Final Project

STAT 26100

Nicolas Vila

Introduction.

In the following sections, I'll present an analysis of the Central England Temperature (CET) series and I'll be addressing the main questions about current temperature trends.

Before doing so, I'll first present the dataset used in the aforementioned analysis. The CET records mean monthly temperatures from 1659 to date. It is possibly the longest temperature dataset recorded with over 360 years of data. Having such a huge sample of data makes it better to get more robust and reliable results. I have to note first that the dataset didn't include mean temperature values for November and December of 2022. For the sake of simplicity while analysing the time series, I used the mean temperature for those months in 2021 as the mean temperature in 2022. I will use different modelling techniques such as ARIMA models and spectral analysis to try to observe past and future trends in the temperature of Central England.

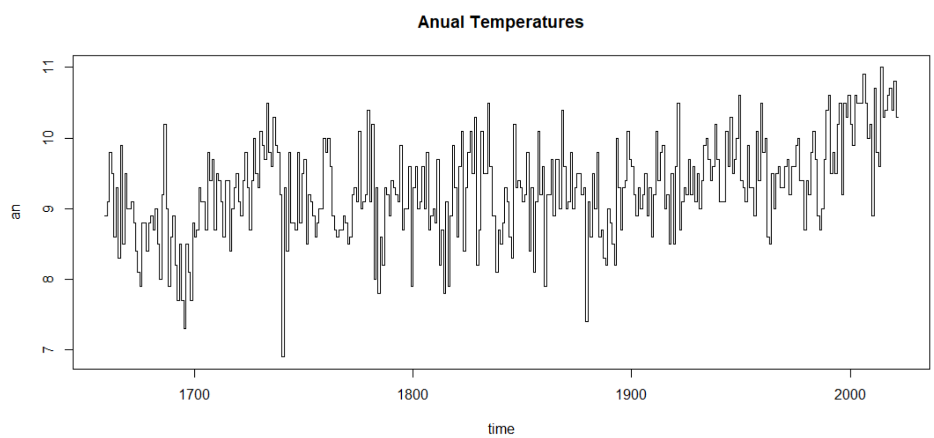


Figure 1

1. Long Term Trend Analysis.

An interesting observation of Figure 1.1 is that the time series seems to have different trends. Moreover, it seems like in the past 50 years, the trend is undoubtedly increasing faster. To observe such phenomena I will fit smoothing techniques to this dataset to show the previous trends which will set us in a starting point of understanding to the process that I'll be modelling it in the next sections.

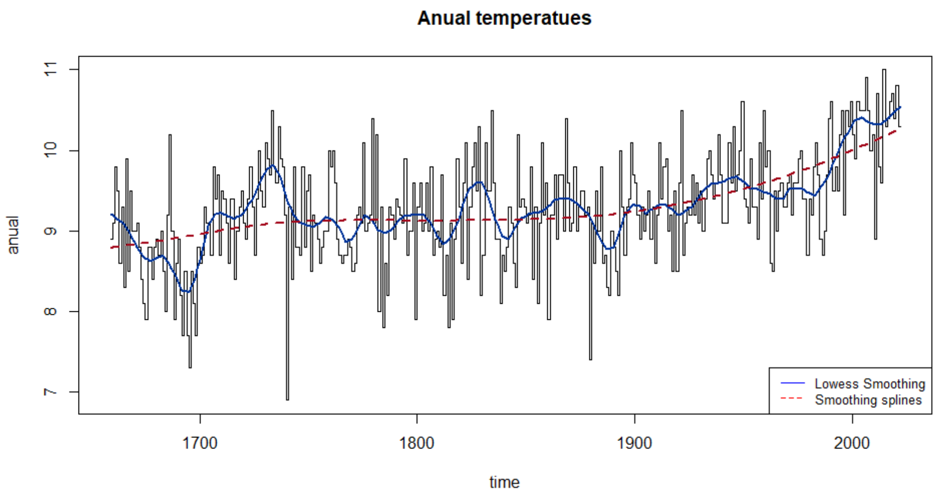


Figure 2

As it's seen from the plot above, there's clearly an increasing trend, growing faster than in the last 250 years approximately(there's an important increasing trend starting in the 1700s). An important part of this work is to model this dataset and being able to interpret the results to answer questions such as if global warming slowed down in the 2000s, or if there has been increasingly warmer temperatures during the past years. All of those questions will be addressed at the end of the report.

2. Analysis with ARIMA models.

2.1 Annual Data analysis

The first tool I'll analyse the annual process with will be the ARIMA(p,d,q) model. Before fitting the model, it is necessary that we perform a data transformation such that we obtain a non-stationary process. To do so, I'll perform a first difference transformation as it can be seen in Figure 1.2.
Let x_t be the time series. Then, let the transformed time series be $y_t = x_t - x_{t-1}$.

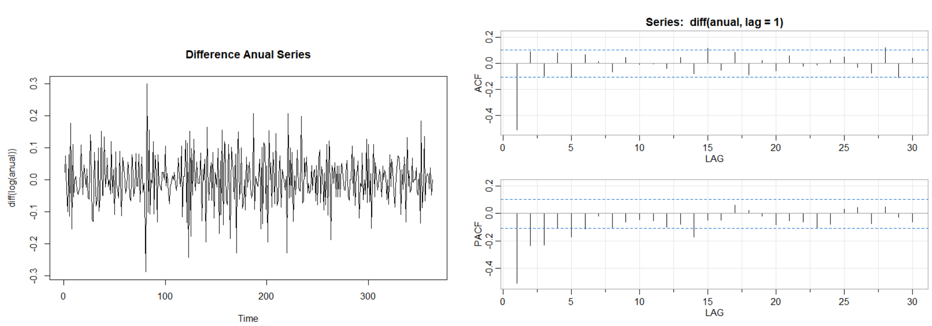


Figure 3

After observing such plots and looking at Figure 1.3 one could argue that the ACF plot cuts-off after lag=1 and PACF tails off. However, we shall try different order models and compare them in terms of AICc, AIC, BIC and mean error values. The chosen model will be the one with the lowest values.

