Statistics for spatio-temporal data
Final report

Student: Duc Duong (871492)

Instructor: Prof. Carlo Gaetan

Venice, July 2017

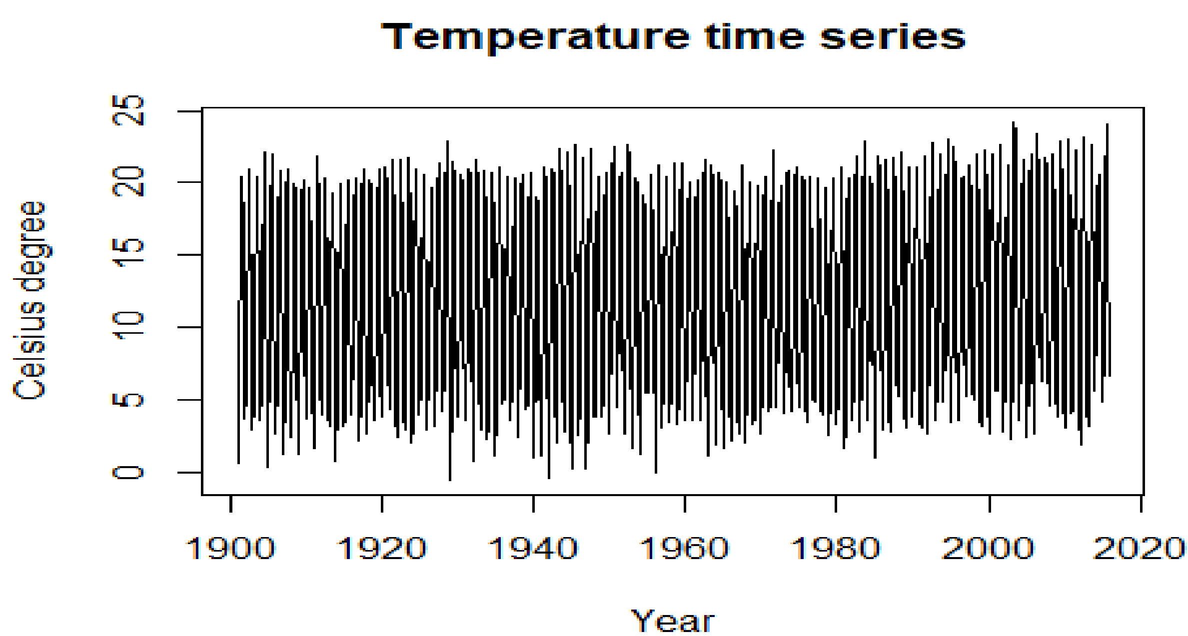
1. Motivation for studying the problem

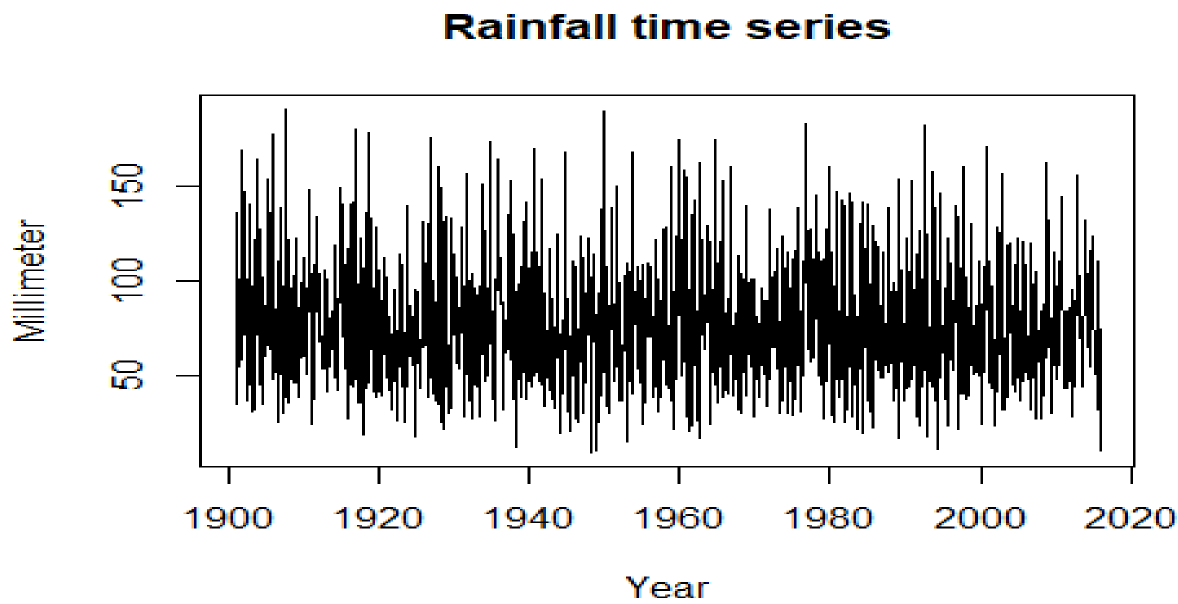
Global warming and its impacts on the environment, human society attracts many scientific research and also social and political interests. That motivates me to do some analysis on the temperature in Italy in order have a better understanding of the problem. On the other hand, I also analysis the rainfall to figure out whether there is a relationship between temperature and rainfall.

