CONSIDERATIONS FOR REAL-TIME DATA ANALYSIS USING MULTIPLE

MAGNETOMETER SOURCES FOR GIC STUDIES TO IMPROVE THE

SITUATIONAL AWARENESS OF AN ELECTRIC GRID MODEL

A Thesis

by

RAMYAA RATHNA KUMAR

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,   Adam Birchfield
Co-Chair of Committee, Thomas Overbye
Committee Members,   Le Xie
                       Weiping Shi
                       Jennifer Marshall
Head of Department,   Miroslav Begovic

August 2021

Major Subject: Electrical Engineering

ABSTRACT

The electric power grid which interconnects generation sources, to the consumer ends is a combination of several power system components. Events such as geomagnetic disturbances (GMD) inflict damage on these components depending on their degree of severity. Hence, it is essential to obtain proper knowledge of the electric grid model for analyzing the vulnerable regions and accordingly devising mitigation strategies to improve grid resiliency, thereby minimizing the extent of losses.

This thesis aims to effectively utilize actual historical and real-time data to enhance electric grid model knowledge by reconstructing GMD scenarios. Magnetic and electric field data used for the analysis are generated and calculated from Texas A&M Magnetometer Network (TAMUMN). Days of GMD activity (G2, G1) during the past year are selected to reconstruct events with similar data on a synthetically developed version of the Texas electric grid with 7000 bus network. Upon integrating the data to a simulated power system model, the impact of geomagnetically induced currents (GIC) can be determined. By performing certain power system analysis techniques, the most affected regions, magnitudes of maximum current at substations and the transmission lines with highest activity can be obtained. This is greatly useful for planning purposes and studies as it directs the user to focus on a specific section of the grid model and works toward strengthening it.

A significant focus of the thesis is on data analysis, visualization, and interpretation of the causal data to enhance the user's knowledge and familiarity with each dataset and

its differences for various levels of GMD activity. The resulting induced voltages on each transmission line of the system is calculated using TAMUMN data. The entire process is automated for all the events and sensitivity analysis is performed on the generated data to improve the situational awareness of a given electric grid model.

The final sections of the thesis discuss the importance of the analysis on a given electric grid model and the future scope of improvement in areas of mathematical models for real-time data which can be followed by event detection and GIC monitoring using real time data for vulnerability assessment.

ACKNOWLEDGEMENTS

I thank my adviser Prof. Adam Birchfield, for his constant support and motivation throughout this journey. As a great mentor, he directed me towards this thesis, encouraging my ideas with constant feedback. I am extremely grateful to Prof. Thomas Overbye for discovering me as a graduate student capable of research, even while I was in self-doubt. His futuristic vision has always inspired me to take up challenging problems and look for its real-world applications. I also thank each person under Overbye research group for always being approachable, making me feel comfortable from the start and helping me learn new topics.

I am thankful to my mother and father, Manjula Rathna Kumar and S. Rathna Kumar for always being my pillars of support and for their unconditional love and prayers, along with my little brother Sudharsan Rathna Kumar who never stopped cheering me and always made me feel the best. I am incomplete without them in every sphere of life.

I am thankful to all my friends here at Texas A&M University who were with me through thick and thin in the past two years and for sharing a bond no less than a family. Finally, I am humbled and grateful to the Almighty for providing me with a great family, friends, institution, and professors. I hope to become a good engineer, uplifting the values embedded in me by my parents and teachers.

CONTRIBUTORS AND FUNDING SOURCES

Contributors and Funding Sources

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Electric energy generated at power plants, moves through a complex system, known as the electric grid, which is an interconnected network for electricity delivery from producers to consumers. It consists of essential components such as substations, transformers and power lines which span over large distances, connecting thousands of generation sources to hundreds of millions of electricity customers across the country. According to electric grid reliability standards, a power system should be able to deliver electricity to all points of consumption, without degradation or failure. To ensure reliable electricity supply, it is necessary to improve grid resiliency, primarily by generalizing the system and understanding its response to disturbances, secondly through the development of proper tools for event monitoring and management.[1][2]

High impact low frequency (HILF) events such as geomagnetic storms [3][4] have negative repercussions [5] on the electric power grid. Such events are responsible for altering the Earth's magnetic field which is in turn responsible for inducing electric fields into the grid [6]. High voltage bulk power transformers, including generator step-up and major substation step-down transformers, are critical assets within the grid, because of their large area of impact, cost, and time to replace [7]. For these systems, the effects of currents due to such geomagnetic disturbances (GMD) can include harmonic currents that can cause relays to trip equipment, fringing magnetic fields that can create heating in the transformer, increased reactive power consumption that can cause the system to collapse

1

due to voltage instability along with damaging customer equipment due to power quality disturbances [8]. History of occurrences of such GMD events can be traced back to the 1940 event at Philadelphia Electric Company, which recorded strong reactive power swings and voltage surges throughout the electric grid [9]. The 1989 Hydro-Quebec GMD event at Canada, affected six million customers, who lost power for more than nine hours. Similar to this, the 2003 GMD event, cost the loss of a $450M research satellite at South Africa.

The magnitude and direction of these fields depend on the storm, as well as the conductivity of the Earth, resulting in geomagnetically induced currents (GIC) [10]. They flow in the Earth and high voltage transmission lines, due to the lines' low dc resistance, as well as metal piping, telecommunication cables and railway infrastructure. The presence of geoelectric fields in the power system induces the flow of such low frequency currents, (GICs) which have the potential to cause damage to majority of power system components such as transformers [11][12].

This thesis focuses on achieving a reasonable understanding of the electric grid model to increase system awareness and remain prepared in the events of geomagnetic disturbances where assets and customers are at a greater risk of damage and loss of electricity. Existence of magnetic field data and electric grid model is utilized here to gain system knowledge and determine the most vulnerable regions, hence, increasing the efficiency of the analysis by redirecting the user's attention towards a small fraction of the entire grid. Throughout this work, historical and real-time data is extracted, pre-processed, and visualized using multiple paradigms. Integrating data to power system helps to

recreate the actual scenario, which serves as an environment for analysis and monitoring purposes.

1.2. Existing Research and Current Challenges

Previous research has focused on analyzing GMD events and performing simulations to enable real time measurement and monitoring of these events [13]. A large fraction of this work uses emulated sensor measurements for magnetic field data [14]. There have been several approaches to develop ground models for determining the electric field of a region. So far, for most power system studies, typically a uniform conductivity model, or a layered 1D model with different conductivity values at different depths, have been used [15][16]. The geophysics literature describes other more detailed methods of calculating electric fields. One of them is using the electromagnetic transfer functions (EMTFs) approach that represents the 3D conductivity response of the Earth [17].

This thesis focuses on using recently created EPRI (Electric Power Research Institute) ground conductivity models, which are based on 1D transfer functions that differ across regions. Updated 1D regional boundaries of different states based on regional averages of local empirical transfer functions is documented by EPRI. Further, actual measurements of magnetic field data, captured by four sparsely located magnetometers of Texas A&M Magnetometer Network (TAMUMN) is extracted and mapped to all power system components in the electric grid, located within a certain spatial range of the latitude and longitude considerations for developing the GMD case scenarios. A synthetic version

of the electric grid model of Texas, developed with a 7000-bus design is augmented for this research.

Studying the effects of GMD on the electric grid using NERC (North American Electric Reliability Corporation) benchmark events has been dealt in several previous works with the aim of creating the worst-case scenario of GMD on the electric grid model [18][19]. Although this approximation can be useful to set a threshold on the power system values and the GIC current limits [20], it often overestimates or produces a result of extrapolated values. This is not particularly useful to identify elements such as transmission lines and transformers which are in the verge of vulnerability and have sufficient tolerances for only a certain specific range of values. Hence it is essential that the system be tested with the exact data that affected it in the past, to understand the trends in the system.

Accessing magnetic field data for each location can be further enhanced with greater number of magnetometers [21], which will shorten the range for assumptions of magnetic field over large landscapes. Integrating data with a power system software tool has few challenges that cannot be generalized and depends on the format of each data, range of values and its dimensions. Much of the initial section is spent on pre-processing sensor generated magnetic field and electric field data resulting from applying ground models to the former [22].

1.3. Thesis Organization

The main contribution of the work done in the thesis is to leverage existing data for understanding the electric grid model's behavior during events of disturbances. This document is divided overall into three major sections, which describes the workflow undertaken. Employing data analysis metrics and visualization for a huge data set to reconstruct scenarios for power system analysis on the electric grid is a holistic approach for improving its situational awareness. All the sections add value and scope to future scalability, for incorporating larger scenarios. Each section below summarizes the segment of work involved in a brief sense.

The initial section is on data preprocessing and visualization which considers magnetic field data corresponding to all days of activity, preprocessing, detecting outliers, and developing real-time visualization. Further, a comparison of plots from all sources of TAMUMN for different scenarios is also performed and a geographical plot for the spatial distribution of the electric and magnetic field is obtained. This also validates the ground model configurations used [15].

The next section deals with electric grid modeling and power system analysis on the synthetic grid model using the data. This serves as a testbed to integrate real-time data for evaluating all the scenarios under consideration. This also provides an overview of the regions which are always at a larger risk of damage due to the events, suggesting possible remedial action schemes [23]. Different GMD scenarios are constructed, and the resultant GIC currents are analyzed.

To work on strengthening the situational awareness of the system, vulnerability assessment of the areas is enhanced by developing sensitivity analysis for the selected and most affected transformers. This determines the proximity of the most affected branches and performs calculations to evaluate the impact of GIC currents [24], concluding the final section. A quick summary of each of the chapters is listed below.

Chapter 2 navigates through the preliminary tasks done on the magnetic field data. This includes providing a brief overview on the location and existence of local (TAMUMN) magnetometers and its comparison with few other closely located magnetometers. Further, the description of the GMD events, that are reconstructed for analysis, is elaborated. This chapter addresses major topics done with the data such as visualization and statistical analysis.

The need for electric field, as a primary data to effectively demonstrate the GIC analysis is discussed in Chapter 3. Here, the focus is on different ground models and its usage to calculate resultant electric fields. The scope is extended to develop visualization for different electric field plots, to compare multiple magnetometers and ground models for all the scenarios under consideration.

In Chapter 4, the discussion is routed towards building the grid model. A synthetic version of the actual Texas electric grid model is originally developed as a 7000-bus system. The addition of synthetic substation grounding resistances, transformer configurations and relevant power system parameters to the grid model is elaborated here. A larger part of this chapter is devoted towards the issue of reconstructing a GMD scenario and its extension for all the cases.

Chapter 5 deals with performing GIC sensitivity analysis on the electric grid model developed for all the cases. This is done to deliver results on the most affected transformer, its metrics, and the vulnerable regions in the grid. A summary and contribution of this work and future directions are discussed in Chapter 6.

2. MAGNETIC FIELD DATA*

This chapter focuses on illustrating and interpreting the magnetic field data obtained from the Texas A&M Magnetometer Network (TAMUMN). Data correlation, statistical analysis and visualization is elaborated here.

2.1. Texas A&M Magnetometer Network (TAMUMN)

The Texas A&M Magnetometer Network (TAMUMN) captures the magnetic field at specific locations, as shown in Fig 2.1. The installations consist of a magnetic field sensor, data acquisition and transmission unit, and a solar power system. This data is recorded at a cadence of 1 Hz and transmitted in real time (latency < 10 ms) to a data center (server) at TAMU, for the purposes of display, access, and archiving [25].

Figure 2.1 Texas A&M Magnetometer Network Locations: Amarillo (AMR), Beaumont (BMT), Beeville (BVL), Overton (OVR), RELLIS (RLS), and Stephenville (STP). Odessa (ODS) station in Far West Texas was installed under a prior project; its data will be added to the TAMUMN shortly [25].

In selecting the sites for the six magnetometers of TAMUMN, the idea was to distribute them across the state of Texas for maximum coverage. Another consideration was to locate them close to at least some existing GIC monitors. According to previous studies [26], a magnetometer should be located reasonably close to a substation for the calculated geoelectric field (and hence GIC) to be close to the actual field at the substation. Determining how close they need to be is one of the goals of the research for which this data is being collected. Considering all these factors, the magnetometers were sited at some of the Texas A&M AgriLife campus locations, that are present throughout Texas. These locations are Amarillo (AMR), Beaumont (BMT), Beeville (BVL), Overton (OVR), and Stephenville (STP). The sixth magnetometer is at the RELLIS (RLS) campus in College Station, TX and is also where the server and all the data is hosted. Fig 2.1 shows all these magnetometers, with a seventh in far west Texas shown in blue representing an older (currently offline) installation done by the institute. Work is ongoing to upgrade this station and add its data to the TAMUMN, to improve estimation of the geoelectric field across Texas.

2.2. GMD Description and Days of Activity

Days of geomagnetic activity throughout the past year are obtained from the National oceanic and atmospheric administration (NOAA) [27][28]. Fig 2.2 represents the geomagnetic activity chart downloaded from NOAA for a period of 15 days. Days of G1 and G2 level activity are taken under consideration and the magnetic field data (**B** field) obtained from Texas A&M Magnetometer Network (TAMUMN) corresponding to these

9

days is preconditioned and visualized. The level of activity is defined according to space weather prediction (SWPC) standards which justifies the nature of geomagnetic storms based on the 'Kp Index' parameter [29]. When the Kp Index is 5, the event is classified as a 'G1 event' which can impact the power system to experience minor power grid fluctuations. Events with Kp Index as 6 are categorized as 'G2' level storms. In this situation, high latitude power systems may experience voltage alarms, long duration storms may even cause transformer damage.
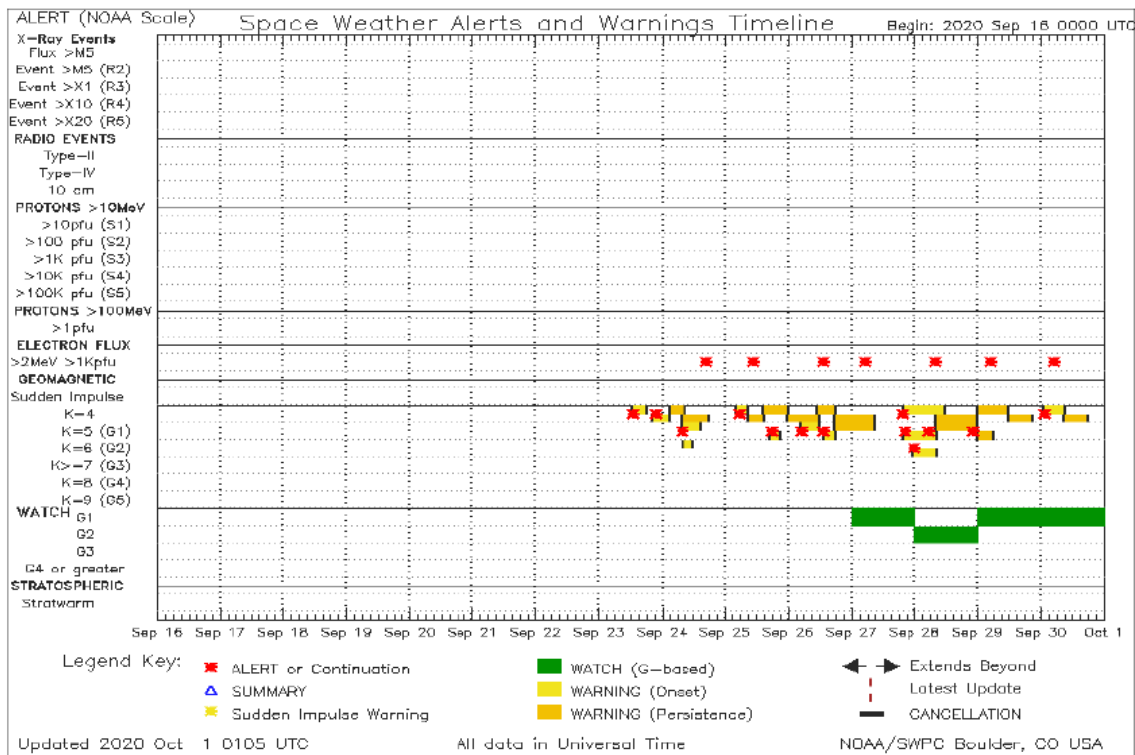
Figure 2.2 Sample geomagnetic activity chart, downloaded from NOAA [33].

Days of geomagnetic activity observed during the past year, that are reconstructed using the similar magnetic field data are on 09/28/2020 (G2 event), 10/23/2020 (G1 event) and for comparison purposes, a quiet day without any geomagnetic activity is considered

on 10/09/2020. The upcoming sections deals with analysis techniques on the magnetic field data for each of the scenarios.

2.3. Data Specifications

Magnetic field data captured by TAMUMN is streamed online to a secure website which stores the data, downloadable in a CSV format, with a 1 second resolution. The historical record is available from the time of installation and is downloaded for the required days of observation. Initial preprocessing is performed on the data to generate 10 seconds spaced magnetic field data along the three axes of the magnetometers, Bx, By and Bz. Data generated by Amarillo, Beaumont, Beeville, and Stephenville magnetometers out of the TAMUMN are utilized for all the analysis purposes. Historical and real-time data visualization is performed. Prior to any statistical and graphical analysis, the data is also observed for any outliers which are identified and eliminated from the dataset. For further validation, magnetic field data of three USGS (United States Geological Survey) observatories are obtained from Intermagnet database [30] and preprocessed to match the format of the local TAMUMN data for correlation purposes. A flowchart depicted in Fig 2.3 summarizes the process and relationship of TAMUMN magnetometers for potential power system monitoring, control, and analysis.
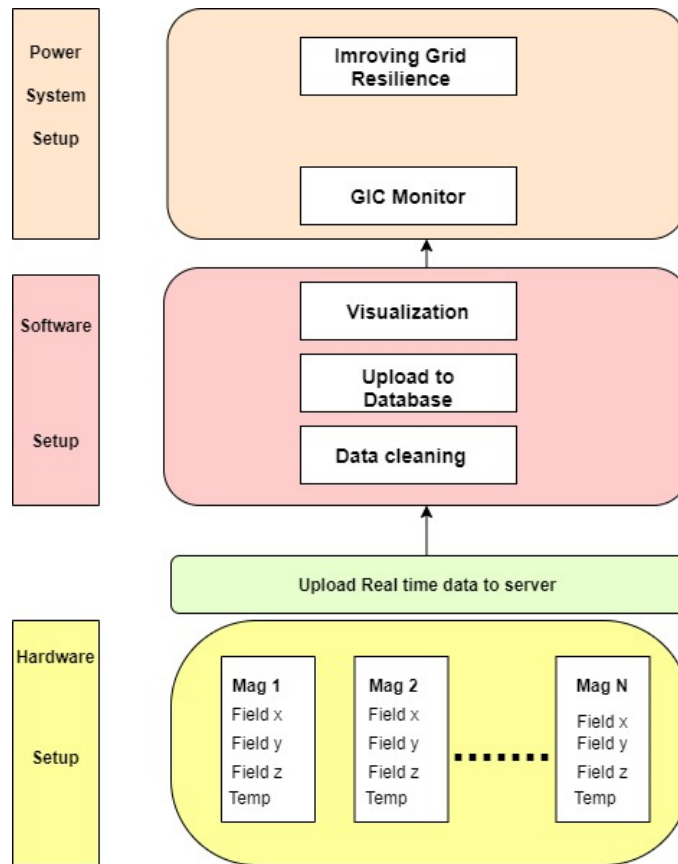
Figure 2.3 Flowchart describing software and hardware setup for magnetometers [25].