	AR(p)	MA(q)	ARMA(p, q)
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

Figure 4

The tested models in this section will be the ARIMA(0,1,1), ARIMA(3,1,0), ARIMA(2,1,1) and ARIMA(3,1,1). We shall analyse independently the results of each model:

1. ARIMA(0,1,1):

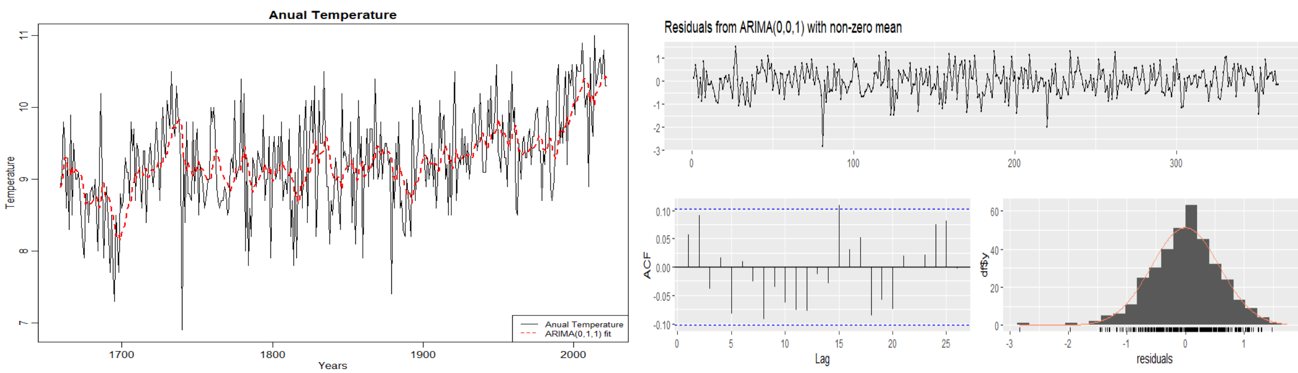


Figure 5

This model fits the data properly in a long term view. It can model the trends properly but can't fit the large variations in the short term. Regarding the residuals, we can say that they behave like white noise

2. ARIMA(3,1,1)

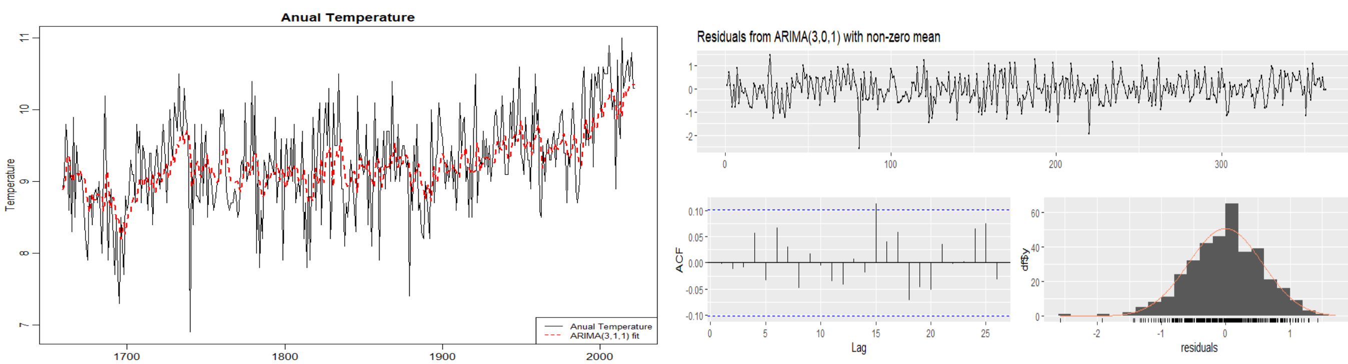


Figure 6

This fit behaves very similarly to the one seen before, the same conclusions are made here.

3. ARIMA(3,1,0)

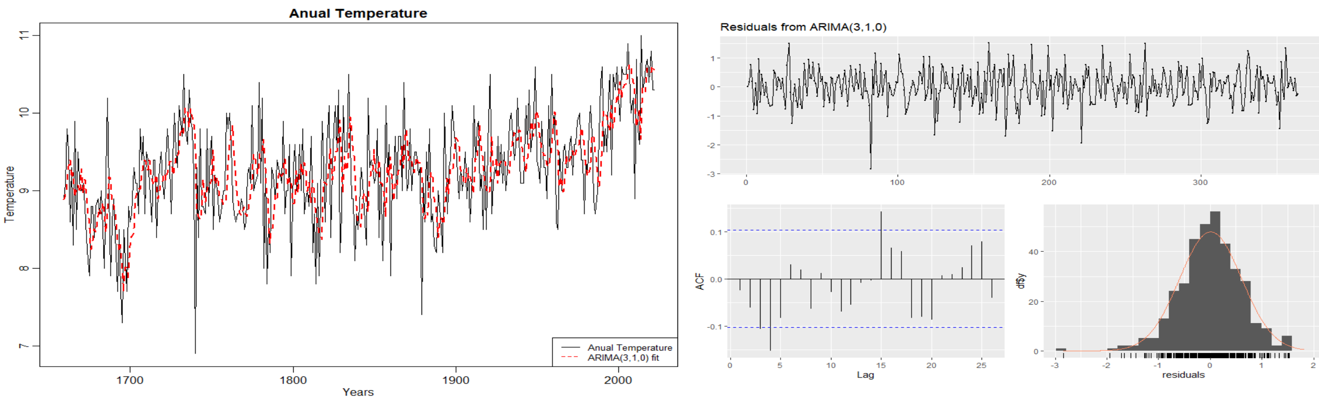


Figure 7

In this case, the fit looks more true to the data than the other models. There are also small correlations in the residuals. Nevertheless one could reasonably argue that the residuals behave very much like white noise given the diagnostics above.

Following we can observe a table with different performance parameters of every model.

	AICc	AIC	BIC	ME
ARIMA(0,1,1)	642,4774	642,4441	654,1273	0,022118
ARIMA(3,1,0)	671,1679	671,0561	690,5281	0,0088923
ARIMA(3,1,1)	638,2026	638,0344	661,4008	0,03454305
ARIMA(2,1,1)	636,2151	636,1033	655,5753	0,03444805

Table 1

From the table we can see that there’s a trade off between AICc, AIC, BIC and mean error. Given these parameters, I will use a model that has a balanced trade-off between such parameters. The chosen model will therefore be an ARIMA(0,1,1). I will use this method to forecast the time series.

