Covid19

Julien JACQUES

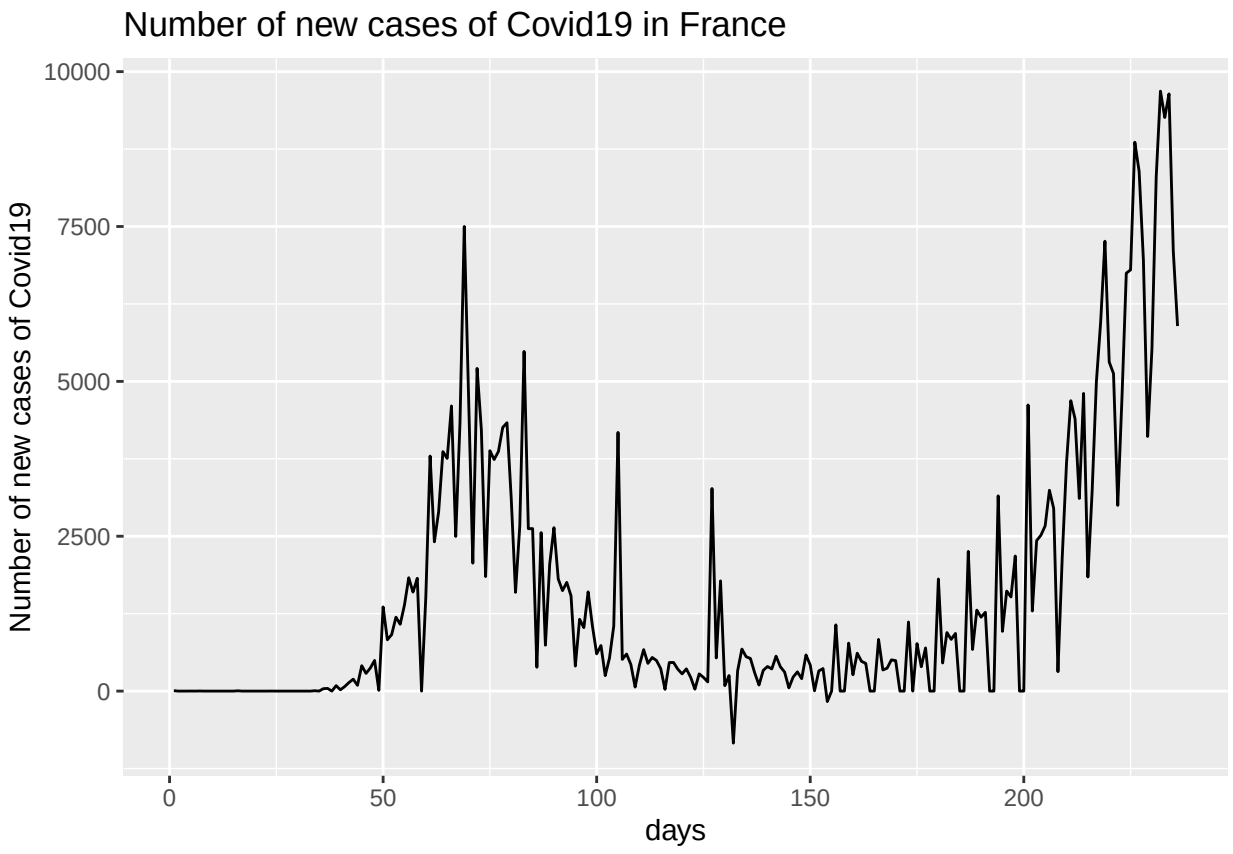
```
covid19 = read.csv("~/Enseignement/Formations/Times series/data/WHO-COVID-19-global-data.csv")
covid19_F=covid19[covid19$Country=="France",]
covid19_F_nc=ts(covid19_F$New_cases)
covid19_F_nd=ts(covid19_F$New_deaths)
```