2. A description of data

My data is the record of the monthly average temperature (by Celsius degree) and rainfall (by millimeter) in Italy in 115 years (from 01/1901 to 12/2015) Source: [Climate change knowledge portal] (http://sdwebx.worldbank.org/climateportal/index.cfm)

Brief of data:

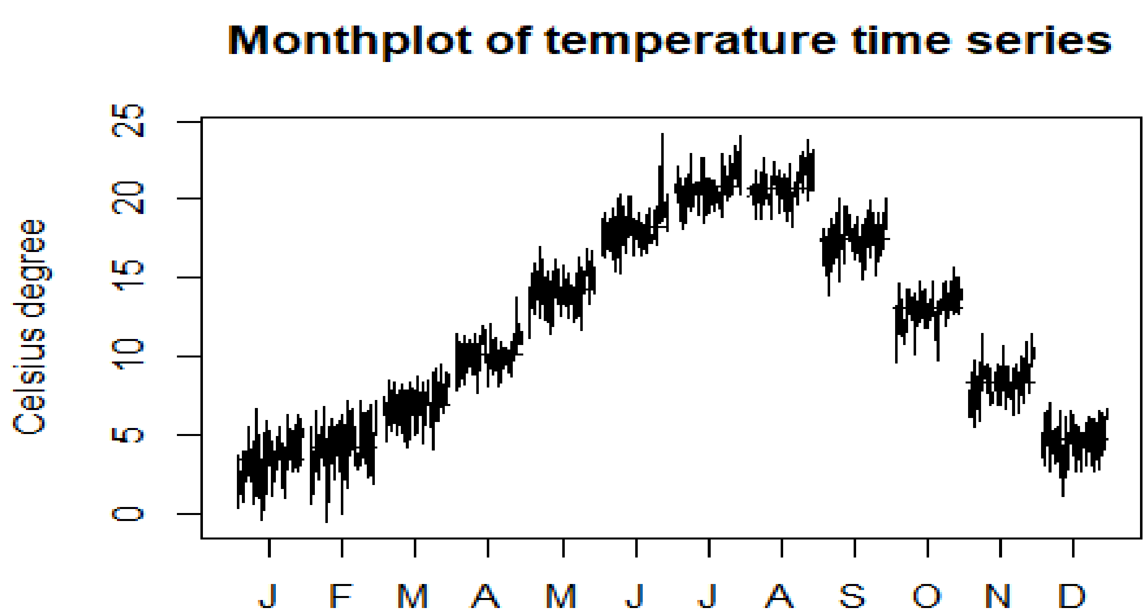