	Predicted Value	Low Interval	High Interval
2023	10,682	9,502419	11,86151
2024	10,81126	9,573742	12,04877
2025	10,71047	9,383721	12,03722
2026	10,85029	9,452911	12,24767

Table 2

The mean temperature value in 2022 was 11.2ºC. However, the model fitted indicated a mean temperature of 10.56ºC for 2022. Therefore, I shall take this value instead of the real value as a benchmark to compare the predicted temperatures. As we can see in the table above, the predicted temperatures for 2023 and 2024 indicate an increasing trend. The model predicts a small slow down in 2025 but then increases again in 2026. Even though this model is susceptible to errors(we can see that the confidence intervals are quite wide), what we can take from the predictions is that we expect the mean annual temperatures in Central England to keep increasing in the next few years.

2.2 Monthly Data Analysis

On the other hand, we can analyse the monthly data of the time series. Since it ranges over a long period of time, plotting monthly data over +350 isn't very informative. However, it is important to check ACF and PACF of such series:

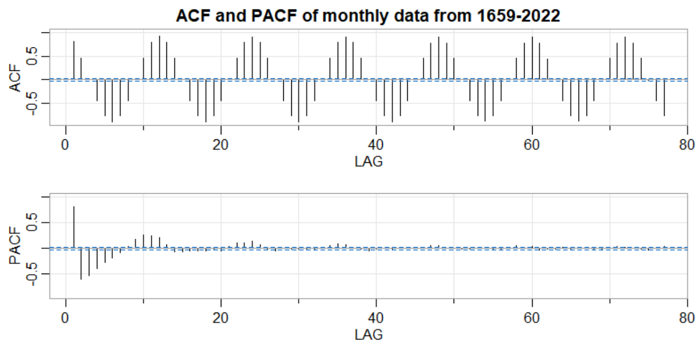


Figure 8

We clearly observe a sinusoidal form in the ACF implying that the data is seasonal. Indeed, it is seasonal since every season of each year will be highly correlated with the data from the same season in the other years(we can observe a seasonal period with a peak at lag 12). In order to model this behaviour we could firstly apply a difference to the original time series of lag = 12. After doing so, we can see that the ACF loses the periodical form that we saw previously.

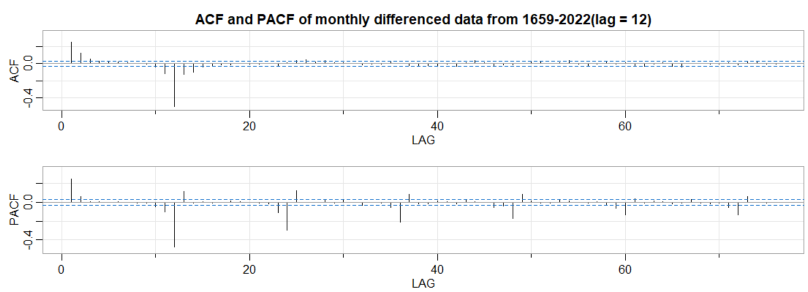


Figure 9

We can observe two peaks in the ACF, the first of them at lag 1 and a second one at lag 12. Since one could interpret the graphs saying that the ACF is cutting off after lag 12 while the PACF tails off, one could argue about fitting an MA(12) model to the data. However, I believe that trying to model the monthly seasonal process would be much harder and less informative than analysing each season independently.

Therefore a more interesting exercise is to also analyse each season independently and try to model them in order to obtain a better understanding from the data without a season factor. Hereinafter, I will do the same as I've been doing but with the data from the Winter(December, January and February), Spring(March, April, May), Summer(June, July, August) and Fall(September, October, November) independently from each other. Each season time series has been calculated as a mean between the monthly data.

To begin with, I’ll simply plot the different series using a 75% exponential moving average and a lowess smoothing method. From such graphs we can detect that there’s an undoubted increase in the temperatures in all seasons from the 1990s to date.

