PROCEEDINGS OF SCIENCE

Statistical treatment of solar energetic particle forecasting through supervised learning approaches

Simone Benella,[a,*] **Mirko Stumpo,**[b,a] **Monica Laurenza,**[a] **Tommaso Alberti,**[a] **Giuseppe Consolini**[a] **and Maria Federica Marcucci**[a]

[a] *INAF - Institute for Space Astrophysics and Planetology, Via del Fosso del Cavaliere 100, I-00133, Rome, Italy*

[b] *Department od Physics, University of Rome Tor Vergata, Via della Ricerca Scientifica 1, I-00133, Rome, Italy*

E-mail: simone.benella@inaf.it

Solar energetic particles (SEPs) represent one of the most hazardous events in space weather. In the last decades, a great variety of techniques have been developed for the prediction of SEP occurrence, mainly based on the statistical association between the >10 MeV proton flux and some precursors (e.g., solar flares, coronal mass ejections, etc.). In this paper we focus on the Empirical model for Solar Proton Event Real Time Alert (ESPERTA), a model which makes a prediction for an SEP event after the occurrence of a ≥M2 solar flare by considering three input parameters: the flare source region longitude, the soft X-ray fluence and the radio fluence at ∼1 MHz. Here, we recast the ESPERTA model in the supervised learning framework and we perform the cross validation of the predictive model also applying rare event corrections (i.e., data oversampling and loss function weighting) because of the highly unbalanced nature of the SEP occurrence. The best performances are obtained by using the Synthetic Minority Oversampling Technique, leading to a probability of detection of 0.83 and a false alarm rate (FAR) of 0.39. Nevertheless, the improvement of the validation scores with respect to the unbalanced case is small. A relevant FAR on the SEP prediction comes as a natural consequence of the sample base rates. In summary we give evidence that the statistical approach to the forecasting of SEP events should take into account the following considerations: 1) the model need to be calibrated with respect to the expected occurrence of SEP events, 2) the decision threshold strongly affects model performance and 3) the features used in the model, when taken individually, are unable to fully separate the classes of events in the parameter space, thus the use of techniques for handling unbalanced problems does not guarantee a better performance.

27th European Cosmic Ray Symposium,
25-29 July 2022
Van der Valk Hotel Nijmegen Lent Hertog Eduardplein 4, 6663 AN Nijmegen, The Netherlands.

*Speaker

PoS(ECRS)014

1. Introduction

An early definition of *Space Weather* provided by the U.S. National Space Weather Plan (created in 1995, reviewed and redefined in March 2019) describes it as the set of «conditions on the Sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health». One of the most important hazards for human activities and health are represented by solar energetic particle (SEP) events. SEPs mainly consist of high energy protons, electrons and heavier ions sporadically emitted by the Sun in association with solar transient events such as solar flares or coronal mass ejections (CMEs) [1–5]. SEPs are detectable as sudden increases of the particle intensity mostly in space-based observations. Nevertheless, during the most severe events, solar particles can reach relativistic energies producing ground level enhancements (GLEs), which can be observed by the ground-based neutron monitors. Radiation storms may have an important impact on human activities and health. For instance, during moderate and severe storms, data from operational instruments can be contaminated by noise associated with high-energy particles. Moreover, SEPs can be at the origin of damages to spacecraft operations and electronics, radio communication issues due to ionization of the atmosphere, and strong hazard also for human health in terms of radiation dose for astronauts (e.g., during International Space Station operations) and/or aircraft crews on polar routes during GLEs [6–9]. For all those reasons, reliable SEP prediction models integrated within an effective alert system are of fundamental importance in the context of *Space Weather*. Many forecasting models have been developed in the last two decades and they can be divided in three classes: *physics-based* models, *empirical* models and *machine learning* (ML) models [10]. Naturally, all these approaches are constrained by the short Sun to Earth transit time of protons in the 10 MeV – 20 GeV energy range, which generally goes from 15 min up to few hours.

