ECON 144 Project 2

Andy Ho

2024-11-14

I. Introduction

Solar radiation, another way to call sunlight or light resource from the Sun, is electromagnetic radiation emitted by the Sun. As the Sun emits energy in radiation of short wavelengths, clouds and other gas molecules that are present weakens it as the radiation passes through the atmosphere. Once it reaches the Earth's surface, it can get reflected or absorbed, which is when humans may be able to capture it and utilize it for energy. Humans use solar radiation everyday in different forms such as heat and electricity. Therefore, its use is important for many technologies that have been developed over the years. Once it is reflected or absorbed on the Earth's surface, it returns back to space in radiation of long wavelengths.

Solar radiation has played an impactful role in regards to the Earth's climate. Solar radiation directly keeps the atmosphere worm and it creates wind patterns as solar energy is not evenly distributed across Earth's surface. Cloud formation and temperatures are also directly influenced by the Sun. The Sun also drives photosynthesis for plants, which absorbs carbon dioxide and ultimately cools the climate throughout the process. Solar flares increase solar radiation because they are intense bursts of electromagnetic radiation emitted from the Sun's atmosphere, releasing large amounts of energy across the electromagnetic spectrum. Overall, the Sun and its solar radiation has a significant influence in climate patterns.

However, solar radiation over the years have caused many scientists to be concerned about its effects on climate change. There has been many changes to the planet including an increase in solar irradiance, which is the amount of solar energy that reaches a surface per unit of area. This leads to increase warming of the planet, which directly leads to the global warming part of climate change that many scientists have been observing. The increase of solar irradiance liekly results in the release of methane and carbon dioxide that is stored in the Earth's ocean and ice caps. The ice caps itself presents a positive feedback loop in climate change. As more ice melts, more energy is abosrbed since the high albedo ice melts into the darker ocean waters. As a result, there is further warming and this causes even more ice to melt. Therefore, it is important to measure solar radiation because it provides vital information for understanding and predicting climate change since solar radiation has a huge influence in climate patterns and it can cause further increase in Earth's temperatures.

The data set that we are working with provides daily observations of the estimated DGSR (daily global solar radiation) at Zhongshan Station in Antarctica, one of Earth's ice caps. The long-term estimated DGSR provides us a detailed table of radiation data that allows us to observe how solar radiation and ice caps play a large role in global climate change. The data has four columns which includes the year, month, day, and the estimated DGSR value. The values range from March 1st, 1989 to May 26th, 2020. We start off with numerical values for all four columns for the entire data set, where Months 1 to 12 are respectively describing January to December. We will use the data to see how average solar radiation has varied over time and see if there are seasonal patterns. This will help us observe if there are changes to solar radiation and if certain periods has more solar radiation present, which can help us know when there will be greater changes to the Earth's climate. To make it easier, we will modify the original data set so that we are observing monthly data of the average daily estimated DGSR for our analysis. We will also include a second variable from another data set called Radio Flux 10.7cm, which is data related to solar radio flux at a wavelength of 10.7 centimeters. Solar radio flux measurements are used as a proxy for solar activity, with higher values indicating increased solar activity. This second variable is within a data set that comes from the Space Weather Prediction Center. We will start from the first month that is present in both data sets, January 1997, to the last month that is present, May 2020. Note that June is not present throughout all the years for the measurement of solar radiation in Antarctica. Let the overall hypothesis be that solar radiation and radio influx (10.7cm) increases throughout the years.

Uploading the Data

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
##    Year Month Day Estimated.DGSR
## 1  1989     3   1      10.157780
## 2  1989     3   2       5.944928
## 3  1989     3   3      13.550971
## 4  1989     3   4       5.153736
## 5  1989     3   5       9.636340
## 6  1989     3   6       8.573378
## 7  1989     3   7      13.659337
## 8  1989     3   8      12.408861
## 9  1989     3   9       3.528212
## 10 1989     3  10       3.710640
## 11 1989     3  11      12.056321
## 12 1989     3  12       6.274177
## 13 1989     3  13       4.966673
## 14 1989     3  14       7.096776
## 15 1989     3  15      10.839193
## 16 1989     3  16       4.095352
## 17 1989     3  17       7.268062
## 18 1989     3  18       3.954773
## 19 1989     3  19       9.497548
## 20 1989     3  20       8.737299
## 21 1989     3  21       4.412003
## 22 1989     3  22       3.064991
## 23 1989     3  23       4.083718
## 24 1989     3  24       4.802231
```

```
## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 24 × 3
## # Groups:   Year [3]
##     Year Month Monthly_Avg_DGSR
##    <int> <int>            <dbl>
##  1  1989     3           6.51
##  2  1989     4           2.05
##  3  1989     5           0.238
##  4  1989     6           0.169
##  5  1989     7           0.0990
##  6  1989     8           1.17
##  7  1989     9           4.88
##  8  1989    10          12.9
##  9  1989    11          23.3
## 10  1989    12          25.0
## # ℹ 14 more rows
```

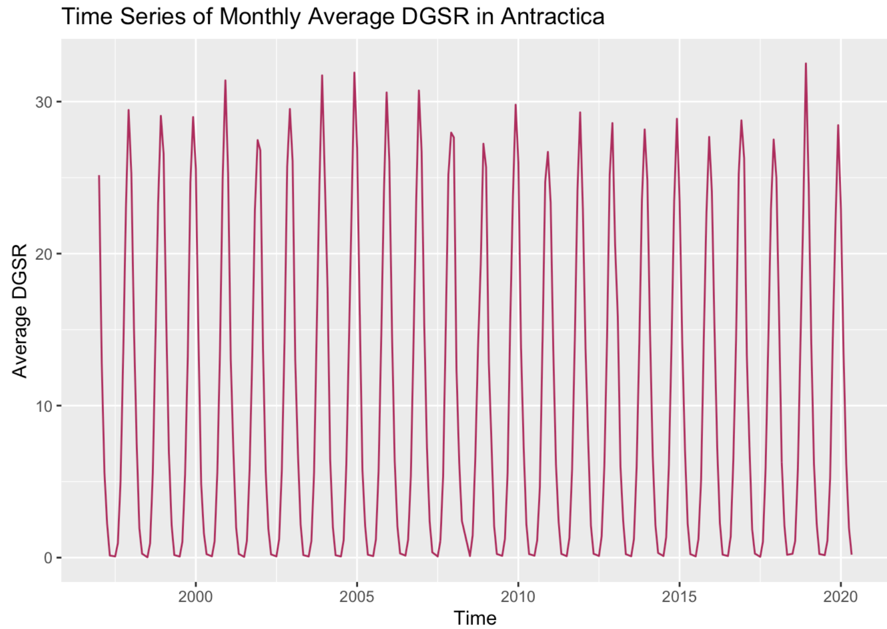
```
## # A tibble: 12 × 2
##    Month   Avg_Radio_Flux
##    <chr>            <dbl>
##  1 1997-01           74
##  2 1997-02           73.8
##  3 1997-03           73.6
##  4 1997-04           74.6
##  5 1997-05           74.8
##  6 1997-06           71.8
##  7 1997-07           71.2
##  8 1997-08           79.2
##  9 1997-09           96.2
## 10 1997-10           85
## 11 1997-11           99.5
## 12 1997-12           98.8
```

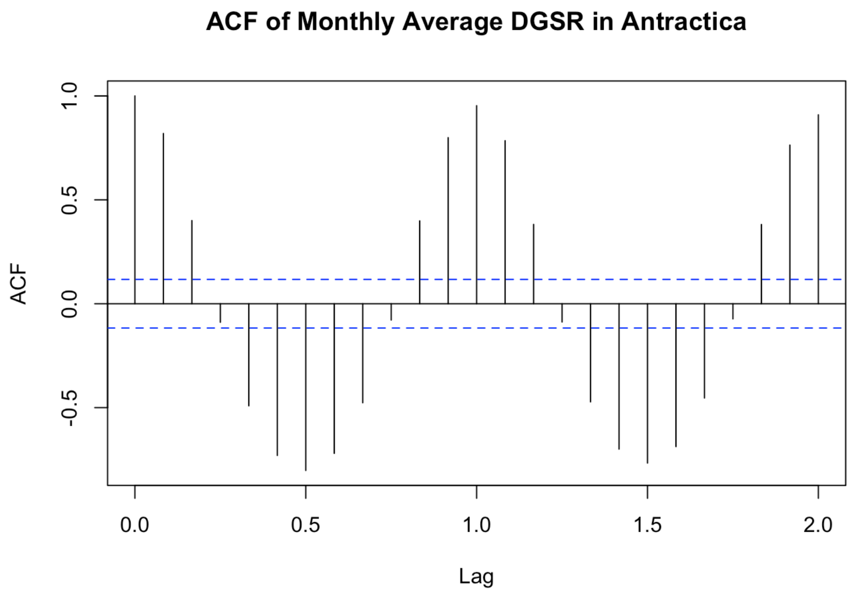
II. Results

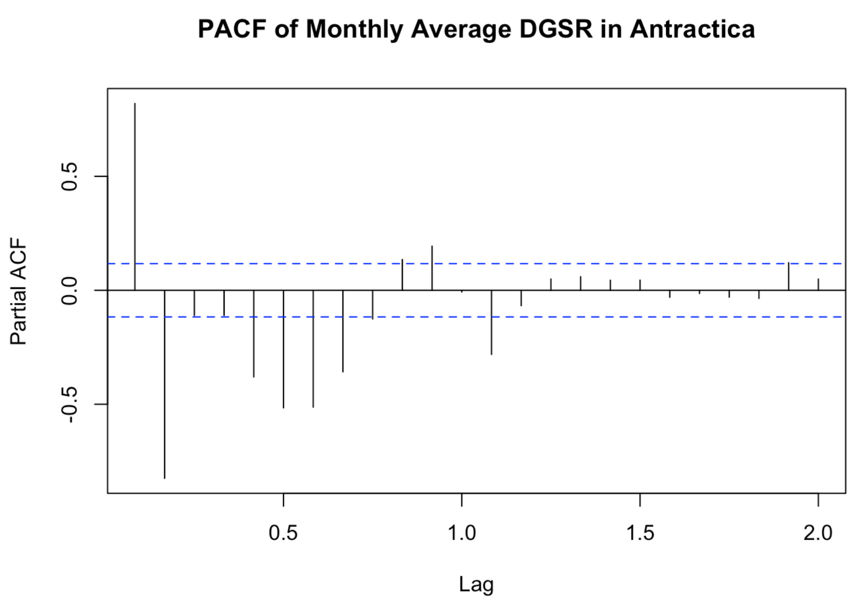
1. Modeling and Forecasting Trend, Seasonality, and Cycles

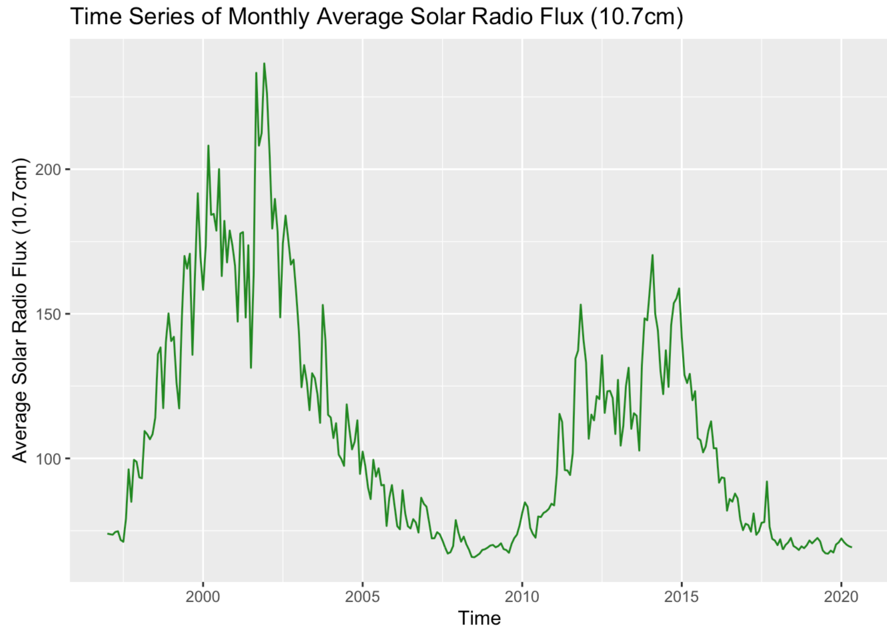
(a) Produce a time-series plot of your data including the respective ACF and PACF plots.

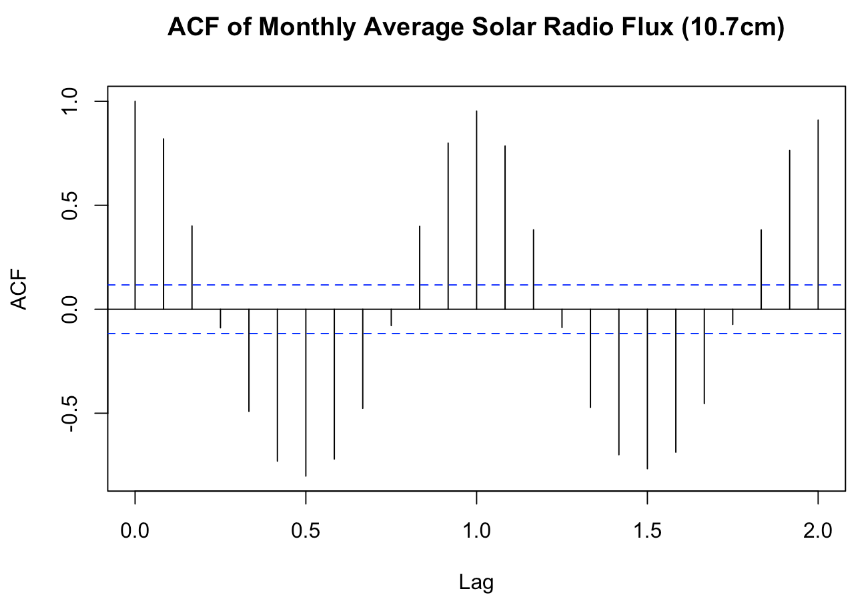
```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

