.
- Kassen, B. und G. Bell (1998). "Experimental evolution in *Chlamydomonas*. IV. Selection in environments that vary through time at different scales." *Heredity* 80: 732-741.
- Kassen, R. (2002). "The experimental evolution of specialists, generalists, and the maintenance of diversity." *Journal of Evolutionary Biology* 15: 173-190.
- Kawecki, T. J. (2000). "The evolution of genetic canalization under fluctuating selection." *Evolution* 54(1): 1-12.
- Kimura, M. (1983). *The neutral theory of molecular Evolution*. Cambridge, Cambridge University Press.
- Levins, R. (1968). *Evolution in Changing Environments*. Princeton, New Jersey, Princeton University Press.
- Lewontin, R. C. (1971). "The effect of linkage on the mean fitness of a population." *Proceedings of the National Academy of Sciences of the United States of America* 68: 984-986.
- Müller, J. (1993). *Räumliche und Zeitliche Variabilität der genetischen Struktur natürlicher Cladocerenpopulationen (Crustacea:Cladocera)*. Fachbereich Biologie. Mainz, Johannes Gutenberg-Universität: 72.
- Powell, J. A. (2001). "Longest insect dormancy: *Yucca* moth larvae (Lepidoptera : Prodoxidae) metamorphose after 20, 25, and 30 years in diapause." *Annals of the*

- Entomological Society of America 94(5): 677-680.
- Rice, J. (1995). *Mathematical Statistics and Data Analysis*. Belmont, CA, Wadsworth.
- Roche (1999). Grippe - Erstzulassung eines neuartigen Grippemedikamentes in Pillenform - Tamiflu® von Roche, http://www.presseerklaerungen.de/presseerklaerungen/texte/pharma_medikamente/roche/roche_071099.html, Roche Deutschland Holding GmbH. 2002.
- StatSoft, I. (2001). *STATISTICA für Windows* [Software-System für Datenanalyse], www.statsoft.com.
- Stern, K. und P. M. A. Tigerstedt (1974). *Ökologische Genetik*. Stuttgart, Gustav Fischer Verlag.
- Tuljapurkar, S. (1990). *Population Dynamics in Variable Environments*. New York, Berlin, Heidelberg, London, Paris, Tokyo, Hong Kong, Springer Verlag.
- Turelli, M., D. W. Schemske und P. Bierzychudek (2001). "Stable two-allele polymorphisms maintained by fluctuating fitnesses and seed banks: Protecting the blues in *Linanthus parryae*." *Evolution* 55(7): 1283-1298.
- Weider, L. J. (1992). "Disturbance, competition and the maintenance of clonal diversity in *Daphnia pulex*." *Journal of Evolutionary Biology* 5: 505-522.
- Willis, K. J., A. Kleczkowski und S. J. Crowhurst (1999). "124,000-year periodicity in terrestrial vegetation change during the late Pliocene epoch." *Nature* 397(6721): 685-688.
- Winkle, S. (2002). *Die Kulturgeschichte der Seuche*. München, Komet Ma - Service und Verlagsgesellschaft.

9 Anhang

9.1 Detaillierte Herleitung der Fitness-Set-Theorie nach Levins 1968

Die Gesamtfitness in heterogener Umwelt hängt von der Fitness in den Einzelumwelten ab. Das Muster des Zusammenwirkens der Einzelumwelten bestimmt die Gesamtfitness in der heterogenen Umwelt. Daher definiert man die Adaptive Funktion $A(W_1, W_2)$ als eine monoton steigende Funktion ihrer Argumente (Fitness W_i), die die Überlebenswahrscheinlichkeit $P(t)$ der Tiere in der heterogenen Umwelt beschreibt.

Sei nun die Sterbewahrscheinlichkeit $m(t)$ für einen Zeitschritt Δt konstant, wobei $m(t)$ einem von zwei Werten aufgrund der gerade herrschenden Umwelt einnimmt. Somit genügt die Wahrscheinlichkeit $P(t)$, bis zu einem Zeitpunkt t überlebt zu haben, der Zuweisung

$$A(W_1, W_2) = P(t+\Delta t) = P(t) [1 - m(t) \Delta t] \quad (\text{Levins 1968 : 2.2})$$

Unterscheidet man nun die Fälle von grobkörniger Umwelt, in der ein Tier nur selten einen Wechsel der Umwelt erfährt, von feinkörniger Umwelt, in der die heterogene Umwelt als Mischung zweier Umwelten auftritt, so ergeben sich folgende Grenzverhalten der Überlebenswahrscheinlichkeit:

Grobkörnige Umwelt: Sei Δt eine ganze Generation

$$\rightarrow P(t) = \prod_{t=0}^{t_{\max}} (1 - m(t) \Delta t) \quad (\text{Levins 1968 : 2.3})$$

$$= \prod_{t=0}^{t_1} (1 - m_1 \Delta t) \quad \bullet \quad \prod_{t=t_1}^{t_{\max}} (1 - m_2 \Delta t)$$

Umwelt 1

Umwelt 2

$$= [1 - m_1 \Delta t]_1^t \quad \bullet \quad [1 - m_2 \Delta t]_{\max}^t \cdot 1^{-t}$$

Die Überlebenswahrscheinlichkeit entspricht der Fitness in der Umwelt. Also lassen sich den beiden obigen Faktoren (Fitness in der Umwelt) diese auch anteilmäßig zuordnen:

$$P(t) = W_1^p \bullet W_2^{(1-p)} \quad (\text{Levins 1968 : 2.3f})$$

Feinkörnige Umwelt: Ein Tier begegnet beiden Umwelten ständig, z.B. als verschiedene Nahrung, Tag/Nacht, etc., d.h. die Zeitschritte sind sehr klein. Als Grenzwert ergibt sich für $\Delta t \rightarrow 0$:

$$P(t+\Delta t) = P(t) [1 - m(t) \Delta t] = P(t) - P(t) m(t) \Delta t \quad | -P(t)$$

$$P(t+\Delta t) - P(t) = -P(t) m(t) \Delta t \quad | : \Delta t$$

$$\frac{P(t+\Delta t) - P(t)}{\Delta t} = -P(t) m(t) \quad \text{für } \Delta t \rightarrow 0 \text{ gilt nun}$$

$$dP(t)/dt = P'(t) = -P(t) m(t) \quad | : P(t)$$

$$\frac{P'(t)}{P(t)} = -m(t)$$

nun gilt (Substitution, Sonderfall)

$$\int \frac{P'(t)}{P(t)} dt = \ln P(t) + C = \int -m(t) dt \quad | -C ; \text{„exp“}$$

$$P(t) = e^{-C - \int m(t) dt} = e^{-C} e^{-\int m(t) dt} \quad \text{für } e^{-C} := P_0$$

$$P(t) = P_0 e^{-\int m(t) dt} \quad (\text{Levins 1968 : 2.4})$$

Dies ist maximal, wenn das Integral $\int m(t) dt$ am kleinsten ist. Dies ist dann der Fall, wenn das arithmetische Mittel von $p \cdot W_1$ und $(1-p) \cdot W_2$ am größten ist. Natürlich können auch mittelkörnige Umwelten auftreten. Dabei verläuft die Adaptive Funktion zwischen linearen und multiplikativen (bzw. logarithmischen) Graphen.

9.2 Motivation der Quantitativen Fitness-Set-Theorie

9.2.1 Notationsfehler im Originaltext

Levins führt im Original keine Indizes mit, wodurch ihm der Fehler unterläuft, in der Erläuterung der Gleichung (Levins 1968 2.3) von ganzzahligen Laufvariablen (t_i) auf Verhältnisse zu schließen. Er setzt somit

$$[1 - m_1 \Delta t]^{t_1} \cdot [1 - m_2 \Delta t]^{t_{\max} - t_1} = W_1^p \cdot W_2^{(1-p)} \quad (9.2.1)$$

und $t_1 = p$ und $t_{\max} - t_1 = (1-p)$. Dies ist offensichtlich keine wahre Aussage, denn

$$t_i \in \mathbb{N}^+ \text{ und } p \in [0..1]. \quad (9.2.2)$$

9.2.2 Extremfallbetrachtung

Levins verweist darauf, dass neben fein- und grobkörnige Umwelten auch mittelkörnige existieren können ohne diesen (allgemeinen) Fall zu formulieren. Genau dieser Entscheidung liegt jedoch eine subjektive Bewertung zugrunde, die "aus der Sicht des Tieres" getroffen werden muss und eine wichtige Information für die quantitative Anwendung der Fitness-Set-Theorie bildet.

