Time Series HW 8

Allison Theobold, Andrea Mack, and Kenny Flagg

November 29, 2016

Read Monserud and Marshall (2001) "Time-series analysis of..." All answers to the following should be short even though the paper is long and contains a lot of information. It ends up introducing concepts that we will spend the rest of the semester digging into and does that pretty well. Focus on the statistical methods employed, reasons for using them, and how they discuss their results. This should be done in groups of up to 3. You will lose 2% if you do not work in a group of 2 and get a bonus of 2% if you work in a group of 3.

1. *What do they provide for a reason for removing the quadratic trend? Why do they want to use a trend model with limited flexibility?*

   The authors fit both a linear and a quadratic trend to the original data, with the residuals from both fits having a constant mean and variance. They then "used standard statistical procedures (e.g., t-test) to decide between the linear and quadratic forms." Supposedly, from this t-test the p-value for the quadratic trend was deemed "too high" and the trend was removed. The authors describe that "care should be taken to describe and remove the trend in the data over time", with the goal of "goal is to isolate the influence of climate on tree growth". The authors describe the cautioning in fitting trends so that one "remove(s) just enough of the trend so that the residuals are stationary, without heavy-handedly removing the additional information (underlying signal) we seek to analyze".

2. *How do they determine the order of the trend model from none to quadratic? Are there any potential criticisms of the way they do this? Do not discuss the potential issue with doing these tests without accounting for potential autocorrelation – we know that the SEs would change if these tests were conducted using models that account for autocorrelation.*

   The authors determine the order of the trend in the model by fitting a quadratic trend, and using a t-test, with $\alpha = 0.05$, to determine if that term is "necessary" in the trend model. They determine the "significant" trends for each of the trees, as well as the mean of the species, for each of the three responses of interest (δ_p, $\frac{c_i}{c_a}$, and, Δ). The authors described in the *Gas Exchange Variables* section, the dependence of each of these three responses on each other, "these series are not independent of each other. The ratio $\frac{c_i}{c_a}$ is directly proportional to the difference $\delta_a - \delta_p$, and Δ is dominated by the numerator $\delta_a - \delta_p$. Thus, we can conclude

that the time series properties ... will be almost identical, because the variates differ only be a factor of proportionality."

Hence, the reader should be cautious of testing for trends for each of the different variables $(\delta_p, \frac{c_i}{c_a}, c_a - c_i(\text{tree-ring}), c_a - c_i(\text{leaf}), \text{and}, \Delta)$, as they are all related to one another. The result is a total of 15 tests for trends per tree, where each test is dependent on the others. This will result in a spurious detection of a *false* trend about once per tree, which can be seen in the the results in *Table 2*. For example, tree D102 has different terms included in the trend model for each of the different response variables, $\delta_p, \frac{c_i}{c_a}, \text{and}, c_i - c_a$. As these variables are directly proportional to one another, the same time trend should describe each of the responses, eliminating the need for multiple testing of the trends for each response, and a large resulting false detection rate.

3. *Figure 3 contains two ACF plots. Discuss the patterns in the two SACFs (left two panels of Figure 3).*

    Figure 3 shows the SACF plots for the mean $\delta^{13}C$ series for Douglas fir and western white pine. Both SACF plots exhibit decaying oscillations indicative of AR(2) autocorrelation. Both plots have a large lag 2 correlation (exceeding the $\alpha = 0.05$ bounds). The western white pine SACF has a large lag 1 correlation, implying that consecutive years have similar mean $\delta^{13}C$. In contrast, the Douglas fir SACF has a small lag 1 correlation, suggesting more complicated behavior where the mean $\delta^{13}C$ is related to the mean $\delta^{13}C$ from two years prior, but consecutive means are not so strongly related.

4. *Note that their bounds for testing in Figure 3 are different from the formula provided in the book – we will discuss the modification I think they used later. What is the sample size (length of time series!) for the mean Douglas-fir and mean western white pine time series. Based on these sample sizes, what cut-offs would you plot on Figure 3 based on the approximate result we discussed in class? How do theirs differ?*

    Table 2 shows that the mean Douglas fir series has data for 85 years and the mean white pine series includes data from 77 years. On the SACF plot, we would use the cutoffs

    $$\pm \frac{2}{\sqrt{85}} = \pm 0.217$$

    for the mean Douglas fir series and

 $$\pm \frac{2}{\sqrt{77}} = \pm 0.228$$

    for the mean white pine series. Rather than using these constant values, the cut-offs in Figure 3 generally increase for larger lags, presumably to account for the lower number of pairs used to estimate the sample autocorrelation at larger lags.

