Forecasting GICs and geoelectric fields from solar wind data using LSTMs: application in Austria

Rachel L. Bailey[1,1], Roman Leonhardt[1,1], Christian Moestl[2,2], Ciaran Beggan[3,3], Martin Reiss[2,2], Ankush Bhaskar[4,4], and Andreas Jeffrey Weiss[5,5]

[1]ZAMG
[2]Austrian Academy of Sciences
[3]British Geological Survey Edinburgh Office
[4]ISRO/Vikram Sarabhai Space Centre
[5]Space Research Institute

November 30, 2022

Abstract

The forecasting of local GIC effects has largely relied on the forecasting of dB/dt as a proxy and, to date, little attention has been paid to directly forecasting the geoelectric field or GICs themselves. We approach this problem with machine learning tools, specifically recurrent neural networks or LSTMs by taking solar wind observations as input and training the models to predict two different kinds of output: first, the geoelectric field components Ex and Ey; and second, the GICs in specific substations in Austria. The training is carried out on the geoelectric field and GICs modelled from 26 years of one-minute geomagnetic field measurements, and results are compared to GIC measurements from recent years. The GICs are generally predicted better by an LSTM trained on values from a specific substation, but only a fraction of the largest GICs are correctly predicted. This model had a correlation with measurements of around 0.6, and a root-mean-square error of 0.7 A. The probability of detecting mild activity in GICs is around 50%, and 15% for larger GICs.

Forecasting GICs and geoelectric fields from solar wind data using LSTMs: application in Austria

R. L. Bailey[1], R. Leonhardt[1], C. Möstl[2], C. Beggan[3], M. A. Reiss[2], A. Bhaskar[4], A. J. Weiss[2,5]

[1]Conrad observatory, Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria
[2]Space Research Institute, Austrian Academy of Sciences, Graz, Austria
[3]British Geological Survey, Edinburgh, UK
[4]Space Physics Laboratory, ISRO/Vikram Sarabhai Space Centre, Trivandrum, India
[5]Institute of Physics, University of Graz, Universitätsplatz 5, 8010 Graz, Austria

Key Points:

- The aim is to directly forecast GICs rather than dB/dt, which is often used as a proxy.
- Results from LSTMs predicting either Ex and Ey or substation GICs from solar wind data are compared.
- GIC forecasting seems to work best when the LSTM model is trained directly on GIC data.

Corresponding author: Rachel Bailey, `r.bailey@zamg.ac.at`

Abstract

The forecasting of local GIC effects has largely relied on the forecasting of dB/dt as a proxy and, to date, little attention has been paid to directly forecasting the geoelectric field or GICs themselves. We approach this problem with machine learning tools, specifically recurrent neural networks or LSTMs by taking solar wind observations as input and training the models to predict two different kinds of output: first, the geoelectric field components E_X and E_Y; and second, the GICs in specific substations in Austria. The training is carried out on the geoelectric field and GICs modelled from 26 years of one-minute geomagnetic field measurements, and results are compared to GIC measurements from recent years. The GICs are generally predicted better by an LSTM trained on values from a specific substation, but only a fraction of the largest GICs are correctly predicted. This model had a correlation with measurements of around 0.6, and a root-mean-square error of 0.7 A. The probability of detecting mild activity in GICs is around 50%, and 15% for larger GICs.

Plain Language Summary

Using satellites, we measure the state of the solar wind a short distance away from the Earth (at the so-called Lagrange-1 or L1 point) to see what is coming towards us at any given moment. Changes in the solar wind such as an increase in wind speed or a strong magnetic field can potentially impact satellite operation in orbit and power grid infrastructure on the ground - in extreme cases, solar storms can damage power grids and transformers by inducing electrical currents in the power lines. These are called geomagnetically induced currents (GICs). Here, we attempt to forecast the scales of GICs by applying machine learning methods, specifically Long-Short-Term-Memory recurrent neural networks, to take the solar wind data measured at the L1 point and predict the currents that would be seen in power grids in Austria. This gives us a lead time of around 10 to 40 minutes in the forecast. We discuss whether it is best to attempt to predict the ground electric field that leads to the GICs or the GICs themselves, and discuss the difficulties in this kind of prediction and the shortfalls in the model.

1 Introduction

Geomagnetically induced currents (GICs) have long been known to affect power grids, transformers and any earthed conductive networks spanning large distances (for an overview, see Boteler et al., 1998; Boteler & Pirjola, 2017; Kelbert, 2020). GICs can cause problems in power grid operation such as transformer overheating or permanent transformer damage and system collapse in extreme cases (Molinski, 2002), leading to further societal and economic harm (Eastwood et al., 2018). Although studies of GICs were restricted to high latitudes where the consequences are more pronounced, mid-latitudes are being paid increasingly more attention as local effects such as transformer overheating are discovered (Barbosa et al., 2015; Butala et al., 2017; Lotz & Danskin, 2017; Gil et al., 2019; Caraballo et al., 2020; Svanda, Michal et al., 2020, among others).

The forecasting of GICs has developed alongside studies into the effects of regional GICs (Pulkkinen et al., 2006). Forecasting in particular is a complex problem due to the chain of cascading induction effects from the impingement of solar wind at the bow shock down to currents flowing between the earth and power grids on the surface. Improving predictive GIC modelling is listed as one of the open questions still to address to achieve GIC readiness (Pulkkinen et al., 2017).

Most studies so far have focused on predicting geomagnetic activity - such as dB/dt, which is often used as a proxy for GICs - from solar wind data measured at L1 or in near-Earth space. The earliest studies addressing this problem with neural network architecture are Wintoft (2005) and Wintoft et al. (2015), followed by Lotz and Cilliers (2015) and recently Keesee et al. (2020) and Tasistro-Hart et al. (2021). The $Dst/SYMH$ index in particular has received a lot of attention from geophysicists and machine learning engineers alike (e.g. Lu et al., 2016; Bhaskar & Vichare, 2019; Wintoft & Wik, 2021).

While dB/dt is often used as a proxy for GICs, it does not provide the whole picture. The downside of modelling with this approach is that dB/dt only functions as a useful indicator of GIC activity. The relationship between dB/dt and E (which is the primary factor determining the scale of the GICs) depends on the magnetotelluric transfer function, which is frequency dependent (Chave & Jones, 2012). Single values of the time derivative of the magnetic field can only be useful GIC proxies if further assumptions on the frequency content are made (Pulkkinen et al., 2006).

What do we do if we want to develop a model that provides forecasts that power grid operators can work with? One approach would be to directly forecast the surface geoelectric field, from which GICs at different stations can be calculated. In comparison to the many studies into forecasting dB/dt and Dst, little effort has been devoted to forecasting geoelectric fields thus far. Pulkkinen et al. (2009, 2010) studied the forecasting of GICs from remote solar observations, allowing a few days warning before larger events. Modelling of geoelectric fields from solar wind to ground using full MHD modelling has been carried out by Pulkkinen, Hesse, et al. (2007), Zhang et al. (2015) and Honkonen et al. (2018), and with empirical modelling in Lotz et al. (2017).

In this study, we aim to tackle this problem from another angle and forecast regional GICs from L1 solar wind data using a machine learning method, and we compare the results to observations of GICs in Austria. We try this with two different approaches: in the first, we train a model to forecast the geoelectric field and calculate the GICs from there, and in the second we forecast the GICs directly. Predictions from both methods are evaluated and compared using data from recent years.

This study is structured as follows. Section 2 describes the data used in this study, including an analysis of 26 years of geomagnetic measurements used to model GICs in the region of Austria and a case study looking at the 2003 Halloween storm. Section 3 then goes on to describe the models built to forecast GIC values, and the results are presented in Section 4, discussed in Section 5 and summarised in Section 6.

2 Data

This analysis relies on INTERMAGNET-quality geomagnetic observatory data, which ensures a high quality of data with few data gaps or disturbances. We use data with a cadence of one minute because these are available for a long time period (26 years), which is not possible with 1 Hz data. Data with 1-minute resolution should be representative of most important GIC content (Pulkkinen et al., 2006). Due to Austria's small size (roughly 280 x 600 km), we assume that the geomagnetic variations are roughly constant across it both latitudinally and longitudinally, and therefore only select and use geomagnetic variations from one station at a time.

In the following, we describe the data sets used in this study. Geomagnetic field variations from observatory measurements were used to calculate the ground geoelectric field in Austria. GICs at any power grid substation can be calculated from the geoelectric field, and the equations for two specific substations are determined using a linear fit to observed GICs. In terms of the geomagnetic and geoelectric field components, x and y refer to the geographic northward and eastward directions respectively.

2.1 Geomagnetic observatory data from WIC and FUR

The Conrad Observatory (WIC), situated at a geomagnetic latitude of 42.95° and longitude of 89.94° according to AACGM-v2 (Shepherd, 2014), is located southwest of

Vienna near the town of Muggendorf in Lower Austria. High quality geomagnetic measurements have been carried out here since the official opening mid-2014, providing six years of data for analysis. We extend the time range using data from Fürstenfeldbruck (FUR) in Bavaria, Germany. Initial studies are done using WIC data, and studies of long-term measurements are carried out using FUR data. A map showing the location of the two stations can be found in **Fig. 1**.

The Fürstenfeldbruck Geomagnetic Observatory (geomagnetic lat: 43.06°, lon: 85.93°) is one of the closest INTERMAGNET-quality geomagnetic observatories to the Conrad Observatory. It is situated almost directly west of WIC and separated by 348 km. This station is a very good proxy for geomagnetic field variations in Austria due to its proximity and the similar geomagnetic latitude and geological setting. Measurements at a quality high enough for this analysis have been carried out since 1995, providing twenty-six years of data or 13.7 million data points at a 1-minute resolution.

An analysis of the coherence between WIC and FUR data has been carried out for the overlapping years of measurements (2015-2021), in which the Pearson's correlation coefficient (PCC) between the two time series doesn't drop below 0.99 for either the x or y variables over all six years. The correlation in variations (dBx/dt and dBy/dt) is slightly lower, with the lowest values (0.91) seen in the dBy/dt values.

2.2 Geoelectric field

In order to model the expected levels of GICs, we need knowledge of the ground geoelectric field in the region. The geoelectric field for the past 26 years is modelled directly from the 1-minute geomagnetic field variations at FUR. The model approach used is the one-dimensional plane wave method (e.g. Boteler & Pirjola, 2017) using the EURHOM model number 39 (Ádám et al., 2012) to describe the one-dimensional layers of resistivity going into the Earth. We assume the time series is representative across the country, which is a reasonable approach for small areas but not for larger countries. The plane wave approach was used in favour of the thin-sheet approach used in previous studies (Bailey et al., 2017, 2018) for the shorter computation times with similar levels of accuracy. The calculation results in the horizontal geoelectric field components E_X and E_Y. Note that the x-component in the geoelectric field corresponds to the y-component geomagnetic field variations, and vice versa.

2.3 Geomagnetically induced currents

To evaluate the levels of GICs over the 26 years of available FUR data, we do not follow the standard modelling procedure of putting the geoelectric field components through the full power grid network, which would be computationally heavy, but instead find a direct linear fit of the geoelectric field components to measurements of GICs to find the current at station j, i.e.

$$GIC_j = a_j \cdot E_x + b_j \cdot E_y \tag{1}$$

where a_j and b_j are station-specific real coefficients (with units A·km/V). This approach can only be used on transformer stations with measurements since the coefficients must be determined from a linear fit to the data, but it often has similar or better accuracy than results from a network model. See Pulkkinen, Pirjola, and Viljanen (2007) or Torta et al. (2012) for more discussion on this method and for the equations determining a_j and b_j.

The fit for Eq. 1 was applied to measurements of direct currents from multiple transformer neutral points in Austrian power grid substations provided by the Graz University of Technology. In this study, only measurements from two substations were used:

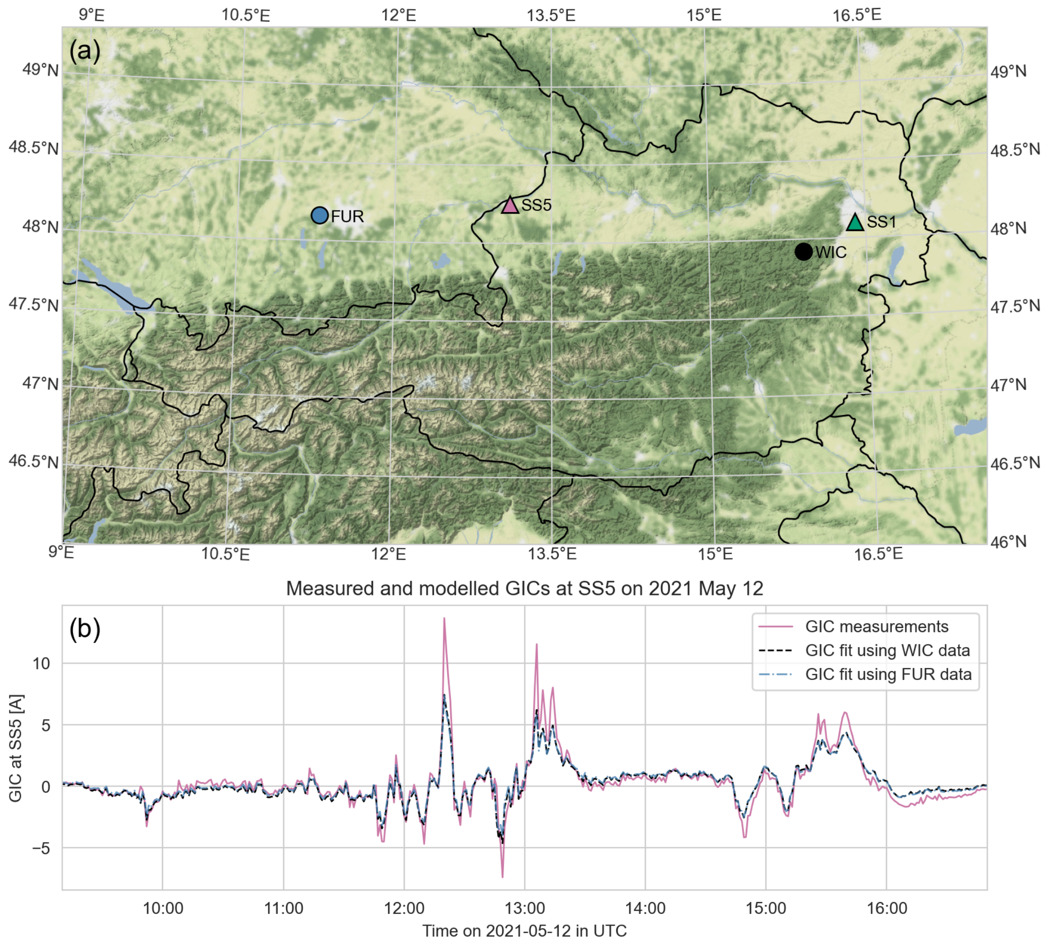

Figure 1. (a) A map showing the locations of two power grid substations (triangles) and the two geophysical observatories (circles) used for geoelectric field modelling, and (b) an example of GIC fit from modelled geoelectric field values for a geomagnetic storm in May 2021. The solid line (purple) shows transformer neutral point current measurements that have been offset-corrected and resampled via interpolation to a 1-minute sampling rate (from 1-second). The two dashed lines show the GICs calculated from E using WIC (black) and FUR (blue) data, which are nearly identical. Note that the largest GIC values are almost always underestimated despite the otherwise good agreement between model and measurements.

one near Vienna (hereafter referred to SS1 for Substation 1) and another north of Salzburg (SS5), both with sampling rates of one second. The data was resampled to a one minute sampling rate for use in this study using a sliding window median. These two stations are of interest because they are in the high-voltage network and experience larger GICs than the other stations with measurements. As such they are useful examples for depicting the expected maximum scales of GICs that could be seen across the grid. We choose three geomagnetically active periods and use the geoelectric field components E_X and E_Y modelled from FUR data to derive the following equations:

$$GIC_{SS1} = 3.77 \cdot 10^{-2} \cdot E_x + 3.19 \cdot 10^{-2} \cdot E_y \qquad (2)$$

$$GIC_{SS5} = 0.44 \cdot 10^{-2} \cdot E_x + 5.55 \cdot 10^{-2} \cdot E_y \quad (3)$$

We see that the x-component of the geoelectric field contributes roughly the same amount to the GICs seen in SS1 as the y component. The y-component of the geoelectric field mostly dominates the currents in SS5 and contributes ten times more than the x-component. The differences in contributions from geoelectric field components stem from the varying grid layout and connections at each substation. An analysis shows that the GICs calculated from these equations are slightly more accurate than those from the full network model. Comparing to measurements at SS1, the Pearson's correlation coefficients for both GICs from the network model and GICs from Eq. 1 are 0.86, while at SS5 the correlation improves from 0.85 to 0.88. In both cases the amplitudes of the GICs are better matched and the root-mean-square-errors drop from 0.24 to 0.12 A at SS1 and 0.46 to 0.12 A at SS5. These measures were calculated from a fit of the geoelectric field data to measurements using eight days of geomagnetically active periods (including the September 2017 storm). This includes the most recent active period, meaning the measurements should represent the current grid configuration and we exclude fitting only to grid noise by using a geomagnetically active period. A fit applied to the geoelectric field modelled from WIC rather than FUR data produces slightly different coefficients but results in the same level of accuracy when compared to GIC measurements. An example of the measurements and GIC fits can be seen in **Fig. 1b**.

Regardless of which time range the fit is applied to, the GICs calculated using Eq. 1 (as well as those from the network model) tend to underestimate the peaks of the largest GICs by up to a factor of two (see e.g. **Fig 1b**, 12:20 or 13:05 UTC). We assume this is a result of attenuation of the modelled geoelectric field due to the lower sampling rate used for field modelling (Grawe et al., 2018) or the oversimplification of using a uniform geoelectric field and 1D model of the subsurface resistivity (Ngwira et al., 2015; Sun & Balch, 2019; Weigel, 2017). Despite this, the very good agreement between model and measurements means that any results based on the modelled geoelectric fields will still be reasonable.

In addition to the absolute GIC values, we also look at the cumulative absolute GICs over an hour, GIC_{sum1h}. GIC_{sum1h} is taken as the sum of minute values over the hour as a separate indicator for geomagnetic activity, more representative of sustained GICs than large spikes, both of which can have different (but similarly detrimental) effects on transformers (Bolduc, 2002; Gaunt & Coetzee, 2007). Using the accumulated sum of GICs or geoelectric field has seen usage in other studies, although not often - Lotz and Danskin (2017) used the accumulated E over varying periods and Viljanen et al. (2014) also worked with daily GIC sum averaged across nodes. The scale of GIC_{sum1h} will vary depending on the sampling rate of the data used, but in the case of minute data in Austria, 0 to 50 Ah can be seen during quiet times, and values above that generally represent more active times.

2.4 Distribution of values

In order to determine how best to forecast GICs, we first look at the 26 years of available data and the distributions of both geomagnetic variations and modelled GICs. **Figure 2a** presents the distribution of FUR minute dBx/dt and dBy/dt variations. There are very few values populating the tail of the distribution where the largest values are found. High values for this region are at 80 nT/min and upwards. The largest variations occur most commonly in the x-direction (leading to larger $\mathrm{E_Y}$) rather than the y-direction, implying that stations in the power grid sitting on east-west lines are already more susceptible to larger GICs.

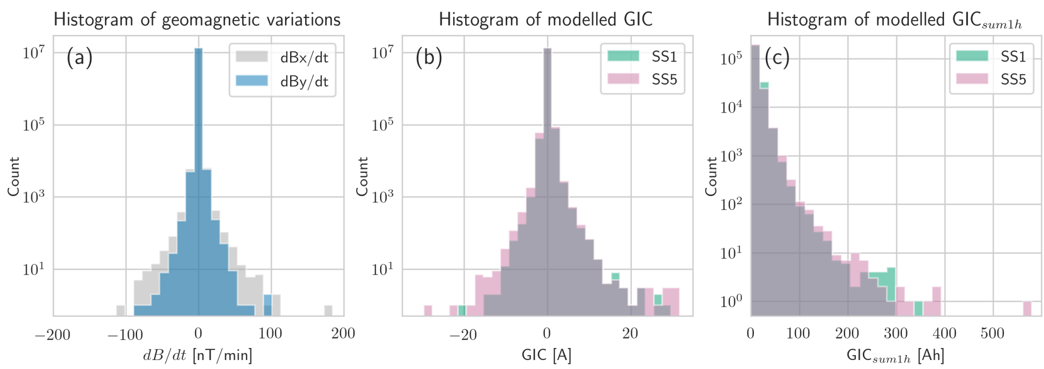

Figure 2. Histograms showing the distribution of the values in (a) the geomagnetic variations at FUR, (b) the GICs modelled from dB/dt at two substations, and (c) the hourly cumulative modelled GICs at two substations for all data, GIC$_{sum1h}$. The y-axes have logarithmic scales.

In **Figs. 2b and 2c**, the GICs observed at SS5 are larger than those at SS1. While the size of the currents depends largely on the network topology and grounding resistance, we noted in Section 2.3 that the currents at SS5 are mostly determined by the y-component of the geoelectric field (or x-component of the geomagnetic field variations), which generally sees larger variations.

2.5 Most active days

In **Table 1**, the 10 most active days in the 26 years of data according to different measures of activity dBx/dt and dBy/dt at FUR, modelled |GIC| and GIC$_{sum1h}$ at both SS1 and SS5 are listed. There are many overlapping days between the different measures, making a total of 19 days. Bold font highlights the ten largest values in each column.

A similar table for largest GIC days in Central Europe was produced in Viljanen et al. (2014, Table 4), and we see that the tables are very much in agreement with 17 shared dates, even though the table in Viljanen et al. (2014) is only based on one variable. They used a value akin to the GIC$_{sum1h}$ used here, namely the daily sum of GICs averaged across all nodes. Similarly, 17 of the days listed here also appear in Juusola et al. (2015), Table 3, where an analysis of the days with largest GICs was carried out for Northern Europe. Other larger storms that have occurred since those studies (March 2015 and September 2017) do not stand out in comparison to those from the last solar cycle with the exception of the storm from June 2015.

The largest values in each measure are clearly centered around the 2003 Halloween storm. Large values in dBx/dt tend to go alongside large GIC values in SS5, and days with large GIC$_{sum1h}$ usually coincide with days with larger |GIC|, as expected. Some exceptions are 2000-09-17, 2001-04-08, 2005-01-07 and 2005-08-24, which only show high cumulative GICs but do not stand out in dB/dt-values and peak GICs. A comparison of these events shows they have large and unidirectional geomagnetic field variations (with total field changes of 100 to 300 nT) that occur over an hour or more. These in particular lead to sustained GICs in stations susceptible to geomagnetic field changes in that direction. The variations on 2000-09-17 are shown as an example of this kind of behaviour in **Fig. 3**. Although not extremely geomagnetically active, they show that power grid transformers would have been subjected to large amounts of cumulative GICs sustained over an hour at least.

Table 1. Table showing the ten most active days according to the maximum values in three measures: leftmost are the horizontal geomagnetic field variations (dBx/dt and dBy/dt), in the centre the absolute GICs (|GIC|) at two different transformer stations (SS1 and SS5), and rightmost the cumulative GICs over an hour at two transformer stations (GIC$_{sum1h}$). Bold font highlights the ten largest values seen in that measure. The largest values are seen during the Halloween Storm on 2003 October 29-31 (italicised).

