Analysis and Prediction of Precipitation Model in Exeter

Bochen Ren

690054351

Analysis and Prediction of Precipitation Model in Exeter

Abstract

In recent years, due to global warming, the global precipitation situation has undergone significant changes. These changes are mainly reflected in socio-economic and ecological issues, especially in areas that have increased fragility. Therefore, it is necessary to estimate short-term and long-term future precipitation by analysing long-term observation station data. This essay aims to predict the precipitation in the next ten years by analysing the precipitation model in the Exeter region of the United Kingdom. In this research, I use time series to analyze data. Model fitting was performed on the data from 1921 to 2010, and then a model test was performed with the data from 2011 to 2020. Finally, the tested model was used to predict the next ten years. In addition, a linear relationship between temperature data and precipitation data is constructed to assist in the analysis of precipitation data. Through the comparison of HW, ARIMA and VAR, three forecasting models, the best model is selected for forecasting. The precipitation data is collected from 85,000 stations with data records over a decade around the world. This data is compiled and published by Global Precipitation Climatology Centre, and the temperature data is collected from the Physical Sciences Laboratory, which includes two sets of data, Global Historical Climatology Network version 2 MMSAT dataset and Climate Anomaly Monitoring System dataset. The analysis results in this essay show that the prediction results of the vector autoregressive model are more accurate than the other two methods and that the precipitation in Exeter will slowly increase in the next ten years. In addition, through the auxiliary analysis of temperature data, it is concluded that the rainfall in Exeter has a strong correlation with the temperature in Exeter, but the correlation with the overall temperature of England is weak. Moreover, this essay discusses and analyzes the relationship between temperature rising and extreme precipitation and concludes that to a certain extent that precipitation will increase with increasing temperature without considering the influence of other factors.

Keywords: time series, precipitation prediction, the Seasonal-Trend decomposition procedure based on Loess, the Kwiatkowski, Phillips, Schmidt, and Shin algorithm, the Holt-Winters algorithm, the Autoregressive Integrated Moving Average model, the vector autoregressive model

Table of Contents

1. Introduction

In the context of global climate change, climate factors have changed to varying degrees in both time and space. As one of the most important climatic factors affecting human life, rainfall has also undergone significant changes (Stocker, 2014). Rainfall variation directly affects floods, droughts, water resources planning, and safe use, making it one of the critical factors impacting the sustainable development of society (Wei, 2007). In terms of rainfall research, with the global warming and urbanization process, the global terrestrial rainfall has increased regionally (Field et al., 2012). The most severe impact is that humid areas are wetter, and the dry regions are drier. In other words, the place that initially had more precipitation tends to have more precipitation, and the place that initially had less precipitation tends to have less precipitation (Yang et al., 2013). The development of climate extremes in the context of global warming has also led to signficant changes in the temporal distribution of rainfall (Trenberth, 2011). Therefore, the research and prediction of rainfall models have played a vital role in the development of human society.

Prediction of precipitation is of great significance to human society. It will directly or indirectly affect many fields such as agriculture, biodiversity, and natural disaster warning. First, precipitation has a huge impact on agriculture, especially rainfall. Water is a necessary condition for the survival of all plants. As the most effective watering method, precipitation plays an integral role in agriculture. Although precipitation is very important to the health of plants, they require a specific amount. Too little precipitation can cause drought and cause a large number of plant deaths (Boken et al., 2005). Excessive precipitation will lead to over irrigation, and this humid environment will promote the overgrowth of harmful fungi and cause plant death (Burns, 2007). Different plants need different amounts of irrigation to survive. Therefore, the prediction of future rainfall can effectively help farmers choose crops that need to be planted later or protect current crops and avoid food shortages caused by extreme weather (Van Liere et al., 1994). Then, precipitation prediction plays a vital role in the study of biodiversity. Predicting the response of biodiversity to climate change is very important. Precipitation is a climatic factor that restricts complex areas. Therefore, we can explore biodiversity through the analysis and prediction of rainfall models, such as bird range-scale abundance changes (Illán et al., 2014). Last but not least, precipitation prediction also has critical applications for early warning of natural disasters. Precipitation mainly affects two kinds of natural disasters, one is drought, and the other is extreme rainfall. Extreme precipitation will cause many natural disasters such as floods, mudslides, landslides, and so on. The prediction of rainfall in advance can effectively reflect the occurrence probability of related natural disasters and accurately predict it to reserve time for human beings to make corresponding countermeasures to reduce the loss of human society due to natural disasters as much as possible.

This essay will focus on the precipitation model in Exeter from 1921 to 2020 and predict the precipitation of Exeter in the next ten years through the analysis of the model. In the first paragraph, I will briefly introduce the background of this research and why this research is worthy of doing. Next, I will study the research of others and analyze the research methods and results of others. Thirdly, the data in this research will be described, which include the source and processing method of data. Then, I will explain the research method of this experiment. I will show the predicted results and discuss them in the fifth and sixth parts. Last but not least, the fundamental research will be concluded.

2. Literature Review

2.1 Background

Precipitation plays an essential role in human society. According to data analysis, today's precipitation presents a gradual upward trend, and most of it is extreme precipitation. Now there are even heavy downpours where it never rains, and these phenomena increase the risk of flooding (Trenberth, 2011). The balance of atmospheric and surface energy plays a vital role in the hydrological cycle. Although the current precipitation pattern will not change much, it also causes arid regions to become arider and humid regions to become more humid (Bates et al., 2008). All of the above impacts have varying degrees of impact on various industries. Therefore, the research and prediction of rainfall patterns have become a crucial topic.

2.2 Main Methods

A good prediction is still the basis of all science (Makridakis & Hibon, 2000). Time series models are now widely used in various scientific fields, including hydrology. As one of the best methods, time series has an integral role in the analysis and prediction of rainfall models. Some of the main advantages of the time series model are due to the systematic search capabilities of identification, estimation, and diagnostic checks (Mishra & Desai, 2005). The time series model contains many analysis methods and forecasting methods, including data analysis methods: the Seasonal-Trend decomposition procedure based on Loess (STL), the Kwiatkowski, Phillips, Schmidt, and Shin algorithm (KPSS), and so on. and forecasting methods: the Holt-Winters algorithm (HW), the Autoregressive Integrated Moving Average model (ARIMA), the vector autoregressive model (VAR), and so on.
As for the prediction itself, if the model is specified incorrectly, they are always prone to wrong prediction results (Diebold & Lopex, 1996). Therefore, the choice of model is crucial for forecasting. This problem has been solved by using the functions of various predictive models to obtain specific aspects of data for the analysis and selection of models. The decomposition process can be analyzed by decomposing the time series into feature-based sub-

sequences to increase the accuracy of the model. Theodosius is based on decomposing time series components by STL decomposition program, decomposing the extrapolation of linear combinations of subsequences, and extrapolating re-aggregation to obtain estimates of the global sequence. They analyzed the time series of NN3 and M1 competition data and obtained more accurate results by comparing other standard statistical techniques, ARIMA, Theta, Holt-Winters' and Holt's Damped Trend methods (Theodosious, 2011). Moreover, Chen et al. (2020) pointed out that using the Seasonal-Trend decomposition based on Loess - the single long short-term memory (STL-LSTM), a hybrid algorithm, can reduce the impact of irregular fluctuations, and the result will be better than just using LSTM. Although the subject of this article is the short-term prediction of subway passenger flow, this algorithm is more effective in mitigating the impact of irregular fluctuations. In my research, the irregular fluctuations of precipitation are not so obvious. Thus, I chose to use the STL to process the data (Chen et al., 2020).

Thasneem et al. (2019) used KPSS to check the data stability of the representative concentration paths RCP 4.5 and 8.5 on the changes in extreme precipitation in the Chalyar River Basin under climate change. After discussing the uncertainty of the two sets of data sets, a collection of five global climate model regional climate model (GCM–RCM) combinations is used to obtain forecast results (Thasneem et al., 2019).

Besides KPSS, Philips and Perron test and Augmented Dickey Fuller test are also commonly used methods to test data stationarity. Um et al. (2018) compared statistical methods. They used four statistical tests, which included two stationarity tests, the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and Philips and Perron (PP) tests, and two trend tests, the Mann-Kendall (MK) and regression tests to evaluate Four time series types which included stationary in variance and trend in mean (S_T), stationary in variance and no trend in mean (S_NT), nonstationary in variance and trend in mean (NS_T), and nonstationary in variance and no trend in mean (NS_NT). After bringing them into the precipitation data of eight cities in the United States for observation to analyze and predict, the following conclusions have got (Um et al., 2018):

1. For S_T data, the PP test usually produces a better plateau result than the KPSS test.
2. For S_NT data, the smoothness rate of the PP test is higher than that of the KPSS test.
3. For NS_T data, as the sample size and slope increase, but the standard deviation decreases, the smoothness rate of KPSS and PP tests are both obvious declines
4. For the NS_NT data, the PP test produces better performance than the KPSS test in all cases.

HW is one of the exponential smoothing methods. The exponential smoothing method includes three methods, Single Exponential Smoothing Method, Double Exponential Smoothing and Triple Exponential Smoothing Method. Usually, we call Triple Exponential Smoothing Method HW (Kalekar, 2004). Raha and Gayen (2020) used the exponential smoothing model to carry out a meteorological drought simulation study in the Bangla district of Bengal, India. A comparative analysis of several time steps using standardized precipitation index (SPI) using double exponential (DE) and Holt–Winter exponential smoothing model (HW) found that DE is a more accurate model (Raha & Gayen, 2020). In addition, HW will also be used in the field of temperature prediction. Indriani et al. (2020) used HW to predict the temperature of the Cilacap Regency. By comparing the analysis results, they concluded that the HW prediction is accurate (Indriani., 2020).

The ARIMA method includes three methods: moving average process, autoregressive process and autoregressive moving average process (Box & Pierce, 1970). In other words, ARIMA is a new method optimized by moving average process, autoregressive process and autoregressive moving average process.

Bari et al. (2015) used ARIMA to analyze and predict rainfall in the Sylhet area. Similar to the method in this essay, for the data set from 1980 to 2010, they chose to use the data from 1980 to 2006 for model development and the data from 2007 to 2010 for model verification. Finally, they use the validated ARIMA model to make predictions (Bari et al., 2015).

Furthermore, there are many other methods similar to ARIMA, such as the Discrete Autoregressive Moving Average and Seasonal Autoregressive and Moving Average, which are optimized by the three basic methods. Chang et al. (1984) chose the discrete autoregressive moving average (DARMA) to analyze the daily river water data in Indiana. The structure of the DARMA series is specified by the marginal distribution of independent and identically distributed (iid) discrete random variables and their random linear combinations. A basic attribute of the DARMA series is that its related structure can be specified independently of its marginal density. Under the change of marginal density, the related structure of DARMA remains unchanged. Experimental results show that the Markov chain, a popular model for the daily precipitation sequences, is a specified model within the DARMA family (Chang et al., 1984).

Wang et al. (2013) found that rainfall has a strong auto-correlation time series in the seasonal characteristics, so using Seasonal Autoregressive and Moving Average (SARIMA) to statistically analyze the rainfall data in Shouguang City, Shandong Province, and get a better prediction fit Spend. And it is believed that the model can be used for actual precipitation forecast and warning (Wang et al., 2013)

Wiah (2017) used the VAR model to study the cocoa production in Ghana and studied cocoa production in Ghana through the four factors of maximum temperature, minimum temperature, precipitation and number of rainy days. Through the analysis of the VAR model, it is concluded that there is a causal relationship between the output of Ghanaian cocoa and the maximum temperature, the minimum temperature and the precipitation, and there is no causal relationship between the output and the number of rainy days. The maximum temperature and the number of rainy days have a negative impact on the yield, and the rainfall and the minimum temperature have a positive impact on the yield (Wiah, 2017).

In addition to time series, the Spatiotemporal Regression Kriging (STRK) method is one of the methods for predicting precipitation models. It processes regression and residual Kriging separately. Firstly, STRK predicts the regression part on a fine grid and then extracts the residuals of all observations. Next, it fits the global product sum variogram model and then inserts the residual error into the same grid and add it to the forecast trend. Finally, the final grid map is produced by residual Kriging. STRK can produce more accurate and detailed monthly precipitation forecasts by combining ground-measured precipitation and MODIS time series of NDVI images. However, different regions have different local trends. So local estimates should be made for different regions to analyze which predictive factors are more dominant in different parts of the study area and the degree of changes in these parameters in space and time. In addition, the high computational requirements are also the shortcomings of STRK. The regression of all subsets, the fitting of the variogram, the prediction of time and space, and the output map must go through a lot of calculations and consume a lot of time (Hu et al., 2017).

2.3 Contribution

I roughly categorize most of the forecasting methods in time series and select three methods that I think are more representative for forecasting comparison. First, I choose the three-time exponential smoothing method, HW, in the exponential smoothing method. Secondly, ARIMA is selected as the representative of basic methods moving average process, autoregressive process and autoregressive moving average process and other variant methods such as Seasonal Autoregressive Integrated Moving Average (SARIMA) and Seasonal Autoregressive Comprehensive Moving Average of Exogenous Variables (SARIMAX), etc. Finally, VAR was chosen as the representative of similar methods such as Vector autoregressive moving average (VARMA) and Vector autoregressive moving average of Exogenous Variables (VARMAX), etc. Furthermore, by searching the literature, it is found that VAR is usually used in other fields and has very few applications in the field of precipitation. Therefore, this article chooses to compare VAR with HW and ARIMA. The comparison results show that VAR has better prediction results. This result proves the feasibility of VAR in the field of rainfall prediction and fills the vacancy of VAR in the field of rainfall prediction.

In addition, this article also discusses the relationship between temperature and precipitation in depth. I will establish a connection between extreme weather and extreme weather, such as heatwaves and extreme precipitation, and conduct a more in-depth study on it. Nowadays, due to the frequent occurrence of extreme rainfall all over the world, and in some cases, we did not make correct predictions, which resulted in no time for us to deploy corresponding emergency plans to avoid the impact of extreme precipitation. Therefore, we urgently need to study the trigger conditions of extreme precipitation to predict extreme precipitation and reduce the casualties and losses caused by extreme precipitation to human society. Temperature is one of the main factors triggering conditions of extreme precipitation. A large amount of evidence shows that a heatwave may be accompanied by extreme precipitation. Therefore, it is very important to discuss the relationship between temperature and precipitation.

3. Data

In this essay, the data is divided into two parts: precipitation and temperature.

3.1 Precipitation Data

The precipitation data is from the Global Precipitation Climatology Centre (GPCC) (download link: https://opendata.dwd.de/climate_environment/GPCC/html/fulldata-monthly_v2020_doi_download.html). This data is based on the collection of approximately 85,000 sites around the world with recording durations of 10 years or more. The cover ranges are from 15000 to more than 50000 sites per month (Schneider et al., 2021). The data obtained from the GPCC official website is from 1900 to 2020. Plotting the Exeter data, we can find that the data from 1900 to 1920 is abnormal compared with all subsequent data (figure 1). Thus, in this research, I will use the data from 1921 to 2020 but not the whole data.

