

## Practical Test for Goodness of Fit of Low-Order AR Models

Karim J. Rahim\*

David J. Thomson\*

### Abstract

This paper presents a practical test for goodness-of-fit of low-order Autoregressive (AR) models. The test compares the maximum absolute deviation between the estimated integrated spectrum and the theoretical integrated spectrum of an AR model. We demonstrate using simulation that the multitaper method results accurate estimates of AR model coefficients. Simulation is used to study the effectiveness of the goodness-of-fit test. Recently Tourre et al. (2011) presented a spectral analysis of the Burgundy Pinot Noir Grape Harvest Date (GHD) series using the multitaper method, and this analysis used an AR(1) model to prewhiten the data. We fit several AR models to the Pinot Noir series, then using our goodness-of-fit test, we determine that several AR models, including an AR(1), are reasonable for the GHD series.

**Key Words:** Time Series, Spectral Estimation, Goodness of Fit, Multitaper Method, Autoregressive Coefficient

### 1. Introduction

In this paper we: (a) use simulation to show that multitaper spectral estimation in conjunction with the Levinson–Durbin recursions provide accurate selection of autoregressive (AR) coefficients, (b) propose a practical test for assessing the goodness-of-fit of AR coefficients, (c) test, using simulation, the proposed goodness-of-fit test, and (d) fit several AR models the Burgundy Pinot Noir Grape Harvest Date (GHD) time series. Our tests find that several models are acceptable for this series. This paper does not consider the problem of AR coefficient order selection, and in practical application we use the Akaike Information Criterion (AIC).

AR models are used in many applications, for example they are used to prewhiten data in engineering applications (Thomson 1977), and to prewhiten data in climate science prior to harmonic analysis (Mann and Lees 1996). The choice of AR model and the estimated AR coefficients used in prewhitening can affect the residuals and the subsequent harmonic analysis, thus masking or enhancing features of the spectrum. Using simulation, we show that AR coefficients obtained by different methods are not equally distributed, and we find in our simulations that the Levinson–Durbin recursions with multitaper spectral estimates and Burg’s algorithm produce unbiased low variance estimates. We fit and compare the goodness-of-fit of several AR models to the Burgundy Pinot Noir GHD series, which has been used in the climate literature (Chuine et al. 2004; Tourre et al. 2011).

Goodness of fit for AR models have been proposed and discussed in the literature, (Priestley 1981, pp. 475–494) and Anderson (1997). Our test is based on the maximum absolute deviation of the integrated spectrum, originally proposed by Bartlett (Priestley 1981, p. 479), and as a practical point, our test use simulation to determine approximate p-values.

There are multiple methods for fitting AR coefficients, and we review two popular methods: (a) solving the Yule–Walker equations with Levinson–Durbin recursions, and (b) Burg’s recursions using forward and backward estimators. The former can be improved by using a better (multitaper) estimator of the autocorrelation function.

---

\*Dept. of Math and Stats, Queen’s University, Jeffery Hall University Ave., Kingston ON K7L 3N6  
email: karim@mast.queensu.ca

This paper is organized in the following manner. Section 2 discusses some of the basic theory of AR coefficients, and it reviews two of the procedures used in calculation of the coefficients, section 3 reviews some general cautionary notes regarding the use of AR models, section 4 compares several methods used in obtaining AR coefficients, section 5 discusses goodness-of-fit tests for AR coefficients, section 6 presents a simulation analysis of goodness-of-fit tests, section 7 compares various fitted AR models for the Burgundy GHD series, and section 8 gives concluding remarks and suggests future work.

## 2. Calculation of AR Coefficients

### 2.1 Preliminaries

**Definition 1.** If  $\{Z_t\}$  is a purely random process with zero mean and variance  $\sigma_z^2$ , indexed by  $t = 1, 2, \dots, N$ . The process  $\{X_t\}$  is an AR process of order  $p$  (denoted as an AR( $p$ ) process), and we have

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t. \quad (1)$$

Subject to certain constraints AR( $p$ ) processes are often considered second-order stationary, meaning the first and second moments are time invariant. See Chatfield (2004, pp. 43–44) for the constraints. The coefficients  $\phi_1, \phi_2, \dots, \phi_p$  are called *autocorrelation* coefficients.

*Remark 1.* Equation (1) notes an analogy between an AR( $p$ ) model and a regression problem; however, instead of independent variables the right side of equation (1) has lagged copies of the dependent variable.

*Remark 2.* The  $p^{\text{th}}$  coefficient of an AR( $p$ ) process is called a *reflection* or *partial autocorrelation* coefficient. One can use the notation  $\hat{\phi}_{1,p}$  to indicate the first sample autocorrelation coefficient in an order  $p$  model, and thus indicate it differs from  $\hat{\phi}_{1,1}$  which would be the only autocorrelation coefficient in an AR(1) process. We will use the two subscript notation if the order,  $p$ , is unclear or changing.

**Definition 2.** The partial autocorrelation coefficient  $\phi_{j,j}$  represents the correlation between  $X_t$ , and  $X_{t-h}$  with the linear dependence of the interceding terms,  $X_{t-1}, X_{t-2}, \dots, X_{t-h+1}$  removed, or *partialled* out.

*Remark 3.* An alternative notation for AR( $p$ ) often used in engineering and in Priestley (1981) processes is:

$$Y_t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} = Z_t, \quad (2)$$

where  $\alpha_j = -\phi_j$ .