According to Whitman *et al.* [10], physics-based models include particle acceleration and transport processes on the Sun and in the interplanetary space in order to predict the main features of an SEP event (e.g. the arrival time, the temporal particle intensity profile, etc.). In general, such models are not suitable for providing real-time forecasting because they require significant computational resources. Some input parameters of the models may also be poorly characterized, so they may not be easily incorporated into an operational workflow. In contrast, empirical models are based on identifying the statistical relationships present between SEP events and characteristics of the precursors, which are related to the underlying physical processes, but are not intended to model them. For this reason, these models provide quick forecasts and can be easily incorporated into operations. The output of these model can be a binary prediction of occurrence of an SEP event in a given time window or they can provide an estimate of some quantity of interest (e.g., SEP event onset time, peak intensity, etc.). More recently, ML models are being studied and applied giving rise to a new class of SEP prediction models that can provide quick outputs with improved accuracy [11–14]. Similar to empirical models, ML algorithms are able to identify relationships between SEP events and other parameters on, at least ideally, large datasets without any knowledge of the underlying physical processes. In this sense, ML models can be regarded as an upgrade of the empirical models, from which they are inspired in approach and purpose, but they differ in the use of complex and refined ML algorithms (e.g., neural networks).

Model	POD	FAR	Reference
AFRL PPS (SXR)	43%	50%	Smart & Shea [19], Kahler *et al.* [20]
Protons (SXR, Radio)	57%	55%	Balch [21]
FORSPEF (SXR, Radio)	55%	42%	Papaioannou *et al.* [22]
Wind/Waves Type II/III data	62%	22%	Winter & Ledbetter [23]
ESPERTA (SXR, Radio)	62%	39%	Laurenza *et al.* [24], Alberti *et al.* [25]
FORSPEF (SXR, CME)	71%	41%	Anastasiadis *et al.* [26]
REleASE (Electrons)	63%	30-35%	Posner [27]
UMASEP (SXR, Protons)	80%	34%	Nunez [16]
SWPC Forecaster	88%	18%	1995-2005

Table 1: Comparison between POD and FAR scores of some models for SEP forecasting based on different features/SEP precursors.

At present, nearly three dozen SEP models have been or are being developed in the scientific research community, although not all of them have been validated over a long SEP dataset, (see [10] for a complete review on the available models). Commonly used scores enabling us to quantify the model performances are the probability of detection (POD)

$$\mathrm{POD} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{1}$$

and the false alarm rate (FAR)

$$\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}}, \tag{2}$$

where TP stands for "true positive", FN for "false negative" and "FP" for "false positives". The model performances in terms of POD and FAR are mainly determined by the set of observations (features) the model relies on. Table 1 shows the scores of some statistical models based on different precursors. For instance, methods based on CME parameters exhibit a better performance than those using the flare ones, although fast CMEs which have not associated SEP events could increase the FAR.

Another important metric for evaluating the performances of forecasting models is the warning time, which is fundamental for an early alert that can mitigate the effects of an incoming SEP event [15]. In this respect, flare-based and also CME-based models show a significant temporal advantage in average with respect to other models, especially those including also the high-energy proton intensity as a feature (i.e., autoregressive models). Indeed, SEP forecasting techniques that depend on particle intensity [e.g., 16–18] may issue late alerts if the maximum forecast intensity is close to the forecasting threshold. Nevertheless, the key observations of CMEs in the solar corona are not available in real time and this makes CME-based models less effective for a near-real-time forecasting.

The purpose of this paper is to present a discussion of some relevant issues in the statistical models for SEP forecasting using flare precursors as input variables by focusing on the Empirical model for Solar Proton Event Real Time Alert (ESPERTA) model [12, 24, 25, 28]. All the main considerations associated with the SEP prediction in the framework of the ESPERTA model are

Table 2: POD and FAR validation metrics of the ESPERTA model for for the three time intervals.

	1995-2005 (training) [24]	2006-2014 (test) [25]	2015-2017 (test) [29]
POD	47/75 = 0.63	19/32 = 0.59	4/5 = 0.80
FAR	34/81 = 0.42	8/27 = 0.30	2/6 = 0.33

pretty general and must be carefully taken into account in developing empirical models. The paper is organized as follows: in Section 2 we make an overview of the ESPERTA model, whereas in Section 3 we show the model cross-validation in the general framework of a ML approach. Discussions and conclusions are drawn in Section 4.