9.2.3 Nicht explizite Definition von Fitness

Levins definiert die Fitness eines Individuums als die Wahrscheinlichkeit $P(t)$ den Zeitraum t bis $t+\Delta t$ zu überleben. Dieses ist eine Negativdefinition, die durch das Gegenereignis, die Wahrscheinlichkeit zu sterben $m(t)$ im Zeitraum Δt bestimmt ist. Daher beruht seine Herleitung implizit auf der Annahme, dass Fitness immer in $[0,1]$ liegt, da sich ansonsten ein anderer graphischer Verlauf als der angegebene bilden kann. Für Anwendungen, bei denen oft Reproduktionsraten als Fitnesswerte genutzt werden, treten u.U. negative Werte auf. Daher ist die o.g. Nebenbedingung wichtig.

9.3 Beweis der Konstanz der Gesamtfitnessnormierung

Sei \vec{f}_r^n der normierte relative Fitnessvektor (über alle Genotypen) zwischen den Generationen n und $n-1$ und damit $\max(\vec{f}_r^n) = 1$. Es ergibt sich die Gesamtfitness \vec{G}^n aus

$$\overrightarrow{G}^n = \frac{\overrightarrow{f_r^{n-1}} \bullet \overrightarrow{f_r^n}}{\max(\overrightarrow{f_r^{n-1}} \bullet \overrightarrow{f_r^n})} \quad (9.3.1)$$

$$\text{Beh.: } \overrightarrow{G}^n = \frac{\overrightarrow{f_r^{n-1}} \bullet \overrightarrow{f_r^n}}{\max(\overrightarrow{f_r^{n-1}} \bullet \overrightarrow{f_r^n})} = \frac{\prod_{i=1}^n \overrightarrow{f_r^i}}{\max(\prod_{i=1}^n \overrightarrow{f_r^i})} \quad (9.3.2)$$

Bew.: (Vollständige Induktion)

$$\overrightarrow{G}^1 = \overrightarrow{f_r^1} \quad (9.3.3)$$

$$\overrightarrow{G}^2 = \frac{\overrightarrow{f_r^1} \bullet \overrightarrow{f_r^2}}{\max(\overrightarrow{f_r^1} \bullet \overrightarrow{f_r^2})} \quad (9.3.4)$$

$$\overrightarrow{G}^3 = \frac{\overrightarrow{G}^2 \bullet \overrightarrow{f_r^3}}{\max(\overrightarrow{G}^2 \bullet \overrightarrow{f_r^3})} = \frac{\frac{\overrightarrow{f_r^1} \bullet \overrightarrow{f_r^2}}{\max(\overrightarrow{f_r^1} \bullet \overrightarrow{f_r^2})} \bullet \overrightarrow{f_r^3}}{\max\left(\frac{\overrightarrow{f_r^1} \bullet \overrightarrow{f_r^2}}{\max(\overrightarrow{f_r^1} \bullet \overrightarrow{f_r^2})} \bullet \overrightarrow{f_r^3}\right)} \stackrel{\text{in Produktschreibweise}}{=} \frac{\frac{\prod_{i=1}^3 \overrightarrow{f_r^i}}{\max\left(\prod_{i=1}^2 \overrightarrow{f_r^i}\right)}}{\max\left(\frac{\prod_{i=1}^3 \overrightarrow{f_r^i}}{\max\left(\prod_{i=1}^2 \overrightarrow{f_r^i}\right)}\right)} \quad (9.3.5)$$

⋮

$$\overrightarrow{G}^m = \frac{\frac{\prod_{i=1}^m \overrightarrow{f_r^i}}{\max\left(\prod_{i=1}^{m-1} \overrightarrow{f_r^i}\right)}}{\max\left(\frac{\prod_{i=1}^m \overrightarrow{f_r^i}}{\max\left(\prod_{i=1}^{m-1} \overrightarrow{f_r^i}\right)}\right)} \stackrel{\text{sei } \max\left(\prod_{i=1}^{m-1} \overrightarrow{f_r^i}\right) = c_o}{=} \frac{\frac{\prod_{i=1}^m \overrightarrow{f_r^i}}{c_{m-1}}}{\max\left(\frac{\prod_{i=1}^m \overrightarrow{f_r^i}}{c_{m-1}}\right)} = \frac{1}{c_{m-1}} \bullet \frac{\prod_{i=1}^m \overrightarrow{f_r^i}}{\max\left(\prod_{i=1}^{m-1} \overrightarrow{f_r^i}\right)} \quad (9.3.6)$$

$$= \frac{\prod_{i=1}^m \overrightarrow{f_r^i}}{\max\left(\prod_{i=1}^m \overrightarrow{f_r^i}\right)} \quad \text{q.e.d} \quad (9.3.7)$$

9.4 Rückführung des QFM auf die Ergebnisse von Levins (1968)

9.4.1 Für grobkörnige Umwelt

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \prod_{i=1}^n W(r(i, \tau_1, \tau_2), \tau_1, \tau_2, \lambda, f_1, f_2) \quad (2.3)$$

für grobkörnige Umwelt gilt $\tau_1=1$,

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \prod_{i=1}^n W(r(i, 1, \tau_2), 1, \tau_2, \lambda, f_1, f_2) \quad (9.4.1.1)$$

$$= \prod_{i=1}^n W([(i-1) \bmod \tau_2], 1, \tau_2, \lambda, f_1, f_2) \quad (9.4.1.2)$$

$$= \prod_{i=1}^n \left[\frac{1}{1} \sum_{x=[(i-1) \bmod \tau_2]}^{[(i-1) \bmod \tau_2]+1-1} w(x, \tau_2, \lambda, f_1, f_2) \right] \quad (9.4.1.3)$$

$$= \prod_{i=1}^n [w([(i-1) \bmod \tau_2], \tau_2, \lambda, f_1, f_2)] \quad (9.4.1.4)$$

Unter der Annahme, dass $(n \cdot \tau_1) \bmod \tau_2 = 0$, bzw. irrelevant klein ist (für große n und kleine τ_2 wohl gültig), dann lässt sich $m = (n \cdot \tau_1) \div \tau_2$ definieren als die Anzahl der Perioden der Umweltozillation im Gesamtzeitraum. Damit lässt sich das Produkt darstellen als:

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \left(\prod_{i=1}^{\tau_2} [w([(i-1) \bmod \tau_2], \tau_2, \lambda, f_1, f_2)] \right)^m \quad (9.4.1.5)$$

und die beiden Umwelten lassen sich mathematisch trennen in:

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \left(\prod_{i=1}^{\lambda} [w([(i-1) \bmod \tau_2], \tau_2, \lambda, f_1, f_2)] \cdot \prod_{i=\lambda+1}^{\tau_2} [w([(i-1) \bmod \tau_2], \tau_2, \lambda, f_1, f_2)] \right)^m \quad (9.4.1.6)$$

Für w gilt nun

$$w([(i-1) \bmod \tau_2], \tau_2, \lambda, f_1, f_2) = \begin{cases} f_1 \forall [(i-1) \bmod \tau_2] \bmod \tau_2 \in [0.. \lambda[& (9.4.1.7) \\ f_2 \forall [(i-1) \bmod \tau_2] \bmod \tau_2 \in [\lambda.. \tau_2[& \end{cases}$$

$$= \begin{cases} f_1 \forall [(i-1) \bmod \tau_2] \in [0.. \lambda[& (9.4.1.8) \\ f_2 \forall [(i-1) \bmod \tau_2] \in [\lambda.. \tau_2[& \end{cases}$$

hierbei fällt der Doppel Modulo weg, da $(a \bmod b) \bmod b = (a \bmod b)$ ist. Nun gilt für alle Werte aus $[0.. \lambda[$, dass $w=f_1$ und aus $[\lambda.. \tau_2[$, dass $w=f_2$, und die Gesamtfitnessausdruck

reduziert sich im grobkörnigen Fall zu

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \left(\prod_{k=0}^{\lambda} [f_1] \cdot \prod_{o=\lambda}^{\tau_2-1} [f_2] \right)^m = ([f_1]^\lambda \cdot [f_2]^{\tau_2-\lambda})^m = f_1^{\lambda \cdot m} \cdot f_2^{(\tau_2-\lambda)m} \quad (9.4.1.9)$$

$$\text{für } (m \bmod \tau_2) = 0 \text{ gilt } g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = f_1^{\frac{\lambda \cdot n}{\tau_2}} \cdot f_2^{\frac{(\tau_2-\lambda)n}{\tau_2}} = \left(f_1^{\frac{\lambda}{\tau_2}} \cdot f_2^{\frac{(\tau_2-\lambda)}{\tau_2}} \right)^n \quad (9.4.1.10)$$

Eine analoge Formel findet sich im Originaltext.