**Definition 3.** The autocovariance sequence (acvs) for lag  $\tau$  is defined as:

$$\gamma_\tau = E\{[X_t - \mu][X_{t-\tau} - \mu]\}, \quad (3)$$

where  $\mu$  is the mean of the process  $X_t$ , and  $\tau = 0, 1, \dots, N - 1$ . We see that if  $\tau = 0$ , then  $\gamma(0)$  is simply the variance.

**Definition 4.** The autocorrelation sequence (acs) is defined as:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \quad (4)$$

and  $\rho_0 = 1$ .

*Remark 4.* A plot of the sample autocorrelation coefficients over increasing lag,  $\tau$ , is known as a correlogram.

**Definition 5.** The typical *biased* estimator of the acvs is

$$\hat{\gamma}_\tau^B = \frac{1}{N} \sum_{\tau=0}^{N-|\tau|} [X_t - \bar{X}][X_{t+|\tau|} - \bar{X}]. \quad (5)$$

*Remark 5.* If one replaces  $\bar{X}$  with  $\mu$ , and multiplies by  $\frac{N}{N-|\tau|}$  one would have an unbiased estimate; however, simply making the second substitution while using an estimator of  $\mu$  will not produce an unbiased estimator. See Percival and Walden (1993, pp. 190–191), hereinafter abbreviated as (P&W 1993), and Anderson (1971, pp. 448–449) for more details.

**Theorem 1.** *The sequence formed by (5) is positive definite if and only if the realizations of  $X_1, X_2, \dots, X_N$  are not all identical (P&W 1993, p. 195).*

*Remark 6.* Generally Fast Fourier Transforms (FFTs) are used instead of calculating the autocovariance sequence in equation (5) directly. See remark 8.

**Definition 6.** We will often examine and estimate of the Spectral Density Function (sdf), which is the Fourier transform of the acvs,

$$S(f) = \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{-i2\pi f\tau}. \quad (6)$$

In equation (6), we allow  $\tau \in \mathbb{Z}$ , thus we are considering a process with infinite past and future. Frequency,  $f$ , takes values in  $[0, 1/2]$ . We note the above equality is true only in mean-square sense, but it can be considered point-wise in all practical applications (P&W 1993, p. 132).

*Remark 7.* The sdf for a stationary AR( $p$ ) process is

$$S_{AR}(f) = \frac{\sigma_Z^2}{\left|1 - \sum_{j=1}^p \phi_j e^{-i2\pi f j}\right|^2}, \quad \text{for } |f| \leq 1/2, \quad (7)$$

when  $\Delta t = 1$ . In practice this equation is calculated using the FFT.

**Definition 7.** The customary estimator of the sdf is the direct spectral estimator:

$$\hat{S}_D(f) = \left| \Delta t \sum_{t=1}^N h_t x_t e^{-i2\pi f t \Delta t} \right|^2. \quad (8)$$

In the above estimator  $\Delta t$  is the change in time step,  $t$ , and  $h_t$  is a data taper. If we allow  $h_t = \sqrt{1/n}$  the direct spectral estimator becomes so-called *periodogram*, which we denote as  $\hat{S}(f)$ . The raw periodogram asymptotically unbiased, but the bias can exist even with large sample sizes in practical applications (Thomson 1982, p. 1058). Additionally, the periodogram is an *inconsistent* statistical estimator, that is the variance does not decrease as the sample size increases (Chatfield 2004, p. 129). It can be shown that for real data this estimator has a  $\chi_2^2$  distribution when  $h_t = \sqrt{1/N}$  for all frequencies except  $f = 0$  and  $f = 1/2$  which have contain only real values and thus have a  $\chi_1^2$  distribution (P&W 1993, p. 222).

*Remark 8.* The periodogram and the biased estimator of the autocovariance sequence, equation (5), are Fourier transform pairs,

$$\{\hat{\gamma}_\tau^B\} \longleftrightarrow \{\hat{S}(f)\}$$

**Definition 8.** We will use a set of orthonormal Discrete Prolate Spheroidal Sequences (DPSS), also known as *Slepian sequences*, as data tapers. These sequences are defined as solutions to the system of equations (Slepian 1978):

$$\sum_{t'=0}^{N-1} \frac{\sin[2\pi W(t-t')]}{\pi(t-t')} v_{t',k}(N,W) = \lambda_k(N,W) v_{t,k}(N,W) \quad (9)$$

for  $t, t' = 0, 1, \dots, N-1$ . These sequences are discrete time analogs of real functions that are optimally concentrated in time and frequency (P&W 1993, pp. 75–81). The parameter  $W$  represents the effective bandwidth, which is often included in the time–bandwidth parameter  $NW$ , and  $k$  represents the current taper. Typically there are  $k = 0, 1, \dots, K-1$  tapers where  $K = 2NW$ .

**Definition 9.** We use a set of orthonormal Slepian sequences in constructing multitaper spectral estimate. If we let  $\hat{S}_k(f)$  represent the direct spectral estimator in equation (8) formed using the Slepian sequence of order  $k$ , then the simplest form if the multitaper spectral estimate becomes

$$\hat{S}^{(MT)} \equiv \frac{1}{K} \sum_{k=0}^{K-1} \hat{S}_k(f). \quad (10)$$

The individual  $\hat{S}_k(f)$  estimates are often referred to as *eigenspectra*, and the averaged estimator, equation (10), is distributed as  $\chi_{2K}^2$  for  $f \neq 0$  and  $f \neq 1/2$ .

