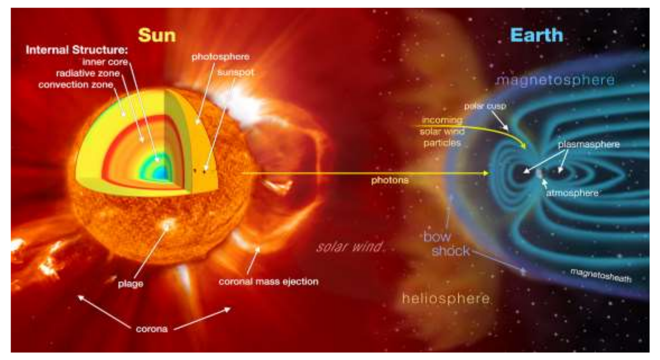
# MACHINE LEARNING MODELING AND RISK ANALYSIS OF GEOMAGNETICALLY INDUCED CURRENTS FROM SPACE WEATHER AND CORONAL MASS EJECTIONS

Jet Propulsion Laboratory
California Institute of Technology

**STUDENTS:** SEAN JUNG, JUI-WEI CHANG, JIAQI DONG, SHIYUN ZHOU, LONG THANH NGUYEN, MAZZY CHEN, TUNG THANH DANG

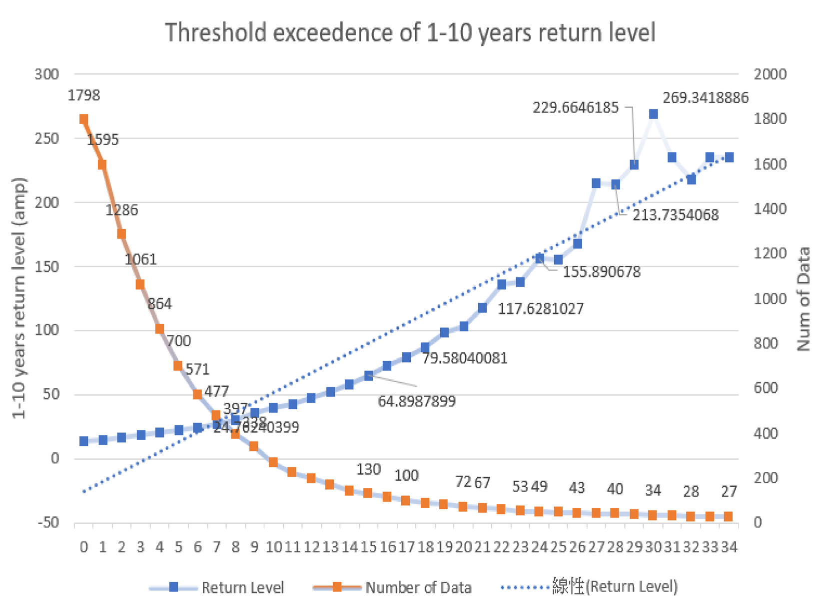
## Coronal Mass Ejections

- **Coronal Mass Ejections (CME)** are significant release of plasma and magnetic energy from the Sun, known to cause large disturbances in the Earth's magnetosphere. The resulting magnetic variations create large geomagnetically induced currents (GIC), known to cause massive power outages and significant damage to society's electrical infrastructure (i.e power grid, communication lines).




*Source: NASA Scientific Visualization Studio*

## Extreme Value Analysis

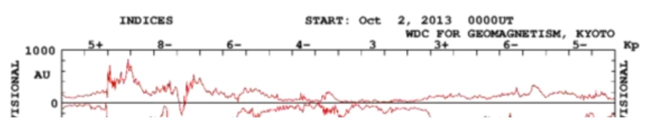



- The **Generalized Extreme Value (GEV) model** seeks to predict and quantify the stochastic nature of these extreme GIC events through historical data.
- Our GEV model conveys a direct correlation between higher thresholds and higher return values, indicating a strong dependency of extreme values on the chosen threshold.
- This sensitivity highlights the importance of selecting appropriate thresholds to capture the most relevant extreme events while minimizing statistical bias.

