Chapter 2
Generalized state-space models for modeling nonstationary EEG time-series

A. Galka, K.K.F. Wong, and T. Ozaki

2.1 Introduction

Contemporary neuroscientific research has access to various techniques for recording time-resolved data relating to human brain activity: electroencephalography (EEG) and magnetoencephalography (MEG) record the electromagnetic fields generated by the brain, while other techniques, such as near-infrared spectroscopy (NIRS) and functional magnetic resonance imaging (fMRI) are sensitive to the local metabolic activity of brain tissue.

Time-resolved data contain valuable information on the dynamical processes taking place in brain. EEG and MEG time-series are especially promising, since the electromagnetic fields of the brain are directly reflecting the activation of neural populations; furthermore these time-series can be recorded with high temporal resolution. Extraction of the dynamic changes captured by EEG/MEG recordings is an ideal application for *time-series analysis* [10].

From the multiplicity of concepts and methods for time-series analysis that have been applied to neuroscientific time-series, we focus here on *predictive modeling*, i.e., finding a predictor for future time-series values, based on present and past values. More precisely, we will discuss a particular class of predictive modeling that is attracting considerable attention due to its wide applicability: the *state-space* model [2, 3, 6, 12, 13].

Because nonstationary phenomena—such as sudden phase transitions relating to qualitative changes in dynamical behavior—cannot be modeled well using standard

Andreas Galka
Department of Neurology, University of Kiel, Schittenhelmstrasse 10, 24105 Kiel, Germany.
e-mail: a.galka@neurologie.uni-kiel.de

Kevin Kin Foon Wong
Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA

Tohru Ozaki
Tohoku University, 28 Kawauchi, Aoba-ku, Sendai 980-8576, Japan.

D.A. Steyn-Ross, M. Steyn-Ross (eds.), *Modeling Phase Transitions in the Brain,*
Springer Series in Computational Neuroscience 4, DOI 10.1007/978-1-4419-0796-7_2,

state-space approaches, in this chapter we present a generalization of state-space modeling appropriate for this purpose. This generalized algorithm may also serve as a detector for phase transitions.

2.2 Innovation approach to time-series modeling

Let the data be denoted by $y(t), t = 1,\ldots,T$, where T denotes the length of the time-series, i.e., the number of time points at which the data were sampled. In this chapter we will assume the case of univariate (scalar) data, although the modeling algorithms to be presented can also be applied to multivariate (vector) data; techniques like EEG and MEG usually provide multivariate time-series, resulting from a set of up to a few hundred sensors. By confining the analysis to a single channel, we confine our attention to the local brain area for which the chosen sensor is most sensitive.

At a given time point $t-1$ we intend to predict $y(t)$, employing the data $y(\tau), \tau = t-1, t-2, t-3, \ldots$ The optimal predictor is given by the conditional expectation $\mathscr{E}\big(y(t) \mid y(t-1), y(t-2), \ldots\big)$, such that the data model is given by

$$y(t) = \mathscr{E}\big(y(t) \mid y(t-1), y(t-2), \ldots\big) + v(t), \tag{2.1}$$

where $v(t)$ denotes the prediction error or *innovation*. The art of time-series modeling then lies in finding a good approximation to $\mathscr{E}\big(y(t) \mid y(t-1), y(t-2), \ldots\big)$. For an optimal predictor, any correlation structure in the data $y(t)$ is employed for the purpose of prediction, such that, in the time-series of innovations, no correlation of any kind remains, i.e., the innovations are a *white-noise* series. The concept of mapping given data to white innovations represents the core idea of the *innovation approach* to time-series modeling [11].

The theory of innovation approach modeling of Markov processes has been elaborated mainly by Levy [14] and Kailath [12]; one of the main results states that under mild conditions, including continuity of the dynamics, a predictor exists such that the innovations time-series will have a multivariate normal (Gaussian) distribution. We refrain from giving details here; instead the reader is referred to [18].

2.3 Maximum-likelihood estimation of parameters

A parametric function of present and past data, $y(t-1), y(t-2), \ldots$, may be chosen as an approximation to $\mathscr{E}\big(y(t) \mid y(t-1), y(t-2), \ldots\big)$, i.e., as a predictor; it will typically depend on a set of model parameters, collected in a vector ϑ. Following the concept of maximum-likelihood estimation of statistical parameters, we need to maximize the likelihood defined by the conditional probability distribution

$$L\big(\vartheta; y(1),\ldots,y(T)\big) = p\big(y(1),\ldots,y(T) \,\big|\, \vartheta\big)\,; \tag{2.2}$$

equivalently, the logarithm of the likelihood, $\log L\big(\vartheta; y(1),\ldots,y(T)\big)$, may be maximized. We will now derive an expression for $\log L\big(\vartheta; y(1),\ldots,y(T)\big)$, to be used in the innovation approach. The joint probability distribution of the data can be expanded as a product

$$\begin{aligned} &p\big(y(1),\ldots,y(T) \,\big|\, \vartheta\big) = \\ &\qquad p\big(y(1) \,\big|\, \vartheta\big)\, p\big(y(2) \,\big|\, y(1),\vartheta\big)\cdots p\big(y(T) \,\big|\, y(T-1),\ldots,y(1),\vartheta\big)\,, \end{aligned} \tag{2.3}$$

where we have used the fact that the data must obey *causality*. The joint probability distribution of the *innovations* has a simpler shape, due to the white-noise property which removes any conditioning on previous values:[1]

$$p\big(\nu(1),\ldots,\nu(T) \,\big|\, \vartheta\big) = p\big(\nu(1) \,\big|\, \vartheta\big)\, p\big(\nu(2) \,\big|\, \vartheta\big)\cdots p\big(\nu(T) \,\big|\, \vartheta\big)\,. \tag{2.4}$$

We can employ this simpler expression for deriving the likelihood of the data. The relationship between $p\big(y(1),\ldots,y(T) \,\big|\, \vartheta\big)$ and $p\big(\nu(1),\ldots,\nu(T) \,\big|\, \vartheta\big)$ can be found from the function linking these two sets of variables; it is given by Eq. (2.1). According to the standard rules for transforming probability distributions, the Jacobi determinant of this function then arises as a correction to be multiplied with $p\big(\nu(1),\ldots,\nu(T) \,\big|\, \vartheta\big)$; however, note that from Eq. (2.1) we have

$$\frac{\partial \nu(t)}{\partial y(\tau)} = \begin{cases} 1 & \text{for} \quad t = \tau \\ 0 & \text{for} \quad \tau > t\,, \end{cases} \tag{2.5}$$

where we have used the fact that also the predictor must obey causality. Consequently, the Jacobi determinant is unity, and the joint probability of the given data must be equal to the joint probability of the corresponding innovations,

$$p\big(\nu(1),\ldots,\nu(T) \,\big|\, \vartheta\big) = p\big(y(1),\ldots,y(T) \,\big|\, \vartheta\big) \tag{2.6}$$

although the functional form of these two distributions may differ very much. Finally this gives us for the logarithmic likelihood, employing a normal (Gaussian) distribution for the innovations, as argued above,

$$\log L\big(\vartheta; y(1),\ldots,y(T)\big) = -\frac{1}{2}\left(T\log\sigma_\nu^2(t) + \sum_{t=1}^{T}\frac{\nu^2(t)}{\sigma_\nu^2(t)} + T\log(2\pi)\right), \tag{2.7}$$

where $\sigma_\nu^2(t)$ denotes the variance of the innovations.

[1] Here the problem arises that, for the first data value $y(1)$, no previous values exist which could be employed by a predictor. But for sufficiently long time-series, the contribution of the first, or the first few, data values to the likelihood can be neglected.

2.4 State-space modeling

In state-space modeling [2, 3, 6, 12, 13], the data $y(t)$ are modeled by a system of two equations,

$$\mathbf{x}(t) = \mathsf{A}\mathbf{x}(t-1) + \eta(t) \tag{2.8}$$
$$y(t) = \mathsf{C}\mathbf{x}(t) + \varepsilon(t)\,, \tag{2.9}$$

where $\mathbf{x}(t)$ denotes the M-dimensional *state vector*, $\eta(t)$ the *dynamical noise* term and $\varepsilon(t)$ the *observation noise* term; the model parameters are given by the *state transition matrix* A and the *observation matrix* C. Furthermore, there are the covariance matrices S_η and σ_ε^2 of the noise terms (where for univariate data σ_ε^2 is a single variance parameter instead of an actual covariance matrix). Alternatively, the dynamical model, Eq. (2.8), could be chosen as a continuous-time model, i.e., as a stochastic differential equation.

When interpreted as an input–output model, the state-space model of Eqs (2.8, 2.9) produces one output signal $y(t)$ from two input signals $\eta(t)$ and $\varepsilon(t)$. This mapping is not invertible, i.e., the original inputs $\eta(t)$ and $\varepsilon(t)$ cannot be reconstructed from the output $y(t)$. However, it is possible to define a transformed model, such that instead of two input signals just one is present, appearing both in the positions of the dynamical noise and the observation noise; it turns out that this input signal is given by the innovations $\nu(t)$ [11]. While the innovations can directly replace observation noise, they need to be multiplied by a problem-specific gain matrix (the *Kalman gain matrix*), before they can replace dynamical noise; in the case of univariate data, this matrix will be an $(M \times 1)$-dimensional vector.

