

Predicting the CME arrival time based on the recommendation algorithm

Yu-Rong Shi (石育榕)^{1,2,3}, Yan-Hong Chen (陈艳红)^{1,3}, Si-Qing Liu (刘四清)^{1,2,3}, Zhu Liu (刘柱)⁴,
Jing-Jing Wang (王晶晶)^{1,3}, Yan-Mei Cui (崔延美)^{1,3}, Bingxian Luo (罗冰显)^{1,2,3}, Tian-Jiao Yuan (袁天
娇)^{1,3}, Feng Zheng (郑锋)⁴, Zisiyu Wang (王子思禹)^{1,2,3}, Xin-Ran He (何欣燃)^{1,2,3} and Ming Li (李铭)^{1,2,3}

¹ National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China; chenyh@nssc.ac.cn

² University of Chinese Academy of Sciences, Beijing 100049, China

³ Key Laboratory of Science and Technology on Environmental Space Situation Awareness, Chinese Academy of Sciences, Beijing 100190, China

⁴ Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen 518055, China

Received 2021 January 21; accepted 2021 March 17

Abstract CME is one of the important events in the sun-earth system as it can induce geomagnetic disturbance and an associated space environment effect. It is of special significance to predict whether CME will reach the Earth and when it will arrive. In this paper, we firstly built a new multiple association list for 215 different events with 18 characteristics including CME features, eruption region coordinates and solar wind parameters. Based on the CME list, we designed a novel model based on the principle of the recommendation algorithm to predict the arrival time of CMEs. According to the two commonly used calculation methods in the recommendation system, cosine distance and Euclidean distance, a controlled trial was carried out respectively. Every feature has been found to have its own appropriate weight. The error analysis indicates the result using the Euclidean distance similarity is much better than that using cosine distance similarity. The mean absolute error and root mean square error of test data in the Euclidean distance are 11.78 and 13.77 h, close to the average level of other CME models issued in the CME scoreboard, which verifies the effectiveness of the recommendation algorithm. This work gives a new endeavor using the recommendation algorithm, and is expected to induce other applications in space weather prediction.

Key words: Sun: coronal mass ejections (CMEs) — method: recommendation algorithm

1 INTRODUCTION

CMEs (coronal mass ejections) are massive plasma eruptions carrying a magnetic field occurring from the solar atmosphere into the heliosphere (Webb & Howard 2012). Observations of CMEs began in the 1970s, the earliest research satellite that was developed by the European Space Agency (ESA) and National Aeronautics and Space Administration (NASA) was the Solar and Heliospheric Observatory (SOHO) to study the structure of the Sun, chemical composition, dynamics of the solar interior, structure of the outer solar atmosphere (density, temperature, velocity fields, etc.) and the relationship between solar wind and solar atmosphere. The coronagraphs (Large Angle and Spectrometric Coronagraph Experiment, LASCO) (Brueckner et al. 1995) aboard SOHO have made tremendous contributions to CMEs observations.

The interactions between CMEs and interplanetary disturbance can cause violent outbursts in the space environment. In general, the CMEs toward Earth may cause near-Earth space environment effects within 1–3 days after its eruption. Therefore, whether CMEs reach the Earth and when they arrive are specially significant in space weather prediction. Currently, there are many different models to estimate the CMEs arrival time (Zhao & Dryer 2014; Verbeke et al. 2018). These models fall into three categories: empirical models are fitted by looking for the relationship between CME velocities and their arrival time (Vandas et al. 1996; Wang et al. 2002; Manoharan 2006; Schwenn et al. 2005; Xie et al. 2004; Núñez et al. 2016; Paouris & Mavromichalaki 2017), drag-based models fully consider the interaction between the CME and the background solar wind (Subramanian et al. 2012; Vrřnak et al. 2013; Hess & Zhang 2015; Möstl et al. 2015),

and MHD models use the observed data as boundary conditions to predict the CME arrival time (Smith & Dryer 1990; Dryer et al. 2001; Moon et al. 2002; Tóth et al. 2005; Detman et al. 2006; Feng & Zhao 2006; Feng et al. 2007; Riley et al. 2012, 2013; Odstrcil et al. 2004; Poedts et al. 2020; Jin et al. 2017; Sokolov et al. 2013; van der Holst et al. 2014; Wang et al. 2018). With the development of computers, a lot of research is devoted to the prediction of CME arrival time using the machine learning method. Sudar et al. (2016) based on FCNN to fit the relationship between CME arrival time, initial velocity and central meridian. Liou et al. (2018) used SVM (support vector machines) to predict the arrival time based on CME features. Wang et al. (2019) employed the CNN (convolutional neural network) region model to obtain the CME arrival time by white-light observation.

Currently, the existing means of observation are the field observation to measure the plasma and magnetic field data by satellites, and using the coronagraph carried by satellites to image CMEs. However, since the imaging of CMEs is a projection effect on a two-dimensional plane, the shape and structure of images observed cannot reflect the real propagation shape of CMEs in three-dimensional space. The observation results of CMEs are limited by the projection effect and observation angle, which greatly weakens our research on the physical mechanism of CME and the development of prediction work. Therefore, in routine work, forecaster's experience is still useful in predicting the geo-effectiveness of CMEs, as a complementary of the CME modeling. In order to provide the forecaster a reference in the real forecasting service, a recommendation system based on machine learning was designed to adequately excavate the historical CME events effects and find similar CME events to guide current prediction work. Then machine learning has achieved more excellent results than other methods in computer vision (Zheng et al. 2018), speech recognition (Schultz et al. 2021), control system (Liu et al. 2021), physical chemistry (Kang et al. 2021), biology (Huang et al. 2020) and other natural science fields (Ham et al. 2019; Wang et al. 2021) for data modeling. Related methods are increasingly used to build models from the increasing volume of space data to find the natural laws and meet the needs of our production and life as well as scientific research (Wang et al. 2017). CAT-PUMA (Liu et al. 2018) model used a support vector machine (SVM) algorithm, taking the Interplanetary Coronal Mass Ejection (ICME) physical parameters as the input characteristic parameters to predict CMEs arrival time, which use the SOHO satellite observations of 182 CME events before through meticulously analyzing characteristics of CMEs and the solar wind parameters. The recommendation system is

a system for reasonably filtering the existing massive information, predicting the user's rating or preference for items, and giving reasonable recommendations for them. Currently, the application of recommendation algorithm in earth science is still in its infancy, especially in the application of space weather. In this paper, a new model was adopted, which is based on historical data, recommending similar historical events for current CME events and forecasting CME recommended arrival time simultaneously. The CME data setting and normalization will be highlighted in Section 2. Section 3 introduces how to calculate the distances between historical events and current CME events. The recommendation result and discussion were carried out in Section 4. Section 5 was the summary and conclusion.

2 DATA

Since the propagation time of CMEs is related to various factors, the characteristic parameters of CMEs, which describe the CME's direction, angle width, speed etc. should be taken into account. In addition, the source region coordinates of CMEs on the Sun also related to the direction of CMEs propagation. According to previous experience, CMEs escape from the Sun and interact with interplanetary media. Therefore, the interplanetary physical quantities of background solar wind have a great influence on the propagation of CMEs. All the physical parameters mentioned above are described below. Finally, a multiple association list of CME, which includes 215 CME events and 18 physical parameters, is determined and used for the further analysis. The following gives a detailed introduction of the 18 characteristic parameters.

2.1 Seven Characteristic Parameters Determined from CME List and ICME List

From 1996 to 2020, the CME catalog¹ includes 30 321 events which is obtained by LASCO observations and maintained at the Coordinated Data Analysis Workshops (CDAW) data center (Yashiro et al. 2004). Seven features were used in this paper extracted from the list: CPA (central position angle), MPA (the position angle of the fastest moving segment of the CME leading edge), angle width (the sky-plane width of CMEs), and four speeds: the linear speed, quadratic speed at the time of initial and final (last possible) height measurement respectively, and second-order polynomial fitting speed evaluated when the CMEs are at a height of 20 Rs (solar radius).

For the same time period, all observed geo-effective CMEs was established by the following near-Earth

¹ https://cdaw.gsfc.nasa.gov/CME_list/UNIVERSAL/text_ver/

ICME database: the Richardson and Cane list² (Cane & Richardson 2003; Richardson & Cane 2010; Ameri & Valtonen 2017). The time of CMEs disturbance to the ground is defined as the arrival time of a shock at Earth. Then, propagation time is obtained by subtracting the outbreak time with CMEs. There are 215 geo-effective CME events associated with CMEs in history. So far, we have obtained a geo-effective CME list including seven characteristics of CMEs and the propagation time of CMEs. We obtained a more comprehensive list of CMEs associated with flares by integrating the information from these lists. At the same time, it also laid a foundation for the next step of searching for the active region of CMEs.

