submitted to **Journal of Space Weather and Space Climate**

The Probabilistic Solar Particle Event foRecasting (PROSPER) Model

Athanasios Papaioannou[1], Rami Vainio[2], Osku Raukunen[2,3], Piers Jiggens[4], Angels Aran[5,6,7], Mark Dierckxsens[8], Sotirios A. Mallios[1], Miikka Paassilta[2], and Anastasios Anastasiadis[1]

[1] Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens, I. Metaxa & Vas. Pavlou St. GR-15236, Penteli, Greece
e-mail: atpapaio@astro.noa.gr
[2] Department of Physics and Astronomy, University of Turku, 20014 Turku, Finland
[3] Aboa Space Research Oy, Tierankatu 4B, 20520 Turku, Finland
[4] European Space Research and Technology Centre (ESTEC), Space Environment and Effects Section, Keplerlaan 1, 2200AG Noordwijk, The Netherlands
[5] Dep. Física Quàntica i Astrofísica (FQA), Universitat de Barcelona (UB), c. Martí i Franquès, 1, 08028 Barcelona, Spain
[6] Institut de Ciències del Cosmos (ICCUB), Universitat de Barcelona (UB), c. Martí i Franquès, 1, 08028 Barcelona, Spain
[7] Institut d'Estudis Espacials de Catalunya (IEEC), c. Gran Capità, 2-4, 08034 Barcelona, Spain
[8] Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Avenue Circulaire 3, 1180 Uccle, Belgium

ABSTRACT

The Probabilistic Solar Particle Event foRecasting (PROSPER) model predicts the probability of occurrence and the expected peak flux of Solar Energetic Particle (SEP) events. Predictions are derived for a set of integral proton energies (i.e. E>10, >30 and >100 MeV) from characteristics of solar flares (longitude, magnitude), coronal mass ejections (width, speed) and combinations of both. Herein the PROSPER model methodology for deriving the SEP event forecasts is described and the validation of the model, based on archived data, is presented for a set of case studies. The PROSPER model has been incorporated into the new operational Advanced Solar Particle Event Casting System (ASPECS) tool to provide nowcasting (short term forecasting) of SEP events as part of ESA's future SEP Advanced Warning System (SAWS). ASPECS also provides the capability to interrogate PROSPER for historical cases via a run on demand functionality.

Key words. solar flares – coronal mass ejections – solar energetic particles – operational tool – validation – radiation storms

1. Introduction

Solar energetic particle (SEP) events constitute a significant component of the near Earth radiation environment and consist of protons, electrons and heavier ions (Vainio et al. 2009). Such events originate from particle acceleration in solar flares (SFs) and/or shocks associated with coronal mass ejections (CMEs) (e.g. Reames 2015; Vlahos et al. 2019). Once energetic particles are accelerated and injected into open magnetic field lines, they are consequently routed through the interplanetary magnetic field (IMF) (Belov et al. 2005; Cane and Lario 2006). Subsequently, when an observer (i.e. spacecraft) is magnetically connected to the source of the particles, enhancements by several orders of magnitude above the pre-event background are observed *in-situ*, with SEP events often observed in a broad range of solar longitudes (Rouillard et al. 2012; Lario et al. 2016). SEPs can last from a few hours to several days and their relative composition varies by many orders of magnitude from event to event (Reames 2013; Desai and Giacalone 2016). The classical paradigm divides SEP events into two categories, see e.g. Reames (1999): those particles that are accelerated at SFs (Aschwanden 2002) are known as impulsive events, and other particle populations that are accelerated by near-Sun CME-driven shocks are termed as gradual (Reames 1999, 2002; Kahler 2001; Cane and Lario 2006). However, this “two class” picture does not match the diversity and wealth of the observed SEP event properties which indicate a more complex nature (Cane et al. 2010; Papaioannou et al. 2016; Vlahos et al. 2019).

SEP events have a direct space weather impact on electronics and humans (Baker 2004). For example, the survivability of a spacecraft is directly affected by the total energy deposited by the passage of energetic particles. This may result in the degradation and ultimate failure of its electronic components due to ionisation or displacement damage mechanisms (Daly et al. 1996; Feynman and Gabriel 2000). Effects on electronics also include single event effects (SEEs) (Pellish et al. 2010), these appear when particles deposit sufficient energy or charge in a sensitive region of a component, with non-permanent (soft) errors such as bit flips, as well as, permanent (hard) errors such as latchups (i.e. SELs) and burnouts (SEBs) taking place (Sexton 2003; Harboe-Sørensen 2013). Furthermore, SEP events are a major threat to human spaceflight outside the protective shield of the Earth’s magnetosphere (Bizzarri et al. 2017; Townsend 2021) and pose a severe danger for aircrews and passengers on polar flights (Tobiska et al. 2015; Mishev et al. 2015; Miroshnichenko 2018). The ionizing particle radiation can lead to damage to human cells and DNA alternations (Azzam et al. 2012). There are two basic parameters that have a decisive role in the radiation effects in humans: the strength of the exposure and the specific organ/tissue that was encountered. For example, a short-term exposure to high doses of radiation depositing energy in the eye can lead to acute radiation effects such as cataracts. Long-term exposure to low doses of radiation deposing energy in bone marrow may progressively lead to leukaemia or other types of cancer. The former effects are categorized as deterministic while the latter as stochastic (Cucinotta et al. 2003). The high-energy tail of the SEP spectrum, in which ions are accelerated to relativistic energies, is dominant in spacecraft orbits well within the magnetosphere. In addition, high-energy particles upon interaction with matter, from e.g. the shielding material of a spacecraft, produce secondaries that can enhance the radiation effect of SEP events and are of major concern in heavily shielded environments such as human spaceflight (Raukunen et al. 2018). Nonetheless, these high-energy particles can also reach the Earth’s atmosphere and generate secondaries through nuclear reactions. Consequently, these secondaries are recorded as a significant sudden increase at ground level, as detected by e.g. neutron monitors (NMs). Such events are termed as Ground-level enhancements (GLEs) (Bütikofer et al. 2009; Asvestari et al. 2017; Mishev et al. 2018). This results in enhanced ionisation together with modifications of the local chemistry of the high polar atmosphere of the Earth (Usoskin et al. 2011; Mironova and Usoskin 2014).

SEPs, along with trapped radiation in planetary environments, are responsible for cumulative (dose) effects on spacecraft electronics and materials under low and moderate amounts of shielding. These are modelled statistically to produce specification (climatological) models to predict degradation over the duration of a mission. Such models make use of fits to SEP flux data in the form of lognormal, truncated power law, and exponential cut-off power law distributions to derive fluxes as a function of mission duration and confidence (Jiggens et al. (2018); Raukunen et al. (2018) and references therein). However, these models cannot make short-term forecasts of the SEP environment to provide warnings for operational missions, human spaceflight, launch operators and aircraft operators.

A small fraction of all solar flares and CMEs lead to SEP events and as a result the prognosis of SEPs is not a trivial task, since this is a highly imbalanced problem (see the relevant discussion in Lavasa et al. 2021). The scientific questions that one needs to address in such studies (see e.g. Anastasiadis et al. 2019) are summarized as: *If we know the characteristics of the parent solar events, could the probability of SEP occurrence be reliably inferred* and *how do the characteristics of SEP events (e.g. peak flux) map to the characteristics of their parent solar events* ? Both questions are being largely addressed through the implementation of databases and the establishment of empirical and/or semi-empirical statistical relations (see e.g. Gopalswamy et al. 2003, 2004; Garcia 2004a,b; Belov et al. 2005; Laurenza et al. 2009, 2018; Núñez 2011; Trottet et al. 2015; Dierckxsens et al. 2015; Anastasiadis et al. 2017; Papaioannou et al. 2018b; Kahler and Ling 2018; Richardson et al. 2018). The underlying idea is to identify a proper proxy (or combinations of proxies) that can be used for the unfolding of patterns and relationships among the parameters of SEP events and their parent solar events, using observational evidence at hand.In turn, such empirical relations point to the underlying physical processes of the SEP generation (Balch 2008). For example, Laurenza et al. (2009) provides prognosis of the SEP occurrence based on solar flare location, size (i.e. soft X-ray (SXR) fluence) and evidence of particle escape (i.e. radio fluence at ∼1 MHz); Papaioannou et al. (2018b) provides short-term forecasts of the SEP occurrence and the corresponding peak flux utilizing CME characteristics (e.g. width and speed); Kahler and Ling (2018) has used the SXR peak flux ratio, following Garcia (2004a), to establish the probability of SEP occurrence and Richardson et al. (2018) used the CME speed and the direction relative to the observer's site (i.e. magnetic connection) to predict the peak intensity of protons. In addition, the work from Posner (2007), has proven the concept of short-term forecasting of the appearance and intensity of solar ion events by means of relativistic electrons, making use of the higher speed of these electrons propagating from the Sun to 1 AU. Hence, it appears that such empirical or semi-empirical relations, can be used for the forecasting of solar radiation storms. Given the complexity and the incomplete knowledge of the underlying physical mechanisms at work, recent studies attempt to make use of higher order statistical relations (see e.g. Papaioannou et al. 2018a) and machine learning approaches (see Lavasa et al. 2021) in order to infer the probability of SEP occurrence. Such studies make use of the complete parameter space at hand, utilizing both solar flare and CME characteristics and have shown promising results. In this work, the Probabilistic Solar Particle Event Forecasting (PROSPER) Model, is presented. This model is integrated into the Advanced Solar Particle Event Casting System (ASPECS) operational tool that provides predictions of the probability of occurrence of SEP events, the expected peak flux and the resulting SEP time profile (http://phobos-srv.space.noa.gr/). PROSPER applies a novel, data-driven methodology to predict SEP events probabilistically for a set of integral energies, namely E>10, >30 and >100 MeV[1]. As it is going to be presented in detail, PROSPER takes advantage of the Bayes theorem (Bayes and Price 1763), taking into account all observational evidence at hand, without any bias. Bayesian approaches have been increasingly applied to the field of solar physics (Arregui

---

[1] PROSPER was further extended to E>60 MeV but the results are not shown through the ASPECS operational system

2022) and to solar flare forecasting, in particular (see e.g. Wheatland 2004, 2005, and reference there in). The model comes with three modes of operation depending on the available inputs : (a) CME characteristics (width, speed); (b) SF characteristics (longitude, magnitude) and (c) combinations of both CME and SF characteristics. The methodology is outlined and discussed in Section 2. The application of the methodology to all modes of operation is detailed in Section 3. Validation results for a set of case studies are presented in Section 4. Finally, the results are summarised and discussed in Section 5.

2. Analysis

2.1. Data

For the development of the PROSPER model, a catalogue of SEP events that includes 314 SEP events from 1984–2013 was utilized (e.g. Papaioannou et al. 2016). This SEP event catalogue is based on Geostationary Operational Enviromental Satellite (GOES)/Energetic Particle Sensor (EPS) data (for details on the data set see Sandberg et al. (2014)) and it includes key information on the proton peak flux and the total fluence of the identified SEP events in the three integral energy channels (namely, E > 10; >30; >100 MeV)[2]. It further includes the associated solar sources of the SEP events in terms of solar flares and CME characteristics. In particular, it includes solar SXR flux measurements, provided by GOES, recorded in the same period including 35306 C, M and X class flare events (ftp://ftp.ngdc.noaa.gov/STP/space-weather/solar-data/solar-features/solar-flares/x-rays/goes/). From the initial sample of SXR flares 14818 events for which the location was not available were excluded, leading to a sample of 20429 C, M and X class flares. At the same time, the CME identifications (i.e. plane-of-sky speeds and angular widths - AW) are made by the Large Angle and Spectroscopic Coronagraph (LASCO) (Brueckner et al. 1995) onboard the Solar and Heliospheric Observatory (SoHO) in the period from 1997–2013. The initial CME sample consists of 22143 events, which was reduced to 3693 events when association criteria between CMEs and solar flares were applied (see details in Papaioannou et al. (2016))[3]. The CME identifications, utilized in this study, are included in the Coordinated Data Analysis Web (CDAW) online CME Catalog (https://cdaw.gsfc.nasa.gov/CME_list/; Gopalswamy et al. 2009).

2.2. Mathematical formulation of PROSPER

In this section the general methodological approach for the implementation of PROSPER is presented. PROSPER provides both the Probability of SEP Occurrence P(SEP) (subsection 2.2.1) and the expected Peak Proton Flux (subsection 2.2.2). The first part, starts with the construction of Cumulative Distribution Functions (CDFs) by the data, continues with the implementation of Probability Density Functions (PDFs) and concludes with the application of the Bayes formula, from which the P(SEP) is directly provided. The second part, presents the establishment of the peak proton flux with the construction of the relevant CDFs.

2.2.1. Probability of SEP occurrence

This section presents the procedure for the establishment of P(SEP) within PROSPER mode of operation (a), i.e. utilising CME characteristics alone. PROSPER is applied to E>10, >30, (>60) and >100 MeV and for three modes of operation i.e., (a), (b) and (c) in Table 1 which

---

[2] PROSPER will further be extended to E>300 MeV as part of the ASPECS tool

[3] by construction, CMEs associated to far side solar flares are not included in our sample

provides an overview of the modes of operation for PROSPER, the inputs used, as well as, the bins applied based on the characteristics of their parent solar events. In particular, CMEs are grouped with respect to their angular width (see details in Papaioannou et al. 2018b), while solar flares by their longitude. Solar events originating in the west of the Sun, as observed by an observer on Earth, are more likely to produce large SEPs detectable at Earth (see e.g. Swalwell et al. 2017), especially those that lie within W20^o-W80^o (see Cliver et al. 2012, and references therein). From our sample roughly half of the SEP events (~46%) are associated with a solar flare originating west of W20^o, hence our choice of positional requirement in the algorithms. Such events are termed as: "well connected" whereas those east of W20^o as "poorly connected".

Table 1. Details on PROSPER's modes of operation, inputs and initial binning

Mode of operation	**Inputs**	**Bins**	**Resulting function per bin**
(a) CME	speed & AW	continuous speed AW = 360^o (Halo - H) 120^o ≤ AW < 360^o (Partial Halo - PH) AW < 120^o (Non-Halo - NH)	$f(V_{CME})$ [3]
(b) flare	Soft X-ray flare magnitude & longitude	continuous magnitude lon. ≥20^o (well connected - WC) lon. <20^o (poorly connected - PC)	$f(F_{SXR})$ [2]
(c) flare & CME	all of the above	[H or PH or NH] & WC [H or PH or NH] & PC	$f(V_{CME}, F_{SXR})$ [6]
Acronyms & bins			
AW = angular width	*H = Halo*	*PH= Partial Halo*	*NH = Non-Halo*
lon.= longitude	*WC = well connected*	*PC = poorly connected*	
with respect to the	*(lon. ≥20^o)*	*(lon. <20^o)*	
Sun-Earth line			

Column 1 denotes the mode of operation, column 2 declares the inputs used per mode, column 3 details the initial bins applied to the data based on the inputs and column 4 presents the resulting number of functions per bin. The numbers in the brackets of column 4 denote the number of resulting functions per mode and integral energy. These functions are the P(SEP) ones which are going to be detailed in the following steps for each of the selections. Hence for each integral energy (e.g. E>10 MeV) there are 3 $f(V_{CME})$, 2 $f(F_{SXR})$ and 6 $f(V_{CME}, F_{SXR})$ functions in accordance to column 3. PROSPER is applied independently to each of the integral energies considered.

Implementation of Cumulative Distribution Functions First, we bin the data in accordance to Table 1. The empirical cumulative distribution function (CDF) for each sub-sample was constructed by sorting the events in the ascending order of the continuous parameter (CME speed and/or solar flare magnitude; see also Table 1), and using the standard fractional ordinates as probabilities. As parent solar events (i.e. parameterized by flares and/or CMEs) can be either associated with SEP events or not, it is possible to construct two CDFs: one for each sample, i.e. one for all parent solar events (i.e. parameter, Λ) of the specific binning: $P(\Lambda < \lambda)$[4] and one for all parent solar events that were associated with an SEP event in that

---

[4] λ refers to the continuous parameter of Table 1 per case

bin, $P(\Lambda < \lambda \,|\, \mathrm{SEP})$.

Next, the resulting (filtered) data were fit with a log-normal distribution (see e.g. Raukunen et al. 2018), which was found to be the most robust of the flux distributions applied to SEP fluxes for specification models, in this work. This is given by the following equation:

$$F(\lambda) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{\log_{10}(\lambda) - \mu}{\sigma\sqrt{2}}\right)\right], \tag{1}$$

where $F(\lambda)$ is the probability of Λ being lower than λ, $\mathrm{erf}(x)$ is the error function, and μ and σ are the mean and standard deviation of $\log_{10}(\lambda)$, respectively.

