Analysis of the Central England Temperature Record

Michael Frasco

March 15, 2015

Summary

In this paper, I conclude that a warming trend does exist in the Central England Temperature (CET) record. I find that it is appropriate to fit the data as noise around a stationary and causal trend, rather than just as a random walk. I attempt to confirm and expand on claims made in previous publications. I found that the model that best captured the trend and dependence structure in the CET record used local polynomial kernel regression. I chose the smoothing parameter for this model using Whaba's leave one out cross validation. I also investigated whether periodic components existed in the CET record. Unlike Benner (1999), I was unable to confirm that any possible frequencies had significant peaks in the standard fourier periodogram. However, incorporating a periodic component did result in an improved model.

Literature Review

Before I begin my analysis of the Central England Temperature record, I will review the existing literature that I read on this topic, which inspired my own analysis. Upon reading about the lack of consistent temperature measurments before 1708 in Manley (1974), I have decided to follow Harvey and Mills (2003) and limit the focus of my analysis to the record after 1723.

Jones and Bradley (1992a) have one of the first papers about the temperature trends in the CET record, in which they conclude that the CET "hardly shows any warming between the early 1700s and 1980" and that warming is least evident in the summer season. Jones and Bradely claim that a linear fit yields a 0.51 degree increase in the annual temperature from 1751 to 1980. A major component of deciding whether the recent uptick in annual temperature is understanding how variable annual temperatures have been for the past 300 years. To this point, Jones and Bradley state that the variability tends to be the greatest in the winter, followed by spring, autumn and summer. However, they do not provide a quantiative analysis.

Benner (1999) claims that the results of his analysis on the CET record "support a possible warming trend, especially in recent years." Although he observes that the early part of the record experienced a cold spell and the middle part was relatively stable, Benner states that simple quadratic fits applied to the entire data record, to the period after 1800, and to the period since 1900 all support an overall warming trend. Harvey and Mills (2003) directly refute Benner on the issue of a general increase in the CET. They claim that it is important to understand the stochastic process generating the data, rather than just fitting a simple linear trend. The recent increase in annual temperature could be due to periodic oscillations in the series. Harvey and Mills argue that the CET trends are highly nonlinear and need to be modelled non-parametrically. They observe periods of cooling, stability, and warming, and find no compelling evidence of a general upward trend in the CET over the last 300 years.

No one denies that the most recent decades have been the warmest decades in the entire span of the CET. However, what is uncertain is whether the increase in the annual temperature over the last 50 years falls within the normal variability of the temperature record. Furthermore, as Benner (1999) remarks, "any potential trends must be separated from the long-period oscillations in the data." Benner (1999) provides a thorough examination of the periodic frequencies found in the CET record, discovering oscillations from three years to 113 years present. He evaluates the red noise significance level, which I will do also, of these frequencies and also tries to determine if they persist throughout the entire series.

Analysis

In this section, I aim to justify the claims of Jones and Bradley (1992a), Benner (1999), and Harvey and Mills (2003) that I mentioned in the literature review. I will do this by replicating their results.

Claim 1

Jones and Bradley (1992a) claim that the annual temperature change accounted for by the linear trend starting at 1751 and ending at 1980 is 0.51 degrees. I take this to mean that when they fit a linear trend to the data, the difference in the predicted temperature in 1980 and is 0.51 degrees. Similarly they found that a linear fit starting at 1801 accounted for 0.49 degrees and one starting at 1851 accounted for 0.49 degrees. I replicated these models using the data for each starting period up to 1980, as well as for each starting period up to 2014. Below is a table of the values.

```
##              Starting in 1751  Starting in 1801 Starting in 1851
## Up to 1980               0.51              0.55             0.53
## Up to 2014               0.93              1.01             1.05
```

Notice in the in the table above that the temperature increase accounted for by the linear model is much greater when we use the most recent data. This means that the recent data provides greater evidence for warming. While Benner and Jones said that there was hardly any evidence of a warming trend in the data up to 1980, I wonder if they would be persuaded if they saw the continued temperature increase.

Claim 2

Jones and Bradley (1992a) used linear models to fit any posible trend in the data. Benner (1999) claims that simple quadratic fits applied to the entire data record, to the period after 1800, and to the period since 1900 all support an overall warming trend. To investigate and compare these claims, I plotted the best linear and quadratic fit on top of the data. I used the data from 1723 to 2014 for both models.

**Annual Temperature
From 1723 to 2015**
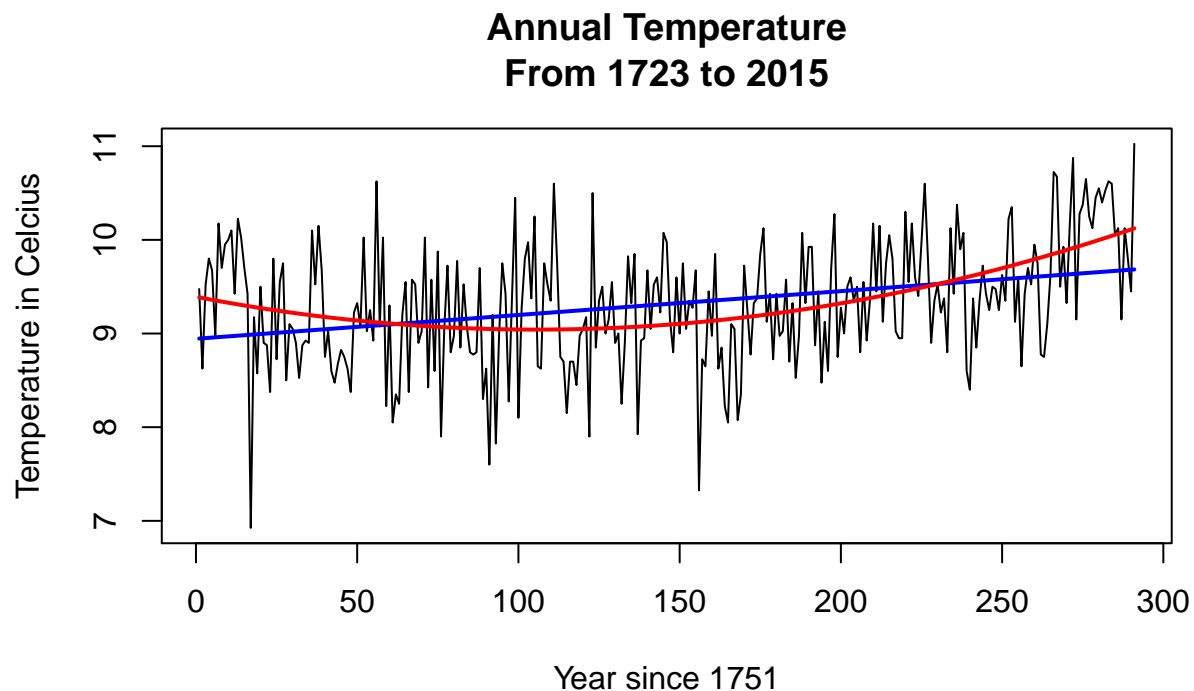
From the plot above, we can see that the quadratic model seems to capture the recent rise in temperature better than the linear model. However, both models fail to account for the fluctuations in the middle of the CET record. To get a better sense of the numbers from the resulting best quadratic fit, I have printed the summary table of the regression.

```
##
## Call:
## lm(formula = annual1723 ~ Year + Year_Squared)
##
## Residuals:
```

```
##      Min     1Q  Median      3Q     Max
## -2.3635 -0.3825 -0.0283  0.4506  1.5571
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.39e+00   1.06e-01   88.54  < 2e-16 ***
## Year         -6.64e-03   1.68e-03   -3.96  9.6e-05 ***
## Year_Squared 3.14e-05   5.56e-06    5.65  3.9e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.599 on 288 degrees of freedom
## Multiple R-squared:  0.193,  Adjusted R-squared:  0.187
## F-statistic: 34.4 on 2 and 288 DF,  p-value: 3.86e-14
```