2.4. Visualizations

The initial step towards reconstructing GMD scenarios is to understand and familiarize with the causal data. Several plots are developed for visualizing the historical magnetic field data of TAMUMN devices. A common pattern is observed based on the magnitudes of the spread of the data. In the underlying subsections, the distribution of the data over the entire day is visualized to observe the trends and deviations from the assumed hypothesis of the events.

2.4.1. Magnetic Field (B field) Visualizations of TAMUMN for Different Events

The three events of G1, G2 and a normal day is considered, and the respective **B** field data obtained across all the TAMUMN devices is plotted on a single graph. Fig 2.4, 2.5 and 2.6 displays the variation of **B** field throughout each day for different events. Along the vertical axis is the **B** field captured by TAMUMN in nT. The scale along the horizontal axis shows the variation of time within a day in the 24-hour format of HH: MM: SS. To avoid crowding of all points on the horizontal axis, every 1000 points is chosen for display, hence showing few time points.
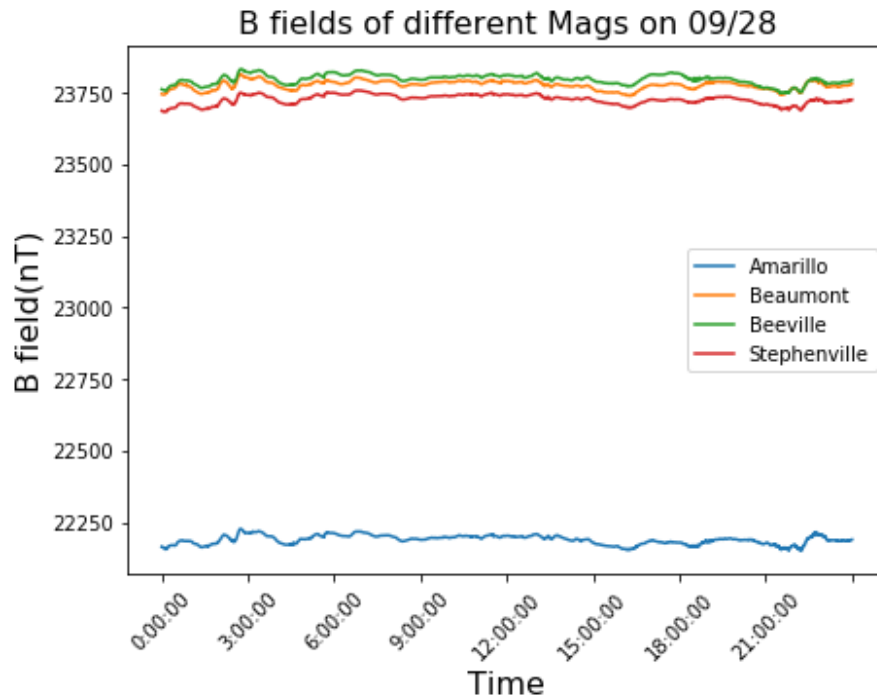
Figure 2.4 B field across TAMUMN devices during a G2 event.
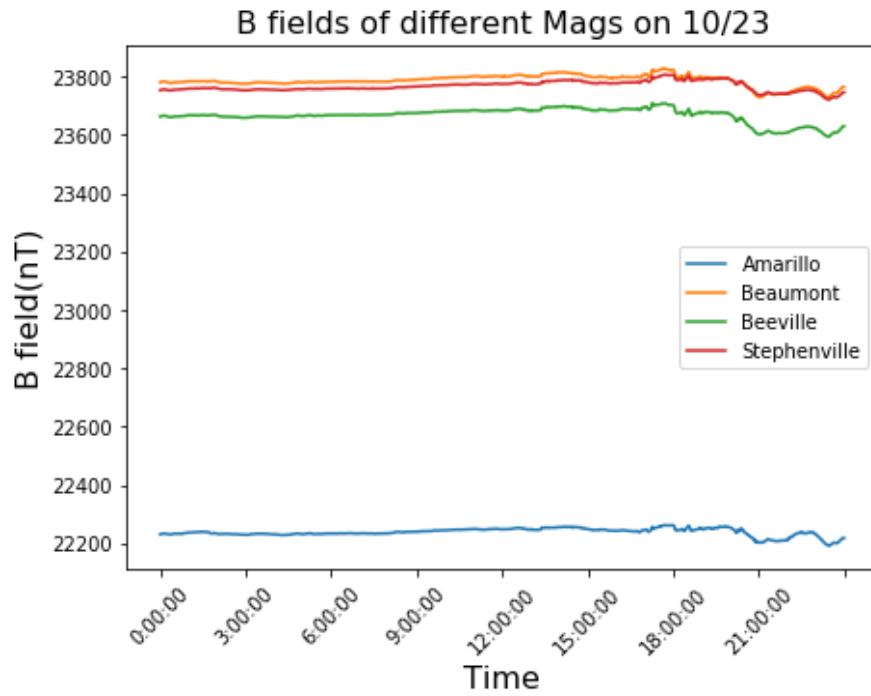
Figure 2.5 B field across TAMUMN devices during a G1 event.
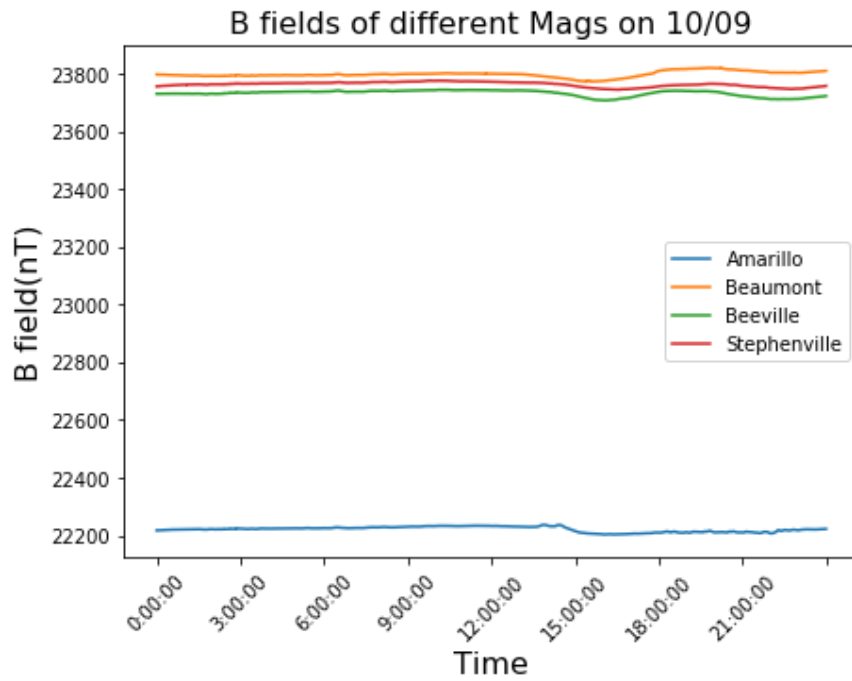
Figure 2.6 B field across TAMUMN devices during a quiet day.

14

A general observation from the graphs above is that all the lines indicating the **B** field values of magnetometers show a common overall pattern. The **B** field values varies within 1600nT for the stations recording the least (AMR) and the maximum value (STP). This small change between the ranges can be attributed to the differences in the location and geographical coordinates. BVL, BMT, and STP have the ranges of latitude and longitude between 28°- 30° and 95°-98° respectively, however AMR has coordinates of 35° (latitude) and 103° (longitude). The relative difference in its spatial location, is clearly visible on Fig 2.1. This could be a possible cause for the minor differences in the scale of **B** field values of AMR with the others. Greater presence of peaks in the graphs are observed during a G2 event than a normal day, which exhibits relatively flat lines. Plots for G1 event shows deviations from the flat line, towards the end of the day, indicating the occurrence of a minor geomagnetic activity. The existence of peaks and valleys in G2 and G1 plots proves the presence of a differential of the **B** field. This plot gives an insight on the electric field which theoretically is a derivative of the magnetic field, depending on the ground models, specific to a location. Hence, TAMUMN sensor data visualization projects the magnetic field in different events, directing towards further investigation.
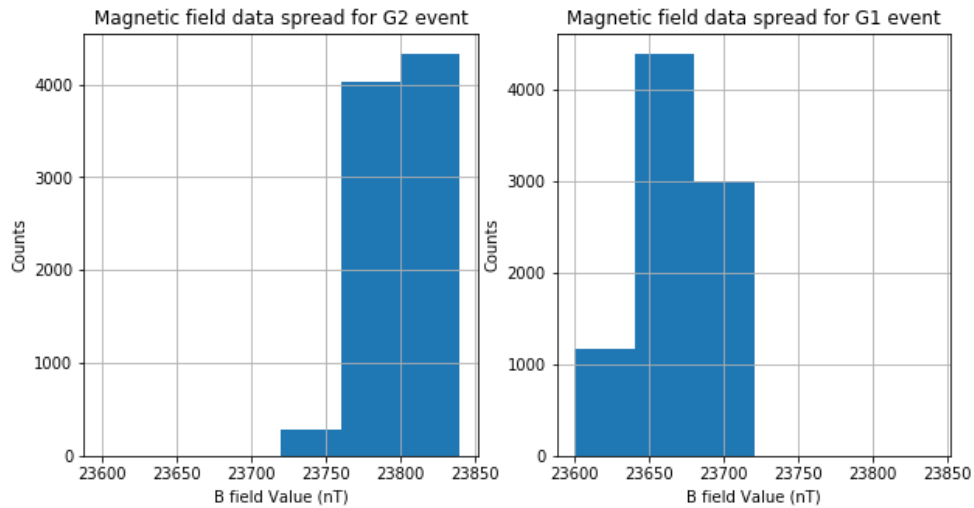
Figure 2.7 Histogram for G2 and G1 event.

Fig 2.7 depicts the frequency of distribution of the **B** field data, of BVL. For all time points of observation, the y axis takes in the counts of the **B** field measurement in nT, which is displayed along the horizontal axis. The insights obtained from the graph is that a higher occurrence of **B** field values within the range of 23760- 23840 nT during a G2 day, and between 23650-23700 nT during a G1 day is observed. The information captured from this histogram is to leverage the spread of data over a wide range of values for further calculations. This proves that G2 event which is characterized by frequent existence of peaks have greater occurrences of higher magnitudes of **B** field data.

2.4.2. Event Based Magnetometer B Field Comparison.

To visualize the response of a single magnetometer during all the events, plots of the **B** field for each magnetometer is created in Fig 2.8 and Fig 2.9. Since similar information is conveyed by all TAMUMN magnetometers, only the plots at the stations Amarillo and Beeville are shown. Clearly the range of values captured by both the TAMUMN devices, Amarillo, and Beeville, have a small difference for similar events. This is observed by the limits of the vertical axis on both the graphs. The information extracted from these graphs is based on observation that. G2 and G1 events are characterized by a peak or drop in the **B** field, indicating frequent changes in the values from an assumed steady state level. Minor variations in the **B** field of a normal day does not account for any established event. Since GMD activities occur during certain intervals of a day, these graphs can also be used as an indicator to trace the periods of such occurrences and estimate the impact levels.
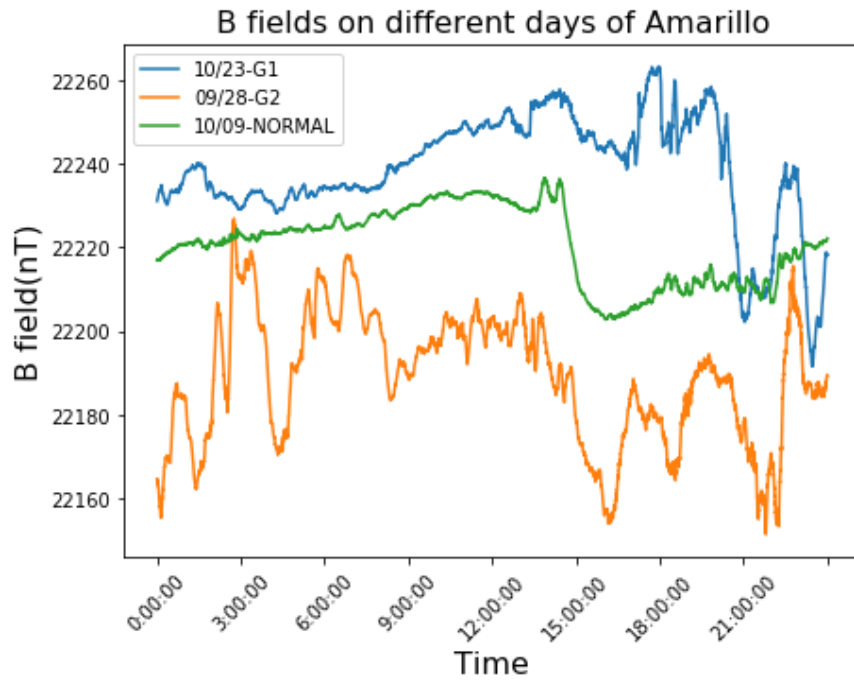
Figure 2.8 B field variation across TAMUMN Amarillo for all events considered.
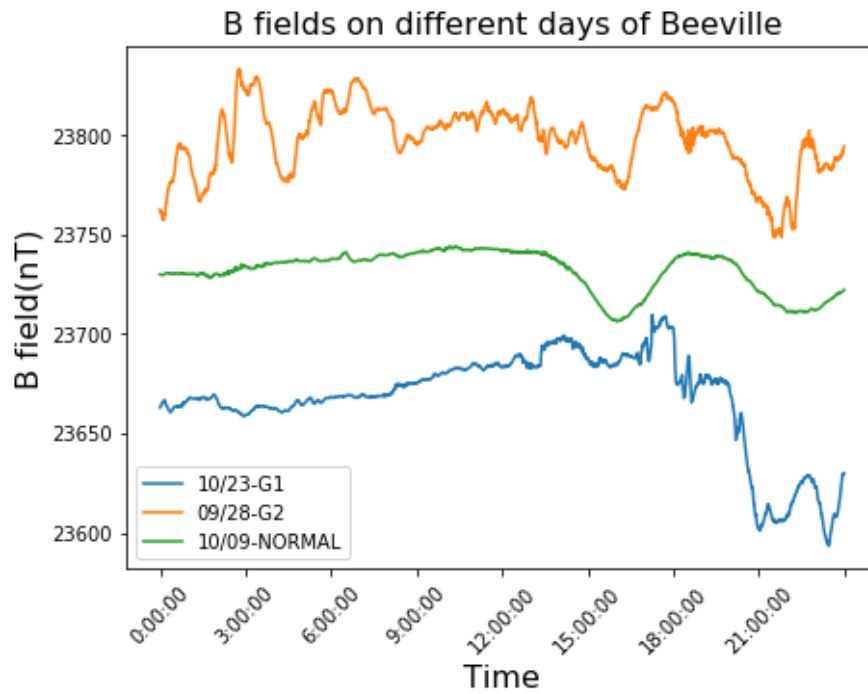
Figure 2.9 B field variation across TAMUMN Beeville for all the events considered.

18

2.4.3. Real-Time Visualization

The data from each station of TAMUMN is streamed online, which is queried every second using a script written in Python language that runs in real time and converts it to JSON format [31]. A dashboard is developed using the Open-Source tool Grafana [32] which receives magnetic field data as an input, with values updated every second to display a time-varying graph. Small- and large-time frames can be chosen using the interface, to view real-time variations in the magnetic field over the Texas footprint. Fig 2.10 visualizes multiple data streams i.e., data from five magnetometers namely AMR (Amarillo), BMT (Beaumont), BVL (Beeville), OVR (Overton) and RLS (RELLIS). A snapshot for the time frame of an entire day is captured from the website (Grafana) for the sake of displaying here. The figure consists of the magnetic field along the horizontal direction (Bx) and the vertical direction (By) captured by the five devices in the upper half and lower half sections of the graph, respectively. The horizontal axis keeps changing according to the time at a given instant and the vertical axis provides the corresponding magnitude of the magnetic field in nT. The purpose of such a real-time visualization is to demonstrate the magnetic field value at any selected periods of time. On the right of each graph, statistics on the magnetic field values are displayed for each of the TAMUMN devices used in the visualization chart. The visualization is dated on 05/26/2021, during when a G1 event was observed. The limits and ranges of the magnetic field Bx and By differs according to the inherent characteristics of the magnetic north and magnetic east components of the Earth respectively. The change in the magnetic field is depicted by the

19

deviations in the graph, from a steady value, which can be attributed to the presence of a

GMD event, that causes a change in the Earth's magnetic field.
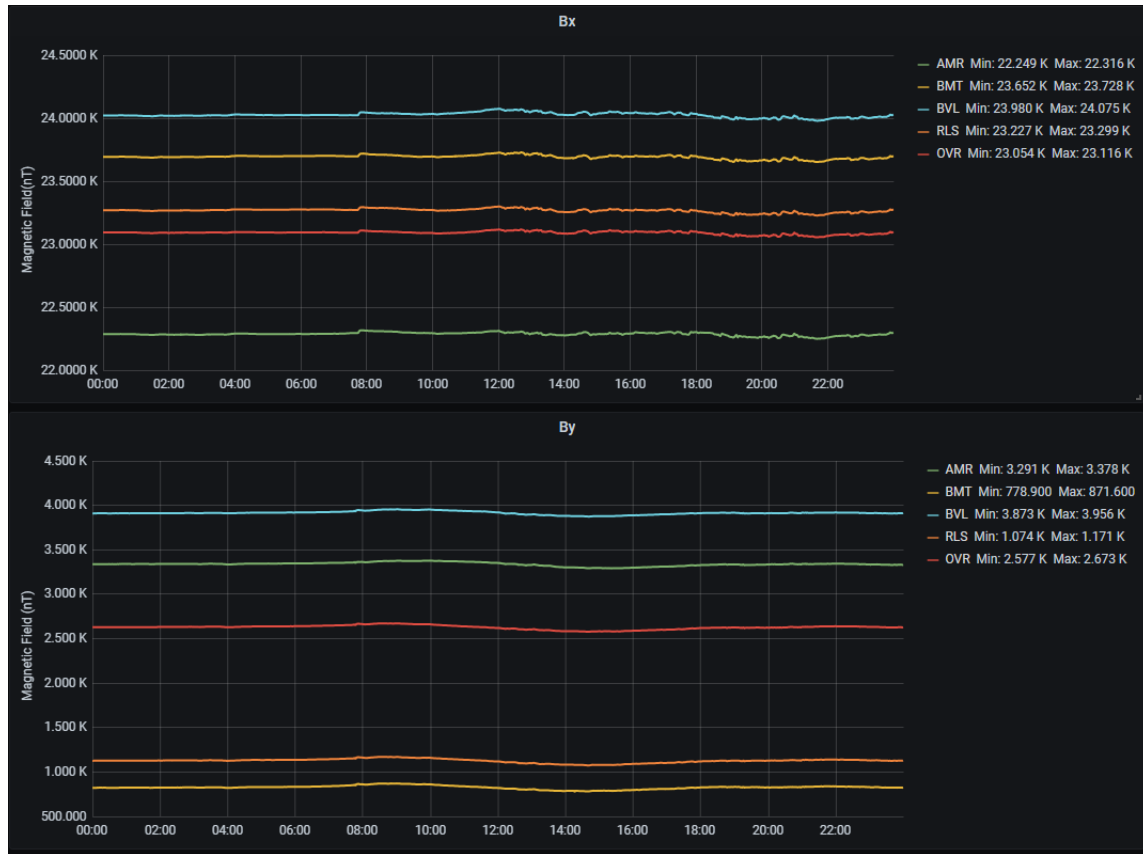
Figure 2.10 Real-time visualization of B field data of TAMUMN devices, taken over a snapshot of a single day during G1 activity.

