

Ensemble Estimation of Network Parameters: A Tool to Improve the Real-time Estimation of GICs in the South African Power Network

M.J. Heyns^{1,2}, S.I. Lotz¹, P.J. Cilliers¹, C.T. Gaunt²

¹South African National Space Agency, Space Science Directorate, Hermanus, South Africa

²Department of Electrical Engineering, University of Cape Town, Cape Town, South Africa

E-mail: mheyns@sansa.org.za

Abstract. It has long been known that large grounded conducting networks on the surface of Earth, such as power lines or pipelines, are affected by solar activity and subsequent geomagnetic storms. With the increased use of electrical technologies, society has become more and more dependent on electrical power and power networks. These power networks form extensive grounded conductors which are susceptible to geomagnetically induced currents (GICs). GICs at any specific node in a power network are assumed to be linearly related to the horizontal vector components of an induced plane-wave geoelectric field by a pair of network parameters. These network parameters are not easily measured in the network, but may be estimated empirically. In this work, we present a new approach of using an ensemble of network parameters estimates. The ensembles include a huge number of parameter pair estimates calculated from simultaneously solving pairs of time instances of the governing GIC equation. Each individual estimate is not the true state of the system, but a possible state. Taking the ensemble as a whole though gives the most probable parameter estimate. The most probable parameter estimate for both network parameters, as defined by their respective ensembles, is used directly in the modelling of GICs. The ensembles themselves however allow for further analysis into the nature of GICs. An improvement is seen when comparing the out-of-sample performance of the ensemble estimates with previous GIC modelling in the South African power network during the Halloween Storm of 2003. For the first time, it is shown that errors in the GIC modelling chain are absorbed into the network parameter estimates. Using a range of estimates from the ensemble, a GIC prediction band is produced. This band corresponds to an error estimate for predicted GIC. Furthermore, it has been explicitly shown for the first time that estimated network parameters vary with GIC magnitude during an event. This behaviour is then used to refine the parameter estimation further and allow for real time dynamic network parameter estimation.

1. Introduction

Modelling geomagnetically induced currents (GICs) entails not just modelling a single system, but rather a chain of coupled systems. The general driver of a geomagnetic storm disrupts current systems in the near-Earth environment and hence the geomagnetic field. Fluctuations in the geomagnetic field through a closed path including the power network induces a geoelectric field and an electromotive force (EMF), which ultimately drives the GIC. The effects of the resulting GICs in these networks can be particularly damaging and costly [1], especially in high-latitude regions. Lesser, but still significant effects have also been observed at mid-latitudes [2–4].

There are a few governing assumptions made in GIC modelling. It is assumed that the induced geoelectric field is driven by a spatially constant plane-wave geomagnetic field over the system

[5]. Such an induced geoelectric field would be uniform if the conductivity of Earth is dependent only on depth (laterally constant). Typical causes of such a geomagnetic field would be a very long (relative to the system) uniform sheet or line current. These assumptions are idealised cases which can be thought of as first order approximations [6]. In reality, the ionospheric current systems are exceptionally complex, non-uniform and highly variable. At mid-latitudes the useful plane-wave geomagnetic field assumption is more accurate than in auroral regions where the auroral electrojet is the dominant current system. This current system is more variable and closer compared to the mid-latitude drivers, making the interactions more complicated [3, 7, 8].

1.1. Network Parameters

Due to the conductivity and extent of the Earth, the main component of the driving EMF comes from the induced geoelectric field at the surface of the Earth. This results in a linear relation between the geoelectric field and the resulting GIC through Ohm's law, defined by the projection of the effective geoelectric field onto the network and the sum of all resistances in the induction loop (which is dominated by network resistances). This results in the governing GIC equation in Cartesian coordinates,

$$GIC(t) = aE_x(t) + bE_y(t), \quad (1)$$

where E_x and E_y are the horizontal geoelectric field components in the Northerly and Easterly directions respectively. Since the network parameters are scaling factors that also penalise non-alignment of the geoelectric field vector with the network, the preferred direction for the geoelectric field to produce large GICs can be found. The ratio of network parameters b/a as used in previous work [9] is encoded in this direction (in this case effectively the angle from North). Any deviation or non-alignment of the geoelectric field from the preferred direction would result in a fraction of the total geoelectric field magnitude contributing to the measured GIC. These traditional network parameters, with units of Akm/V, are usually determined analytically using available information about the network, which is often inaccurate.

