

Validation of the CME Geomagnetic Forecast Alerts Under the COMESEP Alert System

Mateja Dumbović^{1,2}  · Nandita Srivastava^{3,4} · Yamini K. Rao^{3,5} · Bojan Vršnak¹ · Andy Devos⁶ · Luciano Rodriguez⁶

Received: 18 October 2016 / Accepted: 2 June 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract Under the European Union 7th Framework Programme (EU FP7) project Coronal Mass Ejections and Solar Energetic Particles (COMESEP, <http://comesep.aeronomy.be>), an automated space weather alert system has been developed to forecast solar energetic particles (SEP) and coronal mass ejection (CME) risk levels at Earth. The COMESEP alert system uses the automated detection tool called Computer Aided CME Tracking (CACTus) to detect potentially threatening CMEs, a drag-based model (DBM) to predict their arrival, and a CME geoeffectiveness tool (CGFT) to predict their geomagnetic impact. Whenever CACTus detects a halo or partial halo CME and issues an alert, the DBM calculates its arrival time at Earth and the CGFT calculates its geomagnetic risk level. The geomagnetic risk level is calculated based on an estimation of the CME arrival probability and its likely geoeffectiveness, as well as an estimate of the geomagnetic storm duration. We present the evaluation of the CME risk level forecast with the COMESEP alert system based on a study of geoeffective CMEs observed during 2014. The validation of the forecast tool is made by comparing the forecasts with observations. In addition, we test the success rate of the automatic forecasts (without human intervention) against the forecasts with human intervention using advanced versions of the DBM and CGFT (independent tools available at the Hvar

Earth-affecting Solar Transients

Guest Editors: Jie Zhang, Xochitl Blanco-Cano, Nariaki Nitta, and Nandita Srivastava

✉ M. Dumbović
mdumbovic@geof.hr; mateja.dumbovic@uni-graz.at

✉ N. Srivastava
nandita@prl.res.in; url: <http://www.cessi.in>

¹ Hvar Observatory, Faculty of Geodesy, University of Zagreb, Zagreb, Croatia

² Institute of Physics, University of Graz, Graz, Austria

³ Udaipur Solar Observatory, Physical Research Laboratory, Udaipur, India

⁴ Centre for Excellence in Space Sciences, Kolkata, India

⁵ Present address: Indian Institute of Technology (BHU), Varanasi, India

⁶ Solar-Terrestrial Center of Excellence – SIDC, Royal Observatory of Belgium (ROB), Brussels, Belgium

Observatory website, <http://oh.geof.unizg.hr>). The results indicate that the success rate of the forecast in its current form is unacceptably low for a realistic operation system. Human intervention improves the forecast, but the false-alarm rate remains unacceptably high. We discuss these results and their implications for possible improvement of the COMESEP alert system.

Keywords Coronal mass ejections, initiation and propagation · Coronal mass ejections, interplanetary · Magnetosphere, geomagnetic disturbances

1. Introduction

Coronal mass ejections (CMEs) and their interplanetary counterparts (interplanetary coronal mass ejections, ICMEs) are regarded as the main drivers of space weather that cause the most intense geomagnetic storms (*e.g.* Gonzalez *et al.*, 1998; Plunkett *et al.*, 2001; Zhang *et al.*, 2003; Koskinen and Huttunen, 2006; Echer *et al.*, 2008; Yermolaev *et al.*, 2012, and references therein). Therefore, one of the main aspects of space weather prediction is the estimation of the possible CME geomagnetic impact, *i.e.* its geoeffectiveness. Based on the widely accepted concept proposed by Dungey (1961), the enhanced CME geoeffectiveness is related to the effective reconnection with the geomagnetic field and therefore with the southern component of the ICME magnetic field, B_s , and the corresponding y component of the convective electric field, $E_y = vB_s$ (where v is the solar wind speed). This is confirmed with statistical studies based on *in situ* properties of ICMEs and indices of geomagnetic activity, such as the disturbance storm time index, Dst (*e.g.* Kane, 2005; Richardson and Cane, 2011; Verbanac *et al.*, 2013, and references therein).

The present prediction scheme of the geomagnetic storm magnitude is generally reliable and depends on measured ICME *in situ* properties. The prediction of the Dst index is done in real time by various groups (*e.g.* Feldstein, 1992; Fenrich and Luhmann, 1998; O'Brien and McPherron, 2000) on the basis of inputs from the original formula of Burton, McPherron, and Russell (1975). However, *in situ* measurements of ICMEs are provided at the L1 Lagrangian point, *i.e.* about one hour before the start of the disturbance (for typical ICME speed), providing very limited “response time” (*e.g.* Koskinen and Huttunen, 2006; Richardson and Cane, 2011). Our current knowledge restricts us from predicting the crucial B_s component of the ICME magnetic field at earlier times, *e.g.* from remote solar observations. Some studies have tried to compare the magnetic field of the ICME to its solar source region magnetic field in the initiation phase (*e.g.* Bothmer and Schwenn, 1994; Möstl *et al.*, 2008). However, even if the original orientation of the magnetic field inside the CME were known in the initiation phase, the prediction of the B_s component at Earth would be severely hampered by the fact that CMEs rotate while propagating (*e.g.* Vourlidas *et al.*, 2011; Isavnin, Vourlidas, and Kilpua, 2013).

From the solar perspective, analysis of many years of solar images recorded by the *Solar and Heliospheric Observatory* (SOHO: Domingo, Fleck, and Poland, 1995) has revealed that fast full-halo CMEs associated with strong flares and originating close to the central meridian and low and mid-latitudes are potentially favourable candidates for producing strong geomagnetic storms (see, *e.g.*, Zhang *et al.*, 2003; Srivastava and Venkatakrishnan, 2004; Gopalswamy, Yashiro, and Akiyama, 2007; Zhang *et al.*, 2007; Richardson and Cane, 2010; Dumbović *et al.*, 2015, and references therein). The prediction schemes that rely on the remotely measured CME properties are found to be much less reliable when the interplanetary conditions are not taken into account (*e.g.* Srivastava, 2005; Valach *et al.*, 2009; Kim *et al.*, 2010; Uwahoro, McKinnell, and Habarulema, 2012), however, the advantage

is in the early warning. This type of advance forecasting requires the identification of key solar parameters that determine the geoeffectiveness of a CME and an empirical probabilistic model to estimate its geomagnetic impact. The development of the Coronal Mass Ejections and Solar Energetic Particles (COMESSEP, <http://comessep.aeronomy.be>) alert system is a significant step in this direction. This operational space weather alert system has been running since the beginning of 2014, fully automatically, without human intervention, and among other things, it forecasts geomagnetic storms and performs risk analysis for selected user groups.

