








Upgrades of the ESPERTA forecast tool for solar proton events

Monica Laurenza^{1,*} , Mirko Stumpo¹ , Pietro Zucca² , Mattia Mancini² , Simone Benella¹ ,
Liam Clark², Tommaso Alberti³ , and Maria Federica Marcucci¹ 

¹ Institute of Space Astrophysics and Planetology – INAF, Via del Fosso del Cavaliere 100, 00133 Roma, Italy

² ASTRON, The Netherlands Institute for Radio Astronomy, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands

³ Istituto Nazionale di Geofisica e Vulcanologia, via di Vigna Murata 605, 00143 Roma, Italy

Received 8 September 2023 / Accepted 14 March 2024

Abstract—The Empirical model for Solar Proton Events Real Time Alert (ESPERTA) exploits three solar parameters (flare longitude, soft X-ray fluence, and radio fluence) to provide a timely prediction for the occurrence of solar proton events (SPEs, i.e., when the >10MeV proton flux is ≥ 10 pfu) after the emission of a $\geq M2$ flare. In addition, it makes a prediction for the most dangerous SPEs for which the >10 MeV proton flux is ≥ 100 pfu. In this paper, we study two different ways to upgrade the ESPERTA model and implement it in real time: 1) by using ground based observations from the LOFAR stations; 2) by applying a novel machine learning algorithm to flare-based parameters to provide early warnings of SPE occurrence together with a fine-tuned radiation storm level. As a last step, we perform a preliminary study using a neural network to forecast the proton flux 1-hour ahead to complement the ESPERTA tool. We evaluate the models over flare and SPE data covering the last two solar cycles and discuss their performance, limits, and advantages.

Keywords: Solar energetic particles / Machine learning

1 Introduction

Solar proton events (SPEs) constitute a hazardous condition in the interplanetary space and near-Earth environment in terms of ionizing radiation that can endanger technology systems and life (e.g. Bothmer et al., 2007; Vainio et al., 2009; Temmer, 2021). For instance, SPEs can disrupt spacecraft operations and instruments on board. In extreme cases, satellites may be rendered useless, memory impacts can cause loss of control, serious noise can affect image data, star trackers may be unable to locate sources, and permanent damage to solar panels is possible. SPEs can also affect the polar ionosphere, causing absorption of high-frequency (HF) radio waves (Hunsucker, 1992; Hargreaves, 2005). This impact can consequently disrupt long-distance radio communications and radar systems. Even a complete blackout of HF communications can occur through the polar regions and position errors can make navigation operations extremely difficult. In addition, SPEs can induce biological effects and pose a major risk to manned space flights. Astronauts on deep space missions and in high-inclination low Earth orbit (LEO) may be exposed to high radiation dose, as well as aircrews and passengers of airlines in polar routes when SPEs reach relativistic energies. Therefore, a warning

system is required to predict SPE occurrence and severity and mitigate the impacts.

Many models have been developed in the past years to forecast SPEs (also termed $\geq S1$ events according to the NOAA scale¹, i.e., when the >10 MeV proton flux exceeds the 10 pfu threshold), both physics-based and empirical (e.g. Whitman et al., 2023, and references therein). Physics-based models, such as SOLPENCO (Aran et al., 2006), SEPSTER (Richardson et al., 2018), and SEPMOD (Luhmann et al., 2007), are based on the modelling of processes of particle acceleration and propagation from the Sun to the interplanetary space in order to predict the properties of the resulting SPE. A big drawback of these models is that they are computationally intensive, and their usage in real-time prediction tools may be limited by the resources available, so they may not be easily put into operation.

On the contrary, empirical models, which include also machine learning techniques, mainly exploit correlations between SPE observations and precursor parameters. Early empirical models to predict SPEs include: the “proton prediction system” PPS76 (Smart & Shea, 1979, 1989; Kahler et al., 2007) based on solar flare parameters (microwave or X-ray flux, flare location); the model (Balch, 1999, 2008) used by NOAA Space Weather Prediction Center, which also exploits metric radio type II and type IV bursts (indicating the presence of a coronal

*Corresponding author: monica.laurenza@inaf.it

¹ <https://www.swpc.noaa.gov/noaa-scales-explanation>.

Table 1. Evolution of the ESPERTA model in terms of the provided forecast (severity of the SPEs and time when the forecast is issued), validation period and type, and scores (probability of detection-POD, false alarm rate-FAR, median, and average (AVG) warning time-WT).

References	Forecast	Forecast time	Validation	Median (AVG) WT	POD	FAR
Laurenza et al. (2009)	\geq S1 SPEs	10 min after \geq M2 flare peak time	1995–2005	55 min*	63%	42%
Alberti et al. (2017)	\geq S1 SPEs	10 min after \geq M2 flare peak time	2006–2014	~2 h (~7 h)	59%	30%
Laurenza et al. (2018)	\geq S1 SPEs	10 min after \geq M2 flare peak time	1995–2014	~4.8 h (~9 h)	63%	38%
Laurenza et al. (2018)	\geq S2 SPEs	S1 threshold crossing	1995–2014	~1.7 h (4 h)	75%	24%
Alberti et al. (2019)	\geq S1 SPEs	10 min after \geq M2 flare peak time	1995–2017	6.7 h (–)	63%	38%
Alberti et al. (2019)	\geq S2 SPEs	S1 threshold crossing	1995–2017	–	75%	23%
Stumpo et al. (2021)	\geq S1 SPEs	10 min after \geq M2 flare peak time	1995–April 2017; stratified cross-validation	–	63%**	42%

* Based on the 31–50 MeV protons onset times published by Posner (2007).

** In Stumpo et al. (2021) a theoretical POD of 76% was computed without including the 21 missed SPEs due to <M2 flares. Here we considered them to obtain the operational POD, which is comparable with the other ESPERTA ones.

mass ejection-CME driven shock) as input parameters. Some empirical models incorporate CME information as input (St. Cyr et al., 2017; Papaioannou et al., 2018, 2022), whereas others, such as REleASE (Posner, 2007; Núñez et al., 2018) and UMASEP (Núñez, 2011), rely on the arrival times of relativistic electrons to predict SPEs in the 30–50 MeV energy range or high-energy protons at 1 AU compared to lower >10 MeV energy protons, respectively. Other machine learning-based approaches have been employed such as decision tree models based on GOES soft X-ray (SXR) and high-energy proton observations (Boubrahimi et al., 2017; Núñez & Paul-Pena, 2020; Lavasa et al., 2021).

