STAT 26100: Time Series Analysis on the Central England Temperature

Caroline Jiang

12/7/2022

```
library(tidyr)
library(astsa)
```

```
## Warning: package 'astsa' was built under R version 4.1.3
```

1. Introduction

The Central England Temperature (CET) record measures the monthly mean surface air temperatures of the midlands region of England in degrees Celsius from the year 1659 to the present. It was originally published by Professor Gordon Manley in 1953.

This record is the longest series of monthly temperature observations in existence. It is, therefore, an invaluable data set for climate scientists to study long-term temperature change. As global warming becomes a topics concerning all human beings in the modern world, I hope to gain insights on how temperature has changes from the 17 century to the 21 century, using the Central England Temperatures as my research sample.

I utilized a number of time series analysis techniques in this paper. First, I focused on the annual average temperatures. In this process, I fit both linear regression models and quadratic models to identify the general trend of how the temperature has changed. Second, I divided the time series into smaller sub-series to identify different trends on different dimensions, specifically centurial variation and seasonal variation. Both linear models and quadratic models were included in this process. Lastly, I conducted spectral analysis to identify periodical patterns within this data set, and I studied both overall oscillations and oscillations by centuries.

2. General Trend

2.1 Linear Model

I first trimmed the original data set so that it contains the annual average temperatures from year 1659 to 2021. The 2022 temperatures have not been fully collected, and therefore I excluded the year 2022. I then conducted a 5-year Kernel Smoothing over the raw data in order to wash out the smaller oscillations while preserving the overall trend.

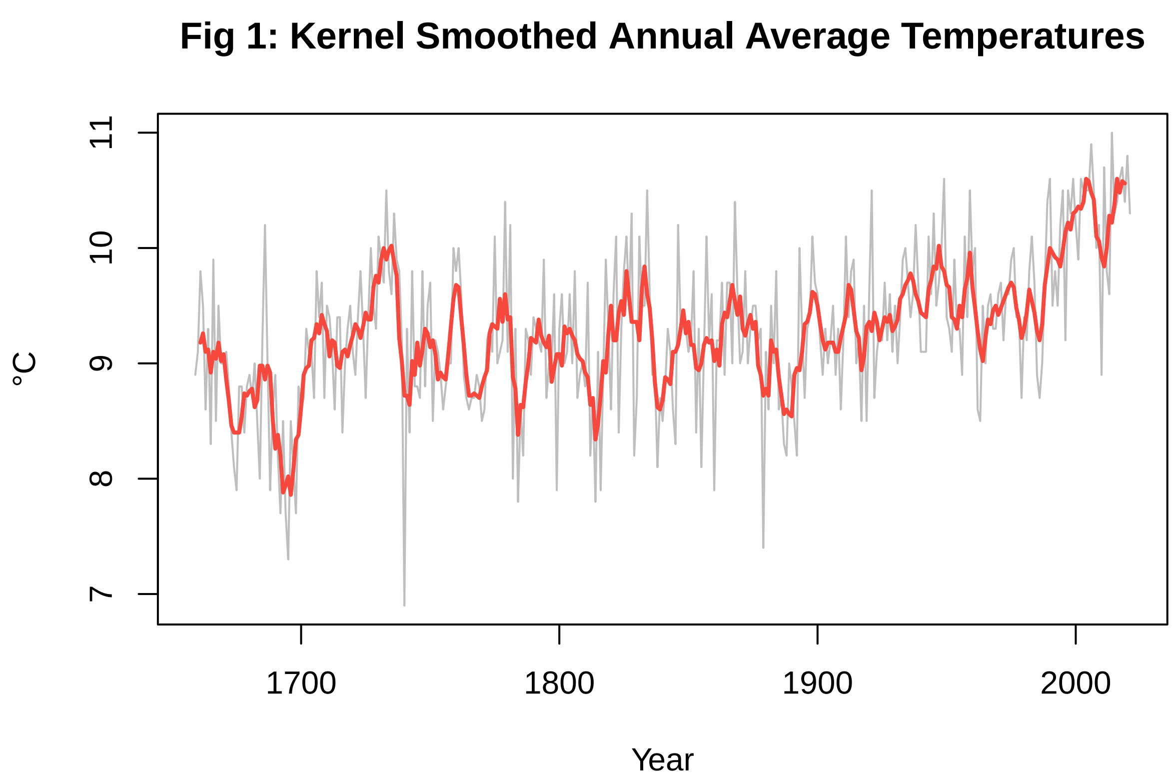
Looking at Fig 1, it is easy to identify an upward trend in the annual average temperatures. One can also notice a significant drop in the temperature around year 1700, and a significant peak in the temperetaure around year 2000.

```
#Data loading and cleaning
df <- read.table("meantemp_monthly_totals.txt", header = TRUE)
cet_ann = df[, c("Year", "Annual")]
head(cet_ann)
```

```
##   Year Annual
## 1 1659    8.9
## 2 1660    9.1
## 3 1661    9.8
## 4 1662    9.5
## 5 1663    8.6
## 6 1664    9.3
```

```r
cet_ann <- ts(cet_ann$Annual, frequency=1, start = 1659, end = 2021)
```

```r
#Kernel smoothing over 5 years
cet_ann_ma <- stats::filter(cet_ann, filter = rep(1 / 5, 5), sides = 2)
plot(
  cet_ann,
  col = "grey",
  xlab = "Year",
  ylab = "°C",
  main = "Fig 1: Kernel Smoothed Annual Average Temperatures"
)
lines(cet_ann_ma, col = 2, lwd = 2)
```



I then fitted a simple linear regression model to the data set. The model can be written as:

$$T = \beta_0 + \beta_1 t + w_t$$

After model fitting, the model becomes:

$$T = 3.95074 + 0.00288t + w_t$$

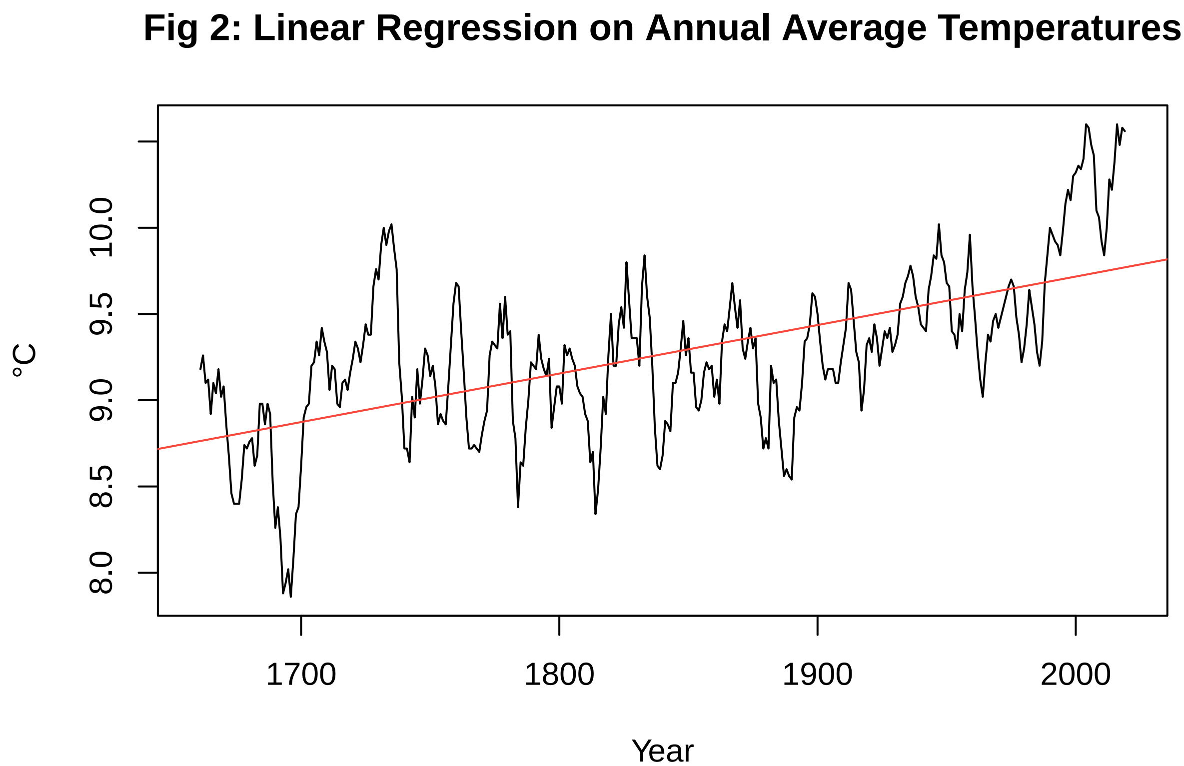
The confidence interval of β_1 is [0.00243, 0.00312] with a small p-value. This shows a upward trend in the time series. The temperature is getting about 0.29 degree Celsius higher each century, and has risen about 1.05 degree Celsius from 1659 to 2021.

