

**Direkte EM- und Loglikelihood-Profil-Verfahren**  
**in Intervallkartierungsmodellen**  
**der Pflanzenzucht**

**Dissertation**

zur Erlangung des Grades  
einer Doktorin der Naturwissenschaften  
der Universität Dortmund

Dem Fachbereich Statistik der Universität Dortmund

vorgelegt von

**Katharina Emrich**

aus Herdecke/NRW

Dortmund 2002

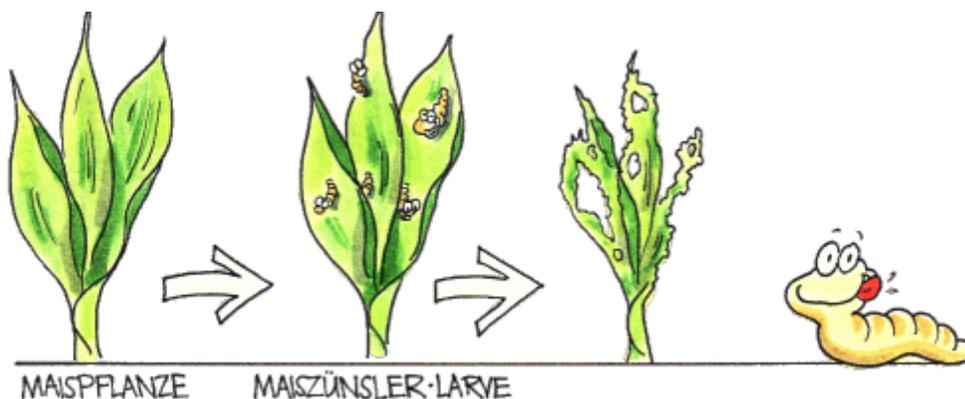
# Inhaltsverzeichnis

1. Einleitung .....	3
2. Genetische Grundlagen .....	7
3. Genkartierungsmethoden in der Pflanzenzucht .....	14
3.1 Klassische QTL-Kartierungsmethoden .....	14
Traditioneller Ansatz der quantitativen Genetik.....	14
Maximum Likelihood Methode .....	16
3.2 Intervallkartierung.....	18
4. Der EM-Algorithmus .....	22
4.1 Allgemeine Eigenschaften des EM-Algorithmus.....	22
4.2 Der EM-Algorithmus für gemischte Normalverteilungen.....	26
4.3 EM-Algorithmus und Kovarianzen für Intervallkartierungsmodelle.....	30
Der EM-Algorithmus im Profil-Intervallkartierungsverfahren .....	30
Der EMV-Algorithmus für Intervallkartierungsverfahren .....	35
5. Experimentelle Daten.....	39
5.1 Die Daten der Feldexperimente .....	39
5.2 Behandlung der fehlenden Beobachtungen ("missing values").....	41
5.3 Auswahl der Markerkofaktoren des statistischen Modells.....	44
6 Datenauswertung und Vergleich verschiedener Kartierungsverfahren.....	46
6.1 Der EM-und der ECM-Algorithmus .....	47
6.2 Vergleich von Maximum-Likelihood-Ansätzen anhand tabellarischer und graphischer Auswertung der QTL-Kartierungsdaten.....	49
Tabellarische und graphische Auswertung für Chromosom 1 .....	51
Graphische Analyse von Chromosom 1 .....	53
Tabellarische und graphische Auswertung von Chromosom 2 .....	61

Tabellarische und graphische Auswertung Chromosom 3 .....	65
Tabellarische und graphische Auswertung Chromosom 4 .....	68
Tabellarische und graphische Auswertung Chromosom 5 .....	72
Tabellarische und graphische Auswertung Chromosom 6 .....	75
Tabellarische und graphische Auswertung Chromosom 7 .....	78
Tabellarische und graphische Auswertung Chromosom 8 .....	81
Tabellarische und graphische Auswertung Chromosom 9 .....	84
Tabellarische und graphische Auswertung Chromosom 10 .....	87
6.3 Die tiefliegende Lösung im einfachen Intervallkartierungsmodell.....	90
Vergleich mit der linearen Regression (F-Statistik-Profile).....	90
6.4 Die oszillierende Lösung im SIM-Modell.....	103
6.5 Das zusammengesetzte Intervallkartierungsmodell (CIM-Modell).....	104
6.6 Die Basislösung und die F-Statistik für das CIM-Modell.....	107
7. Resultate .....	114
7.1 Diskussion der Ergebnisse .....	114
7.2 Zusammenfassung und Schlussfolgerung .....	122
8. Appendix .....	124
8.1 Appendix A: numerische Tabellen.....	124
Appendix A.1 Startwerte und Funktionswerte unter $H_0$ : .....	125
8.2 Appendix M .....	135
9. Literatur .....	141

# 1. Einleitung

Ständiger weltweiter Wandel in Kultur, Landwirtschaft und Klima erfordert neue Ansätze zur Entwicklung von Zuchtverfahren für neue Kulturpflanzensorten. Vor 100 Jahren gab es weltweit 1 000 000 000 Menschen, heute hat sich diese Anzahl versechsfacht! Es erscheint nicht möglich, die Ernährung der wachsenden Weltbevölkerung ausschließlich auf rein traditionellem Wege sicher zu stellen. Selektionsauswahl von Individuen mit wünschenswerten Eigenschaften haben in der Pflanzenzucht eine Jahrtausende alte Tradition. Heute versucht man, Methoden zu finden, um zielgerichteter gewünschte Eigenschaften in neuen Sorten zu kombinieren. Die direkte Lokalisierung quantitativer Gene im Genom der Pflanzen könnte zielgerichtete Neuzuchten ermöglichen. Insbesondere veränderte Umweltbedingungen (Wandel des Wetters, Erwärmung des Weltklimas mit nachfolgender Veränderung großer Wetterströmungen) können in Zukunft in weiten Landstrichen zu Dürre und Verkarstung führen. In Zusammenhang mit erhöhtem Bevölkerungsdruck sind Forschungen in diesem Bereich zwingend notwendig, um binnen kürzerer Fristen (d. h. innerhalb von Jahren statt Jahrzehnten) auf wandelnde Umweltbedingungen zu reagieren. Das Ziel sollte in der langfristigen Versorgung aller Menschen mit dem Lebensnotwendigen bestehen.



In der quantitativen Genetik versucht man, auf der Basis von bekannten Zusammenhängen auf unbekannte Gene rückzuschließen, welche in ihrer Summe quantitative Eigenschaften

hervorrufen. Beispielhaft zu nennen sind Resistenzmechanismen gegen amerikanische Abarnten des Maiszünlers, welcher in Mittel - und Südamerika für hohe Ernteausfälle von bis zu 80 % der Maisernte verantwortlich ist. Gelingt es, die Lage (und Ausprägung) quantitativer Gene festzustellen, so kann ein Zuchtziel viel effektiver geplant und überprüft werden.

Gensequenzierung führte dazu, dass für die meisten Kulturpflanzen die Polynucleotidsequenzen vorliegen oder feststellbar sind. Die aus (unterschiedlich langen) Sequenzen gebildeten Gene, welche zusammenwirkend landwirtschaftliche Eigenschaften realisieren, müssen jedoch erst lokalisiert werden.

Fragestellung: Neuere Verfahren der Intervallkartierung suchen das Genom von Pflanzen schrittweise nach Hinweisen auf das Vorhandensein quantitativer Gene ab. Dabei werden in Abhängigkeit der Kreuzungsexperimentdesigns genetische Modelle zumeist auf der Basis von gemischten Normalverteilungen gewählt, welche Verteilungsannahmen hinsichtlich der Aufspaltung der unbeobachtbaren quantitativen Gen-Realisationen modellieren. Verwendung finden Modelle mit zufälligen und Modelle mit zufälligen und festen Effekten. Der einfacher in Standardsoftware integrierbare Regressionsansatz wird oft verwendet, indem Mittelwerte der zufälligen Effekte genutzt werden. Für die einfachste Form dieses Ansatzes erhält man sehr ähnliche Ergebnisse im Vergleich zu Maximum-Likelihood-Schätzern. In dieser Arbeit wird untersucht werden, inwieweit die Maximum-Likelihood-Schätzung der Parameter mit Hilfe von iterativen EM-Algorithmen zusätzliche, von den Regressionslösungen abweichende Schätzer liefert. Es muss betont werden, dass der iterative EM-Algorithmus als Funktion des Startwertes gerade *keine* eindeutigen Ergebnisse sondern beispielsweise bei Profilverfahren mehrere Zweige einer Lösungsmannigfaltigkeit liefert und somit zusätzliche Information in Abhängigkeit von gewählten Startwerten resultiert. Die hier beispielhaft dargestellte Berechnung von EM- und EMV-Maximum Likelihood-Schätzern für ein einfaches lineares Modell eines F2-Generations-Kreuzungs-schemas kann auf jedes andere genetische lineare Modell, welches in der quantitativen Genanalyse Verwendung findet, angewandt werden.

In Kapitel 2 wird zunächst eine Einführung in grundlegende genetische Begriffe und Zusammenhänge gegeben. Diese sind zum näheren Verständnis der nachfolgend beschriebenen Verfahren notwendig.

Dann folgt in Kapitel 3 ein Überblick klassischer Genkartierungsmethoden, welche in den letzten Jahrzehnten entwickelt wurden und eine Einführung in die Idee der Intervallkartierung.

In Kapitel 4 wird sodann der EM-Algorithmus allgemein und im Zusammenhang mit gemischten Normalverteilungen beschrieben. Sodann erfolgt die Übertragung der vorher beschriebenen Sachverhalte auf ein genetisches F<sub>2</sub>-Generationsmodell. Hier wird zunächst eine Möglichkeit beschrieben, EM-Schätzwerte für die sogenannten Profilverfahren der Intervallkartierung zu berechnen. Sodann wird ein Formalismus hergeleitet, an Stelle von Profilen<sup>1</sup> direkte "EMV"-Schätzer zu erhalten, somit tritt an die Stelle eines Profils je aktuellem Untersuchungsintervall im Genom ein einziger Maximierungsschritt der Loglikelihood hinsichtlich der Gesamtdimension.

In Kapitel 5 erfolgt eine Einführung in die Experimentaldaten mit Beschreibung der Datenvorbereitung.

Kapitel 6 beschreibt die Datenauswertung der Experimentaldaten; sie erfolgt graphisch und anhand von Tabellen der Schätzwerte.

Resultate werden in Kapitel 7 dargestellt: Hier ist zu betonen, dass man zusätzlich zu den analog zur Regressionsanalyse auftretenden "Basislösungen" weitere Maximum-Likelihood-Schätzer findet. Diese müssen auf ihre Praxistauglichkeit hin untersucht werden. Insbesondere die als "Kofaktorlösung" bezeichneten Maximum-Likelihood-Schätzer sind aller

---

<sup>1</sup> mit einer Reduktion der Dimension der Loglikelihood um 1 und einer anschließenden Maximierung der Loglikelihood hinsichtlich dieser Dimension mit Hilfe eines diskreten Profils

Wahrscheinlichkeit nach nicht spurios<sup>2</sup> und legen nahe, im genetischen Modell alle Marker (oder aber: keine Marker) zu Kofaktoren zu machen, um eine Willkür, welche beispielsweise bei der Vorauswahl von Kofaktoren mit Regressionsansätzen geschieht, auszuschließen.

---

<sup>2</sup> Eine spuriöse Lösung ist eine mathematische Lösung, die keine Entsprechung zu naturwissenschaftlichen (oder sonstigen empirischen) Daten hat.

## 2. Genetische Grundlagen

Einige grundlegende Begriffe und genetische Zusammenhänge werden kurz erklärt, da sie zum Verständnis der später beschriebenen Methoden notwendig sind (vgl. Kull, Knodel (1980), Hennig (1998)). Wenn in einer Ausgangsgeneration (Parental Generation) eines Kreuzungsexperimentes zwei homozygote<sup>3</sup>, monogen hybride<sup>4</sup> Pflanzen (im Experiment: Inzuchtlinien) miteinander gekreuzt werden, so kann man ein Kreuzungsschema erstellen. Die hier beispielhaft dargestellte Eigenschaft sei die Farbe der Samen: gelb (dominant, in der Parentalgeneration mit Genotyp YY) und rot (rezessiv, in Parentalgeneration mit Genotyp yy), wobei im Phänotyp (Erscheinungsbild) ein rezessives Allel des Genotyps von einem dominanten Allel überlagert wird. Der Genotyp eines Gens umfasst jeweils zwei Allele. Im Beispiel mögliche Allele sind Y und y.

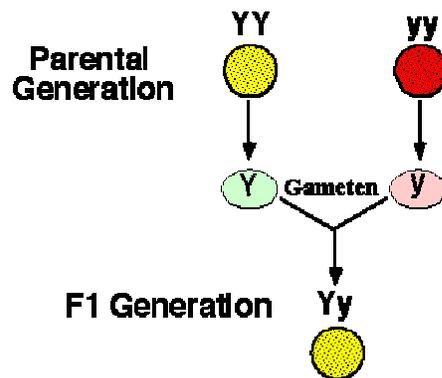


Abbildung 2.1: Parental Generation eines Kreuzungsexperimentes und F1 Generation

<sup>3</sup> gleicherbig

<sup>4</sup> d. h. nur ein Gen berücksichtigende und sich in einem Merkmal unterscheidende

Die Gameten sind in Abbildung 2.1 grün (Y) und rot (y) dargestellt. Nach Mendels Uniformitätsgesetz treten in der ersten Filialgeneration F1 nur Pflanzen mit gelben Samen auf. Kreuzt man Pflanzen dieser Generation wieder (üblicherweise über Selbstbefruchtung) miteinander, so erhält man nach dem Spaltungsgesetz von Mendel

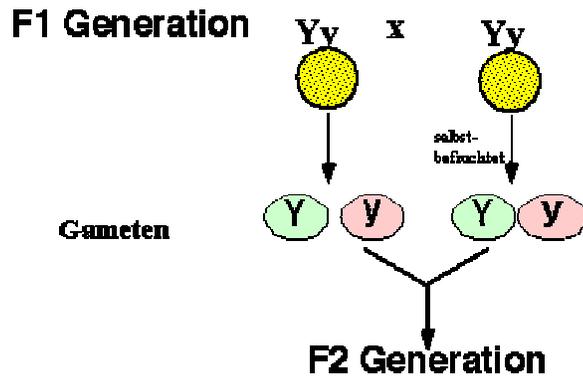


Abbildung 2.2: Selbstbefruchtung der F1-Pflanzen

in der zweiten Filialgeneration die in Abbildung 2.2 dargestellte Aufspaltung des Phänotyps im Verhältnis 3:1 und des Genotyps im Verhältnis 1:2:1.

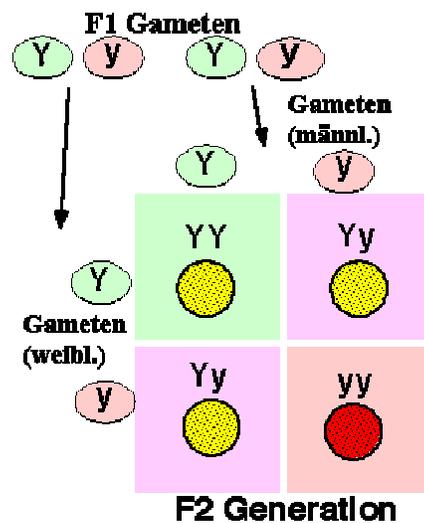


Abbildung 2.3: Aufspaltung der Eigenschaft Samenfarbe der Nachkommen in der F2 Generation

Der Phänotyp einer Pflanze eines dominant rezessiven Erbgangs erlaubt es nicht, Rückschlüsse auf den Genotyp der Pflanze zu ziehen, d. h. die F2 Pflanzen mit Genotyp YY haben das gleiche phänotypische Erscheinungsbild wie die Pflanzen mit Genotyp Yy (also mit den Allelen Y und y). Den Genotyp einer gelben F2 Pflanze kann man ausschließlich über eine Testkreuzung feststellen. Hier wird eine Testkreuzung mit einer rezessiv homozygoten Pflanze dargestellt. Die Testkreuzung dieses Beispiels kann als eine Rückkreuzung betrachtet werden, da der rezessiv homozygote Genotyp dem des rezessiven Elter aus der Parentalgeneration entspricht.

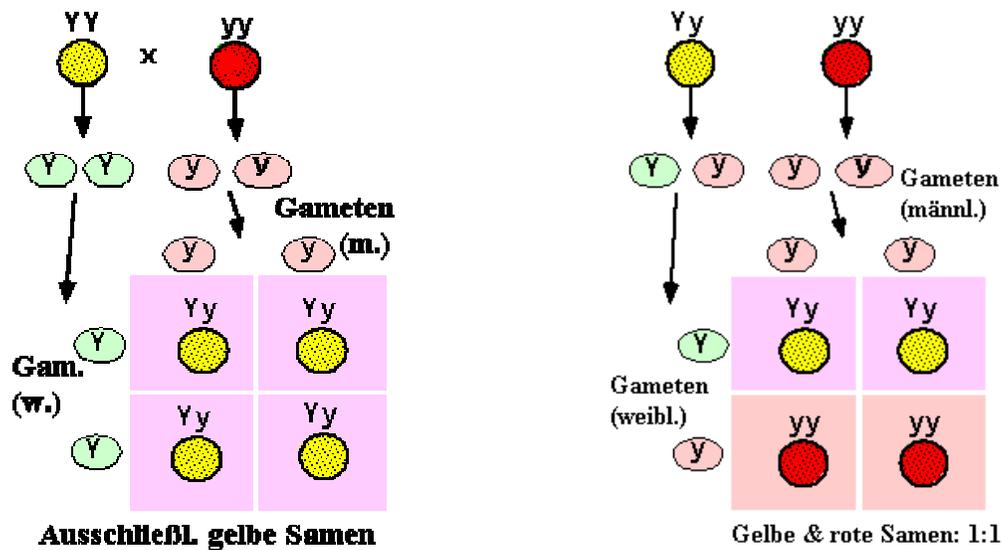
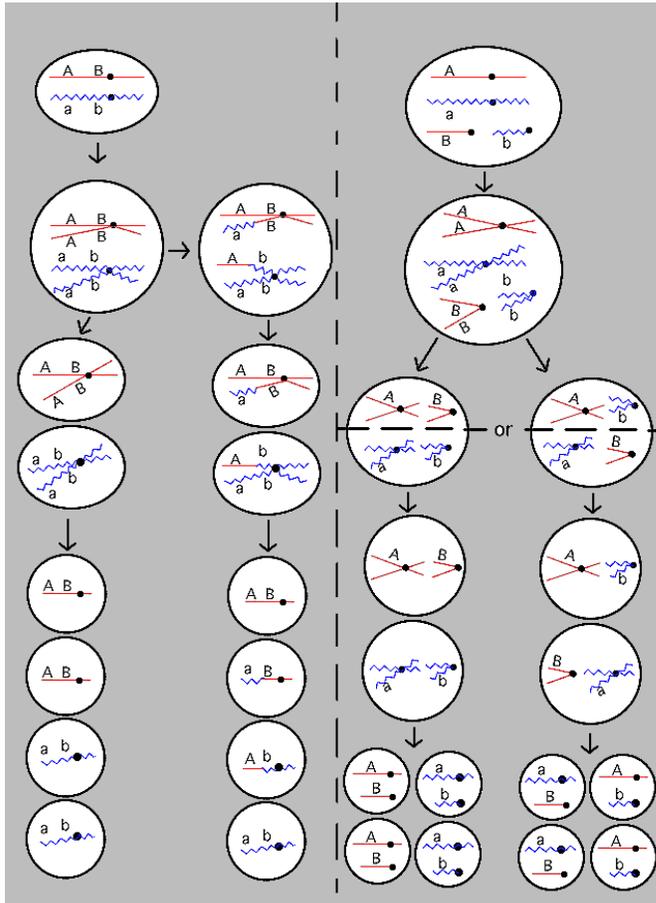


Abbildung 2.4: Testkreuzung einer homozygot (Genotyp YY)-und einer heterozygot (Yy)-dominanten Pflanze mit einer homozygot rezessiven Pflanze (yy)

Der homozygot dominante Genotyp (mit gelben Samen) wird ausschließlich gelbbesamte Nachkommen haben, da sich wieder das Uniformitätsgesetz realisiert (Abbildung 2.4, links). Der heterozygot dominante Genotyp (ebenfalls gelb besamt) wird dagegen im Verhältnis 1:1 gelb- und rot-besamte Nachkommen hervorbringen (Abbildung 2.4, rechts).

Es gibt selbstverständlich viele weitere Erbgänge als den beispielhaft dargestellten dominant-rezessiven. Bei einem kodominanten Erbgang beeinflussen beide Allele den Phänotyp in unterschiedlicher Weise, so dass der Genotyp direkt aus dem Phänotyp ablesbar ist; das Zahlenverhältnis des Genotyps in einem Kreuzungsschema entspricht dann dem des Phänotyps.

Betrachtet man dihybride oder polyhybride F<sub>2</sub> Kreuzungen, bei welchen sich die Pflanzen in 2 oder mehr Merkmalen voneinander unterscheiden, so erhält man ebenfalls feste Aufspaltungsverhältnisse der Merkmalskombinationen der Nachkommenschaft.



**Abbildung 2.5: Zwei Gene A und B (mit Allelen A, a und B, b): Crossing over (rechts) und unabhängige Allelverteilung (links) in der Meiose (Keimzellenbildung)**

Dabei kann es vorkommen, dass zusätzlich zu den von Mendel beschriebenen Zahlenverhältnissen noch eine geringe Anzahl von sog. Rekombinanten auftritt. Dies geschieht immer dann, wenn zwei der untersuchten Eigenschaften (Merkmale) ihre Gene auf dem gleichen Chromosom (in der gleichen Kopplungsgruppe, d. h. im gleichen Chromosomenabschnitt) haben. Dann kommt es bei der Bildung der Keimzellen (Gameten) in der Meiose zu sog. Kopplungsbrüchen oder Crossing over (Abbildung 2.5, rechts).

Die Anzahl auftretender Rekombinanten zwischen zwei Genorten gibt Auskunft über relative Abstände der beiden Genorte zueinander und ermöglicht letztendlich die Konstruktion von Genkarten. Zwei sehr nah beieinander liegende Genorte werden weniger Crossing over, und damit weniger Rekombinanten eines Kreuzungsexperimentes hervorbringen, als zwei weiter voneinander entfernt liegende Gene. In einer Genkarte werden bekannte Gene des Genoms in linearer Reihenfolge angeordnet dargestellt. Die Anzahl von Rekombinanten wird meist in Centimorgan (cM) angegeben (Czihak et. al., 1990):

$$m = \frac{\text{Zahl der Rekombinanten}}{\text{Gesamtzahl der Nachkommen}}, \text{ cM} = m \cdot 100 \text{ [cM]} \quad (2.1)$$

$m$  wird als Morgans Kartierungsfunktion bezeichnet. Morgans Kartierungsfunktion (mapping function) ist nur für *kleine* Genabstände sinnvoll, da bei großen Abständen zwischen zwei Genen die Frequenz von Doppel-Crossing over steigt. Es gilt für 3 Genorte A, B und C (Liu, 1998):

$$r_{AC} = r_{AB} + r_{BC} - 2 \cdot r_{AB} \cdot r_{BC} \quad (2.2)$$

$r_{AB} \cdot r_{BC}$  ist die Frequenz für Doppel-Crossing over. Die Einheit cM kann mit Hilfe von Kartierungsfunktionen in Rekombinationsfrequenzen überführt werden, welche für die später beschriebenen QTL- Kartierungsmethoden benötigt werden. Für *große* Genabstände verwenden viele Kartierungsansätze Haldanes Kartierungsfunktion:

$$r = \frac{1}{2} (1 - e^{-2 \cdot m}) \quad (2.3)$$

Die maximale Rekombinationsfrequenz  $r$  ist dann gegeben als  $r=0,5$  und steht für eine unabhängige Verteilung der Allele zweier Gene bei der Weitergabe an die Filialgeneration.

### **3. Genkartierungsmethoden in der Pflanzenzucht**

#### **3.1 Klassische QTL-Kartierungsmethoden**

In der Vergangenheit wurden eine Vielzahl von Methoden zur Schätzung von Gen(-Ort)-en (sog. QTL: quantitative trait loci) quantitativer Eigenschaften von Pflanzen entwickelt. Es war und ist in der Pflanzenzucht von großem Interesse, festzustellen, von welchen Genen eine Eigenschaft wie die Wuchshöhe oder der Ertrag einer Pflanzensorte beeinflusst werden.

Auf den folgenden Seiten wird ein kurzer Einblick in die Entwicklungsgeschichte von Methoden der quantitativen Genanalyse gegeben.

#### **Traditioneller Ansatz der quantitativen Genetik**

Beginnend mit Sachs (1923) wurden in der quantitativen Genetik Verfahren entwickelt, um *in der Nähe* eines bekannten Markers ein Gen/ QTL zu lokalisieren, welches die Ausprägung einer untersuchten Eigenschaft beeinflusst.

Bei Inzuchtexperimenten kann man in verschiedenen Kreuzungsschemata als bekannt voraus setzen, dass die Nachkommen eines Kreuzungsexperimentes bestimmte Markergenotypkombinationen aufweisen. In einem (monohybriden) Rückkreuzungsexperiment ist bekannt, dass die Nachkommen in der einen Gruppe den Markergenotyp AA und in der anderen Gruppe von Nachkommen den Genotyp Aa besitzen. Zunächst ging man so vor, dass die phänotypischen Mittelwerte zweier Gruppen von Nachkommen eines Pflanzenzuchtexperimentes verglichen wurden.

Man berechnet in einem ersten einfachen Regressionsansatz einen Substitutionseffekt (auch additiver Effekt genannt) des Austauschs des Allels A gegen das Allel a durch Schätzen der Differenz der Gruppenmittel. Dieser phänotypische Effekt einer Allelsubstitution wurde mit einem einfachen t-Test (Ann.: Normalverteilung der zugrunde liegenden Populationen) mit folgender Hypothese überprüft:

$$H_0 : \mu_D = (\mu_{AA} - \mu_{Aa}) = 0 \quad \text{vs.} \quad H_1 : \mu_D \neq 0 \quad (3.4)$$

Es wurde also getestet, ob die Differenz der Gruppenmittel der Nachkommengruppen mit Markergenotyp AA und Aa gleich 0 ist, um das Vorhandensein eines phänotypischen Effekts im Kreuzungsexperiment zu überprüfen.

Lander und Botstein (1989) beschrieben einige Nachteile dieser Methode:

(i) Falls der quantitative Merkmalsgenort (QTL) nicht in der direkten Umgebung des Markerlocus zu finden ist, so kann der phänotypische Effekt stark unterschätzt werden. Für eine Rekombinationsfrequenz (nach Haldanes Kartierungsfunktion) von

$$r = \begin{cases} (\text{Abstand zwischen Marker und QTL in cM=Centimorgan}) \cdot 0,01, & \text{falls der Abstand} \leq 50 \text{ cM} \\ 0,5 & \text{sonst} \end{cases} \quad (3.5)$$

kann der phänotypische Effekt einen Bias von  $(1 - (2 \cdot r))$  haben, bzw. um diesen Faktor nach unten verzerrt sein. Beweis: Falls die beiden QTL Genotypklassen phänotypische Mittelwerte von 0 und 1 haben, so haben die beiden Markerklassen phänotypische Mittelwerte von  $r$  und  $(1-r)$ . Dies folgt aus der Definition der Rekombinationsfrequenz.

(ii) Falls der QTL nicht auf dem Markerlocus zu finden ist, so wächst die Nachkommenzahl, welche zum Nachweis des QTL benötigt wird, deutlich an. Insbesondere gilt, dass die durch den Marker erklärte Varianz um  $(1 - 2 \cdot r)^2$  abnimmt und

die Anzahl benötigter Nachkommen dementsprechend um den Faktor  $1/(1-2\cdot r)^2$  zunimmt.

(iii) Der Ansatz definiert *keine* mögliche oder wahrscheinliche Position des QTL. Enge Kopplung zu einem QTL mit geringem Effekt lässt sich nicht unterscheiden von entfernter Kopplung zu einem QTL mit großem Effekt.

(iv) Die im Ansatz verwendete Fehlerwahrscheinlichkeit 1. Art mit  $\alpha = 0,05$  berücksichtigt nicht die Tatsache, dass an vielen Markern getestet wird, ob ein QTL vorhanden ist. Das Risiko, ein QTL fälschlicherweise zu lokalisieren beträgt (nach Gleichung (3.1)) an *einem* Marker nur 5%, allerdings ist natürlich die Chance, *irgendwo* im Genom fälschlicherweise ein QTL zu lokalisieren sehr viel höher.

Die hier aufgeführten Nachteile des traditionellen Ansatzes entstehen dadurch, dass man die Marker *einen nach dem anderen* analysiert.

Parallel zur traditionellen Methode entstand die

### **Maximum Likelihood Methode**

Man kann den traditionellen Ansatz der linearen Regression des Phänotyps auf den Genotyp als einen Spezialfall der Maximum Likelihood Methode ansehen. Es gilt dann folgender Zusammenhang:

$$\phi_i = a + b \cdot g_i + \varepsilon_i \quad (3.6)$$

$\phi_i$  seien die Realisationen der quantitativen Eigenschaft,  $g$  sei eine (0,1)-Indikatorvariable, welche die Anzahl von  $a$  Allelen des rückgekreuzten Elters im oben genannten

Rückkreuzungsexperiment angibt<sup>5</sup>,  $\varepsilon$  ist eine normalverteilte Zufallsvariable mit Erwartungswert 0 und Varianz  $\sigma^2$ ,  $i$  ( $=1,2,\dots,n$ ) ist die Nummer des untersuchten Studienobjekts einer Stichprobe von  $n$  Pflanzen des Experimentes,  $a$ ,  $b$  und  $\sigma^2$  sind unbekannte Parameter.  $b$  steht hier für den geschätzten phänotypischen Effekt einer einzelnen Allel-Substitution an einem vermuteten QTL. In einem solch einfachen linearen Modell entsprechen die Lösungen der linearen Regression  $(\hat{a}, \hat{b}, \hat{\sigma}^2)$  den Maximum Likelihood Schätzern (MLEs) der Parameter. Es handelt sich also um die Werte, welche die Likelihoodfunktion  $L(a, b, \sigma^2)$  maximieren, wobei man die Likelihoodfunktion als die Wahrscheinlichkeit ansehen kann, dass die beobachteten Daten in dieser Form aufgetreten sind (es handelt sich um die Dichtefunktion der Zufallsvariable, welche die untersuchte Eigenschaft kreiert unter der Annahme von variablen Parametern und festen Realisationen der Zufallsvariable).

$$L(a, d, \sigma^2) = \prod_{i=1}^n f((\phi_i - a - b \cdot g_i), \sigma^2) \quad (3.7)$$

und

$$f(x, \sigma^2) = (2 \cdot \pi \cdot \sigma^2)^{-1/2} \exp(x^2 / 2 \cdot \sigma^2) \quad (3.8)$$

ist Normaldichte von  $x$  mit Varianz  $\sigma^2$ .

Für Maximum Likelihood Schätzer wird häufig ein Likelihoodquotienten Test unter der Hypothese  $H_0 : b = 0$  durchgeführt. Aus der Tradition praktischer Anwendungen in der

---

<sup>5</sup> Mögliche Genotypen des rückgekreuzten Elters sind AA und Aa.

KARTIERUNG VON GENORTEN QUANTITATIVER EIGENSCHAFTEN

Humangenetik heraus verwendet man dabei in der quantitativen Genetik den sog. LOD Score mit dem dekadischen Logarithmus als Teststatistik:

$$\text{LOD} = \log_{10} \left( \frac{L(\hat{a}, \hat{b}, \hat{\sigma}^2)}{L(\hat{a}_0, 0, \hat{\sigma}_0^2)} \right), \quad (3.9)$$

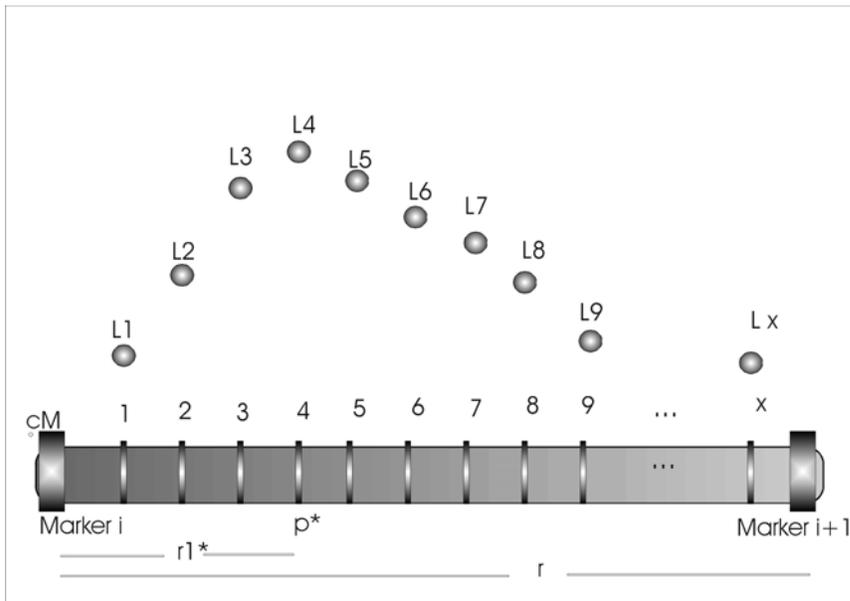
Ebenso gut kann man selbstverständlich den natürlichen Logarithmus ( $\ln(\cdot)$ ) verwenden (es gilt:  $\ln(x) = \ln(10) \cdot \log_{10}(x)$ ). Der Ausdruck (3.9) wird groß, wenn die Loglikelihood unter der Hypothese  $H_0$  deutlich kleiner als die Loglikelihood-Funktion am Schätzwert (unter der Alternative) ist. Wenn ein Testwert  $T$  überschritten wird, so nimmt man an, dass ein QTL am untersuchten Marker vorliegt. Die Maximum Likelihood Methode wurde an einer Vielzahl von Markern angewendet und beinhaltet somit dieselben Probleme wie die traditionelle Methode.

### **3.2 Intervallkartierung**

In den 80-er Jahren des letzten Jahrhunderts entwickelte man Verfahren, welche die Lokalisierung einer Vielzahl von Markergenen im Genom von Kulturpflanzen ermöglichten. Führt man heute Kreuzungsexperimente mit Kulturpflanzen (z. B. Maissorten) durch, so kann man zumeist die Genotypen einer großen Anzahl von Markergenen feststellen. Viele der benötigten Verfahren sind aufwändig und teuer, da das Erbmaterial der Pflanzen im Labor untersucht werden muss. In der Datenauswertung in Kapitel 6 wird ein Beispiel einer RFLP (restriction fragment length polymorphism)-Markergenkarte angegeben. Markergenkarten werden anhand der Häufigkeit auftretender Rekombinanten zwischen verschiedenen Markergenen geschätzt. Die verwendeten Rekombinationsfrequenzen geben die relative Lage der Markergene zueinander im Genom

wieder. Allerdings hängt es vom verwendeten Verfahren ab, welche Markerreihenfolge ermittelt wird, so dass diese nicht eindeutig ist.

Die Standardsoftware der Intervallkartierung Mapmaker 3.0 nach Lander et. al. (1987), Lincoln et. al.(1992) und (1993)) ermöglicht das Schätzen von Markergenkarten und wird in der Praxis sehr häufig verwendet.



**Abbildung 3.1: Intervallkartierung: Ein Markerintervall der Länge  $x$  cM mit untersuchten Positionen, den zugehörigen Profilwerten  $L1, L2, \dots, Lx$  der Schätzwerte an Positionen  $1, 2, \dots, x$  und Lageschätzer des QTL  $p^*$ .**

Die Methoden der Intervallkartierung verwenden eine vorhandene Markergenkarte, um das gesamte Genom (oder interessierende Teile des Genoms) mit Markerintervallen zu überdecken. Jeweils zwei benachbarte Markergene bilden dann ein Markerintervall. Abbildung 3.1 zeigt ein schematisiertes Beispiel eines Markerintervalls zwischen den Markern Nr.  $i$  und Nr.  $(i+1)$ . Das Intervall hat die Länge von  $x$  cM, so dass man mit Haldanes Kartierungsfunktion die Rekombinationsfrequenz  $r$  zwischen dem linken und rechten Marker berechnen kann. Dann untersuchen Intervallkartierungsverfahren

schrittweise jede mögliche Position *innerhalb* des Intervalls (in Abständen von 1, 2 oder auch 5 cM), um festzustellen, ob das mutmaßlich vorhandene quantitative Merkmalsgen innerhalb des Markerintervalls einen Genort (QTL= quantitative trait locus) besitzt. Hierzu werden Parameterschätzmethoden auf Modelle mit Zufallseffekten (einfache Intervallkartierung, SIM) bzw. auf Modelle mit gemischten (zufälligen und festen) Effekten (zusammengesetzte Intervallkartierung, CIM) angewandt. Die zufälligen Effekte werden für Verteilungen der ansonsten unbeobachtbaren QTL-Genotypen benötigt. Regressionsschätzverfahren für SIM-Modelle wurden von Knapp et. al. (1990), Haley and Knott (1992), Knott und Haley (1992), Haley et. al. (1994) Martinez und Curnow (1992), Jansen (1992) und (1993) diskutiert. Hier werden die zufälligen Effekte eines einfachen linearen Modells durch ihre Erwartungswerte ersetzt. Lander und Botstein (1989) und Jansen (1993) schlugen vor, die Intervallkartierung mit Maximum-Likelihood-Methoden durchzuführen, dabei spielt der EM-Algorithmus zur Schätzung der unbekanntgen genetischen Parameter eine wichtige Rolle. CIM-Modelle entwickelte man, indem in SIM-Modelle Markerinformation in Form einer Kofaktormatrix integriert wurde. Diese Verfahren wurden zuerst von Jansen (1993) und Zeng ((1993) und (1994)) angeregt. Zudem fanden viele Diskussionen statt hinsichtlich der Wahl entsprechender genetischer Modelle, welche in die statistischen Intervallkartierungsmodelle integriert werden. Es gibt daher für verschiedene Kreuzungsexperimente und experimentelle Designs eine Vielzahl von Methoden. Doerge et. al. (1997) geben einen Überblick bestehender Verfahren.

Probleme hinsichtlich fehlender Markerinformationen ("missing values") wurden für Regressionsmodelle von Martinez and Curnow (1994) diskutiert (vgl. auch Little, (1992), Little and Rubin (1986), Jiang and Zeng (1997)). Ebenso wurden (insbesondere für Regressionsanalyse- Modelle) Ansätze entwickelt, um Umwelteinflüsse in die Auswertung

zu integrieren (Sari-Gorla et. al. (1997) (1) und (2)). Es existieren Ansätze, um die Anzahl quantitativer Gene in ein Modell zu integrieren (Jansen (1994), Jansen and Stam (1994), Sillanpää and Arjas (1998)).

Auch Verfahren für multiple QTL-Modelle wurden bereits diskutiert (Moreno-Gonzalez (1992), Jansen (1993), Rodolphe und Lefort (1993), Kao et al (1999)). In Kapitel 4 wird gezeigt, wie EM-Schätzer für ein F<sub>2</sub>-Kreuzungsexperiment hergeleitet werden können. Dabei wird ein CIM-Modell verwendet, das SIM kann man für einen solchen Modellierungsansatz immer als einen Spezialfall des CIM-Modells ansehen (bei welchem die Faktorkomponente wegfällt). In der Datenauswertung (Kapitel 6) werden sowohl CIM- als auch SIM-Modelle berechnet und in Kapitel 7 diskutiert. Primär wird dabei das Verhalten der Maximum-Likelihood-Schätzverfahren mit Hilfe des EM-Algorithmus analysiert.

## 4. Der EM-Algorithmus

### 4.1 Allgemeine Eigenschaften des EM-Algorithmus

Der EM-Algorithmus ist ein iteratives Verfahren zur Maximum-Likelihood-Schätzung, wenn auf Grund fehlender Beobachtungen oder wegen unbeobachtbarer, den Beobachtungen zugeordneten Variablen ein sogenannter "missing data"-Ansatz formuliert werden kann (Dempster, Laird, Rubin (1977), Wu, (1983), McLachlan und Krishnan (1997), Meng und Rubin (1993)). Für Verfahren der Intervallkartierung kann der Merkmalsbeobachtungsvektor  $y$  gemeinsam mit den nicht beobachtbaren Realisationen des (unbekannten) quantitativen Gens als ein hypothetisch vollständiger Datensatz behandelt werden, so dass der EM-Algorithmus in der nun folgenden Definition anwendbar ist.

Seien  $X$  und  $Y$  zwei Stichprobenräume und es existiere eine surjektive ("many to one") Abbildung von  $X$  nach  $Y$ . Sei  $y$  eine Realisation von  $Y$ . Die  $x$  aus  $X$  seien nicht direkt, sondern nur indirekt über  $y = y(x)$  in  $Y$  beobachtbar. Dann sei eine Familie von Zufallsdichten  $f(x|\theta)$  für  $x$ , mit Parametervektor  $\theta \in \Omega$  gegeben und

$$g(y|\theta) = \int_{X(y)} f(x|\theta) dx \quad (4.1)$$

sei Dichte von  $y$ . Dabei sei  $X(y) = \{x : y(x) = y\}$ .

Die Maximum-Likelihood-Schätzer werden berechnet, indem die Dichte der beobachteten Daten  $g(y|\theta)$  über  $\theta \in \Theta$  maximiert wird. Dieses Vorgehen ist in vielen Anwendungen jedoch sehr aufwändig oder nur numerisch möglich. In solchen Fällen ist jedoch die Maximierung der den vollständigen Daten zugeordneten Dichte  $f(x|\theta)$  zumeist einfacher

durchzuführen. Es gilt allerdings, dass  $x$  unbeobachtbar ist. Daher wird  $\ln(f(x|\theta))$  durch den *bedingten Erwartungswert* für vorgegebene  $y$  und vorgegebene vorangegangene Schätzwerte des iterativen Verfahrens ersetzt.

Der  $(t+1)$ 'ten E-(expectation)-Schritt des EM-Algorithmus berechnet

$$Q(\theta^{(t)}) = E_{\theta^{(t)}} \left[ \ln(f(x|\theta)) \mid x \in X(y) \right] \quad (4.2)$$

Im  $(t+1)$ 'ten M-(maximization)-Schritt des EM-Algorithmus wird diese bedingte Erwartung maximiert:

$$\max_{\theta \in \Theta} \left( E_{\theta^{(t)}} \left[ \ln(f(x|\theta)) \mid x \in X(y) \right] \right) \quad (4.3)$$

Die Iterationen werden durchgeführt, bis die Differenz der Folge von Funktionswerten  $|Q(\theta^{(t+1)}) - Q(\theta^{(t)})|$  "klein" ist (auch andere, äquivalente, Konvergenzkriterien sind möglich). Am Konvergenzpunkt  $\hat{\theta}$  ist dann unter Restriktionsbedingungen nicht nur ein Maximum von  $Q(\hat{\theta})$  gegeben, sondern auch ein Maximum der beobachteten Loglikelihood-Funktion.

Für Exponentialfamilien gelten besonders einfache Herleitungen des EM-Algorithmus über suffiziente Statistiken. Die hier diskutierten gemischten Normalverteilungen stammen jedoch nicht aus Exponentialfamilien, sondern sind Mischungen von Exponentialfamilien, so dass viele besonders einfache Aussagen zum Konvergenzverhalten des EM-Algorithmus nicht angewendet werden können.

Restriktionsbedingungen und das Konvergenzverhalten des EM-Algorithmus werden ausführlich von den oben genannten Autoren diskutiert. Insbesondere werden folgende Vorteile des EM-Algorithmus gegenüber anderen iterativen Verfahren (Newton-Raphson- und Fishers- Scoring-Methode) genannt:

- Der EM-Algorithmus ist numerisch stabil und erhöht die Loglikelihood mit jedem Iterationsschritt.
- Der EM-Algorithmus lässt sich leicht implementieren, da sich die vollständige Loglikelihood-Funktion häufig in eine explizit maximierbare Form bringen lässt.
- Selbst wenn die explizite Form sich nicht herstellen lässt, ist oft eine Maximierung mit Standardsoftwarepaketen im M-Schritt möglich.

Allerdings kann der EM-Algorithmus die Konvergenz gegen ein globales Maximum nicht garantieren, dies gilt für alle iterativen Verfahren. Sind mehrere Maximalwerte vorhanden, so hängt es von der Wahl der Startwerte ab, welches Maximum gefunden wird.

Es wird kein Schätzer der Kovarianzmatrix automatisch "mitgeliefert" (wie in der Scoring Methode). Theoretisch kann es vorkommen, dass ein Sattelpunkt statt eines Maximums gefunden wird (s. insbesondere Wu (1983)), so dass die Maximumseigenschaft überprüft werden muss. In Kombination mit dem nun dargestellten Verfahren zur Berechnung von asymptotischen Kovarianzmatrizen von EM-Schätzern ist es jedoch leicht möglich, die negative Definitheit der Hessematrix der Loglikelihood-Funktion zu überprüfen.

Für Situationen, in welchen der EM-Algorithmus Anwendung findet, hat Louis (1982) jedoch eine Möglichkeit hergeleitet, um über die Loglikelihood-Funktion der vollständigen Daten eine Informationsmatrix der beobachteten Loglikelihood-Funktion herzuleiten. Somit wird für eine solche Situation die Berechnung einer *asymptotischen* Kovarianzmatrix ermöglicht.

Gegeben sei die Loglikelihood der beobachteten Daten

$$\ln(g(y|\theta)) = \ln\left(\int_{x(y)} f(x|\theta) dx\right) \quad (4.4)$$

und die Loglikelihood der vollständigen Daten  $\ln(f(x|\theta))$ . Dann wird im Maximierungsschritt des EM-Algorithmus

$$\max_{\theta \in \Theta} \left( E_{\theta^{(t)}} \left[ \ln(f(x|\theta)) \mid x \in X(y) \right] \right) \quad (4.5)$$

berechnet, woraus folgt, dass

$$\frac{\partial}{\partial \theta} \left( E_{\theta^{(t)}} \left[ \ln(f(x|\theta)) \mid x \in X(y) \right] \right) \stackrel{!}{=} 0 \quad (4.6)$$

( d. h. man sucht die Nullstellen des Gradienten der erwarteten Loglikelihood-Funktion).

Sei  $S(x, \theta) = \frac{\partial}{\partial \theta} (\ln(f(x|\theta)))$  Gradient von  $\ln(f(x|\theta))$  und

$$S^*(y|\theta) = \frac{\partial}{\partial \theta} (\ln(g(y|\theta))) = \frac{\int_{X(y)} f'(x|\theta) dx}{\int_{X(y)} f(x|\theta) dx} \quad (4.7)$$

Gradient von  $\ln(g(y|\theta))$ . Multiplikation des Zählers von (4.7) mit  $f(x|\theta)/f(x|\theta)$  ergibt:

$$S^*(y|\theta) = E_{\theta} \left[ \ln(f(x|\theta)) \mid x \in X(y) \right], \text{ zudem gilt für den EM-Schätzer } \hat{\theta}: \quad (4.8)$$

$$S^*(y|\hat{\theta}) = 0.$$

Die den Gradienten zugeordneten Hessematrizen (2. Ableitungsmatrizen) werden bezeichnet als  $B(x|\theta)$  und  $B^*(y|\theta)$

Dann folgt aus dem "Missing Information Principle", vgl. Orchard and Woodbury (1972)

bzw. aus der direkten Ableitung und Umformung von  $S^*(y|\theta)$  (Louis, 1982):

$$I_Y(\theta) = E_{\theta} (B(x|\theta) \mid x \in X(y)) - E_{\theta} (S(x|\theta) \cdot S^T(x|\theta) \mid x \in X(y)) + S^*(y|\theta) \cdot S^{*T}(y|\theta) \quad (4.9)$$

Somit kann die Informationsmatrix der beobachteten Loglikelihood als Substraktion der Informationsmatrix der vollständigen Loglikelihood mit der Informationsmatrix der "missing data" Loglikelihood berechnet werden:

$$I_Y = I_X - I_{X|Y}. \quad (4.10)$$

Dieser Zusammenhang gilt nur *in direkter Umgebung des Maximums*  $\hat{\theta}$ , Invertierung von  $I_Y$  liefert die asymptotische Kovarianzmatrix.

## 4.2 Der EM-Algorithmus für gemischte Normalverteilungen

Die gemischte Normalverteilung findet Anwendung, wenn die untersuchte Population in  $g$  unterscheidbare Gruppen zerfällt, wobei die Proportionen der Unterpopulationen in der Grundgesamtheit unbekannt sind. Dies ist bei Intervallkartierungsmethoden in genetischen Kreuzungsexperimenten der Fall, da das vermutete quantitative Gen in den untersuchten Nachkommen in Abhängigkeit der benachbarten Markergenotypen nach Mendels Spaltungsgesetz in verschiedenen Genotypenrealisationen auftritt. Daher wird zunächst die gemischte Normalverteilung diskutiert (Titterington et. al. (1985), McLachlan und Krishnan (1997) und McLachlan und Peel (2000)).

Eine univariate gemischte Normaldichte werde definiert:

Gegeben sei ein Zufallsvektor  $Y$  aus einer gemischten Normalverteilung mit  $g$  Komponenten:

Dann ist die gemischte Normaldichte anzugeben als:

$$f(y; \psi) = \sum_{i=1}^g p_i f_i(y; \theta_i), \quad (4.11)$$

wobei<sup>6</sup>  $\psi = (\theta^T, p_1, p_2, \dots, p_g)^T$ ,  $\theta_i = (\mu_i, \sigma^2)$ ,  $\theta^T = (\mu_1, \mu_2, \dots, \mu_g, \sigma^2)$

$$\text{und } f_i(y; \theta_i) = (2 \cdot \pi \cdot \sigma^2)^{-\frac{1}{2}} \cdot \exp\left(-\frac{(y - \mu_i)^2}{2 \cdot \sigma^2}\right) \quad (4.12)$$

sei Normaldichte mit Erwartungswert  $\mu_i$  und Varianz  $\sigma^2$ .

Die beobachtete Zufallsstichprobe der gemischt normalverteilten Population sei gegeben

$$\text{durch: } y = (y_1, y_2, \dots, y_n). \quad (4.13)$$

Die Loglikelihood-Funktion der *beobachteten* Daten für  $\psi$  ist dann gegeben als:

$$\ln(L(\psi|\theta)) = \sum_{j=1}^n \ln(f(y_j; \psi)) = \sum_{j=1}^n \ln\left(\sum_{i=1}^g p_i \cdot f_i(y_j; \theta_i)\right) \quad (4.14)$$

Um die Maximum-Likelihood-Schätzer der obigen Funktion zu berechnen, muss die Loglikelihood-Funktion nach allen Parametern im Vektor  $\psi$  abgeleitet werden. Die Funktion (4.14) ist zwar differenzierbar, aber die Berechnung einer expliziten Lösung des partiellen Ableitungssystems ist sehr kompliziert (obwohl zumindest numerisch möglich). In solchen Fällen hat sich der EM-Algorithmus als adäquate Wahl zur expliziten Berechnung der Maximum-Likelihood-Schätzer erwiesen. Dempster, Laird und Rubin (1977) haben in ihrem Artikel zum EM-Algorithmus auch gemischte Normalverteilungen untersucht.

Hierzu wird das vorhandene Problem in eine "unvollständige Datenlage" überführt ("incomplete- data- problem"). Gegeben sei der unbeobachtbare oder "missing data" Vektor

$z = (z_1^T, z_2^T, \dots, z_n^T)^T$ , wobei  $z_j^T$  ( $j=1, 2, \dots, n$ ) ein  $g$ -dimensionaler Indikatorvektor sei

---

<sup>6</sup> Der Exponent T steht für das Transponieren des Vektors.

und  $z_{ij} = (z_i)_j$  die Werte 1 oder 0 annehme, je nachdem, ob  $y_j$  der  $i$ 'ten Komponente der gemischten Normalverteilung entstammt, oder nicht.

Dann lässt sich ein "vollständiger Daten"-Vektor definieren als:

$$\mathbf{x} = \left( \mathbf{y}^T \mid \mathbf{z}^T \right)^T. \quad (4.15)$$

Die Loglikelihood-Funktion des vollständigen Datensatzes hat dann multinomiale Form:

$$\ln(L_c(\boldsymbol{\psi} \mid \mathbf{x})) = C + D \quad (4.16)$$

$$\text{wobei } C = \sum_{j=1}^n \sum_{i=1}^g z_{ij} \cdot \ln(p_i) \quad (4.16a)$$

$$\text{und } D = \sum_{j=1}^n \sum_{i=1}^g z_{ij} \cdot \ln(f_i(y_j; \theta_i)) \quad (4.16b)$$

Die Anwendung des EM-Algorithmus besteht aus 2 Iterationsschritten:

i) Im E-Schritt ("expectation step") wird der Erwartungswert der Loglikelihood der vollständigen Daten  $E[\ln(L_c)]$  berechnet.

ii) Im M-Schritt ("maximization step") wird die zuvor hergeleitete erwartete Loglikelihood maximiert.

Die Schritte i) und ii) werden wiederholt, bis der Abstand (oder der normierte Abstand) zwischen den erwarteten Loglikelihood-Werten des  $(t)$ 'en und des  $(t+1)$ 'en Schrittes extrem klein geworden ist (Konvergenzkriterium).

Die Funktion (4.16a) ist linear in den unbeobachtbaren Daten  $z_{ij}$ , daher gilt im  $(t+1)$ 'en Iterationsschritt, dass sich i) über die Berechnung des bedingten Erwartungswertes der Zufallsvariable  $Z_{ij}$ , welche zu  $z_{ij}$  gehört, durchführen lässt:

$$E_{\psi^{(t)}}(Z_{ij}|y) = P_{\psi^{(t)}}(Z_{ij} = 1|y) = \pi_{ij}^{(t)} = \frac{p_i^{(t)} f_i(y_j|\theta^{(t)})}{\sum_{i=1}^g p_i^{(t)} f_i(y_j|\theta^{(t)})}, \quad (4.17)$$

$$i=1,2,\dots,g, j=1,2,\dots,n$$

$\pi_{ji}^{(t)}$  ist die posteriori Wahrscheinlichkeit, dass das (j)'te Stichprobenelement mit dem Beobachtungswert  $y_j$  zur (i)'te Komponente der Mischung gehört.

Das Maximum der a priori Wahrscheinlichkeiten ergibt sich dann im M-Schritt intuitiv aus den a posteriori Wahrscheinlichkeiten:

$$p_i^{(t)} = \left( \sum_{j=1}^n \pi_{ij}^{(t)} \right) / n \quad (4.18)$$

Anmerkung: Die Schätzwerte für  $p_i$  lassen sich auch durch direkte Ableitung der Loglikelihood-Funktion der beobachteten Daten herleiten (McLachlan und Krishnan, 1997).

Die Gleichung (4.16b) der Loglikelihood-Funktion des vollständigen Datensatzes kann umgeformt werden zu:

$$(4.16b) = -\frac{1}{2} \cdot n \cdot \ln(2 \cdot \pi) - \frac{1}{2} \cdot \sum_{j=1}^n \sum_{i=1}^g z_{ij} \cdot \left( \ln(\sigma^2) + (y_j - \mu_i)^2 / \sigma^2 \right) \quad (4.19)$$

Somit erhält man die erwartete vollständige Loglikelihood-Funktion des (t+1)'ten Iterationsschrittes als:

$$E(\ln(L_c(\psi^{(t+1)}))) = \sum_{j=1}^n \sum_{i=1}^g \pi_{ij}^{(t)} \cdot \ln(p_i) - \frac{1}{2} \cdot n \cdot \ln(2 \cdot \pi) - \frac{1}{2} \cdot \sum_{j=1}^n \sum_{i=1}^g \pi_{ij}^{(t)} \cdot \left( \ln(\sigma^2) + (y_j - \mu_i)^2 / \sigma^2 \right) \quad (4.20)$$

Im E-Schritt (i) reicht es also wiederum aus, die a posteriori Wahrscheinlichkeiten zu berechnen und einzusetzen, im M-Schritt (ii) müssen  $\mu_1^{(t+1)}, \mu_2^{(t+1)}, \dots, \mu_g^{(t+1)}, \sigma^{(t+1)2}$  als Maximalwerte von (4.20) berechnet werden. Werden die partiellen Ableitungen von (4.20) gleich 0 gesetzt, so erhält man folgende Lösungen des (t+1)'en Maximierungsschrittes (die

einzelnen partiellen Ableitungen sind unabhängig voneinander, insbesondere fällt  $\sigma^2$  aus der gleich 0 gesetzten partiellen Ableitung nach  $\mu_i$  heraus, Emrich und Urfer (1999)):

$$\mu_i^{(t+1)} = \sum_{j=1}^n \pi_{ij}^{(t)} \cdot y_j / \sum_{j=1}^n \pi_{ij}^{(t)} \quad (i=1,2,\dots,g) \quad (4.21)$$

und

$$\sigma^{(t+1)2} = \frac{\sum_{j=1}^n \sum_{i=1}^g \pi_{ij}^{(t)} \cdot (y_j - \mu_i^{(t+1)})^2}{n}, \quad (4.22)$$

wobei die a posteriori Wahrscheinlichkeiten aus Gleichung (4.19) stammen.