Implementation of the Probability Density Functions These CDFs were consequently used in order to construct probabilities for which the CME speed/flare magnitude, λ lies within a certain range defined by an upper and a lower limit, i.e. $\in [\lambda_1, \lambda_2]$. The resulting independent probability of *observing* α, i.e. the probability of Λ to fall within a certain range of values is given by the subtraction of two CDFs constructed by all data points in a sample one for the lower limit, i.e. $\Lambda < \lambda_1$ and one for the upper one, i.e. $\Lambda < \lambda_2$ leading to $P\,(\Lambda < \lambda_2)$ and $P\,(\Lambda < \lambda_1)$. The subtraction of which results in:

$$P(\alpha) = P\,(\Lambda \in [\lambda_1, \lambda_2]) = P\,(\Lambda < \lambda_2) - P\,(\Lambda < \lambda_1)\,. \tag{2}$$

The independent probability of *observing* β, i.e. the probability of SEP events in a sample, is based on the number of SEP associated events (N_{SEP}) and the total number of events (N_{total}) in the sample and is defined as:

$$P(\beta) = P(SEP) = N_{SEP}/N_{total}. \tag{3}$$

The conditional probability of *observing* α under the condition of *observing* β, $P(\alpha\,|\,\beta)$, i.e. the probability of the Λ to fall within a certain range of λ, under the condition that these values refer to SEP events. This is also given as the subtraction of two CDFs. In particular, the first CDF refers to the lower limit, i.e. $\Lambda < \lambda_1 \,|\, \mathrm{SEP}$ and the second to the upper one, i.e. $\Lambda < \lambda_2 \,|\, \mathrm{SEP}$ leading to $P\,(\Lambda < \lambda_1 \,|\, \mathrm{SEP})$ and $P\,(\Lambda < \lambda_2 \,|\, \mathrm{SEP})$. As a result the conditional probability is defined as:

$$P(\alpha\,|\,\beta) = P\,(\Lambda \in [\lambda_1, \lambda_2] \,|\, \mathrm{SEP}) = P\,(\Lambda < \lambda_2 \,|\, \mathrm{SEP}) - P\,(\Lambda < \lambda_1 \,|\, \mathrm{SEP})\,. \tag{4}$$

Using the Bayes theorem (Bayes and Price 1763; Joyce 1999) one may define the inverse conditional probability:

$$P(\beta\,|\,\alpha) = P\,(\mathrm{SEP} \,|\, \Lambda, \in [\lambda_1, \lambda_2]) \tag{5}$$

which is a PDF and provides the probability of SEP occurrence when $\Lambda \in [\lambda_1, \lambda_2]$. The Bayes formula dictates that:

$$P(\beta\,|\,\alpha) = \frac{P(\alpha\,|\,\beta)P(\beta)}{P(\alpha)} \tag{6}$$

Merging Equations 2–6, the following formula is obtained:

$$P\,(\mathrm{SEP} \,|\, \Lambda \in [\lambda_1, \lambda_2]) = \frac{[P\,(\Lambda < \lambda_2 \,|\, \mathrm{SEP}) - P\,(\Lambda < \lambda_1 \,|\, \mathrm{SEP})]\, P(\mathrm{SEP})}{P\,(\Lambda < \lambda_2) - P\,(\Lambda < \lambda_1)}. \tag{7}$$

Hence, Equation 7 shows the probability to have an SEP event under the condition that $\Lambda \in [\lambda_1, \lambda_2]$. For example, in the case of CMEs, substituting Λ with V_{CME}, λ_1 with V_1 and λ_2 with V_2, Equation 7 becomes:

$$P\,(\mathrm{SEP}\,|\,V_{\mathrm{CME}} \in [V_1, V_2]) = \frac{[P\,(V_{\mathrm{CME}} < V_2\,|\,\mathrm{SEP}) - P\,(V_{\mathrm{CME}} < V_1\,|\,\mathrm{SEP})]\,P(\mathrm{SEP})}{P\,(V_{\mathrm{CME}} < V_2) - P\,(V_{\mathrm{CME}} < V_1)}. \tag{8}$$

Moreover, setting $V_1 = V$ and $V_2 = V + \mathrm{dV}$ and dividing both nominator and denominator by dV while letting $\mathrm{d}V \to 0$ results to:

$$P\,(\mathrm{SEP}\,|\,V_{\mathrm{CME}} = V) = \frac{P'\,(V_{\mathrm{CME}} < V\,|\,\mathrm{SEP})}{P'\,(V_{\mathrm{CME}} < V)} P(\mathrm{SEP}) = \frac{f_{\mathrm{SEP}}(V)}{f(V)} P(\mathrm{SEP}). \tag{9}$$

Both Equations 8 and 9 provide the probability of occurrence of an SEP event given a CME of speed, V.

In Equation 9, $f(V) = P'(V_{\mathrm{CME}} < V)$ is the probability density function (PDF) of the CME speeds for the binned sub-sample of choice (see Table 1) and $f_{\mathrm{SEP}}(V)$ is the PDF for those CMEs of the bin that are associated with SEP events. In order to construct the PDFs (i.e. $f(V)$ & $f_{\mathrm{SEP}}(V)$) from the relevant CDFs, there are two ways: (a) apply numerical differentiation (labeled as *Method 1*) and (b) fit the data points with an empirical CDF and directly take the derivatives of the relevant distributions, as a function of the CME speed (V_{CME}, $[km/s]$) (labeled as *Method 2*). As it can be seen in Figure 1, for the case of Halo CMEs and E>10 MeV SEP events, the output is equivalent. Therefore, from the empirical CDFs, which are based on data, and the applied log-normal fits, one can directly obtain the relevant PDFs and from the application of the Bayes formula (Equation 9) obtain an analytical expression that answers to the question: *What is the probability that an SEP event [P(SEP)] will occur if one knows the CME speed (V_{CME}) (and width) of the driving CME* ?

Equation 9 also stands for the case of solar flares. In this case, the application of the Bayes formula based on solar flare longitude and flux (F) results in:

$$P\,(\mathrm{SEP}\,|\,F_{\mathrm{SXR}} = F) = \frac{P'\,(F_{\mathrm{SXR}} < F\,|\,\mathrm{SEP})}{P'\,(F_{\mathrm{SXR}} < F)} P(\mathrm{SEP}) = \frac{f_{\mathrm{SEP}}(F)}{f(F)} P(\mathrm{SEP}). \tag{10}$$

At this point, it is noted that Equation 6 acquires very large values and tends to infinity as $P(\alpha) \to 0$. For this reason, using some elementary mathematical properties of sets and probabilities, Equation 6 can be written in a form which is more stable and appropriate for numerical calculations, as:

$$P(\beta\,|\,\alpha) = \frac{P(\alpha\,|\,\beta)P(\beta)}{P(\alpha|\beta)P(\beta) + P(\alpha\,|\,\beta^c)P(\beta^c)}, \tag{11}$$

with β^c being the complementary event of β (Maritz and Lwin 2018). Although the denominators of Eq. 6 and Eq. 11 are equivalent, the former $\to$ 0 faster than the latter one. As a result Equation 11 has an optimal numerical behaviour and allows one to obtain outputs at a wider range compared to Equation 6 before the denominator of Equation 6 $\to 0$. Therefore, the formulas used in PROSPER calculations are the more complex expressions of Equations 7–10 based on substitution of Equations 2–5 in Equation 11.

Additionally, in the case of a solar flare and an associated CME the Bayes formula used in PROSPER gets the general form:

$$P(\beta_{i,j}\,|\,V,F) = \frac{f_{\beta_i}(V) f_{\beta_j}(F) P(\beta_{i,j})}{[f_{\beta_i}(V) f_{\beta_j}(F) P(\beta_{i,j}) + f_{\beta_i^c}(V) f_{\beta_j^c}(F) P(\beta_{i,j}^c)]}. \tag{12}$$

where i refers to CME AW, i.e. Halo, Partial Halo, Non Halo and j to the solar flare longitude ("well" and/or "poorly" connected); see details per selection in Table 1). $P(\beta_{i,j}$ and $P(\beta_{i,j}^c)$ are calculated directly by the measurements.

2.2.2. Peak flux estimation

PROSPER provides a peak flux estimation for a given (user-defined) confidence based on the P(SEP) described here above. First the data are filtered, based on the P(SEP) that was derived in subsection 2.2.1. In particular, once P(SEP) is obtained, peak flux data are filtered based on the integral energy that the SEP event is expected to reach. That said, if i.e. an SEP event is expected to reach E>100 MeV only the events from the list with significant peak fluxes reaching E>100 MeV are employed across all integral energies. The same stands for filtering the data samples based on other integral energies based on the obtained P(SEP) (i.e. if an event is expected to reach E>30 MeV). At any case, the presence of a high energy population is an important filter for lower energy fluxes since it ensures spectral coherence. Next, additional binning is applied using the available characteristics of the parent solar events (i.e. SF magnitude & location, CME speed & width). Once the data samples are retrieved we utilize the tabulated peak proton flux (PPF) for the SEP events in the list employed in Papaioannou et al. (2016) in order to derive the relevant distributions.

The filtered peak fluxes were fit with the exponential cut-off power law that was found to be the most robust of the flux distributions applied for specification models (consistent with findings from both Jiggens et al. (2018) and Raukunen et al. (2018)). This is given by the following equation:

$$P\left(\mathrm{F_p} \geq \mathrm{PPF_0}\,|\,\mathrm{SEP}\right) = 1 - \frac{PPF_0^{-\gamma}\exp\left(\frac{x_{low}}{x_{lim}}\right)}{x_{low}^{-\gamma}\exp\left(\frac{PPF_0}{x_{lim}}\right)}, \tag{13}$$

with γ being the power-law exponent, x_{low} a parameter related to the lower limit of the F_P distribution, and x_{lim} a parameter related to the upper limit of the F_P distribution.

Equation 13 answers the question: *What is the expected probability that the peak flux will exceed a certain threshold, (PPF_0) for a given integral energy, provided that it is certain an SEP event will occur*? From Equation 13, one can find the expected peak flux that corresponds to a specific probability threshold, solving Equation 13 for PPF_0 for a given value of $P\,(\mathrm{F_P} \geq \mathrm{PPF_0}\,|\,\mathrm{SEP}) = P_{thres}$. Following this, the $P(\mathrm{SEP}\,|\,\Lambda = \lambda_0)$ is used to modulate the prediction of the expected peak flux. In essence, a weighted average is employed, driven by the obtained $P(\mathrm{SEP}\,|\,\Lambda = \lambda_0)$. There are two components in this averaging: the first one is the PPF_0 for a specific user defined threshold multiplied by $P(\mathrm{SEP}\,|\,\Lambda = \lambda_0)$ and the second one is the background value of the specific integral energy multiplied by the 1 - $P(\mathrm{SEP}\,|\,\Lambda = \lambda_0)$. The averaging leads to PPF_{thres} and the obtained $P(\mathrm{SEP}\,|\,\Lambda = \lambda_0)$ is used as a weight. To this end PROSPER provides $F_{P\mathrm{thres}}$ obtained from Equation 14:

$$F_{P\mathrm{thres}} = PPF_0 \cdot P(\mathrm{SEP}\,|\,\Lambda = \lambda_0) + background\,flux \cdot [1 - \mathrm{P}(\mathrm{SEP}\,|\,\Lambda = \lambda_0)] \tag{14}$$

In the limiting cases of the binary extremes of $P(\mathrm{SEP}|\Lambda = \lambda_0)$, the following expression is obtained:

$$F_{P\mathrm{thres}} = \begin{cases} PPF_0 & , P(\mathrm{SEP}\,|\,\Lambda = \lambda_0) = 1 \\ background\ flux & , P(\mathrm{SEP}\,|\,\Lambda = \lambda_0) = 0 \end{cases} \tag{15}$$

The *background flux* was calculated on the basis of the recorded intensity of each integral energy of interest (i.e. E>10-; >30-; and >100 MeV)[5] and corresponds to the mean value obtained in September 2009 (e.g. during the solar minimum), which is: 0.23, 0.122, 0.089 and 0.050 pfu, respectively (see Figure 2). Evidently, depending on the CL, $thres$, of choice (by default in the ASPECS tool, PROSPER provides $thres$=50%, and 90%, which were defined by the user community), and thus the model provides PPF_{50} and PPF_{90}.

3. Results

3.1. Coronal Mass Ejections

In the case of CMEs, PROSPER's concept was applied across three CME AW bins (see Table 1), for all integral energies of our database (E > 10; >30; (>60) and >100 MeV) and for both the distribution of all CMEs speeds (V_{CME}) in each sub-sample, as well as, the distribution of the speeds of CMEs related to SEPs. Figure 3 depicts the obtained empirical CDFs from the data (i.e. $P(V_{\mathrm{CME}} < V)$). From top to bottom, on the left hand side, the column refers to *all* Halo, *all* Partial Halo and *all* Non Halo CMEs of our sample. Additionally, similar plots for $P(V_{\mathrm{CME}} < V\,|\,\mathrm{SEP})$ for an integral energy of E>10 MeV is provided on the right hand side of the plot, for each case, following the same labeling. Each sub-plot of Figure 3 presents the data points (in black color) for each distribution. Additionally, the log-normal fit (from Equation 1) to these points is presented as a red-line, with the fraction of the data (in %) represented by the fit. Finally, the mean absolute error (MAE)[6] was calculated at each case and is imprinted on every plot. Although the related results are not shown in Figure 3, it is noteworthy that when shifting to higher energies, i.e. from E>10 MeV to E>100 MeV, a reduction in the number of SEP events that reach higher energies is apparent. In turn, fits for these energies rely on fewer points.

A combined representation of these CDFs, but for all integral energies of interest (e.g. E>10; >30; (>60) and >100 MeV) is given in Figure 4[7]. Each panel provides the fitted CDFs for the $P(V_{\mathrm{CME}} < V)$ (black line) and $P(V_{\mathrm{CME}} < V\,|\,\mathrm{SEP})$ distributions - the latter for each integral energy, color coded as: E>10 MeV - red line; E>30 MeV - blue line; E>60 MeV - green line and E>100 MeV - orange line. From top to bottom, panels refer to Halo, Partial Halo and Non Halo CMEs, respectively. The PDFs from the derivatives of the distributions $f(V)$ and $f_{\mathrm{SEP}}(V)$ are consequently presented in Figure 5, following the same reasoning (and labeling) as in Figure 4. It should be noted that the X-axis (i.e. V_{CME}) in all panels extend to a simulated value of 10000 km/s. However, the actual data point to an upper limit V_{CME} of ~3000 km/s (see Figure 3). That said, the derived Bayes $P(\mathrm{SEP})$ is obtained up to that V_{CME}. However, we present the whole evolution of the fit and we overlay a gray border hatched rectangle area after that limit.

The final step in this formulation is to employ the Bayes formula (Equation 9) and retrieve the fits that reply to the question: *what is the probability of SEP occurrence from a CME with known*

---

[5] data taken from https://satdat.ngdc.noaa.gov/sem/goes/data/avg/

[6] MAE=$\sum_{i=1}^{n} |y_i - x_i|/n$, with n being the number of pairs, y and x the fitted and the actual value, respectively

[7] Although E>60 MeV is not implemented in the ASPECS tool, we present the obtained fits for $P(\mathrm{SEP})$ in this and consequent Figures for consistency

width (AW) and speed (V_{CME})? These fits are presented in Figure 6. Each panel corresponds to one integral energy of interest, and each line within every panel refers to Halo (blue line); Partial Halo (red line) and Non Halo (green line) CMEs. Hence, for each integral energy a total of 3 $f(V_{CME})$ functions are obtained (see Table 1). The abscissa of Figure 6 gives the V_{CME} and the ordinate provides the probability of SEP occurrence. Hence, for a CME with known AW, one may select the proper fit and from the known V_{CME} directly obtain the expected probability, P(SEP). It can be noted that up to the limit of V_{CME} ~3000 km/s for any given CME speed, the largest P(SEP) is obtained for Halo (blue line), followed by Partial Halo (red line) and Non Halo (green line). Moreover, both the red and the green curve would have a similar behaviour as the blue curve, if the X-axis would have been expanded to >10000 km/s. In addition, when crossing the $V_{CME} \sim 3000$ km/s limit it can be seen that Partial Halo (red line) and Non Halo (green line) CMEs lead to a higher P(SEP) compared to Halo CMEs (blue line). This is the output of the application of Equation 9 and it points to the possibility to get a higher P(SEP) for Partial and Non Halo CMEs provided that these CMEs are faster than any CME ever observed. Nonetheless, the application of an upper limit for V_{CME} that is inferred by the actual sample used, results into higher probabilities for Halo CMEs up to that limit. That said, all observed CMEs with $V_{CME} > 3000$ km/s are treated as having a $V_{CME} = 3000$ km/s.

As noted in subsection 2.2.2, based on user consultations, a focused probability range from 50-90% has been selected, while taking into account, as a filter, the probability to derive an SEP event at a respective energy. If an SEP event is expected to reach e.g. E>100 MeV the data for all energies is binned based on this energy filtering and consequently the probabilities are derived based on CME speed and width. The obtained results for the case of CME inputs are depicted in Figure 7. There are three different filters based on the predicted integral energy that the particles will reach and different bins based on the CME characteristics. In particular, panels (a) and (b) of Fig. 7 refer to SEP events that are expected to reach an integral energy of E>10 MeV. The red (blue) lines depict the fast (slow) bins and each of the panels refer to Halo (panel (a)) and Not Halo (panel (b)) CMEs. In addition, panels (c)-(f) of Fig. 7 refer to SEP events that are expected to reach an integral energy of E>30 MeV. The blue (orange) lines in each of the panels present the E>10 (E>30) MeV fits, whereas each of the panels stands for a different selection of AW and CME speed – panel (c): Halo and slow; panel (d) Halo and fast; panel (e) not Halo and slow and panel (f) not Halo and fast CMEs. Finally, panel (g) of Fig. 7 represents the fits for a filtering at E>100 MeV and each line corresponds to an integral energy of interest color coded as: red (E>10 MeV); green (E>30 MeV) and blue (E>100 MeV). All bins and selections are detailed in Table 2.