In the present article, we have analysed all the COMESSEP alerts issued in 2014 with the objective to test the performance of the COMESSEP forecasting tools, in particular, those implemented to forecast the geomagnetic impact of the geoeffective CMEs. Verification of the forecast of space weather mainly implies a quantitative and qualitative assessment of the performance of space weather prediction alerts, as has been done for terrestrial weather forecasting in the past. Recent efforts on various space weather forecast validations include those performed by, *e.g.*, Berghmans *et al.* (2005), Balch (2008), Crown (2012), and Devos, Verbeeck, and Robbrecht (2014). In addition, the evaluation is continuously performed for several world-wide space weather prediction centres (*e.g.* the Space Weather Prediction Center, SWPC, in US, or Solar Influences Data Analysis Centre, SIDC, in Belgium) by the National Institute of Information and Communications Technology (NICT, http://seg-web.nict.go.jp/cgi-bin/forecast/eng_forecast_score.cgi). This type of analysis not only reveals the strength and the limitations of the tools used for forecasting, but also provides scope for advancing new and improved versions of tools to provide a better and accurate forecasting of space weather.

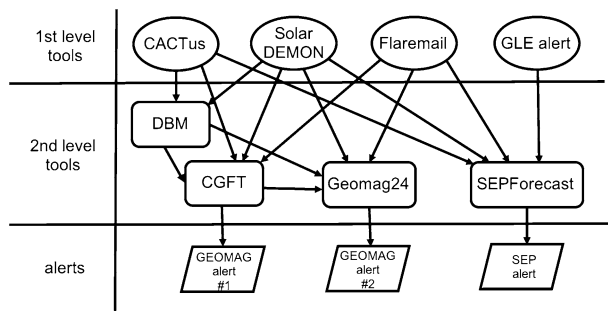
2. Description of Forecast Tools

The COMESSEP alert system is the first fully automatic system for the detection of CMEs and solar flares, forecasting of CME arrival, and assessing their potentially hazardous impact (Crosby *et al.*, 2012). It was developed within a European Union 7th Framework Programme (EU FP7) project Coronal Mass Ejections and Solar Energetic Particles (COMESSEP) and runs fully automatically, *i.e.* without human intervention. It consists of several interconnected tools, model or data based, that work together to automatically provide geomagnetic and solar energetic particles (SEP) radiation storm alerts. In general, the tools running under the COMESSEP alert system can be divided into two categories: first-level producers (first-level tools) and tools that are both consumers and producers (second-level tools). First-level tools use near real-time data to monitor and automatically detect potentially hazardous events, based on which, they issue the alert. The alerts issued by first-level tools trigger second-level tools, which then, based on the input provided by first-level tools, produce their own alerts. The process is visualised in the flow diagram in Figure 1.

In this article we focus on the part of the COMESSEP alert system that forecasts geomagnetic impact of CMEs several days in advance. It consists of two second-level tools:

- The Drag-based Model (DGM) is a model for heliospheric propagation of CMEs, based on the assumption that the dominant force in the heliospheric dynamics of ICMEs is the magnetohydrodynamical equivalent of the aerodynamic drag. In its basic form (which is implemented in the COMESSEP system), it provides the ICME Sun–Earth transit time, the arrival time, and the impact speed for a given set of input parameters (see Vršnak *et al.*, 2013, 2014). The tool and its description are available at <http://oh.geof.unizg.hr/DBM/dbm.php>, as is the advanced form of the DBM, which provides this output for any target in the heliosphere and also takes the shape of the ICME

Figure 1 Flow diagram of the tools used in the COMESEP alert system (for an explanation, see the main text).



into account by employing the so-called cone geometry (described in Žic, Vršnak, and Temmer, 2015).

- The CME Geomagnetic Forecast Tool (CGFT) is an empirical statistical model that determines the CME geomagnetic risk level based on a probability estimation of the CME arrival and its likely geoeffectiveness (both are derived based on remote parameters of a CME and its associated flare). In addition, the estimated storm duration is based on the estimated geoeffectiveness and the month of the eruption, with the start time determined by the DBM ICME arrival. A more comprehensive description of the tool is given below.

Figure 1 shows that these second-level tools receive alerts from three first-level tools that detect a potentially hazardous (*i.e.* Earth-directed) CME and flare event:

- The Computer Aided CME Tracking (CACTus) is a software package that autonomously detects CMEs (see, *e.g.*, Berghmans, Foing, and Fleck, 2002; Robbrecht and Berghmans, 2004) in image sequences from the *Large Angle Spectroscopic Coronagraph* (LASCO: Brueckner *et al.*, 1995) onboard SOHO, and whenever a halo or partial-halo CME is detected, it sends an alert to the COMESEP system. The catalogue of CMEs detected by CACTus and its description are available at <http://sidc.be/cactus/>.
- The Solar Dimming and EUV Wave Monitor (Solar DEMON) is a software package that automatically and in real time detects solar flares using data from the *Atmospheric Imaging Assembly* (AIA: Lemen *et al.*, 2012) onboard the *Solar Dynamics Observatory* (SDO: Pesnell, Thompson, and Chamberlin, 2012) and provides information such as location, time, and relative intensity (Kraaikamp and Verbeeck, 2015). The event catalogue of the tool and its description are available at <http://solardemon.oma.be>.
- Flaremail is a tool that monitors GOES soft X-ray flux data in near real-time, and whenever an M- or X-class flare is detected, it sends an alert to the COMESEP system (<http://sidc.oma.be/products/flaremail/>).

To supplement the early geomagnetic impact forecast given by CGFT, there is also a tool that estimates the risk level of geomagnetic storm occurrence for the next 24 hours (Geomag24, <http://www.spaceweather.space.dtu-dk/forskning/Geomag24>). Geomag24 takes into account recent alerts issued by the DBM and CGFT, observations of the last months of geomagnetic activity, and solar wind observations combined with estimates of the background solar wind speed and coronal hole area (to include the risk of high-speed solar wind streams), as well as current *in situ* solar wind and geomagnetic data to estimate the risk of a geomagnetic storm for the next 24 hours. In addition, there are supplementary tools for SEP forecast. The first-level producer called the ground-level enhancement tool (GLE alert, available at <http://cosray.phys.uoa.gr/index.php/glealertplus>) monitors possible GLE events and upon detection issues an alert to the COMESEP system, and the SEP forecast tool predicts the probability and level for a radiation storm with proton energies > 10 MeV and

> 60 MeV resulting from a flare. It primarily uses an input from Flaremail, but when available, it also uses input from Solar DEMON, CACTus, and GLE alert. Since the Geomag24 and SEP forecast tools are not the focus of this study, further details are not given, but they can be found at the COMESEP webpage (<http://comesep.aeronomy.be>).

