

Adaptive Modelle  
zur Struktur- und Mustererkennung in stochastischen Zeitreihen  
und ihre Anwendung in der Biosignalanalyse

DISSERTATION

zur Erlangung des akademischen Grades  
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt dem Rat der Fakultät für Mathematik und Informatik  
der Friedrich-Schiller-Universität Jena

von Dipl.-Math. Eva Möller geb. Barešová  
geboren am 20. November 1965 in Košice/Slowakei

## INHALTSVERZEICHNIS

<b>0 ANLIEGEN DER ARBEIT</b>	<b>1</b>
<b>1. MODELLE DER STOCHASTISCHEN APPROXIMATION</b>	<b>5</b>
<b>1.1 Konsistente rekursive Verfahren         zur Schätzung statistischer Parameter</b>	<b>5</b>
<b>1.2 Schätzung der Parameter eines ARMA-Prozesses</b>	<b>8</b>
1.2.1 Definition und Eigenschaften von ARMA-Prozessen	9
1.2.2 Eigenschaften des rekursiven Schätzverfahrens	14
1.2.3 ARMA-Modellierung von Biosignalen	21
<b>2 ADAPTIVE VERFAHREN</b>	<b>27</b>
<b>2.1 Wahl der Steuerfolge in rekursiven Schätzverfahren</b>	<b>28</b>
<b>2.2 Beschreibung und Eigenschaften adaptiver         rekursiver Algorithmen</b>	<b>35</b>
<b>2.3 Konstruktion eines adaptiven Schätzverfahrens         zur Anpassung von ARMA-Modellen</b>	<b>42</b>
2.3.1 Vergleich der adaptiven Modellanpassung mit konsistenten Schätzverfahren	47
2.3.2 Adaptivität und Konvergenzverhalten der adaptiven ARMA-Modellierung	48
<b>2.4 Bestimmung der Steuerfolge         in rekursiven Quantilschätzungen</b>	<b>53</b>

<b>3 EINSATZMÖGLICHKEITEN ADAPTIVER VERFAHREN</b>	
<b>IN DER BIOSIGNALANALYSE</b>	63
<b>3.1 Einsatzmöglichkeiten der adaptiven ARMA-Modellierung</b>	63
<b>3.2 Anwendung der adaptiven Trendschätzung</b>	
<b>zur Berechnung der Durchblutungsgeschwindigkeit</b>	
<b>in retinalen Gefäßen des Auges</b>	65
<b>3.3 Adaptive Segmentation von Langzeit-EEG-Signalen</b>	72
3.3.1 Aufgabenstellung, bekannte Lösungsansätze	72
3.3.2 Konstruktion adaptiver Schwellen	78
3.3.3 Anwendung adaptiver rekursiver Algorithmen	
<b>im Segmentationsverfahren</b>	80
<b>Literaturverzeichnis</b>	86
<b>Abbildungen</b>	

## 0. ANLIEGEN DER ARBEIT

In vielen Bereichen der Praxis werden Signale gemessen, die charakteristische Zustände eines beobachteten Objektes widerspiegeln. Zur Beschreibung der Signale werden Modelle und Methoden aus der Theorie von Zeitreihen angewandt, deren Kenngrößen charakteristische Muster und Strukturen im Signal quantitativ erfassen. Natürlich beschreiben die Werte der Kenngrößen die Signaleigenschaften nur dann korrekt, wenn die Modellvoraussetzungen vom Meßsignal erfüllt wurden.

Da biologische Signale, wie z.B. Elektroenzephalogramm, -kardiogramm, -myogramm u.a., stochastischen Charakter besitzen, können ihre Kenngrößen mit Hilfe von Schätzfunktionen, die sich aus der Analyse von zufälligen Prozessen ergeben, bestimmt werden. Eine wesentliche Voraussetzung weit verbreiteter Modelle ist die Stationarität und Ergodizität der Prozesse, d.h., ihre charakteristischen Parameter sind zeitinvariant und können aus einer Realisierung des Prozesses geschätzt werden. Sie erlauben weiter die Analyse der Zeitreihen im spektralen Bereich (vgl. z.B. [2], [31], [1], [16] u.a.). Da sich aber die Zustände eines biologischen Objektes im allgemeinen mit der Zeit verändern, kann eine Beobachtung des Signals nicht als Realisierung eines stationären Prozesses aufgefaßt werden. In der Biosignalanalyse handelt es sich überwiegend um Signale mit einem stark instationären Charakter. Werden für ihre Beschreibung Parameter stationärer Modelle angewandt, ergeben die Schätzverfahren bei einer konkreten Anwendung nur mittlere Parameter, die keine Informationen über einzelne Muster und Strukturen im Signal enthalten. Andere Modelle nehmen oft eine bestimmte Art der Instationarität an und schränken damit die Klasse der Signale, für die sie angewandt werden können, ein (vgl. z.B. [63], [16], [1] u.a.). Aus diesen Gründen werden neue Methoden benötigt, die robust gegenüber der Strukturvariabilität des Signals sind und eine breite Klasse von Signalen sachgerecht beschreiben können. Die Entwicklung der Meß- und Rechentechnik, die eine immer schnellere Aufnahme von ständig wachsenden Datenmengen gestattet, stellt bei der Analyse der Meßsignale zusätzliche Ansprüche an die Effizienz der Auswertalgorithmen.

In dieser Arbeit werden Verfahren, die auf der Theorie der stochastischen Approximation aufbauen, untersucht. Für stationäre schwach abhängige Prozesse können mit ihrer Hilfe Schätzfunktionen der wichtigsten statistischen Parame-

ter mit sehr guten Eigenschaften konstruiert werden (vgl. [15], [47]). Ihre rekursive Berechnungsvorschrift ist von besonderem Vorteil. Die Werte der Parameter können sehr schnell und effektiv berechnet und bei einer sequentiellen Messung stets aktualisiert werden. Bei Verzicht auf Konvergenz können nach den Vorschlägen aus [22] die Verfahren so gestaltet werden, daß die Schätzfolgen einen dynamischen Charakter gewinnen und sich den Eigenschaften der vorliegenden Zeitreihe anpassen. Ihr adaptives Verhalten ermöglicht es, Strukturbrüche im Signal zu erkennen. In stationären Abschnitten befinden sich die Schätzfolgen dennoch in der Nähe der mittleren Werte. Somit können die Verfahren für die Auswertung beliebiger Signale angewandt werden. Ihre Konstruktionsweise wird zur adaptiven Gestaltung der Schätzfolgen von statischen Parametern und zur Entwicklung neuer dynamischer Kenngrößen genutzt. Aufbauend auf Ergebnisse der Arbeiten [22], [23], [24] und [53] werden in dieser Arbeit die Eigenschaften der Verfahren weiter untersucht. Im Mittelpunkt steht die Entwicklung neuer adaptiver Algorithmen und die Bestimmung der sogenannten Steuerfolge, welche die Eigenschaften der adaptiven Schätzungen wesentlich beeinflusst. Neben zahlreichen allgemeinen Hinweisen zur Anwendung der entwickelten Verfahren in der Analyse von Biosignalen werden Lösungen einiger ausgewählter Probleme detailliert beschrieben.

Im ersten Kapitel werden Schätzverfahren, die aus der Theorie der stochastischen Approximation abgeleitet werden können, dargestellt. Zu diesen Verfahren gehören auch die Algorithmen zur Schätzung der Parameter eines ARMA-Prozesses von Günther [28]. ARMA-Modelle, die Zeitreihen mit wenigen Parametern erfassen, zählen zu den wichtigsten Methoden zur Beschreibung von zufälligen stationären Zeitreihen und sind auch für die Analyse von Biosignalen bekannt. Da der Algorithmus aus [28] bislang bzgl. seiner praktischen Effizienz nur wenig getestet wurde, ist es erforderlich, ihn für die Zwecke der Biosignalanalyse zu untersuchen. Von Bedeutung ist vor allem die Güte der Schätzung für Modelle höherer Ordnung. Anschließend wird anhand dieses Verfahrens die ARMA-Modellierung von ausgewählten Biosignalen demonstriert.

Im 2. Kapitel werden adaptive rekursive Verfahren behandelt. In [22] wird die Adaptivität der Verfahren der stochastischen Approximation erreicht, in dem für sie eine konstante zeitunabhängige Steuerfolge gewählt wird. Ziel dieser Arbeit ist es, Kriterien zur Wahl der Steuerkonstanten aufzuzeigen und zu untersuchen. Das Ausgangsmodell wird erweitert, in dem für die einzelnen Ver-

fahren Steuerfolgen konstruiert werden, die das gewünschte, oben beschriebene Verhalten der Schätzfolgen begünstigen. Es wird angestrebt, durch geeignete signalabhängige Steueralgorithmen die Länge der Adaptionsphase der Schätzfolge sowie ihre Streuung in stationären Abschnitten zu minimieren. Steuerfolgen, die von den Werten der Zeitreihe abhängen, wurden für die konsistenten Verfahren zur Konvergenzbeschleunigung und Minimierung der Streuung der Schätzfolgen konstruiert (vgl. [38], [47]). Ausgehend von diesen Methoden werden solche Steueralgorithmen gesucht, welche die bekannten adaptiven Verfahren in dynamische selbstlernende Operatoren überführen. Eine solche Steuerung wird auch bei der Konstruktion eines adaptiven rekursiven Algorithmus zur Schätzung der Parameter eines ARMA-Prozesses vorgenommen. Das Verfahren wird bzgl. seiner Eigenschaften detailliert untersucht. Der Einfluß der einzelnen Steuerparameter wird an zahlreichen Beispielen demonstriert.

Im 3. Kapitel werden Lösungen konkreter Aufgaben aus der Biosignalanalyse beschrieben, die durch die Anwendung der adaptiven rekursiven Verfahren erreicht wurden. Die adaptive Trendschätzung wird bei der Berechnung der Durchblutungsgeschwindigkeit in Gefäßen am Augenhintergrund angewandt. Eine Verbesserung der adaptiven Segmentation von Langzeit-EEG-Messungen, also die Zerlegung der EEG-Meßsignale in Epochen, die gleichartige Muster und Strukturen enthalten, wird durch die Anwendung der dynamischen rekursiven Kenngrößen, die momentane Eigenschaften des Signals beschreiben, und durch die Konstruktion einer adaptiven Schwelle erreicht.

An dieser Stelle möchte ich Herrn Prof. Gert Griebach für das Thema und die Betreuung der Arbeit danken. Weiter danke ich dem Direktor des Instituts für Medizinische Statistik, Informatik und Dokumentation, Herrn Prof. Herbert Witte, für die Unterstützung bei der Erstellung der Arbeit. Mein besonderer Dank gehört Frau Dr. Bärbel Schack, mit der ich zum Thema der adaptiven ARMA-Modellierung zusammengearbeitet habe, und Herrn Dr. Vladimír Krajča für zahlreiche Diskussionen zum Thema der adaptiven Segmentation.

## Bezeichnungen

$\mathbb{N}$	- Menge der natürlichen Zahlen
$\mathbb{R}$	- Menge der reellen Zahlen
$\mathbb{C}$	- Menge der komplexen Zahlen
$x, \xi$	- Zufallsgrößen
$Ex$	- Erwartungswert von $x$
$X = \{x_t\}$	- zufälliger Prozeß (kursiv)
$X = \{x_t\}$	- Realisierung eines zufälligen Prozesses (roman)
$c, r, z,$	- Vektoren (Fettschrift)
$Z = \{z_t\}$	- Vektorprozeß
$\hat{a}, \hat{r}, \hat{\sigma}$	- Schätzung eines Parameters
$\mathcal{M}^c, \mathcal{L}^d, \mathcal{E}, \mathcal{G}$	- Operatoren
$\mathbb{1}_A = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases}$	- Indikatorfunktion
$\langle a, b \rangle = \sum_{i=1}^n a_i b_i$	- Skalarprodukt
mathematische Stichprobe	- Folge unabhängiger identisch verteilter Zufallsgrößen

## 1 MODELLE DER STOCHASTISCHEN APPROXIMATION

Mathematische Methoden und Verfahren der Zeitreihenanalyse sind eine wesentliche Grundlage der Auswertung technischer oder biologischer Meßsignale. Dabei ist man an Kenngrößen interessiert, die die charakteristischen Eigenschaften der Zeitreihe erfassen, sich möglichst genau und effektiv bestimmen lassen und gewissen statistischen Gütekriterien genügen.

Aus der Theorie der stochastischen Approximation lassen sich sowohl für mathematische Stichproben als auch für schwach abhängige stationäre Zeitreihen zahlreiche Schätzfunktionen entwickeln, die die wichtigsten statistischen Kenngrößen rekursiv bestimmen. Es sind z.B. Schätzfunktionen für den Erwartungswert, Quantilwert oder für die Parameter eines ARMA-Prozesses. Aufgrund der rekursiven Berechnungsvorschrift können die geschätzten Parameter in ihrer Entwicklung beobachtet werden. Ihre Vorteile liegen in der günstigen rechentechnischen Umsetzung, da die Berechnung aus dem letzten Schätz- und dem aktuellen Meßwert erfolgt.

Die Schätzfolgen der Verteilungsparameter können aus dem Grundsatz der stochastischen Approximation abgeleitet werden, wie im nächsten Abschnitt gezeigt wird. Sie lassen sich einfach realisieren und liefern bei der Anwendung an realen Signalen, die der Voraussetzung der Stationarität genügen, gute Schätzwerte. Die Anpassung von ARMA-Modellen an Biosignale ist bekannt. Da aber bei der Modellierung realer Prozesse nur in den seltensten Fällen Kenntnisse über die in den Modellen benötigten theoretischen Voraussetzungen vorliegen, sind in jedem Anwendungsgebiet Untersuchungen zur Modellwahl und zur Güte der Schätzverfahren durchzuführen. Die Möglichkeit, ARMA-Modelle mit Hilfe des Algorithmus aus [28], der eng mit den Modellen der stochastischen Approximation zusammenhängt, an Biosignale anzupassen, wird im zweiten Teil dieses Kapitels untersucht.

### 1.1 Konsistente rekursive Verfahren zur Schätzung statistischer Parameter

Die Theorie der stochastischen Approximation beschäftigt sich mit der Bestimmung von Nullstellen einer Abbildung  $R: \mathbb{R}^n \rightarrow \mathbb{R}^n$ , deren Werte nicht exakt bekannt sind, man hat nur eine zufällige Folge  $Y = \{y_t\}_{t=0}^{\infty}$  zur Verfügung, für die  $Ey_t = R(x_t)$ ,  $t \in \mathbb{N}$ , gilt. Unter speziellen Bedingungen an die Abbildung  $R$

und die Folge  $Y$  kann eine rekursive zufällige Folge  $X = \{x_t\}_{t=0}^{\infty}$  konstruiert werden, die mit Wahrscheinlichkeit Eins gegen die Nullstelle von  $R$  konvergiert. Diese Herangehensweise ermöglicht für die wichtigsten statistischen Kenngrößen, wie z.B. Momenten-, Quantilwertschätzung oder Schätzung der Dichtefunktion, eine Entwicklung rekursiver Schätzfunktionen, die sowohl auf mathematische Stichproben als auch auf stark mischende stochastische Prozesse anwendbar sind.

Die ersten Arbeiten zu dieser Problematik für den eindimensionalen Fall sind z.B. [51], [67] oder [39]. Eine methodische Zusammenfassung der existierenden Ergebnisse wurde von Nevelson und Chasminski [47] durchgeführt. Hier sind auch die notwendigen Bedingungen an  $R$  und  $Y$ , die Konstruktion der Folge  $X$  und der Beweis ihrer starken Konsistenz zu finden. Dieser Satz wurde von Borodin [15] auf mehrdimensionale stark mischende zufällige Prozesse erweitert. Die Konvergenz der Folge  $X$  mit Wahrscheinlichkeit Eins wurde mit Hilfe der Martingaltheorie und der Theorie der Markowschen Prozesse bewiesen. Für diese Arbeit ist es ausreichend, den Satz über die starke Konsistenz der rekursiven Schätzung von Borodin für den eindimensionalen Fall zu formulieren.

Es sei  $\{\xi_t(\omega)\}_{t=0}^{\infty}$  eine Folge stationärer Zufallsgrößen über dem Wahrscheinlichkeitsraum  $[\Omega, \mathcal{F}, \mathcal{P}]$  mit Werten in  $\mathbb{R}$ . Die Folge sei stark mischend, d.h. der Mischungskoeffizient

$$\alpha(\tau) = \sup_{s>0} \sup_{A \in \mathcal{F}_0^s, B \in \mathcal{F}_{s+\tau}^{\infty}} (\mathcal{P}(A \cap B) - \mathcal{P}(A)\mathcal{P}(B)) \quad (1.1)$$

konvergiert mit  $\tau \rightarrow \infty$  gegen Null, wobei die Familien von  $\sigma$ -Algebren  $\mathcal{F}_s^t$  durch  $\mathcal{F}_s^t = \sigma(\xi_s, \dots, \xi_t)$  definiert sind (Mit  $\sigma$  wird die bzgl.  $\xi_s, \dots, \xi_t$  kleinste  $\sigma$ -Algebra bezeichnet).

Weiter seien  $R(x)$  eine reelle und  $G(t, x, \omega)$  eine zufällige Funktion mit Werten in  $\mathbb{R}$ .  $G$  sei für  $t \geq 0$  und  $\omega \in \Omega$  in jedem Punkt  $x$  meßbar. Es gelte

$$EG(t, x, \omega) = 0 \quad \text{für } x \in \mathbb{R} \text{ und } t \geq 0. \quad (1.2)$$

Im weiteren werden folgende Bedingungen vorausgesetzt ( $C$  bezeichne eine in den folgenden Formeln nicht notwendig gleiche Konstante).

i) Die Gleichung

$$R(x) = 0 \quad (1.3)$$

besitzt eine eindeutige Lösung  $r$ .

ii) Es existiert  $\delta G/\delta x$  und ein  $\gamma > 0$ , so daß gilt:

$$\mathbf{E} \sup_{x \in \mathbb{R}} |G(t, x, \omega)|^\gamma < C \quad \text{und} \quad \mathbf{E} \sup_{x \in \mathbb{R}} \left| \frac{\delta G}{\delta x}(t, x, \omega) \right| < C. \quad (1.4)$$

iii) Für alle  $x \in \mathbb{R}$  gelte

$$R(x) (x-r) \leq 0. \quad (1.5)$$

und

$$|R(x)| \leq C(1+|x|). \quad (1.6)$$

iv) Für den Mischungskoeffizienten  $\alpha(\tau)$  gilt

$$\alpha(\tau) \leq C (\ln \tau)^{-2(m+2)(1+h)/m}, \quad m \in \mathbb{N} \quad (1.7)$$

und

$$\alpha(\tau) \leq C \tau^{-h} \quad h > 0. \quad (1.8)$$

**Satz 1.1:**

Sei  $R$  eine reellwertige Funktion auf  $\mathbb{R}$  und  $G(t, x, \omega)$  eine zufällige Funktion auf  $\mathbb{R}$ , die für alle  $x$  meßbar ist. Es gelten die Bedingungen (1.2) bis (1.8) und  $c_t = \frac{c}{t+1}$ ,  $c = \text{const.}$  Dann konvergiert die Folge  $X = \{x_t\}_{t=0}^\infty$ , die mittels

$$\begin{aligned} x_0 &= x \quad (x \in \mathbb{R}) \\ x_{t+1} &= x_t + c_t (R(x_t) + G(t+1, x_t, \omega)) \end{aligned} \quad (1.9)$$

definiert wird, mit Wahrscheinlichkeit Eins gegen die Nullstelle  $r$  der Funktion  $R$ .

In der allgemeinen Betrachtung von  $n$ -dimensionalen Prozessen wird in [15] die Bedingung (1.5) viel schärfer formuliert. Für den eindimensionalen Fall bedeutet sie folgendes. Es existiert ein  $\lambda > 0$ , so daß für alle  $x \in \mathbb{R}$  die Bedingung  $R(x) (x-r) < -\lambda (x-r)^2$  gilt. In [53] wurde der Satz in der Form, wie er hier angegeben ist, bewiesen und für die Konstruktion einer rekursiven Schätzung des Quantilwertes angewandt.

**Satz 1.2:**

Sei  $\{\xi_t(\omega)\}_{t=0}^\infty$  ein stark stationärer und stark mischer zufälliger Prozeß über dem Wahrscheinlichkeitsraum  $[\Omega, \mathcal{F}, \mathcal{P}]$ . Für den Mischungskoeffizienten gelten die Bedingungen (1.7) und (1.8). Falls  $\mathbf{E} \xi_t^{2k+1} < \infty$  ist, dann existiert das  $k$ -te Moment  $\mu_k = \mathbf{E} \xi_t^k$  und die Folge

$$\begin{aligned} m_0 &= m \\ m_{t+1} &= m_t + \frac{c}{t+1} (\xi_{t+1}^k - m_t), \quad c = \text{const.} \end{aligned} \quad (1.10)$$

konvergiert mit Wahrscheinlichkeit Eins gegen  $\mu_k$ , wobei  $m$  ein beliebiger aber fester Startwert ist.

Im Beweis wird die Funktion  $R$  als

$$R(x) = \mu_k - x$$

definiert. Für die Korrekturfolge  $Y$  mit

$$y_{t+1}(m_t, \omega) := R(m_t) + G(t+1, m_t, \omega) = \xi_{t+1}^k(\omega) - m_t$$

können die Eigenschaften (1.2) bis (1.6) einfach nachgewiesen werden (vgl. [22]).

**Satz 1.3:**

Sei  $\{\xi_t(\omega)\}_{t=0}^{\infty}$  eine Folge stark mischender Zufallsgrößen mit der absolut stetigen Verteilungsfunktion  $F_{\xi}$  und der Dichtefunktion  $f_{\xi}$ . Für den  $\alpha$ -Quantil  $q_{\alpha}$  von  $F_{\xi}$  gelte  $f_{\xi}(q_{\alpha}) > 0$ . Dann konvergiert die Folge  $\{q_{\alpha}(t)\}_{t=0}^{\infty}$ , die durch

$$\begin{aligned} q_{\alpha}(0) &= q \\ q_{\alpha}(t+1) &= q_{\alpha}(t) + \frac{c}{t+1} \left( \alpha - I_{(-\infty, q_t(t))}(\xi_{t+1}) \right) \end{aligned} \quad (1.11)$$

definiert ist, mit Wahrscheinlichkeit Eins gegen  $q_{\alpha}$ .

Beweis: [22], [53]

Es kann gezeigt werden, daß für  $c = 1$  die Momentenschätzung (1.10) nach  $t$  Rekursionsschritten dem empirischen Moment einer Stichprobe des Umfangs  $t$  entspricht (vgl. [22]). Die Funktion (1.11) bildet eine neue Schätzung der Quantile, die im Vergleich zum empirischem Quantil rekursiv mit jedem neuen Stichproben- bzw. Prozeßwert aktualisiert werden kann.

## 1.2 Schätzung der Parameter eines ARMA-Prozesses

Eine verbreitete und in der Biosignalanalyse bekannte Klasse von Modellen sind die linearen parametrischen Prozesse, unter denen autoregressive Prozesse gleitender Summen eine zentrale Rolle einnehmen. Unter Verwendung des zweiten Theorems von Weierstraß kann bewiesen werden, daß sie stationäre Prozesse mit stetiger Spektraldichte hinreichend gut approximieren (vgl. [2],

S. 445). Erfahrungen weisen aus, daß derartige Annahmen auch für Biosignale getroffen werden können und damit gute Anpassungen möglich sind.

Bei der Anpassung von ARMA-Modellen an stationäre Prozesse ist die wichtigste Aufgabe, zuverlässige Verfahren zur Schätzung der Modellparameter zu entwickeln. Für diese Arbeit sind die rekursiven Verfahren von Günther [28] von Bedeutung, die in Anlehnung an Methoden aus der Theorie der stochastischen Approximation entwickelt wurden. Im folgenden soll die Güte der Parameterschätzung für endliche simulierte ARMA-Zeitreihen höherer Ordnung sowie die Güte der Schätzung der Spektraldichte für simulierte Signale und reale Biosignale untersucht werden. Im letzten Teil werden Möglichkeiten zur Beschreibung ausgewählter Biosignale durch lineare parametrische Modelle behandelt. Zuvor sollen dennoch die notwendigen Definitionen, Bezeichnungen und Eigenschaften der ARMA-Prozesse bereitgestellt werden.

### 1.2.1 Definition und Eigenschaften von ARMA-Prozessen

Betrachtet wird eine Folge von reellen Zufallsgrößen  $E = \{\varepsilon_t\}_{t=0}^{\infty}$  über dem Wahrscheinlichkeitsraum  $[\Omega, \mathcal{F}, \mathcal{P}]$  mit folgenden Eigenschaften:

1.  $E\varepsilon_t = 0$ , für alle  $t \in \mathbb{N}$ ,
2.  $E\varepsilon_t \varepsilon_s = R_{\varepsilon}(t-s) = \sigma_{\varepsilon}^2 \delta_{t,s}$  für  $t, s \in \mathbb{N}$ .

Es seien  $p$  und  $q$  nichtnegative ganze Zahlen und  $a_0, \dots, a_p, b_0, \dots, b_q \in \mathbb{R}$  mit:

$$a_0 > 0, \quad a_p \neq 0 \quad (1.12)$$

und  $b_0 > 0, \quad b_q \neq 0. \quad (1.13)$

Weiterhin seien

$$P(z) = \sum_{k=0}^p a_k z^{p-k} \quad \text{und} \quad Q(z) = \sum_{l=0}^q b_l z^{q-l}, \quad z \in \mathbb{C}$$

zwei Polynome mit folgenden Eigenschaften:

$$\text{aus } P(z) = 0 \quad \text{folgt} \quad |z| < 1, \quad (1.14)$$

$$\text{aus } Q(z) = 0 \quad \text{folgt} \quad |z| < 1, \quad (1.15)$$

$$\text{aus } P(z) = 0 \quad \text{und} \quad Q(z') = 0 \quad \text{folgt} \quad z \neq z'. \quad (1.16)$$

#### Satz 1.4:

Es seien  $p$  und  $q$  nichtnegative ganze Zahlen und  $(a_0, \dots, a_p)$  und  $(b_0, \dots, b_q)$  zwei Folgen reeller Zahlen mit den Eigenschaften (1.12) bis (1.16). Ist

$\{\varepsilon_t\}_{t=0}^{\infty}$  ein Prozeß, für den die Eigenschaften 1. und 2. gelten, dann wird durch die Beziehung

$$\sum_{k=0}^p a_k x_{t-k} = \sum_{l=0}^q b_l \varepsilon_{t-l}, \quad t \in \mathbb{N}, \quad (1.17)$$

ein schwach stationärer, invertierbarer und unverkürzbarer Prozeß  $X = \{x_t\}_{t=0}^{\infty}$  mit  $E x_t = 0$  und  $\text{Var}(x_t) < \infty$  definiert und eindeutig bestimmt.

Zum Beweis vgl. z.B. [2], [1] oder [3].

Der Prozeß  $X$  ist ein autoregressiver Prozeß gleitender Summen oder kurz ein ARMA(p,q)-Prozeß. Für  $q = 0$  erhält man eine sogenannte autoregressive Folge der Ordnung  $p$ , kurz als AR(p)-Prozeß bezeichnet. Für  $p = 0$  ist  $X$  eine Folge von gleitenden Summen der Ordnung  $q$ , kurz als MA(q)-Prozeß (moving average process) bezeichnet. ARMA(p,q) ist das gemischte Modell, wobei  $p$  die Ordnung des AR- und  $q$  des MA-Anteils ist. Zur Vereinfachung des Modells wird ohne Beschränkung der Allgemeinheit im weiteren  $a_0 = b_0 = 1$  gesetzt.

#### Eigenschaften:

1. Zu jeder gegebenen Autokovarianzfunktion eines ARMA-Prozesses gibt es genau einen invertierbaren ARMA-Prozeß. (Siehe [2])
2. Ein ARMA-Prozeß, der den Bedingungen des Satzes 1.4 genügt, besitzt eine absolut stetige Spektralfunktion und damit eine Spektraldichte  $f$ , deren Darstellung sich aus (1.17) herleiten läßt:

$$f(\lambda) = f_{\varepsilon}(\lambda) \cdot \frac{\left| 1 + \sum_{k=1}^q b_k e^{-i\lambda k} \right|^2}{\left| 1 + \sum_{k=1}^p a_k e^{-i\lambda k} \right|^2}, \quad \lambda \in [-\pi, \pi], \quad (1.18)$$

wobei  $f_{\varepsilon}(\lambda) = \sigma_{\varepsilon}^2 / 2\pi$  die Spektraldichte des Prozesses  $E = \{\varepsilon_t\}_{t=0}^{\infty}$  ist [1].

3. Nach dem Dekompositionstheorem von Wold kann ein beliebiger ARMA(p,q)-, bzw. MA(q)-Prozeß als ein AR-Prozeß möglicherweise unendlicher Ordnung dargestellt werden. Ebenso läßt sich jeder ARMA(p,q)-, bzw. AR(p)-Prozeß als ein MA-Prozeß möglicherweise unendlicher Ordnung darstellen (vgl. z.B. [69] oder [16]).

Die Verwendung von ARMA-Modellen zur Beschreibung zufälliger Zeitreihen ist im folgenden Satz begründet.

**Satz 1.5:**

Sei  $g$  eine stetige Spektraldichte eines reellen Prozesses. Dann existiert für alle  $\varepsilon > 0$  ein  $p$  und ein  $q$ ,  $p, q \in \mathbb{N}$ , und ein ARMA( $p, q$ )-Prozeß mit der Spektraldichte

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{\left| \sum_{l=0}^q b_l e^{-il\lambda} \right|^2}{\left| \sum_{k=0}^p a_k e^{-ik\lambda} \right|^2},$$

so daß für alle  $\lambda \in [-\pi, \pi)$  gilt:  $|g(\lambda) - f(\lambda)| \leq \varepsilon$ .

Dieser Satz folgt aus dem zweiten Approximationstheorem von Weierstraß (vgl. z.B. [43]). Für den Beweis vgl. [2].

Der Entwicklung zuverlässiger Verfahren zur Schätzung der Modellparameter wurde in den letzten Jahren große Aufmerksamkeit gewidmet. Die meisten der Algorithmen ergeben sich aus der folgenden Aufgabenstellung.

Sei  $\{x_t\}_{t=0}^\infty$  ein stationärer Prozeß und  $c_r \in \mathbb{R}^r$  mit  $r = p + q$  ein Parametervektor,  $c_r = (c_1, \dots, c_r)^\top$ . Gesucht wird ein  $c_r^0 \in \mathbb{R}^r$ , für den gilt:

$$E|e_t(c_r^0)|^2 = \min_{c_r \in \mathbb{R}^r} E|e_t(c_r)|^2, \quad (1.19)$$

wobei

$$e_t = x_t + \sum_{k=1}^p c_k x_{t-k} - \sum_{l=1}^q c_{p+l} e_{t-l} \quad (1.20)$$

ist. Wie bekannt, ist dafür

$$\text{grad } E|e_t(c_r^0)| = 0 \quad (1.21)$$

notwendig. Sei

$$z_t = (x_t, \dots, x_{t-p+1}; -e_t, \dots, -e_{t-q+1})^\top, \quad (1.22)$$

dann ist

$$e_t = x_t + \langle c_r, z_{t-1} \rangle, \quad (1.23)$$

$$E|e_t(c_r)|^2 = E|x_t|^2 + 2 E x_t \langle c_r, z_{t-1} \rangle + E \langle c_r, z_{t-1} \rangle^2 \quad (1.24)$$

und schließlich

$$\text{grad } E|e_t(c_r)|^2 = 2 E x_t z_{t-1} + 2 E \langle c_r, z_{t-1} \rangle z_{t-1}. \quad (1.25)$$

Somit gilt nach (1.10) für  $c_r^\circ$

$$E x_t z_{t-1} + E \langle c_r^\circ, z_{t-1} \rangle z_{t-1} = 0 \quad (1.26)$$

oder auch

$$E e_t(c_r^\circ) z_{t-1} = 0. \quad (1.27)$$

Ist  $\{x_t\}_{t=0}^\infty$  ein ARMA(p,q)-Prozeß mit dem Parametervektor  $\Theta_{p,q} = (a_1, \dots, a_p; b_1, \dots, b_q)$ , so bekommt man (1.26) durch Multiplizieren von (1.17) mit  $z_{t-1}$  und durch die Bildung des Erwartungswerts. Somit nimmt die Gleichung (1.24) für  $\Theta_{p,q}$  ein Minimum an und es gilt

$$\min_{c_r \in \mathbb{R}^r} E|e_t(c_r)|^2 = E|\varepsilon_t|^2 = \sigma_\varepsilon^2. \quad (1.28)$$

Für den Fall eines AR-Prozesses ( $q = 0$ ) ist (1.26) äquivalent zu dem System der Yule-Walker-Gleichungen

$$R_{XX} a_p = r_{XX}$$

mit  $a_p = (a_1, \dots, a_p)$ ,  $R_{XX} = (R_{XX}(|k-l|))_{k,l=1}^p$  und  $r_{XX} = (R_{XX}(1), \dots, R_{XX}(p))$ , wobei mit  $R_{XX}$  die Autokovarianzfunktion bezeichnet wird (vgl. [1], [16], [66], [75] u.a.). Es kann mit dem Levinson-Durbin-Algorithmus sukzessiv gelöst werden (vgl. [37], [49], [16], [45], [20], [42] u.a.). Das Verfahren besitzt gute numerische Eigenschaften (vgl. [19]) und ist auch von theoretischer Bedeutung (siehe [17], [44]).

Eine andere Möglichkeit zur Lösung der Gleichung (1.19) bieten iterative numerische Verfahren, zu denen auch das rekursive Gradientenverfahren gehört. Aus ihm ergibt sich die folgende Berechnung des Parametervektors

$$c_r(t) = c_r(t-1) + k_t \text{grad } E|e_t(c_r(t-1))|^2. \quad (1.30)$$

Aus (1.25), (1.27) und (1.30) folgt:

$$c_r(t) = c_r(t-1) + 2k_t E z_{t-1} e_t(c_r(t-1)). \quad (1.31)$$

Die Konvergenz des Verfahrens wird für  $k_t \in (0, 2/\lambda_{\max})$  gewährleistet, wobei  $\lambda_{\max}$  der größte Eigenwert der Matrix  $E z_{t-1} z_{t-1}^T$  ist (vgl. z.B. [50], Kap.2). Die Lösung der Gleichung (1.27) ist eindeutig, wenn die Matrix  $E z_{t-1} z_{t-1}^T$  positiv definit ist. Diese Bedingung ist für Prozesse mit  $R_{XX}(0) > 0$  erfüllt (s. [16] Kap. 7).

Für die Schätzung der Parameter bei einer unbekanntem Kovarianzfunktion wurden zahlreiche Verfahren mit unterschiedlicher statistischer Güte entwickelt. Dabei sind sowohl der Parametervektor als die Modellordnung  $p$  und  $q$  zu schätzen. Am häufigsten wird bei der Konstruktion solcher Verfahren der Levinson-Durbin-Algorithmus als eine gemeinsame Grundlage verwendet (vgl. z.B. [16], [49], [37], u.a.). Im einfachsten Fall werden  $q = 0$  gewählt und für die Autokorrelationsmatrix  $\mathbf{R}_{XX}$  die empirische Schätzung  $\hat{\mathbf{R}}_{XX}$  mit den Koeffizienten

$$\hat{\mathbf{R}}_{XX}(k,N) = \frac{1}{N} \sum_{t=k}^{N-1} x_{t-k} x_t \quad k < N. \quad (1.37)$$

verwendet. Für  $\hat{\mathbf{R}}_{XX}(0,N) > 0$  ist die Matrix  $\hat{\mathbf{R}}_{XX}$  positiv definit (vgl. [16]). Der nach dem Levinson-Durbin-Algorithmus bestimmte Parametervektor  $\hat{\mathbf{a}}_n = (\hat{a}_1, \dots, \hat{a}_n)$  hat die folgenden Eigenschaften. Wenn  $p$  die Ordnung des Prozesses ist und  $n \geq p$ , dann gilt  $\mathcal{P}(\lim_{N \rightarrow \infty} \hat{a}_k = a_k) = 1$ ,  $k = 1, \dots, n$ , wobei  $a_k = 0$  für  $p < k \leq n$  ist (siehe [44], Lemma 1(h)). Die Schätzung  $\hat{\sigma}_p^2$  des Parameters  $\sigma_\varepsilon^2$  ist eine asymptotisch erwartungstreue und konsistente Schätzung (siehe [1]). Der Levinson-Durbin-Algorithmus ist für eine gut konditionierte Matrix  $\hat{\mathbf{R}}_{XX}$  (d.h.  $\det(\hat{\mathbf{R}}_{XX}) > 0$ ) stabil (vgl. [19]). Numerische Instabilität tritt auf, wenn  $\det(\hat{\mathbf{R}}_{XX}) \approx 0$  ist. Dieser Fall liegt dann vor, wenn die untersuchte Zeitreihe eine Überlagerung komplexer Sinusoide ist. Das Verfahren benötigt  $O(n^2)$  Operationen. Für die Schätzung der MA-Parameter sind ebenfalls zahlreiche Verfahren bekannt. Es ist dennoch wesentlich komplizierter, für diese Parameter Schätzungen mit ähnlich guten statistischen Eigenschaften zu erreichen (vgl. [16], [37], [49]).

In engem Zusammenhang mit der Gradientenmethode (1.30) stehen die rekursiven Verfahren aus [28] zur Schätzung der Parameter eines ARMA-Prozesses, die im nächsten Abschnitt behandelt werden.

Für die Schätzung der Modellordnung sind zahlreiche Funktionen  $\phi_N(k,l)$  bekannt, die für  $k = p$  und  $l = q$  mit  $N \rightarrow \infty$  ihr Minimum annehmen (siehe [44]). Diese Verfahren besitzen unterschiedliche statistische Güte und es empfiehlt sich, für eine konkrete Anwendung in Abhängigkeit von der Länge der Zeitreihen und dem Analyseziel die einzelnen Kriterien zu testen. Z.B. zeigen empirische Untersuchungen, daß für endliche simulierte AR-Zeitreihen eines kleinen Umfangs ( $N \leq 512$ ) die konsistente Schätzung von Hannan und Quin [33] oft eine Ordnung ermittelt, die kleiner ist als die theoretische. Das

AIC-Kriterium  $AIC(k) = N \ln \hat{\sigma}_k^2 + 2k$ ,  $k=1, \dots, n$ , liefert für diesen Fall die richtige bzw. eine größere Ordnung (siehe [62] und [3]). Da für Schätzungen mit einer Ordnung, die größer ist als die theoretische, positive statistische Aussagen vorliegen (siehe [44]), empfiehlt es sich, für solche Zeitreihen das AIC-Kriterium zu wählen. Für Schätzungen mit einer kleineren Ordnung als  $p$  sind keine Aussagen über die Schätzungsgüte möglich (vgl. [44], Lemma 1). Für die Ordnungsschätzung gemischter Modelle ist das konsistente Kriterium von Hannan und Rissanen bekannt (siehe [34]).

### 1.2.2 Eigenschaften des rekursiven Schätzverfahrens

Im vorangegangenen Abschnitt wurde gezeigt, daß der Parametervektor eines ARMA-Prozesses die Nullstelle vom Gradienten des mittleren quadratischen Fehlers ist, der im allgemeinen nicht bekannt sein muß. In der Sprache der stochastischen Approximation bedeutet das, daß die Funktion  $R$  als eine Abbildung von  $\mathbb{R}^{p+q}$  auf  $\mathbb{R}^{p+q}$  mittels

$$R(\mathbf{c}_r) = \text{grad } E |e_t(\mathbf{c}_r)|^2$$

definiert wird. Nach (1.16) ist

$$\text{grad } E |e_t(\mathbf{c}_r)|^2 = 2E e_t(\mathbf{c}_r) \mathbf{z}_{t-1}.$$

Es gilt weiter

$$y_t(\hat{\mathbf{c}}_r(t-1), \xi_t) = e_t(\hat{\mathbf{c}}_r(t-1)) \mathbf{z}_{t-1} = (x_t - \langle \hat{\mathbf{c}}_r(t-1), \mathbf{z}_{t-1} \rangle) \mathbf{z}_{t-1},$$

wobei  $\{\xi_t\}_{t=0}^{\infty}$  mit  $\xi_t = (x_t, \mathbf{z}_{t-1})$  ein  $(p+q+1)$  dimensionaler Prozeß ist. Da er nicht stationär ist, kann der Satz der stochastischen Approximation von Borodin [16] nicht zum Beweis der Konvergenz der Schätzung angewandt werden; deshalb ist ein eigenständiger Beweis notwendig. Dieser Fall wurde von Günther (siehe [28] und [29]) vollständig behandelt, die folgenden Sätze wurden bewiesen und einige Aussagen zur Konvergenzgeschwindigkeit sind gemacht worden.

Zuerst werden einfache Kriterien für die Stationarität und Invertierbarkeit eines ARMA-Prozesses angegeben, die für die Konstruktion der Schätzalgorithmen von Bedeutung sind.

**Lemma 1.1:**

i) Für die Stationarität eines ARMA(p,q)-Prozesses ist hinreichend, daß gilt  $\sum_{j=1}^p |a_j| < 1$ . Für die Invertierbarkeit ist die Bedingung  $\sum_{j=1}^q |b_j| < 1$  hinreichend.

ii) Für das Polynom  $P_p(\theta, z) = 1 + \sum_{j=1}^p \theta_j z^j$  mit  $\theta = (\theta_1, \dots, \theta_p)^T \in \mathbb{R}^p$  und  $\theta_p \neq 0$  gelte  $P_p(\theta, z) \neq 0$  für  $|z| \leq 1$ . Dann gilt  $|\theta_j| < \binom{p}{j}$ .

Sind für die Parametervektoren  $(a_1, \dots, a_p)$  und  $(b_1, \dots, b_q)$  die Bedingungen aus i) erfüllt, so genügen sie auch den Bedingungen von ii).

Beweis: siehe [28].

Somit kann die Eigenschaft der Beschränktheit der Parameter vorausgesetzt und in dem Schätzalgorithmus benutzt werden.

Die Parameter werden nach dem folgenden rekursiven Verfahren geschätzt:

$$\begin{aligned} e_t &:= 0, & t < 1 \\ e_t &:= x_t + \sum_{j=1}^p \hat{a}_j(t-1) x_{t-j} + \sum_{j=1}^q \hat{b}_j(t-1) e_{t-j}, & t \geq 1 \end{aligned} \tag{E}$$

$$\begin{aligned} \hat{a}_j(t) &:= 0, \quad \text{für } t \leq j, \quad j = 1, \dots, p, \\ \hat{a}_j(t) &:= \begin{cases} d = \hat{a}_j(t-1) - c_{t-1} e_t x_{t-j}, & |d| < \binom{p}{j} \\ \binom{p}{j} \operatorname{sgn}(d), & \text{sonst} \end{cases}, \\ &\text{für } t > j, \quad j = 1, \dots, p, \end{aligned} \tag{A}$$

$$\begin{aligned} \hat{b}_j(t) &:= 0, \quad \text{für } t \leq j, \quad j = 1, \dots, q, \\ \hat{b}_j(t) &:= \begin{cases} d = \hat{b}_j(t-1) - c_{t-1} e_t e_{t-j}, & |d| < \rho_j \\ \rho_j \operatorname{sgn}(d), & \text{sonst} \end{cases}, \\ &\text{für } t > j, \quad j = 1, \dots, q, \end{aligned} \tag{B}$$

$$\hat{\sigma}_\varepsilon^2(t) := \hat{\sigma}_\varepsilon^2(t-1) - 1/t (\hat{\sigma}_\varepsilon^2(t-1) - e_t^2), \quad t = 1, 2, \dots \tag{S}$$

Bemerkung: Wie aus (E) deutlich wird, haben die geschätzten Gleitmittelparameter im Vergleich zu (1.17) das umgekehrte Vorzeichen.

Für den Vektor  $(\rho_1, \dots, \rho_q)$  gilt:

$$\rho_j > 0 \quad (j=1, \dots, q) \quad \text{und} \quad \sum_{j=1}^q \rho_j \leq 1.$$

Die Folge  $\{c_t\}_{t=1}^{\infty}$  steuert die Schätzung unter Berücksichtigung des Vorhersagefehlers  $e_t$ . Diese sogenannte Steuerfolge kann für das angegebene Verfahren auf zwei Arten gewählt werden.

Als erstes sei  $\{c_t\}_{t=1}^{\infty}$  eine Folge reeller Zahlen, die den folgenden Bedingungen genügt:

$$\begin{aligned} \text{(i)} \quad & \sum_{t=1}^{\infty} c_t = \infty \\ \text{(ii)} \quad & \sum_{t=1}^{\infty} c_t^2 < \infty \end{aligned} \tag{1.32}$$

**Satz 1.6:**

Sei  $X = \{x_t\}_{t=0}^{\infty}$  ein ARMA(p,q)-Prozeß, für den die Bedingungen des Satzes 1.1 gelten. Zusätzlich sei  $E = \{\varepsilon_t\}_{t=0}^{\infty}$  eine Folge unabhängiger identisch verteilter Zufallsgrößen mit  $E\varepsilon_t = 0$ ,  $\text{Var } \varepsilon_t = \sigma_{\varepsilon}^2$  und  $E\varepsilon_t^4 < \infty$ . Dann konvergiert das Verfahren (E), (A), (B) mit der deterministischen Steuerfolge  $\{c_t\}_{t=0}^{\infty}$ , die nach (1.32) berechnet wird, im Quadratmittel gegen den Parametervektor  $(a_1, \dots, a_p; b_1, \dots, b_q)$  und die mittels (S) erzeugte Schätzfunktion ist eine asymptotisch erwartungstreue Schätzung von  $\sigma_{\varepsilon}^2$ .

Zum Beweis vgl. [29] sowie [28], Theorem 2.1 und Folgerung 2.1.

Eine in der stochastischen Approximation üblicherweise benutzte Zahlenfolge, die der Bedingung (1.32) genügt, ist

$$c_t = c/t, \quad c = \text{const.} \quad t = 1, 2, \dots \tag{1.33}$$

Eine für die Konvergenzgeschwindigkeit günstige Wahl der Konstanten c hängt von den Eigenschaften des Prozesses  $\{x_t\}_{t=0}^{\infty}$  ab, insbesondere von seiner Streuung, und setzt einige Vorinformationen über die Zeitreihe voraus. In Abbildung 1.1 wird das unterschiedliche Verhalten der Schätzfolgen für eine ungünstige und eine angepaßte Wahl der Konstante c dargestellt.

Weiter betrachte man  $\{c_t\}_{t=1}^{\infty}$  als eine Folge von Zufallsgrößen, definiert durch

$$\begin{aligned} c_t &= s_t^{-1} & t &= 1, 2, \dots \\ s_1 &= 1 + x_1^2 + c & & \\ s_t &= s_{t-1} + x_t^2 + c & c &\geq 0. \end{aligned} \tag{1.34}$$

**Satz 1.7:**

Sei  $X = \{x_t\}_{t=0}^{\infty}$  ein AR(p)-Prozeß. Weiter gelten die Bedingungen des Satzes 2.4.1. Wird die Steuerfolge nach (1.34) konstruiert, so konvergieren die Folgen (A) und (S) mit Wahrscheinlichkeit Eins gegen die Parameter  $a_1, \dots, a_p$  und  $\sigma_{\varepsilon}^2$ .

Gilt zusätzlich, daß  $E\varepsilon_t^8 < \infty$ , dann konvergieren die Folgen (A) und (S) auch mit deterministischer Steuerfolge (1.32) mit Wahrscheinlichkeit Eins gegen die theoretischen Parameter.

Zum Beweis vgl. [28], Theoreme 2.2 bis 2.4 und die Folgerung 2.2.

**Satz 1.8:**

Sei  $X = \{x_t\}_{t=0}^{\infty}$  ein MA(1)-Prozeß, wobei  $E = \{\varepsilon_t\}_{t=0}^{\infty}$  eine Folge unabhängiger identisch verteilter Zufallsgrößen mit  $E\varepsilon_t = 0$ ,  $\text{Var } \varepsilon_t = \sigma_{\varepsilon}^2$  und  $E\varepsilon_t^8 < \infty$  ist. Dann konvergiert das Verfahren (E), (B) und (S) im Quadratmittel gegen die Parameter  $b_1$  und  $\sigma_{\varepsilon}^2$ , wobei die Folge  $\{c_t\}_{t=0}^{\infty}$  nach (1.34) berechnet wird.

Zum Beweis vgl. [28], Theorem 2.5.

Die zufällige Steuerfolge  $\{c_t\}$ , definiert in (1.34), ist mehr an die konkreten Eigenschaften der Zeitreihe angepaßt. Es erweist sich, daß sie in einem bestimmten Zusammenhang mit der deterministischen Steuerfolge  $c_t = \frac{c}{t}$  steht. Man kann nämlich zeigen, daß

$$c_t = (t (\hat{\sigma}_x^2(t)+c))^{-1} \tag{1.35}$$

gilt, wobei  $\hat{\sigma}_x^2(t)$  das zweite empirische Moment der Folge  $\{x_t\}_{t=0}^{\infty}$  ist und damit eine Schätzung der Varianz der zentrierten Zeitreihe. Aufgrund der Stationarität nähert sich  $\hat{\sigma}_x^2(t)$  einer Konstanten für genügend große  $t$ . Damit entspricht die Größe  $(\hat{\sigma}_x^2(t)+c)^{-1}$  näherungsweise dem konstanten Faktor im deterministischen Fall.

Mehrere Untersuchungen des Approximationsverfahrens an simulierten Daten und Biosignalen ergaben, daß die Verwendung der deterministischen Steuerfolge bezüglich der Konvergenzgeschwindigkeit und der Streuung der Parameterschätzungen nicht problemlos ist. In Abbildung 1.2 sind die Schätzfolgen des Gleitmittelparameters  $b_1$  eines MA(1)-Prozesses mit stochastischer und deterministischer Steuerfolge dargestellt.

### Simulationsergebnisse

Zur Güte und Konvergenzgeschwindigkeit der Schätzungen der ARMA-Parameter wurden konkrete Untersuchungen nach dem Verfahren (E), (A), (B) und (S) für simulierte Zeitreihen niedriger Ordnung ( $p = 1, 2$ ,  $q = 1$ ) durchgeführt (vgl. [28]). Für Zeitreihen aus Wirtschaftsprozessen erweist sich eine Modellierung mit AR-Prozessen mit niedriger Ordnung als ausreichend. Es ist aber bekannt, daß für die Analyse von Biosignalen AR-Modelle der Ordnung  $p \geq 6$  erforderlich sind. Aus diesem Grund wurde das Verfahren an einer Vielzahl simulierter Daten untersucht. Für die Untersuchungen wurden AR(p)-, MA(q)- und ARMA(p,q)-Folgen simuliert, wobei  $\varepsilon_0, \varepsilon_1, \dots$  als unabhängige gleichverteilte Pseudozufallszahlen mit dem Computern erzeugt wurden. Nach dem oben eingeführten Schätzalgorithmus wurden die entsprechenden Iterationen durchgeführt und nach 10, 100, 256 und 512 Schritten ausgewertet. Voruntersuchungen, deren Ergebnisse in den Abbildungen 1.1 und 1.2 demonstriert wurden, zeigen, daß das Verfahren für eine deterministische Steuerfolge (1.33) im allgemeinen zu langsam konvergiert. Daher wurde in allen Untersuchungen die stochastische Steuerfolge  $\{c_t\}$  nach (1.34) gewählt. Bei der Bewertung der Güte der Schätzung wurde die Abweichung der geschätzten Werte von den theoretischen durch den relativen Fehler

$$\alpha_t(\theta) = \left| \frac{\hat{\theta}(t) - \theta}{\theta} \right| \quad (1.36)$$

beschrieben, wobei  $\theta$  der wahre Parameter und  $\hat{\theta}(t)$  die Schätzung nach  $t$  Iterationsschritten sind.

