Signal Processing 80 (2000) 1607–1627

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

Some wavelet-based analyses of Markov chain data

D.R. Brillinger[a,*], P.A. Morettin[b], R.A. Irizarry[c], C. Chiann[b]

[a]*Statistics Department, University of California, Berkeley, CA 94720, USA*
[b]*Statistics Department, University of São Paulo, SP 05315-970, Brazil*
[c]*Biostatistics Department, Johns Hopkins University, Baltimore, MD 21205, USA*

Received 19 April 1999; received in revised form 21 December 1999

Abstract

This work considers signals whose values are discrete states. It proceeds by expressing the transition probabilities of a nonstationary Markov chain by means of models involving wavelet expansions and then, given part of a realization of such a process, proceeds to estimate the coefficients of the expansion and the probabilities themselves. Through choice of the number of and which wavelet terms to include, the approach provides a flexible method for handling discrete-valued signals in the nonstationary case. In particular, the method appears useful for detecting abrupt or steady changes in the structure of Markov chains and the order of the chains. The method is illustrated by means of data sets concerning music, rainfall and sleep. In the examples both direct and improved estimates are computed. The models include explanatory variables in each case. The approach is implemented by means of statistical programs for fitting generalized linear models. The Markov assumption and the presence of nonstationarity are assessed both by change of deviance and graphically via periodogram plots of residuals. © 2000 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Diese Arbeit betrachtet Signale, deren Werte diskrete Zustände sind. Sie fährt fort, indem die Übergangswahrscheinlichkeiten einer nichtstationären Markov-Kette anhand von Modellen, die Wavelet-Entwicklungen beinhalten, ausgedrückt werden, und macht dann damit weiter, die Koeffizienten der Entwicklung und der Wahrscheinlichkeiten selbst zu schätzen, wobei ein Teil einer Realisierung eines solchen Prozesses gegeben sei. Durch die Wahl, wieviel und welche der Wavelet-Terme zu berücksichtigen sind, liefert diese Vorgehensweise eine flexible Methode, um wertdiskrete Signale im nichtstationärem Fall zu behandeln. Insbesondere scheint die Methode nützlich zu sein, um abrupte oder stetige Änderungen in der Struktur von Markov-Ketten und die Ordnung der Ketten zu entdecken. Die Methode wird anhand von Musik-, Regen- und Schlafdaten veranschaulicht. In den Beispielen werden sowohl direkte als auch verbesserte Schätzungen berechnet. Die Modelle beinhalten in allen Fällen erklärende Variablen. Die Methode wird mit Hilfe von statistischen Programmen zur Anpassung verallgemeinerter linearer Modell implementiert. Die Markov-Annahme und die Gegenwart der Nichtstationarität werden sowohl durch die Änderung der Abweichung als auch graphisch durch Periodogrammdarstellungen der Residuen bewertet. © 2000 Elsevier Science B.V. All rights reserved.

Résumé

Ce travail considère les signaux dont les valeurs sont des états discrets. Il procède en exprimant les probabilités de transition d'une chaîne de Markov non stationnaire au moyen de modèles impliquant des expansions en ondelettes et

* Corresponding author. Tel.: + 1-510-642-2781; fax: + 1-510-642-7892.
E-mail address: brill@stat.berkeley.edu (D.R. Brillinger).

PII: S0165-1684(00)00097-9

ensuite, étant donné une partie de la réalisation d'un tel processus, procède à l'estimation des coefficients de l'expansion et des probabilités elles-mêmes. Par le choix du nombre et des coefficients à inclure, l'approche fournit une méthode flexible pour manipuler des signaux à valeurs discrètes dans un cas non stationnaire. En particulier, le modèle se révèle utile pour détecter des changements abrupts et réguliers dans des chaînes de Markov et l'ordre des chaînes. La méthode est illustrée au moyen des ensembles de données concernant la musique, la chute de pluie et le sommeil. Dans les exemples nous calculons à la fois les estimateurs directs et améliorés. L'approche est implémentée au moyen de programmes statistiques pour l'ajustement de modèles linéaires généralisés. La supposition de Markov et de non stationnarité est évaluée à la fois par un changement de déviation et de façon graphique via des courbes de périodogrammes des résidus.

Keywords: Wavelets; Transition probabilities; Nonstationary processes; Markov chains; Improved estimates

This work presents empirical analyses of nonstationary Markov chain models, based on wavelet expansions, for signals taken from musicology, meteorology and sleep research, respectively. A basic goal is looking for time-varying characteristics of the various series, such as trend and/or changing (seasonal) effects. The work proceeds from an initial analysis of the transition probabilities into the coefficients of a wavelet expansion. This is followed by an estimation of the coefficients and a synthesis to obtain estimates of the transition probabilities themselves. The fitted characteristics may be used to assess stationarity, e.g. detecting points of change amongst other things. Through choice of the number of and just which wavelet terms to include in the linear predictor the approach provides a flexible method for handling signals of discrete-state-valued observations amongst other possibilities.

The work may be viewed as involving a nonlinear model within a linear model setup. Specifically transition probabilities, $P_{ab}(t)$, of movement from state a to state b are expressed as functions of a linear predictor in t, by means of wavelet expansions and link functions. Generalized linear model methodology and computing programs are employed in the empirical analyses.

Markov processes, in particular Markov chains, have long been basic to signal processing. One can mention their use in cryptology, coding, networks, speech, control, image processing for example. In the last decade wavelets have also become basic to many areas of signal processing. Since wavelets provide economical expansions for a wide class of functions, this implies for example that they provide good compression of signals and images.

In this work Markov chains and wavelet techniques are married together to deal with nonstationary processes. These two fields have been joined together before, e.g. by Crouse and Baraniuk [15] which concerns hidden Markov modes, but the present work concerns discrete-valued processes and has a different intent.

The next section provides pertinent basic background on Markov chains, wavelets, the model and its analysis. Section 3 describes the data sets, Section 4 presents the results of the analyses and the paper ends with some general discussion.

1. Background

1.1. The Markov chain case

The concern is signals that take on a discrete set of values. A homogeneous or stationary Markov chain with A states is a random process, $Y(t)$, taking on values in the set $\{1,2,\dots,A\}$, such that the conditional probabilities of taking on values at the next time step, given the whole history of the process, depend solely on the present value. Specifically,

$$\begin{aligned}\mathrm{Prob}\{Y(t+1)=b|Y(t)=a,\,Y(t-1)=a_{-1},\\ Y(t-2)=a_{-2},\dots\}\\ =\mathrm{Prob}\{Y(t+1)=b|Y(t)=a\}=P_{ab} \qquad (1)\end{aligned}$$

for $b,a,a_{-1},a_{-2},\dots\in\{1,\dots,A\}$ and $t=1,2,\dots$. The circumstance (1) is called the Markov property and also appears in the dynamic equation of the state space model, so common in signal processing today.

The matrix $\boldsymbol{P} = [P_{ab}]$ is called the transition probability matrix. It, with a set of initial conditions $\text{Prob}\{Y(t) = a\} = P_a$, determines the process in the sense that probabilities of sample realizations $\{Y(0), Y(1), Y(2), \ldots, Y(T-1)\}$ may be set down. Results have been developed concerning first passage times, limiting behavior, communicating states, etc. by various authors, e.g. Feller [22] and Dynkin [20]. Applications may be found in [1,2].

If the probabilities P_a and P_{ab} depend on the particular time point, the chain is nonstationary and $P_{ab}(t)$ will denote the conditional probability of being in state b at time t, given that the process was in state a at time $t-1$ while $P_a(t)$ will denote the marginal probability of being in state a at time t. It will be supposed that the state of the process has been observed at the T successive times, $t = 1,2,\ldots,T$.

In many cases a set of parameters, reduced from the full set $\{P_a(t), P_{ab}(t)\}$, is required, particularly if A is not small and the amount of data is limited. The approach adopted here is to employ a linear parameterization of some function of the P's, e.g. to write

$$\text{logit}\{P_{ab}(t)\} = \sum_{j,k} \beta_{abjk} \psi_{jk}(t) \tag{2}$$

with the β's unknown parameters to be estimated and the ψ's given functions (here use of $\text{logit}\{\pi\} = \log(\pi/(1-\pi))$ provides a simple manner to ensure that the probability stays between 0 and 1). At the next step this expression is substituted into a likelihood function such as (3) below and the β's estimated by maximizing the likelihood. There may be a further step of shrinkage of the coefficient estimates, that is replacement of an estimate $\hat{\beta}$ by a value closer to 0 in an attempt to improve the estimate. In expansion (2), in this work, ψ's are the functions of some wavelet basis as discussed below.

In defining the likelihood function it is convenient to replace the process $Y(t)$, $t = 0,1,2,\ldots$ by a vector-valued process $\boldsymbol{X}(t) = [X_a(t)]$ where $X_a(t) = 1$ if $Y(t) = a$ and $X_a(t) = 0$ otherwise. It satisfies $\sum X_a(t) = 1$ and $P_{ab}(t) = \text{Prob}\{X_b(t) = 1 | X_a(t-1) = 1\}$. Also one sets $P_a = \text{Prob}\{X_a(0) = 1\}$.

Let $X_{ab}(t) = 1$, if the process is in state a at time $t-1$ and in state b at time t, and $X_{ab}(t) = 0$ otherwise. Given the data and parametric forms for $P_a(t), P_{ab}(t)$ the likelihood is now

$$\left[\prod_{a=1}^{A} P_a^{X_a(0)}\right]\left[\prod_{t=1}^{T}\prod_{a=1}^{A}\prod_{b=1}^{A} P_{ab}(t)^{X_{ab}(t)}\right] \tag{3}$$

viewed as a function of the parameters. In the case that $A = 2$ things may be simplified. Write $\pi_1(t) = P_{11}(t)$, $\pi_2(t) = P_{22}(t)$, then $P_{12}(t) = 1 - \pi_1(t)$, $P_{21}(t) = 1 - \pi_2(t)$ and the likelihood is

$$P_1^{X_1(0)} P_2^{X_2(0)} \prod_{t=1}^{T} \{\pi_1(t)^{X_{11}(t)}[1-\pi_1(t)]^{X_{12}(t)} \pi_2(t)^{X_{22}(t)} \times [1-\pi_2(t)]^{X_{21}(t)}\}. \tag{4}$$

In a variety of cases, e.g. T large, the first two terms, may be neglected. This will be done in the results presented. The estimation criterion then becomes

$$\prod_{t=1}^{T} \{\pi_1(t)^{X_{11}(t)}[1-\pi_1(t)]^{X_1(t-1)-X_{11}(t)} \pi_2(t)^{X_{22}(t)} \times [1-\pi_2(t)]^{X_2(t-1)-X_{22}(t)}\} \tag{5}$$

as a function of the unknown parameters. When consideration below turns to estimation, it is useful to note that this has the form of a likelihood based on independent Bernoullis, that is random variables taking on the values 0,1 with some probability π. In consequence, the log of the criterion is the sum of a term in $\pi_1(t)$ and one in $\pi_2(t)$ each corresponding to a binomial distribution. Standard statistical packages, allowing generalized linear model fitting of Binomials, may now be employed to compute estimates of the β's of (2).

A variety of properties of maximum-likelihood estimates have been developed for Markov chains in the large sample case. For example, Billingsley [3] developed consistency and asymptotic normality results for a stationary finite-dimensional parameter Markov chain. Foutz and Srivastava [23] and Ogata [35] derived the large sample distribution of the maximum-likelihood estimate in the stationary ergodic case. Bishop et al. [5] suggested some methods for assessing empirically whether a Markov chain is stationary. Fahrmeir and Srivastava [21] indicated how nonstationary Markov chain models might be included within the generalized linear modelling methodology. Details of this are provided below. Coe and Stern [14] presented empirical analyses involving nonstationary Markov chain models. McCullagh and Nelder [31, Section 8.4.3], discussed the Coe and Stern work.

