Kaipio, Jari. Simulation and Estimation of Nonstationary EEG. Kuopio University Publications C. Natural and Environmental Sciences 40. 1996. 110 p.
ISBN 951-780-538-1
ISSN 1235-0486

ABSTRACT

Electroencephalograms (EEG) are electric signals that originate in brain and are measured from scalp. EEG is widely used in the assessment of certain neurophysiological states and disorders. Although it is well known that the statistical properties of EEG are not time-independent (stationary), most methods that are used in EEG analysis rely at least implicitly on that the statistical properties of EEG do not depend on time.

One of the most difficult topics in the practical analysis of nonstationary processes such as EEG is that it is difficult to evaluate the performance of estimation methods since there are usually no known references. One is then usually obliged to resort to simulations and Monte Carlo analyses.

In this thesis a systematic method for the generation of realistic simulations for nonstationary background EEG was developed. The method is based on the realization of EEG as a time-varying autoregressive (AR) process and is capable of realizing various types of variabilities of EEG. As an example, the applicability of the method to the simulation of a two state EEG was then verified.

In this thesis a new method for time-varying AR modelling was also developed. The method is called the modified time-varying autoregressive least squares (MTVARLS) method. The method is based on the existence of a hypothetical model for the variability of the phenomenon that is to be estimated. This model is expressed as an ensemble of optimal predictor or AR parameter evolutions and the corresponding prediction errors. The model is capable of taking into account correlations between different kinds of features in the expectable predictor evolutions. This method was then applied to the estimation of event related synchronization changes. It was found that the method was able to track for example small delays in the synchronization.

The case in which the hypothetical ensemble can not be formed was also studied. This is called blind estimation of optimal MTVARLS subspaces. In this case adaptive predictors were used to obtain evolution estimates for the optimal predictors. It was found that in most practical cases the required number of observations from the process is too high to enable the blind estimation scheme.

AMS (MOS) Classification: 93E10, 62M10
National Library of Medicine Classification: WL 150, QT 36
INSPEC Thesaurus: electroencephalography; time-varying systems; modelling; simulation; identification

To Paula and Maija

Acknowledgements

The explicit part of this study was carried out in the Department of Applied Physics at the University of Kuopio during 1993–1995.

I wish to express my gratitude to my supervisors Professor Lauri Patomäki, Ph.D., and Ari Pääkkönen, Ph.D., for their support and encouragement. Especially I would like to thank Professor Patomäki for the opportunity to work in the Department of Applied Physics and for the confidence he has shown in me at my various duties in the department. He has also made it possible that a research group specialized in biomedical modelling has evolved in the department.

I wish to thank the official reviewers Docent Pekka Loula, Ph.D., and Jouko Lampinen, Ph.D., for their constructive criticism and suggestions. Thanks to Marko Juntunen, M.Sc., Pasi Karjalainen, M.Sc., Harri Saarnisaari, M.Sc., and Marko Vauhkonen, M.Sc., who also read the manuscript and made valuable suggestions are also due. Especially I would like to thank Mr. Karjalainen for the long-standing comradeship and co-operation in all research fields of our mutual interest. He has also contributed to this thesis by participating in the programming and simulations in Chapters 4 and 5.

I wish also to thank the staffs of the Department of Applied Physics, the Department of Computer Science and Applied Mathematics at University of Kuopio and Puumalainen Research Laboratories Ltd. with whom I have had the privilege to work with. Thanks to The Department of Clinical Neurophysiology, University of Kuopio and the department head Docent Juhani Partanen, M.D., Ph.D., for co-operation and support to the group are also due.

I want also to thank all my students. I have learnt a lot when I have explained various matters to them. I would also like to thank all other researchers with whom I have had the privilege to work with in projects that are not directly connected with this thesis. Especially I wish to express my gratitude to Associate Professor Erkki Somersalo, Ph.D, with whose co-operation the research of our group has gained much depth, and Docent Jouko Tervo, Ph.D., whose efforts in both teaching and research are greatly appreciated.

My love and appreciation go to my wife Tuula, my daughters Paula and Maija and to my parents Liisa and Pekka Kaipio. I am also very happy to have so many both academic and non-academic friends that I do not have the smallest possibility to thank them here. Without all these people I would possibly have (sooner) ended in a box with soft walls and a shirt with one sleeve.

Abbreviations

AIC	Akaike's information criterion
AR	Autoregressive
ARMA	Autoregressive moving average
CSA	Compressed spectral array
EEG	Electroencephalogram, electroencephalography
EMG	Electromyelogram, electromyelography
EOG	Electro-oculogram, electro-oculography
ERD	Event related desynchronization
ERS	Event related synchronization
FFT	Fast Fourier transform
FPE	Final prediction error
KF	Kalman filter
LMS	Least mean square
LP	Low pass
LS	Least squares
MA	Moving average
MAP	Maximum a posteriori
MDL	Minimum description length
MTVARLS	Modified time-varying autoregressive least squares
PCA	Principal component analysis
REM	Rapid eye movement
RLS	Recursive least squares
SP	Smoothness priors
SPA	Spectral parameter analysis
STD	Standard deviation
STFT	Short-time Fourier transform
TV	Time-varying
TVAR	Time-varying autoregressive
TVARLS	Time-varying autoregressive least squares
TVARMALS	Time-varying autoregressive moving average least squares
YW	Yule-Walker

CONTENTS

CHAPTER I

Introduction

Few sources of information in the real world conform to the implicit and explicit assumptions of methods that are used to analyze them. Sometimes these assumptions are passed unnoticed and the results of the application of the methods are interpreted as if the (unknown) assumptions had been valid. For the most part this is an unfortunate side-effect of the enhanced accessibility to sophisticated software.

The correct approach to deal with the incompability between the source characteristics and assumptions of the standard methods is naturally to develop these methods to take into account the characteristics. Due to the differences in source characteristics it is evident that the better a method is tailored to suit a specific application the worse and more unpredictable its behavior will be in applications with different characteristics. This means that in most applications it is necessary to develop methods that are not just simple modifications of old ones.

In a real world setting we are hardly ever able to assess the performance of a method by just applying it to observed data (a sample). In many cases a method to be tested is compared with an old one that is assumed to give the correct answers and thus serves as a reference. This approach is agreeable for example in longitudinal investigations. The new methods are then, however, limited by the performance of the previous methods. One obvious solution is to use simulations to test the new method. It turns out that in many cases the generation of realistic simulations is very problematic and most methods are tested with such unrealistic simulations that are easy to generate. A further ingredient of a good simulation is that the optimal performance of the method should be well defined and calculable from the known characteristics of the simulation. In other words, the simulation should be such that the estimates given by the method can be compared to optimal estimates. The above requirements mean that the ability to generate the simulations requires the knowledge of a good model for the information source. It is then evident that the optimal method should be based on this model.

One of the key problems in the tailoring of methods to specific applications is how to express the prior knowledge in a mathematical formalism. This can turn

out to be a very difficult task. And even when this can be accomplished there is still the further task of encoding this information in a form that can be used by the method.

The aims and contents of the thesis

Throughout this thesis the personal pronoun "we" is used to denote the (sole) author of this thesis. The original aim of the thesis was to develop an estimation method for the event related synchronization changes that are observed in electroencephalograms (EEG's) and correlate for example with the level of visual stimulation. The method was to be able to estimate changes and trends in consecutive tests. The need for such a method was obvious since the methods that were used in the estimation of event related synchronization changes were statistically infeasible for single sample estimation.

When the performance of the method that was developed for this end was to be evaluated, it became clear that the analysis methods used with EEG have been very seldom evaluated objectively and that there were no systematic algorithms for the generation of the required realistic simulations. For this reason such a systematic method was developed in this thesis.

Both the new methods for simulation and estimation are based on time-varying autoregressive (time-varying AR, TVAR) modelling. We have, however, not used the concept TVAR in the title of the thesis since the new ideas are also applicable to other parametric simulation and estimation methods.

In Chapter 2 we discuss the aspects of the theory of stationary and nonstationary processes and time series models that are relevant to the thesis. In Chapter 3 we give a short review of selected topics of the the applications of parametric time series models to the analysis of EEG. We also discuss basic problems of the use of the models and the interpretation of the results.

The Chapters 4–6 constitute the novel part of this thesis. In Chapter 4 we discuss the simulation of EEG and present a complete system of generating realistic simulations of nonstationary EEG. The method is based on the description of the nonstationary EEG as a time-varying autoregressive process with smoothly varying parameter evolution. In Chapter 5 we propose a major modification to the time-varying autoregressive least squares (TVARLS) estimation scheme that is able to take into account prior information on the evolution of the model parameters. The encoding of the prior information is based on the formation of a set of expected evolutions. In Chapter 6 we discuss the scheme of Chapter 5 in the case in which we do not have enough information to be able to form a set of expected evolutions. As a solution, the estimation of the set of expected evolutions with the aid of adaptive predictors is proposed. This approach is called here *blind estimation* since the approximating subspace of the modified time-varying autoregressive least squares (MTVARLS) scheme is estimated from the data itself with no prior assumptions.

Since the proposed methods are somewhat diverse we discuss their extensions as well as possible further investigations at the end of the respective chapters rather than in a separate chapter.

In this thesis we concentrate on the methodological aspect of the modelling. Moreover, in this stage we are more interested in the theoretical and interpretational substance of the proposed methods than in the average performance. We feel that new methodologies are too often applied too early before adequate understanding of the methods in a general level is acquired. We feel also that the theoretical characteristics of the methods proposed in this thesis should be investigated further before the methods are applied to EEG at large. For these reasons we use an extremely small amount of real EEG data in this thesis and concentrate on the performance analysis of the methods with this data only. For the same reason we do not study the multivariate (multichannel) extensions of the proposed methods here.

CHAPTER II

Autoregressive models

In this chapter we give a short review of stationary and nonstationary autoregressive models. There are many good textbooks that discuss most of the topics discussed in this chapter. It must be said, however, that most of the misinterpretations concerning the results given by parametric modelling are due to general misunderstanding of stochastic processes. This is turn is due to the fact that most of the entry-level books exhibit the *signal processing* paradigm [185] that focuses on deterministic signals and systems instead of the *time series analysis* paradigm [200, 33] in which the true stochastic nature of the estimates is taken into account and the statistics are taken as estimates from observations (realizations). There is also an intermediate level of approach that we could describe as the *stochastic process* paradigm [187, 188] that consideres time series but takes the statistics as known.

In a review on spectral analysis of EEG [52] there is a warning concerning the interpretation of results when the statistical nature of estimation is overlooked.

We do not discuss the statistics of the estimates here. To keep the presentation short we assume that all processes are normally distributed so that certain notions such as orthogonality and independence are interchangeable. Also the minimum mean square estimators coincide with the linear minimum mean square estimators and we do not have to discuss conditional expectations. This review is also very selective and discusses different topics on very different levels.

2.1 Stationary ARMA models

Stationary autoregressive moving average ARMA(p, q) processes x_t can be expressed as

$$\begin{aligned} x_t &= \sum_{k=1}^{p} \phi_k x_{t-k} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j} \\ (1 - \Phi(z))x_t &= (1 + \Theta(z))e_t \\ \Phi(z) &= \sum_{k=1}^{p} \phi_k z^{-k} \end{aligned}$$

$$\Theta(z) \quad = \quad \sum_{j=1}^{q} \theta_j z^{-j} \ ,$$

where e_t is a stationary orthogonal process whose autocovariance is thus the impulse sequence $\gamma_e(h) = \sigma^2\delta(h)$ and z is the shift operator $z^k x_t = x_{t+k}$ for all integers k. The polynomial operators $1 - \Phi(z)$ and $1 + \Theta(z)$ define the autoregressive (AR) and moving average (MA) recursions of the description, respectively. If $q = 0$, x_t is an AR(p) process and if $p = 0$, x_t is a MA(q) process. In what follows the notions variance and power are used interchangeably.

We adopt the following terminology.

- For an ARMA(p, q) *process* x_t the above is true, that is, there are finite p and q and $\phi_k,\ k = 1, \ldots, p$ and $\theta_k,\ k = 1, \ldots, q$ such that x_t is the solution of the stochastic recursion equation when $E\{e_t e_{t+h}\} = \gamma_e(h) = \sigma_e^2\delta(h)$. In this case the innovations process e_t is also the minimum variance prediction error process.

- An ARMA(p, q) *approximation* for a process x_t is the description

$$x_t = \sum_{k=1}^{p} \phi_k x_{t-k} + e_t + \sum_{j=1}^{q} \theta_j e_{t-j} \ ,$$

  where $\Phi(z)$ and $\Theta(z)$ are determined so that the variance σ_e^2 of the prediction error process e_t is minimal in some sense. Usually the criterion of minimality is taken to be the mean square criterion which in this case means that the aim is to minimize the variance σ_e^2 If the process is not an ARMA(p, q) process, the prediction error process does not coincide with the innovations process and $E\{e_t e_{t+h}\} = \gamma_e(h) \neq \sigma_e^2\delta(h)$ although the conventional ends to the approximation, such as spectral estimation, usually require that $\gamma_e(h) \approx \sigma_e^2\delta(h)$. In [187, 188] this kind of an approximation is called mean square estimation.

- An ARMA(p, q) *estimate* for a realization (observation) $x_t, t = 1, \ldots, T$ of a stationary stochastic process is a function of the observed data that gives the estimates $\widehat{\Phi}(z)$, $\widehat{\Theta}(z)$ and $\widehat{\sigma}_e^2$. These estimates are functions of random variables and are thus themselves random variables and depend on the realization. The random variables $\widehat{e}_t, t = 1, \ldots, T$ are usually called *residuals* and some of the criteria for the validation of the model depend on how much the sample autocovariance estimate for $\widehat{e}_t$ differs from $\widehat{\sigma}_e^2\delta(h)$.

Note that in the first two cases the true (second order) statistics is known and in the third case the statistics is not known. In this chapter we assume that the parameters are constants, that is, they do not depend on the underlying probability space. The last two cases are different aspects of ARMA *modelling*.

An ARMA(p,q) process (approximation, estimate) x_t is causal if it assumes the representation

$$x_t = \sum_{k=0}^{\infty} \psi_k e_{t-k} = \Psi(z) e_t \ , \tag{2.1}$$

where $\Psi(z)$ is a stable causal operator, that is, $\sum_{k=0}^{\infty} |\psi_k| < \infty$. A necessary and sufficient condition for the causality and stationarity of x_t is that all the roots of $\Psi(z)$ have modulus less than unity. An ARMA(p,q) process (approximation, estimate) x_t is invertible if its innovations assume the representation

$$e_t = \sum_{k=0}^{\infty} \xi_k x_{t-k} = \Xi(z) x_t \ , \tag{2.2}$$

where the operator $\Xi(z)$ is a stable causal operator. A necessary and sufficient condition for the invertibility of x_t is that all the roots of $\Xi(z)$ have modulus less than unity. The forms (2.1) and (2.2) are called the MA(∞) and AR(∞) representations, respectively.

2.1.1 Estimation of AR models

We discuss only the AR model estimation here. Although most of the textbooks on signal and time series analysis give some methods for the modelling of ARMA processes, we refer only to [40] that is an up-to-date and comprehensive reference.

Let $\gamma(h)$ be the (true) autocovariance sequence of x_t. Define the vector $\gamma_k^m = (\gamma(k), \ldots, \gamma(m))^{\mathrm{T}}$ and the $(m+1) \times (m+1)$ symmetric Toeplitz matrix Γ_m whose first column equals γ_0^m. For the optimal (linear) mean square predictor $\widehat{x}_t = \alpha(z) x_t$ for x_t based on the variables (regressors) $x_{t-k}, k = 1, \ldots, p$ we write

$$x_t = \widehat{x}_t + e_t = \alpha(z) x_t + e_t \ .$$

The prediction error variance $\sigma_e^2 = E\left\{e_t^2\right\}$ is to attain the minimum under this predictor structure. It is well known that in this case the prediction error is orthogonal to the regressors [187] and the predictor coefficients α_k are obtained as the solution of the normal equations

$$\Gamma_{p-1} \alpha = \gamma_1^p \ , \tag{2.3}$$

where $\alpha = (\alpha_1, \ldots, \alpha_p)^{\mathrm{T}}$. The prediction error variance is then

$$\sigma_e^2 = \gamma(0) - \alpha^{\mathrm{T}} \gamma_1^p \ .$$

In the Hilbert space setting of random variables we say that $\widehat{x}_t$ is the orthogonal projection of x_t onto the subspace spanned by the processes $x_{t-k}, k = 1, \ldots, p$. For a general account on the estimation and optimization in the Hilbert space setting, see [160, 217].

This is the optimal predictor when the covariance is known. If x_t is an AR(p) process, we have $\alpha_k = \phi_k, k = 1, \ldots, p$ and $\alpha_k = 0, k > p$, that is, the optimal

predictor coefficients coincide with the AR(p) coefficients. Thus the estimation of optimal p–order predictor is equivalent to the estimation of an AR(p) model. From now on we will denote the predictor estimate with Φ whether or not it is assumed to contain the parameters of the AR model or predictor coefficient only.

The Yule-Walker estimate for the predictor is obtained by replacing the covariances in (2.3) by the biased (triangle windowed) sample autocovariance estimates $\widehat{\Gamma}^{\mathrm{YW}}$ and $\widehat{\gamma}^{\mathrm{YW}}$. The least squares (LS) estimate is obtained by writing the prediction equations

$$X_t = H\Phi + r \ ,$$

where $X_{t-k} = (x_{p+1-k}, \dots x_{T-k})^{\mathrm{T}}$, $r = (r_{p+1}, \dots, r_T)^{\mathrm{T}}$ is the residual and the regressor matrix

$$H = \begin{pmatrix} x_p & \cdots & x_1 \\ \vdots & & \vdots \\ x_{T-1} & \cdots & x_{T-p} \end{pmatrix}$$

and determining Φ so that the square of the norm of the residual $\|r\|^2 = r^{\mathrm{T}} r$ attains its minimum. The prediction equations can also be written in the form

$$X_t = \sum_{k=1}^{p} H_k \phi_k + r = \sum_{k=1}^{p} X_{t-k} \phi_k + r \ ,$$

where $H_k = X_{t-k}$ are the columns of H. The solution is obtained when X_t is projected (orthogonally) onto the subspace spanned by $X_{t-k},\ k = 1, \dots, p$. Note that these are observations of the processes $x_{t-k}, k = 1, \dots, p$. The residual is $r = X_t - H\Phi$ and

$$\begin{aligned} H^{\mathrm{T}}(X_t - H\Phi) &= 0 \\ H^{\mathrm{T}} H \Phi &= H^{\mathrm{T}} X_t \ , \end{aligned}$$

the formal solution to which is

$$\Phi = (H^{\mathrm{T}} H)^{-1} H^{\mathrm{T}} X_t$$

although it is advisable to compute the solution with an orthogonalization scheme.

The Yule-Walker and LS estimates are thus both of the form of (2.3) and differ only with respect to the covariance estimates and we can write

$$\begin{aligned} \widehat{\Gamma}^{\mathrm{YW}}_{p-1} \Phi_{\mathrm{YW}} &= (\widehat{\gamma}^{\mathrm{YW}})_1^p \\ \widehat{\Gamma}^{\mathrm{LS}}_{p-1} \Phi_{\mathrm{LS}} &= (\widehat{\gamma}^{\mathrm{LS}})_1^p \ , \end{aligned}$$

where the LS covariance estimates are $\widehat{\Gamma}^{\mathrm{LS}}_{p-1} = (T-p)^{-1} H^{\mathrm{T}} H$ and $(\widehat{\gamma}^{\mathrm{LS}})_1^p = (T-p)^{-1} H^{\mathrm{T}} X_t$. These are consistent unbiased estimates but although $\widehat{\Gamma}^{\mathrm{LS}}_{p-1}$ is non-negative definite, it is not a symmetric Toeplitz matrix and thus is not structurally the covariance of any stationary process. The LS estimate is sometimes called the covariance estimate for the AR parameters [164].

Since the forward and backward predictor coefficients coincide for stationary processes, the prediction equations can be augmented with the backward prediction equations. The resulting estimate is called the forward-backward or the modified covariance estimate [164].

2.1.2 Spectral estimation

A very large class of stationary processes can be expressed in the form

$$\begin{aligned} x_t &= \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} e^{it\omega}\, d\zeta(\omega) \\ F(\omega) &= E\left\{|d\zeta(\omega)|^2\right\} \ , \end{aligned}$$

where $d\zeta(\omega)$ is a process defined on the continuous index set $\omega \in (-\pi, \pi)$ and whose increments $(d\zeta(\omega_k+\delta\omega)-d\zeta(\omega_k))$ and $(d\zeta(\omega_j+\delta\omega)-d\zeta(\omega_j))$ are orthogonal for all $\omega_k \neq \omega_j$. If the spectral distribution function $F(\omega), \omega \in (-\pi, \pi)$ of x_t is continuously differentiable, we can write $dF(\omega) = f(\omega)\, d\omega$, where $f(\omega)$ is the spectral density, or spectrum, for short. It can be shown that

$$\gamma(h) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ih\omega} f(\omega)\, d\omega \tag{2.4}$$

$$f(\omega) = \sum_{k=-\infty}^{\infty} \gamma(k) e^{-i\omega k} \ . \tag{2.5}$$

See [33] for the detailed formulation and results concerning the existence. The autocovariance sequence is a non-negative definite sequence and the spectrum is therefore a non-negative function.

Let $x_t = H(z)v_t$, where v_t is stationary and $H(z)$ is a stable causal invertible operator (filter). The spectrum $f_x(\omega)$ of a x_t is then

$$f_x(\omega) = |H(\exp(i\omega))|^2 f_v(\omega) \ ,$$

where $f_v(\omega)$ is the spectrum of v_t. An AR(p) process can be written in the form

$$\begin{aligned} x_t &= H(z)e_t \\ H(z) &= (1 - \Phi(z))^{-1} \end{aligned}$$

and we have

$$\begin{aligned} f_x(\omega) &= \frac{f_e(\omega)}{\left|1 - \Phi(\exp(i\omega))\right|^2} \\ &= \frac{\sigma_e^2}{\left|1 - \Phi(\exp(i\omega))\right|^2} \ , \end{aligned}$$

since $\gamma_e(h) = \sigma_e^2 \delta(h)$.

2.1.3 Approximate spectral factorization

A useful property of the ARMA(p,q) models is that they can be approximated by sums of AR(1) and AR(2) processes and a finite order MA(q) process. The parameters involved in the factorization are often called spectral parameters (Spectral Parameter Analysis, SPA).

Since the practical computation of the approximate spectral factorization is not readily found in literature we discuss it here in some detail in the case of AR models.

For a spectral density function $f(\omega)$ define the function $S(z),\ z \in \mathbb{C}$ so that $f(\omega) = S(z)$ for $z = \exp(i\omega)$. Also the function $S(z)$ is usually referred to as the spectrum. For a real-valued AR(p) process $(1-\Phi(z))x_t = e_t$ we have $|1-\Phi(z)|^2 = (1-\Phi(z))(1-\Phi(z^{-1}))$ and

$$S(z) = \frac{1}{2\pi}\,\frac{\sigma_e^2}{(1-\Phi(z))(1-\Phi(z^{-1}))}\ .$$

It can be shown [236] that the covariance of an ARMA(p,q) process with $p \geq q$ can be written in the form

$$\gamma(\tau) = \sigma_e^2\left(\sum_{k=1}^{p_1}\kappa_k^1\alpha_k^{-|\tau|} + \sum_{k=1}^{p_2}|\beta_k|^{|\tau|}\left(\kappa_k^2\cos(\tau\omega_k) - \kappa_k^3\sin(|\tau|\omega_k)\right)\right)\ , \qquad (2.6)$$

where the real roots of $(1-\Phi(z))$ are $\alpha_k,\ k = 1,\ldots,p_1$ and the complex roots are $\beta_k\exp(\pm i\omega_k),\ k = 1,\ldots,p_2$ and $\kappa_k^1,\ \kappa_k^2$ and κ_k^3 are constants. The autocovariance can thus be interpreted to consist of factors corresponding to the roots of $1-\Phi(z)$. Further, we can write for the variance of x_t

$$\gamma(0) = \sum_{k=1}^{p_1+p_2}\gamma^k(0)\ ,$$

where the terms $\gamma^k(0)$ come directly from (2.6) with $\tau = 0$ and can be interpreted as variances corresponding to the factors.

In the frequency domain we can perform the corresponding factorization as

$$f(\omega) = \sum_{k=1}^{p_1+p_2}f^k(\omega) = \sum_{k=1}^{p_1+p_2}S^k(z)\big|_{z=\exp(i\omega)} = \sigma_e^2\sum_{k=1}^{p_1+p_2}\left.\frac{\Theta_t^k(z)}{\Phi_t^k(z)}\right|_{z=\exp(i\omega)}\ , \qquad (2.7)$$

where the real functions $f^k(\omega)$ are called the spectral factors.

We note, however, the following. The autocovariance factors $\gamma^k(\tau)$ are not necessarily non-negative definite sequences. This means that the process x_t can not be thought of as a sum of processes with covariances $\gamma^k(\tau)$. The non-negative definiteness of $\gamma^k(\tau)$ implies that the spectral factors $S^k(\omega)$ are not necessarily positive for all ω. However, when the roots of the model are not too near to each other, these factors make sense in that the spectral factors are almost positive and can be said to represent (approximate) either first or second order processes.

The spectral factors can be calculated using the residue theorem as follows (see [188] for a general discussion on the use of residue theorem with sequences). Calculate the roots z_k of $1 - \Phi(z)$. For the factor corresponding to a real root z_k calculate the residues ζ_{kj} , $j = 1, 2$ of $\sigma_e^2(1 - \Phi(z))(1 - \Phi(z^{-1}))z^p$ at z_k and z_k^{-1}, respectively. The spectral factor is now

$$S^k(z) = \zeta_{k1}(z - z_k)^{-1} + \zeta_{k2}(z - z_k^{-1})^{-1}$$

and $\gamma^k(0)$ can be obtained as the sum of residues of $S^k(z)z^{-1}$ at z_k and 0.

For the factor corresponding to a complex root (pair) z_k calculate the residues ζ_{kj} , $j = 1, 2$ of $\sigma_e^2(1 - \Phi(z))(1 - \Phi(z^{-1}))z^p$ at z_k and z_k^{-1}, respectively. The spectral factor is now

$$S^k(z) = \zeta_{k1}(z - z_k)^{-1} + \overline{\zeta}_{k1}(z - \overline{z}_k)^{-1} + \zeta_{k2}(z - z_k^{-1})^{-1} + \overline{\zeta}_{k2}(z - \overline{z}_k^{-1})^{-1} \ ,$$

where the overbar denotes complex conjugation and $\gamma^k(0)$ can be obtained as the sum of residues of $S^k(z)z^{-1}$ at z_k, $\overline{z}_k$ and 0.

2.2 Time-varying autoregressive models

With time-varying models we refer to such models whose parameters depend on time. The corresponding processes are thus nonstationary. As for the terminology, we note that the term nonstationary models refers sometimes to models with unit roots, *e.g.* models with seasonal characteristics. In these models the parameters are not time-varying. See [30, 66] for analysis of such processes. The notions of deterministic and stochastic regression in the estimation of time-varying parameters are probably due to Gersch [75].

2.2.1 Nonstationary prediction

Let $x_t, t = 1, \ldots, T$ be a nonstationary process with zero mean and covariance matrix Γ that is symmetric but does not exhibit Toeplitz structure. Denote the diagonal blocks of Γ by Γ_k^m with entries $E\{x_i x_j\} = \gamma(i, j), i, j = k, \ldots, k + m$. Define the vector $\gamma_k^m = (\gamma(k, k-1), \ldots, \gamma(k, k-m))^{\mathrm{T}}$. Also this vector is a block of Γ.

The time-varying minimum mean square predictor $\Phi(t; z)$ is obtained as follows. Write

$$x_t = \widehat{x}_t + e_t = \Phi(t; z)x_t + e_t$$

for $t = p + 1, \ldots, T$ separately. By employing the orthogonality principle we can write

$$E\{(x_t - \widehat{x}_t)x_{t-\ell}\} = 0$$

for $\ell = 1, \ldots, p$. The resulting systems of equations are in the matrix form

$$\begin{pmatrix} \gamma(t-1, t-1) & \cdots & \gamma(t-1, t-p) \\ \vdots & \ddots & \vdots \\ \gamma(t-p, t-1) & \cdots & \gamma(t-p, t-p) \end{pmatrix} \begin{pmatrix} \phi_1(t) \\ \vdots \\ \phi_p(t) \end{pmatrix} = \begin{pmatrix} \gamma(t, t-1) \\ \vdots \\ \gamma(t, t-p) \end{pmatrix}$$

and using the above defined notations

$$\Gamma_{t-p}^{t-1}\Phi(t) = \gamma_t^p \ .$$

The prediction error variance is then

$$\sigma_e^2(t) = \gamma(t,t) - \Phi^{\mathrm{T}}(t)\gamma_t^p \ .$$

The vector valued variable $\Phi(t) \in \mathbb{R}^p$ is called the predictor evolution. The time-varying AR(p) process with parameter evolution $\Phi(t)$ and innovations variance $\sigma_e^2(t)$ can now be defined as a process for which these coincide with the predictor evolution and prediction error variance. The innovations are orthogonal but the variance is not necessarily a constant.

There are also other predictor structures, for example we could use the variables $\{x_{t-1}, \ldots, x_{t-p}, e_{t-1}, \ldots, e_{t-q}\}$ as regressors and the predictor structure would be recursive. See [96, 3, 191] for analysis on time-varying predictors.

Available ensemble data

If an ensemble $x_t(\zeta_k), k = 1, \ldots, N$ from the process is observed, we can readily estimate the time-varying covariance matrix by

$$\widehat{\Gamma} = N^{-1} \sum_{k=1}^{N} X_k X_k^{\mathrm{T}} \ ,$$

where $X_k = (x_1(\zeta_k), \ldots, x_t(\zeta_k))^{\mathrm{T}}$. It must be noted that each predictor estimate $\widehat{\Phi}(t)$ is a p-parameter estimate from data of size N and the prediction error estimate is

$$\widehat{\sigma}_e^2(t) = \frac{N}{N-p}\left(\widehat{\gamma}(t,t) - \widehat{\Phi}^{\mathrm{T}}(t)\widehat{\gamma}_t^p\right) \ .$$

We are thus not able to construct a time-varying predictor easily from a small ensemble since we should keep $N/(N-p)$ as close to unity as possible. See [68] for a discussion on the estimation (smoothing) of a time-varying covariance estimate when the bandwidth of the process is known.

2.2.2 Deterministic regression: TVARLS scheme

The time-varying autoregressive least squares (TVARLS) scheme was first introduced in 1970 [206]. The approach was to approximate the parameter evolution by truncated Taylor expansion of order M and to solve the weighted least squares estimate of the expansion coefficients. This is equivalent to forcing the parameters to be polynomials of order M. The idea has been further developed, *e.g.* in [147, 90, 95, 77] and several sets of basis functions have been proposed.

Time-varying autoregressive models for a process (sample) $x_t, t = 1, \ldots, T$ can be written as

$$x_t = \sum_{k=1}^{p} \phi_k(t) x_{t-k} + e_t \ , \tag{2.8}$$

where e_t are the residuals and p is the order of the AR model (predictor). One of the criteria for the validity of the model is that the residuals can be modeled as zero mean independent random variables with variance $\sigma_e^2(t)$. In the conventional TVARLS scheme the time-varying prediction coefficients (the AR parameters) $\phi_k(t)$ are separately constrained to a (usually smooth) subspace $\mathcal{S} \subset \mathbb{R}^T$ with dimension M and basis $\psi_1(t), \psi_2(t), \ldots, \psi_M(t)$ where the choice $\psi_1(t) \equiv 1$ is usually made to include the stationary model as a special case. The time-varying parameters are thus of the form

$$\phi_k(t) = \sum_{m=1}^{M} c_{mk}\psi_m(t) \ . \tag{2.9}$$

Denote the p–order linear predictor of x_t by $\widehat{x}_t$ with

$$\begin{aligned} \widehat{x}_t &= \sum_{k=1}^{p} \phi_k(t) x_{t-k} \\ &= \sum_{m=1}^{M}\sum_{k=1}^{p} c_{mk}\psi_m(t)x_{t-k} \end{aligned} \tag{2.10}$$

so that $x_t = \widehat{x}_t + e_t$.