We focus our analysis on CGFT alerts that are triggered by DBM and receive input from CACTus, Solar DEMON, and Flaremail (when available). CGFT can be divided into three modules: module I, which estimates the probability of CME arrival, module II, which estimates CME geo effectiveness, and module III, which estimates the storm duration based on the estimated geoeffectiveness and the month of the eruption (based on an analysis similar to Yokoyama and Kamide, 1997). In this study we analyse and estimate the correctness of the forecast of the geomagnetic impact of the CME and not the estimated storm duration, *i.e.* we analyse the forecast capability of only the first two CGFT modules.

CGFT module I: The probability of a CME arrival is estimated through a statistical model that relates the flare source position with a potential CME arrival at Earth. Each bin of longitude is associated with a probability of arrival. The probability of arrival is categorised into the following levels: very unlikely (0–10%), unlikely (10–40%), possible (40–70%), likely (70–90%), and very likely (90–100%). The arrival estimation of CMEs is based on the position of the source location, *i.e.* bin of longitude: very unlikely (< -60 or > 60 deg), unlikely (-60 to -30 or 30 to 60 deg), possible (-30 to -10 or 10 to 30 deg), and very likely (-10 to 10 deg). The source position is known for an association with a detected flare of at least X-ray class M. In case of absence of an associated flare, the likelihood of arrival is set to “possible”. This statistical model is based on the general observation that CMEs originating closer to the centre of the solar disc are more likely to arrive at Earth (*e.g.*, Gopalswamy *et al.*, 2000), but is not quantified based on a robust statistical measure. Therefore, the ranges of values for specific longitudinal bins, as well as the corresponding probabilities, are chosen in a somewhat arbitrary way.

CGFT module II: The geoeffectiveness of a CME is related to the disturbance storm time index, Dst, and is determined by a statistical model relating each single solar parameter to the absolute value of Dst, |Dst|. The full set of parameters that can be used by this tool includes the CME apparent width and speed, the associated flare soft X-ray class and source position, and the CME-CME interaction parameter (for details see Dumbović *et al.*, 2015). Since in a fully automatic system the calculation of the CME-CME interaction parameter is not trivial, this parameter is not used in the currently operating COMESEP alert system. In addition, the tool can work with only a partial set of available parameters (*e.g.* only with CME speed). The probabilities for each parameter are combined to one probability for each |Dst| bin (for details see Dumbović *et al.*, 2015). The probability distribution across the |Dst| bins is converted into the estimate of one single |Dst| bin, using specific thresholds on calculated probabilities, P . For example, the set of thresholds [0, 0.20, 0.12, 0.10] contains a threshold 0 on $P(|Dst| < 100)$, 0.20 on $P(100 < |Dst| < 200)$, 0.12 on $P(200 < |Dst| < 300)$, and 0.10 on $P(300 < |Dst| < 400)$. These steps are applied in this specific order to assign the estimated impact to a single |Dst| bin. When a condition is fulfilled, that |Dst| bin is chosen and the process is stopped. The thresholds are as follows: $P(300 < |Dst| < 400) > 0.10 \rightarrow 300 < |Dst| < 400$, $P(200 < |Dst| < 300) > 0.12 \rightarrow 200 < |Dst| < 300$, $P(100 < |Dst| < 200) > 0.20 \rightarrow 100 < |Dst| < 200$, and $P(50 < |Dst| < 100) > 0 \rightarrow 50 < |Dst| < 100$.

For each combination of arrival probability and storm level, a risk level is defined according to the CME geomagnetic risk matrix (Figure 2). There are four possible risk levels that provide an indication of the severity of the alerts sent: extreme risk (red), high risk (orange), moderate or medium risk (yellow), and low risk (green). The CGFT issues an alert when a risk level is estimated to be greater than “low”. The process is started by a CME alert

STORM LEVEL (Dst)	SEVERE (Dst >300)	MODERATE	HIGH	HIGH	HIGH	EXTREME
	STRONG (Dst =200-300)	MODERATE	MODERATE	MODERATE	HIGH	HIGH
	MODERATE (Dst =100-200)	LOW	MODERATE	MODERATE	MODERATE	HIGH
	MINOR (Dst <100)	LOW	LOW	LOW	MODERATE	MODERATE
ARRIVAL PROBABILITY		very unlikely (0-10%)	unlikely (10-40%)	possible (40-70%)	likely (70-90%)	very likely (90-100%)

Figure 2 CME geomagnetic risk matrix.

from CACTus that is sent to both the CGFT and DBM. The DBM calculates the arrival time of the CME at Earth and triggers the CGFT, but the CGFT does not use DBM parameters as an input. The CGFT receives the CME parameter width and speed from CACTus, and when the CME is associated with a flare, the flare strength is provided by Flaremail and the position of the source region of the flare from Solar DEMON. The CGFT modules I and II calculate arrival probability and geoeffectiveness levels, respectively, and estimate the risk of a geomagnetic storm, based on the CME geomagnetic risk matrix shown in Figure 2. If the estimated risk is higher than “low”, the CGFT issues an alert.

3. Data and Method

We compiled a list of all alerts issued by COMESEP/CGFT in 2014 (a total of 72 alerts). For each event, COMESEP/DBM issued a forecast of CME arrival time and speed, and COMESEP/CGFT issued an alert with geomagnetic storm risk level based on arrival probability and estimated storm level. We cross-checked the list with that of Richardson and Cane (2010), which is continuously updated at <http://www.srl.caltech.edu/ACE/ASC/DATA/level3/icmetable2.htm>, and with *in situ* measurements by the *Magnetic Field Investigation* (MFI: Lepping *et al.*, 1995) and *Solar Wind Experiment* (SWE: Ogilvie *et al.*, 1995) instruments onboard the *Wind* spacecraft in a time frame of ±24 hours around the forecast DBM arrival time. In time periods when *Wind* observations were unavailable, we used the *Advanced Composition Explorer* (ACE: Stone *et al.*, 1998) spacecraft measurements. In addition, we checked the Dst levels (available at http://wdc.kugi.kyoto-u.ac.jp/dst_provisional/index.html) around the observed ICME arrival times (when the ICME arrival is observed), or around forecast ICME arrivals when the ICME arrival is not observed. In this way, for each issued alert, we obtained information whether it was correctly forecast. In order to make a full evaluation of the forecasts we also needed the information on geomagnetic storms which occurred, but were not forecast. In this way, we compiled a list of all geomagnetic storms with Dst < −90 nT caused by ICMEs in 2014. Although the Dst threshold in the CGFT module II for relevant geomagnetic activity is −100 nT, here we allow a −10 nT error in the determination of the geoeffectiveness level and set −90 nT as a threshold value for the relevant geomagnetic activity. We found that in 2014 there were only three geomagnetic storms with Dst < −90 nT, a −116 nT storm on 19 February, a −91 nT storm on 20 February, and a −94 nT storm on 28 February. All three storms were forecast by COMESEP/CGFT,