Mit sukzessivem Einsetzen dieser Lösungen in die erwartete Likelihoodfunktion des vollständigen Datensatzes in jedem Iterationsschritt liefert der EM-Algorithmus Konvergenzpunkte, meist Maximalwerte, der ursprünglichen Loglikelihood-Funktion der beobachteten Daten.

### 4.3 EM-Algorithmus und Kovarianzen für Intervallkartierungsmodelle

#### Der EM-Algorithmus im Profil-Intervallkartierungsverfahren

Gegeben sei ein genetisches Modell eines F2-Kreuzungsexperimentes der Form:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_2 \\ \mathbf{G}_1 \\ \mathbf{G}_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \beta_0 + \begin{bmatrix} 1 & -1/2 \\ 0 & 1/2 \\ -1 & -1/2 \end{bmatrix} \begin{bmatrix} a \\ d \end{bmatrix} = \mathbf{1}_{3 \times 1} \beta_0 + \mathbf{D} \mathbf{E} \quad (4.23)$$

$\mathbf{E} = \begin{bmatrix} a \\ d \end{bmatrix}$

$G_i$  mit  $i=0,1,2$  seien unbeobachtbare QTL-Genotypwerte,  $\mathbf{D}$  sei genetische Designmatrix.

Die Datenauswertung der F2-Kreuzungsexperimentdaten erfolgt dann mit dem zusammengesetzten Intervallkartierungsmodell (CIM):

$$y_j = ax_j^* + dz_j^* + X_j\beta + \varepsilon_j \quad (4.24)$$

Es sei:  $y_j$ , ( $j=1,2,\dots,n$ ) Realisationen einer Zufallsvariable  $y$ , (z. B. Wuchshöhe der Maispflanzen),  $a$  und  $d$ : genetische Effekte,  $a$  ist additiver (phänotypischer)-,  $d$  ist Dominanzeffekt,  $x_j^*$  and  $z_j^*$  sind diskrete unbeobachtbare Zufallsvariablen mit:

$$x_j^* = \begin{cases} 1 & \text{für QTL mit Genotyp } Q_1/Q_1 \\ 0 & \text{für QTL mit Genotyp } Q_1/Q_2, j=1,2,\dots,n \\ -1 & \text{für QTL mit Genotyp } Q_2/Q_2 \end{cases} \quad (4.25)$$

und  $z_j^* = \begin{cases} 1/2 & \text{falls das QTL den Genotyp } Q_1/Q_2 \text{ besitzt,} \\ -1/2 & \text{sonst, } j=1,2,\dots,n \end{cases}$

$\beta$  ist Regressionskoeffizientenvektor der Dimension  $k$ ,  $X_j \subset X$ , wobei  $X$  eine Matrix ist, deren  $M$  Spalten die Markerrealisierungen der  $M$  Marker enthalten und  $\varepsilon_j \sim N(0, \sigma^2)$ .

Zudem wird ein einfaches Intervallkartierungsmodell (SIM), welches  $X_j\beta$  nicht enthält, betrachtet. Ziel der Datenanalyse ist es, das Schätzverhalten des EM-Algorithmus in der Praxis zu analysieren und dieses mit dem ECM sowie mit Regressionsschätzverfahren zu vergleichen.

Die Loglikelihood-Funktion der beobachteten Daten eines Intervallkartierungsansatzes hat für ein F2-Kreuzungsexperiment die folgende Form (Kao und Zeng 1997):

$$\ln L(\theta) = \sum_{j=1}^n \left[ \ln \left( \sum_{i=1}^3 p_{ji} \cdot f(y_j; \mu_{ji}, \sigma^2) \right) \right] \quad (4.26)$$

$f_{ji}$  ist Normaldichte mit Erwartungswert  $\mu_{ji}$  und Varianz  $\sigma^2$ , die Erwartungswerte der Mischung von Normalverteilungen lauten im F2-Ansatz:

$$\begin{aligned} \mu_{j1} &= a - d/2 + X_j\beta \\ \mu_{j2} &= d/2 + X_j\beta \\ \mu_{j3} &= -a - d/2 + X_j\beta \end{aligned} \tag{4.27}$$

Die  $p_{ji}$  sind einfache quadratische Funktionen von  $p=r_A/r^7$ , in Abhängigkeit der benachbarten Markergenotypen des Intervalls. Daher kann man (unter der Annahme, dass keine Doppel- Crossing over stattfinden), diese bedingten QTL-Genotypwahrscheinlichkeiten der Tabelle 4.1 entnehmen:

**Tabelle 4.0.1** Bedingte QTL-Genotypwahrscheinlichkeiten in Abhängigkeit der Markergenotypen des Untersuchungsintervalls: **A**: homozygot erster Genotyp, **H**: heterozygoter Genotyp, **B**: homozygot zweiter Genotyp, die Genorte sind diallel.

Markergenotypen	QTL-Genotypen			Stichproben- Größe <b>n</b>
	Q <sub>1</sub> /Q <sub>1</sub>	Q <sub>1</sub> /Q <sub>2</sub>	Q <sub>2</sub> /Q <sub>2</sub>	
AA <sup>8</sup> =M1	1	0	0	n <sub>1</sub>
AH=M2	1 - p	p	0	n <sub>2</sub>
AB=M3	(1 - p) <sup>2</sup>	2 · p · (1 - p)	p <sup>2</sup>	n <sub>3</sub>
HA=M4	p	1 - p	0	n <sub>4</sub>
HH <sup>9</sup> =M5	c · p · (1 - p)	1 - 2 · c · p · (1 - p)	c · p · (1 - p)	n <sub>5</sub>
HB=M6	0	1 - p	p	n <sub>6</sub>
BA=M7	p <sup>2</sup>	2 · p · (1 - p)	(1 - p) <sup>2</sup>	n <sub>7</sub>
BH=M8	0	p	1 - p	n <sub>8</sub>

<sup>7</sup> d. h. vom Verhältnis von Rekombinanten zwischen dem linken Marker und dem QTL zur Gesamtrekombinanzzahl des untersuchten Intervalls.

<sup>8</sup> Der erste Markergenotyp M1 hat Markergenotypkombination AA, d. h.: Am linken und rechten Markergenort des Intervalls realisiert sich der homozygot erste Genotyp. Der zweite Markergenotyp M2=AH bedeutet: Am linken Markergenort ist der homozygot erste, am rechten Markergenort der heterozygote Genotyp realisiert.

<sup>9</sup>  $c = (r^2) / (r^2 + (1-r)^2)$

BB=M9	0	0	1	$n_9$
-------	---	---	---	-------

Die Loglikelihood-Funktion ist somit ein Summand von Loglikelihood-Funktionen aus den "Populationen" mit Markergenotyp M1, M2, ..., Mr (den obigen Markergenotypen und zusätzlichen Genotypwahrscheinlichkeiten, wenn die Genotyprealisationen eines Markergens nicht gemessen werden konnten, vgl. Kap. 5.2 ("missing values"), und  $\sum_{s=1}^r n_s = n$ ),

d.h.:

$$\begin{aligned} \ln L(\theta) &= \sum_{j \in \{M1, M2, \dots, Mr\}} \left[ \ln \left( \sum_{i=1}^3 p_{ji} f(y_j; \mu_{ji}, \sigma^2) \right) \right] \\ &= \sum_{j \in \{M1\}} \left[ \ln \left( \sum_{i=1}^3 p_{ji} f(y_j; \mu_{ji}, \sigma^2) \right) \right] + \dots + \sum_{j \in \{Mr\}} \left[ \ln \left( \sum_{i=1}^3 p_{ji} f(y_j; \mu_{ji}, \sigma^2) \right) \right] \end{aligned} \tag{4.28}$$

Für dieses Modell leiteten Kao und Zeng (1997) eine Maximum-Likelihood-Schätzmethode mit einem ECM-Algorithmus (Meng und Rubin, (1993)) her. Emrich und Urfer (1999) verwendeten diesen Ansatz zur Herleitung direkter *expliziter* Formeln zur Maximum Likelihood-Schätzung mit dem EM-Algorithmus. Die erwartete Loglikelihood-Funktion des E-Schrittes, die expliziten Formeln der Maximierung des M-Schrittes des EM-Algorithmus, sowie der Gradientenvektor von  $\ln(L_c)$  werden im Appendix M aufgeführt.

Beide Schätzverfahren werden einander in der Datenanalyse gegenüber gestellt.

Die in Kapitel 4.2 vorgestellte Likelihood-Funktion der "vollständigen Daten"  $L_c$  des EM-Ansatzes, welche zur Berechnung der asymptotischen Kovarianzmatrix von EM-ML-Schätzern benötigt wird, kann als die Likelihood eines hierarchischen Experimentes angesehen werden (Louis (1982) brachte ein Beispiel für eine gemischte Normalverteilung aus 2 Populationen):

- Die Realisationen einer unbeobachtbaren Zufallsvariable  $z_{ji}$  ( $j=1,2,\dots,n$ ) werden durch ein trinomiales Zufallsexperiment erzeugt.
- Den drei Realisationen der Zufallsvariable werden 3 mögliche QTL-Genotypen zugeordnet, die zugehörigen Beobachtungen besitzen Normaldichten mit Erwartungswert  $\mu_{ji}$  und Varianz  $\sigma^2$ . Dann ergibt sich die Loglikelihood-Funktion als:

$$\ln(L_c) = \sum_{j=1}^n \sum_{i=1}^3 \left[ z_{ji} \cdot \ln(p_{ji}) + z_{ji} \cdot \ln\left(f(y_j | \mu_{ji}, \sigma^2)\right) \right] \quad (4.29)$$

$z_{ji}$ : unbeobachtbare Indexvariable, mit  $z_{ji} = \begin{cases} 1 & \text{falls QTL-Genotyp } i \text{ vorliegt} \\ 0 & \text{sonst} \end{cases}$

Da die vollständige Loglikelihood-Funktion linear in  $z_{ji}$  ist, wird der Erwartungswert des (t+1)'ten Iterationsschrittes berechnet als:

$$Q(\theta | \theta^{(t)}) = E(\ln(L_c)) = \sum_{j=1}^n \sum_{i=1}^3 \left[ \pi_{ji}^{(t)} \cdot \ln(p_{ji}) + \pi_{ji}^{(t)} \cdot \ln\left(f(y_j | \mu_{ji}, \sigma^2)\right) \right] \quad (4.30)$$

Diese Funktion wird in Profil-Intervallkartierungsverfahren in jedem Iterationsschritt über  $\theta = (a, d, \sigma^2, \beta_1, \beta_2, \dots, \beta_k)$  maximiert:

Die expliziten Formeln für diesen EM-Ansatz werden im Appendix M angegeben in Form des linearen Gleichungssystems  $A_t \cdot \theta^{(t+1)} = b_t$ . Dieses lineare Gleichungssystem lässt sich aus dem gleich Null gesetzten linearen Gleichungssystem der partiellen Ableitungen nach allen Komponenten des Parametervektors  $\theta$  herleiten. Man erhält dann im (t+1)'en Maximierungsschritt des EM-Algorithmus eines Intervallkartierungs-Profilverfahrens die Lösung:

$$\theta^{(t+1)} = A_t^{-1} \cdot b_t, \text{ mit } A_t = A[\theta^{(t)}], \text{ } b_t = b[\theta^{(t)}]. \quad (4.31)$$

und für  $t \rightarrow \infty$   $\hat{\theta} = A^{-1} \cdot b$

wobei 
$$\hat{\theta} = \lim_{t \rightarrow \infty} \theta^{(t)}, A = \lim_{t \rightarrow \infty} A_t, b = \lim_{t \rightarrow \infty} b_t,$$

Der EM-Algorithmus liefert einen Maximum-Likelihood-Schätzer, wenn die Folge der Differenzen von  $Q(\theta|\theta^{(t+1)}) - Q(\theta|\theta^{(t)})$  gegen Null geht. Der Konvergenzpunkt  $\hat{\theta}$  ist dann sowohl Maximum der vollständigen Loglikelihood-Funktion  $\ln(L_c)$  als auch der beobachteten Loglikelihood-Funktion  $\ln(L)$  und somit Maximum Likelihood-Schätzer für  $\theta = (a, d, \sigma^2, \beta_1, \beta_2, \dots, \beta_k)$ .

### Der EMV-Algorithmus für Intervallkartierungsverfahren

Die beobachtete Loglikelihood-Funktion  $\ln(L)$  und die zugehörige vollständige Loglikelihood-Funktion  $\ln(L_c)$  können nicht nur als Funktionen von  $\theta = (a, d, \sigma^2, \beta_1, \beta_2, \dots, \beta_k)$  betrachtet werden, sondern auch als Funktionen von  $\theta = (p, a, d, \sigma^2, \beta_1, \beta_2, \dots, \beta_k)$ . Dann wird im Intervallkartierungsansatz das Maximum nicht mehr in jedem Markerintervall über ein diskretes Profil von Funktionswerten der Loglikelihood- (oder Lod-Score-) Funktion für eine Anzahl fester vermuteter QTL-Positionen berechnet (vgl. Kap 2), sondern in jedem Markerintervall wird ein Maximum der Loglikelihood-Funktion über alle Parameter in  $\theta = (p, a, d, \sigma^2, \beta_1, \beta_2, \dots, \beta_k)$  berechnet.

Profilverfahren der Intervallkartierung reduzieren also den m-dimensionalen Parameterraum auf einen (m-1)-dimensionalen Parameterraum und betrachten dann den Grat der Loglikelihood-Funktion bezüglich des m' ten Parameters. Die diskrete Grat- oder Profilkurve an festen Werten des (m)' ten (Positions-) Parameters wird graphisch maximiert, d. h. es wird der höchste diskrete Wert der Funktion im Intervall ermittelt.

Statt dieses Ansatzes kann man jedoch auch eine direkte Maximierung mit dem EM-Algorithmus über alle Parameter (bzw. Variablen der Loglikelihood-Funktion) durchführen. Dieses Verfahren wird im Sinne des "vollständigen" m-dimensionalen Parameter-raums als EMV-Algorithmus bezeichnet und findet natürlich genau einmal in jedem Inter-vall Anwendung.

Die Berechnung einer beobachteten Informationsmatrix  $I_Y$  nach Louis (1982, Kap 4.1) kann dann ohne weiteres erfolgen. In Profilverfahren wird demgegenüber  $p$  als ein fester vorab bestimmter Schätzer der Position des QTL angesehen.

Betrachtet man den Maximierungsschritt des EM-Algorithmus bezüglich  $p$ , so ist nur der von  $p$  abhängige Teil der Funktion  $Q(p|\theta^{(t)}) = \sum_{j=1}^n \sum_{i=1}^3 \pi_{ji}^{(t)} \cdot \ln(p_{ji})$  von Interesse, da der zweite Summand von  $Q(\cdot)$  unabhängig von  $p$  ist! Die  $\pi_{ji}^{(t)}$  sind in jedem Iterationsschritt *fixe* Konstanten.

Der Maximierungsschritt des EM-Algorithmus wird dann erweitert, indem dem obigen linearen Gleichungssystem partieller Ableitungen die gleich 0 gesetzte partielle Ableitung nach  $p$  zugefügt wird.

$$\frac{\partial Q}{\partial p} = \frac{\partial}{\partial p} \left( \sum_{j \in M1} \sum_{i=1}^3 \ln(p_{ji}) \cdot \pi_{ji}^{(t)} + \sum_{j \in M2} \sum_{i=1}^3 \ln(p_{ji}) \cdot \pi_{ji}^{(t)} + \dots + \sum_{j \in Mr} \sum_{i=1}^3 \ln(p_{ji}) \cdot \pi_{ji}^{(t)} \right) = 0 \quad (4.32)$$

Der Tabelle 4.1 ist zu entnehmen, dass beispielsweise gilt:

$$\sum_{j \in M2} \ln(p_{j1}) = \sum_{j \in M8} \ln(p_{j3}) \quad (4.33)$$

Somit lässt sich  $Q(p|\theta^{(t)})$  darstellen als eine Funktion der Form

$$\begin{aligned}
 Q(p|\theta^{(t)}) = & \ln(1) \cdot \left( \sum_{j \in M1} \pi_{j1}^{(t)} + \sum_{j \in M9} \pi_{j3}^{(t)} \right) \\
 & + \ln(p) \cdot \left( \sum_{j \in M2} \pi_{j2}^{(t)} + \sum_{j \in M8} \pi_{j2}^{(t)} + \sum_{j \in M4} \pi_{j1}^{(t)} + \sum_{j \in M6} \pi_{j3}^{(t)} \right) \\
 & + \ln(1-p) \cdot \left( \sum_{j \in M2} \pi_{j1}^{(t)} + \sum_{j \in M4} \pi_{j2}^{(t)} + \sum_{j \in M6} \pi_{j2}^{(t)} + \sum_{j \in M8} \pi_{j3}^{(t)} \right) \\
 & + \sum_{\kappa} v_{\kappa}(p) \cdot \left( \sum_{j \in M_{v_{\kappa}}} \pi_{ji}^{(t)} \right)
 \end{aligned} \tag{4.34}$$

Mit  $v_{\kappa}(p)$  seien die unterscheidbaren bedingten QTL-Wahrscheinlichkeiten bezeichnet, welchen die zugehörigen a posteriori Summen über die jeweiligen Marker-QTL-Genotypkombinationen zugeordnet werden.

Diese Funktion lässt sich in jedem Iterationsschritt mit festen a posteriori Wahrscheinlichkeiten berechnen, so dass die direkte Maximierung der Funktion  $Q(p|\theta^{(t)})$  nach  $p$  beispielsweise mit dem Newton-Verfahren oder einer Quasi-Newton Maximierung (S-Plus) durchgeführt werden kann. Dies geschieht in der praktischen Auswertung durch einen zusätzlichen numerischen Maximierungsschritt von  $Q(p|\theta^{(t)})$ .

Man erhält somit den EMV-Schätzer:  $\hat{\theta} = (\hat{p}, \hat{a}, \hat{d}, \hat{\sigma}^2, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)$ .

Die asymptotische Kovarianzmatrix für  $\hat{\theta}$  kann dann mit Hilfe der in Kapitel 4.1 (Gleichungen (4.9), (4.10)) beschriebenen Herleitung der Informationsmatrix beobachteter Daten eines EM-ML-Schätzers berechnet werden. Die Berechnung der Informationsmatrix der beobachteten Daten aus der Differenz der Informationsmatrizen der vollständigen Daten und der "fehlenden" Daten ist recht aufwändig, und wurde von Kao und Zeng (1997) für ihre ECM-Schätzer durchgeführt (unter der Annahme, dass der Lageparameter vorab

bestimmt wurde und somit vorliegt). Die Formeln der Informationsmatrizen werden in der Datenanalyse benötigt und daher im Appendix M angegeben.

## 5. Experimentelle Daten

### 5.1 Die Daten der Feldexperimente

Prof. H. F. Utz von der Universität Hohenheim (Fachgebiet Angewandte Genetik und Pflanzenzüchtung) hat Prof. W. Urfer (Fachbereich Statistik, Universität Dortmund) freundlicher



Weise QTL-Kartierungsdaten von Mais-Pflanzenexperimenten zur Verfügung gestellt. Diese wurden bereits in den letzten Jahren mehrfach unter verschiedenen Gesichtspunkten ausgewertet (z. B. Schön et. al. (1994) und Melchinger et. al. (1998)). Fichtner (2000) hat sich am Fachbereich Statistik in Dortmund im Rahmen seiner Diplomarbeit mit Fragestellungen der Versuchsplanung der genannten Feldversuche be-

schäftigt.

Die in Dortmund verwendeten Daten stimmen teilweise überein mit den in Melchinger (1998) analysierten QTL-Kartierungsdaten. In der dortigen Studie wurden zwei Maissorten ("early maturing European elite flint inbreds") KW1265 und D146, bezeichnet als P1 und P2, gekreuzt. Die Pflanzenexperimente fanden an verschiedenen Standorten in Süddeutschland in den Jahren 1990/1991 statt (unter Verwendung eines 40 zu 10  $\alpha$ -Designs, vgl. auch Urfer et. al. (1999)).

344 F2 Pflanzen wurden (wie in Schön (1994) beschrieben,) für 89 bekannte Markergenorte genotypisiert. Die Lage der 89 Markergenorte im Genom der verwendeten Sorten ist auf Grund von speziellen Schätzverfahren (Kopplungsanalyse mittels Mapmaker 3.0, (Lander et. al. (1987), Lincoln et. al. (1992) und (1993)) bestimmt worden und wird als bekannt vorausgesetzt.

Eine schematische Darstellung der Markergenorte wird in Abbildung 6.1 in Kapitel 6 angegeben.

M1, M2, ... ,M89 sind also die Markergenorte auf den 10 Chromosomen, die Intervalle werden durch senkrechte Striche voneinander getrennt und die Intervallgröße wird in jedem Intervall in cM angegeben. Die 10 Chromosomen werden in vereinfachter Form und Struktur gezeigt, wobei jedes Chromosom in Abbildung 6.1 am ersten geschätzten Markergenort des Chromosoms beginnt und mit dem letzten bekannten Marker Locus endet.

Es handelt sich um sog. RFLP (restriction fragment length polymorphism)-Marker, deren Genotypen sich nicht im Phänotyp der Pflanzen, sondern biochemisch im Labor (auf Grund von Aufspaltungen der DNS-Sequenzen an festen Punkten der Chromosome mittels bestimmter Enzyme) feststellen lassen. Die untersuchten landwirtschaftlichen Merkmale der Pflanzen wurden an Mittelwerten von Nachkommensgruppen aus Testkreuzungen mit einer unabhängigen weiteren Maissorte (KW4115) als Pollenlieferant und den F3- bzw. den Parentalpflanzen erhoben. Diese Mittelwerte werden - wie in der Praxis üblich - als Realisationen von F2-Pflanzenmerkmalen angesehen. Die Nachkommen wurden auf folgende agronomischen Eigenschaften hin untersucht: Kornertrag pro Hektar, Feuchtigkeitsanteil im Ertrag, das Korngewicht und die Wuchshöhe der Pflanzen vom Boden bis zum Blütenansatz ("*tassel branch*"). Diese letztere Eigenschaft war die in den in Dortmund vorliegenden Daten agronomische Eigenschaft. Im vorliegenden Datensatz sind 344 "Pflanzen" mit ihren Merkmals- und Markergenotyprealisationen erfasst. Die Merkmalsrealisationen enthielten keine fehlenden "Beobachtungen" ("missing values"), allerdings gab es fehlende Markergenotyprealisationen. Aufgrund praktischer und technischer Problemen innerhalb der Feldversuche und im Labor konnten 6,7% der Markerrealisationen nicht erhoben werden. Abbildung 5.1 zeigt, dass in den meisten Markerrealisationsvektoren mindestens 20 fehlende Werte vorhanden sind. Daher muss vor der Datenauswertung entschieden werden, wie die "missing values" zu behandeln sind.

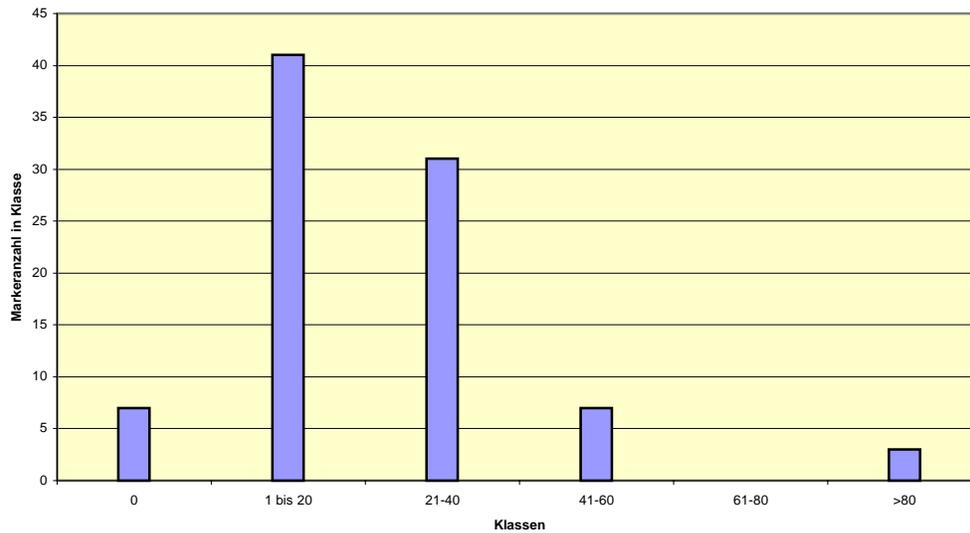


Abbildung 5.1: Anzahl von Markern mit fehlenden Markerrealisationen

## 5.2 Behandlung der fehlenden Beobachtungen ("missing values")

Vor Beginn der Datenauswertung muss entschieden werden, wie fehlende Realisationen des beobachteten Merkmals und der Marker behandelt werden sollen. Fehlende Merkmalsbeobachtungen wurden bereits bei der früher erfolgten Auswertung aussortiert. Das in Kapitel 4 (Gl. (4.23)) beschriebene genetische Modell erfasst eine Teilmenge der 89 Markerrealisationsvektoren als Kofaktoren in einer Kofaktormatrix. Die fehlenden Markerrealisationen der ausgewählten Kofaktoren werden - analog zu anderen Softwareprogrammen wie Plabqtl (Utz und Melchinger (1996)) - durch ihre Erwartungswerte ersetzt, dieses Vorgehen wurde auch von Little (1992) diskutiert.

Alle anderen Markerrealisationen werden in dem Intervallkartierungsverfahren ausschließlich zur Berechnung der QTL (quantitative trait locus)-Genotypwahrscheinlichkeiten ( $p_{ji}$ ) in Abhängigkeit der Markergenotypen in den einzelnen Intervallen benötigt. Die ( $p_{ji}$ ) hängen von  $p = r_A / r$  ab, wobei  $r_A$  die Rekombinationsfrequenz zwischen dem linken Marker und

dem QTL und  $r$  die Rekombinationsfrequenz zwischen dem linken und rechten Marker des Intervalls darstellt.

Es wurde entschieden, die QTL-Genotypwahrscheinlichkeiten mit fehlenden rechten, linken oder beidseitigen Markerrealisationen mittels des Verfahrens von Fisch et. al. (1996) zu berechnen. Dieser Ansatz verwendet jedoch keine Funktion in Abhängigkeit von  $p = r_A/r$ , sondern berechnet die Wahrscheinlichkeiten mit Hilfe einer Matrix von Übergangswahrscheinlichkeiten (vgl. Fahrmeir (1981)) aus den Rekombinationsfrequenzen zwischen dem linken Marker ( $r_1$ ) und dem QTL und dem rechten Marker und dem QTL ( $r_2$ ). Somit wird in diesem Ansatz - im Gegensatz zu den anderen hier verwendeten Methoden - die Existenz von Doppel-Crossing over angenommen, so dass es sich für große Markerintervalle tatsächlich nur um eine Schätzung der entsprechenden Wahrscheinlichkeiten handelt. Diese Wahrscheinlichkeiten müssen nun zur Integration in die verwendete EM-Methode als Funktionen von  $p$  dargestellt werden. Dies erfolgte über einen Regressionsansatz der Form  $p_{ji} = f(p) = a \cdot p^2 + b \cdot p + c$ . Die Wahl dieses Modells bietet sich an, da auch die anderen QTL-Genotypwahrscheinlichkeiten als einfache quadratische Gleichungen von  $p$  beschrieben werden können. Die Schätzung der Parameter  $a$ ,  $b$  und  $c$  muss in Abhängigkeit der Intervallgrößen (bzw. der Rekombinationsfrequenz  $r$ ) für alle möglichen Kombinationen von fehlenden Markergenotypen und QTL-Genotypen erfolgen, da insbesondere der doppelt heterozygote Markergenotyp HH (mit Allelen  $A_1, A_2$  am linken Markerloкус und  $B_1, B_2$  am rechten) über  $c = r^2 / (r^2 + (1-r)^2)$  von der Intervallgröße abhängt. Das oben genannte polynomiale Modell ergibt eine genaue Anpassung für die rechten Randwerte der Intervalle, aus denen  $r$  gewählt wird; tatsächlich ist für kleine Markerintervalle  $R^2 = 1$ ; nur im Fall von  $r > 0.5$  und einem rechtsseitig fehlenden Marker ergibt sich eine schlechtere Anpassung, welche mit steigender Intervallgröße verstärkt wird, wie man im zweiten Teil der Tabelle 5.1 anhand des Bestimmtheitsmaßes  $R^2$  ablesen kann.

Fehlen beide angrenzenden Marker, so gilt für alle Rekombinationsfrequenzen  $r$  die "unabhängige" Genotypverteilung (0,25;0,5;0,25). Mit diesen beiden Tabellen sind alle möglichen Markergenotypkombinationen für fehlende Randmarker abgedeckt, da eine Symmetrie der Koeffizienten  $a$ ,  $b$  und  $c$  für verschiedene Markergenotypkombinationen vorliegt. So gilt für die Koeffizienten (wie in der Tabelle angegeben)  $a=0,0004$ ,  $b=-0,0004$  und  $c=1$ , dass diese Kombination für MQ1A als auch für MQ3B gilt. Zur visuellen Verdeutlichung wird die Funktion/ Verteilung der QTL-Genotypen  $f(p)$  für  $r=0,3$  und für den linken Markergenotypen A1A1 und mit rechtem unbekanntem Markergenotypen in der folgenden Abbildung 5.2 angegeben:

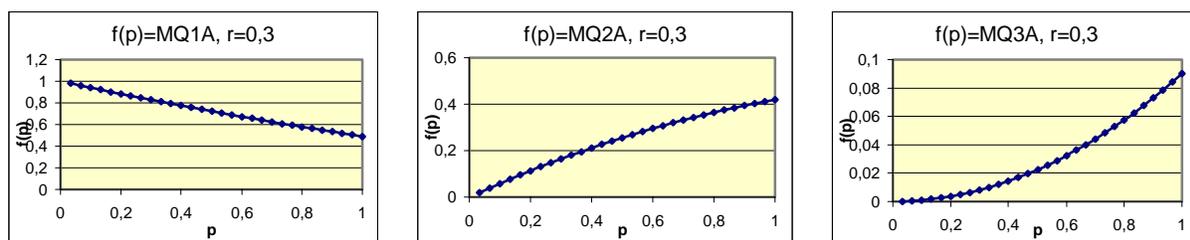


Abbildung 5.2: Graphische Darstellung der Wahrscheinlichkeitsverteilung  $f(p)$  für MM, B1B1

Tabelle 5.1: Koeffizienten für  $f(p)$  zur Berechnung der bed. Genotypw. bei fehlenden Markerreal.

r und Intervallgröße	MQ1A <sup>10</sup> = MQ3B			MQ2A = MQ2B			MQ3A = MQ1B			MQ1A = MQ3B			MQ2A = MQ2B		
	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
bis 0,02	0,0004	-0,0004	1	-0,0008	0,0200	0	0,0004	0	0	-0,0004	0,0200	0	0,0008	-0,0400	1
bis 0,05	0,0025	-0,1000	1	-0,0050	0,1000	0	0,0025	0	0	-0,0025	0,0500	0	0,0050	-0,1000	1
bis 0,1	0,0100	-0,2000	1	-0,0200	0,2000	0	0,0100	0	0	-0,0100	0,1000	0	0,0200	-0,2000	1

<sup>10</sup> M steht für einen fehlenden Genotyp an einem Markergenort (d. h. beide Allele sind unbekannt). A=homozygot erster Genotyp für linken oder rechten Markergenort, je nach aufgeführter Reihenfolge, B=homozygot zweiter Markergenotyp für linken oder rechten Markergenort, H=heterozygoter Genotyp für links oder rechts, Q1, Q3, Q2 sind homozygot erster, zweiter bzw. heterozygoter Markergenotyp für den QTL.

bis 0,2	0,0400	-0,4000	1	-0,0800	0,4000	0	0,0400	0	0	-0,0400	0,2000	0	0,0800	-0,4000	1
bis 0,3	0,0900	-0,6000	1	-0,1800	0,6000	0	0,0900	0	0	-0,0900	0,3000	0	0,1800	-0,6000	1
bis 0,4	0,1600	-0,8000	1	-0,3200	0,8000	0	0,1600	0	0	-0,1600	0,4000	0	0,3200	-0,8000	1
bis 0,5	0,2500	-1,0000	1	-0,5000	1,0000	0	0,2500	0	0	-0,2500	0,5000	0	0,5000	-1,0000	1
0,5 und bis cM 60	0,2500	-1,0000	1	-0,5000	1,0000	0	0,2500	0	0	-0,2500	0,5000	0	0,5000	-1,0000	1
0,5 und bis cM 70	0,2500	-1,0000	1	-0,5000	1,0000	0	0,2500	0	0	-0,2500	0,5000	0	0,5000	-1,0000	1
0,5 und bis cM	0,2500	-1,0000	1	-0,5000	1,0000	0	0,2500	0	0	-0,2500	0,5000	0	0,5000	-1,0000	1

r	AQ1M = BQ3M			AQ2M = MQ2B			AQ3M = BQ1M			HQ1M = HQ3M			HQ2M		
	a=	b	c	a	b	c	a	b	c	a	b	c	a	b	c
u. Int.															
gr.	a=	b	c	a	b	c	a	b	c	a	b	c	a	b	c
0,02	0,0004	0,0392	0,9604	-0,0008	-0,0384	0,0392	0,0004	-0,0008	0,0004	-0,0004	-0,0192	0,0192	0,0008	0,0384	0,9608
0,05	0,0025	0,0950	0,9025	-0,0050	-0,0900	0,0950	0,0025	-0,0050	0,0025	-0,0025	-0,0450	0,0475	0,0050	0,0900	0,9050
0,1	0,0100	0,1800	0,8100	-0,0200	-0,1600	0,1800	0,0100	-0,0200	0,0100	-0,0100	-0,0800	0,0900	0,0200	0,1600	0,8200
0,2	0,0400	0,3200	0,6400	-0,0800	-0,2400	0,3200	0,0400	-0,0800	0,0400	-0,0400	-0,1200	0,1600	0,0800	0,2400	0,6800
0,3	0,0900	0,4140	0,4761	-0,1800	-0,2280	0,4278	0,0900	-0,1860	0,0961	-0,0900	-0,1140	0,2139	0,1800	0,2280	0,5722
0,4	0,1600	0,4720	0,3481	-0,3200	-0,1440	0,4838	0,1600	-0,3280	0,1681	-0,1600	-0,0720	0,2419	0,3200	0,1440	0,5162
0,5	0,2500	0,5000	0,2500	-0,5000	0,0000	0,5000	0,2500	-0,5000	0,2500	-0,2500	0,0000	0,2500	0,5000	0,0000	0,5000
0,5cM															
bis 60	0,6064	0,0322	0,2404	-0,6824	0,3367	0,4667	0,0761	-0,3690	0,2929	-0,3412	0,1684	0,2334	0,6824	-0,3367	0,5333
R <sup>2</sup>		0,9808			0,9546			0,9844			0,9546			0,9546	
0,5bis															
cM 70	1,0324	-0,5530	0,3047	-0,7806	0,5489	0,4351	-0,2518	0,0041	0,2602	-0,3903	0,2745	0,2175	0,7806	-0,5489	0,5649
R <sup>2</sup>		0,9139			0,8077			0,9832			0,8077			0,8077	
0,5 ab															
70 cM	1,2471	-0,9108	0,3596	-0,7409	0,5817	0,4260	-0,5061	0,3291	0,2144	-0,3705	0,2909	0,2130	0,7409	-0,5817	0,5740
R <sup>2</sup>		0,7826			0,6095			0,9521			0,6095			0,6095	

### 5.3 Auswahl der Markerkofaktoren des statistischen Modells

Ursprünglich war beabsichtigt, die selbstprogrammierte Datenanalyse mit den Ergebnissen anderer Programmpakete zu vergleichen. Daher wurden 8 Kofaktoren ausgewählt, welche als sinnvolle Wahl für Modell-Kofaktoren sowohl im Regressionsansatz (Plabqtl, Utz und

Melchinger (1996)) als auch im ECM-Ansatz (QTL-Cartographer, Basten et. al. (1994)) angegeben wurden. Plabqtl trifft die Auswahl der Kofaktoren mit Hilfe von schrittweiser Regression, wobei die partiellen F-Statistiken unter Hinzunahme jeweils eines weiteren Markers betrachtet werden, QTL-Cartographer hingegen wählt Kofaktoren für den CIM-Ansatz auf der Basis einer SIM-Regressionsvoranalyse aus.

Dementsprechend wurden die Marker mit den Nummern 5, 10, 28, 39, 41, 47, 61 und 86 ausgewählt. Außerdem wurden einige Marker mit dominant-rezessiven Genotyprealisationen nicht in die Analyse integriert, da der Informationsgehalt dominant-rezessiver Markerrealisationen geringer ist als der Informationsgehalt von kodominanten Realisationen (vgl. Lorieux et. al. 1995 (1) und (2)) und für diesen Zweck zusätzlich zu den QTL-Genotypverteilungen in Tabelle 4.1 weitere QTL-Genotypverteilungen hätten berechnet werden müssen. Somit wurden die Intervalle, welche an Marker Nr. 6, 9, 29, 30, 53, 54, 58 grenzten, so zusammengefasst, dass die genannten Marker in den Intervallen liegen.

Die Marker mit Nr. 17, 49 und 67 wurden bereits in früheren Auswertungen nicht in die vorgegebene Markerkarte integriert.

## 6 Datenauswertung und Vergleich verschiedener Kartierungsverfahren

Die in Abbildung 6.1 dargestellte schematische Markerkarte soll eine Hilfestellung bei der Interpretation der folgenden Auswertung liefern. Über den schematischen Chromosomen stehen Intervallnummern, im Inneren der Markerintervalle werden die Intervallgrößen in cM angegeben.

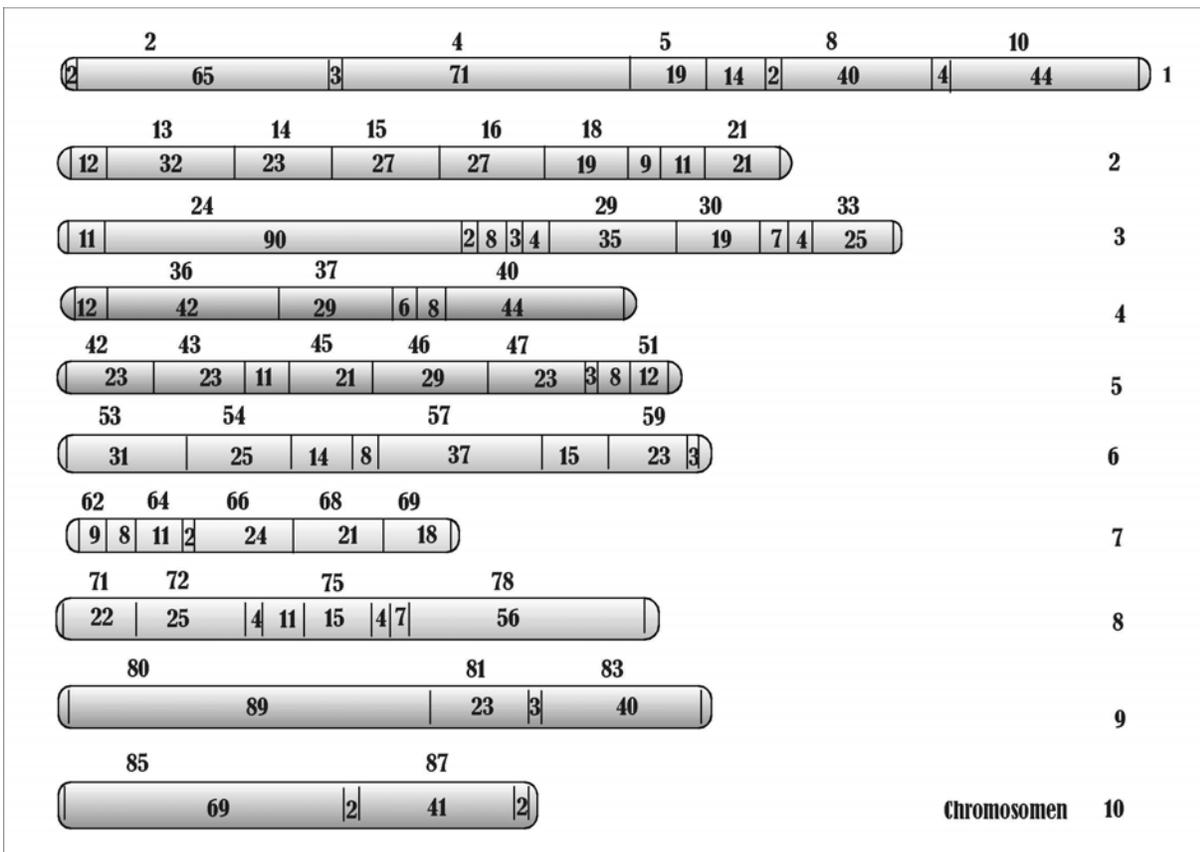


Abbildung 6.1: Schematische Darstellung der Markerintervalle auf den 10 Chromosomen

Die Datenauswertung wurde parallel mit verschiedenen QTL-Kartierungsansätzen durchgeführt. Zum einen wurden Standardverfahren verwendet, welche dann mit dem hier entwickelten EMV-Ansatz ergänzt werden können. Der auf Regressionsanalyse beruhende

Ansatz von Haley and Knott (1994) wurde für ein SIM und ein CIM-Modell durchgeführt. Des Weiteren wurden diese SIM - und CIM-Modelle verwendet, um Maximum-Likelihood-Schätzer mittels des EM-Algorithmus zu berechnen. Der EM-Algorithmus wurde ebenso wie der EMV-Algorithmus mit Hilfe von Lösungen linearer Gleichungssystemen durchgeführt (s. Kap. 4.3.2 und Appendix M). Diese Lösungen habe ich zudem mit denen des ECM-Algorithmus (Kao, Zeng 1997) verglichen.

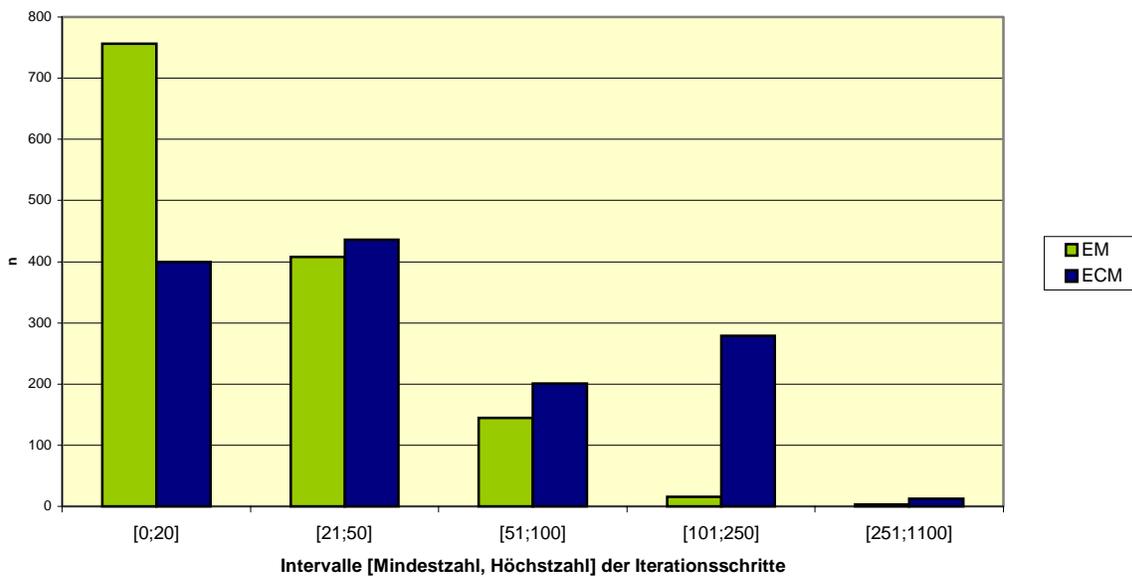
### **6.1 Der EM-und der ECM-Algorithmus**

Kao und Zeng (1997) beschreiben ein iteratives Verfahren zur Berechnung von Maximum-Likelihood-Schätzern für ein QTL-Kartierungsmodell, welches zufällige und feste Effekte enthält und einen ECM-Algorithmus zur Parameterschätzung integriert (Meng & Rubin (1993)). Simultan zu dem dort vorgegebenen Datenproblem kann man auch einen direkten EM-Algorithmus herleiten, welcher auf der Lösung von linearen Gleichungssystemen der Form  $A\theta = b$  im Maximierungs-Teil jedes Iterationsschritts beruht (Emrich & Urfer (1999)). Die einzelnen Schritte sind hierbei etwas aufwändiger zu berechnen, da eine Invertierung einer Matrix der Dimension  $(k+2)$  durchgeführt wird, während der ECM-Algorithmus mit der Invertierung einer  $k$ -dimensionalen Matrix auskommt.

Allerdings zeichnet sich der ECM-Algorithmus durch ein langsames Konvergenzverhalten aus. Davon abgesehen, hatten jedoch beide Verfahren ähnliche Konvergenzbereiche und führten im vorliegenden Datenproblem zu denselben Maximum-Likelihood-Schätzern.

Das langsamere Konvergenzverhalten des ECM-Algorithmus diskutiere ich hier anhand der Anzahl von Iterationsschritten, welche bei einem vorgegebenen festen Anfangs-Startwert für die iterativen Verfahren benötigt wurden. Dieser Startwert wurde anhand einer Voranalyse ausgewählt, und hatte insbesondere die Eigenschaft, stabil gegen einen Maximum-Likelihood-Schätzer zu konvergieren, ohne frühzeitig in Folge von Instabilitäten (meist

indefiniten Matrizen) abzubrechen. Die Voranalyse wurde mit einigen hundert Startwerten auf Chromosom 1 durchgeführt und zeigte, dass der ECM-Algorithmus weniger stabil ist, als der EM-Algorithmus (d. h. es kam häufiger zu frühzeitigen Abbrüchen in Iterationsschritten). Den Vergleich der Anzahl von Iterationsschritten bis zur Konvergenz an den 1329 untersuchten mutmaßlichen Gen-Positionen auf den 10 Chromosomen (im Abstand von 1 cM) diskutiere ich zunächst anhand eines Histogramms der benötigten Anzahl von Iterationsschritten auf allen Chromosomen.



**Abbildung 6.2:** Anzahl n der berechneten Genpositionen, in denen auf der Abzisse angegebenen Intervallen, hell : EM-Verfahren, dunkel : ECM-Verfahren

**Tabelle 6.1:** Anteile von Schätzungen in Iterationsgruppen

Anzahl von Iterationen	EM	ECM
[0;20]	0,569	0,301
[21;50]	0,307	0,328
[51;100]	0,109	0,151

[101;250]	0,012	0,210
[251;1100]	0,002	0,010

In Abb. 6.2 und in Tabelle 6.1 wird verglichen, wie häufig beide Verfahren bis zu 20, 21 bis 50, 51 bis 100 etc. Iterationsschritte (n) benötigen, um bei gleichem Abbruchkriterium zu konvergieren. Der EM-Algorithmus benötigt für den vorliegenden Datensatz in 57% der Schätzungen weniger als 20 Iterationsschritte, um zu konvergieren, während der ECM-Algorithmus in dieser Klasse mit bis zu 20 Iterationsschritten nur 30% der Schätzungen anbietet, in der Klasse von 21 bis 50 Iterationsschritten haben beide Verfahren etwa 30% ihrer Berechnungen. Über 100 Iterationen werden bei Verwendung des EM-Algorithmus nur in 1,4 % der Fälle benötigt, bei Verwendung des ECM-Algorithmus dagegen in 22 % der Fälle.

## **6.2 Vergleich von Maximum-Likelihood-Ansätzen anhand tabellarischer und graphischer Auswertung der QTL-Kartierungsdaten**

**Vorbemerkung:** Die meisten Intervallkartierungsverfahren betrachten Lod-Score-Profile über das Genom.

Der Lod-Score ist definiert durch:

$$\text{Lod-Score} = \frac{2}{\ln(10)} \cdot \ln(\text{LQ}) = 2 \cdot \log_{10} \left( \frac{L}{L_0} \right), \quad (6.10)$$

wobei  $\ln(\text{LQ})$  der logarithmierte Likelihoodquotient ist. Dieser ist gegeben durch:

$$\ln(\text{LQ}) = \ln(L) - \ln(L_0), \quad (6.11)$$

wobei hier (vgl. Tab. A1)

$$\ln(L_0) = \begin{cases} -2386,3375 & \text{für IM und} \\ -1882,813 & \text{für CIM} \end{cases} \quad (6.12)$$

mit

$$\begin{aligned} \ln(L_0) &= \max_{\theta \in \Theta_0} \ln(L) \Big|_{H_0} = \max_{\theta \in \Theta_0} \sum_{j=1}^n \ln(f_j) \cdot (p_{j1} + p_{j2} + p_{j3}) \\ &= \max_{\theta \in \Theta_0} \sum_{j=1}^n \ln(f_j), \end{aligned} \quad (6.13)$$

da unter  $H_0$ :  $a=0$  und  $d=0$  gilt:

$$\begin{aligned} \mu_{j1} &= \mu_{j2} = \mu_{j3}, \\ \text{also auch } f_{j1} &= f_{j2} = f_{j3} = f_j \end{aligned} \quad (6.14)$$

Hieraus folgen die oben angegebenen numerischen Werte.

Alternativ zur Berechnung von Loglikelihood-Profilen über das gesamte Genom ist es auch möglich, in jedem Intervall den in Kapitel 4 hergeleiteten EMV-Algorithmus zu verwenden, um nach jeweiligen Maximalwerten zu suchen. Dabei finden sich dann einzelne Extremwerte, welche dann daraufhin überprüft werden müssen, ob es sich um Maximalwerte handelt. Im Prinzip gilt auch für Lod-Score und Loglikelihood-Profile, dass diese aus Sattelpunkten bestehen können. Selbst am Maximum einer diskreten Profildfunktion könnte ein Sattelpunkt (hinsichtlich einer oder mehrerer der anderen Variablen) vorliegen. Die Überprüfung geschieht am einfachsten durch Berechnung der Eigenwerte der Kovarianzmatrix der beobachteten Daten. Diese wird über die Inverse der Informationsmatrix der beobachteten Daten berechnet. Louis (1983) beschrieb für einen ML-Schätzer des EM-Algorithmus ein Verfahren zur Herleitung der Kovarianzmatrix über die Informationsmatrix der beobachteten Daten. Die Informationsmatrix ist wiederum laut Definition der negative Erwartungswert der Hessematrix der Loglikelihood-Funktion der beobachteten Daten. Somit gilt, dass die Kovarianzmatrix positiv definit sein muss, damit gefolgert werden kann, dass die Hessematrix der Loglikelihood-Funktion negativ definit ist. Dies ist wiederum die hinreichende Bedingung

für ein Maximum der Loglikelihood-Funktion. Die notwendige Bedingung für ein Maximum ist bereits durch die Anwendung des EM-Algorithmus erfüllt.

Die EM - bzw. EMV-Methoden wurden auf ein einfaches Intervallkartierungsmodell (SIM, auch IM genannt) und für ein zusammengesetztes Intervallkartierungsmodell (CIM) angewendet. Die Datenvorbereitung und die Auswahl der Kofaktoren für das CIM wurde wie in Kapitel 5 beschrieben durchgeführt.

Der EMV-Algorithmus wird, wie in Kapitel 4 hergeleitet, in jedem Intervall zunächst mit festen, "stabilen" Anfangswerten durchgeführt, dann erfolgte eine intensivere Suche nach weiteren Schätzwerten, da ja der EM-Algorithmus i. A. - wie andere iterative Schätzverfahren auch - als Funktion der Anfangsschätzer mehrere (oder sogar zahlreiche) verschiedene Konvergenzpunkte (lokale Maxima) liefert.

Zunächst werden die Ergebnisse von Chromosom 1 ausführlich diskutiert, die folgenden Chromosomen werden dann nach ähnlichem Muster dargestellt.

## **Tabellarische und graphische Auswertung für Chromosom 1**

Die Ergebnisse des EMV-Algorithmus werden in Tabelle 6.2 gegeben, wobei beachtet werden muss, dass hier nur Maximalwerte mit "großen" Loglikelihood-Werten von weiterem Interesse sind, da nur sie am jeweils gefundenen Schätzwert des Lageparameters  $p$  auf die Anwesenheit eines quantitativen Gens hindeuten. Zur Definition häufig verwendeter Startwerte wird auf Tabelle A1 verwiesen.

Tabelle 6.2: "Hauptmaximalwerte" für Chrom. 1 mit Schätzwerten und Varianzen für Startw. 1 bis 4

Schätzer EMV mit Startwert 1,2,3    Schätzwerte EMV mit Startwert 4

Intervalle	2	5 (Kofaktor)	8	10 (Kof.)	2	4	5 (Kof.)	8	10 (Kof.)
$\hat{a}$	-14,42	-222,10	-234,37	-234,37	0,68	0,71	1,30	1,81	0,42
$\hat{a}$	-15,28	-9,24	3,19	3,19	-482,39	-483,14	-484,75	-486,23	-482,57
$\hat{\beta}_1$	31,90	227,99	4,21	4,21	2,25	1,76	1,18	1,95	2,34
$\hat{\beta}_2$	33,63	7,95	238,64	238,64	3,62	3,66	3,40	2,01	3,29
$\hat{\beta}_3$	22,20	1,70	-1,18	-1,18	-1,59	-1,57	-1,68	-1,69	-1,60
$\hat{\beta}_4$	16,19	1,13	0,19	0,19	0,80	0,82	0,82	0,82	0,79
$\hat{\beta}_5$	26,75	2,83	1,17	1,17	1,64	1,70	1,70	1,72	1,70
$\hat{\beta}_6$	32,26	3,14	2,89	2,89	1,17	1,16	1,26	1,13	1,20
$\hat{\beta}_7$	35,36	3,55	2,95	2,95	1,33	1,29	1,26	1,26	1,31
$\hat{\beta}_8$	28,70	-1,26	0,29	0,29	-1,47	-1,50	-1,46	-1,47	-1,46
$\hat{\sigma}$	56,01	17,12	10,15	10,15	4,77	4,78	4,71	4,63	4,81
$\hat{p}$	0,53532	0,05044	0,97328	0,03050	0,518	0,496	0,497	0,530	0,484
Iter.	586,00	11,00	10,00	9,00	33,00	54,00	57,00	51,00	69,00
Loglikelih.	-1878,46	-1524,03	-1332,33	-1329,41	-1278,19	-1282,33	-1382,46	-1299,97	-1282,92
Varianzen									
$K(\hat{p}, \hat{p})$	0,04467	0,00028	0,00012	0,00012	0,000936	0,000933	0,000785	0,000790	0,000855
$K(\hat{a}, \hat{a})$	32,75	13,87	5,11	5,11	0,12	0,28	0,29	0,26	0,34
$K(\hat{a}, \hat{a})$	175,43	3,49	1,23	1,23	4,78	5,72	6,14	6,18	5,92
$K(\hat{\sigma}, \hat{\sigma})$	68313,71	499,70	61,63	61,63	3,11	3,22	3,30	3,21	3,16
$K(\hat{\beta}_1, \hat{\beta}_1)$	20,31	12,65	0,68	0,68	0,16	0,36	0,38	0,16	0,16
$K(\hat{\beta}_2, \hat{\beta}_2)$	17,32	1,72	4,46	4,46	0,14	0,14	0,15	0,35	0,42
$K(\hat{\beta}_3, \hat{\beta}_3)$	17,24	1,68	0,60	0,60	0,13	0,13	0,13	0,13	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	23,63	2,19	0,78	0,78	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_5, \hat{\beta}_5)$	22,51	2,16	0,77	0,77	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	20,27	2,01	0,70	0,70	0,16	0,16	0,16	0,16	0,16
$K(\hat{\beta}_7, \hat{\beta}_7)$	17,26	1,85	0,65	0,65	0,15	0,15	0,15	0,14	0,15
$K(\hat{\beta}_8, \hat{\beta}_8)$	19,61	2,02	0,70	0,70	0,16	0,16	0,16	0,15	0,16

## Graphische Analyse von Chromosom 1

Zunächst werden die Loglikelihood-Profile für Chromosom 1 unter Verwendung des CIM-Modells analysiert. Für das zusammengesetzte Intervallkartierungsmodell wurden im Rahmen der Datenvorbereitung 8 Marker bestimmt, welche als Kofaktoren verwendet werden. Dies geschieht analog zu anderen Intervallkartierungsprogrammen mittels schrittweiser Regression. Dementsprechend wurden folgende Marker zu Kofaktoren des Modells: Marker 5, 10, 28, 39, 41, 47, 61, 86.