Instead of the usual approach to focus on the normal equations we will employ the regression formalism since the modification to the TVARLS scheme that we propose in this paper is more understandable in the latter formalism. To obtain the nonwindowed prediction equations we write (2.8) in matrix form for $t = p+1, \ldots, T$ and define $c = (c_{11}, \ldots, c_{1p}, \ldots, c_{M1}, \ldots, c_{Mp})^{\mathrm{T}}$, $X = (x_{p+1}, \ldots, x_T)^{\mathrm{T}}$, $E = (e_{p+1}, \ldots, e_T)^{\mathrm{T}}$ and the regressor matrix

$$H = \begin{pmatrix} \psi_1(p+1)x_p & \cdots & \psi_1(p+1)x_1 & \cdots & \psi_M(p+1)x_p & \cdots & \psi_M(p+1)x_1 \\ \vdots & & \vdots & & \vdots & & \vdots \\ \psi_1(T)x_{T-1} & \cdots & \psi_1(T)x_{T-p} & \cdots & \psi_M(T)x_{T-1} & \cdots & \psi_M(T)x_{T-p} \end{pmatrix} \tag{2.11}$$

where the (roman) superscript T denotes transposition. The prediction equations take now the form

$$X = Hc + E \ . \tag{2.12}$$

If H is nonsingular, the solution c that minimizes $E^{\mathrm{T}}E$ can be expressed in closed form as $c = (H^{\mathrm{T}}H)^{-1}H^{\mathrm{T}}X$, that is, the explicit solution of the normal equations. The estimated parameter evolution is then obtained from (2.9). In this formalism each parameter evolution is estimated separately as a linear combination of the basis functions and we have pM free coefficients in (2.12).

Several sets of functions have been used as TVARLS bases. These include at least ordinary polynomials, Legendre polynomials, sinusoidal (Fourier) bases, discrete prolate spheroidal sequences, B-splines, Walsh and Haar bases [206, 147, 77, 38, 2, 62], see also [95, 1]. With the exception of the last two items, the

bases (and their linear combinations) can be described as smooth. Efficient computational approaches for the TVARLS scheme are discussed in [147, 28]. The TVARMALS and lattice structure extensions are studied in [90]. The multivariate extension of the TVARLS scheme is given in [75]. See also [176] where an application in which affine basis is used for very short segments.

There is actually no reason whatsoever to restrict oneself to the use of any of these sets of basis functions. A suitable combination *e.g.* of polynomial and sinusoidal functions may well be appropriate in some situations.

2.2.3 Stochastic regression: Adaptive predictors

We do not give a review of adaptive algorithms here. There are good textbooks on the subject from entry level [233] through [151, 100] to a demanding level [14, 39]. We give only a short description of the structure of adaptive algorithms and discuss some general characteristics that all adaptive algorithms exhibit.

A large class of adaptive algorithms can be expressed in the form [14]

$$\theta(t) = \theta(t-1) + \alpha_t \Xi(\theta(t-1), \varphi_t) + \alpha_t^2 \varepsilon_t(\theta(t-1), \varphi_t) \ , \tag{2.13}$$

where $\theta(t)$ are the recursively estimated parameters, φ_t is a sequence of observations (regressors), α_t is a sequence of scalar gains and Ξ_t and ε_t are functions that define how the parameter estimates are to be updated given a new observation. These functions depend on the model structure and the employed criterion of optimality. The criterion of optimality may be either explicit (well defined) or implicit.

In adaptive linear modelling of time series $x_t \in \mathbb{R}$, the estimation scheme can usually (especially in the cases considered in this thesis) be written in the form

$$\begin{aligned} x_t &= \widehat{x}_t + e_t \\ \widehat{x}_t &= \varphi_t^{\mathrm{T}} \theta(t) \ . \end{aligned}$$

This applies to the so-called transversal filter structure, in which the estimate is a linear function of the regressors at each time although the parameters depend on past data in a nonlinear manner. There are also other important filter structures in which this is not the case, for example the algorithm may update the lattice parameters [64], the model polynomial roots [177] or the center frequency of a narrow bandwidth process [144] directly. We discuss only the case of linear prediction and nonrecursive predictor structure, which yields directly a time-varying AR model estimate for the process (observation) x_t. We have then

$$\begin{aligned} \varphi_t &= (x_{t-1}, \ldots, x_{t-p})^{\mathrm{T}} \\ \theta(t) &= \Phi(t) \ . \end{aligned}$$

The most commonly used transversal algorithms are the least mean square (LMS, the one-step stochastic gradient algorithm), the recursive least squares (RLS) and the Kalman filter (KF) algorithms. These can be written in the following forms.

- The LMS algorithm can be written in the form

$$\theta(t) = \theta(t-1) + \alpha_t \varphi_t e_t \ ,$$

  where α_t controls the adaptation rate of the algorithm.

- The standard forgetting factor RLS algorithm can be written the form

$$\begin{aligned} \theta(t) &= \theta(t-1) + K_t e_t \\ K_t &= \frac{\kappa^{-1} P_{t-1} \varphi_t}{1 + \kappa^{-1} \varphi_t^{\mathrm{T}} P_{t-1} \varphi_t} \\ P_t &= \kappa^{-1} P_{t-1} - \kappa^{-1} K_t \varphi_t^{\mathrm{T}} P_{t-1} \ , \end{aligned}$$

  where κ is called the forgetting factor and determines the adaptation rate of the algorithm.

- The Kalman filter corresponding to one-step predictor is of the form

$$\begin{aligned} \theta(t) &= \theta(t-1) + K_t e_t \\ K_t &= \frac{P_{t-1} \varphi_t}{\sigma_e^2(t) + \varphi_t^{\mathrm{T}} P_{t-1} \varphi_t} \\ P_t &= P_{t-1} - \frac{P_{t-1} \varphi_t \varphi_t^{\mathrm{T}} P_{t-1}}{\sigma_e^2(t) + \varphi_t^{\mathrm{T}} P_{t-1} \varphi_t} + \Gamma_v(t) \ , \end{aligned}$$

  where this structure corresponds to the random walk hypermodel for the parameter evolution (state equations) with $\theta(t+1) = \theta(t) + v_t$, where v_t is an orthogonal process with covariance $\Gamma_v(t)$ and $\sigma_e^2(t)$ is the prediction error variance. The adaptation rate is controlled by $\Gamma_v(t)$. Usually the choice $\Gamma_v(t) = \sigma_v^2 I$ is made so that the individual parameter evolutions are assumed to be independent. Note that the prediction error variance is assumed to be known. The adaptive Kalman filter that estimates these variances is given *e.g.* in [107].

We do not give a general treatment of adaptive algorithms here. We note, however, that although the Kalman filter is the maximum a posteriori estimator for the state (predictor parameters) in the case of normal distribution, the optimality is also connected with the assumed hypermodel. It can be shown that *e.g.* LMS and RLS algorithms are optimal a posteriori estimators for the respective implicitly assumed hypermodels [150]. This means that in order to choose an "optimal" adaptive estimator for the parameters, we should be able to determine an appropriate hypermodel in each case. If the assumption on the parameter evolution is that it is smooth in the sense that the second or third differences of individual parameter evolutions are small, a higher order hypermodel would be more appropriate than a first order model. For the determination of the corresponding multi-stage adaptive estimators, see [13, 14].

2.2.4 Time-varying spectral estimation

The spectrum is defined only for stationary processes. A general extension to the time-varying (nonstationary) case is not easily definable. This is due to that in a nonstationary situation many of the interpretational items will be lost. In principle there are many possible candidate definitions for the TV spectrum that balance with the trade-offs inherent in all approaches.

Of fundamental importance is the recognition of the time-frequency uncertainty principle (theorem), that makes it impossible to define a time-varying spectrum with arbitrary time and frequency resolutions [187, 188]. It is also essential to distinguish between deterministic and statistical definitions. There are definitions that work well in the deterministic setting, for example, in the determination of filters for chirp signals in radar signal processing [43]. The notion "frozen" is used to refer to the fixing of time t.

We do not attempt to discuss here all methods that are used to obtain different kinds of definitions concerning time and frequency variations in processes. It must be noted that the definitions and the performance of the estimates concerning these variations can not be easily compared. The most significant approaches that are not discussed here include short time Fourier transforms (sliding window periodograms), the time-frequency distributions, such as the Wigner-Ville distribution, and wavelets and filter banks. For short reviews of these methods in connection with biomedical signal analysis tasks, see [29, 224].

2.2.5 Evolutionary spectrum

Let x_t be nonstationary with covariance (matrix) $\Gamma(s,t)$ such that $\Gamma(t,t) < \infty$ for all t. For stationary processes $\gamma(\tau) = \Gamma(s,t)$ where $\tau = t - s$ depends only on the difference $t - s$. Thus we can write for the relation between covariance and spectral distribution function

$$\begin{aligned} \gamma(\tau) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\tau\omega} dF(\omega) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\omega t} e^{-i\omega s} dF(\omega) \ . \end{aligned}$$

In analogy with this representation we write

$$\Gamma(s,t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \psi_s^*(\omega)\psi_t(\omega) d\mu(\omega) \ ,$$

where the functions $\psi_t(\omega)$ are nonstationary kernels corresponding to $\exp(it\omega)$ of the stationary case. This representation always exists on a finite interval $t \in (1,T)$ and the kernels can be chosen to be the eigenvectors of $\Gamma(s,t)$. All kernels do not, however, carry the interpretations of time and frequency. One choice that enables (in principle) this interpretation, is the family of functions

$$\psi_t(\omega) = A_t(\omega) e^{i\vartheta(\omega)t} \ ,$$

where the (Fourier) spectrum of $A_t(\omega)$ with ω fixed, has absolute maximum at zero. We do not pursue the general theory of evolutionary spectra here. For the practical issues of estimation, selection of nonstationary kernels and the theory of oscillatory processes, see [200, 201]. For a short review of other similar approaches, see [15].

Definition based on innovations

The theoretical elaboration is probably due to Tjøstheim, see [200, 201], but the general idea has been proposed earlier at least in [25]. This approach is based on the representation of the process in terms of the innovations and the associated Green's kernel. If the process is regular (purely nondeterministic), it can be represented in the form

$$x_t = \Theta(t; z)e_t = \sum_{k=0}^{\infty} \theta_k(t)e_{t-k}$$

in accordance with the MA(∞) representation of the stationary processes. Here e_t can be taken as a stationary orthogonal process with constant variance σ_e^2 but we cannot have $\theta_0(t) \equiv 1$. Equivalently we can take e_t just to be orthogonal with variance $\sigma_e^2(t)$ and set $\theta_0(t) \equiv 1$.

Let $G(t, s)$ be the one-sided Green's function (time-varying impulse response) corresponding to the operator $\Theta(t; z)$ [170]. Then

$$x_t = \sum_{k=-\infty}^{t} G(t, t-k)e_k \ .$$

This representation is important since G is a causal operator and we can write

$$X = GE \ ,$$

where $X = (x_1, \ldots, x_T)^{\mathrm{T}}$, $E = (e_1, \ldots, e_T)^{\mathrm{T}}$ and G is a lower triangular matrix.

The covariance $\Gamma(s, t)$ is then

$$\Gamma(s, t) = \sum_{k=0}^{\infty}\sum_{j=0}^{\infty} G(t, t-k)G(s, s-j)\mathrm{cov}(e_{t-k}e_{s-j}) \ .$$

If e_t is orthogonal we obtain further for the variance of x_t

$$\mathrm{var}(x_t) = \sum_{k=0}^{\infty} G^2(t, t-k)\mathrm{var}(e_{t-k}) \ .$$

The time-varying spectrum can now be defined as (with t a fixed parameter)

$$f(t; \omega) = \frac{\sigma_e^2(t)}{2\pi}\left|1 + \sum_{k=0}^{\infty} \theta_k(t)e^{-i\omega t}\right|^2 \ .$$

It can be shown that this definition conforms to the class of evolutionary spectra in which the kernel functions are generated by the corresponding time-varying filters.

Now let $\Phi(t;z)$ be a predictor for x_t so that the prediction error e_t is an orthogonal process and equals thus the innovations of x_t. We have then

$$f(t;\omega) = \frac{\sigma_e^2(t)}{2\pi} \left| 1 - \sum_{k=0}^{\infty} \phi_k(t) e^{-i\omega t} \right|^{-2} .$$

The practical task is then to estimate the predictor and the time-varying prediction error variance $\sigma_e^2(t)$. It turns out that although this definition is both theoretically and practically interesting, there are some practical problems that do not have any general solutions. The most difficult problem arises in the case of estimation of small bandwidth processes. In this case some of the roots of $1 - \Phi(t;z)$ have near unity moduli and small changes in the estimates $\widehat{\Phi}(t;z)$ can induce dramatic changes to the spectrum $f(t;\omega)$. This causes occasionally severe instability to the spectral estimates obtained with this method. The time-varying spectral estimates of Chapter 6 have been obtained with this method and the instability is clearly visible.

We can also consider the approximate spectral factorizations for the frozen models by writing all variables of Section 2.1.3 as functions of time. All the variance-related variables inherit the problem of the instability of spectral estimates.

CHAPTER III

Autoregressive models of EEG

In this chapter we discuss general topics that are relevant to the interpretation and modelling of EEG. We give then a concise review of the applications of parametric methods to the analysis of EEG. The topics and details in this chapter are very selective and the weight is on methodological aspects. Some of the main alternatives to AR models are also mentioned but these are only examples.

In spite of some theoretical similarities between AR processes and some neural models we should never assume that the EEG process can be fully described by an AR process [235]. The most we can and in most cases should do is to investigate whether the EEG can in some cases be approximated with a parametric process. Even this can be argued to be irrelevant. The ends to the EEG analysis are very diverse and it is sometimes argued that the main point is whether an EEG model can extract *features* that are relevant to the specific application. This is mainly true but we have to be very careful so as not to evaluate the applicability of the model based on how well the results given by the model correspond to the results given by the previously used models. If these are interpreted as correct (true) references, the new model and the results given by it may be discarded since they do not give the "correct" results although the main reason for the use of the model was to test whether these results were correct in the first place.

It is possibly safest to regard the parametric models as approximations to the true (observed) EEG process. The applications can then often be divided to the analysis of the quality of the approximations and the subsequent analysis of the data itself.

There are various methods to test the applicability of the parametric models. The most important of these are perhaps those that estimate the order (orders) of the model and those that evaluate the uncorrelatedness of the associated prediction errors. In some methods these two tasks are carried out simultaneously. It must be noted, however, that these methods usually assume *e.g.* regularity of the process. This means that the interpretation of the results may not be correct in case of (partly) chaotic data.

If the model is decided to be a valid approximation to the data, the features given by the model can be used in the application. The most typical features (ends) in EEG analysis are the model parameters themselves, prediction error

(residual) variance, some features extracted from the power spectrum estimate corresponding to the model such as band power estimates or the whole spectrum estimate. In principle the model order could also be used to distinguish between different states of EEG but we have no knowledge of any studies where model order had been used explicitly.

For general as well as historical views and aspects on the automatic analysis of EEG one should see [31, 84, 7, 80, 81, 154]. It must be noted that most of what is said is dependent on methods of analysis. Most of the method-independent performance analyses are based on information theory and even then these are restricted to purely stochastic settings. One should also consult the IFCN committee guidelines from 1994 [182], in which very conservative recommendations for EEG analysis methods are given, see also [181]. For example, in the eyes-closed state, a minimum of 1 sec of artefact-free data should be available for the (center) frequency analysis of EEG. Also in other aspects the discussion of analysis methods focuses heavily on the use of periodograms.

3.1 Preliminary models of EEG

In this section we discuss the *preliminary models* of EEG, that is, whether EEG can be modeled as a stochastic or a chaotic (deterministic) process. Since this thesis concentrates on stochastic models, the stationarity of EEG is subsequently discussed.

3.1.1 Linearity of EEG

Most papers on EEG methodology assume either a statistical framework in which stationarity, distributions and independence are well defined, or a deterministic framework in which the processes are assumed to be outputs of chaotic neuronal circuitry.

As such both frameworks are clearly incorrect. With the current understanding of quantum mechanics it is impossible to think of anything purely deterministic concerning the degrees of freedom of even a small neuronal circuitry. On the other hand, the structure and dynamics of neuronal circuitry is well enough understood to rule out the possibility that EEG would truly be generated by a causal linear filter driven by independent normal innovations. Even with local chaotic generators the EEG measured over the scalp will contain processes over an extended region. In most cases it can not be assumed that all these generators are so connected that the observations (the mapping of all these signals through the volume conductor) could still be represented deterministic.

Anyway, it has been shown that in certain states of the brain the EEG can be described as *e.g.* stationary AR process of relatively low order (even with normal innovations) [195]. Equally as conclusively it has been shown that in some other states the EEG portrays such phase coherence that it is extremely unlikely that such would have been produced by a linear system with innovations of any distribution, see [195, 117].

A common error in trying to prove that EEG conforms to deterministic chaos is to compare it to a normal independent process, for such analyses see for example [59, 10]. To show why this comparison is incorrect is explained below. See [186] for a correct evaluation, that is, comparison of (sleep) EEG to a normal process with the same power spectrum. Implicitly the same comparison was made in [19], where linear and non-linear methods for EEG forecasting were compared. These results indicate that even partly chaotic EEG can often be well approximated by AR models, see the example below. Nevertheless, in [117] it was conjectured that methods based on deterministic chaos would perform better than those conforming to the statistical framework.

Some define linearity so that the prediction error should be normal. This is incorrect since we can always have a linearly generated process with independent identically distributed innovations but with arbitrary non-normal distribution. The error in the above implicit test of linearity is that if the innovations were normal and the system was nonlinear, the prediction error corresponding to a linear predictor would not be normal. See [200] for how to distinguish between these concepts, in particular for an example of a dependent orthogonal process.

The relevant question is then: How adequately can EEG be described by either model, or more practically, do the parameters pertaining to a model include enough information so as to be useful. Such a question is extremely difficult to answer since, as noted earlier, there is no '*The EEG* ' and the failure of both models in some states is obvious as such and has been demonstrated many times.

One possible solution would be to employ a detector that would estimate *e.g.* the incompatibility of a (fixed order) AR model with the EEG. We would then use *e.g.* a chaotic description until the AR model would again be compatible.

A regular process that simulates phase coherence

To show why one should be very careful when using correlation dimensions in the assessment of chaoticity of time series we present a simple method how to generate a regular stochastic process that could easily be judged as deterministic. Consider the following. Select an EEG waveform that would be regarded as deterministic. At least for some time this can be described as periodic. Let the period be T. Calculate the trigonometric Fourier expansion

$$x = \sum_{k=-N}^{N} \langle x, \phi_k \rangle \, \phi_k \ ,$$

where $\phi_k = \exp(-jk\omega_0 t)$, $t = 1, \ldots, T$, $k = -N, \ldots, N$ are the basis functions of the expansion, $x = (x_1, \ldots, x_T)$, $\langle \cdot, \cdot \rangle$ denotes inner product and where N is fixed by the sampling frequency and T. Generate the filters

$$f_{k,\epsilon}(z) = \frac{\sigma_{k,\epsilon}^{-2} \langle x, \phi_k \rangle}{1 - 2(1-\epsilon)\cos(k\omega_0) z^{-1} + (1-\epsilon)^2 z^{-2}} \ ,$$

where

$$\sigma_{k,\epsilon}^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|1 - 2(1-\epsilon)\cos(k\omega_0)\exp(-j\omega) + (1-\epsilon)^2\exp(-2j\omega)\right|^{-2} d\omega \ .$$

Then generate

$$x_t^\epsilon = \sum_{k=1}^{N} f_{k,\epsilon}(z)e_t \ ,$$

where e_t are independent identically distributed with unit variance. Then let ϵ tend to but not reach zero. From the construction it is evident that x_t^ϵ is regular for all $0 < \epsilon < 1$ and that its variance equals $\langle x, x\rangle$. The time domain appearance is not likely to resemble x but its appearance should be as "deterministic" as that of x_t.

The main point was not to try to maintain that this would be a feasible model for EEG but rather to point out that in calculating correlation dimensions the correct "noise reference" is not necessarily white noise. We refer to [201] and [200] for a discussion on the differences between linearity of time series, distribution of innovations and prediction error, and mean square prediction. See also [129] for a review of the methods to generate regular stochastic processes with prescribed (non-normal) distribution and covariance.

3.1.2 Stationarity of EEG

When we consider measured "real world" processes we can hardly ever postulate wide sense stationarity for an observation of an arbitrary time span. We are thus obliged to restrain ourselves to approximate stationarity over time spans of variable duration. The segmentation of the observations to such epochs necessitates measures of changes in the statistical description of the process and the estimates.

It is qualitatively clear that the faster the description changes over time the smaller is the amount of information based on which we have to judge whether a change has happened or not. If we use methods that rely on (implicit) assumptions on stationarity to nonstationary segments we will obtain, as a rule, unpredictable results whose interpretation is more or less impossible. This applies also to EEG. It is customary to segmentate the observation to either fixed or variable span segments that are considered as approximately stationary, or *pseudostationary*. The spans of accepted pseudostationary segments vary from a second to several minutes depending on the situation. It must be noted that the term nonstationary often refers to "disturbances" in a stationary epoch rather than to a (slowly) varying statistical description. These nonstationarities are discussed later in this chapter.

Historically, most early analyses on the stationarity of EEG were done incorrectly using statistical methods that required that the samples were independent and normal, *i.e.* the observation was a normal innovation process. For a review on early studies, see [167]. Other factors in the early studies were insufficient ensemble data, single channel experiments and few subjects.

It is clear that the notions of (amplitude) distribution and nonstationarity can not be separated. This follows from the fact that in order to be able to employ correct degrees of freedom we must know the covariance of the process. This in turn has no meaning if the ensemble is not stationary. This necessitates simultaneous estimation of (degrees of) stationarity and the deviation from a default distribution, most notably the normal distribution.

One such approach using the Kolmogorov–Smirnov test has been described in [167]. Other studies/situations on the stationarity and normality include mental arithmetics and sleep [223]. The results are usually expressed in the form 'X percent of T length epochs in state S were classified as stationary/normal'. The main result is that there is no dramatic "shoulder" in the segment length – percentage graphs that would allow one to choose a prescribed segment length of either approximate stationarity or normal characteristics.

The measures for stationarity depend heavily on the aims and requirements of the analysis methods. In [42], for example, stationarity is claimed to hold for 12 sec in most cases.

As mentioned above, every method that is used to segmentate the process has to incorporate a measure of change. These are not, however, always appropriate as many methods employ an initialization phase after each segment boundary, which means that the process should be stationary after each boundary. This is a requirement that is seldom met.

3.2 Spectral analysis of EEG

EEG is often divided into the frequency bands delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz) and beta (over 13 Hz). These limits are only approximate and other limits are sometimes used. These bands are also sometimes divided further into *e.g.* low and high alpha. Taking into account the attainable durations of stationary epochs and the theoretical resolution and variance of the spectrum estimates it is often questionable whether reliable inference for the sub-bands can be made. Some statistical studies support the hypothesis that frequency bins are heavily correlated and that the number of approximately independent generators would most often be only four.

The spectrum estimation methods can be coarsely divided into 1) Fourier methods, 2) parametric methods and 3) filter bank -type methods. The Fourier methods include the conventional variety of non-parametric methods such as smoothed and non-smoothed periodograms and Blackman-Tukey and other methods in which various windows are applied either directly to data or autocovariance estimate. We associate the Fourier transform approximations such as the Walsh transforms also to this class. The parametric methods and their use in spectral analysis are reviewed in Chapter 2. The filter bank -type methods employ a bank of parallel band pass filters into which the EEG is input. The outputs of these filters are demodulated to give the mean powers of the bands with the energy spectrum of the filters as spectral weights.

3.2.1 Fourier methods

The traditional method (since 1938) to estimate the spectrum of EEG is to use Fourier analysis [89]. See [52] for a review on Fourier transform based spectral analysis methods with special reference to EEG. Also most textbooks on time series analysis consider the basic Fourier transform based spectral estimation methods. It is a common misunderstanding that periodogram is the definition of power spectrum for the observed data. This is probably partly due to entry level textbooks such as [185] that call sample autocovariance just (deterministic) autocovariance or autocorrelation. See Section 2.1.2 for the correct definition of spectrum.

There are basically the three following approaches. Here X_t denotes the measurement vector and X_{-t} its time inverse, $*$ denotes convolution, W is a non-negative valued weight vector, S the power spectrum estimate and $\mathcal{F}$ denotes the (finite) Fourier transform operator. Here the scalings are omitted.

1. Data windowing and subsequent use of the Wiener-Khinchine theorem. $S = |\mathcal{F}(WX)|^2$. With no windowing ($W \equiv 1$) we obtain the most commonly used method and S is called the periodogram. All spectrum estimates are non-negative definite. The periodogram can be smoothed by convolving it with a frequency domain window. The non-windowed periodogram is asymptotically unbiased but not consistent. The consistent variants are not even asymptotically unbiased.

2. Computation of the Fourier transform of the windowed sample autocovariance. $\mathcal{F}(W(X_t * X_{-t}))$. With a triangle window we obtain the common periodogram. The spectrum estimates may not be non-negative definite.

3. Computation of periodograms of segmented data with or without segment overlapping and subsequent averaging over the periodograms. These are called Welch estimates and they are consistent, asymptotically unbiased and non-negative definite. Averaging the corresponding *amplitude* spectrum estimates over an ensemble will yield (possible severely) biased estimates.

Two more items are worth reminding. The first is that the spectrum of a time series is a function defined on the interval $[-\pi, \pi] \subset \mathbb{R}$, that is, on a continuous index set. The values given by any Fourier based spectrum estimator will not generally yield the values of the true spectrum in the uniform grid even in the deterministic case. See [188] for an exact account on the properties of Fourier transforms concerning time and frequency sampling. The other is that interpolation of the estimates (usually carried out by the so-called zero-padding) does not enhance the true resolution of the estimate although the visual appearance would suggest that.

Fourier transform based methods are usually referred to as FFT (Fast Fourier Transform) methods although FFT is only an algorithm that computes the finite Fourier transform in time (computational complexity) $T \log T$ instead of T^2 for the direct calculation.

The selection and analysis of different window families with special reference to EEG characteristics have been analyzed in [125]. A data-dependent window size estimation for the post-windowing of the periodogram is proposed in [173] The fitting of some "hump"-type functions to raw periodograms have been proposed in [140, 189]. These are not equivalent to any windowing.

One of the problems associated with the Fourier methods is the leakage of power between frequency bins. The problem is most relevant when there are spectrally relevant features in nearby bins and the difference between the powers is large. In [16] a pre-whitening procedure is proposed. First an AR model is estimated and the spectrum of the residual is then calculated with periodogram instead of the usual constant spectrum assignment, see Section 2.1.2.

One major problem with Fourier methods is the sensitivity to spikes and artefacts [53]. This follows from the fact that the spectrum of an impulse sequence is constant and therefore affects all frequency bins in the spectrum estimate.

The estimation of the powers of bands is usually carried out by summing over appropriate frequency bins of periodogram. See [52] for how to treat the background. For the statistics of periodogram based band power analysis see [213, 214].

The most appealing feature of the Fourier based methods is that the smoothed versions are quite robust. They do not perform either very well or badly and the interpretation of the results does not require understanding of very complex theories. The interpretation of AR model induced spectral estimates, on the other hand, requires understanding of rather difficult topics especially due to the instability of the estimates when some roots of the estimated model polynomial have near unity moduli.

Walsh transforms

The Walsh transform has been proposed to be used as an approximation to the Fourier transform of EEG [230, 146, 115, 119, 54]. The Walsh transform is the analogue of Fourier transform with an orthonormal kernel that attains only values ± 1. The Walsh kernel functions can be described as the Fourier kernel functions after taking the sign mapping. Thus the computation of the Walsh transform consists only of addings and subtractions.

Estimation of the bispectrum

The (power) spectrum is a function that depends only on the second moments of the process. For normal processes this is all that is needed since all the higher order moments can be calculated from this. The bispectrum is a third order method. It contains information for example about phase coherences between different frequencies and it can be used to distinguish between different non-normal processes that have the same power spectrum. AR models have also be used in the estimation of EEG bispectra [180]. See [51] for the applications of bispectrum estimation with EEG.

3.2.2 Parametric methods

Almost all the studies in which parameric estimates for EEG have been calculated, the corresponding spectral estimates have also been computed. Spectrum estimation can be said to be one of the main applications of AR models to EEG analysis. Most of the early papers on the parametric models of EEG were directly concerned with direct estimation of the spectrum [73, 25, 193] or with the estimation of the autocovariance [61, 236].

The main attraction in the use of parametric models in spectrum estimation is that the degrees of freedom of the model can usually be kept relatively low. For example, an AR(6) model may be adequate in some situations where we have, say, one sample of $T = 120$ points on which the estimate can be based on. A similar reduction in the degrees of freedom for a periodogram yields estimates with very poor frequency resolution. See [134, 164] for general accounts on the spectral estimation based on various approaches. As noted earlier the classical approach to deal with the nonstationarities of EEG is to divide the observation into small segments and assume the segments represent samples from essentially stationary processes. It is possible that the duration of these segments are very short. The ability of the estimation scheme to model short segments is then a central problem. In [118] a comparison between spectral estimation with short length AR models and periodograms was given.

Another feature of AR models is that they are relatively efficient in the estimation of center frequencies of bands. There are some studies in which the center frequency estimates from periodograms and AR models are compared, see for example [104]. The results are clearly in favor of AR models over periodograms.

Parametric spectrum estimates are relatively insensitive to spikes and artefacts. A robust procedure (tolerant to the mis-specifications in the distribution) for spectral estimation is proposed in [171]. The method is based on an iteration consisting of the following steps: AR model estimation, a "soft limiting" of the residuals and the recomputation of the sample.

The statistics of SPA parameters computed from different estimators for AR and ARMA parameters were discussed in [220]. They proposed that in spite of theoretical supremacy of ARMA models, the AR models are more feasible for the estimation of the band parameters of EEG.

We refer sometimes to the AR parameters and sometimes to the prediction coefficients. Although these coincide for normal AR processes and mean square estimates for predictors, the following must be taken into account. When we refer to AR models we assume that the model is a good approximation for the process, in other words, the prediction error process can be assumed to be orthogonal. However, we may still use a "predictor model" for the process when we do not necessarily try to acquire the orthogonality of the prediction error process. The predictor coefficients can then be used *e.g.* as features but we should not form the corresponding spectrum estimate without estimating the spectrum $\widehat{f_e}(\omega)$ of the prediction error process e_t.

Estimates for the model order

As there is no general EEG there is not a general model order that would be universally appropriate. In principle it would be possible to select the smallest order that is adequate for all expectable cases. Ideally the excessive prediction coefficients would then tend to be very small and perhaps effectively zero. This will, however, lead to excessive computations and the stability of the predictor parameter estimates decreases. The worst point is, however, that the instability of the spectrum estimates may increase dramatically. For this reason it is not advisable to select a "maximum expectable order".

Depending on the situation the estimated (used) orders range from $p = 2$ [180] up to $p = 48$ [221] where the latter figure was obtained with the AIC criterion for a frequency band 0 – 51 Hz. Most of the studies on specific applications of AR models to EEG include at least a preliminary estimation for the model order in that situation. The most often proposed model order is perhaps $p = 5$ [229]. See [221] for empirical evaluation of different model order selection criteria in EEG modelling. It must be noted that preprocessing of the data before modelling, such as filtering (typically band pass filtering with 1 – 20 Hz), effects the estimation of AR models severely. Very narrow band pass filtering means that the AR model describes the preprocessing filter more than the actual EEG itself.

3.2.3 Filter bank -type methods

The use of filter banks in EEG analysis dates back at least to 1966 [49]. In that study the problems involved in the selection of time windows, filter bandwidth and variance of the band estimates at the output of the detector are already discussed although in most subsequent applications these problems seem to have been ignored.

Several bandpass filter structures have been employed, for example Butterworth [101] and Chebychev [6] filters. We have no knowledge whether bandpass filters with any optimality structure have been used in EEG analysis.

3.3 Specific applications of autoregressive models to EEG

In this section we give a brief review of the use of parametric models with EEG. The review concentrates on such applications that might be approached with the methods proposed in this thesis. Since this thesis concentrates on the models of EEG we do not review clinical applications except occasionally. It must be noted that the modelling stage of EEG is hardly ever the end to any analysis but rather a building block of a larger system. The parametric models have in some applications acquired the status of discrete Fourier transforms in the sense that these methods are sometimes hardly mentioned even when they are used, see [114] for an example.

The AR and ARMA process models were first applied to the EEG in [236, 61, 73, 232]. In spite of the theoretical advantages of ARMA structure over AR and MA cases, ARMA estimates have not been widely applied to EEG analysis. This

is mostly due to the problems that are inherent in the estimation of the the ARMA parameters. In [175] it was suggested that EEG is not always a minimum phase process which means that the models are not causally invertible. The argument was based on estimated ARMA models.

The applicability of AR models to the EEG has usually been verified only experimentally but there has been some limited interest towards the theoretical justification of this [221, 235, 19]. AR models have been used in a wide collection of tasks concerning EEG, including spectral analysis, classification of brain states, detection of transients, prediction of epileptic seizures and the simulation of EEG. See [110, 155, 116, 41, 113] for general reviews on the use of parametric models with EEG.