This transformed model is known as the *innovation representation* or *Kalman filter representation* of the state-space model. It can be shown that the mapping between $y(t)$ and $\nu(t)$ is invertible [11]. The existence of this representation provides the justification for practical state-space modeling of time-series.

For given model parameters, the famous Kalman filter algorithm can be applied for the purpose of generating estimates of the state vector [13]; improved estimates can be obtained by additional application of a smoother algorithm [19]. While the Kalman filter performs a pass through the time-series data in forward direction of time, the smoother proceeds in backward direction. Since predictions are only possible in forward direction, it is only the Kalman filter which maps the data to innovations and thereby provides a corresponding value for the likelihood of the data.

2.4.1 State-space representation of ARMA models

A well-established class of predictive models for time-series is given by autoregressive moving-average (ARMA) models [5]. As a simple example for univariate data $y(t)$, we consider the following ARMA(2,1) model:

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \eta(t) + b_1 \eta(t-1)\,, \tag{2.10}$$

where $\eta(t)$ denotes again a dynamical noise term, with variance σ_η^2. This model consists of an autoregressive (AR) term of second order, with parameters a_1, a_2, and a moving-average (MA) term of first order, with parameter b_1, therefore it is denoted by ARMA(2,1). We can rewrite Eq. (2.10) as

$$\begin{aligned} y(t) &= a_1 y(t-1) + \zeta(t-1) + \eta(t) \\ \zeta(t) &= a_2 y(t-1) + b_1 \eta(t) \end{aligned} \tag{2.11}$$

which is equivalent to

$$\begin{pmatrix} y(t) \\ \zeta(t) \end{pmatrix} = \begin{pmatrix} a_1 & 1 \\ a_2 & 0 \end{pmatrix} \begin{pmatrix} y(t-1) \\ \zeta(t-1) \end{pmatrix} + \begin{pmatrix} 1 \\ b_1 \end{pmatrix} \eta(t)\,, \tag{2.12}$$

where $\zeta(t)$ denotes an auxiliary state variable which can be interpreted as a slightly odd predictor of $y(t+1)$ [2]. We define a state vector as $\mathbf{x}(t) = \big(y(t), \zeta(t)\big)^\dagger$ (where † denotes matrix transpose) and obtain the state-space model

$$\mathbf{x}(t) = \begin{pmatrix} a_1 & 1 \\ a_2 & 0 \end{pmatrix} \mathbf{x}(t-1) + \begin{pmatrix} 1 \\ b_1 \end{pmatrix} \eta(t) \tag{2.13}$$

$$y(t) = (1,0)\,\mathbf{x}(t)\,. \tag{2.14}$$

The dynamical noise term of this model is given by $(1, b_1)^\dagger \eta(t)$; the corresponding covariance matrix follows as

$$\mathsf{S}_\eta = \begin{pmatrix} 1 & b_1 \\ b_1 & b_1^2 \end{pmatrix} \sigma_\eta^2\,. \tag{2.15}$$

In Eq. (2.14) observation noise is absent, $\sigma_\varepsilon^2 = 0$; however, as a generalization we may (and will) allow for nonzero σ_ε^2.

The specific form of the state transition matrix $\left(\begin{smallmatrix} a_1 & 1 \\ a_2 & 0 \end{smallmatrix}\right)$ is known as *left companion form*, or (in the language of control theory) *observer canonical form* [12]; it is a characteristic property of the state-space model corresponding to this form that the MA parameter b_1 is accommodated in the covariance matrix of the dynamical noise, while the observation matrix $\mathsf{C} = (1,0)$ keeps a very simple form.

Note that the scaling of the components of the state vector in Eq. (2.13) is directly controlled by the variance σ_η^2; since the model is linear, this degree of freedom can be shifted to the observation matrix which then becomes $\mathsf{C} = (c_1, 0)$ while the dynamical noise variance can be normalized to $\sigma_\eta^2 = 1$. While in the case of univariate data this is a possible, but not necessary choice, it provides the appropriate generalization for the case of multivariate data; for this reason, we will adopt this choice in this chapter.

The construction leading to the model of Eqs (2.13, 2.14) is easily extended to ARMA$(p, p-1)$ models with higher order $p > 2$, yielding a state-space model

$$\mathbf{x}(t) = \begin{pmatrix} a_1 & 1\,0\,\ldots\,0 \\ a_2 & 0\,1\,\ldots\,0 \\ \vdots & \vdots\,\vdots\,\ddots\,\vdots \\ a_{p-1} & 0\,0\,\ldots\,1 \\ a_p & 0\,0\,\ldots\,0 \end{pmatrix} \mathbf{x}(t-1) + \begin{pmatrix} 1 \\ b_1 \\ \vdots \\ b_{p-2} \\ b_{p-1} \end{pmatrix} \eta(t) \tag{2.16}$$

$$y(t) \;=\; (1,0,\ldots,0,0)\,\mathbf{x}(t)\,. \tag{2.17}$$

The covariance matrix of the dynamical noise term of this model follows as

$$\mathsf{S}_\eta = \begin{pmatrix} 1 & b_1 & b_2 & \ldots & b_{p-1} \\ b_1 & b_1^2 & b_1 b_2 & \ldots & b_1 b_{p-1} \\ b_2 & b_1 b_2 & b_2^2 & \ldots & b_2 b_{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{p-1} & b_1 b_{p-1} & b_2 b_{p-1} & \ldots & b_{p-1}^2 \end{pmatrix} \sigma_\eta^2\,. \tag{2.18}$$

2.4.2 Modal representation of state-space models

The dynamics of any linear state-space model can be characterized by the set of eigenvalues of its state transition matrix A; the eigenvalues are found by transforming A into a diagonal matrix. If M denotes the dimension of the state-space, there will be M eigenvalues; a certain subset of these eigenvalues will be real, denoted by $a^{(1)},\ldots,a^{(m_1)}$ (where m_1 denotes the number of real eigenvalues), while the remaining eigenvalues will form pairs of complex-conjugated eigenvalues (assuming that all elements of A are real), denoted by $(\psi_{(1)},\psi^*_{(1)},\ldots,\psi_{(m_2)},\psi^*_{(m_2)})$ (where m_2 denotes the number of pairs of complex-conjugated eigenvalues). Then we will have $M = m_1 + 2m_2$.

Real eigenvalues $a^{(k)}$ of A correspond to autoregressive models of first order, AR(1):

$$y(t) = a^{(k)} y(t-1) + \eta(t)\,. \tag{2.19}$$

Each complex-conjugated pair of eigenvalues $\psi_{(k)},\psi^*_{(k)}$ can be interpreted as an oscillatory eigen-mode of the dynamics, with a resonance frequency $\phi_{(k)}$ (corresponding to the phase of the complex eigenvalues) and an accompanying damping coefficient $\rho_{(k)}$ (corresponding to the modulus of the complex eigenvalues):

$$\psi_{(k)} = \rho_{(k)} \exp i\phi_{(k)}\,, \tag{2.20}$$

where $i = \sqrt{-1}$.

Consider a complex-conjugated pair of eigenvalues ψ,ψ^* within the diagonalized state transition matrix; it corresponds to a (2×2)-block $\left(\begin{smallmatrix} \psi & 0 \\ 0 & \psi^* \end{smallmatrix}\right)$ on the diagonal. It is always possible to transform such a block to left companion form $\left(\begin{smallmatrix} a_1 & 1 \\ a_2 & 0 \end{smallmatrix}\right)$ by a linear transform; therefore each complex-conjugated pair of eigenvalues can be

represented by an ARMA(2,1) model, according to Eq. (2.13). The autoregressive parameters follow from phase and modulus of the complex eigenvalues by

$$a_1^{(k)} = 2\rho_{(k)} \cos \phi_{(k)} \quad , \quad a_2^{(k)} = -\rho_{(k)}^2 \,. \tag{2.21}$$

This transformation has the benefit of removing the complex numbers from the diagonalized state transition matrix.

Finally, the modal representation [23, 24] of the state-space model is given by the transformed state transition matrix:

$$\tilde{\mathsf{A}} = \begin{pmatrix} a^{(1)} & 0 & \dots & 0 & 0 & 0 & 0 & 0 \dots & 0 & 0 \\ 0 & a^{(2)} & \dots & 0 & 0 & 0 & 0 & 0 \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \ddots & \vdots & \vdots \\ 0 & 0 & \dots & a^{(m_1)} & 0 & 0 & 0 & 0 \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & a_1^{(1)} & 1 & 0 & 0 \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & a_2^{(1)} & 0 & 0 & 0 \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & a_1^{(2)} & 1 \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & a_2^{(2)} & 0 \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \dots & a_1^{(m_2)} & 1 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 \dots & a_2^{(m_2)} & 0 \end{pmatrix} \tag{2.22}$$

where we have ordered the dimensions of the transformed state-space, such that dimensions corresponding to real eigenvalues come first, followed by dimensions corresponding to complex eigenvalues.[2]

Note that this matrix is block-diagonal, such that no dynamical interactions between blocks, and therefore between the corresponding AR(1) and ARMA(2,1) components, will occur; however, it has to be kept in mind that in general the dynamical noise covariance matrix S_η of the state-space model will not be block-diagonal, thereby creating instantaneous correlations between components.

