

5 Autoregressive-Moving-Average Modeling

5.1 Purpose.

Autoregressive-moving-average (ARMA) models are mathematical models of the persistence, or autocorrelation, in a time series. ARMA models are widely used in hydrology, dendrochronology, and many other fields. There are several possible reasons for fitting ARMA models to data. Modeling can contribute to *understanding the physical system* by revealing something about the physical process that builds persistence into the series. For example, a simple physical water-balance model consisting of terms for precipitation input, evaporation, infiltration, and groundwater storage can be shown to yield a streamflow series that follows a particular form of ARMA model. ARMA models can also be used to *predict* behavior of a time series from past values alone. Such a prediction can be used as a baseline to evaluate possible importance of other variables to the system. ARMA models are widely used for prediction of economic and industrial time series. ARMA models can also be used to *remove* persistence. In dendrochronology, for example, ARMA modeling is applied routinely to generate *residual* chronologies – time series of ring-width index with no dependence on past values. This operation, called *prewhitening*, is meant to remove biologically-related persistence from the series so that the residual may be more suitable for studying the influence of climate and other outside environmental factors on tree growth.

5.2 Mathematical Model

ARMA models can be described by a series of equations. The equations are somewhat simpler if the time series is first reduced to zero-mean by subtracting the sample mean. Therefore, we will work with the *mean-adjusted* series

$$y_t = Y_t - \bar{Y}, \quad t = 1, \dots, N \quad (1)$$

where Y_t is the original time series, \bar{Y} is its sample mean, and y_t is the mean-adjusted series. One subset of ARMA models are the so-called *autoregressive*, or AR models. An AR model expresses a time series as a linear function of its past values. The *order* of the AR model tells how many lagged past values are included. The simplest AR model is the first-order autoregressive, or *AR(1)*, model

$$y_t + a_1 y_{t-1} = e_t \quad (2)$$

where y_t is the mean-adjusted series in year t , y_{t-1} is the series in the previous year, a_1 is the lag-1 autoregressive coefficient¹, and e_t is the *noise*. The noise also goes by various other names: the *error*, the *random-shock*, and the *residual*. The residuals e_t are assumed to be random in time (not autocorrelated), and normally distributed. Be rewriting the equation for the AR(1) model as

$$y_t = -a_1 y_{t-1} + e_t \quad (3)$$

we can see that the AR(1) model has the form of a regression model in which y_t is regressed on its previous value. In this form, a_1 is analogous to the regression coefficient, and e_t to the regression residuals. The name *autoregressive* refers to the regression on self (auto).

¹ Some authors (e.g., Chatfield, 2004) write the equation for an AR(1) process in the form $y_t = a_1 y_{t-1} + e_t$, which implies a positive coefficient a_1 for positive first-order autocorrelation. But as written in (2), positive autocorrelation goes with a negative coefficient a_1 . There is no confusion as long as the equation being used for describing the process is presented along with values of parameters. The convention used in this chapter follows Ljung (1995) and MATLAB®'s System Identification Toolbox.

Higher-order autoregressive models include more lagged y_t terms as predictors. For example, the second-order autoregressive model, AR(2), is given by

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} = e_t \quad (4)$$

where a_1, a_2 are the autoregressive coefficients on lags 1 and 2. The p^{th} order autoregressive model, AR(p) includes lagged terms on years $t-1$ to $t-p$.

The *moving average* (MA) model is a form of ARMA model in which the time series is regarded as a moving average (unevenly weighted) of a random shock series e_t . The *first-order moving average*, or MA(1), model is given by

$$y_t = e_t + c_1 e_{t-1} \quad (5)$$

where e_t, e_{t-1} are the residuals at times t and $t-1$, and c_1 is the first-order moving average coefficient. As with the AR models, higher-order MA models include higher lagged terms. For example, the second order moving average model, MA(2), is

$$y_t = e_t + c_1 e_{t-1} + c_2 e_{t-2} \quad (6)$$

The letter q is used for the order of the moving average model. The second-order moving average model is MA(q) with $q=2$.

We have seen that the autoregressive model includes lagged terms on the time series itself, and that the moving average model includes lagged terms on the noise or residuals. By including both types of lagged terms, we arrive at what are called *autoregressive-moving-average*, or ARMA, models. The order of the ARMA model is included in parentheses as ARMA(p,q), where p is the autoregressive order and q the moving-average order. The simplest, and most frequently used ARMA model is ARMA(1,1) model

$$y_t + a_1 y_{t-1} = e_t + c_1 e_{t-1} \quad (7)$$

5.3 Steps in modeling

ARMA modeling proceeds by a series of well-defined steps. The first step is to *identify* the model. Identification consists of specifying the appropriate structure (AR, MA or ARMA) and order of model. Identification is sometimes done by looking at plots of the acf and partial autocorrelation function (pacf). Sometimes identification is done by an automated iterative procedure -- fitting many different possible model structures and orders and using a goodness-of-fit statistic to select the best model.

The second step is to *estimate* the coefficients of the model. Coefficients of AR models can be estimated by least-squares regression. Estimation of parameters of MA and ARMA models usually requires a more complicated iteration procedure (Chatfield 2004). In practice, estimation is fairly transparent to the user, as it accomplished automatically by a computer program with little or no user interaction.

The third step is to *check* the model. This step is also called *diagnostic checking*, or *verification* (Anderson 1976). Two important elements of checking are to ensure that the residuals of the model are random, and to ensure that the estimated parameters are statistically significant. Usually the fitting process is guided by the principal of *parsimony*, by which the best model is the simplest possible model -- the model with the fewest parameters -- that adequately describes the data.

