(L)TI (T)raite de l'(I)nformation

Participants:

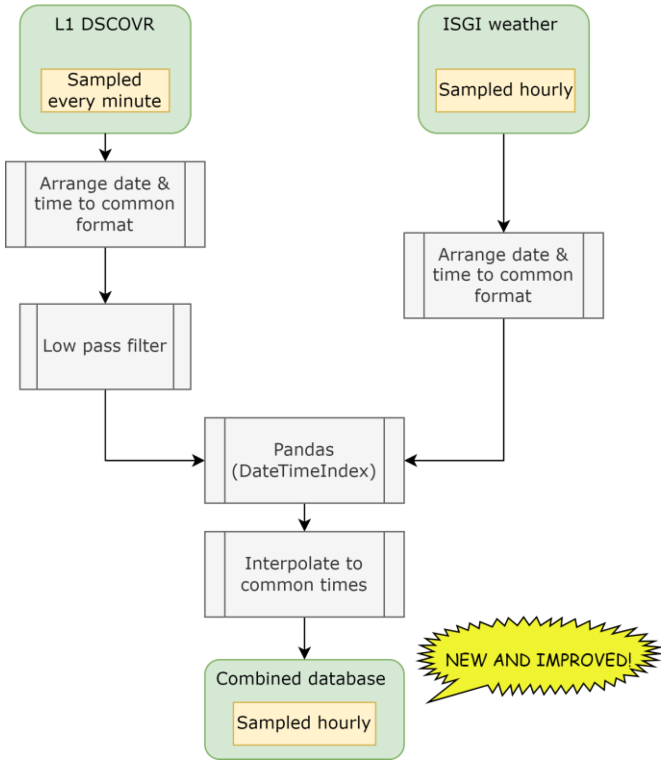
- Alexandre Beaulieu
- Mathieu Bergeron
- Samuel Fortin

DSCOVR DATA: PROCESSING, INTERPRETATION & PREDICTION

INFORMATIQUE + GÉNIE

SOFTWARE + ENGINEERING

DATA WRANGLING & PREPROCESSING



Data sources

DSCOVR dataset (L1 data)

ISGI hourly space weather (aa, am, Kp, Dst, PCN, PCS, AE, AU, AL, AO, SC, SFE)

Transformation

In both databases, datetime is not in the same format. Steps are taken to make them similar.

Since we aim to resample to an hourly rate, low pass filtering of L1 DSCOVR data is applied to reduce the SNR as much as possible.

Combination

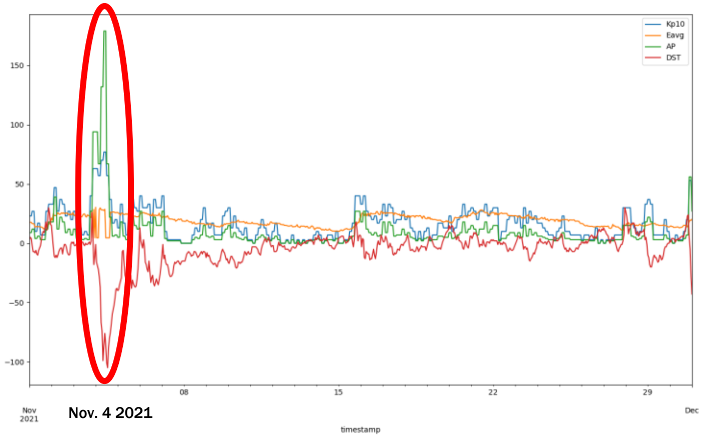
Using DateTimeIndex functionality of the pandas library, both databases are merged.

Resampling

The merged dataset is then re-sampled to common times at an hourly rate. When the time is not just, are then resampled to the desired rate.

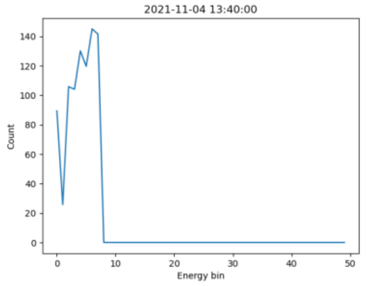
ANOMALY DETECTION

Space weather metrics during November 2012

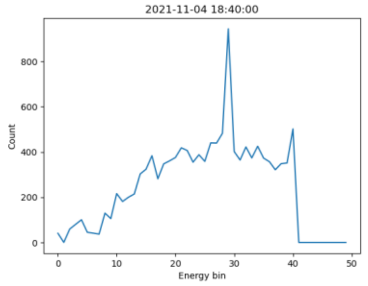


L1 DSCOVR data columns 4 to 54

Spectrum during solar event



Spectrum after the solar event



Hypothesis:

Solar events can cause errors in the sensors of the DSCOVR probe.