We plot the number of new cases

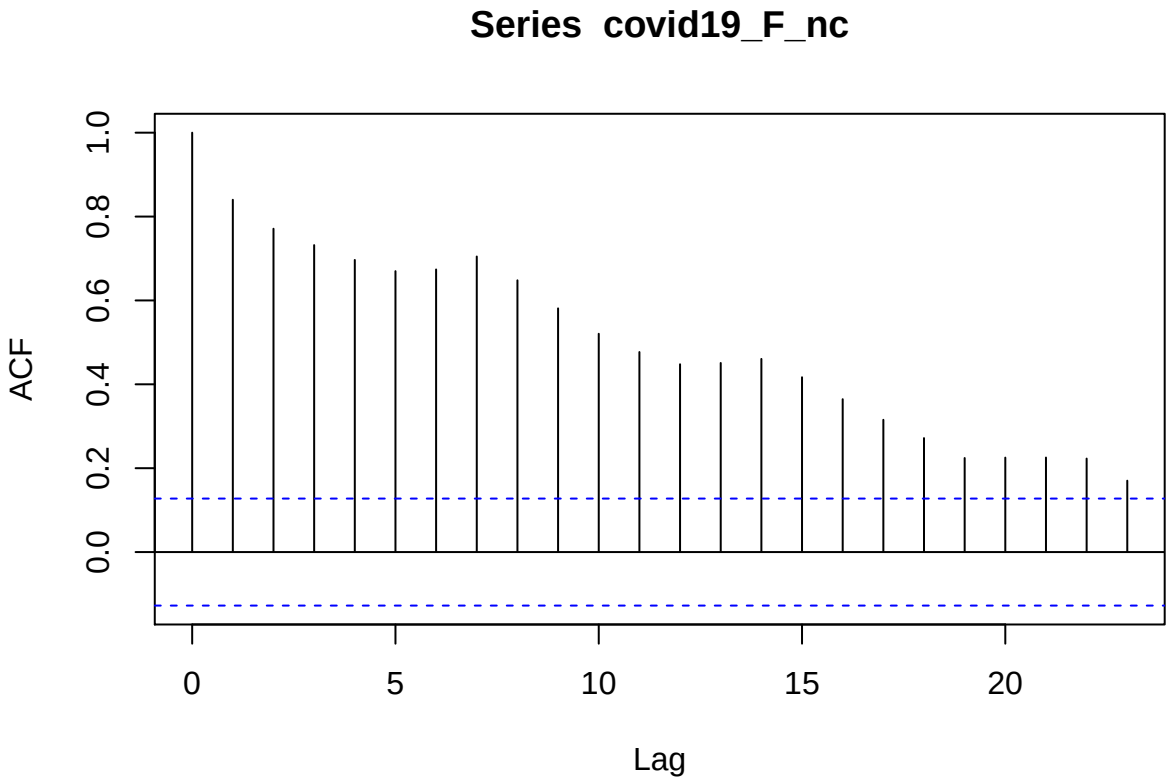
```
library(forecast)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.2
```

```
autoplot(covid19_F_nc) +
ggtitle('Number of new cases of Covid19 in France')+ xlab('days')+
ylab('Number of new cases of Covid19')
```