2.4.3 The dynamics of AR(1) and ARMA(2,1) processes

We shall briefly discuss some dynamical properties of the components defined in the previous section. For an ARMA(2,1) process, as defined by Eq. (2.10) or in state-space representation by Eq. (2.13), the corresponding pair of eigenvalues should lie inside the unit circle of the complex plane, otherwise the dynamics would be unstable, i.e., there is a stability condition for the modulus of the eigenvalues,

[2] In the case of repeated eigenvalues, the transformation to the modal representation will not be possible, but this case is unlikely to arise for real-world data.

$0.0 < \rho < 1.0$. The closer ρ approaches the unit circle, the sharper the resonance will become; a sine wave corresponds to the limit case of $\rho = 1.0$.

The frequency-domain transfer function of an ARMA(2,1) process with AR parameters a_1, a_2 and MA parameter b_1 is given by

$$h(f) = \frac{1 + b_1 \exp(-2\pi i f)}{1 - a_1 \exp(-2\pi i f) - a_2 \exp(-4\pi i f)}, \tag{2.23}$$

where $i = \sqrt{-1}$ and $0 \leq f \leq 0.5$. The behavior of the real part of this function is shown in Fig. 2.1 for a fixed value of ϕ and a set of values for ρ. It can be seen that only for values of ρ close to 1.0 a sharp resonance peak appear. The first-order moving average term $b_1 \eta(t-1)$ produces a distortion of the curves; for the case $b_1 = 1.0$ this distortion is most pronounced, since the numerator of Eq. (2.23) becomes zero at $f = 0.5$. We remark that for ARMA(p,q) models with MA model order $q > 1$ the MA component may impose more complicated changes on the transfer function, since then zeros of the numerator may occur at any frequency.

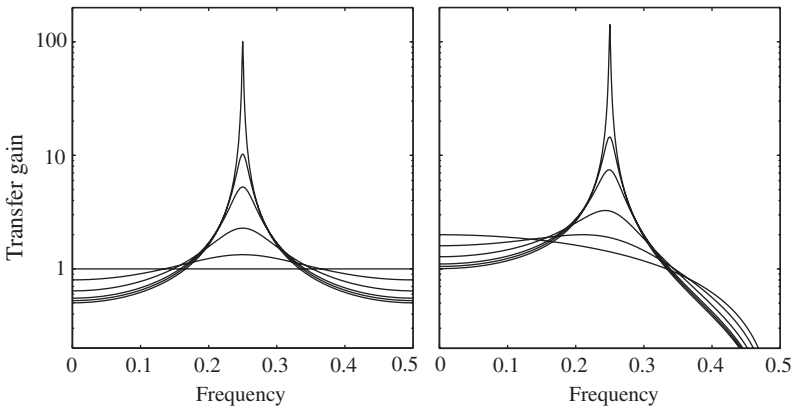

Fig. 2.1 Real part of the transfer function of an ARMA(2,1) process for resonant frequency $\phi = 0.25$, damping coefficients $\rho = 0.0, 0.5, 0.75, 0.9, 0.95, 0.995$ (curves from bottom to top at frequency 0.25) and moving average parameters $b_1 = 0.0$ (left figure) and $b_1 = 1.0$ (right figure). Note the logarithmic scale of the vertical axis.

For AR(1) processes, according to Eq. (2.19), there is only a single real eigenvalue of the transition matrix, which is equal to the first-order autoregressive parameter itself; here we denote this parameter simply by a, for ease of notation. It is obvious that also a real eigenvalue should lie inside of the unit circle, i.e., it should fulfill the stability condition $|a| < 1.0$. The case $a = 1.0$ corresponds to a random walk. AR(1) components cannot have resonant frequencies,[3] but they can serve the

[3] A somewhat pathological exception is the case $a < 0.0$ which corresponds to an oscillation with precisely the Nyquist frequency; however, this oscillation will not produce an actual resonance peak.

purpose of describing random-walk-like behavior, such as slow drifts and trends in the data, especially if a is close to 1.0.

2.4.4 State-space models with component structure

The modal representation of state-space models corresponds to a model that is organized into sets of AR(1) and ARMA(2,1) components; we can generalize this structure by allowing for higher-order components, i.e., ARMA$(p, p-1)$ components with $p > 2$, as given by Eqs (2.16), (2.17) and (2.18). For each p, up to some maximum model order, we may choose a set of n_p ARMA$(p, p-1)$ components and arrange their individual $(p \times p)$-dimensional state transition matrices on the diagonal of the state transition matrix of the state-space model with component structure. The new state dimension will then result as $M = \sum n_p p$, and the state transition matrix will again have a block-diagonal structure, with all remaining elements vanishing.[4] ARMA$(p, p-1)$ components with $p > 2$ can be regarded as summarizing a subset of the eigenvalues of the state transition matrix within one $(p \times p)$-dimensional block in left companion form.

If we intend to design a state-space model consisting of mutually independent components, we should choose for the covariance matrix of the dynamical noise S_η the same block-diagonal structure as for the state transition matrix. The corresponding blocks are then given simply by nonzero values for the variances of the AR(1) components, and by $(p \times p)$-dimensional block matrices, as shown in Eq. (2.18), for the ARMA$(p, p-1)$ components, while again all elements outside of these blocks vanish. For this model structure, there exist no ways by which correlations, instantaneous or delayed, could arise between components, except for coincidental correlations due to limited data-set size.

Finally, the $(1 \times M)$-dimensional observation matrix of the state-space model with component structure is given by

$$\mathsf{C} = \Big(c_1^{(1)}, c_2^{(1)}, \ldots, c_{n_1}^{(1)}, c_1^{(2)}, 0, c_2^{(2)}, 0, \ldots, c_{n_2}^{(2)}, 0, \\ c_1^{(3)}, 0, 0, c_2^{(3)}, 0, 0, \ldots, c_{n_3}^{(3)}, 0, 0, \ldots \Big), \quad (2.24)$$

where the $c_i^{(p)}$ are model parameters, if the corresponding dynamical noise variances σ_η^2 have been normalized to unity.

[4] As a generalization, it would be possible to use some of the elements outside the blocks for introducing coupling between components.

2.5 State-space GARCH modeling

The state-space model, as presented above, is sufficient for modeling a given time-series under the assumption of *stationarity*. In the case of *nonstationarity*,[5] the dynamical properties of the time-series data change with time; in this case, some of the model parameters would have to change their values as well, in order to adapt to these changing properties. This additional freedom may either be given to the deterministic part of the model (the first term on the rhs of Eq. (2.13), i.e., the AR term) or to the stochastic part (the second term, i.e., the MA term).

Here we choose the second option, i.e., we allow the dynamical noise covariance to change with time. By this step we approach the concept of *stochastic volatility* modeling [21], which consists of defining the dynamical noise (co)variance itself as a set of new state variables, obeying a separate stochastic dynamical model. For this additional dynamical model a new dynamical noise term is required, which renders this model estimation problem considerably more complicated; however, there exists a famous approximation to full stochastic volatility modeling, known as *generalized autoregressive conditional heteroscedastic*[6] (GARCH) modeling [4, 7]. GARCH modeling was introduced in the field of financial data analysis.

Originally, GARCH modeling was developed only for the direct modeling of data through AR/ARMA models; its core idea is to use the innovation at the previous time point, $v(t-1)$, as an estimate of the noise input to the additional volatility model. Recently, the method has been generalized to the situation of state-space modeling [8, 25]. The main problem in this generalization is given by the fact that, in the case of state-space models, we would need to employ *state prediction errors* as an estimate of the noise input, but all that is available is the set of *data prediction errors*, i.e., innovations.

2.5.1 State prediction error estimate

In order to derive a state-space version of GARCH modeling, it is necessary to derive a suitable estimator $\hat{v}_{\mathbf{x}}(t)$ of the state prediction error. The first choice for a simple estimator is given by $\hat{v}_{\mathbf{x}}(t) = \mathsf{K}(t)v(t)$, where $\mathsf{K}(t)$ denotes the $(M \times 1)$-dimensional Kalman gain matrix of a Kalman filter, used for estimating states from given time-series data; $\mathsf{K}(t)$ can be regarded as a regularized pseudo-inverse of the observation matrix C. However, in practical applications this simple estimator displays poor performance, whence we will use a refined estimator, derived in [25]:

$$\hat{v}^2_{\mathbf{x}}(t) = \mathsf{S}_\eta(t) - \mathsf{S}_\eta(t)\mathsf{C}^\dagger\sigma_v^{-2}(t)\mathsf{C}\mathsf{S}_\eta(t) + \mathsf{K}(t)v^2(t)\mathsf{K}^\dagger(t) \tag{2.25}$$

[5] We note that, within the framework of linear modeling, *nonlinearity* may be indistinguishable from *nonstationarity*.