Table 1 Event statistics for 2014 for a) COMESEP alert system (without human intervention), and b) COMESEP + human scheme (with human intervention). CGeFT stands for CME Geoeffectiveness Forecast Tool, an independent, more recent version of the module II of CGFT (for an explanation, see the main text).

Total number of detected CMEs (CACTus catalogue)	1855
Number of detected CMEs with width > 120 deg (CACTus catalogue)	98
Number of $ Dst < -90$ nT storms	3
a) COMESEP	
Total number of issued CGFT alerts	72
Number of CGFT alerts with medium risk	65
Number of CGFT alerts with high risk	7
Number of CGFT alerts with module II prediction $ Dst > -100$ nT	47
Number of CGFT alerts with module II prediction -200 nT < $ Dst < -100$ nT	0
Number of CGFT alerts with module II prediction -300 nT < $ Dst < -200$ nT	23
Number of CGFT alerts with module II prediction $ Dst < -400$ nT	2
b) COMESEP + human	
Total number of CGeFT alerts	33
Number of CGeFT alerts with $ Dst > -100$ nT	23
Number of CGeFT alerts with -200 nT < $ Dst < -100$ nT	8
Number of CGeFT alerts with $ Dst < -300$ nT	3

although the actual estimated $|Dst|$ level was not correct in all three events (in two events the COMESEP/CGFT module II calculated $|Dst| < 100$ nT, but because of module I calculation and application of the risk matrix, alerts were issued). Finally, we compiled a list of all CACTus CMEs with angular width larger than 120 deg to determine the number of potential alerts that could have been issued, but were not issued because the risk level was calculated to be “low”. The corresponding numbers are given in Table 1.

For each COMESEP/CGFT alert we checked what the forecast would be with human intervention, *i.e.* if upon receiving the alert from CACTus, the forecaster would use the most recent and advanced version of DBM and CGFT together with all available observations (*e.g.* measurements from the *Solar-Terrestrial Relations Observatory*, STEREO: Kaiser *et al.*, 2008). For this purpose, we used a number of available online event catalogues under the assumption that the same information would be obtained by forecasters using available satellite observation. The SOHO/LASCO CME catalogue (Yashiro *et al.*, 2004, http://cdaw.gsfc.nasa.gov/CME_list/) was used as an observer catalogue of CMEs and their parameters (plane-of-sky speed, apparent angular width). In addition, we used the COR1 CME catalogue from the two STEREO spacecraft (available at <http://cor1.gsfc.nasa.gov/catalog/>) to eliminate backside events (in 2014 both STEREO spacecraft had a field of view on the back side of the Sun) and movies from the *Sun Earth Connection Coronal and Heliospheric Investigation* (SECCHI) suite onboard STEREO (available at http://secchi.nrl.navy.mil/index.php?p=js_secchi) for time periods where the COR1 CME catalogue had no entries. In order to estimate the source region of the CME, we used the Solar DEMON flare catalogue (available at http://solardemon.oma.be/science/flares.php?min_seq=1&min_flux_est=0.00000001&days=0&science=1), the NOAA GOES flare list (available at <http://www.ngdc.noaa.gov/stp/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/xrs/>), and the SDO/AIA fil-

ament eruption list (available at <http://aia.cfa.harvard.edu/filament/>). These catalogues provided information needed to forecast whether a CME was expected to arrive at Earth, together with its arrival and geoeffectiveness level. To forecast the arrival of CMEs, an advanced version of the DBM was used (ADBM, available at <http://oh.geof.unizg.hr/DBM/dbm.php>, for details see Žic, Vršnak, and Temmer, 2015, or the same webpage under “Documentation”), and for the geoeffectiveness level, we used a more recent version of the CGFT module II (CGeFT, available at <http://oh.geof.unizg.hr/CGEFT/cgeft.php>, for details see the same webpage under “Documentation”). It should be noted that CGeFT makes a rougher estimation of the $|\text{Dst}|$ level than CGFT module II with the aim to be more reliable (possible $|\text{Dst}|$ levels are < 100 nT, $100\text{--}200$ nT, and > 200 nT). The number of alerts in this COMESEP + human scheme is reduced to 66 because the CGeFT also takes the CME-CME interaction parameter into account (which is determined by the observer, in this case by M. Dumbović). In several cases, we found that COMESEP/CGFT issued two alerts for two separate CMEs, whereas the observer would issue only one alert because the two detected CMEs are likely or very likely to interact and result in one storm event. In total we found that out of these 66 events, 13 were likely or very likely involved in a CME-CME interaction. Furthermore, out of these 66 alerts, four were found to be CACTus false alerts (reported CMEs were not found on either the revised level 2 CACTus or SOHO/LASCO CME lists), and 29 were found to be DBM false alerts (either backside events or the advanced version of DBM, which takes into account CME geometry, calculated that the CME will miss Earth). Therefore, in this COMESEP + human scheme there were in fact only 33 alerts that would be triggered using CGeFT. The corresponding numbers for COMESEP + human scheme are given in Table 1.