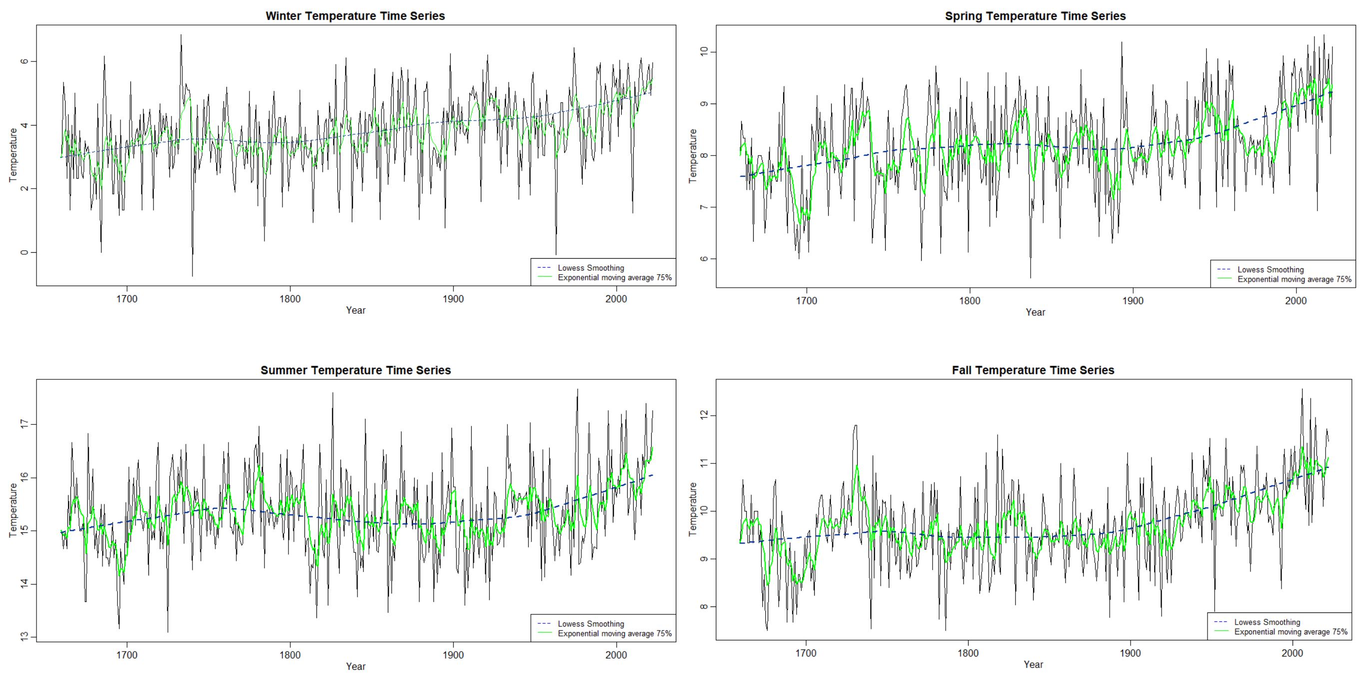


Figure 10

In addition, I shall analyse the ACFs of the first difference, which I argue that it transforms the series into a non-stationary one. From there, it will be possible to choose an ARIMA model that fits the data correctly.

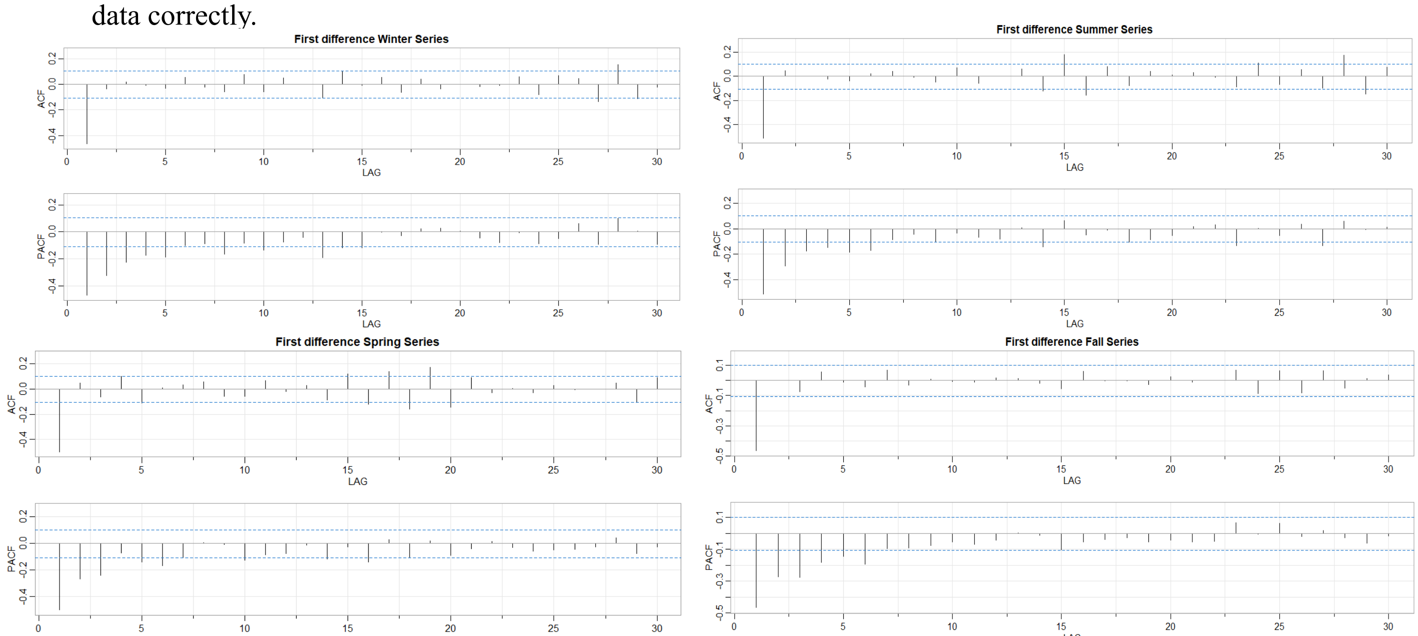


Figure 11

For these time series I tested several orders in order to find the optimal ARIMA model that fits each process better. The model fitted for the Winter series is an ARIMA(1,1,0). We can judge by the plot below that it is a fairly accurate model of the original series even though it can’t capture some of the spikes. The residuals behave like white noise which can be seen from the diagnostics below.

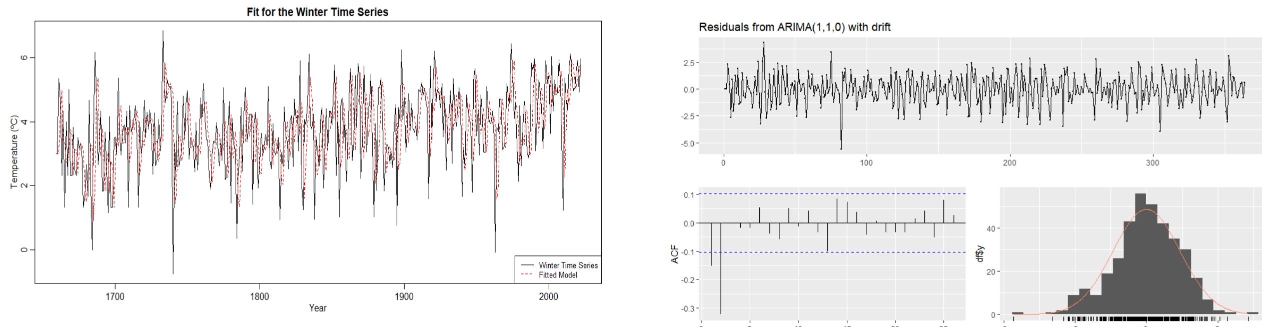


Figure 12

As for the Spring series, the model utilised was an ARIMA(3,1,0). We can observe that the fit doesn’t capture most of the peaks. However, it does capture the short term trends. Regarding the residuals, even though a small correlation can be seen at lag 4, we could interpret the residuals as white noise.