2.5. Correlation Analysis

2.5.1. Description

Correlation analysis is a statistical method used to evaluate the relationship between two quantitative variables, where results are assigned numerical values between -1 and 1. A high correlation means that two or more variables have a strong relationship

with each other. Meanwhile a weak correlation indicates that the variables are hardly related. In [33] correlation is performed between magnetic field data of different observatories, which reveals the evolution of the **B** field over the Earth's surface, during a storm. Here, the closest magnetometers to the transformer neutral GIC measurements of interest were over 800 miles away and approximations using interpolation were made to perform model validation.

Here, correlation analysis is used to assess the closeness of magnetic field data obtained from the TAMUMN magnetometers. In addition, data from a few other magnetometers installed across the US, as depicted by the green pins in Fig 2.11, are also analyzed.
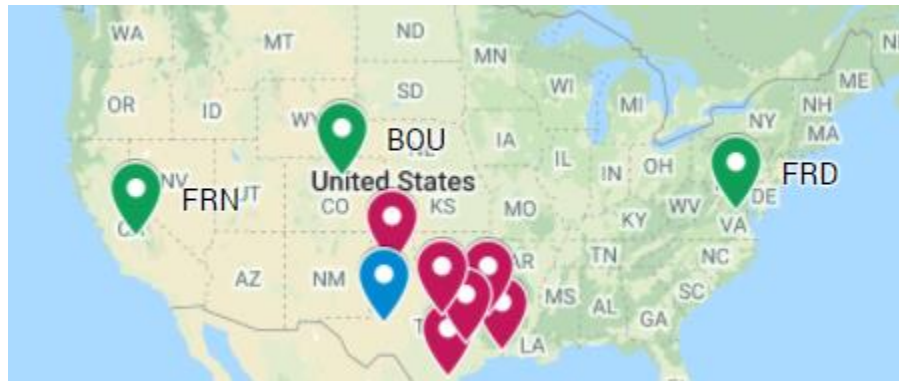
Figure 2.11 US Magnetometers used in correlation analysis along with TAMUMN. FRN: Fresno, BOU: Boulder, FRD: Fredericksburg [25].

These are USGS observatories namely, 1) Fresno (FRN), 2) Boulder (BOU), 3) Fredericksbrug (FRD) [34]. This is done to further test the validity and quality of the TAMUMN data on a wider scale, compared to existing magnetic observatory quality data, and to investigate local anomalies, if any. The analysis is done on the 10-second magnetic

field measurements over a 24-hr period using the Pearson's correlation coefficient, which is the covariance of the two variables divided by the product of their standard deviations. Consider X and Y as random variables which indicate the x-component of the magnetic field at two different stations. Since all the three components of magnetic field show similarity in data, the x-component is chosen as it represents the geomagnetic north. This is usually defined as the direction that the local field points in the horizontal plane, either on average or on a "quiet" day when there is short time scale (< 1 day) variation. The equation for the coefficient is given by,

$$\rho_{x,y} = \frac{\text{cov}(x,y)}{\sigma_x \sigma_y} \; \text{----[2.1]}$$

Here $\sigma_x \sigma_y$ pertains to the product of standard deviation of the entire data for both the stations selected. The covariance between the stations is divided by this value.

2.5.2. Procedure and Results

Magnetic field data for other stations BOU (Boulder), FRD (Fredericksburg), FRN(Fresno) is obtained in an encrypted format from the Intermagnet database [38]. Initial preprocessing is performed on this raw data to match its format to that obtained from TAMUMN servers (AMT, BMT, BVL, STP) for performing correlation analysis during all the GMD events. By generating the overall dataset for similar timepoints, a line plot, comprising all the stations is initially plotted. Upon obtaining the differences in the line plots for all the magnetometer readings, the magnetic field data is detrended for all values. A larger deviation between plots for different magnetometers is eliminated by

detrending successive values. A boxplot is then created, which specifies the outliers among the data. Correlation analysis is now performed by keeping and eliminating the outliers depicted in the boxplots. This shows that all the devices capture the outliers in the same trend and validates measurements across multiple magnetometers.

Fig 2.12 depicts the original distribution of the **B** field data from all the stations for a G2 event. Since this does not provide an accurate idea on the occurrences of peaks, the detrended graph as shown in Fig 2.13 explains all the variations within a conceivable vertical axis limit. For values exceeding beyond 1.5 times the interquartile range, points are eliminated from the overall dataset. The distribution of points is determined by the boxplot shown in Fig 2.14. Points exceeding the upper and lower limits are shown by black dots that do not fall between the horizontal bars of the graph.
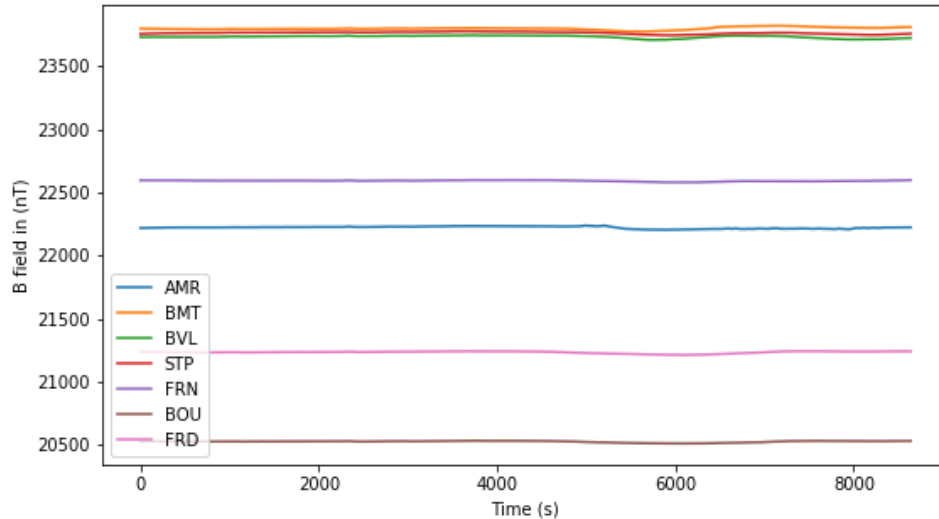
Figure 2.12 B field data from all the stations for a G2 event.
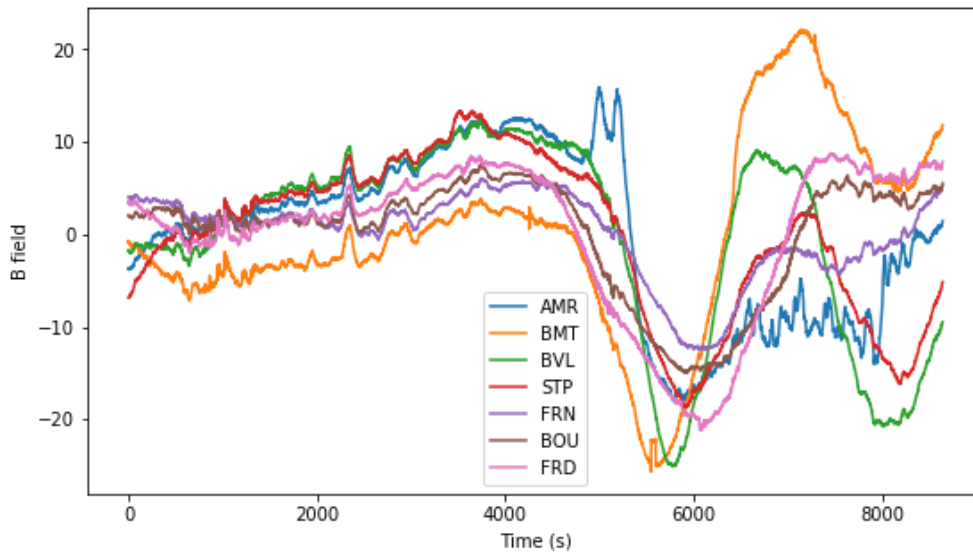
Figure 2.13 Detrended value of B field data from all the stations for a G2 event.
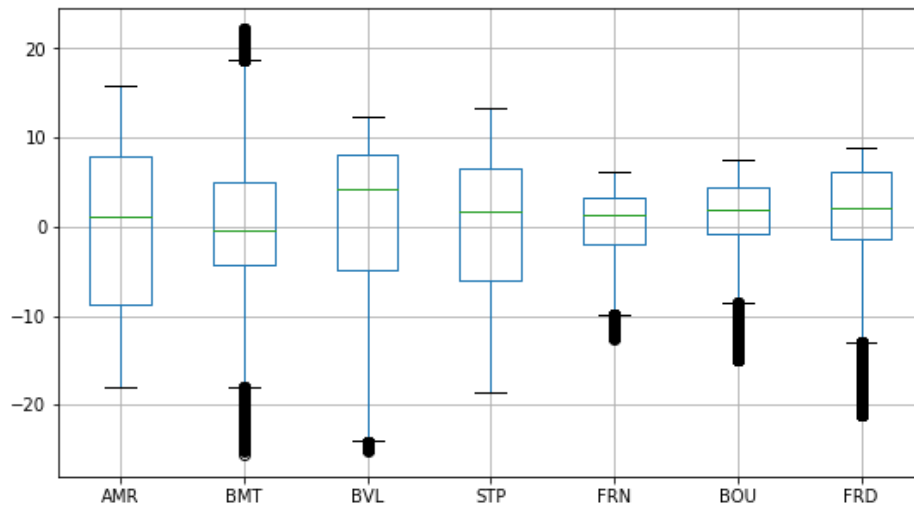
Figure 2.14 Boxplot depicting the presence of outliers beyond the interquartile range.

After all the specified modifications to the dataset, correlation is now performed using the equation provided in (2.1). A general observed fact is that the strength of correlation decreases with the level of GMD activity. The correlation matrix for a G2 level

event is shown in Fig 2.15 and a heatmap to visualize this metric is shown in Fig 2.16. It can be observed that a higher correlation exists between most of the local TAMUMN stations (> 0.6) on all days of activity and they are also well correlated with most of the other stations (>0.5), with some interesting exceptions. When two variables are correlated, the relative changes in their values appear to be linked. All correlations are positive, indicating that the change in values among all the stations lies in the same direction. In the presence of a geomagnetic disturbance of a higher magnitude, such as a G2 event, few deviations are observed in the correlations among the stations, in comparison to a G1 and quiet day. This hypothesis stems from the fact that GMD events typically cover a wide area, such as on the continental scale. The similar procedure is performed on the dataset for the G1 event and a heatmap depicting its correlation matrix is shown in Fig 2.17. Similarly, Fig 2.18 depicts the correlation among all stations during a quiet day.

	AMR	BMT	BVL	STP	FRN	BOU	FRD
AMR	1.000000	0.931792	0.825337	0.905980	0.740407	0.960940	0.652219
BMT	0.931792	1.000000	0.858138	0.893525	0.643500	0.881753	0.753016
BVL	0.825337	0.858138	1.000000	0.903284	0.804648	0.801337	0.482020
STP	0.905980	0.893525	0.903284	1.000000	0.836415	0.885743	0.518651
FRN	0.740407	0.643500	0.804648	0.836415	1.000000	0.785797	0.125555
BOU	0.960940	0.881753	0.801337	0.885743	0.785797	1.000000	0.593991
FRD	0.652219	0.753016	0.482020	0.518651	0.125555	0.593991	1.000000

Figure 2.15 Correlation matrix among all stations for a G2 event.
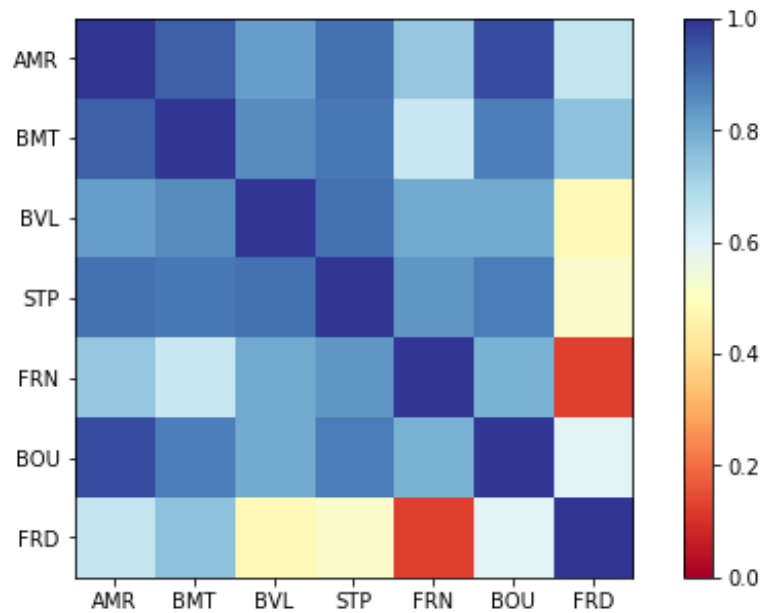
Figure 2.16 Correlation between TAMUMN and other magnetometers during a G2 event (09/28/2020).
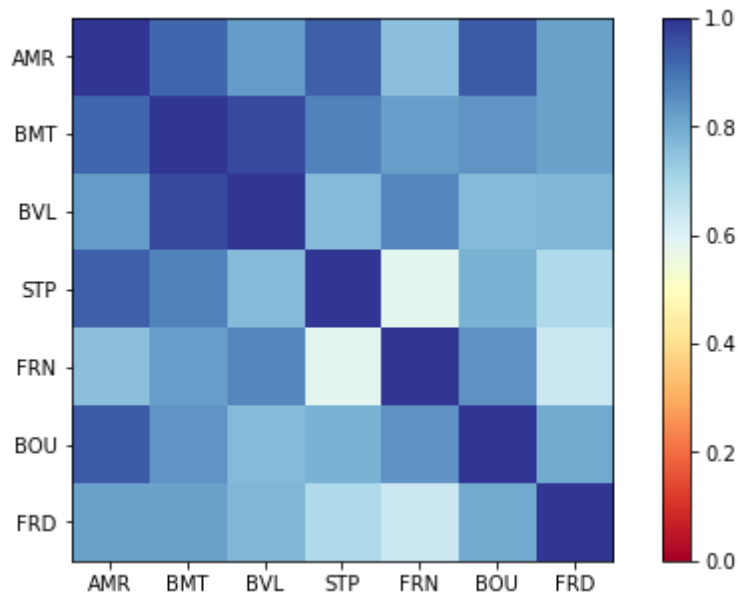
Figure 2.17 Correlation between TAMUMN and other magnetometers during a G1 event (10/23/2020).
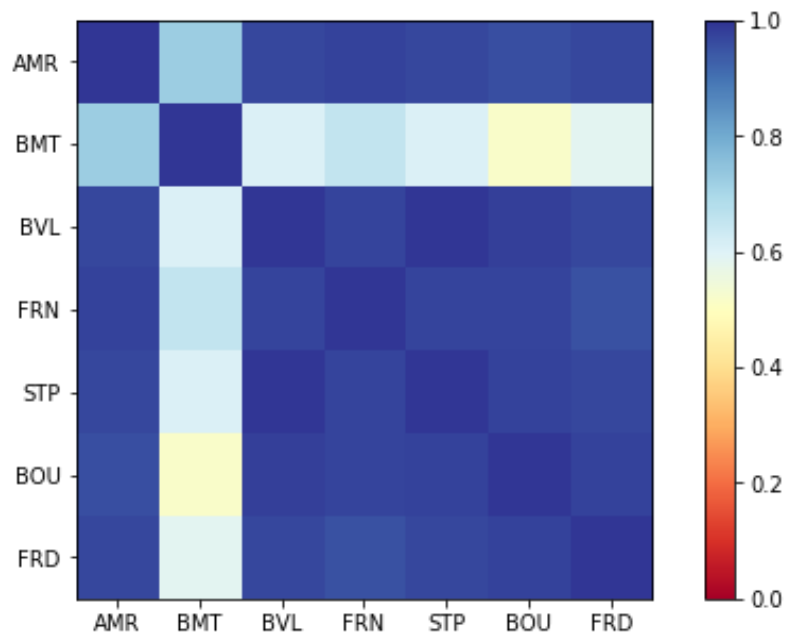
Figure 2.18 Correlation between TAMUMN and other magnetometers during a quiet day (10/09/2020).

From the above heatmaps indicating correlation among different stations, it can be observed that on a G1 day with relatively lower activity levels, all stations stay correlated with a higher magnitude. The four TAMUMN magnetometers, match each other with dark blue shades, indicating higher magnitudes of correlation, all of which are greater than 0.76. Further, the correlation with other stations too shows good results. The lowest value that can be inferred from the matrix forming the heatmap for a G2 event, shown in Fig 2.16 is 0.12, between FRD and FRN which explains the correlation of distantly located station at both the ends of the map. On comparing Fig 2.16 and 2.17, it is evident that the **B** field values are strongly correlated with a higher magnitude during G1 event than a G2 event with the lowermost limit of correlation being 0.5. Correlation between most of the stations is maximum on a quiet day, as observed in Fig 2.18 in comparison to other days. A quiet

27

day is characterized by no observable activity on the GMD scale, hence eliminating any regional variation. This gives an idea on the general dependance of correlation coefficient in the absence of an event. Further several other statistical techniques can be implemented as future works to handle exceptions of the case.

