CME detection onboard Venus Express

Agatha DUZAN — Sebastien CHAHOUD — Nastasia MOREL
Department of Computer Science, EPFL, Switzerland
Under the supervision of professor O. Witasse (European Space Agency, ESTEC, Noordwijk, Netherlands)

***Abstract*—While predicting Space Weather forecasting and protecting spacecraft against radiation are increasingly crucial for planetary exploration, only a few spacecraft are equipped with dedicated tools for these tasks. However, every spacecraft, whether planetary or astronomical, is fitted with numerous housekeeping sensors. Some of these sensors can be useful in identifying radiation hazards caused by solar particle events. Specifically, certain sensors like the Error Detection and Correction (EDAC) memory counters can pinpoint energetic particles affecting detectors and subsystems. These counters often exhibit a sudden spike in error counts corresponding to the arrival of energetic particles [1] [2]. Furthermore, during the impact of a Coronal Mass Ejection (CME), the sensors can detect a forbush decrease associated with a reduction in the slope or plateau of the EDAC counter. Other sensors such as the magneto can also detect some changes with the spacial weather since a sudden change of the value of the magnetic field usually correspond to a CME event.**

Our study involves analyzing the engineering data set from the European Space Agency (ESA) Solar System missions of the Venus Express spacecraft. We perform a feasibility study on detecting CME events using EDAC counters and magneto sensors.

The study's results illustrate how engineering sensors can provide insights into the solar particle environment at a spacecraft location.

I. Introduction

The data used in this project comes from the ESA database of the Venus Express mission.We have the dataset containing the EDAC counter daily & every 12 minutes, the magneto data sensors including the intensity of the magnetic field in the three direction x, y & z with its distance from the surface of Venus and another dataset containing the CME start & end time.

It was our job to use this data to build a model able to predict a CME event happening in live. In practice, we were able to predict a CME event with an accuracy of 99.8% and a F1 score of 95.5%.

II. Data fetching and Pre-processing

* *Fetching the data:*
  The first steps of our job required fetching data from the VEX mission:
  - EDAC counters
  - Magnetometer measurements
  - CME start and end times

  We acquired the data from the ESA database [3] and various other sources (public & private) from ESA missions. This task involved developing web scrapping algorithms as the data was stored in separate files for each day and aggregated over each month, creating a nested web link structure.

* *Sanitizing the data:*
  CME related data was obtained from difference sources and in different formats. To have data consistency, we converted the data such as date-time features into usable formats.

* *Remove counters restarts:*
  During the VEX mission, EDAC counters onboard the space craft were occasionally reset to zero. Hence as we can see in Figure 1, we corrected the counter by removing these resets as they made by human intervention, and thus irrelevant to our data.

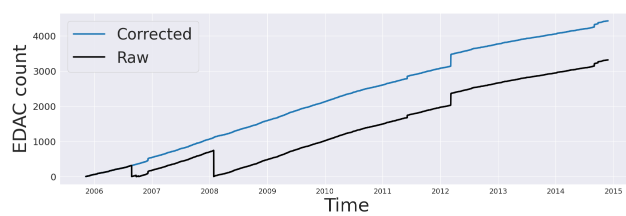

Fig. 1: Raw and Corrected EDAC counter of the VEX mission

III. First look at the data

Once we had usable data, we performed some basic analysis before training any models, to assess potential challenges that we might face. Here are the notable results that we had :

A. Class imbalance

CME events are very rare so they are under-represented in our dataset: only 1.78% of our data corresponds to CMEs. This class imbalance can lead to models that are biased toward the majority class (always predicting that there is no CME event) and might not perform well in accurately predicting the minority class (high false negative rate resulting in a low F1 score).

When training our models, we will have to implement additional techniques to mitigate this imbalance (we will give more detail on this later, when we explain the training phase).

B. Low feature correlation

Before even training a model that predicts CMEs by looking at EDAC, we looked at the correlation between these two.

We first looked at it just graphically: in Figure 2, we see that although some CMEs do seem to correspond with a jump in EDAC, some other CMEs happen with no EDAC change, and sometimes there is an increase in EDAC without it being linked to a CME .