We evaluated the forecast by comparing the predicted value with the observed value using verification measures for binary events, *i.e.* events with two possible outcomes. Each binary event can have only two possible forecast outcomes (it was either forecast or not) and two possible observed outcomes (it was either observed or not), which were then combined into four possible outcomes: a hit (forecast and observed), a miss (not forecast, but observed), a false alarm (forecast, but not observed), and a correct rejection (not forecast and not observed). For a list of events (*i.e.* an evaluation sample), the number of hits, misses, false alarms, and correct rejections can be represented in a contingency table for a binary event, which is then used to calculate verification measures (see, *e.g.*, Devos, Verbeeck, and Robbrecht, 2014; Dumbović, Vršnak, and Čalogović, 2016). As verification measures we used the probability of detection (POD), the false-alarm ratio (FAR), bias (BIAS), and Heidke skill score (HSS), where each gives only a partial information on the quality of forecast. The POD measures the success rate (regardless of the number of false alarms), the FAR measures the false-alarm rate (regardless of the success rate), BIAS measures the ratio of the frequency of forecasts to the frequency of observations (*i.e.* whether the system is more inclined to forecast more or fewer events), and the HSS estimates the accuracy of the forecast relative to that of random chance (for details see, *e.g.*, Devos, Verbeeck, and Robbrecht, 2014; Dumbović, Vršnak, and Čalogović, 2016).

Since CGFT makes a forecast of $|\text{Dst}|$ in four different ranges, a dichotomisation must be applied, *i.e.* events need to be separated into two classifications in order to produce a contingency table for a binary event. A geomagnetic storm can be regarded as a binary event by defining a specific threshold – it occurs or it does not occur, where for the CGFT the value of the threshold can be taken as the CGFT threshold for relevant geoeffectiveness $|\text{Dst}| > 100$ nT, *i.e.* $|\text{Dst}| > 90$ nT when we allow ± 10 nT error in the observation (for simplicity and clarity reasons we refer to this threshold as $|\text{Dst}| > 100$ nT). With this dichotomisation, a hit is defined as an event where $|\text{Dst}| > 100$ nT was both forecast and observed, a miss is

Table 2 Contingency table for a binary event.

Observation	Forecast	
	Yes	No
Yes	Hit:	Miss:
	$ \text{Dst} > 100$ nT forecast	$ \text{Dst} < 100$ nT forecast
	$ \text{Dst} > 100$ nT observed	$ \text{Dst} > 100$ nT observed
	CGFT-1	CGFT-2
	CGeFT-2	CGeFT-1
No	False alarm:	Correct rejection:
	$ \text{Dst} > 100$ nT forecast	$ \text{Dst} < 100$ nT forecast
	$ \text{Dst} < 100$ nT observed	$ \text{Dst} < 100$ nT observed
	CGFT-24	CGFT-71
	CGeFT-9	CGeFT-86

an event that was forecast as $|\text{Dst}| < 100$ nT, but observed as $|\text{Dst}| > 100$ nT, a false alarm is an event forecast as $|\text{Dst}| > 100$ nT, but observed as $|\text{Dst}| < 100$ nT, and correct rejection is an event where $|\text{Dst}| < 100$ nT was both forecast and observed. A contingency table for a binary event based on this dichotomisation is given in Table 2. It should be noted that this dichotomisation approach does not discern between events with $|\text{Dst}| > 100$ nT, which have $|\text{Dst}|$ in different ranges (*e.g.* $100 \text{ nT} < |\text{Dst}| < 200 \text{ nT}$ and $|\text{Dst}| > 200 \text{ nT}$), but since in the chosen time period only three $|\text{Dst}| > 100$ nT events were observed, it is the most suitable one.

4. Evaluation Results and Discussion

The verification measures for COMESEP/CGFT alerts (without human intervention) and CGeFT alerts, *i.e.* COMESEP + human scheme (with human intervention), are shown in Figure 3. The figure shows that the POD for the COMESEP + human scheme is much higher (two times) than that of COMESEP/CGFT without human intervention. However, the FAR (given in Figure 3b) is quite high in both cases, and for COMESEP/CGFT without human intervention, it is very close to the value of the worst forecast. This indicates that the system (regardless of the human intervention) has a tendency to forecast more events and issue false alerts, which is supported by the result for BIAS, which is much larger than 1 in both cases (see Figure 3c). Finally, Figure 3d shows the HSS, which is almost 0 for the COMESEP/CGFT system without human intervention, but improves to 0.25 with human intervention. These results indicate that without human intervention the forecast of the system is not much better than a random guess, but shows some skill when a human observer intervenes.

The evaluation results for the CGFT module II using verification measures for the binary events reveal many shortcomings. While the automatic system shows some ability to detect geomagnetic storms (as seen in POD), the false-alarm rate is unacceptably high for a realistic operation system, resulting in an overall skill that is not better than a random guess (as seen in HSS). The forecast skill somewhat improves with human intervention, but even with human intervention, the FAR remains unacceptably high. The reasons for this poor early prediction probably lie in the fact that the system 1) does not use “real” CME input

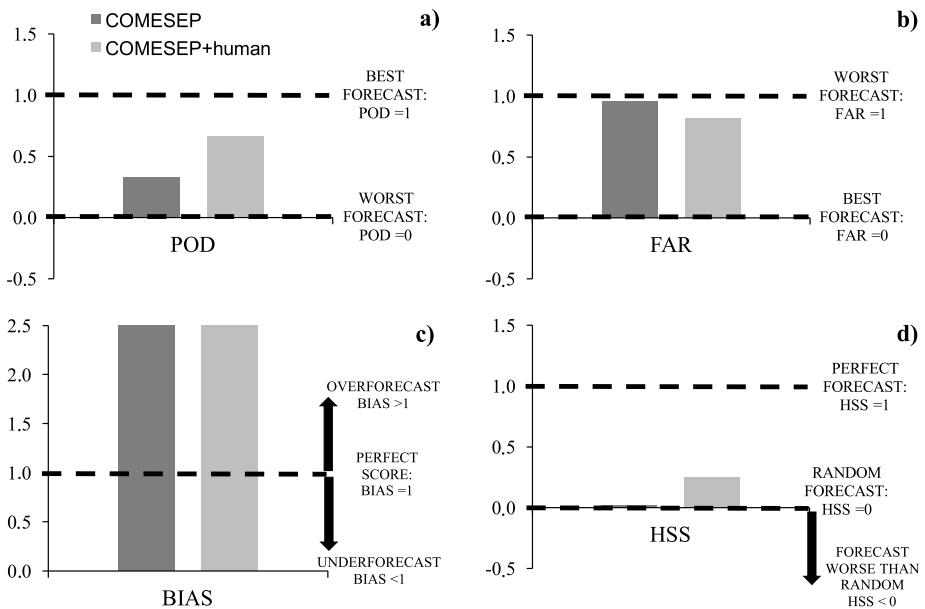


Figure 3 Evaluation measures for COMESEP alerts (without human intervention) and COMESEP + human scheme (with human intervention).

values, but projected ones, 2) does not incorporate CME propagation and evolution effects in the calculation, and most notably 3) cannot distinguish between different magnetic field orientations in CMEs.