But when looking at figure 3, which is the residual plot of the linear model, I realized the residuals didn't distribute randomly around 0. Instead, they form a convex curve especially from year 1750 to year 2021. Therefore, I decided to fit a quadratic model to improve model performance.

```r
#Linear model for the overall trend
year = time(cet_ann_ma)
fit = lm(cet_ann_ma ~ 1 + year, na.action = na.exclude)
summary(fit)
```

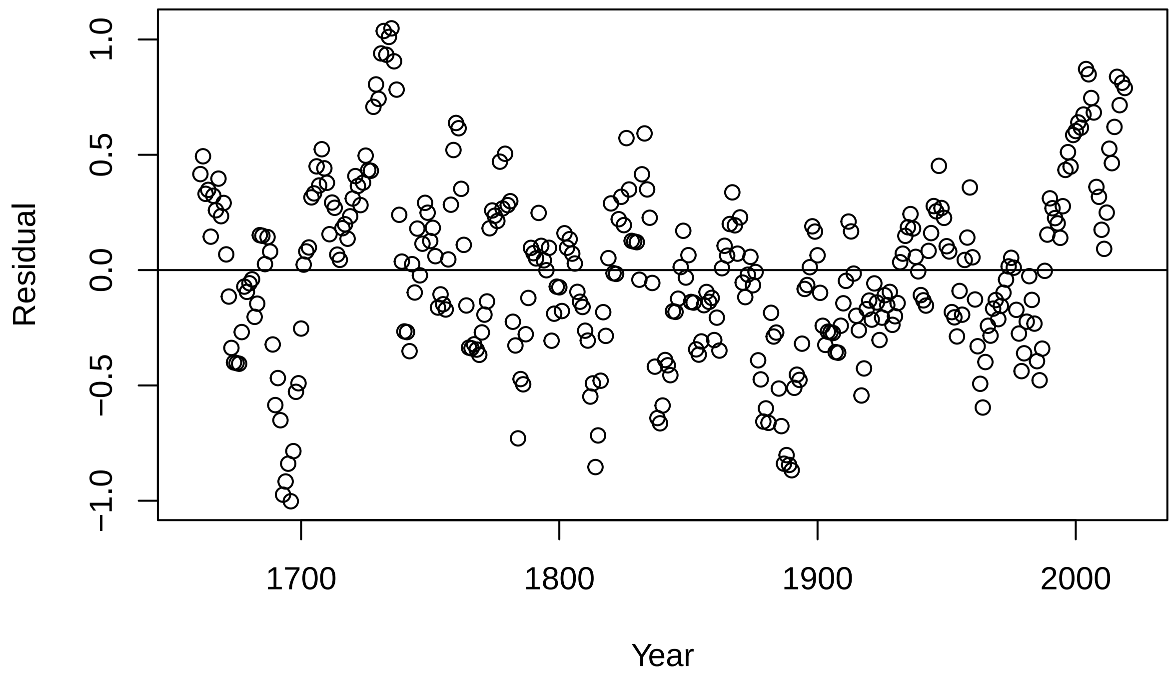
```
##
## Call:
## lm(formula = cet_ann_ma ~ 1 + year, na.action = na.exclude)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.00213 -0.25709 -0.00547  0.23475  1.04819
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 4.0928215  0.3550614   11.53   <2e-16 ***
## year        0.0028121  0.0001927   14.60   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3783 on 357 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.3737, Adjusted R-squared:  0.372
## F-statistic:   213 on 1 and 357 DF,  p-value: < 2.2e-16
```

```r
plot(cet_ann_ma,
     xlab = "Year",
     ylab = "°C",
     main = "Fig 2: Linear Regression on Annual Average Temperatures")
abline(fit, col = 2)
```




```
plot(
  x = time(cet_ann_ma),
  y = resid(fit),
  xlab = "Year",
  ylab = "Residual",
  main = "Fig 3: Residual for Linear Regression Model"
)
abline(0, 0, col = 1)
```

Fig 3: Residual for Linear Regression Model



2.2 Quadratic Model

I constructed the following quadratic model for the data set:

$$T = \beta_0 + \beta_1 t + \beta_2 t^2 + w_t$$

After model fitting, I get:

$$T = 46.25 - 3.977 * 10^{-2} t + 1.157 * 10^{-5} t^2 + w_t$$

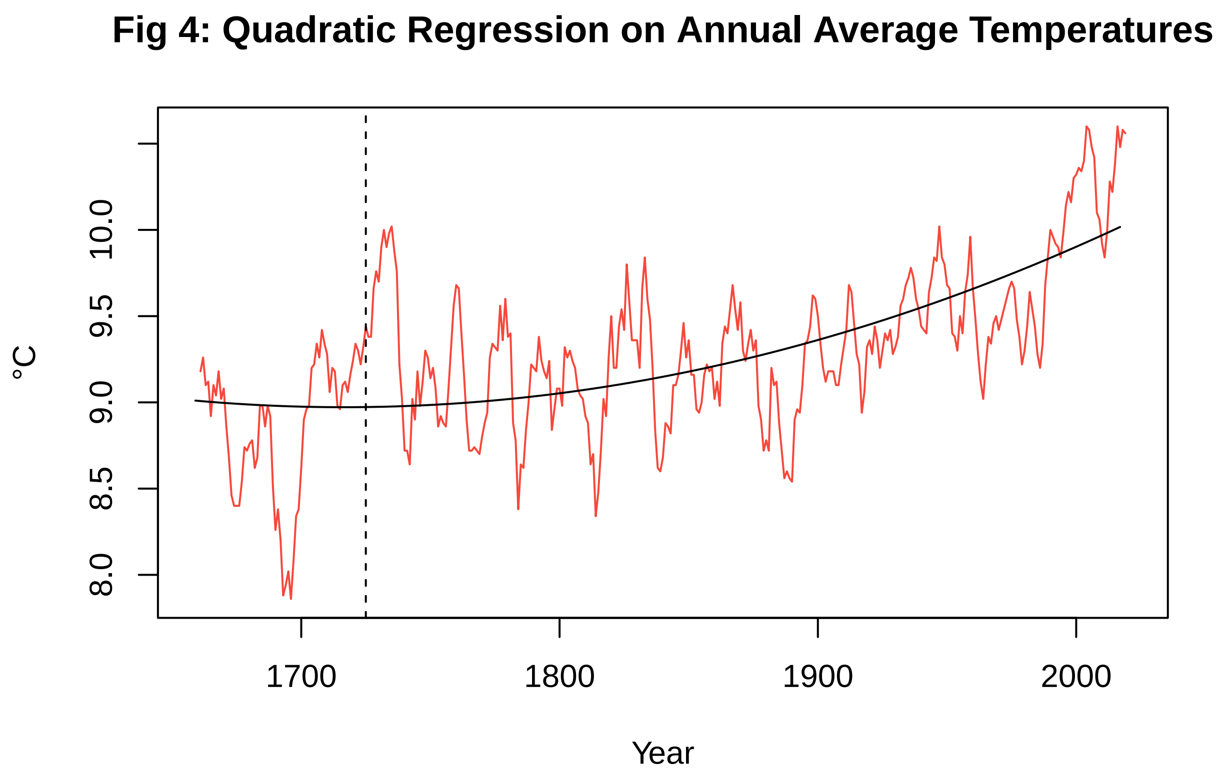
The p-values of both coefficients are small. Figure 4 shows that the turning point of the model is at year 1725. Temperature is dropping from year 1659 to year 1725, but it begins to rise after year 1725. The residual plot of the quadratic model (figure 5) displays a more random pattern then the residual plot of the linear model, which shows that the quadratic model is a better fit for the data.

```
#Quadratic model for the overall trend
yearsq = time(cet_ann_ma) ^ 2
fit2 <- lm(cet_ann_ma ~ 1 + year + yearsq, na.action = na.exclude)
summary(fit2)
```

```
##
## Call:
## lm(formula = cet_ann_ma ~ 1 + year + yearsq, na.action = na.exclude)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.11780 -0.22836  0.00527  0.24050  1.04490
##
```

