EGU2018-7329
[X4.207]

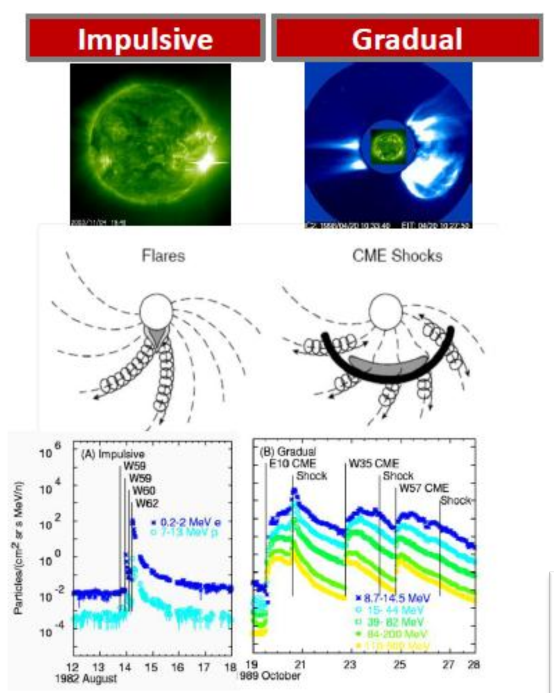
Nowcasting Solar Energetic Particle (SEP) events using Principal Components Analysis (PCA)

A. Papaioannou[1], A. Anastasiadis[1], A. Kouloumvakos[2], M. Paassilta[3], R. Vainio[3], E. Valtonen[3], A. Belov[4], E. Eroshenko[4], V. Yanke[4], M. Abunina[4], A. Abunin[4]

[1] Institute of Astronomy, Astrophysics, Space Applications and Remote Sensing, National Observatory of Athens, Greece
[2] IRAP, Université de Toulouse, CNRS, CNES, UPS, Toulouse, France
[3] Department of Physics and Astronomy, University of Turku, Finland,
[4] Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation (IZMIRAN), 42092 Troitsk, Moscow Region, Russia

Abstract: Solar Energetic Particle (SEP) events and their parent solar events (e.g. solar flares - SFs and coronal mass ejections - CMEs) are closely related. A wealth of statistical studies has indicated the dependence of the probability of occurrence of SEP events on the magnitude and the longitude of the SF, as well as the velocity and the width of the CME. However, most studies are limited to two dimensional correlations. In addition, similar coefficients are identified for the pair-wise correlation of the SEP peak intensity to both the SF magnitude and the CME speed. The situation is further complicated by the fact that the solar parameters are not independent. In this work, we perform a principal component analysis (PCA) on a set of six (6) solar variables (i.e. CME width and velocity, logarithm of the SF magnitude, SF longitude, duration and rise time), and we further apply logistic regression to infer the possible prediction of SEP events. In our analysis, we utilize 126 SEP events with complete solar information. Each SEP event is a vector in six dimensions (corresponding to the six solar variables used in this work). PCA transforms the input vectors into a set of orthogonal components. We applied logistic regression with a single categorical predictor, as well as, single or multiple explanatory variables. Furthermore, we validated our findings with the implementation of categorical scores (Probability of Detection - POD, False Alarm Rate –FAR). We present and interpret the obtained scores and we discuss the strengths and weaknesses of the different implementations

Motivation



> **Solar Energetic Partiicle (SEP) events** *are associated to* ***solar flares (SFs)*** *and* ***coronal mass ejections (CMEs)***. We currently know that:

Solar flares @ W45-W70 are associated to SEP events → Magnetic connection

The higher the flare flux the higher the probability of SEP occurrence @ all energies