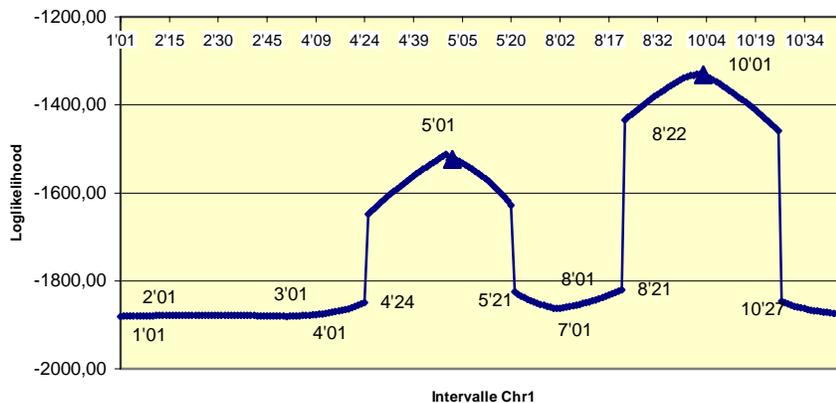


Abbildung 6.3: Loglikelihood-Profil EM-Algorithmus mit Modell CIM und Startwert 1<sup>11</sup>

Das Loglikelihood-Profil für Chromosom 1 mit Startwert 1 in Abbildung 6.3 zeichnet sich zunächst dadurch aus, dass es zweimal in der Mitte von Markerintervallen un stetige Sprünge gibt, in welchen die Loglikelihood-Wert plötzlich sehr groß und später wieder ebenso sprunghaft klein werden.

<sup>11</sup> In Abb. 6.3 sind die Intervallanfänge angegeben, 5'01 steht für Intervall 5, erste Position bei 1 cM und 4'24, 5'21, 8'21 und 10'27 sind Positionen, an welchen die "Sprünge" stattfinden.

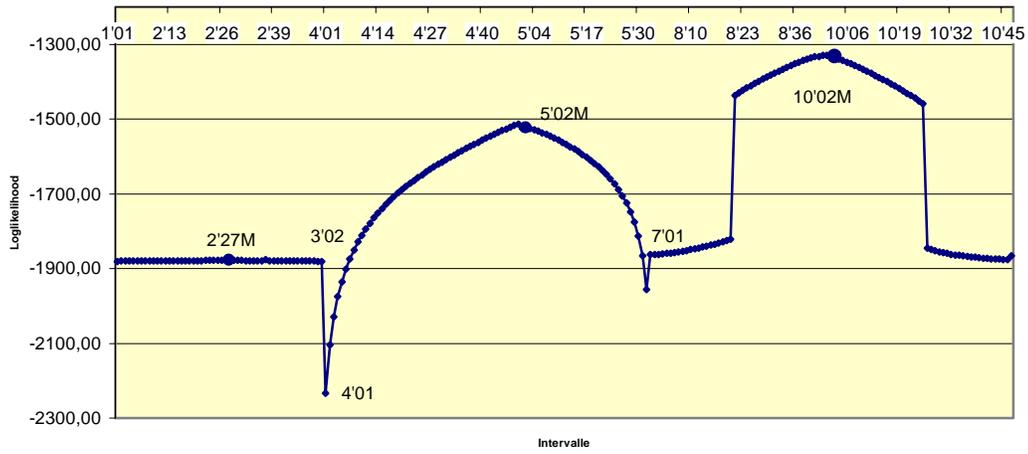


Abbildung 6.4: Loglikelihood-Profil EM-Algorithmus mit Modell CIM und Startwert  $2^{12}$

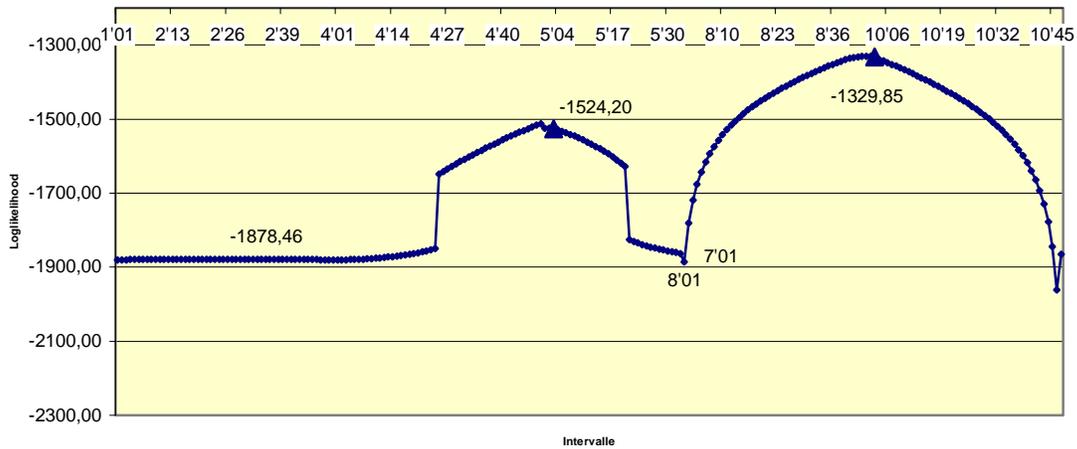


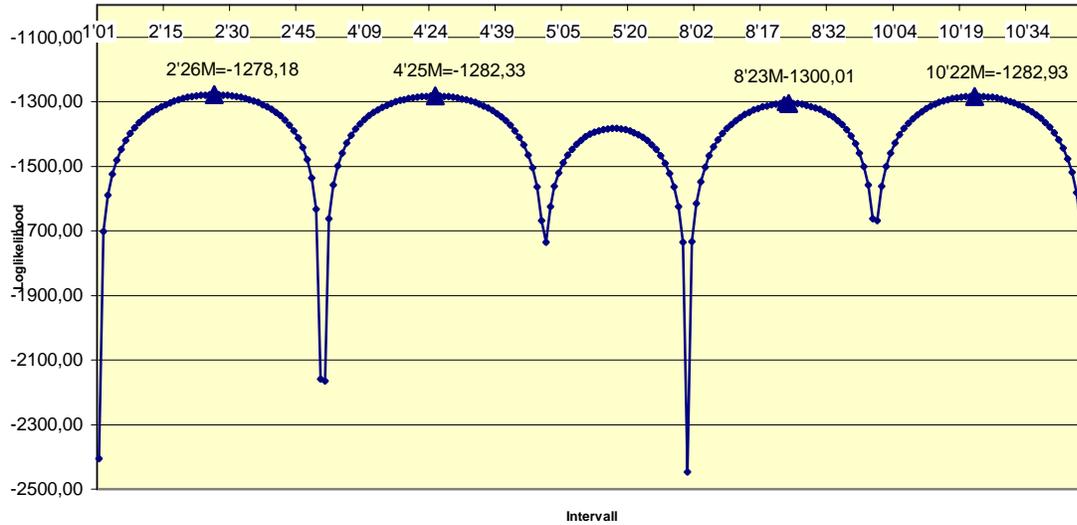
Abbildung 6.5: Loglikelihood-Profil EM-Algorithmus mit Modell CIM und Startwert  $3^{13}$

Auch eine Verkleinerung der Schrittgröße von 1 cm auf 1/10 cm (oder noch kleiner) ändert an dem unstetigen Verhalten der Funktion an diesen Stellen nichts.

<sup>12</sup> In Abb. 6.5 werden erwähnte Randpunkte und die Funktionswerte der Maxima angegeben, maximaler Profilwert bei Intervall 10, Position 2 entspricht 10'02M in Abb. 6.4.

<sup>13</sup> In Abb. 6.6 werden die Funktions- (Loglikelihood-) Werte der Maxima angegeben.

Wenn man dagegen die Ergebnisse von Startwert 2 betrachtet (Abbildung 6.4), dann verschwinden die beiden linken Funktionssprünge, während bei Startwert 3 (Abbildung 6.5) gerade die beiden rechten verschwinden.



**Abbildung 6.6: Loglikelihood-Profil EM-Algorithmus mit Modell CIM und Startwert 4**

Ein Zusammenfügen der Ergebnisse von Startwert 2 und 3 ergibt offenbar ein stetiges Verhalten eines Funktionsastes (eventuell mit Ausnahme der Ränder bei Intervall 3/ Intervall 4 (3'02 und 4'01) bzw. Intervall 7/ Intervall 8 (7'01 und 8'01) bei denen sehr kleine Funktionswerte (eventuell  $-\infty$ ) auftreten).

Einen vollständig anderen Funktionszweig mit völlig neuen Maxima liefert dagegen der Startwert 4 (Abb. 6.6). Hier findet der EM-Algorithmus in jedem größeren Intervall ein Maximum in etwa der Mitte des Intervalls.

In Abbildung 6.7 sind die Lösungen aller 4 Startwerte zusammen aufgetragen. Offenbar gruppieren sich die Konvergenzpunkte in 3 verschiedenen Funktionszweigen:

i) Die "oszillatorische Lösung", mit Maxima in etwa der Mitte (Schätzwerte für Parameter  $p$  zwischen 0,53 und 0,48, vgl. Tab. 6.2) der Intervalle I2, I4, I5, I8 und I10.

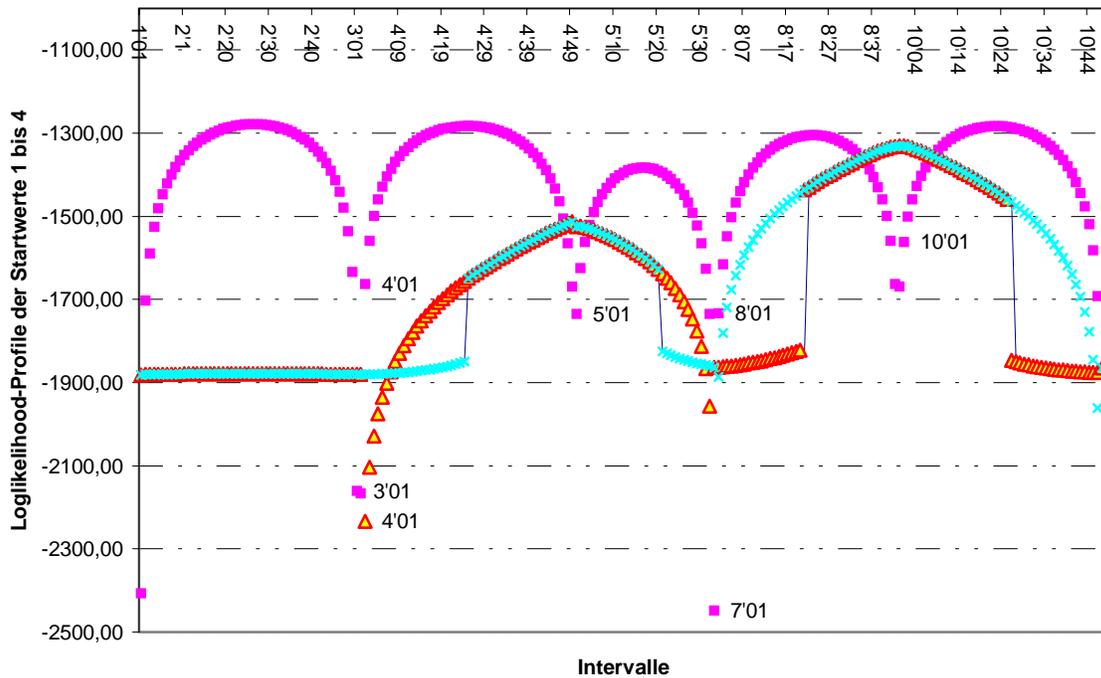


Abbildung 6.7: Überblick: Loglikelihood-Profile von Chromosom 1 mit verschiedenen Startwerten

ii) Die "Kofaktoren-Lösung" mit Suprema an den Intervallgrenzen I4/ I5 und I8/ I10, bei welchen genau die Kofaktoren des CIM-Modells auftreten. Sie enthält Maxima in I5 und I10, in unmittelbarer Nähe des jeweiligen Supremums (Abbildung 6.7a).

iii) Eine in relativ engen Grenzen variierende, bei kleinen Werten auftauchende Lösung (die "tiefliedende Lösung") mit einem (nur in entsprechender Vergrößerung sichtbaren) Maximum in Intervall 2 bei Werten von  $\hat{p}$  zwischen 0,53 und 0,54.

Man vergleiche dieses mit den EMV- Ergebnissen der Tabelle 6.2. Man erhält hier zum Startwert 4 die Maxima der Lösung i). Bezüglich der Varianzen, und der Schätzer der Parameter  $a$ ,  $d$ ,  $\beta_i$  ( $i=1,2,\dots,8$ ) und  $\sigma$  bilden diese eine gegen alle anderen Lösungen klar

abgegrenzte Gruppe. Das gilt auch für die "Koordinatenwerte"  $\ln(L)=-1289\pm 11$  und  $p=0,50\pm 0,03$  mit Ausnahme des Loglikelihood-Wertes in Intervall 5 ( $\ln(L)=-1382$ ).

In der Gruppe ii) erhält man in Tabelle 6.2 ein Maximum mehr als bei dem Profilverfahren, nämlich das in I8 (bei  $\hat{p}=0,97328$ , vgl. Abbildung 6.7a).

Die Gruppe ii) zerfällt jedoch in 2 Untergruppen:

a) Die beiden Maxima in I8 ( $\hat{p}=0,97328$ ) und I10 ( $\hat{p}=0,03050$ ) mit innerhalb der numerischen Genauigkeit identischen Schätzern der Parameter  $a$ ,  $d$ ,  $\beta_i$  ( $i=1,2,\dots,8$ ) und  $\sigma$ , und fast identischen asymptotischen Varianzen  $K(\hat{\beta}_1, \hat{\beta}_1)$ . Sie besitzen große Werte für  $\hat{\beta}_2$  und relativ große für  $K(\hat{\beta}_2, \hat{\beta}_2)$ .

b) Das Maximum in I5 (mit  $\hat{p}=0,05044$ ), dessen Parameterschätzer und Varianzen erheblich von denen der Untergruppe a) abweichen. Insbesondere sind  $\hat{\beta}_1$  und  $K(\hat{\beta}_1, \hat{\beta}_1)$  groß anstatt  $\hat{\beta}_2$  und  $K(\hat{\beta}_2, \hat{\beta}_2)$ .

Das Maximum bei  $\hat{p}=0,53532$  in Gruppe iii) (aus I2) ist offensichtlich bezüglich der Parameterschätzer und Varianzen am meisten von allen anderen Maxima verschieden.

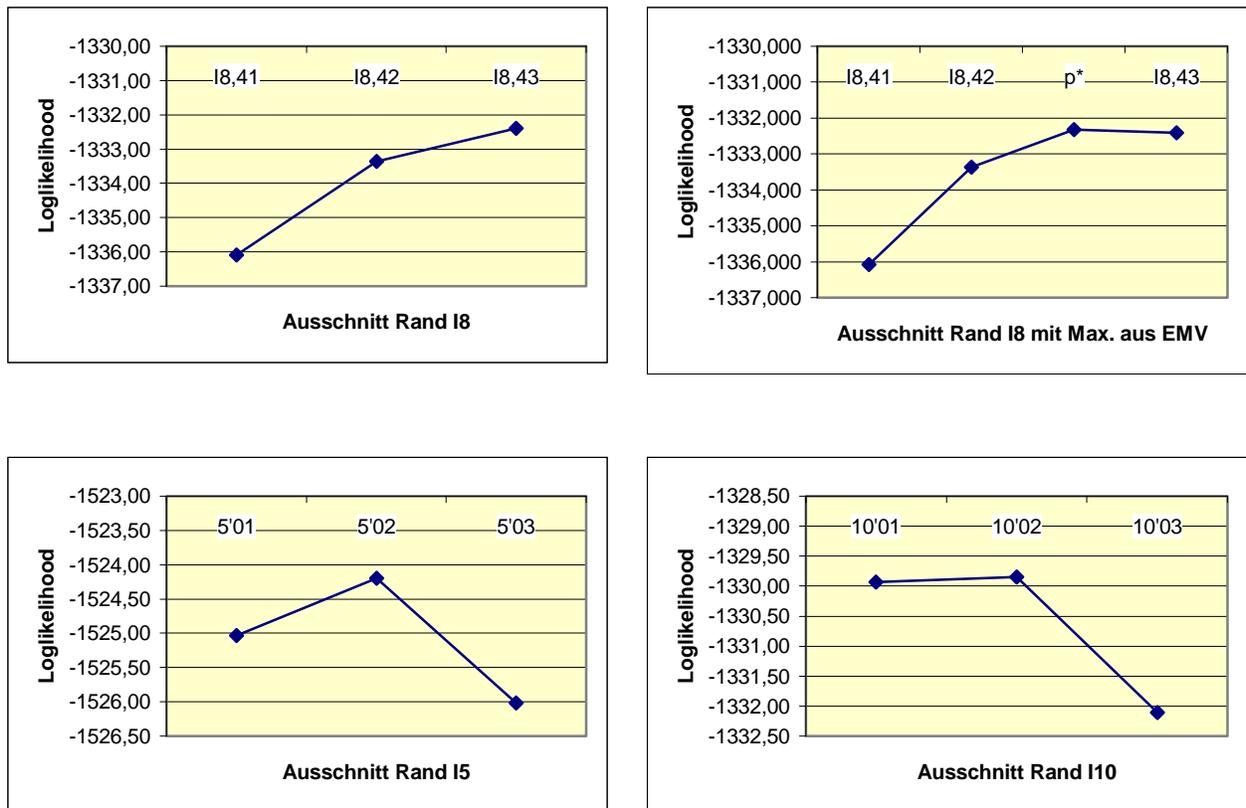


Abbildung 6.7a: Rechter Rand I8 mit und ohne EMV-Max., sowie Ränder I5 und I10.<sup>14</sup>

Der Vergleich der Loglikelihood-Profil (oder Lod-Score-) mit der EMV-Methode zeigt insgesamt, dass diese beiden Methoden weitgehend äquivalent sind.

Die Tatsache, dass die Profilmethode ein Maximum der Kofaktorengruppe (I8,  $\hat{p}=0,973$ ) nicht liefert, sollte nicht überschätzt werden. Mehr Rechenpunkte (ein feinerer Diskretisierungsgrad) der Variablen  $p$  nahe der Intervallgrenze I8/ I10 würde dieses Maximum ohne Zweifel liefern. In der Praxis bedeutet dies allerdings, dass die Diskretisierung deswegen oft sehr viel feiner sein muss, als  $1 \text{ cM}^{15}$ , man vergleiche Abbildung 6.7a, linkes und rechtes oberes Diagramm.

<sup>14</sup> lies:  $p^*=\hat{p}$  des EMV!

<sup>15</sup> In der Praxis werden i. A. 1 bis 5 cM verwendet

Für das EMV-Verfahren spricht auch der geringere Rechenaufwand und die höhere Rechengenauigkeit.

Für das Profil-Verfahren spricht dessen höhere Anschaulichkeit, z. B. taucht die "oszillatorische Lösung" in allen 5 größeren Intervallen auf. Ein solches Verhalten kann in praxi durchaus zum Auffinden weiterer Maxima führen. Jedoch muss hier vor Übereinfachung gewarnt werden: Es liegt z. B. nahe, die Lösung ii) in das Gebiet der Intervalle I1 bis I3 fortzusetzen. Trotz einer Wahl von 30 000 Anfangsschätzern mit einer Monte-Carlo-Methode wurde kein weiterer Konvergenzpunkt in diesem Intervall gefunden! Das liegt höchstwahrscheinlich, wenn auch nicht mit letzter Sicherheit, daran, dass dort kein weiterer Kofaktor auftaucht, daher auch die Namensgebung für die Gruppe ii).

Diese Diskussion zeigt, dass die beiden Verfahren sich - trotz ihrer prinzipiellen Äquivalenz - in der Praxis gut ergänzen.

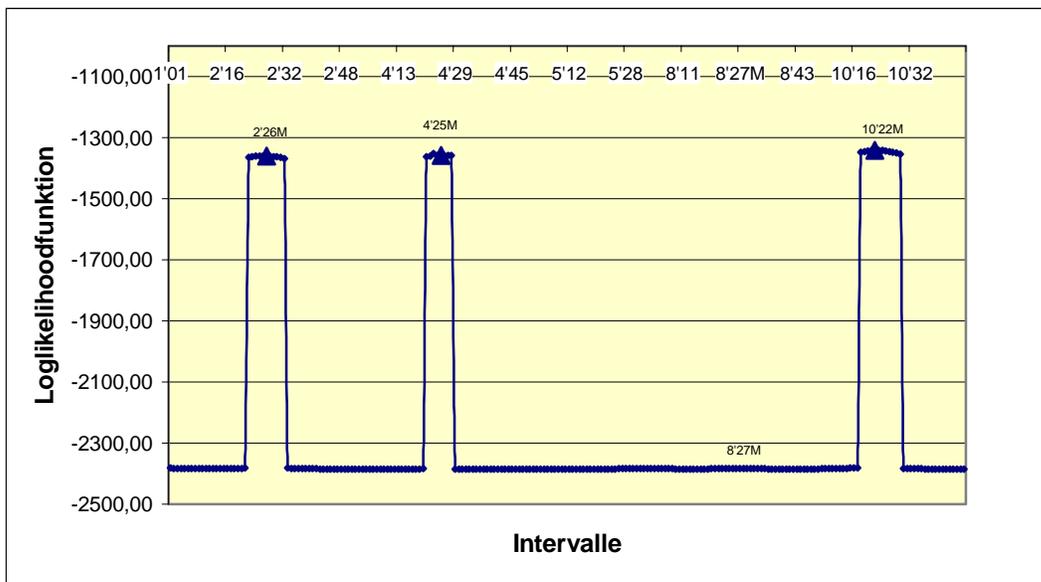
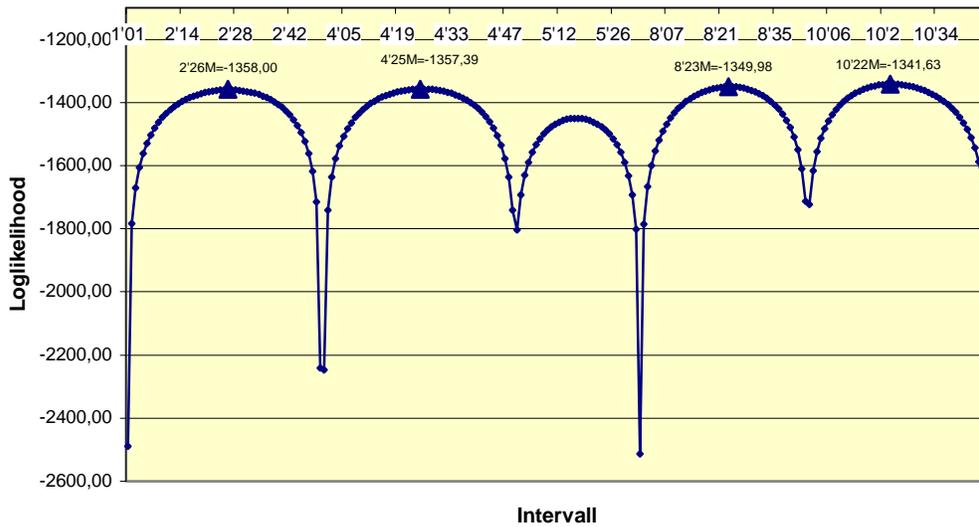


Abbildung 6.8: Loglikelihood-Profil für Chromosom 1, SIM, stabiler Startwert ( $a=1$ ,  $d=2$ ,  $\sigma=10$ )



**Abbildung 6.9: Loglikelihood-Profil für Chromosom 1, SIM, oszillierende Lösung, Startwert OL**

Das Loglikelihood-Profil des EM-Algorithmus unter Annahme des einfachen Intervallkartierungsmodells (IM) ohne Kofaktoren zeigt bei dem "stabilen Startwert" (Tab. A1) ebenfalls Unstetigkeitsstellen und findet in der Mitte von Intervall 2, Intervall 4 und Intervall 10 drei Hauptmaximalwerte (siehe Abbildung 6.8). Zudem ist auch ein Maximum in Intervall 8 zu finden, welches aber offensichtlich sehr kleine Funktionswerte besitzt.

Die 3 Hauptmaxima (Abbildung 6.8) entsprechen den 3 Hauptmaxima der oszillatorischen Lösung des CIM im EM-Algorithmus für das Chromosom 1 (vgl. Tab.6.2). An dieser Stelle ist es naheliegend, in Analogie zu CIM einen anderen Startwert zu versuchen, ein günstiger Wert ist in Tab. A1 gegeben und mit OL bezeichnet. Es ergibt sich eine oszillierende Lösung in Analogie zu der oszillierenden Lösung des CIM. Der Vergleich von Abbildung 6.9 mit Abbildung 6.6 zeigt, dass in CIM das Maximum in I2 das höchste ist, gefolgt von denen in I4, I10 und I8, während in SIM das Maximum in I10 das höchste ist, gefolgt von denen in I8, I4 und I2. Die Positionen der Maxima stimmen innerhalb der Genauigkeit von 1 cM jeweils miteinander überein. Neben dieser Lösung taucht auch in SIM eine "tiefliegende Lösung" auf

(s. Abbildung 6.8), die nur in relativ engen Grenzen variiert. Aus trivialen Gründen (keine Kofaktoren) kann die Kofaktorenlösung in SIM nicht erscheinen.

Die Auswertung der weiteren Chromosomen erfolgt mit den Startwerten 1 und 4 für das Loglikelihood-Profil des CIM Modells. Beide Startwerte liefern wie bei Chromosom 1 sehr unterschiedliche Maxima.

### Tabellarische und graphische Auswertung von Chromosom 2

In Tabelle 6.3 werden die höheren Maxima des EMV-Algorithmus mit Varianzen (Hauptdiagonalelementen der geschätzten Kovarianzmatrix) und Schätzwerten gegeben. Eine vollständige Darstellung aller Maxima mit Schätzwerten der Parameter  $a$ ,  $d$ ,  $p$ ,  $\sigma$ ,  $\beta_i$ , den Kovarianzen und den Eigenwerten der Kovarianzmatrix ist in den Tabellen A2 und A3 zu finden.

**Tabelle 6.3: Maxima des EMV für Chr 2 mit Varianzen und Schätzwerten für Startwerte 1 und 4. Für Eigenwerte der Kovarianzmatrix und genauere Werte der angegebenen Schätzer und Varianzen, vgl.**

**Tabellen A2 und A3**

**Startwert 1      Startwert 4**

Intervalle	16	12	13	14	15	16	18	19	20	21
$\hat{a}$	-18,68	-1,09	-0,93	-0,79	-0,43	-0,74	-1,24	-1,41	-1,07	-1,07
$\hat{d}$	5,46	-482,18	-481,93	-481,82	-481,54	-480,94	-480,88	-480,98	-481,47	-481,24
$\hat{\beta}_1$	30,23	2,33	2,32	2,37	2,36	2,32	2,28	2,29	2,35	2,37
$\hat{\beta}_2$	35,07	3,59	3,64	3,70	3,72	3,83	3,84	3,81	3,71	3,68
$\hat{\beta}_3$	23,69	-1,61	-1,59	-1,63	-1,58	-1,51	-1,49	-1,49	-1,56	-1,53
$\hat{\beta}_4$	16,19	0,84	0,76	0,81	0,78	0,78	0,82	0,87	0,89	0,89
$\hat{\beta}_5$	25,10	1,72	1,73	1,71	1,73	1,73	1,65	1,70	1,67	1,69
$\hat{\beta}_6$	31,22	1,21	1,28	1,22	1,21	1,28	1,36	1,32	1,28	1,30
$\hat{\beta}_7$	37,62	1,31	1,35	1,35	1,37	1,41	1,38	1,34	1,36	1,34
$\hat{\beta}_8$	29,62	-1,57	-1,55	-1,51	-1,47	-1,41	-1,38	-1,46	-1,53	-1,46
$\hat{\sigma}$	56,18	4,70	4,73	4,76	4,80	4,76	4,66	4,61	4,70	4,70
$\hat{p}$	0,54956	0,4886	0,4908	0,5364	0,5009	0,4895	0,5050	0,5044	0,4981	0,5359
Iter.	534,00	32,00	31,00	33,00	29,00	33,00	28,00	34,00	34,00	29,00
Loglikelih.	-1876,25	-1696,14	-1357,30	-1445,4	-1444,4	-1472,1	-1574,55	-1825,95	-1741,98	-1472,30
Varianzen										
$K(\hat{p}, \hat{p})$	0,0635	0,0008	0,0009	0,0009	0,0008	0,0007	0,0007	0,0008	0,0008	0,0008
$K(\hat{a}, \hat{a})$	27,10	0,11	0,11	0,12	0,11	0,12	0,10	0,11	0,11	0,10

$K(\hat{\alpha}, \hat{\alpha})$	63,47	4,62	4,65	4,68	4,77	4,85	4,63	4,59	4,64	4,64
$K(\hat{\sigma}, \hat{\sigma})$	59979,74	3,05	3,06	3,11	3,12	3,10	2,95	2,96	3,05	3,01
$K(\hat{\beta}_1, \hat{\beta}_1)$	19,48	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15
$K(\hat{\beta}_2, \hat{\beta}_2)$	16,99	0,14	0,14	0,14	0,14	0,14	0,14	0,13	0,14	0,13
$K(\hat{\beta}_3, \hat{\beta}_3)$	17,02	0,13	0,13	0,13	0,13	0,13	0,13	0,13	0,13	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	23,43	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_5, \hat{\beta}_5)$	22,10	0,17	0,17	0,17	0,17	0,17	0,17	0,16	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	19,61	0,15	0,16	0,16	0,16	0,16	0,16	0,15	0,16	0,16
$K(\hat{\beta}_7, \hat{\beta}_7)$	17,01	0,14	0,15	0,15	0,15	0,15	0,14	0,14	0,14	0,14
$K(\hat{\beta}_8, \hat{\beta}_8)$	19,24	0,15	0,16	0,16	0,16	0,16	0,15	0,15	0,15	0,15

Diese Maximalwerte ergänzen die Loglikelihood-Profile in Abbildung 6.10 und Abb. 6.11. In Abbildung 6.10 fällt auf, dass der Loglikelihood-Funktionswert des einzigen Maximums sehr klein ist im Vergleich zu denen der "oszillatorischen Lösung" in Abbildung 6.11, so dass man nicht vom Vorhandensein eines Gens an dieser Position ausgehen kann.

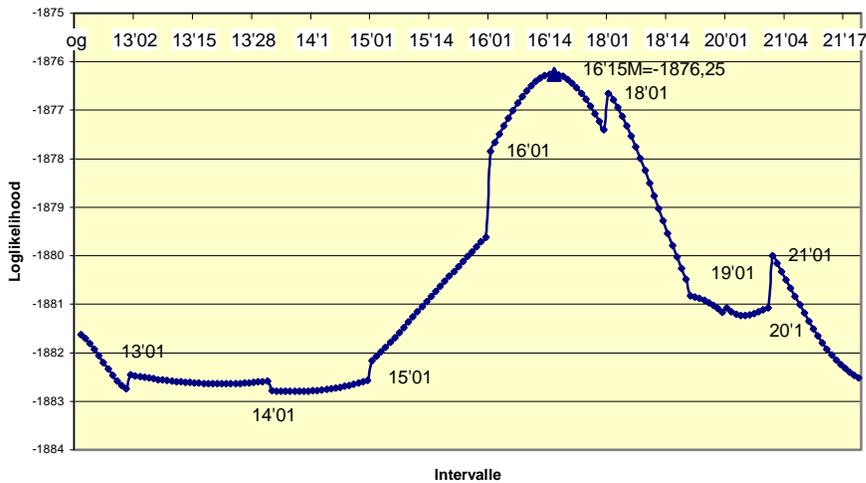
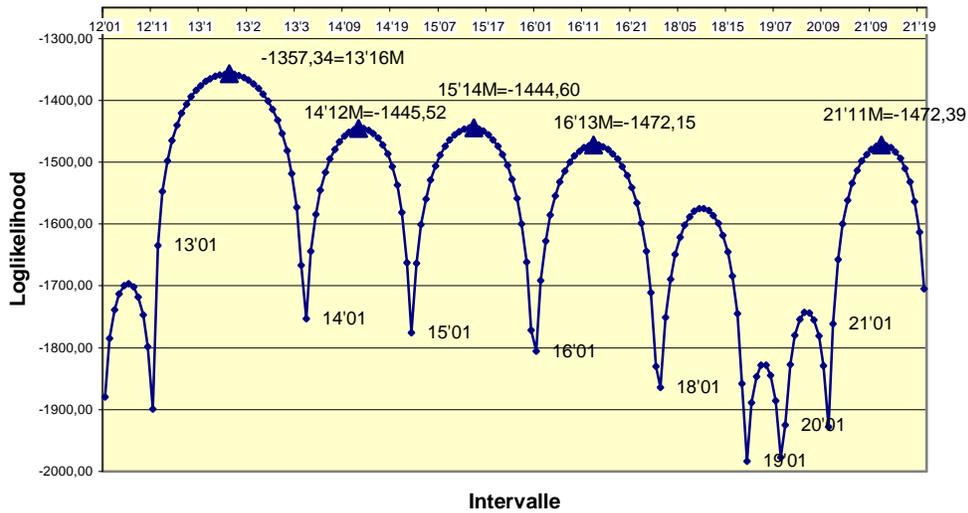


Abbildung 6.10: Loglikelihood-Profil Chromosom 2, CIM mit Startwert 1



**Abbildung 6.11: Loglikelihood-Profil Chromosom 2, CIM mit Startwert 4**

In Abbildung 6.11 liegt ein gut ausgeprägtes Hauptmaximum in der Mitte des Intervalls 13. Auch hier findet die "oszillatorische Lösung" in jedem größeren Intervall ein Maximum, allerdings sind die Maximalwerte der Likelihood in I14 bis I21 um viele Größenordnungen kleiner als diejenigen der oszillatorischen Lösung auf Chromosom 1.

Betrachtet man das Loglikelihood-Profil des EM-Algorithmus unter Annahme des SIM-Modells, so findet sich kein Maximum bei dem stabilen Startwert (Abb. 6.12). Entsprechende Spitzen des Profils sind Randwerte, was durch die Angabe der jeweils ersten Positionen der einzelnen Markerintervalle verdeutlicht werden soll.

Verwendet man jedoch den in Chromosom 1 bereits verwendeten Startwert für die Oszillatorische Lösung (OL), dann erhält man diese wiederum in völliger Analogie zu dem CIM-Modell, man vergleiche Abbildung 6.13 mit Abbildung 6.11. Die starke Dominanz des Maximum in I13 ist evident. Die Position der Maxima von Abbildung 6.11 und Abbildung 6.13 stimmen innerhalb der Genauigkeit von 1 cM jeweils überein.

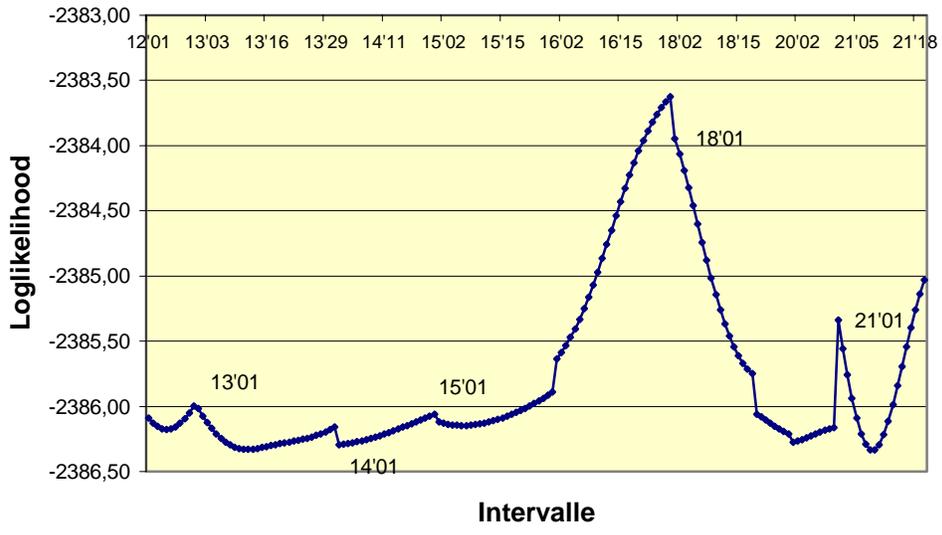


Abbildung 6.12: Loglikelihood-Profil Chromosom 2, SIM, stabiler Startwert

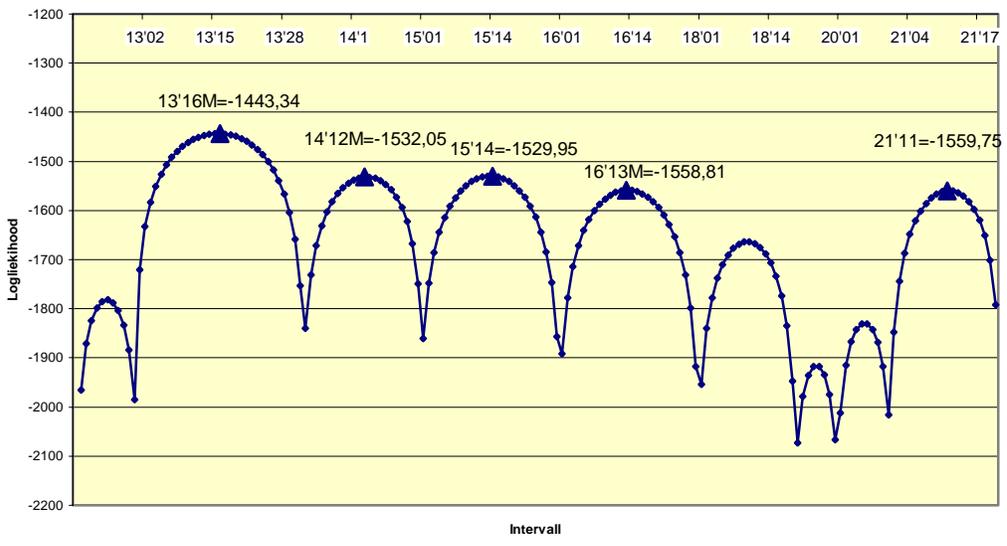


Abbildung 6.13: Loglikelihood-Profil Chromosom 2, SIM, oszillatorische Lösung, Startwert OL

### Tabellarische und graphische Auswertung Chromosom 3

Tabelle 6.4: Wie Tabelle 6.3 für Chromosom 3

Startwert 1

Startwert 4

Intervalle	24	25	26	28 KOF.!	23	24	26	28 KOF. !	33
$\hat{a}$	-39,02	-43,83	-171,11	-236,43	-1,04	-0,93	0,34	-1,45	-1,14
$\hat{d}$	-14,62	-12,84	-9,00	0,94	-481,81	-481,07	-482,36	-480,00	-482,22
$\hat{\beta}_1$	29,59	28,55	12,89	3,40	2,32	2,35	2,33	2,34	2,42
$\hat{\beta}_2$	28,52	25,65	11,05	4,40	3,69	3,64	3,69	3,68	3,69
$\hat{\beta}_3$	42,50	51,83	174,11	235,61	-1,46	-1,04	-1,90	-0,43	-1,62
$\hat{\beta}_4$	15,12	14,86	5,21	1,41	0,80	0,80	0,78	0,75	0,75
$\hat{\beta}_5$	24,39	23,95	9,16	2,30	1,78	1,74	1,74	1,72	1,77
$\hat{\beta}_6$	31,13	28,85	11,65	2,06	1,19	1,22	1,21	1,06	1,04
$\hat{\beta}_7$	32,10	30,13	11,00	0,54	1,29	1,25	1,32	1,40	1,45
$\hat{\beta}_8$	27,52	27,65	9,31	-0,61	-1,59	-1,56	-1,48	-1,60	-1,59
$\hat{\sigma}$	52,02	53,42	32,88	8,35	4,71	4,75	4,81	4,68	4,69
$\hat{p}$	0,9014	0,2811	0,7953	0,0131	0,4514	0,5222	0,5278	0,4673	0,5096
Iter.	203,00	102,00	57,00	10,00	39,00	63,00	93,00	55,00	26,00
Loglikelihood.	-1866,15	-1858,39	-1791,28	-1255,28	-1666,78	-1275,41	-1819,77	-1288,26	-1453,96
Varianzen									
$K(\hat{p}, \hat{p})$	0,0064	0,1088	0,0025	0,00004	0,0008	0,0013	0,0008	0,0009	0,0008
$K(\hat{a}, \hat{a})$	46,06	41,00	63,73	3,51	0,13	0,23	0,48	0,24	0,10
$K(\hat{d}, \hat{d})$	60,70	35,45	13,06	0,83	4,64	4,85	6,01	4,92	4,61
$K(\hat{\sigma}, \hat{\sigma})$	68552,64	48310,08	7973,59	28,24	3,12	3,30	3,17	3,31	2,96
$K(\hat{\beta}_1, \hat{\beta}_1)$	17,91	17,53	7,08	0,46	0,15	0,15	0,16	0,15	0,15
$K(\hat{\beta}_2, \hat{\beta}_2)$	16,05	16,10	7,21	0,42	0,14	0,14	0,14	0,14	0,13
$K(\hat{\beta}_3, \hat{\beta}_3)$	29,43	34,41	63,05	3,23	0,13	0,21	0,52	0,28	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	22,13	21,28	8,13	0,52	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_5, \hat{\beta}_5)$	21,28	20,13	7,94	0,52	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	18,60	17,67	8,02	0,48	0,16	0,16	0,16	0,16	0,16
$K(\hat{\beta}_7, \hat{\beta}_7)$	16,44	16,14	7,19	0,45	0,14	0,15	0,15	0,14	0,14
$K(\hat{\beta}_8, \hat{\beta}_8)$	18,30	17,26	8,09	0,47	0,16	0,16	0,16	0,16	0,15

Tabelle 6.4 ist zu entnehmen, dass die dem Loglikelihood-Profil zugrunde liegende Funktion unter Startwert 1 ein sehr hohes Maximum (Loglikelihood=-1255,2) zu Beginn des Intervalls 28, und zwar bei  $\hat{p}=0,013$ , aufweist. Dieses Maximum, das höchste von allen auf Chromosom 3, wird vom Loglikelihood-Profil in Abbildung 6.14 unter der vorgegebenen

Auflösung nicht angezeigt! Auch hier müsste an dieser Stelle (wie in Intervall 8 bei Chromosom 1) ein feinerer Diskretisierungsgrad als 1 cM gewählt werden.

Offensichtlich liefern die Loglikelihood-Profilverfahren (Abbildung 6.14, Abbildung 6.15) und das EMV-Verfahren (Tab. 6.4) wieder die 3 klar voneinander abgegrenzten, schon bei Chromosom 1 gefundenen Gruppen, nämlich die oszillatorische Lösung mit zwei Hauptmaxima in etwa der Mitte der Intervalle 24 und 28 (Abbildung 6.15), die Kofaktorlösung mit dem höchsten Maximum überhaupt am linken Rand von Intervall 28 (Tab. 6.4) und die tiefliegende Lösung (Abbildung 6.14) bei relativ niedrigen bzw. sehr niedrigen Werten.

Die Übereinstimmung der oszillierenden Lösungen von CIM und SIM (Abb. 6.16) ist, abgesehen von einer Verschiebung der Ordinatenwerte, frappierend; die Verschiebung der Abszissenwerte beträgt an den Positionen der Maxima höchstens 2 cM.

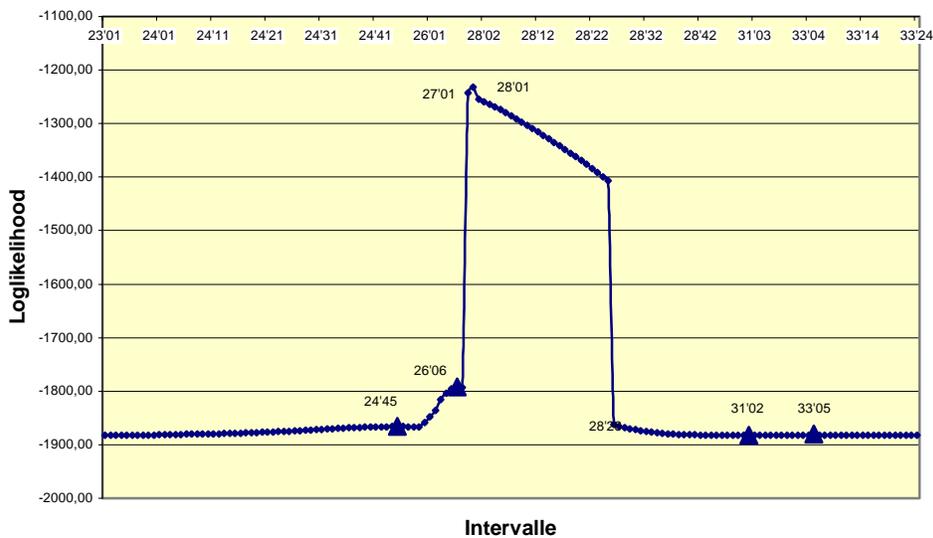


Abbildung 6.14 : Loglikelihood-Profil Chromosom 3, CIM mit Startwert 1

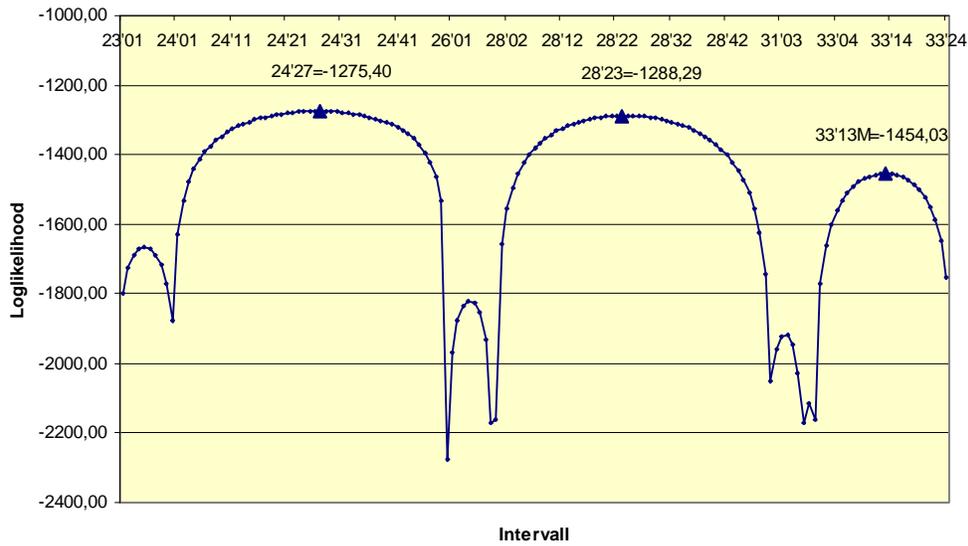


Abbildung 6.15: Loglikelihood-Profil Chromosom 3, CIM mit Startwert 4. Die 90 cM im Intervall 24 (s. Abb. 6.1) werden durch die Kartierungsvorschrift reduziert auf 50 Einheiten in der graphischen Darstellung. Die Intervalle 28, 29 und 30 werden zusammen gelegt (entsprechend 58 cM und 50 "Positionen" in der Graphik).

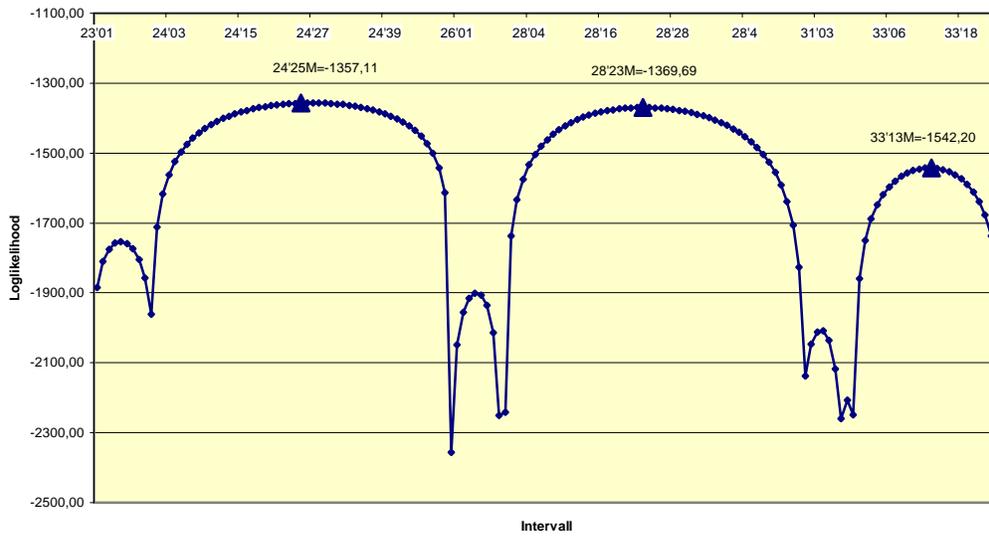


Abbildung 6.16: Loglikelihood-Profil Chromosom 3, SIM, oszillatorische Lösung, Startwert OL

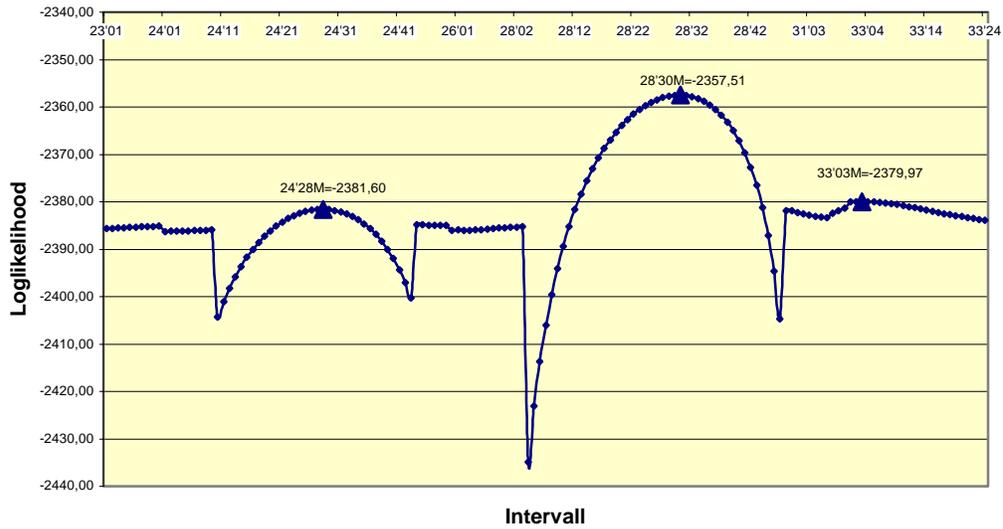


Abbildung 6. 17: Loglikelihood-Profil Chromosom 3, SIM, stabiler Startwert

## Tabellarische und graphische Auswertung Chromosom 4

Tabelle 6.5: Wie Tabelle 6.3 für Chromosom 4

Startwert 1

Startwert 4

Intervalle	39 (KOF)	40*	35	36	37	38	39 (KOF)	40*
$\hat{a}$	-236,53	-240,97	0,49	0,48	0,56	-0,82	0,88	1,09
$\hat{d}$	-0,67	0,54	-481,87	-482,02	-482,34	-480,30	-483,62	-484,82
$\hat{\beta}_1$	3,05	2,39	2,34	2,37	2,39	2,32	2,35	2,31
$\hat{\beta}_2$	4,48	4,13	3,65	3,66	3,65	3,70	3,73	3,72
$\hat{\beta}_3$	-1,18	-1,33	-1,62	-1,61	-1,60	-1,60	-1,61	-1,64
$\hat{\beta}_4$	238,07	-1,03	0,74	0,61	0,33	1,59	-0,07	0,23
$\hat{\beta}_5$	1,37	243,29	1,78	1,78	1,79	1,74	1,65	0,87
$\hat{\beta}_6$	1,08	1,28	1,25	1,25	1,25	1,20	1,17	1,16
$\hat{\beta}_7$	2,12	1,51	1,33	1,34	1,38	1,30	1,33	1,31
$\hat{\beta}_8$	-0,81	-1,42	-1,51	-1,49	-1,47	-1,49	-1,47	-1,51
$\hat{\sigma}$	8,06	7,04	4,80	4,80	4,80	4,78	4,77	4,76
$\hat{p}$	0,0207	0,9552	0,5055	0,4913	0,4917	0,5104	0,5079	0,5039
Iter.	3,00	31,00	31,00	30,00	42,00	106,00	101,00	61,00
Loglikelih.	-1214,49	-1210,61	-1666,54	-1294,60	-1397,24	-1994,30	-1949,00	-1312,27

\* Der Marker 41 am rechten Rand von Intervall 40 ist Kofaktor

Varianzen								
$K(\hat{p}, \hat{p})$	0,00042	0,00018	0,00082	0,00098	0,00090	0,00074	0,00069	0,00075
$K(\hat{a}, \hat{a})$	3,27	2,55	0,11	0,12	0,19	0,53	0,48	0,36
$K(\hat{d}, \hat{d})$	0,80	0,60	4,72	4,75	4,91	6,54	6,74	7,38
$K(\hat{\sigma}, \hat{\sigma})$	24,56	14,25	3,11	3,12	3,14	3,39	3,36	3,24
$K(\hat{\beta}_1, \hat{\beta}_1)$	0,43	0,33	0,15	0,15	0,16	0,16	0,16	0,15
$K(\hat{\beta}_2, \hat{\beta}_2)$	0,39	0,30	0,14	0,14	0,14	0,14	0,14	0,14
$K(\hat{\beta}_3, \hat{\beta}_3)$	0,38	0,29	0,13	0,13	0,13	0,13	0,13	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	3,31	0,38	0,17	0,19	0,30	0,68	0,62	0,27
$K(\hat{\beta}_5, \hat{\beta}_5)$	0,48	2,41	0,17	0,17	0,17	0,17	0,18	0,39
$K(\hat{\beta}_6, \hat{\beta}_6)$	0,46	0,34	0,16	0,16	0,16	0,16	0,16	0,16
$K(\hat{\beta}_7, \hat{\beta}_7)$	0,41	0,32	0,15	0,15	0,15	0,15	0,15	0,15
$K(\hat{\beta}_8, \hat{\beta}_8)$	0,44	0,34	0,16	0,16	0,16	0,16	0,16	0,16

Das Loglikelihood-Profil des CIM-Verfahrens für Startwert 1 (Abb. 6.18) weist in der Nähe der Kofaktormarker 39 und 41, nämlich am linken Rand des Intervalles 39 und am rechten Rand des Intervalles 40, Maxima der Kofaktorenlösungen auf. Auf Grund des unzureichenden Diskretisierungsgrades wird das Maximum in Intervall 39 bei  $\hat{p}=0,0207$  im Profil wiederum nicht erkannt und muss also über den EMV (oder eine feinere Diskretisierung) bestimmt werden. Das Maximum (der Kofaktorenlösung) in Intervall 40 für  $\hat{p}=0,9552$  kann auch näherungsweise aus dem Profil abgelesen werden. Diese beiden Maxima sind die weitaus höchsten. Die oszillatorische Lösung (Abb. 6.19) findet etwa in der Mitte der Intervalle 36, 37 und 40 Maximalwerte. Diese Maxima treten an gleicher Stelle im SIM-Profil der Loglikelihood-Funktion mit Startwert OL auf (Abb. 6.21). Die Unterschiede zwischen den Funktionswerten der Maxima sind in CIM etwas größer als in SIM.