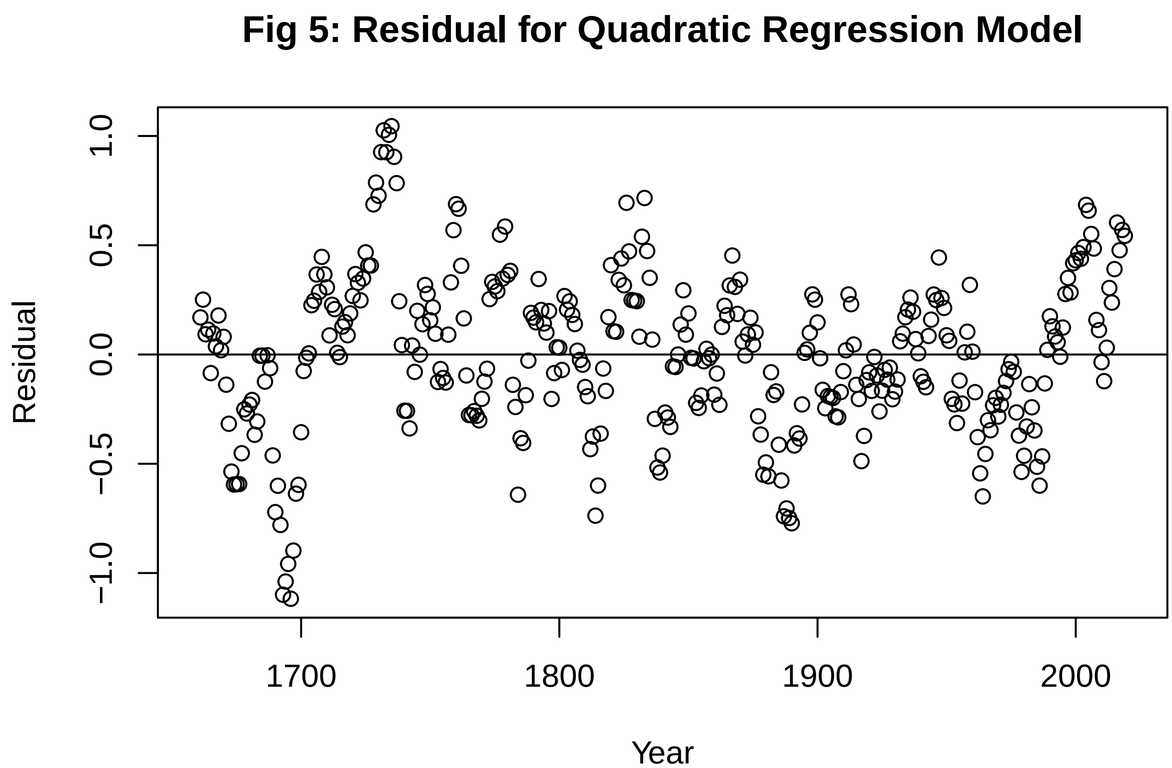
```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.315e+01  6.721e+00   6.419 4.38e-10 ***
## year        -3.977e-02  7.322e-03  -5.432 1.04e-07 ***
## yearsq       1.157e-05  1.989e-06   5.818 1.33e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.362 on 356 degrees of freedom
##   (4 observations deleted due to missingness)
## Multiple R-squared:  0.4281, Adjusted R-squared:  0.4249
## F-statistic: 133.2 on 2 and 356 DF,  p-value: < 2.2e-16
```

```
plot(
  cet_ann_ma,
  col = 2,
  xlab = "Year",
  ylab = "°C",
  main = "Fig 4: Quadratic Regression on Annual Average Temperatures"
)
lines(ts(fit2$fitted.values, start = 1659), type = "l")
abline(v = 1725, lty = 2)
```




```
plot(
  x = time(cet_ann_ma),
  y = resid(fit2),
  xlab = "Year",
```

```
  ylab = "Residual",
  main = "Fig 5: Residual for Quadratic Regression Model"
)
abline(0, 0)
```



2.3 ARIMA Model

To further improve model fitting, I fitted an ARIMA model to the CET data on hand. I first studied ACF and PACF plots to determine what model I should use. Figure 6 and figure 7 show that neither ACF nor PACF of the data set cut off after a certain lag. On the other hand, Figure 8 and figure 9 show that when CET data is differenced, its ACF cuts off after lag 1, which indicates the choice of an ARIMA model that has a moving average component. Therefore I fitted the model ARIMA(0, 1, 1). The model can be written as:

$$x_t = x_{t-1} + w_t + \phi w_{t-1}$$

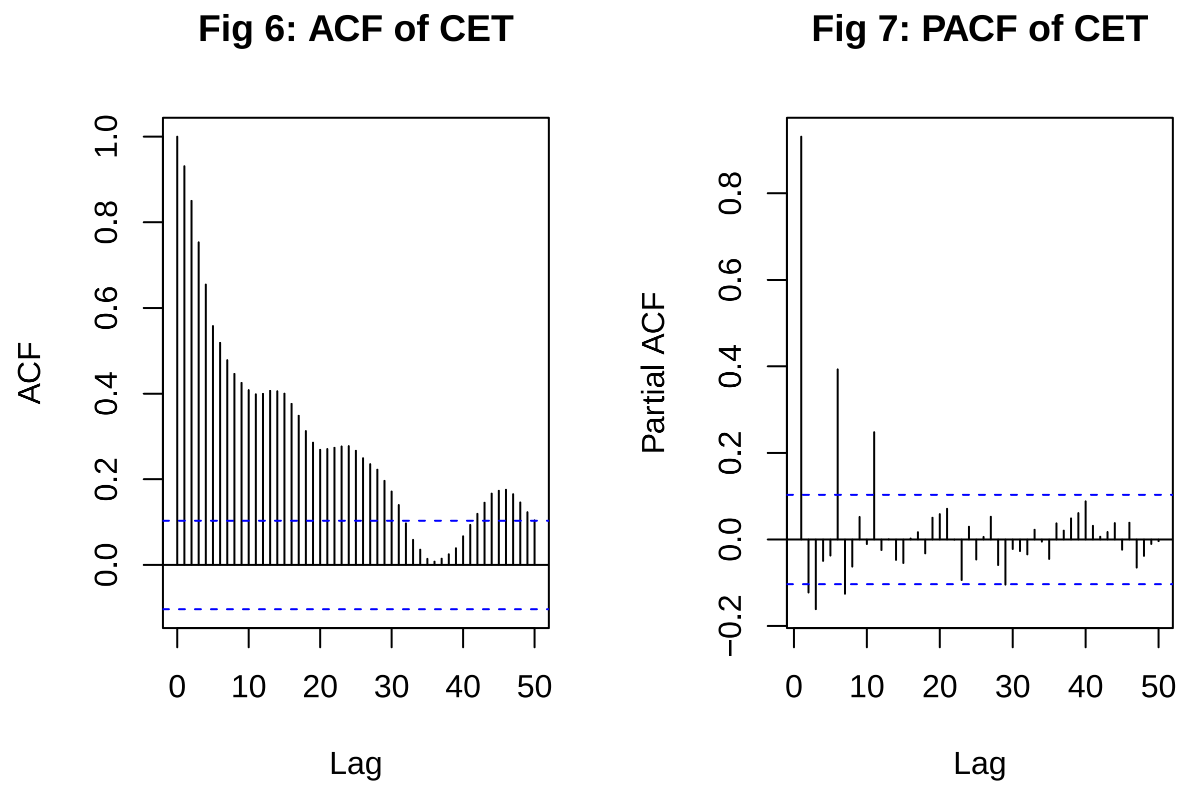
The model fitting renders a ϕ of 0.074 and a standard error of 0.046. Therefore the fitted ARIMA(0, 1, 1) model is

$$x_t = x_{t-1} + w_t + 0.074 w_{t-1}$$

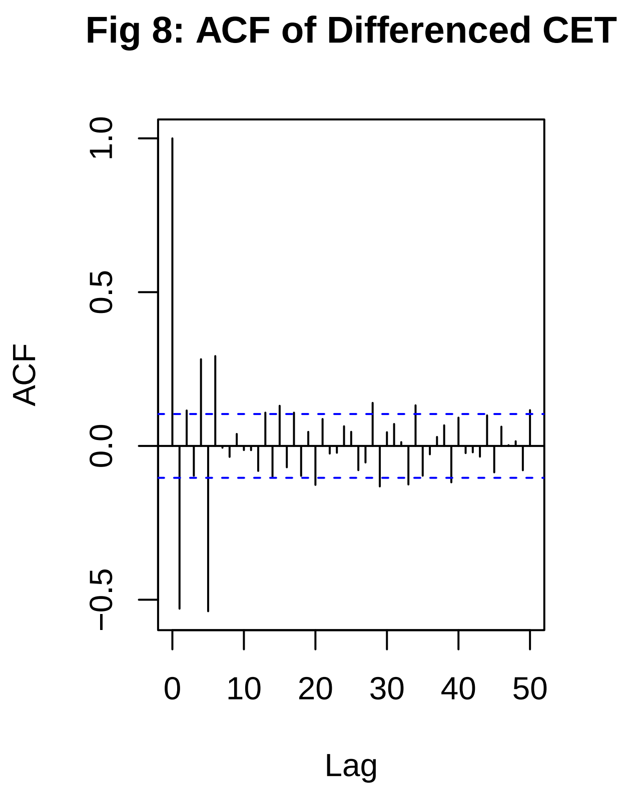
The residuals given in figure 10 are normally distributed, which is a sign of good model fitting. Another measure of the performance of a model is the Akaike Information Criterion. The AIC score for this model is -278.3642, calculated using the AIC function.

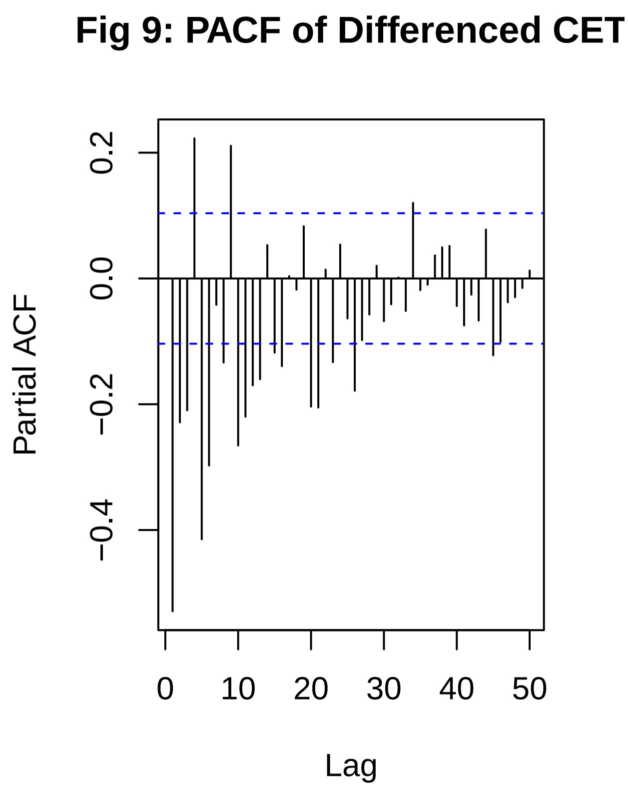
```
#ACF and APCF check
par(mfrow = (c(1, 2)))
acf(cet_ann_ma,
    lag.max = 50,
```

```
    na.action = na.exclude,
    main = "Fig 6: ACF of CET")
pacf(cet_ann_ma,
     lag.max = 50,
     na.action = na.exclude,
     main = "Fig 7: PACF of CET")
```




```
acf(diff(diff(cet_ann_ma)),
    lag.max = 50,
    na.action = na.exclude,
    main = "Fig 8: ACF of Differenced CET")
pacf(diff(diff(cet_ann_ma)),
     lag.max = 50,
     na.action = na.exclude,
     main = "Fig 9: PACF of Differenced CET")
```