Date	dBx/dt [nT/min]	dBy/dt [nT/min]	|GIC1| [A]	|GIC5| [A]	GIC1$_{sum1h}$ [Ah]	GIC5$_{sum1h}$ [Ah]
1998-05-04	52.0	**46.0**	**11.10**	9.24	139.5	165.3
2000-04-06	42.9	**43.7**	9.47	11.00	176.7	192.5
2000-07-15	**184.7**	28.5	**17.67**	**28.39**	**265.2**	**364.5**
2000-09-17	34.5	19.9	10.19	8.89	**238.9**	**252.0**
2001-03-31	**82.4**	**40.7**	9.46	**16.69**	216.7	190.8
2001-11-06	**85.1**	**38.1**	**11.82**	**13.66**	**226.6**	**292.4**
2001-11-24	**62.4**	33.3	**12.79**	**17.42**	**262.0**	**251.6**
2003-10-29	**102.9**	**92.3**	**27.76**	**31.02**	**330.0**	**534.1**
2003-10-30	33.1	**40.3**	**16.82**	**16.69**	**268.0**	**282.6**
2003-10-31	**91.5**	**56.2**	**13.55**	**16.68**	131.8	**229.3**
2003-11-20	19.8	31.4	**11.90**	10.35	**284.9**	**280.3**
2004-07-26	**78.5**	8.5	8.74	**14.71**	77.8	77.8
2004-11-07	43.0	**37.7**	7.15	8.24	158.4	157.2
2004-11-08	24.7	28.9	**10.49**	8.64	**250.8**	212.9
2004-11-09	**76.1**	**49.9**	**14.99**	13.05	**261.5**	205.8
2005-05-15	36.3	35.1	10.38	**13.70**	**230.6**	**364.4**
2005-08-24	41.6	31.9	9.48	12.89	215.9	**349.1**
2005-09-11	**60.7**	30.7	8.04	12.07	70.8	94.1
2015-06-22	**63.0**	12.8	8.30	**15.85**	138.7	198.6

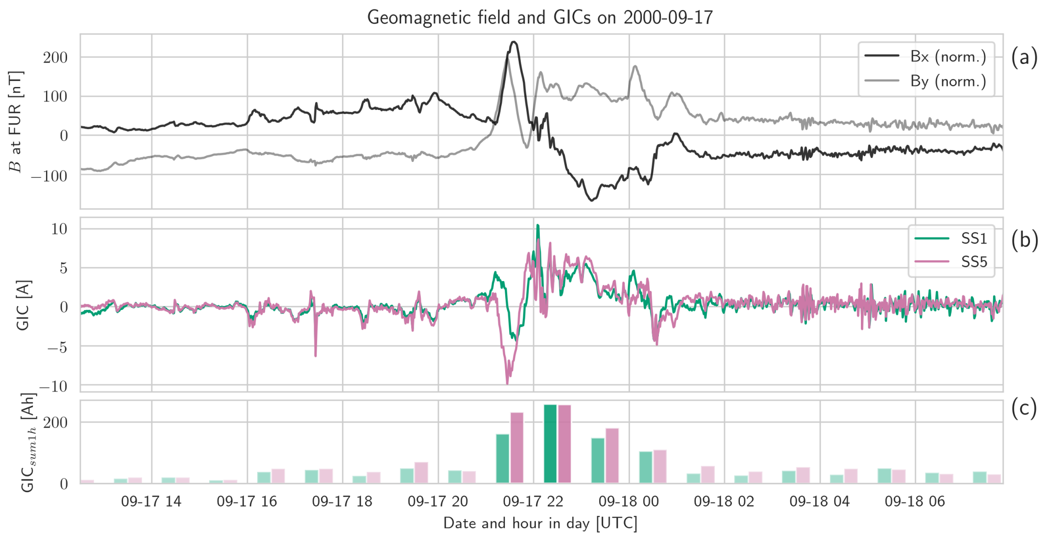

Figure 3. Plot of (a) geomagnetic variations at FUR (normalised to around zero by subtracting the mean field strength), (b) modelled GICs at two substations, and (c) cumulative hourly GICs on 2000-09-17 as an example of a day with no extreme GIC values but large cumulative hourly GICs.

2.6 Case study: 2003 Halloween Storm

In **Fig. 2**, almost all of the values in the tail end of the distribution resulted from the "Halloween storm", which lasted from 2003 October 29 to November 1. These also make up the largest values in **Table 1**, with maximum GIC values almost twice as large as the other values seen. We now conduct a detailed analysis of the behaviour during this storm and the GICs that were likely present in the power grid as an example of the problems that can arise when using only dB/dt as a proxy for GICs. We see that both large GICs and sustained GICs appear without large dB/dt values.

The geomagnetic storm that occurred at the end of October in 2003 was the result of a series of fast and geoeffective coronal mass ejections hitting the Earth during a particularly active period around the maximum of solar cycle 23 (e.g., Gopalswamy et al., 2005). In Eastwood et al. (2018), this storm was classified as a 1-in-10 year event, and is not considered an exceptionally rare example. No event of this or a higher magnitude has occurred since 2003 (with the exception of a CME directed away from Earth on July 2012, see Ngwira et al., 2013; Baker et al., 2013; Liu et al., 2014), and such events are somewhat more probable during the solar maxima (Owens et al., 2021), but have also occurred at any point throughout the solar cycle.

A brief evaluation of this storm for Austria was carried out in Bailey et al. (2018), in which a maximum GIC of 14 A was modelled. Using an updated model with newer data allows us to get a more accurate estimate of GICs during stronger events, and using the method from Section 2.3 for SS1 and SS5 we see the values reaching 25−30 A. Taking into account that the GIC peaks modelled using minute data generally underestimate the observations, these could also have reached up to 60 A.

Figure 4 compares the geomagnetic field and the modelled GICs for the 2003 Halloween Storm. Panels (a) and (b) show the geomagnetic field variations in the x and y directions. The thick lines plotted below the field show the presence of various levels of dB/dt variations (as they might be shown using a forecasting method). Light grey shows a level of 10 nT/min, and this increases going upwards to 25 nT/min, 50 nT/min and 75 nT/min. The thickness of the line shows how often the value was exceeded within a time frame of 30 minutes (with a maximum being 30 times). Panels (c) and (e) show the GICs calculated from the modelled geoelectric field at the substations SS1 and SS5, and the panels (d) and (f) show the cumulative sum of absolute GIC values (GIC_{sum1h}) over 1-hour periods.

Four time intervals, highlighted in yellow on the plot, have been picked out for discussion. Intervals 1 and 2 have been selected because, as can be seen in the high levels of dB/dt in both components, these were the most active periods. Intervals 3 and 4, in contrast, were chosen because of continuously low levels of dB/dt but lack of higher ($>$ 50 nT/min) values.

Interval 1 shows a large GIC value, which is fairly short-lived. Interval 2, in contrast, shows a consistent level of moderate GICs, though it does not reach an extremely high value. Interval 3 has a similar level of sustained GIC_{sum1h} as Interval 2 despite it having a comparatively smaller amount of dB/dt over the same period. In Interval 4, SS1 experiences the second highest value of GIC (17 A) throughout the whole storm, even though there is only continuous low-level dBx/dt and dBy/dt (10 to 25 nT/min), most of it unidirectional (comparable to the type of signal seen in **Fig. 3**). On top of that, the cumulative GICs are also some of the highest.

In summary, we see there are large differences between periods that have short-lived but large GICs (Intervals 1 and 4) and those that have longer periods of sustained GICs (Intervals 2 and 3), and both large GICs and sustained GICs can appear without large dB/dt because the ground geoelectric field responds at a range of frequencies not captured by dB/dt intensity alone. Each scenario could lead to different problems if it were

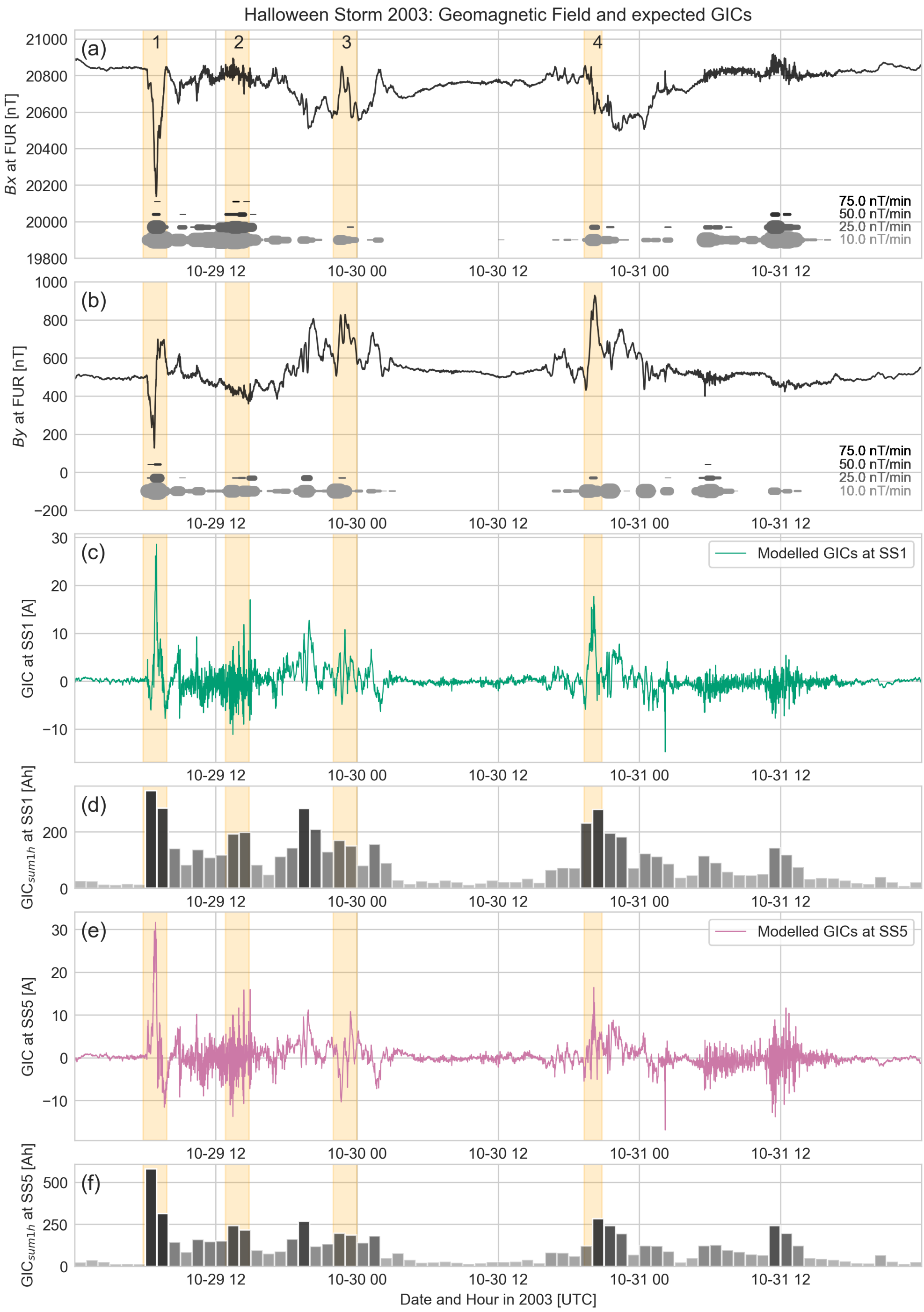

Figure 4. The Halloween storm from 2003 October 29 till 2003 November 1, during which some of the largest geomagnetic variations of the last few decades were seen. (a) and (b) show the geomagnetic variations at FUR in the x and y directions. Plotted below are levels of activity (10, 25, 50, and 75 nT/min) with line thickness showing how often these values were exceeded over a certain time range. (c) and (e) show the modelled GICs at the substations SS1 and SS5, and (d) and (f) show the cumulative GICs over each hour at each substation.

to occur in a transformer to any large degree (Price, 2002; Gaunt & Coetzee, 2007; Bolduc, 2002).

3 Building a Forecasting Model

From the analysis of past data, we deduce that, in order to forecast a comprehensive summary of expected GIC behaviour, we need to forecast either both geoelectric field components or the GICs directly. While the magnitude of the field is most important, the direction also plays an important role. From Eqs. 2 and 3, we see that a large value in E_X at SS5, for example, could be cancelled out by a smaller negative one in the E_Y value, and the opposite could be true elsewhere, making a station-by-station approach advantageous.

We now move on to build a forecasting model based on these conclusions. Three machine learning methods were put through an initial comparison for evaluation: a standard feed-forward neural network (NN) with three layers (32 neurons initially), a gradient boosting regressor based on XGBoost in Python (with 400 decision trees), and a recurrent neural network (specifically, a Long-Short-Term Memory RNN or LSTM) with three layers (32 blocks initially) and a basic Attention mechanism. sThe models were compared according to a set of metrics for model evaluation (root-mean-square error, Pearson's correlation coefficient, probability of detection). From these first comparisons, the LSTM with Attention showed the most promise and was developed into the final model, although due to the myriad machine learning methods available these days there may well be other approaches equally suited for this task.

3.1 Data preparation

The input to the machine learning model is solar wind data measured at L1 and forward-propagated to the bow shock. This means that, assuming we take measurements from satellites situated at L1, we have a varying forecast lead time between 15 and 60 minutes depending on the solar wind speed. The high resolution OMNI data set (see section on Data Availability for details) was used for solar wind measurements (speed, density, and magnetic field components) at a minute cadence combined with the local time and day in year to make up the features, while the model target was either the geoelectric field (E) modelled from FUR data or the GICs modelled from the E_X and E_Y components.

Taking solar wind measurements that have already been propagated forward to the bow shock, we use the two hours prior to the time we wish to forecast as input. This goes from $t - 120$ minutes to $t - 0$, where t is the forecast time. The range of 120 minutes for past data was decided on through experimentation, where the period was increased until longer periods did not lead to any improvements in the forecasting skill. To reduce the size and complexity of the input data, it is subsampled to a 10-minute resolution by picking every 10th point (rather than interpolation and/or fitting, which we found led to a loss in forecast skill), resulting in sequences of length 12. These sequences are used as input to forecast the maximum value of E or GICs over 40 minutes from $t - 10$ to $t+30$. This step of ten minutes into the "past" (which reduces the lead time by ten minutes) is to account for possible timing errors in propagating the solar wind forward to the bow shock.

Sampling the modelled geoelectric field or GIC data to produce a balanced data set for model training is challenging because there is a clear bias towards quiet times and not enough data from geomagnetically active times (with a factor of roughly $10^7 : 1$ for quiet to active). An initial approach using the entire data set led to a trained model that predicted only quiet times, which could not be remedied without additional data handling or large changes to the training methods. The target data set was therefore selectively sampled to reduce the imbalance. The distribution of samples was undersampled in the range of E = 0 to 100 mV/km (GIC = 0 to 8 A). Above that, we applied some data augmentation by duplicating the samples by 2 to 5 times and applying a random offset in time to the input data of each to avoid identical samples. The offset was randomly sampled without replacement from values between -10 and +10 minutes, which shifts the input solar wind data that the model sees, and means that the maximum value was either closer to the start or the end of the following 40-minute forecast window. Otherwise, all samples had a minimum time difference of 60 minutes between them. The resulting distribution is close to a one-sided Gaussian distribution. Roughly the same number of samples (9000) were used in training for each target.

The samples were split into training and testing sets by time. The years 2000, 2001 were reserved for validation to aid in model selection during training, while 2017, 2019 and 2020 were reserved for testing, and the remaining 21 years were used in training. The presence of data gaps longer than 15 consecutive minutes in the OMNI data set led to samples being excluded from the analysis - this led to 8 to 15% sample exclusion, depending on the years used. Data gaps shorter than 15 minutes were linearly interpolated over.

We reduced all values of E $>$ 200 mV/km (GIC $>$ 15 A) to 200 mV/km (15 A) because the larger values were only present in roughly 100 of the 13.7 million time-steps (or five to seven events in the 25-year period) and heavily skewed the distribution, in which all values were scaled between 0 and 1. Rescaling points above this limit greatly improved the level to which the model could learn the problem but also means that the maximum forecast the model can realistically produce is for 200 mV/km.

3.2 Training the LSTM

To approach this forecasting problem, we use a four-layer LSTM with an Attention layer. The Attention mechanism is meant to simulate human attention (first developed in Bahdanau et al., 2015), which can be understood intuitively as a mechanism that picks out the most important part of a sequence and discards the parts that are considered irrelevant. It is a tool now commonly applied in natural language processing for example (Galassi et al., 2020). The model is structured so that the input first goes through an LSTM layer and then through the Attention mechanism. The data is then fed into another LSTM layer before going through a final feed-forward layer to reduce the output to a single value.

For geoelectric field prediction, the LSTM branches into two: the left side deals with a regression problem, namely forecasting the maximum magnitude of the geoelectric field. We chose a custom loss function for the regression problem where events (peaks) are rare in the data, and where the scale of the peaks is important. A min-max scaling factor used as a penalty term meant that training to match the peak value would drive the loss down. The right side of the LSTM forecasts the sign of the geoelectric field in a classification problem, which in this case is the sign of the maximum field value used for the regression problem. Here, the binary cross-entropy loss function was used. Training worked better when the two were trained as separate targets, rather than attempting to forecast E without taking the absolute value first. The regression problem appears to be not too difficult a task, but the model had far more problems trying to forecast the direction. In training, the weights of the two problems are, when scaled, about 15 : 1 for regression to classification. The classification problem to determine the sign is given secondary importance because even an LSTM dedicated to this problem had trouble achieving a good level of accuracy. A diagram of the different LSTM architectures, the loss functions and the hyperparameters used for the training of each model can be found in the supporting information. Iteration through the various possible hyperparameters was carried out for all four models for optimisation. Similar sets of hyperparameters were found for each LSTM application, with some minor differences between them, although the choice of the same hyperparameters for all applications also led to reasonable models in all cases. Regularisation was applied in the form of dropout.

Multiple models were trained to evaluate the best approach for forecasting GICs. Those trained to forecast the geoelectric field components are referred to as LSTM-E, while nets trained to forecast the GICs directly are referred to as LSTM-GIC. Both neural nets are only trained on the output of geophysical models (in the case of E, the result of FUR variations put through the plane-wave model, and for GICs, these are the currents calculated in power grid transformers from E) because we don't have measurements of E or GIC over long enough periods and because, as described in Sec. 2.3, GICs from geophysical models reach a good enough accuracy to be a reasonable substitute in training. Both models predict the absolute value of the target, but the LSTM-E predicts the sign (positive or negative) in addition.

3.3 Evaluating the model skill

Each model was trained on its respective training set and the best LSTM parameters were chosen based on model behaviour when presented with the validation set. Following training, we ran the model on the test data set in a virtual 'real-time mode' providing updates to the input data every 15 minutes, and giving an output with a 15-minute cadence. The comparison to the ground truth (either the modelled geoelectric field or measured GICs) is performed point-to-point as well as by looking at events, where the event-based analysis is given the most importance. In order to have a benchmark for comparison, we produced a real-time persistence approach which takes the maximum of the geoelectric field or GICs in the 20 minutes before the solar wind measurement time to forecast the maximum when the solar wind would reach Earth. As such, the persistence model (PERS) also uses a varying forecast lead time. The machine-learning forecast model should be able to beat persistence in most measures.

Our event-based analysis follows the recommendations put forward by Pulkkinen et al. (2013) and Welling et al. (2018) for dB/dt forecasting. An "event" in the data is classified as a value that exceeds a certain threshold, while all values below that threshold are non-events. By defining a threshold, we can calculate the confusion matrix (Wilks, 2011), which includes the number of correctly-predicted events or true positives (TP), missed events or false negatives (FN), incorrectly-predicted events or false positives (FP), and the correctly-predicted non-events or true negatives (TN). The metrics proposed in Pulkkinen et al. (2013) include the Probability of Detection (POD), which is the fraction of measured events correctly predicted as events, also called the true positive rate (TPR or TP/(TP+FN)). Similarly, we include the probability of False Detection (POFD), the fraction of measured non-events incorrectly predicted as events, which is equivalent to the false positive rate (FPR or FP/(FP+TN)). In addition, the Heidke Skill Score (HSS) and True Skill Statistic (TSS) are also considered, both of which are derived from all variables in the confusion matrix (see e.g. Heidke, 1926; Bloomfield et al., 2012). Both the HSS and TSS show no model skill at 0, and better model skill when approaching 1. The TSS has the benefit over the HSS of being unbiased by event/non-event ratios. We also include the bias (BS), which shows if the model tends to over-predict (more false positives, BS $>$ 1) or under-predict (more false negatives, BS $<$ 1).

4 Results

We present the results split in two parts: in the first part, we test our model's forecasting ability with regards to the the geoelectric field components. The results are compared to the geoelectric field modelled from geomagnetic variations at FUR (see Sec. 2.2). In the second part, we test the forecasting ability for GICs. These are calculated using (1) the geoelectric field components predicted from LSTM-E to calculate the GICs at the two substations we picked for analysis, and (2) directly from LSTM-GIC for each substation. The comparison between the model results and measurements of GICs is carried out for the years 2017, 2019 and 2020.

For the evaluation of geoelectric field forecast, we compute the scores for three event thresholds: these are 30, 60, and 90 mV/km in both E_X and E_Y. In GICs, the level of 60 mV/km corresponds to a current of roughly 4 A through either SS1 or SS5, and we use similar thresholds of 2, 4 and 6 A. It is difficult to determine the minimum level of GICs above which transformers may experience adverse effects because these are heavily dependent on transformer type and the presence of DC-handling mechanisms. We have too few measurements of GICs exceeding higher levels such as 10 A to make an analysis at this level useful, but 4 A is crossed often during geomagnetically active times. The results are described in the next section.

Figure 5 gives a graphical representation of the model behaviour at each threshold using receiver-operator characteristic (ROC) and detection-error tradeoff (DET) curves. Both depict the model's ability to forecast events at varying thresholds. The ROC curve shows the trade-off between the true positive rate (also POD) and false positive rate (also POFD) at different event thresholds. Usually, when the threshold is low, the TPR is high but we also see an increased FPR, which is unwanted - a model that captures the observed behaviour shows a curve that keeps close to the upper left corner. The area-under-the-curve (AUC in the legend) shows good model skill as it approaches 1. On the other hand, the DET curve shows the relationship between the false negative rate (fraction of all predicted non-events that were measured events misclassified as non-events, or FN/(FN+TN)) and false positive rate, the number of which usually goes up as the other goes down depending on where the threshold for an event is set. Here, the best model behaviour is seen as the curves approach the lower left corner. It is useful in error minimisation to deduce the rate at which the FNR improves with regards to an increase in FPR rate (and vice-versa).

