**Challenge: Develop the oracle of DSCOVR**

The National Oceanic and Atmospheric Administration’s (NOAA’s) space weather station, the Deep Space Climate Observatory (DSCOVR), can measure the strength and speed of the solar wind in space, which enables us to predict geomagnetic storms that can severely impact important systems like GPS and electrical power grids on Earth. It acts like a sensor buoy at sea that warns of an oncoming tsunami—DSCOVR can warn forecasters 15 to 60 minutes before a storm of particles and magnetic field, known as a coronal mass ejection (or CME) reaches earth. However, it continues to operate past its expected lifetime and produces occasional faults that may themselves be indicators of space weather (the launch was in Feb. 11 2015, the arrival l at Sun–Earth L1 Lagrange point in Jun 8, 2015, and was expected to last 5 years)

A screenshot of a computer

Description automatically generated

DSCOVR orbits about a million miles from Earth in a unique location called Lagrange point 1, which basically allows it to hover between the Sun and our planet. From that vantage point, DSCOVR measures the plasma that may cause geomagnetic storms hours before it reaches us– ideally providing an early warning of what’s coming our way. The time that it takes for that plasma to reach Earth and trigger a geomagnetic storm might be anywhere from about 15 minutes to a few hours.

A diagram of the sun and planets

Description automatically generated

DSCOVR uses measurements of the solar wind density, temperature, speed, and magnetic field to run computer simulations of the Earth's magnetic field and atmosphere. Based on those simulations, NOAA forecasts when a geomagnetic storm will occur and how strong it will be. The strength of the geomagnetic storm is measured on a scale called the Planetary K-index (Kp).

The instrument onboard DSCOVR that measures the solar wind's magnetic field continues to function very well, the instrument that measures the solar wind density, temperature, and speed has lost sensitivity and experiences faults and anomalies from time to time.

Challenge: use the **"raw" data** from DSCOVR—faults and all—to predict geomagnetic storms on Earth.

Currently, NOAA relies on stable, well-calibrated level 2 data for its forecasts. We challenge you to develop your own geomagnetic activity forecast using the raw DSCOVR data directly as input.

The DSCOVR Faraday Cup is the instrument that provides the solar wind density, speed, and temperature used by NOAA to run its forecast models. NOAA refers to the densities, speeds, and temperatures as "level 2" data. These quantities are not measured directly, however; the instrument actually measures the entire spectra of solar wind particles over time and then computes those quantities from the spectra. These spectral data are referred to as "level 1" data, or sometimes as "raw" data.

To work well, a space science experiment must be stable and well-calibrated to translate the "raw" data into the more useful "level 2" data. DSCOVR no longer meets this criteria; in its old age, certain electrical anomalies and faults have been observed that shift the calibration and introduce noisy signals in unpredictable ways into DSCOVR raw data. When these faults occur, NOAA generally attempts to identify them and switches to a backup weather station rather than attempting to recalibrate or mitigate the error to produce the level 2 data accurately.

DSCOVR uses 2 main space weather instruments

* Magnetometer: Measures solar wind magnetic field vector.
* Faraday Cup: Provide real-time measurement of solar wind proton density, speed, velocity, temperature, etc... NASA and NOAA (SWPC and OSPO) are actively monitoring Faraday Cup behavior in various solar wind conditions, and developing or updating flight software solutions to optimize the instrument behavior. The most recent modification was performed in August 2020, and additional work is ongoing.

In our challenge we will deal with PlasMAG detectors and FC. The PlasMAG detectors do not take data all of the time, and the Faraday cup does not make measurements over its full range every minute. Whenever and wherever no data are available, the field is filled in with an integer 0. **We recommend converting these to "NaN" in your computing environment after you load the data.**

DSCOVR SPACE WEATHER DATA PORTAL -> <https://www.ngdc.noaa.gov/dscovr/portal/#/#swi>

Initial data that we should use for the challenge: <https://www.spaceappschallenge.org/develop-the-oracle-of-dscovr-experimental-data-repository/>

Regarding the data of the experimental repository:

* **1st Column:** Refers to the minute reading in UTC
* **2nd to 4th Columns:** Refers to the (x, y, z) components of the magnetic field vector (nT units) in the GSE reference frame. Here's a detailed explanation of these measurements: The interplanetary magnetic field (IMF), which is the magnetic field from the Sun, is measured in Bx, By, and Bz components, which are vector quantities with Bx and By oriented parallel to the poles (ecliptic) and Bz oriented perpendicular to the ecliptic. The Bt value indicates the total strength of the three Bx, By and Bz components of the IMF. When the Bz component of CMEs is positive (northward), it has less effect on the Earth’s magnetosphere. However, when the Bz component is negative (southward), it opposes the direction of Earth’s magnetic field, causing changes and reconnection of IMF and the Earth’s magnetic field
* **5th+ columns:** These are the "raw" measurements of the spectrum from the plasma detector so they are the "level 1" data. We're reading it as each value is the flow strength of the solar wind, with each column being an interval in a range of flow speeds at which the solar wind is traveling at.