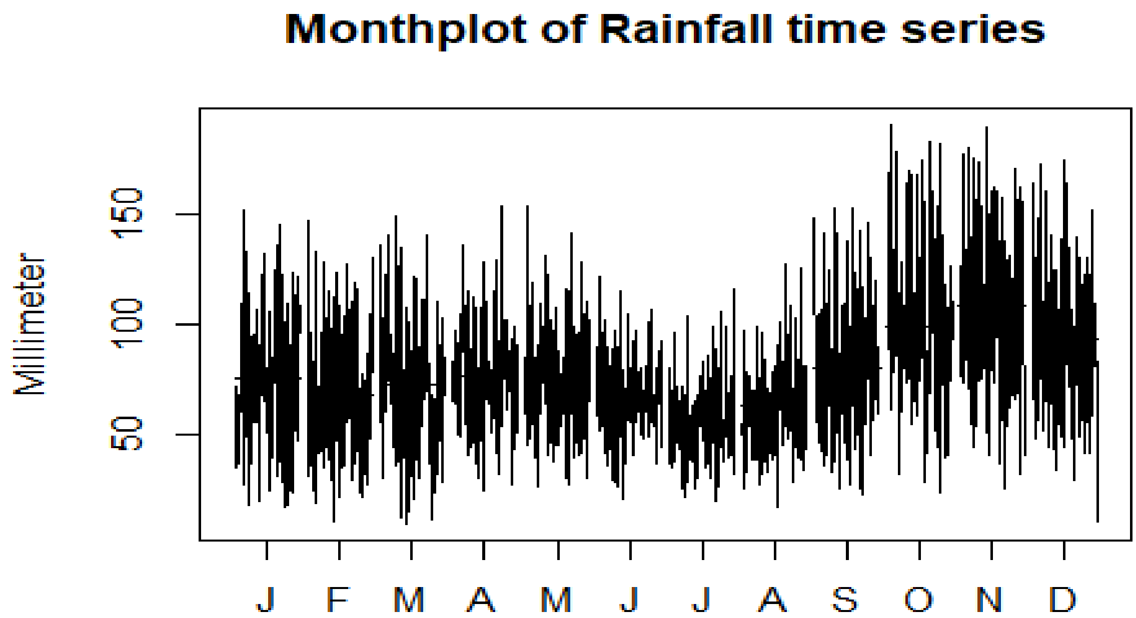


3. Identify possible trends and seasonality. Check the relevant between temperature and rainfall

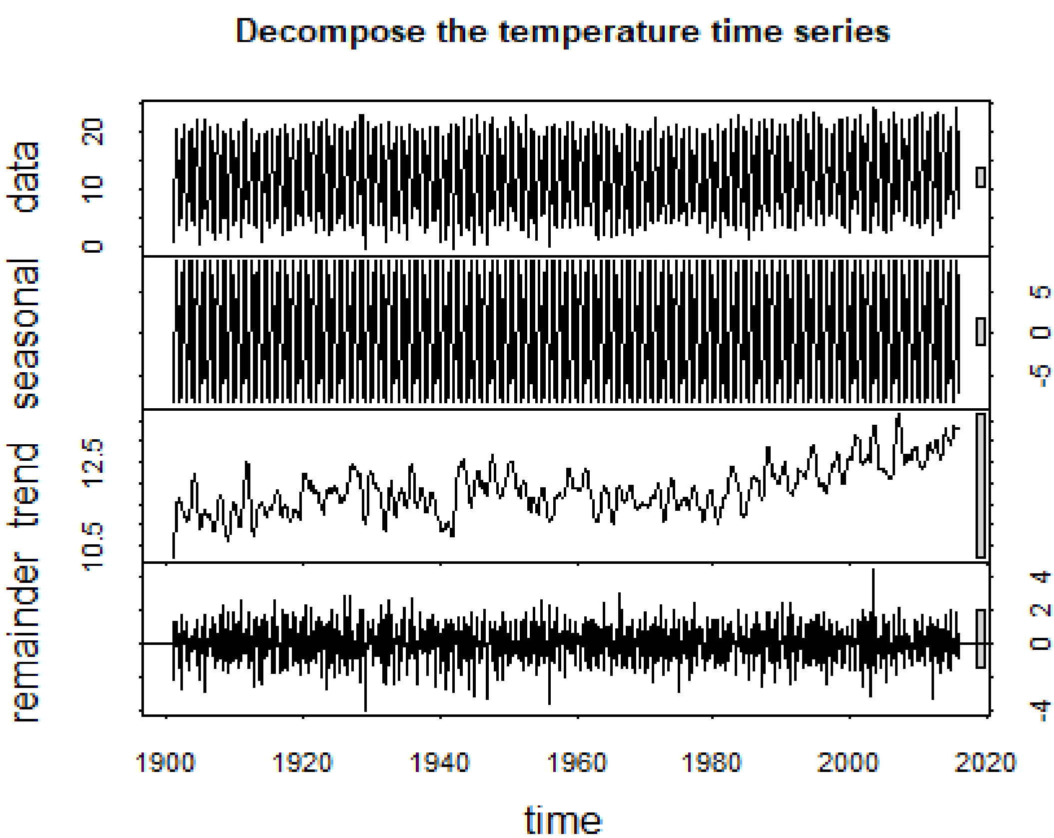
We are considering the monthly time series. As result, my first step is to take a look at the month plot of both time series.