Note that all three terms (the intercept, the coefficient for time as a linear input, and the coefficient for time as a quadratic input) are all highly significant. It is interesting to note that the coefficient for the linear input is negative. This accounts for the decrease in predicted temperature during the 1700s that we see in the graph. And this also highlights the fact that fitting a global model to the CET record isn't a very appropriate thing to do, since this model would predict extreme temperatures in the year 1000 or 3000 BC.

```
##             AIC   BIC
## Linear    561.2 572.2
## Quadratic 532.6 547.3
## Cubic     534.5 552.9
```

Above I compared the AIC and BIC values for the linear, quadratic and cubic model over the entire data set. Notice that the quadratic model performs the best on both criteria. I must also point out that that none of the coefficients (except for the intercept) were significant at even a 10% level for the cubic model and the cubic model's fit were almost identical to that of the quadratic model. Even though there appears to more nonlinearity in the data than the quadratic model can account for, the quadratic model is the best of the three.

Remarks

Of course, simply fitting a linear or quadratic trend, as has been done above, may be inappropriate since it is of crucial importance to analyse the underlying stochastic process generating the data and its periodic oscillations. Before I approach these problems, let me take a moment to address one of theoretical issues at stake here.

If you believe that the geophysical process generating the temperature series can be modeled as a random walk with drift, the recent the annual temperature increase does not represent a permanent change in climate conditions. On the other hand, if you believe that the CET record is better modeled as a stationary process around a global trend, then the temperature increase signifies that global warming is a serious threat to the climate. These two different opinions are at the heart of the debate around global warming. By using unit root tests such as the DF, ADF, and PP tests, we can evaluate the strength of the null hypothesis (random walk with drift) against the alternative hypothesis (stationary around a global trend).

```
##
##  Augmented Dickey-Fuller Test
##
## data:  annual1723
## Dickey-Fuller = -7.426, Lag order = 2, p-value = 0.01
## alternative hypothesis: stationary
```

```
##
##  Phillips-Perron Unit Root Test
##
## data:  annual1723
## Dickey-Fuller Z(alpha) = -282.8, Truncation lag parameter = 5,
## p-value = 0.01
## alternative hypothesis: stationary
```

The p-values for the null hypothesis are all much smaller than 0.01, so we have reason to believe that the CET record is stationary around a trend. However, we should not put too much weight into these tests because they make an extremely simplifying assumption. The ADF and PP tests assume that the trend can be modeled by an AR(p) plus linear trend process. As we have seen from the data, a nonlinear model appears to be far more appropriate. Nevertheless, the unit root tests can provide some amount of support behind the general idea that is best to model the CET record as more than just a random walk with drift.
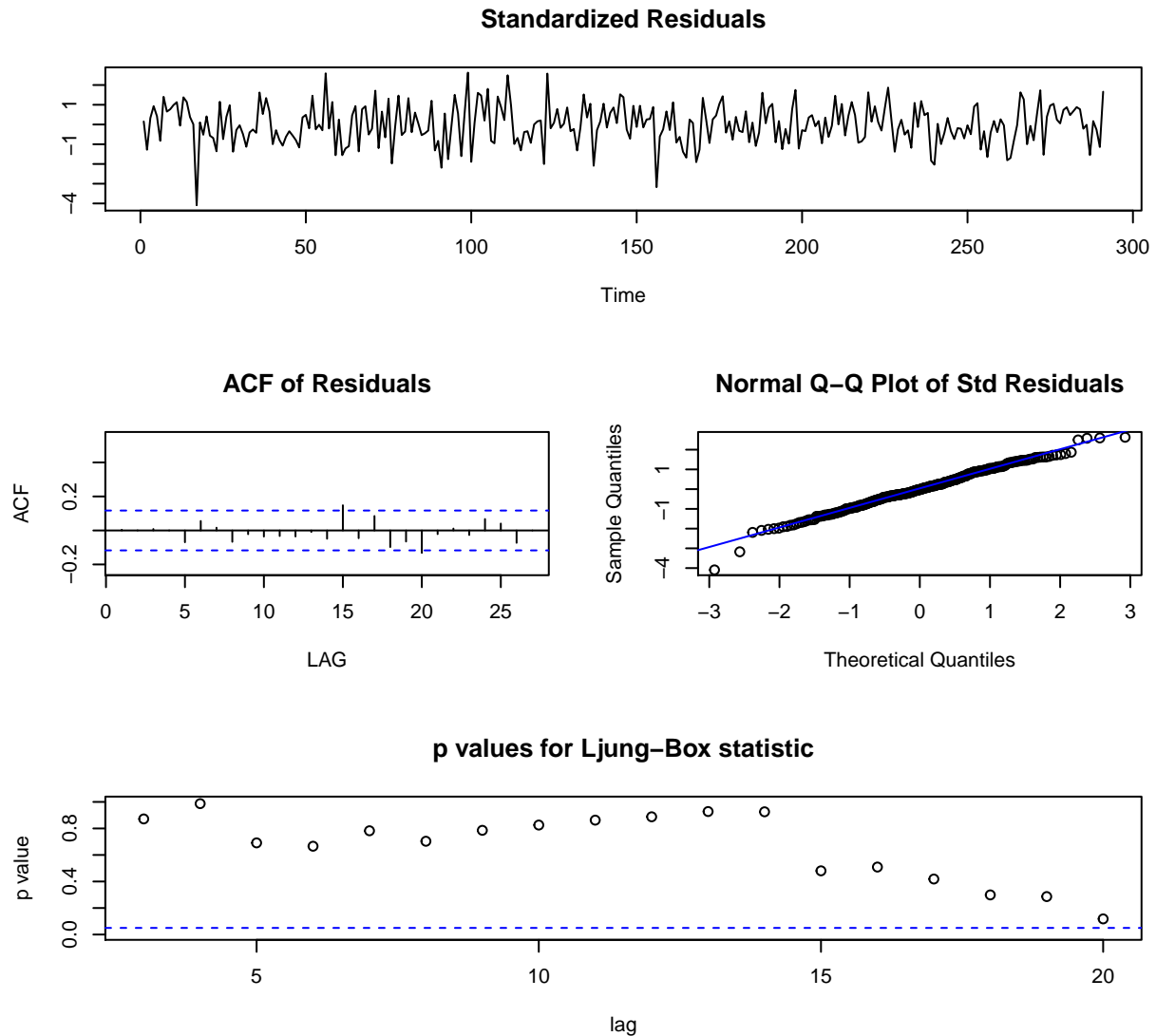
Claim 4