3.3.1 Detection of transients

In this setting a transient refers to a short disturbance in the process that can otherwise be described as stationary. Detection of other kinds of short-time waveforms is usually not best done using the method described below. Some waveforms such as spike-and-wave and K-complexes are usually detected *e.g.* with template matching procedures [211, 50] or complex sequential nonlinear procedures [32]. Other methods that are proposed to be used as transient detectors include time domain peak-trough analysis [8], syntactic methods [58], time domain descriptors [218, 85], nonlinear filters [204] and fractal dimension calculation [5]. Of greater interest is a method for the detection of vertex waves and K-complexes that is based on a generator model and the corresponding detector [47]. See [88] for a review of spike detection algorithms in connection with epilepsy.

The detection of transients (disturbances, nonstationarities) with the aid of parametric models is usually called *inverse filtering*. We write the process in the form $x_t = y_t + \nu_t$ where y_t is the EEG and ν_t is the transient. Let $\Phi(z)$ be at least an approximation to the optimal predictor of y_t. Then $(1 - \Phi(z))y_t$ is approximately equivalent to the innovations of y_t with variance σ_e^2. Now assume that $\nu_t = \delta_{t-t_0}$. Then $(1 - \Phi(z))x_t = e_t + \widetilde{\Phi}_{t-t_0}$ where $\widetilde{\Phi}_t = \Phi_t$ when $0 \leq t \leq p$ and zero otherwise. This means that the inverse filtering reduces y_t to its innovations and that the impulse disturbation triggers a signal equivalent to the impulse response of the inverse filter. This is basically the idea presented *e.g.* in [157].

A modification of this was studied in [4] where the parameters were tracked with the LMS algorithm and the simple treshold detector was substituted with more general nonlinear functions. In [17] a modification was presented in which effectively the prediction error of the second difference of EEG was tracked.

In many cases the bandwidth of the EEG is such that the variance ratio σ_y^2/σ_e^2 is of the order 10 to 100 and the maximum of the impulse response is typically of the order 1. For the calculation of sensitivities and specifities and detection theory in general, see for example [169].

This method can also easily be used to suppress the transients by *e.g.* hard limiting the prediction error process to yield $\widehat{e}_t$ and calculating $\widehat{x}_t = (1 - \widehat{\Phi}(z))^{-1}\widehat{e}_t$. This can be done in the cases when ν_t is not considered as part of the EEG such

as an epileptiform spike or a spike-and-wave pattern but rather a (measurement) disturbance. This approach to parametric modelling is sometimes called robust estimation, see [102] for a general review on robust methods and [171] for an application to AR modelling of EEG.

3.3.2 Adaptive segmentation

Segmentation of nonstationary processes is another example of the use of inverse filtering. Consider a nonstationary process whose statistical description may change slowly in time so that the prediction error variance is time invariant. Let the optimal predictor of the process at time t_0 be $\Phi_0(z)$. Let the prediction error variance of this time-frozen predictor be $\sigma_e^2(t)$. It is then clear that $\sigma_e^2(t_0)$ is the minimum attainable prediction error variance and that $\sigma_e^2(t) > \sigma_e^2(t_0)$ always when the optimal predictor at time t differs from $\Phi_0(z)$. It is then possible to construct a statistics for the detection of a change point in the process based on the statistics of a (weighted) sample of the prediction errors from $\Phi_0(z)$.

In practice it is necessary to estimate the process variance and/or the deviation of the spectrum of the prediction errors from white noise to be able to construct a detector of a change point or a segmenter.

The approach taken in the classical papers [22, 21] is to estimate an AR model at the begin of each segment using a fixed window of about 200 points. Then a moving window AR model is estimated and the segment boundary is assigned to a time where the two models differ considerably from each other. The distance measure that was used was the mean quadratic difference between the spectra corresponding to the two models. It has been shown that the estimated transition time has high variance and that it is asymmetrical with respect to the direction of the transition, see [11].

In [111, 112] the Kalman filter was used to track the AR parameters of a time-varying ARMA process, where the aggregate model was an ARMA process in additive independent noise. The change-point detection was based on a short-time (sliding window) estimate of the approximate likelihood function and an estimate on the variance of the additive noise. They also discuss the tracking of the statistics of [22, 21] with Kalman filter.

In [72] the changes in the estimated cross spectra between two different channels were tracked. The criterion was selected to enable the use of a vector valued AR model. Also other features than prediction error variances and spectral parameters can be used as the tracking variables. In [132, 190] the problem of spindle detection was approached by tracking the EEG with a time-varying predictor of structure ARMA(6,2). The roots of the corresponding denominator polynomial were tracked and the EEG was segmented based on the movement of the roots in the complex plane.

There are many problems with the segmentation of EEG. The types of change are very diverse including almost stepwise changes such as sleep EEG changes sometimes are, slow drift-type changes or trends such as the changes in EEG when the effect of anaesthetics diminishes. The segmenters then often yield a larger number of segments that visual inspection would suggest [154]. The changes

may also be in the spectral characteristics or just in the variance; there might be a change between two different lengthy pseudostationary segments or a very short "excursion" from a single pseudostationary state.

Perhaps the most difficult task for a segmenter (segmentation algorithm) is to handle such cases where the changes can not be modeled as stepwise changes but the rate of change is so great that there cannot be assumed to be any pseudostationary segments long enough to allow for the initialisation, that is, the estimation of the reference predictor. In these cases the initial estimates could be based on a severely nonstationary segments. Such a case is the event related synchronization of occipital EEG that is studied in Section 5.2.2.

Other approches to the segmentation of EEG include information theoretic measures [207] and finite state automata [91]. See [94] and the references therein for a review on the current state of the art of segmentation with model based methods.

3.3.3 Classification and clustering

The task of the classification of EEG can be divided into two stages. First some features are extracted from the data for each segment and then the features are compared to representations of the classes to decide to which class the segment belongs. There are several principles (classifiers) on which the classification can be based on, see for example [65, 169]. In unsupervised clustering procedure the class representers are not initially given but the algorithm will itself generate classes based on some fixed measures of how much the features are allowed to differ from each other within a class. Reviews of classification and clustering methods with special reference to the applications in EEG are [82, 152].

Often we have to know the density functions for the features for each class to be able to compute *e.g.* the likelihoods for each class given the features of a segment. In practice this involves also the selection of the number of the classes and preselection of the segments belonging to each class to enable the determination of the class statistics. Since assumptions on the class statistics such as normality can often not be done, the size of the training set can turn out to be very large and impractical.

A central problem in classification is the selection of features. It is common jargon in EEG classification literature to say that different features yield different "discrimination powers", that is, different percentages of agreement between the classifier and the known classes. When the implementation of classifiers is considered it is clear that a set of features and the same features under a nonlinear bijection may perform very differently although in principle the situation does not change. This is due to that the discriminant functions may have radically different topologies. An example of a mapping inducing nontrivial changes to topologies is the mapping from polynomial coefficients to the roots. A common approach in the selection of features is to start with a large set and to discard some of these *e.g.* with the aid of stepwise discriminant analysis [196].

The traditional features used with the classification of EEG include mean power or amplitude and powers of some frequency bands. These are usually

calculated from FFT's of the segments [216]. Since the early 1980's it has also been popular to calculate these and other so-called spectral parameters from the respective AR models. This is due to studies that were somewhat in favor of AR parameters or some other spectral parameters obtained from the respective AR models [98, 123, 121, 20, 124, 45].

Other features include cepstral coefficients and center frequency [23, 44] roots of the AR model polynomials [79, 132, 190] reflection coefficients corresponding to the lattice representation of the AR model [126, 127] and correlation coefficients [35].

Some studies have been done in which it has been assessed that the parameters of the approximate spectral decomposition correlate strongly with the visual appearance of the EEG [106]. As noted previously, this is still considered important in EEG analysis.

Sleep scoring

An important subfield of the EEG classification is sleep scoring, or sleep stage classification. Sleep is usually classified into five sleep stages and the awake stage. These classes are called the Reschenhofer-Kales classes. The predictor coefficients were first suggested to be used to classify sleep stages in [79]. Since then sleep scoring has been a popular application of AR modelling of EEG [123, 122, 120, 121]. In [69, 70] Bodenstein-Praetorius algorithm was used to divide the EEG into variable duration segments and the prediction coefficients among some other features were used to cluster and classify the epochs. In [208] predictor coefficients for fixed length segments were calculated by averaging estimates from Kalman filter. The features were subsequently fed into the self-organizing map for clustering and classification. For a review of methods for the classification of sleep EEG see [219].

Also in sleep scoring the utilization of periodograms is the standard choice. In [67, 86, 128] periodogram was used to extract band powers that are used as features in sleep stage classification. This was done also in [83, 215, 183] where neural networks were used as classifiers. In [228] it was maintained that periodogram estimates were more suitable in the estimation of low amplitude mixed frequency EEG that is typical during wakefulness. There are also some other classification schemes where sleep/vigilance is divided into a greater number of states [158, 228].

3.3.4 Tracking of anaesthesia

Parametric modeling has also been applied to the tracking of the effect of drugs, especially the depth of anaesthesia. It must be noted that somewhat unlike sleep EEG, the EEG under anaesthesia is more like a continuous transition phenomenon. For practical reasons, however, the depth of anaesthesia is usually divided into classes [203].

Also in this application the first time series method that was used was periodogram [168, 166]. Autoregressive models were first suggested to be used in the estimation of the depth of anaesthesia in [78] where Kullback-Leibler nearest

neighbor classifier was used. Here distances of samples from the classes were calculated as the prediction error variances of EEG for the predictors representing the classes. In [37] fixed duration segments were modelled with ARMA(p, q) and AR(p) models of different orders. Rather than model parameters they examined the roots of the model polynomials which was done earlier also in [79]. Clustering with AR parameters as features has been done also in [226]. Also in [48, 12] ARMA models were estimated from fixed duration segments. Subsequently, in [48] the change points were estimated with the aid of several linear regression estimates for the time variation of the alpha band power. In [36] the patient's EEG was tracked with Kalman filter (predictor) and the state (prediction coefficients) was used to assess the effect of anaesthetic drugs and so aided in the respective control. The state was, however, not used as a direct control variable.

In [202] four different sets of features were used for the classification: band powers from FFT, AR parameters from Kalman filter, Hjorth descriptors and zero crossing statistics. In this study the discriminatory power of the different sets of features were almost equal.

3.3.5 Other applications

In [227] a method for eye movement artifact suppression was proposed. The method is based on parametric modeling of observed EEG as a sum of true EEG (AR process) and electro-oculogram that was handled as an exogenous input, see [148] for a unified treatment on methods for the estimation of parametric models with exogenous inputs. In [103] this was done with the aid of RLS algorithm connected to estimate the EEG with electro-oculograms (EOG) as regressors. The use of Kalman filter in the reductions of additive noise and EMG is described in [130, 9, 212].

AR models have also been used in the estimation of information flow between different parts of the brain. All these applications use thus methods for multivariate models. The most important application of this has been the localization of epileptic foci with stationary AR models [74, 225] and nonstationary AR models [192]. AR models were used in other information related studies also in [210, 105, 159]. For a short review on the applications, see [165]

In [137, 136] a preliminary attempt to realize man–machine communication with EEG was made. A severely paralyzed patient was asked to think of performing some tasks such as opening the door and rotating a wheel to turn left. The EEG during the times of concentration was recorded, modeled as an AR model and classified according to the task. After this the patient was equipped with EEG apparatus and was able to control a wheelchair with some degree of success. The wheelchair was equipped with a computer and on-line analysis and classification software and appropriate control actuators and telemetry devices.

3.4 Nonstationary analysis of EEG

It is impossible to define exactly what nonstationary analysis refers to. In this thesis we, loosely speaking, take nonstationary analysis to mean "continuously"

time-to-time changing estimates. In this setting, such estimates that are inherently time-varying but that are subsequently converted to segment estimates by averaging are not considered nonstationary analyses. Such an approach has been adopted *e.g.* in [36] in which the predictor estimates from Kalman filter were averaged over either a fixed or variable duration segments.

Even in cases of relatively rapidly varying situations it is still a common approach to divide the EEG into short segments and to model these as samples from stationary processes, for examples see [104, 71, 209]. In these cases the segments are usually modelled with AR models, see [118] for analysis of the estimation of EEG spectra from very short segments.

The main problem in nonstationary analyses is that the amount of "information" can become larger than that of the original data. For example, if we use the Kalman filter to estimate six time-varying prediction coefficients, we end up with six parameter evolution processes whose traces are practically impossible to interpret as such. On the other hand, we might look for an estimate for the time-varying power of the alpha waves. This is a variable whose time trace is clearly easily interpretable. This can in principle be estimated from a nonstationary parametric model [236] and is also done in Chapter 6.

The earliest nonstationary analysis of EEG is probably [25], where Kalman filter was used to track the prediction coefficients. This is an example of a stochastic regression method. The analysis was pursued further in [26]. Here the assumed hypermodel was the first order homogenous Markov model, or random walk, and was assigned only to the AR parameters. Thus the parameter (state) evolutions were assumed to be independent. In this and most later studies the ARMA extension was also studied, but the MA part was frozen to some prescribed state. Most notably, the estimation of state and observation noise (co)variances, $\Gamma_v = \mu^2 I$ and σ_e^2, respectively, by seeking the maximum likelihood of these parameters iteratively, is also discussed. An index of nonstationarity d is then proposed

$$d = \sqrt{\mathrm{var}(x_t)}\mu/\sigma_e \ .$$

Here d is considered as a constant and thus $\mathrm{var}(x_t)$ is actually also a constant. This is an implicit constraint to the parameter evolution processes. In [26] numerical and implementational considerations were also given. This was also the approach taken in [106, 109] where a larger class of EEG data was examined.

In [112] Kalman filter was used to track very slowly varying background and was then used to aid in the use of Bodenstein-Praetorius type change point detector. Slow potential shifts and the corresponding residuals were estimated with AR and ARMA models in [198].

The deterministic regression approach (TVARLS scheme) was proposed in [2] to be used as an aid in the segmentation of sleep EEG. This is probably the situation where the TVARLS scheme is least appropriate since the concatenation of the parameter evolutions is not "continuous" and constraint extensions that would yield meaningful estimates at the presegment boundaries have not been proposed. See [75] for the multivariate extension of the TVARLS scheme and applications to

EEG. A modified TVARLS scheme is discussed in Section 5.1 where the modelling of event-related syncronization changes is discussed.

In [133] a short-time Fourier transform scheme (a modified sliding window periodogram) was proposed and the method was then applied to visual event related desynchronization and synchronization (ERD/ERS) estimation. This is a non-parametric scheme where the stability of the algorithm is controlled by the duration of the sliding window. This kind of an approach stands no chance to estimate the dynamics of the transition phases as it was deduced. The spectrum estimates are often arranged into a matrix structure whose display is called the compressed spectral array (CSA) and is often interpreted as a data reduction method although the amount of data in the CSA is often greater than in the original data [205].

The use of cascades consisting of a band-pass filter and a demodulator (a rectifier and a low-pass filter) has been popular in the extraction of the (time-varying) powers of bands [194]. A short description of the problems related to this kind of approaches is given in Section 5.3.

One of the difficult topics in nonstationary EEG analysis is the performance evaluation of the tracking and other estimation algorithms. Since EEG itself does not provide any reference one usually has to resort to simulations. In many studies the tracking algorithms have been tested with time-varying sinusoids, for example, see [112]. This is not very appropriate since sinusoids form a very special class of processes, that is, the class of predictable processes. The performance of algorithms with such processes is usually much too good to enable any inference to be made with respect to non-predictable (regular) processes. A better approach to evaluate the performance of tracking algorithms is taken in [108, 231] where a time-varying AR process was generated. Here the parameters of the approximate spectral factorization were made time-varying and the description of the process was realistic. In Chapter 4 we discuss a systematic method for generation of simulations (realizations) of nonstationary EEG.

CHAPTER IV

Simulation of nonstationary EEG

In the estimation of the properties of stationary processes we have several methods with which we can obtain consistent estimates if the process is ergodic, which means that the time averages can be used to estimate the ensemble averages. We can then make our estimates more accurate by using more data to calculate them.

In the case of nonstationary processes we often cannot use time averages. To estimate a property of the process at a time instant we have, in the strict sense, just one point of the sample to build our estimate on. To get over this problem the assumption of slow time variation is often adopted. This gives us a possibility to use *e.g.* adaptive segmentation which will chop the sample to quasi-stationary (almost stationary) epochs to which we would then apply methods that are intended to be used with samples from ergodic processes [22, 111]. The applicability of this approach depends heavily on the rate of time variation, the criteria on which the segmentation is based and the correctness, or rather, the adequacy of the time series model.

There are cases in which the time variation is so rapid that the process cannot be approximated with a concatenation of stationary epochs but can not be decribed as an abrupt change from one stationary state to another either. The lengths of the quasi-stationary segments would then be too short to enable adequate estimates. This applies to both estimation and simulation of such processes. On the other hand, we cannot allow for a time variation that is totally "unpredictable". This would mean that if we have obtained a sample of the process and aim to estimate its time-varying characteristics, the model may have, loosely speaking, more degrees of freedom than the sample the estimates are based on [76].

If we can make an assumption of *smoothness* of the time variation we can still obtain meaningful estimates and at the same time allow reasonably rapid changes of process characteristics. This follows from the fact that one can have a smooth sequence of parameters describing the process, *i.e.* a sequence with small second and/or third differences, that has large first differences. This means that the parameter rate of change does not have to be small as such. This applies also to the building of models for time-varying processes and the use of these models to the generation of realizations of these processes.

4.1 Earlier approaches

The simulation of EEG using an AR model was suggested early in the seventies by [60, 232, 237], see also Chapter 3. Simulations of time-varying AR processes were used by [231] to test the modeling and tracking capabilities of the Kalman filter. The aim in this paper was to assess the applicability of the Kalman filter to the estimation of the time-varying EEG and the simulations were made to approximate the classical band structure of EEG. They used an approximate factorization of AR model to first and second order spectral factors that have been shown to correspond to the classical bands of EEG, see Section 2.1.3. To produce time-varying realizations they let the parameters of these factors, *i.e.* the power, bandwidth and the center frequency, to be time-varying. They did not, however, discuss how the time variability should be systematically implemented to produce realizations that would exhibit the time-varying characteristics of the EEG. In addition, the classical band structure of EEG does not always apply, especially when the EEG exhibits at least partly chaotic characteristics. They also discussed the use of simulations to aid in the recognition of certain spectral parameters of the model using visual inspection [106].

4.2 Simulation with TVAR representation

In this section we present a systematic method which can be used to generate realizations for the simulation of time-varying EEG. The method is purely phenomenological and does not rely on models for the neuronal generation of EEG as in [145].

The method is based on dividing the EEG to classes that are described by estimated distributions of their predictor (AR model) coefficients and prediction error variances. If the class evolution is not predetermined, this evolution is then simulated using *e.g.* estimated state transition probabilities and Markov models or state lengths (durations). The parameter representers of the classes are then drawn from the respective distributions according to the state evolution and concatenated in time. This piecewise stationary parameter evolution is projected onto a set of smoothly time-varying functions, which projection is finally used to generate the realizations by time-varying filtering.

4.2.1 Estimation and simulation of the state evolution and class representers

We assume that the time-varying EEG can be approximated with a progression of states $s_j,\ j = 1, \ldots$ each of which is a member of a class $\mathcal{C}_k,\ k = 1, \ldots, K$, and that within these states the process can be approximated with an AR(p) process. For the selection of an appropriate order p in different situations, see [229]. The final selection of p should be the maximum of the adequate orders of all classes. We also assume that within each state the coefficients $\phi_i^j,\ i = 1, \ldots, p$ of the predictor and the prediction error variance σ_j^2 corresponding to the AR(p) model are drawn from the density that can be approximated with a $p + 1$ –dimensional

normal distribution. Each class has its own distribution. We call the vector $\theta^j = (\phi_1^j, \ldots, \phi_p^j, \sigma_j^2)^{\mathrm{T}}$ a *representer* of the state s_j. We will further assume that the representer of the state s_j is independent of the representers of the states s_ℓ for all $j \neq \ell$ if the class to which s_j belongs is known.

There are several methods with which to estimate such a class evolution if the (probabilistic) description is not known a priori . One such a method is to model the class evolution as a Markov model and estimate the parameters of the model. This model can then be used to simulate the class evolution. The states are then drawn from the respective distributions in the order that the class evolution dictates. The estimation of a continuous-time Markov model and the simulation of sleep EEG class evolution (hypnogram) has been discussed in [139, 138]. In addition to the class evolution, the model should produce the durations T_j of the states.

The assumption of approximate normality of representers implies that we only need to estimate the means and covariances of θ^k for each class $\mathcal{C}_k$ to be able to generate representers.

If the means and covariances of θ^k are not known, they can be estimated from existing segmented and classified real EEG data *e.g.* by using any preferred method to calculate the parameters of the AR(p) model and the associated residual error variance for each segment. These can be used *in lieu* of the predictor coefficients and the prediction error variances.

Let there be $n_k > p$ estimates $\widehat{\theta^k}$ for θ^k. Estimates for the mean $\widehat{\mu}_k$ and covariance $\widehat{\Gamma}_k$ of $\widehat{\theta^k}$ can then be obtained as

$$\widehat{\mu}_k = n_k^{-1} \sum_{i=1}^{n_k} \widehat{\theta}_i^k \tag{4.1}$$

$$\widehat{\Gamma}_k = (n_k - 1)^{-1} \sum_{i=1}^{n_k} (\widehat{\theta}_i^k - \widehat{\mu}_k)(\widehat{\theta}_i^k - \widehat{\mu}_k)^{\mathrm{T}} \ . \tag{4.2}$$

We can now generate representers from the thus estimated $p+1$-variate normal distributions

$$\mathcal{N}_k(\theta) = (2\pi)^{-p/2} |\widehat{\Gamma}_k|^{-p/2} \exp\left(-(\theta - \widehat{\mu}_k)^{\mathrm{T}} \widehat{\Gamma}_k^{-1} (\theta - \widehat{\mu}_k) \right) \ ,$$

where $|\widehat{\Gamma}_k|$ denotes the determinant of the $\widehat{\Gamma}_k$, as follows: Generate a $p+1$-vector g of independent normal variables with zero mean and unit variance. Then the variables $\theta = L_k g + \widehat{\mu}_k$ are distributed with density $\mathcal{N}_k(\theta)$, where $\widehat{\Gamma}_k = L_k^{\mathrm{T}} L_k$ is the Cholesky decomposition of the $\widehat{\Gamma}_k$ [187].

Note that if the representers have been estimated from relatively short segments, the covariances $\widehat{\Gamma}_k$ may include a non-negligible contribution that is due to the small sample estimate properties. In such a case we can reduce the variances of the classes simply by dividing the covariance matrices by a number that is greater than unity. While this method is not the correct way to diminish the small sample contribution, it maintains the eigenvector structure of $\widehat{\Gamma}_k$ and

only diminishes the eigenvalues. This means that the only result of this operation is that the deviances of the simulated θ^ks from the means $\widehat{\mu}_k$ will be smaller. This will also decrease the possibility of obtaining a temporarily unstable model.

An AR(p) model is stable if and only if the roots $\zeta_i,\ i = 1,\ldots,p$ of the polynomial $1-\Phi(t;z)$ fulfil $|\zeta_i| < 1$ for all t. For some classes the roots of $1-\Phi(t;z)$ with $\phi_i = \mu_i,\ i = 1,\ldots,p$ may have almost unit modulus or the covariance may be large, which means that the probability of obtaining an unstable representer is non-negligible. This is why the stability of each representer should be verified by calculating the roots of the polynomial. All roots with modulus greater than or equal to unity should be transferred to lie inside the unit circle and the predictor reassembled. This can be done by factoring the polynomial to (complex) roots and multiplying each unstable root ζ_i by $|\zeta_i|^{-2}$. This operation will approximately restore the shape of the spectrum in most practical cases [149, subroutine fstab].

4.2.2 Generation of the time-varying AR model

We will discuss first two types of direct concatenation that are not feasible to generate time-varying realizations and how this unfeasibility can be overcome. Then we discuss the details of the method that was shortly described in the beginning of this chapter.

Direct concatenation

The majority of time series methods that are designed to track the characteristics of time-varying processes are based on either direct or indirect minimization of the prediction (residual) error [148]. It is thus desirable that the optimal predictor of the realization behaves in a way that it makes sense to compare the estimated model with it. For example, if the coefficients of the optimal predictor will decay relatively fast, we would use as the definition of time-varying spectral density (2.2.5).

There are at least two trivial ways to use concatenation to generate realizations of nonstationary processes which ways are not directly usable. The first is to generate separately stationary segments and then concatenate them directly. There are two major problems with this kind of realizations. In this case we have actually a non-overlapping sum of independent processes. Let x_t be the realization with zero mean for all t and t^* be the first instant of a new segment. Then, irrespective of the values at time t^*-1, the correlation of x_{t^*} with the entire history of x_t is zero, the predictor for x_{t^*} is zero and the spectrum estimate $f(t^*;\omega) = \sigma_e^2(t^*)$. This is clearly not desirable. In [143] a direct concatenation of AR(2) realizations were used to simulate EEG and test the performance of a segmenting algorithm. The accuracy of the segmenting was therefore excellent, in all cases less than the inverse of the maximum frequency of the spectrum, that is, the approximate "wavelength".

On the neuronal level this kind of a sudden change in process characteristics would correspond to a situation in which one activity inducing neuron cluster would shut down abruptly and at the same time another cluster would gain its

full effect. This is a totally unrealistic situation.

The second type of concatenation involves the concatenation of the parametric representations of the segments. This concatenation could then be used as a time-varying filter with abrupt filter coefficient changes at the segment borders. As above, this kind of change would correspond to an unrealistic abrupt change in the physical and chemical state of the neuron cluster. In addition, it turns out that the calculation of the optimal predictor is far from straightforward.

For this reason we have chosen to build a smoothly changing parametric representation for the process. We use a time-varying AR(p) model as the representation and generate the realizations by feeding white noise to the corresponding filter. If the coefficients of the model do not change very much during the correlation time of the process, the optimal predictor is approximately equal to the coefficients of the AR(p) at each time. Thus the reference against which the estimates are compared is directly accessible.

The generation of smooth parameter evolution

We start from a sequence of representers (predictor coefficients and prediction error variances $\phi_k^i,\ k = 1, \ldots, p$ and σ_i^2, respectively) which are the result of the simulated state evolution as described above. Let the corresponding sequence of the lengths (durations) of the segments be $T_i,\ i = 1, \ldots$. We call the concatenate of the segments a *block* which is of length $T = \sum T_i$. First we build the coefficient matrix Φ^{c} of piecewise constant parameter evolution

$$\Phi^{\mathrm{c}} = \begin{pmatrix} \phi_1^1 & \phi_2^1 & \cdots & \phi_p^1 \\ \vdots & & & \vdots \\ \phi_1^1 & \phi_2^1 & \cdots & \phi_p^1 \\ \phi_1^2 & \phi_2^2 & \cdots & \phi_p^2 \\ \vdots & & & \vdots \\ \phi_1^2 & \phi_2^2 & \cdots & \phi_p^2 \\ \phi_1^3 & \phi_2^3 & \cdots & \phi_p^3 \\ \vdots & & & \vdots \end{pmatrix} \begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} T_1 \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} T_2 \\ \\ \\ \end{matrix} = [\Phi_1^{\mathrm{c}}\ \Phi_2^{\mathrm{c}}\ \ldots] \ .$$

Next we smooth each individual parameter process $\Phi_j^{\mathrm{c}}\ j = 1, \ldots, p$ by projecting them separately onto a smooth subspace $\mathcal{S}$ of $\mathbb{R}^T$, where $T = \sum T_i$. Let us for the moment suppose, that the smooth subspace $\mathcal{S}$ is such that there are projections onto it which restore the approximate "shapes" of Φ_j^{c}. Let now $S = [\psi_1, \psi_2, \ldots, \psi_M]$ be a $T \times M$ matrix containing a basis of $\mathcal{S}$ in its columns.

The orthogonal projections Φ_j of Φ_j^{c} onto $\mathcal{S}$ are

$$\Phi_j = \Phi_j^{\mathrm{c}}\big|_{\mathcal{S}} = S(S^{\mathrm{T}}S)^{-1}S^{\mathrm{T}}\Phi_j^{\mathrm{c}} = P_{\mathcal{S}}\Phi_j^{\mathrm{c}} \ ,$$

where $P_{\mathcal{S}}$ is the orthogonal projector onto $\mathcal{S}$. The invertibility of $S^{\mathrm{T}}S$ is trivial since S is assumed to contain a basis for $\mathcal{S}$. The projection can be performed

simultaneously for all j giving

$$\Phi = P_{\mathcal{S}}\Phi^{\mathrm{c}} \ .$$

The predictor coefficients $\phi_k(t)$ are now of the form

$$\phi_k(t) = \sum_{l=1}^{M} q_{kl}\psi_l(t) \ ,$$

where $\{q_{kl}\} = Q = P_{\mathcal{S}}\Phi^{\mathrm{c}}$. To complete the specification we need the correspondingly smoothed evolution of the prediction error variance. As above, concatenate σ_i^2 to give

$$\Upsilon^{\mathrm{c}} = (\overbrace{\sigma_1^2, \ldots, \sigma_1^2}^{T_1}, \overbrace{\sigma_2^2, \ldots, \sigma_2^2}^{T_2}, \sigma_3^2, \ldots)^{\mathrm{T}} \ .$$

Then using the same basis we obtain

$$\Upsilon = P_{\mathcal{S}}\Upsilon^{\mathrm{c}} \ .$$

Next we generate a white noise sequence of zero mean and unit variance $e_t,\ t = 1, \ldots, T$. The realization is then obtained as

$$(1 - \Phi(t; z))x_t = \sqrt{\Upsilon(t)}e_t \ .$$

Due to the highly nonlinear mapping of polynomial coefficients to the roots, it would be very burdensome to add the stability condition $|\zeta_k(t)| < 1$ to the projection as a constraint in the least squares problem corresponding to the projection. This constrained least squares problem can be expressed as

$$\min_{\Phi}\{\|\Phi - \Phi^{\mathrm{c}}\|,\ \Phi_j \in \mathcal{S}\ \forall j,\ |\zeta_k(t)| < 1\ \forall k, t\} \ .$$

Since Φ tends to overshoot the stationary representers in parameter space, it is probable that in some cases the roots $\zeta_k(t)$ will be temporarily unstable. This is, however, not a problem if the overshoot and its extent in time are not large since in this case x_t does not diverge much.

The choice of basis for $\mathcal{S}$

It is always necessary to include the constant basis as one of the functions so that the stationary case falls into this setting as a special case whatever the basis. For convenience we will thus assume that $\psi_1 \equiv 1$.

The basis function selection has been discussed in the literature concerning time-varying AR modeling, for a review see [75] and Section 2.2.2. The problem of estimating optimal time-varying AR (TVAR) model can be viewed to be the inverse to the one at hand. In TVAR modeling the parameter evolution is constrained so that each parameter process is in $\mathcal{S}$.

The basis functions suggested to be used in conjunction with TVAR models include general polynomial, Fourier, Haar, Walsh, Legendre, spline and prolate

spheroidal wave bases. With the exceptions of the Haar and Walsh basis, which are block pulse bases, the suggested bases are able to model smooth changes in the parameter evolution. This ability is dependent on the choice of basis and the parameters concerning it, the dimension of $\mathcal{S}$ in particular.

The problem with most bases is that they are global within the block so that a small change anywhere in the block induces changes that extend overall within the block. This is obviously not desirable. What we wish to have is to maintain the local nature of the basis and simultaneously obtain a smooth parameter evolution.

We have chosen to use a basis consisting of shifted and scaled (sampled) Gaussian functions $\psi_i(t)$:

$$\psi_i(t) = \exp\left(-(t-t_i)^2/d_i^2\right), \quad i = 2, \ldots, M \ .$$

Obviously there can be no fixed general choice for M, t_i or d_i that would be optimal in some sense. The choice of these parameters depends heavily on the underlying situation and the demands. In the case of simulation the feasible parameters depend on the number and lengths (durations) of the stationary segments, or more generally, the distribution of these two.

According to our experience the Gaussian basis fulfils the requirements of locality and smoothness of evolution with relatively simple selection rules of M, t_i or d_i for a wide class of evolutions. The trivial choice for t_i and d_i is such that the half-widths of ψ_i will exactly cover the length of the block. This leads to $t_i = (i-1)T/(M-1)$, $2Md_i = T$, where T is the length of the block. The selection of M is yet left open. If the block consists of D segments of length L each, then $T = DL$ and we could choose $M = D + 1$. This selection leads to projections that are smooth but tend to overshoot the design classes heavily in the parameter space.

We have found that the selection

$$\begin{aligned} M &\approx 6D \\ d_i &\approx 3T/M \\ t_i &= (i-1)T(M-1) \end{aligned}$$

will produce reasonable projections if the segment length distribution is not very wide.