Consideration now turns to the wavelet methodology basic to the model being studied.

1.2. Wavelets

Wavelets are contemporary approximation tools, alternative to existing basis systems such as sines and cosines, Walsh functions, etc.

The basic fact about wavelets is that they are *localized* in time (and space), contrary to what happens with the trigonometric functions used in Fourier analysis. This behavior makes wavelets ideal for the analysis of nonstationary signals, particularly those with transients or singularities. Fourier bases are localized in frequency but not in time; small changes in some of the observations may induce substantial changes in almost all the components of a Fourier expansion, a fact that does not hold for basic wavelet expansions and can be a real disadvantage.

In elementary wavelet analysis there are two basic functions, the *scaling function* (or father wavelet) ϕ and the *wavelet* ψ. Here ϕ is a solution of the two-scale difference equation

$$\phi(t) = \sqrt{2}\sum_{k\in Z} h_k \phi(2t-k) \tag{6}$$

and normalized via $\int\phi(t)\,\mathrm{d}t = 1$, while ψ is defined by

$$\psi(t) = \sqrt{2}\sum_{k\in Z} (-1)^k h_{1-k}\phi(2t-k). \tag{7}$$

Here Z is the set of all integers and the h_k's are filter coefficients which can be chosen in such a way that one has wavelets with desirable properties. Defining $\phi_{\ell k}(t) = 2^{\ell/2}\phi(2^{\ell}t-k)$ and $\psi_{jk}(t) = 2^{j/2}\psi(2^j t-k)$, for example the system $\{\phi_{\ell k}(t)\}_{k\in Z}\cup\{\psi_{jk}(t)\}_{j\geqslant\ell;k\in Z}$ forms an orthonormal basis for the space of square integrable functions on the real line $L_2(\Re)$, under some additional conditions on the filter coefficients. Accordingly, any $f\in L_2(\Re)$ can be expanded as

$$f(t) = \sum_{k\in Z}\alpha_{\ell k}\phi_{\ell k}(t) + \sum_{j\geqslant\ell}\sum_{k\in Z}\beta_{jk}\psi_{jk}(t), \tag{8}$$

where ℓ is the "coarse" level of the approximation and the wavelet coefficients are given by

$$\alpha_{\ell k} = \int f(t)\phi_{\ell k}(t)\,\mathrm{d}t, \qquad \beta_{jk} = \int f(t)\psi_{jk}(t)\,\mathrm{d}t, \tag{9}$$

following the orthonormality. On occasion nonorthogonal functions are used and one speaks of frame analysis.

In practice for each wavelet analysis, empirical versions of the wavelet coefficients are defined. For example, they can be defined as least-squares estimates, say $\hat{\alpha}_{\ell k},\hat{\beta}_{jk}$, that is, minimizers of

$$\sum_{t=1}^{n}\left[f(t) - \sum_{k=0}^{2^{\ell}-1}\alpha_{\ell k}\phi_{\ell k}(t) - \sum_{j=\ell}^{J-1}\sum_{k=0}^{2^{j}-1}\beta_{jk}\psi_{jk}(t)\right]^2 \tag{10}$$

with J appropriately chosen. In this work nonlinear smoothing (thresholding or shrinkage) rules are applied to the coefficients $\hat{\beta}_{j,k}$ to obtain improved estimators.

Several issues are of interest here:

(i) the choice of the wavelet basis,
(ii) the choice of a shrinkage policy,
(iii) the estimation of the scale parameter (noise level).

A brief discussion of these follows. For further details see for example [10–13,30,33].

(i) Concerning the choice of the wavelet basis, some possibilities are the Haar functions and the compactly supported wavelet bases of Daubechies [16]. Other examples are the Morlet and Mexican hat wavelets, which generate frames under specific conditions. For these wavelets equations (6)–(8) no longer hold and the coefficients in (9) are obtained using a dual frame. See [16] for details on frames.

The problem and the form of the signal to be analyzed may suggest a particular basis. In the examples to be presented here, the Haar expansion will be used, having in mind its simplicity of interpretation and its ability to detect abrupt temporal changes. The Haar expansion is based on the choices

$$\phi(t) = 1, \quad 0 \leqslant t < 1, \tag{11}$$

$$\psi(t) = \begin{cases} 1, & 0 \leqslant t < \frac{1}{2}, \\ -1, & \frac{1}{2} \leqslant t < 1. \end{cases} \tag{12}$$

Expansion (8) is then, more simply

$$f(t) = \alpha_{00} + \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \beta_{jk}\psi_{jk}(t) \tag{13}$$

for some J. It may be remarked that in this case the fitted values simply correspond to assuming the function is constant at the coarsest resolution employed.

(ii) By *shrinkage* is meant the replacement of an estimated coefficient, $\hat{\beta}_{jk}$, by a shrunken value $\hat{\beta}^*_{jk} = w(\hat{\beta}_{jk}/s_{jk})\hat{\beta}_{jk}$, for some function $w(.)$ such that $w(u) \approx 1$ for large $|u|$ and ≈ 0 for small $|u|$, and with s_{jk} an estimated standard error of $\hat{\beta}_{jk}$. The function $w(.)$ is meant to dampen down the variability of $\hat{\beta}_{jk}$ and not to introduce too much bias. The estimated function will be, in the Haar case,

$$\hat{f}^*(t) = \hat{\alpha}_{00} + \sum_j \sum_k \hat{\beta}^*_{jk}\psi_{jk}(t). \tag{14}$$

Various criteria have been suggested for the choice of $w(.)$. For example Blow and Crick [6], using a mean-squared error criterion, were led to the function

$$w(u) = \frac{\sqrt{\pi}}{2}\left[I_0\left(\frac{u^2}{2}\right) + I_1\left(\frac{u^2}{2}\right)\right]\mathrm{e}^{-u^2/2} \tag{15}$$

with the I_j Bessel functions. Tukey [37] suggested the use of

$$w(u) = (1 - 1/u^2)_+, \tag{16}$$

which weights to zero any terms with $|\hat{\beta}_{jk}|$ less than its standard error and smoothly downweights larger values. This is the $w(\cdot)$ used in the examples presented below.

Donoho and Johnstone [17–19], motivated by considerations of risk, work with functions of the form $\delta_{\lambda_n}(\hat{\beta}_{jk})$, with $\lambda_n \to \infty$ as $n \to \infty$, e.g. $\lambda_n = s_{jk}\sqrt{2\log n}$. Here s_{jk} is the estimated standard deviation of $\hat{\beta}_{jk}$. Other forms of shrinkage rules might be used to improve estimates, as the Sure Shrink [18] and a cross-validation procedure [34]. See [33] for further suggestions. It remains to be learned when these various choices are particularly appropriate and for which practical situations.

In practice, ranges of values of j, k in (2) need to be selected. Here the various j, k terms will have varying weights, as a result of employing shrinkage, and in a sense this alleviates the problem of choice of range for j, k.

(iii) In the case of (15) or (16), s_{jk}, an estimate of the standard deviation of $\hat{\beta}_{jk}$, is needed. For a signal plus stationary noise model, Brillinger [8] bases such estimates on an estimate of the power spectrum of the errors. In the present examples output from a standard generalized linear model program may be used. Details are given below.

The present work will consider principally a logit for the probabilities and a wavelet-based regression function, as in (2). Of course functions other than the logit may be used, see McCullagh and Nelder [31].

In practice in the definitions the time period of observation will be shrunk to the unit interval by working in terms of the variate t/T.

1.3. The model and its implementation

Given a stretch of data from a two-state Markov chain, with transition probabilities $P_{ab}(t)$, in the empirical examples presented the estimation criterion (5) will be used. What is further needed is a specific model for the $\pi_a(t)$, $a = 1,2$.

Fahrmeir and Kaufmann [21] and Kaufmann [29] present a maximum-likelihood approach for statistical inference concerning categorical-valued time signals possessing certain forms of Markov structure. Their model allows the inclusion of explanatory variables. These authors develop consistency and asymptotic normality properties of the estimates amongst other things. The model may be written as

$$\mathrm{Prob}\{X_a(t) = 1 \mid \boldsymbol{X}(t-1), \boldsymbol{X}(t-2), \dots\} = h_a(\boldsymbol{Z}(t)^{\tau}\boldsymbol{\beta})$$

for $a = 1, \dots, A-1$, where $\boldsymbol{X}(t) = [X_a(t)]$, $\boldsymbol{h}: R^{A-1} \to R^{A-1}$, is one-to-one, and $\boldsymbol{Z}(t)$ is a function of past observations and fixed explanatories and $\boldsymbol{\beta}$ is a vector of unknown parameters. In this case, it will include the wavelet coefficients (see

Eq. (19) below). Higher-order Markov chains may be included by inserting interaction terms such as $X_a(t-1)X_b(t-2)$ into the linear predictor, $\boldsymbol{Z}(t)^{\tau}\boldsymbol{\beta}$.

To be specific, consider the two-state ($A=2$) and Haar wavelets case. Model (2) may be written as

$$\text{Prob}\{Y(t)=a|Y(t-1)=a\}$$
$$=\pi_a(t)=h\left\{\alpha_a+\sum_{j=0}^{J_a}\sum_{k=0}^{2^j-1}\beta_{ajk}\psi_{jk}(t)\right\}, \qquad (17)$$

where $a=1,2$ with h for example the inverse of the logit transform as in (2). This model falls within the framework of the Fahrmeir–Kaufmann work defining $X_a(t)=1$ if $Y(t)=a$ and $X_a(t)=0$ otherwise for $a=1,\dots,A-1$. Assuming that the preceding model is correct and that J_a is finite, the results of Fahrmeir and Kaufmann show that the usual maximum-likelihood large sample standard error formulae are appropriate asymptotically.

The s_{ajk}, i.e. the standard error estimates for the $\hat{\beta}_{ajk}$, will be required in the formation of shrunken estimates. They are typically part of the output of maximum-likelihood programs. These values (and estimated covariances) may be used to estimate the variances of derived estimates, e.g. of the transition probabilities of the Markov model. This is what has been done in the examples presented below.

In [7,8] it is proposed to estimate the uncertainty of a shrunken wavelet estimate of a mean function by acting as if the weights, $w(s_{ajk}/\hat{\beta}_{ajk})$ are constant, really more nearly constant in the sense that the major variability comes from the $\hat{\beta}_{ajk}$. This is what has been done in the examples presented below. It is also acted as if the J_a were constant.

It is crucial to assess the goodness of fit of models employed as, for example, the Markov assumption is basic for the development. The nonstationarity described by wavelet expansions leads to a generalized linear model, so techniques proposed for that case may be employed. These include: deviance analysis and various types of residual analysis. In particular, since temporal dependence is a principal basic concern, an examination of the periodogram of the residuals may prove insightful in considering alternatives of stationary dependence.

Further details of the computations and definitions are given in the appendix.

2. The data sets

Consideration now turns to applying the above modelling procedure to some observed signals of interest.

2.1. Music

Markov processes have been used in finding structure in music, see for example [36,26,28]. For example musicologists have tried to model melodies as kth order Markov chains. These methods have generally failed to capture the essence of melodies for two reasons. Firstly, they miss the global structure of the music and secondly, they assume stationarity, a characteristic that melodies definitely do not seem to possess.

In [27] a stochastic composition is created using a five-state Markov model (big jump up, small jump up, no jump, small jump down, big jump down) to generate the intervals between notes of the melody. A 5×5 transition probability matrix, estimated from simple melodies, is used. It was noticed that, although the melody sounded fine for small stretches of time, it lacked direction and seemed repetitive. Use of a nonstationary transition probability matrix may "improve" such stochastic compositions. In this work, as a preliminary study, a simple two-state (jump, no jump) model will be employed. A jump occurring at time t is related to a note starting at that time. This representation is then equivalent to the rhythm of the melody. Stretches with many consecutive notes can be referred to as an *intense* part of the melody.

