STAT 33600 Project

Yuhui Ni 12174192

1 Introduction

As Gayathri Vaidyanathan said in his article [1], The debate between researchers and doubters about whether or not the global warming trend has slowed down has reached a crescendo in the past few years. Some scientists argue that the global warming trend has not slowed down in 2000s, while others hold the opposite opinion. Here, we use the central England temperature as an example to clarify whether or not the climate is getting warm. The second part is parametric modeling of annual central England temperature and season central England temperature separately. The third part we use nonparametric method to model annual central England temperature. The fourth part deals with season central temperature for each season based on nonparametric methods. The last part is conclusion about warming trend analysis results using these methods.

2 Parametric modeling of the temperature

2.1 annual temperature data modeling

First, we consider annual average temperature data by taking the average of four seasonal data, since there are more variability in quarterly temperature data, which unavoidably decreases the resolution of model construction and temperature trend analysis. At the same time, we use the polynomial regression to fit the temperature trend. Here, we consider the following model:

$$x_t = u_t + y_t$$

We suppose $u_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_n t^n$, then use least square regression to fit this model trend with increasing polynomial degree. Here, we increase polynomial degree one by one until new predictor is no long statistically significant.

First, we consider simple linear trend $u_t = \beta_0 + \beta_1 t$:

```
 Estimate    Std. Error  t value     Pr(>|t|)
(Intercept) 4.260896431 0.5940368057 7.172782 4.308299e-12
t           0.002704857 0.0003226905 8.382200 1.232161e-15
```

As we can see, the p-values of intercept and t are both statisticly significant at 5% level. Again, we consider quadratic polynomial $u_t = \beta_0 + \beta_1 t + \beta_2 t^2$:

```
 Estimate    Std. Error   t value     Pr(>|t|)
(Intercept)  4.105933e+01 1.165774e+01  3.522065 0.0004842434
t           -3.746325e-02 1.271313e-02 -2.946816 0.0034233964
I(t^2)       1.092712e-05 3.457326e-06  3.160571 0.0017103688
```

The above summary results suggest that quadratic ploynomial also makes senses here.

Then we consider cubic ploynomial $u_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$

```
                 Estimate   Std. Error   t value     Pr(>|t|)
(Intercept) -8.618031e+02 2.315071e+02 -3.722577 0.0002294238
t            1.441788e+00 3.790441e-01  3.803747 0.0001679165
I(t^2)      -7.954096e-04 2.065322e-04 -3.851262 0.0001395227
I(t^3)       1.462344e-07 3.745092e-08  3.904695 0.0001130321
```

The predictors are all statistically significant at 5% level. And the predictors are no longer significant if we use higher degree polynomials. Hence, cubic ploynomial may be the best polynomial here.

Then we use unit root test to examine the stationary of the detrended temperature data $y_t = x_t - u_t$.

```
Title:
Augmented Dickey-Fuller Test
Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -10.4869
  P VALUE:
    0.01
```

According to above result, we reject the null hypothesis that there exists a unit root in y_t, which means we could treat y_t as stationary process. Next, we implement the examination of sample autocorrelation function and sample partial autocorrelation function of the detrended data $y_t = x_t - u_t$. The result suggests that a MA(2) process maybe appropriate for fitting the autocorrelation structure.