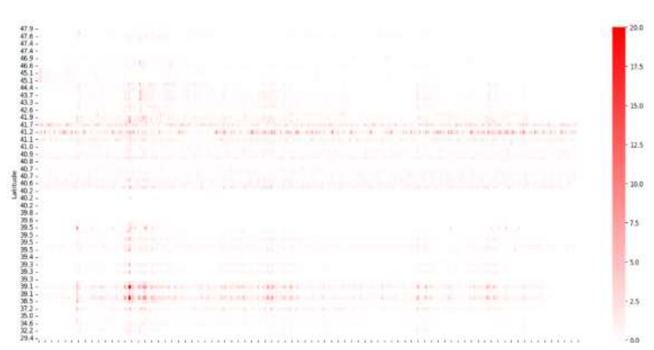
## Wavelet Decomposition & Cross Correlation

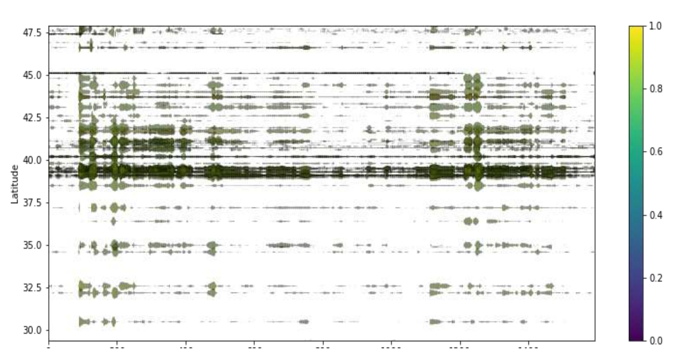
- A 4-level wavelet decomposition of historical GIC data on each station was conducted utilizing **Haar wavelets** to detect the greatest rate of change of GICs. Other wavelets were also applied to compare smoothness of coefficients, between wavelet types.
- After performing **Maximal Overlap Discrete Wavelet Transform (MODWT)**, a **sliding window** aggregated a small timeframe (30 min) of preprocessed data to create time-based nodal graphs with eigenvector centrality measures. These graphs utilized a set of minimum **device-specific thresholds** to determine the correlation connections versus a global variable, in order to prevent bias towards highly active geomagnetic monitors. Equations used for **optimizing threshold**:

$$A^*_{ij}(C_T,t) = \Phi(|C_{ij}(t)| - C_T). \qquad n_i(C_T) = \left(\sum_t^T \sum_j^{N(t)} \frac{A^*_{ij}(t,C_T)}{N(t)-1}\right)/T$$




