Some Wavelet-Based Analyses of Markov Chain Data

David R. Brillinger*, Pedro A. Morettin†, Rafael A. Irizarry‡ and Chang Chiann§

Version 5.0: September 9, 1998

Abstract

This paper expresses 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 observations in the nonstationary case. In particular the method appears useful for detecting abrupt or steady changes in the structure of Markov chains. The method is illustrated by means of data sets concerning music, rainfall and sleep. In the examples both direct and shruken estimates are computed. The approach is implemented by means of programs for fitting generalized linear models. The goodness of fit and the presence of nonstationarity are assessed both by change of deviance and graphically via periodogram plots.

1 Introduction

This work presents empirical analyses of nonstationary Markov chain models, based on wavelet expansions, for time series data sets taken from musicology, meteorology and sleep research, respectively. A basic goal is looking

*David R. Brillinger is Professor, Department of Statistics, University of California, Berkeley, CA 94720

†Pedro A. Morettin is Professor, Department of Statistics, University of São Paulo, SP 05315-970, Brazil.

‡Rafael A. Irizarry is Assistant Professor, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205.

§Chang Chiann is currently a Post-Doctoral Fellow, Department of Statistics, University of São Paulo, SP 05315-970, Brazil.

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, eg. 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 sequences of discrete state valued observations amongst other possibilities.

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

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.

2 Background

2.1 The Markov Chain Case

Consider a nonstationary Markov chain. Suppose time is discrete, $t = 1, 2, ...$ and that the chain has A states indexed by the labels $a,\ b$. Let $P_{ab}(t)$ denote the conditional probability of being in state b at time t, given that the process was in state a at time $t-1$. Let $P_b(t)$ denote the marginal probability of being in state b at time t. The process may be denoted $Y(t),\ t = 1, 2, ...$ with $Y(t) \in \{1, ..., A\}$. It will be supposed that the state of the process has been observed at the T successive times, $t = 1, 2, ..., T$.

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

$$\mathrm{logit}\{P_{ab}(t)\} = \sum_{j,k} \beta_{abjk}\psi_{jk}(t), \tag{1}$$

with the β's unknown parameters to be estimated. At the next step this expression is substituted into a likelihood function and the unknowns estimated by maximizing a likelihood. There may be a further step of shrinkage. In the expansion (1), the ψ's are the functions of some wavelet basis, as discussed below.

A likelihood function, on which estimates can be based, may be set down as follows. Define $Y_a(0) = 1$, if the chain starts in state a and $Y_a(0) = 0$ otherwise, with $\sum_a Y_a(0) = 1, EY_a(0) = P_a = Prob\{Y_a(0) = 1\}$. Similarly, define $Y_{ab}(t) = 1$, if the process is in state a at time $t-1$ and in state b at time t, and $Y_{ab}(t) = 0$ otherwise and lastly define $Y_b(t) = 1$, if in state b at time t, $Y_b(t) = 0$ otherwise.

Given the data and parametric forms for $P_a(t),\ P_{ab}(t)$ the likelihood is

$$[\prod_{a=1}^{A} P_a^{y_a(0)}][\prod_{t=1}^{T}\prod_{a=1}^{A}\prod_{b=1}^{A} P_{ab}(t)^{y_{ab}(t)}], \tag{2}$$

where the $y_a,\ y_{ab}$ refer to observed data values (as opposed to employing the notation $Y_a,\ Y_{ab}$ for the corresponding random quantities).

In the case that $A = 2$ things simplify. With $\pi_1(t) = P_{11}(t), P_{12}(t) = 1 - \pi_1(t), \pi_2(t) = P_{22}(t), P_{21}(t) = 1 - \pi_2(t)$ the likelihood is

$$P_1^{y_1(0)} P_2^{y_2(0)} \Pi_{t=1}^{T} \{\pi_1(t)^{y_{11}(t)}[1-\pi_1(t)]^{y_{12}(t)}\pi_2(t)^{y_{22}(t)}[1-\pi_2(t)]^{y_{21}(t)}\}. \tag{3}$$

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

$$\Pi_{t=1}^{T}\{\pi_1(t)^{y_{11}(t)}[1-\pi_1(t)]^{y_1(t-1)-y_{11}(t)}\pi_2(t)^{y_{22}(t)}[1-\pi_2(t)]^{y_2(t-1)-y_{22}(t)}\}. \tag{4}$$

When consideration below turns to estimation, it is useful to note that this has the form of a likelihood based on independent Bernoullis with $y_1(t), y_2(t)$ taking on the values 0 or 1 depending on whether the process is in state 1 or state 2 at time t. 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 (1).

A variety of properties of maximum likelihood estimates have been developed for Markov chains in the large sample case. For example, Billingsley (1961) developed consistency and asymptotic normality results for a stationary finite dimensional parameter Markov chain. Foutz and Srivastava

(1979) and Ogata (1980) derived the large sample distribution of the maximum likelihood estimate in the stationary ergodic case. Bishop et al (1975) suggested some methods for assessing empirically whether a Markov chain is stationary. Fahrmeir and Kaufmann (1987), Kaufmann (1987) indicated how nonstationary Markov chain models might be included within the generalized linear modelling methodology. Details of this are provided below. Coe and Stern (1982) presented empirical analyses involving nonstationary Markov chain models. McCullagh and Nelder (1989), Section 8.4.3, discussed the Coe and Stern work.

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

2.2 Wavelets

Wavelets are contemporary 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.

Depending on the situation the functions of a wavelet basis may be orthogonal or not. In developing orthonormal bases it is convenient to start with a *father wavelet* or *scaling* function ϕ, such that

$$\phi(t) = \sqrt{2} \sum \ell_k \phi(2t - k) \tag{5}$$

for some coefficients, l_k, and normalized via $\int \phi(t) dt = 1$. A *mother wavelet* ψ is then obtained through

$$\psi(t) = \sqrt{2} \sum h_k \phi(2t - k), \tag{6}$$

where

$$h_k = (-1)^k \ell_{1-k} \tag{7}$$

The equations (5) and (6) are called *dilation equations*. The coefficients ℓ_k, h_k are low-pass and high-pass filters, respectively and appear in the so-called quadrature mirror filters used in fast algorithms to compute the wavelet transform.

