STAT-S 650 TIME SERIES ANALYSIS

Forecasting Disturbance Storm-Time Index

Final Project Report

Spring 2024

Authored by:

Ritika Shrivastava

Introduction

Information provided by the Earth's magnetic field is of primary importance for navigation and the pointing of technical devices such as antennas, satellites and smartphones. The excellent results from this challenge hold immediate promise for the space weather community.
— Manoj Nair, Research Scientist, NOAA/CIRES Geomagnetism Group

The transfer of energy from solar wind to Earth's magnetic field can cause massive geomagnetic storms, wreaking havoc on key infrastructure systems like GPS, satellite communication, and electric power transmission.

The severity of these geomagnetic storms is measured by the Disturbance Storm-time Index, or Dst. In the past three decades, empirical, physics-based, and machine learning models have made advances in forecasting Dst from real-time solar wind data. However, predicting extreme geomagnetic events remains especially hard, and robust solutions are needed that can work with raw, real-time data streams under realistic conditions like sensor malfunctions and noise.

The aim of this project is to develop a model for forecasting *Dst* (Disturbance Storm-Time Index) that 1) pushes the boundary of predictive performance 2) under operationally viable constraints 3) using specified real-time solar-wind data feeds. The goal is to predict the Disturbance Storm-Time Index (*Dst*), a measure of magnetic activity, from the provided data up to the time of prediction.

Dst values are measured by 4 ground-based observatories near the equator. These values are then averaged to provide a measurement of *Dst* for any given hour. However, these values are not always provided in a timely manner.

Dataset Description

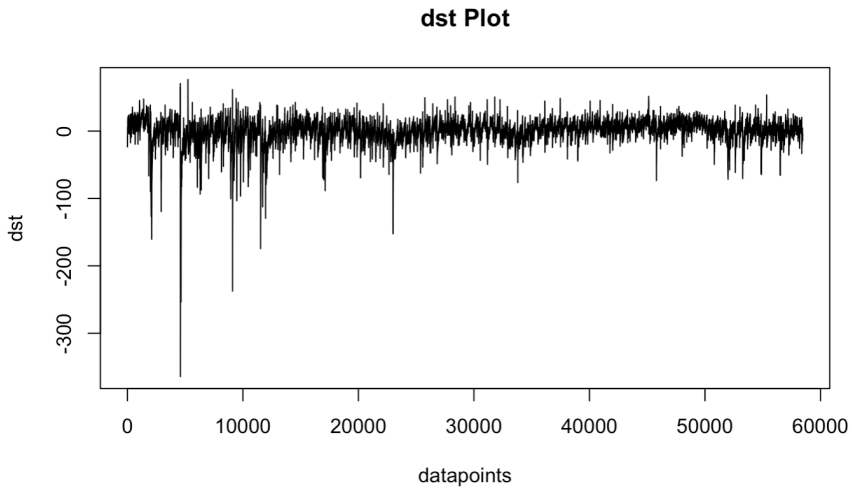
- bx_gse - Interplanetary-magnetic-field (IMF) X-component in geocentric solar ecliptic (GSE) coordinate (nanotesla (nT))
- by_gse - Interplanetary-magnetic-field Y-component in GSE coordinate (nT)
- bz_gse - Interplanetary-magnetic-field Z-component in GSE coordinate (nT)
- theta_gse - Interplanetary-magnetic-field latitude in GSE coordinates (defined as the angle between the magnetic vector B and the ecliptic plane, being positive when B points North) (degrees)
- phi_gse - Interplanetary-magnetic-field longitude in GSE coordinates (the angle between the projection of the IMF vector on the ecliptic and the Earth–Sun direction) (degrees)
- bx_gsm - Interplanetary-magnetic-field X-component in geocentric solar magnetospheric (GSM) coordinate (nT)
- by_gsm - Interplanetary-magnetic-field Y-component in GSM coordinate (nT)
- bz_gsm - Interplanetary-magnetic-field Z-component in (GSM) coordinate (nT)
- theta_gsm - Interplanetary-magnetic-field latitude in GSM coordinates (degrees)
- phi_gsm - Interplanetary-magnetic-field longitude in GSM coordinates (degrees)
- bt - Interplanetary-magnetic-field component magnitude (nT)
- density - Solar wind proton density (N/cm^3)
- speed - Solar wind bulk speed (km/s)
- temperature - Solar wind ion temperature (Kelvin)

Hypothesis to test

1. Target variable dst follows a seasonal trend over the years.
2. Target variable dst is influenced by the lagged values of features associated with the solar winds.
3. Target variable dst depends on the previous day's dst.

