STATS 326
Applied Time Series
ASSIGNMENT THREE
ANSWER GUIDE

Question One:

```
> red.Quarter = factor(c(rep(1:4,18),(1:3)))

> SF.fit1 = lm(red.CO2.ts[-1]~red.Time[-1]+red.Time.break[-1]+
   red.Quarter[-1]+red.CO2.ts[-75])

> summary(SF.fit1)

Call:
lm(formula = red.CO2.ts[-1] ~ red.Time[-1] + red.Time.break[-1] +
    red.Quarter[-1] + red.CO2.ts[-75])

Residuals:
     Min       1Q   Median       3Q      Max
-0.48990 -0.12209 -0.00581  0.11306  0.53995

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        108.82449   29.39775   3.702 0.000435 ***
red.Time[-1]         0.14548    0.03860   3.769 0.000349 ***
red.Time.break[-1]   0.04182    0.01196   3.496 0.000843 ***
red.Quarter[-1]2     0.43876    0.07593   5.778 2.14e-07 ***
red.Quarter[-1]3     1.14763    0.07477  15.348  < 2e-16 ***
red.Quarter[-1]4     0.41510    0.06543   6.344 2.22e-08 ***
red.CO2.ts[-75]      0.70187    0.08043   8.727 1.18e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1889 on 67 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 4.282e+04 on 6 and 67 DF,  p-value: < 2.2e-16
```

```
> t.76.pred = SF.fit1$coef[1]+SF.fit1$coef[2]*76+
  SF.fit1$coef[3]*26+SF.fit1$coef[6]+SF.fit1$coef[7]*red.CO2.ts[75]
> t.76.pred
(Intercept)
   406.0347

> t.77.pred = SF.fit1$coef[1]+SF.fit1$coef[2]*77+
  SF.fit1$coef[3]*27+SF.fit1$coef[7]*t.76.pred
> t.77.pred
(Intercept)
   406.1401

> t.78.pred = SF.fit1$coef[1]+SF.fit1$coef[2]*78+
  SF.fit1$coef[3]*28+SF.fit1$coef[4]+SF.fit1$coef[7]*t.77.pred
> t.78.pred
(Intercept)
   406.8401

> t.79.pred = SF.fit1$coef[1]+SF.fit1$coef[2]*79+
  SF.fit1$coef[3]*29+SF.fit1$coef[5]+SF.fit1$coef[7]*t.78.pred
> t.79.pred
(Intercept)
   408.2276

> SF.pred = c(t.76.pred,t.77.pred,t.78.pred,t.79.pred)
> names(SF.pred) = c("2018.4","2019.1","2019.2","2019.3")
> SF.pred
  2018.4   2019.1   2019.2   2019.3
406.0347 406.1401 406.8401 408.2276

> SF.RMSEP = sqrt(1/4*sum((actual-SF.pred)^2))
> SF.RMSEP
[1] 0.2384888
```

The Seasonal Factor model included a trend term, a trend break term, a seasonal factor and a lagged response to take care of autocorerelation detected in the Residual Series. The Residual Series showed reasonably constant scatter about 0 with a slight increasing trend for the first 2 – 3 years. The Residual Series had a larger negative residual for time period 38 (2009.2) and a larger positive residual for time period 66 (2016.2). The plot of the autocorrelation function of the Residual Series showed significant lags at 1, 11 and 16 but they were small enough to ignore. The residuals appeared to follow a normal distribution (Shapiro-Wilk P-value = 0.852) although the residuals were slightly right skewed due to the large positive residual discussed above.

Quarters 2 – 4 CO2 concentrations were all higher than the omitted baseline level (Quarter 1) with Quarter 3 being the highest (1.15 ppm).

The RMSEP was 0.24 ppm.