Identification by visual inspection of acf and pacf. The classical method of model identification as described by Box and Jenkins (1970) is judge the appropriate model structure and order from the appearance of the plotted acf and partial autocorrelation functions. We have already discussed the acf, and know that the acf at lag k measures the correlation of the series with itself lagged k years. The *partial autocorrelation function* (pacf) at lag k is the

autocorrelation at lag k after first removing autocorrelation with a $AR(k-1)$ model. If the $AR(k-1)$ model effectively whitens the time series, the pacf at lag k will be zero. The identification of ARMA models from the acf and pacf plots is difficult and requires much experience for all but the simplest models. Let's look at the diagnostic patterns for the two simplest models: $AR(1)$ and $MA(1)$.

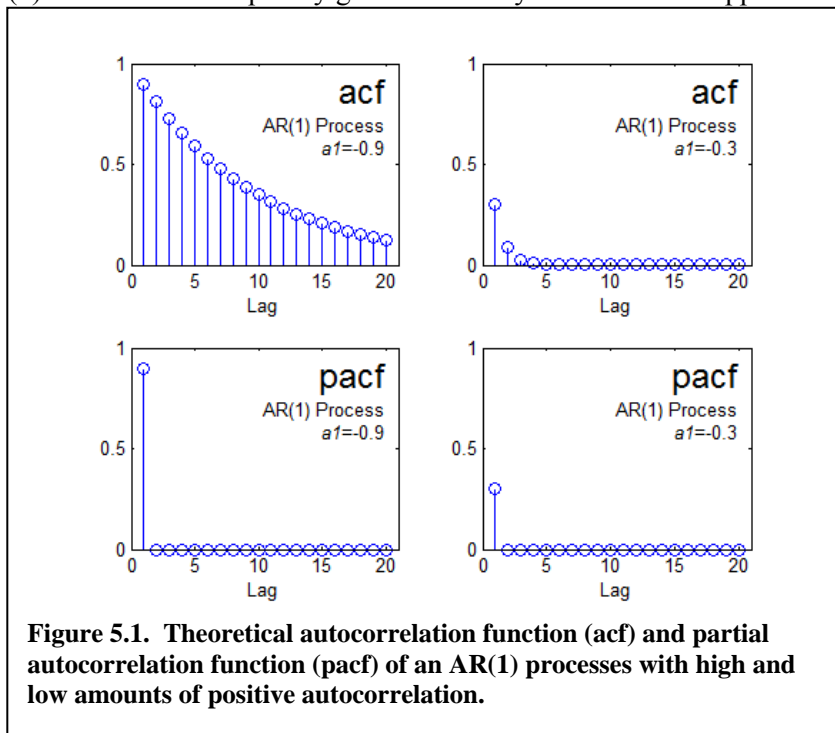
The acf of an $AR(1)$ model declines geometrically as function of lag. For example, the acf of series that follows an $AR(1)$ model with coefficient $a_1 = 0.5$ is $\{0.5, 0.5^2, 0.5^3, 0.5^4\}$ at lags 1-4. The pacf of the $AR(1)$ process at lags $k > 1$ is zero, because if the model is $AR(1)$, all autocorrelation is removed by the $AR(1)$ model. In summary, the diagnostic patterns of acf and pacf for an $AR(1)$ model are:

- Acf: declines in geometric progression from its highest value at lag 1
- Pacf: cuts off abruptly after lag 1

The opposite types of patterns apply to an $MA(1)$ process:

- Acf: cuts off abruptly after lag 1
- Pacf: declines in geometric progression from its highest value at lag 1

Theoretical acf and pacf for two $AR(1)$ processes with large and small autoregressive coefficients are shown in Figure 5.1. The acf and pacf in both cases follows the diagnostic patterns described above. The persistence is specified by the size of the coefficient. For an $AR(1)$ model, the square of the autoregressive coefficient is analogous to R^2 in regression. Accordingly, for a model with $a_1 = -0.9$ persistence explains 81% of the variance, and for a model with $a_1 = -0.3$ persistence explains 9% of the variance. Note that the lower the coefficient for the $AR(1)$ model the more quickly geometric decay leads to an acf approaching zero.



The theoretical patterns of decay of the acf and pacf can be visually compared with the estimated acf and pacf of a time series to decide whether the series was likely generated by an AR(1) process. The theoretical decay patterns can be difficult to discern for actual time series, especially if the series is short. Higher-order AR, MA, and ARMA processes also have characteristic theoretical patterns of decay of the acf and pacf. The decay patterns are more complicated than those illustrated for the AR(1) model. Generally, however, the pacf cuts off after lag p for an AR(p) process, and the acf cuts off abruptly after lag q for an MA(q) process.

Short memory and long memory processes. Though the acf may die off very slowly for an AR(1) process with a high AR coefficient (Figure 5.1), the AR(1) process, and indeed all stationary AR, MA and ARMA processes are termed “short memory” processes. Such processes satisfy a condition of eventual dying off of the acf. The condition is given by $|\rho_k| < C\lambda^k$, where ρ_k is the theoretical autocorrelation at lag k , and C and λ are constants with $0 < \lambda < 1$ (Chatfield 2004, p. 261). For such processes, $\sum_{k=0}^{\infty} |\rho_k|$ converges. On the other hand, there is another class of processes, called “long-memory” processes, for which $\sum_{k=0}^{\infty} |\rho_k|$ does not converge. See Chatfield (2004, p. 260) for discussion of long-memory processes.

