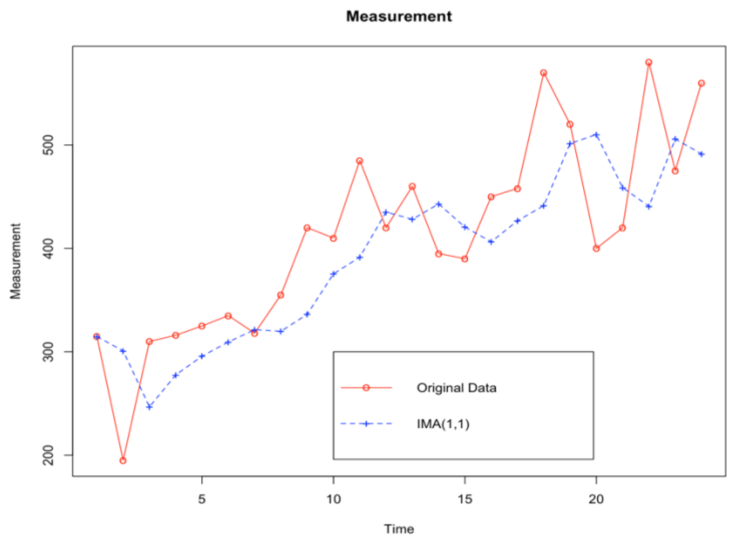
DSC 275/475: Time Series Analysis and Forecasting (Fall 2019)
HW #3
Chunlei Zhou

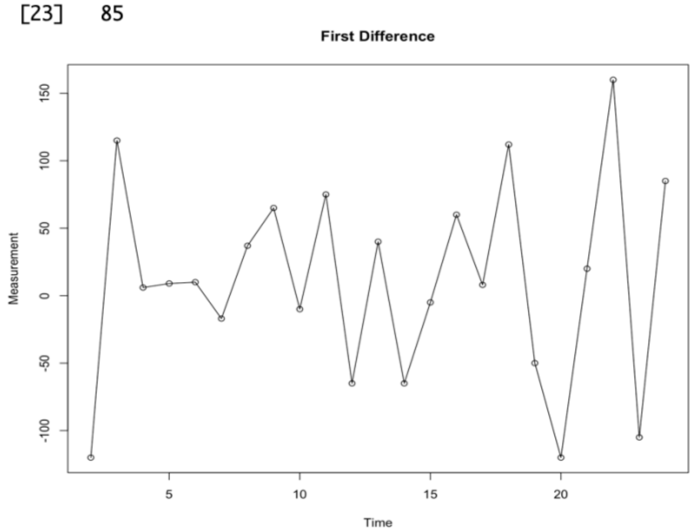
1. **The data in provided in the file Measurement_Q1.xls exhibits a linear trend. Apply the following models to the data.**

   a) **Develop an IMA(1,1) model for the data. Overlay the model output on the original time series.**

