Abstract

Historically, variations of the Earth’s magnetosphere due to Coronal Mass Ejections (CME) are known to induce geomagnetically-induced currents (GIC) that can cause significant damage to society’s electrical infrastructure. Predicting and understanding the frequency, characteristics, and subsequent effects of GICs requires deliberate data framework and analysis to find meaningful insight. In response, this research has created a data pipeline that aggregates the GIC measurements amongst public domain datasets (i.e SuperMAG, OMNIWeb, NERC) to structured schemas in facilitating extreme value and network analysis. Generalized Extreme Value Model outlines the estimates the 1-in-100 year catastrophic solar events, resulting in inversely, non-linear trend between return values and pre-defined return period (T). Quantile-Quantile plots analyzes the characteristics of regions, differentiating between supernodes of greater GIC activity by magnitude. Connecting the overall GIC response required an additional, time-based network analysis, utilizing wavelet decomposition to identify long and short-term trends prior to cross-correlation via the sliding window method. Cross-correlation with threshold optimization in lieu of global threshold found to be computational expensive, yet solidified the long east-west connections (Orr 2021) upon increased GIC activity. Overlaying of physical grid to visualize the potential impact of GIC variations provided uninteresting conclusions, however, highlighted additional areas of interests through eigenvalue centrality measures beyond active regions surrounding super-nodes. Predictions of these GIC events and magnetic perturbations were modeled using three machine learning (ML) models, including multi-linear regressions (MLR), random forest regression (RFR), and long-short term model (LSTM). Given a 5-min interval events, LSTM had the best performance in modeling using ReLu activation and Root Means Square Error (RMSE) loss function with Adam optimization. Batch size was set at 360, since smaller batch size induces a slower and less stable training process, while a larger batch size leads to poorer generalization.

Are most GICs measured in a node usually normally distributed?

Are most GICs measured in a node usually extremely distributed?

And also label the blue histogram to show GIC (I’m assuming it’s the GIC)

Let me know when you’re free this week.

I think it’ll be better if we can just talk it through.

-Sean-

Mazzy

Great writeup! I think it’ll