The example to be considered involves the first 128 measures of the rhythm of the soprano line of J.S. Bach's unfinished fugue, *Contrapunctus XIV* from Die Kunst der Fuge. To begin, it is necessary to put such data into the form considered in the paper. To this end temporal subdivisions of a measure are set up. The smallest has been called a *tatum* [4]. In this particular fugue the smallest subdivision of the beat is a 16th note (a note of one-sixteenth the duration of a measure). However, 16th notes are used only as embellishments so to be able to study the structure of the piece in terms of the intense parts, a tatum will be defined to be an eighth-note and a two-state time series will be

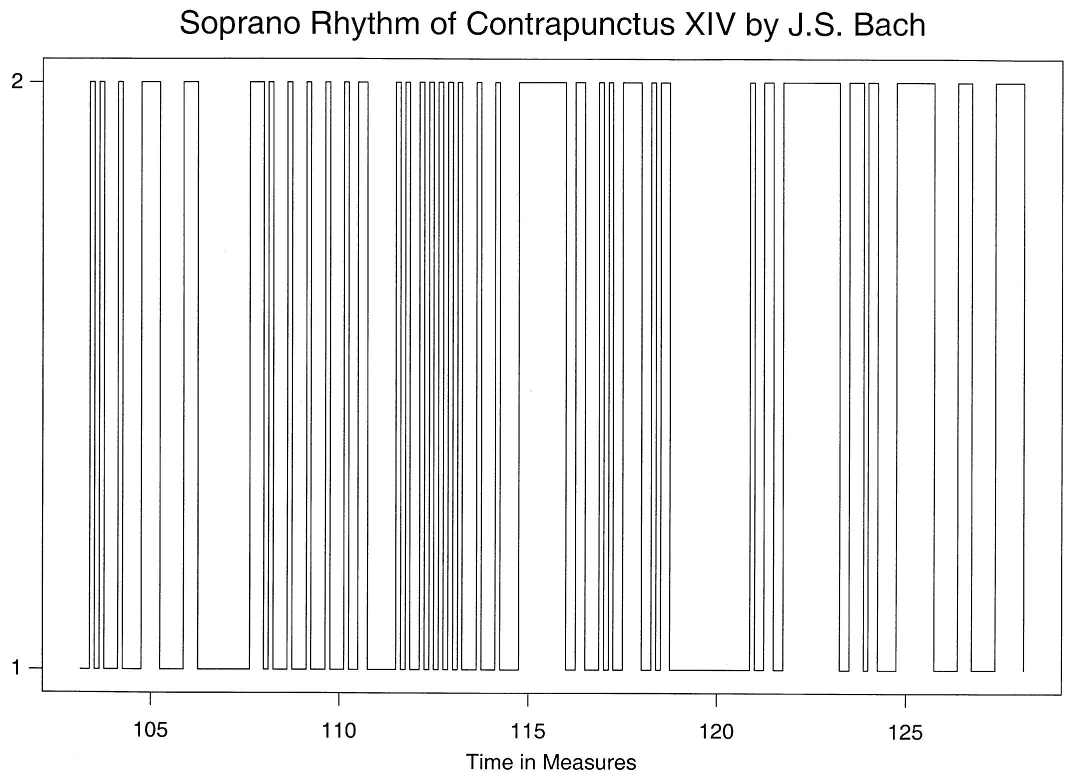

Fig. 1. Times of the beginnings of notes for the Soprano line in Bach's unfinished fugue for measures 80–111. The value 2 corresponds to a new note starting.

defined via

$$Y(t) = \begin{cases} 2 & \text{if the beginning of a note} \\ & \quad \text{occurs in tatum } t, \\ 1 & \text{no new note in tatum } t. \end{cases} \tag{18}$$

There are then $T = 1024$ observations in total. Fig. 1 presents some data from towards the end of the piece. The event of a new note starting corresponds to the level 2. One notices, for example, a number of stretches of constant level.

Questions that might be addressed here include: can wavelet analysis usefully describe nonstationarity present? Does temporal dependency exist beyond the modeled nonstationary Markov?

Brillinger and Irizarry [9] and Irizarry [27] contain more details on the quantification and statistical analysis of music.

2.2. Snoqualmie Falls rain

For the present work Peter Guttorp provided daily data concerning whether or not at least 0.01 in of rain had occurred at Snoqualmie Falls, Washington, for each day for the period 1963 to 1977. He had analyzed the January data [24] and in particular fit two-state stationary Markov chains of orders 1 and 2. Guttorp restricted consideration to January values in order to obtain realizations of an approximately stationary process. In the present work all the days and months are studied.

The data for the year 1963 is graphed in Fig. 2 with $Y = 1$ when no rain and $Y = 2$ when rain. One sees stretches of both wet and dry in winter and summer as was to be anticipated.

Questions of interest include: Is the seasonal, that is annual changing effect? Are there changes in the signal structure?

2.3. Sleep research

Mallo et al. [32] investigated the sleep–awake behavior of a boy from the age of five weeks to four years. The procedure consisted of recording waking and sleep states via direct observation by the mother or eventually by a maid. When carried out

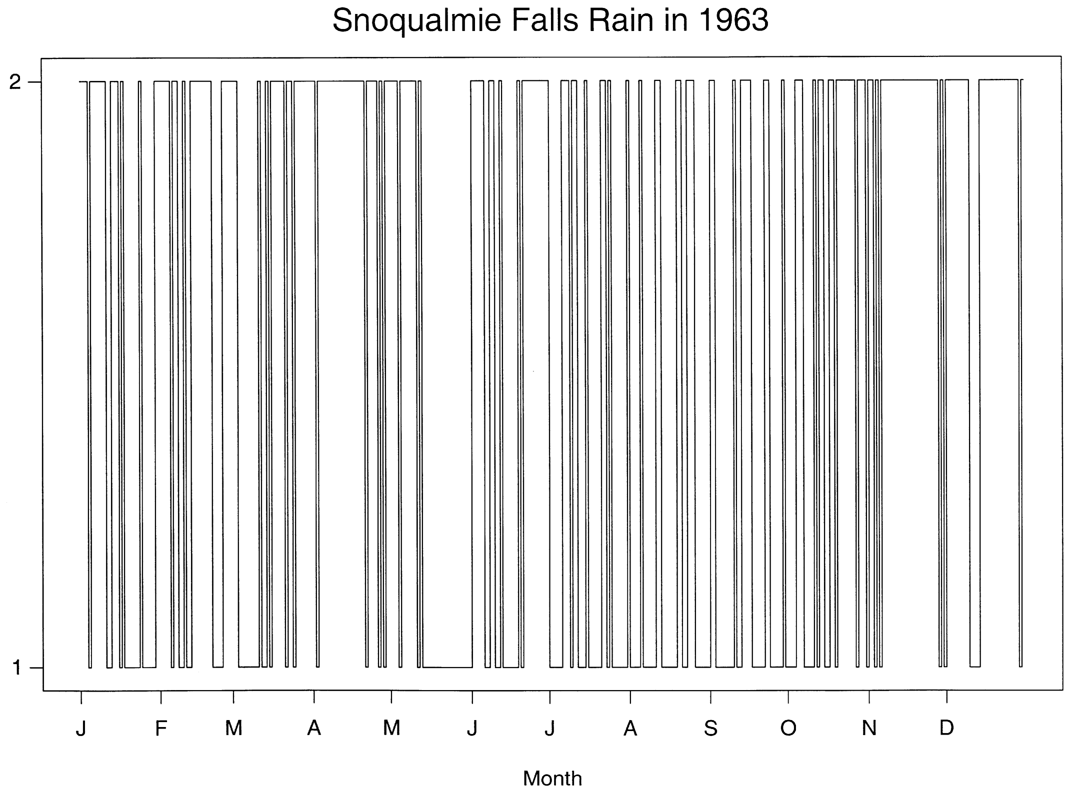

Fig. 2. The rain data. Value 2 corresponds to a day with rain and 1 to a day with none.

the measurements were done at intervals of 10 min. The values 2 and 1 were assigned to the sleep and awake states, respectively. In the present work only the data for the age of five to six weeks, are studied. There are $T = 2016$ values. Fig. 3 shows the plot of a segment of the data. Once again stretches of constancy may be noted with the child asleep and awake for approximately equal lengths of time. Examination of the data, for example by periodogram analysis, shows a period of 24 h as might have been anticipated.

Questions of interest include: Is a simple Markov process an acceptable model? Is the 24 h periodicity changing in character?

3. Results of fitting the Markov models

3.1. The music data

Fig. 1 provided a segment of some baroque music data. To create the data each measure was divided into eight tatums. In most of western music tatums are given names within the measure. In this case, tatums 1,3,5, and 7 are called beats 1,2,3, and 4, respectively. Beats 1 and 3 are called strong beats and beats 2 and 4 are called weak beats. The remaining tatums (2,4,6, and 8), the tatums between beats, are called subdivisions of the beat. In baroque music it is common to have notes starting on the beat, rather than the subdivisions of the beat. Furthermore it is usually more likely to have a note start on a strong beat rather than a weak beat. The terms $\gamma_{as}x_{as}(t), \ldots$ in the model below are "beat" explanatories inserted to handle these phenomena. Specifically define the indicator variables

$$x_{as}(t) = 1 \quad \text{when } t \bmod 4 = 1,$$

$$x_{aw}(t) = 1 \quad \text{when } t \bmod 4 = 3,$$

$$x_{asd}(t) = 1 \quad \text{when } t \bmod 4 = 0 \text{ or } 2$$

with s referring to strong, w referring to weak and sd to subdivision of the beat. The variate values are

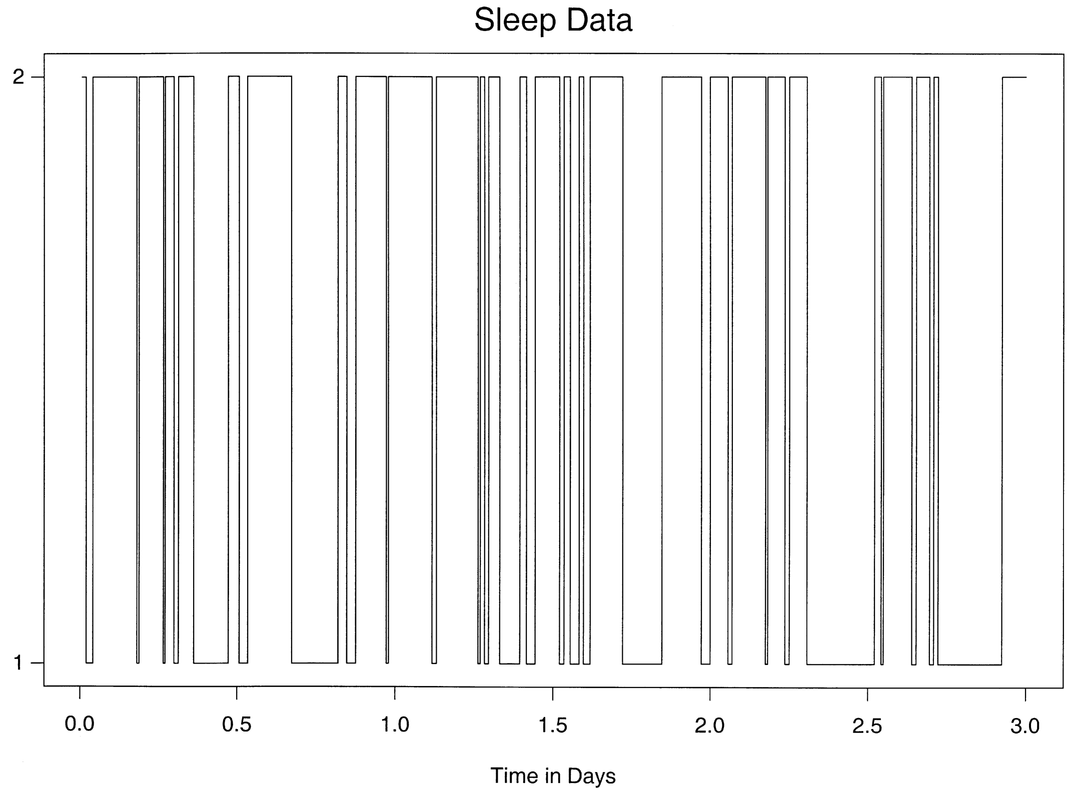

Fig. 3. The sleep data. The value 2 corresponds to the child being asleep, 1 to awake.