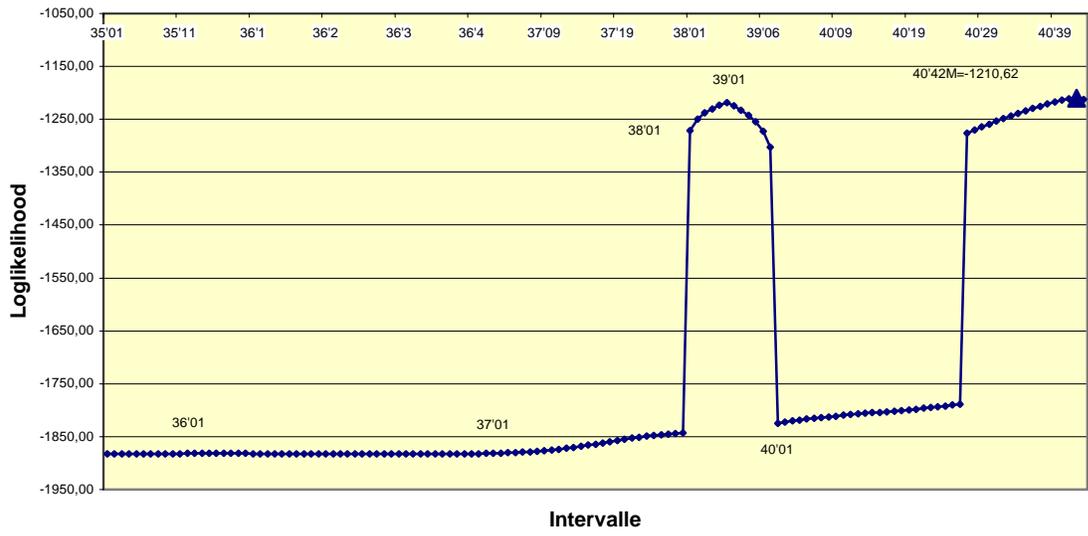


Abbildung 6. 18: Loglikelihood-Profil Chromosom 4, CIM mit Startwert 1

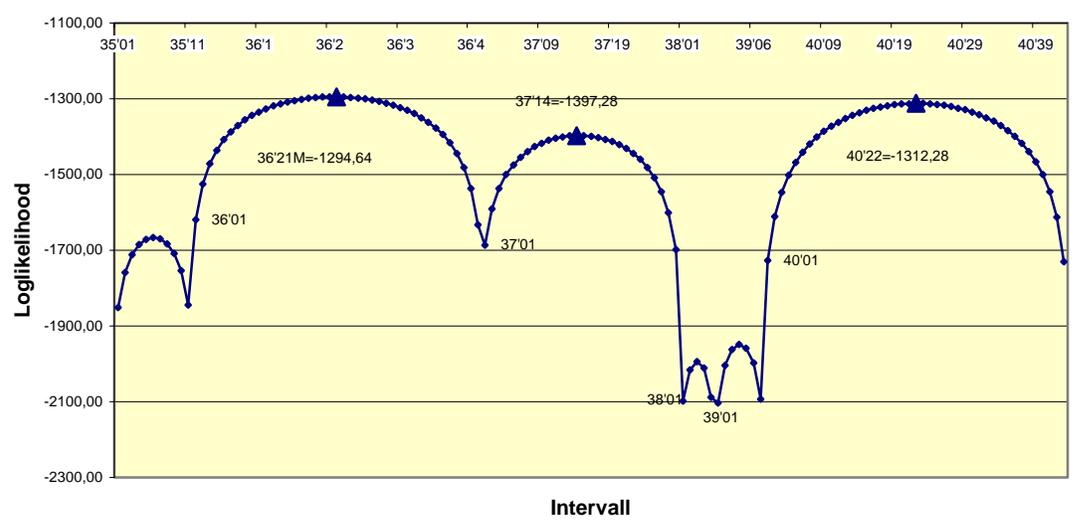


Abbildung 6. 19: Loglikelihood-Profil Chromosom 4, CIM mit Startwert 4

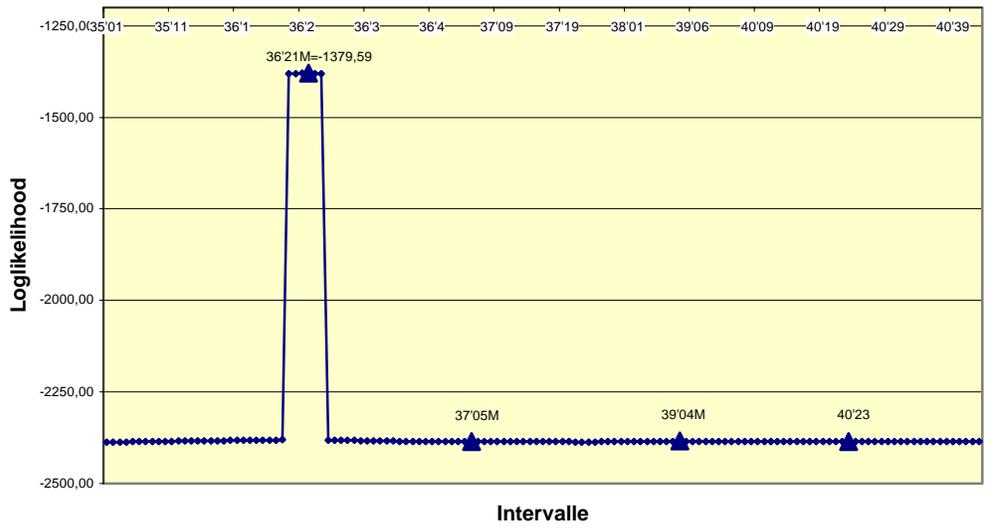


Abbildung 6.20: Loglikelihood-Profil Chromosom 4, SIM, stabiler Startwert

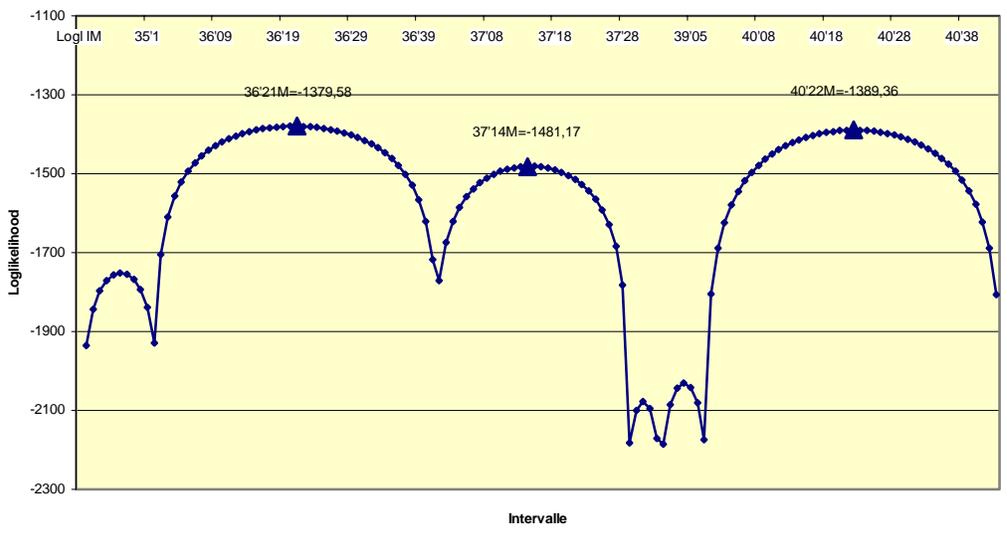


Abbildung 6.21: Loglikelihood-Profil Chromosom 4, SIM, oszillatorische Lösung, Startwert OL

## Tabellarische und graphische Auswertung Chromosom 5

Tabelle 6.6: Wie Tabelle 6.3 für Chromosom 5

Startwert 1

Startwert 4

Intervalle	42*	45	47 (Kof)	50	42*	43	45	46	47 (KoF)	51
$\hat{a}$	-7,40	-23,14	-232,83	-28,44	0,45	0,17	0,34	0,35	0,74	0,38
$\hat{d}$	1,53	-0,69	1,98	6,10	-482,09	-481,94	-482,28	-482,47	-483,47	-482,20
$\hat{\beta}_1$	30,92	29,73	4,17	29,67	2,34	2,35	2,33	2,34	2,33	2,35
$\hat{\beta}_2$	33,44	31,17	4,75	30,98	3,65	3,66	3,67	3,69	3,65	3,66
$\hat{\beta}_3$	22,06	21,39	0,17	19,98	-1,58	-1,59	-1,60	-1,59	-1,59	-1,59
$\hat{\beta}_4$	17,72	16,51	1,81	16,27	0,74	0,78	0,79	0,78	0,81	0,81
$\hat{\beta}_5$	25,25	22,84	2,58	24,14	1,77	1,74	1,76	1,76	1,72	1,73
$\hat{\beta}_6$	32,16	42,73	234,26	45,60	1,13	1,16	1,01	0,86	0,48	1,01
$\hat{\beta}_7$	37,01	34,33	2,11	34,26	1,25	1,29	1,29	1,31	1,29	1,30
$\hat{\beta}_8$	29,39	28,98	-1,17	28,20	-1,44	-1,47	-1,49	-1,49	-1,52	-1,50
$\hat{\sigma}$	57,40	55,85	10,31	55,17	4,80	4,82	4,81	4,81	4,79	4,81
$\hat{p}$	0,73056	0,98846	0,01551	0,07817	0,5097	0,4823	0,4867	0,4925	0,5101	0,5026
Iter.	994,00	697,00	4,00	327,00	28,00	31,00	32,00	49,00	58,00	43,00
Loglikelihood.	-1881,57	-1872,28	-1313,91	-1869,27	-1437,57	-1454,11	-1492,11	-1427,80	-1507,35	-1667,1
Varianzen										
$K(\hat{p}, \hat{p})$	0,52734	0,05259	0,00012	0,03549	0,00083	0,00085	0,00081	0,00076	0,00074	0,00081
$K(\hat{a}, \hat{a})$	23,44	27,89	5,21	29,58	0,11	0,12	0,13	0,28	0,30	0,16
$K(\hat{d}, \hat{d})$	57,03	39,59	1,28	43,75	4,77	4,82	5,01	5,75	6,26	4,91
$K(\hat{\sigma}, \hat{\sigma})$	63399,92	59096,98	65,66	55988,36	3,12	3,13	3,13	3,15	3,20	3,14
$K(\hat{\beta}_1, \hat{\beta}_1)$	19,68	18,89	0,70	18,36	0,15	0,16	0,16	0,16	0,15	0,15
$K(\hat{\beta}_2, \hat{\beta}_2)$	17,31	16,53	0,63	16,18	0,14	0,14	0,14	0,14	0,14	0,14
$K(\hat{\beta}_3, \hat{\beta}_3)$	17,54	16,64	0,61	16,41	0,13	0,13	0,13	0,13	0,13	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	23,95	22,73	0,79	22,38	0,17	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_5, \hat{\beta}_5)$	22,99	21,99	0,78	21,35	0,17	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	20,33	25,09	4,61	25,81	0,16	0,16	0,20	0,41	0,44	0,19
$K(\hat{\beta}_7, \hat{\beta}_7)$	17,62	16,82	0,68	16,42	0,15	0,15	0,15	0,15	0,15	0,15
$K(\hat{\beta}_8, \hat{\beta}_8)$	19,85	18,70	0,72	18,33	0,16	0,16	0,16	0,16	0,16	0,16

Die Kofaktorenlösung liefert in EMV (Tab. 6.6) das weitaus höchste Maximum im Intervall 47, wo ein Kofaktormarker auftritt. Dieses Maximum taucht in Abb. 6.22 bei einem

\* Der Marker am linken Rand von Intervall 42 ist Kofaktor

Diskretisierungsgrad von 1 cM nicht auf, erscheint aber, wenn am linken Rand des Intervalles 47 eine hinreichend feine Diskretisierung gewählt wird. Die oszillatorischen Lösungen sind in SIM (Abb. 6.25) und CIM (Abb. 6.23) extrem ähnlich, d. h. sie besitzen mit hoher Genauigkeit dieselben Positionen (Differenzen kleiner als 1 cM) und sehr kleine Unterschiede in den Differenzen der Maximalwerte, z. B. zwischen dem höchsten und zweithöchsten Maximum jeweils in CIM und SIM.

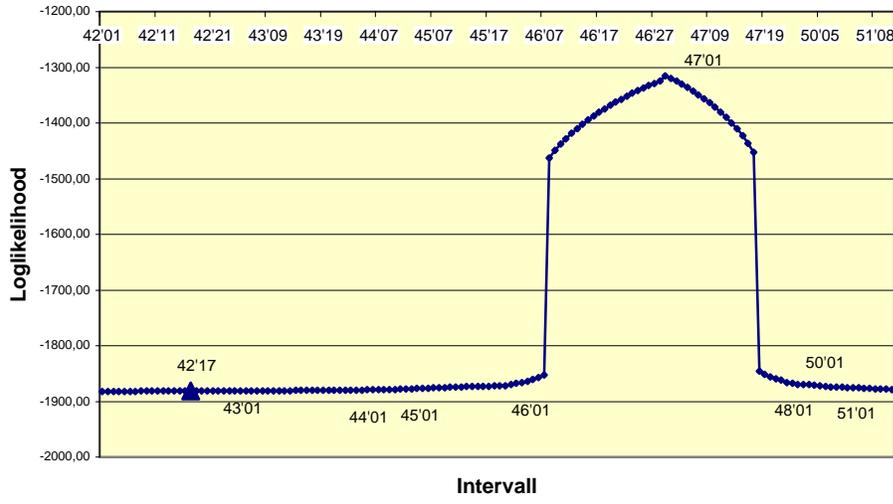


Abbildung 6.22: Loglikelihood-Profil Chromosom 5, CIM mit Startwert 1

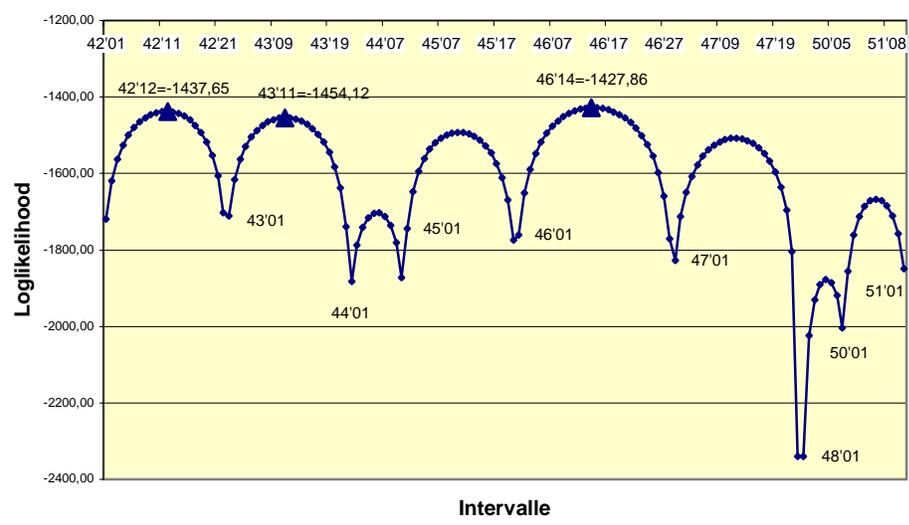


Abbildung 6.23: Loglikelihood-Profil Chromosom 5, CIM mit Startwert 4

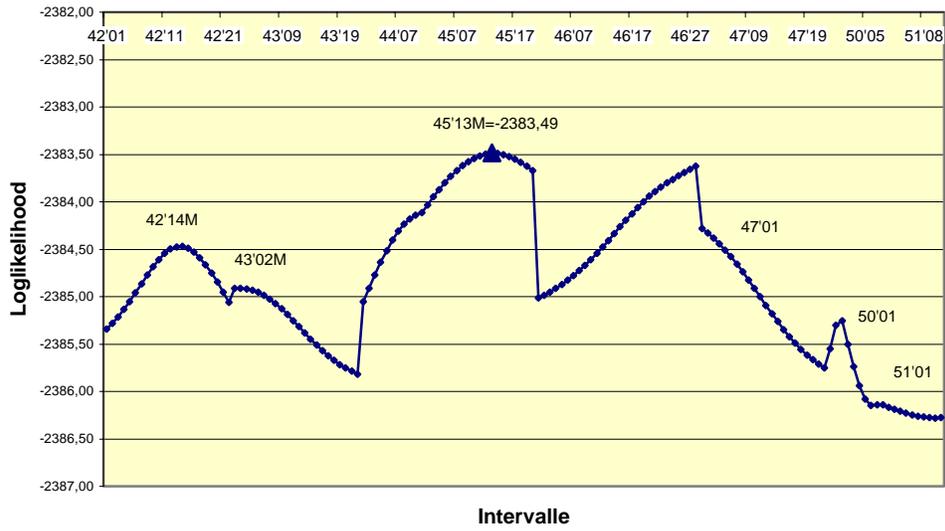


Abbildung 6.24: Loglikelihood-Profil Chromosom 5, SIM (kleine Funktionswerte!), stabiler Startwert

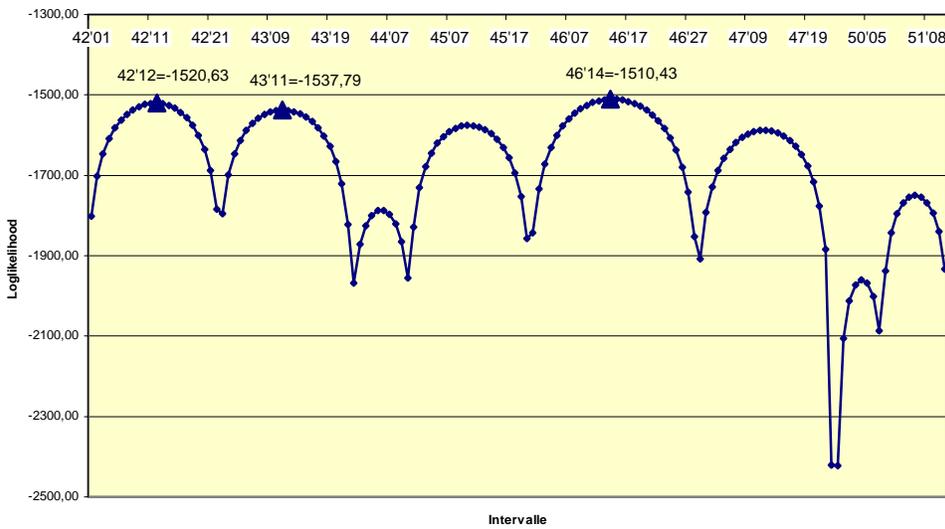


Abbildung 6.25: Loglikelihood-Profil Chromosom 5, SIM, oszillatorische Lösung, (Startwert OL)

## Tabellarische und graphische Auswertung Chromosom 6

Tabelle 6.7: Wie Tabelle 6.3 für Chromosom 6

Startwert 1                      Startwert 4

Intervalle	55	59	55	56	57	59	60 <sup>16</sup>
$\hat{a}$	4,98	-181,59	0,33	0,29	-0,04	-0,26	-0,99
$\hat{d}$	-13,68	4,91	-481,67	-481,69	-481,79	-481,39	-479,70
$\hat{\beta}_1$	31,13	11,08	2,36	2,37	2,35	2,36	2,35
$\hat{\beta}_2$	32,35	11,75	3,65	3,66	3,67	3,67	3,70
$\hat{\beta}_3$	22,33	2,00	-1,57	-1,58	-1,60	-1,60	-1,61
$\hat{\beta}_4$	17,60	6,91	0,81	0,81	0,79	0,79	0,82
$\hat{\beta}_5$	25,93	4,53	1,73	1,74	1,73	1,72	1,72
$\hat{\beta}_6$	31,03	4,90	1,20	1,19	1,19	1,17	1,18
$\hat{\beta}_7$	37,20	191,90	1,31	1,30	1,33	1,56	2,31
$\hat{\beta}_8$	30,12	8,67	-1,45	-1,45	-1,48	-1,47	-1,45
$\hat{\sigma}$	57,13	26,56	4,81	4,81	4,82	4,82	4,77
$\hat{p}$	0,60908474	0,92061081	0,5011	0,4828	0,4767	0,5141	0,4953
Iter.	305,00	6,00	35,00	34,00	35,00	83,00	359,00
Loglikelihood.	-1880,51	-1679,10	-1591,67	-1820,67	-1294,97	-1494,91	-2248,35
Varianzen							
$K(\hat{p}, \hat{p})$	0,1298455	0,0005963	0,000844	0,000805	0,000850	0,000733	0,000755
$K(\hat{a}, \hat{a})$	21,59	25,87	0,12	0,12	0,12	0,41	1,67
$K(\hat{d}, \hat{d})$	55,32	8,62	4,76	4,76	4,75	5,72	12,73
$K(\hat{\sigma}, \hat{\sigma})$	62761,68	2892,60	3,13	3,13	3,14	3,15	4,65
$K(\hat{\beta}_1, \hat{\beta}_1)$	19,65	4,57	0,16	0,16	0,16	0,16	0,16
$K(\hat{\beta}_2, \hat{\beta}_2)$	17,30	4,05	0,14	0,14	0,14	0,14	0,14
$K(\hat{\beta}_3, \hat{\beta}_3)$	17,39	4,11	0,13	0,13	0,14	0,14	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	23,71	5,21	0,17	0,17	0,17	0,17	0,18
$K(\hat{\beta}_5, \hat{\beta}_5)$	22,50	5,20	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	19,93	4,84	0,16	0,16	0,16	0,16	0,16
$K(\hat{\beta}_7, \hat{\beta}_7)$	17,59	22,65	0,15	0,15	0,16	0,51	1,85
$K(\hat{\beta}_8, \hat{\beta}_8)$	19,56	4,56	0,16	0,16	0,16	0,16	0,16

Die oszillierende Lösung (Abb. 6.27) liefert ein stark dominantes Maximum. Die Ähnlichkeit der CIM-Lösung und der SIM-Lösung (Abb. 6.29) ist wiederum erstaunlich.

---

<sup>16</sup> Der Marker 61 am rechten Rand von Intervall 60 ist Kofaktor!

Im Intervall 60 steigt die Loglikelihood-Funktion mit Startwert 1 (CIM) sehr stark an, und zwar über Werte von -1050 hinaus (vgl. Abb. 6.26). Sie besitzt offenbar ein Supremum am rechten Rand dieses Intervalles - höchstwahrscheinlich deswegen, weil an der rechten Intervallgrenze von I60 der Kofaktormarker 61 auftritt. Maxima besitzt dieser Funktionszweig jedoch nur in den Intervallen 59 und 55 mit kleinen bzw. sehr kleinen Werten.

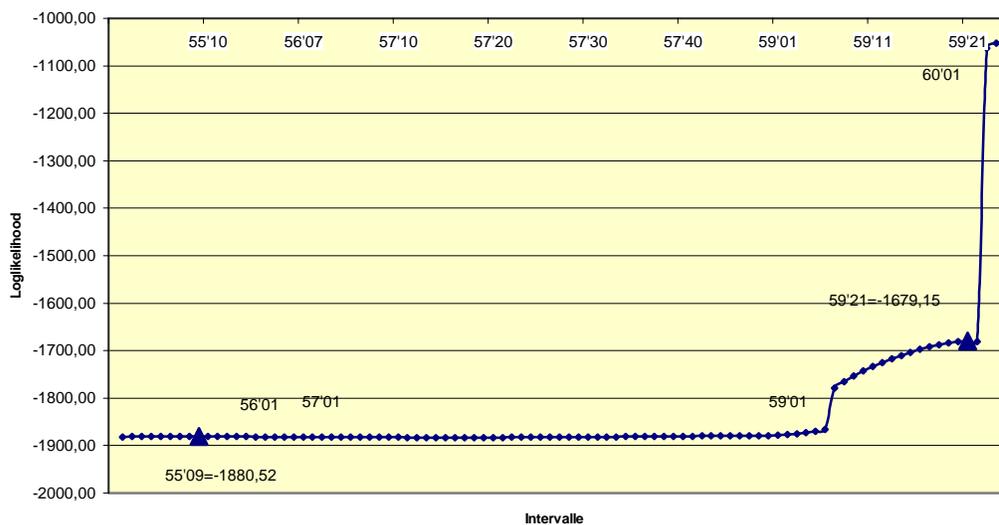


Abbildung 6.26: Loglikelihood-Profil Chromosom 6, CIM mit Startwert 1

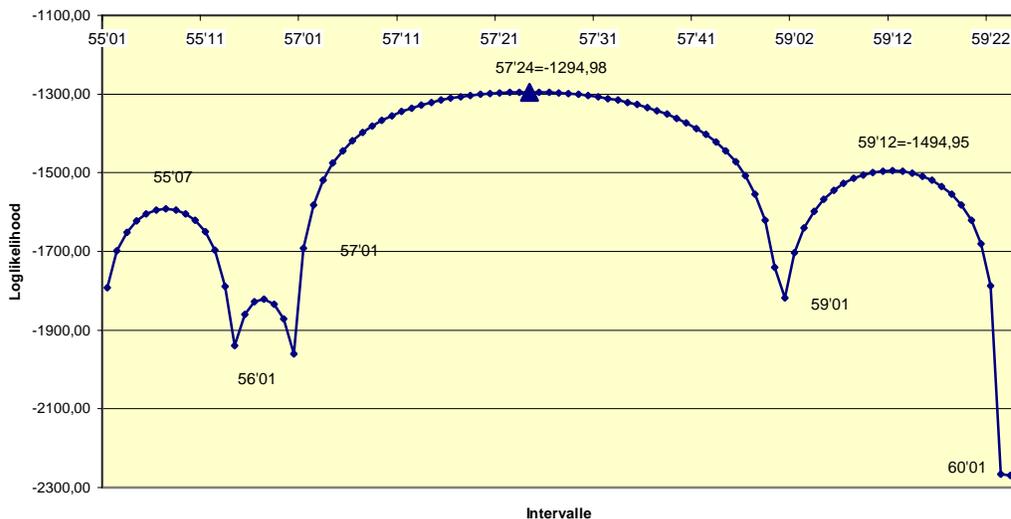


Abbildung 6.27: Loglikelihood-Profil Chromosom 6, CIM mit Startwert 4

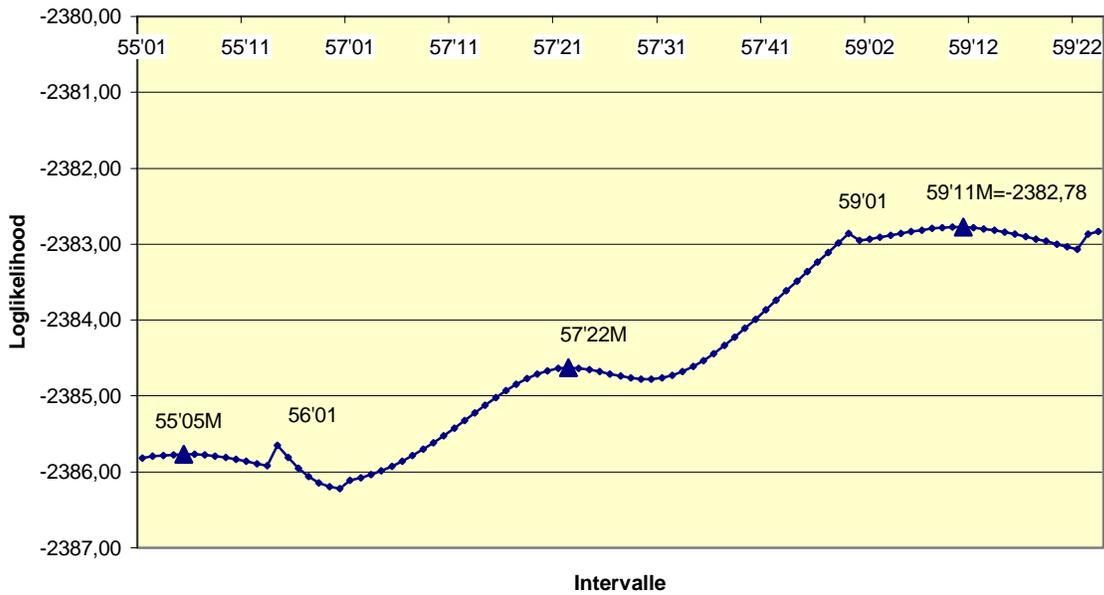


Abbildung 6.28: Loglikelihood-Profil Chromosom 6, SIM (kleine Funktionswerte), stabiler Startwert

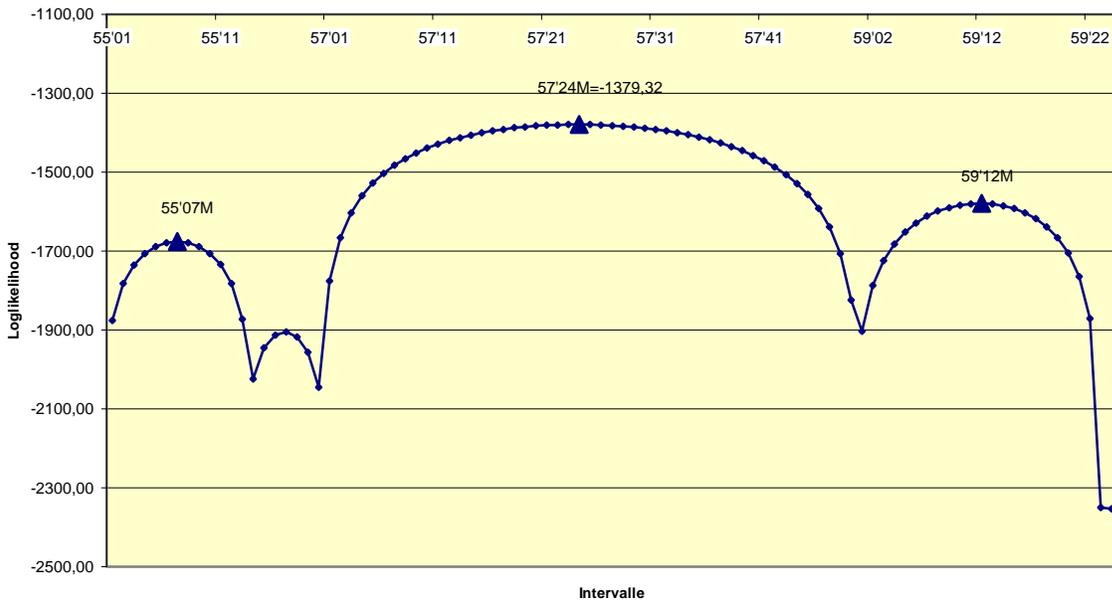


Abbildung 6.29: Loglikelihood-Profil Chromosom 6, SIM, oszillatorische Lösung, Startwert OL.

## Tabellarische und graphische Auswertung Chromosom 7

Tabelle 6.8: Wie Tabelle 6.3 für Chromosom 7

Startwert 1

Startwert 4

Intervalle	I62	I63	I70	I62	I63	I64	I65	I66	I68	I69
$\hat{a}$	-4,74	-4,75	-8,53	0,34	0,07	-0,56	-0,72	-1,22	-1,61	-1,46
$\hat{d}$	0,44	4,31	0,85	-481,91	-481,82	-481,65	-481,75	-481,20	-481,05	-481,44
$\hat{\beta}_1$	31,19	30,81	31,23	2,34	2,35	2,30	2,27	2,26	2,38	2,34
$\hat{\beta}_2$	33,17	33,44	33,05	3,67	3,67	3,69	3,69	3,70	3,78	3,71
$\hat{\beta}_3$	22,31	22,03	22,35	-1,58	-1,59	-1,63	-1,66	-1,70	-1,66	-1,58
$\hat{\beta}_4$	17,24	17,23	17,60	0,80	0,79	0,78	0,76	0,77	0,88	0,84
$\hat{\beta}_5$	26,61	26,43	25,41	1,70	1,73	1,78	1,80	1,84	1,68	1,57
$\hat{\beta}_6$	31,29	31,53	31,78	1,20	1,20	1,22	1,24	1,34	1,23	1,33
$\hat{\beta}_7$	36,43	36,53	36,42	1,29	1,31	1,34	1,33	1,37	1,36	1,29
$\hat{\beta}_8$	29,83	29,98	30,17	-1,49	-1,48	-1,42	-1,41	-1,30	-1,34	-1,30
$\hat{\sigma}$	57,55	57,51	57,33	4,81	4,82	4,79	4,77	4,67	4,55	4,60
$\hat{p}$	0,02935	0,65129	0,59004	0,4861	0,5184	0,4989	0,4865	0,4963	0,5260	0,5064
Iter.	876,00	8,00	578,00	33,00	34,00	39,00	38,00	29,00	31,00	32,00
Loglikelihood.	-1882,28	-1882,11	-1881,28	-1765,42	-1882,78	-1846,66	-2536,28	-1496,18	-1528,00	-1544,92
Varianzen										
$K(\hat{p}, \hat{p})$	3,4981	2,8514	0,4097	0,0008	0,0007	0,0007	0,0007	0,0008	0,0007	0,0008
$K(\hat{a}, \hat{a})$	22,51	28,20	24,69	0,13	0,12	0,13	0,13	0,10	0,09	0,10
$K(\hat{d}, \hat{d})$	47,16	59,35	59,75	4,75	4,76	4,73	4,70	4,61	4,48	4,50
$K(\hat{\sigma}, \hat{\sigma})$	63834,91	63702,48	63205,77	3,13	3,14	3,13	3,13	2,97	2,82	2,91
$K(\hat{\beta}_1, \hat{\beta}_1)$	19,98	19,82	19,70	0,15	0,16	0,16	0,16	0,15	0,15	0,15
$K(\hat{\beta}_2, \hat{\beta}_2)$	17,47	17,35	17,25	0,14	0,14	0,14	0,14	0,13	0,13	0,13
$K(\hat{\beta}_3, \hat{\beta}_3)$	17,76	17,66	17,47	0,13	0,14	0,13	0,13	0,13	0,13	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	24,02	24,18	24,28	0,17	0,17	0,17	0,17	0,17	0,16	0,16
$K(\hat{\beta}_5, \hat{\beta}_5)$	23,06	22,99	22,75	0,17	0,17	0,17	0,17	0,17	0,16	0,16
$K(\hat{\beta}_6, \hat{\beta}_6)$	20,19	20,79	20,10	0,16	0,16	0,16	0,16	0,16	0,15	0,15
$K(\hat{\beta}_7, \hat{\beta}_7)$	17,88	17,92	17,43	0,15	0,15	0,15	0,15	0,14	0,14	0,14
$K(\hat{\beta}_8, \hat{\beta}_8)$	19,96	19,84	19,71	0,16	0,16	0,16	0,16	0,16	0,15	0,15

Die oszillatorische Lösung (Abb. 6.31) ist in CIM dominant und besitzt eine außerordentlich große Ähnlichkeit mit der oszillatorischen Lösung des SIM (Abb. 6.33), wenn auch die Differenzen zwischen den drei höchsten Maximalwerten beim Übergang von CIM zu SIM etwas verändert werden.

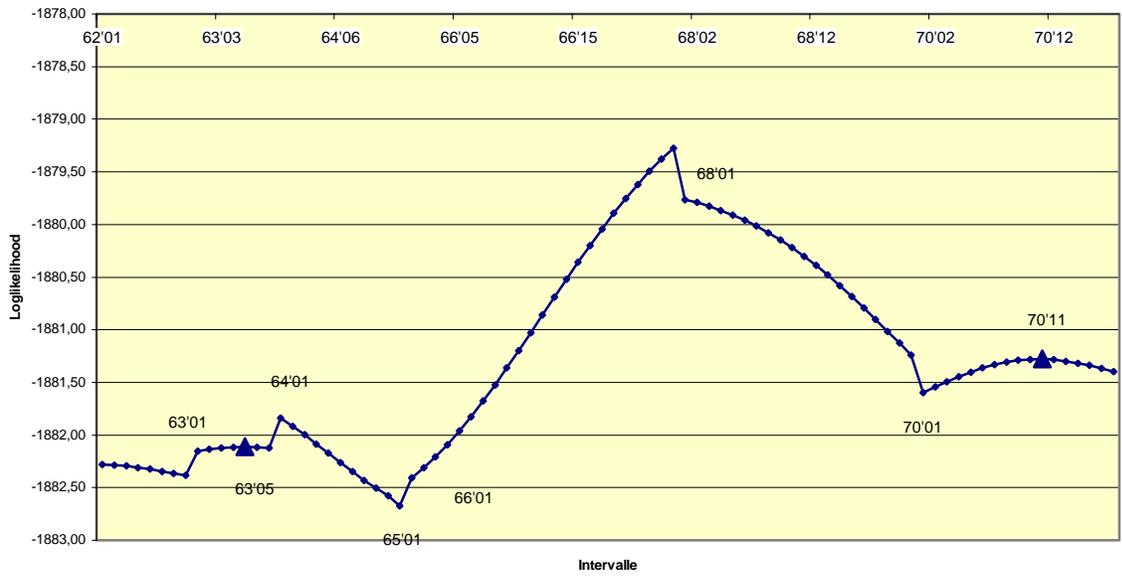


Abbildung 6.30: Loglikelihood-Profil Chromosom 7, CIM mit Startwert 1 (kleine Funktionswerte)

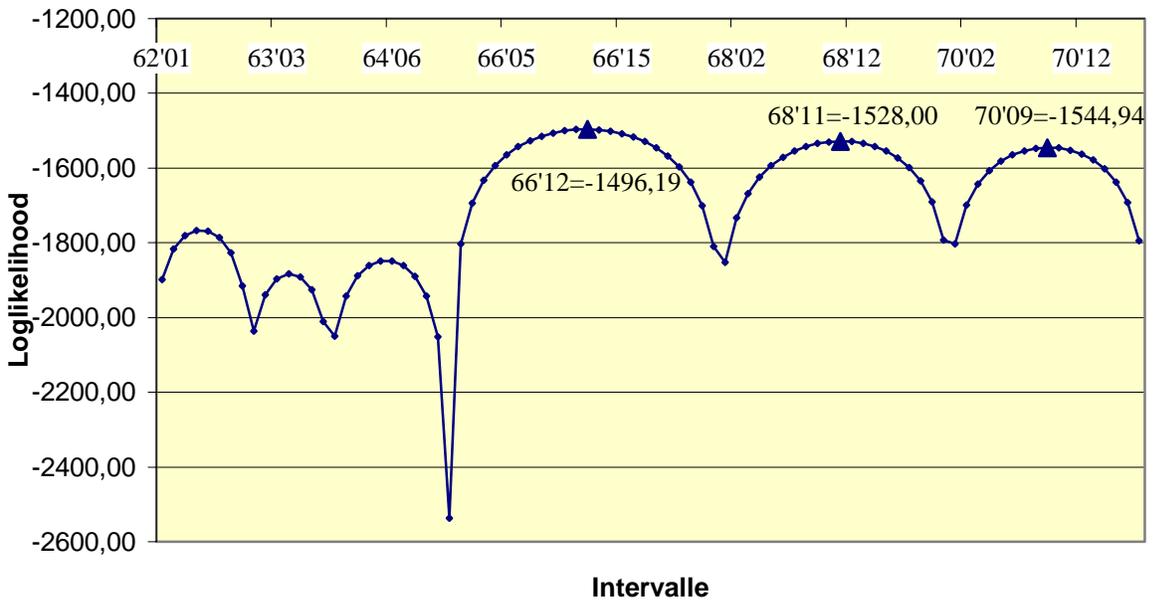


Abbildung 6.31: Loglikelihood-Profil Chromosom 7, CIM mit Startwert 4

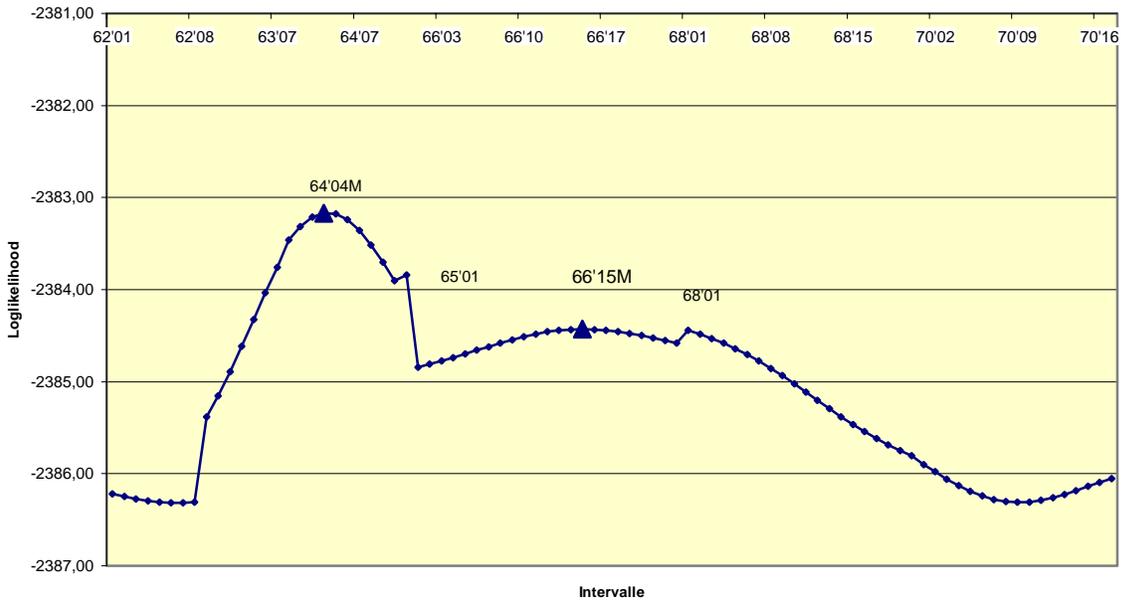


Abbildung 6.32: Loglikelihood-Profil Chromosom 7, SIM (kleine Funktionswerte), stabiler Startwert

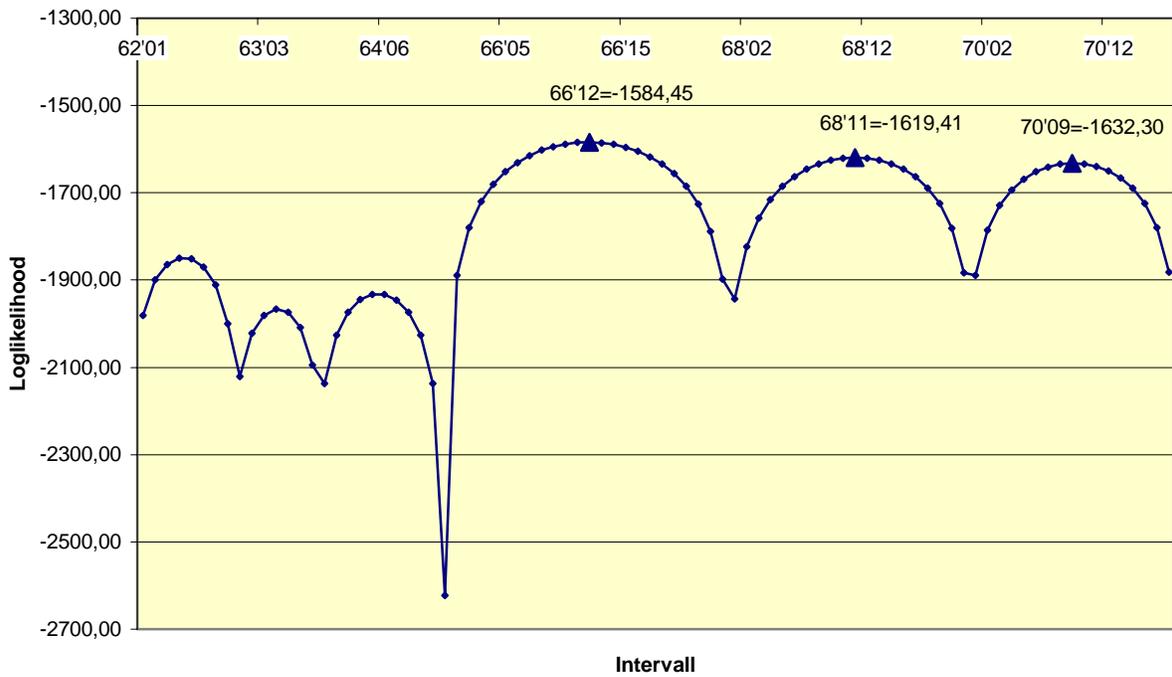


Abbildung 6.33: Loglikelihood-Profil Chromosom 7, SIM, oszillatorische Lösung, Startwert OL

## Tabellarische und graphische Auswertung Chromosom 8

Tabelle 6.9: Wie Tabelle 6.3 für Chromosom 8

	Startwert 1		Startwert 4							
Intervalle	72	78	71	72	73	74	75	76	77	78
$\hat{a}$	10,92	13,46	0,72	0,65	0,44	0,43	-0,54	-0,54	-0,18	-0,34
$\hat{d}$	-5,10	-12,56	-481,47	-481,36	-481,54	-481,51	-482,32	-482,43	-482,00	-482,13
$\hat{\beta}_1$	29,95	29,00	2,31	2,31	2,31	2,31	2,39	2,37	2,36	2,37
$\hat{\beta}_2$	32,22	33,07	3,65	3,64	3,66	3,66	3,66	3,64	3,66	3,66
$\hat{\beta}_3$	22,56	22,71	-1,58	-1,57	-1,57	-1,57	-1,63	-1,64	-1,61	-1,63
$\hat{\beta}_4$	17,03	17,65	0,78	0,80	0,80	0,80	0,75	0,78	0,78	0,77
$\hat{\beta}_5$	27,28	27,37	1,76	1,79	1,79	1,78	1,66	1,64	1,71	1,70
$\hat{\beta}_6$	31,38	30,64	1,20	1,20	1,21	1,22	1,17	1,19	1,20	1,20
$\hat{\beta}_7$	36,21	36,15	1,37	1,35	1,32	1,33	1,28	1,27	1,29	1,29
$\hat{\beta}_8$	30,54	30,50	-1,42	-1,39	-1,44	-1,43	-1,52	-1,52	-1,49	-1,52
$\hat{\sigma}$	57,05	56,37	4,77	4,78	4,80	4,80	4,79	4,79	4,82	4,81
$\hat{p}$	0,80806	0,10264	0,5440	0,4999	0,5157	0,4940	0,5028	0,5217	0,4618	0,4526
Iter.	1665,00	892,00	31,00	30,00	39,00	31,00	44,00	51,00	65,00	33,00
Loglikelih.	-1879,79	-1876,64	-1528,69	-1455,16	-2105,76	-1716,02	-1608,45	-2072,08	-1736,87	-1260,74
Varianzen										
$K(\hat{p}, \hat{p})$	0,2082	0,0482	0,0007	0,0008	0,0008	0,0008	0,0008	0,0008	0,0009	0,0011
$K(\hat{a}, \hat{a})$	29,10	21,44	0,11	0,11	0,14	0,14	0,15	0,16	0,15	0,12
$K(\hat{d}, \hat{d})$	59,39	79,54	4,72	4,75	4,78	4,79	4,86	4,95	4,94	4,85
$K(\hat{\sigma}, \hat{\sigma})$	62447,75	62898,13	3,09	3,10	3,14	3,14	3,15	3,17	3,14	3,13
$K(\hat{\beta}_1, \hat{\beta}_1)$	19,75	19,56	0,15	0,15	0,16	0,16	0,16	0,15	0,16	0,16
$K(\hat{\beta}_2, \hat{\beta}_2)$	17,19	16,95	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,14
$K(\hat{\beta}_3, \hat{\beta}_3)$	17,28	17,00	0,13	0,13	0,13	0,13	0,13	0,14	0,14	0,14
$K(\hat{\beta}_4, \hat{\beta}_4)$	23,67	23,29	0,17	0,17	0,17	0,17	0,17	0,17	0,17	0,17
$K(\hat{\beta}_5, \hat{\beta}_5)$	22,63	22,19	0,17	0,17	0,17	0,17	0,17	0,18	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	19,94	19,62	0,16	0,16	0,16	0,16	0,16	0,16	0,16	0,16
$K(\hat{\beta}_7, \hat{\beta}_7)$	17,34	17,08	0,15	0,15	0,15	0,15	0,15	0,15	0,15	0,15
$K(\hat{\beta}_8, \hat{\beta}_8)$	19,62	19,13	0,16	0,16	0,16	0,16	0,16	0,16	0,16	0,16

Die oszillatorische Lösung, wiederum sehr ähnlich in SIM (Abb. 6.37) und CIM (Abb. 6.35), liefert das sehr stark dominante Maximum im Intervall 78. Die Abbildungen 6.35 und 6.36 zeigen (man vgl. Abb. 6.1) sehr deutlich, wie stark die Höhe der Oszillatormaxima mit den Intervallbreiten verknüpft sind.

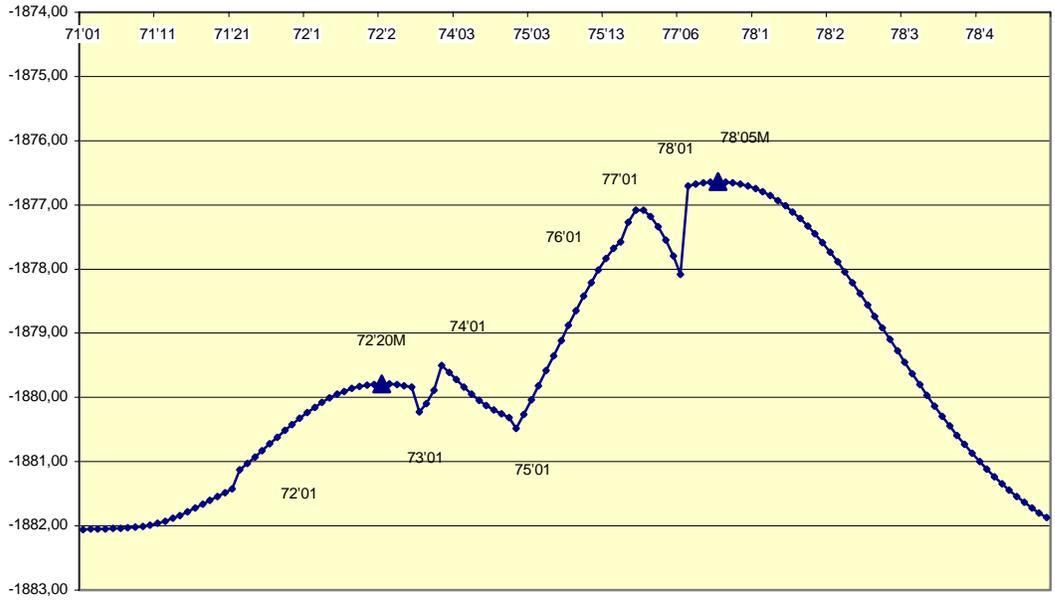


Abbildung 6.34: Loglikelihood-Profil Chromosom 8, CIM mit Startwert 1 (kleine Funktionswerte!)

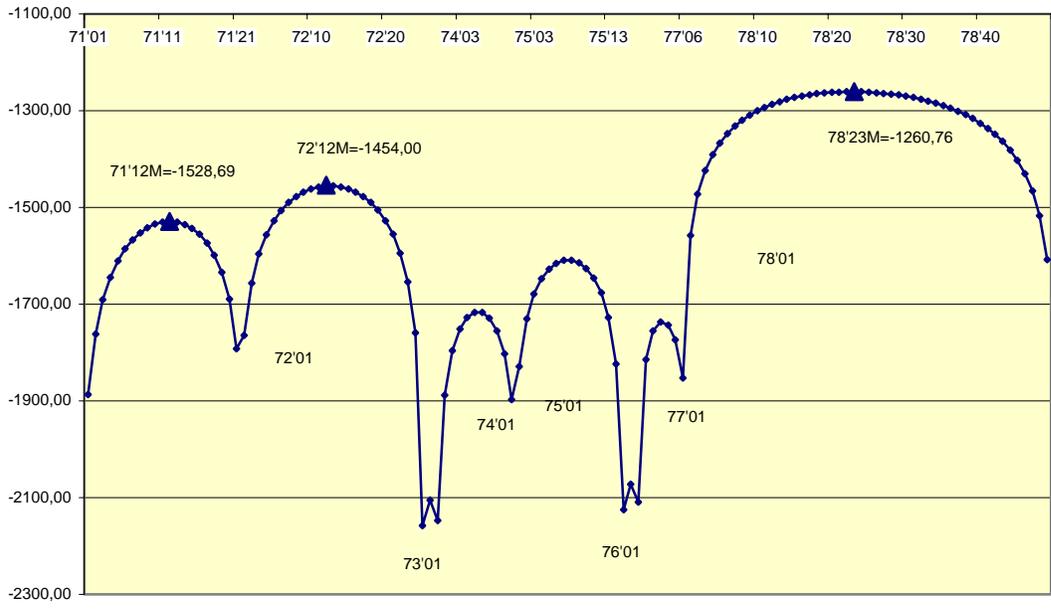


Abbildung 6.35: Loglikelihood-Profil Chromosom 8, CIM mit Startwert 4

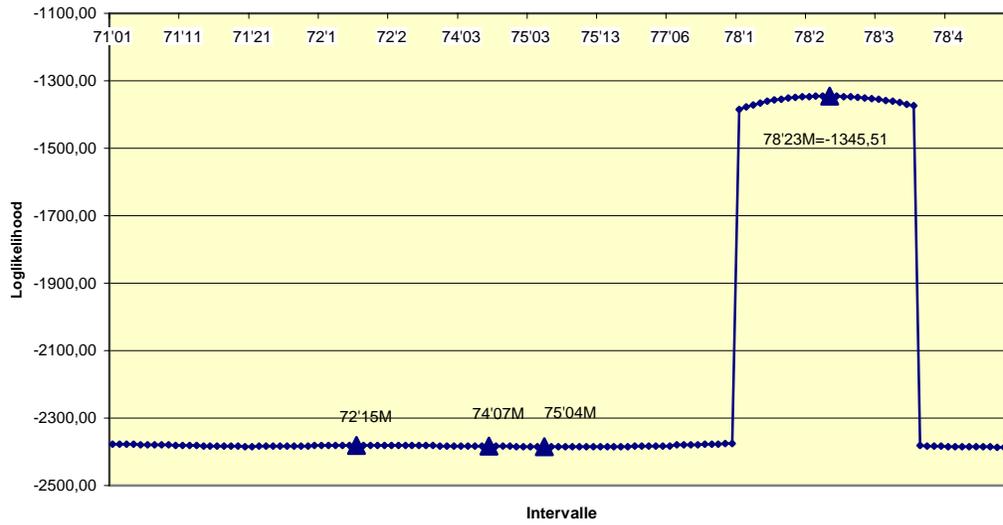


Abbildung 6.36: Loglikelihood-Profil Chromosom 8, SIM, stabiler Startwert

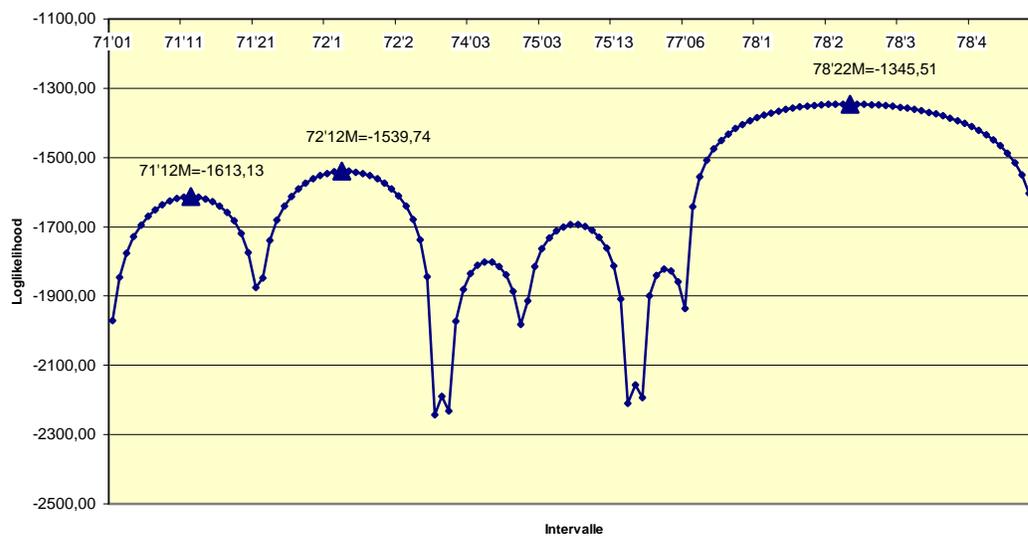


Abbildung 6.37: Loglikelihood-Profil Chromosom 8, SIM, oszillatorische Lösung, Startwert OL

## Tabellarische und graphische Auswertung Chromosom 9

Tabelle 6.10: Wie Tabelle 6.3 für Chromosom 9

Startw. 1                      Startwert 4

Intervalle	83	80	81	82	83
$\hat{a}$	-6,80	-1,11	-1,08	-1,15	-1,11
$\hat{d}$	-23,88	-482,03	-481,81	-481,70	-481,30
$\hat{\beta}_1$	30,13	2,32	2,30	2,39	2,29
$\hat{\beta}_2$	32,72	3,66	3,75	3,77	3,73
$\hat{\beta}_3$	22,16	-1,55	-1,57	-1,64	-1,57
$\hat{\beta}_4$	17,30	0,73	0,77	0,80	0,75
$\hat{\beta}_5$	26,55	1,72	1,67	1,65	1,74
$\hat{\beta}_6$	30,43	1,19	1,10	1,14	1,20
$\hat{\beta}_7$	36,94	1,30	1,44	1,40	1,46
$\hat{\beta}_8$	31,31	-1,50	-1,50	-1,49	-1,40
$\hat{\sigma}$	56,20	4,69	4,70	4,68	4,69
$\hat{p}$	0,805271	0,4899	0,4799	0,4410	0,5643
Iter.	312,00	24,00	29,00	46,00	38,00
Loglikelihood.	-1879,16	-1273,66	-1477,05	-2060,32	-1313,85
Varianzen					
$K(\hat{p}, \hat{p})$	0,031803	0,000915	0,000786	0,000845	0,000926
$K(\hat{a}, \hat{a})$	32,39	0,09	0,10	0,13	0,12
$K(\hat{d}, \hat{d})$	101,53	4,59	4,61	4,63	4,66
$K(\hat{\sigma}, \hat{\sigma})$	69698,74	2,95	3,01	3,14	3,09
$K(\hat{\beta}_1, \hat{\beta}_1)$	19,55	0,15	0,15	0,15	0,15
$K(\hat{\beta}_2, \hat{\beta}_2)$	16,97	0,13	0,14	0,14	0,14
$K(\hat{\beta}_3, \hat{\beta}_3)$	17,19	0,13	0,13	0,13	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	23,43	0,17	0,17	0,17	0,17
$K(\hat{\beta}_5, \hat{\beta}_5)$	22,36	0,17	0,17	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	19,79	0,15	0,16	0,15	0,15
$K(\hat{\beta}_7, \hat{\beta}_7)$	17,51	0,14	0,15	0,14	0,15
$K(\hat{\beta}_8, \hat{\beta}_8)$	19,63	0,15	0,15	0,15	0,15

Die Chromosomen 9 und 10 stellen die einfachsten Fälle dar, weil hier erstens die oszillatorischen Lösungen klar dominieren und diese zweitens nur drei bzw. 2 Oszillationen enthalten. Chromosom 9 zeigt hier wieder die übliche gute Übereinstimmung der oszillatorischen Lösungen in SIM (Abb. 6.41) und CIM (Abb. 6.39).