A state-of-the-art empirical model is ESPERTA (Empirical model for Solar Proton Events Real Time Alert), developed by Laurenza et al. (2009). ESPERTA is based on the logistic regression analysis on three solar parameters, viz., the flare location, 1–8 Å SXR and 1 MHz Type III fluences, being the SXR fluence a measure of flare size/energy and the Type III solar radio bursts the signatures of fast electron beams streaming outward, along the open magnetic field lines. ESPERTA predictions for \geq S1 events are made 10 min after the peak time of a greater or equal to M2 class flare, in order to maximise the warning time, with a good performance. Table 1 includes all different applications over time of ESPERTA in terms of goals and scores. After the development, ESPERTA was validated on an independent data set (Alberti et al., 2017), reaching scores similar to those obtained on the dataset over which it was developed. Moreover, the ESPERTA model has been adjusted (Laurenza et al., 2018; Alberti et al., 2019) to provide early forecasts of the largest SPEs, for which the >10 MeV proton flux exceeds the 100 pfu threshold at the Earth (such events are defined as \geq S2 events, according to the NOAA scale). It is worthwhile to note that ESPERTA is the only model which has been tested in terms of the event severity, with a median warning time of about 2 h for the prediction of \geq S2 events over an extended data set covering the period 1995–2014. Moreover, the ESPERTA model has been reinterpreted in the framework of machine learning and validated through stratified cross-validation (Stumpo et al., 2021), which is more robust than the evaluations of the model performed previously. The performance obtained was comparable with past results.

Although ESPERTA was designed to run in real-time in Space Weather operations, the Wind/WAVES (Bougeret et al., 1995) radio data necessary to compute the radio fluence

are currently not available in real-time. In this paper, we present upgrades to the ESPERTA model to possibly implement it in real-time operation, by using new radio data from ground-based observatories, and provide also a reliable forecasting of the radiation storm flux, by using two machine learning approaches: a binary classification algorithm and a Long Short Term Memory (LSTM)-based regression approaches.

2 Exploitation of ground-based radio observations

ESPERTA was designed to predict the occurrence of an SPE 10 min after the peak of any flare of class greater or equal to M2. To achieve this, ESPERTA is based on three variables:

1. Helio-longitude of the flare;
2. Time-integrated SXR flux (i.e. SXR fluence);
3. Time-integrated radio flux (i.e. radio fluence).

The first two variables were obtained by using information from Hz flares from the former National Geophysical Data Center as in Laurenza et al. (2009) and the 1 min GOES 1–8 Å flux, respectively. The third variable was computed by using the 1 MHz emission from the WIND/Waves instrument. Here we conducted a study to verify the possibility of replacing the 1 MHz fluence from WIND/Waves, with data recorded with the LOW-Frequency ARray (LOFAR; van Haarlem et al., 2013) telescope. Each LOFAR station has two separate antenna arrays: the low-band antennas (LBA), operating below 80 MHz, and the high-band antennas (HBA), operating above 110 MHz (the FM radio band prohibits observations between 80 and 110 MHz). For this study, we used data from a single station in the LBA range.

Replacing space-based with ground-based data has obvious advantages for operations and real-time forecasting tools. Thus, we have recalculated the probability curves of the ESPERTA model as in Laurenza et al. (2009), by replacing the 1 MHz fluence with 30 MHz fluence. First, we used the flare list for the period 1995–April 2017 from Stumpo et al. (2021), including 989 events (55 of which with data gaps), and extracted 200 events for which we were able to obtain data at 30 MHz. As we do not have LOFAR observations before 2014, we also used

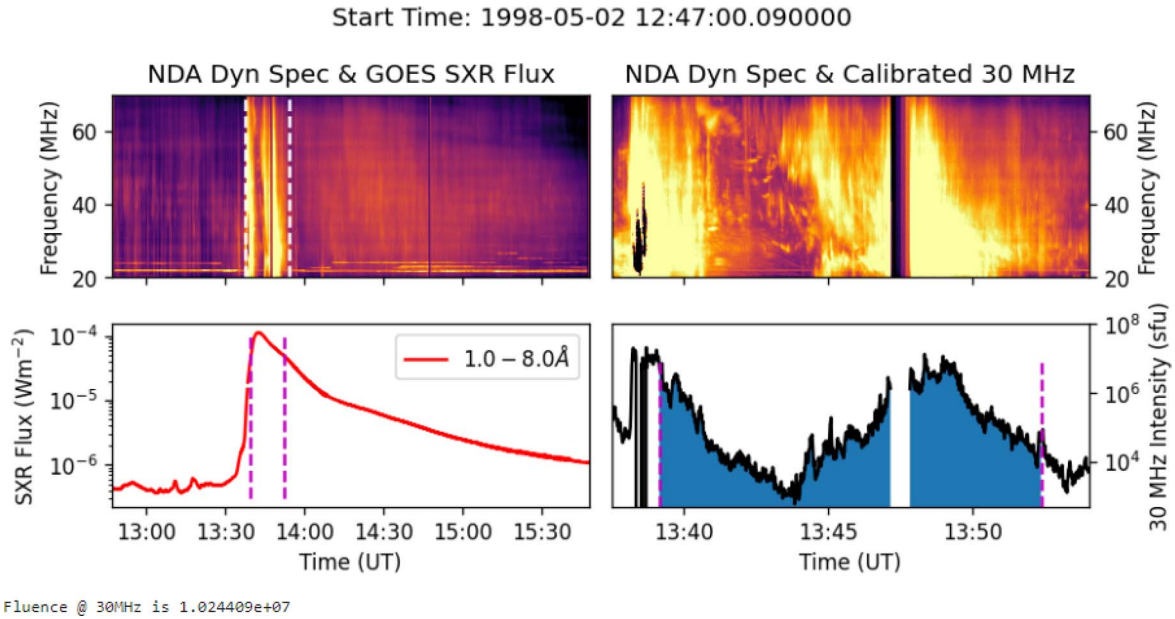


Figure 1. Example event, showing the calibrated dynamic spectrum of the NDA (top left panel) used for one of the 200 selected flares. The radio emission is shown also in detail during the period used for the radio fluence computation (top right panel) as well as the 30 MHz intensity (bottom right panel). The related GOES SXR burst is also shown (bottom-left). The time intervals for the fluence calculation for both SXR and radio data at 30 MHz are marked with a dashed violet line (bottom panels), while the integrated value used to calculate the fluence is marked with a blue area (bottom-right).