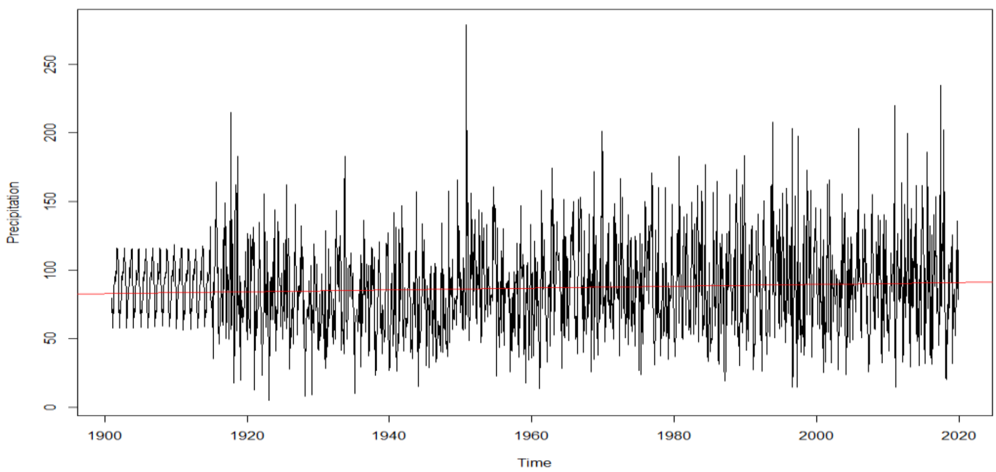

Figure 1

Figure 1| The precipitation in Exeter from 1900 to 2020.

As two cities in England, Exeter is a coastal city and London is a central city. Here we choose London as the main climate in England to compare with Exeter. From figure 1, it is clear to see that the precipitation in Exeter is generally significantly higher than that in London in these 80 years due to the fact that being close to the ocean has a very large impact on precipitation.

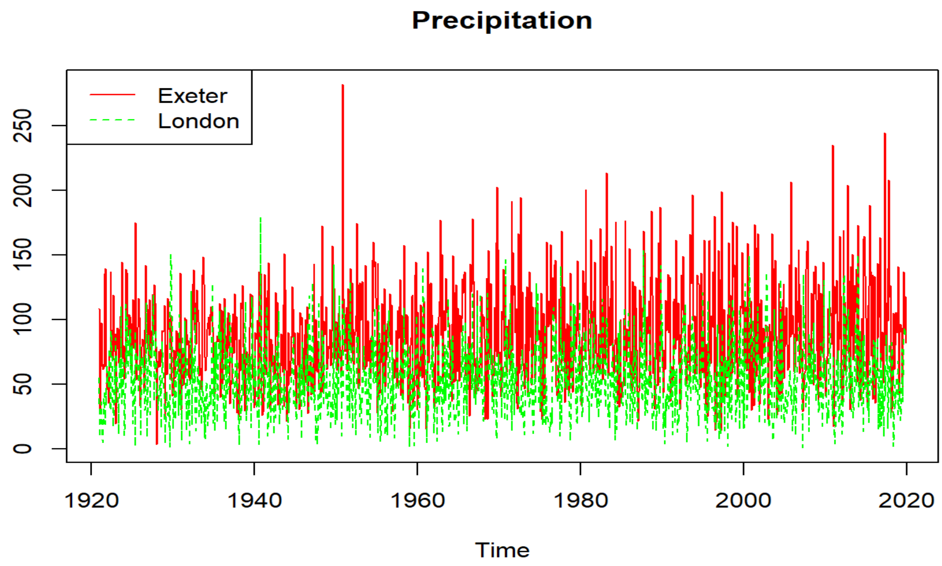

Figure 2

Figure 2 | The precipitation in Exeter and London from 1921 to 2020. The red line is the precipitation for Exeter and the green line is the precipitation for London.

The precipitation time series of Exeter and London is taken out separately for analysis. The overall trend in figure 3 is slowly rising. This result is the same as that reported by the Intergovernmental Panel on Climate Change (IPCC) in 2018. Although the annual increase is not obvious, IPCC pointed out in the report that rainfall in almost all parts of the world is increasing (IPCC, 2018). The peak appeared in 1950, and it is obviously higher than others which makes it seems strange. In the following part, I will try to find out the reason why the precipitation was so high in 1950.

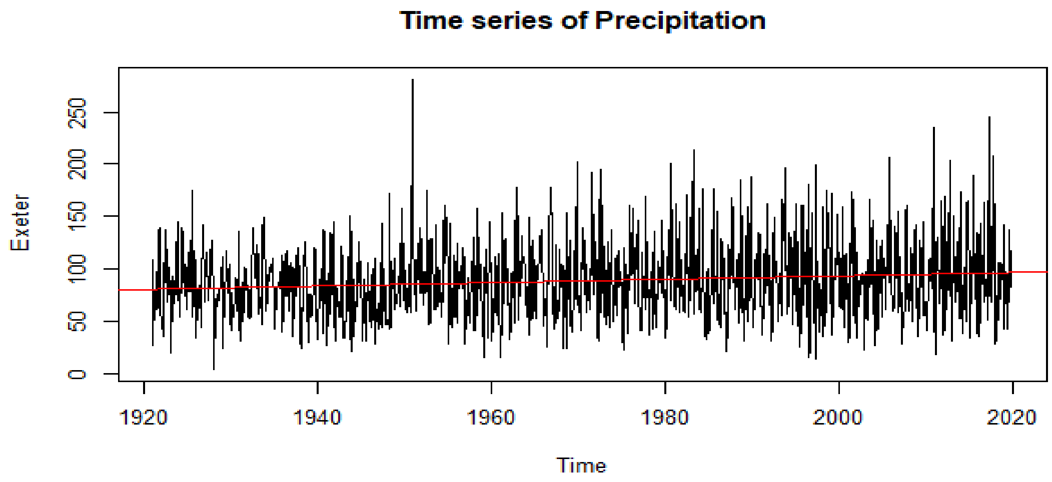

Figure 3

Figure 3| Time series of precipitation in Exeter from 1921 to 2020. The red line is the overall trend.

However, in London, the overall trend seems to be steady. In these 100 years, the precipitation has not changed significantly, unlike what it showed in Exeter. It is worth noting that there was also a peak in 1940, but this peak is not as prominent as the peak in Exeter in 1950. Thus, in this research, I will ignore this peak.

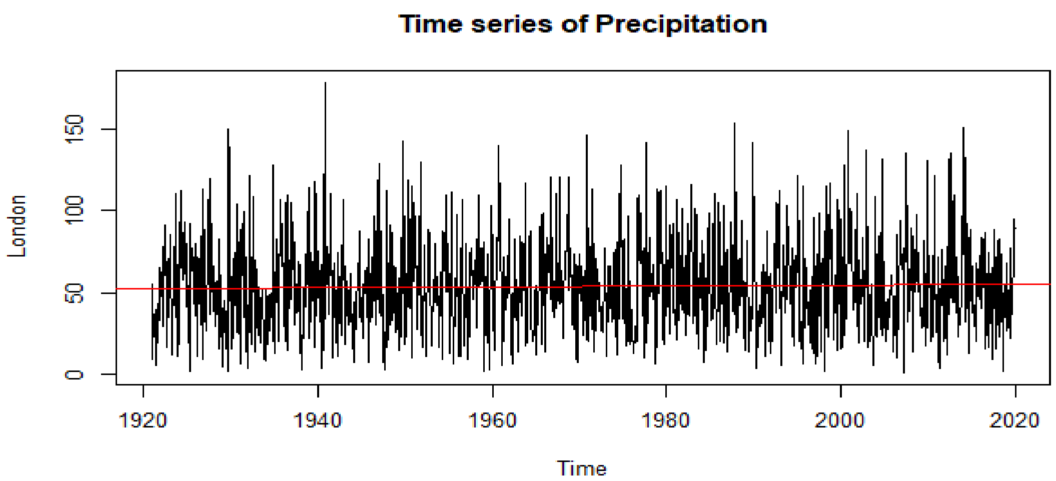

Figure 4

Figure 4| Time series of precipitation in London from 1921 to 2020. The red line is the overall trend.

By drawing a box plot (Figure 5) to analyze the difference of the data, we can see that the Exeter data from the highest value, upper quartile, median, lower quartile and lowest value are significantly higher than the London data. A part of higher values over the 1.5 times the difference value between the upper and lower four points in both Exeter and London.

Figure 5

Figure 5| The box plot for the precipitation data for Exeter and London. The red part is for Exeter and the blue part is for London.

3.2 Temperature Data

The temperature data is from the Physical Sciences Laboratory (PSL) (download link: https://psl.noaa.gov/data/gridded/data.ghcncams.html). This data includes two MMSAT station datasets. One is the Global Historical Climatology Network (GHCN) version 2 MMSAT dataset. GHCN has more than 30 different data sources, 7280 sites worldwide and a time scale of more than a century. GHCN will update regularly and add additional sites to the data set in some specific situations. The other is the Climate Anomaly Monitoring System (CAMS) dataset. CAMS is used in the Climate Prediction Center (CPC) of the National Center for Environmental Prediction (NCEP). The CAMS has 6158 sites all over the world, and some of them began to record the temperature in the 19th century. The CAMS station observations are based on two data sources. One comes from the historical records of the period before 1981 collected and edited by the National Center for Atmospheric Research (NCAR). Another real-time data is from the post-Global Telecommunication System (GTS) period. GAMS is also updated in real time. PSL merged the data from these two data sets after 1948 to get at this set of data (Fan & Van den Dool, 2008). The data obtained from the PSL official website is from 1948 to 2020. Since the precipitation in Exeter reached an abnormal maximum in 1950, in order to avoid the influence of the abnormal value on the research, I selected the temperature from 1951 to 2020. Note that because the temperature is the monthly average temperature, this temperature will be much different than the actual daily body temperature. This phenomenon is normal. Figure 6 shows the specific temperature in Exeter and London in these 70 years. The red line is the temperature data for Exeter and the green line is the temperature data for London. Different from the precipitation data, there is not much difference in temperature between these two places. It may be because both Exeter and London are in a temperate marine climate.

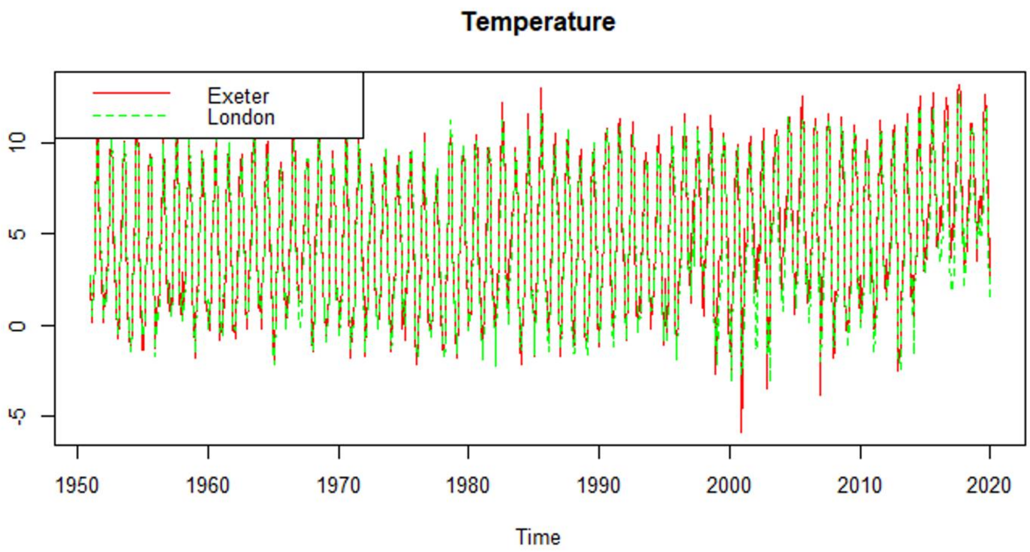

Figure 6

Figure 6| The temperature in Exeter and London from 1951 to 2020. The red line is for Exeter and the green line is for London.

I split and compare these two time series. In figure 7, the time series for Exeter, we can see that the maximum value appears around, but it is not significant than another high temperature month. However, the minimum value was very prominent in 2001. Due to the fact that the centre of this essay is focused on precipitation and the temperature data is an auxiliary study of precipitation, so we will not discuss it in depth here. The overall trend is rising, which coincides with the general trend of global warming.

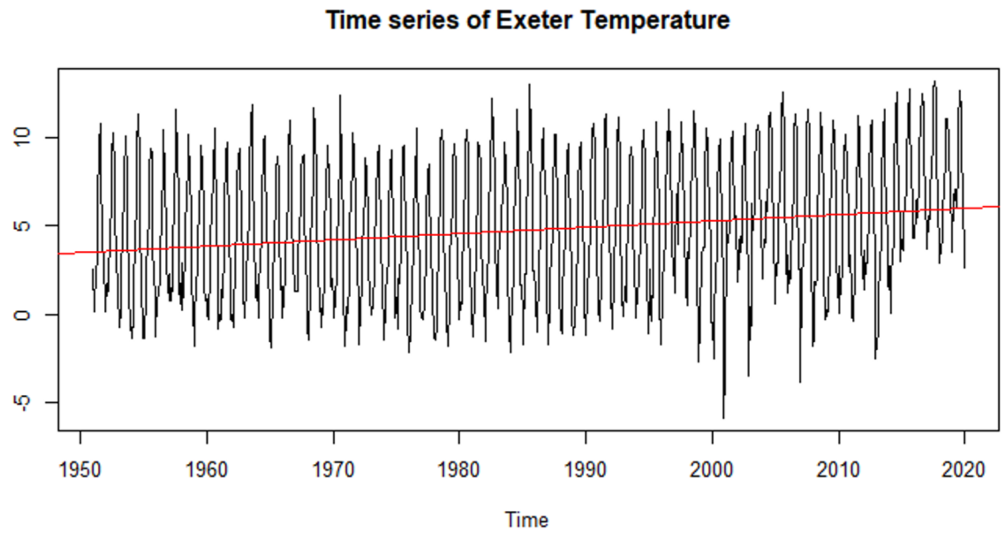

Figure 7

Figure 7| Time series of temperature in Exeter from 1951 to 2020. The red line is the overall trend.

Similarly, in figure 8, the maximum value appeared in 2018, and the minimum value appears in 2000 in London. Both of them are not significant. The overall trend is also increasing.

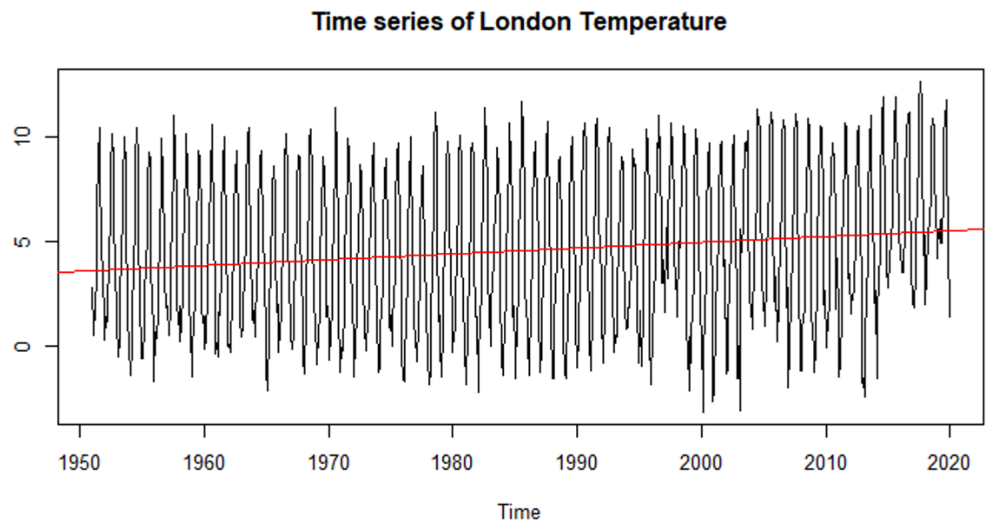

Figure 8

Figure 8| Time series of temperature in London from 1951 to 2020. The red line is the overall trend.