Example:

On November 4th, 2021 , a solar storm caused what we interpret as saturation of the spectrum sensor of the probe in all bands. We observed this as abnormally high variation in the average energy (E_{avg}) and the energy variance of the spectra.

$$E_{avg} = \frac{1}{\sum_{i=0}^{50} N(i)} \sum_{i=0}^{50} (N(i) * (i + 0.5))$$

$$E_{var} = \frac{1}{\sum_{i=0}^{50} N(i)} \sum_{i=0}^{50} \left(N(i) * \left(i + 0.5 - E_{avg}\right)^2\right)$$

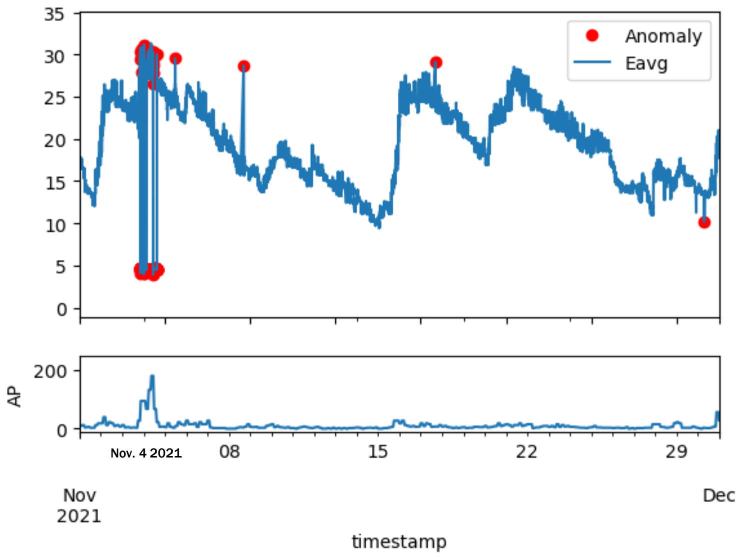
Observing and comparing the energy measurements between the spectra during and outside of the event shows differences in their distributions.

Note: Kp10 in these plots is 10 times Kp. It is the format in which the IGSI saves the metric.

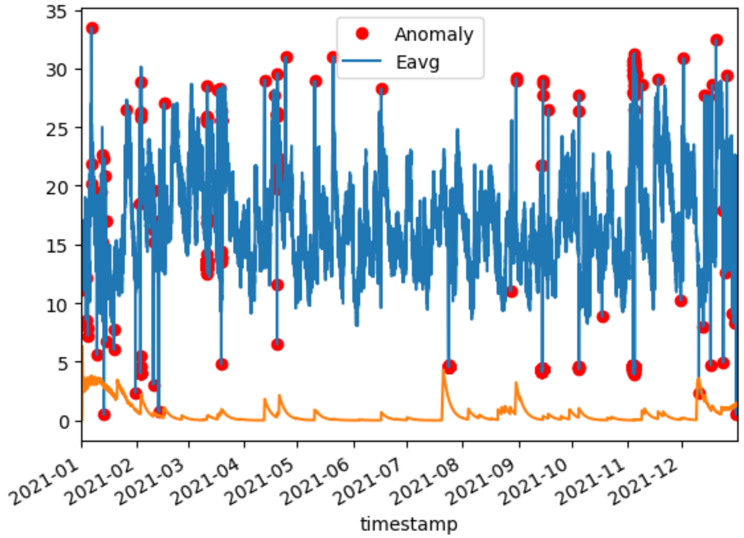
Decision:

Ap seems to be (at a glance) less noisy than Kp. From now, it will be our reference measurement to determine the level of solar activity. A peak high Ap is interpreted as a solar event.

