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

A. Papaioannou¹, A. Anastasiadis¹, A. Kouloumvakos², M. Paassilta³, R. Vainio³, E. Valtonen³, A. Belov⁴, E. Eroshenko⁵, V. Yanke⁶, M. Abunina⁷, A. Abunin⁴

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 Particle (SEP) events are associated with solar flares (SFs) and coronal mass ejections (CMEs). We currently know that:

- Solar flares @ W45-70 is associated to SEP events
- Magnetic connection
- The higher the flare 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?

Classification of SEP events

- 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 plots the original data space such that the axes of the new coordinate system point into the directions of highest variance of the data

Data & Methods

- We have identified a complete parametric grid of six (6) solar variables (i.e. CME with and without a velocity, a logarithm of the SF magnitude, SF longitude, duration and rise time), covering the time period from 1997-2013.
- This resulted in a total of 3693 records with complete information for all six (6) variables, out of which 126 were SEP events and 3567 were non-SEP events.

Nowcasting SEP events:

- Several different configurations of the index I and the related categorical scores

Categorical Scores

- Several different configurations of the index I and the related categorical scores

Conclusions

- We applied Principal Component Analysis (PCA) to the SEP events of our sample and showed that significant radiation storms, categorized as SFs, 5A and 5B, are related to fast halo CMEs, as well as 5A of X-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 prognostic potential for SEP events.
- By 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 of 0.40 leads to a FAR of 24.49% while correctly predicting 58.78%. The optimal POD-PF threshold was obtained for I (32).