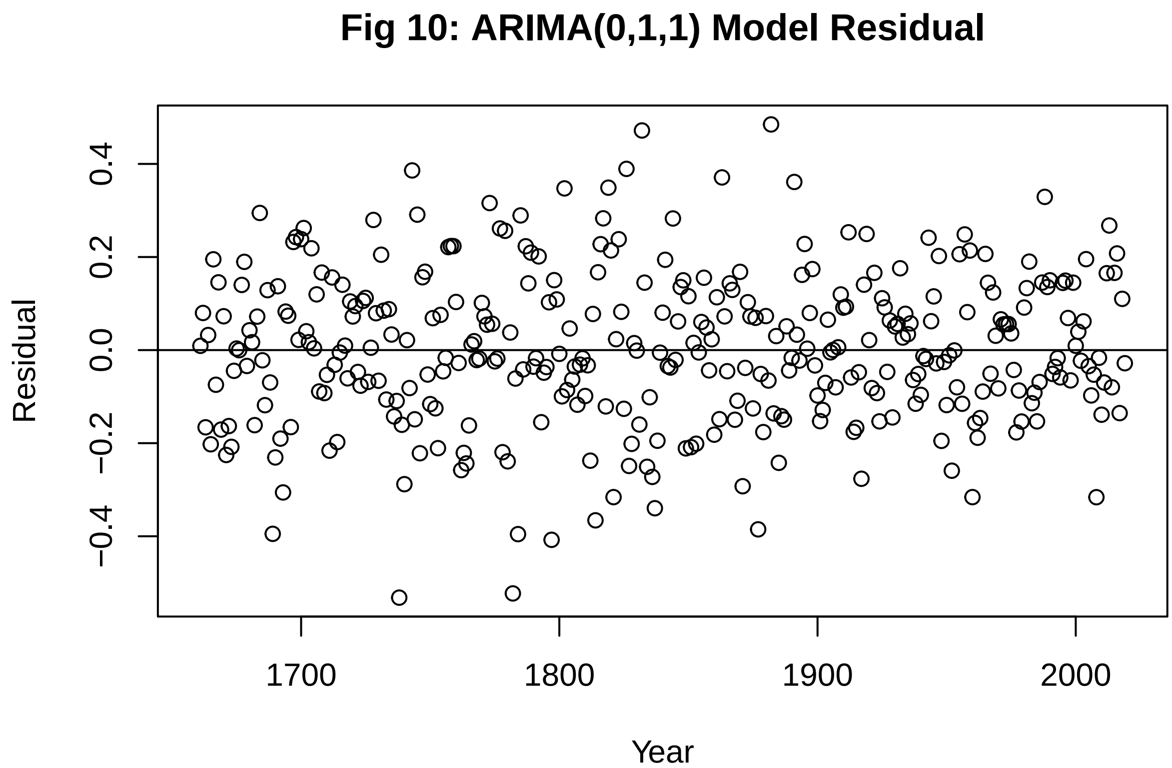





```r
#ARIMA model
arima(cet_ann_ma, order = c(0, 1, 1))
```

```
##
## Call:
## arima(x = cet_ann_ma, order = c(0, 1, 1))
##
## Coefficients:
##          ma1
##       0.0743
## s.e.  0.0460
##
## sigma^2 estimated as 0.02661:  log likelihood = 141.18,  aic = -278.36
```

```r
plot(
  arima(cet_ann_ma, order = c(0, 1, 1))$residuals,
  type = "p",
  ylab = "Residual",
  xlab = "Year",
  main = "Fig 10: ARIMA(0,1,1) Model Residual"
)
abline(h = 0)
```




```
#AIC score
aic1 = AIC(arima(cet_ann_ma, order = c(0, 1, 1), method = "ML"))
aic1
```

```
## [1] -278.3642
```

3. Specified Trends

3.1 Trends by Centuries

I divided the data set into several smaller pieces to examine the temperature trends by centuries. Specifically, I divided the time series into the following time periods: the 18th century, the 19th century, and the 20th century. I then fitted both linear regression and quadratic regression to all three centuries.

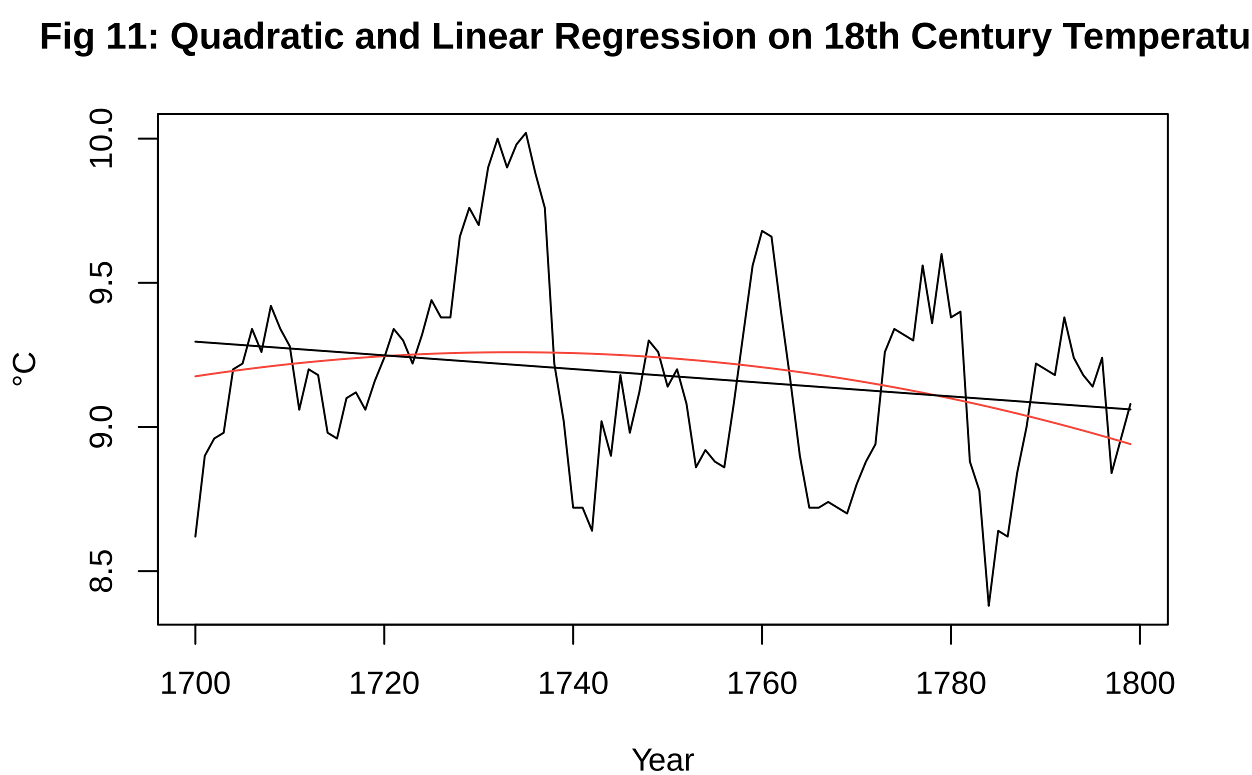
It is clear from figure 4, figure 5 and figure 6 that in the 18th and 19th century, the temperature change is relatively smooth while in the 20th century, the temperature rises drastically. After excluding the temperature drop just before year 1700, there is a downward trend in the 18th century. The trend in the 19th century is almost flat, but in the 20th century, there is a clear upward trend.

The rapid raise in temperature could be a result of the completion of the Second Industrial Revolution at the beginning of the 20th century. During the Second Industrial Revolution, the expansion of rails and the wide-spread usage of electricity all lead to an increase in the carbon-dioxide level in the atmosphere, which in turn leads to a warming weather.