Automated identification by the FPE criterion. A different way of identifying ARMA models is by trial and error and use of a goodness-of-fit statistic. In this approach, a suite of candidate models are fit, and goodness-of-fit statistics are computed that penalize appropriately for excessive complexity. (Think of adjusted R^2 in regression.) Akaike’s Final Prediction Error (FPE) and Information Theoretic Criterion (AIC) are two closely related alternative statistical measures of goodness-of-fit of an ARMA(p,q) model. Goodness of fit might be expected to be measured by some function of the variance of the model residuals: the fit improves as the residuals become smaller. Both the FPE and AIC are functions of the variance of residuals. Another factor that must be considered, however, is the number of estimated parameters $n = p + q$. This is so because by including enough parameters we can force a model to perfectly fit any data set. Measures of goodness of fit must therefore compensate for the artificial improvement in fit that comes from increasing complexity of model structure. The FPE is given by

$$FPE = \frac{1 + n/N}{1 - n/N} * V \quad (8)$$

where V is the variance of model residuals, N is the length of the time series, and $n = p + q$ is the number of estimated parameters in the ARMA model. The FPE is computed for various candidate models, and the model with the *lowest* FPE is selected as the best-fit model.

The AIC (Akaike Information Criterion) is another widely used goodness-of-fit measure, and is given by

$$AIC = \log \left[V \left(1 + \frac{2n}{N} \right) \right] \quad (9)$$

As with the FPE, the best-fit model has *minimum* value of AIC.² Neither the FPE nor the AIC directly addresses the question of the model residuals being without autocorrelation, as they

² The System Identification Toolbox in MATLAB© has functions for the FPE and the AIC. (Computational note: MATLAB© computes the variance V used in the above equations with $N-1$ rather than N in the denominator of the sum-of-squares term.)

ideally should be if the model has removed the persistence. A strategy for model identification by the FPE is to 1) iteratively fit several different models and find the model that gives approximately minimum FPE **and** 2) apply diagnostic checking (see below) to assure that the model does a good job of producing random residuals.

Checking the model – are the residuals random? A key question in ARMA modeling is does the model effectively describe the persistence? If so, the model residuals should be random –or uncorrelated in time – and the autocorrelation function (acf) of residuals should be zero at all lags except lag zero. Of course, for sample series, the acf will not be exactly zero, but should fluctuate close to zero.

The acf of the residuals can be examined in two ways. First, the acf can be scanned to see if any individual coefficients fall outside some specified confidence interval around zero. Approximate confidence intervals can be computed. The correlogram of the true residuals (which are unknown) is such that r_k is normally distributed with mean

$$E(r_k) = 0 \quad (10)$$

and variance

$$\text{var}(r_k) = \frac{1}{N} \quad (11)$$

where r_k is the autocorrelation coefficient of the ARMA residuals at lag k . The appropriate confidence interval for r_k can be found by referring to a normal distribution cdf. For example, the 0.975 probability point of the standard normal distribution is 1.96³. The 95% confidence interval for r_k is therefore $\pm 1.96 / \sqrt{N}$. For the 99% confidence interval, the 0.995 probability point of the normal cdf is 2.57. The 99% CI is therefore $\pm 2.57 / \sqrt{N}$. An r_k outside this CI is evidence that the model residuals are not random.

A subtle point that should be mentioned is that the correlogram of the estimated residuals of a fitted ARMA model has somewhat different properties than the correlogram of the true residuals – which are unknown because the true model is unknown. As a result, the above approximation (11) *overestimates* the width of the CI at low lags when applied to the acf of the residuals of a fitted model (Chatfield, 2004, p. 68). There is consequently some bias toward concluding that the model has effectively removed persistence. At large lags, however, the approximation is close.

A different approach to evaluating the randomness of the ARMA residuals is to look at the acf “as a whole” rather than at the individual r_k 's separately (Chatfield, 2004). The test is called the *portmanteau lack-of-fit test*, and the test statistic is

$$Q = N \sum_{k=1}^K r_k^2 \quad (12)$$

This statistic is referred to as the *portmanteau statistic*, or “ Q ” *statistic*. The Q statistic, computed from the lowest K autocorrelations, say at lags $k = 1, 2, \dots, 20$, follows a χ^2 distribution with $(K - p - q)$ degrees of freedom, where p and q are the AR and MA orders of the model and N is the length of the time series. If the computed Q exceeds the value from the χ^2 table for some specified significance level, the null hypothesis that the series of autocorrelations represents a random series is rejected at that level.

The p -value gives the probability of exceeding the computed Q by chance alone, given a random series of residuals. Thus non-random residuals give *high* Q and *small* p -value. The significance level is related to the p -value by

³ The MATLAB© `disttool` function is a handy interactive graphics tool for getting probability points of the cdf for various distributions

$$\text{significance level (\%)} = 100(1 - p) \quad (13)$$

A significance level greater than 99%, for example, corresponds to a *p*-value *smaller* than 0.01.

Checking the model – are the estimated coefficients significantly different from zero?

Besides the randomness of the residuals, we are concerned with the *statistical significance of the model coefficients*. The estimated coefficients should be significantly different than zero. If not, the model should probably be simplified, say, by reducing the model order. For example, an AR(2) model for which the second-order coefficient is not significantly different from zero might be discarded in favor of an AR(1) model. Significance of the ARMA coefficients can be evaluated by comparing estimated parameters with the standard deviations. For an AR(1) model, the estimated first-order autoregressive coefficient, \hat{a}_1 , is normally distributed with variance

$$\text{var}(\hat{a}_1) = \frac{(1 - \hat{a}_1^2)}{N} \quad (14)$$

where N is the length of the time series. The approximate 95% confidence interval for \hat{a}_1 is therefore two standard deviations around \hat{a}_1 :

$$\hat{a}_1 \pm 2\sqrt{\text{var}(\hat{a}_1)} \quad (15)$$

For example, if the time series has length 300 years, and the estimated AR(1) coefficient is $\hat{a}_1 = 0.60$, the 95% confidence interval is

$$\begin{aligned} 95\% \text{ CI} &= \hat{a}_1 \pm 2\sqrt{\text{var}(\hat{a}_1)} \\ &= 0.60 \pm 2\sqrt{\text{var}(\hat{a}_1)} \\ &= 0.60 \pm 2\sqrt{\frac{(1 - \hat{a}_1^2)}{N}} \\ &= 0.60 \pm 2\sqrt{\frac{(1 - 0.60^2)}{300}} = 0.60 \pm 2\sqrt{\frac{0.64}{300}} \\ &= 0.60 \pm 2\sqrt{.00213} = 0.60 \pm 0.092 \end{aligned} \quad (16)$$

