

Economics 144 - Project 2

Ian Chisholm & Niyant Gurwara

5/31/2017

I: Introduction

For an earlier project, we examined temperatures in Southern California. Stemming from the encouraging results of that project, we wanted to continue our researching and working with meteorological data. There were two reasons for this. Firstly, the data is abundantly available, reaching back more than a century. Second, We wanted to identify environmental trends prevalent today. Considering that climate change is one of the fundamental problems of our generation, we wanted to focus on some aspect of this global process. Our goal was to statistically prove the existence of a physical process that represents climate change. To this effect, we chose to analyze sea levels in relation to rising global temperatures.

Possibly the most simplified explanation for global warming is as follows: accumulation of greenhouse gases is depleting the ozone layer and thereby leading to an increase in mean temperatures around the world. Rising temperatures are melting the polar ice caps, thereby leading to a rise in global mean sea levels. We wanted to prove this relationship between greenhouse gases and sea levels through statistical analysis. Our first idea was to use some measure of mean global sea levels as our dependent variable and Carbon Dioxide (CO₂) levels as our independent variable in a Variable Autoregressive (VAR) Model.

However, upon running some primary diagnostics, our data did not suggest causality between the carbon dioxide and sea levels. After further research, we concluded that this is because radiative forcing, a process that measures the change in energy balance due to greenhouse gases, is significantly delayed. It would be difficult to prove causality between the two variables without thoroughly examining each of the links in between. In order to more accurately predict a relationship, we shifted our attention to another process instead. We decided to model mean sea level as a function of the average temperature in Northwestern California (the dataset we analyzed in project 1). We expected that temperatures and sea levels should be positively correlated, with temperatures causing sea levels to rise.

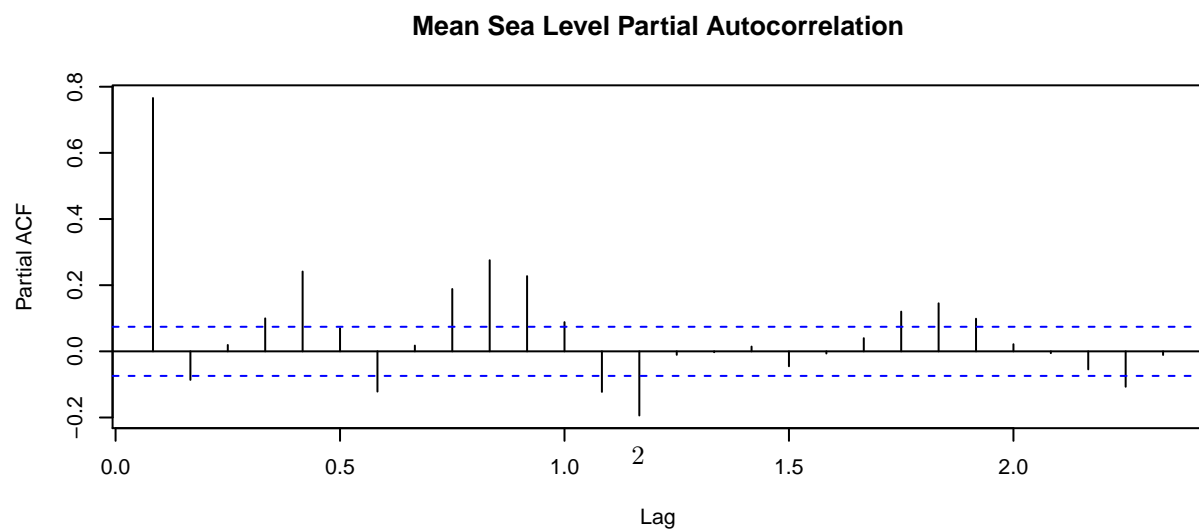
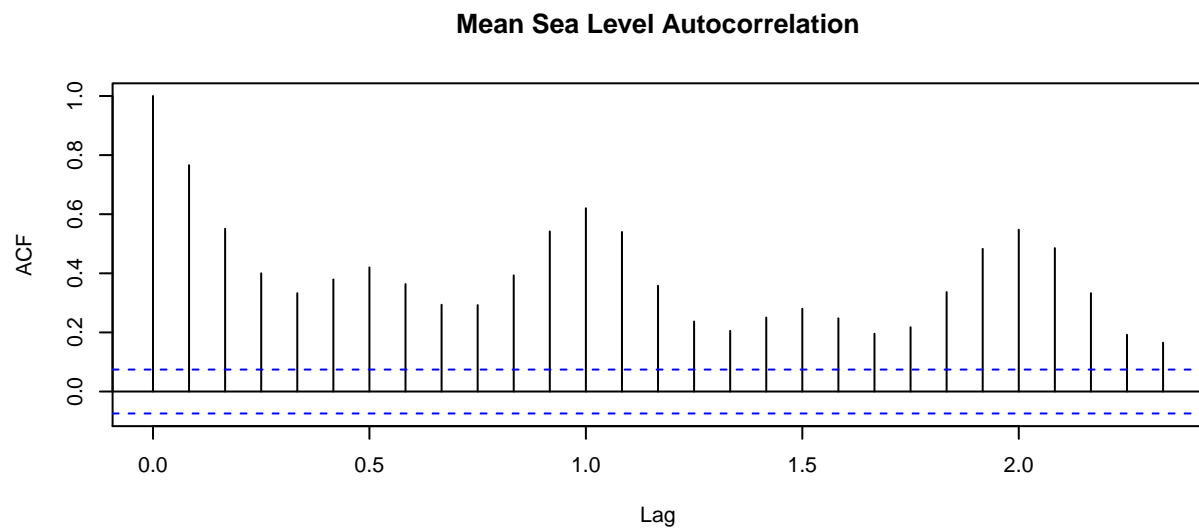
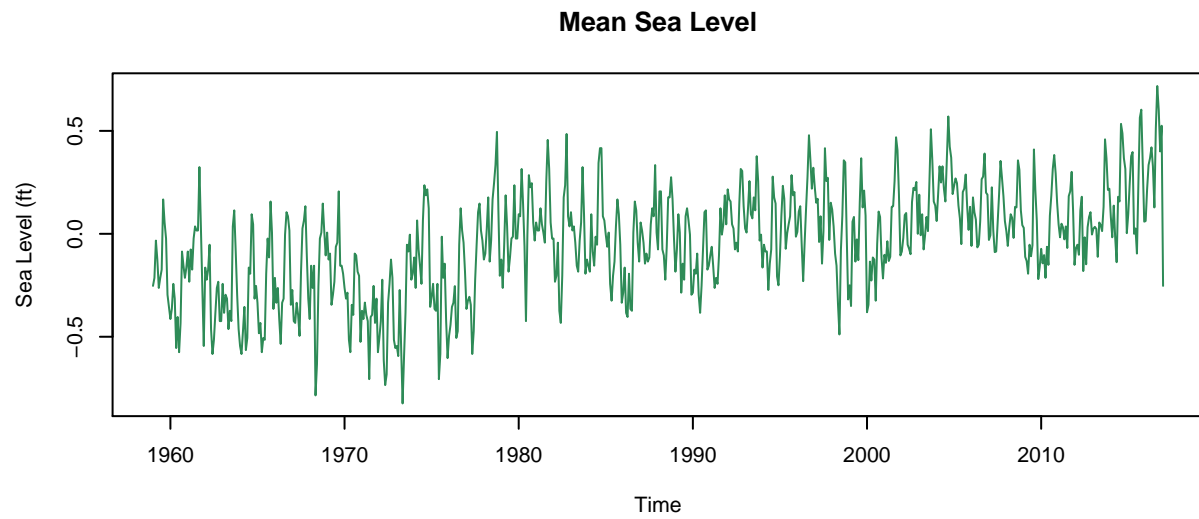
Our data consists of monthly observations for both the time series chosen. Sea level is measured by height in feet. Temperature is measured in Fahrenheit. Observations for both series begin in 1959 and run through 2016. The data was acquired from the National Oceanic and Atmospheric Administration (see references).

The source code for each sub-part of a topic can be found under the corresponding title and serial number in Part V of this report.

II: Results

Trend, Seasonality and Cycles

(a) Time Series Plots



(b) Model with Trend, Seasonal and Cyclical components

We estimated models with different components to ultimately build a model that appropriately fits the data. We begin with the trend (linear or otherwise). After that we move onto seasonality, and ultimately cycles.

Linear Trend

```
##
## Call:
## lm(formula = sl ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65163 -0.15034  0.00142  0.14352  0.62086
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.712e+01  9.472e-01  -18.07  <2e-16 ***
## t             8.586e-03  4.764e-04   18.02  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2109 on 695 degrees of freedom
## Multiple R-squared:  0.3185, Adjusted R-squared:  0.3175
## F-statistic: 324.8 on 1 and 695 DF,  p-value: < 2.2e-16
```

The linear model had two statistically significant coefficients at the 99% confidence interval. Next, we attempted to fit a quadratic trend.