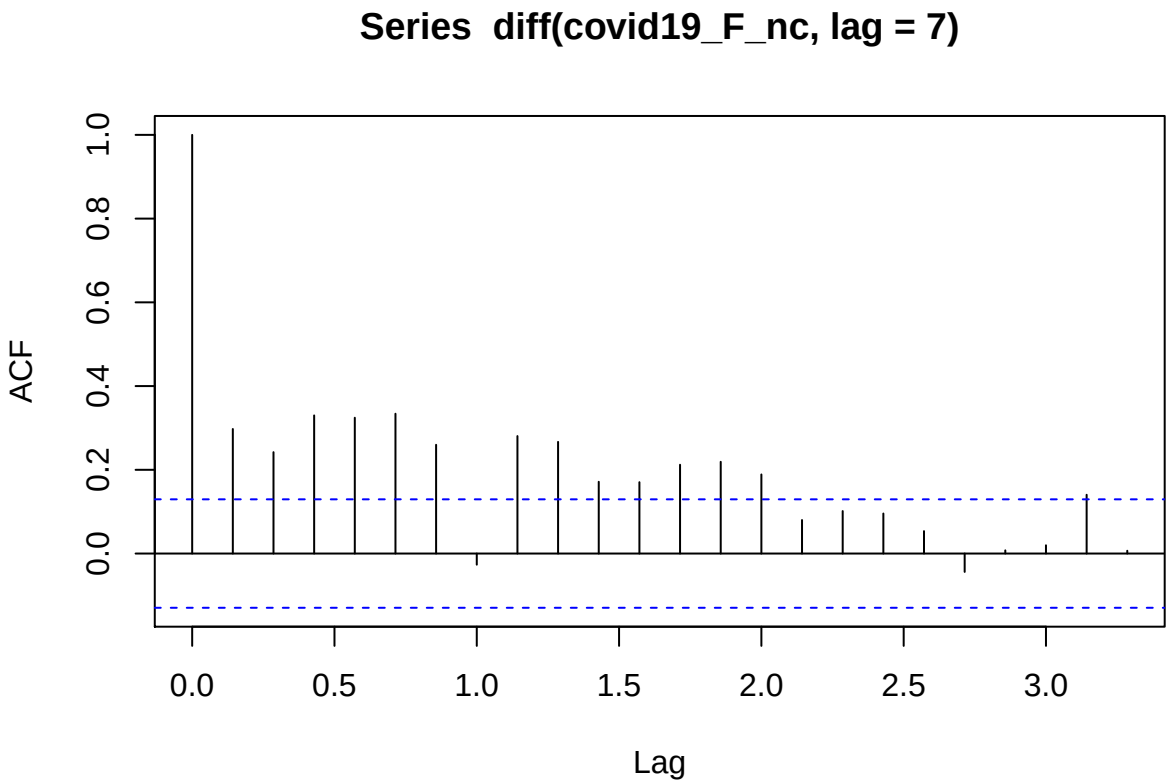



```
acf(covid19_F_nc)
```



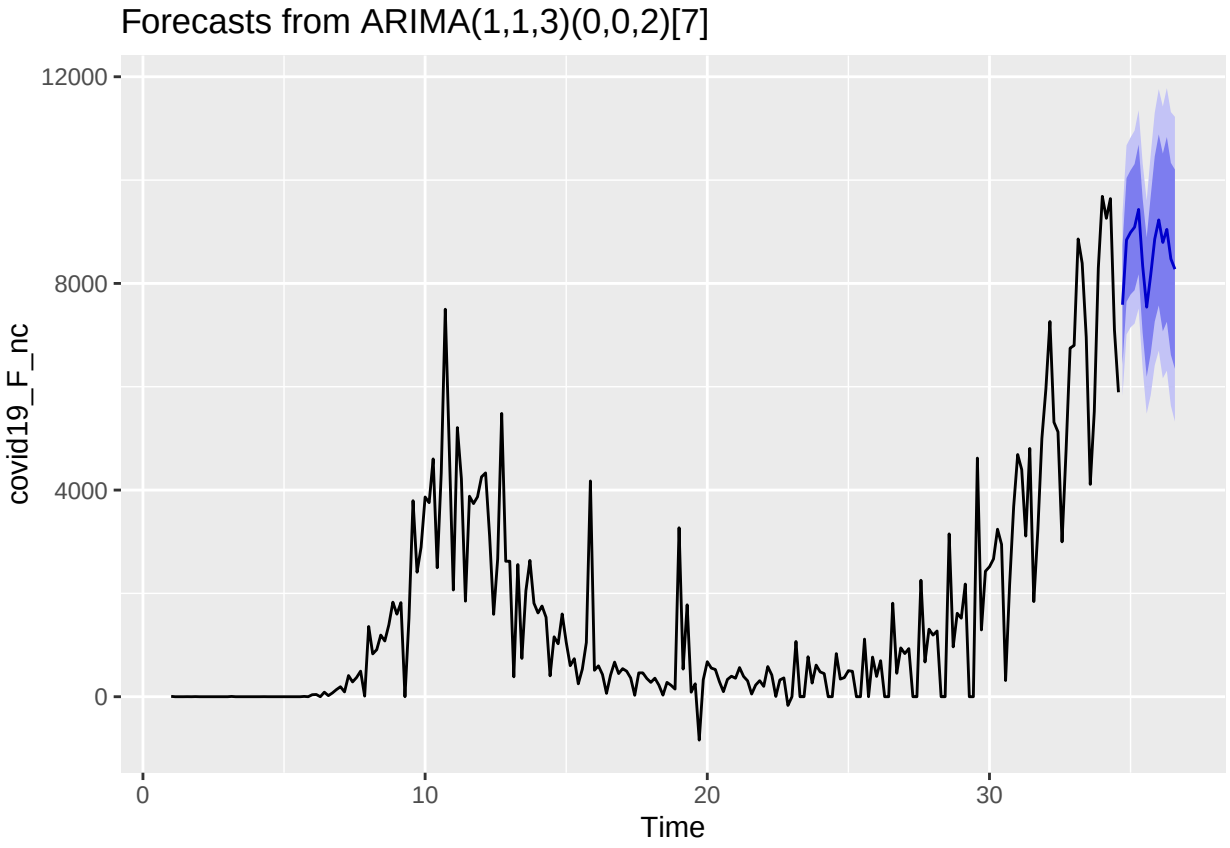
It looks like if there is a trend and a seasonnal pattern of period equal to 7

```
covid19_F_nc=ts(covid19_F_nc,freq=7)
acf(diff(covid19_F_nc,lag=7))
```



Forecasting of the number of new cases

```
fit=auto.arima(covid19_F_nc)
autoplot(forecast(fit,h=14))
```




```
print(forecast(fit,h=14))
```