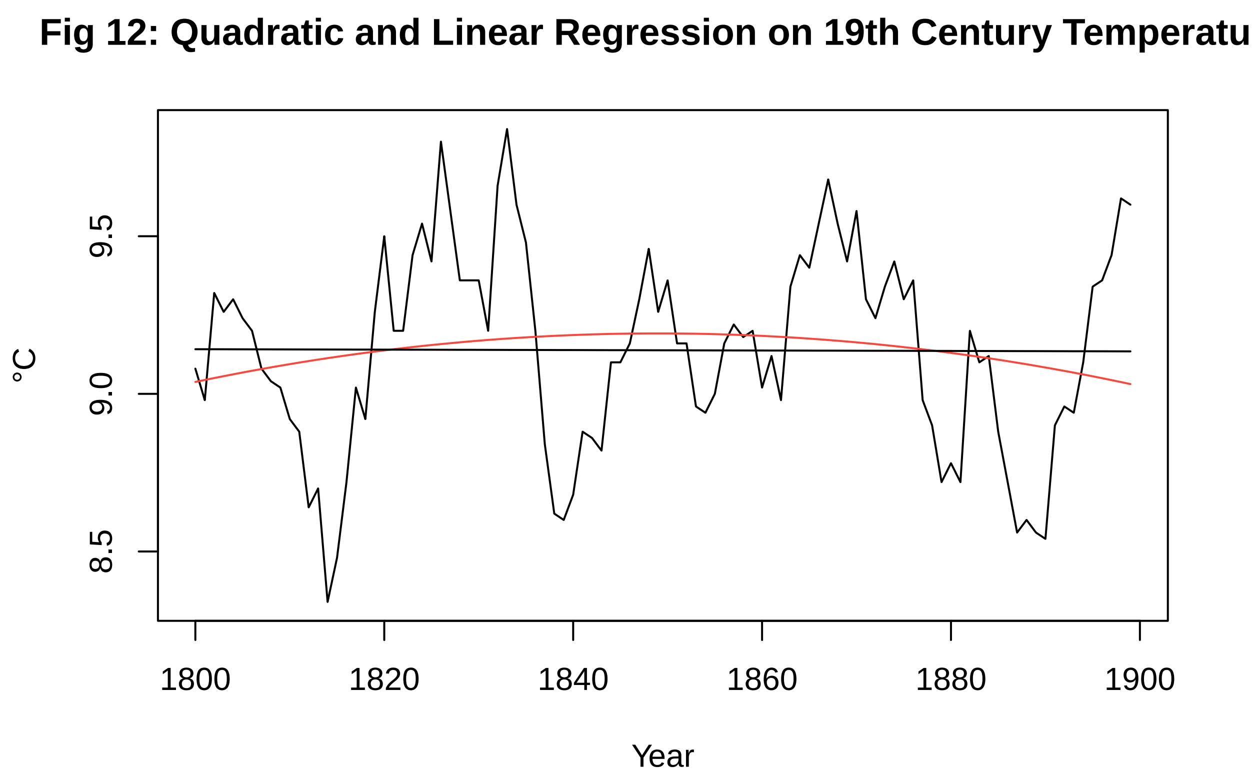
```
#Linear and quadratic regressions of the three centuries
cet_17 = window(cet_ann_ma, start = 1700, end = 1799)
cet_18 = window(cet_ann_ma, start = 1800, end = 1899)
cet_19 = window(cet_ann_ma, start = 1900, end = 1999)
```

```r
year_17 = time(cet_17)
yearsq_17 = time(cet_17) ^ 2
year_18 = time(cet_18)
yearsq_18 = time(cet_18) ^ 2
year_19 = time(cet_19)
yearsq_19 = time(cet_19) ^ 2

fit_17 = lm(cet_17 ~ 1 + year_17, na.action = na.exclude)
fit2_17 = lm(cet_17 ~ 1 + year_17 + yearsq_17, na.action = na.exclude)

fit_18 = lm(cet_18 ~ 1 + year_18, na.action = na.exclude)
fit2_18 = lm(cet_18 ~ 1 + year_18 + yearsq_18, na.action = na.exclude)

fit_19 = lm(cet_19 ~ 1 + year_19, na.action = na.exclude)
fit2_19 = lm(cet_19 ~ 1 + year_19 + yearsq_19, na.action = na.exclude)
```

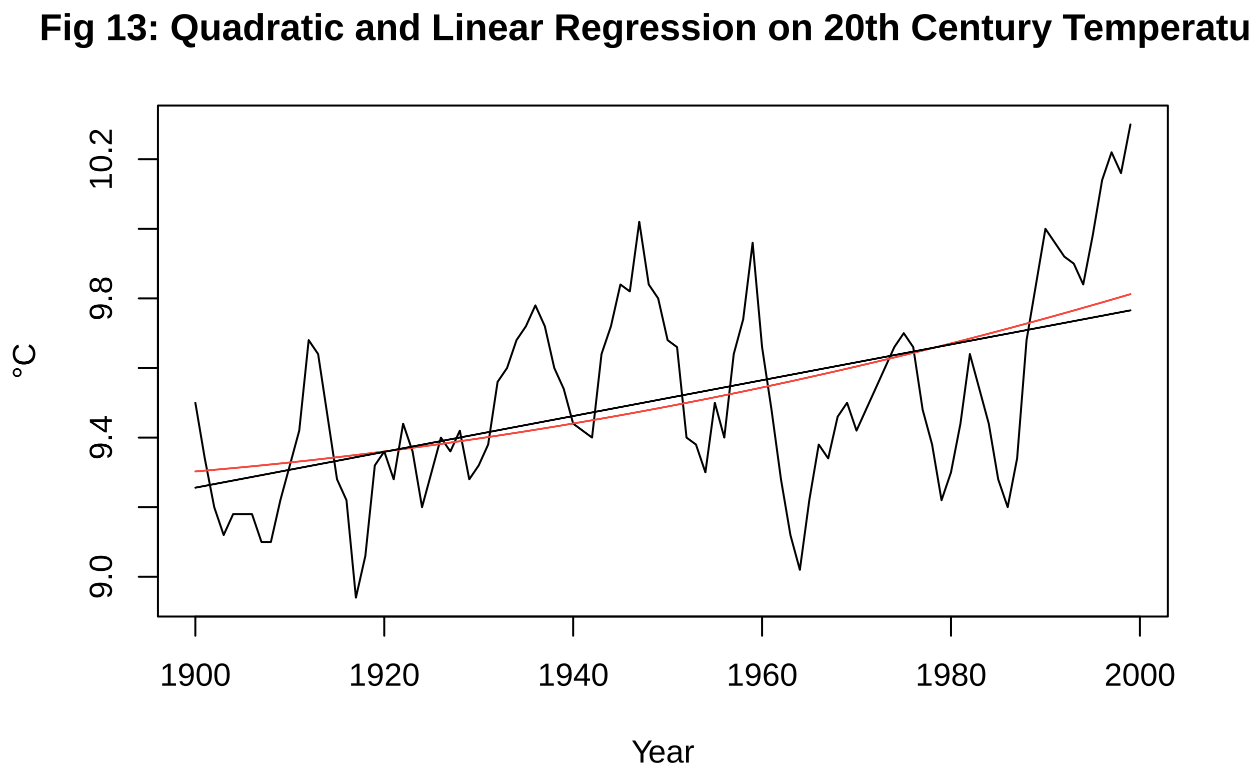
```r
#Draw plots for the three centuries
plot(cet_17,
     xlab = "Year",
     ylab = "°C",
     main = "Fig 11: Quadratic and Linear Regression on 18th Century Temperature")
lines(ts(fit2_17$fitted.values, start = 1700),
      type = "l",
      col = 2)
lines(ts(fit_17$fitted.values, start = 1700), type = "l")
```



```
plot(cet_18,
     xlab = "Year",
     ylab = "°C",
     main = "Fig 12: Quadratic and Linear Regression on 19th Century Temperature")
lines(ts(fit2_18$fitted.values, start = 1800),
      type = "l",
      col = 2)
lines(ts(fit_18$fitted.values, start = 1800), type = "l")
```




```
plot(cet_19,
     xlab = "Year",
     ylab = "°C",
     main = "Fig 13: Quadratic and Linear Regression on 20th Century Temperature")
lines(ts(fit2_19$fitted.values, start = 1900),
      type = "l",
      col = 2)
lines(ts(fit_19$fitted.values, start = 1900), type = "l")
```



3.2 Subsequent Trends

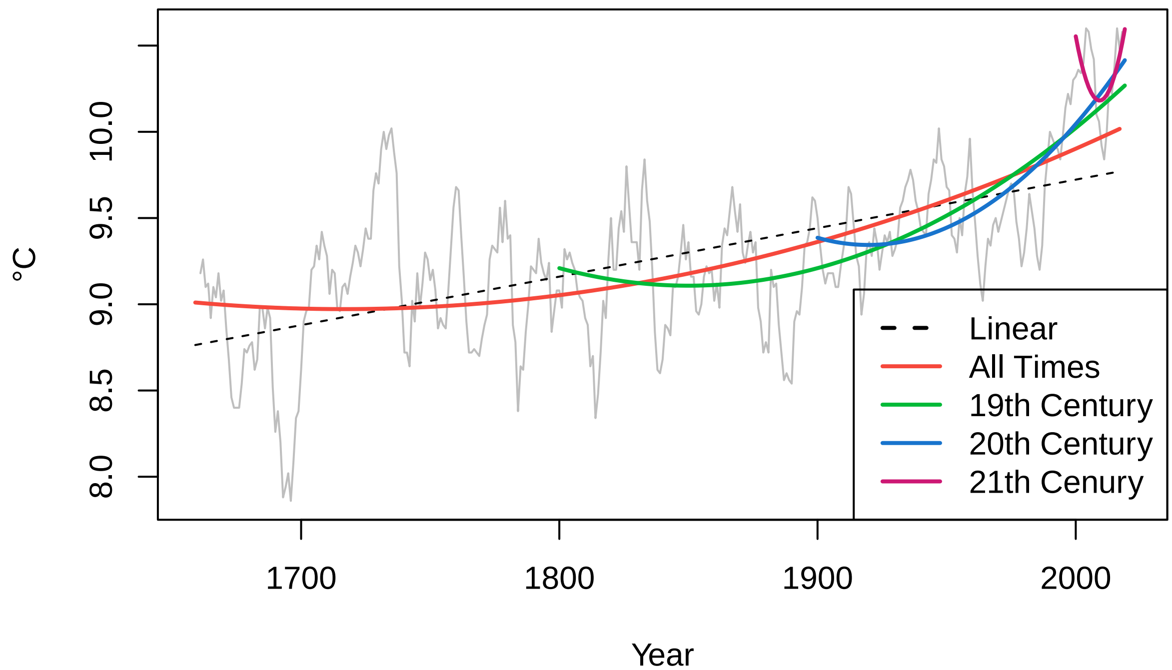
Benner (1999) conducted quadratic fitting to various parts of the time series data: the entire data set, data after 1800, and data after 1900, in order to show how the temperature trend has gradually changed over time. As new data has become available, I expanded Benner's subsequent trend analysis to include the piece of data after 2000.