0 otherwise. The model is now the following:

$$\pi_a(t) = h\left\{ \sum_{j=1}^{J_a} \sum_{k=0}^{2^j-1} \beta_{ajk}\psi_{jk}(t) + \gamma_{as}x_{as}(t) + \gamma_{aw}x_{aw}(t) + \gamma_{asd}x_{asd}(t) \right\}, \tag{19}$$

where $a = 1{,}2$ with h the inverse of the logit transform and with $J_1, J_2 = 3$.

Fig. 4 is based on the data of measures 95 and 96 of the piece and provides the transition probabilities as estimated by substituting the maximum-likelihood estimates of the β, γ into (19). For example $\hat{P}_{11}(t)$ gives the probability of the process remaining in state 1 at the next time point, while $\hat{P}_{12}(t)$ estimates the probability of moving from state 1 to 2. The estimates for measures 80–95 will be the same as for measure 95 because they have the same Z's associated with them (All Haar expansions will be constant in that interval). Similarly 96–111 will have the same estimated transitions as 96. For this reason in Fig. 4 it is only necessary to show 95 and 96, as opposed to 80–111. Showing the whole stretch would not be successful because of the substantial variation. In the plot S refers to a strong beat and W to a weak one. The Fig. 5 provides the wavelet part of the linear predictor. This figure is useful for examining the non-stationarity of the data as in particular it includes marginal $\pm$ 2 s.e. limits about the beat level. In the present case, as was anticipated from the context, there is evidence of nonstationary transition probabilities. At the same time various values are within, or nearly within the $\pm$ 2 s.e. limits, suggesting that improved estimates might be obtained via shrinkage.

Fig. 6 is the same as the previous figure, but with the shrunken estimates. The figure has narrowed, but stretches remain outside the $\pm$ 2 s.e. limits. Using the shrunken estimate and the beat factor, estimates of the transition probabilities may be constructed. However, the difference between the maximum-likelihood estimates of the "beat" factors, $\hat{\gamma}_{1s} = 1.779$, $\hat{\gamma}_{1w} = 2.176$, $\hat{\gamma}_{1sd} = 0.864$, $\hat{\gamma}_{2s} = -1.041$, $\hat{\gamma}_{2w} = -0.020$, $\hat{\gamma}_{2sd} = 2.129$ makes the transition probability estimates highly variable

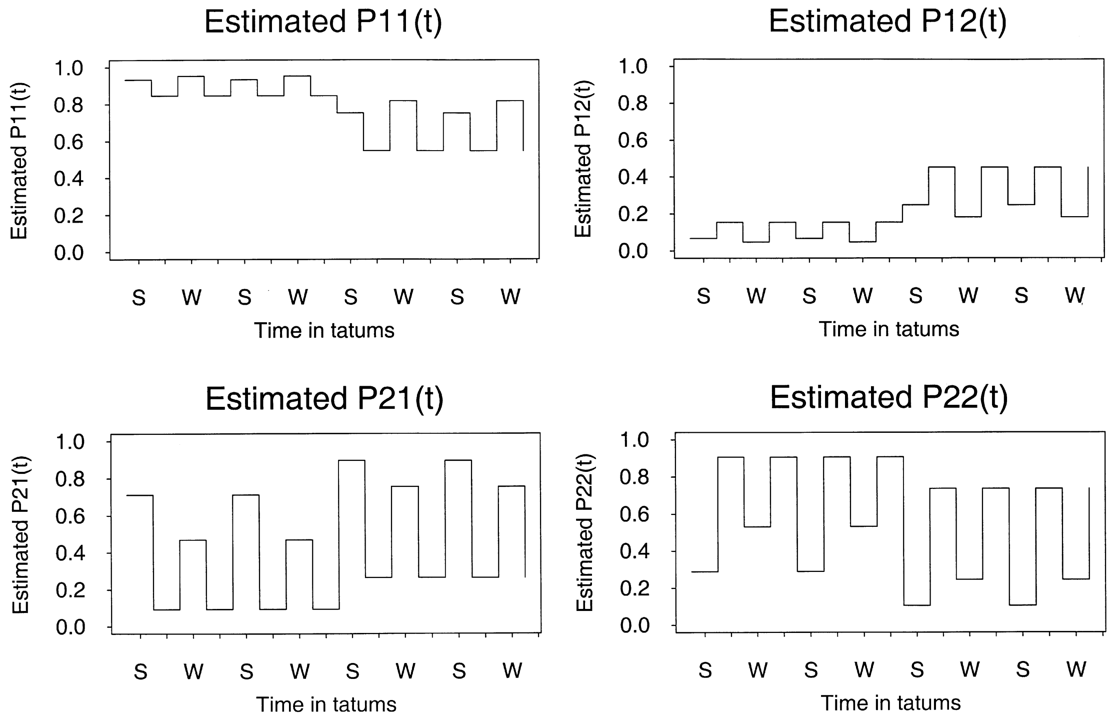

Fig. 4. Estimated transition probabilities for the music data using the data from measures 95 and 96. S refers to a strong beat and W to a weak one.

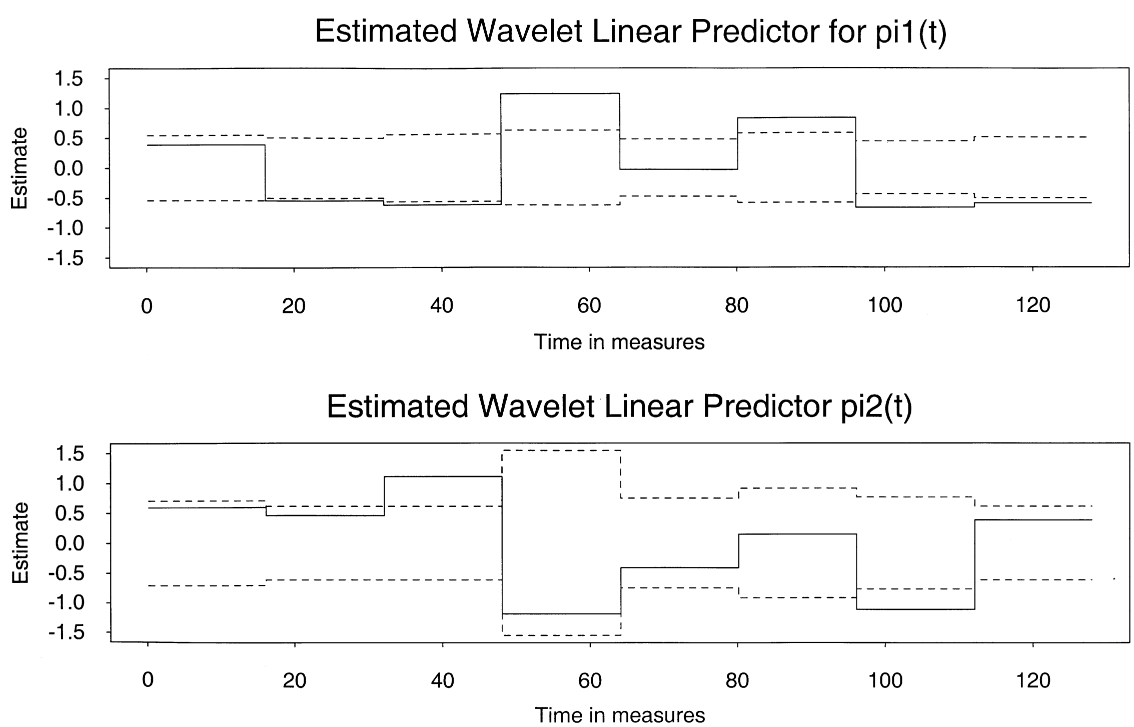

Fig. 5. Fitted values of the linear predictor, excluding the beat explanatories, for the music data. Marginal ± 2 s.e. limits are included.

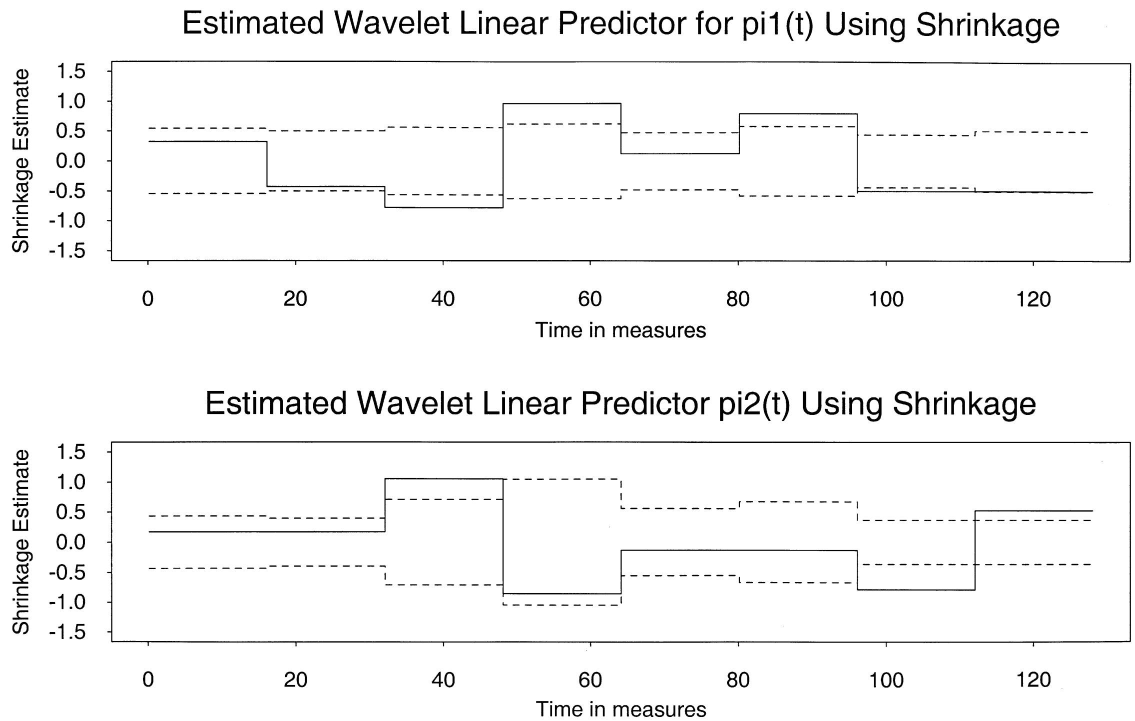

Fig. 6. The shrunken linear predictors, excluding the beat explanatories, for the music data and marginal ± 2 s.e. limits.

Table 1
Deviances resulting from fitting the stationary, then the model with beats and then the model with wavelets (19) to the music data

ANODEV table

Source	Deviance	DF
Music state 1		
Stationary model	492.1	364
Adding beat	377.6	362
Wavelet model	348.4	355
Music state 2		
Stationary model	697.0	658
Adding beat	664.6	656
Wavelet model	617.4	649

within each measure, as seen in Fig. 4. For this reason it is not practical to plot estimates of the transition probabilities for all measures. To appreciate the non-stationarity suggested by the estimates, plots of the linear predictor about the beat levels are provided. Now, for example, the fitted model can be used to generate and listen to further music of this type.