In addition, several other facts have to be taken into consideration. Firstly, evaluation measures were calculated for module II, which estimates the possible storm level, while CGFT also takes the probability of arrival into consideration in issuing alerts. Table 1 shows that out of 72 issued CGFT alerts in 47 cases, the CGFT module II estimated a low storm risk with $|Dst| < 100$ nT. Based on the risk matrix given in Figure 2, this means that in these cases, the alert was issued solely because the probability of the arrival was high. Out of 47 cases of low storm activity calculated by the CGFT module II, in 45 $|Dst| < 100$ nT was indeed observed, *i.e.* in these cases no alert was necessary. We therefore conclude that including the arrival probability to determine the CME risk can lead to a large number of unnecessary alerts, and it might be useful for the COMESEP alert system to revise the CGFT CME risk matrix. Secondly, the evaluation was made for 2014, where the geomagnetic activity was quite low, since there were only three major geomagnetic storms and none of them was stronger than $|Dst| > 200$ nT. The system should also be tested for periods with higher geomagnetic activity to verify its prediction capability for such severe geomagnetic storms.

On the other hand, 2014 corresponds to the second peak of Solar Cycle 24, which had a quite high CME activity. Table 1 shows that CACTus detected 1855 CMEs in 2014, and of these, 98 were halo and partial-halo CMEs, which are regarded as potentially threatening, but only for 72 of them alerts were issued. In addition, the CGFT module II determined that only 25 out of 72 events would be relevantly geoeffective, whereas with human intervention this number is additionally reduced to only 11 forecasts. Therefore, we conclude that the automatic COMESEP system has a high success rate regarding correct rejections. This is the most important aspect of the system, which is basically fed with a huge amount of data, out of which it is supposed to resolve which data or events need to be discarded.

Finally, our results show that with human intervention, the forecast is somewhat improved, but it should be taken into account that in this analysis the observer (M. Dumbović) had STEREO spacecraft observations available, which could more easily identify backside events. For this specific time period (2014), the view of the back side of the Sun was possible because of the STEREO spacecraft location, which might not be the case in the future, when the observer will have to rely on the same observation used by the COMESEP alert system. In addition, we used the more recent and advanced version of the tool. Furthermore, human intervention resulted in a much better input for the CGFT, where CMEs could be associated with flares with soft X-ray classes lower than M, and the CME-CME interaction parameter could also be derived. This strongly suggests that the COMESEP/CGFT forecast could be improved in the future by advancing the input it receives and by implementing more recent and advanced tools.

5. Summary and Conclusion

We evaluated the performance of the fully automatic COMESEP/CGFT tool for issuing alerts on geoeffective and possibly hazardous CMEs. The evaluation revealed many shortcomings of the tool, such as an exceedingly high false-alarm rate and low forecast skill, which is not much better than the random forecast. In addition, because it is combined with the CME arrival probability in order to assess the risk, the system quite often issues unnecessary alerts. The main value of the system (in its current state) lies in its ability to refine the number of potentially hazardous CMEs, as it shows a relatively good correct rejection rate, which additionally improves with human intervention. In general, we find that the CME geoeffectiveness estimation is improved with human intervention, primarily because the human observer is not limited to the observation tools implemented in the system and thus uses a more reliable and extensive input, as well as advanced tools. However, even with human intervention, the false-alarm rate is unacceptably high for a realistic operation system. It should be noted, however, that the performance of the system is yet to be tested for a time period where the geomagnetic activity is much higher. The system should be improved for possible future applications by using the most reliable possible input and incorporating CME propagation and evolution effects, but most notably by including the CME magnetic field orientation as one of the most important aspects of the geomagnetic storm forecast. Since the system is designed such that different tools communicate with each other, it could be easily upgraded by implementing new tools and more recent versions of tools that are already implemented into the system. This and the fact that it is fully automatic and works completely without human intervention makes it a promising starting point for a future operation system for space weather early prediction.

Acknowledgements Open access funding is provided by University of Graz. This work has received funding from the European Commission FP7 Project COMESEP (263252). M.D. and B.V. acknowledge the support of Croatian Science Foundation under the project 6212 “Solar and Stellar Variability” and of European social fund under the project “PoKRet”. The contribution by Y.K.R. is part of her work as a Project Student at Udaipur Solar Observatory/Physical Research Laboratory, Udaipur. L.R. acknowledges support from the Belgian Federal Science Policy Oce through the ESA-PRODEX programme. This research has been partially funded by the Interuniversity Attraction Poles Programme initiated by the Belgian Science Policy Oce (IAP P7/08 CHARM).