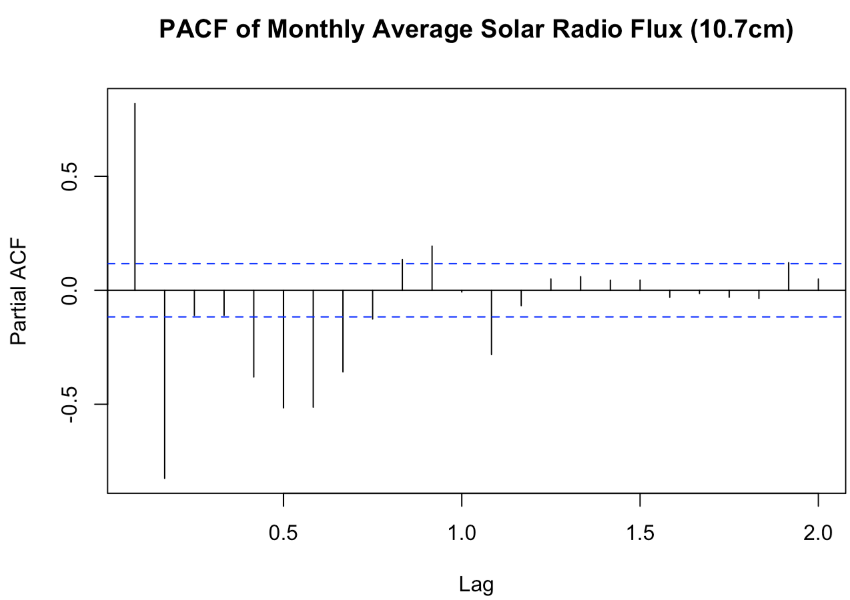

Time Series of Monthly Average DGSR in Antractica
Average DGSR
0
10
20
30
2000
2005
2010
2015
2020
Time


ACF of Monthly Average DGSR in Antractica
ACF
-0.5
0.0
0.5
1.0
0.0
0.5
1.0
1.5
2.0
Lag

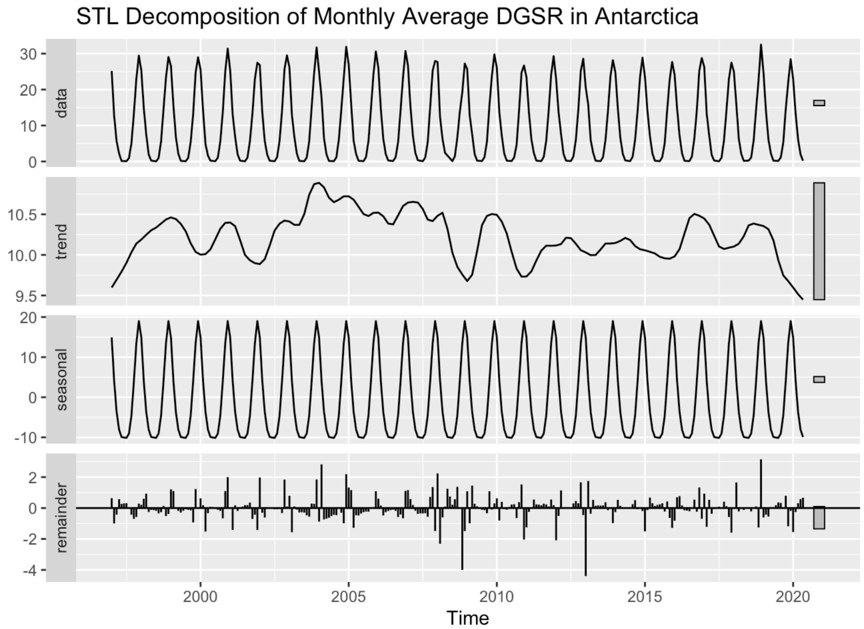

PACF of Monthly Average DGSR in Antractica
Partial ACF
-0.5
0.0
0.5
0.5
1.0
1.5
2.0
Lag

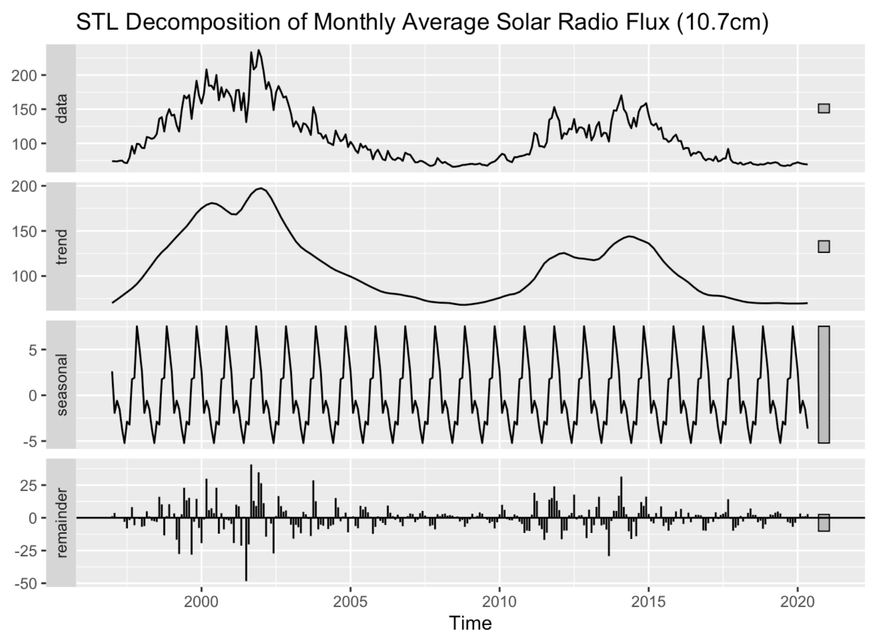






(b) Plot the stl decomposition plot of your data, and discuss the results.