Der relative Fehler (1.36) gestattet eine Aussage nur über die einzelnen Parameter, er vermittelt aber keine Information über die Schätzung der Spektraleigenschaften der Zeitreihe, die von Bedeutung sind. Als weiteres Kriterium zur Beschreibung der Anpassungsgüte wird also der Vergleich der auf den geschätzten Parametern beruhenden parametrischen Spektraldichte mit der theo-

retischen Spektraldichte und mit anderen spektralen Schätzmethoden angewandt. Die Schätzung der Spektraldichte mit Hilfe des Periodogramms ist in der Praxis sehr verbreitet und kann als eine Standardmethode betrachtet werden. Als eine zweite Vergleichsmethode wurde die Spektraldichtenschätzung aus der Anpassung von einem AR-Prozeß mit Hilfe des Levinson-Durbin-Algorithmus und mit der Ordnungsschätzung nach dem AIC-Kriterium gewählt. Diese Modellierung wurde in [3] ausführlich dargelegt und für die Biosignalanalyse untersucht. Eine Zusammenfassung der Ergebnisse wird im folgenden dargestellt. Die einzelnen Simulationsergebnisse sind in [6] detailliert aufgeführt.

#### Schätzung der Parameter von AR-Modellen

Das Schätzverfahren erwies sich bei den Simulationsuntersuchungen als sehr effizient. Oft konnte schon nach 100 Iterationsschritten die Güte der Schätzung als ausreichend betrachtet werden. Der relative Fehler lag nach 256 bzw. 512 Schritten unter 20%, oft unter 10%. Dieses Resultat wurde für sehr unterschiedlich gewählte Konstanten  $c$  in (1.34) und Streuungen des Prozesses  $E = \{\varepsilon_t\}_{t=0}^{\infty}$  ( $0.1 \leq \sigma_{\varepsilon}^2 \leq 10\,000$ ) erreicht. Die Untersuchungen wurden bis zur Ordnung 8 durchgeführt. An dieser Stelle wird das folgende Beispiel ausgeführt. Es wurde ein AR(7)-Prozeß mit den Parametern  $a_1 = -0.1$ ,  $a_2 = -0.2$ ,  $a_3 = 0.1$ ,  $a_4 = 0.05$ ,  $a_5 = -0.06$ ,  $a_6 = 0.2$ ,  $a_7 = -0.15$  und  $\sigma_{\varepsilon}^2 = 200$  simuliert. Die Ergebnisse der Schätzung sind in der nachstehenden Tabelle dargestellt:

t:	10	100	256	512
$\alpha(a_1)$ :	125.0%	12.5%	51.2%	8.9%
$\alpha(a_2)$ :	15.4%	44.9%	22.8%	7.2%
$\alpha(a_3)$ :	191.0%	36.0%	56.9%	25.5%
$\alpha(a_4)$ :	193.0%	74.0%	24.4%	0.6%
$\alpha(a_5)$ :	402.0%	22.3%	93.2%	3.2%
$\alpha(a_6)$ :	79.5%	7.5%	15.0%	16.0%
$\alpha(a_7)$ :	260.0%	20.0%	0.5%	10.2%
$\alpha(\sigma_{\varepsilon}^2)$ :	60.0%	24.0%	9.0%	11.0%

Tabelle 1.1: Relativer Fehler als Gütekriterium der rekursiven Schätzung von Parametern eines AR(7)-Prozesses

Zu diesem Beispiel ist ein Vergleich der Spektraldichteschätzungen in der Abbildung 1.3 dargestellt.

### Schätzung der Parameter von MA- und ARMA-Modellen

Für MA- bzw. ARMA-Modelle beliebiger MA-Ordnung  $q > 1$  ist das rekursive Verfahren nicht uneingeschränkt anwendbar. Die Konvergenz des Verfahrens mit der stochastischen Steuerfolge (1.34) wurde nur für  $q = 1$  gezeigt (Satz 1.8). Praktische Untersuchungen zeigen aber, daß die Schätzung mit der deterministischen Steuerfolge (1.33) stark von der Wahl der Konstanten  $c$  abhängt. Zusätzlich ist im Unterschied zu autoregressiven Parametern die Beschränkung der Folgen der zu schätzenden Parameter  $\{\hat{b}_k(t)\}_{k=0}^q$  für  $q > 1$  nicht allgemein durch Algorithmen regelbar, für die Konvergenz aber notwendig (vgl. [29]). Trotz dieser Einschränkungen wurden Simulationen für  $q > 1$  durchgeführt, um zu testen, ob sich für simulierte Daten die Folgen der Parameterschätzung sehr schnell im eingeschwungenen Zustand befinden. Die Untersuchungen brachten nur in den Fällen auswertbare Ergebnisse, in denen die Folge  $\{e_t\}$  ein schnelles Konvergenzverhalten zeigte.

Bei MA-Modellen erster Ordnung führten die rekursiven Schätzungen zu sehr guten Ergebnissen. Die Parameter waren schon nach 100 Iterationsschritten im eingeschwungenen Zustand. Nach 256 Iterationen lag der relative Fehler bei etwa 15%, nach 512 unter 10%. Für Modelle höherer Ordnung war der relative Fehler größer, vor allem für kleine Parameter ( $|b_k| < 0.1$ ). Wurde  $q = 2$  oder 3 gewählt, so lag der relative Fehler nach 256 Schritten bei etwa 45%, nach 512 bei 30%. Für höhere Ordnungen führte der Schätzfehler in der Spektraldichteschätzung zur Verschiebung lokaler Extrema, bzw. zur Entstehung falscher Extrema (vgl. [6]). Ein Beispiel ist in der folgenden Tabelle dargestellt.

n:	10	100	256	512
$\alpha(b_1)$ :	124.3%	26.7%	22.1%	9.9%
$\alpha(b_2)$ :	108.8%	62.2%	25.7%	31.5%
$\alpha(\sigma_\varepsilon^2)$ :	24.6%	2.5%	1.0%	5.0%

Tabelle 1.2: Relativer Fehler der Schätzung der Parameter eines MA(2)-Prozesses mit  $b_1 = 0.4$ ,  $b_2 = 0.3$  und  $\sigma_\varepsilon^2 = 0.1$ .

Bei der Schätzung der Parameter von simulierten ARMA-Prozessen waren die Schätzfolgen nach 512 Werten nicht unbedingt im eingeschwungenen Zustand und der relative Fehler war dann dementsprechend groß. Zum Vergleich der Spek-

traldichten wurde der simulierten Zeitreihe zusätzlich ein AR-Modell mit Hilfe des Levinson-Durbin-Algorithmus angepaßt, wobei die Ordnung  $p$  aus dem AIC-Kriterium ermittelt wurde. Die ARMA-Spektraldichte, die aus den nach dem Verfahren (E), (A), (B), (S) geschätzten Parametern bestimmt wurde, lag meistens näher an der theoretischen Spektraldichte als die AR-Spektraldichte, die aus der Levinson-Durbin-Schätzung berechnet wurde (vgl. [6]).

Die Ergebnisse der Untersuchungen rechtfertigen die Anwendung des rekursiven Algorithmus (E), (A), (B) und (S) zur Auswertung von Biosignalen. Konkrete Beispiele werden im nächsten Abschnitt dargestellt.

Es soll noch bemerkt werden, daß in den früheren Arbeiten von Günther der Satz 1.6 für die folgende Schätzung der MA-Parameter gezeigt wurde:

$$\begin{aligned} \hat{b}_j(t) &:= 0, \quad \text{für } t \leq j, \quad j = 1, \dots, q, \\ d &:= \hat{b}_j(t-1) - \frac{c}{t} e_t e_{t-j} \quad \text{und} \\ \hat{b}_j(t) &:= \begin{cases} d, & \text{wenn } \sum_{k=1}^{j-1} |\hat{b}_k(t)| + |d| + \sum_{k=j+1}^p |\hat{b}_k(t-1)| < 1 \\ \hat{b}_j(t-1), & \text{sonst} \end{cases}, \quad (B^*) \\ &\quad \text{für } t > j, \quad j = 1, \dots, q, \end{aligned}$$

Im Beweis wurde aber zusätzlich angenommen, daß die Summe der Absolutwerte der AR-Parameter kleiner als eins ist (siehe [30]). Ob diese Einschränkung von praktischer Bedeutung ist, kann aufgrund empirischer Untersuchungen entschieden werden. Die Schätzung läßt sich praktisch einfach realisieren. Soll sie aber bessere Ergebnisse bringen als die Schätzung (B), sind Voruntersuchungen zur Wahl der Konstanten  $c$  notwendig.

### 1.2.3 ARMA-Modellierung von Biosignalen

In der medizinischen Diagnostik und in der Therapieverlaufskontrolle werden Biosignale unterschiedlichster Art wie z.B. EEG-, EKG- oder EMG-Signale aufgenommen und im Zeit- und Spektralbereich ausgewertet. Die statistische Analyse solcher Signale über die Anpassung geeigneter linearer parametrischer Modelle ist auf medizinischem Gebiet weit verbreitet. Biosignale, deren Lei-

stungsverteilung sich mit einer stetigen Spektraldichte charakterisieren läßt, können mit Hilfe von ARMA-Prozessen modelliert werden. Nach dem Satz 1.5 können stetige Spektraldichten mit der rationalen ARMA-Spektraldichtefunktion beliebig gut approximiert werden. Ein wesentlicher Vorteil dieser Modellierung besteht auch darin, daß ein ARMA-Prozeß durch wenige Parameter beschrieben werden kann. Sie erweisen sich dann als sehr günstig, wenn für die vorliegende Zeitreihe die geschätzten Parameter zu weiteren Untersuchungen bzw. zur Entwicklung neuer Kenngrößen dienen können. Die parametrische Darstellung der Spektraldichte konnte zum Beispiel sehr effizient zur Konstruktion eines Wiener Optimalfilters zur Trennung eines evozierten Potentials und der Hintergrundaktivität (eines EEG-Signals) genutzt werden. Die spezifischen Eigenschaften einer ARMA-Spektraldichte, vor allem die Stetigkeit und strenge Positivität, sicherten die Stabilität des Filters (vgl. [7]). Häufig sind spektrale Kenngrößen, wie z.B. Effektivwerte bestimmter Frequenzbänder, von einem besonderen diagnostischen Interesse. In der medizinischen Diagnostik kommen zunehmend Mappingverfahren zum Einsatz, die die topographische Verteilung bestimmter charakteristischer Parameter im Bild wiedergeben. Zur Auswertung des EEG und EMG werden z.B. Spektralleistungsmaps genutzt, die die räumliche Verteilung der Hirn- bzw. Muskelaktivität darstellen. Diese Maps werden aus den Leistungen innerhalb eines Frequenzbandes an den verschiedenen Elektroden berechnet (vgl. z.B. [72], [73], [74], [25]).

Die Anpassungsgüte der Modellierung eines realen Signals hängt von den Eigenschaften der gewählten Schätzverfahren ab. Die untersuchten Signale müssen aber die Modellvoraussetzungen erfüllen, damit die geschätzten Kenngrößen die Signaleigenschaften sachgerecht widerspiegeln. Um ausreichende Ergebnisse zu erzielen, ist also eine Datenvoruntersuchung und -vorverarbeitung notwendig. Im vorliegenden Abschnitt werden dazu einige wesentliche theoretische Erkenntnisse und praktische Anwendungen für Biosignale dargestellt.

Die wichtigste Kenngröße, die mit Hilfe der geschätzten Parameter bestimmt wird, ist die Spektraldichte, welche die Leistungsverteilung des Signals widerspiegelt. Die Spektraldichte eines ARMA-Prozesses ist durch die Funktion (1.18) gegeben. Die analytischen Eigenschaften dieser Funktion stehen im engen Zusammenhang mit einigen Eigenschaften des Prozesses im Zeitbereich und sollen näher untersucht und dargestellt werden. Die Spektraldichte  $f$  ist eine

stetige reellwertige rationale Funktion. Das beinhaltet, daß ARMA-Prozesse nicht aus rein periodischen Anteilen bestehen. Nach dem zweiten Theorem von Weierstraß läßt sich jede stetige Spektraldichte beliebig gut durch eine ARMA-Spektraldichte approximieren (vgl. Satz 1.5). Aufgrund der Invertierbarkeitsbedingung (1.15) ist die Spektraldichte  $f$  echt positiv, also in keinem Frequenzpunkt ist sie gleich Null. Für eine praktische Realisierung ist die folgende Darstellung, die mit (1.18) äquivalent ist, von Nutzen.

**Lemma 1.2:**

Sei  $f$  die Spektraldichte eines ARMA( $p,q$ )-Prozesses mit den Parametern  $(1, a_1, \dots, a_p; \sigma_\varepsilon^2)$  und  $(1, b_1, \dots, b_q)$ ,  $0 \leq p, q < \infty$ . Dann ist für  $f$  die folgende Darstellung richtig:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{\sum_{k=0}^q b_k^\circ \cos k\lambda}{\sum_{k=0}^p a_k^\circ \cos k\lambda}, \quad \lambda \in [-\pi, \pi) \quad (1.38)$$

wobei 
$$b_0^\circ = 1 + \sum_{k=1}^q b_k^2, \quad (1.39)$$

$$b_k^\circ = 2 \left( b_k + \sum_{j=1}^{q-k} b_j b_{j+k} \right), \quad k = 1, \dots, q$$

und 
$$a_0^\circ = 1 + \sum_{k=1}^p a_k^2, \quad (1.40)$$

$$a_k^\circ = 2 \left( a_k + \sum_{j=1}^{p-k} a_j a_{j+k} \right), \quad k = 1, \dots, p$$

sind.

Vgl. auch [49] oder [45], zum Beweis s. z.B. [9].

Von besonderem Interesse sind die lokalen Extrema der Funktion, da sie die Frequenzen mit maximalem bzw. minimalem Leistungsanteil des ARMA-Prozesses beschreiben. Die maximal mögliche Anzahl der Extrema ist mit den Werten von  $p$  und  $q$  gegeben. Sie kann z.B. aus folgenden Betrachtungen bestimmt werden.

Die lokalen Extrema der Funktion  $f$  werden durch die Nullstellen ihrer ersten Ableitung beschrieben. Man führt folgende Bezeichnung für die Darstellung (1.38) ein:

$$f(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{B(\lambda)}{A(\lambda)}, \quad \lambda \in [-\pi, \pi),$$

mit  $A(\lambda) = \sum_{k=0}^q a_k^\circ \cos k\lambda$  und  $B(\lambda) = \sum_{k=0}^p b_k^\circ \cos k\lambda$ . Damit gilt für die erste Ableitung der Funktion  $f$ :

$$f'(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \frac{A(\lambda)B'(\lambda) - A'(\lambda)B(\lambda)}{[A(\lambda)]^2}, \quad \lambda \in [-\pi, \pi). \quad (1.41)$$

Nach einer Umformung ergibt sich

$$A(\lambda)B'(\lambda) - A'(\lambda)B(\lambda) = \sum_{k=0}^p \sum_{l=0}^q \left( 1 a_k^\circ b_l^\circ \sin(k-l)\lambda + (k-l) a_k^\circ b_l^\circ \sin k\lambda \cos l\lambda \right). \quad (1.42)$$

In [9] wurde gezeigt, daß (1.42) im Intervall  $[0, \pi)$  höchstens  $(p+q)$  Nullstellen besitzt. Im Fall  $p = q$  sind es höchstens  $(2p-1)$ . Wegen der Symmetrie der Funktion sind es im Intervall  $[-\pi, \pi)$   $2(p+q)$  bzw.  $2(2p-1)$ , wobei im Nullpunkt immer ein Extremum liegt. Somit ist die maximale Anzahl der Extrema einer ARMA(p,q)-Spektraldichte mit  $(p+q)$  und im Fall  $p = q$  mit  $(2p-1)$  beschränkt. Aus der Darstellung der Funktion  $f$  lassen sich für höhere Ordnungen ( $p + q > 3$ ) keine allgemeinen Aussagen über Art und Lage der lokalen Extrema (Minimum, Maximum oder Wendepunkt) machen, der Zusammenhang zwischen der Modellordnung und der möglichen Gestalt der Spektraldichte wurde aber verdeutlicht. Eine Zerlegung der Spektraldichtefunktion in ihre wesentlichen Anteile wurde mit Hilfe der Zerlegung der Kovarianzfunktion in [76] durchgeführt und die Anwendung in der Analyse von EEG-Signalen in [68] untersucht.

### Einige Untersuchungsbeispiele

In den folgenden Beispielen werden angepaßte parametrische Spektraldichteschätzungen unterschiedlicher Biosignale (EEG, EMG) mit den entsprechenden geglätteten Periodogrammen verglichen. Dafür wurden Datensätze mit 512 Werten untersucht. Es wurden zur Schätzung der Parameter das Verfahren (E), (A), (B) und (S) und der Levinson-Durbin-Algorithmus durchgeführt. Es wird weiter an ausgewählten Beispielen demonstriert, welche Fehler auftreten können, wenn die untersuchten Signale die Modellvoraussetzungen nicht erfüllen.

Als erstes betrachtet man EEG-Signale verschiedener Elektrodenpositionen, die mit einer Abtastrate von 100 Hz aufgenommen wurden. Die Gehirnaktivität kann vor allem im Frequenzbereich von 1.5 Hz bis 20 bzw. 25 Hz nachgewiesen

werden. In der Abbildung 1.4 wird gezeigt, daß die parametrische Spektraldichteschätzung die gleichen inhaltlichen Interpretationen zuläßt wie die Schätzung der Leistungsdichte über die Glättung des Periodogramms. Die lokalen Extrema der Spektraldichtefunktion beschreiben in analoger Weise die charakteristischen Aktivitäten im EEG-Bereich. Der Frequenzbereich von 1.5 Hz bis 20 bzw. 25 Hz wird zur EEG-Analyse gewöhnlich in vier Spektralbänder unterteilt, in denen sich die charakteristischen Vorgänge widerspiegeln. Die Abbildungen 1.5 und 1.6 zeigen das  $\alpha$ -Band (8-13 Hz) an der Elektrode O1, wobei das EEG mit offenen und geschlossenen Augen jeweils in Ruhesituation aufgenommen wurde.

In den folgenden Beispielen werden EMG-Signale analysiert, die die Kaumuskelaktivität (M.masseter) beim Kauen beschreiben. In der Abbildung 1.7 ist eine EMG-Registrierung dargestellt, die keine typische Muskelaktivität enthält und unter anderem dadurch gekennzeichnet ist, daß alle Frequenzen relativ gleichmäßig auftreten. Die Schätzung der Ordnung des angepaßten AR-Prozesses ergab niedrige Werte,  $p = 2$ . Im Gegensatz dazu liegt im Kanal 15 in den Frequenzbändern von 55 Hz bis 98 Hz und von 105 Hz bis 148 Hz eine charakteristische Muskelaktivität vor, die sowohl im Periodogramm als auch im AR-Spektrum deutlich erkennbar ist. Die Schätzung der AR-Ordnung erhöhte sich auf  $p = 15$  (siehe Abb. 1.8). Das entspricht den theoretischen Betrachtungen des Zusammenhangs zwischen der Modellordnung und der möglichen Gestalt der Spektraldichtefunktion.

In Abbildung 1.9 sind zwei EMG-Maps dargestellt, die zum einen aus der Schätzung der Leistung über das Periodogramm und zum anderen über die angepaßte AR-Spektraldichte resultieren. Beide Verfahren erbringen die gleichen inhaltlichen Ergebnisse, die Maps sind optisch fast identisch. Zu bemerken ist dabei, daß die Berechnung der parametrischen Bandleistung direkt mit den AR-Parametern erfolgt und damit eine enorme Datenreduktion erzielt werden kann.

Um Auswertergebnisse in der oben demonstrierten Qualität zu erreichen, müssen die aufgenommenen Daten sachgerecht vorverarbeitet werden. Die Vorverarbeitung der gemessenen Signale ist notwendig, um sie als Realisierung eines stationären Prozesses mit stetiger Spektraldichte betrachten zu können. Als Unstetigkeit wirkt sich im Bereich der niederfrequenten Wellen des Spektrums eine Nulllinienverschiebung (Unstetigkeit in der Frequenz Null) sowie ein Trend im Signal (Unstetigkeit in der Frequenz  $\lambda_1$ ) aus. Die Aktivität in die-

sen Frequenzen ist von keiner diagnostischen Bedeutung und kann auf Meßfehler zurückgeführt werden. Sie kann mit Hilfe einfacher Verfahren (wie z.B. Zentrierung, exponentielle Glättung) eliminiert werden. Werden diese Fehler nicht eliminiert, so können die Verfahren eine falsche Schätzung liefern. So konvergiert z.B. im Levinson-Durbin-Algorithmus die Schätzung  $R(0,N)$  bei rein periodischen Signalen gegen Null. Damit wird das Gleichungssystem instabil (vgl. Abbildung 1.10). Wird ein Signal mit stetigem Spektrum mit einem Signal mit diskretem Spektrum (z.B. einem Trend) überlagert, so erhöht sich die geschätzte Modellordnung wesentlich. Im Verfahren (E), (A), (B) und (S) wird die geschätzte Streuung des mittleren quadratischen Fehlers zu groß (siehe Abbildung 1.11 - 1.13).

Im folgenden Beispiel wird die Notwendigkeit einer sachgerechten Anpassung demonstriert, um eine Fortpflanzung der inhaltlichen Fehler und eine damit verbundene Verfälschung der Endlösung zu vermeiden. Von diagnostischem Interesse sind Erhöhungen der Amplituden in langwelligen Bereichen des EEG ( $\delta$ - und  $\theta$ -Frequenzbereich des EEG-Spektrums), die an eine gleichzeitige Verminderung der höherfrequenten Bestandteile ( $\alpha$ -,  $\beta$ -Frequenzbereich) gekoppelt sind. Eine solche dominante  $\delta$ - oder  $\theta$ - Aktivität läßt sich z.B. bei Hirntumoren nachweisen. Eine Lokalisierung wird über die topographische Darstellung mittels Spektralmapping ermöglicht. Die Abbildungen 1.14 und 1.15 zeigen das Spektralmap im  $\delta$ -Band, berechnet aus dem Periodogramm bzw. über eine ARMA-Anpassung, in dem ein Tumor analog zum CT-Bild um die Elektroden C4 und P4 lokalisiert werden kann. Bei der ursprünglichen Messung ist im Kanal 14 (Elektrode T6) ein Trend deutlich zu erkennen (siehe Abbildung 1.16). Im Spektrum ist er als Peak in der Frequenz  $\lambda_1$  deutlich. Verzichtet man in diesem Signal auf eine Trendeliminierung, so kann dies zur Verdeckung der diagnostisch wichtigen Aktivität im  $\delta$ -Band an den benachbarten Elektroden führen, wie es aus den Leistungsmaps in den Abbildungen 1.17 (Schätzung über das Periodogramm) und 1.18 (Berechnung der Spektraldichte aus geschätzten AR-Parametern) deutlich wird. Da ARMA-Modelle eine stetige Spektraldichte approximieren, kann die Leistung dieses Trends ins  $\delta$ -Band "durchsickern", auch wenn der Peak außerhalb des Leistungsbands liegt. Bei einer Schätzung der Spektraldichte mit dem Periodogramm werden Glättungsverfahren angewandt (um bessere Ergebnisse zu erreichen), was bei dieser Schätzmethode zu gleichen Effekten führen kann.

## 2 ADAPTIVE VERFAHREN

Eine Voraussetzung für die Nutzung von Methoden der stochastischen Approximation ist die starke Stationarität der Zeitreihen. Es ist aber bekannt, daß biologische Signale wie das EEG u.a. nicht stationär sind, sondern nur im Sprachgebrauch der angewandten Ingenieurwissenschaften stationäre Abschnitte oder Epochen besitzen. Das Auftreten derartiger Instationaritäten, wie z.B. die Veränderung der Frequenzband- oder Gesamtleistung des Signals oder der Wechsel von langsamen und schnellen Wellen, führt zur Verfälschung der konvergenten Anpassungsverfahren. Somit entsteht die Frage, ob es möglich ist, unter Verzicht auf die Konvergenz Anpassungsverfahren zu entwickeln, die die inhaltlichen Sachverhalte richtig widerspiegeln, gleichzeitig auf qualitative und quantitative Veränderungen der Zeitreihe reagieren und die Schätzung der neuen Situation anpassen.

In [22] wurde gezeigt, daß die aus der stochastischen Approximation abgeleiteten rekursiven Algorithmen sich mittels neuer Wichtung der Korrekturterme zu günstigen robusten Verfahren entwickeln lassen, die nicht mehr an die Stationarität der Zeitreihen gebunden sind. Durch die Verwendung einer konstanten zeitunabhängigen Steuerfolge wird erreicht, daß die Schätzfolgen sich schnell einer Änderung des gesuchten Parameters anpassen. In diesem Kapitel wird diese Konstruktionsweise beschrieben, die bekannten Verfahren darstellt und ihre Eigenschaften in Abhängigkeit von der Steuerkonstanten untersucht. Es werden weiter für die einzelnen Schätzoperatoren Steuerfolgen konstruiert, welche das dynamische Verhalten der untersuchten Zeitreihe berücksichtigen und in die Schätzung eingehen lassen. Dafür werden Verfahren zur Beschleunigung der Konvergenz im konsistenten Fall betrachtet (vgl. [38]) und für die adaptiven Schätzungen modifiziert. Ausgehend von den Untersuchungen aus dem vorangegangenen Abschnitt wird eine adaptive Schätzung der Parameter eines ARMA-Modells konstruiert und ihre Eigenschaften untersucht. Bei der Bestimmung ihrer Steuerfolge werden die Betrachtungen zur Konvergenz im konsistenten Fall (Satz 1.5) berücksichtigt. Für die Steuerung der adaptiven Quantilschätzung werden Ansätze numerischer Verfahren sowie ein Satz über die Konstruktion der wirksamsten Schätzungen aus [47] angewandt. Zuerst werden zwei konsistente Verfahren mit einer für jedes der Verfahren charakteristi-

schen zeit- und signalabhängigen Steuerfolge konstruiert. Sie können unter Verzicht auf die Zeitabhängigkeit der Steuerfolge zu selbstadaptierenden Operatoren überführt werden.

## 2.1 Wahl der Steuerfolge in rekursiven Schätzverfahren

In der Rekursionsvorschrift aus dem Satz 1.1:

$$x_t = x_{t-1} + c_{t-1} y_t(x_{t-1}, \xi_t(\omega))$$

bildet die Folge  $Y = \{y_t\}$ , die durch

$$y_t(x_{t-1}, \omega) = R(x_{t-1}) + G(t, x_{t-1}, \omega)$$

definiert ist, einen sogenannten Korrekturterm. Sein Einfluß auf die Schätzfolge wird durch die Folge  $\{c_t\}$  bestimmt, die Steuerfolge genannt wird.

Man betrachte den Fall, wenn

$$y_t(x_{t-1}, \omega) = f(\xi_t(\omega)) - x_{t-1} \quad (2.1)$$

ist, wobei  $f$  eine meßbare Funktion, z.B.  $f(\xi_t^k) = \xi_t^k$ ,  $k \in \mathbb{N}$ , bezeichnet. Dann läßt sich die geschätzte Folge in geschlossener Form darstellen:

$$x_t = x_0 \prod_{j=0}^{t-1} (1-c_j) + \sum_{i=1}^{t-1} c_{i-1} \left( \prod_{j=k}^{t-1} (1-c_j) \right) f(\xi_i) + c_{t-1} f(\xi_t). \quad (2.2)$$

Für  $c_t = c/(t+1)$ ,  $c > 0$ , vereinfacht sich die Formel zu

$$x_t = x_0 \prod_{j=0}^{t-1} \frac{j+1-c}{j+1} + \sum_{i=1}^{t-1} \frac{c}{i} \prod_{j=k}^{t-1} \frac{j+1-c}{j+1} f(\xi_i) + \frac{c}{t} f(\xi_t). \quad (2.3)$$

Ist  $c = 1$ , so liefert die Schätzung die wohlbekannte empirische Größe

$$x_t = \frac{1}{t} \sum_{i=1}^t f(\xi_i), \quad (2.4)$$

die für mathematische Stichproben die wirksamste Schätzung der Größe  $Ef(\xi_t)$  ist.

Für den Fall  $c_t = c/(t+1)$  kann man beobachten, daß der Korrekturterm einen immer kleineren Einfluß auf die Schätzung hat und damit die Folge langsam gegen den theoretischen Parameter "zieht" und wenig Dynamik zeigt. Bei schlechten Startwerten ist eine lange Stichprobe notwendig, um in die Nähe des korrekten Wertes zu kommen.

Eine praktische Verbesserung der Konvergenz wurde von Kesten [38] vorgeschlagen und die Beibehaltung der starken Konsistenz der Schätzung bei der Beschleunigung bewiesen. Dabei wird die Steuerfolge nach der folgenden Vorschrift gestaltet:

$$c_t = c/k(t), \quad t \in \mathbb{N},$$

wobei

$$k(1) = 1, \quad k(2) = 2 \tag{2.5}$$

und

$$k(t) = 2 + \sum_{i=1}^t |_{(-\infty, 0)} \left( (x_i - x_{i-1}) (x_{i-1} - x_{i-2}) \right).$$

Inhaltlich bedeutet das, daß  $c_t$  nur dann verkleinert wird, wenn sich in der Folge  $x_t$  die "Korrekturrichtung" verändert.

Die Vorteile der rekursiven Gestalt der Schätzfunktionen liegen in der rechen-technischen Umsetzung der Verfahren. Mittels einfacher Rechenoperationen wird aus wenigen Daten (der letzte Schätz- und der aktuelle Wert) die Schätzung in jedem Punkt berechnet. Die einzige aufwendige Operation ist die Division. Von Griebach [22] wurde vorgeschlagen, sie durch eine angepaßte Wahl der Steuerfolge zu umgehen, wobei die Konsistenz der Schätzung nicht verletzt wird. Sei

$$c_t^p = 2^{-\min(k, 2^{k-p} \langle t \rangle)}, \quad p \in \mathbb{N}. \tag{2.6}$$

Das heißt z.B. für  $p = 0$

$$\begin{aligned} c_1^0 &= 1 \\ c_2^0 &= c_3^0 = 2^{-1} \\ c_4^0 &= c_5^0 = c_6^0 = c_7^0 = 2^{-2} \quad \text{etc.} \end{aligned}$$

Es gilt unter anderem, daß  $c_{2^k}^0 = c_{2^k}^0$  ist, wobei  $c_t^0 = 1/t$ . Damit wird die allgemeine Division durch die Division mit Zweierpotenzen ersetzt, die sich rechen-technisch mit einer Bitrotation realisieren läßt. Eine Beschleunigung der Konvergenz durch (2.5) ist zusätzlich möglich.

Aus den obigen Betrachtungen ist ersichtlich, daß die Qualität der Schätzung wesentlich von der Gestalt der Steuerfolge abhängt. So entsteht die Frage, ob sich über eine bestimmte Wahl der Folge  $\{c_t\}$  ein robuster Algorithmus entwickeln läßt, der auf die Änderungen in der Zeitreihe reagiert und die inhaltlichen Sachverhalte richtig widerspiegelt. Diese Idee wurde von Griebach [22] untersucht und es wurde vorgeschlagen, die Steuerfolge  $c_t = c$  konstant

zu lassen. Eine Konvergenz im mathematischen Sinne liegt für diese Schätzung nicht mehr vor, es können jedoch einige nützliche Eigenschaften einer solchen Schätzung beobachtet werden. Wird die rekursive Schätzung mit einer konstanten Steuerfolge für eine simulierte Zeitreihe mit bekanntem Wert des gesuchten Parameters realisiert, so schwankt die Schätzfolge um den theoretischen Wert des Parameters, ihre Streuung bleibt aber konstant. Wird in der simulierten Zeitreihe der Wert des geschätzten Parameters geändert, konvergiert die Schätzfolge gegen den neuen Wert. Die Streuung der Schätzfolge nimmt mit wachsendem  $c$  zu, die Adaptionsphase nach einer Änderung des Parameterwertes wird aber kürzer. Dieses Verhalten wurde in [22] für die Schätzung des Erwartungswerts auch exakt beschrieben (vgl. Satz 3.3.8). Es kann für den allgemeineren Fall, wenn sich der Korrekturterm  $y_t$  nach (2.1) darstellen läßt, ebenfalls gezeigt werden. Betrachtet man die Formel (2.2) für  $c_t = c = \text{const.}$ , so enthält sie die Form

$$x_t = (1-c)^t x_0 + \sum_{j=1}^t c(1-c)^{t-j} f(\xi_j). \quad (2.7)$$

Analog zu [22] Satz 3.3.8 läßt sich zeigen, daß  $x_t$  eine asymptotisch erwartungstreue (für  $x_0 = f(\xi_0)$  erwartungstreue) Schätzung der Funktion  $Ef(\xi_j)$  ist, wenn  $0 < c \leq 1$  ist:

$$Ex_t = (1-c)^t Ex_0 + (1 - (1-c)^t) Ef(\xi_j).$$

Ihre Varianz ist beschränkt. Man betrachte folgende Spezialfälle.

**Satz 2.1:**

Sei  $\{\xi_t\}_{t=0}^{\infty}$  eine mathematische Stichprobe und  $f$  eine meßbare Funktion mit  $\text{Var}(f(\xi_t)) < \infty$ . Dann konvergiert die Größe  $E(x_t - Ef(\xi_t))^2$  für  $0 < c < 1$  und für jeden Startwert  $x_0$ , der unabhängig von der Stichprobe ist, und es gilt

$$\lim_{t \rightarrow \infty} E(x_t - Ef(\xi_t))^2 = \frac{c}{2-c} \text{Var}(f(\xi_t)).$$

Beweis:

Es gilt, daß

$$(1-c)^t + \sum_{j=1}^t c(1-c)^{t-j} = 1$$

ist. Die Größe  $Ef(\xi_1)$  kann also als

$$E f(\xi_1) = (1-c)^t E f(\xi_1) + \sum_{j=1}^t c(1-c)^{t-j} E f(\xi_j)$$

dargestellt werden. Nach (2.7) kann gezeigt werden, daß

$$\begin{aligned} E(x_t - E f(\xi_1))^2 &= E \left[ (1-c)^t x_0 - E f(\xi_1) + \sum_{j=1}^t c(1-c)^{t-j} f(\xi_j) \right]^2 \\ &= E \left[ (1-c)^t (x_0 - E f(\xi_1)) + \sum_{j=1}^t c(1-c)^{t-j} (f(\xi_j) - E f(\xi_j)) \right]^2 \end{aligned}$$

ist. Wegen Unabhängigkeit gilt

$$\begin{aligned} E(x_t - E f(\xi_1))^2 &= (1-c)^{2t} E(x_0 - E f(\xi_1))^2 + \sum_{j=1}^t c^2 (1-c)^{2(t-j)} (f(\xi_j) - E f(\xi_j))^2 \\ &= (1-c)^{2t} (x_0 - E f(\xi_1))^2 + c \frac{1 - (1-c)^{2t}}{2 - c} \text{Var } f(\xi_1). \end{aligned}$$

Daraus folgt  $\lim_{t \rightarrow \infty} E(x_t - E f(\xi_1))^2 = \frac{c}{2-c} \text{Var}(f(\xi_1)).$  □

Für stark mischende Prozesse lassen sich ähnliche Abschätzungen vermuten, deren Herleitung jedoch einer späteren Arbeit vorbehalten werden soll.

Bemerkung:

Für  $f(\xi_j) = \xi_j^k$  mit  $E \xi_j^k = \mu_k$  und  $E \xi_j^{2k} = \mu_{2k} < \infty$  ist

$$\lim_{t \rightarrow \infty} E(x_t - \mu_k)^2 = \frac{c}{2-c} (\mu_{2k} - \mu_k^2).$$

Ist für  $k = 1$  die Varianz  $\sigma_\xi^2$  von  $\xi_t$  bekannt, so gilt

$$\lim_{t \rightarrow \infty} E(x_t - \mu_\xi)^2 = \frac{c}{2-c} \sigma_\xi^2$$

(siehe auch [22], Satz 3.3.8).

Aus dem obigen Beweis ist ersichtlich, daß nach einer Einschwingphase der Einfluß des Startwertes  $x_0$  exponentiell abnimmt und die Schätzfolge  $\{x_t\}$  um den Wert des Parameters  $E f(\xi_1)$  mit einer endlichen positiven Streuung schwingt. Nach einem Strukturbruch liegt die gleiche Situation vor, wie bei einer ungünstigen Wahl des Anfangswertes, also der letzte Schätzwert ist ein

schlechter Startwert für den neuen Abschnitt, aber er beeinflusst die Schätzung nur in einem beschränkten Zeitintervall. Die Länge dieses Intervalls und die Streuung der Schätzung hängen von der Wahl der Konstanten  $c$  ab. Diese Konstante wird im weiteren als Adaptionfaktor bezeichnet. Für andere Verfahren (z.B. die Quantilwertschätzung) kann ein solches Verhalten aus der folgenden Gleichung erklärt werden. Für die Folge  $\{x_t\}$  aus dem Satz 1.1 gilt aufgrund von (1.5)

$$E(x_t - x_{t-1} | x_{t-1} = x) = c_t R(x) \begin{cases} \geq 0, & x \leq r \\ < 0, & x > r \end{cases},$$

wenn  $c_t > 0$  ist. Für  $c_t = c = \text{const.}$  schwankt die Schätzfolge  $\{x_t\}$  im Mittel um den gesuchten Wert  $r$ , bei seiner Änderung zieht sie mit einer konstanten Schrittlänge gegen den neuen Wert.

Mit der Wahl einer konstanten Steuerfolge in den rekursiven Algorithmen wurde ein Konstruktionsprinzip definiert, mit dessen Hilfe Verfahren zur Bestimmung verschiedener Kenngrößen von Zeitreihen entwickelt werden können, wobei die Zeitreihen nicht unbedingt die strengen Voraussetzungen der Sätze der stochastischen Approximation erfüllen müssen. Sie werden adaptive rekursive Verfahren genannt. Die einzelnen Verfahren und ihre Eigenschaften wurden z.B. in [23], [24], [26], [52], [55] u.a. beschrieben. Sie werden in den folgenden Abschnitten dieses Kapitels dargestellt, soweit sie für diese Arbeit von Bedeutung sind.

In allen Untersuchungen der Eigenschaften der adaptiven Verfahren und deren Einsatzmöglichkeiten in der Signalanalyse ist es von Bedeutung, den Einfluß des Adaptionfaktors zu beschreiben und Kriterien für seine Wahl zu bestimmen. Diese sind meistens für die einzelnen Verfahren spezifisch. In diesem Abschnitt werden zuerst allgemeine Hinweise für Verfahren mit dem Korrekturterm  $y_t$  nach (2.1) untersucht. Beim Vergleich der Formeln (2.4) und (2.7) wird deutlich, daß im konvergenten Fall jede Prozeßkomponente  $\xi_k$ ,  $k < t$ , in der Schätzung das gleiche Gewicht  $1/t$  besitzt. Dagegen nehmen in dem adaptiven Verfahren die Gewichte exponentiell ab, der aktuelle Wert hat den stärksten Einfluß und die "weit zurückliegenden" Ereignisse beeinflussen die Schätzung nicht mehr.

Sei  $x_0 = 0$ . Man betrachte eine positive reelle Zahl  $\vartheta \approx 0$  als untere Schranke für das Gewicht  $c(1-c)^k$ , es sei z.B.  $\vartheta$  die kleinste echt positive auf dem Rechner codierbare Zahlendarstellung. Wenn  $k_0$  als

$$k_0 = \min \{k \in \mathbb{N}, c(1-c)^k < \vartheta\}$$

definiert wird, dann ist bei einer rechentechnischen Umsetzung der Verfahren  $(t-k_0)$  der Zeitpunkt, der zur Zeit  $t$  die Schätzung nicht mehr beeinflusst. Es gilt, daß  $k_0$  die kleinste natürliche Zahl bestimmt, die größer als

$$B = \log_{1-c} (\vartheta/c)$$

ist. Sei  $\vartheta$  eine linksbündige Festkommazahl, d.h.  $\vartheta = 2^{1-p}$ , wobei  $p$  die Bitanzahl ist. Dann gilt für  $B$

$$B = \frac{1-p - \log_2 c}{\log_2 (1-c)} \quad (2.8)$$

Für den Spezialfall, daß  $c = 2^{-d}$ , mit  $d < p$ , eine Zweierpotenz ist, vereinfacht sich (2.8) zu

$$B = \frac{1-p + d}{\log_2 (2^d - 1) - d}$$

c	1/2	1/4	1/8	1/16	1/32	1/64	1/128
$k_0$	15	32	63	119	219	397	707

Tabelle 2.1: Bestimmung von  $k_0$  für verschiedene Werte von  $c$  bei  $p = 16$ .

Im Regelfall sind der Grad der Adaptivität und das Konvergenzverhalten der rekursiven Verfahren zwei sich gegenläufig beeinflussende Faktoren. Somit gestaltet man mit der konvergenten und der konstanten Steuerfolge zwei verschiedene Verfahren. Die obigen Untersuchungen machen deutlich, daß ein einmal festgelegter Adaptionsfaktor entweder die Adaptivität oder die Streuungseigenschaften des Verfahrens verbessert. Erstrebenswert wäre aber ein adaptives Verfahren, das in Einschwing Situationen bzw. beim Übergang von einem stationären Zustand zum nächsten stationären Zustand eine schnelle Anpassung bietet und innerhalb einer längeren stationären Phase die Schwingungen der Schätzfolgenfunktionen dämpft. Dafür kann aber die Steuerfolge nicht konstant sein, sondern sie muß sich in Abhängigkeit von den dynamischen Eigenschaften des Signals und der abgeleiteten Schätzfolgenfunktion in entsprechender Weise verändern.

Unter diesem Gesichtspunkt betrachtet, ist folgende Bestimmung der Folge  $\{c_t\}_{t=0}^{\infty}$  sinnvoll. Seien  $c^1$  und  $c^2$  zwei Konstanten mit

$$0 < c^1 \leq c^2 < 1$$

und

$$c^2 = c^1 + kc, \quad c \in \mathbb{R}, k \in \mathbb{N}.$$

Weiter werden zwei Zählfolgen definiert. In der Funktion  $r$ , definiert durch

$$r(t) = \sum_{j=1}^{z_1} I_{(-\infty, 0)} \left( (x_{t-j} - x_{t-j-1}) (x_{t-j-1} - x_{t-j-2}) \right),$$

wird die Anzahl der Richtungswechsel in der Schätzung in den letzten  $z_1$  Iterationsschritten aufsummiert. Dagegen zählt

$$q(t) = \sum_{j=1}^{z_2} I_{(0, \infty)} \left( (x_{t-j} - x_{t-j-1}) (x_{t-j-1} - x_{t-j-2}) \right),$$

wie oft in den letzten  $z_2$  Schritten kein Richtungswechsel der Schätzfolge vorkam. Dann wird  $\{c_t\}_{t=0}^{\infty}$  wie folgt berechnet:

$$c_0 = c_1 = c_2 = \dots = c_{\max\{z_1, z_2\}} = (c^1 + c^2)/2$$

und für  $t > \max\{z_1, z_2\}$

$$c_t = \begin{cases} c_{t-1} - c, & \text{wenn } r(t) = z_1 \text{ und } c_{t-1} > c^1 \\ c_{t-1} + c, & \text{wenn } q(t) = z_2 \text{ und } c_{t-1} < c^2 \\ c_{t-1}, & \text{sonst.} \end{cases} \quad (2.9)$$

Inhaltlich kann der Algorithmus so gedeutet werden, daß die Steuerfolge sich um den Wert  $c$  verkleinert, wenn die Schätzfolge  $z_1$ -mal hintereinander die Korrekturrichtung gewechselt hat, und sich um  $c$  vergrößert, wenn sie in den letzten  $z_2$ -Schritten die gleiche Korrekturrichtung behielt. Die Konstanten  $c^1$  und  $c^2$  drücken den Grad der Adaptivität ( $c^2$ ) bzw. der Streuung ( $c^1$ ) des Verfahrens aus. Sie sind aus Voruntersuchungen zu ermitteln, ebenfalls die Zahlen  $z_1$ ,  $z_2$  und  $c$ .

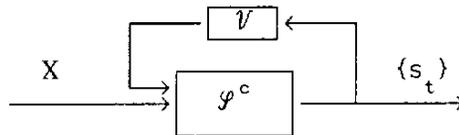
Da in stationären Abschnitten die Steuergröße  $c_t$  immer kleiner wird, kann eine statistisch gute Schätzung angestrebt werden. Bei einem Strukturbruch liegt die gleiche Situation vor, wie bei einer ungünstigen Wahl des Anfangswertes. Da die Steuerfolge nicht nur konstant bleibt, sondern in solchen Abschnitten auch zunehmen kann, hängt die Schätzung für den folgenden stationären Abschnitt kaum von der Vergangenheit ab. Offensichtlich liegt auch hier keine Konsistenz im mathematischen Sinne vor, dennoch werden die Lernfähigkeit und die Anpassung des Verfahrens an die konkrete Zeitreihe begünstigt.

## 2.2 Beschreibung und Eigenschaften adaptiver rekursiver Algorithmen

Eine Wahl der konstanten Steuerfolge führt zur Konstruktion robuster Verfahren, die für Signale anwendbar sind, welche nicht mehr an die scharfen Bedingungen des Satzes 1.1 gebunden sind. In einer groben Klassifizierung sollen vor allem stochastische Signale, deterministische Signale, deterministische Signale mit zufälligen Störungen, stochastische Signale mit Trend und transiente Signale genannt werden. Somit können die betrachteten diskreten Signale als reellwertige Zahlenfolgen  $X = \{x_t\}_{t=0,1,\dots}$  und das Verfahren der allgemeinen Gestalt

$$\begin{aligned} s_0 &= s && \text{(Startwert)} \\ s_{t+1} &= s_t + c K(s_t, x_{t+1}) \end{aligned} \quad (2.10)$$

mit  $c > 0$  als ein Operator  $\mathcal{P}^c$  im Raum der reellen Zahlenfolgen aufgefaßt werden, wobei  $K$  ein Korrekturterm ist. Dieser Operator kann mit Hilfe eines Blockschemas verdeutlicht werden:



Dabei bezeichnet  $\mathcal{V}$  mit  $\mathcal{V}(X)(t) = x_{t-1}$  den Verzögerungsoperator (vgl. auch [24]). Der Operator kann für die Bestimmung verschiedener Kenngrößen angewandt werden, wobei jede Kenngröße sich durch einen charakteristischen Korrekturterm auszeichnet. An dieser Stelle werden die Verfahren dargestellt, die für die Beschreibung von Biosignalen von Bedeutung sind und in dieser Arbeit auch verwendet werden.

Als erstes wird der Operator (2.10) für die Mittelwertbestimmung betrachtet, der als  $\mathcal{M}^c$  mit  $\mathcal{M}^c(X)(t) = m_t$  und

$$\begin{aligned} m_0 &= m && \text{(Startwert)} \\ m_{t+1} &= m_t + c (x_{t+1} - m_t) \end{aligned} \quad (2.11)$$

definiert wird. Wie in [22] Satz 3.3.8 gezeigt wurde, ist  $\mathcal{M}^c$  eine nichtkonsistente Schätzung des Erwartungswertes eines stationären stark mischenden Prozesses. Für nicht stationäre Folgen kann  $\mathcal{M}^c$  als ein lernfähiger Operator zur Anpassung der Mittelwertfunktion angewandt werden (vgl. z.B. [24]).

An dieser Stelle entsteht die Frage nach dem Einfluß des Adaptionsfaktors  $c$

auf die Anpassungsgüte der Schätzung. Eine Möglichkeit, ihn zu beschreiben, ergibt sich aus der Untersuchung der spektralen Eigenschaften des Operators  $M^c$ . Aus der zu (2.11) äquivalenten Darstellung

$$m_{t+1} = c x_{t+1} + (1-c) m_t,$$

ist ersichtlich, daß  $M^c$  ein lineares rekursives Filter erster Ordnung mit der Übertragungsfunktion

$$f_c(\lambda) = \frac{c^2}{1 + (1-c)^2 - 2(1-c) \cos(2\pi\lambda/\nu)} \quad (2.12)$$

mit  $\lambda \in [0, \nu/2]$  und der Abtastfrequenz  $\nu$ , realisiert (vgl. z.B. [40] oder [48]). Die Funktion  $f_c$  ist monoton fallend mit dem Maximum  $f_c(0) = 1$  und dem Minimum  $f_c(\nu/2)$ . Es handelt sich also um die Übertragungsfunktion eines Tiefpaßfilters. Mit fallendem  $c$  wird  $f_c$  immer steiler (wenn  $c_1 < c_2$  ist, dann ist  $f_{c_1}(\lambda) < f_{c_2}(\lambda)$  für  $\lambda \in (0, \nu/2]$ ), d.h. die Bandbreite des Filters wird kleiner und höhere Frequenzen werden stärker gedämpft. Bei der Bestimmung der Bandgrenze dieses Filters wird die folgende Definition verwendet. Wenn die Bandgrenze  $\lambda_0$  durch die Gleichung

$$\int_0^{\lambda_0} f_c(\lambda) d\lambda = D \int_0^{\nu/2} f_c(\lambda) d\lambda \quad (2.13)$$

definiert ist, wobei  $D$  eine Konstante mit  $0 < D < 1$  ist, dann enthält das Intervall  $[0, \lambda_0]$   $100 \cdot D$  % der Gesamtleistung des gefilterten Signals. Nach (2.12) und (2.13) kann die Bandgrenze für den Adaptionfaktor  $c$  aus

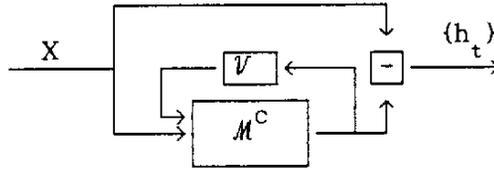
$$\lambda_0 = \frac{\nu}{\pi} \arctan \left( \frac{c}{2-c} \tan \frac{\pi D}{2} \right) \quad (2.14)$$

errechnet werden. Um andererseits eine Filterung bis zur Bandgrenze  $\lambda_0$  zu realisieren ist der Adaptionfaktor

$$c = \frac{2 \tan(\pi\lambda_0/\nu)}{\tan(\pi\lambda_0/\nu) + \tan(\pi D/2)} \quad (2.15)$$

zu wählen (vgl. auch [4] und [24]).