*Remark 9.* When using Slepian sequences, the bandwidth parameter  $W$  is specified by setting the value time-bandwidth value  $NW$ . Typically one sets a bandwidth parameter between 2 and 6 (P&W 1993, p. 335) and noninteger values can be used (Thomson 1982, p. 1086). Judicious selection of the bandwidth parameter can allow for the resolution a lower power harmonic that would otherwise be masked by an adjacent higher power harmonic.

*Remark 10.* In practice we will use the adaptive weighted multitaper spectral estimate,  $\hat{S}^{(AMT)}(f)$ , which uses a sophisticated weighted averaging scheme. This weighting scheme generally down-weights higher order eigenspectra which have a higher bias. See Thomson (1982, pp. 1065–1066) for more details. This weighted averaging scheme provides a non-integer degree-of-freedom estimate at each frequency which is typically slightly below  $2K$ .

## 2.2 Yule-Walker Equations

The Yule-Walker equations are part of the oldest method for estimating the parameters of a zero-mean stationary AR( $p$ ) process  $\{Y_t\}$ . The method involves the following steps:

- (a) Assume the process is stationary. [This step may seem a bit circular.]
- (b) Multiply equation (1) by  $X_{t-k}$  for  $k = 1, 2, \dots, p$ ,
- (c) Take expected values,

$$\gamma_k = \sum_{j=1}^p \phi_j \gamma_{k-j} \quad \text{for all } k > 0. \quad (11)$$

Using the fact that  $Y_{t-k}$  is uncorrelated with noise that occurs after time  $t-k$ , we see that  $E\{Z_t Y_{t-k}\} = 0$ .



### 2.3.2 One-Step Ahead Prediction

**Definition 11.** We write the one-step-ahead AR(p) best linear predictor as

$$\vec{X}_{N+1}(p) = \phi_1 X_N + \phi_2 X_{N-1} + \cdots + \phi_p X_{N-(p-1)}. \quad (19)$$

In matrix notation this can be written as

$$\vec{X}_{N+1}(p) = \boldsymbol{\phi}_p^T \mathbf{X}_p,$$

where  $\mathbf{X}_p = \mathcal{W}_{N-(p-1),N} \mathbf{X}$ , that is  $\mathbf{X}_p$  vector of the last  $p$  elements from  $\mathbf{X} = (X_1, X_2, \dots, X_N)^T$ .

**Definition 12.** The mean-squared one-step-ahead prediction error is given by

$$\begin{aligned} P_{N+1} &= E\{(\vec{X}_{N+1}(p) - X_{N+1})^2\} \\ &= \gamma(0) - \boldsymbol{\gamma}_p^T \Gamma_p^{-1} \boldsymbol{\gamma}_p. \end{aligned} \quad (20)$$

See Shumway and Stoffer (2006, p. 112) for details.

### 2.3.3 Levinson-Durbin Algorithm

The recursions begin by setting

$$\phi_{0,0} = 0, \quad \text{and} \quad P_1 = \gamma(0), \quad (21)$$

then for  $n \geq 1$ , the partial autocorrelation coefficients are updated by

$$\phi_{n,n} = \frac{\rho_n - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho_{n-k}}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho_k}, \quad (22)$$

and the mean-squared one-step-ahead prediction error is from

$$P_{n+1} = P_n(1 - \phi_{n,n}^2). \quad (23)$$

The update of the mean-squared one-step-ahead prediction error in equation (23) is evidently suggested by Burg in 1961. See Burg (1975, p. 14). The autocorrelation coefficients are obtained when  $n \geq 2$  using

$$\phi_{n,k} = \phi_{n-1,k} - \phi_{n,n} \phi_{n-1,n-k}, \quad \text{for } k = 1, 2, \dots, n-1. \quad (24)$$

### 2.3.4 Using Tapered Spectra to Estimate the ACVS

It has been noted that there is no reason to restrict oneself to the acvs computed from untapered spectral estimates (P&W 1993, pp. 396–397), one can use direct spectral estimators and multitaper estimates. It has been shown (P&W 1993, pp. 405–406) that a direct spectral estimator using Slepian sequences with  $NW = 2$  accurately depict the theoretical spectra of a known AR(4) process. We present simulations comparing different estimates for a known AR(4) sequence.

## 2.4 Burg’s Method

### 2.4.1 Overview

Burg’s algorithm is also a solution the Yule–Walker equations; however, it focuses on primarily estimating the partial autocorrelation coefficients without using an estimate of  $\hat{\gamma}_\tau$ . It does this by focusing on minimizing the error in the one–step–ahead and one–step–backwards prediction estimates. In practice, the Burg algorithm has been shown to be more efficient than use of the Levinson–Durbin recursions using the standard biased estimator,  $\hat{\gamma}_\tau^B$ , for smaller sample size (P&W 1993, p. 414). Our simulations indicate the Burg algorithm is considerably more effective than the Levinson–Durbin recursions when using the standard biased estimator,  $\hat{\gamma}_\tau^B$ , but it is not significantly more effective, when a multitaper spectral estimate version of  $\hat{\gamma}_\tau$  is used. Additionally the Burg estimator is not without its own drawbacks (P&W 1993, pp. 525–531).