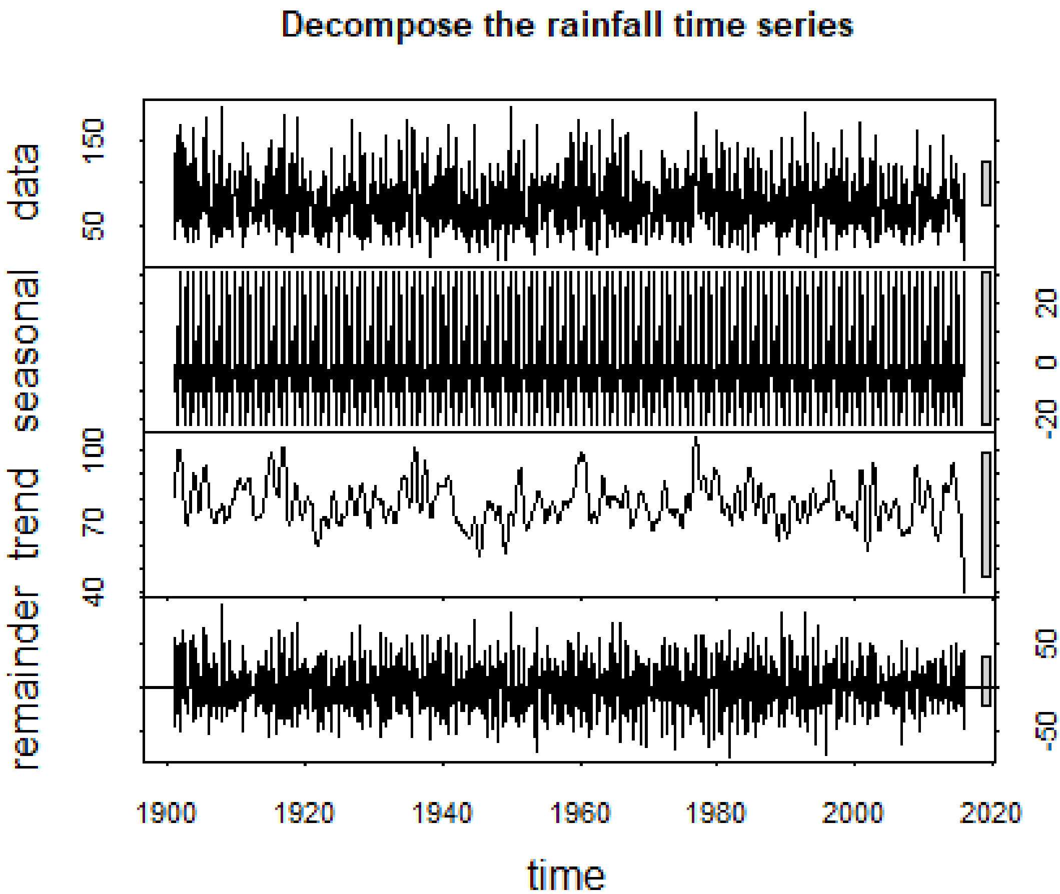
As we expect, the temperature is higher in summer and lower in winter. In contrast, the rainfall is low in July and high from September to the end of each year. It means that there is a seasonality in both time series.