Quadratic Trend

```
##
## Call:
## lm(formula = sl ~ t + I(t^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.65361 -0.15125  0.00248  0.14582  0.61516
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.314e+02  1.256e+02  -1.047   0.296
## t             1.236e-01  1.263e-01   0.979   0.328
## I(t^2)       -2.893e-05  3.177e-05  -0.911   0.363
##
## Residual standard error: 0.2109 on 694 degrees of freedom
## Multiple R-squared:  0.3193, Adjusted R-squared:  0.3173
## F-statistic: 162.8 on 2 and 694 DF,  p-value: < 2.2e-16
```

The quadratic model had no statistically significant coefficients. As a result, it is apparent that the linear trend is a more appropriate fit for this data.

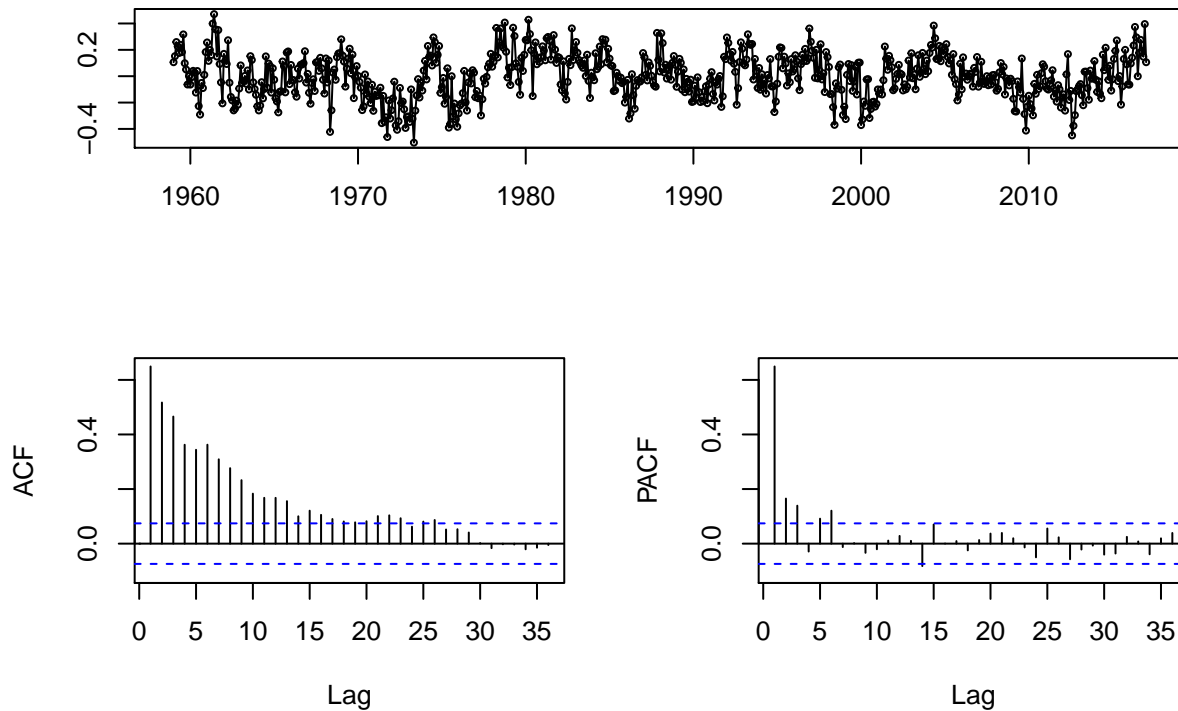
Trend + Seasonality

```
##
## Call:
## tslm(formula = sl ~ t + season)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.50413 -0.11007 -0.00478  0.11950  0.47083
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.704e+01  7.566e-01 -22.518  < 2e-16 ***
## t            8.514e-03  3.806e-04  22.368  < 2e-16 ***
## season2     -7.822e-03  3.115e-02  -0.251  0.801812
## season3      6.545e-02  3.115e-02   2.101  0.035987 *
## season4      5.590e-03  3.115e-02   0.179  0.857626
## season5     -8.290e-02  3.115e-02  -2.661  0.007966 **
## season6     -9.640e-02  3.115e-02  -3.095  0.002051 **
## season7     -2.097e-03  3.115e-02  -0.067  0.946338
## season8      2.039e-01  3.115e-02   6.545  1.17e-10 ***
## season9      3.100e-01  3.115e-02   9.951  < 2e-16 ***
## season10     2.790e-01  3.115e-02   8.955  < 2e-16 ***
## season11     1.190e-01  3.115e-02   3.819  0.000146 ***
## season12    -4.722e-03  3.115e-02  -0.152  0.879572
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1685 on 684 degrees of freedom
## Multiple R-squared:  0.5721, Adjusted R-squared:  0.5646
## F-statistic: 76.22 on 12 and 684 DF,  p-value: < 2.2e-16
```

Next, we fit a model consisting of both, trend and seasonality. A summary of this model indicates that 7 of the 12 seasonal components were statistically significant at the 99% confidence interval level. This means that the data in our time series fluctuates significantly during certain seasons, implying that seasonality plays a big role in the underlying process driving this time series.

To understand the effect of including seasonality in the model, we examined the residuals of this updated model. The residuals provide information regarding any other dynamics that we may have not covered so far.

Trend + Seasonal Model Residuals



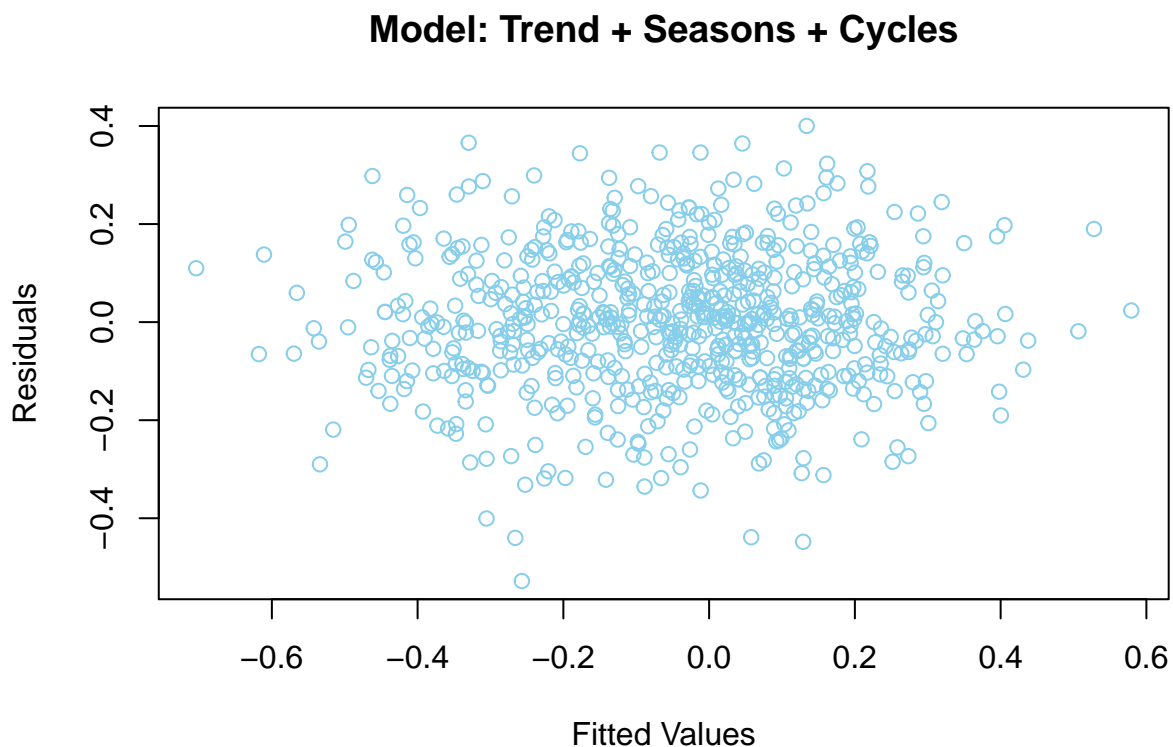
Our seasonal model represented a significant improvement over the linear model (evident in its greater s^2). However, there are clearly cyclical dynamics present in the residuals of this model. The ACF and PACF of the residuals of this model suggest that there still may be autoregressive processes driving the data. We determined that either an AR(3) or AR(5) model would capture these dynamics. Additionally, we included an S-AR(2) component in this new model, choosing $p=2$ because that is the value chosen by R's 'auto.arima' function.

```
##
## Call:
## arima(x = sl, order = c(3, 0, 0), seasonal = list(order = c(2, 0, 0)), xreg = t)
##
## Coefficients:
##          ar1      ar2      ar3      sar1      sar2  intercept          t
##          0.5606  0.0530  0.0465  0.3279  0.2340   -19.4291   0.0098
## s.e.      0.0406  0.0447  0.0403  0.0382  0.0401     3.1541   0.0016
##
## sigma^2 estimated as 0.02:  log likelihood = 372.12,  aic = -728.24
##
## Training set error measures:
##              ME      RMSE      MAE  MPE  MAPE      MASE
## Training set -0.0007379456 0.141437 0.1106599 -Inf  Inf  0.8049066
##              ACF1
## Training set 0.002253995
##
## Call:
## arima(x = sl, order = c(5, 0, 0), seasonal = list(order = c(2, 0, 0)), xreg = t)
##
## Coefficients:
##          ar1      ar2      ar3      ar4      ar5      sar1      sar2  intercept
```

```
##      0.5753  0.0413  0.0526 -0.1045  0.1251  0.3126  0.2214   -19.537
## s.e.  0.0406  0.0448  0.0455   0.0436  0.0386  0.0389  0.0404    3.208
##      t
##      0.0098
## s.e.  0.0016
##
## sigma^2 estimated as 0.01969:  log likelihood = 377.76,  aic = -735.53
##
## Training set error measures:
##              ME      RMSE      MAE  MPE MAPE      MASE
## Training set -0.0008495079  0.1403319  0.1098056 -Inf  Inf  0.7986923
##              ACF1
## Training set -0.01199922
## [1] -691.867
## [1] -690.0575
```