Often $\phi(t)$ and the l_k are such that these functions generate an orthonormal system for $L_2(\Re)$. It can be denoted $\{\phi_{j_0,k}(t)\} \cup \{\psi_{j,k}(t)\}_{j\geq j_0,k}$, with $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$ and $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$ for $j \geq j_0$ with j_0 the so called coarsest scale. Some further properties may hold for these wavelets, such as the admissibility condition $\int \psi(t)dt = 0$, or that the first $(r-1)$ moments of ψ vanish, for some $r \geq 2$. In this case the degree of smoothness of ψ is given by r. For details see Daubechies (1992).

For any $f \in L_2(\Re)$, one may consider the expansion

$$f(t) = \sum_k \alpha_k \phi_{j_0,k}(t) + \sum_{j\geq j_0} \sum_k \beta_{j,k}\psi_{j,k}(t), \tag{8}$$

where the wavelet coefficients are given by

$$\alpha_k = \int f(t)\phi_{j_0,k}(t)dt, \quad \beta_{j,k} = \int f(t)\psi_{j,k}(t)dt \tag{9}$$

following the orthonormality.

An estimate of the function $f(t)$ takes the form

$$\hat{f}(t) \;=\; \sum_k \hat{\alpha}_k \phi_{j_0,k}(t) + \sum_j \sum_k \hat{\beta}_{j,k}\psi_{j,k}(t), \tag{10}$$

where the $\hat{\alpha}_k$, $\hat{\beta}_{jk}$ are estimates of the α_k, β_{jk} of (8).

Several issues are of interest here:

(i) the choice of the wavelet basis,

(ii) the choice of a shrinkage policy,

(iii) the choice of the parameters appearing in the shrinkage scheme,

(iv) the estimation of the scale parameter (noise level).

A brief discussion of these follows. For further details see for example Morettin(1997) and Mallat(1998).

(i) Concerning the choice of the wavelet basis, some possibilities are: the Haar functions, compactly supported wavelet bases (Daubechies, 1992), complex wavelets (Morlet, or modulated Gaussian), Mexican hat (second derivative of Gaussian, but is not an orthogonal system), etc.

The problem and the form of the signal to be analysed may suggest a particular basis. In the examples to be presented principally the Haar expansion will be used having in mind its simplicity of interpretation and the detection of temporal changes. The Haar expansion is based on the choices

$$\phi(t) \;=\; 1,\; 0 \;\leq\; t \;<1, \tag{11}$$

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

The expansion is then, more simply

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

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

(ii) By *shrinkage* is meant the replacement of an estimated coefficient, $\hat{\beta}_{j,k}$, by a shrunken value $\hat{\beta}^*_{jk} \;=\; w(\hat{\beta}_{j,k}/s_{jk})\hat{\beta}_{j,k}$, for some function $w(.)$ with $0 \leq w(.) \leq 1, w(0) = 1$ and with s_{jk} an estimated standard error of $\hat{\beta}_{j,k}$. The function $w(.)$ is meant to dampen down the variability of $\hat{\beta}_{j,k}$. The estimated function will be, in the Haar case,

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

Various criteria have been suggested for the choice of $w(.)$. For example Blow and Crick(1959), using a mean squared error criterion, were led to the function

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

with the I_j Bessel functions. Tukey(1979) suggested the use of

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

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

Donoho and Johnstone(1994,1995,1998), motivated by considerations of risk, work with multipliers of the form $\delta_{\lambda_n}(\hat{\beta}_{j,k})$, with $\lambda_n \to \infty$ as $n \to \infty$, e.g. $\lambda_n = s_{j,k}\sqrt{2\log n}$. Here $s_{j,k}$ is the estimated standard deviation of $\hat{\beta}_{j,k}$. Specifically, they suggested the use of hard and soft thresholds, defined, respectively by

$$\delta_\lambda^H(x) = \begin{cases} 0, & \text{if } |x| \leq \lambda \\ x, & \text{if } |x| > \lambda \end{cases} \tag{17}$$

and

$$\delta_\lambda^S(x) = \begin{cases} 0, & \text{if } |x| \leq \lambda \\ sign(x)(|x| - \lambda), & \text{if } |x| > \lambda \end{cases} \tag{18}$$

for some λ.

The first is of the type so called "kill or preserve", while the second is "kill or reduce". The smooth policy may present larger biases, while the hard one produces smaller biases but larger variances, see Bruce and Gao(1996). This procedure damps down the terms considerably more rapidly than choices (15) and (16). 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 (1) 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) If one uses hard or soft thresholding, one has to choose the form of the parameter λ_n. In some situations it may be level-independent, leading to the so-called *universal* threshold of Donoho and Johnstone, in other situations it may depend of the level j. In the general situation, one might set $\lambda_{j,k} = s_{j,k}\sqrt{2\log n}$, the threshold parameter depending on the level and location. Other proposals are the SureShrink (Donoho and Johnstone,1995) and a cross-validation procedure (Nason,1995).

(iv) In the case of (15) or (16), $s_{j,k}$, an estimate of the standard deviation of $\hat{\beta}_{jk}$, is needed. For a signal plus stationary noise model, Brillinger(1996) 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.

The present work will consider principally a logit link for the probabilities and a wavelet-based regression function, as in (1). Of course other links than the logit may be used.

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

2.3 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 (4) will be used. What is then needed is a specific model for the $\pi_a(t),\ a = 1, 2$.

Fahrmeir and Kaufmann (1987) and Kaufmann (1987) present a maximum likelihood approach for statistical inference concerning categorical-valued time series possessing certain forms of Markov structure. The 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:

$$Prob\{Y_a(t) \ = \ 1 \mid \mathbf{Y}(t-1),\ \mathbf{Y}(t-2),\ ...\} \ = \ h_a(\mathbf{Z}(t)^\tau\ \boldsymbol{\beta}\ )$$

for $a \ = \ 1, ..., A-1$, where $\mathbf{Y}(t) \ = \ [Y_a(t)]$, $\mathbf{h} : R^{A-1} \to R^{A-1}$, is one-to-one, and $\mathbf{Z}(t)$ a function of past observations and fixed explanatories. Higher-order Markov chains are included by inserting interaction terms such as $Y_a(t-1)Y_b(t-2)$ into the linear predictor, $\mathbf{Z}(t)^\tau\ \boldsymbol{\beta}$.