Disclosure of Potential Conflicts of Interest The authors declare that they have no conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Balch, C.C.: 2008, Updated verification of the Space Weather Prediction Center's solar energetic particle prediction model. *Space Weather* **6**, S01001. [DOI](#). [ADS](#).
- Berghmans, D., Foing, B.H., Fleck, B.: 2002, Automated detection of CMEs in LASCO data. In: Wilson, A. (ed.) *From Solar Min to Max: Half a Solar Cycle with SOHO, ESA SP 508*, 437. 92-9092-816-6. [ADS](#).
- Berghmans, D., van der Linden, R.A.M., Vanlommel, P., Warnant, R., Zhukov, A., Robbrecht, E., Clette, F., Podladchikova, O., Nicula, B., Hochedez, J.-F., Wauters, L., Willems, S.: 2005, Solar activity: Nowcasting and forecasting at the SIDC. *Ann. Geophys.* **23**, 3115. [DOI](#). [ADS](#).
- Bothmer, V., Schwenn, R.: 1994, Eruptive prominences as sources of magnetic clouds in the solar wind. *Space Sci. Rev.* **70**, 215. [DOI](#). [ADS](#).
- Brueckner, G.E., Howard, R.A., Koomen, M.J., Korendyke, C.M., Michels, D.J., Moses, J.D., Socker, D.G., Dere, K.P., Lamy, P.L., Llebaria, A., Bout, M.V., Schwenn, R., Simnett, G.M., Bedford, D.K., Eyles, C.J.: 1995, The Large Angle Spectroscopic Coronagraph (LASCO). *Solar Phys.* **162**, 357. [DOI](#). [ADS](#).
- Burton, R.K., McPherron, R.L., Russell, C.T.: 1975, An empirical relationship between interplanetary conditions and Dst. *J. Geophys. Res.* **80**, 4204. [DOI](#). [ADS](#).
- Crosby, N.B., Veronig, A., Robbrecht, E., Vršnak, B., Vennerstrom, S., Malandraki, O., Dalla, S., Rodriguez, L., Srivastava, N., Hesse, M., Odstrcil, D., Consortium, C.: 2012, Forecasting the space weather impact: The COMESEP project. In: Hu, Q., Li, G., Zank, G.P., Ao, X., Verkhoglyadova, O., Adams, J. H. (eds.) *Space Weather: The Space Radiation Environment: 11th Annual International Astrophysics Conference, Amer. Inst. Phys. CP 1500*, 159. [DOI](#).
- Crown, M.D.: 2012, Validation of the NOAA Space Weather Prediction Center's solar flare forecasting look-up table and forecaster-issued probabilities. *Space Weather* **10**, S06006. [DOI](#). [ADS](#).
- Devos, A., Verbeeck, C., Robbrecht, E.: 2014, Verification of space weather forecasting at the Regional Warning Center in Belgium. *J. Space Weather Space Clim.* **4**(27), A29. [DOI](#). [ADS](#).
- Domingo, V., Fleck, B., Poland, A.I.: 1995, The SOHO mission: An overview. *Solar Phys.* **162**, 1. [DOI](#). [ADS](#).
- Dumbović, M., Vršnak, B., Čalogović, J.: 2016, Forbush decrease prediction based on remote solar observations. *Solar Phys.* **291**, 285. [DOI](#). [ADS](#).
- Dumbović, M., Devos, A., Vršnak, B., Sudar, D., Rodriguez, L., Ruždjak, D., Leer, K., Vennerstrøm, S., Veronig, A.: 2015, Geoeffectiveness of coronal mass ejections in the SOHO era. *Solar Phys.* **290**, 579. [DOI](#). [ADS](#).
- Dungey, J.W.: 1961, Interplanetary magnetic field and the auroral zones. *Phys. Rev. Lett.* **6**, 47. [DOI](#).
- Echer, E., Gonzalez, W.D., Tsurutani, B.T., Gonzalez, A.L.C.: 2008, Interplanetary conditions causing intense geomagnetic storms ($Dst \leq -100$ nT) during Solar Cycle 23 (1996–2006). *J. Geophys. Res.* **113**, 5221. [DOI](#). [ADS](#).
- Feldstein, Y.I.: 1992, Modelling of the magnetic field of magnetospheric ring current as a function of interplanetary medium parameters. *Space Sci. Rev.* **59**(1), 83. [DOI](#).
- Fenrich, F.R., Luhmann, J.G.: 1998, Geomagnetic response to magnetic clouds of different polarity. *Geophys. Res. Lett.* **25**, 2999. [DOI](#). [ADS](#).
- Gonzalez, W.D., de Gonzalez, A.L.C., Dal Lago, A., Tsurutani, B.T., Arballo, J.K., Lakhina, G.K., Buti, B., Ho, C.M., Wu, S.-T.: 1998, Magnetic cloud field intensities and solar wind velocities. *Geophys. Res. Lett.* **25**, 963. [DOI](#). [ADS](#).
- Gopalswamy, N., Yashiro, S., Akiyama, S.: 2007, Geoeffectiveness of halo coronal mass ejections. *J. Geophys. Res.* **112**, 6112. [DOI](#). [ADS](#).
- Gopalswamy, N., Lara, A., Lepping, R.P., Kaiser, M.L., Berdichevsky, D., St. Cyr, O.C.: 2000, Interplanetary acceleration of coronal mass ejections. *Geophys. Res. Lett.* **27**, 145. [DOI](#). [ADS](#).
- Isavnin, A., Vourlidis, A., Kilpua, E.K.J.: 2013, Three-dimensional evolution of erupted flux ropes from the Sun ($2-20 R_{\odot}$) to 1 AU. *Solar Phys.* **284**, 203. [DOI](#). [ADS](#).
- Kaiser, M.L., Kucera, T.A., Davila, J.M., St. Cyr, O.C., Guhathakurta, M., Christian, E.: 2008, The STEREO mission: An introduction. *Space Sci. Rev.* **136**, 5. [DOI](#). [ADS](#).
- Kane, R.P.: 2005, How good is the relationship of solar and interplanetary plasma parameters with geomagnetic storms? *J. Geophys. Res.* **110**, A02213. [DOI](#). [ADS](#).

