Using an LSTM and Classification Methods to Determine Risk of dB/dt Threshold Crossings as Proxy for Geomagnetically Induced Currents

Michael Coughlan[1], Amy Keesee[1,2], Victor Pinto[2], Hyunju Connor[3], Jeremiah Johnson[4]

[1]Department of Physics and Astronomy - University of New Hampshire, [2]Institute for the Study of Earth, Oceans, and Space - University of New Hampshire
[3]Department of Physics and Geophysical Institute - University of Alaska Fairbanks, [4]Department of Electrical and Computer Engineering - University of New Hampshire

MAGICIAN
Machine Learning Algorithms for Geomagnetically Induced Currents in Alaska and New Hampshire

Introduction

- The interaction between the solar wind and the Magnetosphere can produce Geomagnetically Induced Currents (GIC's) on the ground, which are capable of causing power outages and damage to crucial infrastructure.
- The ability to predict when and where these events may occur could allow us to avoid the worst of this damage.
- The use of physics-informed machine learning models can offer a computationally inexpensive method of predicting GIC events using horizontal dB/dt as a proxy, though most models thus far have fallen short of consistently accurate predictions. dB/dt was defined as:

$$\frac{dB_H}{dt} = \sqrt{\frac{dE}{dt}^2 + \frac{dN}{dt}^2}$$

With N and E the North and East components of the magnetic field respectively.

- Here, a Long-Short Term Memory (LSTM) model was used to determine the risk of dB/dt going over thresholds of 9, 18, 42, 66, and 90 nT/min for the Ottawa (OTT) ground magnetometer station.
- Three storms were chosen for testing and removed from the training set: March 30, 2001 (~ -211nT), December 14, 2006 (~ -437nT), & August 05, 2011 (~ -126nT).
- The storms were chosen for several reasons; they represent different storm intensities, they occurred at different points in the solar cycle, and there are minimal gaps in the data that needed to be interpolated over.

Method/Data

- The model was trained exclusively on storm time data as defined by a SYM-H value of -50 nT or less for a minimum period of 2 hours.
- The storm data was extracted from a combined data frame of OMNI data and Supermag data from the Ottawa (OTT), mid-latitude station.
- The input features included solar wind speed (VT, Vx, Vy, Vz), IMF_GSE (BT, Bx, By, Bz), proton density, dynamic pressure, solar wind electric field, SYM-H, horizontal magnetic field (N,E), and ground magnetometer sin(MLT) and cos(MLT).
- The LSTM layer utilized 30 minutes of time history to determine if the dB/dt value would go above a series of thresholds, between 30 and 60 minutes into the future.
- The machine learning model consisted of a single LSTM layer with 'RELU' activation and a Dense output layer using 'softmax' activation, implemented using TensorFlow with the Keras backend.
- The 'softmax' activation layer allows the model to interpret the inputs to the layer as discrete probability distributions, allowing us to interpret the outputs of a node as the probability of that node occurring. In this case the output of the node is the probability that dB/dt will cross the given threshold.
- The softmax activation can be described:

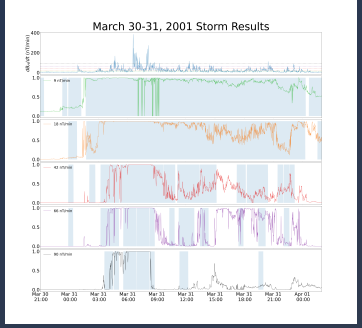
$$softmx(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

- The Probability of Detection (POD), Probability of False Detection (POFD), Frequency Bias (FB), and Heidke Skill Score (HSS) were calculated. These metrics rely on the comparison of whether the actual and predicted values crossed certain thresholds within a defined time period.
- The metrics above utilize a comparison of actual and predicted threshold crossings, where: A is a True Positive, where both the actual and predicted cross the threshold, B is a False Positive, the actual does not cross but the predicted does, C is a False Negative, the actual crosses but the predicted does not, & D is a True Negative, neither actual nor predicted cross the threshold.
- To determine threshold crossings in the actual data, we calculated the maximum value in the 30 minute prediction window and compared that value to the thresholds. Because the softmax activation function outputs a probability, predicted values greater than or equal to 0.5 were considered positive predictions.
- The metrics, as well as the the Precision and Recall metrics used for plotting the Precision-Recall curves, are defined as:

$$POD = \frac{A}{A+C} \qquad FB = \frac{A+B}{A+C} \qquad Precision = \frac{A}{A+B}$$

$$POFD = \frac{B}{B+D} \qquad HSS = \frac{2(AD-BC)}{(A+C)(C+D)+(A+B)(B+D)} \qquad Recall = \frac{A}{A+C}$$

Plots



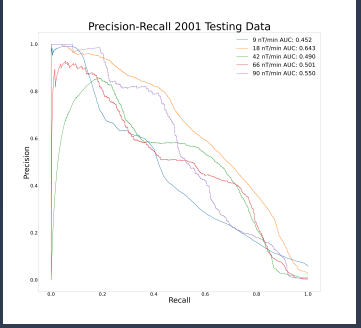
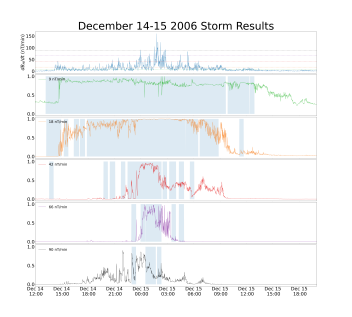


Figure 1a: Model results for the March 2001 storm. First panel on left side is the true dB/dt values with horizontal lines indicating each threshold. The subsequent panels are the results for each threshold with the blue shaded area being a threshold crossing of the real data, and the other colors indicate the model's predicted probability of crossing the respective thresholds 30-60 minutes into the future. Figure 1b: The Precision-Recall curves for the 5 different thresholds., for the full testing data from January 2001 – June 2001.



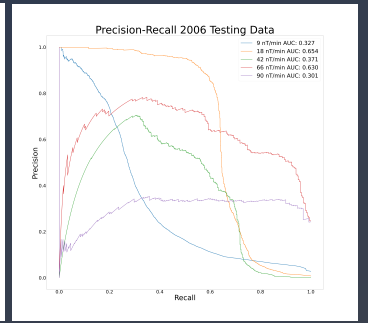
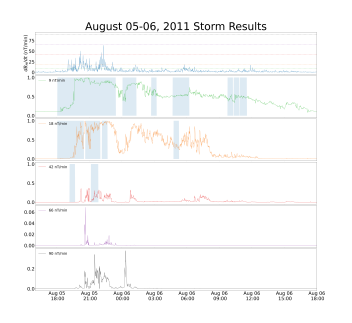


Figure 1a: Model results for the March 2001 storm. First panel on left side is the true dB/dt values with horizontal lines indicating each threshold. The subsequent panels are the results for each threshold with the blue shaded area being a threshold crossing of the real data, and the other colors indicate the model's predicted probability of crossing the respective thresholds 30-60 minutes into the future. Figure 1b: The Precision-Recall curves for the 5 different thresholds., for the full testing data from September 2006 – March 2007.



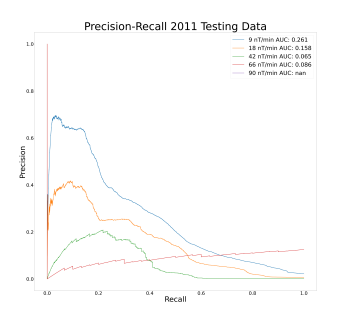


Figure 1a: Model results for the March 2001 storm. First panel on left side is the true dB/dt values with horizontal lines indicating each threshold. The subsequent panels are the results for each threshold with the blue shaded area being a threshold crossing of the real data, and the other colors indicate the model's predicted probability of crossing the respective thresholds 30-60 minutes into the future. Figure 1b: The Precision-Recall curves for the 5 different thresholds., for the full testing data from May 2011 – November 2011.

Metrics/Precision-Recall

- The Precision-Recall curves are a way of evaluating the skill of a model when dealing with an imbalanced dataset. Table 1 shows the percent of positive classes for each threshold in the testing set. For these degrees of imbalance, a model of no skill would have an Area Under the Curve (AUC) of near 0.
- None of the models exhibit the near perfect score of 1, but all of the models show some skill.

Thresholds	2011 Data	2006 Data	2001 Data
9	2.187	2.498	5.860
18	0.488	0.509	2.492
42	0.081	0.164	0.759
66	0.010	0.078	0.441
90	0	0.043	0.268

Table 1: Percent of positive classes for each testing set

- The metric scores in Table 2 were calculated using a greater or less than value of 0.5 for the predicted probability of a threshold crossing. With a predicted probability of 0.5 or greater being considered a 1 and less than 0.5 being a 0.
- Two different models were implemented to optimize metric scores. The 9, 18, and 90 nT/min thresholds used a model that included an additional 30 minutes of the recovery phase of the storm in the training data, while the models for the 42 and 66 nT/min thresholds produced better scores without the extra recovery time.
- All of the models achieved low POFD scores, indicating very few false alarms predicted.
- Several of the models were able achieve high POD scores, with the lowest threshold being above 0.7 and getting slightly worse as the thresholds are increased and the number of crossings is decreased.
- This is born out more in the FB scores, few of which are close to the perfect score of 1. A score below 1 indicates the model is predicting fewer crossings than the real data, which will artificially make the POFD score lower, and a score above 1 means the model is predicting more crossings, inflating the POD scores.
- The August 2011 storm scored the lowest across the board for the HSS scores. This is most likely a product of it being the least intense storm of the three examined. The December 2006 storm preforms the best across the board with the exception of the 90 nT/min threshold where the March 2001 storm scores higher. No HSS scores exceed a score of 0.8, meaning there is more work to be done to improve the models.
- It is important to point out that all of the models miss the initial spike in dB/dt. If this cannot be resolved, it could negate much of the utility of this type of model.