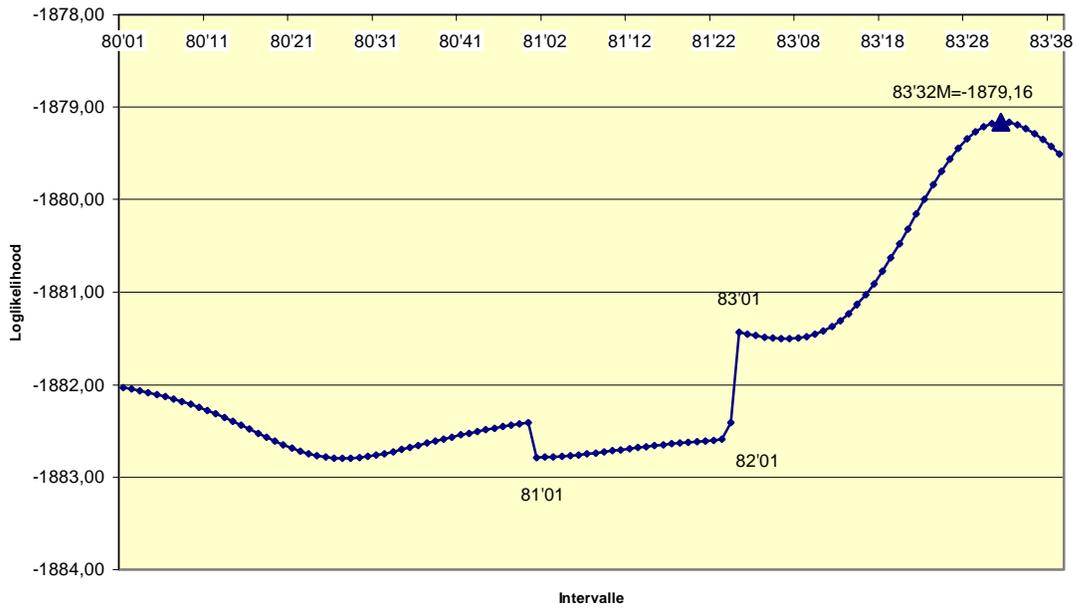


Abbildung 6.38: Loglikelihood-Profil Chromosom 9, CIM mit Startwert 1 (kleine Funktionswerte)

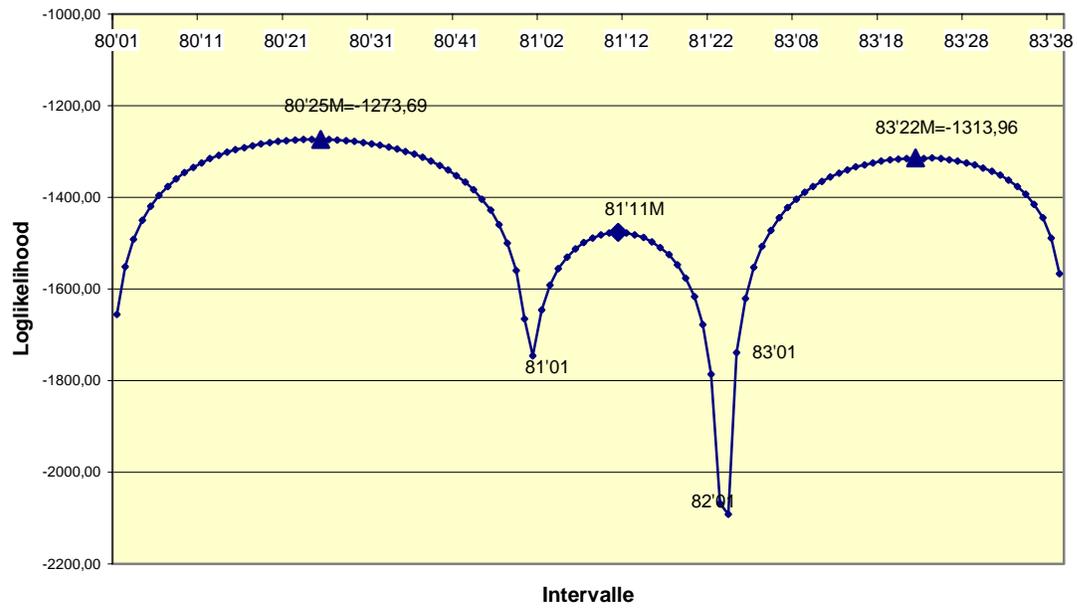


Abbildung 6.39: Loglikelihood-Profil Chromosom 9, CIM mit Startwert 4

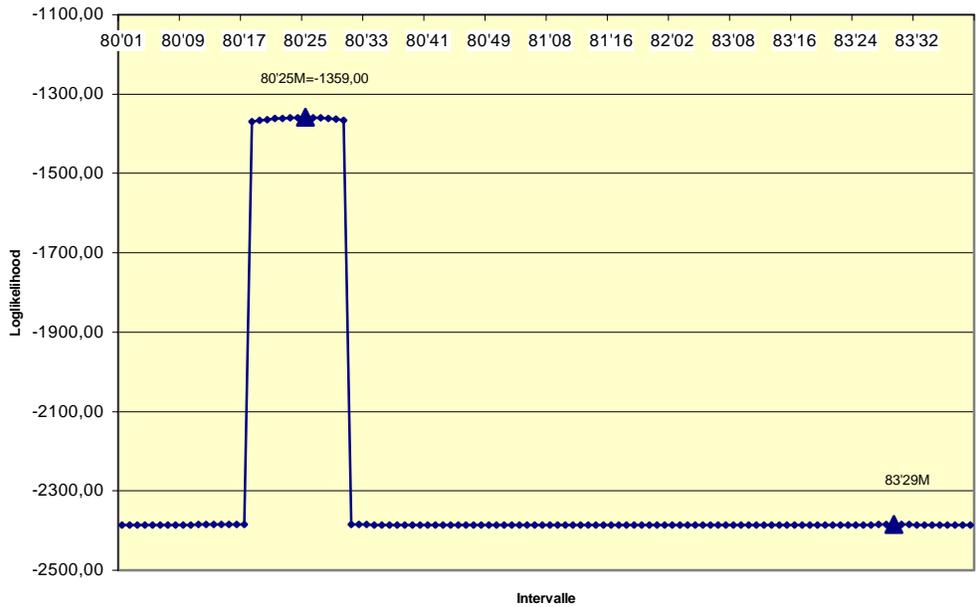


Abbildung 6.40: Loglikelihood-Profil Chromosom 9, SIM, stabiler Startwert

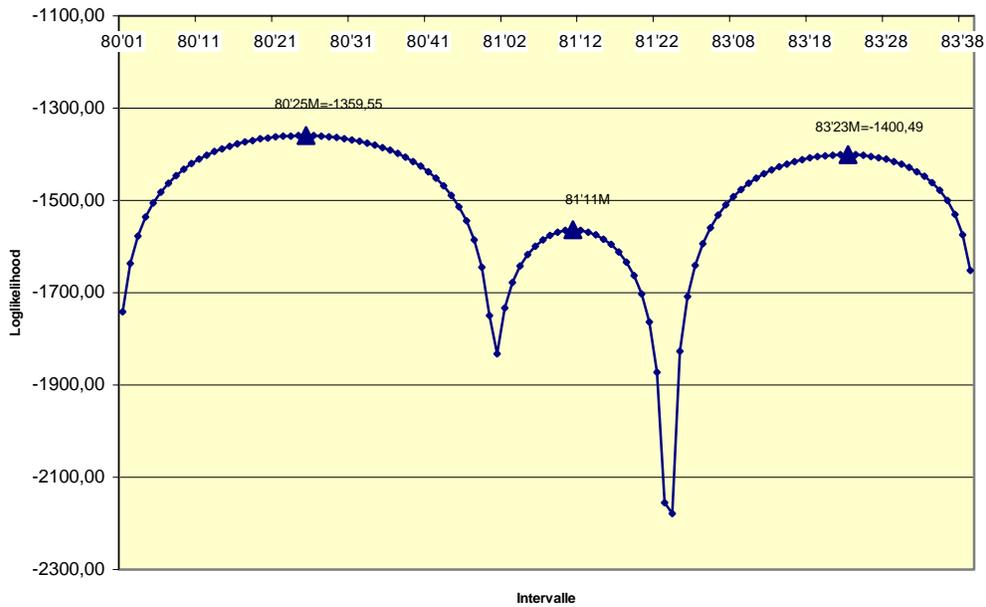


Abbildung 6.41: Loglikelihood-Profil Chromosom 9, SIM, oszillatorische Lösung, Startwert OL

## Tabellarische und graphische Auswertung Chromosom 10

Tabelle 6.11: Wie Tabelle 6.3 für Chromosom 10

Startw. 1      Startwert 4

Intervalle	87*	85	86 Kof.!	87*	88
$\hat{a}$	-201,47	0,26	0,39	-0,15	-0,18
$\hat{d}$	-6,42	-482,14	-482,60	-481,52	-481,61
$\hat{\beta}_1$	7,84	2,34	2,34	2,36	2,35
$\hat{\beta}_2$	9,38	3,68	3,67	3,67	3,67
$\hat{\beta}_3$	2,88	-1,59	-1,60	-1,60	-1,60
$\hat{\beta}_4$	2,74	0,79	0,78	0,80	0,80
$\hat{\beta}_5$	8,56	1,74	1,73	1,74	1,74
$\hat{\beta}_6$	7,46	1,19	1,20	1,20	1,20
$\hat{\beta}_7$	4,49	1,33	1,31	1,32	1,32
$\hat{\beta}_8$	204,20	-1,67	-1,87	-1,35	-1,40
$\hat{\sigma}$	23,32	4,81	4,81	4,82	4,82
$\hat{p}$	0,07166201	0,4664	0,5207	0,5084	0,4913
Iter.	5,00	50,00	199,00	43,00	42,00
Loglikelih.	-1647,43	-1291,26	-2371,94	-1328,61	-2388,59
Varianzen					
$K(\hat{p}, \hat{p})$	0,0003133	0,001072	0,000766	0,000788	0,000757
$K(\hat{a}, \hat{a})$	23,14	0,23	1,04	0,22	0,15
$K(\hat{d}, \hat{d})$	6,55	5,14	9,10	5,55	4,91
$K(\hat{\sigma}, \hat{\sigma})$	1720,40	3,14	3,29	3,14	3,14
$K(\hat{\beta}_1, \hat{\beta}_1)$	3,56	0,16	0,16	0,16	0,16
$K(\hat{\beta}_2, \hat{\beta}_2)$	3,16	0,14	0,14	0,14	0,14
$K(\hat{\beta}_3, \hat{\beta}_3)$	3,10	0,13	0,13	0,13	0,13
$K(\hat{\beta}_4, \hat{\beta}_4)$	4,08	0,17	0,17	0,17	0,17
$K(\hat{\beta}_5, \hat{\beta}_5)$	3,91	0,17	0,17	0,17	0,17
$K(\hat{\beta}_6, \hat{\beta}_6)$	3,62	0,16	0,16	0,16	0,16
$K(\hat{\beta}_7, \hat{\beta}_7)$	3,47	0,15	0,15	0,15	0,15
$K(\hat{\beta}_8, \hat{\beta}_8)$	20,71	0,27	1,20	0,32	0,18

Das Chromosom 10 stellt bezüglich der oszillierenden Lösung den einfachsten Fall schlechthin dar. Die beiden Maxima der oszillatorischen Lösungen sind die weitaus höchsten

---

\* Der Marker 86 am linken Rand von Intervall 87 ist Kofaktor

und diese oszillatorischen Lösungen sind wiederum bezüglich SIM (Abb. 6.45) und CIM (Abb. 6.43) extrem ähnlich.

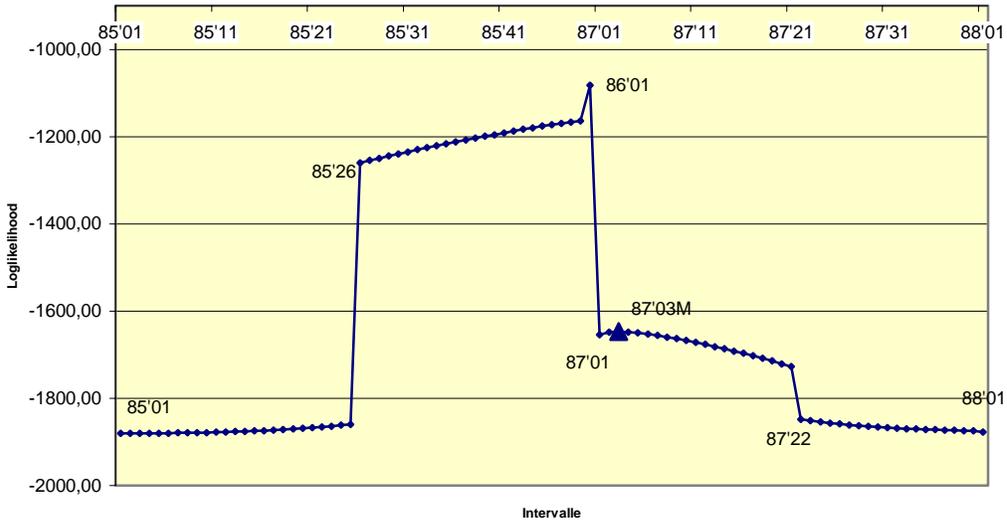


Abbildung 6.42: Loglikelihood-Profil Chromosom 10, CIM mit Startwert 1

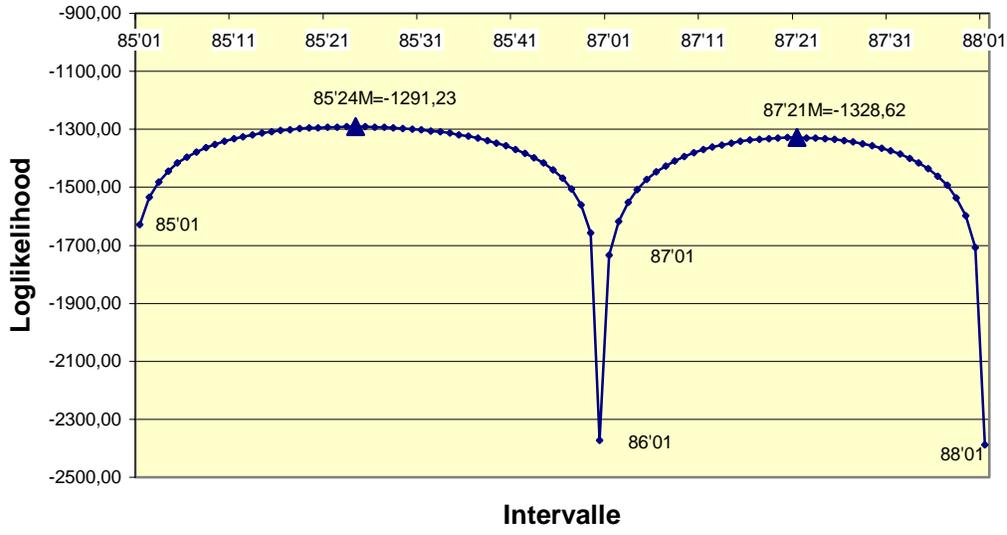


Abbildung 6.43: Loglikelihood-Profil Chromosom 10, CIM mit Startwert 4

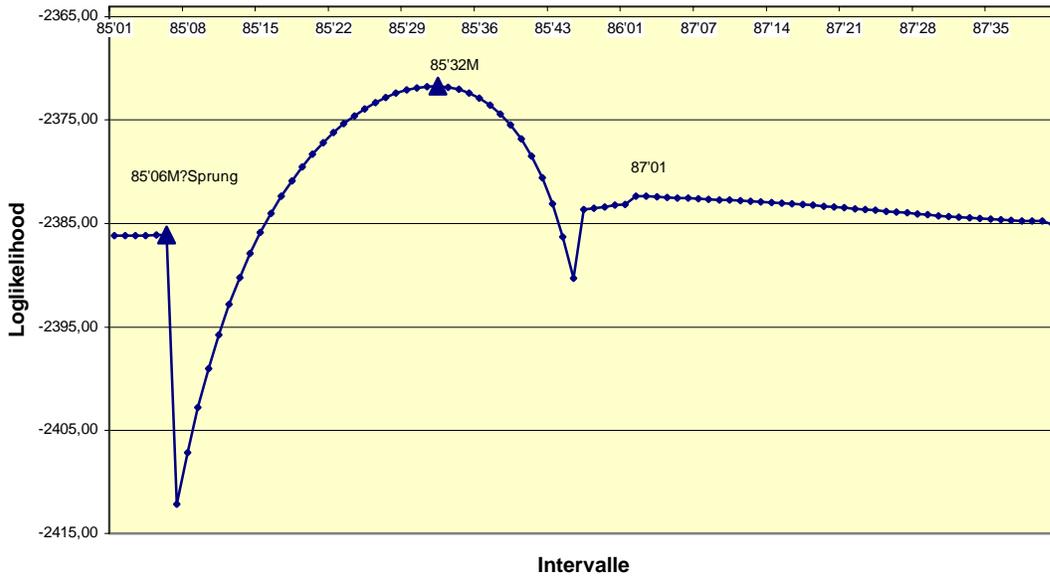


Abbildung 6.44: Loglikelihood-Profil Chromosom 10, SIM, stabiler Startwert

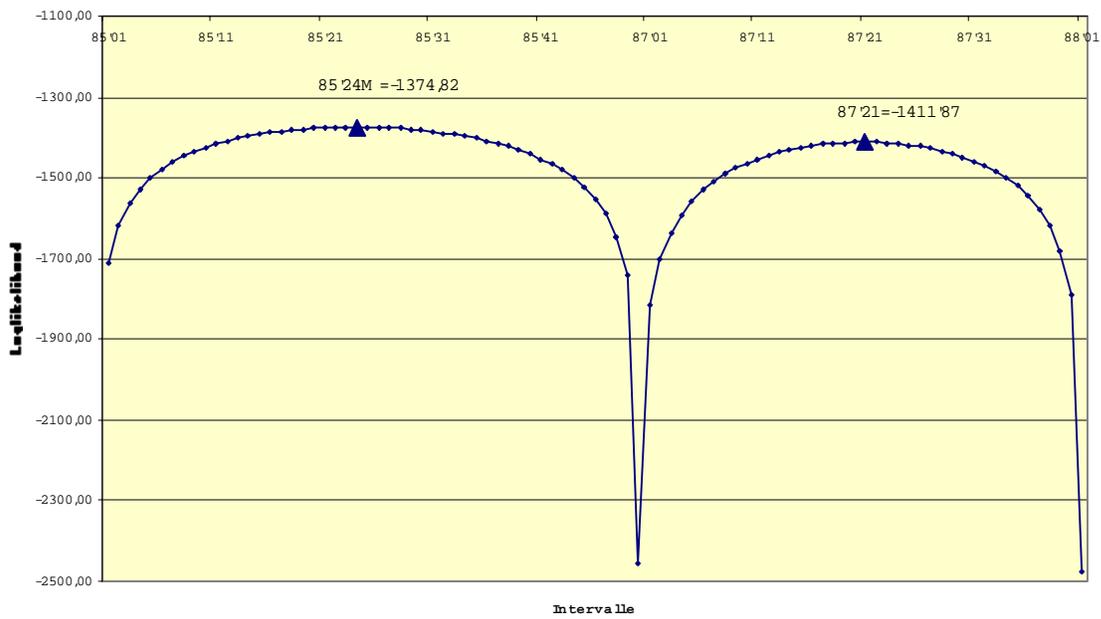


Abbildung 6.46: Loglikelihood-Profil Chromosom 10, SIM, oszill. Lösung, Startwert OL

Im folgenden werden die verschiedenen, bei allen 10 Chromosomen auftauchenden Lösungstypen zusammenfassend erläutert.

### **6.3 Die tiefliegende Lösung im einfachen Intervallkartierungsmodell**

#### **Vergleich mit der linearen Regression (F-Statistik-Profil)**

Es ist bekannt, dass das EM-Verfahren und die lineare Regression (F-Statistik-Profil) sehr ähnliche Ergebnisse liefern, wenn man z. B. die Schätzwerte der Parameter  $a$ ,  $d$  und  $\sigma$  aus dem Regressionsverfahren entnimmt und als Startwerte in das EM-Verfahren für das SIM-Modell einsetzt. In Abbildung 6.46 bis 6.55b ist gezeigt, dass dies in der Tat sowohl für die Grobstruktur als auch viele feine und sehr feine Details in guter, oft sehr guter, Näherung gilt - wenn auch mit Ausnahmen in Teilen der neun Intervalle 2, 4, 10, 24, 28, 36, 78, 80 und 86. Eine Anpassung der Skalen der Ordinaten durch einen Skalenfaktor oder eine Verschiebung wurde nicht vorgenommen.

Wir betrachten jetzt die Intervalle, in welchen die EM-Resultate für  $\ln(LQ)$  von der F-Statistik abweichen. Der Vergleich von Abbildung 6.46 mit Abbildung 6.8 zeigt, dass an den mit Pfeilen gekennzeichneten Positionen von Abbildung 6.46 die  $\ln(LQ)$ -Resultate gegen die oszillatorische Lösung gehen (vgl. auch Abbildung 6.9). Man kann dabei leicht am Rechner erkennen, dass der Parameter  $\sigma$  (als Variable der Loglikelihood im Maximierungsprozess) der treibende Term für das Ansteigen der  $\ln(LQ)$ -Werte ist. Wenn man daher  $\sigma$  bei der Iteration konstant wählt, z. B.  $\sigma = \sigma_0$ , wobei  $\sigma_0$  der zu  $L_0$  gehörende Parameterwert ist ( $\sigma_0 = 249,16$ , Tab. (A1)), dann bleibt die Lösung stabil und bildet Maxima, die in den Intervallen 2 und 10 etwas unterhalb, im Intervall 4 in gleicher Höhe wie die F-Maxima liegen. Interessant ist es auch, die Schätzer  $\hat{a}$ ,  $\hat{d}$  und  $\hat{\sigma}$  zu verfolgen (Tabelle 6.12). Im Bereich des "Sprunges" zwischen Position 2'20 und 2'22 ändern sich alle drei Schätzer (als "Variablen") und  $\ln(L)$  dramatisch,

wobei die Konvergenz sich zunächst beträchtlich verlangsamt, sich aber im Intervall 2'22, wo die oszillierende Lösung schließlich erreicht ist ( $\ln(L)=-1363,88$ ), wieder

**Tabelle 6.12, Chr1, SIM-Resultate an einzelnen Punkten (einzelne Profilwerte)**

Int.	2'07	2'16	2'20	2'21	2'22	2'36	2'45
$\hat{a}$	-56,76	-45,28	-32,75	-27,67	1,28	-23,64	-21,18
$\hat{d}$	9,71	-48,05	-106,54	-138,08	-498,42	-104,15	-40,04
$\hat{\sigma}$	245,96	245,17	241,28	237,55	6,03	242,57	247,77
$\ln(L)$	-2383,06	-2383,02	-2382,43	-2382,18	-1363,88	-2382,95	-2384,98
Iter.	39,00	106,00	225,00	690,00	83,00	105,00	31,00
$\ln(LQ)$	3,277	3,316	3,905	4,155	1022,460	3,390	1,359

Int.	4'06	4'17	4'30	5'05	5'24	8'06	8'11
$\hat{a}$	3,97	5,47	6,49	14,08	42,92	44,92	45,41
$\hat{d}$	-24,49	-97,09	-117,26	22,40	37,74	34,17	14,14
$\hat{\sigma}$	248,84	244,33	242,08	248,73	246,93	246,88	247,12
$\ln(L)$	-2386,07	-2384,76	-2384,64	-2385,91	-2384,50	-2384,19	-2384,56
Iter.	34,00	134,00	249,00	27,00	44,00	36,00	61,00
$\ln(LQ)$	0,264	1,581	1,702	0,426	1,842	2,146	1,777

Int.	8'18	8'22	8'32	8'40	10'07	10'12	10'33
$\hat{a}$	36,57	30,91	27,70	25,90	22,35	20,94	20,45
$\hat{d}$	-35,99	-65,27	-60,59	-27,36	-54,19	-93,94	-93,06
$\hat{\sigma}$	246,84	245,61	246,22	247,99	247,00	243,94	244,05
$\ln(L)$	-2384,39	-2383,99	-2384,07	-2385,02	-2384,33	-2383,00	-2383,60
Iter.	98,00	112,00	64,00	26,00	34,00	63,00	86,00
$\ln(LQ)$	1,945	2,352	2,266	1,316	2,010	3,333	2,733

**Tabelle 6.13: Chr2, SIM-Resultate**

INT.	12,06	13,15	13,27	14,08	14,19	15,09	15,22	16,05
$\hat{a}$	4,24	-1,00	-10,25	8,95	15,10	11,86	14,92	18,83
$\hat{d}$	16,96	-8,89	-1,67	4,67	6,19	13,12	16,90	35,27
$\hat{\sigma}$	249,00	249,12	249,05	249,07	248,92	248,94	248,81	248,25
Logl.	-2386,18	-2386,32	-2386,22	-2386,25	-2386,10	-2386,14	-2385,98	-2385,40
Iter.	23,00	62,00	31,00	33,00	32,00	34,00	25,00	24,00
LQ	0,162	0,021	0,114	0,091	0,234	0,198	0,359	0,933

Int.	16,18	18,03	18,14	19,05	20,06	21,03	21,16	
$\hat{a}$	28,91	21,68	4,11	-6,84	-3,37	-4,76	5,42	
$\hat{d}$	63,88	57,06	39,32	15,60	14,67	33,82	-39,83	
$\hat{\sigma}$	246,53	247,19	248,37	248,99	249,04	248,57	248,32	

Logl.	-2384,13	-2384,19	-2385,54	-2386,15	-2386,21	-2385,76	-2385,54	
Iter.	34,00	20,00	23,00	18,00	20,00	24,00	27,00	
LQ	2,207	2,145	0,793	0,186	0,124	0,577	0,795	

**Tabelle 6.14: Chr3, SIM-Resultate**

INT.	23'05	24'08	24'23	24'35	26'02
$\hat{a}$	23,62	7,07	249,67	251,33	16,63
$\hat{d}$	31,99	-29,64	262,72	257,48	7,18
$\hat{\sigma}$	248,18	248,66	195,11	195,51	248,87
ln(L)	-2385,41	-2386,00	-2382,93	-2384,65	-2385,98
Iter.	28,00	63,00	21,00	23,00	17,00
ln(LQ)	0,929	0,337	3,403	1,687	0,355

INT.	28'03	28'23	28'39	31'02	32'02	33'18
$\hat{a}$	21,17	279,73	274,87	48,20	59,71	58,01
$\hat{d}$	-27,16	292,09	297,16	52,73	31,28	12,20
$\hat{\sigma}$	248,25	183,86	183,96	246,04	245,61	245,84
ln(L)	-2385,27	-2360,56	-2364,95	-2382,58	-2381,84	-2382,72
Iter.	31,00	17,00	15,00	17,00	15,00	34,00
ln(LQ)	1,065	25,774	21,391	3,754	4,500	3,620

**Tabelle 6.15: Chr4, SIM-Resultate**

INT.	35'06	36'07	36'12	36'20	36'36	37'11
$\hat{a}$	-7,56	-26,26	-21,46	0,43	-13,16	1,38
$\hat{d}$	-21,28	-71,94	-110,44	-498,22	-56,41	-28,28
$\hat{\sigma}$	248,87	245,47	242,04	6,15	247,31	248,75
ln(L)	-2386,03	-2383,06	-2381,98	-1379,67	-2384,73	-2386,05
Iter.	22,00	39,00	73,00	74,00	33,00	41,00
ln(LQ)	0,3067	3,2756	4,3603	1006,67	1,6064	0,2890

INT.	37'21	38'03	39'02	39'06	40'12	40'37
$\hat{a}$	9,02	21,16	18,64	22,42	22,61	16,21
$\hat{d}$	-13,76	39,27	48,78	46,28	-0,60	2,34
$\hat{\sigma}$	248,98	248,03	247,70	247,69	248,64	248,90
ln(L)	-2386,17	-2384,97	-2384,57	-2384,58	-2385,87	-2386,05
Iter.	35,00	15,00	15,00	16,00	54,00	32,00
ln(LQ)	0,1706	1,3629	1,7691	1,7587	0,4714	0,2826

Tabelle 6.16: Chr10, SIM-Resultate

INT.	85'04	85'1	85'18	85'32	85'42	85'46	85'49	87'05	87'26
$\hat{a}$	8,97	250,66	256,28	255,86	248,69	36,37	32,54	60,89	54,87
$\hat{d}$	-13,25	275,54	278,69	281,17	282,51	74,60	71,36	39,64	11,68
$\hat{\sigma}$	248,98	192,64	191,00	190,62	191,62	245,50	245,90	245,34	246,19
$\ln(L)$	-2386,17	-2399,03	-2380,89	-2371,73	-2380,58	-2383,69	-2383,26	-2382,54	-2383,92
Iter.	37,00	30,00	21,00	22,00	24,00	83,00	32,00	35,00	105,00
$\ln(LQ)$	0,1637	-12,6971	5,4504	14,6091	5,7544	2,6459	3,0760	3,7976	2,4126

beschleunigt. Außerhalb des Sprungbereiches zwischen 2'20 und 2'30 bleibt der Schätzer des Parameters  $\sigma$  erstaunlich konstant, nämlich zwischen 242,0 und 248,8.

Ein sehr einfaches Bild zeigt dagegen Chromosom 2 (Abb. 6.47). Die drei Lösungen  $\ln(LQ)$ ,  $\ln(LQ)$  für  $\sigma = \sigma_0$  und die F-Statistik sind nahezu identisch und  $\hat{\sigma}$  bleibt annähernd konstant zwischen 246,5 und 249,1 (Tabelle 6.13).

Interessant ist Chromosom 3. Hier gibt es wie Tabelle 6.14 und Abbildung 6.48a zeigen, zwei Lösungstypen. Eine, die mit den F-Werten und den  $\ln(LQ)$ -Werten für  $\sigma = \sigma_0$  gut übereinstimmt und eine, die ein oszillierendes Verhalten zeigt und daher die "zweite oszillierende Lösung" genannt wird. Die erstere weist wiederum einen nahezu konstanten Schätzer  $\hat{\sigma}$  zwischen 245 und 249 und Werte von  $\hat{a}$  und  $|\hat{d}|$  in der Größenordnung von einigen Dutzend auf.

Die letztere dagegen zeigt einen kleineren Wert für  $\hat{\sigma}$ ,  $\hat{d}$ -Werte bei ca. 260 bzw. ca. 295 und  $\hat{a}$ -Werte bei ca. 250 bzw. ca. 277. Interessant ist auch das Verhalten der  $\ln(LQ)$ -Funktion für  $\sigma = \sigma_0$ . Sie zeigt zwischen den Positionen 28'24 und 28'46 andeutungsweise dasselbe Verhalten wie die eigentliche  $\ln(LQ)$ -Funktion, nur viel schwächer ausgebildet (Abbildungen 6.48a und b). Versuchsweise wurde ein anderer Startwert, nämlich für jede in Abbildung 6.48b aufgeführte Position die aus der Regression resultierenden Schätzwerte der Parameter  $a$ ,  $d$  und  $\sigma$  eingesetzt und der jeweilige  $\sigma$ -Wert während des Iterationsverfahrens konstant belassen. Es ergibt sich eine Kurve, die etwas oberhalb der  $\ln(LQ)$ -Funktion mit  $\sigma = \sigma_0$  liegt, aber ansonsten ein völlig analoges Verhalten zeigt (Abbildung 6.48b).

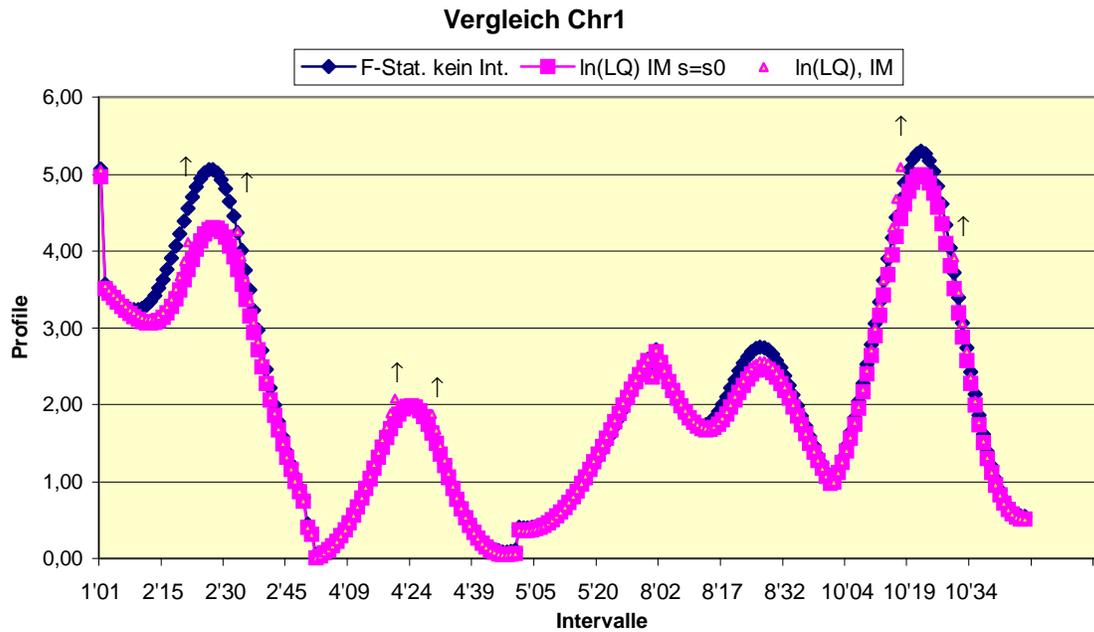


Abbildung 6.46: Für Chrom.1 ist der ln(LQ)-Wert des SIM-Modells angegeben (stabiler Startwert), und zwar zusammen mit den Ergebnissen der linearen Regression und dem ln(LQ)-Wert mit einem in der Iteration konstant belassenem Wert  $\sigma = \sigma_0$  (für s, s0 in Abbildungen lies  $\hat{\sigma}, \sigma_0$ !).

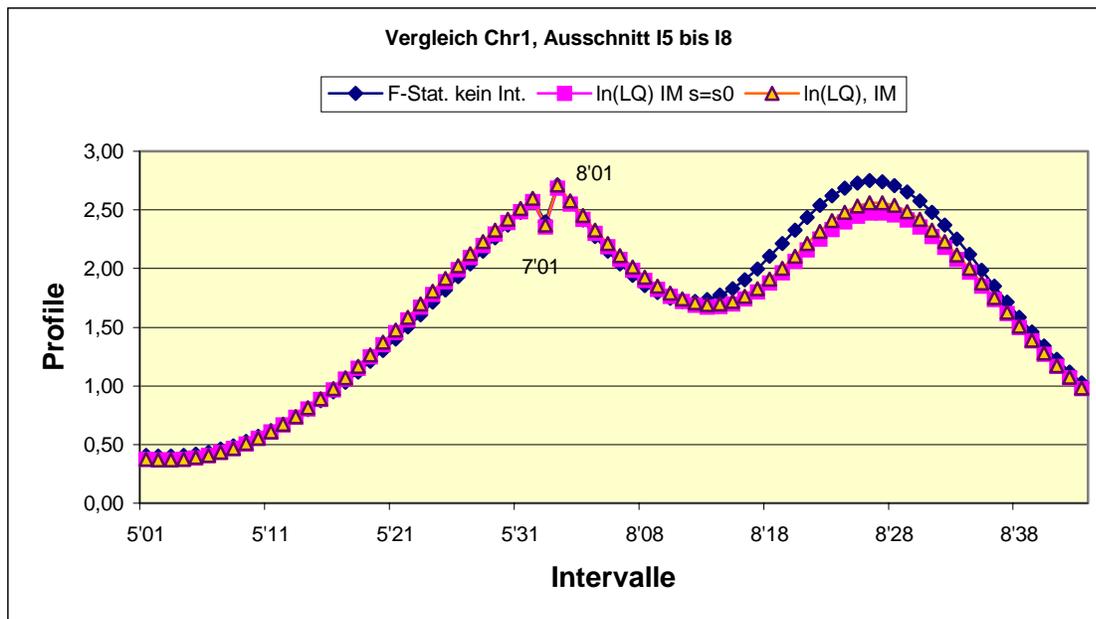


Abbildung 6.46a: Ausschnitt von 6.46

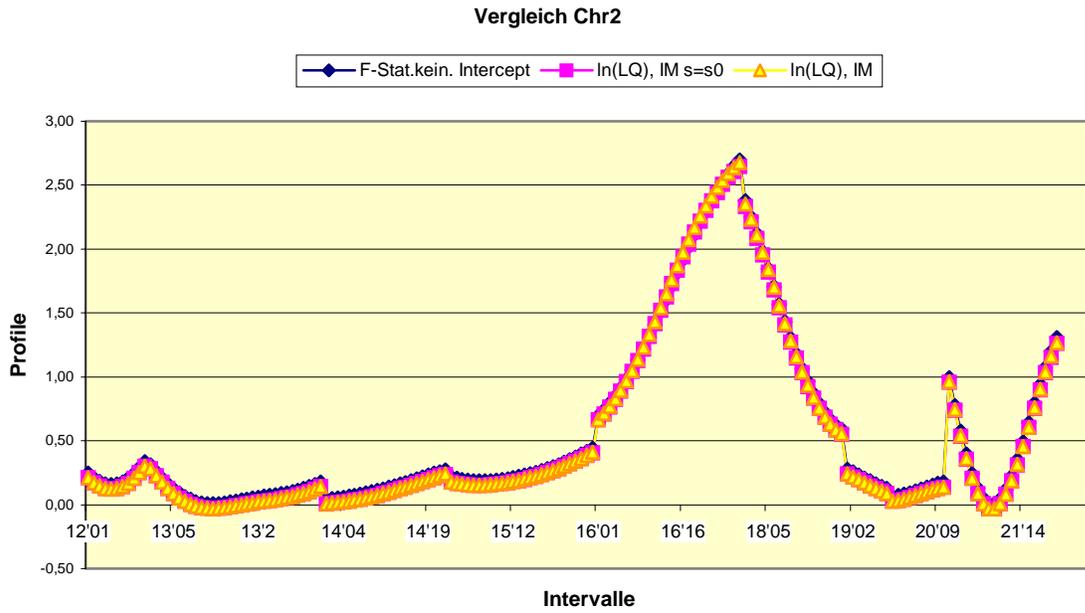


Abbildung 6.47: Wie Abb. 6.46 für Chrom. 2.

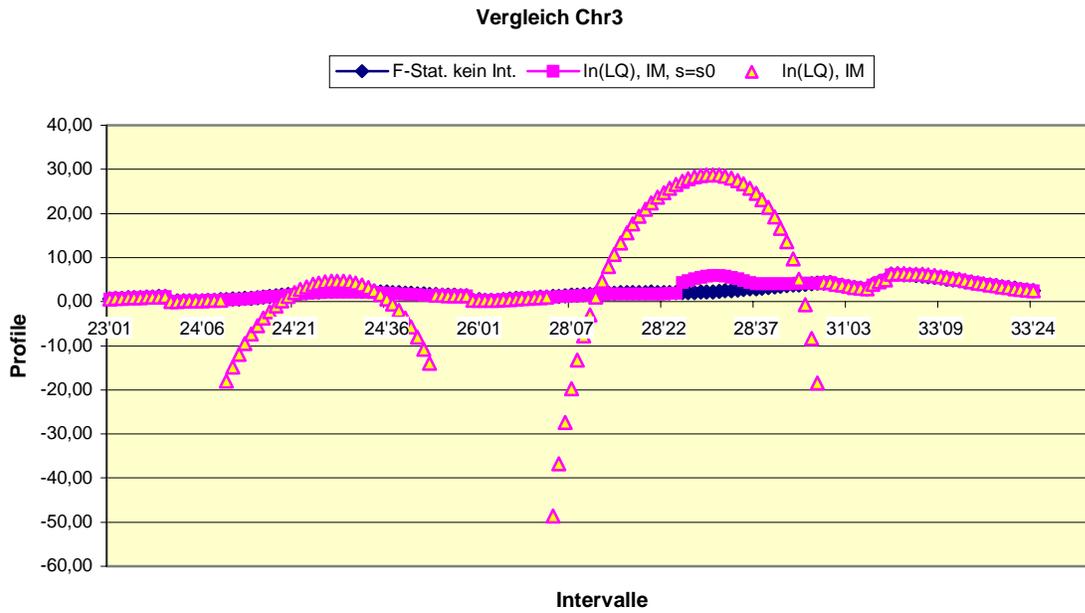


Abbildung 6.48a: Wie Abb. 6.46 für Chrom. 3.

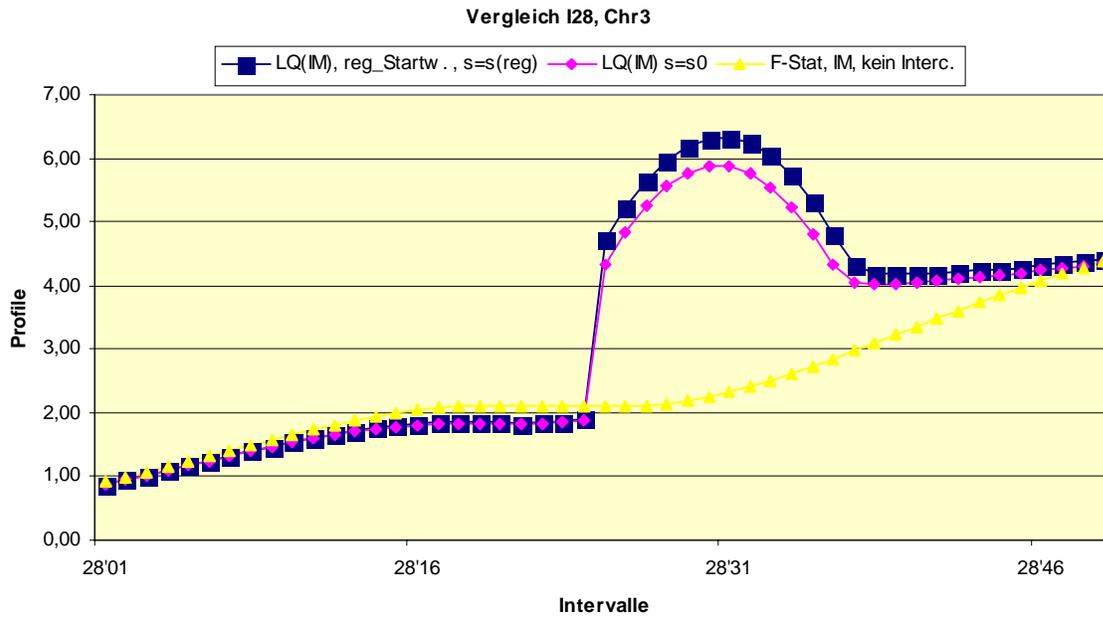


Abbildung 6.48b: Ausschnitt aus Abb. 6.48. Zusätzlich wurde für jede der gerechneten Position  $\ln(LQ)$  mit dem Startwert  $a=a(\text{reg})$ ,  $d=d(\text{reg})$ ,  $\sigma = \sigma(\text{reg})$  berechnet und die so gewählte Standardabweichung  $\sigma$  während der Iteration konstant belassen.

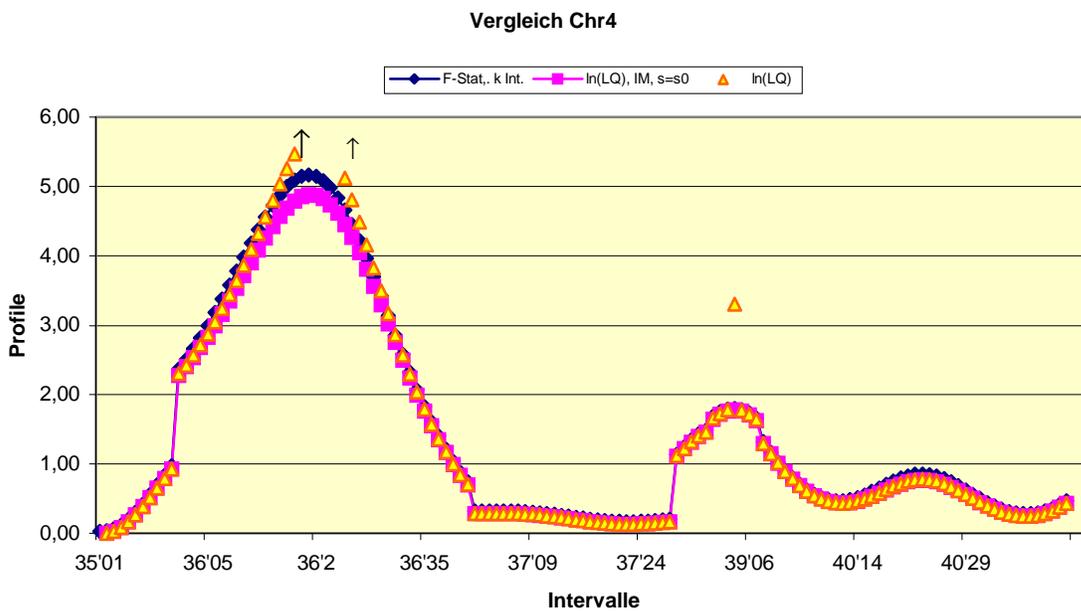


Abbildung 6.49: Wie Abbildung 6.46 für Chrom. 4.

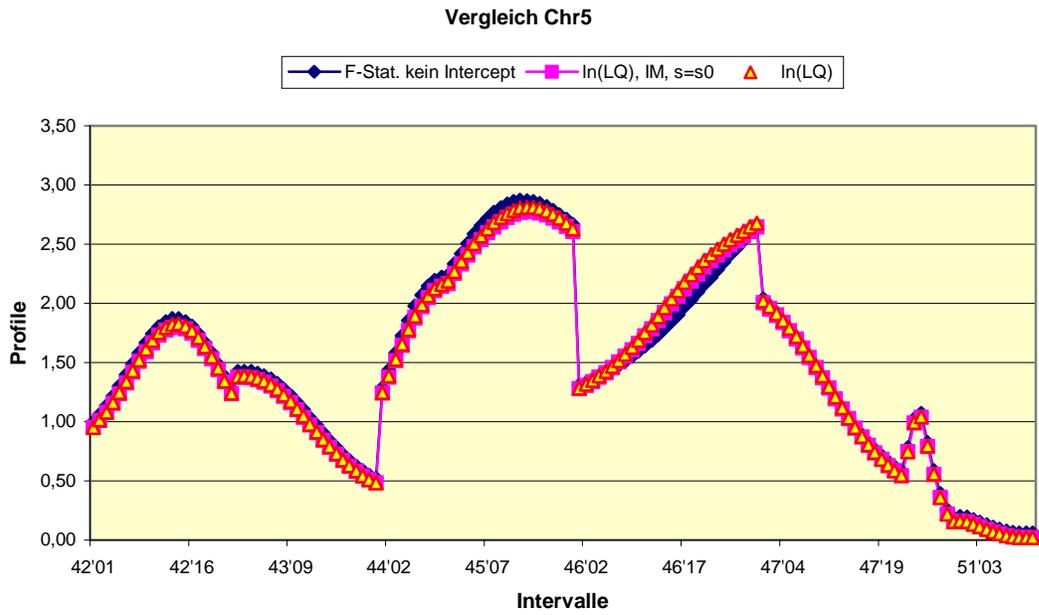


Abbildung 6.50a: Wie Abb. 6.46 für Chromosom 5.

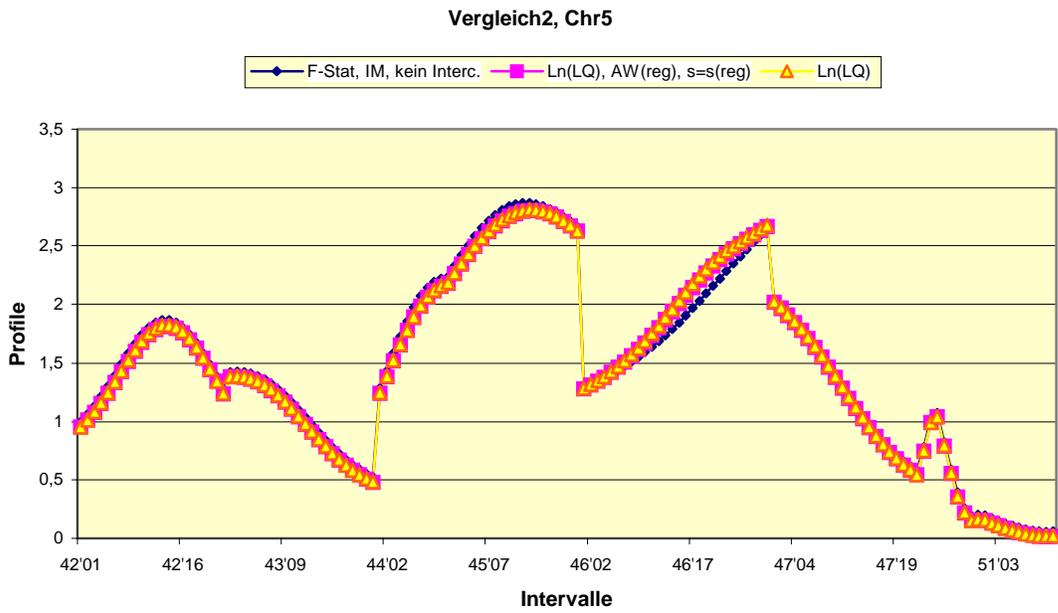


Abbildung 6.50b: Wie Abb. 6.50a, statt  $\sigma = \sigma_0$  wurde  $\sigma = \sigma(\text{reg})$  während der Iteration konstant belassen. Startwert:  $a=a(\text{reg}), d=d(\text{reg}), \sigma = \sigma(\text{reg})$  für alle Punkte.

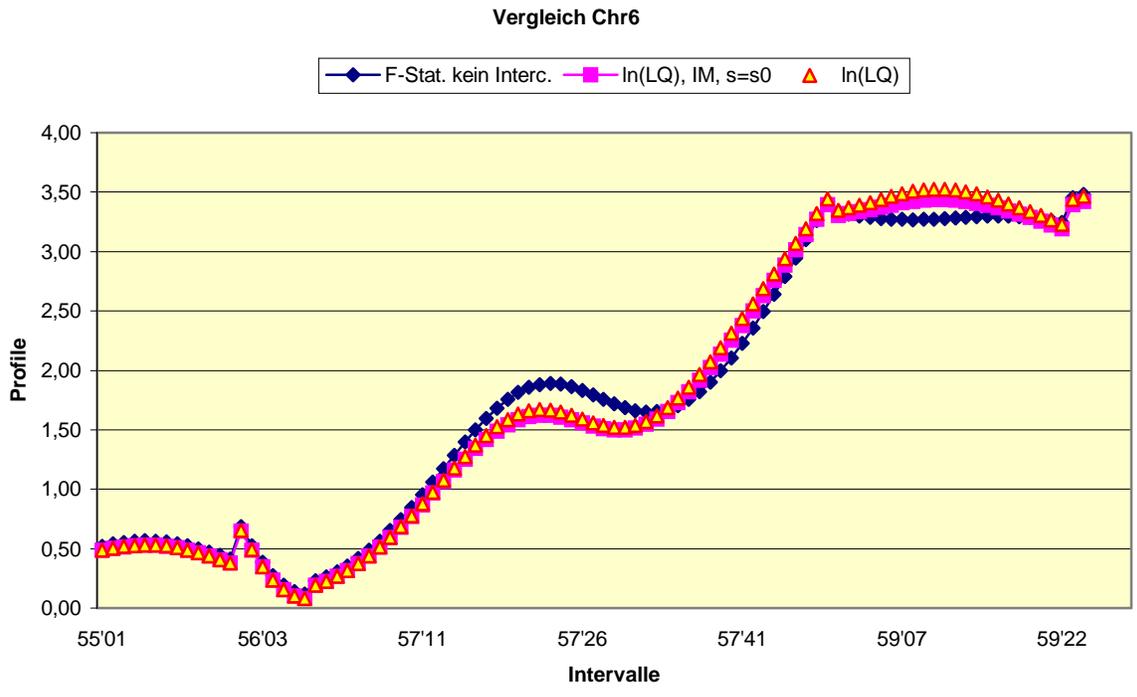


Abbildung 6.51a: Wie Abb. 6.50a, für Chrom. 6

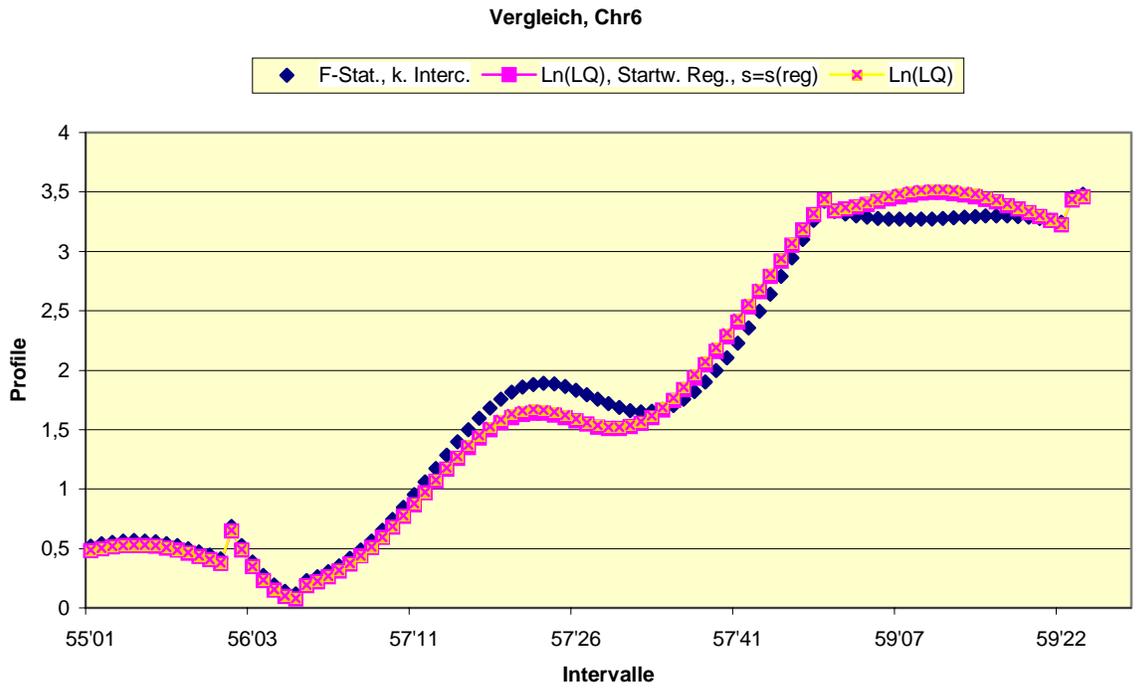


Abbildung 6.51b: Wie Abb. 6.50b für Chrom. 6.