*Source: AE INDEX – Kyoto Geomagnetic Auroral Electrojet*

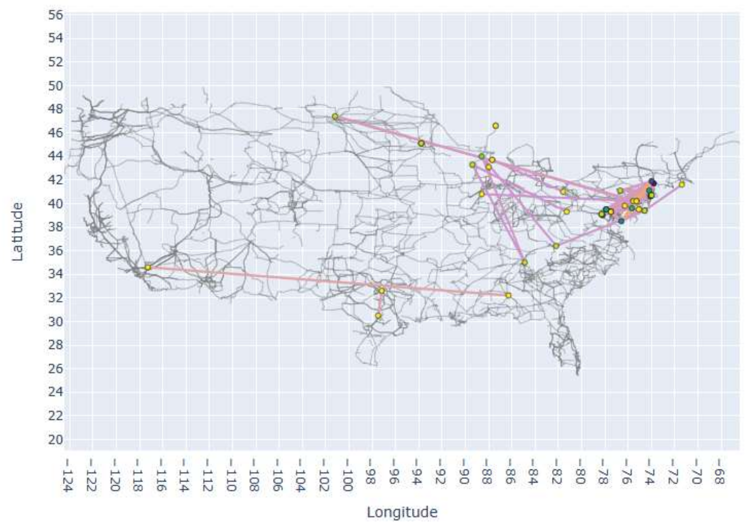




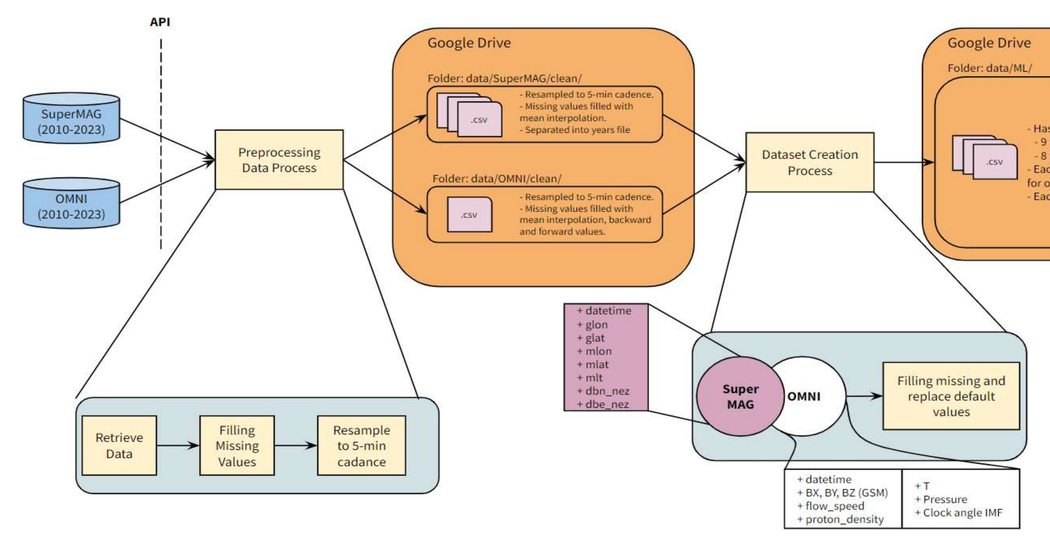

*Source: North American Electric Reliability Corporation GIC Measured Data for 2013 OCT 03*

## Network Analysis

- Sliding window cross-correlations indicated **long-range connections** across east-west regions, notably at the onset of a solar storm.
- This is consistent with previous research, where depression of the magnetosphere due to a CME's arrival often causes an increase in geomagnetic activity amongst similar latitudes regardless of sun-facing orientation.



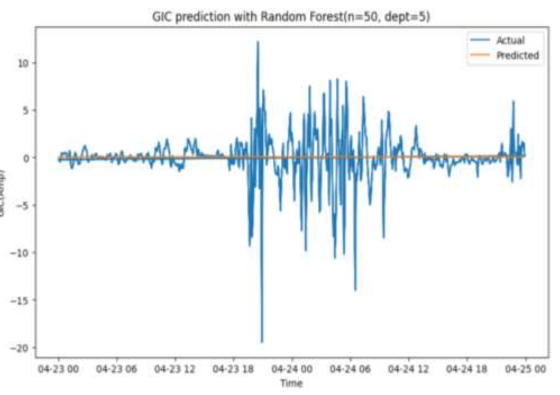

## Machine Learning Data Flow

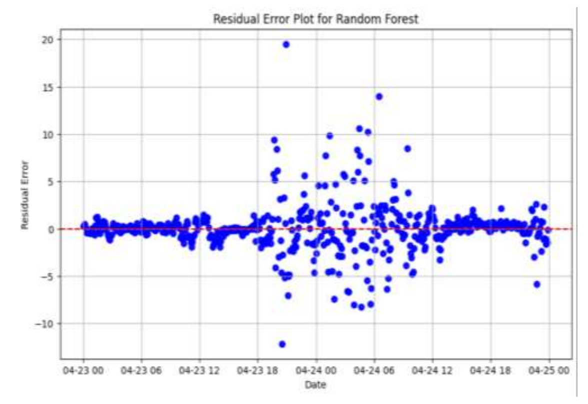



## ML Predictions for Geomagnetic Induced Current (GIC)

### Random Forest Regressor Model

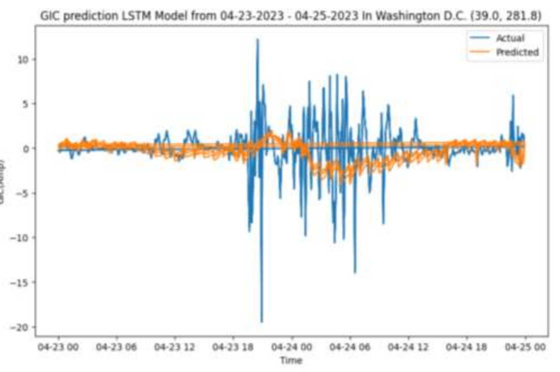
- RFRM performed poorly in predicting GICs with high residual errors.
- Parameters
  - max_depth = 5
  - n_estimator = 50

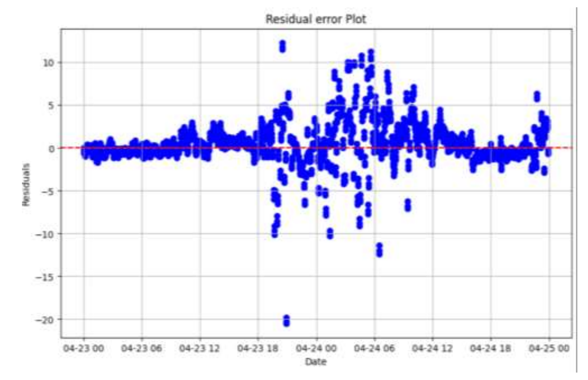



### Long Short Term Memory Model

- LSTM model was more optimal for GIC predictions, reflecting overall trend of data.
- LSTM seeks to reflect the trend of the data.
- Parameters:
  - Epochs = 1,
  - batch_size = default

