Comparison of autoregressive and multitaper spectral analysis for long time series

Imola K. Fodor, James G. Berryman, and Philip B. Stark[1]

keywords: *autoregressive analysis, multitaper analysis, spectral analysis*

ABSTRACT

The periodogram (*i.e.*, the square of the Fourier transform) of a time series generally provides a poor estimate of the spectrum if that spectrum has a wide dynamic range. So the spectrum of any process that includes either one or many resonant modes (sharp peaks) can be expected to be poorly computed by such elementary means. Multitaper spectral analysis is a nonparametric method designed to provide a rigorous method of resolving the spectrum of such complex processes. There are some practical difficulties with this method, such as deciding what tapers to use and how many, that can make the method some what difficult for the uninitiated user. Another approach to spectral analysis is the parametric method known as autoregressive (AR) analysis (related but not identical to maximum entropy spectral analysis). AR analysis can be used to approximate the dominant modes in the spectrum, remove them from the time series and thus prewhiten the remaining time series, thereby eliminating the main problem with analysis based on the periodogram. Furthermore, if the main purpose of the spectral analysis is to determine the number and location of the modes, the AR method provides a set of parameters that can be used to describe the mode structure if desired. The present paper provides a set of examples comparing the use of both the multitaper method and the autoregressive method for analyzing the same (long) time series data. We find that both methods give very comparable results in these examples, thus providing a type of empirical cross-validation showing that both methods are therefore doing an excellent job of estimating the true spectrum of the time series.

INTRODUCTION

The notion of a spectrum of light dates at least from the time of Newton's experiments with prisms and Rowland's experiments with diffraction gratings in which white light is split into its component spectra. However, the notion of a spectrum in acoustics is

[1]**email:** solar@stat.berkeley.edu, berryman@sep.stanford.edu, stark@stat.berkeley.edu

much older, dating at least from the time of Pythagoras (6th Century B.C.) and the musical scale he developed.

The quantitative analysis of acoustical spectra can be accomplished by taking the Fourier transform of the acoustical time series and squaring it. The resulting function of frequency is often called the "periodogram" for the historical reason that it was first plotted, not as a function of frequency but rather, as a function of period. The total energy in a time series is just the integral of the square of the amplitude over time. The periodogram provides a measure of the relative energy contained at each frequency in the power spectrum. For real data with noise, the periodogram, although very well-known, is nevertheless a notoriously bad method of estimating the spectrum (Yuen and Fraser, 1979; Percival and Walden, 1993). Windowing methods (Blackman and Tukey, 1958) were developed to reduce and control the erratic behavior observed in the periodogram. The idea of using multiple windows (or "tapers" as they are now called) and then averaging (*i.e.*, stacking in the frequency domain) has arisen in the last 20 years (Yuen and Fraser, 1979; Thomson, 1982). Multitaper analysis has proven to be a very effective means of producing reliable spectra and also of analyzing the statistical properties of these spectra. We will use long time series, both experimental and simulated, of helioseismic data (sound waves in our Sun) to compare and contrast the multitaper methods to autoregressive methods of spectrum estimation.

Millions of acoustic waves are stochastically excited in the solar convection zone. Those waves travel through the inner layers of the sun, and thus carry information about the conditions of the solar interior. Precise estimation of the many different oscillation mode parameters is therefore crucial in understanding the physical processes of the inner sun. The estimation procedure involves many steps but can be briefly summarized as follows: first, Doppler velocity images of the visible solar disk are taken, then the images are decomposed into time series of spherical harmonic coefficients, then the spectrum of each individual series identified by the spherical harmonics "quantum" numbers l and m is estimated. The spectrum estimation is usually done nonparametrically in the Fourier space. This paper proposes autoregressive (AR) modeling in the time domain as an alternative to the nonparametric methods. Considering an AR model has several justifications: pure sinusoidal variations can be expressed as AR models, and, any continuous spectral density function can be arbitrarily well approximated by AR models of high enough order.

Two datasets are analyzed in the next sections: the actual Global Oscillation Network Group's (GONG) (GONG, 1997) Month 10 ($T = 51840$ values from 1996/03/26 to 1996/04/30) $l = 85$ and $m = 0$ series, and an artificially simulated one that contains the same number of observations as the actual data.

The paper is organized as follows: The autoregressive (AR) model is presented in the next section. Then, methods for selecting the order of the AR process are discussed, followed by descriptions and comparison of three different methods of estimating the parameters of an AR process. Checks on the validity of the AR model assumptions are presented, and then the AR modeling is used to estimate the spectrum by prewhitening the spectrum of the original series. A brief description of a robust nonparametric spectrum estimation procedure (multitapering) is then presented, and the final section compares AR and nonparametric spectrum estimates. Based on these data, the analyses suggest that the AR method and the multitapering method result in very similar spectrum estimates for the long time series considered in the paper.

THE AUTOREGRESSIVE MODEL

The zero mean stationary process $\{X_t\}$ is an autoregressive process of order p, denoted AR(p), if it satisfies the following equation:

$$X_t = \phi_{1,p} X_{t-1} + \phi_{2,p} X_{t-2} + \cdots + \phi_{p,p} X_{t-p} + \epsilon_t, \tag{1}$$

where $\{\epsilon_t\}$ is a white noise process with variance σ_p^2, and the fixed coefficients $\phi_{1,p}, \phi_{2,p}, \cdots, \phi_{p,p}$ are such that all the roots of the polynomial equation $1 - \sum_{j=1}^{p} \phi_{j,p} z^{-j} = 0$ fall inside the unit circle. This latter condition is necessary to ensure the stationarity of the process. The process $\{X_t\}$ may be complex-valued, in which case the $\phi_{j,p}$ coefficients too are complex-valued.

The spectral density function of such a stationary AR(p) model has the form:

$$S(f) = \frac{\sigma_p^2 \Delta t}{\left|1 - \sum_{j=1}^{p} \phi_{j,p} e^{-i2\pi f j \Delta t}\right|^2}, \qquad |f| \leq f_{(N)}, \tag{2}$$

where Δt is the sampling interval and $f_{(N)} \equiv \frac{1}{2\Delta t}$ is the Nyquist frequency. Given estimates $\hat{\sigma}_p^2$ and $\hat{\phi}_{j,p}$, one forms the spectrum estimate:

$$\hat{S}(f) = \frac{\hat{\sigma}_p^2 \Delta t}{\left|1 - \sum_{j=1}^{p} \hat{\phi}_{j,p} e^{-i2\pi f j \Delta t}\right|^2}, \qquad |f| \leq f_{(N)}. \tag{3}$$

There are generally two steps in the estimation procedure: first, determine the order p of the process (which is usually unknown), and second, given that order, estimate the $p + 1$ parameters.

ORDER SELECTION

Of the various procedures proposed in the statistical literature to select the order of an AR model, see Percival and Walden (1993) and references therein, the partial autocorrelation- and the Akaike's information criteria (AIC) methods are used in this paper. Orders selected by the Bayes' information criterion (BIC) are also calculated and compared to the ones selected by the AIC.

The Partial Autocorrelation Method

This order selection procedure is based on the simple observation that for an AR(p) process, the partial autocorrelation sequence $\{\phi_{k,k}\}$ is zero for lags k greater than p. Assuming a Gaussian noise source, approximate confidence limits for the partial autocorrelation coefficients can be set. The order p can then be chosen by looking at the partial autocorrelation plot ($\hat{\phi}_{k,k}$ versus the lag k). See (Ripley and Venables, 1996). The estimates $\hat{\phi}_{k,k}$ are derived by solving the Yule-Walker equations, obtained by multiplying both sides of (1) by X_{t+k}, then taking expectations. The result is:

$$\rho_k = \sum_{j=1}^{p} \phi_{j,p}\rho_{k-j}, \qquad k > 0, \tag{4}$$

where $\rho_k = \text{Cov}\ \{X_t X_{t+k}\}$.

The estimated partial autocorrelation functions for the GONG Month 10 $l = 85$, $m = 0$ data and for a synthetic dataset are plotted on Figure 1. The figures only include terms up to lag 500. The behavior of both functions is the same for greater lags, mainly all the estimated coefficients fall between the rough 95% confidence bands, and so they are omitted for the clarity of the pictures. The horizontal lines giving the normal confidence limits are at $\pm 2/\sqrt{T}$, see Kay and Makhoul (1983) $T = 51840$. Note that the terms corresponding to the second lags are -0.85, and -0.86 for the GONG and for the simulated data correspondingly, but again, the axes limits are truncated so that the rest of the pictures is clearer. Based on these crude diagnostics, the upper limits for the order p of 200 for the GONG data and 150 for the simulated data seem sensible. Beyond those lags, assuming normality, at the 95% confidence level, the null hypotheses that all the partial autocorrelation coefficients are zero cannot be rejected.

