Modeling Long-term Central England Temperature Data

Philip Han
STAT 26100

March 8, 2016

1 Introduction

Earth's long-term climate variability has and continues to be of great interest to scientists and citizens alike. Of particular interest is the study of the phenomenon known as global warming and its possible causes, most notably anthropogenic greenhouse gas emissions. The rate at which the planet is warming and its underlying causes have been most recently questioned following a moderately controversial study by the The National Oceanic and Atmospheric Administration which concluded that global warming has not slowed down in the 2000's (Vaidyanathan 2016).

Long-term temperature data plays a vital role in our understanding of global warming. The central England temperature (CET) data, which dates back to 1659, is one of the longest standing temperature records. This paper will not seek to answer the cause of global warming nor its consequences. Rather, we will seek to model long-term temperatures from the central England temperature data set, and address several claims made by contemporary scientists related to long-term warming trends.

2 Modeling the Central England Temperature Record

2.1 Exploratory Data Analysis

The CET data that this paper uses reports seasonal averages of temperature data dating back to 1659. From a time series plot of the seasonal temperature data (Figure 1), it is evident that there is a high degree of variability and also cyclicality in the data. The most apparent cycle is due seasonal changes in temperature. However, given that global warming is a long term phenomenon, we would like to explore longer-term trends in the temperature data. Plotting the yearly average temperature gives us a cleaner view of the long term data, albeit still with high variability (Figure 2).

We would now like to explore regular or cyclical behavior that the temperature data exhibits. We will do this by examining the periodogram of the data. The periodogram of the raw seasonal data does not provide insight into long-term trends because it is dominated by the seasonal variation in temperature (Figure 3). That is to say, we see only one peak at frequency 0.25, which is indicative of the expected temperature oscillation that occurs with period $\frac{1}{0.25} = 4$ seasons.

The periodogram of the averaged annual data provides us additional insight. We observe the most significant peaks corresponding to small frequency values, suggesting that there are possibly