We tested both of these models with seasonal AR(2) components. Comparing the AIC and BIC of these two models, we found that an AR(3), Seasonal AR(2) model would be the best fit. The amplitude of the residuals of this model was smaller than that without the seasonal AR components.

(c) Residuals vs. Fitted Values



The residuals of our model do not indicate any clear pattern. All of the residuals lie within the range of -0.4 to 0.4, with a majority of them concentrated within the range of -0.2 to 0.2. This is a good indication of the fit of the model. There seem to be no structural breaks in the data as the residuals are quite uniformly spread out within the ranges mentioned above.

(d) Autocorrelation and Partial Autocorrelation Plots

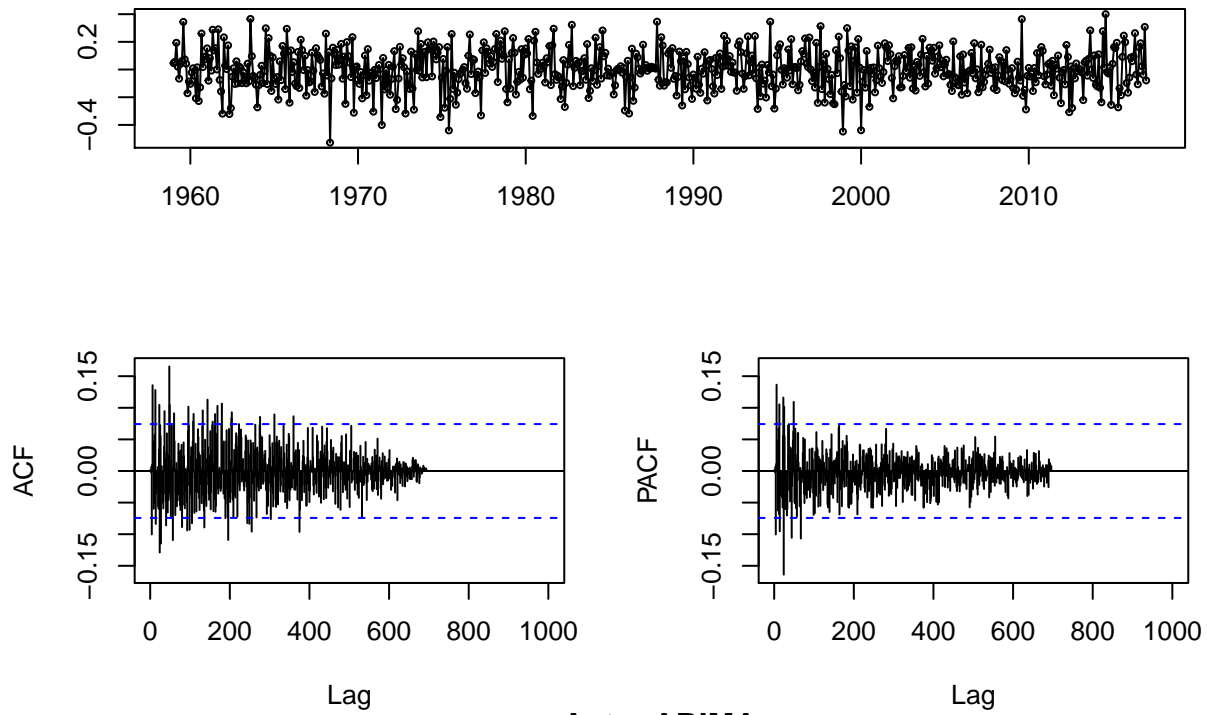
Testing for White Noise

```
##  
## Box-Ljung test  
##  
## data: model$residuals  
## X-squared = 0.0035564, df = 1, p-value = 0.9524  
  
##  
## Box-Pierce test  
##  
## data: model$residuals  
## X-squared = 0.0035411, df = 1, p-value = 0.9525
```

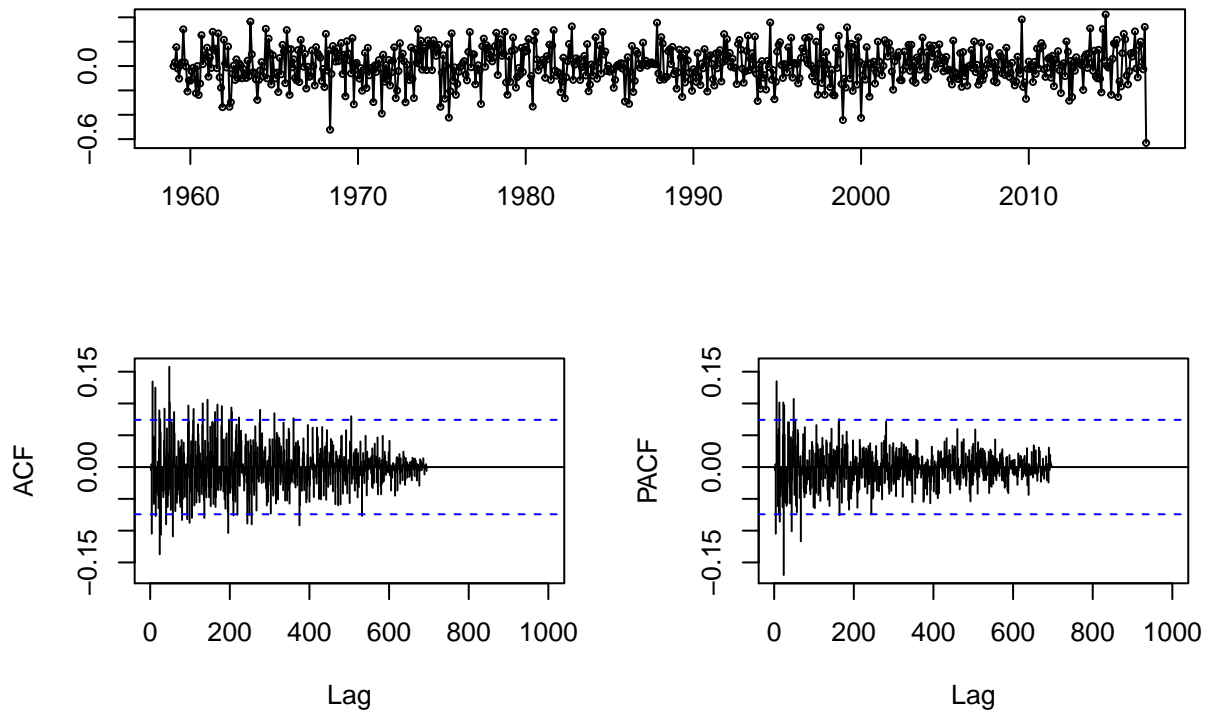
The ACF and PACF of the residuals of our model did not perfectly resemble a white noise process. This is confirmed by the results of the Ljung-Box and Box-Pierce tests for white noise processes. However, the residuals were still significantly small. This created a doubt as to whether we had captured all of the underlying dynamics. To test this, we applied an auto arima function. Its output provides R's estimation of a model that best fits the data chosen. We compared the residuals of such a model with that of the model we had thus far estimated. The purpose was to check the accuracy of the model at this point.