Akaike's Information Criterion

Given that the maximum AR orders to consider are 200, and 150 respectively for the GONG and for the simulated data, are there any lower order AR models that are just as "good", or even "better" than the maximum order models? One standard frequently used in the statistical literature to evaluate the goodness of a model and to compare models of different sizes is the AIC criterion. Heuristically, the higher the order of the process, the less the noise process variability, but the higher the number of parameters to estimate. Any continuous spectrum can be approximated with a large enough order AR model [see Anderson (1971)], but there is a limit to how many parameters can be well estimated from the data. The task is to find a relatively low order model that has a small enough noise process variance estimate. Mathematically, the AIC criterion for an AR(p) process is defined as:

$$AIC(p) \equiv -2\log(\text{maximized likelihood}) + 2\text{p} \approx \text{N}\log(\hat{\sigma}^2_{\text{p}}) + 2\text{p}. \tag{5}$$

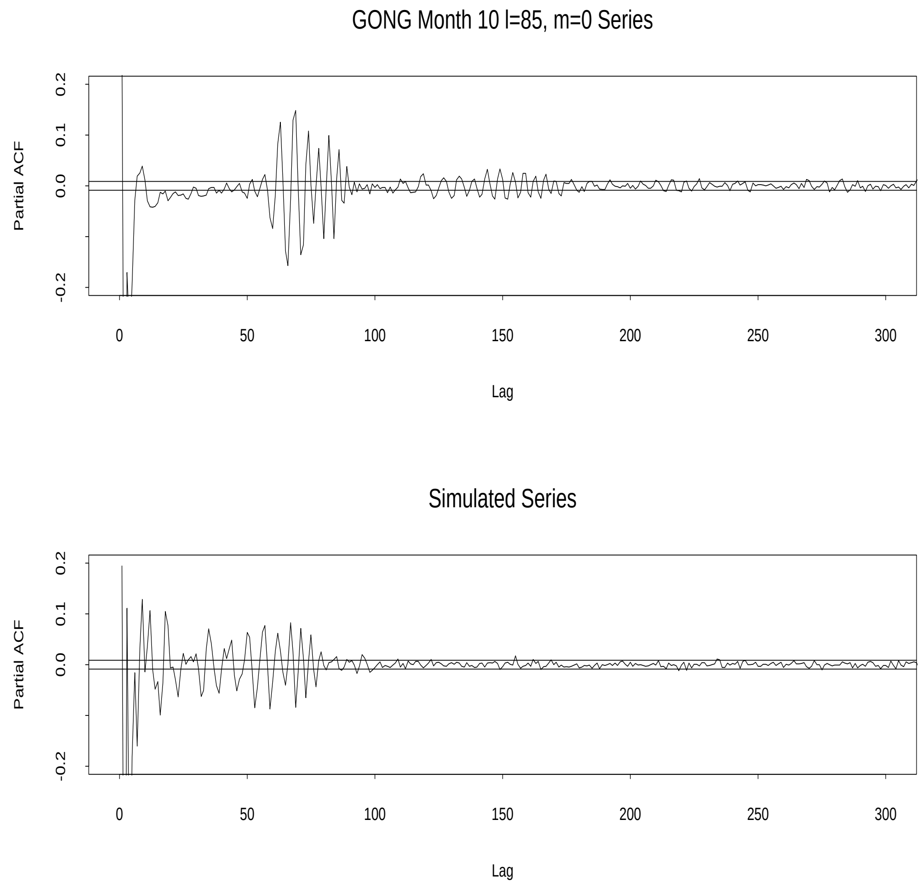

Figure 1: Estimated partial autocorrelation coefficients for GONG data and for simulated data. $\boxed{\text{partacorr}}$ [NR]

To select among competing models, the AIC criterion is first computed for the various models with different orders, and then the model with the least AIC value is selected as the "best" model. Throughout this paper, normal likelihoods are assumed.

Figure 2 shows the AIC(p) values versus orders less than or equal to the highest orders chosen in the previous subsection. The very first plot includes all the orders considered for the GONG data: there is clearly a huge drop in the AIC value at the third order, then another large drop from around order 60 to around 90, then a very slight decrease around order 150, then apparently the plot levels off. The second plot of the top panel presents the section of the first plot between orders 160 and 200. The AIC values drop continually until order 192 then they start to increase. Note however, the almost flat portion of the AIC for orders between 175 and 195. The performances of the different AR models with orders in that range are probably very comparable. The minimum of the AIC is achieved at order 192.

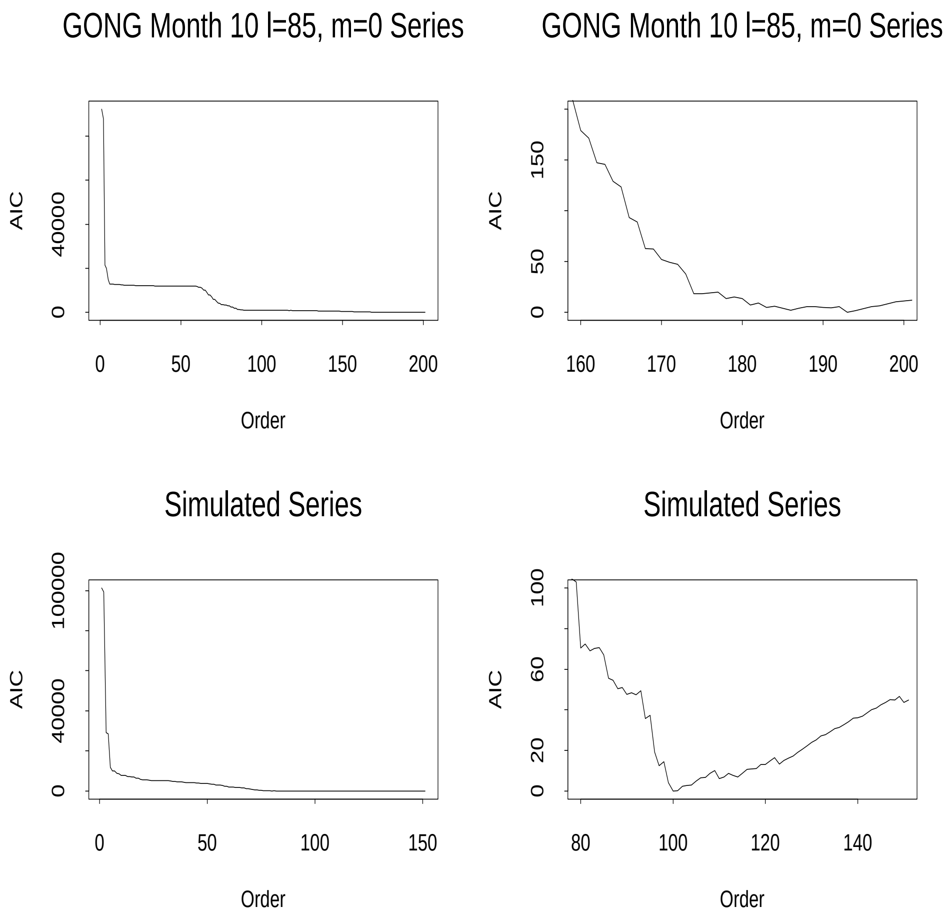

Figure 2: AIC plots for GONG and for simulated data. $\boxed{\text{aic}}$ [NR]

The two plots of the bottom panel illustrate the simulated data AIC. Again, the first figure shows the AIC versus all the lags, in this case up to 150, and the second one only a section around the minimum AIC value. The selected order for this process is 99.

A very similar alternative method to the AIC is the BIC. The BIC objective functional has a slightly different form:

$$BIC(p) \approx log(\hat{\sigma}_p^2) + p\frac{log(N)}{N}. \tag{6}$$

The results based on the BIC are plotted on Figure 3. The BIC selects 168 as the optimal order for the GONG data, and 80 for the simulated data. These selected orders are both smaller than the corresponding orders determined by the AIC. This is to be expected, as the BIC penalizes more for larger models than the AIC does. With a sample size of $T = 51840$, however, it seems like the extra cost of calculating 193 instead of 169 parameters for the GONG data, and 100 parameters instead of 89 for the artificial data is negligible.