The estimated parameter $\hat{a}_1 = 0.60$ is therefore significant as the confidence band does not include zero and in fact is highly significant as the confidence band is far from zero. Equations are also available for the confidence bands around estimated parameters of an MA(1) model and higher-order AR, MA, and ARMA models (e.g., Anderson 1975, p. 70). The estimated parameters should be compared with their standard deviations to check that the parameters are “significantly” different from zero.⁴

5.4 Practical vs statistical significance of persistence

Note from equation (14) that the variance of the estimated autoregressive coefficient for an AR(1) model is inversely proportional to the sample length. For long time series (e.g., many hundreds of observations), ARMA modeling may yield a model whose estimated parameters are significantly different from zero but very small. The persistence described by such a model might actually account for a tiny percentage of the variance of the original time series.

⁴ The `present` function in the MATLAB© System Identification Toolbox is convenient for getting the standard deviations of estimated ARMA parameters

A measure of the *practical significance* of the autocorrelation, or persistence, in a time series is the percentage of the series variance that is reduced by fitting the series to an ARMA model. If the variance of the model residuals is much smaller than the variance of the original series, the ARMA model accounts for a large fraction of the variance, and a large part of the variance of the series is “due to persistence.” In contrast, if the variance of the residuals is almost as large as the original variance, then little variance has been removed by ARMA modeling, and the variance due to persistence is small. A simple measure of fractional variance due to persistence:

$$R_p^2 = 1 - \frac{\text{var}(e_t)}{\text{var}(y_t)} \quad (17)$$

where $\text{var}(y_t)$ is the variance of the original series, and $\text{var}(e_t)$ is the variance of the residuals of the ARMA model.⁵ Whether any given value of R_p^2 is *practically significant* is a matter of subjective judgment and depends on the problem. For example, in a time series of tree-ring index, $R_p^2 = 0.50$ would likely be considered practically significant, as half the variance of the original time series is explained by the modeled persistence. On the other hand, $R_p^2 = 0.01$ might well be dismissed as practically insignificant.

Example. To illustrate the steps in ARMA modeling consider the fitting of a model to a time series of tree-ring index (Figure 5.2). The tree-ring index covers several hundred years, but for illustrative purposes, only a portion of the record has been used in the modeling. The time series plot strongly suggests positive autocorrelation, as successive observations tend to persist above or below the sample mean. The series has a low-frequency spectrum, with much of the variance at wavelengths longer than 10 years. This shape of spectrum is broadly consistent with a positively autocorrelated series. The acf indicates significant positive autocorrelation out to a lag of 4 years, and the pacf “cuts off” after lag 4. These acf and pacf patterns alone are enough to suggest an AR(4) model.

Fitting an AR(4) model to the data results in the equation

$$y_t - 0.3754y_{t-1} - 0.2192y_{t-2} + 0.1199y_{t-3} - 0.2547y_{t-4} = e_t \quad (18)$$

which is in the form of equation (4) extended to autoregressive order $p=4$. Whether the coefficients are significantly different from zero can be evaluated by comparing the estimated coefficients with their standard deviations:

a1: -0.3754 (±0.0956)
a2: -0.2192 (±0.1036)
a3: 0.1199 (±0.1035)
a4: -0.2547 (±0.0968)

Since twice the standard deviation is an approximate 95% confidence interval for the estimated coefficients, all except $a3$ are significantly different from zero. The AR(4) cannot therefore be ruled out because of insignificant coefficients, especially since the highest-order coefficient, $a4$, is significant.

⁵ The equation for the fractional percentage of variance due to persistence uses the one-step-ahead residuals