Other methods to obtain a smooth parameter evolution

Naturally it is possible to tailor the bases individually to follow the state evolution and just to smooth the transients of the parameter evolution. One way to implement this indirectly is to use the sigmoidal basis. An example of a sigmoidal type basis function is

$$\psi_i(t) = (1 + \exp(-c(t - t_i)))^{-1} \ ,$$

where t_i are selected to coincide with the segment borders and c is adjusted to give desired rise times.

It is also easy to build Φ^c and then filter the columns with a (noncausal) zero-phase low-pass filter. Just like with the sigmoidal basis this will, however, lead to a situation where the realization would be a concatenate of stationary and transition regions which is not very realistic.

In addition, it has been shown that the tracking capability of adaptive estimators depends on the type of evolution of the parameters [14]. If the parameters would be time-invariant for a long period of time, the estimators could perform markedly better with the realization than with the real EEG data.

In principle it is possible to realize the parameter evolution as a multivariable AR process. For many situtions this could be assumed to be a feasible model for parameter evolution. The probability density function of the parameters is, however, very difficult to control so that the model remains stable at all times. If the parameter processes are forced to the stability region after generation, the smoothness exhibited *e.g.* by a second order low pass type AR process is difficult to maintain. The testing for stability must be carried out for each t which will be very burdensome with long realizations.

4.3 Simulation of a two state EEG

As an example of the use of the proposed method we simulate the EEG (electrocorticogram) of a drowsy rat. This is a simple example since this EEG can be approximated as a process that toggles between two states. The amount of data on which the statistics we give here is far too small to enable any true inference. The main point of this example is to illustrate the steps of the proposed method in detail.

The state evolution

The data was visually classified to two states. As an aid in the classification we used the spectrum estimates and root locations of the modified covariance estimates of AR(6) model, see Section 2.1.1. The spectrum of rat EEG is often estimated with periodograms [46] and AR models have also been used in this case [162].

The exponential distribution induced by the Markov model did not fit the experimental segment length distribution of either class. We adopted another model for the state evolution. Since there are only two states, state 1 always changes to state 2 and *vice versa* . Thus we can model the state evolution simply by the estimated distributions of the segment lengths. The gamma distribution

$$g(T;\lambda_1,\lambda_2) = \frac{\lambda_2^{\lambda_1+1}}{G(\lambda_1+1)} T^{\lambda_1}\exp(-\lambda_2 T)$$

seemed to fit the observed histograms of both states reasonably well, where G is the gamma function. The estimated probability densities $g_1(T) = g(T;\lambda_1^1,\lambda_2^1)$ and $g_2(T) = g(T;\lambda_1^2,\lambda_2^2)$ are shown in Fig 4.1.

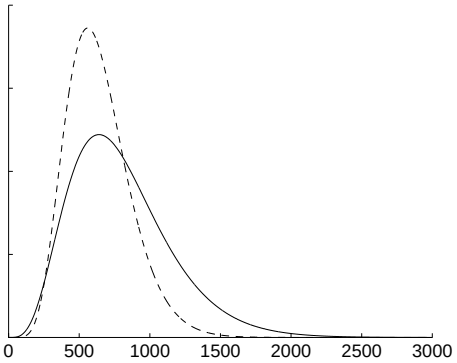


Figure 4.1: The density estimates $g_1(T) = g(T; 4, 0.006)$ and $g_2(T) = g(T; 7, 0.013)$ of the segment length densities for classes 1 (solid) and 2 (dashed).

The representer estimates

We used the forward-backward least squares method to estimate the coefficients of AR(6) models for both states. These coefficients and the corresponding residual variances were used as estimates of the predictor coefficients and prediction error variances. Forty estimates of the representers θ^1 and θ^2 were obtained. The marginal densities of each $\phi_i,\ i = 1, \ldots, 6$ were visually examined to support the normal approximation of the predictor coefficients. The marginal density of the prediction error variances σ^2, however, could not be approximated with the normal distribution, which was to be expected. To use another distribution would be very cumbersome, so we let the joint density be normal but applied hard limits to σ^2 when representers were drawn from the distribution. As hard limits we used the minimum and maximum of the observed residual variances of both classes.

The means $\widehat{\mu}_1$, $\widehat{\mu}_2$ and the covariances $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ are given in Table 4.1. The spectra and root locations corresponding to the means are given in Fig. 4.2. Note that the frequency corresponding to the second peak in the spectrum of state 2 is approximately two times the frequency of the first. This can be taken as a sign of partially chaotic nature of state 2.

The generation of realizations

The segment length sequence was generated by drawing independent random numbers from the distributions g_1 and g_2 in turn. The representer sequence was generated as described in Section 4.2.1.

The segment lengths and representers of a realization are given in Table 4.2. The corresponding concatenated predictor coefficient processes (vectors) Φ_1^c and Φ_2^c and the corresponding smoothed processes Φ_1 and Φ_2 are shown in Fig. 4.3a. The evolution of the prediction error variances Υ^c and Υ are shown in Fig. 4.3b.

TABLE 4.1

The estimated means $\widehat{\mu}_1$, $\widehat{\mu}_2$ and the covariances $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ of classes 1 and 2, see (4.1) and (4.2). The predictor coefficients are ϕ_k and the prediction error variances are σ^2.

	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	σ^2
$\widehat{\mu}_1$	0.5065	-0.0528	0.0619	-0.0828	-0.0009	-0.1431	0.5686
$\widehat{\mu}_2$	0.6722	-0.3583	-0.1741	-0.0562	-0.0344	-0.2811	1.1819
$\widehat{\Gamma}_1$	0.0362	-0.0335	0.0182	-0.0016	0.0049	-0.0055	0.0223
	-0.0335	0.0448	-0.0185	-0.0033	-0.0027	0.0031	-0.0296
	0.0182	-0.0185	0.0283	-0.0029	-0.0034	-0.0005	0.0056
	-0.0016	-0.0033	-0.0029	0.0139	0.0003	-0.0063	-0.0054
	0.0049	-0.0027	-0.0034	0.0003	0.0182	-0.0124	0.0026
	-0.0055	0.0031	-0.0005	-0.0063	-0.0124	0.0183	0.0082
	0.0223	-0.0296	0.0056	-0.0054	0.0026	0.0082	0.0444
$\widehat{\Gamma}_2$	0.0177	-0.0066	-0.0101	0.0127	0.0073	-0.0046	-0.0119
	-0.0066	0.0179	-0.0115	0.0111	-0.0092	0.0097	-0.0072
	-0.0101	-0.0115	0.0398	-0.0343	-0.0026	0.0157	0.0083
	0.0127	0.0111	-0.0343	0.0408	-0.0070	-0.0021	-0.0175
	0.0073	-0.0092	-0.0026	-0.0070	0.0197	-0.0192	0.0066
	-0.0046	0.0097	0.0157	-0.0021	-0.0192	0.0359	-0.0013
	-0.0119	-0.0072	0.0083	-0.0175	0.0066	-0.0013	0.4413

The corresponding realization is shown in Fig. 4.3c and an example of the original EEG in Fig. 4.3d.

To verify the similarity of the quasi-stationary epochs and the transition region between the original rat EEG data and the realizations, we show in Fig. 4.4a a segment of original data and a realization in Fig. 4.4b which is adapted to the original segment as follows: The EEG was divided visually into three parts and the representers for the quasistationary parts were estimated. The temporal extent of the representers was broadened to cover the transition region. The concatenated and smoothed parameter evolution is shown in Fig. 4.4c.

4.4 Discussion

We have presented a systematic method which can be used to simulate nonstationary EEG. The applicability of the method to *e.g.* human sleep EEG simulation is obvious. Apart from EEG that can be described by a stochastic state evolution, this method is also applicable to such processes that exhibit a deterministic type state evolution.

An example is the gradual desynchronization of alpha waves as a response to visual stimulus [163]. In such a case we can take samples of the EEG that are synchronized with the stimulus, estimate representer statistics before the stimulus using one segment $(k_1; T_1)$ and after the stimulus using several segments $(k_2, k_2, \ldots; T_2, T_3, \ldots)$. We can then use the predetermined state order $k_1, k_2, \ldots$ and segment lengths and vary the representers only.

Signals such as EOG, electromyelograms (EMG) and some event related

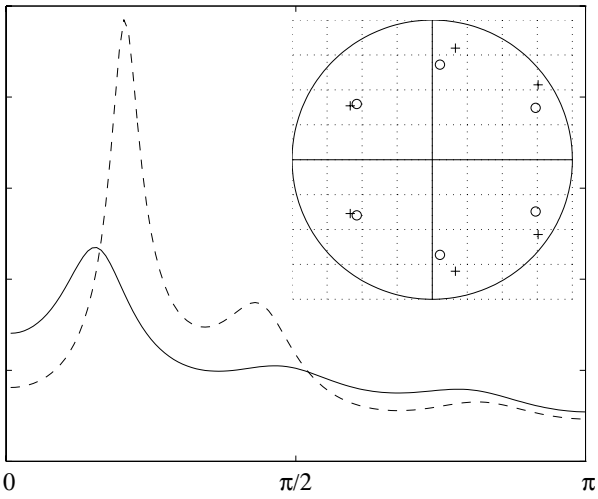


Figure 4.2: The spectra of classes 1 (solid) and 2 (dotted) corresponding to the means μ_1 and μ_2 and the model polynomial root (pole) locations in the complex plane, $\circ$: class 1, $+$: class 2

potentials can be added to the realizations directly. When doing this, however, it must be noted that the occurrences may correlate some states of the background. An example is the correlation of EOG with rapid eye movement (REM) sleep. To achieve a realistic situation the occurrence statistics should be estimated and used accordingly.

It is well known that some epochs of EEG are better described as a mixture of chaotic and stochastic behaviour rather than a regular stochastic process such as an AR(p) model, see *e.g.* [195]. However, for short segments the main difference between chaotic and regular description is that the former exhibits phase correlation whereas the phase of the latter should be independent between any two different frequencies and have uniform distribution between zero and 2π. Short segments of chaotic processes are periodic and can thus be modeled as a limiting case of a regular process having a line spectra with the lines exactly at the multiples of the inverse of the period. Such a process can be approximated also with an AR(p) model. As an example we can draw into attention the state 2 of the rat EEG and the corresponding simulation of Fig. 4.4. The visual appearance of the chaotic state 2 does not differ much from the simulation that is regular by construction. See [19] for a discussion of the applicability of AR(p) model to chaotic EEG.

If this approach is not considered adequate, the prediction error process Υ can be smoothly forced to zero and the thus obtained gap filled by a chaotic process that is multiplied with a taper to avoid the abrupt changes in process characterictics discussed in Section 4.2.2. See [63] for an example of the simulation

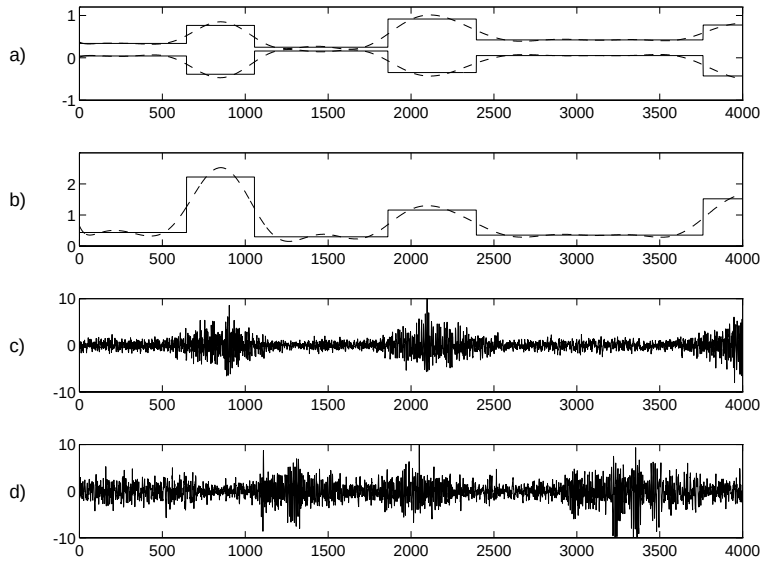


Figure 4.3: a) The concatenated predictor coefficient processes Φ_1^c and Φ_2^c and the corresponding smoothed processes Φ_1 and Φ_2. b) As in a) but for prediction error variances Υ^c and Υ. c) The corresponding realization. d) An example of original rat EEG.

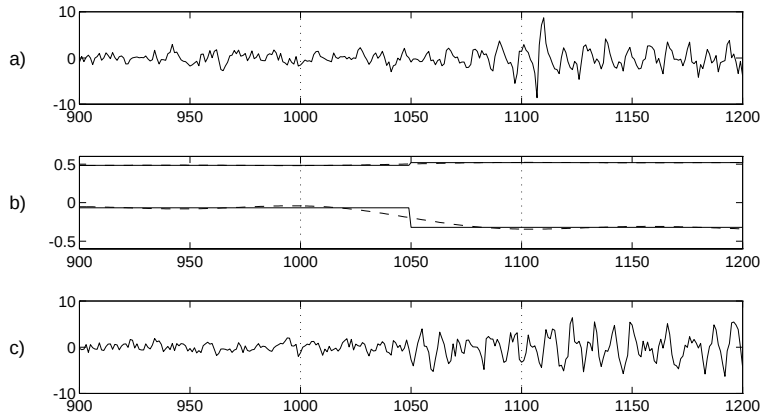


Figure 4.4: a) A segment of original rat EEG data. The dotted lines divide the segment to three parts: state 1, transition and state 2. b) The concatenated parameter evolutions Φ_1^c, Φ_2^c and the corresponding smoothed evolutions. c) An adapted realization.

TABLE 4.2

The segment lengths T_i drawn from the distributions $g_1(T)$ and $g_2(T)$ and the predictor coefficients $\phi_j,\ j = 1, \ldots, 6$ and prediction error variances σ^2 drawn from the normal distribution with means $\hat{\mu}_1$, $\hat{\mu}_2$ and covariances $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$. The prediction error marked with an asterix has been constrained to lie in the observed region.

State	$\sum T_i$	T_i	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6	σ^2
1	0	645.5	0.3402	0.0455	-0.02277	-0.08406	0.0358	-0.2347	0.437
2	645	410.2	0.7642	-0.3862	-0.2483	0.06458	-0.01698	-0.3176	2.224
1	1056	804.8	0.25	0.161	-0.1543	0.06727	0.05747	-0.295	0.3∗
2	1861	532.5	0.9159	-0.3473	-0.53	0.254	0.106	-0.4673	1.159
1	2393	1368	0.4241	0.05525	0.03071	0.01801	0.1109	-0.3348	0.3509
2	3761	555.4	0.774	-0.4261	-0.3595	0.02872	0.1362	-0.5736	1.522

of chaotic EEG. Another method to achieve this is to generate a noise process with constant spectrum but with appropriate phase coherence. This can then be fed to the time-varying filter.

In the above realizations the band processes were not orthogonal since the innovations were fed into the complete model. It has been argued that the band processes should be modeled as (approximately) orthogonal. The proposed method can be easily modified to fulfill this requirement. We just perform the approximate spectral factorization and use the time-varying extension of the approximate spectral factorization and mutually independent innovations for each band process. This was the approach in [237]. In [174] filters were used to generate the different band processes with mutually independent innovations.

CHAPTER V

Modified TVARLS Scheme and Estimation of ERD/ERS

The modeling of EEG with time-varying autoregressive (TVAR) models in which the parameter evolution is constrained to be a linear combination of basis functions is discussed. As a suggestion how to overcome the difficulties inherent in using generic basis functions and separately adjustable coefficients, a method is presented in which the basis is constructed so that the expected parameter evolutions can be optimally approximated by the basis. The method is then applied to the estimation of event related synchronization changes in the EEG. The results show that a properly selected basis enables effective estimation of the parameter evolution if the process to be estimated is compatible with the learning set with which the basis were constructed.

In the conventional TVARLS scheme the coefficients of the basis for each parameter evolution are separately adjustable, see Section 2.2.2. If we assume that the process is stable at all times, we can easily run into trouble with processes that are temporarily of the narrow band type. The roots of the characteristic polynomial of the model will then easily have modulus greater than unity temporarily, which is due to the tendency of the parameters to overshoot in the parameter space. Naturally, unstable models are not always a serious problem *e.g.* if the end to the modeling is prediction since the predictors corresponding to finite order AR models are always stable. However, if the end is *e.g.* to obtain an evolutionary spectrum estimate, this is a major problem.

In some cases more can be assumed of the process than just smoothness or slow variation. One such a case is the event related desynchronization/synchronization (ERD/ERS) of alpha waves of EEG [163, 194]. For example, when a patient has his/her eyes closed, the occipital EEG (primary visual cortex) shows high intensity in the 8–12 Hz region (alpha band, synchronization) while the opening of the eyes this intensity decreases or even vanishes (desynchronization). In this kind of a situation we can assume that the EEG exhibits a more or less rapid transition from a state to another and that the transition starts at some time after the visual stimulation.

While it seems that we had access to an ensemble of realizations by repeating the visual stimulation, this is not necessarily the case. It is well known that in

many EEG experiments of this type the patients are subject to habituation, that is, the responses to stimulations may exhibit various kinds of trends. In such cases the treatment of consecutive responses as an ensemble does not give us a correct view of the situation since we do not actually have an ensemble. Rather, we will obtain some kind of "average" estimates if we *e.g.* calculated the covariance of these "pseudoensembles".

In this chapter we propose an approach to TVARLS modeling in which prior assumptions on the dynamics of the process can be taken into account. This is accomplished by first forming a representative set of expectable parameter evolutions, or a *learning set*. Since the individual parameter evolutions in practically all relevant cases are mutually correlated, the learning set is constructed so that these correlations are not lost. A low dimensional subspace of the space of all expectable parameter evolutions is then extracted in such a way that the evolutions of the learning set can be approximated in this subspace with minimal error. This is accomplished via the eigendecomposition of the covariance matrix of the learning set.

The rest of the chapter is organized as follows. In Section 5.1.1 we discuss the problems inherent in the use of generic bases and the determination of the adjusted bases in general. We also modify prediction equations to give a scheme with non-separately determined basis. In Section 5.2.2 we discuss the formation of the learning set for the visual ERD/ERS data. In Section 5.2.3 we evaluate the performance of the method with simulations and in Section 5.2.4 we apply the method to real ERS data and verify that the method is applicable to the ERD/ERS measurements. We also show how the evolution estimates could be employed to further estimate *e.g.* delays in the start of the synchronization/desynchronization after the trigger. Finally, in Section 5.3 we address the problems and the potentials of the proposed method and discuss some previous methods that have been used in the analysis of ERD/ERS.

5.1 The modified TVARLS scheme

5.1.1 The consept and determination of the optimal subspaces

It is clear that the ability of the TVAR scheme to model the time-varying characteristics of the process is limited by the ability of the chosen set of basis functions to approximate the optimal predictor evolution [2]. Without any prior information on the evolution of the optimal predictor coefficients we are not able to treat the problem of optimality of the basis either.

If the second order statistics of the process were known, the optimal (in the mean square sense) time-varying linear p–order predictor for x_t could be calculated, *e.g.* by using the methods discussed in Section 2.2. Assume that the optimal predictor exists but we have no access to it and use the TVARLS scheme instead. If we now have reason to assume *e.g.* that the evolution of the optimal predictor is constant on a time interval, we should employ such a basis that forces the parameter evolution to be constant on this interval. Generally speaking, the subspace $\mathcal{S}$ should contain only such parameter evolutions that are assumed to

be expected. Naturally, in most cases this kind of prior assumptions can not be made, but when these can be done, they should somehow be taken into account.

Assume further that we have prior estimates on the probability distribution of the predictor at times $t = 1, \ldots, T$ and that these distributions are not separable. If we use separately adjustable coefficients in (2.9) we can not introduce prior information on the densities to the estimates easily.

For these reasons we consider *coefficientwise concatenates* of the parameters in the sequel. This way we can take into account the mutual correlation between the parameters of the predictor. We denote these concatenates by $\Psi \in \mathbb{R}^K,\ K = pT$

$$\Psi = (\phi_1(1), \ldots, \phi_1(T), \ldots, \phi_p(1), \ldots, \phi_p(T))^{\mathrm{T}} \ .$$

The concatenates Ψ are also called evolutions although the indexes to Ψ are not interpreted as previously.

Let us assume that we have access to a set of evolutions $\Psi_n,\ n = 1, \ldots, N$ that can be assumed to represent all expectable evolutions Ψ^* in the following sense. Let $\Theta \subset \mathbb{R}^K$ be the subspace spanned by $\{\Psi_n,\ n = 1, \ldots, N\}$ and P_Θ be the orthogonal projector onto Θ. Of all $\widetilde{\Psi} \in \Theta$ the one nearest to Ψ^* with respect to the Euclidean norm is $\widetilde{\Psi} = P_\Theta \Psi^*$. Assume now that $\|\Psi^* - \widetilde{\Psi}\| < \delta_1$ for some appropriately small δ_1. We could thus use a basis for Θ as the basis for the parameter evolution. However, since the dimension N of Θ can be large, we can easily run into stability problems when we try to estimate the parameter evolution. For this reason we aim next to determine an M-dimensional subspace $\Theta(M) \subset \Theta$ so that the mean of the errors $\|\Psi_n - P_{\Theta(M)}\Psi_n\|^2,\ n = 1, \ldots, N$ is minimal, where $P_{\Theta(M)}$ is the orthogonal projector onto the subspace $\Theta(M)$.

Denote by Γ the non–centered sample covariance matrix of the set $\{\Psi_n,\ n = 1, \ldots, N\}$

$$\Gamma = N^{-1} \sum_{n=1}^{N} \Psi_n \Psi_n^{\mathrm{T}} = N^{-1} \bar{\Psi} \bar{\Psi}^{\mathrm{T}} \ , \tag{5.1}$$

where the $K \times N$ matrix $\bar{\Psi} = (\Psi_1, \ldots, \Psi_N)$. Let v_k and $\lambda_k,\ k = 1, \ldots, K$, be the eigenvectors and eigenvalues, respectively, of Γ, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_K \geq 0$.

It is well known that the mean error criterion is minimized when $\Theta(M)$ is spanned by the M eigenvectors v_m corresponding to the M greatest eigenvalues $\lambda_m,\ m = 1, \ldots, M$ [184, 131]. The subspace $\Theta(M)$ is called the principal subspace of dimension M and if the set $\{\Psi_n,\ n = 1, \ldots, N\}$ was a true ensemble, we would call this approach the sample principal component analysis (PCA). We can thus use these eigenvectors as the basis for parameter evolution with the modifications discussed in Section 5.1.1.

The mean approximation error of the learning set in Θ is

$$J_M = \sum_{m=M+1}^{K} \lambda_m \ .$$

The mean normalized approximation error is $J_M / \mathrm{tr}\, \Gamma$, where tr denotes the trace

of a matrix. If the eigenvalues decay sufficiently fast, we may obtain a small normalized approximation error with a small value of M.

Let the maximum approximation error be $\|\Psi_n - P_{\Theta(M)}\Psi_n\| < \delta_2$ for all n. We have then by triangle inequality that

$$\|\Psi^* - P_{\Theta(M)}\Psi^*\| < \delta_1 + \delta_2$$

for all expectable parameter evolutions Ψ^*. The key problem in the formation of the learning set and in the determination of Θ is to achieve as low a dimension M as possible for a given total approximation error $\delta_1 + \delta_2$. The notion of "optimality" now refers to the optimality of $\Theta(M)$ with respect to the learning set, not to the expected evolutions. The determination of the optimal subspace $\Theta^*(M)$ with respect to the expected evolutions requires that the covariance of these is known, which can very rarely be the case. However, if we can assume that δ_1 is small, $\Theta(M)$ can be assumed to be a good approximation to $\Theta^*(M)$.

The modified prediction equations

The constraining of the (concatenated) parameter evolution to $\Theta(M)$ is equivalent to replacing the functions $\psi_m(t)$ with the respective sections of $v_m(t)$ for each individual parameter and setting $c_{mk} = c_m$ for all $k = 1, \ldots, p$.

Let $v_m = (v_m^1(t)^{\mathrm{T}}, \ldots, v_m^p(t)^{\mathrm{T}})^{\mathrm{T}}$, $m = 1, \ldots, M$ and $t = 1, \ldots, T$. The predictor coefficients are now written in the form

$$\phi_k(t) = \sum_{m=1}^{M} c_m v_m^k(t) \ .$$

Thus, instead of the columns of H in (2.11) we have

$$\varphi_m(t) = \sum_{k=1}^{p} v_m^k(t) x_{t-k}$$

as regressors. The prediction estimate is now of the form

$$\begin{aligned} \widehat{x}_t &= \sum_{m=1}^{M}\sum_{k=1}^{p} c_m v_m^k(t) x_{t-k} \\ &= \sum_{m=1}^{M} c_m \varphi_m(t) \ . \end{aligned}$$

The prediction equations are as in (2.12) but with $c = (c_1, \ldots, c_M)^{\mathrm{T}}$ and

$$H = \begin{pmatrix} \varphi_1(p+1) & \cdots & \varphi_M(p+1) \\ \vdots & & \vdots \\ \varphi_1(T) & \cdots & \varphi_M(T) \end{pmatrix} \ .$$

The dimensionality of the LS problem has now been reduced from pM to M and the estimated parameter evolutions are in $\Theta(M)$ by construction.

It is recommended that the LS solution of c is not done via the normal equations. Instead, the reduction of H to a canonical form via an orthogonal transform will lead to a modification of the prediction equations, the solution of which is numerically more stable [87].

5.2 Modeling of ERD/ERS

5.2.1 Previously used methods for the tracking of alpha type activity in EEG

In [99] a method based on phase locked loop circuitry is presented in which the center frequency of EEG is tracked. This was then applied to a visual learning task. It was noticed that the center frequency was highly nonstationary during periods of less than 5s.

Even relatively recent investigations on time-varying characteristics of EEG rely on segmentation and subsequent stationary AR models of these. These include tracking of alpha frequency before generalized spike-and-wave complexes [104].

In [132, 190] the problem of spindle detection was approached by tracking the EEG with a time-varying predictor of structure ARMA(6,2). The roots of the corresponding denominator polynomial were then tracked with a damped complex Newton method. It was argued that the classification of the EEG between spindle and non-spindle classes was easier to accomplish with a method that is based on these roots than directly on the model parameters. Earlier, the model polynomial roots have been used as features in [79, 36, 37] but the analyses were based on pseudostationary segments.

Several methods that are based on existing integrated circuits of different levels of complexity have been proposed. Usually these methods require the setting of some parameters such as tresholds for which there are not always rules. Often these methods employ some kind of ad hoc post-processing by a computer. Most reports are specific only on the implementation of the circuit and do not discuss the performance analysis of the method. As examples we list the following. In [34] an analog phase-locked loop circuitry was described. A modification of this was reported in [197]. A method based on separate amplitude, period (zero crossing) and pattern channels was proposed in [135]. A method that is based on a simple analog comparator equipped with timers and counters is described in [172].

In [199] the running fractal dimension was used to detect spindle type background activity. Complex demodulation (envelope detection with known center frequency) was used in [97] to detect sleep spindles.

5.2.2 Construction of the learning set

To test the proposed approach we conducted a visual ERD/ERS test using the international 10–20 electrode system. This was carried out by opening and closing the eyes with an auditory stimulus (beep) on 15 second intervals. The subject

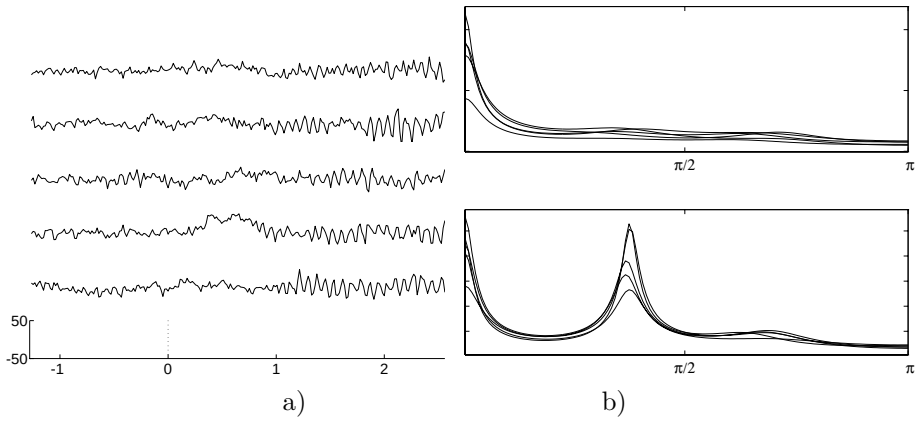


Figure 5.1: a) Five samples of ERS. Vertical scale in microvolts and time in seconds. The trigger is set at time $t = 0$. b) Amplitude spectra of AR(6) models of desynchronized (top) and synchronized (bottom)states for the five samples. Vertical scale is arbitrary and horizontal is in normalized frequency.

was a healthy young female. We chose five samples of the measured EEG at the transition from desynchronized state to synchronized state (eyes open $\rightarrow$ eyes closed). The samples were low–pass filtered to allow for the decimation of the sampling frequency from 250 Hz to 250/4 Hz. The final length of the samples was 960.

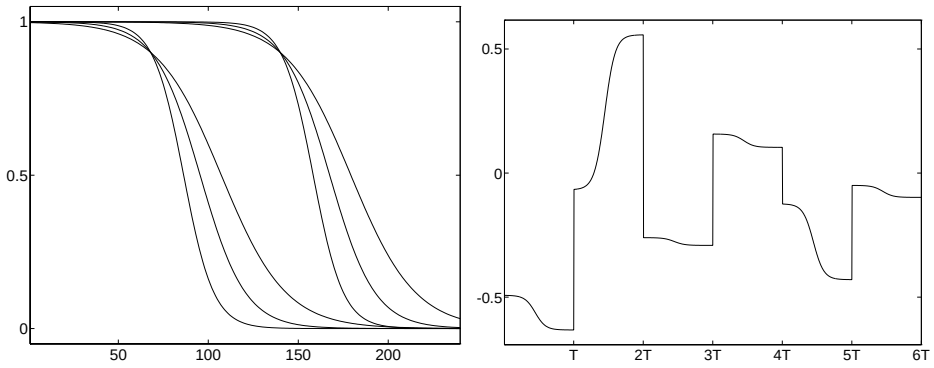
For the phenomenology and clinical relevance concerning alpha rhythms and the ERD/ERS phenomenon, see [163, 194, 234, 141, 27]. For a discussion on the neuroanatomical basis of rhythms, see [222, 156, 153]. For the statistics of the duration of alpha state during the eyes-closed condition, see [24]

We estimated the AR parameters with $p = 6$ for both synchronized and desynchronized states for all five samples using segments of length 240 at the tails of the samples. The parameters were estimated with the modified covariance method [164]. The samples are shown in Fig. 5.1a. The amplitude spectra corresponding the models are shown in Fig. 5.1b. We form the learning set first. We assume that the sets of AR(6) parameters describe the EEG in the synchronized (A_j^{S}) and desynchronized (A_i^{D}) states. We form the parameter evolutions from every A_i^{D} to every A_j^{S} assuming that the transition between the two states is smooth. As the model for transitions we use sigmoid function for which we use three different slopes and 10 different delays of transition. As the length of the parameter evolutions we use $T = 240$. The parameter processes $A_{ijk\ell}(t) \in \mathbb{R}^{p\times T}$ are formed by

$$A_{ijk\ell}(t) = A_i^{\mathrm{S}} + (A_j^{\mathrm{D}} - A_i^{\mathrm{S}})\sigma_{k\ell}(t) \ ,$$

where $i, j = 1, \ldots, 5$, $k = 1, \ldots, 10$, $\ell = 1, 2, 3$ and $\sigma_{k\ell}(t)$ is the sigmoid function

$$\sigma_{k\ell}(t) = (1 + d_0 \exp(d_{1\ell}(t - d_{2k})))^{-1} \ ,$$



a) b)

Figure 5.2: a) The extreme cases of sigmoids used in the construction of the learning set. b) An example of a member Ψ_n of the learning set showing the structure of the concatenated parameter evolutions.

where $d_0 = 1/9$, $d_{11} = 0.056$, $d_{12} = 0.08$, $d_{13} = 0.12$ and $d_{2k} = 68 + 8(k-1)$. The parameter evolutions are monotonous by construction. The extreme cases of the sigmoids $\sigma_{k\ell}(t)$ are shown in Fig. 5.2a. The learning set thus contains $N = 5 \cdot 5 \cdot 10 \cdot 3 = 750$ vectors Ψ_n of length $K = pT = 1440$, that are the componentwise concatenates of $A_{ijk\ell}(t)$. An example of a member Ψ_n of the learning set is shown in Fig. 5.2b.