Peak particle intensities are correlated to the SXR peak flux

Peak particle intensities are correlated with CME speeds

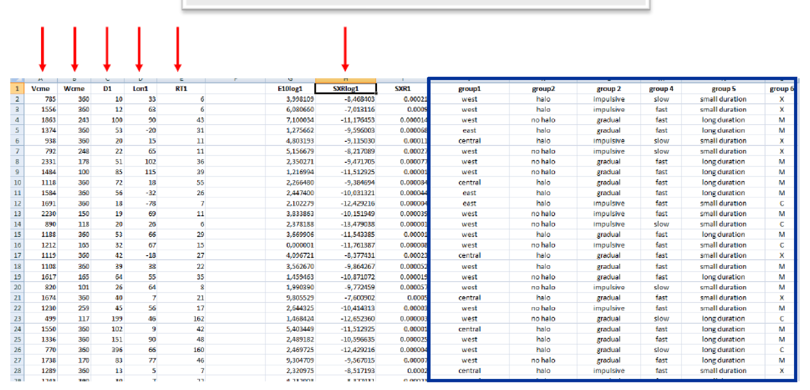
SEPs are associated to type III radio bursts (~100%) and type II radio bursts (~70%)

> We now have several inter-related quantities (solar) that describe (map) SEP events → Can we make use of a higher-order combination of these quantities?

Data & Methods

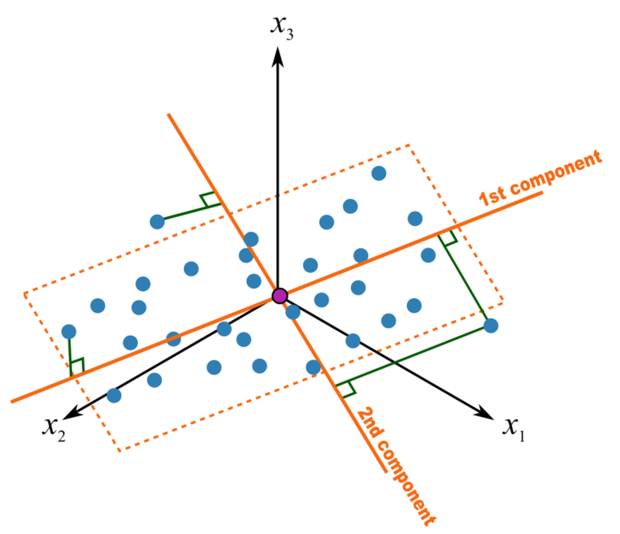
> We have indentified a *complete parametric grid* of **six** (6) **solar variables** (i.e. CME width/size (**s**) and velocity (**u**), logarithm of the SF magnitude (**logSXRs**), SF longitude (**Lon**), duration (**DT**) and rise time (**RT**)), covering the time period from 1997-2013.
> This resulted in a total of *3663 records* with complete information for all **six** (6) **variables**, out of which **126** were **SEP events** and *3537* were *non-SEP events*.

> 126 SEP events
> 6 quantities > several different mappings



> **Principal Component Analysis (PCA) is a multivariate technique that** extracts the ***important information*** from the table, to represent it as a set of *new orthogonal variables* called **principal components**, and to ***display the pattern of similarity*** of the observations and of the variables as points in maps [**score plots**, **loadings plot**].

> **PCA** *rotates the original data space* such that the axes of the new coordinate system point into the directions of highest variance of the data