### 2.4.2 Preliminaries

**Definition 13.** We write the prediction error associated with the one–step–ahead AR(p) predictor, equation (19), as

$$\vec{\epsilon}_t(p) = X_t - \vec{X}_t(p). \tag{25}$$

**Definition 14.** As in definition 11, we write the one–step–back AR(P) best linear predictor as

$$\overleftarrow{X}_t(p) = \phi_1 X_{t+1} + \phi_2 X_{t+2} + \dots + \phi_p X_{t+p}. \tag{26}$$

**Definition 15.** We write the prediction error associated with the one–step–back AR(p) predictor, equation (26), as

$$\overleftarrow{\epsilon}_t(p) = X_t - \overleftarrow{X}_t(p). \tag{27}$$

**Definition 16.** Define  $\mathcal{L}$  as a circular shift operator. If  $\mathbf{v} = (v_1, v_1, \dots, v_N)^T$ , then

$$\mathcal{L} \mathbf{v} = (v_N, v_1, v_1, \dots, v_{N-1})^T$$

**Definition 17.** We plan to fit an AR(p) model to  $\mathbf{X} = (X_1, X_2, \dots, X_N)^T$  and we define the following vector of length  $N + p$

$$\vec{\mathbf{e}}(0) = (X_1, X_2, \dots, X_N, 0, 0, \dots, 0)^T.$$

The vector is  $\mathbf{X}$  concatenated with  $p$  zeros.

**Definition 18.** We also define

$$\overleftarrow{\mathbf{e}}(0) = \mathcal{L} \vec{\mathbf{e}}(0) = (0, X_1, X_2, \dots, X_N, 0, \dots, 0)^T.$$

### 2.4.3 Burg’s Procedure

We define the variance  $\tilde{\sigma}_0^2 = \hat{\gamma}_0^B$ , then for  $k = 1, 2, \dots, p$  we recursively compute the following:

$$\tilde{\phi}_{k,k} = \frac{2 \langle \mathcal{M}_{k+1,N} \vec{\mathbf{e}}(k-1), \mathcal{M}_{k+1,N} \overleftarrow{\mathbf{e}}(k-1) \rangle}{\|\mathcal{M}_{k+1,N} \vec{\mathbf{e}}(k-1)\|^2 + \|\mathcal{M}_{k+1,N} \overleftarrow{\mathbf{e}}(k-1)\|^2} \tag{28}$$

$$\begin{aligned}\tilde{\sigma}_k^2 &= \tilde{\sigma}_{k-1}^2(1 - \tilde{\phi}_{k,k}^2) \\ \vec{\mathbf{e}}(k) &= \vec{\mathbf{e}}(k-1) - \tilde{\phi}_{k,k} \overleftarrow{\mathbf{e}}(k-1) \\ \overleftarrow{\mathbf{e}}(k) &= \mathcal{L}(\vec{\mathbf{e}}(k-1) - \tilde{\phi}_{k,k} \overrightarrow{\mathbf{e}}(k-1)).\end{aligned}$$

Where we use  $\langle \cdot, \cdot \rangle$  to denote vector inner product, and  $\|\cdot\|^2$  to denote the squared norm.

The Burg estimator  $\tilde{\phi}_{k,k}$  differs from the Yule–Walker estimator  $\hat{\phi}_{k,k}$ , but the Burg procedure can be modified to obtain the Yule–Walker estimate. The key point of the Burg estimator is that an estimator of autocovariance sequence, typically  $\hat{\gamma}_\tau^B$  for  $\tau > 0$ , is no longer required, where as for the Yule–Walker equations, an estimator of the autocovariance sequence is required for integer values of  $\tau \leq p$ .

### 3. Cautionary Notes on using AR Spectral Estimates

We generally consider AR models useful for prewhitening data, but we caution against its use in general spectral estimation in the physical sciences. Kaveh and Lippert (1983) believe AR spectral estimation can be patched for use, but Tukey (1984) cautioned against the general use of parametric spectral estimation. For a general overview of the problems of parametric spectral estimation see (P&W 1993, pp. 525–531).

Two specific problems of note when using AR spectral estimates for a sinusoid in additive noise are: (a) the location of the peak in the spectrum is found to depend on the phase of the sinusoid, and (b) two adjacent peaks in the spectrum can appear as one peak (P&W 1993, p. 525). The second problem is known in the literature as *spectral line splitting*.

Two proposed solutions are: (a) replacing the real-valued with an analytic signal (Kay and Marple 1981, p. 1396), and (b) using improved estimates of the autocorrelation function, equation (1). The first solution must consider taking an appropriate Hilbert transform that does not have the same bias properties as estimates based on the raw [biased] periodogram and the process becomes complicated in the presence of multiple lines. We take the latter approach.

### 4. Comparison of Methods for Finding AR Coefficients

We compare selected AR coefficient estimation techniques on simulated data from a high signal-to-noise ratio AR(4) process which has been used in the literature,  $\boldsymbol{\phi} = (2.7607, -3.8106, 2.6535, -0.9238)^T$  (Bishop and Ulrych 1975; Box et al. 1994). Table 1 compares the Mean-Squared-Error (MSE), mean, median, and sample standard deviation, from estimates of the partial autocorrelation coefficient  $\phi_{4,4} = -0.9238$  using the Levinson–Durbin recursions with the biased autocovariance estimator, to an autocovariance estimator based on single Slepian taper  $NW = 5$ , and to an autocovariance estimator constructed using the adaptive multitaper method with  $NW = 5$ , and  $k = 5$ . The acvs estimators were calculated from the estimated spectrum using the property in remark 8. Figure 1 indicates a comparison partial autocorrelation coefficients. In this simulation, the multitaper spectral estimate and the Burg estimate are preferred, and the use of a single Slepian taper is preferred to the standard biased acvs estimator.

