University of Reading

Department of Meteorology

Data Driven Approaches to Improving Space Weather Forecasts for the Power Industry

Carl Andrew Haines

A thesis submitted for the degree of Doctor of Philosophy

January 2021

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Carl Andrew Haines

Abstract

Space weather impacts technological infrastructure in space and on Earth. This thesis focuses on impacts on power systems through geomagnetic activity. During heightened geomagnetic activity, currents can be induced in power lines and cause degradation of transformers. Therefore, it is useful to forecast the severity of geomagnetic storms and the resulting geomagnetically induced currents (GICs) so that mitigating action can be taken. This thesis faces this challenge through three bodies of work. The first two bodies focus on forecasting parameters of geomagnetic storms and the third provides a statistical downscaling scheme to aid forecasting of GICs.

In the first body of work, the duration of geomagnetic storms is investigated. A statistical relationship is established between storm intensity and duration. A skilful and reliable forecast of storm duration (given storm peak intensity) is made, using log-normal distributions.

In the second body of work, two pattern-matching approaches are taken to forecast the occurrence and intensity of geomagnetic storms in geomagnetic index data. The support vector machine and analogue ensemble are implemented for an historical dataset and evaluated using several metrics. It is found that both methods are skilful with respect to climatology and the best method is dependent on the needs of the end-user.

The third body of work provides a downscaling scheme to improve the output of operational magnetospheric models such that a more realistic geoelectric field can be forecast. Using the analogue ensemble approach, a proof-of-concept study is presented which relates variability on a 1-hour timescale to a 1-minute timescale. Implemented using a perfect prognostic approach, the downscaling scheme enables a skilful estimate of geoelectric field with respect to the benchmark of no downscaling.

Acknowledgements

Firstly, I would like to thank my supervisors for mentoring me through this process as well as for the depth of intellectual input they have offered. I am grateful for their tactful guidance that has enabled me to grow in independence as a researcher.

I'm thankful to my wife who has been behind me since the very beginnings of the PhD application phase. She has celebrated my successes and been accepting when I achieve little in the working day. I'm also grateful that she has helped me build working rhythms that bring out my potential.

I would also like to thank EDF and those in the natural hazards team for supporting this work both through the internship they offered and through financial support.

Finally, I am thankful for my fellow PhD students that I have shared hours with in the coffee room, whether physical or virtual. They have taken a genuine interest in my work, offered intellectual input, given IT support and even paddled me home down the Thames. They have been a valuable source of comfort and reassurance on days with a lack of motivation or feelings of inadequacy and I count them as good friends.

Table of Contents

Chapter 1

Introduction

Short term variability in solar activity and the solar wind gives rise to space weather, a natural hazard that impacts technology and human operations, both terrestrially and in space. Our modern-day reliance on technology leaves us particularly vulnerable to space weather effects. This thesis is particularly concerned with the vulnerability of power grids. This chapter gives an introduction into the underlying physics relevant to understanding the research presented in later chapters.

1.1 The Sun, Solar Wind and Magnetosphere

Space weather impacts (see Section 1.3) that occur at Earth are due to a series of processes that begin at the Sun. Plasma and magnetic flux from the solar atmosphere flow through the heliosphere and some of this interacts with the Earth's magnetosphere leading to enhanced magnetospheric and ionospheric conditions that impact infrastructure. In this section we examine details of the Sun, the solar wind, and the magnetosphere that are necessary for understanding the research presented in this thesis. We are interested in the space weather impacts of geomagnetic storms, particularly those driven by Coronal Mass Ejections (CMEs). Within this context,

this section focuses on the solar corona upwards rather than the physical processes that lead to the coronal activity.

1.1.1 Sun

The Sun's atmosphere can be broken into several layers: the photosphere, the chromosphere, the transition region and the corona in order of lowest to highest altitude. Figure 1.1 shows the temperature and density of the solar atmosphere as a function of height from the base of the photosphere. As with Earth's atmosphere, density decreases with increasing altitude following an approximate exponential decay in the photosphere and chromosphere before a sharp drop in the transition region to a much slower decrease in the corona. Temperature varies through these layers with a local minimum through the photosphere, approximately constant in the chromosphere and a sharp rise in the transition region to the corona. Photons from the solar interior take millions of years to reach the photosphere due to repeated scattering. Protons are able to escape from the photosphere which is why it is the main source of the light we observe from the Sun and, in comparison, the chromosphere and corona are optically thin at visible wavelengths.

Sunspots appear on the photosphere as darker regions and are related to the eruption of magnetic fields through the photosphere. In these regions intense magnetic fields limit plasma convection and so the plasma cools more than surrounding plasma (Preist, 2020b). This means sunspots appear with a dark central umbra surrounded by a lighter penumbra. A typical sunspot size is 1-20 millionths of the solar hemisphere (Preist, 2020b). They appear in pairs due to formation by loop emergence from the photosphere and sometimes appear in larger groups.

The solar cycle is characterised by the rise and fall in observed sunspots (Hathaway, 2015). The length of the cycle is variable, from 9 to 14 years with a median length of 10.5 years (van Driel-Gesztelyi and

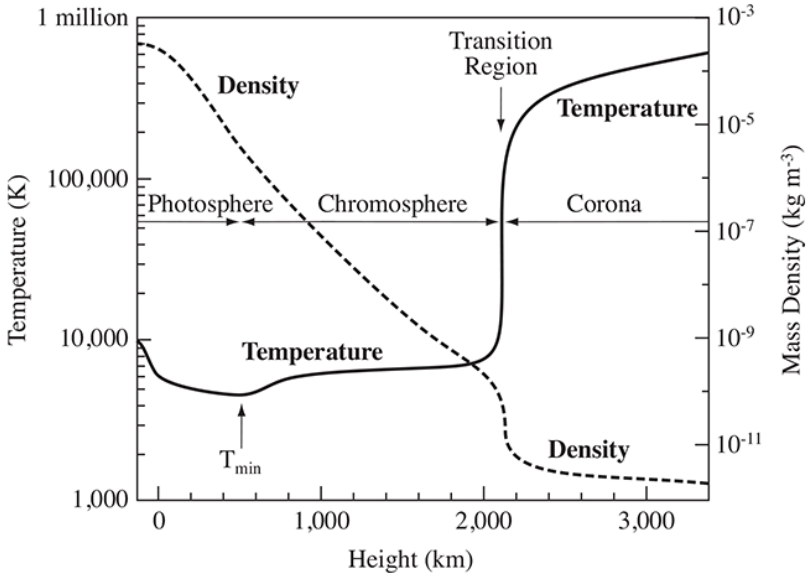


Figure 1.1: The temperature and density of the solar atmosphere at different altitudes starting at the base of the photosphere. This figure was adopted from Preist (2020b).

Owens, 2020). The cyclic behaviour of the sunspot number is shown in Figure 1.2 by the black line. The cycle is a result of the solar dynamo (Charbonneau, 2020) but the reason for the cycle period and variation in cycle period is not well understood. Each solar cycle is comprised of a minimum and a maximum with rising and declining phases between. The different phases are associated with different solar-wind properties and hence varying space weather impacts, discussed further in Section 1.3. At solar minimum the sunspot region occurs above and below the equator with a mean location around ± 25 degrees latitude, although there is significant variablitiy in latitude of individual sunspots (Ossendrijver, 2003), and the location of new sunspots drifts equator-ward through the cycle (Hathaway, 2015). The solar cycle impacts geomagnetic activity with an increase of geomagnetic activity during solar maximum. This is seen by the sunspot number and a geomagnetic index, aa, in Figure 1.2. It has also been shown that the most extreme geomagnetic storms are more likely to occur during solar maximum (Owens et al., 2021).

In the corona there are several structures that owe their existence to

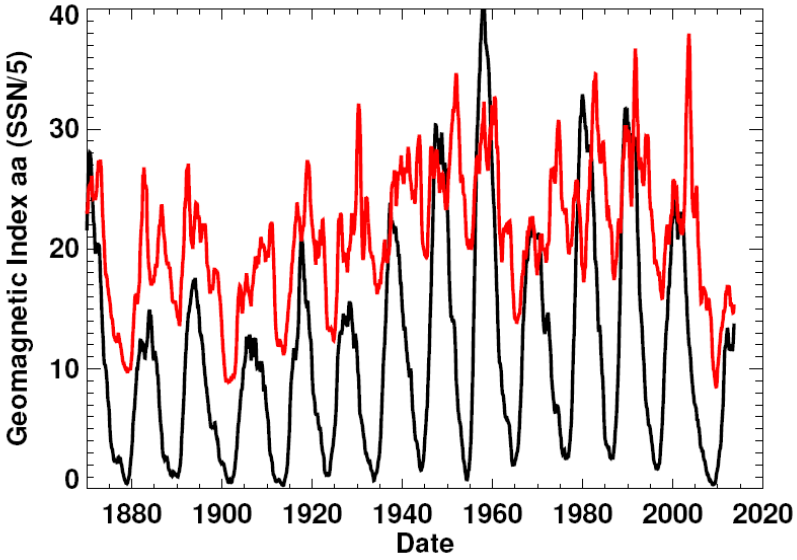


Figure 1.2: A timeseries showing the sunspot number (black) and the geomagnetic index, *aa*, (red). This figure was adopted from Hathaway (2015)

coronal magnetic field structures that are tied to the photosphere, including loops, holes and streamers. Coronal loops are magnetic field lines connecting regions of opposite polarity in the photosphere, commonly in regions of flux emergence. Coronal holes are observed as dark regions in extreme ultraviolet on the Sun due to their inactivity (Cranmer, 2009). They are areas in the corona with a magnetic field rapidly expanding into the heliosphere. This is due to an open magnetic topology allowing coronal plasma to flow out along the magnetic field lines. The solar wind, a continual flow of plasma outward from the corona, consists of fast and slow components. Coronal holes are the source of the fast solar wind (Cranmer, 2009). Coronal streamers sit above active regions and are made of closed magnetic regions topped with open regions spanning into the heliosphere. During solar minimum, streamers appear in a belt around the equator whereas at solar maximum streamers may appear in any location in the corona (Preist, 2020b). The source of the slow solar wind is still not fully understood (Kepko et al., 2016; Antiochos et al., 2011), however, there is evidence that streamers are important in the formation of the slow solar wind (Ofman, 2004).

1.1.2 Solar Wind

The solar wind is a continuous outflow of plasma from the solar corona. It is mainly made up of protons and electrons with some heavier ions. Primarily, structure is divided into "fast" wind originating in the photosphere of coronal holes (Cranmer, 2009; Preist, 2020b) and "slow" wind, the origin of which is still being fully researched (Kepko et al., 2016; Antiochos et al., 2011).

Typical properties of fast and slow solar wind are shown in Table 1.3. As well as greater radial velocity, we see that the fast wind is less dense, hotter and has a weaker associated magnetic field when contrasted to the slow solar wind. The geoeffectiveness of the solar wind is dependant on its density, magnetic field strength, and radial velocity (Owens, Lockwood and Barnard, 2018). The solar wind temperature does not impact geoeffectiveness but is correlated with radial velocity. This geoeffectiveness, G, can be quantified with a coupling function commonly in the form of

$$G = n^{2/3-\alpha} B^{2\alpha} V^{7/3-2\alpha} sin^4(\theta/2), \tag{1.1}$$

where n is the solar wind density, B is the HMF intensity, V is the radial solar wind speed and θ is the HMF non-radial clock angle. Lockwood et al. (2017) empirically estimated the parameter α to be approximately 0.5.

The solar wind has a frozen-in magnetic field, called both the heliospheric magnetic field (HMF) and interplanetary magnetic field (IMF) but referred to here as HMF, which originates from the coronal field and extends out as the solar wind expands through the heliosphere. The HMF has a spiral structure, due to solar wind plasma flowing radially, whilst the foot points of the HMF field lines are fixed in the rotating solar atmosphere. This is known as the Parker spiral (Parker, 1959), depicted in

	Coronal hole	Streamer belt
$V_R{}^a$ km s^{-1}	750	300-400
$n_P{}^b$ cm^{-3}	2-4	5-10
$T_P{}^c$ [x10^5 K]	2-3	0.5-1
$\|B\|^d$ [nT]	3-4	4-8

Figure 1.3: A table of the properties of fast and slow solar wind. Listed is the radial velocity, density, temperature and magnetic field magnitude. The fast solar wind originates in coronal holes and the slow solar wind in the streamer belt. This figure was adopted from Owens (2020).

the ecliptic plane in Figure 1.4 although the spiral also exists outside the ecliptic plane. The angle of the HMF relative to the solar radial direction is a function of the distance from the Sun and the solar wind speed. In a reference frame fixed with the Sun's rotation, such that a coronal source of plasma is fixed, the plasma has a velocity component in both the radial direction, r, and the solar longitudinal direction, ψ. The Parker spiral for a source at ψ_0 and R can be written as

$$r - R = a\frac{u}{\omega}(\psi - \psi_0), \tag{1.2}$$

where u is the solar wind speed and ω is the Sun's angular velocity (Hundhausen, 1995). A typical angle at Earth (1 AU) for a solar wind with speed 400 km/s is 45°. The Parker spiral has been shown to exist by several spacecraft, although transients such as CMEs and Co-rotating interaction regions (CIRs) interrupt the idealised pattern.

CIRs (Richardson, 2018; Gosling and Pizzo, 1999) occur where fast and slow solar wind interact along the same radial line due to the Sun's rotation as shown in Figure 1.5. A fast wind stream (shown in red) catches up with the slow wind stream (shown in blue) but are kept separate according to the frozen flux theorem because of their differing magnetic fields. A shock wave propagates forward into the slow wind and a reverse wave back into the fast wind. The result is a region of increased magnetic field intensity and increased plasma density, both of which in-

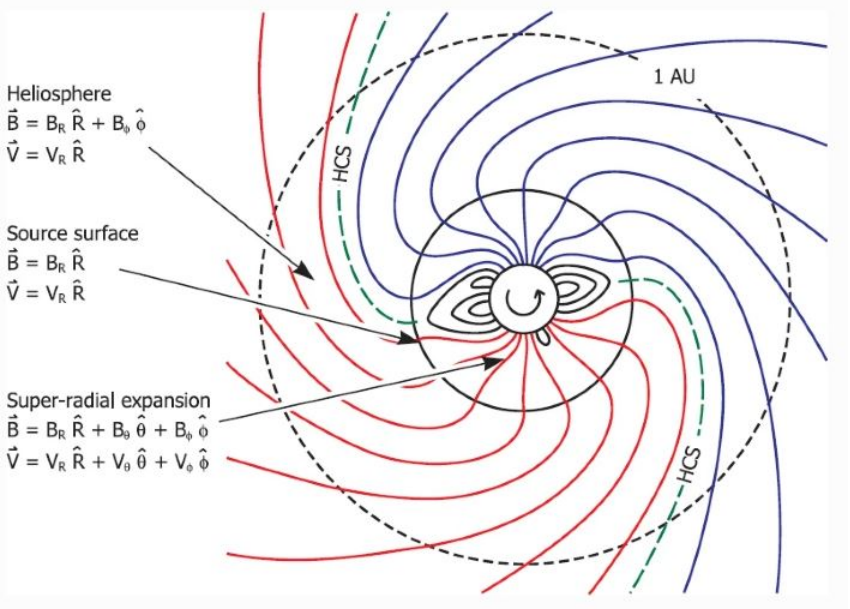


Figure 1.4: A schematic of the HMF in the ecliptic plane. The parker spiral is revealed by the spiralling shape of the magnetic field lines. This figure is adopted from Owens and Forsyth (2013).

crease the geoeffectiveness of the solar wind, as can be understood through Equation 1.1. The stream interface is also inclined to the solar-equatorial plane, so as the field is compressed it deflects out of the plane, producing magnetic field in the B_z direction and this is the major cause of CIR geoeffectiveness. As the fast wind continues to advance it leaves a rarefied region as it outpaces the slow wind behind it. In the Earth's reference frame, an observer will see these interaction regions approximately every 27-day solar rotation period because the coronal holes creating the fast stream are usually long lasting and hence interaction regions are said to co-rotate with the Sun.

CIRs can drive major geomagnetic storms. The "Living with a Star" Coordinated Data Analysis Workshop in March 2005 analysed 79 major geomagnetic storms and found that nine were caused by CIRs. Richardson et al. (2006) found that the maximum size of storms observed due to CIRs was -161 nT in the Dst index (see Section 2.2), a significant event. Due to previously theorised estimates of minimum Dst caused by CIRs, it is unlikely that a storm would far exceed this without being co-driven by

a CME.