4.1 Forecasting E_X and E_Y

We first evaluate the LSTMs trained on the geoelectric field in terms of the root-mean-square-error (RMSE) and the Pearson's correlation coefficient (PCC). Comparing the LSTM-E outputs to modelled E, the RMSE values are 126 mV/km and 111 mV/km for the absolute value of E_X and E_Y, while the PCC values are 0.60 and 0.61. Once the sign of E has been included, the RMSE rises to 261 mV/km and 287 mV/km, while PCC drops to 0.48 and 0.32, so we see that the model's inability to forecast the field direction reliably decreases the accuracy when also considering the field direction.

Table 2 presents an event-based analysis of the LSTM-E results. Multiple thresholds (TH) defining events were considered, and these are listed by the variable "TH" in each line (at 30, 60, and 90 mV/km, representing minor, moderate and strong geomagnetic activity). We see that the skill decreases as the threshold increases (decreasing probability of detection POD and TSS), and that the LSTMs tend towards over-predicting (BS > 1). (The bias for the PERS models is always ~ 1 because the time series being compared are only shifted in time and therefore almost statistically equivalent.) There are always a large number of false positives, although this remains a small fraction of the number of total data points. The LSTM-E models generally outperform the PERS approach, although the Heidke Skill Scores are occasionally smaller in the LSTMs, which implies a worse balance between false positives and true positives. As in the point-to-point values, the E_X component tends to be predicted better than the E_Y component. By evaluating the ROC and DET curves in Fig. 5 (a-b), we see that the LSTM-E models outperforms persistence at all thresholds.

We also conducted a comparison with the results from Honkonen et al. (2018) and Lotz and Danskin (2017), where possible. While the time development of the geoelectric field appears better in the modelling approach in Honkonen et al. (2018), the mag-

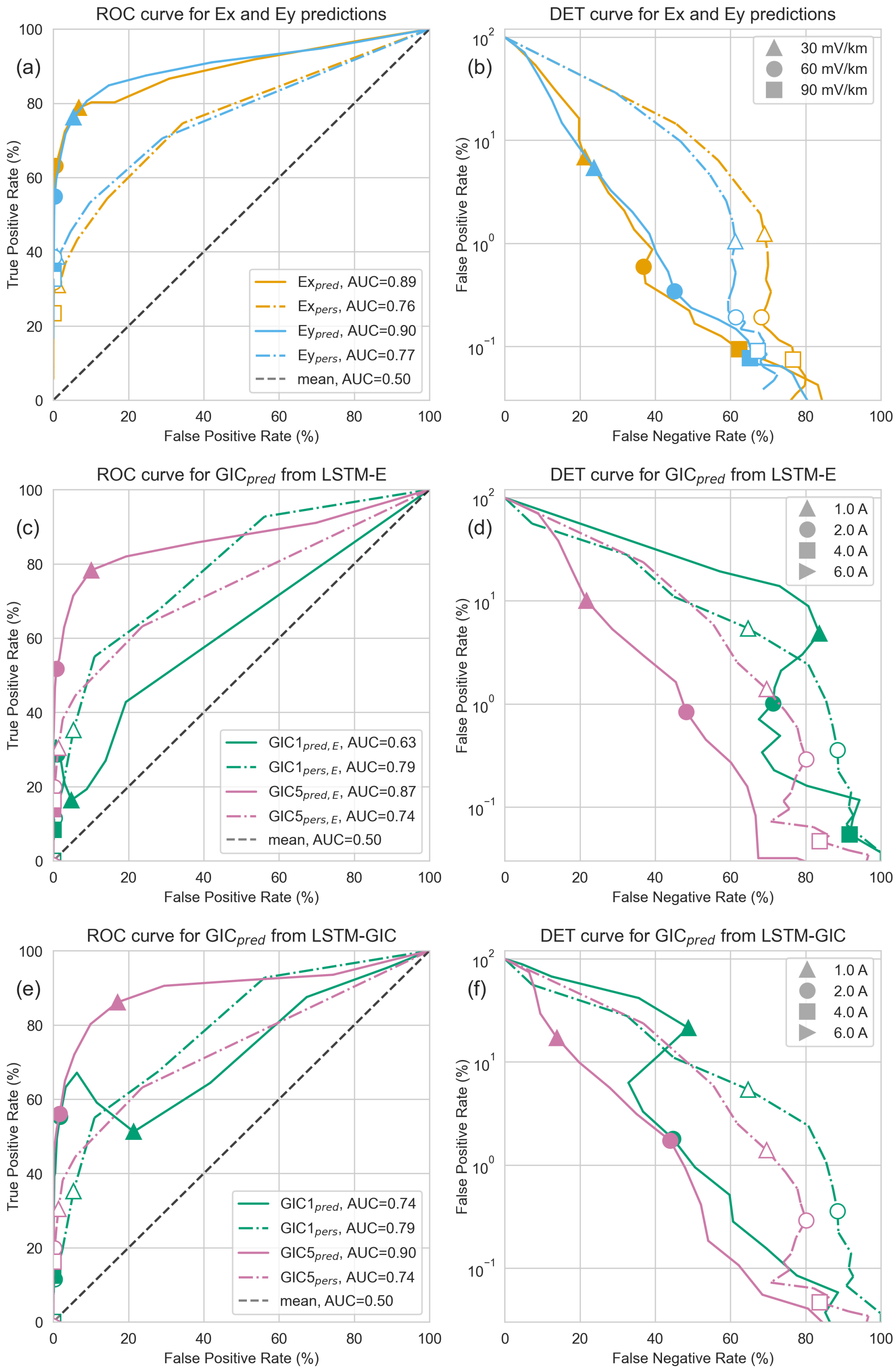

Figure 5. Receiver-operator characteristic (ROC) and detection-error tradeoff (DET) curves for three approaches: (a-b) the geoelectric field, showing the output from the LSTM-E models vs the modelled geoelectric field, (c-d) the GICs calculated from the geoelectric field predicted by LSTM-E compared to measured GICs, and (e-f) the GICs predicted by the LSTM-GIC models compared to measured GICs. SS1 and SS5 are two separate substations in the power grid from which we have measurements. The values for specific event thresholds are labelled with shapes as defined in each legend.

Table 2. Metrics from an event-based analysis of the LSTM-E models applied to the years 2000, 2001, 2017, 2019 and 2020 in a retrospective real-time mode with the model being run at 15-minute intervals. A persistence model (PERS) is included for comparison. The first four columns provide the values for the confusion matrix (where TP, FP, TN and FN are the true positives (hits), false positives, true negatives (misses) and false negatives), the probability of detection (POD), probability of false detection (POFD), Heidke Skill Score (HSS), True Skill Score (TSS), and bias (BS). The variable TH in brackets gives the event threshold used to define events and compute the metrics.

LSTM-E Model	$N_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
$E_{X,pred}$(TH=30)	3092	2436	11749	656	160506	78.8	6.8	0.26	0.72	4.6
$E_{X,pred}$(TH=60)	494	312	1038	182	173815	63.2	0.6	0.34	0.63	2.7
$E_{X,pred}$(TH=90)	175	66	164	109	175008	37.7	0.1	0.33	0.38	1.3
$E_{Y,pred}$(TH=30)	2989	2279	9328	710	163030	76.2	5.4	0.29	0.71	3.9
$E_{Y,pred}$(TH=60)	559	307	600	252	174188	54.9	0.3	0.42	0.55	1.6
$E_{Y,pred}$(TH=90)	241	84	135	157	174971	34.9	0.1	0.36	0.35	0.9
PERS Model	$N_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
$E_{X,pers}$(TH=30)	3092	958	2128	2134	170127	31.0	1.2	0.30	0.30	1.0
$E_{X,pers}$(TH=60)	494	157	335	337	174518	31.8	0.2	0.32	0.32	1.0
$E_{X,pers}$(TH=90)	175	41	130	134	175042	23.4	0.1	0.24	0.23	1.0
$E_{Y,pers}$(TH=30)	2989	1156	1804	1833	170554	38.7	1.0	0.38	0.38	1.0
$E_{Y,pers}$(TH=60)	559	216	335	343	174453	38.6	0.2	0.39	0.38	1.0
$E_{Y,pers}$(TH=90)	241	79	158	162	174948	32.8	0.1	0.33	0.33	1.0

nitudes are not matched as well. An event-based analysis could not be carried out in their case due to the short time series and lack of larger events, but the RMSE and PCC values for E_X and E_Y (reduced to a 15-min sampling rate) come out as 10.5 mV/km and 97.8 mV/km and 0.62 and 0.25, respectively, which is better in the case of E_X but worse in the case of E_Y. Comparing to Lotz and Danskin (2017), we see similar correlations for the geoelectric field components. They found a slightly higher correlation (averaged over three stations and two storms, 0.71 for E_X and 0.53 for E_Y), although they predicted the maximum value for a longer time span (90 mins), making their approach closer to a nowcast than a forecast. The higher RMSE values seen in our study in part derive from the slightly higher levels of daily variation that is forecast even when the field is extremely quiet. Again, in both studies used as comparison we see the northward component of the geoelectric field was predicted better than the eastward component.

4.2 Forecasting GICs

The same results are presented for GICs as for the geoelectric field components in the last section. In the event-based analysis, the thresholds were set at 2, 4 and 6 A, which are roughly equivalent to the thresholds used for the electric field. **Table 3** shows the results of this analysis applied to the test data set years 2017, 2019 and 2020, while **Fig. 5** depicts the ROC and DET curves for the model output versus measured GICs. A comparison between the LSTM-GIC output and the modelled GICs the model was trained on shows similar levels of accuracy as in LSTM-E to the geoelectric field.

We first look at the results for GICs calculated from the geoelectric field components predicted using the LSTM-E models. Note that while the last section mainly looked at the absolute value of the geoelectric fields, in the calculation of GICs the direction of

Table 3. Metrics from an event-based analysis of different model applied to the years 2017, 2019 and 2020 in a retrospective real-time mode with the model being run at 15-minute intervals. GIC1$_{pred,E}$ is the result from the models trained to predict the geoelectric field (LSTM-E), while GIC1$_{pred}$ is the result from the LSTM-GIC. PERS is a persistence model assuming the target (GIC) repeats itself. The first four columns provide the values for the confusion matrix (where TP, FP, TN and FN are the true positives (hits), false positives, true negatives (misses) and false negatives), the probability of detection (POD), probability of false detection (POFD), Heidke Skill Score (HSS), True Skill Score (TSS), and bias (BS). The variable TH in brackets is the event threshold used to define events and compute the metrics."undef." refers to the HSS and TSS at TP=0, which are undefined.

LSTM-E Model	N$_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
GIC1$_{pred,E}$(TH=2)	432	124	1060	308	103697	28.7	1.0	0.15	0.28	2.7
GIC1$_{pred,E}$(TH=4)	24	2	57	22	105108	8.3	0.1	0.05	0.08	2.5
GIC5$_{pred,E}$(TH=2)	307	159	681	148	80649	51.8	0.8	0.27	0.51	2.7
GIC5$_{pred,E}$(TH=4)	43	6	13	37	81581	14.0	0.0	0.19	0.14	0.4
LSTM-GIC Model	N$_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
GIC1$_{pred}$(TH=2)	432	239	1886	193	102871	55.3	1.8	0.18	0.54	4.9
GIC1$_{pred}$(TH=4)	24	3	26	21	105139	12.5	0.0	0.11	0.12	1.2
GIC5$_{pred}$(TH=2)	307	172	1403	135	79927	56.0	1.7	0.18	0.54	5.1
GIC5$_{pred}$(TH=4)	43	7	16	36	81578	16.3	0.0	0.21	0.16	0.5
PERS Model	N$_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
GIC1$_{pers}$(TH=2)	432	50	375	382	104382	11.6	0.4	0.11	0.11	1.0
GIC1$_{pers}$(TH=4)	24	0	26	24	105139	0.0	0.0	undef.	undef.	1.1
GIC5$_{pers}$(TH=2)	307	61	237	246	81093	19.9	0.3	0.20	0.20	1.0
GIC5$_{pers}$(TH=4)	43	7	38	36	81556	16.3	0.0	0.16	0.16	1.0

the geoelectric field is also included, making this an additional error factor if the sign is not predicted accurately. Once the GICs have been calculated using the results from the LSTM-E models and Eq. 1, the absolute value is taken for the rest of the analysis.

As can be seen in **Table 3**, the GICs derived from the LSTM-E models see a considerable drop in accuracy in comparison to the results for E alone in Table 2. Although there were quite reasonable values for POD predicting E, the POD for GICs at the mid-range threshold (60 mV/km or 4 A) drops from around 50% in both components of E to 8% and 16% in substation SS1 and SS5. Evaluating the skill of the model for GICs at high levels is difficult because there are so few events exceeding even a minimal value of 6 A. None of these events (2 at SS1, 12 at SS5 over the three years of data) were predicted using any approach.

In comparing the GIC predictions from the two methods (LSTM-E and LSTM-GIC), we see that the LSTM-GIC seems to perform better but the results are station-specific. The LSTM-GIC performs much better than the LSTM-E at SS1 (e.g. a POD of 55% rather than 29% and higher HSS and TSS values at a threshold of 2 A) and at a similar level at SS5. This is also reflected in a model evaluation using point-to-point metrics. The RMSE values for SS1 and SS5 predicted using LSTM-E are 0.49 A and 0.59 A, while the PCC is 0.35 and 0.67. For GICs predicted using LSTM-GIC, the RMSE values are 0.67 A and 0.78 A (i.e. slightly worse than LSTM-E), but the PCC is 0.56 and 0.64. The accuracy between the two approaches is roughly equivalent for SS5, but using LSTM-GIC rather

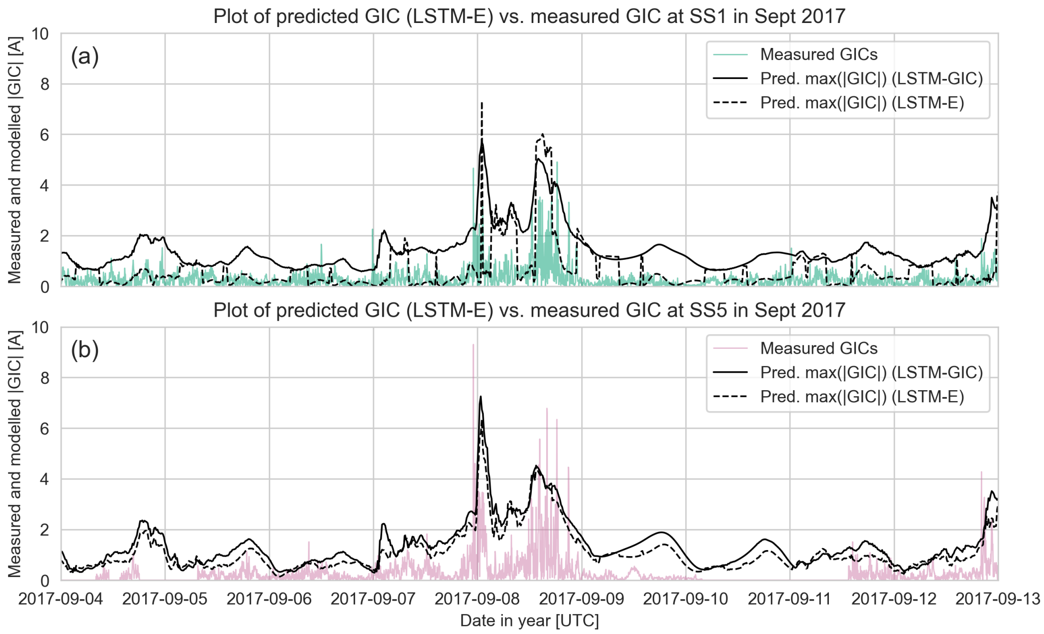

Figure 6. The LSTM-E (dashed line) and LSTM-GIC (solid line) applied to forecasts in an experimental real-time mode and compared to measurements of GICs (coloured lines) at two stations in Austria. The upper panel (a) shows results for SS1 near Vienna, while the lower panel (b) shows results for SS5 near Salzburg (with some data gaps). Although not plotted here, the maximum GIC value computed from the measurements is at the same cadence of 15 minutes to compare to the model forecasts.

than LSTM-E is a definite improvement for SS1 observations. Some of the reason for this can be seen in **Fig. 6**. In SS1, the jumps in values computed from LSTM-E result from changes in the sign of the geoelectric field components, which then cancel each other out and lead to a GIC of zero. (Conversely, ignoring the sign from LSTM-E and taking the absolute values to calculate the GICs in SS1 results in higher correlation and POD but a far larger number of false positives, leaving this as another possibility.) In the best cases, the GIC forecasts only reach a POD of 16% for GICs above a threshold of 4 A, highlighting the difficulty in correctly predicting larger values.

In the ROC and DET curves in **Figure 5** panels (c-d) for GICs from LSTM-E and (e-f) from LSTM-GIC, we also see some of the weak forecasting ability for SS1 primarily represents the LSTM behaviour at low values (GICs < 1 A). At SS1, there is a mostly continuous level of noise around 1 A, and the model does not predict the noise while the persistence model captures it clearly. This is an example of the weakness of ROC curves, where in this case only the lower left corner (showing values greater than 1 A) is of interest to us.

Figure 6 shows the forecast that would have been produced by the model (solid and black dashed lines) against measurements (coloured lines) during the September 2017 storm. The models, particularly the LSTM-GIC approach, do a reasonable job at predicting magnitudes, although the LSTM-E struggles to predict the direction, which is also important for accurate GIC prediction. The storm and the active periods are clearly captured by the forecast, and daily variations from the Sq current are forecasted otherwise. Note that the delayed rise in the forecast of the first peak of the storm does not indicate a timing error. A cross-correlation of the model output shows at maximum an offset in time of 10 minutes and the delay in the figure is simply a feature unique to this storm. While the exact time development of the storm is not captured well, the general scales of GICs are matched well, as is the differentiation between quiet and active times.

In summary, prediction of geoelectric field magnitudes can be achieved with reasonable accuracy (POD of at least 35% even at the highest event threshold), but the prediction of elevated levels of GICs proves difficult with any approach used. The LSTMs usually outperform the persistence models, except in the bias, where the persistence model has the benefit of being statistically equivalent to the data it is being compared to. The persistence model also generally has a lower POFD and higher HSS value at low thresholds (e.g. TH=30 V/km for LSTM-E) because quiet periods tend to persist over time. The LSTMs, however, outperform persistence at the higher thresholds, which are more important for forecasting purposes.

5 Discussion

We have attempted to forecast GICs from solar wind data using LSTMs with two different approaches. We now look at some of the reasons behind the particular difficulty in forecasting GICs.

Some of the low skill seen when comparing predictions to GIC measurements is down to four reasons, mostly related to our GIC data: firstly, there is noise in the GIC observations, particularly at SS1, which has a consistent level of 1 A noise during the day - this is not predicted by the model. Secondly, GIC observations until 2021 had a maximum cutoff point of 3.4 A in the positive direction, removing some peaks from our event list, and these have not been accounted for. Thirdly, the model struggles to predict the direction of the geoelectric field values, which are likely driven by smaller-scale ionospheric currents (Dimmock et al., 2020). Fourthly, as noted in Sec. 2.3, the peaks of observed GICs are often underestimated by geophysical modelling, meaning peaks in the GIC measurements after the cut-off level was removed were often much larger than modelled. This is a problem related to the geoelectric field modelling that may affect the LSTM's ability to learn the problem due to insufficient accuracy in the field modelling. While minute cadence data does capture most of the variability in the GICs, the lack of higher frequency content appears to the primary cause of underestimated peaks, a problem discussed before in Grawe et al. (2018) and recently for the specific problem of GIC estimates in Beggan et al. (2021). As such, it is not surprising that the LSTMs tend to underestimate the actual GICs, and a correction would have to be applied to the target data to account for this.

Outside of the data-specific problems, there are also some timing errors, meaning some peaks arrived slightly later or earlier than they were observed, and as such are not logged as correct predictions even though an event threshold was crossed.

In an application of the model in operations, one caveat is that the maximum possible forecast is 200 mV/km due to a self-imposed limit to improve the model's ability to learn. We assume that in practise, this would be negligible because all values above a certain level (e.g. 100 mV/km) would be of interest, regardless of how large they become. As also discussed in Wintoft et al. (2016), the scale of geomagnetic variations during extreme events can theoretically become so large that it is effectively unbounded for the purpose of this discussion. In the future, this 200 mV/km limit could be improved on by training a model specifically for large value forecasting, which can be switched to if the original model forecasts $E > 150$ mV/km.

In an ideal case, a forecasting model would be developed while taking a cost-loss analysis (Murphy, 1977) such as that used in a space weather context in Owens et al. (2014) into consideration. In the case of network protection, this is a very complex scenario due to the varying impacts and costs associated with transformer damage or power grid outage, many of which are currently nearly impossible to estimate. This is something that can hopefully be developed further as studies into GIC risk progress (Eastwood et al., 2018).

Another, more general problem in forecasting any measure of ground geomagnetic activity from solar wind measurements without further input from magnetospheric modelling is that not all geomagnetic variations are driven by the solar wind directly (see e.g. Kamide et al., 1998; Eastwood et al., 2015). Many variations will result from reconnection in the magnetotail and chaotic processes and would not be relatable through our model, which is essentially a coupling function from the solar wind at the bow shock to the geoelectric field in Austria. Although the machine learning approach described here works at a basic level and could be more promising than forecasts of dB/dt alone, it would need to be coupled with either data from space-borne monitors observing the Earth's magnetosphere, more complex physical models of magnetospheric behaviour, or both to escape this limitation.

The calculations and measurements of the GICs shown in this study are for a specific grid configuration, even though the power grid is continually being upgraded and changed. These changes can have large effects on individual GIC scales over long time ranges. The results shown in Table 1 extend far into the past, for which we do not have a detailed history of grid changes, so the values listed could have been much smaller or much larger depending on how the grid was set up. For the LSTM predictions, we have conducted our analysis with the comparison to measurements over a considerably shorter time range of a few years, where the grid has not changed to any great degree, but the predictions may not be valid in the future for a different grid configuration. In this case, a new fit would need to be found for Eq. 1, and either the LSTM-GIC model would need to be retrained on the updated GIC data, or the GIC values could be calculated anew from the otherwise unchanged LSTM-E output.

Our aim was to develop a model that can provide useful forecasts for power grid operators by providing estimates of the scales of GICs. The difference between this and former studies such as Lotz et al. (2017) and Honkonen et al. (2018), who also predicted ground geoelectric fields from solar wind data, is that we have approached the problem with a new tool (a recurrent neural network) and have been able to forecast GICs directly along with the geoelectric field, with the results compared to measured GICs. We have had some success, particularly with forecasting the geoelectric field, and have tried forecasting substation-specific GICs for the first time, but there are still many problems to be addressed to turn this method into a useful forecast.

6 Summary

We have developed a machine learning approach to forecast GICs in Austria. Using data from the past 26 years and the 2003 Halloween storm as a case study, we argued that forecasts of dB/dt alone, which have been the focus of most past studies, are not sufficient to make actionable GIC forecasts.

From this initial analysis, we set out to forecast maximum expected GICs (over a forty minute window) either directly for specific substations in the power grid or more generally from forecasts of the regional geoelectric field components. From a small set of initial machine learning approaches, an LSTM (recurrent neural network) with an Attention mechanism showed the most promise in forecasting skill and this was developed into a more complex approach.