2.6. ANOVA

To compare different sets of magnetometer data obtained from all TAMUMN sensors, a statistical approach, known as ANOVA (analysis of variance) is used to analyze the differences among the means of different forms of data used for the measurement. The ANOVA test is performed on Minitab software [35]. Data from all TAMUMN sensors are compared for the three different scenarios of a G2, G1 events and a quiet day, to obtain conclusive results on validating the nature of the ANOVA model and the uniqueness of the data from each sensor. ANOVA test is generally performed to validate multiple measurements from different sensors for the same quantity. The results for a G2 event is shown in Fig 2.19 and similar observations are found for the other scenarios. Model summary and Tukey pairwise comparison is depicted in Fig 2.19 and Fig 2.20.

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Factor	3	16162130631	5387376877	21145848.47	0.000
Error	34556	8803912	255		
Total	34559	16170934543			

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
15.9616	99.95%	99.95%	99.95%

Means

Factor	N	Mean	StDev	95% CI
Amarillo_Bx	8640	22189.0	15.5	(22188.6, 22189.3)
Stephenville_Bx	8640	23726.9	16.3	(23726.6, 23727.3)
Beeville_Bx	8640	23797.5	17.2	(23797.2, 23797.9)
Beaumont_Bx	8640	23777.0	14.6	(23776.6, 23777.3)

Pooled StDev = 15.9616

Figure 2.19 ANOVA model summary.

Grouping Information Using the Tukey Method and 99% Confidence

Factor	N	Mean	Grouping			
Beeville_Bx	8640	23797.5	A			
Beaumont_Bx	8640	23777.0		B		
Stephenville_Bx	8640	23726.9			C	
Amarillo_Bx	8640	22189.0				D

Means that do not share a letter are significantly different.

Figure 2.20 Tukey pairwise comparison description.

S represents the standard deviation of the distance between data values and the fitted values. Hence a lower value of S suggests larger validity on the description of the response by the model. Comparing the S values of all three scenarios, it explains that the ANOVA model used for predicting works better with all the available data. The P value of the ANOVA test has a value lower than the significance level. This is used to provide a conclusion on the nature of all the different sensor variations. The value demonstrates

that all data are significantly different. Technically, this highlights that that the magnetic field data obtained from sensors differ and do not qualify for null hypothesis. The null hypothesis in ANOVA explains that there is no difference in means. The research or alternative hypothesis is always that the means are not all equal. Fig 2.21 shows the Tukey pairwise comparison plot including all data points of the TAMUMN sensors.
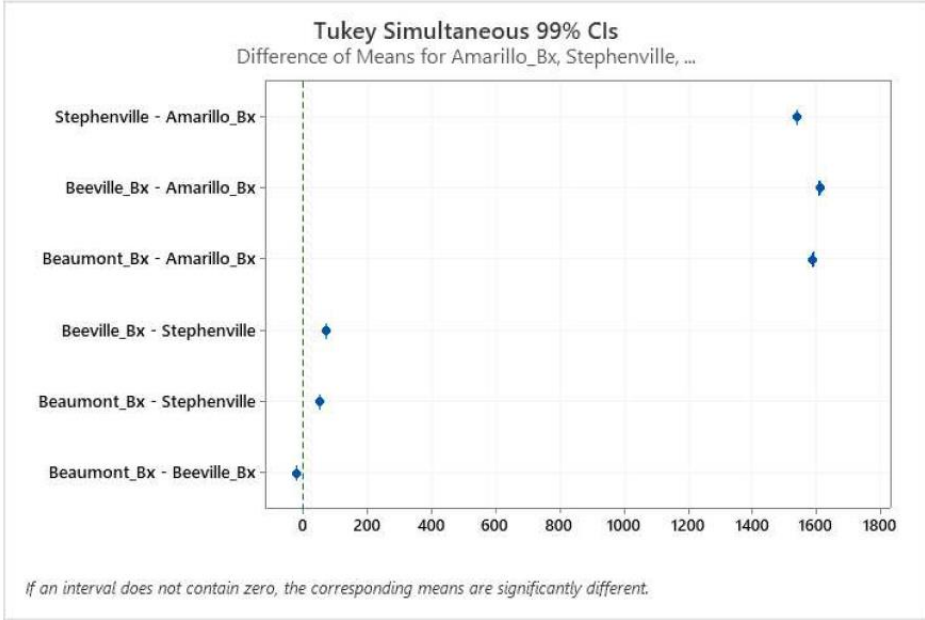
Figure 2.21 Tukey pairwise comparison plot.

While the results of a one-way ANOVA will indicate if there is a main effect of the explanatory variable, the initial results will not provide information on which groups are different from one another. In order to determine which groups are different from one another, a post-hoc test is needed. Post-hoc tests are conducted after a one-way ANOVA to determine which groups differ from each another. There are many different post-hoc analyses that could be performed following a one-way ANOVA. If the ANOVA leads to an evidence that the group means differ, it leads to the direction on investigating which of

the means are different. This is where the Tukey multiple comparison test is used. The test compares the differences between each pair of means with appropriate adjustment for the multiple testing. The results are presented as a matrix showing the result for each pair, either as a P-value or as a confidence interval. The Tukey multiple comparison test, like both the t-test and ANOVA, assumes that the data from the different groups come from populations where the observations have a normal distribution, and the standard deviation is the same for each group.

With respect to ANOVA, it is necessary that the pairwise comparison intersect the 0 line to show that they qualify for the null hypothesis (or rather their means are similar), else the hypothesis is discarded, and eventually, all the data is categorized into different groups. This confirms the reason for the data from each of the TAMUMN magnetometers to be categorized into four different groups by the Tukey comparison.

3. ELECTRIC FIELD DATA

The focus of this chapter is on extracting information from the electric field data and establishing its relation to the sensor generated magnetic field values. Background knowledge on ground models is provided here. Visualizations developed are illustrated to indicate the differences among multiple ground models and magnetometers.

3.1. Ground Models

Geoelectric field induction has a strong dependence on deep Earth geology. The accuracy of Earth conductivity models is an essential factor that determines how well geomagnetically induced currents are estimated for hazard mitigation in planning and operations [15]. Since the electric field obtained is location specific, the Earth's response in each region must be described.

The three commonly used descriptions of Earth's response, increasing in complexity and local accuracy may be thought of in a three "layer" paradigm, as follows [16]:

   • Conductivity (beta) scaling factors

   • Regional 1D conductivity models

   • Use of empirical 2x2 impedance tensors of horizontal geoelectric field response to magnetic field fluctuations also referred to as "three-dimensional" (3D) tensors or EMTFs.

Recent research focuses on using updated 1D models which are based on regional averages of local empirical transfer functions EMTF rather than the conductivity versus depth format that is used when derived from geologic models [16]. To investigate the electric field induced in a region due to changes in magnetic field, the ground response of the specific region is taken into consideration.

Several methods exist to describe Earth's response (conductivity) to the electric field. Previous works of authors focuses on using regional 1D and 3D conductivity versus depth format derived from traditional geologic Earth models [36]. In these approaches, resistivity values are obtained at varying depths according to the frequency of operation, which is in turn used to calculate the resulting electric field in each region. In this method, for each frequency of operation, the surface impedance which depends on the Earth's conductivity, is recursively calculated for each layer. The final set of surface impedance values represent the transfer function of Earth relating the electric and magnetic fields.

Updated 1D regional boundaries of different states based on regional averages of local empirical transfer functions documented by EPRI [17] is represented in Fig 3.1. Region 17 covers a large portion of Texas, hence ground parameters of this area is taken for calculation purposes. In the 3D model, ground response is represented by 2X2 impedance tensors. This is empirically based and calculated after repeated measurements of electric field responses to the magnetic field driver and is used to create a set of Electromagnetic Transfer Functions (EMTFs). The goal of this work is to calculate the geomagnetically induced current (GIC) values in Texas footprint using the calculated electric field on the synthetic electric grid model.
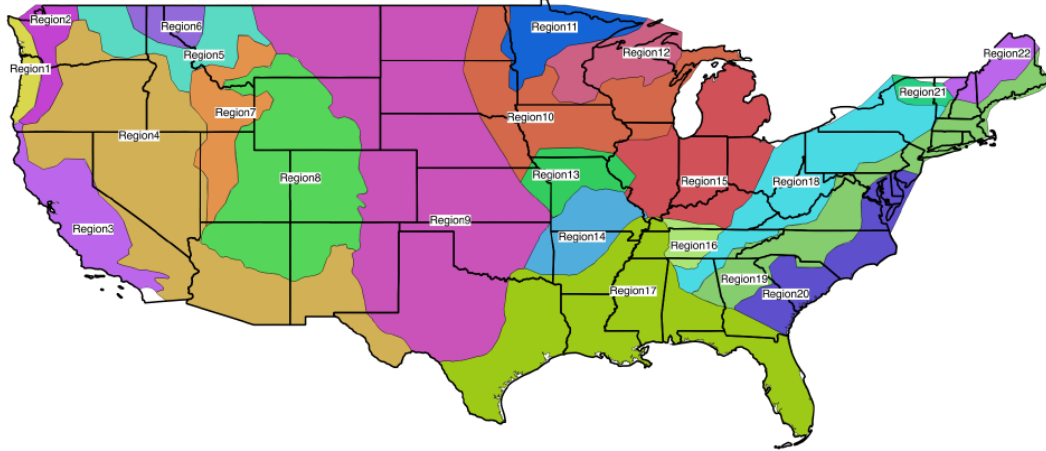
Figure 3.1 Updated 1D regional boundaries of different states documented by EPRI [15].

3.2. Electric Field Visualization

3.2.1. Geoplot

To visualize the spatial distribution of electric field, at a given instance of time, the data for the overall electric field in V/km for each point in the state of Texas, is obtained from the extended TAMUMN server [22]. The data, which is available in JSON format, specifies electric (**E**) field in the east and north directions for all geographical coordinates. A geoplot is developed using this data which depicts the spread of **E** field values, taking the ground model distribution format. Geoplot is a high-level Python geospatial plotting library, which takes a shape file of the region on which the data must be visualized and develops the plot according to user specifications. In this case, since the **E** field contour across the TX area is required to be visualized, a shape file (.shp) of the TX region is downloaded and extracted to a Python script which is developed to include all aspects of visualization. Here the **E** field data is provided, and the script develops a scatter point

34

contour across the entire region, depending on the magnitude intensity. The plot is shown in Fig 3.2, where the vertical and horizontal axis represents the information for the longitude and latitude, respectively. The geoplot in Fig 3.2 is based on the observations taken on a normal day, hence the magnitudes of **E** field do not show any interesting observation however the ranges of the values in each region, takes shape of the ground conductivity models as depicted in Fig 3.1. Since it is a quiet day, the magnetic field data does not demonstrate any pattern which varies in magnitude or location, this provides a relatively constant range of **E** field values over the entire region. The electric field data now calculated is solely influenced by the ground response; hence validating the authenticity of TAMUMN data to the region's ground response.
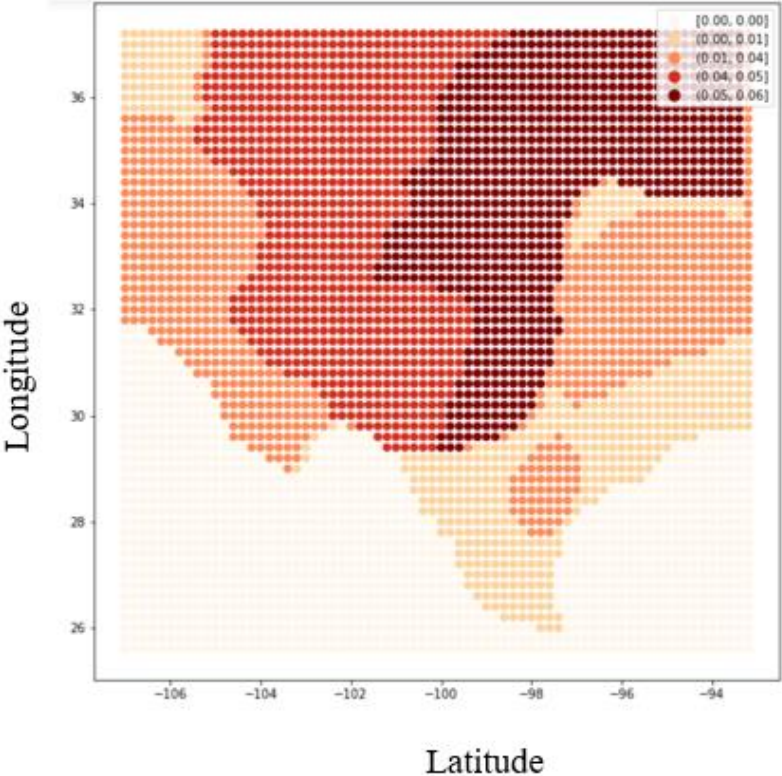
Figure 3.2 Geoplot on Texas region provides distribution of E field intensity.

3.2.2. Electric field (E Field) Visualizations of TAMUMN for Different Events

The three events of G2, G1 and a quiet day is considered, and the respective **E** field data obtained across all the TAMUMN devices is plotted on a single graph. Fig 3.3, 3.4 and 3.5 displays the variation of **E** field throughout each day for different events. The scale along the x axis shows the variation of time within a day in the 24-hour format of HH: MM: SS. To avoid crowding of all points on the horizontal axis, every 1000 points is chosen for display, hence showing few time points.

Along the vertical axis, is the **E** field obtained by using the EPRI ground model on the **B** field data for different days of GMD activity. The horizontal axis describes the hour-based duration of the day. The **E** field data directly influences the existence of GIC currents across the region for the specific intervals. All the three figures have similar scales of axis for comparison purposes. This provides information on the occurrences of peaks during G2 events, at a greater frequency relative to a G1 event and a quiet day. The **E** field data is equivalent to the differential of the **B** field data, along with the ground models taken into consideration. Fig 3.3 exhibits large overshoots, justifying a G2 event with sudden surges which are responsible to cause adverse effects, including increase in current flow in the power grid. Fig 3.4 and 3.5 shows a similarity in the **E** field of Beaumont and Beeville, following a common trend and likewise, **E** field of Amarillo and Stephenville, exhibiting a common pattern. This observation can be attributed to the nature of ground models shared by the TAMUMN magnetometers.
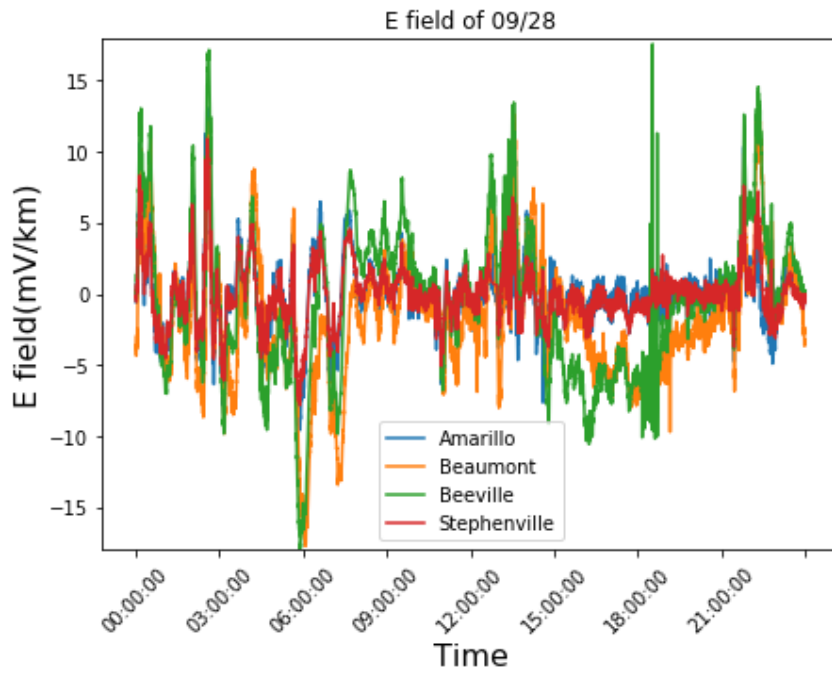
Figure 3.3 E field across TAMUMN devices during a G2 event.
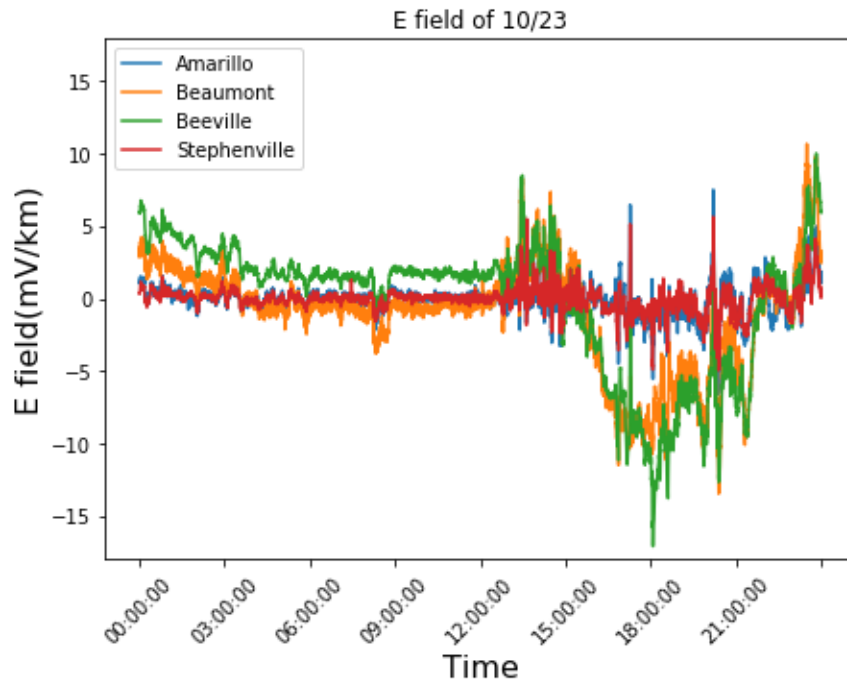
Figure 3.4 E field across TAMUMN devices during a G1 event.
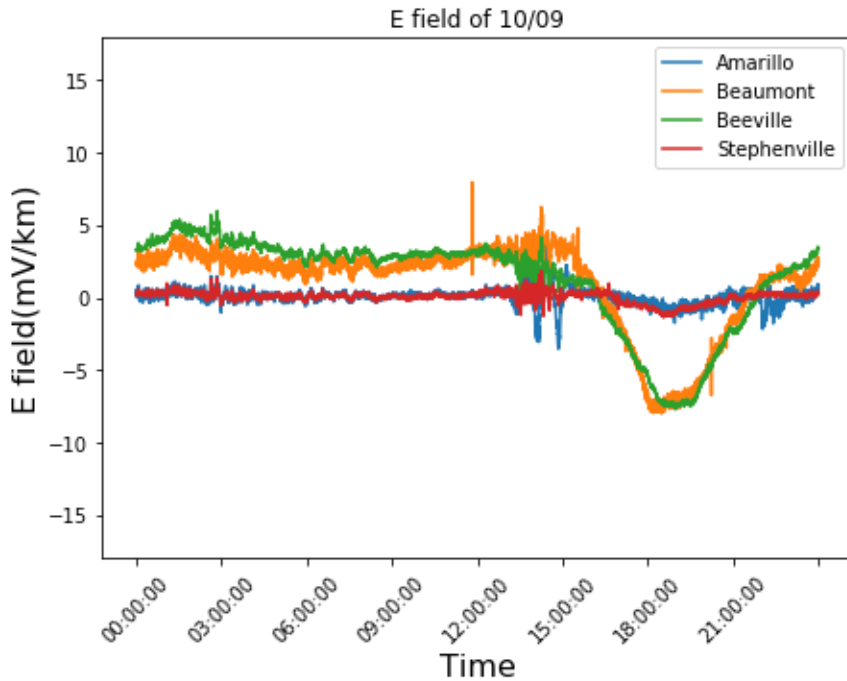
Figure 3.5 E field across TAMUMN during a quiet day.