Before moving to non-parametric methods, Harvey and Mills (2003) improve on the quadratic model of Benner by incorporating an MA(2) component. The choice of using two moving average parameters in the model comes from examining the sample ACF and PACF plots of the detrended CET record. The trend used to detrend the record is the quadratic fit from above. By examining the ACF and PACF of the residuals from the quadratic fit, we can get a sense of the dependence structure still left in the residuals. This will allow us to incorporate that dependence into our model and improve the fit.

Series: Residuals_of_Quadratic_Fit
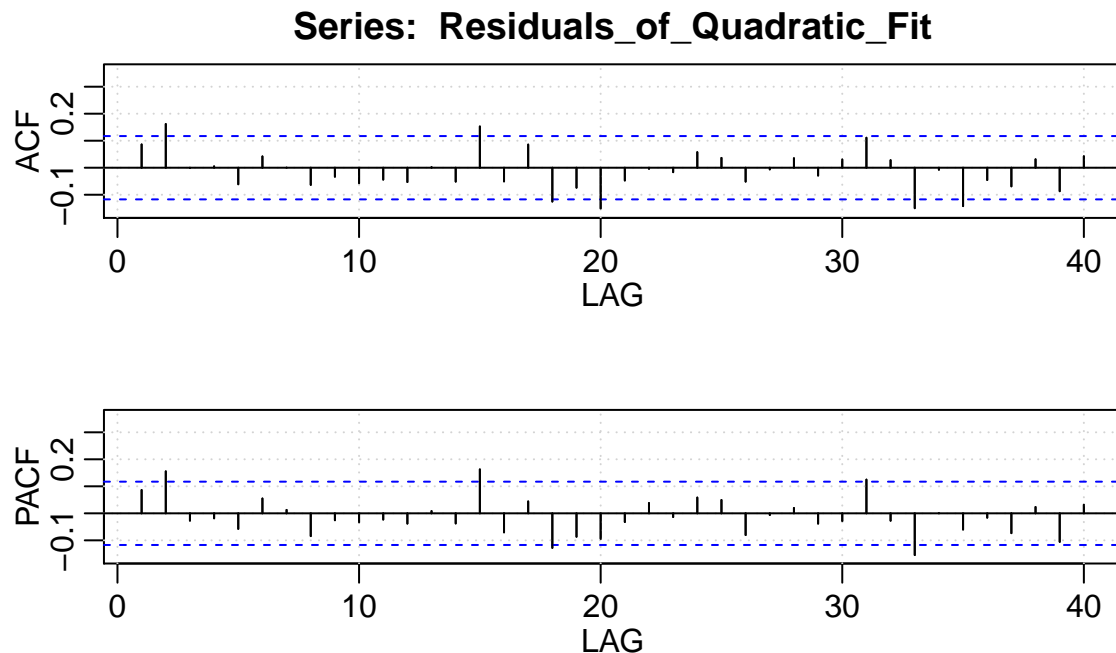
From the plots above, the first thing that jumps out at me is that the ACF plot cuts off after lag 2, which would suggest an MA(2) process. However, if the residuals could be modeled by an MA(2) process then the PACF should tail off. Instead, it also appears to also cut off after lag 2. Furthermore, there is a large amount of dependence between lags 15 and 20 and again around lag 35. This is likely caused by the harmonics of an early lag. Some sort of periodic component that could be captured by a cosine wave of some sort. I will delay a further investigation of this component until I analyze the frequencies.

I decided to fit an MA(2) model plus a quadratic trend to the annual temperature values from 1723 to the present.

Standardized Residuals

ACF of Residuals

Normal Q–Q Plot of Std Residuals

p values for Ljung–Box statistic

Above are some diagnostic plots that evaluate this model by examining the properties of its residuals. In the top graph, note that the plot of the residuals looks very much like white noise. There appears to be constant variance that is centered around the mean. This is confirmed by the lack of lagged correlation in the ACF plots. There is still some correlation between lags 15 and 20, but it is less than before. Also, notice that the p-values for the Ljung-Box statistics are all not significant. The Ljung-Box test checks if the magnitude of the correlation of residuals as a group is too large. It computes a statistic that follows a Chi-Squared distribution with H-p-q degrees of freedom, where H is the lag, to test the null hypothesis. From these diagnostics, we can conclude that our fit does very well. There are no glaring weaknesses. Here is a plot of the fitted values.

