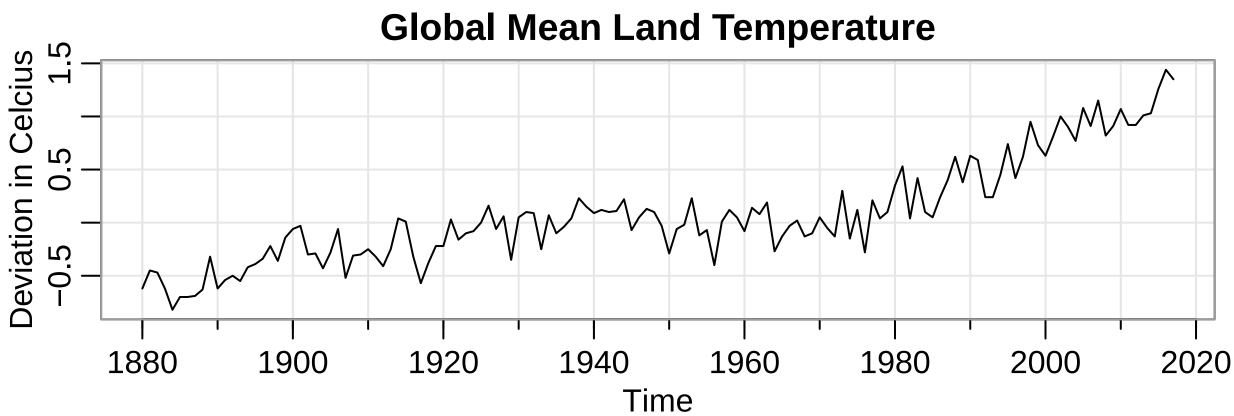
HW 09

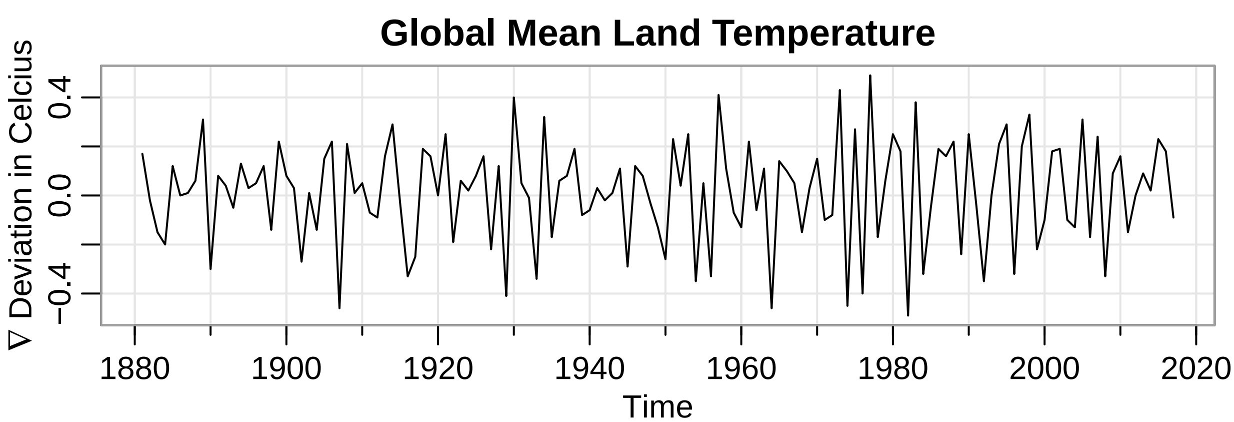
Name: Paul Holaway , NetID: paulch2

Due: 10/27/2022 11:59pm

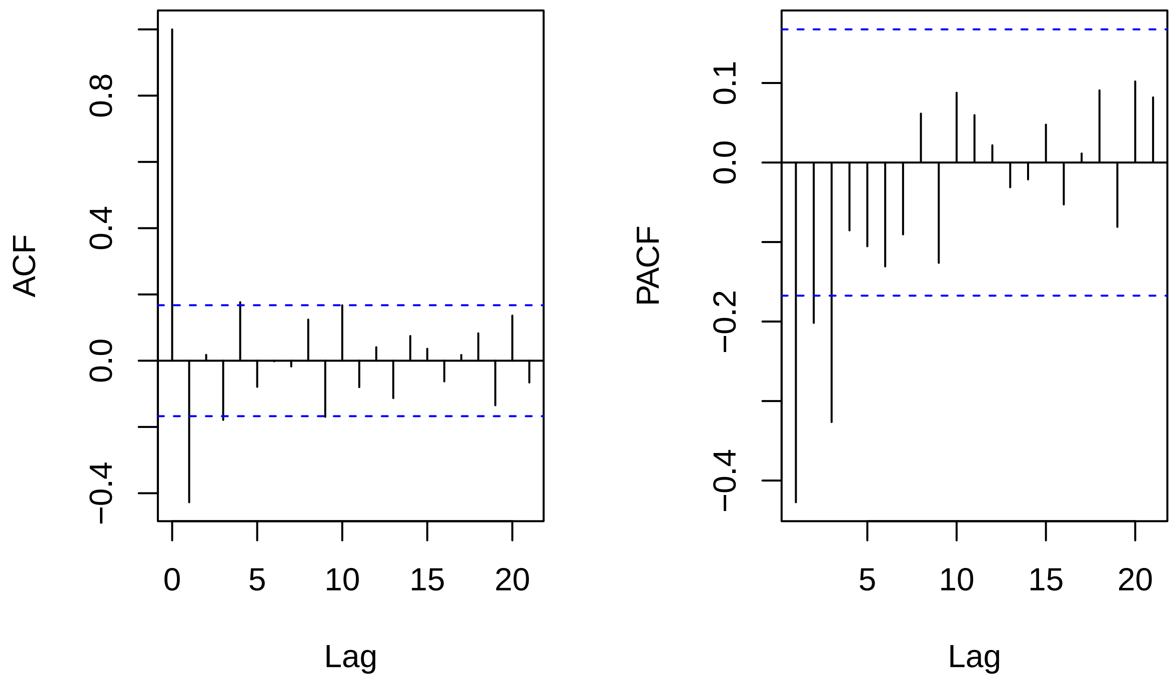
Question 1.

```
par(mfrow = c(2,1))
#Original Data
tsplot(gtemp_land, ylab = "Deviation in Celcius", main = "Global Mean Land Temperature")
#Differenced Data
tsplot(diff(gtemp_land), ylab = TeX(r"($\nabla$ Deviation in Celcius)"),
       main = "Global Mean Land Temperature")
```





```r
par(mfrow = c(1,2))
acf(diff(gtemp_land), main = "")
pacf(diff(gtemp_land), ylab = "PACF", main = "")
```