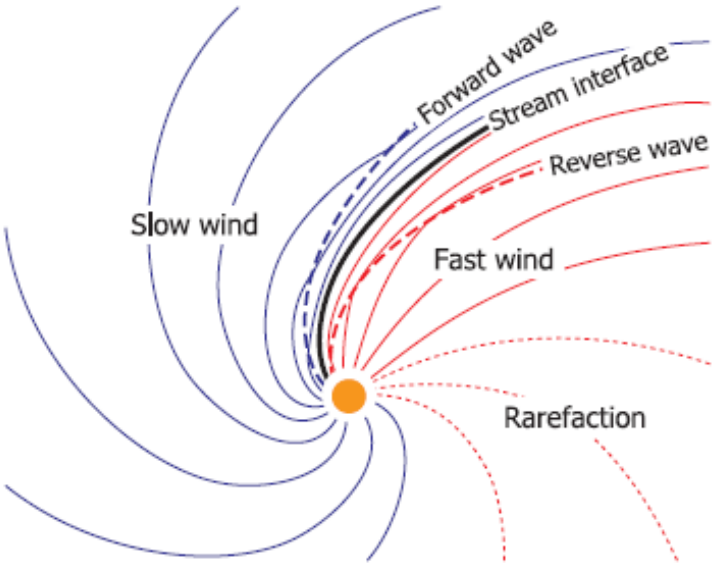


Figure 1.5: The formation of CIRs looking down on the solar-ecliptic plane. The red lines show fast solar wind and the blue show slow. At the front of the fast wind is a compressed region and behind is a rarefied region. A shock wave is shown at the front of the fast wind travelling forward and one travelling back. This figure was adopted from Owens (2020).

Coronal mass ejections (CMEs) are large structures of plasma and magnetic fields that are propelled into the heliosphere by the magnetic field (Preist, 2020b; Kilpua, Koskinen and Pulkkinen, 2017; Webb and Howard, 2012). These large-scale transients are ejected into the background solar wind and cause the most impactful space weather events. Figure 1.6 (left) shows a schematic of a CME flowing into the heliosphere and travelling towards Earth. The CME itself is shown as the large orange loop protruding from the Sun. Its magnetic field is a flux rope, a twisted magnetic field generated by interactions of magnetic loops in the Sun's atmosphere via magnetic reconnection, hereafter referred to as simply reconnection (Preist, 2020b). Ahead of a high speed CME is a shock caused by the movement of the CME through the background solar wind which is travelling at a slower speed. Behind the shock is a sheath region in which plasma and the associated magnetic field is compressed. Figure 1.6 (right)

shows four observed properties of a CME from August 2014 by the Advanced Composition Explorer (ACE) space craft at the L1 Lagrange point (Stone, Frandsen and Mewaldt, 1998). We see that compared to the background solar wind, the sheath region has increased magnetic field magnitude, similarly variable magnetic field angles and increased radial speed. When looking at the CME compared with the background solar wind we see a large increase in magnetic field magnitude followed by a slow decrease back to that of the ambient solar wind. We see magnetic field angles that vary on a lower frequency. The reason we see less point-to-point variability and a systematic trend in the magnetic field angles is because there is a stronger and more coherent magnetic field structure inside the CME. Finally the radial speed begins as that of the sheath region and decreases down to the ambient solar wind. The field is southward at the start and end of this CME so we would expect stronger coupling with the magnetosphere during these times.

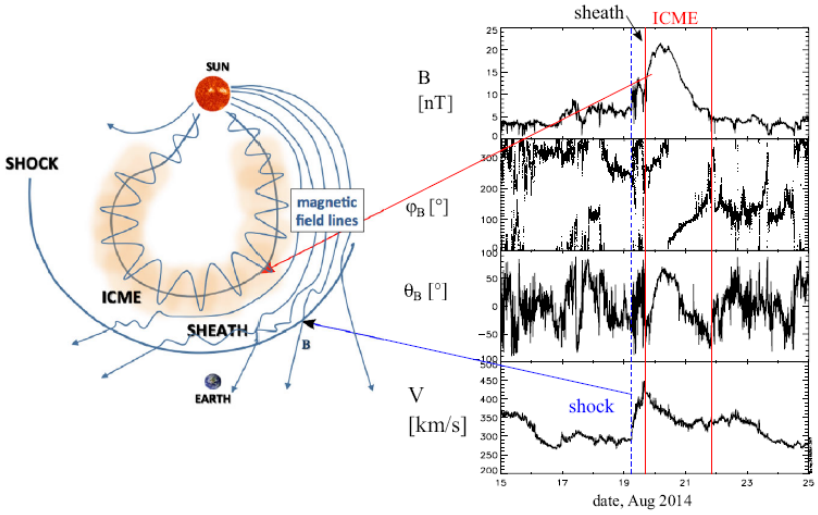


Figure 1.6: A CME travelling through the heliosphere. Left) is a schematic of a CME travelling towards Earth. The magnetic field lines are shown in blue. A shock is shown proceeding the CME with a sheath region behind it. Right) shows solar wind observations of a CME in August 2014 from the ACE spacecraft. Shown from top to bottom is the magnetic field magnitude, magnetic field latitude and longitude angles in geocentric solar magnetospheric coordinates and solar wind speed in the radial direction. This figure was adopted from Kilpua, Koskinen and Pulkkinen (2017).

When there is sufficient shear between the magnetic fields of the solar wind and the magnetosphere, reconnection can occur, coupling together the solar wind and magnetosphere. Due to the structure of Earth's magnetic field, these conditions typically occur when the solar wind magnetic field has a strong southward component, Bz, in geocentric solar magnetospheric reference frame. When enhanced day-side reconnection occurs, the Dungey cycle begins, leading to enhanced geomagnetic activity. Solar wind parameters and geomagnetic indices are shown in Figure 1.7 for an event as an illustrative example. The enhanced speed and magnetic field magnitude in a CME increases the rate of reconnection and hence increase the energy transfer into the magnetosphere, as per Equation 1.1. If, however, B_Z is northward, there will be less reconnection but the CME impact can still compress the magnetosphere which leads to increased geomagnetic activity, albeit less so than when there is enhanced dayside reconnection. The speed of CMEs means that they usually arrive at the Earth two to five days after they can be first observed at the Sun (Preist, 2020b). However, current observational capabilities mean that we cannot obtain reliable observations until the CME reaches the L1 point giving approximately 30-60 minutes warning of its geoeffectiveness (Morley, 2020; Pulkkinen and Rastatter, 2009).

1.1.3 Earth's Magnetosphere

Planets with a strong magnetic field, such as the Earth, will carve out a cavity in the heliosphere that restricts the access of solar wind plasma and the HMF. Such a cavity is called a magnetosphere. The cavity forms in the solar wind because the very high electrical conductivity of the plasmas in space and their very large spatial scales mean that the magnetic field is "frozen-in" to the plasma and does not diffuse through it. This means the plasma in the magnetosphere is constrained by Earth's intrinsic magnetic field and the plasma in interplanetary space is constrained by the exten-

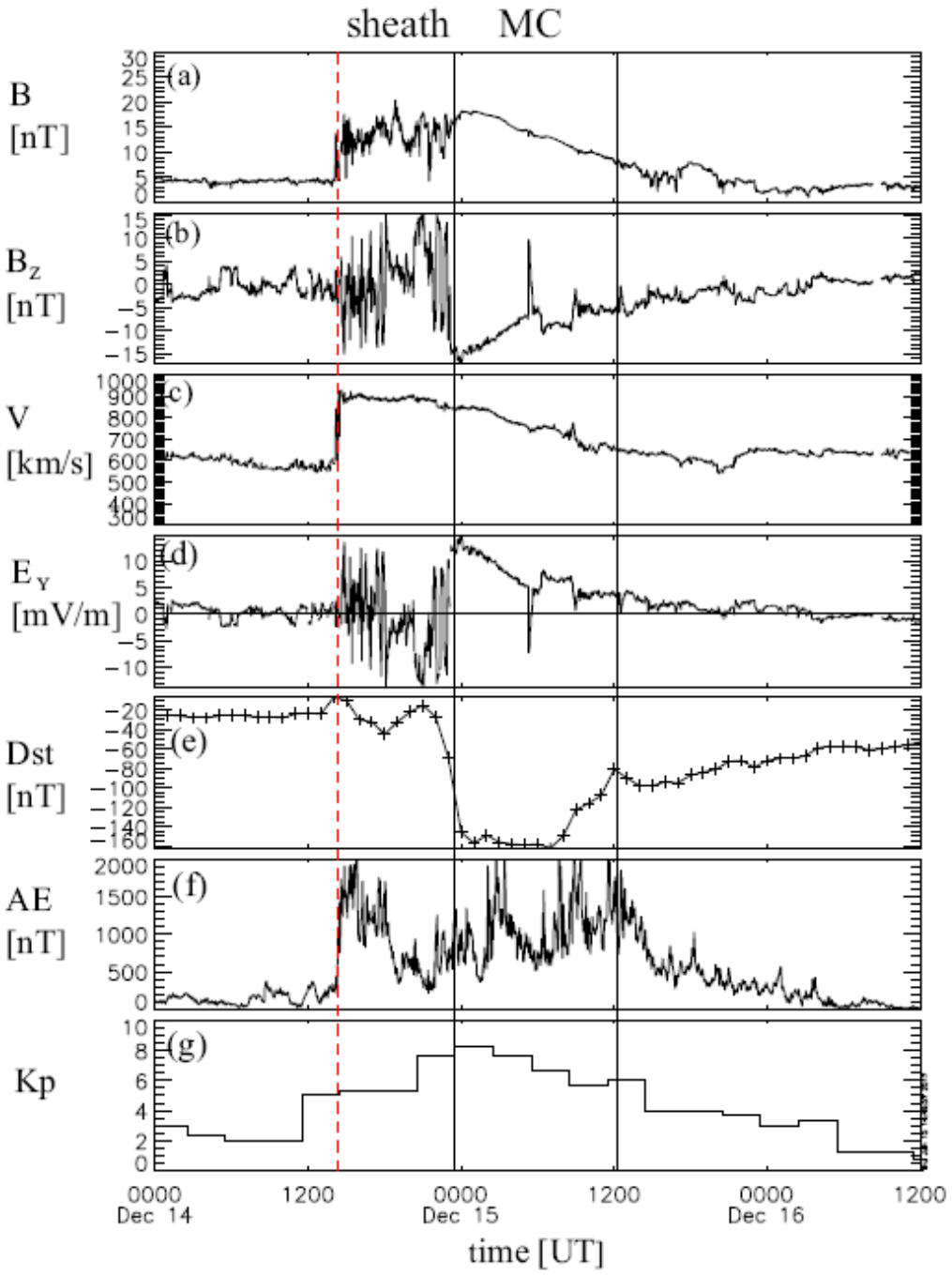


Figure 1.7: A timeseries of solar wind parameters and geomagnetic indices as a CME passes through. This figure was adopted from Kilpua, Koskinen and Pulkkinen (2017).

sion of the solar magnetic field there. To first order, the two cannot mix although breakdowns of the frozen-in flux theorem, in particular reconnection, allow a degree of mixing of solar wind and magnetospheric plasmas. The boundary will form where a balance is reached between the pressure

of the Earth's magnetosphere, where the magnetic pressure dominates, and the solar wind pressure, dominated by the dynamic pressure. This is the magnetopause.

The magnetopause stand off distance can be estimated by comparing the dynamic pressure of the solar wind to the magnetic pressure of Earth's magnetic field by

$$(\rho v^2)_{sw} = \left(\frac{4B(r)^2}{2\mu_0}\right)_m, \tag{1.3}$$

where ρ and v are the density and velocity of the solar wind, $B(r)$ is the magnetic field of the Earth and μ_0 is the permeability of free space. The magnetopause stand off distance corresponds to the value of r that balances this equation. Approximating the intensity of Earth's magnetic field as $B(r) = B_0/r^3$, where B_0 is the magnetic moment, we can express r as

$$r = \sqrt[6]{\frac{2B_0^2}{\mu_0 \rho v^2}}. \tag{1.4}$$

With typical solar wind values the magnetosphere reaches out to approximately 10 R_E (Earth radii) on the day side and is swept out to hundreds of R_E anti-sunward on the night side. This shape is shown in the schematic in Figure 1.8.

Figure 1.8 displays and annotates several features of the magnetosphere. The bow shock is a shock formed as the supersonic and super-Alfvénic solar wind approaches the magnetosphere and is responsible for reducing the flow to subsonic speeds and to near stationary in the Earth's frame at the nose of the magnetosphere. The solar wind plasma travels at speeds greater than the rate at which information is internally transmitted, i.e. the Alfvén Mach number is greater than 1. The bow shock commonly has a mach number of M_A $5-10$ (Sundberg et al., 2017). As the solar wind flow is super Alfvénic, information on the presence of the magnetosphere cannot be communicated to the solar wind flow via waves and instabilities. Consequently, a magnetohydrodynamic (MHD) shock must

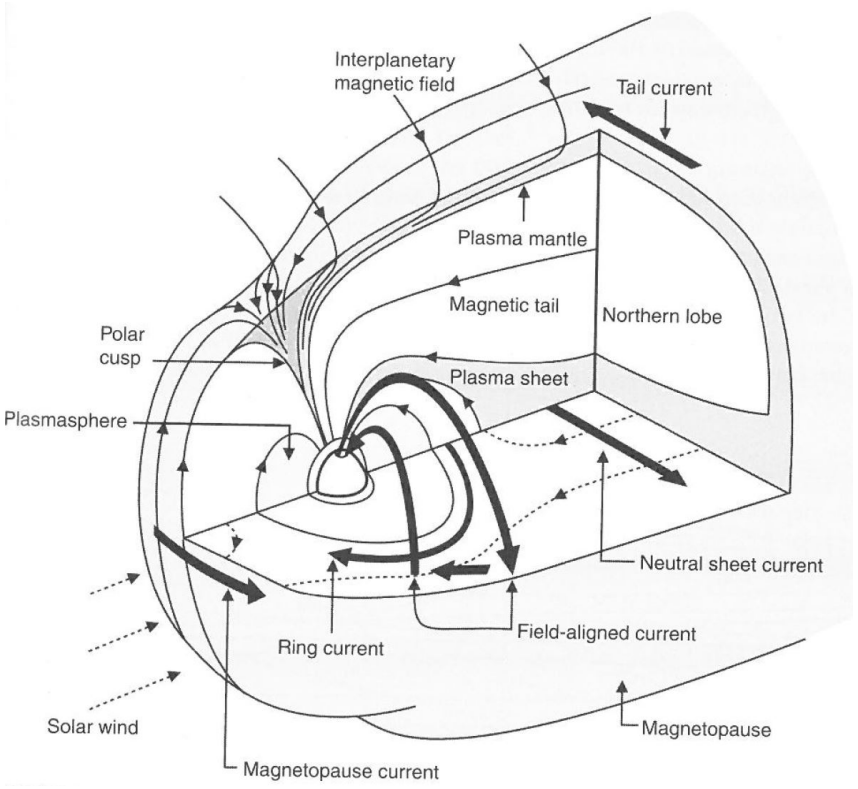


Figure 1.8: A schematic showing the structure of the magnetosphere in the presence of solar wind where the Sun is out of shot to the left of the schematic. This figure was adopted from Russell, Luhmann and Strangeway (2016).

form, in which the solar wind plasma is compressed, slowed and heated behind the bow shock in a region called the magnetosheath. The magnetosheath flow is brought to rest at the nose of the magnetosphere, but flows antisunward around the magnetosphere and eventually becomes indistinguishable from the unshocked solar wind.

Both the bow shock and the magnetopause, the boundary between the magnetosphere and near-Earth plasma, move Earthward as the dynamic pressure of the solar wind is enhanced. Although the magnetopause keeps the two sets of plasma separate, some of the solar wind plasma is able to penetrate through reconnection and viscous interactions at the interface (Arridge, 2020). As a consequence, magnetospheric current systems become enhanced. This is discussed in more detail in Section 2.1.

The Dungey Cycle

This section examines the Dungey cycle, a set of processes caused by a southward magnetic field in the HMF that lead to substorms in the magnetosphere (Dungey, 1961). A schematic for this cycle is shown in Figure 1.9.