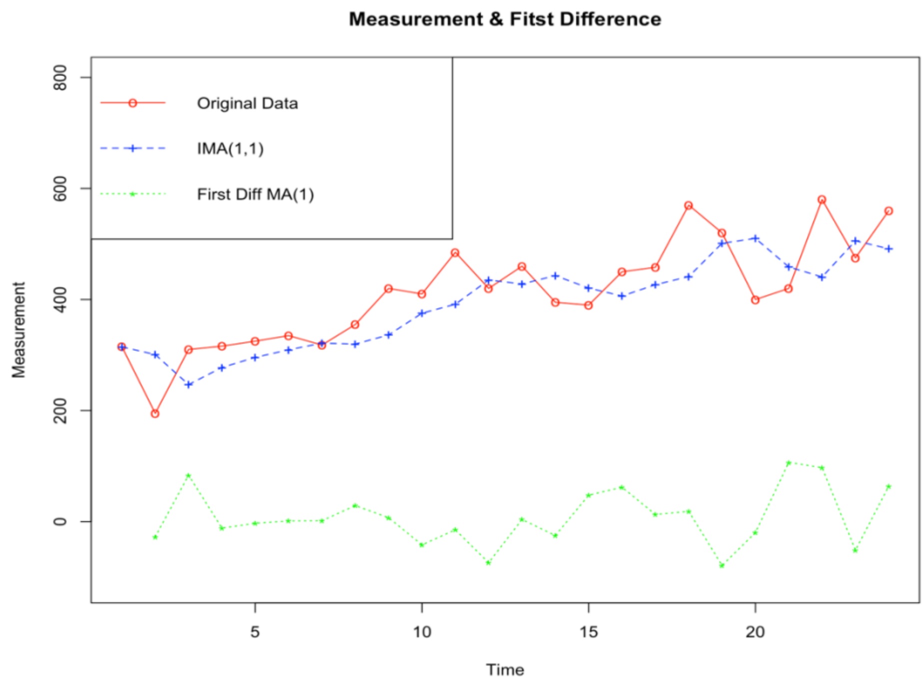


   b) **Compute and plot the first differences of the data.**

```
Time Series:
Start = 2
End = 24
Frequency = 1
 [1] -120  115    6    9   10  -17   37   65  -10   75  -65
[12]   40  -65   -5   60    8  112  -50 -120   20  160 -105
[23]   85
```



c) Develop an MA(1) model on the first difference. Overlay the model output on the first difference computed in (b).



d) Based on the model parameter specifications obtained in (a) and (c), comment on how the two models are related.

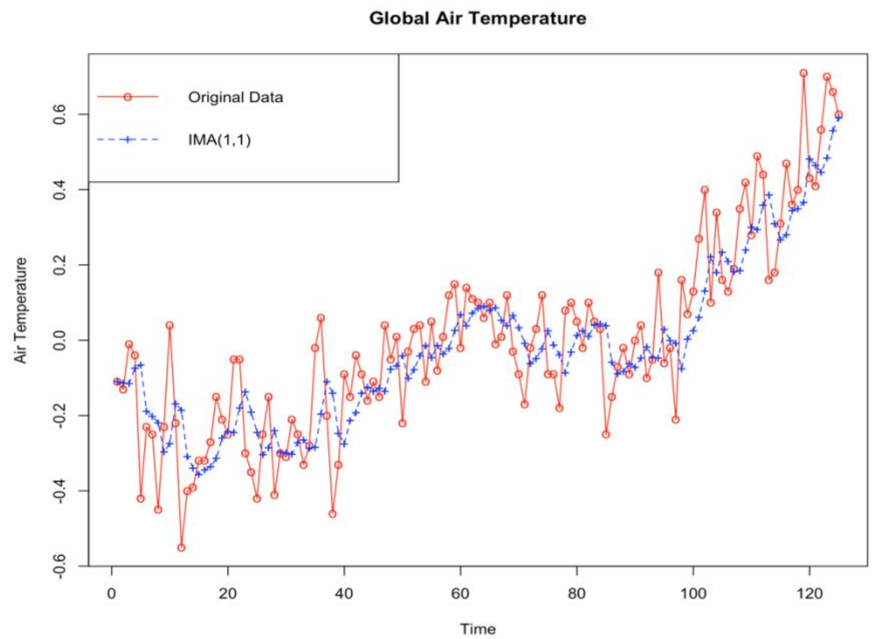
```
> summary(model_1_a)
Series: ts.measurement
ARIMA(0,1,1)

Coefficients:
          ma1
      -0.5332
s.e.   0.1677

sigma^2 estimated as 4517:  log likelihood=-129.07
AIC=262.14   AICc=262.74   BIC=264.41

Training set error measures:
                   ME     RMSE     MAE      MPE     MAPE
Training set 20.14769 64.34419 52.0179 2.941951 13.12147
                  MASE        ACF1
Training set 0.8803619 -0.08236866
```

```
> summary(model_1_c)
Series: first_diff
ARIMA(0,0,1) with non-zero mean

Coefficients:
          ma1     mean
      -1.0000  11.0965
s.e.   0.1392   1.5452

sigma^2 estimated as 3006:  log likelihood=-125.27
AIC=256.55   AICc=257.81   BIC=259.96

Training set error measures:
                   ME     RMSE      MAE     MPE     MAPE
Training set 2.478957 52.38977 41.73513 67.8188 144.9834
                 MASE       ACF1
Training set 0.414525 0.03988509
```