Different from the precipitation, the box plot in figure 9 shows that the temperature data in Exeter and London are similar in all the values. All values are within 1.5 times the difference value between the upper and lower four points.

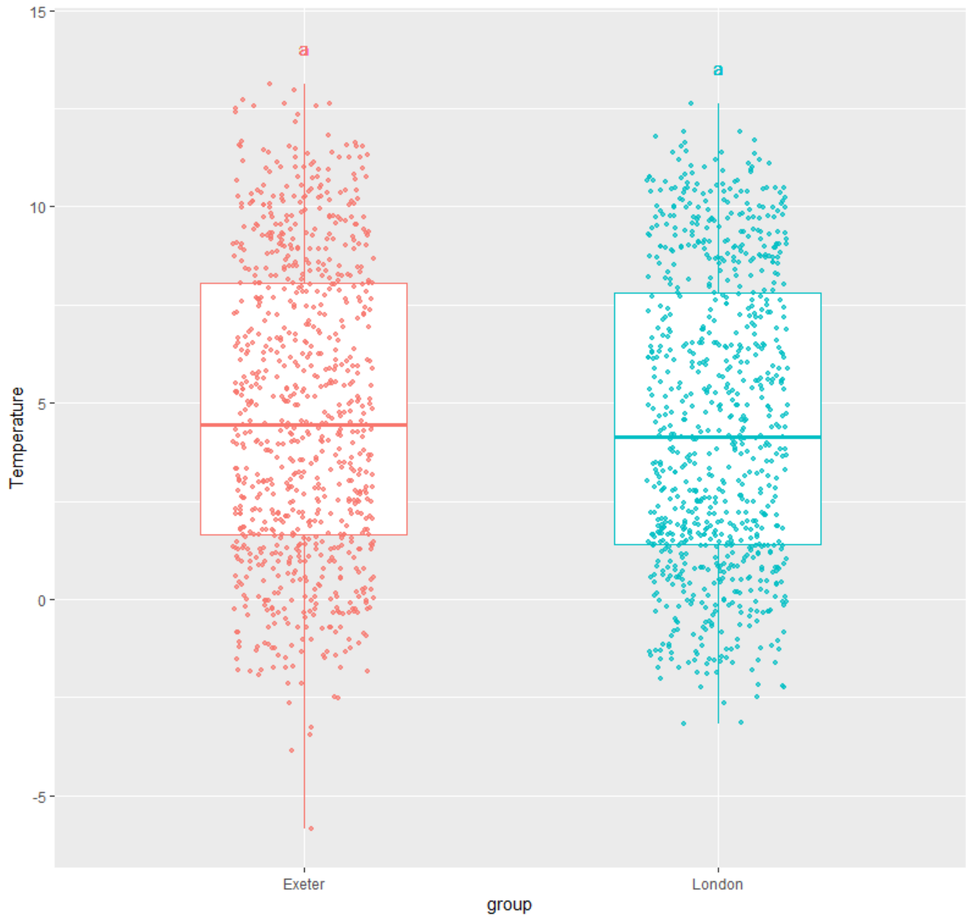

Figure 9

Figure 9| The box plot for the temperature data for Exeter and London. The red part is for Exeter and the blue part is for London.

4. Method

In this project, my main method is time series. Firstly, I extract and integrate data and plot the precipitation time series (figure 2, 3, and 4).
After that, the data is decomposed by the STL. This process can be observed which part has the greatest impact on the data (Chen et al., 2020).

4.1 STL

STL is a common algorithm in the time series decomposition. STL, based on LOESS, decomposes the data at a certain moment (y_v into trend component (t_v, seasonal component (S_v), and remainder component (r_v) three parts:

$$y_v \;=\; t_v + S_v + r_v$$

In this part, we can see the original data and its seasonal components, trend components and residual error components. The process generating the trend can be calculated as a local approximation to a linear trend:

$$t_v = \; t_{v\,\_\,1} + \; \beta_{v\,\_\,1} + \; \eta_v$$
$$\beta_v = \; \beta_{v\,\_\,1} + \; \xi_v$$

where η_v and ξ_v are distributed independently of each other. The mean value is 0, the variance of η_v is σ_η^2, and the variance of ξ_v is σ_ξ^2 over time.

The process generating the seasonal component is:

$$S_v = -\sum_{j=1}^{i-1} S_{v-j} + \omega_v$$

where ω_v is an independently distributed disturbance term. Its mean value is 0, and variance is σ_ω^2 . i is the number of seasons in the year (Harvey & Peters, 1990).
Next, the KPSS algorithm unit root test will be performed on the data and its difference.

4.2 KPSS

KPSS is a commonly used method to verify whether the data is stable. It defined the null hypothesis as the process trend is stable. In other words, the trend of the process is fixed, and the substitution hypothesis is defined as the series having unit roots. KPSS will return the following results: test statistic, p-value and 1%, 2.5%, 5% and 10% confidence interval(CI) critical value. If the test statistic is bigger than the critical value, we reject the null hypothesis, that is, the series is not fixed. If the test statistic is less than the critical value, it fails to reject the null hypothesis. In other words, the series is fixed.
KPSS divides the sequence into three parts, deterministic trend (ξ_t), a random walk (r_t), and a stationary error (ε_t). So, we have:

$$y_t = \xi_t + r_t + \varepsilon_t$$

Here, t = 1, 2, …, T and $r_t = r_{t-1} + u_t$ where u_t are idd(0,σ_u^2). r_0 is the initial value, which is treated as fixed and plays the role of an intercept. The stationarity hypothesis is simply σ = 0. The trend is stable under the null hypothesis because ε_t is assumed to be stable.
Furthermore, if the null hypothesis of level stationarity but not trend stationarity is wished to be tested. It needs to define ε_t as the residual from the regression of y on an intercept, that is $\varepsilon_t = y_t - \bar{y}$ instead of as above (Kwiatkowski et al., 1992).
Fifth, I used the HW algorithm to fit the data. Then, I test whether the residuals of HW fit have autocorrelation and use HW to predict.

4.3 Holt-Winters

HW can also be called Triple Exponential Smoothing. There are two main HW models. One is Additive Seasonal Model, the other one is the Multiplicative Seasonal Model. In this research, I choose to use the former, Additive Seasonal Model, for analysis and predicting (Kalekar, 2004). Actually, similar with STL, HW also divides the data into three parts:

$$y_t = b_1 + b_2 t + S_t, t = 1, 2, \dots, N$$

where b_1 is the permanent component, b_2 is a linear trend component, and S_t is a seasonal additive factor.

Here, we can define R_t as the current depersonalized level of the process at the end of period T. At the end of t, R_t is the estimate of the deseasonalized level, G_t is the estimate of the trend, and S_t is the estimate of the seasonal component (seasonal index).
There are three steps in HW: Over Smoothing, Smoothing of the Trend Factor, and Smoothing of the Seasonal Index.
First of all, over smoothing:

$$R_t = \alpha(y_t - S_{t-L}) + (1 - \alpha) * (R_{t-1} + G_{t-1})$$

where α is a smoothing constant and 0 < α < 1. We depersonalise the data by dividing y_t from S_{t-L}. It can make sure that only the trend component and the prior value of the permanent component enter into the updating process for R_t.
Then, smoothing of the trend factor:

$$G_t = \beta * (S_t - S_{t-1}) + (1 - \beta) * G_{t-1}$$

Where β is the second smoothing constant and 0 < β < 1. The estimate of the trend component is simply the smoothed difference between two consecutive estimates of the deseasonalized level.
Last but not least, smoothing of the seasonal index:

$$S_t = \gamma * (y_t - S_t) + (1 - \gamma) * S_{t-L}$$

where γ is the third smoothing constant and 0 < γ < 1.
After these three steps, we can get the forecast value (Kalekar, 2004):

$$y_t = R_{t-1} + G_{t-1} + S_{t-L}$$

In addition, the ARIMA model will be used to fit the data, and the result model is processed with the same operation as the Holt-Winters model, residual testing and model prediction.

4.4 ARIMA

The ARIMA model refers to transforming a non-stationary time series into a stationary time series and then regressing the input variable into its lag value and the present value and lag value of the random error term to establish a model. ARIMA includes moving average process (MA), autoregressive process (AR) and autoregressive moving average process (ARMA) according to whether the original sequence is stable and the part contained in the regression is different. Each component in ARIMA is used as a parameter with standard symbols. For ARIMA models, the standard notation is ARIMA with p, d, and q, where integer values are substituted for parameters to indicate the type of ARIMA model used. Parameters can be defined as:

p: The number of lagging observations in the model, also known as the lag order.
d: the number of original observation differences, also known as the degree of difference.
q: The size of the moving average window, also known as the order of the moving average.

4.4.1 AR

AR refers to a model that shows a changeable variable that is regressed based on its own lag or prior value.

A p-order AR process can be expressed as:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + a_t$$

Here, c is a constant term and $\{a_t\}$ is "white noise".

After introducing the lag operator L, AR(p) can be written as:

$$Y_t = c + \sum_{i=1}^{p} \phi_i L^i Y_t + a_t$$

or

$$\phi(L) Y_t = c + a_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ (Luetkepoh, 1985).

4.4.2 MA

MA combines the dependence between the observation and the residual of the moving average model applied to the lagging observation.

A q-order MA process can be expressed as:

$$Y_t = c + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} + a_t$$

Here, c is a constant term and $\{a_t\}$ is "white noise".

After introducing the lag operator L, MA(q) can be written as:

$$Y_t = c + \sum_{i=1}^{q} \theta_i L^i a_t + a_t$$

or

$$Y_t = c + \theta(L) a_t$$

where $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ (Ansley et al., 1977).

4.4.3 ARMA

ARMA combines the AR and MA two processes. After merging, an ARMA(p,q) can be obtained. It can be expressed as:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} + a_t$$

Here, c is a constant term and $\{a_t\}$ is "white noise".

After introducing the lag operator L, ARMA(p,q) can be written as:

$$\phi(L) Y_t = c + \theta(L) a_t$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$ and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q$ (Makridakis & Hibon, 1997).

4.4.4 ARIMA

If the sequence Y_t undergoes the difference of d to obtain a stationary sequence W_t, W_t should be modelled by the ARMA (p, q) process. In other words, Wt is an ARMA (p, q) process. We then call Y_t (p, d, q) order

autoregressive single integral moving average process, referred to as ARIMA (p, d, q).
The general ARIMA model can be written:

$$\varphi\ (B)\nabla^{d} z_t = c + \ \theta(B)a_t$$

Here, c is a constant term, $\{a_t\}$ is "white noise", $\varphi(B) = 1 - \varphi_1 B - \varphi_2 B - \cdots - \varphi_p B^p$ and $\theta(B) = 1 - \theta_1 B - \theta_2 B - \cdots - \theta_p B^p$ and $\{a_t\}$ is "white noise" (Box & Pierce, 1970).

Last but not least, after the trend stationarity of the data is analyzed, the VAR model will be used to fit and predict the data.

4.5 VAR

Figure 10 shows the basic process of VAR analysis. It is clear to see that, first of all, we need to set the specification and estimate of the reduced-form VAR model, and then this model will be checked. If the model is rejected, we need to go back to the first step. If the model is not rejected, we can use the model to forecast, analyse the Granger-causality, and conduct the structural analysis. In structural analysis, we need to analyze the forecast scenarios, perform the historical decomposition, analyze the impulse response, and decompose the forecast error variance (Luetkepohl, 2011).

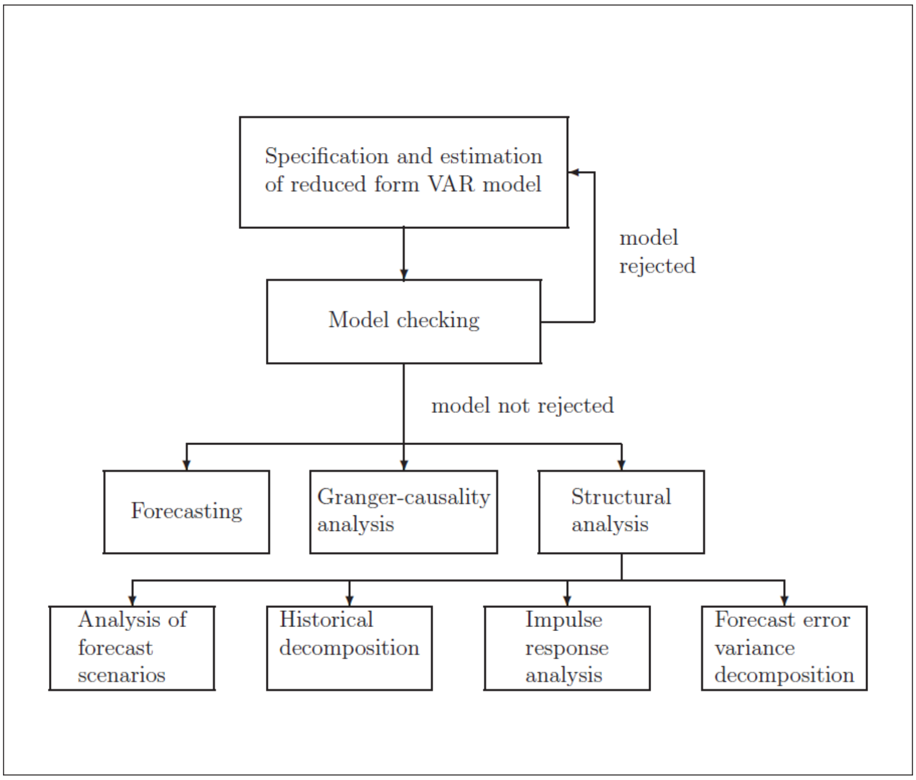

Figure 10

Figure 10| VAR Analysis

4.5.1 The Reduced Form

The VAR model is suitable for describing longer and more frequently observed time series with dynamic structural variables. VAR model can be divided into two parts:

$$y_t = \mu_t + x_t$$

where μ_t is the deterministic part and at most a linear trend $\mu_t = \mu_o + \mu_1 t.$

x_t is a purely stochastic process of order p (VAR(p)):

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + u_t$$

Where the A_i (i = 1,…,p) are (K x K) parameter matrices. u_t = $(u_{1t}, \ldots, u_{Kt})$'

is a K-dimensional white noise process. Its mean value is zero, and it is with covariance matrix $E(u_t u_t') = \sum u$, or we can say $u_t \sim (0, \sum u)$.

Here, $A(L) = I_K - A_1 L - \cdots - A_p L^p$. Thus, $A(L)x_t = u_t$. Now, we can

rewrite the first VAR function as:

$$A(L)y_t = v_0 + v_1 t + u_t$$

or

$$y_t = v_0 + v_1 t + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$$

This process can be known as levels forms of the VAR process (Luetkepohl, 2011).

4.5.2 Model Checking

There are two standard tools for checking residual auto-correlation in VAR models, the Portmanteau and Breusch-Godfrey-LM tests.

All residual autocovariance in Portmanteau test is zero, which is the null hypothesis of the portmanteau test, $H_0: E(u_t u'_{t-i}) = 0, i = 1, 2, \ldots$.