(c) Fit a model that includes, trend, seasonality and cyclical components. Make sure to discuss your model in detail.

Here, we'll fit a seasonal ARIMA model with trend and seasonality.

```
## Series: dgsr_ts
## ARIMA(2,0,0)(0,1,1)[12]
##
## Coefficients:
##          ar1     ar2     sma1
##       0.0432  0.0692  -0.8515
## s.e.  0.0615  0.0610   0.0511
##
## sigma^2 = 0.8951:  log likelihood = -373.03
## AIC=754.06   AICc=754.21   BIC=768.43
##
## Training set error measures:
##                      ME      RMSE       MAE      MPE     MAPE      MASE
## Training set 0.00706925 0.9204907 0.5145053 -2.03336 13.22632 0.7139769
##                      ACF1
## Training set 0.0004009189
```

```
## Series: radio_ts
## ARIMA(2,1,2)(1,0,1)[12] with drift
##
## Coefficients:
##           ar1      ar2      ma1      ma2    sar1     sma1    drift
##       -0.1836  -0.0945  -0.0579  -0.1126  0.3686  -0.3995  -0.0230
## s.e.   0.3463   0.4458   0.3543   0.4320  0.5841   0.5746   0.4495
##
## sigma^2 = 150.3:  log likelihood = -1095.63
## AIC=2207.27   AICc=2207.8   BIC=2236.35
##
## Training set error measures:
##                      ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 0.01250338 12.08533 8.053127 -0.5613536 6.588604 0.3859651
##                      ACF1
## Training set 0.0006447286
```

dgsr_ts model:

For the dgsr_ts model, this model is specified as an ARIMA(2,0,0)(0,1,1)[12], which includes:

AR(2): Two autoregressive terms, suggesting that the current value is affected by the two preceding values.

Seasonal Differencing (1, period 12): One seasonal difference with a period of 12 (likely monthly data), meaning that the model accounts for yearly seasonality.

Seasonal MA(1): A moving average component for the seasonal part, suggesting a seasonal lag of one period (12 months).

The autoregressive coefficients (AR1 and AR2) are 0.0432 and 0.0692, indicating a mild influence from the past two observations. The seasonal moving average coefficient (SMA1) is −0.8515, suggesting a strong seasonal effect in the series. The variance of the residuals (error term) is 0.8951. the AIC, AICc, and BIC criteria (AIC = 754.06, AICc = 754.21, BIC = 768.43) help assess model fit. Lower values are generally better, but we would compare with alternative models to interpret this, which will happen later on.

We get the following training set error values:

ME (Mean Error): A very small value (0.0071), indicating minimal bias in predictions.

RMSE (Root Mean Square Error): 0.9205, representing the average magnitude of prediction error.

MAE (Mean Absolute Error): 0.5145, a measure of the average error magnitude.

MPE (Mean Percentage Error): -2.033%, showing slight under-prediction bias.

MAPE (Mean Absolute Percentage Error): 13.23%, indicating the average prediction error as a percentage.

ACF1 (Lag-1 Autocorrelation of Residuals): Near-zero, suggesting that residuals lack significant autocorrelation, which is desirable.

radio_ts model:

For the radio_ts model, this model is specified as an ARIMA(2,1,2)(1,0,1)[12], which includes:

AR(2) and MA(2): Two autoregressive and two moving average terms in the non-seasonal part, allowing the model to capture more complex temporal dependencies.

Non-seasonal Differencing (1): One non-seasonal difference, which indicates that the series has a trend component.

Seasonal AR(1) and Seasonal MA(1): Seasonal autoregressive and moving average terms with a period of 12, suggesting yearly seasonality.

Drift: The drift term (−0.023) accounts for a slow, consistent trend in the data over time.

The autoregressive coefficients (AR1 and AR2) are -0.1836 and -0.0945, indicating a potential oscillatory behavior in the data. The MA1 and MA2 coefficients are negative, with them being -0.0579 and -0.1126, respectively. This indicates some smoothing effect happening. The seasonal components (SAR1 and SMA1) are 0.3686 and -0.3995, respectively. This suggests the capturing of a yearly seasonal pattern. The AIC, AICc, and BIC criteria are as follows: AIC = 2207.27, AICc = 2207.8, BIC = 2236.35.

We get the following training set error values:

ME (Mean Error): Small bias (0.0125), close to zero.

RMSE (Root Mean Square Error): 12.085, much larger than dgsr_ts, indicating greater average prediction error.

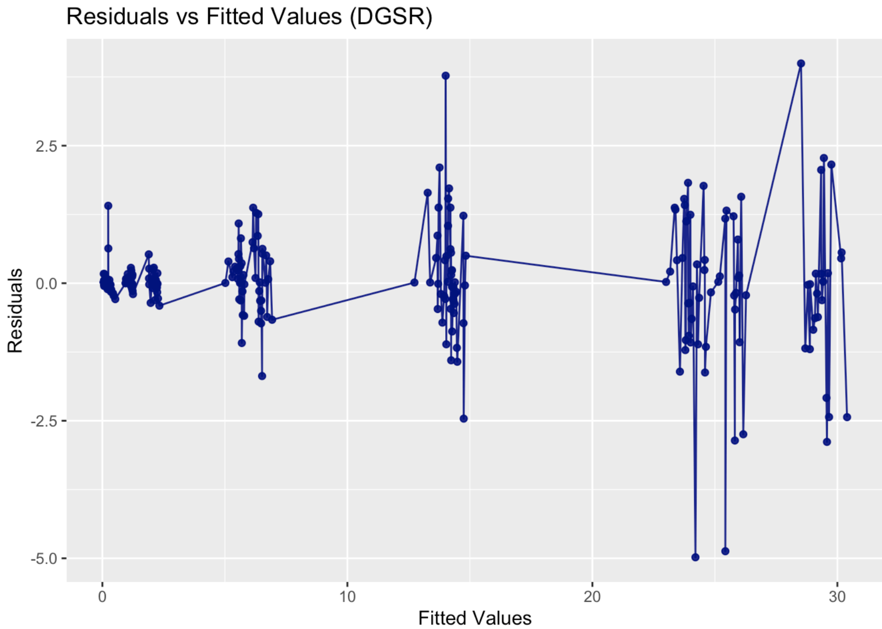
MAE (Mean Absolute Error): 8.053, also larger than dgsr_ts.

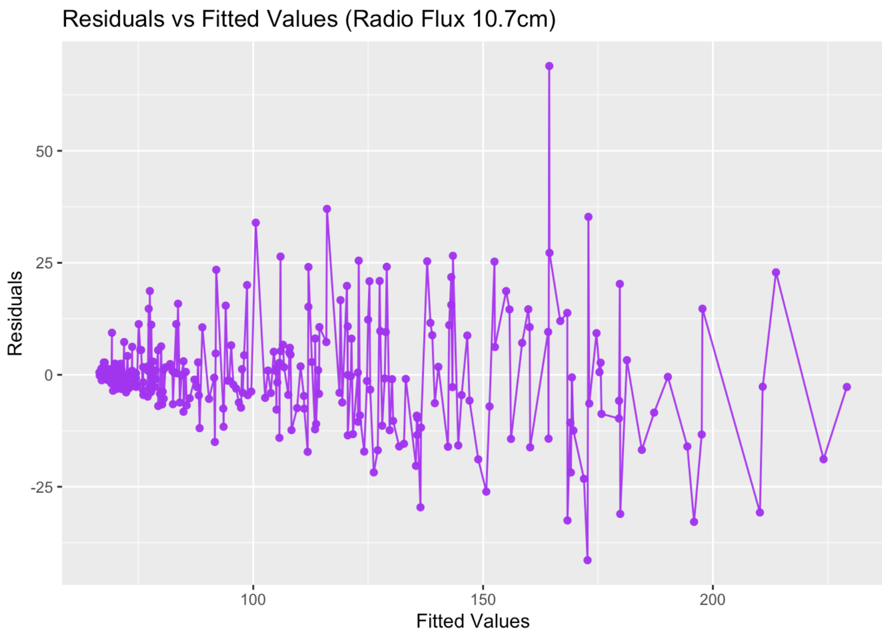
MPE (Mean Percentage Error): Small negative bias (-0.561%), showing slight under-prediction.

MAPE (Mean Absolute Percentage Error): 6.59%, better than dgsr_ts in terms of percentage error.

ACF1 (Lag-1 Autocorrelation of Residuals): Near-zero, indicating no significant autocorrelation in residuals.

(e) Plot the respective residuals vs. fitted values and discuss your observations.

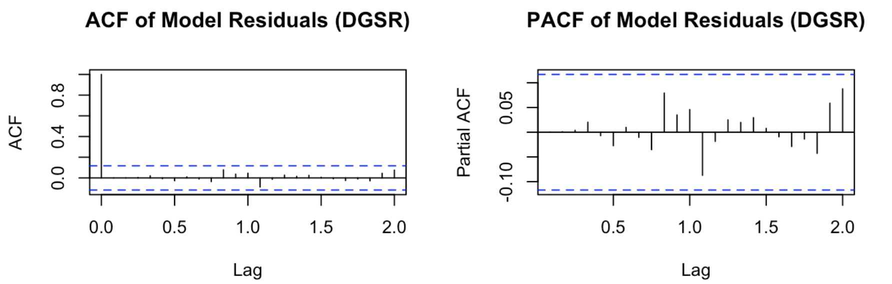


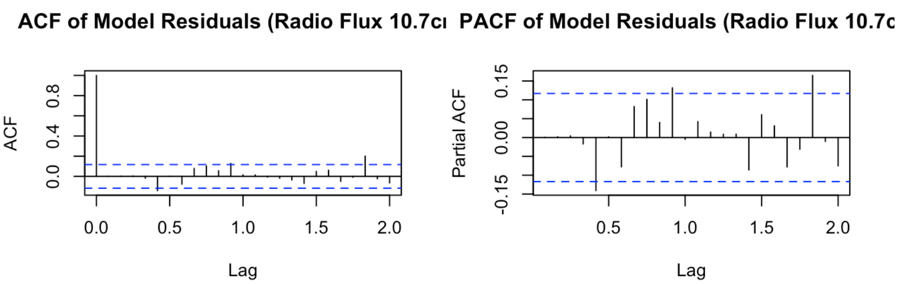


For the first plot concerning DGSR, the residuals in this plot appear to be fairly close to zero for most fitted values but show occasional spikes. There is no strong pattern across the fitted values, which can indicate that the model might be relatively well-fitted in terms of the error distribution. However, some outliers are present, as seen from the few large positive and negative residuals, which may indicate occasional instances where the model fails to predict accurately. If these outliers are not due to errors in data recording or other external factors, this suggests that the model may not fully capture all underlying patterns in the data.