the radio data from the Nançay decametric array (NDA, [Lecacheux, 2000](#)) to cover most of these events (when LOFAR data are not available). We remark that both LOFAR LBA and NDA operate in similar frequency ranges, so this joined dataset can be used to test the method, which afterwards could be potentially used with any radio data at the same frequency. Then, we performed data calibration for the emission at 30 MHz and calculated the radio fluence using the method described in [Laurenza et al. \(2009\)](#).

To perform the calibration the NDA is equipped with a radio emitter with a constant duty cycle and a constant emission spectrum offering the possibility to calibrate the flux emission in absolute terms. An example of the fluence for one of the flares in the list is shown in [Figure 1](#). The top panels of [Figure 1](#) show the dynamic spectrum from NDA, including the details of the emission used for the fluence (top-right). The time intervals for the SXR fluence and for the 30 MHz radio flux are shown in the bottom panels. For the limited number of events from 2014 to 2017 we also used the data from LOFAR, and compared it with the NDA data. They resulted in very good agreement, so the whole dataset is homogeneous. The radio fluence at 30 MHz calculated for 200 events of this dataset, was then used to reproduce the scatter plot replacing the 1 MHz fluence used in [Laurenza et al. \(2009\)](#). We produced new probability curves based on the 30 MHz fluence, and recalculated the Probability of Detection (POD) and False Alarm Rate (FAR) for ESPERTA with the new ground-based radio dataset. In particular, we assumed the probability distribution

$$P(\log X, \log R) = \frac{e^{\eta}}{1 + e^{\eta}} \quad (1)$$

where X and R are the variables SXR and radio fluence, respectively and $\eta(\log X, \log R) = \alpha \log X + \beta \log R + \delta \log$

$X \log R$. The parameters α , β and δ have been estimated via a Montecarlo approach from the observed events and the corresponding fluence at 30 MHz. In contrast with [Laurenza et al. \(2018\)](#) and [Laurenza et al. \(2009\)](#) in this study we did not differentiate for the longitudinal bands of the events, because of the limited number of the considered flares. Adding the information on the longitudinal bands would in principle improve the POD and minimize the FAR. In this study however we mainly want to show if the low-frequency ground-based observations may replace the space-based radio observations originally used in the ESPERTA model.

The resulting probability curves using the 30 MHz data are shown in the left panel in [Figure 2](#). A plot with the correlation between 1 MHz and 30 MHz fluence is shown on the right panel in [Figure 2](#). The linear fit is indicated with a solid black line, the dashed lines represent the 1σ fit confidence while the blue shaded area indicates the 2σ confidence.

We selected a probability threshold to discern positive events (hits, false alarms) from negative events (correct nulls, misses) by maximising the Heidke score (hereafter HSS) ([Heidke, 1926](#)). In particular, the possibility of predicting an event by chance is

$$C = \frac{(\text{hits} + \text{misses}) \cdot (\text{hits} + \text{FAs}) + (\text{FAs} + \text{nulls}) \cdot (\text{misses} + \text{nulls})}{\text{SEPs} + \text{nonSEPs}} \quad (2)$$

where hits and misses are the number of successful or unsuccessful predictions, FAs are the false alarms and nulls are the correct rejections. The resulting Heidke score is expressed then as:

$$\text{HSS} = (\text{hits} + \text{nulls} - C) / ((\text{SEPs} + \text{nonSEPs}) - C) \quad (3)$$

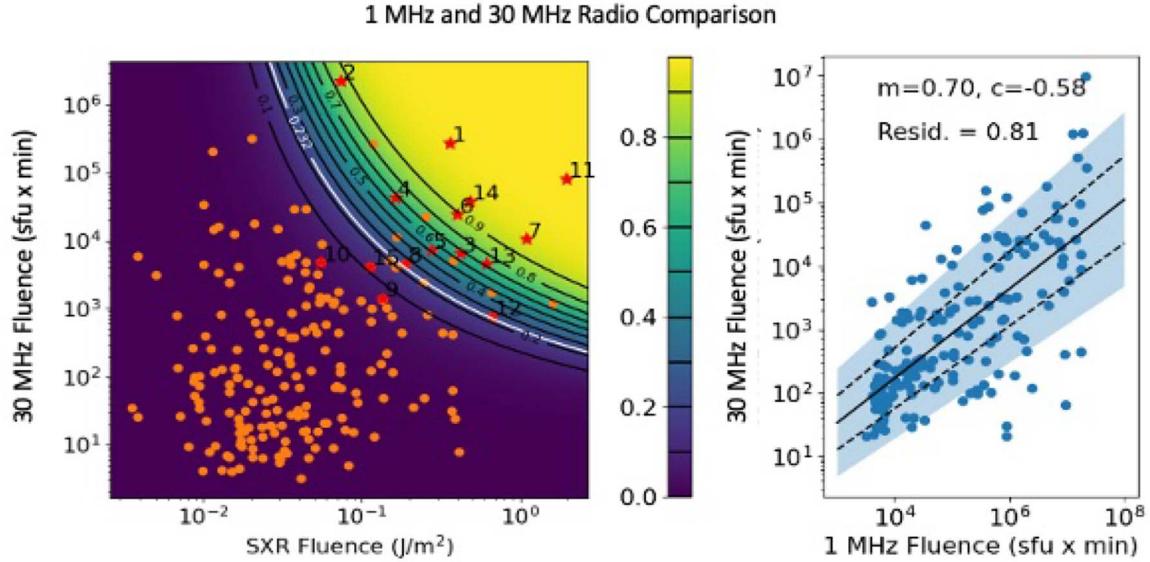


Figure 2. Scatter plot (left) of the 30 MHz fluence vs the SXR one, including also the probability curves (black lines) for ESPERTA recalculated using ground-based radio data at 30 MHz for the 200 flares selected for this study. The probability contour threshold (see the text) is indicated by the white line. Red circles indicate SPE-associated flares, whereas orange circles indicate flares not associated with any SPE. A scatter plot showing the correlation of the radio fluence at 1 MHz and 30 MHz for 200 events of the ESPERTA list extended to 2017 (right). The black dashed lines represent the fit 1σ uncertainty, while the 2σ confidence level is marked with a blue area.