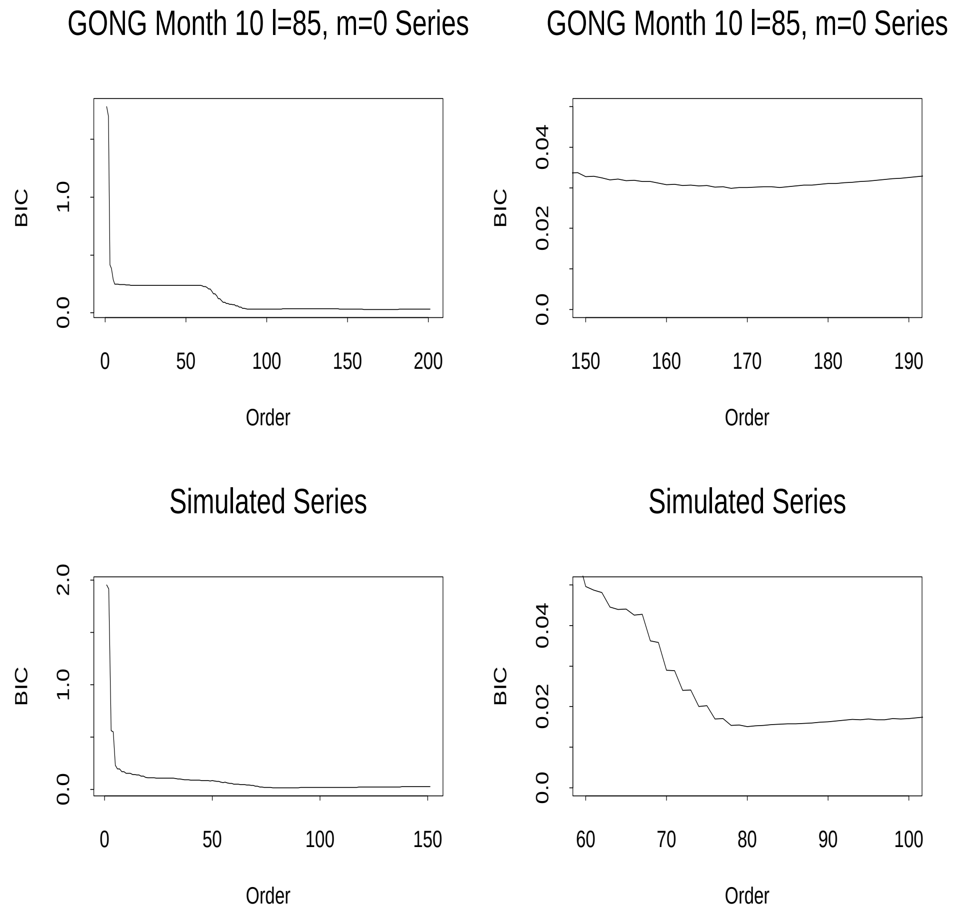

Figure 3: BIC plots for GONG and for simulated data. $\boxed{\text{bic}}$ [NR]

As a comparison of models with order selected by the AIC and by the BIC, Figure 4 overlaps the AR(99) and the AR(80) model for the simulated data.

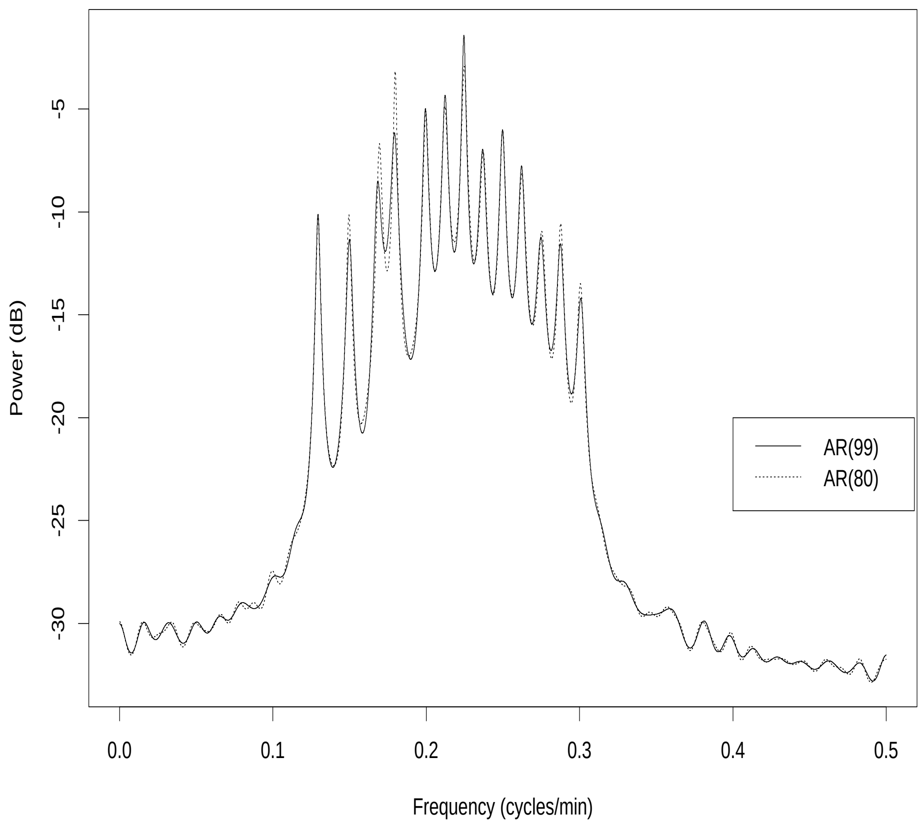

Figure 4: Estimated AR spectra for simulated data: AR(99) selected by the AIC, AR(80) by the BIC. $\boxed{\text{sim.2ar}}$ [NR]

AUTOREGRESSIVE PARAMETER ESTIMATION

There are various ways of estimating the parameters of an AR(p) process, once the order of the process is selected. This paper compares the performance of three different methods of computing the AR coefficients: solving the Yule-Walker equations (Percival and Walden, 1993), using Burg's algorithm (Burg, 1975), and using Marple's algorithm (Marple, 1980).

The Yule-Walker Equations

The oldest way of estimating the parameters of an AR(p) process is by substituting the p nonzero autocorrelation function estimates given by

$$\hat{\rho}_k = \frac{1}{T} \sum_{t=1}^{T-|k|} X_t X_{t+|k|}, \qquad k = 1, \cdots, p \tag{7}$$

into the the Yule-Walker equations (4).

The straightforward matrix algebra solution to the p simultaneous equations requires the inversion of a p by p matrix, and thus can become computationally intensive. The Levinson-Durbin algorithm to solve the Yule-Walker equations explores the Toeplitz structure of the matrix to be inverted, and thus solves for the AR parameters more efficiently. It also cuts the execution time by requiring only $O(p^2)$ operations, as opposed to the brute-force matrix inversion which requires $O(p^3)$ operations. Chapter 9 of (Percival and Walden, 1993) describes the algorithm in detail.

A common objection to estimating the AR parameters with the Yule-Walker equations is the use of the biased autocorrelation estimates $\hat{\rho}_k$ of (7) in (4): substituting the unbiased versions, $T - |k|$ instead of T in the denominator, causes numerical instabilities in the estimation algorithm. Note, however, that better autocorrelation estimates than the ones in (7), like tapered estimates, can be used. The only requirement is that the estimated autocorrelation matrix be positive definite. Chapter 9 of (Percival and Walden, 1993) is a again a good reference.

Burg's Algorithm

Burg argues that knowing p autocorrelation coefficients is not sufficient to identify a process uniquely. Based on p known autocorrelation coefficients, in Burg (1967) he proposes to choose the spectrum of the most random process among all processes whose autocorrelations up to order p agree with the known autocorrelations. However, generally there is no way of knowing a priori the autocorrelation coefficients. Burg (1975) proposes an alternative to the Levinson-Durbin recursions of solving for the AR coefficients: an alternative that unlike the former method does not assume that the process is zero outside the sampling interval. The algorithm is as computationally efficient as the Levinson-Durbin method. The merits and shortcomings of this algorithm are detailed by Percival and Walden (1993). For short series, Lysne and Tjostheim (1987) and Ulrych and Clayton (1976) claim that the Burg method produces more realistic estimates than the Yule-Walker method.