The overall fit of model (21) is now assessed in two fashions: via the final deviances and via the periodograms of the residuals. These two are discussed in the appendix. The results are given for both states 1 and 2 in the analysis of deviance (ANODEV) Table 1 and in Fig. 7, respectively. The final deviances are 348.4 and 617.4 with degrees of freedom 355 and 649. Neither provides evidence for lack of fit, the former on the basis of the comparison of the final deviance to its degrees of freedom (DF) and the latter on the basis of the approximate constancy of the periodogram as a function of frequency. For state 1 the change of deviance in moving from the stationary to the beat model is 114.5 with 2 degrees of freedom and in moving to the wavelet model the change is 29.2 with 7 degrees of freedom. Consistent with Fig. 5, one has evidence of nonstationarity. There is corresponding evidence in the case of state 2. The deviances necessarily decrease as more parameters are added to the model.

The second way overall fit is assessed in this work is via the periodogram of the deviance residuals. This statistic is sensitive to a variety of types of stationary temporal dependence and is crucial for examining the Markov assumption. For example

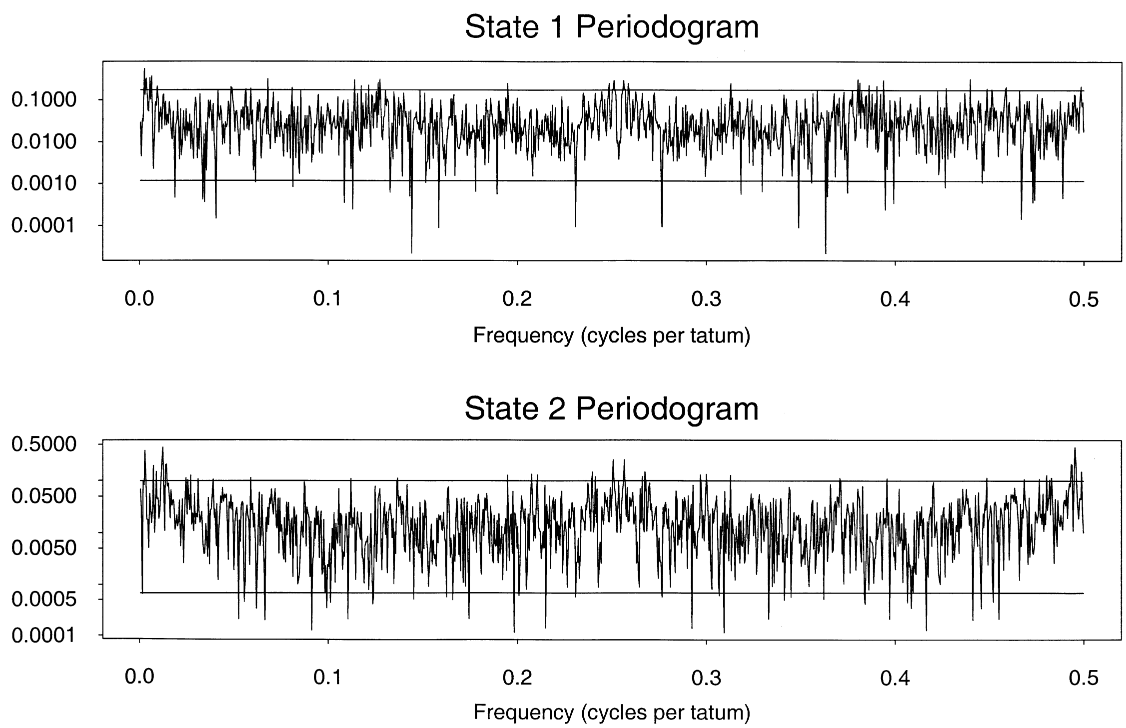

Fig. 7. Periodograms of the deviance residuals of the wavelet fits to the music data. Marginal approximate 95% confidence limits are indicated.

lurking periodicities have an opportunity to show themselves. The periodograms are graphed in Fig. 7 for the two states. The graphs include marginal approximate 95% confidence limits. There is no strong suggestion of remaining temporal dependence.

In summary, a model has been obtained that may be used to generate music pieces of "similar" character to Bach's.

3.2. The Snoqualmie Falls rainfall data

Markov chain analyses of hydrology data, specifically rainfall, were carried out in [14] for example. These authors fit first- and second-order Markov models to the two-state process of {no rain, rain} for four sites scattered about the world. Amongst other models, in the present notation, they fit the model

$$\text{logit}(\pi_a(t)) = \alpha_a + \sum_{l=1}^{L} [\beta_{a\ell} \sin(2\pi\ell t/366) + \gamma_{a\ell} \cos(2\pi\ell t/366)], \quad (20)$$

where $L = 4$, $a = 1{,}2$ and with t in days. They assessed the order of the chain via the change in deviance.

In the present paper the model fit to the Snoqualmie Falls rainfall data, an initial stretch of which was graphed in Fig. 2, is

$$\pi_a(t) = h\left\{\alpha_a + \sum_{l=1}^{L} [B_{al}(t)\sin(2\pi\ell t/365.25) + C_{al}(t)\cos(2\pi\ell t/365.25)]\right\} \quad (21)$$

with

$$B_{al}(t) = \sum_{j,k} \beta_{aljk}\psi_{jk}(t), \qquad C_{al}(t) = \sum_{j,k} \gamma_{aljk}\psi_{jk}(t) \quad (22)$$

with $h\{\}$ the inverse logit transformation. This model allows the amplitudes of the seasonal terms to depend on time. The values $L = 1$, $J_1, J_2 = 4$ and Haar wavelets were employed.

Fig. 8 shows the transition probability estimates for the case of $L = 1$. The estimates fluctuate in a seasonal fashion as was to be expected. The chances of remaining in a state are seen to be high,

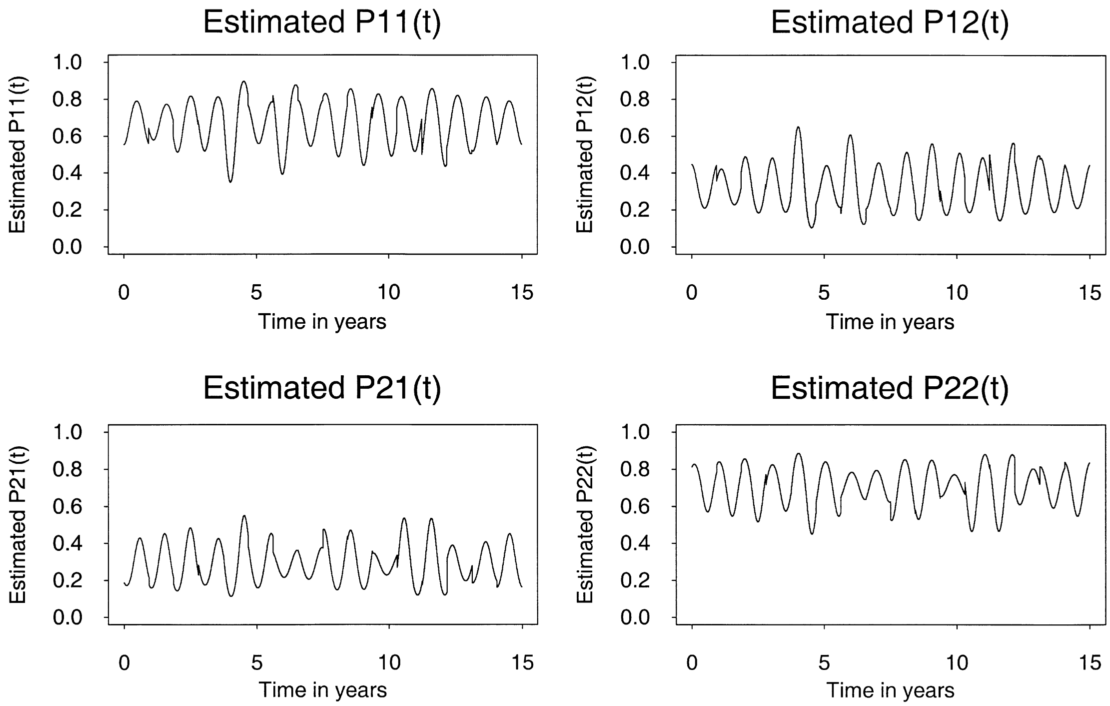

Fig. 8. The wavelet-based transition probability estimates obtained from the model (21), (22) for the rainfall data.

around 0.7, while the chances of changing state are low for both states 1 and 2. This fits with the idea that the Northwest Coast weather shows persistence on a time scale of days. There are minor suggestions of changes in time.

Fig. 9 provides estimates $\hat{\rho}_a(t) = \sqrt{\hat{B}_a(t)^2 + \hat{C}_a(t)^2}$, $a = 1,2$, of the amplitudes. There are no strong suggestions that the amplitude is varying with time.

Fig. 10 provides the transition probability estimates when shrinkage is included. It is to be compared with Fig. 8. The estimates show some variation in shape of the seasonal effect. Had the shrinker put to 0 all coefficients less than twice their standard error there would have been little change from Figs. 8–10.

The deviances obtained from fitting the model (22) with constant $B_a(t), C_a(t)$ and then the model (23), (24) are given in Table 2. The changes in deviance in going between the models are 30.8, 26.6 each with degrees of freedom 30. Neither suggests that bringing time variation of the present type into the seasonal model improves the fit.

The periodograms of the residuals are given in Fig. 11. Had L in model (23) needed to be bigger than the value 1 employed, this might have shown itself here. Neither periodogram shows evidence of remaining temporal dependence, i.e. of lack of validity of the Markov model.

3.3. The sleep data

The following models are fit to the sleep data, part of which appears in Fig. 3,

$$\pi_a(t) = h\left\{\alpha_a + \sum_{l=1}^{L} [B_{al}\sin(2\pi\ell t/144) + C_{al}\cos(2\pi\ell t/144)]\right\}, \quad (23)$$

$$\pi_a(t) = h\left\{\alpha_a + \sum_{l=1}^{L} [B_{al}(t)\sin(2\pi\ell t/144) + C_{al}(t)\cos(2\pi\ell t/144)]\right\} \quad (24)$$

with t in units of 10 min. In the latter case the coefficients are represented by wavelet expansions as in (24). The values $L = 1$, $J_1, J_2 = 6$ are employed.