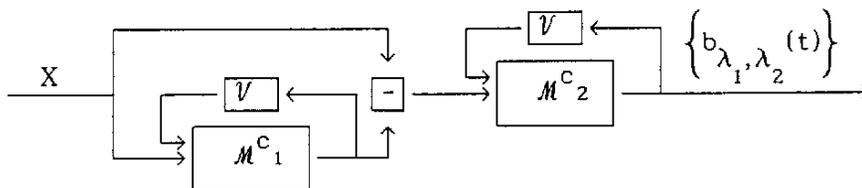
Eine Nulllinienkorrektur kann mit dem Operator  $\mathcal{K}^c = \mathcal{I} - M^c$  beschrieben werden, wobei mit  $\mathcal{I}$  der identische Operator bezeichnet wird, d.h.  $\mathcal{K}^c(X)(t) = h_t$  und  $h_t = x_t - m_t$ . Das zugehörige Blockschema wird dann wie folgt dargestellt:



Da das Verfahren (2.10) als ein Operator über Zahlenfolgen aufgefaßt wird, ist eine Kopplung der verschiedenen Operatoren möglich. Mit  $\mathcal{H}^c$  wird ein Operator definiert, der ein einfaches Hochpaßfilter realisiert (s. z.B. [40]). Ein Bandpaßfilter kann dann mit Hilfe der beiden Operatoren mittels

$$\mathcal{B}_{\lambda_1, \lambda_2} = \mathcal{H}^{c_1} \circ \mathcal{M}^{c_2}$$

konstruiert werden. Für ein bestimmtes Frequenzintervall  $[\lambda_1, \lambda_2]$  lassen sich die Adaptionfaktoren  $c_1$  und  $c_2$  aus (2.15) bestimmen. Das Blockschema hat die folgende Gestalt:



Die Kopplung zweier Mittelwertoperatoren  $\mathcal{M}^{c_1} \circ \mathcal{M}^{c_2}$  führt zur Realisierung eines rekursiven Tiefpaßfilters zweiter Ordnung, das eine bessere Übertragungsfunktion besitzen kann (vgl. z.B. [48]). Zur Wahl der Konstanten  $c_1$  und  $c_2$  wurden einige Vorschläge in [4] gemacht. Es soll noch bemerkt werden, daß es bei der Anwendung des Operators  $(\mathcal{I} - \mathcal{M}^{c_1} \circ \mathcal{M}^{c_2})$  zu einer Verstärkung der Frequenzen an der Bandgrenze  $\lambda_1$  kommen kann.

In [55] wurde die Möglichkeit der Anwendung der adaptiven Mittelwertschätzung zur Detektion und Elimination eines linearen Trends untersucht. Es wurde dabei angenommen, daß das beobachtete Signal  $\{y_t\}_{t=0,1,\dots}$  sich als eine Realisierung des Prozesses

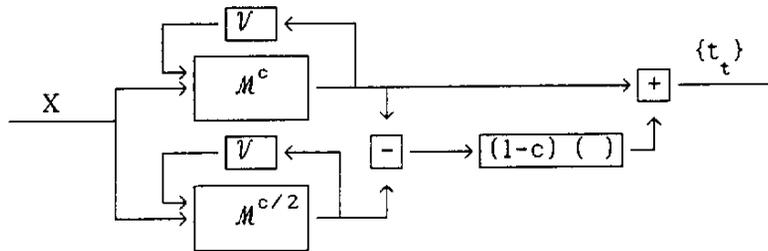
$$y_t = a + bt + x_t$$

mit  $\text{Ex}_t = 0$  und  $a, b \in \mathbb{R}$  darstellen läßt. Es wurde gezeigt, daß  $\mathcal{M}^c(Y)$  eine nicht erwartungstreue Schätzung der linearen Funktion ist und daß die Beziehung

$$EM^c(Y)(t) = a + bt - \frac{1-c}{c} b, \quad (2.16)$$

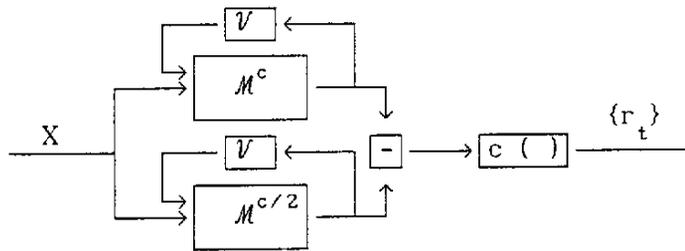
gilt, wobei  $c$  der Adaptionfaktor ist. D.h., der Trend wird mit einer Verzögerung  $\frac{1-c}{c} b$  geschätzt. Um diese Verschiebung zu eliminieren, wurde die Schätzung

$$g^c = M^c + (1-c) (M^c - M^{c/2}) \quad (2.17)$$



vorgeschlagen. Der Anstieg des Trends kann mit dem Operator

$$\mathcal{R}^c = c(M^c - M^{c/2}) \quad (2.18)$$



bestimmt werden. Für die Schätzungen gilt

$$E\mathcal{F}^c(Y)(t) = a + bt \quad \text{und} \quad E\mathcal{R}(Y)(t) = b$$

(vgl. [55]). Der Einfluß des Adaptionfaktors  $c$  auf die Güte der Trendschätzung wird in [55] an zahlreichen Beispielen demonstriert. Ein Kriterium zu seiner Bestimmung wird im Zusammenhang mit einer konkreten Anwendung im 3. Kapitel beschrieben.

Von großem Interesse für die Auswertung von stochastischen Zeitreihen ist die Spektraldichte, die die spektralen Eigenschaften des Signals beschreibt. Methoden und Modelle zu ihrer Schätzung nehmen die Stationarität der Zeitreihe als notwendige Bedingung an. Wie bereits erwähnt, wird dies von realen Signalen im allgemeinen nicht erfüllt. Damit erscheint eine dynamische Spektralanalyse notwendig, die mittels Kurzzeitspektralanalyse oder mit Hilberttrans-

formation realisiert werden kann (vgl. [71]). Beide Methoden benötigen den Algorithmus der Fouriertransformation und es handelt sich um intervallbezogene Analysen. In [24] werden einfache Kenngrößen definiert, die die Spektraleigenschaften des Signals beschreiben und sich als Operatoren in der Form (2.10) bzw. als eine Kopplung solcher Operatoren bestimmen lassen.

Die konsistente Schätzung des zweiten Momentes, die sich aus (1.10) (Satz 1.2) für  $k = 2$  bestimmen läßt, ist eine Schätzung der mittleren Leistung des Signals. Den Wert der mittleren Leistung bestimmen alle Meßwerte mit dem gleichen Gewicht und unabhängig von der Zeit. Wird die Schätzung adaptiv gestaltet, gewinnt man eine Schätzung, die (2.7) mit  $f(x) = x^2$  genügt und die ein Verhalten, wie es im vorangegangenen Abschnitt beschrieben wurde, besitzt. Der aktuelle Meßwert hat den stärksten Einfluß, die Gewichte der zurückliegenden Meßwerte nehmen mit der Zeit exponentiell ab und beeinflussen den Wert der Leistung immer weniger. Auf diese Weise wird also eine dynamische Größe konstruiert, die die momentane Leistung des Signals charakterisiert. Die Aktualität der Schätzung hängt von  $c$  ab. Der Operator der momentanen Leistung wird mit  $\mathcal{L}^c$  bezeichnet, wobei  $\mathcal{L}^c(X)(t) = l_t$  und

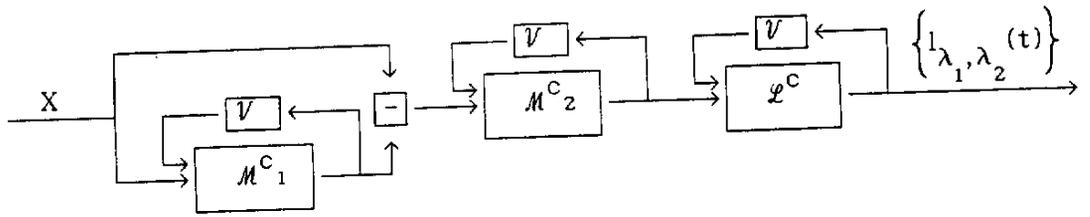
$$\begin{aligned} l_0 &= 1 \\ l_{t+1} &= l_t - c (l_t - x_{t+1}^2) \end{aligned} \quad (2.19)$$

ist. Ein Vergleich mit der Momentanleistung aus der Hilberttransformation wurde in [24] durchgeführt.

Wird die Hilberttransformation durch die Anwendung der Fouriertransformation realisiert, ist eine Filterung des Signals möglich und damit auch eine Schätzung der momentanen Leistung in bestimmten Frequenzintervallen. Um eine adaptive momentane Bandleistung zu bestimmen, werden die Filtereigenschaften des Mittelwertoperators ausgenutzt. Durch die Kopplung der Operatoren der Bandfilterung und der momentanen Leistung wird die adaptive momentane Bandleistung mit dem Operator

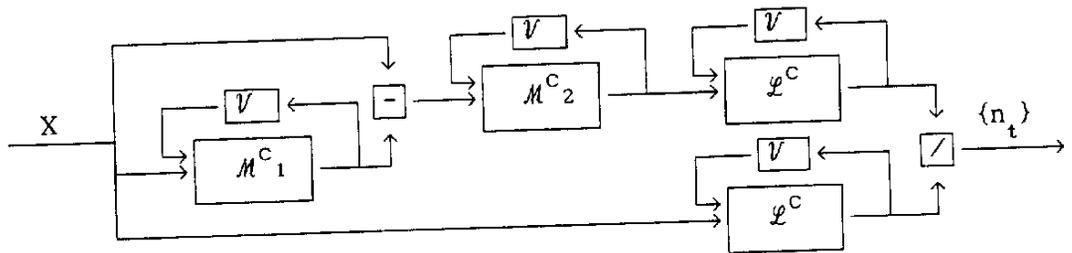
$$\mathcal{L}_{\lambda_1, \lambda_2}^c = \mathcal{B}_{\lambda_1, \lambda_2} \circ \mathcal{L}^c = \mathcal{H}^{c_1} \circ \mathcal{M}^{c_2} \circ \mathcal{L}^c \quad (2.20)$$

definiert, wobei die Konstanten  $c_1$  und  $c_2$  aus (2.15) zu bestimmen sind. Für  $c_1 = c_2 = 1$  wird die adaptive Momentanleistung berechnet. Nach dem folgenden Blockschema können also beide Operatoren realisiert werden:



Die normierte Bandleistung ergibt sich aus der folgenden Kopplung der Operatoren

$$N_{\lambda_1, \lambda_2}^c = \varphi_{\lambda_1, \lambda_2}^c / \varphi^c \quad (2.21)$$



(vgl. [26]).

Eine zweite dynamische spektrale Kenngröße ist die momentane Frequenz. Bei der Konstruktion des adaptiven Operators für ihre Bestimmung wurde eine verallgemeinerte Definition der Frequenz angewandt (vgl. [24]). Üblicherweise wird die Frequenz als ein Parameter sinusförmiger periodischer Funktionen angesehen, der die Anzahl der Perioden pro Zeiteinheit charakterisiert. Für diese Funktionsmenge ist die Frequenz aber auch gleichermaßen bestimmt durch die halbe Anzahl der Nulldurchgänge pro Zeiteinheit. Wird die nun zuletzt genannte Charakterisierung der Frequenz benutzt, muß die Definition nicht mehr zwangsläufig auf sinusförmige und periodische Funktionen eingeschränkt werden. Dieses Konzept der Verallgemeinerung des Begriffes Frequenz wurde beispielsweise erfolgreich für die Ordnung von Walshfunktionen eingesetzt, die ihrerseits im Instrumentarium der Walsh-Hadamard-Transformationen für stufenförmige Signalstrukturen (meistens technischer Signale) erhebliche Bedeutung für die Signalanalyse erlangten. Ausgehend von diesen Überlegungen wurde in [24] als adaptive Momentanfrequenz ein adaptiver Mittelwert über die Anzahl der Überschreitungen (dividiert durch zwei) eines bestimmten Niveaus definiert:

$$f_0 = f$$

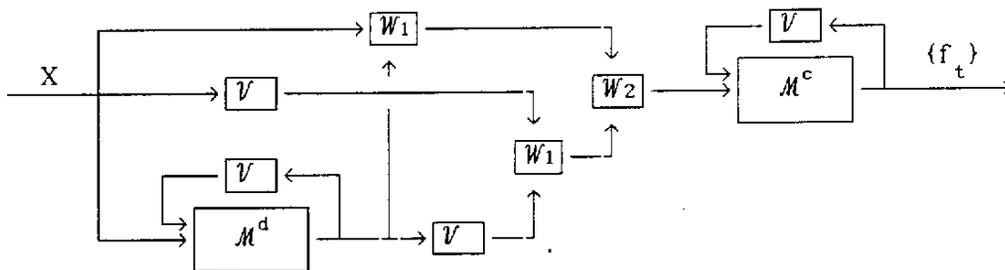
$$f_{t+1} = f_t - c \left( f_t - \int_{(-\infty, 0)} ((x_t - m_t^d) \cdot (x_{t+1} - m_{t+1}^d)) \right) \quad (2.22)$$

Als Niveau wurde allerdings nicht die Nulllinie gewählt. Das hätte zu Folge, daß ein immer positives, aber durchaus variierendes Signal kontinuierlich eine Momentanfrequenz Null zugeordnet bekommen würde. In Anlehnung an die übliche Spektralanalyse, in der oft der Nulllinie des Spektrums (die durch den zeitlich gebildeten Mittelwert des Signals bestimmt wird) keine Beachtung geschenkt wird, wurde in der Definition der adaptiven momentanen Frequenz als Niveau ebenfalls ein zeitlich gebildeter Mittelwert, der adaptive Mittelwert des Signals, gewählt. Die Konstante  $d$  in der Mittelwertschätzung bestimmt die Lage der deterministischen Linie. Ist  $d = 0$ , so werden die Nulldurchgänge gezählt, sonst ist  $d$  aus (2.15) zu berechnen. Die Konstante  $c$  charakterisiert die Aktualität der Schätzung.

Zur Beschreibung des zugehörigen Operators wurden noch die Vergleichsoperatoren  $W_1$  mit  $w_1(x,y) = \int_{(0, \infty)} (x-y)$  und  $W_2$  mit  $w_2(x,y) = \int_{(\mathbb{R} \setminus 0)} (x-y)$  definiert. Dann wird die adaptive Momentanfrequenz mit dem Operator

$$\mathcal{F}^c = M^c \left( W_2 \left( W_1(V, V(M^d)), W_1(\mathcal{F}, M^d) \right) \right) \quad (2.23)$$

berechnet und dem zugehörigen Blockschema rechentechnisch realisiert:



In der Hilberttransformation wird mit der momentanen Leistung und der momentanen Frequenz das Signal eindeutig beschrieben. In welcher Art sie sich mit den entsprechenden adaptiven rekursiven Kenngrößen vergleichen lassen wurde in [24] untersucht.

## 2.3 Konstruktion eines adaptiven Schätzverfahrens zur Anpassung von ARMA-Modellen

Nach den Untersuchungen aus dem 1. Kapitel können stationäre Meßsignale mit Hilfe sachgerecht angepaßter linearer parametrischer Modelle effektiv beschrieben werden. So entsteht die Frage, ob sich diese Modelle auch zur Beschreibung instationärer Zeitreihen anwenden lassen. Die untersuchten Schätzverfahren aus dem Abschnitt 1.2.2 sind aber an strenge statistische Eigenschaften, die die vorliegende Zeitreihe erfüllen soll, gebunden. Ihre Verletzung kann zu einigen Schwierigkeiten führen, wie es im Abschnitt 1.3.3 demonstriert wurde. Da sich das Verfahren (E), (A), (B) und (S) in die rekursiven Methoden aus der stochastischen Approximation einordnen läßt (siehe die Sätze 1.6-1.8), soll die Möglichkeit untersucht werden, ob durch eine adaptive Gestaltung der Steuerfolge aus dieser Rekursionsvorschrift ein robuster Algorithmus konstruiert werden kann, der auf Strukturbrüche reagiert und sich neuen Bedingungen anpaßt.

Die Simulationsergebnisse zum Algorithmus (E), (A), (B) und (S) im Abschnitt 1.2.2 zeigen, daß bei diesem Verfahren nur dann eine günstige Parameterschätzung erfolgt, wenn die Steuerfolge  $\{c_t\}$  den Eigenschaften der Zeitreihe angepaßt ist (siehe (1.34)). Aus diesem Grund ist es nicht ausreichend, sich bei der Konstruktion eines adaptiven Verfahrens auf einen konstanten Adaptionsfaktor zu beschränken, wie es in den Verfahren im vorangegangenen Abschnitt der Fall ist. Da sich ein Zusammenhang zwischen der Steuerfolge nach (1.34) und der Varianz des Signals zeigen läßt (vgl. (1.35)), ist es angebracht, die Steuerfolge in zwei Komponenten zu zerlegen.

Sei  $X = \{x_t\}_{t=0}^{\infty}$  das beobachtete Signal. Die Steuergröße  $c_t$ , im weiteren als Adaptionsvariable bezeichnet, wird dann in der folgenden Form darstellbar

$$c_t = c_x(t) \cdot d, \quad (2.24)$$

dabei realisiert der Ausdruck  $c_x(t)$  die Anpassung an die vorliegende Zeitreihe und der Faktor  $d$  mit  $0 < d < 1$  ist für die Steuerung von Adaptivität bzw. Konvergenz des Schätzverfahrens verantwortlich. In Anlehnung an die Konstruktion (1.34) der Steuerfolge für das konvergente rekursive Verfahren wird der Ausdruck  $c_x(t)$  reziprok zur Varianzschätzung der Zeitreihe gewählt. Dabei kann man  $c_x(t)$  in jedem Schritt anpassen

$$c_x(t) = (\hat{\sigma}_x^2(t))^{-1}, \quad t = 1, 2, \dots \quad (2.25)$$

Nach den Betrachtungen im Abschnitt 2.3 erfolgt die Schätzung der Varianz der Zeitreihe  $\{x_t\}$  adaptiv nach dem Algorithmus

$$\hat{\sigma}_x^2(t) = \hat{\sigma}_x^2(t-1) - c_s (\hat{\sigma}_x^2(t-1) - x_t^2), \quad (2.26)$$

wobei  $c_s$  eine aus Voruntersuchungen gewonnene Adaptionskonstante ist. (Wie bei der Definition des Algorithmus (E), (A), (B) und (S) wird hier angenommen, daß die vorliegende Zeitreihe zentriert ist und die Schätzung des zweiten Momentes mit einer Schätzung der Varianz übereinstimmt.)

Analog wird die Streuung des Vorhersagefehlers nun adaptiv geschätzt:

$$\hat{\sigma}_\varepsilon^2(t) = \hat{\sigma}_\varepsilon^2(t-1) - c_\varepsilon (\hat{\sigma}_\varepsilon^2(t-1) - e_t^2), \quad t = 1, 2, \dots \quad (2.27)$$

An den folgenden Simulationsbeispielen werden einige charakteristische Eigenschaften der adaptiven Schätzfolgen für die Parameter von ARMA-Prozessen demonstriert und der Einfluß der Wahl von  $c_x(t)$  und dem Adaptionsfaktor  $d$  verdeutlicht.

Im ersten Beispiel werden die Parameter  $a_1 = 0.3$ ,  $b_1 = 0.4$  und  $\sigma_\varepsilon^2 = 100$  eines simulierten ARMA(1,1)-Prozesses nach dem beschriebenen Verfahren adaptiv geschätzt. Dabei wurde  $c_s = 0.01$  gewählt,  $c_x(t)$  nach Formel (2.25) berechnet und der Adaptionsfaktor  $d = 0.01$  konstant gelassen.

Es wird deutlich, daß nach einer kurzen Anfangs- oder Lernphase die Schätzfolgen um den wahren Wert schwingen und ihre Umgebung nicht mehr verlassen (vgl. Abbildungen 2.2 und 2.3). Die Adaptionsvariable  $c_t$  verändert sich dabei ständig durch die Anpassung des Termes  $c_x(t)$  an das Verhalten der Zeitreihe, wie aus den Gleichungen (2.25) und (2.26) deutlich wird.

Diese Veränderungen der im weiteren als Normierungsfaktor bezeichneten Größe  $c_x(t)$  sind aber nicht für die Schwankungen der Schätzfolgen der Parameter  $a_1$  und  $b_1$  verantwortlich. Läßt man  $c_x(t)$  nach einer genügend langen Anfangsphase konstant, so verhalten sich die Schätzfolgen der Parameter ähnlich wie vorher. Abbildung 2.1 zeigt die Folge der Adaptionsnorm  $c_x(t)$  berechnet nach (2.25) und (2.26) im Vergleich mit einer nach 100 Schritten konstanten Adaptionsnorm, wobei  $\sigma_x^2(t)$ ,  $t \leq 100$ , rekursiv aus

$$\sigma_x^2(t) = \sigma_x^2(t-1) - (\sigma_x^2(t-1) - x_t^2)/t$$

bestimmt wird.

In Abbildung 2.2 werden die Schätzfolgen des Parameters  $a_1$  für den sich anpassenden Normierungsfaktor  $c_x(t)$  und der konstante Normierungsfaktor  $c_x(t_0)$  miteinander verglichen. Nach einer kurzen Anfangsphase sind beide Kurven fast identisch. In Abbildung 2.3 wird dieselbe Aussage für die Schätzfolgen des Parameters  $b_1$  dargestellt.

Im folgenden Untersuchungsbeispiel wird deutlich, daß der in die Adaptionvariable  $c_t$  eingehende Ausdruck  $c_x(t)$  tatsächlich eine normierende Funktion ausübt und die Bezeichnung Normierungsfaktor zutreffend ist. Der autoregressive Prozeß AR(5) mit den Parametern  $a_1 = 0.1$ ,  $a_2 = -0.2$ ,  $a_3 = -0.05$ ,  $a_4 = 0.1$  und  $a_5 = 0.4$  wurde mit der Rauschstreuung  $\sigma_\varepsilon^2 = 100$  und  $\sigma_\varepsilon^2 = 0.5$  erzeugt. Für beide Prozesse wurden die autoregressiven Parameter mit  $c_s = d = 0.01$  adaptiv bestimmt.

Trotz des großen Unterschiedes der Rauschstreuungen und damit auch der Streuungen der entsprechenden autoregressiven Prozesse sind die adaptiven Schätzalgorithmen sachgerecht, wie es in der Abbildung 2.4 an den Schätzfolgen des Parameters  $a_5$  der beiden Prozesse demonstriert wird.

In den bisherigen Beispielen wurden nur Untersuchungen zum Normierungsfaktor  $c_x(t)$  vorgenommen. Daneben spielt der Adaptionfaktor  $d$  eine steuernde Rolle für die Schnelligkeit der Adaption an den wahren Parameter bzw. des Schwingungsverhaltens der Schätzfolge um diesen. Die Wirkung des Adaptionfaktors  $d$  wird an dem folgenden Simulationsbeispiel eines ARMA(1,1)-Prozesses mit den Parametern  $a_1 = 0.5$ ,  $b_1 = 0.5$  und  $\sigma_\varepsilon^2 = 1$  demonstriert. Die adaptiven Schätzungen wurden mit den Adaptionskonstanten  $c_s = c_\varepsilon = 0.01$  und mit den verschiedenen Adaptionfaktoren  $d_1 = 0.005$ ,  $d_2 = 0.01$  und  $d_3 = 0.02$  durchgeführt.

In den Abbildungen 2.5 und 2.6 sind die entsprechenden Schätzfolgen der Parameter  $a_1$  und  $b_1$  dargestellt. Beide Parameterschätzungen benötigen für den kleinsten Adaptionfaktor  $d_1 = 0.005$  eine sehr lange Phase, um die Umgebung des wahren Parameterwertes überhaupt zu erreichen. Dafür zeigen die Schätzfolgen ein sehr glattes Verhalten, so daß nach erfolgter Konvergenz zum wahren Wert kaum starke Abweichungen von diesem erfolgen. Umgekehrt erfolgt die Einschwingung auf den wahren Parameter für den größten Adaptionfaktor

$d_3 = 0.02$  bei beiden Schätzfolgen sehr schnell, dafür sind stärkere Schwankungen um den entsprechenden Parameter erkennbar.

Hier offenbart sich sichtlich der dynamische Charakter des adaptiven Schätzverfahrens, der allgemein adaptiven Schätzungen auch anderer Prozeßparameter wie z.B. Mittelwert, Streuung, Quantilwerte u. a. zueigen ist (vgl. Abschnitt 2.1 bzw. 2.2).

Der Einfluß der einzelnen Komponenten der Steuerfolge  $\{c_t\}$  wird auch aus der Darstellung der Schätzfolgen der Parameter als Funktion der Steuerfolge und der Zeitreihe bis zur Zeit  $t$  deutlich. Dabei lassen sich die Formeln (A) und (B) (ohne die Randbedingungen) für  $t > \max\{p, q\}$  wie folgt darstellen:

$$\hat{a}_k(t) = c_t x_t x_{t-k} + \sum_{m=p+1}^{t-1} c_m x_m x_{m-k} \prod_{j=m}^{t-1} \left( 1 - c_j \left( \sum_{i=1}^p x_{j-i}^2 + \sum_{i=1}^q e_{j-i}^2 \right) \right) \quad , \quad k = 1, \dots, p$$

$$\hat{b}_k(t) = c_t x_t e_{t-k} + \sum_{m=q+1}^{t-1} c_m x_m e_{m-k} \prod_{j=m}^{t-1} \left( 1 - c_j \left( \sum_{i=1}^p x_{j-i}^2 + \sum_{i=1}^q e_{j-i}^2 \right) \right) \quad . \quad k = 1, \dots, q$$

Im Vergleich zu anderen rekursiven Verfahren, die sich ähnlich darstellen lassen (vgl. (2.1) und (2.2)), hängen die Gewichte

$$d_m = c_m \prod_{j=m}^{t-1} \left( 1 - c_j \left( \sum_{i=1}^p x_{j-i}^2 + \sum_{i=1}^q e_{j-i}^2 \right) \right) \quad (m > p, \text{ bzw } m > q)$$

nicht nur von einer deterministischen Steuerfolge  $\{c_m\}$  ab, sondern auch von den Werten der Zeitreihe. Wird in der adaptiven rekursiven Schätzung die Steuerfolge  $\{c_t\}$  als Produkt einer Konstanten und einer Normierungsgröße gewählt (vgl. (2.24) und (2.25)), so gilt:

$$\hat{a}_k(t) = d \cdot \frac{x_t x_{t-k}}{\hat{\sigma}_x^2(t)} + d \cdot \sum_{m=p+1}^{t-1} \frac{x_m x_{m-k}}{\hat{\sigma}_x^2(m)} \prod_{j=m}^{t-1} \left( 1 - d \frac{\sum_{i=1}^p x_{j-i}^2 + \sum_{i=1}^q e_{j-i}^2}{\hat{\sigma}_x^2(m)} \right) \quad , \quad k = 1, \dots, p$$

$$\hat{b}_k(t) = d \cdot \frac{x_t e_{t-k}}{\hat{\sigma}_x^2(t)} + d \cdot \sum_{m=p+1}^{t-1} \frac{x_m e_{m-k}}{\hat{\sigma}_x^2(m)} \prod_{j=m}^{t-1} \left( 1 - d \frac{\sum_{i=1}^p x_{j-i}^2 + \sum_{i=1}^q e_{j-i}^2}{\hat{\sigma}_x^2(m)} \right) \quad . \quad k = 1, \dots, q$$



### 2.3.1 Vergleich der adaptiven Modellanpassung mit konsistenten Schätzverfahren

Aus den untersuchten Beispielen wurde deutlich, daß sich bei einer sachgerechten Anwendung des adaptiven Verfahrens die Parameterschätzungen schnell in die unmittelbare Umgebung der wahren Parameter bewegen, aber eine Konvergenz nicht vorliegt. Die Schätzungen liegen zwar in der Umgebung der wahren Parameter, verändern sich aber ständig, wobei die Schwankungsbreite wesentlich vom Adaptionsfaktor  $d$  abhängt. In der Studie [8], Abschnitt 4, wurde gezeigt, daß die Genauigkeit der Schätzungen ausreichend für die Darstellung wesentlicher inhaltlicher Zusammenhänge und der charakteristischen Eigenschaften der Zeitreihen ist und für endliche Abschnitte (die Anzahl der Werte war 512) durchaus mit den Ergebnissen konvergenter Schätzverfahren vergleichbar ist. Dazu wurden Realisierungen bekannter ARMA-Prozesse angewandt und die Schätzfolgen des rekursiven adaptiven Verfahrens (2.28)-(2.31) mit denen des rekursiven konvergenten Verfahrens (E), (A), (B), (S) verglichen. Für  $q = 0$  wurden die Endwerte ( $N = 512$ ) der rekursiven Algorithmen mit den Parametern aus der Lösung der Yule-Walker-Gleichungen verglichen.

Im Gegensatz zur rekursiven Schätzung, bei der die Korrekturgrößen mit wachsender Wertezahl immer kleiner werden, kann die adaptive Folge unmittelbar auf die Veränderung der Zeitreihe reagieren aber damit natürlich nicht konvergieren.

Nach dem Einzelvergleich der Parameterschätzungen wurde außerdem ein Vergleich der entsprechenden parametrischen Spektraldichteschätzungen vorgenommen. Als Beispiel wird in der Abbildung 2.7 ein Spektraldichtenvergleich zum AR(5)-Prozeß mit den Parametern aus dem vorangegangenen Beispiel mit den wahren und den nach 512 Werten geschätzten Parametern dargestellt.

Die Form der Spektraldichte wird von allen Schätzverfahren richtig wiedergegeben. Unterschiedlich ist die Höhe des größten lokalen Maximums. Die hier sichtbare Abweichung bei der adaptiven Schätzung ist in der Dynamik des Verfahrens und dabei entstehenden Schwankungen begründet. Neben den Untersuchungen an simulierten Modellen wurden auch Spektraldichteschätzungen von Biosignalen verglichen. Die Abbildungen 2.8 und 2.9 zeigen die Spektraldichteschätzungen eines EEG-Signals (Elektrode O1) eines Probanden im Ruhezustand mit geöffneten und geschlossenen Augen. Deutlich ist bei allen Darstellungen der

Leistungszuwachs im  $\alpha$ -Band (8-13 Hz) bei geschlossenen Augen zu erkennen. Dabei spiegelt die adaptive Schätzung nur eine kurze Vergangenheit wider und ist daher innerhalb kurzer Zeiträume Schwankungen unterworfen. Hierzu ist es angebracht zu untersuchen, inwieweit glättende Verfahren solche Schätzungen stabilisieren.

Somit wurde gezeigt, daß bei sachgerechter Vorgehensweise die Ergebnisse des adaptiven Verfahrens mit denen konsistenter Schätzverfahren, angewandt auf endliche Zeitabschnitte, vergleichbar sind. Damit stehen adaptive parametrische Methoden für die Analyse von Biosignalen zur Verfügung.

### 2.3.2 Adaptivität und Konvergenzverhalten der adaptiven ARMA-Modellierung

Im Unterschied zu bekannten konvergenten Anpassungsverfahren von ARMA-Modellen kann die adaptive ARMA-Modellanpassung so gestaltet werden, daß sie ein stark dynamisches Verhalten aufweist. Die adaptiven Parameterschätzungen reagieren schnell auf Veränderungen der stochastischen Signale, wie z.B. den Übergang von einer stationären Phase zu einer neuen stationären Phase oder das Auftreten von Instationaritäten, und passen sich der neuen Situation selbständig an. Da die adaptive ARMA-Modellanpassung jeweils nur auf den aktuellen Wert und eine kurze Vergangenheit der Zeitreihe zurückgreift, können die Parameter in Echtzeit berechnet werden.

Bevor die Analyse von Biosignalen aufgrund der adaptiven rekursiven ARMA-Modellierung mit Hilfe von (2.28)-(2.31) durchgeführt wird, sollen das Adaptions- und Konvergenzverhalten des Verfahrens an simulierten Zeitreihen untersucht werden. Dabei soll der Einfluß der Adaptionsgrößen  $c_s$ ,  $c_\varepsilon$  und des Adaptionsfaktors  $d$  auf diese Eigenschaften deutlich gemacht werden.

Die Varianzschätzungen des Signales  $\hat{\sigma}_x^2$  und des Vorhersagefehlers  $\hat{\sigma}_\varepsilon^2$  stehen in einem engen funktionalen Zusammenhang (vgl. (2.29) und (2.31)), aus diesem Grund werden die Konstanten  $c_s$  und  $c_\varepsilon$  in gleicher Weise gewählt bzw. verändert und somit im weiteren nur eine Konstante  $c$  mit  $c = c_s = c_\varepsilon$  betrachtet.

Es wurden die folgenden Simulationsbeispiele aus den aufeinanderfolgenden Realisierungen zweier unterschiedlicher AR-Prozesse mit jeweils 256 Werten untersucht. Bei gezielter Wahl der autoregressiven Parameter und der Streuung  $\sigma_\varepsilon^2$  im ersten und zweiten Prozeß, die eine deutliche Veränderung im Spektral-

bereich bewirken, wird das Anpassungsverhalten der adaptiven ARMA-Modellierung an der Sprungstelle im Zusammenhang mit dem Konvergenzverhalten in den stationären Teilbereichen beobachtet.

Im 1. Beispiel sei  $x_1, \dots, x_{256}$  die Realisierung eines AR(2)-Prozesses mit den Parametern  $a_1 = -0.5$ ,  $a_2 = -0.4$ ,  $\sigma_\varepsilon^2 = 1$  und  $x_{257}, \dots, x_{512}$  Realisierungen eines AR(2)-Prozesses mit den Parametern  $a_1 = 0.5$ ,  $a_2 = -0.4$ ,  $\sigma_\varepsilon^2 = 1$ . Für die Spektraldichten dieser Prozesse gilt, daß

$$f_1\left(\frac{\pi}{4} - \lambda\right) = f_2\left(\frac{\pi}{4} + \lambda\right) \quad \text{für } \lambda \in [0, \pi/4],$$

d.h., der Wechsel des ersten autoregressiven Parameters  $a_1$  von  $-0.5$  zu  $0.5$  bewirkt den Übergang von schnellen zu langsamen Schwankungen, wobei die Varianz des Signals sich nicht verändert.

Die Untersuchungen zeigen, daß der Adaptionfaktor  $d$  eine große Bedeutung für die Geschwindigkeit der Anpassung der Schätzfolge des ersten autoregressiven Parameters  $a_1$  nach dem Sprung von  $-0.5$  zu  $0.5$  hat. Die Anpassungsgeschwindigkeit hängt stark von seiner Wahl ab, die Wahl der Adaptionvariablen  $c$  ist hier kaum von Bedeutung (vgl. Abbildung 2.10). Die Schätzfolge des eigentlich unveränderten Parameters  $a_2$  verändert sich in der Übergangsphase ebenfalls, da sie mit der Schätzfolge  $\hat{a}_1(t)$  eng zusammenhängt (Abbildung 2.11).

Die Abbildungen 2.10 und 2.11 zeigen, daß die Schätzfolgen der autoregressiven Parameter für ein größeres  $d$  stärker schwanken. Ausdruck dafür ist die folgende Kenngröße. Man betrachtet die Schätzfolge  $\{\theta(t)\}_{t=0,1,\dots}$  des ARMA-Parametervektors. Sei  $k^* \in \mathbb{N}$  mit  $0 < k^* \leq p+q$ ,  $k^* = k^*(t)$  und

$$|\theta_{k^*}(t) - \theta_{k^*}(t-1)| = \max_{0 < k \leq p+q} |\theta_k(t) - \theta_k(t-1)| \quad \text{für } t \in \mathbb{N}.$$

Dann wird mittels

$$\kappa(t) = \theta_{k^*}(t) - \theta_{k^*}(t-1) \quad (2.32)$$

die maximale Parameterschwankung definiert. Für das 1. Beispiel wird die Folge  $\{\kappa_t\}_{t=1}^N$  in der Abbildung 2.12 für die unterschiedlichen Adaptionvariablen dargestellt. Die Wahl des Adaptionfaktors  $d$  wirkt sich wesentlich auf die Endwerte der Parameterschätzungen ( $N = 512$ ) aus. Sie sind in der folgenden Tabelle zusammengefaßt:

c	d	$\hat{\sigma}_{\varepsilon}^2(N)$	$\hat{a}_1(N)$	$\hat{a}_2(N)$
0.01	0.01	1.275	0.368	-0.298
0.03	0.01	1.434	0.341	-0.290
0.01	0.03	1.226	0.522	-0.411
0.03	0.03	1.386	0.493	-0.404

Tabelle 2.2: Adaptive Schätzung der Parameter von 2 nacheinanderfolgenden AR(2)-Prozessen mit verschiedenen Adaptationsparametern

Es zeigt sich, daß die Veränderung von  $a_1$  für  $d = 0.01$  nicht vollständig erfaßt wird und die Wahl  $d = 0.03$  wesentlich geeigneter ist. Die Erkennung der Wanderung der Frequenzanteile mit hohen Leistungsanteilen von hohen zu niedrigen Frequenzen gelingt für  $d = 0.03$ , aber nicht vollständig für  $d = 0.01$  (vgl. Abbildung 2.13).

Eine Verstärkung des Signals kann durch Vergrößerung des Parameters  $\sigma_{\varepsilon}^2$  erreicht werden. Dabei bleibt die Form der Spektraldichte erhalten (vgl. z.B. (1.18)). Bei geeigneter Wahl der Konstanten  $c$  kann eine schnelle Anpassung der Schätzfolge  $\{\hat{\sigma}_x^2(t)\}$  an die neue Situation und damit eine stabile Schätzung der Parameter erreicht werden.

Man betrachte zuerst eine Folge  $\{x_t\}$ , wobei  $x_1, \dots, x_{256}$  die Realisierung eines AR(2)-Prozesses mit den Parametern  $a_1 = 0.4$ ,  $a_2 = 0.2$ ,  $\sigma_{\varepsilon}^2 = 1$  und  $x_{257}, \dots, x_{512}$  eines AR(2)-Prozesses mit den Parametern  $a_1 = 0.4$ ,  $a_2 = 0.2$ ,  $\sigma_{\varepsilon}^2 = 10$  sind. Das Signal wird also mit den gleichen autoregressiven Parametern erzeugt, aber die Leistung wird nach 256 Werten wesentlich erhöht.

Aus den Eigenschaften der adaptiven Verfahren ist bekannt, daß die Schätzung des Normierungsfaktors  $c_x(t) = (\hat{\sigma}_x^2(t))^{-1}$  sich mit  $c = 0.01$  nur zögernd den neuen Bedingungen anpassen kann. Für  $c = 0.03$  erfolgt die Reaktion der Schätzung auf den Leistungssprung wesentlich schneller, schwankt dann aber auch in der stationären Phase stärker. Das Verhalten der Schätzfolgen der Varianz des Vorhersagefehlers ist ganz analog. Dabei zeigt sich, daß die Wahl des Adaptionsfaktors  $d$  kaum eine Rolle spielt (siehe Abbildung 2.14).

Die Schätzfolgen  $\hat{a}_1(t)$  und  $\hat{a}_2(t)$  schwanken nach der Einschwingphase um ihren wahren Wert und zeigen nur eine kurze Reaktion an der Sprungstelle. Ein kleiner Adaptionsfaktor  $d$  bewirkt dabei in stationären Abschnitten "ruhige" Schätzfolgen, während die Wahl von  $c$  hier kaum einen Einfluß auf diese Schätzungen hat. Betrachtet man die maximale Parameterschwankung  $\langle \kappa_t \rangle$  für unterschiedliche Adaptionsvariablen und Adaptionsfaktoren, so erkennt man, daß die Wahl der Konstanten  $c$  im Normierungsfaktor einen wesentlichen Einfluß auf die Anpassungsgeschwindigkeit der Parameter in der Übergangsphase hat. Eine gleichmäßig kleine Schwankung ist für eine günstige Wahl von  $c = 0.03$  und für einen kleinen Adaptionsfaktor  $d = 0.01$  zu beobachten (vgl. Abbildung 2.15). Für andere Konstanten ist die Schwankung der Parameter an der Sprungstelle im Vergleich zu den stationären Abschnitten merkbar größer, da der Normierungsfaktor zu klein ist.

Dieses Verhalten der geschätzten Parameter kann auch mittels eines Interquantilintervalls um die Folge  $\langle \kappa_t \rangle$  bzw. mittels der Breite dieses Intervalls zum Ausdruck gebracht werden. Diese Kenngröße kann aus (1.11) mit einer konstanten Steuerfolge adaptiv bestimmt werden. Auch sie zeigt für  $c = 0.03$  und  $d = 0.01$  eine gleichmäßig kleine Schwankung und deutet auf eine gute angepasste Verarbeitung der Sprungstelle hin (siehe Abbildung 2.16).

Eine ähnliche Wirkung der Konstanten  $c$  und  $d$  kann bei einem Leistungsabfall beobachtet werden. Im letzten Simulationsbeispiel ist  $x_1, \dots, x_{256}$  die Realisierung eines AR(3) - Prozesses mit den Parametern  $a_1 = 0.4$ ,  $a_2 = -0.3$ ,  $a_3 = 0.2$ ,  $\sigma_\varepsilon^2 = 15$  und  $x_{257}, \dots, x_{512}$  die Realisierung eines AR(3)-Prozesses mit den Parametern  $a_1 = 0.4$ ,  $a_2 = -0.3$ ,  $a_3 = 0.2$ ,  $\sigma_\varepsilon^2 = 3$ . Das Signal wird wieder mit den gleichen autoregressiven Parametern erzeugt, die Leistung fällt nach 256 Werten stark ab. Im Signalverlauf sind damit die Amplitudenschwankungen im zweiten Teil wesentlich kleiner. Die Anpassungsgeschwindigkeit in der Übergangsphase wird in der Folge  $\langle \kappa_t \rangle$  deutlich (Abbildung 2.17). Die geringe Größe der maximalen Parameterschwankung nach der Sprungstelle deutet darauf hin, daß die Korrektur der Parameter nur in eine Richtung erfolgt, bis das gesamte System der Schätzungen wieder eingeschwungen ist. Dabei ist die Adaptionsvariable  $c$  für die Länge dieses Überganges maßgeblich

## 2.4 Bestimmung der Steuerfolge in rekursiven Quantilschätzungen

In die Familie der adaptiven rekursiven Verfahren läßt sich die adaptive Gestaltung der Schätzung von Quantilwerten aus Satz 1.3 einordnen. Dabei erhält die Vorschrift (1.11) die folgende Form:

$$\begin{aligned} q_\alpha(0) &= q \quad (\text{Startwert}) \\ q_\alpha(t+1) &= q_\alpha(t) + c \left( \alpha - I_{(-\infty, q_\alpha(t))}(x_{t+1}) \right), \quad c > 0 \end{aligned} \quad (2.33)$$

wobei mit  $\{x_t\}$  ein Zeitsignal bezeichnet wird. Sie wurde bei der Bestimmung der Konfidenzintervalle in den Untersuchungen der Eigenschaften der adaptiv geschätzten Parameter eines ARMA-Modells im vorangegangenen Abschnitt angewandt. Die Eigenschaften des Schätzverfahrens (2.33) wurden an zahlreichen Beispielen getestet, die Beschreibung der Untersuchungen und die detaillierte Auswertung der Ergebnisse sind in der Studie [54] und in [23] dargestellt. Sie zeigen, daß die Schätzung für mathematische Stichproben mit bekannter Verteilung im Vergleich mit anderen Schätzungen (konsistente rekursive Schätzfolge und empirische Quantilbestimmung) hinreichend gute Werte liefert. Bei der Wahl des Adaptionfaktors  $c$  wurde ein ähnliches Verhalten beobachtet, wie es sich z.B. bei der Momentenschätzung explizit beschreiben läßt. Die Schwankung der Schätzfolge um den theoretischen Wert für ein konstantes  $c$  wird mit wachsendem  $t$  nicht geringer, sie sinkt aber, wenn  $c$  verkleinert wird. Dagegen nimmt bei Strukturbrüchen mit wachsendem  $c$  die Anpassungsgeschwindigkeit zu. Anders als im Fall von Verfahren, die dem Satz 2.1 genügen, ist der funktionale Zusammenhang zwischen der Streuung der Schätzfolge und der Streuung der Zeitreihe bzw. der Konstanten  $c$  im allgemeinen nicht bekannt und zur Wahl des Faktors  $c$  liegen keine Einschränkungen vor. Die Untersuchungen zeigen, daß die Streuung der adaptiven Quantilschätzung durch die Streuung der vorliegenden Zeitreihe nicht beschränkt ist, und daß bei einer ungünstigen Wahl des Adaptionfaktors  $c$  die Schätzung durch ein zu starkes Rauschen verdeckt werden kann. Aus diesen Gründen werden Methoden zur Konstruktion der Steuerfolge gesucht, welche die Eigenschaften der Zeitreihe in die Berechnung einschließen. Als erstes könnte die Steuerung der Folge  $\{c_t\}$  zwischen zwei Konstanten untersucht werden, wie es mit (2.9) beschrieben wurde. Da das adaptive Verfahren (2.33) der Schätzung von Quantilen an sich ein

oszillierendes Verhalten aufweist (s. [23]), ist das in (2.9) verwendete Steuerkriterium für die Quantilschätzung nicht geeignet. Andere Methoden ergeben sich aus den folgenden Betrachtungen.

Aus der numerischen Mathematik sind für die Berechnung einer Nullstelle  $r$  der reellen Funktion  $R: \mathbb{R} \rightarrow \mathbb{R}$  verschiedene Sekantenverfahren bekannt. Es wird gezeigt, daß die Folge  $\{x_t\}_{t=1}^{\infty}$  mit

$$x_{t+1} = x_t - \frac{R(x_t)}{\Delta R(x_t)}$$

unter gewissen Bedingungen gegen die gesuchte Nullstelle konvergiert, wobei mit  $\Delta R(x_t)$  der Anstieg von  $R$  in einer Umgebung von  $x_t$  bezeichnet wird (vgl. z.B. [43]). Hier entsteht die Frage, ob es möglich wäre, mit Hilfe einer Schätzung des Anstiegs der Verteilungsfunktion  $F_{\xi}$  in der Umgebung des aktuellen Schätzwertes  $q_{\alpha}(t)$  die Steuerfolge in der rekursiven Schätzung von Quantilen zu bestimmen. Für die Schätzung des Anstiegs von  $F_{\xi}$  kann der folgende Satz angewandt werden.

**Satz 2.2:**

Sei  $\{\xi_t(\omega)\}_{t=0}^{\infty}$  eine Folge stark mischender Zufallsgrößen mit der absolut stetigen Verteilungsfunktion  $F_{\xi}$  und der Dichtefunktion  $f_{\xi}$ . Die Bedingungen (1.7) und (1.8) seien erfüllt. Ein Intervall  $(a,b]$  wird in disjunkte Klassen  $K_i = [a + (i-1)h, a + ih]$  zerlegt, wobei mit  $h = (b-a)/N$  die Klassenbreite und  $i = 1, \dots, N$  der Klassenindex bezeichnet sind. Zusätzlich werden die Klassen  $K_0 = (-\infty, a]$  und  $K_{N+1} = (b, \infty)$  betrachtet. Die Folge  $\{h_i(t)\}_{t=0}^{\infty}$ , die als

$$\begin{aligned} h_i(0) &= h_0, \quad 0 \leq h_0 \leq 1 \quad (\text{Startwert}) \\ h_i(t+1) &= h_i(t) + \frac{1}{t+1} \left( I_{K_i}(\xi_t) - h_i(t) \right) \end{aligned} \quad (2.34)$$

definiert ist, konvergiert mit Wahrscheinlichkeit Eins und es gilt

$$\lim_{t \rightarrow \infty} h_i(t) = \begin{cases} F_{\xi}(a), & \text{wenn } i = 0 \\ F_{\xi}(a+ih) - F_{\xi}(a+(i-1)h), & \text{wenn } i \in \{1, \dots, N\} \\ 1 - F_{\xi}(b), & \text{wenn } i = N+1 \end{cases} \quad \mathcal{P}\text{-f.s.} \quad (2.35)$$

Beweis:

Der Satz 1.1 wird für den Fall angewandt, daß

$$R(x) = \mathcal{P}(\xi_1 \in K_1) - x \quad \text{und} \quad G(t+1, x, \omega) = |_{K_1}(\xi_{t+1}(\omega)) - \mathcal{P}(\xi_1 \in K_1)$$

sind. Die Eigenschaften (1.2)-(1.6) können einfach nachgewiesen werden, denn es gilt:  $E|_{K_1}(\xi_t(\omega)) = \mathcal{P}(\xi_1 \in K_1)$  und  $|R(x)| \leq 1 + |x|$ .  $\square$

Mit Hilfe der Folgen  $\{h_i(t)\}$ ,  $i = 0, \dots, N+1$ , kann das relative Histogramm

$$H_t(x) = \frac{1}{h} \sum_{i=0}^N h_i(t) |_{K_i}(x) \quad x \in \mathbb{R} \quad (2.36)$$

als eine Schätzung der Klassenwahrscheinlichkeiten definiert werden. Für  $t \rightarrow \infty$  und  $h \rightarrow 0$  konvergiert  $H_t$  mit Wahrscheinlichkeit Eins gegen  $f_\xi$  (vgl. z.B. [21]). Für jedes  $x \in \mathbb{R}$  ist das Histogramm eine Schätzung des Anstiegs von  $F_\xi$  in der Umgebung von  $x$ . Mit Hilfe des Histogramms wird zuerst eine Bestimmung der Steuerfolge für eine konsistente Schätzung des Quantils gezeigt.

**Satz 2.3:**

Man betrachte die Folge  $\{\xi_t\}_{t=0}^\infty$  von unabhängig identisch verteilten Zufallsgrößen mit der absolut stetigen streng monoton wachsenden Verteilungsfunktion  $F_\xi$ . Die Verteilungsdichte  $f_\xi$  ist also eine stetige positive Funktion. Wird in der Folge  $\{q_\alpha(t)\}_{t=0}^\infty$ , die als

$$\begin{aligned} q_\alpha(0) &= q \quad (\text{Startwert}) \\ q_\alpha(t+1) &= q_\alpha(t) + \frac{c_t}{(t+1)} \left( \alpha - |_{(-\infty, q_\alpha(t))}(\xi_{t+1}) \right) \end{aligned} \quad (2.37)$$

gegeben ist, die Folge  $\{c_t\}_{t=0}^\infty$  nach

$$c_t = \begin{cases} \frac{1}{H_t(q_\alpha(t))}, & \text{wenn } H_t(q_\alpha(t)) > 1/h\varepsilon \\ h\varepsilon \quad (\varepsilon > 1), & \text{sonst} \end{cases} \quad (2.38)$$

bestimmt, so konvergiert  $\{q_\alpha(t)\}_{t=0}^\infty$  mit Wahrscheinlichkeit Eins gegen das  $\alpha$ -Quantil  $q_\alpha$  der Verteilungsfunktion  $F_\xi$ .