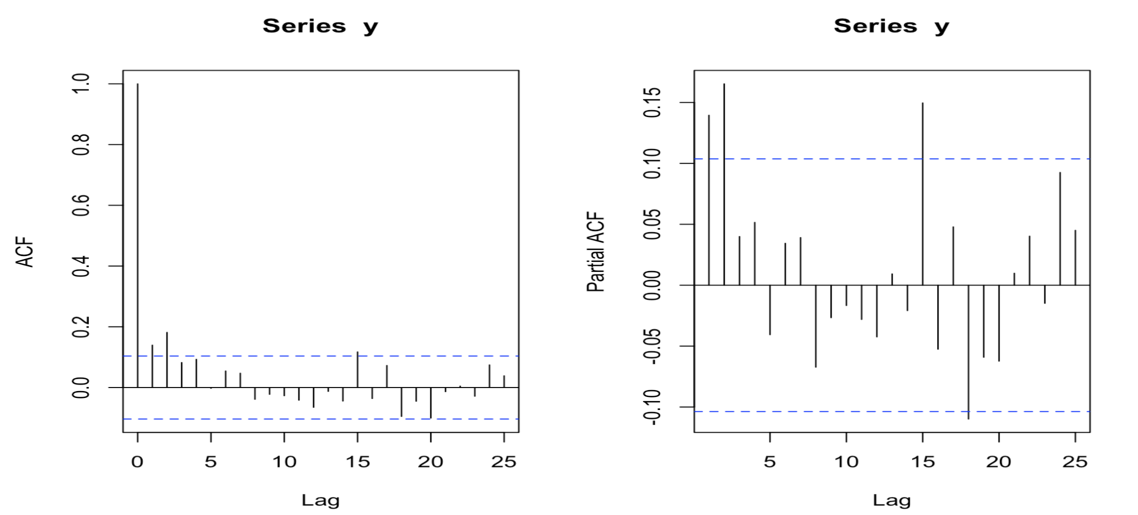


Figure 1: ACF and PACF

According to below result, we can see that

$$y_t = 0.0004 + \epsilon_t + 0.1031\epsilon_{t-1} + 0.1475\epsilon_{t-2} \quad \sigma^2 = var(\epsilon_t) = 0.3517$$

.

```
$coef
          ma1          ma2    intercept
0.1031027760 0.1475298096 0.0004422275
$sigma2
[1] 0.3516558
```

The fitted trend is shown superimposed on the annual temperature series in Figure2. The simple cubic trend almost imposes an increasing, stable, then increasing trend for the temperature data. However, such simplicity fits assuming constant cubic trend does not accord very well with early temperature trend between 1650 and 1700 when temperature trend more likely seems to decline. Furthermore, this model does not take seasonal effect into consideration. Hence, we next take seasonal effect into consideration.

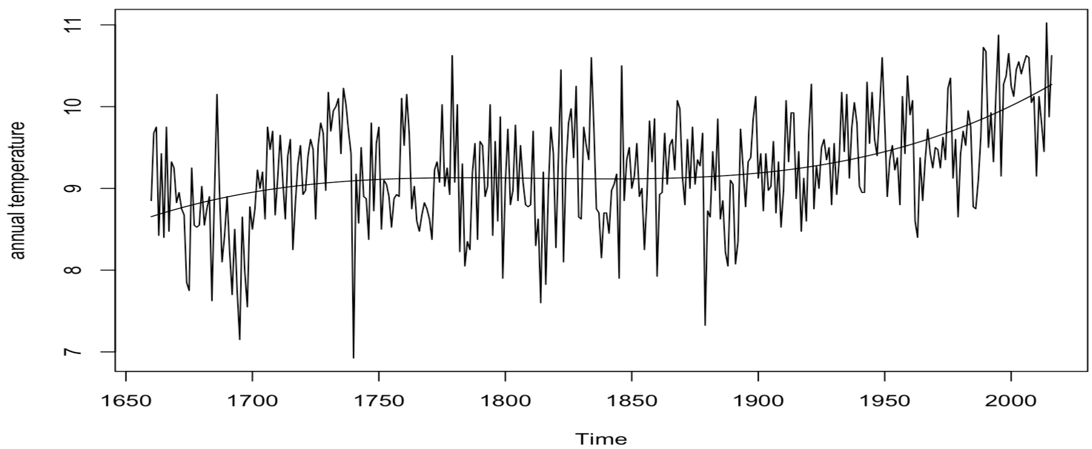


Figure 2: fitted trend for annual temperature

2.2 Structrue model for season temperature data

In this part, we consider a structure,$x_t = T_t + S_t + N_t$, where T_t is a trend component,S_t is a seasonal component, and N_t is noise. In this case, time t is in quarters (1960.00, 1960.25, . . .) so one unit of time is a year. In order to compare with polynomial regression model, we considering the following model:

$$x_t = \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \alpha_1 Q_1(t) + \alpha_2 Q_2(t) + \alpha_3 Q_3(t) + \alpha_4 Q_4(t) + w_t$$