2. The ESPERTA model: an overview

The ESPERTA model was first introduced by Laurenza *et al.* [24] as a novel technique meant for providing short-term warnings of SEP events that meet or exceed the Space Weather Prediction Center threshold of flux ≥10 pfu for protons of energy >10 MeV (i.e. ≥S1 radiation storm in the NOAA scale). ESPERTA is based on the statistical association between ≥S1 SEP events and ≥M2 flares, so it can be classified as an empirical model. In order to provide a quick and reliable prediction of an incoming SEP event, the model takes as input variables only the flare-related features listed below:

1. the heliolongitude of the flare (measured as the averaged heliographic position of the flare central region from observations gathered by several observatories in the $H\alpha$ spectral line);

2. the time integrated soft X-Ray (SXR) flux (X) in the 1-8 Å wavelength range. In order to provide a quick prediction, the time integration is performed until the SXR flux reaches a value which corresponds to the 1/3 of the peak flux within 10 minutes from the SXR peak. For those cases where after 10 minutes the SXR flux has not yet reached the 1/3 of the peak, an exponential fit after the SXR peak is used to extrapolate the 1/3 peak flux point to get an indication of the SXR fluence then used in the analysis. In general, those cases need a further prescription since the SXR fluence tends to be overestimated when the ratio between the fitted SXR intensities at +10 min after the peak time and at the peak time is high (e.g. >0.85) [see 24, for further details];

3. the time integrated radio fluence (R) at a frequency of ∼1 MHz. Due to the highly structured shapes of ∼1 MHz radio flux with respect to the SXR intensity, the time-integration starts 10 minutes after the SXR flux integration and it stops after 10 min from the SXR peak flux.

ESPERTA consists essentially of three models corresponding to the following heliographic regions: east (E120° - E41°), center (E41° - W19°) and west (W20° - W120°). After identifying the heliographic region where a flare is released, the binary prediction of the occurrence/non-occurrence of a SEP ≥S1 after observing a ≥M2 class flare is carried out through a logistic regression algorithm, which takes as input features $\log X$, $\log R$ and the product $\log X \log R$. The first validation of the model was performed on the data interval 1995-2005, covering almost the entire solar cycle 23 [24].

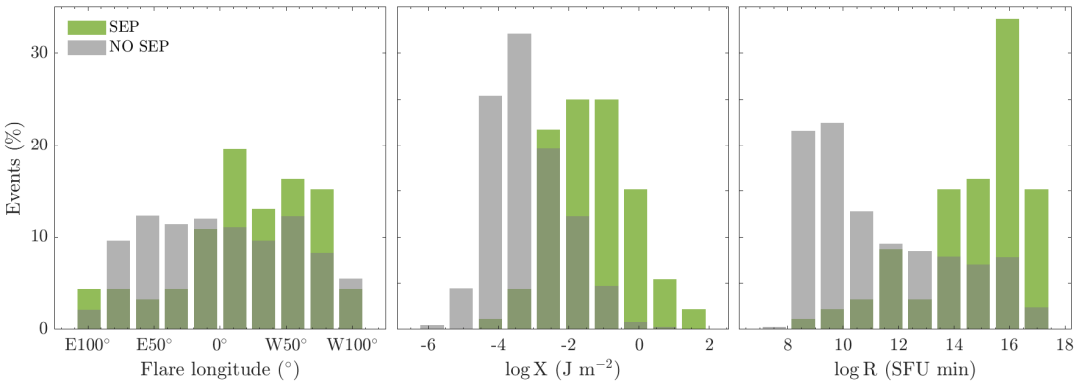

Figure 1: Distributions of the features separated by SEP-associated (green) and non-SEP-associated events (gray).

In this case, the dataset used for training the model is the same used during the testing phase and thus the validation metrics provided at this stage are referred to the training set. Thereafter, two tests of the ESPERTA model performed on unseen SEP data, were carried out later by Alberti *et al.* [25, 29] in the data interval 2006-2014 and 2015-2017, coming to cover the entire solar cycle 24. In these last works authors used the model parameters fitted in the interval 1995-2005 to estimate the model performance over the new test dataset. A summary of the POD and FAR metrics obtained are reported in Table 2. At this stage, the validation metrics provided for this forecasting model depends on the single choice of the training and testing datasets, made on a chronological basis. For this reason a statistical validation of the ESPERTA model became necessary and this procedure will be explained in the next section.