### 5. Goodness-of-fit Test for Autoregressive Processes

The approach for testing the goodness-of-fit of an AR process is based on comparing the observed standardized integrated spectrum to the theoretical standardized integrated spectrum of the selected autoregressive model (Anderson 1997).

	Default	Single Taper	Multitaper	Burg
MSE	0.30903	0.00062	0.00017	0.00017
Mean	-0.4252	-0.9135	-0.9204	-0.9204
Median	-0.4114	-0.9159	-0.9214	-0.9213
Sample SD	0.2457	0.0227	0.0128	0.125

**Table 1:** Comparisons of estimates of  $\phi_{4,4}$  from 100,000 run simulation using the Yule-Walker equations with the biased autocovariance estimator, an autocovariance estimator using one Slepian taper with  $NW = 5$ , and an adaptive weighted multitaper spectral estimate with  $NW = 5$ , and  $k = 8$ , and the partial autocovariance estimator made using Burg's method.

## 5.1 Preliminaries

**Definition 19.** The empirical distribution,  $\hat{F}(x)$  for a random sample of observations of  $X$  is generally  $\hat{F}(x) =$  the proportion of samples observations  $\leq x$ .

**Definition 20.** The integrated spectra  $H(f_0) = \int_{-1/2}^{f_0} S(f) df$ , can be well estimated by

$$\hat{H}(f_0) = \int_{-1/2}^{f_0} \hat{S}_{(D)}(f) df. \quad (29)$$

Note that tests based on the integrated spectrum, standardized or not, are generally not considered to be affected by the bias properties of using the raw periodogram (Priestley 1981, p. 471). In the case of real-valued data, equation (29) can be adjusted to only consider positive frequencies (see: (Priestley 1981, p. 474)).

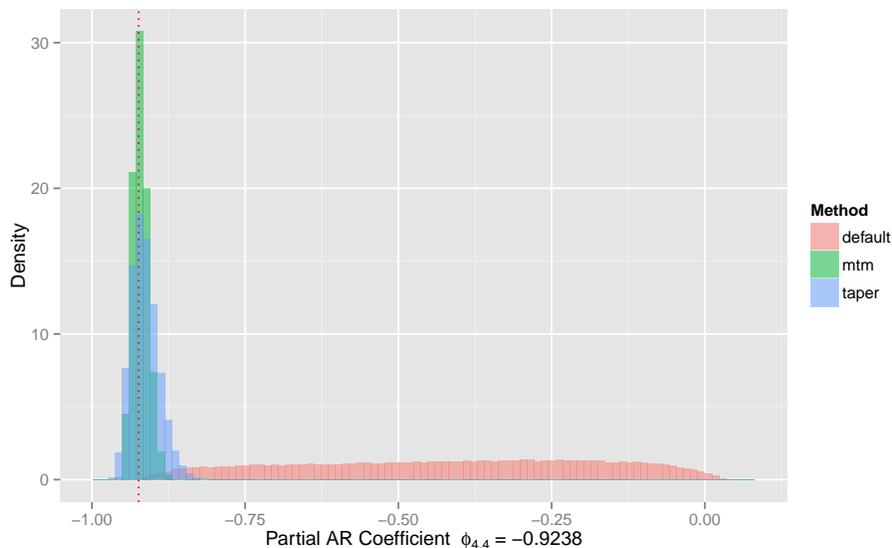
**Definition 21.** The standardized integrated spectrum can be written as

$$F(f) = \frac{\int_{-1/2}^{f_0} S(f) df}{\int_{-1/2}^{1/2} S(f) df}, \quad (30)$$

which we estimate using the standard spectral estimator in equation 8. The goodness-of-fit tests draw on the correspondence between the standardized integrated spectrum and the empirical distribution function, and standardization provides that advantage that asymptotic distributions are valid under more general conditions than those without standardization (Anderson 1997). As with the integrated spectrum, definition 29, this estimator can be constructed from only positive frequencies when restricted to real-valued data.

## 5.2 Goodness-of-fit Tests for AR Processes

An overview of goodness-of-fit tests for AR and Moving Average (MA) models are presented in Priestley (1981, pp. 475–494). We will be using the maximum absolute deviation of the integrated spectrum as a measure of goodness-of-fit, and we will simulations to estimate p-values for the observed maximum absolute deviation. We note Anderson (1997) proposes the same test statistic, the maximum absolute deviation of the integrated spectrum, to test the null hypothesis that the observations are on an AR process of an order not greater than the specified one. In place of asymptotic results linking the Cramér-von Mises, or Kolmogorov-Smirnov statistic, we propose the practical measure of relying on simulations to generate approximate p-values.



**Figure 1:** Estimated fourth reflection coefficient based on a hundred thousand run simulation of an AR(4) process with coefficients 2.7607, -3.8106, 2.6535, -0.9238. Levinson–Durbin estimate using: (a) the default estimate, (b) one DPSS taper with  $NW = 5$ , and (c) an adaptive multitaper estimate with  $k = 8$ . The dashed line indicates -0.9238. Mean estimates were -0.425, -0.914, and -0.920 respectively. The distribution of Burg estimator is very similar to the multitaper spectral estimator and is not shown.