The covariance Γ of the learning set was only implicitly formed as in (5.1). Since the matrix $\bar{\Psi} = (\Psi_1, \ldots, \Psi_N)$ is of size 1440×750, Γ would be of size 1440×1440. On many platforms the calculation of the whole eigendecomposition of Γ would be computationally too burdensome and memory demanding, especially when we only need a few principal eigenvectors. Fortunately, we can use the so–called orthogonal iteration method [87] to calculate the principal subspace in such a way that we do not actually even have to compute Γ. This method was then used to calculate 20 principal eigenvectors and the corresponding eigenvalues. The 20 largest eigenvalues of Γ are shown in Fig. 5.3a. The sections $v_m^2(t),\ m = 1, 3, 5, 9$ of the basis functions corresponding to the coefficient $\phi_2(t)$ are shown in Fig. 5.3b.

Note that by construction of the learning set the times and the rates of change of the transitions as well as the initial and final states are assumed to be independent.

5.2.3 Simulations

To find a feasible dimension M of the basis we generated realizations with members from the learning set and formed the modified TVARLS estimates for different M. The choice $M = 5$ seemed to be a reasonable compromise between approximability and stability. Too low an order M means that the evolutions can not be adequately

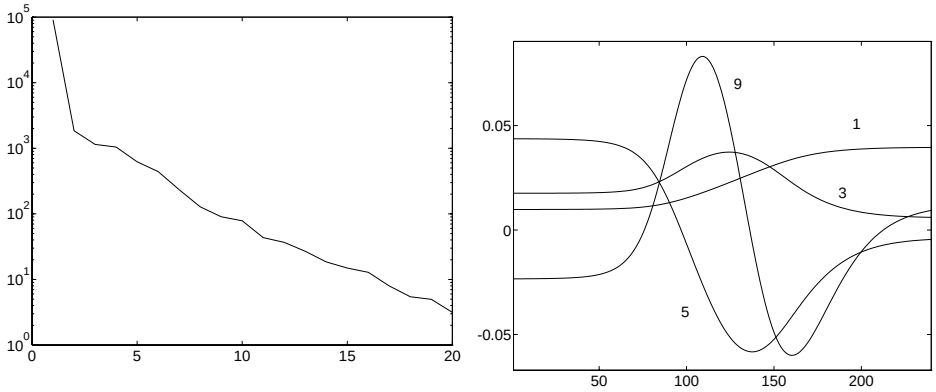


Figure 5.3: a) The 20 first eigenvalues of the covariance matrix Γ of the learning set $\bar{\Psi}$ b) The sections of eigenvectors $v_m^2(t),\ m = 1, 3, 5, 9$ of the covariance matrix Γ corresponding to the parameter $\phi_2(t)$.

approximated by the basis while too high an order introduces instability to the estimates in sense that the realization–wise evolution estimates have non–acceptably high variance.

Since we are more interested in the difference between the true and estimated evolution of the prediction coefficients (mean deviation) than residual error variance we will focus our attention on the mean deviation of the parameter $\phi_2(t)$, the mean change of which is the greatest for the learning set.

Three different evolutions of $\phi_2(t)$ and the corresponding estimates $\widehat{\phi}_2(t)$ from single realizations are shown in Fig. 5.4a. The initial and final states (and so also $\phi_2(1)$ and $\phi_2(T)$) are equal in the three cases. As we can see, the estimates $\widehat{\phi}_2(t)$ can track the transition time differences in $\phi_2(t)$ but tend to be biased towards the "center of the learning set". In addition, the estimated rates of change are too low. These are features that can be explained by the nature of the eigenvectors as shown by Fig. 5.3b in which the eigenvectors that exhibit qualitative differences are shown. The linear combinations of the 2 first eigenvectors have all approximately the same instant and slope of the transition. The features in $v_k(t)$ that allow for changes in the instant of the transition occur from eigenvectors $v_3(t)$ on but all these have the same slope. The corresponding slope features occur in the eigenvectors from $v_9(t)$ on but the features for estimating the instants of transition are still present. Thus the approximability of the slopes of transition of the evolutions would necessitate that we use $M > 5$.

To test the variance of the estimates we calculated the means and variances of parameter evolutions in several cases of the learning set. A typical case is shown in Fig. 5.4b with true and mean evolutions and standard deviation intervals for 20 realizations.

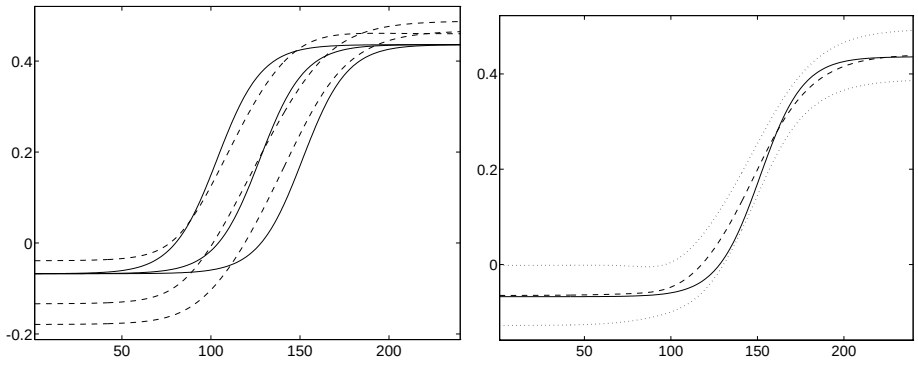


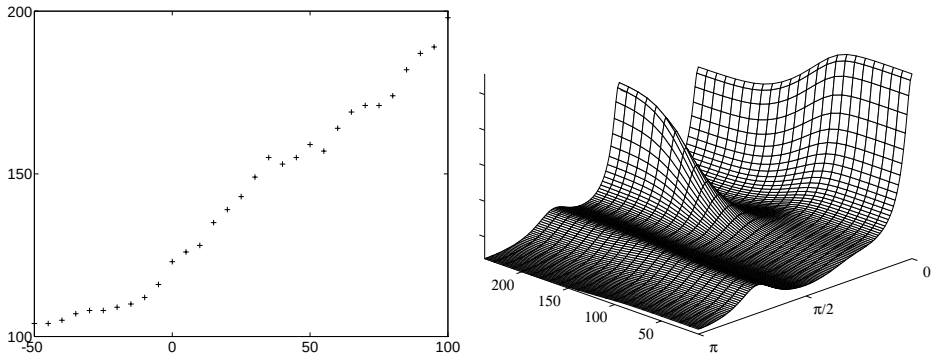
Figure 5.4: a) The parameter processes $\phi_2(t)$ of three examples of the learning set (solid) and the corresponding estimates (dashed) based on single realizations. b) A true parameter process $\phi_2(t)$ (solid), the mean estimate (dashed) and standard deviation interval (dotted) of 20 realizations.

5.2.4 Tracking of ERS

To evaluate the applicability of the method to real data we estimated the predictor evolutions for the data from which the initial and final predictor estimates were calculated, see Fig. 5.1a.

Although we do not propose this method to be used in estimating the instant of an abrupt change in EEG, we conduct an experiment in which we try to evaluate the tracking capability of the estimates. Since there are now no true parameter processes, we can not compare the estimates to these. We take a sample x_t of the measured EEG and form the estimates for $x_{t+T_{\rm tr}}$ where $T_{\rm tr}$ was allowed to extend somewhat beyond the modeling capability of the learning.

We estimate ad hoc the instant of change by the time $\widehat{T}_{\rm tr}$ that is nearest to the mean of the estimated initial and final states, that is, $\widehat{\phi}_2(\widehat{T}_{\rm tr}) \approx (\widehat{\phi}_2(1)+\widehat{\phi}_2(T))/2$. These ad hoc estimates $\widehat{T}_{\rm tr}$ as a function of the shifts $T_{\rm tr}$ are shown in Fig. 5.5a. The relationship between these is approximately $\widehat{T}_{\rm tr} \approx 0.75 \cdot T_{\rm tr} + T_0$ when $\widehat{T}_{\rm tr}$ is compatible with the learning set. The reason that the slope is smaller than unity is that the employed set of basis has a clear bias to estimate the change instant towards the average of the change points in the learning set as explained in Section 5.2.3. This tendency is diminished when the order M is increased. We can also see that the degree of linearity diminishes when $x_{t+T_{\rm tr}}$ is not compatible with the learning set. In Fig. 5.5b we show a typical evolutionary spectrum estimate (spectrogram) that is based on the time–frozen parameters of the estimated TVAR model.



a) b)

Figure 5.5: a) The ad hoc estimate of the transition instants as a function of the shift of a single sample of EEG. The slope is approximately linear ($\widehat{T}_{\rm tr} \approx 0.75 T_{\rm tr} + T_0$) on the interval where the basis allows for changes in the parameters. b) An example of a modified evolutionary (amplitude) spectrum estimate based on the TVARLS model with $\mathrm{var}(e_t) \equiv 1$.

5.3 Discussion

We have discussed a modification of the TVARLS scheme in which the individual parameter evolutions are not separately adjustable, but in which the mutual correlation of the parameters is taken into account. We have also presented a principal component type method for estimating such a basis for the parameter evolution that can optimally approximate the expectable evolutions. We have then applied the proposed method to the modeling of the ERD/ERS phenomenon of EEG.

The crucial problem of the method is the existence of a low dimensional subspace in which the expectable evolutions can be approximated with adequately small errors. If this is not the case, *e.g.* when the eigenvalues of Γ decay too slowly, this method should not be applied. The second problem is the formation of the learning set which may not always be as straightforward as in this case.

It must be noted that the approximability of "marginal" members of the learning set with a basis is poorer than the approximability of more medial members. For this reason it could be recommended that the learning set should contain members that are slightly off the expected region. Otherwise it could be necessary to use a larger dimension M than with this trick if we wish to have a uniform approximability on the whole set. On the other hand, too large a learning set will yield a set of eigenvalues of Γ that will decay slowly, which means that to maintain the approximability we should again increase M which in turn decreases stability of the estimates.

There are two previous approaches to estimate the ERD/ERS evolution. The first is a short time Fourier transform -type (STFT) scheme in which discrete

Fourier transforms are calculated from sliding windowed estimates of sample correlation [133]. In this method the window has to be made relatively short to maintain adequate time resolution and the variances of the time–frequency bins tend to become so large that the usefulness of the estimates suffers. We would thus be compelled to use averaging of the individual estimates.

The other method is to pass the EEG realization–wise through bandpass filters, square the outputs and average over realizations [194]. Thus we obtain *e.g.* an estimate for the time–varying variance (power) of the alpha band. However, the bandpass filtering exhibits some drawbacks. The first is that the decay time of the filter impulse response (from 1 to 0.707) is approximately equal to the inverse of the width of the filter spectrum. If the band is 8–12 Hz, the decay time is about 0.25 s. In practice this means that the evolution estimates are effectively convolved with a window of this duration. The other problem in the filtering approach is that we are not able to track changes in the center frequency of a band within the passband. This implies that, *e.g.* if the alpha band "peak" moves towards the boundaries of the passband, but does not change in variance, the variance of the output of the filter decreases giving us a false indication. To overcome this problem we can divide the alpha band into sub–bands, but this will further decrease the time resolution.

The proposed method does not necessitate ensemble averaging and it can thus track changes in single realizations and thus also trends through the ensemble. The frequency resolution is that of the AR models in general, but with the restrictions imposed by the subspace constraint. Some special information, such as the time–varying variances of the EEG bands can be directly accessed using the approximate spectral decomposition of the model, see Section 2.1.3 for the factorization of stationary parametric models.

One further advantage of the proposed approach is that in some cases different kinds of deviations between parameter evolutions are portrayed in different basis vectors. It is thus possible that it is easier to draw inference from the coefficients of these basis vectors than from coefficients of generic basis.

We are currently studying the Bayesian extension of TVARLS and MTVARLS schemes. In the case of MTVARLS this extension means that we compute the projections of the learning set onto the subspace $\Theta(M)$ and approximate these projections with an M–dimensional density. This is taken to approximate the prior density of the coefficients c_k. We also need the likelihood function (the conditional density of observed x_t on c_k) whose estimation is not quite as simple the approximation of the prior density. However, once this is accomplished, the maximum posterior (MAP) estimate for $c_k, k = 1, \ldots, M$ is calculated. This is an extension that enables the use of a relatively large dimension M so that more features can be estimated. We do not elaborate the extension further here.

CHAPTER VI

Blind estimation of optimal MTVARLS subspaces

In most practical cases we do not have enough information on the evolution of the process to be able to construct a learning set. However, if we can assume that the class of predictor evolutions is adequately tight as above, we can use adaptive predictor estimators to approximate the predictor evolution for each realization. We can then calculate the covariance of these evolution estimates and determine the principal subspace from this.

The reason why we use stochastic regression methods to obtain bases for deterministic regression methods rather than use the former estimates directly is that with this approach we aim to estimate differences in a small pseudoensemble. The size of this pseudoensemble is too small to enable the direct application of the principal component analysis to the time-varying covariance. Since there is no (hypothetical) model for the parameter evolution as was the case in Chapter 5, we call this approach *blind estimation* of modified TVARLS (MTVARLS) subspaces.

6.1 Tracking capability of adaptive algorithms

It is well known that adaptive predictors and filters in general can roughly be said to exhibit two kinds of parameter estimation error that are often referred to as *noise and lag misadjustments*. We could also use the notions instantaneous variance and bias of parameter estimates.

The LMS, RLS and KF algorithms were discussed in Section 2.2.3. From (2.13) it is obvious that the gain α_t controls the rate of change of parameter evolution estimates. If the assumed rate of change is fast we should not set α_t too low to allow the algorithm to be able to track the parameter changes, in other words, to keep the bias low. However, increasing α_t will invariably also increase the instantaneous variance of the estimates. Thus the selection of the gain α_t is a trade-off between tracking speed and "noisiness" of the estimates.

In the sequel we concentrate on the forgetting factor RLS algorithm that is one of the most popular adaptive algorithms due to the facts that the criterion is well understood (weighted least squares solution), it is computationally relatively feasible (matrix inversion lemma), it is easily stabilized (by adjusting the forgetting factor to less than unity) and it does not exhibit different modes of convergence

as do some steepest descent type algorithms such as the LMS algorithm. This is due to the approximate orthogonalizing nature of RLS.

6.1.1 Tracking performance of the RLS algorithm

The performance of an adaptive algorithm can be evaluated by many criteria, such as the mean prediction error and the mean parameter deviation. The most straightforward way is to adjust the parameters of the algorithm to minimize $\sum_t e_t^2$, which was the approach in [26] where the unknown covariances needed by the Kalman filter were adjusted to obtain the minimum mean prediction error. In parameter estimation the most relevant criterion, however, is often the mean square deviation of the parameters, that is

$$\min_{\widehat{\theta}} \|\theta^*(t) - \widehat{\theta}(t)\| \ ,$$

where $\theta^*(t)$ is the true parameters evolution when such an evolution can be said to exist or the optimal one when all the statistics were known. In practice the evolution $\theta^*(t)$ is not known.

In some cases it is possible to assume a model (hypermodel) for the evolution of the parameters. There are several results that can be used as aids in the selection of the parameters of several algorithms when the assumed hypermodel is the random walk model. As usual, the more we can make assumption on the situation, the tighter the results are. The results rely often on such assumptions that are not fulfilled. The most important assumption in many results is the orthogonality of the regressors, that is, $\varphi_t^{\mathrm{T}}\varphi_{t-k} = 0$ for all $k \neq 0$. In the adaptive prediction case we have $\varphi_t(k) = \varphi_{t-j}(k+j)$ for $1 \leq k+j \leq p$, $1 \leq k \leq p$, so the correlation between regressors is obvious and in many cases, especially in the small bandwidth case, the correlation "strength" is such that any analyses assuming orthogonality are clearly useless. We are aware of only one study concerning the performance of the RLS algorithm with this special sequential structure of regressors that evolves in the case of adaptive prediction. Even then the process was a deterministic chirp signal in noise [161].

Other common assumptions include M–dependence, (conditional) zero mean difference of the true parameter evolution, boundedness of observation and regressor processes, implicit assumptions on stationarity and technical assumptions that yield proper conditional expectations. We refer here only to [39, 14, 92, 93, 18, 57, 56] where results with some of the above assumptions are given. The results concern typically such estimates as upper bounds for parameter deviation or upper bounds for the mean parameter tracking error. It is also important to note that tracking a constantly varying predictor (or a system in general) is not the same as transient recovery, *i.e.* the case when the process is stable and stationary between non-frequent stepwise changes.

Covariance of parameter deviation, RLS and random walk hypermodel

Let θ_t^* be the true parameter evolution, or equivalently, the optimal predictor evolution. Assume that e_t is a zero mean process with variance $\sigma_e^2(t)$. Further,

assume that the true parameter evolution satisfies the first order hypermodel (random walk)

$$\theta^*(t) = \theta^*(t-1) + \delta w_t \ ,$$

where $\mathrm{cov}(w_t) = \Gamma_w(t)$, $\delta > 0$ and with $E\{e_t w_t\} = 0$. Let $\widetilde{\theta}(t) = \theta^*(t) - \widehat{\theta}(t)$ be the parameter estimation error (misadjustment), where $\widehat{\theta}(t)$ is the parameter estimate obtained from the RLS algorithm with forgetting factor κ. Define further $\Pi_t = E\{\widetilde{\theta}(t)\widetilde{\theta}(t)^{\mathrm{T}}\}$ and $\Psi(t) = E\left\{\varphi_t\varphi_t^{\mathrm{T}}\right\}$. The exact form for Π_t is very complicated and not very tractable. However, there are a sequence of papers with a decreasingly restrictive assumptions and culminating in [92, 93] in which the approximation for small $\mu = 1 - \kappa$ is given:

$$\begin{aligned} \widehat{\Pi}_0 &= \Pi_0 \\ \widehat{\Pi}_t &= \bar{A}_t\widehat{\Pi}_{t-1}\bar{A}_t^{\mathrm{T}} + \mu^2\bar{P}_t\Psi_t\bar{P}_t\sigma_e^2(t) + \delta^2\Gamma_w(t) \ , \end{aligned}$$

where

$$\begin{aligned} \bar{A}_t &= I - \mu\bar{P}_t\Psi_t \\ \bar{P}_t &= \bar{R}_t^{-1} \\ \bar{R}_t &= (1-\mu)\bar{R}_{t-1} + \mu\Psi_t \\ \bar{R}_0 &= R_0 \ . \end{aligned}$$

The coefficientwise error variances are then obtained from the diagonal of $\widehat{\Pi}_t$.

Although it is not very easy to see, the qualitative nature of the result is that decreasing κ will enlarge (not necessarily monotonously) the variance of the parameter estimates. In our case, the assumption of the orthogonality of w_t and $w_{t-k},\ k \neq 0$ is not fulfilled, since the random walk model is not appropriate here. A more feasible model would be a higher order LP-type hypermodel, but a tracking performance analysis for these is not easily tractable. We are not aware of results of the above type in the case of higher order hypermodels.

Evolution of the mean parameter deviation of RLS

Let $\overline{\theta}(t) = E\{\widehat{\theta}(t)\}$ be the expected evolution of the parameter estimates and redefine $\widetilde{\theta}(t) = \theta^*(t) - \overline{\theta}(t)$. The regressors are allowed to be non-orthogonal with exponentially decaying covariance so that such constants $c > 0$ and $0 < \vartheta < 1$ exist that

$$|E\{x_t x_{t-k}\}| < c\vartheta^{|k|} \ .$$

The forgetting factor RLS is the solution to the problem

$$\widehat{\theta}(t) = \arg\min_{\theta(t)} \sum_{k=1}^{t} \kappa^{t-k}(x_t - \theta^{\mathrm{T}}(t)\varphi_t)^2 \ .$$

Let

$$\begin{aligned} \varrho_t &= \sum_{k=1}^{t} \kappa^{t-k} = \frac{\kappa - \kappa^t}{1-\kappa} \\ \widetilde{\kappa} &= \kappa/\varrho_\infty \ . \end{aligned}$$

We have then [178, 179]

$$\overline{\theta}(t) = \varrho_t^{-1} \sum_{k=1}^{t} \kappa^{t-k} \theta^*(k) + \mathcal{O}(1/\varrho_t) \ , \tag{6.1}$$

where $\mathcal{O}(\cdot)$ is the "big-O" function. This yields asymptotically

$$\overline{\theta}(t) \approx \sum_{k=0}^{\infty} \widetilde{\kappa}^k \theta^*(t-k) \ . \tag{6.2}$$

This can also be expressed in the form [55]

$$\widetilde{\theta}(t) \approx \frac{z^{-1}-1}{1-\kappa z^{-1}} \theta^*(t) \ .$$

These forms are in principle just what are needed to be able to optimize the forgetting factor κ when θ^* is known or there is a realistic hypermodel. However, it turns out that the $\mathcal{O}(1/\varrho_t)$ term is of the form

$$\mathcal{O}(1/\varrho_t) = \frac{ca}{\varrho(1-\vartheta^2)} \ ,$$

where a is a function of second and fourth order moments of φ_t. It is clearly seen that in the case of narrow band processes, that is, processes with $\vartheta \lessapprox 1$, the approximation error term can become very large and the approximation loses its usefulness.

This is the situation also in the synchronized state of the ERD/ERS. Preliminary simulations showed that in this case the approximation (6.1) could not be used in the evaluation of the mean tracking error. However, the qualitative nature of the performance of RLS with respect to κ is clearly seen from (6.2). Increasing κ will make the memory of the algorithm longer and the mean tracking error increases. And *vice versa*, letting $\kappa \to 0$ yields $\overline{\theta}(t) \to \theta^*(t)$. These two results indicate clearly the trade-off between the noise and lag misadjustments.

There are very nice results that would allow one to make an optimal choice of κ with respect to the parameter error covariance. Unfortunately the assumptions are such that these results are not applicable in our case. However, the results involve the minimization of functionals that are of the form

$$\mathcal{O}(\mu + \vartheta^2/\mu^b) \ ,$$

where $b = 1/2, 1$ or some other figure depending on the assumptions. It must still be noted that whether the assumptions are valid in a particular situation or not, it might not be relevant to minimize the above criterion. Rather, some ends might emphasize the need for stability instead of tracking speed.

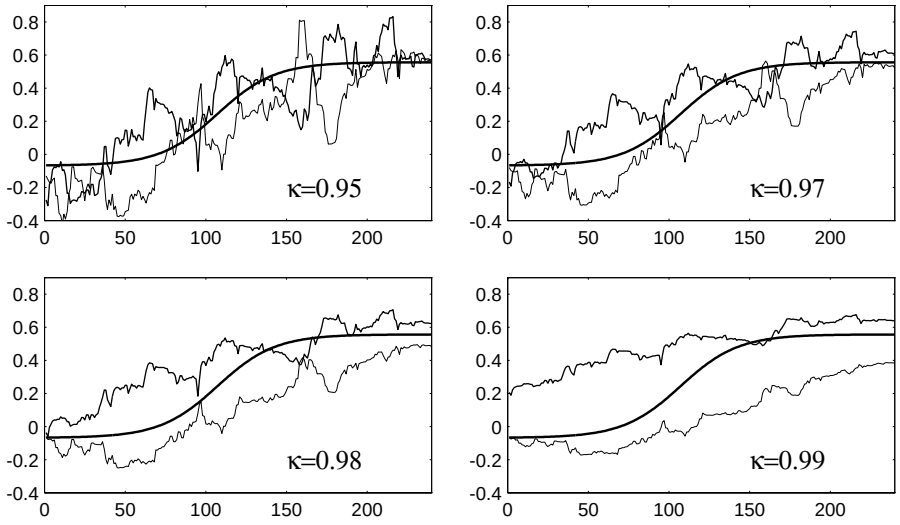


Figure 6.1: Examples of RLS tracking performance with different forgetting factors κ. Bold line: the true evolution $\phi_2(t)$, medium line: backward estimate and weak line: forward estimate.

6.2 A simulation study

We will study the ERS case as follows. We select four forgetting factors $\kappa_1 = 0.95$, $\kappa_2 = 0.97$, $\kappa = 0.98$ and $\kappa_3 = 0.99$. These can be said to correspond to small bias, two compromise and small noise cases, respectively. We also select two sample sizes $N = 1000$ and $N = 20$, that is, the number of realizations that are used to estimate the basis. Of these two the former represents a "large" sample although the obtained eigenspace estimates can not be said to be asymptotical ones. The smaller sample is of interest to us since it is approximately the size that one has access to in the case of real ERD/ERS data.

The parameter evolutions corresponding to the realizations are drawn randomly from the learning set that was constructed in Section 5.2.2. The forward parameter evolution estimates $\widehat{\Phi}^{\mathrm{f}}(t)$ are then computed with RLS and the selected forgetting factors κ for each realization. To reduce the delay of the estimates we form also the backward estimates $\widehat{\Phi}^{\mathrm{b}}(t)$ and use the averages $\widehat{\Phi}(t) = 1/2(\widehat{\Phi}^{\mathrm{f}}(t) + \widehat{\Phi}^{\mathrm{b}}(t))$ as evolution estimates. This trick will reduce the delays of the estimates but does not reduce the "smeariness" of these. To obtain initial state estimates with correct distribution, the tracking was initialized at a time well before the investigated epoch. In Fig. 6.1 typical RLS estimates $\widehat{\Phi}^{\mathrm{f}}(t)$ and $\widehat{\Phi}^{\mathrm{b}}(t)$ together with the true evolution are shown.

As in Chapter 5, the evolution estimates are concatenated to form $\widehat{\Psi}$ and the principal subspaces of the corresponding covariances $\widehat{\Gamma}$ are computed.

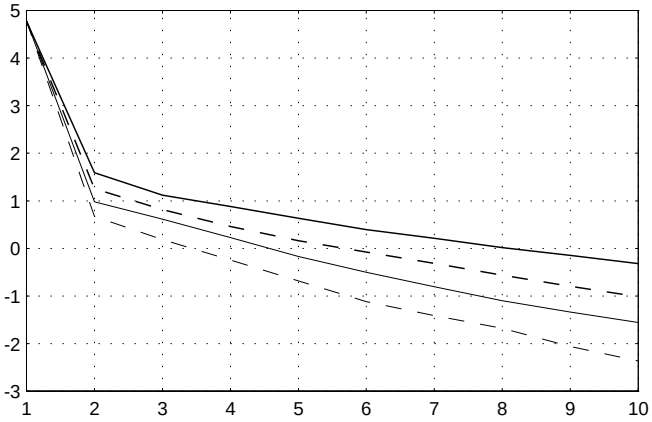


Figure 6.2: The eigenvalue convergence (decay) of the RLS basis estimates, from top to bottom: $\kappa = 0.95$ (solid bold), $\kappa = 0.97$ (dashed bold), $\kappa = 0.98$ (solid weak) and $\kappa = 0.99$ (dashed weak). The vertical scale is logarithmic and the figures denote powers of 10.

6.2.1 Convergence of the principal eigenspaces

It can be conjectured that since the use of a small forgetting factor yields a small bias, the principal eigenvectors $\widehat{v}_m$ corresponding to the covariances of the estimates $\widehat{\Psi}$ would asymptotically converge to somewhere in the vicinity of the true eigenvectors v_m. Also, since the noisiness of the estimates increases with decreasing κ, it can be expected that the eigenvalues would decay faster with large κ than with small κ. It is clear that the required number of realizations would then be very large and it became clear that $N = 1000$ is not enough.

The decays of eigenvalues are shown in Fig 6.2. It is clearly seen that the larger the forgetting factor, the faster the eigenvalue decay is. The fast eigenvalue convergence in the case $\kappa = 0.99$ should not be regarded as an asset in this case. While fast decay of the eigenvalues implies that the corresponding learning set can be approximated with a smaller mean square error for a given subspace dimension and is generally a desired property (fast convergence of the principal subspace), in this case this is due to the lower approximability of the true evolutions in this subspace. The mean eigenvalue decay for sample size $N = 20$ was also calculated. The mean convergence in this case was the same as that for the size $N = 1000$ sample.

The principal eigenvectors exhibit such characteristics as expected. The eigenvectors corresponding to the high κ cases are smoother than the others in both the senses that they are less noisy and the evolutions are slower. The true eigenvectors $v_k(t)$ and the estimates $\widehat{v}_k(\kappa)$ with $k = 1, 3$ and $\kappa = 0.97, 0.99$ are shown in Fig. 6.3 for the case $N = 1000$ and examples of these for the case $N = 20$ in Fig. 6.4. The stability of the estimates that are based on a small sample only

are of importance when the simulations are used to select parameters for the true (measured) ERS data.

According to the simulations the forgetting factor $\kappa = 0.97$ seemed to be a feasible compromise between tracking speed and noisiness. Also the stability of the eigenvector estimates seemed to be adequate with respect to the Standard deviation intervals with size $N = 20$ samples.

6.2.2 Modified TVARLS estimates

There are two possibilities to evaluate the performance of a selected combination (κ, M) that determines a particular basis. One is to calculate the mean prediction error variances for each selection over a large sample. The problem here, however, is that due to the nature of the LS projection, the error will be monotonously decreasing with respect to M and it would seem that increasing M will ever yield better parameter estimates for the processes. This is clearly incorrect since the estimates will only be more "local" and describe the realization instead of the process. With the AR model order p the standard approach would be to employ order selection principles such as Akaike's information criterion (AIC), finel prediction error (FPE), minimum description length (MDL) etc. [40], but these methods are not directly applicable to the selection of M. See [142] for a general discussion of model order selection in the general time-varying case.

The other problem is the sensitivity of the prediction error variance with respect to variations in the estimated parameter processes and the coefficients of the basis vectors. This makes it possible that the prediction error variance may well be smaller in a case where the parameter deviance is larger. Furthermore, in most cases even the mean paramater deviation is not a feasible criterion that should be dictated by the specific application itself. Although in Section 6.3 we calculate and show some approximate spectral factorizations, we do not explicitly aim to any specific application (further use) of the estimated parameter evolutions. For this reason we do not calculate any performance/merit statistics for the estimates with different (κ, M) and concentrate on the raw evolution estimates instead.

In Fig. 6.5 we show a realization corresponding to a member Ψ^* of the learning set. In the sequel we will study the forgetting factors $\kappa = 0.97$ and $\kappa = 0.99$ that are taken to represent the small bias and small noise cases, respectively. All subsequent estimates calculated and shown in this section are calculated for this data.

To help in the visual evaluation of the estimates we show the following evolution estimates together with the true evolution Ψ^*.

1. Projection of Ψ^* onto the span$\{v_1, \ldots, v_M\}$. This is the best possible estimate in the mean square sense in the optimal (correct) subspace.

2. Projection of Ψ^* onto the span$\{\widehat{v}_1(\kappa), \ldots, \widehat{v}_M(\kappa)\}$. This is the best possible estimate in the mean square sense in the blind estimate for the principal subspace.

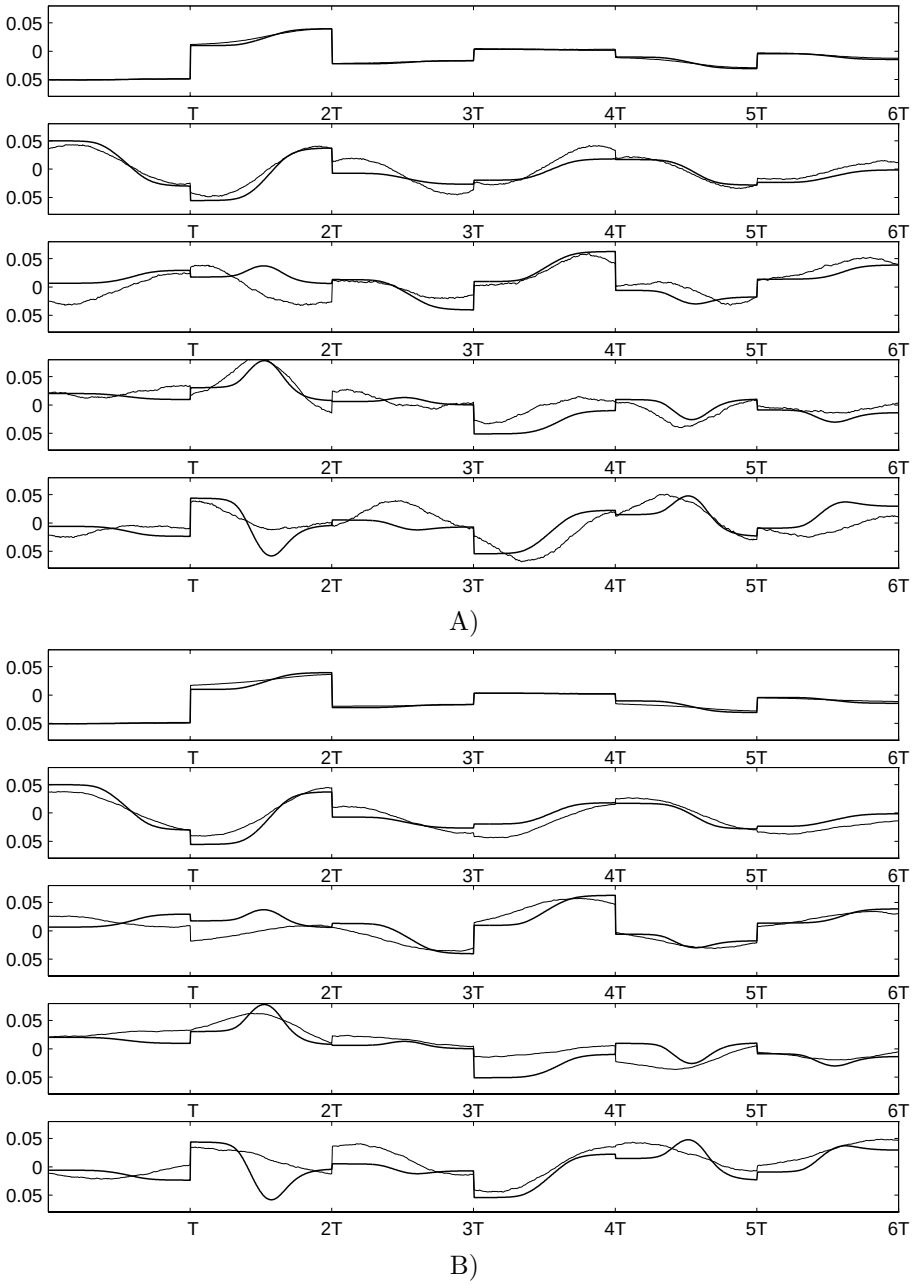


Figure 6.3: The eigenvectors (weak lines) corresponding to the five largest eigenvalues of the RLS covariance matrices with A) $\kappa = 0.97$ and B) $\kappa = 0.99$. In both sets the bold lines are the eigenvectors of the learning set. Number of realizations is $N = 1000$.