Threshold	POD2011	POD2006	POD2001	POFD2011	POFD2006	POFD2001	FB2011	FB2006	FB2001	HSS2011	HSS2006	HSS2001
9	0.706	0.936	0.901	0.298	0.356	0.190	1.212	1.066	0.922	0.391	0.615	0.536
18	0.460	0.857	0.831	0.219	0.127	0.277	1.509	0.948	0.891	0.201	0.722	0.462
42	0.000	0.327	0.377	0.001	0.006	0.102	0.028	0.343	0.444	-0.002	0.409	0.239
66	NaN	0.445	0.552	0.000	0.018	0.325	NaN	0.576	1.133	NaN	0.522	0.220
90	NaN	0.137	0.578	0.000	0.034	0.028	NaN	0.595	0.687	NaN	0.128	0.622

Table 2: The metric scores for the thresholds of the three storms examined. A perfect score for the POD, FB and HSS is 1 and a POFD is 0

Acknowledgements

This work was supported by NSF EPSCoR Award OIA-1920965

Additional MAGICIAN team posters: Connor (2) & Keesee (57)

Applying Machine Learning to MOMA Science Data for Science Autonomy

Victoria Da Poian[1,2] (victoria.dapoian@nasa.gov), Eric I. Lyness[1,3] (eric.i.lyness@nasa.gov),
Ryan M. Danell[4], Melissa G. Trainer[1], Xiang Li[1,5], William B. Brinckerhoff[1] and the MOMA team
[1]Planetary Environments Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD; [2]SURA CRESST, DC; [3]Microtel LLC, MD; [4]Danell Consulting, Inc., Winterville, NC; [5]University of Maryland, Baltimore County, MD

The Mars Organic Molecule Analyser Instrument

Poster ID: 54

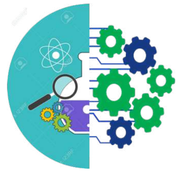
Space Exploration Context

Ground-in-the-loop limitations
Remote destinations and shorter at-target mission lifetimes limit or preclude ground-in-the-loop interactions

Communication challenges
Remote destinations and extreme environments involve longer communication delays and smaller data downlink capacities

Detection challenges
Scientists will not be able to guide spacecrafts' instrumentation in detection opportunistic features of interest

Data prioritization
Future instruments will certainly generate more data: data prioritization is vital to optimize mission science return



Science Autonomy

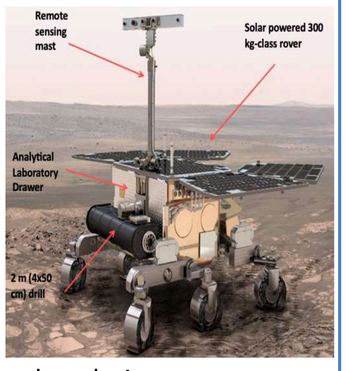
The ability of a science instrument to **analyze its own data** in order:
- to **calibrate** itself
- **optimize ops parameters** based on real-time findings
- **make mission-level decisions** based on scientific observations
- determine which **data** products to **prioritize** and send back first

ExoMars Rover Mission

Objectives

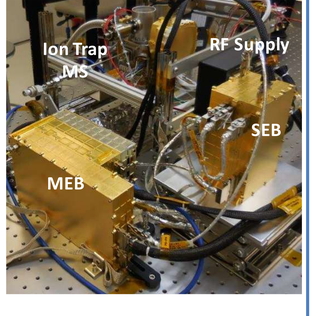
- To search for signs of past and present life on Mars
- To investigate the water/geochemical environment as a function of depth in the shallow subsurface

Instruments



- Drill delivers samples from 2m below the surface to the crushing station
- **Mars Organic Molecule Analyzer (MOMA):** Linear Ion Trap Mass Spectrometer
- **Raman Laser Spectrometer (RLS):** spectral analysis
- **MicrOmega:** imaging the samples (near IR hyperspectral microscope), mineral identification informs MOMA and RLS of regions of interest to target

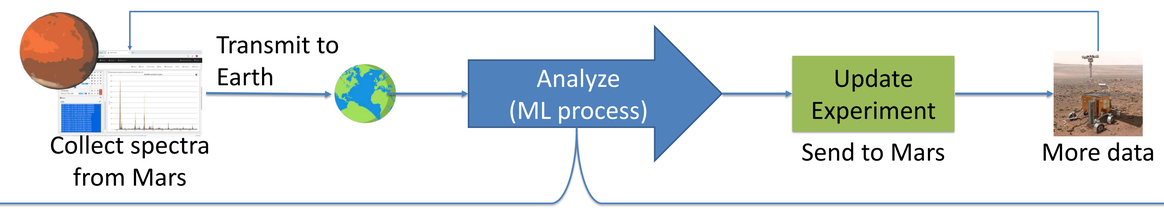
MOMA instrument

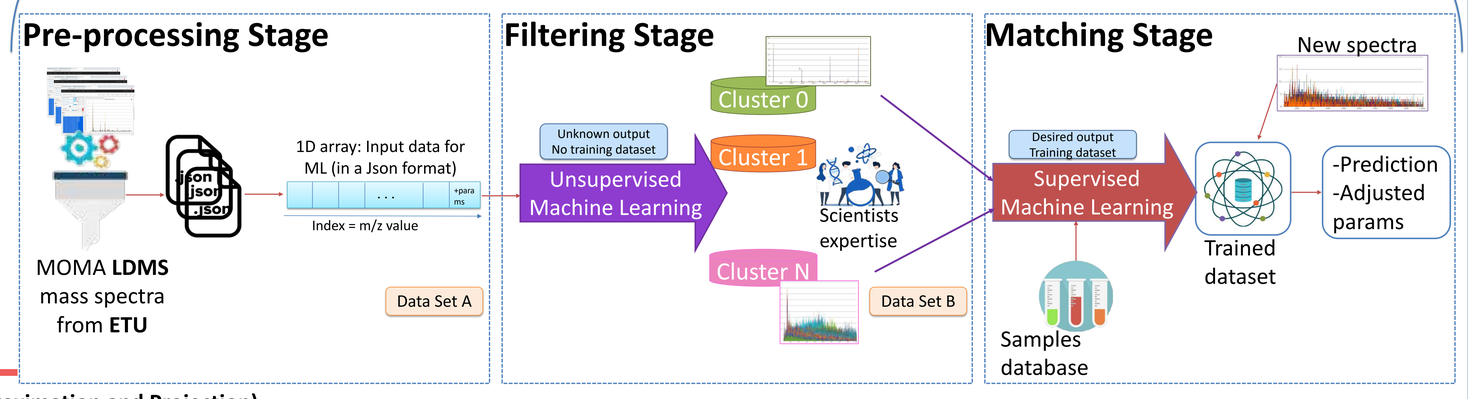


- Dual-source linear ion trap mass spectrometer coupled to pyrolysis/ derivatization-GC (GCMS mode) and UV laser for desorption / ionization (LDMS mode) of crushed rock samples
- Seeks the *molecular* signs of life with *broad* sensitivity to organics and analysis of chirality

Machine Learning process

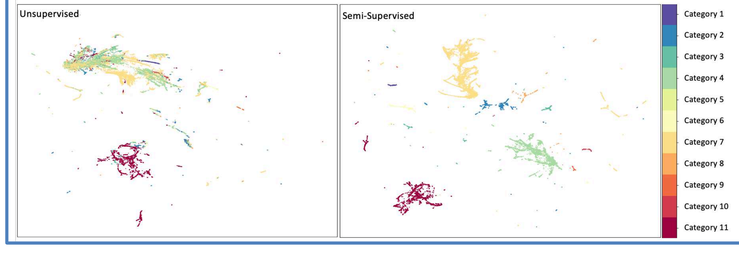
Motivation: The MOMA science team may only have a few hours to analyze the delivered data from Mars and to determine what further experiments should be done to meet the mission's science goals. We are investigating the use of ML to help the science team by matching Flight Model (FM) data from Mars to similar data from tests performed with the Engineering Test Unit (ETU) on Earth.





Initial results

UMAP (Uniform Manifold Approximation and Projection)



- Assists in **high dimension data visualization**
- Provides a **Semi-Supervised** method to help further cluster our data
- Possibility to add **new unseen data** into an existing embedding space
- UMAP Reduced data can be used as a **pre-processing** step of Supervised Learning

Other results: (feel free to contact me)

- **Data Processing:** dimensionality reduction, outlier detection,
- **Filtering Stage:** clustering algorithms
- **Matching Stage:** neural network development, implementation and CAL interface

Key Lessons

Data Volume
Challenging acquisition of large datasets acquisition for ML trainings

Team Efforts
Crucial collaboration between scientists and data science team

Resource Limits
Desired performance and available resources tradeoffs (CPU, memory)

Trusting ML
Dev. of a "Trust Readiness Level" index (like TRL in engineering dev.)

Using Machine Learning to Infer Pre-Entry Properties for Asteroid Threat Analysis

Jonathan Gee, jonathan.r.gee@nasa.gov Ana Maria Tarano, atarano@stanford.edu

Introduction

Accurately assessing asteroid threats relies on knowledge of the asteroid's pre-entry properties such as size, velocity, and mass. Directly measuring these properties can be infeasible due to the sparsity of events and the accuracy and fidelity of various sensors. Current analysis of an asteroid's pre-entry properties involves modeling the asteroid's entry into the Earth's atmosphere. This process can be time consuming and can require manual adjustment of uncertain modeling specific parameters.

NASA Ames has developed a genetic algorithm that can help automate asteroid modeling using the Fragment-Cloud Model (FCM). The algorithm generates realistic energy deposition curves based on actual energy deposition curves from real, observed asteroids. By using these synthetic, labeled energy deposition curves, we developed a one-dimensional convolutional neural network that can predict an asteroid's pre-entry parameters.

Motivation

- Assessing asteroid treats depends on successful characterization of the asteroid's physical properties
- Current methods requires manual modeling of parameters to match the observed energy deposition curves
- Inferring the pre-entry parameters directly from the energy deposition curve rather than modeling entries to match an observed event would greatly improve risk assessments of new asteroids

Data

- The data generated by FCM produces energy deposition curves from 100 km to 0 km with a resolution of 1km
- The dataset has 2.4 million energy deposition curves
- The data is split 88/2/10 for training, training validation, and testing
- The real asteroids are hand-modeled with FCM to match the documented results from other papers and are used for result validation

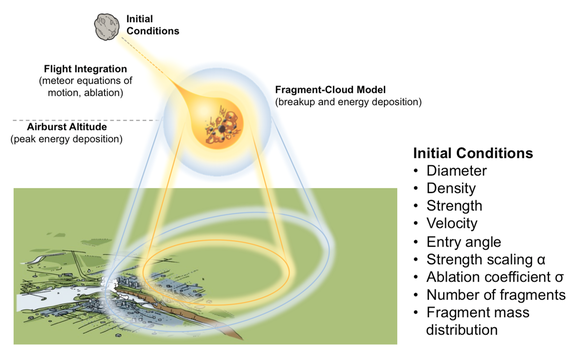
Parameter	Minimum	Maximum	Distribution
Diameter (m)	0.1	50	Uniform
Velocity (km/s)	11	25	Uniform
Entry Angle (°)	10	90	Uniform
Bulk Density (g/cm^3)	1.1	4.0	Uniform
Strength (kPa)	1	15000	Log Uniform

References

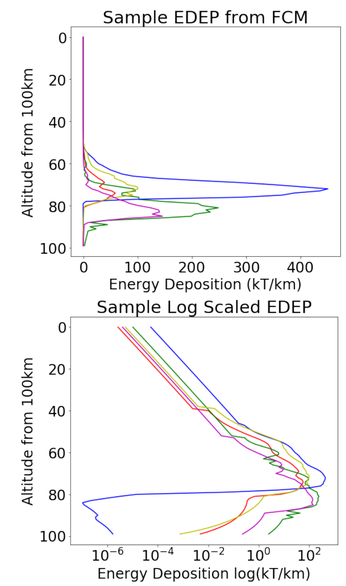
- Tárano, Ana María, et al. "Inference of Meteoroid Characteristics Using a Genetic Algorithm." Icarus, vol. 329, 1 Sept. 2019, pp. 270–281., doi:10.1016/j.icarus.2019.04.002.
- Wheeler, Lorien F., et al. "Atmospheric Energy Deposition Modeling and Inference for Varied Meteoroid Structures." Icarus, vol. 315, Nov. 2018, pp. 79–91., doi:10.1016/j.icarus.2018.06.014.