Time Series Plot

Dst (Target Variable) Time Series Plot



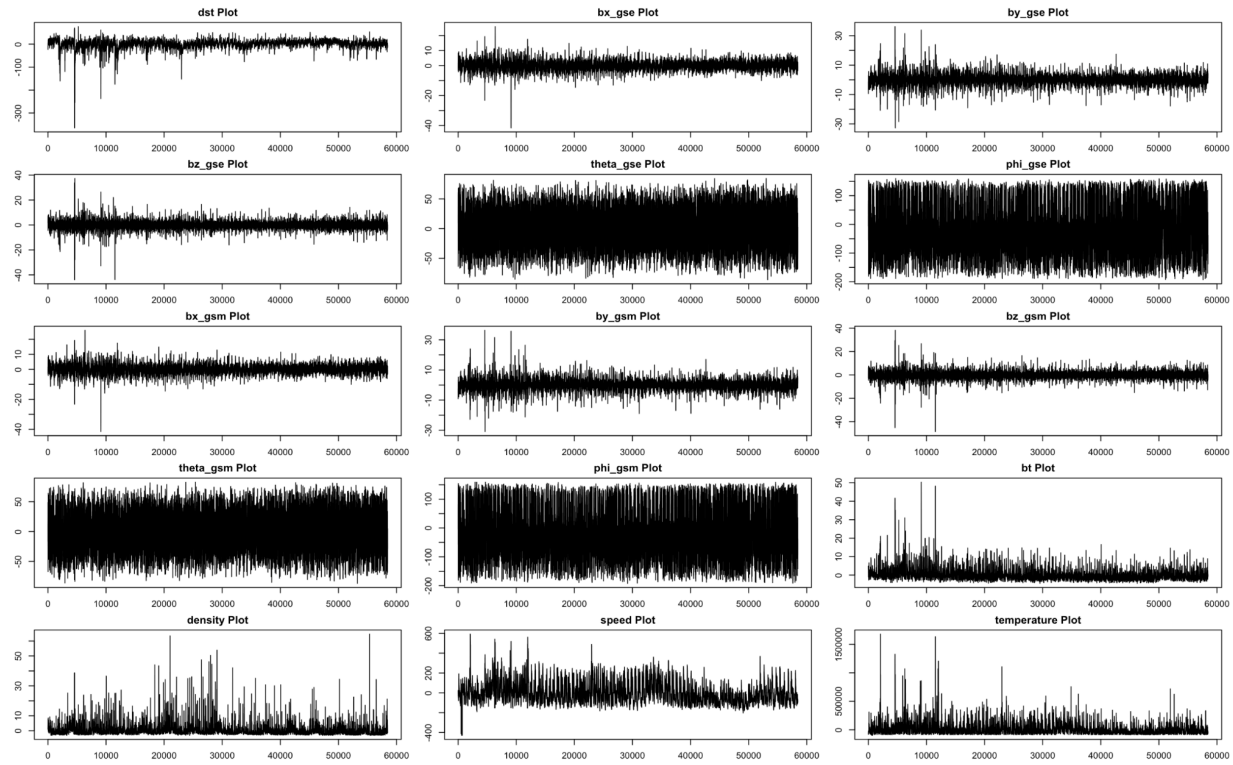
The plot shows volatility in the data, with many spikes particularly evident in the negative direction. The range of the 'dst' values on the y-axis goes from 67 to -374.

The distribution of spikes does not seem to follow a regular pattern, indicating that there is no clear periodicity based solely on this graph. The data points look more like "noise," as we do not see a clear trend upwards or downwards, nor do we see a repeating pattern. Since dst is a geomagnetic disturbance, the variability could suggest the presence of random or chaotic events affecting the readings.

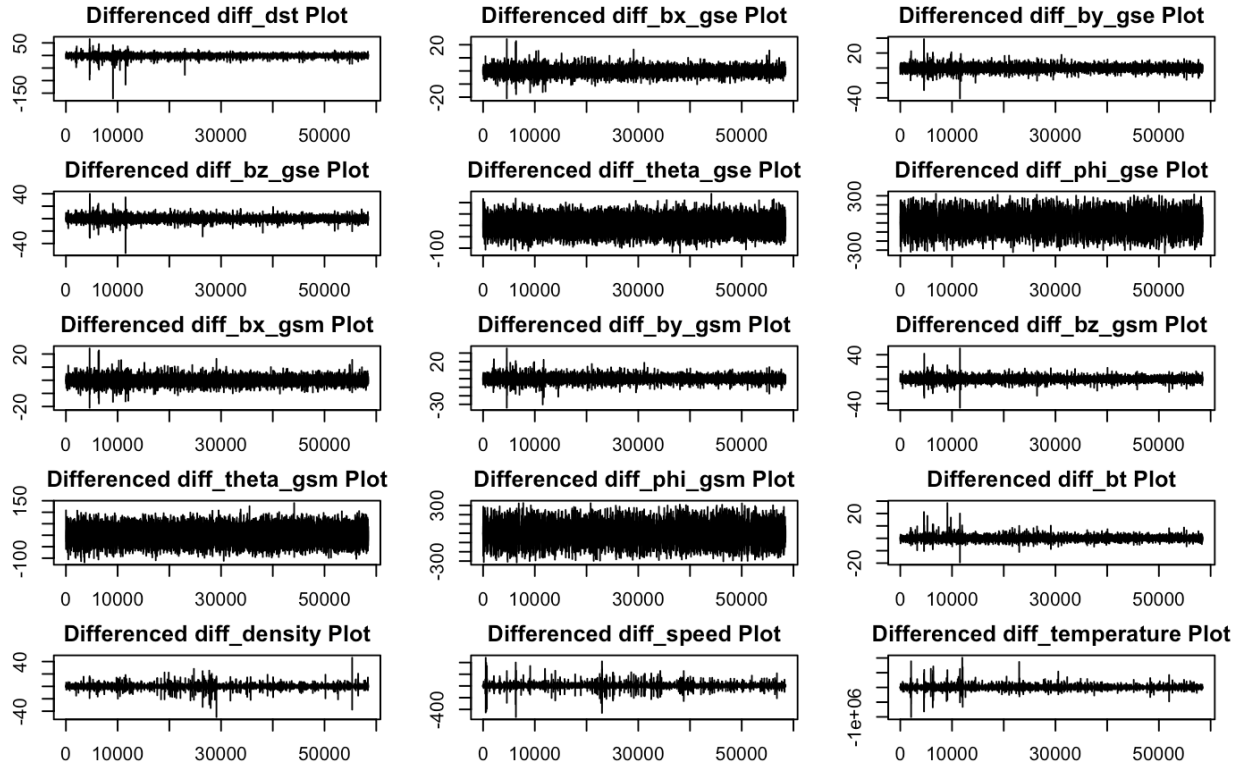
Features Time Series Plot

- bx_gse:: The x-component of the magnetic field in geocentric solar ecliptic (GSE) coordinates, which appears fairly stable.
- by_gse: The y-component in GSE coordinates, showing some variability.
- bz_gse: The z-component in GSE coordinates, also displaying variability that could be associated with solar wind changes.
- theta_gse: The polar angle in GSE coordinates, showing a dense and highly variable dataset.
- phi_gse: The azimuthal angle in GSE coordinates, with a high-density plot indicating rapid changes or a large range of values.