To be specific, consider the two state (A=2) and Haar wavelets case. The model (1) may be written

$$\pi_a(t) = h\{\alpha_a + \sum_{j=0}^{J_a} \sum_{k=0}^{2^j-1} \beta_{ajk}\psi_{jk}(t)\}, \tag{19}$$

$a = 1, 2$ with h for example the inverse of the logit transform as in (1). This model falls within the framework of the Fahrmeir-Kaufmann work. Further the computations may be carried out via programs, such as Splus or Glim, developed for the generalized linear model, in particular for the Binomial case. Assuming that the preceding model is correct and that J_a is finite, the results of Fahrmeir and Kaufmann show that the usual 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, eg. of the transition probabilities of the Markov model. This is what has been done in the examples presented below.

In Brillinger (1994,96) it is proposed to estimate the uncertainty of a shrunken wavelet estimate of a mean function by acting as if the weights, $w(\hat{\beta}_{ajk}/s_{ajk})$ are constant, (really more nearly constant), i.e. that the major variability comes from the $\hat{\beta}_{ajk}$ appearing. This is what has been done in the examples presented below. It is also acted as if the J_a were given.

It is important to assess the goodness of fit of models employed. In the present case the model has two basic characteristics: Markov dependence and nonstationarity described by wavelet expansions. It is a generalized linear model, so techniques proposed for that case may be considered. These include: the final deviance and employing various types of residual analysis. In particular, since temporal dependence is a basic concern, an examination of the periodogram of the residuals may prove insightful in considering alternatives of stationary dependence.

To be specific the estimates may be found 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 equation (4) that $y_{11}(t) = 1$ will be considered a success and $y_{12}(t) = 1$ will be considered a failure. However, $y_{21}(t) = 1$ and $y_{22}(t) = 1$ 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 negligeable rows ocurr. If, for example one is using the function

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

and the rows corresponding to either times t_0 through t_1, or t_1 through t_2 are negligeable, then the corresponding coefficient β_{1jk} is not estimable. Splus resolves this circumstance by assigning 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.

3 The Data Sets

Consideration now turns to applying the above modelling procedure to some observational data sets of interest.

3.1 The Music Data

Markov processes have been used in finding structure in music, see for example Pinkerton (1956), Hiller and Isaacson (1959), Jones (1981). For example musicologists have tried to model melodies as k-th 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 because they assume stationarity, a characteristic that melodies definitely do not seem to possess.

In Irizarry (1998) a stochastic composition is created using a 5-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 by 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 2-state (jump, no jump) model will be employed. A jump ocuring 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 refer to as an *intense* part of the melody.

The example to be considered is 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*, Bilmes (1993). In this particular fugue the smallest subdivision of the beat is a sixteenth. However, sixteenth notes are used only as embelishments so to be able to study the structure of the piece in terms of the intense parts, here a tatum will be defined to be an eighth-note and a two-state time series will be defined via

$$Y(t) = \begin{cases} 2, & \text{if the beginning of a note occurs in tatum } t \\ 1, & \text{no new note in tatum } t. \end{cases} \tag{20}$$

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

Questions that might be addressed here include: can wavelet analysis usefully describe nonstationarity present? Is the piece Markov? Is it Markov of some higher order?

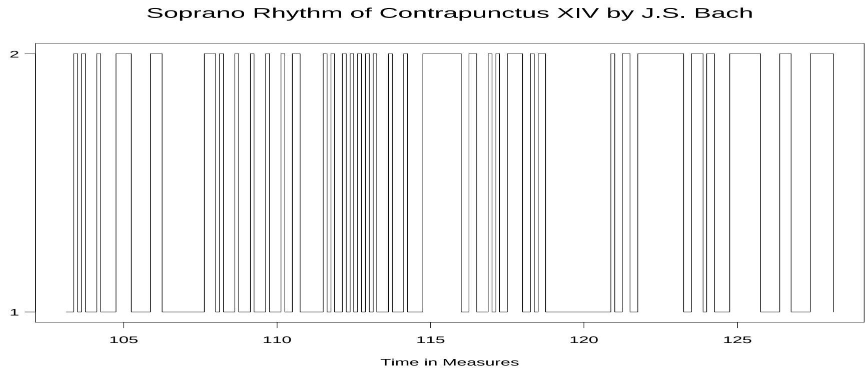

Figure 1: Beginnings of notes for the Soprano line in Bach's unfinished fugue. Th e value 2 corresponds to a new note starting.

Brillinger and Irizarry (1998) and Irizarry (1998) contain more details on the quantification and statistical analysis of music.

3.2 The Snoqualmie Falls Rain Data

For the present work Peter Guttorp provided daily data concerning whether or not at least 0.01 inches of rain had occurred at Snoqualmie Falls, Washington, for each day for the period 1963 to 1977. That is for 15 years. He had analyzed the January data, Guttorp (1995), and in particular fit 2-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 Figure 2 with $Y = 1$ when no rain and $Y = 2$ when rain. One sees stretches of both wet and dry days.

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

3.3 An Example From Sleep Research

Mello et al (1996) 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 his mother or eventually by a maid. When carried out the measurements were done at intervals of 10 minutes. 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 weeks to six weeks, are

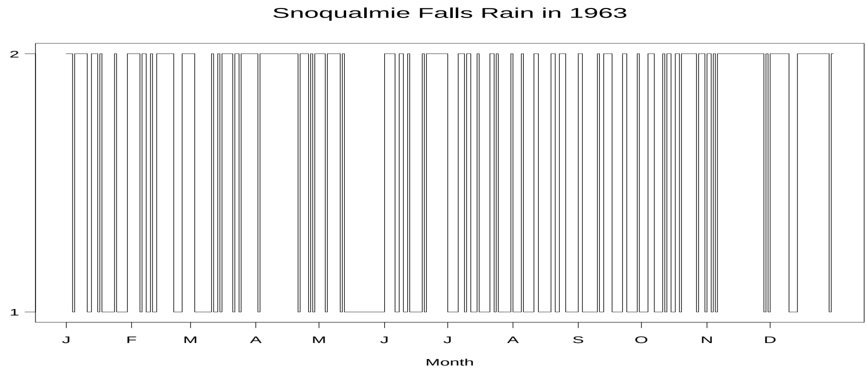

Figure 2: The Rain Data. The value 2 corresponds to a day with rain and 1 to none.

studied. There are $T = 2016$ values. Figure 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 egual lengths of time. Examination of the data, for example by periodogram analysis, shows a period of 24 hours.