Marple's Algorithm

Marple's least squares algorithm (Marple, 1980) estimates the AR parameters by minimizing the sum of the forward and backwards prediction error squares. Its computational complexity is comparable to the complexity of the Levinson-Durbin, and of the Burg algorithms. Marple in (Marple, 1987), and Kay in (Kay, 1988) argue that Marple's algorithm outperforms both the Yule-Walker and the Burg algorithm. Rather than relying on the AIC or other order selection methods, Marple defines two new selection criteria: one based on the ratio of the residual prediction error energy to the total signal energy, the other based on the ratio of the residual prediction error energy at order k to the the residual prediction error energy at order $k-1$. The estimation procedure stops at order p if either the first or the second criterion is less than the corresponding user specified tolerance level. Marple suggests using both tolerance levels between 10^{-2} and 10^{-3}. Following those recommendations however, setting both cutoffs to 10^{-2} in the simulated dataset, results in selecting the order $p = 6$. The spectrum estimate based on that order only shows 2 peaks whereas in reality there should be 13 peaks. Only by decreasing the first tolerance to 10^{-4} and the second one to 10^{-5} does the estimated order increase to a level ($p = 108$) that captures the true structure of the underlying process. The spectrum estimate based on $p = 108$ and on the AR coefficients calculated with the Marple routine is barely distinguishable from the estimate obtained using the AIC selected $p = 99$ and coefficients calculated with the Yule-Walker equations.

Based on the artificial data, the AIC (or BIC) order selection methods appear to be more reliable than those suggested by Marple. To estimate the AR coefficients with the Marple method, first the tolerances are tuned to force the order selection to the order p selected by the AIC criterion. For the synthetic dataset, the values of the two tolerances that guarantee the selected order to be 99 are 0.13366 and 10^{-5}, correspondingly.

Comparing the Methods

First, the order of the process is determined via the AIC, then the coefficients are estimated in the previously described three fashions. The results for the simulated dataset are compared in Figure 5.

Qualitatively and quantitatively, the three spectrum estimates seem very similar. Figure 6 magnifies the details of the previous plot in the range of the known generating

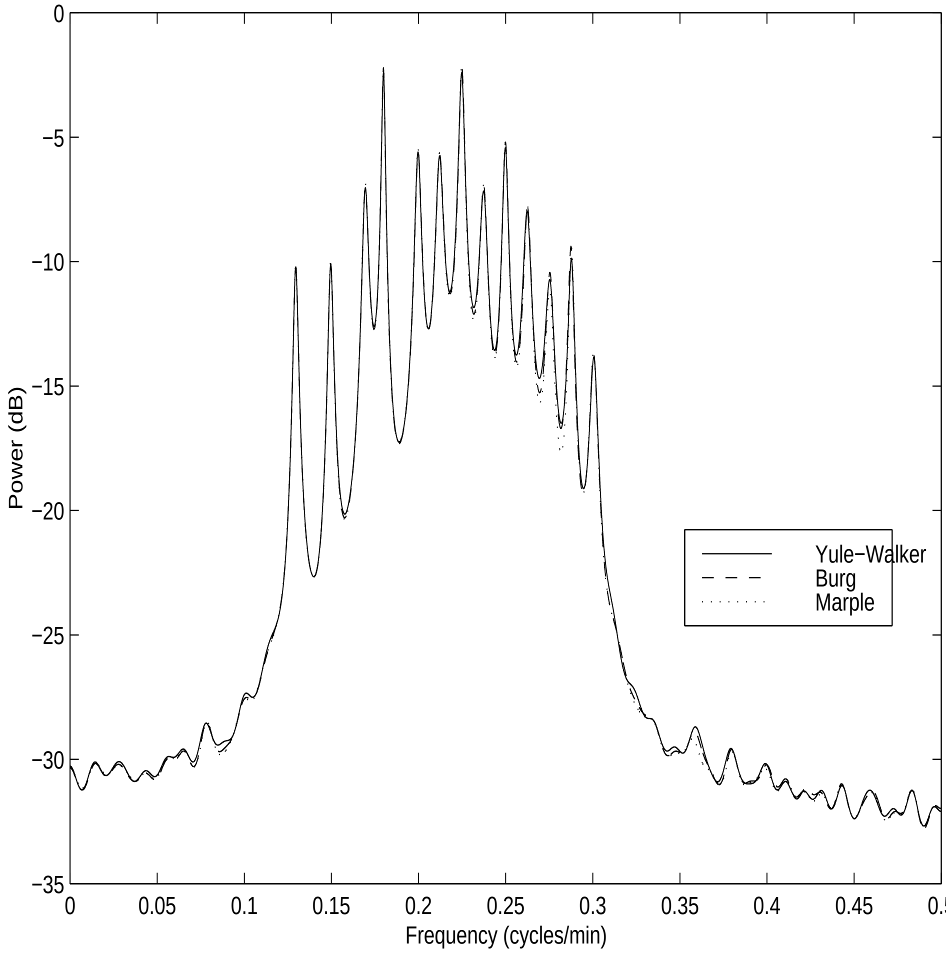

Figure 5: Estimated AR spectra for simulated data using different estimation methods. sim.yw.brg.mrpl.spec [NR]

frequencies. Vertical lines are added at the exact locations of those 13 frequencies. As the plot shows, the three AR spectrum estimates agree to a high degree of accuracy - however, all three of them seem to be slightly biased at some of the frequencies, especially at the peaks located at the highest and lowest frequencies in the band.

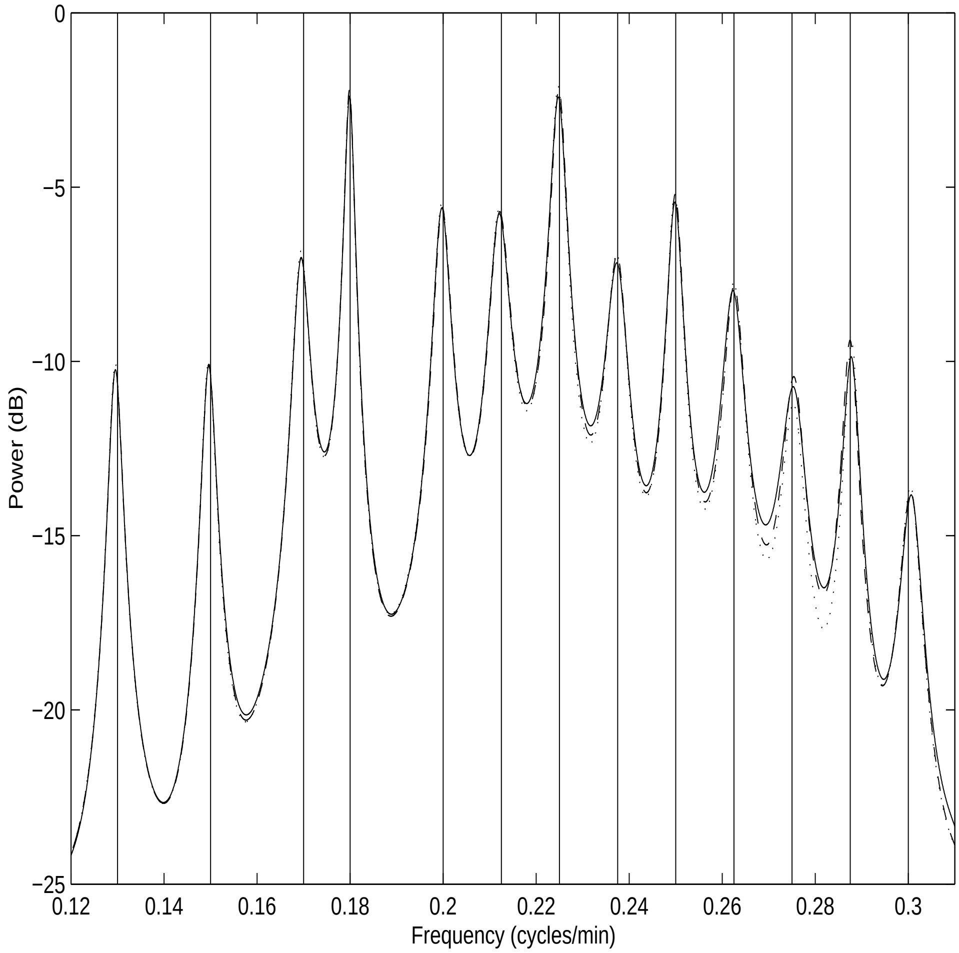

Figure 6: Estimated AR spectra for simulated data using different estimation methods. $\boxed{\text{sim.yw.brg.mrpl.spec2}}$ [NR]

Figure 7 illustrates the estimated AR spectra for the GONG and for the simulated data, both with orders estimated with the AIC, and using the Yule-Walker equations. The simulated data were generated as the superposition of 13 different sinusoidal modes; clearly, the estimated AR(99) spectrum does indicate the presence of 13 frequencies.