[6] The term *heteroscedasticity* refers to the situation in which, within a set of stochastic variables, different variables have different variances. Here, the term *scedasticity*, from Greek *skedasis* for "dispersion", is yet another word for "variance".

which is, strictly speaking, an estimator of the square of the state prediction error; the square is inherited from the square in the definition of (co)variances. In Eq. (2.25), $\sigma_v^2(t)$ denotes the innovation variance, provided by the Kalman filter. From Eq. (2.25), $\hat{v}_{\mathbf{x}}^2(t)$ is a square matrix; in order to obtain the noise estimates for the individual state components, we pick out the diagonal values from this matrix. While this uniquely defines the noise terms for AR(1) components, ARMA$(p, p-1)$ components pose the problem that there are p diagonal elements; here we have chosen to simply average over these elements, but other choices would be possible. The resulting average is denoted by $\hat{v}_{\mathbf{x}}^2(k,t)$ for the kth component of a state-space model with component structure.

2.5.2 State-space GARCH dynamical equation

The design of a state-space GARCH model contains various details of implementation which need to be chosen, and in several cases it is not obvious which choice would be best; instead, practical experience is employed.[7] We found useful the particular implementation which we now describe.

In our implementation, the new time-dependent GARCH state variables correspond roughly to standard deviations, rather than variances; however, in contrast to standard deviations, these variables may also become negative. The state-space GARCH model itself is given by another ARMA(r,s) model,

$$\sigma(k,t) = \sigma(k,0) + \sum_{\tau=1}^{r} \alpha(k,\tau)\sigma(k,t-\tau) + \sum_{\tau=1}^{s} \beta(k,\tau)\hat{v}_{\mathbf{x}}^2(k,t-\tau)\,, \qquad (2.26)$$

such that for each component there is an additional set of state-space GARCH parameters $\sigma(k,0), \alpha(k,1), \ldots, \alpha(k,r), \beta(k,1), \ldots, \beta(k,s)$; these parameters become an additional part of the vector of model parameters ϑ. However, in practice we do not need a state-space GARCH model for each component of a given state-space model, but only for the particular component which actually contains the nonstationary phenomena to be modeled. For the other components we set

$$\sigma(k,0) = 1, \alpha(k,1) = \ldots = \alpha(k,r) = 0, \beta(k,1) = \ldots = \beta(k,s) = 0.$$

The choice of the GARCH model orders r,s forms again part of the model design. In the application examples to be presented in this chapter, we have decided to use $r = 1, s = 10$; experience has shown that sometimes it is advantageous to include a longer history of previous noise estimates into the model. However, in other cases also the choice $r = 1, s = 1$ has yielded good results [25]. In order to simplify the parameter estimation step, we define a constraint $\beta(k,1) = \beta(k,2) = \ldots = \beta(k,10)$,

[7] This situation is not unusual in statistical modeling of data, since it will rarely be possible to set up a model which faithfully reproduces the structure of the underlying natural processes; rather, models have always to be regarded as approximations. At least, this is the situation we are facing in the study of systems of enormous complexity, such as the human brain.

such that in effect we are using an average of the last 10 noise estimates and just one MA parameter. In Refs. [8] and [25], state-space GARCH models were introduced in which the logarithm of the variance, $2\log\sigma(k,t)$, was used as GARCH state variable, but in this chapter we have decided to formulate the model directly in the standard deviations $\sigma(k,t)$.

2.5.3 Interface to Kalman filtering

At each time point t, the current value of the GARCH state variable, $\sigma(k,t)$, is passed through a "nonlinear" observation function by taking the square, thereby becoming a genuine non-negative variance $\sigma^2(k,t)$; this variance then replaces, for component k, the term σ_η^2 which appears in Eqs (2.15, 2.18) of the stationary state-space model. The corresponding dynamical noise covariance matrix of component k then enters the block-diagonal covariance matrix of the state-space model at the appropriate block position of component k, such that this matrix itself becomes time-dependent.

This step represents a major modification of the usual Kalman filter iteration, since the continuous changes of one of the main matrices of the model prevent the filter from reaching its steady state.

2.5.4 Some remarks on practical model fitting

The generalized state-space models discussed in this chapter are parametric models, consisting of a model structure and a parameter vector ϑ. The following table lists the parameter sets contained in ϑ, also giving the dimension of each set:

Description	Symbol	Dimension[8]
state transition matrix parameters[9]	$a^{(k)}, \phi^{(k)}, \rho^{(k)}$	$m_1 + 2m_2$
moving-average parameters	b_i	m_2
observation matrix parameters	c_i	$m_1 + m_2$
observation noise variance	σ_ε^2	1
GARCH parameters	$\sigma(k,0), \alpha(k,\tau), \beta(k,\tau)$	$r+s+1$ or[10] $r+2$
initial state vector	$\mathbf{x}(0)$	$m_1 + 2m_2$

[8] In the table, m_1 and m_2 denote the number of real eigenvalues and of pairs of complex eigenvalues, respectively, regardless of how these eigenvalues are distributed over the ARMA$(p, p-1)$ components of the state-space model.

[9] Optimizing $\phi^{(k)}, \rho^{(k)}$ instead of the corresponding AR parameters $a_1^{(k)}, a_2^{(k)}$ has the advantage that the stability constraint $\rho^{(k)} < 1.0$ can be directly imposed; furthermore, prior knowledge about the frequencies $\phi^{(k)}$ can be conveniently incorporated into the model, or particular frequencies can be excluded from the optimization process.

[10] if the constraint $\beta(k,1) = \beta(k,2) = \ldots = \beta(k,s)$ is applied

For a given choice of ϑ, the Kalman filter provides the corresponding value of the likelihood. Model fitting consists of maximizing the likelihood, or, more conveniently, the logarithmic likelihood, with respect to ϑ by numerical optimization [9]. For this purpose, we are employing standard optimization algorithms, namely the Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton algorithm and the Nelder-Mead simplex algorithm. Sometimes the simplex algorithm can be employed in situations where the quasi-Newton algorithm fails due to numerical problems.

Several optimization steps should be iterated, such that in some steps the optimization is limited to subsets of parameters. With the sole exception of σ_ε^2, all parameters can be uniquely assigned to one of the components; we recommend performing a series of optimization steps such that each step is confined to one component. Some optimization steps may also be confined to state transition matrix parameters, or to observation matrix parameters, etc.

A good initial model is of crucial importance for successful modeling. We recommend fitting an autoregressive model, AR(p), of sufficiently high model order, say $p = 30$, to the given data; fitting of pure AR models, without MA terms, can be done very efficiently by standard least-squares regression [5]. This model is then converted into a state-space model, as discussed above, and the resulting state-space model is transformed into its modal representation; thereby a model consisting of a set of AR(1) and ARMA(2,1) components is obtained. Later, higher-order ARMA components can be created by merging pairs of these AR(1) and/or ARMA(2,1) components. The dynamical noise covariance matrix is constrained to the same block-diagonal structure as the state transition matrix by setting all other elements to zero.

At this point there is a need for subjective interference with the modeling process: usually a subset of the initial components will capture the most important features of the data and of the underlying dynamics, such as frequencies known to play an important role, or prominent time-domain patterns, while other components will describe rather unspecific activity. Only this subset of important components should be selected as initial model, while the remaining components should be discarded. Also, the decision as to which components, if any, are to be merged later to form higher-order ARMA components depends on subjective assessment of the dynamics represented by the components.

Keeping all components from the modal representation would also be possible, but it would result in a very large model with many redundant components; such a model could be employed as an alternative initial model, and later the model could gradually be "pruned" during the optimization process, but this procedure would be very demanding in terms of computational time consumption.

For the observation noise variance σ_ε^2 and the state-space GARCH parameters, no initial values can be obtained by this approach. For σ_ε^2, a small initial value should be chosen, maybe about 10^{-3} times smaller than the variance of the data, unless we have reason to assume that there was considerably more observation noise in the data. Larger initial values for σ_ε^2 may create the risk that the Kalman filter would incorrectly allocate a large fraction of the variance of the data to the

observation noise term. The procedure for the state-space GARCH parameters is described below.

For the application examples which will be discussed below, the dimension of ϑ is about 35; about 10 of these parameters form the initial state vector $\mathbf{x}(0)$. We recommended to keep $\mathbf{x}(0)$ initially at zero and to optimize only the remaining parameters, except for the state-space GARCH parameters, while omitting the contributions of the first (approximately) 20 data points to the likelihood, in order to allow a transient of the Kalman filter to die out. Once a first set of optimization steps has been applied, such that approximate estimates of the main parameter sets of the model have been obtained, the full likelihood is evaluated and the initial state vector is included into the remaining optimization steps.