Question Two:

```
> FH.fit1 = lm(red.CO2.ts[-1]~red.Time[-1]+red.Time.break[-1]+c1[-1]+
  s1[-1]+c2[-1]+red.CO2.ts[-75])

> summary(FH.fit1)

Call:
lm(formula = red.CO2.ts[-1] ~ red.Time[-1] + red.Time.break[-1] +
    c1[-1] + s1[-1] + c2[-1] + red.CO2.ts[-75])

Residuals:
     Min       1Q   Median       3Q      Max
-0.48990 -0.12209 -0.00581  0.11306  0.53995

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        109.32486   29.38088   3.721 0.000408 ***
red.Time[-1]         0.14548    0.03860   3.769 0.000349 ***
red.Time.break[-1]   0.04182    0.01196   3.496 0.000843 ***
c1[-1]              -0.01183    0.04370  -0.271 0.787404
s1[-1]              -0.57381    0.03739 -15.348 < 2e-16 ***
c2[-1]              -0.07344    0.02232  -3.290 0.001597 **
red.CO2.ts[-75]      0.70187    0.08043   8.727 1.18e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1889 on 67 degrees of freedom
Multiple R-squared:  0.9997,    Adjusted R-squared:  0.9997
F-statistic: 4.282e+04 on 6 and 67 DF,  p-value: < 2.2e-16
```

```
> t.76.pred = FH.fit1$coef[1]+FH.fit1$coef[2]*76+FH.fit1$coef[3]*26+
  FH.fit1$coef[4]*cos(2*pi*76*(1/4))+ FH.fit1$coef[5]*sin(2*pi*76*(1/4))+
  FH.fit1$coef[6]*cos(2*pi*76*(2/4))+ FH.fit1$coef[7]*red.CO2.ts[75]
> t.76.pred
(Intercept)
   406.0347

> t.77.pred = FH.fit1$coef[1]+FH.fit1$coef[2]*77+FH.fit1$coef[3]*27+
  FH.fit1$coef[4]*cos(2*pi*77*(1/4))+ FH.fit1$coef[5]*sin(2*pi*77*(1/4))+
  FH.fit1$coef[6]*cos(2*pi*77*(2/4))+ FH.fit1$coef[7]*t.76.pred
> t.77.pred
(Intercept)
   406.1401

> t.78.pred = FH.fit1$coef[1]+FH.fit1$coef[2]*78+FH.fit1$coef[3]*28+
  FH.fit1$coef[4]*cos(2*pi*78*(1/4))+ FH.fit1$coef[5]*sin(2*pi*78*(1/4))+
  FH.fit1$coef[6]*cos(2*pi*78*(2/4))+ FH.fit1$coef[7]*t.77.pred
> t.78.pred
(Intercept)
   406.8401

> t.79.pred = FH.fit1$coef[1]+FH.fit1$coef[2]*79+FH.fit1$coef[3]*29+
  FH.fit1$coef[4]*cos(2*pi*79*(1/4))+ FH.fit1$coef[5]*sin(2*pi*79*(1/4))+
  FH.fit1$coef[6]*cos(2*pi*79*(2/4))+ FH.fit1$coef[7]*t.78.pred
> t.79.pred
(Intercept)
   408.2276

> FH.pred = c(t.76.pred,t.77.pred,t.78.pred,t.79.pred)
> names(FH.pred) = c("2018.4","2019.1","2019.2","2019.3")
> FH.pred
  2018.4   2019.1   2019.2   2019.3
406.0347 406.1401 406.8401 408.2276

> FH.RMSEP = sqrt(1/4*sum((actual-FH.pred)^2))
> FH.RMSEP
[1] 0.2384888
```

The Full Harmonic model produced the same results as the Seasonal Factor model, as was expected. It had the smallest RMSEP (0.24 ppm).

The Full Harmonic model included a trend term, a trend break term, 2 significant harmonics (P-values ≈ 0 and 0.001597) and a lagged response variable to take care of the autocorrelation detected in the Residual Series.

The diagnostic plots are the same as for the Seasonal Factor model.

The Reduced Harmonic model deleting the cosine term with frequency ¼ gave RMSEP of 0.24 ppm (0.2438801) and the Cosine Harmonic model, 0.26 ppm but these values were slightly higher than the Full Harmonic, 0.24 ppm (0.2384888).

Question Three: (Can use either Seasonal Factor model or Full Harmonic model)

The Seasonal Factor model included a trend term, a trend break term, a seasonal factor and a lagged response variable to take care of the autocorrelation detected in the Residual Series.