Method

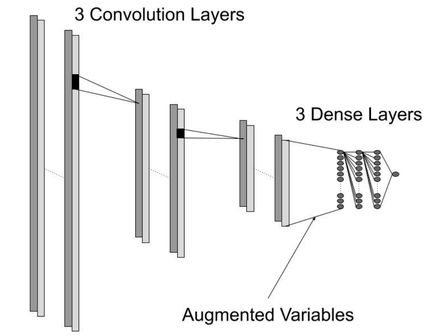
1. Model Asteroid Entries



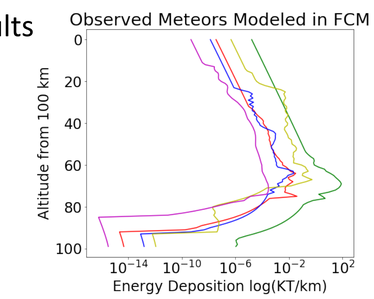
2. Transform Data



3. Train Model



4. Test Results



Model

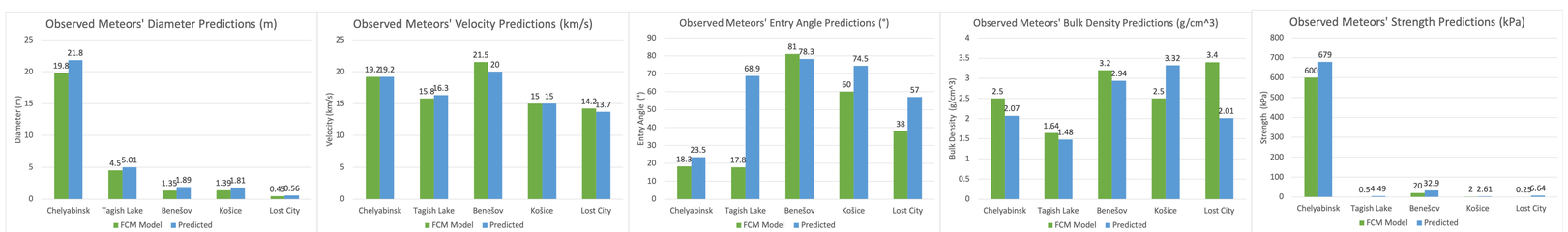
- Input: log scaled energy deposition (EDEP) curve
- Augmented, one-dimensional convolutional neural network
- 3 one-dimensional convolutional layers
- 5 augmented variables: total kinetic energy (KE), mean KE, max KE, and altitude at max KE of the input EDEP curve
- 3 dense layers
- Output: prediction of 1 pre-entry parameter

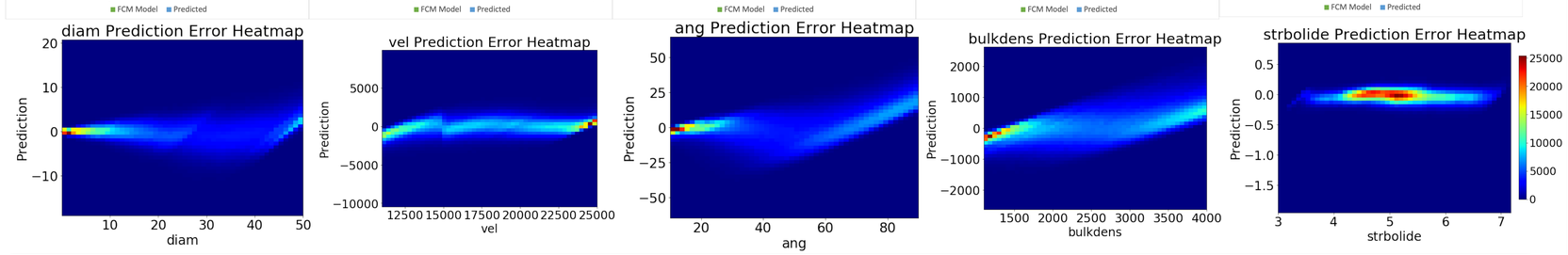
Results

- 5 separate models for each parameter
- Results validated with real modeled asteroids shown in the bar graphs
- Tested with with 10 percent of the synthetic data shown in the error heatmaps and R^2 table

Conclusions

- We can make predictions on real asteroid properties based on a synthetically generated dataset
- We show that our model is quite good at predicting diameter and velocity, and reasonably good at predicting entry angle, bulk density, and strength, for both the real cases as well as the test portion of the dataset
- Extensions to this concept include predicting pre-entry parameters based on light curve data and partial energy deposition curves





R^2 on Test Dataset for each Parameter									
Diameter	0.969	Velocity	0.951	Entry Angle	0.774	Bulk Density	0.628	Strength	0.942

Super-resolution of MDI Solar Magnetograms: Performance Metrics and Error Estimation

Paul J. Wright[1], Xavier Gitiaux[2], Anna Jungbluth[3], Shane A. Maloney[4], Carl Shneider[5], Alfredo Kalaitzis[6], Michel Deudon[7], Atılım Güneş Baydin[3], Yarin Gal[3], & Andrés Muñoz-Jaramillo[8]

[1] Dept. of Physics, Stanford University | pjwright@stanford.edu, [2] George Mason University, [3] Oxford University, [4] Trinity College Dublin, [5] Centrum Wiskunde and Informatica, [6] Element AI, [7] École Polytechnique, [8] SouthWest Research Institute

1. Summary

Aim: Develop an approach to convert and upscale line-of-sight magnetic field data to a reference survey in order to understand long-term variability of the magnetic field on time-scales larger than the lifespan of a single instrument.

ⓐ Data Pre-Processing

1. Standardize the Sun's orientation and distance from the detector such that the solar radius is constant over time
2. Register & shift individual 128" x 128" patches (see in set regions, **Figure 1**) to account for orbital differences.

ⓑ Neural Network Architecture

We use an Encoder-Decoder architecture based on HighRes-Net (see, **github.com/ElementAI/HighRes-net**). The trained Neural Network (NN) output is shown in **Figure 2**.

ⓒ Loss Functions & Metrics

To train our supervised NN, we include a range of terms alongside MSE (mean-squared-error) loss, and evaluate on additional performance metrics.

1. Loss Functions

Histogram: The magnetic field distribution is non-Gaussian; by implementing a *differentiable* histogram, we better preserve the observed distribution of magnetic field.

Structural Similarity Metric (SSIM): Measure the perceived similarity between images.

Gradients: Preserve the gradients of the magnetic field.

2. Performance Metrics

Information Entropy: To understand the informational content of the output over all spatial scales, and to diagnose hallucination in the NN.

ⓓ Error Estimation

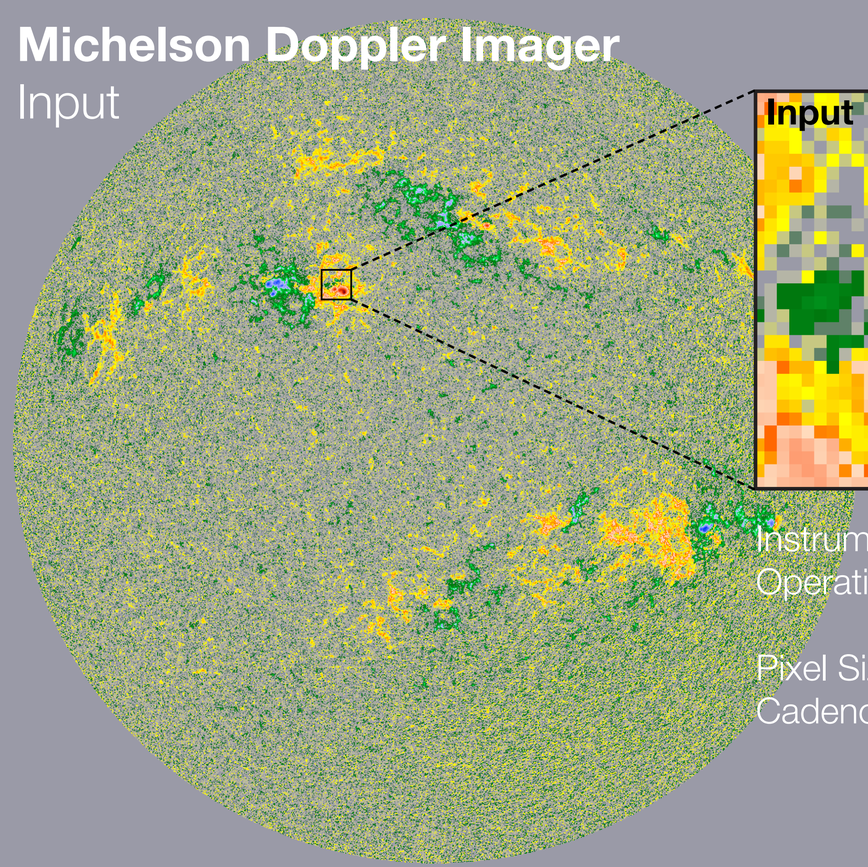
We use a Bayesian framework as in Kendall & Gal (2017) that decomposes uncertainty in to two components: ***epistemic*** (ignorance of the true data generating process), and ***aleatoric*** (the inherent noise). In practice, we implement this by adding Monte Carlo (MC) dropout in each convolutional layer, and track both the mean and variance of the magnetic field values.