A selection of models were trained on 21 years of geoelectric field values modelled from geomagnetic variations at the geomagnetic observatory in Fürstenfeldbruck close to Austria. In the first method, two recurrent neural networks or LSTMs were trained to predict the northward and eastward modelled geoelectric field components and compute the specific substation GICs using a linear equation. In the second method, an LSTM was trained to predict modelled GICs at two substations, which we know correlate very well with the measurements. Five years of data were reserved for testing and evaluating the model. The results were compared to DC measurements at two substations in the Austrian power grid.

The LSTM model worked with reasonable success when predicting the geoelectric field modelled from geomagnetic variations, although translating this success into good GIC forecasts proved difficult. It was possible, however, to outperform a model that simply takes the last observed GICs to forecast future values.

We conclude that forecasting the GICs observed in the power grid from solar wind data measured at L1 is a difficult task, even when the forecasting model does a reasonable job of forecasting the geoelectric field components or modelled GIC. There are many ways to improve the modelling in the future, including using higher-resolution magnetic field measurements (or applying a correction to the modelled geoelectric field before training) to more accurately estimate the peak geoelectric field and GIC values, and by including information on the development of the magnetosphere during storm times.

Although this study has looked specifically at a mid-latitude region, where geomagnetic variations and GICs are not as large as those seen in higher latitude regions such as Scandinavia, we have been able to compare model output directly to measurements and expect that the conclusions drawn will also be valid for other regions with GICs at different scales.

A lower-resolution version of the LSTM-E model will be coupled with the PREDSTORM solar wind forecast (Bailey et al., 2020), which provides forecasts of the ambient solar wind a few days in advance, based on either a recurrence model or data from a spacecraft east of the Sun-Earth line such as STEREO or a future mission to the Lagrange 5 point. We also plan in the future to integrate methods on solar wind B_z forecasting (Reiss et al., 2021) or CME flux rope modelling (Weiss et al., 2021) to advance our capabilities in GIC forecasting for any type of solar wind structures.

7 Data Availability

- INTERMAGNET data for FUR and WIC:
  `https://intermagnet.org/data-donnee/download-eng.php`
- OMNI data: `https://spdf.gsfc.nasa.gov/pub/data/omni/high_res_omni/`
- Open source code for this work (in Python 3 and Jupyter Notebook form):
  `https://github.com/bairaelyn/SOLARWIND2GIC` (**Note:** Zenodo DOI will follow for final version.)
- Exact details on the LSTM structure and hyperparameters used for training can be found in the supporting information for this study.

Acknowledgments

The results presented in this paper rely on the data collected at the Conrad Observatory (ZAMG) in Austria and at Fürstenfeldbruck (LMU), Germany. We thank the Ludwig-Maximilians-Universität München for supporting its operation and INTERMAGNET for promoting high standards of magnetic observatory practice (`www.intermagnet.org`). The data used in this study is publicly available (with the exception of the measurements of GICs in Austria), and details on where to find the data can be found in Section 7. We thank Ilja Honkonen for providing the results from his 2018 study for comparison. R.L.B., C.M., M.A.R., and A.J.W. thank the Austrian Science Fund (FWF) for research funding from projects P31659-N27 and P31521-N27. C.D.B. was funded under UK Natural Environment Research Council Grant NE/P017231/1 "Space Weather Impact on Ground-based Systems (SWIGS)". We thank the two anonymous reviewers for their careful readings of this manuscript and helpful suggestions.

References

Ádám, A., Prácser, E., & Wesztergom, V. (2012). Estimation of the electric resistivity distribution (EURHOM) in the European lithosphere in the frame of the EURISGIC WP2 project. *Acta Geodaetica et Geophysica Hungarica*, *47*(4), 377-387.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd international conference on learning representations.*

Bailey, R. L., Halbedl, T. S., Schattauer, I., Achleitner, G., & Leonhardt, R. (2018). Validating GIC models with measurements in Austria: evaluation of accuracy and sensitivity to input parameters. *Space Weather*.

Bailey, R. L., Halbedl, T. S., Schattauer, I., Römer, A., Achleitner, G., Beggan, C. D., ... Leonhardt, R. (2017). Modelling geomagnetically induced currents in midlatitude Central Europe using a thin-sheet approach. *Annales Geophysicae*, *35*(3), 751.

Bailey, R. L., Möstl, C., Reiss, M. A., Weiss, A. J., Amerstorfer, U. V., Amerstorfer, T., ... Leonhardt, R. (2020). Prediction of Dst during solar minimum using in situ measurements at l5. *Space Weather*, *18*(5), e2019SW002424.

Baker, D. N., Li, X., Pulkkinen, A., Ngwira, C. M., Mays, M. L., Galvin, A. B., & Simunac, K. D. C. (2013, October). A major solar eruptive event in July 2012: Defining extreme space weather scenarios. *Space Weather*, *11*(10), 585-591. doi: 10.1002/swe.20097

Barbosa, C., Alves, L., Caraballo, R., Hartmann, G. A., Papa, A. R., & Pirjola, R. J. (2015). Analysis of geomagnetically induced currents at a low-latitude region over the solar cycles 23 and 24: comparison between measurements and calculations. *Journal of Space Weather and Space Climate*, *5*, A35.

Beggan, C. D., Richardson, G. S., Baillie, O., Hübert, J., & Thomson, A. W. (2021). Geolectric field measurement, modelling and validation during geomagnetic storms in the uk. *Journal of Space Weather and Space Climate*, *11*, 37.

Bhaskar, A., & Vichare, G. (2019). Forecasting of SYMH and ASYH indices for geomagnetic storms of solar cycle 24 including St. Patrick´s day, 2015 storm using NARX neural network. *J. Space Weather Space Clim.*, *9*, A12.

Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. (2012, Mar). Toward Reliable Benchmarking of Solar Flare Forecasting Methods. *The Astrophysical Journal Letters*, *747*(2), L41.

Bolduc, L. (2002). GIC observations and studies in the Hydro-Québec power system. *Journal of Atmospheric and Solar-Terrestrial Physics*, *64*(16), 1793–1802.

Boteler, D. H., & Pirjola, R. J. (2017). Modeling geomagnetically induced currents. *Space Weather*, *15*(1), 258-276.

Boteler, D. H., Pirjola, R. J., & Nevanlinna, H. (1998). The effects of geomagnetic disturbances on electrical systems at the Earth's surface. *Advances in Space Research*, *22*, 17-27.

Butala, M. D., Kazerooni, M., Makela, J. J., Kamalabadi, F., Gannon, J. L., Zhu, H., & Overbye, T. J. (2017). Modeling geomagnetically induced currents from magnetometer measurements: Spatial scale assessed with reference measurements. *Space Weather*, *15*(10), 1357-1372.

Caraballo, R., González-Esparza, J. A., Sergeeva, M., & Pacheco, C. R. (2020). First GIC estimates for the Mexican power grid. *Space Weather*, *18*(2), e2019SW002260.

Chave, A. D., & Jones, A. G. (2012). *The magnetotelluric method: Theory and practice.* Cambridge University Press.

Dimmock, A. P., Rosenqvist, L., Welling, D. T., Viljanen, A., Honkonen, I., Boynton, R. J., & Yordanova, E. (2020). On the regional variability of db/dt and its significance to gic. *Space Weather*, *18*(8), e2020SW002497.
Eastwood, J., Hapgood, M., Biffis, E., Benedetti, D., Bisi, M., Green, L., ... Burnett, C. (2018). Quantifying the economic value of space weather forecasting for power grids: An exploratory study. *Space Weather*, *16*(12), 2052-2067.
Eastwood, J., Hietala, H., Toth, G., Phan, T., & Fujimoto, M. (2015). What controls the structure and dynamics of Earth's magnetosphere? *Space Science Reviews*, *188*(1), 251–286.
Galassi, A., Lippi, M., & Torroni, P. (2020). Attention in natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*.
Gaunt, C. T., & Coetzee, G. (2007). Transformer failures in regions incorrectly considered to have low GIC-risk. In *Power tech, 2007 ieee lausanne* (p. 807-812). IEEE.
Gil, A., Modzelewska, R., Moskwa, S., Siluszyk, A., Siluszyk, M., Wawrzynczak, A., & Zakrzewska, S. (2019). Does time series analysis confirms the relationship between space weather effects and the failures of electrical grids in south poland? *Journal of Mathematics in Industry*, *9*(1), 1–16.
Gopalswamy, N., Yashiro, S., Liu, Y., Michalek, G., Vourlidas, A., Kaiser, M. L., & Howard, R. A. (2005, September). Coronal mass ejections and other extreme characteristics of the 2003 October-November solar eruptions. *Journal of Geophysical Research (Space Physics)*, *110*(A9), A09S15. doi: 10.1029/2004JA010958
Grawe, M. A., Makela, J. J., Butala, M. D., & Kamalabadi, F. (2018). The impact of magnetic field temporal sampling on modeled surface electric fields. *Space Weather*, *16*(11), 1721-1739.
Heidke, P. (1926). Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst. *Geografiska Annaler*, *8*(4), 301–349.
Honkonen, I., Kuvshinov, A., Rastätter, L., & Pulkkinen, A. (2018). Predicting global ground geoelectric field with coupled geospace and three-dimensional geomagnetic induction models. *Space Weather*, *16*(8), 1028-1041.
Juusola, L., Viljanen, A., van de Kamp, M., Tanskanen, E. I., Vanhamäki, H., Partamies, N., & Kauristie, K. (2015). High-latitude ionospheric equivalent currents during strong space storms: Regional perspective. *Space Weather*, *13*(1), 49-60.
Kamide, Y., Baumjohann, W., Daglis, I. A., Gonzalez, W. D., Grande, M., Joselyn, J. A., ... Vasyliunas, V. M. (1998). Current understanding of magnetic storms: Storm-substorm relationships. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17705-17728.
Keesee, A. M., Pinto, V., Coughlan, M., Lennox, C., Mahmud, M. S., & Connor, H. K. (2020). Comparison of deep learning techniques to model connections between solar wind and ground magnetic perturbations. *Frontiers in Astronomy and Space Sciences*, *7*, 72.
Kelbert, A. (2020). The role of global/regional earth conductivity models in natural geomagnetic hazard mitigation. *Surv Geophys*, *41*, 115–166.
Liu, Y. D., Luhmann, J. G., Kajdič, P., Kilpua, E. K. J., Lugaz, N., Nitta, N. V., ... Galvin, A. B. (2014, March). Observations of an extreme storm in interplanetary space caused by successive coronal mass ejections. *Nature Communications*, *5*, 3481. doi: 10.1038/ncomms4481
Lotz, S., & Cilliers, P. (2015). A solar wind-based model of geomagnetic field fluctuations at a mid-latitude station. *Advances in Space Research*, *55*(1), 220 - 230.
Lotz, S., & Danskin, D. W. (2017). Extreme value analysis of induced geoelectric field in South Africa. *Space Weather*, *15*(10), 1347-1356.
Lotz, S., Heyns, M., & Cilliers, P. J. (2017). Regression-based forecast model of in-

duced geoelectric field. *Space Weather*, *15*(1), 180-191.

Lu, J., Peng, Y., Wang, M., Gu, S., & Zhao, M. (2016). Support Vector Machine combined with distance correlation learning for Dst forecasting during intense geomagnetic storms. *Planetary and Space Science*, *120*, 48 - 55.

Molinski, T. S. (2002). Why utilities respect geomagnetically induced currents. *Journal of atmospheric and solar-terrestrial physics*, *64*(16), 1765–1778.

Murphy, A. H. (1977). The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Review*, *105*(7), 803–816.

Ngwira, C. M., Pulkkinen, A., Leila Mays, M., Kuznetsova, M. M., Galvin, A., Simunac, K., … Glocer, A. (2013). Simulation of the 23 July 2012 extreme space weather event: What if this extremely rare CME was Earth directed? *Space Weather*, *11*(12), 671–679.

Ngwira, C. M., Pulkkinen, A. A., Bernabeu, E., Eichner, J., Viljanen, A., & Crowley, G. (2015). Characteristics of extreme geoelectric fields and their possible causes: Localized peak enhancements. *Geophysical Research Letters*, *42*(17), 6916-6921.

Owens, M. J., Horbury, T. S., Wicks, R. T., McGregor, S. L., Savani, N. P., & Xiong, M. (2014). Ensemble downscaling in coupled solar wind-magnetosphere modeling for space weather forecasting. *Space Weather*, *12*(6), 395-405.

Owens, M. J., Lockwood, M., Barnard, L. A., Scott, C. J., Haines, C., & Macneil, A. (2021, May). Extreme Space-Weather Events and the Solar Cycle. *Solar Physics*, *296*(5), 82. doi: 10.1007/s11207-021-01831-3

Price, P. R. (2002). Geomagnetically induced current effects on transformers. *IEEE Transactions on Power Delivery*, *17*(4), 1002-1008.

Pulkkinen, A., Bernabeu, E., Thomson, A., Viljanen, A., Pirjola, R., Boteler, D., … MacAlester, M. (2017). Geomagnetically induced currents: Science, engineering, and applications readiness. *Space Weather*, *15*(7), 828-856.

Pulkkinen, A., Hesse, M., Habib, S., Van der Zel, L., Damsky, B., Policelli, F., … Creamer, E. (2010). Solar shield: forecasting and mitigating space weather effects on high-voltage power transmission systems. *Natural hazards*, *53*(2), 333–345.

Pulkkinen, A., Hesse, M., Kuznetsova, M., & Rastätter, L. (2007). First-principles modeling of geomagnetically induced electromagnetic fields and currents from upstream solar wind to the surface of the Earth. *Annales Geophysicae*, *25*(4), 881–893.

Pulkkinen, A., Pirjola, R., & Viljanen, A. (2007). Determination of ground conductivity and system parameters for optimal modeling of geomagnetically induced current flow in technological systems. *Earth, planets and space*, *59*(9), 999–1006.

Pulkkinen, A., Rastatter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., … Weigel, R. (2013, June). Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations. *Space Weather*, *11*(6), 369-385.

Pulkkinen, A., Taktakishvili, A., Odstrcil, D., & Jacobs, W. (2009). Novel approach to geomagnetically induced current forecasts based on remote solar observations. *Space Weather*, *7*(8).

Pulkkinen, A., Viljanen, A., & Pirjola, R. (2006). Estimation of geomagnetically induced current levels from different input data. *Space Weather*, *4*(8).

Reiss, M. A., Möstl, C., Bailey, R. L., Rüdisser, H. T., Amerstorfer, U. V., Amerstorfer, T., … Windisch, A. (2021, August). Machine learning for predicting the Bz magnetic field component from upstream in situ observations of solar coronal mass ejections. *arXiv e-prints*, arXiv:2108.04067.

Shepherd, S. G. (2014). Altitude-adjusted corrected geomagnetic coordinates: Definition and functional approximations. *Journal of Geophysical Research: Space*

Physics, *119*(9), 7501-7521.
Sun, R., & Balch, C. (2019). Comparison between 1-D and 3-D geoelectric field methods to calculate geomagnetically induced currents: A case study. *IEEE Transactions on Power Delivery*, *34*(6), 2163-2172.
Svanda, Michal, Mourenas, Didier, Zertová, Karla, & Výbost´oková, Tatiana. (2020). Immediate and delayed responses of power lines and transformers in the Czech electric power grid to geomagnetic storms. *J. Space Weather Space Clim.*, *10*, 26.
Tasistro-Hart, A., Grayver, A., & Kuvshinov, A. (2021). Probabilistic geomagnetic storm forecasting via deep learning. *Journal of Geophysical Research: Space Physics*, *126*(1), e2020JA028228.
Torta, J. M., Serrano, L., Regué, J. R., Sánchez, A. M., & Roldán, E. (2012). Geomagnetically induced currents in a power grid of northeastern spain. *Space Weather*, *10*(6).
Viljanen, A., Pirjola, R., Prácser, E., Katkalov, J., & Wik, M. (2014). Geomagnetically induced currents in Europe - modelled occurrence in a continent-wide power grid. *J. Space Weather Space Clim.*, *4*, A09.
Weigel, R. S. (2017). A comparison of methods for estimating the geoelectric field. *Space Weather*, *15*(2), 430-440.
Weiss, A. J., Möstl, C., Amerstorfer, T., Bailey, R. L., Reiss, M. A., Hinterreiter, J., … Bauer, M. (2021). Analysis of coronal mass ejection flux rope signatures using 3DCORE and approximate Bayesian Computation. *The Astrophysical Journal Supplement Series*, *252*(1), 9.
Welling, D. T., Ngwira, C. M., Opgenoorth, H., Haiducek, J. D., Savani, N. P., Morley, S. K., … Liemohn, M. (2018). Recommendations for next-generation ground magnetic perturbation validation. *Space Weather*, *16*(12), 1912-1920.
Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences.* Amsterdam; Boston: Elsevier Academic Press.
Wintoft, P. (2005, July). Study of the solar wind coupling to the time difference horizontal geomagnetic field. *Annales Geophysicae*, *23*(5), 1949-1957.
Wintoft, P., Viljanen, A., & Wik, M. (2016). Extreme value analysis of the time derivative of the horizontal magnetic field and computed electric field. *Annales Geophysicae*, *34*(4), 485–491.
Wintoft, P., & Wik, M. (2021). Exploring three recurrent neural network architectures for geomagnetic predictions. *Frontiers in Astronomy and Space Sciences*, *8*, 72.
Wintoft, P., Wik, M., & Viljanen, A. (2015). Solar wind driven empirical forecast models of the time derivative of the ground magnetic field. *Journal of Space Weather and Space Climate*, *5*, A7.
Zhang, J. J., Wang, C., Sun, T. R., Liu, C. M., & Wang, K. R. (2015). GIC due to storm sudden commencement in low-latitude high-voltage power network in China: Observation and simulation. *Space Weather*, *13*(10), 643-655.

Forecasting GICs and geoelectric fields from solar wind data using LSTMs: application in Austria

R. L. Bailey[1], R. Leonhardt[1], C. Möstl[2], C. Beggan[3], M. A. Reiss[2], A. Bhaskar[4], A. J. Weiss[2,5]

[1]Conrad observatory, Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria
[2]Space Research Institute, Austrian Academy of Sciences, Graz, Austria
[3]British Geological Survey, Edinburgh, UK
[4]Space Physics Laboratory, ISRO/Vikram Sarabhai Space Centre, Trivandrum, India
[5]Institute of Physics, University of Graz, Universitätsplatz 5, 8010 Graz, Austria

Key Points:

- The aim is to directly forecast GICs rather than dB/dt, which is often used as a proxy.
- Results from LSTMs predicting either Ex and Ey or substation GICs from solar wind data are compared.
- GIC forecasting seems to work best when the LSTM model is trained directly on GIC data.

Corresponding author: Rachel Bailey, r.bailey@zamg.ac.at

Abstract

The forecasting of local GIC effects has largely relied on the forecasting of dB/dt as a proxy and, to date, little attention has been paid to directly forecasting the geoelectric field or GICs themselves. We approach this problem with machine learning tools, specifically recurrent neural networks or LSTMs by taking solar wind observations as input and training the models to predict two different kinds of output: first, the geoelectric field components E_X and E_Y; and second, the GICs in specific substations in Austria. The training is carried out on the geoelectric field and GICs modelled from 26 years of one-minute geomagnetic field measurements, and results are compared to GIC measurements from recent years. The GICs are generally predicted better by an LSTM trained on values from a specific substation, but only a fraction of the largest GICs are correctly predicted. This model had a correlation with measurements of around 0.6, and a root-mean-square error of 0.7 A. The probability of detecting mild activity in GICs is around 50%, and 15% for larger GICs.

Plain Language Summary

Using satellites, we measure the state of the solar wind a short distance away from the Earth (at the so-called Lagrange-1 or L1 point) to see what is coming towards us at any given moment. Changes in the solar wind such as an increase in wind speed or a strong magnetic field can potentially impact satellite operation in orbit and power grid infrastructure on the ground - in extreme cases, solar storms can damage power grids and transformers by inducing electrical currents in the power lines. These are called geomagnetically induced currents (GICs). Here, we attempt to forecast the scales of GICs by applying machine learning methods, specifically Long-Short-Term-Memory recurrent neural networks, to take the solar wind data measured at the L1 point and predict the currents that would be seen in power grids in Austria. This gives us a lead time of around 10 to 40 minutes in the forecast. We discuss whether it is best to attempt to predict the ground electric field that leads to the GICs or the GICs themselves, and discuss the difficulties in this kind of prediction and the shortfalls in the model.

1 Introduction

(Bailey et al., n.d.)

Geomagnetically induced currents (GICs) have long been known to affect power grids, transformers and any earthed conductive networks spanning large distances (for an overview, see Boteler et al., 1998; Boteler & Pirjola, 2017; Kelbert, 2020). GICs can cause problems in power grid operation such as transformer overheating or permanent transformer damage and system collapse in extreme cases (Molinski, 2002), leading to further societal and economic harm (Eastwood et al., 2018). Although studies of GICs were restricted to high latitudes where the consequences are more pronounced, mid-latitudes are being paid increasingly more attention as local effects such as transformer overheating are discovered (Barbosa et al., 2015; Butala et al., 2017; Lotz & Danskin, 2017; Gil et al., 2019; Caraballo et al., 2020; Svanda, Michal et al., 2020, among others).

The forecasting of GICs has developed alongside studies into the effects of regional GICs (Pulkkinen et al., 2006). Forecasting in particular is a complex problem due to the chain of cascading induction effects from the impingement of solar wind at the bow shock down to currents flowing between the earth and power grids on the surface. Improving predictive GIC modelling is listed as one of the open questions still to address to achieve GIC readiness (Pulkkinen et al., 2017).

Most studies so far have focused on predicting geomagnetic activity - such as dB/dt, which is often used as a proxy for GICs - from solar wind data measured at L1 or in near-

Earth space. The earliest studies addressing this problem with neural network architecture are Wintoft (2005) and Wintoft et al. (2015), followed by Lotz and Cilliers (2015) and recently Keesee et al. (2020) and Tasistro-Hart et al. (2021). The $Dst/SYMH$ index in particular has received a lot of attention from geophysicists and machine learning engineers alike (e.g. Lu et al., 2016; Bhaskar & Vichare, 2019; Wintoft & Wik, 2021).

While dB/dt is often used as a proxy for GICs, it does not provide the whole picture. The downside of modelling with this approach is that dB/dt only functions as a useful indicator of GIC activity. The relationship between dB/dt and E (which is the primary factor determining the scale of the GICs) depends on the magnetotelluric transfer function, which is frequency dependent (Chave & Jones, 2012). Single values of the time derivative of the magnetic field can only be useful GIC proxies if further assumptions on the frequency content are made (Pulkkinen et al., 2006).

What do we do if we want to develop a model that provides forecasts that power grid operators can work with? One approach would be to directly forecast the surface geoelectric field, from which GICs at different stations can be calculated. In comparison to the many studies into forecasting dB/dt and Dst, little effort has been devoted to forecasting geoelectric fields thus far. Pulkkinen et al. (2009, 2010) studied the forecasting of GICs from remote solar observations, allowing a few days warning before larger events. Modelling of geoelectric fields from solar wind to ground using full MHD modelling has been carried out by Pulkkinen, Hesse, et al. (2007), Zhang et al. (2015) and Honkonen et al. (2018), and with empirical modelling in Lotz et al. (2017).