2.2 Parameters Describing the Source Region

From 1996 to present, the X-ray flow mean value is used to automatically extract the characteristic parameters of solar flare events by using the program in 1 min at two bands of the GOES series satellite. The characteristic parameters of solar flare events list³ is obtained from United States environmental data center. In this solar flare list, the parameters of the active region and associated CMEs are determined manually. The active region numbers of CMEs can be found through this list, so they can correspond to the geo-effective CMEs. In addition to this list, we also referred to the flare working directory of Watanabe et al. (2012)⁴. We obtained a more comprehensive list of CMEs associated with flares by integrating the information from these lists. At the same time, it also laid a foundation for the next step of searching for the active region of CMEs.

The information in the active region list includes the numbers of active regions and the corresponding latitudes and longitudes on the solar disk. Combined with the work mentioned in Section 2.1, the active region information of geo-effective CMEs can be obtained. But these are far from enough, as there are some active regions where locations cannot be determined. For those CMEs with no identified source region information, we refer to a few works (Li & Luhmann 2006; Sinha et al. 2019; Maričić et al. 2020) to determine coordinates. But there are still many undetermined events with no associated flares or active region, and we manually measure the middle position coordinates of the filaments by referencing JHelioviewer (Müller et al. 2017), which is being developed as open source software by the ESA determine the source region information. Finally, the positions of each source region about CMEs were expressed in terms of the

corresponding heliographic coordinate system: longitude and latitude. By integrating the data of various parties and adding manual confirmation, we finally get the data as shown in Appendix Table A.1. The first column is sequence number, the second column is near-Earth ICMEs, contain time information on date, month, year, hours and minute, and the third column is associated with the CME events that also implies a time series. The first three columns are all basic information, and the fourth column is the corresponding active region number. Coordinate information on the surface of the fifth column marks the most important information in the direction of CMEs communication. The sixth and seventh columns are corresponding to the flares or filaments respectively, or both are uncertain information which means some active regions have multiple bursts of solar activity in succession.

2.3 F10.7

Another parameter is the F10.7 index corresponding to the day of the CMEs. After the completion of this work, a list including geo-effective CMEs associated with the F10.7 index of the same day, seven characteristic parameters of CME and the longitude and latitude of the active region can be obtained.

2.4 Eight Solar Wind Parameters

Since the time of CMEs reaching the Earth is directly related to the background solar wind, the input parameters also increase the three magnetic field components of the solar wind (B_x , B_y , B_z), the solar wind velocity, proton density, temperature, flow pressure and plasma beta. The parameters of the solar wind parameters are downloaded in intervals of five minutes from the OMNIWeb⁵. And then we calculated the daily average for each parameter. In addition to the previous parameters, plus the eight solar wind parameters, the final CME list contains a total of 18 parameters.

2.5 Data Preprocessing

After removing partial parameter missing events, final 215 geo-effective ICME events were handed out associated to CMEs, and with 18 characteristics added. Use the respective features on both lists (the CME list and the near-Earth ICME list) to integrate into a new list: multiple association list.

For each near-Earth ICME event, a disturbance time is the time of the associated geomagnetic storm sudden commencement (that is to say: CMEs disturb to the ground). Moreover, we can get the CME start time from

² <http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm>

³ https://hesperia.gsfc.nasa.gov/hessidata/dbase/hessi_flare_list.txt

⁴ https://xrt.cfa.harvard.edu/flare_catalog/

⁵ <https://omniweb.gsfc.nasa.gov>

the CME list, and then get the CME propagation time at Earth.

Normalization is a way of simplifying the calculation, in which the dimensional expression is transformed into a dimensionless expression: a scalar. This method is often used in a variety of computations. Because the magnitude of the 18 characteristic parameters is very wide, we normalize these characteristic parameters and map the data to the range of $-1 \sim 1$, which is more convenient and fast for follow-up data processing. In order to better adapt to different physical characteristics, there are three folds:

The first normalization formula we used is Max-Abs Normalization, because it does not move and aggregate data and therefore does not break any sparsity.

$$X = \frac{(X - X_{\min(\text{axis}=0)})}{(X_{\max(\text{axis}=0)} - X_{\min(\text{axis}=0)})}. \quad (1)$$

The ‘axis=0’ means the normalization operation is performed for each column, that is, the normalization of different characteristic parameters does not affect each other. For the angular width, CME linear speed, CME second order initial speed, CME second order final speed, CME second speed in 20 Rs, F10.7, solar wind speed, proton density, temperature, flow pressure and plasma beta, the above characteristics in total of 11 parameters are normalized in this way.

According to practical experience, CMEs outbreak in the east-west direction of the Sun have completely different effects on the Earth, so the values of CPA and MPA fall within different positive and negative intervals through the sine function, which can reflect the direction of CMEs outbreak. So, the second way of normalization is for CPA and MPA to calculate their sine function. The measuring CPA and MPA rotate counterclockwise from the north of the diurnal surface, so the angle of the eastern hemisphere of the diurnal surface ranges from 0° to 180° , and its sine value falls between 0 and 1, whereas the sine value of the western hemisphere of the diurnal surface is between -1 and 0.

In addition to the above 13 features, there are still five parameters: B_x , B_y , B_z , longitude and latitude. Different from the physical meaning of the characteristics mentioned above, these physical quantities themselves contain positive and negative signs to represent their directional characteristics. Therefore, in order to retain the positive and negative properties of the physical quantities themselves, we adopt a normalized way of dividing them by the maximum value of each physical quantity directly.

In view of all the above normalization steps, we have established a complete unified dimensionless data set, but the manners adopted are not random. We have fully considered the different characteristics of each physical

quantity and tailored different normalization methods for it, so as to facilitate the next model construction.

3 METHOD

3.1 Similarity Determination

The recommendation system (RS) is a technology that provides accurate recommendation to users. The CMEs arrival time forecast based on the recommendation system is measured by calculating the similarity between the concerned event and the rest of the CME events. Through similarity calculation, the most similar historical events corresponding arrival time except itself are found as the final recommended results. According to the 18 characteristics corresponding to each CME event, a 18-dimensional space can be constructed. For each near-Earth ICME event, there is a vector (the cosine distance) or a certain point (the Euclidean distance) in the space that represents every event. By calculating the distance of each point, the similarity between each part of CMEs can be obtained. The smaller the distances between the two events are, the more similar the events. Here we calculate two kinds of commonly used distances: cosine distance and Euclidean distance.

3.1.1 Cosine distance

Cosine similarity measures the difference between two individuals by the cosine of the angle between two vectors in the space. It is more concerned with the difference in direction between two vectors than with distance or length. The formula for n dimensional space is as follows:

$$\cos \theta = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n (A_i \cdot B_i)}{\sqrt{\sum_{i=1}^n A_i^2} \cdot \sqrt{\sum_{i=1}^n B_i^2}}. \quad (2)$$

In the formula, $A(x_1, x_2 \dots x_n)$ and $B(x_1, x_2 \dots x_n)$ represent different CME events corresponding the 18-D space vector. The subscript i represents the i th dimension.

3.1.2 Euclidean distance

The Euclidean distance (also known as Euclidean metric) is a commonly used definition of distance that refers to the true distance between two points in n dimensional space, or the natural length of a vector (that is, the distance from the point to the origin). Again, here is the formula:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}. \quad (3)$$

The interpretation of subscripts like above, x and y represent the spatial coordinates of different near-Earth ICME events.

3.2 Deviation

To evaluate the accuracy of the prediction model, the deviation between the predicted CME arrival time and the actual observation time is defined as follows:

$$\Delta t_i = t_i^f - t_i^o. \quad (4)$$

In this case, t_i^o stands for the i th CMEs observed arrival time and t_i^f is the i th CMEs forecasted arrival time. With this definition, a positive Δt corresponds to when a CME is predicted to arrive latter than it is observed, while a negative Δt corresponds to an early arrival prediction compared to the observations. For different events, the errors of prediction are positive or negative. Generally, we will comprehensively consider the mean error of all events or more reference errors, so there are several error statistics methods (Jolliffe & Stephenson 2012; Verbeke et al. 2018; Riley et al. 2018) to measure the accuracy.

3.2.1 Mean Error (ME)

The metric is the Mean Error, which also defined the accuracy or bias:

$$\text{ME} = \frac{1}{N} \sum_{i=1}^N \Delta t_i. \quad (5)$$

In the formula, N represents the number of events in the experiment, and i represents the specific events. Mean Error is a method of quantifying deviation, which evaluates whether the predicted time is earlier or later than the observed time from the average perspective.

3.2.2 Mean Absolute Error (MAE)

The Mean Absolute Error is more comprehensive to consider the model without the positive and negative error, which is given by :

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\Delta t_i|. \quad (6)$$

In most forecasting models, MAE is the most commonly used metric to measure the forecast ability of a model. Although ME is a metric for model bias, it is not sufficient to measure a model's forecast skill. Because the positive and negative can complement each other, MAE is a more objective to measure the distance between the predicted value and the actual value. However, MAE contains its own limitations and may cause certain evaluation bias. Therefore, in order to make the evaluation more objective and comprehensive, we will also consider the two error measures to be discussed below.