9.4.2 Für feinkörnige Umwelt

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \prod_{i=1}^n W(r(i, \tau_1, \tau_2), \tau_1, \tau_2, \lambda, f_1, f_2) \quad (2.3)$$

Der Zeitschritt, in der die Umweltszillation stattfindet, soll nun nach dem Originaltext „unendlich klein“ werden. Das bedeutet im QFM, dass τ_2 „unendlich klein“ wird. Diesen Zustand nähern wir, da τ_2 nur einen Minimalwert von 1 annehmen kann, durch ein sehr (unendlich) großes τ_1 an. Man definiert ähnlich zum grobkörnigen Fall $m = (\tau_1 \text{ div } \tau_2)$ als die Anzahl der vollständigen Perioden der Umweltszillationen im Zeitraum τ_1 . Zusätzlich definieren einen Rest R , der den nichtvollständigen und variierenden Anteil der Umweltszillation ausmacht:

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \prod_{i=1}^n \frac{m[\lambda f_1 + (\tau_2 - \lambda)f_2] + R(i)}{\tau_1} \quad (9.4.2.1)$$

müssen nun zeigen, dass für $\tau_1 \rightarrow \infty \Rightarrow R(i) \rightarrow 0$ und damit als gleichbleibende realisierte Fitness das Verhältnis der beiden Umwelten zueinander einstellt:

$$\frac{m[\lambda f_1 + (\tau_2 - \lambda)f_2] + R(i)}{\tau_1} = \frac{m[\lambda f_1 + (\tau_2 - \lambda)f_2]}{\tau_1} + \frac{R(i)}{\tau_1} \quad (9.4.2.2)$$

A) Wir wissen: $R(i)$ ist definiert als die Summe der Fitnesswerte, die keine vollständige Umweltperiode ausmachen. Damit ist $R(i)$ immer kleiner als die vollständige Summe innerhalb einer Umweltperiode. Diese beträgt, da jede der τ_2 – verschiedenen Motive einmal vorkommt (und damit das Grundmotiv τ_2 – mal vorkommt):

$$\tau_2[\lambda f_1 + (\tau_2 - \lambda)f_2] = R_{\max} \quad (9.4.2.3)$$

Es gilt nun für ein beliebiges, aber festes R_{\max} , dass $\lim_{\tau_1 \rightarrow \infty} \left(\frac{R_{\max}}{\tau_1} \right) = 0$

Damit ist gezeigt, dass die Faktoren des Produktes konstant sind.

B) Für diesen konstanten Faktor zeigt sich nun nach folgenden Ansatz

$$Faktor = \frac{m[\lambda f_1 + (\tau_2 - \lambda)f_2]}{\tau_1} = \frac{[\lambda f_1 + (\tau_2 - \lambda)f_2]}{\tau_2} = \text{Mittelwert des Umweltmotivs} \quad (9.4.2.4)$$

dass folgender Beweis zu führen ist:

$$\text{Beh.: } \frac{m}{\tau_1} = \frac{1}{\tau_2} \quad (9.4.2.5)$$

Bew:

$$\frac{m}{\tau_1} = \frac{1}{\tau_2} \quad | \text{ für } m = (\tau_1 \text{ div } \tau_2) \quad (9.4.2.6)$$

$$\frac{(\tau_1 \text{ div } \tau_2)}{\tau_1} = \frac{1}{\tau_2} \quad | * \tau_1 \quad (9.4.2.7)$$

$$\tau_1 \text{ div } \tau_2 = \frac{\tau_1}{\tau_2} \quad (9.4.2.8)$$

Diese Aussage ist für große τ_1 gültig. Da dies unserer Annahmen ist, gilt auch der Ansatz.

Damit lässt sich das Produkt zu einer Potenz vereinfachen, da die realisierten Fitnesses konstant bleiben und die Gesamtfitness bildet sich aus

$$g(n, \tau_1, \tau_2, \lambda, f_1, f_2) = \left(\frac{\lambda f_1 + (\tau_2 - \lambda)f_2}{\tau_2} \right)^n \quad (9.4.2.9)$$

Also bildet sich das Verhältnis

$$\frac{\lambda}{\tau_2 - \lambda} = \frac{\lambda}{\tau_2} \cdot \frac{\tau_2 - \lambda}{\tau_2} = \frac{\lambda}{\tau_2} \cdot \left(1 - \frac{\lambda}{\tau_2} \right) \xrightarrow{\text{für } \frac{\lambda}{\tau_2} = p} \frac{p}{(1-p)} \quad (\text{in Levinscher Notation}) \quad (9.4.2.10)$$

9.5 Herleitung der Reihenfolgefunktion

Seien τ_1 und τ_2 Zeitdauern, wobei τ_1 die Fertilitätsdauer einer Tierart und τ_2 die Periode einer Umweltschwankung darstellt. Im Rahmen numerischer Genauigkeit haben beide ein rationales Verhältnis:

$$\tau_1 : \tau_2 = m_1 : m_2, \quad (9.5.1)$$

mit m_1, m_2 teilerfremde natürliche Zahlen. Der Fall irrationaler Verhältnisse wird für sehr große m_1, m_2 angenähert. Es ergibt sich eine größte gemeinsame Schrittlänge

$$\tau = \tau_1/m_1 = \tau_2/m_2. \quad (9.5.2)$$

Im Folgenden setze ich voraus, dass die Umweltbedingungen dadurch gegeben sind, dass in jedem τ -Schritt in τ_2 diese Bedingungen konstant sind, entweder Niveau 1 oder Niveau 2. Weiter ergibt sich eine kleinste gemeinsame Periode

$$\tau_1 * m_2 = \tau_2 / m_1. \quad (9.5.3)$$

Sei o. B. d. A. $\tau_1 < \tau_2$. Wenn die gemeinsame Periode mit τ_1 ausgeschöpft wird, kommt jeder Schritt der Periode τ_2 genau einmal als erster Schritt in τ_1 vor (τ_1 wird m_2 -mal benötigt um die gemeinsame Periode auszuschöpfen). Wenn der i -te Schritt von τ_2 dabei ein zweites Mal als erster Schritt in τ_1 auftritt, müsste sich von da ab alles wiederholen, im Widerspruch zu Obigem. Die Reihenfolge, in der die Schritte von τ_2 als erste Schritte in τ_1 auftreten (Motiv), ergibt sich dabei aus:

$$((k-1) * \tau_1 \bmod \tau_2) \quad (9.5.4)$$

Tabelle 10 Beispiele der Motivabfolge für $\tau_1 = 5, \tau_2 = 7$.

k	1	2	3	4	5	6	7	8	9	10	11	12	13
Motiv	0	5	3	1	6	4	2	0	5	3	1	6	4

An dieser Stelle möchte ich Herrn Prof. Geert Zimmermann danken, der sich die Mühe gemacht hat, diese Herleitung für den diskreten Fall zu entwickeln.

9.6 Funktionen zur grafischen Darstellung der Fitness-Isoklinen in „R“

Das Statistik-Paket „R“ ist die unter der GNU-Public-License frei erhältliche Version der Statistik Programmiersprache „S“. Durch die einfache Programmierbarkeit und umfangreiche Statistik-Zusatzpakete nutzte ich diese zur grafischen Darstellung der Fitness-Isoklinenverläufe.

Die Schlüsselwörter der Programmiersprache und Standard-Routinen von R, bzw. allgemein verfügbarer Programmpakete sind in Arial gehalten. Eigene Variablen und Parameter selbstdefinierter Funktionen sind in Kapitalchen gehalten. Selbstdefinierte Funktionen sind kursiv geschrieben. Als geeignete Darstellung der Urdaten erweist sich die Funktion „scatterplot3d“ (s. in Abb. 5). Die Packages „spatial“ und „scatterplot3d“ müssen geladen werden. Die Raute (#) leitet einen Kommentar ein.