3. The ESPERTA model and the stratified cross-validation

Recently, Stumpo *et al.* [12] reframed ESPERTA in a ML perspective thus providing a statistical cross-validation of the model. In order to perform a binary classification, ESPERTA exploits the supervised learning approach, i.e. the model optimal weights are found by exploiting a set of known target events, labeled as 1 if a SEP follows a ≥M2 flare eruption or 0 otherwise. We define the class of no ≥S1 SEP events associated with an occurring ≥M2 flare as C_0 and the class of ≥S1 SEP events associated with an occurring ≥M2 flare as C_1. By inspecting the distribution functions of the features for the two classes of events, Figure 1, it is evident that there is a remarkable overlapping between the two classes. This represents an important source of error in the prediction, since in those overlapping regions the model uncertainty is large. We define the input vector $\boldsymbol{x}_i$, containing all the features involved in the prediction, and the target variable t_i, which can assume the value 1 if the i-th input vector is an associated SEP event $\in C_1$ or 0 otherwise. The model consists of a function $f(\boldsymbol{x}_i, \boldsymbol{w})$ which maps the input vector into the target variable, i.e. $t_i = f(\boldsymbol{x}_i, \boldsymbol{w})$. This function also depends on a set of unknown parameters $\boldsymbol{w}$, i.e. weights, which can be found by minimizing an error function with respect to a series of examples for which t is known. This is

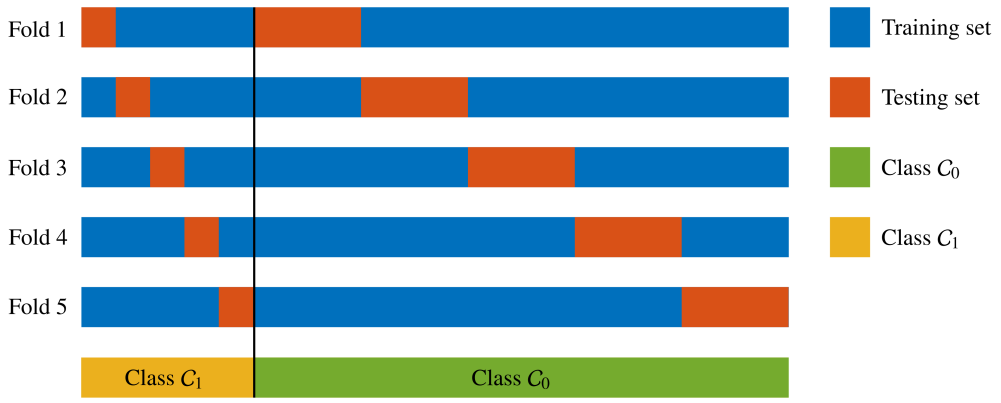

Figure 2: Sketch of the k-fold stratified cross validation performed on the ESPERTA model ($k = 5$). It should be noted how the ratio between the classes C_0 (green) and C_1 (yellow) in the whole dataset is preserved in both training (blue) and testing (red) sets.

done through the maximum-likelihood estimation (MLE) by minimizing the loss function [30]

$$L(\boldsymbol{X}|P) = \sum_{t_i=1} \log P(C_1|\boldsymbol{x}_i, \boldsymbol{w}) + \sum_{t_i=0} \log P(C_1|\boldsymbol{x}_i, \boldsymbol{w}), \tag{3}$$

where $\boldsymbol{X}$ denotes the whole training dataset and $P(C_1|\boldsymbol{x}_i, \boldsymbol{w})$ is the probability of observing a ≥S1 SEP event given the input vector $\boldsymbol{x}_i$ and the weight vector $\boldsymbol{w}$. In this fashion, the model learns the optimal weights directly from the given data. In order to generalize the ESPERTA model, the separation of heliolongitude data in three classes before performing the training of the algorithm is here disregarded and the heliolongitude is taken as a feature for the model in addition to the SXR and radio fluences. Moreover, in order to develop the model in the simplest possible scenario, we disregard also the product term between X and R. ESPERTA exploits the probabilistic approach in the classification of the events by modeling the "actual" probability that the i-th vector is a SEP through the logistic function

$$p(C_1|\boldsymbol{x}_i, \boldsymbol{w}) = \frac{1}{1 + \exp(-w_0 - \boldsymbol{w} \cdot \boldsymbol{x}_i)}. \tag{4}$$

In this case the unknown parameters are optimized by minimizing the functional of Equation (3) and to decide whether or not a given vector is an SEP we introduce a decision function which is based on a selected value of this probability. In an ideal condition (e.g., for balanced classes), a SEP event can be predicted if its estimated probability is larger than 0.5. Nevertheless, this threshold, that we denote with ϵ, can be considered as a parameter for the models and then it needs to be properly tuned for making reliable predictions.

