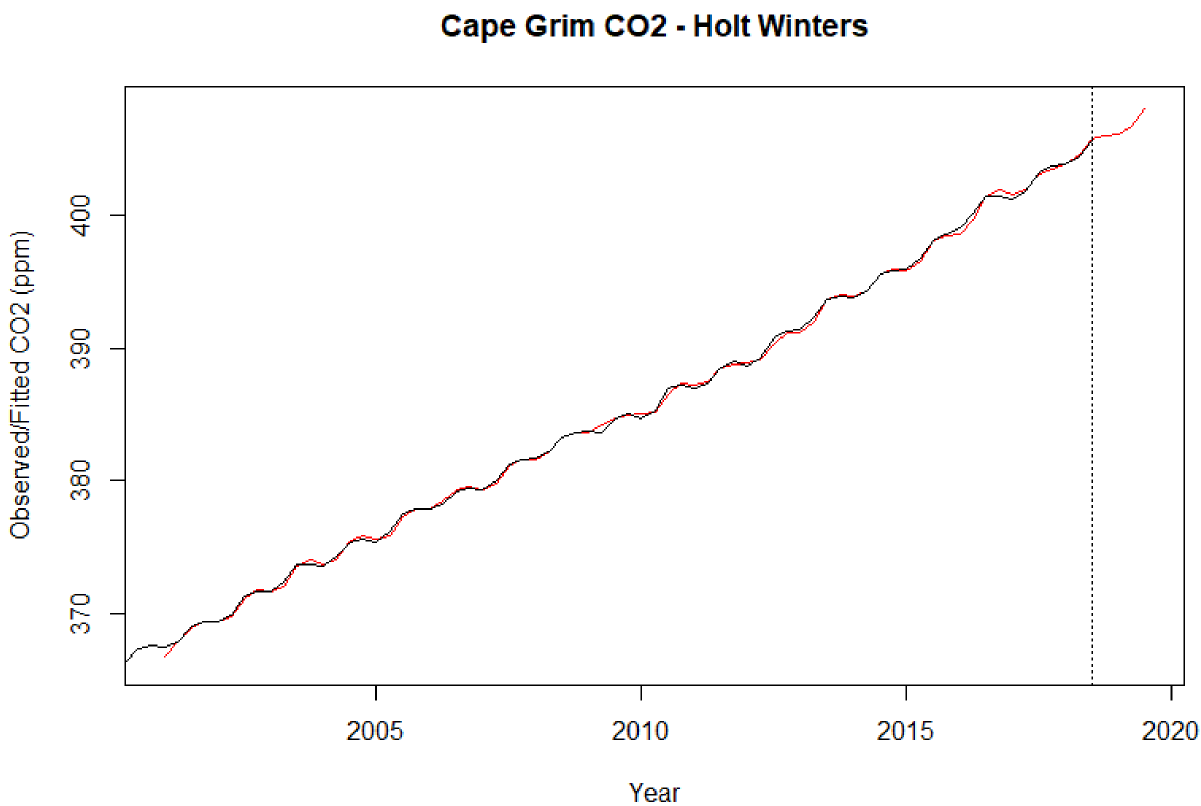
326 Assignment 2

Hasnain Cheena

30/03/2020

Question 1

```
HW.fit = HoltWinters(CO2.fit.ts)
HW.fit

## Holt-Winters exponential smoothing with trend and additive seasonal compon
ent.
##
## Call:
## HoltWinters(x = CO2.fit.ts)
##
## Smoothing parameters:
##  alpha: 0.9267355
##  beta : 0.0813906
##  gamma: 1
##
## Coefficients:
##           [,1]
## a  405.1011655
## b    0.5889777
## s1   0.2311236
## s2  -0.2791367
## s3  -0.2150001
## s4   0.4588345

HW.pred = predict(HW.fit,n.ahead=4)
HW.pred

##          Qtr1     Qtr2     Qtr3     Qtr4
## 2018                            405.9213
## 2019 406.0000 406.6531 407.9159

HW.RMSEP = sqrt(1/4*sum((CO2.pred.ts-HW.pred)^2))
HW.RMSEP

## [1] 0.2214015

plot(HW.fit, HW.pred, main="Cape Grim CO2 - Holt Winters", xlab="Year", ylab=
"Observed/Fitted CO2 (ppm)")
```