3.2. Solar Flares

A similar probabilistic approach was further applied for solar flare inputs replacing V_{CME} with SXR peak flux and imposing two longitudinal source location ranges: well connected ($longitude \geq 20^o$), poorly connected ($longitude < 20^o$) in place of CME angular width. Applying the same methodology we obtained the following fits presented in Figure 8. These fits are similar to Figure 6 but for the SF case. Each panel corresponds to one integral energy of interest, and each line within every panel refers to well connected (red line) and poorly connected (blue line) SFs. As a result, for each integral energy a total of 2 $f(F_{SXR})$ functions are obtained (see Table 1). The X-axis of sub-panels in Figure 8 gives the F_{SXR} and the Y-axis provides the probability of SEP occurrence. Hence, for a solar flare with known longitude, one may select the proper fit and from the magnitude of the solar flare directly obtain the expected probability, P(SEP). In this case, data point to an upper limit F_{SXR} of X28 and thus the derived Bayes

Table 2. Details on PROSPER's peak flux fits for the CME bins, derived by Equation 13. The filtering on the integral energy is first applied and then there are two bins. One for the width of the CME (i.e. AW = 360^oHalo or AW <360^oNot Halo) and one for the CME speed (i.e. V_{CME} being < or ≥ 1250 km/s). These fits are presented in Figure 7.

if an event reaches E>10 MeV			
AW = 360^o (Halo - H)		AW < 360^o (Not Halo)	
V_{CME} < 1250 km/s	$V_{CME} \geq$ 1250 km/s	V_{CME} < 1250 km/s	$V_{CME} \geq$ 1250 km/s
$\gamma = 0.36$ x_{low} = 3.68E+00 x_{lim} = 5.37E+01	$\gamma = 0.26$ x_{low} = 6.38E+00 x_{lim} = 1.18E+04	$\gamma = 0.53$ x_{low} = 2.50E+00 x_{lim} = 1.49E+03	$\gamma = 0.32$ x_{low} = 4.00E+00 x_{lim} = 8.95E+03
if an event reaches E>30 MeV			
AW = 360^o (Halo - H)		AW = 360^o (Halo - H)	
V_{CME} < 1250 km/s	$V_{CME} \geq$ 1250 km/s	V_{CME} < 1250 km/s	$V_{CME} \geq$ 1250 km/s
E>10 MeV		**E>30 MeV**	
$\gamma = 0.30$ x_{low} = 3.68E+00 x_{lim} = 5.27E+01	$\gamma = 0.26$ x_{low} = 6.38E+00 x_{lim} = 1.19E+04	$\gamma = 1.09$ x_{low} = 1.23E+00 x_{lim} = 1.54E+02	$\gamma = 0.30$ x_{low} = 1.24E+00 x_{lim} = 2.27E+03
AW < 360^o (Not Halo)		AW < 360^o (Not Halo)	
V_{CME} < 1250 km/s	$V_{CME} \geq$ 1250 km/s	V_{CME} < 1250 km/s	$V_{CME} \geq$ 1250 km/s
$\gamma = 0.45$ x_{low} = 3.94E+00 x_{lim} = 9.96E+02	$\gamma = 0.30$ x_{low} = 5.72E+00 x_{lim} = 8.55E+03	$\gamma = 0.58$ x_{low} = 9.06E-01 x_{lim} = 3.23E+03	$\gamma = 0.32$ x_{low} = 6.71E-01 x_{lim} = 3.51E+03
if an event reaches E>100 MeV			
No bin			
All data points			
E>10 MeV	**E>30 MeV**	**E>100 MeV**	
$\gamma = 0.26$ x_{low} = 9.50E+00 x_{lim} = 1.19E+04	$\gamma = 0.28$ x_{low} = 2.30E+00 x_{lim} = 2.04E+03	$\gamma = 0.41$ x_{low} = 3.02E-01 x_{lim} = 2.58E+02	

P(SEP) is obtained up to that limit. Again, the total evolution of the simulated fit is presented and we overlay a gray hatched area after that limit.

In the case of the peak flux, the focused range of 50 & 90% is maintained across all modes of operation. Figure 9 provides the outputs per filtered integral energy and bins based on solar flare characteristics. Specifically, Fig. 9(a) refers to a filtering of E>10 MeV with each of the presented fits color coded as: red (F_{SXR} < M3.0); green (M3.0 ≤ F_{SXR} < X1.0) and blue (F_{SXR} ≥ X1.0). Moreover, Fig. 9, panels (b), (d) and (f) refer to a filtering at E>30 MeV. Red (green) fits present the integral energy of E>10 (E>30) MeV, while each of the three panels point to individual solar flare bins. Finally, panels (c), (e) and (g) of Fig. 9 are obtained with a filtering at E>100 MeV. In this case, each of the lines represent an integral energy: red (E>10 MeV); green (E>30 MeV) and blue (E>100 MeV). Again, each of these three panels have to do with a selection based on solar flare magnitude. All of the fits depicted in Figure 9 are detailed in Table 3.

Table 3. Details on PROSPER's peak flux fits for the SF bins, derived by Equation 13. The filtering on the integral energy is first applied and then there are three bins based on the magnitude of the solar flare. These fits are presented in Figure 9.

if an event reaches E>10 MeV		
Flare flux < M3.0	M3.0 ≤ Flare flux < X1.0	Flare flux ≥ X1.0
$\gamma = 0.56$ $x_{low} = 2.63E+00$ $x_{lim} = 1.46E+03$	$\gamma = 0.40$ $x_{low} = 2.97E+00$ $x_{lim} = 1.17E+04$	$\gamma = 0.27$ $x_{low} = 7.69E+00$ $x_{lim} = 1.39E+04$
if an event reaches E>30 MeV		
Flare flux < M3.0	M3.0 ≤ Flare flux < X1.0	Flare flux ≥ X1.0
E>10 MeV		
$\gamma = 0.52$ $x_{low} = 5.66E+00$ $x_{lim} = 1.36E+03$	$\gamma = 0.36$ $x_{low} = 4.81E+00$ $x_{lim} = 1.01E+04$	$\gamma = 0.28$ $x_{low} = 1.29E+01$ $x_{lim} = 1.43E+04$
E>30 MeV		
$\gamma = 0.76$ $x_{low} = 1.13E+00$ $x_{lim} = 3.27E+02$	$\gamma = 0.53$ $x_{low} = 9.29E-01$ $x_{lim} = 5.72E+03$	$\gamma = 0.28$ $x_{low} = 1.59E+00$ $x_{lim} = 2.29E+03$
if an event reaches E>100 MeV		
Flare flux < M3.0	M3.0 ≤ Flare flux < X1.0	Flare flux ≥ X1.0
E>10 MeV		
$\gamma = 0.47$ $x_{low} = 7.74E+00$ $x_{lim} = 2.05E+04$	$\gamma = 0.27$ $x_{low} = 1.06E+01$ $x_{lim} = 9.42E+03$	$\gamma = 0.21$ $x_{low} = 2.66E+01$ $x_{lim} = 1.60E+04$
E>30 MeV		
$\gamma = 0.57$ $x_{low} = 1.91E+00$ $x_{lim} = 5.56E+02$	$\gamma = 0.36$ $x_{low} = 2.74E+00$ $x_{lim} = 2.52E+03$	$\gamma = 0.20$ $x_{low} = 5.33E+00$ $x_{lim} = 2.16E+03$
E>100 MeV		
$\gamma = 0.72$ $x_{low} = 3.67E-01$ $x_{lim} = 5.03E+00$	$\gamma = 0.53$ $x_{low} = 3.33E-01$ $x_{lim} = 1.88E+02$	$\gamma = 0.23$ $x_{low} = 7.12E-01$ $x_{lim} = 1.65E+02$

3.3. Solar Flares & CMEs

For the case of a combination of a solar flare (with known magnitude and longitude) and a CME (with known AW and speed) being identified Equation 12 was employed but enforcing more conditions, depending on the obtained characteristics (e.g. flare longitudinal bin and magnitude, CME width and velocity). In this case, all combinations were considered (e.g. Halo, Partial halo and Non halo, together with well and poorly connected) for all respective energies. Figure 10 depicts derived probability of SEP occurrence using both flare and CME inputs for the case of Halo and well connected events for an integral energy of E>10 MeV. Each line corresponds to a specific solar flare class. In particular, C1.0 (black), M1.0 (blue), X1.0 (red) and X10.0 (magenta) lines are shown. Based in the above, the X-axis of Figure 10 is extended up to V_{CME} = 3000 km/s. From this plot one can derive the probability to get an SEP event, in case the longitude of the parent solar flare falls within the bin (lon $\geq 20^{o}$ - well connected) and the corresponding CME is a Halo one (AW = 360^{o}). Then from the specific magnitude of the flare, the proper curve (fit) is selected and the speed of the CME (X-axis on the plot) leads to the probability of SEP occurrence (Y-axis on the plot). Similar fits have been constructed for all combinations of: (lon. x AW x integral proton energy). As a result, Equation 12 was applied in total in 18 cases (not shown).

Figure 11 shows the obtained fits per filtered energy and bins based on all obtained inputs. Similar to the previous cases the fits are presented in detail in Table 4. As a rule, the dashed lines in the case of E>10 MeV (Fig. 11)(a) and E>100 MeV (Fig. 11)(b), (d), (f) present the bins with the largest CME speed per case for a given solar flare flux bin, while the continuous lines depict the cases with the slower CMEs for the same solar flare bins (see details of the bins in each panel of Fig. 11 and Table 4). Panel (a) of Figure 11 shows E>10 MeV fits for six different selections of solar flares & CMEs (see detail in Table 4). Each pair of lines (i.e. solid and dashed) colored with the same color refer to a similar solar flare bin. In particular, these are: F_{SXR} < M3.0 (red); M3.0 ≤ F_{SXR} <X1.0 (blue) and F_{SXR} ≥ X1.0 (magenta) (see panel (a)). Moreover, panels (b), (d) and (f) of Fig. 11 represent the case of applying a filtering at E>100 MeV. In all these panels each line refers to an integral energy color coded as: red (E>10 MeV); blue (E>30 MeV) and magenta (E>100 MeV). However, there was no event associated with a solar flare of magnitude <X3.0 for any CME speed, reaching an integral energy of E>100 MeV. Therefore, the relevant panel (f) in Figure 11 shows only the fits for events associated with ≥X3.0 for each of the three CME bin (i.e. only dashed lines) (see Table 4). In addition, no reliable fit for SEP events associated with solar flares of magnitude <M6.5 and CME speed V_{CME} < 1350 km/s, at E>10 MeV, when a filter on reaching an integral energy of E>100 MeV is applied, was found (see Fig. 11(b) and Table 4). Hence, there is no solid red line displayed in this panel. Finally, panels (c)-(e), (g)-(i) and (h)-(j) present the cases with a filtering at E>30 MeV. Each panel provides two fits color coded per integral energy as: E>10 MeV (red) and E>30 MeV (blue). Each of the mentioned pairs of panels refer to the same solar flare bin (i.e. F_{SXR} < M3.0; M3.0 ≤ F_{SXR} <X1.0 and F_{SXR} ≥X1.0) while differ in the CME speed as detailed in Table 4. As it can be seen,in two cases that an event reaches E>30 MeV these seems to be a cross-over between the obtained fit at E>10 and E>30 MeV (i.e. panels (c) and (g)). However, these crossings are below a threshold of 40%. Evidently, if a CL <50% is chosen then in this particular case (i.e. E>30 MeV) a spectral incoherence arises. This means that in a future work an update of the selection in PROSPER for the solar flare & CME case that will lead to spectral coherence at all cases for lower CL should be pursued.

Table 4. Details on PROSPER's peak flux fits for the combined solar flares & CME bins, derived by Equation 13. The filtering on the integral energy is first applied and then bins on the solar flare magnitude and CME speed are applied. These fits are presented in Figure 11.

if an event reaches E>10 MeV		
Flare flux < M3.0	M3.0 < Flare flux ≤ X1.0	Flare flux > X1.0
V_{CME} < 1250 km/s	V_{CME} < 1400 km/s	V_{CME} < 1600 km/s
γ = 0.45 x_{low} = 2.98E+00 x_{lim} = 3.85E+01	γ = 0.69 x_{low} = 3.30E+00 x_{lim} = 6.49E+02	γ = 0.22 x_{low} = 4.63E+00 x_{lim} = 4.27E+02
V_{CME} ≥ 1250 km/s	V_{CME} ≥ 1400 km/s	V_{CME} ≥ 1600 km/s
γ = 0.35 x_{low} = 4.23E+00 x_{lim} = 1.58E+03	γ = 0.23 x_{low} = 4.87E+00 x_{lim} = 8.19E+03	γ = 0.21 x_{low} = 2.82E+01 x_{lim} = 1.35E+04

if an event reaches E>30 MeV		
Flare flux < M3.0	M3.0 < Flare flux ≤ X1.0	Flare flux > X1.0
E>10 MeV		
V_{CME} < 1250 km/s	V_{CME} < 1350 km/s	V_{CME} < 1650 km/s
γ = 0.17 x_{low} = 6.14E+00 x_{lim} = 2.98E+01	γ = 0.64 x_{low} = 4.05E+00 x_{lim} = 7.45E+02	γ = 0.26 x_{low} = 7.80E+00 x_{lim} = 4.59E+02
V_{CME} ≥ 1250 km/s	V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1650 km/s
γ = 0.32 x_{low} = 6.06E+00 x_{lim} = 1.50E+03	γ = 0.22 x_{low} = 5.27E+00 x_{lim} = 8.04E+03	γ = 0.21 x_{low} = 2.82E+01 x_{lim} = 1.35E+04
E>30 MeV		
V_{CME} < 1250 km/s	V_{CME} < 1350 km/s	V_{CME} < 1650 km/s
γ = 0.29 x_{low} = 1.26E+01 x_{lim} = 1.48E+00	γ = 0.05 x_{low} = 5.86E+00 x_{lim} = 4.77E+00	γ = 0.40 x_{low} = 1.33E+00 x_{lim} = 2.82E+02
V_{CME} ≥ 1250 km/s	V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1650 km/s
γ = 0.63 x_{low} = 1.19E+00 x_{lim} = 1.23E+03	γ = 0.33 x_{low} = 1.08E+00 x_{lim} = 2.66E+03	γ = 0.18 x_{low} = 2.14E+00 x_{lim} = 2.00E+03

if an event reaches E>100 MeV		
Flare flux < M6.0	M6.0 < Flare flux ≤ X3.0	Flare flux > X3.0
E>10 MeV		
V_{CME} < 1350 km/s	V_{CME} < 1350 km/s	V_{CME} < 1600 km/s
γ= x_{low} = x_{lim} =	γ = 0.48 x_{low} = 8.70E+00 x_{lim} = 8.49E+02	γ = x_{low} = x_{lim} =
V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1600 km/s
γ=0.36 x_{low} = 2.47E+01 x_{lim} = 2.38E+04	γ = 0.18 x_{low} = 1.85E+01 x_{lim} = 1.17E+04	γ =0.18 x_{low} = 6.16E+01 x_{lim} = 1.35E+04
E>30 MeV		
V_{CME} < 1350 km/s	V_{CME} < 1350 km/s	V_{CME} < 1600 km/s
γ=0.41 x_{low} = 2.96E+00 x_{lim} = 9.71E+00	γ = 0.47 x_{low} = 2.38E+00 x_{lim} = 1.01E+02	γ = x_{low} = x_{lim} =
V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1600 km/s
γ=0.24 x_{low} = 2.23E+00 x_{lim} = 4.47E+02	γ = 0.26 x_{low} = 4.97E+00 x_{lim} = 3.14E+03	γ = 0.16 x_{low} = 1.64E+01 x_{lim} = 1.76E+03
E>100 MeV		
V_{CME} < 1350 km/s	V_{CME} < 1350 km/s	V_{CME} < 1600 km/s
γ=1.88 x_{low} = 4.04E-01 x_{lim} = 4.66E+02	γ = 0.77 x_{low} = 4.60E-01 x_{lim} = 4.67E+00	γ = x_{low} = x_{lim} =
V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1350 km/s	V_{CME} ≥ 1600 km/s
γ=0.34 x_{low} = 3.75E-01 x_{lim} = 6.82E+00	γ = 0.32, x_{low} = 2.33E-01 x_{lim} = 1.11E+02	γ = 0.26 x_{low} =1.18E+00 x_{lim} = 2.92E+02

4. Validation

Table 5 provides all inputs of the NASA CCMC SEP scoreboard campaign events[8] that have been investigated in this part of the validation of PROSPER, presented in this work. The purpose of this challenge is to facilitate collaboration of as many as possible SEP modelers and to provide a standardized validation procedure, with specific inputs being used by all parties. Such a challenge is based on curated events and greatly facilitates the comparison of outputs from different prediction methods. Nonetheless, larger samples of events are needed for a more complete evaluation of these methods, including PROSPER. The inputs (i.e. solar flare and/or CME characteristics) have been utilized in order to achieve the PROSPER's P(SEP) per mode of operation. PROSPER peak flux outputs are compared to the derived peak fluxes by the SEP challenge at two integral energies, i.e. E>10 and E>100 MeV. These values are presented in the last two columns of Table 5. The SEP event on 06 January 2014 was not associated with a solar flare (see Table 5), although, a back-sided origin of the driving CME and a (probably occulted) C2.1 solar flare[9] have been reported (Thakur et al. 2014). Moreover, the sequence of SEP events at 04 & 06 September 2017 are considered as one event leading to a single peak flux for the E>10 MeV. An SEP event started on 04 September 2017 with a significant enhancement recorded at E>10 MeV. However, on 06 September 2017 the strongest solar flare of solar cycle 24 was marked (i.e. X9.3) (Jiggens et al. 2019) and an SEP event was clearly visible in higher energies (up to E>100 MeV), while characterized also as a sub-GLE (Mishev et al. 2017). Nonetheless, the E>10 MeV flux continued to be sustained after its initiation on 04 September 2017 and throughout both SEP events lasting several days. Finally, four SEP events did not exceed a threshold of 1 pfu at E>100 MeV and thus are presented in Table 5 as "N/A" entries. It is important to note that in this part of our analysis, the parameters of Table 5 are taken "as is" from the campaign event website and all comparisons and results are driven by the parent solar event parameters and the observed peak proton fluxes at E>10 & E>100 MeV.