Comparing Best Fit Model to Auto Arima

Our Model

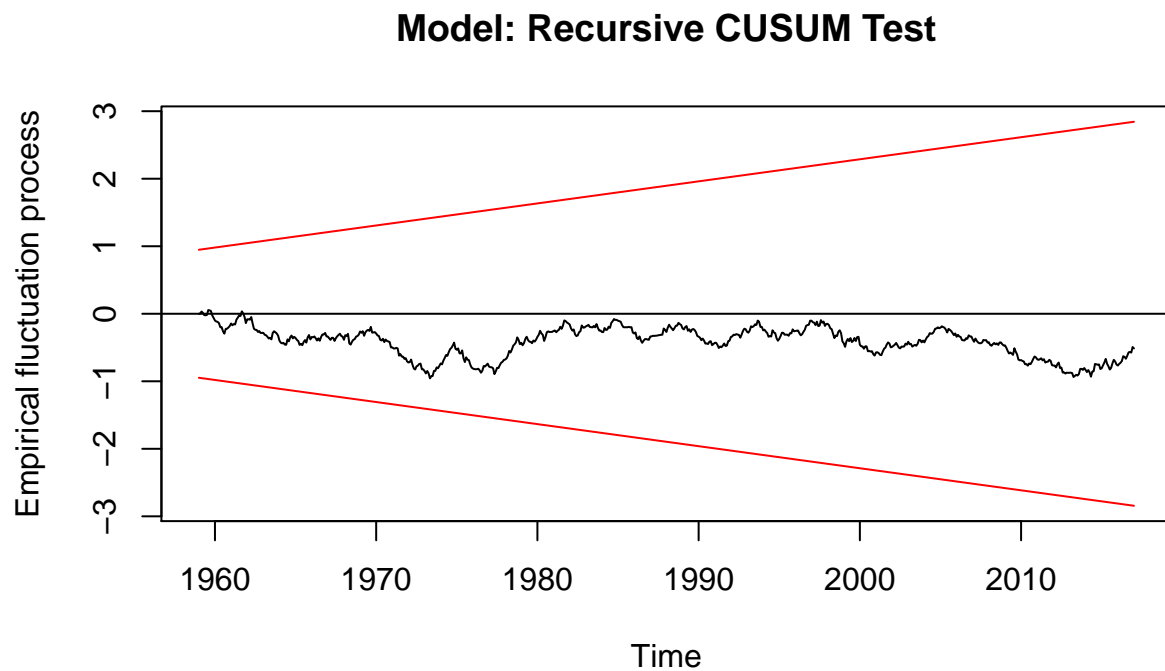


Auto-ARIMA



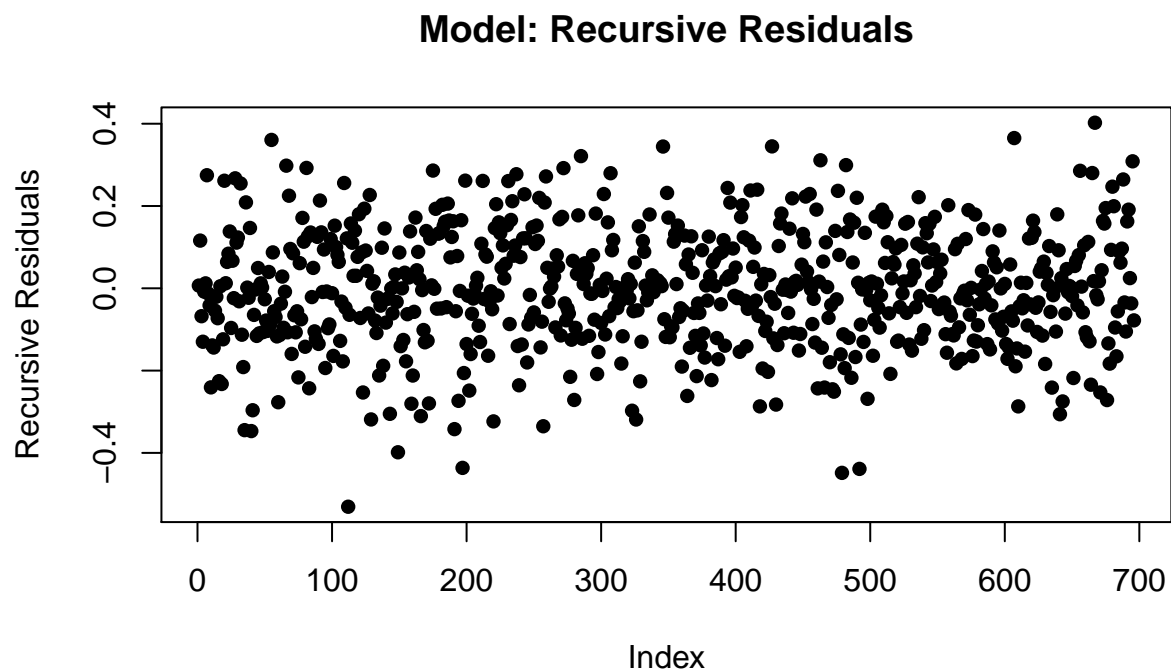
We found that the ACF and PACF of the best fit model according to R were very similar to those of our model. Thus we are confident that we have captured all the underlying dynamics in the chosen time series.

(e) Cumulative Sum Plot



It appears that the model is structurally sound. Everything in the plot lies within the 2 red lines. This means that the underlying engine that drives this time series has not changed. It also means that any structural breaks we may observe are not statistically significant.

(f) Recursive Residuals



The recursive residuals are uniformly spread out. They are not particularly large or small within any given time period. This means that the model is able to fit to the data with practically the same level of accuracy

throughout the length of the series.

(g) Diagnostic Statistics

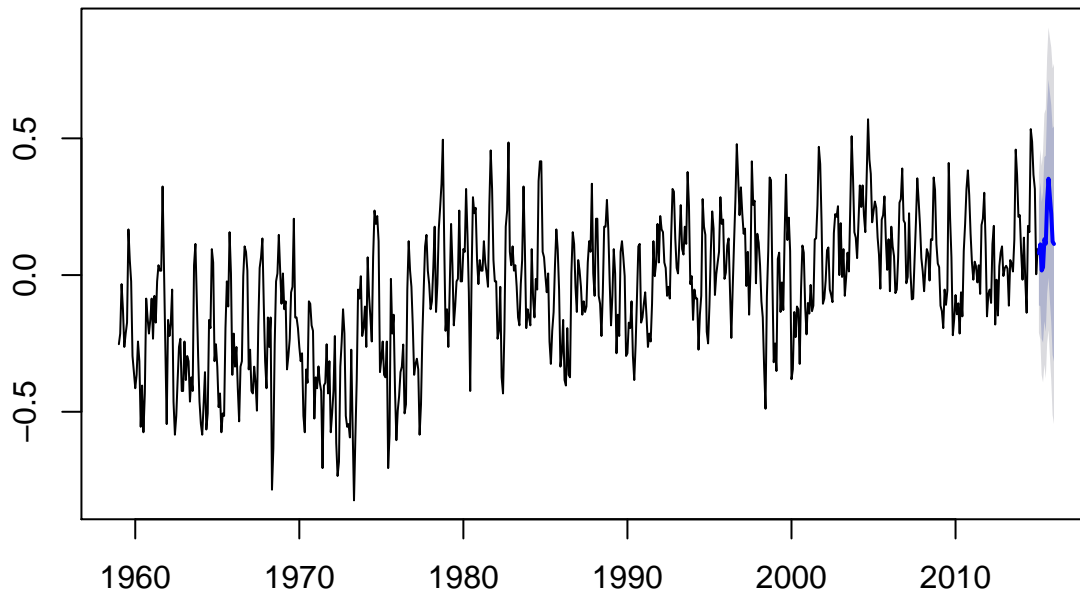
```
##              ME      RMSE      MAE  MPE  MAPE      MASE
## Training set -0.0007379456 0.141437 0.1106599 -Inf   Inf 0.8049066
##              ACF1
## Training set 0.002253995
```

Mean Error: -0.0007379456 Root Mean Squared Error: 0.141437 Mean Annual Error: 0.1106599

The ME and RMSE of our model are both relatively small. This is a good indication as it validates the accuracy of our model.