The overlapping of the classes in the parameter space discussed before is the signature of the multivariate nature of the problem, of the presence of hidden/unknown variables and/or missing information. In general, the overlapping exists also in the multivariate space, meaning that the physics of the problem is not completely caught by the set of variables in use (e.g., we may need combinations between them which are not known *a priori*). To take this effect into account we

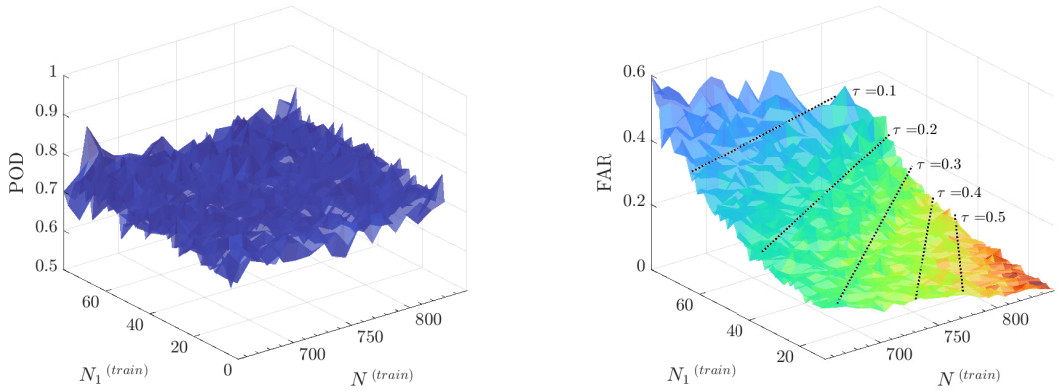

Figure 3: POD (left) and FAR (right) as a function of the number of ≥M2 flares $N^{(train)}$ and the number of ≥S1 SEPs $N_1^{(train)}$ in the training set. Expected values of the FAR score for different choices of the ratio τ are indicated by black dotted lines.

define the decision threshold ϵ as

$$f(\boldsymbol{x}_i;\boldsymbol{w}) = \begin{cases} 1 & \text{if} \quad p(C_1|\boldsymbol{x}_i,\boldsymbol{w}) \geq \epsilon \\ 0 & \text{if} \quad p(C_1|\boldsymbol{x}_i,\boldsymbol{w}) < \epsilon. \end{cases} \tag{5}$$

The calibration of the threshold during the cross-validation is important in order to allow the model to make reliable decisions when input features are located within overlapping regions of the parameter space. Moreover, the unbalance of the dataset might introduce some errors in the correct estimation of the probabilities. Since the loss function consists of two terms related to the two classes in which the dataset is split, Equation (3), if the classes are strongly unbalanced the model tends to learn more from the largest class. Thus, in our case the learning phase favors those flare events which do not have an associated SEP (i.e., C_0 class). In order to adjust this bias we apply two standard rare event correction approaches in order to perform the model cross-validation over a balanced dataset. The method used are:

- the weighting the loss function, i.e., the two terms appearing in Equation (3) are multiplied by the inverse frequency of the classes to counterbalance the importance of the relative contributions;
- the oversampling of the training dataset with the Synthetic Minority Oversampling Technique (SMOTE) in order to rise the statistics of the minority class.

The weighting of the loss function introduces two different weights associated with errors made by the model during the optimization phase, while the SMOTE oversampling replicates though a given protocol the C_1 class in order to meet the frequency of the C_0 class before performing the MLE. In order to make a comparison between the validation scores associated with these corrections, we perform the cross-validation of the model by studying the variation of POD and FAR scores as a function of the decision threshold ϵ. In order to find the best trade-off between POD and FAR we