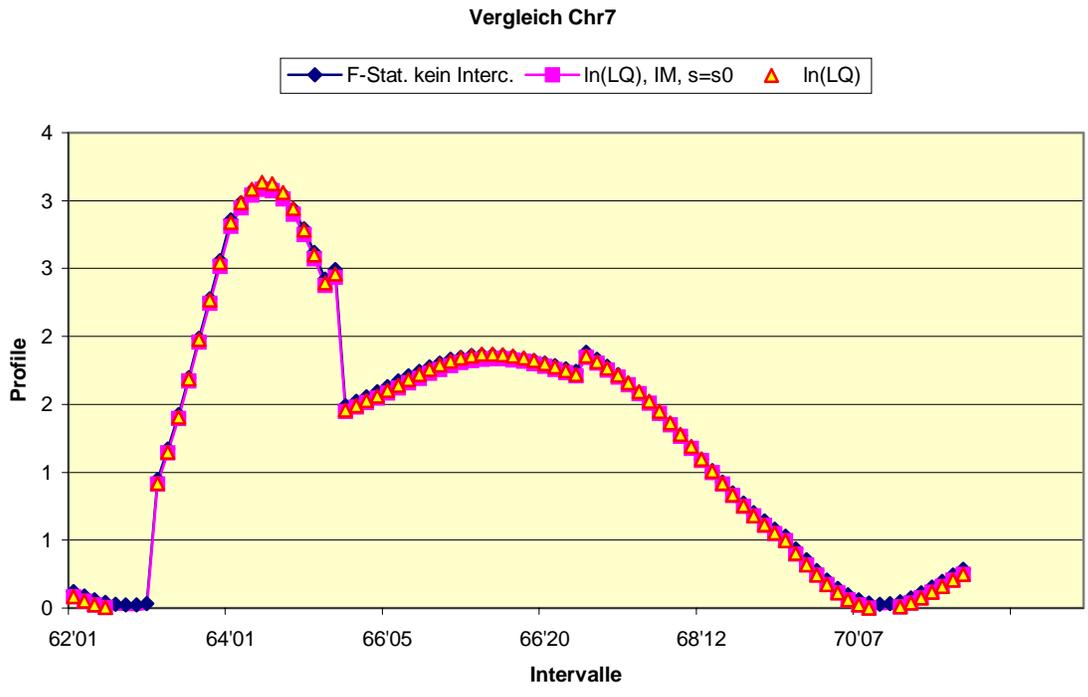


Abbildung 6.52a: Wie Abb. 6.50a, für Chrom. 7.

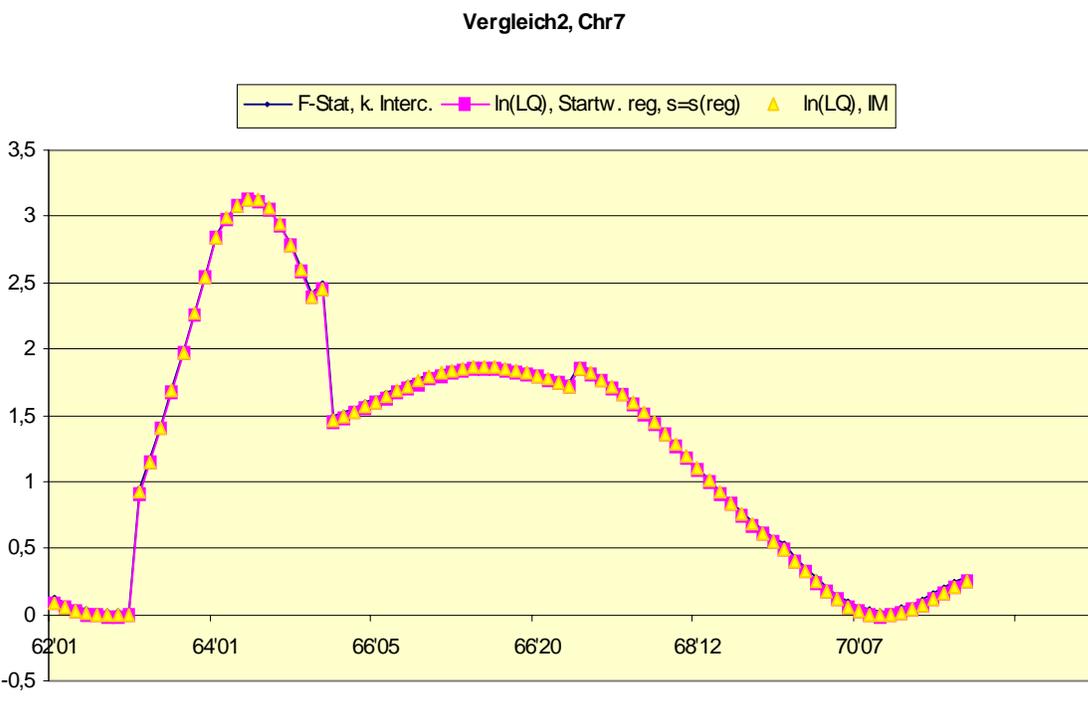


Abbildung 6.52b: Wie Abb. 6.50b für Chrom. 7.

Vergleich Chr8

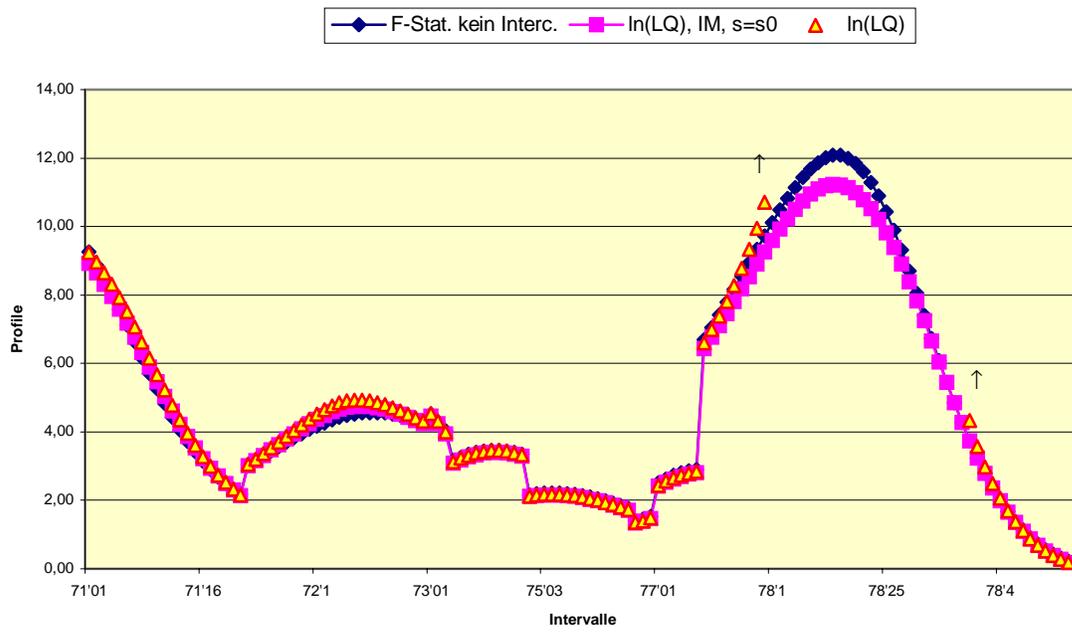


Abbildung 6.53: Wie Abb. 6.46 für Chrom. 8.

Vergleich Chr9

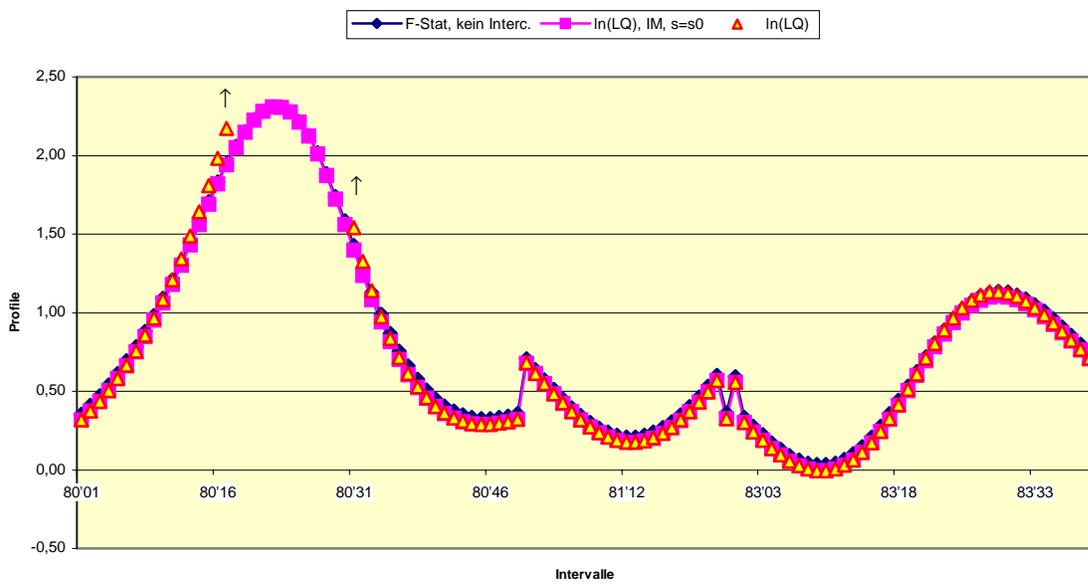


Abbildung 6.54: Wie Abb. 6. 46 für Chrom. 9

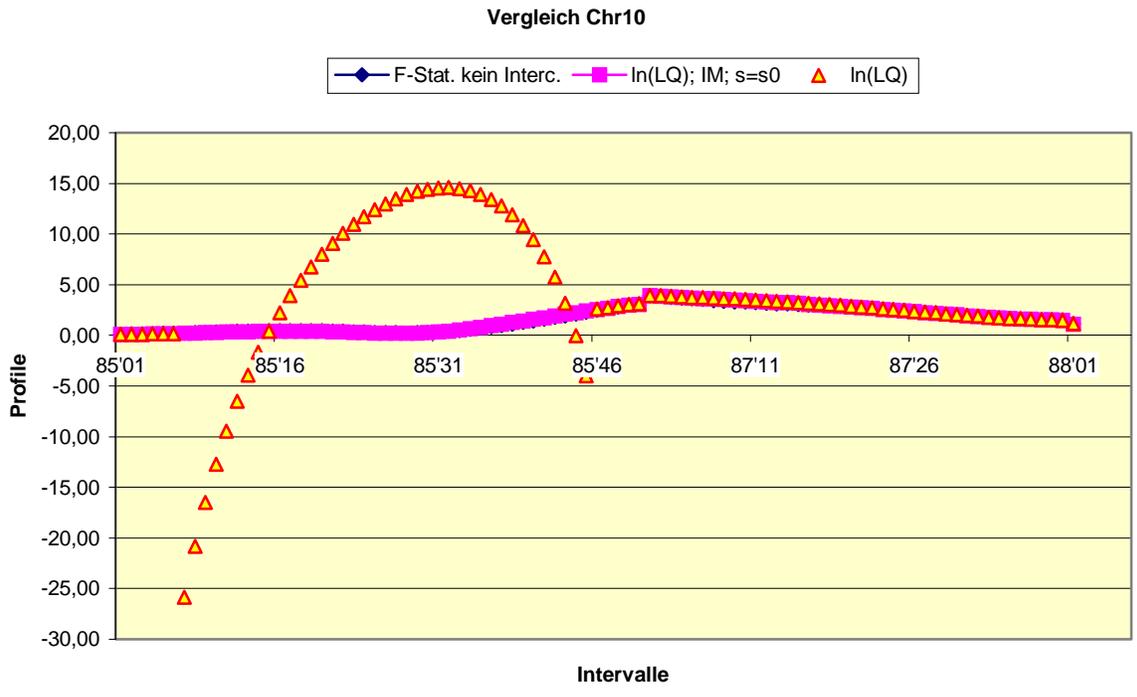


Abbildung 6.55a: Wie Abb. 6.46 für Chrom. 10

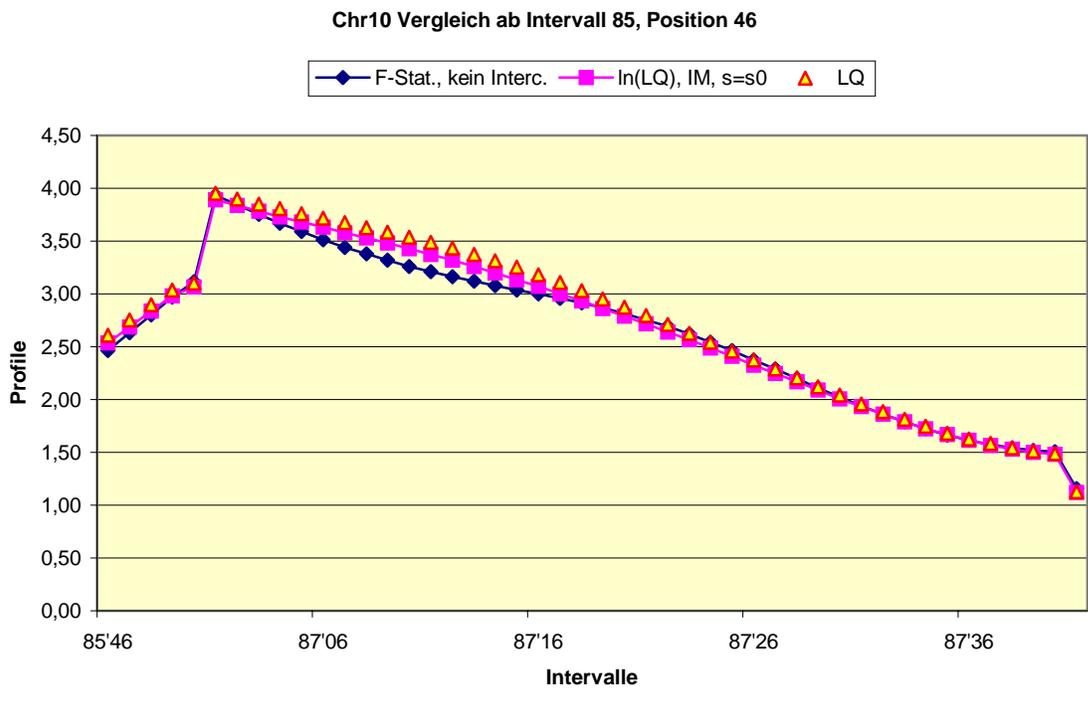


Abbildung 6.55b: Ausschnitt aus Abb. 6.55a.

Bei Chromosom 4 (Abb. 6.49) zeigt sich in Intervall 36 dasselbe Phänomen wie bei Chrom. 1. Die  $\ln(\text{LQ})$ -Lösung mit variiertem  $\sigma$  geht gegen die oszillierende Lösung (Tab. 15), wogegen die  $\ln(\text{LQ})$ -Lösung mit festgehaltenem  $\sigma = \sigma_0$  ein Maximum hat, welches einige Prozent niedriger als das Maximum der F-Statistik ist, aber praktisch an derselben Position auftaucht.

Bei den Chromosomen 5, 6 und 7 erkennt man nur relativ geringe Differenzen zwischen der F-Statistik und den EM-Lösungen mit variablem bzw. konstantem  $\sigma$  (Abb. 6.50a, b, 6.51 a, b und 6.52 a, b). es ist offenbar von marginaler Bedeutung, ob  $\sigma = \sigma_0$  gewählt wird, oder  $\sigma$  aus den Berechnungen der linearen Regression stammt, wie aus dem Vergleich der jeweiligen Abbildungen a und b ersichtlich wird.

Die Chromosome 8 und 9 (Abb. 6.53 und 6.54) zeigen wiederum gute Übereinstimmung zwischen der F-Statistik und den  $\ln(\text{LQ})$ -Werten für  $\sigma = \sigma_0$ , während die  $\ln(\text{LQ})$ -Werte mit nicht konstantem  $\sigma$  im Intervall 78 (Chromosom 8) bzw. Intervall 80 (Chrom. 9) nahezu sprunghaft auf die Werte der oszillierenden Lösung ansteigen, man vergleiche damit auch die Abbildungen 6.36 und 6.37 für Chrom. 8 bzw. 6.40 und 6.41 für Chrom. 9. Während in Abbildung 6.53 die Hauptmaxima für die F-Statistik bzw.  $\ln(\text{LQ})$  mit  $\sigma = \sigma_0$  im Intervall 78 etwas in der Höhe differieren, fallen in Abbildung 6.54 diese beiden Lösungen im Bereich des Hauptmaximums in Intervall 80 praktisch zusammen.

Chromosom 10 zeigt wiederum zwei tiefliegende Lösungen: eine, bei welcher die F-Statistik und die  $\ln(\text{LQ})$ -Werte mit konstantem bzw. variiertem  $\sigma$  sehr ähnlich sind, und in Intervall 85 eine Lösung mit großen  $\hat{a}$  und  $\hat{d}$  (Tabelle 6.16), nämlich die "zweite oszillierende Lösung". Zusammenfassend kann man sagen, dass die tiefliegende Lösung, dort wo sie existiert (d. h. überall, außer in einem Teil der Intervalle 2, 4, 8, 36, 78 und 80) der F-Statistik und der Lösung mit festem  $\sigma = \sigma_0$  sehr ähnlich ist, oder aber durch die zweite oszillierende Lösung repräsentiert wird (Abbildung 6.48a, Abbildung 6.55a). Die erstere Lösung nennen wir ab

jetzt die *Basislösung*. Die letztere, die zweite oszillierende Lösung erscheint nur in den Intervallen 24, 28 und 86.

Diese Einteilung ergibt sich auch aus dem Studium der Tabellen 6.12 bis 6.16. Die Positionen mit großen Schätzwerten der Parameter  $a$  und  $d$  (in den Tabellen:  $280 > \hat{a} > 248$ ,

$298 > \hat{d} > 257$ ) entsprechen der zweiten oszillierenden Lösung. Alle anderen Punkte der tiefliegenden Lösung entsprechen der Basislösung. Für diese gilt  $\hat{\sigma} \approx \sigma_0$ , während die Schätzwerte der Parameter  $a$  und  $d$  eine weite Streuung aufweisen.

#### **6.4 Die oszillierende Lösung im SIM-Modell**

Weit oberhalb der tiefliegenden Lösung erscheint eine völlig andere, in allen Intervallen wohldefinierte Lösung: die hochliegende oszillatorische Lösung, auch erste oszillierende Lösung oder schlechthin die oszillierende Lösung genannte Funktion. Sie ist dadurch gekennzeichnet, dass sie in jedem Intervall ein Maximum besitzt, welches etwa in der Intervallmitte liegt und Dominanzwerte  $\hat{d} \approx -500$  aufweist. Die Position und Höhe der Hauptmaxima sind für alle Chromosome in Tabelle 6.17 gegeben. Mit Ausnahme der Chromosome 5 und 7 liegen diese Maximalwerte zwischen -1341 und -1380 für  $\ln(L)$ , also ist  $\ln(LQ)$  von der Größenordnung 1000. Die Maxima von Tabelle 6.17 stellen mit großer Wahrscheinlichkeit die globalen Maxima der jeweiligen Chromosome im einfachen Intervallkartierungsmodell dar. Sie entsprechen also in diesem Modell den Orten mit der maximalen Likelihood für die entsprechenden Gene, wenn man Lösungen mit kleinem  $\hat{d}$ -Wert ( $\hat{d} \approx -500$ ) akzeptiert (vgl. die Diskussion in Kap 7).

**Tabelle 6.17: Die höchsten Maxima der oszillatorischen Lösung in SIM für alle Chromosomen**

CHROM.	1	2	3	4	5	6	7	8	9	10
Pos.	10'22	13'16	24'25	36'21	46'14	57'24	66'12	78'22	80'25	85'24
Ln(L)	-1341,63	-1357,34	-1357,11	-1379,58	-1510,43	-1379,32	-1584,45	-1345,51	-1359,55	-1374,82

### 6.5 Das zusammengesetzte Intervallkartierungsmodell (CIM-Modell)

Wie schon in der tabellarischen und graphischen Auswertung für die einzelnen Chromosome erwähnt, gibt es drei Lösungstypen: die oszillierende, die Kofaktoren - und die tiefliegende Lösung (man vgl. hierzu auch die Abbildungen in Kapitel 6.2, insbesondere Abbildung 6.7 von Chromosom 1)

**Tabelle 6.18: Die jeweils höchsten CIM-Maxima, der Größe nach geordnet, für Chrom. 1-10.**

ln(L)	-1210,6	-1255,3	-1260,7	-1273,7	-1278,7	-1291,2	-1295,0	-1313,9	-1375,3	-1496,2
Pos.	40'42	28'01	78'23	80'25	2'26	85'24	57'24	47'01	13'16	66'12
Chrom.	4	3	8	9	1	10	6	5	2	7
Typ d. Lösung	Kof	Kof.	Osz.	Osz.	Osz.	Osz.	Osz.	Kof.	Osz.	Osz.
$\hat{p}$	0,955	0,013	0,453	0,490	0,518	0,466	0,477	0,0155	0,491	0,496
$\exp(-\ln(L)/344)$	0,0296	0,0260	0,0256	0,0247	0,0243	0,0234	0,0232	0,0219	0,0195	0,0129

Die Tabellen A2 und A3 geben weitere Aufschlüsse über die Maxima der verschiedenen Lösungstypen. Die tiefliegende Lösung ist gekennzeichnet durch (vgl. Tabelle A1)  $\hat{\sigma}$ -Werte etwas unterhalb (ca. 10%) von  $\sigma_0 = 57,56$  und  $\hat{\beta}_i$  von einigen Dutzend (Tab A3). Die Determinanten der asymptotischen Varianz-Kovarianzmatrizen sind von der Größenordnung  $10^{15}$  bis  $10^{18}$ . Die tiefliegende Lösung ist der für ln(L) mit konstantem  $\sigma = \sigma_0$  ähnlich, taucht aber nicht in allen Intervallabschnitten auf (s. z. B. Abbildung 6.7).

Wir nennen sie im folgenden die *Basislösung*. Ein Analogon zu der zweiten oszillierenden Lösung des einfachen Intervallkartierungsmodells wurde nicht gefunden, auch nicht mit Startwerten  $\hat{a}=\hat{d}=250$  und großen Werten von  $\hat{\sigma}$ .

Die Maxima der oszillatorischen Lösung (Tabelle A2) sind gekennzeichnet durch

$$\hat{p} = 0,50 \pm 0,06$$

$$-479 \geq \hat{d} \geq -486$$

Alle 69 in Tabelle A2 (d. h. mit dem Startwert 4) aufgefundenen Maxima sind von diesem Typus. Obwohl die Determinanten von der Größenordnung  $10^{-11}$  sind (eine Ausnahme ist die Determinante mit  $27 \cdot 10^{-11}$  im Intervall 60), treten Eigenwerte von dieser Größenordnung niemals auf. Der kleinste gefundene Eigenwert ist  $6,83 \cdot 10^{-4}$  (Intervall 65). Das heißt, die Kovarianzmatrix ist immer eindeutig positiv definit und im Besonderen weit davon entfernt, singular zu sein. Für das geometrische Mittel  $M_g$  der zwölf Eigenwerte gilt bei allen 69 gefundenen Maxima

$$0,1595 \geq M_g \geq 0,1181.$$

Die oszillierende Lösung zeichnet sich auch dadurch aus, dass die Schätzwerte der 10 Parameter  $\sigma$ ,  $d$ , und  $\beta_i$  ( $i=1,2,\dots,8$ ) innerhalb eines Intervalls in sehr guter Näherung konstant sind, d. h., dass die Lösungskurve innerhalb eines Intervalls praktisch nur eine Funktion von  $p$  und  $a$  ist.

Von den 27 Maxima in der Tabelle A3 gehören 17 eindeutig dem Typus Basislösung an. Sieben weitere, nämlich die in den Intervallen 5, 8, 10, 28, 39, 40 und 47 gehören eindeutig zum Typus Kofaktorlösung (I5, I8, I10: Tab. 6.2 und Abb. 6.4, 6.5, 6.7a; I28: Tab 6.4 und Abb. 6.14; I39, I40: Tab. 6.5 und Abb. 6.18; I47: Tab. 6.6 und Abb. 6.22). Drei weitere, nämlich die in I26, I59 und I87 müssen weniger anhand des Loglikelihood-Profiles als durch eingehendes Studium der Tabelle A3 als Kofaktormaxima identifiziert werden, und zwar

durch Schätzwerte für Parameter  $a$  unterhalb  $-170$  und große  $\hat{\beta}_i$  beim  $i$ 'ten Kofaktor, z. B. entspricht das Maximum in I26 (ebenso wie das in I28) dem Kofaktor 3. Entsprechend ist  $\hat{\beta}_3$  groß; das Maximum in I59 entspricht dem Kofaktor 7 und hat ein entsprechend großes  $\hat{\beta}_7$ , während das Maximum in I87 mit  $\hat{\beta}_8$  mit dem Kofaktor 8 korrespondiert. Die Werte von  $K(\hat{\sigma}, \hat{\sigma})$  sind bei Kofaktoren-Maxima mindestens um einen Faktor 6 kleiner als bei den Maxima der Basislösung.

Insgesamt muss man sagen, dass das Studium der Tabelle A3 für die Unterscheidung der Kofaktormaxima von den Maxima der Basislösung wesentlich erhellender ist als das Studium des Profilverfahrens: Große Schätzwerte  $|\hat{a}|$  und  $\hat{\beta}_i$  für deren Parameter zeigen an, dass eine Kofaktorlösung vorliegt, die mit dem Kofaktor  $i$  korrespondiert. Auch eine Verwechslung mit den Maxima der Oszillatorlösung ist ausgeschlossen, da diese durch kleine Schätzwerte des Dominanzeffekts  $d$  gekennzeichnet sind.

Tabelle 6.18 zeigt die jeweils höchsten Maxima für jedes Chromosom unter Einschluss der Oszillator - und der Kofaktor-Lösung an, und zwar der Größe nach geordnet.

Ein Vergleich der Tabellen 6.17 und 6.18 zeigt, dass die Oszillatormaxima der Chromosome 2, 6, 7, 9 und 10 ihre Maxima in SIM und CIM innerhalb von 1 cM an derselben Stelle haben, das Maximum des Chromosoms 8 ist nur um 1 cM, das von Chrom. 1 jedoch um nahezu 200 cM verschoben (s. auch Abb. 6.1); die letztere Verschiebung rührt daher, dass in SIM und CIM vier nahezu gleichhohe Maxima für Chrom. 1 auftreten (man vgl. Abb. 6.9 mit Abb. 6.6), so dass der Einfluss der Kofaktoren die Positionen der 4 höchsten Maxima permutieren kann. Die Maxima der Chromosome 3, 4 und 5 werden in CIM von der Kofaktorenlösung geliefert.

Die Maxima der Tab. 6.18 repräsentieren die Maximum-Likelihood-Positionen der Gene im zusammengesetzten Intervallkartierungsmodell mit den hier gewählten Kofaktorpositionen. Die Maximum-Likelihood-Werte, dass ein QTL gefunden wird, sind nach dieser Tabelle im

Intervall 40'42 bei weiten am größten. Der Maximum-Likelihood-Wert an der Position 66'12 ist damit verglichen sehr klein.

Abbildung 6.1 zeigt die Aufteilung der verschieden langen Markerintervalle auf die 10 Chromosome. Es ist evident, dass bei den Chromosomen 2, 3, 5, 6, 7, 8, 9 und 10 das höchste Oszillatormaximum im längsten Intervall liegt, dagegen bei den Chromosomen 1 und 4 im zweitlängsten. Es gibt aber auch 14 Oszillatormaxima unterhalb von  $\ln(L) = -2000$  (Tab. A2) und zwar ausschließlich in Intervallen von 2 bis 4 cM und drei weitere zwischen  $L_0 = -1882,81$  und  $-2000$  in Intervallen zwischen 6 und 8 cM Breite. Das zeigt, ebenso wie beispielsweise die Abbildungen 6.15, 6.23 und 6.35, die starke Abhängigkeit der Höhe der Oszillatormaxima von der Intervallbreite. Auch die tiefliegenden Maxima der Oszillatorlösung sind durch stark negative Schätzwerte  $\hat{d}$  des Dominanzeffektes und kleine Schätzwerte  $|\hat{a}|$  gekennzeichnet, sie stehen nicht im Zusammenhang mit der ebenfalls tiefliegenden 2. oszillatorischen Lösung des SIM-Modelles, da diese durch große  $\hat{a}$  und  $\hat{d}$  eindeutig gekennzeichnet ist.

## **6.6 Die Basislösung und die F-Statistik für das CIM-Modell**

Nach der überwiegend sehr guten Übereinstimmung zwischen linearer Regression und der Basislösung der EM-Methode im SIM-Modell liegt es nahe, einen solchen Vergleich auch im CIM-Modell zu machen.

Da es genau 4 Chromosome gibt, in welchen Kofaktormarker nicht auftreten, nämlich die Chromosome 2, 7, 8 und 9, wird der Vergleich zunächst bei diesen vorgenommen. Der Vergleich von Abbildung 6.56 mit Abbildung 6.10 für Chromosom 2, Abbildung 6.57 mit Abbildung 6.30 für Chromosom 7, Abbildung 6.58 mit Abbildung 6.34 für Chromosom 8 und Abbildung 6.59 mit Abbildung 6.38 für Chromosom 9 zeigt eine bis ins Detail stimmende Formähnlichkeit, aber die Differenzen in den Ordinaten vergleichbarer Positionen sind in der F-Statistik um einen Faktor  $3,6 \pm 0,2$  größer als bei den  $\ln(L)$ -Werten. Diese Streckung und die

Verschiebung um ca. 590 (beachte, in SIM kann man ohne Streckung und Verschiebung die F-Werte direkt mit  $\ln(LQ)$  vergleichen, Abb. 6.46 - 6.55b) beruhen auf dem Einfluss der Kofaktoren. Die Formähnlichkeit besagt aber immerhin, dass die Maxima der Abbildungen 6.56 bis 6.59 praktisch an denselben Stellen liegen, wie die der  $\ln(L)$  - (oder  $\ln(LQ)$  - ) Funktionen.

Wir betrachten jetzt die Chromosome mit Kofaktoren. Zunächst Chromosom 6, bei welchem der Kofaktor ganz am rechten Rand auftaucht (vgl. Abb. 6.1). Abbildung 6.60a und 6.60b geben einen Vergleich ohne den rechten Rand, d. h. ohne Intervall 6.59 und 6.60. Wiederum ergibt sich eine starke Formähnlichkeit, und zwar mit einer Streckung der Ordinatenwerte der F-Statistik um knapp einen Faktor 4. Dieses Bild ändert sich, wenn das Intervall 6.59 mit einbezogen wird. Die Formähnlichkeit wird gering und ist noch geringer, wenn man etwa Chromosom 1 betrachtet, welches Kofaktoren in den Intervallen 5 und 8 aufweist; man vergleiche Abbildung 6.3 mit Abbildung 6.61. Ein noch wesentlich krasserer Beispielfall liefert Chromosom 10 (Abb. 6.42 und Abb. 6.62). Ähnliche Resultate wie bei Chromosom 10 erhält man für die Chromosome 3, 4 und 5.

Es ist somit vollständig klar, dass ein Vergleich der F-Statistik mit dem EM-Verfahren nur in den Bereichen sinnvoll ist, in welchen die Kofaktoren hinreichend weit entfernt sind.

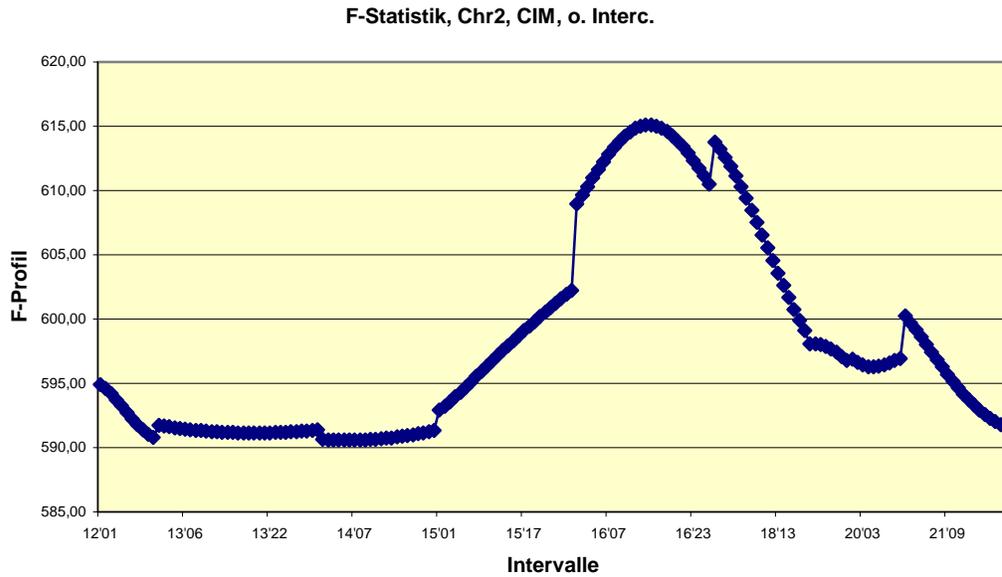


Abbildung 6.56: Die Ergebnisse der linearen Regression ohne Intercept für das CIM Modell bei Chrom. 2. Man vergleiche hierzu Abb. 6. 10 für das ln(L)-Profil dieses Chromosoms.

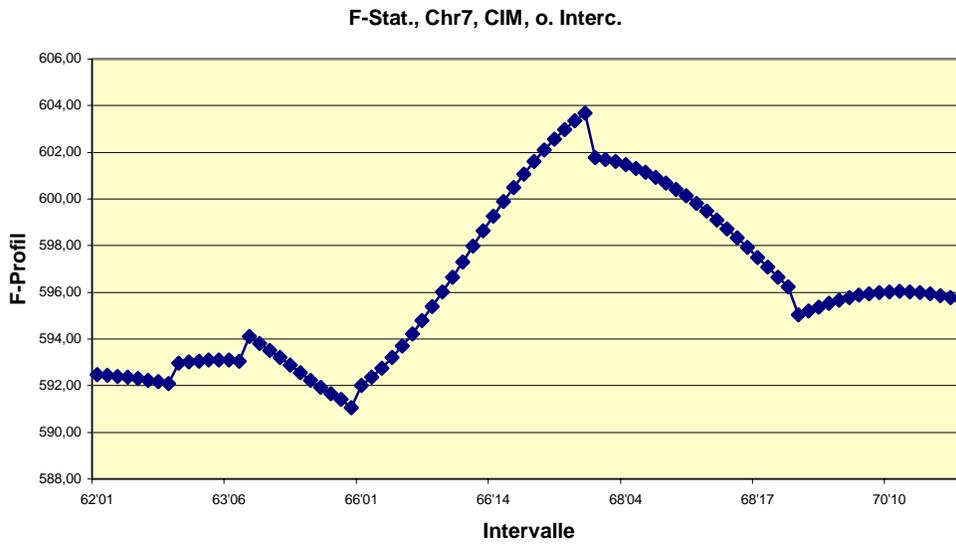


Abbildung 6.57: Wie Abb. 6.56 für Chrom. 7. Man vgl. hierzu die Abb. 6.30 für ln(L).

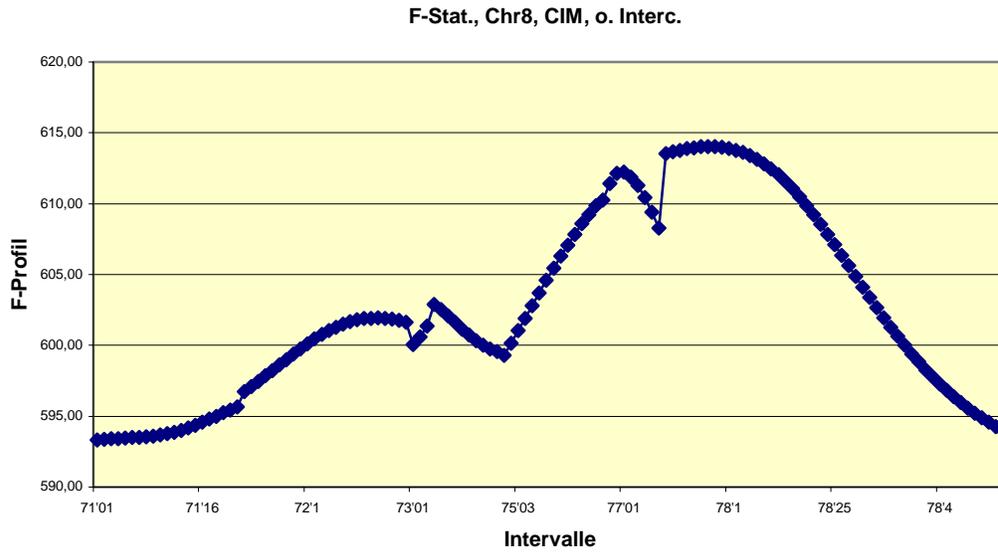


Abbildung 6.58: Wie Abb. 6.56 für Chrom. 8. Man vgl. hierzu die Abb. 6.34 für ln(L).

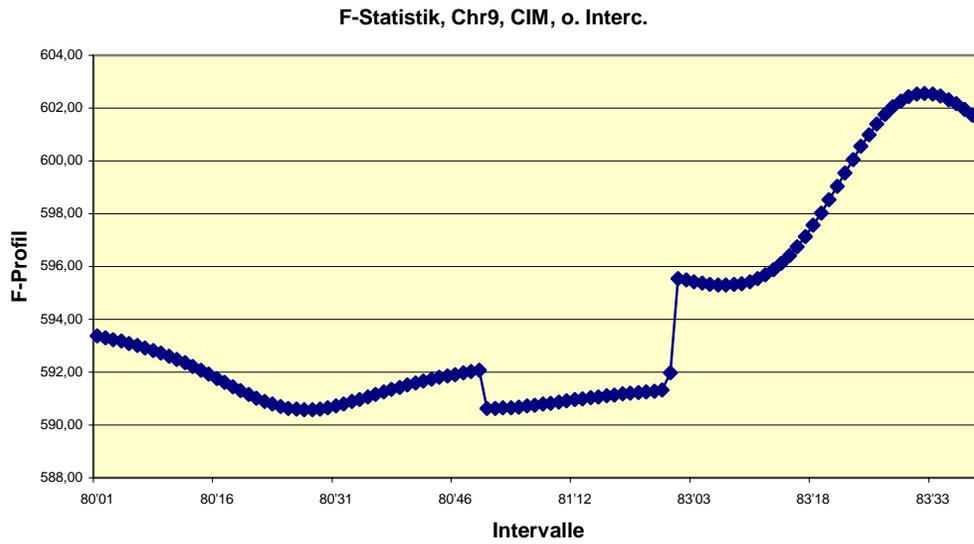


Abbildung 6.59: Wie Abb. 6.56 für Chrom. 9. Man vgl. hierzu die Abb. 6.38 für ln(L).

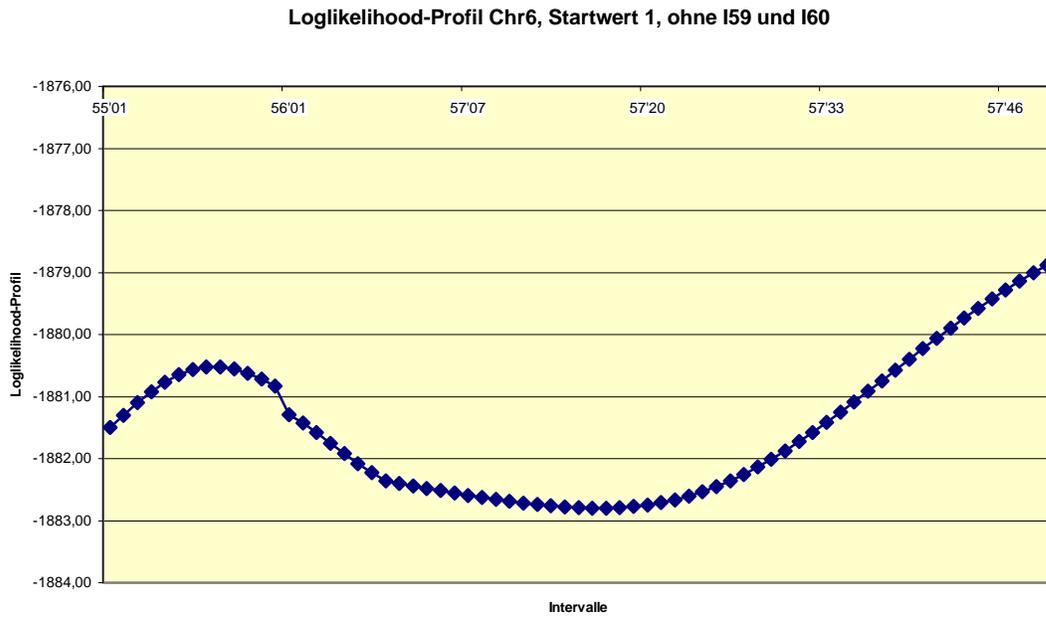


Abbildung 6.60a: Das Loglikelihood-Profil  $\ln(L)$  für Chrom. 6, ohne die Intervalle 59 und 60, vgl. hierzu auch die Abb. 6.26, welche die Resultate für Chrom. 6 in allen Intervallen wiedergibt.

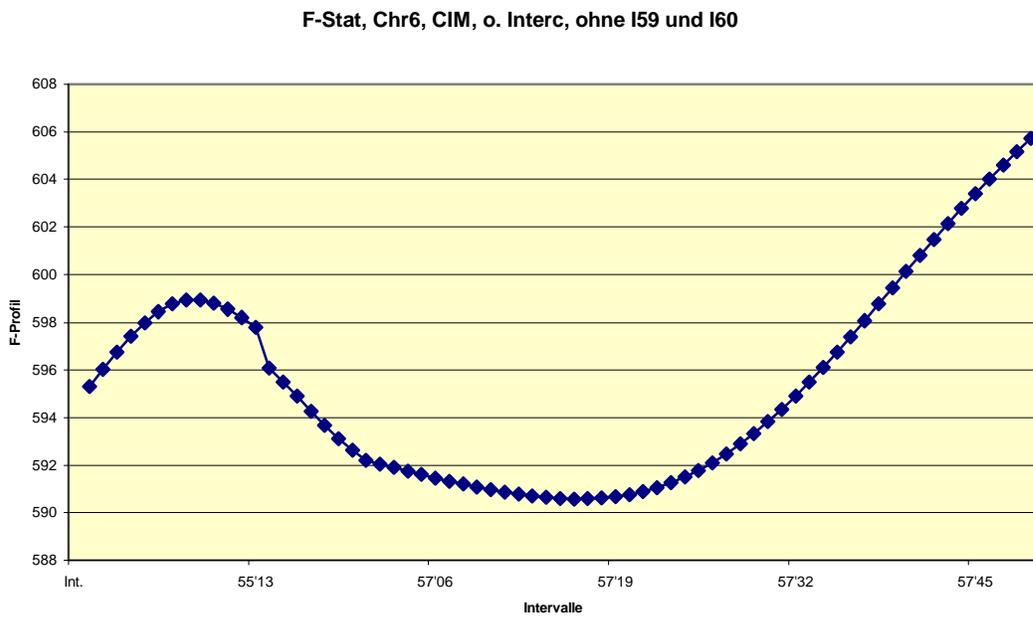


Abbildung 6.60b: Wie Abb. 6.60a, mit dem Unterschied, dass die Ordinatenwerte die Ergebnisse der linearen Regression statt  $\ln(L)$  wiedergeben.

F-Stat, Chr6, CIM, o. Interc.

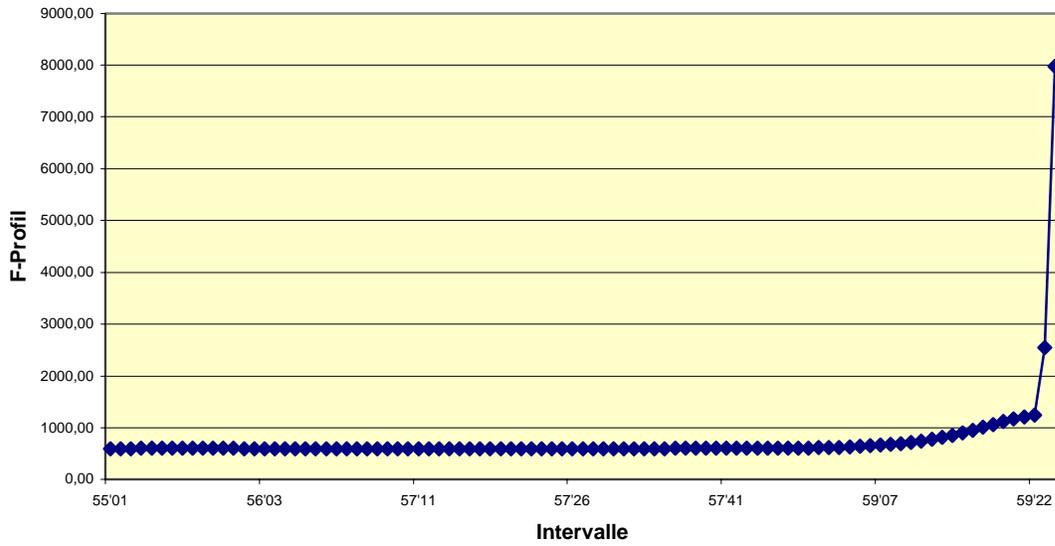


Abbildung 6.60c: Wie Abb. 6.56 für Chrom. 6, man vergleiche hierzu die  $\ln(L)$ -Resultate der Abbildung 6.26.

F-Stat., Chr1, CIM, ohne Intercept

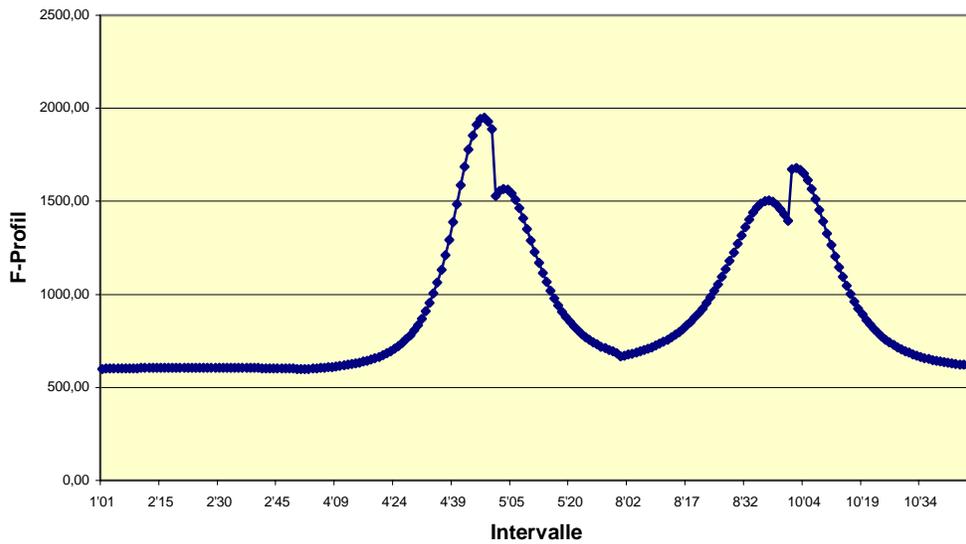


Abbildung 6.61: Wie Abb. 6.56 für Chrom. 1. Man vgl. hierzu die Abb. 6.3 für  $\ln(L)$ .

F-Statistik, Chr10, CIM, o. Interc.

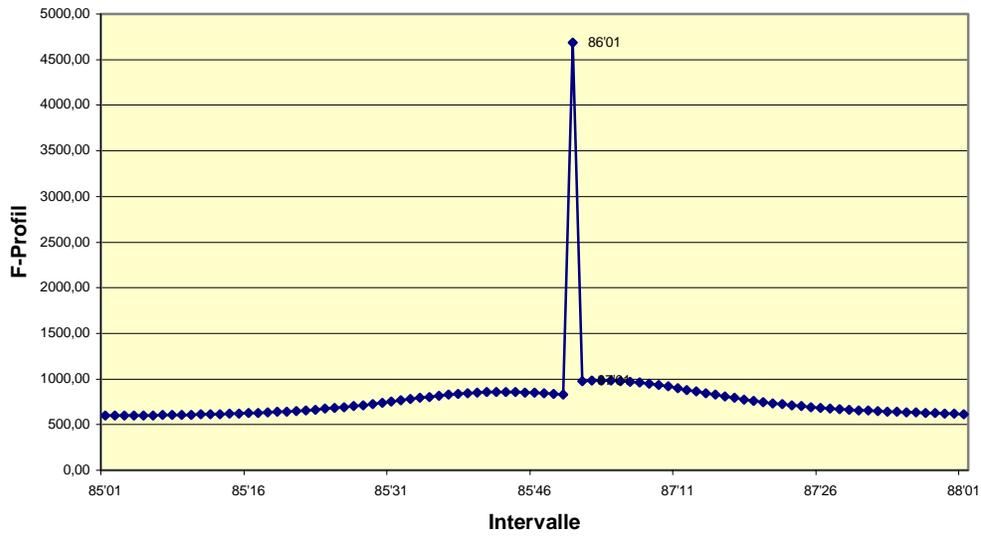


Abbildung 6.62: Wie Abb. 6.56 für Chrom. 10, man vgl. hierzu die Abb. 6.42 für ln(L)

## 7. Resultate

### 7.1 Diskussion der Ergebnisse

Die Gleichung

$$\frac{\partial Q}{\partial \theta_i} \Big|_{\theta_i = \hat{\theta}_i} = 0 \quad (7.1)$$

wobei  $\hat{\theta}_i \in \begin{cases} \{\hat{p}, \hat{a}, \hat{d}, \hat{\sigma}, \hat{\beta}_i, i=1,2,\dots,k\} & \text{für EMV} \\ \{\hat{a}, \hat{d}, \hat{\sigma}, \hat{\beta}_i, i=1,2,\dots,k\} & \text{für EM-Profile} \end{cases}$

führt zu dem Gleichungssystem:

$$A\hat{\theta} = b \quad (7.2)$$

wobei an den Konvergenzpunkten des Iterationsverfahrens

$$A = A[\hat{\theta}], \quad b = b[\hat{\theta}] \quad (7.3)$$

gilt.<sup>17</sup>

Das bedeutet, dass in den Konvergenzpunkten (und nur dort!) ein **nichtlineares** Gleichungssystem gelöst wird, obwohl man in jedem einzelnen Iterationsschritt nur ein lineares inhomogenes Gleichungssystem gelöst hat.

Nichtlineare Gleichungssysteme haben im Allgemeinen mehr als eine Lösung und genau dieses wurde auch hier beobachtet, und zwar in Profilkfunktionen (auch bei festen p!) sowohl in dem einfachen als auch in dem zusammengesetzten Intervallkartierungsmodell<sup>18</sup> (s. etwa Abb. (6.7)).

Wir diskutieren zunächst die Ergebnisse des CIM-Modells.

---

<sup>17</sup> Der Beweis folgt aus ( $A_i = A[\theta^{(i)}], b_i = b[\theta^{(i)}],$  s. Gl. (4.31)):  $A[\theta^{(i)}] \cdot \theta^{(i+1)} = b[\theta^{(i)}]$  für den Grenzwert  $\lim_{i \rightarrow \infty} \theta^{(i)} = \hat{\theta}.$

<sup>18</sup> Es gibt immer mindestens zwei und höchstens drei Lösungen bei festem p.

Man betrachte Abbildung 6.7 für Chromosom 1, welches von allen Chromosomen am grundlichsten studiert wurde. Man erkennt, dass an einigen Teilbereichen drei, an anderen wiederum zwei Lösungen gefunden wurden. Das liegt z. T. daran, dass die Kofaktorlösung in den Intervallen 1 und 2 nicht existiert. Diese Tatsache wurde mit über 30 000 Startwerten überprüft. Offenbar existieren die Kofaktorlösungen nur in Intervallen, in welchen Kofaktoren liegen oder in Intervallen, welche solchen benachbart sind.

Ferner geht die niedrigste Lösung, die "Basislösung" fast sprunghaft ("sprungähnlich") in die Kofaktorlösung über, das gilt insbesondere, wenn man den Startwert 1 verwendet (vgl. Abb. 6.3 auch mit Abb. 6.4 und Abb. 6.5). Obwohl das für die Position der gesuchten Gene nicht relevant ist, kann man sich aus allgemeinen mathematischen Gründen fragen, ob in Abbildung 6.3 das Nichterscheinen der Basislösung zwischen den Positionen 4'24 und 5'21 bzw. 8'21. und 10'27 daran liegt, dass diese Lösung dort nicht existiert, oder doch existiert, aber deswegen nicht gefunden wurde, weil die zu dieser Lösung gehörenden Konvergenzgebiete der EM-Methode sehr klein sind im Vergleich zu den Konvergenzgebieten der Kofaktor- und der Oszillatorlösung (vgl. auch Abb. 6.7). Es dürfte sehr schwer sein, diese Frage im Rahmen der EM-Methode zu lösen (es wurde versucht!). Wenn jedoch eine Lösung existiert, die ähnliche Parameterschätzer  $\hat{a}, \hat{d}, \hat{\sigma}, \hat{\beta}_i, i=1,2,\dots,k$  aufweist, wie benachbarte Punkte der Basislösung, kann man versuchen, die Lösung schrittweise mit Hilfe des Newton-Raphson-Verfahrens zu konstruieren, weil dieses fast immer hervorragende Konvergenzeigenschaften aufweist, wenn man einen Startwert in der Nähe des Konvergenzpunktes hat.

Ich komme nun zu dem wichtigsten Punkt des CIM-Modells. Die von der Basislösung sehr verschiedene Kofaktorenlösung tritt nur in den Kofaktorintervallen und den zu diesen benachbarten Intervallen auf. Das heißt aber, sie hängt in sehr starker Weise von der Auswahl der Kofaktoren ab! Man kann diese Auswahl per Intuition treffen, dies ist jedoch in diesem Falle nur ein anderes Wort für Willkür. Wenn man die Auswahl anhand eines Modells z. B. SIM und eines dazugehörigen Lösungsverfahrens, z. B. EM oder Linearer Regression trifft,

dann stellt das ebenfalls eine Auswahl dar, die ohne Willkür gar nicht getroffen werden kann; das gilt umso mehr, wenn in dem betreffenden Vorauswahl-Modell die Kofaktorlösung überhaupt nicht auftritt (SIM!) oder aber die der *Basislösung* ähnliche Regression des CIM als Grundlage genommen wird.

Es gibt zwei Wege, um aus diesem Dilemma heraus zu kommen: Entweder man erklärt alle Marker zu Kofaktormarkern oder man wählt die triviale Alternative und verzichtet auf Kofaktoren, d. h. bleibt bei dem SIM-Modell.