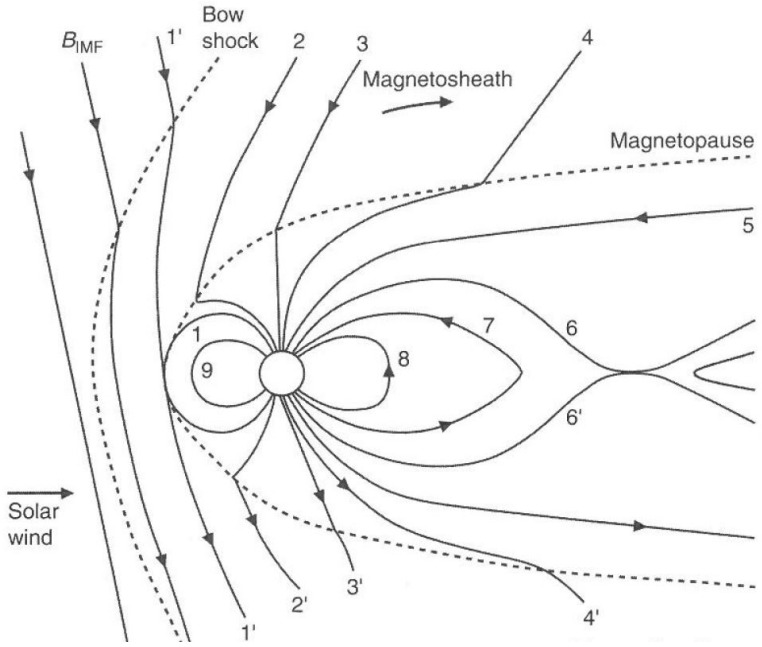


Figure 1.9: A schematic of the magnetosphere being driven by a southward HMF to retell the Dungey cycle. This figure was adopted from Russell, Luhmann and Strangeway (2016).

Due to reconnection on the day side of the magnetosphere (1 in Figure 1.9), some of the interacting plasma is transported to the area of the magnetosphere on the night side of the earth called the magnetotail (5). Here, the magnetic field intensifies as open field lines generated by reconnection accumulate in the tail, raising the magnetic pressure in the tail lobes which thins the cross-tail current sheet that separates them. As the current sheet narrows there is further reconnection in the cross-tail current sheet (6) which re-closes open field lines giving an explosive release of plasma away from the reconnection site. Thus, some is transported down the tail away from Earth and some travels down the magnetospheric field lines into the poles (7). The newly reconnected field lines (8) con-

vect around to the dayside (9). These so called substorms give us the auroras but can also cause GICs. Intense geomagnetic storms are accompanied by intense and frequent substorms, however, intense substorms can occur in the absence of an intense geomagnetic storm (Gonzalez et al., 1994).

Substorms are made up of a growth, expansion and recovery phase (Russell, Luhmann and Strangeway, 2016). During the growth phase, energy is stored in the magnetotail lobes. In the far tail, the lobe magnetic field is set by the static pressure of the solar wind. Therefore, as more flux is stored in the lobe, the tail radius increases to maintain pressure balance with the solar wind. Closer to the Earth, in the near- and mid-tail regions, the tail flares and so there is a component of the solar wind dynamic pressure that acts perpendicular to the magnetopause. This limits the tail expansion and so the magnetic energy density rises in the lobes in the near- and mid-tail regions. During the expansion phase the lobe field decreases as reconnection in the tail closes open field lines and closed field lines move sun-ward and out of the tail under the magnetohydrodynamic magnetic curvature ("tension") force. The energy that was stored is then sent to the tail plasma sheet, the ring current, and the ionosphere. Finally, in the recovery phase the tail reconnection rate drops and the tail stops losing energy. If there is a lower day-side reconnection rate and the tail reconnection tends towards it then the magnetosphere heads towards a quiet state. If, however, the day-side reconnection rate remains high, then a new substorm growth phase begins when the tail reconnection falls below the day-side reconnection rate.

Between the tail lobes is the cross tail current sheet. The current sheet intensifies near the earth as the tail lobe magnetic field increases. The enhanced current close to the Earth is disrupted at the start of the expansion phase. We do not currently fully understand the reason for this disruption and it is an active area of research (Pritchett and Lu, 2018).

What we do know is that the disruption occurs closer to the earth than the tail X-line. The disruption causes the current to flow down field lines into the auroral ionosphere instead of flowing across the tail. The current then flows along the highly-conducting auroral ionosphere and back up field lines to complete the circuit. This current system is the substorm current wedge (see Section 2.1) and the part that flows along the night side auroral ionosphere is the auroral electrojet (see Section 2.1).

1.2 Single Particle Motion and MHD

Plasma is the fourth state of matter complementing solids, liquids and gasses. They are an ionised, conductive gas made up of ions and electrons in approximate balance resulting in quasi-neutrality. The solar wind plasma is a rarefied plasma predominantly containing hydrogen and helium ions along with small amounts of heavier ions. Due to its rarity, collisions between particles are sufficiently infrequent that they are not dynamically important in controlling the evolution of the plasma and it can be treated as collisionless. This allows us to understand the evolution of a single prticle in terms of the net force acting on it, including the electromagnetic field within the plasma. Alternatively, plasmas can be considered as magnetised fluids and their motion understood using MHD (see Section 1.2.2).

1.2.1 Single Particle Motion

We now look at the equations that govern single particle motion. We first look at the Lorentz force

$$F = q[\mathbf{E} + (\mathbf{v} \times \mathbf{B})], \tag{1.5}$$

where **E** is the electric field, **v** is particle velocity, **B** is the magnetic field, and q is the charge on the particle. The Lorentz force governs the motion of charged particles in the presence of electromagnetic fields.

Using Newton's Law of Motion, the motion of a particle under the influence of the Lorentz force is given by

$$q[\mathbf{E}+(\mathbf{v}\times\mathbf{B})]=m\frac{d\mathbf{v}}{dt}, \tag{1.6}$$

where m is the particle mass. In the absence of an electric field and presence of a uniform magnetic field this becomes

$$q[(\mathbf{v}\times\mathbf{B})]=m\frac{d\mathbf{v}}{dt}. \tag{1.7}$$

Taking the dot product of both sides with **v** leads to

$$m\frac{d\mathbf{v}}{dt}\cdot\mathbf{v}=\frac{d\frac{m\mathbf{v}^2}{2}}{dt}=0. \tag{1.8}$$

Therefore, the kinetic energy is constant meaning no work is done on the particle in the absence of an electric field.

Now, consider a particle in a Cartesian coordinate system, with unit vectors i, j and k, with a magnetic field, **B**, aligning the coordinate system such that **B** = Bk, we can write equation 1.7 in its three components as

$$\begin{aligned}\frac{dV_x}{dt}&=\frac{q}{m}V_yB,\\ \frac{dV_y}{dt}&=-\frac{q}{m}V_xB,\\ \frac{dV_z}{dt}&=0.\end{aligned} \tag{1.9}$$

These equations represent circular motion in the x-y plane, with constant speed along the field aligned direction. The appearance of q in the first two of these equations shows that the direction of circular motion will

depend on the charge of the particle so positive ions and electrons will move in opposite directions.

Now we introduce an electric field that has components parallel and perpendicular to the magnetic field such that

$$\mathbf{E} = \mathbf{E}_{\parallel} + \mathbf{E}_{\perp}. \tag{1.10}$$

The parallel component of equation 1.6 is

$$q[\mathbf{E} + (\mathbf{v} \times \mathbf{B})]_{\parallel} = m(\frac{d\mathbf{v}}{dt})_{\parallel}. \tag{1.11}$$

The parallel component of $\mathbf{v} \times \mathbf{B}$ is zero so

$$q\mathbf{E}_{\parallel} = m(\frac{d\mathbf{v}}{dt})_{\parallel}. \tag{1.12}$$

Integrating with respect to time gives

$$V(t)_{\parallel} = \frac{q}{m}\mathbf{E}_{\parallel}t + V_0, \tag{1.13}$$

where V_0 is the field aligned velocity of the particle at time $t = 0$. This shows that in the presence of an electric field with a component parallel to the magnetic field, the particle is accelerated along the field line with direction determined by its charge.

Similarly, when there is a component of the electric field perpendicular to the magnetic field, there is acceleration. In this case, the perpendicular component of equation 1.6 is

$$q[\mathbf{E}_{\perp} + (\mathbf{v}_{\perp} \times \mathbf{B})] = m\frac{d\mathbf{v}_{\perp}}{dt}. \tag{1.14}$$

By separating the velocity into gyrational and drift components, it can be

shown that

$$V_{\perp\ drift} = \frac{\mathbf{E}_{\perp} \times \mathbf{B}}{B^2}. \tag{1.15}$$

Here we see that the drift velocity is not dependent on charge so ions and electrons move in the same direction, therefore, a current is not created.

Gradient Drift

In the previous section we discussed particles in a uniform magnetic field. However, it is common for a magnetic field to vary spatially and have a gradient. As a particle gyrates in a non-uniform magnetic field, it experiences different field strengths in different parts of its gyration. In areas with a stronger field the path will follow a tighter path and a weaker field gives a looser path. Figure 1.10 shows the path of an ion and an electron in a magnetic gradient. The combination of tighter and looser spirals creates an overall drift in opposing directions.

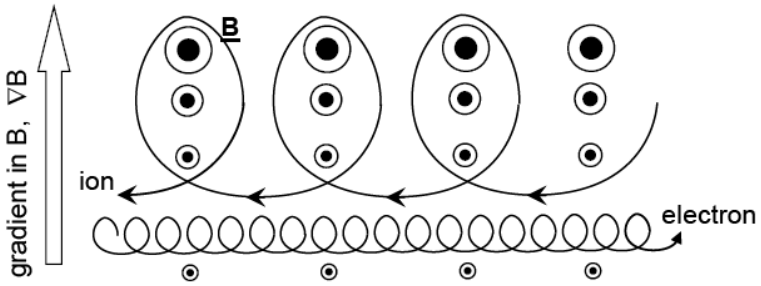


Figure 1.10: A schematic showing gradient drift of an ion and electron in the presence of a magnetic gradient. This figure has been adopted from Lockwood (2019b).

Magnetic Mirror

The magnetic mirror phenomenon occurs when a particle is gyrating around and travelling along a magnetic field and experiencing a gradient in the

magnetic field as depicted in Figure 1.11. As the particle travels through increasing magnetic field strength there is a force applied which acts against its direction of travel.

The forces perpendicular and parallel to the magnetic field on a gyrating particle are

$$F_{\perp} = \mathbf{F}cos(\theta) \tag{1.16}$$

and

$$F_{\parallel} = \mathbf{F}sin(\theta), \tag{1.17}$$

where θ is the angle between the particle's guiding centre and its location in its gyration as shown in Figure 1.11. By dividing these equations we get and rewriting the tan function we get

$$F_{\parallel} = F_{\perp}tan(\theta) = f_{\perp}\frac{dr}{dz}, \tag{1.18}$$

where r is the gyration radius and z is the direction parallel to the magnetic field. We can use that the perpendicular force is the centrifugal force $F_{\perp} = mv_{\perp}^2/r$ to get

$$F_{\parallel} = \frac{mv_{\perp}^2}{r}\frac{dr}{dz}. \tag{1.19}$$

Now, conserving the magnetic flux through a circular cross section we have that

$$c = B_z A = B_z \pi r^2, \tag{1.20}$$

where c is a constant. Then differentiating and rearranging we have

$$\frac{dB_z}{dz} = -2\frac{c}{\pi r^3} = -2\frac{B_z}{r}. \tag{1.21}$$

Rearranging and substituting into 1.19 we get

$$F_{\parallel} = -\frac{mv^2}{2B_z}\frac{dB_z}{dz} \tag{1.22}$$

showing that the force is in the opposite direction to an increasing magnetic gradient. Here, $\frac{mv^2}{2B_z} = \mu_m$ the magnetic moment.

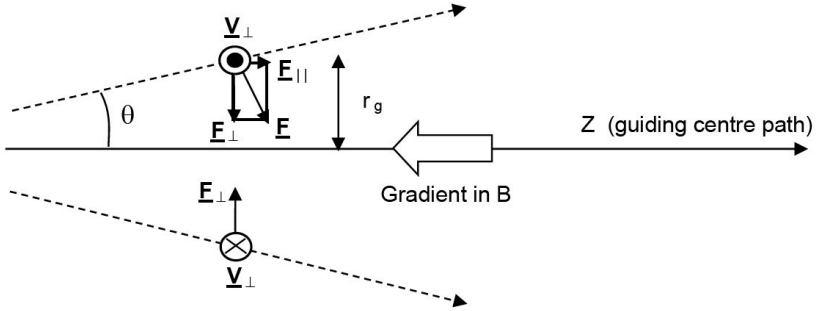


Figure 1.11: A schematic of a particle travelling through a non-uniform magnetic field and gyrating around the field line. This figure has been adopted from Lockwood (2019b).

Curvature Drift

A particle travelling along a geomagnetic field line, bouncing between two mirror points, travels along a curve and thus experiences a centrifugal force perpendicular to the magnetic field. In this situation, the Lorentz force (Equation 1.5) perpendicular to the magnetic field can be simplified to

$$\mathbf{F}_\perp = q\mathbf{E}_\perp. \tag{1.23}$$

We can use this to apply a centrifugal force to Equation 1.14 giving

$$\mathbf{V}_{\perp\mathbf{F}} = \frac{(\mathbf{F} \times \mathbf{B})}{qB^2}. \tag{1.24}$$

This gives a velocity component perpendicular the the magnetic field and centrifugal force with the direction dependent on the sign of the charged particle. Therefore, under curvature drift, electrons will move eastward and positive ions will move westward.

1.2.2 Magnetohydrodynamics (MHD)

In some situations it is useful to model plasma as a magnetohydrodynamic (MHD) fluid. MHD is based on the equations of hydrodynamics, with the additional constraints of solving for evolution of the electromagnetic field by incorporating Maxwell's equations (Preist, 2020a). In 3-D ideal MHD, where all dissipative processes (finite viscosity, electrical resistivity and thermal conductivity) are neglected, there are eight MHD equations that accompany the eight unknowns. These unknowns are mass density, ρ, the isotropic thermal pressure, P, the bulk velocity, $\mathbf{v}$, and the magnetic field $\mathbf{B}$ where $\mathbf{v}$ and $\mathbf{B}$ are three dimensional vectors. The equations are

$$\begin{aligned}
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) &= 0, && \text{Mass continuity eqn.} \\
\rho \frac{\partial \mathbf{v}}{\partial t} &= -\nabla P - \frac{1}{\mu_0} \mathbf{B} \times (\nabla \times \mathbf{B}), && \text{Eqn. of motion} \\
\frac{d}{dt}\left(\frac{P}{\rho^{\gamma}}\right) &= 0, && \text{Energy eqn.} \\
\frac{\partial \mathbf{B}}{\partial t} &= \nabla \times (\mathbf{v} \times \mathbf{B}), && \text{Induction eqn.}
\end{aligned} \tag{1.25}$$

where γ is the ratio of specific heats which is $5/3$ for an ideal gas.

The induction equation and the equation of motion are key to understanding the motion of a plasma that exists in an ideal MHD domain. The induction equation relates the temporal changes in the magnetic field to the presence of a velocity field. The equation of motion relates how the plasma velocity varies with time to the forces placed on it. These forces are the pressure gradient, magnetic pressure, and magnetic tension.

In an ideal plasma, the magnetic flux, ψ, through a surface which moves with the plasma is conserved. We can see this by looking at a surface, S, that is bound by a curve, C. We have that

$$\psi = \int \int_S \mathbf{B} \cdot d\mathbf{S}. \tag{1.26}$$

A temporal change in magnetic flux can be caused either by a change in $\mathbf{B}$ with S fixed or by movement of the curve, C with the plasma. This gives that the time derivative of the magnetic flux is

$$\frac{d\psi}{dt} = \int\int_S \frac{\partial \mathbf{B}}{\partial t} \cdot ds + \oint_C \mathbf{B} \cdot \mathbf{v} \times d\mathbf{l}, \tag{1.27}$$

where $d\mathbf{l}$ is a small increment parallel to C. The second term on the right hand side can be transformed using Stokes' theorem such that we have

$$\frac{d\psi}{dt} = \int\int_S \left(\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{v} \times \mathbf{B}) \right) \cdot d\mathbf{S}. \tag{1.28}$$

From this it is easy to see that, by using the induction equation, the temporal change in flux is zero so the flux is conserved. This is what gives us the frozen-in flux theorem which shows that when plasma moves the magnetic field is dragged with it and when the magnetic field is moved so is the plasma.

1.3 Space Weather Impacts

Changes in the solar wind conditions near-Earth can give rise to space weather impacts which has adverse effects on space- and ground-based technologies. These impacts mainly cause damage and reliability issues to technological systems but also pose a direct threat to human safety in certain circumstances such as solar energetic particle events or large galactic cosmic rays (Tobiska et al., 2015; Wilson et al., 2003). Space weather impacts are a relatively new hazard because it is only in approximately the last 150 years that society's technology has become susceptible to space weather, as discussed by Hapgood (2010). Figure 1.12 shows a range of space weather phenomena and their resulting impacts, several of which are examined in this section. For a description of the impact on power systems see Section 2.3.