Table 5. Details of the SEP events' solar event parameters (i.e. solar flare and CME characteristics) and observed peak proton flux at E>10-; & E>100 MeV, used in the validation, provided under the SHINE/ISWAT/ESWW SEP Model Validation Challenge.

SEP event	**Flare information**			**CME information**			**E>10 MeV**	**E>100 MeV**
Date	Onset Time (HH:MM)	Mag. (W/m^2)	Lon. (°)	Start Time (HH:MM)	Speed (km/s)	Width (°)	Peak Flux (pfu)	Peak Flux (pfu)
2012-03-07	0:02	X5.4	E15	0:24	2684	360	6529.8	69.272
2012-05-17	1:25	M5.1	W89	1:48	1582	360	255.44	20.445
2012-07-12	15:37	X1.4	W02	16:48	885	360	96.08	N/A
2013-04-11	6:55	M6.5	E12	7:24	861	360	113.55	2.0297
2014-01-06	-	-	-	| 8:00	1402	360	42.17	4.079
2014-01-07	18:04	X1.2	W11	18:24	1830	360	1026.1	4.2687
2017-07-14	1:07	M2.4	W33	1:25	1200	360	22.374	N/A
2017-09-04	20:28	M5.5	W16	20:36	1418	360	844.38	N/A
2017-09-06	11:53	X9.3	W34	12:24	1571	360		
2017-09-10	15:35	X8.2	W88	16:00	3163	360	1493.5	68.128

---

[8] https://ccmc.gsfc.nasa.gov/assessment/topics/SEP/campaign2020.php

[9] https://cdaw.gsfc.nasa.gov/CME_list/sepe/

4.1. Probability of detection

Using the associated parent solar events as input parameters (see Table 5), derived the probability of SEP detection per event and for all three PROSPER's modes of operation (see Table 1). Figure 12 presents the obtained outputs. Results are presented in red (flare mode of operation), blue (CME mode of operation) and green (solar flare & CME mode of operation). Each panel corresponds to each integral energy (i.e. E>10-; >30-; and >100 MeV). Within the panel of E>10 MeV (upper panel) the threshold (*pt*) above which the probabilistic forecast is transformed into a categorical one is over-plotted with a dotted horizontal black line. This *pt* was directly obtained by Table 1 of Anastasiadis et al. (2017) and corresponds to the threshold above which an SEP event would have been identified. Such a *pt* is obtained at the perfect skill of a forecasting algorithm and is usually applied for a direct discrete analysis of SEP yes/no. However, it must be noted that PROSPER gives a probabilistic forecast. From the inspection of the E>10 MeV panel in Figure 12, it seems that the PROSPER's mode of operation based on solar flare input (red bars) would spot almost all of the events: ~90% (i.e. 8/9 events for which flare input was available). The SEP event that would not have been spotted by the PROPSER's flare mode of operation is the one that occurred on 14 July 2017. This SEP event was associated with a "well-connected" (W33^o) solar flare but with relatively modest-magnitude (M2.4) that, although led to a probability of SEP occurrence of ~20%, was lower than the *pt* and thus would not have been spotted by PROSPER. Moreover, the P(SEP) seems to be quite successful (100%) for the CME module (i.e. 10/10 events for which CME input was available were identified as SEP events (i.e. crossing the *pt*). Adding to this, the 14 July 2017 event that was "missed" by the flare mode of operation was associated with a halo CME with a speed of 1200 km/s – which enhanced the probability of SEP occurrence to ~ 58% in the CME mode and thus pushed it above the *pt* threshold, resulting into an identification of the event by PROSPER. Finally, the green bars represent PROSPER's outputs for the solar flare & CME mode of operation. As it can be seen, the hit rate is 100% (i.e. 9/9 events for which both flare and CME inputs were available).

These outputs (i.e. P(SEP)) stem from the fits presented in Section 3 here above. In particular, 5/9 (~66%) of the parent solar flares included in Table 5 are "poorly connected" (i.e. longitude <20^o) and 4/9 (~44%) are "well connected" (i.e. longitude ≥ 20^o). Hence, the fits used in these cases are presented in Figure 8. Figure 12 shows that at an integral energy E>10 MeV the largest differences in the achieved P(SEP) are due to the magnitude of the associated solar flare. For example, the first two events on the list are the 07 March 2012 (X5.4/E15 -"poorly connected") and 17 May 2012 (M5.1/W89 - "well connected") with the obtained probability being higher for the former of the two (see Figure 12). When shifting to higher energies (E>100 MeV) the differences in the obtained P(SEP) values (red bars) become larger because of the additional separation of the fits due to the longitudinal difference. Moreover, the two events on 06 & 10 September 2017 result to very high probabilities of SEP occurrence, which seem to be preserved in each integral energy of interest. The specific SEP events are associated with very strong (>X8.0) and "well connected" solar flares, which seem to be dominant. All driving CMEs are Halo and fast, thereby all of the P(SEP) values (blue bars) are obtained from the blue lines (fits) of Figure 6, and the differences are due to the V_{CME} inputs, with larger energies requiring a higher value of V_{CME} in order to lead to a larger P(SEP). That said, for the same V_{CME} the obtained P(SEP) is lower when shifting to higher energies in all tested cases. Finally, for the flare & CME mode of operation (green bars) one can notice that for the cases of 12 July 2012 and 11 April 2013 that are both "poorly connected" ones, associated with comparable CMEs (see Table 5) the difference between the achieved P(SEP) is driven by the magnitude of the associated solar flare and the former case leads to higher P(SEP) at all integral energies. For the SEP event on 04 September 2017 which is also a “poorly connected” one the V_{CME} is almost a factor of ~1.7 larger than the previous two cases, although the magnitude of the associated solar flare is M5.5 (i.e. less strong than the previous two cases). However, the obtained P(SEP) is higher than both cases at all integral energies, pointing to the role of the V_{CME} in the establishment of that probability. Finally, the SEP event on 14 July 2017 which is a “well connected” one and associated with a CME of V_{CME} = 1200 km/s leads to a higher P(SEP) compared to the case of the “poorly connected” SEP event at 11 April 2013 at all integral energies. In the latter case the associated flare is stronger (M6.5 *versus* M2.4) but the CME speed is slower (861 *versus* 1200 km/s). Thus the better magnetic connection and the higher V_{CME} seem to lead to a higher P(SEP).

4.2. Peak proton flux

Figure 13 provides scatter plots of the predicted peak flux versus the observed peak flux at E>10 MeV (top row) and E>100 MeV (bottom row) for each of the three PROSPER’s modes of operation. In particular the first column corresponds to the PROSPER’s outputs based on SF input, the middle column refers to the outputs of PROSPER utilizing CME input alone and the third column represents outputs from PROSPER’s third mode of operation that makes use of both SF and CME inputs. Blue circles denote the obtained outputs at a 50% CL (lower limit) and red circles present the same outputs for a 90% CL (upper limit).

The predicted peak flux at E > 10 MeV as a function of the observed peak flux at the same energy, utilizing only SF information (upper panel on the left hand side of Figure 13), shows a strong correlation especially in the lower limit (i.e. 50% CL) with the correlation coefficient (*cc*) being 0.76 (lower limit) and 0.72 (upper limit). Moreover, more than half of the predicted values seem to be very close to the dichotomous line which indicates a perfect prediction. However, the upper limit predictions (90% CL) show an over-forecasting being above the perfect dichotomous prediction line, with these predictions being within one order of magnitude from the observed peak fluxes, but not always. Nonetheless, the lower and the upper limits seem to capture the actual peak flux quite well at E>10 MeV. When moving to the CME mode of operation (column in the middle) for E>10 MeV, a weaker correlation is observed with the *cc* being 0.33 (lower limit) and 0.32 (upper limit) respectively. The tendency for over predicting is also visible in this case and holds for the majority of the events both at the lower and the upper limit. At the same time, the CME mode of operation, seems not to provide granularity since all of the points are predicted using as input halo and fast CME and with the expectation of an SEP event reaching E>100 MeV. Hence the differences are small between events, since all outputs are derived by the fits of the bottom panel of Figure 10. As a result, those small differences are corroborated with the small differences between the established P(SEP) (see Equation 14). Moving to the results of the combined SF & CME mode of operation (column at the right hand side), a stronger correlation is achieved (*cc* is 0.77 - lower limit & 0.71 - upper limit). Furthermore, the predictions at both the lower and the upper limit seem to be very close to the dichotomous line, with a small tendency for over-forecasting in the upper limit predictions. Nonetheless, the predictions at the lower limit for the strongest SEP events (i.e. those achieving the largest peak flux) at E>10 MeV seem to fall at the perfect dichotomous line.

In the case of E>100 MeV, the peak flux seems to be captured well by all three modes of operation, with the obtained *cc* being on average ≥ 0.90. Especially with the upper limit in the case of the solar flare input (panel at the left hand side at the bottom row) and the flare & CME (panel on the right hand side at the bottom row) modules the agreement is quite reasonable with most of the events predicted at the upper limit (90%) being either on the dichotomous line or very close to it (i.e. within less then 1 order of magnitude).

The results presented in Figure 13, underline the inherent difficulty of the SEP characteristics prognosis. However, the usage of the upper (90%) and lower (50%) limit in PROSPER shows promise and seems to capture the expected peak flux of the SEP events, reasonably well.

5. Discussion & Conclusions

A new probabilistic model, PROSPER, that provides short-term forecasting (nowcasting) of SEP events (i.e. probability of occurrence and corresponding peak flux) has been presented in this work. In particular, PROSPER was implemented based on a straightforward application of the Bayes theorem using probability distribution functions constructed from a large database of solar flares, CMEs and SEPs. The output of the prediction is the probability of a solar event to lead to an SEP event, i.e. P(SEP). As a second step, the estimated peak proton flux of the SEP event is evaluated. PROSPER has three modes of operation based on the inputs received. There is one mode that utilizes solar flare data (magnitude and longitude), one that makes use of CME identifications (AW and speed) and a third one that makes use of the combined input of solar flares and CMEs (see Table 1). The basic implementations of PROSPER presented in this study are summarized as follows:

- **Probability of SEP occurrence**. For the identification of P(SEP) in each case, log-normal fits (Equation 1) to the data were utilized (see e.g. Figures 3 & 4) and the resulting distributions were combined under the Bayes formula (Equation 11). These Bayes fits were obtained for different integral energies spanning from E>10-; E>30- and E>100 MeV (see Figures 6, 8 and 10).
- **Peak fluxes**. The modeling of the peak fluxes was based on exponential cut-off fits (Equation 14) to cumulative distribution functions (CDFs) constructed by the recorded peak fluxes of the SEP events in our sample. In this case, first the data are filtered based on the highest expected integral energy that an SEP event will occur (e.g. applying a thresholding on the P(SEP) obtained in the previous step and thus utilizing the prediction of an event being expected to reach E>100 MeV, E>30 MeV or E>10 MeV). Detailed fits are presented in Figures 7, 9 and 11 as well as in Tables 2, 3 and 4, respectively.

The ramifications of Bayes' rule are countless and have gained wide recognition in data driven, Space Weather related topics. Nonetheless, to our knowledge, PROSPER is the first SEP prediction model that allows the estimation of P(SEP) utilizing Bayes' theorem (see Camporeale 2019, and references there in). PROSPER allows a direct estimation of P(SEP), based on a previous outcome having occurred in similar circumstances and in doing so, it takes into account all available data without any a priori bias. As a result, the benefit of such an approach is the natural and principled way of combining prior information with data, within a solid decision theoretical framework, providing exact inferences that are conditional on data (Hartigan 2012).

Moreover, all PROSPER's modes of operation were validated based on detailed case studies of SEP events selected as part of the NASA CCMC SEP scoreboard challenge (see Table 5). Blind tests with archived parent solar data of these events were applied to PROSPER's modes in order to derive the probabilities of SEP detection per mode of operation and per SEP event. The obtained results were discussed in a comparative manner, showing that the magnetic connectivity (i.e. "well connected" events), strong solar flares and fast CMEs lead to higher achieved

P(SEP) values. The validation results of the PROSPER models are based on a sample of 10 SEP events and are summarized as follows:

- PROSPER's module that is based on solar flare input would spot ~90% (i.e. 8/9) of the events for which flare input was available.
- For the CME module the hit rate is 100% (i.e. 10/10 events for which CME input was available were identified as SEP events).
- The solar flare & CME mode of operation, also has a hit rate of 100% (i.e. 9/9 events for which both flare and CME inputs were available).

Additionally, the largest differences in the achieved P(SEP) in lower energies (i.e. at an integral energy E>10 MeV) are due to the magnitude of the associated solar flare, F_{SXR}. For the higher energies (i.e. E>100 MeV) the differences in the obtained P(SEP) values become larger due to the additional separation of the fits, which is driven by their longitudinal difference. Finally, the solar flare & CME module revealed that the better magnetic connection and the higher V_{CME} seem to lead to a higher P(SEP).

Furthermore, we included scatter plots of observed versus predicted SEP peak fluxes to quantify the reliability of PROSPER's predictions. Although the SEP peak fluxes are difficult to infer based on the characteristics of their parent solar events, the usage of an upper (90%) and lower (50%) limit in PROSPER seemed to capture the expected peak flux of the SEP events, reasonably well. In particular, at E > 10 MeV, for the solar flare module, the predicted peak flux, as a function of the observed peak flux shows a strong correlation especially in the lower limit (i.e. 50% CL) with more than half of the the predicted values being close to the dichotomous line – indicating a perfect prediction. Nonetheless, the CME module at the same energy seems not to capture the expected peak flux with granularity. However, the combined SF & CME mode of operation shows that the predictions at both the lower and the upper limit seem to be very close to the dichotomous line, with a small tendency for over-forecasting in the 90% limit predictions. Nonetheless, the predictions at 50% for the SEP events achieving the largest peak flux at E>10 MeV seem to fall at the perfect dichotomous line. Additionally, in the case of E>100 MeV, the peak flux seems to be captured well by all three modes of operation, especially with the 90% limit in the case of the SF and the flare & CME modules the agreement is quite reasonable with most of the events predicted at 90% being either on the dichotomous line or very close to it (i.e. within less then 1 order of magnitude).

The PROSPER model has been incorporated into the new operational Advanced Solar Particle Event Casting System (ASPECS) tool providing outputs in real-time through the web portal http://phobos-srv.space.noa.gr/. In addition, ASPECS offers the capability to interrogate PROSPER for historical cases via a run on demand functionality. ASPECS is the first realisation of ESA's SEP Advanced Warning System (SAWS) - a modular framework for forecasting solar energetic particle (SEP) events, their characteristics and profiles.

A point that still needs to be addressed in the validation is how the model reacts when solar events not associated with SEPs are introduced as inputs. This is a part of an ongoing follow-up study focusing on the performance of the ASPECS tool. On a preliminary basis, it can be commented that PROSPER achieves categorical scores in line with the expected range for data driven models. The range of scores are taken from Table 1 of Anastasiadis et al. (2017) (i.e. Probability of Detection – POD ~50-70%, False Alarm Rate – FAR ~30-50%). Such a range is realistic and inherently imposed due to the imbalanced dataset that is being used for SEP prediction efforts (see the relevant detail discussion in Lavasa et al. 2021). In fact, recently, Stumpo et al. (2021) have shown that the major drawback in predicting the occurrence of SEPs, i.e. P(SEP), in the framework of statistical forecasting concepts, is the optimization of FAR which directly depends on the imbalance of the dataset used. That means that, the greater the imbalance, the greater the FAR is affected by the presence of false positives. Taking all of these into account, in our next study such a thorough validation step will be implemented.