where $Q_i(t) = 1$ if time t corresponds to $i = 1, 2, 3, 4$,i.e. $Q_i(t)$ is an indicator variable, w_t is noise part.

```
Call:
lm(formula = seasontemp ~ 0 + t + I(t^2) + I(t^3) + Q, na.action = NULL)
Residuals:
    Min      1Q  Median      3Q     Max
-4.5682 -0.5932  0.0225  0.6381  3.1630
Coefficients:
         Estimate Std. Error t value Pr(>|t|)
t       1.458e+00  2.955e-01   4.933 9.03e-07 ***
```

```
I(t^2) -8.042e-04  1.610e-04  -4.996 6.56e-07 ***
I(t^3)  1.479e-07  2.918e-08   5.067 4.57e-07 ***
Q1     -8.767e+02  1.805e+02  -4.858 1.32e-06 ***
Q2     -8.723e+02  1.805e+02  -4.833 1.49e-06 ***
Q3     -8.651e+02  1.805e+02  -4.793 1.81e-06 ***
Q4     -8.708e+02  1.805e+02  -4.825 1.55e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 0.9622 on 1427 degrees of freedom
Multiple R-squared:  0.9911,Adjusted R-squared:  0.9911
F-statistic: 2.27e+04 on 7 and 1427 DF,  p-value: < 2.2e-16
```

Above result shows that all predictors are significant. We here only plot fitted trends. The corresponding result

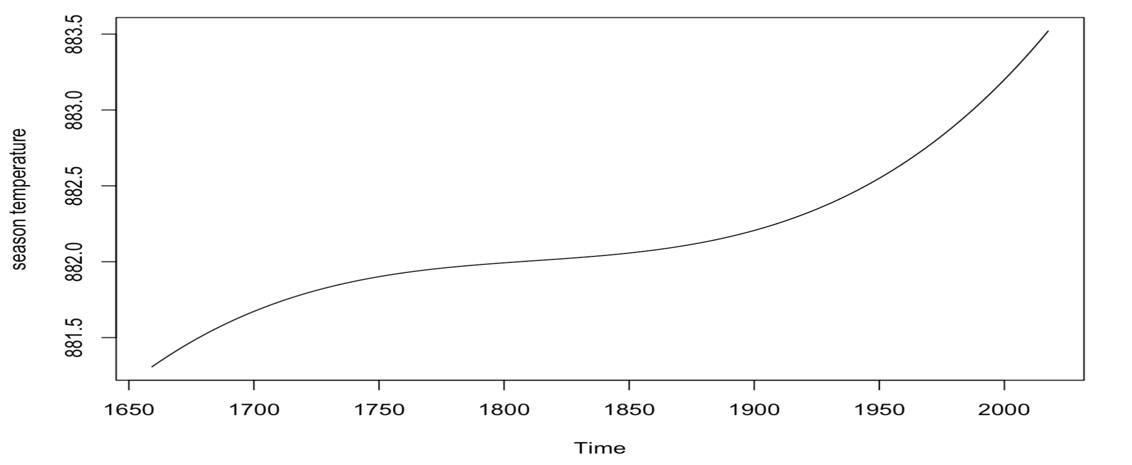


Figure 3: fitted trends for structure model

suggests that the basic trend is slowly increasing before 1750, stable between 1750 and 1900, and increasing faster than early period in 17th century when it enters 20th century. However, the trend value is too big and it does not make sense well. Therefore, we decide to use non-parametric modeling method and fit temperature trend separately for different season.

3 Nonparametric modeling trends

3.1 Kernel smoothing detrending

We assume $x_t = u_t + y_t$. Here, we adopt kernel smoothing method to detrend u_t.Kernel smoothing is a moving average smoother that uses a weight function, or kernel, to average the observations, where u_t is:

$$u_t = \sum_{i=1}^{n} w_i(t) x_i$$

where

$$w_i(t) = K(\frac{t-i}{b}) / \sum_{j=1}^{n} K(\frac{t-j}{b})$$

are the weights and $K(\cdot)$ is kernel function statisfying the conditoins

$$\int k(u)du = 1 \qquad \int uK(u)du = 0 \qquad \int u^2K(u)du > 0$$

b is the bandwidth. Here we use normal kernel $K(z) = \frac{1}{\sqrt{2\pi}}exp(-\frac{z^2}{2})$. The wider the bandwidth, b, the smoother the result. In below figure, we show different bandwidth choices. As we can see, different bandwidth has different trend. For smaller bandwidth $b = 5$, the trend exhibits cycle, it waves up and down. As bandwidth grows larger, the trend becomes smoother and is more similar to the cubic polynomial trend. Hence, the choosing of width is very important.