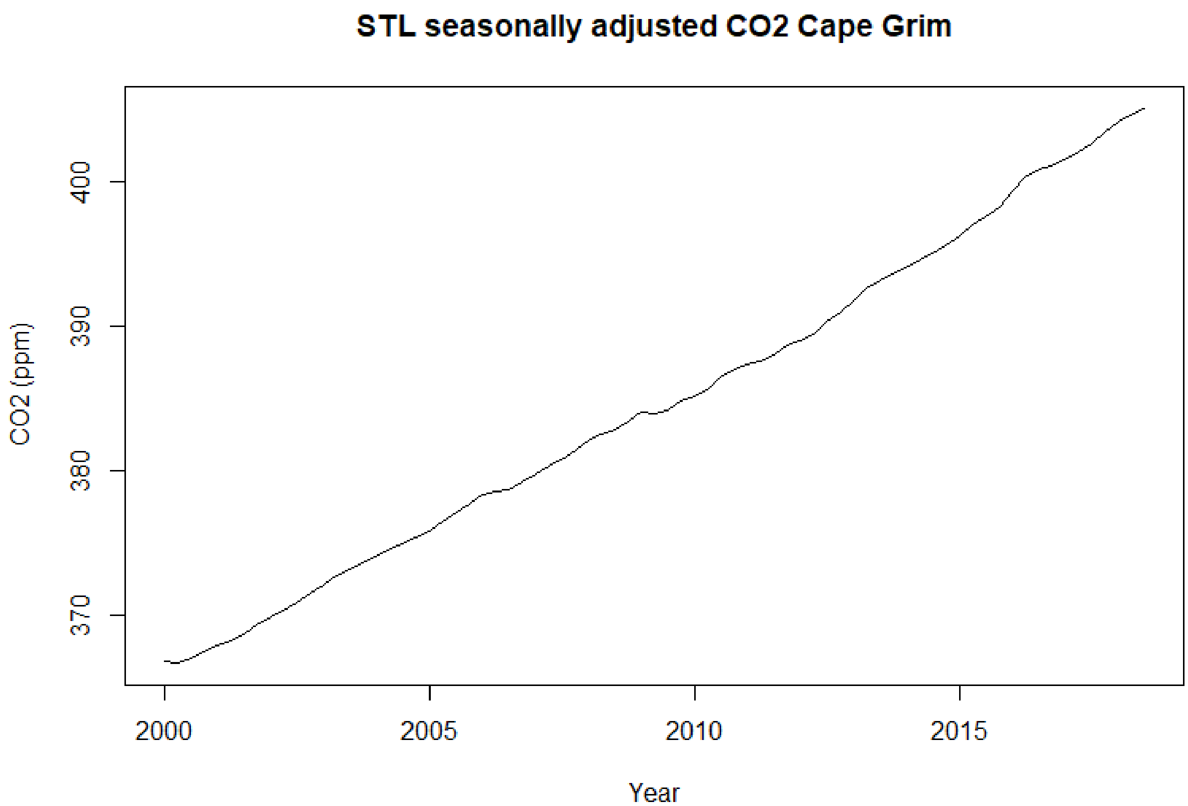


The plot of the Holt-Winters model shows the model is a good fit. This is because the model (red line) is very close to the actual observations (black line) with very little white space between the two lines.

The RMSEP indicates that, on average, the prediction error is 0.22 ppm.

Question 2

```
#calculate and extract seasonal estimates
decomp.stl.CO2 = stl(CO2.fit.ts, s.window="periodic")
#de-seasonalize
stl.CO2.ts = CO2.fit.ts - decomp.stl.CO2$time.series[,1]
plot(stl.CO2.ts, main="STL seasonally adjusted CO2 Cape Grim", xlab="Year", y
lab="CO2 (ppm)")
```