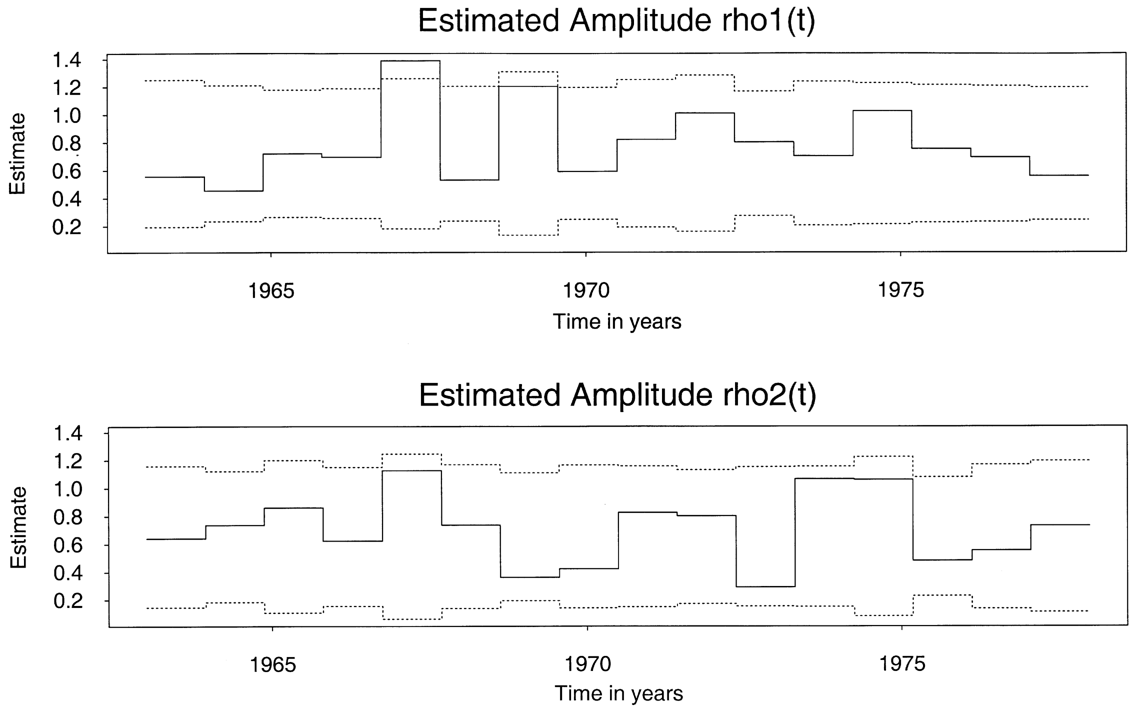

Fig. 9. Wavelet-based estimates of the amplitude $\rho_a(t) = \sqrt{B_a(t)^2 + C_a(t)^2}$ of the model (23), (24) for the rainfall data. Marginal $\pm$ 2 s.e. limits are included.

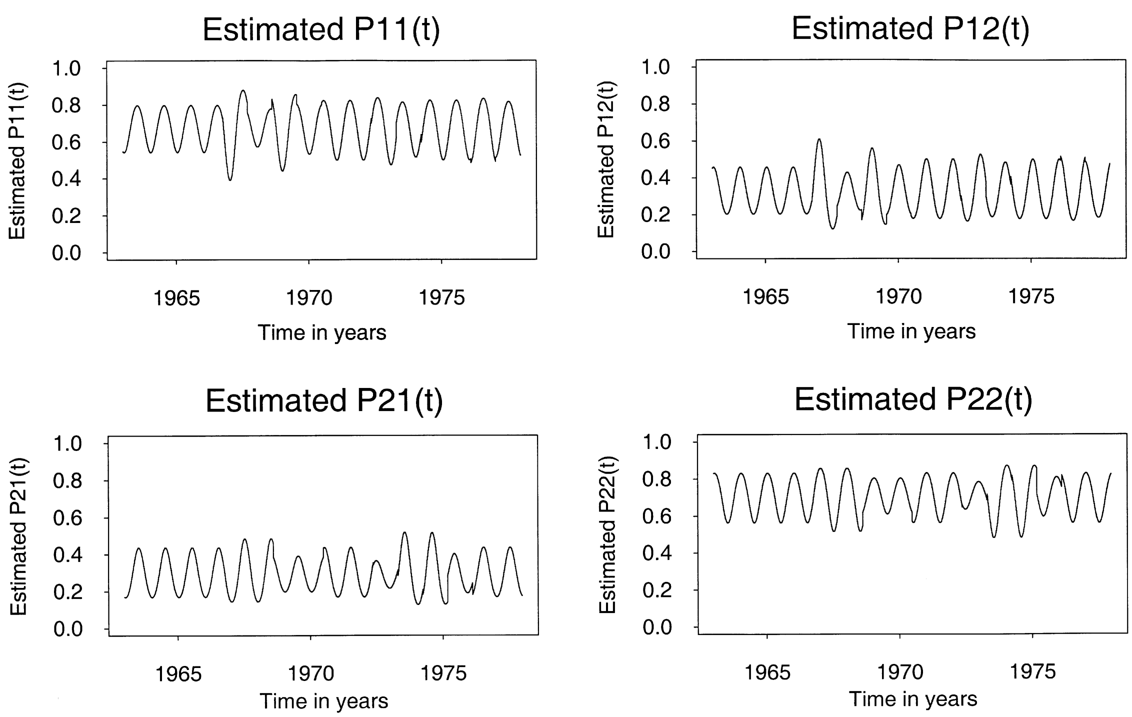

Fig. 10. The result of fitting the model (21), (22) and then applying shrinkage for the rainfall data.

Table 2
Deviances resulting from fitting the constant seasonal model and the model (21), (22) to the rainfall data

ANODEV table		
Source	Deviance	DF
Rain state 1		
Constant coefficient model	3001.6	2606
Wavelet model	2970.8	2576
Rain state 2		
Constant coefficient model	3194.9	2862
Wavelet model	3168.3	2832

Fig. 12 provides the estimated transition probabilities based directly on the maximum-likelihood estimates of model (24), (26). The 24 h period of the fitted probabilities is clear. Also it is apparent that the child tends to remain asleep or awake with high probabilities. Fig. 13 presents the wavelet-based estimates of time-varying amplitudes of the sine and cosine terms, $\rho_a(t) = \sqrt{B_a(t)^2 + C_a(t)^2}$. No evidence of substantial nonstationarity appears. Fig. 14 provides the results of shrinking the estimates towards the constant coefficient estimates after fitting the time varying amplitude model. The results are of more regular appearance.

The deviances found are listed in Table 3. The changes in deviance involved in moving from the constant coefficient to the time-varying model are 24.8 and 18.5, respectively, each with 30 degrees of freedom. Neither provides any evidence for the necessity of inclusion of time varying coefficients, B_a, C_a. Nor do the periodograms of Fig. 15 suggest remaining temporal dependence or that L in the model should be increased from the value 1 employed.

4. Discussion

In the work practical experience has been gained with wavelet-based models and shrinkage estimates for Markov chain data. In particular a variety of departures from stationarity have had an opportunity to show themselves. The Markov assumption is basic to the analysis. The reasonableness of this has been confirmed by examining the

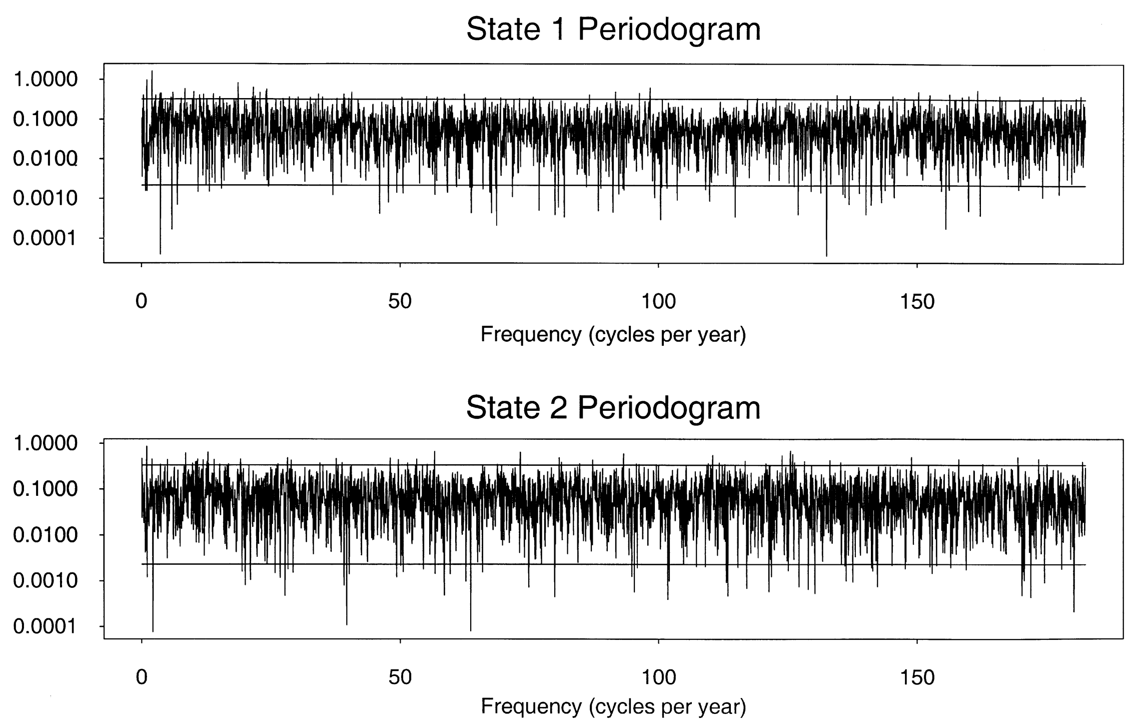

Fig. 11. The periodogram of the deviance residuals for the rainfall data. Marginal approximate 95% confidence intervals are indicated.

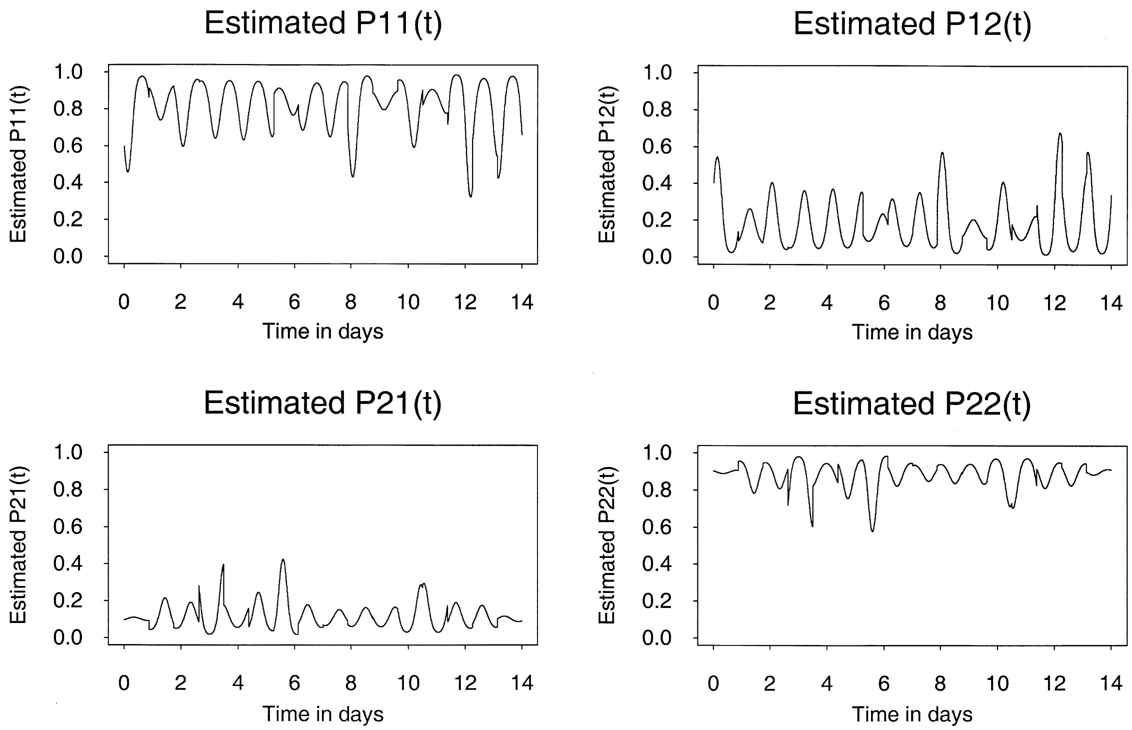

Fig. 12. Wavelet-based transition probability estimates obtained for the period 24 h sleep model.