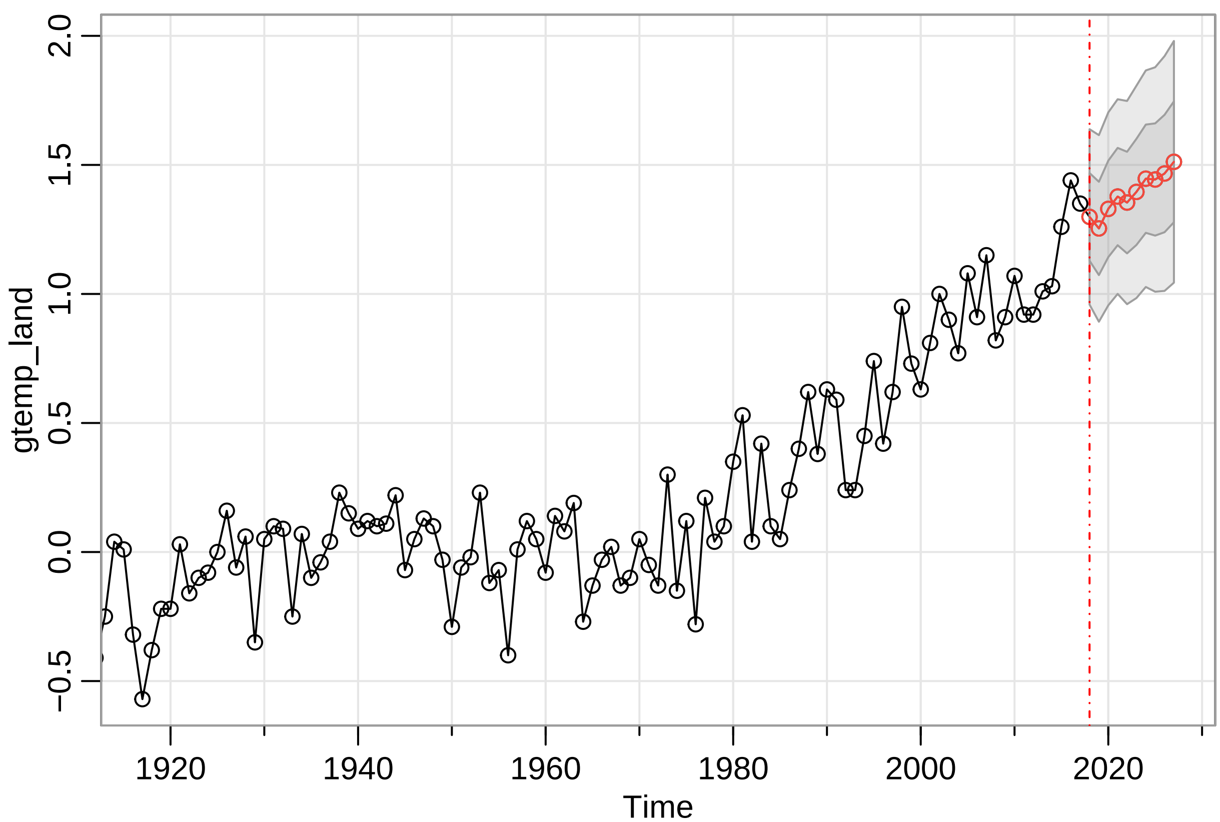
```r
#Use p = 3, d = 1, q = 4
sarima(gtemp_land, p = 3, d = 1, q = 4, no.constant = TRUE, details = FALSE)
```

```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##          ar1      ar2     ar3      ma1     ma2      ma3     ma4
##       0.3503  -0.1041  0.7186  -0.9871  0.2439  -0.8894  0.6816
## s.e.  0.1329   0.1839  0.1949   0.1159  0.1939   0.2498  0.1102
##
## sigma^2 estimated as 0.02886:  log likelihood = 47.5,  aic = -78.99
##
## $degrees_of_freedom
## [1] 130
##
## $ttable
##     Estimate     SE t.value p.value
## ar1   0.3503 0.1329  2.6370  0.0094
## ar2  -0.1041 0.1839 -0.5659  0.5725
## ar3   0.7186 0.1949  3.6873  0.0003
## ma1  -0.9871 0.1159 -8.5182  0.0000
## ma2   0.2439 0.1939  1.2581  0.2106
## ma3  -0.8894 0.2498 -3.5601  0.0005
## ma4   0.6816 0.1102  6.1847  0.0000
##
## $AIC
## [1] -0.5765854
##
## $AICc
## [1] -0.5702481
##
## $BIC
## [1] -0.4060756
```

```r
#Forecasting
sarima.for(gtemp_land, p = 3, d = 1, q = 4, n.ahead = 10, no.constant = TRUE)
```

```
## $pred
## Time Series:
## Start = 2018
## End = 2027
## Frequency = 1
##  [1] 1.298652 1.253831 1.329598 1.377389 1.354039 1.395330 1.446566 1.443441
##  [9] 1.466684 1.511970
##
## $se
## Time Series:
## Start = 2018
## End = 2027
## Frequency = 1
##  [1] 0.1698896 0.1807499 0.1869050 0.1885500 0.1968121 0.2056652 0.2097563
##  [8] 0.2173905 0.2274623 0.2341254
```

```r
abline(v = 2018, lty = 4, col = "red")
```