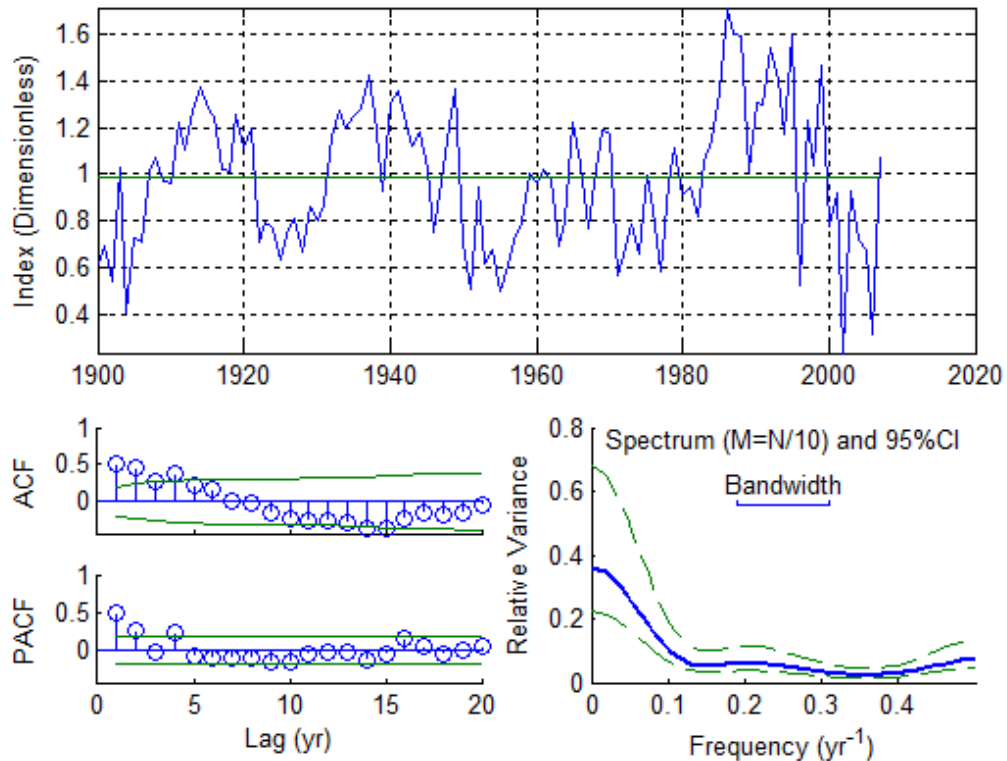
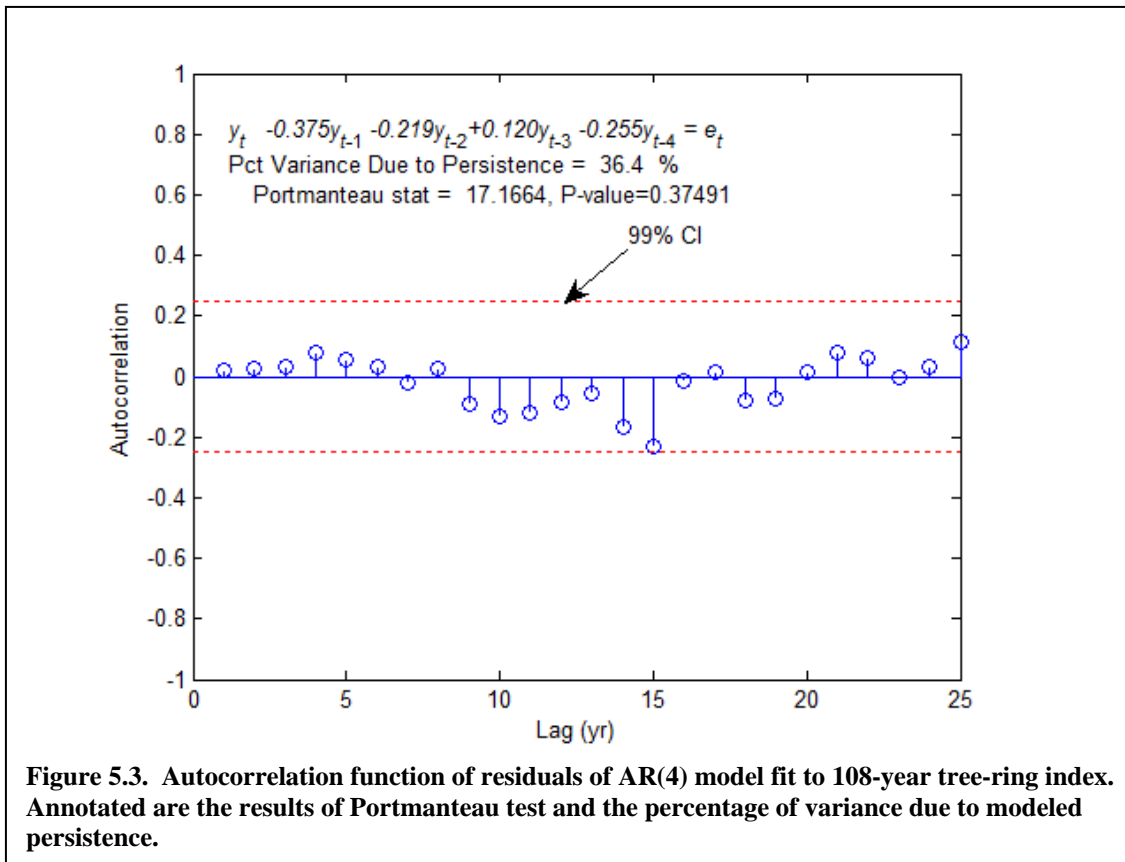


Figure 5.2. Diagnostic plots for ARMA modeling of a 108-year segment of a tree-ring index. Time series is for a *Pinus strobiformis* (Bear Canyon, Jemez Mtns, New Mexico) standard chronology. Top: time plot, 1900-2007. Bottom left: autocorrelation function and partial autocorrelation function with 95% confidence interval. Bottom right: spectrum. Data: *Pinus strobiformis* (Bear Canyon, Jemez Mtns, New Mexico) standard chronology.

Of key importance is the “whiteness” of the residuals – are they non-autocorrelated? A acf plot for the tree-ring example reveals that none of the autocorrelations of the AR(4) model residuals are outside the 99% confidence interval around zero (Figure 5.3). This is a desired result, as effective ARMA modeling should explain the persistence and yield random residuals. Whiteness of residuals can alternatively be checked with the Portmanteau test (see equation (12)). The null hypothesis for the test is all the autocorrelations of residuals for lags 1 to K are zero, where choice of K is up to the user. As annotated on Figure 5.3, for $K=25$ we cannot reject the null hypothesis. The p -value for the test is greater than 0.05 ($p=0.37491$), meaning the test is not significant at the 0.05 α -level. In summary, review of the individual autocorrelations of residuals, the Portmanteau test on those residuals, and the error bars for the estimated AR parameters favor accepting the AR(4) model as an effective model for the persistence in the tree-ring series.