We obtained HSS = 0.70 with a threshold of 0.23, indicated by the white line in the scatter plot on the left panel in Figure 2. A point above this contour is considered to be a hit if a SPE is associated to this flare, otherwise it is considered to be an FA. A point below the contour is considered to be a miss if a SPE is associated with this flare, otherwise, it is considered to be a null. This leads to a POD = 0.69 and FAR = 0.33 for the radio data at 30 MHz. Note that these scores are obtained by considering also the missed SPEs (MISSSES) due to <M2 flares (that the method cannot predict by definition). We estimated the number of MISSSES to be 5, i.e., about 22% of the 21 MISSSES occurred in the period 1995–April 2017 for consistency with the percentage of the 200 analysed flares with respect to the total ones in the same period (934). This allows us to compare the obtained scores with those computed in past studies (see Table 1). We point out that they are computed for the same dataset on which the method has been developed. The obtained scores of POD and FAR are comparable and even slightly better than those using WIND data at 1 MHz (see also next section).

3 ESPERTA model – A supervised learning approach to SPEs forecasting

The ESPERTA model can be reinterpreted in the framework of machine learning (Stumpo et al., 2021; Benella et al., 2023). We used the supervised learning approach to perform the binary classification needed because there are two classes of events, C_0 and C_1 , representing NO-SPEs and SPEs associated flares, respectively. Thus, we defined an input vector \mathbf{x}_i of the three aforementioned ESPERTA features and a binary target variable t_i which can assume the value of one if the input vector is a SPE and 0 in the other case. In this framework, the model is a

function which maps the input vector into the target variable. This function depends on a set of unknown parameters (weights) which can be found by optimizing an error function with respect to a series of examples for which the target variable is given. Once the optimal weights have been learned, the model performance in making predictions is evaluated. The basic idea to do it properly is to split up the original data set into training and testing datasets. The first one is used to learn the best model, while the second one, which is independent from the first, is used to evaluate the performance of the model.

ESPERTA exploits the probabilistic approach by modelling the probability that the i th vector is a SPE with a logistic function. In this case, the unknown parameters are optimized by minimizing the cross-entropy error function:

$$\mathcal{L}(\mathbf{X}|P) = \sum_{i=1} \log P_i + \sum_{i=0} \log(1 - P_i) \quad (4)$$

where

$$P_i(C_1|\mathbf{x}_i, \mathbf{w}) = \frac{1}{1 + e^{-(w_0 + \mathbf{w}^T \mathbf{x}_i)}} \quad (5)$$

analogously to equation (1). Then, to decide whether or not a given vector results in an SPE we define a decision function based on the value of a threshold probability ϵ :

$$t(\mathbf{x}_i) = \begin{cases} 1 & \text{if } P(C_1|\mathbf{x}_i, \mathbf{w}) \geq \epsilon \\ 0 & \text{if } P(C_1|\mathbf{x}_i, \mathbf{w}) < \epsilon \end{cases} \quad (6)$$

Ideally, we can say that it is a SPE if its estimated probability is greater than a threshold of 0.5, but this threshold needs to be reconsidered because the errors in classifying the events may arise essentially from the overlapping of the two classes in the parameter space. This effect is strongly dependent on the degree of unbalancing (see Fig. 1 of Stumpo et al., 2021). To take into

account this effect we have to find the optimal threshold on the probability to decide more accurately the decision boundaries.

Stumpo et al. (2021) computed the validation metrics for ESPERTA over the test set for different values of the threshold (displayed in the top panel of Fig. 3) and obtained the following scores: a critical success index (CSI) of 0.49, a POD of 0.76 and FAR of 0.42. In particular, the CSI metric measures the trade-off between POD and FAR, so they defined the optimal threshold to be one which maximizes the CSI. Note that the above “theoretical” POD was computed without including the 21 missed SPEs due to <M2 flares. If we considered them, we would obtain the “operational” POD = 0.63, as reported in Table 1, which is comparable with the other ESPERTA ones.

3.1 Calibration of the model for predicting \geq S2 SPE events

In order to provide early warnings of SPE occurrence together with an indication of the radiation storm level, we separated the \geq S2 proton events, i.e., those reaching a peak flux of ≥ 100 pfu, over the period 1995–April 2017 (extending the list in Laurenza et al. (2018)). Then we applied the supervised machine learning approach described above by considering the \geq S1 and \geq S2 as two separated classes.

We computed the validation metrics for different threshold probabilities, as shown in the middle panel of Figure 3. Then, we calibrated the model by computing the optimal threshold that maximizes the CSI (yellow line in the middle panel of Fig. 3), which is a trade-off between maximising the POD and minimising the FAR.

As the \geq S2 events are less frequent and we do not have enough statistics, we calibrated the model by using $N - 1$ observations in the training set and 1 event in the test set and repeated this for $N - 1$ times.

The optimization led to the following scores: $f_1 = 0.77$; CSI = 0.63; POD = 0.88; FAR = 0.32, comparable to those obtained by Laurenza et al. (2018). Again, this POD was computed without including the 6 missed \geq S2 SPEs due to <M2 flares. When considering these MISSES, we obtained the “operational” POD = 0.79.

In order to possibly refine the predictions of the radiation storm level, we distinguished also the \geq S3 SPEs, i.e., those reaching a peak flux of ≥ 1000 pfu, over the period 1995–April 2017. Then, we applied the supervised machine learning approach again by considering the \geq S2 and \geq S3 as two separate classes, we computed the metrics (see bottom panel of Fig. 3) and performed a calibration for \geq S3 SPEs. The optimal scores maximising the CSI were found to have the following values: POD = 1.00, FAR = 0.54, $f_1 = 0.63$, CSI = 0.46. Nevertheless, they cannot be considered to be reliable, given the small number of \geq S3 events.