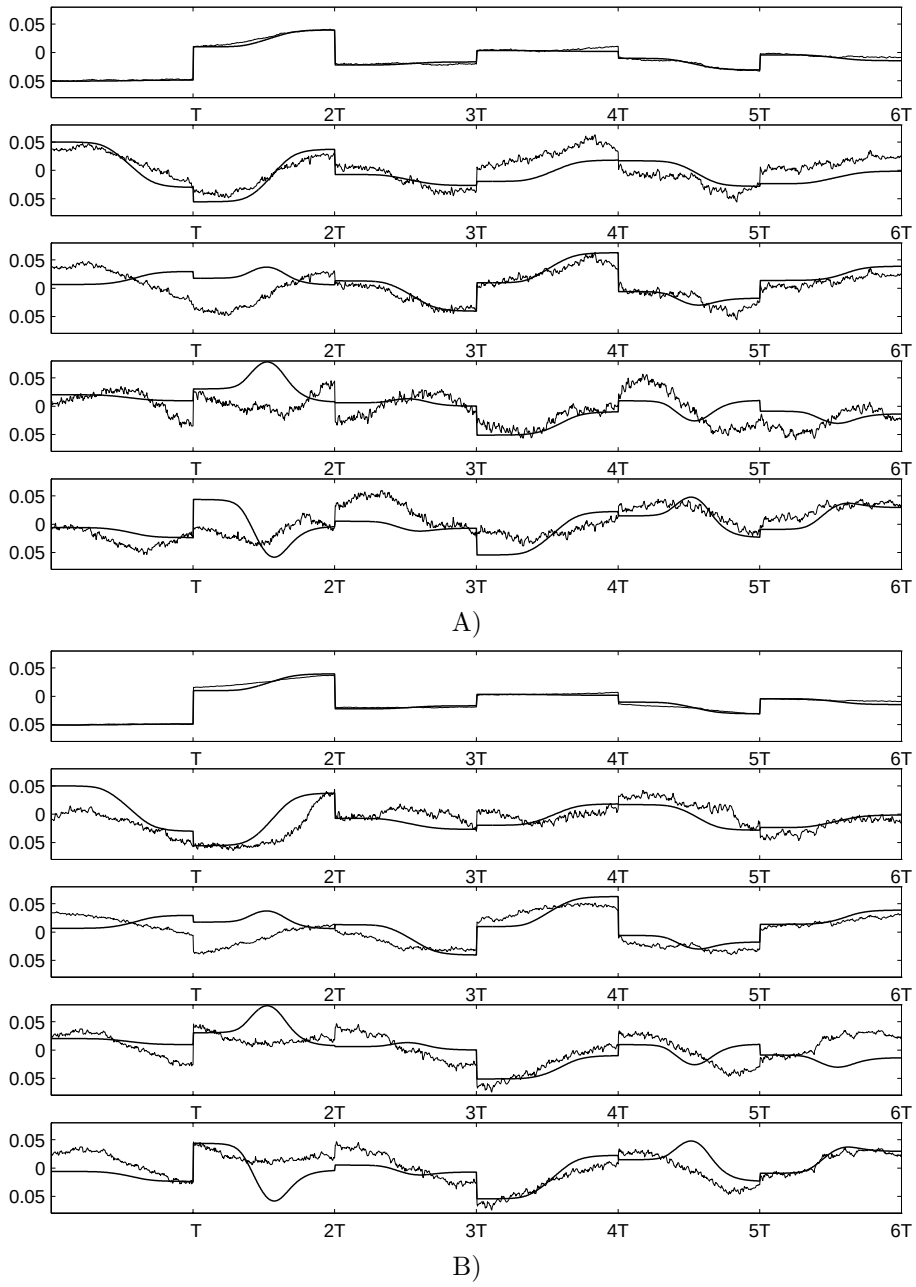


Figure 6.4: As in Fig. 6.3 but the number of realizations is 20.

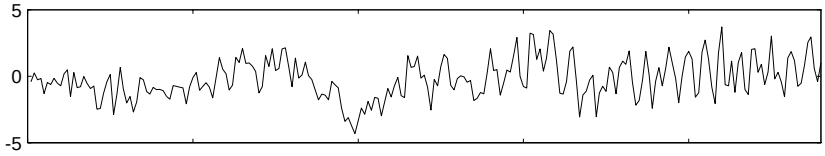


Figure 6.5: A realization corresponding to an evolution of the learning set (hypothetical basis) for which the estimates shown in Figs. 6.6–6.13 are calculated.

3. MTVARLS estimate with the optimal basis $\{v_1, \ldots, v_M\}$. This depends on the realization for Ψ^*.

4. MTVARLS estimate with the blind estimate $\{\widehat{v}_1(\kappa), \ldots, \widehat{v}_M(\kappa)\}$. Also this depends on the realization for Ψ^*.

In Figs. 6.6 and 6.7 these estimates are shown in the case $\kappa = 0.97$ for $M = 1$ and $M = 3$, respectively. In Figs. 6.8 and 6.9 these estimates are shown in the case $\kappa = 0.99$ for $M = 1$ and $M = 3$, respectively.

It must be noted that the estimates shown in Figs. 6.6–6.9 are only examples corresponding to the chosen Ψ^* and a single realization. A large number of estimates was inspected visually for different Ψ^* and $M = 1, \ldots, 5$. It seemed that overall the dimension $M = 3$ should not be exceeded in either of the blind estimates of the basis. On the other hand, the mean difference between approximability (projections) of the evolutions between $M = 2$ and $M = 3$ suggests that we should use $M = 3$ in this case ($N = 1000$). We do not show the estimates with too large M here but refer to Fig. 6.20 where the results of the use of too large an M are clearly visible.

With respect to the forgetting factor in this case, $\kappa = 0.97$ is clearly superior to $\kappa = 0.99$ which is due to the fact that the approximate memory of RLS in the case $\kappa = 0.99$ is 100. This is clearly too large, see Fig. 5.2a. In Figs. 6.10–6.13 we show the projections of Ψ^* onto the blind estimate $\mathrm{span}\{\widehat{v}_1(\kappa), \ldots, \widehat{v}_M(\kappa)\}$ and the MTVARLS estimates with this basis in the case $N = 20$ for $M = 1, 3$ and $\kappa = 0.97, 0.99$.

6.2.3 Overview of the blind estimates

For a pseudoensemble of this nature even a sample size of $N = 1000$ seems to be too small to extract the true eigenvectors except for the first few ones. It is also evident that if the forgetting factor is diminished to reduce the serial correlations of the evolution estimates, the required sample size will easily become too large to be realizable in any circumstances.

The extraction of the particular eigenvectors is, however, not a necessity. More important subjects are the approximability of the true evolutions in a subspace of dimension M and the sensitivity of the MTVARLS scheme with the corresponding

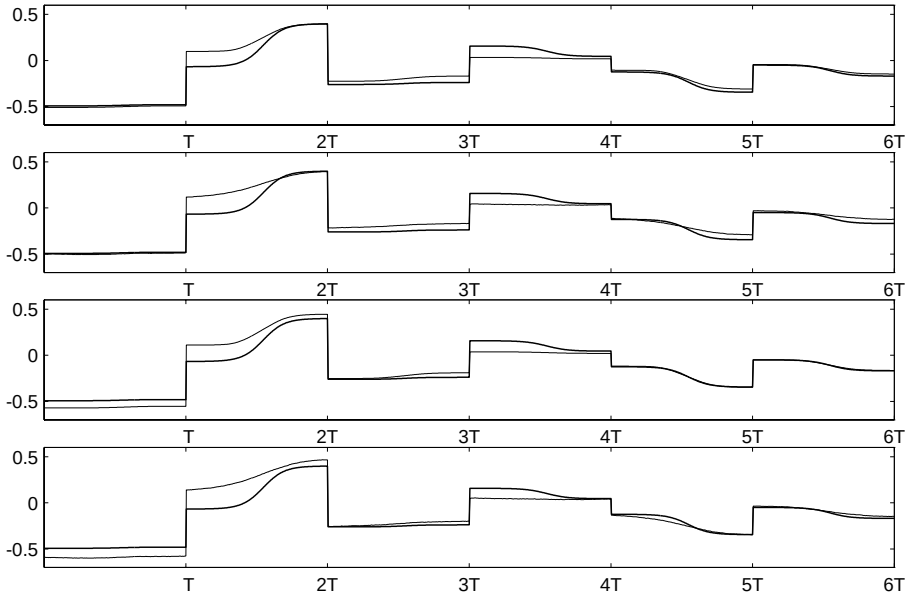


Figure 6.6: True parameter evolution (bold line in all figures) corresponding to the realization shown in Fig. 6.5. The weak lines correspond to (from top to bottom) projection onto optimal subspace, projection onto RLS subspace, MTVARLS estimate with optimal basis and MTVARLS estimate with RLS basis. The forgetting factor is $\kappa = 0.97$, dimension of subspace $M = 1$ and the bases are calculated from a sample of size $N = 1000$.

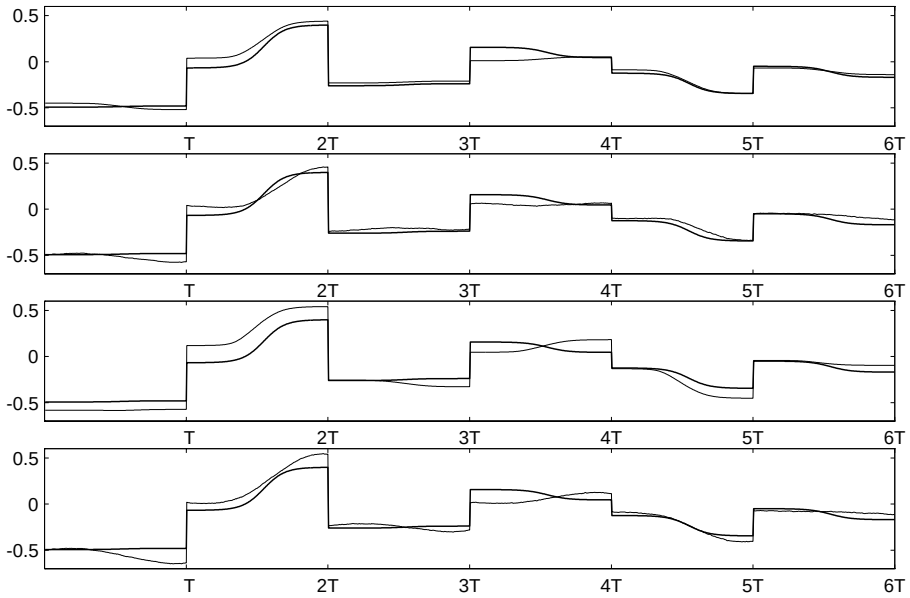


Figure 6.7: As in Fig. 6.6 but with the dimension of subspace $M = 3$.

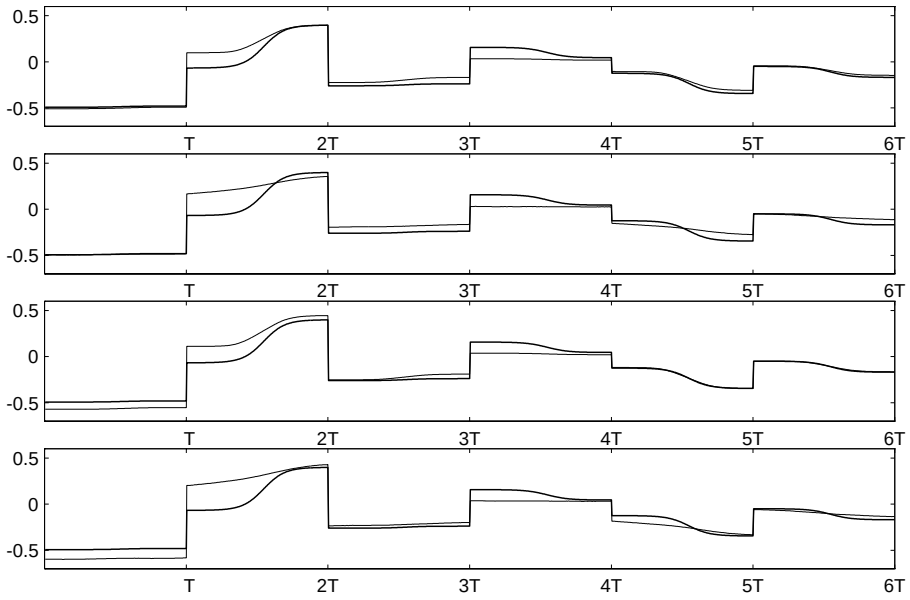


Figure 6.8: As in Fig. 6.6 but the forgetting factor is $\kappa = 0.99$ and the dimension of subspace $M = 1$.

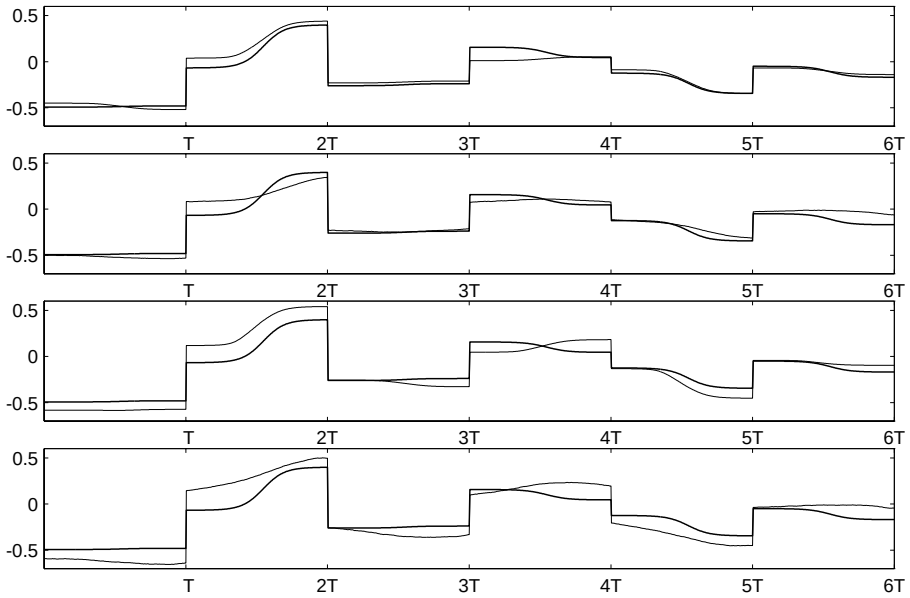


Figure 6.9: As in Fig. 6.8 but with the dimension of subspace $M = 3$.

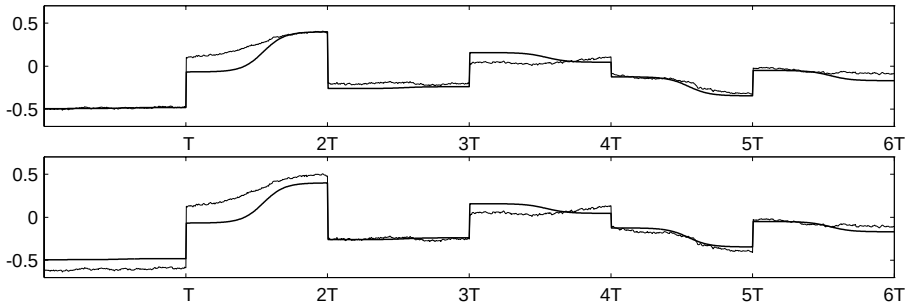


Figure 6.10: True parameter evolution (bold line in both figures) with (weak lines) projection (top) and MTVARLS estimate (bottom); forgetting factor $\kappa = 0.97$, dimension of subspace $M = 1$ and bases calculated from a sample of size 20.

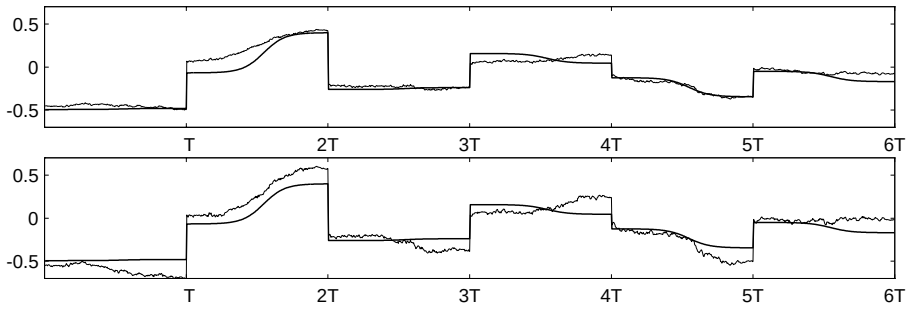


Figure 6.11: As in Fig. 6.10 but with dimension of subspace $M = 3$.

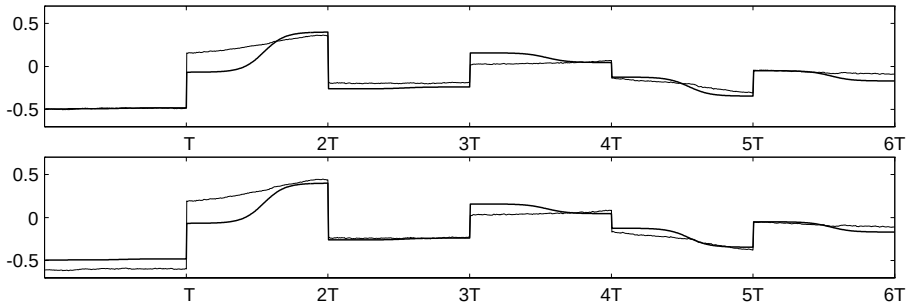


Figure 6.12: As in Fig. 6.10 but the forgetting factor is $\kappa = 0.99$.

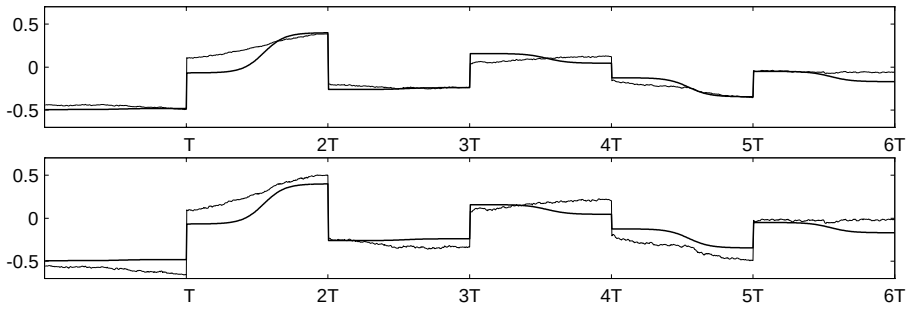


Figure 6.13: As in Fig. 6.12 but with dimension of subspace $M = 3$.

basis. The first subject can be evaluated by computing the projections of the members Ψ of the learning set onto $\mathrm{span}\{v_1, \ldots, v_M\}$ and $\mathrm{span}\{\widehat{v}_1(\kappa), \ldots, \widehat{v}_M(\kappa)\}$ and the corresponding error norms. The ratio of these two can be taken as a measure of the approximating capability of the blind estimate for the basis.

For a principal subspace of dimension M the mean projection error can be computed as

$$E_{\mathrm{opt}}(M) = N^{-1} \sum_{k=1}^{N} \|\Psi_k\|^2 - \sum_{k=1}^{M} \lambda_k \ .$$

For the corresponding mean error $E_{\mathrm{blind}}(M, \kappa)$ with the basis $\{\widehat{v}_k(\kappa)\}$ we have to compute the projections $P_{\Theta(M,\kappa)}\Psi$ directly. The results in the case $\kappa = 0.97$ are shown in Table 6.1. The ratios of the errors support the choice $M = 3$.

TABLE 6.1

The mean projection errors of the learning set onto the optimal subspace ($E_{\mathrm{Opt}}(M)$), the corresponding blind estimates ($E_{\mathrm{blind}}(M, \kappa = 0.97)$) and the ratios of these in the case $\kappa = 0.97$ for different subspace dimensions M. The sample size is $N = 1000$.

M	$E_{\mathrm{Opt}}(M)$	$E_{\mathrm{blind}}(M, \kappa = 0.97)$	$E_{\mathrm{Opt}}(M)/E_{\mathrm{blind}}(M, \kappa = 0.97)$
1	8.3635	7.7483	1.0794
2	6.0667	5.2810	1.1488
3	4.6463	3.7529	1.2381
4	3.8961	2.3606	1.6505
5	2.4316	1.5310	1.5882

The usual method to evaluate the sensitivity of LS solutions is to compute condition numbers for the corresponding regressor matrices. In this case this kind of a comparison does not yield any significant results since the basis vectors are orthonormal in both cases. Instead, we calculate the generalized angles between the subspaces and the corresponding distance measures [87]. The distance between two subspaces $\mathcal{S}_1$ and $\mathcal{S}_2$ is defined as

$$\mathrm{dist}(\mathcal{S}_1, \mathcal{S}_2) = \|P_{\mathcal{S}_1} - P_{\mathcal{S}_2}\| \ ,$$

where $P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_2}$ are orthogonal projectors onto the subspaces and the norm indicates the operator 2-norm. With the aid of the CS decomposition it can be shown that

$$\mathrm{dist}(\mathcal{S}_1, \mathcal{S}_2) = \left(1 - \sigma^2_{\min}(S_1^{\mathrm{T}} S_2)\right)^{1/2} \ ,$$

where S_1 and S_2 are orthonormal bases of the corresponding subspaces and $\sigma_{\min}$ denotes the smallest singular value. The angle $\xi(\mathcal{S}_1, \mathcal{S}_2)$ between the subspaces is then

$$\xi(\mathcal{S}_1, \mathcal{S}_2) = \arcsin \mathrm{dist}(\mathcal{S}_1, \mathcal{S}_2)$$

and is $\pi/2$ (90°) at maximum. If $\xi(\mathcal{S}_1, \mathcal{S}_2) = \pi/2$ this means that there is a nonzero vector in $\mathcal{S}_1$ whose projection onto $\mathcal{S}_2$ is zero. In a way the angle between

subspaces is thus a worst case measure. It turns out that the angles are relatively large. These are shown in Table 6.2.

TABLE 6.2

Generalized angles (in degrees) between the true subspaces and the corresponding blind estimates for different forgetting factors κ and subspace dimensions M. The sample size is $N = 1000$.

κ	$M=1$	$M=2$	$M=3$	$M=4$	$M=5$
0.95	3.8617	26.0467	52.9268	76.6944	60.9379
0.97	4.0563	24.8252	61.1358	66.6164	39.0881
0.98	5.2849	49.2113	66.9481	24.8129	41.0861
0.99	7.8150	79.6824	74.0150	32.7801	41.2169

A more relevant measure would perhaps be the mean distance between the subspaces spanned by the columns of the corresponding MTVARLS regressor matrices. The evaluation of this measure would, however, require a large number of realizations for each Ψ since the differences induced into the regressor matrices are for the most part dependent on the realizations.

6.3 The blind estimates from ERS data

In this section we compute the blind estimates for the basis from ERS data. We use the same ERS data as in the forming of the learning set in the hypothetical evolution case. The data consists of $N = 20$ ERS observations and is shown in Fig. 6.14.

We use the forgetting factor $\kappa = 0.97$ and proceed as in the previous section. The first three eigenvectors are shown in Fig. 6.15. These should be compared to Fig. 6.4A.

The most important difference between these is that with ERS data the first eigenvector (that is almost equal to the mean evolution estimate in this case) shows considerably less any trend in the separate parameter evolutions than with the simulated data. The other clearly visible difference is that the structure of the eigenvectors seems to disappear faster than with the simulations (the fourth eigenvector corresponding to the ERS data can be described as pure noise). This can mean that the differences between the RLS estimates in the ERS case are more irregular or that the differences are smaller. The latter case can be understood if we consider an ensemble consisting of noisy observations of a single (deterministic) evolution. Then the first eigenvector would resemble this evolution and the other eigenvectors could be described as noise with no correlation structure. The other point is that in this case the decay of the eigenvalues will be very slow. It is also possible that the irregularity is partly due to the inadequateness of the 6:th order predictor model.

The mean of $\widehat{v}_k(t)$ also displays slight trend on the interval just before the trigger which could be taken as a waiting state for the event.

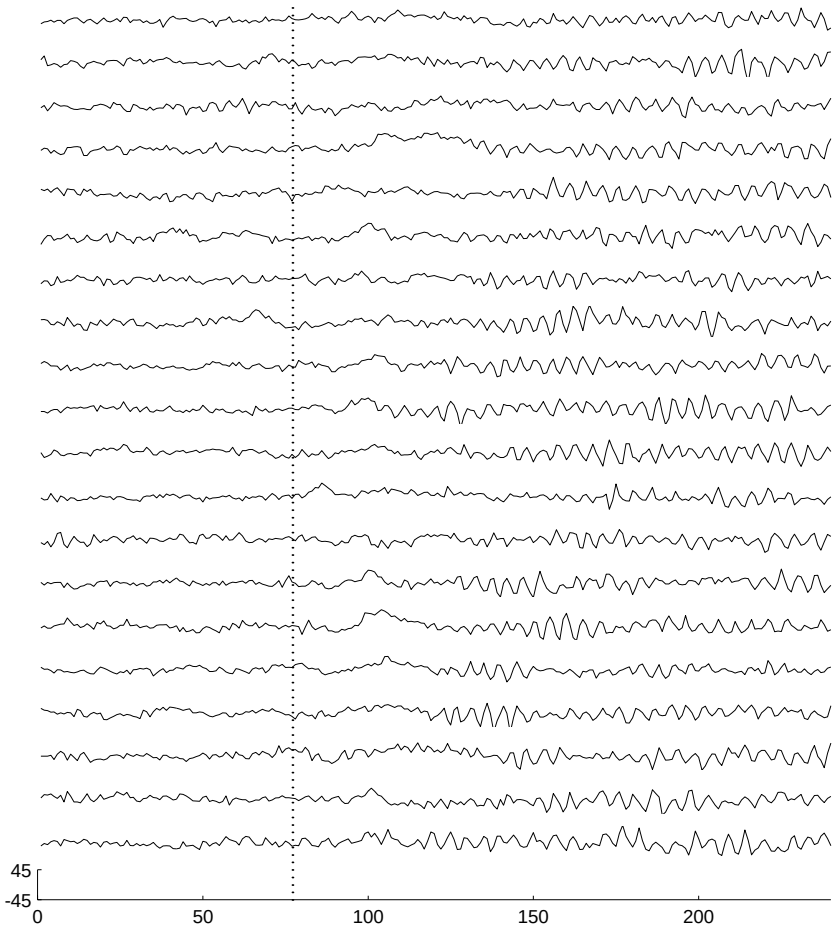


Figure 6.14: The 20 ERS observations for which the RLS estimates are computed. The trigger is at time $t = 80$ and is shown with the dotted line. The vertical scale is in microvolts.

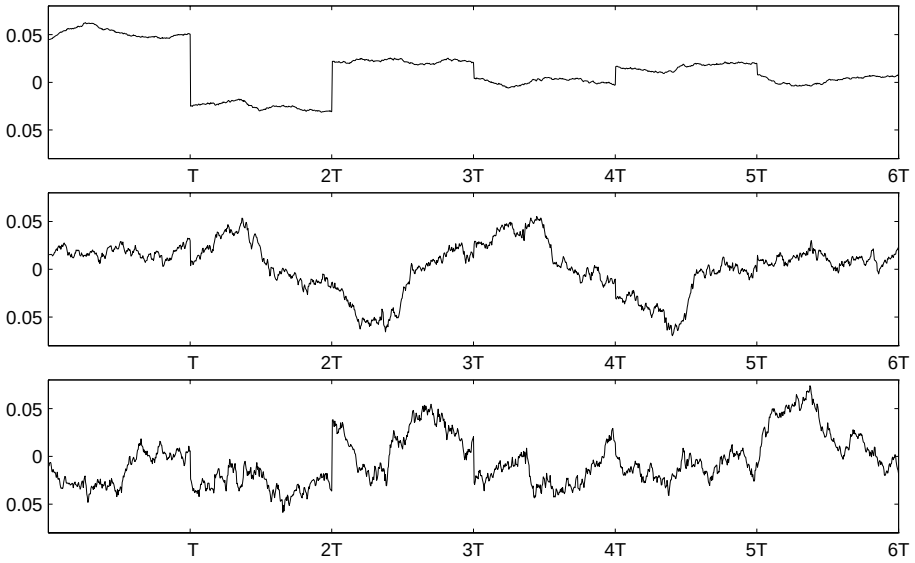


Figure 6.15: The first three eigenvectors corresponding to the covariance of the RLS estimates with $\kappa = 0.97$ and the ERS data shown in Fig. 6.14.

Direct estimation of the mean predictor evolution

We calculated also the mean predictor evolution of order $p = 6$ for the TV covariance of the (pseudo)ensemble x_t^k, $k = 1, \ldots, 20$ directly from the nonstationary covariance estimate as explained in Section 2.2.1. If data were from an ensemble this estimate would asymptotically give the true optimal predictor evolution. In the case of a pseudoensemble we can only obtain a mean predictor evolution. We can, however, assume that since the first eigenvectors should (in this case) closely resemble the mean evolution, we can compare these three estimates.

Since for each time t this is a $p = 6$ parameter estimate based on $N = 20$ observations, the estimates would have a large variance even if the data would form a true ensemble. For this reason we smooth each parameter evolution according to the smoothness priors assumption [76]. These evolutions are shown in Fig. 6.16.

The smoothness priors estimate is computed as follows. Let $\widehat{\Phi}_k$ be the TV covariance estimate of the k:th parameter evolution. The smoothness priors estimate $\widehat{\Phi}_k^{\mathrm{SP}}$ is the solution to the problem

$$\min \left(\|\widehat{\Phi}_k - \widehat{\Phi}_k^{\mathrm{SP}}\|^2 + \alpha^2 \|L\widehat{\Phi}_k^{\mathrm{SP}}\|^2 \right) , \tag{6.3}$$

where L is usually the matrix corresponding to either the second or third order difference operator and α is a constant that controls the trade-off between deviation

and smoothness. To minimize the functional (6.3) we write

$$
\begin{aligned}
\widetilde{\Phi}_k &= (\widehat{\phi}_k(1), \ldots, \widehat{\phi}_k(T), \underbrace{0, \ldots, 0}_{T-3})^{\mathrm{T}} \\
\Phi_k^{\mathrm{SP}} &= (\phi_k^{\mathrm{SP}}(1), \ldots, \phi_k^{\mathrm{SP}}(T))^{\mathrm{T}} \\
E &= (e_1, \ldots, e_{2T-3})^{\mathrm{T}} \\
\widetilde{L} &= \begin{pmatrix}
1 & & & & & \\
& \ddots & & & & \\
& & & & & \\
& & & & \ddots & \\
& & & & & 1 \\
-1 & 3 & -3 & 1 & & \\
& \ddots & & & \ddots & \\
& & -1 & 3 & -3 & 1
\end{pmatrix} \\
W &= \operatorname{diag}\,(\underbrace{T^{-1}, \ldots, T^{-1}}_{T}, \underbrace{\alpha^2, \ldots, \alpha^2}_{T-3})^{\mathrm{T}} \; .
\end{aligned}
$$

We can now restate the problem as a weighted LS problem

$$
\begin{aligned}
&\min E^{\mathrm{T}} W E, \quad \text{where} \\
&\widetilde{\Phi}_k = \widetilde{L}\Phi_k^{\mathrm{SP}} + E \; .
\end{aligned}
$$

The formal solution is

$$
\Phi_k^{\mathrm{SP}} = (\widetilde{L}^{\mathrm{T}} W \widetilde{L})^{-1} \widetilde{L}^{\mathrm{T}} W \widetilde{\Phi}_k \; ,
$$

although it is numerically more stable to compute the conventional LS solution of the modified problem

$$
W^{1/2}\widetilde{\Phi}_k = W^{1/2}\widetilde{L}\Phi_k^{\mathrm{SP}} + W^{1/2}E
$$

with *e.g.* some orthogonalization scheme such as QR decomposition, where $W^{1/2}$ is the Cholesky factor of W, in this case $W^{1/2} = \operatorname{diag}\,(T^{-1/2}, \ldots, T^{-1/2}, \alpha, \ldots, \alpha)^{\mathrm{T}}$.

