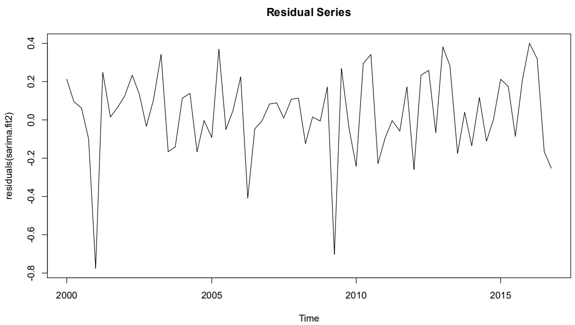
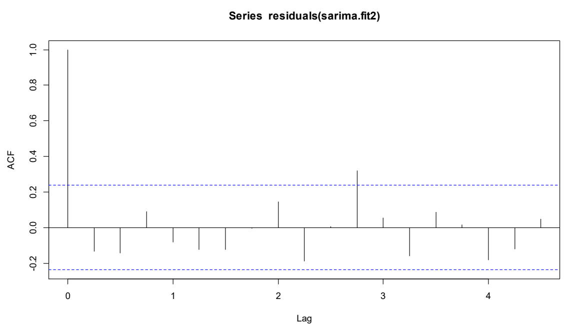
STATS 326
Applied Time Series
ASSIGNMENT FIVE
ANSWER GUIDE

Question One:

```
> sarima.fit2 = arima(red.CO2.ts,order=c(0,1,1),
  seasonal=list(order=c(0,1,1),period=4))
> sarima.fit2

Call:
arima(x = red.CO2.ts, order = c(0, 1, 1), seasonal = list(order =
c(0, 1, 1), period = 4))

Coefficients:
         ma1     sma1
      0.5554  -0.8288
s.e.  0.1313   0.1778

sigma^2 estimated as 0.04362:  log likelihood = 6.68,  aic = -7.37

> plot.ts(residuals(sarima.fit2),main="Residual Series")
```




```
> acf(residuals(sarima.fit2))
```



The Residual Series shows reasonably random scatter about 0, although there are 2 large negative residuals (2001.1 and 2009.2). The autocorrelation function plot of the residuals shows a significant autocorrelation at lag 11 but it is not significant in the pacf (once intermediate correlations are allowed for). The assumptions appear to be satisfied.

```
> SARIMA2.pred = predict(sarima.fit2,n.ahead=4)
> SARIMA2.pred
$`pred`
         Qtr1     Qtr2     Qtr3     Qtr4
2017 401.2802 401.9604 403.2615 403.5376

$se
          Qtr1      Qtr2      Qtr3      Qtr4
2017 0.2089430 0.3863309 0.5048227 0.6003668

> actual = CO2.ts[69:72]
> actual
[1] 401.19 401.77 403.15 403.69

> RMSEP.SARIMA2 = sqrt(1/4*sum((actual-SARIMA2.pred$pred)^2))
> RMSEP.SARIMA2
[1] 0.141468
```

The SARIMA$(0,1,1)\times(0,1,1)_4$ model had the lowest RMSEP (0.14 ppm) of all the SARIMA models tried. The best predicting model from previous assignments was the Cosine Harmonic model with RMSEP of 0.31 ppm.

Since the RMSEP (0.14) is for the SARIMA$(0,1,1)\times(0,1,1)_4$ model is the smallest (by a significant margin) of all the models tried it is the overall best predicting model of all the models tried.

Question Two:

```
> sarima.fit2F = arima(CO2.ts,order=c(0,1,1),
  seasonal=list(order=c(0,1,1),period=4))
> sarima.fit2F

Call:
arima(x = CO2.ts, order = c(0, 1, 1), seasonal = list(order = c(0, 1,
1), period = 4))

Coefficients:
         ma1     sma1
      0.5697   -0.888
s.e.  0.1183    0.140

sigma^2 estimated as 0.04088:  log likelihood = 8.67,  aic = -11.33
```