Based on the parameter specifications listed above in model (a) and model (c), we can see that both models have negatively correlated successive observations. Both are moving average model with order equals to 1. Both models are not integrated with autoregressive model. The difference between the two models is that model (a) uses the original dataset and model (b) uses the first difference dataset.

2. **Consider the global mean surface air temperature anomaly data provided in GlobalAirTemperature.xls.**

    a) **Apply an IMA(1,1) model to this data. Overlay the output on the original data. Calculate the SSE by comparing the model output with the data.**



    SSE = 2.229416

    b) **Now, apply an IMA(1,2) model. Calculate the SSE.**

    SSE = 2.103975

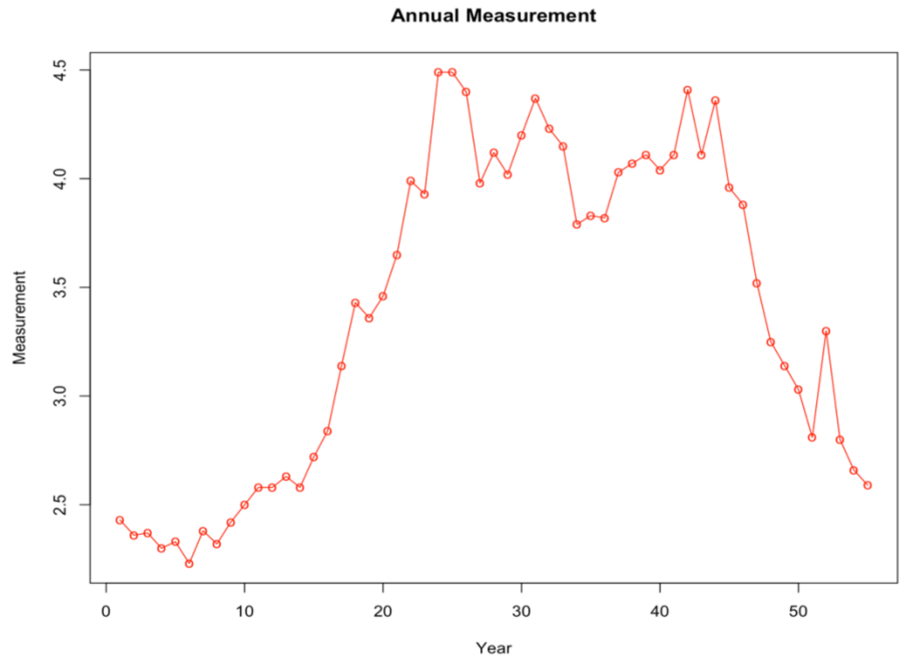
    c) **Comment which model is better suited for this data based on the SSE?**

    Model (b) is better because the SSE of model (b) is smaller than that of model (a).