3.2.3 Standard Deviation (SD)

The standard deviation also known as precision, represents the square root of the variance, and the formula is as follows:

$$\text{SD} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (\Delta t_i - \text{ME})^2}. \quad (7)$$

Here N refers to the number of all events and ME is the Mean Error mentioned above. The advantage of standard deviation over MAE is that it measures the average distance between the predicted value and the observed value to calculate the distribution of the observed value around the mean error.

3.2.4 Root-Mean-Square Error (RMSE)

Additionally, there is a root-mean-square error which is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\Delta t_i)^2}. \quad (8)$$

In general, it is difficult to minimize the RMSE. Because in the formula, the error is quadratic, which gives greater weight for the larger errors to result in increasing the RMSE of the whole model.

3.3 Recommendation Method

For the 215 CME events, 10% of the events were approximately selected as test samples, which is about 20 events, while the remaining 195 events were divided into training sets (156 events) and validation sets (39 events). The recommendation process was divided into three steps. First, in the training process, the distance between two events was determined directly. The weight values of each feature were separately calculated from 0 to 50, and the step size was 1. When the weight of one feature was adjusted, the weight of the rest features was set to 1, which was to ensure the unity and comparison of the experiment. The above operation was carried out based on cosine distance and Euclidean distance, respectively. Eventually, $51 \times 18 = 918$ trials were needed to iterate the 18 feature weights from 0 to 50. By comparing the forecast results in mean error (ME), mean absolute error (MAE), standard deviation (SD) and root-mean-square error (RMSE), the better weights corresponding to the four error parameters could be obtained. Second, the weights corresponding to minimum mean absolute error, standard deviation and root-mean-square error were selected as the weights of input for the verification set, respectively. The final best weight was selected for the model through determining the

minimum of mean absolute error in the verification set. Finally, the test set was used to verify the advantages and disadvantages of the model using the optimal weights.

4 RESULT AND DISCUSSION

4.1 Experimental Result

Through these experiments, the prediction model for CME arrival time including the optimal dependence on the model prediction. In the experiments, the influences of each feature on the model prediction results were measured by analyzing four error measures of mean error, mean absolute error, standard deviation, and root-mean-square error. Figure 1 shows the weights trend of each feature under the cosine distance, in which Figure 1(a) represents each feature under the mean error. As the problem of positive and negative is involved, the mean error cannot clearly express the characteristics of each physical quantity. Figure 1(b), (c) and (d) refer to MAE, SD and RMSE respectively. In these images, the expressions of these features are very similar, especially the standard deviation and RMSE with smaller differences. In three errors, the three velocities of CMEs are particularly obvious manifestation, which are linear velocity, initial velocity and final velocity of CMEs successively. This confirms a common knowledge of physics: an object travel time depends on its speed. Indeed, although the parameters used are not the most accurate, they do reflect that the propagation time of CMEs is related to these velocities.

Same calculation steps for Euclidean distance are processed to find out whether it is more suitable for calculating the recommended similar events. After all, in many other recommended models, Euclidean distance is a very common distance calculation method, which performs well. Such is the case, experiments show that Euclidean distance does perform better than cosine distance in predicting CMEs arrival time in training. The experimental results of Euclidean distance are shown in Figure 2. The mean error under Euclidean distance is shown in Figure 2(a), indicating that the trend of each feature is declining with the increase of weight. However, it is not the only way to evaluate, and we can use other errors to comprehensively evaluate the model results. In Figure 2(b), (c) and (d), it can be found that the trends of some physical features are same as that of cosine distance, the initial velocity and linear velocity of CMEs both illustrating that the errors decrease with the increase of weight.

Tables 1 and 2 respectively reveal the minimum error corresponding to each feature at the cosine distance and Euclidean distance, and the numbers in brackets correspond to the optimal weights. We can find that each

feature has four optimal weight which corresponds to the four different error. Since the mean absolute error and standard deviation are more important, priority is given to the combination of weights corresponding to these two errors. Substituting the combination of these weights into the validation set yields the results in Table 3.

Through the analysis of these two tables, it is found that some optimal weights have common characteristics in the cosine distance and the Euclidean distance. The weights of these four CME velocities, linear velocity, initial velocity, final velocity and V_{20R_s} (the velocity of CMEs at 20 solar radii), are a large proportion, indicating the same importance for the four velocities. In the whole experiment, it indicates that speed is a key factor affecting the time of CMEs reaching the Earth. At the same time, the latitude and longitude of the source region are also the key factors determining the arrival time. Their weights are also within this range, suggesting that position of resource region is one of the factors that influence the ground effect of CMEs. Flow pressure and Plasma beta also have the certain proportion of their weight. It also shows the importance that they have impact on the spread of CME. There are many physical parameters whose optimal weights correspond to values of 0. For example, the B_z components of the magnetic field have weights of 0, indicating it has indeed nothing to do with the time of the CMEs. However, the weights of B_x , B_y and F10.7 is not 0 but they are not big, which means they have a weak impact on the arrival time of CMEs. Furthermore, CPA and MPA represent the direction of the CMEs, but they are not one of the key factors, indicating smaller weights. Moreover, the weight of MPA in the two distances is also 0.

However, in cosines and Euclidean distances, not all the weights behave the same way. For instance, since most CMEs that have hit earth are full halo CMEs with an angular width of 360 degrees, it is also representative that the weight of angular width in Euclidean distance is particularly great. But the angular width does not weigh as much in the cosine distance as it does in Euclidean distance, but it does have some proportion. The same physical quantity that has this property is temperature. However, performance of proton density is the opposite, which is smaller in cosine distance than Euclidean distance. Although the weights of these physical quantities are not exactly the same in the cosine distance and Euclidean distance, they still have a certain reference value.

By applying the optimal weight summarized from the training set, it is notable that the results of the validation data are improved, which are shown in Table 3. After training process, there are four groups of weights correspond to four kinds of errors. ME is usually not

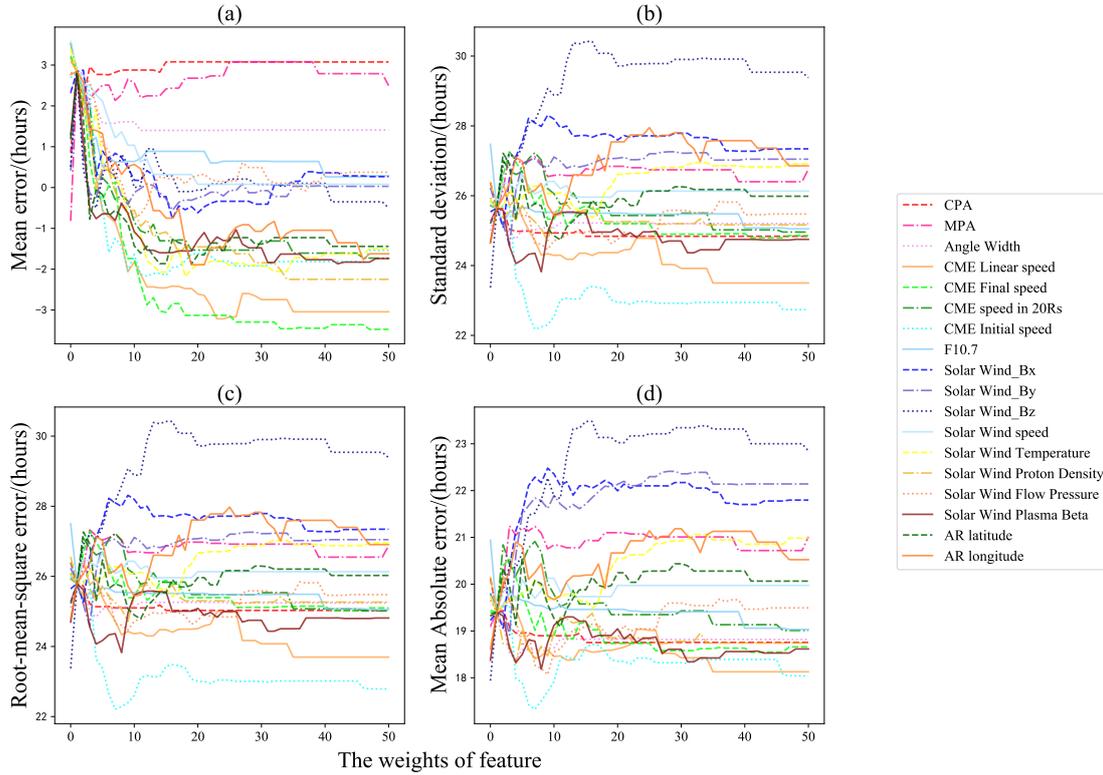


Fig. 1 The relationships between the four errors and the weights of 18 features by calculating the cosine distance, the performance results in mean error (a), mean absolute error (b), standard deviation (c), and root-mean-square error (d) were obtained.