For the second plot regarding the Radio Influx (10.7cm), the residuals show an upward trend with increasing fitted values, indicating potential heteroscedasticity (non-constant variance of residuals). This pattern suggests that the model might be underestimating or overestimating for larger values, which could imply that it's not fully capturing the relationship in the data, particularly at higher fitted values. The residuals appear more dispersed and show greater variability as fitted values increase, meaning the model's error increases with larger predicted values. This could indicate that a transformation of the target variable or a different model might improve the fit.

(f) Plot the ACF and PACF of the respective residuals and interpret the plots.





For the DGSR plots, the ACF plot shows a high autocorrelation at lag 0, as expected, but quickly drops close to zero for subsequent lags. All subsequent lag values fall within the blue confidence bands, which suggests no significant autocorrelation in the residuals. This pattern indicates that the residuals from the DGSR model are approximately uncorrelated, implying that the model has likely captured most of the autocorrelation in the data. The PACF plot also shows most lag values within the confidence bands, with only minor fluctuations. This further suggests that there is no significant partial autocorrelation, which supports the adequacy of the model in capturing the dependencies within the data. Overall, both the ACF and PACF plots for the DGSR model residuals indicate that the residuals are approximately white noise. This is a good sign and suggests that the model is adequately fitted with no remaining patterns in the residuals.

For the Radio Flux (10.7cm) plots, the ACF plot shows high autocorrelation at lag 0 but quickly drops to near-zero values within the confidence bands for subsequent lags. This implies that there is no significant autocorrelation left in the residuals after fitting the model. The PACF plot for Radio Flux (10.7cm) residuals also shows most values within the confidence bands, indicating no significant partial autocorrelation. A few minor peaks fall close to the confidence bands, but they are within acceptable limits, suggesting that the residuals do not have substantial partial autocorrelation. Overall, for the Radio Flux (10.7cm) residuals, both the ACF and PACF plots suggest that the residuals resemble white noise. This is generally a good indication that the model has captured most of the autocorrelation structure in the data.

(g) Plot the respective CUSUM and interpret the plot.

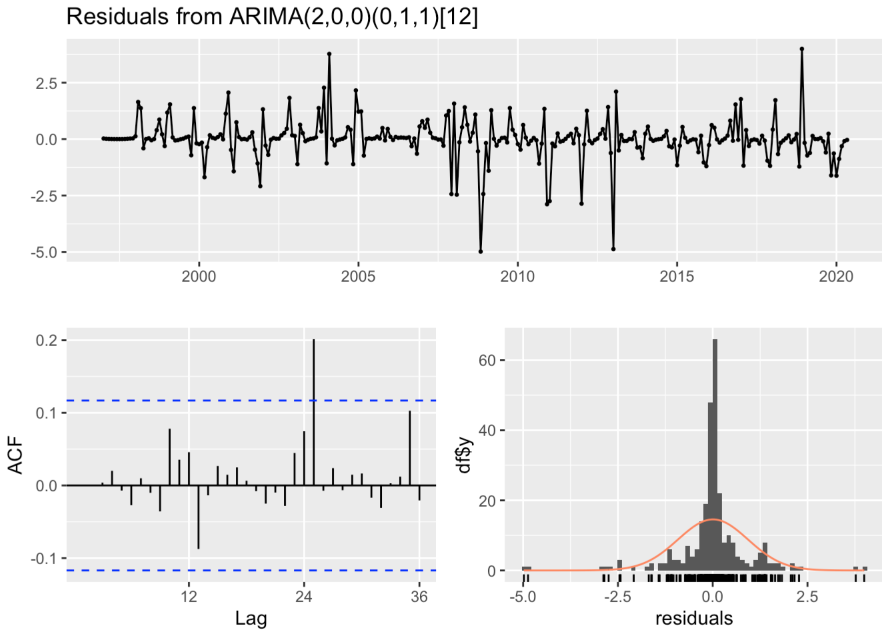
For the DGSR plot, we can see that from around 2005 to 2015, the CUSUM line drops consistently below zero, suggesting a systematic underestimation by the model. This indicates that there might be a trend or change in the underlying data that the model has not captured effectively. The downward trend crossing the lower control limit (around 2010) suggests a significant shift or break in the process, indicating a potential structural change in DGSR during that period.

For the Radio Flux (10.7cm) plot, the CUSUM line initially rises above zero until around 2003, suggesting an overestimation by the model, followed by a consistent downward trend that suggests underestimation afterward. There is a noticeable downward movement starting around 2005, indicating a potential structural break. However, the CUSUM line does not cross the control limits as sharply as in the DGSR plot. This may suggest that the changes in Radio Flux (10.7cm) are less pronounced or the model fit is slightly better in this case.

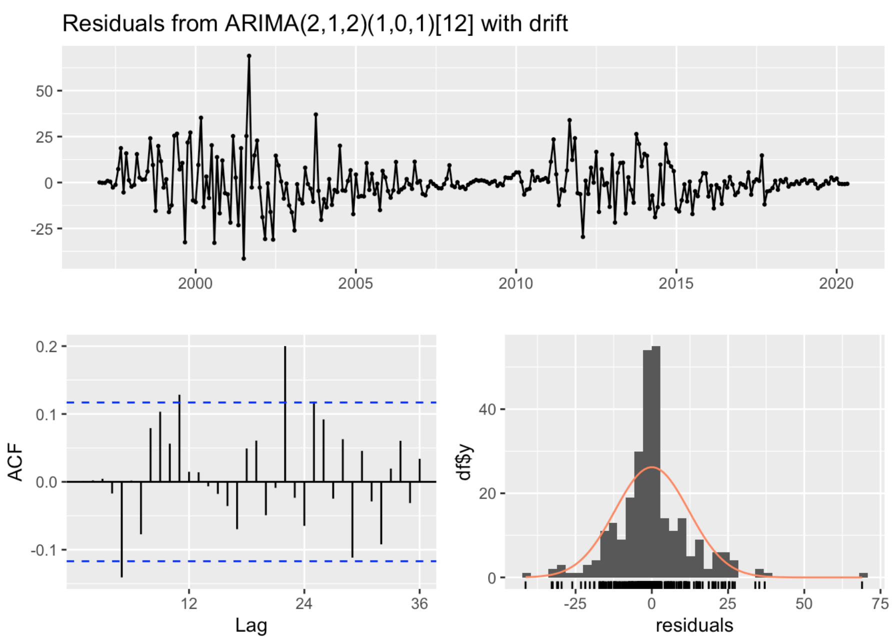
Both plots show signs of structural changes, but the break is more pronounced in the DGSR plot (left), where the CUSUM line crosses the lower control limit. The presence of consistent deviations (upward or downward trends) in the CUSUM plots indicates that the models may not adequately capture changes in the data over time, especially during the periods where the trends are observed.

(h) For your model, discuss the associated diagnostic statistics.

Let’s generate the output to observe the associated diagnostic statistics with our seasonal ARIMA model.




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,0,0)(0,1,1)[12]
## Q* = 9.1889, df = 21, p-value = 0.9876
##
## Model df: 3.   Total lags used: 24
```




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(2,1,2)(1,0,1)[12] with drift
## Q* = 36.73, df = 18, p-value = 0.005681
##
## Model df: 6.   Total lags used: 24
```

DGSR Model Diagnostics:

For the time series plot of residuals, the DGSR residuals appear to fluctuate around zero, but there are notable spikes, especially before 2010. There is some degree of volatility with larger residuals occurring sporadically, indicating potential outliers or irregular variations in the data that the model may not have captured adequately.

The ACF plot shows that most of the autocorrelation values lie within the confidence intervals (blue dashed lines), suggesting that the residuals are mostly uncorrelated. However, there are a few significant spikes (e.g., around lag 24), indicating potential seasonality that the model might not have fully accounted for, or some residual autocorrelation that needs to be addressed.

The histogram of residuals shows a fairly symmetric distribution around zero, which is consistent with normally distributed residuals. The overlaid density curve suggests a slight departure from normality, particularly with a higher peak and some skewness at the tails. This may indicate the presence of extreme values or outliers.