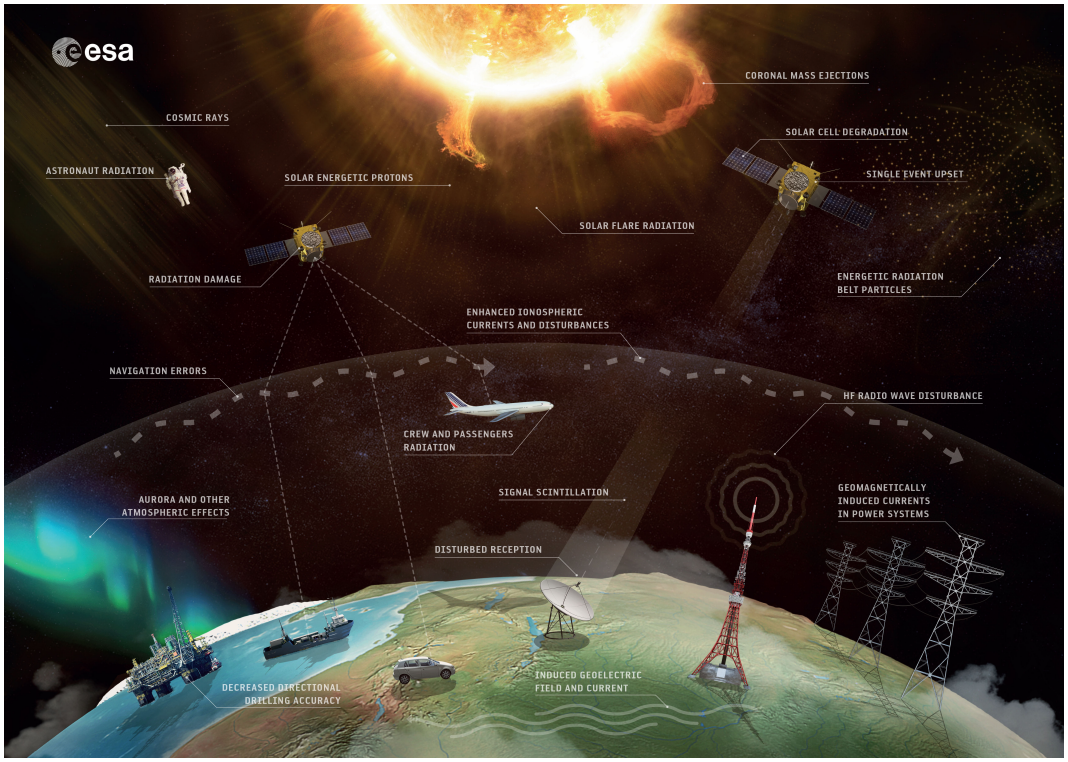


Figure 1.12: A schematic showing a range of space weather phenomena and their resulting impacts. Copyright ESA/Science Office, CC BY-SA 3.0 IGO

Solar energetic particles (SEPs) are high energy charged particles due to solar activity. SEPs have the potential to cause malfunctions in electronic equipment. One type of impact is single event upsets which occur when a charged particle ionises a sensitive part of a circuit and causes it to change state (Pulkkinen, 2007). Commonly this results in a system reboot although can occasionally lead to more permanent damage. Additionally on spacecraft, SEPs can degrade solar panels which can reduce their functionality and life span. SEPs can penetrate the Earth's atmosphere, however many are filtered out by geomagnetic and atmospheric shielding so this impact is more prevalent in aeroplanes and satellites than devices on the ground. SEPs also pose a risk to the health of airline crew and passengers (Malandraki and Crosby, 2018; Cannon et al., 2013). In a single flight during a space weather event, passengers and crew could receive 20 times the recommended annual dose of radiation, a particular concern for crew and frequent flyers (Tobiska et al., 2015; Wilson et al.,

2003). Reynolds et al. (2002) found a significant increase in melanoma and breast cancer in flight attendants.

During a space weather event global navigation satellite systems (GNSS) may become unusable (Sreeja, 2016). In ordinary conditions, GNSS allows for ionospheric plasma to scatter and refract the signal between a satellite and receiving device. However, during a space weather event the ionospheric plasma density and temperature is enhanced and the degree to which GNSS must account for it is unknown. This leads to unreliable location services at best and a complete loss of the GNSS system at worst (Skone, 2001). GNSS can be used to provide timing and synchronisation for technological devices. This means that systems not usually considered vulnerable to space weather can be interrupted if they depend on GNSS timing to run. Cannon et al. (2013) recommended that all systems should be able to operate for up to three days in the absence of GNSS. Mitigating this impact requires modelling the propagation of the signal through the ionosphere (Allain and Mitchell, 2008).

Due either to solar X-rays or solar energetic particles, the bottom layer of the ionosphere can experience an enhancement of temperature and density. High frequency radio communications transmitting through the ionosphere can be disturbed with their path altered or blocked entirely (Pulkkinen, 2007). The enhanced ionosphere acts to reflect some frequencies of high frequency radio and absorbs waves at other frequencies. Similarly, satellite communications, which typically use microwave frequency signals experience scintillation as they pass through an enhanced ionosphere (Elbert, 2008; Aol, Buchert and Jurua, 2020).

1.4 Space Weather Forecasting

Space weather forecasts are used to support activities affected by space weather, particularly to give the opportunity to take mitigating action (Singer,

Heckman and Hirman, 2001; Sharpe and Murray, 2017). Examples of such mitigating action would be temporarily turning satellites off or into safe mode and rescheduling maintenance on power grids to maximise the grid capacity. For a forecast to enable mitigation decisions it needs to both be skilful and give sufficient lead time. Here skill refers to a forecast outperforming a reference forecast, commonly taken as climatology, and lead time refers to how far ahead in time the forecast is made. What counts as sufficient varies depending on the application, for example changing frequencies for radio communications can be very fast but moving humans on space flights to safety can take more time. Current operational geomagnetic forecasts computed manually give a 3-4 day lead time (Morley, 2020; Sharpe and Murray, 2017). Oughton et al. (2019) considers the benefits of a forecast enhanced through improved observations through observational spacecraft on and off the Sun-Earth line and the use of a heliospheric image. They found that the additional lead time possible (up to a week) would allow for more comprehensive mitigating action for the rarer events and increased confidence in the timing of events would allow for clearer cost/benefit analyses for forecast end-users.

In general, forecast techniques can be either physics based or empirically driven. The first relies on our understanding of the underlying physics, and is commonly used for solar wind and magnetosphere forecasts. The models numerically solve a set of equations that govern the properties of the system. These models work best when the input parameters are accurate, the physics is fully understood and there is access to sufficient computational resources. On the other hand, empirically driven models can thrive when these requirements are not met. With a sufficient historical dataset, empirical models can account for consistent observational bias in input parameters, perform without knowledge of the relevant physics and even help to discover unknown physics, and are generally computationally cheap (Owens, Riley and Horbury, 2017a). An advantage of being computationally cheap is that the model can be run as an en-

semble with perturbed input parameters to give an estimate of uncertainly of the forecast, useful for making decisions on mitigating action (Knipp, 2016; Knipp, Hapgood and Welling, 2018; Murray, 2018). While it is possible to do the same using physics based models, it is commonly computationally prohibitive to run large ensembles (Knipp, Hapgood and Welling, 2018). A counter example is a recent model from Owens et al. (2020) who provided a 1-D MHD model for the solar wind which can realistically be run as a large ensemble ($10^5 - 10^6$ ensemble members).

The solar atmosphere and solar wind are constantly monitored for the eruption and propagation of CMEs. Once observed, the evolution of these transients are modelled to give a forecast of whether they will hit Earth and what their properties would be in near Earth space. These properties can then be used as inputs to magnetospheric models that give a forecast of geomagnetic activity, and from there decisions on mitigation can be taken. It typically takes a CME between 1 and 5 days to reach Earth meaning that forecast lead time is limited to this, less any computational time and other processing and data collection and retrieval. Predicting the formation, eruption and evolution of CMEs is an active area of research (Green et al., 2018; Chen, 2017) and outside the scope of this thesis.

CIRs can be forecast using the knowledge that coronal holes, the source of CIR causing fast solar wind, typically last longer than the 27-day rotational period of the Sun (Richardson, 2018). Therefore, a 27-day lead time forecast can be made using a recurrence model which predicts that what happens now will repeat in 27-days (Owens et al., 2013). This method performs better during solar minimum as at solar maximum the corona is more dynamic and so the 27-day recurrence is less strong.

For forecasts of geomagnetic activity, magnetospheric models are used with solar wind parameters as input. Several frameworks for this exist, the majority of which numerically solve the MHD equations and iono-

spheric electrodynamics in a gridded magnetosphere-ionosphere system (for an overview of these frameworks see Welling, 2019). This is examined in more detail in Section 2.4.

Chapter 2

Background

With the focus of this thesis on space weather that leads to GICs, this chapter aims to look at the relevant physics in the magnetosphere that leads to GICs. Section 2.1 looks at the several magnetospheric currents that are known to affect the ground level geomagnetic field. Section 2.3 looks at the processes driving GICs making the link between geomagnetic field and GICs. Finally, Section 2.4 looks at the current forecasting approaches for ground geomagnetic field which would then be used for GIC forecasts.

2.1 Magnetospheric Current Systems

Space plasmas, to a high degree of approximation, are electrically neutral. This is because if significant space charge were to build up, ions and electrons are free to move along field lines and thus to cancel out the non-zero space charge. This "quasi-neutrality" means that currents are divergence-free; i.e. the total current flowing into any volume equals the total flowing out, because otherwise a space charge would build up. Such current continuity means that currents always flow in loops with no sources or sinks.

The magnetosphere contains a network of current systems driven, ultimately, by magnetospheric interaction with the solar wind. Through this interaction, energy is transferred across the magnetopause. The locations and even existence of these current systems varies depending on the state of the solar wind. This Section outlines the main current systems in the magnetosphere.

2.1.1 Ampère's Law

To understand magnetospheric currents we first need to understand Ampere's law. Ampere's law applied to a plasma is

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} \tag{2.1}$$

where **B** is the magnetic field, **J** is current density and μ_0 is the permeability of free space. Like the rest of Maxwell's equations, it is an expression of equivalence rather than causality. That is to say that the curl of the magnetic field is described also by the product of the permeability of free space and the current density but one does not 'cause' the other.

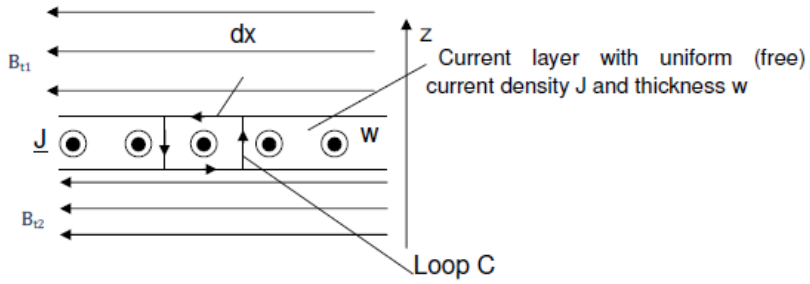


Figure 2.1: A schematic showing a current layer with uniform current density J with thickness w. The magnetic fields above and below the current layer are of strengths B_{t1} and B_{t2} respectively. This figure has been adopted from Finch (2007).

In integral form Ampère's law is

$$\oint_C \mathbf{B} d\ell = \int_A \mu_0 \mathbf{J} dA. \tag{2.2}$$

As an illustration of Ampère's law at work we look at Figure 2.1 in which there exists a thin sheet of current sandwiched between two non-equal magnetic fields. Applying the integral form of Ampère's law to a small area in the current sheet, loop C, we get

$$B_{t1}dx + B_n w - B_{t2}dx - B_n w = \mu_0 J dx w, \tag{2.3}$$

where B_n is the magnetic field normal to the current sheet, and dx and w are as shown in Figure 2.1. By cancelling terms and factoring out dx we have

$$B_{t1} - B_{t2} = \mu_0 J. \tag{2.4}$$

This tells us that the magnitude of the current density is directly proportional to the magnitude of magnetic shear. Magnetic shear appears in several places in the magnetosphere and Ampere's law shows us that currents must be present.

2.1.2 Chapman-Ferraro Current

The boundary of the magnetosphere separates the plasmas of the solar wind and the magnetspheric plasma. In the solar wind, plasma flow momentum is a dominant force whereas in the magnetosphere plasma it is the magnetic field that dominates the total pressure. This difference creates a magnetic shear and a current between the two that lies in the magnetopause. The magnetopause current was conceived by Chapman and Ferraro (1931) hence is often given the name Chapman-Ferraro current.

Figure 2.2 shows a schematic of the Chapman-Ferarro current. On the dayside magnetopause the current flows dusk to dawn at the equator

and loops around what is called the magnetic "cusp" in that hemisphere, from where Earth's magnetic field lines form a funnel-shaped region that extend to earth slightly sunward of the poles.

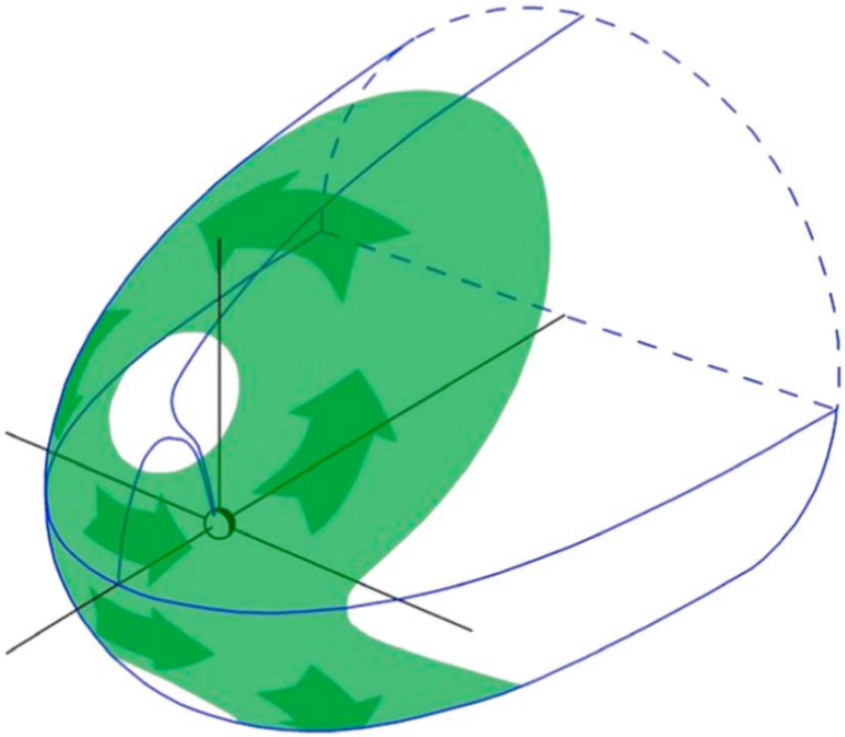


Figure 2.2: A schematic of the Chapman-Ferraro currents on the dayside of the magnetopause. The green ribbons show the path and direction of the current which encircle the cusps in the northern and southern hemispheres. This figure has been adopted from Ganushkina, Liemohn and Dubyagin (2018).

2.1.3 Ring Current

The concept of the ring current grew out of the work of Störmer (1912) on how energetic particles move in Earth's inner magnetosphere where the magnetic field is approximately dipolar. Analysis of the effects of the Lorentz force (see Section 1.2.1) on moving charged particles in a magnetic field shows that in converging field lines energetic particles following helical paths along the field lines can "mirror" (i.e., their field-aligned velocity is slowed and then reversed. See section 1.2.1). Hence a minimum in the field between two regions of larger field form a "magnetic bottle"

which can trap the particles between two such mirror points, Earth's dipole field forms precisely this geometry with a minimum at the magnetic equator and larger field at lower altitudes. This happens at all longitudes and so particles trapped in this bottle bounce from hemisphere to hemisphere and form a band that encircles the Earth and is centred on the equator. A particle gyrating about a field line experiences higher field on the part of its orbit closer to the Earth, so experience gradient drift and travelling along a curved magnetic field line will cause curvature drift (Daglis et al., 1999). A schematic of a ring current particle gyrating around a magnetic field line and being mirrored is shown in Figure 2.3. The longitudinal motions of the trapped ions and electrons in this band are different and so there is a current in this band called the "ring current". The result is that a current runs in a band around the earth roughly above the equator with an inner ring flowing eastward and an outer ring flowing westward as in Figure 2.4.