- Kim, R.-S., Cho, K.-S., Moon, Y.-J., Dryer, M., Lee, J., Yi, Y., Kim, K.-H., Wang, H., Park, Y.-D., Kim, Y.H.: 2010, An empirical model for prediction of geomagnetic storms using initially observed CME parameters at the Sun. *J. Geophys. Res.* **115**(A14), 12108. DOI. ADS.
- Koskinen, H.E.J., Huttunen, K.E.J.: 2006, Geoeffectivity of coronal mass ejections. *Space Sci. Rev.* **124**, 169. DOI. ADS.
- Kraaikamp, E., Verbeek, C.: 2015, Solar Demon – an approach to detecting flares, dimmings, and EUV waves on SDO/AIA images. *J. Space Weather Space Clim.* **5**(27), A18. DOI. ADS.
- Lemen, J.R., Title, A.M., Akin, D.J., Boerner, P.F., Chou, C., Drake, J.F., Duncan, D.W., Edwards, C.G., Friedlaender, F.M., Heyman, G.F., Hurlburt, N.E., Katz, N.L., Kushner, G.D., Levay, M., Lindgren, R.W., Mathur, D.P., McFeaters, E.L., Mitchell, S., Rehse, R.A., Schrijver, C.J., Springer, L.A., Stern, R.A., Tarbell, T.D., Wuelser, J.-P., Wolfson, C.J., Yanari, C., Bookbinder, J.A., Cheimets, P.N., Caldwell, D., Deluca, E.E., Gates, R., Golub, L., Park, S., Podgorski, W.A., Bush, R.I., Scherrer, P.H., Gummin, M.A., Smith, P., Auker, G., Jerram, P., Pool, P., Soufli, R., Windt, D.L., Beardsley, S., Clapp, M., Lang, J., Waltham, N.: 2012, The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 17. DOI. ADS.
- Lepping, R.P., Acuña, M.H., Burlaga, L.F., Farrell, W.M., Slavin, J.A., Schatten, K.H., Mariani, F., Ness, N.F., Neubauer, F.M., Whang, Y.C., Byrnes, J.B., Kennon, R.S., Panetta, P.V., Scheifele, J., Worley, E.M.: 1995, The wind magnetic field investigation. *Space Sci. Rev.* **71**, 207. DOI. ADS.
- Möstl, C., Miklenic, C., Farrugia, C.J., Temmer, M., Veronig, A., Galvin, A.B., Vršnak, B., Biernat, H.K.: 2008, Two-spacecraft reconstruction of a magnetic cloud and comparison to its solar source. *Ann. Geophys.* **26**, 3139. DOI. ADS.
- O'Brien, T.P., McPherron, R.L.: 2000, An empirical phase space analysis of ring current dynamics: Solar wind control of injection and decay. *J. Geophys. Res.* **105**, 7707. DOI. ADS.
- Ogilvie, K.W., Chornay, D.J., Fritzenreiter, R.J., Hunsaker, F., Keller, J., Lobell, J., Miller, G., Scudder, J.D., Sittler, E.C. Jr., Torbert, R.B., Bodet, D., Needell, G., Lazarus, A.J., Steinberg, J.T., Tappan, J.H., Mavretic, A., Gergin, E.: 1995, SWE, a comprehensive plasma instrument for the WIND spacecraft. *Space Sci. Rev.* **71**, 55. DOI. ADS.
- Pesnell, W.D., Thompson, B.J., Chamberlin, P.C.: 2012, The Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 3. DOI. ADS.
- Plunkett, S.P., Thompson, B.J., St. Cyr, O.C., Howard, R.A.: 2001, Solar source regions of coronal mass ejections and their geomagnetic effects. *J. Atmos. Solar-Terr. Phys.* **63**, 389. DOI. ADS.
- Richardson, I.G., Cane, H.V.: 2010, Near-Earth interplanetary coronal mass ejections during Solar Cycle 23 (1996–2009): Catalog and summary of properties. *Solar Phys.* **264**, 189. DOI. ADS.
- Richardson, I.G., Cane, H.V.: 2011, Geoeffectiveness (Dst and Kp) of interplanetary coronal mass ejections during 1995–2009 and implications for storm forecasting. *Space Weather* **9**, 7005. DOI. ADS.
- Robbrecht, E., Berghmans, D.: 2004, Automated recognition of Coronal Mass Ejections (CMEs) in near-real-time data. *Astron. Astrophys.* **425**, 1097. DOI. ADS.
- Srivastava, N.: 2005, A logistic regression model for predicting the occurrence of intense geomagnetic storms. *Ann. Geophys.* **23**(9), 2969. DOI.
- Srivastava, N., Venkatakrishnan, P.: 2004, Solar and interplanetary sources of major geomagnetic storms during 1996–2002. *J. Geophys. Res.* **109**(A18), 10103. DOI. ADS.
- Stone, E.C., Frandsen, A.M., Mewaldt, R.A., Christian, E.R., Margolies, D., Ormes, J.F., Snow, F.: 1998, The Advanced Composition Explorer. *Space Sci. Rev.* **86**, 1. DOI. ADS.
- Uwamahoro, J., McKinnell, L.A., Habarulema, J.B.: 2012, Estimating the geoeffectiveness of halo CMEs from associated solar and IP parameters using neural networks. *Ann. Geophys.* **30**, 963. DOI. ADS.
- Valach, F., Revallo, M., Bochníček, J., Hejda, P.: 2009, Solar energetic particle flux enhancement as a predictor of geomagnetic activity in a neural network-based model. *Space Weather* **7**, 4004. DOI. ADS.
- Verbanac, G., Živković, S., Vršnak, B., Bandić, M., Hojsak, T.: 2013, Comparison of geoeffectiveness of coronal mass ejections and corotating interaction regions. *Astron. Astrophys.* **558**, A85. DOI. ADS.
- Vourlidas, A., Colaninno, R., Nieves-Chinchilla, T., Stenborg, G.: 2011, The first observation of a rapidly rotating coronal mass ejection in the middle corona. *Astrophys. J. Lett.* **733**, L23. DOI. ADS.
- Vršnak, B., Žic, T., Vrbanec, D., Temmer, M., Rollett, T., Möstl, C., Veronig, A., Čalogović, J., Dumbović, M., Lulić, S., Moon, Y.-J., Shanmugaraju, A.: 2013, Propagation of interplanetary coronal mass ejections: The drag-based model. *Solar Phys.* **285**, 295. DOI. ADS.
- Vršnak, B., Temmer, M., Žic, T., Taktakishvili, A., Dumbović, M., Möstl, C., Veronig, A.M., Mays, M.L., Odstrčil, D.: 2014, Heliospheric propagation of coronal mass ejections: Comparison of numerical WSA-ENLIL + cone model and analytical drag-based model. *Astrophys. J. Suppl.* **213**, 21. DOI. ADS.
- Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O.C., Plunkett, S.P., Rich, N.B., Howard, R.A.: 2004, A catalog of white light coronal mass ejections observed by the SOHO spacecraft. *J. Geophys. Res.* **109**, 7105. DOI. ADS.

- Yermolaev, Y.I., Nikolaeva, N.S., Lodkina, I.G., Yermolaev, M.Y.: 2012, Geoeffectiveness and efficiency of CIR, sheath, and ICME in generation of magnetic storms. *J. Geophys. Res.* **117**, A00L07. [DOI](#). [ADS](#).
- Yokoyama, N., Kamide, Y.: 1997, Statistical nature of geomagnetic storms. *J. Geophys. Res.* **102**, 14215. [DOI](#). [ADS](#).
- Zhang, J., Dere, K.P., Howard, R.A., Bothmer, V.: 2003, Identification of solar sources of major geomagnetic storms between 1996 and 2000. *Astrophys. J.* **582**, 520. [DOI](#). [ADS](#).
- Zhang, J., Richardson, I.G., Webb, D.F., Gopalswamy, N., Huttunen, E., Kasper, J.C., Nitta, N.V., Poomvises, W., Thompson, B.J., Wu, C.-C., Yashiro, S., Zhukov, A.N.: 2007, Solar and interplanetary sources of major geomagnetic storms ($\text{Dst} \leq -100$ nT) during 1996–2005. *J. Geophys. Res.* **112**(A11), 10102. [DOI](#). [ADS](#).
- Žic, T., Vršnak, B., Temmer, M.: 2015, Heliospheric propagation of coronal mass ejections: Drag-based model fitting. *Astrophys. J. Suppl.* **218**, 32. [DOI](#). [ADS](#).