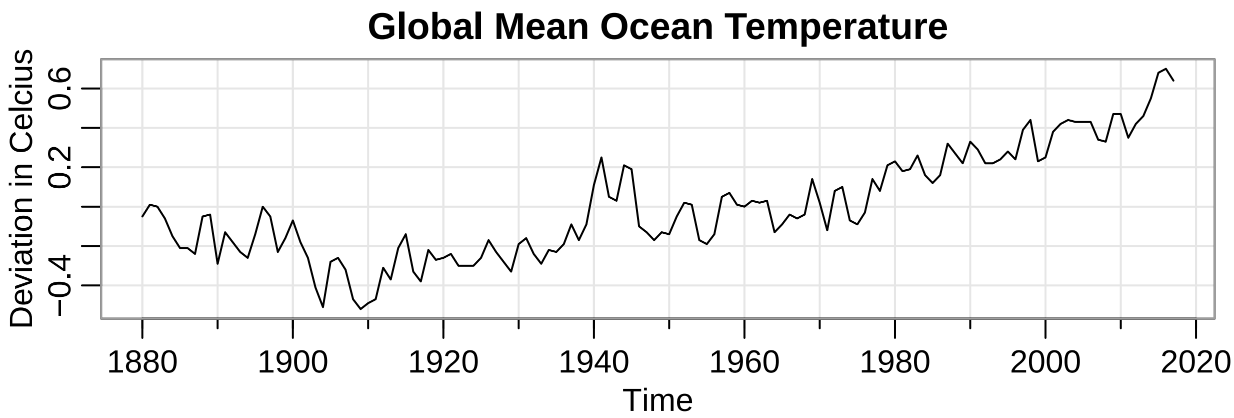


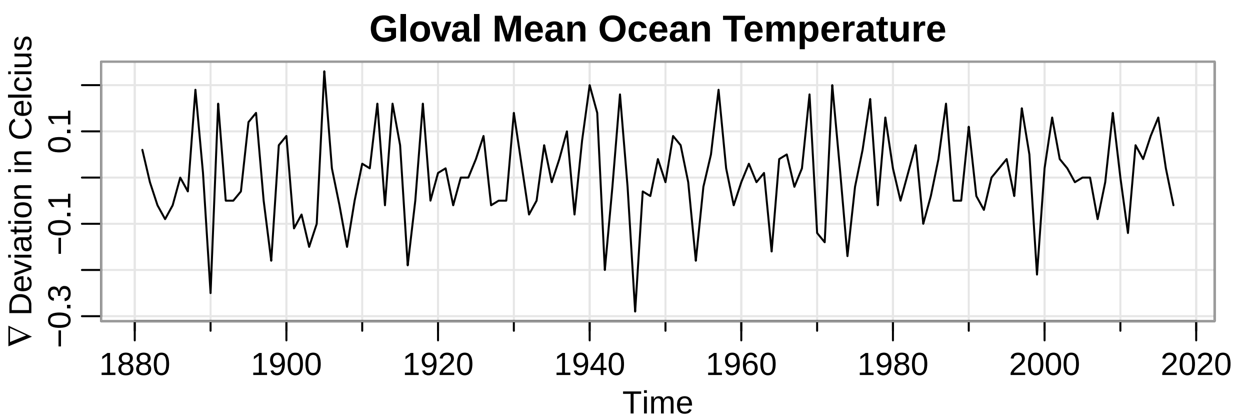
Final Model: $\boxed{\text{ARIMA(3,1,4)}}$

After looking at the original data, it is clear that the data must be differenced. After taking the first difference of the data, the trend in the original data has vanished. Work was done to see if there was a seasonal trend within the data, however after exploring periods between 2 and ten years, no visible improvement was seen in the plot or the model prediction. Therefore an ARIMA model was used as the final model for the data. After looking at the ACF and PACF, we can also see that there does not appear to be any kind of seasonal pattern, which would further support using just an ARIMA model. From the ACF, we can see a slight tailing off with a cutoff at lag 4 and lag 2 being insignificant. From the PACF, we can see a distinct cutoff at lag 3. Multiple ARIMA models were looked at (ARIMA(0,1,4), ARIMA(3,1,0), ARIMA(3,1,1), ARIMA(3,1,4)) and the model with the lowest AIC and BIC was the ARIMA(3,1,4) model. The AIC and BIC were lowest for the ARIMA(3,1,4) model and the remaining terms remained unchanged in terms of significance level, so an ARIMA(3,1,4) model was chosen for the final model. From the model predictions, we would expect the overall average land temperature to increase by as much as a 0.25 $^{\circ}C$ by 2027.