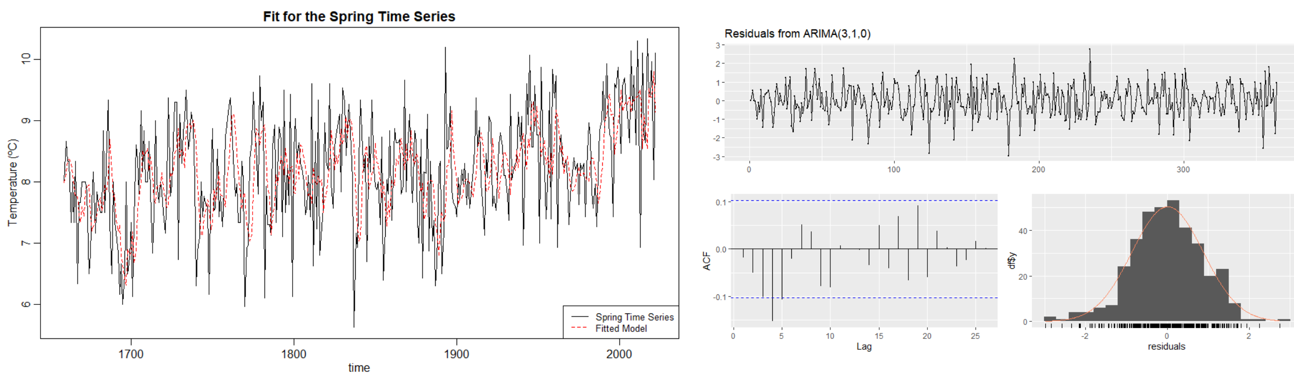


Figure 13

The results for the Summer and Fall seasons look as convincing as the Winter and Spring ones and their fits can be similarly interpreted. We can see such results in Figure 2. and Figure 2.

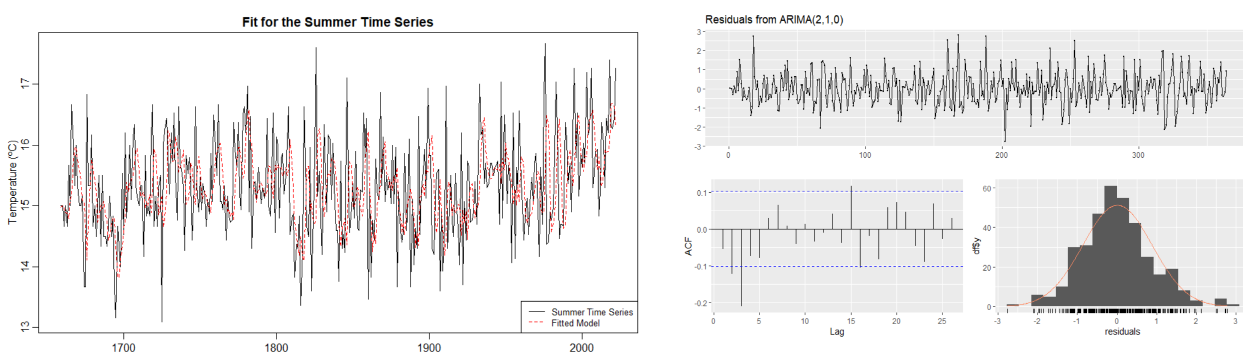
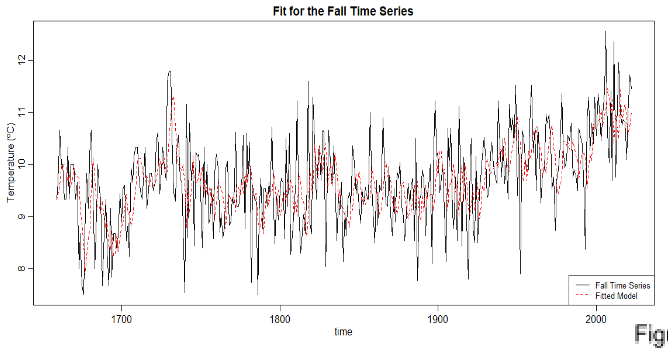


Figure 14



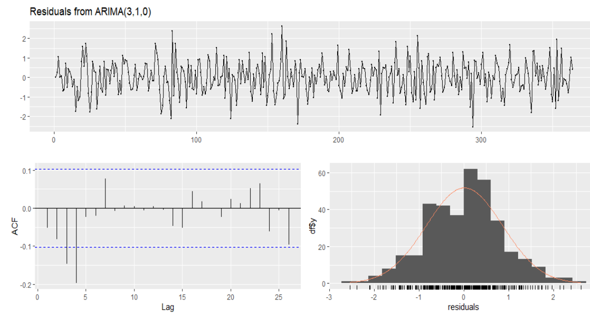


Figure 15

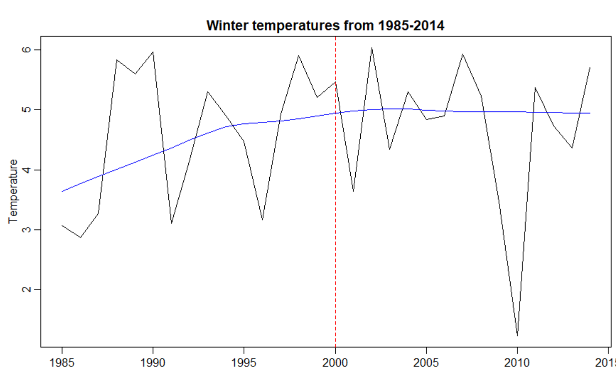
Given the fitted models, I’ve used them in order to predict the temperature values in Central England in the years 2023, 2024, 2025 and 2026. In the table below we can see that the predicted values for such years are lower than the real temperature value in 2022. However, this doesn’t mean that the temperatures will drop. These results counterintuitively imply that the temperatures will rise.
This is because if we take into account as a starting point the predicted temperature of 2022, we can see that the future predictions are in all seasons rising quite fast. Therefore, these results don’t show the exact value of temperatures in the future but show the rising trend that temperatures will experience.