Now, I try to decompose the data to see the effect of each component.





After decomposing data to seasonal, trend and residual. We can see that the seasonal component is quite fit with the comment before. There is no clear trend for rainfall time series and a slight upward trend for temperature.

On the other hand, based on the difference between the seasonality and the trend, I conclude that there is no relevant between the temperature and the rainfall.

4. Check the stationary and evaluate by ARIMA model

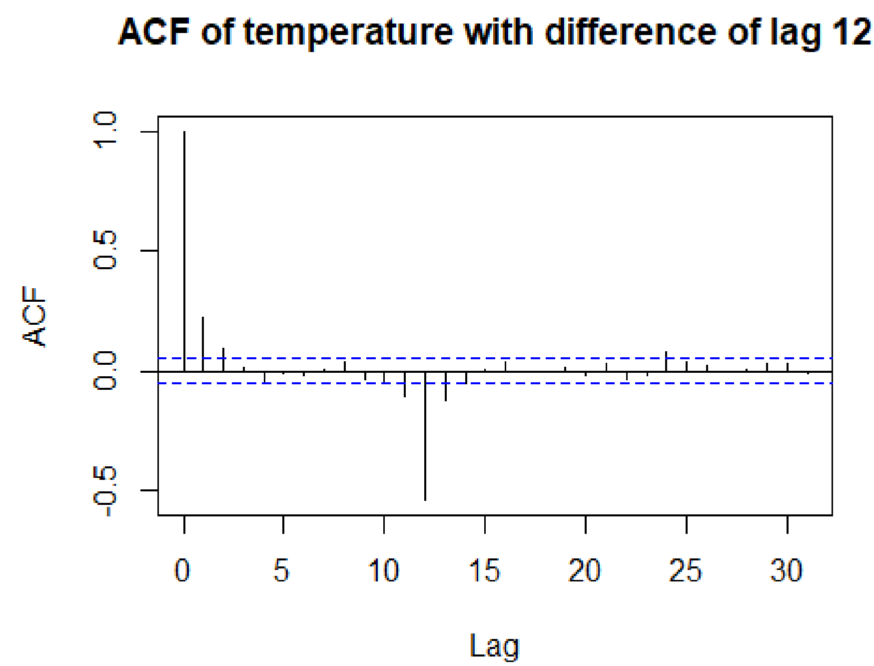
I check the stationary with adf.test() function

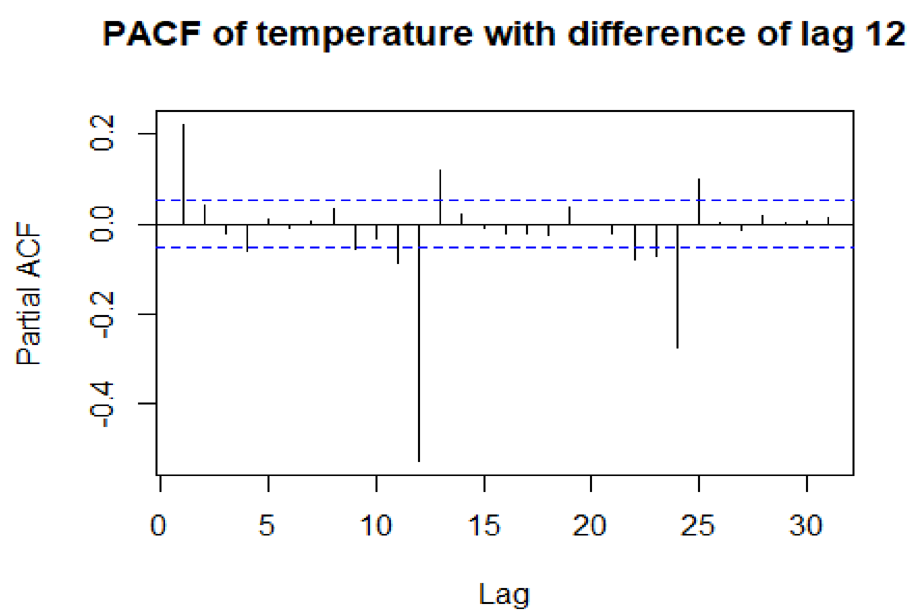
```
##  Augmented Dickey-Fuller Test
## data:  temp.ts
## Dickey-Fuller = -7.7423, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary


##  Augmented Dickey-Fuller Test
## data:  rain.ts
## Dickey-Fuller = -10.387, Lag order = 11, p-value = 0.01
## alternative hypothesis: stationary
```