Im ersteren Falle wählt das Modell anhand des Daten-Inputs selbst aus, welche  $\hat{\beta}_i$  groß und welche klein sind. Eigentlich macht man dann nur eine Verallgemeinerung des SIM-Ansatzes nach dem Schema

$$\begin{array}{ll} \mu_1 = a - \frac{d}{2} & \mu_{j1} = a - \frac{d}{2} + X_j\beta \\ \mu_2 = \frac{d}{2} & \mu_{j2} = \frac{d}{2} + X_j\beta \\ \mu_3 = -a - \frac{d}{2} & \mu_{j3} = -a - \frac{d}{2} + X_j\beta \end{array} \quad \rightarrow \quad (7.4)$$

für alle Marker  $j=1,2,\dots,89$ , d. h. die Anzahl der Kofaktormarker ist identisch mit der Zahl der Marker. Bei einer solchen Prozedur ist auch zu erwarten, dass  $L_0$  erheblich größer wird und dass die  $\ln(LQ)$ -Werte für die Kofaktorlösung und für die oszillatorische Lösung weit weniger groß werden, als sie in der vorliegenden Arbeit mit  $k=8$  sind. Ob dann die Basislösung (mit  $\ln(L) \approx \ln(L_0)$ ) als dritte Lösung des EM-Verfahrens noch auftaucht, muss abgewartet werden.

Bevor die andere "hochliegende" Lösung diskutiert wird, möchte ich noch eine Vorbemerkung machen. Aus den Naturwissenschaften, insbesondere der Physik, kennt man den Begriff der "spuriösen Lösung" (spurious solution), die insbesondere bei der Lösung nichtlinearer Gleichungssysteme als mathematisch einwandfreie Lösung zwar auftreten, aber

keine Entsprechung in der Natur haben, also vom naturwissenschaftlichen Standpunkt aus bedeutungslos sind.

Die in der vorliegenden Arbeit gefundenen beiden besonders interessanten Lösungen für das CIM-Modell, nämlich die Kofaktorlösung und die oszillierende Lösung könnten im Prinzip spuriose Lösungen sein und sind auf diesen Aspekt hin zu untersuchen.

Die Kofaktorenlösung weist als Besonderheit sehr kleine Werte des Additivitätseffektes  $a$  auf, nämlich  $-171 > \hat{a} > -241$ , ferner für den  $i$ 'ten Kofaktor  $\beta_i$  Schätzwerte der Größenordnung 200 ( $i=1,2,\dots,8$ ). Wenn man, wie in der vorliegenden Arbeit, weniger als 10 % der Marker zu Kofaktormarker macht, weist die Kofaktorlösung ausgeprägte Randsuprema auf und Maxima in der Nähe des Randes - mit Ausnahme des Maximums in I26 mit<sup>19</sup>  $\hat{p}=0,797$ . Es ist zu erwarten, dass sich dieses Verhalten in einem beliebigen Intervall ändert, wenn jeder Marker zu einem Kofaktormarker gemacht wird.

Trotz des relativ kleinen (beitragsmäßig großen) Additivitätseffektes habe ich keine Hinweise gefunden, die auf eine Spuriousität der Kofaktorlösung hinweisen.

Die Oszillatorlösung weist beitragsmäßig relativ kleine Schätzwerte der Parameter  $a$  und  $\beta_i$  auf, jedoch extrem kleine Werte für den Dominanzeffekt, nämlich  $\hat{d} \approx -483$ .

Es bleibt abzuwarten, ob die Genetiker solch kleine Werte für biologisch sinnvoll halten.

Alle Punkte des Loglikelihood-Profiles, welche zu einem Intervall gehören (z. B. den Intervallen 2, 4, 5, 8 und 10 von Abb. 6.7) weisen mit sehr hoher Genauigkeit dieselben Werte für  $\hat{a}$ ,  $\hat{d}$ ,  $\hat{\sigma}$ ,  $\hat{\beta}_i$ , ( $i=1,2,\dots,8$ ) auf, d. h. die Loglikelihood-Funktion ist innerhalb eines

---

<sup>19</sup> Die in der Tabelle A3 gegebenen Kofaktor-Maxima sind wahrscheinlich nicht vollständig, da der Startwert 1 nur Kofaktorlösungen in der Nähe der Randsuprema liefert, was für viele andere Startwerte nicht gilt (vgl. Abb. 6.3 mit den Abbildungen 6.4 und 6.5). Mit großer Wahrscheinlichkeit liefert der Startwert 1 aber die höchsten Kofaktormaxima.

Intervalles praktisch nur eine Funktion von  $p$ , und zwar eine sehr einfache, welche ihr Maximum bei ungefähr  $\hat{p}=0,5$  hat, und wobei dieses Maximum in breiten Intervallen sehr hoch und in schmalen sehr niedrig ist.

Dazu kann man jetzt einen "Gedankenversuch" machen. Angenommen, man würde etwa in der Mitte eines recht breiten Intervalles einen neuen Marker finden, dann würde offenbar das Oszillatormaximum in der Mitte des alten Intervalles verschwinden, aber es würden 2 neue, wesentlich niedrigere Maxima in der Mitte der neuen Intervalle auftauchen. Anstelle des alten Maximums, das z. B. das höchste für ein bestimmtes Chromosom sein möge, tritt (treten) aber dann mit ziemlicher Wahrscheinlichkeit ein neues (oder aber zwei neue) Kofaktor-Maximum (-Maxima) auf, welches (welche) durchaus ganz in der Nähe des alten liegen kann (können), weil man ja verlangen muss (s. oben), dass jeder Marker auch Kofaktormarker ist! Daraus folgt, dass es im CIM-Modell z. Z. kein zwingendes Argument gibt, die etwas schematisch aussehende Oszillatorlösung als eine spuriose Lösung zu betrachten, es sei denn die Genetiker hätten Einwände gegen einen stark negativen Dominanzeffekt.

Eine andere Betrachtung ergibt sich dann, wenn durch die Wahl von 89 Kofaktoren bei jedem Chromosom das Kofaktormaximum das höchste ist. Die Frage, ob die oszillatorische Lösung spurios ist oder nicht, wird dann für die Frage der Genpositionen irrelevant sein.

Wir diskutieren jetzt das SIM-Modell.

Das gerade erwähnte "Gedankenexperiment" mit einem, etwa durch verbesserte Experimentiertechnik, zusätzlich hinzukommenden Markergen führt jetzt zu erheblichen Konsequenzen, da es ja Kofaktormaxima, welche annähernd an die Position des alten Maximums treten könnten, nicht gibt. Die Konsequenz hieße dann, dass man, wie hoch die Markerdichte auch sein möge, durch ein zusätzliches Markergen die lokale Position und Höhe der Maxima ganz drastisch geändert würde oder aber, dass - viel realistischer - die oszillierende Lösung im SIM-Modell **spurios** ist.

Das bedeutet, dass die nicht-spuriösen, also statistisch relevanten Daten durch die Abbildungen 6.46 bis 6.55 wieder gegeben werden. Dieses besagt wiederum, dass im SIM-Modell lineare Regression und EM-Profilverfahren fast überall in guter bis hervorragender Übereinstimmung sind. Insbesondere weist die lineare Regression dort, wo die EM-Lösung aus mathematischen, d. h. wohl im Iterationsverfahren begründeten, Ursachen gegen die spuriöse oszillatorische Lösung geht, (Intervall 2, 4, 10 bei Chrom. 1, Intervall 36 bei Chrom. 4, Intervall 78 bei Chrom. 8 und Intervall 80 bei Chrom. 9) ausgeprägte Maxima auf, welche die EM-Basislösung auch an diesen Stellen in sehr vernünftiger Weise interpoliert.

Neue Fragen stellen sich aber bei der Betrachtung der Ergebnisse der Chromosome 3 und 10, Abbildung 6.48a und 6.55a. Die hier auftretende "2. oszillatorische Lösung" ist gekennzeichnet durch große Schätzwerte für  $a$  und  $d$  ( $\hat{a} \approx 250$ ,  $\hat{d} \approx 270$ ), s. Tabellen 6.14 und 6.16.

Die Maxima liegen nicht in der Intervallmitte und sind in den Intervallen 28 und 85 eindrucksvoll hoch. Es gibt keine Hinweise darauf, dass diese mit den Regressionsdaten nicht in Einklang stehenden Resultate einer spuriösen Lösung entsprechen.

Äußerst wichtig bei all diesen Betrachtungen ist, dass bei den EM-Rechnungen sowohl zu dem SIM- als auch zu dem CIM-Modell die Informationsmatrix eindeutig nicht-singulär ist. Daher sind ihre Eigenvektoren linear unabhängig und spannen den ganzen Raum der Variablen  $p, a, d, \sigma$  und  $\beta_i$ , (hier:  $i=1,2,\dots,8$ ) auf.<sup>20</sup>

Die Überprüfung der Nichtsingularität erfolgt am besten durch die Eigenwerte. Der kleinste, in Tabelle A3 bei einer Kofaktorlösung auftretende Eigenwert ist  $4,28 \cdot 10^{-5}$ . Der kleinste in Tabelle A2 bei einer Oszillatorlösung gefundene Eigenwert ist  $6,38 \cdot 10^{-4}$ . Diese Eigenwerte sind hinreichend groß, um Probleme, die in der Nähe einer Singularität auftreten,

---

<sup>20</sup> Dies gilt in jedem Iterationsschritt und somit auch im Konvergenzpunkt.

auszuschließen: alle Lösungen sind vollkommen stabil und entsprechen einem vollständigen Satz linear unabhängiger Eigenvektoren.

Ein Vergleich unserer Ergebnisse mit denen von Kao und Zeng (1997), und QTL-Cartographer gestaltet sich im Detail schwierig, weil im Programm QTL-Cartographer (Basten et al 1994) die SIM- und CIM-Modelle mit Intercept verwendet wurden, während ich die Rechnungen ohne Intercept durchgeführt habe. Bei den SIM-Ergebnissen ist ein indirekter Vergleich via F-Statistiken (via linearer Regression) möglich. Wie oben diskutiert und in Abbildungen 6.46 bis 6.55 dargestellt, ist die Übereinstimmung der Ergebnisse von EM-Verfahren und F-Statistik ohne Intercept eindrucksvoll, wenn auch mit Ausnahme einiger Intervalle. Ein Vergleich der linearen Regression *mit* Intercept mit den EM-Rechnungen der Programme Mapmaker und QTL-Cartographer ergab eine einigermaßen befriedigende Übereinstimmung, wenn auch die von mir errechnete F-Statistik etwas detailreicher ist, weil ich etwas mehr Rechenpunkte als die Standardversionen der oben genannten Programme verwendet habe.

Ein Vergleich der Details der EM(CIM)-Rechnungen mit und ohne Intercept ist zwar nicht möglich, jedoch wurde von Basten et. al. (1999) nichts über die Existenz von Lösungen, welche der Kofaktor- oder der Oszillatorlösung entsprechen, berichtet. Sie erwähnen lediglich, dass in bestimmten "Fenstern", welche um ihre Kofaktoren (5 in der Standardversion) herum liegen, eben diese Kofaktoren vernachlässigt werden; d. h. genauer gesagt, in einem Fenster um den  $i$ 'ten Kofaktor wird  $\beta_i = 0$  gesetzt (oder vernachlässigt).

Dieses Vorgehen könnten wir auch auf die hier vorliegenden Ergebnisse anwenden. Man betrachte z. B. Abbildung 6.3. Das Vorgehen würde hier bedeuten, dass man bei allen Punkten im "Fenster" zwischen 4'24 und 5'21 den Kofaktor 1 vernachlässigt, die anderen 7 Kofaktoren aber beibehält, während man etwa im Fenster zwischen 8'21 und 10'27 den Kofaktor 2 vernachlässigt, den Kofaktor 1 und die Kofaktoren 3 bis 8 aber beibehält. In ähnlicher Weise würde man dann auch bei den anderen Chromosomen vorgehen, und würde

bei den Chromosomen 7, 8 und 9 keine Änderung der Basislösung erhalten, während sie bei den Chromosomen 2, 3, 4, 5, 6 und 10 wieder zusammengestückt werden müsste. Auf diese Weise erhielten wir eine "zusammengestückte Funktion", die außerhalb der Fenster, wo ein  $\beta_i = 0$  gesetzt wird, mit der Basisfunktion übereinstimmte und innerhalb der Fenster die Basisfunktion interpolierte. Selbstverständlich würde auch diese zusammengestückte Funktion erheblich davon abhängen, welche Marker als Kofaktormarker gewählt werden.

Ein solches Verfahren muss für diese Arbeit strikt abgelehnt werden, und zwar aus folgenden Gründen:

i) nichtlineare Gleichungssysteme haben i. A. mehrere Lösungen. Dass wie im vorliegenden Fall "neuartige Lösungen" auftauchen ist interessant und entspricht der hier gestellten Problemstellung (s. Einleitung), nach Lösungen zu suchen, die sich von der Regression (F-Statistik) unterscheiden.

ii) ein Großteil der Fremdartigkeit der Kofaktorlösung stammt wahrscheinlich aus der Tatsache, dass nur ein kleiner Teil der Marker als Kofaktormarker gewählt wird. Das wäre aber nur dann vertretbar, wenn man annehmen könnte, dass bei allen anderen Markern die Schätzwerte der  $\beta_i$  klein wären, wenn man das System vervollständigte, d. h. alle Marker zu Kofaktormarkern machte. Dies muss sich jedoch erst post festum, d. h. anhand der Rechnung, herausstellen, man kann es nicht a priori annehmen. Wenn man jedoch alle Marker zu Kofaktormarkern macht, dann wird höchstwahrscheinlich  $\ln(L_0)$  sehr viel stärker angehoben werden als  $\ln(L)$ , d. h. die  $\ln(LQ)$ -Werte der Kofaktorlösung sind dann wahrscheinlich erheblich kleiner. Ebenso werden dann die "Sprünge" zwischen Basis- und Kofaktorlösung zumindest kleiner, wenn man nicht sogar eine "durchlaufende" Kofaktorlösung erhält.

Zurück zu Basten et. al. (1999). Diese haben sich über die Motivation ihres Vorgehens nicht weiter ausgelassen. Es soll daher keinesfalls unterstellt werden, dass diese Autoren wesentlich eine mathematisch wohldefinierte Kofaktorlösung unterdrückt haben. Man muss beachten, dass an den Sprungstellen von der Basis- zur Kofaktorlösung die Determinante der

Informationsmatrix typischer Weise um einen Faktor  $10^{10}$  bis  $10^{25}$  abfällt. Man kann daher, z. B. wenn man nicht die Eigenwerte betrachtet, durchaus zu dem Schluss kommen, dass die Informationsmatrix irgendwo im Fenster um den Kofaktor singular wird<sup>21</sup>. Da in einem singulären Punkt die Lösung ihren mathematisch wohldefinierten Sinn verliert, und dies i. A. auch in einer Umgebung des singulären Punktes gilt, könnte dies durchaus die Ursache des Vorgehens von Kao und Zeng sein. Auch die geringere Stabilität des ECM-Algorithmus (Kap. 6.1) oder andere, mir unbekannt, Gründe könnten für ihr Vorgehen eine Rolle gespielt haben, zumal bei Modellen mit Intercept kompliziertere Verhältnisse vorliegen können. Bei den hier vorliegenden numerisch wohldefinierten Kofaktorlösungen ist ein solches Vorgehen jedenfalls solange nicht gegeben, als nicht gezeigt worden ist, dass diese Lösung spuriös ist.

## 7.2 Zusammenfassung und Schlussfolgerung

Es wurden für das CIM-Modell bei der Lösung der Gleichung  $\frac{\partial Q}{\partial \theta_i} \Big|_{\theta_i = \hat{\theta}_i} = 0$  nach dem EM-Verfahren zwei neue Lösungstypen gefunden, die oszillatorische Lösung und die Kofaktorlösung, bei welchen sich (zumindest) bei der letzteren keine Anzeichen für eine Spuriösität finden lassen.

Die Kofaktorlösung ist von der Auswahl der Kofaktoren **extrem** stark abhängig. Es bleibt wohl nur die (wohl unzulängliche) Möglichkeit, zu dem SIM-Modell zurückzukehren (s. auch unten), oder aber alle Marker zu Kofaktormarkern zu machen. Es bleibt dann dem CIM-Modell und dem EM-Verfahren überlassen, zu entscheiden, wo und welche Schätzwerte von

---

<sup>21</sup> Da in Profilverfahren, wie sie in den genannten Intervallkartierungsansätzen ausschließlich verwendet werden, das Maximum nicht exakt berechnet wird, können durchaus zusätzliche Probleme mit der Definitheit der Kovarianzmatrix auftreten.

$\beta_i$  klein sind - aber im Nachhinein und nicht a priori! Nach Durchführung einer solchen Rechnung wird die prognostische Kraft des so eingeschränkten CIM-Modelles entscheiden, ob dieses Modell wertvoll ist, oder nicht.

Im SIM-Modell stimmen die Ergebnisse des EM-Verfahrens und der linearen Regression gut überein mit der Ausnahme von 3 von 89 Intervallen, nämlich I24, I28 in Chrom. 3 und I85 in Chrom. 10. Diese Ausnahmen sind allerdings bemerkenswert, weil es sich in I28 (Abb. 6.48a) und I85 (Abb. 6.55a) um eindrucksvolle zusätzliche Maxima handelt.

EMV-Verfahren (incl. ausführlicher Tabellen für Parameterschätzer, Eigenwerte etc.) und  $\ln(L)$ -Profilverfahren liefern im Wesentlichen dieselben Erkenntnisse. Die visuelle Darstellung bringt Vorteile bei der Klassifizierung verschiedener Lösungen. Sind diese Lösungen im Prinzip bekannt, ist das numerisch einfachere EMV-Verfahren durchaus hinreichend. Außerdem liefert es i. A. mehr Maxima als das Profilverfahren mit beschränktem Diskretisierungsgrad von beispielsweise 1 cM. Es wird daher oft genügen, das (z. B. bei 89 Kofaktormarkern) numerisch aufwändige Profilverfahren nur bei einem Chromosom auszuführen, und bei den übrigen Chromosomen die Berechnung (samt Klassifizierung der Lösungen an Hand von Tabellen der Schätzer und Varianzen) mit dem numerisch weniger aufwändigen und genaueren EMV-Verfahren durchzuführen.

Insgesamt stellt das EM-Verfahren eine faszinierende Möglichkeit dar, verschiedene Lösungen von solchen Modellen zu finden, die zu nichtlinearen Gleichungssystemen der Art

$$A[\hat{\theta}] \cdot \hat{\theta} = b[\hat{\theta}] \quad (7.5)$$

an den Konvergenzpunkten führen. Genauer gesagt, kann man auf diese Weise Konvergenzpunkte (fast in allen Fällen Maxima der Loglikelihood) aufspüren, die i. A. zu verschiedenen Funktionszweigen gehören. Profilverfahren können eine wertvolle visuelle Hilfe sein, um die verschiedenen Zweige zu identifizieren.

## 8. Appendix

### 8.1 Appendix A: numerische Tabellen

-Tab. A1: Häufig verwendete Startwerte der SIM- und CIM-Modelle. Ferner sind die Loglikelihood  $\ln(L_0)$  und die zugehörigen Parameterschätzer für SIM- und CIM gegeben.

In den Tabellen A2 und A3 lies:  $\hat{a}$ ,  $\hat{d}$ ,  $\hat{\sigma}$ ,  $\hat{p}$  für  $a$ ,  $d$ ,  $p$ ,  $\sigma$  und  $\hat{\beta}_i$  für  $b_i$ .

-Tab. A2 [(A2).1 bis (A2).6]: Alle Maxima mit Startwert 4. Es handelt sich um 69 Maxima der Oszillatorlösung. Sie sind gekennzeichnet durch:  $486,3 > -\hat{d} > 479,7$ ;  $4,82 > \hat{\sigma} > 4,54$  und  $\hat{p} = 0,50 \pm 0,06$ . Ferner sind  $\hat{\beta}_3$  und  $\hat{\beta}_8$  negativ, alle anderen  $\hat{\beta}_i$  positiv. Der kleinste Eigenwert stimmt in sehr guter Näherung mit  $K(\hat{p}, \hat{p})$  überein.

-Tab. A3 [(A3).1 bis (A3).3]: Alle Maxima mit Startwert 1. Die 10 Maxima in den Intervallen 5, 8, 10, 26, 28, 39, 40, 47, 59 und 87 sind Kofaktormaxima. Sie sind dadurch gekennzeichnet, dass die Schätzwerte von  $-a$  und  $\beta_i$  ( $i$  entspricht dem  $i$ -ten Kofaktor) etwa 200 sind (genauer:  $207 \pm 36$ ), ferner dadurch, dass der kleinste Eigenwert der Kovarianzmatrix in sehr guter Näherung mit  $K(\hat{p}, \hat{p})$  übereinstimmt. Die restlichen 17 Maxima sind vom Basistyp. Sie besitzen eine asymptotischen Varianzschätzer  $\hat{\sigma}$  etwas unterhalb  $\sigma_0 = 57,65$  (genauer:  $\sigma_0 > \hat{\sigma} > 52,0$ ),  $K(\hat{\sigma}, \hat{\sigma})$ -Werte\* oberhalb von 48 000, und ausschließlich positive  $\hat{\beta}_1$ .

---

\* Der größte Eigenwert ähnelt sehr stark  $K(\hat{\sigma}, \hat{\sigma})$  und liegt daher ebenfalls oberhalb von 48 000.

### Appendix A.1 Startwerte und Funktionswerte unter $H_0$ :

#### Startwerte

IM	a	d	$\sigma$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
stabiler Startwert	1	2	10								
oszill. Startw. (OL).	0,42	-498,21	6,15								
CIM	a	d	$\sigma$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$	$\beta_7$	$\beta_8$
Startwert 1	1	2	50	3	4	5	6	7	8	9	10
Startwert 2	-222,10	-9,24	227,99	7,95	1,70	1,13	2,83	3,14	3,55	-1,26	17,12
Startwert 3	-234,37	3,19	4,21	238,64	-1,18	0,19	1,17	2,89	2,95	0,29	10,15
Startwert 4	-333	554	445	543	345	533	444	533	633	317	200

#### Schätz- und Funktionswerte für der Loglikelihood $\ln(L_0)$ unter $H_0$ (IM und CIM)

$\beta_1^0$	$\beta_2^0$	$\beta_3^0$	$\beta_4^0$	$\beta_5^0$	$\beta_6^0$	$\beta_7^0$	$\beta_8^0$	$\sigma_0(\text{CIM})$	$\ln(L_0(\text{CIM}))$	$\sigma_0(\text{IM})$	$\ln(L_0(\text{IM}))$
31.14545	33.1861	22.49632	17.30996	26.13461	31.38462	36.47074	29.80809	57.64801	-1882.813	249.1587	-2386.3375

Kofaktorenmarker der Auswertung: 5, 10, 28 39, 41,47,61 und 86

## Appendix Tabelle (A2).1

	I1	I2	I3	I4	I5	I7	I8	I10	I12	I13	I14	I15
a	0,559	0,680	0,573	0,706	1,302	1,334	1,810	0,418	-1,090	-0,926	-0,786	-0,433
d	-482,156	-482,387	-482,122	-483,143	-484,751	-484,107	-486,227	-482,572	-482,183	-481,926	-481,820	-481,538
b1	2,335	2,254	2,218	1,757	1,179	1,776	1,949	2,343	2,330	2,317	2,366	2,364
b2	3,625	3,624	3,671	3,661	3,404	3,161	2,006	3,294	3,593	3,644	3,705	3,723
b3	-1,608	-1,583	-1,573	-1,570	-1,682	-1,705	-1,691	-1,597	-1,609	-1,593	-1,626	-1,584
b4	0,811	0,803	0,805	0,822	0,816	0,833	0,824	0,793	0,839	0,760	0,808	0,777
b5	1,695	1,641	1,680	1,699	1,704	1,694	1,722	1,702	1,719	1,725	1,710	1,726
b6	1,188	1,175	1,170	1,162	1,257	1,227	1,134	1,200	1,208	1,279	1,220	1,208
b7	1,310	1,334	1,335	1,292	1,259	1,229	1,256	1,311	1,314	1,349	1,350	1,369
b8	-1,487	-1,473	-1,470	-1,498	-1,463	-1,411	-1,466	-1,460	-1,574	-1,550	-1,509	-1,469
σ	4,788	4,773	4,786	4,784	4,713	4,664	4,630	4,809	4,696	4,731	4,756	4,800
p	0,513	0,518	0,486	0,496	0,497	0,507	0,530	0,484	0,489	0,491	0,536	0,501
Iter.	44,00	33,00	43,00	54,00	57,00	49,00	51,00	69,00	32,00	31,00	33,00	29,00
	-2406,09	-1278,19	-2144,40	-1282,33	-1382,46	-2447,68	-1299,97	-1282,92	-1696,14	-1357,30	-1446,40	-1444,40

### Varianzen

K(pp)	0,000746	0,000836	0,000808	0,000933	0,000785	0,000720	0,000790	0,000855	0,000773	0,000906	0,000882	0,000758
K(aa)	0,146035	0,119022	0,145053	0,284291	0,287900	0,169300	0,256617	0,340522	0,111584	0,108817	0,120697	0,114382
K(dd)	4,775186	4,782422	4,761505	5,724814	6,142396	5,170784	6,177104	5,922871	4,617867	4,651227	4,681107	4,770409
K(σσ)	3,148927	3,106331	3,145744	3,222690	3,304548	3,198750	3,205441	3,163331	3,045998	3,056668	3,105455	3,120879
K(b1b1)	0,154395	0,156069	0,161263	0,355099	0,382675	0,181952	0,163547	0,155042	0,150573	0,152232	0,152949	0,154648
K(b2b2)	0,139227	0,137926	0,138000	0,138612	0,147503	0,161635	0,163243	0,148698	0,135893	0,136125	0,137098	0,139915
K(b3b3)	0,133871	0,133060	0,133744	0,133702	0,133970	0,132587	0,130691	0,134388	0,130653	0,132076	0,132966	0,134052
K(b4b4)	0,172535	0,171535	0,172290	0,173955	0,170704	0,168804	0,167493	0,173183	0,168510	0,169922	0,171106	0,172794
K(b5b5)	0,171190	0,171559	0,171595	0,171726	0,168545	0,166770	0,166300	0,173334	0,166359	0,167860	0,169012	0,170665
K(b6b6)	0,158505	0,157902	0,158707	0,159138	0,157968	0,156239	0,155084	0,159139	0,154625	0,157078	0,157294	0,158773
K(b7b7)	0,147008	0,146351	0,147005	0,147246	0,146455	0,145066	0,144132	0,147632	0,143772	0,145132	0,145977	0,149099
K(b8b8)	0,157455	0,156412	0,157256	0,157826	0,156174	0,153663	0,153663	0,158348	0,154534	0,155828	0,156278	0,157572
Eigen=	3,149695	4,915355	3,146861	3,185199	3,184883	3,142924	3,071536	6,237437	4,744065	4,7777184	3,107877	3,121029
werte	4,907252	3,105883	4,893479	6,023924	6,589513	5,408133	6,632767	3,150869	3,048968	3,058839	4,807904	4,901770
	0,000746	0,000934	0,007231	0,431412	0,412495	0,000720	0,363270	0,518778	0,000773	0,000905	0,000881	0,007248
	0,007233	0,007194	0,000808	0,000929	0,007152	0,237531	0,007043	0,000844	0,007082	0,007132	0,007173	0,000758
	0,243071	0,240549	0,240908	0,007234	0,000784	0,007088	0,000788	0,237465	0,236037	0,100199	0,239565	0,241564
	0,115870	0,102258	0,183891	0,235532	0,231268	0,223938	0,231891	0,007258	0,102178	0,237842	0,171358	0,104846
	0,173033	0,172432	0,167634	0,076847	0,069430	0,166826	0,059328	0,068253	0,168782	0,172408	0,159685	0,173114
	0,159667	0,120076	0,110101	0,117332	0,157908	0,081611	0,167690	0,173353	0,114840	0,116040	0,112891	0,118913
	0,125200	0,134052	0,119207	0,137557	0,165354	0,154131	0,114363	0,119272	0,156566	0,157869	0,116254	0,128635
	0,133700	0,155213	0,158758	0,158333	0,115798	0,129902	0,141805	0,144098	0,127632	0,127426	0,136008	0,161138
	0,146616	0,127103	0,136973	0,127883	0,128554	0,140962	0,130500	0,134156	0,143644	0,144073	0,145203	0,136128
	0,142998	0,158476	0,127118	0,167848	0,136483	0,114321	0,153102	0,155562	0,130575	0,133955	0,126022	0,148767
Min(EW)	0,000746	0,000934	0,000808	0,000929	0,000784	0,000720	0,000788	0,000844	0,000773	0,000905	0,000881	0,000758
Det	2,28E-11	2,19E-11	2,43E-11	5,53E-11	4,14E-11	2,03E-11	3,07E-11	6,30E-11	1,39E-11	1,75E-11	2,02E-11	1,84E-11

## Appendix Tabelle (A2).2

	116	118	119	120	121	123	124	125	126	127	128	131
a	-0,744	-1,2393	-1,4058	-1,0725	-1,0675	-1,0378	-0,9342	-0,0048	0,3365	0,5814	-1,4482	-1,5964
d	-480,938	-480,8752	-480,9824	-481,4669	-481,2361	-481,8144	-481,0741	-481,7933	-482,3568	-482,9052	-480,0047	-482,0549
b1	2,321	2,2834	2,2889	2,3479	2,3684	2,3655	2,3469	2,3165	2,3331	2,3370	2,3400	2,3655
b2	3,825	3,8394	3,8054	3,7056	3,6795	3,6859	3,6370	3,6714	3,6896	3,6809	3,6805	3,6976
b3	-1,511	-1,4942	-1,4922	-1,5611	-1,5349	-1,4593	-1,0414	-1,4593	-1,8955	-2,1674	-0,4290	-1,4027
b4	0,777	0,8155	0,8731	0,8856	0,8922	0,8029	0,7978	0,7911	0,7801	0,7896	0,7534	0,7751
b5	1,732	1,6519	1,6960	1,6653	1,6881	1,7782	1,7434	1,7335	1,7399	1,7345	1,7240	1,7367
b6	1,281	1,3580	1,3191	1,2803	1,2972	1,1904	1,2167	1,1973	1,2056	1,2032	1,0624	0,9630
b7	1,410	1,3784	1,3397	1,3603	1,3358	1,2873	1,2506	1,3123	1,3243	1,3181	1,3973	1,4930
b8	-1,408	-1,3847	-1,4631	-1,5290	-1,4605	-1,5902	-1,5573	-1,4739	-1,4761	-1,4790	-1,6002	-1,6539
σ	4,764	4,6616	4,6127	4,6998	4,7010	4,7082	4,7452	4,8195	4,8128	4,8024	4,6769	4,5580
p	0,489	0,5050	0,5044	0,4981	0,5359	0,4514	0,5222	0,5196	0,5278	0,5188	0,4673	0,5165
Iter.	33,00	28,0000	34,0000	34,0000	29,0000	39,0000	63,0000	90,0000	93,0000	163,0000	55,0000	33,0000
	-1472,11	-1574,5519	-1825,9461	-1741,9631	-1472,2993	-1666,7849	-1275,4062	-2276,1555	-1819,7686	-2148,0618	-1288,2593	-1917,8846
Varianzen												
K(pp)	0,000726	0,00070692	0,00075522	0,00075493	0,00081246	0,00082270	0,00130048	0,00083208	0,00079385	0,00081623	0,00086289	0,00073425
K(aa)	0,119326	0,09929113	0,10558172	0,11160841	0,10255047	0,12770601	0,23045383	0,34778995	0,47842187	0,91485149	0,23686862	0,10196238
K(dd)	4,845401	4,63184366	4,59266111	4,63671960	4,64206182	4,64306647	4,85245612	5,36552227	6,01180557	7,97018590	4,92386641	4,48921253
K(σσ)	3,097870	2,94584186	2,96163324	3,04833473	3,00789764	3,11653725	3,29927809	3,13681757	3,17224256	3,34316567	3,31045803	2,87892893
K(b1b1)	0,153476	0,15003922	0,14876303	0,15089948	0,150665729	0,15164863	0,15378168	0,15591927	0,15615871	0,15559282	0,15176402	0,14641391
K(b2b2)	0,142461	0,13515777	0,13334081	0,13510177	0,13471830	0,13547604	0,13794053	0,14176760	0,14055531	0,13934007	0,13549487	0,13185062
K(b3b3)	0,134336	0,13029591	0,12931550	0,13089467	0,13109156	0,13367968	0,21016956	0,33119160	0,52095033	1,02298521	0,27649810	0,12824925
K(b4b4)	0,171250	0,16704319	0,16617003	0,16964563	0,16928302	0,16909117	0,17159453	0,17366392	0,17390644	0,17353503	0,16974774	0,16353960
K(b5b5)	0,169234	0,16611654	0,16387020	0,16751587	0,16680485	0,16762800	0,16979833	0,17152883	0,17169204	0,17155823	0,16793156	0,16173118
K(b6b6)	0,159070	0,15550103	0,15214123	0,15525773	0,15505687	0,15530487	0,15752557	0,15955095	0,15974539	0,15954256	0,15803758	0,15168662
K(b7b7)	0,147801	0,14237157	0,14070494	0,14364801	0,14329339	0,14448328	0,14654348	0,14915734	0,14848190	0,14787272	0,14464810	0,14036011
K(b8b8)	0,157238	0,15287498	0,15071882	0,15365480	0,15327643	0,15606355	0,15791387	0,15860520	0,15819775	0,15818146	0,15550073	0,15112929
Eigen=	4,986259	4,76141675	4,72234807	3,05110730	4,76906435	3,12217204	3,29367649	3,13681627	6,38752661	3,21786027	5,13738049	4,61135947
werte	3,095858	2,94618980	2,96225471	4,76313456	3,00920041	4,76889093	5,02082988	5,57537416	3,15950253	9,13981904	3,29039823	2,88380971
	0,000725	0,00070675	0,00075519	0,00075486	0,00081224	0,00082262	0,32968148	0,52855516	0,70134447	0,95932078	0,36593815	0,00073425
	0,007185	0,23383177	0,00695291	0,23791221	0,23719597	0,00711336	0,00129897	0,00083205	0,00079385	0,00081586	0,00085434	0,22892978
	0,240588	0,00701684	0,23205658	0,00707667	0,00705923	0,23712418	0,00721049	0,00272794	0,00272715	0,23989861	0,23253713	0,00686797
	0,101744	0,08929296	0,16939618	0,10459700	0,17019946	0,16072116	0,23556511	0,23961742	0,06786727	0,00726893	0,00710318	0,08342455
	0,174713	0,17302765	0,15260640	0,17141162	0,09755961	0,17400325	0,16545545	0,07668275	0,24105374	0,06505022	0,06518742	0,16626426
	0,158997	0,15434884	0,11205824	0,13071990	0,15510226	0,10769119	0,08374128	0,16885195	0,16833860	0,16833860	0,16614196	0,15797061
	0,153619	0,11302135	0,14029867	0,14241273	0,11401638	0,14639646	0,14155890	0,12513562	0,12438339	0,16020656	0,12196854	0,11730354
	0,116343	0,14498840	0,12277188	0,12423665	0,13094715	0,13545866	0,12344263	0,14401257	0,14455952	0,12447235	0,15564318	0,13861070
	0,129075	0,12939815	0,09602320	0,11459530	0,14235151	0,12494626	0,15818933	0,15986407	0,16020335	0,14501992	0,14188126	0,122235518
	0,133084	0,12386252	0,12813392	0,15607684	0,12399553	0,11607752	0,12810605	0,12955295	0,12959354	0,12975628	0,12674476	0,12816884
Min(EW)	0,000725	0,000707	0,000755	0,000765	0,000812	0,000823	0,001299	0,000832	0,000794	0,000816	0,000854	0,000734
Det	1,68E-11	1,02E-11	1,05E-11	1,37E-11	1,32E-11	1,79E-11	5,90E-11	6,47E-11	8,43E-11	1,65E-10	3,40E-11	8,58E-12

## Appendix Tabelle (A2).3

	I32	I33	I35	I36	I37	I38	I39	I40	I42	I43	I44	I45
a	-1,5585	-1,1384	0,49844	0,48085	0,56143	-0,81570	0,88486	1,09104	0,45359	0,17467	-0,03278	0,34108
d	-482,0825	-482,2210	-481,86795	-482,02107	-482,34140	-480,29605	-483,61865	-484,82348	-482,08667	-481,93642	-481,77087	-482,28255
b1	2,4237	2,4249	2,33708	2,36528	2,38805	2,32245	2,35301	2,31167	2,33967	2,35284	2,34784	2,33353
b2	3,6816	3,6879	3,64766	3,65714	3,64988	3,69838	3,72823	3,71708	3,64832	3,66201	3,67223	3,67150
b3	-1,4882	-1,6185	-1,61524	-1,60786	-1,59853	-1,60375	-1,61251	-1,63761	-1,57928	-1,58860	-1,59456	-1,59505
b4	0,7796	0,7454	0,74462	0,61070	0,33372	1,58589	-0,06907	0,22819	0,73946	0,78095	0,79160	0,79364
b5	1,7821	1,7673	1,78224	1,78456	1,78709	1,73758	1,65106	0,87213	1,77042	1,74101	1,73131	1,76139
b6	0,9419	1,0388	1,24841	1,24693	1,25242	1,19711	1,17081	1,16266	1,12787	1,15982	1,20801	1,01499
b7	1,4756	1,4469	1,33120	1,34237	1,38382	1,30416	1,32858	1,31292	1,24834	1,28785	1,31470	1,29071
b8	-1,6239	-1,5899	-1,50750	-1,49219	-1,47273	-1,48591	-1,47112	-1,50949	-1,43868	-1,47484	-1,47366	-1,49205
σ	4,5698	4,6883	4,79518	4,79727	4,79669	4,78012	4,77480	4,75909	4,79877	4,81647	4,81942	4,80914
ρ	0,5136	0,5096	0,50552	0,49129	0,49168	0,51037	0,50787	0,50385	0,50973	0,48228	0,50692	0,48674
Iter.	31,0000	26,0000	31,00000	30,00000	42,00000	106,00000	101,00000	61,00000	28,00000	31,00000	36,00000	32,00000
	-2117,3370	-1453,9573	-1666,54420	-1294,60253	-1397,23620	-1994,30467	-1948,99581	-1312,12720	-1437,56673	-1454,11493	-1702,12755	-1492,10830

Varianzen

K(pp)	0,0007	0,0008	0,00082474	0,00097644	0,00089746	0,00074361	0,00069417	0,00075055	0,00082935	0,00085232	0,00079156	0,00081367
K(aa)	0,0994	0,0979	0,11262059	0,11805757	0,19426559	0,53169155	0,47549915	0,36059760	0,11052780	0,11852745	0,13040819	0,13292771
K(dd)	4,5004	4,6061	4,72338393	4,74592098	4,90824183	6,54247404	6,73653788	7,37797089	4,76777821	4,81763387	4,85013091	5,01203064
K(σσ)	2,8647	2,9589	3,11461706	3,11562106	3,13844857	3,39177486	3,35560056	3,23770230	3,11525401	3,13440718	3,13676067	3,12860378
K(b1b1)	0,1475	0,1503	0,15443094	0,15465281	0,15584974	0,15539461	0,15515828	0,15428454	0,15446637	0,15527005	0,15531642	0,15517921
K(b2b2)	0,1324	0,1348	0,13866993	0,13846284	0,13884890	0,13936249	0,14091270	0,13896769	0,13846409	0,13930211	0,13908423	0,13872669
K(b3b3)	0,1279	0,1302	0,13410296	0,13398415	0,13408429	0,13447468	0,13468126	0,13441748	0,13395692	0,13467111	0,13469987	0,13425397
K(b4b4)	0,1643	0,1681	0,17334622	0,18841764	0,29941780	0,67787268	0,61982062	0,26561664	0,17409607	0,17378453	0,17352939	0,17311860
K(b5b5)	0,1621	0,1654	0,17155408	0,17192294	0,17268732	0,17120498	0,17505057	0,39440237	0,17129626	0,17161206	0,17211393	0,17204277
K(b6b6)	0,1527	0,1560	0,15987792	0,15981176	0,16068443	0,15910148	0,15909090	0,15841614	0,16120873	0,16470759	0,17358924	0,19710188
K(b7b7)	0,1406	0,1440	0,14723875	0,14771334	0,15080593	0,14785646	0,14832384	0,14732069	0,14931457	0,15015011	0,14845663	0,14816666
K(b8b8)	0,1507	0,1544	0,15805353	0,15770767	0,15780362	0,15813307	0,15807252	0,15761124	0,15805416	0,15820227	0,15833352	0,15827641
Eigen=	4,6222	2,9603	3,11520567	3,11610954	3,13882719	7,17418994	7,33324258	3,14747757	4,89846835	3,13440326	4,98467519	5,159066549
werte	2,8692	4,7309	4,85137353	4,87403806	5,04899390	0,76457629	0,68544532	7,89956215	3,11539043	4,95063965	3,13675934	3,12780074
	0,0007	0,0008	0,00082472	0,00097566	0,00089710	0,00077089	0,00075982	0,00075005	0,00082934	0,00085115	0,00079132	0,00081297
	0,0069	0,0241	0,00724230	0,00724227	0,00725451	0,00725415	0,00725382	0,00075005	0,24633701	0,00727769	0,24403700	0,00081297
	0,2295	0,0070	0,24370982	0,26200945	0,00089710	0,00074345	0,00069414	0,23899185	0,00724257	0,24369416	0,00727889	0,00726029
	0,1703	0,1796	0,17333231	0,08866042	0,19380827	0,19969629	0,20413288	0,00721518	0,09829630	0,10155557	0,19626566	0,22516841
	0,0849	0,0880	0,09959809	0,16186111	0,07182573	0,07196583	0,06896891	0,06237572	0,17684733	0,18077912	0,16126804	0,16200362
	0,1553	0,1585	0,11672239	0,17389697	0,12234885	0,17104597	0,17087806	0,17202637	0,16242528	0,16086667	0,10071529	0,08549063
	0,1154	0,1420	0,16024154	0,12760547	0,16971696	0,15452168	0,12228433	0,13057942	0,11677377	0,15370876	0,15011530	0,14788595
	0,1396	0,1156	0,14265422	0,15575095	0,15642862	0,14648704	0,14598886	0,14210169	0,13698313	0,13854561	0,12904596	0,12880127
	0,1276	0,1304	0,13034831	0,14592059	0,14631387	0,12350622	0,15646577	0,15674218	0,14767691	0,12945482	0,11742441	0,11712441
	0,1219	0,1236	0,147477738	0,119156708	0,127432975	0,128389771	0,128995822	0,121599251	0,127974103	0,117344176	0,144840609	0,145082120
Min(EW)	0,000777	0,000825	0,000825	0,000976	0,000897	0,000743	0,000694	0,000750	0,000829	0,000851	0,000791	0,000813
Det	8,82E-12	1,16E-11	1,95E-11	2,42E-11	3,72E-11	8,53E-11	7,10E-11	5,47E-11	1,93E-11	2,24E-11	2,31E-11	2,35E-11

## Appendix Tabelle (A2).4

	146	147	148	150	151	155	156	157	159	160	162	163
a	0,35358	0,74439	0,79551	0,35124	0,38466	0,33070	0,28728	-0,03813	-0,26335	-0,98701	0,33517	0,07145
d	-482,47300	-483,46808	-482,99149	-482,26420	-482,20362	-481,67307	-481,68983	-481,79459	-481,38749	-479,70492	-481,91199	-481,81983
b1	2,34106	2,32921	2,31884	2,35612	2,35113	2,36437	2,37091	2,34613	2,35813	2,35477	2,34107	2,34948
b2	3,69013	3,64573	3,66500	3,66797	3,66296	3,64815	3,65296	3,67332	3,67167	3,66832	3,66870	3,66806
b3	-1,58905	-1,59393	-1,56496	-1,59509	-1,58853	-1,56800	-1,58404	-1,59675	-1,60324	-1,60935	-1,57790	-1,58838
b4	0,77794	0,80684	0,78611	0,79770	0,80833	0,81152	0,80540	0,79907	0,78815	0,82012	0,79858	0,79488
b5	1,75585	1,71526	1,73469	1,73275	1,73296	1,72921	1,74402	1,73253	1,71802	1,72254	1,69561	1,72523
b6	0,86062	0,47785	0,68949	0,99079	1,01353	1,19983	1,19438	1,19458	1,17021	1,17803	1,20259	1,19701
b7	1,31191	1,29361	1,28949	1,29415	1,29883	1,31322	1,30113	1,32709	1,56052	2,30801	1,29447	1,30778
b8	-1,49099	-1,51587	-1,51946	-1,49233	-1,49875	-1,45038	-1,45165	-1,47631	-1,47030	-1,45051	-1,49410	-1,48014
σ	4,81170	4,78522	4,76545	4,80864	4,80570	4,80829	4,81104	4,81939	4,81537	4,76618	4,80797	4,81900
p	0,49251	0,51006	0,49819	0,51581	0,50265	0,50115	0,48275	0,47667	0,51405	0,49532	0,48605	0,51836
Iter.	49,00000	58,00000	52,00000	42,00000	43,00000	35,00000	34,00000	35,00000	83,00000	359,00000	33,00000	34,00000
	-1427,79699	-1507,35424	-2319,47248	-1876,90125	-1667,07927	-1591,67295	-1820,66732	-1294,97434	-1494,91180	-2248,34986	-1765,41818	-1882,78074
Varianzen												
K(pp)	0,00076270	0,00073870	0,000716	0,0007588	0,0008067	0,0008444	0,0008052	0,0008502	0,0007330	0,0007552	0,0008272	0,0007368
K(aa)	0,27932184	0,30042630	0,1954065	0,1666460	0,1577454	0,1222789	0,1203014	0,1170857	0,4059587	1,6667752	0,1296174	0,1211452
K(dd)	5,75010793	6,26075891	5,1607780	5,0351244	4,9102038	4,7553468	4,7574627	4,7513635	5,7207308	12,7273337	4,7505833	4,7583769
K(σσ)	3,14702122	3,19539689	3,1917995	3,1396737	3,1407774	3,1309174	3,1317192	3,1366539	3,1527566	4,6531383	3,1340603	3,1364929
K(b1b1)	0,15519568	0,15477226	0,1542671	0,1550693	0,1548228	0,1551627	0,1557448	0,1535863	0,1557651	0,1550460	0,1548724	0,1553705
K(b2b2)	0,13968286	0,13873011	0,1378980	0,1387943	0,1382269	0,1392942	0,1390527	0,1392480	0,1389871	0,1410768	0,1386876	0,1393575
K(b3b3)	0,13447296	0,13415987	0,1341799	0,1343598	0,1342261	0,1349091	0,1344501	0,1353624	0,1351397	0,1346402	0,1345816	0,1352385
K(b4b4)	0,17378486	0,17286003	0,1721557	0,1732250	0,1734100	0,1735839	0,1735215	0,1738931	0,1735160	0,1753397	0,1730917	0,1738092
K(b5b5)	0,17268903	0,17076481	0,1698130	0,1711279	0,1710035	0,1710836	0,1713309	0,1716012	0,1729853	0,1717133	0,1726531	0,1731285
K(b6b6)	0,41203210	0,44041506	0,2387030	0,2168926	0,1946629	0,1590235	0,1591503	0,1599520	0,1636078	0,1593501	0,1591295	0,1594208
K(b7b7)	0,14777843	0,14749716	0,1466879	0,1480340	0,1476106	0,1475030	0,1477623	0,14646971	0,5071405	1,8478988	0,1478893	0,1484636
K(b8b8)	0,15869978	0,15848367	0,1578190	0,1585386	0,1585353	0,1584870	0,1586966	0,1587234	0,1582905	0,1587369	0,1583806	0,1591738
Eigen=	6,01683250	6,63333902	3,1726934	3,1380524	3,1398993	4,8847704	3,1319458	3,1366570	6,0482213	3,3826904	4,8799773	3,1365091
werte	3,14146832	3,15530300	5,3561988	5,1908182	5,0549756	3,1312248	4,8867845	4,8805010	3,1478063	16,6960104	3,1345417	4,8879297
	0,48874106	0,46469181	0,0007163	0,0007588	0,2473001	0,0008444	0,0008051	0,0008500	0,6521919	0,8748955	0,0008272	0,0007368
	0,00076233	0,00725472	0,2953208	0,2676918	0,0008061	0,2432422	0,0072692	0,2428993	0,0007319	0,0007552	0,2431156	0,2453972
	0,24216111	0,07056401	0,2404110	0,2412238	0,2322116	0,1733147	0,1738926	0,1997976	0,2402157	0,0072569	0,2406477	0,0072810
	0,07083867	0,24161966	0,0863456	0,0893713	0,0964610	0,1628387	0,1123623	0,0867649	0,1688900	0,0694327	0,1135250	0,1768661
	0,16206667	0,16105991	0,1171829	0,1174204	0,1625077	0,1108980	0,1631174	0,1683477	0,0695285	0,1702326	0,1599930	0,1220663
	0,11759291	0,11736191	0,1614867	0,1617507	0,1202183	0,1497630	0,1497630	0,1217915	0,1240368	0,1230132	0,1226430	0,1304349
	0,14742211	0,14604930	0,1479690	0,1489535	0,1174340	0,1478933	0,1177885	0,1534194	0,1558543	0,1378647	0,1464704	0,1500465
	0,12981321	0,12918747	0,1453698	0,1451302	0,1376930	0,1358158	0,1297042	0,1378577	0,1559735	0,1335608	0,1335608	0,1351597
	0,146574545	0,147834859	0,1289436	0,1298053	0,1291215	0,1282332	0,1277592	0,1367843	0,1330106	0,1330315	0,1346743	0,1603626
Min(EW)	0,000762	0,000738	0,000716	0,000759	0,000806	0,000844	0,000805	0,000850	0,000732	0,000755	0,000827	0,000737
Det	4,70E-11	4,68E-11	2,83E-11	2,76E-11	2,75E-11	2,24E-11	2,12E-11	2,23E-11	6,60E-11	2,71E-10	2,33E-11	1,99E-11

## Appendix Tabelle (A2).5

	164	165	166	168	169	171	172	173	174	175	176	177
a	-0,56084	-0,72231	-1,21607	-1,60967	-1,45615	0,71609	0,64861	0,43928	0,43008	-0,53835	-0,54002	-0,18350
d	-481,64780	-481,19898	-481,05171	-481,05171	-481,43687	-481,47151	-481,35908	-481,53735	-481,51269	-482,31929	-482,42519	-482,00459
b1	2,30313	2,26691	2,30313	2,38108	2,33916	2,30531	2,30520	2,31473	2,30572	2,39196	2,36667	2,35656
b2	3,68621	3,68758	3,70224	3,77907	3,70556	3,64973	3,64983	3,66111	3,65629	3,66405	3,64471	3,66132
b3	-1,63177	-1,66797	-1,66644	-1,66346	-1,57747	-1,58082	-1,57037	-1,57251	-1,57242	-1,62867	-1,64137	-1,61070
b4	0,77688	0,75838	0,76610	0,88302	0,83508	0,78299	0,79885	0,79987	0,80487	0,75433	0,77524	0,78078
b5	1,77909	1,79573	1,83838	1,67885	1,57291	1,75778	1,79475	1,79969	1,78223	1,65774	1,63845	1,70545
b6	1,21654	1,23946	1,33870	1,23406	1,32945	1,19730	1,20436	1,20526	1,21523	1,16979	1,18694	1,20124
b7	1,34476	1,33277	1,37291	1,36169	1,29034	1,36587	1,34879	1,31634	1,32751	1,28411	1,27077	1,29483
b8	-1,41542	-1,29927	-1,33575	-1,33575	-1,30318	-1,42345	-1,39250	-1,43727	-1,43419	-1,52071	-1,51894	-1,48886
σ	4,78727	4,6610	4,66774	4,54537	4,59771	4,76727	4,77687	4,79992	4,80083	4,79030	4,79017	4,81615
p	0,49886	0,48652	0,49626	0,52602	0,50640	0,54396	0,49986	0,51568	0,49401	0,50276	0,52174	0,46183
Iter.	39,00000	38,00000	29,00000	31,00000	32,00000	31,00000	30,00000	39,00000	31,00000	44,00000	51,00000	65,00000
	-1846,65773	-2536,27569	-1496,18336	-1528,00126	-1544,91748	-1528,68666	-1455,16360	-2105,75614	-1716,02267	-1608,44778	-2072,08190	-1736,87034
Varianzen												
K(pp)	0,0006829	0,0006826	0,0007516	0,0007314	0,00078	0,00071	0,00078	0,00077	0,00078	0,00081	0,00080	0,00093
K(aa)	0,1311437	0,1302941	0,1021082	0,0942693	0,09997	0,11412	0,11348	0,13735	0,13672	0,14902	0,16494	0,15325
K(dd)	4,7262751	4,7021405	4,6111398	4,4804452	4,49708	4,72036	4,75445	4,77886	4,79180	4,85713	4,95448	4,94064
K(σσ)	3,1310093	3,1262524	2,9674230	2,8245976	2,90589	3,09334	3,09895	3,13673	3,13569	3,14852	3,16543	3,13850
K(b1b1)	0,1550533	0,1556269	0,1510803	0,1453568	0,14615	0,15374	0,15414	0,15534	0,15587	0,15527	0,15452	0,15555
K(b2b2)	0,1361583	0,1376972	0,1343538	0,1308296	0,13130	0,13744	0,13774	0,13854	0,13866	0,13813	0,13865	0,13947
K(b3b3)	0,1342230	0,1341177	0,1304325	0,1257726	0,12727	0,13287	0,13287	0,13439	0,13434	0,13427	0,13503	0,13583
K(b4b4)	0,1723828	0,1720839	0,1672127	0,1618117	0,16411	0,17136	0,17149	0,17270	0,17283	0,17318	0,17257	0,17390
K(b5b5)	0,1712004	0,1707421	0,1661479	0,1602450	0,16336	0,16944	0,17093	0,17318	0,17264	0,17314	0,17553	0,17497
K(b6b6)	0,1585273	0,1582755	0,1553800	0,1498465	0,15130	0,15737	0,15767	0,15879	0,15897	0,15876	0,15843	0,15938
K(b7b7)	0,1473290	0,1465882	0,1431725	0,1386352	0,13947	0,14676	0,14671	0,14727	0,14741	0,14725	0,14801	0,14923
K(b8b8)	0,1589053	0,1578228	0,1555549	0,1500039	0,15154	0,15719	0,15880	0,15886	0,15904	0,15824	0,15856	0,15921
Eigen=	4,8552638	4,8298209	2,9686604	2,8276204	4,61944	3,09391	4,88496	3,13711	4,92360	3,14703	5,10373	3,13815
werte	3,1323920	3,1289794	4,7389436	4,6039683	2,90984	4,84924	3,09906	4,90984	3,13579	4,99544	3,16012	5,07994
	0,0006827	0,0006826	0,0007515	0,0007312	0,00078	0,00070	0,00078	0,00077	0,00078	0,00081	0,00080	0,00093
	0,0072343	0,2443294	0,2350098	0,2268438	0,00688	0,00718	0,00720	0,24464	0,24373	0,00723	0,24627	0,00728
	0,2432177	0,0072119	0,0070394	0,0068265	0,22992	0,24006	0,24122	0,00725	0,00725	0,24257	0,00724	0,24525
	0,1089337	0,1719336	0,0888582	0,1622973	0,09012	0,10595	0,10114	0,17299	0,11083	0,11183	0,17323	0,17603
	0,1212075	0,1056631	0,1153809	0,0873373	0,16565	0,17190	0,17220	0,11262	0,17347	0,11971	0,16627	0,16768
	0,1735079	0,1612183	0,1673074	0,1110242	0,11090	0,15879	0,16056	0,11900	0,11757	0,17459	0,11572	0,11651
	0,1613439	0,1205541	0,1274204	0,1507537	0,15650	0,11620	0,11625	0,16131	0,16206	0,14498	0,12238	0,11938
	0,1328506	0,1554690	0,1469651	0,1262632	0,13955	0,13408	0,14701	0,13008	0,15411	0,16178	0,15635	0,15540
	0,1337263	0,1321304	0,1577749	0,1203715	0,12150	0,12911	0,13870	0,14616	0,14492	0,15806	0,12913	0,12850
	0,1545299	0,1343313	0,1306457	0,1385075	0,12714	0,14755	0,12927	0,15103	0,13067	0,12968	0,14571	0,14582
Min(EW)	0,000683	0,000683	0,000751	0,000731	0,00078	0,000705	0,000782	0,000766	0,000781	0,000806	0,000800	0,000932
Det	1,85E-11	1,77E-11	1,16E-11	7,36E-12	9,24E-12	1,57E-11	1,76E-11	2,25E-11	2,28E-11	2,50E-11	2,78E-11	3,17E-11