The magnitude of a geomagnetic storm measured in **planetary K index or Kp index**, which is the **mean of K-indices of twelve observing stations or magnetometers**. Kp-index ranges from 0 to 9; 0 being the lowest and least consequential, and 9 being the highest and most consequential. Kp-indices equal to or above 5 are considered geomagnetic storms. The higher the Kp-index, the more the geomagnetic storm has an effect on the Earth’s system.

Reading this paper is a must (the first pages have helped me greatly for understanding the problem at hand): <https://ccmc.gsfc.nasa.gov/RoR_WWW/SWREDI/contest-presentations/2017/CCMCPaper_ShreeyaKhurana_Final.pdf>

**Next problem**: How do we get the Kp-index? (it is not in the current csv and finally it seems that it is what we want to predict). Some links:

* <https://www.swpc.noaa.gov/products/planetary-k-index>
* <https://kauai.ccmc.gsfc.nasa.gov/DONKI/search/>
* Geomagnetic planetary three-hour index Kp since 1932: <https://kp.gfz-potsdam.de/app/files/Kp_ap_Ap_SN_F107_since_1932.txt>

**CANADA DATA. What is it’s purpose?**

There are 4 different sources related with Canada data:

* [Geospace Observatory (GO) Canada:](https://donnees-data.asc-csa.gc.ca/en/dataset/0176458c-553b-48b4-a5e2-492022c81e85) project that brings together a variety of space weather research tools and infrastructure. The data contained here is basically an inventory of the different tools and instruments as well as the link to each of the instruments. **The data that we can obtain here won’t be useful,** and we will have to drill down to each instrument. This is not needed, as the Datathon org. Has provided us with the instrument to check.
* CARISMA Magnetometer Network: It is **considered part of the GO Canada suite of projects (prior point) but it is being specifically mentioned here due to its pertinence to this challenge** (so, as said, we cant forget about the previous point). Stands for Canadian Array for Realtime Investigations of Magnetic Activity (CARISMA). Series of ground-based sensors that detect disturbances in the Earth's magnetosphere (often related to solar winds). Here, we just have data until 2009 and we have other less specific data until 2017. We have discarded the possibility of using this data.

A screenshot of a computer

Description automatically generated

* [Canadian Solar Flux Archive (F10.7):](https://www.spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-3-en.php)

This data is of the solar flux. Contains measurements of the 10.7cm Flux and daily records of flux monitor output. Each measurement of the 10.7cm Solar Flux is expressed in three values: the observed, adjusted and URSI Series D values.

The data has the next features:

* The Julian Day of the measurement
* The Carrington Rotation Number
* The year
* The month
* The day
* The observed flux
* The adjusted flux
* The Series D flux.

Maybe it can be useful for bringing some of the features (there are 3 records per each day (at 18, 20 and 22 hrs)). Data can be found here: <https://www.spaceweather.gc.ca/forecast-prevision/solar-solaire/solarflux/sx-5-flux-en.php>

* **Cassiope:** It is also part of the GO Canada. The Canadian CASSIOPE satellite, operated by the University of Calgary, carries the Enhanced Polar Outflow Probe (e-POP) suite of scientific instruments to study the ionosphere, where space meets the upper atmosphere. The instruments collect data about the effects of solar storms and, more specifically, their harmful impact on radio communications, satellite navigation and other space and ground-based technologies. More recent data is from 2020. In my opinion it won’t be really useful as the estimated useful life ended in 2020, so the failures should have started (or increased) on 2020.

A screenshot of a computer

Description automatically generated

**Other resources:**

* <https://www.swpc.noaa.gov/products/real-time-solar-wind> just a visualization tool of the measurements in Level 2.
* <https://www.swpc.noaa.gov/products/geospace-magnetosphere-movies> cool visualization tool of some real time measurements
* <https://www.swpc.noaa.gov/products/3-day-geomagnetic-forecast> thats the format in which they predict the forecast