Annual Temperature
From 1723 to 2015
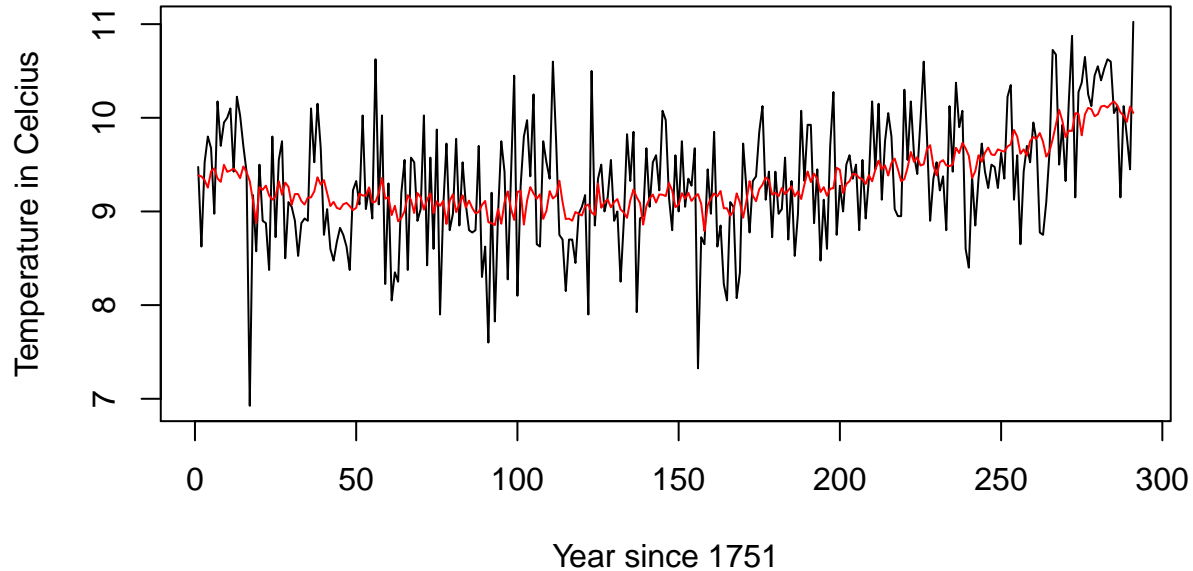

```
##            ma1         ma2     intercept         Year Year_Squared
##      7.067e-02   1.674e-01     9.390e+00   -6.625e-03    3.145e-05
```

Above, I have printed the coefficients for the MA(2) plus quadratic trend model. Notice that the coefficients for the linear and quadratic terms are very similar to those in the simple quadratic model. (The linear coefficient is still negative.) We can also see the effect of the two moving average parameters.

```
##                  AIC     AICc    BIC
## ARMA(0, 1) -0.013  -0.0057  -0.96
## ARMA(0, 2) -0.034  -0.0260  -0.97
## ARMA(1, 1) -0.018  -0.0100  -0.96
## ARMA(1, 2) -0.027  -0.0190  -0.95
```

Since it is difficult to estimate the AR and MA parameters from the ACF and PACF plots, I also created serveral other ARMA models. Each model also incorporates the quadratic trend in order to difference series. Below is a table of AIC, AICc, and BIC values from each of the four models that I experimented with. Note that the MA(2) plus quadratic model has the best results.

While the coefficients of my model to not correspond exactly to that of Harvey and Mills (2003) because I am using more recent data, the structure and relative magnitude of my MA(2) plus quadratic model does match.

Claim 5

Harvey and Mills (2003) rightly criticize the MA(2) plus quadratic model that I replicated above because it assumes that the quadratic trend is constant throughout the entire time period of 1723 to 2014. They recognize that a more flexible trend model that does not assume prior knowledge of the functional form is more appropriate.

Harvey and Mills use local polynomial kernel regression, neigherest neighbors, and low-pass filtering models to create different non-parametric models. In this paper, I will attempt to replicate their use of local polynomial kernel regression, using leave one out cross validation to optimize the smoothing parameter.

I used the locfit package in R to perform the local polynomial kernel regression. This package allowed me to select the smoothing parameter as the fraction of points in the data set that I wanted to use in the local

6

regression. I also followed Harvey and Mills (2003) and used polynomials of degree three. The logic behind choosing cubic instead of quadratic polynomials is that odd-order polynomials tend to perform better at the boundary regions than even-order polynomials. Since we are especially interested in the model's predictions in recent years, we will favor odd order polynomials. Indeed, the AIC for the best degree 2 model is 113.007, while the AIC for the best degree 3 model is 112.8. While this is not a huge difference, it does support the claim that the cubic model is better.

Below is a plot of Wahba's leave one out generalized cross validation score as a function of the fraction of the data series that was used to create each local polynomial. This is a plot for the cubic polynomials. As you can see, the optimal fraction for smoothing is 0.62.

```
## locfit 1.5-9.1     2013-03-22
```

Leave One Out Cross Validation Scores
For Different Smoothing Values

Fraction of Series Used to Smooth
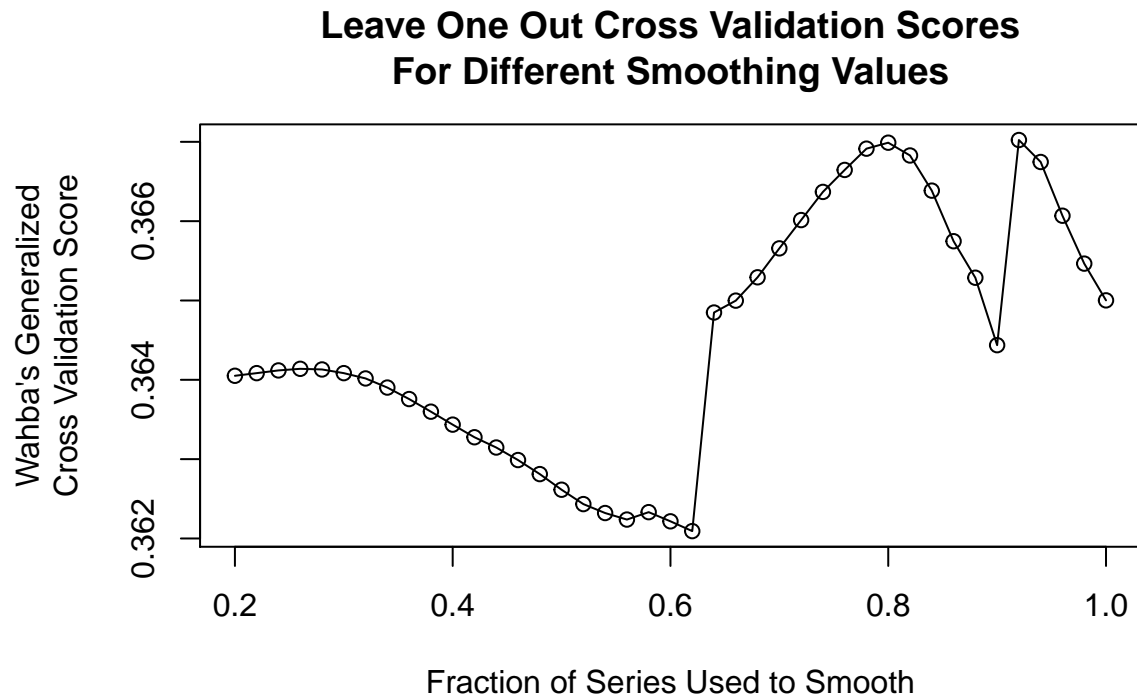
After fitting the specific local polynomial model with the chosen smoothing parameter, I plotted its predicted points on top of the observed data points. Notice that the trend has a substantial increase in the past 50 years. This serves as evidence of the global warming hypothesis.