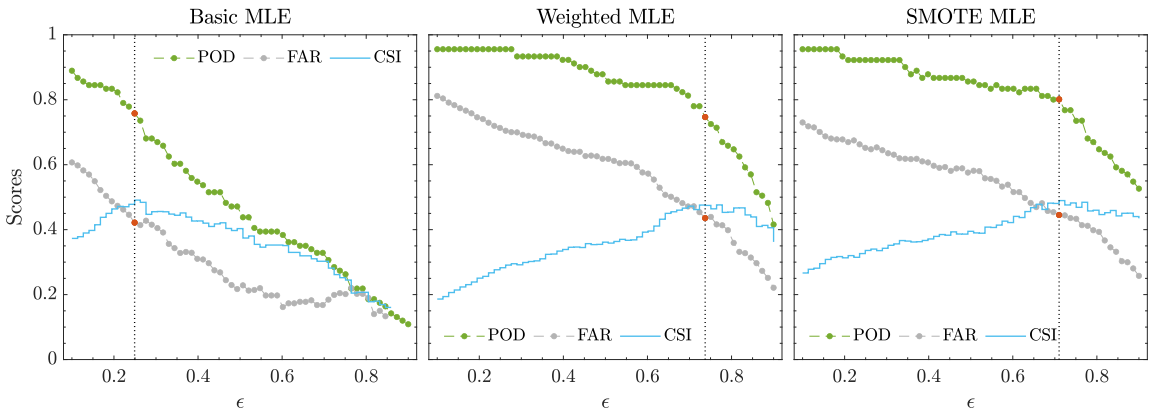

Figure 4: Averaged cross-validated POD (green), FAR (gray) and CSI (blue) scores with respect to the decision threshold ϵ for basic, weighted and SMOTE MLE. The dotted vertical line is placed in correspondence of the maximum of the CSI. The corresponding cross-validated scores are marked by red circles.

maximize the Critical Success Index (CSI), which is defined as:

$$\mathrm{CSI} = \left(\frac{1}{\mathrm{POD}} + \frac{1}{1-\mathrm{FAR}} - 1 \right). \tag{6}$$

For the cross-validation we use the k-fold stratified approach, which consists of splitting the classes C_0 and C_1 in training and testing sets within each fold, and by computing the validation metrics on each of them. We chose $k = 5$ and we require that the frequency of SEPs within the training and testing sets must be preserved. This sets the ground for a realistic scenario as it will be illustrated below. A scheme of the cross-validation procedure used in this work is illustrated in Figure 2.

If we define the fraction of SEPs in the training dataset as $N_1^{(train)}$ and the total number of ≥M2 flares in the training dataset as $N^{(train)}$, we can study how the validation scores may depend on them. In order to investigate this we introduce the fraction of ≥S1 SEP events in the testing dataset as $\tau = N_1^{(test)}/N^{(test)}$ and we cross-validate the model with the k-fold stratified approach for different choices of τ. If we write the fraction τ by using the total number of SEPs in the dataset N_1, the following linear relation between $N_1^{(train)}$ and $N^{(train)}$ can be derived [12]

$$N_1^{(train)} = \tau N N_1 + \tau N^{(train)}, \tag{7}$$

where the total number of ≥M2 flares N and ≥S1 SEPs N_1 are constant. Results obtained for POD and FAR are reported in Figure 3. The trend obtained from Equation (7) for different values of τ is superimposed on the FAR score in the right panel of Figure 3. Whereas the POD remains almost unaffected by the different choices of $N^{(train)}$ and $N_1^{(train)}$, the FAR score exhibit a wide range of variability. For instance, if we set the ratio τ to 0.5 and we consider the correspondent validation metrics, we get a remarkably low value for the FAR (∼10%) which is not representative of the real situation. In fact, the case $\tau = 0.5$ gives us information on how the model would work if the probability of observing an ≥S1 SEP event following an ≥M2 flare was 1/2. This means that, in order to provide a realistic evaluation of the performances, an operational model needs to be calibrated with respect to the expected occurrence of SEPs, otherwise, the model will provide an unrealistic number of false alarms.

Table 3: Cross-validated POD, FAR and CSI scores for the ESPERTA model in the basic, weighted and SMOTE MLE cases respectively.

	POD	FAR	CSI
Basic MLE	0.76	0.42	0.49
Weighted MLE	0.75	0.44	0.48
SMOTE MLE	0.80	0.45	0.49

For this reason the expected occurrence of true positive events, which is about 10% in the whole dataset, is chosen during the calibration. Then, for each value of the threshold ϵ the validation metrics (POD, FAR and CSI) are evaluated. Since the CSI metric provides a trade-off between POD and FAR, the optimal threshold is here defined as the one which maximizes the CSI. The calibration of the decision threshold ϵ is performed for the three cases of basic MLE (without class unbalance correction), weighted MLE and SMOTE MLE (with class unbalance correction). Averaged cross validation scores as a function of ϵ are reported in Figure 4. From these results it is evident how the parameter ϵ strongly affects the model performance. The values obtained in correspondence of the maximum values of the CSI score (dotted lines in Figure 4) are listed in Table 3. We may note that the corrections introduced in the learning phase to handling the bias due to the unbalance between the classes do not get much better performance. Indeed, the best case is represented by the SMOTE MLE which gives an increase of the POD of 4% (80% with respect to 76% of the basic MLE), but, on the other hand, exhibits a worsening of the FAR of 3% (45% with respect to 42% of the basic MLE). This suggests that the main error source is that precursors alone are unable to capture a complete information about the physics underlying the origin of ≥S1 SEP events.