3.2.3. Comparing USGS and EPRI Ground Models.

Since geoelectric field induction is strongly dependent on deep Earth geology, any estimation on the consequences of GIC must characterize the ground response in some way. Previously used USGS models employ a conductivity vs depth format to calculate the ground response of a region [16]. The EPRI model employed in the work done here, derives a frequency dependent surface impedance from the above ground model format. These parameters are also known as Regional Transfer Functions (RTF) are based on regional averages of local empirical EMTF. This provides an estimate of Z(w), which is the Earth's surface impedance, responsible for calculating the ground response in the 1D model.

38

Summarizing the comparisons, both the USGS and EPRI models are 1D ground conductivity models, however the EPRI model of ground response uses the updated regional averaged transfer functions of ground response for contiguous US.

As described previously, the trends captured by Amarillo and Stephenville matches on all events, validating the similarity in ground models. Fig 3.6, Fig 3.7, and Fig 3.8 depicts the commonalities in the behavior of both the USGS and EPRI models.
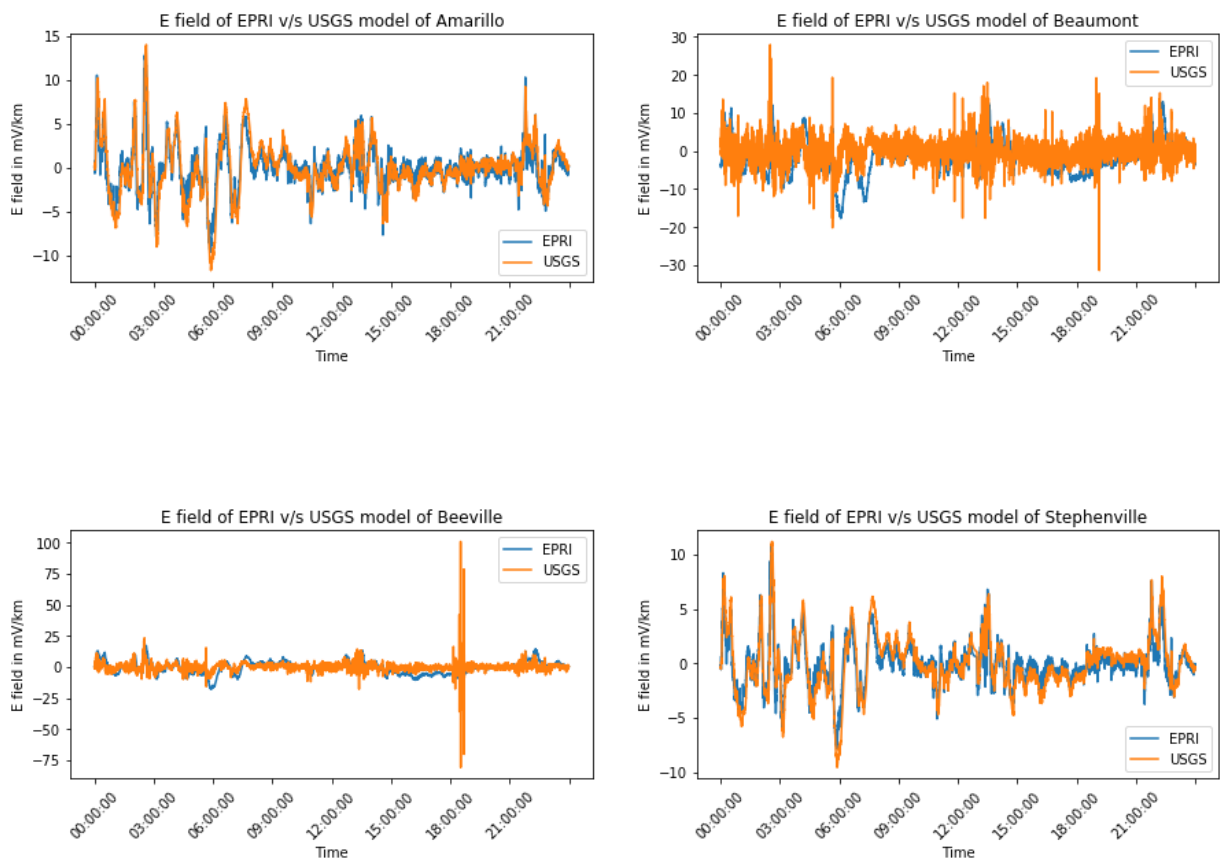
Figure 3.6 Comparison of E field between EPRI and USGS models for a G2 event (09/28).
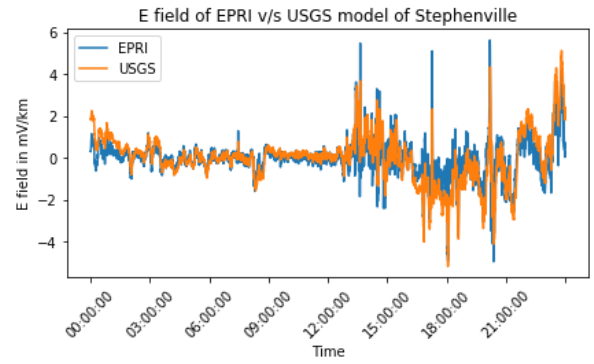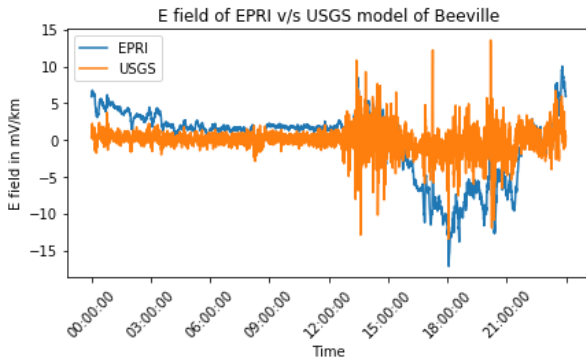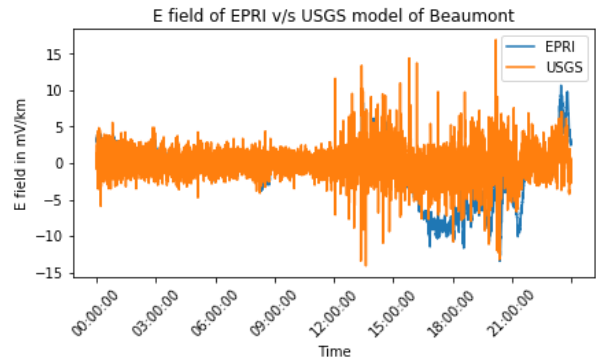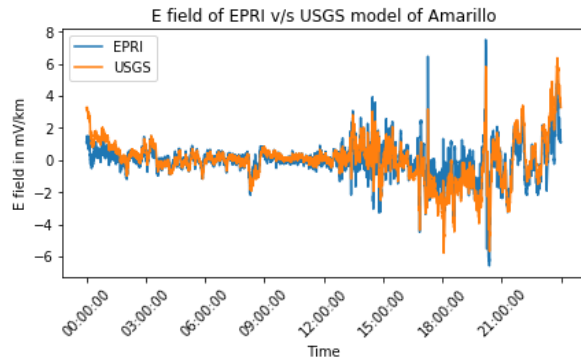
Figure 3.7 Comparison of E field between EPRI and USGS models for a G1 event (10/23).
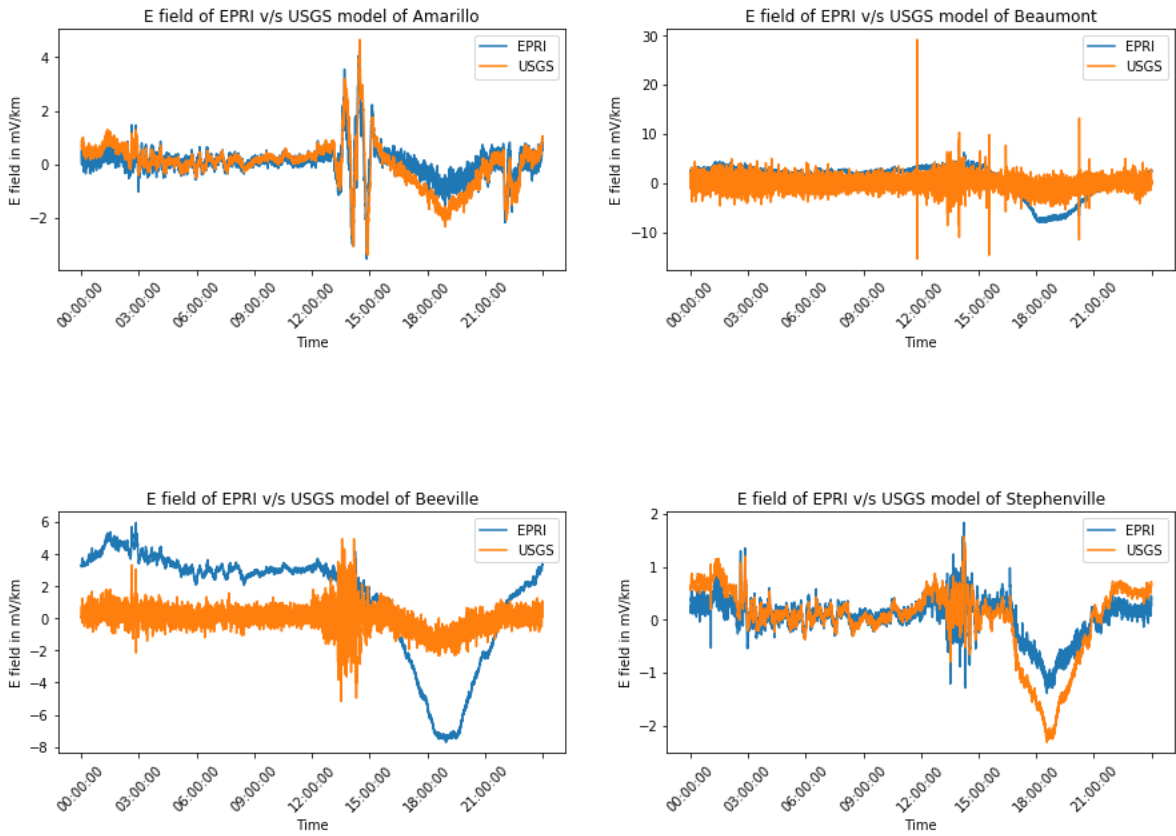
Figure 3.8 Comparison of E field between EPRI and USGS models on a quiet day (10/09).

Fig 3.8 shown above displays the comparison of the E field obtained on using both the models for a non-GMD event. This is characterized by relatively flat lines and elimination of sudden peaks. The trend is very similar to that of a G1 event as shown in Fig 3.7. Fig 3.8 depicting a GMD event shows a completely different trend with the inclusion of frequent variations from the steady state.

41

4. GRID MODEL BUILDING

This chapter focuses on augmenting the synthetic electric grid version built on Texas footprint, consisting of 7000 buses. With the modifications done to the network, it serves as a testbed for reconstructing the geomagnetic storm scenarios using the data, described in the previous sections. The procedures implemented to replicate the simulation of the scenarios is briefly described here.

4.1. Addition of Substation Grounding Resistances

To perform GIC studies on the TX 7000 bus case, the grounding resistances of several substations must be defined as they influence the value of GIC currents that flow across the transmission lines in the system. A probabilistic approach is taken for specifying the resistances of different substations. From [37], the percentage of spread of resistances across the entire power system can be justified, as depicted in Table 4.1.

Table 4.1 Substation grounding resistances, statistics, and validation range [37].

Substation grounding resistance, Ω	Percent (%) of substations, actual case	Percent (%) of substations, validation range
0.01 – 0.05	5.7	0 - 10
0.06 – 0.10	13.8	5 – 20
0.11 – 0.20	22.8	10 – 30
0.21 – 0.50	29.5	15 – 40
0.51 – 1.00	15.5	10 – 25
1.01 – 5.00	12.3	5 – 20
5.01 – 20.00	0.4	0 - 5

Using the values of substation grounding resistance as baseline, a Python script is developed to generate resistances within the range as provided in column 1 of Table 4.1, for the fraction of substations with respect to the overall count of the total substations. The addition of the substation grounding resistances is done to get a full range distribution of resistances across the entire region. This distribution is validated by creating a contour plot, according to the substation resistance magnitude throughout the region which is displayed in Fig 4.1. It is evident that a dominant portion of substation grounding resistances fall within the magnitude of (0-0.21) Ω, which is specified in Table 4.1. There is a sparse distribution among the other values. The legend for the map is accordingly developed and the plot shows all the variations. Fig 4.1 generates the impression that the density of distribution is not focused on any specific region and is randomly spread across the overall electric grid model, avoiding any biased results on further cascading calculations.
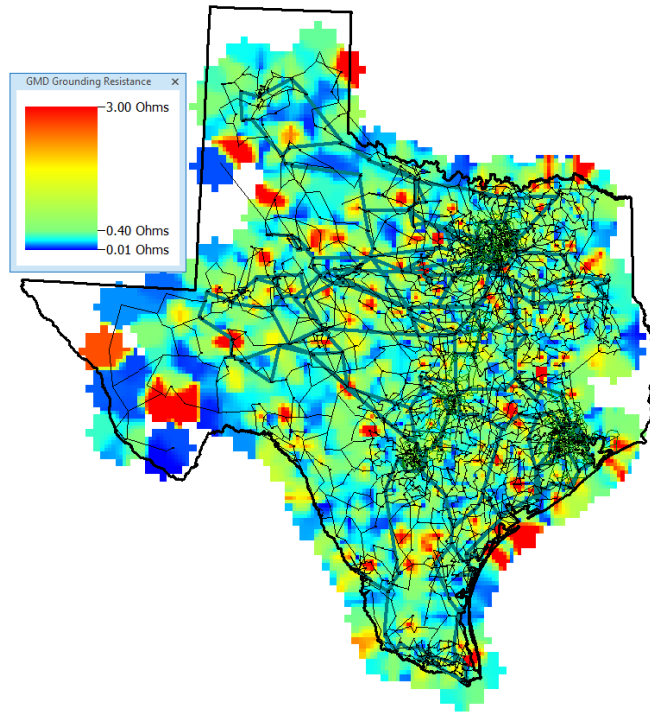
Figure 4.1 Distribution of Substation Grounding Resistances.

4.2. Addition of Transformer Winding Configurations and Presence of Autotransformer

The base case has entries for substation grounding resistances and several essential parameters, such as specifications for generators, buses and switched shunts which lay the foundation for the case. To perform GIC studies, it is necessary to include the winding configuration for all transformers. There is also the possibility for the existence of three winding transformers, which require explicit identification in the case. Further, autotransformer models are also added to the case. A statistical approach is taken to decide on the configuration of the transformer windings, some of which is derived from [37]. The most dominantly existing configuration for transformers in a power system is the wye-

44

delta orientation with wye on the high voltage and delta on the low voltage side respectively. About 75% of the non-autotransformers are assumed to have the wye-delta configuration. Nearly 16% of the non-autotransformers have delta-wye configuration and the rest 9% are designed to have the wye-wye design. The delta-delta design is generally not preferred due to the absence of neutral point in the case of a 3 phase, 4 wire system.

To design the considerations for classifying a given transformer as an autotransformer, the secondary voltage of the transformer is considered. A simple statistical approach is followed, according to which 75% of the total number of transformers whose secondary voltage is greater than 100kV are designed to be autotransformers and 25% of the total transformers with secondary voltage lesser than 100kV are autotransformers. However, all the autotransformers are designed with a wye-wye winding.

The above conditions are coded in Python to generate CSV files on transformer specifications, which are added to the case for visualizing in PowerWorld software. Table 4.2 provides the statistics on the distribution of autotransformers in the case.

Total number of transformers in the case: 1677

Total number of transformers where secondary voltage > 100kV = 313

Total number of transformers where secondary voltage < 100kV = 1362

Table 4.2 Statistics on autotransformers

	Auto Transformers	**Non-Auto Transformers**
Secondary voltage > 100kV	234	79
Secondary voltage < 100kV	340	1022

4.3. Series Voltage across Lines

The electric field data for three events (G2, G1 and a quiet day) occurring on three different days (09/28/2020, 10/23/2020, 10/09/2020) that are obtained using the magnetic field data of four TAMUMN magnetometers are primarily utilized for the analysis. This data is recorded in the resolution of 10 seconds. The synthetic electric grid of Texas (**TX 7k**) with a network of 7000 buses is considered for replicating all the three scenarios. GIC analysis is performed in the PowerWorld software [38]. The voltage drop at each branch, corresponding to the induced electric field for all timepoints is provided as an input to the software, for performing the GIC current calculations on all the transmission lines. The equation specified in (4.1) is used for calculating the voltage drop on each line.

$$V = E_x L_x + E_y L_y \ \text{----(4.1)}$$

Here E_x and E_y refer to the induced electric field along the east and north directions, respectively. Based on the location of the branches, determined by their geographical coordinates, the data of the most closely located TAMUMN sensor for each branch is provided as the **E** field input. L_x and L_y is the length of the branch along the similar directions [23], the procedure for the calculations of which is described below.