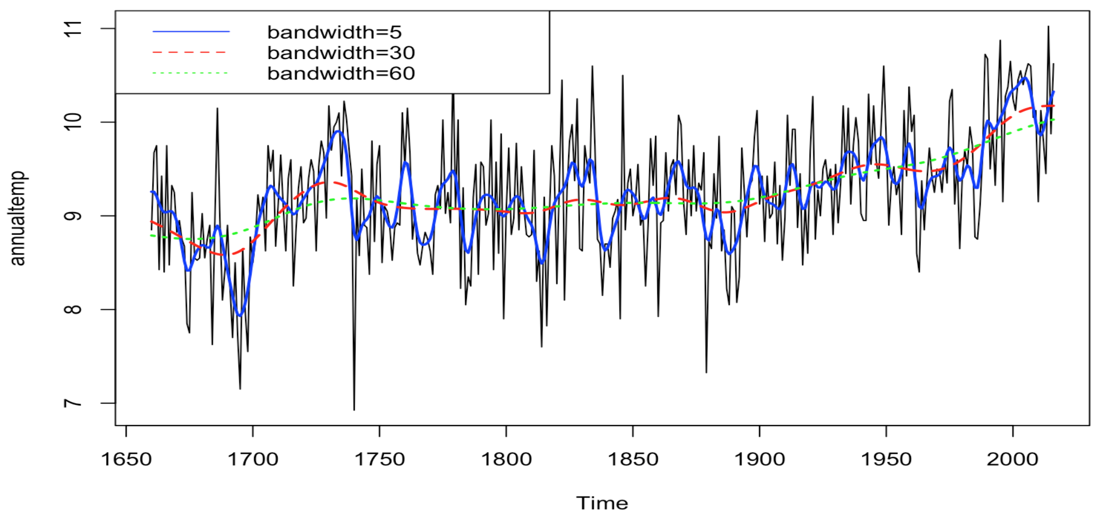


Figure 4: fitted annual temperature trends for different bandwidth

Here, we choose the optimal bandwidth via the method of cross-validation . Based on the principle of ”leave-one-out” prediction, cross-validation minimizes the criterion:

$$CV(b) = \frac{1}{n}\sum_{t=1}^{n}(x_t - \hat{u_t^h})$$

where $\hat{u_t^h}$ is the estimate of $\hat{u}_t$ based on $x_1, x_2, \ldots, x_{h-1}, x_{h+1}, \ldots, x_n$, i.e. with observation h removed.

The figure 5 of Cross-Validation suggests that CV has minimium at $b \approx 5$, but at the same time we notice that CV does not change much even b becomes 30 or even larger(it only changes about 0.02). Hence, we should take more b values into consideration.

When bandwidth, b=5, the kernel smoothing function fits data well. It exhibits periodically. Between 1650 and 1700, it waves down. Between 1700 and 1740, it waves up. After that period, the whole average trend seems to be constant until 1880. It again waves up slowly between 1880 and 1950 and increases faster after 1950. When bandwidth, b=30, the red dashed line exhibits almost pattern except that the periodical pattern becomes less obvious. So we have reason to believe that $b = 5$ and $b = 30$ can both represent the trend of temperature in central England.The former exhibits more periodical pattern and the latter exhibits less, but their whole trends are almost same.