Beweis:

Die Folge  $\{q_\alpha(t), H_t\}$  bildet einen zweidimensionalen Markow-Prozeß im Sinne von [47] §2.3, (s. Theorem 3.1 und die Bedingung A).

Für die Funktion  $H_t$  gilt, daß  $H_t(x) \leq 1/h$  für alle  $x \in \mathbb{R}$  ist. Wegen  $\varepsilon > 1$  gilt für alle  $t \geq 0$ , daß  $c_t \geq h$  ist.

Für den Markov-Prozeß  $\{q_\alpha(t)\}$  betrachtet man die Borel-meßbare nicht negative Funktion  $V(t,q) = (q - q_\alpha)^2$ . Für den erzeugenden Operator von  $\{q_\alpha(t)\}$  bzgl. der Funktion  $V$  kann die folgende Ungleichung gezeigt werden:

$$\begin{aligned} L(V(t,q)) &= \mathbf{E}[(q_\alpha(t+1) - q_\alpha)^2 - (q_\alpha(t) - q_\alpha)^2 | q_\alpha(t) = q] = \\ &= \frac{1}{t+1} (q - q_\alpha) \mathbf{E}c_t \left( \alpha - I_{(-\infty, q)}(\xi_t) \right) + \frac{1}{(t+1)^2} \mathbf{E}c_t^2 \left( \alpha - I_{(-\infty, q)}(\xi_t) \right)^2 \leq \\ &= \frac{h}{t+1} (q - q_\alpha) (\alpha - F_\xi(q)) + \frac{(h\varepsilon)^2}{(t+1)^2} (1 + (q - q_\alpha)^2). \end{aligned} \quad (2.39)$$

Damit konvergiert nach [47] §2.7, Theorem 7.1, die rekursive Schätzung (2.37) mit der Steuerfolge (2.38) gegen das  $\alpha$ -Quantil  $q_\alpha$  von  $F_\xi$   $\mathcal{P}$ -fast sicher.  $\square$

Für die adaptive Gestaltung eines rekursiven Verfahrens ist es von Bedeutung, den Korrekturterm so zu steuern, daß die Schätzfolge auf Strukturbrüche möglichst schnell reagiert, in stationären Abschnitten dennoch eine gute Schätzung liefert. Um diese Eigenschaft bei der Schätzung der Quantile mit einer Steuerfolge, wie sie im Satz 2.3 bestimmt wurde, zu erreichen, werden die adaptiven relativen Häufigkeiten wie folgt bestimmt:

$$\begin{aligned} h_i^c(0) &= 0 \quad (\text{Startwert}) \\ h_i^c(t) &= (1-c)h_i^c(t-1) + c I_{K_i}(\xi_t), \end{aligned} \quad (2.40)$$

$i = 0, \dots, N+1$  und  $0 < c < 1$ . Analog zu (2.36) wird dann die adaptive Schätzung des relativen Histogramms mit

$$H_t^c(x) = \frac{1}{h} \sum_{i=0}^N h_i^c(t) I_{K_i}(x) \quad x \in \mathbb{R}. \quad (2.41)$$

definiert. Ist  $\{\xi_t\}$  eine stark mischende Folge von Zufallsgrößen mit der Verteilungsfunktion  $F_\xi$ , so gilt für  $p_i = \mathcal{P}(\xi_1 \in K_i)$

$$\mathbf{E}h_i^c(t) = c \sum_{k=1}^t (1-c)^{t-k} \mathbf{E}I_{K_i}(\xi_k) = (1 - (1-c)^t) p_i \quad (2.42)$$

(vgl. (2.7)) und somit auch

$$\lim_{t \rightarrow \infty} E H_t^c(x) = \begin{cases} F_\xi(a)/h, & x \in K_0 \\ (F_\xi(a+ih) - F_\xi(a+(i-1)h))/h, & x \in K_i \text{ und } i \in \{1, \dots, N\} \\ (1 - F_\xi(b))/h, & x \in K_{N+1} \end{cases} \quad (2.43)$$

Sind die Zufallsgrößen unabhängig, kann nach dem Satz 2.1 für die Varianz der Schätzung gezeigt werden, daß

$$\lim_{t \rightarrow \infty} E(h_1^c - p_1)^2 = \frac{c}{2-c} p_1(1-p_1), \quad i = 0, 1, \dots, N+1 \quad (2.44)$$

ist.

Für ein beliebiges Signal  $X = \{x_t\}$  wird mit (2.40) ein  $(n+2)$ -dimensionaler Operator definiert, der die momentane Dichte der Werte von  $X$  in den Intervallen  $K_i$  ( $i = 0, \dots, N+1$ ) beschreibt.

Die Steuerfolge  $\{c_t\}$  eines lernfähigen Operators für den  $\alpha$ -Quantilwert wird mit

$$c_t = \frac{d}{\max \left\{ H_t^c(q_\alpha(t)), \frac{1}{h\varepsilon} \right\}} \quad (2.45)$$

definiert, wobei  $d \in (0, 1)$  und  $\varepsilon > 1$  sind. Es gilt für alle  $x \in \mathbb{R}$ , daß

$$H_t^c(x) = \sum_{i=0}^{N+1} \sum_{j=1}^t \frac{c(1-c)^{t-j}}{h} |_{K_i}(x_j) |_{K_i}(x) \leq \frac{1-(1-c)^t}{h} < \frac{1}{h}$$

ist (vgl. (2.7)). Durch eine feste Wahl der Konstanten  $c$ ,  $d$  und  $h$  ist eine untere Schranke der Steuerfolge gegeben:  $c_t > h \cdot d$ . Der Wert von  $c_t$  liegt diesem Wert am nächsten, wenn die Werte der Folge  $X$  dicht um den aktuellen Schätzwert  $q_\alpha(t)$  schwingen. Mit wachsender Streuung von  $X$  nimmt die Steuerfolge zu, die Streuung der Schätzung steigt an. Der maximale Anstieg der Schätzfolge  $h d \varepsilon$  ( $c_t \leq h d \varepsilon$ ), der mit der Wahl von  $\varepsilon$  festgelegt ist, wird an den Stellen der Strukturbrüche von  $\{x_t\}$  erreicht. Mit einer Steuerfolge  $\{c_t\}$  nach (2.45) gewinnt also die adaptive Schätzung von Quantilen ähnliche Eigenschaften wie die Schätzungen, in denen die Folge  $c_t$  zwischen zwei konstanten Werten gesteuert wird, wie es mit (2.9) beschrieben wurde.

Für die Verfahren der stochastischen Approximation wurde ebenfalls der Zusammenhang zwischen dem Anstieg der Funktion  $R$  in der Umgebung der Null-

stelle  $r$  und der Schätzung dieser Nullstelle untersucht. In [47] wurde gezeigt, daß die Varianz der rekursiven Schätzung aus dem Satz 1.1 für mathematische Stichproben von der Wahl der positiven Konstante  $c$  in  $c_t = \frac{c}{t+1}$  abhängt (vgl. §6.6). Sie ist am kleinsten, wenn  $c = -(R'(r))^{-1}$  ist (vgl. §7.5). Für die Schätzung (1.10) des  $k$ -ten Momentes  $\mu_k$  ist die Ableitung der Funktion  $R(x) = \mu_k - x$  konstant gleich  $-1$  und damit die Schätzung mit  $c_t = \frac{1}{t+1}$  die wirksamste. Im allgemeinen ist aber ein solches  $c$  nicht bekannt, es werden jedoch Verfahren untersucht, in denen eine Schätzung für  $R'(r)$  möglich ist. Im folgenden wird ein Satz angegeben, dessen Anwendung zur wirksamsten Schätzung der Nullstelle  $r$  von  $R$  führt.

**Satz 2.4:**

Sei  $\{\xi_t\}_{t=0}^{\infty}$  eine Folge unabhängiger identisch verteilter Zufallsgrößen. Für eine Funktion  $R$  gelten die Bedingungen (1.3) und (1.5). Es existieren zwei reelle Zahlen  $-\infty < r_1 < r_2 < 0$  mit  $r_1 < R'(x) < r_2$ . Die zweite Ableitung  $R''$  sei stetig und beschränkt. Die zufällige Funktion  $G(t,x,\omega)$  genüge den Bedingungen (1.2) und (1.4) und die Ungleichung

$$R^2(x) + EG^2(t,x,\omega) \leq C(1+x^2), \quad C > 0 \text{ und } x \in \mathbb{R}$$

sei erfüllt. Weiter sei eine reelle Folge  $\{a_t\}_{t=0}^{\infty}$  gegeben, für die gilt

$$\frac{a_1}{t^\varepsilon} < a_t < \frac{a_2}{t^\varepsilon}, \quad a_1, a_2 > 0, \quad 0 < \varepsilon < 1/2. \quad (2.46)$$

Man definiert die zufälligen Folgen  $Y = \{y_t\}_{t=0}^{\infty}$ ,  $Z = \{z_t\}_{t=0}^{\infty}$

$$y_t(x,\omega) = R(x) + G(t,x,\omega) \quad (2.47)$$

$$z_t(x,\omega) = \frac{y_t(x + a_t, \omega) - y_t(x - a_t, \omega)}{2a_t} \quad (2.48)$$

und einen zweidimensionalen Markov-Prozeß  $W = \{(v_t, x_t)\}_{t=0}^{\infty}$  mit

$$\begin{aligned} v_0 &= 0 \\ x_1 &= x \quad (\text{Starwert}) \\ v_t &= v_{t-1} + \frac{1}{t} (z_t(x_t, \omega) - v_{t-1}) \end{aligned} \quad (2.49)$$

$$x_{t+1} = x_t + \frac{1}{tv_t^\circ} \frac{y_{t+1}(x_t + a_t, \omega) + y_{t+1}(x_t - a_t, \omega)}{2} \quad (2.50)$$

und

$$v_t^\circ = \begin{cases} -v_t, & v_t < 0 \\ 1, & \text{sonst.} \end{cases} \quad t = 1, 2, \dots$$

Der Prozeß  $W$  konvergiert mit Wahrscheinlichkeit Eins gegen  $(R'(r), r)$ .

Für den Beweis vgl. [47] §7.5, Theorem 5.2. In §7.6, Theorem 6.1 und 6.2, wurde gezeigt, daß die Schätzung für  $\varepsilon > 1/4$  die kleinste Varianz besitzt.

Dieser Satz kann sofort für die wirksamste rekursive Schätzung von Quantilwerten angewandt werden.

**Satz 2.5:**

Man betrachte die Folge  $\{\xi_t\}_{t=0}^\infty$  unabhängiger identisch verteilter Zufallsgrößen mit der absolut stetigen streng monoton wachsenden Verteilungsfunktion  $F_\xi$ . Die Dichtefunktion  $f_\xi$  sei differenzierbar, ihre erste Ableitung sei stetig und beschränkt. Es existieren zwei reelle Zahlen  $-\infty < f_1 < f_2 < 0$  mit  $f_1 < f_\xi(x) < f_2$  für alle  $x$ . Dann konvergiert die Folge  $W = \left\{ (f_t, q_\alpha(t)) \right\}_{t=0}^\infty$ , die für ein  $\varepsilon$  mit  $0 < \varepsilon < 1/2$  mittels der Vorschriften

$$\begin{aligned} f_0 &= 0 \\ q_\alpha(1) &= q \end{aligned}$$

und

$$f_t = \left(1 - \frac{1}{t}\right) f_{t-1} + \frac{1}{2t^{1-\varepsilon}} \left[ q_\alpha(t) - t^{-\varepsilon}, q_\alpha(t) + t^{-\varepsilon} \right] (\xi_t) \quad (2.51)$$

$$q_\alpha(t+1) = q_\alpha(t) + \frac{1}{tf_t^\circ} \begin{cases} \alpha - 1, & \xi_{t+1} < q_\alpha(t) - t^{-\varepsilon} \\ \alpha - \frac{1}{2}, & q_\alpha(t) - t^{-\varepsilon} \leq \xi_{t+1} < q_\alpha(t) + t^{-\varepsilon} \\ \alpha, & q_\alpha(t) + t^{-\varepsilon} \leq \xi_{t+1}, \end{cases} \quad (2.52)$$

$$f_t^\circ = f_t^\circ = \begin{cases} f_t, & f_t > 0 \\ 1, & \text{sonst} \end{cases}$$

definiert ist, gegen  $(f_\xi(q_\alpha), q_\alpha)$   $\mathcal{P}$ -fast sicher.

Beweis:

Daß die Funktionen  $R(q) = \alpha - F_{\xi}(q)$  und  $G(t+1, q, \omega) = F_{\xi}(q) - I_{(-\infty, q]}(\xi_{t+1}(\omega))$  die Bedingungen (1.2) - (1.5) erfüllen, wurde schon im Beweis von Satz 1.3 gezeigt. Es gilt, daß  $R^2(q) + EG^2(t, q, \omega) \leq 2 < 2(1+x^2)$  ist. Da  $F_{\xi}$  streng monoton wachsend ist, gilt für die erste Ableitung  $R'(q) = -f_{\xi}(q) < 0$ . Die reelle Folge  $t^{-\varepsilon}$ ,  $0 < \varepsilon < 1/2$  genügt der Bedingung (2.46). Damit sind die Voraussetzungen vom Satz 2.4 erfüllt. Die Folge  $Y$ , die in (2.47) definiert wurde, ist mit

$$y_t(q, \omega) = \alpha - I_{(-\infty, q]}(\xi_{t+1}(\omega))$$

gegeben. Für die Folge  $Z$  aus (2.48) ergibt sich in diesem Fall

$$z_t(q, \omega) = \frac{-I_{[q-t^{-\varepsilon}, q+t^{-\varepsilon})}(\xi_t(\omega))}{2t^{-\varepsilon}}$$

und für den Korrekturterm in (2.50)

$$\frac{y_{t+1}(q_{\alpha}(t)+t^{-\varepsilon}, \omega) + y_{t+1}(q_{\alpha}(t)-t^{-\varepsilon}, \omega)}{2} = \begin{cases} \alpha - 1, & \xi_{t+1} - q_{\alpha}(t) < -t^{-\varepsilon} \\ \alpha - \frac{1}{2}, & -t^{-\varepsilon} \leq \xi_{t+1} - q_{\alpha}(t) < t^{-\varepsilon} \\ \alpha, & \xi_{t+1} - q_{\alpha}(t) \geq t^{-\varepsilon} \end{cases}$$

Die Konvergenz der Folgen (2.51) und (2.52) ist nach dem Satz 2.4 bewiesen.  $\square$

Nach den Überlegungen zur adaptiven Gestaltung der Schätzung aus dem Satz 2.3 können auch für die Schätzung aus dem Satz 2.5 ähnliche Eigenschaften erreicht werden, in dem sie wie folgt modifiziert wird. Seien  $c, d$  aus  $(0, 1)$ ,  $h > 0$ . Dann wird eine adaptive lernfähige Schätzung des Quantils folgendermaßen definiert:

$$f_0^c = 0$$

$$q_{\alpha}^d(1) = q$$

und

$$f_t^c = (1-c)f_{t-1}^c + \frac{c}{2h} I_{[q_{\alpha}^d(t)-h, q_{\alpha}^d(t)+h)}(\xi_t), \quad (2.53)$$

$$q_{\alpha}^d(t+1) = q_{\alpha}^d(t) + \frac{d}{f_t^c} \begin{cases} \alpha - 1, & \xi_{t+1} < q_{\alpha}^d(t) - h \\ \alpha - \frac{1}{2}, & q_{\alpha}^d(t) - h \leq \xi_{t+1} < q_{\alpha}^d(t) + h \\ \alpha, & q_{\alpha}^d(t) + h \leq \xi_{t+1}, \end{cases} \quad (2.54)$$

wobei  $\hat{f}_t^c = \max \left\{ f_t^c, \frac{1}{2h\varepsilon} \right\}$  und  $\varepsilon > 1$  sind. Es kann mit Hilfe von (2.7) gezeigt werden, daß  $f_t^c < 1/2h$  ist:

$$f_t^c = \sum_{j=1}^t \frac{c(1-c)^{t-j}}{2h} \left| [q_{\alpha}^d(j-h), q_{\alpha}^d(j+h)](\xi_j) \right| \leq \frac{1-(1-c)^t}{2h} < \frac{1}{2h} .$$

Ähnlich wie in der Schätzung mit der Steuerfolge nach (2.45) sind bei einer festen Wahl der Konstanten  $c$ ,  $d$ ,  $h$  und  $\varepsilon$  eine untere und eine obere Schranke der Steuerfolge gegeben:  $2hd < c_t \leq 2hd\varepsilon$ .

In beiden adaptiven Verfahren, in denen die Steuerfolge  $\{c_t\}$  vom Signal abhängig ist, steuert der Faktor  $d$  die Dynamik der Schätzung, während die Schätzung des Anstiegs die Anpassung an die Zeitreihe realisiert. Die Konstante  $h$  ist in Abhängigkeit von der Streuung der untersuchten Zeitreihen zu wählen. Untersuchungen an simulierten Zeitreihen zeigen, daß der Faktor  $c$  in der Anstiegsschätzung nicht kleiner der Faktor  $d$  gewählt werden soll. Dadurch wird eine schnelle Anpassung der Schätzung nach einem Strukturbruch erreicht. Ein ähnliches Verhalten wurde schon in der adaptiven Schätzung der ARMA-Parameter im vorangegangenen Abschnitt beobachtet.

An einem Beispiel sollen die adaptiven Verfahren der Quantilschätzung verglichen werden. Dafür wurde eine Folge von gleichverteilten Pseudozufallszahlen  $\{x_t\}_{t=1}^{1000}$  simuliert, wobei  $x_1 - x_{230}$  im Intervall  $[0,1]$ ,  $x_{231} - x_{730}$  im Intervall  $[6,8]$  und  $x_{731} - x_{1000}$  im Intervall  $[2,5]$  ausgewürfelt wurden. Danach wurde das Interquartilintervall, also das 1. und 3. Quartil, geschätzt. Die Ergebnisse sind in der Abbildung 2.23 dargestellt. Für die adaptive Schätzung mit konstantem Adaptionsfaktor  $c = 0.1$  konnten in den stationären Abschnitten gute Ergebnisse erreicht werden. Dagegen erfolgte mit diesem Wert von  $c$  die Anpassung an einen neuen Abschnitt sehr langsam. In den Schätzungen mit einer lernenden Steuerfolge konnte nach einem Strukturbruch eine sehr rasche Anpassung erreicht werden, wobei in den stationären Abschnitten eine gute Schätzung erreicht wurde. Da in beiden Verfahren der Adaptionsfaktor  $d = 0.05$  gewählt wurde, besitzen beide Schätzungen in stationären Abschnitten annähernd die gleiche Streuung. Die Schätzfolgen (2.53) und (2.54) sind funktional verbunden und die Anpassung der Quantilschätzung (2.54) kann nur so schnell erfolgen, wie schnell die Folge (2.53) an der Stelle eines Strukturbruchs gegen Null konvergiert. Dagegen erfolgt in (2.45) die Histogrammschät-

zung unabhängig von der Schätzfolge des Quantils und eine Anpassung an den neuen Abschnitt kann schneller erreicht werden.

Die empirischen Untersuchungen zeigen, daß eine dem Signal angepaßte Bestimmung der Steuerfolge  $c_t$  die Eigenschaften der adaptiven Schätzung von Quantilen wesentlich verbessern kann. Werden die beiden Methoden verglichen, so ist sofort ersichtlich, daß die adaptive Schätzung (2.53)-(2.54) sich rechentechnisch einfacher realisieren läßt. Untersuchungen an simulierten Zeitreihen mit einer bekannten Verteilungsdichte zeigen aber, daß diese Schätzung stark von der Wahl der Konstanten  $h$  abhängt. Mit wachsendem Wert von  $h$  wird zwar die Streuung der Schätzfolge geringer, für einen "zu großen Wert" von  $h$  schwingt aber die Schätzfolge um einen falschen Wert. Eine optimale Einstellung muß aus Voruntersuchungen für jedes  $\alpha$  bestimmt werden. Dagegen bringt eine Schätzung mit der Steuerfolge nach (2.45) schon für wenige Häufigkeitsklassen gute Ergebnisse. In der Abbildung 2.22 wird ein Vergleich der Schätzung mit einer konstanten Steuerfolge und mit der Steuerfolge nach (2.45) für Biosignale dargestellt.

### 3 EINSATZMÖGLICHKEITEN ADAPTIVER VERFAHREN IN DER BIOSIGNALANALYSE

Die praktische Relevanz der adaptiven rekursiven Verfahren konnte bei der Lösung verschiedener Aufgaben aus der technischen Diagnose nachgewiesen werden (vgl. [22]). Ihre Anwendung bei der Analyse verschiedener Biosignale wurde im vorangegangenen Kapitel sowie in den Arbeiten [23], [55], [24], [27] u.a. angesprochen. Einige Lösungen konkreter Aufgaben werden in diesem Kapitel genauer beschrieben.

#### 3.1 Einsatzmöglichkeiten der adaptiven ARMA-Modellierung

Anknüpfend an die Untersuchungen aus dem Abschnitt 1.2.3, die der ARMA-Modellierung von Biosignalen gewidmet wurden, können die bekannten Auswertemethoden auf die dynamische Analyse von Zeitreihen erweitert werden. Die adaptive rekursive Schätzung von ARMA-Parametern ermöglicht es, die spektralen Eigenschaften eines Signals in ihrer zeitdynamischen Entwicklung zu beobachten und zu analysieren. Da mit der dynamischen ARMA-Modellanpassung zu jedem Zeitpunkt ein vollständiges Modell eines beliebigen Signals zur Verfügung gestellt werden kann, lassen sich mit Hilfe der Operatoren (2.28)-(2.31) unterschiedliche Parameter sowohl im Zeit- als auch im Spektralbereich gewinnen. Die Effektivität der verschiedenen Parameter zur Lösung konkreter Aufgabenstellungen muß jedoch im einzelnen erprobt und untersucht werden.

In der Abbildung 3.1 wird das Spektrum des Signals aus der Abbildung 2.19 in der ersten Ruhephase am 200., in der Kauphase am 400. und in der darauf folgenden Ruhephase im letzten Meßpunkt dargestellt. Die typische Kauaktivität im Frequenzband von 55 Hz bis 98 Hz ist im mittleren Spektrum über die gesamte Zeitreihe, das mit konvergenten Verfahren bestimmt wurde, deutlich (vgl. Abb. 1.8). Sie ist im dynamischen Spektrum am Ende der Kauphase eindeutig erkennbar, am Ende der Messung verschwindet sie allmählich.

Im Abschnitt 1.3.4 wurde die Mappingtechnik als eine bildgebende Auswertemethode eingeführt und in der Abbildung 1.9 an der hier erwähnten Muskelaktivität im Kauprozeß veranschaulicht. Mit Hilfe adaptiver Verfahren läßt sich die dynamische Entwicklung vom Aufbau der Kauphase bis zur Rückkehr in

die Ruhephase detailliert in den aufeinanderfolgenden Maps zeigen (siehe Abbildung 3.2).

Die Berechnung der Spektraldichte in jedem Zeitpunkt (bzw. in sehr nahe liegenden Zeitpunkten) führt zu einer dreidimensionalen Darstellung des Signalspektrums in der Zeit-Frequenz-Ebene. In [52] wurde ein Beispiel der Mikrozirkulationsmessung mittels Laser-Doppler-Flowmetrie bearbeitet. Dabei wurde ein dreidimensionales Spektralgebilde gewonnen, in dem die Veränderung typischer Rhythmen zu erkennen ist (vgl. Abbildung 3.3).

Der mittlere quadratische Vorhersagefehler eines ARMA-Prozesses kann zur Detektion von Instationaritäten im Signal verwendet werden. Benutzt man zu seiner Schätzung die adaptive Berechnungsformel (2.31), so reagiert die Folge  $\{\hat{\sigma}_e^2(t)\}$  auf abrupte Änderungen im Signal mit starkem Zuwachs, bei allmählichen Änderungen bzw. nach einem Strukturbruch verläuft sie konstant (vgl. [56]). Die Geschwindigkeit der Reaktion und der Anpassung kann mit dem Adaptionfaktor gesteuert werden.

Bei der Anwendung eines adaptiven rekursiven Verfahrens ermöglicht die Wahl des Adaptionfaktors bzw. der Adaptionfaktoren die Robustheit und Adaptivität des Verfahrens zu steuern und ihn an die Eigenschaften der untersuchten Zeitreihe anzupassen. Für eine konkrete, möglicherweise automatisch gesteuerte Anwendung muß die Einstellung der Adaptionfaktoren weiteren Untersuchungen unterzogen werden. Dabei soll der Zusammenhang zwischen den Eigenschaften der Zeitreihe und dem Verhalten des gewählten Verfahrens überprüft werden.

Im folgenden Abschnitt wird gezeigt, wie die Anwendung der adaptiven Trend-schätzung bei einer Berechnung der Durchblutungsgeschwindigkeit am Augenhintergrund gegenüber aufwendigen Methoden zu sehr guten Ergebnissen führen kann. Nutzt man die Spektraleigenschaften des Trendoperators aus, so kann ein günstiger Adaptionfaktor exakt berechnet werden. Das ermöglicht, geeignete Lösungsalgorithmen für eine automatische Auswertung zu entwickeln.

Die Erkennung von einzelnen Mustern und Strukturen, die im Meßsignal charakteristische Vorgänge am untersuchten Objekt widerspiegeln, ist eine wichtige Aufgabe in der Signalanalyse. Die Anwendung der dynamischen Kenngrößen, die im Abschnitt 2.3 beschrieben wurden, und eines adaptiven Schwellenoperators bei der adaptiven Segmentierung von Langzeit-EEG-Signalen wird im letzten Abschnitt untersucht.

### 3.2 Anwendung der adaptiven Trendschätzung zur Berechnung der Durchblutungsgeschwindigkeit in retinalen Gefäßen des Auges

Zur Messung der arteriellen Blutgeschwindigkeit in den kleinen Arterien des Auges findet unter anderem die Indikator-technik Anwendung. Ein fluoreszierender Farbstoff wird intravenös injiziert und erscheint nach einigen Sekunden in den Netzhautgefäßen. Die Passage des Farbstoffes durch die Netzhautgefäße kann photoelektrisch aufgezeichnet werden. Registriert man über einem arteriellen Netzhautgefäß mit zwei Meßflächen im Abstand  $\Delta s$  voneinander den zeitlichen Verlauf des die Indikator-konzentration repräsentierenden Fluoreszenzlichtes, ergeben sich zwei Meßkurven.

Die komplexen Zusammenhänge zwischen einer mittleren Blutgeschwindigkeit und den Zeitdifferenzen zwischen gleichen Signalwerten der Meßkurven wurden in einem Modell unter Berücksichtigung nichtlinearer Strömungsverhältnisse, Absorption von Erreger- und Fluoreszenzlicht durch verschiedene Blutbestandteile, der Indikatorverteilung und anderer Einflußgrößen erfaßt (vgl. [65]). Experimentelle Untersuchungen belegen, daß die Beziehungen zwischen Zeitdifferenzen in den beiden Meßkurven und in der mittleren Blutgeschwindigkeit zwischen den beiden arteriellen Meßstellen eine unerwartet gute Näherung der realen Verhältnisse darstellen. Über die Modellbeziehung kann für beliebige Meßbedingungen aus ermittelten Zeitdifferenzen zwischen gleichen Signalwerten beider Meßkurven die mittlere Blutgeschwindigkeit berechnet werden.

Aus dem diskutierten Modell folgt, daß bei normalen Injektions- und Meßbedingungen im Bereich des Signalanstieges ein in ausreichender Näherung linearer Anstieg existiert, dessen Zeitdifferenzen zwischen gleichen Signalwerten der beiden aufeinander im Signalmaximum normierten Meßkurven in einem großen Signalbereich (Signalwerte etwa zwischen 20% bis 80%) konstant sind und in einem definierten Faktor zu der über den Gefäßquerschnitt mittleren Zeitdifferenz aller Blutbestandteile stehen. Aus diesem Zusammenhang läßt sich die mittlere Blutgeschwindigkeit zwischen den Meßflächen berechnen. Die Aufgabe des Auswerteverfahrens für die Meßkurven besteht nunmehr darin, die Zeitdifferenz zwischen den Signalwerten im Bereich dieses linearen Anstieges zu bestimmen.

Die zeitlichen Verläufe der photoelektrisch erfaßten Fluoreszenzlichtströme werden aber durch Störsignale überlagert. Die Meßkurven erhalten konstante

Anteile, die die Signale gegen die Nulllinie verschieben. Sie werden mit unterschiedlichen Faktoren gestaucht und mit für beide Signale verschiedenen zufälligen Rauschsignalen überlagert. Die Störsignale verzerren die Meßkurven und überdecken die gesuchten Zeitdifferenzen zwischen den gleichen Signalwerten beider Meßkurven. Zudem wird es problematisch, den Normierungsfaktor zu bestimmen. Konventionelle Methoden der Signalauswertung wie Mittelwertbildung, Regressionsrechnung, Filter u.a. sind nur bedingt oder nicht verwendbar und vor allem für die Zielstellung einer objektiven automatischen Auswertung der Meßkurven ungeeignet (vgl. z.B. [65]).

Nachfolgend werden adaptive Methoden der Signalverarbeitung untersucht, mit denen die Auswertung der Zeitdifferenz mit linearem Anstieg der Meßkurven ermöglicht wird. Vorgeschlagen wird die Anwendung der adaptiven rekursiven Schätzung eines linearen Trends aus dem Abschnitt 2.3. Zu diesem Ziel wird eine mathematische Beschreibung der Meßkurven hergeleitet.

Betrachtet werden zwei Kurven, von denen jede bis zu einer bestimmten Zeit konstant verläuft, ab diesem Zeitpunkt dann einen positiven Anstieg hat. Aus der Modellbeschreibung folgt, daß dieser Zeitpunkt sowie der Anstieg für beide Kurven verschieden sind. Es wird dennoch angenommen, daß die beiden Kurven vom Verlauf her den gleichen Charakter haben. Gesucht wird die zeitliche Differenz zwischen beiden Kurven im Anstiegsbereich. Man bezeichnet die Reaktionszeiten mit  $t_1$  und  $t_2$ , wobei  $0 < t_1 \leq t_2 < T$  gilt. Mit  $T$  wird die Meßdauer bezeichnet, bzw. die Zeit, in der sich die Meßkurven gemäß dem Modell verhalten, d.h. sie besitzen gleichen Verlauf, sind nur zeitlich verschoben, das Verhältnis der Anstiege ist konstant. Der charakteristische Meßabschnitt im Intervall  $[0, T]$  kann folgendermaßen modelliert werden. Es seien  $f_1$  und  $f_2$  zwei stetige reellwertige Funktionen, für die gilt:

$$\lim_{t \rightarrow t_1} f_1(t) = 0 \quad \text{und} \quad \lim_{t \rightarrow t_2} f_2(t) = 0, \quad (3.1)$$

$$f_1(t) = A f_2(t + \Delta t), \quad (3.2)$$

wobei  $A$  eine Normierungskonstante und  $\Delta t = t_2 - t_1$  eine Zeitverschiebung sind. Die Störungen, die im Meßvorhang entstehen, werden durch stationäre zufällige Prozesse  $X_1$  und  $X_2$  beschrieben, für die  $Ex_i(t) = 0$  ( $i=1,2$ ) gilt.

Zur Erfassung der unterschiedlichen Nulllinien der Meßkurven werden die reellen Zahlen  $k_1$  und  $k_2$  eingeführt. Die Indikatorlösungskurven können als Realisierungen der zufälligen Prozesse  $Y_1$  und  $Y_2$  mit

$$\begin{aligned} y_1(t) &= k_1 + f_1(t) |_{[t_1, T]}(t) + x_1(t) \\ y_2(t) &= k_2 + f_2(t) |_{[t_2, T]}(t) + x_2(t) \end{aligned} \quad (3.3)$$

betrachtet werden.

Zur Bestimmung der Indikatorlaufzeitdifferenz  $\Delta t$  wird angenommen, daß es im idealen Fall Intervalle  $[t_0, t^0]$  bzw.  $[t_0 + \Delta t, t^0 + \Delta t]$  mit  $t_1 < t_0 < t^0 < T$  gibt, in denen sich die Funktionen  $f_1$  bzw.  $f_2$  durch lineare Funktionen ausreichend genug approximieren lassen, d.h., die Messung kann in diesem Abschnitt mit

$$y_1(t) = b_1 t + k_1 + x_1(t) \quad \text{für } t \in [t_0, t^0] \quad (3.4)$$

bzw.

$$y_2(t) = b_2 t + k_2 + x_2(t) \quad \text{für } t \in [t_0 + \Delta t, t^0 + \Delta t]$$

erfaßt werden. Diese Möglichkeit der Vereinfachung wurde in [65] nachgewiesen. Aus den Gleichungen (3.2) und (3.4) folgt, daß in diesem Intervall

$$b_1 = A b_2 \quad (3.5)$$

und

$$\Delta t = \frac{k_1 - A k_2}{b_1} \quad (3.6)$$

gilt.

Zur Hervorhebung der Funktionen  $f_1$  und  $f_2$  aus den Meßkurven  $y_1$  und  $y_2$  wird vorgeschlagen, den Operator  $\mathcal{F}^c$  zur adaptiven Schätzung eines linearen Trends nach (2.17) anzuwenden. Von wesentlicher Bedeutung sind die Eigenschaften der Signale  $\mathcal{F}^c(Y_1)$  und  $\mathcal{F}^c(Y_2)$  und die Frage, welchen Einfluß die Wahl des Adaptionsfaktors  $c$  auf die Schätzung der Differenz  $\Delta t$  im Intervall  $[t_0, t^0]$  hat. Da die Trendschätzung erwartungstreu ist (vgl. [55]), gilt

$$\begin{aligned} E\mathcal{F}^c(Y_1)(t) &= b_1 t + k_1 \\ E\mathcal{F}^c(Y_2)(t + \Delta t) &= b_2 t + b_2 \Delta t + k_2 \end{aligned} \quad t \in [t_0, t^0].$$

Unter Berücksichtigung der Beziehungen aus der Formel (3.5) erhält man für die Differenz  $\Delta t$  aus (3.6):

$$E\mathcal{T}^c(Y_2)(t+\Delta t) - E\mathcal{T}^c(Y_1)(t) = Ab_2 \frac{k_1 - Ak_2}{b_1} + Ak_2 - k_1 = 0, \quad t \in [t_0, t^*].$$

Diese Gleichung gilt für jeden beliebigen Adaptionsfaktor  $c$ ,  $0 < c < 1$ . D.h., die Zeitverschiebung der Trendschätzung der beiden Meßkurven wird von der Wahl des Adaptionsfaktors nicht beeinflusst.

Nutzt man diese Eigenschaft der adaptiven Trendschätzung aus, so muß der Adaptionsfaktor  $c$  nicht so gewählt werden, daß die Trendschätzung sehr schnell den wahren Trend erfaßt, sondern man kann versuchen, das Störsignal jeweils möglichst stark zu unterdrücken und den Trend so gut wie möglich hervorzuheben. Dazu werden für die Wahl eines günstigen Faktors  $c$  die Spektraleigenschaften des Operators  $\mathcal{T}^c$  untersucht. Da er sich als lineare Kombination des Operators  $\mathcal{M}^c$  darstellen läßt (siehe (2.17)), kann er genauso wie dieser als ein rekursives Filter betrachtet werden. Über seine Übertragungseigenschaften gelten folgende Aussagen.

**Satz 3.1:**

Die Übertragungsfunktion des Filters  $\mathcal{T}^c$  besitzt die Darstellung:

$$h_c^T(\lambda) = \frac{b_0 - b_1 \cos(2\pi\lambda/\nu)}{a_0 - a_1 \cos(2\pi\lambda/\nu) + a_2 \cos(4\pi\lambda/\nu)}, \quad (3.7)$$

wobei die Koeffizienten durch

$$\begin{aligned} b_0 &= c^2(18 - 18c + 5c^2), \\ b_1 &= c^2(18 - 18c + 4c^2), \\ a_0 &= 4 + (4 - 3c)^2 + (2 - c)^2 (1 - c)^2, \\ a_1 &= 2(4 - 3c)(4 - 3c + c^2), \\ a_2 &= 4(2 - c)(1 - c) \end{aligned}$$

gegeben sind.

**Beweis:**

Der Operator der adaptiven Mittelwertschätzung  $\mathcal{M}^c$  beschreibt ein rekursives Filter erster Ordnung, dessen z-Transformierte durch

$$H_c^M(z) = \frac{c}{1 - (1-c)z} \quad \text{mit } z \in \mathbb{C}$$

gegeben ist (vgl. z.B. [40], [48]). Aus der Darstellung (2.17) ergibt sich aufgrund der Linearität für die z-Transformation des Operators  $\mathcal{T}^c$

$$H_c^T(z) = \frac{(2-c)c}{1 - (1-c)z} - \frac{(1-c)c}{2 - (2-c)z} = \frac{c((3-c) - (3-2c)z)}{2 - (4-3c)z + (2-c)(1-c)z^2}. \quad (3.8)$$

Seine Übertragungsfunktion ist dann die quadratische Norm der z-Transformation auf dem Einheitskreis, also

$$h_c^T(\lambda) = |H_c^T(e^{-2\pi i \lambda / \nu})|^2, \quad \lambda \in [0, \nu/2],$$

wobei  $\nu$  die Abtastfrequenz ist. Aus den obigen Betrachtungen resultiert nach mehreren Umformungen die Darstellung (3.7).  $\square$

Der Verlauf der Funktion  $h_c^T$  wird in der Abbildung 3.4 verdeutlicht. Aus der Kurvendiskussion ergibt sich, daß die Frequenzen aus dem Intervall  $(0, \lambda_0)$  verstärkt werden und am Rand des Intervalls unverändert bleiben. Dabei ist  $\lambda_0$  die Lösung der Gleichung

$$h_c^T(\lambda) = 1,$$

die ungleich Null ist. Wird für  $h_c^T$  die Darstellung aus (3.7) eingesetzt, so gilt für  $\lambda_0$  ebenfalls die Gleichung

$$(b_0 - a_0 - a_1) + (b_1 - a_1) \cos(2\pi \lambda_0 / \nu) - 2a_2 \cos^2(2\pi \lambda_0 / \nu) = 0.$$

Da eine der beiden Lösungen gleich Null ist (was aus (3.8) leicht zu zeigen ist), kann für die Frequenz  $\lambda_0$  gezeigt werden:

$$\cos \frac{2\pi \lambda_0}{\nu} = \frac{a_1 - b_1}{2a_2} - 1$$

und nach einigen Umformungen dann

$$\lambda_0 = \frac{\nu}{2\pi} \arccos \left( 1 - \frac{c^2}{2} \right). \quad (3.9)$$

Die Frequenz mit der größten Verstärkung  $\lambda^*$  liegt im lokalen Maximum der Übertragungsfunktion und sie läßt sich aus der Gleichung

$$(h_c^T)'(\lambda) = 0,$$

bzw.

$$(b_0 a_1 - b_1 a_0 - b_1 a_2) + 4b_0 a_2 \cos(2\pi \lambda / \nu) - 2b_0 a_2 \cos^2(2\pi \lambda / \nu) = 0. \quad (3.10)$$

bestimmen, wobei für die Lösung  $|\cos(2\pi \lambda^* / \nu)| < 1$  gilt.

Aus dem Modellansatz in [65] kann für die Funktionen  $f_1$  und  $f_2$  angenommen werden, daß sie sich im Analyseabschnitt wie lineare bzw. quadratische Funktionen verhalten. Für die gilt, daß sich ihr wesentlicher Leistungsanteil in der ersten Spektrallinie konzentriert. Die praktische Spektralanalyse der Indikatorlösungskurven bestätigt diese Annahme und zeigt Spektren, die ihr Maximum ebenfalls im ersten Spektralwert haben. Einige Beispiele wurden in [10] ausgeführt.

Um den zugrundeliegenden Trend hervorzuheben, ist der Adaptionfaktor  $c$  nun so zu wählen, daß das lokale Maximum der Übertragungsfunktion  $h_c^T$  in der ersten Spektrallinie liegt, die sich als  $\Delta\lambda = 1/T$  aus der Spektralauflösung der Fouriertransformation im Analyseintervall ergibt. Die Berechnung eines optimalen Adaptionfaktors  $c$  erfolgt aus (3.10) für  $\lambda = \Delta\lambda$ , ist aber nur mit Hilfe aufwendiger numerischer Methoden realisierbar. Eine gute Approximation wird erreicht, wenn das Verstärkungsintervall auf das Spektralintervall  $[0, 2\Delta\lambda]$  beschränkt wird, d.h.:

$$h_{c_{\text{opt}}}^T(2\Delta\lambda) = 1. \quad (3.11)$$

Dann wird der erste Spektralpunkt als eine Approximation der Frequenz mit der größten Verstärkung betrachtet. Aus (3.9) und (3.11) ergibt sich

$$c_{\text{opt}} = \sqrt{2(1 - \cos(4\pi/T))}. \quad (3.12)$$

Abschließend soll die Varianz der Trendschätzung untersucht werden. Man betrachte eine Zeitreihe  $\{y_t\}$  mit

$$y_t = a + bt + x_t$$

mit  $E x_t = 0$  und  $a, b \in \mathbb{R}$ . Es ist ersichtlich, daß die Varianz von  $y_t$  und  $x_t$  identisch ist. Weiter gilt, daß

$$\mathcal{F}^c(Y)(t) = a + bt + \mathcal{F}^c(X)(t)$$

ist und somit

$$E(\mathcal{F}^c(Y)(t) - Ey_t)^2 = E(\mathcal{F}^c(X)(t))^2. \quad (3.13)$$

Ist  $\{x_t\}$  unkorreliert, so kann die Folge  $\{\mathcal{F}^c(X)(t)\}$  für ein genügend großes  $t$  als ein ARMA(2,1)-Prozeß mit der Spektraldichtefunktion  $\frac{\sigma^2}{v} h_c^T$  und der Varianz

$$\text{var}(\mathcal{F}^c(X)(t)) = \frac{\sigma^2}{v} \int_0^v h_c^T(\lambda) d\lambda, \quad (3.14)$$

betrachtet werden, wobei mit  $\sigma^2$  die Varianz von  $x_t$  bezeichnet wird. Bezüglich (3.13) und (3.14) gilt für die asymptotische Varianz des Operators  $\mathcal{F}^c$ , der auf eine mit unkorreliertem Rauschen überlagerte lineare Funktion angewandt wurde, die folgende Formel

$$\lim \mathbf{E}(\mathcal{F}^c(Y)(t) - \mathbf{E}y_t)^2 = c \left( \frac{9}{4-c} - \frac{4}{3-c} \right) \sigma^2 \approx c\sigma^2. \quad (3.15)$$

Um die Differenz  $\Delta t$  aus den gefilterten Signalen  $\mathcal{F}^c(y_1)$  und  $\mathcal{F}^c(y_2)$  bestimmen zu können, ist es notwendig, die Konstanten  $k_1$ ,  $k_2$  und  $A$  zu schätzen. Für dieses Ziel können die spektralen Eigenschaften der Indikatorlösungskurven in folgender Weise genutzt werden. Unterscheiden sich die gemessenen Kurven nur durch eine zeitliche Verschiebung, sind ihre Spektraldichtefunktionen identisch. Die unterschiedlichen Nulllinien spiegeln sich in einer unterschiedlichen Größe der nullten Spektrallinie. Der quadratische Quotient beider Anstiege ist äquivalent zum Quotienten beider Spektraldichten.

Seien  $z_1$  und  $z_2$  die Modellfunktionen nach der Trendfiltration und der Zentrierung im Analyseintervall:

$$z_i(t) = \mathcal{F}^c(Y_i)(t) - m_i, \quad (3.16)$$

wobei  $m_i$  das arithmetische Mittel von  $\mathcal{F}^c(Y_i)$  bezeichnet ( $i = 1, 2$ ). Durch die Zentrierung wird für die Fouriertransformation  $Z_i$  von  $z_i$

$$Z_1(0) = Z_2(0) = 0 \quad (3.17)$$

erreicht und damit auch der Wert Null in der nullten Spektrallinie für beide Meßkurven. Damit gilt für die Nulllinien  $k'_1$  und  $k'_2$  der zentrierten Kurven

$$k'_1 = A k'_2. \quad (3.18)$$

Schließlich gilt auch für die Signale  $z_1$  und  $z_2$  die Beziehung

$$\mathbf{E}z_1(t) = \mathbf{E}z_2(t+\Delta t). \quad (3.19)$$

Zusätzlich gilt aufgrund der Linearität der Fouriertransformation und der Zeitinvarianz der Spektraldichtefunktion die Gleichung

$$|Z_1(\lambda)|^2 = A^2 |Z_2(\lambda)|^2, \quad \lambda \in (0, \nu/2). \quad (3.20)$$

Die Konstante  $A$  kann also aus dem Quotienten beliebiger Paare von Fourierkoeffizienten bestimmt werden. Wie schon erwähnt wurde, ist die wesentliche Leistung der Indikatorlösungskurven in der ersten Spektrallinie konzentriert. Durch die Filtration  $\mathcal{F}^c$  wird die Leistung in dieser Frequenz (für beide Kur-

ven gleich) hervorgehoben, die restliche Leistung wird gedämpft. Um numerische Fehler zu vermeiden, empfiehlt es sich also, die Konstante  $A$  aus

$$A = \frac{|Z_1(\Delta\lambda)|}{|Z_2(\Delta\lambda)|} \quad (3.21)$$

zu bestimmen.

Die Schätzung von  $\Delta t$  kann mit den Methoden, die in [10] dargelegt und diskutiert wurden, als mittlere Zeitdifferenz der Signale  $z_1$  und  $Az_2$  ermittelt werden. Ein Beispiel der Auswertung von einer Indikatorlaufzeitmessung (siehe Abbildung 3.5) nach diesem Verfahren wird in der Abbildung 3.6 demonstriert.

### 3.3 Adaptive Segmentation von Langzeit-EEG-Signalen

#### 3.3.1 Aufgabenstellung, bekannte Lösungsansätze

Bei gewissen Erkrankungen, z.B. Schlafstörungen, Epilepsie u.a., ist die Untersuchung von Langzeit-EEG-Messungen von Bedeutung. Es handelt sich um mehrkanalige Signalaufnahmen von etwa 20 Minuten bis zu mehreren Stunden, deren Abtastrate zwischen 100 und 200 Hz liegt. Die Auswertung solcher Aufnahmen wird üblicherweise von einem erfahrenen Neurologen durchgeführt, indem er visuell nach diagnostisch interessanten Mustern und Strukturen sucht. Eine solche Analyse ist sehr mühsam und stark subjektiv, da sie von den empirischen Erfahrungen aber auch von dem Konzentrationsvermögen des Arztes abhängt. Eine automatische Vorverarbeitung solcher Aufnahmen kann diesen Prozeß unterstützen und objektivieren, wenn dem Arzt die in der Aufnahme enthaltene Information in einer komprimierten und überschaubaren Form angeboten wird. Die diagnostische Beurteilung der Daten wird weiterhin dem Arzt überlassen. In [13] wurde ein Konzept der automatischen Auswertung entwickelt, in dem das EEG-Signal in sogenannte quasistationäre Abschnitte aufgeteilt wird, ähnliche Abschnitte zusammengefaßt und statistisch ausgewertet werden. Es wurde von vielen Autoren weiter verfolgt und die einzelnen Verarbeitungsschritte methodisch verbessert. In diesem Abschnitt werden nur die Methoden der Zerlegung von EEG-Signalen, die in der Literatur als adaptive Segmentation bezeichnet sind, untersucht. Es wird gezeigt, wie dabei die adaptiven rekursiven Verfahren angewandt werden können.

In der Prozedur der adaptiven Segmentation wird vorausgesetzt, daß ein EEG-Signal die Realisierung eines zufälligen Prozesses ist und sich mit Hilfe statistischer Methoden beschreiben läßt. Weiter wird angenommen, daß dieses Signal aus einzelnen Abschnitten besteht, die sich durch ein konstantes Kurzzeitspektrum auszeichnen, und daß zwischen den einzelnen Abschnitten eine sprunghafte Änderung der charakteristischen Kenngrößen erfolgt. Man spricht von einem stückweise- oder quasistationären Verhalten (vgl. auch [14]). Bei der Auswertung eines Langzeit-EEG-Signals ist es von Bedeutung, die einzelnen Abschnitte voneinander zu trennen, damit sie in weiteren Untersuchungen anhand von passenden Strukturmodellen bewertet werden können. Statistische Methoden, die die Stationarität der gesamten Zeitreihe voraussetzen, können bei dieser Aufgabe nicht eingesetzt werden, da sie bei einer konkreten Auswertung nur gemittelte Kenngrößen liefern, in denen Informationen über die einzelnen Muster verloren gehen. Es erweist sich aber als sinnvoll, einzelne Signalabschnitte als Realisierung eines stationären Prozesses zu betrachten und zu ihrer Charakterisierung die bekannten Methoden anzuwenden.

Bei der Aufgabe, ein stückweise stationäres Signal  $X = \{x_t\}$  in einzelne stationäre Abschnitte zu zerlegen, hat sich die folgende Methode als erfolgreich erwiesen. Gesucht werden eine Schätzfunktion

$$G_{t,N} = G(x_{t-N+1}, \dots, x_t), \quad (3.22)$$

die das Signal in dem Intervall  $[t-N+1, t]$  beschreibt und eine symmetrische Vergleichsfunktion

$$V(t_1, t_2) = V\left(G_{t_1, N}, G_{t_2, N}\right) \quad (3.23)$$

mit  $V \geq 0$  und  $V(t_1, t_2) = V(t_2, t_1)$ , die den Unterschied zwischen zwei Abschnitten gleicher Länge zum Ausdruck bringt. Da sich ein quasistationärer Abschnitt durch ein konstantes Kurzzeitspektrum auszeichnet, kann  $G$  als eine Schätzung der Spektraldichtefunktion (s. [13], [61]) bzw. der Kovarianzfunktion (s. [14]) gewählt werden. In einem quasistationären Abschnitt sind die momentane Leistung und die momentane Frequenz ebenfalls konstant und so kann

$$G_{t,N} = (E_{t,N}, F_{t,N})^T, \quad (3.24)$$

zusammengesetzt werden, wobei  $E_{t,N}$  die mittlere Amplitude bzw. die mittlere Leistung und  $F_{t,N}$  die mittlere Frequenz des Signals im Intervall  $[t-N+1, t]$  beschreiben (s. [46], [64]). Dann ist

$$V(G_{t_1,N}, G_{t_2,N}) = g_E V(E_{t_1,N}, E_{t_2,N}) + g_F V(F_{t_1,N}, F_{t_2,N}), \quad (3.25)$$

wobei der Einfluß der Amplituden- und der Frequenzänderung durch die Gewichte  $g_E$  und  $g_F$  gegeben ist. Während in den früheren Arbeiten ein fester Referenzabschnitt ( $t_1 = t_0 + N$ ) mit einem Vergleichsabschnitt ( $t_2 = t_0 + N+1, \dots$ ) verglichen und ein signifikanter Zuwachs der Funktion  $V$  gesucht wurde (s. [13], [46], [14]), wird in [61] und [64] vorgeschlagen, zwei nacheinander folgende disjunkte Abschnitte zu beobachten. Dann besitzt die Funktion

$$W(t) = V(t, t+N), \quad t \in \mathbb{N} \quad (3.26)$$

einen Verlauf, in dem sich die Strukturbrüche im Signal als lokale Maxima der Funktion  $W$  widerspiegeln (vgl. [61]). Die Werte der Funktion  $W$  schwanken innerhalb eines stationären Abschnittes nahe Null. Liegt in dem aktuellen Vergleichsintervall  $[t-N+1, t+N]$  eine Strukturänderung vor, nimmt  $W$  signifikant zu, wobei sie an der Bruchstelle ein lokales Maximum erreicht. Da es sich aber um eine zufällige Funktion handelt, die zahlreichen Schwankungen unterliegt, besitzt sie auch in stationären Abschnitten keinen monotonen Verlauf. Für die Segmentation sind dann nur die lokalen Maxima der Funktion  $W$  von Bedeutung, die sich signifikant aus ihrer Umgebung herausheben.