In this study, we aim to tackle this problem from another angle and forecast regional GICs from L1 solar wind data using a machine learning method, and we compare the results to observations of GICs in Austria. We try this with two different approaches: in the first, we train a model to forecast the geoelectric field and calculate the GICs from there, and in the second we forecast the GICs directly. Predictions from both methods are evaluated and compared using data from recent years.

This study is structured as follows. Section 2 describes the data used in this study, including an analysis of 26 years of geomagnetic measurements used to model GICs in the region of Austria and a case study looking at the 2003 Halloween storm. Section 3 then goes on to describe the models built to forecast GIC values, and the results are presented in Section 4, discussed in Section 5 and summarised in Section 6.

2 Data

This analysis relies on INTERMAGNET-quality geomagnetic observatory data, which ensures a high quality of data with few data gaps or disturbances. We use data with a cadence of one minute because these are available for a long time period (26 years), which is not possible with 1 Hz data. Data with 1-minute resolution should be representative of most important GIC content (Pulkkinen et al., 2006). Due to Austria's small size (roughly 280 x 600 km), we assume that the geomagnetic variations are roughly constant across it both latitudinally and longitudinally, and therefore only select and use geomagnetic variations from one station at a time.

In the following, we describe the data sets used in this study. Geomagnetic field variations from observatory measurements were used to calculate the ground geoelectric field in Austria. GICs at any power grid substation can be calculated from the geoelectric field, and the equations for two specific substations are determined using a linear fit to observed GICs. In terms of the geomagnetic and geoelectric field components, x and y refer to the geographic northward and eastward directions respectively.

2.1 Geomagnetic observatory data from WIC and FUR

The Conrad Observatory (WIC), situated at a geomagnetic latitude of 42.95° and longitude of 89.94° according to AACGM-v2 (Shepherd, 2014), is located southwest of Vienna near the town of Muggendorf in Lower Austria. High quality geomagnetic measurements have been carried out here since the official opening mid-2014, providing six years of data for analysis. We extend the time range using data from Fürstenfeldbruck (FUR) in Bavaria, Germany. Initial studies are done using WIC data, and studies of long-term measurements are carried out using FUR data. A map showing the location of the two stations can be found in **Fig. 1**.

The Fürstenfeldbruck Geomagnetic Observatory (geomagnetic lat: 43.06°, lon: 85.93°) is one of the closest INTERMAGNET-quality geomagnetic observatories to the Conrad Observatory. It is situated almost directly west of WIC and separated by 348 km. This station is a very good proxy for geomagnetic field variations in Austria due to its proximity and the similar geomagnetic latitude and geological setting. Measurements at a quality high enough for this analysis have been carried out since 1995, providing twenty-six years of data or 13.7 million data points at a 1-minute resolution.

An analysis of the coherence between WIC and FUR data has been carried out for the overlapping years of measurements (2015-2021), in which the Pearson's correlation coefficient (PCC) between the two time series doesn't drop below 0.99 for either the x or y variables over all six years. The correlation in variations (dBx/dt and dBy/dt) is slightly lower, with the lowest values (0.91) seen in the dBy/dt values.

2.2 Geoelectric field

In order to model the expected levels of GICs, we need knowledge of the ground geoelectric field in the region. The geoelectric field for the past 26 years is modelled directly from the 1-minute geomagnetic field variations at FUR. The model approach used is the one-dimensional plane wave method (e.g. Boteler & Pirjola, 2017) using the EURHOM model number 39 (Ádám et al., 2012) to describe the one-dimensional layers of resistivity going into the Earth. We assume the time series is representative across the country, which is a reasonable approach for small areas but not for larger countries. The plane wave approach was used in favour of the thin-sheet approach used in previous studies (Bailey et al., 2017, 2018) for the shorter computation times with similar levels of accuracy. The calculation results in the horizontal geoelectric field components $\mathrm{E_X}$ and $\mathrm{E_Y}$. Note that the x-component in the geoelectric field corresponds to the y-component geomagnetic field variations, and vice versa.

2.3 Geomagnetically induced currents

To evaluate the levels of GICs over the 26 years of available FUR data, we do not follow the standard modelling procedure of putting the geoelectric field components through the full power grid network, which would be computationally heavy, but instead find a direct linear fit of the geoelectric field components to measurements of GICs to find the current at station j, i.e.

$$GIC_j = a_j \cdot E_x + b_j \cdot E_y \tag{1}$$

where a_j and b_j are station-specific real coefficients (with units A·km/V). This approach can only be used on transformer stations with measurements since the coefficients must be determined from a linear fit to the data, but it often has similar or better accuracy than results from a network model. See Pulkkinen, Pirjola, and Viljanen (2007) or Torta et al. (2012) for more discussion on this method and for the equations determining a_j and b_j.

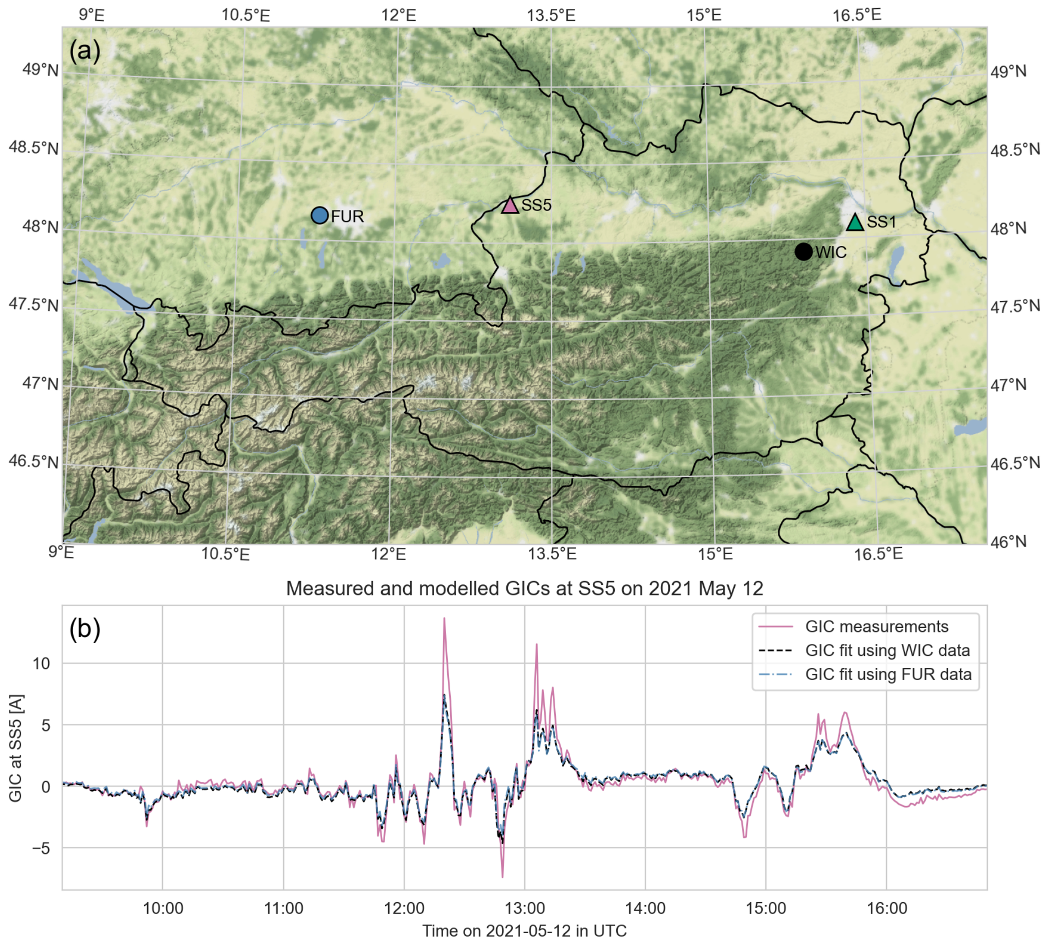

Figure 1. (a) A map showing the locations of two power grid substations (triangles) and the two geophysical observatories (circles) used for geoelectric field modelling, and (b) an example of GIC fit from modelled geoelectric field values for a geomagnetic storm in May 2021. The solid line (purple) shows transformer neutral point current measurements that have been offset-corrected and resampled via interpolation to a 1-minute sampling rate (from 1-second). The two dashed lines show the GICs calculated from E using WIC (black) and FUR (blue) data, which are nearly identical. Note that the largest GIC values are almost always underestimated despite the otherwise good agreement between model and measurements.

The fit for Eq. 1 was applied to measurements of direct currents from multiple transformer neutral points in Austrian power grid substations provided by the Graz University of Technology, a summary of which can be found in Albert et al. (2022). In this study, only measurements from two substations were used: one near Vienna (hereafter referred to SS1 for Substation 1) and another north of Salzburg (SS5), both with sampling rates of one second. The data was resampled to a one minute sampling rate for use in this study using a 1-minute median sliding window. These two stations are of interest because they are in the high-voltage network and experience larger GICs than the other stations with measurements. As such they are useful examples for depicting the expected maximum scales of GICs that could be seen across the grid. We choose three geomagnetically active periods and use the geoelectric field components E_X and E_Y modelled from FUR data to derive the following equations:

$$GIC_{SS1} = 3.77 \cdot 10^{-2} \cdot E_x + 3.19 \cdot 10^{-2} \cdot E_y \quad (2)$$

$$GIC_{SS5} = 0.44 \cdot 10^{-2} \cdot E_x + 5.55 \cdot 10^{-2} \cdot E_y \quad (3)$$

We see that the x-component of the geoelectric field contributes roughly the same amount to the GICs seen in SS1 as the y component. The y-component of the geoelectric field mostly dominates the currents in SS5 and contributes ten times more than the x-component. The differences in contributions from geoelectric field components stem from the varying grid layout and connections at each substation. An analysis shows that the GICs calculated from these equations are slightly more accurate than those from the full network model. Comparing to measurements at SS1, the Pearson's correlation coefficients for both GICs from the network model and GICs from Eq. 1 are 0.86, while at SS5 the correlation improves from 0.85 to 0.88. In both cases the amplitudes of the GICs are better matched and the root-mean-square-errors drop from 0.24 to 0.12 A at SS1 and 0.46 to 0.12 A at SS5. These measures were calculated from a fit of the geoelectric field data to measurements using eight days of geomagnetically active periods (including the September 2017 storm). This includes the most recent active period, meaning the measurements should represent the current grid configuration and we exclude fitting only to grid noise by using a geomagnetically active period. A fit applied to the geoelectric field modelled from WIC rather than FUR data produces slightly different coefficients but results in the same level of accuracy when compared to GIC measurements. An example of the measurements and GIC fits can be seen in **Fig. 1b**.

Regardless of which time range the fit is applied to, the GICs calculated using Eq. 1 (as well as those from the network model) tend to underestimate the peaks of the largest GICs by up to a factor of two (see e.g. **Fig 1b**, 12:20 or 13:05 UTC). We assume this is a result of attenuation of the modelled geoelectric field due to the lower sampling rate used for field modelling (Grawe et al., 2018) or the oversimplification of using a uniform geoelectric field and 1D model of the subsurface resistivity (Ngwira et al., 2015; Sun & Balch, 2019; Weigel, 2017). Despite this, the very good agreement between model and measurements means that any results based on the modelled geoelectric fields will still be reasonable.

In addition to the absolute GIC values, we also look at the cumulative absolute GICs over an hour, GIC_{sum1h}. GIC_{sum1h} is taken as the sum of values over the hour divided by the number of timesteps in an hour (60 for our minute values) to make it independent of sampling rate, and is used as a separate indicator for geomagnetic activity, more representative of sustained GICs than large spikes, both of which can have different (but similarly detrimental) effects on transformers (Bolduc, 2002; Gaunt & Coetzee, 2007). Using the accumulated sum of GICs or geoelectric field has seen usage in other studies, although not often - Lotz and Danskin (2017) used the accumulated E over varying periods and Viljanen et al. (2014) also worked with daily GIC sum averaged across nodes. In Austria, 0 to 0.5 Ah can be seen during quiet times, and values above that generally represent more active times.

2.4 Distribution of values

In order to determine how best to forecast GICs, we first look at the 26 years of available data and the distributions of both geomagnetic variations and modelled GICs. **Figure 2a** presents the distribution of FUR minute dBx/dt and dBy/dt variations. There are very few values populating the tail of the distribution where the largest values are found. High values for this region are at 80 nT/min and upwards. The largest variations occur most commonly in the x-direction (leading to larger $\mathrm{E_Y}$) rather than the y-direction,

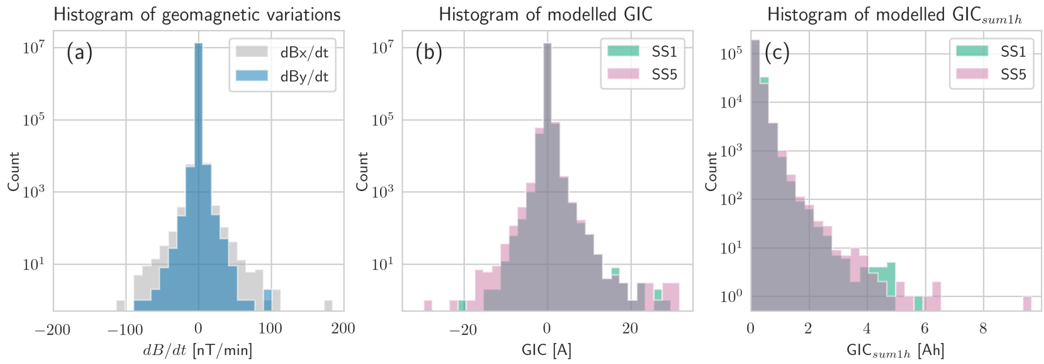

Figure 2. Histograms showing the distribution of the values in (a) the geomagnetic variations at FUR, (b) the GICs modelled from dB/dt at two substations, and (c) the hourly cumulative modelled GICs at two substations for all data, GIC$_{sum1h}$. The y-axes have logarithmic scales.

implying that stations in the power grid sitting on east-west lines are already more susceptible to larger GICs.

In **Figs. 2b and 2c**, the GICs observed at SS5 are larger than those at SS1. While the size of the currents depends largely on the network topology and grounding resistance, we noted in Section 2.3 that the currents at SS5 are mostly determined by the y-component of the geoelectric field (or x-component of the geomagnetic field variations), which generally sees larger variations.

2.5 Most active days

In **Table 1**, the 10 most active days in the 26 years of data according to different measures of activity dBx/dt and dBy/dt at FUR, modelled |GIC| and GIC$_{sum1h}$ at both SS1 and SS5 are listed. There are many overlapping days between the different measures, making a total of 19 days. Bold font highlights the ten largest values in each column.

A similar table for largest GIC days in Central Europe was produced in Viljanen et al. (2014, Table 4), and we see that the tables are very much in agreement with 17 shared dates, even though the table in Viljanen et al. (2014) is only based on one variable. They used a value akin to the GIC$_{sum1h}$ used here, namely the daily sum of GICs averaged across all nodes. Similarly, 17 of the days listed here also appear in Juusola et al. (2015), Table 3, where an analysis of the days with largest GICs was carried out for Northern Europe. Other larger storms that have occurred since those studies (March 2015 and September 2017) do not stand out in comparison to those from the last solar cycle with the exception of the storm from June 2015.

The largest values in each measure are clearly centered around the 2003 Halloween storm. Large values in dBx/dt tend to go alongside large GIC values in SS5, and days with large GIC$_{sum1h}$ usually coincide with days with larger |GIC|, as expected. Some exceptions are 2000-09-17, 2001-04-08, 2005-01-07 and 2005-08-24, which only show high cumulative GICs but do not stand out in dB/dt-values and peak GICs. A comparison of these events shows they have large and unidirectional geomagnetic field variations (with total field changes of 100 to 300 nT) that occur over an hour or more. These in particular lead to sustained GICs in stations susceptible to geomagnetic field changes in that direction. The variations on 2000-09-17 are shown as an example of this kind of behaviour in **Fig. 3**. Although not extremely geomagnetically active, they show that power grid

Table 1. Table showing the ten most active days according to the maximum values in three measures: leftmost are the horizontal geomagnetic field variations (dBx/dt and dBy/dt), in the centre the absolute GICs (|GIC|) at two different transformer stations (SS1 and SS5), and rightmost the cumulative GICs over an hour at two transformer stations (GIC_{sum1h}). Bold font highlights the ten largest values seen in that measure. The largest values are seen during the Halloween Storm on 2003 October 29-31 (italicised).

Date	dBx/dt [nT/min]	dBy/dt [nT/min]	|GIC1| [A]	|GIC5| [A]	$\mathrm{GIC1}_{sum1h}$ [Ah]	$\mathrm{GIC5}_{sum1h}$ [Ah]
1998-05-04	52.0	**46.0**	**11.37**	9.56	2.48	2.81
2000-04-06	42.9	**43.7**	8.78	11.34	3.00	3.43
2000-07-15	**184.7**	28.5	**20.30**	**29.47**	**4.79**	**6.25**
2000-09-17	34.5	19.9	10.45	9.89	**4.21**	**4.21**
2001-03-31	**82.4**	**40.7**	10.85	**17.55**	3.69	3.18
2001-11-06	**85.1**	**38.1**	**12.73**	**13.72**	3.95	**5.24**
2001-11-24	**62.4**	33.3	**14.20**	**17.81**	**4.51**	**4.18**
2003-10-29	**102.9**	**92.3**	**28.57**	**31.67**	**5.77**	**9.66**
2003-10-30	33.1	**40.3**	**17.68**	**16.44**	**4.66**	**4.71**
2003-10-31	**91.5**	**56.2**	**14.75**	**16.88**	2.41	**4.03**
2003-11-20	19.8	31.4	**11.63**	10.73	**4.82**	**4.62**
2004-07-26	**78.5**	8.5	10.15	**15.33**	1.20	1.53
2004-11-07	43.0	**37.7**	7.33	8.60	2.54	2.67
2004-11-08	24.7	28.9	9.77	9.42	**4.17**	3.57
2004-11-09	**76.1**	**49.9**	**14.21**	13.70	**4.28**	3.46
2005-05-15	36.3	**35.1**	**11.45**	**13.96**	**4.38**	**6.31**
2005-08-24	41.6	31.9	10.51	13.18	**4.01**	**6.16**
2005-09-11	**60.7**	30.7	8.81	12.50	1.24	1.65
2015-06-22	**63.0**	12.8	9.94	**16.67**	2.56	3.47

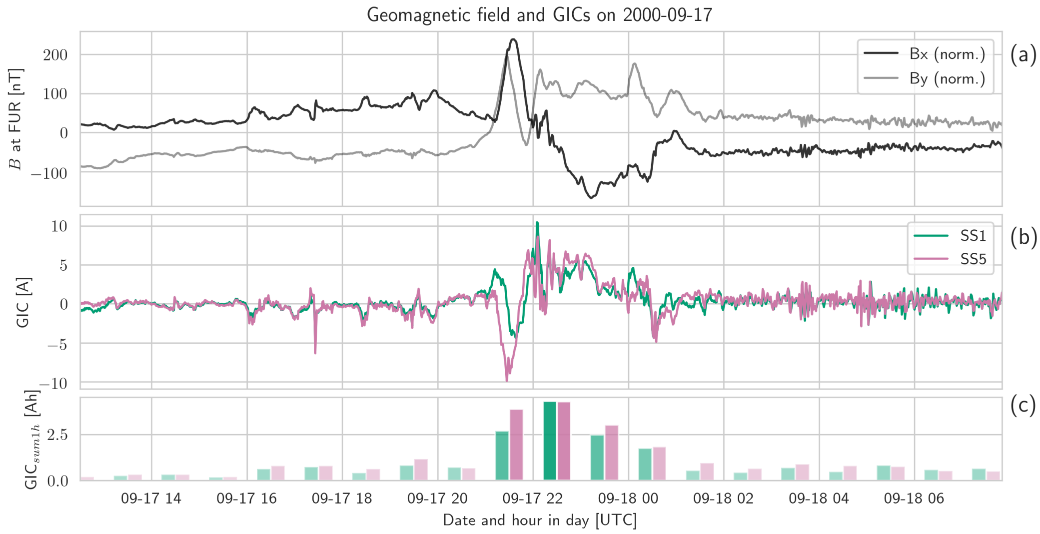

Figure 3. Plot of (a) geomagnetic variations at FUR (normalised to around zero by subtracting the mean field strength), (b) modelled GICs at two substations, and (c) cumulative hourly GICs on 2000-09-17 as an example of a day with no extreme GIC values but large cumulative hourly GICs.

transformers would have been subjected to large amounts of cumulative GICs sustained over an hour at least.

2.6 Case study: 2003 Halloween Storm

In **Fig. 2**, almost all of the values in the tail end of the distribution resulted from the "Halloween storm", which lasted from 2003 October 29 to November 1. These also make up the largest values in **Table 1**, with maximum GIC values almost twice as large as the other values seen. We now conduct a detailed analysis of the behaviour during this storm and the GICs that were likely present in the power grid as an example of the problems that can arise when using only dB/dt as a proxy for GICs. We see that both large instantaneous GICs and sustained GICs appear without large dB/dt values.

The geomagnetic storm that occurred at the end of October in 2003 was the result of a series of fast and geoeffective coronal mass ejections hitting the Earth during a particularly active period around the maximum of solar cycle 23 (e.g., Gopalswamy et al., 2005). In Eastwood et al. (2018), this storm was classified as a 1-in-10 year event, and is not considered an exceptionally rare example. No event of this or a higher magnitude has occurred since 2003 (with the exception of a CME directed away from Earth on July 2012, see Ngwira et al., 2013; Baker et al., 2013; Liu et al., 2014), and such events are somewhat more probable during the solar maxima (Owens et al., 2021), but have also occurred at any point throughout the solar cycle.

A brief evaluation of this storm for Austria was carried out in Bailey et al. (2018), in which a maximum GIC of 14 A was modelled. Using an updated model with newer data allows us to get a more accurate estimate of GICs during stronger events, and using the method from Section 2.3 for SS1 and SS5 we see the values reaching 25−30 A. Taking into account that the GIC peaks modelled using minute data generally underestimate the observations, these could also have reached up to 60 A.