GENERALISATION



Anomalies shown superimposed on the E_{avg} signal and aligned in time with the A_p signal. We see that the number of anomalies is high when there is a high density of peaks.



Results of the anomaly detection algorithm over a year. The orange signal is a proxy for the density of missing data in the spectra. We see that the anomalies in the signal do not always correspond to periods of high missing data.

Anomaly detection

Assuming the spectrometer signals are unreliable during solar events, there should be a higher density of peaks and anomalous samples in the E_{avg} and E_{var} signals for its duration. Another hypothesis was that solar storms yielded a large amount of missing data. The orange time series of the rightmost plot do not support this hypothesis.

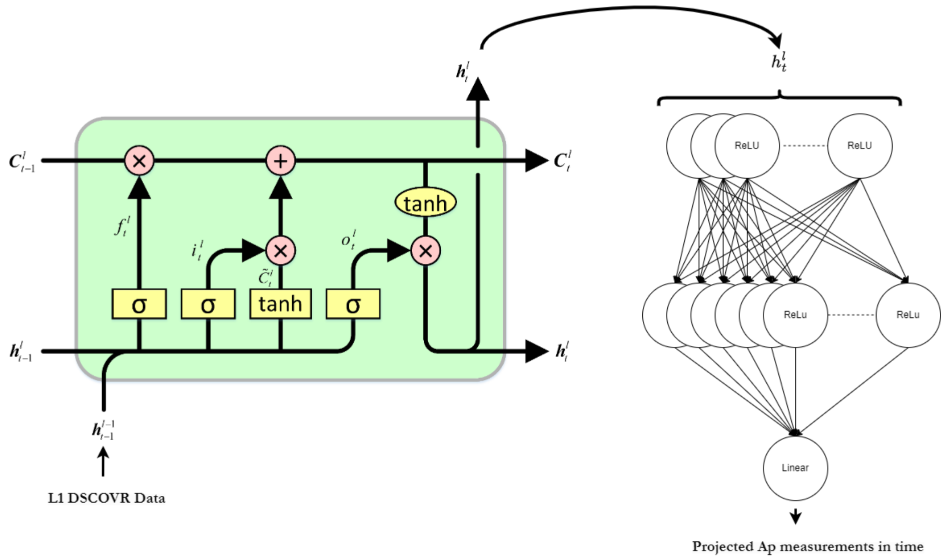
Tool

The DBSCAN algorithm is used to detect anomalies in the E_{avg} & E_{var} signals.

Conclusion

While it is true that the Nov 4th solar event produced a large number of anomalies in the data, the anomaly exhibited also tagged a many samples that do not correspond to solar events

MACHINE LEARNING MODEL



Hypothesis:
A link exists between the present measurements of space weather, their future value and the L1 DSCOVR dataset.

Goal:
Predict the value of A_p in the future over a fixed window.

Means:

Why a Long Short-term Memory (LSTM)?

The network topology allows it to make links with its past inputs. For physical system representation, this property is a requirement.

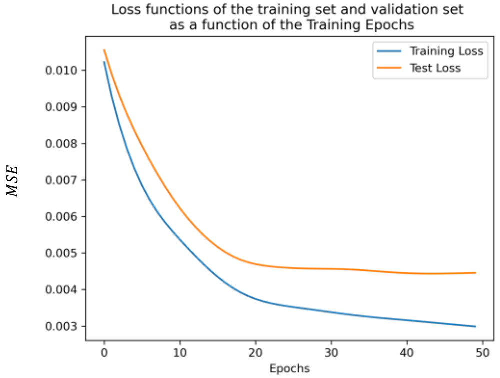
Why a Feed Forward Neural net (FFNN)?

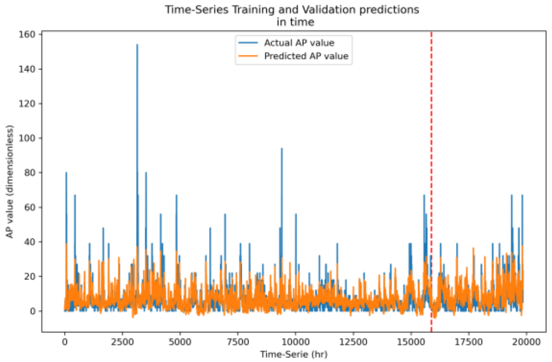
Adding a FFNN to the LSTM layer allows for added complexity of the output. This is principally because the DSCOVR data is not stable*.