2. Model Development

From simultaneous GIC and $E_{x,y}$ data, a single estimate of the network parameters can be derived empirically [8, 10]. It is assumed that these parameters are constant over the time scale of a geomagnetic event, only changing with major alterations of power network related hardware or operations [11]. Due to associated errors at each point in the GIC modelling chain, it has been observed that different empirical values of a and b may be derived for different data subsets defined by magnitude or conductivity [8, 10]. This suggests that a number of errors from different drivers are absorbed into these parameters. To acknowledge these errors and borrowing notation from Wik et al. [8], where α and β represent the empirically derived network parameters, we can redefine the governing equation to,

$$\Gamma(t) \approx \alpha E_x(t) + \beta E_y(t), \text{ where} \quad (2)$$

$$\Gamma(t) \equiv GIC(t) + GIC(t)_{err} \text{ (or the GIC as measured),} \quad (2a)$$

$$\alpha \equiv a(1 + E_x(t)_{err}/E_x(t)) \text{ and} \quad (2b)$$

$$\beta \equiv b(1 + E_y(t)_{err}/E_y(t)). \quad (2c)$$

Given a time-series of n time instances, each a measured state of the system, taking any two relevant and comparable time instances would allow for α and β to be solved for simultaneously. Using all possible combinations of time instance pairs would result in $n(n-1)/2 \approx n^2/2$ (for large n) sets of empirical network parameter estimates. These estimates are collected into the ensembles α and β . Using a similar approach of statistical mechanical ensembles, each estimate in the ensemble represents a possible state of the real system. When all states are considered together and the result normalised appropriately, an ensemble becomes a probability distribution of system states.

Previous empirical estimation has generally included a fit of sorts that results in a single estimate. One approach is to use a least squares routine and fit equation (1) to the data [8]. Another approach is to select near-zero crossings of the geoelectric field for a single geoelectric field component and solve equation (1). This generates a number of single parameter estimates and a linear fit is then applied (excluding outliers) to find the final estimate [10]. It has also been shown that the ratio of network parameters can be found empirically from data [12]. If a single parameter is known, this ratio can be used to find the other parameter in the parameter pair [9]. All these methods estimate the network parameters and absorb effects analytical methods don't, but do not give an indication of any variation in the parameters (in which errors are propagated), whereas the spread of the resulting ensembles do.

2.1. Data Sources and Selection

This work analyses the Grassridge (GRS) substation in the South African power network, where previous GIC modelling has been done. Measured 2-sec transformer neutral GIC data for 31 March 2001 and 29-31 October 2003 is used. This data range spans geomagnetic storms and no network changes were made during this time (comparable time instances). The geoelectric field data used in this work is derived from 1-min magnetometer data measured at Hermanus (which has been shown to be 'local' enough for GIC modelling [13]). This process makes use of the magnetotelluric method with a layered Earth conductivity model [12]. The conductivity profiles used in this work include the local empirically derived 10-layer profile for Grassridge [9] and the non-local 5-layer for Québec (QUE) [14]. To make use of relevant data, a process of data selection was implemented. In previous work this entailed using only GIC data which satisfies $|GIC| > 0.1 \times RMS(GIC)$, where $RMS = \sqrt{\sum_{i=1}^n GIC(t_i)^2/n}$ (root mean square), and further selection of significant geoelectric field time instances (which can lead to biasing). Since ensemble estimation is robust and makes use of all possible combinations of time instances, this selection can be relaxed and varied, with the only criterion being that there need to be enough time instances to create a large representative ensemble.

3. Ensemble Estimation Results

In order to compare ensemble estimation to previous work the same GIC data selection criterion is used initially. The data from the Halloween Storm of 29 October 2003 is kept out of the ensemble training as a validation set (not done in previous work). Figure 1 shows the resulting normalised α ensemble histogram (with a similar result applicable for β). It should be noted that although the probability distribution has the expected bell-shape as predicted from equation (2b), it also has significantly heavy tails (variance and higher order moments are undefined). In this case, the median and not the mean is the best measure of central tendency and hence the most probable network parameter estimate. Also shown is the effect of using different conductivity profiles, with the local empirically derived profile having less spread but no shifting of the peak. This shows that errors made in the derivation of the geoelectric field are absorbed into the empirical network parameters.