1) Funktion „fitnessplot“

Ziel: Diese Funktion übernimmt drei Vektoren fit1, fit2, fittot. Diese drei Vektoren spannen die Fitnessfunktion auf (fit1 für Umwelt 1, fit 2 für Umwelt 2, fittot für die Gesamtfitness). Darüber wird eine Fläche gelegt, die als Repräsentation der Fitnessfunktion interpretiert wird. Im Text finden sich verschiedene Abbildungstypen: Die Darstellung der Fitnessisoklinen in Graustufen (Version 1) ist in Abb. 1 gezeigt, als Contourlinien (Version 2) wurde in Abb. 15 eingesetzt. Auch die Darstellung als 3-dimensionales Gitter ist möglich. (s. in Abb. 5). Die drei Funktionen sind hier in eine Funktion zusammengefasst und jeweils auskommentiert.

Variablen:

GITTER: enthält die (x,y,z)-Werte mit äquidistanten (gefitteten) Werten zur grafischen Darstellung (s. „*machegitter*“)

col: Parameter, enthält die Farbabstufungen der Fitnesswerte

Algorithmus:

```
function (FIT1,FIT2,FITTOT) {  
  GITTER <- machegitter(FIT1,FIT2,FITTOT)  
  # Version 1 : male fitnessisoklinen  
  filled.contour(GITTER, col = heat.colors(20)) # Ausgefüllte Fitnessflächen  
  # Version 2: Nur Fitnessisoklinen, nicht ausgefüllt  
  # contour(GITTER, col = heat.colors(20))  
  # Version 3: 3-dimensionale Funktionsdarstellung, Gitternetz  
  # persp(GITTER, zlim=c(0,1),axes=TRUE)  
}
```

2) Funktion „machegitter“

Ziel: Die Funktion extrapoliert aus den beliebig verstreuten Fitnesswerten (fit1, fit2, fittot) äquidistante Gitterpunkte aus denen Standard-Grafikroutinen aus „R“ dreidimensionale Grafiken erzeugen können.

Variablen:

lsqmodellfit enthält das gefittete Modell (4-dimensionales Polynom)

gitter: enthält die (x,y,z)-Werte mit äquidistanten (gefitteten) Werten zur grafischen Darstellung, entsprechend den Parametern der Standard-Routine tmat.

Algorithmus:

```
function (FIT1,FIT2,FITTOT) {
  #normiere fitnessvektoren, so daß max(fitness)=1
  FIT1 <- FIT1 /max(FIT1)
  FIT2 <- FIT2 /max(FIT2)
  FITTOT<- FITTOT /max(FITTOT)
  # fitte ein sehr variables modell darüber :)
  LSQMODELLFIT <- surf.ls(4,FIT1,FIT2,FITTOT)
  # generiere aus dem modell äquidistante gitterpunkte
  GITTER <- trmat(LSQMODELLFIT,0,1,0,1,25)
  GITTER # Ausgabe der Gitterpunkte
}
```

3) QFM-Implementation (Erläuterung der Parameter s. Kapitel 1)

Aktuelle fitness „w“:

```
function(X,TAU2,LAMBDA, FU1,FU2) {
  COND <- ( X %% TAU2) < LAMBDA)
  if (COND) ERG <- FU1
  else ERG <- FU2
  ERG
}
```

Realisierte Fitness „W“

```
function(TIMEST,TAU1,TAU2,LAMBDA,FU1,FU2) {
  SUMME <- rep(0,length(FU1)) # Nullsetzen des Vektors
  for (i in TIMEST:(TIMEST+(TAU1-1)))
  { SUMME <- SUMME+ w(i,TAU2,L,FU1,FU2)
    # summe der fitness zum Zeitpunkt t
  }
  SUMME <- SUMME/TAU1 ## Mittelwert bilden
  SUMME
}
```

Funktion „reihenfolge“

```
function(i,TAU1,TAU2) {
  POSITION <- ((i-1) * TAU1) %% TAU2
  POSITION
}
```

Gesamtfitness „g“

```
function(N,TAU1,TAU2,LAMBDA,FU1,FU2) {
  PRODUKT <- rep(1,length(FU1))
  FU1 <- FU1/max(FU1)
  FU2 <- FU2/max(FU2)
  for (i in 1:N)
  { POSITION <- reihenfolge(i,TAU1,TAU2)
    PRODUKT <- PRODUKT*W(POSITION,TAU1,TAU2,LAMBDA,FU1,FU2)
  }
  PRODUKT <- PRODUKT/max(PRODUKT)
  PRODUKT
}
```

9.7 Module und Kernprozeduren aus dem QFM-Analyseprogramm

9.7.1 Datenstrukturen

Die verwendeten Datenstrukturen sind im hier abgedruckten Datentypmodul „Datenstrukturen“ deklariert. Konstantendeklarationen sind, sofern nicht selbsterklärend, mit `c<Bezeichner>`, Typendeklarationen mit `t<Bezeichner>` gekennzeichnet. Einige auf die Datentypen bezogene Funktionen (z.B. `initpop`, `ParToStr`), die in verschiedenen Modulen genutzt werden, wurden zugefügt.

Kernelemente sind

- „tfs“ : beschreibt einen Fitness-Set
- `tfitlist` : enthält die Parameterkombinationen, den Anpassungsfehler (wenn ein „fitting“ vorgenommen wurde), Anzahl der Nebenlösungen und ein Ähnlichkeitsparameter, der beim Sortieren der Lösungen genutzt wird.

```
unit Datenstrukturen;

interface

const tab=chr(9);
      cmaxmotive=100000; // Maximale Arraygrösse für Parameterkombinationen

type fitness=double;
      tparameter=record
          n, // Anzahl an Generationen
          tau1, // Fertilitätsdauer
          tau2, // Dauer der gesamten Umweltperiode
          l:integer; // Dauer der Umwelt 1 ( $\rightarrow \lambda$ )
      end;
      tgtfit =record // Fitnesskombination eines Genotyps
          f1, // fitness in Umwelt 1
          f2, // fitness in Umwelt 2
          G :fitness; // Gesamtfitness
      end;

      tfs = array of tgtfit; // enthält einen Fitness-Set, d.h. für jeden
                             // Genotyp die Fitnesses in zwei Umwelten
      tfitresult= record
          gof: double; // Anpassungsfehler
          par: tparameter; // Parameterkombination
          anz:integer; // Anzahl "gleicher" Parameterkombinationen
                       //  $\rightarrow$  Nebenlösungen
          aehnlich: double; // Ähnlichkeitsmaß für Parameter
      end;

      tfitlist= array of tfitresult;

      tsyntax = array[0..cmaxmotive] of record
          s : string;
      end;
          // Farbsyntax der Umweltmotive für
          // graphische Darstellungen, z.B. Abb. 3 und Abb. 4

      tfselement=record
          p:tparameter;
          anz:integer;
```

```

        end;

        tpop=array of double;

        tfslist  =array of tfselement;
        tstringarr= array of string;          //

function issamedouble(a,b:double):boolean;
// True wenn die Differenz < 1E-10 ist, sonst sind die Zahlen verschieden (False)

function fitresultttostr(res:Tfitresult):string;
// Gibt die Parameter, Anzahl Nebenlösungen, Anpassungsfehler etc. als String aus
// Für Ausgaben in Memo-Feldern, Ergebnisdateien, ...

function ParToStr(p:tparameter):string;
// Ausgabe der reinen Parameterwerte als String

function ParameterToWert(p:tparameter):double;
// Zum Sortieren wird der Parameterkombination ein Wert zugewiesen

procedure initpop(var pop:Tpop; len:integer);
// Initialisiert die Population mit 0

function pvergleich(p1:tparameter; p2:tparameter):boolean;
// vergleicht zwei Parameterkombinationen im Sinne: Ist p2 element der "menge mit
// maximum p1"? Wird zum Erzeugen einer „Unter-Hauptlösungsmenge“ aus einer vorher
// erzeugten grossen Hauptlösungsmenge verwendet

implementation

uses sysutils;

function pvergleich(p1:tparameter; p2:tparameter):boolean;
var ergebnis:boolean;
begin
    if p1.n<p2.n
    then ergebnis:=false
    else if p1.taul<p2.taul
        then ergebnis:=false
        else if p1.tau2<p2.tau2
            then ergebnis:=false
            else if p1.l<p2.l
                then ergebnis:=false
                else ergebnis:=true;
    pvergleich:=ergebnis;
end;

function issamedouble(a,b:double):boolean;
const epsilon=0.0000000001;
var dif:double;
    erg:boolean;
begin
    dif:=a-b;
    if abs(dif) < epsilon
    then erg:=true
    else erg:=false;
    result := erg;
end;

function fitresultttostr(res:Tfitresult):string;
var s:string;
begin
    str(res.gof:1:5,s);
    s:=
        inttostr(res.par.n)   + tab
        + inttostr(res.par.taul)+ tab
        + inttostr(res.par.tau2)+ tab
        + inttostr(res.par.l)   + tab
        + inttostr(res.anz)    + tab
        + s;
    result:=s;
end;

function parametertowert(p:tparameter):double;
begin
    result:=p.l*1.0 + p.tau2*100.0 + p.taul*10000.0 + p.n*1000000.0;
end;

function ParToStr(p:tparameter):string;
begin