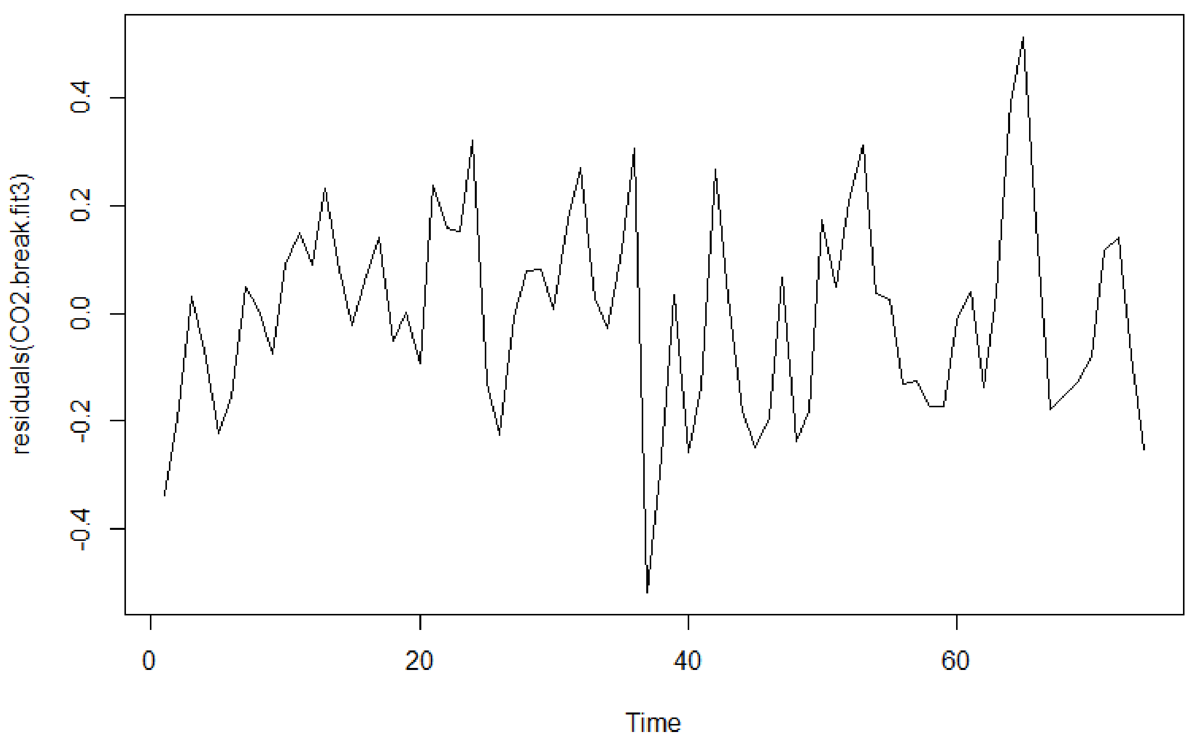


The plot of the seasonal-trend-lowess seasonally adjusted series shows a change in slope around t=50. The slope after the break in trend seems to be slightly steeper than before the break. Both slopes have are reasonably linear with an increasing trend.

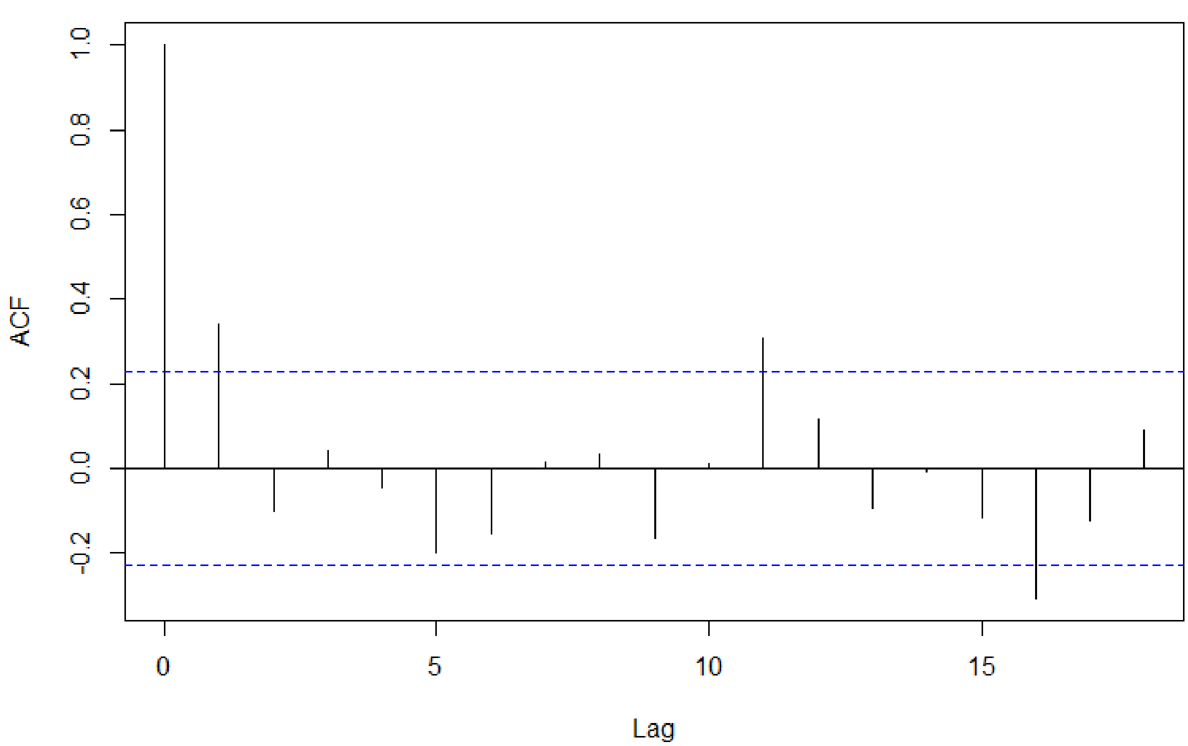
```
#seasonal estimates
decomp.stl.CO2$time.series[1:4,1]

## [1] -0.3906880 -0.3028320  0.4787971  0.2147230
```