It should be mentioned that, PROSPER's CME and flare & CME modes of operation are subject to the CME input data used. In particular, while developing PROSPER, CME data from the CDAW CME catalogue have been used. It was pointed out by Richardson et al. (2015) that different CME catalogues provide different estimations of the same CME event, hence given that PROSPER is a data driven model, it is inherently affected by such differences. On top of the CDAW CME catalogue, other options include: (a) the Computer-Aided CME Tracking (CACTUS) catalog (https://wwwbis.sidc.be/cactus/), which is compiled using a specialized software package (Robbrecht and Berghmans 2004); (b) the SEEDS (Solar Eruptive Event Detection System) catalog (http://spaceweather.gmu.edu/seeds/) (Olmedo et al. 2008); (c) the CORIMP (coronal image processing) method catalog (http://alshamess.ifa.hawaii.edu/CORIMP/) (Byrne et al. 2012) and (d) the Space Weather Database Of Notification, Knowledge, Information (DONKI) developed by CCMC (https://ccmc.gsfc.nasa.gov/donki/). Therefore, a natural next step of this study is to adapt PROSPER to each of these catalogues and quantify differences/changes following the study by Richardson et al. (2015).

The usage of CME identifications (e.g. angular width and speed) for the derivation of the SEP occurrence probabilities (P(SEP)) and expected peak proton flux have been explored, already, in a few operational efforts (e.g. Dierckxsens et al. 2015; Papaioannou et al. 2018b; Richardson et al. 2018). On top of that statistical studies have pointed out that P(SEP) increases as a function of solar flare magnitude (F_{SXR}), longitude and CME speed (V_{CME}) - especially when considering solar flare-CME couples situated on the west part of the visible solar disk (i.e. Dierckxsens et al. 2015). In addition, it was recently showed that F_{SXR}, V_{CME} and SXR fluence have the largest potential for discriminating solar events that are associated with SEP events (Lavasa et al. 2021), while as many as targeted parameters need to be taken into account, in order to build a reliable and more accurate SEP predicting system (see e.g. Papaioannou et al. 2018b). PROSPER takes advantage of such findings and comes with three modes of operation. Nevertheless, its performance certainly benefits from the provision of reliable CME estimates or a proper proxy for the CME speed and width, which would be very much desirable to obtain in near-real time.

Acknowledgements. This work was supported through the ESA Contract No. 4000120480/NL/LF/hh "Solar Energetic Particle (SEP) Advanced Warning System (SAWS)". Athanasios Papaioannou and Angels Aran acknowledge the support from the project MDM-2014-0369 of ICCUB (Unidad de Excelencia "María de Maeztu"). The CME Catalog used in this work 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. Athanasios Papaioannou, Rami Vainio & Anastasios Anastasiadis acknowledge the International Space Science Institute and the supported International Team 441: High EneRgy sOlar partICle Events Analysis (HEROIC, http://www.issibern.ch/teams/heroic/). Finally, Athanasios Papaioannou acknowledges support from NASA/LWS project NNH19ZDA001N-LWS. The authors would like to thank Dr Manolis Georgoulis for stimulating discussions, Dr Katie Whitman for fruitful collaboration and exchange of ideas as concerns the validation of PROSPER and Mr George Vasalos for valuable technical assistance. Furthermore, the authors would like to thank the anonymous referees for a critical and constructive reading of the manuscript and for valuable comments that improved the contents of the paper.

References

Anastasiadis, A., D. Lario, A. Papaioannou, A. Kouloumvakos, and A. Vourlidas. Solar energetic particles in the inner heliosphere: status and open questions. *Philosophical Transactions of the Royal Society A*, **377**(2148), 2019. URL https://doi.org/10.1098/rsta.2018.0100. 1

Anastasiadis, A., A. Papaioannou, I. Sandberg, M. Georgoulis, K. Tziotziou, A. Kouloumvakos, and P. Jiggens. Predicting flares and solar energetic particle events: The FORSPEF tool. *Solar Physics*, **292**(9), 1–21, 2017. URL https://doi.org/10.1007/s11207-017-1163-7. 1, 4.1, 5, 12

Arregui, I. Recent applications of Bayesian methods to the solar corona. *Frontiers in Astronomy and Space Sciences*, 29, 2022. URL https://doi.org/10.3389/fspas.2022.826947. 1

Aschwanden, M. J. Particle acceleration and kinematics in solar flares–A synthesis of recent observations and theoretical concepts (Invited Review). *Space Science Reviews*, **101**(1-2), 1–227, 2002. DOI:10.1023/A:1019712124366, URL http://dx.doi.org/10.1023/A:1019712124366. 1

Asvestari, E., T. Willamo, A. Gil, I. Usoskin, G. Kovaltsov, V. Mikhailov, and A. Mayorov. Analysis of Ground Level Enhancements (GLE): Extreme solar energetic particle events have hard spectra. *Advances in Space Research*, **60**(4), 781–787, 2017. URL https://doi.org/10.1016/j.asr.2016.08.043. 1

Azzam, E. I., J.-P. Jay-Gerin, and D. Pain. Ionizing radiation-induced metabolic oxidative stress and prolonged cell injury. *Cancer letters*, **327**(1-2), 48–60, 2012. URL https://doi.org/10.1016/j.canlet.2011.12.012. 1

Baker, D. N. Specifying and forecasting space weather threats to human technology. In I. Daglis, ed., Effects of space weather on technology infrastructure, NATO Science Series II: Mathematics, Physics and Chemistry, 1–25. Springer,Dordrecht, 2004. URL https://doi.org/10.1007/1-4020-2754-0_1. 1

Balch, C. C. Updated verification of the Space Weather Prediction Center's solar energetic particle prediction model. *Space Weather*, **6**(1), 2008. DOI:10.1029/2007SW000337, URL http://dx.doi.org/10.1029/2007SW000337. 1

Bayes, M., and M. Price. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London Series I*, **53**, 370–418, 1763. 1, 2.2.1

Belov, A., H. Garcia, V. Kurt, H. Mavromichalaki, and M. Gerontidou. Proton enhancements and their relation to the X-ray flares during the three last solar cycles. *Solar Physics*, **229**(1), 135–159, 2005. DOI:10.1007/s11207-005-4721-3, URL http://dx.doi.org/10.1007/s11207-005-4721-3. 1

Bizzarri, M., M. G. Masiello, R. Guzzi, and A. Cucina. Journey to Mars: A Biomedical Challenge. Perspective on future human space flight. *Organisms. Journal of Biological Sciences*, **1**(2), 15–26, 2017. URL https://doi.org/10.13133/2532-5876/14197. 1

Brueckner, G., R. Howard, M. Koomen, C. Korendyke, D. Michels, et al. The large angle spectroscopic coronagraph (LASCO). *Solar Physics*, **162**(1-2), 357–402, 1995. DOI:10.1007/BF00733434, URL http://dx.doi.org/10.1007/BF00733434. 2.1

Bütikofer, R., E. O. Flückiger, L. Desorgher, M. R. Moser, and B. Pirard. The solar cosmic ray ground-level enhancements on 20 January 2005 and 13 December 2006. *Advances in space research*, **43**(4), 499–503, 2009. URL https://doi.org/10.1016/j.asr.2008.08.001. 1

Byrne, J. P., H. Morgan, S. R. Habbal, and P. T. Gallagher. Automatic Detection and Tracking of Coronal Mass Ejections. II. Multiscale Filtering of Coronagraph Images. *ApJ*, **752**(2), 145, 2012. 10.1088/0004-637X/752/2/145, 1207.6125. 5

Camporeale, E. The Challenge of Machine Learning in Space Weather: Nowcasting and Forecasting. *Space Weather*, **17**(8), 1166–1207, 2019. 10.1029/2018SW002061, 1903.05192. 5

Cane, H., and D. Lario. An introduction to CMEs and energetic particles. *Space science reviews*, **123**(1-3), 45–56, 2006. DOI:10.1007/s11214-006-9011-3, URL http://dx.doi.org/10.1007/s11214-006-9011-3. 1

Cane, H., I. Richardson, and T. von Rosenvinge. A study of solar energetic particle events of 1997–2006: Their composition and associations. *Journal of Geophysical Research: Space Physics*, **115**(A8), 2010. DOI:10.1029/2009JA014848, URL http://dx.doi.org/10.1029/2009JA014848. 1

Cliver, E. W., A. G. Ling, A. Belov, and S. Yashiro. Size distributions of solar flares and solar energetic particle events. *ApJ*, **756**(2), L29, 2012. 10.1088/2041-8205/756/2/L29. 2.2.1

Cucinotta, F. A., H. Wu, M. R. Shavers, and K. George. Radiation dosimetry and biophysical models of space radiation effects. *Gravitational and Space Biology*, **16**(2), 11–19, 2003. 1

Daly, E., G. Drolshagen, A. Hilgers, and H. Evans. Space environment analysis: Experience and trends. In Environment Modeling for Space-Based Applications, vol. 392, 15, 1996. 1

Desai, M., and J. Giacalone. Large gradual solar energetic particle events. *Living Reviews in Solar Physics*, **13**(1), 3, 2016. URL https://doi.org/10.1007/s41116-016-0002-5. 1

Dierckxsens, M., K. Tziotziou, S. Dalla, I. Patsou, M. Marsh, N. Crosby, O. Malandraki, and G. Tsiropoula. Relationship between solar energetic particles and properties of flares and CMEs: Statistical analysis of solar cycle 23 events. *Solar Physics*, **290**(3), 841–874, 2015. DOI:10.1007/s11207-014-0641-4, URL http://dx.doi.org/10.1007/s11207-014-0641-4. 1, 5

Feynman, J., and S. Gabriel. On space weather consequences and predictions. *Journal of Geophysical Research: Space Physics*, **105**(A5), 10,543–10,564, 2000. URL https://doi.org/10.1029/1999JA000141. 1

Garcia, H. Forecasting methods for occurrence and magnitude of proton storms with solar hard X rays. *Space Weather*, **2**(6), 2004a. DOI:10.1029/2003SW000035, URL http://dx.doi.org/10.1029/2003SW000035. 1

Garcia, H. Forecasting methods for occurrence and magnitude of proton storms with solar soft X rays. *Space Weather*, **2**(2), 2004b. DOI:10.1029/2003SW000001, URL http://dx.doi.org/10.1029/2003SW000001. 1

Gopalswamy, N., S. Yashiro, S. Krucker, G. Stenborg, and R. A. Howard. Intensity variation of large solar energetic particle events associated with coronal mass ejections. *Journal of Geophysical Research: Space Physics*, **109**(A12), 2004. DOI:10.1029/2004JA010602, URL http://dx.doi.org/10.1029/2004JA010602. 1

Gopalswamy, N., S. Yashiro, A. Lara, M. Kaiser, B. Thompson, P. Gallagher, and R. Howard. Large solar energetic particle events of cycle 23: A global view. *Geophysical research letters*, **30**(12), 2003. DOI:10.1029/2002GL016435, URL http://dx.doi.org/10.1029/2002GL016435. 1

Gopalswamy, N., S. Yashiro, G. Michalek, G. Stenborg, A. Vourlidas, S. Freeland, and R. Howard. The SOHO/LASCO CME Catalog. *Earth Moon and Planets*, **104**(1-4), 295–313, 2009. 10.1007/s11038-008-9282-7, URL https://dx.doi.org/10.1007/s11038-008-9282-7. 2.1

Harboe-Sørensen, R. 40 years of radiation single event effects at the european space agency, ESTEC. *IEEE Transactions on Nuclear Science*, **60**(3), 1816–1823, 2013. URL https://dx.doi.org/10.1109/TNS.2013.2247630. 1

Hartigan, J. A. Bayes theory. Springer Science & Business Media, 2012. 5

Jiggens, P., C. Clavie, H. Evans, T. O'Brien, O. Witasse, et al. In situ data and effect correlation during September 2017 solar particle event. *Space Weather*, **17**(1), 99–117, 2019. URL https://doi.org/10.1029/2018SW001936. 4

Jiggens, P., D. Heynderickx, I. Sandberg, P. Truscott, O. Raukunen, and R. Vainio. Updated model of the solar energetic proton environment in space. *Journal of Space Weather and Space Climate*, **8**, A31, 2018. URL https://doi.org/10.1051/swsc/2018010. 1, 2.2.2

Joyce, J. M. The foundations of causal decision theory. Cambridge University Press, 1999. 2.2.1

Kahler, S. The correlation between solar energetic particle peak intensities and speeds of coronal mass ejections: Effects of ambient particle intensities and energy spectra. *Journal of Geophysical Research: Space Physics*, **106**(A10), 20,947–20,955, 2001. DOI:10.1029/2000JA002231, URL http://dx.doi.org/10.1029/2000JA002231. 1

Kahler, S. W., and A. G. Ling. Forecasting Solar Energetic Particle (SEP) events with Flare X-ray peak ratios. *Journal of Space Weather and Space Climate*, **8**, A47, 2018. URL https://doi.org/10.1051/swsc/2018033. 1

Lario, D., R.-Y. Kwon, A. Vourlidas, N. Raouafi, D. Haggerty, et al. Longitudinal Properties of a Widespread Solar Energetic Particle Event on 2014 February 25: Evolution of the Associated CME Shock. *The Astrophysical Journal*, **819**(1), 72, 2016. DOI:10.3847/0004-637X/819/1/72, URL http://dx.doi.org/10.3847/0004-637X/819/1/72. 1

Laurenza, M., T. Alberti, and E. Cliver. A short-term ESPERTA-based forecast tool for moderate-to-extreme solar proton events. *The Astrophysical Journal*, **857**(2), 107, 2018. URL https://doi.org/10.3847/1538-4357/aab712. 1

Laurenza, M., E. Cliver, J. Hewitt, M. Storini, A. Ling, C. Balch, and M. Kaiser. A technique for short-term warning of solar energetic particle events based on flare location, flare size, and evidence of particle escape. *Space Weather*, **7**(4), 2009. DOI:10.1029/2007SW000379, URL http://dx.doi.org/10.1029/2007SW000379. 1

Lavasa, E., G. Giannopoulos, A. Papaioannou, A. Anastasiadis, I. Daglis, A. Aran, D. Pacheco, and B. Sanahuja. Assessing the Predictability of Solar Energetic Particles with the use of Machine Learning techniques. *Solar Physics*, **296**(7), 1–47, 2021. URL https://doi.org/10.1007/s11207-021-01837-x. 1, 5

Maritz, J. S., and T. Lwin. Empirical Bayes Methods. Routledge, 2018. 2.2.1

Mironova, I., and I. Usoskin. Possible effect of strong solar energetic particle events on polar stratospheric aerosol: a summary of observational results. *Environmental Research Letters*, **9**(1), 015,002, 2014. URL https://doi.org/10.5194/acp-13-8543-2013. 1

Miroshnichenko, L. I. Retrospective analysis of GLEs and estimates of radiation risks. *Journal of Space Weather and Space Climate*, **8**, A52, 2018. URL https://doi.org/10.1051/swsc/2018042. 1

Mishev, A., F. Adibpour, I. Usoskin, and E. Felsberger. Computation of dose rate at flight altitudes during ground level enhancements no. 69, 70 and 71. *Advances in Space Research*, **55**(1), 354–362, 2015. URL https://doi.org/10.1016/j.asr.2014.06.020. 1

Mishev, A., S. Poluianov, and I. Usoskin. Assessment of spectral and angular characteristics of sub-GLE events using the global neutron monitor network. *Journal of Space Weather and Space Climate*, **7**, A28, 2017. URL https://doi.org/10.1051/swsc/2017026. 4

Mishev, A., I. Usoskin, O. Raukunen, M. Paassilta, E. Valtonen, L. Kocharov, and R. Vainio. First analysis of ground-level enhancement (GLE) 72 on 10 September 2017: Spectral and anisotropy characteristics. *Solar Physics*, **293**(10), 1–15, 2018. URL https://doi.org/10.1007/s11207-018-1354-x. 1

Núñez, M. Predicting solar energetic proton events (E> 10 MeV). *Space Weather*, **9**(7), 2011. DOI:10.1029/2010SW000640, URL http://dx.doi.org/10.1029/2010SW000640. 1

Olmedo, O., J. Zhang, H. Wechsler, A. Poland, and K. Borne. Automatic Detection and Tracking of Coronal Mass Ejections in Coronagraph Time Series. *Sol. Phys.*, **248**(2), 485–499, 2008. 10.1007/s11207-007-9104-5. 5

Papaioannou, A., A. Anastasiadis, A. Kouloumvakos, M. Paassilta, R. Vainio, E. Valtonen, A. Belov, E. Eroshenko, M. Abunina, and A. Abunin. Nowcasting solar energetic particle events using principal component analysis. *Solar Physics*, **293**(7), 1–23, 2018a. URL https://doi.org/10.1007/s11207-018-1320-7. 1

Papaioannou, A., A. Anastasiadis, I. Sandberg, and P. Jiggens. Nowcasting of Solar Energetic Particle Events using near real-time Coronal Mass Ejection characteristics in the framework of the FORSPEF tool. *Journal of Space Weather and Space Climate*, **8**, A37, 2018b. URL https://doi.org/10.1051/swsc/2018024. 1, 2.2.1, 5