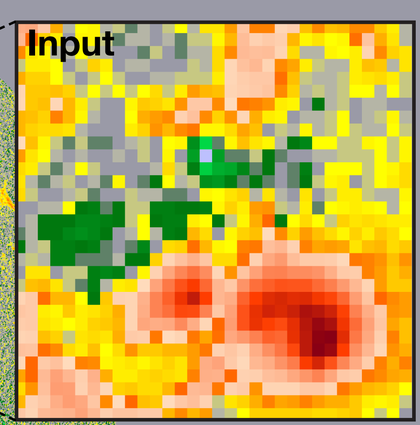
ⓔ Conclusions & Future Work

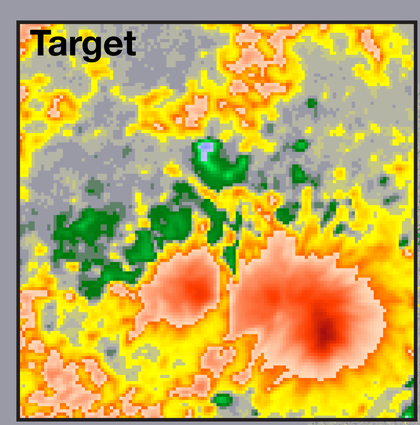
- To our knowledge, this is the first application of Bayesian Neural Networks to a super-resolution problem.
- Earlier versions of this work were published in workshops at NeurIPS 2019 (Gitiaux *et al* 2019, arxiv: **1911.01486**; Jungbluth *et al* 2019, arxiv: **1911.01490**).
- **Shortly, we will provide test users with the super-resolution output to understand the suitability for various science tasks.**

2. Data Overview

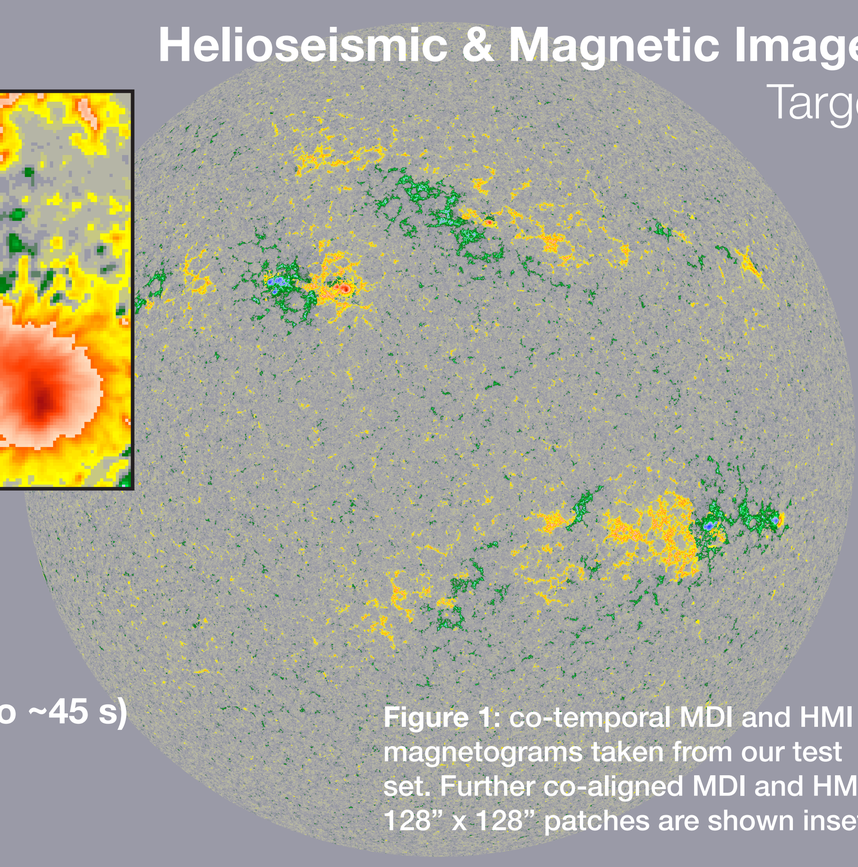
Michelson Doppler Imager
Input







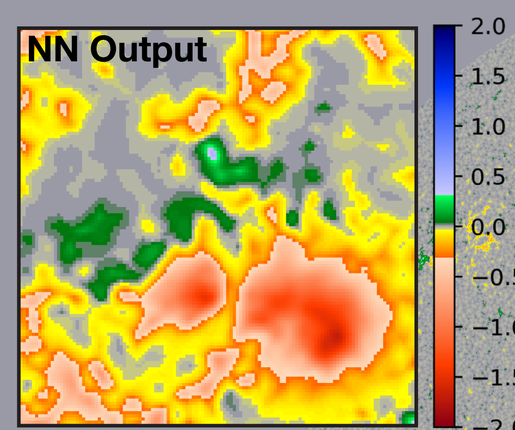
Helioseismic & Magnetic Imager
Target



	MDI	HMI
Instrument:	**MDI**	**HMI**
Operation:	**1995 - 2011**	**2010 - present**
	(one year of overlap)	
Pixel Size:	**2″**	**0.5″**
Cadence:	**96 min.**	**12 min. (down to ~45 s)**

Figure 1: co-temporal MDI and HMI magnetograms taken from our test set. Further co-aligned MDI and HMI 128" x 128" patches are shown inset.

3. Neural Net. Output



Super-resolution MDI
Neural Network Output

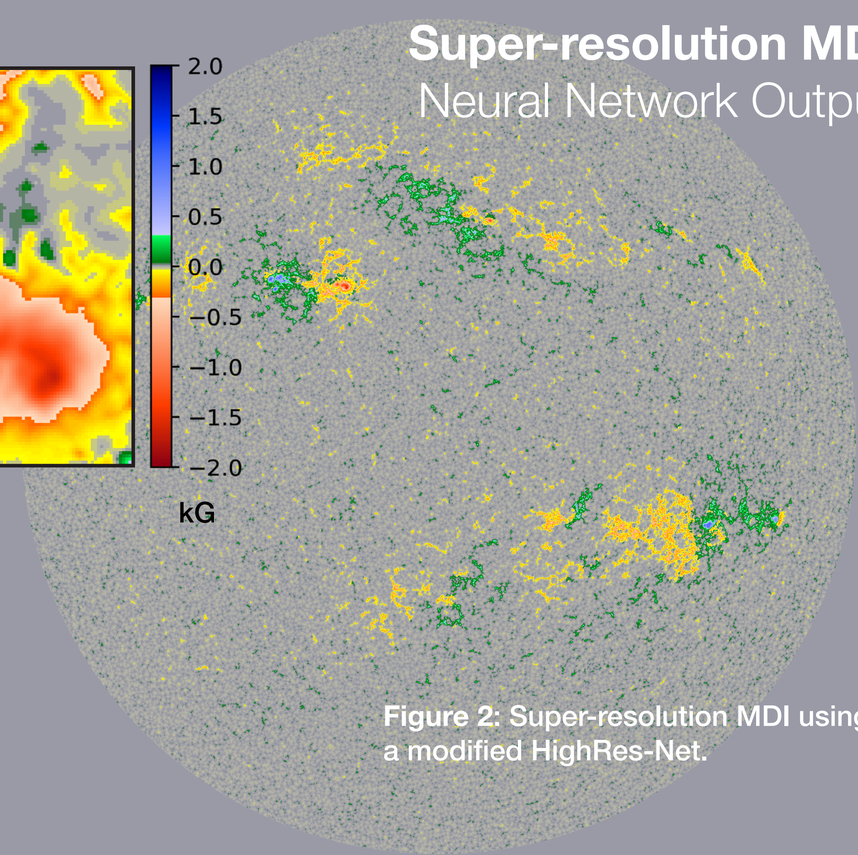


Figure 2: Super-resolution MDI using a modified HighRes-Net.

57: Comparison of Time Series Techniques to Model Connections Between Solar Wind Input and Geomagnetically Induced Currents

Amy Keesee[1a], Victor Pinto[1], Michael Coughlan[1], Connor Lennox[1], Md Shaad Mahmud[1], Hyunju Connor[2]
[1]*University of New Hampshire, Durham, NH 03824*
[2]*University of Alaska Fairbanks, Fairbanks, AK 99775*
[a] Amy.Keesee@unh.edu

MAGICIAN
Machine Learning Algorithms for Geomagnetically Induced Currents in Alaska and New Hampshire

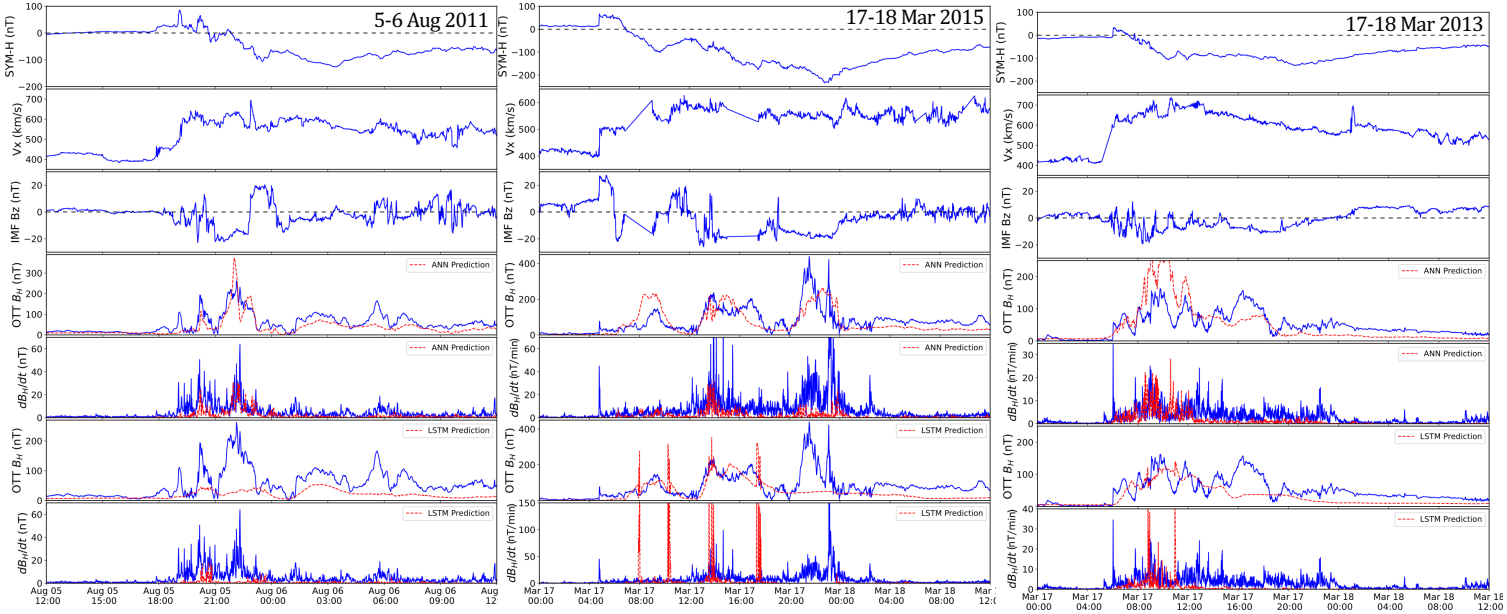
Abstract

Geomagnetically induced currents (GIC) can drive power outages and damage power grid components while also affecting pipelines and train systems. Developing the ability to predict local GICs is important to protecting infrastructure and limiting the impact of geomagnetic storms on public safety and the economy. While GIC data is not readily available, variations in the magnetic field, dB/dt, measured by ground magnetometers can be used as a proxy for GICs. We are developing a set of neural networks to predict the east and north components of the magnetic field, B_E and B_N, from which the horizontal component, B_H, and its variation in time, dB_H/dt, are calculated. We apply two techniques for time series analysis to study the connection of solar wind and interplanetary magnetic field properties obtained from the OMNI dataset to the ground magnetic field perturbations. The analysis techniques include a feed-forward artificial neural network (ANN) and a long-short term memory (LSTM) neural network. Here we present a comparison of both models' performance when predicting the B_H component of the Ottawa (OTT) ground magnetometer for the year 2011 and 2015 and then when attempting to reconstruct the time series of B_H for two geomagnetic storms that occurred on 5 August 2011 and 17 March 2015.