```

```

    result:= inttostr(p.n*1000000+p.tau1*10000
                    +p.tau2*100+p.l*1);
end;

procedure initpop(var pop:Tpop; len:integer);
var i:integer;
begin
    finalize(pop);
    setlength(pop, len);
    for i:=0 to len-1
    do pop[i]:=0;
end;
end.

```

9.7.2 Implementation des QFM

Das QFM ist im Funktionsmodul „fsmodel“ implementiert. Die einzelnen Fitnesssebenen sind zugreifbar.

```

unit fsmodel;

interface
    uses Datenstrukturen;

function Gesamtfitness(p:tparameter; f:tgfit): fitness;
    {Gesamtfitness über mehrere Generationen realisierter Fitness → Gleichung 2.4
    Rückgabewert -> Fitness → [0;1]
    Eingabewerte
    p enthält
        n: Anzahl der Generationen,
        tau1: Dauer der Fertilitätsperiode,
        tau2: Dauer der Umweltperiode,
        l: Dauer der ersten Umwelt,
    f enthält:
        f1: fitness in erster Umwelt,
        f2: fitness in zweiter Umwelt}

function reihenfolge(i:integer; tau1:integer; tau2:integer):integer;
    {Reihenfolgefunktion: Die Abfolge der realisierten Fitnesswerte kann
    wichtig sein ! → Gleichung 2.5
    Rückgabewert: Einstiegszeitpunkt im Intervall 0..tau2-1
    Eingabewerte:
        i: Zeitpunkt
        tau1: Dauer der Fertilitätsperiode,
        tau2: Dauer der Umweltperiode}

function realisiertefitness(t:integer;tau1:integer;tau2:integer;l:integer;
                           f1:fitness;f2:fitness): fitness;
    {fitnessfunktion für realisierte Fitness: Gibt die mittlere Fitness zwischen
    einem Zeitpunkt x und x+tau1 zurück → Gleichung 2.3
    Rückgabewert -> [0;1]
    t: Anfangszeitpunkt,
    tau1: Dauer der Fertilitätsperiode,
    tau2: Dauer der Umweltperiode,
    l: Dauer der ersten Umwelt,
    f1: fitness in erster Umwelt,
    f2: fitness in zweiter Umwelt}

function w(x:integer;tau2:integer;l:integer;f1:fitness;f2:fitness): fitness;
    {Rückgabe: aktuelle Fitness zum Zeitpunkt x (Gleichung 2.1)
    Rückgabewert -> [0;1]
    t: Anfangszeitpunkt,
    tau1: Dauer der Fertilitätsperiode,
    tau2: Dauer der Umweltperiode,
    l: Dauer der ersten Umwelt,
    f1: fitness in erster Umwelt,
    f2: fitness in zweiter Umwelt}

implementation

function w(x:integer;tau2:integer;l:integer;f1:fitness;f2:fitness): fitness;

```

```

var z: integer; // zeitpunkt als projektion in 0 - tau2
    ergebnis:double;
Begin
    z := x mod tau2;
    if z < 1
    then ergebnis:=f1
    else ergebnis:=f2;
    w:=ergebnis;
end;

function realisiertefitness(t:integer;taul:integer;tau2:integer;l:integer;
                           f1:fitness;f2:fitness): fitness;
// Gleichung 2.2
var x:integer; // Zeitpunkt für w
    summe:fitness;
begin
    summe:=0;
    for x:=t to t+taul-1 // Mittelwert von bis
    do summe:=summe+w(x,taul,tau2,l,f1,f2);
    realisiertefitness:=summe/taul
end;

function reihenfolge(i:integer; taul:integer; tau2:integer):integer;
// Gleichung 2.4 (Reihenfolgefunktion)
begin
    reihenfolge:=(i-1)*taul mod tau2;
end;

function Gesamtfitness(p:tparameter; f:tgtfit): fitness;
// Gleichung 2.3
var i:integer; // Laufvariable für die Generationen
    produkt: fitness;
begin
    produkt:=1;
    with p
    do with f
    do begin
        for i:=1 to n // alle generationen
        do produkt:=produkt*realisiertefitness(reihenfolge(i,taul,tau2),
                                                taul,tau2,l,f1,f2);
        end;
    Gesamtfitness:=produkt;
end;
end.

```

9.7.3 Ermitteln der Parameterkombination, „Fitting“

Das Funktionsmodul „fsfit“ stellt die Funktionen „bestekombination“ und „gebebestekombination“ zur Verfügung. Beide Funktionen ermitteln die Parameterkombination, die den Fitness-Set am besten erklären. Die Prozedur „bestekombination“ sortiert die gesamte Parameterliste nach der Größe des Anpassungsfehlers. Bei Verwenden der Prozedur ist somit die Ausgabe ähnlicher Parameterkombinationen vereinfacht. Die Funktion „gebebestekombination“ sortiert nicht, sondern gibt nur die Lösung zurück. Die Funktion ist somit schneller.

```

unit fsfit;

interface

uses Datenstrukturen, fsmodel, SysUtils, math;

procedure bestekombination(fs:tfs; var fitliste:tfitlist);
{ Berechnet alle gegebenen Parametermöglichkeiten und sortiert nach
  bestem passenden Eingabeparameter:
  fitliste: enthält im record „.par“ die zu fittenden Parameterkombinationen
  fs: Fitness-Set der Population
  Ausgabeparameter
  fitlist: Parameterkombinationen, nach Anpassungsfehlergrösse sortiert}

function gebebestekombination(fs:tfs; var fitliste:tfitlist):Tfitresult;

```

```
{ Algorithmisch analog zu „bestekombination“, nur sortiert es die fitliste nicht,
sondern gibt als Ergebnis den Parametersatz mit kleinstem Anpassungsfehler
zurück}
```

implementation

```
function fitallgt(fs:tfs; p:tparameter):double;
var gt:integer;
    gof:double;
    maxG:fitness;
    G : array of fitness;
begin //fitallgt
    maxG:=0;
    setlength(G, length(fs)); // Zum Merken der Gesamtfitnesswerte
    for gt:=0 to length(fs)-1 // Berechne alle Gesamtfitnesses
    do begin
        G[gt]:= Gesamtfitness(p,fs[gt]); // und merke diese für das Fitting
        maxG := max(maxG,G[gt]); // suche größte Fitness zum Normieren
    end;

    gof:=0; // goodness of fit initialisieren
    for gt:=0 to length(fs)-1 // berechne fit über alle GT
    do gof:=gof + abs(fs[gt].G - G[gt]/maxG); // Differenz Normierter Fitnesses
    gof:=gof/length(fs); // berechne Mittlere absolute Abweichung
    if issamedouble(gof,0) then gof:=0;
    result:=gof;
end;

procedure bestekombination(fs:tfs; var fitliste:tfitlist);
var i:integer;

    procedure QSort(Links, Rechts : LongInt; var table:Tfitlist);
    var
        i, j : LongInt;
        x, w : tfitresult;
    begin
        i := Links;
        j := Rechts;
        x := Table[(Links + Rechts) div 2];
        repeat
            while Table[i].gof < x.gof do Inc(i);
            while Table[j].gof > x.gof do Dec(j);
            if i <= j
            then begin
                w := Table[i]; // austauschen
                Table[i] := Table[j];
                Table[j] := w;
                Inc(i);
                Dec(j);
            end;
        until i > j;
        if Links < j then QSort(Links,j,table);
        if Rechts > i then QSort(i,Rechts,table);
    end;

begin
    // für jede parameterkombi berechne goodness of fit
    for i:=0 to length(fitliste)-1
    do fitliste[i].gof:=fitallgt(fs,fitliste[i].par);
    //sortiere fitlist nach gof
    qsort(0,length(fitliste)-1,fitliste);
end;

function gebebestekombination(fs:tfs; var fitliste:tfitlist):Tfitresult;
var i:integer;
    bestebisher:tfitresult;
begin
    // für jede parameterkombi berechne Anpassungsfehler
    bestebisher.gof:=99999;
    for i:=0 to length(fitliste)-1
    do begin
        fitliste[i].gof:=fitallgt(fs,fitliste[i].par);
        if bestebisher.gof>fitliste[i].gof
        then bestebisher:=fitliste[i];
    end;
    result:=bestebisher;
end;

end.
```