5. *We discussed four different reasons for doing time series modeling at the beginning of the semester (pages 1 and 2 of lecture notes). Of those four reasons, which one does their use of ARMA models fall into and why? There may be multiple choices that are reasonable here, so the argument you make is important. I am not talking about what you could do with their models, just what they did do with them and their reason for employing them.*

Monserud and Marshall use ARMA models to describe and understand the trends in carbon isotope ratios. In the summary section, the authors emphasize that the residuals of an appropriate ARMA model will satisfy the independence assumption and that hypothesis tests yield biased results when "significant" autocorrelation is present but not accounted for. The remainder of the paper proceeds to describe AR and MA models, look at visual diagnostics for these types of models, and discuss the trends that were estimated in the context of biological theory. The bigger picture for this research is that scientists will need to accommodate dependency in models used for forecasting or exploring the relationships between carbon isotope levels and various predictors, but in this study the authors focus on explicating the trends and dependence structures that may be useful in such models.

6. *Use `arima.sim` to simulate from Delta response AR(2) models for either the mean D or mean W discussed in Table 4 generating the same number of observations as in their time series (see Table 2). You can use a standard deviation for the white noise process of 0.4. This simulation is just concerned with the residual process after removing the trend.*

   Plot a simulated fake time series, the SACF from the fake time series, and compare the results to those that you can find in the paper in Figure 3.
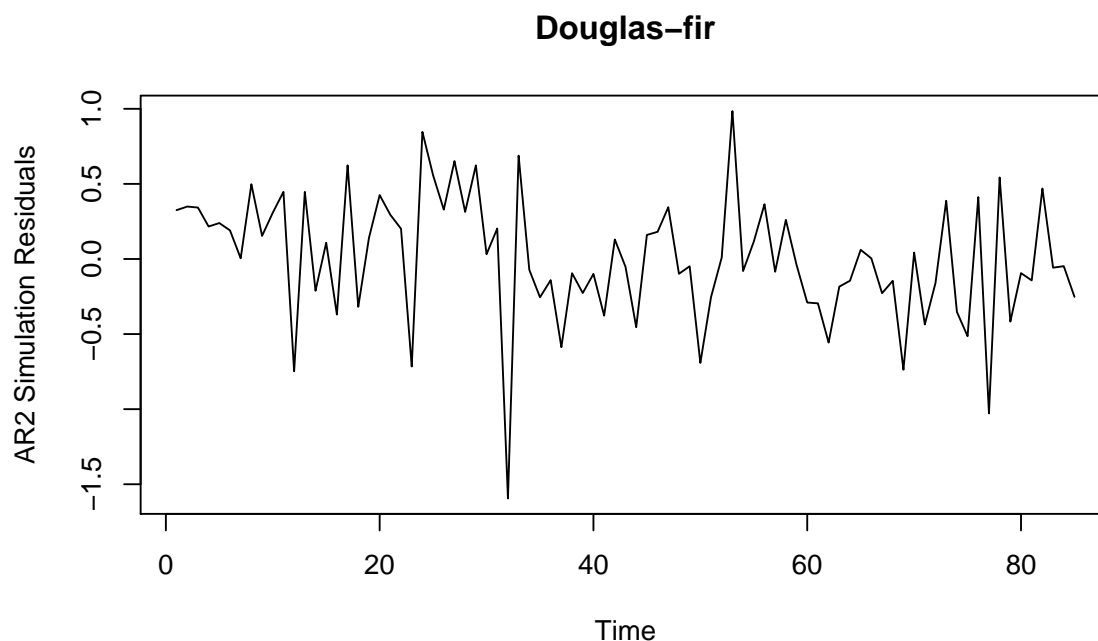
Douglas–fir