3.1. Comparison with Previous Work

A comparison is made with the empirical network parameter estimates of two previous studies, namely Ngwira et al. [9] and Matandirotya et al. [10]. Both these studies used the first day of the Halloween Storm, 29 October 2003, for GIC modelling. During this day, the period between 06:00 and 12:00 is highly disturbed according to SYM-H and the period between 19:00 and 24:00 is relatively disturbed (see Figure 3). The periods 06:00-12:00 and 19:00-24:00, will be referred to as the Primary and Secondary Storm Phases respectively. In addition to this, Matandirotya et al. [10] used the finite element method (FEM) to further improve the derived geoelectric field results (not done in this work).

To measure the performance of the modelling a number of error metrics are used, namely the *RMSE* (root mean square error), *RE* (relative error) and Pearson's correlation coefficient

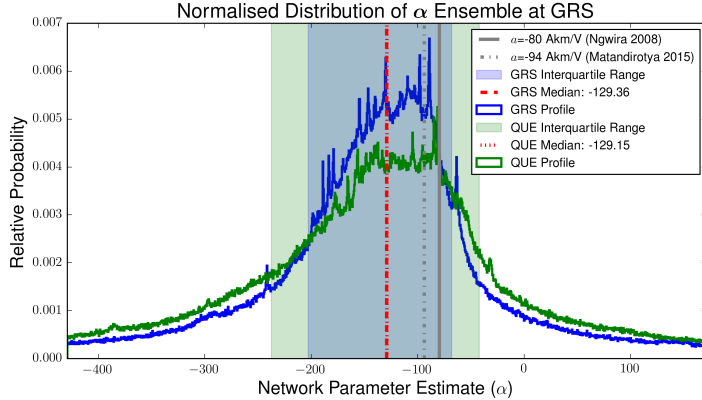


Figure 1. Comparison of the α distributions derived from the local (GRS) and Québec (QUE) conductivity profiles. This behaviour also holds for the β distributions. The interquartile range shown is used to model a prediction band (Figure 3). Also shown are the network parameter estimates from previous work.

ρ . The first two metrics quantify the error in absolute amplitude and are defined by,

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (GIC_{obs}(t_i) - GIC_{mod}(t_i))^2}{n}} \quad \text{and} \quad RE = \frac{GIC_{obs} - GIC_{mod}}{GIC_{obs}}. \quad (3)$$

Using the previous work's definition, only the median RE for $|GIC| > 1$ A is considered and the result is shown as a percentage [10]. Pearson's correlation coefficient on the other hand is defined in the normal way and quantifies the correlation. The results summarised in Table 1 below show that ensemble estimation does significantly better in general (an overall improvement of more than 10% using the locally based conductivity profile and more than 5% using the non-local profile).

Table 1. Ensemble estimation results compared to previous work.

Data	RMSE [A] (ρ)			RE%
	06:00-12:00	19:00-24:00	00:00-24:00	
<i>Ngwira Set</i> ($a=-80, b=1$ A km/V)				
FEM	0.96	1.07	1.35	51
<i>Matandirotya Set</i> ($a=-94, b=24$ A km/V)				
FEM	1.38	1.11	0.98	41
<i>Grassridge Profile</i> ($\alpha=-129.36, \beta=7.90$ A km/V)				
GRS	1.42 (0.88)	0.54 (0.97)	0.86 (0.88)	30
<i>Québec Profile</i> ($\alpha=-129.15, \beta=5.61$ A km/V)				
QUE	1.78 (0.78)	0.81 (0.93)	1.12 (0.79)	35

3.2. Network Parameter Dependence on GIC Magnitude

Relaxing the criterion for using significant GIC data (considering that already only data corresponding to geomagnetic storms is used), we can use different percentile windows relating to GIC magnitude to quantify the previously observed behaviour that different empirical network parameters are obtained for different data subsets. Taking a 25% percentile window for GRS results in more than 1 million parameters pairs per window, satisfying the condition that a representative ensemble is produced. Arranging the percentile range estimates along increasing values of measured GIC in Figure 2 shows that α and β are apparently not constant and vary with GIC intensity as the storm evolves (since GIC magnitude varies with time). At small magnitude GICs, the network parameters are negligible. This suggests that the preferred direction is indeterminate and that these cases are often a result of miss-alignment and/or a small geoelectric field. As the GIC strength grows, the network parameters become more and more relevant. This is not the case with the β parameter at GRS which stays close to zero. This is a

function of the directionality weighting from the network. GRS is located at a stable endpoint in the network with only a single, approximately North-South directed line. Therefore the β parameter, which scales the Eastward geoelectric field component E_y , is small. This suggests that GICs at GRS are practically independent of the Eastward component of the geoelectric field, which has been seen in previous analysis [2, 9, 10]. Even with this variation, the preferred direction as defined by the ratio of network parameters remains constant.

