Contents

1. Executive Summary

Objective

For this time-series project, we will be analyzing multiple series of the Actuaries Climate Index, which is intended to provide a useful monitoring tool—an objective indicator of the frequency of extreme weather and the extent of sea level change. The ACI is available for the United States and Canada and 12 subregions thereof, but the focus of this project will be the combined value of U.S. and Canada. At first glance, the index seems to infer progressive climate change in North America. But we are interested whether we can make accurate statistical inference by using time series methodology.

For more information, see the home page[1], and data download[2]. The ACI is sponsored by the American Academy of Actuaries, Canadian Institute of Actuaries, Casualty Actuarial Society, and Society of Actuaries.

Data Overview

A total of 239 data points are available since 1961, with the last data point in the period June – August 2020. The values are standardized with the mean and standard deviation from the reference period, 1961 – 1990. The dataset contains 7 series in total:

CDD

Drought – Maximum number of consecutive dry days (daily precipitation less than 1mm) per year, with linear interpolation between yearly values to approximate monthly and seasonal values

Rx5Day

Heavy rain – Maximum 5-day rainfall in the month or season

Sea Level

Sea level – Sea level measurements are on a monthly basis via tide gauges located at 76 stations with reliable time series. The tide gauges measure sea level relative to the land below, but since the land is moving in many places, this component measures the combined effect on coastal shorelines of land movements and sea level changes.

T10

Low temperatures – Frequency of daily temperatures below the 10th percentile

T90

High temperatures – Frequency of daily temperatures above 90th percentile

WP90

High wind – Frequency of daily mean wind speeds above the 90th percentile, as measured by wind power, which has been shown to be proportional to wind damages. Wind power is defined as:

[1] https://actuariesclimateindex.org/about/
[2] https://actuariesclimateindex.org/data/

(1/2)ρw3, where w is the daily mean wind speed and ρ is the air density (taken to be constant at 1.23 kg/m3).

ACI

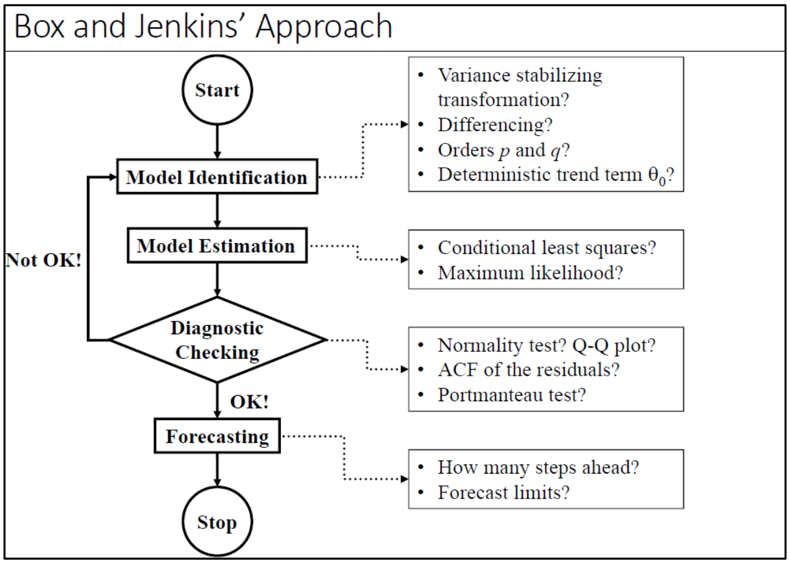
ACI = **mean**($T90_{std} - T10_{std} + P_{std} + D_{std} + W_{std} + S_{std}$)

Where 'std' stands for standardized anomaly, which represents the deviation from the mean in terms of standard deviation.

For the time being, only 3 datasets (CDD, Rx5Day, Sea Level) will be analyzed.

Methodology

The Box and Jenkin's approach will be used in general. In one case we automated the BJ apporach using grid search for an ARMA(p, q) model.



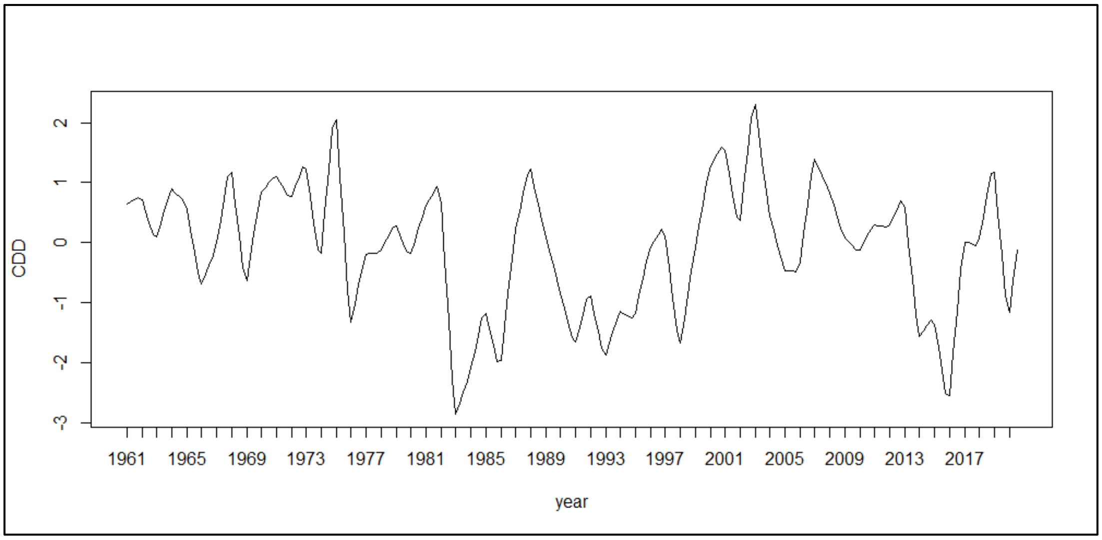
Findings

	Concluding Description	Applied Models
Consecutive Dry Days	has periods of high and low volatility, but in general it is stationary	Series: MA(10) Residuals: ARCH(5)
Maximum 5-Day Rainfall	a random walk with a small drift component	ARIMA(0, 1, 1)
Sea Level	Deterministic along time since year 2000	Series: Deterministic Trend Residuals: ARMA(1, 7)

The most worrying finding among the 3 is the rising sea level, we expect the sea level to be higher by 7 times the standard deviation in the reference period 1961 – 1990.

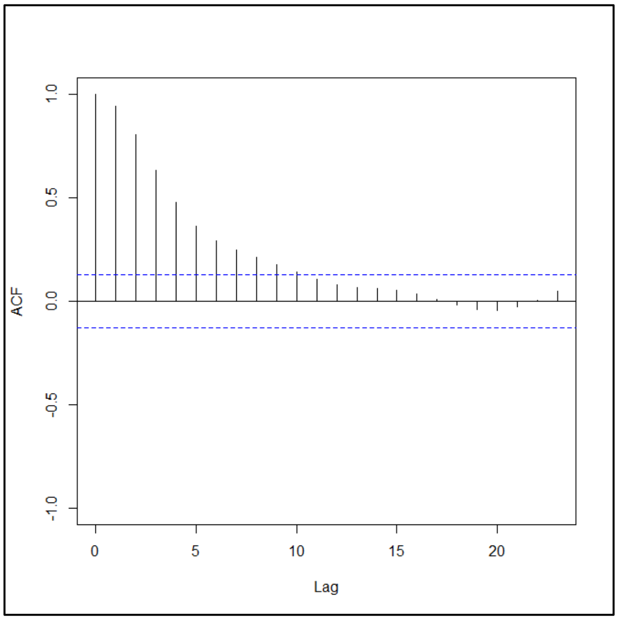
2. Consecutive Dry Days

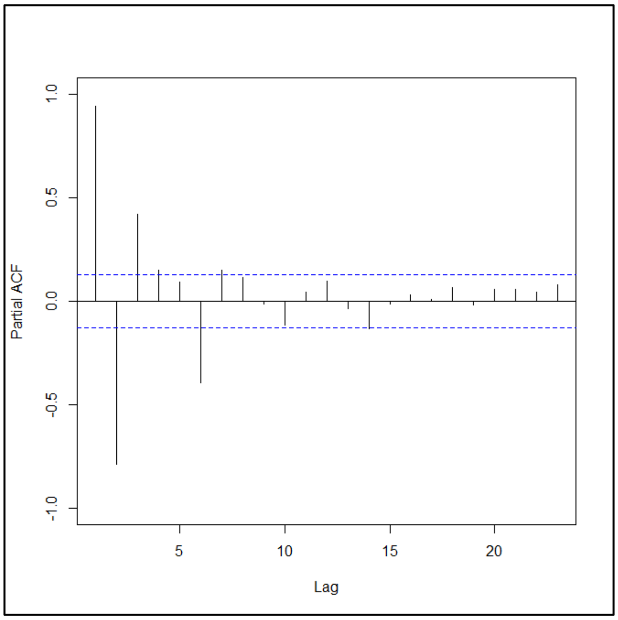
Overview



The data looks roughly stationary, does not warrant differencing.

Model Identification





Observation	ACF is decaying / tail-off, while PACF shows alternating pattern.
Interpretation	Possible candidates are Random Walk model or Pure Moving Average model. However, considering that ACF is vanishing at LAG 18, a random walk is unlikely. Hence, we should consider an MA(10) model.

Model Estimation

Fitting an MA(10) model, we see that the parameters are significantly different from zero, except the last three.