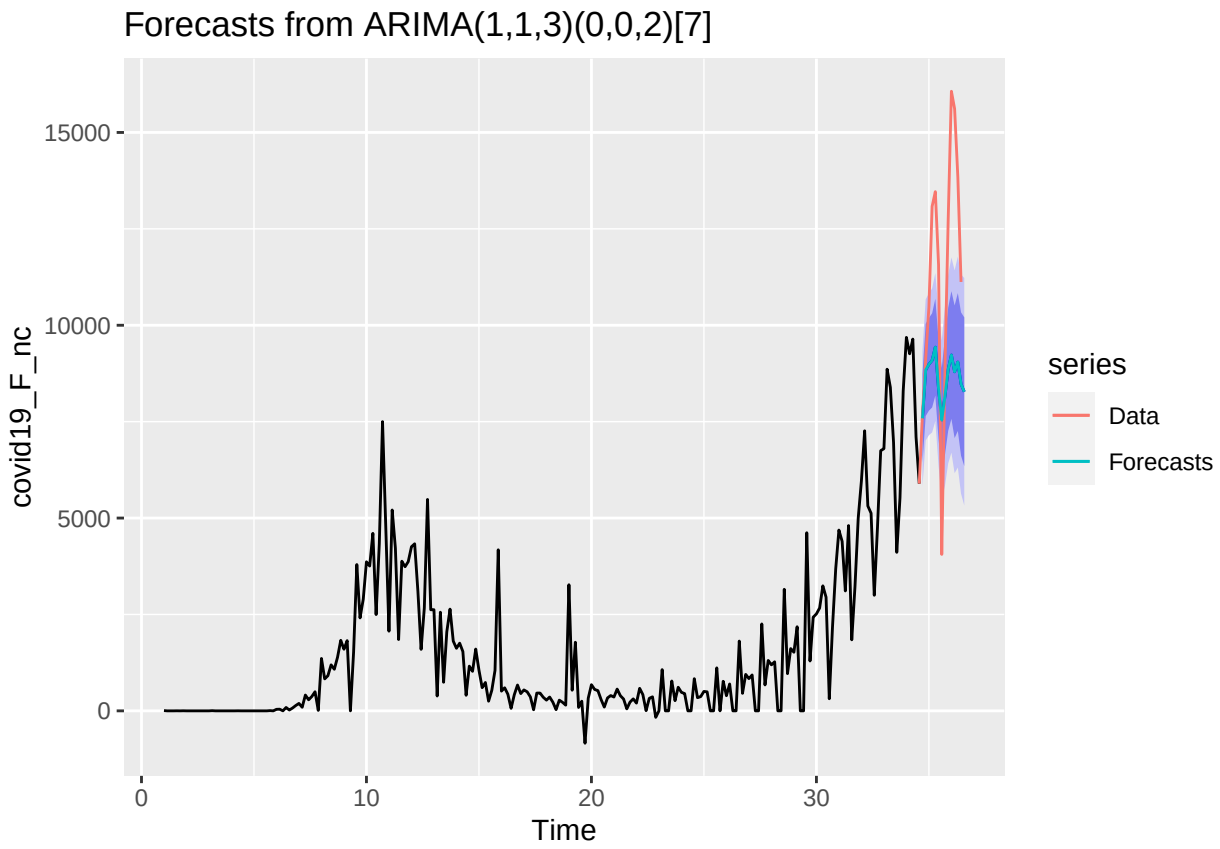
```
##          Point Forecast    Lo 80     Hi 80    Lo 95     Hi 95
## 34.71429       7586.987 6435.536  8738.438 5825.994  9347.980
## 34.85714       8842.091 7645.119 10039.062 7011.481 10672.701
## 35.00000       8986.545 7784.993 10188.097 7148.929 10824.161
## 35.14286       9089.448 7868.535 10310.362 7222.223 10956.674
## 35.28571       9432.374 8177.930 10686.818 7513.867 11350.881
## 35.42857       8340.200 7042.081  9638.319 6354.898 10325.502
## 35.57143       7541.099 6193.126  8889.072 5479.552  9602.645
## 35.71429       8164.517 6642.949  9686.085 5837.480 10491.554
## 35.85714       8859.255 7261.977 10456.534 6416.429 11302.082
## 36.00000       9228.608 7574.163 10883.053 6698.352 11758.864
## 36.14286       8793.009 7073.781 10512.236 6163.677 11422.340
## 36.28571       9045.476 7257.573 10833.379 6311.115 11779.837
## 36.42857       8475.186 6617.152 10333.220 5633.568 11316.804
## 36.57143       8275.703 6347.564 10203.841 5326.869 11224.536
```

We can compare with the true data observed during the 2 course sessions

```
covid19_2 = read.csv("~/Enseignement/Formations/Times series/data/WHO-COVID-19-global-data2.csv")
covid19_F_2=covid19_2[covid19_2$Country=="France",]
covid19_F_2=covid19_F_2[-(1:21),]
```