Radio Flux (10.7cm) Model Diagnostics:

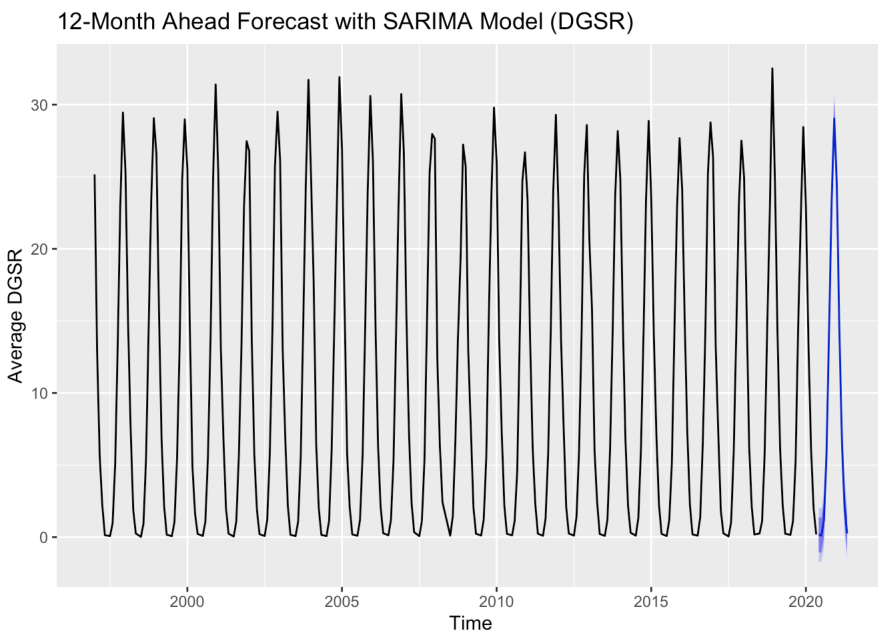
For the time series plot of residuals, the residuals show greater volatility compared to the DGSR model, particularly before 2010. There are larger fluctuations, indicating that the model may not be fitting the data well in certain periods. The volatility decreases after 2010, suggesting a change in variance that the model did not account for.

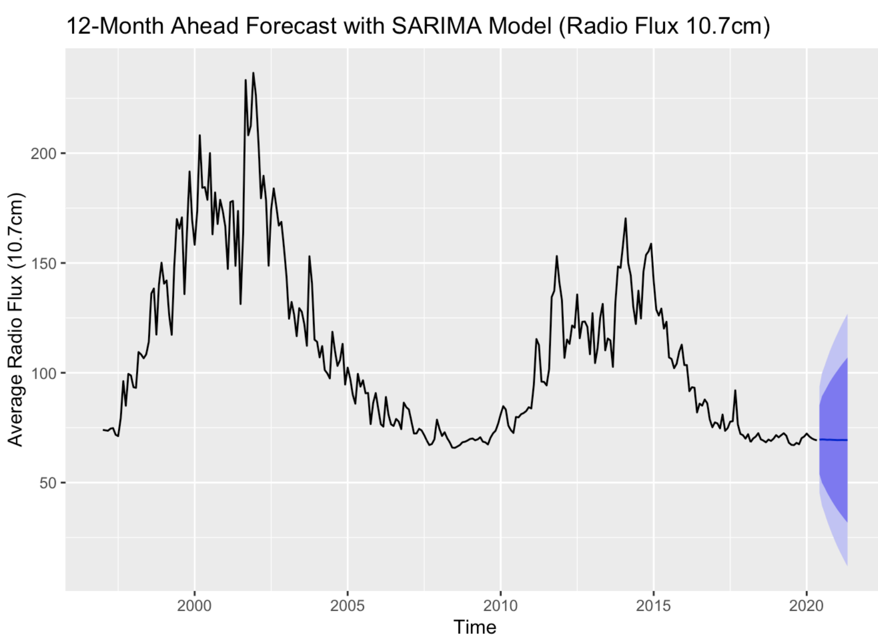
The ACF plot indicates significant autocorrelation at several lags (e.g., lag 24), and the pattern is more pronounced than in the DGSR model. This suggests that the residuals are not purely random and that the model has not fully accounted for some underlying seasonal or periodic components.

The histogram shows a noticeable departure from normality, with the residuals being heavily skewed and showing a high peak. The right tail is longer, suggesting the presence of extreme positive residuals. This asymmetry points to potential issues with model fit, particularly during certain periods or due to external shocks.

Overall, both models exhibit some issues with autocorrelation, though it is more pronounced in the Radio Flux model. The DGSR model shows better residual behavior with fewer signs of volatility and less skewness. For the Radio Flux (10.7cm) model, the presence of drift and higher volatility suggests a need for a more complex model to capture the underlying data dynamics.

(i) Use your model to forecast 12-steps ahead. Your forecast should include the respective error bands.





(j) Compare your forecast from (i) to the 12-steps ahead forecasts from auto.arima model. Which model performs best in terms of MAPE?

```
## [1] "DGSR MAPE: 10.4581896528297"
```

```
## [1] "DGSR Auto ARIMA Model MAPE: 10.4581896528297"
```

```
## [1] "Radio Flux MAPE: 2.16184933801712"
```

```
## [1] "Radio Flux Auto ARIMA Model MAPE: 2.16184933801712"
```

From the above output both the seasonal ARIMA model and the auto ARIMA model have the same MAPE values. Therefore, both models are similar in terms of performance for prediction accuracy. However, the models for predicting Radio Flux (10.7cm) seems to be better than the models for predicting DGSR, since 2.16184933801712 is less than 10.4581896528297.

(k) Combine the two forecasts and comment on the MAPE from this forecasts vs., the individual ones.

```
## [1] "DGSR Combined MAPE: 10.4581896528297"
```

```
## [1] "Radio Flux Combined MAPE: 2.16184933801712"
```

The MAPE for this forecast vs. the individual ones seem to be exactly the same. Therefore, the combined forecasts in this scenario likely has similar perfomance in prediction accuracy when compared to the individual forecasts.

(l) Fit an appropriate VAR model using your two variables. Make sure to show the relevant plots and discuss your results from the fit.

```
## Loading required package: MASS
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
## Loading required package: strucchange
```

```
## Loading required package: sandwich
```

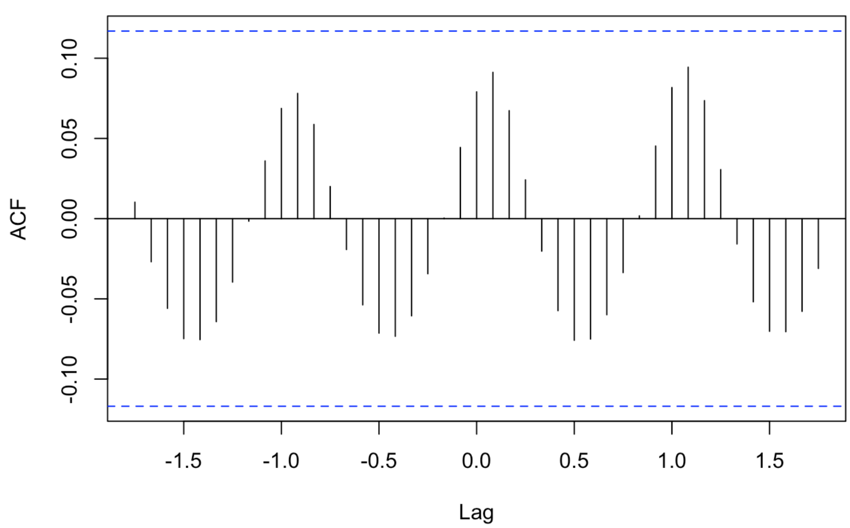
```
## Loading required package: urca
```