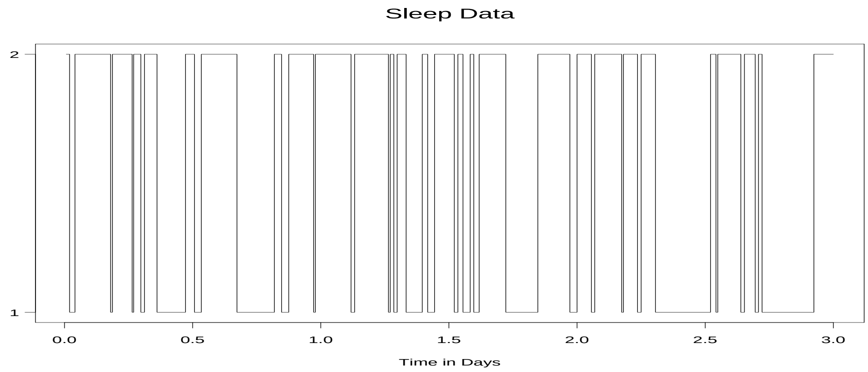

Figure 3: Sleep Data. The value 2 corresponds to the child being asleep, 1 to awake.

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

4 Results of Fitting the Markov Chain Models

4.1 The Music Data

Figure 1 provided a segment of some baroque music. In this type of music it is common to have notes starting on the beat (tatums 1,3,5,7 in the 8 tatums within a measure), rather than the subdivisions of the beat (tatums 2,4,6,8). Further more it is more likely that a note starts on a strong beat (tatums 1 and 5) rather than a weak beat (tatums 3 and 7). The terms $\gamma_{as}x_{as}(t), \ldots$ in the model below are "beat" explanatories inserted to handle this phenomenon. Specifically write

$$\begin{cases} x_{as}(t) = 1 \ when \ t \ mod \ 4 = 1 \\ x_{aw}(t) = 1 \ when \ t \ mod \ 4 = 3 \\ x_{asd}(t) = 1 \ when \ t \ mod \ 4 = 0 \ or \ 2 \end{cases}$$

with s referring to strong, w referring to weak and sd to subdivision. [RAFA - what's that?] The model fit is the following

$$\pi_a(t) = h\{\sum_{j=1}^{J_a}\sum_{k=0}^{2^j-1} \beta_{ajk}\psi_{jk}(t) + \gamma_{as}x_{as}(t) + \gamma_{aw}x_{aw}(t) + \gamma_{asd}x_{asd}(t)\} \quad (21)$$

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

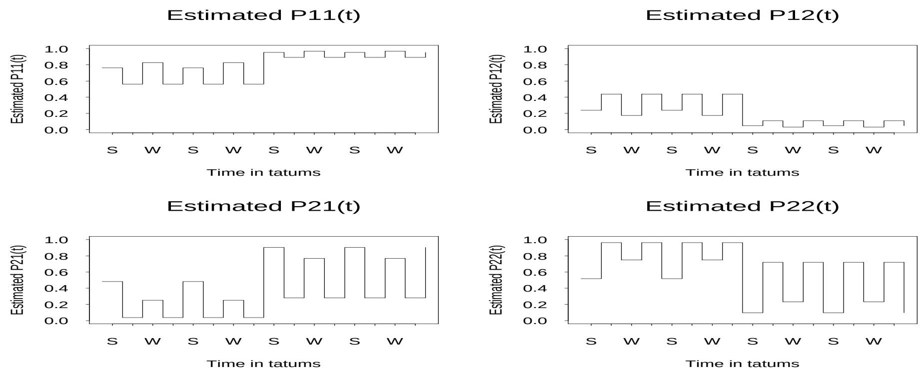

Figure 4: Estimated transition probabilities for the music data. S refers to a strong beat and W to a weak one.

Figure 4 uses the data of measures 47 and 48 of the piece and provides the transition probabilities as estimated by substituting the maximum likelihood

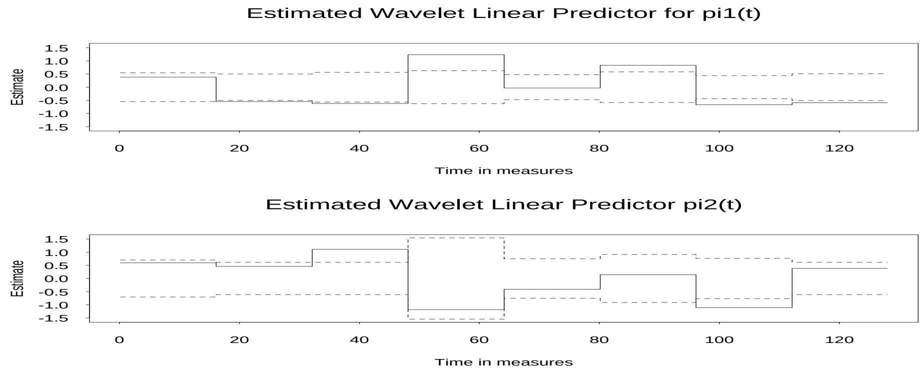

Figure 5: Fitted values of the linear predictor for the music data. Marginal ± 2 s.e. limits are included.

estimates of the β, γ into (21). In the plot S refers to a strong beat and W to a weak one. There is an apparent effect.

Figure 5 provides the wavelet part of the linear predictor. Figure 5 is useful for examining the nonstationarity of the data as in particular it includes marginal ± 2 s.e. limits about the beat level. In the present case, as was anticipated from the context, there is evidence of nonstationarity transition probabilities. At the same time various values are within, or nearly within, the ± 2 s.e. limits suggesting that improved estimates might be obtained via shrinkage.