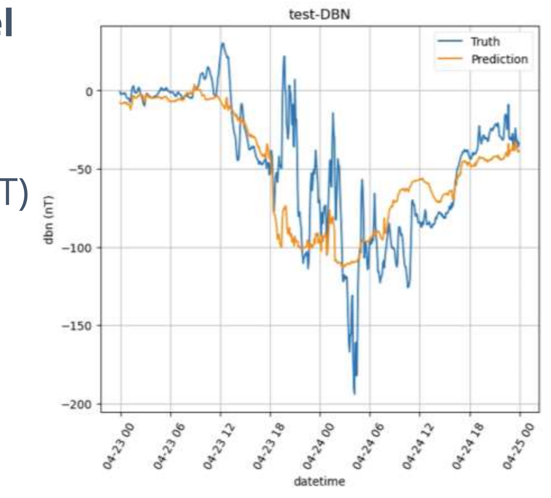


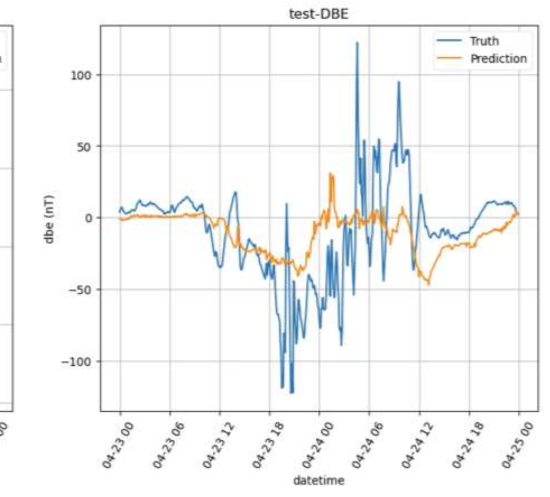


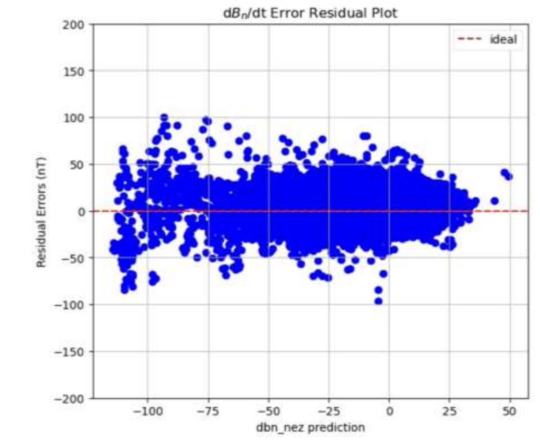

## ML Predictions for Magnetic Perturbation (dB/dt)

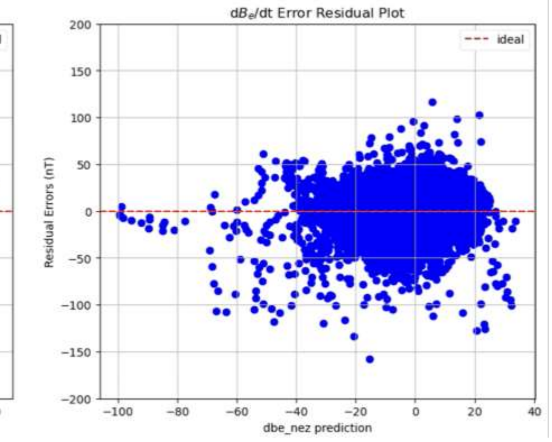
### Random Forest Regressor Model

- RFRM resulted in mediocre performance in predicting magnetic perturbation values (nT) for the North and East axes.
- However, the results are essential in understanding the dynamics and the characteristics of the data.
- Model Parameters:
  - max_depth = 90
  - max_features = 0.5
  - min_samples_leaf = 8
  - min_samples_split = 5
  - n_estimators = 911
  - Loss = RMSE
- The model captures low-magnitude and non-linear trends, however, fails to learn the high-magnitude dynamics of the data.