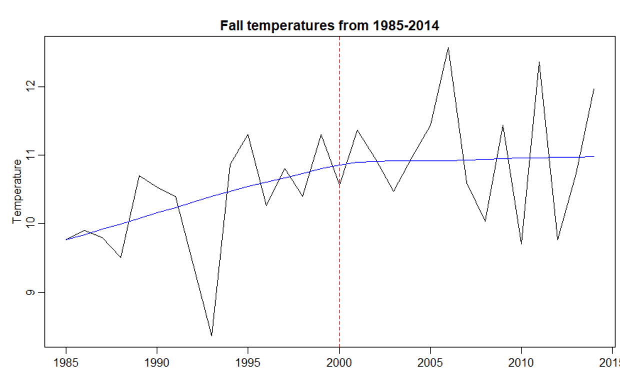
	Real Temperature 2022
Winter	5,96667
Spring	10,1
Summer	17,26667
Fall	11,46667

	Winter			Spring			Summer			Fall		
	Predicted Value	Low Interval	High Interval	Predicted Value	Low Interval	High Interval	Predicted Value	Low Interval	High Interval	Predicted Value	Low Interval	High Interval
2022	**5,394188**			**9,143453**			**16,3422**			**11,04507**		
2023	5,496107	2,7354	8,2556786	9,332381	7,6255	11,03923	16,66297	14,9502	18,37573	11,06716	9,4011	12,73318
2024	5,725921	2,5959	8,855938	9,468287	7,6847	11,25183	16,81789	15,0117	18,62407	11,23764	9,4821	12,99315
2025	5,629566	1,8749	9,384225	9,196792	7,3046	11,08898	16,89354	14,9081	18,87894	11,3733	9,5290	13,21761
2026	5,68511	1,5235	9,846743	9,514099	7,5171	11,51114	16,79735	14,5767	19,01798	11,31906	9,3851	13,25301

Table 3

Moreover it has been a controversial statement, the one made by the NOAA group saying that global warming didn’t slow down in the 2000s. In Figure 2. can be seen every season’s mean temperature from 1985 to 2014 in order to see its behaviour before and after the year 2000. I have also plotted a lowest smoothing method to the mentioned time series to observe changes in the trend. It becomes apparent that except for the Spring season, the trend clearly slowed in the 2000s.
We can only state that this happened in the Central England region and can’t make any kind of generalisation that this behaviour was the same in global temperatures. I believe I can neither affirm nor refute such a statement because analysing data from Central England one can’t generalise for the entire planet. Nevertheless, I must say that such a slow down in the rising temperatures trend during the 2000s indeed happened in the region of Central England.





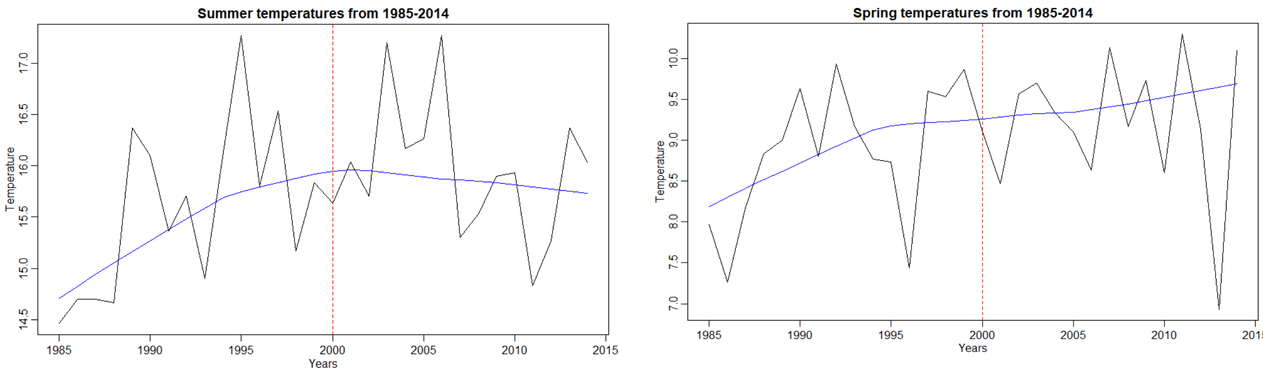


Figure 16

We can conclude this section by stating that from the analysis and modelling performed to this series, we can affirm that we are experiencing a trend of warmer temperatures every year in all seasons and especially, the trend increases even faster starting in the 1990s to date. It is reasonable to say that the mentioned increasing trend slowed down during the 2000s but that doesn't change the bigger picture, which is that we are experiencing warmer temperatures every year.

3. Spectral Observations

We've seen in section 2.2 that the sequential monthly data from 1659-2022 had an important seasonal component and its ACF behaved in a sinusoidal form. A good method to analyse this type of periodic signals is to plot the spectrum of the series to find its main frequencies. From there we could model a time series with the frequencies found.

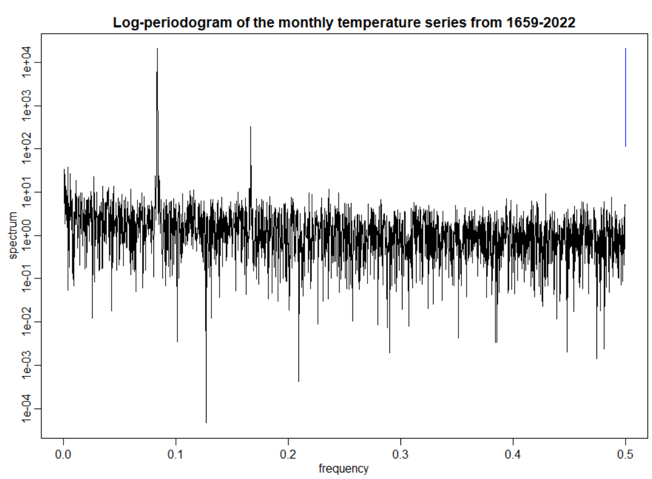


Figure 17

We can clearly see two dominant frequencies in the spectrum. The larger one is at frequency 0.083 and the smaller at 0.167. This implies that there are two main periodic signals, one with a period of 12 months and the other with a period of 6 months. I must note that it is a log spectrum, which means that the larger peak is actually much larger than the second one.

What I mean with this is that the period of 12 months is more important than the one of 6. However, such analysis in the frequency domain just affirms what we already knew, which is that there was a periodic component with a period of one year. This is because temperatures recorded one year apart from each other are very likely to be similar than the ones recorded shortly after.

3. Conclusion

Throughout the different analysis performed to this dataset both annually, monthly and seasonal, I've come up to the conclusion that the increasing trend was clear in every analysis made. All the predictions made by every model fitted predicted an increase in the temperatures in the following years which is somewhat concerning. Nevertheless this doesn't mean that the rate at which temperatures rise can't slow down and that nothing can be done to prevent that, as happened in the 2000s in the Central England region, where we experienced a slow down in the rising trend.