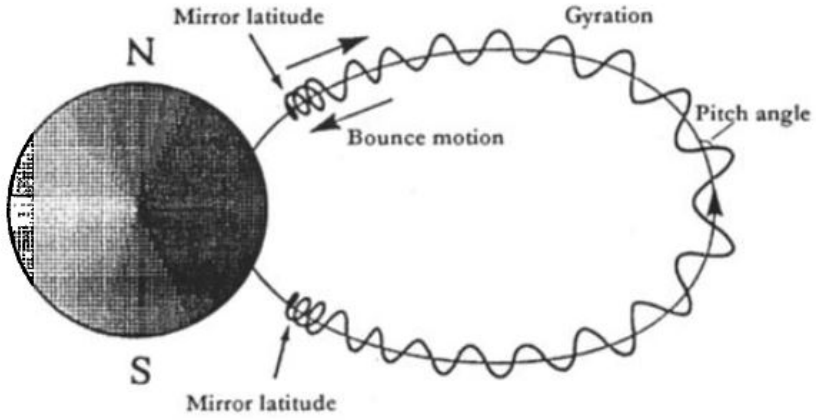


Figure 2.3: A schematic of a particle in the ring current gyrating around a magnetic field line. The particle drifts along this line and is reversed at the mirror points. As it drifts along the field line it experiences the gradient and curvature drifts. This figure has been adopted from Daglis et al. (1999).

The Active Magnetospheric Particle Tracer Explorers/Charge Composition Explorer (Krimigis et al., 1982) satellite gave the first comprehensive in-situ measurements of the ring current. It obtained plasma pressure profiles in the radial direction from Earth giving a cross-sectional picture of

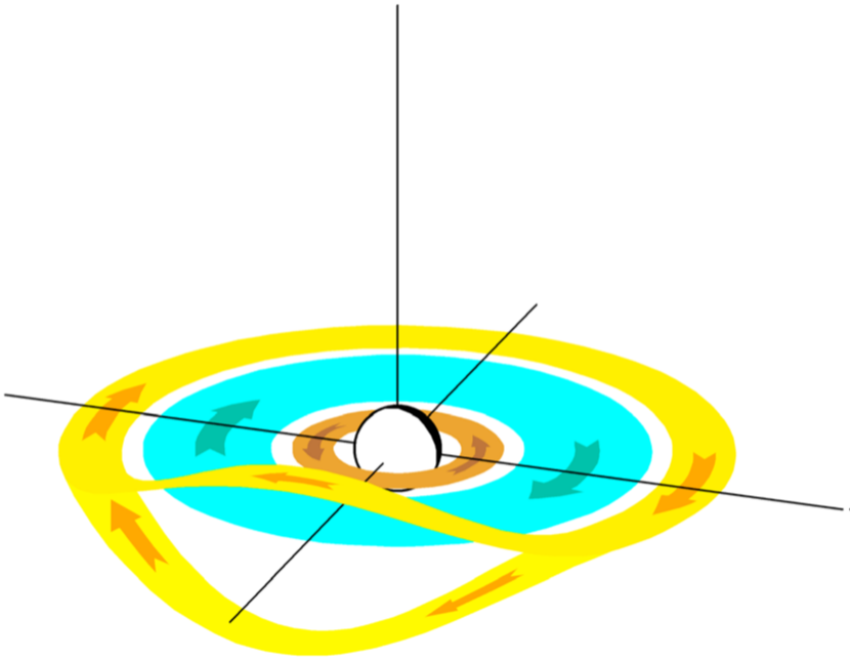


Figure 2.4: A schematic of the ring current as viewed from the day-side above the earth. The main current is the westward ring current in blue and there is also an eastward current of less importance show in brown. A representation of the split ring current is shown in yellow but this is outside the scope of this work as its existence is still an active area of research (Ganushkina et al., 2015) that is not the focus of this thesis. This figure has been adopted from Ganushkina, Liemohn and Dubyagin (2018).

the pressure in the ring current. It was seen that plasma pressure peaks at around 3 R_E, as seen in Figure 2.5 for the reference and storm cases, (e.g. Figure 2 of Lui, McEntire and Krimigis, 1987) meaning that a pressure gradient, and hence a current, exists on either side of this peak. Since the pressure gradient is of opposite sign either side of the peak so is the current. Thus, we have an eastward and westward flowing ring current. These are shown in Figure 2.4 as the brown and blue regions respectively.

The westward ring current, located at approximately 3-8 R_E, is caused by charged particles in the magnetosphere experiencing the Lorentz force

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \tag{2.5}$$

where q is the charge of the particle, $\mathbf{E}$ is an electric field, $\mathbf{B}$ is a mag-

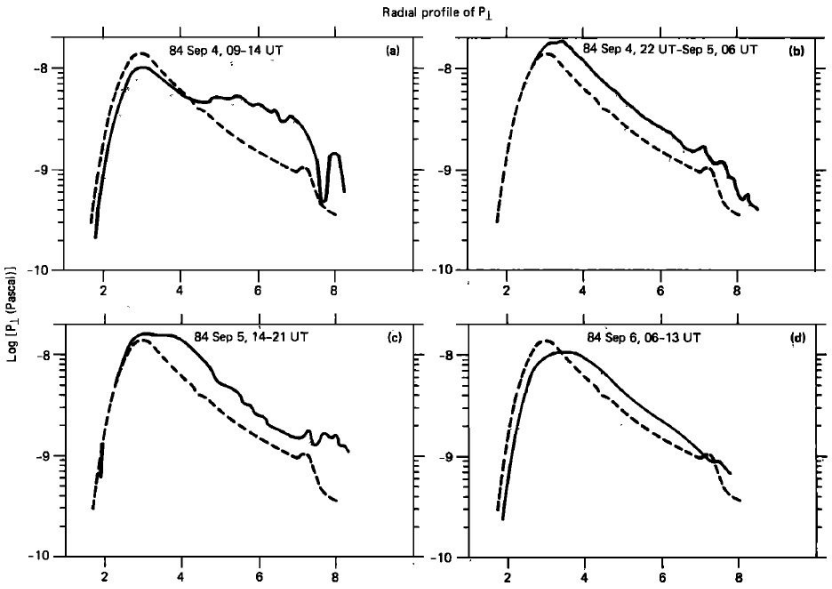


Figure 2.5: The radial profiles of the pressure perpendicular to the magnetic field for four passes of the satellite during a geomagnetic storm in 1984. The solid line shows the readings and the dashed line shows the observations made in a reference pass. The horizontal axis has unite of Earth radii. This figure has been adopted from Lui, McEntire and Krimigis (1987)

netic field and **v** is the particle velocity. From this equation we see that charged particles with a velocity not parallel to the magnetic field will experience a force acting perpendicular to the magnetic field and the particles velocity. Hence the charged particle will gyrate around a magnetic field line. If a charged particle has a velocity component parallel to the magnetic field, this will remain unchanged and the particle will move in a helix. The mirror effect that traps the particles into the ring current is caused by the pitch of that helix decreasing and then changing sign because of a field-aligned gradient in the magnetic field that means the particle moves into higher field intensity as it moves along the field.

In a non-uniform magnetic field with field gradient perpendicular to the magnetic field, as during times of geomagnetic activity, a particle will experience a changing Lorentz force in different parts of its route along the helix. This means that instead of the helix being a combination of the motion along the field line (bouncing from hemisphere to hemisphere) and a circular gyration, the orbit around the field has a smaller radius closer

to the Earth where the field is larger. This will cause the particle to drift perpendicular to both the magnetic field and the field gradient. The direction of drift depends on the sense of gyration that, in turn depends on the sign of the charge, thus oppositely charged particles drift opposite directions creating a current. This is called gradient drift

In addition, there is a drift caused by curvature of magnetic field lines. As a particle travels along a curved field line, a centripetal force is experienced and the gyrating cyclotron motion increases away from the centre. This movement through a magnetic field results in a force perpendicular to the magnetic field and direction of motion. As with gradient drift, the direction of drift depends on the charge of the particle so this "curvature drift" creates a current.

In summary, the magnetic bottle effect, due to the convergence of the magnetic field at lower latitudes, holds charged particles in an approximate torus around the earth. Gradient drift and curvature drift work together in moving positively charge particles westward and negatively charged particles eastwards giving the westward ring current.

There also exits an eastward ring current which is minor in comparison to the westward ring current, flows anticlockwise around the equator when looking down from the north and is located at approximately 2-3 R_E. The anticlockwise current flow causes a magnetic field in the same direction as the Earth's (Ganushkina, Liemohn and Dubyagin, 2018) meaning an enhancement in the eastward ring current would be measured as a positive magnetic change on the ground. The eastward ring current can undergo minor enhancements and shifts in location during geomagnetic storms (Lui, McEntire and Krimigis, 1987).

If the ring current were the same at all longitudes around the Earth it could be called "symmetric", however, it has been shown that the current densities are not the same at all magnetic local times and so is described as being, in general, asymmetric. Liemohn et al. (2001) showed

that the ring current is typically more symmetric during quiet time but is partial during geomagnetic storms. In particular Liemohn et al. (2001) saw that, during the main and early recovery phases of a storm, the ring current closed through field aligned currents into the ionosphere and this is the cause of the observed asymmetry.

During a geomagnetic storm there is significant and rapid intensification of the current. This is captured by the *Dst* index (see Section 2.2) as a decrease during the main phase of a geomagnetic storm. The decrease is because, contrary to the eastward ring current, the clockwise direction of the current causes a magnetic field with opposite polarity to the Earth's. An example of a *Dst* timeseries during a geomagnetic storm is shown in Figure 2.6.

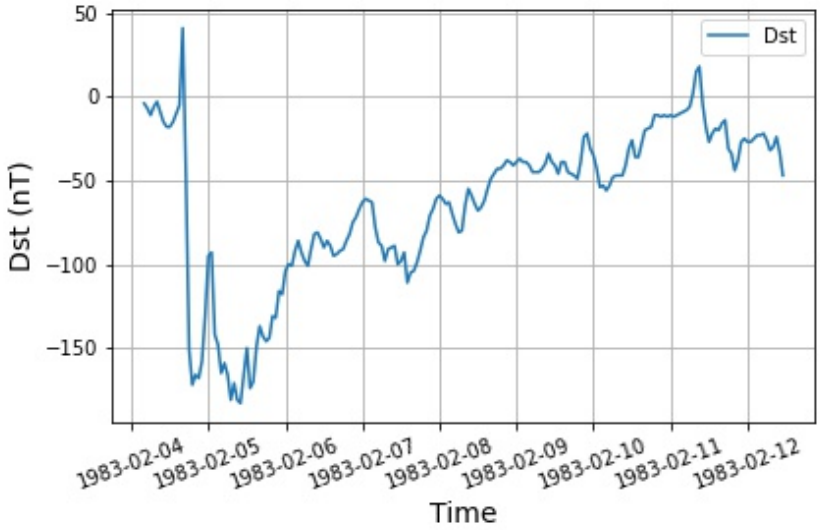


Figure 2.6: An example of a geomagnetic storm from February 1983 recorded in the *Dst* index lasting for several days.

Figure 2.7 shows the current intensity in the ring current for different levels of *Dst*, a geomagnetic index (see Section 2.2). Visible are the eastward and westward ring currents with the westward ring current generally having greater intensity and width. The inner eastward current has a much smaller effect on the magnetic field measured at the surface of the Earth, the effect being almost always with a westward current as the

intensity increase is greater (Le, Russell and Takahashi, 2004).

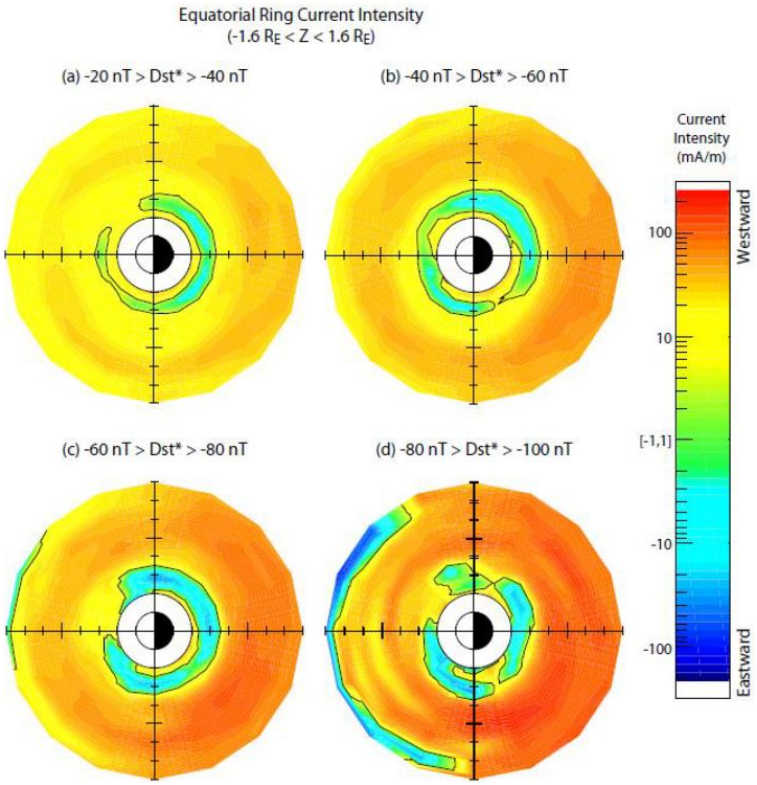


Figure 2.7: Intensity of the ring current as a function of magnetic local time and distance from the dipole axis. This figure has been adopted from Le, Russell and Takahashi (2004).

2.1.4 Field Aligned Current Systems

In the E region of the ionosphere, ions and electrons collide with neutrals interrupting their gyration around a field line. In the presence of an electric field, the ions and electrons diffuse along the electric field causing currents perpendicular to the magnetic field. The currents perpendicular to the magnetic and electric fields are called Hall currents and those perpendicular to the magnetic field but parallel to the electric field are Pedersen currents (Carlson and Egeland, 1995).

When the ionospheric conductivity is horizontally uniform, Hall currents form closed loops in the ionosphere as shown in Figure 2.8. There-

fore, Pedersen currents, which flow perpendicular to Hall currents, cannot form and so connect to field aligned currents that flow out into the magnetosphere.

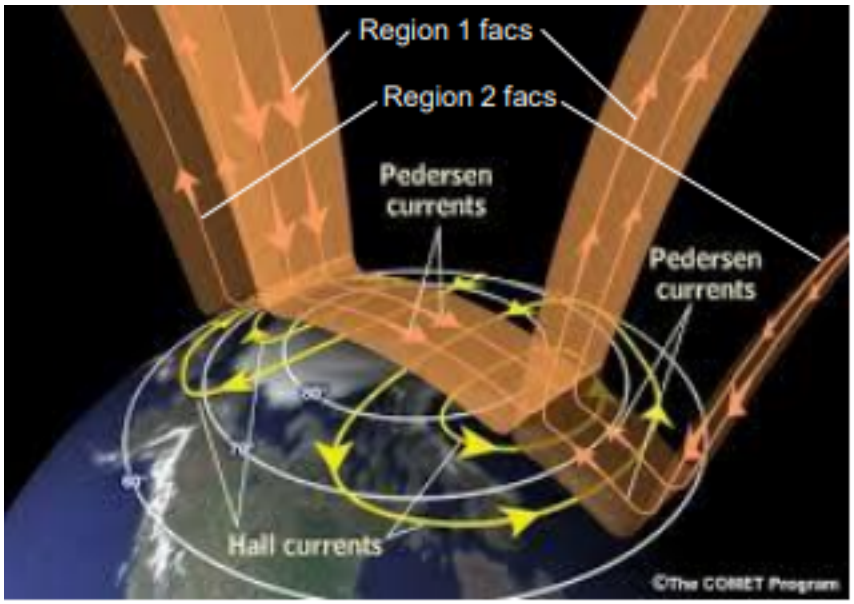


Figure 2.8: A schematic of transpolar and auroral Pedersen currents closing via field aligned currents in the presence of Hall currents. The origin of this figure is unknown.

Transpolar Pedersen currents flowing dusk to dawn across the poles connect to the so-called Region 1 field aligned currents. These flow up on the dawn side and down on the dusk side and connect to the Chapman-Ferraro current in the magnetosphere. This is shown by the left hand side red ribbon in Figure 2.9. Ganushkina, Liemohn and Dubyagin (2018) explains that this current can be diverted during geomagnetically active times to the cross-tail current sheet and close on the magnetopause, as shown by the right hand side red ribbon in Figure 2.9. This will be explained in Section 2.1.5.

Auroral Pedersen currents flow with opposite polarity to the transpolar Pedersen currents across the auroral region as shown by Figure 2.8. These then connect to region 2 field aligned currents on the equatorward edge of the auroral oval as shown in Figure 2.10. The region 2 field aligned currents then close via the ring current in the magnetosphere.