During the first part of the model fitting procedure, there should not yet be any state-space GARCH models, i.e., the state-space GARCH parameters should be fixed as $\sigma(k,0) = 1$, $\alpha(k,1) = 0$, $\beta(k,1) = 0$, $\beta(k,2) = 0$, ... After the estimates of the other parameter sets have converged to stable values, it can be decided which component should be given a state-space GARCH model. Usually, the nonstationary behavior to be modeled is represented only by one or possibly two components, and only these components should be given state-space GARCH models. Experience has shown that if state-space GARCH models are given to all components of a state-space model, components tend to become blurry and featureless, since too much freedom is available to each component. After estimates of the state-space GARCH parameters have been obtained, also all other model parameters need to be refitted, since the introduction of state-space GARCH models may considerably change the dynamics of the complete model.

In many cases, we probably cannot expect to reliably find global maxima in a 25-dimensional, highly heterogeneous parameter space. After the optimization procedure, the Hessian matrix at the obtained solution should routinely be computed, in order to check for the possibility of saddle points; nevertheless we may find only local maxima. Refined studies of the geometry of these parameter spaces would be needed, in order to obtain additional insight into this problem. However, we expect that for practical purposes a good solution will be almost as useful as the perfect solution. In the end, the properties of the innovations will always allow an assessment of the quality of the obtained model; major problems during the optimization step will usually also be reflected in the innovations.

2.6 Application examples

In the remaining part of this chapter we will discuss the application of state-space modeling, with component structure and state-space GARCH components, to three examples of EEG time-series; all contain nonstationary phenomena: in the first example, due to the transition from the conscious state to anesthesia; in the second, due to the transition from one sleep stage into another; and in the third, due to the occurrence of an epileptic seizure.

2.6.1 Transition to anesthesia

As the first example we choose an EEG time-series recorded from a patient being anesthesized (with propofol) prior to surgery. Sampling rate was $f_s = 100$ Hz. From the full clinical data set we select the time-series recorded at the T4 electrode versus average reference; we select $T = 2000$ sample points, starting with the moment when induction of anesthesia was begun. The data are shown in Fig. 2.2A; the same data were also analyzed in [25]. It can be seen that the qualitative appearance of the trace changes within the 20 seconds covered by this time-series, i.e., the data contain pronounced nonstationarity: high-amplitude oscillations in the delta frequency range gradually become stronger, corresponding to the loss of consciousness. The transition from the conscious state to anesthesia may be regarded as a phase transition in brain dynamics [22].

We model the data by a state-space model consisting of $m_2 = 5$ mutually independent ARMA(2,1) components; the model is fitted by maximizing the log-likelihood until convergence. It is found that one of the components represents the gradually increasing delta range oscillation; in a second modeling step, a state-space GARCH model is added to this component, but not to the remaining four components. The state-space GARCH model orders are $r = 1$, $s = 10$, but we apply to the MA parameters the constraint introduced above, $\beta(k,1) = \beta(k,2) = \ldots = \beta(k,10)$. The three additional parameters of the state-space GARCH model are also fitted by maximizing the log-likelihood; then the other sets of model parameters are refitted, starting at their previous non-GARCH values, in order to allow the model to adapt to the presence of the state-space GARCH model. Joint and alternate optimization of state-space GARCH model parameters and other parameters are iterated a few times, again until convergence.

The resulting five components are shown in Fig. 2.2B; together they represent a decomposition of the data of Fig. 2.2A. The figure shows smoothed state estimates, as obtained by a standard Rauch-Tung-Striebel smoother [19] which performs a backward pass through the time-series; during optimization only the forward pass of the Kalman filter is performed, since it is this pass which transforms the data into innovations and thereby produces a value for the likelihood.

Note that in Fig. 2.2B all components are displayed with the same variance, such that their dynamical properties can be compared; their actual variances in state-space will differ considerably, since we have chosen to normalize the variances of the dynamical noises to 1, instead of the variances of the estimated states[11].

In Fig. 2.2B components are ordered according to increasing frequency; this is possible since all components are modeled by ARMA(2,1) processes, such that there is a single resonance frequency for each component. At the top we find the nonstationary delta range component (labeled c1), with frequency[12] $f = 0.422$ Hz and

[11] The effective variances of the time-series of estimated state components do not represent model parameters, therefore they would be inaccessible for the purpose of normalization.

[12] The physical frequency f is related to the phase ϕ of the corresponding pair of complex eigenvalues, as defined by Eq. (2.20), by $\phi = 2\pi f/f_s$, where f_s denotes the sampling frequency of the data.

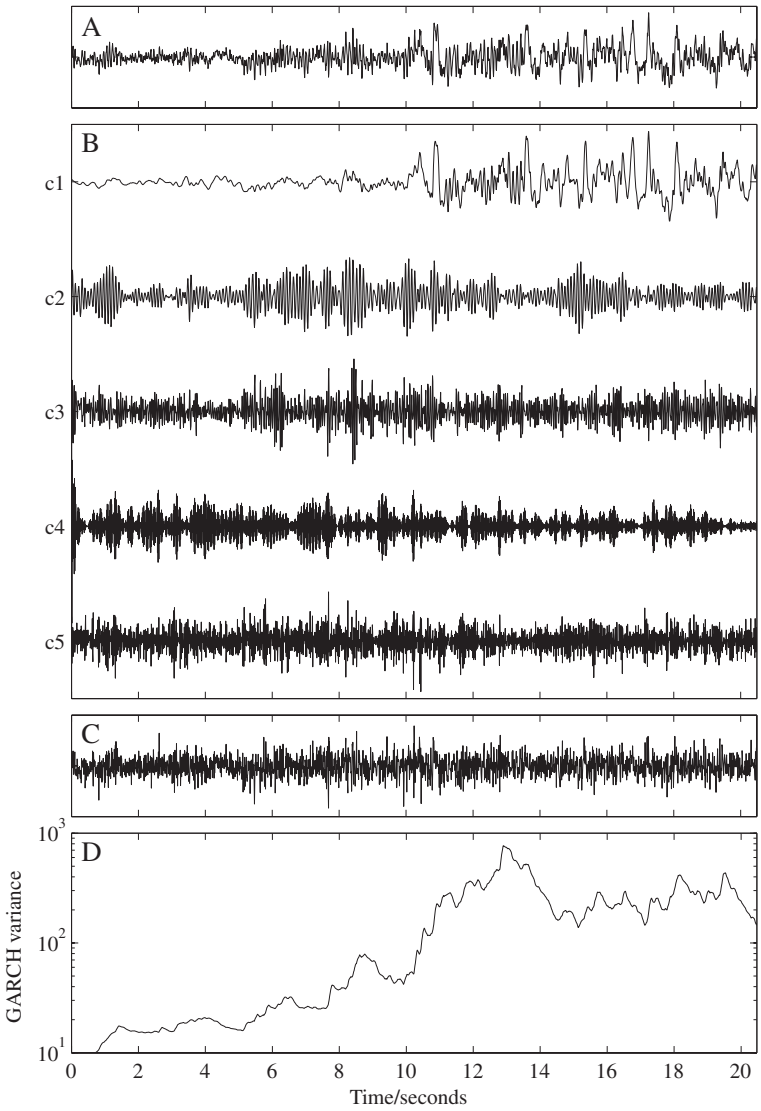

Fig. 2.2 EEG time-series with transition to anesthesia: Data (A); state-space decomposition (B); innovations (C); and state-space GARCH variance of component c1 (D). Vertical axes for all graphs in subfigures A, B and C have been rescaled individually for convenience of graphical display. Resonance frequencies f and damping coefficients ρ of components: c1: $f = 0.422$ Hz, $\rho = 0.690$; c2: $f = 10.495$ Hz, $\rho = 0.946$; c3: $f = 17.463$ Hz, $\rho = 0.772$; c4: $f = 45.34$ Hz, $\rho = 0.910$; c5: $f = 48.649$ Hz, $\rho = 0.292$.

damping coefficient $\rho = 0.690$; the gradual increase of delta amplitude is clearly visible. The next two components (c2 and c3) represent alpha and beta range components, with $f = 10.495$ Hz, $\rho = 0.946$ for c2; and $f = 17.463$ Hz, $\rho = 0.772$ for c3. The remaining two components (c4 and c5) represent high-frequency noise components, with $f = 45.34$ Hz, $\rho = 0.910$ for c4; and $f = 48.649$ Hz, $\rho = 0.292$ for c5. Note that for this data set the Nyquist frequency lies at $f_s/2 = 50$ Hz.

In Fig. 2.2C the weighted innovations are shown, confirming that little, if any, structure has remained in the innovations. The raw innovations (prediction errors) of the state-space model have been weighted by being divided at each time point by the square root of the corresponding innovation variance, as provided by the Kalman filter; remember that in presence of a state-space GARCH model the Kalman filter will not reach its steady state, such that also the innovation variance (or, more generally, covariance) will not converge to a constant value.