- bx_gsm: The x-component of the magnetic field in geocentric solar magnetospheric (GSM) coordinates, which appears stable similar to the GSE equivalent.
- by_gsm: The y-component in GSM coordinates, showing variability.
- bz_gsm: The z-component in GSM coordinates, displaying a pattern that may be indicative of consistent fluctuations or cyclic behavior.
- theta_gsm: The polar angle in GSM coordinates, with significant variability and dense plotting.
- phi_gsm: The azimuthal angle in GSM coordinates, displaying a wide range of values with dense plotting.
- bt: The total magnetic field strength, with spikes that might correspond to solar events or interplanetary shocks.
- density: The plasma density, which has large spikes potentially indicating solar flare events or coronal mass ejections.
- speed: The solar wind speed, showing variability and some prominent peaks, possibly from high-speed streams emanating from coronal holes.
- temperature: The solar wind temperature, with extreme values in some instances which could correlate with solar activity.



Features Time Series Plot after Differencing



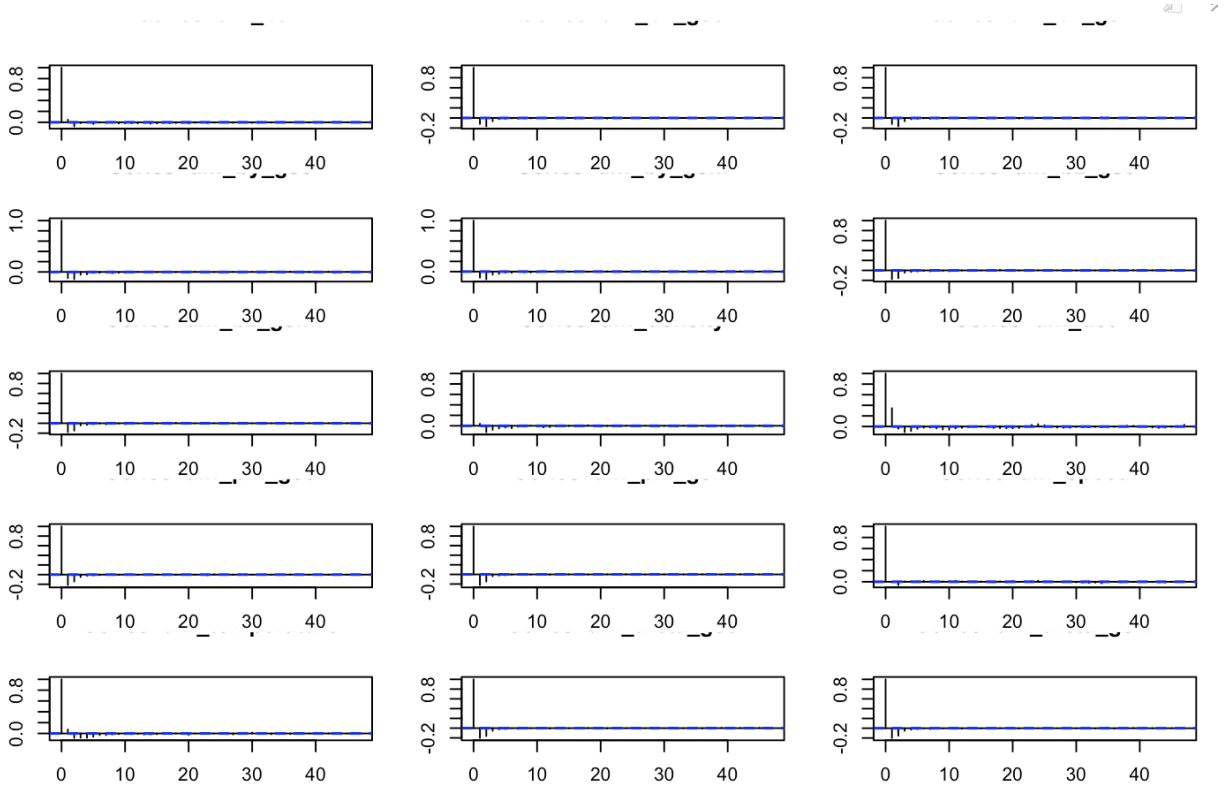
- Differenced dst: The changes in the dst variable appear to fluctuate around zero without a discernible pattern or trend, which is typical after differencing a time series that has a trend.
- Differenced bx_gse, by_gse, and bz_gse: The magnetic field components in GSE coordinates (bx_gse, by_gse, bz_gse) show variability around zero, similar to dst. This indicates that the differencing may have removed any linear trend, and the resulting series is mean stationary.
- Differenced theta_gse and phi_gse: Both angular measurements in GSE coordinates (theta_gse and phi_gse) exhibit high-frequency fluctuations after differencing, with no visible trend, which is consistent with a differenced stationary process.
- Differenced bx_gsm, by_gsm, and bz_gsm: For the GSM coordinates, the magnetic field components also show variations around zero. This suggests that these series also have no strong trends after differencing.
- Differenced theta_gsm and phi_gsm: The differenced angular measurements in GSM coordinates are similar to their GSE counterparts, with a lot of variability but no clear trend, indicating a mean stationary

process.

- Differenced bt: The total magnetic field strength (bt) exhibits a similar pattern to the other magnetic field components, with no apparent trend after differencing.
- Differenced density: Plasma density changes are scattered around zero, but with some larger spikes, suggesting occasional larger changes from one measurement to the next.
- Differenced speed: Solar wind speed changes show a similar pattern to density, with some outliers that may indicate large changes or events.
- Differenced temperature: The temperature of the solar wind shows the most pronounced spikes after differencing, indicating significant variability and occasional large shifts from one time point to another.