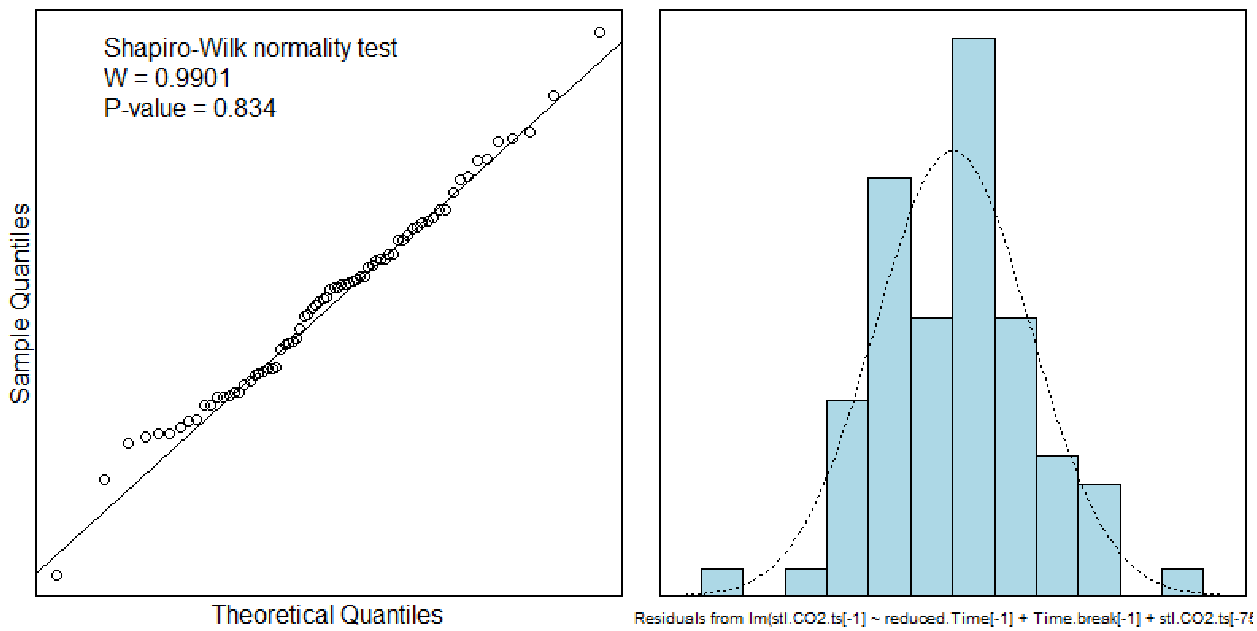
The seasonal estimates for Quarter 1 & Quarter 2 are negative and below the trend. In comparision Quarter 3 & Quarter 4 seasonal estimates are positive and above the trend. Quarter 3 has the highest seasonal estimate (0.48 ppm) while Quarter 1 is the lowest, 0.39 ppm.