Bartlett related the asymptotic distribution of the mean absolute deviation between the estimated normalized spectrum and the theoretical spectrum,

$$\max_{0 \leq f \leq 1/2} \sqrt{N} |\hat{F}_+(f) - F_+(f)|, \tag{31}$$

to the Kolmogorov–Smirnov statistic which has been used in testing the goodness–of–fit in empirical distributions (Priestley 1981, p. 480). We use the subscript positive sign + to indicate we are constructing the estimate solely on positive frequencies, (see definition 21).

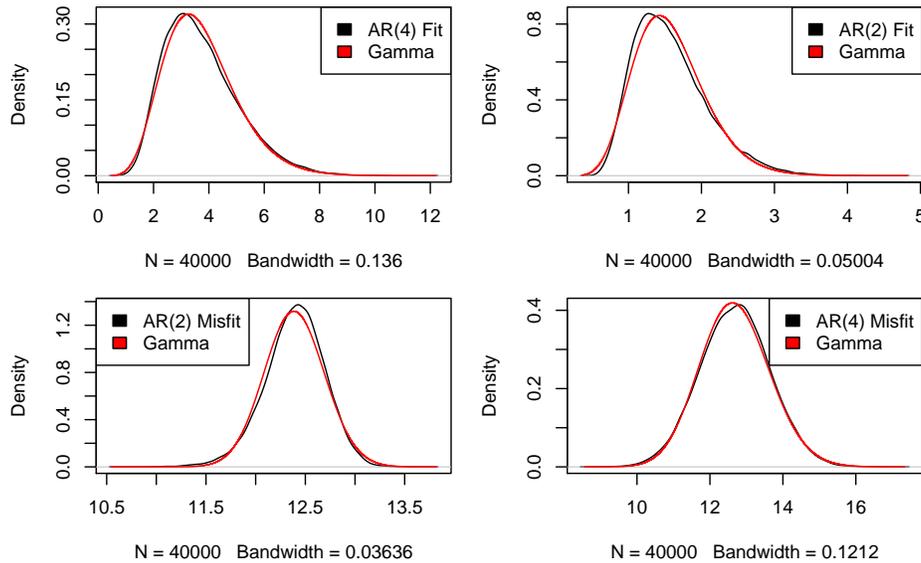
### 5.3 Proposed Methodology

Limiting distributions for the goodness–of–fit tests have been studied Anderson (1997); but practical software solutions are not readily available, and we propose a simple simulation based statistical test. Additionally simulations do not constrain us to a one–size–fits all approach. We propose (a) careful fitting of AR coefficients, (b) plotting the estimated spectra against the theoretical spectra, see equation (7), for the selected AR model, and (c) comparing the estimated standardized integrated spectrum to the theoretical spectra for the AR using the maximum absolute deviation as a test. We then use simulations to assess the significance of the observed distance. In constructing the theoretical AR spectrum used in the standardized integrated spectrum, we estimate  $\sigma_Z^2$  in equation (7), from the data.

## 6. Simulations of Goodness–of–fit

We assess the goodness of fit for AR models for two AR models used in the literature, the AR(4) model discussed in figure 1, and the AR(2) model  $\phi = (0.75, -0.5)^T$  (P&W 1993, p. 45). Figure 2 compares empirical distributions of the distance, showing four comparisons

in our simulations, two cases where the simulated AR process matches the theoretical, and two cases where we simulate mismatches, that is the AR process simulated does not match the theoretical. The top two plots show the distributions of the distances where the models accurately fit, and the lower two show distributions of misfit models. Comparing the top two plots in figure 2 to the bottom two, one can see considerable  $x$ -axis values change. The misfit models generate larger distances. The red line indicates fitted Gamma distributions and table 2 indicates the shape and rate parameters of the fitted Gamma distributions.

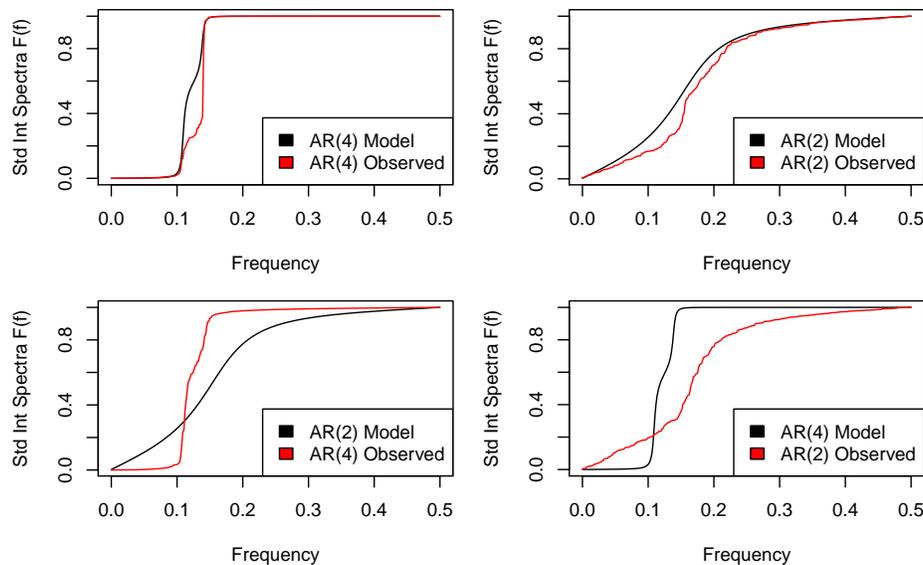


**Figure 2:** This figure shows the observed maximum absolute distance observed from 40000 simulations. The top left is from comparing a simulated AR(4) to the theoretical AR(4), the top right is from comparing a simulated AR(2) to the theoretical AR(2), the bottom left is from comparing a simulated AR(4) to the theoretical AR(2), and the bottom right is from comparing a simulated AR(4) to a theoretical AR(2)