Thus here we propose a forecasting scheme where forecasts of \geq S1 SEP events are given 10 min after the flare peak time, and are followed by forecasts of the ≥ 10 MeV solar proton events with a peak flux ≥ 100 pfu (i.e., \geq S2) at the time when the >10 MeV proton flux reaches the S1 level (i.e., at the end of the third consecutive 5-minute interval with average flux >10 pfu). The median (mean) warning time of \geq S1 forecast is about 6 (8) h (Alberti et al., 2017), whereas the median (mean) for \geq S2 forecast is 1.7 (4) h (Laurenza et al., 2018). Thus, we would be able to predict the occurrence of an SPE soon after

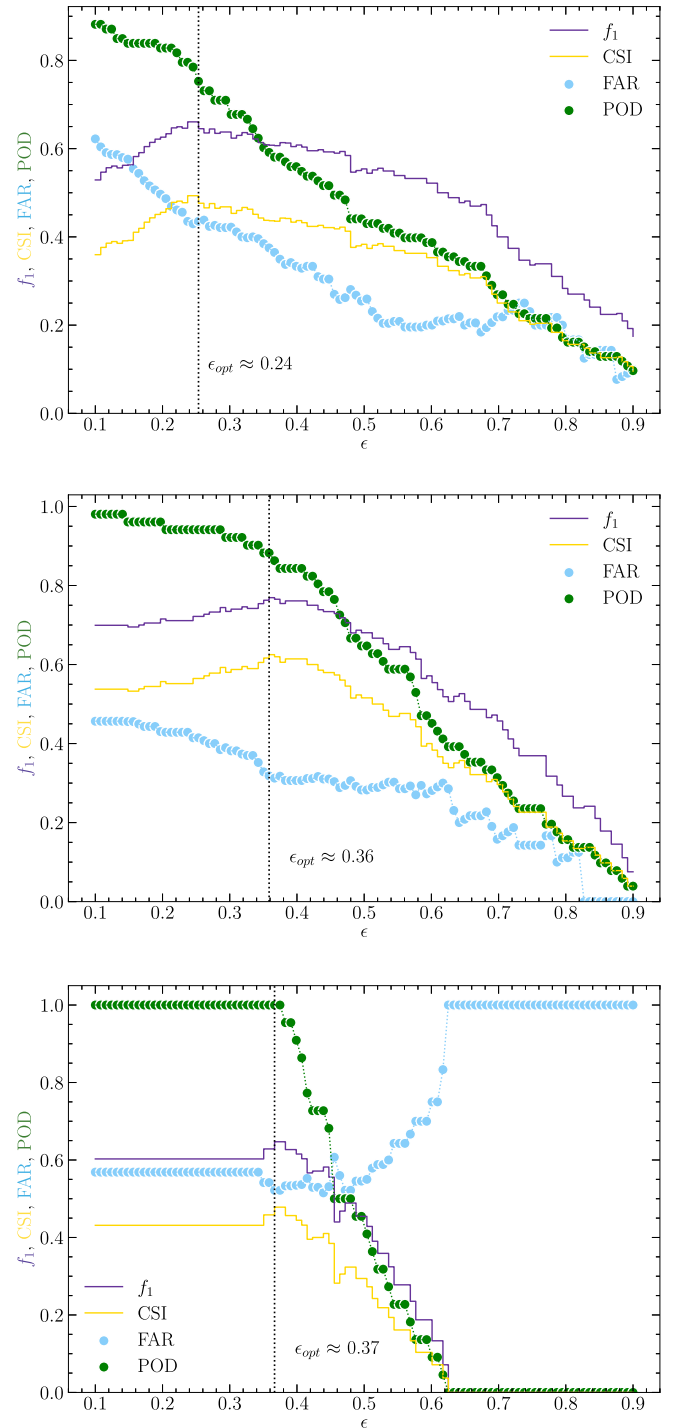


Figure 3. Averaged cross-validated scores with respect to the decision threshold ϵ for forecasting \geq S1 SPEs (top), \geq S2 SPEs (middle), and \geq S3 SPEs (bottom).

the flare and, if it actually occurs, we could give an indication that it is going to reach or surpass the S2 level in about 2 h.

In order to implement this approach in real time we can use the aforementioned LOFAR data, as currently LOFAR is observing with a dedicated station a calibrated solar radio spectrum (see Sect. 5). Another possibility is to exploit the STEREO-A beacon data, as the spacecraft is currently close