MODEL CHECKING

The four plots of Figure 8 present the periodogram, the quantile-quantile (qq) plot versus the standard normal distribution, the cumulative periodogram, and the esti-

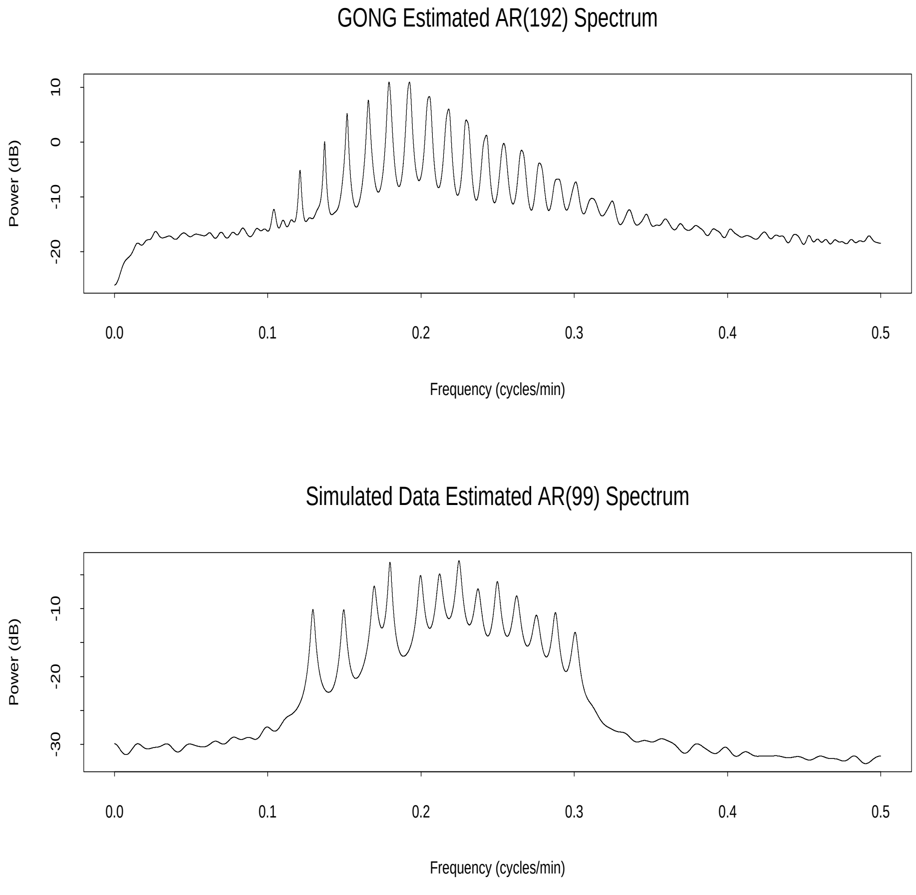

Figure 7: Estimated AR spectra for GONG and for simulated data. gong.sim.ar [NR]

mated autocorrelation function for the residuals from the AR(192) model fitted to the GONG data. The relative flatness of the periodogram suggests a white noise process. Note however, the dynamic range of the process: most of the power is between about -30 and -10 decibels, but with extreme values ranging from -50 to -5 decibels.

The qq plot indicates that, apart from a few outliers, the empirical distribution of residuals seems to be consistent with a normal distribution. The cumulative periodogram tests for Gaussian white noise using the Kolmogorov-Smirnov goodness of fit test (Conover, 1980). As the normalized periodogram values fall between the upper and lower limits, there is not sufficient evidence at the 95% significance level to reject the null hypothesis of white noise.

The estimated residual autocorrelation function plot, however, suggests that the residuals are correlated at many lags.

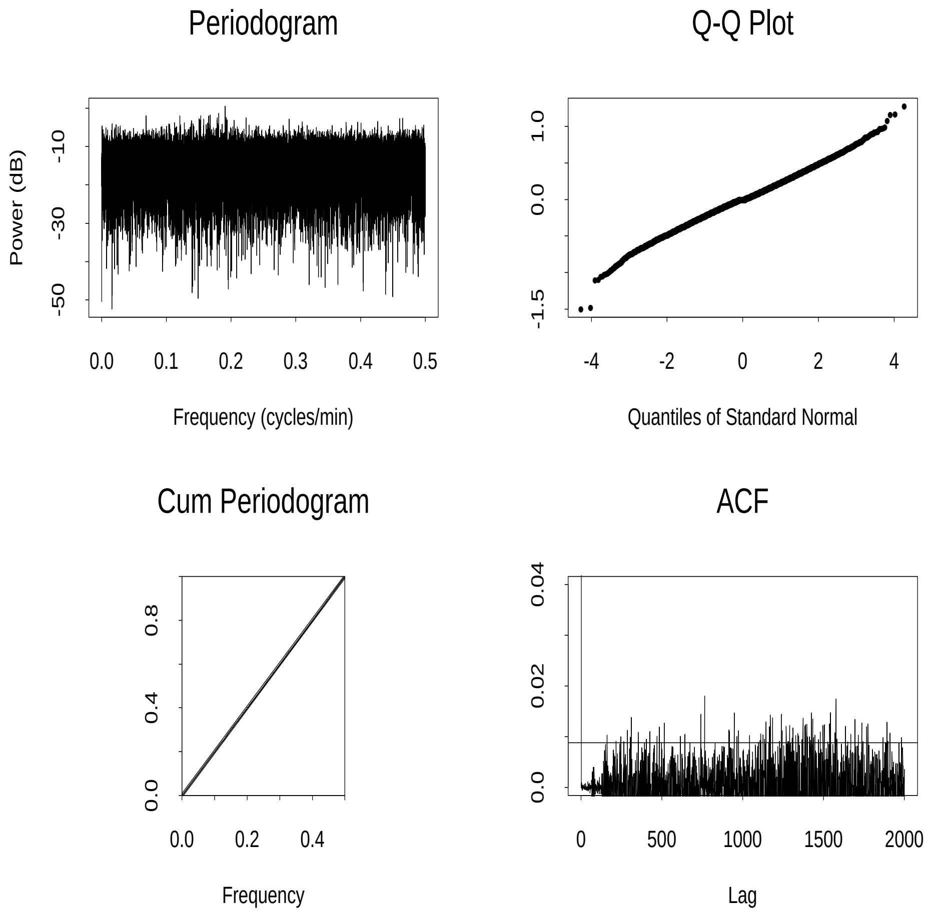

Figure 8: Different diagnostics for the residuals from the fitted AR(192) model to GONG data. $\boxed{\text{gong.ar192.res}}$ [NR]

Similar diagnostics for the simulated data are shown on Figure 9. In this case the spectrum is white again, but the qq plot indicates departure from normality. As in the GONG data case, the residuals are correlated.

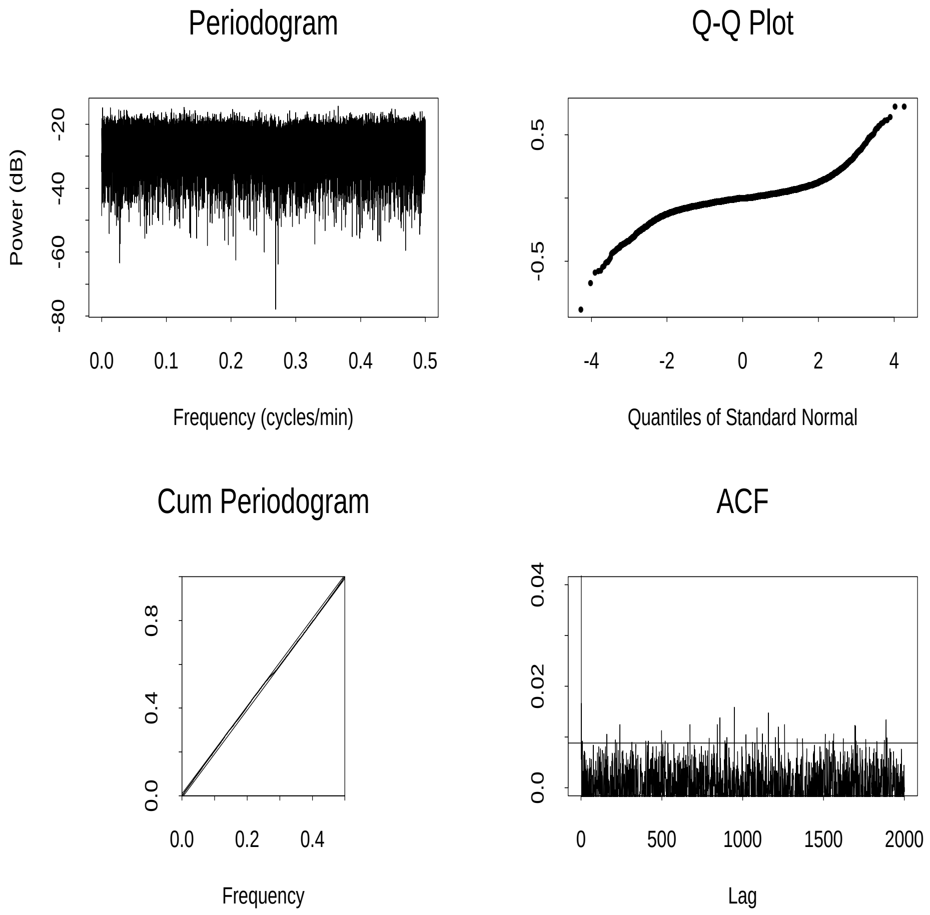

Figure 9: Different diagnostics for the residuals from the fitted AR(99) model to the simulated data. $\boxed{\text{sim.ar99.res}}$ [NR]