## Appendix Tabelle (A2).6

	178	180	181	182	183	185	186	187	188
a	-0,33715	-1,10725	-1,08055	-1,146707191	-1,110861820	0,263968683	0,392197840	-0,147190497	-0,176190409
d	-482,12509	-482,03197	-481,81343	-481,703793677	-481,295442381	-482,141261766	-482,598299468	-481,518724577	-481,613712140
b1	2,36768	2,31682	2,30398	2,387458665	2,294383130	2,344732803	2,344578467	2,356153122	2,353477870
b2	3,66658	3,66176	3,75352	3,769708929	3,734032759	3,678170783	3,670058413	3,669080479	3,668436707
b3	-1,63495	-1,55040	-1,56857	-1,643332988	-1,565744225	-1,589323500	-1,595183656	-1,598243667	-1,599654466
b4	0,76581	0,73123	0,76894	0,798662625	0,753934946	0,796592751	0,784437717	0,801166967	0,801696213
b5	1,70290	1,71960	1,67247	1,649994099	1,741424205	1,742042302	1,731462201	1,737464800	1,742608772
b6	1,20279	1,18528	1,09872	1,141976723	1,204174085	1,187136023	1,196615956	1,196311511	1,203076651
b7	1,28891	1,29572	1,44259	1,395850089	1,459478687	1,331551693	1,313006873	1,315055766	1,315830122
b8	-1,51834	-1,50012	-1,50240	-1,488385172	-1,400422271	-1,669379705	-1,866667741	-1,346394772	-1,402562043
σ	4,80808	4,69102	4,69889	4,682555908	4,691443360	4,814055691	4,810618274	4,818043665	4,816654385
p	0,45264	0,48988	0,47988	0,441037243	0,564340561	0,466355003	0,520689719	0,508415459	0,491324382
Iter.	33,00000	24,00000	29,00000	46,000000000	38,000000000	50,000000000	199,000000000	43,000000000	42,000000000
	-1260,74421	-1273,65755	-1477,04510	-2060,318707053	-1313,849632570	-1291,262849578	-2371,940922081	-1328,608608094	-2388,586624541
Varianzen									
K(pp)	0,00113	0,00091	0,00079	0,000845287	0,000926440	0,001071556	0,000765651	0,000788494	0,000757367
K(aa)	0,12308	0,09053	0,10453	0,131367287	0,121677686	0,225661025	1,042129326	0,219157213	0,147819462
K(dd)	4,84986	4,59411	4,61032	4,631223351	4,659340972	5,137723365	9,098355489	5,50827538	4,914618377
K(σσ)	3,12914	2,94569	3,01106	3,139661637	3,089102907	3,139964113	3,289995231	3,137349591	3,136896435
K(b1b1)	0,15531	0,15023	0,15121	0,151342821	0,150726482	0,155144513	0,155335066	0,156053166	0,155358922
K(b2b2)	0,13890	0,13463	0,13594	0,136574425	0,135920271	0,138990016	0,139027088	0,139101016	0,139040559
K(b3b3)	0,13596	0,13016	0,13094	0,131408985	0,130920901	0,134669419	0,134685571	0,134770023	0,134672956
K(b4b4)	0,17370	0,16760	0,16870	0,168931986	0,168379128	0,173540762	0,173911817	0,174481271	0,173965044
K(b5b5)	0,17209	0,16554	0,16649	0,167166061	0,165841902	0,171468473	0,171534687	0,171642774	0,171820305
K(b6b6)	0,15903	0,15344	0,15562	0,154957609	0,154649106	0,159649210	0,159389865	0,159397095	0,159528986
K(b7b7)	0,14816	0,14280	0,14521	0,144587564	0,145859540	0,148898476	0,147848168	0,147938716	0,147872030
K(b8b8)	0,16003	0,15315	0,15363	0,154060582	0,154395756	0,273136752	1,202098988	0,322945273	0,182918249
Eigen=	3,12892	4,71898	3,01348	4,757039069	4,791022097	5,309702671	10,554685278	5,766157930	3,136717474
werte	4,98470	2,94661	4,73532	3,147436413	3,090731928	3,138931907	3,184818430	3,136722687	5,05825948
	0,00113	0,00091	0,00079	0,000845287	0,000926064	0,001070744	0,942061720	0,381114813	0,000757367
	0,00726	0,00703	0,00707	0,007090458	0,235331109	0,368903432	0,000765651	0,000786522	0,242975222
	0,24260	0,23403	0,23576	0,237432970	0,007073447	0,242561634	0,007267041	0,242493071	0,007274019
	0,10586	0,08815	0,17021	0,170582535	0,169120036	0,007276581	0,242733423	0,007285288	0,214187053
	0,11709	0,16791	0,09307	0,110413778	0,106060129	0,085968555	0,077637825	0,075621371	0,100628566
	0,17328	0,11419	0,15967	0,114751938	0,116763946	0,172972887	0,172648181	0,172001831	0,171893413
	0,14375	0,15506	0,12556	0,158317835	0,130511960	0,120028705	0,118694677	0,136766438	0,121189018
	0,15415	0,14209	0,13077	0,146028210	0,155691498	0,148343700	0,136789599	0,119108042	0,136901623
	0,16000	0,13025	0,14586	0,134562364	0,123472977	0,136331593	0,127888727	0,127947945	0,148124860
	0,12767	0,12359	0,11688	0,127626742	0,150135901	0,127825271	0,149046396	0,148466232	0,128794129
Min(EW)	0,001128	0,000913	0,000785	0,000845	0,000926	0,001071	0,000766	0,000787	0,000757
Det	3,01E-11	1,25E-11	1,32E-11	1,83E-11	1,81E-11	5,36E-11	1,77E-10	3,84E-11	2,49E-11

## Appendix: Tabelle (A3).1

	12	15	17	18	110	116	124	125	126	128	131	133
a	-14,41671795	-222,10412510	-33,38484974	-2,34,36657784	-2,34,36657784	-18,67896975	-39,01537123	-43,82662896	-171,113835888	-236,42700066	-6,6070849	5,07038368
d	-15,28281880	-9,24213422	-0,44542766	3,19296912	3,19296912	5,46408105	-14,61893145	-12,83593233	-9,001986731	0,93949726	3,86356064	6,54585768
b1	31,90229129	227,98747673	41,78034412	4,21060178	4,21060178	30,22104507	29,59254929	28,55335283	12,889838133	3,40264217	31,15851595	30,82159016
b2	33,62693989	7,95165870	42,19479550	238,63956497	238,63956497	35,07232663	28,51940366	25,65114167	11,048232074	4,39770288	33,36227901	33,02529550
b3	22,19742619	1,70117966	22,61477866	-1,18398991	-1,18398991	23,69394027	42,49572923	51,82945101	174,113745994	235,61259999	22,65751651	22,03240173
b4	16,18924159	1,13434160	14,07444361	0,18932737	0,18932737	16,19309054	15,11704751	14,85815244	5,212239474	1,40708622	17,26604603	17,41296375
b5	26,74841765	2,82765456	24,52504353	1,17467915	1,17467915	25,10009106	24,38507495	23,94722594	9,163125019	2,29726267	25,99413732	26,11737714
b6	32,26089579	3,14429627	27,20698120	2,88920090	2,88920090	31,22201673	31,13176074	28,84803200	11,647907303	2,06004739	31,04757188	31,58358280
b7	35,36243579	3,54907803	34,41433126	2,94574394	2,94574394	37,62194753	32,10423670	30,12998172	11,000698062	0,54238845	36,60499273	36,20432991
b8	28,70356752	-1,26476613	24,20488937	0,29295612	0,29295612	29,61681820	27,52301134	27,65296786	9,307794723	-0,60896064	29,71288043	30,19326942
s	56,00912226	17,12222381	54,29589951	10,14662953	10,14662953	56,17523063	52,02333692	53,42415185	32,884228171	8,34803286	57,61332998	57,48513923
p	0,53532141	0,05043560	0,46230678	0,97328294	0,03050306	0,54955943	0,90135073	0,28106025	0,795266607	0,01311493	0,18229406	0,17800311
Iter.	586,00	11,00	262,00	10,00	9,00	534,00	203,00	102,00	57,00	10,00	2768,00	3406,00
Logl	-1878,457813	-1524,029194	-1862,855920	-1332,329245	-1329,410574	-187,6252745	-1866,148250	-1858,385080	-1791,279023	-1255,282765	-1882,628743	-1882,113892
Varianzen												
K(pp)	0,0446743	0,0002765	0,2827440	0,0001180	0,0001152	0,0634943	0,0064499	0,1088461	0,0024828	0,0000428	10,6866799	0,2519067
K(aa)	32,7495468	13,8745578	26,5531524	5,1145394	5,1145394	27,0986431	46,0597441	41,0019562	63,7305751	3,5069072	26,9472659	31,7938971
K(dd)	175,4313876	3,4948041	38,0372024	1,2282477	1,2282477	63,4718429	60,6985506	35,4478212	13,0603246	0,8340762	51,2161501	58,2062752
K(ss)	68313,7146	499,7037063	50884,0204167	61,6252607	61,6252607	59979,7409573	68552,6443304	48310,0756	7973,58506	28,23632	64093,3770	63879,8797
K(b1b1)	20,3067303	12,6463553	20,4579297	0,6809054	0,6809054	19,4798286	17,99292485	17,5527456	7,0804018	0,4635126	19,8612124	19,8086999
K(b2b2)	17,3184343	1,7165238	17,3323707	4,4648865	4,4648865	16,9918206	16,0504589	16,0990570	7,21002005	0,4163014	17,6357659	17,3194839
K(b3b3)	17,2399410	1,6752186	15,8503752	0,5960539	0,5960539	17,0210621	29,4324223	34,4099068	63,0457800	3,2342739	18,5743311	17,6487986
K(b4b4)	23,6286693	2,1932110	21,7021967	0,7753777	0,7753777	23,4295559	22,1283196	21,2806822	8,1331881	0,5206633	24,0353210	23,9896660
K(b5b5)	22,5063622	2,1648356	20,3995494	0,7653650	0,7653650	22,0993706	21,2763845	20,1289786	7,9361980	0,5154386	23,3489245	22,8091448
K(b6b6)	20,2726437	2,0079574	18,6824133	0,7002189	0,7002189	18,6045031	17,6719863	17,6719863	8,0215261	0,4765468	20,72229379	20,5458825
K(b7b7)	17,2584812	1,8528658	15,7769789	0,6504487	0,6504487	16,4389760	16,1352006	16,1352006	7,1863417	0,4525447	18,1271963	18,0365848
K(b8b8)	19,6106639	2,0171336	18,4242171	0,7019241	0,7019241	19,2400812	18,2974712	17,2583262	8,0905233	0,4725723	20,0354689	19,8968148
Eigen-	165,8628666	499,7037063	50884,0276748	0,0001180	0,0001152	59980,0333	68563,2473	48310,179	7980,8588401	0,0000428	64093,3874049	63879,9900
werte	68338,27253	0,0002765	42,9153241	61,6252607	61,6252607	64,3046645	55,7117187	0,1026057	124,0583999	28,236324	1,0604533	61,0879586
	0,0429778	27,1005248	0,2686575	9,7967783	9,7967783	61,7213749	70,0868945	70,0868945	0,0024676	6,9352681	55,8144173	1,0701777
	1,0398694	3,5286710	34,2395130	1,2762857	1,2762857	0,0619307	0,0051467	35,6729618	13,1475093	0,8357489	34,7697228	0,2110084
	30,8289599	0,0942135	0,9539442	1,0492638	1,0492638	0,9681934	0,9421479	0,9421479	0,3515968	0,7181076	7,6302509	33,2156569
	34,6444740	2,9445781	30,9432430	0,0330909	0,0362658	33,7671153	30,8456122	29,7661973	3,2015997	0,0223965	29,9837200	34,9332980
	24,0739513	0,9693186	15,1024141	0,7695052	1,0492638	15,8980458	11,9911655	9,7199237	11,2014224	0,2014772	15,3789183	16,3207035
	16,3412098	2,1374646	11,9823438	0,3062658	0,5277290	24,0056013	21,8427491	20,8534212	7,9540456	0,5036956	25,8319726	23,9923555
	20,1701355	2,0167871	18,4152171	0,6377965	0,7695052	19,4980037	21,0142162	18,2756187	7,5841234	0,3717770	22,2438995	21,0645616
	18,6087037	1,6194811	21,6174010	0,5914156	0,6377965	17,3438985	16,2271892	16,0705013	5,8135876	0,4354126	18,4296600	18,2794341
	17,8632987	1,7448544	16,9391055	0,5277290	0,6898365	18,5611944	18,6542725	19,8705889	6,8374491	0,4806582	20,9584372	20,8157441
	22,3333092	1,4875702	20,1147084	0,6898365	0,5914156	21,9640505	17,3141339	15,6111071	6,0713776	0,3882952	19,0974007	19,2079719
min(EW)	0,0429778	0,0002765	0,2686575	0,0001180	0,0001152	0,0619307	0,0051467	0,1026057	0,0024676	0,0000428	1,0604533	0,2110084
Det	3,19E+16	6,44E+01	1,46E+16	1,02E-04	9,97E-05	1,26E+16	1,05E+15	6,42E+15	5,90E+09	3,46E-07	1,97E+18	6,16E+16

## Appendix: Tabelle (A3).2

	139	140	142	145	147	150	155	159	162	163	170	172
a	-236,53022395	-2,40,9730633	-7,40116821	-23,14010901	-232,83165869	-28,44450807	4,98164800	-18,158701801	-4,74184611	-4,75330288	-8,53434204	10,92121547
d	-0,66871050	0,53633940	1,53194558	-0,68553813	1,98399890	6,09574866	-13,67639019	4,90621813	0,44127531	4,31095841	0,84797127	-5,09993108
b1	3,04920573	2,39333838	30,91630871	29,73339505	4,17359653	29,66717066	31,13084668	11,08324895	31,18703106	30,80535170	31,22690958	29,94984554
b2	4,47670124	4,13442838	33,43735679	31,16857043	4,74591187	30,98051269	32,35478435	11,74707743	33,16850186	33,34983043	33,05463099	32,21654913
b3	-1,18285171	-1,32919131	22,06246805	21,38581422	0,17483047	19,97564116	22,32935471	1,99992144	22,31049686	22,03435017	22,34824361	22,56424532
b4	238,07339416	-1,02886429	17,71507493	16,50745496	1,80909325	16,27046818	17,60094295	6,91155424	17,23612712	17,22746219	17,59580681	17,02622247
b5	1,37043147	243,2865297	25,25026909	22,83801123	2,58312224	24,13659110	25,93197694	4,53017583	26,61486410	26,42769871	25,41208413	27,28371954
b6	1,07756481	1,8094992	32,15760426	42,73136627	234,26480354	45,60159507	31,02886212	4,90054018	31,28690001	31,53376300	31,77991839	31,38106488
b7	2,12151319	1,50559882	37,00688128	34,32669127	2,10585576	34,25782401	37,20390659	19,190147753	36,43147096	36,53486364	36,42376140	36,20961088
b8	-0,80841419	-1,42199972	29,39147442	28,98160500	-1,17158469	28,20460742	30,12396338	8,66886223	29,83152382	29,97649867	30,17393993	30,54414510
s	8,06181967	7,03604946	57,39696206	55,85254402	10,30893903	55,16850846	57,13129645	26,55851336	57,51407169	57,51407169	57,33102149	57,04982054
p	0,02067515	0,95522029	0,73056030	0,98845652	0,01551019	0,07817176	0,60908474	0,92061081	0,02934821	0,65128793	0,59004019	0,80806177
Iter.	3,00	31,00	994,00	697,00	4,00	327,00	305,00	6,00	876,00	8,00	578,00	1665,00
Logl	-1214,491887	-1210,614239	-1881,565742	-1872,283333	-1313,914259	-1869,270137	-1880,514428	-1679,103594	-1882,275481	-1882,113370	-1881,278844	-1879,789716
Varianzen												
K(pp)	0,0004200	0,0001802	0,5273370	0,0525889	0,0001192	0,0354939	0,1298455	0,0005963	3,4981403	2,8514443	0,4096754	0,2081834
K(aa)	3,2668608	2,5527587	23,4381650	27,8898038	5,2082267	29,5831539	21,5933712	25,8635996	22,5101325	28,1991146	24,6851156	29,0980300
K(dd)	0,7954976	0,5968331	57,0287735	39,5887401	1,2760189	43,7512002	55,3171501	8,6221717	47,1635045	59,3455676	59,7542539	59,3942423
K(ss)	24,558615	14,249089	63,999222651	59,0969798739	65,6640155	55,988,3572	62,761,6804727	28,92,5968009	63,834,9085	63,702,4801	63,205,76695	62,447,74660
K(b1b1)	0,4331007	0,3310830	19,6821075	18,8897684	0,7047831	18,3562853	19,6516925	4,5746632	19,9822885	19,8203376	19,7021565	19,7548534
K(b2b2)	0,3889663	0,2969260	17,3054106	16,5326556	0,6324514	16,1822634	17,2993084	4,0474443	17,4654097	17,3471879	17,2524383	17,1915785
K(b3b3)	0,3781847	0,2868410	17,5401857	16,6438715	0,6094003	16,4128042	17,3905952	4,1107117	17,6490076	17,6567347	17,4685324	17,2803767
K(b4b4)	3,3101287	0,3755167	23,9540748	22,7338461	0,7929252	22,3800449	23,7084159	5,2127923	24,0206178	24,1776089	24,2822823	23,6730395
K(b5b5)	0,4820360	2,4147685	22,9874417	21,9885073	0,7828195	21,3549090	22,5016647	5,2048679	23,0550433	22,9928982	22,7540646	22,6330719
K(b6b6)	0,4555278	0,3419014	20,3277571	25,0883947	4,6058001	25,8112615	19,9257805	4,8406474	20,1906548	20,7939970	20,0951234	19,9446838
K(b7b7)	0,4130003	0,3155930	17,6173192	16,8161444	0,6769198	16,4183577	17,5902804	2,64967241	17,8789353	17,9189244	17,4280637	17,3435721
K(b8b8)	0,4431885	0,3377146	19,8508580	18,6967925	0,7249743	18,3343022	19,5562754	4,5594760	19,9580877	19,8373131	19,7097812	19,6164707
Eigen-	0,0004200	0,0001802	63,999,951	59,097,179	65,6640155	55,988,6691	62,761,8795	28,92,5968	63,834,914	63,702,4856	63,205,8440	62,447,8438
werte	6,7251549	14,2490886	0,4842151	44,0106494	0,0001192	48,8005571	0,1222304	0,0005963	48,9204837	64,0159126	61,4891639	0,1774423
	24,5586151	5,0665389	57,3062589	0,0420724	9,9918987	0,0316617	55,8612588	49,18997	2,9522311	2,0069853	0,3854836	63,2008103
	0,8255581	0,6147061	1,0680860	38,5594936	0,0341734	0,9909012	34,2781931	0,2264950	1,0619779	1,0644797	34,7034384	1,0505986
	0,0208811	0,0159073	35,6592077	31,6806775	1,2975887	40,5179490	1,0547347	9,1683398	34,4184801	35,1802899	1,0618669	1,0505986
	0,2172327	0,4500030	26,5075903	1,0062677	1,1044151	31,6682045	24,7522600	7,1276423	25,9061605	27,7717946	24,1471128	26,8425909
	0,5605191	0,1635485	16,6883148	22,0706915	0,3621319	12,9422018	23,8994211	2,1747865	22,9111833	16,6058683	25,0986592	17,3648829
	0,3425037	0,3607115	23,9490487	14,3473672	0,5397982	15,6573218	16,5172035	5,0603669	18,2029090	16,6247460	16,6247460	16,5430770
	0,3584128	0,3001235	17,9598253	20,1131676	0,7375915	21,4036068	21,5592640	3,7352083	23,2675768	23,2675768	22,2764586	20,6761071
	0,4770778	0,2837509	20,1179868	17,4565789	0,5929926	18,5041169	18,5041169	4,2056536	18,1572910	20,2969032	18,9966682	22,4152028
	0,4288669	0,2595030	18,2686634	15,8434134	0,6855668	17,0957602	19,9678553	4,7434153	21,5032761	22,9170943	18,1310675	18,6755305
	0,4102844	0,3351144	22,2412664	19,5914881	0,6681624	19,8553439	17,9488098	4,0559769	19,1697910	19,1976360	20,6150097	24,3911728
min(EW)	0,0004200	0,0001802	0,4842151	0,0420724	0,0001192	0,0316617	0,1222304	0,0005963	1,0619779	1,0644797	0,3854836	0,1774423
Det	1,50E-06	2,50E-08	1,04E+17	4,64E+15	1,50E-04	3,13E+15	2,16E+16	4,18E+06	5,43E+17	5,44E+17	8,78E+16	4,16E+16

### Appendix: Tabelle (A3).3

	178	183	187
a	13,45908609	-6,80430803	-201,47011542
d	-12,55979937	-23,87979542	-6,42452132
b1	28,99668636	30,12545508	7,83619744
b2	33,07036684	32,71657462	9,38428801
b3	22,70679822	22,16265333	2,88490799
b4	17,65155541	17,29746503	2,74377038
b5	27,37438706	26,54805086	8,55601527
b6	30,63943288	30,42843944	7,45965089
b7	36,15146519	36,94354091	4,48925358
b8	30,49544883	31,30910193	204,20182725
s	56,36530430	56,19510908	23,32327260
p	0,10264426	0,805271	0,07166201
Iter.	892,00	312,00	5,00
Logl	-1876,638632	-1879,160707	-1647,433834
Varianzen			
K(pp)	0,0481933	0,031803	0,0003133
K(aa)	21,4404452	32,3909220	23,1432818
K(dd)	79,5357955	101,53161	6,5497193
K(ss)	62898,1348274	69698,73899	1720,4002880
K(b1b1)	19,5623207	19,5529160	3,5595107
K(b2b2)	16,9471626	16,9661142	3,1639962
K(b3b3)	17,0008529	17,1876101	3,0958402
K(b4b4)	23,2871489	23,4256524	4,0794055
K(b5b5)	22,1934405	22,3604313	3,9087278
K(b6b6)	19,6176307	19,7925812	3,6229080
K(b7b7)	17,0800744	17,5119097	3,4669904
K(b8b8)	19,1339582	19,6318617	20,7071742
Eigen-	62900,2772	69709,6218	0,0003133
werte	0,0315012	92,8705855	1720,4002880
	78,2116436	0,0253706	6,6950289
	1,0311213	1,0431514	44,4301145
	33,7572848	32,4031068	0,1746466
	24,2268573	33,8037394	5,6114082
	16,3419172	16,1947035	1,8782715
	23,3030347	24,2078622	4,0398851
	16,9314222	19,9148677	3,1997066
	20,1597292	17,8373065	2,7833987
	17,7812716	18,7890042	2,9961245
	21,9288802	22,4108660	3,4889694
min(EW)	0,0315012	0,0253706	0,0003133
Det	6,62E+15	1,10E+16	1,11E+05

## 8.2 Appendix M

### Abschnitt M.1

Die Loglikelihoodfunktion des vollständigen Datensatzes lautet (nach Kap. 4, Gl. 4.16):

$$\ln(L_c(\theta)) = \left[ \sum_{j=1}^n \sum_{i=1}^3 z_{ji} \cdot \ln(p_{ji}) + z_{ji} \cdot \ln(f_{ji}) \right], \quad (M15)$$

$f_{ji}$  ist Normaldichte mit Erwartungswert  $\mu_{ji}$  und Varianz  $\sigma^2$ .

$$z_{ji} = \begin{cases} 1 & \text{falls der QTL-Genotyp Nr. } i \text{ eintritt, (m\u00f6glich ist: } Q_1Q_1, Q_1Q_2, Q_2Q_2 \text{)} \\ 0 & \text{sonst} \end{cases},$$

$$z_{ji} \cdot z_{jk} = 0 \text{ f\u00fcr } i \neq k!$$

Da die Loglikelihoodfunktion  $\ln(L_c)$  linear in  $z_{ji}$  ist, berechnet sich der Erwartungswert im

(t+1). E-Schritt des EM-Algorithmus als:

$$E \left[ \ln(L_c(\theta | \theta^{(t)})) \right] = \sum_{j=1}^n \sum_{i=1}^3 \left[ \pi_{ji}^{(t)} \cdot \ln(p_{ji}) + \pi_{ji}^{(t)} \cdot \ln(f_{ji}) \right], \quad (M16)$$

$f_{ji}$  ist Normaldichte mit Erwartungswert  $\mu_{ji}$  und Varianz  $\sigma^2$ .

$$\pi_{ji}^{(t)} = \frac{p_{ji} f(y_j; \mu_{ij}^{(t)}, \sigma^{2(t)})}{\sum_{v=1}^3 p_{jv} f(y_j; \mu_{jv}^{(t)}, \sigma^{2(t)})} \quad (M17)$$

Anmerkung: In den EM- und ECM- Intervallkartierungsmethoden sind die  $p_{ji}$  fest, in der "EMV-Methode" h\u00e4ngen die  $p_{ji}^{(t)}$  ebenso wie alle anderen Variablen vom vorangegangenen Iterationsschritt ab!

Der (t+1)' Maximierungsschritt des EM-Algorithmus besteht in der Bildung des Gleichungssystem partieller Ableitungen der erwarteten Loglikelihoodfunktion nach  $a$ ,  $d$ ,  $\beta_1, \dots, \beta_k$ . Dieses lineare Gleichungssystem wird gleich 0 gesetzt, so dass es umgeformt werden kann zu:

$$\mathbf{A}\boldsymbol{\theta}^{(t+1)} = \mathbf{b}, \quad \text{wobei } \boldsymbol{\theta}^{(t+1)} = \left( \mathbf{a}^{(t+1)}, \mathbf{d}^{(t+1)}, \beta_1^{(t+1)}, \dots, \beta_k^{(t+1)} \right)^T. \quad (\text{M18})$$

mit:

$$\begin{pmatrix} \sum_{j=1}^n (\pi_{j1}^{(t)} + \pi_{j3}^{(t)}) & \frac{1}{2} \sum_{j=1}^n (\pi_{j3}^{(t)} - \pi_{j1}^{(t)}) & \sum_{j=1}^n (\pi_{j1}^{(t)} - \pi_{j3}^{(t)}) X_{j1} & \dots & \sum_{j=1}^n (\pi_{j1}^{(t)} - \pi_{j3}^{(t)}) X_{jk} \\ \frac{1}{2} \sum_{j=1}^n (\pi_{j3}^{(t)} - \pi_{j1}^{(t)}) & \frac{1}{4} \sum_{j=1}^n (\pi_{j1}^{(t)} + \pi_{j2}^{(t)} + \pi_{j3}^{(t)}) & \frac{1}{2} \sum_{j=1}^n (-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}) X_{j1} & \dots & \frac{1}{2} \sum_{j=1}^n (-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}) X_{jk} \\ \sum_{j=1}^n (\pi_{j1}^{(t)} - \pi_{j3}^{(t)}) X_{j1} & \frac{1}{2} \sum_{j=1}^n (-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}) X_{j1} & \sum_{j=1}^n X_{j1} X_{j1} & \dots & \sum_{j=1}^n X_{jk} X_{j1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{j=1}^n (\pi_{j1}^{(t)} - \pi_{j3}^{(t)}) X_{jk} & \frac{1}{2} \sum_{j=1}^n (-\pi_{j1}^{(t)} + \pi_{j2}^{(t)} - \pi_{j3}^{(t)}) X_{jk} & \sum_{j=1}^n X_{j1} X_{jk} & \dots & \sum_{j=1}^n X_{jk} X_{jk} \end{pmatrix} \begin{pmatrix} \mathbf{a}^{(t+1)} \\ \mathbf{d}^{(t+1)} \\ \beta_1^{(t+1)} \\ \vdots \\ \beta_k^{(t+1)} \end{pmatrix} = \begin{pmatrix} \sum_{j=2}^n (\pi_{j1}^{(t)} - \pi_{j3}^{(t)}) y_j \\ \frac{1}{2} \sum_{j=1}^n (-\pi_{j1}^{(t)} + \pi_{j1}^{(t)} - \pi_{j3}^{(t)}) y_j \\ \sum_{j=1}^n X_{j1} y_j \\ \vdots \\ \sum_{j=1}^n X_{jk} y_j \end{pmatrix}$$

Das obige Gleichungssystem ist unabhängig von  $\sigma^2$ , daher kann die Lösung für diesen Parameter unabhängig von den anderen partiellen Ableitungen berechnet werden:

$$\sigma^{2(t+1)} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^3 (y_j - \mu_{ji}^{(t+1)})^2 \pi_{ji}^{(t)} \quad (\text{M19})$$

Die Lösung der Maximierung des (t+1)' Iterationsschrittes lautet dann:

$$\boldsymbol{\theta}^{(t+1)} = \left( \mathbf{a}^{(t+1)}, \mathbf{d}^{(t+1)}, \beta_1^{(t+1)}, \dots, \beta_k^{(t+1)}, \sigma^{2(t+1)} \right)^T \quad (\text{M20})$$

Der Gradientenvektor der vollständigen Loglikelihood-Funktion ist gegeben durch:

$$\mathbf{S}(\mathbf{x}) = \begin{pmatrix} \sum_{j=1}^n \sum_{i=1}^3 p_{ji}^{(1)} \cdot z_{ji} \\ \frac{1}{\sigma^2} \sum_{j=1}^n [(y_j - \mu_{j1}) z_{j1} - (y_j - \mu_{j3}) z_{j3}] \\ \frac{1}{2\sigma^2} \sum_{j=1}^n [-(y_j - \mu_{j1}) z_{j1} + (y_j - \mu_{j2}) z_{j2} - (y_j - \mu_{j3}) z_{j3}] \\ \frac{1}{\sigma^2} \sum_{j=1}^n \sum_{i=1}^3 (y_j - \mu_{ji}) z_{ji} X_{jm}, \quad \forall m=1,2,\dots,k \\ -\sum_{j=1}^n \sum_{i=1}^3 \left[ \frac{1}{2} \cdot \frac{1}{\sigma^2} \right] + \frac{1}{2 \cdot \sigma^4} \sum_{j=1}^n \sum_{i=1}^3 (y_j - \mu_{ji}) z_{ji} \end{pmatrix} \quad (\text{M21})$$

Die  $p_{ji}^{(1)}$  sind die ersten Ableitungen der bedingten QTL-Genotypwahrscheinlichkeiten in Tabelle 4.1 (nach p). Für die Informationsmatrizen werden auch die zweiten Ableitungen  $p_{ji}^{(2)}$  benötigt. Diese sind der Arbeit von Kao und Zeng (1997) zu entnehmen und wurden analog in Emrich und Urfer (1999) angegeben.

Tabelle M.2: Dargestellt sei  $I_1$ , die vollständige Informationsmatrix sei  $I_X = \frac{1}{\sigma^2} \cdot I_1$

Parameter	p	a	d	$\sigma^2$	$\beta_w^{\#}$
p	$-\sigma^2 \cdot \sum_{j=1}^n \sum_{i=1}^3 p_{ji}^{(2)} \cdot \pi_{ji}$	0	0	0	0
a	0	$\sum_{j=1}^n \pi_{j1} + \pi_{j3}$	$\frac{1}{2} \cdot \sum_{j=1}^n (-\pi_{j1} + \pi_{j3})$	$\frac{1}{\sigma^2} \cdot \sum_{j=1}^n \left( \begin{array}{l} (y_j - \mu_{j1}) \cdot \pi_{j1} \\ - (y_j - \mu_{j3}) \cdot \pi_{j3} \end{array} \right)$	$\sum_{j=1}^n (\pi_{j1} - \pi_{j3}) \cdot X_{jw}$
d	0	$\sum_{j=1}^n \frac{1}{2} \cdot (-\pi_{j1} + \pi_{j3})$	$\frac{n}{4}$	$\frac{1}{2 \cdot \sigma^2} \cdot \sum_{j=1}^n \left( \begin{array}{l} -(y_j - \mu_{j1}) \cdot \pi_{j1} \\ + (y_j - \mu_{j2}) \cdot \pi_{j2} \\ - (y_j - \mu_{j3}) \cdot \pi_{j3} \end{array} \right)$	$\frac{1}{2} \cdot \sum_{j=1}^n (-\pi_{j1} + \pi_{j2} - \pi_{j3}) \cdot X_{jw}$
$\sigma^2$	0	$\frac{1}{\sigma^2} \cdot \sum_{j=1}^n \left( \begin{array}{l} (y_j - \mu_{j1}) \cdot \pi_{j1} \\ - (y_j - \mu_{j3}) \cdot \pi_{j3} \end{array} \right)$	$\frac{1}{2 \cdot \sigma^2} \cdot \sum_{j=1}^n \left( \begin{array}{l} -(y_j - \mu_{j1}) \cdot \pi_{j1} \\ + (y_j - \mu_{j2}) \cdot \pi_{j2} \\ - (y_j - \mu_{j3}) \cdot \pi_{j3} \end{array} \right)$	$\frac{n}{2 \cdot \sigma^2}$	$\frac{1}{\sigma^2} \cdot \sum_{j=1}^n \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jw} \cdot \pi_{ji}$
$\beta_v^{\#}$	0	$\sum_{j=1}^n (\pi_{j1} - \pi_{j3}) \cdot X_{jv}$	$\frac{1}{2} \cdot \sum_{j=1}^n (-\pi_{j1} + \pi_{j2} - \pi_{j3}) \cdot X_{jv}$	$\frac{1}{\sigma^2} \cdot \sum_{j=1}^n \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jv} \cdot \pi_{ji}$	$\sum_{j=1}^n X_{jv} \cdot X_{jw}$

# v, w=1,2,...,k

Tabelle M.3 enthält Komponente  $I_2$ , entsprechender Teil der Informationsmatrix sei:  $I_{X|Y} |_{\text{Teil 1}} = \frac{1}{\sigma^4} \cdot I_2$

p	p	a
$\sigma^4$	$\sigma^4 \cdot \left\{ \sum_{j=1}^n \sum_{i=1}^3 (p_{ji}^{(1)})^2 \cdot \pi_{ji} + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{mi}^{(1)} \cdot \pi_{mi} \right] \cdot \left[ \sum_{i=1}^3 p_{ji}^{(1)} \cdot \pi_{ji} \right] \right\}$	$\sigma^2 \cdot \left\{ \sum_{j=1}^n \left[ p_{ji}^{(1)} \cdot \pi_{ji} \cdot (y_j - \mu_{ji}) - p_{j\beta}^{(1)} \cdot \pi_{j\beta} \cdot (y_j - \mu_{j\beta}) \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{ji}^{(1)} \cdot \pi_{ji} \right] \cdot \left[ (y_m - \mu_{mi}) \cdot \pi_{mi} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta} \right] \right\}$
a	$\sigma^2 \cdot \left\{ \sum_{j=1}^n \left[ p_{ji}^{(1)} \cdot \pi_{ji} \cdot (y_j - \mu_{ji}) - p_{j\beta}^{(1)} \cdot \pi_{j\beta} \cdot (y_j - \mu_{j\beta}) \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{ji}^{(1)} \cdot \pi_{ji} \right] \cdot \left[ (y_m - \mu_{mi}) \cdot \pi_{mi} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta} \right] \right\}$	$\sum_{j=1}^n \left[ (y_j - \mu_{ji})^2 \cdot \pi_{ji} + (y_j - \mu_{j\beta})^2 \cdot \pi_{j\beta} \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ (y_j - \mu_{ji}) \cdot \pi_{ji} - (y_j - \mu_{j\beta}) \cdot \pi_{j\beta} \right] \cdot \left[ (y_m - \mu_{mi}) \cdot \pi_{mi} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta} \right]$
d	$\sigma^2 \cdot \left\{ \frac{1}{2} \cdot \left[ \sum_{j=1}^n - (y_j - \mu_{ji}) \cdot \pi_{ji} \cdot p_{j1}^{(0)} + (y_j - \mu_{j2}) \cdot \pi_{j2} \cdot p_{j2}^{(0)} - (y_j - \mu_{j\beta}) \cdot \pi_{j\beta} \cdot p_{j\beta}^{(0)} \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{ji}^{(1)} \cdot \pi_{ji} \right] \cdot \left[ \frac{1}{2} \cdot \left[ -(y_m - \mu_{mi}) \cdot \pi_{mi} + (y_m - \mu_{m2}) \cdot \pi_{m2} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta} \right] \right] \right\}$	$\frac{1}{2} \cdot \left\{ \sum_{j=1}^n - (y_j - \mu_{ji})^2 \cdot \pi_{ji} + (y_j - \mu_{j\beta})^2 \cdot \pi_{j\beta} \right\} + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ (y_j - \mu_{ji}) \cdot \pi_{ji} - (y_j - \mu_{j\beta}) \cdot \pi_{j\beta} \right] \cdot \left[ \frac{-(y_m - \mu_{mi}) \cdot \pi_{mi} + (y_m - \mu_{m2}) \cdot \pi_{m2} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta}}{(y_m - \mu_{m2}) \cdot \pi_{m2} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta}} \right]$
$\sigma^2$	$\sigma^2 \cdot \left\{ \sum_{j=1}^n \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot p_{ji}^{(1)} \cdot \pi_{ji} \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{mi}^{(1)} \cdot \pi_{mi} \right] \cdot \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right] \right\}$	$\sum_{j=1}^n \left[ \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{ji}) \cdot \pi_{ji} - \left( \frac{(y_j - \mu_{j\beta})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{j\beta}) \cdot \pi_{j\beta} \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ (y_m - \mu_{mi}) \cdot \pi_{mi} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta} \right] \cdot \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right]$
$\beta_v^{\#}$	$\sigma^2 \cdot \left\{ \sum_{j=1}^n \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jv} \cdot p_{ji}^{(1)} \cdot \pi_{ji} \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{mi}^{(1)} \cdot \pi_{mi} \right] \cdot \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jv} \cdot \pi_{ji} \right] \right\}$	$\sum_{j=1}^n (y_j - \mu_{ji})^2 \cdot X_{jv} \cdot \pi_{ji} - (y_j - \mu_{j\beta})^2 \cdot X_{jv} \cdot \pi_{j\beta} + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ (y_m - \mu_{mi}) \cdot \pi_{mi} - (y_m - \mu_{m\beta}) \cdot \pi_{m\beta} \right] \cdot \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jv} \cdot \pi_{ji} \right]$

# v, w=1,2,...,k

**Tabelle M.4 zeigt Komponente I<sub>3</sub>. Der zweite Teil der Informationsmatrix fehlender Werte lautet  $I_{X|Y} |_{\tau_{\text{el}2}} = \frac{1}{\sigma^4} \cdot I_3$**

Par.	d	$\sigma^2$	$\beta_w$
p	$\sigma^2 \cdot \frac{1}{2} \cdot \sum_{j=1}^n \left[ -(y_j - \mu_{ji}) \cdot \pi_{ji} \cdot p_{ji}^{(1)} + (y_j - \mu_{j2}) \cdot \pi_{j2} \cdot p_{j2}^{(1)} - (y_j - \mu_{j3}) \cdot \pi_{j3} \cdot p_{j3}^{(1)} \right]$ $\sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{ji}^{(1)} \cdot \pi_{ji} \cdot \frac{1}{2} \cdot \begin{bmatrix} -(y_m - \mu_{m1}) \cdot \pi_{m1} \\ + (y_m - \mu_{m2}) \cdot \pi_{m2} \\ - (y_m - \mu_{m3}) \cdot \pi_{m3} \end{bmatrix} \right]$	$\sigma^2 \cdot \left\{ \sum_{j=1}^n \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2\sigma^2} - \frac{1}{2} \right) \cdot p_{ji}^{(1)} \cdot \pi_{ji} \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{mi}^{(1)} \cdot \pi_{mi} \right] \cdot \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right] \right\}$	$\sigma^2 \cdot \left\{ \sum_{j=1}^n \left( \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jw} \cdot p_{ji}^{(1)} \cdot \pi_{ji} \right) + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 p_{mi}^{(1)} \cdot \pi_{mi} \right] \cdot \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jw} \cdot \pi_{ji} \right] \right\}$
a	$\frac{1}{2} \cdot \sum_{j=1}^n \left[ \left( -(y_j - \mu_{ji})^2 \cdot \pi_{ji} + (y_j - \mu_{j3})^2 \cdot \pi_{j3} \right) \right]$ $+ \sum_{\substack{j,m=1 \\ m \neq j}}^n \left\{ \frac{1}{2} \cdot \begin{bmatrix} -(y_m - \mu_{m1}) \cdot \pi_{m1} \\ + (y_m - \mu_{m2}) \cdot \pi_{m2} \\ - (y_m - \mu_{m3}) \cdot \pi_{m3} \end{bmatrix} \right.$ $\left. \cdot \left[ (y_j - \mu_{ji}) \cdot \pi_{ji} - (y_j - \mu_{j3}) \cdot \pi_{j3} \right] \right\}$	$\sum_{j=1}^n \left[ \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{ji}) \cdot \pi_{ji} - \left( \frac{(y_j - \mu_{j3})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{j3}) \cdot \pi_{j3} \right]$ $+ \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ (y_m - \mu_{m1}) \cdot \pi_{m1} - (y_m - \mu_{m3}) \cdot \pi_{m3} \right] \cdot \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right]$	$\sum_{j=1}^n \left[ (y_j - \mu_{ji})^2 \cdot X_{jw} \cdot \pi_{ji} - (y_j - \mu_{j3})^2 \cdot X_{jw} \cdot \pi_{j3} \right] + \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ (y_m - \mu_{m1}) \cdot \pi_{m1} - (y_m - \mu_{m3}) \cdot \pi_{m3} \right] \cdot \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jw} \cdot \pi_{ji} \right]$
d	$\frac{1}{4} \cdot \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \begin{matrix} -(y_j - \mu_{ji}) \cdot \pi_{ji} \\ + (y_j - \mu_{j2}) \cdot \pi_{j2} \\ - (y_j - \mu_{j3}) \cdot \pi_{j3} \end{matrix} \right] \cdot \left[ \begin{matrix} -(y_m - \mu_{m1}) \cdot \pi_{m1} \\ + (y_m - \mu_{m2}) \cdot \pi_{m2} \\ - (y_m - \mu_{m3}) \cdot \pi_{m3} \end{matrix} \right]$	$\sum_{j=1}^n \left[ - \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{ji}) \cdot \pi_{ji} + \left( \frac{(y_j - \mu_{j2})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{j2}) \cdot \pi_{j2} - \left( \frac{(y_j - \mu_{j3})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{j3}) \cdot \pi_{j3} \right]$ $+ \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \frac{1}{2} \cdot \begin{matrix} -(y_m - \mu_{m1}) \cdot \pi_{m1} \\ + (y_m - \mu_{m2}) \cdot \pi_{m2} \\ - (y_m - \mu_{m3}) \cdot \pi_{m3} \end{matrix} \right] \cdot \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right]$	$\frac{1}{2} \cdot \sum_{j=1}^n \left[ -(y_j - \mu_{ji})^2 \cdot X_{jw} \cdot \pi_{ji} + (y_j - \mu_{j2})^2 \cdot X_{jw} \cdot \pi_{j2} - (y_j - \mu_{j3})^2 \cdot X_{jw} \cdot \pi_{j3} \right]$ $+ \sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \frac{1}{2} \cdot \begin{matrix} -(y_m - \mu_{m1}) \cdot \pi_{m1} \\ + (y_m - \mu_{m2}) \cdot \pi_{m2} \\ - (y_m - \mu_{m3}) \cdot \pi_{m3} \end{matrix} \right] \cdot \left[ \sum_{i=1}^3 (y_m - \mu_{mi}) \cdot X_{mw} \cdot \pi_{mi} \right]$

	$\frac{1}{2} \cdot \sum_{j=1}^n \left[ \left( \frac{(y_j - \mu_{j1})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{j1}) \cdot \pi_{j1} + \left( \frac{(y_j - \mu_{j2})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{j2}) \cdot \pi_{j2} - \left( \frac{(y_j - \mu_{j3})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{j3}) \cdot \pi_{j3} \right]$ $+ \sum_{\substack{j,m=1 \\ m \neq j}}^n \frac{1}{2} \cdot \left[ \frac{-(y_m - \mu_{m1}) \cdot \pi_{m1}}{-(y_m - \mu_{m3}) \cdot \pi_{m3}} \cdot \left[ \sum_{i=1}^3 \left( \frac{(y_i - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right] \right]$	$\sum_{j=1}^n \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right] +$ $\sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 \left( \frac{(y_m - \mu_{mi})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{mi} \right] \cdot$ $\left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right]$	$\sum_{j=1}^n \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot (y_j - \mu_{ji}) \cdot X_{jw} \cdot \pi_{ji} +$ $\sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 \left( \frac{(y_m - \mu_{mi})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{mi} \right]$ $\cdot \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jw} \cdot \pi_{ji} \right]$
$\beta_v^{\#}$	$\frac{1}{2} \cdot \sum_{j=1}^n \left[ -(y_j - \mu_{j1})^2 \cdot X_{jv} \cdot \pi_{j1} + (y_j - \mu_{j2})^2 \cdot X_{jv} \cdot \pi_{j2} - (y_j - \mu_{j3})^2 \cdot X_{jv} \cdot \pi_{j3} \right]$ $+ \sum_{\substack{j,m=1 \\ m \neq j}}^n \frac{1}{2} \cdot \left[ \frac{-(y_m - \mu_{m1}) \cdot \pi_{m1}}{-(y_m - \mu_{m3}) \cdot \pi_{m3}} \cdot \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jv} \cdot \pi_{ji} \right] \right]$	$\sum_{j=1}^n \left[ \sum_{i=1}^3 \left( \frac{(y_j - \mu_{ji})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{ji} \right] \cdot (y_j - \mu_{j1}) \cdot X_{jv} \cdot \pi_{j1} +$ $\sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 \left( \frac{(y_m - \mu_{mi})^2}{2 \cdot \sigma^2} - \frac{1}{2} \right) \cdot \pi_{mi} \right] \cdot \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jv} \cdot \pi_{ji} \right]$	$\sum_{j=1}^n (y_j - \mu_{j1})^2 \cdot X_{jv} \cdot X_{jw} \cdot \pi_{j1} +$ $\sum_{\substack{j,m=1 \\ m \neq j}}^n \left[ \sum_{i=1}^3 (y_j - \mu_{ji}) \cdot X_{jv} \cdot \pi_{ji} \right] \cdot \left[ \sum_{i=1}^3 (y_m - \mu_{mi}) \cdot X_{mw} \cdot \pi_{mi} \right]$

# v, w=1,2,...,k

## 9. Literatur

Basten, C.J., B.S. Weir, and Z.-B. Zeng. (1999). QTL Cartographer. Department of Statistics, North Carolina State University, Raleigh.

Czihak, Gerhard, Langer, H., Ziegler, H. [Hrsg.] (1995): Biologie. 4. verb. u. erw. Aufl., Berlin [u.a.], Springer.

Dempster, A. P., Laird, N. M. and Rubin, D. B., (1977): Maximum Likelihood from Incomplete Data via the EM Algorithm. J. R. Statist. Soc. B 39, pp. 1-38.

Doerge, R. W., Zeng, Z.-B. and Weir, B. S. (1997): Statistical Issues in the Search for Genes Affecting QTL in Experimental Populations. Statistical Science; Vol 12 No. 3, pp. 195-219.

Emrich, K., 1999: Schätzung genetischer Parameter unter Verwendung molekularbiologischer Daten in der Pflanzenzüchtung. Diplomarbeit, Fachbereich Statistik, Universität Dortmund.

Emrich, K. and Urfer, W., (1999): Estimation of genetic parameters using molecular markers and EM algorithms. Technical Report 48/1999. Department of Statistics, University of Dortmund. Available from the world wide web: <http://www.statistik.uni-dortmund.de/sfb475/sfblit.htm>

Fahrmeir, L., Kaufmann, H. L., Ost, F. (1981): Stochastische Prozesse, Carl Hanser, München  
Wien,

Fichtner ,A. (2000): Auswertung eines Mais-Feldexperimentes zur Untersuchung von Genotypeneffekten bei quantitativen Merkmalen. Dortmund, Univ., Diplomarbeit, 2000

Fisch, R. D., Ragot, M. and Gay, G. (1996): A Generalization of the Mixture Model in the Mapping of Quantitative Trait Loci for Progeny From a Biparental Cross of Inbred Lines. *Genetics* 143, pp. 571-577.

Haley, C.S., and S.A. Knott. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: pp. 315-324.

Haley, C.S., Knott, S. A. and Elsen, J.M. (1994): Mapping QTL in crosses between outbred lines using least squares. *Genetics* 136, pp. 1195-1207.

Hennig, W. (1998): *Genetik*, 2. Auflage, Springer Verlag

Jansen, R. C. (1992): A general mixture model for mapping QTL by using molecular markers. *Theoretical and Applied Genetics* 85, pp. 252-260

Jansen, R.C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics* 135, pp. 205-211.

Jansen, R. C. (1994): Controlling the type I and II errors in mapping QTL. *Genetics* 138, pp. 871-881.

Jansen, R.C., and Stam, P.. (1994). High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* 136, pp. 1447-1455.

Jiang, C. and Zeng, Z.-B. (1995): Multiple trait analysis of genetic mapping for QTL. *Genetics* 140, pp. 1111-1127.

Johnson, R. A. and Wichern, D. W. (1999): *Applied Multivariate Statistical Analysis*. Ausg. 4, Prentice Hall, New Jersey.

Kao, C.-H., and Zeng, Z.-B., (1997): General formulas for obtaining the MLE's and the asymptotic variance-covariance matrix in mapping quantitative trait loci when using the EM algorithm. *Biometrics* 53. pp. 653-665.

Kao, C.-H., Zeng, Z.-B. and Teasdale, R. D., (1999): Multiple Interval Mapping for Quantitative Trait Loci. *Genetics* **152**, pp. 1203-1216.

Knapp, S. J., Bridges, W. C. and Birkes, D. (1990): Mapping QTL using molecular marker linkage maps. *Theoretical and Applied Genetics* 79, pp. 583-592.

Knott, S. A. and Haley, C. S. (1992): Aspects of maximum likelihood methods for the mapping of QTL in line crosses. *Genetical Research* 60, pp. 139-151.

Kull, H./Knodel, U. (1980): *Genetik und Molekularbiologie*, 2. Auflage, Stud. R. Biol. 1.

Lander, E. S. and Botstein, D. (1989): Mapping Mendelian Factors Underlying Quantitative Traits Using RFLP Linkage Maps. *Genetics* 121: pp. 185-199.

Lander, E.S., Green, P. Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E. and Newburg, L. (1987): MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: pp. 174-181

Lincoln S., Daly M., Lander E. (1992). Mapping genes controlling quantitative traits with MAPMAKER/QTL. Whitehead Institute Technical Report. 2nd edition.

Lincoln, S.E., Daly, M. J., and Lander, E. S. (1993): Constructing genetic linkage maps with MAPMAKER/EXP version 3.0: A tutorial and reference manual. Whitehead Institute for Biomedical Research, Cambridge, MA.

Little, R. J. A. (1992): Regression With Missing X's: A Review. *JASA*, Vol. 87, No. 420, pp. 1227-1237.

Little, R. J. A. and Rubin, D. B. (1986): *Statistical Analysis with Missing Data*. Wiley, New York.

Liu, B.-H. (1998): *Statistical Genomics: Linkage, Mapping, and QTL Analysis*. CRC Press

Lorieux, M., Goffinet, B., Perrier, X., de Gonzalez, Leon and Lanaud, C. (1995): Maximum-Likelihood models for mapping genetic marker markers showing segregation distortion. 1. Backcross populations. *Theor. Appl. Genet.* 90: pp. 73-80.

Lorieux, M., Perrier, X., Goffinet, B., de Gonzalez and Lanaud, C. , Leon (1995): Maximum-Likelihood models for mapping genetic marker markers showing segregation distortion. 2. F2 populations. *Theor. Appl. Genet.* 90: pp. 73-80.

Louis, T. A. (1982): Finding the Observed Information Matrix when Using the EM Algorithm, *J. R. Statist. Soc. B*, 44, No. 2, pp 226-233.

Martinez, O., and R. N. Curnow, (1992): Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* 85, pp 480-488.

Martinez, O., and R. N. Curnow, (1994): Missing markers when estimating QTL using regression mapping. *Heredity* 73: pp. 198-206.

McLachlan, G., Krishnan, T. (1997): *The EM Algorithm and Extensions*, Wiley

McLachlan, G., Peel, D. (2000): *Finite Mixture Models*, Wiley

Melchinger, A.E., Utz, H.F. and Schön, C.C., (1998): Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. *Genetics* 149, pp 383-403.

Meng, X.-L. and D. B. Rubin, (1993). Maximum likelihood estimation via the ECM algorithm: A general framework. *Biometrika* 80 pp. 267–268.

Moreno-Gonzalez, J. (1992): Genetic models to estimate additive and non-additive effects of marker-associated QTL using multiple regression techniques. *Theoretical and Applied Genetics* 85, pp. 435-444.

- Orchard, T. and Woodbury, M. A. (1972): A missing information principle: theory and applications. Proc. Sixth Berkeley Symp. Math. Stat. & Prob., 1, pp. 698-715.
- Rodolphe, F. and Lefort, M. (1993): A multi-marker model for detecting chromosomal segments displaying QTL activity. Genetics 134, pp. 1277-1288.
- Sachs, K. (1923): The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8, pp. 552-560.
- Sari-Gorla, M., Krajewski, P., Binelli, G., Frova, C., Taramino, G. and Villa, M. (1997): Genetic dissection of herbicide tolerance in maize by molecular markers. Molecular Breeding 3: pp. 481-493.
- Sari-Gorla, M., Calinski, T., Kaczmarek, Z. and Krajewski, P. (1997): Detection of QTL  $\times$  environment interaction in maize by a least squares interval mapping method. Heredity 78, pp. 146-157.
- Schoen, C.C., Melchinger, A. E., Boppenmaier, Brunklaus-Jung, E., Herrmann, R. G. and Seitzer, J. F. (1994). RFLP mapping in maize. Quantitative trait loci affecting testcross performance of elite European flint lines. Crop Sci. 34: pp. 379-389.
- Sillanpää, M. J. and Arjas, E. (1998): Bayesian Mapping of Multiple Quantitative Trait Loci From Incomplete Inbred Line Cross Data. Genetics 148, pp. 1373-1388.
- Titterton, D. M., Smith, A. F. M., Makov, U. E. (1985): Statistical Analysis of Finite Mixture Models. Wiley, New York

Urfer, W., Emrich, K., Mejza, S. and Hering, F. (1999): Quantitative trait loci mapping in plant genetics by  $\alpha$ -design experiments and molecular genetic marker systems. Technical Report 34/1999, University of Dortmund. Available from the world wide web: <http://www.statistik.uni-dortmund.de/sfb475/sfbliit.htm>.

Utz, H.F. and Melchinger, A.E., (1996): PLABQTL: A program for composite interval mapping of QTL. J. Quant. Trait Loci. 2, pp. 1-5.

Wu, C. F. J. (1983): On the convergence properties of the EM algorithm, Ann. Statist. 11, pp. 95- 103.

Zeng, Z.-B. (1993). Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA. 90, pp. 10972-10976.

Zeng, Z.-B., (1994). Precision mapping of quantitative trait loci. Genetics 136: pp. 1457-1468

S-Plus ist registrierter Markenname von Mathsoft Inc., Seattle, Washington.

Betreuer der Dissertation und 1. Gutachter: Prof. Dr. W. Urfer

2. Gutachter: Prof. Dr. F. Hering

Prüfungskommission:

Vorsitz: Dekan Prof. Dr. J. Kunert

Prof. Dr. W. Urfer, : Prof. Dr. F. Hering, Dr. K.-H. Loesgen

## **Danksagung**

Ich danke vor Allem Herrn Prof. Dr. W. Urfer für seine hilfreiche und aufmunternde Betreuung, Herrn Prof. Dr. F. Hering für wertvolle Anmerkungen und Herrn Prof. Dr. H. F. Utz (Universität Hohenheim) für die Bereitstellung der Experimentaldaten.

Die finanzielle Unterstützung durch die DFG wird dankbar anerkannt.

Prüfungstermin: 14. 8. 2002