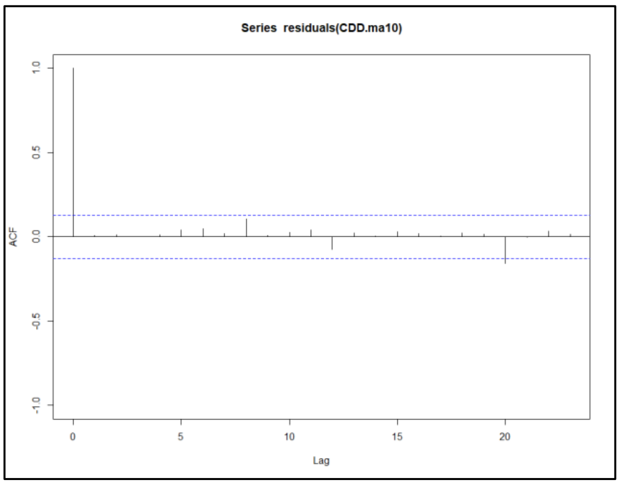
```
arima(x = data$CDD, order = c(0, 0, 10))

Coefficients:
         ma1     ma2     ma3     ma4     ma5     ma6     ma7     ma8     ma9    ma10  intercept
      2.3739  3.7266  4.9788  4.5610  3.3928  2.1958  1.0258  0.4513  0.2379  0.1083    -0.0551
s.e.  0.0666  0.1740  0.3085  0.4487  0.5158  0.5077  0.4268  0.2864  0.1647  0.0633     0.1641

sigma^2 estimated as 0.01147:  log likelihood = 183.05,  aic = -342.11
```

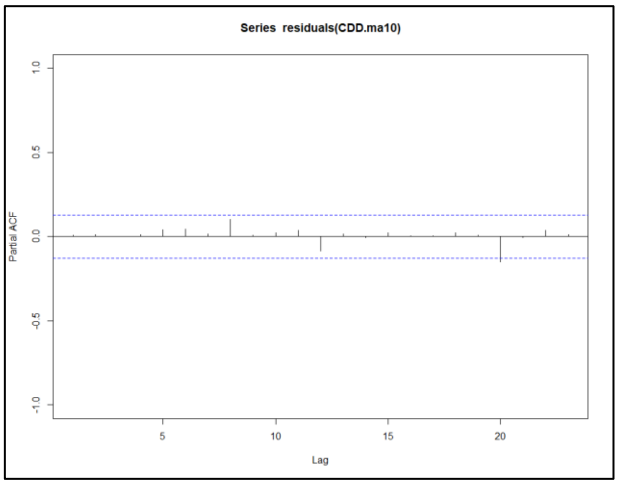
Diagnostic Checking

Selection Criteria:

```
> AIC(CDD.ma10)
[1] -342.1066

> BIC(CDD.ma10)
[1] -300.389
```

A plot of the residuals shows clean ACF and PACF except at lag 20.





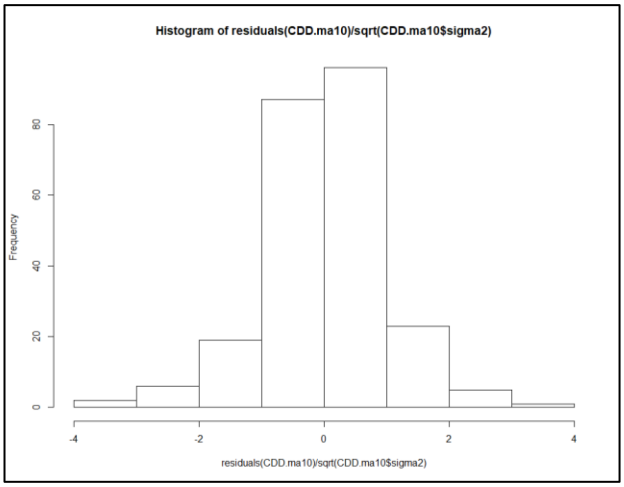
We should test the model adequacy against the joint null hypotheses,

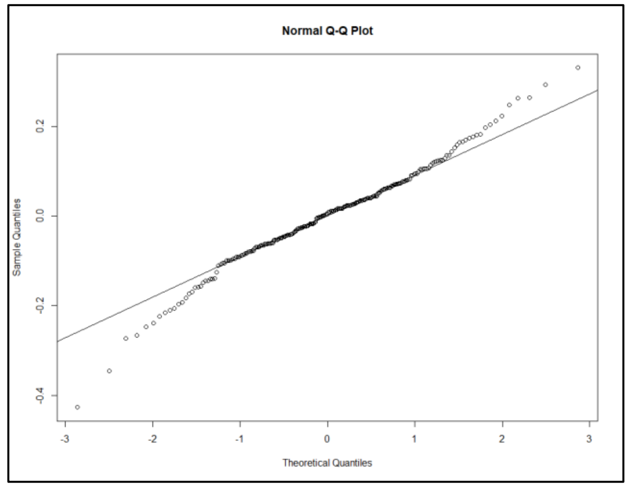
$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_K = 0$$

```
        Box-Pierce test

data:  residuals(CDD.ma10)
X-squared = 12.706, df = 12, p-value = 0.3907
```

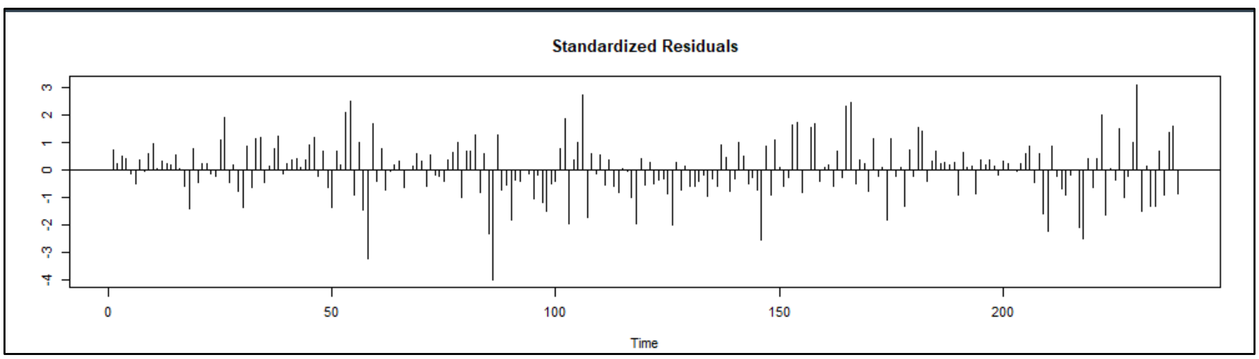
```
        Box-Ljung test

data:  residuals(CDD.ma10)
X-squared = 13.665, df = 12, p-value = 0.3226
```

It seems the model fits reasonably well, and based on the p-value we will accept (not reject) the MA(10) model at 5% significance.



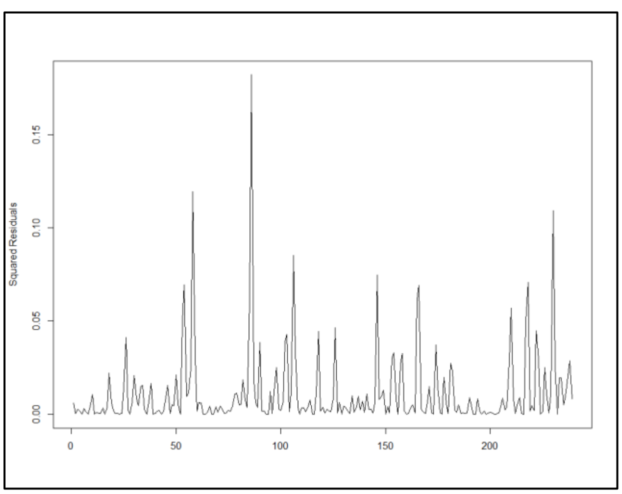


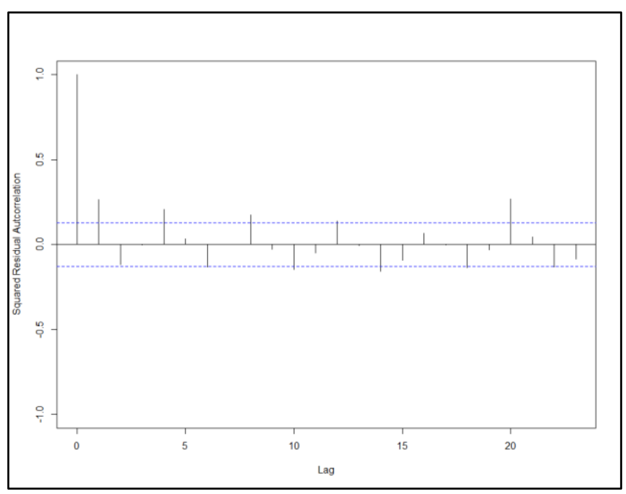
The qq-plot shows that the residuals are heavy-tailed on both ends. The normality assumption seems questionable, a “leptokurtic” distribution better desribes the residuals.



The plot of standardized residuals shows that it may not have a constant variance. (for example compare magnitude for time < 50 versus time >= 50)

Conditional Heteroscedasticity





We see periods of higher and lower spikes among the squared residuals. And with a plot of the squared residual autocorrelation, we see more than a few significant spikes. We can see that the ACF is quite significant even at lag = 20. A GARCH may better serve the estimation but due to technical limit we resort to an ARCH(5) model.

```
lm(formula = y ~ x1 + x2 + x3 + x4 + x5)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.039267 -0.008699 -0.005166  0.002437  0.150554 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.008368   0.001872   4.471 1.23e-05 ***
x1           0.335221   0.065966   5.082 7.79e-07 ***
x2          -0.192674   0.068369  -2.818  0.00525 ** 
x3           0.024504   0.069521   0.352  0.72481    
x4           0.203185   0.068362   2.972  0.00327 ** 
x5          -0.086020   0.065996  -1.303  0.19375    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01995 on 228 degrees of freedom
Multiple R-squared:  0.1511,	Adjusted R-squared:  0.1325 
F-statistic: 8.115 on 5 and 228 DF,  p-value: 4.574e-07
```

From the estimation we see lag = 1, 2, 4 are signifiant. Consistent with what we have seen in the ACF plot. Below is the Lagrange-Multiplier test statistic.

```
> (LM.stat <- 239*summary(lm(y~x1+x2+x3+x4+x5))$r.squared)
[1] 36.1063

> qchisq(0.95,5)
[1] 11.0705
```

We can see the statistc exceed theoretical value at 5% significane. A Portmanteau test shows the same.

```
	Box-Ljung test

data:  error.2
X-squared = 49.03, df = 10, p-value = 4.021e-07
```

Forecast

There are two parts to forecast: (1) The variance of future data (2) The value of future data.