Therefore, the alternative is that at least one autocovariance and one autocorrelation is not zero. The test statistic is based on the residual

autocovariance $\hat{C}_j = T^{-1} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}'_{t-j}$, where the $\hat{u}_t$ is the estimated

residual after adjusting the mean value. We can express the Portmanteau test as (Luetkepohl, 2011):

$$Q_h = T \sum_{j=1}^{h} tr(\hat{C}'_j \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1})$$

If the cointegrating rank is unknow, the Portmanteau test is not recommended to use (Brüggemann et al., 2006).

The Portmanteau test is usually used to test for autocorrelation with high order. To the low order, however, the Breusch-Godfrey LM test is more suitable. It can be seemed as a test for zero coefficient matrices in a VAR model for the residuals,

$$u_t = B_1 u_{t-1} + \cdots + B_h u_{t-h} + e_t,$$

where the e_t is white noise, the error term. Thus, a test of:

$$H_0: B_1 = \cdots = B_h = 0 \text{ vs } H_1: B_i \neq 0, i \in \{1, \ldots, h\}$$

It can be used to check the u_t is white noise (Luetkepohl, 2011). If the cointegrating rank is known, Breusch-Godfrey LM test is not recommended to use (Brüggemann et al., 2006).

4.5.3 Forecasting

The reduced form VAR model applies to forecast because it represents a random process of the conditional mean. The conditional expectation of y_{T+h} given y_t is:

$$y_{Y+h|T} = E(y_{T+h}|y_T, y_{T-1}, \dots)$$

$$= v_0 + v_1(T + h) + A_1 y_{T+h-1|T} + \cdots + A_p y_{T+h-p|T}, t \leq T$$

where the $y_{Y+h|T} = y_{T+j}$ for j ≤ 0. $y_{Y+h|T}$ has the best result if the white noise is iid. In other words, it has the minimum mean squared error (MSE) h-step ahead forecast in period T. The relationships between forecast error and h-step forecast are:

$$y_{Y+h} - y_{Y+h|T} = u_{T+h} + \phi_1 u_{T+h-1} + \cdots + \phi_{h-1} u_{T+1}$$

where the ϕ_i is:

$$\phi_i = \sum_{j=1}^{i} \phi_{i-j} A_j, i = 1, 2, \dots$$

with $\phi_0 = I_k$ and $A_j = 0$ for j > p. The mean value of error is zero, and the forecast error covariance (MSE matrix) is:

$$\sum_y (h) = E\left[(y_{Y+h} - y_{Y+h|T})(y_{Y+h} - y_{Y+h|T})'\right] = \sum_{j=0}^{h-1} \phi_j \sum\nolimits_u \phi_j'$$

or we can say (Luetkepohl, 2011):

$$y_{Y+h} - y_{Y+h|T} \sim \left(0, \sum_y (h)\right)$$

4.5.4 Granger-Causality Analysis

VAR model describes the joint generation process of a number of variables. Thus it can be used to research the relationships between variables (Luetkepohl, 2011). If the past values and present values of y_{2t} are helpful to improve the forecast of y_{1t}, then we can say the y_{2t} causal for y_{1t} (Granger, 1969):

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \sum_{i=1}^{p} \begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} \begin{pmatrix} y_{1,t-i} \\ y_{2,t-i} \end{pmatrix} + u_t$$

where if $a_{12,i} = 0, i = 1, 2, \dots, p$, the y_{2t} is not Granger-causal for y_{1t}.

4.5.5 Structural Analysis

Structural analysis can be divided into four parts: Impulse Response Analysis, Forecast Error Variance Decompositions, Historical Decomposition of Time Series, and Analysis of Forecast Scenarios.

Firstly, the impulse response analysis, innovations or shocks enter from the residual vector $u_t = (u_{1t}, \dots\ u_{Kt})'$. Thus, we can give:

$$y_t = A(L)^{-1}u_t = \phi(L)u_t = \sum_{j=0}^{\infty} \phi_j u_{t-j}$$

where $\phi(L) = \sum_{j=0}^{\infty} \phi_j L^j = A(L)^{-1}$.

The forecast error variance decompositions, secondly, can be represented as:

$$y_{Y+h} - y_{Y+h|T} = \psi_0 v_{T+h} + \psi_1 v_{T+h-1} + \cdots + \psi_{h-1} v_{T+1}$$

Using $\Sigma_0 = I_k$, the forecast error variance of the *k*th component of y_{Y+h} should be:

$$\sigma_k^2(h) = \sum_{j=0}^{h-1} \left(\psi_{k1,j}^2 + \cdots + \psi_{kK,j}^2\right) = \sum_{j=1}^{K} \left(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2\right)$$

where $\psi_{nm,j}$ means the (n,m)th element of ψ_j. $\left(\psi_{kj,0}^2 + \cdots + \psi_{kj,h-1}^2\right)$ means the contribution of the *j*th shock to the h-step forecast error variance of k.

In 1958, Burbidge and Harrison proposed the historical decomposition of time series. We can get the contributions of the structural shocks by decomposing the time series. The historical decomposition of time series can be represented as:

$$y_t = \sum_{i=0}^{t-1} \phi_i u_{t-i} + A_1^{(t)} y_0 + \cdots + A_p^{(t)} y_{1-p}$$

$$= \sum_{i=0}^{t-1} \psi_i v_{t-i} + A_1^{(t)} y_0 + \cdots + A_p^{(t)} y_{1-p}$$

where the ϕ_i and ψ_i are the moving average coefficient matrices. $A_i^{(t)}$ is the element in a (pK x pk) matrix A^t, where

$$A^t = \begin{bmatrix} A_1 & \cdots & A_{p-1} & A_p \\ I_k & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & I_K & 0 \end{bmatrix}$$

Finally, the VAR model will use for analysing different forecast scenarios or conditional forecasts. To describe this process, the analysis of forecast scenarios can be represented as:

$$y_{Y+h} = \sum_{i=0}^{h-1} \psi_i v_{T+h-i} + A_1^{(h)} y_T + \cdots + A_p^{(h)} y_{T+1-p} \quad , h = 1, 2, \ldots$$

This function is the same as the historical decomposition of time series (Luetkepohl, 2011).

5. Result

First of all, I use STL to separate the precipitation data in Exeter. The result is in figure 11. It is clear to see that the data has a strong and regular periodicity. Through the analysis of the trend, we can see that the overall data is slowly rising amidst fluctuations. To the remainder, this part has a weak influence on the data. In this research, the remaining part will be ignored.

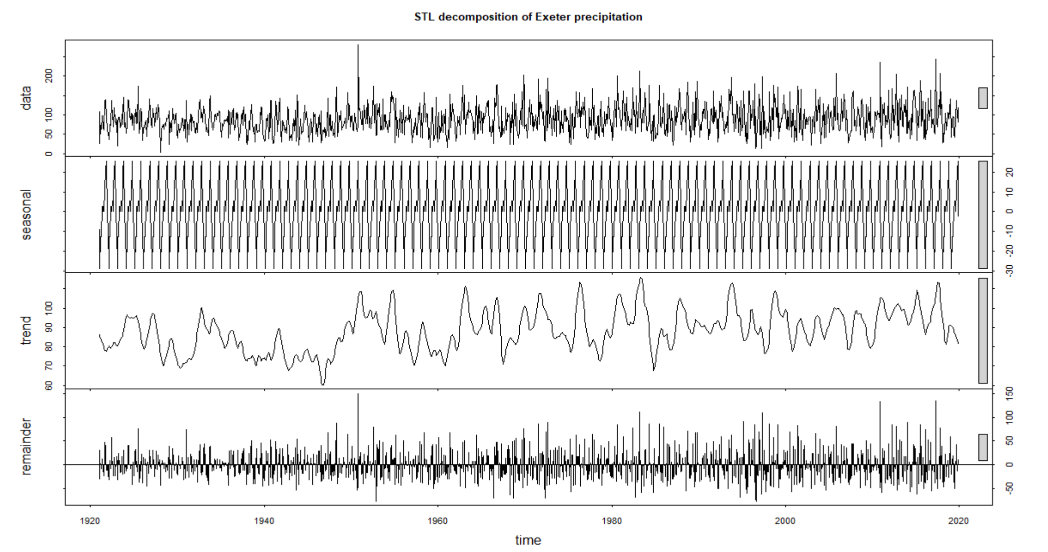

Figure 11

Figure 11| The STL decomposition of Exeter precipitation from 1920 to 2020

After analysing the trend, I analyzed the data again. Here, I perform level stationarity test and trend stationarity test on the data, respectively. The result is in table 1. When the KPSS test is performed on the original data, the p-value of the Level is 0.01, which rejects the null hypothesis of KPSS, thus indicating that the Level of the original data is unstable. The p-value of the Trend is 0.1, which does not reject the null hypothesis of KPSS, so it shows that the Trend of the original data is stable. On this basis, we perform differential processing on the original data and perform a KPSS test on differential precipitation. The p-value of Level is 0.1, and the null hypothesis of KPSS is not rejected. Therefore, it is explained that the Level of differential precipitation is stable. The p-value of the Trend is 0.1, which does not reject the null hypothesis of KPSS, so it shows that the Trend of differential precipitation is stable. The smaller the value of KPSS, means that the data is more stable. Although the trend of both precipitation and differential precipitation is stable, the trend of differential precipitation is more stable than it of precipitation by comparing the KPSS value.

		KPSS	Truncation lag parameter	p-value
Precipitation	Level	1.7842	7	0.01
	Trend	0.056155	7	0.1
Differential precipitation	Level	0.0038185	7	0.1
	Trend	0.0037215	7	0.1

Table 1

Table 1| The KPSS result

Next, HW is used to analyse and predict the precipitation model. Here, I also separate the precipitation data in figure 12. This step is similar with the STL. The level has a significant increasing trend, but the trend initially decreased and then gradually stabilised. Though the seasonal part is not as clear as it in STL, we still can clearly see the periodicity. Both STL and HW decompose the data, and both of them concluded that the data showed regular periodicity. However, though their periodicity is similar, the data differed greatly. From figure 11 and figure 12, we can see that the seasonal decomposition of HW is more scientific because the data has certain fluctuations. This fluctuation is more similar to the fluctuation of the original data. The seasonal decomposition of STL is too ideal. Next, Through the analysis of the Residual density histogram and density estimation, Normal Q-Q Plot and Residual Series (all these figures are in Appendix part, please see figure 19, figure 20 and figure 21), we got the residuals are in accordance with the normal distribution. Finally, figure 13 shows the predicted result of HW. In the figure, the black line is the original data and the blue line is the predicted data. The predicted precipitation fluctuates cyclically from 75 mm to 125 mm approximately from 2011 to 2020.

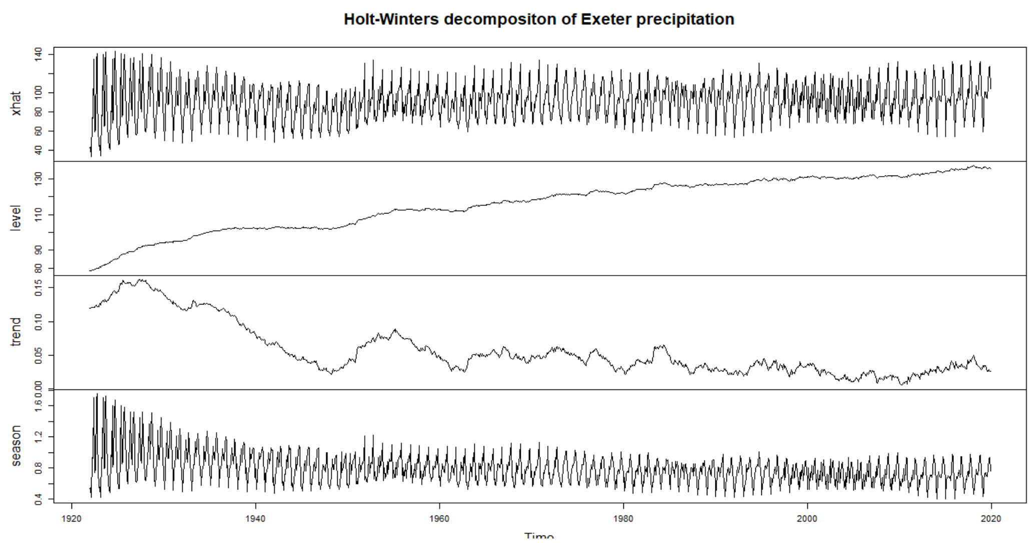

Figure 12

Figure 12| Holt Winters decomposition of Exeter precipitation from 1920 to 2020

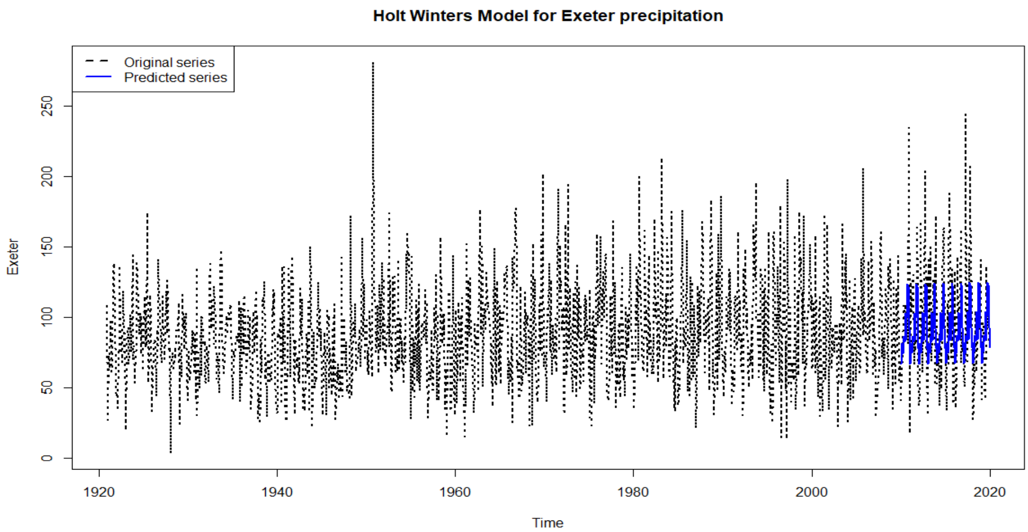

Figure 13

Figure 13| Holt Winters Model for Exeter precipitation. The black line is the original data, the blue line is the predicted data.

Then, I use the second method, ARIMA, to analyze and predict the model. ARIMA automatically fits the model. By comparing the bic value, the selected model is ARIMA(0,0,2)(1,0,0). Figure 14 is the forecast result. In the figure, the black line is the original data and the blue line is the predicted data. The dark blue zone and the grey zone are the CI. Dark blue is 75% CI and grey is 95% CI. It can be seen from the figure that the CI of ARIMA's forecast results is abnormal. The 75% CI and 95% CI are very large, and the maximum value of the CI has a rising trend with the growth of the year, and the minimum value of the CI has a downward trend or even less than 0 with the growth of the year. This result is problematic. I speculate that the reason for the abnormal result may be because ARIMA does not fit this data set, but the specific reasons still need to be investigated in a more scientific way.