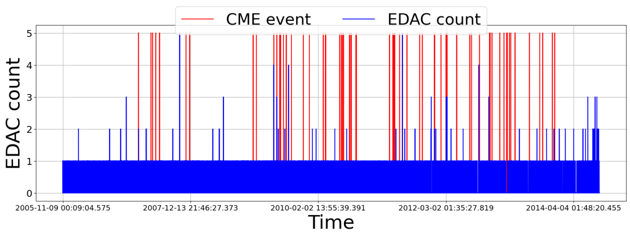

Fig. 2: Overlay of EDAC counter with CME events

To deepen this analysis, we also computed the correlation coefficients of the two variables. Since we didn't know if the correlation was linear or not, we looked at both Pearson and Spearman correlation indicators:

∗ *Pearson Correlation Coefficient:*
  This measures the linear correlation between two variables. It ranges from -1 to +1, with -1 indicating a perfect negative linear correlation, +1 indicating a perfect positive linear correlation, and 0 indicating no linear correlation.

∗ *Spearman's Rank Correlation:*
  This assesses monotonic relationships: it also ranges from -1 to +1, with -1 indicating a perfect negative association between ranks (as one variable increases, the other variable decreases), +1 indicating a perfect positive association (as one variable increases, the other variable also increases) and 0 indicating no association between ranks. If the data is not normally distributed or the relationship is not linear, it might be more appropriate.

The results for the correlation between the features EDAC and CME are summarized in Table I :

Pearson Correlation Coefficient	0.013
Spearman's Rank Correlation	0.010

TABLE I: EDAC and CME correlation Coefficients

Given these almost-zero correlation values, it's clear that EDAC counts alone have very little association with CME events.

This is why in our models later, we will look at more complex patterns of EDAC counts: to predict a CME at a given time, we will look at the time series of the last X hours of EDAC counts. This should lead to better prediction.

C. Magnetometer data

CME event directly influence the magnetic field around them. Hence during a CME, we observe fluctuation in the magnetic field measured of instruments onboard the VEX spacecraft, as seen in in Figure 3. For this particular reason, we will utilize magnetometer data to predict CME and compare it's performance with the EDAC only prediction models.

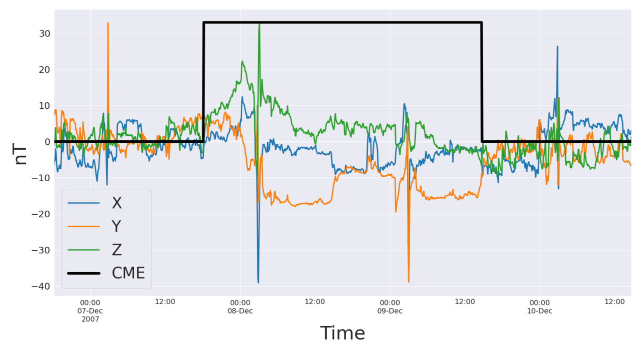

Fig. 3: Magnetometer measurements during a CME event

VEX spacecraft orbits near the planet Venus. Hence the magnetometer measurements are influenced by the magnetosphere of the planet, resulting in periodic spikes in the data, as shown in Figure 4. This observation led us to add the VEX-sun distance as another feature for our classification task.

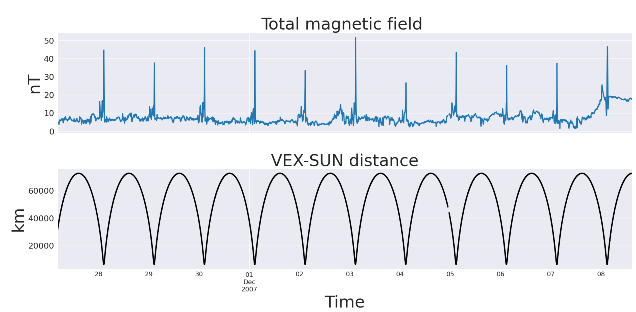

Fig. 4: Relation between Magnetometer measurements spikes and VEX-SUN distance

IV. Baseline models

We started by training "easy" models, to serve as an indicator for later improvement, and to find the best hyper parameters with lower cost.