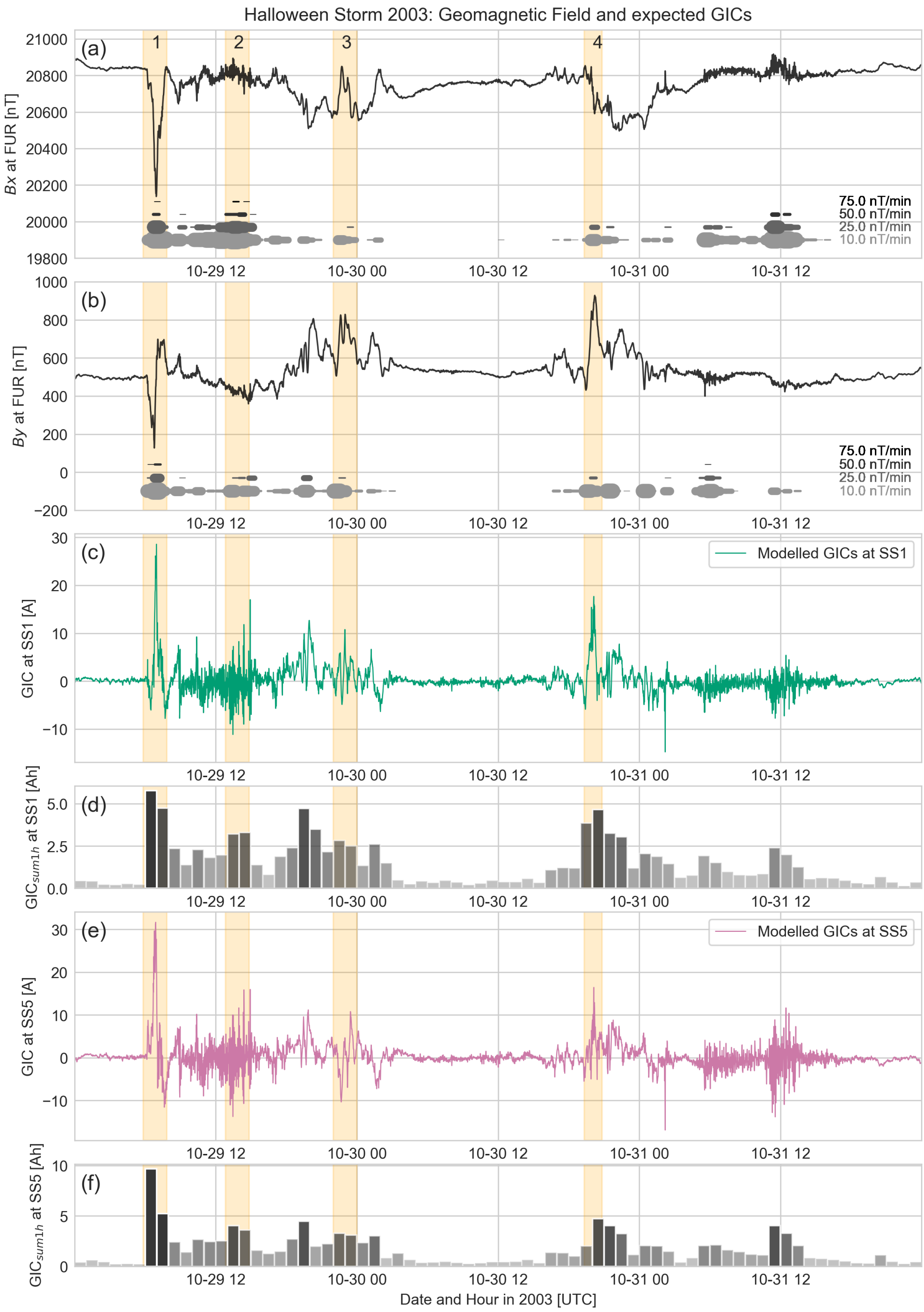

Figure 4. The Halloween storm from 2003 October 29 till 2003 November 1, during which some of the largest geomagnetic variations of the last few decades were seen. (a) and (b) show the geomagnetic variations at FUR in the x and y directions. Plotted below are levels of activity (10, 25, 50, and 75 nT/min) with line thickness showing how often these values were exceeded over a certain time range. (c) and (e) show the modelled GICs at the substations SS1 and SS5, and (d) and (f) show the cumulative GICs over each hour at each substation.

Figure 4 compares the geomagnetic field and the modelled GICs for the 2003 Halloween Storm. Panels (a) and (b) show the geomagnetic field variations in the x and y directions. The thick lines plotted below the field show the presence of various levels of dB/dt variations (as they might be shown using a forecasting method). Light grey shows a level of 10 nT/min, and this increases going upwards to 25 nT/min, 50 nT/min and 75 nT/min. The thickness of the line shows how often the value was exceeded within a time frame of 30 minutes (with a maximum being 30 times). Panels (c) and (e) show the GICs calculated from the modelled geoelectric field at the substations SS1 and SS5, and the panels (d) and (f) show the cumulative sum of absolute GIC values (GIC_{sum1h}) over 1-hour periods.

Four time intervals, highlighted in yellow on the plot, have been picked out for discussion. Intervals 1 and 2 have been selected because, as can be seen in the high levels of dB/dt in both components, these were the most active periods. Intervals 3 and 4, in contrast, were chosen because of continuously low levels of dB/dt but lack of higher ($>$ 50 nT/min) values.

Interval 1 shows a large GIC value, which is fairly short-lived. Interval 2, in contrast, shows a consistent level of moderate GICs, though it does not reach an extremely high value. Interval 3 has a similar level of sustained GIC_{sum1h} as Interval 2 despite it having a comparatively smaller amount of dB/dt over the same period. In Interval 4, SS1 experiences the second highest value of GIC (17 A) throughout the whole storm, even though there is only continuous low-level dBx/dt and dBy/dt (10 to 25 nT/min), most of it unidirectional (comparable to the type of signal seen in **Fig. 3**). On top of that, the cumulative GICs are also some of the highest.

In summary, we see there are large differences between periods that have short-lived but large GICs (Intervals 1 and 4) and those that have longer periods of sustained GICs (Intervals 2 and 3), and both large GICs and sustained GICs can appear without large dB/dt because the ground geoelectric field responds at a range of frequencies not captured by dB/dt intensity alone. Each scenario could lead to different problems if it were to occur in a transformer to any large degree (Price, 2002; Gaunt & Coetzee, 2007; Bolduc, 2002).

3 Building a Forecasting Model

From the analysis of past data, we deduce that, in order to forecast a comprehensive summary of expected GIC behaviour, we need to forecast either both geoelectric field components or the GICs directly. While the magnitude of the field is most important, the direction also plays an important role. From Eqs. 2 and 3, we see that a large value in $\mathrm{E_X}$ at SS5, for example, could be cancelled out by a smaller negative one in the $\mathrm{E_Y}$ value, and the opposite could be true elsewhere, making a station-by-station approach advantageous.

We now move on to build a forecasting model based on these conclusions. Three machine learning methods were put through an initial comparison for evaluation: a standard feed-forward neural network (NN) with three layers (32 neurons initially), a gradient boosting regressor based on XGBoost in Python (with 400 decision trees), and a recurrent neural network (specifically, a Long-Short-Term Memory RNN or LSTM) with three layers (32 blocks initially) and a basic Attention mechanism. The three types or architecture were set up in size and hyperparameter choice to be somewhat comparable in basic accuracy on an initial subset of the training data set, then were provided the full, identical data sets (scaled and shaped according to each method) and compared according to a set of metrics for model evaluation (root-mean-square error, Pearson's correlation coefficient, probability of detection). From these first comparisons, the LSTM with Attention showed the most promise and was developed into the final model, although due to the myriad machine learning methods available these days there may well be other approaches equally suited for this task. Details on the comparison can be found in the Jupyter Notebook #4 listed in Sec. 7.

3.1 Data preparation

The input to the machine learning model is solar wind data measured at L1 and forward-propagated to the bow shock. This means that, assuming we take measurements from satellites situated at L1, we have a varying forecast lead time between 15 and 60 minutes depending on the solar wind speed. The high resolution OMNI data set (see section on Data Availability for details) was used for solar wind measurements (speed, density, and magnetic field components) at a minute cadence combined with the local time and day in year to make up the features, while the model target was either the geoelectric field (E) modelled from FUR data or the GICs modelled from the E_X and E_Y components.

Taking solar wind measurements that have already been propagated forward to the bow shock, we use the two hours prior to the time we wish to forecast as input. This goes from $t - 120$ minutes to $t - 0$, where t is the forecast time. The range of 120 minutes for past data was decided on through experimentation, where the period was increased until longer periods did not lead to any improvements in the forecasting skill. To reduce the size and complexity of the input data, it is subsampled to a 10-minute resolution by picking every 10th point (rather than interpolation and/or fitting, which we found led to a loss in forecast skill), resulting in sequences of length 12. These sequences are used as input to forecast the maximum value of E or GICs over 40 minutes from $t - 10$ to t+30. This step of ten minutes into the "past" (which reduces the lead time by ten minutes) is to account for possible timing errors in propagating the solar wind forward to the bow shock.

Sampling the modelled geoelectric field or GIC data to produce a balanced data set for model training is challenging because there is a clear bias towards quiet times and not enough data from geomagnetically active times (with a factor of roughly $10^7 : 1$ for quiet to active). An initial approach using the entire data set led to a trained model that predicted only quiet times, which could not be remedied without additional data handling or large changes to the training methods. The target data set was therefore selectively sampled to reduce the imbalance. The distribution of samples was undersampled in the range of E = 0 to 100 mV/km (GIC = 0 to 8 A). Above that, we applied some data augmentation by duplicating the samples by 2 to 5 times and applying a random offset in time to the input data of each to avoid identical samples. The offset was randomly sampled without replacement from values between -10 and +10 minutes, which shifts the input solar wind data that the model sees, and means that the maximum value was either closer to the start or the end of the following 40-minute forecast window. Otherwise, all samples had a minimum time difference of 60 minutes between them. The resulting distribution is close to a one-sided Gaussian distribution. Roughly the same number of samples (9000) were used in training for each target.

The samples were split into training and testing sets by time. The years 2000, 2001 were reserved for validation to aid in model selection during training, while 2017, 2019 and 2020 were reserved for testing, and the remaining 21 years were used in training. The presence of data gaps longer than 15 consecutive minutes in the OMNI data set led to samples being excluded from the analysis - this led to 8 to 15% sample exclusion, depending on the years used. Data gaps shorter than 15 minutes were linearly interpolated over.

We reduced all values of E $>$ 200 mV/km (GIC $>$ 15 A) to 200 mV/km (15 A) because the larger values were only present in roughly 100 of the 13.7 million time-steps (or five to seven events in the 25-year period) and heavily skewed the distribution, in which all values were scaled between 0 and 1. Rescaling points above this limit greatly improved the level to which the model could learn the problem but also means that the maximum forecast the model can realistically produce is for 200 mV/km. This was tested by evaluating a model trained on data clipped at 200 mV/km versus one trained on the original data, and the model trained on clipped data performed better on both clipped and unclipped test data sets.

3.2 Training the LSTM

To approach this forecasting problem, we use a four-layer LSTM with an Attention layer. The Attention mechanism is meant to simulate human attention (first developed in Bahdanau et al., 2015), which can be understood intuitively as a mechanism that picks out the most important part of a sequence and discards the parts that are considered irrelevant. It is a tool now commonly applied in natural language processing for example (Galassi et al., 2020). The model is structured so that the input first goes through an LSTM layer and then through the Attention mechanism. The data is then fed into another LSTM layer before going through a final feed-forward layer to reduce the output to a single value.

For geoelectric field prediction, the LSTM branches into two: the left side deals with a regression problem, namely forecasting the maximum magnitude of the geoelectric field. We chose a custom loss function for the regression problem where events (peaks) are rare in the data, and where the scale of the peaks is important. A min-max scaling factor used as a penalty term meant that training to match the peak value would drive the loss down. The right side of the LSTM forecasts the sign of the geoelectric field in a classification problem, which in this case is the sign of the maximum field value used for the regression problem. Here, the binary cross-entropy loss function was used. Training worked better when the two were trained as separate targets, rather than attempting to forecast E without taking the absolute value first. The regression problem appears to be not too difficult a task, but the model had far more problems trying to forecast the direction. In training, the weights of the two problems are, when scaled, about $15 : 1$ for regression to classification. The classification problem to determine the sign is given secondary importance because even an LSTM dedicated to this problem had trouble achieving a good level of accuracy. A diagram of the different LSTM architectures, the loss functions and the hyperparameters used for the training of each model can be found in the supporting information. Iteration through the various possible hyperparameters was carried out for all four models for optimisation. Similar sets of hyperparameters were found for each LSTM application, with some minor differences between them, although the choice of the same hyperparameters for all applications also led to reasonable models in all cases. Regularisation was applied in the form of dropout.

Multiple models were trained to evaluate the best approach for forecasting GICs. Those trained to forecast the geoelectric field components are referred to as LSTM-E, while nets trained to forecast the GICs directly are referred to as LSTM-GIC. Both neural nets are only trained on the output of geophysical models (in the case of E, the result of FUR variations put through the plane-wave model, and for GICs, these are the currents calculated in power grid transformers from E) because we don't have measurements of E or GIC over long enough periods and because, as described in Sec. 2.3, GICs from geophysical models reach a good enough accuracy to be a reasonable substitute in training. Both models predict the absolute value of the target, but the LSTM-E predicts the sign (positive or negative) in addition.

3.3 Evaluating the model skill

Each model was trained on its respective training set and the best LSTM parameters were chosen based on model behaviour when presented with the validation set. Following training, we ran the model on the test data set in a virtual 'real-time mode' providing updates to the input data every 15 minutes, and giving an output with a 15-minute cadence. The comparison to the ground truth (either the modelled geoelectric field or measured GICs) is performed point-to-point as well as by looking at events, where the event-based analysis is given the most importance. In order to have a benchmark for comparison, we produced a real-time persistence approach which takes the maximum of the geoelectric field or GICs in the 20 minutes before the solar wind measurement time to forecast the maximum when the solar wind would reach Earth. As such, the persistence model (PERS) also uses a varying forecast lead time. The machine-learning forecast model should be able to beat persistence in most measures.

Our event-based analysis follows the recommendations put forward by Pulkkinen et al. (2013) and Welling et al. (2018) for dB/dt forecasting. An "event" in the data is classified as a value that exceeds a certain threshold, while all values below that threshold are non-events. By defining a threshold, we can calculate the confusion matrix (Wilks, 2011), which includes the number of correctly-predicted events or true positives (TP), missed events or false negatives (FN), incorrectly-predicted events or false positives (FP), and the correctly-predicted non-events or true negatives (TN). The metrics proposed in Pulkkinen et al. (2013) include the Probability of Detection (POD), which is the fraction of measured events correctly predicted as events, also called the true positive rate (TPR or TP/(TP+FN)). Similarly, we include the probability of False Detection (POFD), the fraction of measured non-events incorrectly predicted as events, which is equivalent to the false positive rate (FPR or FP/(FP+TN)). In addition, the Heidke Skill Score (HSS) and True Skill Statistic (TSS) are also considered, both of which are derived from all variables in the confusion matrix (see e.g. Heidke, 1926; Bloomfield et al., 2012). Both the HSS and TSS show no model skill at 0, and better model skill when approaching 1. The TSS has the benefit over the HSS of being unbiased by event/non-event ratios. We also include the bias (BS), which shows if the model tends to over-predict (more false positives, BS > 1) or under-predict (more false negatives, BS < 1).

4 Results

We present the results split in two parts: in the first part, we test our model's forecasting ability with regards to the the geoelectric field components. The results are compared to the geoelectric field modelled from geomagnetic variations at FUR (see Sec. 2.2). In the second part, we test the forecasting ability for GICs. These are calculated using (1) the geoelectric field components predicted from LSTM-E to calculate the GICs at the two substations we picked for analysis, and (2) directly from LSTM-GIC for each substation. The comparison between the model results and measurements of GICs is carried out for the years 2017, 2019 and 2020.

For the evaluation of geoelectric field forecast, we compute the scores for three event thresholds: these are 30, 60, and 90 mV/km in both E_X and E_Y. In GICs, the level of 60 mV/km corresponds to a current of roughly 4 A through either SS1 or SS5, and we use similar thresholds of 2, 4 and 6 A. It is difficult to determine the minimum level of GICs above which transformers may experience adverse effects because these are heavily dependent on transformer type and the presence of DC-handling mechanisms. We have too few measurements of GICs exceeding higher levels such as 10 A to make an analysis at this level useful, but 4 A is crossed often during geomagnetically active times. The results are described in the next section.

Figure 5 gives a graphical representation of the model behaviour at each threshold using receiver-operator characteristic (ROC) and detection-error tradeoff (DET) curves. Both depict the model's ability to forecast events at varying thresholds. The ROC curve shows the trade-off between the true positive rate (also POD) and false positive rate (also POFD) at different event thresholds. Usually, when the threshold is low, the TPR is high but we also see an increased FPR, which is unwanted - a model that captures the ob-

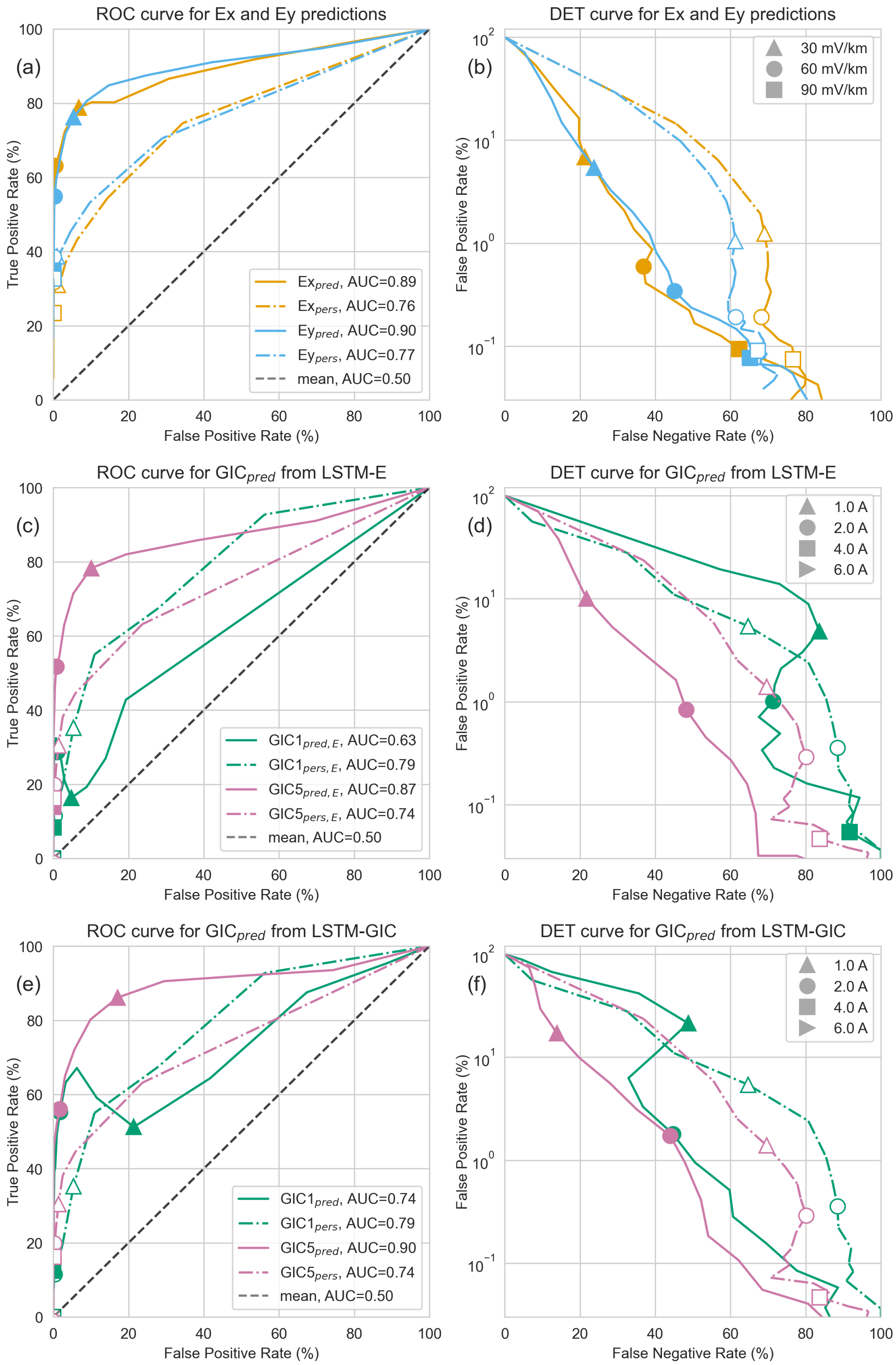

Figure 5. Receiver-operator characteristic (ROC) and detection-error tradeoff (DET) curves for three approaches: (a-b) the geoelectric field, showing the output from the LSTM-E models vs the modelled geoelectric field, (c-d) the GICs calculated from the geoelectric field predicted by LSTM-E compared to measured GICs, and (e-f) the GICs predicted by the LSTM-GIC models compared to measured GICs. SS1 and SS5 are two separate substations in the power grid from which we have measurements. The values for specific event thresholds are labelled with shapes as defined in each legend.

served behaviour shows a curve that keeps close to the upper left corner. The area-under-the-curve (AUC in the legend) shows good model skill as it approaches 1. On the other hand, the DET curve shows the relationship between the false negative rate (fraction of all predicted non-events that were measured events misclassified as non-events, or FN/(FN+TN)) and false positive rate, the number of which usually goes up as the other goes down depending on where the threshold for an event is set. Here, the best model behaviour is seen as the curves approach the lower left corner. It is useful in error minimisation to deduce the rate at which the FNR improves with regards to an increase in FPR rate (and vice-versa).

Table 2. Metrics from an event-based analysis of the LSTM-E models applied to the years 2000, 2001, 2017, 2019 and 2020 in a retrospective real-time mode with the model being run at 15-minute intervals. A persistence model (PERS) is included for comparison. The first four columns provide the values for the confusion matrix (where TP, FP, TN and FN are the true positives (hits), false positives, true negatives (misses) and false negatives), the probability of detection (POD), probability of false detection (POFD), Heidke Skill Score (HSS), True Skill Score (TSS), and bias (BS). The variable TH in brackets gives the event threshold used to define events and compute the metrics.

LSTM-E Model	$N_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
$E_{x,pred}$(TH=30)	3092	2436	11749	656	160506	78.8	6.8	0.26	0.72	4.6
$E_{x,pred}$(TH=60)	494	312	1038	182	173815	63.2	0.6	0.34	0.63	2.7
$E_{x,pred}$(TH=90)	175	66	164	109	175008	37.7	0.1	0.33	0.38	1.3
$E_{y,pred}$(TH=30)	2989	2279	9328	710	163030	76.2	5.4	0.29	0.71	3.9
$E_{y,pred}$(TH=60)	559	307	600	252	174188	54.9	0.3	0.42	0.55	1.6
$E_{y,pred}$(TH=90)	241	84	135	157	174971	34.9	0.1	0.36	0.35	0.9
PERS Model	$N_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
$E_{x,pers}$(TH=30)	3092	958	2128	2134	170127	31.0	1.2	0.30	0.30	1.0
$E_{x,pers}$(TH=60)	494	157	335	337	174518	31.8	0.2	0.32	0.32	1.0
$E_{x,pers}$(TH=90)	175	41	130	134	175042	23.4	0.1	0.24	0.23	1.0
$E_{y,pers}$(TH=30)	2989	1156	1804	1833	170554	38.7	1.0	0.38	0.38	1.0
$E_{y,pers}$(TH=60)	559	216	335	343	174453	38.6	0.2	0.39	0.38	1.0
$E_{y,pers}$(TH=90)	241	79	158	162	174948	32.8	0.1	0.33	0.33	1.0

4.1 Forecasting E_x and E_y

We first evaluate the LSTMs trained on the geoelectric field in terms of the root-mean-square-error (RMSE) and the Pearson's correlation coefficient (PCC). Comparing the LSTM-E outputs to modelled E, the RMSE values are 126 mV/km and 111 mV/km for the absolute value of E_x and E_y, while the PCC values are 0.60 and 0.61. Once the sign of E has been included, the RMSE rises to 261 mV/km and 287 mV/km, while PCC drops to 0.48 and 0.32, so we see that the model's inability to forecast the field direction reliably decreases the accuracy when also considering the field direction.