Autocorrelations and Cross Correlations

Autocorrelation Plots



The autocorrelation function (ACF) plots illustrate the temporal behavior of various differenced variables encompassing magnetic field components, plasma density, solar wind parameters, and geomagnetic indices. Across this array of measurements, a consistent trend emerges: there is a conspicuous absence of significant autocorrelation at any lag. This absence implies a lack of systematic relationship between observations at different time intervals, indicating that the variables exhibit a degree of randomness or stochastic behavior. Such findings are crucial in understanding the dynamic nature of space weather phenomena, as they suggest that the observed fluctuations in these parameters do not persistently influence future states. This underscores the complex and inherently unpredictable nature of space weather dynamics, emphasizing the need for continuous monitoring and analysis to comprehend and forecast its impacts on Earth and space-based systems.

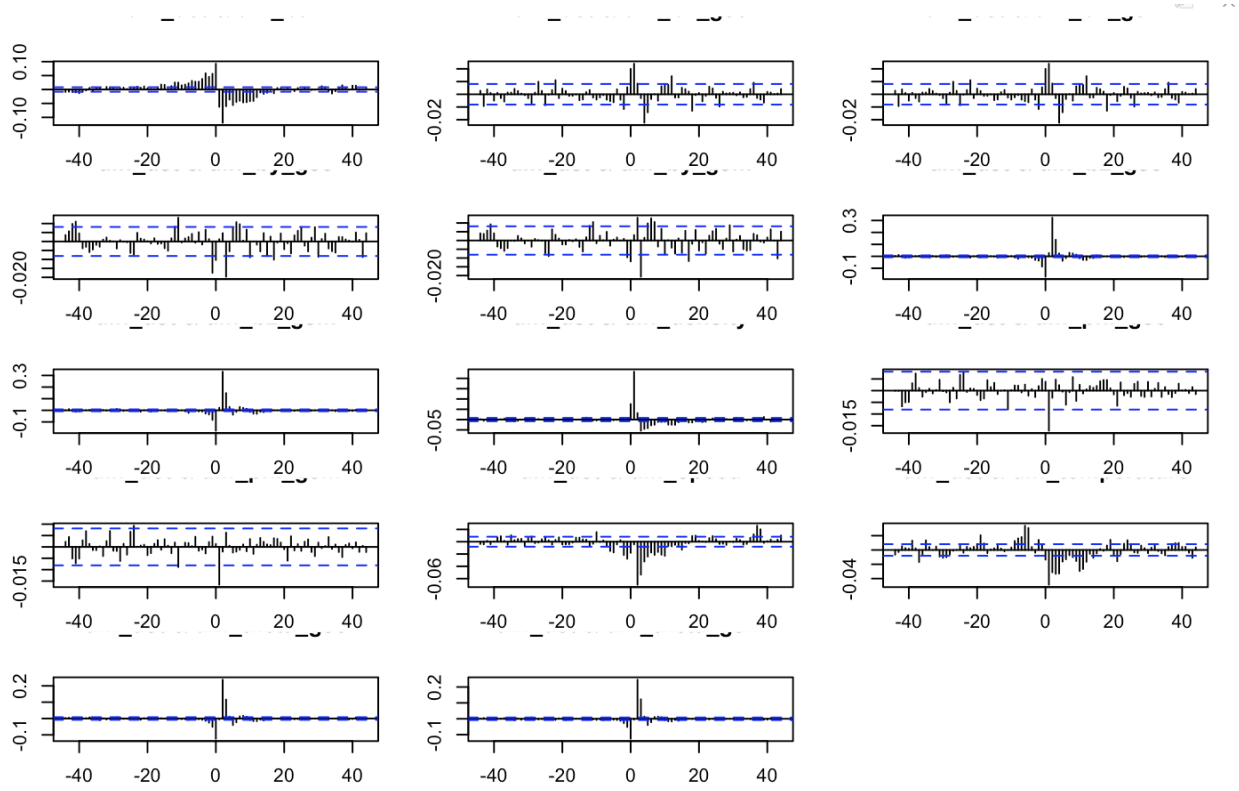
Cross Correlation Plots

Across all plots, the cross-correlations are minimal and generally within the confidence bounds, suggesting no significant lead or lag relationships between the differenced DST time series and the other differenced variables.

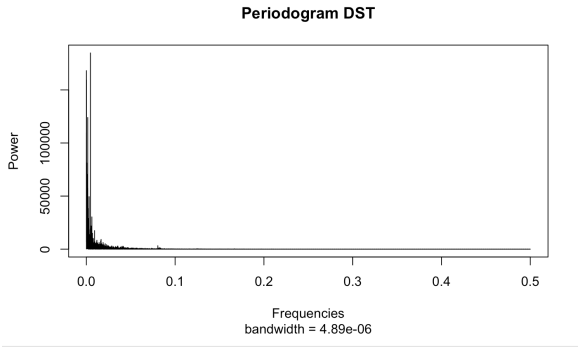
This could indicate that the changes in DST do not have a direct or delayed correlation with changes in the other measured variables, at least not within the range of lags presented in these plots.

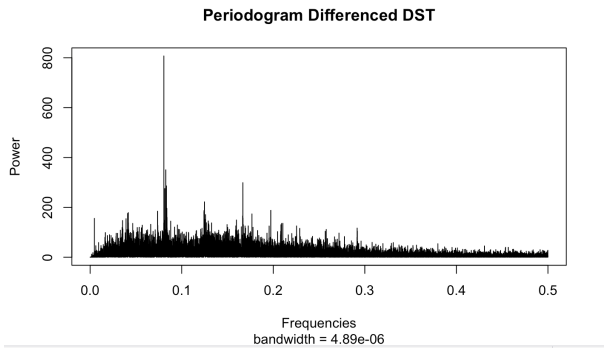
If these variables were expected to be related, the lack of correlation could suggest that any relationship is not linear or that it might occur at lags not covered by these plots. It's also possible that the differencing process has removed or obscured any relationships by equalizing mean levels and

emphasizing transitory fluctuations between successive observations.