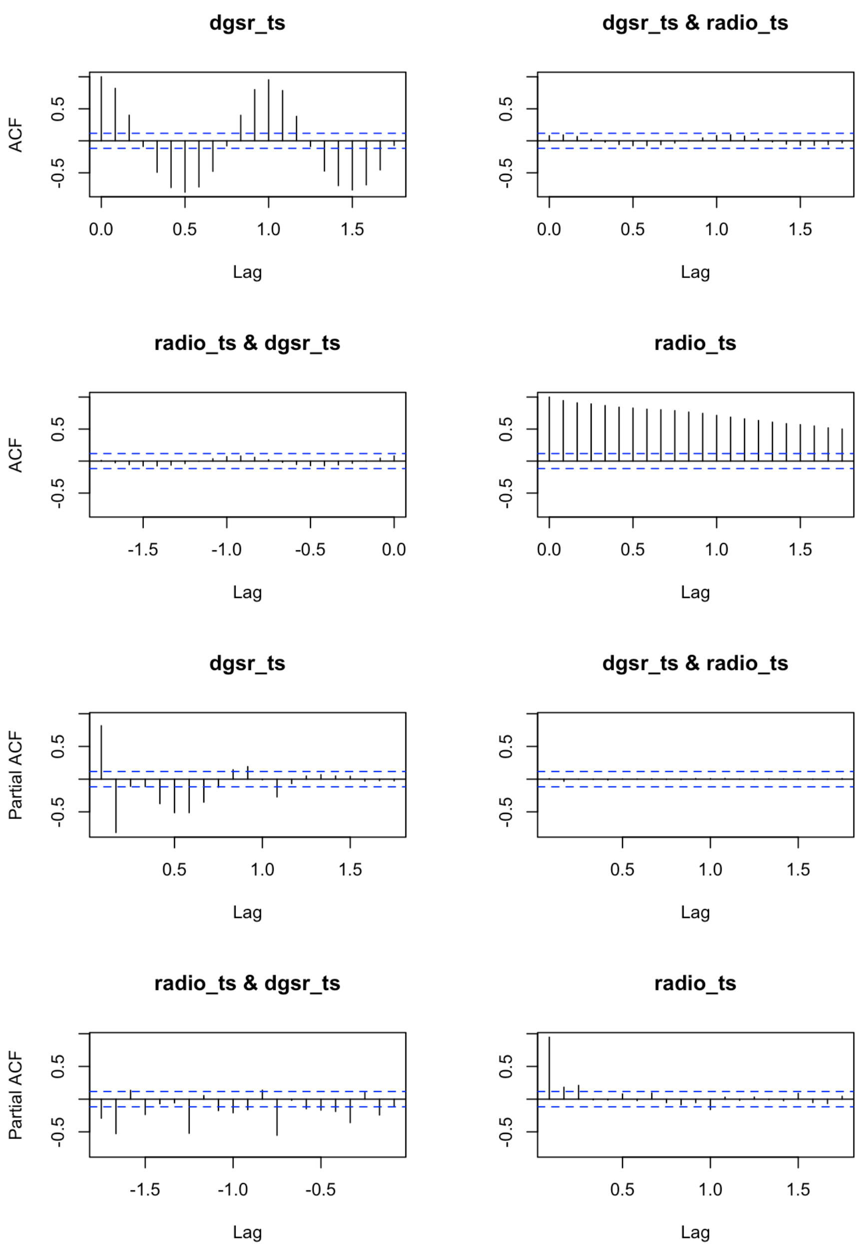
```
## Loading required package: lmtest
```

```
##
## VAR Estimation Results:
## =========================
## Endogenous variables: dgsr_ts, radio_ts
## Deterministic variables: const
## Sample size: 279
## Log Likelihood: -1449.252
## Roots of the characteristic polynomial:
## 0.9658 0.2887 0.2796 0.1188
## Call:
## VAR(y = solar_var_data, p = 2, season = 12L)
##
##
## Estimation results for equation dgsr_ts:
## ========================================
## dgsr_ts = dgsr_ts.l1 + radio_ts.l1 + dgsr_ts.l2 + radio_ts.l2 + const + sd1 + sd2 + sd3 + sd4 + sd5 + sd6 + sd
7 + sd8 + sd9 + sd10 + sd11
##
##              Estimate Std. Error t value Pr(>|t|)
## dgsr_ts.l1    0.073601   0.061296   1.201    0.231
## radio_ts.l1   0.002731   0.004516   0.605    0.546
## dgsr_ts.l2    0.066697   0.061174   1.090    0.277
## radio_ts.l2  -0.002792   0.004515  -0.618    0.537
## const         8.801587   0.871292  10.102  < 2e-16 ***
## sd1          -5.113921   0.702139  -7.283 3.79e-12 ***
## sd2         -15.939358   0.953761 -16.712  < 2e-16 ***
## sd3         -22.686886   0.952962 -23.807  < 2e-16 ***
## sd4         -25.852055   1.086496 -23.794  < 2e-16 ***
## sd5         -26.831583   1.399993 -19.166  < 2e-16 ***
## sd6         -26.489397   1.592744 -16.631  < 2e-16 ***
## sd7         -26.457743   1.645472 -16.079  < 2e-16 ***
## sd8         -25.381448   1.651489 -15.369  < 2e-16 ***
## sd9         -20.961020   1.600521 -13.096  < 2e-16 ***
## sd10        -12.795976   1.344285  -9.519  < 2e-16 ***
## sd11         -4.246262   0.802506  -5.291 2.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.916 on 263 degrees of freedom
## Multiple R-Squared: 0.9926,  Adjusted R-squared: 0.9922
## F-statistic:  2359 on 15 and 263 DF,  p-value: < 2.2e-16
##
##
## Estimation results for equation radio_ts:
## =========================================
## radio_ts = dgsr_ts.l1 + radio_ts.l1 + dgsr_ts.l2 + radio_ts.l2 + const + sd1 + sd2 + sd3 + sd4 + sd5 + sd6 + s
d7 + sd8 + sd9 + sd10 + sd11
##
##             Estimate Std. Error t value Pr(>|t|)
## dgsr_ts.l1   -0.56156    0.81941  -0.685  0.49374
## radio_ts.l1   0.76431    0.06037  12.660  < 2e-16 ***
## dgsr_ts.l2   -1.32348    0.81778  -1.618  0.10678
## radio_ts.l2   0.19422    0.06036   3.218  0.00145 **
## const        23.78553   11.64759   2.042  0.04214 *
## sd1          13.72701    9.38632   1.462  0.14481
## sd2          16.48532   12.75004   1.293  0.19716
## sd3          10.41965   12.73936   0.818  0.41415
## sd4          -9.37949   14.52446  -0.646  0.51899
## sd5         -23.82455   18.71534  -1.273  0.20414
## sd6         -30.38223   21.29208  -1.427  0.15479
## sd7         -28.97628   21.99696  -1.317  0.18889
## sd8         -30.97444   22.07739  -1.403  0.16180
## sd9         -25.74172   21.39604  -1.203  0.23002
## sd10        -25.32573   17.97064  -1.409  0.15993
## sd11        -10.02290   10.72803  -0.934  0.35102
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 12.25 on 263 degrees of freedom
## Multiple R-Squared: 0.9073,  Adjusted R-squared: 0.9021
## F-statistic: 171.7 on 15 and 263 DF,  p-value: < 2.2e-16
##
##
##
## Covariance matrix of residuals:
##          dgsr_ts radio_ts
```

```
## dgsr_ts   0.8391  -0.6654
## radio_ts -0.6654 149.9452
##
## Correlation matrix of residuals:
##           dgsr_ts radio_ts
## dgsr_ts   1.00000 -0.05933
## radio_ts -0.05933  1.00000
```

CCF for DGSR and Radio Influx (10.7cm)





For the VAR model, we can see that The model explains about 90.73% of the variance in the dependent variables, which is very high, indicating a good fit. The adjusted R-squared being at 90.21% confirms the model's robustness even after adjusting for the number of predictors. The F-statistic is 171.7 with a p-value less than $2.2*10^{-16}$, showing that the model is statistically significant as a whole. For sd8, sd9, sd10, and sd11, these are coefficients for different time lags or possibly seasonal differencing. None of the coefficients are statistically significant (all p-values > 0.05). This suggests that these specific lags do not have a substantial impact on the dependent variables in the model. The residual standard error (12.25) indicates the average deviation of the observed values from the model's predicted values. The covariance matrix indicates that the relationship between the residuals of the two time series (dgsr_ts and radio_ts). The negative covariance (-0.6654) between the two suggests a moderate inverse linear relationship. The correlation coefficient (-0.059) between the residuals of dgsr_ts and radio_ts is quite low, implying minimal residual correlation between the two time series.

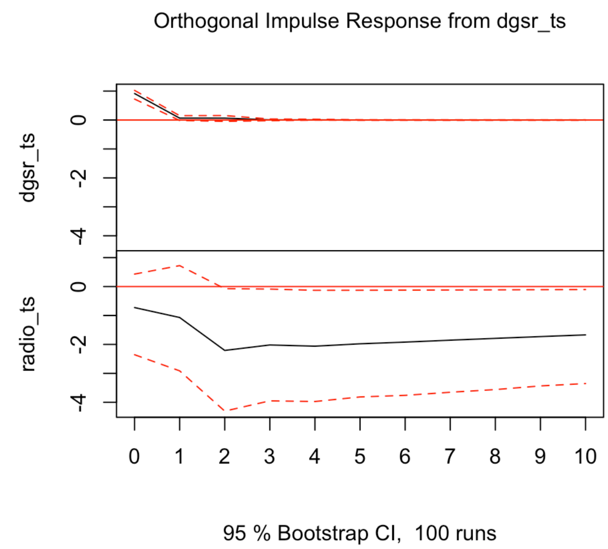
For the CCF plot, there seems to be some cross-correlation between DGSR and Radio Flux at certain lags, both positive and negative, but most correlations are within the significance bounds.

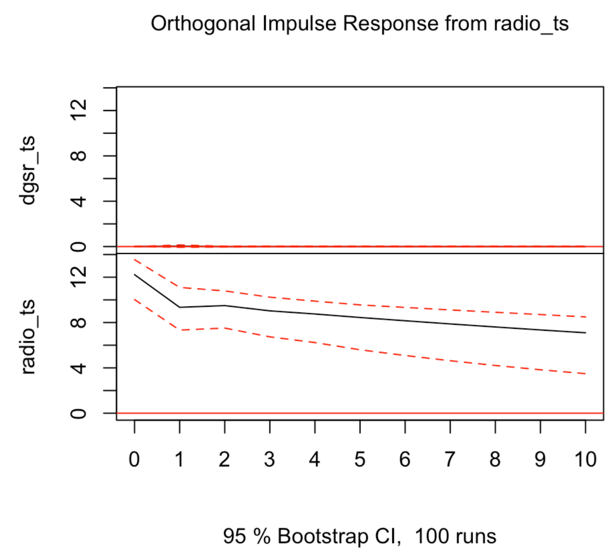
With the ACF plots, the dgsr_ts model displays significant autocorrelation at lag 1 but drops below the significance bounds for higher lags, suggesting minimal autocorrelation after the first lag. The radio_ts model shows significant autocorrelation across multiple lags, indicating persistence or memory in the Radio Flux time series. Overall, there is minimal significant cross-autocorrelation between the two variables for most lags, suggesting a weak linear relationship in the cross-lagged structure.

With the PACF plots, The PACF plot for dgsr_ts shows a significant spike at lag 1, which suggests that this time series has a strong autoregressive component, meaning the value at time t is closely correlated with the previous time step (lag 1). The values drop to near-zero at higher lags, indicating that the impact of prior lags decreases quickly and doesn't have a prolonged autoregressive influence. This might imply that an AR(1) model could be a good fit for this time series. This plot might represent the cross-PACF between radiation_ts and radio_ts, though without explicit labels it's an assumption. For the dgsr_ts and radio_ts PACF plot, the values are close to zero, suggesting little to no correlation between the two series at the lags tested. This indicates that the two variables may not directly influence each other, or that any relationship may be weak or non-linear. For the radio_ts and dgsr_ts plot, it is similar to the previous cross-PACF plot, with reversed order. Again, the values are close to zero, which reinforces the idea that these two time series might not be linearly related or have minimal interaction at the lags considered. The PACF plot for radio_ts shows a signiciant spike at lag 1, similar to dgsr_ts. This also implies an autoregressive component at lag 1, suggesting that radio_ts might similarly be modeled with an AR(1) structure. The lack of significant correlations at higher lags indicates that any influence from past values dissipates quickly.

(m) Compute, plot, and interpret the respective impulse response functions.

```
plot(irf(solar_var_model))
```





DGSR Impulse Response Function:

Looking at the plot, we can see that a positive shock to DGSR has an immediate and strong effect on itself, with the response starting at a high positive value. The response decays slightly over time but remains relatively stable, suggesting that a shock to DGSR has a persistent effect on itself over the observed time horizon.