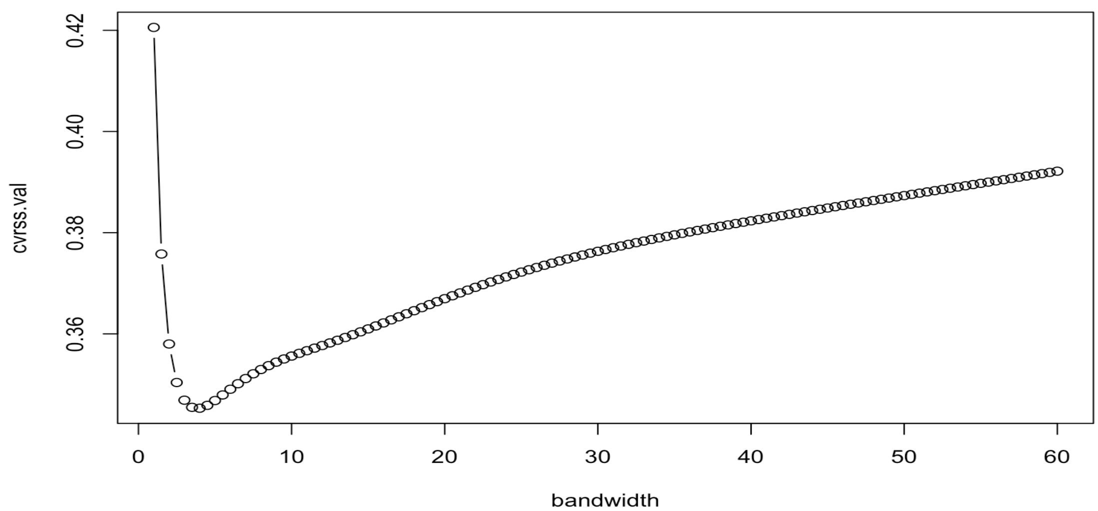


Figure 5: cross validation

3.2 lowess detrending

In this section, we use the lowess method to fit the temperature trend. First, a certain proportion of nearest neighbors to x_t are included in a weighting scheme; values closer to x_t in time get more weight. Then, a robust weighted regression is used to predict x_t and obtain the smoothed values u_t . The larger the fraction of nearest neighbors included, the smoother the fit will be.

Here, R uses robust locally weighted regression. According to Cleveland, W. S. [2]'s definition, let W be a weight function with the following properties:

- W(x)>0 for $|x| < 1$
- W(-x)=W(x)
- W(x) is a nonincreasing function for $x \geq 0$
- W(x)=0 for $|x| \geq 1$

Given data $x_1, x_2, \cdots, x_n$, for each i let h_i be the distance from x_i to the r th nearest neighbor of x_i, which means h_i represents the r th smallest number among $|x_i - x_j|$, for $j = 1, \cdots, n$. For $k = 1, \cdots, n$, let

$$w_k(x_i) = W(h_i^{-1}(x_k - x_i))$$

For each i compute the estimates, $\hat{\beta}_j, j = 0, \cdots, d$ of the parameters in a polynomial regression of degree d of y_k on x_k, which is fitted by weighted least squares with weight $w_k(x_i)$ for (x_k, y_k). Thus the $\hat{\beta}_j(x_i)$ are the values of β_j that minimize

$$\sum_{k=1}^{n} w_k(x_i)(y_k - \beta_0 - \beta_1 x_1 - \cdots - \beta_d x_k^d)^2$$

The smoothed point at x_i is $(x_i, \hat{x_i})$, where $\hat{y_i}$ is the fitted value value of the regression at x_i. Thus

$$\hat{y_i} = \sum_{j=0}^{d} \hat{\beta}_j(x_i) x_i^j$$

R uses lowess with robust locally weighted regression. We try different smoother span of $f = 0.1, 0.5, 0.8$ of the data to get the trend. As we can see, the fitted temperature trend gets smoother as span becomes larger. For $f = 0.5$ and