Annual Temperature
From 1723 to 2015
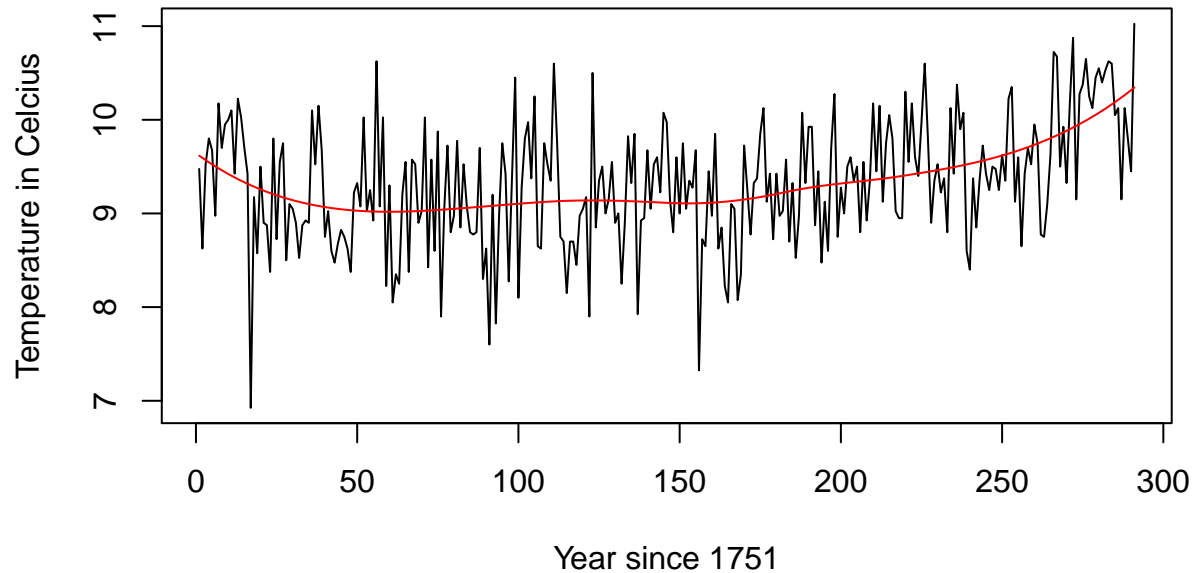

Claim 6

The next claim that I want to verify is that of Benner (1999) regarding the oscillations found in the data set. The idea behind this approach is that there may be some sinusodial variations in the trend that we have been modeling. Recall the dependence remaining the residuals of the MA(2) plus quadratic model around lag 15. While Benner uses four independent methods to find the frequencies occuring in the CET record, I will use the standard Fourier spectral analysis that we learned in class. The disadvantage of this approach is that we have poor frequency resolution at lower frequencies (since there are fewer data points to support these frequencies). Unfortunately, these low frequencies are probably the ones that we are most interested in investigating, because they would suggest long term temperature oscillations. Nevertheless, we will proceed with this standard approach.

Before we begin, we will detrend the nonstationarity in the CET record by evaluating the residuals from the MA(2) plus quadratic trend. If we did not do this, the trend could introduce extremely low frequency components in the periodogram which would obscure the appearance at higher frequencies.

Series: Model_Residuals
Raw Periodogram

frequency
bandwidth = 0.000962
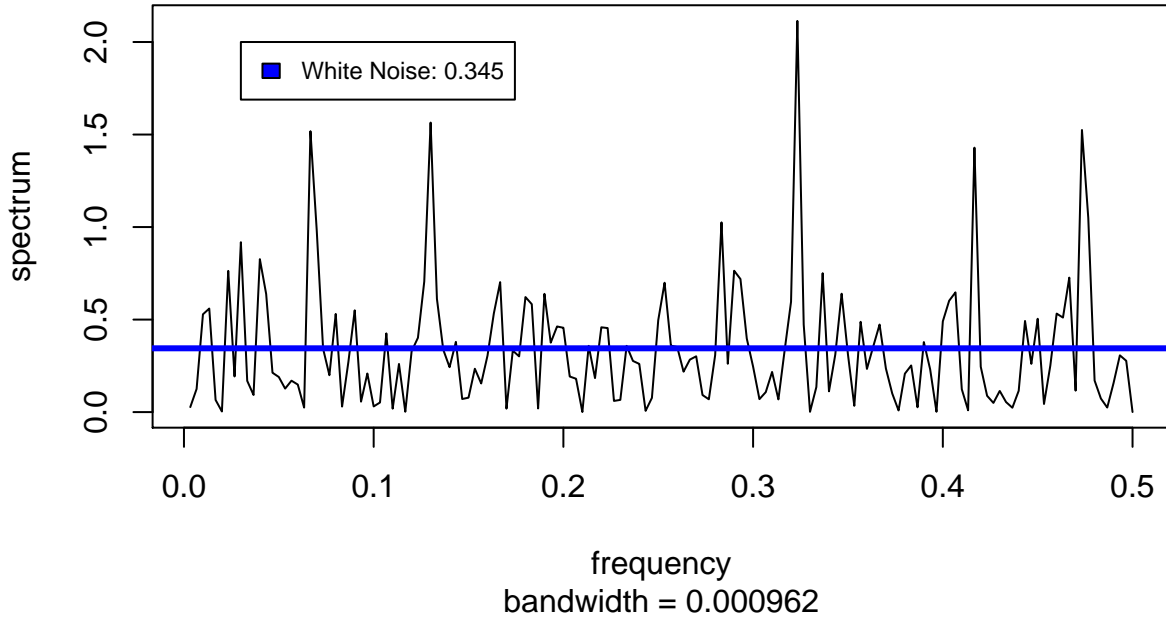
Above is the raw periodogram. I also plotted the white noise variance of the residuals, which represents a possible null hypothesis for the spectrum. Five frequencies immediately jump out at me as deviating from the null hypothesis. The results of the raw periodogram are displayed in the table below, along with 95% confidence intervals to test their significance. The confidence intervals are based on the chi-squared distribution. We used one sided confidence intervals since we are only interested in seeing if the lower bound of the interval is above the noise generated by the data. Furthermore, since we are considering 5 peaks as potential frequencies, we will use the bonferroni correction to solve the issue of multiple hypothesis testing. Thus, we will test each a 1% significance rate so that the family-wise significance rate is 5%.