(h) 12-Step Ahead Forecast

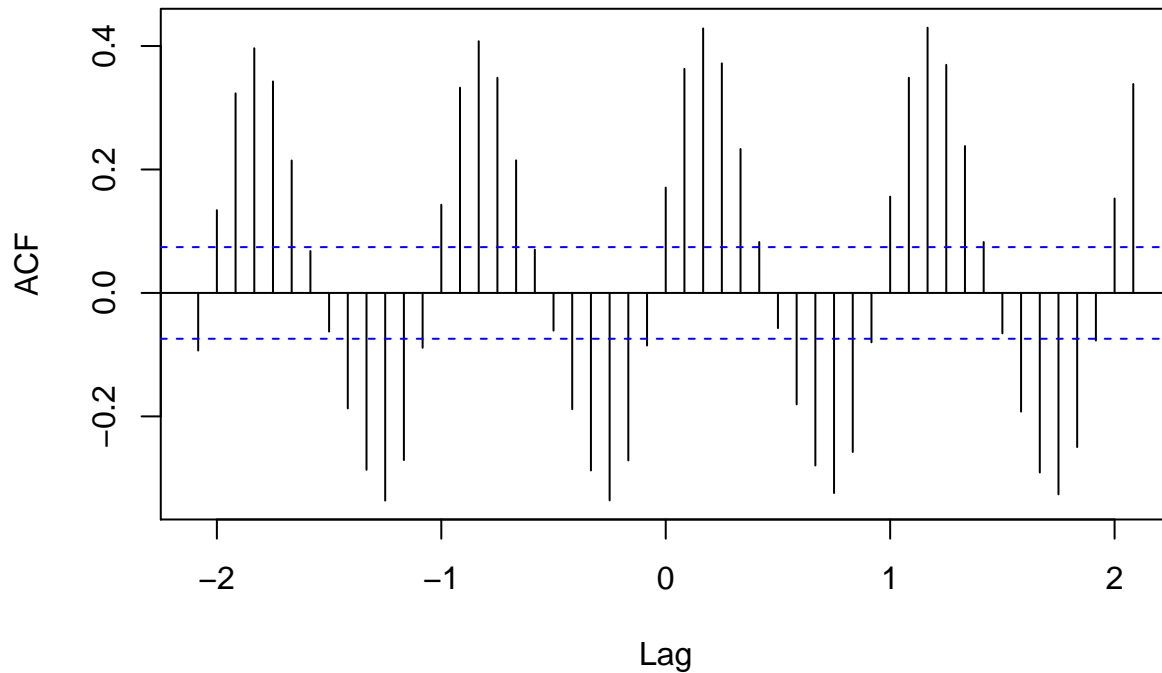
Forecasts from ARIMA(3,1,0)(2,0,0)[12]



(i) VAR Model

Next, we appropriated a VAR model. The primary purpose of this was to prove causality between sea levels and temperature.

Cross-Correlation: Sea Levels and Temperature



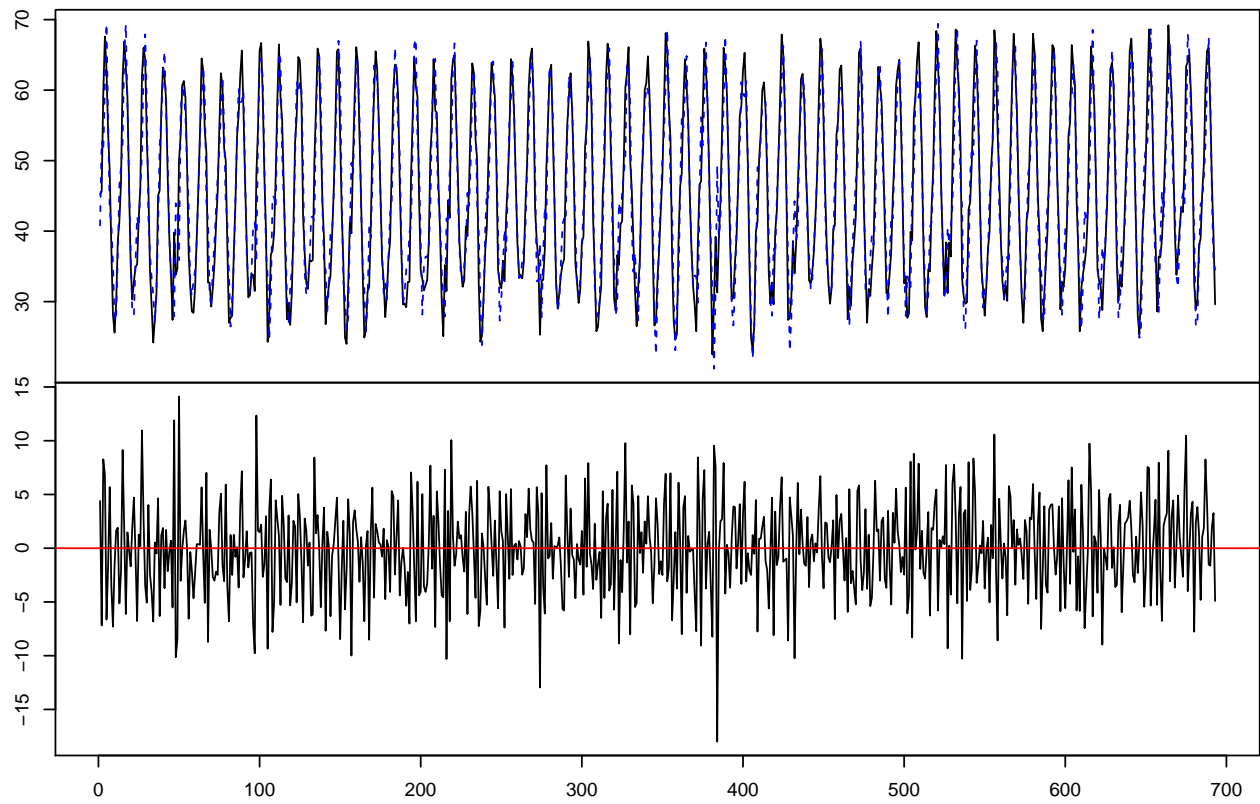
```
##
## VAR Estimation Results:
## =====
## Endogenous variables: temp, sl
## Deterministic variables: const
## Sample size: 693
## Log Likelihood: -1664.804
## Roots of the characteristic polynomial:
## 0.944 0.944 0.8705 0.6184 0.4027 0.4027
## Call:
## VAR(y = var.df, p = 3)
##
##
## Estimation results for equation temp:
## =====
## temp = temp.l1 + sl.l1 + temp.l2 + sl.l2 + temp.l3 + sl.l3 + const
##
##      Estimate Std. Error t value Pr(>|t|)
## temp.l1  1.02983    0.03197  32.211 < 2e-16 ***
## sl.l1    -7.22237    1.11628  -6.470 1.86e-10 ***
## temp.l2   0.12232    0.05036   2.429  0.0154 *
## sl.l2     0.86508    1.39674   0.619  0.5359
## temp.l3  -0.55485    0.03237 -17.143 < 2e-16 ***
## sl.l3     5.70381    1.12469   5.071 5.09e-07 ***
## const    18.26475    0.84039  21.734 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
```

```

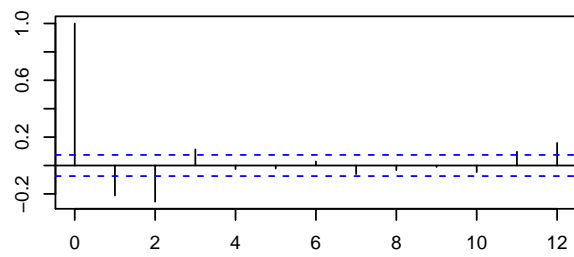
## Residual standard error: 4.394 on 686 degrees of freedom
## Multiple R-Squared: 0.8825, Adjusted R-squared: 0.8815
## F-statistic: 859.1 on 6 and 686 DF, p-value: < 2.2e-16
##
##
## Estimation results for equation sl:
## =====
## sl = temp.l1 + sl.l1 + temp.l2 + sl.l2 + temp.l3 + sl.l3 + const
##
##           Estimate Std. Error t value Pr(>|t|)
## temp.l1  0.0060132  0.0010823   5.556 3.95e-08 ***
## sl.l1    0.7089338  0.0377888  18.760 < 2e-16 ***
## temp.l2  0.0005743  0.0017049   0.337  0.7363
## sl.l2   -0.0543870  0.0472830  -1.150  0.2504
## temp.l3 -0.0019765  0.0010956  -1.804  0.0717 .
## sl.l3    0.1605182  0.0380735   4.216 2.82e-05 ***
## const   -0.2173526  0.0284493  -7.640 7.33e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1488 on 686 degrees of freedom
## Multiple R-Squared: 0.6645, Adjusted R-squared: 0.6616
## F-statistic: 226.4 on 6 and 686 DF, p-value: < 2.2e-16
##
##
## Covariance matrix of residuals:
##           temp      sl
## temp 19.31133 0.01744
## sl    0.01744 0.02213
##
## Correlation matrix of residuals:
##           temp      sl
## temp 1.00000 0.02668
## sl    0.02668 1.00000