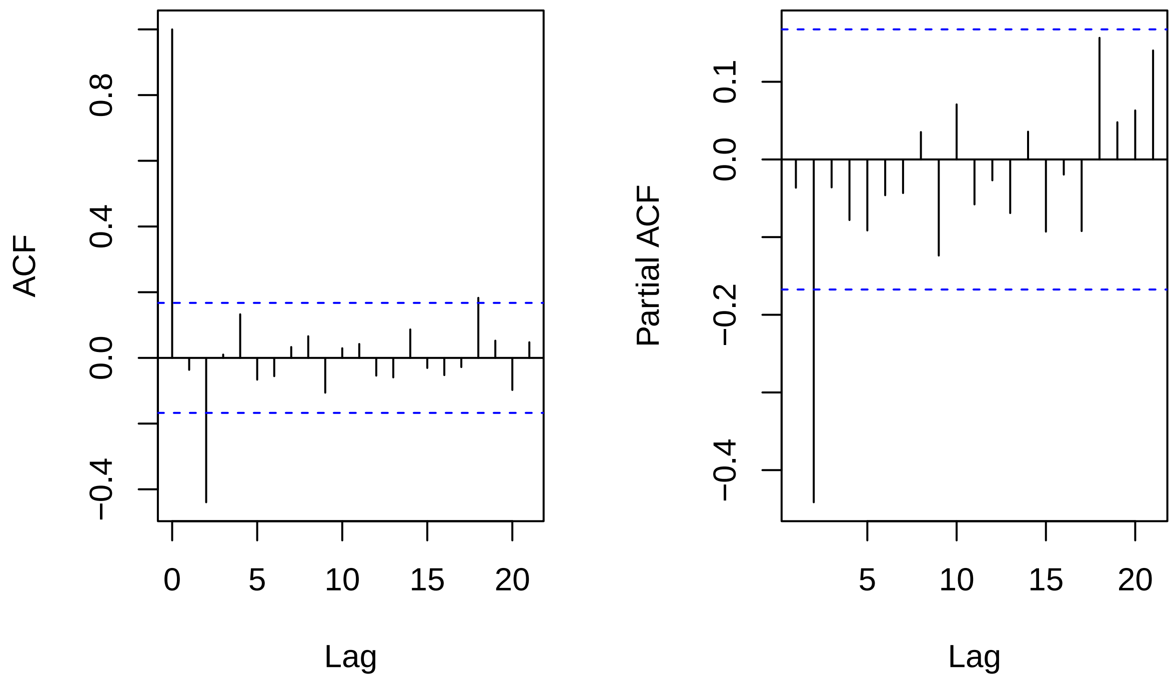
Question 2.

```
par(mfrow = c(2,1))
#Original Data
tsplot(gtemp_ocean, ylab = "Deviation in Celcius", main = "Global Mean Ocean Temperature")
#Differenced Data
tsplot(diff(gtemp_ocean), ylab = TeX(r"($\nabla$ Deviation in Celcius)"),
       main = "Gloval Mean Ocean Temperature")
```





```r
par(mfrow = c(1,2))
acf(diff(gtemp_ocean), main = "")
pacf(diff(gtemp_ocean), main = "")
```