5.5 Prewhitening

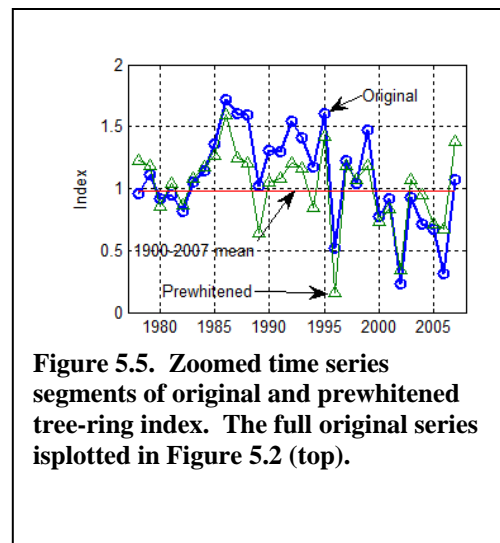
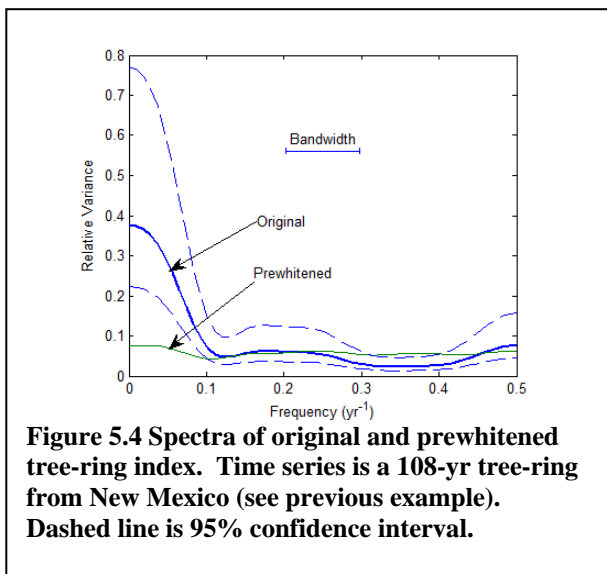
Prewhitening refers to the removal of autocorrelation from a time series prior to using the time series in some application. In dendroclimatology standard indices of individual cores are prewhitened as a step in generating “residual” chronologies (Cook 1985). Prewhitening can also be applied at the level of the site chronology to remove the persistence (Monserud 1986). In the context of ARMA modeling, the prewhitened series is equivalent to the ARMA residuals⁶. As prewhitening aims at removal of persistence, there is an expected effect on the spectrum. Specifically, for a persistent (positively autocorrelated) series, the spectrum of the prewhitened series should be flattened relative to the spectrum of the original series. A positively autocorrelated series has a low-frequency spectrum, while white noise – the objective of ARMA modeling – has a flat spectrum.

The effect of prewhitening on the spectrum is clear for our tree-ring example (Figure 5.4). The original series has a low-frequency spectrum and the prewhitened series has a nearly flat spectrum. Note also that the spectrum of the prewhitened spectrum is lower overall than that of the original series. The area under the spectrum is proportional to variance, and the removal of variance due to persistence will result in a spectrum with a smaller area. For the tree-ring series, a substantial fraction of the variance (more than 1/3) is due to persistence.

The flattening of the spectrum, a frequency-domain effect, is reflected by changes in the time domain. For a series with positive autocorrelation, prewhitening acts to damp those time series features that are characteristic of persistence. Thus we expect that broad swings above and below

⁶ In practice, the original mean of the tree-ring index is usually restored, so the prewhitened chronology has the same mean as the standard chronology, rather than a mean of zero

the mean (broad peaks and troughs) will be reduced in amplitude, or damped. The damping is evident in a zoomed portion of the time series for the tree-ring example (Figure 5.5). For example, the swing above the mean for the period 1983-1995 has been lessened in amplitude, and positive departures in 1989 and 1994 have been converted to negative departures. Specific differences in an original and prewhitened series can be readily explained by referring to the equation of the fitted ARMA model used to prewhiten. For the tree-ring example, the fitted AR(4) model is equation (18). Differences in the original and prewhitened index (Figure 5.5) derive directly from this equation. Recall first of all that y_t in the equation is a departure from the mean (red line in Figure 5.5). The residual is computed by summing over the departures of the original series from the mean for the previous 4 years, after multiplying those residuals by the estimated ARMA coefficients. The prewhitened index for 1988 is therefore drawn closer to the mean because negative coefficients are applied to fairly large positive anomalies 1987, 1986, and 1984.



5.6 Simulation and Prediction

In addition to helping to objectively describe persistence and providing an objective way to remove it, ARMA modeling can also be used in simulation and prediction. A brief introduction to these topics is given here. More extensive treatment can be found elsewhere (e.g., Chatfield 2004, Wilks 1995).

Simulation is the generation of synthetic time series with the same persistence properties as the observed series. Prediction is the extension of the observed series into the future based on past and present values. The AR(1) model