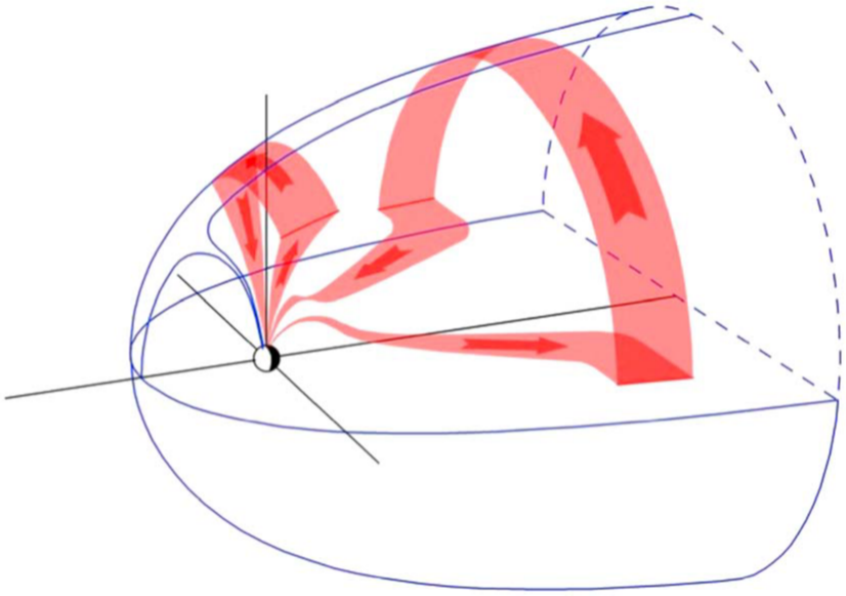


Figure 2.9: Two current flow loops in which field aligned currents connect the Chapman-Ferraro currents in the magnetopause to ionospheric currents. The left-hand loop shows the Region-1 field-aligned currents connecting to the ionosphere near the boundary between the polar cap and the auroral oval. These currents are always present but vary in strength. The right hand loop shows the substorm current wedge that forms when the cross-tail current is disrupted and field-aligned currents flow into the ionosphere and along the auroral oval in the auroral electrojet. This only happens in substorm expansion phases. This figure has been adopted from Ganushkina, Liemohn and Dubyagin (2018).

2.1.5 Cross-tail Current and Substorm Current Wedge

The magnetotail stretches far into space. The field lines connect to Earth near the poles meaning that those connecting to the northern hemisphere are directed Earthward whereas those connecting to the southern hemisphere are directed anti-Earthward creating a region of magnetic shear. As dictated by Ampere's law, separating the magnetic fields of opposite direction is an equatorial current sheet that stretches out into the magnetotail as depicted in Figure 2.11 by the flat blue surface with the current running from dusk to dawn. The current is closed via the tail-ward section of the magnetopause.

The location of the cross-tail current sheet varies during a geomagnetic storm (McPherron, 1995). During the growth phase of a substorm

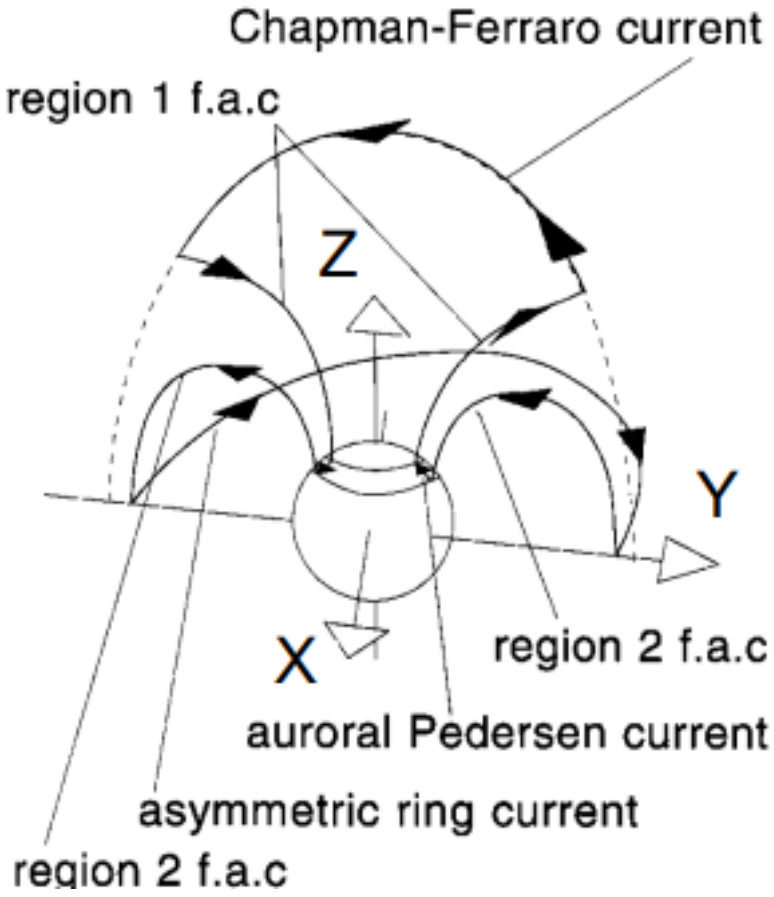


Figure 2.10: A schematic of the region 1 and 2 field aligned currents. The region 1 field aligned current connects the poleward side of the auroral oval to the Chapman-Ferraro current. The region 2 field aligned current connects the equatorward edge of the auroral oval to the night-side ring current. This figure has been adopted from Lockwood (2019b)

there is a higher rate of creation of open field lines on the dayside magnetosphere than there is destruction in the magnetotail. This adds magnetic flux to the tail and hence intensifies the cross tail current and moves it closer to the earth.

In the expansion phase of a substorm, reconnection occurs in the near-earth tail and the cross-tail current is interrupted as the substorm current wedge is formed (McPherron, Russell and Aubry, 1973). The substorm current wedge connects the cross-tail current to the highly conducting auroral ionosphere in both hemispheres. The section flowing in the auroral ionosphere is called the auroral electrojet. The substorm current

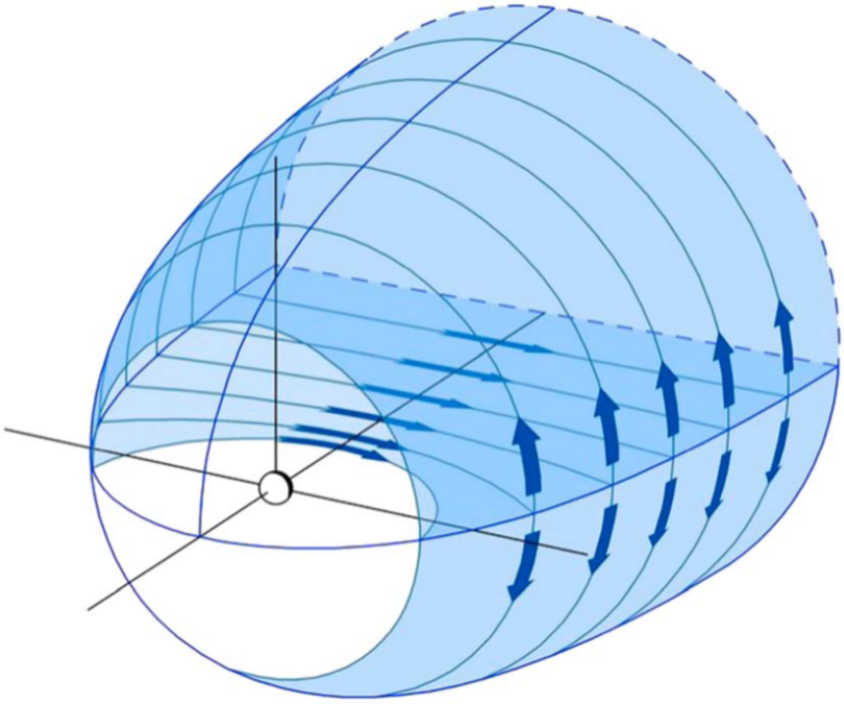


Figure 2.11: A schematic of the cross-tail current shown by the blue ribbons. The current flows horizontally across separating opposite polarity of the magnetic field. The current closes through the magnetopause current. This figure has been adopted from Ganushkina, Liemohn and Dubyagin (2018).

wedge system is depicted for the northern hemisphere in Figure 2.12 by the pink ribbon. It shows that the cross-tail current sheet is interrupted and rerouted into the auroral oval where it flows as the auroral electrojet from dusk to dawn then returns to the cross-tail current sheet.

2.2 Geomagnetic Indices

The current systems discussed in this section all have an impact on the ground level geomagnetic field. However, the magnetic effect of a current measured is a function of distance to the measuring location. For example, for a steady current, the contribution to the total magnetic field of a current element is given by the Biot-Savart law, with a $1/r^2$ drop-off. This means that close currents have the largest effect but further currents can also be felt but to a much lesser extent. Therefore, indices do not

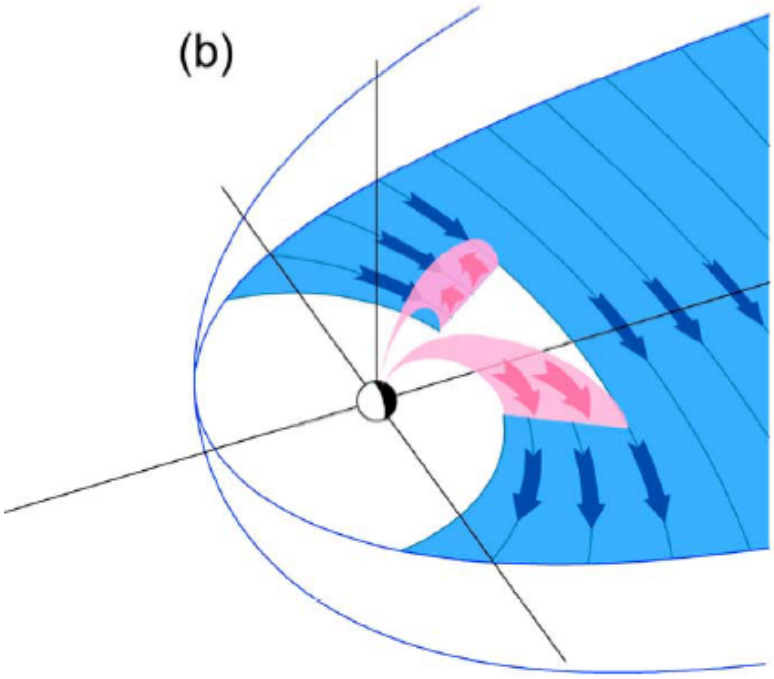


Figure 2.12: A schematic of the substorm current wedge is shown in pink. The cross-tail current is interrupted and the current is directed into the auroral oval where it flows as the auroral electrojet. This figure has been adopted from Ganushkina, Liemohn and Dubyagin (2018).

necessarily measure just one current system.

Geomagnetic indices use observations of the temporal change in magnetic field (dB/dt) measured with ground based magnetometers. These relatively cheap pieces of equipment now have a good coverage of Earth's surface. A number of indices have been developed to quantify geomagnetic activity. Each index uses a different combination of measurements and a different network of observing stations to monitor different aspects of the magnetospheric currents.

I here present an overview of four common indices.

- *Dst* (Disturbance Storm Time) is an index made from the hourly means of four equatorial magnetometers available from 1957 onwards. The low latitude of the stations means that *Dst* is primarily a mea-

sure of the ring current, an equatorward magnetospheric current system (see Section 2.1). The major disturbances of Dst are negative as ring current intensification causes a decrease in the equatorial geomagnetic field. This intensification is caused by the injection of energetic particles into the inner magnetosphere by substorm expansions and enhanced magnetospheric convection. They are seen in the main phase of geomagnetic storms. However the equatorial stations can also sense the Chapman-Ferraro currents flowing in the dayside magnetopause and this can give a small positive perturbation ahead of the main phase of the storm as enhanced solar wind, often in the form of a coronal mass ejection, compresses the magnetosphere and brings the currents in the magnetopause closer to Earth. For this reason, it is common for a geomagnetic storm to cause a slight rise in *Dst* before a sharp negative fall.

- The *AE* (auroral electrojet) index aims to measure the flow of the auroral electrojet currents in the high-latitude ionosphere. For this reason the measurements are taken from 12 high-latitude stations spaced out longitudinally such that both the eastward and westward auroral electrojets are measured at any given time. The stations are in the northern hemisphere as in the southern hemisphere much of the auroral oval, where the stations need to be located, is over sea or ocean. *AU* is defined as the maximum recorded positive disturbance at any station and *AL* is the minimum recorded disturbance. *AE* is a single measure of the electrojets defined as the difference between *AU* and *AL*. Data is available from 1957 onwards at a 2.5 minute resolution.

- *am* is a 3-hour range index available from 1959. It is compiled using data from approximately 24 stations across the mid-latitudes although the exact stations used has varied through time. The measured K values (Bartels, Heck and Johnston, 1939), a quasi-logarithmic

scale from 0-9 based on station specific look-up tables, are averaged for each longitudinal sector and then the northern and southern hemispheres to give a global measure of geomagnetic activity. Although compiled from mid-latitude stations the range indices are mainly sensitive to enhancements to the midnight auroral electrojet which moves equatorward as well as growing in intensity during substorm expansion phases and major storms.

- The *aa* index is a long running index available from 1868 to present. It is compiled in a similar way to am but with only two approximately antipodal stations (UK and Australia). There have been three stations in each country over the time period of the index. Lockwood et al. (2018a,b) corrected *aa* for the changing geographic location of stations and the drift of the geomagnetic pole with time, to produce aa_H. Using a model of station response and sensitivity, they also reduced the time-of-day and day-of-year effects which result from creating a global index from stations with limited longitudinal sampling.

2.3 GICs

Intensification of magnetospheric current systems resulting from geomagnetic storms induce electric fields on the surface of the Earth which in turn induce currents in systems containing long conductors such as power grids and pipelines. These currents are named geomagnetically induced currents (GICs) (Ngwira and Pulkkinen, 2019). GICs flowing in technological systems cause issues which, in mild cases, leads to increased maintenance and, in extreme cases, to system failure (Rajput et al., 2021).

The variability in magnetospheric currents during a storm causes a variation in the geomagnetic field at the surface of the Earth, with the strength of the perturbation following an inverse distance squared law be-

tween the current and the location where the geomagnetic field is observed. This leads to currents being induced in all conductors experiencing a changing magnetic field.

The task of estimating GICs can be separated into two components (Pirjola, 2000; Ngwira and Pulkkinen, 2019). The first is the estimation of the ground-level geoelectric field resulting from space weather. Faraday's Law of Induction links the activity in the magnetosphere and ionospheric current systems with GICs at ground level. It is

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}. \tag{2.6}$$

This states that a time-varying magnetic field accompanies an electric field through electromagnetic induction. Because magnetic field fluctuations are more intense closer to current systems, it is expected, and observed, that GICs of relevant magnitude are more likely to occur at latitudes in closer proximity to them. Currents such as the auroral ovals and the substorm current wedge occur at high latitudes and so we would expect GICs to be a greater issue at high latitudes. However, it should be noted that during strong geomagnetic storms the ring current intensifies over low-latitudes meaning GICs could occur there (Abda et al., 2020; Lui, McEntire and Krimigis, 1987; Mohd Anuar et al., 2019). Mathematically, this is shown by the Biot-Savart law:

$$\mathbf{B} = \frac{\mu_0 \mathbf{I}}{4\pi} \int \frac{ds \times \hat{r}}{r^2}, \tag{2.7}$$

where r is the radial distance between current and magnetic field location, $\mathbf{I}$ is the current, and s is the path of the current. For such a magnetic field observed at the Earth's surface, temporal variations will be accompanied by an electric field. This can lead to a current flowing on the

ground following Ohm's law,

$$\mathbf{J} = \sigma \mathbf{E}, \tag{2.8}$$

where $\mathbf{J}$ is the current density related to $\mathbf{I}$ by the expression $I = AJ$, where A is the cross-sectional area of the current, and σ is the conductivity of the ground. Therefore, ground conductivity plays a key role in the degree to which the geomagnetic field penetrates the earth and induces an electric field. A more conductive ground allows the field to penetrate deeper. The penetration depth also depends on the frequency of the geomagnetic field variations with lower frequency variations penetrating deeper. Because of the dependence on ground conductivity, which varies with depth due to layers of different materials, full modelling of GICs requires knowledge of the ground structure. Kelbert (2020) provides an overview of the current capabilities in GIC and electric field modelling.

Magnetotellurics is a method that can determine the distribution of the conductivity of the ground by simultaneously monitoring the ground level magnetic and electric fields then processing the data. In the UK, Beggan et al. (2021) used six months worth of uninterrupted magnetic and electric field observations to derive the impedence of the Earth at three sites (Lerwick, Eskdalemuir and Hartland). From this a magnetotelluric-(MT-) transfer function was created to estimate the electric field from a magnetic field at these three locations.