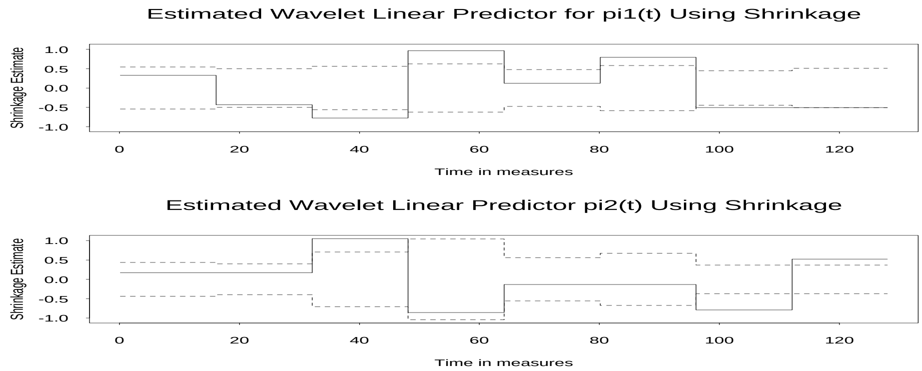

Figure 6: The shrunken linear predictors for the music data and marginal ± 2 s.e. limits.

Figure 6 is the same as the previous figure, but with the shrunken estimates. Using this estimate and the beat factor useful estimates of the transition probabilities may be constructed. [DO IT?]

The overall fit of the model (21) is assessed in two fashions: via the final deviances and via the periodograms of the residuals. The results are given for both states 1 and 2 in Table 1 and Figure 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. For state 1 the change of deviance in moving from the stationary to the beat model is 114.5 and in moving to the wavelet model the change is 29.2 with 7 degrees of freedom with 2 degrees of freedom. Consistently with Figure 5 one has evidence of nonstationarity. There is corresponding evidence in the case of state 2.

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. The periodograms are graphed in Figure 7 for the two states. The graphs include marginal approximate 95% confidence limits. There is no strong suggestion of remaining temporal dependence.

ANODEV Table - Music State 1

Source	Deviance	DF
Stationary Model	492.1	364
Adding Beat	377.6	362
Wavelet Model	348.4	355

ANODEV Table - Music State 2

Source	Deviance	DF
Stationary Model	697.0	658
Adding beat	664.6	656
Wavelet Model	617.4	649

Table 1: Deviances resulting from fitting the stationary, then the nonstationary model (21) to the music data.

In the present case there were some difficulties of estimation of the coefficients of the type refered to at the end of Section 2.3.

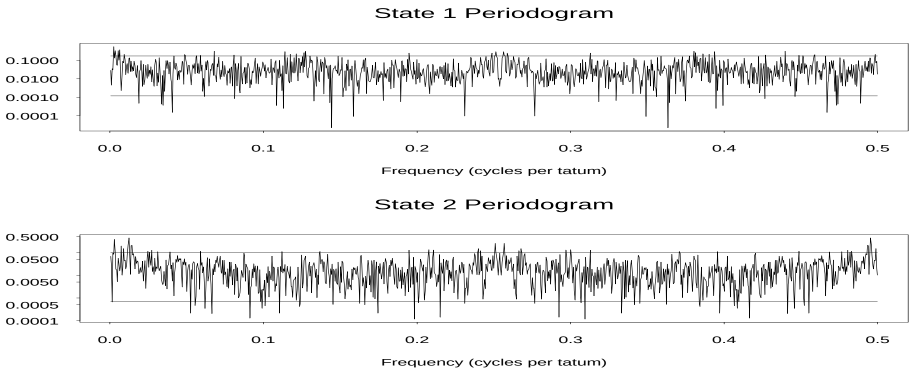

Figure 7: Periodograms of the deviance residuals of the wavelet fits to the music data. Marginal approximate 95% confidence limits are indicated.

4.2 The Snoqualmie Falls Rainfall Data

Markov chain analyses of rainfall data were carried out in Coe and Stern (1982) 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

$$logit(\pi_a(t)) \;=\; \alpha_a \;+\; \sum_{l=1}^{L}[\beta_{a\ell}\sin(2\pi\ell t/366) + \gamma_{a\ell}\cos(2\pi\ell t/366)] \qquad (22)$$

$L = 4$, $a = 1, 2$ 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 Figure 2, is

$$\pi_a(t) = h\{\alpha_a + \sum_{l=1}^{L}[B_{al}(t)\sin(2\pi\ell t/365.25) + C_{al}(t)\cos(2\pi\ell t/365.25)]\} \quad (23)$$

with

$$B_{al}(t) = \sum_{j,k}\beta_{aljk}\psi_{jk}(t),\; C_{al}(t) = \sum_{j,k}\gamma_{aljk}\psi_{jk}(t) \qquad (24)$$

It allows the amplitudes of the seasonal terms are allowed to depend on time. The values $L = 1$, $J_1, J_2 = 4$ and Haar wavelets were employed.

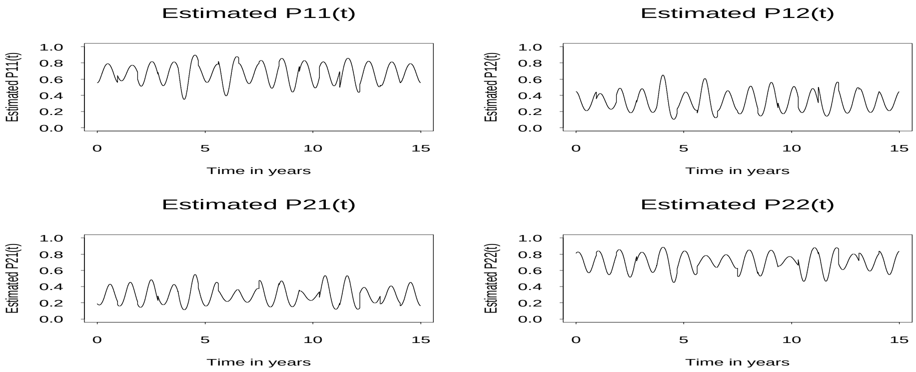

Figure 8: The wavelet-based transition probability estimates obtained from the model (23,24) for the rainfall data.

Figure 8 shows the transition probability estimates for the case of $L = 1$. They fluctuate in a seasonal fashion as was to be expected. The chances of remaining in a state appear high and of changing state, low for both states 1 and 2. This fits with the idea that the North West Coast weather shows persistence on a time scale of days. One sees some suggestions of changes in structure.