Papaioannou, A., I. Sandberg, A. Anastasiadis, A. Kouloumvakos, M. K. Georgoulis, K. Tziotziou, G. Tsiropoula, P. Jiggens, and A. Hilgers. Solar flares, coronal mass ejections and solar energetic particle event characteristics. *Journal of Space Weather and Space Climate*, **6**, A42, 2016. URL https://doi.org/10.1051/swsc/2016035. 1, 2.1, 2.2.2

Pellish, J. A., M. A. Xapsos, C. A. Stauffer, T. M. Jordan, A. B. Sanders, R. L. Ladbury, T. R. Oldham, P. W. Marshall, D. F. Heidel, and K. P. Rodbell. Impact of spacecraft shielding on direct ionization soft error rates for sub-130 nm technologies. *IEEE Transactions on Nuclear Science*, **57**(6), 3183–3189, 2010. 1

Posner, A. Up to 1-hour forecasting of radiation hazards from solar energetic ion events with relativistic electrons. *Space Weather*, **5**(5), 2007. DOI:10.1029/2006SW000268, URL http://dx.doi.org/10.1029/2006SW000268. 1

Raukunen, O., R. Vainio, A. J. Tylka, W. F. Dietrich, P. Jiggens, D. Heynderickx, M. Dierckxsens, N. Crosby, U. Ganse, and R. Siipola. Two solar proton fluence models based on ground level enhancement observations. *Journal of Space Weather and Space Climate*, **8**, A04, 2018. URL https://doi.org/10.1051/swsc/2017031. 1, 2.2.1, 2.2.2

Reames, D. V. Particle acceleration at the Sun and in the heliosphere. *Space Science Reviews*, **90**(3-4), 413–491, 1999. DOI:10.1023/A:1005105831781, URL http://dx.doi.org/10.1023/A:1005105831781. 1

Reames, D. V. Magnetic topology of impulsive and gradual solar energetic particle events. *The Astrophysical Journal Letters*, **571**(1), L63, 2002. DOI:10.1086/341149, URL http://dx.doi.org/10.1086/341149. 1

Reames, D. V. The Two Sources of Solar Energetic Particles. *Space Sci. Rev.*, **175**(1-4), 53–92, 2013. 10.1007/s11214-013-9958-9, 1306.3608. 1

Reames, D. V. What Are the Sources of Solar Energetic Particles? Element Abundances and Source Plasma Temperatures. *Space Science Reviews*, **194**(1-4), 303–327, 2015. DOI:10.1007/s11214-015-0210-7, URL http://dx.doi.org/10.1007/s11214-015-0210-7. 1

Richardson, I., M. Mays, and B. Thompson. Prediction of solar energetic particle event peak proton intensity using a simple algorithm based on CME speed and direction and observations of associated solar phenomena. *Space Weather*, **16**(11), 1862–1881, 2018. URL https://doi.org/10.1029/2018SW002032. 1, 5

Richardson, I. G., T. T. von Rosenvinge, and H. V. Cane. The Properties of Solar Energetic Particle Event-Associated Coronal Mass Ejections Reported in Different CME Catalogs. *Sol. Phys.*, **290**(6), 1741–1759, 2015. 10.1007/s11207-015-0701-4, 1505.03071. 5

Robbrecht, E., and D. Berghmans. Automated recognition of coronal mass ejections (CMEs) in near-real-time data. *A&A*, **425**, 1097–1106, 2004. 10.1051/0004-6361:20041302. 5

Rouillard, A., N. Sheeley, A. Tylka, A. Vourlidas, C. Ng, et al. The longitudinal properties of a solar energetic particle event investigated using modern solar imaging. *The Astrophysical Journal*, **752**(1), 44, 2012. DOI:10.1088/0004-637X/752/1/44, URL http://dx.doi.org/10.1088/0004-637X/752/1/44. 1

Sandberg, I., P. Jiggens, D. Heynderickx, and I. Daglis. Cross calibration of NOAA GOES solar proton detectors using corrected NASA IMP-8/GME data. *Geophysical Research Letters*, **41**(13), 4435–4441, 2014. DOI:10.1002/2014GL060469, URL http://dx.doi.org/10.1002/2014GL060469. 2.1

Sexton, F. W. Destructive single-event effects in semiconductor devices and ICs. *IEEE Transactions on Nuclear Science*, **50**(3), 603–621, 2003. URL https://doi.org/10.1109/TNS.2003.813137. 1

Stumpo, M., S. Benella, M. Laurenza, T. Alberti, G. Consolini, and M. F. Marcucci. Open Issues in Statistical Forecasting of Solar Proton Events: A Machine Learning Perspective. *Space Weather*, **19**(10), e02794, 2021. 10.1029/2021SW002794. 5

Swalwell, B., S. Dalla, and R. W. Walsh. Solar Energetic Particle Forecasting Algorithms and Associated False Alarms. *Sol. Phys.*, **292**(11), 173, 2017. 10.1007/s11207-017-1196-y, 1710.08156. 2.2.1

Thakur, N., N. Gopalswamy, H. Xie, P. Mäkelä, S. Yashiro, S. Akiyama, and J. Davila. Ground level enhancement in the 2014 January 6 solar energetic particle event. *The Astrophysical Journal Letters*, **790**(1), L13, 2014. URL https://doi.org/10.1088/2041-8205/790/1/L13. 4

Tobiska, W. K., W. Atwell, P. Beck, E. Benton, K. Copeland, et al. Advances in atmospheric radiation measurements and modeling needed to improve air safety. *Space Weather*, **13**(4), 202–210, 2015. URL https://doi.org/10.1002/2015SW001169. 1

Townsend, L. W. Effects of Space Radiation on Humans in Space Flight. *Space Weather Effects and Applications*, 63–78, 2021. URL https://doi.org/10.1002/9781119815570.ch3. 1

Trottet, G., S. Samwel, K.-L. Klein, T. D. de Wit, and R. Miteva. Statistical evidence for contributions of flares and coronal mass ejections to major solar energetic particle events. *Solar Physics*, **290**(3), 819–839, 2015. URL https://doi.org/10.1007/s11207-014-0628-1. 1

Usoskin, I., G. Kovaltsov, I. Mironova, A. Tylka, and W. Dietrich. Ionization effect of solar particle GLE events in low and middle atmosphere. *Atmospheric Chemistry and Physics*, **11**(5), 1979–1988, 2011. URL https://doi.org/10.5194/acp-11-1979-2011. 1

Vainio, R., L. Desorgher, D. Heynderickx, M. Storini, E. Flückiger, et al. Dynamics of the Earth's particle radiation environment. *Space science reviews*, **147**(3-4), 187–231, 2009. DOI:10.1007/s11214-009-9496-7, URL http://dx.doi.org/10.1007/s11214-009-9496-7. 1

Vlahos, L., A. Anastasiadis, A. Papaioannou, A. Kouloumvakos, and H. Isliker. Sources of solar energetic particles. *Philosophical Transactions of the Royal Society of London Series A*, **377**(2148), 20180,095, 2019. 10.1098/rsta.2018.0095, 1903.08200, URL https://doi.org/10.1098/rsta.2018.0095. 1

Wheatland, M. S. A Bayesian Approach to Solar Flare Prediction. *ApJ*, **609**(2), 1134–1139, 2004. 10.1086/421261, astro-ph/0403613. 1

Wheatland, M. S. A statistical solar flare forecast method. *Space Weather*, **3**(7), S07003, 2005. 10.1029/2004SW000131, astro-ph/0505311. 1

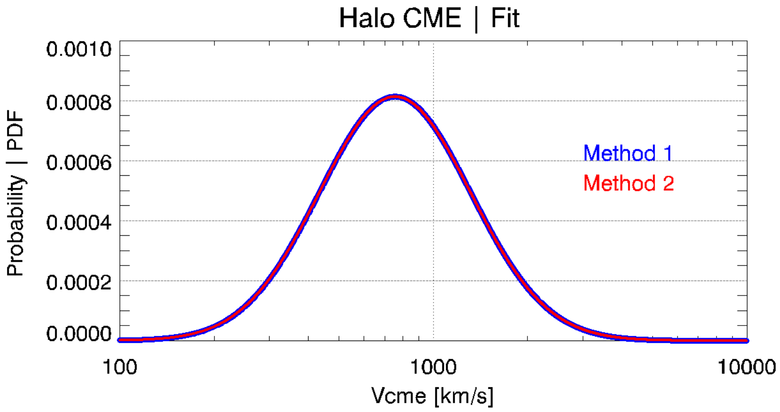

Fig. 1. An illustration of the two methods (*Method 1*; (blue color) and *Method 2*; (red color)), as a superposition investigated in this work. Both fits in the plots correspond to the Probability Distribution Functions (PDFs) for the case of Halo CMEs, with respect to the CME speed V_{CME}. See text for details

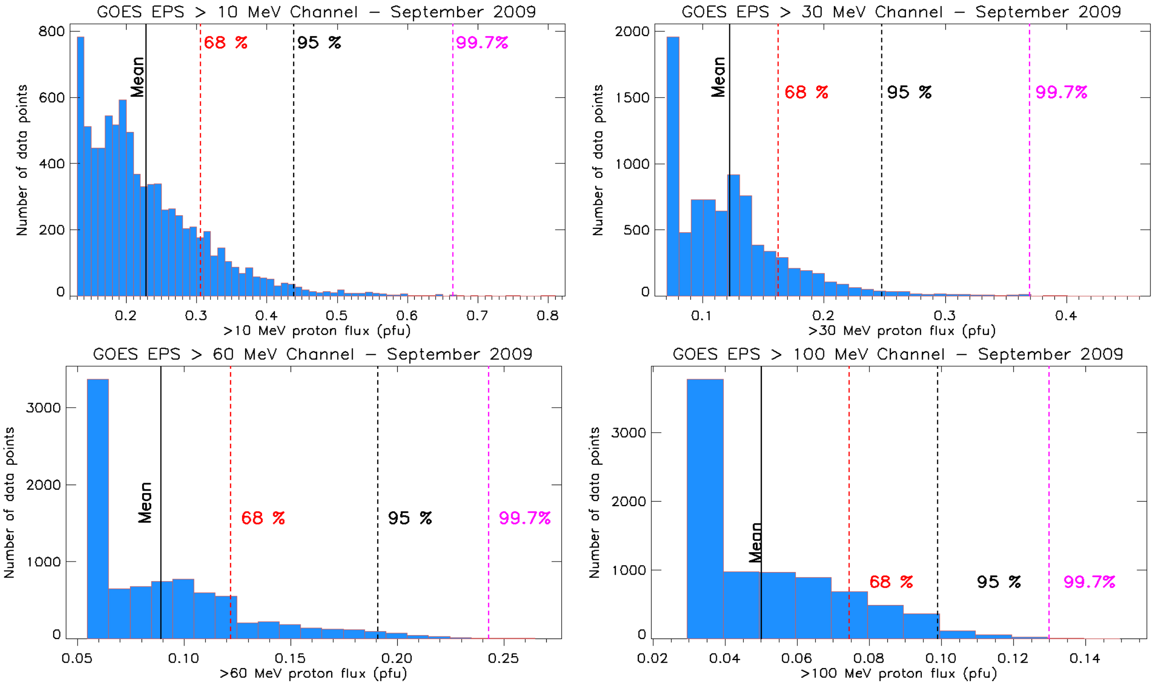

Fig. 2. The distributions of the GOES/EPS data per integral energy of interest for September 2009. The relevant percentiles of 68%-95% & 99.7% of the distributions are depicted as vertical dashed lines. The mean values of these distributions (presented as solid vertical lines) are used as the background level per channel in Equation 14.

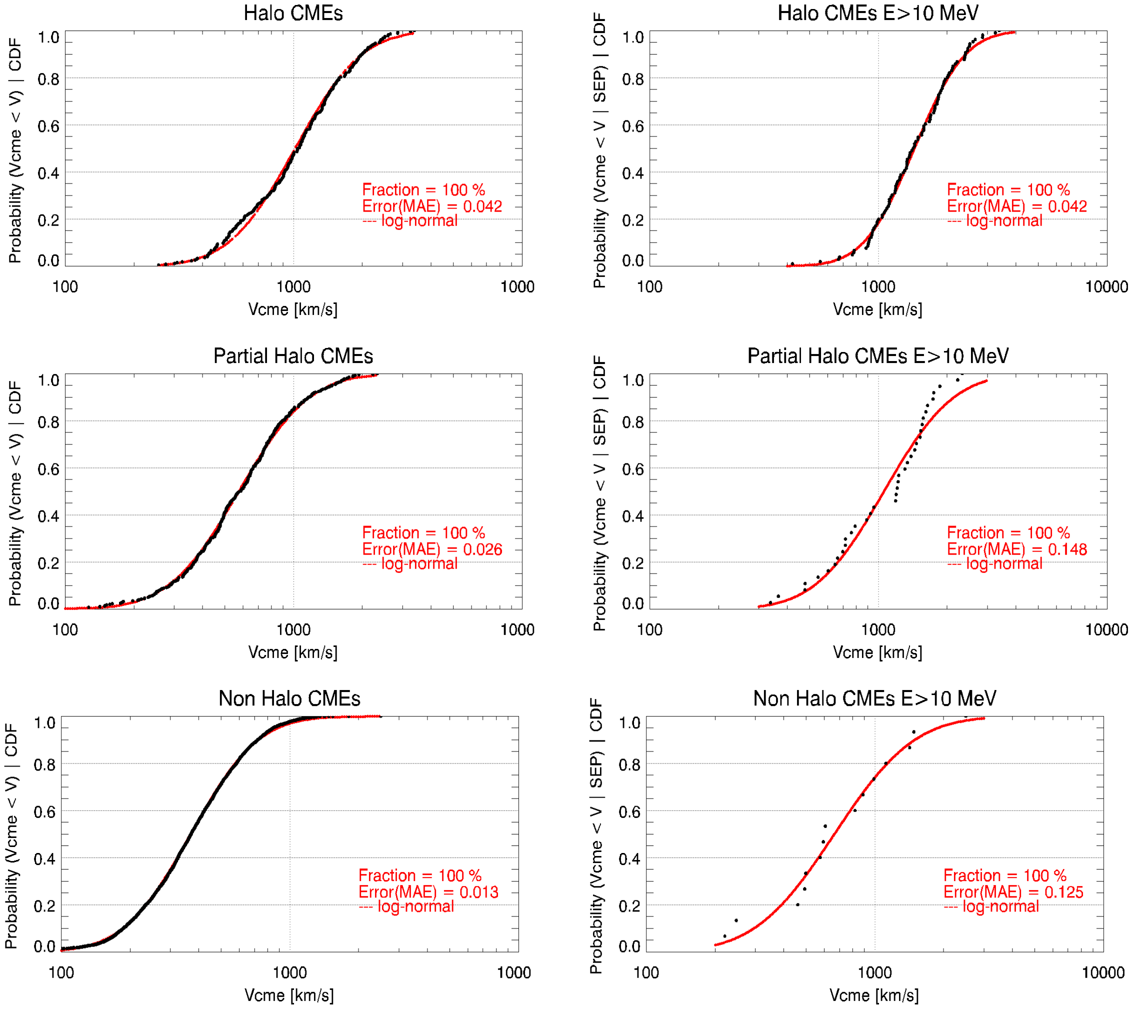

Fig. 3. The observational Cumulative Distribution Functions (CDFs) constructed by the database (i.e. all black points correspond to actual data). The red line depicts the log-normal fit to the data per case. Each plot further includes the fraction of the data explained by the fit (in [%]) [i.e. a 100% fraction means that all data points have been used in the corresponding fit], the mean absolute error (MAE) and the fit that was used (i.e. log-normal). From top to bottom the column on the left hand side corresponds to the CDFs per CME width (Halo, Partial Halo, Non-Halo) for all CMEs in each sample. The column on the right hand side corresponds to the CDFs in case that the CME was associated with an SEP at an energy of E>10 MeV, at each width bin. Similar fits (and plots) have been constructed for all other SEP integral energies of interest but are not shown.