*Source: tutorial

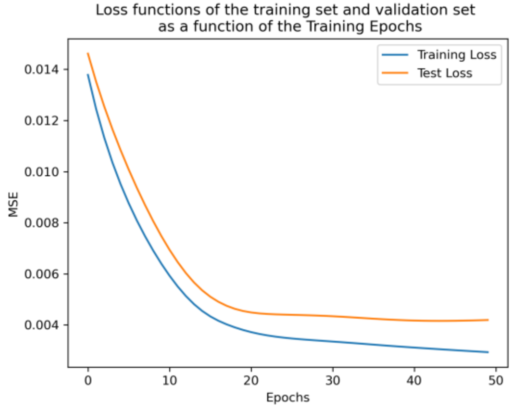
TRAINING

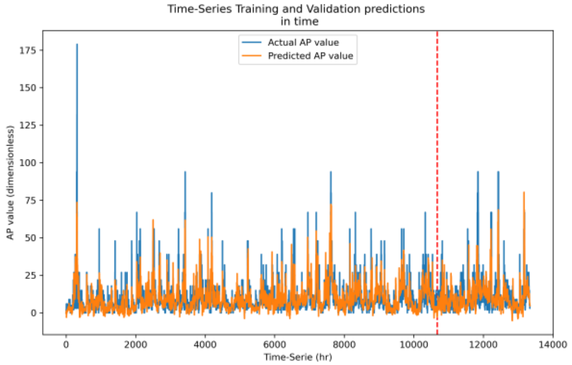
DSCOVR data only





Concatenated DSCOVR & weather data





Training & Testing set

Imported from the pre-processing phase.

Input

1) DSCOVER L1 data.

2) Concatenation of existing space weather & DSCOVER L1 data.

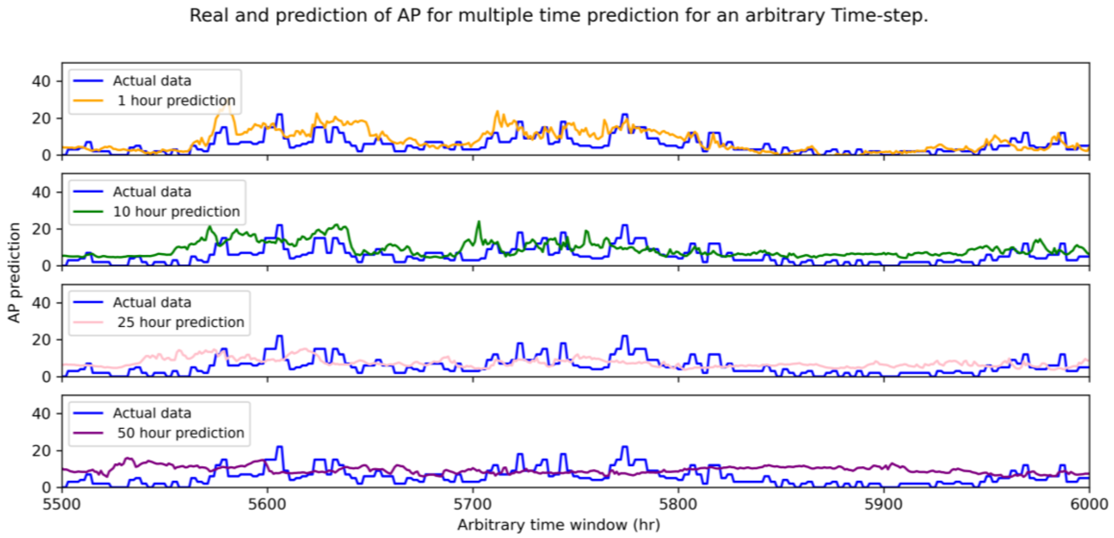
[[L1 DSCOVR], aa, am, Kp, Dst, PCN, PCS, AE, AU, AL, AO, SC, SFE]

Ouput

Prediction of the Ap over the next 50 hours from the time of the last data point.

RESULTS

Prediction using ONLY the DSCOVR data as input



Prediction using the DSCOVR data & past space weather metrics as inputs