9.7.4 Routinen zur Auswertung (Anpassungsfehler E_{obs} und Schätzfehler E_{real})

Das Funktionsmodul „statistik“ berechnet die zur statistischen Auswertung benötigten Maße, z.B. Anpassungsfehler E_{obs} und Schätzfehler E_{real} . Die Funktion „reduziereFS“ entfernt die in der Natur nicht vorhandene Information über ausgestorbene Genotypen.

```
unit statistik;

interface
  uses datenstrukturen, fsmodel;

  function abweichungParameter_Fs(p:tparameter; fs:tfs):double;
  // Eingabe: parameterkombination p und Fitness-Set fs
  // Berechnet den Anpassungsfehler  $E_{\text{obs}}$  → Rückgabewert

  procedure reduziereFS(fs:Tfs; var fsziel:Tfs);
  // Nur vorhandene Genotypen gehen in das Fitting ein → fsziel wird auf vorhandene
  // Genotypen reduziert
  // Eingabe: fs
  // Ausgabe: fsziel

  function abweichungParameter_Parameter(p1,p2:tparameter;fs:tfs):double;
  // Berechnet den Schätzfehler  $E_{\text{real}}$  - allgemein bedeutet dies das berechnen des
  // Abstandes zweier Parameterkombinationen an einem vorgegebenen Fitness-Set
  // Eingabe Parameterkombinationen p1, p2 und Fitness-Set
  // Ausgabe: Rückgabewert

implementation

  function abweichungParameter_Fs(p:tparameter; fs:tfs):double;
  var i:integer;
      abweichung:double;
  begin
    abweichung:=0;
    for i:=0 to length(fs)-1
    do abweichung:=abweichung+abs(fs[i].G-Gesamtfitness(p,fs[i]));
    result:=abweichung/length(fs);
  end;

  function abweichungParameter_Parameter(p1,p2:tparameter;fs:tfs):double;
  var i:integer;
      abweichung:double;
  begin
    abweichung:=0;
    for i:=0 to length(fs)-1
    do abweichung:=abweichung+abs(Gesamtfitness(p1,fs[i])-Gesamtfitness(p2,fs[i]));
    result:=abweichung/length(fs);
  end;

  procedure reduziereFS(fs:Tfs; var fsziel:Tfs);
  var i, ziel:integer;
  begin
    ziel:=0;
    setlength(fsziel,length(fs));

    for i:=0 to length(fs)-1
    do if fs[i].G>0
    then begin
        fsziel[ziel]:=fs[i];
        inc(ziel);
      end;

    setlength(fsziel,ziel);
  end;
end.
```

9.7.5 Darstellung der Umweltvariation und Körnigkeit

Das Funktionsmodul „grafischumwelt“ enthält die Funktionen und Prozeduren zur Erzeugung

der Umweltabfolgen, ihrer Syntax zur grafischen Darstellung, sowie ihrer „Bewertung“ zur Identifikation von Nebenlösungen, sowie der Körnigkeit.

- **Syntax:** Jede Parameterkombination beschreibt ein eigenes Zusammenspiel aus Umweltabfolgen und Generationsdauer. Visualisiert werden diese z.B. in Abb. 3 und Abb. 4. Blaue Vierecke entsprechen einem Zeitschritt Umwelt 1, rote Vierecke entsprechen einem Zeitschritt Umwelt 2. Die schwarzen Ränder entsprechen der Dauer einer Generation. Jedes Viereck läßt sich also dadurch beschreiben, ob es blau oder rot ist (Grossbuchstaben/Kleinbuchstaben), und ob es keinen schwarzen Rand hat („ohne“), normal ist (oben und unten einen schwarzen Rand), oder eine links oder rechts begrenztes Viereck darstellt. Diese Syntax wird mit einer Abfolge von Buchstaben dargestellt (Prozedure „erzeugesyntax“).
- **Bewertung zur Identifikation von Nebenlösungen:** „Blau“ und „Rot“ werden Werte zugeordnet, und die Mittelwerte „innerhalb einer Generation“ werden multipliziert (Analog dem QFM). Gleiche Lösungen erzeugen dann gleiche Werte (Funktion „anasyntax“).
- **Körnigkeit:** „Blau“ und „Rot“ erhalten Werte zugeordnet. Nun wird die Varianz über die Mittelwerte berechnet und an der größtmöglichen Varianz normiert. (Funktion `varianz_anasyntax`).

```

unit graphischumwelt;

interface
uses Datenstrukturen, sysutils;

const cL = chr(ord('l')); // Umwelt - Syntax // Linke Begrenzung, Umwelt 2
      cR = chr(ord('r')); // Rechte Begrenzung, Umwelt 2
      cB = chr(ord('b')); // Beide Seiten mit Begrenzung, Umwelt 2
      cN = chr(ord('n')); // nur oben & unten mit Begrenzung, Umwelt 2
      cO = chr(ord('o')); // ohne Begrenzung, Umwelt 2

      cbL = chr(ord('L')); // s.o. jedoch Umwelt 1, etc.
      cbR = chr(ord('R'));
      cbB = chr(ord('B'));
      cbN = chr(ord('N'));
      cbO = chr(ord('O'));

      choch = 2.3; // Wert für Umwelt 1
      cniedrig = 1.1; // Wert für Umwelt 2
      cvarnorm = (choch*choch+cniedrig*cniedrig)/2 - sqrt((choch+cniedrig)/2);
      // Maximale Varianz eines Umweltmotivs

procedure erzeugesyntax(p:tparameter;var h:string);
{Für Abb. 3 und Abb. 4 habe ich eine Umwelt-Syntax eingeführt:
- Grossbuchstaben repräsentieren einen Zeitschritt, in dem Umwelt 1 herrscht.
- Kleinbuchstaben repräsentieren einen Zeitschritt, in dem Umwelt 2 herrscht.
- „L“ und „l“ sind der Generationsbeginn (da grafisch eine „Linke“-Begrenzung (=Linie) eingezeichnet ist.
- „R“ und „r“ ist das Generationsende, da Rechts eine Linie einzuzeichnen ist.
- „N“ und „n“ ist der Normalfall (=während der Generation), ohne linke & rechte Begrenzung, nur oben und unten ist eine Linie einzuzeichnen

```


- „B“ und „b“ ist der Spezialfall, $\tau_1=1$, hier sind beide Seiten begrenzt.
- „O“ und „o“ sind Fälle außerhalb des Leserasters, und daher „ohne“ Linien.

```

function reterzeugesyntax(p:tparameter):string;
{Identisch zu „erzeugesyntax“, jedoch als Funktion, damit Aufrufe in I/O- und
Sortiererroutinen vereinfacht geschrieben werden können}

function anasyntax(s:string):double;
{Der Syntax-String erhält einen Wert zugeordnet, mit dem Nebenlösungen erkannt
werden können. Die Syntax wird in Blöcke von realisierten Fitnesses analysiert:
Analog zum QFM wird innerhalb der realisierten Fitness der Mittelwert gebildet und
die realisierten Fitnesses multipliziert. Umwelt 1 erhält den Wert hoch und Umwelt
2 den Wert niedrig.}

```

```

function varianz_anasyntax(s:string):double;
{Berechnet die Körnigkeit einer Umweltsyntax; dabei wird analog zu „anasyntax“
verfahren, wobei jedoch statt der Multiplikation der realisierten Fitnesses die
Varianz über alle realisierten Fitnesses berechnet wird → Gleichung 2.5}

