Canadian Space Agency | Agence spatiale canadienne

GO Canada – Geomagnetic Data Integration

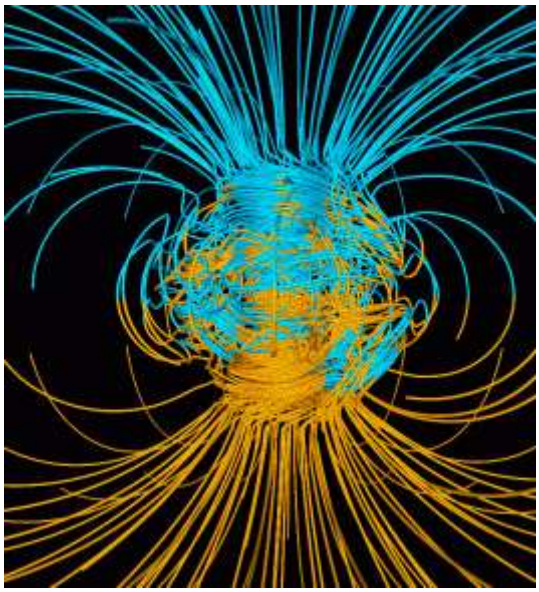

Figure 1. Model of the Earth's Magnetic field.
Credit: Wikimedia

THE CHALLENGE

Assist scientists in improving Canadian geomagnetic data sets - essential for monitoring the Earth's magnetic field responsible for protecting the population from solar flares - by (1) helping point out erroneous data collected by magnetometers across Canada and/or (2) determine geomagnetic activity across Canada and map it.

Access the data subset :
ftp://ftp.asc-csa.gc.ca/users/OpenData_DonneesOuvertes/pub/Space%20Apps%20Challenge%202019/GO%20Canada/

THE ISSUE

The data collected by magnetometers sometimes presents erroneous entries. This might be due to many factors such as thunderstorms occurring near the measurement site, and the presence of nearby high-voltage lines or electrical equipment.

The GO Canada magnetometers are widespread and irregularly dispatched across Canada. CSA would therefore also benefit in mapping the Canadian geomagnetic activity and use it to estimate its condition at any location in Canada.

THE NEEDS (you can choose to tackle one or more)

1. Identify parasitic magnetic fields

Many factors can influence the behaviour of a magnetometer, such as power lines, electrical equipment or electric storms. In order to avoid confusing an artificial magnetic fluctuation with a real geomagnetic storm, CSA needs to be able to point out erroneous magnetic measurements. The main characteristic of these parasitic disturbances is that they are local, unlike the disturbances caused by a solar storm. Good interpolation would make it possible to estimate geomagnetic activity across Canada and flag measurements that are not coherent with other sites or observatories. Such local perturbations may be caused by lightning, or equipment interfering with the measurements at a particular site. It would be valuable to be able to flag such data that is not scientifically significant.

Potential Output

Developing a validation algorithm for a dataset : you could produce a script that takes the dataset1.txt data set as input and returns an output report of the measurements whose reliability is questionable, including a quantitative index of the reliability of this measurement. The format (layout, extension...) of this report is up to you.

How to Get Started

Identify false variation in the magnetic field. This can be fairly easily identified by exploiting the locality of the parasitic disturbances comparing it with measurements collected at nearby sites.

- If geomagnetic activity varies similarly for a few sites located hundreds or even thousands of kilometres away, this variation can be attributed to solar activity.
- On the other hand, if a peak of magnetic activity is observed at a location, but the magnetic field remains constant 100 kilometres away during the same period, it should be suspected that the peak is not physically related to solar activity.

2. Mapping Canadian geomagnetic activity

CSA needs to properly model geomagnetic activity on a Canadian scale.

GO Canada magnetometers are irregularly dispersed throughout Canada. They are sometimes separated by thousands of kilometres, which increases the difficulty of properly mapping the magnetic field to determine the geomagnetic activity between existing sites. The second objective of this challenge is to produce a map of the Canadian Geomagnetic Activity, to be able to estimate it anywhere in Canada.

Potential Output

You could perform 2D mapping with irregular mesh size using an interpolation.

How to get started

The challenge is composed of two steps: the solution must be able to read in the input files, identify data that is not correlated with the rest of the dataset, and output the results graphically, enabling the user to browse the data in space and time, and to produce virtual measurements from surrounding observations.

Dig the data to identify true magnetic activity and then map it.

- Obtaining a good map of magnetic field variations makes it possible to estimate the variation of the magnetic field at any time and in any place on Canadian soil.
- A set of training data is available to participants so that they can test the different methods and choose the most accurate one to interpolate the missing site from the "holed" dataset (dataset2_holed.txt).

To determine the best method, you should make sure that:

- The chosen method provides the minimum error using the training data set
- Magnetic field variations are never negative on the mesh (by definition)
- The mapping is not overfitted.

BACKGROUND

Global interest in space weather has grown significantly in recent decades, largely due to the potential consequences of major space weather events on the world's population. The Canadian Geospatial Observatory (GO Canada) is a program developed by the Canadian Space Agency in collaboration with Natural Resources Canada (NRCAN) and several Canadian universities to bring together a range of measurements from more than 60 stations located in Canada and the northern United States. Among these instruments is a vast network of magnetometers that identify, classify and even predict geomagnetic storms, a major space-meteorological threat. This network of magnetometers operates continuously and without supervision, so there is no way to recognize measurement errors caused by parasitic magnetic fields or power failures. It is also impossible to know precisely the geomagnetic activity in a region without measuring instruments.

The input files

The input files were produced by processing data from NRCan and the University of Alberta's CARISMA arrays. To simplify access to the data for this particular challenge, the dataset was reduced to three files, containing magnetic variation indicators computed from the source data. The location of the sites are provided with the site ID on the second line, and the maximum magnetic variations start on the 5th line. The magnetic variation indicator is related to the geomagnetic activity and is computed from the root mean square of the maximum hourly variation of the x,y,z magnetic field components taken from the GO Canada dataset. The date column is not accurate as the dates and times were shifted so the candidates do not refer to the source data to complete the challenge.

dataset1.txt contains some data that was artificially altered, and the challenge is to automatically identify suspicious data and estimate confidence levels.

dataset2_holed.txt contains unaltered data, but a location is missing. The challenge is to generate the best estimate of the actual data from the other data available.

dataset2_full.txt is a training set, containing the same type of data, untampered, for a different period.

OTHER RELEVANT INFORMATION AND DATA SETS

CSA's Expert Workshop

Canadian Geospatial Observatory (GO Canada)

Mathematical basis of 2D interpolation

2D regression methods and examples using Python

Wikipedia page for polynomial regression

Wikipedia page for overfitting

Technical information on magnetometers and related data sets:

- CARISMA (University of Alberta)
- CARISMA Data
- Natural Resources Canada
- NRCAN data

Additional information on space weather:

- Space Weather
- Space Weather Canada