(1) ARCH(5) Variance forecast (forecast origin = 2020-3, developed recursively)

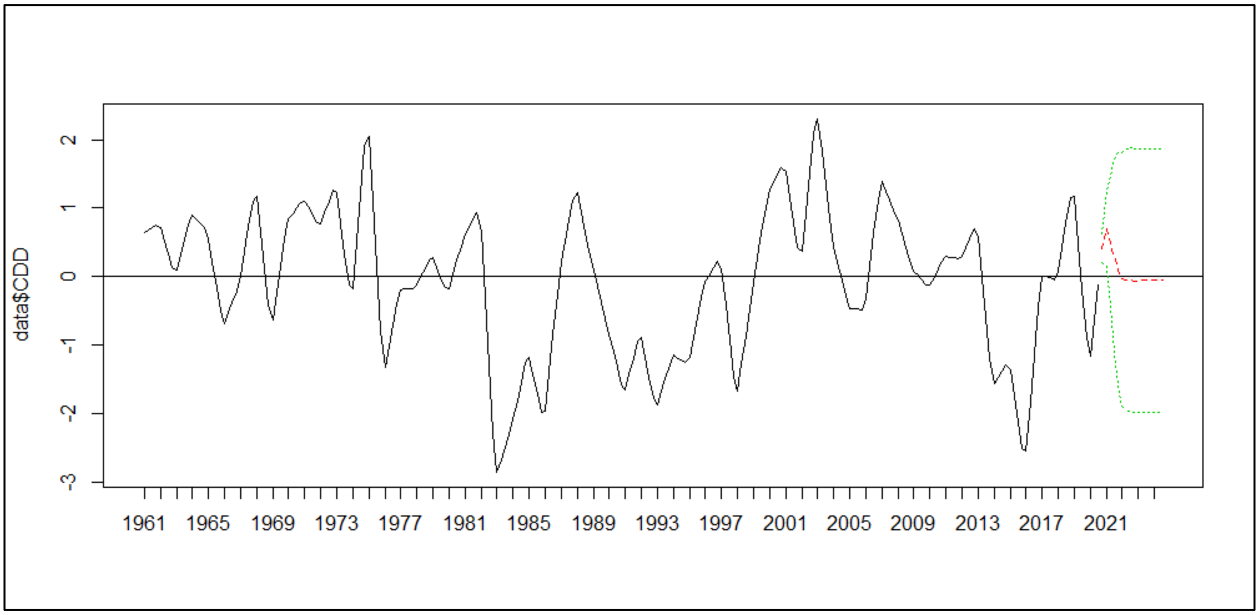
2019-4	2020-1	2020-2	2020-3	2020-4	2021-1	2021-2	2021-3
0.009814	0.020684	0.028782	0.008438	0.007645	0.013508	0.017271	0.01327

Based on linear regressive model specified as

x4	x3	x2	x1	intercept
0.203185	0	-0.19267	0.335221	0.008368

Based on the forecast, we can expect the squared residuals to increase in the coming quarters. In other words, the volatility in the data will increase. Calculation made with Excel.

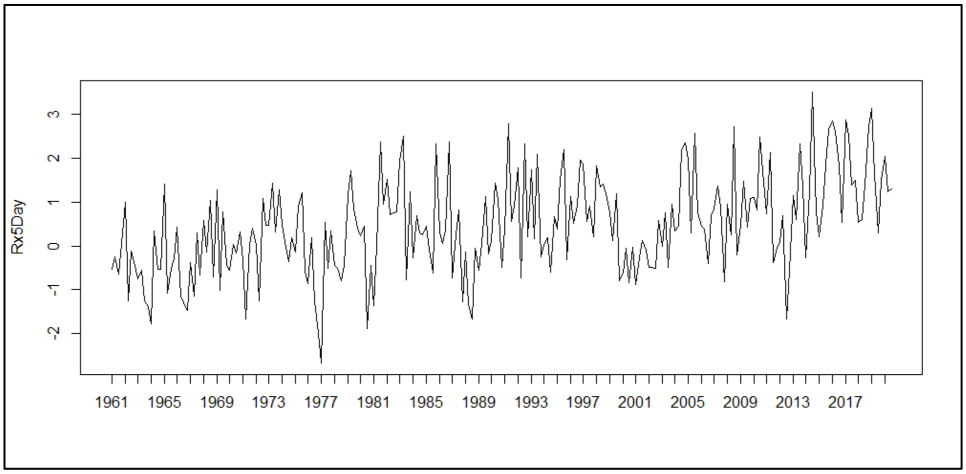
(2) MA(10) forecast



As can be seen from the red dotted line, we expect a rise in consecutive dry days into 2021, before going back to normal level. The green dotted line (5% boundaries) stops expanding after 2022.

3. Maximum 5-day rainfall

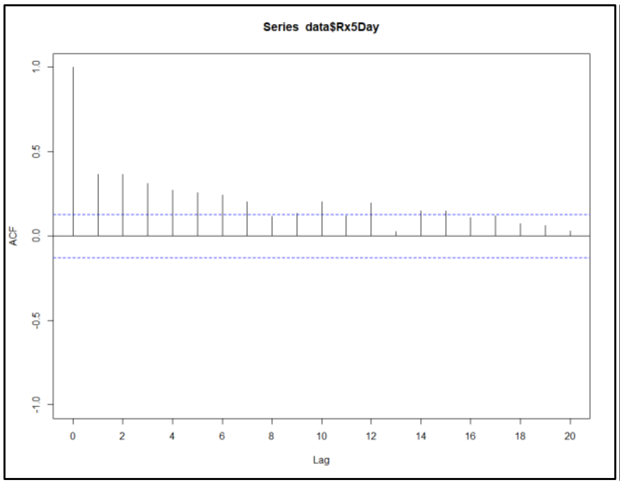
Overview

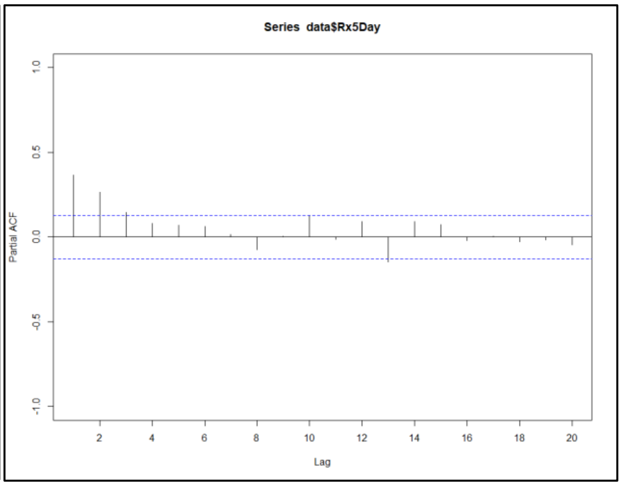


At first sight, there seems to be a weak trend, the data is more volatile and hence may subject to seasonality unlike the previous series. We will see on differenced series if there is seasonality with s = 4. We should perform simple differencing due to the trend.

Model Identification

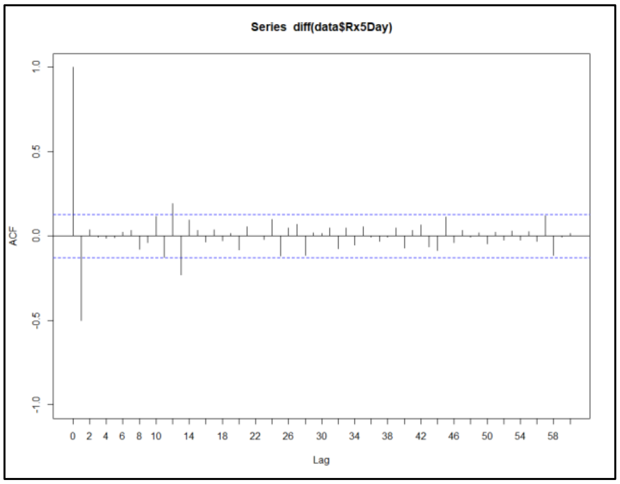
The ACF and PACF of the non-differenced series are as follow.

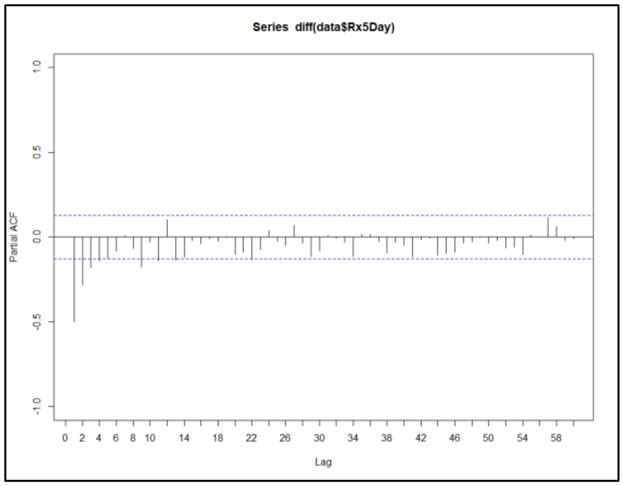




Observation	From the ACF and PACF of the original data, we do not see any spike indicating a seasonal component. PACF reaches zero more quickly than ACF. We may see this as ACF tailing off and PACF cutoff.
Interpretation	Possible candidate is AR(6).

The ACF and PACF of the differenced series are as follow.





Observation	The differencing on the series seems to have eliminated much of the autocorrelation. For ACF, there are no spikes at lag multiples of 4, indicating no year-on-year seasonality. There is a negative spike at lag = 1. While the PACF tails off.
Interpretation	Possible candidate is ARIMA(0, 1, 1).

Model Estimation

Here two models will be fitted: AR(6) and ARIMA(0, 1, 1).