P-value = 0.01 mean that the time series is stationary. It lead to d=0 in ARIMA model. Because of both monthly time series has seasonality. I will take the difference of lag 12 and then analysis the ACF and PACF graphs:

For the temperature:

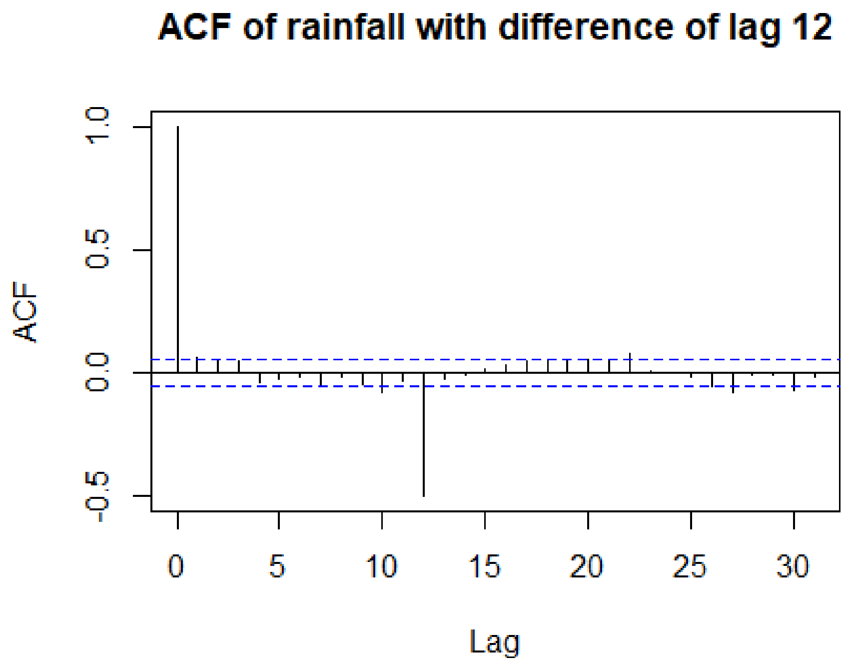


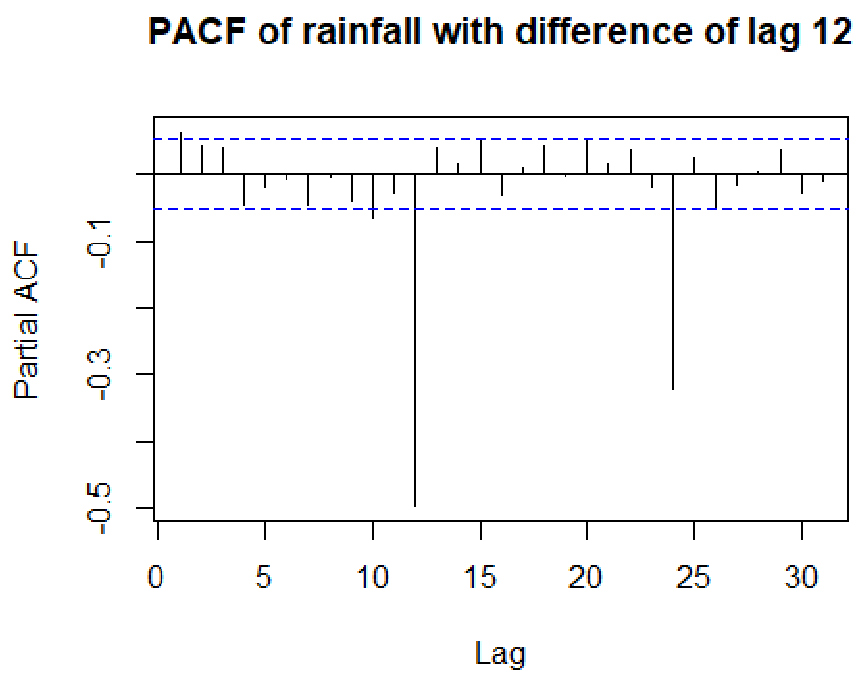


Non-seasonal behavior: The PACF shows a cut off at lag 1 (some strikes at lag 11 and 13 may be affected by the seasonality) and the ACF is exponentially decaying. A non-seasonal AR (1) may be a useful part of the model.
Seasonal behavior: I only consider around lags 12, 24… In the ACF, there are some (negative)

spikes around lag 12 (lag 12 is much higher). The PACF tapers in multiples of 12, which mean the PACF has significant lags at 12, 24. So, I consider seasonal MA (1).