Classification of SEP events

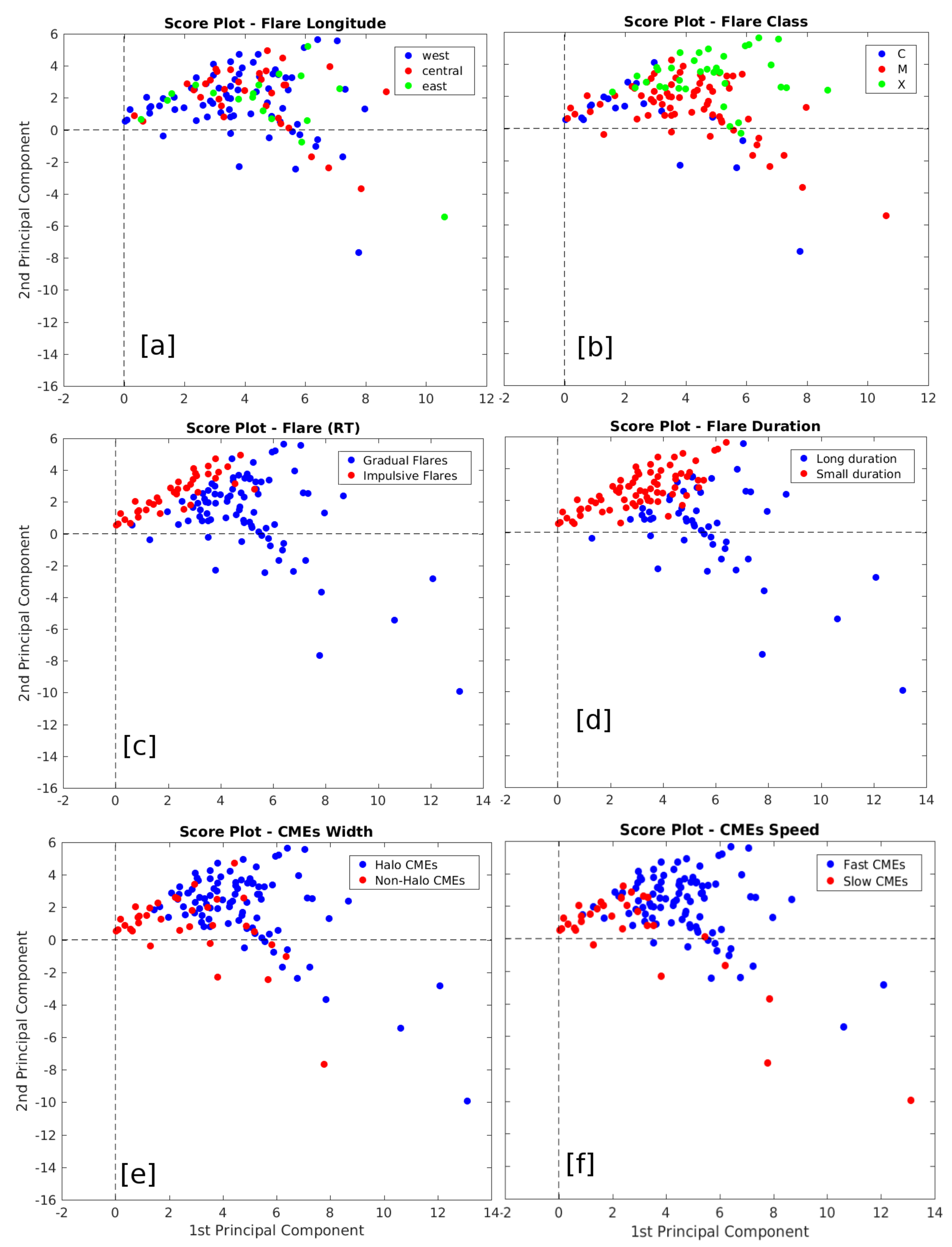


__Figure 1__. Results of the PCA. From top to bottom, seven score plots, colour coded on the basis of different groupings of the variables, while the bottom panel on the right depicts the loadings of the primary components.

> **Figure 1** presents the **score plots** of the SEP sample for different groups of the initial parameters/variables (from the top panel on the left, labeled with [**a**], to the bottom panel on the right, labeled with [**f**]), as well as the score plot (bottom panel on the right, labeled with [**g**]).

> The ***first four panels*** of focus on the variables that stem from **solar flares**, while the following two panels, i.e., [**e**] and [**f**] display the obtained score plots on the basis of the **CME characteristics**.