Douglas–fir AR2
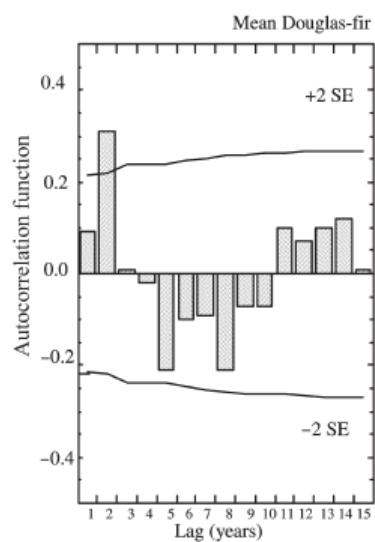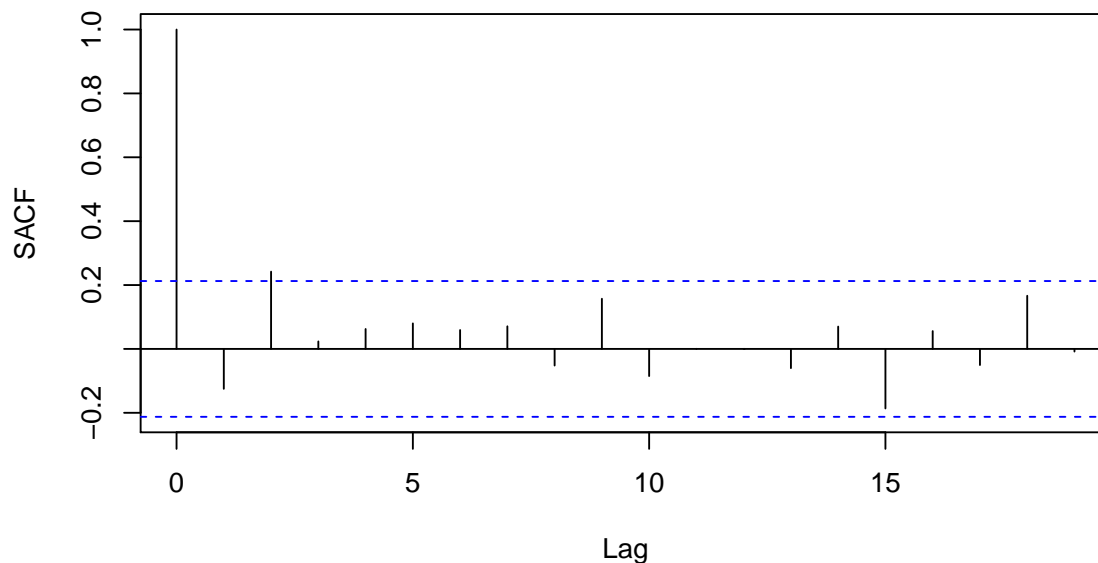
Figure 3

The SACF shows a significant correlation at lag 2. The ACF showed a significant correlation at lag 2 and approaching significant correlations at lags 5 and 8.

How are your results similar or different from their results?

In terms of the correlations at lags 2, the results are similar, both suggesting a significant correlation. Successive patterns in significant correlations are different between the SACF and ACF plots. In the ACF plot, lags 5 and 8 appear to almost have significant correlations
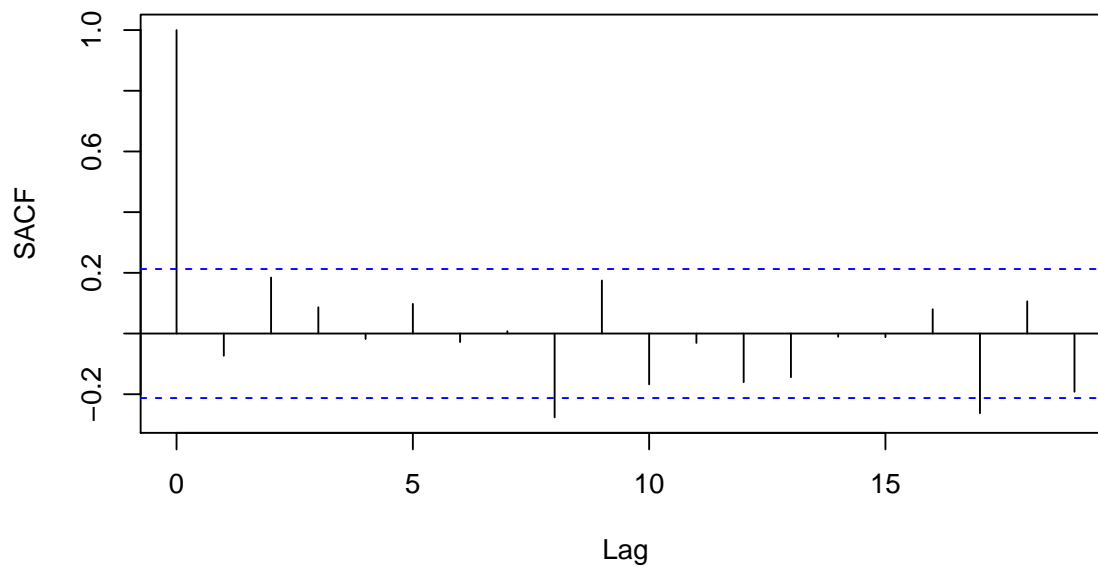
4

while in the SACF, lag 9 is nearing having a significant correlation with subsequent spurious correlations that are almost significant also appearing.

Simulate a second series and discuss the differences in your two simulations.