Methodology

- OMNI data and SuperMag data from the Ottawa (OTT) station from the years 1995-2010 were used. A full linear interpolation was done on the missing data points.
- Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM) models were trained in Tensorflow-Keras.
- The input feature vector includes solar wind speed (V, Vx, Vy, Vz), IMF (B, Bx, By, Bz), proton density, dynamic pressure, temperature and solar wind electric field using a 1-min cadence for the first 12 preceding minutes (i.e., up to t – 12) plus 10-min averages over the previous 2 hours. The ground magnetometer sin(MLT) and cos(MLT) values have been included to ensure a cyclical dependence over the Earth's rotation and solar zenith angle as a proxy of both latitude and yearly seasonality.
- The output value is not included as an input feature in the LSTM for direct comparison to the ANN. We also want to determine the ability for a forecasting model, requiring a time delay for the model input.

Storm Results



Metrics

- Each storm interval is divided into 20-min windows, within which the maximum measured dB_H/dt is determined to see whether it crosses threshold values of 18, 42, 66, and 90 nT/min.
- The modeled dB_H/dt is calculated with two methods described by Pulkinnen et al, 2013 and Tóth et al., 2014, respectively:
  - $\left(\frac{dB}{dt}\right)_H = \sqrt{\frac{dB_N}{dt}^2 + \frac{dB_E}{dt}^2}$
  - $\left(\frac{dB}{dt}\right)_H = \left(\frac{B_H}{248\,\mathrm{nT}}\right)^{1.04} \frac{\mathrm{nT}}{\mathrm{s}}$
- These values are used to determine the probability of detection (POD), probability of false detection (PFD), percentage correct (PC), and Heidke Skill Score (HSS).

		Pulkkinen				Tóth			
		18	42	66	90	18	42	66	90
2011 Storm	ANN POD	0.33	0.00	—	—	0.50	0.50	—	—
	LSTM POD	0.08	0.00	—	—	0.00	0.00	—	—
	ANN PFD	0.02	0.00	0.00	0.00	0.02	0.03	0.03	0.01
	LSTM PFD	0.08	0.00	—	—	0.00	0.00	—	—
	ANN PC	0.88	0.97	1.00	1.00	0.90	0.96	0.97	0.99
	LSTM PC	0.85	0.97	1.00	1.00	0.84	0.97	1.00	1.00
	ANN HSS	0.41	0.00	—	—	0.58	0.37	—	—
	LSTM HSS	0.13	0.00	—	—	0.00	0.00	—	—
2015 Storm	ANN POD	0.09	0.00	0.00	0.00	0.69	0.77	0.00	0.00
	LSTM POD	0.13	0.11	0.20	0.00	0.38	0.33	0.20	0.00
	ANN PFD	0.00	0.00	0.00	0.00	0.13	0.13	0.00	0.00
	LSTM PFD	0.05	0.07	0.07	0.07	0.19	0.08	0.03	0.00
	ANN PC	0.73	0.92	0.95	0.98	0.82	0.86	0.95	0.98
	LSTM PC	0.71	0.86	0.90	0.92	0.68	0.87	0.94	0.98
	ANN HSS	0.13	0.00	0.00	0.00	0.56	0.42	0.00	0.00
	LSTM HSS	0.09	0.04	0.10	0.00	0.19	0.23	0.19	0.00

Four threshold values in nT/min are used.

Conclusions

- There is some ability for each of the models to predict the timing of magnetic field perturbations, though this ability is not consistently better for either model between the storms and neither is able to predict the magnitude of the enhancements or predict enhancements later in the storm.
- Validation metrics indicate that the LSTM is barely more skilled than random or constant predictions, and that using an empirical fitting improves HSS as it does for first principles-based models.

Further information

MAGICIAN Team Posters at this workshop: Connor (2), Coughlan (53)

This work is Funded by NSF EPSCoR Award OIA-1920965

Keesee AM, Pinto V, Coughlan M, Lennox C, Mahmud MS and Connor HK (2020) *Front. Astron. Space Sci.* 7:550874. doi: 10.3389/fspas.2020.550874

Poster 58

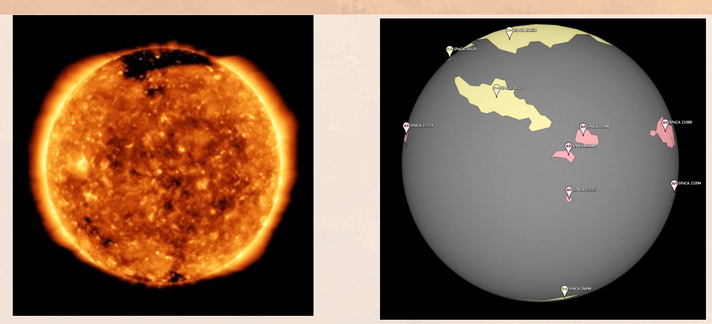
At a Glance

- Using a multi-channel solar dataset from SDO and labels using conventional image processing, we trained a neural net to segment solar features.
- We found using an LSTM-Unet hybrid created better convergence over a Unet.
- Including well known physics, like the PDF of sunspot locations, increased the accuracy of the model.
- The LSTM-Unet approach is only as good as the labels. We are currently working on a "fuzzy-labeling approach to introduce uncertainties to the training data.

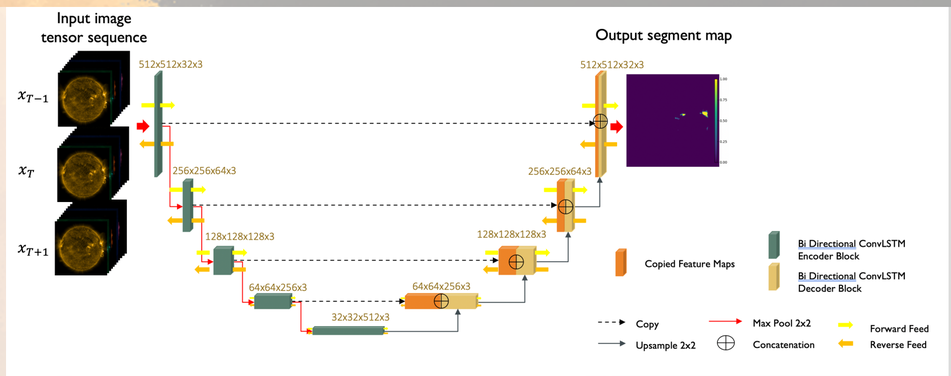
For more information contact
michael.s.kirk@nasa.gov

Developing Deep Learning for Solar Feature Recognition

Michael S. Kirk, Raphael Attie, James Stockton, Matt Penn, David Hall, Barbara Thompson



The Solar Dynamics Observatory (SDO) images the sun in extreme ultraviolet wavelengths (left) to reveal several features that evolve and influence the Earth: sunspots, coronal holes, and active regions (labels of these are shown in the image on the right). Our goal was to encode classical methods of feature labeling using ML to be able to run on the full 200 million image archive where conventional methods would not be able to perform.

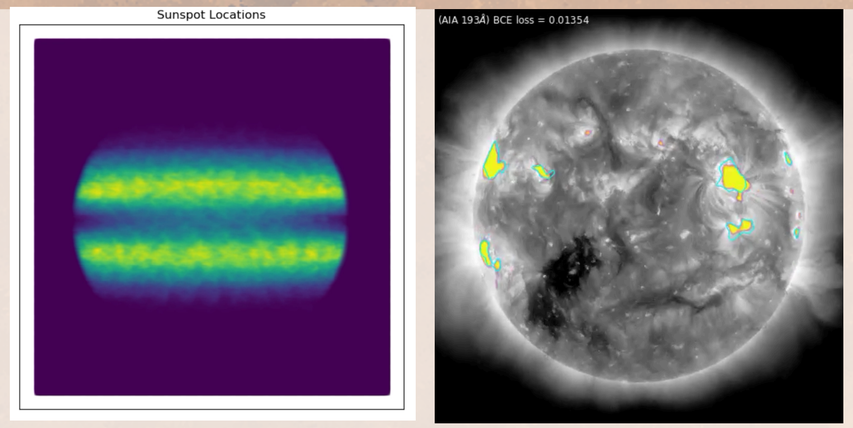


A Unet architecture is well suited to the high-resolution SDO images to detect solar features of various size. In training the Unet, we down-sampled the images by a factor of 4 to be able to include all 11 wavelength bands SDO produces. Not every solar feature is observed in each channel, but we wanted to make sure the Unet was judging the saliency of each channel correctly.

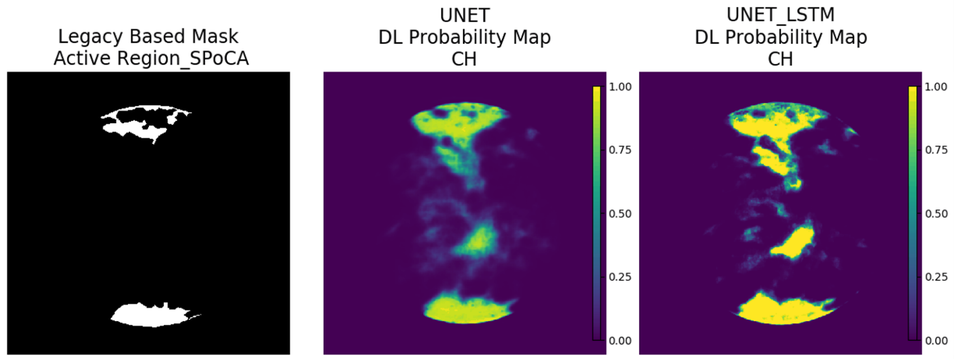
Because solar features evolve in time, we implemented a long short-term memory (LSTM) architecture on each layer of the Unet by passing a tensor of `[t-1, t, t+1]` through the pooling and up-sampling steps (shown above). This hybrid LSTM-Unet approach greatly improved the accuracy and convergence of the model (shown below). The drawback of the hybrid approach is that it requires 3-times the memory for training, which makes full resolution, 4k x 4k, images difficult to work with.

GRAY BOX MODEL

Encode existing scientific knowledge into a machine learning architecture.