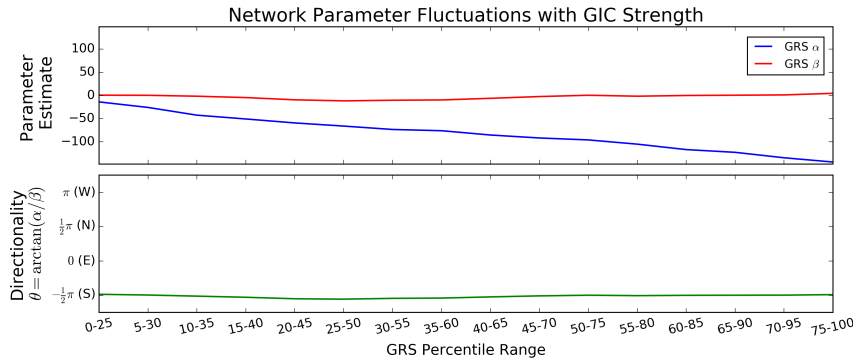


Figure 2. Plot of empirical network parameter estimates at GRS for different magnitude defined GIC percentile ranges. Also shown is the constant ratio of network parameters (network defined directionality).

3.3. GIC Prediction Band

The spread in the distributions of estimated network parameters suggests that using single values of α and β to relate GIC to induced geoelectric field is not correct. Instead of using this approach, we can use the interquartile range of α and β (see Figure 1) to predict an associated range of GICs. This network parameter range would span the typical error propagated in the GIC modelling chain, without straying into the heavy tails. Since we are most concerned about the largest GICs, we then use the values of α and β in this interquartile range that will maximise the range of resulting GIC. This approach results in a GIC band, as shown in Figure 3, instead of a single estimate. This gives a handle on the uncertainty in the prediction at any time.