Douglas–fir
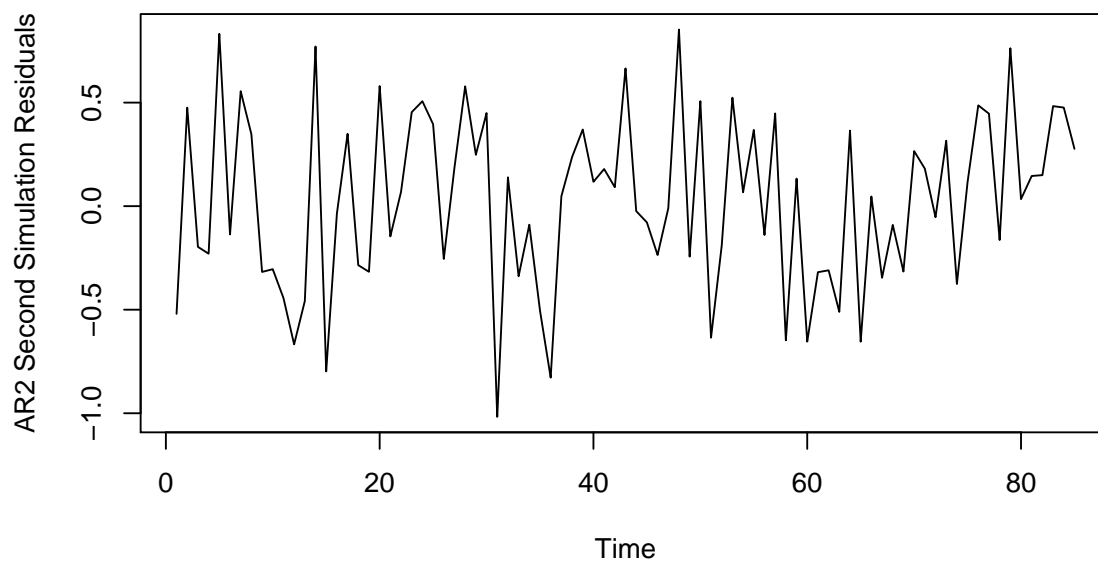
Douglas–fir Second AR2

The first plot of the simulation residuals shows a more apparent positive autocorrelation pattern, although not strong, as residuals at time points closer together are more similar in sign. There appears to be more extreme spikes either positive or negative occurring every 15 to 20 time points out.

The second plot of simulation residuals has much more extreme subsequent simulated residuals although the positive autocorrelation is still apparent. The more extreme spikes still appear every 15 to 20 time points out until around time lag 60, at which time, a increasing positive overall trend is seen until the end of the time series.

Both plots of simulated residuals look quite rough.

In terms of SACF plots, the first simulation showed a significant correlation at lag 2, while the only significant correlation (that is possibly not spurious) is at lag 8.

7. *Use R to find the roots of your selected AR(2) model based on their published coefficients and note what that tells you about the properties of the AR(2) process under consideration.*

```
          Roots
1  1.676501+0i
2 -1.864001+0i
```

Since the roots are not complex, there is not a quasi-periodic structure.re are complex roots, the process has a quasi-periodic structure.

Note that $\phi_1 = 0.06$ and $\phi_2 = 0.32$. The AR(2) process is stationary iff:

- $\phi_1 + \phi_2 < 1$

  $0.06 + 0.32 = 0.38 < 1$

- $\phi_2 - \phi_1 < 1$

  $0.32 - 0.06 = 0.26$

- $|\phi_2| < 1$

  $0.32 < 1$

Therefore, the process is stationary.

R Code

Problem 6

```r
# mean d has 85 years, n = 85
# use 0.4 for sd
# get ar components from table 5

set.seed(53608)
dar2 <- arima.sim(n = 85, list(ar = c(0.06,0.32)), sd = 0.4)


plot(dar2, ylab = "AR2 Simulation Residuals", main = "Douglas-fir")

acf(dar2, main = "Douglas-fir AR2", ylab = "SACF")


dar2_2 <- arima.sim(n = 85, list(ar = c(0.06,0.32)), sd = 0.4)

plot(dar2_2, ylab = "AR2 Second Simulation Residuals", main = "Douglas-fir")

acf(dar2_2, main = "Douglas-fir Second AR2", ylab = "SACF")
```

Problem 7

```r
# change parameter estimates here also depending on the delta he means
# Should be polyroot(c(1, -0.06, -0.32))
# See CC page 71 or Mark's notes page 9 -- The coefficients of the polynomial are the negative phi values
ar2_roots <- data.frame(polyroot(c(1,-0.06,-0.32)))
colnames(ar2_roots) <- "Roots"

ar2_roots
```