AUTOREGRESSIVE MODELING AS PREWHITENING FILTER

The pure parametric approach of AR modeling requires strong indications that the process under investigation is truly an AR process. If such an assumption is questionable, AR modeling can still be of use as a prewhitening filter. If the original process has high dynamic range, then power leakage can bias the classical nonparametric spectral estimates. A means of controlling the bias is to prewhiten the series: construct a new series that is a linear filtered version of the original series such that the new one has smaller dynamic range. Then estimate the spectrum of the new series in a nonparametric way, say using direct spectral estimators, and finally use the theory of linear filters to construct a spectrum estimate of the original series based on the spectrum estimate of the filtered series. The main difference between this combined parametric and nonparametric approach and the pure parametric approach is that while the latter assumes a white residual noise process, the former actually estimates the residual spectrum in a nonparametric way. The AR order selection is not as crucial in the former approach as it is in the latter; in the combined procedure the nonparametric part is actually able to compensate for any inaccuracies caused by the parametric part. In the pure AR case, however, the right order selection can be very important; studies suggest that being off just by a couple of orders can lead to very different final spectrum estimates.

Suppose an AR(p) model is fit to the data. Call the resulting parameter estimates $\hat{\phi}_{1,p}, \cdots, \hat{\phi}_{p,p}$ and $\hat{\sigma}_p^2$. Then form the residual series as:

$$e_t(p) \equiv X_t - \hat{\phi}_{1,p} X_{t-1} - \cdots - \hat{\phi}_{p,p} X_{t-p}, \qquad t = p+1, \cdots, N. \tag{8}$$

The relationship between the residual and the original series spectral density functions, $S_e(f)$ and $S_X(f)$ respectively, is given by:

$$S_e(f) = \left| 1 - \sum_{j=1}^{p} \hat{\phi}_{j,p} e^{-i2\pi f j \Delta t} \right|^2 S_X(f). \tag{9}$$

Given an estimate $\hat{S}_e(f)$ with good bias and variance properties, one obtains:

$$\hat{S}_X(f) = \frac{\hat{S}_e(f)}{\left| 1 - \sum_{j=1}^{p} \hat{\phi}_{j,p} e^{-i2\pi f j \Delta t} \right|^2}. \tag{10}$$

Figure 10 presents the prewhitened spectrum estimates for the GONG data and for the simulated data.

MULTITAPERED SPECTRUM ESTIMATES

The previous sections describe the pure parametric and the mixed parametric and nonparametric approaches to spectrum estimation, both methods assuming, to some degree at least, that the process under study is a realization of an AR process. In this section, the results of a purely nonparametric spectral estimation procedure are presented.

Suppose one is interested in estimating the spectrum $S(f)$ of a zero mean, stationary random process $\{X_t\}$. The general multitaper estimate based on the realization $x_0, x_1, ..., x_{T-1}$ is of the form:

$$S(\hat{f}) = \sum_{k=0}^{K-1} w_k \left| \sum_{t=0}^{T-1} v_t^{(k)} x_t e^{-i2\pi f t} \right|^2, \tag{11}$$

where $\mathbf{v}^{(k)} = (v_0^{(k)}, ..., v_{T-1}^{(k)})$ for $k = 0, ..., K-1$ are K tapers. The estimate above is a combination of K eigenspectrum estimates. The w_k are the corresponding weights

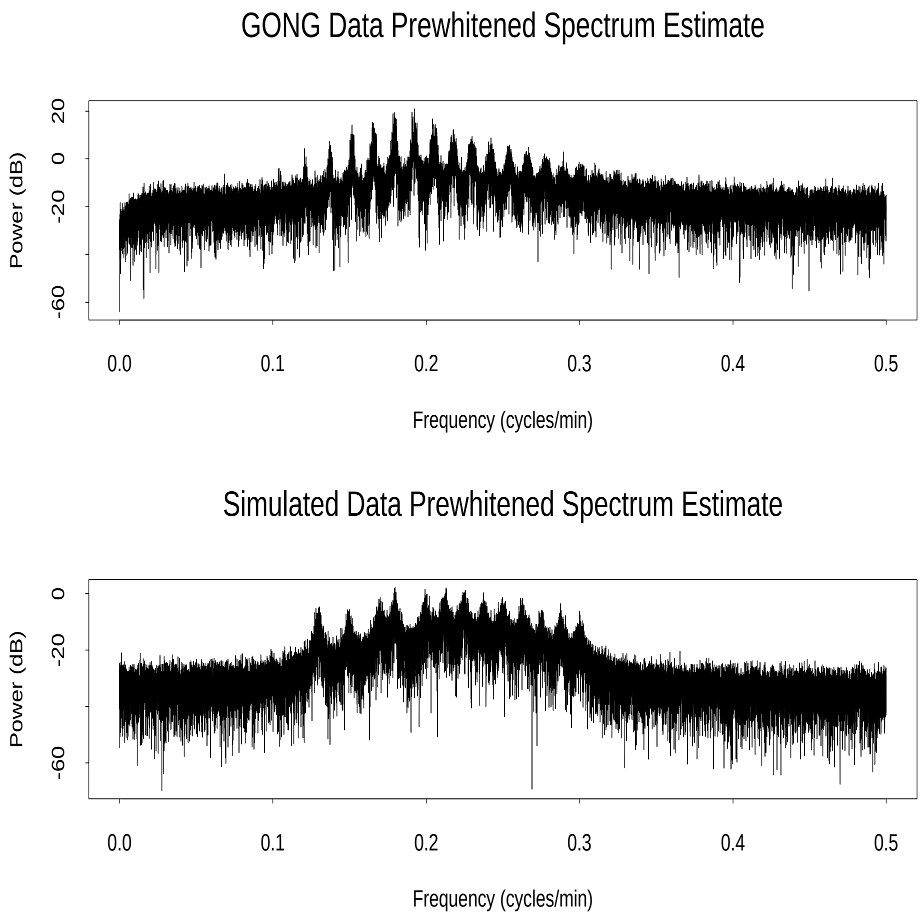

Figure 10: Prewhitened GONG and simulated data spectrum estimates. gong.sim.prwhtn [NR]

the individual tapered estimates get in the final multitapered version. The simplest version of multitapering is to use equal weights $w = 1/K$.

First fix K=1 in (1). Then $w_0 = 1$ and $\mathbf{v}^{(0)} = (1/\sqrt{2\pi T}, ..., 1/\sqrt{2\pi T})$ corresponds to the familiar periodogram. The periodogram is inconsistent and generally a biased estimate of the true spectral density. It is inconsistent because its variance does not approach zero as the number of observations increases. The remedy is to "smooth" it, which is equivalent to multiplying the data by a taper before computing the periodogram. Because of the side-lobes of the Fejér kernel power is transferred among different regions of the spectrum (leakage) and cause the periodogram to be biased. The formula below illustrates the point:

$$E\{S(\hat{f})\} = (2\pi T)^{-1} \int_{-1/2}^{1/2} F(f-\mu) S(\mu) d\mu, \tag{12}$$

where $F(\lambda) = (2\pi T)^{-1} \frac{\sin^2(N\lambda/2)}{\sin^2(\lambda/2)}$ is the Fejér kernel.

Unless the spectrum of interest is nearly flat, employing a non-constant taper reduces the broad-band bias due to leakage and thus increases the resolution of the estimate.