Looking at figure 7, it is clear that the quadratic curves are getting more convex as time approaches the present. There is also a more and more upward trend in temperatures going from the 19th Century to the 21th Century. Within the 21th Century, there is a strong convex trend where the drop and the raise offset one another, but compared to the years before 2000, the temperatures after 2000 are at a historical high.

```
#Quadratic models for three periods
cet_18up = window(cet_ann_ma, start = 1800)
cet_19up = window(cet_ann_ma, start = 1900)
cet_20up = window(cet_ann_ma, start = 2000)

year_18up = time(cet_18up)
year_19up = time(cet_19up)
year_20up = time(cet_20up)
yearsq_18up = time(cet_18up) ^ 2
yearsq_19up = time(cet_19up) ^ 2
yearsq_20up = time(cet_20up) ^ 2

fit2_18up = lm(cet_18up ~ 1 + year_18up + yearsq_18up, na.action = na.exclude)
fit2_19up = lm(cet_19up ~ 1 + year_19up + yearsq_19up, na.action = na.exclude)
fit2_20up = lm(cet_20up ~ 1 + year_20up + yearsq_20up, na.action = na.exclude)
```

```r
#Draw plots
plot(
  cet_ann_ma,
  type = "l",
  col = "grey",
  xlab = "Year",
  ylab = "°C",
  main = "Fig 14: Quadratic Regressions on Different Time Periods"
)
lines(ts(fit$fitted.values, start = 1659), lty = 2)
lines(ts(fit2$fitted.values, start = 1659),
      col = 2,
      lwd = 2)
lines(ts(fit2_18up$fitted.values, start = 1800),
      col = 3,
      lwd = 2)
lines(ts(fit2_19up$fitted.values, start = 1900),
      col = 4,
      lwd = 2)
lines(ts(fit2_20up$fitted.values, start = 2000),
      col = 6,
      lwd = 2)
legend(
  "bottomright",
  legend = c(
    "Linear",
    "All Times",
    "19th Century",
    "20th Century",
    "21th Cenury"
  ),
  col = c(1, 2, 3, 4, 6),
  lty = c(2, 1, 1, 1, 1),
  lwd = 2
)
```

Fig 14: Quadratic Regressions on Different Time Periods



3.3 Seasonal Analysis

Seasonal factors also play crucial roles in thermometry analysis. Therefore, besides the monthly CET data, I also utilized quarterly CET data to identify whether the temperature trends vary for different seasons.

I fit linear regression models to the time series data of each season, and obtain the following coefficients:

$$Winter = 0.00409 Spring = 0.00299 Summer = 0.00117 Autumn = 0.00318$$

From figure 15, it is clear that although temperature has been rising in all four seasons, the slope is the steepest for winter and the flattest for summer. This indicates a significant seasonal difference in temperature change.

One possible explanation bases on regional characteristics of England. As a result of the cool weather, England traditionally depends on coal-based heating in winter, while the majority of households don't use air-conditioners in summer. As the population expands over the years, more coals are consumed in winter while the energy use doesn't change much in summer due to the general lack of demand. Thus, the temperature trend in the winter becomes distinct from that in the summer.

```r
#Load seasonal data and kernel smooth
df2 <- read.table("meantemp_seasonal_totals.txt", header = TRUE)
df2[1, 2] = NA
df2 = subset(df2, select = -c(Annual))
cet_win <- ts(df2$Win,
              frequency = 1,
              start = 1659,
              end = 2022)
cet_sum <- ts(df2$Sum,
```

```r
              frequency = 1,
              start = 1659,
              end = 2022)
cet_spr <- ts(df2$Spr,
              frequency = 1,
              start = 1659,
              end = 2022)
cet_aut <- ts(df2$Aut,
              frequency = 1,
              start = 1659,
              end = 2022)

cet_win_ma <-
  stats::filter(cet_win, filter = rep(1 / 5, 5), sides = 2)
cet_spr_ma <-
  stats::filter(cet_spr, filter = rep(1 / 5, 5), sides = 2)
cet_aut_ma <-
  stats::filter(cet_aut, filter = rep(1 / 5, 5), sides = 2)
cet_sum_ma <-
  stats::filter(cet_sum, filter = rep(1 / 5, 5), sides = 2)
```

```r
#Fit linear models to seasonal data
fit_win = lm(cet_win_ma ~ 1 + time(cet_win_ma), na.action = na.exclude)
print(paste(
  "The time coefficient for winter is",
  summary(fit_win)$coefficients[2, 1]
))
```

```
## [1] "The time coefficient for winter is 0.00408890306717916"
```

```r
fit_spr = lm(cet_spr_ma ~ 1 + time(cet_spr_ma), na.action = na.exclude)
print(paste(
  "The time coefficient for spring is",
  summary(fit_spr)$coefficients[2, 1]
))
```

```
## [1] "The time coefficient for spring is 0.00299330241745692"
```

```r
fit_sum = lm(cet_sum_ma ~ 1 + time(cet_sum_ma), na.action = na.exclude)
print(paste(
  "The time coefficient for summer is",
  summary(fit_sum)$coefficients[2, 1]
))
```

```
## [1] "The time coefficient for summer is 0.00117081664724779"
```

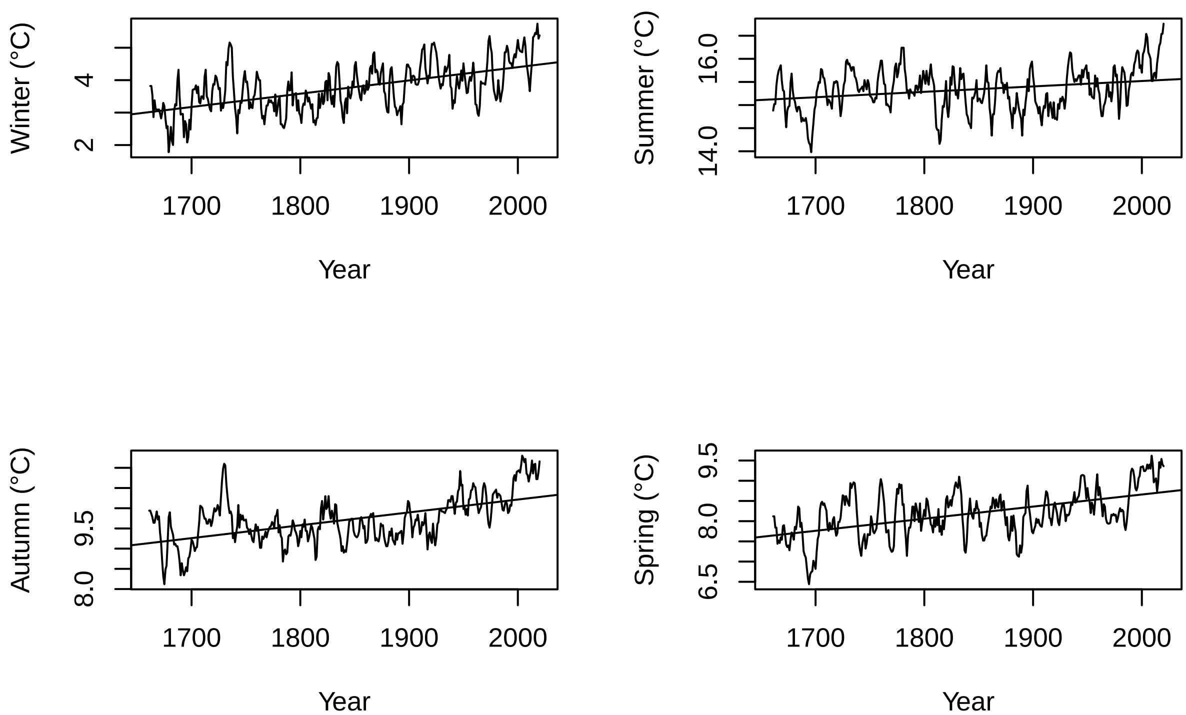
```r
fit_aut = lm(cet_aut_ma ~ 1 + time(cet_aut_ma), na.action = na.exclude)
print(paste(
  "The time coefficient for autumn is",
  summary(fit_aut)$coefficients[2, 1]
))
```

```
## [1] "The time coefficient for autumn is 0.00318322415039211"
```

```r
#Draw plots
par(mfrow = (c(2, 2)))
plot(cet_win_ma, xlab = "Year", ylab = "Winter (°C)")
```

```
abline(fit_win)
plot(cet_sum_ma, xlab = "Year", ylab = "Summer (°C)")
abline(fit_sum)
plot(cet_aut_ma, xlab = "Year", ylab = "Autumn (°C)")
abline(fit_aut)
plot(cet_spr_ma, xlab = "Year", ylab = "Spring (°C)")
abline(fit_spr)
mtext(
  "Fig 15: Linear Regression for Four Seasons",
  side = 3,
  line = -1,
  outer = TRUE
)
```