A shock to DGSR causes a small initial decline in radio flux, followed by a gradual decrease over the first few lags. The response remains negative for the first few periods but appears to start returning to zero toward the end of the horizon. The confidence intervals indicate that this effect may not be very strong, as the response stays within or near the confidence bands, suggesting that the influence of radio flux on DGSR is weak or short-lived.

Radio Flux (10.7cm) Impulse Response Function:

Looking at the plot, we can see that a positive shock to radio flux initially increases radio_ts substantially, with the effect gradually decreasing over time. This pattern suggests that the impact of a shock to radio flux on itself is strong initially but diminishes over the period, though it doesn't return to zero immediately, indicating some persistence.

Additionally, a shock to radio flux has minimal impact on DGSR, as the response remains close to zero over time. The confidence intervals suggest that the response is not statistically significant, implying that shocks to radio flux do not meaningfully influence radio flux

(n) Perform a Granger-Causality test on your variables and discuss your results from the test.

```
grangertest(dgsr_ts ~ radio_ts, order = 2)
```

```
## Granger causality test
##
## Model 1: dgsr_ts ~ Lags(dgsr_ts, 1:2) + Lags(radio_ts, 1:2)
## Model 2: dgsr_ts ~ Lags(dgsr_ts, 1:2)
##   Res.Df Df      F Pr(>F)
## 1    274
## 2    276 -2 1.5512 0.2138
```

```
grangertest(radio_ts ~ dgsr_ts, order = 2)
```

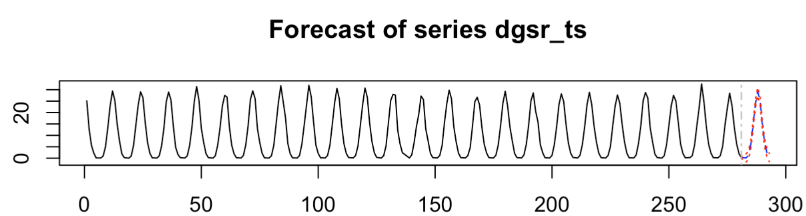
```
## Granger causality test
##
## Model 1: radio_ts ~ Lags(radio_ts, 1:2) + Lags(dgsr_ts, 1:2)
## Model 2: radio_ts ~ Lags(radio_ts, 1:2)
##   Res.Df Df      F  Pr(>F)
## 1    274
## 2    276 -2 3.2051 0.04208 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

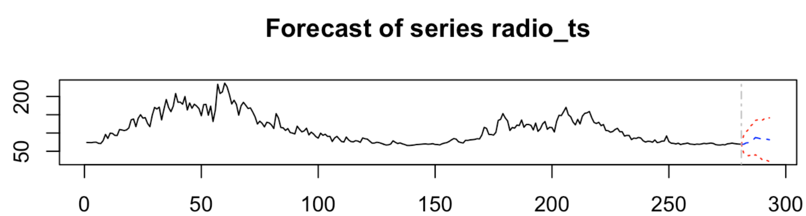
Let the null hypothesis be that the lags of radio_ts do not Granger cause dgsr_ts, or in other words, radio_ts is useful in forecasing dgsr_ts. Looking at the results, we can see that for the that the p-value is 0.2138 which is greater than the common significance level of 0.05. Therefore, we fail to reject the null hypothesis. This suggests that radio_ts does not Granger-cause dgsr_ts.

Additionally, let the null hypothesis be that the lags of dgsr_ts do not Granger cause radio_ts, or in other words, dgsr_ts is useful in forecasing radio_ts. Looking at the results, we can see that for the that the p-value is 0.04208 which is less than the common significance level of 0.05. Therefore, we can reject the null hypothesis. This indicates that dgsr_ts Granger-causes radio_ts at the 5% significance level.

(o) Use your VAR model to forecast 12-steps ahead. Your forecast should include the respective error bands. Comment on the differences between the VAR forecast and the other ones obtained using the different methods.

```
solar_var_forecasts = predict(object = solar_var_model, n.ahead = 12)
plot(solar_var_forecasts)
```





In this model, we can see that the forecast of dgsr_ts exhibits a clear periodic pattern with consistent peaks and troughs, indicating a strong seasonality. The forecasting model seems to capture this cyclical pattern well, suggesting that dgsr_ts is highly regular and predictable. In the forecasted region (right end of the plot, around time 300), the model predicts the continuation of this seasonal pattern with a peak and a dip. This regularity likely reflects a seasonal influence, which could correspond to something like solar radiation or another cyclical natural phenomenon.

For radio_ts, we can see that the VAR model displays a forecast that shows a more irregular and decreasing trend over time, with some fluctuations. Unlike dgsr_ts, there is no strong periodic pattern; instead, it appears to fluctuate at lower amplitudes with an overall downward trend. The forecasted values in the right end (near time 300) have a much smaller predicted range, indicating limited expected change. The forecast intervals (in dashed lines) show some uncertainty but are narrower than those in dgsr_ts, which might suggest that radio_ts has more stable and predictable fluctuations at this point in the time series.

Comparing the VAR model forecasts to the other ones obtained using different methods from earlier, we can see that the forecasts are very similar. Like the original DGSR model, the VAR model displays a continuing seasonal pattern that increases and then decreases for the next 12 steps. This correlates with the seasonal pattern that has been happening already. Additionally, for radio flux, the VAR forecast also comes into a steady state like the original model from earlier. Additionally, although almost steady, there is still a sign of a slight decrease like what we saw in the other model. The only main difference is that the VAR model has slightly larger error band ranges than compared to the original models.

III. Conclusions and Future Work

For this project, we wanted to see if there was any patterns seen in solar radiation and radio influx (10.7cm) over the years of 1997 to 2020. The overall hypothesis was that solar radiation and radio influx (10.7cm) increases over the years. Therefore, we need to find policies to mitigate these increases so that we can prevent further climate change from happening. We should also be able to see which periods of time tend to see higher solar radiation and radio influx, so we can know when is it better to apply these policies. Ideally, scientists should be able to forecast when is it appropriate to reduce our technology usage and resources in order to balance the increased solar radiation and radio influx that are present.

From the analysis above, we found out that there is actually not much of a trend for solar radiation, and there is actually a slight decreasing trend in radio influx. However, there is high seasonal patterns in solar radiation and a possible cyclic behavior happening for radio influx. The VAR model also showcases this behavior, with larger error band intervals that increases our confidence of the values. Overall, this means that we should still look into policies that help reduce energy usage in specific periods of time in order to prevent further climate change from happening.

Additionally, for future work, there is definitely many aspects of our study that can be improved on. For one, our CUSUM plots show trends that suggest the need for model adjustments, such as including additional explanatory variables or allowing for time-varying parameters. Additionally, in regards to the diagnostic plots, we should consider further adjustments to account for remaining autocorrelation (e.g., adding more seasonal AR or MA terms). We should also address potential non-normality by investigating the sources of extreme residuals, possibly through outlier detection or robust modeling techniques. Finally, we should also re-evaluate the stationarity of the data, especially for the radio flux model, to ensure that differencing has adequately stabilized the mean and variance.

IV. References

Energy.gov, (https://www.energy.gov/eere/solar/solar-radiation-basics (https://www.energy.gov/eere/solar/solar-radiation-basics))

Iberdrola, (https://www.iberdrola.com/social-commitment/solar-radiation# (https://www.iberdrola.com/social-commitment/solar-radiation#):~:text=Solar%20radiation%20is%20the%20energy%20emitted%20by%20the%20Sun%2C%20which,influences%20atmospheric%20and%20climat

UCAR Center for Science Education, (https://scied.ucar.edu/learning-zone/sun-space-weather/sun-and-climate-change (https://scied.ucar.edu/learning-zone/sun-space-weather/sun-and-climate-change))

National Snow and Ice Data Center, (https://nsidc.org/learn/parts-cryosphere/sea-ice/quick-facts-about-sea-ice# (https://nsidc.org/learn/parts-cryosphere/sea-ice/quick-facts-about-sea-ice#):~:text=As%20explained%20above%2C%20sea%20ice,the%20Arctic's%20overall%20energy%20balance.)

Imperial College London, (https://www.imperial.ac.uk/media/imperial-college/grantham-institute/public/publications/briefing-papers/Solar-Influences-on-Climate---Grantham-BP-5.pdf (https://www.imperial.ac.uk/media/imperial-college/grantham-institute/public/publications/briefing-papers/Solar-Influences-on-Climate---Grantham-BP-5.pdf))

V. R Source Code

```
# Load libraries
library(dplyr)
library(tidyr)
library(zoo)

# Load the DGSR data
radiation_data <- read.csv("DGSR_Data.csv")
head(radiation_data, n = 24) # Preview of the data set

# Calculate the monthly average of daily estimated DGSR values
monthly_avg_dgsr <- radiation_data %>%
  group_by(Year, Month) %>%
  summarize(Monthly_Avg_DGSR = mean(`Estimated.DGSR`, na.rm = TRUE))

# Create a complete sequence of Year and Month with NA for missing entries (including June)
complete_monthly_avg_dgsr <- monthly_avg_dgsr %>%
  complete(Month = 1:12, fill = list(Monthly_Avg_DGSR = NA))

# Interpolation to help impute reasonable values for June (Month 6)
complete_monthly_avg_dgsr <- complete_monthly_avg_dgsr %>%
  arrange(Year, Month) %>%
  mutate(Monthly_Avg_DGSR = na.approx(Monthly_Avg_DGSR, na.rm = FALSE))

# Removing the rest of NA values (before March 1989 and after May 2020)
complete_monthly_avg_dgsr <- complete_monthly_avg_dgsr %>%
  filter(is.na(Monthly_Avg_DGSR) == FALSE)