Carefully chosen tapers reduce leakage by replacing the Fejér kernel with kernels possessing smaller side-lobes. Note, however, the variance-bias tradeoff. Reducing the broad-band bias (reducing the side-lobes of the kernel) amounts to increasing the variance and the local bias (increasing the central lobe of the kernel). A good spectral estimate should have both small bias and small variance. This is where multitapering, K greater than 1, proves to be helpful. Combining several eigenspectrum estimates with good bias properties preserves the small bias and reduces the variance of the resulting multitaper spectral estimate. If the tapers are mutually orthogonal, then, under the conditions of Theorem 4.4.2 of Brillinger (1981b), their Fourier transforms will be asymptotically independent zero-mean normal variates, so that the sum of the eigenspectra will be asymptotically distributed as a χ^2 variable. Confidence intervals can then be easily constructed for the multitapered spectral estimate. Internal measures of variability such as the jackknife, free of any distributional assumptions, were proposed in the literature by Thomson and Chave (1991) as alternative ways to calculating variance estimates of the multitaper spectrum estimate. References to nonparametric spectral estimation include Brillinger (1981a), Brillinger (1981b), and Percival and Walden (1993).

The Generalized Sinusoidal Tapers

An orthogonal taper family based on sinusoids that requires little computational effort is proposed by Riedel and Sidorenko (1995). Those tapers are approximations to the so-called minimum bias tapers, also proposed by Reidel and Sidorenko (1995), as a family with good variance and bias properties, and therefore as excellent candidate for multitapering data with no gaps in the observations. The GONG data however, like any other helioseismic data, contain occasional missing observations. Define the gap structure $\mathbf{I} = (I_0, ..., I_{T-1})$ as the sequence of indicators specifying whether the observation at time t is available. The case that all the realizations are available corresponds to $I_t = 1$ for all $t \in \{0, 1, ..., T-1\}$. The generalized sinusoidal taper family has K sequences of length T as members: $\mathbf{v}^{(k)} = (v_0^{(k)}, ..., v_{T-1}^{(k)})$, $k = 0, ..., K-1$, the k-th order generalized sine taper $\mathbf{v}_s^{(k)}(T)$, being:

$$\mathbf{v}_s^{(k)}(T) = (v^k_{(s,0)}(T), ..., v^k_{(s,T-1)}(T)), \tag{13}$$

with

$$v^{(k)}_{(s,t)}(T) = I_t \sqrt{\frac{2}{T+1}} \sin \frac{\pi(k+1)t}{T+1}. \tag{14}$$

The orthogonality of the sine tapers is not necessarily preserved by the above operation. To obtain the final version of the generalized sine tapers, an additional step of re-orthogonalization is needed. Figure 11 shows the first two generalized sine tapers corresponding to the GONG Month 10 gap structure.

PARAMETRIC AND NON-PARAMETRIC SPECTRUM ESTIMATES

Figure 12 presents the parametric AR(99) and the nonparametric 100 sine tapered spectrum estimates. There is little difference between the two very differently obtained estimates! Both of them recover the overall shape of the underlying process and the 13 peaks at the generating frequencies. Note that the multitapered estimate is almost as smooth as the AR(99) estimate, particularly in the region of the peaks. In the region away from the generating frequencies, the multitapered estimate is slightly rougher than the smooth AR(99) estimate, but it is remarkable how similar in shape they are. It is certainly very encouraging that two fundamentally different methods produce spectrum estimates that are so alike, both close to the true spectrum. Does it follow that spectrum estimation via multitapering and via AR modeling are exchangeable? To answer that question, it is necessary to investigate the assumptions of the two procedures.

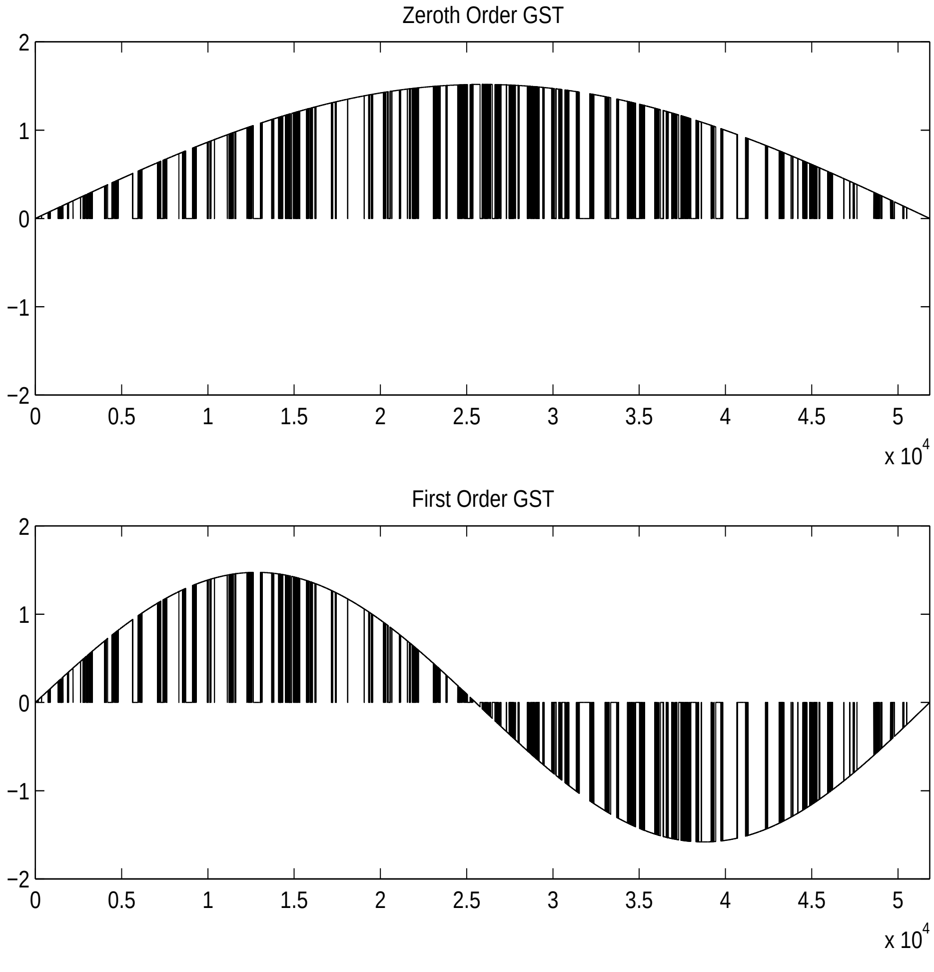

Figure 11: Generalized sinusoidal tapers. $\boxed{\text{st2}}$ [NR]

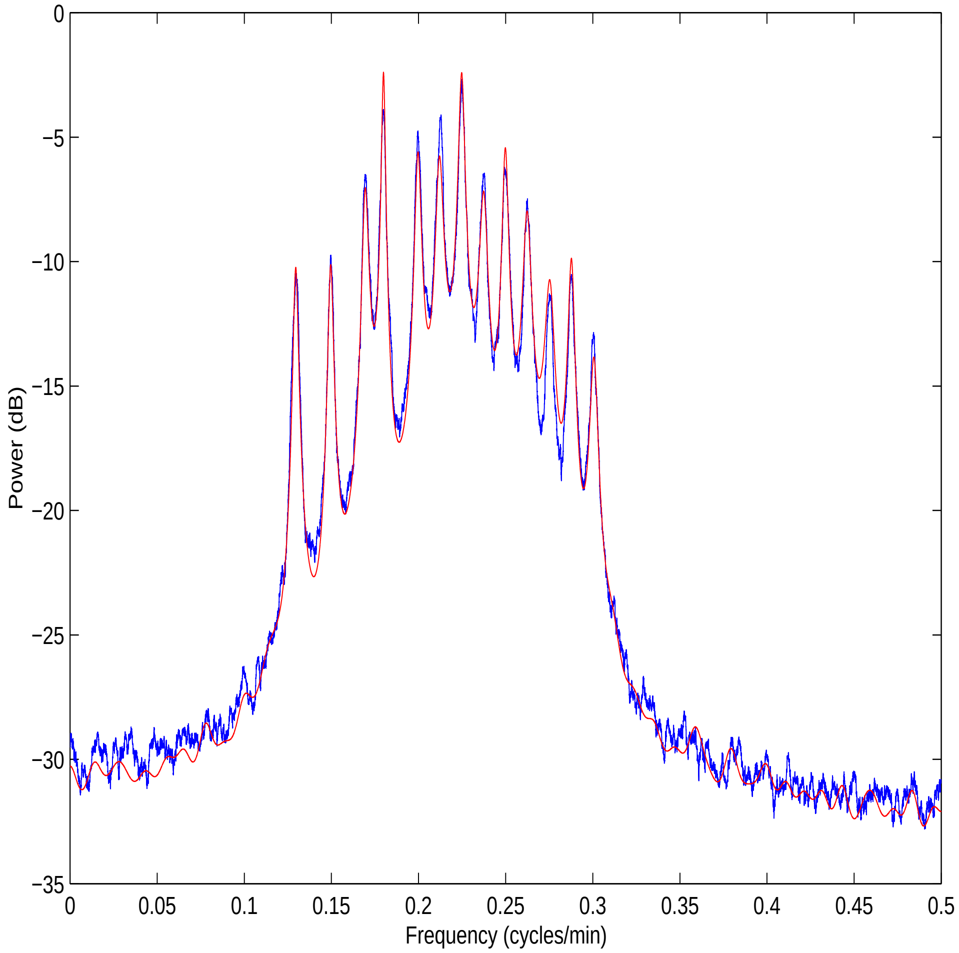

Figure 12: Simulated data AR(99) (red) and 100 sine tapered (blue) spectrum estimates. $\boxed{\text{x13.2r.100st.yw.spec}}$ [NR]