Figure 1: CET

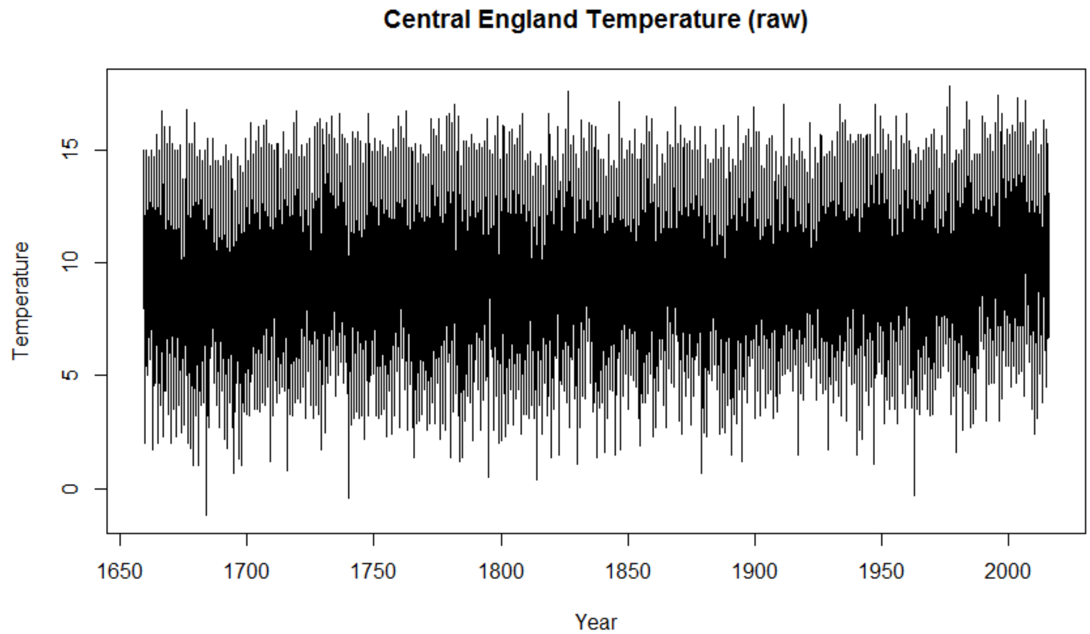
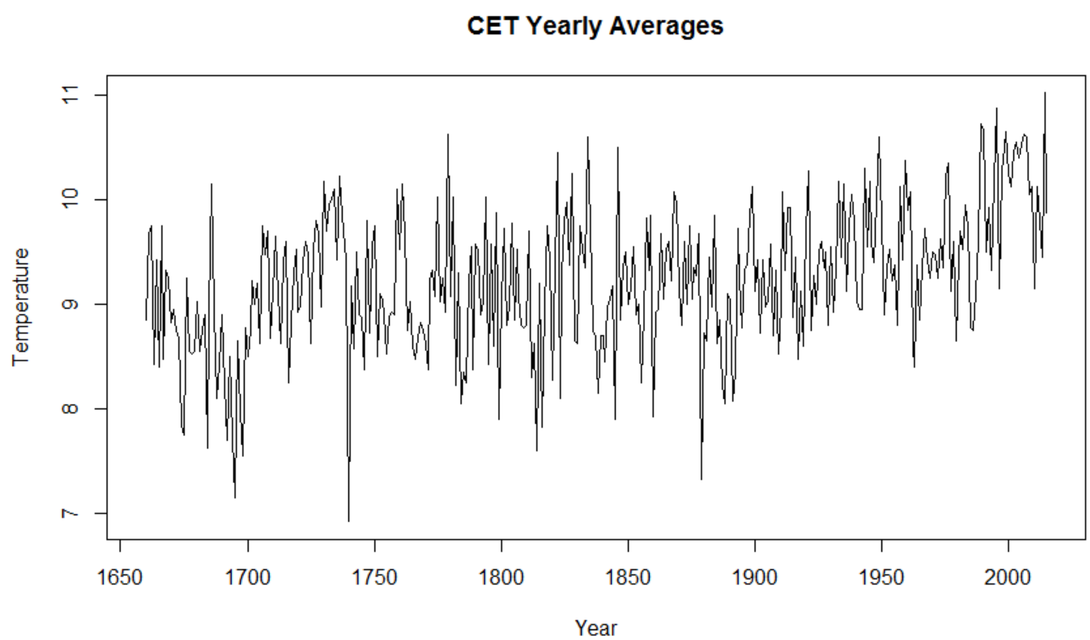


Figure 2: CET Yearly Averages



recurring patterns that occur on a longer time frame (i.e. decades to centuries). The highest peak occurs at frequency ≈ 0.04 which corresponds to a pattern that occurs every 25 years. There is quite a large band of significant activity, suggesting that the possible cyclical temperature patterns are irregular. However, these patterns need to be explored further.

We will conclude our discussion of periodicity by examining the ACF and PACF of the yearly averaged data. A cyclical pattern is fairly evident with oscillations in both the ACF and PACF over time (Figure 5).