Spectral Analysis



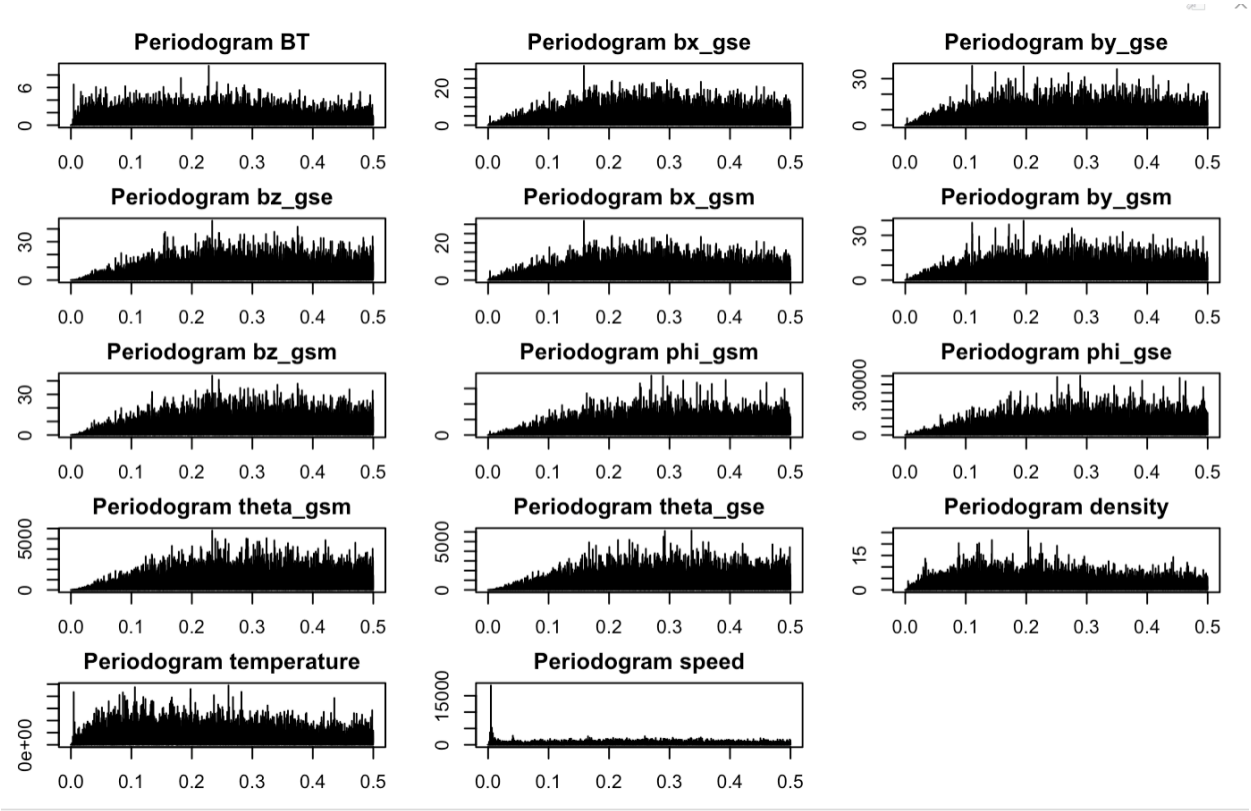


The spike or peak in the periodogram could indicate the presence of a periodic signal or pattern related to geomagnetic disturbances or storm activity. Some

potential reasons for a spike in the periodogram of the differenced DST index could be:

1. Solar activity cycles: Solar activity, such as sunspots and solar flares, follows periodic cycles (e.g., the 11-year solar cycle), which can modulate geomagnetic activity and the DST index.
2. Recurrent geomagnetic disturbances: Certain types of geomagnetic disturbances, like coronal mass ejections or high-speed solar wind streams, may exhibit periodicities due to the rotation of the Sun or the structure of the interplanetary magnetic field.
3. Seasonal or annual patterns: There could be seasonal or annual patterns in geomagnetic activity and the DST index due to factors like the Earth's orbital position relative to the Sun or the tilt of the Earth's magnetic dipole.

The process of differencing has reduced the dominance of the trend component in the DST data, as evidenced by the decrease in power at the low-frequency range in the periodogram. The presence of smaller peaks in the differenced data suggests that there may be other underlying structures or seasonal effects in the DST index that were not as apparent before differencing. The periodograms provide a clear visual representation of the impact of differencing on the time series data, moving it towards stationarity—a requirement for many time series analysis methods, including ARIMA modeling.



- Periodogram BT: The flat spectrum across frequencies suggests that differencing has removed any trend in the total magnetic field strength (BT) data, as no strong periodic components are evident.
- Periodogram bx_gse: The lack of distinct peaks implies that differencing has effectively mitigated any non-stationary behavior in the x-component of the magnetic field in GSE coordinates.
- Periodogram by_gse: A uniform spread of power indicates that trends or seasonality in the y-component of GSE coordinates have likely been addressed through differencing.
- Periodogram bz_gse: The widespread spectral density without dominant frequencies suggests that differencing has neutralized any clear periodicity in the z-component in GSE coordinates.
- Periodogram bx_gsm: Similar to the GSE coordinate counterpart, the bx_gsm data after differencing show no significant periodic components.

- Periodogram by_gsm: The absence of pronounced peaks in the by_gsm dataset suggests that differencing has removed non-stationary elements, such as trends or seasonality.
- Periodogram bz_gsm: A widespread spectral density indicates that differencing has likely achieved stationarity for the z-component in GSM coordinates.
- Periodogram phi_gsm: The even spectral density across frequencies implies that differencing has addressed any non-stationary behavior in the azimuthal angle in GSM coordinates.
- Periodogram phi_gse: The lack of standout peaks in the phi_gse dataset suggests that differencing has mitigated periodic or seasonal patterns.
- Periodogram theta_gsm: The dispersed spectral density for the polar angle in GSM coordinates post-differencing indicates the removal of trends and potential cyclic behavior.
- Periodogram theta_gse: An even distribution of power across frequencies for the polar angle in GSE coordinates suggests that differencing has normalized the data.
- Periodogram density: The absence of strong periodic signals in the plasma density periodogram suggests that differencing has been effective in creating a stationary time series.
- Periodogram speed: The flat spectrum across frequencies for the solar wind speed indicates that differencing has likely removed non-stationary elements from the data.
- Periodogram temperature: The widespread spectral density in the temperature data suggests that differencing has neutralized trends and periodicities.