Fig 15: Linear Regression for Four Seasons



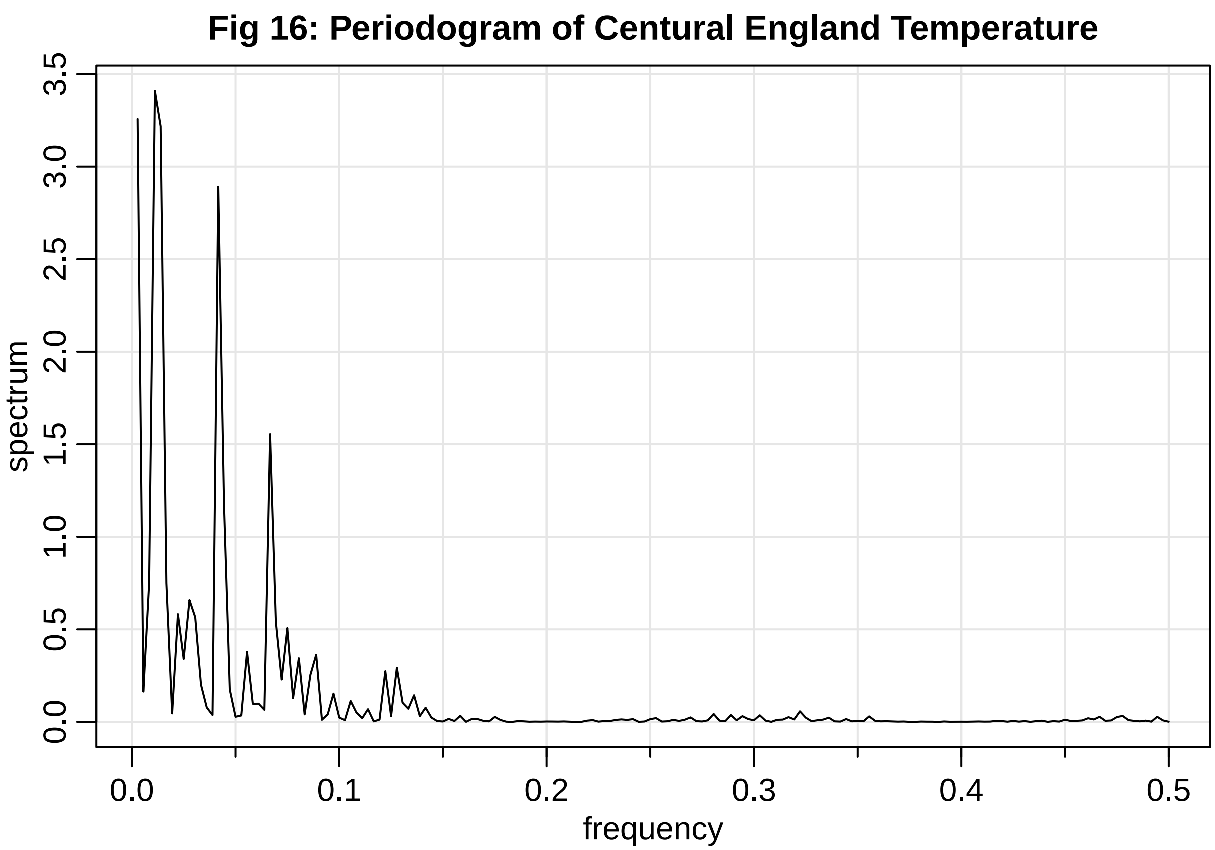
4. Spectral Analysis

4.1 Overall Spectural Analysis

The temperature of a certain region is often affected by a number of periodic factors such as tides and the activity of sunspots. Therefore, I conducted a spectral analysis on the CET data to identify any periodical trends over the centuries.

Figure 16 shows a periodogram of the CET data set. I listed the most significant peaks in a descending order. Some of the most significant peaks I obtained are consistent with what Benner (1999) obtained through FFT, LSP, and SSA spectrum analysis, such as the 68-year cycle versus Benner's 72-year cycle, the 24-year cycle versus Benner's 24-year cycle, and the 120-year cycle versus Benner's 110-year cycle.

```r
#Draw periodogram
cet.per = mvspec(cet_ann_ma,
                 plot = TRUE,
                 na.action = na.exclude,
                 main = "Fig 16: Periodogram of Centural England Temperature")
```




```r
#Present peaks
sort_peaks = cet.per$details[order(cet.per$details[, 3], decreasing = TRUE),]
print(sort_peaks[1:10,])
```

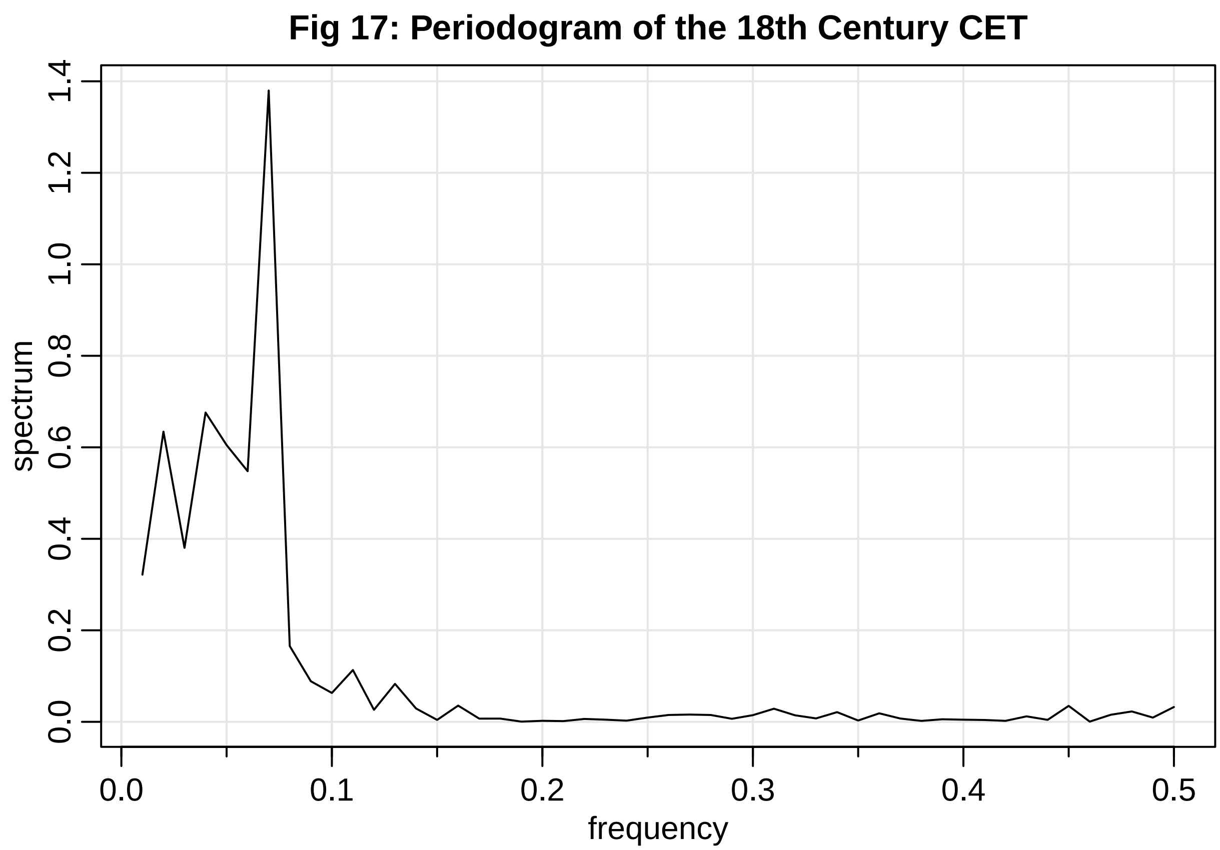
```
##       frequency period spectrum
##  [1,]    0.0111   90.0   3.4096
##  [2,]    0.0028  360.0   3.2572
##  [3,]    0.0139   72.0   3.2177
##  [4,]    0.0417   24.0   2.8915
##  [5,]    0.0667   15.0   1.5545
##  [6,]    0.0444   22.5   1.1713
##  [7,]    0.0083  120.0   0.7498
##  [8,]    0.0167   60.0   0.7495
##  [9,]    0.0278   36.0   0.6580
## [10,]    0.0222   45.0   0.5819
```

4.2 Spectural Analysis by Centuries

Similar to section 3.2, I am interested in how the periodical patterns differ in different centuries. Due to the possibility of certain oscillations existing only in certain periods but not persisting throughout, it is essential to break down the time series for a closer study. Therefore, I conducted spectral analysis on the 18th, the 19th, and the 20th century data to identify the most significant oscillations in each time period.

Judging from the periodograms and the peaks listed below, there are several oscillations persistent throughout all three centuries: 15 years, 50 years, 20 years, and 25 years.