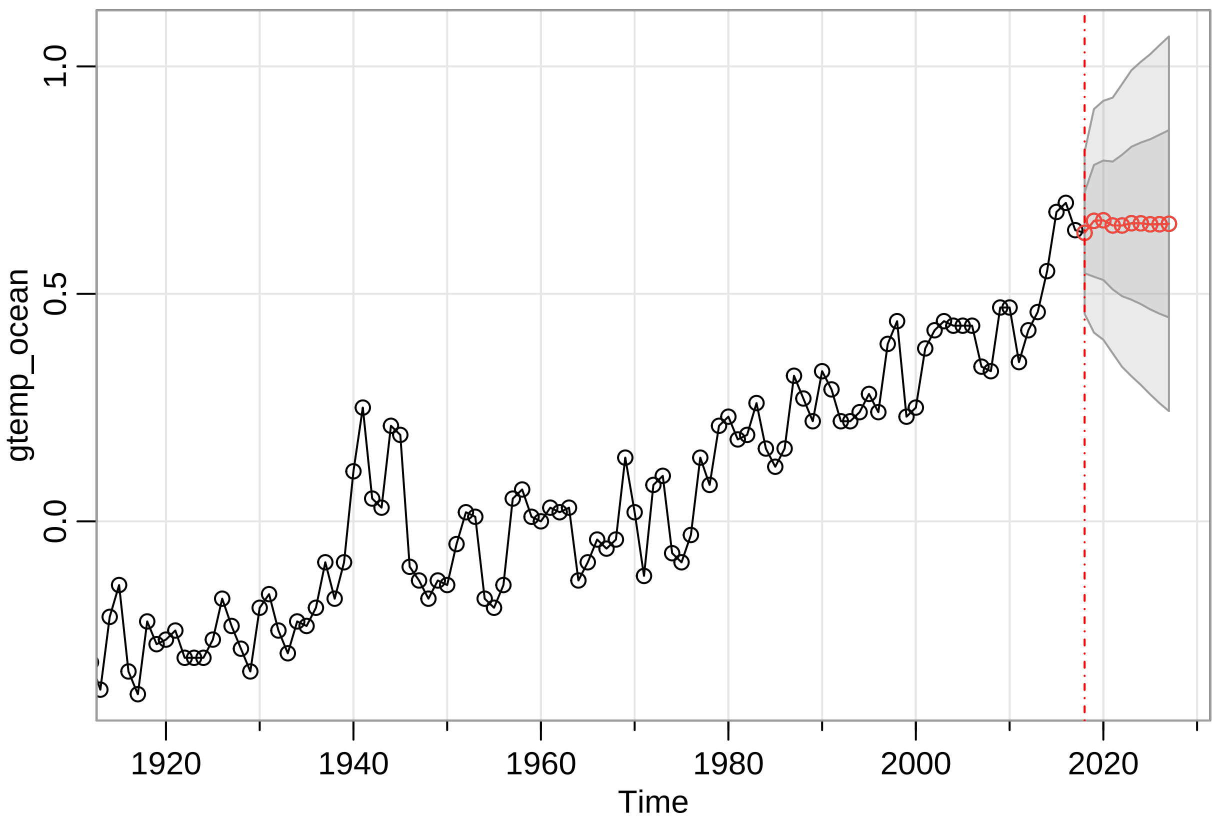
```r
#Use p = 2, d = 1, q = 2
sarima(gtemp_ocean, p = 2, d = 1, q = 0, no.constant = TRUE, details = FALSE)
```

```
## $fit
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##           ar1      ar2
##       -0.0481  -0.4334
## s.e.   0.0769   0.0762
##
## sigma^2 estimated as 0.007916:  log likelihood = 136.86,  aic = -267.72
##
## $degrees_of_freedom
## [1] 135
##
## $ttable
##     Estimate     SE t.value p.value
## ar1  -0.0481 0.0769 -0.6259  0.5324
## ar2  -0.4334 0.0762 -5.6856  0.0000
##
## $AIC
## [1] -1.954133
##
## $AICc
## [1] -1.953479
##
## $BIC
## [1] -1.890192
```

```r
#Forecasting
sarima.for(gtemp_ocean, p = 2, d = 1, q = 0, n.ahead = 10, no.constant = TRUE)
```

```
## $pred
## Time Series:
## Start = 2018
## End = 2027
## Frequency = 1
##  [1] 0.6342199 0.6604996 0.6617399 0.6502917 0.6503051 0.6552656 0.6550211
##  [8] 0.6528832 0.6530920 0.6540084
##
## $se
## Time Series:
## Start = 2018
## End = 2027
## Frequency = 1
##  [1] 0.08897274 0.12283648 0.13128685 0.14050038 0.15543591 0.16812967
##  [7] 0.17755419 0.18692851 0.19676198 0.20590788
```

```r
abline(v = 2018, lty = 4, col = "red")
```