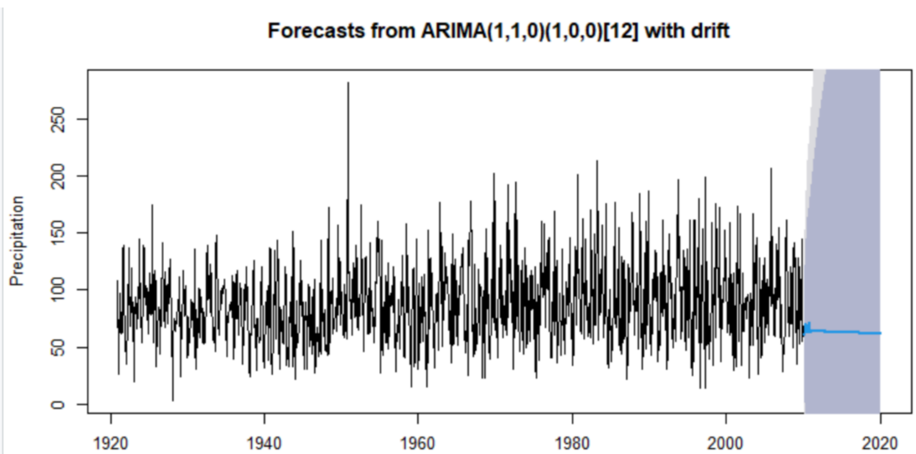

Figure 14

Figure 14| Forecasts result from ARIMA(0,0,2)(1,0,0) of Exeter precipitation from 2011 to 2020

Last but not least, I use var to predict the precipitation model. The prediction result has been shown in figure 15. The blue line is the predicted data, and the red line is the 95% CI. Similar to ARIMA, theCI in the prediction result of VAR is also abnormal. From the figure, we can see that the prediction result has a minimum value around 2014, but the CI of minimum value is less than 0. But if ignoring these two points, the overall prediction result is still good.

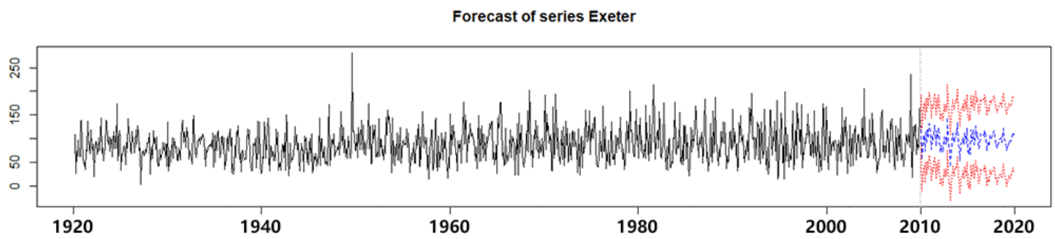

Figure 15

Figure 15| Forecasts result for VAR of Exeter precipitation from 2011 to 2020

The temperature data is used to assist in the study of precipitation data. Here, I try to find the relationship between temperature and precipitation by constructing a linear relationship between temperature and precipitation. I constructed a linear relationship between the precipitation in Exeter and the temperature in Exeter (figure 16), the precipitation in London and the temperature in London (figure 16), the precipitation in Exeter and the temperature in London (figure 17) separately. Finally, I got three functions:

1. P(Exeter,t) = 0.01165 * T(Exeter,t) + 4.040739 R squared: 0.0101
2. P(London,t) = 0.00576 * T(London,t) + 4.238277 R squared: 0.0009011
3. P(Exeter,t) = 0.009592 * T(London,t) + 4.102695 R squared: 0.006285

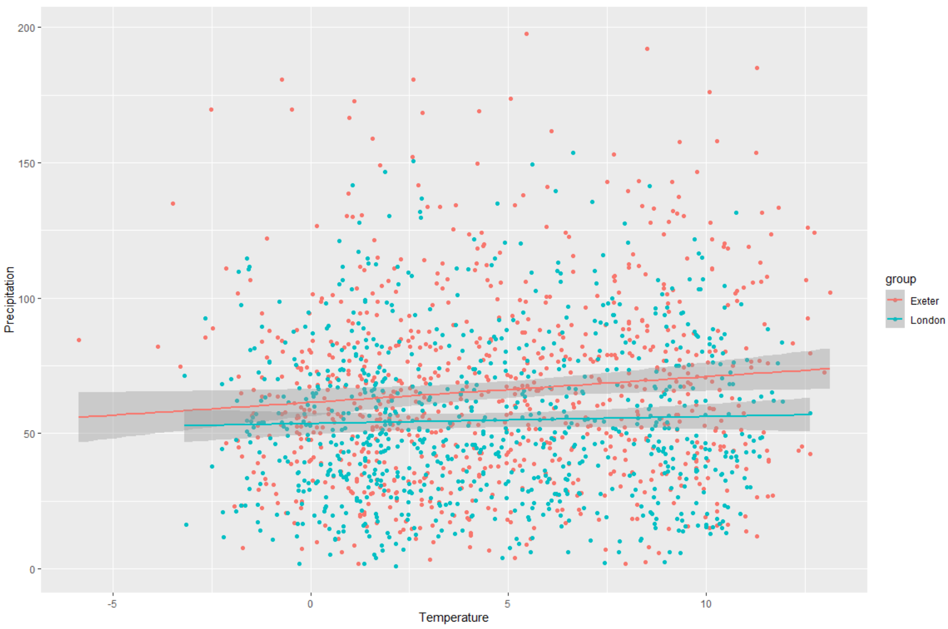

Figure 16

Figure 16| The linear relationship between temperature and precipitation of Exeter and London. The red point is the precipitation in Exeter corresponding to the temperature in Exeter. The red line is the linear relationship of the temperature in Exeter and precipitation in Exeter. The blue point is the precipitation in London corresponding to the temperature in London. The blue line is the linear relationship between the temperature in London and precipitation in London.

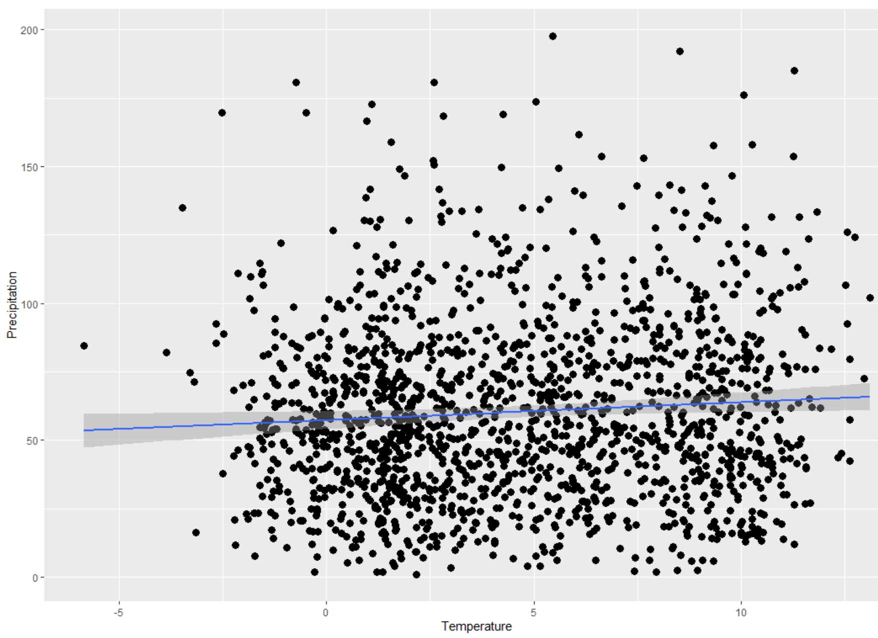

Figure 17

Figure 17| The linear relationship between temperature in London and precipitation in Exeter. The black point is the precipitation in Exeter corresponding to the temperature in London. The blue line is the linear relationship between the temperature in London and precipitation in Exeter.

6. Discussion

In this part, I will focus on three parts, the peak point of precipitation for Exeter in 1950, precipitation prediction and auxiliary research on temperature to discuss my research.

6.1 Peak precipitation for Exeter

Throughout the history of Exeter, I found that there were no special events in 1950 that caused Exeter's precipitation to reach its extreme value. Interestingly, however, Exeter experienced a flood in 1960. On October 27, 1960, there was a flood in the River Exe (Furse, 1992). The river pours down from the banks of the Exe River above Exwick, flows through St Thomas, and towards Alphington. The precipitation in the Exe River catchment area exceeded 380 mm in that month, which is half of the annual average. On October 26, it rained 60 mm (Brierley et al., 1965). The flood destroyed more than 2,000 buildings. People referred to this flood as the St. Thomas flood and this day as Black Thursday (Booth, 1961). To explain this phenomenon, only precipitation in a short period of time will affect the flood, and the annual precipitation, or the long-term precipitation, cannot be used as a direct judgment condition for the flood.

6.2 Precipitation Prediction

Aside from the abnormal CIs of ARIMA and VAR, we only select reasonable parts for comparison. By comparing the forecast results of HW, ARIMA and VAR with the actual data from 2011 to 2020, it is clear to see that the overlap between the prediction result of VAR and the actual data from 2011 to 2020 is larger than the other two methods. The predicted result of HW has a lower coincidence degree with the original data, and it has too significant regularity, which makes the result too ideal. In these three methods, HW can get second place. To ARIMA, the data fit this method so poorly compatible with the data. The CI of it is incorrect. The predicted data has a lower coincidence degree than the original data and it has almost no seasonal fluctuations. This phenomenon is inconsistent with the facts. Moreover, the predicted overall trend is completely opposite to the overall trend of the original data. The trend of the original data is gradually increasing, while the predicted overall trend is indeed decreasing. Thus, ARIMA performs the worst of these three methods. However, the forecast result of VAR has the best coincidence degree with the original data. The data has natural fluctuations and obvious seasonal changes, which make it closer to the real situation. Though two points of the CI are less than 0, they can be ignored. The performance of VAR can get first place in these three methods. Therefore, I chose to use VAR to predict the rainfall model of Exeter rainfall. Figure 18 shows the rainfall data predicted by VAR from 2021 to 2030. The blue line is the forecast data, and the red line is the 95% CI. From the figure, we can see that the forecast data fluctuates significantly from 2021 to 2025, and after 2025, the data will gradually become regular and stable. Therefore, I speculate that from 2025 to 2030, climate fluctuations in Exeter are likely to stabilize, and the possibility of extreme precipitation is very low. More specific analysis of the data from 2021 to 2025, we can see that the maximum value appears around 2023 and 2025, and the minimum value appears around 2021 and 2022. Unlike the peak in 1950, both of them are not significant than other predicted data. Though the maximum values and the minimum values are not significant to others, it is still very important to prepare in advance to deal with extreme weather such as drought and extreme precipitation. The prediction results in this essay can only be used as an auxiliary reference and cannot achieve very accurate predictions. Thus, for the occurrence of extreme climates, relevant departments still need to use more accurate methods to make more accurate predictions. The overall trend of these ten years is still rising. Thus, I predict that the precipitation in Exeter will be higher in the following ten years. In this prediction result, there is still the problem of the abnormal CI. At the minimum position around 2021 and 2022, the lowest value of the CI is less than 0. Theoretically, the CI of the precipitation prediction result should always be greater than 0. Only when the predicted value is very low, and the choice of the CI is very wide, the lowest value of the CI can be infinitely close to 0. The specific reason for this phenomenon needs to be investigated.

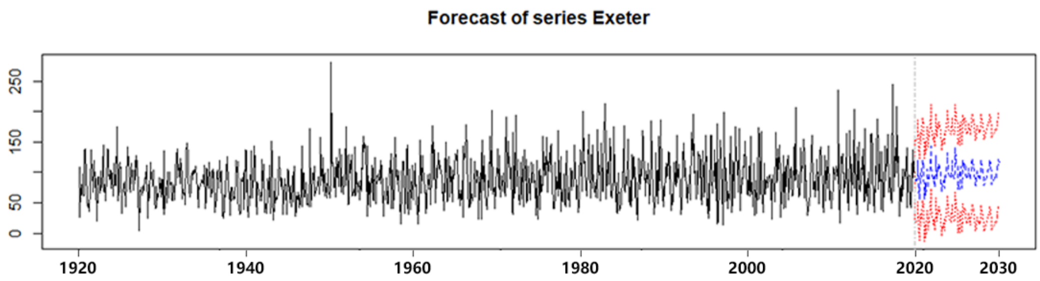

Figure 18

Figure 18| The predicted result of precipitation in Exeter from 2021 to 2030. The blue line is the predicted data. The red line is 95% CI.

6.3 Temperature Research

From figure 16, we have got the linear relationship of precipitation in Exeter and the temperature in Exeter, the precipitation in London and the temperature in London and the precipitation in Exeter and the temperature in London, by comparing the R squared values of the three linear relationships, we can get R squared (1)> R squared (3)> R squared (2). Therefore, we can say that the precipitation in Exeter has a strong positive correlation with the temperature in Exeter, and the precipitation in Exeter has a weaker positive correlation with the temperature in the main climate in England. Furthermore, by comparing the main trends in figure 3 and figure 7, we can find that the precipitation is also increasing as the temperature rises. Therefore, without considering the influence of other factors, we can conclude that precipitation will increase with increasing temperature to a certain extent. The higher the temperature, the more water the atmosphere can hold, which means more rainfall (Witze, 2018). When the rainfall reaches a critical value, it forms one of the most well-known weather events in the world, extreme precipitation. According to calculations, for each degree (K) increase in global temperature, water vapour will increase by 7%, leading to an increase in precipitation. Intense rainfall will precipitate and remove a large amount of water vapour that can be used for light and moderate precipitation. Therefore, rising global temperature leads to an increase in heavy precipitation, while light and moderate precipitation decrease (Trenberth et al., 2003). According to the climate model predictions provided by the IPCC, in most parts of the world, the number of days of heavy precipitation events and the number of days of dry days will increase, including many areas where both extremes will increase (Solomon, 2007). In researching global data, we found that increasing temperature and increasing rainfall are global phenomena. Alexander conducted research on temperature data and precipitation data from 1900 to 2003 and found that in the past 100 years, the temperature of cold nights and warm nights has been rising, the number of cold nights has been decreasing, and the number of warm nights has constantly been increasing. Significant warming occurred throughout the 20th century, and the precipitation index showed a trend of getting wet throughout the 20th century. (Alexander, 2006). A lot of extreme precipitation caused by temperature rise frequently occurs around the world. In 2005, Hurricane Katrina struck the United States from August 23 to August 31, causing more than 1,800 deaths and 125 billion dollars in damage (Didlake & Houze, 2009). In 2017, Hurricane Irma (Cangialosi et al., 2018) and Hurricane Maria (Pasch et al., 2018) swept the Atlantic Ocean from August 30 to October 2, causing more than 3,900 deaths and 168.77 billion dollars in damage. Lawrence Berkeley National Laboratory in California used allowable convection models to study these three hurricanes and found that climate change increased precipitation in these storms to 9%. The warming will bring more extreme precipitation from similar hurricanes in the future (Patricola & Wehner, 2018). In 2018, Hurricane Florence hit the Atlantic Ocean from August 31 to September 18, which caused 54 deaths and 242.3 billion dollars in damage (Stewart & Berg, 2019). In the same year, a flood occurred in Kerala, India, which caused 483 deaths and 56 billion dollars in damage (Mishra & Shah, 2018). If greenhouse gas emissions continue to rise, Hurricane Florence, floods in India and other devastating rains will recur in the future. Angeline Pendergrass, an atmospheric scientist at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, stated that all signs indicate that extreme precipitation will be more severe in the next 20 years than in the past 20 years. If any measures are taken to allow the global temperature to continue to rise, extreme precipitation will have an irreversible impact on the living environment of human beings (Witze, 2018). In 2021, the flood swept Europe.