4. Discussion and Conclusions

The SEP forecasting by means of empirical models and/or ML algorithms integrated in an operational workflow represents one of the most important aims in the *Space Weather* science. SEPs, in fact, represent one of the most hazardous events for human activities and health, hence correct and reliable predictions of such energetic events are desirable in order to coordinate Earth and space operations on which they could have a major impact. In this framework the ESPERTA model represents an important tool with a remarkable performances in terms of validation metrics (POD and FAR) and average warning time [24, 25, 29].

In this paper we pointed out the importance of properly calibrating and cross-validating forecasting models in order to get reliable performances. We have shown that the optimization of the weights of the model by using the basic MLE for ESPERTA, as well as the weighted MLE and the SMOTE MLE do not provide a significant improvement of the performances, especially in terms of the FAR score. The POD score is defined on elements belonging to the same class, Equation (1), and thus it is not affected by the unbalancing of the dataset. In contrast, the FAR score mixes the classes C_0 and C_1 in its definition, Equation (2), and it can be expressed as a function of the fraction τ [12]. This fact highlights the extreme importance of a correct calibration of the model in order to get realistic performances in an operational scenario.

In summary, the most important considerations and warnings of this work can be stated as follows:

1. any statistical forecasting model dealing with a small and unbalanced dataset such as SEPs needs to be calibrated with respect to the expected occurrence of the classes C_i;
2. the decision threshold ϵ strongly affects the model performance;
3. the use of different techniques for handling unbalanced data do not significantly improve the performance of the model especially if the available features, when taken individually, are unable to fully separate the classes of events within the parameter space.

Acknowledgements

The data used in this work are publicly available at the web link http://hdl.handle.net/20.500.12386/30967. This research has been carried out in the framework of the CAESAR (Comprehensive spAce wEather Studies for the ASPIS prototype Realization) 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 (ASI Space weather InfraStructure) prototype of scientific data centre for Space Weather. M.S. acknowledges the PhD course in Astronomy, Astrophysics and Space Science of the University of Rome "Sapienza", University of Rome "Tor Vergata" and Italian National Institute for Astrophysics (INAF), Italy.

References

[1] S.W. Kahler, J. Sheeley, N. R., R.A. Howard, D.J. Michels, M.J. Koomen, R.E. McGuire et al., *Associations beteen coronal mass ejections and solar energetic proton events*, *Journal of Geophysical Research* **89** (1984) 9683.

[2] M.A. Shea and D.F. Smart, *A Summary of Major Solar Proton Events*, *Solar Physics* **127** (1990) 297.

[3] M.-B. Kallenrode, *Neutral lines and azimuthal "transport" of solar energetic particles*, *Journal of Geophysical Research* **98** (1993) 5573.

[4] D.V. Reames, *Solar energetic particle variations*, *Advances in Space Research* **34** (2004) 381.

[5] M. Desai and J. Giacalone, *Large gradual solar energetic particle events*, *Living Reviews in Solar Physics* **13** (2016) 3.

[6] N. Iucci, A.E. Levitin, A.V. Belov, E.A. Eroshenko, N.G. Ptitsyna, G. Villoresi et al., *Space weather conditions and spacecraft anomalies in different orbits*, *Space Weather* **3** (2005) 01001.

[7] M.-H.Y. Kim, G. De Angelis and F.A. Cucinotta, *Probabilistic assessment of radiation risk for astronauts in space missions*, *Acta Astronautica* **68** (2011) 747.

[8] J.P. Eastwood, E. Biffis, M.A. Hapgood, L. Green, M.M. Bisi, R.D. Bentley et al., *The Economic Impact of Space Weather: Where Do We Stand?*, *Risk Analysis* **37** (2017) 206.