Table 1 The summary of errors and optimal weights in train samples for mean error, mean absolute error, standard deviation, root-mean-square error in cosine distance.

Feature	CPA	MPA	AW	Linear Speed	Initial Speed	Final Speed	V_{20Rs}	F10.7	Bx	By	Bz	Solar wind speed	Proton density	Temperature	Flow pressure	Plasma beta	Latitude	Longitude
ME	0.85	-0.82	1.4	-3.22	-2.17	-3.47	-1.73	0.26	-0.76	-0.56	-0.5	0.08	-2.25	-2.18	-0.74	-1.87	-1.87	-1.89
Weight	0	0	11	23	12	47	46	40	16	16	50	20	34	18	10	42	19	19
SD	24.84	25.27	24.76	13.5	22.17	24.79	24.96	25.05	25.1	24.58	23.37	25.64	24.71	25.37	24.46	23.82	24.73	24.65
Weight	15	0	4	35	7	41	46	47	2	3	0	1	9	16	9	8	11	0
RMSE	25.03	25.28	24.84	23.7	22.21	25.02	25.02	25.05	25.26	24.58	23.38	25.8	24.71	25.42	24.47	23.83	24.79	24.8
Weight	15	0	4	35	7	41	46	47	2	3	0	1	9	16	9	8	11	0
MAE	18.75	19.1	18.39	18.13	17.33	18.55	19.01	19.03	19.06	18.91	17.94	19.4	18.31	19.4	18.07	18.19	19.09	18.61
Weight	15	0	4	35	7	41	46	47	2	3	0	1	4	1	9	8	11	0

The first row of the table represents the 18 features, with the order as follows: CPA, MPA, angular width (AW), CME linear speed, CME initial speed, CME final speed, CME speed in 20RS (V_{20Rs}), F10.7, the x -axis component of the magnetic field (B_x), the y -axis component of the magnetic field (B_y), the z -axis component of the magnetic field (B_z), average daily speed of solar wind, solar wind daily average proton density, solar wind average daily temperature, average daily flow pressure of solar wind, plasma beta, latitude and longitude coordinates of the active region, unit of errors: h. The data in the table correspond to the time errors of each feature, and the next line indicates the corresponding optimal weight.

considered, so the remaining three weights correspond to three errors. Under cosine distance, the weights corresponding to standard deviation and root-mean-square error in training data are same, which can be regarded as an experiment. So there are two experiments in cosine distance and three experiments in Euclidean distance. For cosine distance, the best results are in first line, the mean error, standard deviation, root-mean-square error and mean absolute error of validation sets are -6.48 , 18.54 , 19.36

and 15.24 h respectively, whose weights are corresponding to MAE in training data. For Euclidean distance, the best results are the second experiments in Euclidean. The ME, SD, RMSE and MAE are -1.46 , 19.15 , 19.21 and 15.13 h respectively.

In the testing set, we showed a significant advantage reducing the mean absolute error to about 11 h, and the forecasting ability is excellent and speed is very fast. A test set of 20 CME events was also used to compare the

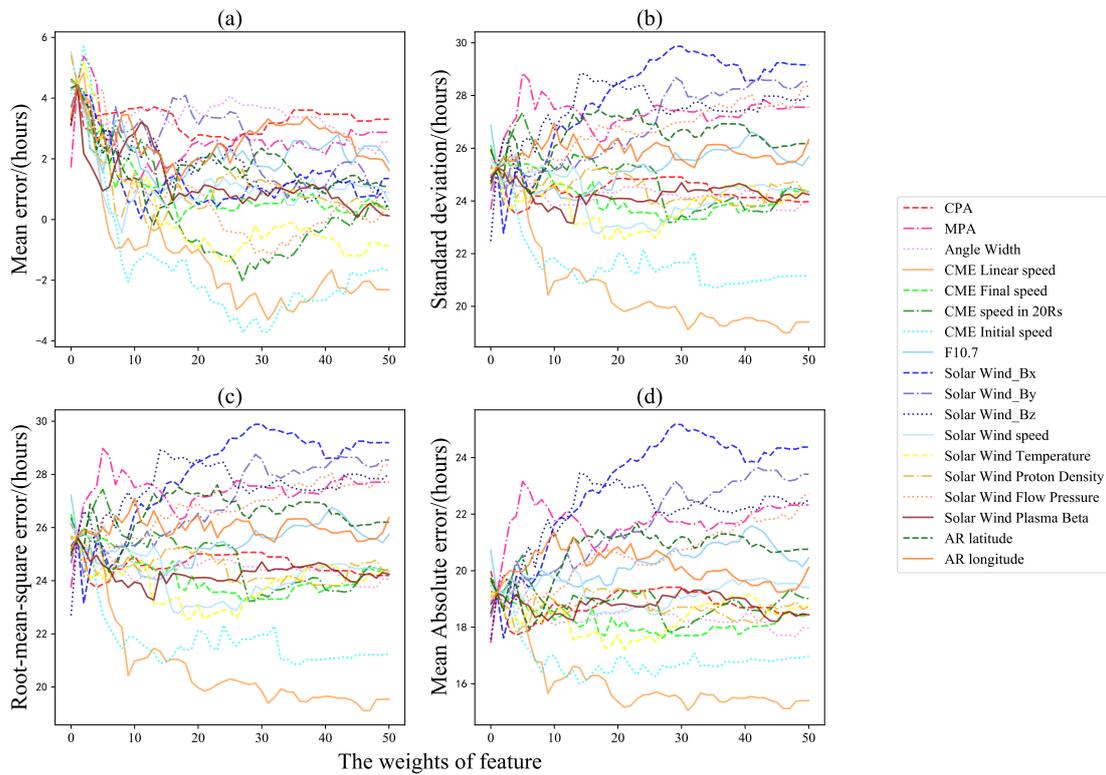


Fig. 2 The relationships between the four errors and the weights of 18 features by calculating the Euclidean distance, the performance results in mean error (a), mean absolute error (b), standard deviation (c), and root-mean-square error (d) were obtained.

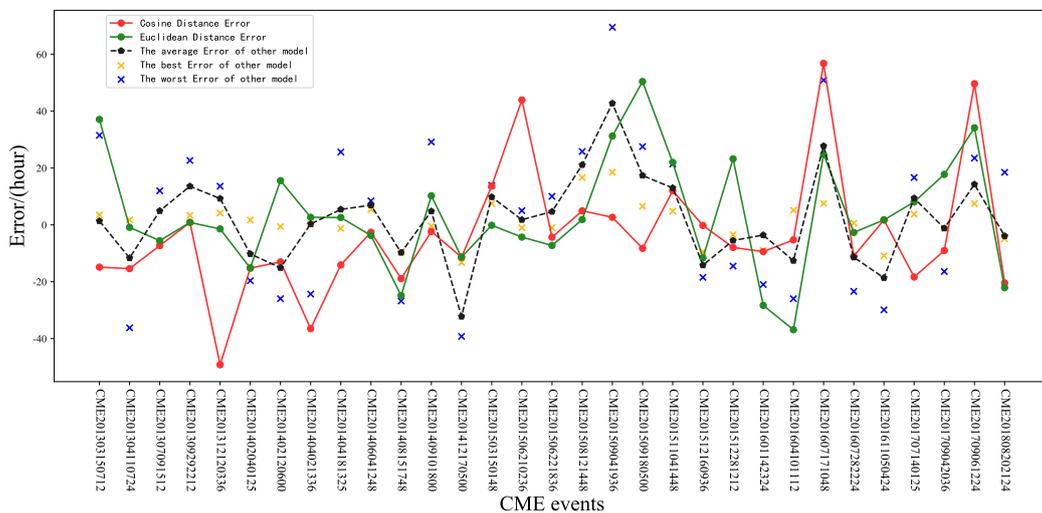


Fig. 3 Exhibit the errors calculated severally by recommended model of the cosine distance and Euclidean distance, and the errors of other models in average value, the best results, and the worst results.

predicted arrival times as shown in Table 4. The first row is the result of cosine distance, and the second row is the result of Euclidean distance. Compared with the results of the training set, the performance of the test set shows the SD, RMSE and MAE are smaller than those for training set. The total number of CME events that have hit earth

in history is only more than 200. Although, choosing 10% of them as the test set is bound to have some contingency and instability, the result still can verify the ability of the method to some extent.