Many countries such as Germany, Austria, and so on were affected. This flood occurred after an unprecedented heatwave in the Pacific Northwest and Northern Europe (Austin, 2021). Carl-Friedrich Schluisner, head of the research group at the Institute of Geography of the Humboldt University of Berlin, and Zurich Switzerland Sebastian Hippel of the Institute of Climate and Atmosphere Research at the Federal Institute of Technology believes that man-made climate change is heating up our atmosphere, which causes the extra moisture caused by warming to cause more rainfall, especially extreme precipitation (Bleiker, 2021). After this, London also suffered a flood on July 26. Dozens of streets and subway stations were flooded. Meteorological Bureau meteorologist Steven Kitts said the flood was caused by extreme precipitation caused by the evaporation of water due to high temperatures on the ground and the encounter of colder air higher in the atmosphere (BBC, 2021).

7. Conclusion

In conclusion, in this research, I analyzed and predicted the Exeter precipitation model. I chose London as the main climate model for England to compare with it and used temperature data to assist in the analysis of precipitation. First, I used STL and HW, two methods to decompose and analyze the data in cycles and trends and then used KPSS to analyze and process the data. After that, HW, ARIMA, and VAR three methods were used to predict the precipitation model. By comparing the prediction results with the original data from 2011 to 2020, I got the result that the prediction results of VAR were more accurate compared with the other two methods. Therefore, I finally chose to use VAR to predict the precipitation from 2021 to 2030, and the prediction results are shown in Figure 18. Finally, I constructed a linear relationship between the precipitation in Exeter and the temperature in Exeter, the precipitation in London (the main climate model in England) and the temperature in London (the main climate model in England), the precipitation in Exeter and the temperature in London (the main climate model in England) separately. I got the result that there is a significant positive correlation between the precipitation in Exeter and the temperature in Exeter, and the precipitation in Exeter has a weak positive correlation with the temperature in London (the main climate model in England). Afterwards, I discussed the entire research and found that to the study of Exeter's extreme precipitation points in 1950, Exeter only had a flood in 1960 but not in 1950. Then the relationship between temperature and rainfall was investigated and concluded to a certain extent that precipitation will increase with increasing temperature without considering the influence of other factors. Therefore, in order to reduce the impact of extreme climates such as extreme precipitation and drought on humans, it is very necessary to reduce the temperature rise mainly caused by the greenhouse effect through energy-saving, emission-reduction, afforestation, and so on. The research in this essay is limited. First of all, I only used three methods in this essay and chose VAR to predict the precipitation. This can only show that VAR is the best forecasting method compared to HW and ARIMA. But not VAR is the best method to predict the precipitation. It must have a better method than VAR to predict. Then, in this essay, I chose London, the central city of England, as the overall climate model for England. However, it does not collect the climate of various regions of England for averaging processing. Therefore, the choice of London as the main climate model for England in this project is not representative. A more scientific model should be used to obtain the average precipitation, temperature and climate model for England. This also led to problems when analysing the linear relationship between temperature and precipitation. The degree of correlation between the precipitation in London and the temperature in London is not as strong as the precipitation in Exeter and the temperature in Exeter and the precipitation in Exeter and the temperature in London. Furthermore, the daily precipitation and temperature will be better than the monthly precipitation and temperature. A shorter period of time will make the analysis and forecast more accurate because the monthly precipitation is the sum of the precipitation of each day in each month, but the monthly average temperature is the average temperature of each day in each month. This has a greater impact on the data. Last but not least, the abnormal CIs in ARIMA and VAR are also very important issues. We should explore the reasons for abnormal CIs in more depth, and on the basis of the two methods, they should be optimized and changed to eliminate the abnormal CIs. Thus, in the future study, I will focus on optimization of existing models and more methods to predict and analyze and look for a more scientific method to get various data on the main climate in England instead of directly choosing the central city London as the main climate model in England. The data also needs to be optimized. Data with shorter time intervals will be used, such as daily data.

8. Reference

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., ... & Vazquez - Aguirre, J. L. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research: Atmospheres*, *111*(D5).

Ansley, C. F., Spivey, W. A., & Wrobleski, W. J. (1977). On the structure of moving average processes. *Journal of Econometrics*, *6*(1), 121-134.

Austin, H. (2021, July 16). *Rapid rain and slow response blamed for high number of flooding deaths in Europe.* NBCNews.com..

Bari, S. H., Rahman, M. T., Hussain, M. M., & Ray, S. (2015). Forecasting monthly precipitation in Sylhet city using ARIMA model. *Civil and Environmental Research*, *7*(1), 69-77.

Bates, B., Kundzewicz, Z., & Wu, S. (2008). *Climate change and water*. Intergovernmental Panel on Climate Change Secretariat.

BBC (2021, July 26). *UK weather: London flooding caused by heavy rain and storms.* BBC News. https://www.bbc.co.uk/newsround/57967356.

Bleiker, C. (2021, July 16). *Hochwasser: Diese rolle SPIELEN Jetstream und Mondumlaufbahn*. https://www.dw.com/de/hochwasser-diese-rolle-spielen-jetstream-und-mondumlaufbahn/a-58292603. Booth, R. E. (1961). Looking Back on 1960. *Weather*, *16*(6), 193-196.

Boken, V. K., Cracknell, A. P., & Heathcote, R. L. (2005). *Monitoring and predicting agricultural drought: a global study*. Oxford University Press.

Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, *65*(332), 1509-1526.

BRIERLEY, J., BISWAS, A., LEGGET, R., HALL, D., & GOSLING, R. (1965). DISCUSSION. FLOODING IN THE EXE VALLEY, 1960. *Proceedings of the Institution of Civil Engineers*, *32*(1), 109-116.

Brüggemann, R., Lütkepohl, H., & Saikkonen, P. (2006). Residual autocorrelation testing for vector error correction models. *Journal of Econometrics*, *134*(2), 579-604.

Burns, R. (2007, June 6). *Texas Crop and Weather.* https://web.archive.org/web/20100620134950/http://agnewsarchive.tamu.edu/dailynews/stories/CROP/Jun0607a.htm.

Cangialosi, J. P., Latto, A. S., & Berg, R. (2018). National Hurricane center tropical cyclone report: Hurricane Irma. *National Oceanic and Atmospheric Administration. Report no. AL112017*.

Chang, T. J., Kavvas, M. L., & Delleur, J. W. (1984). Daily precipitation modeling by discrete autoregressive moving average processes. *Water Resources Research*, *20*(5), 565-580.

Chen, D., Zhang, J., & Jiang, S. (2020). Forecasting the Short-Term Metro Ridership with Seasonal and Trend Decomposition Using Loess and LSTM Neural Networks. *IEEE Access*, *8*, 91181-91187.

Didlake Jr, A. C., & Houze Jr, R. A. (2009). Convective-scale downdrafts in the principal rainband of Hurricane Katrina (2005). *Monthly Weather Review*, *137*(10), 3269-3293.
Diebold, F. X., & Lopez, J. A. (1996). 8 Forecast evaluation and combination. *Handbook of statistics*, 14, 241-268.
Fan, Y., & Van den Dool, H. (2008). A global monthly land surface air temperature analysis for 1948–present. *Journal of Geophysical Research: Atmospheres*, *113*(D1).
Field, C. B., Barros, V., Stocker, T. F., & Dahe, Q. (Eds.). (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change*. Cambridge University Press.
Furse, J. (1992). Exhibitions: EDWARD ALLINGTON. *Art Monthly (Archive: 1976-2005)*, (162), 25.
Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424-438.
Harvey, A. C., & Peters, S. (1990). Estimation procedures for structural time series models. *Journal of forecasting*, *9*(2), 89-108.
Hu, D., Shu, H., Hu, H., & Xu, J. (2017). Spatiotemporal regression Kriging to predict precipitation using time-series MODIS data. *Cluster Computing*, *20*(1), 347-357.
Illán, J. G., Thomas, C. D., Jones, J. A., Wong, W. K., Shirley, S. M., & Betts, M. G. (2014). Precipitation and winter temperature predict long‐term range‐scale abundance changes in Western North American birds. *Global Change Biology*, *20*(11), 3351-3364.
Indriani, R., Sugandha, A., Tripena, A., Larasati, N., Rokhman, A. F., & Bon, A. T. (2020). Forecasting of Air Temperature in Cilacap Regency with Triple Exponential Smoothing (Holt-Winter) Method.
Intergovernmental Panel on Climate Change. (2018). *Global warming of 1.5° C: an IPCC special report on the impacts of global warming of 1.5° C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*. Intergovernmental Panel on Climate Change.
Kalekar, P. S. (2004). Time series forecasting using holt-winters exponential smoothing. *Kanwal Rekhi school of information Technology*, *4329008*(13), 1-13.
Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?. *Journal of econometrics*, *54*(1-3), 159-178.
Luetkepoh, H. (1985). Comparison of criteria for estimating the order of a vector autoregressive process. *Journal of time series Analysis*, *6*(1), 35-52.

Luetkepohl, H. (2011). *Vector Autoregressive Models* (No. ECO2011/30).
Makridakis, S., & Hibon, M. (1997). ARMA models and the Box–Jenkins methodology. *Journal of forecasting*, *16*(3), 147-163.
Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International journal of forecasting*, *16*(4), 451-476.
Mishra, A. K., & Desai, V. R. (2005). Spatial and temporal drought analysis in the Kansabati river basin, India. *International Journal of River Basin Management*, *3*(1), 31-41.
Mishra, V., & Shah, H. L. (2018). Hydroclimatological perspective of the Kerala flood of 2018. *Journal of the Geological Society of India*, *92*(5), 645-650.
Pasch, R. J., Penny, A. B., & Berg, R. (2018). Hurricane Maria (AL152017). *National Hurricane Center, USA*.
Patricola, C. M., & Wehner, M. F. (2018). Anthropogenic influences on major tropical cyclone events. *Nature*, *563*(7731), 339-346.
Raha, S., & Gayen, S. K. (2020). Simulation of meteorological drought using exponential smoothing models: A study on Bankura District, West Bengal, India. *SN Applied Sciences*, *2*(5), 1-24.
Schneide, U., Ziese, M., Becker, A., Finger, P., & Elke, R. (2021). *Global Precipitation Analysis Products of the GPCC.* Global Precipitation Climatology Centre at Deutscher Wetterdienst.
Solomon, S. (2007). The physical science basis: Contribution of Working Group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. *Intergovernmental Panel on Climate Change (IPCC), Climate change 2007*, *996*.
Stewart, S., & Berg, R. (2019). National Hurricane Center Tropical Cyclone Report Hurricane Florence. *National Hurricane Center, nd, https://www. nhc. noaa. gov/data/tcr/AL062018_Florence. pdf*.
Stocker, T. (Ed.). (2014). *Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge university press.
Thasneem, S. A., Chithra, N. R., & Thampi, S. G. (2019). Analysis of extreme precipitation and its variability under climate change in a river basin. *Natural Hazards*, *98*(3), 1169-1190.
Theodosiou, M. (2011). Forecasting monthly and quarterly time series using STL decomposition. *International Journal of Forecasting*, *27*(4), 1178-1195.
Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1-2), 123-138.
Trenberth, K. E., Dai, A., Rasmussen, R. M., & Parsons, D. B. (2003). The changing character of precipitation. *Bulletin of the American Meteorological Society*, *84*(9), 1205-1218.
Van Liere, M. J., Ategbo, E. A. D., Hoorweg, J., Den Hartog, A. P., & Hautvast, J. G. (1994). The significance of socio-economic characteristics for adult seasonal body-weight fluctuations: a study in north-western Benin. *British*

Journal of Nutrition, *72*(3), 479-488.
Wang, S., Feng, J., & Liu, G. (2013). Application of seasonal time series model in the precipitation forecast. *Mathematical and Computer modelling*, *58*(3-4), 677-683.
Wei, F (2007). *Modern climate statistical diagnosis and prediction technology [M].* Beijing: Meteorological Publishing House.
Wiah, E. N. (2017). Impact of climate change on cocoa yield in Ghana using vector autoregressive model. *Ghana Journal of Technology*, *1*(2), 32-39.
Witze, A. (2018). Why extreme rains are gaining strength as the climate warms. *Nature*, *563*(7732), 458-461.
Yang, X., Ferrat, M., & Li, Z. (2013). New evidence of orographic precipitation suppression by aerosols in central China. *Meteorology and Atmospheric Physics*, *119*(1), 17-29.

9. Acknowledgement

I would like to thank my parents for their financial support during my research and my girlfriend and friends for their spiritual support. Special thanks to my two supervisors, John Bruun and Stefan Siegert, all the teachers in the ADSS team, and my tutor for all their help in my life and study.