For AR(6),

```
arima(x = data$Rx5Day, order = c(6, 0, 0))

Coefficients:
         ar1     ar2     ar3     ar4     ar5     ar6  intercept
      0.2047  0.1941  0.1106  0.0574  0.0550  0.0678     0.4472
s.e.  0.0644  0.0655  0.0667  0.0667  0.0659  0.0649     0.1977

sigma^2 estimated as 0.9441:  log likelihood = -332.47,  aic = 680.94
```

For ARIMA(0, 1, 1),

```
arima(x = data$Rx5Day, order = c(0, 1, 1))

Coefficients:
          ma1
      -0.8139
s.e.   0.0586

sigma^2 estimated as 0.9697:  log likelihood = -334.59,  aic = 673.17
```

Diagnostic Checking

For AR(6),

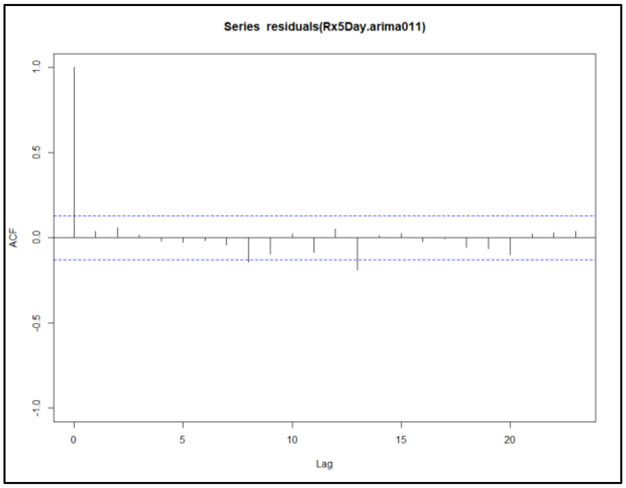
```
> AIC(Rx5Day.ar6)
[1] 680.9438

> BIC(Rx5Day.ar6)
[1] 708.7555
```

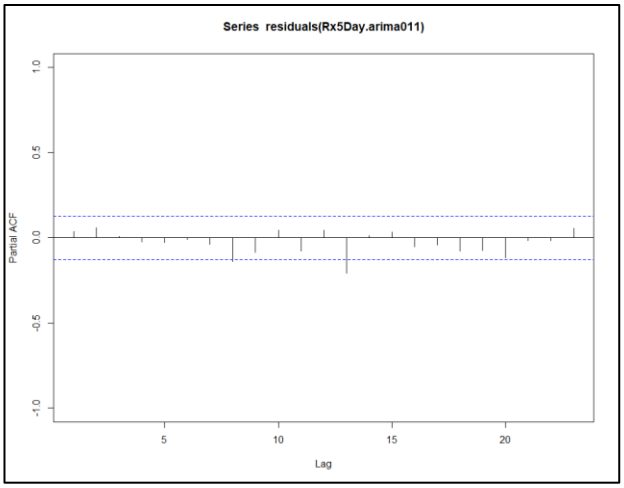
For ARIM(0, 1, 1),

```
> AIC(Rx5Day.arima011)
[1] 673.1707

> BIC(Rx5Day.arima011)
[1] 680.1152
```

They have comparable scores. Based on the criteria, we will choose ARIMA(0, 1, 1) over AR(6). The remaining work does not consider AR(6).

The ACF and PACF for ARIMA(0, 1, 1) are as follow.





The ACF and PACF seems clean except at lag = 8 and lag = 13.

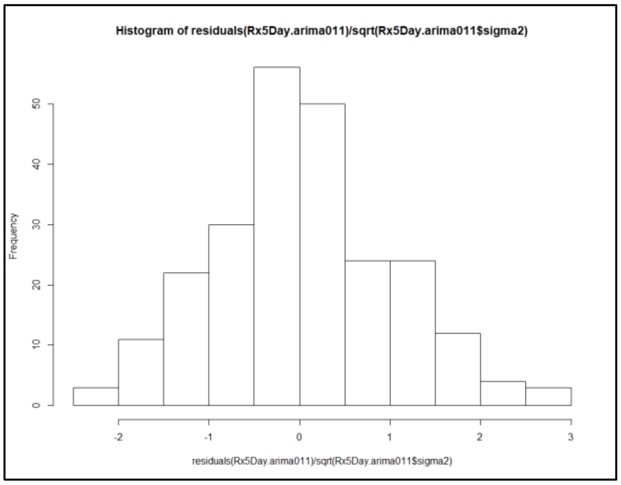
We should test the model adequacy against the joint null hypotheses,

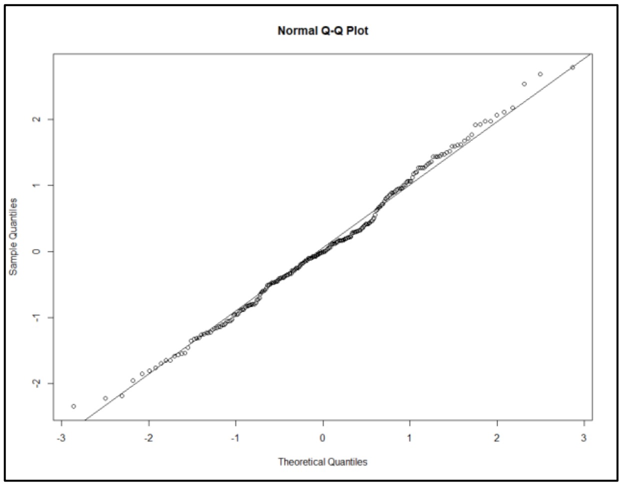
$$H_0 : \rho_1 = \rho_2 = \ldots = \rho_K = 0$$

```
        Box-Pierce test

data:  residuals(Rx5Day.arima011)
X-squared = 24.856, df = 19, p-value = 0.1654
```

```
        Box-Ljung test

data:  residuals(Rx5Day.arima011)
X-squared = 26.39, df = 19, p-value = 0.1197
```

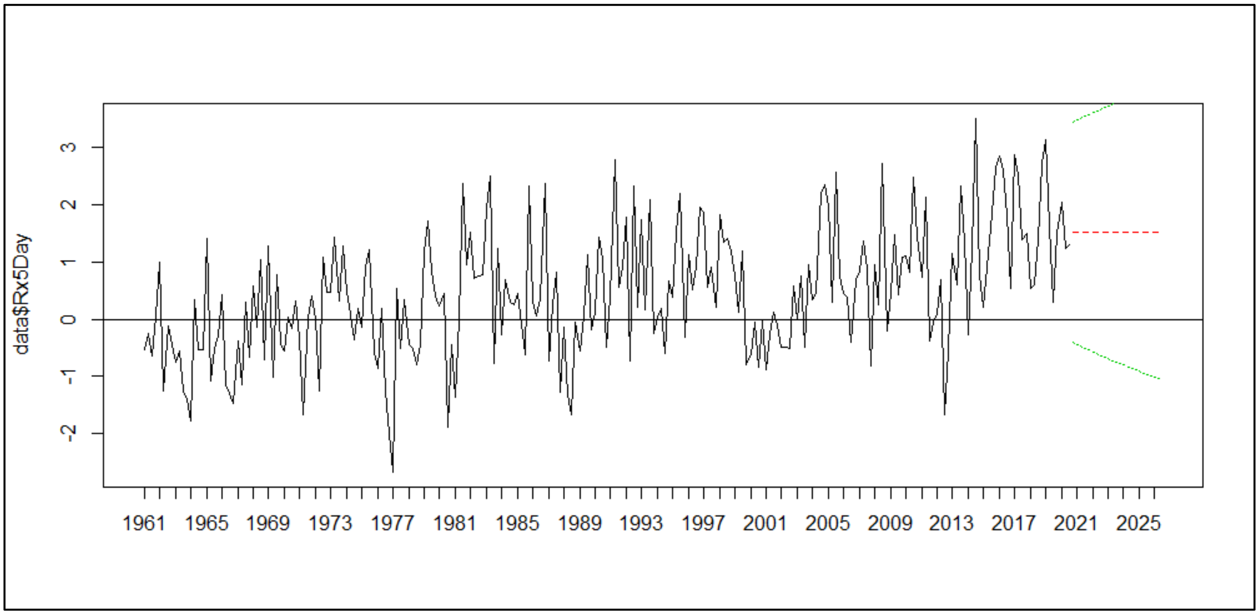
It seems the model fits reasonably well, and based on the p-value we will accept (not reject) the ARIMA(0, 1, 1) model at 5% significance.





The histogram and Q-Q plot show reasonably normal pattern.

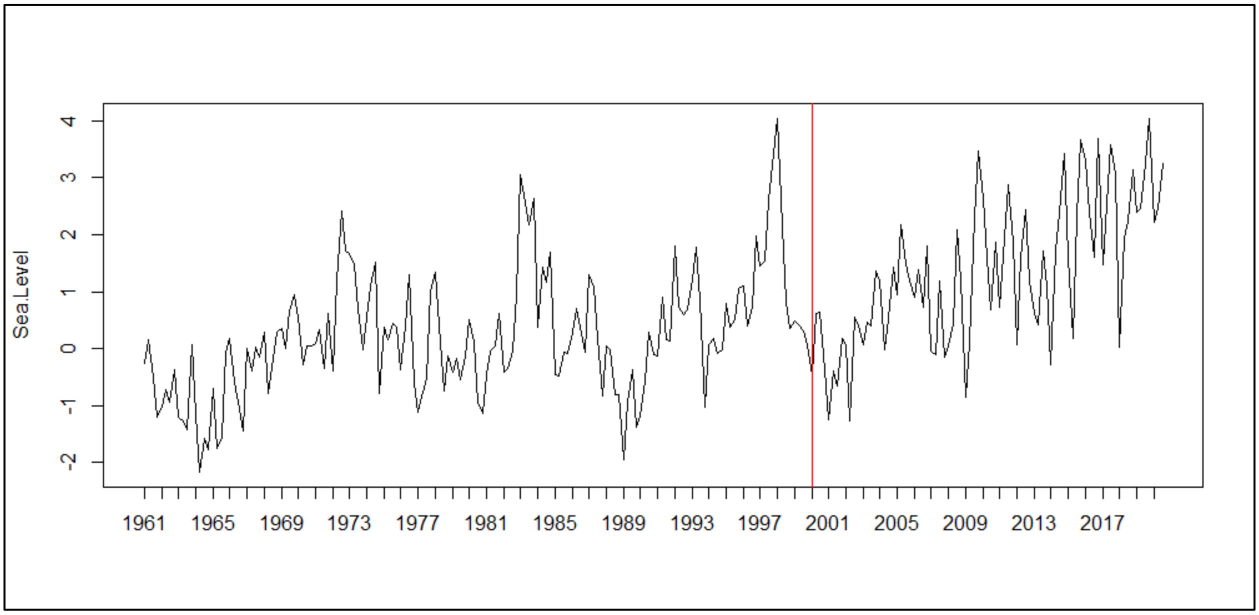
Forecast



We can see that the 6-year forecast has a very slow uptrend, with two expanding 5% boundaries.