Table 2 presents an event-based analysis of the LSTM-E results. Multiple thresholds (TH) defining events were considered, and these are listed by the variable "TH" in each line (at 30, 60, and 90 mV/km, representing minor, moderate and strong geomagnetic activity). We see that the skill decreases as the threshold increases (decreasing probability of detection POD and TSS), and that the LSTMs tend towards over-predicting (BS > 1). (The bias for the PERS models is always ~ 1 because the time series being compared are only shifted in time and therefore almost statistically equivalent.) There are always a large number of false positives, although this remains a small fraction of the number of total data points. The LSTM-E models generally outperform the PERS approach, although the Heidke Skill Scores are occasionally smaller in the LSTMs, which implies a worse balance between false positives and true positives. As in the point-to-point values, the E_X component tends to be predicted better than the E_Y component. By evaluating the ROC and DET curves in Fig. 5 (a-b), we see that the LSTM-E models outperforms persistence at all thresholds.

We also conducted a comparison with the results from Honkonen et al. (2018) and Lotz and Danskin (2017), where possible. While the time development of the geoelectric field appears better in the modelling approach in Honkonen et al. (2018), the magnitudes are not matched as well. An event-based analysis could not be carried out in their case due to the short time series and lack of larger events, but the RMSE and PCC values for E_X and E_Y (reduced to a 15-min sampling rate) come out as 10.5 mV/km and 97.8 mV/km and 0.62 and 0.25, respectively, which is better in the case of E_X but worse in the case of E_Y. Comparing to Lotz and Danskin (2017), we see similar correlations for the geoelectric field components. They found a slightly higher correlation (averaged over three stations and two storms, 0.71 for E_X and 0.53 for E_Y), although they predicted the maximum value for a longer time span (90 mins), making their approach closer to a nowcast than a forecast. The higher RMSE values seen in our study in part derive from the slightly higher levels of daily variation that is forecast even when the field is extremely quiet. Again, in both studies used as comparison we see the northward component of the geoelectric field was predicted better than the eastward component.

4.2 Forecasting GICs

The same results are presented for GICs as for the geoelectric field components in the last section. In the event-based analysis, the thresholds were set at 2, 4 and 6 A, which are roughly equivalent to the thresholds used for the electric field. **Table 3** shows the results of this analysis applied to the test data set years 2017, 2019 and 2020, while **Fig. 5** depicts the ROC and DET curves for the model output versus measured GICs. A comparison between the LSTM-GIC output and the modelled GICs the model was trained on shows similar levels of accuracy as in LSTM-E to the geoelectric field.

We first look at the results for GICs calculated from the geoelectric field components predicted using the LSTM-E models. Note that while the last section mainly looked at the absolute value of the geoelectric fields, in the calculation of GICs the direction of the geoelectric field is also included, making this an additional error factor if the sign is not predicted accurately. Once the GICs have been calculated using the results from the LSTM-E models and Eq. 1, the absolute value is taken for the rest of the analysis.

As can be seen in **Table 3**, the GICs derived from the LSTM-E models see a considerable drop in accuracy in comparison to the results for E alone in Table 2. Although there were quite reasonable values for POD predicting E, the POD for GICs at the mid-range threshold (60 mV/km or 4 A) drops from around 50% in both components of E to 8% and 16% in substation SS1 and SS5. Evaluating the skill of the model for GICs at high levels is difficult because there are so few events exceeding even a minimal value of 6 A. None of these events (2 at SS1, 12 at SS5 over the three years of data) were predicted using any approach.

In comparing the GIC predictions from the two methods (LSTM-E and LSTM-GIC), we see that the LSTM-GIC seems to perform better but the results are station-specific. The LSTM-GIC performs much better than the LSTM-E at SS1 (e.g. a POD of 55% rather than 29% and higher HSS and TSS values at a threshold of 2 A) and at a similar level at SS5. This is also reflected in a model evaluation using point-to-point metrics. The RMSE values for SS1 and SS5 predicted using LSTM-E are 0.49 A and 0.59 A, while the PCC

Table 3. Metrics from an event-based analysis of different model applied to the years 2017, 2019 and 2020 in a retrospective real-time mode with the model being run at 15-minute intervals. $GIC1_{pred,E}$ is the result from the models trained to predict the geoelectric field (LSTM-E), while $GIC1_{pred}$ is the result from the LSTM-GIC. PERS is a persistence model assuming the target (GIC) repeats itself. The first four columns provide the values for the confusion matrix (where TP, FP, TN and FN are the true positives (hits), false positives, true negatives (misses) and false negatives), the probability of detection (POD), probability of false detection (POFD), Heidke Skill Score (HSS), True Skill Score (TSS), and bias (BS). The variable TH in brackets is the event threshold used to define events and compute the metrics."undef." refers to the HSS and TSS at TP=0, which are undefined.

LSTM-E Model	$N_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
$GIC1_{pred,E}$(TH=2)	432	124	1060	308	103697	28.7	1.0	0.15	0.28	2.7
$GIC1_{pred,E}$(TH=4)	24	2	57	22	105108	8.3	0.1	0.05	0.08	2.5
$GIC5_{pred,E}$(TH=2)	307	159	681	148	80649	51.8	0.8	0.27	0.51	2.7
$GIC5_{pred,E}$(TH=4)	43	6	13	37	81581	14.0	0.0	0.19	0.14	0.4
LSTM-GIC Model	$N_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
$GIC1_{pred}$(TH=2)	432	239	1886	193	102871	55.3	1.8	0.18	0.54	4.9
$GIC1_{pred}$(TH=4)	24	3	26	21	105139	12.5	0.0	0.11	0.12	1.2
$GIC5_{pred}$(TH=2)	307	172	1403	135	79927	56.0	1.7	0.18	0.54	5.1
$GIC5_{pred}$(TH=4)	43	7	16	36	81578	16.3	0.0	0.21	0.16	0.5
PERS Model	$N_{events,obs}$	TP	FP	FN	TN	POD	POFD	HSS	TSS	BS
$GIC1_{pers}$(TH=2)	432	50	375	382	104382	11.6	0.4	0.11	0.11	1.0
$GIC1_{pers}$(TH=4)	24	0	26	24	105139	0.0	0.0	undef.	undef.	1.1
$GIC5_{pers}$(TH=2)	307	61	237	246	81093	19.9	0.3	0.20	0.20	1.0
$GIC5_{pers}$(TH=4)	43	7	38	36	81556	16.3	0.0	0.16	0.16	1.0

is 0.35 and 0.67. For GICs predicted using LSTM-GIC, the RMSE values are 0.67 A and 0.78 A (i.e. slightly worse than LSTM-E), but the PCC is 0.56 and 0.64. The accuracy between the two approaches is roughly equivalent for SS5, but using LSTM-GIC rather than LSTM-E is a definite improvement for SS1 observations. Some of the reason for this can be seen in **Fig. 6**. In SS1, the jumps in values computed from LSTM-E result from changes in the sign of the geoelectric field components, which then cancel each other out and lead to a GIC of zero. (Conversely, ignoring the sign from LSTM-E and taking the absolute values to calculate the GICs in SS1 results in higher correlation and POD but a far larger number of false positives, leaving this as another possibility.) In the best cases, the GIC forecasts only reach a POD of 16% for GICs above a threshold of 4 A, highlighting the difficulty in correctly predicting larger values.

In the ROC and DET curves in **Figure 5** panels (c-d) for GICs from LSTM-E and (e-f) from LSTM-GIC, we also see some of the weak forecasting ability for SS1 primarily represents the LSTM behaviour at low values (GICs < 1 A). At SS1, there is a mostly continuous level of noise around 1 A, and the model does not predict the noise while the persistence model captures it clearly. This is an example of the weakness of ROC curves, where in this case only the lower left corner (showing values greater than 1 A) is of interest to us.

Figure 6 shows the forecast that would have been produced by the model (solid and black dashed lines) against measurements (coloured lines) during the September 2017

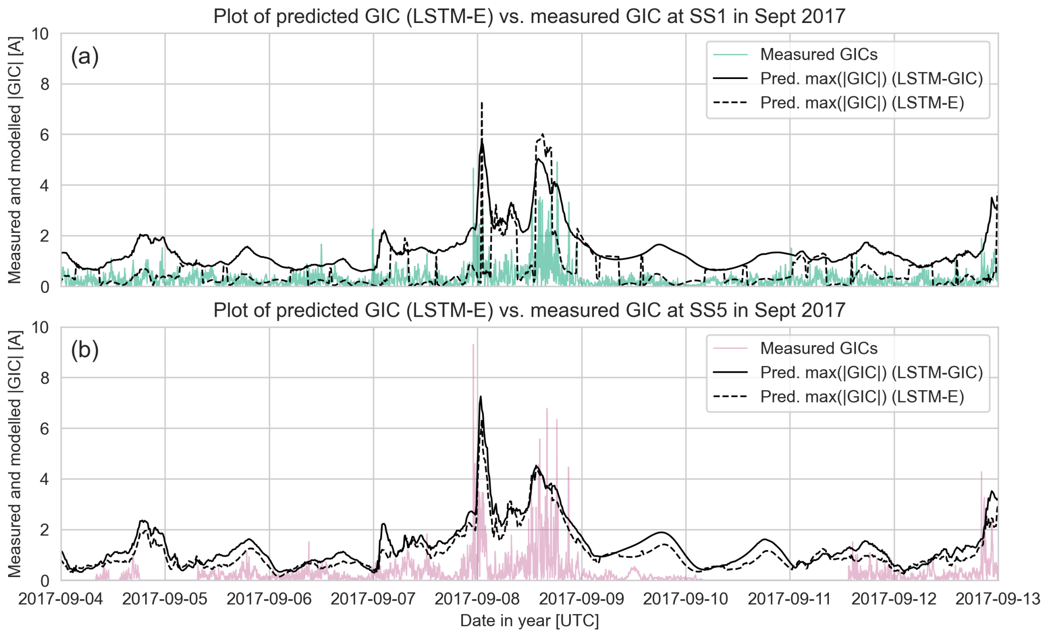

Figure 6. The LSTM-E (dashed line) and LSTM-GIC (solid line) applied to forecasts in an experimental real-time mode and compared to measurements of GICs (coloured lines) at two stations in Austria. The upper panel (a) shows results for SS1 near Vienna, while the lower panel (b) shows results for SS5 near Salzburg (with some data gaps). Although not plotted here, the maximum GIC value computed from the measurements is at the same cadence of 15 minutes to compare to the model forecasts.

storm. The models, particularly the LSTM-GIC approach, do a reasonable job at predicting magnitudes, although the LSTM-E struggles to predict the direction, which is also important for accurate GIC prediction. The storm and the active periods are clearly captured by the forecast, and daily variations from the Sq current are forecasted otherwise. Note that the delayed rise in the forecast of the first peak of the storm does not indicate a timing error. A cross-correlation of the model output shows at maximum an offset in time of 10 minutes and the delay in the figure is simply a feature unique to this storm. While the exact time development of the storm is not captured well, the general scales of GICs are matched well, as is the differentiation between quiet and active times.

In summary, prediction of geoelectric field magnitudes can be achieved with reasonable accuracy (POD of at least 35% even at the highest event threshold), but the prediction of elevated levels of GICs proves difficult with any approach used. The LSTMs usually outperform the persistence models, except in the bias, where the persistence model has the benefit of being statistically equivalent to the data it is being compared to. The persistence model also generally has a lower POFD and higher HSS value at low thresholds (e.g. TH=30 V/km for LSTM-E) because quiet periods tend to persist over time. The LSTMs, however, outperform persistence at the higher thresholds, which are more important for forecasting purposes.

5 Discussion

We have attempted to forecast GICs from solar wind data using LSTMs with two different approaches. We now look at some of the reasons behind the particular difficulty in forecasting GICs.

Some of the low skill seen when comparing predictions to GIC measurements is down to four reasons, mostly related to our GIC data: firstly, there is noise in the GIC observations, particularly at SS1, which has a consistent level of 1 A noise during the day - this is not predicted by the model. Secondly, GIC observations until 2021 had a maximum cutoff point of 3.4 A in the positive direction, removing some peaks from our event list, and these have not been accounted for. Thirdly, the model struggles to predict the direction of the geoelectric field values, which are likely driven by smaller-scale ionospheric currents (Dimmock et al., 2020). Fourthly, as noted in Sec. 2.3, the peaks of observed GICs are often underestimated by geophysical modelling, meaning peaks in the GIC measurements after the cut-off level was removed were often much larger than modelled. This is a problem related to the geoelectric field modelling that may affect the LSTM's ability to learn the problem due to insufficient accuracy in the field modelling. While minute cadence data does capture most of the variability in the GICs, the lack of higher frequency content appears to the primary cause of underestimated peaks, a problem discussed before in Grawe et al. (2018) and recently for the specific problem of GIC estimates in Beggan et al. (2021). As such, it is not surprising that the LSTMs tend to underestimate the actual GICs, and a correction would have to be applied to the target data to account for this.

Outside of the data-specific problems, there are also some timing errors, meaning some peaks arrived slightly later or earlier than they were observed, and as such are not logged as correct predictions even though an event threshold was crossed.

In an application of the model in operations, one caveat is that the maximum possible forecast is 200 mV/km due to a self-imposed limit to improve the model's ability to learn. We assume that in practise, this would be negligible because all values above a certain level (e.g. 100 mV/km) would be of interest, regardless of how large they become. As also discussed in Wintoft et al. (2016), the scale of geomagnetic variations during extreme events can theoretically become so large that it is effectively unbounded for the purpose of this discussion. In the future, this 200 mV/km limit could be improved on by training a model specifically for large value forecasting, which can be switched to if the original model forecasts $E > 150$ mV/km.

In an ideal case, a forecasting model would be developed while taking a cost-loss analysis (Murphy, 1977) such as that used in a space weather context in Owens et al. (2014) into consideration. In the case of network protection, this is a very complex scenario due to the varying impacts and costs associated with transformer damage or power grid outage, many of which are currently nearly impossible to estimate. This is something that can hopefully be developed further as studies into GIC risk progress (Eastwood et al., 2018).

Another, more general problem in forecasting any measure of ground geomagnetic activity from solar wind measurements without further input from the magnetosphere-ionosphere system is that not all geomagnetic variations are driven by the solar wind directly (see e.g. Kamide et al., 1998; Eastwood et al., 2015). Many of the ground variations, particularly at shorter timescales (Alberti et al., 2017), are not directly driven by the solar wind but are instead the consequence of other processes being triggered. These can include complex magnetospheric dynamics such as reconnection in the magnetotail, as well as random, chaotic processes. Such processes can not be related in detail through our model, which is essentially a coupling function from the solar wind at the bow shock to the geoelectric field in Austria. Some of the dynamics will be represented to some degree, but it is difficult to ascertain exactly which in a black-box machine learning model. A further difficulty in improving predictions lies in the fact that GICs can only be calculated accurately with knowledge of magnetic field variations at timescales of seconds (Grawe et al., 2018), ideally, and the LSTM must make approximations of what kind of variations are expected due to the conditions rather than deriving the variations precisely. Although the machine learning approach described here works at a basic level and could be more promising than forecasts of dB/dt alone, to create a model that can also account for complex magnetospheric processes it would need to be coupled with either data from space-borne monitors observing the Earth's magnetosphere, more complex physical models of magnetospheric behaviour (developing a so-called grey-box model as recommended in Camporeale (2019), for example), or both.

The calculations and measurements of the GICs shown in this study are for a specific grid configuration, even though the power grid is continually being upgraded and changed. These changes can have large effects on individual GIC scales over long time ranges. The results shown in Table 1 extend far into the past, for which we do not have a detailed history of grid changes, so the values listed could have been much smaller or much larger depending on how the grid was set up. For the LSTM predictions, we have conducted our analysis with the comparison to measurements over a considerably shorter time range of a few years, where the grid has not changed to any great degree, but the predictions may not be valid in the future for a different grid configuration. In this case, a new fit would need to be found for Eq. 1, and either the LSTM-GIC model would need to be retrained on the updated GIC data, or the GIC values could be calculated anew from the otherwise unchanged LSTM-E output.

Our aim was to develop a model that can provide useful forecasts for power grid operators by providing estimates of the scales of GICs. The difference between this and former studies such as Lotz et al. (2017) and Honkonen et al. (2018), who also predicted ground geoelectric fields from solar wind data, is that we have approached the problem with a new tool (a recurrent neural network) and have been able to forecast GICs directly along with the geoelectric field, with the results compared to measured GICs. We have had some success, particularly with forecasting the geoelectric field, and have tried forecasting substation-specific GICs for the first time, but there are still many problems to be addressed to turn this method into a useful forecast.

6 Summary

We have developed a machine learning approach to forecast GICs in Austria. Using data from the past 26 years and the 2003 Halloween storm as a case study, we argued that forecasts of dB/dt alone, which have been the focus of most past studies, are not sufficient to make actionable GIC forecasts.

From this initial analysis, we set out to forecast maximum expected GICs (over a forty minute window) either directly for specific substations in the power grid or more generally from forecasts of the regional geoelectric field components. From a small set of initial machine learning approaches, an LSTM (recurrent neural network) with an Attention mechanism showed the most promise in forecasting skill and this was developed into a more complex approach.

A selection of models were trained on 21 years of geoelectric field values modelled from geomagnetic variations at the geomagnetic observatory in Fürstenfeldbruck close to Austria. In the first method, two recurrent neural networks or LSTMs were trained to predict the northward and eastward modelled geoelectric field components and compute the specific substation GICs using a linear equation. In the second method, an LSTM was trained to predict modelled GICs at two substations, which we know correlate very well with the measurements. Five years of data were reserved for testing and evaluating the model. The results were compared to DC measurements at two substations in the Austrian power grid.

The LSTM model worked with reasonable success when predicting the geoelectric field modelled from geomagnetic variations, although translating this success into good GIC forecasts proved difficult. It was possible, however, to outperform a model that simply takes the last observed GICs to forecast future values.

We conclude that forecasting the GICs observed in the power grid from solar wind data measured at L1 is a difficult task, even when the forecasting model does a reasonable job of forecasting the geoelectric field components or modelled GIC. There are many ways to improve the modelling in the future, including using higher-resolution magnetic field measurements (or applying a correction to the modelled geoelectric field before training) to more accurately estimate the peak geoelectric field and GIC values, and by including information on the development of the magnetosphere during storm times.

Although this study has looked specifically at a mid-latitude region, where geomagnetic variations and GICs are not as large as those seen in higher latitude regions such as Scandinavia, we have been able to compare model output directly to measurements and expect that the conclusions drawn will also be valid for other regions with GICs at different scales.

A lower-resolution version of the LSTM-E model will be coupled with the PREDSTORM solar wind forecast (Bailey et al., 2020), which provides forecasts of the ambient solar wind a few days in advance, based on either a recurrence model or data from a spacecraft east of the Sun-Earth line such as STEREO or a future mission to the Lagrange 5 point. We also plan in the future to integrate methods on solar wind B_z forecasting (Reiss et al., 2021) or CME flux rope modelling (Weiss et al., 2021) to advance our capabilities in GIC forecasting for any type of solar wind structures.

7 Data Availability

- INTERMAGNET data for FUR and WIC:
  `https://intermagnet.org/data-donnee/download-eng.php`
- OMNI data: `https://spdf.gsfc.nasa.gov/pub/data/omni/high_res_omni/`
- Open source code for this work (in Python 3 and Jupyter Notebook form):
  `https://doi.org/10.5281/zenodo.5704715`
- Exact details on the LSTM structure and hyperparameters used for training can be found in the supporting information for this study.
- A subset of the data set used to derive the results, namely the the GIC observations and model forecasts used to produce Figure 6, have also been included in the supporting information and saved in an online repository:
  `https://doi.org/10.6084/m9.figshare.19102772.v1`

Acknowledgments

The results presented in this paper rely on the data collected at the Conrad Observatory (ZAMG) in Austria and at Fürstenfeldbruck (LMU), Germany. We thank the Ludwig-Maximilians-Universität München for supporting its operation and INTERMAGNET for promoting high standards of magnetic observatory practice (`www.intermagnet.org`). The data used in this study is publicly available (with the exception of the measurements of GICs in Austria), and details on where to find the data can be found in Section 7. An excerpt of GIC measurements and model output has been provided in the Supporting Information, and we thank both the TU Graz and Austrian Power Grid for supporting these measurements. We thank Ilja Honkonen for providing the results from his 2018 study for comparison. R.L.B., C.M., M.A.R., and A.J.W. thank the Austrian Science Fund (FWF) for research funding from projects P31659-N27 and P31521-N27. C.D.B. was funded under UK Natural Environment Research Council Grant NE/P017231/1 "Space Weather Impact on Ground-based Systems (SWIGS)". We thank the two anonymous reviewers for their careful readings of this manuscript and helpful suggestions.

References

Ádám, A., Prácser, E., & Wesztergom, V. (2012). Estimation of the electric resistivity distribution (EURHOM) in the European lithosphere in the frame of the EURISGIC WP2 project. *Acta Geodaetica et Geophysica Hungarica*, *47*(4), 377-387.

Albert, D., Schachinger, P., Bailey, R. L., Renner, H., & Achleitner, G. (2022). Analysis of long-term gic measurements in transformers in austria. *Space Weather*, *20*(1), e2021SW002912.

Alberti, T., Consolini, G., Lepreti, F., Laurenza, M., Vecchio, A., & Carbone, V. (2017). Timescale separation in the solar wind-magnetosphere coupling during st. patrick's day storms in 2013 and 2015. *Journal of Geophysical Research: Space Physics*, *122*(4), 4266-4283.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. In *3rd international conference on learning representations.*

Bailey, R. L., Halbedl, T. S., Schattauer, I., Achleitner, G., & Leonhardt, R. (2018). Validating GIC models with measurements in Austria: evaluation of accuracy and sensitivity to input parameters. *Space Weather*.

Bailey, R. L., Halbedl, T. S., Schattauer, I., Römer, A., Achleitner, G., Beggan, C. D., … Leonhardt, R. (2017). Modelling geomagnetically induced currents in midlatitude Central Europe using a thin-sheet approach. *Annales Geophysicae*, *35*(3), 751.

Bailey, R. L., Leonhardt, R., Möstl, C., Beggan, C., Reiss, M. A., Bhaskar, A., & Weiss, A. J. (n.d.). Forecasting gics and geoelectric fields from solar wind data using lstms: application in austria. *Space Weather*, *n/a*(n/a), e2021SW002907. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021SW002907` (e2021SW002907 2021SW002907) doi: https://doi.org/10.1029/2021SW002907

Bailey, R. L., Möstl, C., Reiss, M. A., Weiss, A. J., Amerstorfer, U. V., Amerstorfer, T., … Leonhardt, R. (2020). Prediction of Dst during solar minimum using in situ measurements at l5. *Space Weather*, *18*(5), e2019SW002424.