A. The "dumb" model

Let's consider a "dumb" model that would always predict zero (i.e. no CME event): this model would have high accuracy, in this case, 98%, but a terrible F1 score: zero (no true positive). This is why it's essential to consider both of these metrics for a comprehensive evaluation of our models, rather than relying solely on accuracy. Our objective will be to optimize the F1 score while also keeping a satisfactory accuracy (ideally not too much below the dumb model accuracy).

B. Basic Neural Network

This is our basic architecture: it will be a a baseline to compare its performance with the performance of the more complex LSTMs later. Basic neural networks are faster to train than the LSTMs so we will use this model for the hyperparameter search (since it requires many training runs). We also experiment with many techniques to mitigate class imbalance, and will only keep the ones that work best on the LSTMs.

* *Architecture details:*
  We created all the neural net models using the library TensorFlow [4]. The models took as input a time series (on the last X hours) of the EDAC counts, it had two hidden layers, and a final layer with the prediction. We used ADAM for the optimizer, and binary cross entropy for the loss since it is better suited for binary classification

* *Finding the best window size:*
  We created various models with different window sizes X, i.e. they predicted CME events at a given time by looking at the last X hours of data. The window size we tried spanned from just one hour to one month. We looked at the accuracy and F1 score and found that the best was to look at all data from the last X=3 hours.

* *Mitigating class imbalance:*
  Because of the severe class imbalance, the neural networks tended to learn the "dumb" model (always predicting zero). We tried two approaches to mitigate this: data augmentation (duplicating instances of the minority class) and adding class weights (give more importance to instances of minority class during training). Data augmentation led to strong overfitting on the training data and bad performance on the validation data so for the future models we only kept class weights.

C. First results

Table II summarizes the results we had for our basic models. Because of the severe class imbalance, the neural network tended to learn the "dumb" model (always predicting zero) thus resulting in a null F1 score. The model with class weights had a slightly better F1 score but a terrible accuracy (it always predicted the minority class).

Model	Accuracy	F1 score
"Dumb"	98%	0%
Neural Net	98%	0%
Neural net with Class weights	2%	4%

TABLE II: Baseline results of the simpler models

Overall, any technique trying to increase the F1 score led to a model that always predicted the minority class. This is why we moved on to a more complex architecture, better suited to deal with time series data and detect more complex patterns.

V. LSTM ARCHITECTURE

We chose to use the Long Short-Term Memory (LSTM) model [5] to detect the CMEs using time series data from spacecraft magneto and EDAC sensors. This model is an improved version of the recurrent neural networks (RNNs) since it can handle sequential data, improves the long-term dependency learning of the model and mitigates the vanishing gradient problem inherent from traditional RNNs.

The LSTM architecture includes memory cells, input gates, forget gates, and an output gate as seen in Figure 5. The memory cells serve as storage units, allowing the model to accumulate information over extended sequences. The input gates regulate the flow of new information into the memory cells, determining which elements are relevant for learning. Simultaneously, the forget gates manage the retention or discarding of existing information, enabling the model to selectively remember or forget patterns based on their significance. The output gate then governs the information that is passed on to subsequent time steps or the final prediction, contributing to the model's ability to forecast time series.

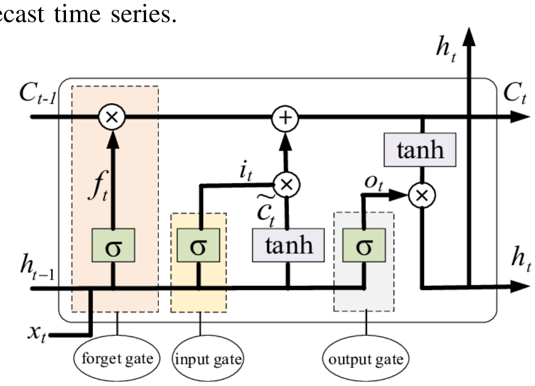

Fig. 5: Architecture of the Long Short-Term Memory model

Thanks to the memory cells and the different gates, the LSTM architecture can understand complex temporal patterns within the magneto and EDAC time series and improves the detection accuracy of CME events.

VI. Final model

For the final models that we trained, we created LSTMs using the PyTorch library [6].