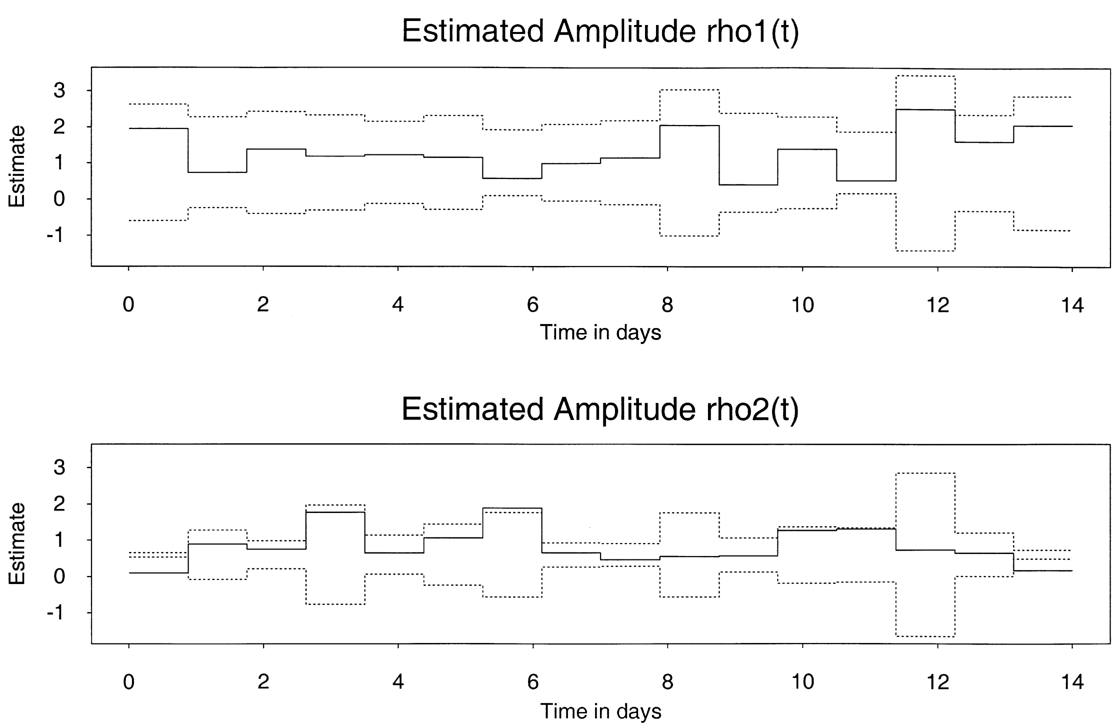

Fig. 13. Wavelet-based estimates of the amplitudes, $\rho_a(t)$, of the period 24 components of the sleep data. Marginal $\pm$ 2 s.e. limits are included.

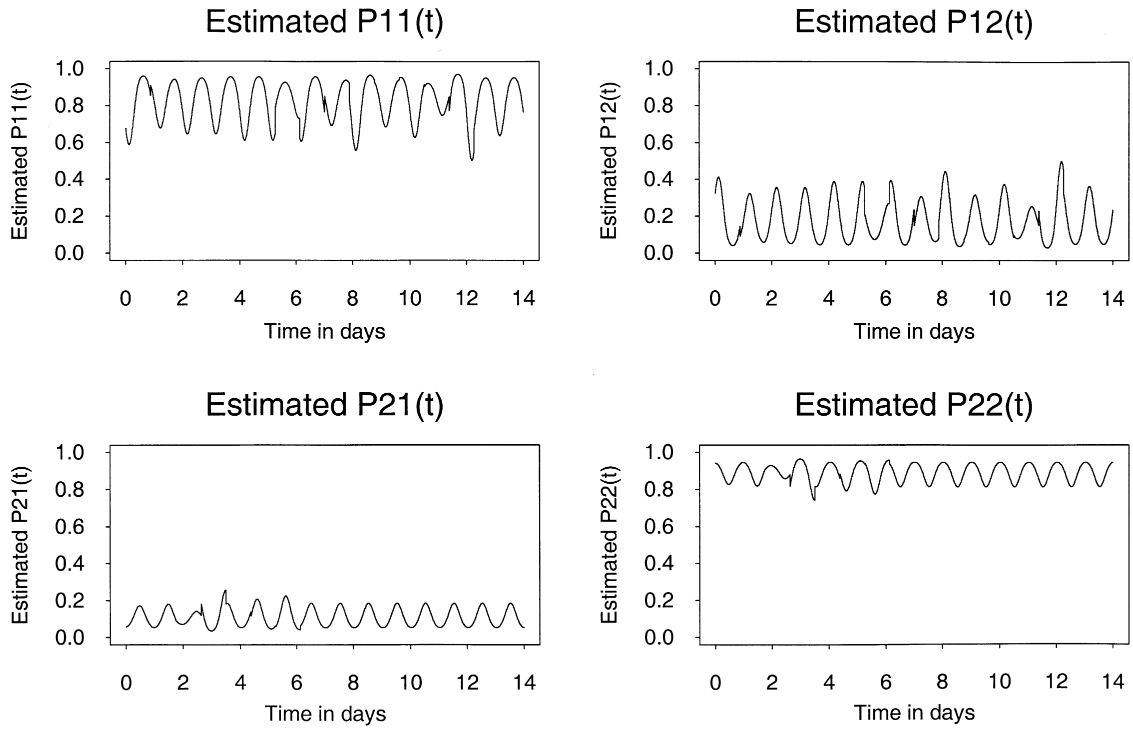

Fig. 14. The results of fitting the model (24) to the sleep data and then applying shrinkage.

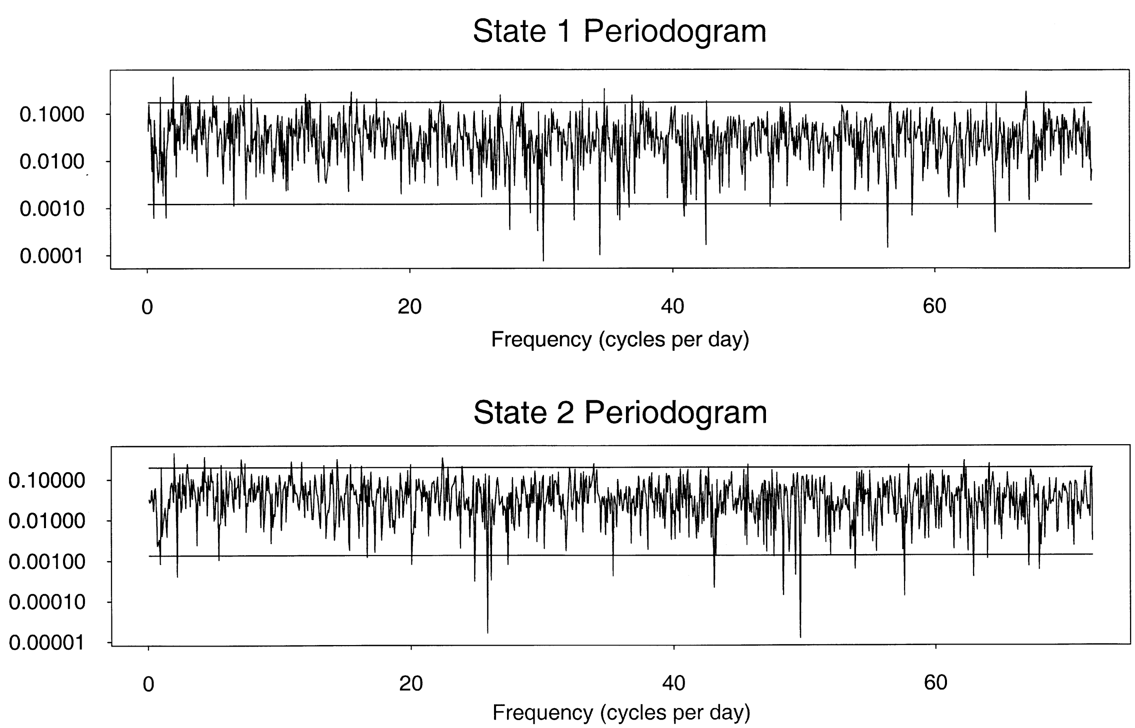

Fig. 15. Periodograms of the residuals of the fit of the sleep model. Also included are marginal approximate 95% confidence limits.

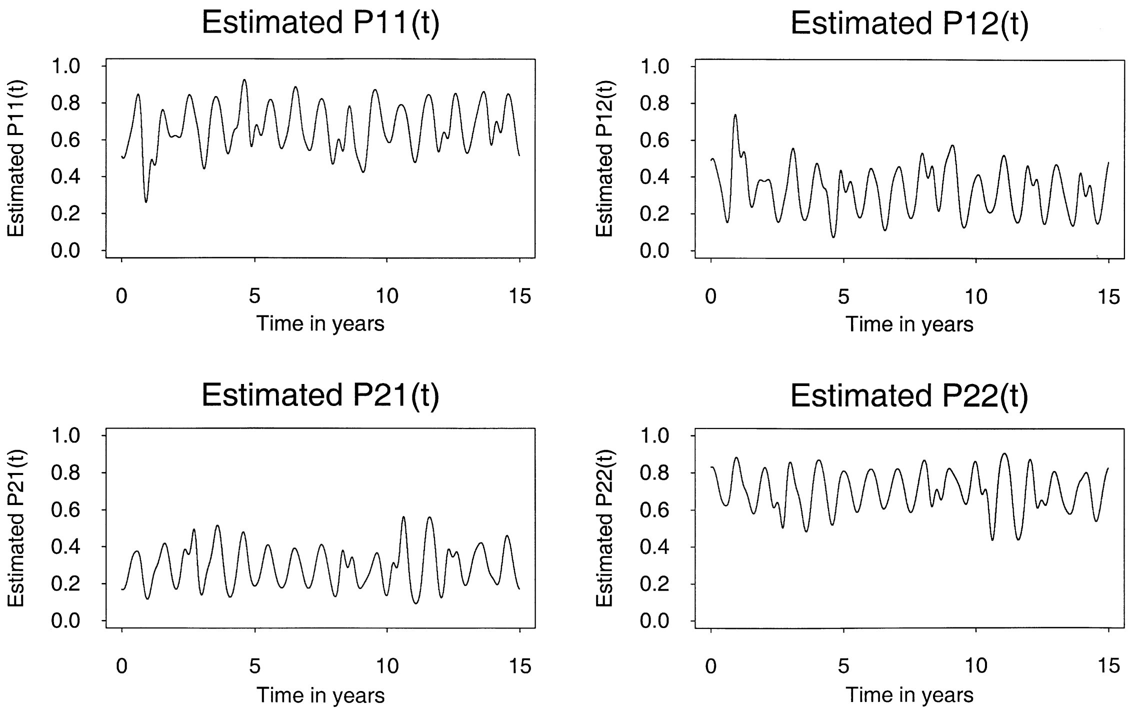

Fig. 16. The results of employing the sombrero function in estimating the transition probabilities for the main data.