```

implementation

```

procedure erzeugesyntax(p:tparameter;var h:string);
var gen,t,i:integer;
begin
  h:='';
  for gen:=1 to p.n // zeichne Leseraster
  do begin
    if p.taul=1
    then h:=h+cB // links und rechts begrenzen, da 1 generation=ein leseraster
    else begin
      h:=h+cL; // links begrenzen = Generationsbeginn
      for t:=2 to p.taul-1
      do h:=h+cN; // „normal“ = innerhalb einer Generation
      h:=h+cR; // rechts begrenzen = Generationsende
    end;
  end;

  for i:=1 to length(h) // setze Umwelten im Leseraster
  do if ((i-1) mod p.tau2) < p.l
  then h[i]:=upcase(h[i]);

  for i:=(length(h) mod p.tau2) to p.tau2-1 // vervollständige Umweltszillation
  do if i < p.l
  then h:=h+upcase(c0)
  else h:=h+c0;
end;

function reterzeugesyntax(p:tparameter):string;
var h:string;
begin
  erzeugesyntax(p,h);
  result:=h;
end;

function anasyntax(s:string):double;
var tau2, gen,
    g,u, leserasterende :integer;
    stringwert, leserasterwert:double;
    c:char;
begin
  if s<>''
  then begin
    tau2:=0;
    repeat // ermittle tau2
      inc(tau2);
      c:=upcase(s[tau2]);
    until c in [cbB,cbR];

    leserasterende:=length(s)+1;
    repeat
      dec(leserasterende);
      c:=upcase(s[leserasterende]);
    until c in [cbB,cbR];
    gen := leserasterende div tau2; // ermittle generationsanzahl

    stringwert:=1; // quantifizierung des gesamtstrings,

```

```

// 1 → neutrales element der multiplikation
for g:=1 to gen
do begin
  leserasterwert:=0; // quantifizierung eines leserasters
  for u:=1 to tau2
  do case s[(g-1)*tau2+u] of
    cL, cR, cB, cN : leserasterwert:=leserasterwert+choch;
    cbL, cbR,cbB,cbN : leserasterwert:=leserasterwert+cniedrig;
    cb0, cO : leserasterwert:=leserasterwert;
  end;
  leserasterwert:=leserasterwert/tau2; // Mittelwert bilden
  stringwert:=stringwert*leserasterwert; // aufmultiplizieren
end;
end
else stringwert:=-9999;
if issamedouble(0,stringwert)=true
then result:=0
else result:=stringwert;
end;

function varianz_anasyntax(s:string):double;
// Analysiere die syntax der Umweltvariation und quantifiziere diese zur
// einschätzung der Körnigkeit der Umwelten
var tau2, gen,
    g,u, leserasterende :integer;
    stringwert, leserasterwert:double;
    c:char;
    sumwert, sumsqwert,mean :double;
begin
  if s<>' '
  then begin
    tau2:=0;
    repeat // ermittle tau2
      inc(tau2);
      c:=upcase(s[tau2]);
    until c in [cbB,cbR];

    leserasterende:=length(s)+1;
    repeat // ermittle tau2
      dec(leserasterende);
      c:=upcase(s[leserasterende]);
    until c in [cbB,cbR];
    gen := leserasterende div tau2; // ermittle generationsanzahl

    sumwert :=0;
    sumsqwert:=0;
    for g:=1 to gen
    do begin
      leserasterwert:=0; // quantifizierung eines leserasters
      for u:=1 to tau2
      do case s[(g-1)*tau2+u] of
        cL, cR,cB,cN : leserasterwert:=leserasterwert+choch;
        cbL, cbR,cbB,cbN : leserasterwert:=leserasterwert+cniedrig;
        cb0, cO : leserasterwert:=leserasterwert;
      end;
      leserasterwert:=leserasterwert/tau2; // Mittelwert bilden
      sumwert:=sumwert+leserasterwert; // Summe
      sumsqwert:=sumsqwert+ leserasterwert*leserasterwert; // Quadratsumme
    end;
    mean:=sumwert/gen;
    stringwert:=sumsqwert/gen - sqr(mean); // Varianzberechnung
  end
  else stringwert:=-9999;
  if issamedouble(0,stringwert)=true
  then result:=0
  else result:=stringwert;
end;
end.

```

9.7.6 Reduktion auf die Hauptlösungsmenge

Im folgenden sind verschiedene Routinen zur Bearbeitung der Parameterdateien und deren

Reduktion auf eine Hauptlösungsmenge aufgeführt. Diese waren ursprünglich nicht in einem Modul zusammengefasst, da sie in verschiedene Programmen eingebettet waren (Parametererzeugung und Analyseprogramm).

Die Kernprozedur heißt „speicheresensitivitaetsanal1daten“: Diese Prozedur sortiert eine Parameterliste erst nach dem Umweltwert („anasyntaxwert“). Identische Umweltwerte beschreiben „Blöcke gleicher Lösungen“. Diese Blöcke werden nach dem Prinzip der kleinsten Lösung sortiert. Diese Liste wird gespeichert. Nach dem Einladen einer solchen Liste kann „extrahierehauptloesungen“ jeweils den ersten Eintrag eines Blockes als Hauptlösung erkennen und in die Hauptlösungsdatei übertragen. Nähere Erläuterungen zur Definition von Haupt- und Nebenlösung finden sich auch in Kapitel 2.2.3.

```

procedure QSortg0(Links, Rechts : LongInt; var table:Tfslist);
// Angepasster Qicksort Algorithmus. Das vorgegebene Array „table“ wird nach dem
// „Umweltwert“ sortiert. „Table“ enthält die zu sortierenden
// Parameterkombinationen
var i, j      : LongInt;
      x, w     : tfselement;
begin
  i := Links;
  j := Rechts;
  x := Table[(Links + Rechts) div 2];
  repeat
    while anasyntax(reterzeugesyntax(Table[i].p)) <
      anasyntax(reterzeugesyntax(x.p))      do Inc(i);
    while anasyntax(reterzeugesyntax(Table[j].p)) >
      anasyntax(reterzeugesyntax(x.p))      do Dec(j);
    if i <= j
      then begin
        w := Table[i];           // autauschen
        Table[i] := Table[j];
        Table[j] := w;
        Inc(i);
        Dec(j);
      end;
  until i > j;
  if Links < j then QSortg0(Links,j,table);
  if Rechts > i then QSortg0(i,Rechts,table);
end;

```

```

procedure QSortinnen(Links, Rechts : LongInt; var table:Tfslist);
// Angepasster Qsort Algorithmus, Das vorgegebene Array „table“ wird zwischen
// „Links“ und „rechts“ nach dem Prinzip der kleinsten Lösung sortiert
// „Sortieren von Blöcken gleicher Lösungen“
var i, j : LongInt;
      x, w : tfselement;

function innen(p:tparameter):double;
begin
  result:=p.l*1.0 + p.tau2*100.0 + p.taul*10000.0 + p.n*1000000.0;
end;

```

```

begin
  i := Links;
  j := Rechts;
  x := Table[(Links + Rechts) div 2];
  repeat
    while innen(Table[i].p) < innen(x.p) do Inc(i);
    while innen(Table[j].p) > innen(x.p) do Dec(j);
    if i <= j
      then begin
        w := Table[i];           // autauschen
        Table[i] := Table[j];
        Table[j] := w;
      end;
  until i > j;
end;