Figure 3: Seasonal Periodogram

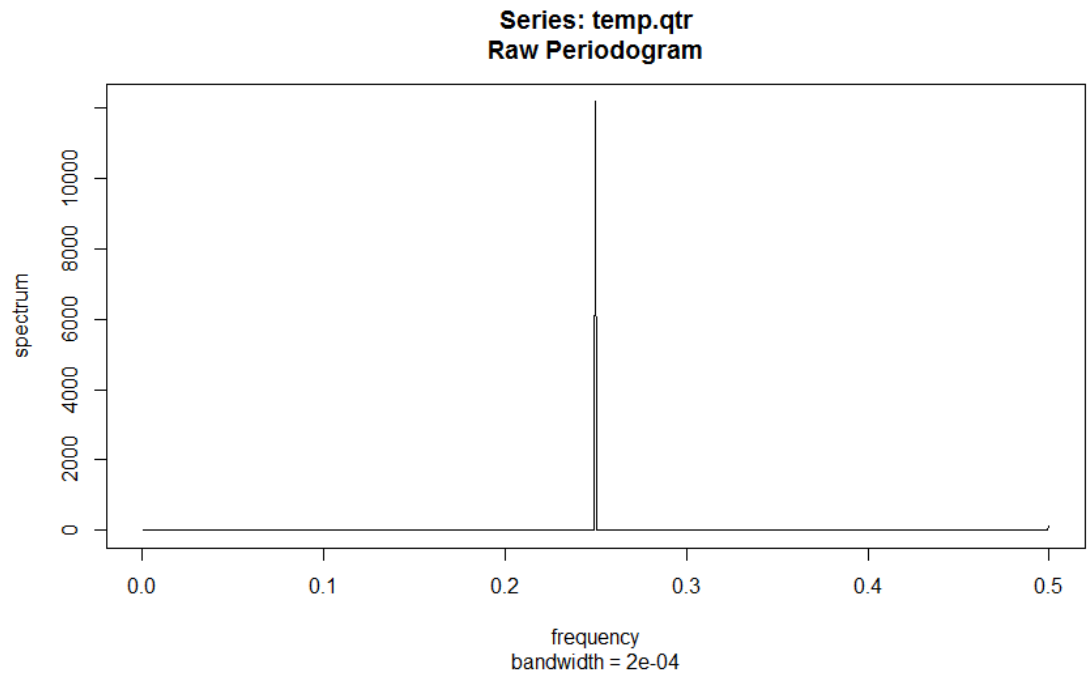


Figure 4: Yearly Average Periodogram

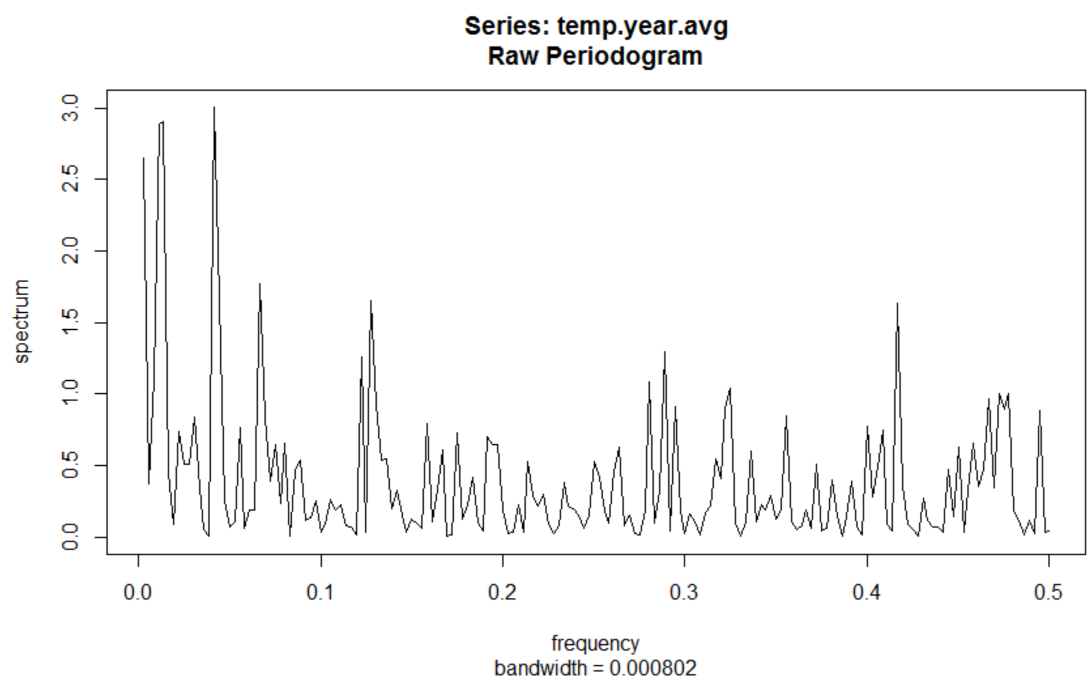
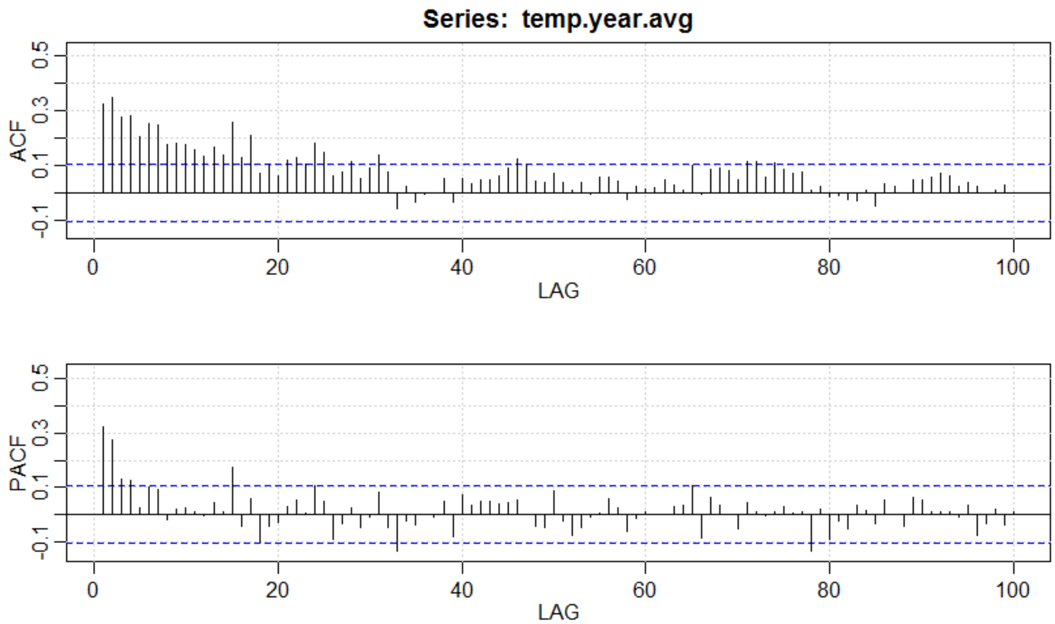