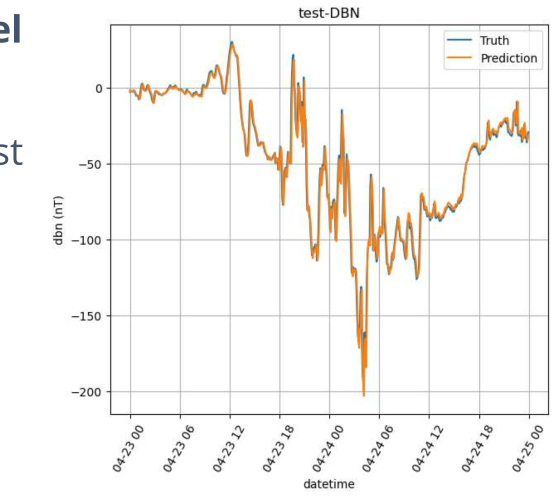



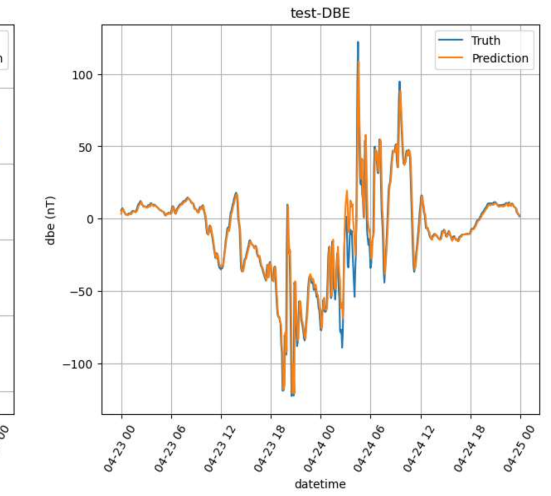



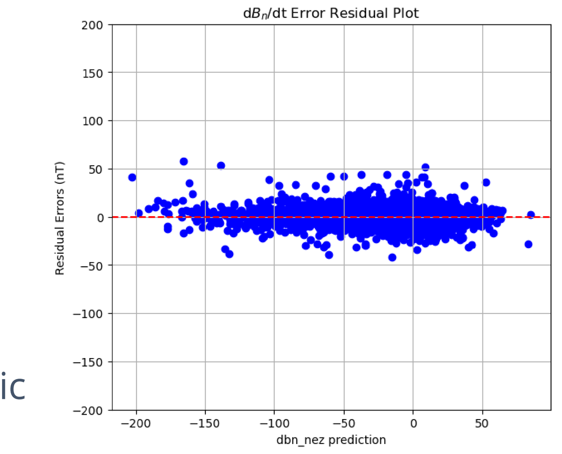



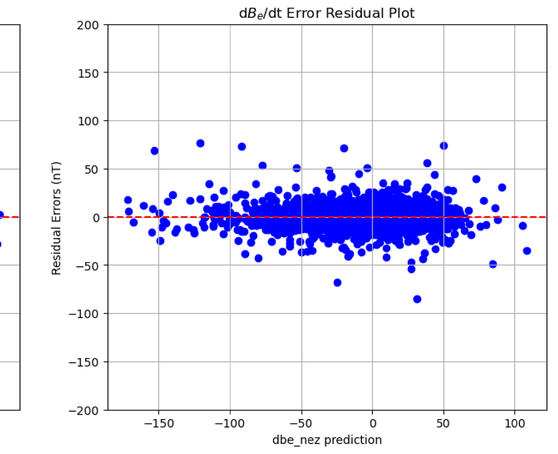

### Long Short Term Memory Model

- LSTM model had the highest performance over random forest and multilinear regressions.
- Model Parameters:
  - Batch_size = 360
  - Optimization = Adam
  - Loss = MSE
  - Epochs = 50
- Comparison between models' residual plot indicates LSTM's greater performance through more centralized graphs.
- Most notably, LSTM model has superior performance when tested at small time intervals, able to predict accurate magnetic perturbation of 5 min intervals.




## Future Work, References, and Acknowledgments

- Integrating extreme value and network analysis in fine-tuning ML models during training for increased nowcasting/forecasting.
- Simulate the resiliency of power grid based on network analysis.

[1] L.Orr, S.C Chapman, C.D. Beggan. *Wavelet and Network Analysis of Magnetic Field Variation and Geomagnetically Induced Currents During Large Storms*. AGU Publication. August 2021
[2] V. Upendran, P Tigas, B. Ferdousi et. al *Global Perturbation Forecasting Using Deep Learning*. AGU Publication. November 2022
[3] M. Blandin, H. Connor, D. Ozturk, A. Keesee, et al. *Multi-variate LSTM Prediction of Alaska Magnetometer Chain Utilizing a Coupled Model Approach.* Frontiers in Astronomy and Space Sciences. May 2022

ELECTRICAL & COMPUTER ENGINEERING
UNIVERSITY of WASHINGTON

**ADVISER:** Ryan McGranaghan, Jonathan Sauder, Arindam Das, Kavya Balasubramanian

**SPONSOR:** NASA Jet Propulsion Laboratory