$$y_t + a_1 y_{t-1} = e_t \quad (19)$$

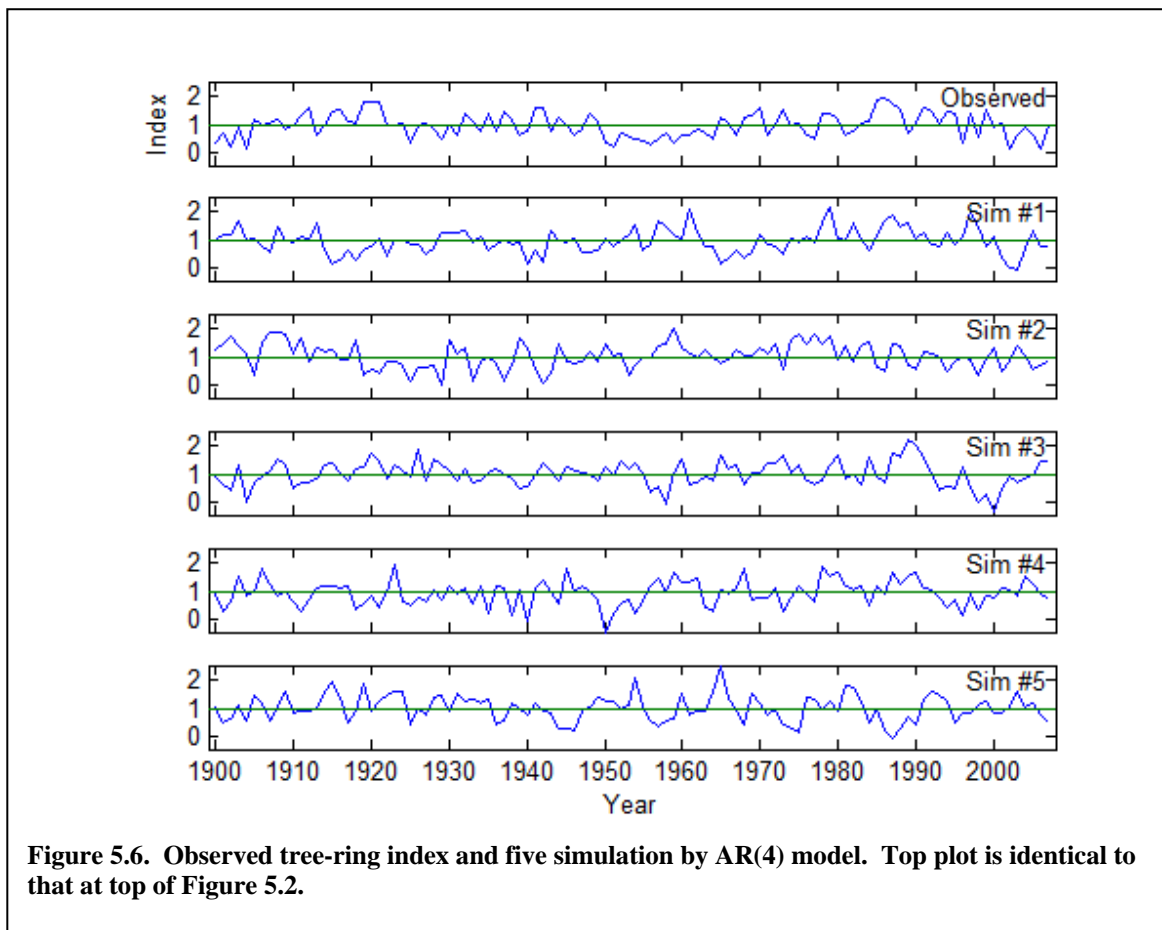
Where y_t is the departure of a time series from its mean, a_1 is the autoregressive coefficient, and e_t is the noise term, can be rearranged as

$$y_t = -a_1 y_{t-1} + e_t \quad (20)$$

A simulation of y_t can be generated from equation (20) by the following steps: 1) estimate the autoregressive parameter by modeling the time series as an AR(1) process, 2) generate a time

series of random noise, e_t , by sampling from an appropriate distribution, 3) assume some starting value for y_{t-1} , and 4) recursively generate a time series of y_t . The appropriate distribution for the noise is typically a normal distribution with mean 0 and variance equal to the variance of the residuals from fitting the AR(1) model to the data.

For example, five simulations of the tree-ring index are plotted along with original time series in Figure 5.6. The simulations appear to effectively mimic the low-frequency behavior of the observed series. Any synchrony in timing of variations in the various series is due to chance alone, as the simulations have been randomly generated. A possible application of such series is to generate empirical (as opposed to theoretical) confidence intervals of the relationship between the observed time series and a climate variable. For such an application, many (e.g., 10,000) simulated tree-ring series of length 108 years can be generated, correlation coefficients computed between each simulation and the climate series, and 95% confidence interval for significant correlation established as the 0.025 and 0.975 probability points of the 10,000 sample correlations. If the correlation between the observed time series and the climate series is outside the confidence interval, a significant statistical relationship is inferred.



Prediction differs from simulation in that the objective of prediction is to estimate the future value of the time series as accurately as possible from the current and past values. Unlike simulations, predictions utilize past values of the observed time series. A prediction form of the AR(1) model is

$$\hat{y}_t = -\hat{a}_1 y_{t-1} \quad (21)$$

where the “^” indicates an estimate. The equation can be applied “one step ahead” to get estimate \hat{y}_t from observed y_{t-1} . A “ k -step-ahead” prediction can also be made by recursive application of equation (21). In recursive application, the observed y at time 1 is used to generate the estimated \hat{y} at time 2. That estimate is then substituted as y_{t-1} to get the estimated \hat{y} at time 3, and so on. The k -step-ahead predictions eventually converge to zero as the prediction horizon, k , increases⁷.

Prediction is illustrated in Figure 5.7 for the tree-ring example. Recall that the index for the 108-year period 1900-2007 was fit with an AR(4) model, which explained about 1/3 the variance of the series. The segments of observed and predicted index beginning with 1990 are plotted in Figure 5.7. The observed index ends in 2007, and for the years 1990-2007 the predicted values plotted are one-step-ahead predictions. Such predictions make use of all observed data for times $t \leq (t - 1)$ to make the prediction for year t . The predictions for years beginning with 2008 (future) are k -step-ahead predictions. These predictions use observed index for times $t \leq (t - k)$ and predicted index for later years before year t in making a prediction for year t . The predicted index for 2008 is a one-step-ahead prediction, for 2009 a 2-step-ahead prediction, for 2010 a 3-step-ahead prediction and so forth. The persistence in the data combined with the low-growth period starting about 2000 leads to predictions of below normal tree-ring index for the 10 years of prediction horizon beyond the end of the observed data. The predictions converge toward the mean as the horizon lengthens, however, because the influence of the past gradually fades.