Residual Series



ACF Plot of Residual Series






```
#model summary
summary(CO2.break.fit3)

##
## Call:
## lm(formula = stl.CO2.ts[-1] ~ reduced.Time[-1] + Time.break[-1] +
##     stl.CO2.ts[-75])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51687 -0.13691  0.01665  0.11813  0.51237
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)      113.68138   29.12337   3.903 0.000216 ***
## reduced.Time[-1]   0.15131    0.03826   3.955 0.000181 ***
## Time.break[-1]     0.04304    0.01189   3.620 0.000554 ***
## stl.CO2.ts[-75]    0.68994    0.07972   8.654 1.14e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.189 on 70 degrees of freedom
## Multiple R-squared:  0.9997, Adjusted R-squared:  0.9997
## F-statistic: 8.543e+04 on 3 and 70 DF,  p-value: < 2.2e-16
```

```r
#forecast 2018 Q4
t76.sa.pred = CO2.break.fit3$coefficients[1] + CO2.break.fit3$coefficients[2]
* 76 + CO2.break.fit3$coefficients[3] * 26 + CO2.break.fit3$coefficients[4]
* stl.CO2.ts[75]
t76.pred = t76.sa.pred + stl.seasonal.estimates[4,1]

#forecast 2019 Q1
t77.sa.pred = CO2.break.fit3$coefficients[1] + CO2.break.fit3$coefficients[2]
* 77 + CO2.break.fit3$coefficients[3] * 27 + CO2.break.fit3$coefficients[4]
* t76.sa.pred
t77.pred = t77.sa.pred + stl.seasonal.estimates[1,1]

#forecast 2019 Q2
t78.sa.pred = CO2.break.fit3$coefficients[1] + CO2.break.fit3$coefficients[2]
* 78 + CO2.break.fit3$coefficients[3] * 28 + CO2.break.fit3$coefficients[4]
* t77.sa.pred
t78.pred = t78.sa.pred + stl.seasonal.estimates[2,1]

#forecast 2019 Q3
t79.sa.pred = CO2.break.fit3$coefficients[1] + CO2.break.fit3$coefficients[2]
* 79 + CO2.break.fit3$coefficients[3] * 29 + CO2.break.fit3$coefficients[4]
* t78.sa.pred
t79.pred = t79.sa.pred + stl.seasonal.estimates[3,1]

results.df

##      Time Seasonally.Adjusted Predictions
## 1 2018 Q4            405.7808    405.9955
## 2 2019 Q1            406.4578    406.0671
## 3 2019 Q2            407.1193    406.8164
## 4 2019 Q3            407.7700    408.2488

#calculte RMSEP
stl.pred.ts = ts(results.df$Predictions, start=c(2018,4), frequency = 4)

#RMSEP
STL.RMSEP = sqrt(1/4*sum((CO2.pred.ts-stl.pred.ts)^2))
STL.RMSEP

## [1] 0.1951761
```

Question 3 - Technical Notes

The seasonal estimates show that the CO_2 concentration is below the overall trend for Quarter 1 & Quarter 2 with Quarter 1 being the lowest (-0.39 ppm) and the CO_2 concentration is above the overall trend for Quarter 3 & Quarter 4, with Quarter 3 being the highest (0.48 ppm).

The plot of the seasonally adjusted series shows a change in the slope around t=50. The trend after t=50 is slightly higher than previously.

The final model fitted to the seasonal-trend-lowess seasonally adjusted series included a break-in trend term and a lagged response to account for the autocorrelation detected in the Residual Series.

For the final model the Residual Series appears to be random scatter about 0 with a slight upward trend in the early part of the series. The slight upward trend occurs in the early part of the series and is therefore not a concern. The plot of the autocorrelation function shows significant autocorrelation at lag 1, lag 11 and lag 16. However, this is not a concern as lag 11 and lag 16 are unusual lags to be significant in quarterly data. Furthermore, even though the plot shows significant positive autocorrelation at lag 1, our model has a lagged response variable to account for autocorrelation at lag 1 and it is only slightly significant; hence not a worry. The residuals appear to be normally distributed (min=0.52, max=0.51 and a Shaprio-Wilk test p-value = 0.83). Therefore, the model assumptions are satisfied.