Adding in known systematics and physical phenomena greatly increased the accuracy of the model. We included a probability density function of sunspots (shown left), a function to model the degradation of the detectors, and six temporal signatures that are well known – both physical signals and systematic errors. The accuracy could be gauged by comparing predicted vs. measured labels of active regions (shown on the right).



Despite the robust models, there are still significant disagreements between the legacy labels and the DL probability maps (show above). Examining the input images raises the question if there is some error in the legacy labels. Currently we are working to create a probability map of training labels to better capture the uncertainty inherent in the solar features.

This work was a partnership between NASA, NVIDIA, and Altamira.

ALTAMIRA NASA NVIDIA

FRONTIER DEVELOPMENT LAB 2019

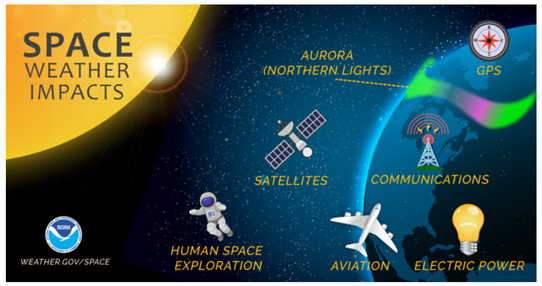
59. CORRELATION OF AURORAL DYNAMICS AND GNSS SCINTILLATIONS WITH AN AUTOENCODER

Kara D. Lamb[1,2,*], Garima Malhorta[3,*], Athanasios Vlontzos[4,*], Edward Wagstaff[5,*], Atilim Günes Baydin[5], Anahita Bhiwandiwalla[6], Yarin Gal[5], Alfredo Kalaitzis[7], Anthony Reina[8], Asti Bhatt[9]
*equal contributions

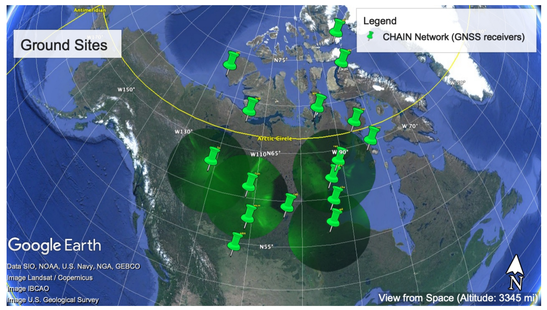
1). CIRES 2). NOAA ESRL 3). University of Michigan 4). Imperial College London 5). University of Oxford 6). Intel AI Lab 7). Element AI 8). Intel AIPG 9). SRI International

PROBLEM

PROBLEM SETTING



Signals from Global Navigation Satellite System (GNSS) satellites are altered in phase and amplitude by ionospheric scintillations. These scintillations can cause a loss of spatial tracking and time information.



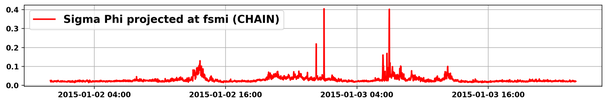
Scintillations are known to correlate with visible aurora. To investigate this correlation, we use data from several observation networks in N. Canada

QUESTIONS

- Are specific structures within the visible aurora more likely to correlate with the occurrence of GNSS phase scintillations?
- Can an unsupervised approach to aurora image classification improve our understanding of this correlation?

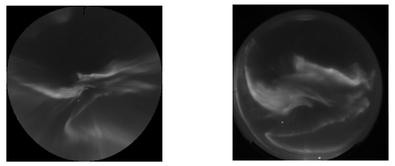
DATA SOURCES

TIME SERIES



We use measurements of the ionospheric phase scintillation index (σ_Φ), the standard deviation of the detrended carrier phase, averaged over 60 s.

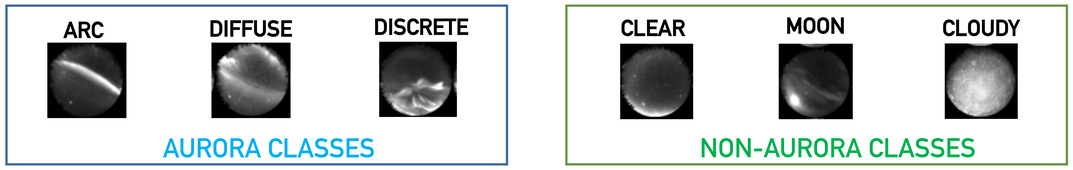
AURORAL IMAGERY



35,277 raw images from THEMIS all-sky imagers are projected onto a latitude-longitude grid. This data set was used to train the auto-encoder

THEMIS IMAGE CLASSES

7700 manually annotated auroral images, classified among 6 classes

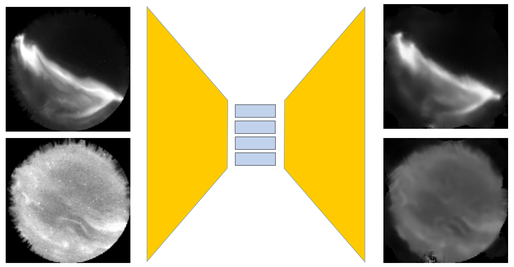


ACKNOWLEDGEMENTS

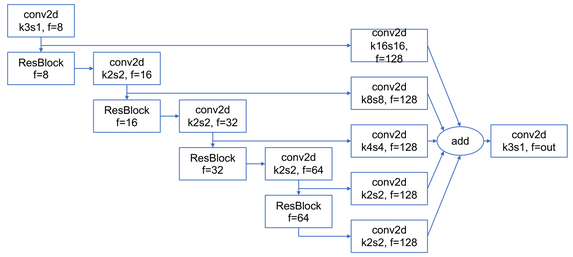
We would like to thank Google Cloud for providing the computational infrastructure and resources for this project.

SOLUTION

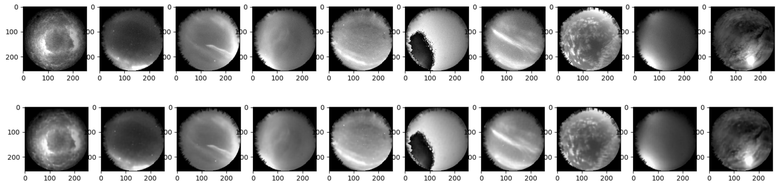
AUTOENCODER



The encoder and decoder uses a U-Net like architecture in order to learn image structures at different scales.

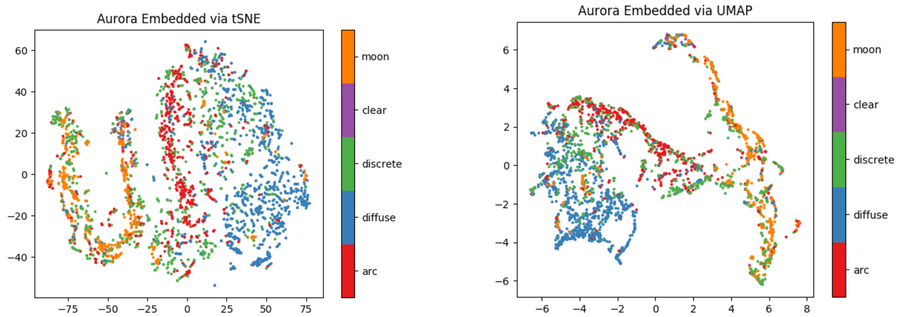


Res-AE architecture for encoder.

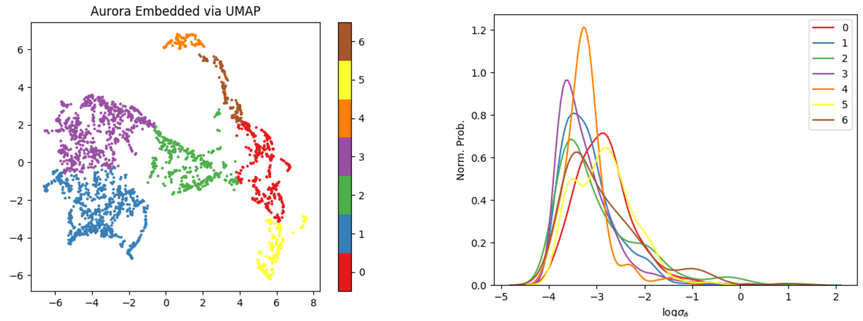


Original (top) and reconstructed (bottom) images

RESULTS



Visualization of aurora images in the latent space using tSNE (left) and UMAP (right). Unsupervised clusters typically correlate with human-annotated image classes and show similar clusters using both tSNE and UMAP.



Some clusters in the latent space (left) are more likely to be associated with higher phase scintillations (right). Clusters containing discrete and arc aurora classes are associated with higher phase scintillations.

For additional information see our paper: https://arxiv.org/abs/1910.03085

Complex Data Explorer (CODEX) – A multi-use Machine Learning Powered Tool for Rapid Data Exploration

Jack Lightholder, Lukas Mandrake, Josh Rodriguez, Rob Tapella, Patrick Kage
jack.a.lightholder@jpl.nasa.gov

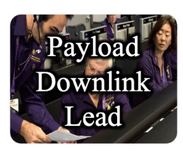
Abstract

Modern science datasets from missions like OCO-2 and telemetry records in Ops may have 500+ simultaneous measurements at each of millions of time samples. Scientists would often like to look through the record and discover not only expected trends but ones they did not initially guess, while Ops personnel perform the same task under serious time pressure should an anomaly occur. In both cases, the optimal environment for this rapid exploration large data would be one where visualizations were clear, interactive, and responsive, permitting the investigator to "play" with the data and gain rapid insight, falsify hypothesis, and make discoveries. Machine Learning (ML) has proven invaluable in providing some of these key data insights, but to do so in a statistically robust and reliable manner requires a data science professional and a lot of custom Python code, losing any sense of interaction and play. CODEX will address these concerns by providing a desktop-like environment with standard scientific graph types that are robust to rapid, powerful exploration.