Die Konstruktion der Schätzfunktion  $G$  und der Vergleichsfunktion  $V$  ist der Gegenstand vieler Studien (vgl. z.B. [13], [46], [14], [61], [64], [12] u.a.), wobei zwei gegeneinander wirkende Kriterien die Güte der Methode bestimmen. Zum einen ist es die Detektionsfähigkeit des Verfahrens und zum anderen die Geschwindigkeit der Berechnung. Dabei muß bei der gewählten Methode eine mehrkanalige parallele Auswertung möglich sein. Und schließlich kann die Wahl eines Schwellwertes, der die natürliche Schwankung der Funktionswerte  $W(t)$  von den signifikanten Änderungen trennt, die Segmentierungsprozedur wesentlich beeinflussen. In den vergleichenden Studien von Krajčá et al. (siehe [41]) wurde eine Kombination der erwähnten Verfahren ausgewählt, bei der eine ausreichende Detektion der sprunghaften Änderungen

in EEG-Signalen gewährleistet ist, wobei eine schnelle Verarbeitung von langen mehrkanaligen Datenaufnahmen möglich ist.

Die Kenngrößen aus [14] lassen sich in das obige Schema wie folgt einordnen. Man bezeichne mit  $X_{t,N}$  die Schätzung der Spektraldichte und mit  $R_{t,N}$  die Schätzung der Kovarianzfunktion im Abschnitt  $[t, t-N+1]$ . Dann ist

$$W(t) = \frac{\sum_{j=0}^{N/2} \left( X_{t,N}(j) - X_{t+N,N}(j) \right)^2}{\sum_{j=0}^{N/2} X_{t,N}(j) \cdot \sum_{j=0}^{N/2} X_{t+N,N}(j)} = \frac{\sum_{k=-K}^K \left( R_{t,N}(k) - R_{t+N,N}(k) \right)^2}{R_{t,N}(0) R_{t+N,N}(0)} \quad (3.27)$$

Die Funktion  $W$  bringt die relative Änderung im Signal zum Ausdruck. Die Normierung ermöglicht, einen festen Schwellwert  $\kappa$  zu wählen, so daß ein signifikanter Unterschied zwischen den Abschnitten dann vorliegt, wenn die Kurzzeitspektren sich mindestens um den Faktor  $\kappa$  unterscheiden, wobei auf der Stelle des lokalen Maximums über der Schwelle der Strukturbruch liegt. Die Methode ist gegen Änderungen bei schwacher Leistung besonders empfindlich, sie wird aber robuster, wenn man eine minimale Leistung zu den geschätzten Werten addiert. Bei einer günstigen Anzahl der Kovarianzkoeffizienten können mit dieser Methode sowohl für simulierte Beispiele als auch für reelle Daten sehr gute Segmentationsergebnisse erreicht werden. (Von den Autoren wurde  $K = 4$  vorgeschlagen.) Ähnlich wird im Verfahren aus [61] das Spektrum beider Abschnitte verglichen. Da hier der mittlere relative Unterschied zwischen den Spektralkoeffizienten nach

$$W(t) = \sum_{j=0}^{N/2} \frac{\left( X_{t,N}(j) - X_{t+N,N}(j) \right)^2}{X_{t,N}(j) \cdot X_{t+N,N}(j)}$$

betrachtet wird, muß demzufolge in jedem Schritt die Spektraldichte geschätzt werden. Dagegen können in (3.27) die Kovarianzkoeffizienten, die sich rekursiv aktualisiert lassen, zum Vergleich der Spektren angewandt werden. Aufgrund der rechentechnisch aufwendigen Vergleichsfunktion wird aber auch hier die Geschwindigkeit des Algorithmus trotz einer schnellen Berechnung der Kovarianz beeinträchtigt.

Eine wesentlich einfachere und sehr schnelle Methode liefert die Anwendung der Kenngrößen aus [64], die durch

$$\begin{aligned}
E_{t,N} &= \sum_{k=t-N+1}^t |x_k|, \\
F_{t,N} &= \sum_{k=t-N+1}^{t-1} |x_{k+1} - x_k|
\end{aligned}
\tag{3.28}$$

definiert sind. Als Vergleichsfunktion wird

$$W(t) = |E_{t,N} - E_{t+N,N}| + \rho |F_{t,N} - F_{t+N,N}|
\tag{3.29}$$

betrachtet, wobei  $\rho = 7$  für EEG-Signale empirisch ermittelt wurde. Da die Funktion  $W$  explizit von der Amplitude des Signals abhängt, wird eine Schwelle entweder empirisch am Anfang der Auswertung gesetzt oder als ein Teil des Mittelwertes der Werte von  $W$  über ein langes Intervall abschnittsweise definiert. Eine gute Detektion von Strukturbrüchen in beliebigen stochastischen Signalen konnte mit (3.28)-(3.29) nicht erreicht werden. Eine der Ursachen besteht darin, daß die Varianz der Funktion  $W(t)$  nach (3.29) mit der Varianz des Signals wächst. Das führt bei der Segmentierung mit einer konstanten Schwelle zu falschen Detektionen im Bereich hoher Signalleistung. Für die Segmentierung spezieller EEG-Signale zeigt sich dieser Vergleich als ausreichend (vgl. [41]) und wird wegen der einfachen und sehr schnellen Berechnung bevorzugt.

Die Untersuchungen zeigen, daß die Leistungsänderung im Signal relativ einfach mit Hilfe der Varianz erkannt werden kann. Die Kenngrößen zur Detektion der Frequenzänderung sind aber entweder rechentechnisch aufwendig oder wenig wirksam. In Anlehnung an die Definition der Momentanfrequenz im Abschnitt 2.3 wird vorgeschlagen, die Funktion  $F_{t,N}$  als mittlere Anzahl der Nulldurchgänge des zentrierten Signals nach

$$F_{t,N} = \frac{1}{N-1} \sum_{k=t-N+1}^{t-1} |_{(-\infty,0)}(x_{k+1} - \bar{x}_{t,N}) \cdot (x_k - \bar{x}_{t,N})|,
\tag{3.30}$$

zu gestaltet, wobei  $\bar{x}_{t,N}$  den empirischen Mittelwert im Intervall  $[t-N+1, t]$  bezeichnet. Mittels (3.30) wird eine einfache, dennoch im Gegensatz zur Schätzung nach [64] von der Amplitude des Signals unabhängige Schätzung der Frequenz, erreicht. Die Funktion  $E_{t,N}$  wird nach

$$E_{t,N} = \frac{1}{N} \sum_{k=t-N+1}^t (x_k - \bar{x}_{t,N})^2,
\tag{3.31}$$

berechnet. Bei der Konstruktion der Vergleichsfunktion  $V_E$  kann die folgende Überlegung nützlich sein. Für den speziellen Fall, daß  $\{x_t\}$  die Realisierung einer Folge unabhängiger identisch verteilter Zufallsgrößen mit dem Erwartungswert  $\mu$  und der Varianz  $\sigma^2$  ist, gilt für die Schätzfunktion (3.31)  $\text{Var}(E_{t,N}) = 2\sigma^4(N-1)/N^2$  (vgl. z.B. [21], Abschnitt 9.5). Würde die Vergleichsfunktion als absolute Differenz zwischen  $E_{t,N}$  und  $E_{t+N,N}$  definiert, so wäre  $\text{Var}|E_{t,N} - E_{t+N,N}| = 4\sigma^4(N-1)/N^2$  direkt abhängig von der Varianz der Zeitreihe. Andererseits gilt unter der zusätzlichen Annahme der Normalverteilung von  $x_t$ , daß die Funktion  $E_{t,N}/E_{t+N,N}$  die F-Verteilung besitzt und ihre Varianz nur von N abhängig ist. Aus dieser Überlegung und in Anlehnung an die Methoden aus [46] wird für die Vergleichsfunktion die folgende Berechnungsvorschrift vorgeschlagen:

$$V_E(t,t+N) = \frac{\max \langle E_{t,N}, E_{t+N,N} \rangle + E_0}{\min \langle E_{t,N}, E_{t+N,N} \rangle + E_0}, \quad (3.32)$$

wobei durch die Addition einer konstanten Leistung  $E_0$  die Stabilität und Robustheit der Schätzung erreicht wird. Denn für  $\min \langle E_{t,N}, E_{t+N,N} \rangle \leq E_0$  gilt

$$V_E(t,t+N) - 1 \leq \frac{1}{2} \left( \frac{\max \langle E_{t,N}, E_{t+N,N} \rangle}{\min \langle E_{t,N}, E_{t+N,N} \rangle} - 1 \right).$$

Der relative Vergleich wird auch für die Schätzfunktionen der Frequenz gewählt, um Unterschiede im nieder- und hochfrequenten Bereich gleichermaßen zu erkennen:

$$V_F(t,t+N) = \frac{\max \langle F_{t,N}, F_{t+N,N} \rangle}{\min \langle F_{t,N}, F_{t+N,N} \rangle}. \quad (3.33)$$

Die Stabilität der Schätzung ist durch die Vorfilterung der Meßdaten im medizinisch relevanten Spektralband 1.5 - 30 Hz gegeben. Die Funktionen  $V_E$  und  $V_F$  liegen im Intervall  $[1, \infty)$ . Mit der Funktion W, die als

$$W(t) = \max \langle V_E(t,t+N), V_F(t,t+N) \rangle - 1 \quad (3.34)$$

definiert wird, ist ein schärferes Maß der Unterschiede im Signal gegeben als mit dem mittleren Wert nach (3.25). Ein lokales Maximum einer der beiden Kenngrößen bleibt erhalten, solange der Unterschied in der anderen Größe nicht höher liegt. Eine Wichtung der beiden Größen ist weiterhin möglich. Die

Funktionen (3.30) und (3.31) liefern eine zuverlässige Schätzung der Leistung und der Frequenz, wobei die Berechnung schnell genug erfolgt. Da die Frequenzschätzung eindimensional ist, kann auch die Vergleichsfunktion effizient berechnet werden.

Die Methoden können an einem Beispiel verglichen werden (siehe Abbildung (3.7)). Während die Folge (3.29) sehr zufällige Reaktion aufweist, ist der Verlauf der Funktionen (3.27) und (3.34) ähnlich. Die Unterschiede in der Frequenzschätzung führen lediglich zum erhöhten Rauschen von  $W$ . Ein zusätzliches Problem, das sich bei dem Vergleich der Methoden ergibt, ist die Wahl der Schwelle. Die Detektion der Sprungstellen mit einem konstanten Schwellwert kann nur schwer automatisch erfolgen. Da ein EEG-Signal bzgl. Personen- und Krankheitsbildern spezifisch ist, sind die im Signal auftretenden Unterschiede allgemein nicht bekannt. Außerdem hängt die Varianz von  $W$  von der Länge  $N$  der vergleichenden Abschnitte ab. Die Konstruktion einer adaptiven Schwellwertfunktion, die sich den Signaleigenschaften anpaßt, ist daher notwendig.

### 3.3.2 Konstruktion adaptiver Schwellen

Die Detektion einer signifikanten Änderung im Signal ist ein allgemeines in der Signalanalyse sehr oft auftretendes Problem. Soll die Erkennung automatisch erfolgen, muß das Verhalten des Signals apriori bekannt sein. Z.B. kann bei bekannter Verteilung der einzelnen Meßwerte, bzw. der beobachteten statistischen Größe ein fester Quantilwert dieser Verteilungsfunktion als Schwellwert angewandt werden. Es ist aber nur in den seltensten Fällen möglich, eine Aussage über die Verteilung eines Zeitreihenparameters zu machen. Wird ein  $\alpha$ -Quantil aus den Meß- bzw. Schätzwerten approximiert, so erreicht man im konkreten Fall eine Schwelle, bei der  $100 \cdot (1-\alpha)$  % der Werte der Schätzfunktion größer als diese Schwelle sind, wobei die Größe dieser Werte die Schätzung nicht beeinflusst. Es kann dazu führen, daß in stationären Abschnitten falsche Werte als Instationaritäten (Ausreißer) detektiert werden. Ein geschätzter Quantilwert kann trotzdem zur Detektion von Instationaritäten angewandt werden, wenn die absolute Größe der Werte berücksichtigt wird. Dabei ist auch für nichtsymmetrische Daten der Interquantilabstand ein gutes Maß für die Bereichsspanne der Meßwerte.

Sei  $q_\alpha$  eine Schätzfunktion des  $\alpha$ -Quantils. Eine Schwelle wird mit

$$s = q_\alpha + \gamma (q_{(1+\beta)/2} - q_{(1-\beta)/2}) \quad (3.34)$$

definiert, wobei mit  $s$  für  $\alpha \geq 1/2$  und  $\gamma > 0$  eine obere und für  $\alpha \leq 1/2$  und  $\gamma < 0$  eine untere Schwelle gegeben sind. Während mit  $\alpha$  der prozentuale Anteil möglicher Ausreißer gegeben ist, bestimmt die Wahl von  $\gamma$  die Höhe der Schwelle bezüglich der Bereichspanne, die als der Interquantilabstand

$$k_\beta = q_{(1+\beta)/2} - q_{(1-\beta)/2}$$

angenommen wird. In der beschreibenden Statistik charakterisiert z.B. der Interquartilabstand  $k_{0.5} = q_{0.75} - q_{0.25}$  die Variabilität der mittleren Hälfte der Beobachtungswerte. Als nicht typisch (Ausreißer) werden dann die Werte betrachtet, die größer als  $q_{0.75} + 1.5 k_{0.5}$  bzw. kleiner als  $q_{0.25} - 1.5 k_{0.5}$  sind (vgl. [36]). Diese Beschreibung einer Stichprobe ist als "box and whisker plot" bekannt. Für eine konkrete Anwendung in der Signalanalyse ist es dennoch notwendig, die Einstellung von  $\alpha$ ,  $\beta$  und  $\gamma$  der Aufgabenstellung anzupassen. Eine Schwelle, die nach der Formel (3.34) konstruiert wird, paßt sich den Eigenschaften der Zeitreihe bzw. des aus ihren Werten geschätzten Parameters an, so daß nur die Werte als Instationaritäten erkannt werden, die sich aus der Zeitreihe deutlich hervorheben.

Zur Schätzung des Quantils einer i.a. nicht bekannten Verteilungsfunktion kann z.B. die rekursive Vorschrift (1.11) aus dem Satz 1.3 angewandt werden. Für Funktionen, die ein quasistationäres Verhalten aufweisen, führt eine adaptive Gestaltung des Algorithmus, wie es im Abschnitt 2.5 vorgeschlagen wurde, zu Schätzungen, die sich dem dynamischen Verhalten der zufälligen Funktion anpassen. Man bezeichne mit  $Q_\alpha$  einen adaptiven Operator zur Schätzung der Quantilwerte, dann kann der Operator zur Bestimmung einer Schwelle  $\mathcal{S}$  der Folge  $Y = \{y_t\}$  (der Meßwerte bzw. ihrer Funktionswerte) wie folgt definiert werden:

$$\mathcal{S}(Y)(t) = Q_\alpha(Y)(t) + \gamma \mathcal{K}_\beta(Y)(t), \quad t = 0, 1, \dots \quad (3.35)$$

wobei mit  $\mathcal{K}_\beta$  der Operator

$$\mathcal{K}_\beta = \max \left\{ Q_{(1+\beta)/2} - Q_{(1-\beta)/2}, K_0 \right\}$$

gegeben ist. Der Schwellenoperator  $\mathcal{S}$  wird gleichzeitig während der Schätzung von  $Y$  bestimmt und dessen Überschreitung (für  $\gamma > 0$ ) bzw. Unterschreitung (für  $\gamma < 0$ ) sofort detektiert. Konvergiert die Folge  $Y = \{y_t\}$  gegen einen konstanten Wert, so konvergiert  $\mathcal{K}_\beta(Y)(t)$  gegen Null. Für diesen Fall ist es

notwendig, eine minimale Länge des  $\beta$ -Interquantilintervalls  $K_0 > 0$  zu definieren, damit der Schwellenoperator seine Eigenschaften behält. Bei der Wahl der Adaptionfaktoren ist es wichtig, die Adaptivität des Operators  $K_\beta$  gegenüber  $Q_\alpha$  zu begünstigen, um eine schnelle Anpassung zu erreichen. Die praktische Relevanz der auf diese Weise konstruierten Schwellen konnte erfolgreich bei der Detektion von Gefäßen in Bildern des Augenhintergrundes gezeigt werden (vgl. [57]), wobei hier eine konstante Steuerfolge bei der Schätzung der Quantile verwendet wurde. Für die komplexe Aufgabe der adaptiven Segmentierung soll die Schätzung der Quantilwerte lernfähig gestaltet werden.

### 3.3.3 Anwendung adaptiver rekursiver Algorithmen im Segmentationsverfahren

Die Prozedur der adaptiven Segmentation wie sie am Anfang dieses Abschnittes beschrieben wurde, kann durch die Anwendung adaptiver rekursiver Verfahren an verschiedenen Stellen verbessert werden. Die Umsetzung der Frequenzdefinition aus dem Abschnitt 2.3 in der Formel (3.30) kann die Prozedur beschleunigen. Durch die Anwendung des Operators  $\mathcal{P}$  nach (3.35) für die Konstruktion einer Schwelle wird eine bessere automatische Erkennung von lokalen Maxima der Funktion  $W$  möglich, indem sich die Höhe der Schwelle dem Verlauf von  $W$  anpaßt. Durch die Adaption der Schwelle werden die lokalen Maxima erkannt, in deren Umgebung die Funktion  $W$  einen scharfen Anstieg besitzt. Sie treten als kurze Instationaritäten von  $W$  auf, denen sich die Schwelle nicht anpaßt. Damit werden redundante Segmente vermieden, die z.B. durch zufällige Schwankungen oder durch eine dicht aneinander folgende Frequenz- und Leistungsänderung entstehen können. Im Verfahren (3.28)-(3.29), in dem die Differenz der Kenngrößen betrachtet wird, kann durch die Adaption der Schwelle erreicht werden, daß auch im Bereich der hohen Leistung nur die lokalen Extrema von  $W$  detektiert werden, die einer deutlichen Änderung im Signal entsprechen.

Für eine automatische Erkennung soll die Schwelle kanal- und patientenunabhängig konstruiert werden. D.h., bei einer festen Einstellung der Abschnittslänge  $N$ , bleiben die Schwellenparameter ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $K_0$  und die Adaptionfaktoren der Steuerfolgen) sowohl für alle Kanäle einer Messung als auch für alle Messungen einer Art konstant. Zahlreiche empirische Untersuchungen zeigten, daß eine solche Einstellung möglich ist, wobei die Steuerfolgen bei der adap-

tiven rekursiven Bestimmung der Quantile nach der Formel (2.45) gewählt wurden. Für die Histogrammschätzung wurden 2-5 Klassen im Intervall  $[1, D]$  gebildet. Für  $W$  nach (3.27) und (3.34) wurde  $D = 10$  gewählt, für  $W$  nach (3.28) konnte  $D$  aus dem maximalen möglichen Meßwert und der Länge  $N$  des Analyseabschnitts ermittelt werden. Bei der Wahl der Adaptionfaktoren muß berücksichtigt werden, daß mit wachsendem  $N$  der Anstieg von  $W$  kleiner wird. Dementsprechend ist die Adaption der Schwelle mittels eines kleineren Adaptionfaktors zu verlangsamen. In der Abbildung 3.8 wird die Segmentation nach (3.34) mit einer adaptiven und konstanten Schwelle verglichen.

Das Ergebnis der Segmentierung nach der beschriebenen Methode hängt wesentlich von der Länge der Analyseabschnitte ab. Für ein großes  $N$  ist der Anstieg der Funktion  $W$  sehr klein, weniger deutliche Änderungen werden nicht erkannt. Liegen zusätzlich zwei und mehrere Strukturbrüche im Analyseabschnitt, müssen nicht alle erkannt werden, bzw. das Maximum der Funktion  $W$  liegt dann nicht unbedingt an der Stelle des Bruches. Andererseits bewirkt ein kleines  $N$  ein höheres Schwanken der Funktion  $W$ , so daß ihre lokalen Maxima durch ein Rauschen verdeckt werden. Das Verhalten der Funktion  $W$  wird an einem simulierten sinusförmigen Signal demonstriert, in dem das Grundmuster durch ein anderes Segment, dessen Länge kleiner als  $N$  ist, unterbrochen wurde (im Beispiel wurde die Amplitude erhöht, siehe Abbildung 3.9a). Die Funktion bleibt konstant, solange das kurze Segment vollständig in einem der Beobachtungsabschnitte liegt. In diesem Fall können die Segmentgrenzen nicht eindeutig erkannt werden. Mit wachsendem  $N$  wird die konstante Phase von  $W$  länger, ihr Wert in dieser Phase kleiner. Ein eindeutiges lokales Minimum entsteht erst, wenn  $N$  kleiner oder gleich der Länge des Abschnitts ist. (In der Abbildung 3.9b wird das gleiche Verhalten von  $W$  für eine Folge unabhängiger gleichverteilter Zufallsgrößen demonstriert.) Liegen in den Vergleichsabschnitten mehrere Muster, bekommt  $W$  den Verlauf einer stufenartigen Funktion, wobei nur für spezielle Werte von  $N$  sich eindeutige lokale Extrema ausbilden (s. Abb. 3.9c).

Der Einfluß der Größe  $N$  auf das Segmentationsergebnis kann abgeschwächt werden, indem die Berechnung von  $E_t$  und  $F_t$  dynamisch gestaltet wird. Man betrachte im Zeitpunkt  $t$  die folgende Schätzung der momentanen Leistung

$$E_0 = 0 \quad (3.36)$$

$$E_t = E_{t-1} + c (x_t^2 - E_{t-1}), \quad t = 1, 2, \dots, T$$

und der momentanen Frequenz

$$F_0 = 0 \quad (3.37)$$

$$F_t = F_{t-1} + c (I_{(-\infty, 0)}(x_t \cdot x_{t-1}) - F_{t-1}), \quad t = 1, 2, \dots, T$$

aus den vorangegangenen Meßwerten und die Schätzung

$$E'_T = 0 \quad (3.38)$$

$$E'_t = E'_{t+1} + c (x_t^2 - E'_{t+1}), \quad t = T-1, T-2, \dots, 0$$

$$F'_T = 0 \quad (3.39)$$

$$F'_t = F'_{t+1} + c (I_{(\infty, 0)}(x_t \cdot x_{t+1}) - F'_{t+1}) \quad t=T-1, T-2, \dots, 0$$

aus den darauffolgenden Meßwerten, wobei  $c \in (0, 1)$  und  $T$  die Länge der Messung bzw. eines Messungsblocks sind. Dann wird mit der Vergleichsfunktion

$$W(t) = \max \left\{ \frac{\max \{E_t, E'_{t+1}\} + E_0}{\min \{E_t, E'_{t+1}\} + E_0}, \frac{\max \{F_t, F'_{t+1}\}}{\min \{F_t, F'_{t+1}\}} \right\} - 1 \quad (3.40)$$

ein Segmentationsverfahren konstruiert, mit dem die Signale aus der Abbildung 3.9 für verschiedene Werte des Faktors  $c$  eindeutig segmentiert werden (siehe Abbildung 3.10). Die Verbesserung wird dadurch erzielt, daß nicht der mittlere Unterschied von zwei benachbarten Abschnitten verfolgt wird, sondern die Meßwerte abhängig von der Entfernung zum aktuellen Zeitpunkt betrachtet werden. Während in der ersten Methode zum Zeitpunkt  $t$  genau  $N$  Meßwerte aus der Vergangenheit und ebenfalls  $N$  Meßwerte aus der Zukunft in die Schätzung gleichmäßig mit dem Gewicht  $c_k = 1/N$  ( $k = 1, \dots, N$ ) eingegangen sind, werden in der Berechnung (3.36)-(3.39) die Werte  $x_{t-k}$  und  $x_{t+k+1}$  ( $k = 0, 1, \dots$ ) mit  $c_k = c(1-c)^k$  exponentiell gewichtet. Sei  $c = d/N$ ,  $d > 1$ . Dann gilt für  $j < N$  mit  $c(1-c)^j = 1/N$ , daß

$$j = \frac{\ln(d)}{\ln(N/N-d)} \quad (3.41)$$

ist und sich durch

$$\frac{\ln(d)}{d} - \frac{\ln(d)}{N} \leq \frac{j}{N} < \frac{\ln(d)}{d} < \frac{1}{e} \approx 0.368.$$

abschätzen läßt. Weiter kann für ein genügend großes  $N$  gezeigt werden:

$$c(1-c)^k = \frac{d}{N} \left(1 - \frac{d}{N}\right)^k > \frac{de^{-d}}{N} \text{ für } k < N$$

und

$$c(1-c)^k = \frac{d}{N} \left(1 - \frac{d}{N}\right)^k \leq \frac{de^{-d}}{N} \text{ für } k \geq N$$

(für  $N$  gilt  $(1 - d/N)^N \approx e^{-d}$ ). Wurde also in der ersten Methode ein  $N$  gefunden, so daß die Länge der Analyseabschnitte möglichst klein ist, um alle Änderungen zu erkennen und dennoch groß genug, um ausreichende Schätzergebnisse zu bekommen, erreicht man für  $1.5 \leq d < 10$ , daß die Gewichte von etwa dem ersten Drittel der Werte größer, die restlichen kleiner als  $1/N$  sind (s. Abbildung 3.11). Am deutlichstem werden die Meßwerte  $x_t$  und  $x_{t+1}$  mit  $c_0 = \frac{d}{N} > \frac{1}{N}$  gewichtet, das Gewicht der Werte  $x_{t-N+1}$  und  $x_{t+N}$  wird durch  $c \approx \frac{de^{-d}}{N} < 1/eN$  abgeschätzt. Andererseits kann man aus (3.41) und (3.42) für jeden Faktor  $c$  den Einfluß der Meßwerte  $x_{t-N+1}$  und  $x_{t+N}$  mit  $d = cN$  bestimmen. Die Größen (3.36)-(3.39) lassen sich mit Hilfe der Operatoren  $\mathcal{L}^c$  und  $\mathcal{F}^c$  bestimmen (vgl. Abschnitt 2.3, Formel (2.19) und (2.22)), wobei die Funktionen  $E'_t$  und  $F'_t$  durch die Anwendung der Operatoren auf die Folge  $x'_k = x_{t+T-k}$ ,  $k=0, \dots, T-1$ , erreicht wird. Eine einfache Korrektur der Nulllinienverschiebung, eines Trends und ein Unterdrücken der Wellen im Bereich der Störsignale kann aufgrund der Filtereigenschaften der adaptiven Mittelwertschätzung (siehe (2.12)-(2.15)) rekursiv gleichzeitig mit der Auswertung erfolgen. Wird die adaptive momentane Leistung aus (2.19) durch die adaptive momentane Bandleistung (2.20) ersetzt und in der Schätzung der adaptiven momentanen Frequenz eine optimale deterministische Linie gewählt, so ist eine Datenvorverarbeitung nicht mehr notwendig. Die Kenngrößen werden über Signalabschnitte mit der Länge  $T$  (die von den rechentechnischen Möglichkeiten abhängig ist) bestimmt. Diese Abschnitte überlappen sich um  $k_0$ -Meßwerte, um eine korrekte Schätzung von  $E'_T$  und  $F'_T$  am Ende eines Abschnittes zu bekommen. Dabei wird  $k_0$  aus (2.8) berechnet.

Die Segmentierung von reellen EEG-Signalen mit Hilfe der Methode (3.36)-(3.40) wird am Abschnitt aus einem Kanal über 48 sec (Abtastrate 128 Hz, 6144 Meßwerte) gezeigt und mit der Segmentierung nach (3.30)-(3.34) verglichen (siehe Abbildung 3.12). Dazu wurde für beide Verfahren die beste empirisch ermittelte Einstellung der Parameter gewählt. Für die Segmentierung über feste Signalabschnitte wurde  $N = 100$  in der adaptiven Schätzung der Kenngrößen  $c = 0.03$  bestimmt. Die weiter diskutierten Segmentgrenzen wurden in den

Abbildungen 3.12a (adaptive Schätzung) und 3.12b (feste Abschnitte) mit nummerierten Pfeilen gekennzeichnet. Mit 1 und 3 werden zwei Stellen markiert, an denen kurze (etwa 50 Punkte lange) aber deutliche Wellen mit Hilfe adaptiver Verfahren erkannt wurden. Im Verlauf der zugehörigen Vergleichsfunktionen (Abbildung 3.12c) ist das Bild aus dem Beispiel über ein simuliertes Signal wieder zu finden (vgl. Abb. 3.9a und 3.10a). Während sich in der Schätzung nach (3.34) über dem Muster zwei Plateaus bilden, können mit (3.50) die Segmentgrenzen eindeutig bestimmt werden. Es ist erkennbar, daß der Zuwachs der Funktion  $W$  in beiden Fällen um so größer ist, je deutlicher sich das Muster aus der Umgebung hervorhebt (vgl. Verlauf von  $W$  für Pfeil 1 und Pfeil 3). Bei der adaptiven Schätzung aber können die Grenzen durch ein eindeutiges lokales Maximum exakt detektiert werden. Bei der Segmentierung über feste Abschnitte hängt die Position der Grenze von zufälligen Einflüssen und von dem sich adaptierenden Verhalten der Schwelle ab. In der Abbildung 3.14d wird nochmal deutlich, daß sich auch bei längeren Abschnitten mit der adaptiven Schätzung die Segmentgrenzen eindeutig bestimmen lassen (Pfeil 2). Da die Funktion  $W$  zufällig ist, kann ihre Anstiegsphase durch mehrere Peaks begleitet werden, als Segmentgrenze wird die Stelle des größten lokalen Maximums erkannt (Pfeil 4). Die Untersuchungen zeigen, daß mit Hilfe adaptiver Schätzung der Leistung und der Frequenz alle Muster erkannt werden, wobei die Lage der Grenzen zwischen den Mustern und Strukturen sich exakt durch eindeutige lokale Maxima der Vergleichsfunktion bestimmen läßt. Weiter wird das Verhalten der Vergleichsfunktion gezeigt, wenn sich das Signal annähernd wie eine gedämpfte Schwingung verhält oder wenn sich die Frequenz stetig verändert. Ein Segmentende kann erst dann erkannt werden, wenn es zu einer sprunghaften Änderung einer der Kenngrößen kommt. Man kann beobachten, daß beide Methoden die Segmentgrenzen unterschiedlich definieren. So verschiebt sich z.B. die Grenze 11 (Abb. 3.12b) bei der Schätzung über feste Signalabschnitte an die Stelle der Grenze 6 (Abb. 3.12a) in der adaptiven Schätzung. Im gleichen Zusammenhang können möglicherweise auch die Grenzen 3 und 8 betrachtet werden. Die Segmentgrenzen 5 und 7 in der Abb. 3.12a haben kein Gegenstück in der Abb. 3.12b, was sich mindestens bei der Grenze 5 als Nachteil auswirkt (vgl. Abb. 3.12e). In der Schätzung über feste Abschnitte kann ein Anstieg von  $W$  beobachtet werden. Da sich hier aber kein eindeutiges Extremum ausbildet, paßt sich die Schwelle diesem Verlauf an. Schließlich kann in der Abb. 3.12f wie-

derholt die unterschiedliche Lage der ersten Segmentgrenze auf die nicht eindeutige Lage des lokalen Maximum von  $W$  nach (3.34) zurückgeführt werden. Das Fehlen der Grenzen 9 und 10 in der adaptiven Schätzung entsteht durch die Adaption der Schwelle über einer Reihe flacher lokaler Maxima im Verlauf der Funktion  $W$ .

Die Eigenschaften der Segmentierung können durch die Wahl der Parameter beeinflusst werden. Die Methode ist für einen kleinen Faktor  $c$  robust gegenüber kurzen Änderungen. Dabei sind die Anstiegsphase in der Umgebung eines lokalen Maximums von  $W$  länger und der Anstieg kleiner, der Wert von  $W$  an der Stelle des lokalen Maximums ist ebenfalls klein. Mit wachsendem Faktor  $c$  wird die Methode empfindlicher gegen kurze Änderungen, der Verlauf von  $W$  ist stärker verrauscht, die lokalen Maxima aber deutlicher. Ähnliches Verhalten zeigt die Funktion  $W$  auch beim Vergleich von zwei Abschnitten fester Länge, wenn die Länge  $N$  größer wird. Die Funktion  $W$  bekommt aber mit wachsendem  $N$  den Verlauf einer stufenartigen Funktion und die Lage der lokalen Maxima wird immer undeutlicher. Durch den Parameter der adaptiven Schwelle kann die Segmentation folgendermaßen beeinflusst werden. Durch den Parameter  $\gamma$  und durch den minimalen Schwellenwert  $K_0$  kann die Höhe der Schwelle gesteuert werden. Mit der Wahl des Adaptionfaktors wird der Anstieg von  $W$  bestimmt, bei welchem die lokalen Maxima detektiert werden und bei welchem sich die Schwelle dem Verlauf von  $W$  anpaßt.

## LITERATURVERZEICHNIS

- [1] Anděl J., Statistische Analyse von Zeitreihen, Akademie-Verlag, Berlin, 1984.
- [2] Anderson T.W., Statističeskij analiz vremennyh rjadov, Verlag Mir, Moskau, 1986.
- [3] Barešová E., Parametrische Spektralanalyse schwach stationärer Prozesse über die Anpassung autoregressiver Modelle und ihre Anwendung in der technischen und medizinischen Diagnostik, Friedrich-Schiller-Universität Jena, (Diplomarbeit) 1989.
- [4] Barešová E., Griebach G., Filtereigenschaften adaptiver Schätzverfahren und ihre Anwendung in der Biosignalanalyse, AG Biomathematik der Abteilung für Biomedizintechnik, Friedrich-Schiller-Universität Jena, Forschungsbericht, 1991.
- [5] Barešová E., Schack B., Lernfähige adaptive ARMA-Modellierung, AG Biomathematik der Abteilung für Biomedizintechnik, Friedrich-Schiller-Universität Jena, Forschungsbericht Nr. 5, 1992.
- [6] Barešová E., Schack B., Zur Wahl der Steuerfolge bei der adaptiven Parameterschätzung von ARMA-Prozessen, AG Biomathematik der Abteilung für Biomedizintechnik, Friedrich-Schiller-Universität Jena, Forschungsbericht Nr.1, 1992.
- [7] Barešová E., Hoyer D., Schack B., Hoyer H., Ein Filterverfahren zur Hervorhebung wesentlicher Anteile evozierter Potentiale am Beispiel der P100 visuell evozierter Potentiale, Medizintechnik 31, 7-9, 1991.
- [8] Barešová E., Schack B., Entwicklung einer adaptiven ARMA-Modellierung stochastischer Zeitreihen, AG Biomathematik der Abteilung für Biomedizintechnik, Friedrich-Schiller-Universität Jena, Forschungsbericht Nr. 4, 1992.
- [9] Barešová E., Schack B., Untersuchungen zur ARMA-Modellierung von Biosignalen, AG Biomathematik der Abteilung für Biomedizintechnik, Friedrich-Schiller-Universität Jena, Forschungsbericht Nr. 3, 1992.
- [10] Barešová E., Griebach G., Schack B., Vilser W., Bräuer-Burchardt Ch., Filtering Properties of an Adaptive Trend Operator and their Application for Determining the Arterial Blood Velocity in Retinal Vessels, Medical Progress through Technology, No. 21, 126-134, 1996.

- [11] Barešová E., Griebach G., Schack B., Vilser W., Bräuer-Burchardt Ch., Untersuchungen von Filtereigenschaften adaptiver Trendschätzungen und ihre Anwendung bei der Analyse von Indikatorlöskurven in der Augenheilkunde, Biomedizinische Technik 39 (Ergänzungsband), 1994.
- [12] Basseville M., Benveniste A., Sequential Detection of Abrupt Changes in Spectral Characteristics of Digital Signals, IEEE Transactions on Information Theory, Vol. 29, No. 5, Sept. 1983.
- [13] Bodenstein G., Praetorius H.M., Feature Extraction from the Electroencephalogram by Adaptive Segmentation, Proceedings of IEEE, Vol. 65, No. 5, May 1977.
- [14] Bodenstein G., Schneider W., Malsburg C. v.D., Computerized EEG Pattern Classification by Adaptive Segmentation and Probability-Density-Function Classification. Description of the Method, Comput. Biol. Med. Vol. 15, No. 5, pp. 297-313, 1985.
- [15] Borodin A.N., Procedura stochastičeskoj aproximaciji pri nabludenjach, udovletborjajuščich usloviju slaboj zavisimosti, Teoria Verojatnostej i jej Primenenia, Moskva, 1979.
- [16] Brockwell P.J., Davis R.A., Time Series: Theory and Methods, Springer Verlag, New-York, Inc, 1991.
- [17] Burg J.P., Maximum entropy spectral analysis, in Proc. 37th Meeting Society of exploration Geophysicists, Oklahoma City, OK, 31.10.1967.
- [18] Chatfield C., Analyse von Zeitreihen, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1982.
- [19] Cybenko G., The numerical stability of the Levinson-Durbin-Algorithm for Toeplitzsystem of equations, SIAM J. Sci. Stat. Comput., Vol. 1, No.3, pp. 303 - 319, 1980.
- [20] Durbin J., The fitting of time series models, Rev. Inst. Int. Statist., Vol. 28, No. 3, pp. 233 - 243, 1960.
- [21] Fisz M., Wahrscheinlichkeitsrechnung und mathematische Statistik, VEB Deutscher Verlag der Wissenschaften, Berlin, 1989.
- [22] Griebach G., Computerorientierte Meßstochastik in der technischen Diagnose und dem Signalmapping, Friedrich-Schiller-Universität Jena, (Habilitationsschrift) 1990.

- [23] Griebbach G., Schack B., Adaptive Quantile Estimation and its Application in Analysis of Biological Signals, *Biom. J.* 35(2), pp. 165-179, 1993.
- [24] Griebbach G., Schack B., Putsche P., Barešová E., Bolten, J., The Dynamic Description of a Stochastic Signal by its Adaptive Momentary Power and Momentary Frequency Estimation and its Application in Analysis of Biological Signals, *Med. & Biol. Eng. & Computing*, 32 (Vol.6), pp. 632- 637, 1994.
- [25] Griebbach G., Witte H., Rechenzeit- und speicherplatzoptimale Aufbereitung des 4-NN- (4-nearest neighbours) Interpolationsalgorithmus für EEG-Spektralmapping, *Medizintechnik* 28, 100-102, 1988.
- [26] Griebbach G., Witte H., Schack B., Die Kenngröße adaptiver Leistungsquotient im EEG-Mapping, AG Biomathematik der Abteilung für Biomedizintechnik, Friedrich-Schiller-Universität Jena, Forschungsbericht, 1992.
- [27] Griebbach G., Witte H., Schack B., Möglichkeiten des Einsatzes von adaptiven dynamischen Spektralgrößen in der Biosignalanalyse, *Biomedizinische Technik (Erg.bd)* 38, 81-82, 1993.
- [28] Günther R., Adaptive Parameterschätzung und Vorhersage in ARMA- und verwandten Modellen, Friedrich-Schiller-Universität Jena, (Habilitationsschrift) 1989.
- [29] Günther R., On the Convergence of Some Adaptive Estimation Procedures, *Math. Operationsforsch. Statist., Ser. Stat.*, Vol. 14, No. 4, pp 535-550, 1983.
- [30] Günther R., Adaptive Vorhersage von stationären zufälligen Folgen mit rationaler Spektraldichte, Friedrich-Schiller-Universität Jena, (Promotionsschrift) 1976.
- [31] Hannan E.J., *Time Series Analysis*, Wiley, New York, 1960.
- [32] Hannan E.J., *Multiple time series*, Wiley, New York, 1970.
- [33] Hannan E.J., Quinn B.G., The Determination of the Order of an Autoregression, *J. R. Statist. Soc. Ser. B*, vol.41, no.2, pp. 190-195, 1979.
- [34] Hannan E.J., Rissanen J., Recursive estimation of mixed autoregressive moving average order, *Biometrika* 69, 81-94, 1982.
- [35] Haykin S., *Adaptive Filter Theory*, Prentice-Hall, New Jersey, 1986.

- [36] Hoaglin D.C., Mosteller F., Tukey J.W., Understanding Robust and Extrapolary Data Analysis, John Wiley & Sons, Inc., New York u.a., Chapter 3, pp 58-65, 1983.
- [37] Kay S.M., Marple S.L., Spectrum Analysis - a Modern Perspektive, Proceeding of the IEEE, Vol. 69, No. 11, 1981.
- [38] Kesten H., Accelerated Stochastic Approximation, Annals of Math. Stat. 29, pp. 41 - 59.
- [39] Kiefer E., Wolfowitz J., Stochastic Estimation of the Maximum of a Regression Function, Ann. Math. Statist., No. 23, pp. 462-466, 1952.
- [40] Koopmans L.H., The Spectral Analysis of Time Series, Academic Press, New York - San Francisco - London, 1974.
- [41] Krajča V., Petráněk S., Patáková I., Värri A., Automatic Identification of Significant Graphoelements in Multichannel EEG Recordings by Adaptive Segmentation and Fuzzy Clustering, Int J Biomed Comput. 28, pp. 71-89, Elsevier Scientific Publishers Ireland Ltd., 1991.
- [42] Levinson N., The Wiener Root Mean Square Error Criterion in Filter Desing and Prediction, J. Math. Phys., Vol. 25, No. 4, pp. 261 - 278, 1944.
- [43] Maess G., Vorlesung über numerische Mathematik, Akademie-Verlag, Berlin, 1984.
- [44] Maier H., Methoden zur Schätzung der Ordnung bei autoregressiven Modellen, BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1989.
- [45] Makhoul J., Linear Prediction, A Tutorial Review, Proceeding of the IEEE, Vol. 63, No. 4, 1975.
- [46] Michael D., Houchin J., Automatic EEG Analysis: A Segmentation Procedure Based on the Autocorrelation Function, Electroenceph. Clin. Neurophysiol. 46, pp. 219 - 225, 1979.
- [47] Nevelson M.B., Chasminskij P.Z., Stochasticeskaja approximamacia i rekurentnoje ocenivanie, Nauka Moskau 1972.
- [48] Otnes R.K., Enochson L., Applied Time Series Analysis, Vol. 1, Wiley, New York, 1978 (russ. Mir, Moskau, 1982).
- [49] Proakis J.G., Manolakis D.G., Introduction to Digital Signal Processing, Macmillan Publishing Company, New York, 1989.
- [50] Pšeničnyj B.N., Danilin J.M., Numerische Methoden für Extremalaufgaben, Deutscher Verlag der Wissenschaften, Berlin, 1982.

- [51] Robbins H., Monro S., A Stochastic Approximation Method, *Ann. Math. Statistic*, 22, 1, pp. 400-407, 1951.
- [52] Schack B., Barešová E., Griebach G., Witte H., *Methods of Dynamic Spectral Analysis by Self-exciting ARMA Models and their Application to Analysing Biosignals*, *Med. & Biol. Eng. & Computing* (im Druck), 1995.
- [53] Schack B., Griebach G., Die stochastische Approximation von Quantilwerten stark mischender zufälliger Prozesse, AG Biomathematik der Abteilung für Biomedizintechnik, FSU Jena, Forschungsbericht Nr. 3, 1991.
- [54] Schack B., Griebach G., Eigenschaften adaptiver Quantilwertschätzungen. Untersuchungen an einigen Simulationsbeispielen, AG Biomathematik der Abteilung für Biomedizintechnik, FSU Jena, Forschungsbericht Nr. 4, 1991.
- [55] Schack B., Griebach G., Adaptive Methods of Trend Detection and their Application in Analysing Biosignals, *Biom. J.* 36(4), pp.429-452, 1993.
- [56] Schack B., Griebach G., Barešová E., The Dynamic Description of Nonstationary EEG on the Basis of Adaptive Methods und Models, In: *Quantitative and Topological EEG and EMG Analysis*, Universitätsverlag Jena, pp. 249-253, 1995.
- [57] Schack B., Vilser W., Barešová E., Bräuer-Burchardt Ch., Griebach G., Witte H., Arnold M., Putsche P., Anwendung adaptiver Verfahren zur Erfassung der Position und des Durchmessers von Blutgefäßen am Augenhintergrund, Tagungsband Mustererkennung zum 16. DAGM Symposium und 18. Workshop der ÖAGM, 474-481, 1994.
- [58] Schack B., Witte H., Griebach G., Parametrische Methoden der dynamischen Spektralanalyse und ihre Anwendung in der Biosignalanalyse, *Biomedizinische Technik (Erg.bd)* 38, 79-80, 1993.
- [59] Schumann N.P., Witte H., Scholle H.Ch., Zwiener U., Spektrales EMG-Mapping am M.masseter: Topographie der EMG-Aktivität in Abhängigkeit von der Muskelfunktion, In: Harzer W. (ed.), *Kieferorthopädischer Gewebeumbau*, Quintessenz-Verlag, 191-197, 1990.
- [60] Shibata R., Selection of the order of an autoregressive model by Akaike's information criterion, *Biometrika* 63, 117-126, 1976.
- [61] Silin D.Y, Skrylev K.M., Avtomatičeskaja segmentacija EEG, *Žurnal vysšej nervnoj dejatelnosti* 36, pp 1152-1155 (in russisch), 1986.

- [62] Steinberg H.-W., Gasser Th., Franke J., Fitting Autoregressive Models to EEG Time Series, An Empirical Comparison of Estimates of the Order, IEEE Transformations On Acoustics, Speech, And Signal Processing, Vol. ASSP-33, No. 1, Feb. 1985.
- [63] Tong H., Non-linear Time Series: A Dynamical System Approach, Oxford University Press, Oxford - New York - Toronto, 1990.
- [64] Värri A., Digital Processing of the EEG in Epilepsy, Licentiate Thesis, Tampere University of Technology, Tampere, Finland, 1988.
- [65] Vilser W., Möglichkeiten und Grenzen der retinalen Durchblutungsdiagnostik auf der Basis von Indikatortechnik und Längenmessungen. Theoretische und praktische Untersuchungen zu einem ophtalmologischen Arbeitsplatz für die Diagnostik retinaler Durchblutungsstörungen, Technische Hochschule Ilmenau (Habilitationsschrift), 1994.
- [66] Walker G., On Periodicity in Series of Related Terms, Proc. Roy. Soc. London, Series A, vol. 131, pp. 518-532, 1931.
- [67] Wasan M.T., Stochastic Approximation, Cambridge, 1969.
- [68] Wennberg A., Zetterberg L.H., Application of a computer-based model for EEG-analysis, Electroenceph. and clin. Neurophysiol., 31: 457-468, 1971.
- [69] Whitman E.C., The Spectral Analysis of Diskrete Time Series in terms of linear regressive models, Naval Ordnance Labs Rep. NOLTR-70-109, White Oak, MD, June 23, 1974.
- [70] Witte H., Ansorg J., Eiselt M., Griebbach G., Stallknecht K., Strategies of image and data processing in EEG-EP-EMG-Mapping, Bild und Ton 43, pp. 144-147, 1990.
- [71] Witte H., Griebbach G., Krajčů V., Rother M., Stallknecht K., Putsche P. Möglichkeiten des Einsatzes der diskreten Hilberttransformation in der Biosignalanalyse - Nachweis der Effizienz durch Signalmodelle. Medizintechnik 29, 65-70, 1989.
- [72] Witte H., Griebbach G., Zwiener U., Rother M., EEG- Leistungsspektral-Mapping - ein funktionell bildgebendes Verfahren in der Neuroelektrodiagnostik. I. Grundlagen der Meßwerterfassung und -verarbeitung, Bild und Ton 5, 136-143, 1989.

- [73] Witte H., Griebach G., Zwiener U., Rother M., EEG- Leistungsspektral-Mapping - ein funktionell bildgebendes Verfahren in der Neuroelektrodiagnostik. II. Grundlagen der Bilderzeugung und Perspektiven der Bild- verarbeitung, Bild und Ton 6, 172-176, 1989.
- [74] Witte H., Griebach G., Schack B., Stallknecht K., Krajča V., Patáková I., Nutzung der Fouriertransformation als Basisalgorithmus für die Datenvorverarbeitung in der automatischen EEG-Analyse. Medizintechnik 30, 65-71, 1990.
- [75] Yule G.U., On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer's Sunspot Numbers, Philosophical Trans. Roy. Soc. London, Series A, vol 226, pp. 267-298, July 1927.
- [76] Zetterberg L.H., Estimation of Parameter for Linear Difference Equation with Application to EEG Analysis, Math. Biosciences, 5: 277:275, 1969.