Finally, in Fig. 2.2D the time-dependent variance of the delta range component is shown, as described by the state-space GARCH model. Note that the vertical axis of this figure is logarithmic. This graph should be studied together with the delta range component itself, the top graph in Fig. 2.2B. It can be seen that the variance increases from values around 20 in the first few seconds to values around 200–300 at the end of the time-series; this increase may be interpreted as a data-derived quantitative representation of the phase transition process.

At the beginning of the time-series, the dynamics of the variance of the delta range component was initialized at an arbitrary value of 1.0, from which it has to rise to more realistic values during a short transient which is not explicitly resolved in the figure. The maximum-likelihood estimates of the state-space GARCH parameters are $\sigma(k,0) = 0.0837, \alpha(k,1) = 0.975, \beta(k,1\ldots10) = 4.111 \times 10^{-5}$.

2.6.2 Sleep stage transition

The second example is given by an EEG time-series recorded from the surface of a fetal sheep brain (144 days gestation age). The original sampling rate was 250 Hz, but we decide to subsample the data to $f_s = 125$ Hz. A single electrode is selected. Out of a longer data set, a subset of $T = 50000$ sample points (at 125 Hz) is selected, covering a transition between slow-wave sleep (SWS) to REM sleep. The data are shown in Fig. 2.3A. The transition is discernible by a decrease of signal amplitude with concomitant fading of the characteristic slow-wave activity.

For modeling the sleep data we choose the same model structure as used for the anesthesia study, i.e., we choose a state-space model consisting of $m_2 = 5$ mutually independent ARMA(2,1) components; the model is fitted by maximizing the log-likelihood until convergence. Again it is found that only one of the components captures the nonstationary behavior representing the sleep stage transition; in a second modeling step, a state-space GARCH model is added to this component, but not to the remaining four components. State-space GARCH model orders are the same

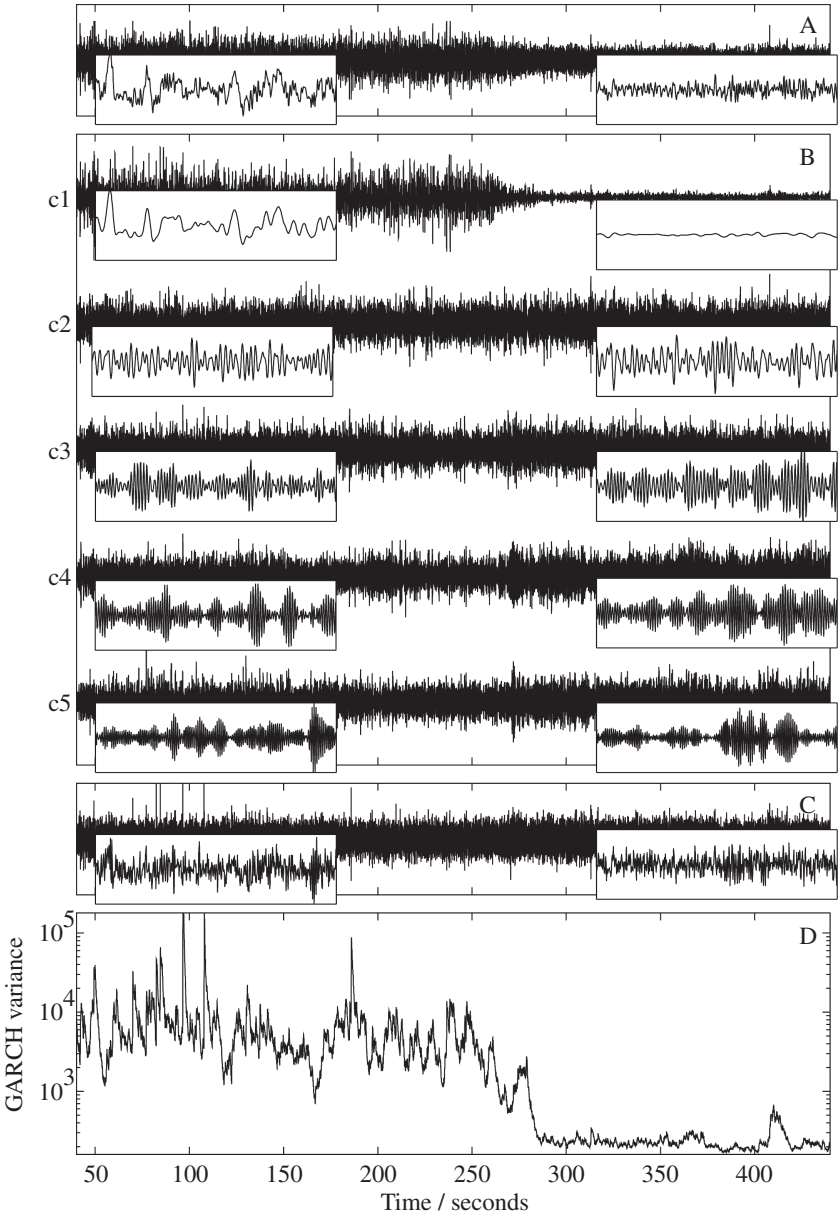

Fig. 2.3 EEG time-series from fetal sheep brain with transition from slow-wave sleep to REM sleep: Data (A); state-space decomposition (B); innovations (C); and state-space GARCH variance of component c1 (D). Vertical axes for all graphs in subfigures A, B and C have been rescaled individually for convenience of graphical display. Resonance frequencies f and damping coefficients ρ of components: c1: $f = 3.811$ Hz, $\rho = 0.910$; c2: $f = 11.465$ Hz, $\rho = 0.882$; c3: $f = 18.796$ Hz, $\rho = 0.926$; $f = 24.896$ Hz, $\rho = 0.951$; c5: $f = 30.801$ Hz, $\rho = 0.945$. Insets show enlarged parts of data, state-space components and innovations: 100–104 s (left) and 400–404 s (right).

as for the anesthesia example: $r = 1$, $s = 10$, and the same constraint for the MA parameters is employed. Model parameters are optimized until convergence.

The resulting five components are shown in Fig. 2.3B, ordered according to increasing frequency; again, smoothed state estimates are shown, rescaled to the same variance. The first component, labeled c1, is the nonstationary component; its frequency and damping coefficient is $f = 3.811$ Hz, $\rho = 0.910$. For the remaining components, frequencies and damping coefficients are $f = 11.465$ Hz, $\rho = 0.882$ for c2; $f = 18.796$ Hz, $\rho = 0.926$ for c3; $f = 24.896$ Hz, $\rho = 0.951$ for c4; and $f = 30.801$ Hz, $\rho = 0.945$ for c5. The Nyquist frequency lies at $f_s/2 = 62.5$ Hz. Note that damping coefficients for all components are fairly close to 1.0, indicating pronounced oscillatory behavior.

The weighted innovations and the time-dependent variance of the nonstationary component are shown if in Figs. 2.3C and 2.3D, respectively. It can be seen that the variance decreases from values around 5000 in the first part of the time-series (representing slow-wave sleep) to values around 200 in the latter part (representing REM sleep). If the variance is used as a quantitative measure for the transition between the two sleep stages, the time point at which the transition occurs can be identified to within a time-interval of no more than 5 s; however, note that also within each of the two sleep stages there are slow changes of the variance which may reflect changes of the underlying physiological state.

Also for this model, the dynamics of the variance of the nonstationary component was initialized at a value of 1.0, from which it rises to appropriate values around 5000 during a short transient. The maximum-likelihood estimates of the state-space GARCH parameters are $\sigma(k,0) = 0.176, \alpha(k,1) = 0.985, \beta(k,1\ldots10) = 2.68 \times 10^{-6}$.

In this time-series, we have the example of a nonstationarity where a state with large variance passes to a state with smaller variance; we remark that we were also able to model data displaying the opposite situation, i.e., the transition from REM sleep to slow-wave sleep, from the same experiment (the same fetus) with the same model class.