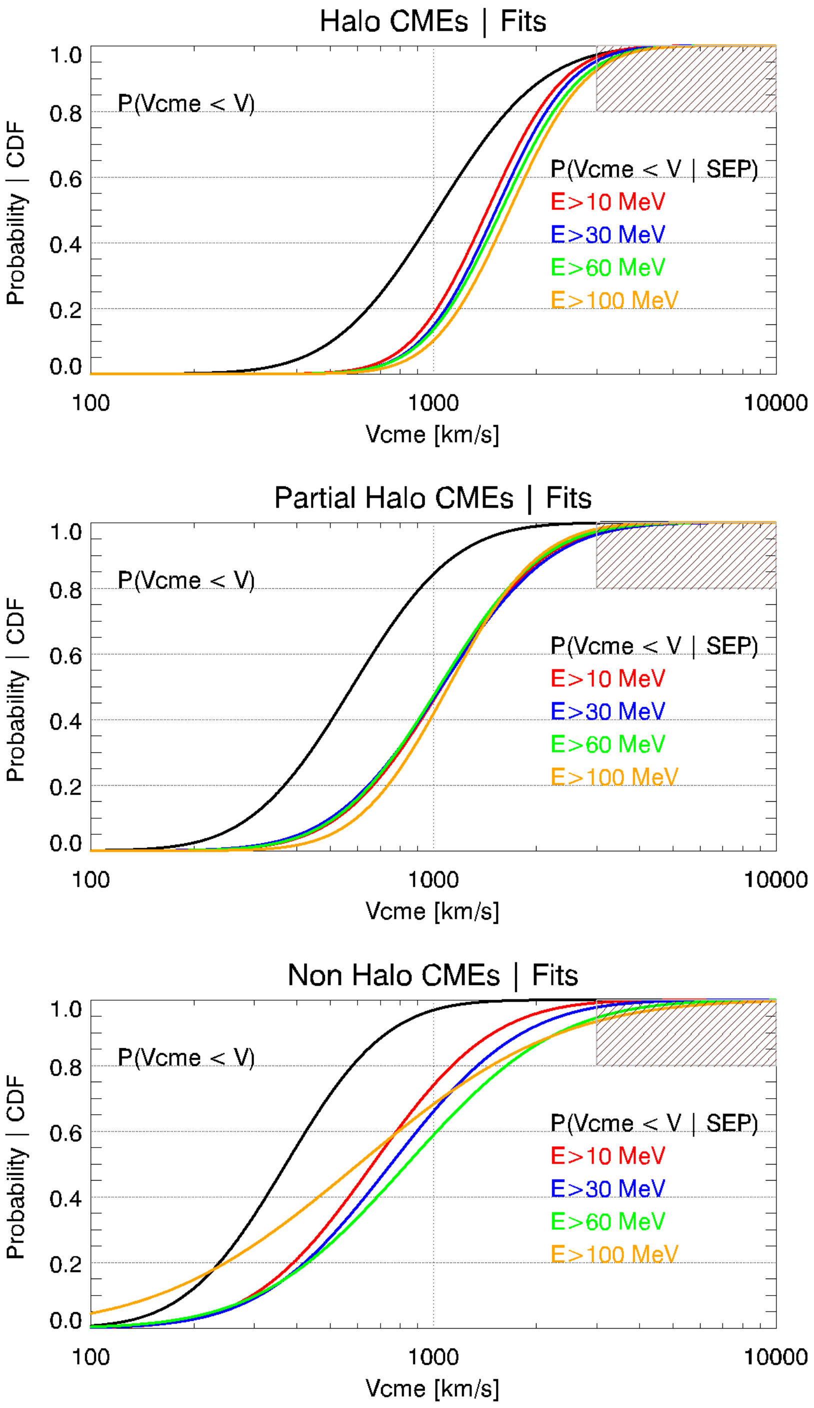

Fig. 4. The Cumulative Distribution Functions (CDFs) for the case of Halo CMEs (top panel); Partial Halo CMEs (middle panel) and Non-Halo CMEs (bottom panel). The black color represents all CMEs in the respective sample, whereas CMEs associated with SEP events for different integral energies are colour coded. The gray border hatched rectangle area provides the limit for the V_{CME}.

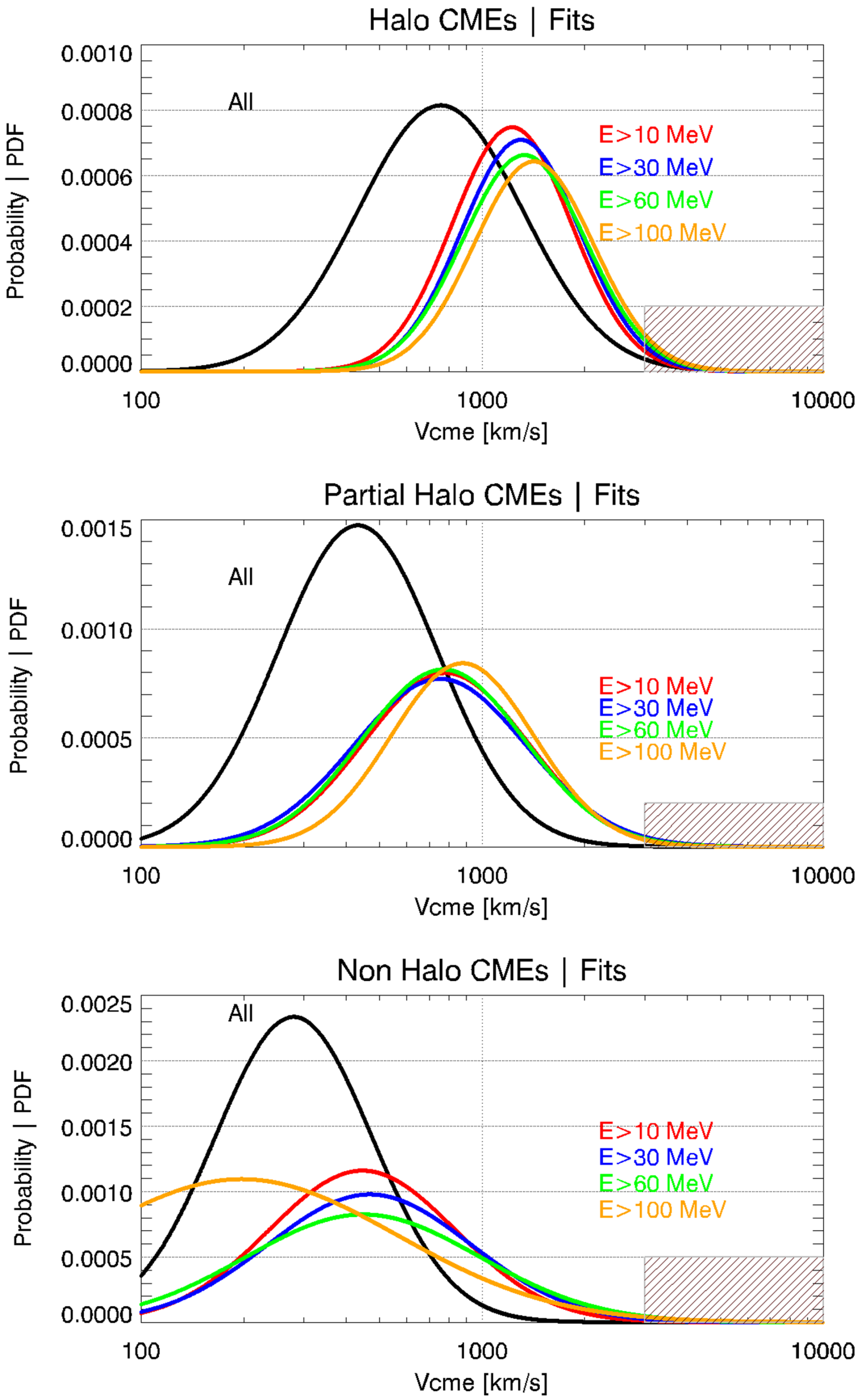

Fig. 5. The Probability Distribution Functions (PDFs) for the case of Halo (top panel), Partial Halo (middle panel) and Non Halo (bottom panel) CMEs, as derived by the CDFs, applying *Method 2*. The black color represents all CMEs in the sample, whereas CMEs associated with SEP events for different integral energies are colour coded. The gray border hatched rectangle area provides the limit for the V_{CME}.

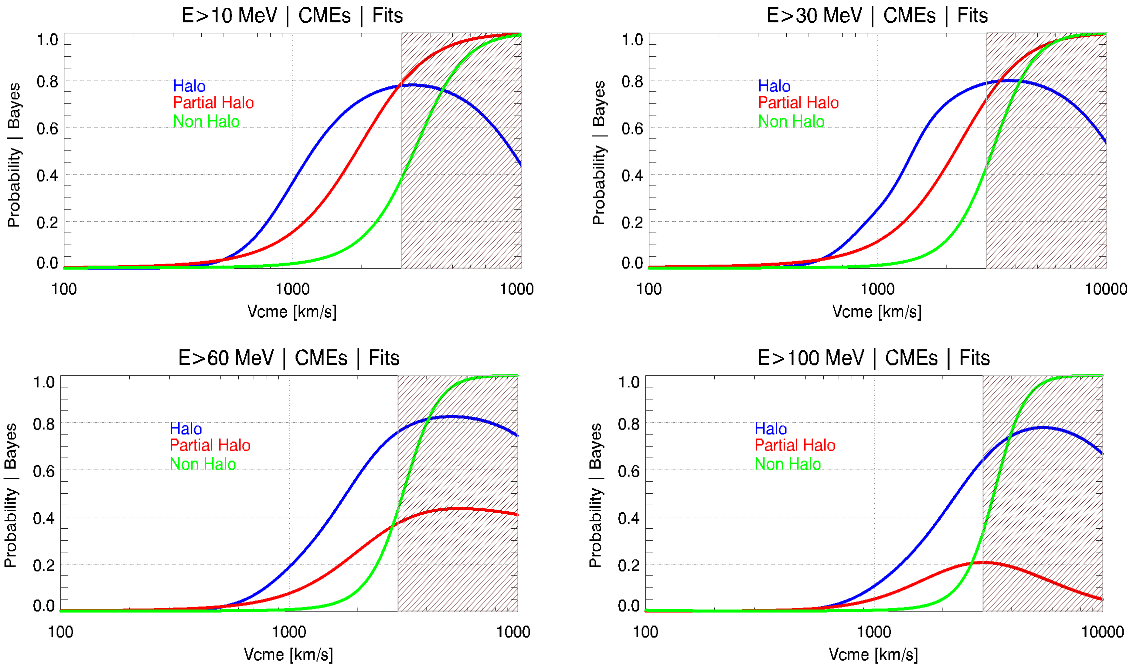

Fig. 6. The curves for the estimation of the probability of SEP occurrence when both the speed (V_{CME}, X-axis) and the width (Halo, Partial Halo and Non-Halo) of a CME are known. Each panel corresponds to an integral energy (i.e. E>10-; >30-; >60- and >100 MeV). Within each panel three fits are presented, one per CME width - colour coded as: Halo (blue), Partial Halo (red), Non Halo (green) CMEs. The gray border hatched rectangle area provides the limit for the V_{CME}. See text for details.

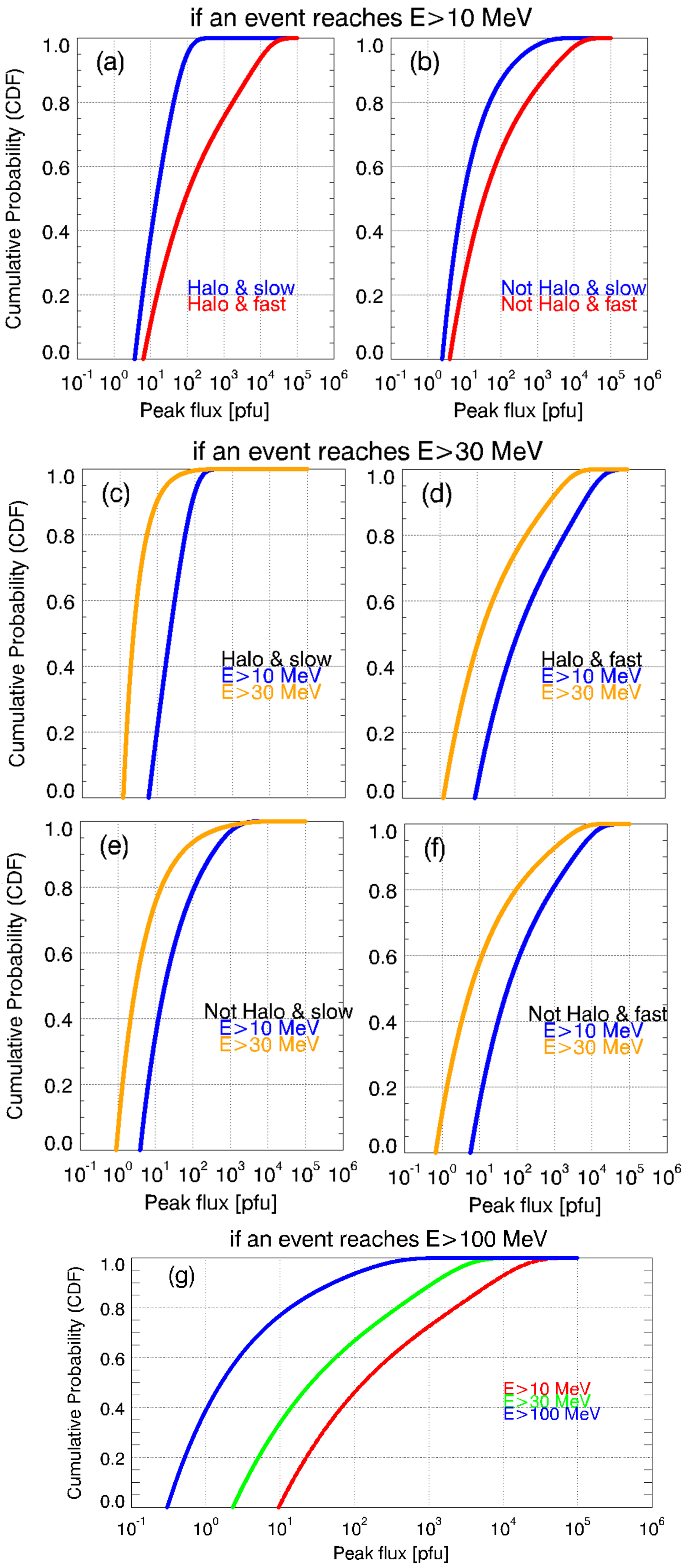

Fig. 7. The CDFs constructed by the database (i.e. all points correspond to actual data) for each integral energy of choice, for the estimation of the peak flux (Peak flux [pfu], X-axis), when both the speed and the width of a CME are known. Lines depict the exponential cut-off fits to the data points per case (from Equation 13). Panels: (a) and (b) correspond to a filtering integral energy at E>10MeV; (c)-(f) to E>30 MeV and (g) to E>100 MeV). See text and Table 3 for details on the bins and the obtained parameters of the fit per case.

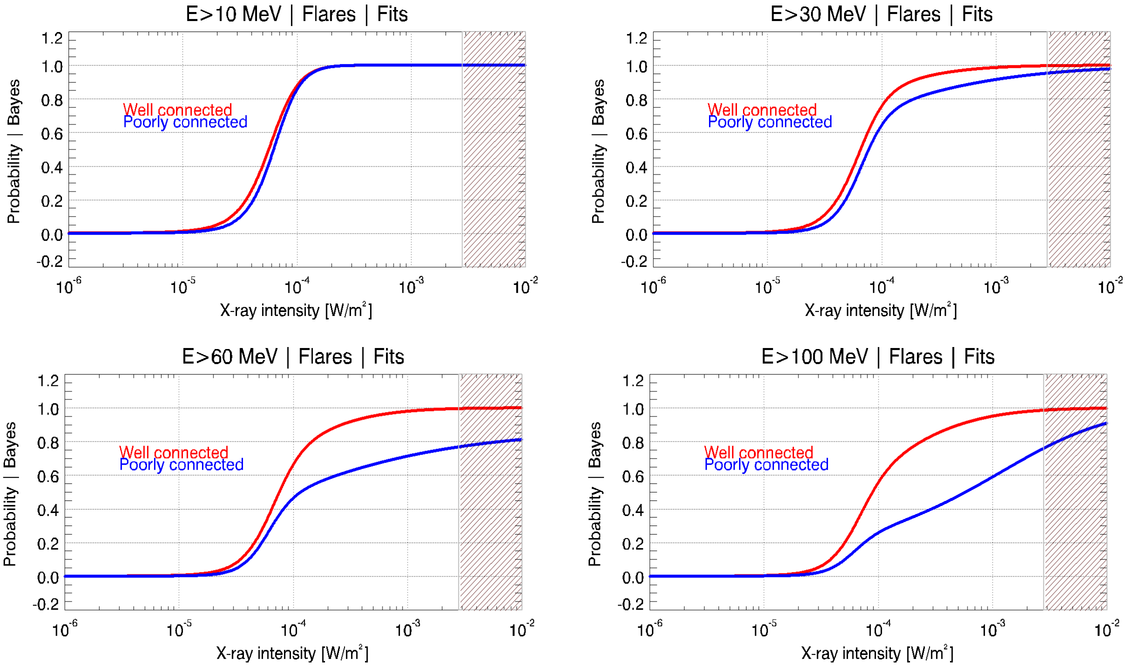

Fig. 8. The curves for the estimation of the probability of SEP occurrence when both the SXR magnitude and the longitude of a solar flare are known. Each panel corresponds to an integral energy (i.e. E>10-;>30-;>60- and >100 MeV). Within each panel two fits are presented, one per longitudinal bin - colour coded as: well connected (red), and poorly connected (blue) solar flares. See text for details. The gray border hatched rectangle area provides the limit for the F_{SXR}. This is similar to Figure 6.