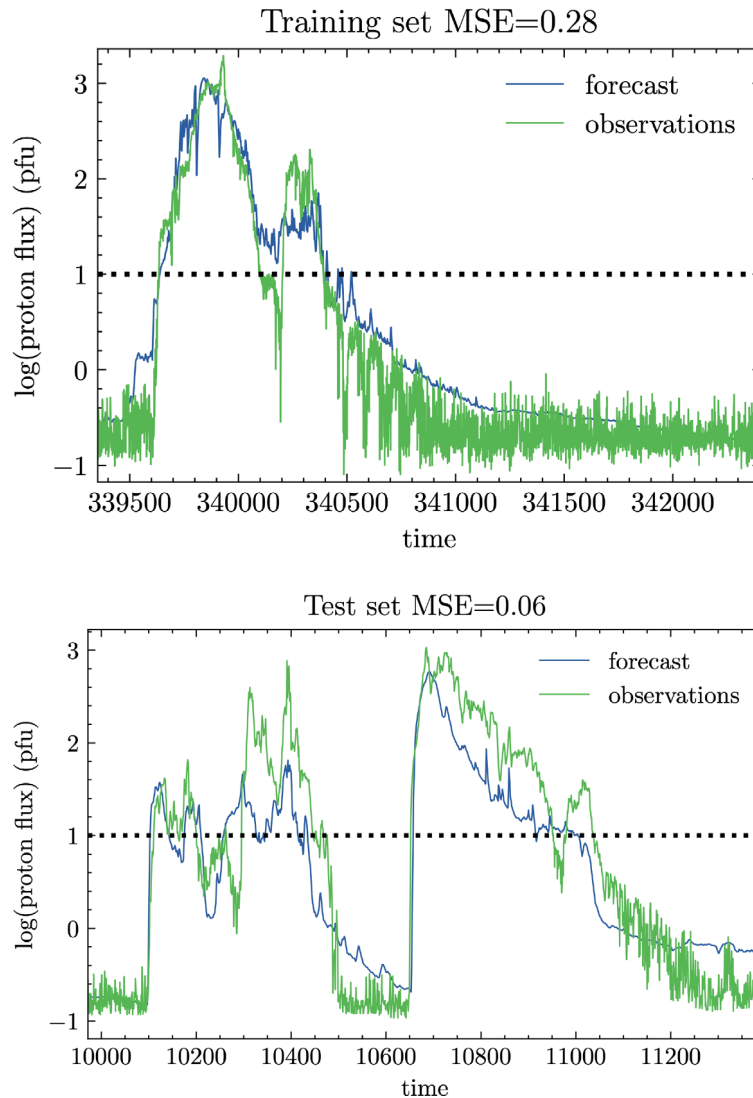


Figure 4. Predictions (blue line) of the proton flux by employing the LSTM compared with observations (green line) for one event in the training set (top) and one in the test set (bottom). The predicted profiles are shifted 1 h ahead. The RMSE is given by the square root of the MSE.

to the Earth furnishing real-time beacon data which are similar to the WIND/WAVES ones. Nevertheless, this aspect will be investigated deeper in the near future.

4 Long short-term memory

As detailed in [Stumpo et al. \(2021\)](#), the major challenge for the design of classification algorithms based on flare precursors, such as ESPERTA, is the optimization of the FAR. This problem is due to the fact that the distributions of the features remain overlapped. The overlap creates a region of uncertainty, which is all the more important the more unbalanced the classes are. Furthermore, the classification does not provide quantitative information on the peak of the proton flux. As we showed previously, in order to adapt a classification algorithm to give a prediction of SPE severity, it is necessary to introduce three- or four-class problems. In that case, some classes can have quite

poor statistics, such as the one of $\geq S3$ SPEs and, as a consequence, the relative unbalancing is greater. For this reason, creating a classification model for the prediction of the $\geq S3$ SPE severity is unreliable.

Thus, we explored an alternative approach that allows for forecasting the whole SPE profile through the LSTM networks, which are a special kind of recurrent neural networks, capable of learning long-term dependencies ([Hochreiter & Schmidhuber, 1997](#)).

The main advantage is that we have direct access to the predicted time series at time $t + n$, which is extremely helpful for real-time monitoring and space weather operations. The idea of the LSTM is to use past data of the input time series say at time $t - 1, t - 2, \dots, t - m$, in order to predict the value assumed by the proton flux at time $t + n$, where m and n are the lookback and lookforward, respectively. The network depends on a series of parameters w (weights) which weight the importance of each point in the past until $t - m$, in order

Table 2. Operational probability of detection-POD and theoretical false alarm rate-FAR computed for the different developed models.

Model	Forecast	Validation	POD	FAR
ESPERTA-based with 30 Mhz fluence	$\geq S1$	1995–April 2017	69%	33%
LSTM-based	$\geq S1$	2012–April 2017	67%	10%
ESPERTA-based with 1 Mhz fluence	$\geq S2$	1995–April 2017	79%	32%

to predict the value assumed at $t + n$. The weights w are initialized and optimized with respect to the mean squared error function computed over a series of examples (learning phase). As input time series we could not use two out of three of the ESPERTA parameters, i.e., the heliologitude as it is not a time series and the radio fluence, because of the unavailability of WIND/WAVES data in real-time and the limited number of ground-based observations. Thus we introduced the quasi-relativistic electron flux as a feature. In particular, we used the 1 min electron flux data in the energy range 0.25–0.7 MeV obtained from COSTEP/EPHIN experiment onboard SOHO from 1996 to 2019 for 103 SPEs. It is worthwhile to note that the quasi-relativistic electrons arrive on the Earth about 1 h before the >30 MeV protons and are already successfully exploited in the HESPERIA RELeASE forecasting technique (Posner, 2007; Malandraki & Crosby, 2018).

We trained the model on the dataset from 1996-01-01 to 2011-12-31 and tested it on the dataset from 2012-01-01 to 2019-12-31. We evaluated the performance by computing the root mean squared error (RMSE) over the test dataset. We obtained an RMSE of 0.24 pfu, which is referred to the logarithm of the proton flux.

Figure 4 shows an example of the comparison between observations (green line) and forecast (blue line) for an event in the training dataset and one in the testing dataset (top and bottom panels, respectively). It can be seen that both the profiles and the onset time are quite well reproduced. The predicted profiles are provided one hour in advance of the actual ones. As far as the SEP intensity is concerned, the agreement is generally good and in any case, the discrepancy remains on average bounded within the RMSE. We emphasise that the LSTM algorithm also allows for a dynamic forecast of the SPE features (e.g. onset time, peak flux, and peak time) up to 1-hour in advance. From the LSTM-predicted proton flux we identified how many events could be predicted and performed a cross-comparison with ESPERTA over the same testing period January 2012–April 2017 (for a total of 27 SPEs). We obtained the following results: POD = 59% (16/27) and FAR = 16% (3/19) for ESPERTA, whereas POD = 67% (18/27) and FAR = 10% (2/20) for the LSTM. The LSTM results are better in terms of classification, although with a lower warning time (up to 1 h vs 6.7 h median warning time of ESPERTA).

5 Conclusions

Our study aimed at advancing the SPE forecasting capabilities (Laurenza et al., 2023) based on the well-proven concept of the ESPERTA model (Laurenza et al., 2009, 2018) and implement a real-time tool. Our findings can be summarised as follows.