4. Sea Level

Overview



There seems to be a trend in rising sea level since year 2000. Hence, for this time series, we will only consider the later part which is the range 2000 – 2020. (83 data points)

Unlike previous pages, we are going to investigate whether the movement is a random walk with drift or a deterministic trend. For deterministic trend, we cannot use ARIMA model.

Model Identification, Estimation and Checking

This section involves iterative procedures, and hence we do not divide into three sections.

Unit Root tests are performed with results presented below.

(1) Dickey-Fuller Test with Trend

```
lm(formula = Y.diff ~ Y.lag1 + TIME)

Residuals:
     Min       1Q   Median       3Q      Max 
-2.56838 -0.67283  0.05085  0.62375  1.96938 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.106966   0.210241  -0.509    0.612    
Y.lag1      -0.782748   0.109897  -7.123 4.35e-10 ***
TIME         0.027893   0.005807   4.804 7.28e-06 ***
```

(2) Augmented Dickey-Fuller Test with Trend and p = 1

```
lm(formula = Y.diff ~ Y.lag1 + TIME + Y.diff.lag1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2632 -0.5485  0.1635  0.7251  1.7287

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.206463   0.207511  -0.995  0.32288
Y.lag1      -1.018561   0.135084  -7.540 7.77e-11 ***
TIME         0.037087   0.006374   5.818 1.29e-07 ***
Y.diff.lag1  0.307242   0.108115   2.842  0.00574 **
```

(3) Augmented Dickey-Fuller Test with Trend and p = 2

```
lm(formula = Y.diff ~ Y.lag1 + TIME + Y.diff.lag1 + Y.diff.lag2)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2722 -0.5403  0.1785  0.7104  1.7458

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.218808   0.217590  -1.006   0.3178
Y.lag1      -0.983516   0.180099  -5.461 5.90e-07 ***
TIME         0.036206   0.007789   4.648 1.41e-05 ***
Y.diff.lag1  0.278261   0.136866   2.033   0.0456 *
Y.diff.lag2 -0.036115   0.114993  -0.314   0.7543
```

Dicky-Fuller tests with Trend		
Lag, p	t_{DF}	10% Critical Value
-	-7.123	-3.228
1	-7.540	-3.230
2	-5.461	-3.233

Based on the numbers, we reject the H_0 that the series is a random walk with drift. The series will be estimated as a deterministic trend model, which means that the ARIMA model is not suitable.

In light of the deterministic trend, we will first (1) conduct a linear regression, and (2) check on the residual components.

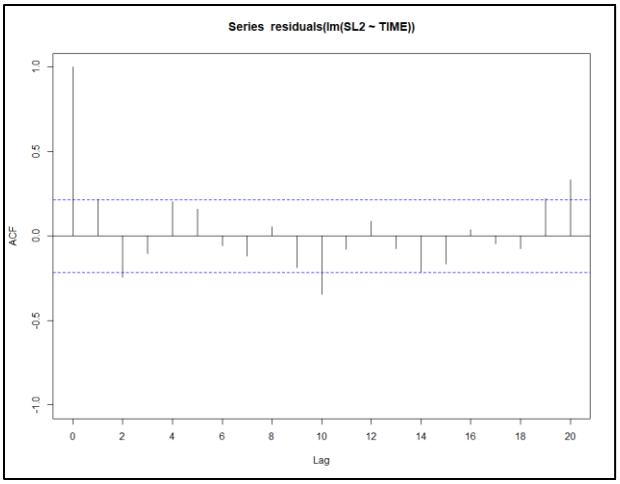
(1) The model is fit with the time component being significant.

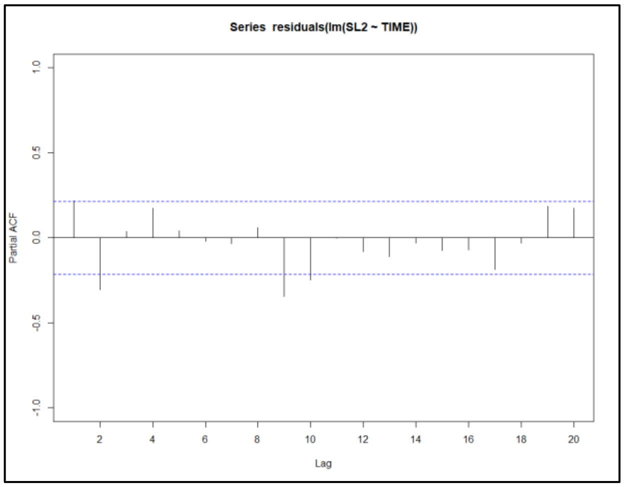
```
lm(formula = SL2 ~ TIME)

Residuals:
     Min       1Q   Median       3Q      Max
-2.42633 -0.58412  0.05754  0.63541  2.20856

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.164787   0.206761  -0.797    0.428
TIME         0.035906   0.004276   8.397 1.24e-12 ***
```

(2) Portmanteau test (assuming we did not fit any parameter) suggests the residuals are not white noise.

```
        Box-Pierce test

data:  residuals(lm(SL2 ~ TIME))
X-squared = 54.5, df = 22, p-value = 0.0001398
```

ACF and PACF of the residuals are as follow.





Observation	Both ACF and PACF are non-vanishing
Interpretation	We will fit an ARMA(p, q) model using a grid search.

AIC Table (only models passing Portmanteau tests are shown)

p \ q	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1							216.07	217.22	219.18	219.87
2							217.48	219.21		221.87
3							218.97		215.86	
4							218.52	219.72	217.19	223.22
5							219.74	221.71		
6										
7							212.81			
8										
9	212.97	214.96		215.92	214.14		217.23			
10	214.95	216.84	217.86	216.09	215.96					

BIC Table (only models passing Portmanteau tests are shown)

p \ q	1	2	3	4	5	6	7	8	9	10
1							240.26	243.83	248.20	251.31
2							244.08	248.23		255.73
3							248.00		249.73	
4							249.96	253.58	253.47	261.92
5							253.60	257.99		
6										
7							251.51			
8										
9	241.99	246.41		252.20	252.84		260.77			
10	246.40	250.70	254.15	254.79	257.08					

It seems that ARMA(7, 7) or ARMA(1, 7) will fit best. See summary as follows.

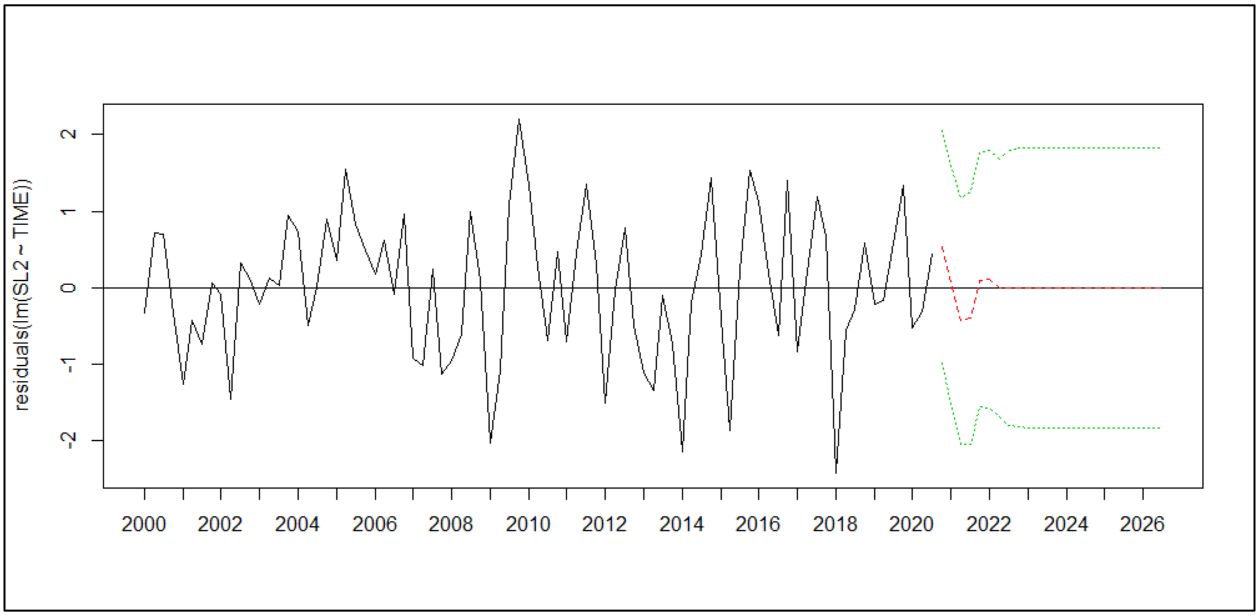
	ARMA(7, 7)	ARMA(1, 7)
AIC	212.8094	216.0677
BIC	251.5109	240.2561
Box-Pierce p-value	0.167	0.3431
Box-Ljung p-value	0.06783	0.1389

Overall, we'll choose ARMA(1, 7). Summary of model as follows.

```
arima(x = residuals(lm(SL2 ~ TIME)), order = c(1, 0, 7))

Coefficients:
         ar1      ma1      ma2      ma3     ma4     ma5      ma6      ma7  intercept
      0.5058  -0.2975  -0.3776  -0.1354  0.2168  0.1363  -0.1592  -0.3832    -0.0010
s.e.  0.1982   0.1953   0.1111   0.1267  0.1058  0.1149   0.1187   0.1362     0.0222

sigma^2 estimated as 0.5953:  log likelihood = -98.03,  aic = 216.07
```

Forecast

Since this is a deterministic trend, we will forecast longer into the future (2050) to see what would happen. First we see what happens in the short term (with the predicted residuals).



It seems the residual component will have a decline until mid-2021, followed by a reversion to the mean. After which it is unclear what would happen with the residual component.

For the deterministic component is merely a straight line, when we reach the year 2050, the sea level would be expected to go to (-0.164787) + (83 + 4 x 30) (0.035906) = 7.124131. Meaning an unprecedented increase in sea level.

5. Conclusion

Based on the analysis, we epxect more volatility in the consecutive dry days in North America in the remaining year. And it is likely the maximum consecutive dry days will be longer than last year. On the other hand, we neither predict an increase or decrease in the maximum 5-day rainfall. We predict a temporary decrease in sea level into mid-2021 but in the longrun we believe the sea level is on a steady and dangerous rise.