First, there is a fundamental "philosophical" difference: the multitapering method is completely free of distributional assumptions, while the AR method assumes a certain functional form for the spectral density to be estimated. If there is prior knowledge that the data under investigation is a realization of an AR process, then assuming an AR model does make sense. The nonparametric multitapering method can be used in the absence of any distributional assumptions, whenever the spectrum of interest has a large dynamic range. As any continuous spectrum can be approximated by a long enough AR process, AR modeling can be used to estimate the spectrum of such process, but care must be exercised in determining the order of the model and in quantifying the goodness of the fit.

Second, the multitapering and the AR approaches have different free parameters: The number of tapers in the multitapering method, and the order of the AR process in the AR modeling. How to select the optimal number of tapers to use is a classical variance-bias trade-off problem that can be quantified (Percival and Walden, 1993): Increasing the number of tapers amounts to decreasing the variability, decreasing the broad-band bias due to power leakage, and increasing the local bias. The more tapers used, the smoother the resulting spectrum estimate is. Picking the order based on a realization of an assumed AR process is of different nature: Any estimated AR spectrum is smooth by definition; the order determines the details of the estimate. If too low of an order is selected, structure in the underlying spectral distribution might be missed, if too high an order is selected, spurious peaks might be introduced in the estimate.

Third, the bias and the standard error of the multitapered spectrum estimate can be obtained using the jackknife technique. Discussed in detail by Efron and Tibshirani (1993), this statistical procedure provides an empirical way, completely free of any distributional assumptions, to estimate the variability of the multitapered estimate. Essentially, if there are K tapers, a multitapered estimate based on $K-1$ tapers is computed K times, each time leaving one taper out. Then, the properly scaled standard error of those K "leave-one-out" estimates provides an estimate of the standard error of the K multitapered spectrum estimate. Traditional variability estimates, both for the multitapered and for the AR methods, rely on distributional assumptions and asymptotics.

Fourth, the issue of missing data needs to be addressed in the AR modeling. The AR estimation algorithms used in this paper assume the data values of zero whenever there were gaps in the GONG Month 10 series. The codes need to be adjusted to deal with this issue properly. The tapers, on the other hand, were specifically designed to take Month 10's gap structure into account, and therefore are constrained to be zero at the locations of the missing values - see Figure 11. Note that other orthogonal taper families, such as the discrete prolate spheroidal sequences proposed by Thomson (1982), and discussed in detail by Percival and Walden (1993), and the minimum bias tapers proposed by Riedel and Sidorenko (1995), and their generalized versions for data with missing observations do exist. The former depend on the bandwidth parameter W, $W \leq 0.5$, and are the vectors with maximal energy concentration in the frequency band $[-W, W]$, designed to suppress power leakage from the resonant modes. The latter minimize the local bias, defined as the integral of the product of frequency squared times energy at that frequency.

Finally, the computational complexity of calculating the general sine multitapered estimate and calculating the AR estimate is comparable. Because of the closed form expression of the sine tapers, they can be evaluated almost instantly even for very long datasets. The additional step of reorthogonalization in the missing values case does not significantly increase the calculation time. Calculating K of the "best" prolate or minimum bias tapers however, amounts to solving for the K largest eigenvalues and associated eigenvectors of a T by T matrix, where T is the length of the series. Using the block Lanczos algorithm (Parlett, 1980) with reorthogonalization, and special algorithms that exploit the Toeplitz structure of the large matrices, the running time to obtain 100 tapers, each of length $T = 51840$, takes more than 200 hours on a Sun Sparc Ultra 2 machine. Computationally, therefore, the cost of getting the prolate or the minimum bias tapers and then calculating the multitapered estimate with those tapers is higher than calculating the AR spectrum estimate.

In conclusion, the autoregressive and the generalized sine multitaper spectral analysis of the solar oscillations data considered in this paper result in very comparable estimates. Based on the simulation data, where the location of the modes is known, it is determined that both methods recover the true structure of the spectrum. The spectrum of the observed data thus might be estimated using either methods, provided that the process that generates the GONG data resembles the process underlying the synthetic data. It is up to the investigator to know the advantages and limitations of both methods, and to make an informed decision about which method to use.

REFERENCES

Anderson, T. W., 1971, The statistical analysis of time series: John Wiley and Sons, New York.

Blackman, R. B., and Tukey, J. W., 1958, The measurement of power spectra: Dover Publications, New York.

Brillinger, D. R., 1981a, The key role of tapering in spectrum estimation: IEEE Trans. Acoust. Speech and Signal Processing, **29**, no. 5, 1075–1076.

Brillinger, D. R., 1981b, Time series, data analysis and methods: Holden-Day, San Francisco, California.

Burg, J. P., 1967, Maximum entropy spectral analysis: Maximum entropy spectral analysis:, Proceedings of the 37th Meeting of the Society of Exploration Geophysicists.

Burg, J. P., 1975, Maximum entropy spectral analysis: Ph.D. thesis, Department of Geophysics, Stanford University.

Conover, W. J., 1980, Practical nonparamteric statistics. second edition.: John Wiley and Sons, New York.

Efron, B., and Tibshirani, R. J., 1993, An introduction to the bootstrap: Chapman and Hall, New York.

GONG. http://www.gong.noao.edu/index.html:, 1997.

Kay, S. M., and Makhoul, J., 1983, On the statistics of the estimated reflection coefficients of an autoregressive process: IEEE Transactions on Acoustics, Speech, and Signal Processing, **31**, 1447–1455.

Kay, S. M., 1988, Modern spectral estimation: Theory and application: Prentice Hall, Englewood Cliffs, New Jersey.

Lysne, D., and Tjostheim, D., 1987, Loss of spectral peaks in autoregressive spectral estimation: Biometrika, **74**, 200–206.

Marple, S. L., 1980, A new autoregressive spectrum analysis algorithm: IEEE Transactions on Acoustics, Speech, and Signal Processing, **28**, 441–454.

Marple, S. L., 1987, Digital spectral analysis with applications: Prentice Hall, Englewood Cliffs, New Jersey.

Parlett, B. N., 1980, The symmetric eigenvalue problem: Prentice Hall, Englewood Cliffs, New Jersey.

Percival, D. B., and Walden, A. T., 1993, Spectral analysis for physical applications. multitaper and conventional univariate techniques: Cambridge University Press, Cambridge, England.

Riedel, K. S., and Sidorenko, A., 1995, Minimum bias multiple taper spectral estimation: IEEE Transactions on Signal Processing, **43**, no. 1, 188–195.

Ripley, B. D., and Venables, W. N., 1996, Modern applied statistics with s-plus: Springer Verlag, Berlin.

Thomson, D. J., and Chave, A. D., 1991, Jackknifed error estimates for spectra, coherences, and transfer functions, *in* Haykin, S., Ed., Advances in Spectrum Analysis and Array Processing: Prentice Hall, 1, 58–113.

Thomson, D. J., 1982, Spectrum estimation and harmonic analysis: Proc. IEEE, **70**, 1055–1096.

Ulrych, T. J., and Clayton, R. W., 1976, Time series modeling and maximum entropy: Phys. Earth and Plan. Int., **12**, 188–200.

Yuen, C. K., and Fraser, D., 1979, Digital spectral analysis: CSIRO/Pitman, East Melbourne, Australia.