The periodograms of various parameters, including total magnetic field strength (BT), magnetic field components in both GSE and GSM coordinates, azimuthal

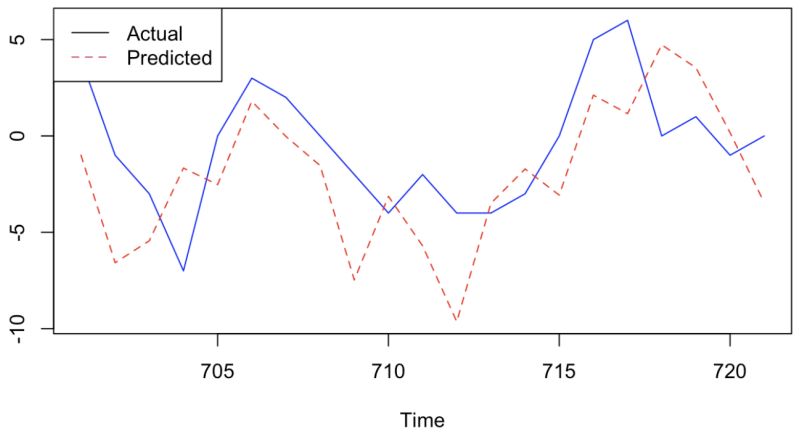
and polar angles, plasma density, solar wind speed, and temperature, after differencing reveal consistent patterns. The absence of distinct peaks or prominent frequencies in these periodograms indicates that differencing effectively removes trends, seasonality, and non-stationary behavior from the data. Instead, the spectral density appears uniformly spread across frequencies, suggesting that the differenced time series have become more stationary. This normalization process enables a clearer examination of the inherent variability in the data, highlighting the effectiveness of differencing in preparing the time series for further analysis and modeling.

Modeling and Results

I created 5 different ARIMA models. The first model consisted of variables up to 4 lags. The second model consisted of variables up to 3 lags. The third model consisted of variables up to 2 lags. The fourth model consisted of variables up to 1 lag.

The last model, the simple one, consisted of just the variables, no lagged variables. I also developed a SARIMA model whose performance was similar to the first ARIMA model.

The best performing model is as follows:



The ARIMA model with up to 4 lagged variables is the best model in terms of prediction because it has the lowest AIC score of 3312.78. The BIC score is 3581.29.

The mathematical equation for the model can be written as:

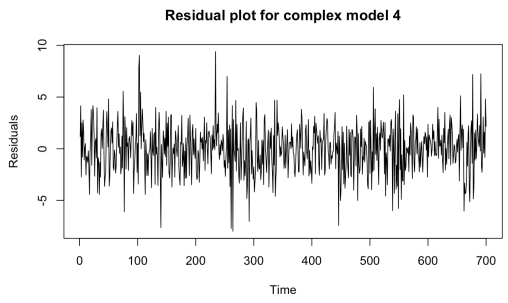
$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$$

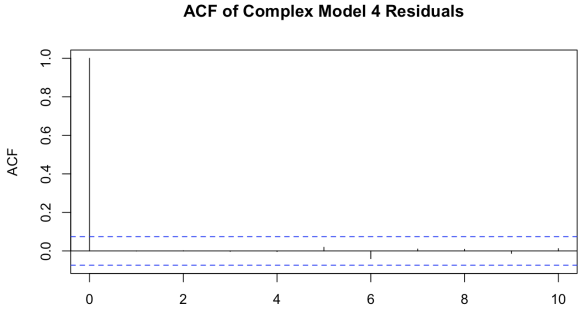
- Y_t represents the differenced time series at time t.
- c is the constant term.
- ϕ_1 and ϕ_2 are the autoregressive coefficients for the first and second lagged values, respectively.
- θ_1 is the moving average coefficient for the first lagged error term.
- ε_t is the error term at time t.

I am choosing AIC over BIC score because AIC is more suitable for short term forecasting (as in the case of DST, is hourly).

```r
AIC(complex_model_lag4)
BIC(complex_model_lag4)
AIC(complex_model_lag3)
BIC(complex_model_lag3)
AIC(complex_model_lag2)
BIC(complex_model_lag2)
AIC(average_model_lag1)
BIC(average_model_lag1)
AIC(simple_model)
BIC(simple_model)
AIC(sarima_model_auto)
BIC(sarima_model_auto)
```

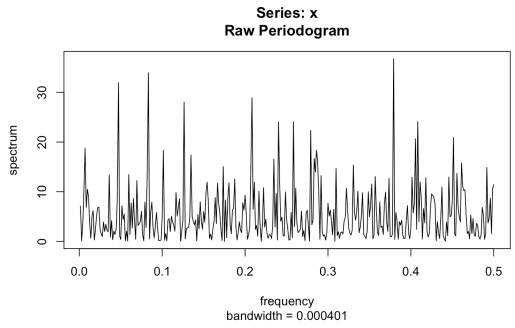
```
[1] 3312.779
[1] 3581.293
[1] 3296.001
[1] 3528.106
[1] 3318.721
[1] 3500.764
[1] 3486.078
[1] 3590.753
[1] 3575.847
[1] 3625.909
[1] 3311.869
[1] 3575.831
```





The residuals of the model fluctuate around zero, which is generally a good sign. However, there are some large spikes, indicating possible outliers or moments of high volatility not captured by the model. It follows a random walk and they are white noise.

