Classification of High Energy Particle Precipitation Events Using Computer Vision

Caleb Bettcher

caleb.bettcher@colorado.edu

Robert Marshall

robert.marshall@colorado.edu

Problem Statement and Research Objectives

Earth's radiation belts are home to a large number of orbiting particles with varying energy levels. Certain events, such as Electromagnetic Ion Cyclotron (EMIC) waves, disrupt the balance of these regions and result in particles falling towards the Earth's atmosphere. On a large scale, Energetic Particle Precipitation (EPP) can deplete atmospheric ozone, interrupt telecommunications, and destroy sensitive electronic components across the energy sector causing serious effects to society's delicate infrastructure and our planet. Data collected by the twin ELFIN satellites (launched in 2018) aims to provide researchers with a deeper understanding of this region of space and assist in developing techniques to better predict large solar weather events and mitigate complications with infrastructure systems. We aim to further this field of research by constructing Machine Learning (ML) and Computer Vision (CV) models to analyze the vast amount of data from the ELFIN satellites to determine the time frame and source of EPP events. Armed with our models and findings, we hope future researchers will develop a better understanding of how to predict EPP-causing events and protect against them.

Methodology

In order to use CV to recognize signatures of particle precipitation data, we must construct a training set from images of our data. We've chosen to use a graph displaying the J_{prec}/J_{trap} ratio (the number particles which fall into the atmosphere over the number of particles which are trapped in orbit) by energy level over time which highlights several characteristics of various EPP sources.

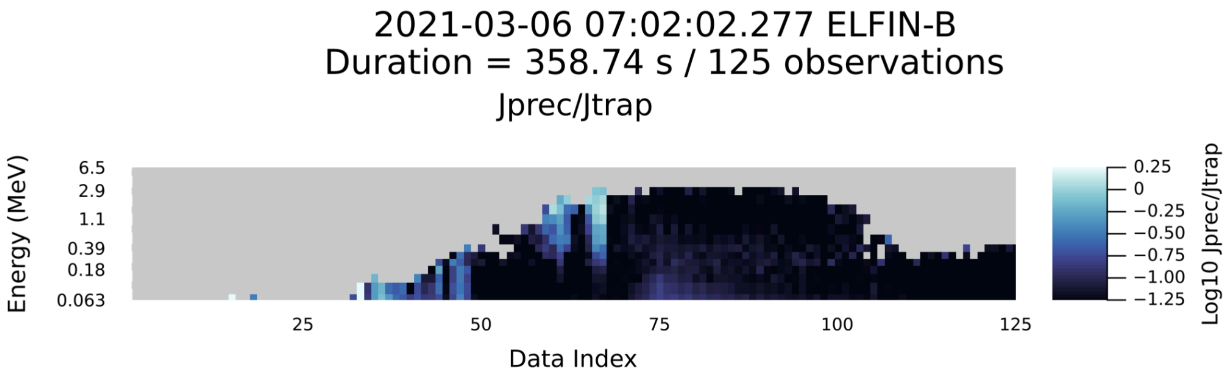

Figure 1. Two EMIC wave signatures around 60 and 65 on the Data Index axis indicated by the elevated precipitation rate at higher energy levels with a declined rate as energy level decreases.

To provide a proof of concept and determine if this approach is worth further pursuing, we will first focus on EPP events caused by EMIC waves. We've chosen to place a particular emphasis on EMIC wave-induced events due to their tendency to cause a comparatively large amount of precipitation of the highest energy particles which have the greatest impact on the aforementioned effects of EPP. A comparatively large amount of research and documentation has been done specifically on EMIC-induced precipitation measured on ELFIN, allowing us to construct an effective training dataset. After isolating EMIC wave events, plotting them against a standardized graph, and introducing seeded randomness to improve model performance, the data will be labeled and used to train a CV model. Initially, we have decided to use the YOLO CV model due to its performance, versatility, and simplicity. If our smaller scale test produces promising results for EMIC waves, we will introduce other sources of EPP and increase the complexity of our model to cover the entirety of the ELFIN dataset before analyzing results and presenting our findings.

Technical Background

Coming into this research position, I had limited prior exposure to plasma physics. However, on the computer science side of this project, I've had opportunities to hone my coding abilities at a professional level through various software development internships and had previous experiences working with various ML models through a combination of classes and personal projects such as building a LLM revolving around global sentiment towards climate change to predict political affiliation. Being a subset of ML, the key concepts behind CV are the same; as such, adjusting my thinking towards CV required less work than understanding the complex physics behind the data being analyzed. Fortunately, a deep understanding of graduate level plasma physics was not a required skill set and, due to the nature of classifying EPP events based on their unique data signatures, a comparatively simple understanding of the physics behind these processes was sufficient. This fact allows me to stay focused primarily on the computer science and ML aspect of the research, with ample resources available to bring me up to speed on necessary physics information.

In the context of using Computer Vision for data classification, our group with the LAIR lab is the among first to apply CV to space physics. As a result, outside of the standard documentation for CV models such as YOLO by Ultralytics, there is a limited amount of information available for the application of CV models to classify satellite data. Fortunately, one member of our research group, Paraksh Vankawala, is using a similar approach. Using a combination of Paraksh's research and our own additions, we hope the models we train are able to accurately and reliably classify various sources of EPP.

Facilities, Equipment or Other Requirements

Our project will be coded using various languages such as Python and Julia, and will utilize various third-party packages to enhance these languages and make use of existing ML architectures. To conduct our research in accordance with the guidelines for practicing good research, any code written and data used will be replicable, documented, and follow general coding guidelines including using version control and package management. Beyond building ML training sets and initial setup, we will use supercomputing resources to train the CV models in order to achieve the best results in a reasonable amount of time. CU provides supercomputing resources for research through the ALPINE supercomputer, and the aerospace department has the BLANCA supercomputer, on which our lab owns several nodes. By leveraging these resources, we hope to achieve improved results and save valuable research time.

Timeline

By the end of December and the first semester of school, we hope to have tested the first versions of our YOLO model trained from preexisting EMIC wave data and expect to see promising results. By February, we hope to experiment with altering our initial training set and various training parameters to improve model performance. If our initial models produce promising results, we will include other sources of EPP into training sets and perform a similar process. By April, we will write software tools which can be used to automatically apply trained models over the entirety of the ELFIN data and present results of the model's analysis in a usable form. Finally, by the end of May and the spring semester, we will condense our analyzed results into a presentable product and determine findings.