Figure 5: Yearly Average ACF/PACF



2.2 Building a Simple Model

From the periodograms, ACF, and PACF in the preceding section, we know that there may be long-term cyclical patterns of interest. The periodogram indicates that there may be patterns that occur over centuries; we will fit a fairly large AR model to accommodate these patterns in our long-term analysis. In particular, we will fit an AR(100) model on the averaged annual data. This simplistic model will enable us to conduct a rudimentary analysis of long-term temperature behavior.

We can validate our model by evaluating its predictive power. We exclude the n most recent years from the data and regress our model. Then, using our regression, we predict the values of the temperatures in the n most recent years and compare those to the actual data. Based on the actual temperatures (lines) vs. the predicted temperatures (dashed lines) in Figures 6-8, we can see that the model exhibits moderate predictive power despite its simplicity. With a regression model this large, not all terms will be significant. However, the residual plot appears to be random, so the model does not appear to be invalidated.

3 Evaluating Claims by Other Authors

We will now seek to evaluate several claims made by by Jones, Vaidyanthan, and Benner.

(i) Long-term Warming Trend
Jones claims that there was hardly any warming between the years 1700 and 1980, whereas authors cited by Vaidyanthan and Benner both argue that the warming trend is real. We will evaluate this claim with a simple kernel smoothing of the temperature data to observe long-term trends. Figure 10, which was obtained by smoothing the yearly average temperature data with a bandwidth of 10, appears to show a warming trend from 1700 to 1980.