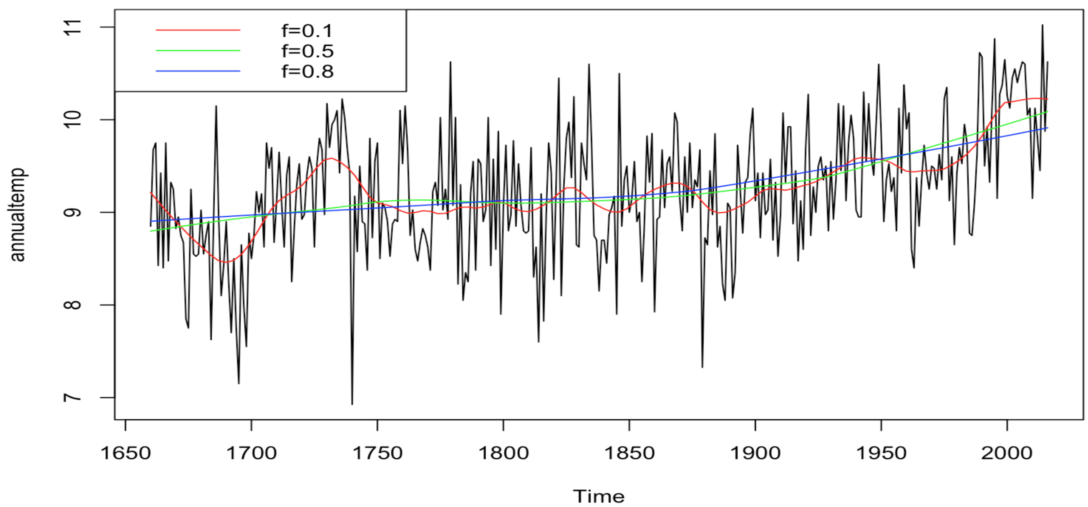


Figure 6: different smoother span for lowess

$f = 0.8$, the temperature trend is just like an increasing line, which grows slower before 1850 and grows faster after that. For $f = 0.1$, similar to kernel smoothing method, lowess fitting shows that it is pattern of declining, increasing, stable and then increasing. Specially, we see that the trend has a small downward between 1650 and 1700, and then it goes up until 1700. And it goes down between 1700 and 1750. In other words, it behaves cyclically. After that it stays almost stable between 1750 and 1900. Finally, it begins to increase faster especially after 1970.

Larger span causes the fitted trend to be over smoothed, hence, we have reason to believe that smaller span could better represent the inner trend of central England temperature data. Therefore, there is evidence for us to believe that the temperature is getting warm in 2000s according to the fitted temperature trend.

3.3 Smoothing spline detrending

Similar to polynomial regression, smoothing spline minimizes a compromise between the fit and the degree of smoothness given by:

$$\sum_{t=1}^{n}(x_t - u_t)^2 + \lambda \int (m_t^{''})^2 dt$$

where u_t is a spline with a knot at each t and primes denote differentiation. The degree of smoothness of degree is controlled by $\lambda > 0$. λ is seen as a trade-off between linear regression (completely smooth) and the data itself (no smoothness). The larger the value of λ, the smoother the fit. Here we separately fit the temperature trend using ordinary leave-one-out cross-validation(CV) and generalized cross-validation(GCV) to determine the best smooth parameter spar.

From Figure 7, we see that CV and GCV almost get the same result and we can directly use CV method for choosing the best spar. At the same time, the figure suggests that the temperature exhibits cyclical behavior with relatively large amplitude between 1650 and 1750. And it stays relatively stable between 1750 and 1900. After that it is increasing faster.