```

Diagram of fit and residuals for temp



ACF Residuals



PACF Residuals

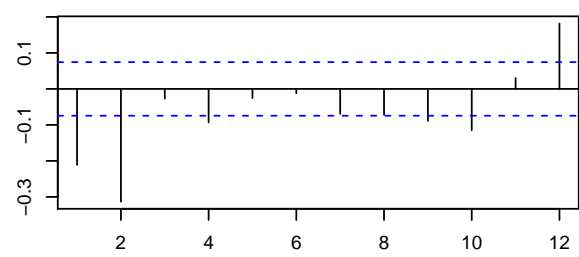
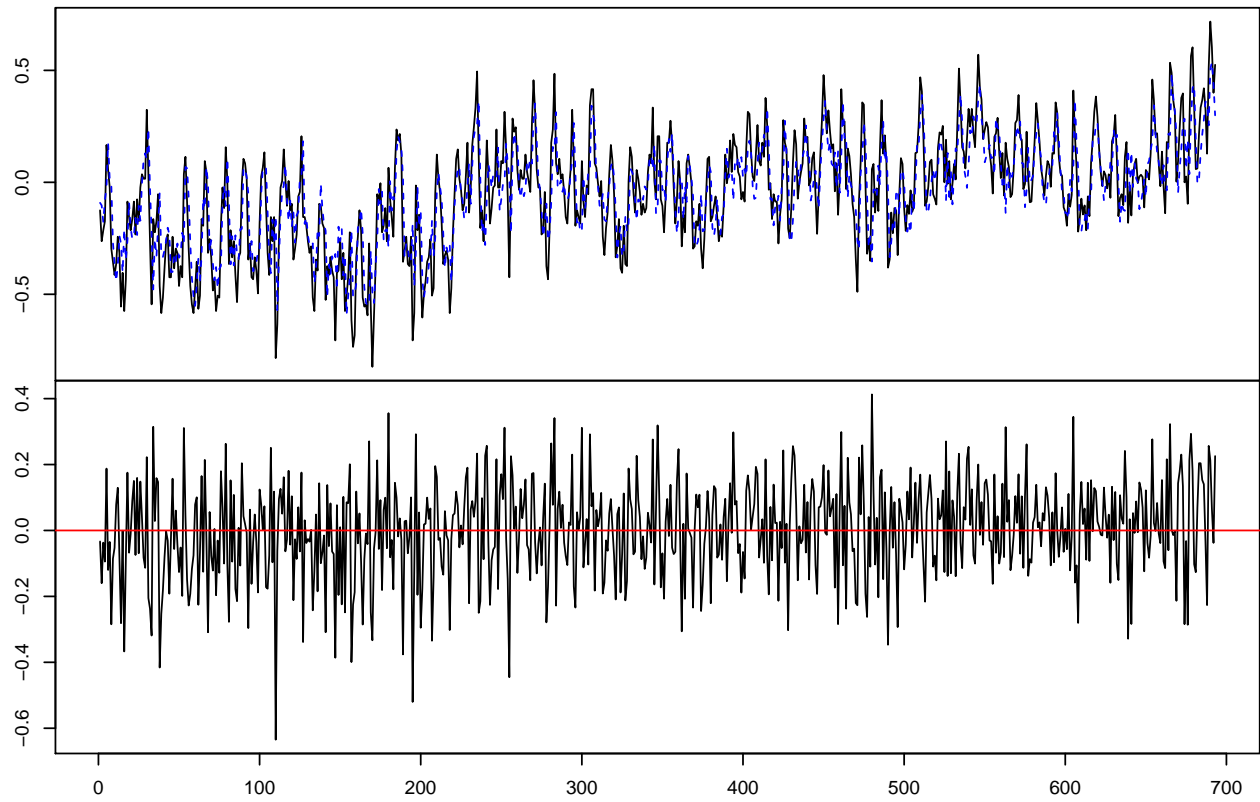
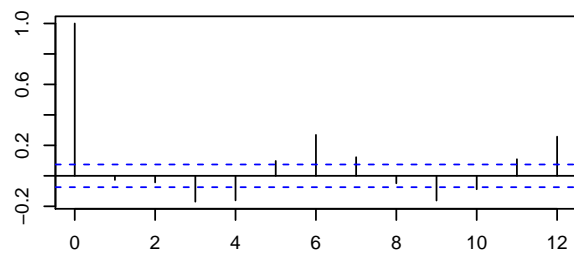


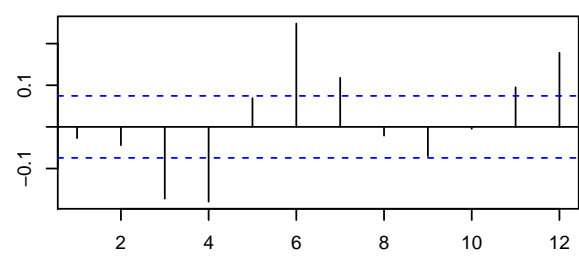
Diagram of fit and residuals for sl



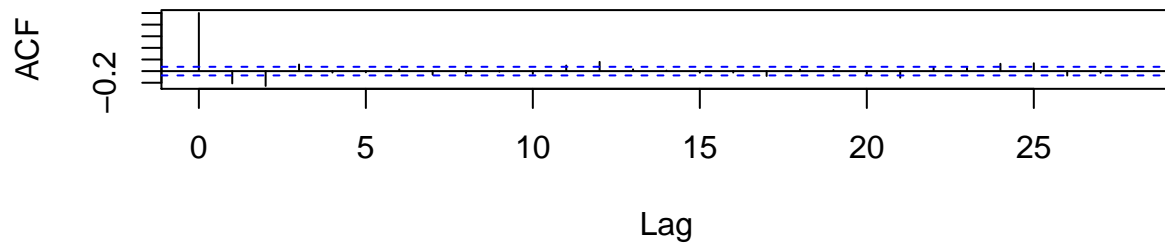
ACF Residuals



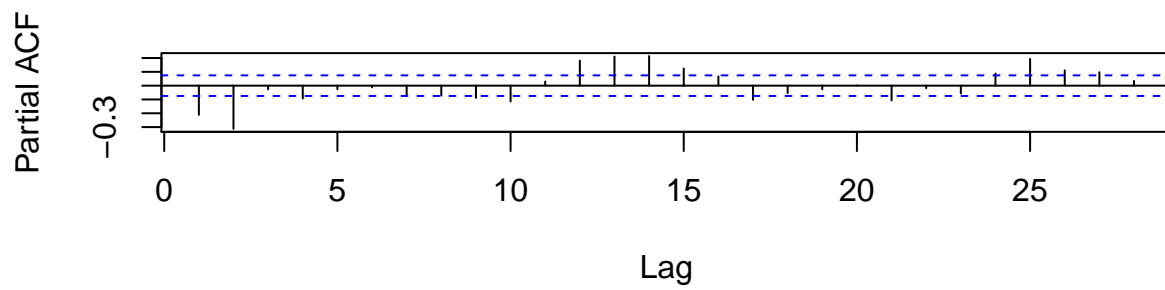
PACF Residuals



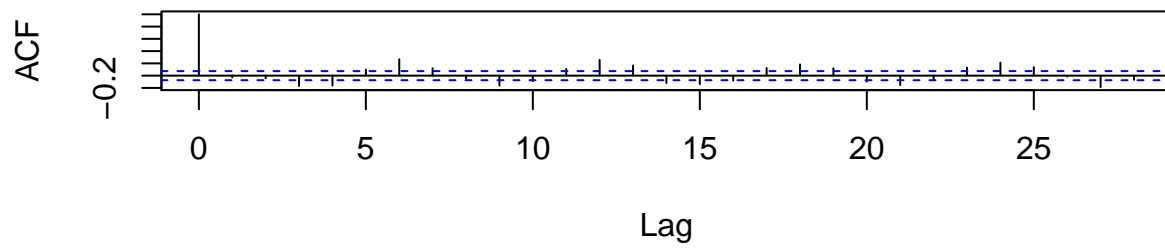
y=Temperature



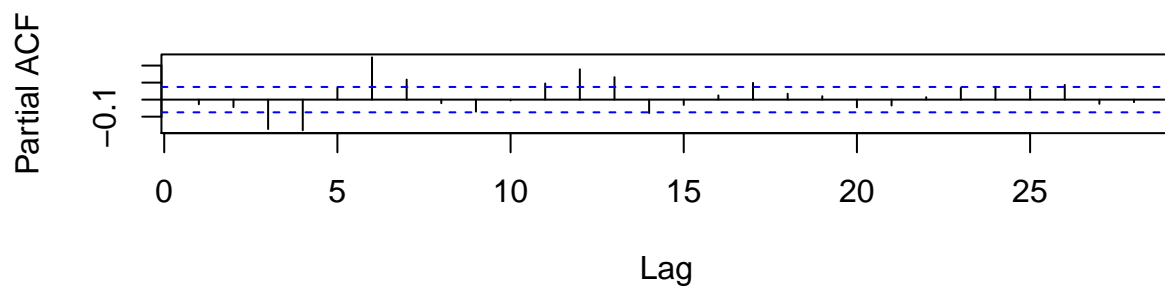
Series residuals(v1)[, 1]