Currently, there are limited methods for predicting CME arrival time. The recommendation method was firstly

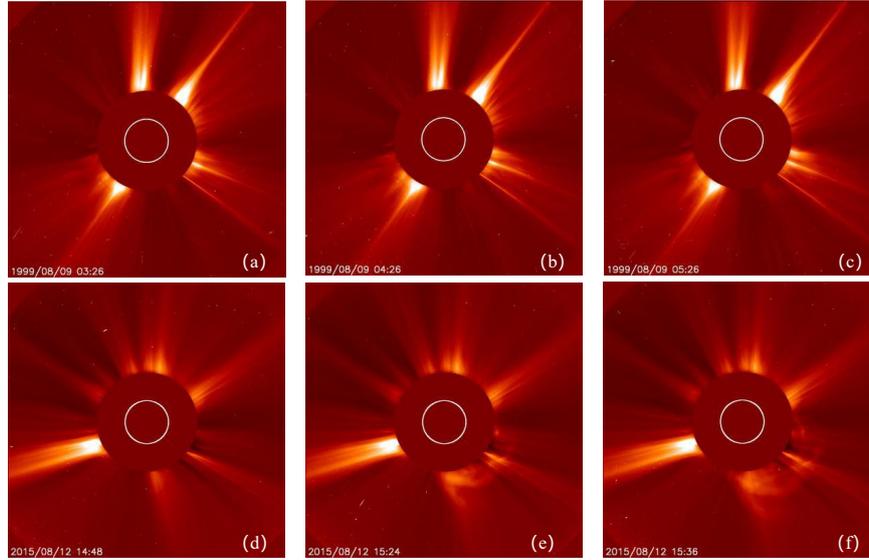


Fig. 4 Top: From left to right, snapshots of the CME event occurring at 2015 August 12 14:48 UT provided by SOHO LASCO C2. Bottom: From left to right, snapshots of the CME event occurring at 1999 August 9 3:26 UT.

Table 2 The summary of errors and optimal weights in train samples for mean error, mean absolute error, standard deviation, root-mean-square error in Euclidean distance.

Feature	CPA	MPA	AW	Linear Speed	Initial Speed	Final Speed	V_{20R_s}	F10.7	Bx	By	Bz	Solar wind speed	Proton density	Temperature	Flow pressure	Plasma beta	Latitude	Longitude
ME	2.59	-1.21	2.17	-3.3	-3.73	-0.07	-2.04	0.94	-0.03	-0.7	-0.33	0.44	-0.37	-1.41	-1.11	-0.07	-0.17	1.29
Weight	-19	8	48	31	27	17	27	14	11	34	50	8	13	24	31	45	13	17
SD	23.5	23.68	23.68	19	20.72	23.2	23.18	24.03	22.78	23.84	22.47	22.76	23.79	22.53	24.91	23.15	24.25	24.97
Weight	4	0	48	46	35	27	28	13	2	2	0	17	27	21	2	13	6	0
RMSE	23.75	23.74	23.78	19.11	20.86	23.21	23.23	24.05	22.15	24.15	22.69	22.77	23.81	22.54	25.09	23.27	24.32	25.19
Weight	4	0	48	46	35	27	28	14	2	2	0	17	42	21	8	13	6	0
MAE	17.72	17.44	17.66	15.05	16	17.61	17.81	19.16	17.62	18.35	17.57	18.42	17.78	17.21	18.82	18.2	17.96	18.92
Weight	4	0	48	31	14	27	28	2	2	0	0	16	4	21	2	45	6	0

Table 3 The summary of the statistics errors for validation samples in cosine distance and Euclidean distance (The scale in the second column is the optimal combination of weights for all the features, error unit: h).

Model	Test Number	Weights	ME	SD	RMSE	MAE
Cosine distance	1	15:0:4:35:7:41:46:47:2:3:0:1:4:1:9:8:11:0	-6.48	18.54	19.63	15.24
	2	15:0:4:35:7:41:46:47:2:3:0:1:9:16:9:8:11:0	-5.7	21.33	22.08	17.46
Euclidean distance	1	4:0:48:31:14:27:28:2:2:0:0:16:4:21:2:45:6:0	0.42	21.2	21.2	17.15
	2	4:0:48:46:35:27:28:13:2:2:0:17:27:21:2:13:6:0	1.46	19.15	19.21	15.13
	3	4:0:48:46:35:27:28:14:2:2:0:17:42:21:8:13:6:0	2.4	19.67	19.82	15.92

used and compared the prediction results with that of the models issued on the CME Scoreboard⁶ (developed at the Community Coordinated Modeling Center, CCMC). These results can be seen in Table 5. In this table, the first column shows the CMEs temporal information, the second column showing the observed CMEs arrival time to earth, the third column showing the predicted time and time error in cosine distance, the fourth column also showing the predicted time and error in Euclidean distance, and the last

three columns of the table show the results of the other models. It can be manifested that the two predicted results in cosine and Euclidean distance fall between the worst and the optimal errors of other model, and evenly distributed around the average values. Among the last three columns, the average values of other models indicate a very small error. After all, the difference of error between each model is relatively large, so the average value will be reduced. The remaining two columns correspond to the best and worst results of the other models, respectively. An interesting finding is that there are events in the cosine distance and

⁶ <https://kawaii.ccmc.gsfc.nasa.gov/CMEscoreboard/>

Table 4 The Results of Test Set for Cosine Distance and Euclidean Distance (Error Unit: h)

Model	Weights	ME	SD	RMSE	MAE
Cosine distance	15:0:4:35:7:41:46:47:2:3:0:1:4:1:9:8:11:0	-0.93	16.75	16.77	13.89
Euclidean distance	4:0:48:46:35:27:28:13:2:2:0:17:27:21:2:13:6:0	0.68	13.75	13.77	11.78

Table 5 The results of the test set using recommendation model, the average error, best result and worst result of 32 models submitted to the CME scoreboard (Since the scorecard only has data from 2013, there are only 31 events for comparison in the end, unit: h).

CME events	Arrival time	Predicted results (Cosine distance) prediction error time	results error	Predicted results (Euclidean distance) prediction error time	results error	The average value of other models	The best result of other models	The worst result of other models
201303150712	46.78	31.92	-14.87	83.9	37.12	1.38	3.53	31.53
201304110724	63.5	48.12	-15.38	62.63	-0.87	-11.65	1.8	-36.2
201307091512	74.03	66.75	-7.28	68.48	-5.55	4.95	-7.12	12
201309292212	51.7	52.57	0.87	52.57	0.87	13.6	3.37	22.67
201312120336	81.4	32.2	-49.2	79.98	-1.42	9.27	4.18	13.63
201402040125	87.67	72.48	-15.18	72.48	-15.18	-10.2	1.73	-19.62
201402120600	79.27	66.25	-13.02	94.83	15.57	-15.08	-0.58	-25.95
201404021336	68.4	31.92	-36.48	68.48	2.68	0.33	0.33	-24.32
201404181325	45.52	31.43	-14.08	48.12	2.6	5.45	-1.22	25.63
201406041248	76.07	73.5	-2.57	72.35	-3.72	6.94	5.4	8.47
201408151748	85.15	66.25	-18.9	60.28	-24.87	-9.75	-9.37	-26.77
201409101800	45.88	43.55	-2.33	56.12	10.23	4.82	-0.43	29.18
201412170500	110.18	98.83	-11.35	98.83	-11.35	-32.22	-13.22	-39.22
201503150148	50.95	64.63	13.68	50.82	-0.13	9.83	7.52	13.92
201506210236	39.95	83.9	43.95	35.65	-4.3	1.82	-0.98	5.02
201506221836	42.88	38.55	-4.33	35.65	-7.23	4.68	-0.95	10.05
201508121448	65.68	70.63	4.95	67.57	1.88	21.12	16.7	25.83
201509041936	66.4	69.08	2.68	97.67	31.27	42.78	18.53	69.53
201509180500	49.07	40.82	-8.25	99.5	50.43	17.38	6.55	27.55
201511041448	51.5	63.5	12	73.5	22	13	4.85	21.43
201512160936	78.67	78.48	-0.18	67.03	-11.63	-14.17	-9.45	-18.45
201512281212	60.63	52.7	-7.93	83.9	23.27	-5.48	-3.48	-14.48
201601142324	94.55	85.15	-9.4	66.25	-28.3	-3.57	-8.97	-20.97
201604101112	92.38	87.1	-5.28	55.5	-36.88	-12.58	5.17	-25.98
201607171048	60.28	117.12	56.83	85.15	24.87	27.77	7.57	50.92
201607282224	111.6	100.6	-11	108.95	-2.65	-11.38	0.62	-23.38
201611050424	97.67	99.42	1.75	99.42	1.75	-18.62	-10.87	-29.87
201707140125	52.57	34.23	-18.33	60.63	8.07	9.42	3.77	16.67
201709042036	50.62	41.58	-9.03	68.4	17.78	-1.18	-1.13	-16.38
201709061224	34.23	83.9	49.67	68.4	34.17	14.27	7.5	23.5
201808202124	100.6	80.17	-20.43	78.52	-22.08	-3.9	-4.87	18.5

the Euclidean distance that was recommended the same historical events, and the error is very small. The two events took place on 2013 September 29 at 22:12 UT and 2016 November 5 at 4:24 UT. The recommend historical events are 2017 July 14 1:25 UT and 2016 October 8 18:36 UT.