The Residual Series showed reasonably constant scatter about 0 with a slight increasing trend for the first 2 – 3 years. The Residual Series had a larger negative residual for time period 38 (2009.2) and a larger positive residual for time period 66 (2016.2). The plot of the autocorrelation function of the Residual Series shows significant lags at 1, 11 and 16, but they are small enough to ignore. The residuals appear to follow a normal distribution (Shapiro-Wilk *P-value* = 0.852). The residuals are slightly right skewed due to the large residual discussed above.

We have strong evidence against the hypothesis that the coefficient associated with the Time variable is 0 (*P-value* = 0.000349) and strong evidence that the coefficient associated with the Time.break variable is 0 (*P-value* = 0.000843). We have very strong evidence that Quarters 2 to 4 are above the omitted baseline level (*P-values* ≈ 0) with Quarter 3 being the highest. 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 CO2 concentration (*P-value* ≈ 0). The Multiple R^2 is almost 1 (0.9997) indicating that nearly all the variation in the seasonally adjusted CO2 concentration is explained by the model.

The Residual Standard Error is 0.19 ppm so prediction intervals should be narrow. The model predictions can be relied on as the assumptions appear to be satisfied.

The RMSEP for the predictions was 0.24 ppm which was the same as for the Full Harmonic model.

Our predictions for 2018.4 and 2019.1-3 were:

Quarter 4: 406.03 ppm
Quarter 1: 406.14 ppm
Quarter 2: 406.84 ppm
Quarter 3: 408.23 ppm

Question Four:

```
> full.Quarter = factor(c(rep(1:4,19),(1:3)))

> SF.fit = lm(full.CO2.ts[-1]~full.Time[-1]+full.Time.break[-1]+
   full.Quarter[-1]+full.CO2.ts[-79])
> summary(SF.fit)

Call:
lm(formula = full.CO2.ts[-1] ~ full.Time[-1] + full.Time.break[-1] +
    full.Quarter[-1] + full.CO2.ts[-79])

Residuals:
     Min       1Q   Median       3Q      Max
-0.50285 -0.12867 -0.00066  0.12102  0.53787

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         108.36491   28.78665   3.764 0.000341 ***
full.Time[-1]         0.14500    0.03785   3.831 0.000273 ***
full.Time.break[-1]   0.04082    0.01114   3.665 0.000474 ***
full.Quarter[-1]2     0.46150    0.07438   6.205 3.25e-08 ***
full.Quarter[-1]3     1.16834    0.07242  16.132  < 2e-16 ***
full.Quarter[-1]4     0.41786    0.06380   6.549 7.79e-09 ***
full.CO2.ts[-79]      0.70309    0.07876   8.927 3.20e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1888 on 71 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9997
F-statistic: 5.127e+04 on 6 and 71 DF,  p-value: < 2.2e-16

> t.80.pred = SF.fit$coef[1]+SF.fit$coef[2]*80+
SF.fit$coef[3]*30+SF.fit$coef[6]+SF.fit$coef[7]*full.CO2.ts[79]
> t.80.pred
(Intercept)
   408.6447

> t.81.pred = SF.fit$coef[1]+SF.fit$coef[2]*81+
   SF.fit$coef[3]*31+SF.fit$coef[7]*t.80.pred
> t.81.pred
(Intercept)
   408.6901

> t.82.pred = SF.fit$coef[1]+SF.fit$coef[2]*82+
   SF.fit$coef[3]*32+SF.fit$coef[4]+SF.fit$coef[7]*t.81.pred
> t.82.pred
(Intercept)
   409.3694

> t.83.pred = SF.fit$coef[1]+SF.fit$coef[2]*83+
   SF.fit$coef[3]*33+SF.fit$coef[5]+SF.fit$coef[7]*t.82.pred
> t.83.pred
(Intercept)
   410.7396

> SF.Full.pred = c(t.80.pred,t.81.pred,t.82.pred,t.83.pred)
> names(SF.Full.pred) = c("2019.4","2020.1","2020.2","2020.3")
> SF.Full.pred
  2019.4   2020.1   2020.2   2020.3
408.6447 408.6901 409.3694 410.7396
```

The model including all the data to 2019.3 has very similar estimates to our reduced time frame model. The Residual Standard Error is (0.19 ppm) so the prediction intervals should be reasonably narrow. Our predictions should be reliable.

Question Five:

The best predicting model is the Seasonal Trend Lowess Seasonally Adjusted model with an RMSEP of 0.20 ppm.