Baker, D. N., Li, X., Pulkkinen, A., Ngwira, C. M., Mays, M. L., Galvin, A. B., & Simunac, K. D. C. (2013, October). A major solar eruptive event in July 2012: Defining extreme space weather scenarios. *Space Weather*, *11*(10), 585-591. doi: 10.1002/swe.20097

Barbosa, C., Alves, L., Caraballo, R., Hartmann, G. A., Papa, A. R., & Pirjola, R. J. (2015). Analysis of geomagnetically induced currents at a low-latitude region over the solar cycles 23 and 24: comparison between measurements and calculations. *Journal of Space Weather and Space Climate*, *5*, A35.

Beggan, C. D., Richardson, G. S., Baillie, O., Hübert, J., & Thomson, A. W. (2021). Geolectric field measurement, modelling and validation during geomagnetic storms in the uk. *Journal of Space Weather and Space Climate*, *11*, 37.

Bhaskar, A., & Vichare, G. (2019). Forecasting of SYMH and ASYH indices for geomagnetic storms of solar cycle 24 including St. Patrick´s day, 2015 storm using NARX neural network. *J. Space Weather Space Clim.*, *9*, A12.

Bloomfield, D. S., Higgins, P. A., McAteer, R. T. J., & Gallagher, P. T. (2012, Mar). Toward Reliable Benchmarking of Solar Flare Forecasting Methods. *The Astrophysical Journal Letters*, *747*(2), L41.

Bolduc, L. (2002). GIC observations and studies in the Hydro-Québec power system. *Journal of Atmospheric and Solar-Terrestrial Physics*, *64*(16), 1793–1802.

Boteler, D. H., & Pirjola, R. J. (2017). Modeling geomagnetically induced currents. *Space Weather*, *15*(1), 258-276.

Boteler, D. H., Pirjola, R. J., & Nevanlinna, H. (1998). The effects of geomagnetic disturbances on electrical systems at the Earth's surface. *Advances in Space Research*, *22*, 17-27.

Butala, M. D., Kazerooni, M., Makela, J. J., Kamalabadi, F., Gannon, J. L., Zhu, H., & Overbye, T. J. (2017). Modeling geomagnetically induced currents from magnetometer measurements: Spatial scale assessed with reference measurements. *Space Weather*, *15*(10), 1357-1372.

Camporeale, E. (2019). The challenge of machine learning in space weather: Nowcasting and forecasting. *Space Weather*, *17*(8), 1166-1207.

Caraballo, R., González-Esparza, J. A., Sergeeva, M., & Pacheco, C. R. (2020). First GIC estimates for the Mexican power grid. *Space Weather*, *18*(2), e2019SW002260.

Chave, A. D., & Jones, A. G. (2012). *The magnetotelluric method: Theory and practice.* Cambridge University Press.

Dimmock, A. P., Rosenqvist, L., Welling, D. T., Viljanen, A., Honkonen, I., Boynton, R. J., & Yordanova, E. (2020). On the regional variability of db/dt and its significance to gic. *Space Weather*, *18*(8), e2020SW002497.

Eastwood, J., Hapgood, M., Biffis, E., Benedetti, D., Bisi, M., Green, L., … Burnett, C. (2018). Quantifying the economic value of space weather forecasting for power grids: An exploratory study. *Space Weather*, *16*(12), 2052-2067.

Eastwood, J., Hietala, H., Toth, G., Phan, T., & Fujimoto, M. (2015). What controls the structure and dynamics of Earth's magnetosphere? *Space Science Reviews*, *188*(1), 251–286.

Galassi, A., Lippi, M., & Torroni, P. (2020). Attention in natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*.

Gaunt, C. T., & Coetzee, G. (2007). Transformer failures in regions incorrectly considered to have low GIC-risk. In *Power tech, 2007 ieee lausanne* (p. 807-812). IEEE.

Gil, A., Modzelewska, R., Moskwa, S., Siluszyk, A., Siluszyk, M., Wawrzynczak, A., & Zakrzewska, S. (2019). Does time series analysis confirms the relationship between space weather effects and the failures of electrical grids in south poland? *Journal of Mathematics in Industry*, *9*(1), 1–16.

Gopalswamy, N., Yashiro, S., Liu, Y., Michalek, G., Vourlidas, A., Kaiser, M. L., & Howard, R. A. (2005, September). Coronal mass ejections and other extreme characteristics of the 2003 October-November solar eruptions. *Journal of Geophysical Research (Space Physics)*, *110*(A9), A09S15. doi: 10.1029/2004JA010958

Grawe, M. A., Makela, J. J., Butala, M. D., & Kamalabadi, F. (2018). The impact of magnetic field temporal sampling on modeled surface electric fields. *Space Weather*, *16*(11), 1721-1739.

Heidke, P. (1926). Berechnung des erfolges und der güte der windstärkevorhersagen im sturmwarnungsdienst. *Geografiska Annaler*, *8*(4), 301–349.

Honkonen, I., Kuvshinov, A., Rastätter, L., & Pulkkinen, A. (2018). Predicting global ground geoelectric field with coupled geospace and three-dimensional geomagnetic induction models. *Space Weather*, *16*(8), 1028-1041.

Juusola, L., Viljanen, A., van de Kamp, M., Tanskanen, E. I., Vanhamäki, H., Partamies, N., & Kauristie, K. (2015). High-latitude ionospheric equivalent currents during strong space storms: Regional perspective. *Space Weather*, *13*(1), 49-60.

Kamide, Y., Baumjohann, W., Daglis, I. A., Gonzalez, W. D., Grande, M., Joselyn, J. A., … Vasyliunas, V. M. (1998). Current understanding of magnetic storms: Storm-substorm relationships. *Journal of Geophysical Research: Space Physics*, *103*(A8), 17705-17728.

Keesee, A. M., Pinto, V., Coughlan, M., Lennox, C., Mahmud, M. S., & Connor, H. K. (2020). Comparison of deep learning techniques to model connections between solar wind and ground magnetic perturbations. *Frontiers in Astronomy and Space Sciences*, *7*, 72.

Kelbert, A. (2020). The role of global/regional earth conductivity models in natural geomagnetic hazard mitigation. *Surv Geophys*, *41*, 115–166.

Liu, Y. D., Luhmann, J. G., Kajdič, P., Kilpua, E. K. J., Lugaz, N., Nitta, N. V., ... Galvin, A. B. (2014, March). Observations of an extreme storm in interplanetary space caused by successive coronal mass ejections. *Nature Communications*, *5*, 3481. doi: 10.1038/ncomms4481

Lotz, S., & Cilliers, P. (2015). A solar wind-based model of geomagnetic field fluctuations at a mid-latitude station. *Advances in Space Research*, *55*(1), 220 - 230.

Lotz, S., & Danskin, D. W. (2017). Extreme value analysis of induced geoelectric field in South Africa. *Space Weather*, *15*(10), 1347-1356.

Lotz, S., Heyns, M., & Cilliers, P. J. (2017). Regression-based forecast model of induced geoelectric field. *Space Weather*, *15*(1), 180-191.

Lu, J., Peng, Y., Wang, M., Gu, S., & Zhao, M. (2016). Support Vector Machine combined with distance correlation learning for Dst forecasting during intense geomagnetic storms. *Planetary and Space Science*, *120*, 48 - 55.

Molinski, T. S. (2002). Why utilities respect geomagnetically induced currents. *Journal of atmospheric and solar-terrestrial physics*, *64*(16), 1765–1778.

Murphy, A. H. (1977). The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situation. *Monthly Weather Review*, *105*(7), 803–816.

Ngwira, C. M., Pulkkinen, A., Leila Mays, M., Kuznetsova, M. M., Galvin, A., Simunac, K., ... Glocer, A. (2013). Simulation of the 23 July 2012 extreme space weather event: What if this extremely rare CME was Earth directed? *Space Weather*, *11*(12), 671–679.

Ngwira, C. M., Pulkkinen, A. A., Bernabeu, E., Eichner, J., Viljanen, A., & Crowley, G. (2015). Characteristics of extreme geoelectric fields and their possible causes: Localized peak enhancements. *Geophysical Research Letters*, *42*(17), 6916-6921.

Owens, M. J., Horbury, T. S., Wicks, R. T., McGregor, S. L., Savani, N. P., & Xiong, M. (2014). Ensemble downscaling in coupled solar wind-magnetosphere modeling for space weather forecasting. *Space Weather*, *12*(6), 395-405.

Owens, M. J., Lockwood, M., Barnard, L. A., Scott, C. J., Haines, C., & Macneil, A. (2021, May). Extreme Space-Weather Events and the Solar Cycle. *Solar Physics*, *296*(5), 82. doi: 10.1007/s11207-021-01831-3

Price, P. R. (2002). Geomagnetically induced current effects on transformers. *IEEE Transactions on Power Delivery*, *17*(4), 1002-1008.

Pulkkinen, A., Bernabeu, E., Thomson, A., Viljanen, A., Pirjola, R., Boteler, D., ... MacAlester, M. (2017). Geomagnetically induced currents: Science, engineering, and applications readiness. *Space Weather*, *15*(7), 828-856.

Pulkkinen, A., Hesse, M., Habib, S., Van der Zel, L., Damsky, B., Policelli, F., ... Creamer, E. (2010). Solar shield: forecasting and mitigating space weather effects on high-voltage power transmission systems. *Natural hazards*, *53*(2), 333–345.

Pulkkinen, A., Hesse, M., Kuznetsova, M., & Rastätter, L. (2007). First-principles modeling of geomagnetically induced electromagnetic fields and currents from upstream solar wind to the surface of the Earth. *Annales Geophysicae*, *25*(4), 881–893.

Pulkkinen, A., Pirjola, R., & Viljanen, A. (2007). Determination of ground conductivity and system parameters for optimal modeling of geomagnetically induced current flow in technological systems. *Earth, planets and space*, *59*(9), 999–1006.

Pulkkinen, A., Rastatter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., ... Weigel, R. (2013, June). Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations. *Space Weather*, *11*(6), 369-385.

Pulkkinen, A., Taktakishvili, A., Odstrcil, D., & Jacobs, W. (2009). Novel approach to geomagnetically induced current forecasts based on remote solar observations. *Space Weather*, *7*(8).

Pulkkinen, A., Viljanen, A., & Pirjola, R. (2006). Estimation of geomagnetically induced current levels from different input data. *Space Weather*, *4*(8).

Reiss, M. A., Möstl, C., Bailey, R. L., Rüdisser, H. T., Amerstorfer, U. V., Amerstorfer, T., ... Windisch, A. (2021, August). Machine learning for predicting the Bz magnetic field component from upstream in situ observations of solar coronal mass ejections. *arXiv e-prints*, arXiv:2108.04067.

Shepherd, S. G. (2014). Altitude-adjusted corrected geomagnetic coordinates: Definition and functional approximations. *Journal of Geophysical Research: Space Physics*, *119*(9), 7501-7521.

Sun, R., & Balch, C. (2019). Comparison between 1-D and 3-D geoelectric field methods to calculate geomagnetically induced currents: A case study. *IEEE Transactions on Power Delivery*, *34*(6), 2163-2172.

Svanda, Michal, Mourenas, Didier, Zertová, Karla, & Výbost´oková, Tatiana. (2020). Immediate and delayed responses of power lines and transformers in the Czech electric power grid to geomagnetic storms. *J. Space Weather Space Clim.*, *10*, 26.

Tasistro-Hart, A., Grayver, A., & Kuvshinov, A. (2021). Probabilistic geomagnetic storm forecasting via deep learning. *Journal of Geophysical Research: Space Physics*, *126*(1), e2020JA028228.

Torta, J. M., Serrano, L., Regué, J. R., Sánchez, A. M., & Roldán, E. (2012). Geomagnetically induced currents in a power grid of northeastern spain. *Space Weather*, *10*(6).

Viljanen, A., Pirjola, R., Prácser, E., Katkalov, J., & Wik, M. (2014). Geomagnetically induced currents in Europe - modelled occurrence in a continent-wide power grid. *J. Space Weather Space Clim.*, *4*, A09.

Weigel, R. S. (2017). A comparison of methods for estimating the geoelectric field. *Space Weather*, *15*(2), 430-440.

Weiss, A. J., Möstl, C., Amerstorfer, T., Bailey, R. L., Reiss, M. A., Hinterreiter, J., ... Bauer, M. (2021). Analysis of coronal mass ejection flux rope signatures using 3DCORE and approximate Bayesian Computation. *The Astrophysical Journal Supplement Series*, *252*(1), 9.

Welling, D. T., Ngwira, C. M., Opgenoorth, H., Haiducek, J. D., Savani, N. P., Morley, S. K., ... Liemohn, M. (2018). Recommendations for next-generation ground magnetic perturbation validation. *Space Weather*, *16*(12), 1912-1920.

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences.* Amsterdam; Boston: Elsevier Academic Press.

Wintoft, P. (2005, July). Study of the solar wind coupling to the time difference horizontal geomagnetic field. *Annales Geophysicae*, *23*(5), 1949-1957.

Wintoft, P., Viljanen, A., & Wik, M. (2016). Extreme value analysis of the time derivative of the horizontal magnetic field and computed electric field. *Annales Geophysicae*, *34*(4), 485–491.

Wintoft, P., & Wik, M. (2021). Exploring three recurrent neural network architectures for geomagnetic predictions. *Frontiers in Astronomy and Space Sciences*, *8*, 72.

Wintoft, P., Wik, M., & Viljanen, A. (2015). Solar wind driven empirical forecast models of the time derivative of the ground magnetic field. *Journal of Space Weather and Space Climate*, *5*, A7.

Zhang, J. J., Wang, C., Sun, T. R., Liu, C. M., & Wang, K. R. (2015). GIC due to storm sudden commencement in low-latitude high-voltage power network in China: Observation and simulation. *Space Weather*, *13*(10), 643-655.

Space Weather

Supporting Information for

Forecasting GICs and geoelectric fields from solar wind data using LSTMs: application in Austria

R. L. Bailey[1], R. Leonhardt[1], C. Möstl[2], C. Beggan[3], M. A. Reiss[2,3], A. Bhaskar[4], A. J. Weiss[2,5]

[1]Conrad observatory, Zentralanstalt f¨ur Meteorologie und Geodynamik, Vienna, Austria
[2]Space Research Institute, Austrian Academy of Sciences, Graz, Austria
[3]British Geological Survey, Edinburgh, UK
[4]Space Physics Laboratory, ISRO/Vikram Sarabhai Space Centre, Trivandrum, India
[5]Institute of Physics, University of Graz, Universit¨atsplatz 5, 8010 Graz, Austria

Contents of this file

Additional Supporting Information (Files uploaded separately)

None.

Introduction

This document contains supporting information for the manuscript, "Forecasting GICs and geoelectric fields from solar wind data using LSTMs: application in Austria" by Bailey, R. L. et al., submitted to *Space Weather*.

This document describes the model architecture and hyper parameters used for training LSTMs for two purposes:

1) LSTM-E: An LSTM for predicting the geoelectric field.
2) LSTM-GIC: An LSTM for predicting substation-specific GICs in the Austrian power grid.

The LSTMs were trained using the Python package *keras*. The code used to define the custom BasicAttention layer is included in this supporting information, otherwise all layers and objects referred to in the diagrams are *keras*-specific objects.

Figure S1 - LSTM-E diagram.

The general structure of the model for forecasting geoelectric fields has two branches that branch out from the initial two LSTM layers that process the input features initially. These both go into separate but identical BasicAttention layers. The left-hand side of the LSTM-E tackles a regression problem to predict the magnitude of the geoelectric field (ignoring direction), while the right-hand side of the LSTM deals with the classification problem of attempting to predict the direction of the field.

Two LSTMs of this type were trained: one for the x-component of the geoelectric field (LSTM-Ex) and one for the y-component of the geoelectric field (LSTM-Ey). Through hyper parameter tuning, different parameters were chosen for each variable.

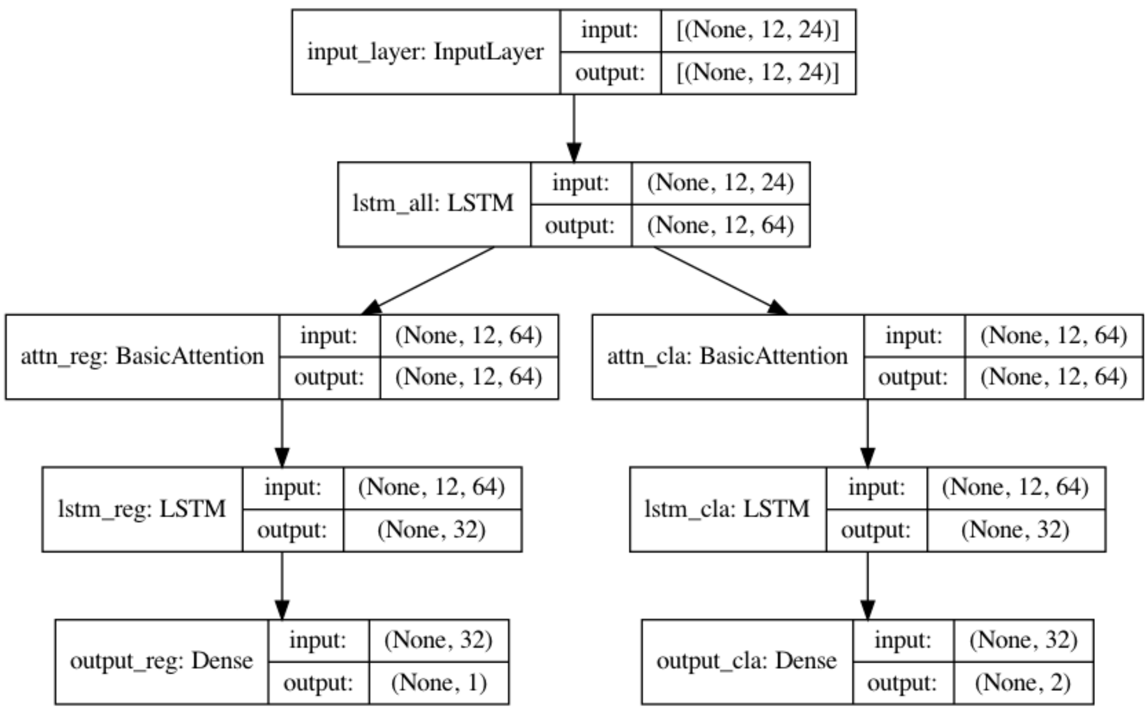

Table S2 - Hyperparameters for LSTM-E training. The hyper parameters used for training two different LSTMs: one for Ex and one for Ey.

	LSTM-Ex	**LSTM-Ey**
Loss weighting [regression, classification]	[1000, 1]	[2000, 1]
Batch size	64	32
Epochs	10	10
Fraction of LSTM dropout	0.1	0.1
Number of LSTM hidden states	32	32

Figure S3 - LSTM-GIC diagram.

In comparison to the LSTM used to predict the geoelectric field, in the case of GICs observed at specific transformers we ignore the classification problem and instead focus onto on predicting the magnitude. The samples go through one LSTM layer before being put through an Attention layer, which returns the sequences. These are fed into another LSTM layer before being reduced to a single value as output using a feed-forward Dense layer.

Two LSTMs of this type were trained: one for GICs seen in a substation near Vienna (LSTM-GIC1) and one for GICs seen in a substation near Salzburg (LSTM-GIC5). Through hyper parameter tuning, different parameters were chosen for each target variable.

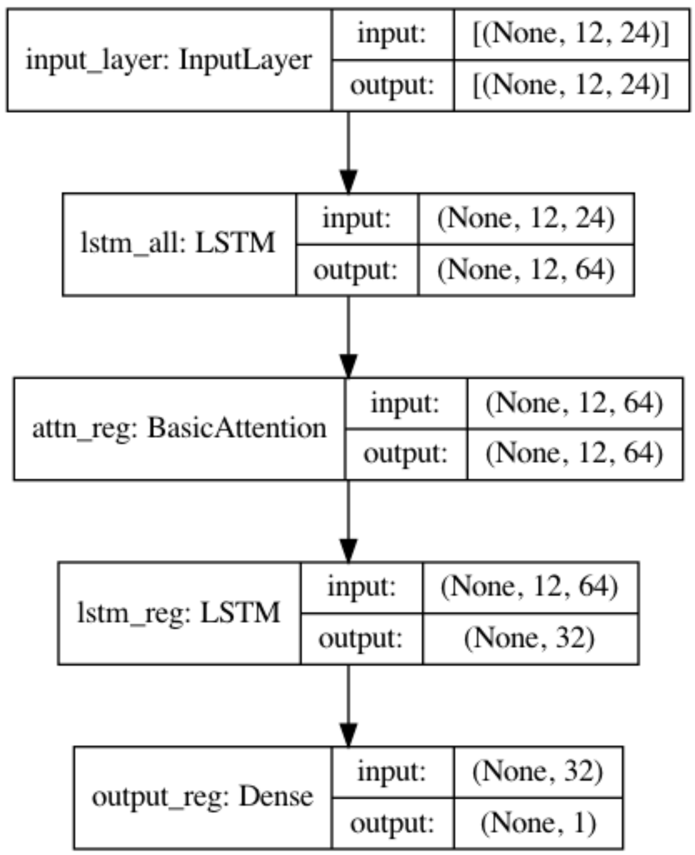

Table S4 - Hyperparameters for LSTM-GIC training. The hyper parameters used for training two different LSTMs: one for GICs at substation #1 near Vienna (GIC1) and for a power grid substation near Salzburg (GIC5).

	LSTM-GIC1	**LSTM-GIC5**
Batch size	32	64
Epochs	10	20
Fraction of LSTM dropout	0.1	0.3
Number of LSTM hidden states	32	64

Text S5 - Python Object: BasicAttention layer. This is the Python script for a custom Attention layer object included in the LSTM architecture.

```
class BasicAttention(Layer):
    '''Basic Self-Attention Layer built using this resource:
    https://towardsdatascience.com/create-your-own-custom-atten-
tion-layer-understand-all-flavours-2201b5e8be9e'''

    def __init__(self, return_sequences=True, n_units=1,
w_init='normal', b_init='zeros', **kwargs):
        self.return_sequences = return_sequences
        self.n_units = n_units
        self.w_init = w_init
        self.b_init = b_init
        super(BasicAttention,self).__init__(**kwargs)

    def build(self, input_shape):
        self.n_features = input_shape[-1]
        self.seq_len = input_shape[-2]

        self.W=self.add_weight(name="att_weight", shape=(self-
.n_features,self.n_units),
                               initializer=self.w_init)
        self.b=self.add_weight(name="att_bias", shape=(self.se-
q_len,self.n_units),
                               initializer=self.b_init)

        super(BasicAttention,self).build(input_shape)

    def call(self, x):

        e = K.tanh(K.dot(x,self.W)+self.b)
        a = K.softmax(e, axis=1)
        output = x*a

        if self.return_sequences:
            return output

        return K.sum(output, axis=1)

    def get_config(self):
        config = super(BasicAttention, self).get_config()
        config["return_sequences"] = self.return_sequences
        config["n_units"] = self.n_units
        config["w_init"] = self.w_init
        config["b_init"] = self.b_init
        #config["name"] = self.name
        return config
```

```
@classmethod
def from_config(cls, config):
    return cls(**config)
```