Models	Shape	SD (Shape)	Rate	SD (Rate)
AR(4)	7.9970	0.0554	2.1444	0.0153
AR(2)	10.2543	0.07136	6.5002	0.0464
Misfit AR(2)	1676.2664	11.8357	135.2800	0.9553
Misfit AR(4)	176.7686	1.2486	13.9327	0.0986

**Table 2:** Shape and rate parameters for the fitted Gamma distributions shown in figure 2. Both the shape and rate parameters are considerably higher for the case where the simulated AR model did not match the theoretical model.

In order to get a sense of how the simulated integrated spectra compare to the theoretical, the top left plot in figure 3 show the observed integrated spectrum from that AR(4) simulation run that had the extreme (largest) value for maximum absolute deviation of the 40000 simulated against the theoretical integrated spectrum for the AR(4) process. The top right plots the AR(2) simulation run that had the extreme (largest) value for maximum absolute deviation of the 40000 simulated against the theoretical integrated spectrum for the AR(2) process. These two plots demonstrate extreme (largest) differences when the AR process is appropriate. The bottom two plots indicate the extreme (smallest) observed difference between simulation and theoretical when the simulated process did not match



**Figure 3:** We ran 40000 simulations each comparing a simulated AR(4) to the theoretical AR(4), top left, a simulated AR(2) to the theoretical AR(2), top right, a simulated AR(2) to the theoretical AR(2), bottom left, and a simulated AR(4) to the theoretical AR(2), bottom right. The top two plots indicate the *worst fit* of the 40000 runs when the simulations were from the same model as the theoretical, and the bottom two plots indicate the *best fit* of the 40000 runs when the simulations are from a different model than the theoretical.

the theoretical. Comparing the bottom left plot and the bottom right plot one wonders if a misfit to a AR(4) process is easier to detect than the misfit to the AR(2) process.

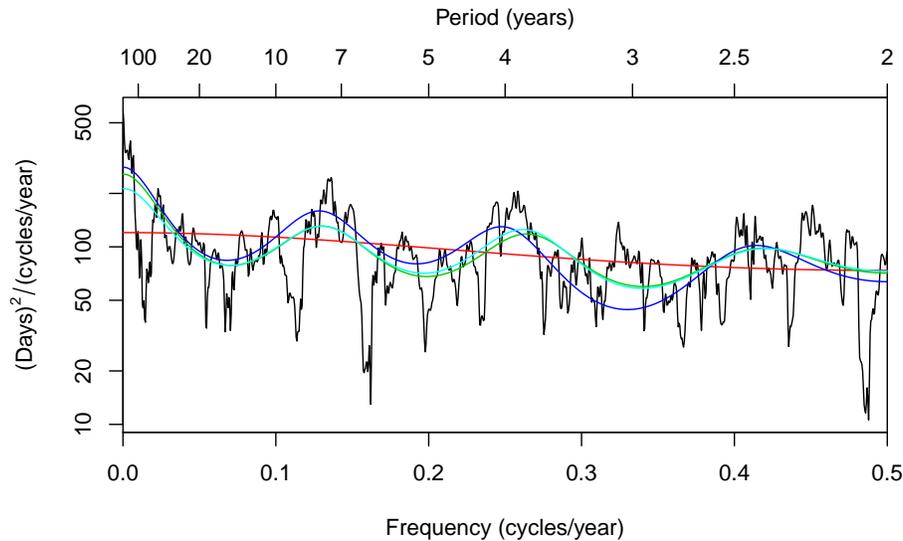
## 7. Burgundy Grape Harvest Dates

A robust linkage between Burgundy GHD and European climate fluctuations has been proposed Tourre et al. (2011), and Pinot Noir is considered to be highly sensitivity to climate variations. As we mentioned, in the physical sciences it is customary to use AR models to prewhiten the data (Mann and Lees 1996), before the residuals from the AR model are used in harmonic analysis. Figure 4 presents the raw spectrum of the GHD series, and the associated spectra several AR models, including models of the same order where different techniques were used to obtain the AR coefficients. It does appear from plot that selection of AR model prewhitener can affect harmonic analysis of residuals. We proceed to fit several AR models to the Burgundy GHD series and test them for goodness-of-fit.

Using our method for comparing AR goodness-of-fit, table 3 shows the observed maximum absolute deviation of the sample integrated spectrum from the theoretical integrated spectrum of the model, and simulated  $p$ -values testing the hypothesis that the maximum absolute deviation is greater than would be expected if the data matched the theoretical AR model. While looking at the spectrum in figure 4 it appears the choice of prewhitner can affect the significance of the harmonic components; however, this test does not enable us to distinguish between models.