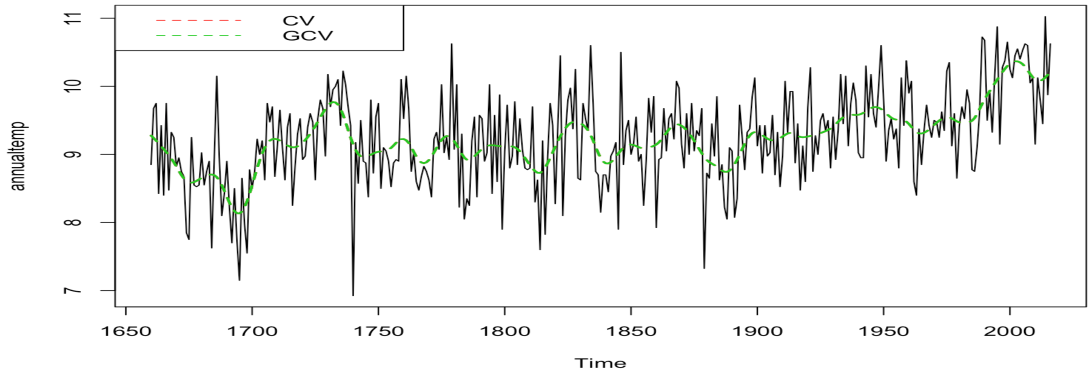
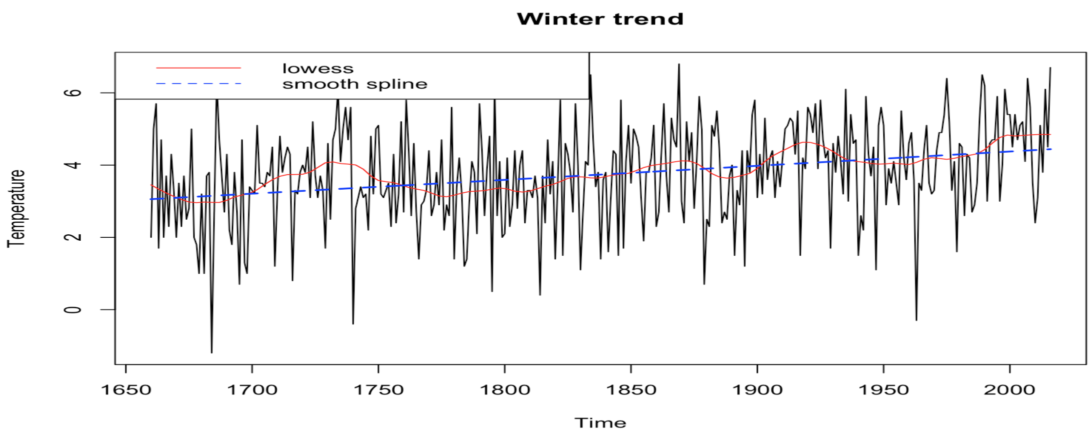
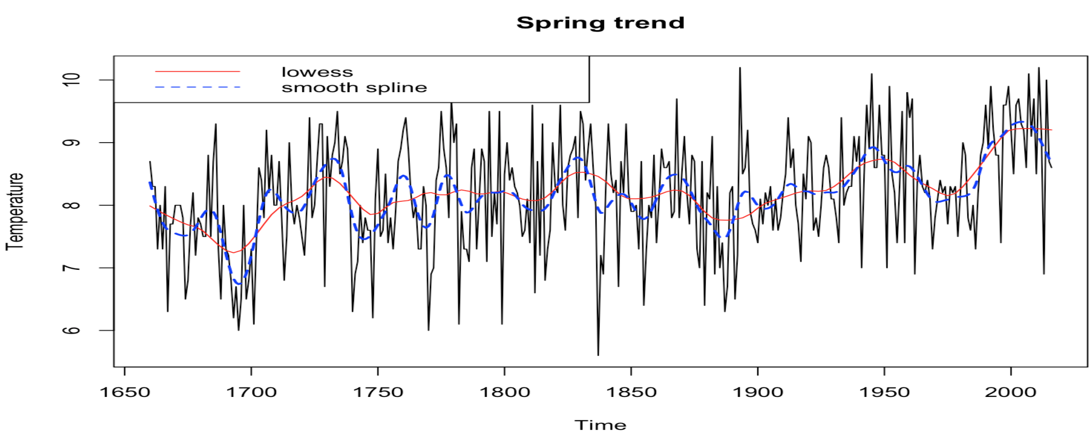


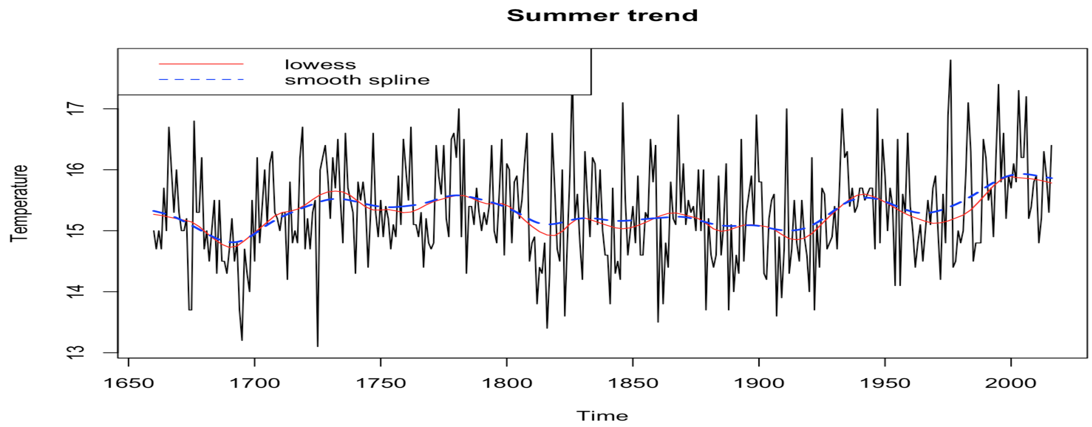
Figure 7: smoothing spline using CV and GCV