10. Appendix

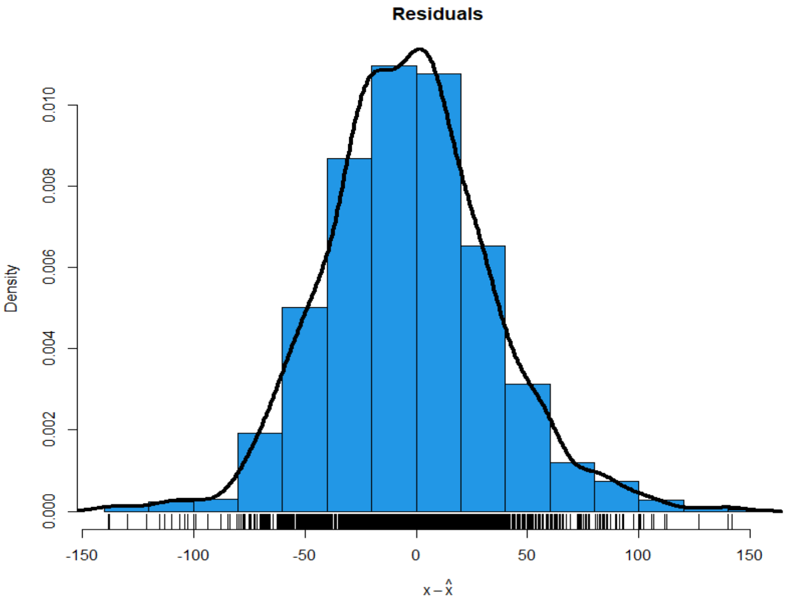

Figure 19

Figure 19| Residual density histogram and density estimation of HW

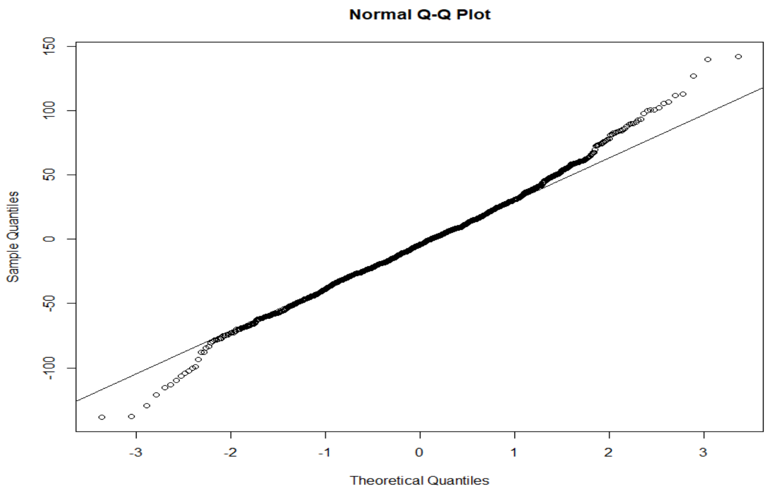

Figure 20

Figure 20| Normal Q-Q Plot of HW

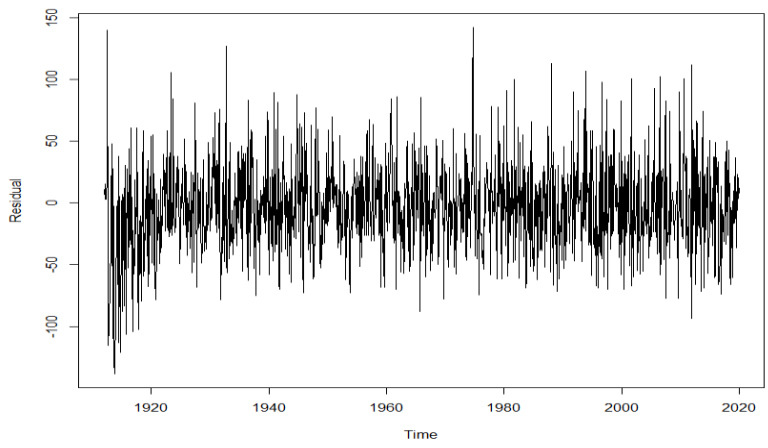

Figure 21

Figure 21| Residual Series of HW

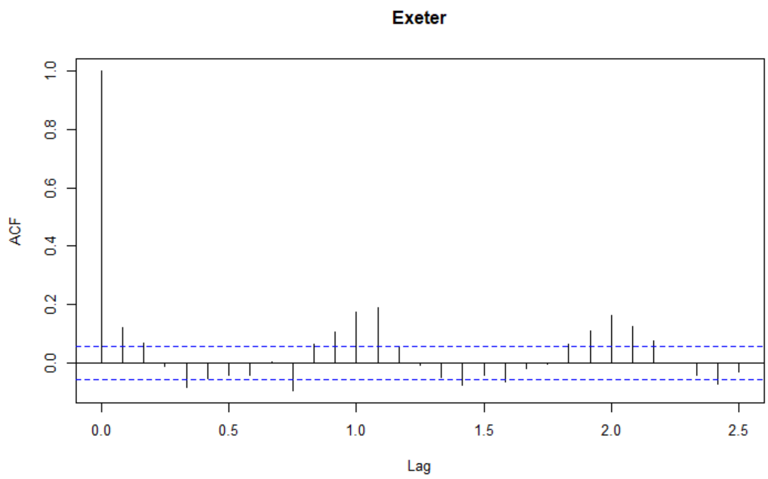

Figure 22

Figure 22| AFC of Exeter precipitation time series

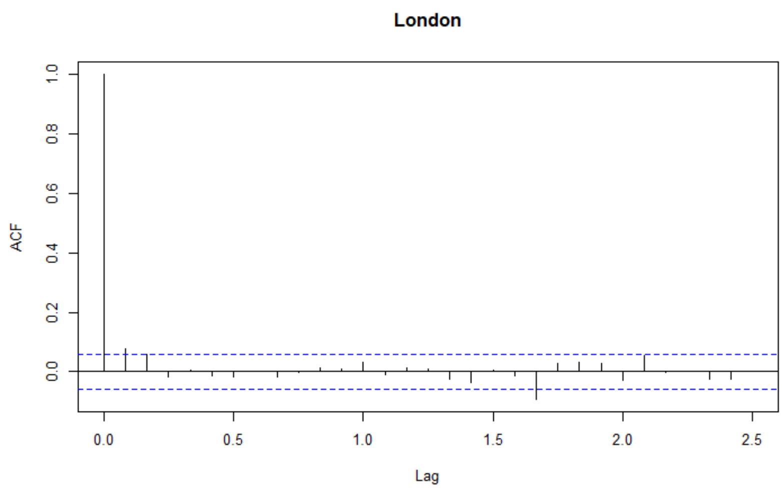

Figure 23

Figure 23 | AFC chart of London precipitation time series

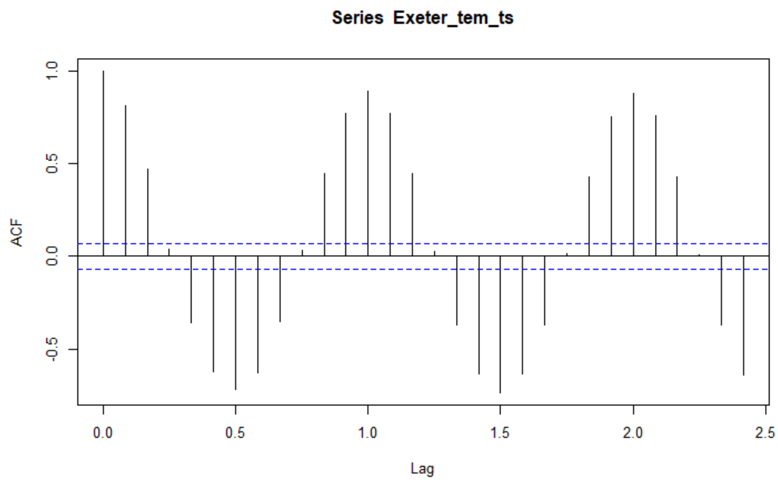

Figure 24

Figure 24 | AFC chart of Exeter temperature time series

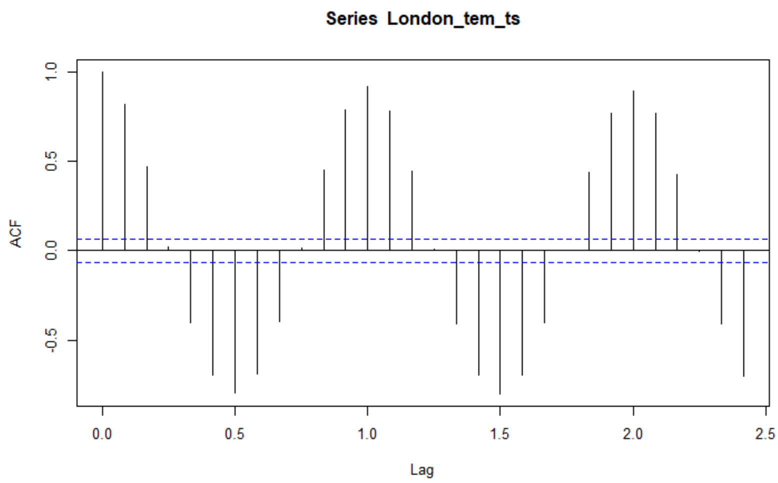

Figure 25

Figure 25 | AFC chart of London temperature time series

```r
knitr::opts_chunk$set(echo = TRUE)
#package
library(RColorBrewer)
library(lattice)
library(trend)
library(magrittr)
library(raster)
library(ggplot2)
library(rnaturalearth)
library(ncdf4)
library(cowplot)
library(dplyr)
rm(list = ls())

####Time Series of Precipitation Data analysis####
#read Data
ncdata1 <- nc_open("full_data_monthly_v2020_1921_1930_05.nc")
ncdata2 <- nc_open("full_data_monthly_v2020_1931_1940_05.nc")
ncdata3 <- nc_open("full_data_monthly_v2020_1941_1950_05.nc")
ncdata4 <- nc_open("full_data_monthly_v2020_1951_1960_05.nc")
ncdata5 <- nc_open("full_data_monthly_v2020_1961_1970_05.nc")
ncdata6 <- nc_open("full_data_monthly_v2020_1971_1980_05.nc")
ncdata7 <- nc_open("full_data_monthly_v2020_1981_1990_05.nc")
ncdata8 <- nc_open("full_data_monthly_v2020_1991_2000_05.nc")
ncdata9 <- nc_open("full_data_monthly_v2020_2001_2010_05.nc")
ncdata10 <- nc_open("full_data_monthly_v2020_2011_2019_05.nc")

#get precipitation data
precip1 <- ncvar_get(ncdata1,'precip',start = c(1,1,1))
precip2 <- ncvar_get(ncdata2,'precip',start = c(1,1,1))
precip3 <- ncvar_get(ncdata3,'precip',start = c(1,1,1))
precip4 <- ncvar_get(ncdata4,'precip',start = c(1,1,1))
precip5 <- ncvar_get(ncdata5,'precip',start = c(1,1,1))
precip6 <- ncvar_get(ncdata6,'precip',start = c(1,1,1))
precip7 <- ncvar_get(ncdata7,'precip',start = c(1,1,1))
precip8 <- ncvar_get(ncdata8,'precip',start = c(1,1,1))
precip9 <- ncvar_get(ncdata9,'precip',start = c(1,1,1))
precip10 <- ncvar_get(ncdata10,'precip',start = c(1,1,1))


# Precipitation of Exeter
precip_e1 <- precip1[232,79,]
precip_e2 <- precip2[232,79,]
precip_e3 <- precip3[232,79,]
precip_e4 <- precip4[232,79,]
precip_e5 <- precip5[232,79,]
precip_e6 <- precip6[232,79,]
precip_e7 <- precip7[232,79,]
precip_e8 <- precip8[232,79,]
precip_e9 <- precip9[232,79,]
precip_e10 <- precip10[232,79,]

# Precipitation of London
```

```
precip_L1 <- precip1[359,77,]
precip_L2 <- precip2[359,77,]
precip_L3 <- precip3[359,77,]
precip_L4 <- precip4[359,77,]
precip_L5 <- precip5[359,77,]
precip_L6 <- precip6[359,77,]
precip_L7 <- precip7[359,77,]
precip_L8 <- precip8[359,77,]
precip_L9 <- precip9[359,77,]
precip_L10 <- precip10[359,77,]

Exeter <- data.frame(c(precip_e1,precip_e2,precip_e3,precip_e4,precip_e5,precip_e6,
                       precip_e7,precip_e8,precip_e9,precip_e10))

London <- data.frame(c(precip_L1,precip_L2,precip_L3,precip_L4,precip_L5,precip_L6,
                       precip_L7,precip_L8,precip_L9,precip_L10))

colnames(Exeter)<- "Exeter"
colnames(London)<- "London"

precip <- cbind(Exeter,London)

#plot time series of Exeter and London (together)
precip <- ts(precip,frequency = 12, start = c(1921,1))
plot(precip,plot.type="single",lty=1:2,ylab="",col=c("red","green"))
title("Precipitation")
legend("topleft",colnames(precip),lty=1:2,col = c("red","green"))

#plot time series of Exeter and London (separate)
#Exeter
par(mfrow=c(2,2))
Exeter <- ts(Exeter,frequency = 12,start = c(1921,1))
plot(Exeter)
abline(lm(Exeter~time(Exeter)),col = "red")
title("Time series of Precipitation")
acf(Exeter)
#London
London <- ts(London,frequency = 12,start = c(1921,1))
plot(London)
abline(lm(London~time(London)),col = "red")
title("Time series of Precipitation")
acf(London)

# STL
stl_e <- stl(precip[,1],"per")
stl_l <- stl(precip[,1],"per")
plot(stl_e,main="STL decomposition of Exeter precipitation")
plot(stl_l,main="STL decomposition of London precipitation")

# HW
e1 <- HoltWinters(Exeter,seasonal = "multiplicative")
plot(e1$fit,main="Holt-Winters decompositon of Exeter precipitation")
L1 <- HoltWinters(London,seasonal = "multiplicative")
```

```r
plot(L1$fit,main="Holt-Winters decompositon of London precipitation")

# the Residual density histogram and density estimation, Normal Q-Q Plot and Residual Series of HW in E
er <- e1$x-e1$fitted[,1]
par(mfrow=c(1,3))
hist(er,main="Residuals",col =4,
     xlab=expression(x-hat(x)),probability=T)
lines(density(er),lwd=2)
rug(er)
qqnorm(er)
qqline(er)
plot(er,ylab="Residual")

# using HW to predict
e <- window(Exeter,end=c(2010,1),frequency=12)
ae <- HoltWinters(e,seasonal = "multiplicative")
Ye <- predict(ae,n.ahead=120,prediction.interval = FALSE)
plot(Exeter,ylim=range(c(e,Ye)),lty=3,lwd=2)
lines(Ye,lwd=2,col="blue")
title("Holt Winters Model for Exeter precipitation")
legend("topleft",c("Original series","Predicted series"),
       lty = c(2,1),lwd=2,col = c("black","blue"))

# ARIMA
#Fit Exeter data using automatic ARIMA method
library(forecast)
re1 <- auto.arima(Exeter,ic="bic")
re1
# Test the residuals of the model fitted by the ARIMA method
library(portes)
par(mfrow=c(2,2))
plot(MahdiMcLeod(re1$residuals,1:60,order = 2)[,4],xlab="Lag",ylab="p-value",
     main="General variance tests",pch=16)
abline(h=0.05,lty=2)
plot(LjungBox(re1$residuals,1:60,order = 2)[,4],xlab="Lag",ylim=c(0,1),
     ylab="p-value",main="Ljung-Box tests",pch=16)
abline(h=0.05,lty=2)
acf(re1$residuals,main="ACF of residuals",lag.max = 60)
plot(re1$residuals,type="o",ylab="Residuals",pch=16)
title("Residual series")
abline(h=0,lty=2,col="red")

# Using ARIMA to predict
e <- window(Exeter,end=c(2010,1),frequency=12)
a <- auto.arima(e)
z <- forecast(a,h=120,prediction.interval = FALSE)
plot(z,ylim=c(min(e),max(e)),ylab="Precipitation")

# Perform KPSS unit root test on precipitation data
library(tseries)
for(i in 1:2){
        print(kpss.test(precip[,i],null = "Level"))
        print(kpss.test(precip[,i],null = "Trend"))
```

```
        print(kpss.test(diff(precip[,i]),null = "Level"))
        print(kpss.test(diff(precip[,i]),null = "Trend"))
}
# VAR
library(vars)
# predict from 2011 to 2020
precip1 <- ts(precip,start = c(1920,1),end = c(2010,12),frequency = 12)
z1 <- VAR(precip1,p=100,type="both")
summary(z1)
zp <- predict(z1,n.ahead=120,ci=0.95)
plot(zp)
# predict from 2021 to 2030
precip2 <- ts(precip,start = c(1920,1),frequency = 12)
z2 <- VAR(precip2,p=100,type="both")
summary(z2)
zp2 <- predict(z2,n.ahead=120,ci=0.95)
plot(zp2)

#### Time Series of temperature Data analysis####
# read Data
airt <- nc_open("air.mon.mean.nc")
air <- ncvar_get(airt,"air",start = c(1,1,1))
air1 <- air[,,1]
# get Exeter data and change it to time series data
Exeter_tem <- air[368,77,25:852]-273
Exeter <- as.data.frame(Exeter_tem)
Exeter_tem_ts <- ts(Exeter_tem,frequency = 12,start = c(1951,1))
# get London data and change it to time series data
London_tem <- air[359,77,25:852]-273
London_tem_ts <- ts(London_tem,frequency = 12,start = c(1951,1))

temp <- cbind(Exeter_tem_ts,London_tem_ts)
colnames(temp)<- c("Exeter","London")

# plot time series of Exeter and London (together)
plot(temp,plot.type="single",lty=1:2,ylab="",col=c("red","green"))
title("Temperature")
legend("topleft",colnames(temp),lty=1:2,col = c("red","green"))

# plot time series of Exeter and London (separate)
#Exeter
plot(Exeter_tem_ts,ylab="")
abline(lm(Exeter_tem_ts~time(Exeter_tem_ts)),col = "red")
title("Time series of Exeter Temperature")
acf(Exeter_tem_ts)
#London
plot(London_tem_ts,ylab="")
abline(lm(London_tem_ts~time(London_tem_ts)),col = "red")
title("Time series of London Temperature")
acf(London_tem_ts)

# STL
stl_et <- stl(temp[,1],"per")
```

```
stl_lt <- stl(temp[,1],"per")
plot(stl_et,main="STL decomposition of Exeter Temperature")
plot(stl_lt,main="STL decomposition of London Temperature")

# Perform KPSS unit root test on temperature data
library(tseries)
for(i in 1:2){
        print(kpss.test(temp[,i],null = "Level"))
        print(kpss.test(temp[,i],null = "Trend"))
        print(kpss.test(diff(temp[,i]),null = "Level"))
        print(kpss.test(diff(temp[,i]),null = "Trend"))
}


#### The correlation between precipitation and temperature####
rm(list = ls())
#### temperature data####
# read data
airt <- nc_open("air.mon.mean.nc")
air <- ncvar_get(airt,"air",start = c(1,1,1))
# Exeter data
Exeter_tem <- air[368,77,25:852]-273
# London data
London_tem <- air[359,77,25:852]-273

Exeter_tem <- data.frame(Exeter_tem)
Exeter_tem$group <- "Exeter"
colnames(Exeter_tem) <- c("Temperature","group")
London_tem  <- data.frame(London_tem )
London_tem $group <- "London"
colnames(London_tem ) <- c("Temperature","group")
data <- rbind(Exeter_tem,London_tem)
summary(data)

#### precipitation data ####
ncdata5 <- nc_open("full_data_monthly_v2020_1951_1960_05.nc")
ncdata6 <- nc_open("full_data_monthly_v2020_1961_1970_05.nc")
ncdata7 <- nc_open("full_data_monthly_v2020_1971_1980_05.nc")
ncdata8 <- nc_open("full_data_monthly_v2020_1981_1990_05.nc")
ncdata9 <- nc_open("full_data_monthly_v2020_1991_2000_05.nc")
ncdata10 <- nc_open("full_data_monthly_v2020_2001_2010_05.nc")
ncdata11 <- nc_open("full_data_monthly_v2020_2011_2019_05.nc")

precip5 <- ncvar_get(ncdata5,'precip',start = c(1,1,1))
precip6 <- ncvar_get(ncdata6,'precip',start = c(1,1,1))
precip7 <- ncvar_get(ncdata7,'precip',start = c(1,1,1))
precip8 <- ncvar_get(ncdata8,'precip',start = c(1,1,1))
precip9 <- ncvar_get(ncdata9,'precip',start = c(1,1,1))
precip10 <- ncvar_get(ncdata10,'precip',start = c(1,1,1))
precip11 <- ncvar_get(ncdata11,'precip',start = c(1,1,1))

# Precipitation of Exeter
precip_e5 <- precip5[367,78,]
```

```r
precip_e6 <- precip6[367,78,]
precip_e7 <- precip7[367,78,]
precip_e8 <- precip8[367,78,]
precip_e9 <- precip9[367,78,]
precip_e10 <- precip10[367,78,]
precip_e11 <- precip11[367,78,]

# Precipitation of London
precip_L5 <- precip5[359,77,]
precip_L6 <- precip6[359,77,]