The ACF plot shows that autocorrelations are within the confidence bounds for most lags, suggesting that there is no significant autocorrelation in the residuals. This indicates that the model has captured the time series' structure well, at least in terms of linear relationships.



The absence of strong peaks suggests that no single frequency dominates the residuals, which is another indicator of a good model fit.

```
        Box-Pierce test

data:  complex_resids
X-squared = 0.26864, df = 5, p-value = 0.9982
```

The Box-Pierce test results suggest that there is no significant autocorrelation in the residuals at lags up to 5, as the p-value is high (0.9982).This means we fail to reject the null hypothesis of no autocorrelation in the residuals, which is what we want in a well-fitting model.

```
Call:
arima(x = diff_dst[0:700], order = c(2, 0, 1), xreg = data_arima4[0:700, ])

Coefficients:
         ar1      ar2      ma1  intercept  dst_lag1  dst_lag2  dst_lag3  dst_lag4  bx_gse_lag1
      0.7935  -0.2500  -0.9658     0.0078    0.3369    0.0919    0.0967    0.0846     114.9259
s.e.  0.1271   0.1003   0.0231     0.0078    0.1275    0.0601    0.0502    0.0358     157.7715
      bx_gse_lag2  bx_gse_lag3  bx_gse_lag4  by_gse_lag1  by_gse_lag2  by_gse_lag3  by_gse_lag4
         373.0479      29.4624      -7.8766       0.5491       0.5895       0.9011       1.2451
s.e.     204.2324     203.9568     158.4049       0.3559       0.3784       0.3960       0.3725
      bz_gse_lag1  bz_gse_lag2  bz_gse_lag3  bz_gse_lag4  bx_gsm_lag1  bx_gsm_lag2  bx_gsm_lag3
          -0.3882      -0.5968       0.425       0.0646    -114.8177    -372.9630     -29.5416
s.e.       0.5869       0.5816       0.567       0.5471     157.7680     204.2247     203.9460
      bx_gsm_lag4  by_gsm_lag1  by_gsm_lag2  by_gsm_lag3  by_gsm_lag4  bz_gsm_lag1  bz_gsm_lag2
           7.8203      -0.4602       0.2934      -0.6764      -1.1139       0.0853       0.9170
s.e.     158.3995       0.3259       0.3373       0.3297       0.3314       0.6027       0.5905
      bz_gsm_lag3  bz_gsm_lag4  theta_gse_lag1  theta_gse_lag2  theta_gse_lag3  theta_gse_lag4
          -0.3497      -0.2520         -0.0303          0.0984         -0.0715          0.0013
s.e.       0.5853       0.5593          0.0477          0.0480          0.0489          0.0452
      theta_gsm_lag1  theta_gsm_lag2  theta_gsm_lag3  theta_gsm_lag4  density_lag1
              0.0494         -0.0733          0.0823          0.0058        0.5397
s.e.          0.0493          0.0495          0.0490          0.0464        0.0958
      density_lag2  density_lag3  density_lag4  temperature_lag1  temperature_lag2
            0.1158       -0.3695        -0.036                 0                 0
```

- The first AR term (ar1) is significant with a coefficient of 0.7935 and a standard error of 0.1271, indicating a strong positive correlation with the first lag of the differenced series.
- The second AR term (ar2) is also significant but negative with a coefficient of -0.2500 and a standard error of 0.1003, suggesting a partial corrective effect from the second lag.The MA term (ma1) is highly significant with a coefficient of -0.9658 and a small standard error, indicating the model accounts for a substantial amount of noise from the previous term.
- Among the external regressors, certain lags seem to have significant coefficients, such as bx_gse_lag2, by_gse_lag3, by_gsm_lag4, theta_gsm_lag2, and density_lag1.
- The significance of the AR and MA terms alongside some external regressors suggests a relatively good fit to the historical data.