Considering a transmission line to span between substations A and B, Ly is the distance along the NS direction and Lx is the length of the transmission line along the East-West direction. Assuming a spherical Earth, the NS distance (Ly) is calculated from the difference in the latitude of the substations A and B. For calculating the EW distance (Lx), no direct relation exists since the lines of longitudes converge as they approach the pole. Hence it is necessary to consider the latitude of the substations while converting their

longitudinal separation into distance. $\Delta \, lat$ is the difference in latitude (degrees) between the substations A and B, as given in (4.2) ϕ is defined in (4.3) as the average of two latitudes. Similarly, $\Delta \, long$ is the difference in longitudes (degrees) between the substations Using all the above parameters, the NS and EW distance can be empirically calculated according to (4.5) and (4.6).

$$\Delta \, lat = Lat \, A - Lat \, B \text{-----}(4.2)$$

$$\phi = \frac{Lat \, A + Lat \, B}{2} \text{-----}(4.3)$$

$$\Delta \, long = Long \, A - Long \, B \text{-----}(4.4)$$

$$Ly = (111.133 - 0.56 \cos(2\phi)) \, \Delta \, lat \text{----}(4.5)$$

$$Lx = (111.5065 - 0.1872 \, \cos(2\phi) \,) \, \cos\phi \, \Delta long \ \text{-----}(4.6)$$

For each branch in the **TX 7k** case, V is calculated using (4.1), with E field parameters obtained from closely located TAMUMN magnetometers, Lx and Ly as calculated from (4.5) and (4.6). This process of identifying the closely located TAMUMN device and performing the series voltage calculations for each branch is automated by developing a script in Python. On providing the series voltage as inputs to each branch, the GIC tool of PowerWorld software is run to calculate the geomagnetically induced currents at each branch. This is repeated for all scenarios and the summary of GIC currents along with their visualizations is shown below.

Table 4.3 provides a summary of the GIC current ranges obtained on implementing the analysis using the data. The substation GIC currents are plotted in a Geographic Data View (GDV) format, which displays the magnitude of GIC currents across various

substations using a color contour, where the size of the substation scales according to the magnitude of the current. GDV plot is developed for all the three scenarios, using a common scale of reference. This helps to compare the magnitude of currents during events and determine the largely impacted regions in the electric grid model. There is a sharp difference in the magnitudes of the maximum transformer currents during the G2, G1 events and a normal day. The substation currents similarly vary for both G2 and G1 events, which is at a different range in comparison to a normal day. However, in both G1 and G2 events, the substations which experience the peaks of current during each event are not necessarily the same but belong to a common region in the electric grid model.

Table 4.3 Summarizing GIC currents for different GMD scenarios.

	Transformer Neutral Current (A) [min: max]	Substation GIC Amps to Neutral Current (A) [min: max]
G2 event (09/28/2020)	[-1.05: 0.95]	[-1.25: 1.63]
G1 event (10/23/2020)	[-0.421: 0.324]	[-0.77: 0.5]
Normal day (10/09/2020)	[-0.247: 0.253]	[-0.28: 0.36]

The details of the vulnerable substations can be taken into consideration for planning purposes to develop mitigation strategies at each location such as installation of GIC blocking devices and designing transformer protection. Fig 4.2, Fig 4.3 and Fig 4.4 depict the GDV plot for a G2 event, G1 event and a normal day respectively. The legend in the plot relates the scaling ratio specified for the input current. For instance a GIC current of 0.4 A is scaled to 120 times its actual size in the contour.

Input Value	Characteristic Value
0.00	0
0.40	120.
0.70	250.
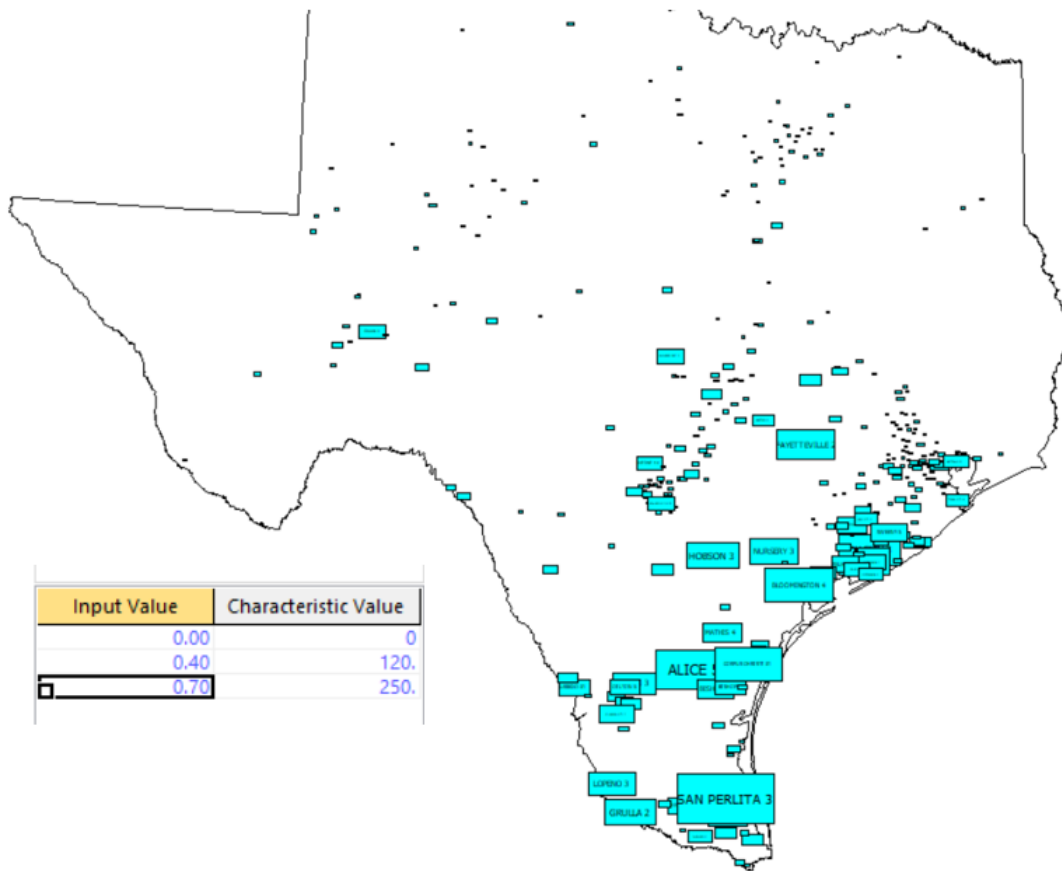
Figure 4.2 GDV plot of the TX 7k region for a G2 event, each box indicates a substation, and its size corresponds to the magnitude of GIC current.

From the above figure, few substations such as San Perlita 3, Alice 5, Bloomington, and Corpus Christi are dominantly visible, explaining the existence of a higher value of GIC current at these regions during the G2 event. Several other substations share similar higher value, which is justified by scaled current value indicated by the rectangular boxes.

Input Value	Characteristic Value
0.00	0
0.40	120.
0.70	250.
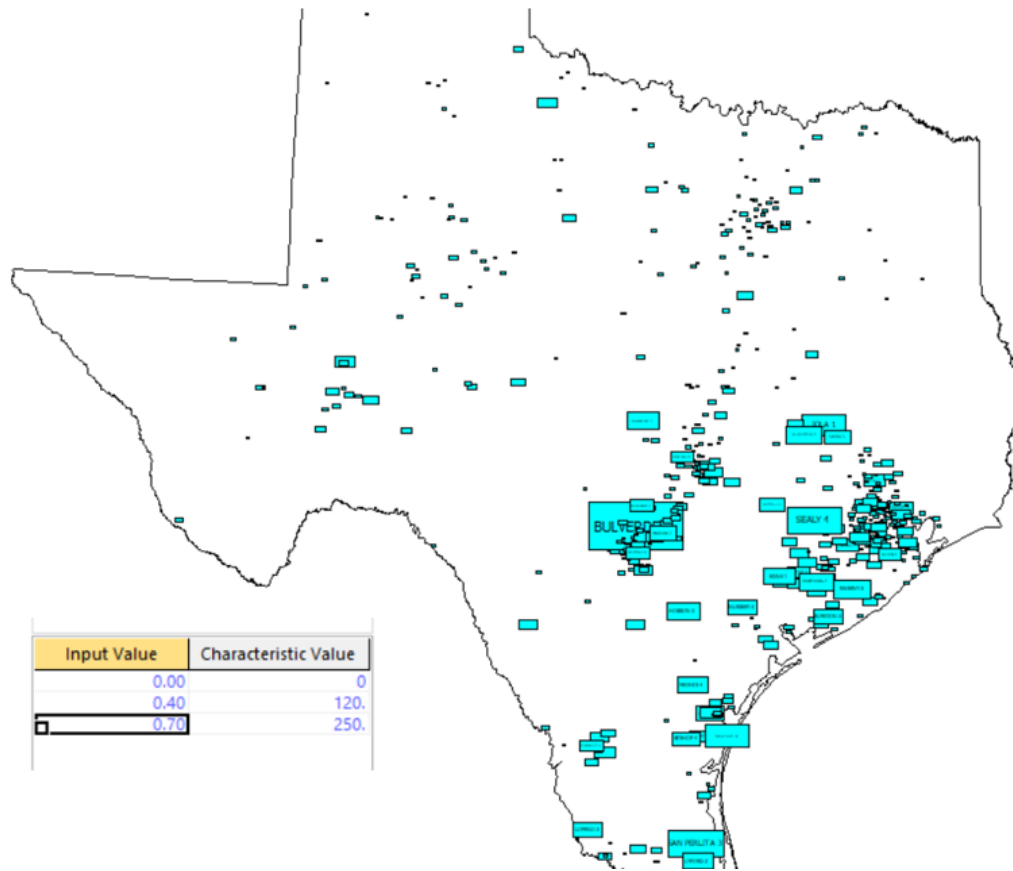
Figure 4.3 GDV plot of the TX 7k region for a G1 event, each box indicates a substation, and its size corresponds to the magnitude of GIC current.

Input Value	Characteristic Value
0.00	0
0.40	120.
0.70	250.
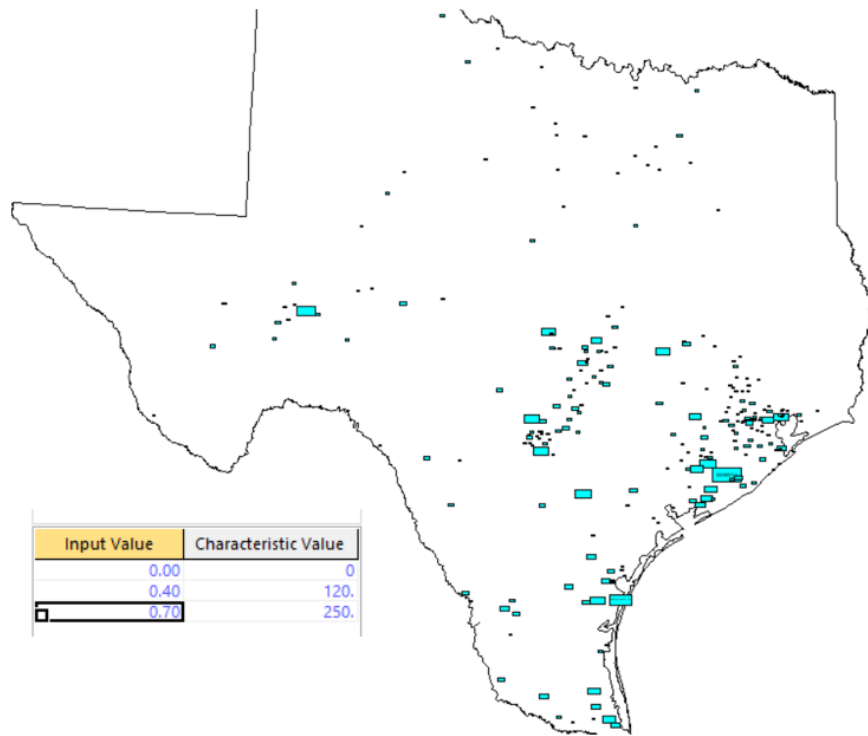
Figure 4.4 GDV plot of the TX 7k region for a quiet day, each box indicates a substation, and its size corresponds to the magnitude of GIC current.

From Fig 4.3, it is observable that few substations like Bulverde 6 and San Perlita 3 have higher values of GIC current, indicated by the size of the rectangular boxes, however there are not many substations with higher values of current in comparison to a G2 event. The number of affected substations (with a higher value of GIC currents) remain to be effectively fewer in G1 than a G2 event.

In contrast, Fig 4.4, depicting a normal day has a lower range of limits for the current value. Hence with the similar scale for all three figures, the only effective detail obtained from Fig 4.4 is that the behavior of the system on a quiet day is drastically different than that to a day with considerable GMD activity. The reasons for higher current

at a specific substation in comparison to others, can be attributed to several factors such as storm magnitude and direction, geographical location, or even the configurations of the components. Since, the cumulative factors cannot be decided based on such overall assumptions, this process of event reconstruction helps to identify all such regions involved.

5. SENSITIVITY ANALYSIS

This chapter focuses on performing sensitivity analysis on the most affected transformers resulting out of the reconstruction of GMD scenarios. The outcome is to generate the extent of current flow on the transmission lines to the response of induced voltages.

5.1. Identifying the Vulnerable Transformers

Sensitivity analysis is performed to determine the effect of transformer currents on neighboring and closely affected branches [39][40]. The transformers with the largest values of effective GIC currents, during both G2 and G1 events are identified as severely affected devices. The primary step in the analysis is to identify the transformers with predominantly higher currents during both the scenarios. Table 5.1 and 5.2 provides the details on the transformer per phase effective GIC Amps during G2 and G1 events, respectively.

Table 5.1 Per phase GIC effective currents of transformers on G2 event

Bus Name High	Bus Name Med	Transformer Per Phase Effective GIC, Amps
ALICE 5 1	ALICE 5 2	0.235
SAN PERLITA 3 1	SAN PERLITA 3 3	0.204
BLOOMINGTON 4 1	BLOOMINGTON 4 2	0.177
BRUNI 3 1	BRUNI 3 3	0.154
BAY CITY 3 1	BAY CITY 3 2	0.14
BAY CITY 3 1	BAY CITY 3 2	0.14
BRUNI 2 1	BRUNI 2 3	0.119

Table 5.2 Per phase GIC effective currents of transformers on G1 event

Bus Name High	Bus Name Med	Transformer Per Phase Effective GIC, Amps
SAN PERLITA 3 1	SAN PERLITA 3 3	0.14

BULVERDE 6 1	BULVERDE 6 2	0.129
BRUNI 3 1	BRUNI 3 3	0.119

On evaluating the entries in Table 5.1 and 5.2, the transformers which are inflicted with a larger current in both the scenarios are considered as most vulnerable. Further, the entries common to both the tables are selected for being more prone to overall disturbances. Table 5.3 mentions all transformers selected out of the scenarios for the analysis. A label is used for each of the transformer to ease the detailing in the upcoming calculations. Fig 5.1 displays the geographical location for each of the transformer listed in Table 5.3. AT represents an autotransformer, 3WT stands for a three-winding transformer.

Table 5.3 Description of transformers used for analysis.

TRANSFORMER LABEL	FROM BUS	TO BUS	VOLTAGE LEVEL (kV)
Transformer A	SAN PERLITA 3 1	SAN PERLITA 3 3	345/22
Transformer B (AT)	BULVERDE 6 2	BULVERDE 6 1	138/345
Transformer C (AT)	ALICE 5 1	ALICE 5 2	345/138
Transformer D	BRUNI 3 1	BRUNI 3 3	345/13.8
Transformer E (3WT)	BLOOMINGTON 4 1	BLOOMINGTON 4 2 BLOOMINGTON 4 3	138/13.8
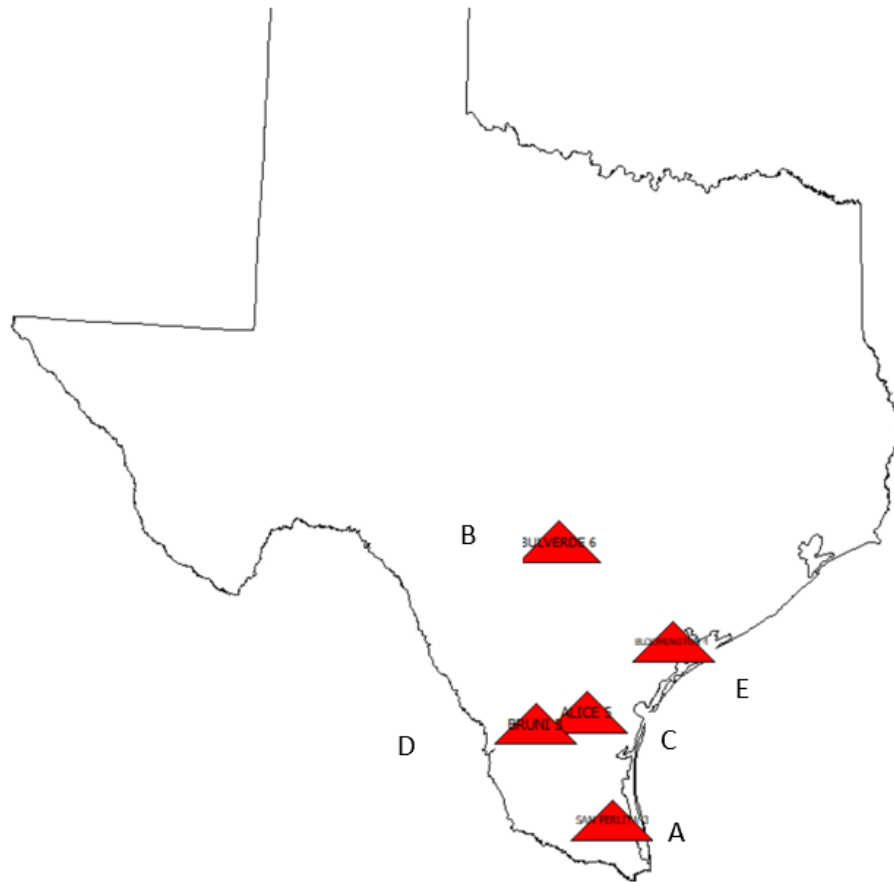
Figure 5.1 Location of transformers used for sensitivity analysis.

5.2. Sequence for Sensitivity Analysis

For each of the transformer mentioned in Table 5.3, the analysis is sequentially performed where a 1V/km electric field is provided in the specified direction to the system, resulting in the creation of geomagnetically induced voltage. This causes the flow of currents in the transformer windings. Sensitivity Analysis performed, helps to determine the extent of the contribution (impact) on each of the lines for the given flow of effective current in the transformers. The sensitivity metric, dIeffective/ dVoltage, is defined as the ratio of the change in the transformer per phase effective current to the induced GMD

voltage. This metric has a value between -1:1 and it explains the amount of transformer effective GIC current that flows in a particular branch for a 1V assumed input of electric field due to GMD event. The sign indicates the direction of the current flow into the branches, while considering a standard from and to bus notation. Overall, the absolute magnitude is taken for analysis.