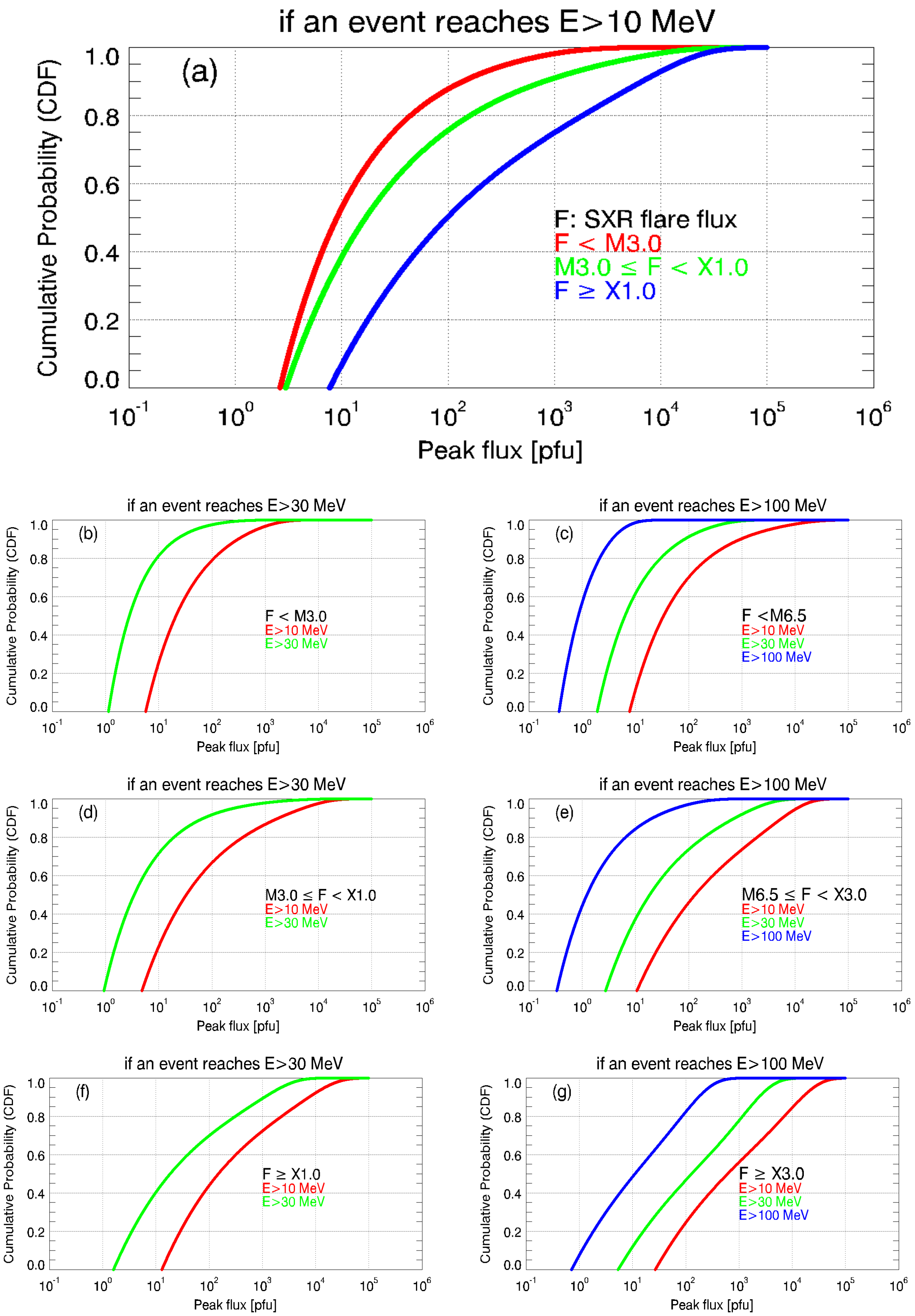

Fig. 9. The CDFs constructed by the database for each integral energy of choice, for the estimation of the peak flux (Peak flux [pfu], X-axis), when the SXR solar flare magnitude is known. Lines depict the exponential cut-off fits to the data points per case. Panel (a) corresponds to a filtering of the integral energy at E>10 MeV. The left hand column (panels (b), (d), (f)) corresponds to a filtering integral energy at E>30 MeV and the right hand column (panels (c), (d), (g)) to a filtering integral energy at E>100 MeV. Black color denotes the SXR flare flux bin in each panel. See text and Table 3 for details on the bins and the obtained parameters of the fit per case.

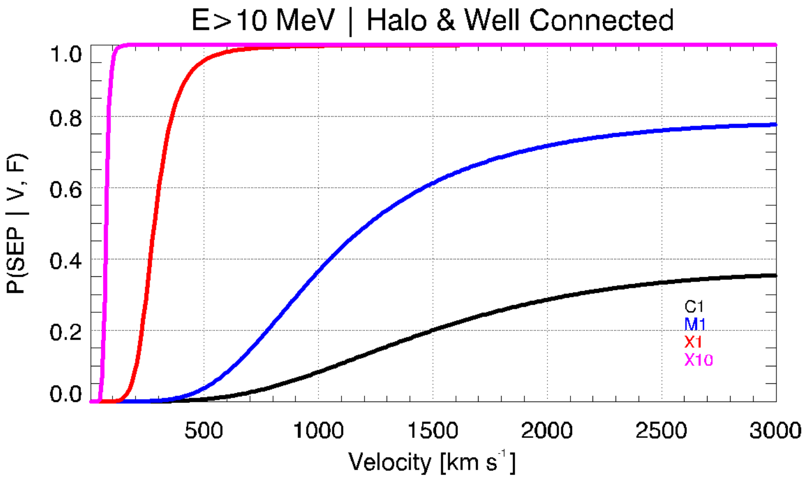

Fig. 10. The curves for the estimation of the probability of SEP occurrence when both a CME and a solar flare have taken place and thus the SXR magnitude (X-axis) and the longitude of a solar flare, as well as, the CME speed and AW are known. The figure presents the output of PROSPER for the case of E>10 MeV with Halo CMEs (AW = 360^o) and well connected flares (lon ≥20^o). Each fit corresponds to a solar flare class, color coded as: C1.0 (black line), M1.0 (blue line), X1.0 (red line) and X10 (magenta line). The X-axis shows the speed of the CME (in km s^{-1}).

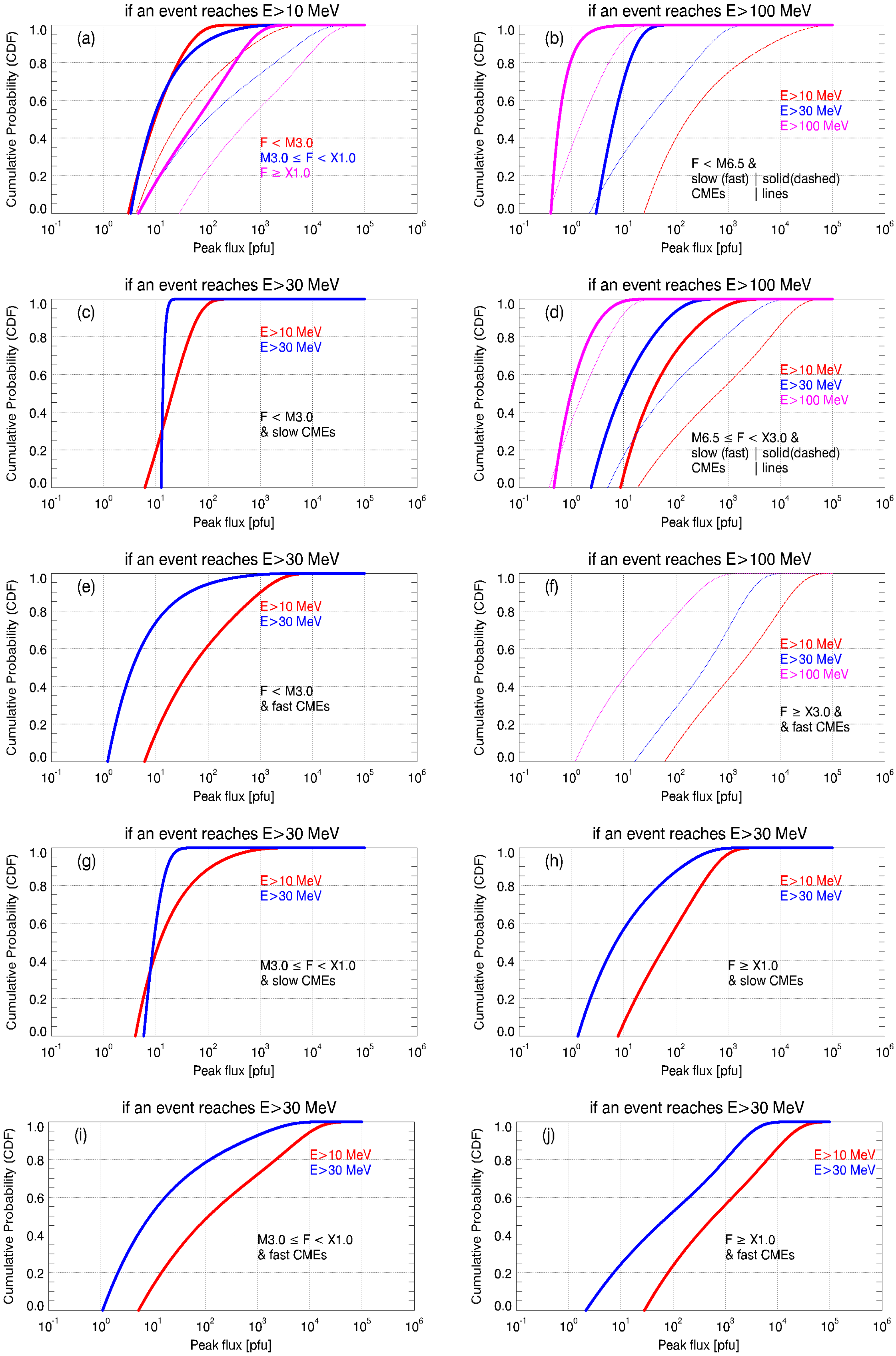

Fig. 11. The CDFs constructed by the database (i.e. all points correspond to actual data) for each integral energy of choice. See text and Table 4 for details on the bins and the obtained parameters of the fit per case. Each panel further displays the selected bins and the displayed energies, if more than one are included in a panel.

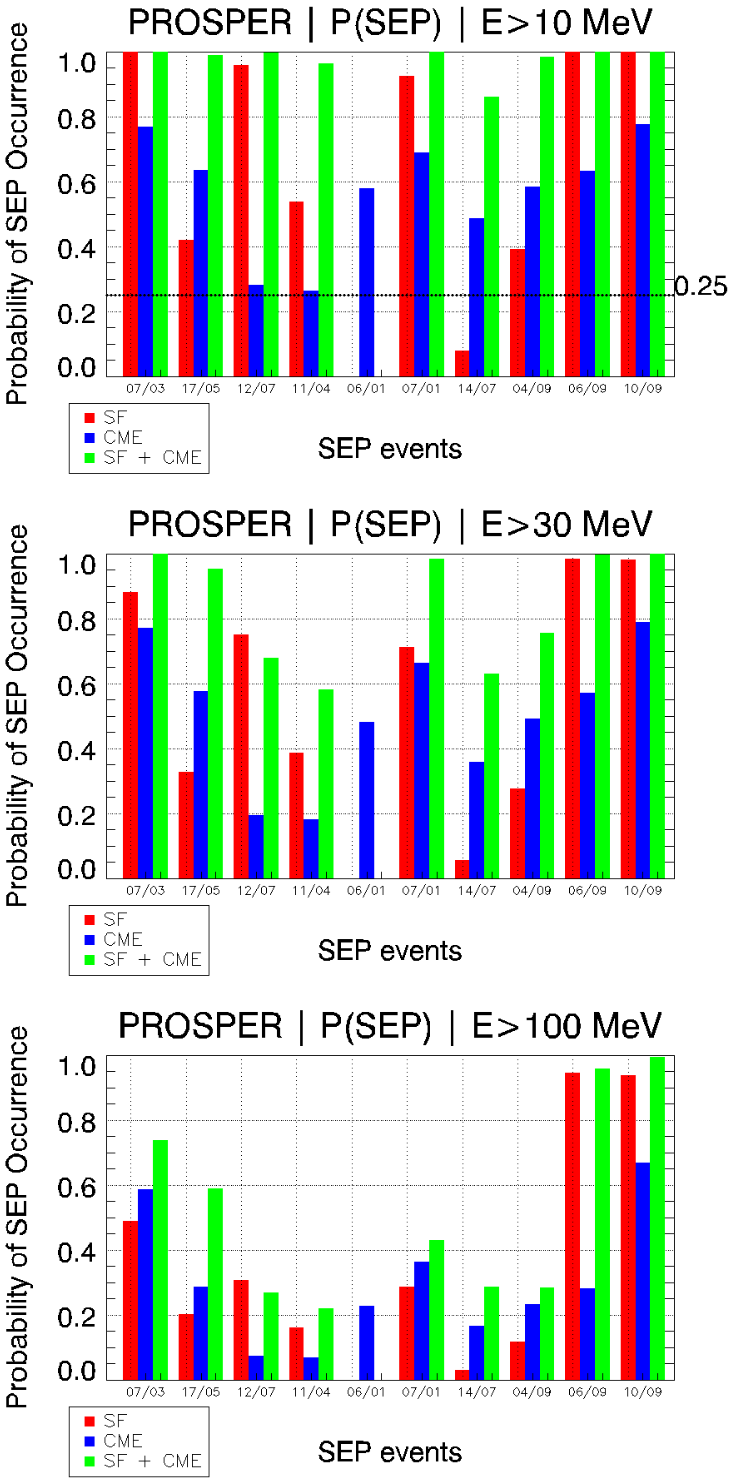

Fig. 12. Comparison of the predicted probabilities of SEP detection per PROSPER's mode of operation for all cases included in Table 5, for each integral energy of interest (i.e. E>10-; E>30-; and E>100 MeV). In each panel, the red histogram represents the outputs of PROSPER based on SF data, the blue histogram represents the outputs of the model based on CME data, and the green histogram represents the outputs of the model based on both SF & CME inputs. In the upper panel the horizontal line represent the threshold above which most SEP prediction modules would issue a notification for a forthcoming SEP event – see details in Table 1 of Anastasiadis et al. (2017).

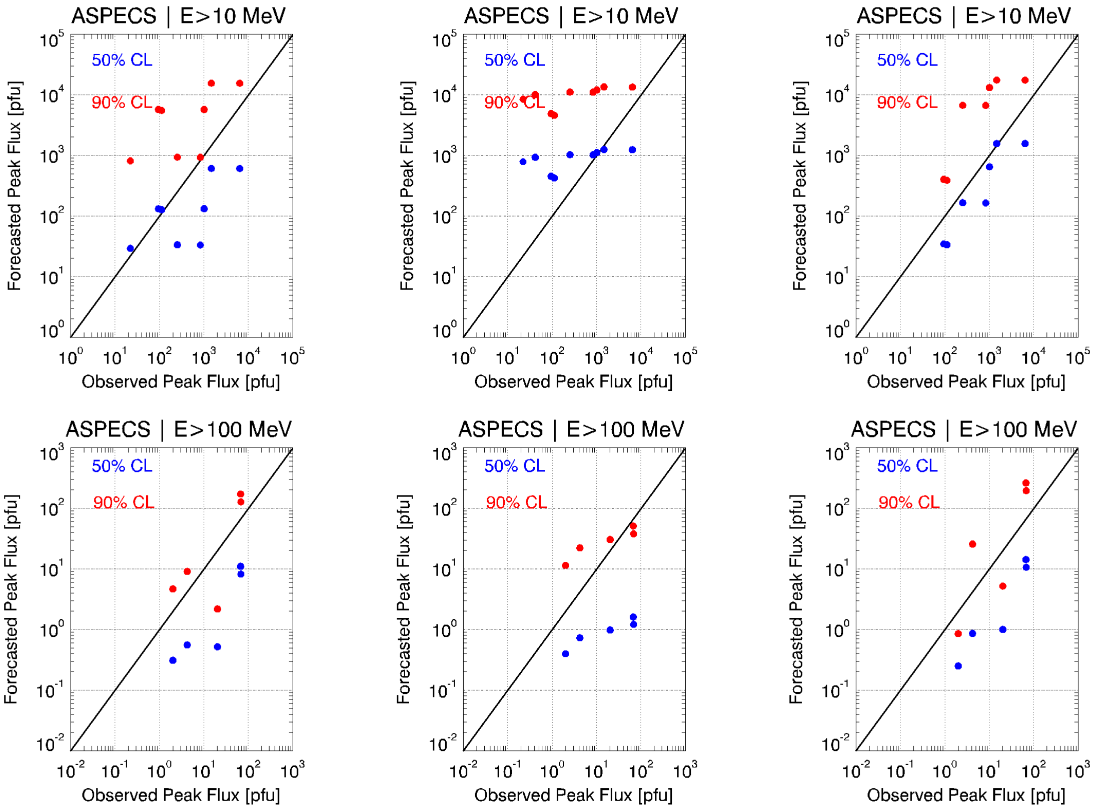

Fig. 13. Comparison of predicted with observed peak flux for the SEP events of Table 5, based on solar flare (column at the left hand side), CME (column in the middle) and solar flare & CME (column at the right hand side) input data. The top row refers to an integral energy of E>10 MeV and the bottom to E>100 MeV.