1. Ground-based radio observations represent a viable option to provide real-time alerts.
2. The performance of the developed models for the $\geq S1$ prediction is quite good and in agreement with past results. Table 2 summarizes the obtained scores. We were able to estimate an operational POD. On the contrary, the computed FAR can be considered to be theoretical, as it strongly depends on the imbalance of the dataset (see Stumpo et al., 2021, for a thorough analysis).
3. The machine learning-based ESPERTA modified model is the only validated forecasting technique predicting the occurrence of the SPEs producing radiation storm level $\geq S2$. The performance of the method is high with the following validation metrics: theoretical (operational) POD = 88% (79%), FAR = 32% with a good median warning time (~ 2 h). The estimation of the warning time remains theoretical as it depends on the data availability in real time.

These findings are very promising for the implementation of an ESPERTA-based tool with advanced prediction capabilities for $\geq S2$ SPEs by using ground-based radio data. One possible approach to implement the ESPERTA tool in real-time is to use LOFAR data. In particular, as the new project IDOLS (Incremental development of LOFAR Space-weather) is providing real time calibrated radio data using one dedicated LOFAR station (see e.g. <https://spaceweather.astron.nl/SolarKSP/data/website/>), the ESPERTA model may be used at least during the European day time in real-time operations in the near future. One limitation of this approach is however the limited observing time resulting from a single LOFAR station. Other radio observations in other parts of the world would be needed to achieve 24-hour coverage. This study wants to show how any low-frequency radio dataset properly calibrated, would be able to be integrated into the ESPERTA tool, providing that the probability curves are calculated for the specific frequency used in the observations.

As a last step, we provided the first preliminary results about the forecasting of the SPE flux profile 1 h in advance using an LSTM-based neural network. Moreover, the comparison performed between ESPERTA and the LSTM over the same time span (January 2012–April 2017), shows that the LSTM has better performance, but lower warning time. Thus, the ESPERTA model could be complemented with the prediction of the flux profile through the LSTM neural network when exploiting the additional feature of quasi-relativistic electrons. We envisage an operational model consisting of different modules so that first ESPERTA would issue a forecast about the $\geq S1$ SPE occurrence 10 min after a $\geq M2$ flare. Such prediction would be likely followed (because of the lower warning time) by: 1) the LSTM dynamic forecast, which includes the forecast of the SPE features (e.g. the peak flux), and 2) the

$\geq S2$ SPE ESPERTA forecast presumably later than the LSTM one. Note that the radio data at 30 MHz would be available within the 10-minute forecasting timeframe of the first module as they are recorded in real-time with a cadence of 4 data points per second, although only during daytime in Europe. Of course, a decision rule would have to be established in case of different predictions of the different modules. However, a deeper investigation is necessary in the future, especially to assess the LSTM-based operational tool.

Acknowledgements

This research has been carried out in the framework of the CAESAR project, supported by the Italian Space Agency and the National Institute of Astrophysics through the ASI-INAF n. 2020-35-HH.0 agreement for the development of the ASPIS prototype of the scientific data centre for Space Weather. All data used in this analysis are publicly accessible from NASA (Wind/WAVES) and NOAA (GOES). The Wind/WAVES data were obtained from <ftp://solar-radio.gsfc.nasa.gov/>. GOES data have been downloaded from the <http://www.ngdc.noaa.gov/stp/satellite/goes/dataaccess.html> repository. The editor thanks Marlon Nunez and an anonymous reviewer for their assistance in evaluating this paper.