Based on the prediction results of recommended model and other models, we list the four types of errors for the 31 events listed above in Table 6. Our mean error is -3.50 h under the cosine distance and 3.50 h under the Euclidean distance, respectively. After averaging the errors of all events for other models, the mean error of all events in average values is 1.9 h. The mean absolute error of cosine distance is 15.52 h, and the mean absolute error of Euclidean distance is 14.86 h, which has certain comparability with other models. Although the result is not

Table 6 Four Errors in Different Models (unit: h)

Models	ME	MAE	SD	RMSE
Cosine distance	-3.50	15.52	21.40	21.68
Euclidean distance	3.50	14.86	19.65	19.96
The average value of other models	1.90	11.57	14.72	14.84
The best result of other models	0.85	5.54	7.18	7.23
The worst result of other models	3.37	24.12	26.92	27.13

as good as the average errors of other models, whose mean absolute mean is 11.57 h. The main reason is that most of the 31 events are distributed in the training set, and the optimal weights are selected from the verification set. The errors for these events were ordinary, but the time error for the test set mentioned above was small.

Figure 3 shows the errors of the recommended model and other models. The black dotted line among image represents a statistical average of other model, and

yellow crosses express the optimal results of the other models. In contrast, blue crosses represent the worst result. Furthermore, our statistical results of cosine distance correspond to the red dots and green dots correspond to Euclidean distance, which fall nearly between the fork of blue and yellow, more balancing distribution near the black dotted line. The model prediction results for different events were fairly balanced, with no significant deviations, while other models do not guarantee that the prediction results were perfect every time.

4.2 Discussion

The purpose of using recommendation algorithm is not only giving the prediction of arrival time of a certain CME, but also introduce a similar event to operational forecasters, who can give a modified prediction according to the results of models and experiences. Such as, on 2015 August 12 14:48 UT, there was a partial halo CME erupting as shown in Figure 4(a)–(c). If the forecaster confirms that the CME will reach Earth, but does not know when it will arrive. We can input its parameters into the recommendation system and the system will recommend a similar historical CME on 1999 August 9 3:26 UT, which is shown in Figure 4(d)–(f). By comparing these two CME events, the propagation time of the former event is 65.68 h, and the arrival time of the latter is 67.57 h. Therefore, the recommended error is around 1.88 h, which is acceptable. Through the comparison of Figure 4, we can also find the similarities between the two events. First of all, both of the two eruptions are partial halo CME, and the CPA and MPA is very similar. In addition to this, the four CME speeds of two events are around 400 km s^{-1} . So the time of their arrival on Earth was very close. By directly comparing the images of the recommended events, the reliability of the model is proved on the other hand.

The events used in this study are only for the CMEs arriving on Earth, which limited the sample number used. Because there are so few cases, there is no guarantee that every event will have a similar historical outcome. Due to the limited number of samples, the accuracy of the model will be greatly limited. At present, the recommendation algorithm is outstanding in many directions, but no one has done similar experiments in the space physics, so our model is also a new attempt that used simple methods to achieve comparable results. Currently, our model was not yet able to predict whether CMEs can reach the Earth or not. Further attempts are expected to be made in the future, trying to use a large number of CME data in the history for training and matching. A more complete model can predict whether CME will hit the Earth firstly. If it will hit the Earth, the model will then give the forecast time and

recommend CME with a similar history for the forecast work reference.

5 SUMMARY

In this paper, we implemented a recommended algorithm to predict the arrival time of a new CME event based on the CME events in the history. Firstly, a multiple association list containing 18 characteristic parameters was made by integrating the CME list, near-Earth ICME list, solar flare list, active region and corresponding solar wind parameter. Secondly, the prediction method of CME arrival time was performed by applying a recommendation algorithm for the first time. The similarity between two CME events is calculated using cosine distance and Euclidean distance, respectively. The error analysis of test data indicates that the MAE is 11.78 h in the Euclidean distance test set, better than the result from cosine distance calculation. Thirdly, the result of the recommendation method was compared to that of other models issued on the CME scoreboard, which verifies the model reasonable. The result based on Euclidean distance can keep up with other models in the CME scoreboard. The recommendation model can not only predict the arrival time of CMEs, but also provide the information of similar historical events to operational space weather forecasters, which will be a helpful reference for artificial empirical prediction of CME effects.

Acknowledgements This CME catalog is generated and maintained at the CDAW Data Center by NASA and The Catholic University of America in cooperation with the Naval Research Laboratory. SOHO is a project of international cooperation between ESA and NASA. For near-Earth ICME list, we acknowledge all the various experimenters who have made their data available through the ACE Science Center and other sources. A portion of this work is supported by a NASA Heliophysics Guest Investigator Grant. The flare catalog is developed by the Smithsonian Astrophysical Observatory and is based on the Hinode Flare Catalog maintained by ISAS/JAXA and the Institute for Space-Earth Environmental Research (ISEE) at Nagoya. This work was supported by the National Natural Science Foundation of China (Grant Nos. 12071166 and 42074224).

Appendix A: NEAR-EARTH ICME EVENT LIST FOR THE ACTIVE REGION

Table A.1 Near-Earth ICME Event List for Active Region

No.	Near-Earth ICME	CME	NOAA ^a	Coordinate	Flare	Filament ^b
	Date and Time	Date and Time	Active Region		Date and Time	
1	199701100400	199701061510		S18E06	Flare	
2	199702100200	199702070030		S40W5		Y
3	199704110600	199704071427	8027	S28E12	199704071407	
4	199705150900	199705120530	8038	N21W19	199705120455	
5	199705261600	199705212100		N06W12	Only Sunspot	
6	199708031300	199707300445		N35W12		Y
7	199709031300	199708300130	8078	N14E05	199708292332	
8	199709212100	199709172028	8084	N22W88	199709171803	
9	199710011600	199709280108		N40W06		Y
10	199710102200	199710061528	8090	S28W10		Y
11	199710270000	199710231126		N45E39		Y
12	199711070400	199711040610	8100	S21W39	199711040558	
13	199712101800	199712061027		N50E05		Y
14	199712301000	199712260231		S20E14	Only Sunspot	
15	199801070100	199801022328		N30W27		Y
16	199801210600	199801170409	8137	S15E10	Only Active Region	
17	199801292000	199801251526	8144	N13E37		Y
18	199802171000	199802140655	8156	S24E16	Only Sunspot	
19	199803041300	199802281248	8169	S14W44	199803011302	
20	199805020500	199804291658	8210	S17E18	199804291637	
21	199805041000	199805021406	8210	S17W22	199805021342	
22	199806241600	199806210535	8251	N16W53	Only Active Region	
23	199810190400	199810151004		N22W01		Y
24	199811072200	199811040754	8375	N18W07	199811040719	
25	199811090100	199811052044	8375	N18W21	199811051955	
26	199811130200	199811091817	8378	N14W03		Y
27	199904161800	199904130330	8511	S16W33	Only Active Region	
28	199904210400	199904180830		N39E07		Y
29	199906272200	199906241331	8595	N39W08	199906241412	
30	199907062100	199907031954	8616	N31W86	Only Active Region	
31	199907271700	199907232130		N13W09	Only Sunspot	
32	199907311900	199907280906	8651	N24E65	199907280814	
33	199908120300	199908090326	8662	S16E21	Relate to 4 flare	
34	199909221900	199909200606		S21E01	199909200122?	
35	200001221700	200001181754	8831	S18E01	200001181726	
36	200002111600	200002080930	8858	N28E14	200002080900	
37	200002121200	200002100230	8858	N27W10	200002100208	
38	200002141200	200002120431	8858	N25W38	200002120410	
39	200002210500	200002171931	8878	S22W43	Only Active Region	
40	200004070600	200004041632	8933	N17W70	200004041541	
41	200004182000	200004151035	8963	N16E49	Only Active Region	
42	200005022000	200004290430	8976	S06E01	Only Active Region	
43	200005131700	200005102006	8990	N14E20	Only Active Region	
44	200005162300	200005131226	8996	S22E57	Only Active Region	
45	200005230900	200005201450		S11W22	Only Sunspot	
46	200005241200	200005210726		N18W03	Only Sunspot	
47	200006042200	200005310806	9974	N18E31	Only Active Region	
48	200006081200	200006061554	9026	N20E10	200006061525	
49	200007110200	200007071026		N04E00	Only Active Region	
50	200007131300	200007111327	9077	N18E33	200007111310	
51	200007141700	200007122030	10029	S27W10	Only Active Region	
52	200007151900	200007141054	9077	N18W09	200007141024	
53	200007200100	200007170854	10031	N14E72	Only Active Region	
54	200007270200	200007230530	9091	S12W06		Y
55	200007281200	200007250330	9097	N08W15	200007250249	
56	200008101900	200008062230	9104	S19W67	Only Active Region	
57	200008120500	200008091630	9122	N21E21	Only Active Region	
58	200009022200	200008291830	9142	N15E09	Only Active Region	
59	200009081200	200009050554		N29E12		Y
60	200009180000	200009160518	9165	N13W14	200009160426	
61	200010031000	200010020350	9176	S09E07	200010020013	
62	200010051300	200010022026	9176	S09E00	200010022004	
63	200010131600	200010092350	9182	N02W18	200010092343	
64	200010282100	200010250826		N10W66	200010251125	
65	200011061700	200011031826	9213	N03W09	200011031902	
66	200011270800	200011241530	9236	N22W07	200011241513	
67	200012230000	200012181150	9269	N14W06	200012181111	