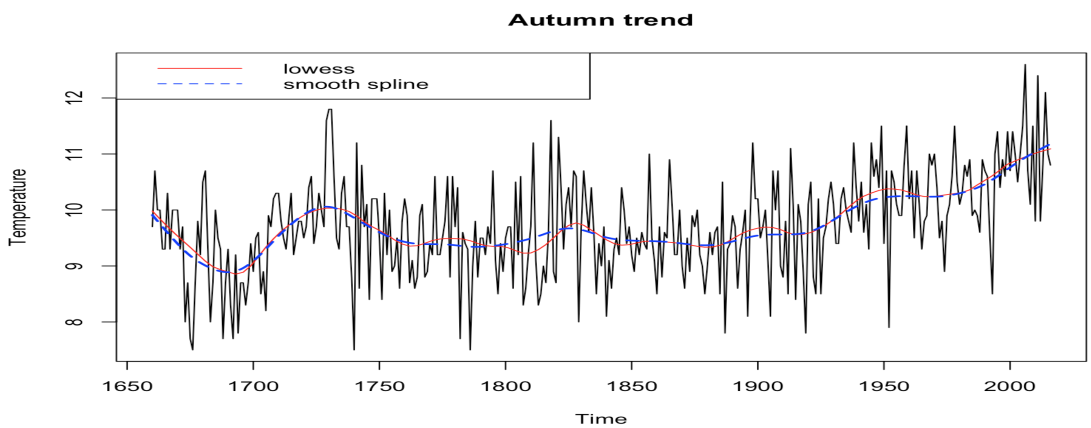
4 Seasonal trends modeling

In previous discussion, we does not take seasonal effect into consideration. However, trend patterns are more clear when observed through seasonal data. Here, we use lowess and smoothing spline to remodeling seasonal trends separately for each season just as Harvey, David I., and Terence C. Mills [3] did.









For lowess method, the general pattern is of decreasing, increasing, stable and then increasing trend temperatures. However, the pattern is not significant for all seasons. As we can see, the winter and summer temperature trend show no enough evidence to exhibit this pattern. For winter temperature trend, it has a very slowly increasing trend. For summer temperature trend, it looks more like a stable trend so that we have not enough evidence to say that the summer is getting warm in 2000s. Additionally, the summer temperatures in recent years are not higher than 1700s. For spring and autumn trend temperature, they indeed exhibits more upward trend after 1900 and we can believe that there exists an upward temperature trend in spring and autumn.

As for smooth spline method, it exhibits less cyclical patterns for winter and more cyclical pattern for spring, but it shows almost same whole trend as lowess method.

5 Conclusion

Based on different analysis methods, we conclude that annual temperature exhibits pattern of decreasing, increasing, stable and then increasing trend, which is proved by both parametric and nonparametric method. So we believe that the annul temperature is getting warm in 2000s. As for season temperatures, they behaves slightly differently. Spring and autumn temperature shows almost same trend pattern as annual temperature. However, summer and winter temperature tends to be stable, at least we have not enough evidence to believe that winter and summer temperature is getting warm, especially for summer temperature. To sum, the annual average climate is getting warm in 2000s, but it exhibits slightly different for different season.

References

[1] https://www.scientificamerican.com/article/did-global-warming-slow-down-in-the-2000s-or-not/

[2] Cleveland, William S. "Robust locally weighted regression and smoothing scatterplots." Journal of the American statistical association 74, no. 368 (1979): 829-836.

[3] Harvey, David I., and Terence C. Mills. "Modelling trends in central England temperatures." Journal of Forecasting 22, no. 1 (2003): 35-47.

[4] Shumway, Robert H., and David S. Stoffer. Time series analysis and its applications: with R examples Fourth Edition. Springer Science & Business Media, 2016.