$SARIMA(0,1,1)\times(0,1,1)_4$

$(1 - B)(1 - B^4)y_t = (1 + \alpha B)(1 + AB^4)\varepsilon_t$
$(1 - B - B^4 + B^5)y_t = (1 + \alpha B + AB^4 + \alpha AB^5)\varepsilon_t$
$y_t - y_{t-1} - y_{t-4} + y_{t-5} = \varepsilon_t + \alpha\varepsilon_{t-1} + A\varepsilon_{t-4} + \alpha A\varepsilon_{t-5}$
$y_t = y_{t-1} + y_{t-4} - y_{t-5} + \varepsilon_t + \alpha\varepsilon_{t-1} + A\varepsilon_{t-4} + \alpha A\varepsilon_{t-5}$
$y_t = y_{t-1} + y_{t-4} - y_{t-5} + \varepsilon_t + 0.5697\varepsilon_{t-1} - 0.888\varepsilon_{t-4} - 0.5058936\varepsilon_{t-5}$

```
> CO2.ts
       Qtr1   Qtr2   Qtr3   Qtr4
2000 366.39 366.40 367.42 367.70
......
2016 398.97 400.10 401.41 401.40
2017 401.19 401.77 403.15 403.69

> residuals(sarima.fit2F)
             Qtr1         Qtr2         Qtr3         Qtr4
2000  0.211535282  0.094608972  0.062797936 -0.097113207
......
2016  0.408694027  0.343373905 -0.166843061 -0.237515542
2017 -0.056352710 -0.007455011  0.099525388  0.198209623
```

$y_{t+1} = y_t + y_{t-3} - y_{t-4} + \varepsilon_{t+1} + 0.5697\varepsilon_t - 0.888\varepsilon_{t-3} - 0.5058936\varepsilon_{t-4}$
$= 403.69 + 401.19 - 401.40 + 0 + 0.5697(0.198209623) -$
$0.888(-0.056352710) - 0.5058936(-0.237515542)$
$= 403.7631$

$y_{t+2} = y_{t+1} + y_{t-2} - y_{t-3} + \varepsilon_{t+2} + 0.5697\varepsilon_{t+1} - 0.888\varepsilon_{t-2} - 0.5058936\varepsilon_{t-3}$
$= 403.7631 + 401.77 - 401.19 + 0 + 0.5697(0) -$
$0.888(-0.007455011) - 0.5058936(-0.056352710)$
$= 404.3782$

$y_{t+3} = y_{t+2} + y_{t-1} - y_{t-2} + \varepsilon_{t+3} + 0.5697\varepsilon_{t+2} - 0.888\varepsilon_{t-1} - 0.5058936\varepsilon_{t-2}$
$= 404.3782 + 403.15 - 401.77 + 0 + 0.5697(0) -$
$0.888(0.099525388) - 0.5058936(-0.007455011)$
$= 405.6736$

$y_{t+4} = y_{t+3} + y_t - y_{t-1} + \varepsilon_{t+4} + 0.5697\varepsilon_{t+3} - 0.888\varepsilon_t - 0.5058936\varepsilon_{t-1}$
$= 405.6736 + 403.69 - 403.15 + 0 + 0.5697(0) -$
$0.888(0.198209623) - 0.5058936(0.099525388)$
$= 405.9872$

Question Three:

My brief was to predict the atmospheric concentration of carbon dioxide at Cape Grim in Tasmania, Australia (in parts per million) for 2018.

We need to be a little careful with our predictions and their reliability as we have a Time Series with only 72 observations. However the model that is used is a good model so the predictions should be reliable.

I built several different models using the first seventeen years of data (2000 – 2016) and used each model to predict the 4 quarters of 2017. Each model's predictions were compared with the actual 2017 values to find the model that produced the most accurate predictions.

Once the best predicting model was found, it was re-run with all the data (2000 – 2017) and predictions done for 2018, as requested.

I predict the carbon dioxide concentration in the atmosphere for 2018 at Cape Grim in Tasmania, Australia will be:

403.76 ppm for Quarter 1, 2018
404.38 ppm for Quarter 2, 2018
405.67 ppm for Quarter 3, 2018
405.99 ppm for Quarter 4, 2018