A. Input features

Each LSTM took as input a timeseries of features of the last 3 hours (with a granularity of 5 minutes) and predicted whether it corresponded to a CME event. We trained different models that took different features as input to see which ones worked best for our task. The features we used are:

* EDAC counts
* Magnetic field vector magnitude (BT)
* Magnetic field vector coordinates (BX, BY and BZ)
* Distance between VEX spacecraft and the sun (RSC)

B. Model architecture

Each model was composed of two hidden LSTM layers stacked on top of each other: in a multi-layered LSTM, the output of one layer of LSTM cells is fed as input to the next layer. This stacking allows the network to learn more complex features at different levels of abstraction. Each LSTM cell in the network has a hidden state of size 50, i.e. the internal memory of each LSTM cell (which is crucial for capturing and retaining information over time) can store and process information using 50-dimensional vectors.
Here are further details about the models :

* Activation function: sigmoid
* Loss: Binary Cross Entropy
* Optimizer: Adam with a learning rate of 0.001
* Batch size: 128
* Number of epoch: 10

C. Class imbalance mitigation

Lastly, we trained all models with class weights. For each model, the F1 score was always higher with class weights compared to the same model trained without class weights, while keeping a similar accuracy. We provide an example of this improvement in Table III, for the model whose input features are the magnetic field coordinates.

Model	Accuracy	F1 score
Without Class weights	99.60%	88.30%
With Class weights	99.80%	94.15%

TABLE III: Effect of class weights on model performance

D. Models' performance

From table IV, we can see that the best model is the one that predicts based on both magnetic field coordinates and the solar distance.

CMEs seem to be detected with abnormal changes in the magnetometer data. Adding the VEX-Sun distance features allows for the model to compensate disturbance from Venus' magnetosphere.

Input Features	Accuracy	F1 score
EDAC	98.00%	0.00%
BT	95.67%	18.00%
BX BY BZ	99.80%	94.15%
BX BY BZ RSC	**99.84%**	**95.50%**

TABLE IV: Performance of the LSTM models

VII. Conclusion

This project aimed to study the possibility of recognizing CME events using EDAC counter onboard spacecrafts, focusing on the Venus Express mission. The motivation to use EDAC was because the counter are present onboard most spacecraft and thus CME event could be detected without the need for specific instruments. Unfortunately, we determined EDAC counters are not a significant indicator of the presence of a CME event near the spacecraft. We then changed our focus towards magnetometer data acquired during the Venus Express Mission. We observed a strong correlation between these magnetic field disruption and CME events, allowing us to build machine learning models to robustly detect these Solar events.

VIII. Future work and limitations

Our model only predicts CME events from the Venus Express spacecraft and we could try to see how our model would behave on other spacecrafts in order to know if our model over-fitted on the Venus orbit or it generalizes to other orbits as well. However, the CME labelization for the other spacecrafts such as the Mars Express and Rosetta are incomplete, so training or testing our model on them would have poor integrity.

IX. Ethical considerations

For our project, the main stakeholder would be ESA since they own the dataset we used to train our model. Another one could be the environment since spacecrafts are usually not environment friendly but the dataset was collected for other purposes.

This research strictly respects ethical principles, ensuring responsible and transparent use of data throughout the study (see Figure 6). The primary dataset utilized in this project is sourced from the ESA website, which provides open-access information. The use of openly available data aligns with principles of transparency and accountability in scientific research. Moreover, while some additional data from ESA is not currently public, it is slated for future release and it is important to note that this data does not involve any confidential or private information. The focus of our project is the detection of CMEs, inherently it avoids any violation of individual privacy as it pertains solely to space weather phenomena. We acknowledge and respect the data-sharing policies of ESA, ensuring that our research maintains the confidentiality and privacy of any non-public information.

References

[1] E. W. Knutsen, O. Witasse, B. Sanchez-Cano, M. Lester, R. F. Wimmer-Schweingruber, M. Denis, J. Godfrey, and A. Johnstone, “Galactic cosmic ray modulation at mars and beyond measured with edacs on mars express and rosetta,” 2021, https://doi.org/10.1051/0004-6361/202140767.