Generally, for GIC analysis purposes, the convention commonly used is to consider the per phase current for transformers and transmission lines and the total three phase current for the substation neutral current. $I_{Effective,r}$ for a transformer, is the effective per phase current value that depends on the type of the transformer. For a grounded wye-delta transformer, it is the current in the grounded winding, which is generally the higher voltage end, whereas for an autotransformer the effective current depends on the sum of the currents going into both the coils, taking the transformer's winding ratio into account. This value serves as the metric for classifying the most vulnerable transformers among the scenarios.

$I_{max,r}$ for a given transformer represents the absolute worst-case maximum $I_{Effective,r}$ value for the given induced field condition. The standard chosen is a 1V/km electric field input and for different orientations of the electric field, the maximum current experienced by the transformer provides the above value. This quantifies, the maximum possible damage to the system for a specified condition. The number of lines contributing to certain percentages of the maximum current, helps to investigate the influence of the transmission lines and the extent of the spread of GICs across the region, by developing a

contour that assesses the distances of the regions experiencing the effect of the transformer currents.

The value of $I_{max,r}$ depends on the assumed direction of the field and its value is obtained on identifying the orientation of the field that produces the maximum value of effective current. However, the values of GIC differs significantly based on the location of the transformer. GICs are lower for transformer located in the interior of the network compared to those located on the edge since the former experience cancellation of GIC currents flowing into one side of the transformer's substation. A metric is used to evaluate the relation of $I_{Effective,r}$ and $I_{max,r}$, independent of the magnitude of the electric field, Λ_r which is defined in (5.1)

$$\Lambda_r = \frac{I_{Effective,r}}{I_{max,r}} \text{ --- (5.1)}$$

Higher values of the above metric indicates that the transformers are closer to the network edge. Table 5.4 summarizes the results of the sensitivity analysis performed on all transformers listed in Table 5.3.

Table 5.4 Sensitivity analysis results on transformers specified in Table 5.3.

Transformer Name	Transformer A	Transformer B	Transformer C	Transformer D	Transformer E
$I_{max,r}$ (A)	5.549 ($\theta = 60$)	4.978 ($\theta = 90$)	20.966 ($\theta = 84$)	7.515 ($\theta = 90$)	5.861 ($\theta = 90$)
Highest % contribution to $I_{max,r}$ by a line	39.6	23.9	5.8	38.5	20.6
No. of lines contributing to 75% of $I_{max,r}$	151	155	224	137	315

Proximity of the 75% contributing lines	17 – 54 km	8 – 46 km	2 – 30 km	8-149 km	3 – 35 km
No. of lines contributing to 90% of $I_{max,r}$	181	186	269	165	380
$I_{Effective,r}$ for a 0-degree, 1 V/km electric field (A) ($I_{Effective}$,r0)	2.967	2.427	2.232	2.471	2.471
$\Lambda_{r0} =$ $I_{Effective}$, r0/ $I_{max,r}$	0.5346	0.487	0.1064	0.3288	0.4216
$I_{Effective,r}$ for a 90-degree, 1 V/km electric field (A) ($I_{Effective}$,r90)	4.695	4.978	20.847	7.515	5.861

It is observed that the most sensitive line for a particular transformer contributes to a higher percentage of the overall GICs. The table above indicates that out of all the existing transmission lines in the network, less than 3% of the total number of lines (9140 branches) account for 95% of the GIC in most of the transformers. This ultimately means that while performing a GMD analysis, an accurate information of the GMD induced voltages on only a small subset of transmission lines would be sufficient. Some general observations from Table 5.4 are Transformer D has approximately 40% of its $I_{Effective,r}$ contributed by a single line. This finding leverages further analysis to focus on this line and all nearby lines within the span of the proximity regions. A total of 165 lines has a

greater contribution to the total current amongst the 9140 lines. For a utility operator, this information is highly useful to monitor certain regions frequently and increase their resiliency in advance. Transformer E, on the other hand is a three winding transformer and hence has about 380 lines contributing to a larger fraction of its total effective current. This necessitates the fact that this transformer must be well prepared and contain installations such as GIC blocking devices in order to avoid negative repercussions on a larger fraction of transmission lines. Transformer B, which is an autotransformer, located in the middle of the network has fewer lines influencing its effective current, this can also be attributed to the cancellation of GIC currents from opposite networks into the transformer substation.

Fig 5.2, Fig 5.3, Fig 5.4, Fig 5.5, and Fig 5.6 depict the sensitivity contours for Transformer A, B, C, D, and E, respectively. These figures illustrate the transmission line impacted by the GIC currents. The location of the transformer in these figures is indicated by a cyan oval shaped box. The red and blue colored regions indicate the branches contributing to 95% of sensitivity in both positive and negative directions, respectively. The lines which are least affected with very low sensitivity values are shown with a grey color scale. Branches having no contribution towards the GIC current, having 0 sensitivity are indicated with white background. 345kV transmission lines are depicted green in color and 138kV lines are indicated as black in color.

	Branches with positive sensitivity adding to 95% of GIC
	Branches with negative sensitivity adding to 95% of GIC
	Branches accounting very low sensitivity values
	Branches with 0 sensitivity to GIC
	345 kV transmission lines
	138 kV transmission lines
	Transformer of interest
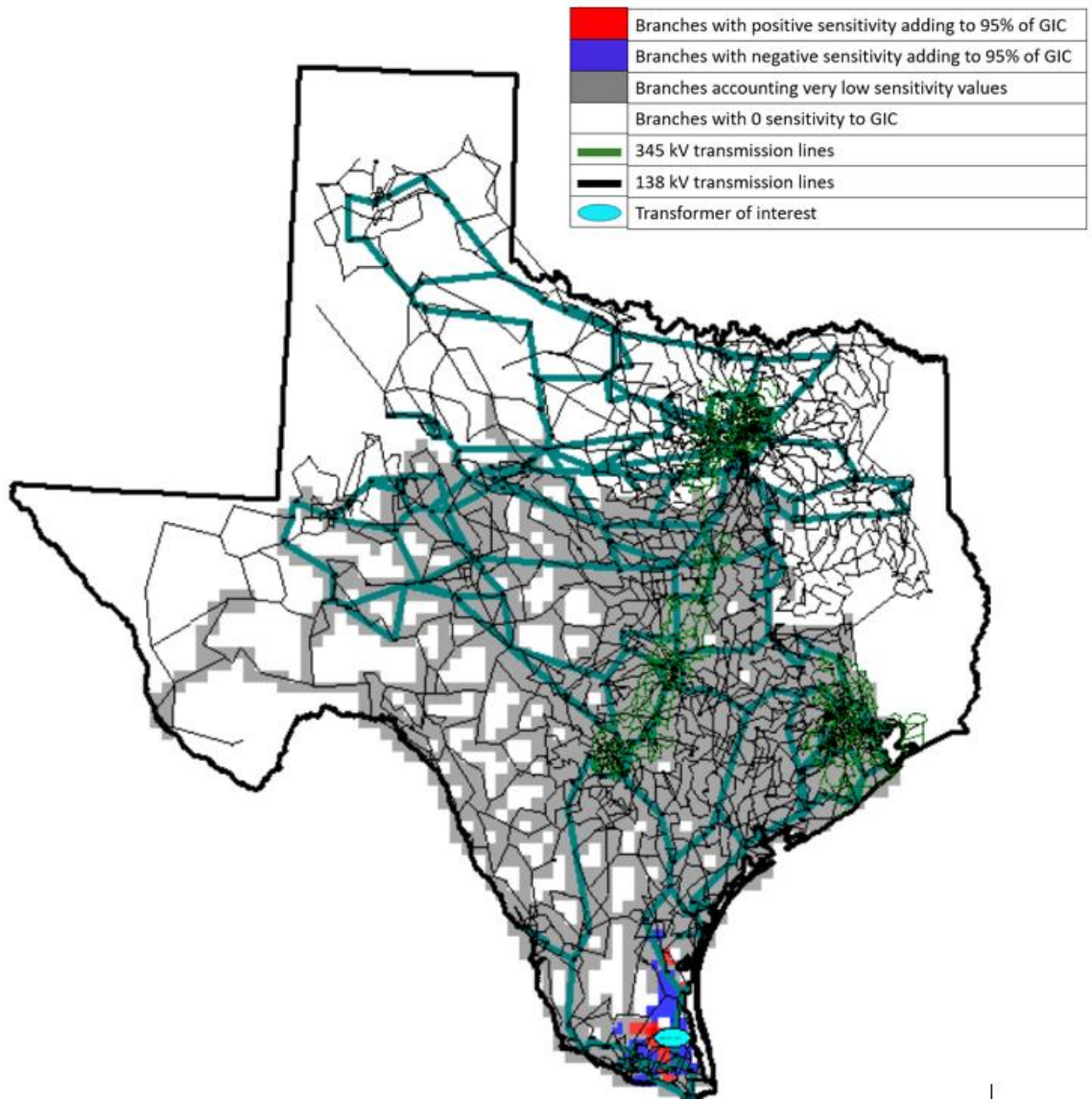
Figure 5.2 Sensitivity analysis contour showing GIC current flow in transmission lines for transformer A.
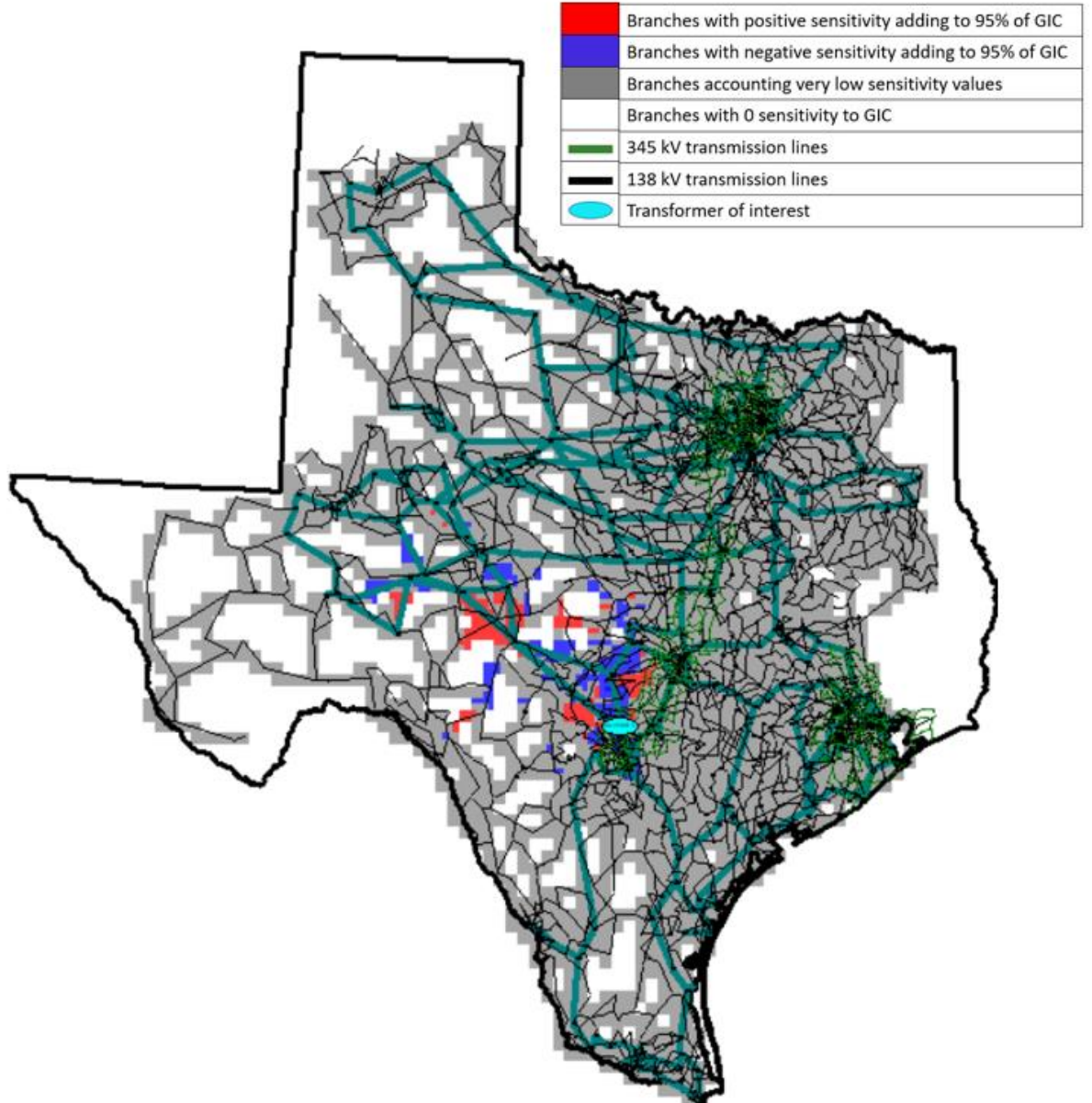
Figure 5.3 Sensitivity analysis contour showing GIC current flow in transmission lines for transformer B.
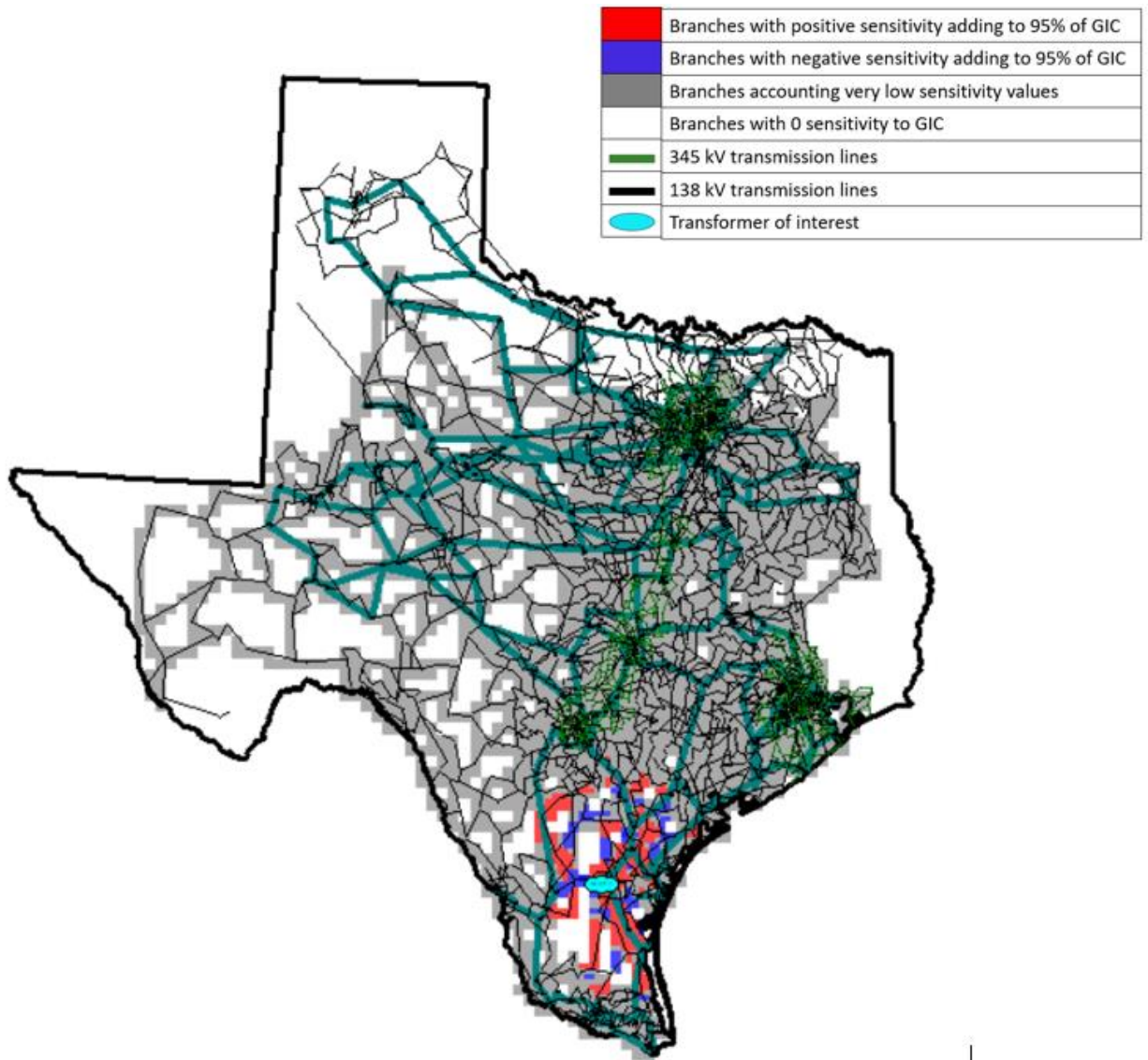
Figure 5.4 Sensitivity analysis contour showing GIC current flow in transmission lines for transformer C.
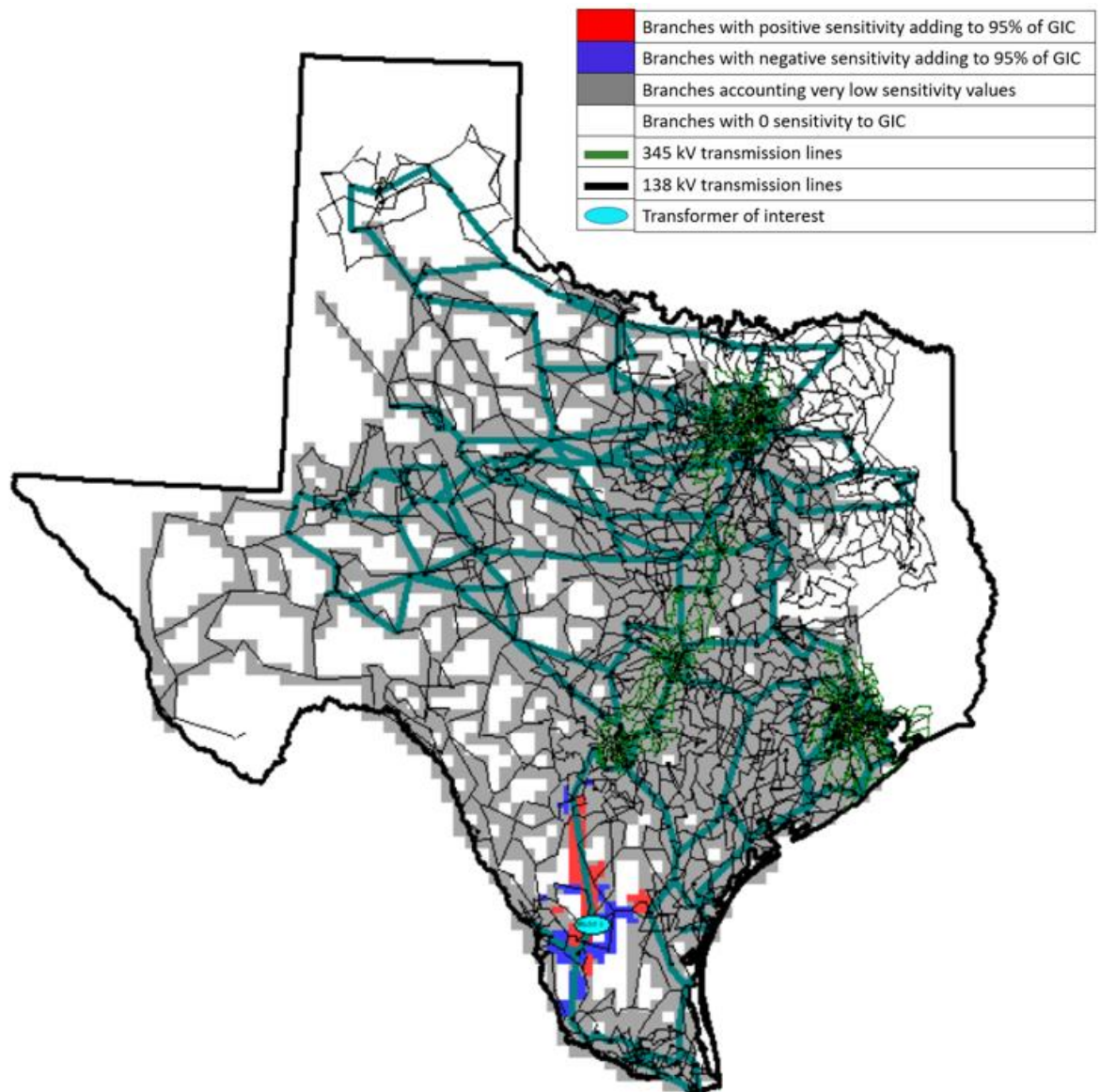
Figure 5.5 Sensitivity analysis contour showing GIC current flow in transmission lines for transformer D.
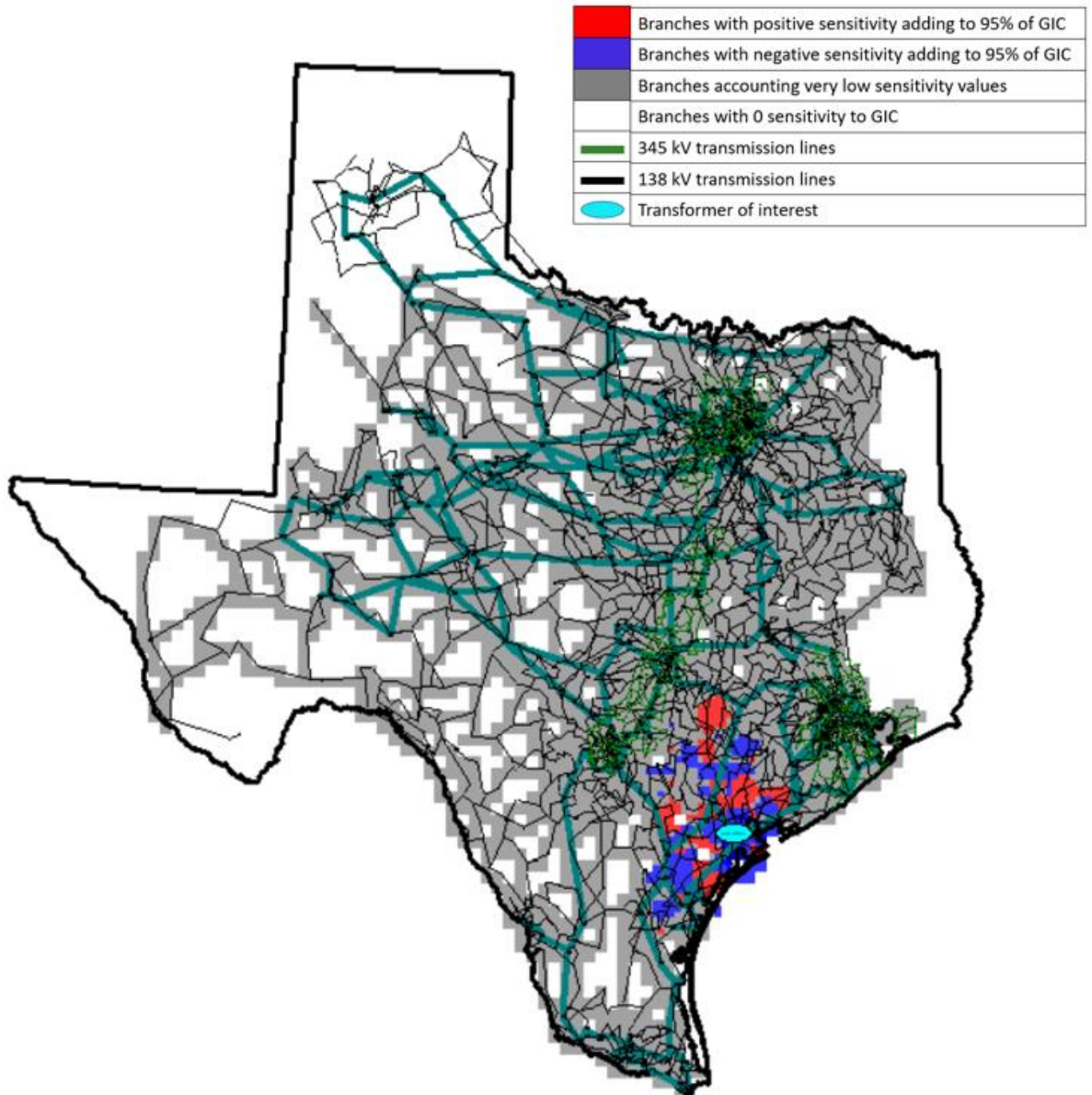
Figure 5.6 Sensitivity analysis contour showing GIC current flow in transmission lines for transformer E.