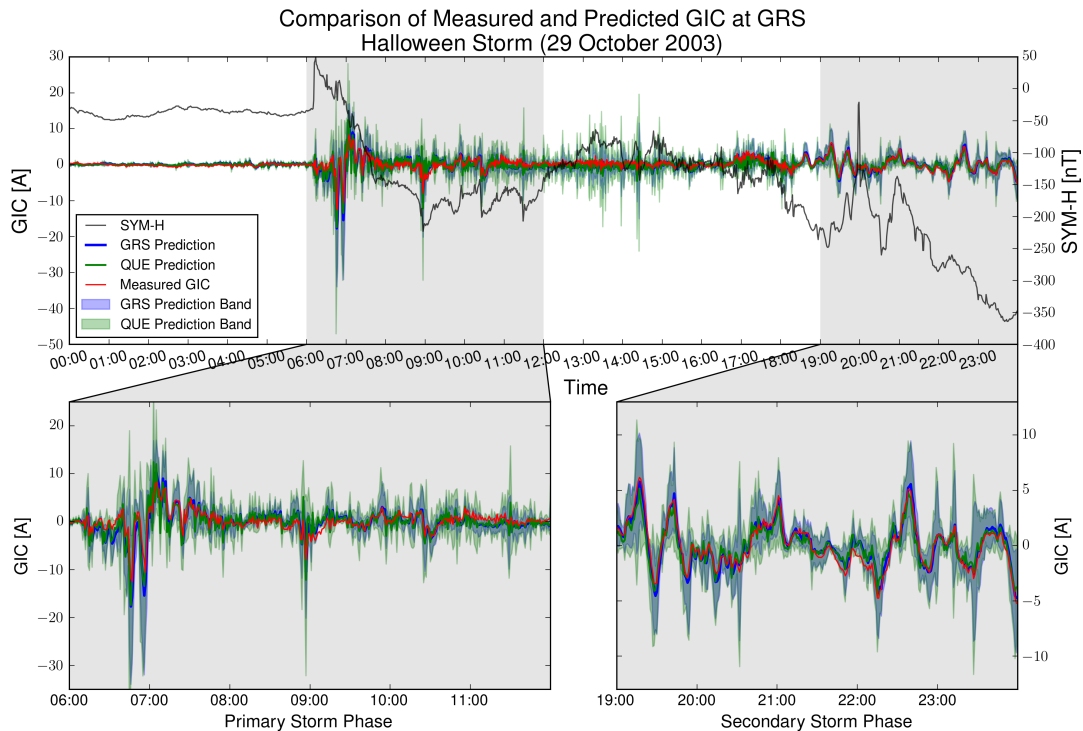


Figure 3. A GIC prediction band produced using a range of estimates.

3.4. Dynamic Network Parameters

Using the observed correlation of the network parameters with GIC magnitude, dynamic network parameters can be created to further improve modelling. The percentile window shifted through all the percentiles results in an overlap of estimates for a given GIC magnitude. Using a weighted mean of these overlapping estimates would result in a single representative estimate for a corresponding GIC magnitude. Calculating an arbitrary GIC magnitude's relevant percentile in the ensemble would then allow the representative network parameter estimate to be mapped to it. Using the GRS data with its low resolution and the non-local QUE profile (that should do a lot worse than a locally derived profile) we see an improvement nonetheless (see Table 2). Also shown are the highest percentile range parameters, appropriate for extreme value modelling.

Table 2. Dynamic network parameter estimation using GIC magnitude.

Québec Profile	RMSE [A] (ρ)			RE%
	06:00-12:00	19:00-24:00	00:00-24:00	
Static Parameters	1.78 (0.78)	0.81 (0.93)	1.12 (0.79)	35
Dynamic Parameters	1.75 (0.82)	0.72 (0.94)	1.05 (0.82)	36
Extreme Parameters	1.97 (0.79)	0.78 (0.93)	1.21 (0.79)	36

4. Discussion and Conclusion

Not only does ensemble estimation give a much better prediction compared to previous analytical and empirical methods, but the uncertainty in GIC modelling is also quantified. It has been shown how errors made in the derivation of the geoelectric field components are reflected in the network parameters. Taking this into account and a range of values for each of the network parameters, a band of GIC may be calculated instead of a single estimate. Further analysis has also explicitly shown for the first time that the network parameters are not constant and are correlated with GIC strength and hence storm phase. Using this to define dynamic or extreme parameter estimates further improves GIC modelling accuracy and usefulness. These approaches can be incorporated into a real-time probabilistic prediction scheme, with direct application to mitigation schemes for power utilities and other relevant parties.

References

- [1] Oughton E J, Skelton A, Horne R B, Thomson A W P and Gaunt C T 2017 *Space Weather* **15** 65–83
- [2] Koen J 2002 *Geomagnetically Induced Currents in the Southern African Electricity Transmission Network* PhD Thesis University of Cape Town
- [3] Kappenman J G 2003 *Space Weather* **1**
- [4] Gaunt C T and Coetzee G 2007 *2007 IEEE Lausanne Power Tech* 807–812
- [5] Pirjola R 1982 *Geophysica* **18** 1–161
- [6] Boteler D H and Pirjola R J 2017 *Space Weather* **15** 258–276
- [7] Boteler D H and Pirjola R J 1998 *Geophysical Journal International* **132** 31–40
- [8] Wik M, Viljanen A, Pirjola R, Pulkkinen A, Wintoft P and Lundstedt H 2008 *Space Weather* **6**
- [9] Ngwira C M, Pulkkinen A, McKinnell L A and Cilliers P J 2008 *Space Weather* **6**
- [10] Matandirotya E, Cilliers P J and Van Zyl R R 2015 *Space Weather* **13** 185–195
- [11] Viljanen A, Pirjola R, Prácsér E, Ahmadzai S and Singh V 2013 *Space Weather* **11** 575–584
- [12] Pulkkinen A A, Pirjola R and Viljanen A 2007 *Earth, Planets and Space* **59** 999–1006
- [13] Ngwira C M, McKinnell L A, Cilliers P J, Viljanen A and Pirjola R 2009 *Space Weather* **7**
- [14] Boteler D and Pirjola R 1998 *IEEE Transactions on Power Delivery* **13** 1303–1308