# Preview the imputed data
head(complete_monthly_avg_dgsr, n = 24)

# Do the same with the radio flux data
radio_data <- read.csv("daily_solar_data.csv")

# Convert Date to Date type if necessary
radio_data$Date <- as.Date(radio_data$Date, format="%Y-%m-%d")

# Calculate monthly averages
radio_monthly_data <- radio_data %>%
  mutate(Month = format(Date, "%Y-%m")) %>%
  group_by(Month) %>%
  summarise(Avg_Radio_Flux = mean(Radio.Flux.10.7cm, na.rm = TRUE))

# Preview the result
head(radio_monthly_data, n = 12)

# Get data from January 1997 to May 2020
complete_monthly_avg_dgsr <- complete_monthly_avg_dgsr %>%
  filter(Year >= 1997)
radio_monthly_data <- radio_monthly_data[1:281,]

# (a)

# Load libraries
library(ggplot2)
library(forecast)

# Create a Date column for proper time series format
complete_monthly_avg_dgsr$Date <- as.Date(paste(complete_monthly_avg_dgsr$Year, complete_monthly_avg_dgsr$Month,
1, sep = "-"))

# Convert it to a time series object
dgsr_ts <- ts(complete_monthly_avg_dgsr$Monthly_Avg_DGSR, start = c(min(complete_monthly_avg_dgsr$Year), 1), freq
uency = 12)

# Time Series Plot (DGSR)
autoplot(dgsr_ts, col = "maroon") +
  labs(title = "Time Series of Monthly Average DGSR in Antractica", x = "Time", y = "Average DGSR")

# ACF and PACF plots (DGSR)
acf(dgsr_ts, main = "ACF of Monthly Average DGSR in Antractica")
pacf(dgsr_ts, main = "PACF of Monthly Average DGSR in Antractica")

# Time series object for radio_flux
radio_ts <- ts(radio_monthly_data$Avg_Radio_Flux, start = c(min(complete_monthly_avg_dgsr$Year), 1), frequency =
12)
```

```
# Time Series Plot (Radio Flux)
autoplot(radio_ts, col = "forestgreen") +
  labs(title = "Time Series of Monthly Average Solar Radio Flux (10.7cm)", x = "Time", y = "Average Solar Radio F
lux (10.7cm)")

# ACF and PACF plots ((Radio Flux))
acf(dgsr_ts, main = "ACF of Monthly Average Solar Radio Flux (10.7cm)")
pacf(dgsr_ts, main = "PACF of Monthly Average Solar Radio Flux (10.7cm)")

# (b)

# Perform STL decomposition (DGSR)
radiation_stl_decomp <- stl(dgsr_ts, s.window = "periodic")

# Plot the STL decomposition (DGSR)
autoplot(radiation_stl_decomp) +
  labs(title = "STL Decomposition of Monthly Average DGSR in Antarctica")

# Perform STL decomposition (Radio Flux)
radio_stl_decomp <- stl(radio_ts, s.window = "periodic")

# Plot the STL decomposition (Radio Flux)
autoplot(radio_stl_decomp) +
  labs(title = "STL Decomposition of Monthly Average Solar Radio Flux (10.7cm)")

# (c)

# Fit a seasonal ARIMA model to capture trend and seasonality (DGSR)
radiation_sarima_model <- auto.arima(dgsr_ts, seasonal = TRUE)

# Summary of the model (DGSR)
summary(radiation_sarima_model)

# Fit a seasonal ARIMA model to capture trend and seasonality (Radio Flux)
radio_sarima_model <- auto.arima(radio_ts, seasonal = TRUE)

# Summary of the model (Radio Flux)
summary(radio_sarima_model)

# (e)

# Get residuals vs. fitted values and storing it in a data frame (DGSR)
radiation_residuals_df <- data.frame(Fitted_Values = radiation_sarima_model$fitted, Residuals = radiation_sarima_
model$residuals)

# Plot of the residuals vs. fitted values (DGSR)
ggplot(radiation_residuals_df, aes(x = as.numeric(Fitted_Values), y = as.numeric(Residuals))) +
  geom_line(color = "navy", alpha = 0.9) +
  geom_point(color = "navy", alpha = 0.9) +
  ggtitle("Residuals vs Fitted Values (DGSR)") +
  xlab("Fitted Values") +
  ylab("Residuals")

# Get residuals vs. fitted values and storing it in a data frame (Radio Flux)
radio_residuals_df <- data.frame(Fitted_Values = radio_sarima_model$fitted, Residuals = radio_sarima_model$residu
als)

# Plot of the residuals vs. fitted values (Radio Flux)
ggplot(radio_residuals_df, aes(x = as.numeric(Fitted_Values), y = as.numeric(Residuals))) +
  geom_line(color = "purple", alpha = 0.9) +
  geom_point(color = "purple", alpha = 0.9) +
  ggtitle("Residuals vs Fitted Values (Radio Flux 10.7cm)") +
  xlab("Fitted Values") +
  ylab("Residuals")

# (f)

par(mfrow = c(2, 2))

# Plot ACF and PACF of residuals (DGSR)
acf(radiation_residuals_df$Residuals, main = "ACF of Model Residuals (DGSR)")
pacf(radiation_residuals_df$Residuals, main = "PACF of Model Residuals (DGSR)")

# Plot ACF and PACF of residuals (Radio Flux)
acf(radio_residuals_df$Residuals, main = "ACF of Model Residuals (Radio Flux 10.7cm)")
pacf(radio_residuals_df$Residuals, main = "PACF of Model Residuals (Radio Flux 10.7cm)")

# (g)
```

```
par(mfrow = c(1, 2))

# Plot the CUSUM (DGSR)
plot(efp(radiation_residuals_df$Residuals~1, type = "Rec-CUSUM"), main = "CUSUM Plot of Model Residuals (DGSR)")

# Plot the CUSUM (Radio Flux)
plot(efp(radio_residuals_df$Residuals~1, type = "Rec-CUSUM"), main = "CUSUM Plot of Model Residuals (Radio Flux 1
0.7cm)")

# (h)

# DGSR
checkresiduals(radiation_sarima_model)

# Radio Flux
checkresiduals(radio_sarima_model)

# (i)

# Forecast 12 months ahead (DGSR)
radiation_forecast_12 <- forecast(radiation_sarima_model, h = 12)

# Plot the forecast (DGSR)
autoplot(radiation_forecast_12) +
  labs(title = "12-Month Ahead Forecast with SARIMA Model (DGSR)", y = "Average DGSR")

# Forecast 12 months ahead (Radio Flux)
radio_forecast_12 <- forecast(radio_sarima_model, h = 12)

# Plot the forecast (Radio Flux)
autoplot(radio_forecast_12) +
  labs(title = "12-Month Ahead Forecast with SARIMA Model (Radio Flux 10.7cm)", y = "Average Radio Flux (10.7c
m)")

# (j)

# Calculate the MAPE for the 12-step forecast (DGSR)
actuals <- window(dgsr_ts, start = end(dgsr_ts) - c(0, 11))
radiation_mape <- mean(abs((as.numeric(actuals) - as.numeric(radiation_forecast_12$mean)) / actuals)) * 100
print(paste("DGSR MAPE:", radiation_mape))

# Fit auto.arima model and forecast (DGSR)
radiation_auto_model <- auto.arima(dgsr_ts)
radiation_forecast_auto <- forecast(radiation_auto_model, h = 12)
radiation_auto_mape <- mean(abs((as.numeric(actuals) - as.numeric(radiation_forecast_auto$mean)) / actuals)) * 10
0
print(paste("DGSR Auto ARIMA Model MAPE:", radiation_auto_mape))

# Calculate the MAPE for the 12-step forecast (Radio Flux)
actuals2 <- window(radio_ts, start = end(radio_ts) - c(0, 11))
radio_mape <- mean(abs((as.numeric(actuals2) - as.numeric(radio_forecast_12$mean)) / actuals2)) * 100
print(paste("Radio Flux MAPE:", radio_mape))

# Fit auto.arima model and forecast (Radio Flux)
radio_auto_model <- auto.arima(radio_ts)
radio_forecast_auto <- forecast(radio_auto_model, h = 12)
radio_auto_mape <- mean(abs((as.numeric(actuals2) - as.numeric(radio_forecast_auto$mean)) / actuals2)) * 100
print(paste("Radio Flux Auto ARIMA Model MAPE:", radio_auto_mape))

# (k)

# Combined forecasts (DGSR)
radiation_combined_forecast <- (radiation_forecast_12$mean + radiation_forecast_auto$mean) / 2
radiation_mape_combined <- mean(abs((as.numeric(actuals) - as.numeric(radiation_combined_forecast)) / actuals)) *
100
print(paste("DGSR Combined MAPE:", radiation_mape_combined))

# Combined forecasts (DGSR)
radio_combined_forecast <- (radio_forecast_12$mean + radio_forecast_auto$mean) / 2
radio_mape_combined <- mean(abs((as.numeric(actuals2) - as.numeric(radio_combined_forecast)) / actuals2)) * 100
print(paste("Radio Flux Combined MAPE:", radio_mape_combined))

# (l)

# Load var library
library(vars)

# Create a two-variable time series matrix
solar_var_data <- cbind(dgsr_ts, radio_ts)
```

```r
# Fit the VAR model
solar_var_model <- VAR(solar_var_data, p = 2, season = 12)

# Summary of the VAR model
summary(solar_var_model)

# Relevant Plots

# CCF
ccf(dgsr_ts, radio_ts, main = "CCF for DGSR and Radio Influx (10.7cm)")

# ACF and PACF Plots
acf(solar_var_data)
pacf(solar_var_data)

# (m)

plot(irf(solar_var_model))

# (n)

grangertest(dgsr_ts ~ radio_ts, order = 2)
grangertest(radio_ts ~ dgsr_ts, order = 2)

# (o)

solar_var_forecasts = predict(object = solar_var_model, n.ahead = 12)
plot(solar_var_forecasts)
```