	Estimate	Std. Error	z value	Pr(>|z|)	
ar1	7.9352e-01	1.2706e-01	6.2453	4.230e-10	***
ar2	-2.4996e-01	1.0028e-01	-2.4927	0.0126792	*
ma1	-9.6578e-01	2.3054e-02	-41.8922	< 2.2e-16	***
intercept	7.7574e-03	7.7799e-03	0.9971	0.3187123	
dst_lag1	3.3692e-01	1.2750e-01	2.6426	0.0082277	**
dst_lag2	9.1859e-02	6.0101e-02	1.5284	0.1264116	
dst_lag3	9.6725e-02	5.0188e-02	1.9273	0.0539470	.
dst_lag4	8.4613e-02	3.5803e-02	2.3633	0.0181125	*
bx_gse_lag1	1.1493e+02	1.5777e+02	0.7284	0.4663487	
bx_gse_lag2	3.7305e+02	2.0423e+02	1.8266	0.0677622	.
bx_gse_lag3	2.9462e+01	2.0396e+02	0.1445	0.8851417	
bx_gse_lag4	-7.8766e+00	1.5840e+02	-0.0497	0.9603419	
by_gse_lag1	5.4907e-01	3.5588e-01	1.5429	0.1228651	
by_gse_lag2	5.8949e-01	3.7838e-01	1.5579	0.1192478	
by_gse_lag3	9.0108e-01	3.9597e-01	2.2756	0.0228684	*
by_gse_lag4	1.2451e+00	3.7255e-01	3.3422	0.0008312	***
bz_gse_lag1	-3.8822e-01	5.8692e-01	-0.6615	0.5083228	
bz_gse_lag2	-5.9683e-01	5.8163e-01	-1.0261	0.3048332	
bz_gse_lag3	4.2502e-01	5.6701e-01	0.7496	0.4535000	
bz_gse_lag4	6.4605e-02	5.4706e-01	0.1181	0.9059929	
bx_gsm_lag1	-1.1482e+02	1.5777e+02	-0.7278	0.4667588	
bx_gsm_lag2	-3.7296e+02	2.0422e+02	-1.8262	0.0678143	.
bx_gsm_lag3	-2.9542e+01	2.0395e+02	-0.1449	0.8848291	
bx_gsm_lag4	7.8203e+00	1.5840e+02	0.0494	0.9606241	
by_gsm_lag1	-4.6021e-01	3.2588e-01	-1.4122	0.1578983	
by_gsm_lag2	2.9336e-01	3.3730e-01	0.8697	0.3844462	
by_gsm_lag3	-6.7640e-01	3.2969e-01	-2.0516	0.0402061	*
by_gsm_lag4	-1.1139e+00	3.3143e-01	-3.3609	0.0007769	***
bz_gsm_lag1	8.5267e-02	6.0269e-01	0.1415	0.8874925	
bz_gsm_lag2	9.1701e-01	5.9048e-01	1.5530	0.1204265	
bz_gsm_lag3	-3.4966e-01	5.8529e-01	-0.5974	0.5502299	
bz_gsm_lag4	-2.5196e-01	5.5932e-01	-0.4505	0.6523596	
theta_gse_lag1	-3.0288e-02	4.7657e-02	-0.6355	0.5250817	
theta_gse_lag2	9.8366e-02	4.8020e-02	2.0484	0.0405172	*
theta_gse_lag3	-7.1452e-02	4.8878e-02	-1.4619	0.1437807	
theta_gse_lag4	1.3040e-03	4.5155e-02	0.0289	0.9769612	
theta_gsm_lag1	4.9436e-02	4.9277e-02	1.0032	0.3157507	
theta_gsm_lag2	-7.3319e-02	4.9532e-02	-1.4803	0.1388054	
theta_gsm_lag3	8.2349e-02	4.9019e-02	1.6800	0.0929667	.
theta_gsm_lag4	5.8017e-03	4.6410e-02	0.1250	0.9005164	
density_lag1	5.3974e-01	9.5844e-02	5.6314	1.787e-08	***
density_lag2	1.1577e-01	1.2163e-01	0.9519	0.3411635	
density_lag3	-3.6954e-01	1.0131e-01	-3.6476	0.0002647	***
density_lag4	-3.6031e-02	1.1096e-01	-0.3247	0.7453949	
temperature_lag1	-1.7306e-05	3.9929e-05	-0.4334	0.6647190	
temperature_lag2	8.3544e-06	3.2963e-05	0.2534	0.7999239	
temperature_lag3	-2.6732e-06	3.3428e-05	-0.0800	0.9362627	
temperature_lag4	9.7606e-07	4.3464e-05	0.0225	0.9820837	
speed_lag1	1.6878e-03	3.1487e-03	0.5360	0.5919390	
speed_lag2	-1.9123e-02	3.2317e-03	-5.9174	3.270e-09	***
speed_lag3	-1.1238e-03	3.8556e-03	-0.2915	0.7706824	
speed_lag4	-3.3342e-03	3.2067e-03	-1.0398	0.2984428	
bt_lag1	-2.3587e-02	1.6697e-01	-0.1413	0.8876616	
bt_lag2	2.8830e-01	1.6923e-01	1.7036	0.0884596	.
bt_lag3	-5.9628e-01	1.7316e-01	-3.4435	0.0005743	***
bt_lag4	-3.1388e-01	1.7600e-01	-1.7834	0.0745230	.
sin_pred	1.3732e-02	1.2832e-02	1.0701	0.2845552	
cos_pred	2.4609e-02	1.3684e-02	1.7983	0.0721273	.

Conclusion

The ARIMA model with external regressors, which included lagged values of solar wind variables and DST itself, showed that certain past values are indeed relevant in predicting DST. The significance of these lags indicates some form of dependency, supporting the idea that DST is influenced by past solar wind conditions and its own past values, though this influence may be more nuanced than a straightforward day-to-day relationship.

The model diagnostics, including residual checks and the Box-Pierce test, suggest that the model captured the underlying process adequately, with residuals behaving as white noise, indicating that there is no obvious structure left in the residuals that the model has failed to capture.

Hypothesis 1: DST follows a seasonal trend over the years.

The initial periodogram of the DST data before differencing suggested the presence of a strong low-frequency component, which could indicate a seasonal trend or a long-term pattern. However, after differencing, this prominent peak was removed, suggesting that any seasonality was effectively addressed by the differencing process. While this supports the hypothesis of seasonality, the absence of significant seasonal parameters in the final ARIMA model could mean that the seasonality is not a dominant feature after differencing or may not be periodic within the range of the ARIMA model's lags.

Hypothesis 2: DST is influenced by the lagged values of features associated with solar winds.

The cross-correlation functions between differenced DST and the differenced solar wind variables did not reveal strong correlations at various lags. This suggests that the linear relationship between daily changes in DST and the daily changes in the solar wind variables may not be significant. However, the ARIMA model's coefficients for certain lags of solar wind variables were significant, indicating some predictive power. The relationship may be complex and not entirely linear, or it may manifest over different lags than those initially considered.

Hypothesis 3: DST depends on the previous day's dst.

The ACF plot for the differenced DST data indicated that the autocorrelations for most lags were within the confidence bounds, suggesting no significant autocorrelation after differencing. This would imply that the differenced DST is not dependent on its immediate past values once it has been made stationary. Nonetheless, the ARIMA model coefficients for lagged DST values (dst_lag1, dst_lag2, etc.) were significant, pointing to some dependency, albeit likely of a more complex nature than direct day-to-day influence.

The findings are somewhat mixed with regard to the initial hypotheses. While there is evidence supporting the impact of past solar wind conditions on DST and some indication of seasonality and autocorrelation, the relationships are not straightforward. They may involve more complex dynamics than simple linear dependencies, requiring a nuanced modeling approach. Further refinement of the model could include examining non-linear relationships, considering additional or different lags, and incorporating other potential predictors that might help in understanding the DST index variations better. Other techniques such as Machine Learning or Long Short Term Memory models would be helpful in providing better forecasts.