6. Appendix

Dataset

Year	Season	CDD	Rx5Day	Sea Level	T10	T90	WP90
1961	1	0.65	-0.53	-0.26	-1.25	0.16	-1.15
1961	2	0.68	-0.25	0.15	0.25	-0.87	0.58
1961	3	0.71	-0.64	-0.49	-0.85	0.49	-0.55
1961	4	0.75	0.06	-1.19	0.75	-0.05	2.16
1962	1	0.71	0.99	-1.03	1.09	-0.22	0.26
1962	2	0.52	-1.26	-0.72	0.39	-0.34	-1.35
1962	3	0.33	-0.12	-0.94	0.12	-1.36	-1.15
1962	4	0.14	-0.44	-0.37	-1.03	0.62	-1.2
1963	1	0.09	-0.75	-1.2	0.71	1.16	0.54
1963	2	0.31	-0.55	-1.26	-0.51	0.05	-0.11
1963	3	0.54	-1.26	-1.43	-0.41	-0.17	-0.19
1963	4	0.76	-1.38	0.08	-1.67	3.52	-1.9
1964	1	0.89	-1.78	-1.15	-0.31	-0.37	-0.37
1964	2	0.83	0.34	-2.17	1.66	-0.82	0.63
1964	3	0.78	-0.54	-1.57	0.88	-0.37	0.48
1964	4	0.72	-0.54	-1.77	0.02	-0.57	-0.75
1965	1	0.56	1.42	-0.7	0.91	-0.45	1.19
1965	2	0.21	-1.08	-1.74	1.62	-0.28	-1.01
1965	3	-0.15	-0.51	-1.54	1.47	-1.66	-0.57
1965	4	-0.5	-0.3	-0.06	0.45	-0.67	-0.16
1966	1	-0.69	0.42	0.19	0.21	-0.61	-2.8
1966	2	-0.53	-1.15	-0.58	1.1	-0.27	0.01
1966	3	-0.37	-1.3	-0.94	0.09	-0.48	-1.54
1966	4	-0.22	-1.47	-1.44	-0.19	-0.37	0.34
1967	1	0.01	-0.37	0	-0.64	-0.41	-0.17
1967	2	0.37	-1.15	-0.39	2.23	-0.47	0.82
1967	3	0.74	0.3	0.03	0.38	-0.57	-1.2
1967	4	1.1	-0.67	-0.15	-0.56	0.31	0.4
1968	1	1.17	0.59	0.3	-0.02	0.23	0.16
1968	2	0.64	-0.15	-0.79	-0.55	0.1	0.69
1968	3	0.11	1.03	-0.16	1.68	-0.82	1.65

1968	4	-0.42	-0.7	0.3	-1.01	-0.73	-0.22
1969	1	-0.64	1.28	0.36	0.65	-1.49	1.02
1969	2	-0.24	-1.01	0.01	0.15	-0.41	-1.82
1969	3	0.15	0.78	0.66	1.59	0.38	0.44
1969	4	0.55	-0.42	0.94	0.03	0.11	-0.84
1970	1	0.84	-0.55	0.56	-0.93	-0.07	-1.65
1970	2	0.92	0.03	-0.27	-0.12	-1	-1.02
1970	3	1	-0.17	0.06	-0.68	0.73	-0.68
1970	4	1.08	0.31	0.04	0.96	-0.17	0.35
1971	1	1.1	-0.31	0.1	0.46	-0.6	0.18
1971	2	1	-1.67	0.33	0.5	-1.58	0.77
1971	3	0.89	0.03	-0.34	0.39	-0.24	-1.98
1971	4	0.79	0.4	0.63	-0.07	0.05	-0.75
1972	1	0.78	0.02	-0.38	1.01	-0.73	1.21
1972	2	0.94	-1.25	1.22	0.98	-0.18	-1.48
1972	3	1.1	1.09	2.41	1.44	-0.57	-0.5
1972	4	1.26	0.47	1.72	1.3	-2.36	-0.21
1973	1	1.22	0.48	1.67	0.92	-0.32	-0.4
1973	2	0.78	1.43	1.49	-0.7	-0.32	0.92
1973	3	0.33	0.31	0.7	-0.16	-0.3	1.2
1973	4	-0.12	1.29	-0.01	-0.08	-0.14	0.8
1974	1	-0.18	0.5	0.56	-0.54	-0.44	0.21
1974	2	0.51	0.08	1.11	0.14	-0.5	1.68
1974	3	1.21	-0.36	1.51	0.67	-0.67	2.14
1974	4	1.9	0.18	-0.79	0.72	-0.69	-0.5
1975	1	2.05	-0.14	0.37	-0.68	-0.89	-0.55
1975	2	1.1	0.91	0.16	1.42	-1.69	1.5
1975	3	0.15	1.22	0.44	0.17	-0.39	1.42
1975	4	-0.8	-0.57	0.38	0.25	-0.09	-0.09
1976	1	-1.33	-0.85	-0.36	-0.64	0.65	-0.11
1976	2	-1.02	0.19	0.39	-0.28	-0.33	0.03
1976	3	-0.71	-1.25	1.29	1.5	-0.6	0.84
1976	4	-0.4	-2	-0.59	1.58	-0.59	-0.6
1977	1	-0.19	-2.68	-1.12	-0.18	-0.03	-0.96
1977	2	-0.18	0.53	-0.76	-1.29	1.84	0.3
1977	3	-0.18	-0.52	-0.51	-0.86	0.66	-1.34
1977	4	-0.17	0.34	1.01	-0.85	-0.14	-0.78
1978	1	-0.13	-0.45	1.35	0.45	-0.87	-0.07
1978	2	0	-0.52	0.4	-0.69	-0.39	-1.78
1978	3	0.12	-0.79	-0.73	-0.25	-0.16	-0.54
1978	4	0.24	-0.45	-0.13	-0.17	0.03	-0.96
1979	1	0.27	1.17	-0.4	2.72	-2.07	-0.35
1979	2	0.13	1.72	-0.16	-0.23	-0.37	-0.67

1979	3	-0.01	0.85	-0.54	0.24	-0.5	-0.49
1979	4	-0.16	0.41	-0.12	-1.23	1.98	-0.84
1980	1	-0.18	0.24	0.52	-0.97	0.61	1.02
1980	2	0.03	0.46	0.12	-1.04	0.35	0.06
1980	3	0.24	-1.89	-0.93	-1.09	1.06	0.89
1980	4	0.45	-0.44	-1.14	-0.6	0.44	0.03
1981	1	0.62	-1.37	-0.45	-0.58	3.27	0.71
1981	2	0.72	0.02	-0.04	-1.62	1.23	-1.11
1981	3	0.83	2.38	0.05	-0.83	0.27	0.08
1981	4	0.93	0.95	0.63	-0.87	0.27	1.02
1982	1	0.67	1.53	-0.42	0.55	-0.99	1.08
1982	2	-0.31	0.72	-0.35	0.83	-1.26	0.86
1982	3	-1.29	0.76	-0.03	0.48	-1.4	-0.52
1982	4	-2.27	0.78	0.95	0.28	-0.62	-0.48
1983	1	-2.86	1.94	3.06	-2.13	0.73	-0.68
1983	2	-2.67	2.5	2.58	0.69	-0.34	0.19
1983	3	-2.49	-0.77	2.18	-1.4	1.54	-0.73
1983	4	-2.3	1.24	2.64	-0.98	0.53	0.42
1984	1	-2.08	-0.27	0.37	0.79	0.48	-0.62
1984	2	-1.81	0.69	1.44	0.1	-0.13	-0.02
1984	3	-1.54	0.3	1.17	-1.33	-0.06	0.35
1984	4	-1.26	0.25	1.7	1.14	-0.53	1.29
1985	1	-1.17	0.44	-0.45	0.46	-0.03	0.03
1985	2	-1.44	-0.08	-0.47	-1.04	0.3	0.2
1985	3	-1.71	-0.62	-0.05	1.4	-0.73	0.19
1985	4	-1.98	2.33	-0.09	2.83	-1.11	1.95
1986	1	-1.97	0.32	0.3	-0.23	1.47	0.51
1986	2	-1.4	0.05	0.71	-0.64	1.37	0.13
1986	3	-0.83	0.37	0.38	-0.27	-0.47	0.62
1986	4	-0.26	2.37	-0.06	1.02	-0.35	-1.22
1987	1	0.22	-0.74	1.31	-1.84	1.13	-1.44
1987	2	0.52	0.26	1.08	-1.43	2.28	-1.32
1987	3	0.82	0.82	0.18	-0.45	1.08	0.83
1987	4	1.12	-1.27	-0.82	-1.2	0.55	-0.74
1988	1	1.22	-0.14	0.04	-1.08	-0.15	0.7
1988	2	0.94	-1.32	-0.02	-1.1	1.28	1.25
1988	3	0.65	-1.67	-0.8	-1.57	3.13	0.72
1988	4	0.37	-0.06	-0.81	-0.67	-0.63	0.95
1989	1	0.1	-0.55	-1.95	0.27	0.08	2.13
1989	2	-0.14	0.06	-0.94	0.2	0.92	-0.31
1989	3	-0.37	1.13	-0.36	-1.19	0.63	-0.61
1989	4	-0.61	-0.18	-1.38	0.52	0.6	0.81
1990	1	-0.85	0.13	-1.17	0.81	0.81	0.36