```

```

        Inc(i);
        Dec(j);
    end;
until i > j;
if Links < j then QSortinnen(Links,j,table);
if Rechts > i then QSortinnen(i,Rechts,table);
end;

procedure sortgleicheloesungen(var fslist:tfslist);
{Die Liste ist nach „Anasyntaxwerten“ vorsortiert (qsortg0) und soll nun nach Haupt
und Nebenlösungen nachsortiert werden, so dass die erste Lösung in einem "Block
gleicher Lösungen" die sog. Hauptlösung ("einfachste Lösung") ist. Zudem
wird zu jeder Lösung die Anzahl errechnet}
var blockanzeiger:string;
    gleichelsg,
    banf,bend:integer;
    anaanf:double;
    anaend:double;
    laenge:integer;
    ent:boolean;
begin
    blockanzeiger:='';
    banf:=0;
    anaend:=-9999; // vorbelegen mit unmöglichem Index
    while banf<=length(fslist)-1 // Solange der Anfang nicht das Ende erreicht hat
    do begin
        blockanzeiger:=reterzeugesyntax(fslist[banf].p);
        if issamedouble(anasyntax(blockanzeiger),
            anasyntax(reterzeugesyntax(fslist[banf].p)))
        then begin //habe Blockanfang(banf) gefunden
            bend:=banf; // nun suche blockende (bend)
            anaanf:=anasyntax(blockanzeiger);
            laenge:=length(fslist)-1;
            if bend<=laenge // wenn bend nicht das Ende überschritten hat
            then anaend:=anasyntax(reterzeugesyntax(fslist[bend].p));
            ent:=(bend<=laenge) and issamedouble(anaanf,anaend);
            while ent
            do begin
                inc(bend); // bend sitzt auf dem nächsten blockanfang
                laenge:=length(fslist)-1;
                anaanf:=anasyntax(blockanzeiger);
                if bend<=laenge
                then anaend:=anasyntax(reterzeugesyntax(fslist[bend].p));
                ent:=(bend<=laenge) and issamedouble(anaanf,anaend);
            end;
            // sortiere zwischen blockanfang (banf) und blockende (bend)
            qsortinnen(banf,bend-1,fslist);
            // Merke Anzahl der Lösungen in ".anz"
            for gleichelsg:=banf to bend-1
            do fslist[gleichelsg].anz:=bend-banf;

            banf:=bend; // schaue ab dem ende nach einem neuen block
        end
        else begin // singuläre Hauptlösung = kein block von Lösungen
            fslist[banf].anz:=1; // Merke anzahl Lösungen
            inc(banf);
        end;
    end;
end;

procedure speichereresensitivitaetsanaldaten(name:string; fslist:tfslist);
// Sortiert fslist, so dass die erste Lösung in einem Block gleicher Lösungen die
// Hauptlösung ist.
// Eingabe: name: Filename der Parameterdatei, in der die sortierte fslist
// gespeichert werden soll.
// fslist: Parameterliste
// Ausgabe: Textfile mit dem Namen „name“, enthält Zeilenweise die sortierten
// Parameter
begin
    // Blöcke gleicher Lösungen bilden = sortiere nach anasyntaxwert („g0“)
    qsortg0(0,length(fslist)-1,fslist);

    // sortiere nach l+tau2*100+tau1*10000+n*1000000
    // = Sortiere die Blöcke identischer Lösung nach Prinzip der
    // einfachsten Lösung (klein -> groß)
    sortgleicheloesungen(fslist);

    // und erzeuge die Parameterdatei

```

```

    schreibeinsfile(name,fslist);
end;

procedure extrahierehauptloesungen(inlist:tfitlist; var outlist:tfitlist);
// Überträgt die Hauptlösungen von „inlist“ nach „outlist“
// Eingabe: eine mit „speicheresensitivitaetsanaldaten“ erzeugte Parameterdatei,
//           die in „inlist“ geladen wurde.
// Ausgabe: outlist: Liste von Hauptlösungen
var vorgaenger,jetzt:fitness;
    i,outindex      : integer;
begin
    vorgaenger:=0; // extrahiere die Hauptlösungen = Finde 1. Eintrag in einem Block
                    // gleicher Lösungen
    finalize(outlist);
    outindex:=0;
    for i:=0 to length(inlist)-1
    do begin
        jetzt:=inlist[i].aehnlich;
        if not issamedouble(jetzt, vorgaenger) // die "einfachste" Hauptlösung
                                                // gefunden
                                                // (erste im block)
        then
            begin
                setlength(outlist,outindex+1); // Verlängere die Liste
                outlist[outindex]:=inlist[i]; // Kopiere den Listeneintrag
                inc(outindex);
                vorgaenger:=jetzt;           // Aktualisiere den Vergleichswert
            end;
        end;
    end;
end;

```

9.7.7 Das Simulationsmodell zur Generierung von Testdaten

Das Funktionsmodul „sensesim“ enthält die Prozedur „sensitivitätslauf“ zur Durchführung eines Simulationslaufes. Die Genotypfrequenzen der Population werden dabei gemäß der Parameter für das QFM von Generation zu Generation verändert. Die Populationsgrösse bleibt konstant. Nähere (grafische) Erläuterungen findet sich z.B. als Abb. 8 in Kapitel 3.

```

unit sensesim;

interface
    uses datenstrukturen, fsmodel;

    procedure startwertepop(var pop:tpop; popgroesse:integer);
// Produziert eine Gleichverteilung der Individuen über alle Genotypen als
// Startverteilung
// Eingabe popgroesse
// Ausgabe pop

    Procedure sensitivitaetslauf(p:tparameter; fs:tfs; popgroesse:integer;
                                var pop:Tpop);
// Die in pop enthaltene Genotypverteilung wird der oszillierenden Selektion
// ausgesetzt.
// Eingabe Parameterkombination p, Fitness-Set fs, popgroesse
// Ausgabe Genotypverteilung pop

    function sampling(pop:tpop;samplesize:integer):tpop;
// entnimmt aus pop eine Stichprobe der groesse „samplesize“
// Eingabe: pop, samplesize
// Ausgabe: Rückgabewert

    Procedure driftlauf(p:tparameter; fs:tfs; popgroesse:integer; var pop:Tpop);
// Diese Prozedur ist ein Analogon zu „sensitivitaetslauf“, bei der jedoch KEINE
// Selektion agiert.
// Eingabe Parameterkombination p, Fitness-Set fs, popgroesse
// Ausgabe Genotypverteilung pop

implementation

procedure startwertepop(var pop:tpop; popgroesse:integer);
var i:integer;
    anzgt:integer;
begin

```

```

anzgt:=length(pop);
for i:=0 to anzgt-1
do pop[i]:=trunc (popgroesse/anzgt+0.5); // Runden auf ganze Tiere
end;

function realfitness(fs:tfs; gt:integer; gen:integer; p:tparameter):double;
begin
result:=realisiertefitness(reihenfolge(gen,p.tau1,p.tau2),
p.tau1,p.tau2,p.l,fs[gt].f1, fs[gt].f2);
end;

function nextgen(var pop:tpop;gen:integer;p:tparameter;fs:tfs):tpop;
// "Vermehrung", erzeugt die Erwartete Verteilung der Genotypen nach einer
// Generation durch Multiplikation der realisierten Fitness mit der Genotypfrequenz
var i:integer;
begin
for i:=0 to length(pop)-1
do pop[i]:=pop[i]*realfitness(fs, i, gen,p);
result:=pop;
end;

function sampling(pop:tpop;samplesize:integer):tpop;
var cumfreq: array of double;
Zielpop:Tpop;
poplen,i,anzahl:integer;
r:double;
begin
// Erzeuge kummuliertes Häufigkeitsarray
poplen:= length(pop);
setlength(cumfreq,poplen);
initpop(zielpop,poplen);

cumfreq[0]:=pop[0];
for i:=1 to poplen-1
do cumfreq[i]:=cumfreq[i-1]+pop[i]; // Aufaddieren

for i:=0 to poplen-1
do cumfreq[i]:=cumfreq[i]/cumfreq[poplen-1]; // Normieren

for anzahl:=1 to samplesize
do begin
// zufällige Auswahl eines Individuums
// (realisiert als zufällige Auswahl
// eines Genotyps abhängig von seiner Frequenz) → Ziehen mit Zurücklegen
r:=random;
i:=0;
while r>=cumfreq[i] do inc(i);
// Erhöhe die Zielpopulation des Genotyps
zielpop[i]:=zielpop[i]+1;
end;
result:=zielpop;
end;

Procedure sensitivitaetslauf(p:tparameter; fs:tfs; popgroesse:integer; var
pop:Tpop);
var gen:integer;
begin
for gen:=1 to p.n
do begin
pop:=nextgen(pop,gen,p,fs); // "Vermehrung" unter Selektion
pop:=sampling(pop,popgroesse); // durch resampling -> konstante popgroesse
end;
end;

Procedure driftlauf(p:tparameter; fs:tfs; popgroesse:integer; var pop:Tpop);
var gen:integer;
begin
for gen:=1 to p.n
do pop:=sampling(pop,popgroesse); // durch resampling -> konstante popgroesse
end;
end.

```