## ABBILDUNGEN

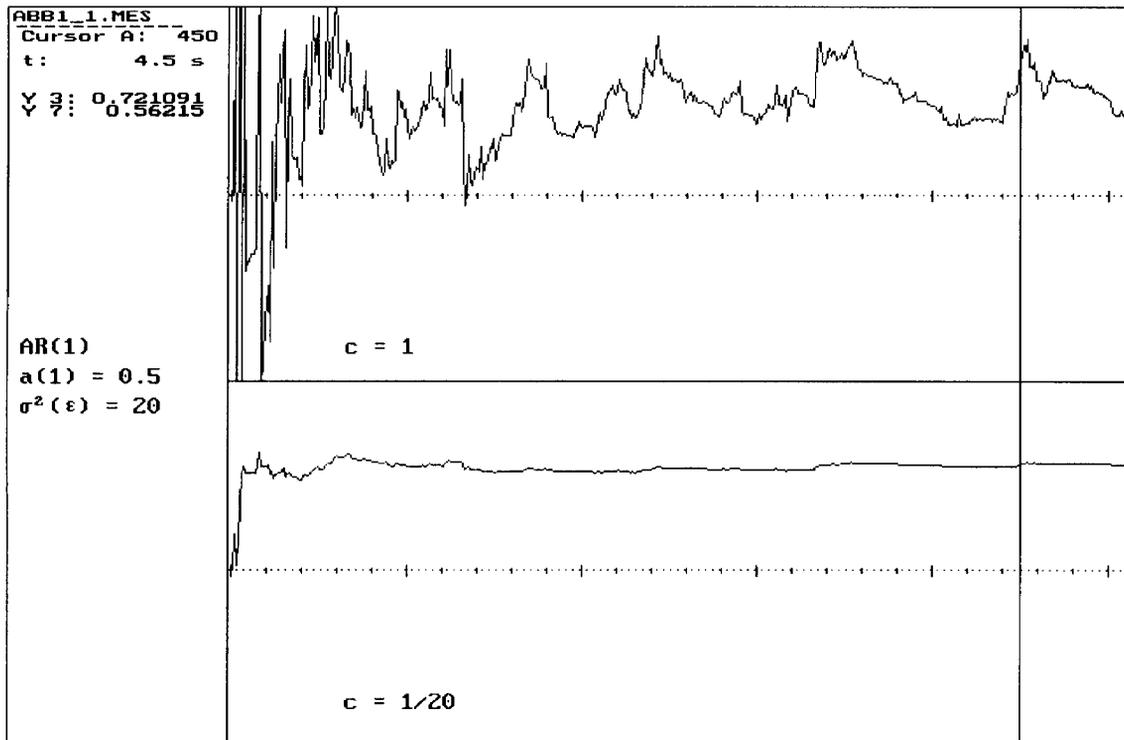


Abb.1.1: Konvergenzverhalten der Schätzfolge  $\hat{a}_1(t)$  eines AR(1)-Prozesses, die mit einer deterministischen Steuerfolge bestimmt wurde

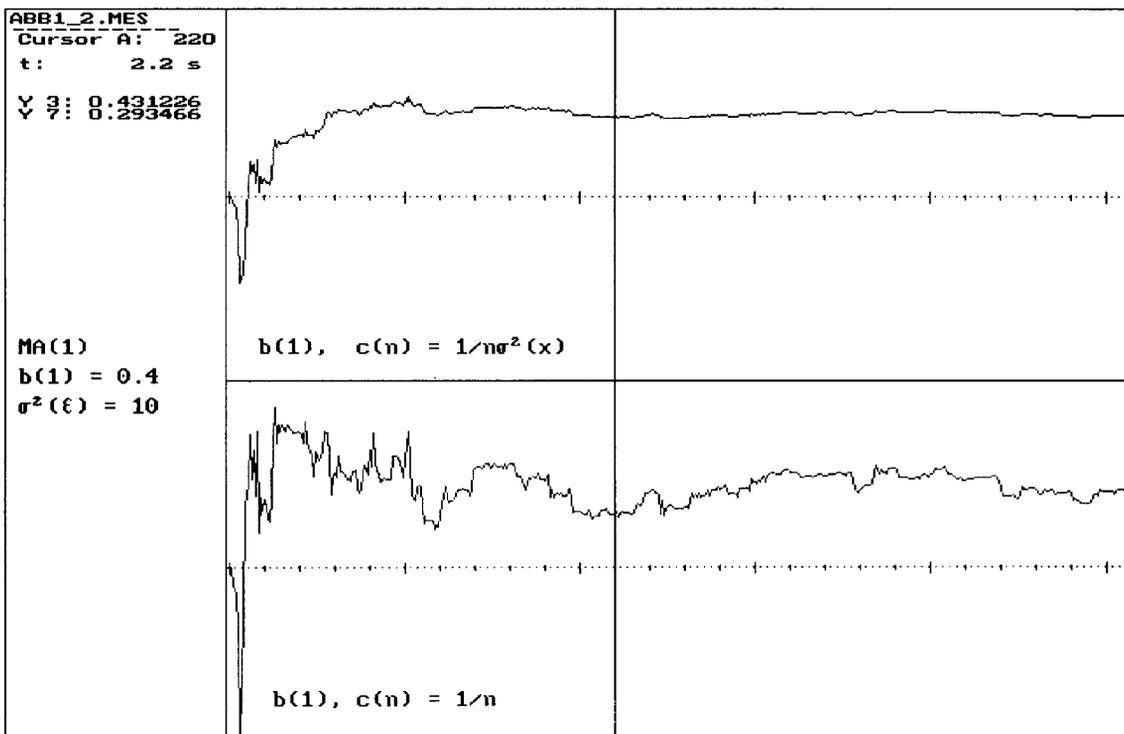


Abb.1.2: Verbesserung der Parameterschätzung durch eine an die vorliegende Zeitreihe angepaßte Wahl der Steuerfolge  $\{c_i\}$  entsprechend (1.27)

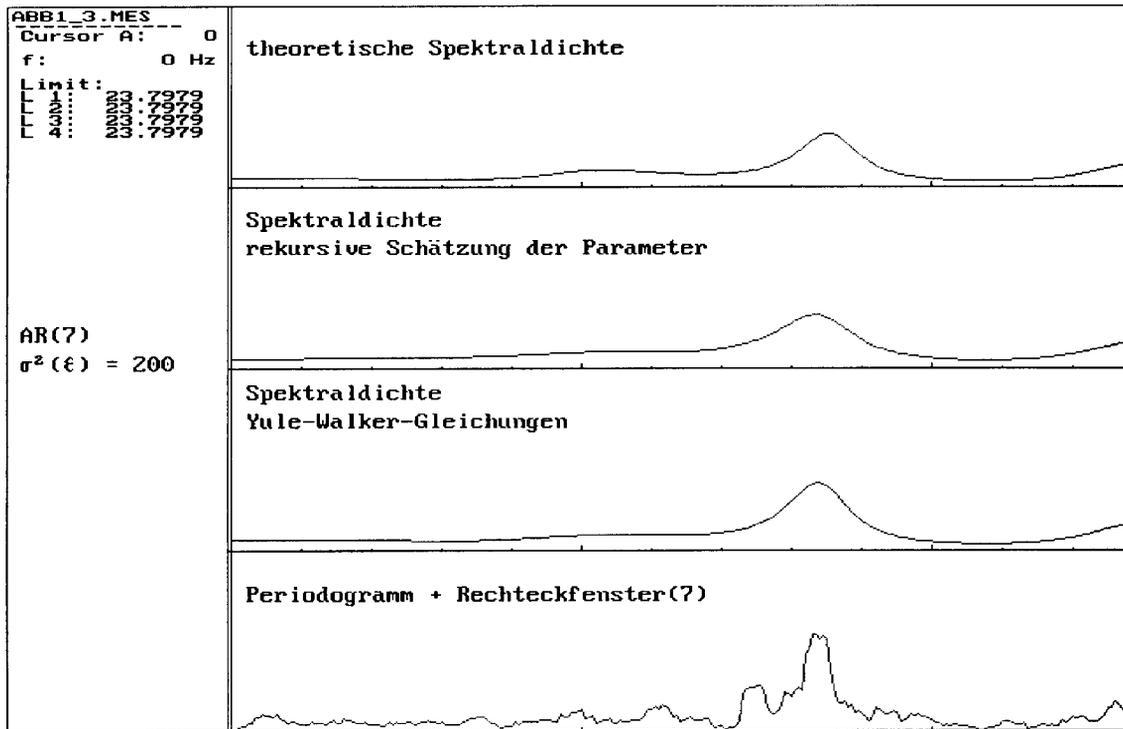


Abb.1.3: Vergleich der Spektraldichteschätzungen

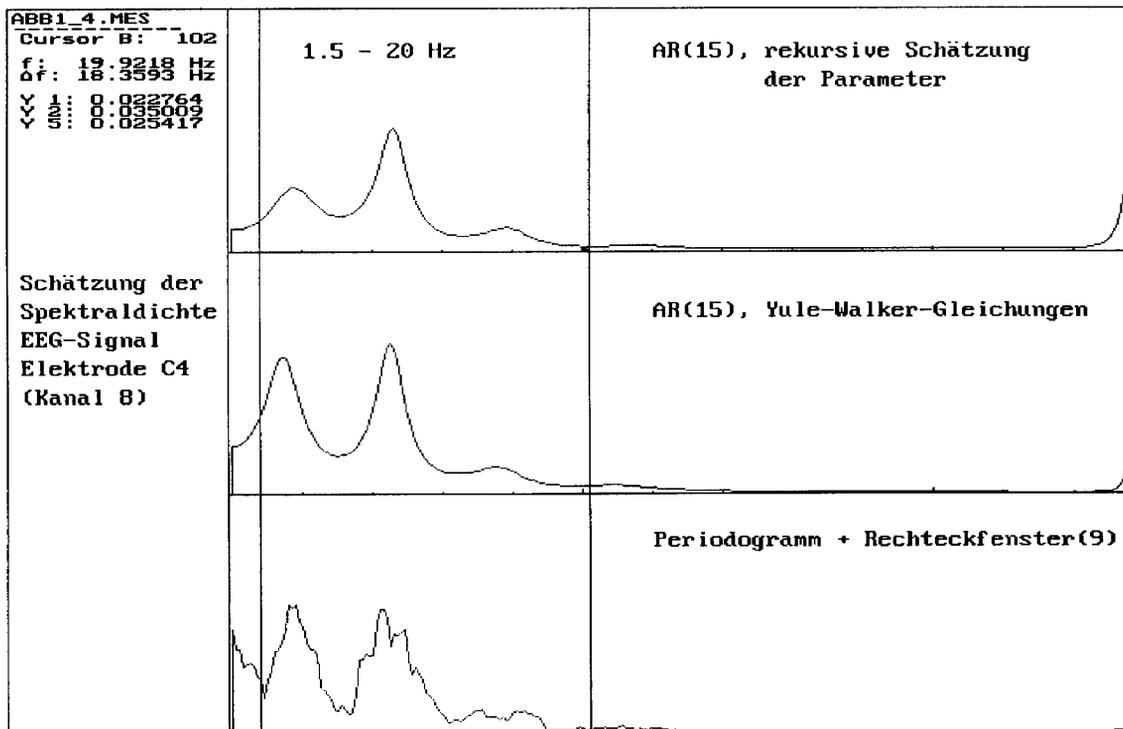


Abb.1.4: Vergleich von Schätzungen der Spektraldichte eines EEG-Signals

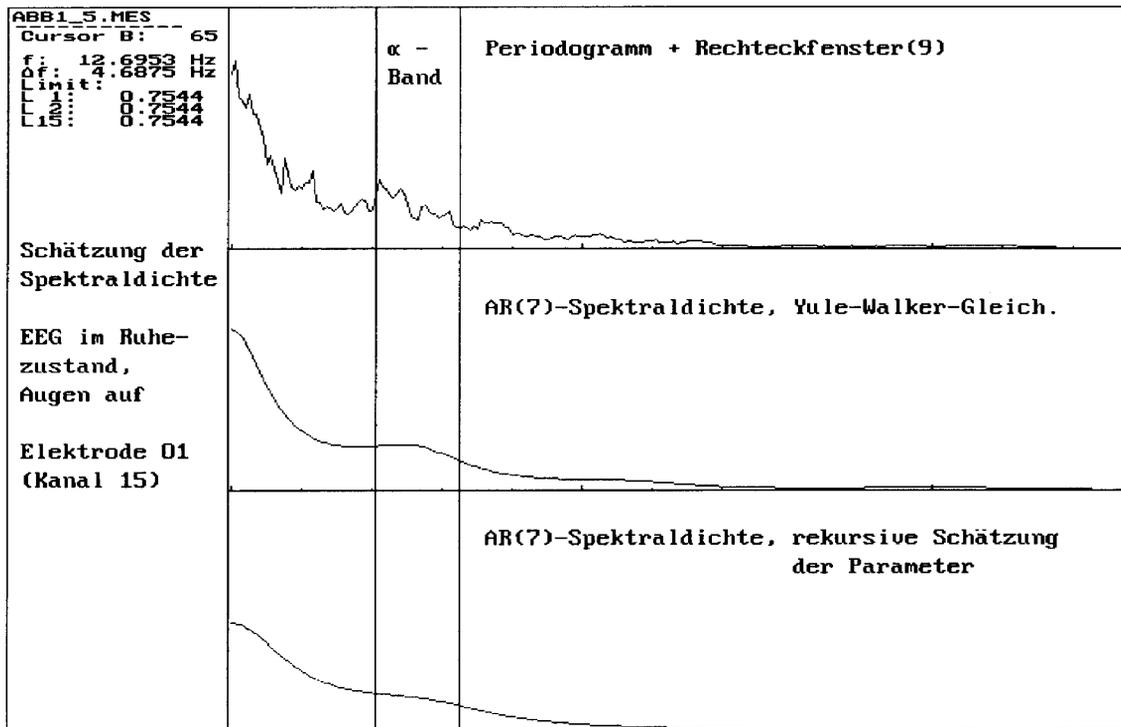


Abb.1.5: Bei geöffneten Augen ist im  $\alpha$ -Band des EEG- Spektrums an der Elektrode O1 keine Aktivität zu erkennen

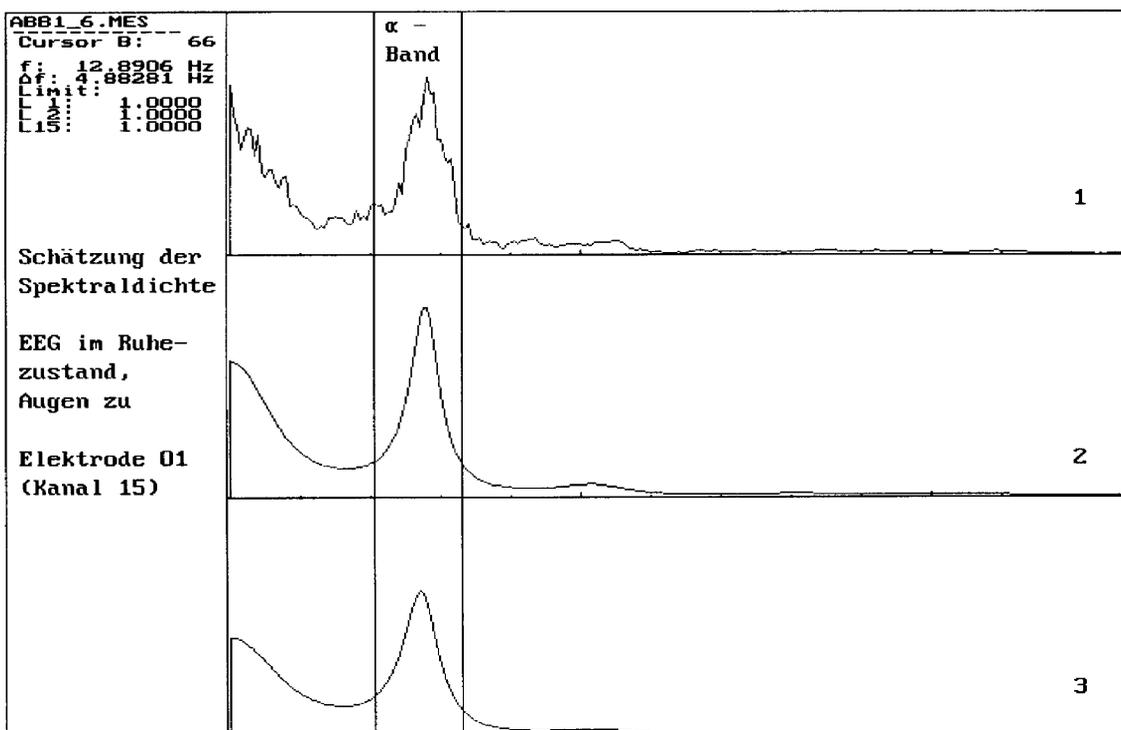


Abb.1.6: Der typische Leistungszuwachs im  $\alpha$ - Band des EEGs an der Elektrode O1 bei geschlossenen Augen wurde in der parametrischen Spektraldichte richtig widerspiegelt. 1: Periodogramm, (geglättet mit einem Rechteckfenster der Breite 9); 2: AR(11)-Spektraldichte, Parameterschätzung mit Hilfe des Levinson-Durbin-Algorithmus; 3: AR(11)-Spektraldichte, rekursive Schätzung der Parameter

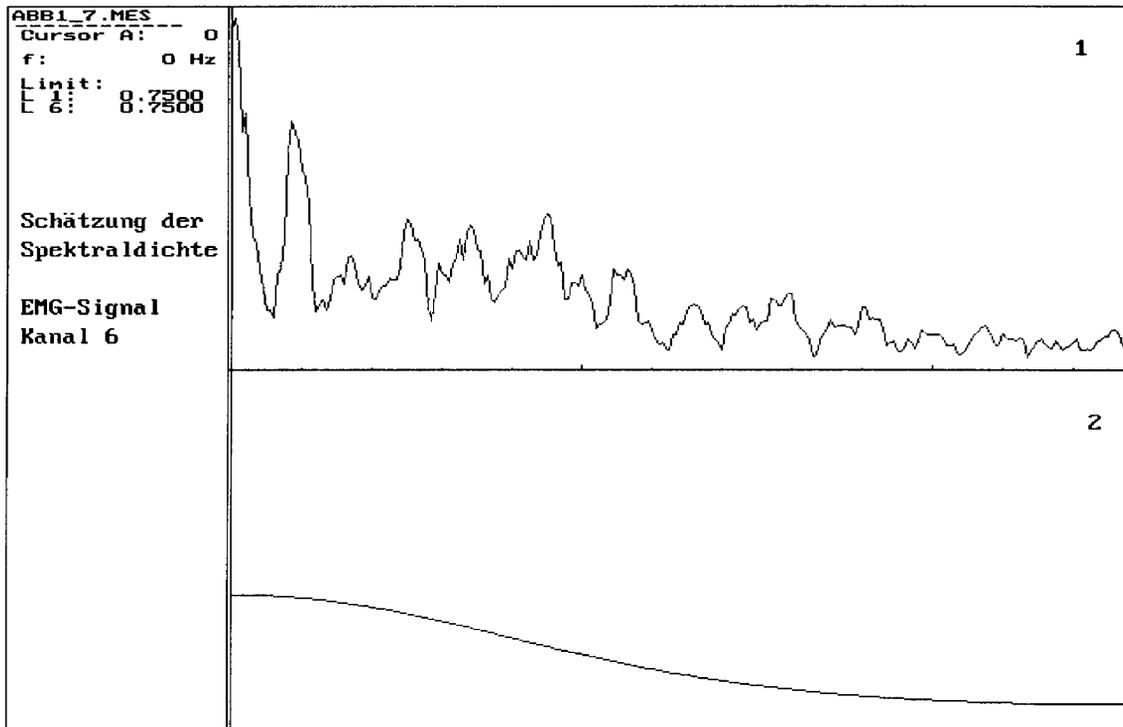


Abb.1.7: Im Kanal 6 des EMG-Signals ist keine wesentliche Muskelaktivität erkennbar. Schätzung der Spektraldichte. 1: Periodogramm (geglättet mit einem Rechteckfenster der Breite 7); 2: AR(2)-Spektraldichte, Parameterschätzung mit Hilfe des Levinson-Durbin-Algorithmus

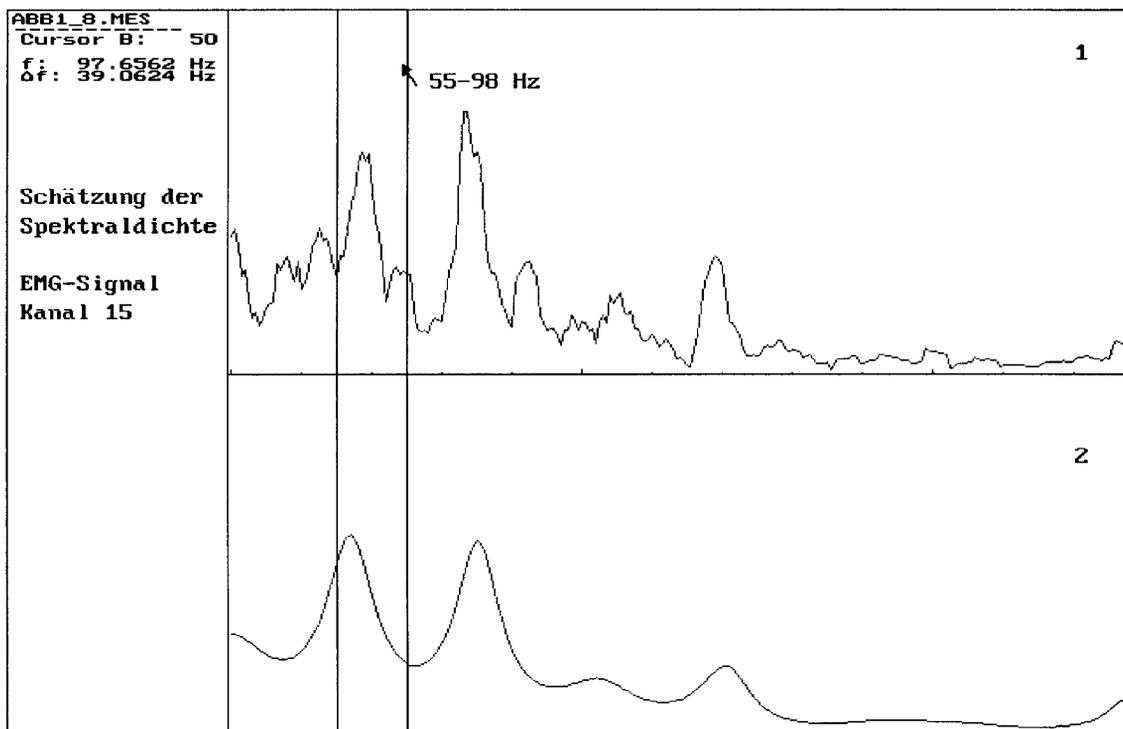


Abb.1.8: Charakteristische Muskelaktivität im Spektralband 55-98 Hz ist deutlich erkennbar. 1: Periodogramm (geglättet mit einem Rechteckfenster der Breite 7); 2: AR(14)-Spektraldichte, Parameterschätzung mit Hilfe des Levinson-Durbin-Algorithmus

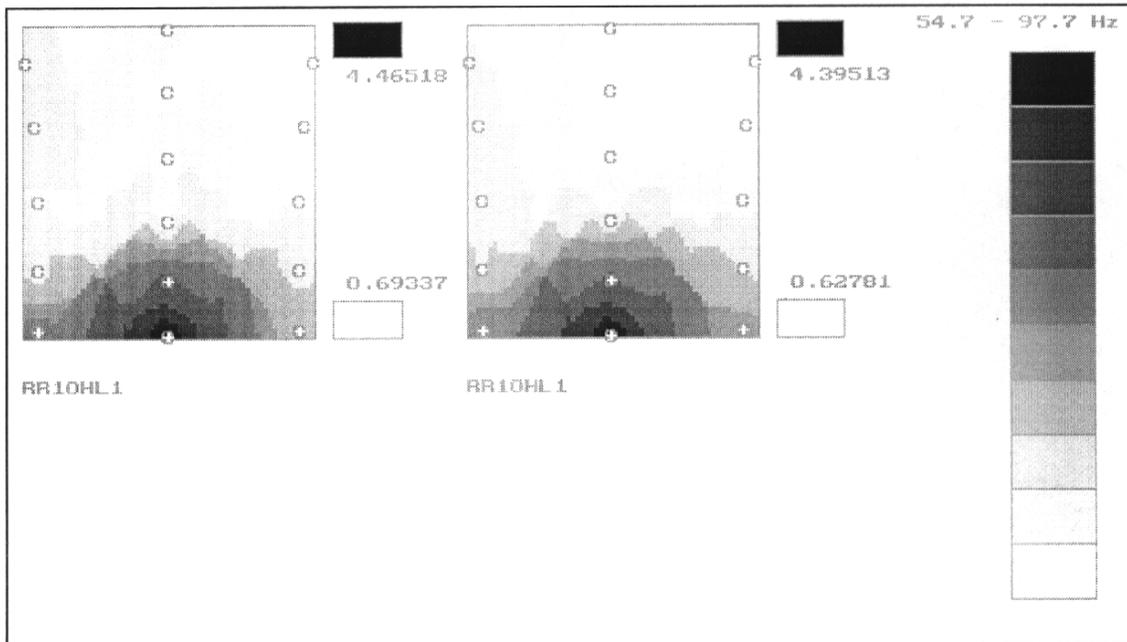


Abb.1.9: Links: Map über AR-Spektraldichte, Rechts: Map über Periodogramm

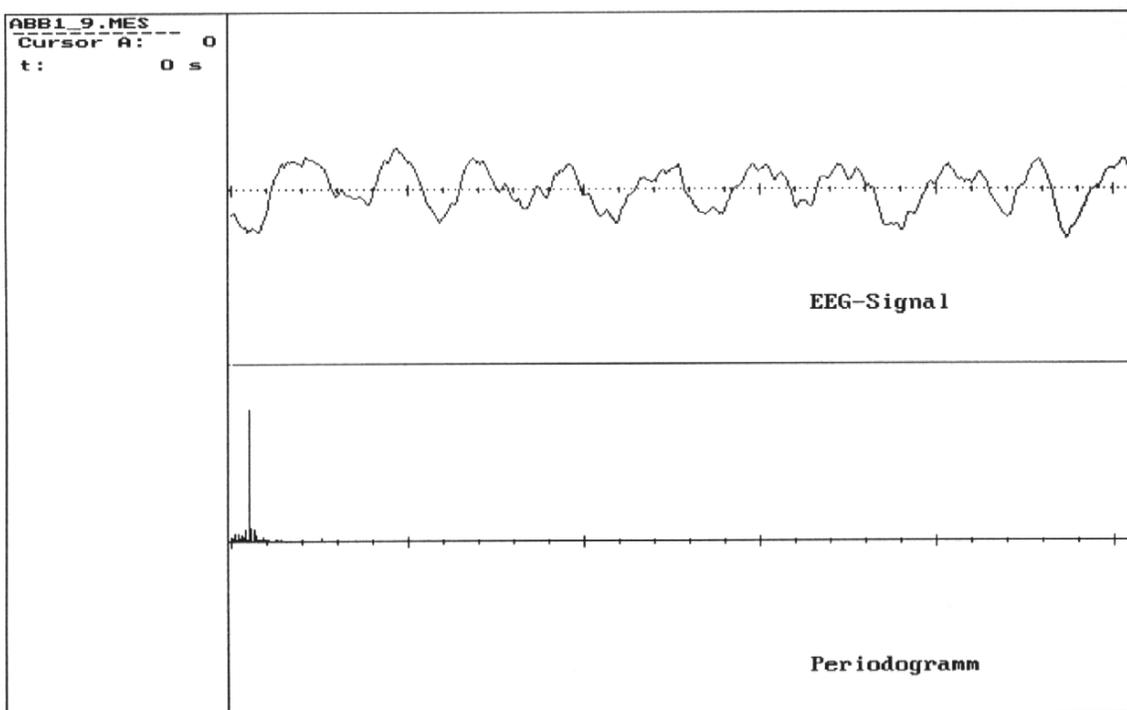


Abb.1.10: Bei diesem Signal wird der Levinson-Durbin-Algorithmus im 3. Schritt instabil

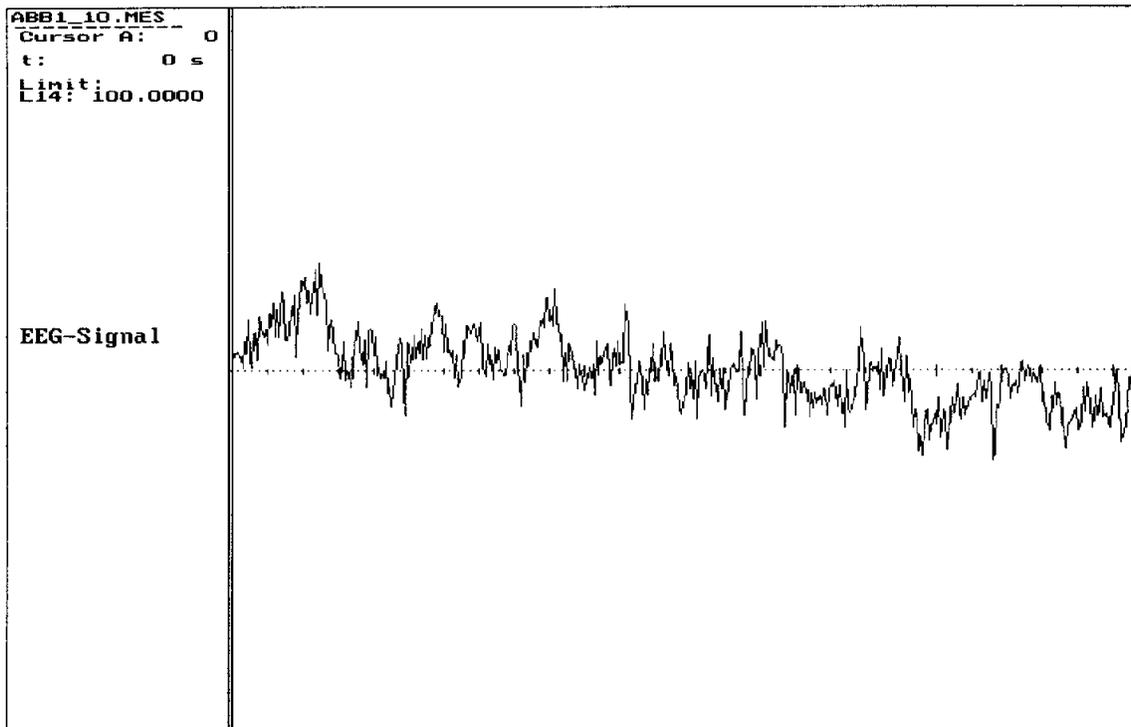


Abb.1.11: EEG-Signal mit Trend

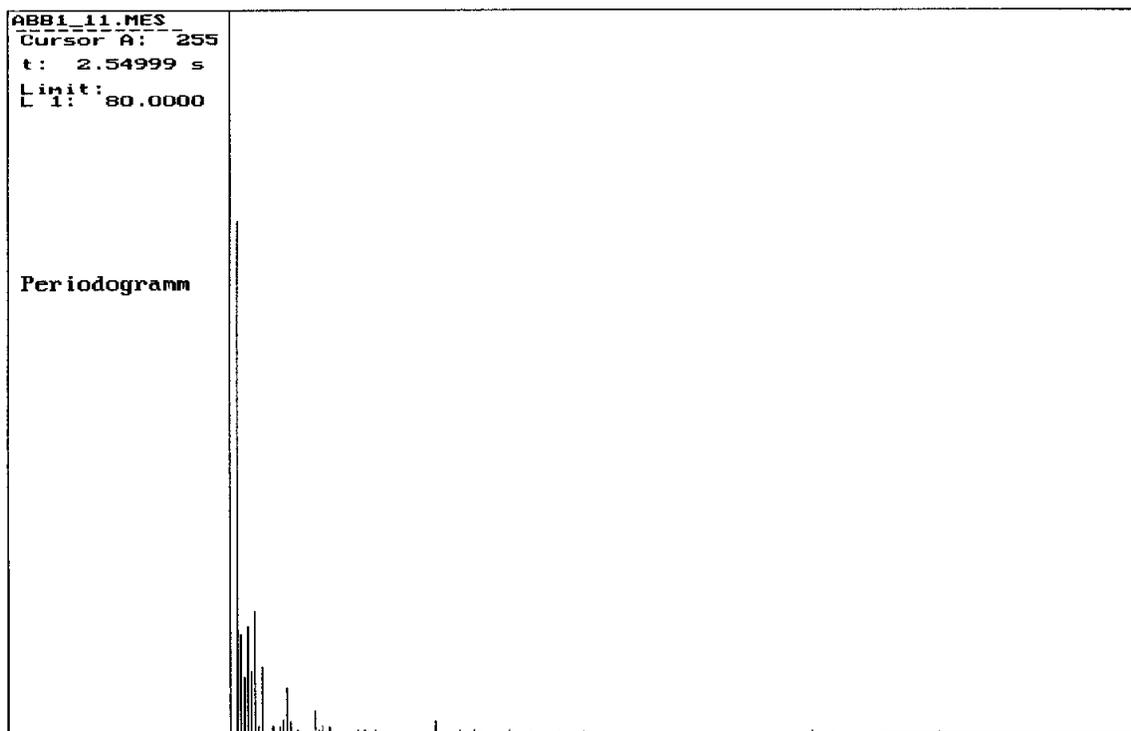


Abb.1.12: Periodogramm des EEG-Signals aus Abb.1.11

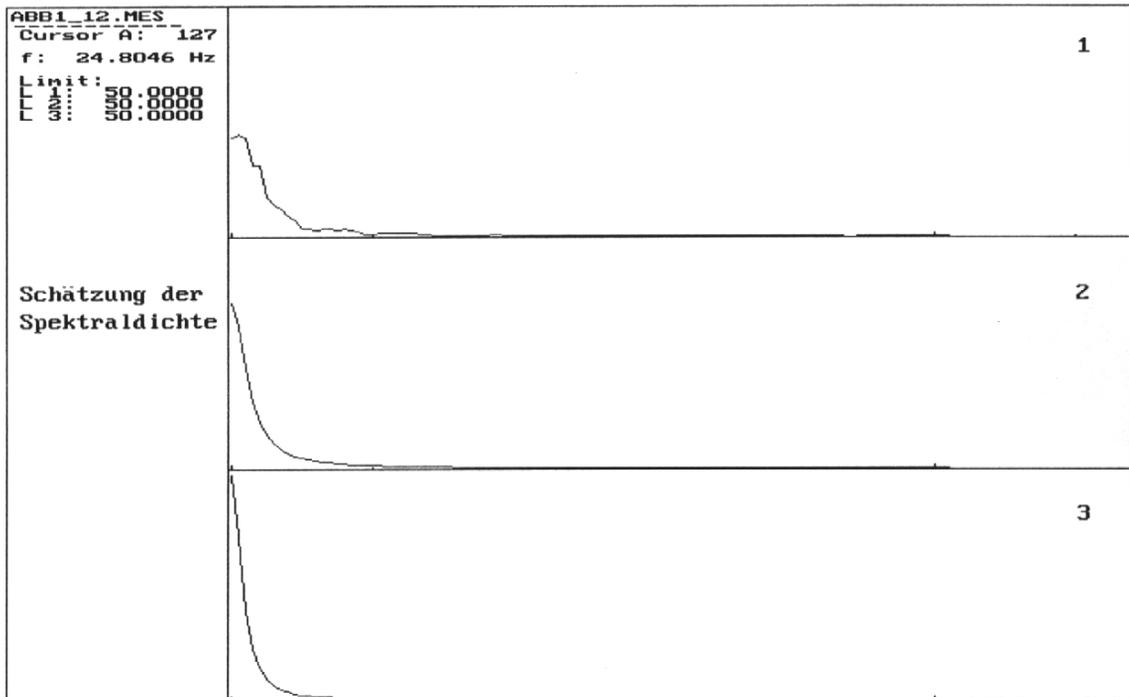


Abb.1.13: Spektraldichtenvergleich zu Abb. 1.11. 1: Periodogramm (geglättet mit einem Rechteckfenster der Breite 7); 2,3: AR(14)-Spektraldichte, Parameterschätzung mit Hilfe des Levinson-Durbin-Algorithmus,  $\hat{\sigma}_\varepsilon^2 = 24.08$  (2), rekursive Parameterschätzung,  $\hat{\sigma}_\varepsilon^2(512) = 38.1$  (3)

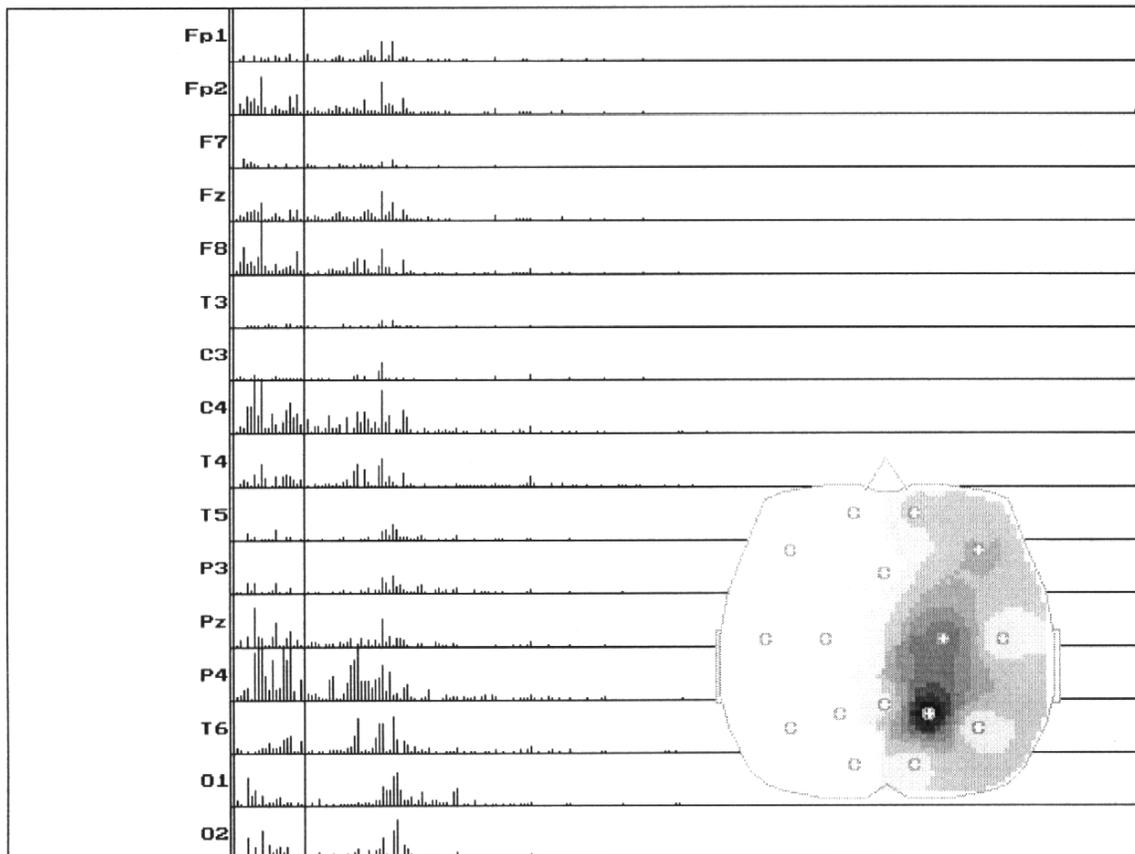


Abb.1.14: EEG-Spektralmap im  $\delta$ -Band bei einem Gehirntumor, berechnet aus dem Periodogramm

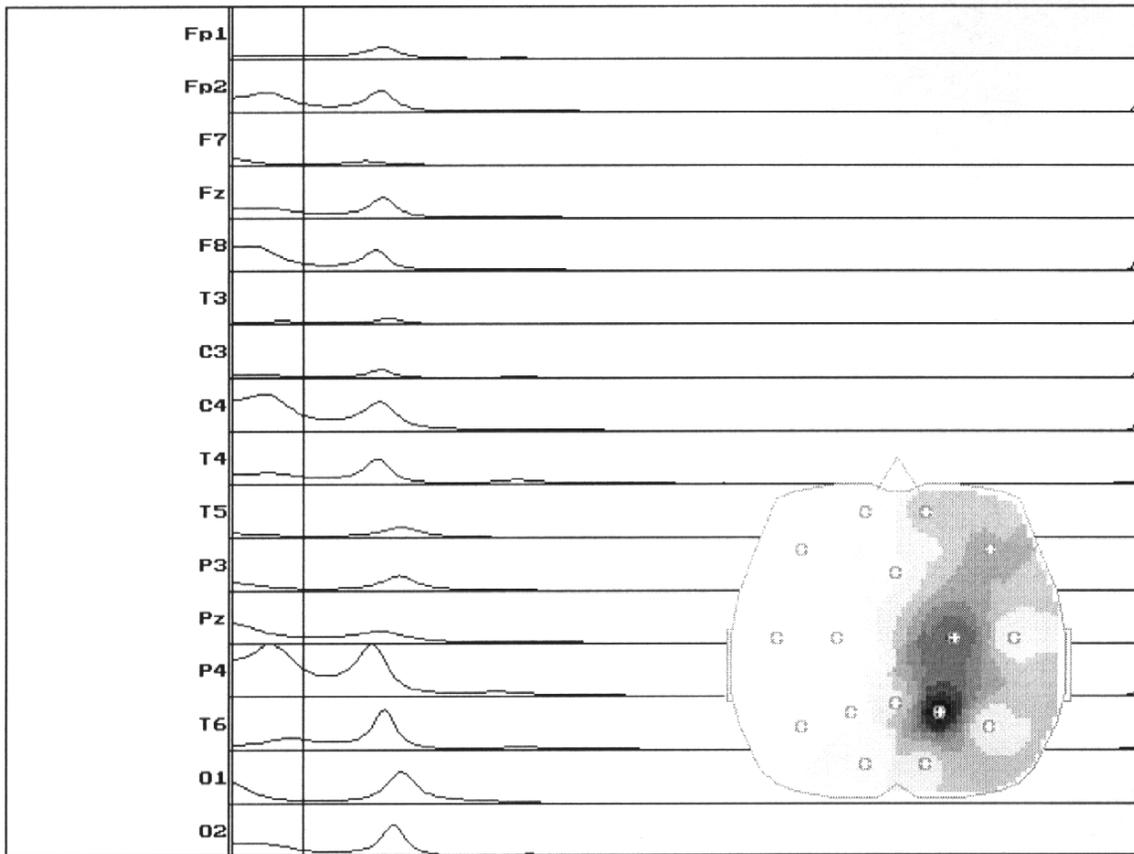


Abb.1.15: EEG-Spektralmap im  $\delta$ -Band eines Gehirntumors berechnet aus dem angepaßten ARMA-Modell

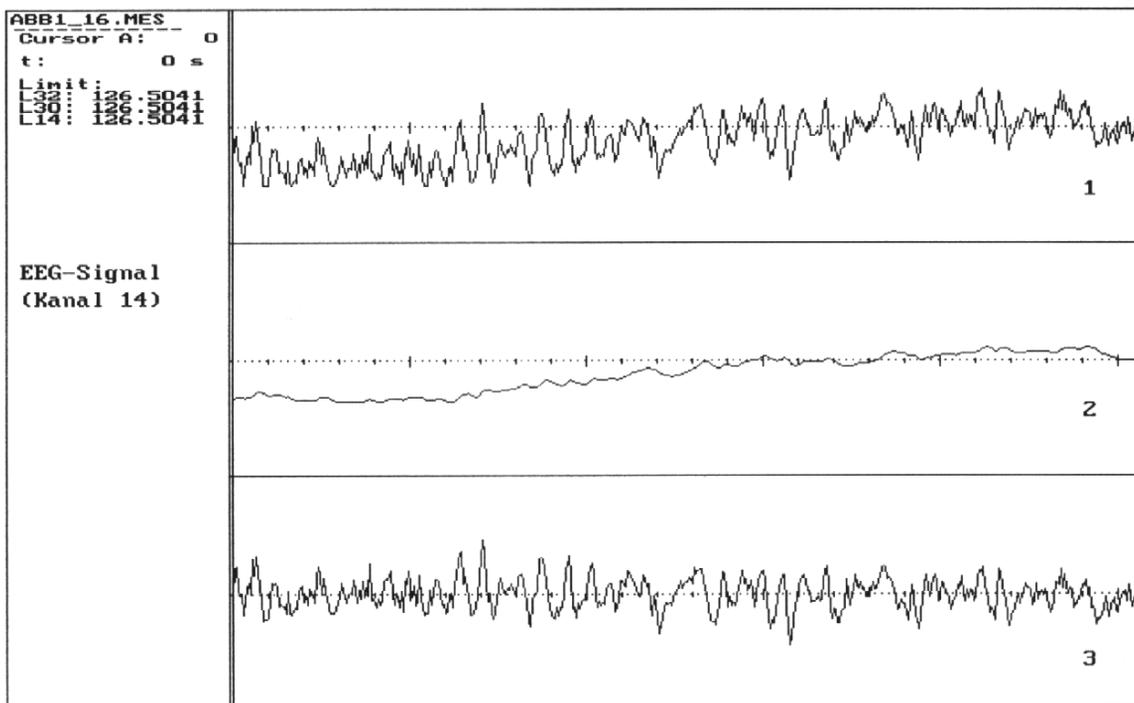


Abb.1.16: EEG-Signal an der Elektrode T6 überlagert mit einem Trend und die Trenddetektion und -elimination

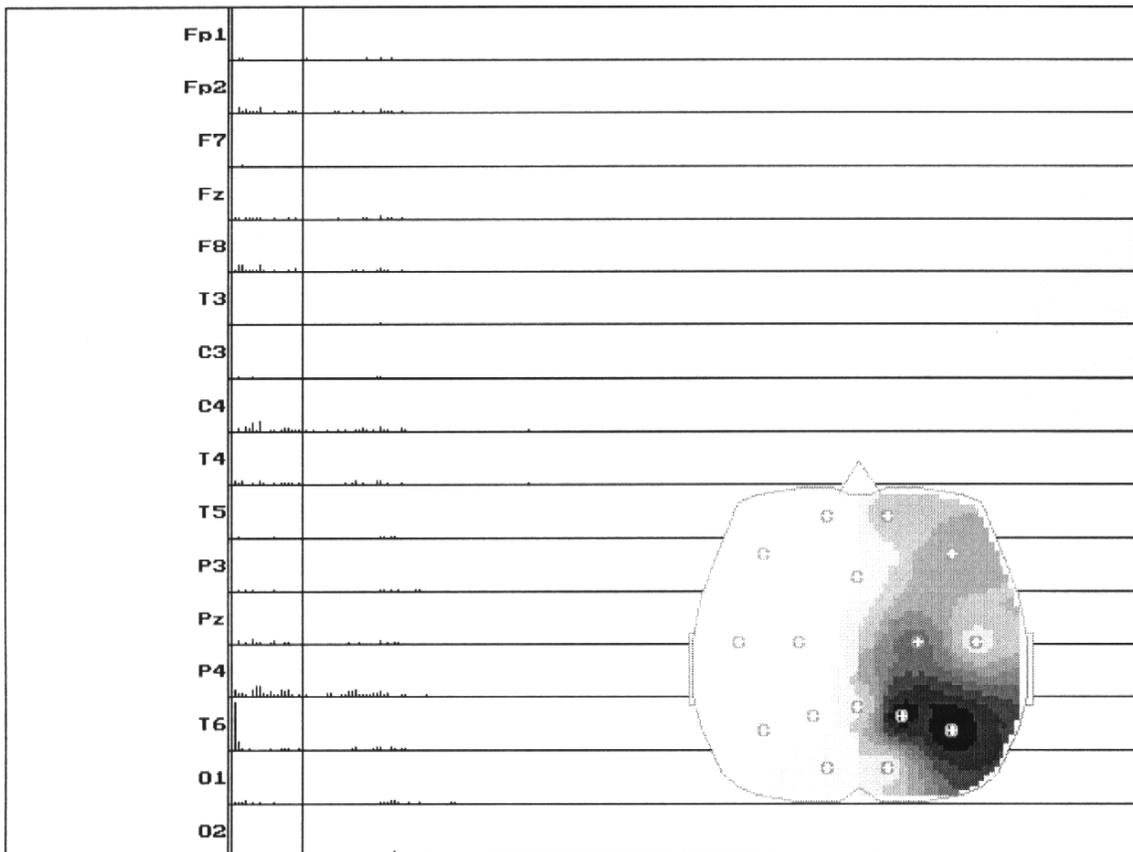


Abb.1.17.: EEG-Spektralmap im d-Band eines Gehirntumors berechnet aus dem Periodogramm ohne Trendelimination im Zeitbereich

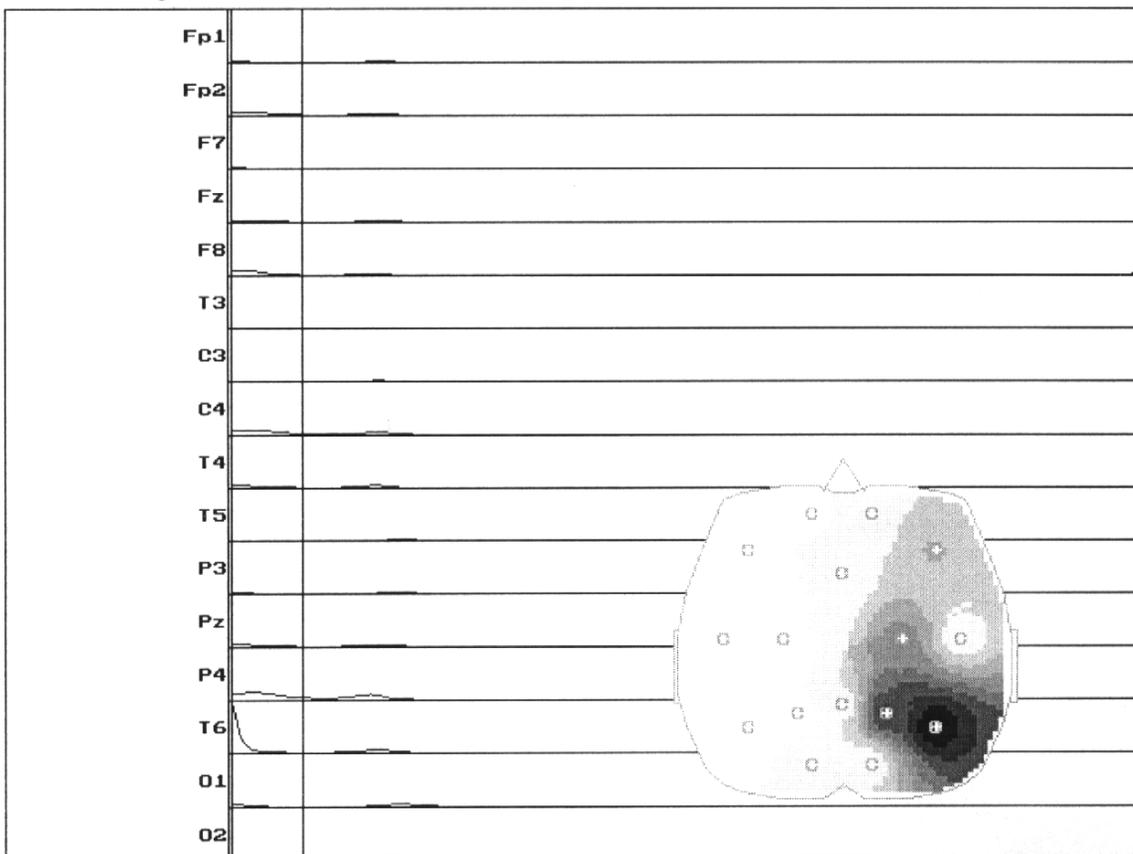


Abb.1.18.: EEG-Spektralmap im d-Band eines Gehirntumors berechnet aus dem angepaßten ARMA-Modell ohne Trendelimination im Zeitbereich