1990	2	-1.09	1.44	-0.57	-0.98	1.82	1.4
1990	3	-1.33	1.06	0.29	-1.18	1.51	0.77
1990	4	-1.58	-0.49	-0.1	-0.65	0.8	1.71
1991	1	-1.66	0.38	-0.12	-0.42	0.77	1.35
1991	2	-1.42	2.79	0.91	-1.65	1.19	0.71
1991	3	-1.19	0.55	0.16	-1.57	0.49	0.57
1991	4	-0.95	0.9	0.14	1.17	0.28	1.22
1992	1	-0.89	1.78	1.8	-1.87	1.03	-0.34
1992	2	-1.18	-0.72	0.74	0.26	1.28	-1.78
1992	3	-1.47	2.33	0.6	3.41	-1.24	-0.1
1992	4	-1.76	0.21	0.68	1.14	-1.14	1.36
1993	1	-1.88	1.75	1.22	0.19	-0.79	1.47
1993	2	-1.68	0.17	1.78	-0.95	0.6	-0.62
1993	3	-1.48	2.09	0.86	0.81	-0.27	0.87
1993	4	-1.28	-0.24	-1.03	0.5	-1.38	1.35
1994	1	-1.15	0.04	0.07	0.23	-0.59	-0.44
1994	2	-1.19	0.18	0.19	-1.28	0.99	-0.6
1994	3	-1.22	-0.6	-0.08	-1.71	0.7	0.59
1994	4	-1.26	0.67	-0.01	-1.31	-0.41	1.44
1995	1	-1.18	0.39	0.8	-1.64	1.06	-1.04
1995	2	-0.89	1.53	0.37	-0.18	0.49	-0.54
1995	3	-0.59	2.21	0.5	-0.98	1.32	0.77
1995	4	-0.3	-0.32	1.06	-0.28	0.45	0.72
1996	1	-0.07	1.12	1.11	0.62	0.92	2.13
1996	2	0.03	0.52	0.39	0.69	-0.06	1.7
1996	3	0.13	0.93	0.74	-1.21	0.01	0.35
1996	4	0.23	1.96	1.97	1.57	-2.15	1.34
1997	1	0.12	1.87	1.45	-0.65	0.18	1.97
1997	2	-0.41	0.56	1.55	0.45	-0.58	-0.91
1997	3	-0.93	0.91	2.62	-1.1	-0.03	-0.42
1997	4	-1.46	0.2	3.42	-0.61	0.59	0.36
1998	1	-1.68	1.82	4.03	-2.25	1.36	-0.73
1998	2	-1.27	1.34	2.16	-1.41	1.97	-0.84
1998	3	-0.87	1.41	0.81	-2	1.85	0.39
1998	4	-0.47	1.2	0.35	-2.26	2.23	0.39
1999	1	-0.08	0.76	0.48	-1.35	1.47	1.42
1999	2	0.28	0.13	0.43	-1.18	0.01	0.97
1999	3	0.65	1.19	0.31	-0.67	1.22	0.76
1999	4	1.01	-0.8	0.07	-1.4	1.2	0.22
2000	1	1.28	-0.63	-0.46	-2.13	1.76	1.33
2000	2	1.38	-0.06	0.62	-1.71	1.16	0.1
2000	3	1.49	-0.83	0.64	-0.56	-0.05	1.62
2000	4	1.59	-0.04	-0.36	0.69	0.08	0.92

2001	1	1.53	-0.89	-1.25	-0.56	-0.72	-0.16
2001	2	1.16	-0.26	-0.38	-0.95	0.39	1.09
2001	3	0.79	0.13	-0.66	-1.97	1.13	0.2
2001	4	0.42	-0.1	0.18	-1.66	1.37	0.72
2002	1	0.37	-0.5	0.08	-1.89	1.36	-1.1
2002	2	0.94	-0.5	-1.27	2.19	-0.34	1.5
2002	3	1.52	-0.52	0.56	-1.5	1.95	1.18
2002	4	2.09	0.58	0.39	-0.16	1.53	0.25
2003	1	2.31	-0.01	0.08	-1.21	0.91	0.82
2003	2	1.83	0.75	0.47	-0.08	0.62	0.27
2003	3	1.34	-0.5	0.41	-2.14	1.72	0.08
2003	4	0.85	0.95	1.36	-0.39	2.29	0.79
2004	1	0.45	0.35	1.18	-1.38	-0.61	0.88
2004	2	0.2	0.46	-0.02	-1.26	1.51	0.06
2004	3	-0.05	2.2	0.53	1.28	0.72	-1.37
2004	4	-0.29	2.36	1.44	-1.25	0.09	1.39
2005	1	-0.46	1.99	0.95	-1.45	1.32	-0.2
2005	2	-0.47	0.3	2.17	-1.19	1.2	0.71
2005	3	-0.47	2.56	1.49	-2.57	1.61	-0.6
2005	4	-0.48	0.76	1.19	-1.89	1.87	1.27
2006	1	-0.32	0.45	0.91	-1.1	2.67	0.34
2006	2	0.16	0.39	1.38	-1.68	1.43	-0.08
2006	3	0.64	-0.41	0.72	-3.33	2.19	-0.52
2006	4	1.11	0.7	1.8	-0.03	0.13	0.96
2007	1	1.39	0.85	-0.04	-0.84	1.06	1.12
2007	2	1.25	1.37	-0.11	-0.41	1.41	0.83
2007	3	1.11	0.89	1.19	-2.54	2.2	-1.29
2007	4	0.97	-0.81	-0.15	-1.93	1.36	0.6
2008	1	0.82	0.95	0.06	-0.86	-0.71	-0.35
2008	2	0.62	0.26	0.44	-0.13	-0.99	0.19
2008	3	0.43	2.72	2.09	-1.34	-0.32	1.36
2008	4	0.24	-0.21	1.22	-0.88	0	0.45
2009	1	0.08	0.48	-0.86	-0.43	-0.25	-0.06
2009	2	0.02	1.48	0.12	-0.09	-0.28	-0.01
2009	3	-0.05	0.43	2.31	0.42	0.37	-0.59
2009	4	-0.12	1.08	3.48	-0.7	1.37	0.64
2010	1	-0.13	1.11	2.64	-0.9	-0.52	-0.76
2010	2	-0.01	0.82	1.61	-1.59	1.97	1.09
2010	3	0.11	2.48	0.69	-2.9	1.88	0.47
2010	4	0.22	1.46	1.88	-1.54	1.72	1.5
2011	1	0.3	0.74	0.74	-0.56	-0.36	0.74
2011	2	0.28	2.14	1.94	-0.55	-0.08	0.28
2011	3	0.27	-0.38	2.87	-3.23	2.45	0.28

2011	4	0.26	-0.06	1.86	-1.34	1.62	1.28
2012	1	0.29	0.1	0.08	-1.96	1.46	0.24
2012	2	0.43	0.7	1.6	-1.98	3.44	0.04
2012	3	0.56	-1.67	2.45	-2.94	2.69	-0.17
2012	4	0.7	-0.56	1.2	-0.66	0.63	0.83
2013	1	0.59	1.15	0.63	-1.55	0.01	-1.53
2013	2	-0.01	0.61	0.43	1.66	-0.1	-0.02
2013	3	-0.6	2.32	1.71	-1.74	2.26	-1.12
2013	4	-1.2	1.35	1.11	-1.03	1.22	0.24
2014	1	-1.57	-0.28	-0.27	0.88	-0.2	-0.51
2014	2	-1.47	1.24	1.72	0.16	-0.15	-0.93
2014	3	-1.38	3.52	2.39	-1.39	0.39	-0.07
2014	4	-1.29	0.78	3.42	-0.1	1.44	0.5
2015	1	-1.36	0.22	1.61	-0.74	1.93	-1.16
2015	2	-1.75	0.92	0.19	-1.27	2.27	-1.98
2015	3	-2.13	1.86	2.32	-2.36	2.05	-0.23
2015	4	-2.51	2.66	3.67	-2.05	3.59	0.79
2016	1	-2.54	2.86	3.29	-2.37	2.9	-1.05
2016	2	-1.84	2.62	2.4	-1.93	3.08	-1.33
2016	3	-1.14	1.83	1.61	-3.71	2.98	0.55
2016	4	-0.45	0.54	3.68	-2.58	4.24	0
2017	1	0.01	2.88	1.47	-1.06	1.91	0.87
2017	2	0	2.52	2.56	-0.78	1.07	0.45
2017	3	-0.02	1.38	3.57	-2.66	1.12	0.01
2017	4	-0.04	1.49	3.08	-1.03	2.98	0.89
2018	1	0.07	0.54	0.03	-0.38	1.36	0.01
2018	2	0.43	0.6	1.94	-0.59	0.92	-0.45
2018	3	0.79	1.38	2.24	-3.24	2.63	0.33
2018	4	1.15	2.72	3.14	0.16	1.8	-1.39
2019	1	1.17	3.13	2.39	-0.31	0.28	-0.31
2019	2	0.48	1.62	2.46	-0.3	0.81	-1.17
2019	3	-0.2	0.29	3.26	-2.99	1.6	-0.71
2019	4	-0.88	1.49	4.05	0.1	2	0.52
2020	1	-1.16	2.04	2.21	-1.92	0.54	-2.12
2020	2	-0.64	1.24	2.46	-0.14	0.74	-0.07
2020	3	-0.13	1.31	3.26	-3.42	2.5	0.52

Consecutive Dry Days Codes

```
data = read.csv(file.choose())
#length of time series is 239
plot(1:239,data$CDD,type="l",xlab="year",ylab="CDD",xaxt="n")
axis(1,at=seq(1,length(data$Rx5Day),4),labels=1961:2020)

y.acf <- acf(data$CDD,main="",ylim=c(-1,1))
y.pacf <- pacf(data$CDD,main="",ylim=c(-1,1))

#Model Identification *****************************
CDD.ma10 <- arima(data$CDD,c(0,0,10))
CDD.ma10
#Diagnostic Checking *********************************
AIC(CDD.ma10)
BIC(CDD.ma10)
acf(residuals(CDD.ma10),ylim=c(-1,1))
pacf(residuals(CDD.ma10),ylim=c(-1,1))
Box.test(residuals(CDD.ma10),lag=22,fitdf=10)
Box.test(residuals(CDD.ma10),lag=22,type="Ljung-Box",fitdf=10)
hist(residuals(CDD.ma10)/sqrt(CDD.ma10$sigma2))
qqnorm(residuals(CDD.ma10))
qqline(residuals(CDD.ma10))
tsdiag(CDD.ma10,gof.lag=22)

#ARCH *****************************
error <- residuals(CDD.ma10)
plot(1:length(error),error,type="l",xlab="",ylab="Residuals")
abline(a=0,b=0)
error.2 <- error^2
plot(1:length(error.2),error.2,type="l",xlab="",ylab="Squared Residuals")
acf(error.2,ylab="Squared Residual Autocorrelation",main="",ylim=c(-1,1))
# Lagrange-Multipler Test with p = 5
y <- error.2[6:239]
x1 <- error.2[5:238]
x2 <- error.2[4:237]
x3 <- error.2[3:236]
x4 <- error.2[2:235]
x5 <- error.2[1:234]
round(cbind(y,x1,x2,x3,x4,x5),4)
summary(lm(y~x1+x2+x3+x4+x5))
(LM.stat <- 239*summary(lm(y~x1+x2+x3+x4+x5))$r.squared)
qchisq(0.95,5)
# Portmanteau Test
Box.test(error.2,lag=10,type="Ljung-Box",fitdf=0)

#Forecasting ***********************************

CDD.ma10.pred <- predict(CDD.ma10,n.ahead=16)
CDD.ma10.pred
i <- 1:length(data$CDD)
plot(i,data$CDD,type="l",xlab="",xaxt="n",xlim=c(1,(length(data$CDD)+16)))
abline(a=0,b=0)
axis(1,at=seq(1,(length(data$CDD)+16),4),labels=1961:2024)
i <- (length(data$CDD)+1):(length(data$CDD)+16)
lines(i,CDD.ma10.pred$pred,col=2,lty=2)
lines(i,CDD.ma10.pred$pred+1.96*CDD.ma10.pred$se,col=3,lty=3)
```

```
lines(i,CDD.ma10.pred$pred-1.96*CDD.ma10.pred$se,col=3,lty=3)
legend(96,13,legend=c("data$CDD","Forecasts","Forecast
Intervals"),lty=c(1,2,3),col=c(1,2,3))
```