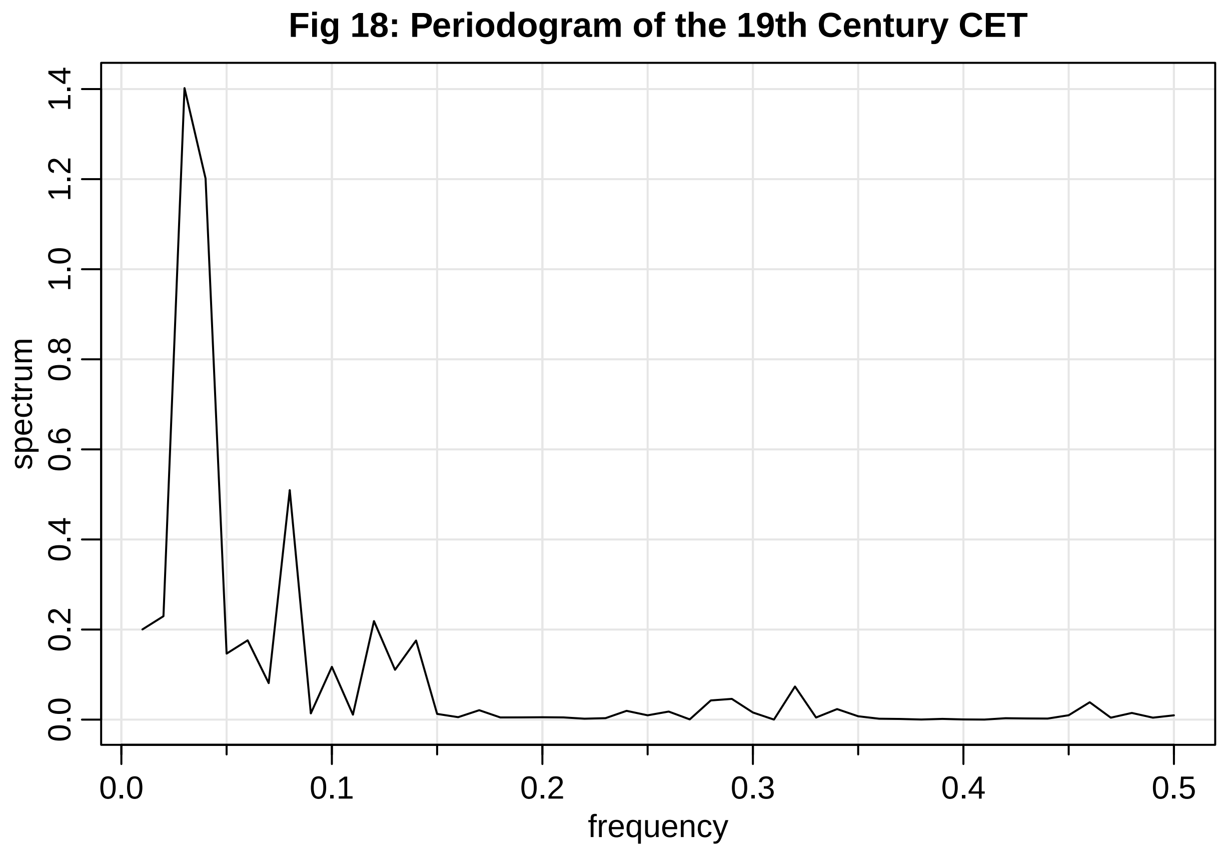
```
#Periodogram and peaks in the three centuries
cet.pre1 = mvspec(cet_17,
                  plot = TRUE,
                  na.action = na.exclude,
                  main = "Fig 17: Periodogram of the 18th Century CET")
```




```
sort_peaks1 = cet.pre1$details[order(cet.pre1$details[, 3], decreasing =
                                       TRUE), ]
sort_peaks1[1:5, ]
```

```
##      frequency  period spectrum
## [1,]      0.07 14.2857   1.3799
## [2,]      0.04 25.0000   0.6760
## [3,]      0.02 50.0000   0.6342
## [4,]      0.05 20.0000   0.6048
## [5,]      0.06 16.6667   0.5478
```

```
cet.pre2 = mvspec(cet_18,
                  plot = TRUE,
                  na.action = na.exclude,
                  main = "Fig 18: Periodogram of the 19th Century CET")
```

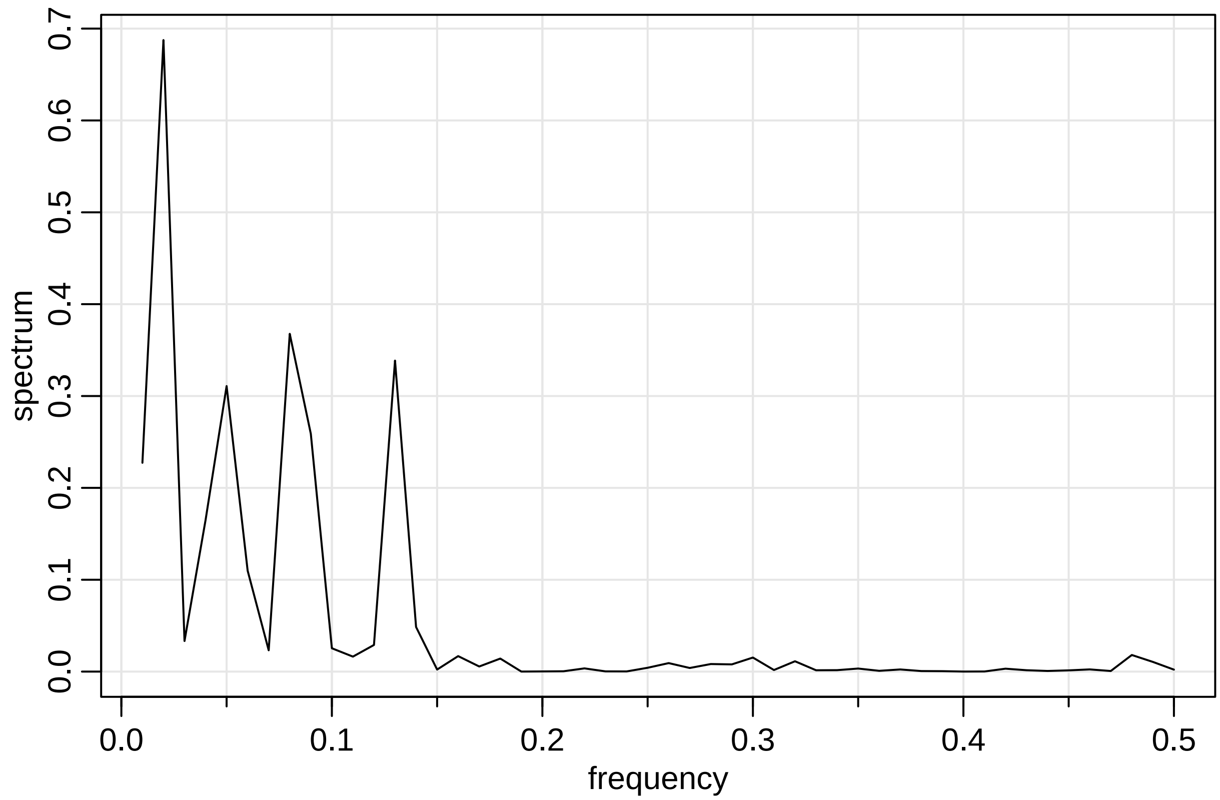



```
sort_peaks2 = cet.pre2$details[order(cet.pre2$details[, 3], decreasing =
                                                           TRUE),]
sort_peaks2[1:5,]
```

```
##      frequency  period spectrum
## [1,]      0.03 33.3333   1.4021
## [2,]      0.04 25.0000   1.2015
## [3,]      0.08 12.5000   0.5095
## [4,]      0.02 50.0000   0.2296
## [5,]      0.12  8.3333   0.2188
```

```
cet.pre3 = mvspec(cet_19[1:121],
                  plot = TRUE,
                  na.action = na.exclude,
                  main = "Fig 19: Periodogram of the 20th Century CET")
```

Fig 19: Periodogram of the 20th Century CET




```r
sort_peaks3 = cet.pre1$details[order(cet.pre3$details[, 3], decreasing =
                                                      TRUE), ]
sort_peaks3[1:5, ]
```

```
##      frequency  period spectrum
## [1,]      0.02 50.0000   0.6342
## [2,]      0.08 12.5000   0.1656
## [3,]      0.13  7.6923   0.0829
## [4,]      0.05 20.0000   0.6048
## [5,]      0.09 11.1111   0.0887
```

5. Conclusion

The Central England Temperature data set is an invaluable source of information when studying global warming and temperature change. After analyzing the CET data using linear models, quadratic models, and ARIMA models, I was able to identify an overall warming trend, especially beginning in the 20th century. The warming trend in winter is more severe then that in summer, possibly due to coal consumption for heating in winter. Furthermore, using spectral analysis, I identified several significant oscillations including the 15-year cycle and the 25-year cycle, which are highly likely to be related to tidal movements or solar activities.

This paper is limited in the sense that all data is collected in central England instead of globally. It is possible that temperatures of other regions in the world display patterns distinct from the CET data. Further analysis can be done on the correlation between CET and other weather-related factors, in order to study what caused the upward trend and oscillations in the CET data.

6. References

1. BENNER, T. C. (1999). Central England temperatures: Long-term variability and teleconnections. Int. J. Climatol. 19, 391¯a403.
2. Gayathri Vaidyanathan (February 25, 2016) Did Global Warming Slow Down in the 2000s, or Not? Scientists clarify the recent confusion. http://www.scientificamerican.com/article/did-global-warming-slow-down-inthe-2000s-or-not/
3. Parker, D.E., T.P. Legg, and C.K. Folland. 1992. A new daily Central England Temperature Series, 1772-1991. Int. J. Clim., Vol 12, pp 317-342