Another important note to take into account is that the models presented here are very simple models in comparison to state of the art models that try to predict future trends and temperatures. Therefore the reliability of the models presented in this report is very questionable given that temperatures may be influenced by many other parameters other than past temperatures.

4. References

1. JONES, P. D. & HULME, M. (1997). The changing temperature of Central England. In Climates of the British Isles, 460 Present, Past and Future, Ed. M. Hulme and E. Barrow, London: Routledge. pp. 173ā95.

2. JONES, P. D. & BRADLEY, R. S. (1992a). Climatic variations in the longest instrumental records. In Climate Since A.D. 1500, Ed. R. S Bradley and P. D. Jones, London: Routledge. pp. 246ā68.

3. BENNER, T. C. (1999). Central England temperatures: Long-term variability and teleconnections. Int. J. Climatol. 19, 391ā403.

4. HARVEY, D. I. & MILLS, T. C. (2003). Modelling trends in central England temperatures. J. Forecasting 22, 35ā47.

5. JONES, P. D. & BRADLEY, R. S. (1992b). Climatic variations over the last 500 years. In Climate Since A.D. 1500, Ed. R. S Bradley and P. D. Jones, London: Routledge. pp. 649ā65

Appendix

A.1: Annual Analysis

```
>>plot.ts(time, anual, type="single",main ="Anual temperatures")
>>lines(lowess(time,anual, f=.05), lwd=2, col="#003399")
>>lines(smooth.spline(time, anual, spar= 1), lty=2, lwd=2, col="#A3081B")
>>legend("bottomright", legend=c("Lowess Smoothing", "Smoothing
splines"),col=c("blue", "red"), lty=1:2, cex=0.8)

>>plot.ts(diff(log(anual)), main="Difference Anual Series")
>>acf2(diff(anual,lag=1)) ## Cuts off after lag 1
>>x = diff(anual)
>>reg_1 = arima(anual, order=c(0,1,1))
>>reg_2 = arima(anual, order=c(3,1,0))
>>reg_3 = arima(anual, order=c(3,1,1))
>>reg_4 = arima(anual, order=c(4,1,2))
>>reg_5 = arima(anual, order=c(2,1,1))
>>plot.ts(time, anual, type="l", main="Anual Temperature", xlab="Years",
ylab="Temperature")
>>lines(time, fitted(reg_2), col="red", lwd=2, lty=2)
>>plot.ts(time, anual, type="l", main="Anual Temperature", xlab="Years",
ylab="Temperature")
>>lines(time, fitted(reg_5), col="red", lwd=2, lty=2)
>>plot.ts(time, anual, type="l", main="Anual Temperature", xlab="Years",
ylab="Temperature")
>>lines(time, fitted(reg_1), col="red", lwd=2, lty=2)
>>plot.ts(time, anual, type="l", main="Anual Temperature", xlab="Years",
ylab="Temperature")
>>lines(time, fitted(reg_3), col="red", lwd=2, lty=2)

>>aicc = function(model){
  n = model$nobs
  p = length(model$coef)
  aicc = model$aic + 2*p*(p+1)/(n-p-1)
  return(aicc)
}

>>forecast::checkresiduals(reg_1)
>>summary(reg_1)
>>reg_1$aic
>>forecast::checkresiduals(reg_2)
>>summary(reg_2)
>>reg_2$aic
>>forecast::checkresiduals(reg_3)
>>summary(reg_3)
>>reg_3$aic
>>forecast::checkresiduals(reg_5)
>>summary(reg_5)
>>reg_5$aic
```