Final Model: $\boxed{\text{ARIMA(2,1,0)}}$

After looking at the original data, it is clear that the data must be differenced. After taking the first difference of the data, the trend in the original data has vanished. Work was done to see if there was a seasonal trend within the data, however after exploring periods between 2 and ten years, no visible improvement was seen in the plot or the model prediction (just like in `gtemp_land`). Therefore an ARIMA model was used as the final model for the data. After looking at the ACF and PACF, we can also see that there does not appear to be any kind of seasonal pattern, which would further support using just an ARIMA model. From the ACF, we can see a distinct cutoff at lag 2 and lag 1 being insignificant. From the PACF, we can see a distinct cutoff at lag 2 with lag 1 also being insignificant. Multiple ARIMA models were looked at ($d = 1$, $p = 0, 1, 2$, $q = 0, 1, 2$ combinations) and the model with the lowest AIC and BIC was the ARIMA(2,1,0) model. However, its predictions did not change much beyond two to three years in the future. While it does give us a prediction, the SE of those predictions is also quite high. I would expect the one to three year predictions to be the most accurate using this model while anything beyond that to not be as accurate. With the overall global ocean temperature change having so many factors influencing it, it would make sense that any model made would have great difficulty predicting anything beyond the immediate future.

Question 3.

Consider the MA(1) series,

$$X_t = W_t + \theta W_{t-1},$$

where W_t is white noise with variance σ_w^2.

(a) Derive the minimum mean square error one-step forecast based on the infinite past, and determine the mean-square error for this forecast. Hint: see (3.86) of [TSA4].

$$X_t = (1+\theta B)W_t \implies W_t = (1+\theta B)^{-1}X_t$$

$$\text{By Geometric Series} \implies W_t = \sum_{j=0}^{\infty}(-\theta)^j X_{t-j}$$

$$X_{n+1} = W_{n+1} + \theta W_n = W_{n+1} + \theta\sum_{j=0}^{\infty}(-\theta)^j X_{t-j}$$

$$\implies X_{n+1}^n = W_{n+1}^n + \theta\sum_{j=0}^{\infty}(-\theta)^j X_{t-j}^n$$

$$\text{With } X_n^n = X_n \text{ and } W_{n+1}^n = 0 \implies \boxed{X_{n+1}^n = \theta\sum_{j=0}^{\infty}(-\theta)^j X_{t-j}}$$

$$P_{n+1}^n = \mathbb{E}[(X_{n+1} - X_{n+1}^n)^2] = \mathbb{E}[(W_{n+1} + \theta\sum_{j=0}^{\infty}(-\theta)^j X_{t-j} - \theta\sum_{j=0}^{\infty}(-\theta)^j X_{t-j})^2] = \mathbb{E}[(W_{n+1})^2]$$

$$\sigma_w^2 = \mathbb{E}[(W_{n+1})^2] - [\mathbb{E}(W_{n+1})]^2 \text{ and } \mathbb{E}(W_{n+1} = 0) \implies \sigma_w^2 = \mathbb{E}[(W_{n+1})^2]$$

$$\implies \boxed{P_{n+1}^n = \sigma_w^2}$$

(b) Let X_{n+1}^n be the truncated one-step-ahead forecast, as in (3.92) of [TSA4].

Show that

$$E[(X_{n+1} - X_{n+1}^n)^2] = \sigma_w^2(1+\theta^{2+2n}).$$

$$\mathbb{E}[(X_{n+1} - X_{n+1}^n)^2] = \mathbb{E}[(W_{n+1} + \theta\sum_{j=0}^{n}(-\theta)^j X_{t-j})^2]$$

$$= \mathbb{E}[(W_{n+1} + \theta(1+\theta^{n+1})W_n)^2]$$

$$= \mathbb{E}[(W_{n+1})^2 + 2\theta(1+\theta^{n+1})W_n W_{n+1} + \theta^2(1+\theta^{n+1})^2 W_n^2]$$

$$= \sigma_w^2 + \sigma_w^2\theta^{2n+2} = \boxed{\sigma_w^2(1+\theta^{2n+2})}$$