[2] B. Sánchez-Cano, O. Witasse, E. W. Knutsen, D. Meggi, S. Viet, M. Lester, R. F. Wimmer-Schweingruber, M. Pinto, R. Moissl, J. Benkhoff, H. Opgenoorth, U. Auster, J. de Brujine, P. Collins, G. D. Marchi, D. Fischer, Y. Futaana, J. Godfrey, D. Heyner, M. Holmstrom, A. Johnstone, S. Joyce, D. Lakey, S. Martinez, D. Milligan, E. Montagnon, D. Müller, S. A. Livi, T. Prusti, J. Raines, I. Richter, D. Schmid, P. Schmitz, H. Svedhem, M. G. G. T. Taylor, E. Tremolizzo, D. Titov, C. Wilson, S. Wood, and J. Zender, “Solar energetic particle events detected in the housekeeping data of the european space agency’s spacecraft flotilla in the solar system,” 2023, https://doi.org/10.1029/2023SW003540.

[3] Esa archives. [Online]. Available: https://archives.esac.esa.int/psa/ftp/

[4] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, “Tensorflow: A system for large-scale machine learning,” 2016.

[5] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” 1997, https://www.researchgate.net/publication/13853244_Long_Short-term_Memory.

[6] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, “Pytorch: An imperative style, high-performance deep learning library,” 2019.

<table>
<tr><th>Dataset</th><th colspan="2">Beneficence</th><th colspan="2">Non-maleficence</th></tr>
<tr><td rowspan="9">□ Who created the dataset?
The European Space Agency.
□ For what purpose was the dataset created?
To collect space data.
□ What mechanisms or procedures were used to collect the data?
The data was provided mainly from the public ESA dataset. While some additional data from ESA is not currently public, it is slated for future release.
□ Who was involved in the data collection process?
The European Space Agency.
□ Over what timeframe was the data collected?
The data was collected from 2006 to 2014.
□ Was any preprocessing of the data done?
There was a tedious work done on the data preprocessing.
□ Are there any missing data or data errors?
There is data missing such as when the spacecraft entered hibernation.
□ Where is the data stored?
The data is stored on our private computers and for the small files, on GitHub as well.</td><td colspan="2" rowspan="2">□ What are the expected benefits of analyzing this data? For whom?
Predicting space weather forecasting so we can protect the spacecrafts against radiation.</td><td>Risks</td><td>Mitigation</td></tr>
<tr><td colspan="2">□ Does the dataset contain unsafe data (violence, nudity…)? No
□ What kind of impacts can errors in the data or in the analysis have? Wrong prediction of space weather, not severe impact
□ Could the data or the conclusions from the analysis be used in harmful ways? No</td></tr>
<tr><th colspan="2">Sustainability</th><th colspan="2">Fairness</th></tr>
<tr><td>Risks</td><td>Mitigation</td><td>Risks</td><td>Mitigation</td></tr>
<tr><td colspan="2">□ What is the carbon and water footprint generated by the storage of the data and by the computation in the analysis process? The data was of very reasonable size and the analysis process does not require abnormal computation expenses.
□ What type of human manual labor is involved in the data (e.g. labeling)? The data is directly fetched from the ESA website, so it does not require human manual labor.
□ Does the data or the analysis require updates? Since we used the LSTM model, it is preferable to have the latest timeseries data since our model depend on the solar cycle.</td><td colspan="2">□ Is the data representative from a larger set (population)? How are subgroups represented? No
□ What kinds of biases may affect the data? We only trained our model on Venus Express so if the model is used on other spacecrafts, since the distance from the sun will differ, the results would not be as accurate.
□ Can the outcomes of the analysis be different for different groups? It could be different for different spacecrafts.
□ Could the data or analysis results contribute to discrimination against people or groups? No</td></tr>
<tr><th colspan="2">Privacy</th><th colspan="2">Empowerment</th></tr>
<tr><td>Risks</td><td>Mitigation</td><td>Risks</td><td>Mitigation</td></tr>
<tr><td colspan="2">□ Does the data contain personal or sensitive information? No
□ Can personal or sensitive information be derived or inferred from the data or from the analysis? No</td><td colspan="2">□ How are the people concerned involved with the data or the analysis: have they been notified, have they consented? No one is involved with the data or analysis.
□ Are the people concerned able to make choices (e.g. revoke consent, modify or delete data) regarding the data or the analysis? –</td></tr>
</table>

Fig. 6: Assessing Ethical Risks Table