[9] J.B.L. Jones, R.D. Bentley, R. Hunter, R.H.A. Iles, G.C. Taylor and D.J. Thomas, *Space weather and commercial airlines*, *Advances in Space Research* **36** (2005) 2258.

[10] K. Whitman, R. Egeland, I.G. Richardson, C. Allison, P. Quinn, J. Barzilla et al., *Review of solar energetic particle models*, *Advances in Space Research* (2022) .

[11] M. Núñez and D. Paul-Pena, *Predicting >10 MeV SEP Events from Solar Flare and Radio Burst Data*, *Universe* **6** (2020) 161.

[12] M. Stumpo, 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** (2021) e2021SW002794.

[13] E. Lavasa, G. Giannopoulos, A. Papaioannou, A. Anastasiadis, I.A. Daglis, A. Aran et al., *Assessing the Predictability of Solar Energetic Particles with the Use of Machine Learning Techniques*, *Solar Physics* **296** (2021) 107.

[14] S. Aminalragia-Giamini, S. Raptis, A. Anastasiadis, A. Tsigkanos, I. Sandberg, A. Papaioannou et al., *Solar Energetic Particle Event occurrence prediction using Solar Flare Soft X-ray measurements and Machine Learning*, *Journal of Space Weather and Space Climate* **11** (2021) 59.

[15] O. St. Cyr, A. Posner and J. Burkepile, *Solar energetic particle warnings from a coronagraph*, *Space Weather* **15** (2017) 240.

[16] M. Núñez, *Predicting solar energetic proton events (E > 10 MeV)*, *Space Weather* **9** (2011) S07003.

[17] M. Núñez, *Real-time prediction of the occurrence and intensity of the first hours of >100 MeV solar energetic proton events*, *Space Weather* **13** (2015) 807.

[18] M. Núñez, P.J. Reyes-Santiago and O.E. Malandraki, *Real-time prediction of the occurrence of GLE events*, *Space Weather* **15** (2017) 861.

[19] D. Smart and M. Shea, *Pps-87: A new event oriented solar proton prediction model*, *Advances in Space Research* **9** (1989) 281.

[20] S. Kahler, E. Cliver and A. Ling, *Validating the proton prediction system (pps)*, *Journal of atmospheric and solar-terrestrial physics* **69** (2007) 43.

[21] C.C. Balch, *Updated verification of the Space Weather Prediction Center's solar energetic particle prediction model*, *Space Weather* **6** (2008) S01001.

[22] A. Papaioannou, A. Anastasiadis, I. Sandberg, M.K. Georgoulis, G. Tsiropoula, K. Tziotziou et al., *A Novel Forecasting System for Solar Particle Events and Flares (FORSPEF)*, vol. 632, p. 012075, Aug., 2015, DOI.

[23] L.M. Winter and K. Ledbetter, *Type ii and type iii radio bursts and their correlation with solar energetic proton events*, *The Astrophysical Journal* **809** (2015) 105.

[24] M. Laurenza, E. Cliver, J. Hewitt, M. Storini, A. Ling, C. Balch et al., *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** (2009) .

[25] T. Alberti, M. Laurenza, E. Cliver, M. Storini, G. Consolini and F. Lepreti, *Solar activity from 2006 to 2014 and short-term forecasts of solar proton events using the esperta model*, *The Astrophysical Journal* **838** (2017) 59.

[26] A. Anastasiadis, A. Papaioannou, I. Sandberg, M. Georgoulis, K. Tziotziou, A. Kouloumvakos et al., *Predicting flares and solar energetic particle events: The forspef tool*, *Solar Physics* **292** (2017) 1.

[27] A. Posner, *Up to 1-hour forecasting of radiation hazards from solar energetic ion events with relativistic electrons*, *Space Weather* **5** (2007) .

[28] M. Laurenza, T. Alberti and E. Cliver, *A short-term esperta-based forecast tool for moderate-to-extreme solar proton events*, *The Astrophysical Journal* **857** (2018) 107.

[29] T. Alberti, M. Laurenza and E. Cliver, *Forecasting solar proton events by using the esperta model*, *Il nuovo cimento C* **42** (2019) 1.

[30] P. McCullagh and J. Nelder, *Generalized Linear Models, Second Edition*, Chapman and Hall/CRC Monographs on Statistics and Applied Probability Series, Chapman & Hall (1989), 10.1201/9780203753736.