We have strong evidence against the hypothesis that the coefficient associated with the time variable is 0 (p-value = 0.0002). Further, we have strong evidence against the hypothesis that the coefficient associated with the break-in trend time variable is 0 (p-value = 0.0006). Additionally, we have very strong evidence against the hypothesis of no autocorrelation (p-value ≈ 0).

The F-statistic provides extremely strong evidence against the hypothesis that none of the variables are related to the seasonally adjusted CO_2 concentration (p-value ≈ 0). The Multiple R^2 is almost 1 (0.9997) indicating that nearly all the variation in the seasonally adjusted CO_2 concentration is explained by the model.

The Residual Standard Error is 0.19 ppm so the prediction intervals will be reasonably narrow. The model predictions can be relied upon as the assumptions are satisfied. The RMSEP for the predictions from 2018 Q4 to 2019 Q3 was 0.20 ppm which was smaller than that of the Moving Average model which was 0.23 ppm. The predictions for 2018 Q4 to 2019 Q3 are:

2018 Q4: 406.0 ppm
2019 Q1: 406.07 ppm
2019 Q2: 406.82 ppm
2019 Q3: 408.25 ppm

Question 4

```
#rerun the STL model using all avaliable data
CO2.break.fit4 = lm(stl.CO2.ts[-1] ~ Time[-1]+Time.break[-1] + stl.CO2.ts[-79
])

#model summary
summary(CO2.break.fit4)

##
## Call:
## lm(formula = stl.CO2.ts[-1] ~ Time[-1] + Time.break[-1] + stl.CO2.ts[-79])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5294 -0.1417  0.0186  0.1271  0.5095
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     114.64012   28.54054   4.017 0.000140 ***
## Time[-1]          0.15269    0.03754   4.067 0.000118 ***
## Time.break[-1]    0.04254    0.01109   3.837 0.000260 ***
## stl.CO2.ts[-79]   0.68731    0.07813   8.797 4.02e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1896 on 74 degrees of freedom
## Multiple R-squared:  0.9998, Adjusted R-squared:  0.9997
## F-statistic: 1.016e+05 on 3 and 74 DF,  p-value: < 2.2e-16

#forecast 2019 Q4
t80.sa.pred = CO2.break.fit4$coefficients[1] + CO2.break.fit4$coefficients[2]
* 80 + CO2.break.fit4$coefficients[3] * 30 + CO2.break.fit4$coefficients[4]
* stl.CO2.ts[79]
t80.pred = t80.sa.pred + stl.seasonal.estimates[4,1]

#forecast 2020 Q1
t81.sa.pred = CO2.break.fit4$coefficients[1] + CO2.break.fit4$coefficients[2]
* 81 + CO2.break.fit4$coefficients[3] * 31 + CO2.break.fit4$coefficients[4]
* t80.sa.pred
t81.pred = t81.sa.pred + stl.seasonal.estimates[1,1]

#forecast 2020 Q2
t82.sa.pred = CO2.break.fit4$coefficients[1] + CO2.break.fit4$coefficients[2]
* 82 + CO2.break.fit4$coefficients[3] * 32 + CO2.break.fit4$coefficients[4]
* t81.sa.pred
t82.pred = t82.sa.pred + stl.seasonal.estimates[2,1]

#forecast 2020 Q3
t83.sa.pred = CO2.break.fit4$coefficients[1] + CO2.break.fit4$coefficients[2]
```

```
* 83 + CO2.break.fit4$coefficients[3] * 33 + CO2.break.fit4$coefficients[4]
* t82.sa.pred
t83.pred = t83.sa.pred + stl.seasonal.estimates[3,1]

#Predictions
stl.pred.df = data.frame(CO2=c(t80.pred,t81.pred,t82.pred,t83.pred))
stl.pred.ts = ts(stl.pred.df, start=c(2019,4), frequency = 4)
stl.pred.ts
```

```
##          Qtr1     Qtr2     Qtr3     Qtr4
## 2019                            408.5988
## 2020 408.6075 409.3381 410.7559
```

The seasonal-trend-lowess model was the best predicting model out of the three models fitted to the dataset. This is because it has the lowest RMSEP (0.2 ppm) relative to the other models.

The summary shows that the seasonal-trend-lowess model parameter estimates for the full timeframe model ($CO2.break.fit4$) are similar to those of the reduced timeframe model ($CO2.break.fit3$).

As the model assumptions were satisfied, the predictions should be reasonably reliable.