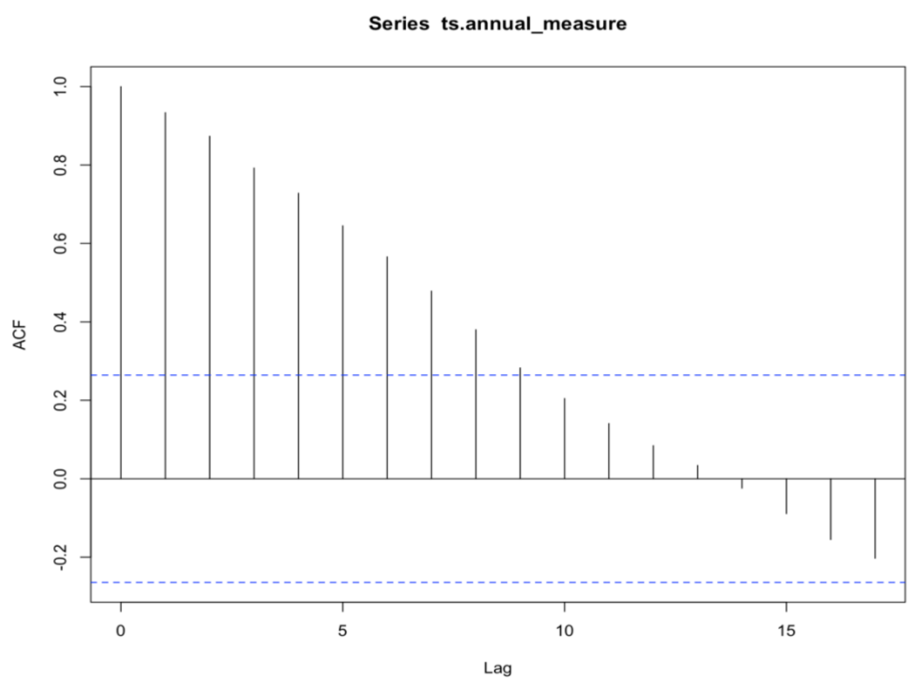
3. **Review the dataset contained in the file Measurement_Q3.xls which contains a measurement recorded annually over close to 50 years.**

   a) **Plot the time series, ACF, and Partial AutoCorrelation Function (PACF).**

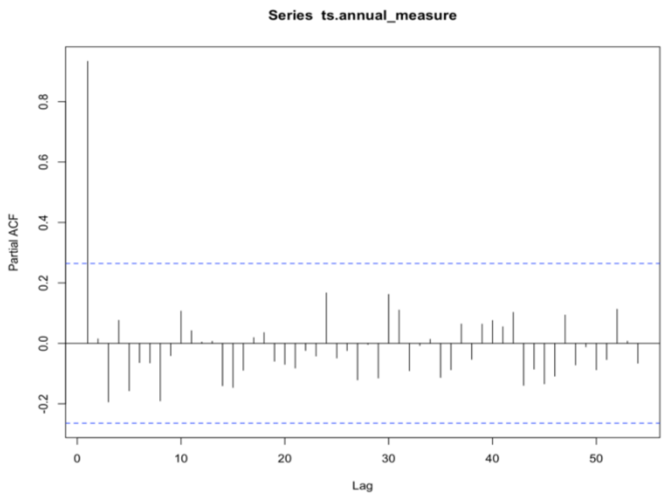
      Time Series:



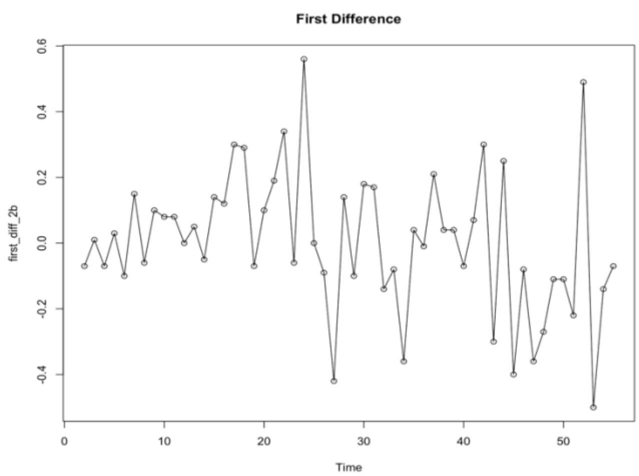
      ACF:



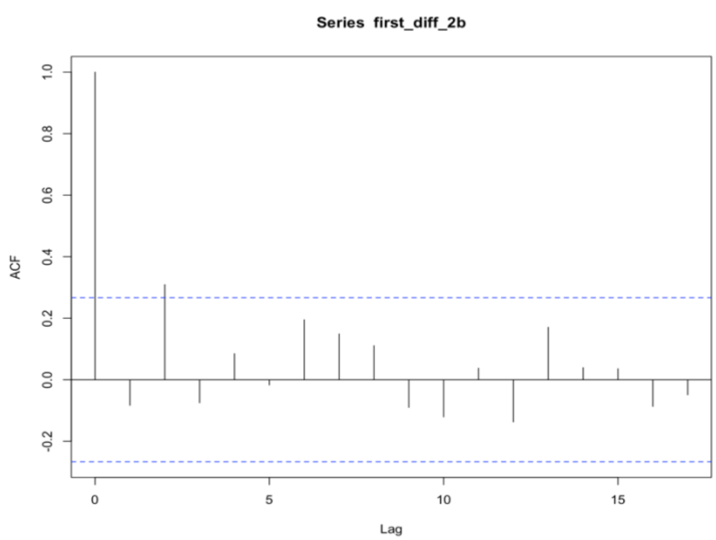
PACF:



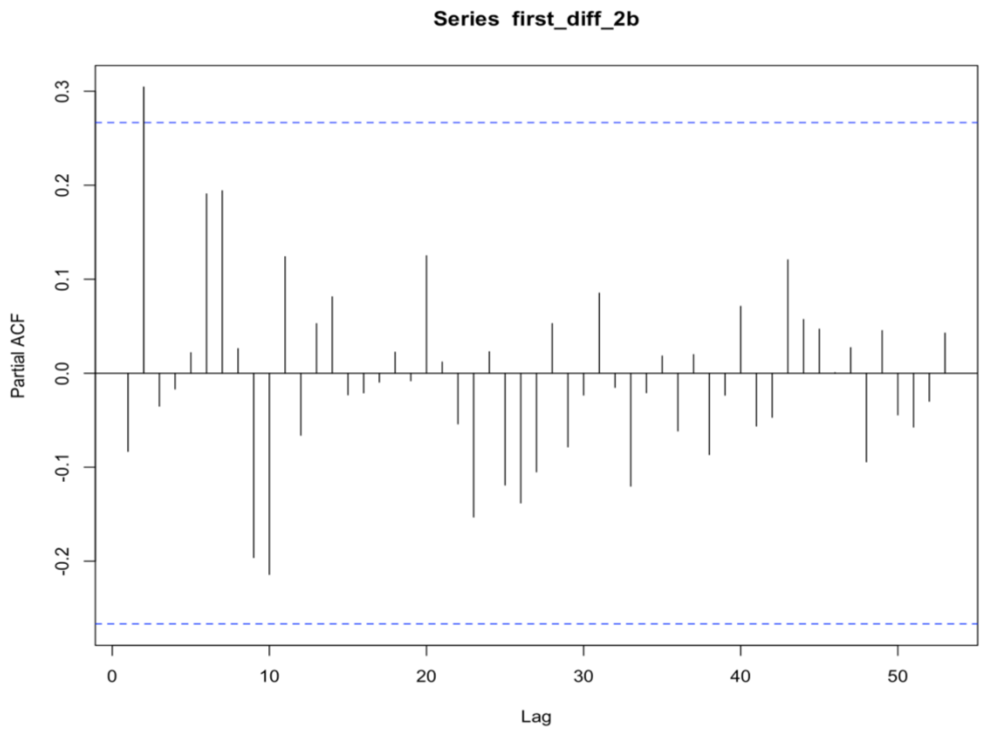
b) Plot the first difference, ACF, and PACF. (6 pts)
First Difference:



ACF:



PACF:



c) **What model order, i.e. p,d,q in ARMA(p,d,q), would you recommend for the above time series. Provide a justification for your answer.**

I would recommend the ARMA(2,1,0) model or ARMA(0,1,2) model. Reasons are as follow:

Given the ACF plot for the original data and the ACF plot for the first difference dataset, it is clear that the first difference dataset provides stationary data, while the original data is nonstationary. Thus, I recommend d = 1.

Based on the ACF plot for the first difference dataset, the lag = 2 is significant, so I recommend p = 2.

Based on the PACF plot for the first difference dataset, the lag = 2 is significant, so I recommend q = 2.