SEARCHING A SOLUTION TO CORONAL MASS EJECTION (CME) EFFECTS

Nathan Ansel & Kay Krister
Supervisor: Kelby Tirtabudi, S.T., M.T.
Center for Young Scientists. Cahaya Bangsa Classical School, Indonesia. ansel.2022@student.cahayabangsa.org
Silver Medal in Physics Category YSC 2018
Special Award in Physics Category LPB Nasional 2018

I. Introduction

Weather takes place in earth's atmosphere, but did you know, weather occurs in outer space as well? One of these weathers include an eruption namely Coronal Mass Ejections (CME). These strong explosions eventuate with the transport of matter from the sun into interplanetary space. A CME could critically affect life on Earth by creating geomagnetic storms and ionosphere anomalies. A geomagnetic storm once created a 9-hour blackout in Canada, and an ionospheric anomaly once created a GBAS error (Ground Based Augmentation System) as much as 800mm/km. Errors as such may result in a large scaled casualty. This is why a way to predict CME occurrences is needed. If predicting CME occurrences is possible, humans can be prepared to stop it. Hence, this study aims to discover a way to predict a CME occurrence, and thus solve this issue.

II. Experimental Design

In order to solve this issue, a statistical analysis of CME variables that include Plasma Temperature (Kelvin), Proton Density (N/cm^3), Plasma Speed (Km/sec), Sunspot Number, and Disturbance Storm Time Index (nT) is needed. The data can be retrieved directly from NASA's web in omniweb.gsfc.nasa.gov. The data retrieved are processed to predict through calculating descriptive statistics, coefficients of correlation, determination, and covariance, moving average, and linear regression. Finding the values from descriptive statistics would help analyse the characteristics of the variables and how fluctuative they are. Finding the coefficients of correlation, determination, and covariance would tell about the strength of the relationship of the different variables- whether independent or dependent, how strong they are related, and are they directly or inversely proportional. Finding the moving average value would help predict the value for the next period of a particular variable. Last but not least, finding the value from the linear regression method would predict the values of the variables.

The experiment from this research would be about finding the right condition for a certain method to predict the upcoming values from a specific variable. This experimental design focuses on trial and error, testing out a number of 9 types of moving average and 15 types of periodic cycles. The error percentage then will be calculated, and the condition with the lowest error percentage would be the key to predicting the upcoming values most accurately.

III. Result & Analysis

Descriptive statistics calculate the mean, median, mode, minimum, maximum, variance, and standard deviation of the data. The values are shown in the following table.

95230
6
432
74
-13

90627
6
426
64
-12

83663
6
440
2
-12

38401
3.6
328
0
-50

200926
10.8
610
256
8

29458
1.2
48.9
57.1
9.3

867785757.9
1.4
2391.9
3259.8
86.3

Legend

T : Temperature (K - Kelvin) *SN: Sunspot Number*
PD: Proton Density (N/cm$_3$) *PS : Plasma Speed (Km/sec)*
Dst: Disturbance Time Index (nT)

To link the analysis of the prediction methods, the experiment needs to calculate the coefficient of correlation, determination and covariance of the variables. The coefficient of correlation shows how strongly related two variables are and in what way they are related. The coefficient has a range of -1 to 1. The closer the correlation to 1 or -1, the more correlated the

variables are. Positive coefficient means direct proportion of the variables, and negative coefficient means the other way around. The coefficient of determination shows the strength of the correlationship of two variables. The coefficient of determination, which is also known as R-squared, allows the dependent variable variation to be explained because of the independent variable; it is the proportion of the variance in the dependent variable that is predictable from the independent variable. While covariance shows the strength of the product of the deviations of two variates from their respective means.

CORRELATION						DETERMINATION				
	T	PD	PS	SN	Dst	T	PD	PS	SN	Dst
T	1					1				
PD	-0.4	1				0.2	1			
PS	0.9	-0.5	1			0.9	0.2	1		
SN	0.2	-0.05	0.1	1		0.05	0.003	0.009	1	
Dst	-0.5	0.07	-0.5	-0.4	1	0.2	0.005	0.3	0.1	1

COVARIANCE					
	T	PD	PS	SN	Dst
T	864730174				
PD	-15184.9	1.4			
PS	1334717.2	-26.7	2383.5		
SN	364638.8	-3.6	265.2	3248.4	
Dst	-134202.0	0.8	-232.3	-190.0	86.0

After knowing how strong the variables are related and which relationships are most reliable to use to predict the upcoming data, the MA method and linear regression can be used to calculate the upcoming data. Some significant facts found from the analysis are:

(1)The plasma temperature [T] is most related to the plasma speed [PS] , they have a correlation of 0.9.