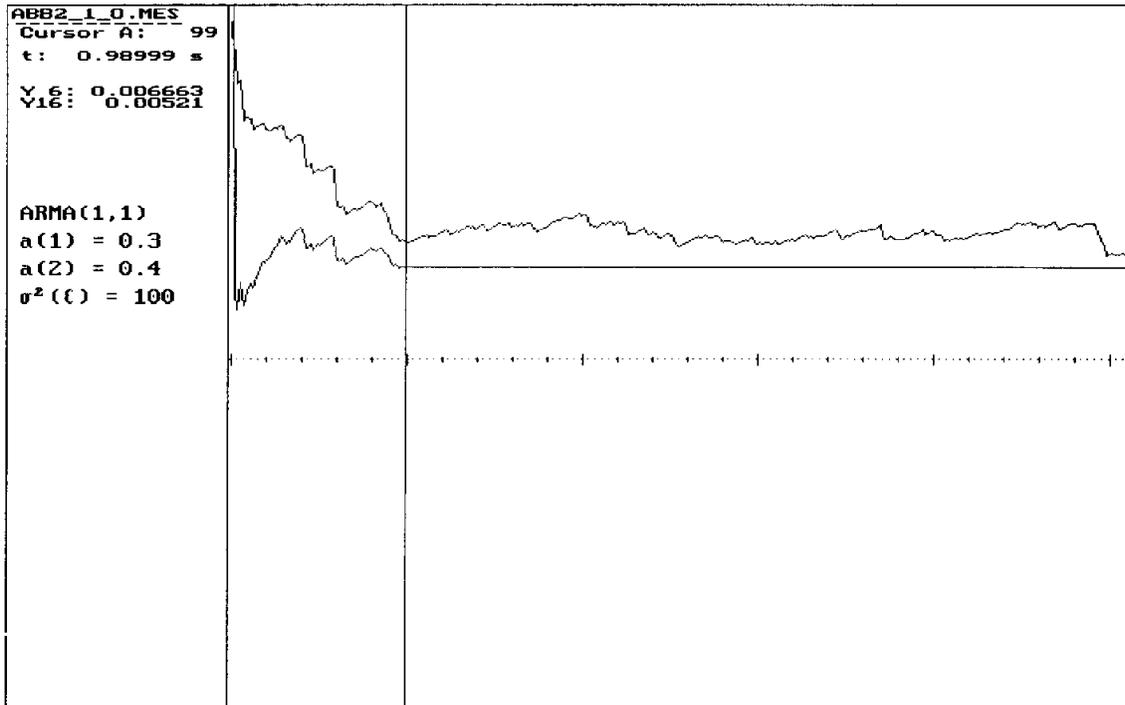


Abb.2.1: Verlauf der Adaptionnorm  $c_x(n) = (\hat{\sigma}_x^2(n))^{-1}$  bei adaptiver Schätzung der Varianz (obere Kurve) und bei mittlerer empirischer Varianz der ersten 100 Meßwerte (untere Kurve)

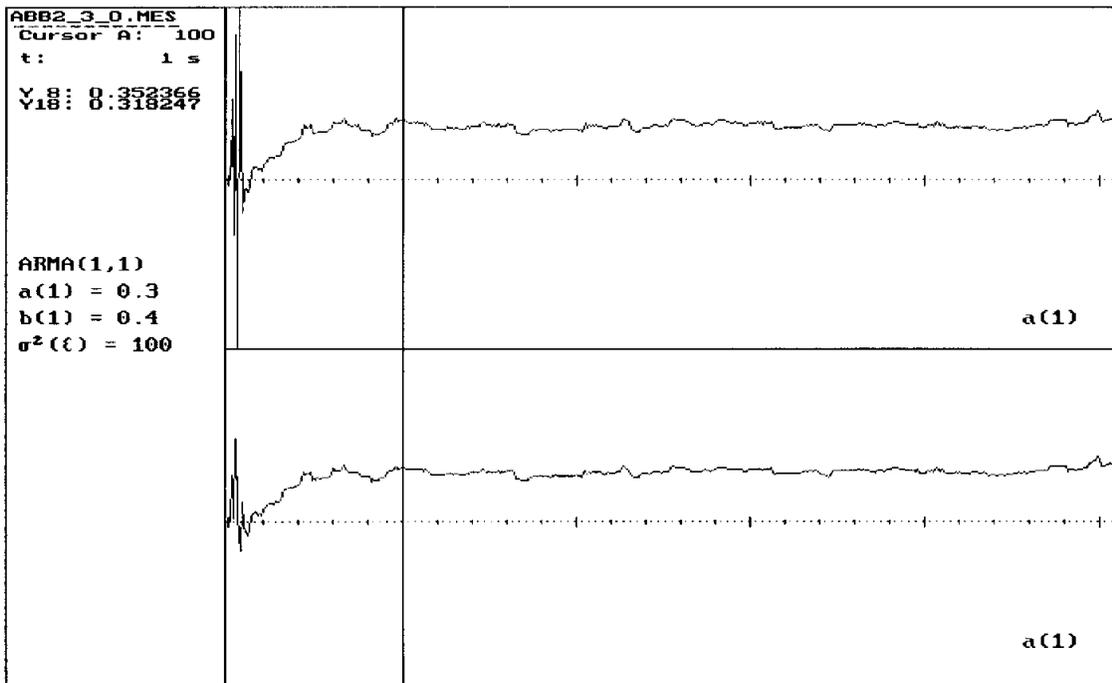


Abb.2.2: Schätzfolge  $\hat{a}_1(t)$  für adaptive (obere Kurve) und konstante Adaptionnorm (untere Kurve)

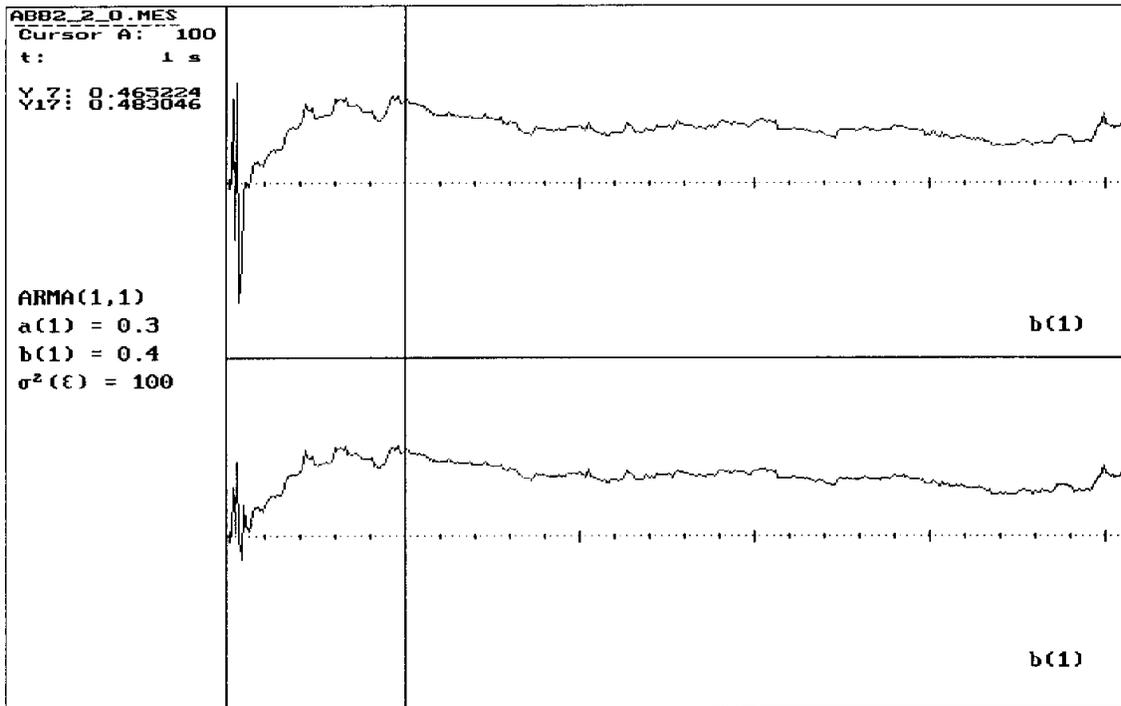


Abb.2.3: Schätzfolge  $\hat{b}_1(t)$  für adaptive (obere Kurve) und konstante Adaptionnorm (untere Kurve)

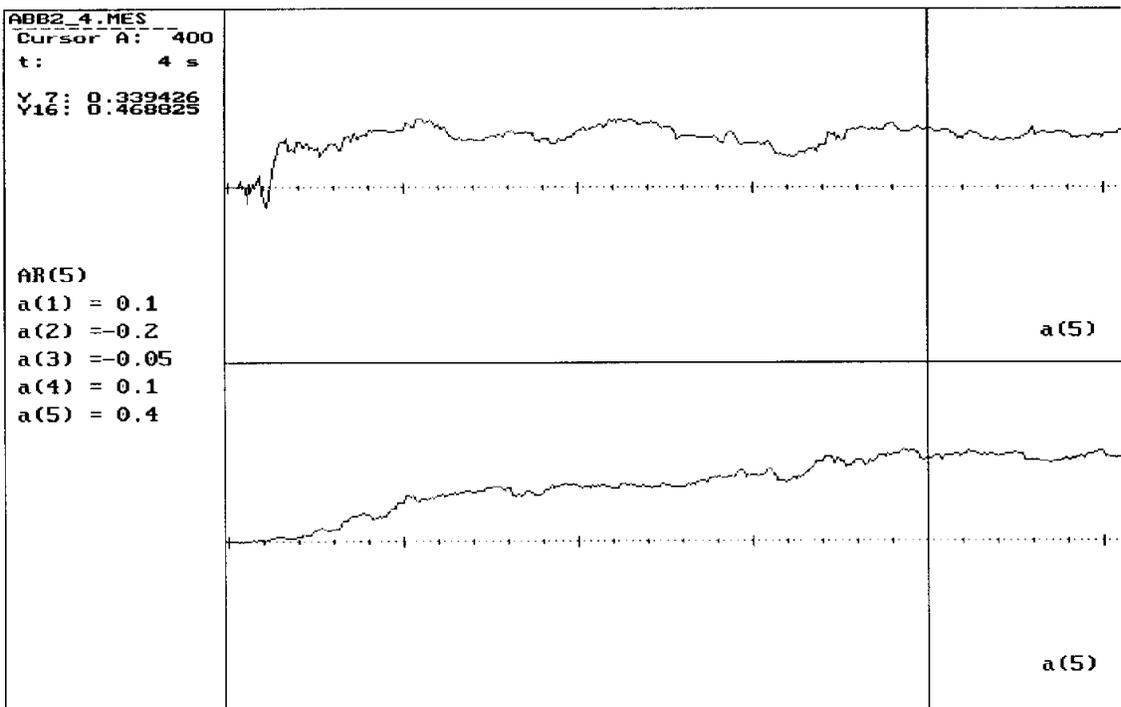


Abb.2.4: Schätzfolge  $\hat{a}_5(t)$  eines AR(5)-Prozesses mit  $\sigma_\epsilon^2 = 100$  (obere Kurve) und  $\sigma_\epsilon^2 = 0.5$  (untere Kurve)

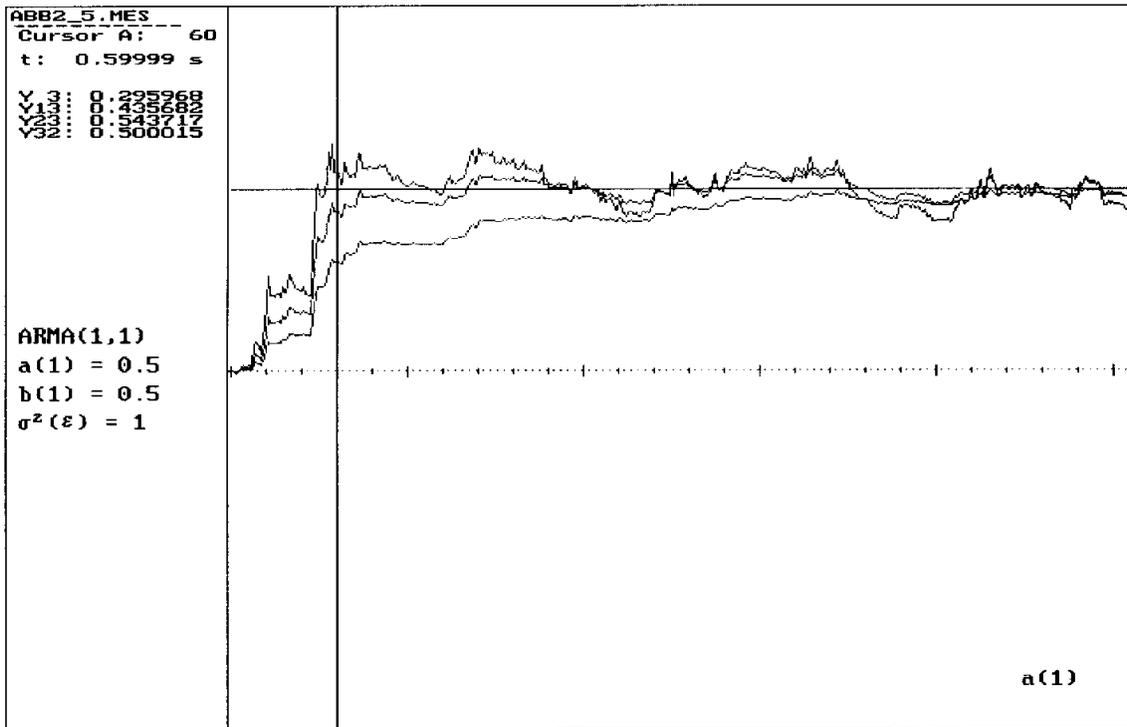


Abb.2.5: Schätzfolge  $\hat{a}_1(t)$  für verschiedene Werte des Adaptionfaktors:  $d = 0.005$  (flachste Kurve),  $d = 0.01$  (mittlere Kurve) und  $d = 0.02$  (obere Kurve) im Vergleich mit dem wahren Parameter  $a_1 = 0.5$  (Konstante)

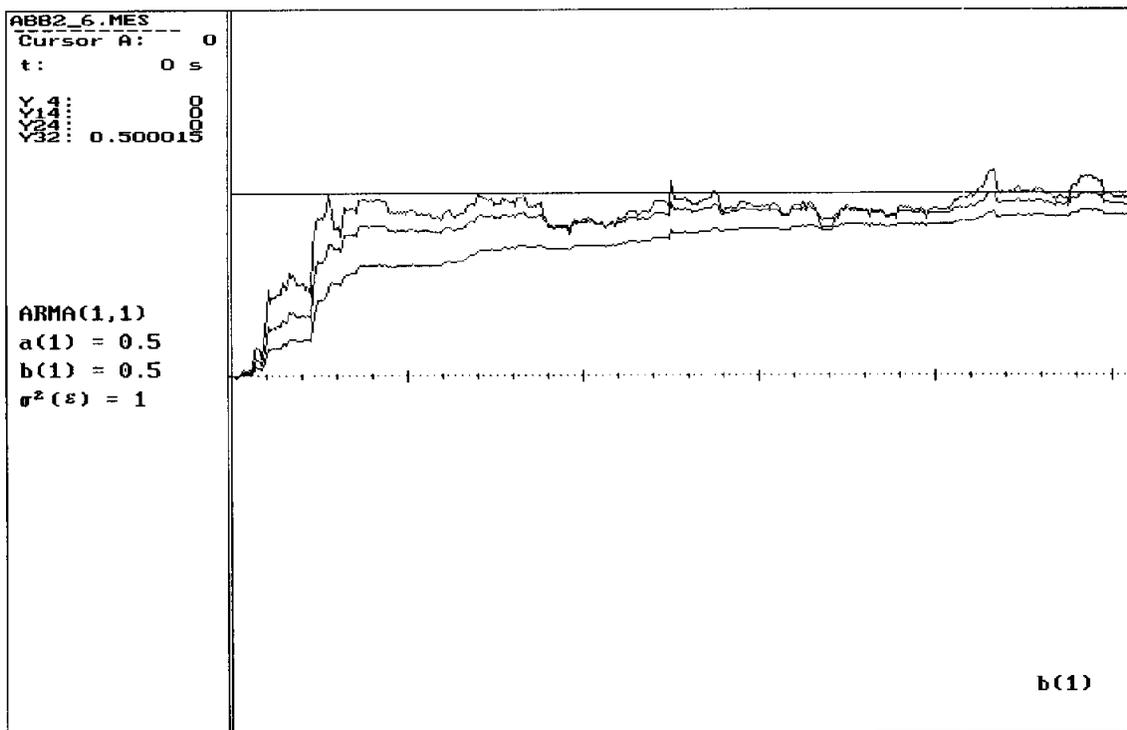


Abb.2.6: Schätzfolge  $\hat{b}_1(t)$  für verschiedene Werte des Adaptionfaktors:  $d = 0.005$  (flachste Kurve),  $d = 0.01$  (mittlere Kurve) und  $d = 0.02$  (obere Kurve), verglichen mit dem wahren Parameter  $b_1 = 0.5$  (Konstante)

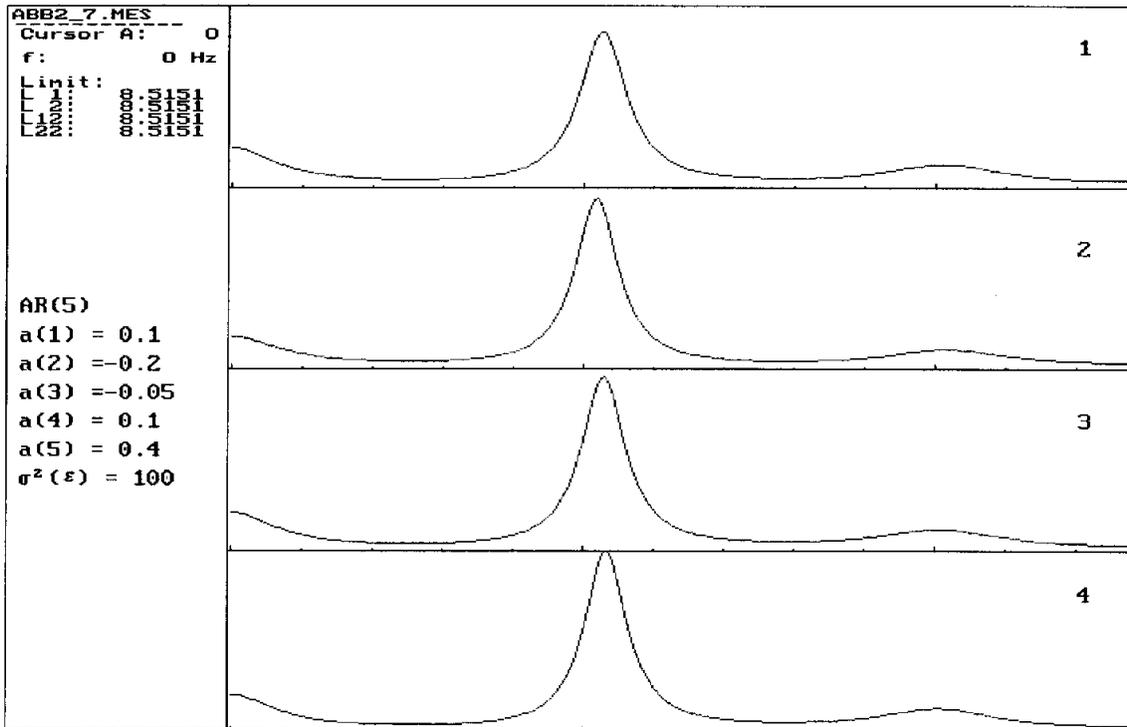


Abb.2.7: Spektraldichte eines AR(5)-Prozesses (4. Kurve) und ihre verschiedene Schätzungen: adaptive Schätzung der Parameter (1. Kurve ); rekursive Schätzung der Parameter (2. Kurve) Parameterschätzung mit Hilfe des Levinson-Durbin-Algorithmus (3. Kurve)

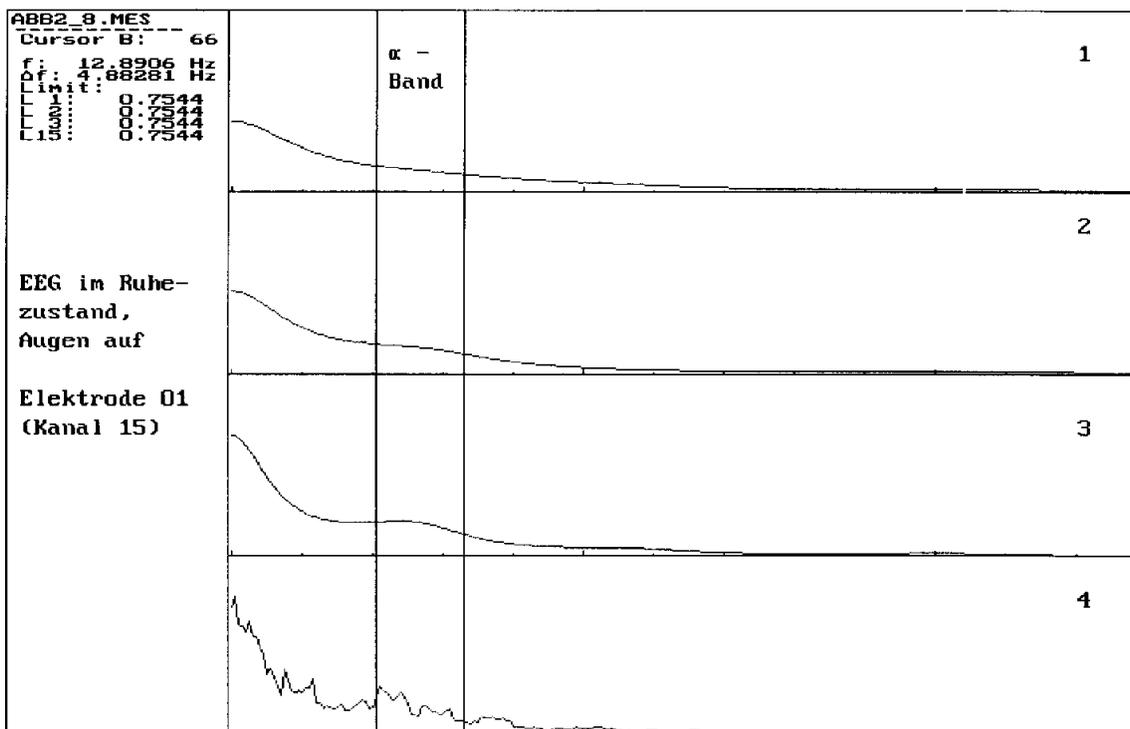


Abb.2.8: Spektraldichteschätzungen des EEGs (Elektrode O1) bei geöffneten Augen. 1: adaptive Schätzung der Parameter; 2: rekursive Schätzung der Parameter; 3: Parameterschätzung mit Hilfe des Levinson-Durbin-Algorithmus; 4: Periodogramm (geglättet durch ein Rechteckfenster der Breite 9)

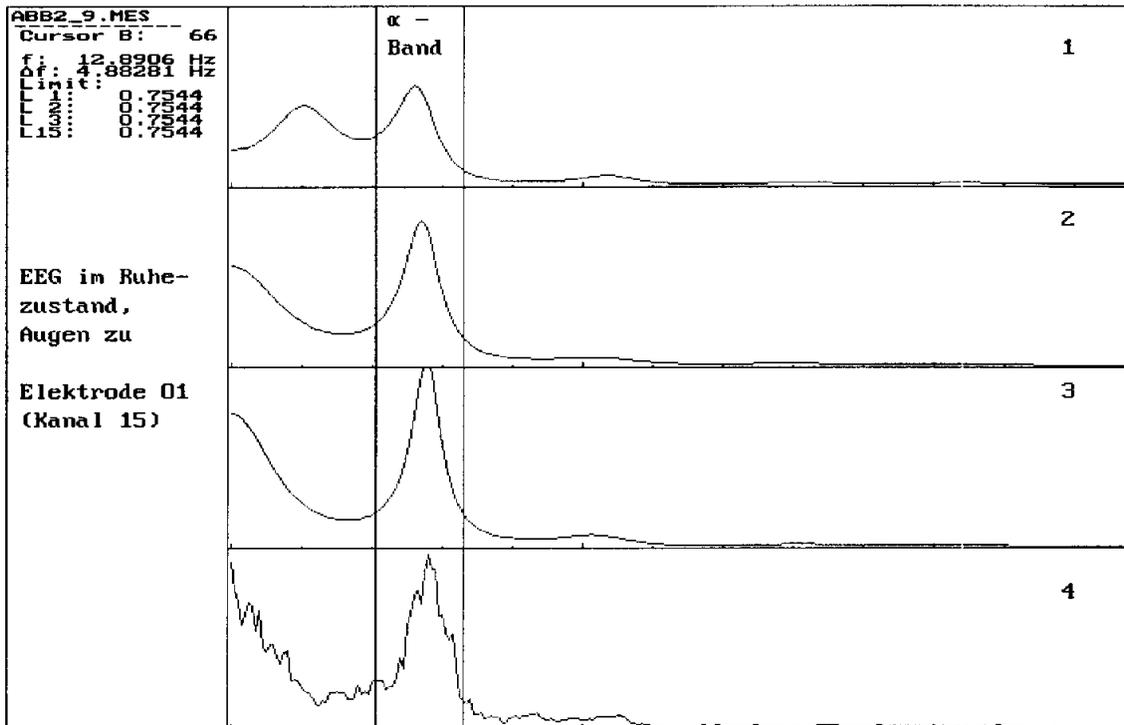
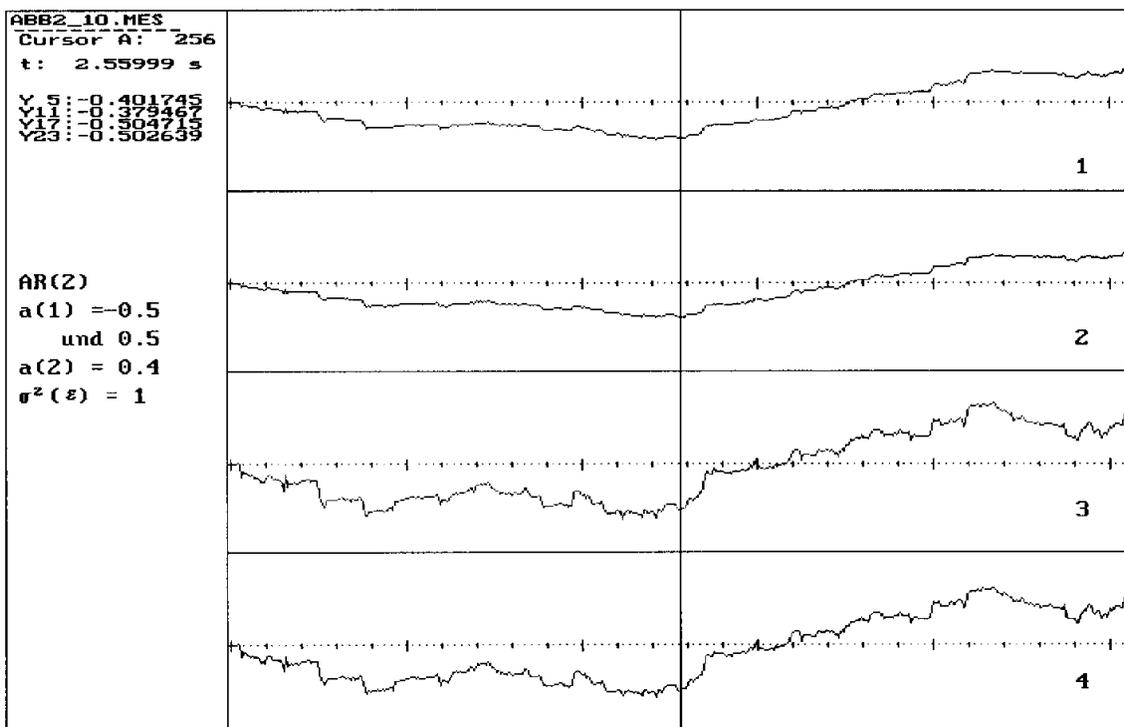


Abb.2.9: Spektraldichteschätzungen des EEGs (Elektrode O1) bei geschlossenen Augen.  
 1: adaptive Schätzung der Parameter; 2: rekursive Schätzung der Parameter; 3: Parameterschätzung mit Hilfe des Levinson-Durbin-Algorithmus; 4: Periodogramm



(geglättet durch ein Rechteckfenster der Breite 9)

Abb.2.10: Schätzfolgenfunktionen  $\hat{a}_i(t)$  mit unterschiedlichen Adaptionen. 1:  $c = 0.01$ ,  $d = 0.01$ ; 2:  $c = 0.03$ ,  $d = 0.01$ ; 3:  $c = 0.01$ ,  $d = 0.03$ ; 4:  $c = 0.03$ ,  $d = 0.03$

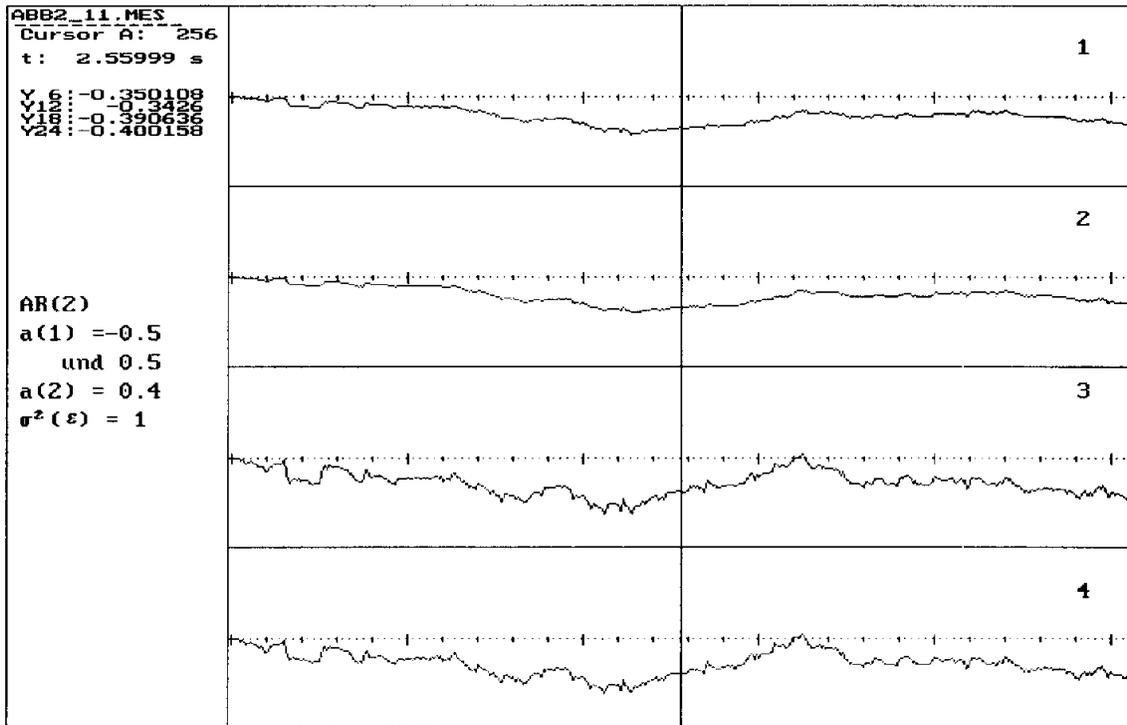


Abb.2.11: Schätzfolgenfunktionen  $\hat{a}_j(t)$  mit unterschiedlichen Adaptiongrößen. 1:  $c = 0.01$ ,  $d = 0.01$ ; 2:  $c = 0.03$ ,  $d = 0.01$ ; 3:  $c = 0.01$ ,  $d = 0.03$ ; 4:  $c = 0.03$ ,  $d = 0.03$

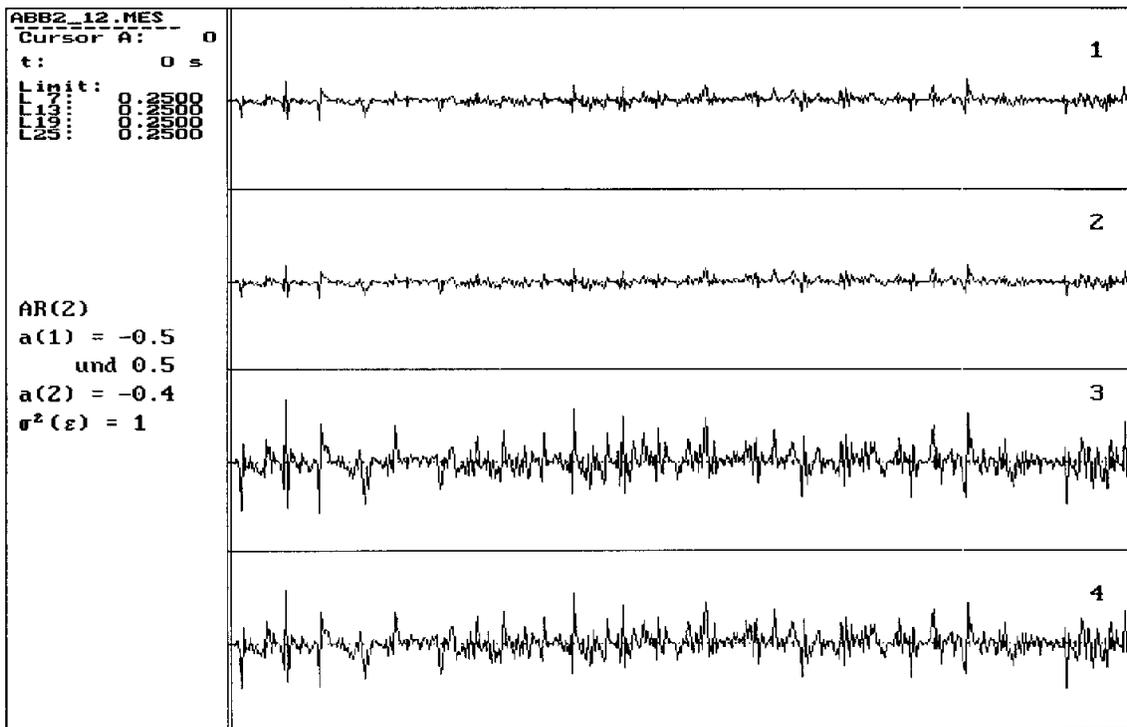


Abb.2.12: Maximale Parameterschwankung  $\kappa(t)$  für unterschiedliche Adaptionvariablen. 1:  $c = 0.01$ ,  $d = 0.01$ ; 2:  $c = 0.03$ ,  $d = 0.01$ ; 3:  $c = 0.01$ ,  $d = 0.03$ ; 4:  $c = 0.03$ ,  $d = 0.03$

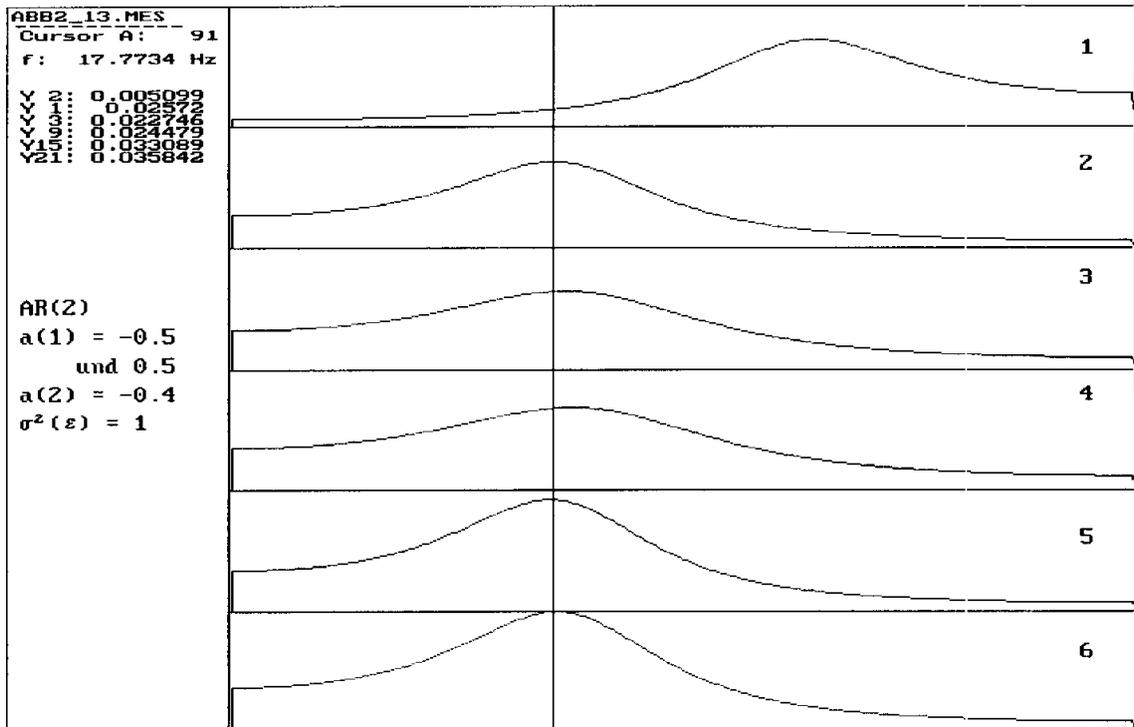


Abb.2.13: Spektraldichten mit den theoretischen Parametern und den Endwerten der adaptiven Schätzungen. 1: theoretische Spektraldichte des ersten AR-Prozesses; 2: theoretische Spektraldichte des zweiten AR-Prozesses; adaptive Schätzung mit 3:  $c = 0.01$ ,  $d = 0.01$ ; 4:  $c = 0.03$ ,  $d = 0.01$ ; 5:  $c = 0.01$ ,  $d = 0.03$ ; 6:  $c = 0.03$ ,  $d = 0.03$

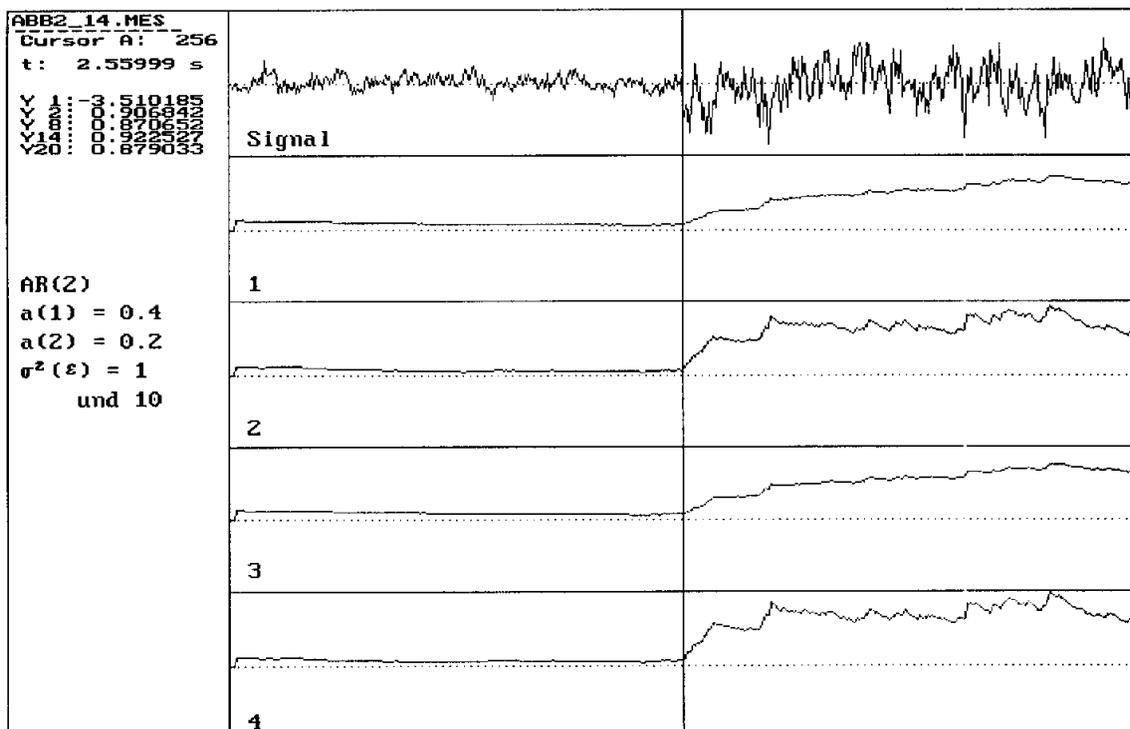


Abb.2.14: Signal mit Leistungssprung und adaptive Schätzfolgen des mittleren quadratischen Vorhersagefehlers  $\sigma_f^2$ . 1:  $c = 0.01$ ,  $d = 0.01$ ; 2:  $c = 0.03$ ,  $d = 0.01$ ; 3:  $c = 0.01$ ,  $d = 0.03$ ; 4:  $c = 0.03$ ,  $d = 0.03$

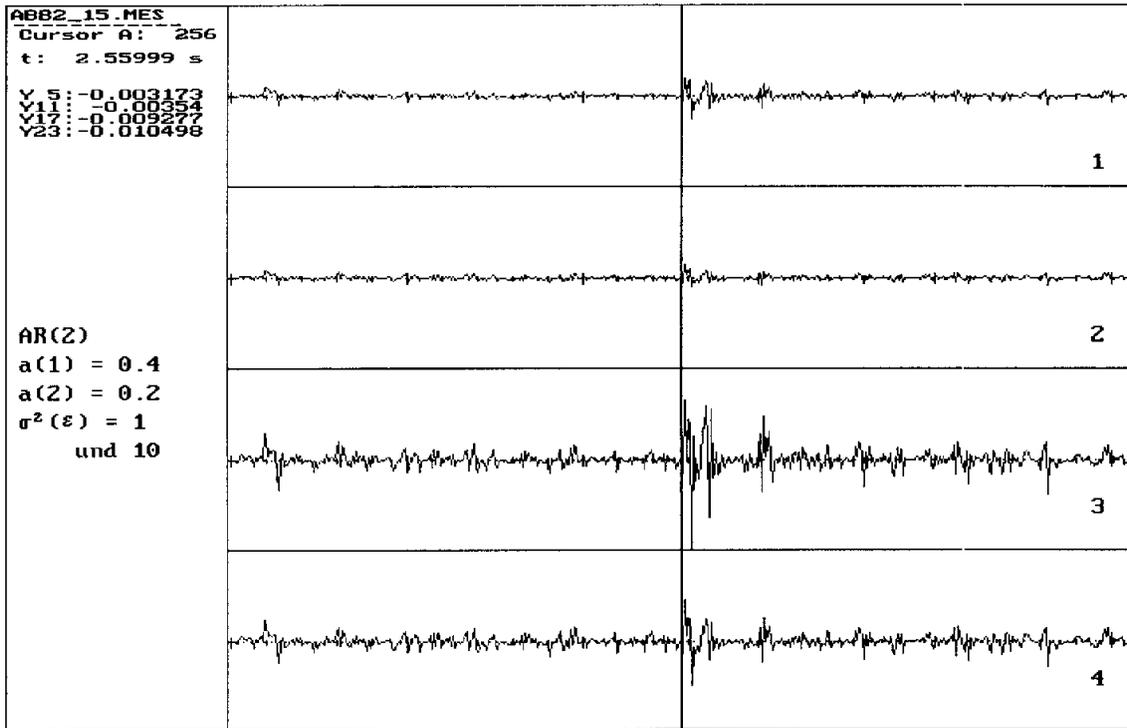


Abb.2.15: Maximale Parameterschwankung  $\kappa(t)$  mit unterschiedlichen Werten von  $c$  und  $d$ .  
 1:  $c = 0.01, d = 0.01$ ; 2:  $c = 0.03, d = 0.01$ ; 3:  $c = 0.01, d = 0.03$ ; 4:  $c = 0.03, d = 0.03$

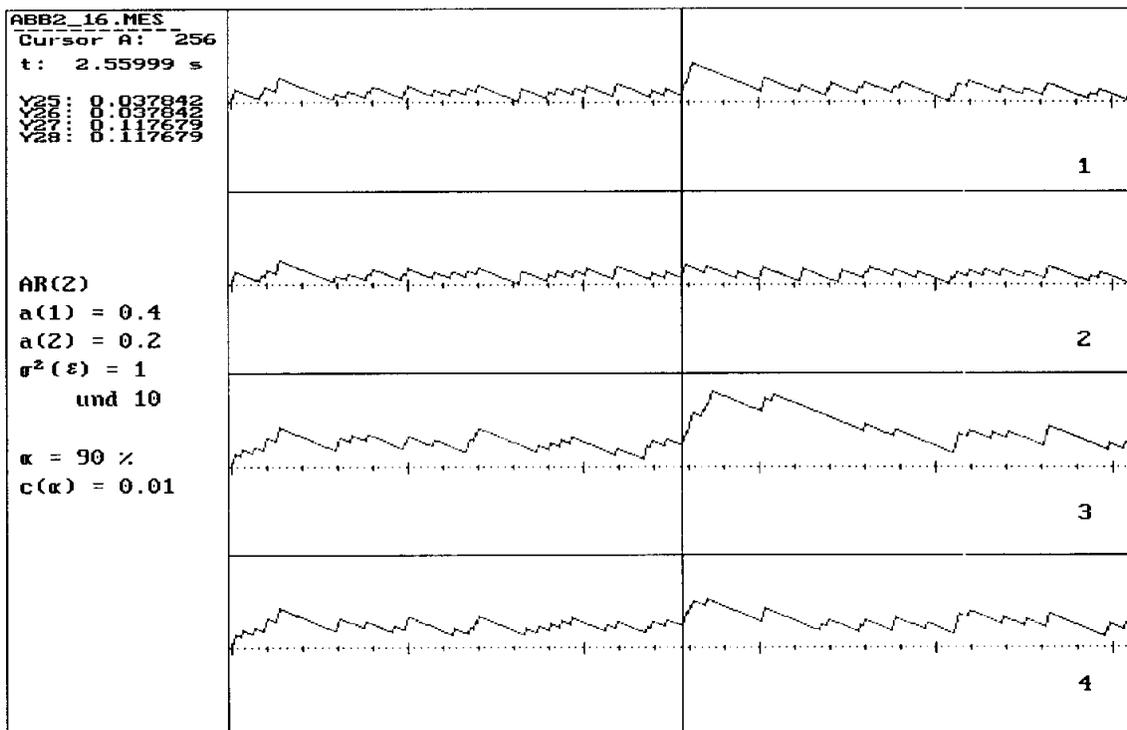


Abb.2.16: Breite adaptiver Interquantilintervalle um die maximale Parameterschwankung.  
 1:  $c = 0.01, d = 0.01$ ; 2:  $c = 0.03, d = 0.01$ ; 3:  $c = 0.01, d = 0.03$ ; 4:  $c = 0.03, d = 0.03$

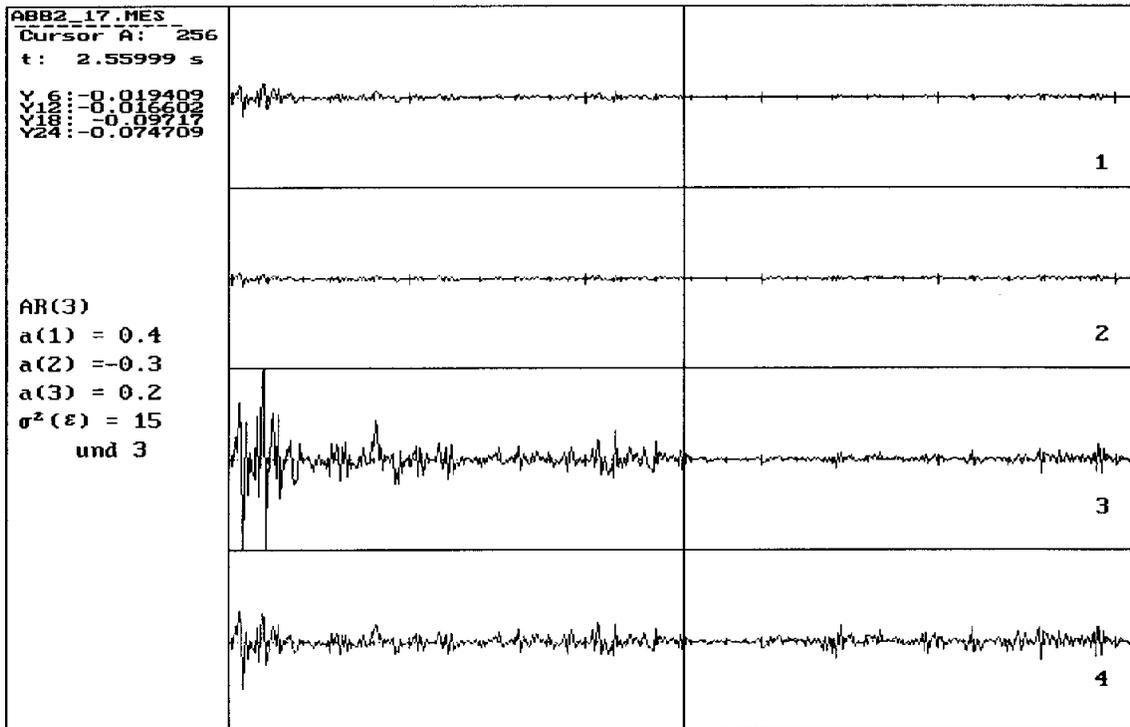


Abb.2.17: Maximale Parameterschwankung  $\kappa(t)$  mit unterschiedlichen Werten von  $c$  und  $d$ .  
 1:  $c = 0.01, d = 0.005$ ; 2:  $c = 0.04, d = 0.005$ ; 3:  $c = 0.01, d = 0.025$ ; 4:  $c = 0.04, d = 0.025$