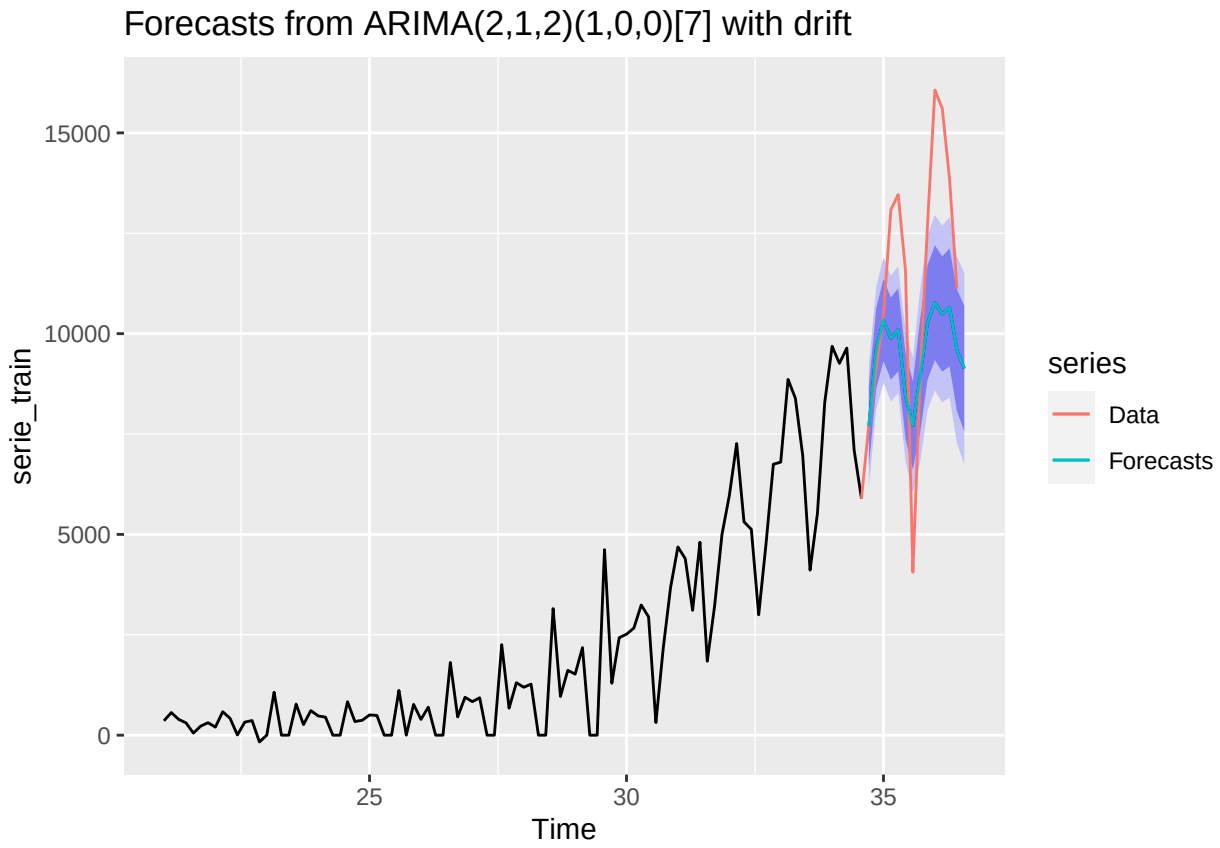
```
covid19_F_2_nc=ts(covid19_F_2$New_cases)
covid19_F_2_nd=ts(covid19_F_2$New_deaths)
covid19_F_2_nc=ts(covid19_F_2_nc,freq=7)
```

```
prev=forecast(fit,h=14)
autoplot(prev) + autolayer(tail(covid19_F_2_nc,14), series="Data")+
  autolayer(prev$mean, series="Forecasts")
```