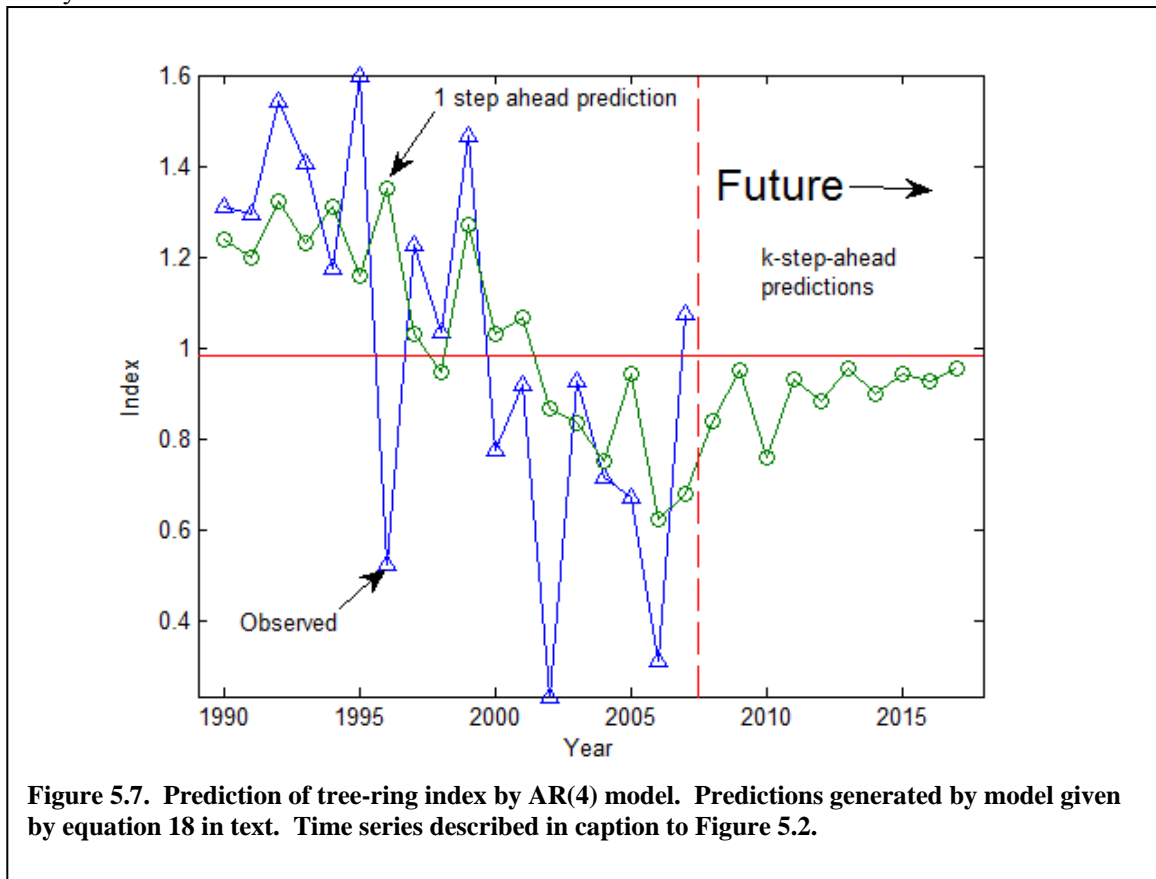


Figure 5.7. Prediction of tree-ring index by AR(4) model. Predictions generated by model given by equation 18 in text. Time series described in caption to Figure 5.2.

⁷ Because the modeling assumes y_t is a departure from the mean, this convergence in terms of the original time series is convergence toward the mean

5.7 Extension to nonstationary time series

ARMA modeling assumes the time series is weakly stationary. With the appropriate modification, nonstationary series can also be studied with ARMA modeling. Trend defined as a deterministic function of time can be removed by curve-fitting prior to ARMA modeling. That is the approach in dendrochronology, where a smooth curve describing the “growth trend” is removed in converting ring widths to ring indices. Periodic time series are a special case in which the trend is cyclical. An example of a periodic series is a monthly time series of air temperature with its annual cycle. The mean is clearly nonstationarity in that it varies in a regular pattern depending on month. One way of handling such a series with ARMA modeling is to remove the annual cycle by expressing the monthly data as departures from their long-term monthly means. Another way is by applying periodic ARMA models, in which separate parameters are simultaneously estimated for each month of the year. Periodic ARMA models are discussed by Salas et al. (1980).

A trend itself might be a stochastic feature of a time series. For example, a random walk time series wanders with a time varying population mean. The series does not tend to return to any specific preferred level. Such series can be detrended by first-differencing before ARMA modeling – a modeling approach called autoregressive-integrated-moving-average (ARIMA) modeling. Further discussion of ARIMA modeling can be found elsewhere (Anderson 1976; Box and Jenkins 1976; Salas et al. 1980)

5.8 References

- Anderson, O., 1976, Time series analysis and forecasting: the Box-Jenkins approach: London, Butterworths, p. 182 pp.
- Box, G.E.P., and Jenkins, G.M., 1976, Time series analysis: forecasting and control: San Francisco, Holden Day, p. 575 pp.
- Chatfield, C., 2004, The analysis of time series, an introduction, sixth edition: New York, Chapman & Hall/CRC.
- Cook, E.R., 1985, A time series approach to tree-ring standardization, Ph. D. Diss., Tucson, University of Arizona.
- , Shiyatov, S., and Mazepa, V., 1990, Estimation of the mean chronology, *in* Cook, E.R., and Kairiukstis, L.A., eds., Methods of dendrochronology, applications in the environmental sciences: In.: Kluwer Academic Publishers, p. 123-132.
- Ljung, L., 1995, System Identification Toolbox, for Use with MATLAB, User's Guide, The MathWorks, Inc., 24 Prime Park Way, Natick, Mass. 01760.
- Monserud, R., 1986, Time series analyses of tree-ring chronologies, *Forest Science* 32, 349-372.
- Salas, J.D., Delleur, J.W., Yevjevich, V.M., and Lane, W.L., 1980, Applied modeling of hydrologic time series: Littleton, Colorado, Water Resources Publications, p. 484 pp.
- Wilks, D.S., 1995, Statistical methods in the atmospheric sciences: Academic Press, 467 p.