Other work has focused on creating 3-D models which give a more comprehensive measure of the ground. This can either be done by using MT-transfer functions at particular location then interpolating the electric field or it can be done by interpolating the magnetic field on to a grid and obtain an MT-transfer function at each gridpoint through MT-inversion (Kelbert, 2020; Avdeev, 2005) However, these models are of limited availability and thus have limited applications at present. As examples, a list of regional 3-D models for the US can be found in Table 8.2 of Kelbert,

Bedrosian and Murphy (2019).

The next part in the GIC problem is the engineering step. This takes the geoelectric field and estimates the level of GIC in a power grid. This is completely dependent on the set up parameters of the grid and so is usually carried out by engineers. GICs can be parameterised in terms of

$$GIC(t) = aE_x(t) + bE_y(t) \tag{2.9}$$

where a and b are grid set up parameters and E_x and E_y are the two components of the horizontal electric field. a and b are unique to each transformer and power line and depend on the resistance and geometry of the system (Pirjola, 2000). Without specific knowledge of a and b, it is the task of scientists to provide accurate geoelectric field estimates and forecasts.

In power grids, GICs flow between the ground and the system through earthing cables in transformer substations. This occurs particularly in areas with low ground conductivity as the power grid offers a path with significantly lower resistance. These currents have a very low-frequency of 0.01–0.001 Hz (quasi DC) with average magnitudes of 10–15 A and peaks of up to 100 A for 1–2 min (Heindl et al., 2011). These low frequency currents behave similarly to a direct current (DC) in a 50 Hz alternating current (AC) system. In a transformer, this quasi-DC signal can lead to half-cycle saturation, where the maximum magnetic flux of the transformer core is achieved every half-cycle. During saturation, excess flux leaks out of the core and induces eddies in other components, causing them to heat (Marti, Rezaei-Zare and Narang, 2013). Sustained heating can lead to degradation and, in extreme cases, destruction of the transformer. In addition to transformer heating, GIC effects can cause transformers to operate in a non-linear regime, where current is no longer proportional to voltage, resulting in relay trips, voltage instability and drops in power causing blackouts (Pirjola, 2000).

Transformer insulation damage is a cumulative process and can be accrued over the course of many GICs. When systems fail, the cause may be in conjunction with, and often attributed to other causes (Molinski, 2002). Thus the cumulative processes may not have been given the right level of importance. Studies have found that the Generator Step-Up transformers in the north-eastern United States (the region of the US with the highest GIC susceptibility) have a 60% higher failure rate than the rest of the country, and a faster mean time-to-failure rate, indicating early ageing of the transformers (Kappenman, 1996). The failures appear to occur in phase with the timing of geomagnetic storms, while the frequency of failures in other regions can be represented with a normal distribution in time (Kappenman, 1996). Forbes and St. Cyr (2004) investigated the electrical power market in the year 2000, concluding that geomagnetic storms increased the wholesale electricity price by a total of approximately $500 million.

Another factor in transformer damage is geomagnetic storm duration which plays a significant role in the potential for GIC damage. The longer the geomagnetic field fluctuates at a great intensity, the longer the consequent GIC is sustained in a power system, and given that said GIC is of sufficient magnitude, the more potential there is for system instability and/or damage. The risk of exposure to GICs is both a function of time and magnitude (Kappenman, Ave and Suite, 2010).

Extreme space weather events are more likely to give rise to large GICs. The 2003 Halloween event caused GICs that were likely the cause of the failure of the Swedish high-voltage transmission system (Pulkkinen et al., 2005) and damage to a South African transformer (Gaunt and Coetzee, 2007). An even more extreme event in 1989 lead to the collapse of the Hydro-Quebec network in Canada (Boteler, 2001).

Forbes and St.Cyr (2017) investigated whether the UK power grid is effected by GICs. They looked at the reliability of the electricity grid

during the declining phase of solar cycle 23 using the net imbalance volume (NIV), a metric of electric grid stability. Their results suggested that there was an impact on NIV due to GICs on National Grid operations.

Pulkkinen et al. (2017) reviewed the current understanding of GICs. They concluded that the most likely result of a large GIC event is voltage instability leading to blackouts rather than permanent transformer damage on a large scale. That said, industry is not only interested in the most probable events, but the combination of the probability and the severity. For example, the power industry may be more concerned about permanent transformer damage that takes months to replace than a few hours blackout, even if its is much less probable.

Milder GIC events are also of interest because they occur more frequently. Schrijver (2015) and Schrijver et al. (2014a) looked at the impact of moderate space weather events using insurance claims data. In particular, Schrijver (2015) shows that high frequency, low impact events may be comparable in cumulative economic cost to low frequency, high impact events. Similarly, Schrijver et al. (2014a) identifies significant rises in electrical equipment insurance claims on both the top 5% and top third of geomagnetically active days. While these studies focused on geomagnetic activity as a whole, it is reasonable to assume that this holds for GICs specifically.

The conclusion of Pulkkinen et al. (2017) was that the key future challenge is to be able to specify the spatiotemporal evolution of the geoelectric field during extreme events. They also pointed to the need for GIC forecasting with a multi-day lead time and for estimates of upper limits of extreme GICs.

2.4 GIC Forecast Capabilities

As discussed in the previous section, the science problem for forecasting GICs is to forecast the ground level magnetic perturbations, ΔB, and convert this into geoelectric field. The forecasting of ΔB can be achieved in two ways. The first is using physics-based first-principles models, typically MHD, to model the response of the magnetosphere to solar wind inputs (Welling, 2019). The second is to use empirical models that link the solar wind conditions directly to ΔB (Weimer, 2019). In the discussion in this section the solar wind inputs are those observed at the L1 point meaning that these forecasts have short lead times and are effectively nowcasts.

The MHD-model approach numerically solves the MHD equations shown in Chapter 1 on temporal and spatial grid boxes roughly 0.1 to 1.2 Earth radii (El-Alaoui et al., 2012). Welling (2019) reviewed the validity of MHD models of the magnetosphere. They found that current MHD models are good at reproducing the shape of the magnetosphere, at reproducing large-scale ionospheric features and at capturing magnetospheric convection. However, they note that MHD models generally struggle with properly capturing magnetotail dynamics, key to substorms. They also found that a higher spatial grid resolution provides better results, as seen in Pulkkinen et al. (2010) and Wiltberger et al. (2017). This must be weighed up with the increased computational costs as halving the average spacial scale creates a factor of eight increase in number of grid cells and also an increase in temporal resolution due to the Courant–Friedrichs–Lewy (CFL) condition (Courant, Friedrichs and Lewy, 1956).

Global MHD models typically have an inner boundary above the ionosphere because **B** increases closer to the surface and hence the wave speed increases requiring a prohibitaively high time resolution due to the CFL condition. Therefore Biot-Savart integration is used to retrieve ground level magnetic field from the model at particular locations. These so called "virtual magnetometers" allow global MHD models to be directly validated

against ground based observations and, in principle, to be used as forecasts for GICs.

Pulkkinen et al. (2010) evaluated three configurations of the space weather modelling framework (SWMF Tóth et al., 2005, 2012) and one configuration of Open Global Generalized Circulation Model (OpenGGCM) (Raeder, Berchem and Ashour-Abdalla, 1998) on their ability to produce ΔB using virtual magnetometers during four storm events. They made evaluations using prediction efficiency with all models coming out with a negative score in each event. They also examined the power spectra of ΔB and found that the models lacked spectral power for some events, particularly for frequencies between 10^{-3} to 10^{-2}.

Empirical models have the advantage that they are computationally efficient compared with MHD models and are thus cheaper and easier to run in real time. This opens up the possibility for models to be run as an ensemble to quantify forecast confidence based on observation uncertainty. Another advantage is that, in settings where there is unknown or unresolved physics, empirical models can (potentially) still perform well.

An example of an empirical model for predicting ΔB is the Weimer model described in Weimer (2013). The model uses spherical cap harmonic analysis (Haines, 1985) which determines the geomagnetic field at a given location by fitting a spherical cap to the derivatives of a potential function. The spherical cap is extended from the poles to the equator, thus covering an entire hemisphere. In the equations for each of the vector components of ground level magnetic perturbations are two parameters that are empirically derived as functions of the solar wind parameters. These parameters are observed by the DSCOVER spacecraft at the L1 Lagrange point, meaning that the Weimer model can give an approximate 1-hour lead time. The Weimer model was evaluated for its ability to forecast an event from July 2004. At all 11 locations tested it well produced ΔB, however, it lacked the high-frequency signal (Weimer, 2013). This can be

seen in Figure 2.13 which shows the Weimer model output in blue and the observations in red at the 11 locations.

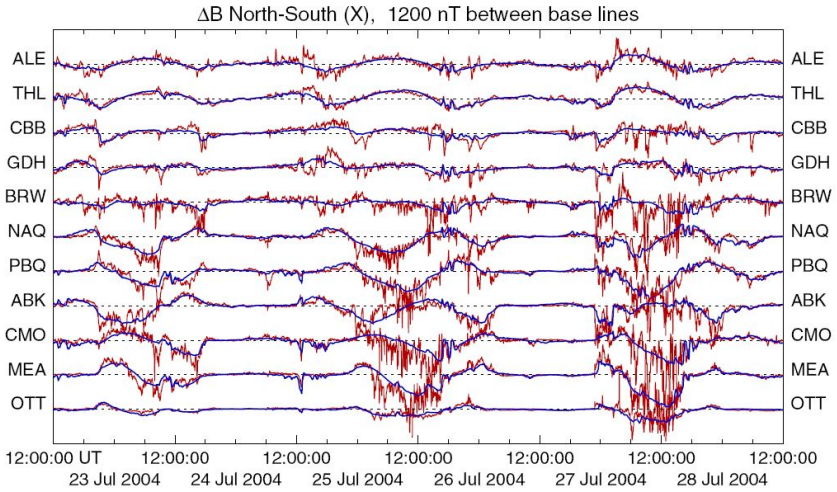


Figure 2.13: The results of the Weimer model for the north-south magnetic field component are shown for a July 2004 event at 11 different locations. The blue line shows the model output and the red shows the observations. The model does well at capturing the low resolution variations but struggles with the high.

2.5 Thesis Objectives and Outline

The main aim of this thesis is to contribute towards space weather forecasting, particularly in a way that is operationally relevant to the power industry. Ideally, infrastructure operators would be provided with a long lead-time GIC forecast, however GIC forecasting is still in its infancy. This thesis takes a two pronged approach towards improving our forecast capability for the power industry.

The first prong is a focus on geomagnetic activity as measured by geomagnetic indices (see Section 2.2). While forecasting indices is not equivalent to forecasting GICs, geomagnetic indices are nevertheless a useful proxy for magnetospheric activity and have long historical records allowing for a statistical understanding and empirical forecast methods. In Chapter 3 we examine the duration of geomagnetic storms in a geomag-

netic index and establish a statistical relationship between duration and peak intensity of the storm. A model based on log-normal distributions is presented giving a probabilistic estimate of the duration of a storm for a given peak storm intensity. It is hoped this is useful to infrastructure operators, as they are better able to plan if they have an estimate of the duration of the disruption. It also gives an insight into the level of transformer damage that could be done as this is a time integrated effect.

In Chapter 4 we use two pattern-matching forecast techniques, namely the analogue ensemble (AnEn) and support vector machine (SVM). Both methods are implemented to take the recent time history of the index as input and give a prediction for up to a one week lead time.

The second prong focuses more specifically on GICs. Current GIC forecast is limited by the resolution of operational magnetospheric models. Typically, the magnetospheric models underestimate the power of the high frequency variability in the geomagnetic field causing an underestimation of the magnitude of corresponding geoelectric field which is needed to forecast GICs. To address this problem, Chapter 5 presents a temporal downscaling model of the geomagnetic field as a proof of concept study. Employing the analogue ensemble technique, a time series of geomagnetic field is downscaled from a 1-hour resolution to a 1-minute resolution.

Chapter 3

The Variation Of Geomagnetic Storm Duration With Intensity

This chapter presents the work of my first project during my PhD. The study was conceptualised by my supervisors, however I had significant input into the direction of the study as it moved from establishing a statistical relationship to building and testing a model. I conducted all of the data handling and analysis required and wrote the resulting paper. This Chapter is published in *Solar Physics* as Haines et al. (2019).

C. Haines, M.J. Owens, L. Barnard, M. Lockwood, A. Ruffenach

Abstract

Variability in the near-Earth solar wind conditions can adversely affect a number of ground- and space-based technologies. Such space weather impacts on ground infrastructure are expected to increase primarily with geomagnetic storm intensity, but also storm duration, through time-integrated effects. Forecasting storm duration is also necessary for scheduling the resumption of safe operating of affected infrastructure. It is therefore impor-

tant to understand the degree to which storm intensity and duration are correlated. The long-running, global geomagnetic disturbance index, *aa*, has recently been recalibrated to account for the geographic distribution of the component stations. We use this aa_H index to analyse the relationship between geomagnetic storm intensity and storm duration over the past 150 years, further adding to our understanding of the climatology of geomagnetic activity. Defining storms using a peak-above-threshold approach, we find that more intense storms have longer durations, as expected, though the relationship is nonlinear. The distribution of durations for a given intensity is found to be approximately log-normal. On this basis, we provide a method to probabilistically predict storm duration given peak intensity, and test this against the aa_H data set. By considering the average profile of storms with a superposed epoch analysis, we show that activity becomes less recurrent on the 27-day timescale with increasing intensity. This change in the dominant physical driver, and hence average profile, of geomagnetic activity with increasing threshold is likely the reason for the non-linear behaviour of storm duration.

3.1 Introduction

A geomagnetic storm is a significant disturbance in the Earth's magnetic field (e.g. Gonzalez et al., 1994) due to specific sets of conditions in the near-Earth solar wind. Southward-orientated interplanetary magnetic field (IMF) can reconnect with the geomagnetic field of the dayside magnetosphere resulting in the storage of energy in the lobes of the magnetotail (Dungey, 1961). Reconnection associated with disturbances in the magnetotail current sheet releases the stored energy and currents in the upper atmosphere are enhanced. This can result in adverse effects in a number of ground- and space-based technologies, as well as posing a threat to the health of astronauts and flight crew and passengers (Baker, 1998). In particular, this substorm cycle of tail lobe energy storage and release can result in enhanced ionospheric currents which can in turn lead to Geomagnetically Induced Currents (GICs), a quasi-DC signal, flowing through ground infrastructure such as power grids and pipelines (Patel et al., 2016; Cannon et al., 2013). An induced DC current in AC power transformers may create a half cycle saturation leading to degradation and breakdown. Thus geomagnetic storms can affect today's technologically centred society to great disruption and at great cost (Eastwood et al., 2017a; Oughton et al., 2017a; Riley et al., 2018; Cannon et al., 2013).

GICs result from rapidly varying local geomagnetic fields, on the timescale of minutes or less. Therefore geomagnetic indices (e.g. Dst, AE and Kp), which summarise global geomagnetic variability on the timescale of hours, do not directly relate to the GIC drivers. However, there is coupling across these timescales and spacial scales, with the largest GICs occurring during geomagnetic storms (Trichtchenko and Boteler, 2004, 2007).

Space-weather forecasting often focuses on estimating the onset and intensity of geomagnetic storms (Abunina et al., 2013; Lundstedt, Gleisner and Wintoft, 2003; Joselyn, 1995). But storm duration can also be an important secondary factor through time-integrated effects (Mourenas, Arte-

myev and Zhang, 2018; Balan et al., 2016). Indeed, Lockwood et al. (2016) showed that the time-integrated value of solar wind forcing, which measures the net geomagnetic disturbance, was influenced both by the amplitude and duration of storms. Therefore, a forecast of storm duration would be beneficial in the development of mitigation plans for sensitive infrastructure and services, in terms of estimating when normal operations can resume. Thus it is important to quantify the relationship between storm intensity and duration.

Surveys of the largest geomagnetic storms in the *Dst* index have included estimates of storm duration (Echer, Gonzalez and Tsurutani, 2008; Balan et al., 2016), though the relation with storm intensity was not explicitly quantified. Using the *Dst* index, Yokoyama and Kamide (1997) showed that the average duration of main and recovery phases of storms increased with storm intensity across three broad categories (weak, moderate and strong) for 8 years of observations. A similar analysis was presented in Hutchinson, Wright and Milan (2011) for solar cycle 23 but concluded that storm main phase duration increases with intensity to a point, but then the relationship reverses for storms more intense than a peak $SYM-H$ index disturbance of $-150nT$. Recently, Walach and Grocott (2019) analysed the $SYM-H$ index between 2010 and 2016 in a similar way to Hutchinson, Wright and Milan (2011) but concluded that there was no clear ordering of intensity by storm duration. Conversely, Lefevre et al. (2016) used the *aa* index to demonstrate an increase in average storm duration across five peak intensity levels. The weakest storms had a median duration of 6 hours and the most intense storms 30 hours, noting that the duration of the most intense storms ranged from 12 to 93 hours. Xie et al. (2008) showed that, for large storms, preconditioning of the magnetospheric system by prior solar wind conditions can increase storm duration but found no evidence that it increases peak amplitude.