precip_L7 <- precip7[359,77,]
precip_L8 <- precip8[359,77,]
precip_L9 <- precip9[359,77,]
precip_L10 <- precip10[359,77,]
precip_L11 <- precip11[359,77,]

Exeter <- data.frame(c(precip_e5,precip_e6,precip_e7,precip_e8
                       ,precip_e9,precip_e10,precip_e11))

London <- data.frame(c(precip_L5,precip_L6, precip_L7,precip_L8,
                       precip_L9,precip_L10,precip_L11))

colnames(Exeter)<- "Exeter"
colnames(London)<- "London"

Exeter <- data.frame(Exeter)
Exeter$group <- "Exeter"
colnames(Exeter) <- c("Precipitation","group")

London <- data.frame(London)
London$group <- "London"
colnames(London) <- c("Precipitation","group")
data_p <- rbind(Exeter,London)
summary(data_p)
data$precipitation <- data_p$Precipitation

# fit together
fit <- lm(Temperature~precipitation,data)
summary(fit)
par(mfrow=c(2,2))
plot(fit)

# fit separate
#Exeter
fit_E <- lm(Temperature~precipitation,data[1:876,])
summary(fit_E)
par(mfrow=c(2,2))
plot(fit_E)

#London
fit_L<- lm(Temperature~precipitation,data[877:1752,])
summary(fit_L)
par(mfrow=c(2,2))
```

```
plot(fit_L)


#using different color to plot
ggplot(data,aes(y=precipitation,x=Temperature,colour=group)) + geom_point(size=3,shape=19)
ggplot(data,aes(y=precipitation,x=Temperature)) + geom_point(size=3,shape=19)


#95%CI
p0.99 <- ggplot(data,aes(x=Temperature,y=precipitation,colour=group)) + geom_point(shape=19) + xlab("Te
p0.99

p0 <- ggplot(data,aes(y=precipitation,x=Temperature)) + geom_point(size=3,shape=19)+ xlab("Temperature"
p0
p1 <- p0+ theme_bw() +
        theme(panel.grid.major=element_line(colour=NA),
              panel.background = element_rect(fill = "transparent",colour = NA),
              plot.background = element_rect(fill = "transparent",colour = NA),
              panel.grid.minor = element_blank())

p1


p <- p0.99+ theme_bw() +
        theme(panel.grid.major=element_line(colour=NA),
              panel.background = element_rect(fill = "transparent",colour = NA),
              plot.background = element_rect(fill = "transparent",colour = NA),
              panel.grid.minor = element_blank())
p

#### Temperature difference analysis####
library("agricolae")
library("ggpubr")
library("vegan")
library(EasyStat)
library(ggplot2)
#### temperature####
library(ncdf4)
# temperature data
airt <- nc_open("air.mon.mean.nc")
air <- ncvar_get(airt,"air",start = c(1,1,1))
# Exeter data
Exeter_tem <- air[368,77,25:852]-273
# London data
London_tem <- air[359,77,25:852]-273

Exeter_tem <- data.frame(Exeter_tem)
Exeter_tem$group <- "Exeter"
colnames(Exeter_tem) <- c("Temperature","group")
London_tem  <- data.frame(London_tem )
London_tem $group <- "London"
colnames(London_tem ) <- c("Temperature","group")
data <- rbind(Exeter_tem,London_tem)
```

```r
summary(data)

#Statistical analysis based on a single indicator
#Normal test and analysis of homogeneity of variance NorNorCVTest
#Use Cases
NorCV = NorNorCVTest(data = data, i= 1, method_cv = "leveneTest")
#Extract normal test results
NorCV[[1]]
#Extract the results of the homogeneity of variance test
NorCV[[2]]

#Non-parametric test (KwWlx)
res = KwWlx(data = data, i= 1)
# Call non-parameter pairwise comparison result: letter mark display
res[[1]]
#Using table shows the difference between the two results
res[[2]]

#Box plot showing the results of analysis of variance or non-parametric test (aovMuiBoxP)
#Use Cases
library(tidyverse)
PlotresultBox = aovMuiBoxP(data = data, i=1, sig_show ="abc", result = res[[1]])
#Extract pictures
p = PlotresultBox[[1]]
p


#### Difference analysis of rainfall data ####
rm(list = ls())
#read data
ncdata1 <- nc_open("full_data_monthly_v2020_1921_1930_05.nc")
ncdata2 <- nc_open("full_data_monthly_v2020_1931_1940_05.nc")
ncdata3 <- nc_open("full_data_monthly_v2020_1941_1950_05.nc")
ncdata4 <- nc_open("full_data_monthly_v2020_1951_1960_05.nc")
ncdata5 <- nc_open("full_data_monthly_v2020_1961_1970_05.nc")
ncdata6 <- nc_open("full_data_monthly_v2020_1971_1980_05.nc")
ncdata7 <- nc_open("full_data_monthly_v2020_1981_1990_05.nc")
ncdata8 <- nc_open("full_data_monthly_v2020_1991_2000_05.nc")
ncdata9 <- nc_open("full_data_monthly_v2020_2001_2010_05.nc")
ncdata10 <- nc_open("full_data_monthly_v2020_2011_2019_05.nc")

#get precipitation data
precip1 <- ncvar_get(ncdata1,'precip',start = c(1,1,1))
precip2 <- ncvar_get(ncdata2,'precip',start = c(1,1,1))
precip3 <- ncvar_get(ncdata3,'precip',start = c(1,1,1))
precip4 <- ncvar_get(ncdata4,'precip',start = c(1,1,1))
precip5 <- ncvar_get(ncdata5,'precip',start = c(1,1,1))
precip6 <- ncvar_get(ncdata6,'precip',start = c(1,1,1))
precip7 <- ncvar_get(ncdata7,'precip',start = c(1,1,1))
precip8 <- ncvar_get(ncdata8,'precip',start = c(1,1,1))
precip9 <- ncvar_get(ncdata9,'precip',start = c(1,1,1))
precip10 <- ncvar_get(ncdata10,'precip',start = c(1,1,1))
```

```r
# Precipitation of Exeter
precip_e1 <- precip1[232,79,]
precip_e2 <- precip2[232,79,]
precip_e3 <- precip3[232,79,]
precip_e4 <- precip4[232,79,]
precip_e5 <- precip5[232,79,]
precip_e6 <- precip6[232,79,]
precip_e7 <- precip7[232,79,]
precip_e8 <- precip8[232,79,]
precip_e9 <- precip9[232,79,]
precip_e10 <- precip10[232,79,]

# Precipitation of London
precip_L1 <- precip1[359,77,]
precip_L2 <- precip2[359,77,]
precip_L3 <- precip3[359,77,]
precip_L4 <- precip4[359,77,]
precip_L5 <- precip5[359,77,]
precip_L6 <- precip6[359,77,]
precip_L7 <- precip7[359,77,]
precip_L8 <- precip8[359,77,]
precip_L9 <- precip9[359,77,]
precip_L10 <- precip10[359,77,]

Exeter <- data.frame(c(precip_e1,precip_e2,precip_e3,precip_e4,precip_e5,precip_e6,
                       precip_e7,precip_e8,precip_e9,precip_e10))

London <- data.frame(c(precip_L1,precip_L2,precip_L3,precip_L4,precip_L5,precip_L6,
                       precip_L7,precip_L8,precip_L9,precip_L10))

colnames(Exeter)<- "Exeter"
colnames(London)<- "London"

Exeter <- data.frame(Exeter)
Exeter$group <- "Exeter"
colnames(Exeter) <- c("Precipitation","group")

London <- data.frame(London)
London$group <- "London"
colnames(London) <- c("Precipitation","group")
data_p <- rbind(Exeter,London)
summary(data_p)


#Statistical analysis based on a single indicator
#Normal test and analysis of homogeneity of variance NorNorCVTest
#Use Cases
NorCV = NorNorCVTest(data = data_p, i= 1, method_cv = "leveneTest")
#Extract normal test results
NorCV[[1]]
#Extract the results of the homogeneity of variance test
NorCV[[2]]
```

```
#Non-parametric test (KwWlx)
res = KwWlx(data = data, i= 1)
# Call non-parameter pairwise comparison result: letter mark display
res[[1]]
#Using table shows the difference between the two results
res[[2]]

#Box plot showing the results of analysis of variance or non-parametric test (aovMuiBoxP)
#Use Cases
library(tidyverse)
PlotresultBox = aovMuiBoxP(data = data_p, i=1, sig_show ="abc", result = res[[1]])
#Extract pictures
p = PlotresultBox[[1]]
p

####The relationship of Exeter's precipitation with London's Temperature.####
rm(list = ls())
# temperature data
airt <- nc_open("air.mon.mean.nc")
air <- ncvar_get(airt,"air",start = c(1,1,1))
# Exeter data
Exeter_tem <- air[368,77,25:852]-273
# London data
London_tem <- air[359,77,25:852]-273

Exeter_tem <- data.frame(Exeter_tem)
Exeter_tem$group <- "Exeter"
colnames(Exeter_tem) <- c("Temperature","group")
London_tem  <- data.frame(London_tem )
London_tem $group <- "London"
colnames(London_tem ) <- c("Temperature","group")
data <- rbind(Exeter_tem,London_tem)
summary(data)

#### precipitation data####
ncdata5 <- nc_open("full_data_monthly_v2020_1951_1960_05.nc")
ncdata6 <- nc_open("full_data_monthly_v2020_1961_1970_05.nc")
ncdata7 <- nc_open("full_data_monthly_v2020_1971_1980_05.nc")
ncdata8 <- nc_open("full_data_monthly_v2020_1981_1990_05.nc")
ncdata9 <- nc_open("full_data_monthly_v2020_1991_2000_05.nc")
ncdata10 <- nc_open("full_data_monthly_v2020_2001_2010_05.nc")
ncdata11 <- nc_open("full_data_monthly_v2020_2011_2019_05.nc")

precip5 <- ncvar_get(ncdata5,'precip',start = c(1,1,1))
precip6 <- ncvar_get(ncdata6,'precip',start = c(1,1,1))
precip7 <- ncvar_get(ncdata7,'precip',start = c(1,1,1))
precip8 <- ncvar_get(ncdata8,'precip',start = c(1,1,1))
precip9 <- ncvar_get(ncdata9,'precip',start = c(1,1,1))
precip10 <- ncvar_get(ncdata10,'precip',start = c(1,1,1))
precip11 <- ncvar_get(ncdata11,'precip',start = c(1,1,1))

# Precipitation of Exeter
precip_e5 <- precip5[367,78,]
```

```
precip_e6 <- precip6[367,78,]
precip_e7 <- precip7[367,78,]
precip_e8 <- precip8[367,78,]
precip_e9 <- precip9[367,78,]
precip_e10 <- precip10[367,78,]
precip_e11 <- precip11[367,78,]

# Precipitation of London
precip_L5 <- precip5[359,77,]
precip_L6 <- precip6[359,77,]
precip_L7 <- precip7[359,77,]
precip_L8 <- precip8[359,77,]
precip_L9 <- precip9[359,77,]
precip_L10 <- precip10[359,77,]
precip_L11 <- precip11[359,77,]

Exeter <- data.frame(c(precip_e5,precip_e6,precip_e7,precip_e8,
                       precip_e9,precip_e10,precip_e11))

London <- data.frame(c(precip_L5,precip_L6, precip_L7,precip_L8,
                       precip_L9,precip_L10,precip_L11))

colnames(Exeter)<- "Exeter"
colnames(London)<- "London"

Exeter <- data.frame(Exeter)
Exeter$group <- "Exeter"
colnames(Exeter) <- c("Precipitation","group")

London <- data.frame(London)
London$group <- "London"
colnames(London) <- c("Precipitation","group")
data_p <- rbind(Exeter,London)
summary(data_p)
data$precipitation <- data_p$Precipitation

#fit Exeter precipitation and London Temperature
fit_E <- lm(Exeter[,1]~London_tem[,1])
plotdata <- Exeter
plotdata$Temper <- London_tem$Temperature
plotdata%>%ggplot(aes(x=Precipitation,y=Temper))+geom_point()+
        xlab("Preciptation of Exeter")+ylab("Temperature of London")+
        geom_smooth(aes(x=Precipitation,y=Temper),method = lm,linetype=1,se=TRUE,span=1)+
        theme_bw() +theme(panel.grid.major=element_line(colour=NA),
                          panel.background = element_rect(fill = "transparent",colour = NA),
                          plot.background = element_rect(fill = "transparent",colour = NA),
                          panel.grid.minor = element_blank())
summary(fit_E)
fitted(fit_E)
residuals(fit_E)
par(mfrow=c(2,2))
plot(fit_E)
```