Also other TV covariance smoothing methods could be used. The most common ones are sliding window estimates, see [133, 201, 200]. The smoothness priors estimate, however, does not necessarily slow down the evolutions at all if the matrix L corresponds to at least the second difference operator. For this reason the smoothness priors estimate can be argued to suit better to the case at hand.

In Fig. 6.17 we show the predictor evolutions corresponding to the first eigenvector of the hypothetical evolution, the smoothed TV covariance estimate and the first eigenvector of the blind estimate. As can be seen, the TV covariance follows more closely the hypothetical evolution than the blind estimate with respect to the parameters ϕ_2 and ϕ_5. In the case of parameters ϕ_3 and ϕ_6 the situation is reversed. Most notable is the behavior of the RLS estimate for the parameter

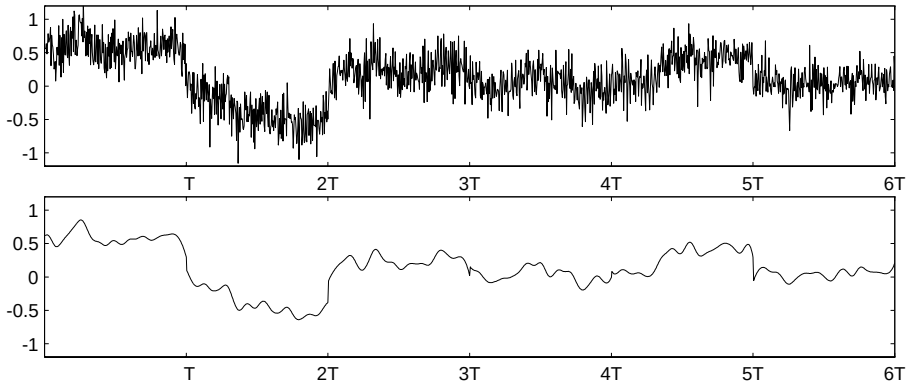


Figure 6.16: Top: the predictor evolution corresponding to the TV covariance estimate. Bottom: The corresponding smoothed evolution.

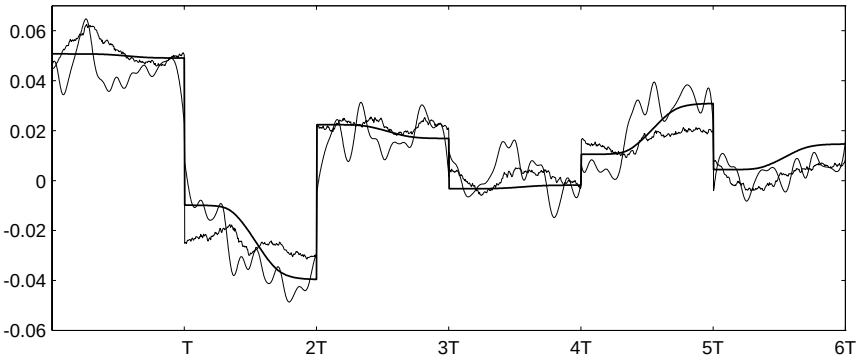


Figure 6.17: The time varying predictors (normalized) corresponding to the first eigenvector of the hypothetical evolutions (bold), first eigenvector of the RLS estimates (medium) and the TV covariance estimate (weak).

ϕ_2 that has the greatest effect on the location of the model pole that corresponds to the peak at alpha band and therefore also to the bandwidth of the process. The change in this parameter is much smaller than in the hypothetical and TV covariance cases, the latter two of which seem to follow each other well. This is a demonstration of the complex serial and inter-parameter correlations when an adaptive predictor is used. The phenomenon can be partly explained by the common behavior of the adaptive filters that the tracking speed is not symmetrical with respect to the transitions from large process (prediction error) variance epochs to small variance epochs and vice versa.

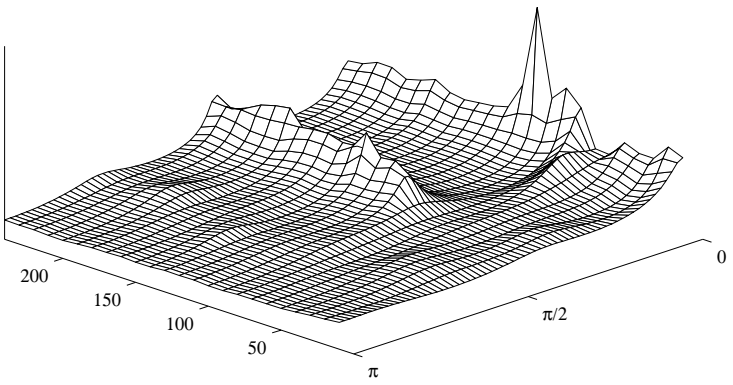


Figure 6.18: The TV (amplitude) spectrum estimate corresponding to the TV covariance estimate. Vertical scale is arbitrary.

The time-varying spectrum estimates

The modified TV spectrum estimates for the smoothed TV covariance estimate and the evolution corresponding to the first eigenvector of the blind estimate, that is, $\sqrt{\lambda_1}\widehat{v}_1(\kappa)$ are shown in Figs. 6.18 and 6.19, respectively. The notion "modified" is to be taken in the sense of [164], that is, the prediction error variance $\sigma_e^2(t)$ is set to unity for all t. The reason for this here is that the estimation of the time-varying process (prediction error) variance is a very difficult problem that has no general solution. We do not attempt to solve this problem here. Several *ad hoc* methods evolve immediately, such as LP filtering or smoothness priors estimate of the mean squared prediction errors $N^{-1}\sum_k (e_t^k)^2$. However, it turns out that in this case it is very difficult to make these methods robust. It is possible that the projection of the mean squared prediction errors onto an appropriately selected subspace could be a feasible approach in this case. This is, however, no solution to this problem in general.

The main problem with AR model based TV spectrum estimation can clearly be seen from these figures and Fig. 6.20. This is the sensitivity of the spectra to small changes in the modulus of the model roots near the unit circle. Nevertheless, at least the qualitative nature of the evolution of the alpha activity is clearly visible in Figs. 6.18 and 6.19 although the latter contains more alpha activity on the pre-stimulus period. The effect of the small upward swing in $\widehat{v}_1$ corresponding to $\phi_2(t)$ that can be seen in Fig. 6.17 at about the time of the trigger ($t = 80$) is clearly seen Fig. 6.19 as a downward swing in the alpha activity.

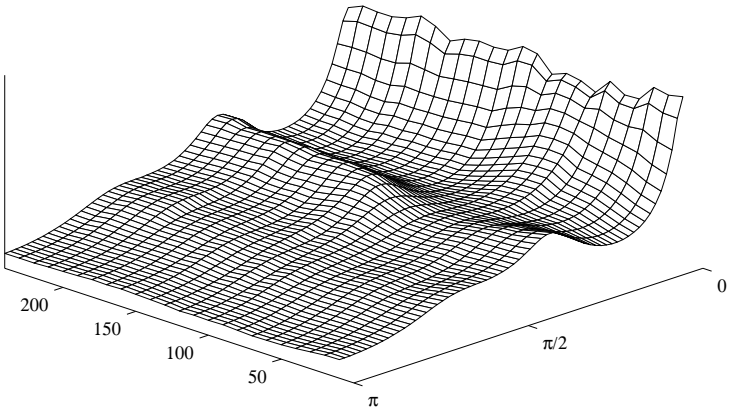


Figure 6.19: The TV (amplitude) spectrum estimate corresponding to the first eigenvector of the RLS covariance estimate. Vertical scale is arbitrary.

The estimation of the alpha activity

It is clear that the blind estimates that are based on a sample of size $N = 20$ are not suitable for the estimation of the differences between the evolutions, and therefore not for evolution of the alpha activity. As an example we show in Fig. 6.20 an observation together with the alpha component activities (STD's) corresponding to the TV covariance estimate and MTVARLS estimates with Blind estimates for the basis for $M = 1, \ldots, 4$. It is clearly seen that the case $M = 2$ is no different from the case $M = 1$ and that the cases $M = 3, 4$ are clearly meaningless. In Fig. 6.18 we show the approximate alpha component spectrum corresponding to the TV covariance estimate.

6.4 Discussion

We have discussed the possibility of using adaptive predictors (filters) in the aid in estimating basis functions for the modified TVARLS scheme. To this end we evaluated the ability of the RLS algoritm to extract the true eigenspace structure using simulations. This was done in the case of a relatively large sample size ($N = 1000$) and a small sample size ($N = 20$) that is relevant in the case of the application of the method to the estimation of ERD/ERS dynamics.

It was found that even in the large sample case the forgetting factor RLS algorithm was able to extract only first few eigenvectors. This was due to (at least) the following reasons.

1. Although the uniform distribution of the hypothetical evolutions does not induce uniformity of the distribution of the parameter estimates, the sample size $N = 1000$ seems to be too small.

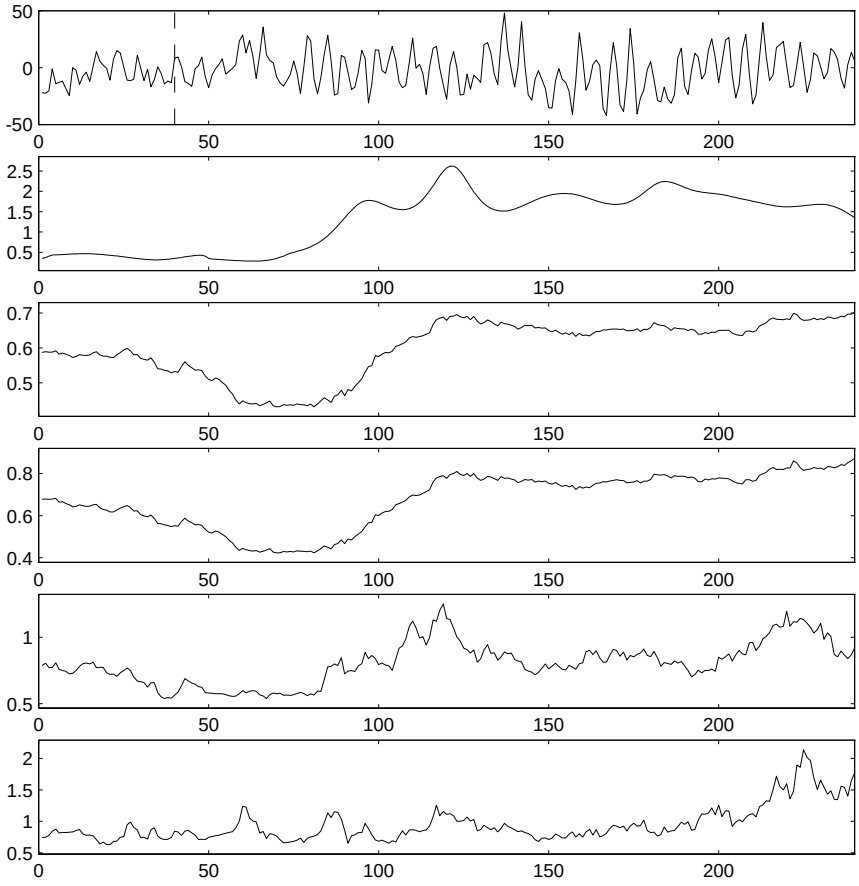


Figure 6.20: Top (1st): An observation (ERS), 2nd: The STD (square root of power) evolution corresponding to the TV covariance estimate, 3rd-6th: the MTVARLS estimates of the ERS with $M = 1, \ldots, 4$, respectively. Vertical scale is arbitrary.

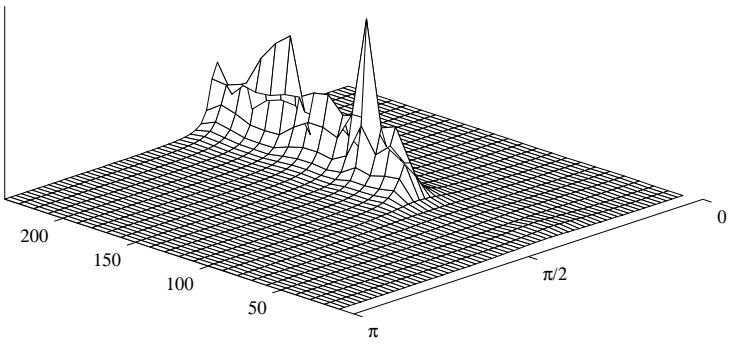


Figure 6.21: The TV (amplitude) spectrum estimate corresponding to the alpha band factor obtained from the approximate spectral factorization of the (smoothed) TV covariance estimate

2. The maximum rate of evolution seems to necessitate the use of a forgetting factor $\kappa = 0.95$ or smaller. This would in turn require an even larger number of realizations since the realizations-wise variance of estimates would become larger.

It must be remembered that one of the main motivations for this study was that the transitions could not be assumed either to be abrupt or slow in the classical sense.

Comparison of eigenvalue decays

It is also interesting to compare the eigenvalue decay rates of the hypothetical case (learning set), the blind estimates of the simulations and the ERS data. These are shown in Table 6.3. The fact that the decay rate in the case of the simulations is slower than with the optimal subspaces is clear since the adaptive predictor estimates have to add to the variability of the original evolutions if these evolutions are to be estimated in the first place. This would not be the case when $\kappa \to 1$ in which case all individual predictor estimates would tend to constant evolutions.

More notable is the structure of the eigenvalue decay in the case of blind estimates of ERS data. If we take the simulation case to decay slowly but steadily, in the ERS case λ_2 is much larger than with the simulations and there is a pronounced step between λ_2 and λ_3 after which the decay is slow. The eigenvalues $\lambda_k, k = 3, \ldots, 5$ are, however, considerably smaller than in the case of simulations. There are many possible explanations for this phenomenon but these can not be

verified without knowledge on the distributions of the original data and the blind estimates.

However, it can be said that the phenomenon relates to the variability of the evolution estimates. Thus the evolutions of the ERS data can be approximated with subspaces of smaller dimension than the simulation estimates. Since the increase in the variability due to the RLS estimates is the same in both cases, it can be said that the variability in the hypothetical pseudoensemble is larger than in true ERS data. Since the initial and final states were estimated from the same data at least the following explanations are possible: the initial and final state distribution estimates were too wide, the transitions and states were correlated in ERS data, the hypothetical transition dynamics was too complex and/or the true differences in the optimal predictor evolutions of ERS data are of more random nature than the hypothetical evolutions. These possibilities could be investigated if enough data could be collected to obtain very accurate estimates for, say, $M = 10$ largest eigenvalues and the corresponding eigenvectors.

TABLE 6.3

The eigenvalue decays corresponding to the optimal hypothetical subspace, the corresponding simulations and the ERS estimates. The figures are scaled so that $\lambda_1 = 1$ in all three cases.

Case	λ_2	λ_3	λ_4	λ_5
Optimal	0.0204	0.0126	0.0115	0.0069
Simulations, $\kappa = 0.97$	0.0287	0.0189	0.0178	0.0154
ERS, $\kappa = 0.97$	0.0349	0.0141	0.0122	0.0087

Separate basis estimates

It is possible that the separation of the estimation of the basis for different parameters could be a feasible alternative. This would mean that we would use the conventional TVARLS scheme and estimate the basis function for each parameter separately. The gains and losses as compared to the MTVARLS scheme are evident. We could have, say, four basis functions for each parameter and would in principle be able to model the initial and final states, transition times and transition rates for each parameter evolution separately. The total degrees of freedom would then be $4 \cdot 6 = 24$ that is just one tenth of the data on the observation interval. This ratio can be considered to be somewhat too small. Again, the separate bases would allow the changes in individual parameter evolutions to happen at different times. This is not possible with the MTVARLS basis corresponding to the hypothetical evolution.

On the other hand, the correlations between the initial and final states would largely be lost and the final states would approximately equal to stationary AR model estimates that are calculated from the initial and final segments. In the hypothetical evolution of Chapter 5 the initial and final states were assumed to be

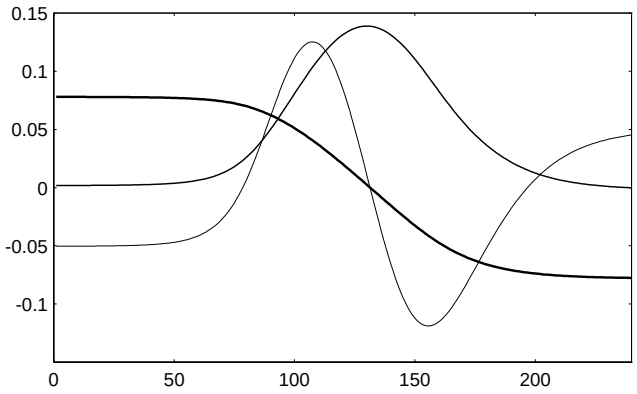


Figure 6.22: The first three eigenvectors (in decreasing boldness of line styles) corresponding to the covariance of the centered sigmoids.

independent but this was just an *ad hoc* choice. In the case of MTVARLS scheme and blind estimates for the basis this would not be the case.

With the hypothetical evolutions, the first basis vector would be set to unity and the eigenvectors for $k > 1$ would be obtained from the covariance of $\sigma_{k\ell}(t) - 1/2$ (the sigmoidal model evolutions, see Fig. 5.2). The first three of these eigenvectors are shown in Fig. 6.22. It can be seen that for the most part the coefficients corresponding to the constant function and the first eigenvector determine the initial and final states, the second eigenvector determines the transition instant and the third eigenvector determines the rate of change of the transition.

CHAPTER VII

Conclusions

Since the 1970's the trend in time series analysis and system identification has been towards nonstationary and nonlinear (not normally distributed) analyses. Several classes of methodologies have been developed for both cases. Although the methods used in EEG analysis usually assume both stationarity and linearity, it is well known that these assumptions are not valid and in many cases not even approximately valid. This thesis is concerned with the analysis of nonstationary EEG.

In a general setting there are limits in the performance of nonstationary analyses. For example, in nonstationary spectral estimation the time-frequency uncertainty principle is a fundamental limit. This principle does not depend on the estimation method but is a consequence of the definition of spectrum and the Cauchy-Schwartz inequality. The situation is different when prior information on the process is available. The problem is then usually that the information is such that it can not be implemented into the estimation algorithms easily. For example in nonstationary spectrum estimation the inherent uncertainties can be partially avoided when certain structural information on the process evolution is available.

The adaptive algorithms that are able to track the process or system parameters continuously have been proposed to be used in nonstationary time series analysis from the 1960's on when they were first conceived. Several adaptive algorithms have been suggested to be used also to estimate nonstationary EEG. Till the end of 1980's the performance analysis of adaptive algorithms was concentrated on transient analysis while the studies on true tracking performance were uncommon. Since that it has turned out that the analytical results on true tracking performance are not easy to use to optimize and analyze the algorithm performance. One has then usually to resort to simulations in the evaluation of the algorithm performance. The problem can then be that there are no easy ways of producing realistic nontrivial simulations with adequate variability.

In this thesis we have presented new methods for the simulation and estimation of nonstationary processes. Although the main application of the methods in this thesis is EEG it is by no means the only possible application. On the other hand, these new methods were originally motivated and developed by the specific problems in EEG analysis, that is, generating simulations for Monte Carlo type

performance evaluation of EEG analysis methods and the estimation of the event related synchronization changes.

The proposed method for the simulation of scalp EEG is the first systematic approach to the problem. Although some details of the method have been used earlier, the realization of the state variabilities, the smoothing of the predictor evolution and the overall composition are new. It was also demonstrated that acceptable simulations can be realized with information based on very small amount of data. It is also worth of noting that the method was developed in the first place since there were no such methods previously and these were necessary for the evaluation of the MTVARLS method which was the original aim of this thesis.

In the second novel part of this thesis a new estimation method was developed that is capable of taking into account prior information on the phenomenon. This is called the modified time-varying autoregressive least squares (MTVARLS) method. We believe that it is worth to explain shortly the train of thoughts that led to the MTVARLS method. The MTVARLS method and the consept of optimal subspaces grew gradually when we gained understanding of the TVARLS method. One of the first problems associated with the TVARLS method was that all the proposed bases (subspaces) were more or less of *ad hoc* type. A question of possible optimality of the bases was then pursued. After some experimental work and theoretical considerations it became clear that in general cases (such as the prolate spheroidal wave bases in the low pass type parameter evolution case) the performance of any basis with some optimality structure would not be considerably better than, say, of generic polynomial basis. However, the optimality of a basis scheme could be pursued if the variability of the nonstationary covariances would not be too big. The event related synchronization changes are examples of such cases. Once it was conceived that in this case it was possible to construct a set of representative individual parameter evolutions, the use of principal component analysis with this hypothetical ensemble to obtain low-dimensional approximations arose immediately. The problem was still that although, in a way, the individual parameter dynamics could be controlled in this way, the information on the interrelations and correlations between *e.g.* initial and final states could not be implemented in the ordinary TVARLS scheme in any feasible way. It was then realized that the parameter evolutions could not be treated and estimated individually. This meant that the parameter evolutions had to be concatenated and that these concatenations should be treated as primary variables. The principal component analysis of these variables led to a subspace constraint of the TVARLS scheme that we call the MTVARLS scheme. This is the theory presented in this thesis. The obvious Bayesian extension of the approach is not pursued in this thesis. The method was then applied to the estimation of event related synchronization.

The final part of the thesis consists of a preliminary evaluation of the MTVARLS method in a situation in which it is not possible to form a hypothetical ensemble of expectable parameter evolutions. As a solution to this problem we proposed the estimation of the parameter evolutions from samples using adaptive

predictors and to use these estimates as the ensemble. This approach was called blind estimation of MTVARLS subspaces. After discussing the possibilities for analytical studies we studied the method experimentally to find whether the principal component analysis of the adaptive predictor estimates could extract the structure of the principal subspace of the true ensemble. This was done in both large and small sample cases. It was found that although the blind estimation for the basis could in principle be done, the required sample size can turn out to be too large in most cases, especially in the case of event related synchronization changes.

All methods proposed in this thesis can be said to be on theoretically sound bases. The complexity of the ideas, as is often the case with nonstationary analyses, does not, however, allow for any simple universal interpretation of the results obtained with these methods. Furthermore, the applicability of the proposed methods depends heavily on the characteristics of the end application and the verification of the applicability of these methods can turn out to be tedious in many cases.

REFERENCES

[1] N.A. Abdraddo. A new approach to the problem of estimating spectral parameters of non-stationary time series models. *Journal of Computers in Applied Mathematics*, 5:125–129, 1979.

[2] N. Amir and I. Gath. Segmentation of EEG during sleep using time varying autoregressive modeling. *Biol Cybern*, 61:447–455, 1989.

[3] B.D.O. Anderson and P.J. Moylan. Spectral factorization of a finite-dimensional nonstationary matrix covariance. *IEEE Trans Automat Contr*, 19:680–691, 1974.

[4] K. Arakawa, D. H. Fender, H. Harashima, H. Miyakawa, and Y. Saitoh. Separation of a nonstationary component from the EEG by a nonlinear digital filter. *IEEE Trans Biomed Eng*, 33:724–726, 1986.

[5] J.E. Arle and R.H. Simon. An application of fractal dimension to the detection of transients in the electroencephalogram. *Electroencephalogr Clin Neurophysiol*, 75:269–305, 1990.

[6] I. Bankman and I Gath. Chebychev filter bank for estimation of frequency band powers in EEG. *Med Biol Eng Comput*, 29:94–98, 1991.

[7] J. Barlow. Computerized clinical electroencephalography in perspective. *IEEE Trans Biomed Eng*, 26:377–391, 1979.

[8] R. Barr, J. Ackmann, and J. Sonnenfeld. Peak-detection algorithm for EEG analysis. *Int J Bio-Med Comput*, 9:465–476, 1978.

[9] F. Bartoli and S. Cerutti. An optimal linear filter for the reduction of noise superimposed to the EEG signal. *J Biomed Eng*, 5:274–280, 1983.

[10] E. Basar. *Chaos in Brain Function*. Springer-Verlag, 1990.

[11] M. Basseville and A. Benveniste. Sequential detection of abrupt changes in spectral characteristics of digital signals. *IEEE Trans Inform Theory*, 29:709–724, 1983.

[12] R. Bender, B. Schultz, A. Shultz, and I. Pichlmayr. Identification of EEG patterns occurring during anesthesia by means of autoregressive parameters. *Biomed Tech*, 36:236–240, 1991.

[13] A. Benveniste and J.-J. Fuchs. Single sample modal identification of a nonstationary staochastic process. *IEEE Trans Automat Contr*, 30:66–74, 1985.

[14] A. Benveniste, M. Metivier, and P. Priouret. *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, 1990.

[15] C.S.K. Bhagavan. On non-stationary time series. In *Handbook of Statistics, Vol.*

5. Elsevier, 1985.

[16] G. Birch, P. Lawrence, J. Lind, and R. Hare. Application of prewhitening to AR spectral estimation of EEG. *IEEE Trans Biomed Eng*, 35:640–645, 1988.

[17] W. Birkemeier, A.B. Fontaine, G. Celesia, and K. Ma. Pattern recognition techniques for the detection of epileptic transients in EEG. *IEEE Trans Biomed Eng*, 25:213–217, 1978.

[18] S. Bittanti and M. Campi. Adaptive RLS algorithms under stochastic excitation - L^2 convergence analysis. *IEEE Trans Automat Contr*, 36:963–967, 1991.

[19] K. Blinowska and M. Malinowski. Non-linear and linear forecasting of the EEG time series. *Biol Cybern*, 66:159–165, 1991.

[20] K.J. Blinowska, L.T. Czerwosz, W. Drabik, P.J. Franaszczuk, and H. Ekiert. EEG data reduction by means of autoregressive representation and discriminant analysis procedures. *Electroencephalogr Clin Neurophysiol*, 51:650–658, 1981.

[21] G. Bodenstein and H. Praetorius. Pattern recognition of EEG by adaptive segmentation. In W. Perkins, editor, *Biomedical Computing*, pages 20–31. Pitman Medical, 1977.

[22] G. Bodenstein and H.M. Praetorius. Feature extraction from the electroencephalogram by adaptive segmentation. *Proc IEEE*, 65:642–652, 1977.

[23] G. Bodenstein, W. Schneider, and C. Malsburg. Computerized EEG pattern classification by adaptive segmentation and probability-density-function classification. Description of the method. *Comput Biol Med*, 15:297–313, 1985.

[24] Z. Bohdanezky, P. Lansky, and T. Radil-Weiss. EEG alpha and non-alpha intervals alteration. *Biol Cybern*, 30:109–113, 1978.

[25] T. Bohlin. Analysis of EEG signals with changing spectra. Technical Report Tech. Paper 18.212, IBM Nordic Laboratory, Sweden, 1971.

[26] T. Bohlin. Analysis of EEG signals with changing spectra using a short-word Kalman estimator. *Math Biosci*, 35:221–259, 1977.

[27] F. Boiten, J. Sergeant, and R. Geuze. Event-related desynchronization: the effects on energetic and computational demands. *Electroencephalogr Clin Neurophysiol*, 82:302–309, 1992.

[28] M. Boudaoud and L. Chaparro. Composite modeling of nonstationary signals. *Journal of the Franklin Institute*, 324:113–124, 1987.

[29] G.F. Boudreaux-Bartels and R. Murray. Time-frequency signal representations for the biomedical signals. In *The Biomedical Engineering Handbook*, pages 866–887. CRC Press, 1995.

[30] G. Box and G. Jenkins. *Time Series Analysis. Forecasting and Control*. Holden-Day, 1976.

[31] M. Brazier. The application of computers to electroencephalography. In R. Stacy and B. Waxman, editors, *Computers in Biomedical Research*, pages 295–315. Academic Press, 1965.

[32] G. Bremer, J. Smith, and I. Caracan. Automatic detection of the K-complexes in sleep electroencephalograms. *IEEE Trans Biomed Eng*, 17:314–323, 1970.

[33] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, 1991.

[34] R. Broughton, T. Healey, J. Maru, D. Green, and B. Pagurek. A phase locked loop device for automatic detection of sleep spindles and stage 2. *Electroencephalogr Clin Neurophysiol*, 44:677–680, 1978.

[35] J. Campbell, E. Bower, S. Dwyer, and G. Lago. On the sufficiency of autocorrelation functions as EEG descriptors. *IEEE Trans Biomed Eng*, 14:49–52,

1967.

[36] S. Cerutti, D. Liberati, G. Avanzini, S. Franceschetti, and F. Panzina. Classification of the EEG during neurosurgery. Parametric identification and Kalman filtering compared. *J Biomed Eng*, 8:244–254, 1986.

[37] S. Cerutti, D. Liberati, and P. Mascellani. Parameter extraction in EEG processing during riskful neurosurgical operations. *Signal Processing*, 9:25–35, 1985.

[38] R. Charbonnier, M. Barlaud, G. Alengrin, and J. Menez. Results on AR-modelling of nonstationary signals. *Signal Processing*, 12:143–151, 1987.

[39] H.-F. Chen and L. Guo. *Identification and Stochastic Adaptive Control*. Birkhäuser, 1991.

[40] B. Choi. *ARMA Model Identification*. Springer-Verlag, 1992.

[41] A. Cohen. *Biomedical Signal Processing*, volume I. CRC Press, 1986.

[42] B. C. Cohen and A. Sances Jr. Stationarity of the human electroencephalogram. *Med Biol Eng Comput*, 15:513–518, 1977.

[43] E. Cook and M. Bernfeld. *Radar signals*. Academic Press, 1967.

[44] O.-D. Creutzfeldt, G. Bodenstein, and J.S. Barlow. Computerized EEG pattern classification by adaptive segmentation and probability density function classification. Clinical evaluation. *Electroencephalogr Clin Neurophysiol*, 60:373–393, 1985.

[45] D.H. Crowell, R.H. Jones, L.E. Kapuniai, and P. Leung. Autoregressive representation of infant EEG for the purpose of hypotheses testing and classification. *Electroencephalogr Clin Neurophysiol*, 43:317–324, 1977.

[46] M. Culic, B. Jankovic, J. Milenkovic, and L. Rakic. EEG power spectra in the acute rat model of epilepsy. *Periodicum Biologorum*, 93:229–230, 1991.

[47] A.C. Da Rosa, B. Kemp, F.H. Lopes Da Silva, and H.A.C. Kamphuisen. A model-based detector of vertex waves and K complexes in sleep electroencephalogram. *Electroencephalogr Clin Neurophysiol*, 78:71–79, 1991.

[48] M. Deistler, O.Prohaska, E.Reschenhofer, and R. Vollmer. Procedure for identification of different stages of EEG background activity and its application to the detection of drug effects. *Electroencephalogr Clin Neurophysiol*, 64:294–300, 1986.

[49] J.J. Denier van der Gon and J. Strackee. Some aspects of EEG frequency analysis. *IEEE Trans Biomed Eng*, 13:120–130, 1966.

[50] E. D'Hollander and G. Orban. Spike recognition and on-line classification by unsupervised learning system. *IEEE Trans Biomed Eng*, 26:279–284, 1979.

[51] G. Dumermuth and Th. Gasser. The computation of EEG bispectra. *Comput Programs Biomed*, 8:235–242, 1978.

[52] G. Dumermuth and L. Molinari. Spectral analysis of EEG background activity. In *Methods of Analysis of Brain Electrical and Magnetic Signals*, volume 1 of *Handbook of Electroencephalography and Clinical Neurophysiology*, pages 85–130. Elsevier, 1987.

[53] R. Dzwonczyk and M. Howie. Comparison of a set of power distribution parameters for intraoperative EEG monitoring. *IEEE Trans Biomed Eng*, 33:792–795, 1986.

[54] R. Dzwonczyk, M. Howie, and J. McDonald. A comparison between Walsh and Fourier analysis of the electroencephalogram for tracking the effects of anesthesia. *IEEE Trans Biomed Eng*, 31:551–556, 1984.