A.2: Monthly Analysis

```
#Variable an_m: monthly sequential data
>>plot.ts(an_m)
>>acf2(an_m)
>>acf2((diff(an_m, lag=12)), main ="ACF and PACF of monthly differenced data from
1659-2022(lag = 12)")
>>fit1 = arima(diff(an_m, lag=12), c(0,0,12))
>>forecast::checkresiduals(fit1)
```

```
>>fit2 = arima(diff(an_m, lag=12), c(0,0,1))
>>forecast::checkresiduals(fit2)

>>fit3 = arima(diff(an_m, lag=12), c(1,0,13))
>>forecast::checkresiduals(fit3)

>>fit1$aic
>>fit2$aic
>>fit3$aic

#WINTER
>>jan <-meantemp_monthly_totals[[2]]
>>feb <-meantemp_monthly_totals[[3]]
>>dec <-meantemp_monthly_totals[[13]]
>>dec[364]=dec[363]
>>winter = (jan + feb+dec)/3
>>acf2(diff(winter))
>>ema=c(winter[1])
>>for(i in 2:364){
  ema[i]=(1-0.75)*winter[i]+0.75*ema[i-1]
}
>>par(mfrow=c(1,1))
>>plot.ts(time,winter, type="l",xlab="Year", ylab="Temperature", main="Winter
Temperature Time Series")
>>lines(time,ema, col="green")
>>lines(lowess(time,winter, f=.40), lty=2, col="#003399")
>>legend("bottomright", legend=c("Lowess Smoothing", "Exponential moving average
75%"),col=c("blue", "green"), lty=2:1, cex=0.8)

#SPRING
>>mar <-meantemp_monthly_totals[[4]]
>>apr <-meantemp_monthly_totals[[5]]
>>may <-meantemp_monthly_totals[[6]]
>>spring = (mar+apr+may)/3
>>plot.ts(spring)

>>ema=c(spring[1])
>>for(i in 2:364){
  ema[i]=(1-0.75)*spring[i]+0.75*ema[i-1]
}
>>par(mfrow=c(1,1))
>>plot.ts(time,spring, type="l",xlab="Year", ylab="Temperature", main="Spring
Temperature Time Series", lwd=1.25)
>>lines(time,ema, col="green", lwd= 2)
>>lines(lowess(time,spring, f=.40), lty=2, col="#003399", lwd=2)
>>legend("bottomright", legend=c("Lowess Smoothing", "Exponential moving average
75%"),col=c("blue", "green"), lty=2:1, cex=0.8)

#SUMMER
>>jun <-meantemp_monthly_totals[[7]]
>>jul <-meantemp_monthly_totals[[8]]
>>aug <-meantemp_monthly_totals[[9]]
>>summer = (jun+jul+aug)/3
>>plot.ts(summer)
>>ema=c(summer[1])
>>for(i in 2:364){
  ema[i]=(1-0.75)*summer[i]+0.75*ema[i-1]
}
>>par(mfrow=c(1,1))
```

```
>>plot.ts(time,summer, type="l",xlab="Year", ylab="Temperature", main="Summer
Temperature Time Series", lwd=1.25)
>>lines(time,ema, col="green", lwd= 2)
>>lines(lowess(time,summer, f=.40), lty=2, col="#003399", lwd=2)
>>legend("bottomright", legend=c("Lowess Smoothing", "Exponential moving average
75%"),col=c("blue", "green"), lty=2:1, cex=0.8)


#FALL
>>nov <-meantemp_monthly_totals[[12]]
>>nov[364]=nov[363]
>>sep <-meantemp_monthly_totals[[10]]
>>oct <-meantemp_monthly_totals[[11]]
>>fall = (sep+oct+nov)/3
>>plot.ts(fall)
>>ema=c(fall[1])
>>for(i in 2:364){
  ema[i]=(1-0.75)*fall[i]+0.75*ema[i-1]
}
>>par(mfrow=c(1,1))
>>plot.ts(time,fall, type="l",xlab="Year", ylab="Temperature", main="Fall Temperature
Time Series", lwd=1.25)
>>lines(time,ema, col="green", lwd= 2)
>>lines(lowess(time,fall, f=.40), lty=2, col="#003399", lwd=2)
>>legend("bottomright", legend=c("Lowess Smoothing", "Exponential moving average
75%"),col=c("blue", "green"), lty=2:1, cex=0.8)

>>par(mfrow=c(1,1))
>>acf2(diff(winter), main="First difference Winter Series")
>>acf2(diff(spring), main="First difference Spring Series")
>>acf2(diff(summer), main="First difference Summer Series")
>>acf2(diff(fall), main="First difference Fall Series")

>>fitwinter = auto.arima(winter, d=1)
>>par(mfrow=c(1,1))
>>plot.ts(type="l",time, winter, xlab="Year", main="Fit for the Winter Time Series",
ylab="Temperature (°C)")
>>lines(time,fitted(fitwinter), col='Red',lty=2)
>>legend("bottomright", legend=c("Winter Time Series", "Fitted Model"),col=c("black",
"red"), lty=1:2, cex=0.8)
>>fc = (forecast(fitwinter, h = 4))
>>print(fc)
>>accuracy(fitwinter)
>>summary(fitwinter)
>>forecast::checkresiduals(fitwinter)
>>print(fitted(fitwinter)[364])

>>fitspring = arima(spring, c(3,1,0))
>>par(mfrow=c(1,1))
>>plot.ts(time, spring, main="Fit for the Spring Time Series", ylab="Temperature
(°C)", type="l")
>>lines(time, fitted(fitspring), col='Red',lty=2)
>>legend("bottomright", legend=c("Spring Time Series", "Fitted Model"),col=c("black",
"red"), lty=1:2, cex=0.8)
>>fc = (forecast(fitspring, h = 4))
>>print(fc)
>>summary(fitspring)
>>print(fitted(fitspring)[364])
>>forecast::checkresiduals(fitspring)

>>fitsummer = arima(summer, c(2,1,0))
```

```
>>par(mfrow=c(1,1))
>>plot.ts(time,type="l",summer, main="Fit for the Summer Time Series",
ylab="Temperature (°C)")
>>lines(time, fitted(fitsummer), col='Red',lty=2)
>>legend("bottomright", legend=c("Summer Time Series", "Fitted Model"),col=c("black",
"red"), lty=1:2, cex=0.8)
>>fc = (forecast(fitsummer, h = 4))
>>print(fc)
>>print(fitted(fitsummer)[364])
>>forecast::checkresiduals(fitsummer)
>>summary(fitsummer)

>>fitfall = arima(fall, c(3,1,0))
>>par(mfrow=c(1,1))
>>plot.ts(time, type='l',fall, main="Fit for the Fall Time Series", ylab="Temperature
(°C)")
>>lines(time, fitted(fitfall), col='Red',lty=2)
>>legend("bottomright", legend=c("Fall Time Series", "Fitted Model"),col=c("black",
"red"), lty=1:2, cex=0.8)
>>fc = (forecast(fitfall, h = 4))
>>print(fc)
>>summary(fitfall)
>>forecast::checkresiduals(fitfall)
>>print(fitted(fitfall)[364])

>>time = c(1985, 1986, 1987, 1988, 1989, 1990, 1991, 1992, 1993, 1994, 1995, 1996,
1997, 1998, 1999)
>>Years = c(time, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010,
2011, 2012, 2013, 2014)

>>plot(Years, winter[327:356], type="l", main="Winter temperatures from 1985-2014",
ylab="Temperature")
>>abline(v=2000, col="red", lty=2)
>>lines(lowess(Years, winter[327:356]), col="blue")

>>plot(Years, fall[327:356], type="l", main="Fall temperatures from 1985-2014",
ylab="Temperature")
>>abline(v=2000, col="red", lty=2)
>>lines(lowess(time2, fall[327:356]), col="blue")

>>plot(Years, spring[327:356],  type="l", main="Spring temperatures from 1985-2014",
ylab="Temperature")
>>abline(v=2000, col="red", lty=2)
>>lines(lowess(time2, spring[327:356]), col="blue")

>>plot(Years, summer[327:356],  type="l", main="Summer temperatures from 1985-2014",
ylab="Temperature")
>>abline(v=2000, col="red", lty=2)
>>lines(lowess(time2, summer[327:356]), col="blue")
```

A.2: Spectrum

```
>>sp = spec.pgram(an_m, log="yes", main="Periodogram of the monthly temperature series
from 1659-2022")

>>sorted = sort(sp$spec, decreasing = TRUE)[c(1,2,3,4,7)];sorted
#Largest peaks
>>p1 = sp$freq[sp$spec==sorted[1]];p1
>>p3 = sp$freq[sp$spec==sorted[5]];p3
```