## 8. Conclusions and Future Work

This paper demonstrated using simulation that various methods of calculating autoregressive coefficients affect the values of the estimated coefficients, and that multitaper spectral



**Figure 4:** Adaptive multitaper spectrogram of the Harvest Date series. The parameters used are:  $NW = 3$ ,  $k = 5$ . Plotted over the spectrum, we have the standard AR(1) in red, the standard AR(8) in green, the DPSS tapered AR(8) in blue, and the multitaper AR(8) spectrum in cyan.

AR( $p$ ) Model	Max Abs Dist	Simulated P-value
AR(1) no taper	0.9335	0.2634
AR(8) no taper	0.9291	0.2073
AR(8) 1 DPSS	0.9286	0.1839
Misfit AR(4)	0.9305	0.2256

**Table 3:** Maximum absolute deviation of the observed GHD standardized integrated spectrum to the theoretical standardized integrated spectrum for the various models and approximate p-values based on simulations testing the null hypothesis that the maximum absolute deviation small enough for the model to be appropriate. Based on this goodness-of-fit criterion, we see little difference in the choice of models, and certainly no significant difference. We conclude each of the four models fit reasonably well.

estimation used with Levinson–Durbin recursions are as effective as Burg’s recursions in obtaining AR coefficients for a high signal-to-noise ratio AR(4) process. We proposed a practical method of testing the goodness-of-fit of AR estimators using the maximum absolute deviation between the standardized integrated spectrum from the estimated data and the theoretical standardized integrated spectra from the theoretical AR model, and we tested this method on two AR processes with simulations. We selected different AR models for the Burgundy GHD data set and used our goodness-of-fit tests to see if any are not appropriate. We concluded that the four selected models fit reasonably well to the GHD series.

There are several areas for future work: (a) test, with simulations, different but closely related simulated AR models in order to determine how the test works, and test simulations with mixed spectra which include discrete line components, (b) consider other ways of comparing two spectra, for example the  $L^2$  distance would be more sensitive to overall differences, whereas maximum absolute deviation may be more sensitive to high power

line components, (c) questions of stationarity and change point exist in climate series such as the Burgundy GHD series, thus one could section the series at a change point and make multiple comparisons of spectra before and after the change point to each other, and to AR model for the entire series.

## References

- Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, Wiley New York.
- (1997), “Goodness-of-fit Tests for Autoregressive Processes,” *Journal of Time Series Analysis*, 18, 321–339.
- Bishop, T. and Ulrych, T. (1975), “Maximum Entropy Spectral Analysis and Autoregressive Decomposition,” *Reviews of Geophysics and Space Physics*.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control*, Wiley, 3rd ed.
- Burg, J. P. (1975), “Maximum Entropy Spectral Analysis,” Ph.D. thesis, Stanford University.
- Chatfield, C. (2004), *The Analysis of Time Series: an Introduction*, CRC Press.
- Chuine, I., Yiou, P., Viovy, N., Seguin, B., Daux, V., and Ladurie, E. L. R. (2004), “Historical Phenology: Grape Ripening as a Past Climate Indicator,” *Nature*, 432, 289–290.
- Kaveh, M. and Lippert, G. A. (1983), “An Optimum Tapered Burg Algorithm for Linear Prediction and Spectral Analysis,” *ASSP-31*, 438–444.
- Kay, S. M. and Marple, Jr, S. L. (1981), “Spectrum Analysis—A Modern Perspective,” *Proceedings of the IEEE*, 69, 1380–1419.
- Mann, M. E. and Lees, J. M. (1996), “Robust Estimation of Background Noise and Signal Detection in Climatic Time Series,” *Climatic Change*, 33, 409–445.
- Percival, D. B. and Walden, A. T. (1993), *Spectral Analysis for Physical Applications*, Cambridge University Press New York, NY, USA.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series. Volume 1: Univariate Series. Volume 2: Multivariate Series, Prediction and Control*, Probability and Mathematical Statistics.
- Shumway, R. H. and Stoffer, D. S. (2006), *Time Series Analysis and its Applications: With R Examples*, Springer, 2nd ed.
- Slepian, D. (1978), “Prolate Spheroidal Wave Functions, Fourier Analysis, and Uncertainty. V—the Discrete Case,” *Bell System Tech. J.*, 57, 1371–1430.
- Thomson, D. J. (1977), “Spectrum Estimation Techniques for Characterization and Development of WT4 Waveguide,” *Bell Syst. Tech. J.*, 56, 1769–1815.
- (1982), “Spectrum Estimation and Harmonic Analysis,” *Proceedings of the IEEE*, 70, 1055–1096.
- Tourre, Y. M., Rousseau, D., Jarlan, L., Le Roy Ladurie, E., and Daux, V. (2011), “Western European Climate, and Pinot Noir Grape Harvest Dates in Burgundy, France, since the 17th Century,” *Climate Research*, 46, 243.
- Tukey, J. W. (1984), “Styles of Spectrum Analysis,” in *A Celebration in Geophysics and Oceanography – 1982 In Honor of Walter Munk*, La Jolla, CA: Scripps Institution of Oceanography, Reference Series 84-5, March, 1984, pp. 100–103, pp 1143-1153 of **The Collected Works of J. W. Tukey** Vol II, D. R. Brillinger, Ed., Wadsworth, Monterey, Ca., 1984.