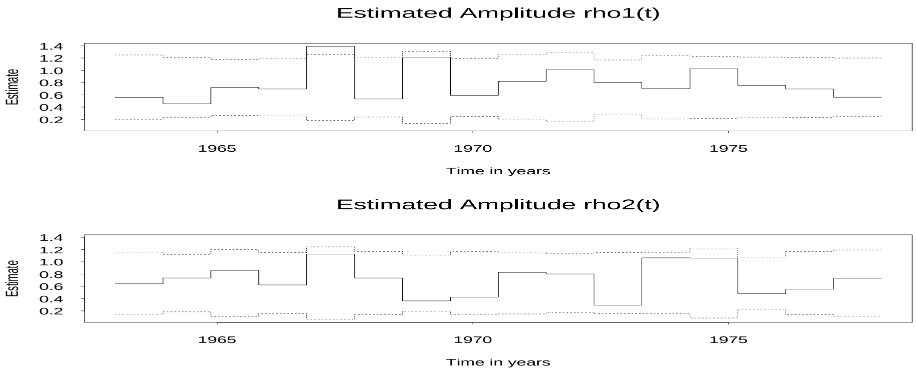

Figure 9: Wavelet-based estimates of $\rho_a(t) = \sqrt{B_n(t)^2 + C_n(t)^2}$ of the model (23,24) for the rainfall data. Marginal ± 2 s.e. limits are included.

Figure 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

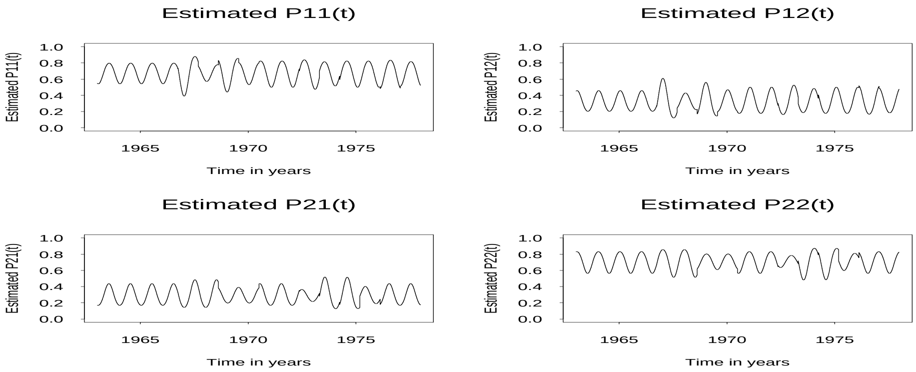

Figure 10: The result of fitting the model (23,24) and then applying shrinkage for the rainfall data.

Figure 10 provides the transition probability estimates when shrinkage is included. It is to be compared with Figure 8. The estimates show some changes of 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 Figure 8 to Figure 10.

ANODEV Table - Rain State 1

Source	Deviance	DF
Constant Coefficient Model	3001.6	2606
Wavelet Model	2970.8	2576

ANODEV Table - Rain State 2

Source	Deviance	DF
Constant Coefficient Model	3194.9	2862
Wavelet Model	3168.3	2832

Table 2: Deviances resulting from fitting the constant seasonal model and the model (23,24) to the rainfall data.

The deviances resulting from fitting the model with constant $B_a(t), C_a(t)$ and the model (23,24) are given in Table 2. The changes in deviance in going

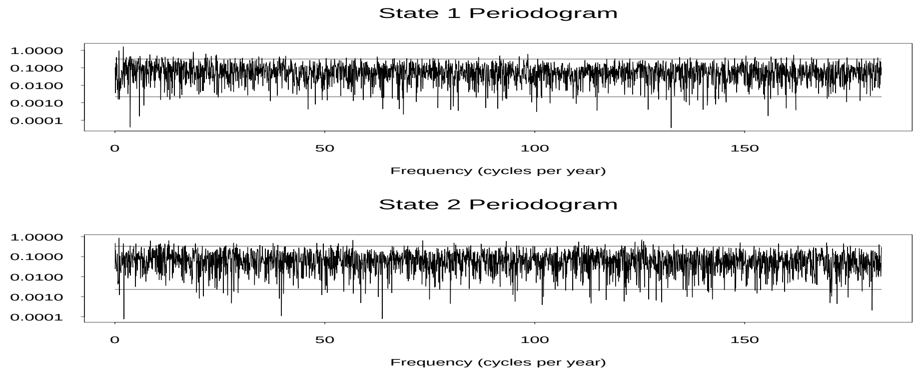

Figure 11: The periodogram of the deviance residuals for the rainfall data. Marginal approximate 95% confidence intervals are indicated.

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 improves the fit.

The periodograms of the residuals, given in Figure 11, Neither shows evidence of remaining temporal dependence.

4.3 The Sleep Data

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

$$\pi_a(t) = h\{\alpha_a + \sum_{l=1}^{L}[B_{al}\sin(2\pi\ell t) + C_{al}\cos(2\pi\ell t)]\}, \tag{25}$$

$$\pi_a(t) = h\{\alpha_a + \sum_{l=1}^{L}[B_{al}(t)\sin(2\pi\ell t) + C_{al}(t)\cos(2\pi\ell t)]\} \tag{26}$$

In the latter the coefficients are represented by wavelet expansions as in (25). The values $L = 1, J_1, J_2 = 6$ were employed.

Figure 12 provides the estimated transition probabilities. The 24 hour period of the fitted probabilities is clear. Also it is apparent that the child tends to remain in the sleep or awake state it already occupies.

Figure 13 presents the wavelet-based estimates of time varying amplitudes of the sine and cosine terms. No evidence of substantial nonstationarity appears.

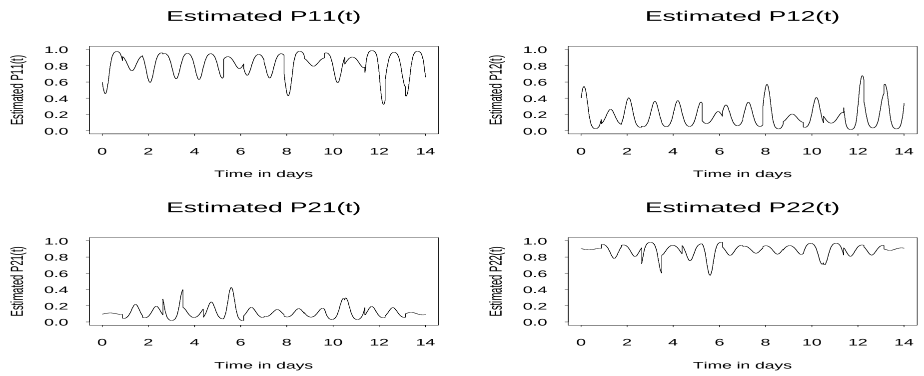

Figure 12: Wavelet-based transition probability estimates obtained for the period 24 hr sleep model.