From this initial analysis, we fail to see the longer periods that Benner discovered with his analysis. However, there is a lot of noise in the periodogram. And the wide bandwidth at the very small frequencies suggests that a very long period may exist, even if it is irregular.

```
##   Frequency Period Lower Bound Spectrum
## 1      0.13   7.50      0.3301     1.52
## 2      0.26   3.95      0.3387     1.56
## 3      0.65   1.55      0.4582     2.11
## 4      0.83   1.20      0.3105     1.43
## 5      0.95   1.06      0.3301     1.52
```

The 95% confidence interval lower bounds for these five frequencies are all not greater than the noise generated in the other frequencies. This leads us to caution against the idea that these frequencies represent any real periodicity captured in the data.

Claim 7

In his analysis, Benner (1999) tested the significance of the peaks of the periodogram using red noise significance levels. The idea behind red noise significance levels is that the null hypothesis of a constant spectrum (from white noise) is inappropriate for climate data. Instead, we should compare the periodogram to a null hypothesis of red noise, which states that lower frequencies will be more prevalent in the data.

In order to replicate Benner's work, I first needed to obtain estimates for the particular power spectrum that represents the red noise. That is I needed to estimate the parameters A and B in the equation I =

A * w ^ -B, where I is the value from the periodogram for a given frequency w. To do this, I transformed the periodogram into the log space and fit a linear model. After obtaining my estimate for A, I adjusted it to account for the bias in the periodogram introduced by taking the log. I was then able to construct my significance threshold by adding upper and lower bounds to the estimated null spectrum. Below, I plot the data from the periodogram in the log space, along with the upper and lower bounds. Notice how none of the peaks break the upper bound. This leads me to once again conflict with Benner's conclusions about the significance of these periodic components.

```
## (Intercept)      freq
##    -1.61612    0.04129
```

Logged Periodogram and Red Noise Confidence Bands
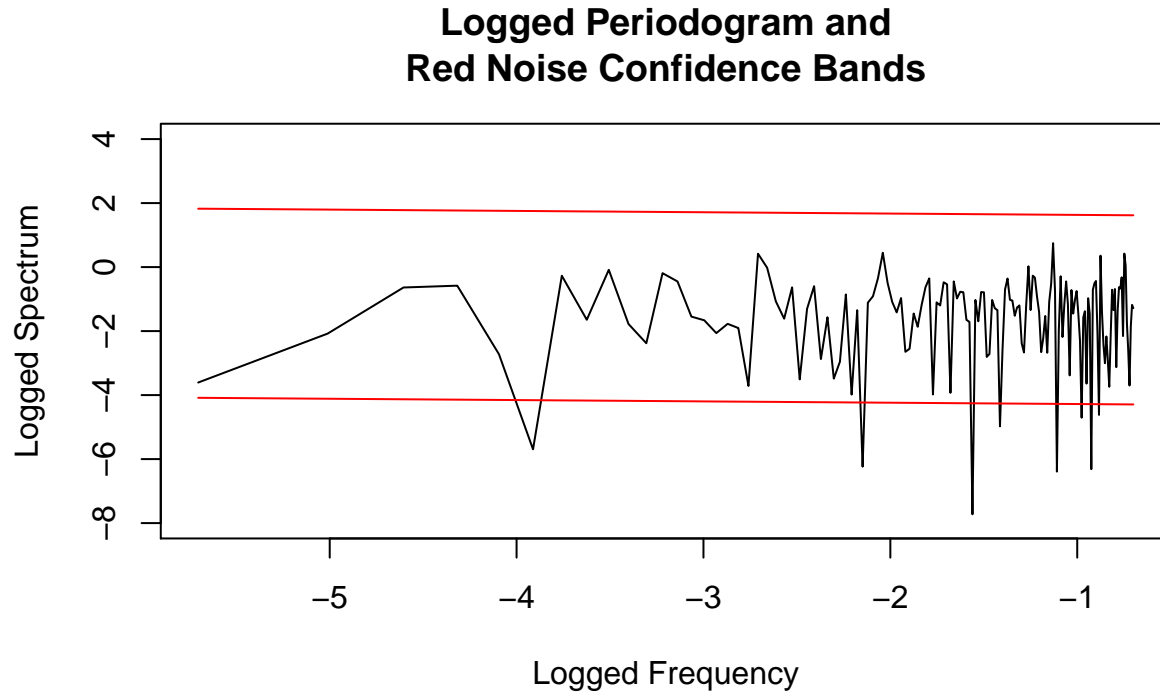
Smoothed Periodogram