(2)The pairs that are inversely quite related are (*using abbreviations*) T with PD, T with DST, PD with PS, DST with PS, and SN with DST; each pair has a coefficient ranging from -0.5 to -0.4.

(3)The rest of the pairs are not related at all.

The next method is the Moving Average Method (from MA 2 to MA 10). The MA prediction method works in a way that the next data, or the prediction, will not be far from the data previously; it calculates the average of some numbers to find the next number. Since MA 2 has the lowest percentage of error, it will be used to predict CME occurrences through calculating its variables. However, the MA method has a weakness; it can only predict a short-term data and the predicted numbers must stay update to the real data in order to predict the next data, because the prediction have errors of its own.

In other words, long term prediction is not available using this method. This table shows the most accurate type of MA, their error percentage, and the prediction.

T	PD	PS	SN	DST
MA 2	**MA 4**	**MA 5**	**MA 5**	**MA 5**
16.44%	7.51%	4.43%	31.46%	64.58%
81489	7	440	22	-15

The last prediction method used is linear regression method. Linear regression works with the data that will be predicted using the linear line (*x* and *y* intercept relationship). The weakness of linear regression is that it is a limited linear relationship; it looks at the mean of the dependent variable only, sensitive to outliers, and data must be independent, while some CME data are not independent. The experiment calculates 15 types cycles-meaning that the first cycle, cycle 2, have 2 *x* variables and the last cycle, cycle 16, have 16 *x* variables. The cycle with low error for each variable would be the key to predicting the pattern. Here are the best cycles, their error percentage, and their predicted values for the next 8 terms.

T	PD	PS	SN	DST
Cycle 9	**Cycle 16**	**Cycle 14**	**Cycle 16**	**Cycle 14**
25.7%	15.26%	8.39%	395%	114.5%
98044	6.4	430.45	72.1	-11.3
97554	6.27	419.25	72.7	-10.8
96807	6.3	409.55	74.9	-9.3
98044	6.1	427.9	85.8	-14.4
95023	5.9	441.9	78.2	-19.4
94971	6.1	441.15	73.8	-17.6
89139	6.3	433.45	73.8	-14.2
94920	6.4	442	71.3	-12.3

Overall, the most reliable prediction comes from plasma speed, plasma temperature, and proton density. The sunspot no. is not reliable because it is not correlated with any variables. However, the sunspot no. follows the solar cycle. The linear regression method cannot be applied to confirm this fact because it requires to take over 140 variables, which requires complex calculations. As for the dst index, it is not predictable because it is relative to the condition of Earth's magnetosphere.

IV. Conclusion & Recommendation

Hence, these 3 prediction methods are necessary to predict the future data. However, a CME would happen surprisingly, and the prediction would not be able to predict when a CME will come, but when CMEs are likely to come. Humans must use the prediction method to predict the seasons where CMEs often appear. There are two solutions to solve a disastrous CME. The first solution is to create or use an available magnetic field to stop the plasma from affecting Earth and its people. This could be done by creating an electric current that creates a magnetic dipole. The second solution is to make a plasma stream (because plasma is made of negative and positive electric charges, hence, when the electric charge is negative, it would be like electricity in the cable, and it can create a magnetic field. Plasma can

be created by heating an object to thousands or even tens of thousands of degrees above). However, since applying these solutions might be inefficient due to the fact that these solutions require expensive fundings, the prediction method is necessary to use, for it allows people to be aware of what might be coming.

V. Reference

- http://www.spacedaily.com/reports/Magnetic_Storms_Affect_Humans_As_Well_As_Telecommunications.html
- https://cdaw.gsfc.nasa.gov/
- https://omniweb.gsfc.nasa.gov/
- https://www.statisticshowto.datasciencecentral.com/probability-and-statistics/correlation-coefficient-formula/