From the above figures, it is evident that Fig 5.2, shows less coverage on the contour map, which is evident from Table 5.4 that very few transmission lines (~181) carry the GIC current compared to the rest of the figures. Although many transmission

lines contribute to the GIC currents in Fig 5.3 on transformer C, there is the possibility of cancellation of GIC currents due to the strategic location of the transformer in the interior of the system. The purpose of this analysis and the plots demonstrated above is to confirm that the knowledge of a few lines is sufficient for calculating the extent of the impact of GIC currents.

6. CONCLUSIONS

6.1. Summary and Contributions

The focus of the thesis is to leverage the advantage of a huge set of historical data to develop better understanding of a designed electric grid model. Required statistical analysis and visualization is developed to study the patterns of data and validate it. Power system analysis is performed on integrating the data with the synthetic grid model to generate GIC currents, replicating different levels of storm scenarios. With the help of data and the electric grid model, system awareness is increased, and an appropriate assessment of the vulnerable regions is determined by implementing sensitivity analysis.

The entire work done follows a sequence of steps for assessing the vulnerability of power system to several disturbances. This thesis describes the procedural analysis for any given electric grid model to enhance its system knowledge. Each chapter in the thesis, elaborates on the tasks performed towards this analysis.

Chapter 2 validates data obtained TAMUMN magnetometers, along with a few other USGS stations. Several visual tools are introduced, and a combination of graphical and statistical approach is implemented for familiarizing with the data.

Chapter 3 directs the attention towards the necessity and procedure to obtain the electric field data, the influence of ground models and a few visualizations to reflect on some commonalities in electric field patterns relevant to the ground models.

Contents in chapter 4 contributes towards grid model augmentation and thoroughly elaborates on steps used to replicate the GMD scenarios in the grid model. The result of which generates huge data, that is investigated in Chapter 5 where power system

sensitivity analysis is demonstrated, and the results interpreted are used to generalize the system parameters.

Primary takeaways from the thesis includes effectively using multiple sources of data for a given measurement, by proper validation and pre-processing. Further, building upon this data and integrating it with the system model extends the benefits of historical data to enhance situational awareness of the model. As the title of the thesis suggests, existing data from available magnetometers is maximized to facilitate data-driven, real time analysis, for generating informed decisions about the system.

Hence with all the background information obtained about the system, real time analysis is benefitted by focusing on specific sections of a huge network and carefully monitoring its parameters prior to and during the occurrence of any events. This can also be useful for planning operations where device installations and mitigation strategies are taken into consideration to increase the resiliency of the system.

6.2. Future Work

The work opens several directions for future research on improving real time situational awareness for different specifications of electric grid model.

With regards to data usage, future work can use the real time visualizations elaborated here to develop event detection strategies that would use both system information and present data to identify disturbances and assess vulnerable regions of a grid model. The information obtained from historical data could be used as a preliminary tool to aid the process of focusing on required sections of an electric grid network [41].

Regarding visualization, there always exists future scope of improvement to combine historical and real time data that enables suggested operator actions and additional decision-making information. Since a large portion of the work is based on time series data, the next steps could be to work on machine learning models that explains and forecasts the time series data such as the electric field using the original magnetic field data [42]. This could also be manipulated to create fake scenarios for investigating the threshold of the electric grid parameters during worst case events.

The benefits of this research can be extended for system planning, research, and educational reasons for the purpose of improving the situational awareness and decision making for large and complex systems during a GMD event.

REFERENCES

[1] "TPL-007-2 Transmission System Planned Performance for Geomagnetic Disturbance Events," North American Electric Reliability Corporation (NERC), Tech. Rep., Oct 2017

[2] "GAO Critical Infrastructure Protection, Protecting the Electric Grid from Geomagnetic Disturbances", Dec 19, 2018. [Online]. Available: https://www.gao.gov/assets/gao-19-98.pdf.

[3] "Effects of geomagnetic disturbances on the bulk power system," North American Electric Reliability Corporation (NERC), Feb 2012.

[4] "High-Impact, Low-Frequency Event Risk to the North American Bulk Power System," North American Electric Reliability Corporation (NERC), Tech. Rep., June 2010.

[5] "Geomagnetic disturbance effects on power systems," IEEE Transactions on Power Delivery, vol. 8, no. 3, pp. 1206–1216, 1993.

[6] V. D. Albertson, J. M. Thorson, R. E. Clayton, and S. C. Tripathy, "Solar-Induced-Currents in Power Systems: Cause and Effects," IEEE Transactions on Power Apparatus and Systems, vol. PAS-92, pp. 471–477, March 1973

[7] K. Shetye and T. Overbye, "Modeling and analysis of GMD effects on power systems: An overview of the impact on large-scale power systems.," IEEE Electrification Magazine, vol. 3, pp. 13–21, Dec 2015.

[8] S. Meliopoulos, J. Xie and G. Cokkinides, "Power system harmonic analysis under geomagnetic disturbances," *2018 18th International Conference on Harmonics and Quality of Power (ICHQP),* 2018, pp. 1-6, doi: 10.1109/ICHQP.2018.8378913.

[9] U.S Department of Energy, "Geomagnetic Disturbance Monitoring Approach and Implementation Strategy", January 2019. [Online]. Available: https://www.energy.gov/sites/prod/files/2019/06/f64/DOE_GMD_Monitoring_January2019_508v2.pdf

[10] L. Marti, C. Yiu, A. Rezaei-Zare, and D. Boteler, "Simulation of Geomagnetically Induced Currents with Piecewise Layered-Earth Models," IEEE Transactions on Power Delivery, vol. 29, pp. 1886–1893, Aug 2014.

[11] T. J. Overbye, T. R. Hutchins, K. Shetye, J. Weber and S. Dahman, "Integration of geomagnetic disturbance modeling into the power flow: A methodology for large-scale system studies," 2012 North American Power Symposium (NAPS), Champaign, IL, USA, 2012, pp. 1-7, doi: 10.1109/NAPS.2012.6336365

[12] D. H. Boteler and R. J. Pirjola, "Modeling geomagnetically induced currents," Space Weather, vol. 15, pp. 258–276, Jan 2017.

[13] M.Rezaei-Zare, M.Ariannik, and A.Rezaei-Zare, (2020), Real-time detection of vulnerable power system areas to geomagnetic disturbance. IET Gener. Transm. Distrib., 14: 3838-3845. https://doi.org/10.1049/iet-gtd.2019.1871.

[14] C. Klauber, K. S. Shetye, Z. Mao, T. J. Overbye, J. Gannon and M. Henderson, "Real-Time Monitoring Applications for the Power Grid under Geomagnetic Disturbances,"

2020 IEEE Electric Power and Energy Conference (EPEC), Edmonton, AB, Canada, 2020, pp. 1-5, doi: 10.1109/EPEC48502.2020.9320114.

[15] A. Pulkkinen, R. Pirjola, & A.Viljanen. Determination of ground conductivity and system parameters for optimal modeling of geomagnetically induced current flow in technological systems. Earth Planet Sp **59,** 999–1006 (2007).

[16] K. S. Shetve, A. B. Birchfield, R. H. Lee, T. J. Overbye and J. L. Gannon, "Impact of 1D vs 3D Earth Conductivity Based Electric Fields on Geomagnetically Induced Currents*," 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe),* 2018, pp. 1-6, doi: 10.1109/ISGTEurope.2018.8571514

[17] "EPRI document on ground models", https://www.epri.com/research/products/3002014856

[18] R. Sharma and J. D. McCalley, "Extreme Value Analysis of Geomagnetically Induced Currents Based on Historical Magnetic Field Data," *2019 North American Power Symposium (NAPS)*, 2019, pp. 1-6, doi: 10.1109/NAPS46351.2019.9000198

[19] K. S. Shetye, T. J. Overbye, Q. Qiu, and J. Fleeman, "Geomagnetic disturbance modeling results for the AEP system: A case study," in 2013 IEEE Power Energy Society General Meeting, pp. 1–5, July 2013.

[20] C. Klauber, K. Shetye, T. J. Overbye and K. Davis, "A GIC Estimator for Electric Grid Monitoring During Geomagnetic Disturbances," in IEEE Transactions on Power Systems, vol. 35, no. 6, pp. 4847-4855, Nov. 2020, doi: 10.1109/TPWRS.2020.3000746

[21] "Mag-13 Three-Axis Magnetic Field Sensors," Bartington Instruments. [Online]. Available: https://www.bartington.com/wp-content/uploads/pdfs/datasheets/Mag-13 DS3143.pdf

[22] "Extended TAMUMN electric field server from CPI", http://gloc.gino.gicmagnetics.com/static/efield.b3d

[23] R. Horton, D. Boteler, T. J. Overbye, R. Pirjola and R. C. Dugan, "A Test Case for the Calculation of Geomagnetically Induced Currents," in IEEE Transactions on Power Delivery, vol. 27, no. 4, pp. 2368-2373, Oct. 2012, doi: 10.1109/TPWRD.2012.2206407

[24] A. Yan, D. Zhou, and L. Marti, "Analysis of geomagnetically induced currents," in 2013 IEEE Power Energy Society General Meeting, pp. 1–6, July 2013.

[25] K. S. Shetye *et al.*, "Development and Electric Grid Applications of a Magnetometer Network," in *IEEE Open Access Journal of Power and Energy,* vol. 8, pp. 77-84, 2021, doi: 10.1109/OAJPE.2021.3054396.

[26] "Improving Conductivity Models for Geomagnetically Induced Current Estimation," Electric Power Research Institute (EPRI), Tech. Rep, March 2020.

[27] "Geomagnetic Activity SWPC", https://www.swpc.noaa.gov/content/wmo/geomagnetic-activity

[28] "World Magnetic Model 2020," NOAA National Centers for Environmental Information, May 2020. [Online]. Available: https://www.ngdc.noaa.gov/geomag/WMM/

[29] "Kp Index parameters", https://www.spaceweatherlive.com/en/help/the-kp-index.html

[30] "INTERMAGNET Data", https://www.intermagnet.org/datadonnee/download-eng.php#

[31] "The JSON Data Interchange Syntax," Standard ECMA-404, 2017.

[32] "Grafana Data," 2020. [Online]. Available: https://grafana.com/docs/grafana/latest/features/datasources/influxdb/

[33] M. Kazerooni, H. Zhu, and T. J. Overbye, "Use of sparse magnetometer measurements for geomagnetically induced current model validation," in 2015 North American Power Symposium (NAPS), 2015, pp. 1–6.

[34] "USGS Geomagnetic Observatories." https://www.usgs.gov/natural-hazards/geomagnetism/science/observatories?qt-science center objects=0#qt-science center objects.

[35] "Minitab for ANOVA", https://support.minitab.com/en-us/minitab/18/help-and-how-to/modeling-statistics/anova/supporting-topics/basics/what-is anova/#:~:text=Learn%20more%20about%20Minitab%2018,at%20the%20different %20factor%20levels

[36] J. L. Gannon, A. B. Birchfield, K. S. Shetye, and T. J. Overbye, "A Comparison of Peak Electric Fields and GICs in the Pacific Northwest Using 1-D and 3-D Conductivity," Space Weather, vol. 15, no. 11, pp. 1535–1547, 2017.

[37] A. Birchfield, "The Creation, Validation, and Application of Synthetic Power Grids", Ph.D. dissertation, ECEN Dept, Texas A&M University, College Station, Texas, USA, 2018.

[38] "PowerWorld GIC currents",https://www.powerworld.com/knowledge-base/gic-time-varying-electric-field-example

[39] T. J. Overbye, K. S. Shetye, T. R. Hutchins, Q. Qiu, and J. D. Weber, "Power grid sensitivity analysis of geomagnetically induced currents," IEEE Transactions on Power Systems, vol. 28, pp. 4821–4828, Nov 2013.

[40] Y. Zhang, K. S. Shetye, A. B. Birchfield and T. J. Overbye, "Grid Impact Evaluation of Localized Geomagnetic Field Enhancements Using Sensitivity Analysis," 2019 North American Power Symposium (NAPS), Wichita, KS, USA, 2019, pp. 1-6, doi: 10.1109/NAPS46351.2019.9000197.

[41] D. H. Boteler, L. Trichtchenko, R. Pirjola, J. Parmelee, S. Souksaly, A. Foss, and L. Marti, "Real-time simulation of geomagnetically induced currents," in 2007 7th International Symposium on Electromagnetic `Compatibility and Electromagnetic Ecology, pp. 261–264, June 2007.

[42] Y. Ge, A. J. Flueck, D. Kim, J. Ahn, J. Lee and D. Kwon, "Power System Real-Time Event Detection and Associated Data Archival Reduction Based on Synchrophasors," in IEEE Transactions on Smart Grid, vol. 6, no. 4, pp. 2088-2097, July 2015, doi: 10.1109/TSG.2014.2383693.