A Common Need

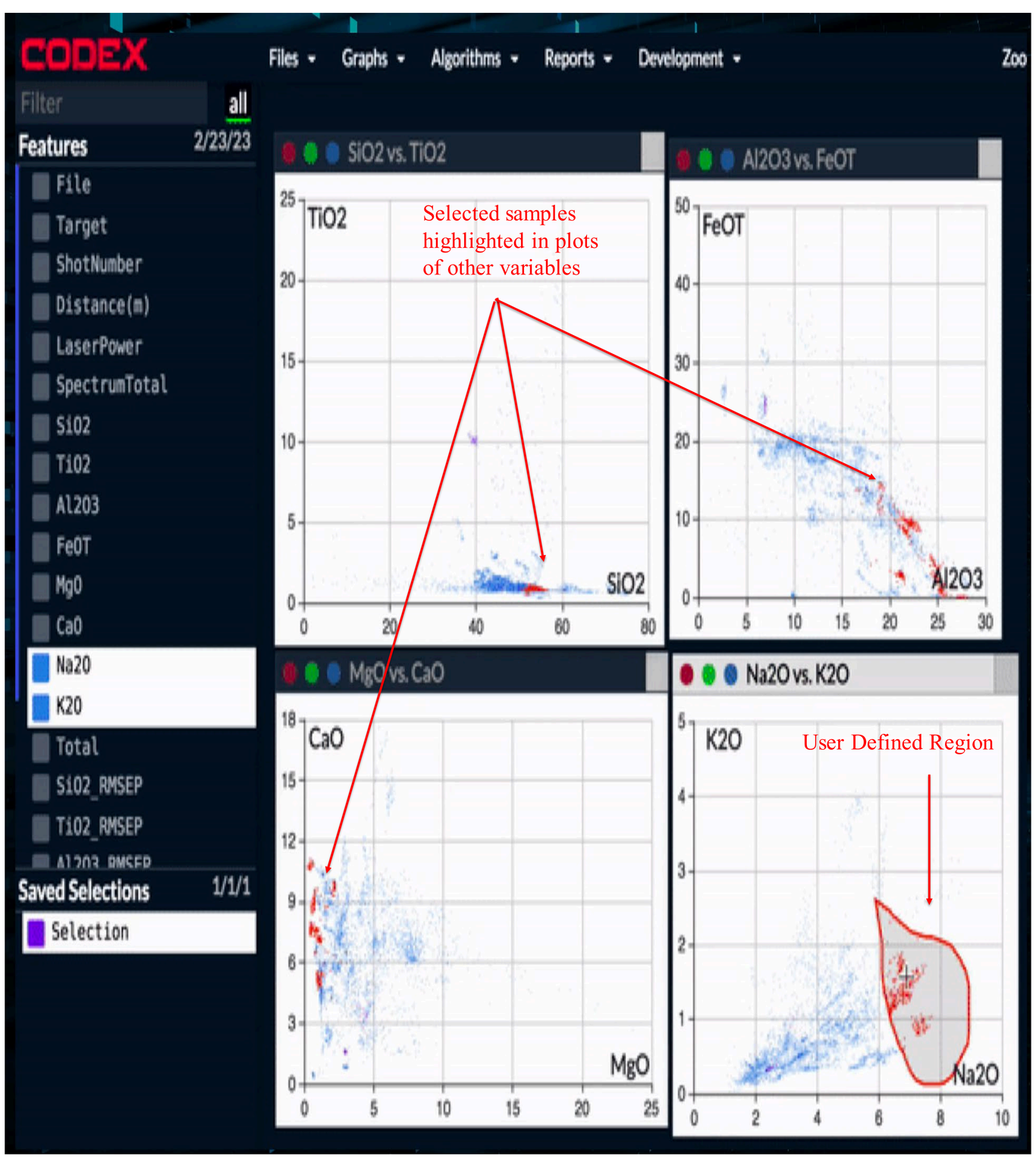
- High Dimensional time-series data
- Identify strange, outlier, or invalid values
- Interactively explore data
- Build/falsify hypotheses
- Interrogate relationships between cols
- Find more events like this
- Provide simple recipe to recognize events
- Create predictive, explanatory models
- How many families of data are present?

Machine Learning was made for this!



Intuitive exploration of the impact of algorithms, and their parameters settings, on your data. Rapidly explore algorithm categories to see which meet the needs of your application, without writing custom code every time!

Fast Hypothesis Testing



Intuitive exploration of multiple data components. Draw plot regions to see where those data points are in other aspects of your data for rapid intuition building.

Guiding Principles

- **Easy, interactive graphing**
  - scatter, heatmap, histogram, line, bar
  - linked: data here can be found everywhere else
- **Fast interactivity**
  - Humans learn best by manipulating playing
  - Slow batch analyses lose context & attention…
- **Continuous, visual guidance**
  - No question without rich support to guide answer
  - Permit visual selection & previews
- **Never stop working**
  - Long analyses run in background
  - Always foreworn of time & memory

Capabilities

- Clustering
- Regression
- Classification
- Dimensionality Reduction
- Feature Selection
- Endmember Analysis
- Anomaly Finding
- Interactive / Linked Graphing

Conclusions

- Fast discovery of data issues & problems
- Fast intuition building
- Powerful ML techniques made visual
- Guidance for every step of exploration
- Doesn't replace Python or MATLAB
- Does start you off ready to do great work

Future Infusion & Applications

- Mission operations data – build quick intuition about vehicle health and conduct on the fly trending.
- Science data – explore data sets to locate areas of interest for deeper study.
- Complex data interactions – leverage machine learning to gain insights to data characteristics which separate populations of data.

Contact: Jack Lightholder (jack.a.lightholder@jpl.nasa.gov) - Poster 60

Jet Propulsion Laboratory
California Institute of Technology

"Look and you will find it – what is unsought will go undetected." - Sophocles

COSMIC

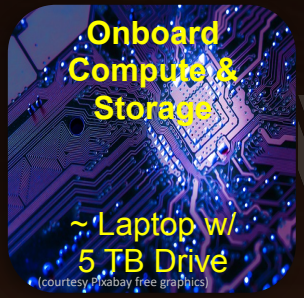
Content-based On-board Summarization to Monitor Infrequent Change

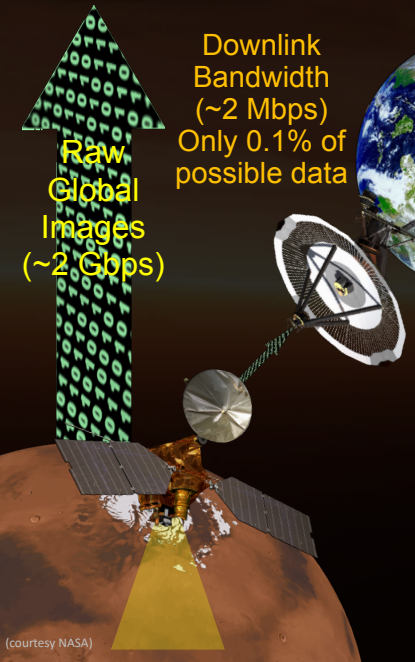
PI: Dr. Lukas Mandrake
Co-I: Gary Doran (398), Masha Liukis (398), Steven Lu (398),
Umaa Rebbapragada (398), Jimmie Young (398), Kiri Wagstaff (398)
Interns: Aaron Alvarado, Valentin Bickel, Anneliese Braunegg, Tina Cao, Paul Horton, Daniel Jeong, Michael Munje, Asher Trockman

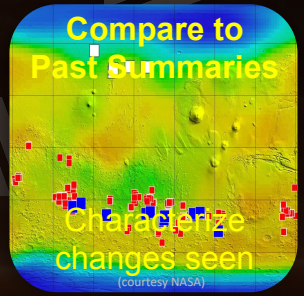
RPC 122 - Virtual Research Presentation Conference
Program: Strategic Initiative (6x)
Assigned Presentation #18
CL#: 20-5803

Poster #62

Innovation: Onboard Summarization for Science







Innovative Concepts

- Leverage Upcoming Space Computing
- Leave camera on & process everything
- Trustable science autonomy
- Store content-based summarizations
- Perform change detection between summaries
- Can now trade onboard compute for bandwidth!
- Many kinds of change can be captured
- Camera + compute + COSMIC is new smart instrument!

Scientists in Control

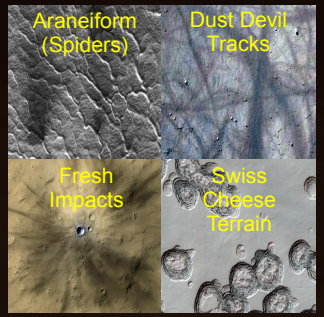
- Scientists reconfigure COSMIC on-the-fly
- COSMIC sends reports on global change
- Scientists more informed to react:
  - Request targeting of observed change?
  - Request tiny snapshots already waiting?
  - Request high-res monitoring of a site?
- Miss a comm pass? COSMIC keeps working.

Change Detection Capabilities & Progress

Known Science Targets

(surface classification)

- Know what, but not where/when
- Can look for any surface feature
- Needs example images for training
- Alerts provide time/space map
- Works on entire globe



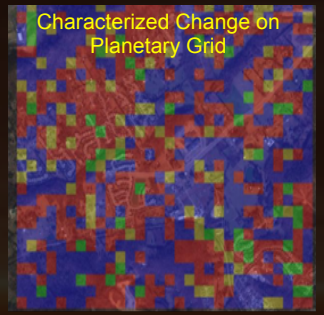
TRL ~5

- Tested on entire CTX database
- Actual science discoveries attained!
- Related task infused into PDS archive
- Ready for flight-oriented maturation

Agnostic Change Detection

(image-summary-based change)

- Don't know what, where, or when
- Discover new transients
- Global reports on change characterization
- Agnostic to expectation or known targets



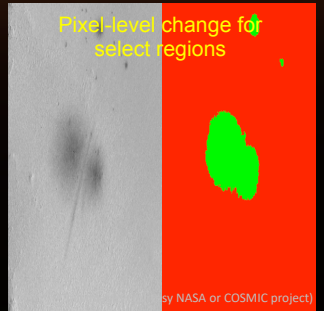
TRL ~3

- Feasibility study complete on real data
- Computational goals met
- New change-detection req: Stability
- Refinement needed for specific science use-cases and rigorous V&V

Region Monitoring

(Pixel change for small areas)

- Know where, but not what/when
- Classical change detection
- Only a ~1k small regions at a time
- Alerts when something starts to happen



TRL ~3.5

- Excellent performance on small tests using real & synthetic data
- Tolerant to expected operational environment
- Ready for rigorous V&V

A Deep Learning Approach to GNSS-R: Predicting Soil Moisture with Delay-Doppler Maps

Dr. T. Maximillian Roberts†, Ian Colwell, Dr. Rashmi Shah, Dr. Stephen Lowe, Dr. Clara Chew‡

Jet Propulsion Laboratory, ‡University Corporation for Atmospheric Research

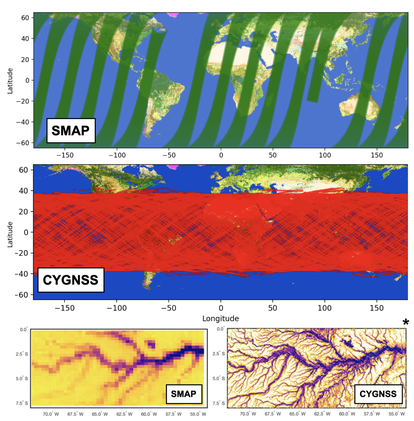
Jet Propulsion Laboratory
California Institute of Technology

Abstract

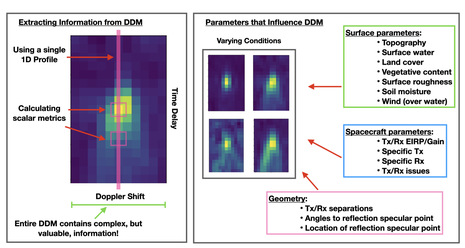
GNSS reflection measurements can be calibrated with data from SMAP to yield estimates of soil moisture with increased spatiotemporal resolution, useful to certain hydrological/meteorological studies. Current approaches which use simple models of the relation between the DDM (delay-Doppler map) and soil moisture which can fail in certain regions. Complex information contained in the complete 2D DDM could help in these regions, and can be extracted through the application of deep learning based techniques. This approach simultaneously provides the ability to incorporate additional contextual information from external datasets. Our work explores the data-driven approach of convolutional neural networks (CNNs) to determine complex relationships between the reflection measurement and surface parameters. CYGNSS DDMs were aligned with SMAP soil moisture values and ancillary datasets, a network was developed and trained with these measurements, the results of which are analyzed and compared to existing global soil moisture products.