Because we're looking at 12th differences, the model we might try for the original series is ARIMA (1,0,0)×(0,1,1)[12]. I'm also try ARIMA (1,0,0)×(1,1,1)[12] to compare.

For the rainfall:





Non-seasonal behavior: The pattern is not clear. PACF have a strike at lag 1, 10. In ACF there are unclear strikes at lag 1 and maybe 2. So, I will try both (1,0,1) and (1,0,2) for this part. Seasonal behavior: In contrast, this behavior is very clear. ACF show a (negative) spikes at lag 12.PACF has significant and exponentially decaying lags at 12, 24. So, I consider seasonal MA (1).

So, the model for this part is ARIMA (1,0,0)×(0,1,1)[12] and ARIMA (2,0,0)×(0,1,1)[12]

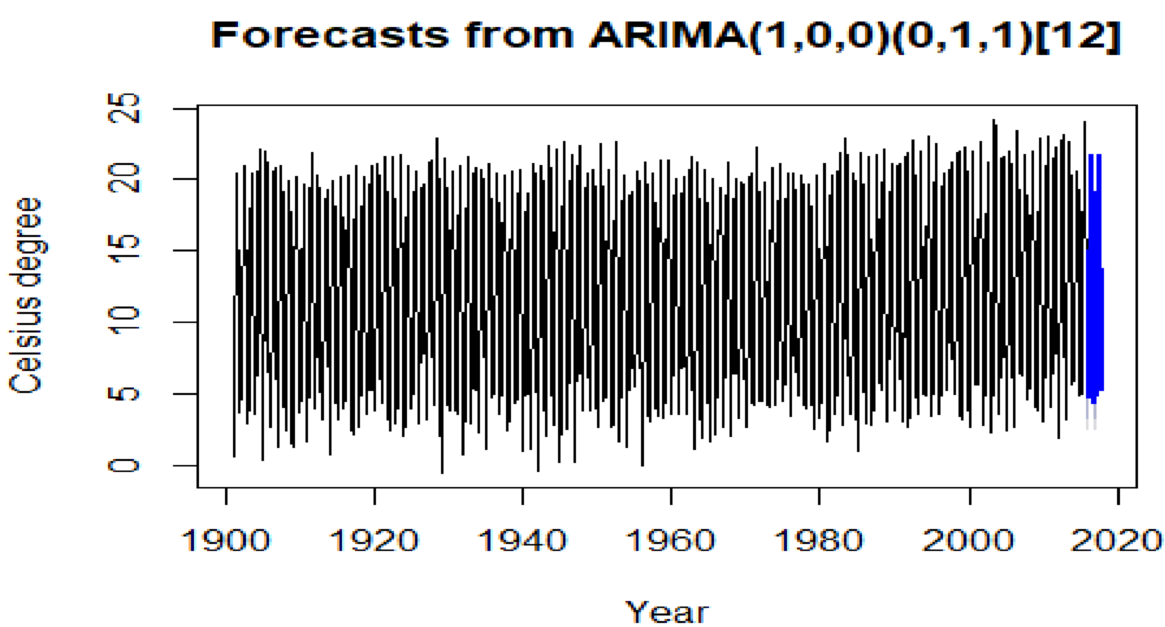
I use AIC and BIC values to evaluate those models. The results are showed in the below table:

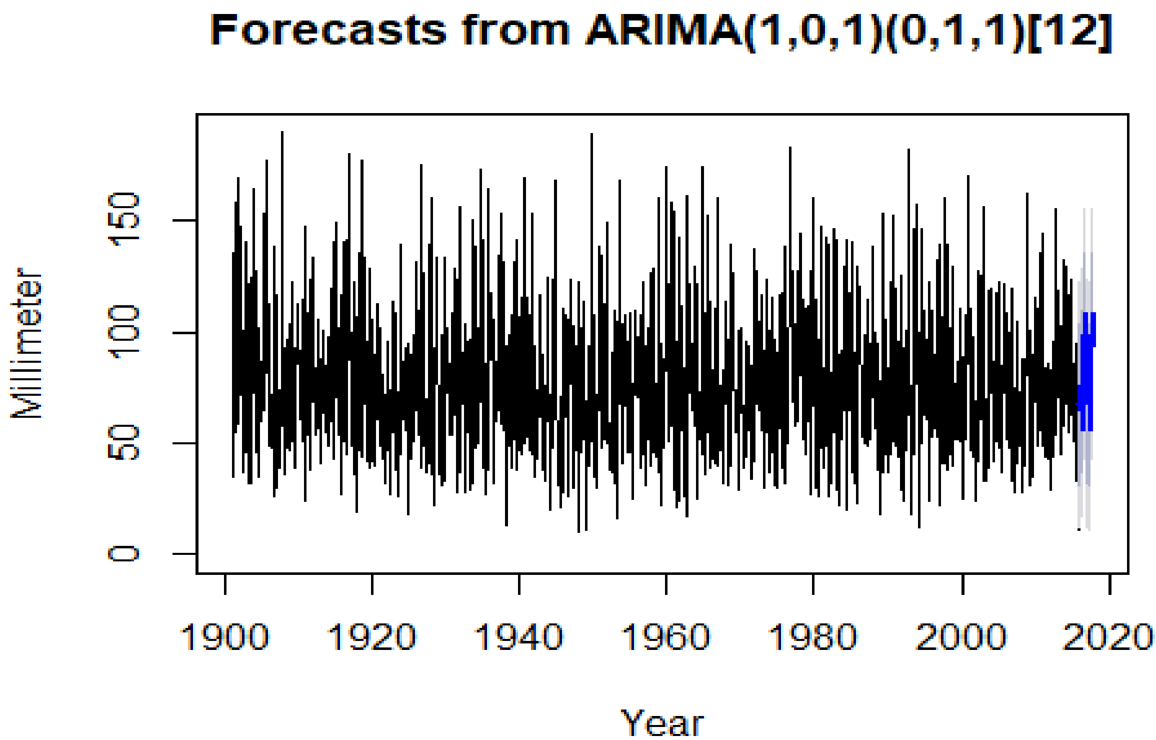
Data	Model	AIC	BIC
Temperature	ARIMA (1,0,0)×(0,1,1)[12]	4269.819	4285.482
	ARIMA (1,0,0)×(1,1,1)[12	4270.098	4290.982
Rainfall	ARIMA (1,0,0)×(0,1,1)[12]	13101.62	13122.5
	ARIMA (2,0,0)×(0,1,1)[12]	13103.62	13129.72

Based on these results, the most suitable models are ARIMA (1,0,0)×(0,1,1)[12] for temperature and ARIMA (1,0,0)×(0,1,1)[12] for rainfall.

I check the residuals of that models. It look like white noise.

5. Forecast some periods





6. Conclusions

Both time series have seasonality, it fits with our knowledge about climate.

In more than 100 years in Italy, the rainfall still remains while the temperature slightly increases. It means that the environment is hotter and hotter.