While the periodogram does provide an unbiased estimator of the spectral density function, it can have large uncertainties because each estimate for a given frequency is merely the sum of squares of only two random variables (the cosine and sine components) for any sample size. The periodogram's variance does not decrease as the sample size increases.
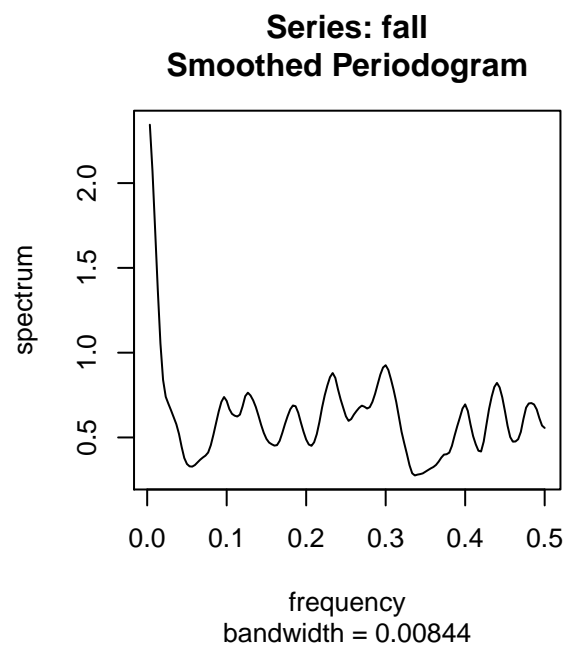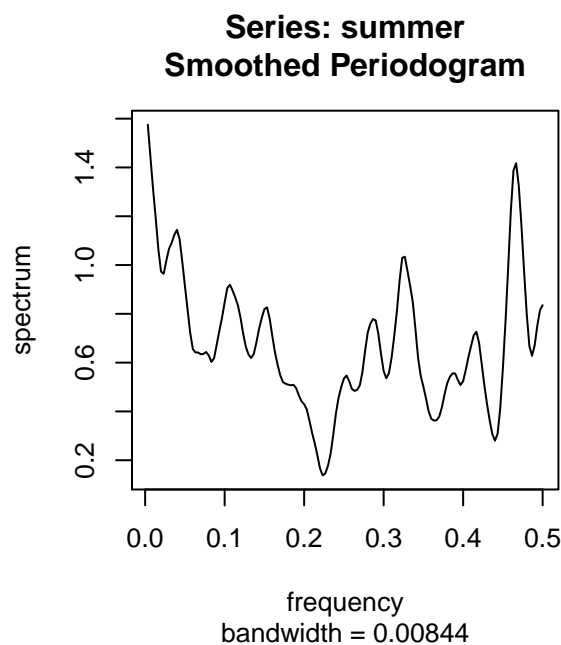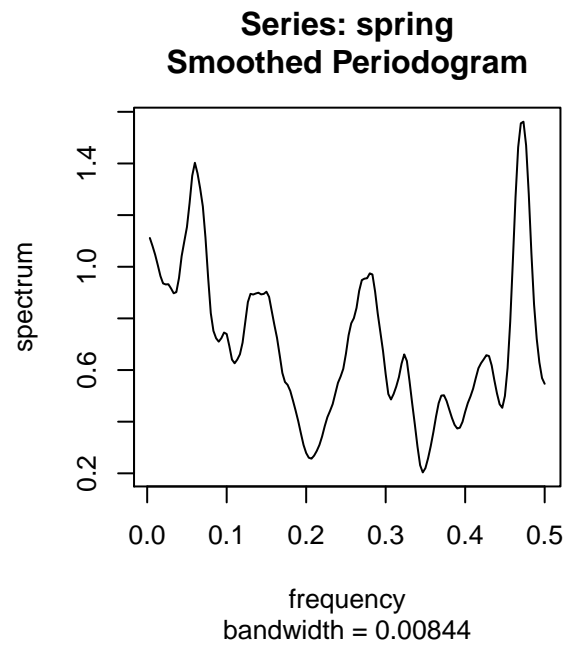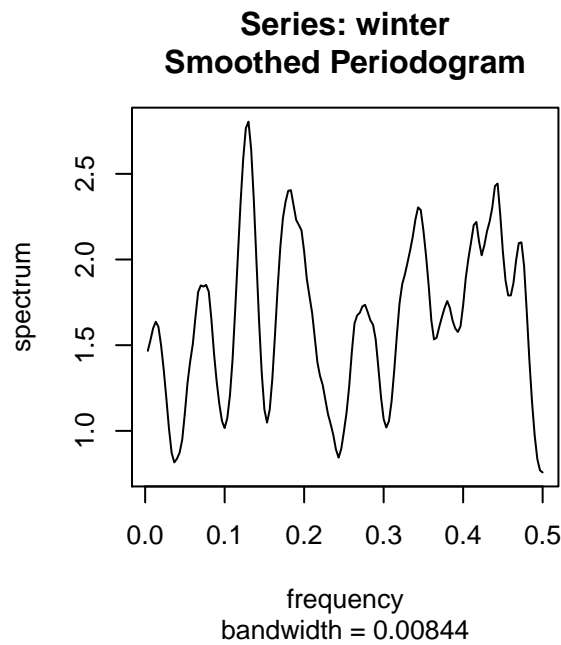
We can improve the variance by averaging the values for the periodogram over a frequency band, in order to create a smoothed estimate of the spectral density function. We used a modified daniell kernel with tapering over 20% of the series to create the smoothed periodogram below.

10

Series: Model_Residuals
Smoothed Periodogram

frequency
bandwidth = 0.00844
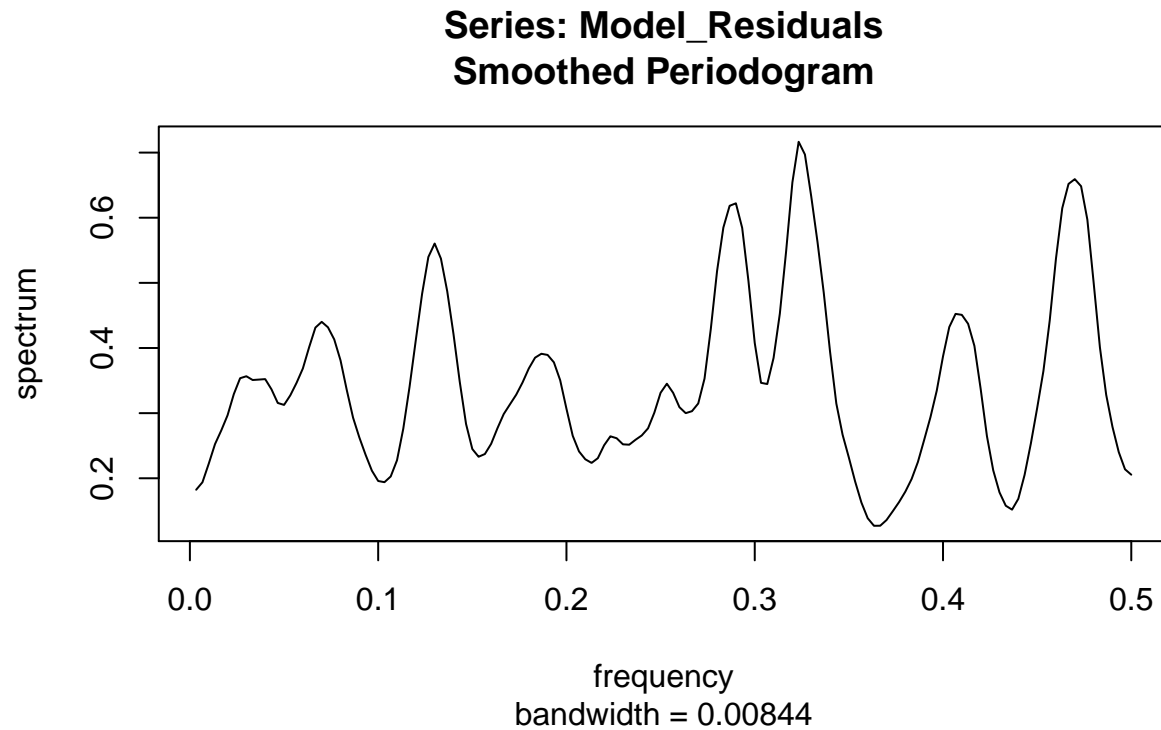
Although the peaks of the smoothed periodogram appear more clearly in this graph, they are no more significant than before. And there are certainly no peaks at the low frequencies that Benner claimed.