References

- Alberti T, Laurenza M, Cliver EW. 2019. Forecasting solar proton events by using the ESPERTA model. *Nuovo Cimento C Geophys Space Phys C* **42** (1): 40. <https://doi.org/10.1393/nccf/2019-19040-y>.
- Alberti T, Laurenza M, Cliver EW, Storini M, Consolini G, Lepreti F. 2017. Solar activity from 2006 to 2014 and short-term forecasts of solar proton events using the ESPERTA model. *Astrophys J* **838**(1): 59. <https://doi.org/10.3847/1538-4357/aa5cb8>.
- Aran A, Sanahuja B, Lario D. 2006. SOLPENCO: A solar particle engineering code. *Adv Space Res* **37**(6): 1240–1246. <https://doi.org/10.1016/j.asr.2005.09.019>.
- Balch CC. 1999. SEC proton prediction model: verification and analysis. *Radiat Meas* **30**(3): 231–250. [https://doi.org/10.1016/S1350-4487\(99\)00052-9](https://doi.org/10.1016/S1350-4487(99)00052-9).
- Balch CC. 2008. Updated verification of the space weather prediction center's solar energetic particle prediction model. *Space Weather* **6**(1): S01001. <https://doi.org/10.1029/2007SW000337>.
- Benella S, Stumpo M, Laurenza M, Alberti T, Consolini G, Marcucci MF. 2023. Statistical treatment of solar energetic particle forecasting through supervised learning approaches. In: *Proceedings of Science, 27th European Cosmic Ray Symposium*. p. 14. <https://doi.org/10.22323/1.423.0014>.
- Bothmer V, Daglis IA, Bogdan TJ. 2007. Space weather: Physics and effects. *Phys Today* **60**(12): 59. <https://doi.org/10.1063/1.2825074>.
- Boubrahimi SF, Aydin B, Martens P, Angryk R. 2017. On the prediction of >100 MeV solar energetic particle events using GOES satellite data. In: *2017 IEEE International Conference on Big Data (Big Data)*. IEEE. pp. 2533–2542. <https://doi.org/10.1109/BigData.2017.8258212>.
- Bougeret JL, Kaiser ML, Kellogg PJ, Manning R, Goetz K, et al. 1995. Waves: The radio and plasma wave investigation on the wind spacecraft. *Space Sci Rev* **71**(1–4): 231–263. <https://doi.org/10.1007/BF00751331>.
- Hargreaves JK. 2005. A new method of studying the relation between ionization rates and radio-wave absorption in polar-cap absorption events. *Ann Geophys* **23**(2): 359–369. <https://doi.org/10.5194/angeo-23-359-2005>.
- Heidke P. 1926. Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst. *Geogra Ann* **8**(4): 301–349. <https://doi.org/10.2307/519729>.
- Hochreiter S, Schmidhuber J. 1997. Long short-term memory. *Neural Comput* **9**(8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Hunsucker RD. 1992. Auroral and polar-cap ionospheric effects on radio propagation. *IEEE Trans Antennas Propag* **40**(7): 818–828. <https://doi.org/10.1109/8.155747>.
- Kahler S, Cliver E, Ling A. 2007. Validating the proton prediction system (PPS). *J Atmos Sol Terr Phys* **69**(1–2): 43–49. <https://doi.org/10.1016/j.jastp.2006.06.009>.
- Laurenza M, Alberti T, Cliver E. 2018. A short-term ESPERTA-based forecast tool for moderate-to-extreme solar proton events. *Astrophys J* **857** (2): 107. <https://doi.org/10.3847/1538-4357/aab712>.
- Laurenza M, Cliver E, Hewitt J, Storini M, Ling A, Balch C, Kaiser M. 2009. 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). <https://doi.org/10.1029/2007SW000379>.
- Laurenza M, Del Moro D, Alberti T, Battiston R, Benella S, et al. 2023. The CAESAR project for the ASI space weather infrastructure. *Remote Sens* **15** (2): 346. <https://doi.org/10.3390/rs15020346>.
- Lavasa E, Giannopoulos G, Papaioannou A, Anastasiadis A, Daglis IA, Aran A, Pacheco D, Sanahuja B. 2021. Assessing the predictability of solar energetic particles with the use of machine learning techniques. *Sol Phys* **296**(7): 107. <https://doi.org/10.1007/s11207-021-01837-x>.
- Lecacheux A. 2000. The Nançay decameter array: A useful step towards giant, new generation radio telescopes for long wavelength radio astronomy. *Geophys Monogr Ser* **119**: 321. <https://doi.org/10.1029/GM119p0321>.
- Luhmann JG, Ledvina SA, Krauss-Varban D, Odstroil D, Riley P. 2007. A heliospheric simulation-based approach to SEP source and transport modeling. *Adv Space Res* **40**(3): 295–303. <https://doi.org/10.1016/j.asr.2007.03.089>.
- Malandraki OE, Crosby NB. 2018. The HESPERIA HORIZON 2020 project and book on solar particle radiation storms forecasting and analysis. *Space Weather* **16**(6): 591–592. <https://doi.org/10.1029/2018SW001950>.
- Núñez M. 2011. Predicting solar energetic proton events (E>10 MeV). *Space Weather* **9**(7): S07003. <https://doi.org/10.1029/2010SW000640>.
- Núñez M, Paul-Pena D. 2020. Predicting >10 MeV SEP events from solar flare and radio burst data. *Universe* **6**(10): 161. <https://doi.org/10.3390/universe6100161>.
- Núñez M, Klein K-L, Heber B, Malandraki OE, Zucca P, Labrens J, Reyes-Santiago P, Kuehl P, Pavlos E. 2018. HESPERIA forecasting tools: Real-time and post-event. In: *Solar particle radiation storms forecasting and analysis*. Springer, Cham. pp. 113–131. https://doi.org/10.1007/978-3-319-60051-2_7.
- Papaioannou A, Anastasiadis A, Sandberg I, Jiggins P. 2018. Nowcasting of solar energetic particle events using near real-time coronal mass ejection characteristics in the framework of the FORSPEF tool. *J Space Weather Space Clim* **8**: A37. <https://doi.org/10.1051/swsc/2018024>.
- Papaioannou A, Vainio R, Raukunen O, Jiggins P, Aran A, Dierckxens M, Mallios SA, Paassilta M, Anastasiadis A. 2022. The probabilistic solar particle event forecasting (PROSPER) model. *J Space Weather Space Clim* **12**: 24. <https://doi.org/10.1051/swsc/2022019>.
- Posner A. 2007. Up to 1-hour forecasting of radiation hazards from solar energetic ion events with relativistic electrons. *Space Weather* **5**(5). <https://doi.org/10.1029/2006SW000268>.
- Richardson IG, Mays ML, Thompson BJ. 2018. 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. <https://doi.org/10.1029/2018SW002032>.
- Smart DF, Shea MA. 1979. PPS76: A computerized event mode solar proton forecasting technique. In: Vol. 1 of *NOAA Solar-Terrestrial Predictions Proceedings*. Donnelly RF (Ed.) National Oceanic and Atmospheric Administration Environmental Research Laboratories, Boulder CO. pp. 406–427.
- Smart D, Shea M. 1989. PPS-87: A new event oriented solar proton prediction model. *Adv Space Res* **9**(10): 281–284. [https://doi.org/10.1016/0273-1177\(89\)90450-X](https://doi.org/10.1016/0273-1177(89)90450-X).
- St Cyr OC, Posner A, Burkepille JT. 2017. Solar energetic particle warnings from a coronagraph. *Space Weather* **15**(1): 240–257. <https://doi.org/10.1002/2016SW001545>.
- Stumpo M, Benella S, Laurenza M, Alberti T, Consolini G, Marcucci MF. 2021. Open issues in statistical forecasting of solar proton events: a machine learning perspective. *Space Weather* **19**(10): e2021SW002794. <https://doi.org/10.1029/2021SW002794>.
- Temmer M. 2021. Space weather: the solar perspective. *Living Rev Sol Phys* **18**(1): 4. <https://doi.org/10.1007/s41116-021-00030-3>.
- Vainio R, Desorgher L, Heynderickx D, Storini M, Flückiger E, et al. 2009. Dynamics of the Earth's Particle Radiation Environment. *Space Sci Rev* **147**(3–4): 187–231. <https://doi.org/10.1007/s11214-009-9496-7>.

van Haarlem MP, Wise MW, Gunst AW, Heald G, McKean JP, et al. 2013. LOFAR: The LOw-Frequency ARray. *A&A* **556**: A2. <https://doi.org/10.1051/0004-6361/201220873>.

Whitman K, Egeland R, Richardson IG, Allison C, Quinn P, et al. 2023. Review of solar energetic particle models. *Adv Space Res* **72(12)**: 5161–5242.

Cite this article as: Laurenza M, Stumpo M, Zucca P, Mancini M, Benella S, et al. 2024. Upgrades of the ESPERTA forecast tool for solar proton events. *J. Space Weather Space Clim.* **14**, 8. <https://doi.org/10.1051/swsc/2024007>.