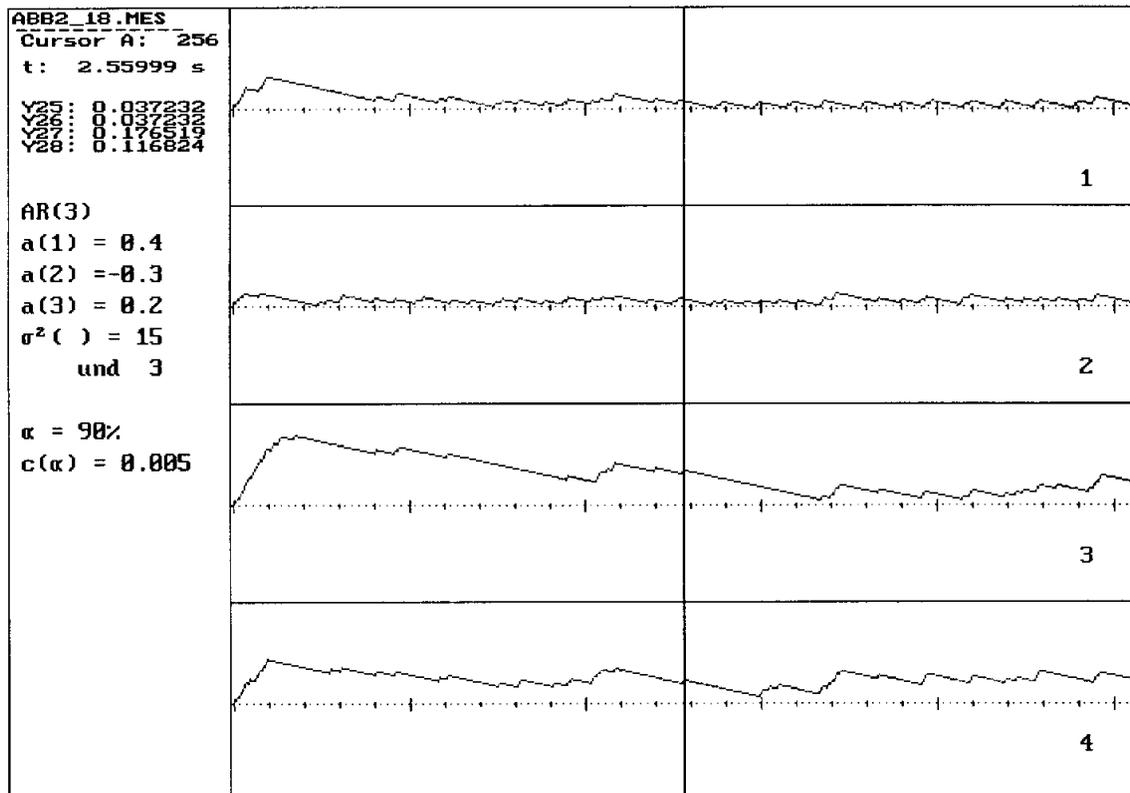


Abb.2.18: Breite adaptiver Interquantilintervalle um die maximale Parameterschwankung.  
 1:  $c = 0.01, d = 0.005$ ; 2:  $c = 0.04, d = 0.005$ ; 3:  $c = 0.01, d = 0.025$ ; 4:  $c = 0.04, d = 0.025$



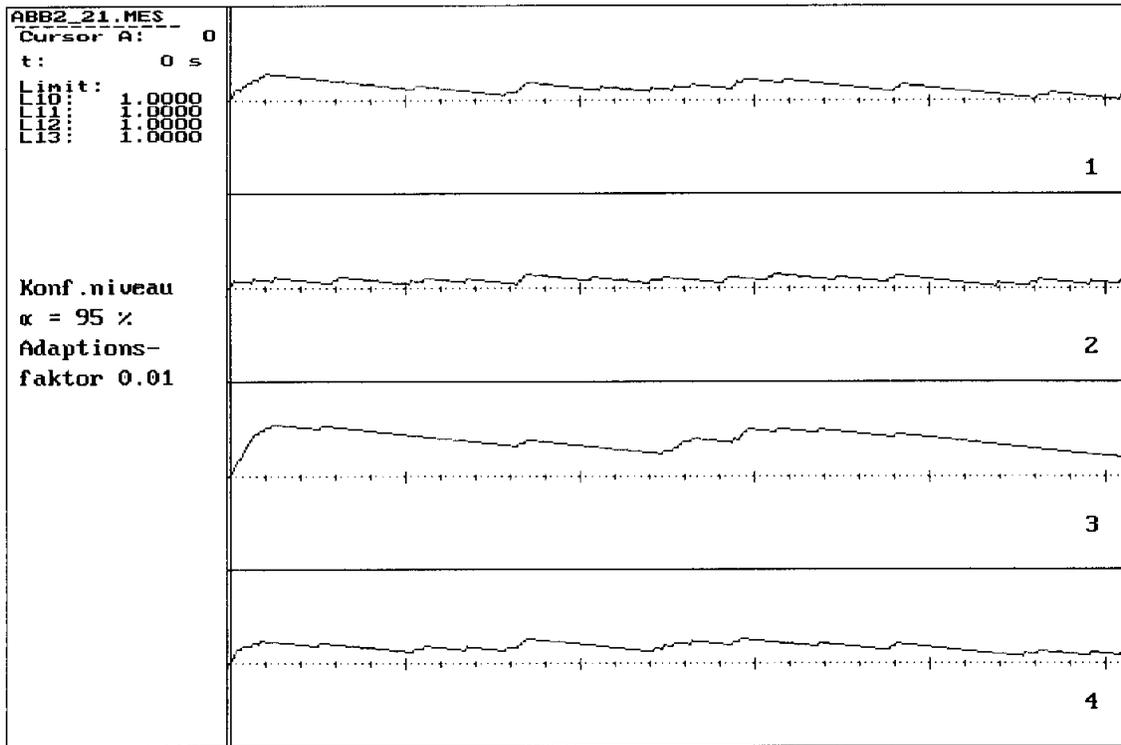


Abb.2.21: Breite adaptiver Interquantilintervalle um die maximale Parameterschwankung für verschiedene Werte der Faktoren  $c$  und  $d$ . 1:  $c = 0.01$ ,  $d = 0.01$ ; 2:  $c = 0.04$ ,  $d = 0.01$ ; 3:  $c = 0.01$ ,  $d = 0.025$ ; 4:  $c = 0.04$ ,  $d = 0.025$

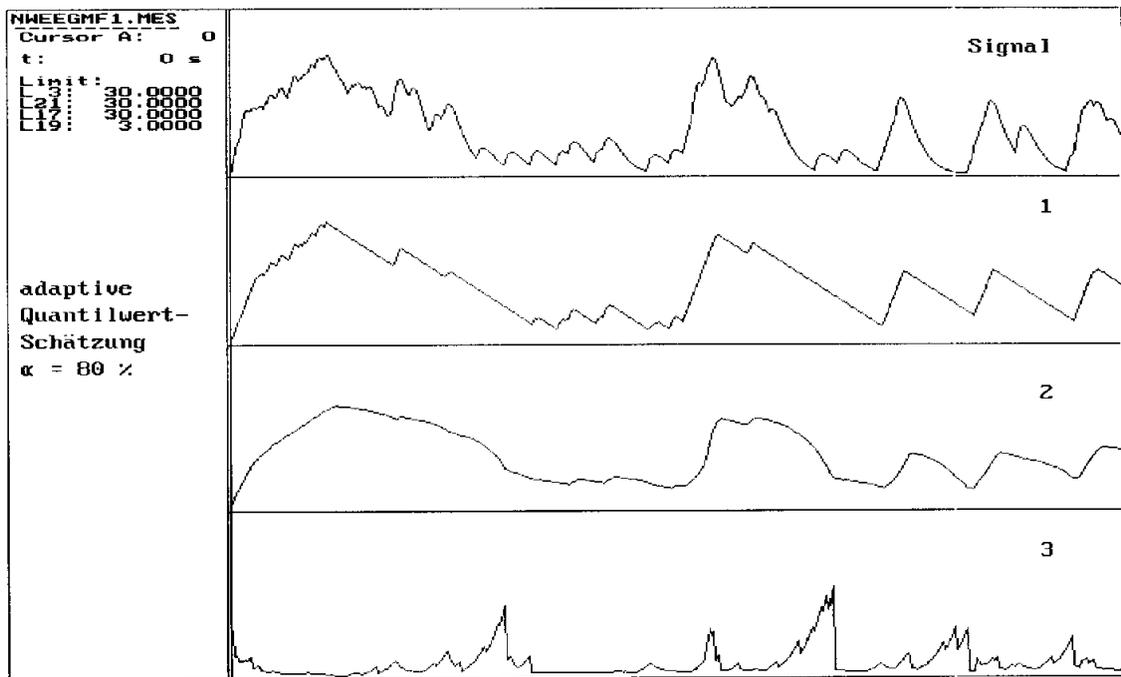
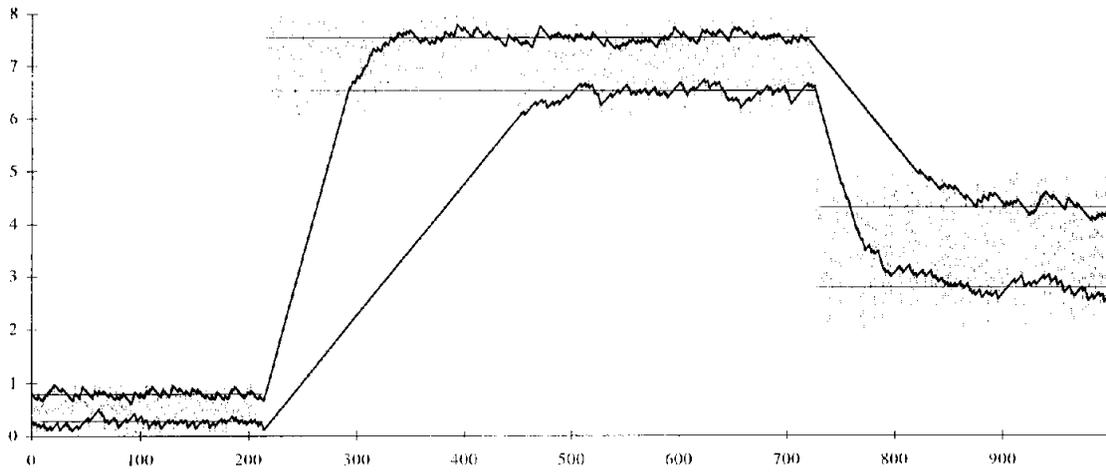
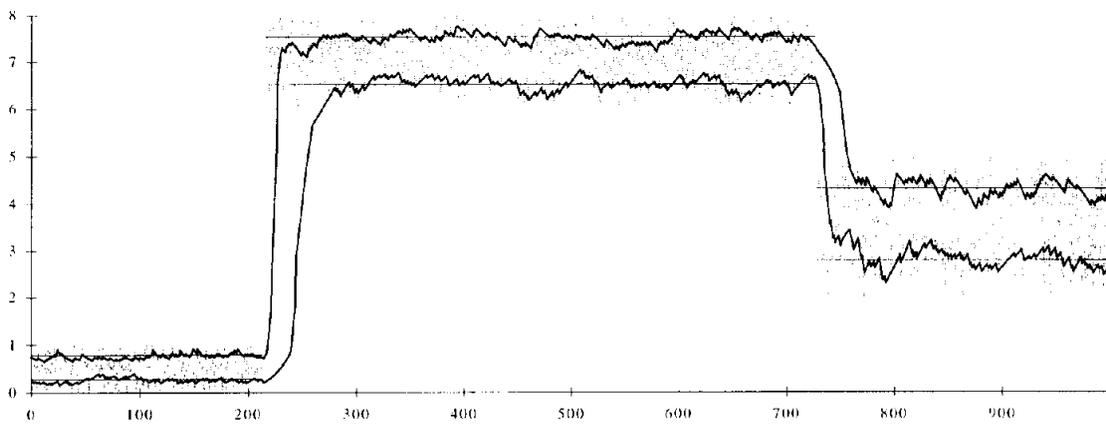


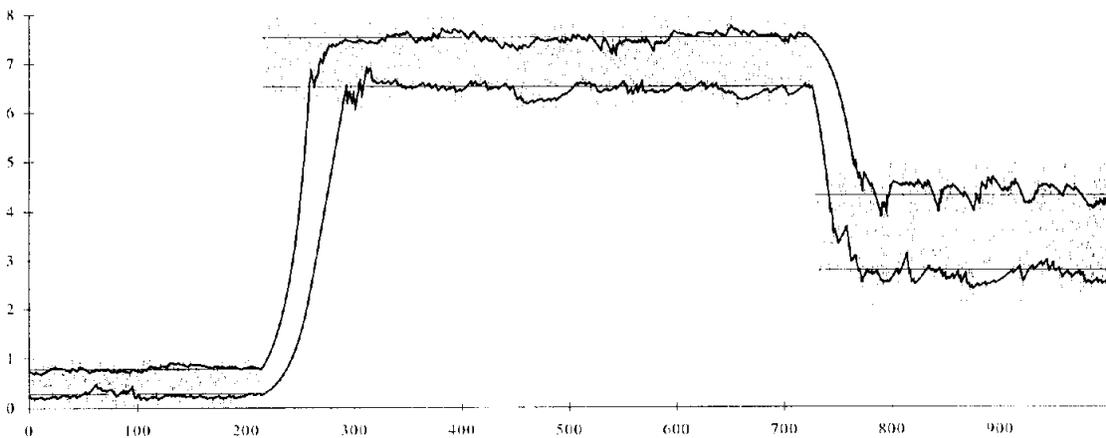
Abb.2.22: Adaptive Quantilwertschätzung,  $\alpha = 80\%$ . Signal: Verlauf der momentanen Frequenz eines EEG-Signalabschnittes. 1: Schätzung mit konstantem Adaptionsfaktor  $c = 0.01$ , 2: Schätzung mit der Steuerfolge  $c_i = d/H^c(q_\alpha(t))$ ,  $d = 0.01$ , Histogrammklassen in  $[0,20]$ ,  $h = 2$ ,  $c = 0.05$ , 3: Verlauf der zugehörigen Steuerfolge  $c_i$



a) Adaptive Schätzung mit konstanter Steuerfolge mit  $c = 0.1$

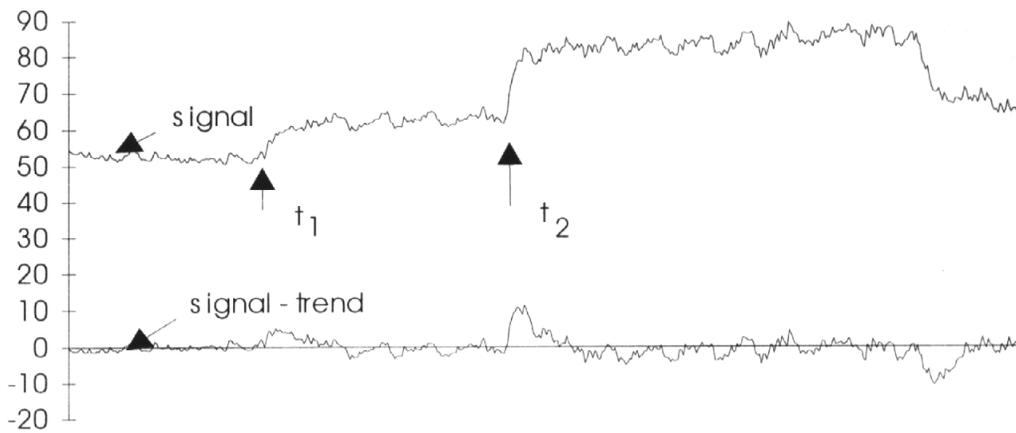


b) Adaptive Schätzung mit der Steuerfolge nach (2.71) mit dem Adaptionfaktor  $d = 0.05$  und der Histogrammschätzung mit 10 Klassen im Intervall  $[0,8]$  und dem Adaptionfaktor  $c = 0.05$ .

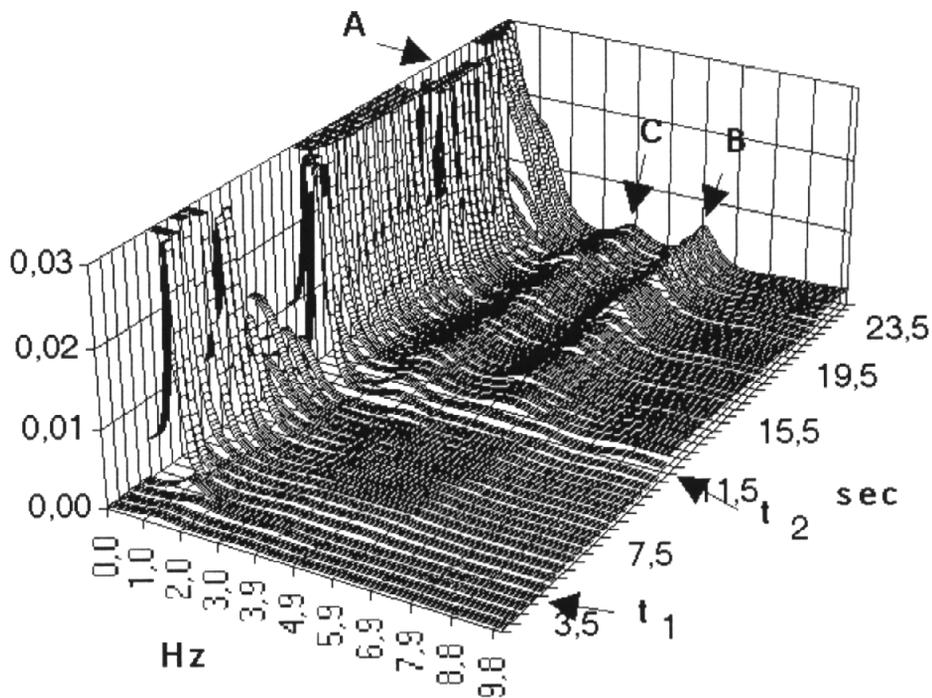


c) Adaptive Schätzung nach (2.79)-(2.80) mit  $c = d = 0.05$ ,  $h = 0.1$  und  $\epsilon = 50$

Abb.2.23: Vergleich von adaptiven Schätzungen der Quantilwerte am Beispiel des 50%-Konfidenzintervalls für eine simulierte Zeitreihe von unabhängigen gleichverteilten Pseudozufallszahlen



a) Laser-Doppler-Flowmetrie-Signal vor und nach der Trendeliminierung



b) Dynamische Spektralanalyse während einer schrittweisen Veränderung der Mikro-  
zirkulation: A - Atmungsrythmus, B - rhythmische Komponente des Blutdruckes  
bedingt durch den Herzschlag, C - 2. Harmonische des Atmungsrythmus

Abb.3.3: Mikro-  
zirkulation auf der Kortexoberfläche fortlaufend registriert (Tierexperiment).  
Durch Abklemmen der linken und rechten Arterie carotis communis konnte die  
Beeinflussung der Mikro-  
zirkulation vorgenommen werden. Die Wiederherstellung  
normaler Bedingungen wurde nach Öffnen der linken ( $t_1$ ) und der rechten ( $t_2$ ) Arterie  
mit Hilfe der dynamischen Spektralanalyse untersucht.

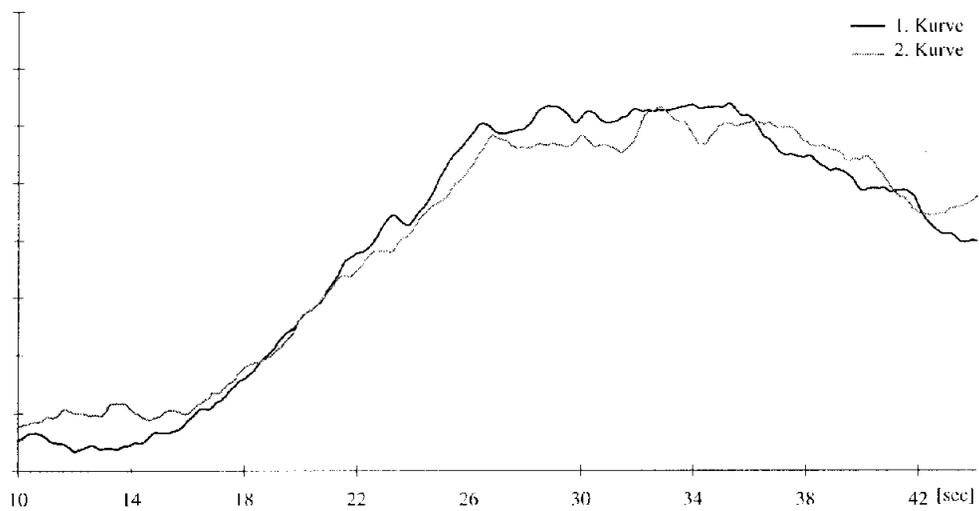


Abb.3.5: Beispiel einer Messung von Indikatorlösungskurven (Ausschnitt 10.- 44. sec)

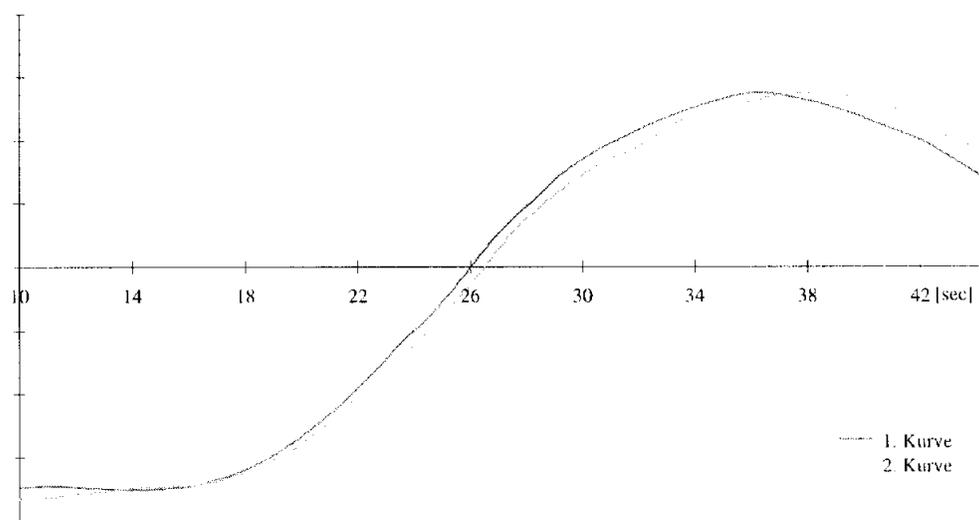
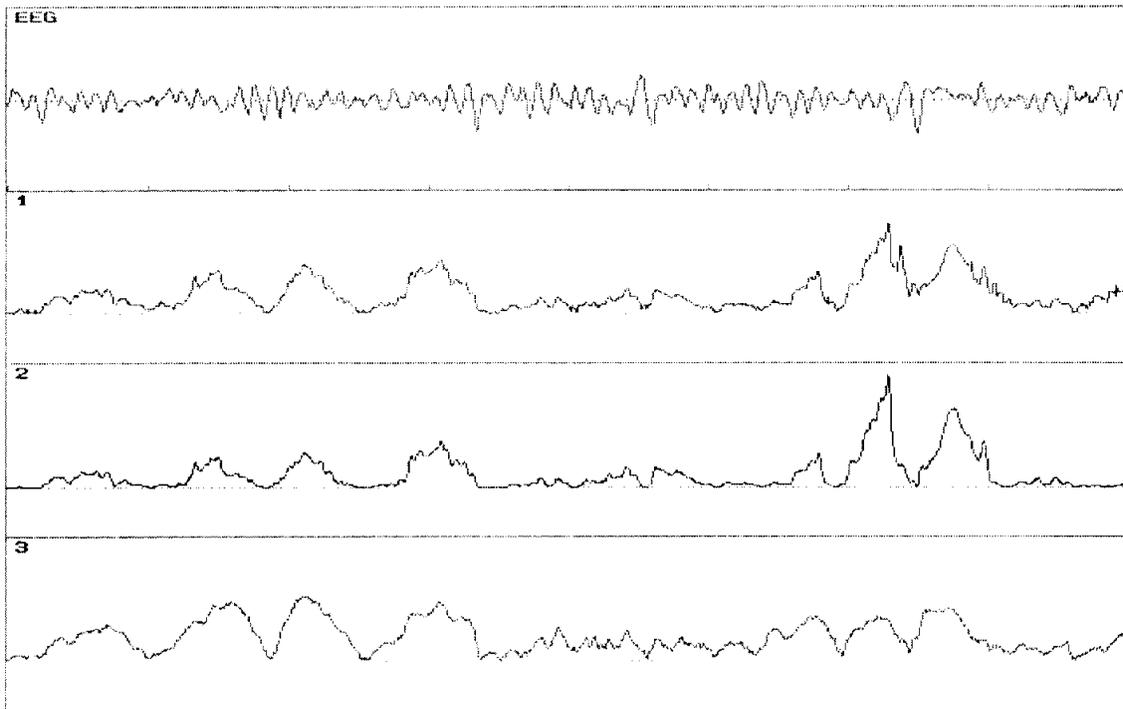
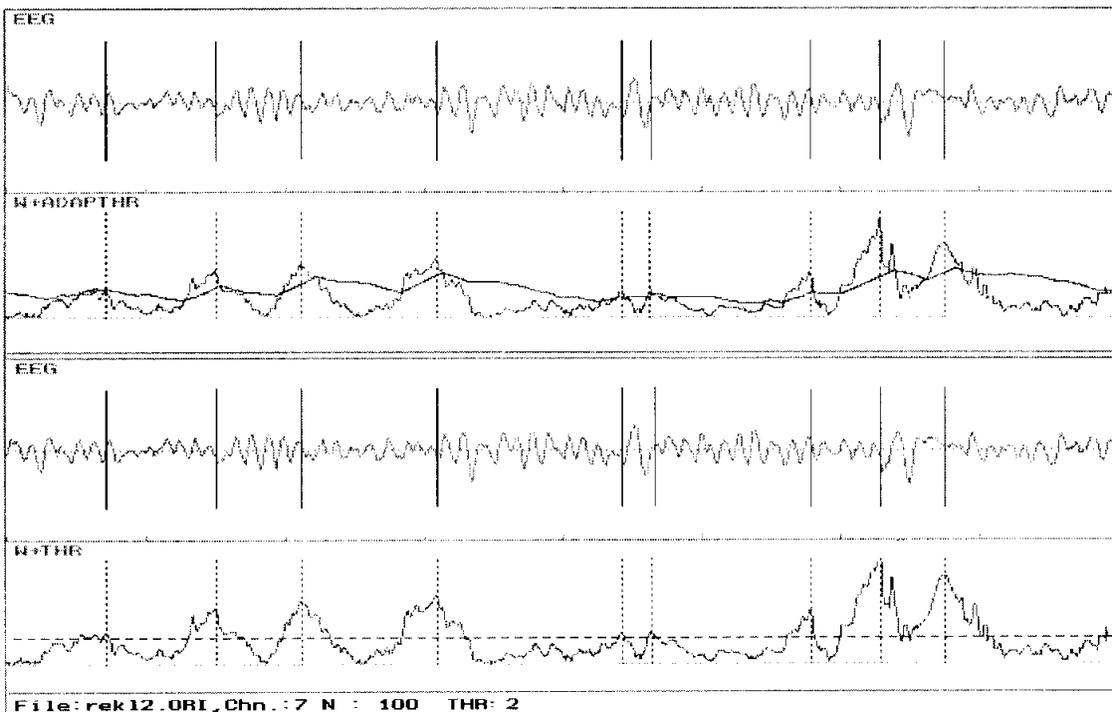


Abb.3.6: Indikatorlösungskurven nach der Vorverarbeitung (Filterung, Zentrieren, Normieren)



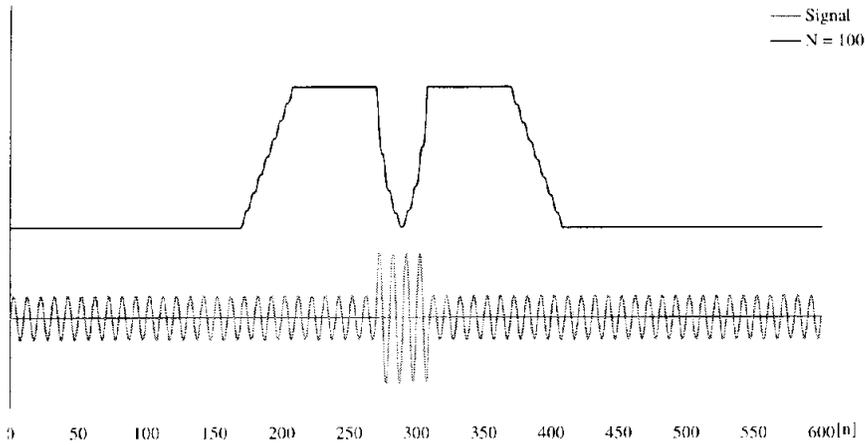
File: rek12.ORI.Chn.:7 N: 100

Abb.3.7: Verlauf der Funktion  $W$  für verschiedene Segmentationsverfahren am Beispiel eines EEG-Abschnitts über 8 sec (1600 Meßwerte), Länge der Vergleichsabschnitte  $N = 100$ . 1: Schätzung nach (3.34); 2: Schätzung nach (3.27) ( $K = 4$ ); 3: Schätzung nach (3.29)

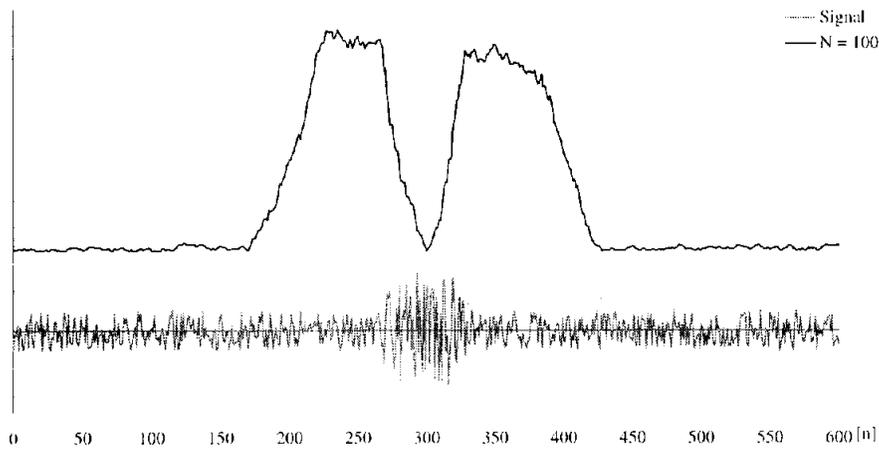


File: rek12.ORI.Chn.:7 N: 100 THR: 2

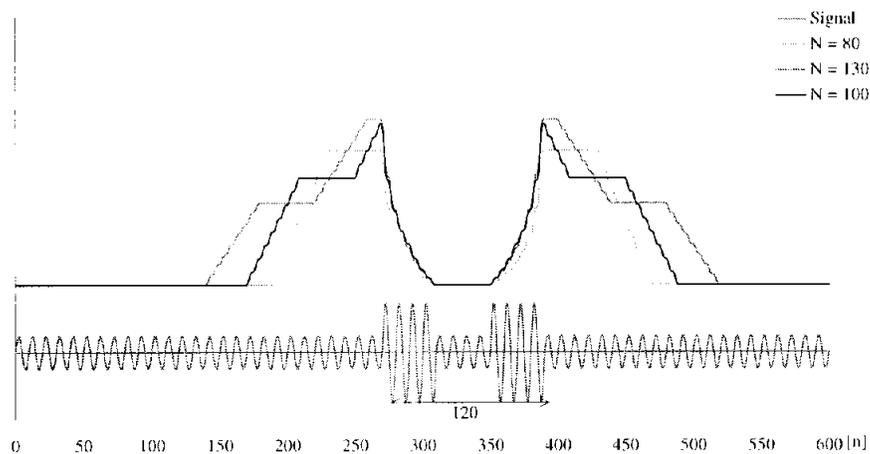
Abb.3.8: Adaptive Segmentation eines EEG-Signalabschnittes über 8 sec mit Hilfe der Funktion  $W$  aus (3.34) bei der Anwendung einer adaptiven Schwelle (Bild oben) und einer konstanten Schwelle  $\kappa = 2$  (Bild unten). Die Parameter der adaptiven Schwelle sind  $\alpha = 70\%$ ,  $\beta = 40\%$ ,  $\gamma = 1$ ,  $K_0 = 1$ . Für beide Quantilwertfolgen wurden die gleichen Parameter gewählt: Adaptionsfaktor  $d = 0,005$ , Histogrammklassen im Intervall  $[1,10]$ ,  $h = 3$ , Adaptionsfaktor  $c = 0,025$ .



a) Signal:  $x_n = 0.5\sin(2\pi n/32)$ ,  $n = 0, \dots, 270, 311, \dots, 600$ ;  $x_n = 1.5\sin(2\pi n/32)$ ,  $n = 271, \dots, 310$

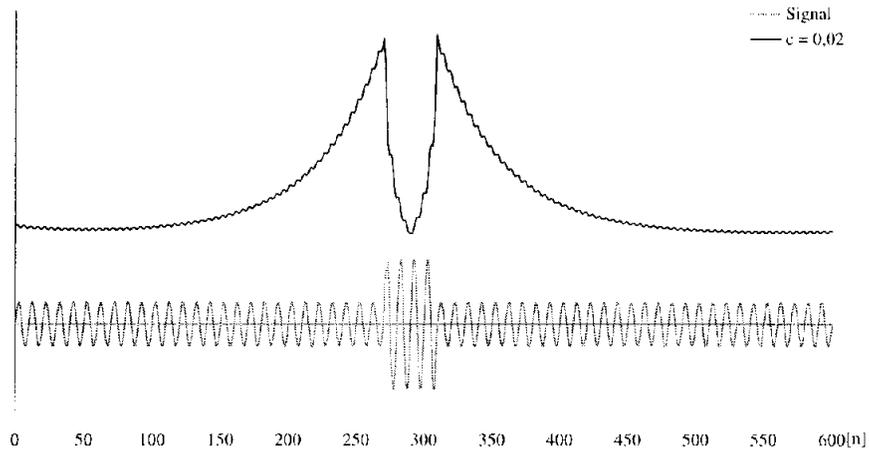


b) Signal:  $\{x_n\}$  ist eine Folge unabhängiger gleichverteilter Zufallszahlen. Für  $n = 0, \dots, 270, 331, \dots, 700$  liegen die Zahlen im Intervall  $[-0.5, 0.5]$ , für  $n = 271, \dots, 330$  im Intervall  $[-1.5, 1.5]$

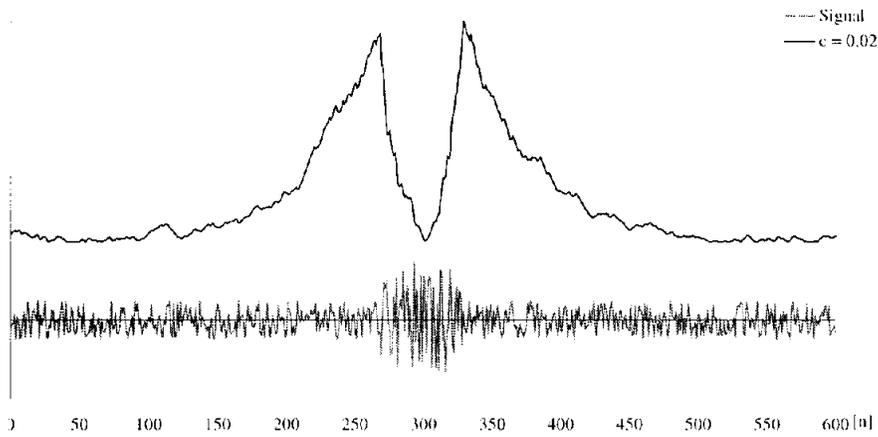


c) Signal:  $x_n = 0.5\sin(2\pi n/32)$ ,  $n = 0, \dots, 270, 311, \dots, 350, 391, \dots, 600$ ;  
 $x_n = 1.5\sin(2\pi n/32)$ ,  $n = 271, \dots, 310, 351, \dots, 390$

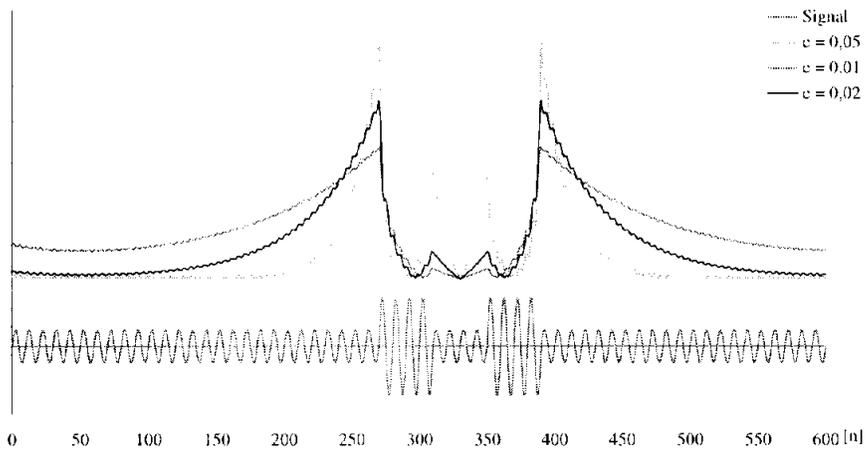
Abb.3.9: Verlauf der Funktion  $V_E$  nach (3.32) für simulierte Zeitreihen, in denen die Amplitude in kurzen Abschnitten auf das Dreifache erhöht wurde



a) Signal:  $x_n = 0.5\sin(2\pi n/32)$ ,  $n = 0, \dots, 270, 311, \dots, 600$ ;  $x_n = 1.5\sin(2\pi n/32)$ ,  $n = 271, \dots, 310$



b) Signal:  $\{x_n\}$  ist eine Folge unabhängiger gleichverteilter Zufallszahlen. Für  $n = 0, \dots, 270, 311, \dots, 600$  liegen die Zahlen im Intervall  $[-0.5, 0.5]$ , für  $n = 271, \dots, 310$  im Intervall  $[-1.5, 1.5]$



c) Signal:  $x_n = 0.5\sin(2\pi\omega n)$ ,  $n = 0, \dots, 270, 311, \dots, 350, 391, \dots, 600$ ;  
 $x_n = 1.5\sin(2\pi\omega n)$ ,  $n = 271, \dots, 310, 351, \dots, 390$

Abb.3.10: Verlauf der Funktion  $W$  nach (3.50) für simulierte Zeitreihen, in denen die Amplitude in kurzen Abschnitten auf das Dreifache erhöht wurde

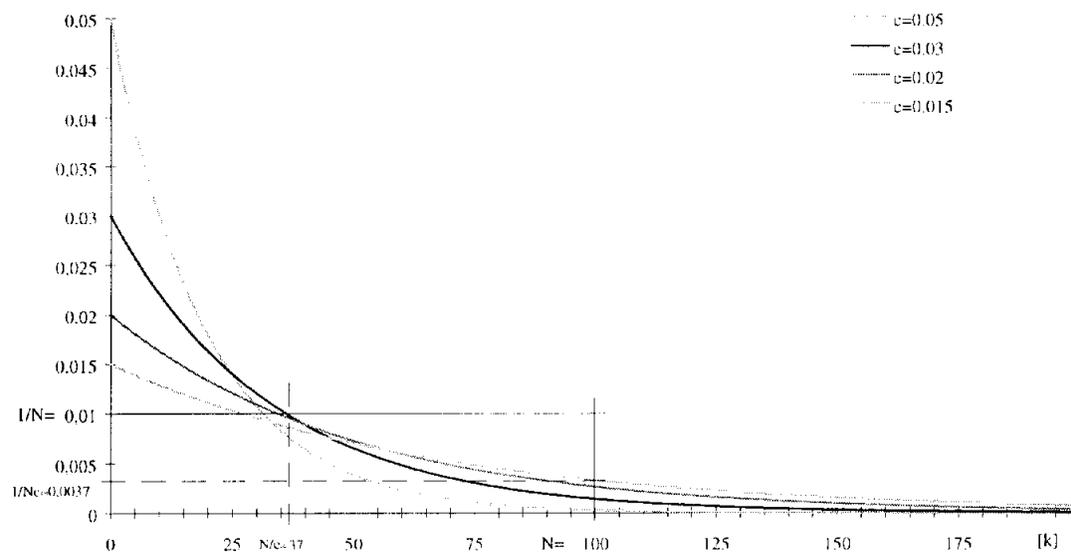
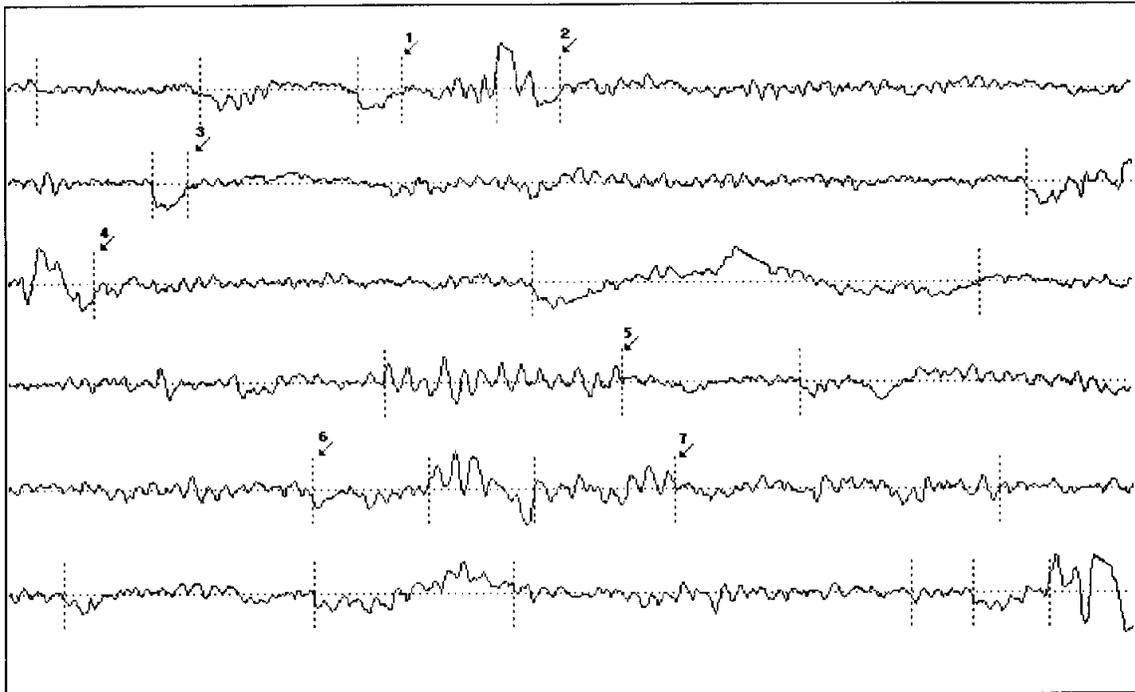
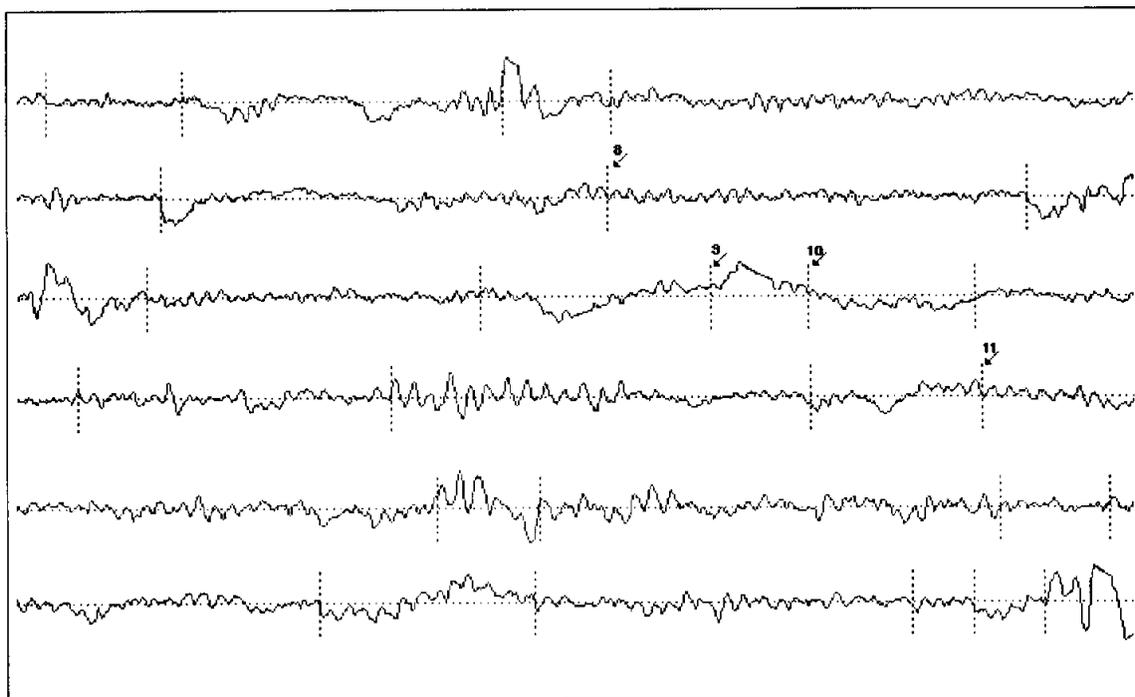


Abb.3.11: Die Gewichte bei der adaptiven Schätzung der momentanen Leistung und der momentanen Frequenz für verschiedene Werte des Adaptionfaktors  $c$  im Vergleich zu Gewichten bei Berechnung der mittleren Parameter in einem Intervall der Länge  $N = 100$

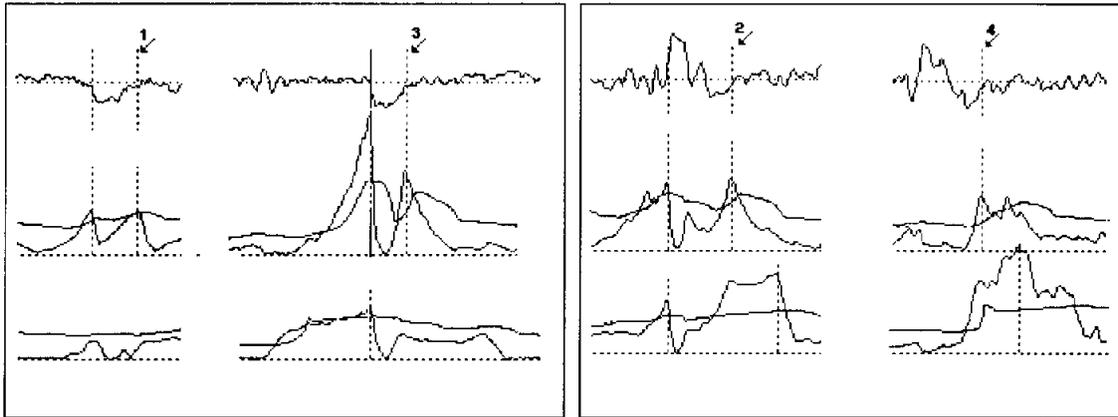


a) Segmentation nach (3.46)-(3.49), adaptive Schätzung der Parameter mit Faktor  $c = 0,03$



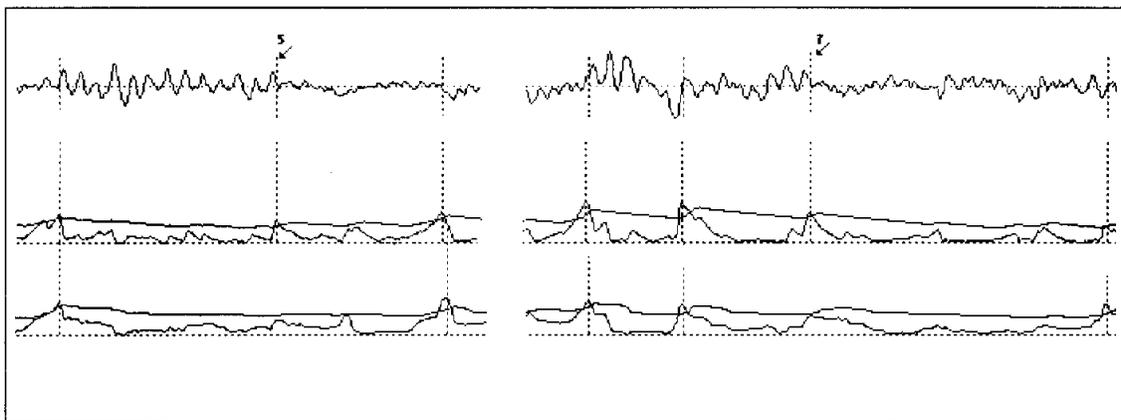
b) Segmentation nach (3.30)-(3.33), Schätzung der Parameter über Abschnitte der Länge  $N = 100$

Abb.3.12: Adaptive Segmentation von einem EEG-Signal über 48 sec. (1 Kanal), 8 sec. (1024 Meßpunkte) pro Zeile. Parameter für adaptive Schwelle in beiden Fällen:  $\alpha = 70\%$ ,  $\beta = 40\%$ ,  $\gamma = 1$ ,  $K_0 = 1$ . Für beide Quantilwertfolgen wurden die gleichen Parameter gewählt: Adaptionfaktor  $d = 0,02$ , Histogrammklassen im Intervall  $[1,10]$ ,  $h = 3$ , Adaptionfaktor  $c = 0,1$  (Im Vergleich zur Abbildung 3.10 wurden die Adaptionfaktoren vergrößert, da die Signale mit einer kleineren Abtastrate aufgenommen wurden)

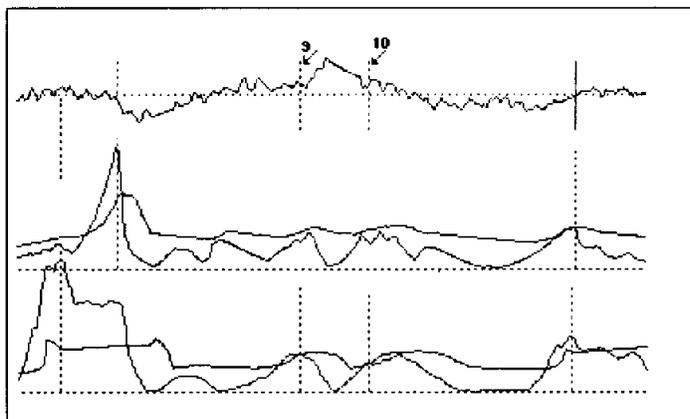


c)

d)



e)



f)

Abb.3.12c-f: Verlauf der Funktion  $W$  für ausgewählte Abschnitte aus dem vorangegangenen Signal: 1. Zeile: Signalabschnitt  
 2. Zeile: Verlauf von  $W$  für Segmentation mit adaptiver Schätzung der Parameter  
 3. Zeile: Verlauf von  $W$  für Segmentation, in der die Parameter über feste Abschnitte geschätzt wurden

## Selbständigkeitserklärung

Ich erkläre, daß ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Hilfsmittel und Literatur angefertigt habe.

Jena 22. 3. 1996

## LEBENS LAUF

Möller, Eva geb. Barešová

20. 11. 1965 Košice / Slowakei  
verheiratet seit 22. 5. 1996
- 1972 - 1980 Grundschule in Košice/Slowakei
- 1980 - 1983 Gymnasium in Košice/Slowakei
- 1983 - 1984 Gymnasium in Banská Štiavnica/Slowakei
- 1984 Abitur
- 1984 - 1989 Studium, Mathematik,  
Friedrich-Schiller-Universität Jena
20. 7. 1989 Diplom, Dipl.-Math.  
Friedrich-Schiller-Universität Jena
- 1989 - 1993 interne Aspirantur an der Mathematischen Fakultät  
der Friedrich-Schiller-Universität Jena
- seit 1993 wissenschaftlicher Mitarbeiter im Institut für  
Medizinische Statistik, Informatik und Dokumentation  
Friedrich-Schiller-Universität Jena
- Jena, 3. 3. 1997