2.6.3 Temporal-lobe epilepsy

As the third example we choose an EEG time-series recorded from a patient suffering from temporal-lobe epilepsy, during awake resting state with open eyes. Sampling rate was $f_s = 200$ Hz. From the full clinical data set we select the time-series recorded at the Fz electrode versus linked earlobes; out of a longer data set, we select $T = 2048$ sample points, covering one short generalized epileptic seizure of a type characteristic for temporal-lobe epilepsy. The data are shown in Fig. 2.4A. In the figure, it can be seen that at time near 7 s the qualitative appearance of the trace changes abruptly, with a series of periodic high-amplitude spike-wave patterns emerging; these patterns are typical of the ictal regime (containing the seizure), while the earlier part of the trace represents the preictal regime. The transition from

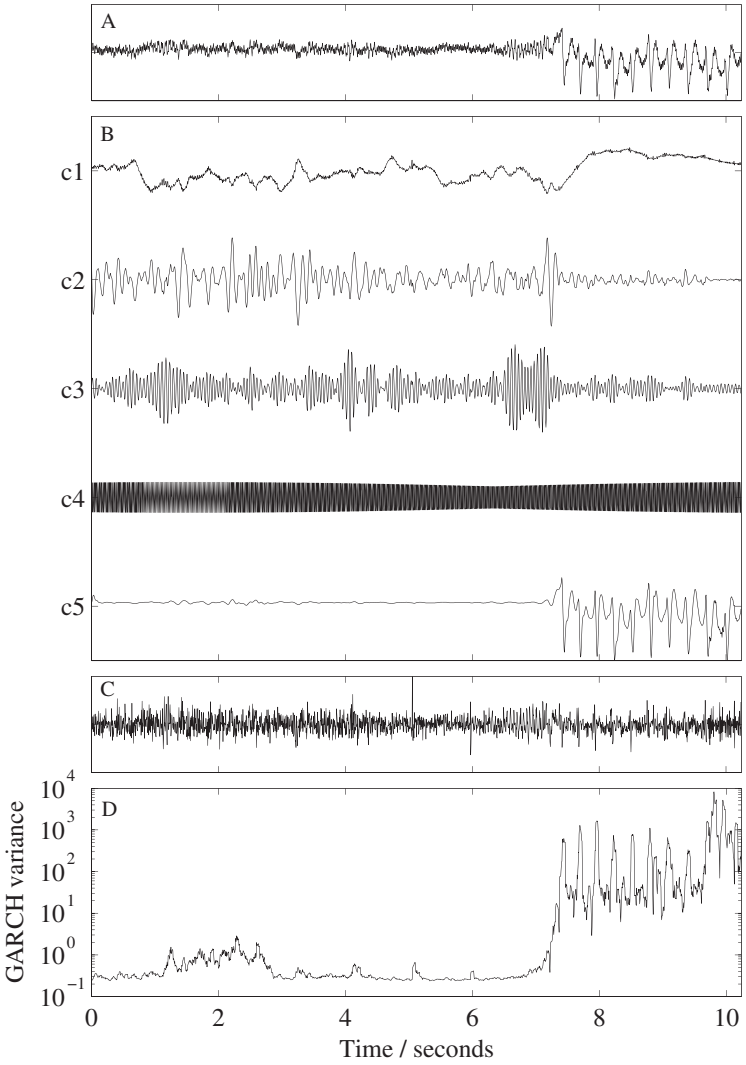

Fig. 2.4 EEG time-series with epileptic seizure: Data (A); state space decomposition (B); innovations (C); and state-space GARCH variance of component c5 (D). Vertical axes for all graphs in subfigures A–C have been rescaled individually for convenience of graphical display. AR-parameter of component c1 is $a_1 = 0.985$. Resonance frequencies f and damping coefficients ρ of other components: c2: $f = 7.978\,\mathrm{Hz}$, $\rho = 0.876$; c3: $f = 17.288\,\mathrm{Hz}$, $\rho = 0.976$; c4: $f = 50.025\,\mathrm{Hz}$, $\rho = 1.0$; component c5 has frequencies $f = 3.274\,\mathrm{Hz}$, $f = 18.171\,\mathrm{Hz}$ and corresponding damping coefficients $\rho = 0.870$, $\rho = 0.883$.

the preictal to the ictal regime has recently been discussed by Milton *et al.* [15] in analogy with phase transitions in physics.

We model the data by a state-space model consisting of one AR(1), three ARMA(2,1) and one ARMA(4,3) components (corresponding to $m_1 = 1$ real eigenvalues and $m_2 = 5$ complex-conjugated pairs of eigenvalues); this structure is chosen according to the transformation of an initial AR model into modal representation which reveals at least two components representing the epileptic seizure; these two components are merged into a single fourth-order component. Again, the initial state-space model is fitted by maximizing the log-likelihood until convergence. The ARMA(4,3) component, representing the seizure activity, is then provided with a state-space GARCH model, while the remaining four components are not. Again, the state-space GARCH model orders are $r = 1$, $s = 10$, and the same constraint for the MA parameter as before is employed. Fitting of the three additional parameters and refitting of the other sets of model parameters proceeds in the same way as for the earlier anesthesia and fetal sleep examples.

The resulting five components are shown in Fig. 2.4B, ordered according to increasing frequency, with the ARMA(4,3) component at the bottom of the figure; together these components represent a decomposition of the data of Fig. 2.4A. Also this figure shows smoothed state estimates; again, for convenience of graphical display, variances of components have been normalized to the same value.

At the top of the figure, the single AR(1) component is shown, labeled c1; its state transition parameter a_1 is 0.985, and thereby well suited for describing slow drifts and trends in the data. In this time-series, there seems to be a slow shift of potential during the seizure; we see that the AR(1) component captures this shift well, thereby facilitating the modeling of the oscillatory pattern during the seizure by another component. In the preictal regime, the first-order component also captures some unspecific low-frequency activity.

Below the AR(1) component, we see in Fig. 2.4B the three ARMA(2,1) components, with frequencies and damping coefficients $f = 7.978$ Hz, $\rho = 0.876$ for c2, $f = 17.288$ Hz, $\rho = 0.976$ for c3, and $f = 50.025$ Hz, $\rho = 1.0$ for c4; the Nyquist frequency lies at $f_s/2 = 100$ Hz. Components c2 and c3 represent alpha- and beta-range components, respectively; the beta activity is clearly visible in the data. Component c4 represents the frequency of the electrical power supply, i.e., an artifact of technical origin; the damping coefficient of $\rho = 1.0$ clearly reveals an undamped oscillation.

At the bottom of Fig. 2.4B, the ARMA(4,3) component representing the epileptic seizure is shown; it can be seen that this component displays only weak activity until the seizure commences. The seizure itself is well extracted, without leakage into the other components. The two frequencies of this component are $f = 3.274$ Hz and $f = 18.171$ Hz; the corresponding damping coefficients are $\rho = 0.870$ and $\rho = 0.883$. It is obvious that the first of these frequencies describes the main periodicity of the ictal spike–wave patterns.

In Fig. 2.4C the weighted innovations are shown; again they are weighted by being divided at each time point by the square root of the corresponding innovation variance. While it can be seen that most of the structure has been removed, there are

still some remains of seizure-related structure in the innovations. This can be seen most clearly from the series of sharp spikes in the innovations which correspond well with the epileptic spikes in the data. The last 40 samples of the innovations are probably dominated by muscle artifact effects.

Finally, in Fig. 2.4D the time-dependent variance of the epileptic seizure component is shown, as described by the state space GARCH model. Note that again the vertical axis of this figure is logarithmic. This graph should be studied together with the epileptic seizure component itself, the bottom graph in Fig. 2.4B. Again, at the beginning of the time-series, the dynamics of the variance was initialized at an arbitrary value of 1.0; the variance then drops to a somewhat smaller value and mostly stays close to this value for several seconds, until the seizure commences. The maximum-likelihood estimates of the state space GARCH parameters are $\sigma(k,0) = 0.465, \alpha(k,1) = 5.044 \times 10^{-3}, \beta(k,1 \ldots 10) = 3.941 \times 10^{-3}$; from these values it is not surprising that the variance stays close to the constant term $\sigma(k,0) = 0.465$, as long as the innovations remain small.

However, as soon as the seizure starts, the variance rises to values of almost 10^3; then, the variance oscillates roughly between 10 and 10^3, thereby following the spike–wave oscillation of the seizure. We thus have two regimes of different behavior of variance, preictal and ictal; if the transition between these two regimes is regarded as a phase transition, the concurrent rise of the variance may again be interpreted as a data-derived quantitative representation of this phase transition process. We emphasize that no prior information—relating to either the components in the data or the timing of seizure onset—was given to the algorithm.

Also in the preictal regime, the time-dependent variance shows some structure, such as a transient increase of variance between 1.0 and 3.5 s into the time-series; whether this structure actually reveals relevant information on the epileptic seizure component cannot be decided on the basis of the analysis of just a single seizure.

2.7 Discussion and summary

For centuries, the ability to make quantitative predictions has been regarded as one of the ultimate goals of science. Our present work, which aims to construct predictive models for particular brain phenomena that are accessible to direct observation, is motivated by the same goal.

Much is now known about the elementary constituents of the human brain: the neurons, synapses, neurotransmitters and ion channels. In principle, it should be possible to use this knowledge to set up a detailed model of the dynamics of brain; such a model would allow reliable predictions of the observable phenomena generated by the brain. However, due to the enormous numbers of these constituents and the complexity of their interconnections, this is not (yet) a practicable task.

Alternatively, a predictive model may be set up predominantly or exclusively based on the available data, and this is the path we have followed in this chapter. More specifically, we have studied how such a model can be set up for the

purpose of efficient description of transitions between qualitatively different dynamical states, i.e., of *nonstationary* behavior. The resulting models summarize various useful statistics about the data, mainly encoded in the properties of the state-space components, into which the data are decomposed. Each component is characterized by one or several resonant frequencies, but also by the corresponding damping coefficients; furthermore the total power in the data is distributed over the components in a specific way. Nonstationary components are modeled by additional state-space GARCH models, and the time-dependent variance information provided by such models offers additional information on the processes underlying the data; it may be used also for purposes of automatic classification and event detection. In particular, phase transitions represent an example of nonstationary processes; thus the time-dependent variance may serve as a data-derived quantitative representation of the underlying phase transition processes.