Table A.1 *Continued.*

No.	Near-Earth ICME	CME	NOAA ^a	Coordinate	Flare	Filament ^b
	Date and Time	Date and Time	Active Region		Date and Time	
68	200101240900	200101202130	9313	S07E42	200101202120	
69	200103040400	200102281450	9360	S11W37	Only Active Region	
70	200103191700	200103160350		N37W50		Y
71	200103281700	200103251706	9402	N17E16	200103251636	
72	200103310500	200103281250	9393	N13E00	200103281240	
73	200104010400	200103291026	9393	N17W18	200103291015	
74	200104041800	200104022206	9393	N16W70	Only Active Region	
75	200104081400	200104061930	9415	S21E33	200104061921	
76	200104112200	200104100530	9415	S23W19	200104100526	
77	200104130900	200104111331	9415	S22W31	200104111326	
78	200104281400	200104261230	9433	N17W25	200104261312	
79	200107090200	200107050354	9516	N12W92	Only Active Region	
80	200108172000	200108141601	9571	N07W29	Only Active Region	
81	200109240000	200109201931	9619	N17W67	Only Active Region	
82	200109291100	200109270454	9627	S04W39	Only Active Region	
83	200110010800	200109280854	9636	N14E13	200109280830	
84	200110020400	200109291154		N00E07	flare	
85	200110120400	200110091130	9653	S22E11	200110091111	
86	200110212000	200110191650	9661	N16W35	200110191630	
87	200110270300	200110221826	9672	S18E13	200110221508	
88	200110292200	200110251526	9672	S18W27	200110251502	
89	200111061200	200111041635	9684	N05W28	200111041619	
90	200111192200	200111170530	9704	S18E28	200111170525	Y
91	200111241400	200111222330	9704	S18W38	200111222329	
92	200112300000	200112260530	9742	N08W54	200112260504	
93	200203190500	200203152306	9866	S09W06	200203152310	
94	200204171600	200204150350	9906	S15W14	200204150355	
95	200204200000	200204170826	9906	S14W41	200204170824	
96	200205111500	200205081350	9934	S16W19	200205081327	
97	200205201000	200205160050	9948	S22E01	200205150813	
98	200205232000	200205220350	10000	S12W60	200205220354	Y
99	200207181200	200207152030	10030	N18E00	200207152008	
100	200208191200	200208161230	10069	S07E11	200208161232	
101	200209080400	200209051654	10102	N08E26	200209051706	
102	200209192000	200209170806	114	S11W43	Only Active Region	
103	200305300200	200305280050	10365	S07W32	200305272307	
104	200305302200	200305290127	10365	S07W45	200305290022	
105	200306171000	200306140154		N25W27		Y
106	200308180100	200308142006	10431	S10E02	200308140611	
107	200310242100	200310220830	10484	N07E25	200310211922	
108	200310280230	200310261754	10484	N04W43	200310261721	
109	200310291100	200310281130	10486	S17E04	200310281110	
110	200310310200	200310292054	10486	S16W11	200310292049	
111	200311201000	200311180850	10501	N03E09	200311180831	
112	200401220800	200401200006	10540	S13W09	200401200045	
113	200401232300	200401210454		S30E39	200401210511	Y
114	200407221800	200407201331	10652	N10E32	200407201232	
115	200407241400	200407220731	10652	N08E06	Only Active Region	
116	200407252000	200407231606	10652	N08W10	Only Active Region	
117	200407270200	200407251454	10652	N08W35	200407251514	
118	200409141500	200409120036	10672	N05E33	200409120056	
119	200409181200	200409141012	10672	N05E10	200409140930	
120	200411072200	200411042330	10691	N09E19	200411042248	
121	200411092000	200411071654	10696	N08W22	200411071605	
122	200411120800	200411100226	10696	N08W62	200411100213	
123	200412122200	200412082026	10709	N04W11	200412081959	
124	200501082100	200501051530	10715	N04W33	Only Active Region	
125	200501161400	200501131754	10718	S07E07	200501131712	
126	200501171300	200501152306	10720	N13W03	200501152258	
127	200501182300	200501170930	10720	N13W30	200501170610	
128	200501211900	200501200654	10720	N14W70	200501200701	
129	200502201200	200502170006	10734	S05W34	200502162338	
130	200505200300	200505161350	10759	N11W35	200505161246	
131	200505300100	200505261506	10767	S08E12	200505261420	
132	200505310400	200505262126	10767	S08E12	Only Active Region	
133	200507101000	200507071706	10786	N12W04	200507071629	

Table A.1 Continued.

No.	Near-Earth ICME	CME	NOAA ^a	Coordinate	Flare	Filament ^b
	Date and Time	Date and Time	Active Region		Date and Time	
134	200508090000	200508050854	10795	N13E14	Only Active Region	
135	200508241400	200508220131	10798	S11W62	200508220133	
136	200509021800	200508311130	10803	N11W13	200508311151	
137	200509110500	200509091948	10808	S09E54	200509092002	
138	200509151400	200509132000	10808	S09E10	200509131905	
139	200604131500	200604100606	10869	S05W20	200504100842	
140	200607102100	200607060854	10898	S08W41	200607060836	
141	200608201300	200608161630	10904	S12W15	200608161617	
142	200608302000	200608262057	10905	S08E05	200608261952	
143	200612142200	200612130254	10930	S05W33	200612130239	
144	200612170000	200612142230	10930	S05W47	200612142214	
145	200711192300	200711151850		S07W18	No active region in solar disk	
146	200812170300	200812120854	1009	S25W90	Only Active Region	
147	200912191300	200912160430	1035	N30W18	200912160124	
148	201002110800	201002070354	1045	N23W01	201002070234	
149	201004051200	201004031033	1059	S22W15	201004030954	
150	201004120100	201004080454	1061	N17W44	Only Active Region	
151	201005281900	201005241406	AR	N18W32	20100524144	Y
152	201102181900	201102150224	1158	S21W27	201102150156	
153	201103060900	201103030612		S14W16	Only Active Region	
154	201108050500	201108020636	1261	N15W21	Only Active Region	
155	201108062200	201108040412	1261	N15W49	201108040357	
156	201109100300	201109062305	1283	N14W18	201109062220	
157	201109262000	201109241248	1302	N12E47	201109241320	
158	201111020100	201110271200	1330	N08E05	Not clear	
159	201111131000	201111091336	1342	N17E22	201111091335	
160	201111290000	201111260712	1353	N08W49	201111260710	
161	201201210600	201201181224		S19E03		Y
162	201201222300	201201191436	1402	N29E15	201201191605	
163	201203090300	201203070024	1429	N17E15	201203070024	
164	201203151700	201203131736	1429	N18W62	201203131741	
165	201205161600	201205120000	1447	S15E20	flare	Y
166	201206162300	201206141412	1504	S16E01	201206141435	
167	201207050000	201207020836	1515	S17E06	201207020703	
168	201207090000	201207041724	1513	N17W36	201207041639	
169	201207150600	201207121648	1520	S16W09	201207120805	
170	201209050600	201209020400	1560	N03W05	201209020158	
171	201210010000	201209280012	1577	N08W41	201209272357	
172	201210081800	201210050248	1584	S22W40	201210050317	
173	201211010000	201210271648	1598	S15W11	flare	Y
174	201211130800	201211091512	1608	S21E10	Only Active Region	
175	201211241200	201211201200	1618	N08E14	Only Active Region	
176	201211261200	201211231348		S39E10		Y
177	201301171600	201301131200	1640	N28W35		Y
178	201303171500	201303150712	1692	N09W03	201303150658	
179	201304141700	201304110724	1719	N10W00	201304110716	
180	201306280200	201306232236	1776	N11W63	Only Active Region	
181	201307130500	201307091512	1785	N18E15	201307091325	
182	201310022300	201309292212		N17W27	201309292339	Y
183	201310090900	201310061443		S32E46	201310061424	
184	201312151600	201312120336		S24W40	Only Active Region	
185	201402080100	201402040125	1967	S12W24	flare	
186	201402160500	201402120600	1974	S12W12	201402120425	
187	201402191200	201402161000	1977	S11E01	201402160926	
188	201404052200	201404021336	2027	N12E42	201404021405	
189	201404210700	201404181325	2036	S16W41	201404181303	
190	201406082000	201406041248	2080	S12E46	Only Active Region	
191	201408191600	201408151748	2144	S17W33	Only Active Region	
192	201409122200	201409101800	2158	N15W00	201409101745	
193	201409170200	201409121824		N16W26		Y
194	201412220400	201412170500	2242	S18W01	201412170451	
195	201501070700	201501030312	2253	S07E09	Only Active Region	
196	201503171300	201503150148	2297	S18W38	201503150213	
197	201506230200	201506210236	2371	N13W00	201506210142	
198	201506251000	201506221836	2371	N13W13	201506221823	