Maximum 5-day rainfall Codes

```
data = read.csv(file.choose())
#length of time series is 239

# Time series plots of original series
i <- 1:length(data$Rx5Day)
plot(i,data$Rx5Day,type="l",xlab="",ylab="Rx5Day",xaxt="n")
axis(1,at=seq(1,length(data$Rx5Day),4),labels=1961:2020)
#ts.plot(diff(data$Rx5Day),type="l",xlab="") # not shown

# ACF and PACf of original and 1st differenced series
acf(data$Rx5Day,ylim=c(-1,1),lag.max=20,xaxt="n")
axis(1,at=seq(0,20,2))
pacf(data$Rx5Day,ylim=c(-1,1),lag.max=20,xaxt="n")
axis(1,at=seq(0,20,2))

acf(diff(data$Rx5Day),ylim=c(-1,1),lag.max=60,xaxt="n")
axis(1,at=seq(0,60,2))
pacf(diff(data$Rx5Day),ylim=c(-1,1),lag.max=60,xaxt="n")
axis(1,at=seq(0,60,2))

# Model Estimation ****************************************

#AR(6)
Rx5Day.ar6 <- arima(data$Rx5Day,c(6,0,0))
Rx5Day.ar6

#ARIMA(0,1,1)
Rx5Day.arima011 <- arima(data$Rx5Day,c(0,1,1))
Rx5Day.arima011

# Diagnostic Checking ************************************

AIC(Rx5Day.ar6)
BIC(Rx5Day.ar6)

AIC(Rx5Day.arima011)
BIC(Rx5Day.arima011)

acf(residuals(Rx5Day.arima011),ylim=c(-1,1))
pacf(residuals(Rx5Day.arima011),ylim=c(-1,1))

Box.test(residuals(Rx5Day.arima011),lag=22,fitdf=3)
Box.test(residuals(Rx5Day.arima011),lag=22,type="Ljung-Box",fitdf=3)

hist(residuals(Rx5Day.arima011)/sqrt(Rx5Day.arima011$sigma2))
qqnorm(residuals(Rx5Day.arima011))
qqline(residuals(Rx5Day.arima011))

#Forecasting **********************************************
```

```
Rx5Day.arima011.pred <- predict(Rx5Day.arima011,n.ahead=24)
Rx5Day.arima011.pred
i <- 1:length(data$Rx5Day)
plot(i,data$Rx5Day,type="l",xlab="",xaxt="n",xlim=c(1,(length(data$Rx5Day)+24
)))
abline(a=0,b=0)
axis(1,at=seq(1,(length(data$Rx5Day)+24),4),labels=1961:2026)
i <- (length(data$Rx5Day)+1):(length(data$Rx5Day)+24)
lines(i,Rx5Day.arima011.pred$pred,col=2,lty=2)
lines(i,Rx5Day.arima011.pred$pred+1.96*Rx5Day.arima011.pred$se,col=3,lty=3)
lines(i,Rx5Day.arima011.pred$pred-1.96*Rx5Day.arima011.pred$se,col=3,lty=3)
legend(96,13,legend=c("data$Rx5Day","Forecasts","Forecast
Intervals"),lty=c(1,2,3),col=c(1,2,3))
```

Sea Level Codes

```
data = read.csv(file.choose())
#length of time series is 239
i <- 1:length(data$Sea.Level)
plot(i,data$Sea.Level,type="l",xlab="",ylab="Sea.Level",xaxt="n")
axis(1,at=seq(1,length(data$Sea.Level),4),labels=1961:2020)
abline(v = 157, col=2)

#reduced length is 83
TIME <- 1:83
SL2 <- data$Sea.Level[c(157:239)]

# Model Identification, Estimation and Checking
*************************************

Y.lag1 <- c(NA,SL2[-83])
Y.diff <- c(NA,diff(SL2))
Y.diff.lag1 <- c(NA,NA,diff(SL2)[-82])
Y.diff.lag2 <- c(NA,NA,NA,diff(SL2)[-(81:82)])
cbind(TIME,SL2,Y.lag1,Y.diff,Y.diff.lag1,Y.diff.lag2)

# Dickey-Fuller Test with Trend
summary(lm(Y.diff~Y.lag1+TIME))

# Augmented Dickey-Fuller Test with Trend and p = 1
summary(lm(Y.diff~Y.lag1+TIME+Y.diff.lag1))

# Augmented Dickey-Fuller Test with Trend and p = 2
summary(lm(Y.diff~Y.lag1+TIME+Y.diff.lag1+Y.diff.lag2))

# Linear Regression against time
summary(lm(SL2~TIME))
Box.test(residuals(lm(SL2~TIME)),lag=22,fitdf=0)
acf(residuals(lm(SL2~TIME)), ylim=c(-1,1), lag.max=20, xaxt="n")
axis(1,at=seq(0,20,2))
pacf(residuals(lm(SL2~TIME)), ylim=c(-1,1), lag.max=20, xaxt="n")
axis(1,at=seq(0,20,2))

# AIC Grid Search for ARMA(p, q)
AICtable = matrix(nrow=10, ncol=10)
for (p in 1:10) {
```

```
  for (q in 1:10) {
    if (p==4 & q==3) next #avoid error
    if (p==9 & q==3) next #avoid error
    if (p==9 & q==8) next #avoid error
    Sea.Level.arima101 <- arima(residuals(lm(SL2~TIME)),c(p,0,q))
    boo1 = Box.test(residuals(Sea.Level.arima101),lag=22,fitdf=p+q)
    boo2 = Box.test(residuals(Sea.Level.arima101),lag=22,type="Ljung-
Box",fitdf=p+q)
    if (boo1$p.value > 0.05 & boo2$p.value > 0.05) {
      AICtable[p, q] = AIC(Sea.Level.arima101)
    }
  }
}
AICtable

# BIC Grid Search for ARMA(p, q)
BICtable = matrix(nrow=10, ncol=10)
for (p in 1:10) {
  for (q in 1:10) {
    if (p==4 & q==3) next #avoid error
    if (p==9 & q==3) next #avoid error
    if (p==9 & q==8) next #avoid error
    Sea.Level.arima101 <- arima(residuals(lm(SL2~TIME)),c(p,0,q))
    boo1 = Box.test(residuals(Sea.Level.arima101),lag=22,fitdf=p+q)
    boo2 = Box.test(residuals(Sea.Level.arima101),lag=22,type="Ljung-
Box",fitdf=p+q)
    if (boo1$p.value > 0.05 & boo2$p.value > 0.05) {
      BICtable[p, q] = BIC(Sea.Level.arima101)
    }
  }
}
BICtable

#ARMA(7,7) summary
Sea.Level.arima707 <- arima(residuals(lm(SL2~TIME)),c(7,0,7))
Sea.Level.arima707
AIC(Sea.Level.arima707)
BIC(Sea.Level.arima707)
Box.test(residuals(Sea.Level.arima707),lag=22,fitdf=14)
Box.test(residuals(Sea.Level.arima707),lag=22,type="Ljung-Box",fitdf=14)

#ARMA(1,7) summary
Sea.Level.arima107 <- arima(residuals(lm(SL2~TIME)),c(1,0,7))
Sea.Level.arima107
AIC(Sea.Level.arima107)
BIC(Sea.Level.arima107)
Box.test(residuals(Sea.Level.arima107),lag=22,fitdf=8)
Box.test(residuals(Sea.Level.arima107),lag=22,type="Ljung-Box",fitdf=8)

# Forecasting
**********************************************************************
# residual ARMA(1,7) model
Sea.Level.arima107.pred <- predict(Sea.Level.arima107,n.ahead=24)
Sea.Level.arima107.pred
i <- 1:length(residuals(lm(SL2~TIME)))
plot(i,residuals(lm(SL2~TIME)),type="l",xlab="",xaxt="n",xlim=c(1,(length(res
iduals(lm(SL2~TIME)))+24)))
```

```
abline(a=0,b=0)
axis(1,at=seq(1,(length(residuals(lm(SL2~TIME)))+24),4),labels=2000:2026)
i <- (length(residuals(lm(SL2~TIME)))+1):(length(residuals(lm(SL2~TIME)))+24)
lines(i,Sea.Level.arima107.pred$pred,col=2,lty=2)
lines(i,Sea.Level.arima107.pred$pred+1.96*Sea.Level.arima107.pred$se,col=3,lt
y=3)
lines(i,Sea.Level.arima107.pred$pred-
1.96*Sea.Level.arima107.pred$se,col=3,lty=3)
legend(96,13,legend=c("residuals(lm(SL2~TIME))","Forecasts","Forecast
Intervals"),lty=c(1,2,3),col=c(1,2,3))

# linear model
Sea.Level.2050pred <- (-0.164787) + (83 + 4 * 30) * (0.035906)
Sea.Level.2050pred
```