In this chapter, we have sketched a systematic approach to building state-space models for univariate time-series data; the generalization to multivariate data is straightforward. State-space models are predictive models, mapping the time-series data to a time-series of prediction errors, denoted as *innovations*. The innovation approach to data modeling aims at whitening the data, i.e., at removing all correlations from the innovations; this is the condition for the validity of the expression for the logarithmic likelihood of the data given by Eq. (2.7).

The innovations are also a source of information for further improvement of models; a good example is given by the third application example of this chapter. Epileptic spike-wave patterns are known to be difficult to model by autoregressive models [16]; the strongly anharmonic waveforms, in combination with poor stability of the main frequency, pose considerable challenges. An improved model for the epileptic seizure component, possibly incorporating also nonlinear elements, should be able to reduce the amount of seizure-related residual structure in the innovations which is visible in Fig. 2.4C; alternatively, or additionally, the design of the state-space GARCH model may be further improved.

The choice of the model order of certain components represents a question of model design, i.e., the choice of model structure; a related problem is that of model comparison. This is a much more difficult problem than estimating model parameters within a fixed model structure, and a full discussion would go beyond the scope of this chapter. For the purpose of time-series decomposition and characterization of nonstationarities, we have found the approach of fitting a set of mutually independent ARMA$(p, p-1)$ components useful; the choice of the number of components and their model orders will, to some extent, remain a subjective decision. However, such subjective decisions may be partly based on prior knowledge about the properties of physiologically meaningful components, or of well-known artifacts.

Fitting larger models with larger numbers of model parameters will usually improve the likelihood, when compared with smaller models. It is well known that this effect invites the risk of overfitting, against which the maximum-likelihood method itself has no protection. Information criteria like the Akaike Information Criterion (AIC) [1] or the Bayesian Information Criterion (BIC) [20] have been introduced, for replacing the likelihood $L\big(\vartheta; y(1), \ldots, y(T)\big)$, or, more precisely,

replacing $-2\log L\big(\vartheta; y(1),\ldots,y(T)\big)$; these criteria contain a penalty term for the number of model parameters, such that it can be decided whether the improvement of likelihood resulting from extending a model is worth the price of additional model complexity. Recently, *logarithmic evidence* has been proposed as an alternative for AIC and BIC [17].

For the application examples presented in this chapter we have not reported detailed values of log-likelihood, AIC or BIC; but we remark that the comparison of both AIC and BIC for the best non-GARCH models with the final models including state-space GARCH modeling has consistently favored the latter models.

Information criteria like AIC or BIC are best known as tools for estimating optimal model orders for model classes like AR(p) models; but in fact these measures permit the comparison of the performance of models in a much wider setting, such as non-nested models, or even, with respect to their structure, mutually unrelated models. Then, in principle, the process of model design could be based completely on comparison of such criteria, instead on subjective decisions; the problem here is that, for each competing model, a time-consuming numerical optimization procedure would have to take place before the values of the criteria would become available; this would make such an approach very time-consuming. But the power of information criteria for quantitative model comparison should be kept in mind.

Also, other design choices of the modeling algorithm discussed in this chapter could be investigated in the light of information criteria. As an example, we again mention details of the implementation of state-space GARCH modeling, such as model orders, or the choice of the estimator for the state prediction errors, Eq. (2.25); the quadratic estimator which we have employed, following [25], draws its main justification from its superior performance in practical applications, also in terms of information criteria, as compared to other estimators.

Use of state-space GARCH modeling to describe nonstationary structure in time-series—in the absence of prior information on the timing of the nonstationary changes—represents a comparatively new approach that will require more study, both in simulations and applications, in order to become an established tool for time-series analysis. In this chapter we have demonstrated its rich potential for modeling phase transitions and other nonstationary behavior in electroencephalographic time-series data.

Acknowledgments The work of A. Galka was supported by Deutsche Forschungsgemeinschaft (DFG) through SFB 654 “Plasticity and Sleep”. The anesthesia EEG data set was kindly provided by W.J. Kox and S. Wolter, Department of Anesthesiology and Intensive Care Medicine, Charité-University Medicine, Berlin, Germany, and by E.R. John, Brain Research Laboratories, New York University School of Medicine, New York, USA. The fetal sleep data set was kindly provided by A. Steyn-Ross, Department of Engineering, University of Waikato, New Zealand. The epilepsy EEG data set was kindly provided by K. Lehnertz and C. Elger, Clinic for Epileptology, University of Bonn, Germany.

References

1. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Contr. **19**, 716–723 (1974)
2. Akaike, H., Nakagawa, T.: Statistical Analysis and Control of Dynamic Systems. Kluwer, Dordrecht (1988)
3. Åström, K.J.: Maximum likelihood and prediction error methods. Automatica **16**, 551–574 (1980)
4. Bollerslev, T.: Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics **31**, 307–327 (1986), doi:10.1016/0304-4076(86)90063-1
5. Box, G.E.P., Jenkins, G.M.: Time Series Analysis, Forecasting and Control, 2. edn. Holden-Day, San Francisco (1976)
6. Durbin, J., Koopman, S.J.: Time Series Analysis by State Space Methods. Oxford University Press, Oxford, New York (2001)
7. Engle, R.F.: Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. Econometrica **50**, 987–1008 (1982), doi:10.2307/1912773
8. Galka, A., Yamashita, O., Ozaki, T.: GARCH modelling of covariance in dynamical estimation of inverse solutions. Physics Letters A **333**, 261–268 (2004), doi:10.1016/j.physleta.2004.10.045
9. Gupta, N., Mehra, R.: Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations. IEEE Trans. Autom. Contr. **19**, 774–783 (1974), doi:10.1109/TAC.1974.1100714
10. Hamilton, J.D.: Time Series Analysis. Princeton University Press, Princeton, New Jersey (1994)
11. Kailath, T.: An innovations approach to least-squares estimation – Part I: Linear filtering in additive white noise. IEEE Trans. Autom. Control **13**, 646–655 (1968), doi:10.1109/TAC.1968.1099025
12. Kailath, T.: Linear Systems. Information and System Sciences Series. Prentice-Hall, Englewood Cliffs (1980)
13. Kalman, R.E.: A new approach to linear filtering and prediction problems. J. Basic Engin. **82**, 35–45 (1960)
14. Lévy, P.: Sur une classe de courbes de l'espace de Hilbert et sur une équation intégrale non linéaire. Ann. Sci. École Norm. Sup. **73**, 121–156 (1956)
15. Milton, J.G., Chkhenkeli, S.A., Towle, V.L.: Brain connectivity and the spread of epileptic seizures. In: V.K. Jirsa, A.R. McIntosh (eds.) Handbook of Brain Connectivity, pp. 477–503. Springer-Verlag, Berlin, Heidelberg, New York (2007)
16. Ozaki, T., Valdes, P., Haggan-Ozaki, V.: Reconstructing the nonlinear dynamics of epilepsy data using nonlinear time-series analysis. J. Signal Proc. **3**, 153–162 (1999)
17. Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J.: Comparing dynamic causal models. NeuroImage **22**, 1157–1172 (2004), doi:10.1016/j.neuroimage.2004.03.026
18. Protter, P.: Stochastic Integration and Differential Equations. Springer-Verlag, Berlin, Heidelberg, New York (1990)
19. Rauch, H.E., Tung, G., Striebel, C.T.: Maximum likelihood estimates of linear dynamic systems. American Inst. Aeronautics Astronautics (AIAA) Journal **3**, 1445–1450 (1965)
20. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978)
21. Shephard, N.: Statistical aspects of ARCH and stochastic volatility. In: D.R. Cox, D.V. Hinkley, O.E. Barndorff-Nielsen (eds.) Time Series Models in Econometrics, Finance and Other Fields, pp. 1–67. Chapman & Hall, London (1996)
22. Steyn-Ross, M.L., Steyn-Ross, D.A., Sleigh, J.W., Wilcocks, L.C.: Toward a theory of the general-anesthetic-induced phase transition of the cerebral cortex. I. A thermodynamics analogy. Phys. Rev. E **64**, 011917 (2001), doi:10.1103/PhysRevE.64.011917
23. Su, G., Morf, M.: Modal decomposition signal subspace algorithms. IEEE Trans. Acoust. Speech Signal Proc. **34**, 585–602 (1986)

24. West, M.: Time series decomposition. Biometrika **84**, 489–494 (1997)
25. Wong, K.F.K., Galka, A., Yamashita, O., Ozaki, T.: Modelling nonstationary variance in EEG time-series by state space GARCH model. Computers Biol. Med. **36**, 1327–1335 (2006), doi:10.1016/j.compbiomed.2005.10.001

http://www.springer.com/978-1-4419-0795-0

Modeling Phase Transitions in the Brain
(Eds.)D.A. Steyn-Ross; M. Steyn-Ross
2010, Approx. 350 p. 103 illus., 24 in color., Hardcover
ISBN: 978-1-4419-0795-0