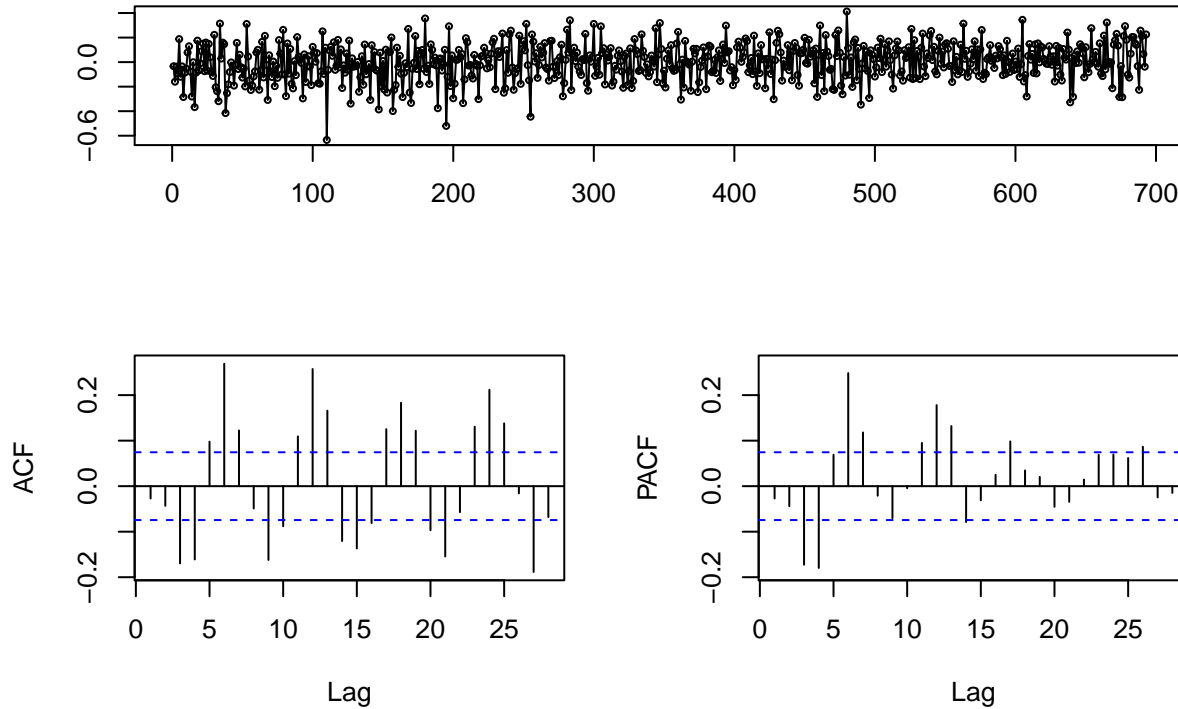
y=Sea Level



Series residuals(v1)[, 2]



$$\text{sea levels} = \text{temperature}(t-k) + \text{sea levels}(t-k)$$



We fit a VAR model of order = 3 using the two variables, sea level and temperature. The summary of the model provides the following information:

Sea Level is dependent on temperature at lag = 1 at the 99% confidence level Sea level is dependent on temperature at lag = 3 at the 90% confidence level Sea level is dependent on itself at lag = 1 and 3 at the 99% confidence level.

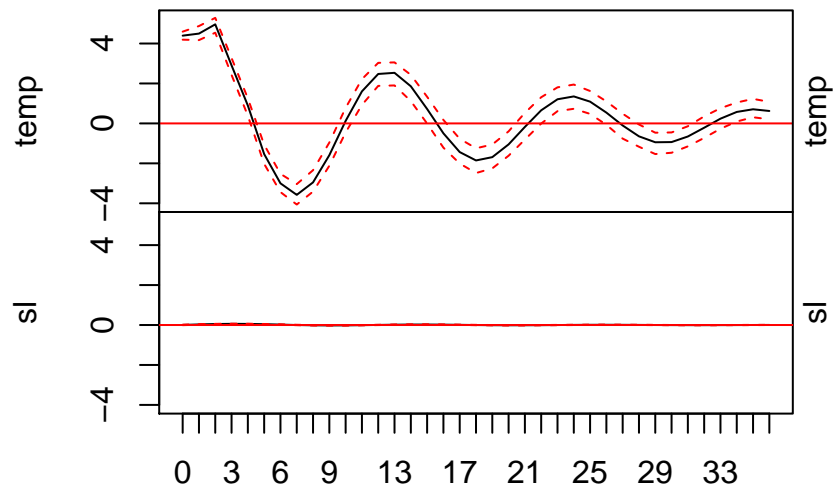
Strangely enough, the model suggests stronger causality in the opposite direction. However, intuitively we know that this cannot be the case. (Temperature causes sea level rise and not the other way round.)

Examining the ACF and PACF plots of the VAR model with sea level as the dependent variable, we see that they do not resemble the ACF and PACF plots synonymous with white noise. This means that the VAR model of order = 3 does not entirely capture all of the underlying dynamics within the series. It suggests that the model can be improved further, perhaps by including seasonal components.

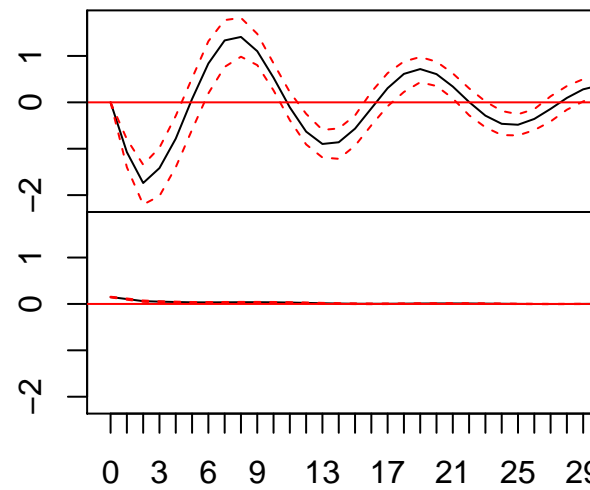
(j) Impulse Response Functions

The plot of the Impulse Response Function shows that an increase in temperature causes a change in the sea level. While these changes may be of an extremely small size, they do still occur. The graph shows that a 1 unit increase in temperature will lead to a less than 1 unit increase in the sea level. It will then fall slightly below the original level before it rises back up again. It will ultimately stabilize around its original value roughly 3 years after the initial impulse.

Orthogonal Impulse Response from temp

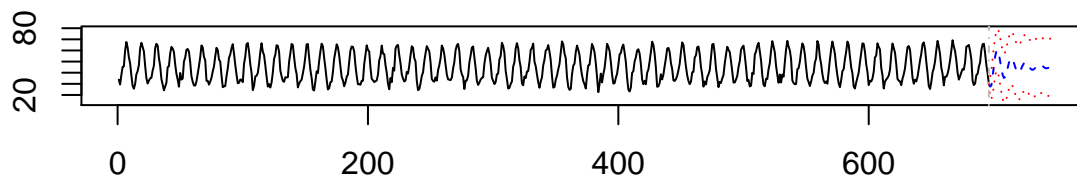


Orthogonal Impulse Response from sl



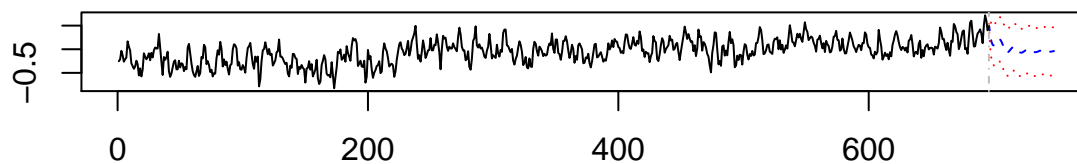
95 % Bootstrap CI, 100 runs

Forecast of series temp



95 % Bootstrap CI, 100 runs

Forecast of series sl



(k) Granger-Causality

```
## Granger causality test
##
## Model 1: sl ~ Lags(sl, 1:3) + Lags(tmp, 1:3)
## Model 2: sl ~ Lags(sl, 1:3)
##   Res.Df Df       F    Pr(>F)
## 1      687
```