Figure 6: Predicting 5 Most Recent Years

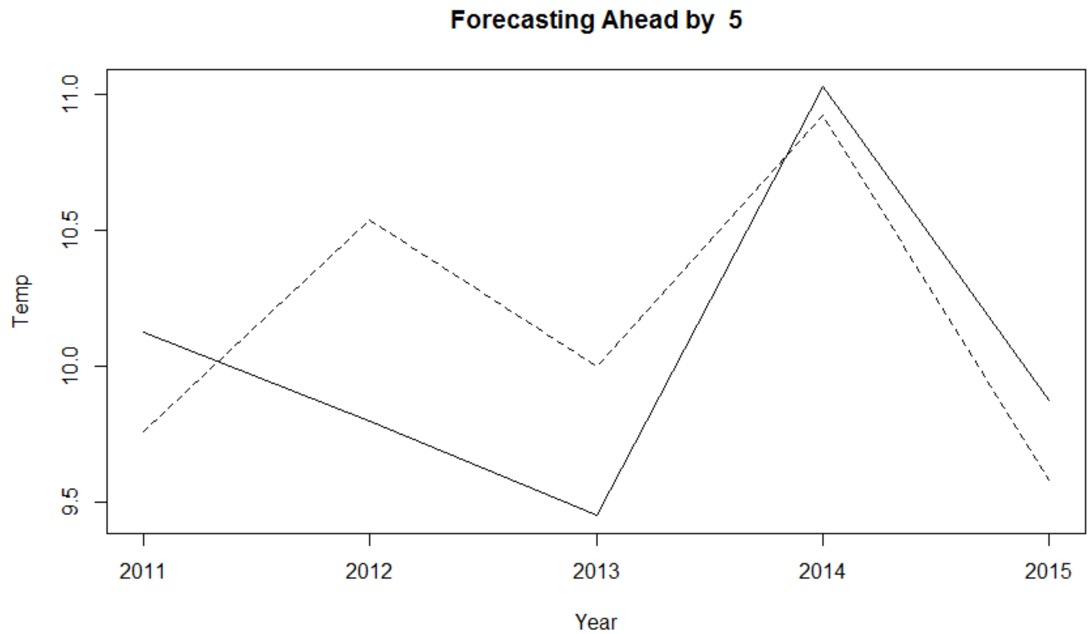
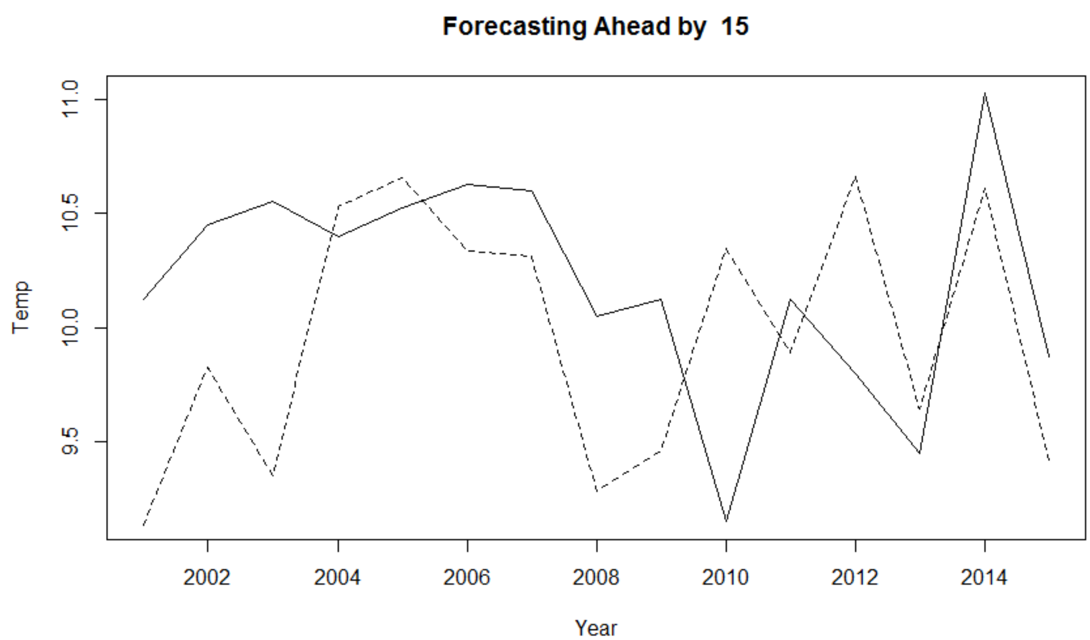


Figure 7: Predicting 15 Most Recent Years



(ii) Oscillations

Jones claims that decadal warming/cooling is very present in the data. Vaidyanthan cites authors who argue that oscillations are a possibility, whereas Benner argues that oscillations seem to be present but don't persist and are non-stationary. This claim is difficult to validate/invalidate without additional information. In particular, it would be insightful to compare temperature data with the purported cause of the oscillation (for instance, compare