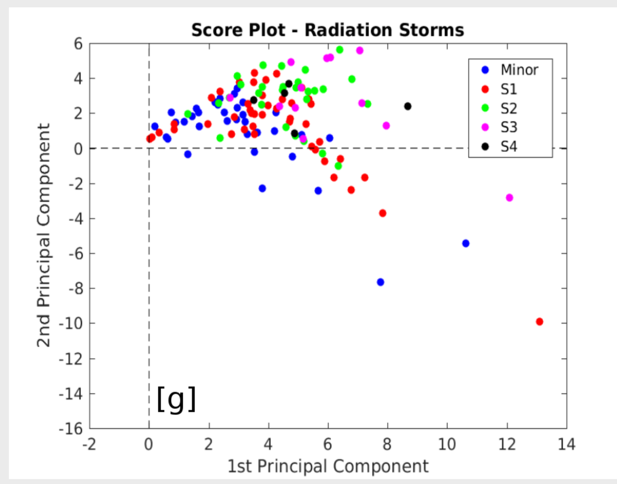
> Panel [**a**] is colour coded on the basis of the position of the parent solar are, i.e., green stands for *western longitudes* (**W20-W120**); blue for *central longitudes* (**E20-W20**) and brick for *eastern* (**E100-E20**) *longitudes* of the SEP associated solar flares.

> Panel [**b**] is colour coded on the basis of the GOES peak photon flux with blue colour presenting **C class**; brick colour **M class** and green colour **X class** solar flares.

> Panel [**c**] is colour coded on the basis of the solar are rise time, with blue colour denoting gradual flares (i.e. rise time > 13 min) and brick colour standing for impulsive solar flares (i.e. rise time < 13 min).

> Panel [**d**] is colour coded with respect to the **duration of the solar flare**. Blue colour stands for long duration solar flares, while brick colour for short duration solar flares.

> Panel [**e**] presents **halo** (Earth directed – 360 width) **CMEs** in blue colour and all other **non halo CMEs** in brick colour. Panel [**f**] depicts **fast CMEs** (1000 Km s[-1]) in blue and **slow CMEs** (<1000 Km s[-1]) in brick colour.



> Panel [**g**] presents the score plot of all 126 events colour coded as a function of their solar radiation scale, e.g. S1, in brick colour; S2, in green; S3, in purple; S4, in grey and minor events (E> 10 MeV <10 pfu) in blue colour.

Nowcasting SEP events:

Papaioannou *et al*., Solar Physics, *accepted*, 2018

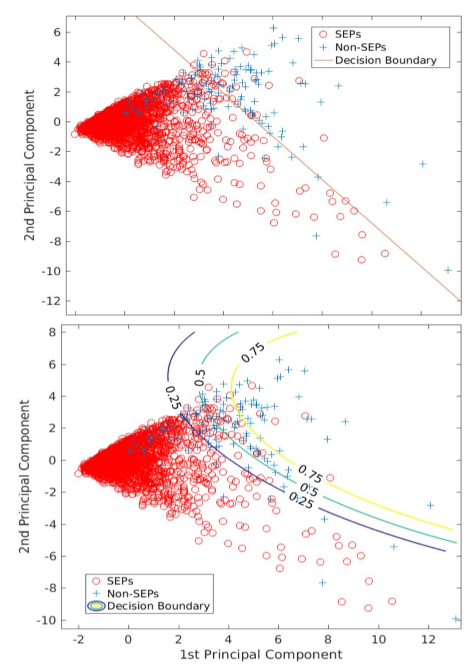


__Figure 2__. Scatter plot of the SEP (blue crosses) and Non-SEP (red circles) events, as they map on the projected space of PC1 and PC2.