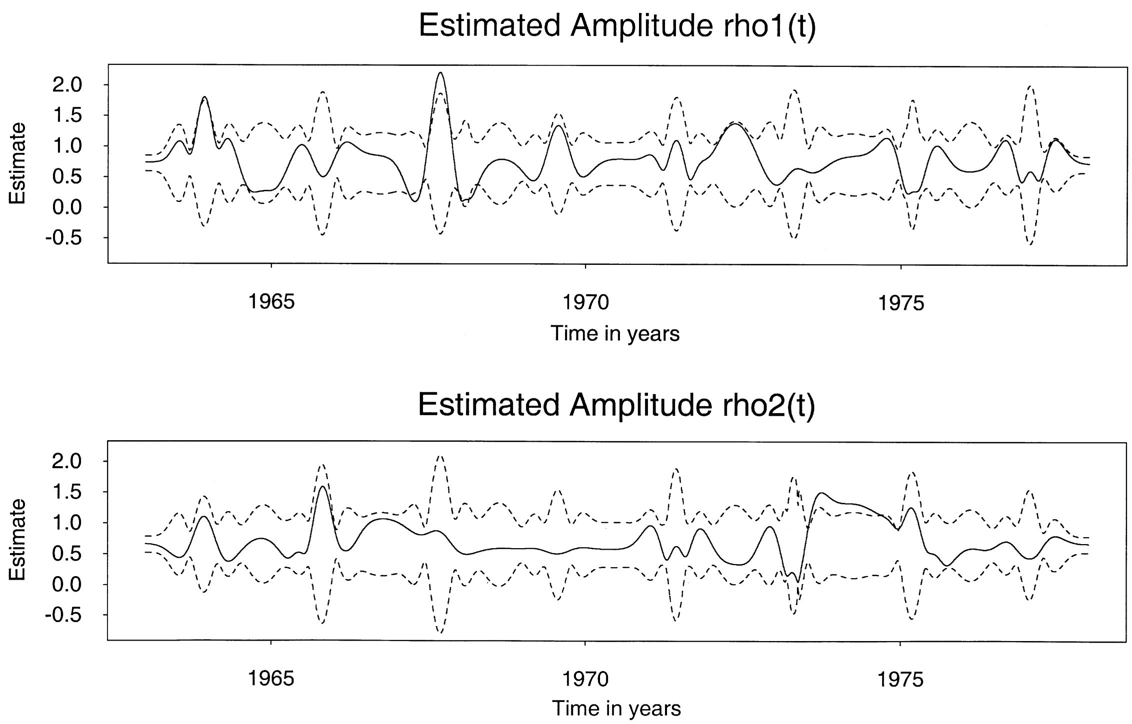

Fig. 17. The estimated amplitudes for the linear predictor when the sombrero function is employed.

periodogram of residuals appropriate to the binary nature of the data.

The initial estimates computed were maximum likelihood, but in an attempt to improve upon them marginal shrinkage has been employed. Covariates were included in the analyses with no difficulty.

Haar wavelets were employed, because of simplicity of interpretation and to search for abrupt

Table 3
Deviances obtained when modeling the sleep data

ANODEV table		
Source	Deviance	DF
Sleep state 1		
Constant coefficient model	640.4	899
Wavelet model	615.6	869
Sleep state 2		
Constant coefficient model	715.9	1111
Wavelet model	697.4	1081

changes. We have examined the effect of employing smoother families of orthonormal wavelets and of nonorthogonal regular functions. In particular, figures similar to Fig. 8 were obtained for the rainfall data using the S8 (symmlet) wavelet (which generates a compactly supported orthonormal system) and the Mexican hat (which generates a nonorthogonal system). Apart from the fact that the figures become smooth, there are no changes in conclusions. (See Figs. 16 and 17.)

The examples presented are all for the case of a process with two states, but extensions to the higher-order case are immediate. Extensions to chain-type processes remembering further back in time have also been also indicated.

Acknowledgements

We thank Peter Guttorp for providing the Snoqualmie Falls data and Luiz Menna-Barreto for providing the sleep data. The work was supported in part by the NSF Grants DMS-9625774 and INT-9600251, the CNPq Grant 910011/96-6 and the FAPESP Grant 97/11631-7.

Appendix

The computations needed in the work may be carried out via programs such as Splus or Glim, developed for generalized linear models and in particular for the binomial case.

The estimates of the coefficients are obtained by maximizing criterion (5) having taken the function $h(\cdot)$ of (19) to be

$$h(\eta) = \exp(\eta)/(1+\exp(\eta)),$$

i.e. the inverse logit. This type of estimate falls into the category of the so-called generalized linear model, see McCullagh and Nelder [32], Hastie and Pregibon [25], Venables and Ripley [38]. This model provides extensions of many of the basic concepts of regression analysis. An important extension is of residual to deviance residual.

In the present setup there are deviance residuals for states 1 and 2. The former is given by

$$d_t = \sqrt{2}\,\mathrm{sgn}\left(\frac{X_{11}(t)}{X_1(t)} - \hat{\pi}_1(t)\right)\left[X_{11}(t)\log\frac{X_{11}(t)}{X_1(t)\hat{\pi}_1(t)} + (X_1(t)-X_{11}(t))\log\left(\frac{X_1(t)-X_{11}(t)}{X_1(t)(1-\hat{\pi}_1(t))}\right)\right]^{1/2}.$$

Here log(0) is taken to be zero and $\hat{\pi}_1(t)$ is the fitted value of the probability under the model. State 2 residual deviances, e_t, are given by a similar formula. The final deviance is given by

$$\sum_{t=1}^{T}(d_t^2+e_t^2).$$

The use of these quantities is in assessing the fit of the model.

There is empirical evidence to suggest that the final deviance has a distribution that is approximately χ^2 (chi-square) with $T-p$ DF, where p is the number of parameters estimated. Further likelihood ratio theory indicates that changes of deviance will be approximately χ^2 with DF the number of null parameters added to the model.

Continuing, if the model is fitting well the values $d_1, d_2, \ldots, d_T$ should be approximately independent. The alternative is some form of temporal dependence. An effective way of picking up temporal dependence is examining the periodogram. The deviance residual periodogram is given by the modulus squared Fourier transforms

$$\frac{1}{2\pi T}\left|\sum_{t=1}^{T} d_t \mathrm{e}^{-\mathrm{i}\lambda t}\right|^2, \qquad \frac{1}{2\pi T}\left|\sum_{t=1}^{T} e_t \mathrm{e}^{-\mathrm{i}\lambda t}\right|^2.$$

In the case of (approximate) independence the expected values of these will be constant in frequency λ and the distributions χ^2 with 2 DF. This last may be used to set approximate confidence intervals in the figures. Examples are given in Section 3.

The estimates may be computed using the function glm( ) from Splus. For the binomial case, glm( ) takes data in the form of a two column matrix in which a 1 in the first column and 0 in the second denotes a success and a 0 in the first column and 1 in the second denotes a failure. In the case of estimating, say $\pi_1(t)$, one sees from Eq. (5) that $X_{11}(t) = 1$ will be considered a success and $X_{12}(t) = 1$ will be considered a failure. However, $X_{21}(t) = 1$ and $X_{22}(t) = 1$ are neither a success nor a failure. The function glm( ) can handle this type of situation by having 0 in both columns in the row corresponding to time t. A problem arises when too many such negligible rows occur. If, for example one is using the function

$$\psi_{jk}(t) = \begin{cases} 1, & t_0 \leqslant t < t_1, \\ -1, & t_1 \leqslant t < t_2 \end{cases}$$

and the rows corresponding to either times t_0 through t_1, or t_1 through t_2 are negligible, then the corresponding coefficient β_{1jk} is not estimable. Splus resolves this circumstance by assigning Not Available (NA) to the estimate of β_{1jk}. This presents a problem at the shrinkage step. In the examples presented to resolve this problem wavelet terms corresponding to NA estimates are removed from the regression matrix and the glm( ) fit reinitiated.

References

[1] J.R. Benjamin, C.A. Cornell, Probability, Statistics and Decision for Civil Engineers, McGraw-Hill, New York, 1970.

[2] A.T. Bharucha-Reid, Elements of the Theory of Markov Processes and Their Applications, McGraw-Hill, New York, 1960.

[3] P. Billingsley, Statistical Inference for Markov Processes, University of Chicago Press, Chicago, 1961.

[4] J. Bilmes, Timing is of the essence, Masters Thesis, MIT, 1993.

[5] Y.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis, MIT Press, Cambridge, MA, 1975.

[6] D.M. Blow, F.H.C. Crick, The treatment of errors in the isomorphous replacement method, Acta Crystalogr. 12 (1959) 794–802.

[7] D.R. Brillinger, Some river wavelets, Environmetrics 5 (1994) 211–220.

[8] D.R. Brillinger, Some uses of cumulants in wavelet analysis, J. Nonparametr. Statist. 6 (1996) 93–114.

[9] D.R. Brillinger, R.A. Irizarry, An investigation of the second- and higher-order spectra of music, Signal Process. 39 (1998) 161–179.

[10] A.G. Bruce, H.-Y. Gao, S + Wavelets: User's Manual, StatSci, Seattle, WA, 1994.

[11] A.G. Bruce, H.-Y. Gao, Understanding waveshrink: variance and bias estimation, Biometrika 83 (1996) 727–745.

[12] C. Chiann, Wavelet analysis in time series, Ph.D. Thesis, University of São Paulo, 1997 (in Portuguese).

[13] C. Chiann, P.A. Morettin, A wavelet analysis for time series, J. Nonparametr. Statist. 10 (1998) 1–46.

[14] R. Coe, R.D. Stern, Fitting models to daily rainfall data, J. Appl. Meteorol. 21 (1982) 1024–1031.

[15] M.S. Crouse, R.G. Baraniuk, Contextual hidden Markov models for wavelet-domain signal processing, Proceedings of the 31st Asilomar Conference on Signals, Systems and Computers, IEEE Computer Society, 1998.

[16] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, 1992.

[17] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1994) 425–455.

[18] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. 90 (1995) 1200–1224.

[19] D.L. Donoho, I.M. Johnstone, Minimax estimation via wavelet shrinkage, Ann. Statist. 26 (1998) 879–921.

[20] E.B. Dynkin, Markov Processes, Springer, Berlin, 1965.

[21] L. Fahrmeir, H. Kaufmann, Regression models for nonstationary categorical time series, J. Time Ser. Anal. 8 (1987) 147–160.

[22] W. Feller, An Introduction to Probability Theory and Its Applications, Wiley, New York, 1957.

[23] R.V. Foutz, R.C. Srivastava, Statistical inference for Markov processes when the model is incorrect, Adv. Appl. Probab. 11 (1979) 737–749.

[24] P. Guttorp, Stochastic Modelling of Scientific Data, Chapman & Hall, London, 1995.

[25] T.J. Hastie, D. Pregibon, Generalized linear models, in: J.M. Chambers, T.J. Hastie (Eds.), Statistical Models in S, Wadsworth, Pacific Grove, 1992, pp. 195–247.

[26] L. Hiller, L. Isaacson, Experimental Music, McGraw-Hill, New York, 1959.

[27] R. Irizarry, Statistics and music: fitting a local harmonic model to musical sound signals, Ph.D. Thesis, University of California, Berkeley, 1998.

[28] K. Jones, Compositional applications of stochastic processes, Comput. Music J. 5 (1981) 381–396.

[29] H. Kaufman, Regression models for nonstationary categorical time series: asymptotic estimation theory, Ann. Statist. 15 (1987) 79–98.

[30] S. Mallat, A Wavelet Tour of Signal Processing, Academic Press, San Diego, 1998.

[31] P. McCullagh, J.A. Nelder, Generalized Linear Models, 2nd Edition, Chapman & Hall, London, 1989.

[32] L. Mello, A. Isola, F. Louzada, L. Menna-Barreto, A four-year follow-up study of the sleep-wake cycle of an infant, Biol. Rhythm Res. 27 (1996) 291–298.
[33] P.A. Morettin, Wavelets in statistics, Reviews of the Institute of Mathematics and Statistics, University of São Paulo, Vol. 3, 1997, pp. 211–272.
[34] G.P. Nason, Wavelet function estimation using cross-validation, in: A. Antoniadis, G. Oppenheim (Eds.), Wavelets and Statistics, Springer, New York, 1993, pp. 261–280.
[35] Y. Ogata, Maximum likelihood estimates of incorrect Markov models for time series and the derivation of AIC, J. Appl. Probab. 17 (1980) 59–72.
[36] R. Pinkerton, Information theory and melody, Sci. Amer. 194 (1956) 77–84.
[37] J.W. Tukey, Introduction to the dilemmas and difficulties of regression, Report, Statistics Department, Princeton University, 1979.
[38] W.N. Venables, B.D. Ripley, Modern Applied Statistics with SPLUS, 2nd Edition, Springer, New York, 1997.