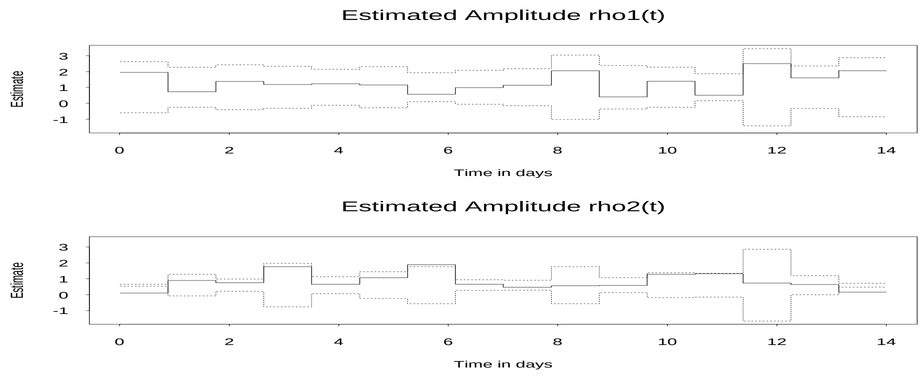

Figure 13: Wavelet-based estimates of the amplitudes, $\rho_a(t)$, of the period 24 components of the sleep data. Marginal ± 2 s.e. limits are included.

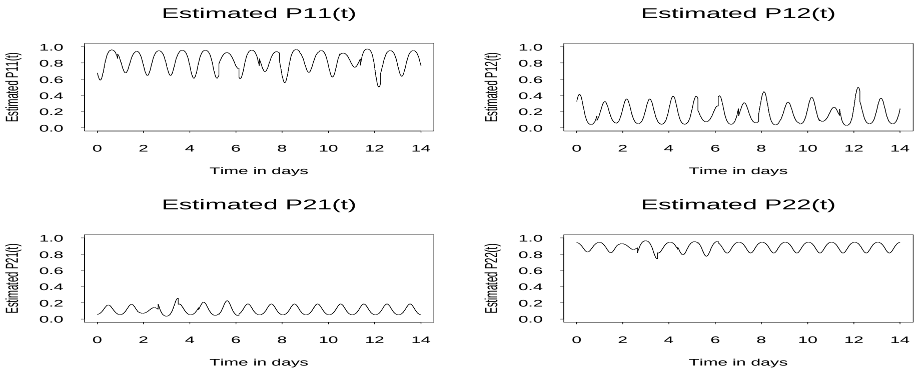

Figure 14: The results of fitting the model (26) to the sleep data and then applying shrinkage.

Figure 14 provides the results of shrinking the estimates to the constant coefficient estimates after fitting the time varying amplitude model. The result is of much more regular appearance.

ANODEV Table - Sleep State 1

Source	Deviance	DF
Constant Coefficient Model	640.4	899
Wavelet Model	615.6	869

ANODEV Table - Sleep State 2

Source	Deviance	DF
Constant Coefficient Model	715.9	1111
Wavelet Model	697.4	1081

Table 3: Deviances obtained when modelling the sleep data.

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 freedom. Neither provides any evidence of for the inclusion of time varying coefficients, B_a, C_a. Nor do the periodograms of Figure 15 suggest remaining temporal dependence.

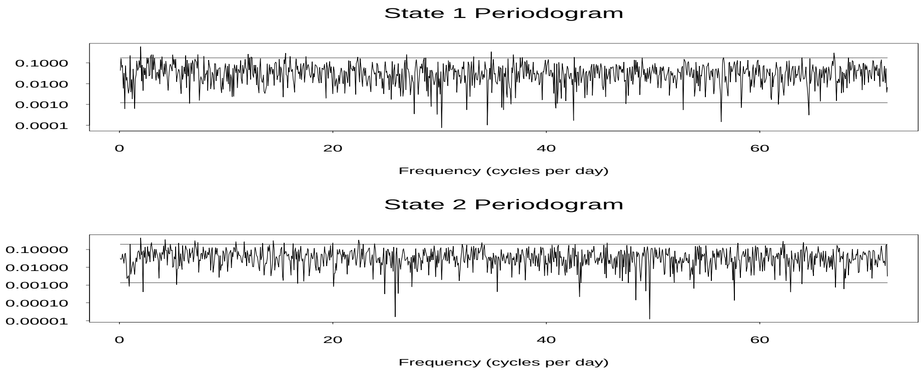

Figure 15: Periodograms of the residuals of the fit of the sleep model. Also included are marginal approximate 95% confidence limits.

In summary ...

4.4 A Continuous Wavelet Fit

The Haar wavelets involve jump discontinuities. They can be expected to be particularly useful when abrupt changes are taking place. However it seems worth recording the results of employing the sombrero function. This will be done for the rain data. The sombrero function is given by

$$(x^2 - 1)e^{-x^2/2} \tag{27}$$

Figure 16 is to be compared with Figure 8. The visible changes have now become smooth, rather than abrupt, as was to be anticipated. Figure 17 is ...

5 Discussion

In the work practical experience has gained with wavelet-based models for the Markov chain data. In particular a variety of departures from stationarity have had an opportunity to show themselves. Principally Haar wavelelts were employed, because of simplicity of interpretation and to highlight abrupt changes. The initial estimates computed were maximum likelihood, but in an attempt to improve upon them shrinkage has been employed. Covariates may be inclued in the analysis simply, and this was done in the music example.

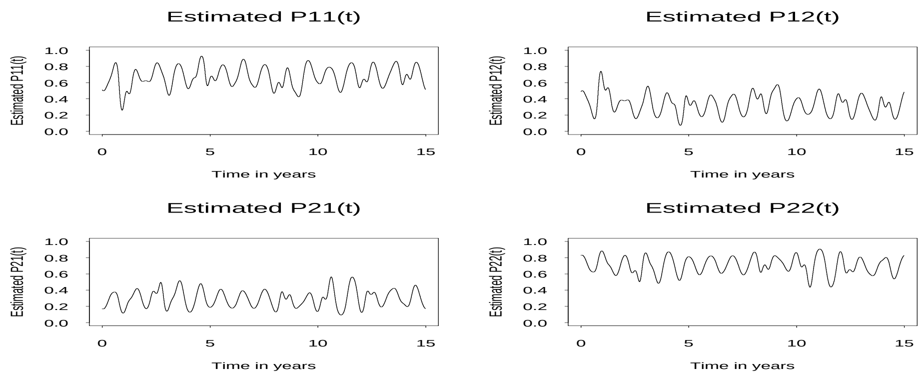

Figure 16: The estimated amplitudes for the linear predictor when the sombrero function is employed.