> An index **I** based on **PCA**

$$I = PC1 = A_1 * (\mathrm{logSXR}) + A_2 * (\mathrm{Long}) + A_3 * (\mathrm{RT}) + A_4 * (\mathrm{DT}) + A_5 * (\mathrm{u}) + A_6 * (\mathrm{w})$$

$$h_\theta(g(x)) = \frac{1}{1 + e^{-\theta^T g(x)}}$$

> When we applied **PCA** and then a **Logistic Regression,** by a *simple threshold* **pth=50%** we could identify most of the SEP events in our sample (*__Figure 2__* top panel).
> When applying a **higher order regression** we obtained *several different decision boundaries* for different **pth**, depicted as contours (*__Figure 2__* bottom panel)

Categorical scores

> Several different configurations of the index **I** and the related categorical scores

index	Form (Scheme)	POD (%)	FAR (%)	HSS
$I^{(1)}$	[PC1]	26.98	44.26	0.3490
$I^{(2)}$	[PC1, PC2]	54.76	30.30	0.6013
$I^{(3)}$	[PC1, PC2, PC3]	55.56	28.57	0.6134
$I^{(4)}$	[PC1, PC2, PC3, PC4]	53.97	29.17	0.6007
$I^{(5)}$	[PC1, ..., PC5]	53.17	29.47	0.5943
$I^{(6)}$	[PC1, ..., PC6]	53.17	28.72	0.5973
$I^{(2+O^2)}$	$[PC1, PC2, PC1^2, PC2^2, PC1 * PC2]$	56.35	31.07	0.6080
$I^{(3+O^2)}$	$[PC1, PC2, PC3, PC1^2, PC2^2, PC3^2, PC1 * PC2]$	58.73	24.49	0.6502

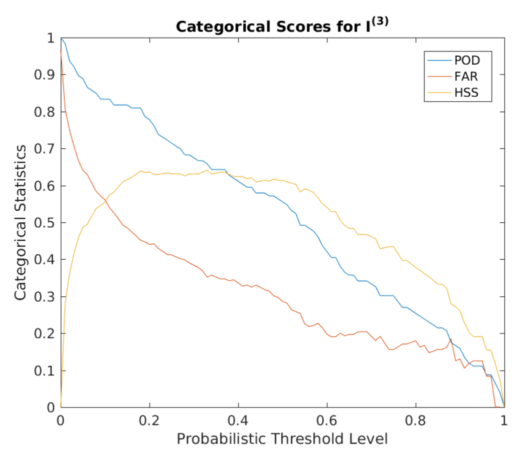


__Figure 3__.Categorical Scores (POD, FAR, HSS for I(3)

Conclusions

> We applied **Principal Component Analysis** (**PCA**) to the **SEP events** of our sample and showed that *significant radiation storms*, categorized as **S4**, **S3** and **S2**, are *related* to **fast** and **halo CMEs**, as well as **SFs** of **>M-class**
> Using the **outputs of PCA**, a new index (**I**) was introduced and tested with respect to its *predictive capabilities*. It was demonstrated that ***it actually holds prognosis potential for SEP events***.
> Employing the **logistic regression analysis**, we introduced *several different schemes* for the (**I**) index. We found that the *classification* of **SEP events** *versus* **Non-SEP** ones, for a threshold ***pth=50%*** leads to a **FAR** of **24.49 %** while correctly predicting **58.73 %**. The optimal **POD=77.78 %** was obtained for I(3+O2).

References: **Papaioannou et al.,** ***Nowcasting Solar Energetic Particle (SEP) Events using Principal Components Analysis (PCA)***, **Solar Physics,** ***accepted***, **2018**

Contact: **atpapaio@astro.noa.gr**

Acknowledgement: **This post-doctoral research was carried out by A.P. with an IKY scholarship funded by the action "Supporting post-doctoral researchers" from the resources of the b.p. "Human Resources Development Education and Lifelong Learning" with Priority Axes 6,8,9 and co-funded by the European Social Fund and the Greek government. AA would further like to acknowledge the "SPECS: Solar Particle Events and foreCasting Studies research grant of the National Observatory of Athens. M.P. acknowledges funding from ESA contract 4000120480/17/NL/LF/hh.**