Figure 8: Predicting 50 Most Recent Years

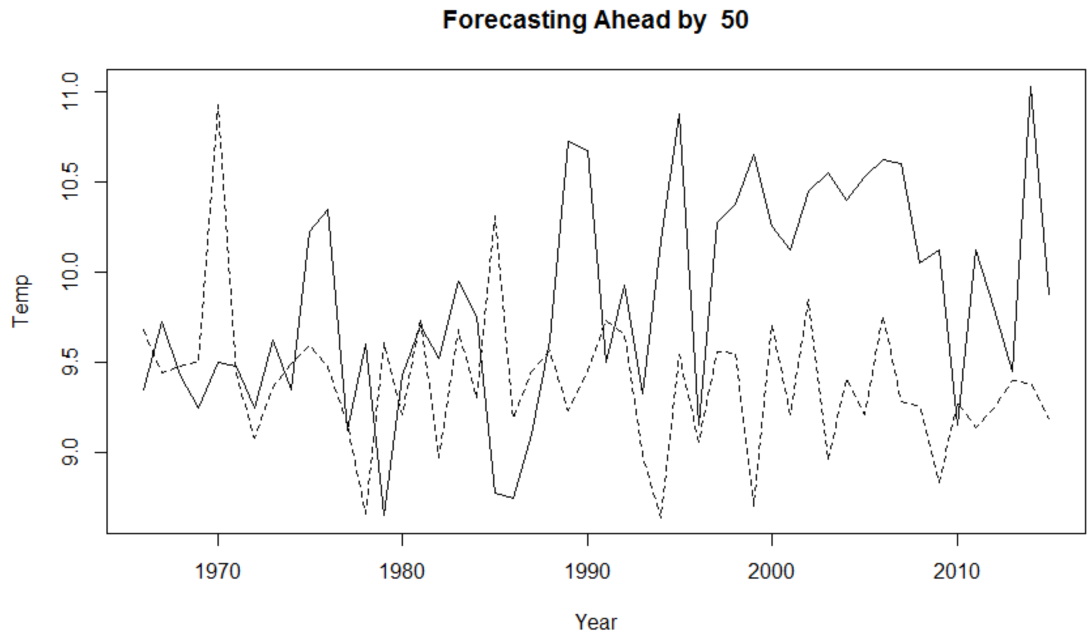
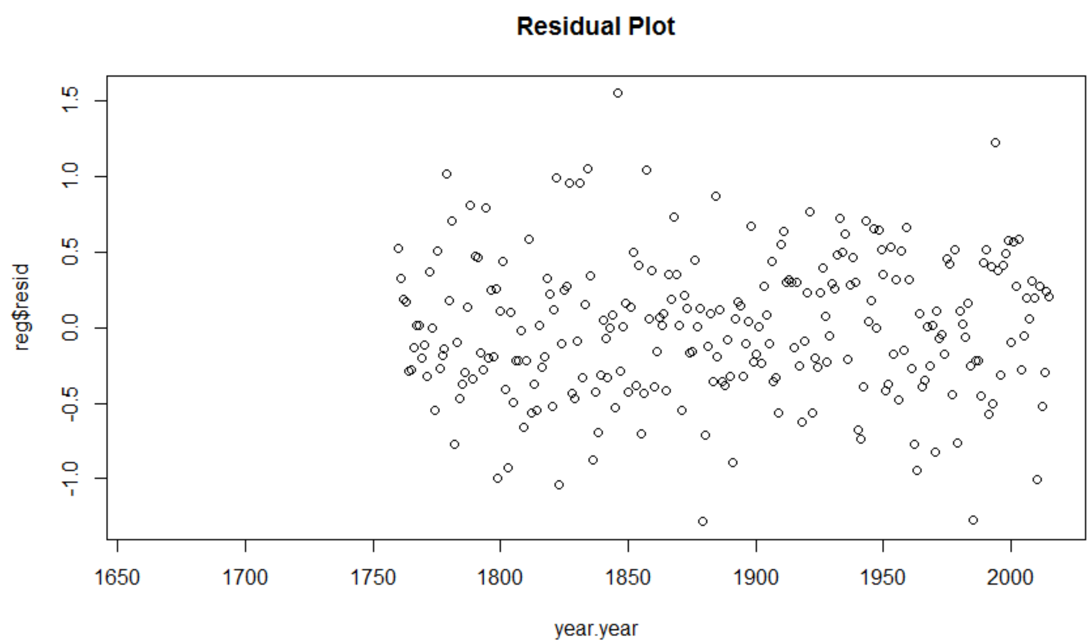
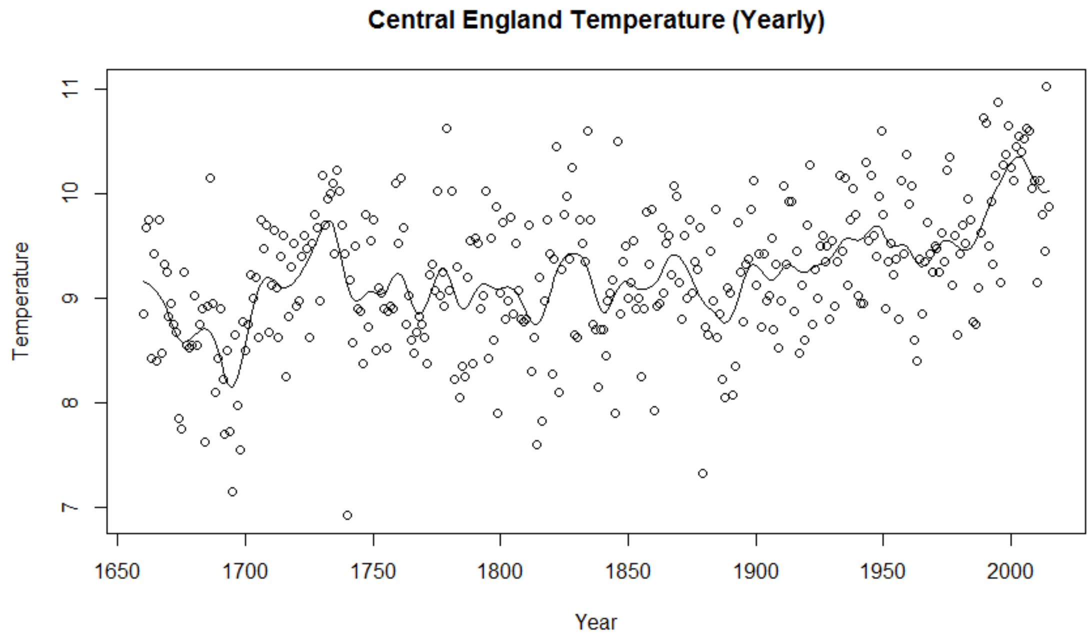