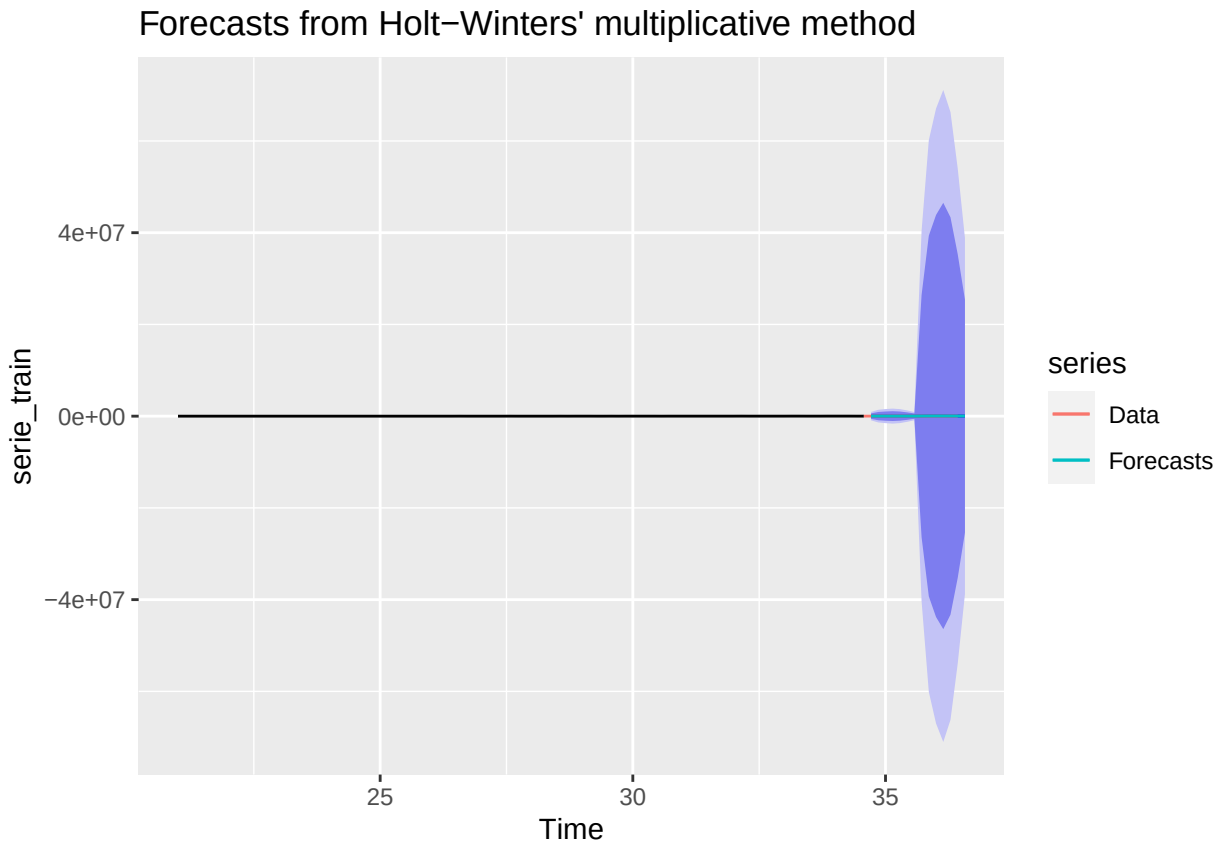
It seems that there were a multiplicative effet of the seasonal part, that have not been taken into account. It seems also that the slope of the trend has been under-evaluated. It may be cause by the fact that the regime of the serie is different in the first 20 weeks. We can try the forecast by removing these elements.

```
serie_train=window(covid19_F_2_nc,start=c(21,1),end=c(34, 5))
serie=window(covid19_F_2_nc,start=c(21,1))
fit=auto.arima(serie_train)
prev=forecast(fit,h=14)
autoplot(prev) + autolayer(tail(covid19_F_2_nc,14), series="Data")+
  autolayer(prev$mean, series="Forecasts")
```



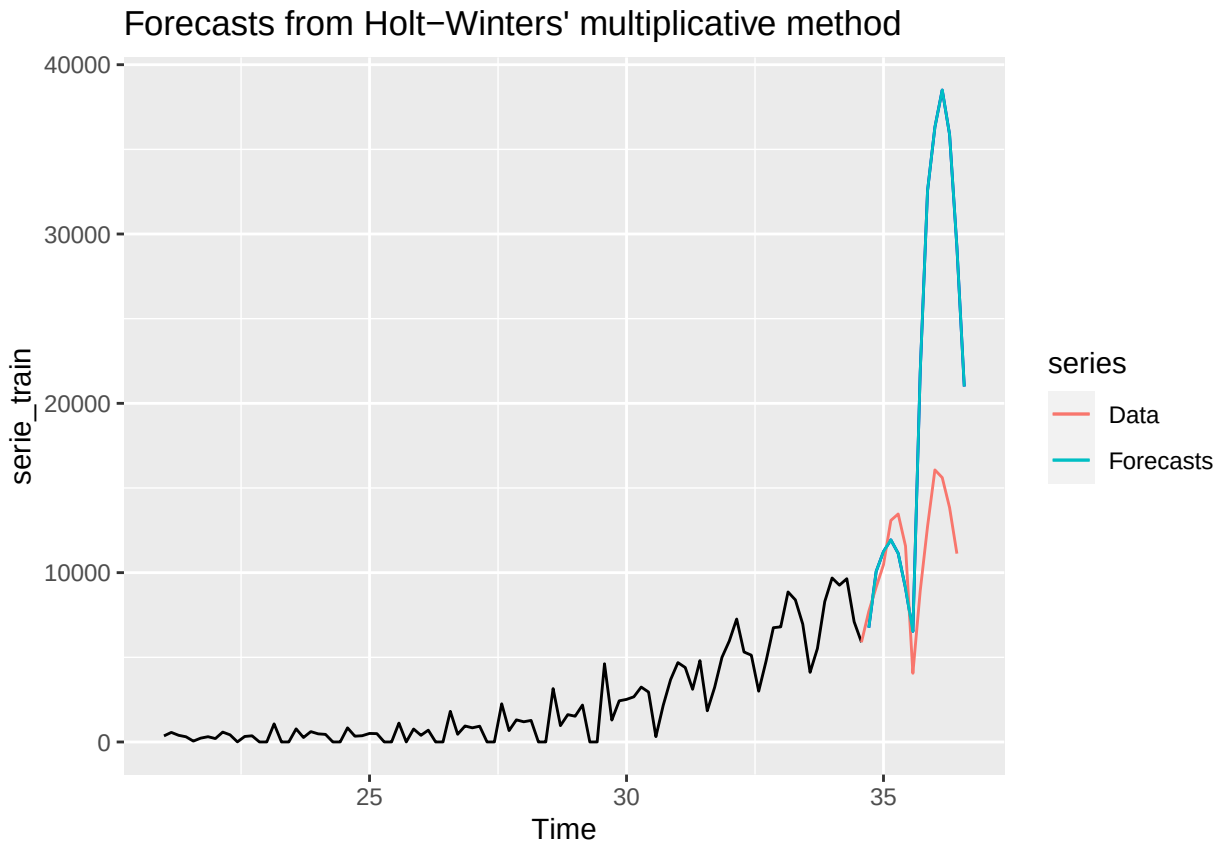
The prevision does not take into account the fact that the amplitude of the periodic pattern increase. We can try a multiplicative exponential smoothing model

```
serie_train[serie_train<1]=0.1
fit2=hw(serie_train,seasonal='multiplicative',h=14)
autoplot(fit2)+ autolayer(tail(covid19_F_2_nc,14), series="Data")+
  autolayer(fit2$mean, series="Forecasts")
```