[55] E. Eleftheriou and D.D. Falconer. Tracking properties and steady state performance of RLS adaptive filter algorithms. *IEEE Trans Acoust, Speech Signal Processing*, ASSP-34:1097–1110, 1986.

[56] E. Eweda and O. Macchi. Convergence of the RLS and LMS adaptive filters. *IEEE Trans Circuits Syst*, CAS-34:799–803, 1987.

[57] D.C. Farden. Tracking properties of adaptive signal processing algorithms. *IEEE Trans Acoust, Speech Signal Processing*, ASSP-29:439–446, 1981.

[58] C. Faure. Attributed strings for recognition of epilectic transients in EEG. *Int J Bio-Med Comput*, 16:217–229, 1985.

[59] J. Fell. Deterministic chaos and the first positive Lyapunov exponent: a nonlinear analysis of the human electroencephalogram during sleep. *Biol Cybern*, 69:139–146, 1993.

[60] P. Fenwick, P. Mitchie, J. Dollimore, and G. Fenton. Mathematical simulation of the electroencephalogram using an autoregressive series. *Int J Bio-Med Comput*, 2:281–307, 1971.

[61] P.B.C. Fenwick, P. Mitchie, J. Dollimore, and G.W. Fenton. Application of the autoregressive model to E.E.G. analysis. *Agressologie*, 10:553–564, 1969.

[62] M. Flaherty. Application of polynomial splines to a time-varying autoregressive model of speech. In *ICASSP-88*, pages 2220–2223, 1988.

[63] W.J. Freeman. Simulation of chaotic EEG patterns with a dynamic model of the olfactory system. *Biol Cybern*, 56:139–150, 1987.

[64] B. Friedlander. Lattice implementation of some recursive parameter-estimation algorithms. *Int J Contr*, 37:661–684, 1983.

[65] K. Fukunaga. *Introduction to statistical pattern recognition*. Wiley, 1972.

[66] W.A. Fuller. Nonstationary autoregressive time series. In *Handbook of Statistics, Vol. 5*. Elsevier, 1985.

[67] G. Gandolfo, L. Glin, G. Lacoste, M. Rodi, and C. Gottesmann. Automatic sleep-wake scoring in the rat on microcomputer Apple II. *Int J Bio-Med Comput*, 23:83–95, 1988.

[68] W.A. Gardner. Correlation estimation and time-series modeling for nonstationary processes. *Signal Processing*, 15:31–41, 1988.

[69] I. Gath and E. Bar-On. Computerized method for scoring of polygraphic sleep recordings. *Comput Programs Biomed*, 11:217–223, 1980.

[70] I. Gath and E. Bar-On. Classical sleep stages end the spectral content of the EEG signal. *Int J Neurosci*, 22:147–156, 1983.

[71] I. Gath, C. Feuerstein, D.T. Pham, and G. Rondouin. On the tracking of rapid dynamic changes in seizure EEG. *IEEE Trans Biomed Eng*, 39:952–958, 1992.

[72] I. Gath, A. Michaeli, and C. Feuerstein. A model for dual channel segmentation of the EEG signal. *Biol Cybern*, 64:225–230, 1991.

[73] W. Gersch. Spectral analysis of EEG's by autoregressive spectral decomposition of time series. *Math Biosci*, 7:205–222, 1970.

[74] W. Gersch. Causality or driving in electrophysiological signal analysis. *Math Biosci*, 14:177–196, 1972.

[75] W. Gersch. Non-stationary multichannel time series analysis. In *Methods of Analysis of Brain Electrical and Magnetic Signals*, volume 1 of *Handbook of Electroencephalography and Clinical Neurophysiology*, pages 261–296. Elsevier, 1987.

[76] W. Gersch. Smoothness priors. In *New Directions in Time Series Analysis, Part II*, pages 113–146. Sringer-Verlag, 1991.

[77] W. Gersch, A. Gevins, and G. Kitagawa. A multivariate time varying autoregressive modeling of nonstationary covariance time series. In *IEEE Conf. CDC-83*, pages 579–584, 1983.

[78] W. Gersch, F. Martinelli, and J. Yonemoto. A Kullback-Leibler -nearest neighbor rule classification of EEGs: The EEG population screening problem, an anesthesia level EEG classification application. *Comput Biomed Res*, 13:283–296, 1980.

[79] W. Gersch, J. Yonemoto, and P. Naitoh. Automatic classification of multivariate EEGs using an amount of information measure and the eigenvalues of parametric time series model features. *Comput Biomed Res*, 10:297–318, 1977.

[80] A.S. Gevins. Introduction. In *Methods of Analysis of Brain Electrical and Magnetic Signals*, volume 1 of *Handbook of Electroencephalography and Clinical Neurophysiology*, pages 1–14. Elsevier, 1987.

[81] A.S. Gevins. Overview of computer analysis. In *Methods of Analysis of Brain Electrical and Magnetic Signals*, volume 1 of *Handbook of Electroencephalography and Clinical Neurophysiology*, pages 31–84. Elsevier, 1987.

[82] A.S. Gevins. Statistical pattern recognition. In *Methods of Analysis of Brain Electrical and Magnetic Signals*, volume 1 of *Handbook of Electroencephalography and Clinical Neurophysiology*, pages 541–582. Elsevier, 1987.

[83] A.S. Gevins, R.K. Stone, and S.D. Ragsdale. Differentiating the effects of three benzodiazipines on non-REM sleep EEG spectra. *Neuropsychobiology*, 19:108–115, 1988.

[84] A.S. Gevins, C.L. Yeager, S.L. Diamond, J.-P. Spire, G.M. Zeitlin, and A.H. Gevins. Automated analysis of the electrical activity of the human brain (EEG): a progress report. *Proc IEEE*, 63:1382–1399, 1975.

[85] J. Glover, P. Ktonas, N. Raghavan, J. Urunuela, S. Velamuri, and E. Reilly. A multichannel signal processor for the detection of epileptogenic sharp transients in the EEG. *IEEE Trans Biomed Eng*, 33:1121–1128, 1986.

[86] C. Goeller and C. Sinton. A microcomputer-based sleep stage analyzer. *Comput Methods Program Biomed*, 29:31–36, 1989.

[87] G.H. Golub and C.F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 1989.

[88] J. Gotman. *Clinical Applications of Computer Analysis of EEG and Other Neurophyological Signals*, volume 2 of *Handbook of Electroencephalography and Clinical Neurophysiology*, chapter Computer analysis of the EEG in epilepsy, pages 171–204. Elsevier, 1986.

[89] A.M. Grass and F.A. Gibbs. A Fourier transform of the electroencephalogram. *J Neurophysiol*, 1:521–526, 1938.

[90] Y. Grenier. Time-dependent ARMA modeling of nonstationary signals. *IEEE Trans Acoust, Speech Signal Processing*, ASSP-31:899–911, 1983.

[91] W. Grochulski, P. Penczek, and J. Pozielski. Segmentation of the epilectic EEG by means of a finite state automaton. *Int J Bio-Med Comput*, 18:35–44, 1986.

[92] L. Guo and L. Ljung. Exponential stability of general tracking algorithms. *IEEE Trans Automat Contr*, 40:1376–1387, 1995.

[93] L. Guo and L. Ljung. Performance analysis of general tracking algorithms. *IEEE Trans Automat Contr*, 40:1388–1402, 1995.

[94] F. Gustafsson. Segmentation of signals using piecewise constant linear regression models. Technical Report LiTH-ISY-R-1672, University of Linköping, Department of Automatic Control, 1994.

[95] M. Hall, A. Oppenheimer, and A. Willsky. Time-varying parametric modeling of speech. *Signal Processing*, 5:267–285, 1983.

[96] M Hallin and J.-F. Ingenbleek. Nonstationary Yule-Walker equations. *Statist Probab Lett*, 1:189–195, 1983.

[97] Y.-L. Hao, Y. Ueda, and N. Ishii. Improved procedure of complex demodulation and an application to frequency analysis of sleep spindles in EEG. *Med Biol Eng Comput*, 30:406–412, 1992.

[98] A. Hasman, B. Jansen, G. Landeweerd, and A. van Blokland-Vogelesang. Demonstration of segmentation techniques for EEG records. *Int J Bio-Med Comput*, 9:311–321, 1978.

[99] R. Hileman and D. Dick. Detection of phase characteristics of alpha waves in the electroencephalogram. *IEEE Trans Biomed Eng*, 18:379–382, 1971.

[100] M.L. Honig and D.G. Messerschmidt. *Adaptive Filters: Structures, Algorithms and Applications*. Kluwer Academic Publishers, 1984.

[101] T. Hsia and R. Kraft. A digital signal processing system for EEG frequency analysis. *Comput Programs Biomed*, 17:27–34, 1983.

[102] P.J. Huber. *Robust Statistical Procedures*. SIAM, 1977.

[103] E. C. Ifeachor, B. W. Jervis, E. L. Morris, E. M. Allen, and N. R. Hudson. New online method for removing ocular artefacts from EEG signals. *Med Biol Eng Comput*, 24:356–363, 1986.

[104] T. Inouye, H. Sakamoto, K. Shinosaki, S. Toi, and S. Ukai. Analysis of rapidly changing EEG's before generalized spike and wave complexes. *Electroencephalogr Clin Neurophysiol*, 76:205–221, 1990.

[105] T. Inouye, A. Yagasaki, H. Takahashi, and K. Shinosaki. The dominant direction of interhemispheric EEG changes in the linguistic process. *Electroencephalogr Clin Neurophysiol*, 51:265–275, 1981.

[106] A. Isaksson. On time variable properties of EEG signals examined by means of a Kalman filter. Technical Report 95, Royal Institute of Technology, 1975.

[107] A. Isaksson. Identification of time varying systems through adaptive Kalman filtering. In *IFAC 10th Triennial World Congress*, pages 305–310, 1987.

[108] A. Isaksson and A. Wennberg. An EEG simulator - a means of objective clinical interpretation of EEG. *Electroencephalogr Clin Neurophysiol*, 39:313–320, 1975.

[109] A. Isaksson and A. Wennberg. Spectral properties of nonstationary EEG signals, evaluated by means of Kalman filtering: Application examples from a vigilance test. In P. Kellaway and I. Petersen, editors, *Quantitative Analytic Studies in Epilepsy*, pages 389–402. Raven Press, 1976.

[110] A. Isaksson, A. Wennberg, and L.H. Zetterberg. Computer analysis of EEG signals with parametric models. *Proc IEEE*, 69:451–461, 1981.

[111] N. Ishii, A. Iwata, and N. Suzumura. Segmentation of non-stationary time series. *Int J Systems Sci*, 10:883–894, 1979.

[112] N. Ishii, A. Iwata, and N. Suzumura. Detection of abrupt change and trend in the time series. *Int J Systems Sci*, 11:557–566, 1980.

[113] V. Iyer, Y. Ploysongsang, and P. Ramamoorthy. Adaptive filtering in biological signal processing. *CRC Crit Rev Biomed Eng*, 17:531–584, 1990.

[114] V. Jagannathan, J. Bourne, B. Hansen, and J. Ward. Artificial intelligence methods in quantitative electroencephalogram analysis. *Comput Programs Biomed*, 15:249–258, 1982.

[115] B. Jansen. Comments on "Walsh spectral estimates with applications to the classification of the EEG signals". *IEEE Trans Biomed Eng*, 28:667–668, 1981.

[116] B. Jansen. Analysis of biomedical signals by means of linear modeling. *CRC Crit Rev Biomed Eng*, 12:343–392, 1985.

[117] B. Jansen. Quantitative analysis of electroencephalograms: Is there chaos in the future? *Int J Bio-Med Comput*, 27:95–123, 1991.

[118] B. Jansen, J. Bourne, and J. Ward. Autoregressive estimation of short segment spectra for computerized EEG analysis. *IEEE Trans Biomed Eng*, 28:630–638, 1981.

[119] B. Jansen, J. Bourne, and J. Ward. Spectral decomposition of EEG intervals using Walsh and Fourier transforms. *IEEE Trans Biomed Eng*, 28:836–838, 1981.

[120] B. Jansen, A. Hasman, and R. Lenten. Piece-wise EEG analysis: an objective evaluation. *Int J Bio-Med Comput*, 12:17–27, 1981.

[121] B. Jansen, A. Hasman, and R. Lenten. Piecewise analysis of EEGs using AR-modeling and clustering. *Comput Biomed Res*, 14:168–178, 1981.

[122] B. Jansen, A. Hasman, R. Lenten, and R. Pikaar. Automatic sleep staging by means of profiles. In *MEDINFO 80*, pages 385–389, 1980.

[123] B. Jansen, A. Hasman, and S. Visser. Features to segmentate EEG recordings: A comparative study. In *Lecture Notes in Medical Informatics I*, pages 533–544. Springer-Verlag, 1978.

[124] B.H. Jansen, J.R. Bourne, and J.W. Ward. Identification and labeling of EEG graphic elements using autoregressive spectral estimates. *Comput Biol Med*, 12:97–106, 1982.

[125] B. Jervis, M. Coelho, and G. Morgan. Spectral analysis of EEG responses. *Med Biol Eng Comput*, 27, 1989.

[126] R. Jindra. EEG analysis by means of the PARCOR coefficients. *Biol Cybern*, 28:51–54, 1977.

[127] R Jindra. EEG feature extraction using PARCOR coefficients. *Biomedical Journal*, 22(6):553–560, 1980.

[128] M. Jobert, H. Escola, E. Poiseau, and P. Gaillard. Automatic analysis of sleep using two parameters based on principal component analysis of electroencephlography spectral data. *Biol Cybern*, 71:197–207, 1994.

[129] G.E. Johnson. Constructions of particular random processes. *Proc IEEE*, 82:270–285, 1994.

[130] T. Johnson, S. Wright, and A. Segall. Filtering of muscle artifact from the electroencephalogram. *IEEE Trans Biomed Eng*, 26:556–563, 1979.

[131] I.T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 1986.

[132] J.P. Kaipio and E. Koivisto. Analysis of EEG spindles and their automatic detection. In *National Conference on Biophysics and Biotechnology, Kuopio, Finland*, pages 18–19, 1991.

[133] N. Kawabata. A nonstationary analysis of the electroencephalogram. *IEEE Trans Biomed Eng*, 20:444–452, 1973.

[134] S.M. Kay and S.L. Marple. Spectrum analysis - a modern perspective. *Proc IEEE*, 69:1380–1419, 1981.

[135] B. Keane. EEG phasic event detection by microcomputer. *IEEE Trans Biomed Eng*, 25:297–299, 1978.

[136] Z. Keirn and J. Aunon. Man-machine communications through brain-wave processing. *IEEE Eng Med Biol Mag*, 9:55–57, 1990.

[137] Z. Keirn and J. Aunon. A new mode of communication between man and his surroundings. *IEEE Trans Biomed Eng*, 37:1209–1214, 1990.

[138] B. Kemp. *Model-based monitoring of human sleep stages*. PhD thesis, University of Twente, 1987.

[139] B. Kemp and H.A.C. Kamphuisen. Simulation of human hypnograms using a Markov chain model. *Sleep*, 9:405–414, 1986.

[140] Y. Kingma, C. Pronk, and D. Sparreboom. Parameter estimation of power spectra

using Gaussian functions. *Comput Biomed Res*, 9:591–599, 1976.

[141] M. Könönen and J.V. Partanen. Blocking of EEG alpha activity during visual performance in healthy adults. A quantitative study. *Electroencephalogr Clin Neurophysiol*, 87:164–166, 1993.

[142] F. Kozin and F. Nakajima. The order determination problem for linear time-varying AR models. *IEEE Trans Automat Contr*, AC-25:250–257, 1980.

[143] V. Krajza, S. Petranek, I. Patakova, and A. Värri. Automatic identification of significant graphoelements in multichannel EEG recordings by adaptive segmentation and fuzzy clustering. *Int J Bio-Med Comput*, 28:71–89, 1991.

[144] R. Raja Kumar and R. Pal. Tracking of bandpass signals using center-frequency adaptive filters. *IEEE Trans Acoust, Speech Signal Processing*, 38:1710–1721, 1990.

[145] T. Lagerlund and F. Sharbrough. Computer simulation of the generation of the electroencephalogram. *Electroencephalogr Clin Neurophysiol*, 72:31–40, 1989.

[146] H. Larsen and D. Lai. Walsh spectral estimates with applications to the classification of the EEG signals. *IEEE Trans Biomed Eng*, 27:485–492, 1980.

[147] L.A. Liporace. Linear estimation of nonstationary signals. *J Acoust Soc Am*, 58:1288–1295, 1975.

[148] L. Ljung. *System Identification: Theory for the User*. Prentice-Hall, 1987.

[149] L. Ljung. *System Identification Toolbox for MATLAB*, 1988.

[150] L. Ljung, G. Pflug, and H. Walk. *Stochastic Approximation and Optimization of Random Systems*. Birkhäuser, 1992.

[151] L. Ljung and T. Söderström. *Theory and Practice of Recursive Identification*. MIT Press, 1983.

[152] F.H. Lopes da Silva. Computer-assisted EEG diagnosis: Pattern recognition techniques. In *Electroencephalography: Basic principles, clinical applications and related fields*, pages 900–919. Urban & Schwarzenberg, 1987.

[153] F.H. Lopes da Silva. Dynamics of EEGs as signals of neuronal populations: Models and theoretical considerations. In *Electroencephalography: Basic principles, clinical applications and related fields*, pages 16–28. Urban & Schwarzenberg, 1987.

[154] F.H. Lopes da Silva. EEG analysis: Theory and practice. In *Electroencephalography:Basic principles, clinical applications and related fields*, pages 871–890. Urban & Schwarzenberg, 1987.

[155] F.H. Lopes da Silva and N.J.I. Mars. Parametric methods in EEG analysis. In *Methods of Analysis of Brain Electrical and Magnetic Signals*, volume 1 of *Handbook of Electroencephalography and Clinical Neurophysiology*, pages 234–260. Elsevier, 1987.

[156] F.H. Lopes da Silva and A. Van Rotterdam. Biophysical aspects of EEG and magnetoencephalogram generation. In *Electroencephalography: Basic principles, clinical applications and related fields*, pages 30–41. Urban & Schwarzenberg, 1987.

[157] F.H. Lopes da Silva, W. ten Broeke, K. van Hulten, and J. Lommen. EEG nonstationaries detected by inverse filtering in scalp and cortical recordings of epileptics: statistical analysis and spatial display. In *Quantitative Analytical Studies in Epilepsy*, pages 375–387. Raven Press, 1976.

[158] P. Loula. *Nonlinear methods of processing physiological signals in anesthesia and vigilance*. PhD thesis, Tampere University of Technology, 1994.

[159] P. Loula, K. Mäkelä, V. Jäntti, and J. Astola. Correlated noise sources of EEG in anesthesia: an application to average mutual information based multivariate autoregressive model. In *Proc. 2nd Ragnar Granit Symp.: EEG and MEG Signal Analysis and Interpretation*, pages 27–32, 1993.

[160] D.G. Luenberger. *Optimization by Vector Space Methods*. Wiley, 1969.
[161] O. Macchi and N. Bershad. Adaptive recovery of a chirped sinusoid in noise,part 1: Performance of the RLS algorithm. *IEEE Trans Signal Processing*, ASSP-39:583–594, 1991.
[162] P. C. Madhavan, B. E. Stephens, D. Klingberg, and S. Morzorati. Analysis of rat EEG using autoregressive power spectra. *J Neurosci Meth*, 40:91–100, 1991.
[163] O.M. Markand. Alpha rhythms. *J Clin Neurophysiol*, 7:163–189, 1990.
[164] S.L. Marple. *Digital Spectral Analysis with Applications*. Prentice-Hall, 1987.
[165] N.J.I. Mars and F.H. Lopes da Silva. EEG analysis methods based on information theory. In *Methods of Analysis of Brain Electrical and Magnetic Signals*, volume 1 of *Handbook of Electroencephalography and Clinical Neurophysiology*, pages 297–308. Elsevier, 1987.
[166] M. Matejzek and J. Devos. Selected methods of quantitative EEG analysis and their applications in psychotropic drug research. In *Quantitative Analytical Studies in Epilepsy*, pages 183–205. Raven Press, 1976.
[167] J. McEwen and G. Anderson. Modeling the stationarity and Gaussianity of spontaneous electroencephalographic activity. *IEEE Trans Biomed Eng*, 22:361–369, 1975.
[168] J. McEwen, G. Anderson, M. Low, and L. Jenkins. Monitoring the level of anesthesia by automatic analysis of spontaneous EEG activity. *IEEE Trans Biomed Eng*, 22:299–305, 1975.
[169] J.L. Melsa and D.L. Cohn. *Decision and Estimation Theory*. McGraw-Hill, 1978.
[170] K.S. Miller. Nonstationary autoregressive processes. *IEEE Trans Inform Theory*, 15:315–316, 1969.
[171] L. Molinari and G. Dumermuth. Robust spectral analysis of the EEG. *Neuropsychobiology*, 15:208–218, 1986.
[172] T.J. Morrow and K.L.Casey. A microprocessor device for the real-time detection of synchronized alpha and spindle activity in the EEG. *Brain Research Bulletin*, 16:439–442, 1986.
[173] A. Murro, D. King, H. Flanigin, B. Gallagher, and J. Smith. Adaptive smoothing: an improved method for spectral analysis and its application to seizure EEG. *Int J Bio-Med Comput*, 26:63–72, 1990.
[174] S. Narasimhan and D. Dutt. Software simulation of the EEG. *J Biomed Eng*, 7:275–281, 1985.
[175] S.V. Narasimhan. Pole-zero spectral modeling of EEG. *Signal Processing*, 18:17–32, 1989.
[176] K. Nathan, Y.-T. Lee, and H. Silverman. A time varying analysis method for rapid transitions in speech. *IEEE Trans Signal Processing*, 39:815–824, 1991.
[177] A. Nehorai and D. Starer. Adaptive pole estimation. *IEEE Trans Acoust, Speech Signal Processing*, 38:825–838, 1990.
[178] M. Niedzwiecki. First order tracking properties of weighted least squares estimators. *IEEE Trans Automat Contr*, 33:94–96, 1988.
[179] M. Niedzwiecki. On tracking characteristics of weighted least squares estimators applied to nonstationary system identification. *IEEE Trans Automat Contr*, 33:96–98, 1988.
[180] T. Ning and J. Bronzino. Autoregressive and bispectral analysis techniques: EEG application. *IEEE Eng Med Biol Mag*, 9(March):47–50, 1990.
[181] M. R. Nuwer. Quantitative EEG: I. techniques and problems of frequency analysis and topographic mapping. *J Clin Neurophysiol*, 5:1–43, 1988.

[182] M.R. Nuwer, D. Lehmann, F. Lopes da Silva, S. Matsuoka, W. Sutherling, and J.-F. Vibert. IFCN guidelines for topographic and frequency analysis of EEGs and EPs. Report of an IFCN committee. *Electroencephalogr Clin Neurophysiol*, 91:1–5, 1994.

[183] D. O'Boyle, E. Choi, G. Conroy, and M. Turega. Learned classification of EEG power spectra using a neural network. *Journal of Physiology*, 438:345, 1991.

[184] E. Oja. *Subspace Methods in Pattern Recognition*. Wiley, 1983.

[185] A.V. Oppenheim and R.W. Schafer. *Digital Signal Processing*. Prentice-Hall, 1975.

[186] Achermann P, R. Hartmann, A. Gunzinger, W. Guggenbühl, and A. Borbely. All-night sleep and artificial stochastic control signals have similar correlation dimensions. *Electroencephalogr Clin Neurophysiol*, 90:384–387, 1994.

[187] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 1984.

[188] A. Papoulis. *Signal Analysis*. McGraw-Hill, 1984.

[189] R.D. Pascual-Marqui, P.A. Valdes-Sosa, and A. Alvarez-Amador. A parametric model for multichannel EEG spectra. *Int J Neurosci*, 40:89–99, 1988.

[190] L. Patomäki, J.P. Kaipio, and P.A. Karjalainen. Tracking of nonstationary EEG with the roots of ARMA models. In *IEEE Conf. EMBC-95*, 1995.

[191] M.S. Peiris. On prediction with the time dependent ARMA models. *Comm Statist A: Theory and Methods*, 15:3695–3668, 1986.

[192] P. Penczek, W. Grochulski, J. Gryb, and M. Kowalzyk. The use of a multi-channel Kalman filter algorithm in structural analysis of the epileptic EEG. *Int J Bio-Med Comput*, 20:135–151, 1987.

[193] G. Pfurtscheller and G. Haring. The use of an EEG autoregressive model for the time-saving calculation of spectral power density distribution with a digital computer. *Electroencephalogr Clin Neurophysiol*, 33:113–115, 1972.

[194] G. Pfurtscheller, C. Neuper, and W. Mohl. Event-related desynchronization (ERD) during visual processing. *Int J Psychophysiol*, 16:147–153, 1994.

[195] J. Pijn, J. van Veerven, A. Noest, and F. Lopez da Silva. Chaos or noise in EEG signals. Dependence on state and brain site. *Electroencephalogr Clin Neurophysiol*, 79:371–381, 1991.

[196] J. Pinon, P. Rubel, G. Maraval, F. Mauguirre, and M. Revol. Analyse discriminante des grapho-elements de l'EEG. application a la detection des bouffees de pointes-ondes diffuses. *Int J Bio-Med Comput*, 13:237–261, 1982.

[197] R.T. Pivik, F.W. Bylsma, and R.J. Nevins. A new device for automatic sleep spindle analysis: the Spindicator. *Electroencephalogr Clin Neurophysiol*, 54:711–713, 1982.

[198] D. Popivanov. Time series analysis of brain potentials preceding voluntary movements. *Med Biol Eng Comput*, 30:9–14, 1992.

[199] N. Pradhan and D.N. Dutt. Use of running fractal dimension for the analysis of changing patterns in electroencephalograms. *Comput Biol Med*, 23:381–388, 1993.

[200] M.B. Priestley. *Spectral Analysis and Time Series*. Academic Press, 1981.

[201] M.B. Priestley. *Non-linear and non-stationary time series analysis*. Academic Press, 1988.

[202] R. Pronk and A. Simons. Processing of the electroencephalogram in cardiac surgery. *Comput Programs Biomed*, 18:181–190, 1984.

[203] R.A.F. Pronk. Peri-operative monitoring. In *Handbook of Electroencephalography and Clinical Neurophysiology. Clinical Applications of Computer Analysis of EEG and Other Neurophysiological Signals*, pages 93–130. Elsevier, 1986.

[204] J. Qian, J. Barlow, and M. Beddoes. A simplified arithmetic detector for EEG sharp transients - preliminary results. *IEEE Trans Biomed Eng*, 35:11–18, 1988.

[205] S. Quint, D. Michaels, G. Hillard, and J. Messenheimer. A real-time system for the spectral analysis of the EEG. *Comput Methods Program Biomed*, 28:11–22, 1989.

[206] T. Subba Rao. The fitting of non-stationary time series models with time-dependent parameters. *J R Statist Soc*, B 32:312–322, 1970.

[207] E. Reschenhofer, R. Vollmer, and M. Deistler. Segmentation of the EEG by minimization of the total gain of information. *Neuropsychobiology*, 18:155–159, 1987.

[208] S. Roberts and L. Tarassenko. New method of automated sleep quantification. *Med Biol Eng Comput*, 30:509–517, 1992.

[209] Z. Rogowski, I. Gath, and E. Bental. On the prediction of epileptic seizures. *Biol Cybern*, 42:9–15, 1981.

[210] Y. Saito and H. Harashima. Tracking of informations within multichannel EEG record - causal analysis in EEG. In *Recent Advances in EEG and EMG Data Processing*, pages 133–146. Elsevier, 1981.

[211] B. Saltzberg, L.S. Lustick, and R.G. Heath. Detection of focal depth spiking in the scalp EEG of monkey. *Electroencephalogr Clin Neurophysiol*, 31:327–333, 1971.

[212] D. Sauter, B. J. Martin, N. Di Renzo, and C. Vomscheid. Analysis of eye tracking movements using innovations generated by a Kalman filter. *Med Biol Eng Comput*, 29:63–69, 1991.

[213] B. Schack, H. Wagner, G. Griezbach, and H. Witte. Statistical comparison of band spectral power: an aid in stochastic analysis of brain electrical activity. part 2: Construction of confidence intervals of spectral band power and their aplication in EEG analysis. *Biomedical Journal*, 34:827–841, 1992.

[214] B. Schack, H. Wagner, and H. Witte. Statistical comparison of band spectral power: an aid in stochastic analysis of brain electrical activity. part 1: Approximation of the distribution of broad band power by a χ^2 distribution. *Biomedical Journal*, 34:179–192, 1992.

[215] N. Schaltenbrand, R. Lengelle, and J.-J. Macher. Neural network model: Application to automatic analysis of human sleep. *Comput Biomed Res*, 26:157–171, 1993.

[216] B. Sklar, J. Hanley, and W. Simmons. A computer analysis of EEG spectral signatures from normal and dyslectic children. *IEEE Trans Biomed Eng*, 20:20–26, 1973.

[217] C.G. Small and D.L. McLeish. *Hilbert Space Methods in Probability and Statistical Inference*. Wiley, 1994.

[218] J. Smith. Automatic analysis and detection of EEG spikes. *IEEE Trans Biomed Eng*, 21:1–7, 1974.

[219] J.R. Smith. Automated analysis of sleep EEG data. Handbook of Electroencephalography and Clinical Neurophysiology. Clinical Applications of Computer Analysis of EEG and Other Neurophyological Signals, pages 131–147. Elsevier, 1986.

[220] W. D. Smith and D. L. Lager. Evaluation of simple algorithms for spectral parameter analysis of the electroencephalogram. *IEEE Trans Biomed Eng*, 33:352–358, 1986.

[221] H.-W. Steinberg, T. Gasser, and J. Franke. Fitting autoregressive models to EEG time series: an empirical comparison of estimates of the order. *IEEE Trans Acoust,*

Speech Signal Processing, ASSP-33:143–150, 1985.

[222] M. Steriade, P. Gloor, F.H. Lopez de Silva, and M.-M. Mesulam. Basic mechanisms of cerebral rhythmic activities. *Electroencephalogr Clin Neurophysiol*, 76:481–508, 1990.

[223] H. Sugimoto, N. Ishii, A. Iwata, and N. Suzumura. On the stationarity and normality of the electroencephalographic data during sleep stages. *Comput Programs Biomed*, 8:224–234, 1978.

[224] N.V. Thakor and D. Sherman. Wavelet (time-scale) analysis in biomedical signal processing. In *The Biomedical Engineering Handbook*, pages 886–906. CRC Press, 1995.

[225] B. Tharp and W. Gersch. Spectral analysis of seizures in humans. *Comput Biomed Res*, 8:503–521, 1975.

[226] C.E. Thomsen, A. Rosenfolck, and K.N. Christensen. Assesment of anaesthetic depth by clustering analysis and autoregressive modeling of electroencephalogram. *Comput Methods Program Biomed*, 34:125–138, 1991.

[227] M. van der Berg-Lenssen, C. Brunia, and J. Blom. Correction of ocular artefacts in EEGs using an autoregressive model to describe the EEG. A pilot study. *Electroencephalogr Clin Neurophysiol*, 73:72–83, 1989.

[228] A. Värri. *Algorithms and systems for the analysis of long-term physiological signals*. PhD thesis, Tampere University of Technology, 1992.

[229] F. Vaz, P. Guedes de Oliveira, and J. Principe. A study on the best order for autoregressive EEG modelling. *Int J Bio-Med Comput*, 20:41–50, 1987.

[230] B. Weide, L. Andrews, and A. Iannone. Real-time analysis of EEG using Walsh transforms. *Comput Biol Med*, 8:255–263, 1978.

[231] A. Wennberg and A. Isaksson. Simulation of nonstationary EEG signals as a means of objective clinical interpretation of EEG. In *Quantitative Analytical Studies in Epilepsy*, pages 493–509. Raven Press, 1976.

[232] A. Wennberg and L. Zetterberg. Application of a computer-based model for EEG analysis. *Electroencephalogr Clin Neurophysiol*, 31:457–468, 1971.

[233] B. Widrow and S.D. Stearns. *Adaptive Signal Processing*. Prentice-Hall, 1985.

[234] W. Van Winsum, J. Sergeant, and R. Geuze. The functional significance of event-related desynchronization of alpha rhythm in attentional and activating tasks. *Electroencephalogr Clin Neurophysiol*, 58:519–524, 1984.

[235] J. Wright, R. Kydd, and A Sergejev. Autoregression models of EEG. *Biol Cybern*, 62:201–210, 1990.

[236] L. Zetterberg. Estimation of parameters for a linear difference equation with application to EEG analysis. *Math Biosci*, 5:227–275, 1969.

[237] L. Zetterberg. Experience with analysis and simulation of EEG signals with parametric description of spectra. In P. Kellaway and I. Petersen, editors, *Automation of Clinical Electroencephalography*, pages 161–201. Raven Press, 1973.