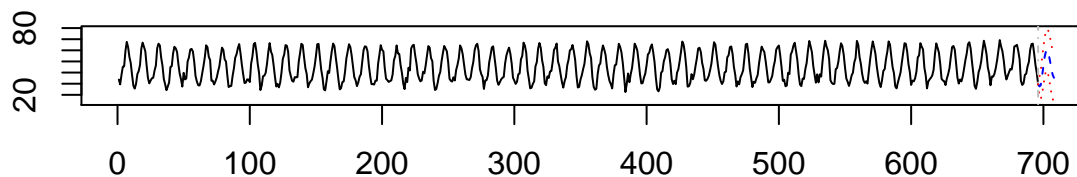
```
## 2      690 -3 43.923 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Granger causality test
##
## Model 1: tmp ~ Lags(tmp, 1:3) + Lags(sl, 1:3)
## Model 2: tmp ~ Lags(tmp, 1:3)
##   Res.Df Df       F    Pr(>F)
## 1      687
## 2      690 -3 21.499 2.559e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

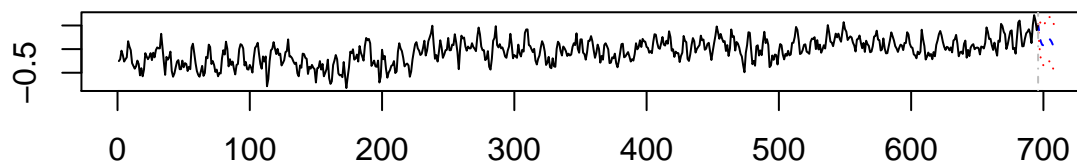
The Granger causality test reveals that temperature Granger causes sea level. Surprisingly, the opposite is also true. This seems slightly counter-intuitive and suggests that the two are part of a negative feedback loop. However, we must keep in mind that sea level and temperature are just two variables in the global climate system. Each of these is likely affected by several other variables that determine their values and general trend. In order to get a more accurate understanding of causality these variables need to be examined in a far wider context.

(1) VAR Forecast

Forecast of series tmp



Forecast of series sl



```
## [1] 0.978284
```

```
## [1] 1.238948
```

We can see from the plots of these models' forecasts that they are predicting almost opposite trends over the course of the next year. It comes as no surprise that the ARIMA model is continuing the recent trend of the data because the AR(3) component introduces considerable persistence. Based on the mean squared errors of these forecasts (compared to observed data), the ARIMA model is a superior forecasting tool.

III: Conclusions

After encountering dual Granger-Causality, our most important conclusion is that it is exceedingly difficult to reduce climactic phenomena to two variables. If we wish to further study climactic causality we must include more variables and study them in the context of the global (or a local) climatic ecosystem.

Breaking down the chosen time series step-by-step revealed a lot about the underlying processes that drive it. We were pleased with our ability decompose the timeseries and successfully capture trend, seasonality and cyclical dynamics. Though the VAR analysis was somewhat underwhelming, dual Granger-causality is more gratifying than none.

This concluded our work with meteorological data thus far. It has revealed the many difficulties that arise with data analysis in this field. The numerous and continuously changing processes that affect environmental phenomena are difficult to predict in today's world. This was our first attempt at doing so. This project has brought to light certain problems that we must learn to deal with if we are to accurately use data analysis as a tool for implementing environmental policies in the future.

IV: References

- (1) "Divisional Data Select". National Oceanic and Atmospheric Administration. www7.ncdc.noaa.gov. Web.
- (2) "Water Levels - NOAA Tides & Currents". tidesandcurrents.noaa.gov. Web.

V: Source Code

```
library(vars)
library(forecast)

rm(list=ls())

setwd('/Users/ian/Desktop/project2')

# ----- Setup -----
df <- read.csv('temp_co2_sl.csv')
df <- df[1:696,]

# ----- Training Data -----

var.df <- data.frame(temp=df$temp,sl=df$sl)

sl <- ts(df$sl,1959,2016,frequency = 12)
co2 <- ts(df$co2,1959,2016,frequency=12)
tmp <- ts(df$temp,1959,2016,frequency=12)
t <- ts(df$date,1959,2016,frequency=12)

# ----- Trend -----
par(mfrow=c(3,1))
plot(sl,col='seagreen',type='l', main='Mean Sea Level', ylab='Sea Level (ft)', xlab='Time')
acf(sl, main='Mean Sea Level Autocorrelation')
pacf(sl, main='Mean Sea Level Partial Autocorrelation')
```

```

fit1 <- lm(sl~t)
summary(fit1)

fit2 <- lm(sl~t + I(t^2)) # No statistically significant coefficients
summary(fit2)

# ----- Seasonality -----
m1 <- tslm(sl ~ t + season)
summary(m1)

# ----- Cycles -----
tsdisplay(m1$residuals, main='Trend + Seasonal Model Residuals') # AR(3) / AR(4) / AR(5)
plot(m1$fitted.values,m1$residuals,col='skyblue',main='Trend + Seasonal Model',xlab='Fitted Values',ylab='Residuals')

tsc1 <- arima(sl, order = c(3,0,0),seasonal=list(order=c(2,0,0)),xreg=t) # need to learn how to implement
tsc2 <- arima(sl, order = c(5,0,0),seasonal=list(order=c(2,0,0)),xreg=t) # need to learn how to implement

summary(tsc1)
summary(tsc2)
BIC(tsc1)
BIC(tsc2) # AIC says tsc2 is best, s2 and BIC says tsc1

model <- tsc1

Box.test(model$residuals, type = "Ljung-Box")
Box.test(model$residuals, type = "Box-Pierce") # NOT white noise

tsdisplay(model$residuals,lag.max=1000)
tsdisplay(auto.arima(sl)$residuals,lag.max=1000) # even auto.arima has similar dynamics

# ----- Trend + Seasonality + Cycles Evaluation -----

# (c) Residuals vs. Fitted Values
plot(fitted(model), model$residuals, col='skyblue', main='Model: Trend + Seasons + Cycles', xlab='Fitted Values', ylab='Residuals')

# (d) Auto and Partial Autocorrelation Plots
par(mfrow=c(2,1))
acf(model$residuals, main='Model Residuals Auto Correlation')
pacf(model$residuals, main='Model Residuals Partial Auto Correlation')

# (e) Cumulative Sum Plot
plot(efp(model$res~1, type = "Rec-CUSUM"),main='Model: Recursive CUSUM Test')

# (f) Recursive Residuals
y= recresid(model$residuals~1)
plot(y, pch=16, ylab="Recursive Residuals", main='Model: Recursive Residuals')

# (g) Diagnostic Statistics
accuracy(model)

# ----- TSC Training Data Eval -----

t.sl <- ts(sl[1:684],1959,2015,frequency = 12)
t.co2 <- ts(co2[1:684],1959,2015,frequency=12)

```

```

t.tmp <- ts(tmp[1:684],1959,2015,frequency=12)
t.t <- ts(t[1:684],1959,2015,frequency=12)

t.model <- arima(t.sl, order = c(3,1,0),seasonal=list(order=c(2,0,0))) # need to learn how to implement
h <- df$sl[685:696]

# (h) 12-Step Ahead Forecast
plot(forecast(t.model,h=12))
fcast <- forecast(t.model,h=12)
fcast <- fcast[[4]]
tsc.err <- h - fcast

# ----- VAR fitting -----
#(i) Var Model
ccf(sl, tmp) # regular seasonal patterns - very gradual decay

v1 = VAR(var.df,p=3)
summary(v1)

plot(v1)

par(mfrow=c(2,1))
acf(residuals(v1)[,1],main='y=Temperature')
pacf(residuals(v1)[,1])

par(mfrow=c(2,1))
acf(residuals(v1)[,2],main='y=Sea Level')
pacf(residuals(v1)[,2])

tsdisplay(residuals(v1)[,2],main ="sea levels = temperature(t-k) + sea levels(t-k)")

#(j) Impulse Response Functions
irf(v1)
plot(irf(v1, n.ahead=36))

#(k) Granger-Causality
grangertest(sl ~ tmp, order = 4)
grangertest(tmp ~ sl, order = 4)

#(l) VAR Forecast
training<- dim(var.df)[1]-12
t.vdf <- var.df[1:training,]

v1 = VAR(var.df,p=3)

var.predict = predict(object=v1, n.ahead=12)
plot(var.predict)

v.fcast <- predict(object=v1, n.ahead=12)
sl.fcast <- v.fcast[[1]][2]
sl.fcast <- sl.fcast[[1]][1:12]

```

```
v.err <- df$sl[685:696]
v.err <- v.err - sl.fcast

mean(sum(tsc.err^2)) # ARIMA error (MSE)

mean(sum(v.err^2)) #VAR error (MSE)
```