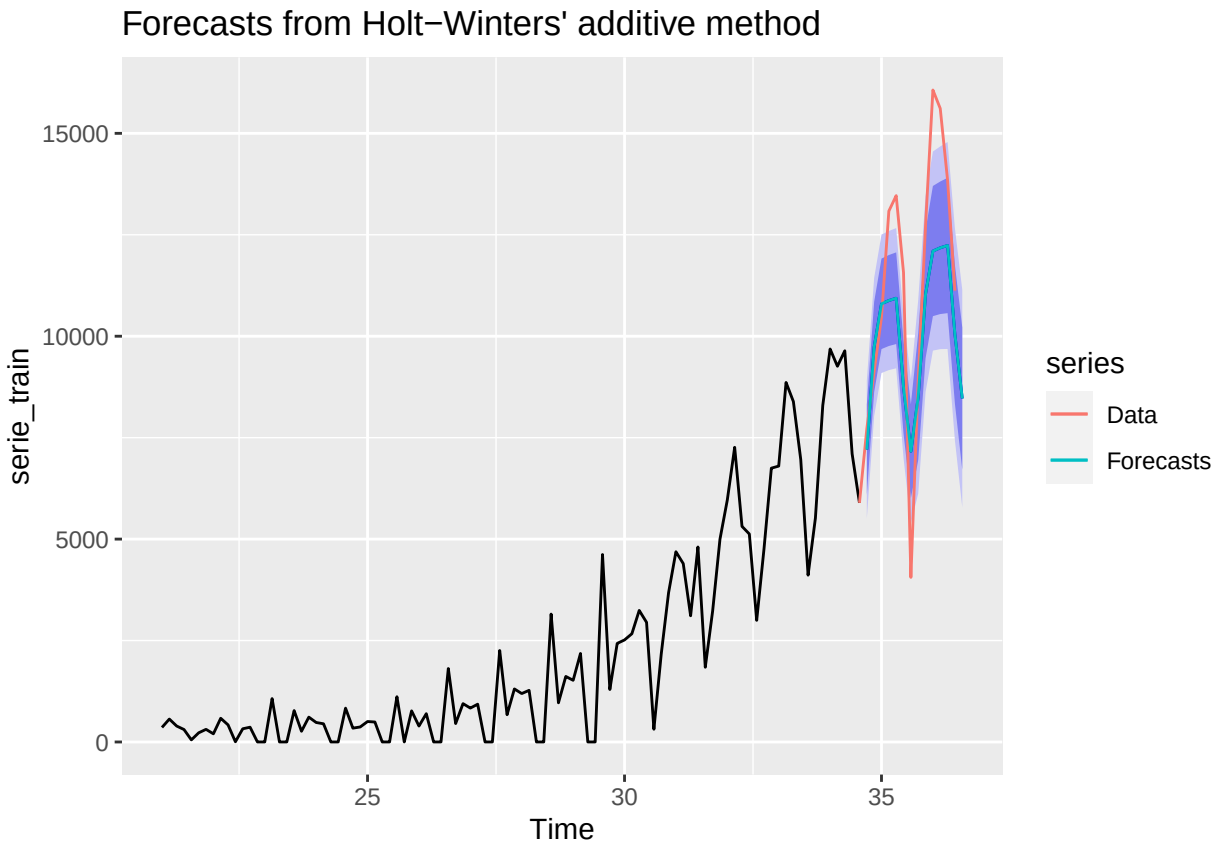
we do not see nothing since the prediction interval is huge. We will plot the forecast without it:

```
autoplot(fit2,PI = F)+ autolayer(tail(covid19_F_2_nc,14), series="Data")+
  autolayer(fit2$mean, series="Forecasts")
```