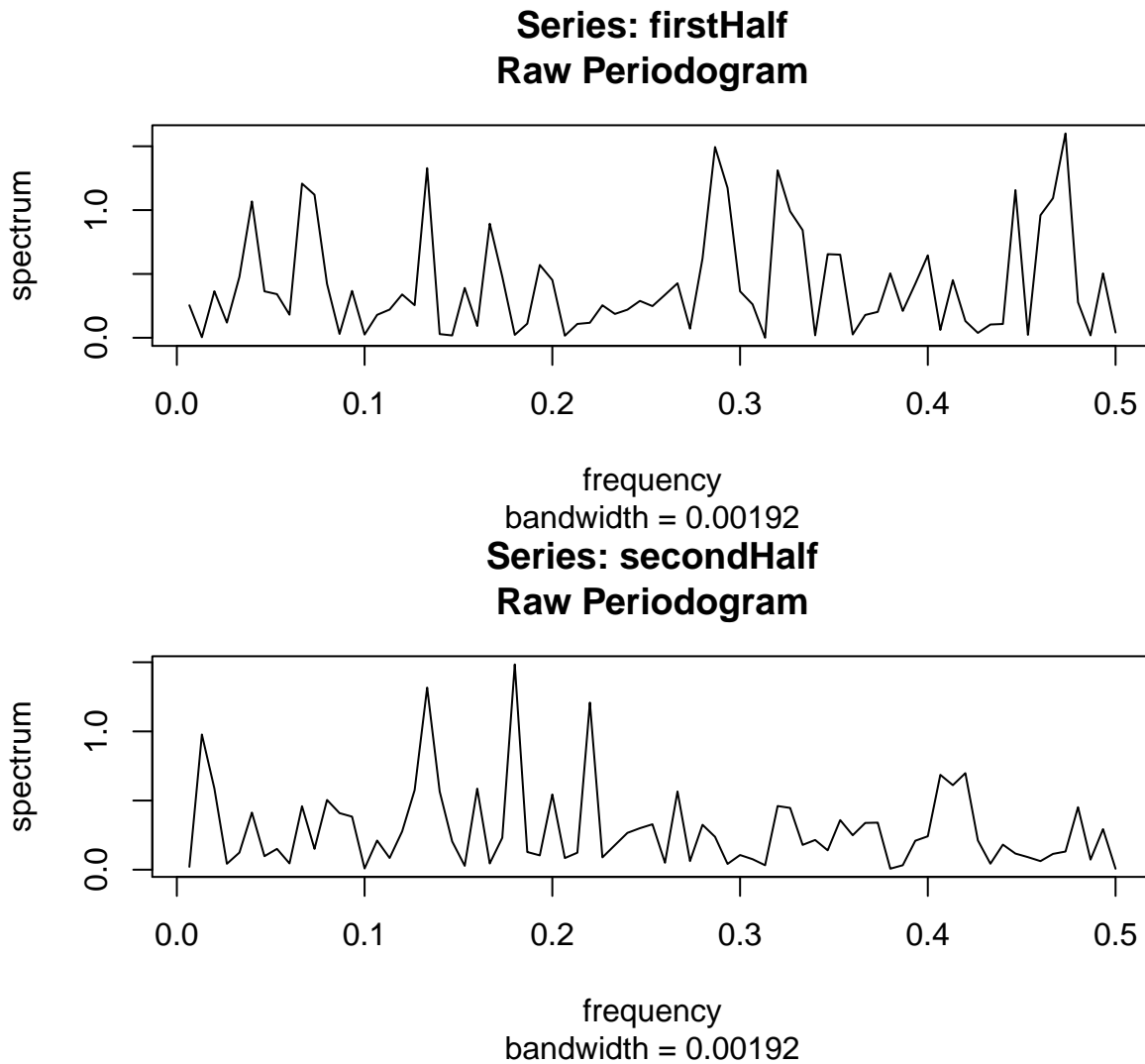**Seasonal Variability Investigations**

Jones and Bradley (1992a) claimed the winter series had the greatest variability while the summer series had the lowest variability. I can confirm these results as winter has a variance of 1.759 degrees, while summer has a variance of 0.672 degrees. Since we failed to find periodicity in the annual temperature series, one possible explanation is that the periodicity is hiding within the seasons. Below I plot the four smoothed periodograms for the four seasons.

Series: winter
Smoothed Periodogram

frequency
bandwidth = 0.00844

Series: spring
Smoothed Periodogram

frequency
bandwidth = 0.00844

Series: summer
Smoothed Periodogram

frequency
bandwidth = 0.00844

Series: fall
Smoothed Periodogram

frequency
bandwidth = 0.00844

While the summer and fall periodograms exhibit peaks at very low frequencies (which correspond to high periods), these peaks are not significant enough to reject the null hypothesis of white noise. The values of the specturm are still very small.

Claim 8

The last claim that I want to investigate is whether certain frequencies exist only in the first half of the record or only in the second half of the record. To investigate this claim, I will split the CET record at the year 1868 and perform a spectral analysis on both records. Once again, I use the residuals from the MA(2) plus quadratic model for the spectral analsysis.

Series: firstHalf
Raw Periodogram

frequency
bandwidth = 0.00192

Series: secondHalf
Raw Periodogram

frequency
bandwidth = 0.00192

Nothing significant jumps out at me in either periodogram. The peaks are still not significant.

Another effort

Disappointed that I was unable to discover any periodic signals in the data, I decided to see what happened to my models if I included a periodic component anyway. I looped through a large number of frequencies and found that if I included a cosine term with a period of just under 6 years, then the AIC of the resulting model is -0.0467. The AIC without the cosine term is -0.0339. The BIC values are about the same for both models. So there might be some evidence to support Benner's claim that a period of about 7-8 years exists in the data.

Conclusion

Although I was unable to confirm Benner's discovery of several significant periodic oscillations in the data, I was able to build a model that fit the annual temperature data well. The model that I chose captured the non-linearity evident in the CET record. From this model, we can see that annual temperatures are clearly trending up. One of the most important results of this paper is the finding that we can reject the hypothesis that the data is generated by a non-stationary random walk process. As seen in our rejection of the unit-root hypothesis, we found that representing the data as trend stationary is more appropriate. This allows us to say that the model we created is legitimate.