Table A.1 *Continued.*

No.	Near-Earth ICME	CME	NOAA ^a	Coordinate	Flare	Filament ^b
	Date and Time	Date and Time	Active Region		Date and Time	
199	201507130600	201507100224	2384	S28E47	201507100121	
200	201508152100	201508121448	2396	S18W65	Only Active Region	
201	201509080000	201509041936	2410	S26E13	Only Active Region	
202	201509210800	201509180500	2415	S19W29	201509180404	
203	201511070600	201511041448	2443	N06W09	201511041352	
204	201512200300	201512160936	2468	S16W13	201512160903	
205	201512311700	201512281212	2473	S22W19	201512281245	
206	201601191000	201601142324		S24W13		Y
207	201604140900	201604101112	2529	N14E49	201604100549	
208	201607200700	201607171048	2565	N05W08	201607170803	
209	201608021400	201607282224		N26W10		Y
210	201610130600	201610081836	2600	N06E49	Only Active Region	
211	201611100000	201611050424	2606	N08E46	Only Active Region	
212	201707161500	201707140125	2667	N12W85	201707132340	
213	201709072000	201709042036	2673	S08W16	201709042033	
214	201709081100	201709061224	2673	S09W45	201709061202	
215	201808251200	201808202124		N17W12		Y

^a Number of active region specified by NOAA; ^b Y indicates that the activity associated with a CME event is filament.

References

- Ameri D., & Valtonen E. 2017, *Sol. Phys.*, 292, 79
- Brueckner, G. E., Howard, R. A., Koomen, M. J., et al. 1995, *Sol. Phys.*, 162, 357
- Cane, H., & Richardson, I. 2003, *Journal of Geophysical Research (Space Physics)*, 108, 1156
- Detman, T., Smith, Z., Dryer, M., et al. 2006, *Journal of Geophysical Research (Space Physics)*, 111, A07102
- Dryer, M., Fry, G., Sun, W., et al. 2001, *Sol. Phys.*, 204, 265
- Feng, X., & Zhao, X. 2006, *Sol. Phys.*, 238, 167
- Feng, X., Zhou, Y., & Wu S. 2007, *ApJ*, 655, 1110
- Ham, Y.-G., Kim, J.-H., & Luo J.-J. 2019, *Nature*, 573, 568
- Hess, P., & Zhang, J. 2015, *ApJ*, 812, 144
- Huang, Y., Zheng, F., Cong, R., et al. 2020, *IEEE Transactions on Image Processing*, 29, 8187
- Jin, M., Manchester, W., van der Holst, B., et al. 2017, *ApJ*, 834, 173
- Jolliffe, I. T., & Stephenson, D. B. 2012, *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, 2nd Edition (Hoboken: John Wiley & Sons)
- Kang, Y., Li, L., & Li, B. 2021, *Journal of Energy Chemistry*, 54, 72
- Li, Y., & Luhmann, J. 2006, *ApJ*, 648, 732
- Liou, K., Sotirelis, T., & Richardson, I. 2018, *Journal of Geophysical Research (Space Physics)*, 123, 485
- Liu, J., Ye, Y., Shen, C., Wang, Y., & Erdélyi, R. 2018, *ApJ*, 855, 109
- Liu, S., Yao, S., Zhu, G., Zhang, X., & Yang, R., 2021, *Journal of Intelligent & Robotic Systems*, 101, <https://doi.org/10.1007/s10846-020-01289-8>
- Manoharan, P. 2006, *Sol. Phys.*, 235, 345
- Maričić, D., Vršnak, B., Veronig, A. M., et al. 2020, *Sol. Phys.*, 295, 91
- Moon, Y. J., Dryer, M., Smith, Z., Park, Y., & Cho, K. 2002, *Geophys. Res. Lett.*, 29, 1390
- Möstl, C., Amerstorfer, T., Frahm, R., et al. 2015, *Nature Communications*, 6, 7135
- Müller, D., Nicula, B., Felix, S., et al. 2017, *A&A*, 606, A10
- Núñez, M., Nieves-Chinchilla, T., & Pulkkinen, A. 2016, *Space Weather*, 14, 544
- Odstrčil, D., Riley, P., & Zhao, X. 2004, *Journal of Geophysical Research (Space Physics)*, 109, A02116
- Paouris, E., & Mavromichalaki, H. 2017, *Sol. Phys.*, 292, 180
- Poedts, S., Lani, A., Scolini, C., et al. 2020, *Journal of Space Weather and Space Climate*, 10, 57
- Richardson, I., & Cane, H. 2010, *Sol. Phys.*, 264, 189
- Riley, P., Linker, J., Lionello, R., & Mikić, Z. 2012, *Journal of Atmospheric and Solar-Terrestrial Physics*, 83, 1
- Riley, P., Linker, J., & Mikić, Z., 2013, *Journal of Geophysical Research (Space Physics)*, 118, 600
- Riley, P., Mays, L., Andries, J., et al. 2018, *Space Weather*, 16, 1245
- Schultz M., Betancourt C., Gong B., et al., 2021, *Philosophical Transactions Series A*, 379, 20200097
- Schwenn, R., dal Lago, A., Huttunen, E., & Gonzalez, W. 2005, *Annales Geophysicae*, 23, 1033
- Sinha, S., Srivastava, N., & Nandy, D. 2019, *ApJ*, 880, 84
- Smith, Z., & Dryer, M., 1990, *Sol. Phys.*, 129, 387
- Sokolov, I., van der Holst, B., Oran, R., et al. 2013, *ApJ*, 764, 23
- Subramanian, P., Lara, A., & Borgazzi, A. 2012, *Geophys. Res. Lett.*, 39, L19107
- Sudar, D., Vršnak, B., & Dumbović, M. 2016, *MNRAS*, 456, 1542
- Tóth, G., Sokolov, I., Gombosi, T., et al., 2005, *Journal of Geophysical Research (Space Physics)*, 110, A12226

- van der Holst, B., Sokolov, I., Meng, X., et al. 2014, *ApJ*, 782, 81
- Vandas, M., Fischer, S., Dryer, M., Smith, Z., & Detman, T. 1996, *J. Geophys. Res.*, 101, 15645
- Verbeke, C., Mays, M. L., Temmer, M., et al. *Space Weather*, 2019, 17, 6
- Vršnak, B., Žic, T., Vrbanec, D., et al. 2013, *Sol. Phys.*, 285, 295
- Wang, C.-P., Kim, H.-J., Yue, C., et al. 2017, *Journal of Geophysical Research (Space Physics)*, 122, 4210
- Wang, J., Ao, X., Wang, Y., et al. 2018, *Journal of Space Weather and Space Climate*, 8, A39
- Wang, Q.-J., Li, J.-C., Guo, L.-Q., 2021, *RAA (Research in Astronomy and Astrophysics)*, 21, 012
- Wang, Y., Liu, J., Jiang, Y., & Erdélyi, R. 2019, *ApJ*, 881, 15
- Wang, Y., Ye, P.Z., Wang, S., et al. 2002, *Journal of Geophysical Research (Space Physics)*, 107, 1340
- Watanabe, K., Masuda, S., & Segawa, T. 2012, *Sol. Phys.*, 279, 317
- Webb, D. F., & Howard, T. A. 2012, *Living Reviews in Solar Physics*, 9, 3
- Xie, H., Ofman, L., & Lawrence, G. 2004, *Journal of Geophysical Research (Space Physics)*, 109, A08103
- Yashiro, S., Gopalswamy, N., St. Cyr, O., et al. 2004, *Journal of Geophysical Research (Space Physics)*, 109, A07105
- Zhao, X., & Dryer, M. 2014, *Space Weather*, 12, 448
- Zheng, F., Shao, L., & Han J. 2018, *IEEE Transactions on Intelligent Transportation Systems*, 19, 3387