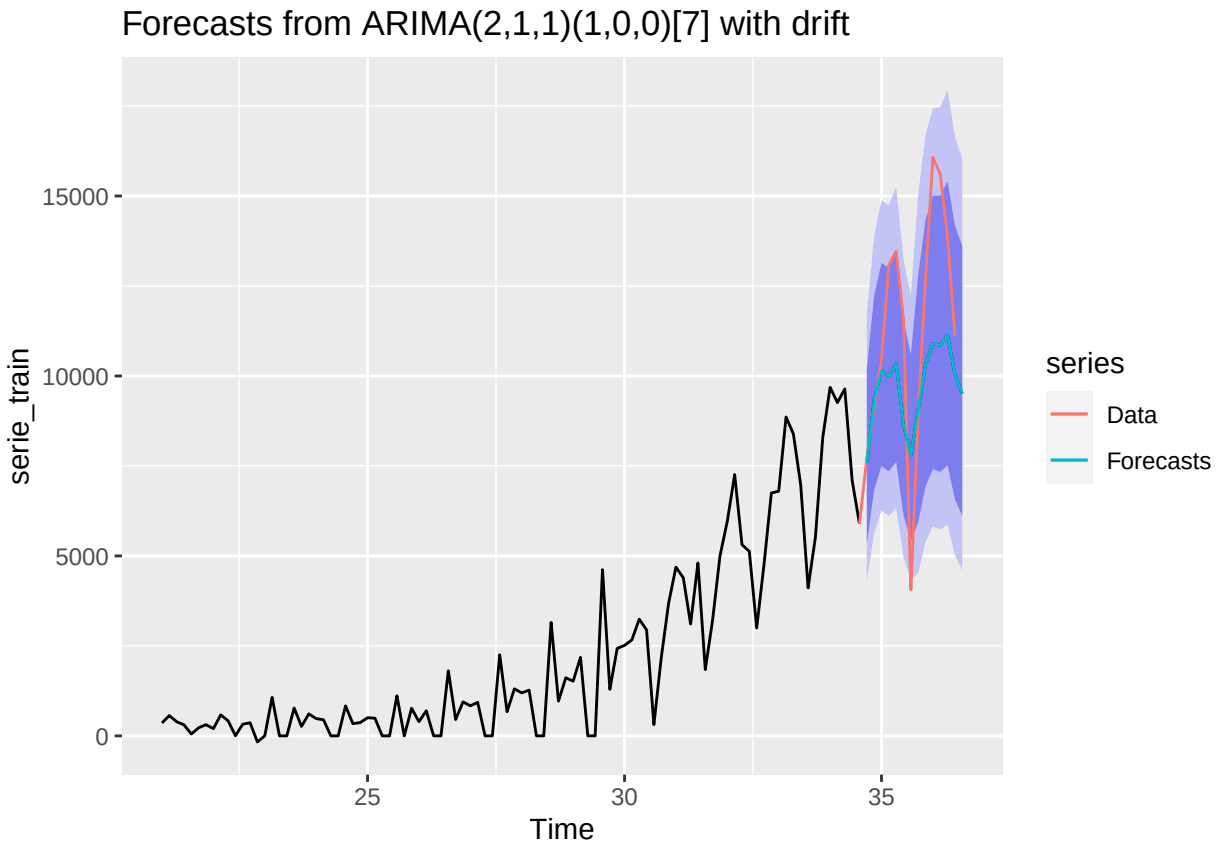
and compare it to an additive :

```
fit3=hw(serie_train,seasonal='additive',h=14)
autoplot(fit3)+ autolayer(tail(covid19_F_2_nc,14), series="Data")+
  autolayer(fit3$mean, series="Forecasts")
```



Finally, we can try to improve our SARIMA model by using the Box.Cox transformation

```
serie_train=window(covid19_F_2_nc,start=c(21,1),end=c(34, 5))
serie=window(covid19_F_2_nc,start=c(21,1))
fit=auto.arima(serie_train,lambda="auto")
prev=forecast(fit,h=14)
autoplot(prev) + autolayer(tail(covid19_F_2_nc,14), series="Data")+
  autolayer(prev$mean, series="Forecasts")
```



The forecast would have been better with such model, but our analyse has been influenced by observing the true future...