Motivation and Concept

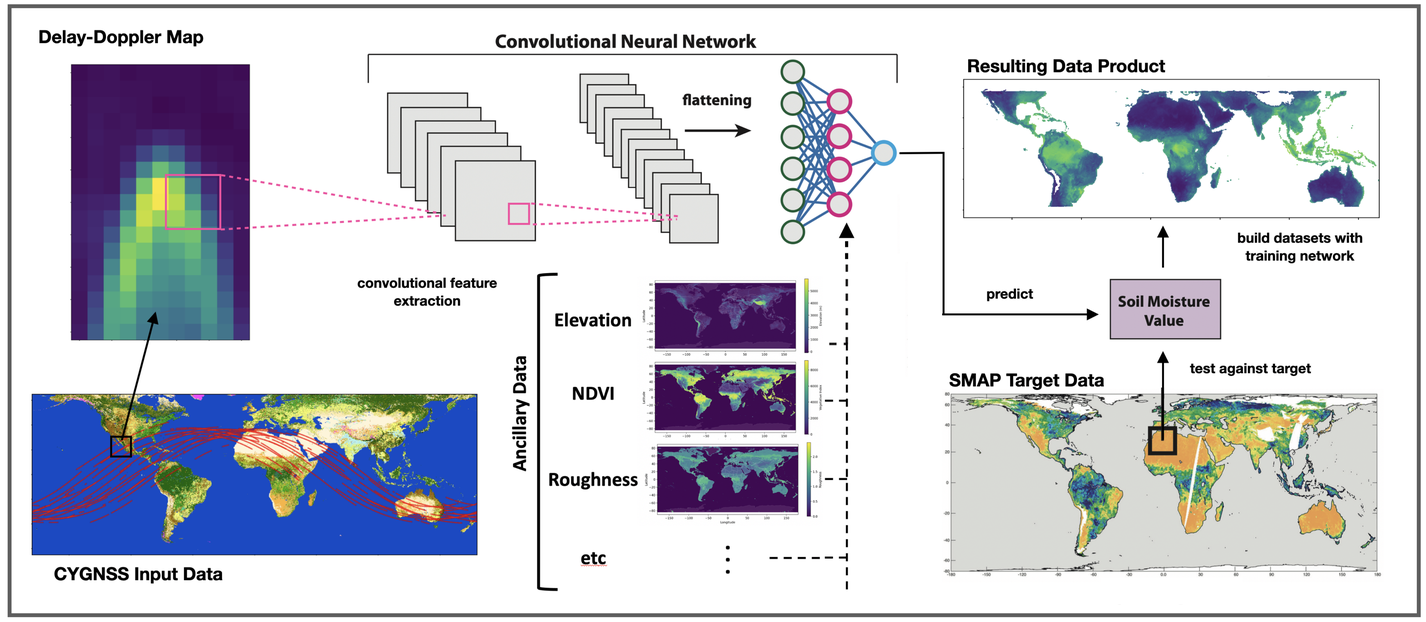
A comparison of the spatial and temporal resolutions attainable with SMAP vs CYGNSS, globally for a single day, and composite over the Amazon river.



CYGNSS DDMs are calibrated with SMAP measurements, and are used to estimate soil moisture. The full interpretation of DDMs is a complex problem, current methods use simplistic models.



CNNs extract complex structural features from the DDMs while integrating contextual information to build data-driven models of the reflecting conditions.

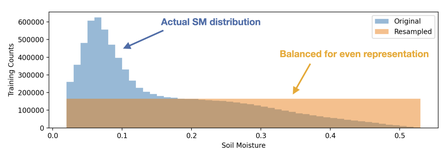


Dataset Development

Parameters which influence the DDM were determined and corresponding datasets were aggregated (see below table). Each dataset was studied to understand the characteristics, outliers, and indicators of problematic values.

Data	Rate	Source	Base Resolution	Used Resolution
Primary Input: CYGNSS DDM	Per DDM	CYGNSS	0.5-7 km	0.5-7 km
Ancillary Inputs:				
CYGNSS SC Info/PRN Number	Per DDM	CYGNSS	DDM-scale	DDM-scale
Angle/Range to Reflection	Per DDM	CYGNSS	DDM-scale	DDM-scale
Gains/EIRP	Per DDM	CYGNSS	DDM-scale	DDM-scale
Latitude/Longitude	Per DDM	CYGNSS	DDM-scale	DDM-scale
Surface Elevation/Slope	Static	SMAP L1-L3 Ancillary	1 km	3 km
NDVI	Daily	SMAP L1-L3 Ancillary	1 km	3 km
Stem Factor/VWC	Static/Daily	Calculated	1 km	3 km
Land Cover	Static	GlobCover	1 km	3 km
Surface Roughness	Static	SMAP L1-L3 Ancillary	1 km	3 km
Precipitation	Daily	SMAP L1-L3 Ancillary	36 km	36 km
Surface Water	Static	Pekel	30 m	3 km
Target: SMAP SM Value	Daily	SMAP	36 km	36 km

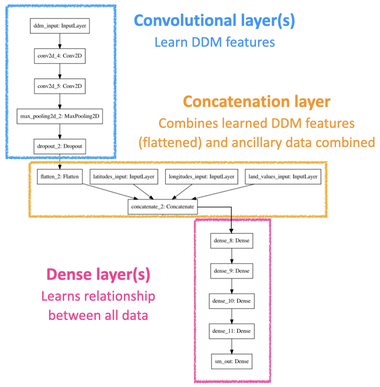
Data from different sources are "aligned", spatially and temporally, to fall within the same 3 km EASE grid cell on the same day. These training samples are compiled into a database with features designed for efficient processing.



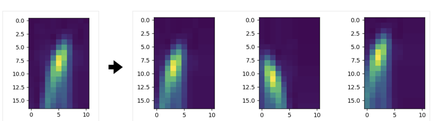
Training sets are filtered, standardized, and "balanced" to optimize for training. The above shows the balancing process, where some values are undersampled and others are oversampled to create an even distribution. Catergorical scalar data (like land type) are encoded to "one-hot" array inputs.

Neural Network Development

Development of the neural network was broken in two; developing a CNN specifically tuned for processing DDMs, and building a complete network for prediction of soil moisture.



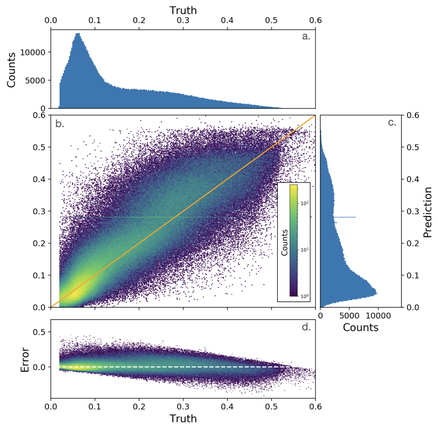
Optimization of DDM-specific CNN was studied with the toy-problem of land-type classification (blue layers only); DDM only input, land value is the target. Common CNN architectures were explored, and DDM augmentation and resolution enhancement was tested.



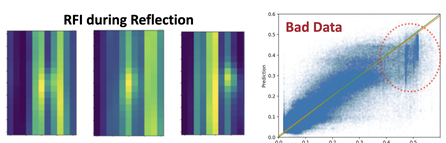
A complete network (blue, yellow, and pink layers) architecture able to accept various ancillary inputs was developed to estimate soil moisture, integrating the DDM-tuned CNN.

Analysis and Results

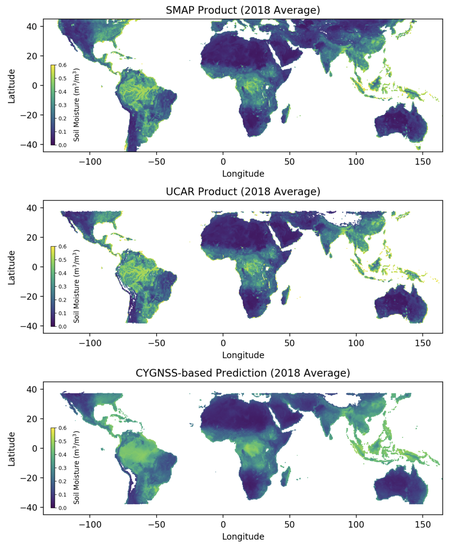
Network performance is established using unseen data split from the original dataset composed of randomly distributed points in space and time. Passing these samples through the network, a strong correlation between the predictions and targets (Pearson coefficient of 0.89) demonstrates the technique's potential. Dataset preprocessing, training, and analysis is repeated iteratively.



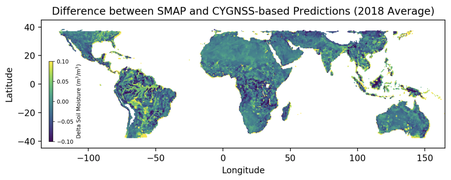
Bias toward underestimation of high SM values can be seen, likely a result of the balancing process. Other issues, such as RFI and problematic SMAP data are also revealed through analysis.



Network predictions are compared on a global scale to existing SM product of SMAP and UCAR averaged to 36 km EASE grid cells for the entire year of 2018. Qualitative comparison shows overall trends in strong agreement. However, areas with expected high SM content display less detailed structure.



Differences between predictions and the UCAR product are seen to strongly correlate with problematic SMAP quality flags removed from training data. Biases are likely created by removing samples over high surface water fraction.



Next Steps

This work has shown CNNs can be used to interpret DDMs directly with opportunity for significant improvement. Immediate follow on studies will further refine dataset filtering and spatiotemporal averaging, and add valuable, missing ancillary datasets (such as "distance-to-water"). More advanced work will create "ensembles" of networks for regional prediction, implement vector inputs for ancillary data (input region of values, not average), and include in situ measurements in training as "high value" targets. Futhermore, this concept is generalizable to other surface retrievals and product development such as "freeze-thaw", flood/inundation, and water masks.

†Contact email: t.maximillian.roberts@jpl.nasa.gov

*Modified from Chew et al. 2019.