Figure 9: Residual Plot from Regression



oceanic temperature data with surface temperature data). Benner cites instances of teleconnections such as solar oscillations and oceanic effects. The periodgram and ACF analysis conducted prior seems to indicate that oscillations do exist (see Figures 4 and 5). However, their significance is difficult to determine.

Figure 10: Long-term Warming Trend



(iii) Warming Growth Rate

Vaidyanthan cites authors that represent both sides of this issue. The NOAA in particular claims that global warming has not slowed down in the 2000s while another group of scientists believes this to be the case. The analysis done here tends to support the claim that global warming has not slowed down. Looking at Figure 7, the previously constructed AR model predicts the past 15 years with a fair degree of accuracy. The regression presupposes the previous rate of warming (because it's regressed on all previous data), so if we suppose our model to be true, we can conclude that actual temperatures are in line with the historical rate of warming. Of course, this requires a specific definition of what the rate of warming actually consists of. Here, we simply claim that if global warming was happening at some given rate (which the model takes into account), and the model predicts the temperatures from the past 15 years fairly well, then the model parameters have not significantly changed.

4 Conclusion

In our study of the central England temperature dataset, we created an AR model with 100 lag terms. We found that although the model fit criteria were not met particularly well, the model did an acceptable job in terms of prediction. Based on our model predictions as well as analysis of the temperature dataset, we find that there appears to be an overall warming trend. Oscillations are present, and quantifying their significance requires additional analysis. Finally, the growth rate does not appear to slow down in the 2000s. Further analysis could include a model that took into account seasonal and intra-year variation, additional smoothing techniques, and explicitly modeling plausible oscillations in the temperature data.