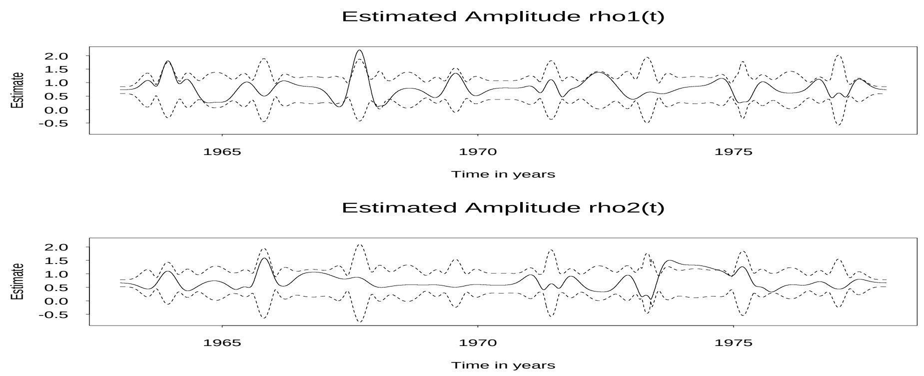

Figure 17: The results of employing the sombrero function in estimating the transition probabilities for the rain data.

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

The shrinkage might have been made automatic by inclusion of a penalty term in the log likelihood.

6 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.

7 References

Billingsley, P. (1961), *Statistical Inference for Markov Processes*, University of Chicago Press, Chicago.

Bilmes, J. (1993), *Timing is of the Essence*, Masters Thesis, MIT.

Bishop, Y. M., Fienberg, S. E. and Holland, P. W. (1975), *Discrete Multivariate Analysis*, MIT Press, Cambridge.

Blow, D. M. and Crick, F. H. C. (1959), "The Treatment of Errors in the Isomorphous Replacement Metod", *Acta Crystalographica*, 12, 794-802.

Brillinger, D. R. (1994), "Some River Wavelets", *Environmetrics*, 5, 211-220.

Brillinger, D. R. (1996), "Some Uses of Cumulants in Wavelet Analysis", *J. Nonparametric Statistics*, 6, 93-114.

Brillinger, D. R. and Irizarry, R. (1998), "An Investigation of the Second- and Higher-Order Spectra of Music", *Signal Processing*, 39, 161-179.

Bruce, A. G. and Gao, H-Y. (1994), *S+ Wavelets: User's Manual*, StatSci, Seattle, WA.

Bruce, A. G. and Gao, H-Y. (1996), "Understanding Waveshrink: Variance and Bias Estimation", *Biometrika*, 83, 727-745.

Chiann, C. (1997), *Wavelet Analysis in Time Series*, Ph.D thesis, University of São Paulo (in Portuguese).

Chiann, C. and Morettin, P. A. (1998), "A Wavelet Analysis for Time Series", To appear, *J. Nonparametric Statistics.*

Coe, R. and Stern, R. D. (1982), "Fitting Models to Daily Rainfall Data", *J. Applied Meteorology*, 21, 1024-1031.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Philadelphia: SIAM.

Donoho, D. L. and Johnstone, I. M. (1994), "Ideal Spatial Adaptation by Wavelet Shrinkage", *Biometrika*, 81, 425-455.

Donoho, D. L. and Johnstone, I. M. (1995), "Adapting to Unknown Smoothness via Wavelet Shrinkage", *Journal of the American Statistical Association*, 90, 1200-1224.

Donoho, D. L. and Johnstone, I. M. (1998), "Minimax Estimation via Wavelet Shrinkage", *Ann. Statist.*, 26, ?-?.

Fahrmeir, L. and Kaufmann, H. (1987), "Regression Models for Non-Stationary Categorical Time Series", *J. Time Series Analysis*, 8, 147-160.

Foutz, R. V. and Srivastava, R. C. (1979), "Statistical Inference for Markov Processes when the Model is Incorrect", *Adv. Appl. Prob*, 11, 737-749.

Guttorp, P. (1995), *Stochastic Modelling of Scientific Data*, London: Chapman and Hall.

Hiller, L. and Isaacson L. (1959), *Experimental Music*, New York: McGraw-Hill

Irizarry, R. (1998), *Statistics and Music: Fitting a Local Harmonic Model to Musical Sound Signals.* Ph. D. Thesis, University of California, Berkeley.

Jones, K. (1981), "Compositional Applications of Stochastic Processes", *Computer Music Journal*, 5, 381-396.

Kaufman, H. (1987), "Regression Models for Nonstationary Categorical Time Series: Asymptotic Estimation Theory", *Ann. Statist*, 15, 79-98.

Mallat, S. (1998), *A Wavelet Tour of Signal Processing*, San Diego: Academic Press.

Mello, L., Isola, A., Louzada, F. and Menna-Barreto, L. (1996), "A Four-Year Follow-up Study of the Sleep-Wake Cycle of an Infant", *Biological Rhythm Research*, 27, 291-298.

McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models* (Second Edition), London: Chapman and Hall.

Morettin, P. A. (1997), "Wavelets in Statistics", *Reviews of the Institute of Mathematics and Statistics, University of São Paulo*, 3, 211-272.

Nason, G. P. (1995), "Wavelet Function Estimation using Cross-Validation", In *Wavelets and Statistics*, 261-280 (Antoniadis, A. and Oppenheim, G., editors), New York, Springer-Verlag, Lecture Notes in Statistics 103.

Ogata, Y. (1980), "Maximum Likelihood Estimates of Incorrect Markov Models for Time Series and the Derivation of AIC", *J. Applied Prob*, 17, 59-72.

Pinkerton, R. (1956), "Information Theory and Melody," *Scientific American*, 194, 77-84.

Tukey, J. W. (1979), " Introduction to the Dilemmas and Dificulties of Regression", Report, Statistics Dept., Princeton University.