The long-running, global geomagnetic disturbance index, *aa* (Mayaud,

1971), has recently been recalibrated to account for the geographic distribution of the component stations (Lockwood et al., 2018a). In this study, we use this aa_H index to analyse the relationship between geomagnetic storm intensity and storm duration over the past 150 years, further adding to our understanding of the climatology of geomagnetic activity. In particular, we construct and test a simple probabilistic forecast of storm duration based on storm intensity.

3.2 Data

Changes in magnetospheric current systems can result in magnetic fluctuations which can be measured at a number of stations around the world with ground-based magnetometers. These measurements are compiled into a variety of indices, including *aa*, *kp*, *Dst* and *AE*, which measure different properties of global geomagnetic activity. The *aa* index, created by Mayaud (1971), is based on the k values devised by Bartels, Heck and Johnston (1939). The k values are made by ranking the range of variation in the observed horizontal or vertical field component (whichever gives the larger value) in each 3-hour period into one of 10 categories. These categories are defined by quasi-logarithmic limits based on the station's proximity to the auroral oval and a k value of 0 to 9 is assigned. The *aa* index is a combination of measurements taken from two mid-latitude stations, in the UK and Australia, since 1868 which, by virtue of being in opposite hemispheres and roughly 10 hours apart in local time, provide a quasi-global measure of geomagnetic activity with particular sensitivity to the substorm current wedge (Lockwood, 2013).

The major disadvantage of *aa* is that it is compiled from just two stations, but that is necessary in order to generate the series back to 1868, which is its major advantage. Lockwood et al. (2018a) and Lockwood et al. (2018b) corrected *aa* for a number of factors. Firstly they

made allowance for the changing geographic location of the midnight sector auroral oval, due to the secular change in Earth's intrinsic geomagnetic field (giving different drifts of the geomagnetic poles in the two hemispheres, as observed), which influences the proximity of the stations to the most relevant current system, the nightside auroral electrojet of the substorm current wedge. This improves the intercalibration of the different stations needed to compile the *aa* series (three have been needed in each hemisphere to make a continuous series since 1868). In addition, they allowed for the time-of-day/time-of-year response pattern of the station (the "station sensitivity") using a model of how it is influenced by proximity to the auroral oval and by local ionospheric conductivity, thereby reducing the errors introduced by using just two stations. The resulting index, called the "homogeneous aa index", aa_H, has been tested by Lockwood et al. (2019) against the *am* index, generated in a similar way to *aa* but using rings of 24 mid-latitude stations in both hemispheres (Mayaud (1981), `http://isgi.unistra.fr/indices_am.php`), and was shown to perform considerably more uniformly in local time than the original *aa* index. The analysis shown in the present paper uses the aa_H data but similar results have been achieved for the *am* index. The aa_H index runs from 1868 to present and definitive *am* data currently runs from 1959 to 2013. Although *am* gives a more accurate representation of the instantaneous state of global geomagnetic activity (Lockwood et al., 2019), the advantage of aa_H is that it provides an extra century of observations which means that better statistics on large events are obtained.

3.3 Storm Definition

There are no universally-agreed upon criteria to classify geomagnetic storms in geomagnetic time series, with definitions depending on the purpose of the study (Riley et al., 2018). Lefevre et al. (2016) defined storms in the 3-hourly *aa* time series as events in which the maximum value exceeds

one of a set of limits which we refer to here as the "upper threshold". They defined the start and end of a storm as the times when *aa* rose above and fell below a threshold of twice the mean of the whole data-set ($40nT$): we here refer to this as the "lower threshold". Thus in their definition of an event, the start of the storm is where *aa* first exceeds $40nT$ prior to the peak, and the end is where it last exceeds $40nT$ after the peak. Thus multiple upper threshold crossings can potentially constitute a single storm (see also Figure 3.1). In the Lefevre et al. (2016) definition, *aa* continuously exceeds the lower threshold in an event and so even a brief drop below it means that a new event is counted as having started. Only storms associated with a sudden storm commencement (SSC) were considered - definitive SSCs are defined by visual searches for sudden increases in the northward component of the field measured at any one of five low-latitude stations. This yielded a set of 2370 storms for the lowest upper threshold they considered ($50nT$). Kilpua et al. (2015) used 3-hourly *aa* data, but with an upper threshold of $100nT$ and a lower threshold of $50nT$. This gave 2073 storms and performed analysis on cumulatively binned categories of storms larger than a certain value, beginning at $100nT$ and increasing in increments of $100nT$ up to $600nT$. This upper and lower threshold method of storm definition was also employed by Riley and Love (2017) for the Dst index.

Other approaches include that of Hutchinson, Wright and Milan (2011) who identified storms in $SYM-H$ using a threshold approach of $-80nT$ and then proceeding to manually inspect individual storms using knowledge of the characteristic storm trace and the solar wind data. This was feasible because the study included only 143 events from a single solar cycle, a relatively small sample compared to that enabled by the aa_H data set.

We follow a similar approach to Kilpua et al. (2015) and Lefevre et al. (2016) for storm definition. However, we will follow a more data-informed approach to threshold selection by following the percentile ap-

proach (Gonzalez et al., 1994). When categorising *Dst* storms (in which storms are negative perturbations), Gonzalez et al. (1994) defined the lowest 25% of *Dst* values as a weak storm, the lowest 8% as a moderate storm, and the lowest 1% as a strong storm. Of course, the issue with this approach is that the percentile-based thresholds will change with the interval considered. This variation will be greatest with short data sequences and very high thresholds (e.g., the top 1% or higher).

Since a strong aa_H storm has a positive value, we define a storm to start when the aa_H value is greater than the $90th$ percentile of the full 1868-2017 3-hour data sequence. The end of the storm will be the last value greater than the $90th$ percentile. The $90th$ percentile of aa_H (1868-2017) is $40.1nT$, similar to that used in previous studies. For some parts of the analysis, we have also used an upper threshold to select storms of different peak intensity. As in previous studies, an event must have a peak value over the upper threshold to be selected, but the storm duration is nevertheless measured as the time spent over the lower threshold (i.e., the upper threshold determines whether an event is classed as a storm and the lower threshold determines its duration). Upper and lower thresholds are depicted in Figure 3.1.

A feature of this approach is that a single measurement that falls just below the lower threshold brings an end to a storm. This is depicted in Figure 3.1 in which the tail at the end of the storm is cut off. Another feature is that two events occurring close to each other with respect to time will be classed as a single event if the index does not fall back below the lower threshold in the interval between them. This double peak effect is depicted in the plot on the right of Figure 3.1. While this approach may sometimes disagree with the interpretation of a human observer, it gives a set of objectively defined storms, making our analysis reproducible and readily applicable to the entire dataset.

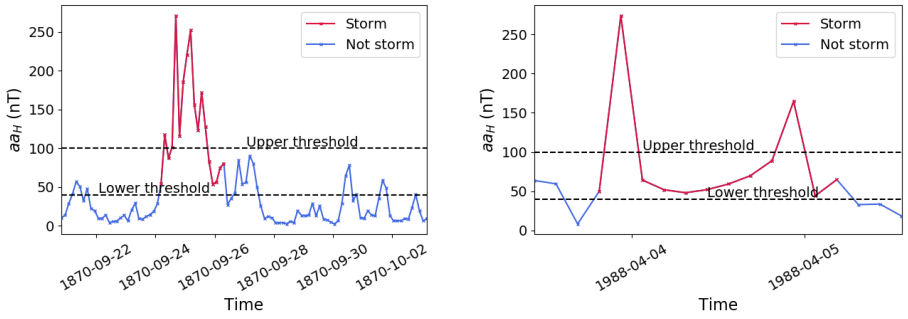


Figure 3.1: Examples of geomagnetic storm definitions using 3-hourly aa_H data. The upper threshold determines if the event is included in the analysis. A lower threshold determines the duration of the event. The period defined as a storm, when using an upper and lower threshold, is shown in red. Left: A slight dip below the lower threshold means what an individual observer may regard as the tail of the storm has been cut off. Right: A double peak shape suggests two storms may have been counted together as a single event.

3.4 Results

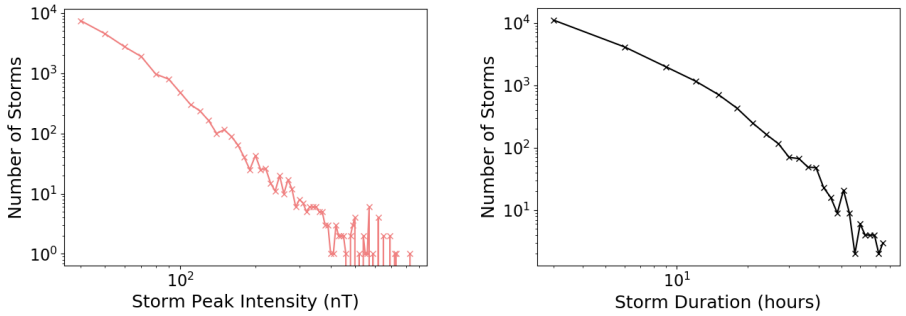


Figure 3.2: Left: The number of storms as a function of peak intensity on a log-log scale. The data was grouped into equal sized bins of 10nT. Right: The number of storms as a function of storm duration on a log-log scale plotted using bin width of three hours, the resolution of aa_H, meaning every possible duration has a unique bin.

The number of storms in the aa_H index can be seen in Figure 3.2 (Left). The data has been grouped into bins of width $10nT$. The log-log scale reveals that the distribution of storms follows an approximate power law. This result is in agreement with that from Riley (2012) who found a power law in the occurrence of geomagnetic storms in the *Dst* index. The sharp drop off associated with a power law can be seen in the probabilities of storm peak intensity revealed by the Complementary

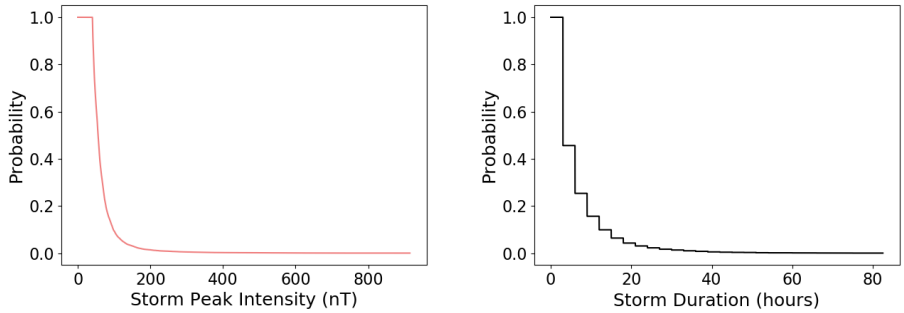


Figure 3.3: Complementary Cumulative Distribution Functions (CCDFs) of peak intensity (left) and durations (right) for all storms in the aa_H index.

Cumulative Distribution Function (CCDF) in Figure 3.3 (Left).

Similarly, the number of storms as a function of duration has a sharp drop off. Figure 3.2 (Right) reveals that the decrease is slightly less than a power law. The accompanying CCDF in Figure 3.3 (Right) shows the drop off in a stepped fashion due to the quantisation of the aa_H index into 3-hour data-points.

The variation of mean storm duration with storm intensity, as measured by the upper threshold, is presented in Figure 3.4. As expected, the general trend is that as the upper threshold is increased, the mean storm duration also increases. Of course, if storms have a common recovery timescale and thus similar saw-tooth-like profile (see Section 3.5), storm duration would be expected to increase monotonically with peak intensity. This result is in qualitative agreement with Hutchinson, Wright and Milan (2011) and Lefevre et al. (2016), who noted an increase in storm duration for increasing storm intensities.

We find the relation between intensity and duration to be non-linear, with a plateau in mean duration towards higher thresholds, at a level which selects approximately the top 50-100 storms. Yokoyama and Kamide (1997) noted a similar effect in a set of *Dst* storms and state that storm intensity increases more than linearly with duration.

The above analysis uses cumulative bins of storm intensity, so there

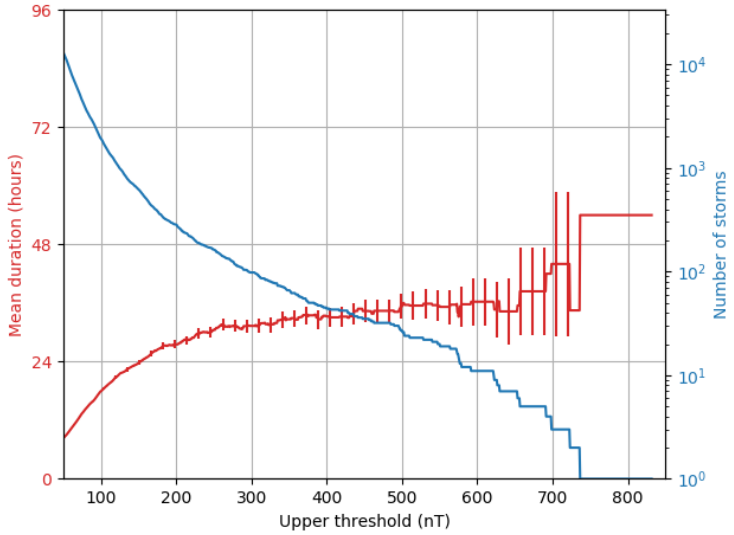


Figure 3.4: Mean storm duration (red line and left axis) and number of storms (blue line and right axis) as a function of increasing storm intensity, as defined by the upper threshold value. Storms have been organised into cumulative intensity bins for the upper threshold. The red line shows how the average duration increases as the upper threshold is increased. Error bars are plus and minus one standard error on the mean. The blue line and right-hand axis show the number of storms in each bin on a log scale.

are common events within bins. We now separate events using differential peak intensity bins of $70-90nT$, $90-110nT$, $110-150nT$, $150-190nT$, $190-230nT$, $230-300nT$, $300-400nT$, and above $400nT$. The number of storms in each bin can be seen in Figure 3.5 and decreases with intensity. Figure 3.5 does not represent the distribution function due to unequal bin widths (the distribution function is shown in Figure 3.2). The bins were chosen to balance the granularity of intensities examined and the number of events in each bin to ensure enough events for the following statistical analysis.

The distribution of the storm durations in each intensity class broadly resembles that of a lognormal distribution. This is shown for storms of all intensity classes in Figure 3.6. A lognormal distribution is defined by the mean of the logarithm of the values, μ, and the standard deviation of the logarithm of the values, σ. These parameters were found from the observed durations in each intensity class through Maximum Likelihood Esti-

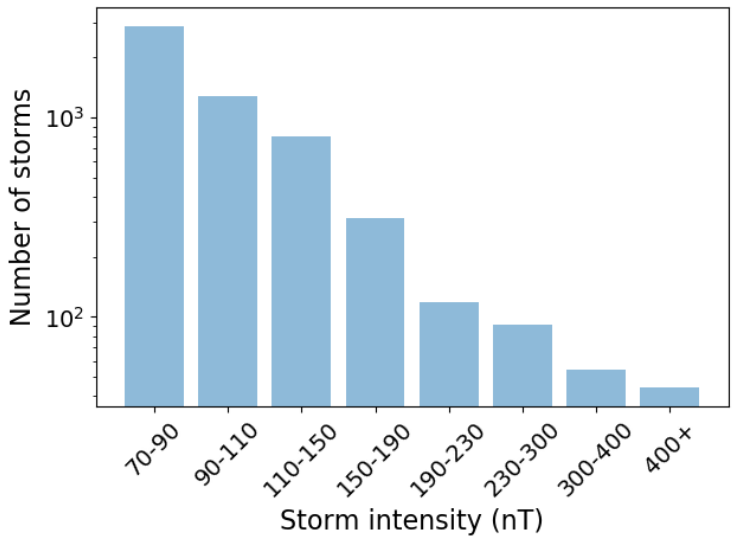


Figure 3.5: The number of storms in each class of peak storm intensity. Due to unequal bin size, this does not represent a distribution function.

mation (MLE) and used to create a lognormal distribution, plotted in Figure 3.6 in dark purple. The light purple distribution shows a histogram of the observed data as an estimate of the probability density function (PDF). By eye, the lognormal distribution provides a reasonable first-order match at all intensity thresholds. However, statistical testing suggests the lognormal distribution may not properly capture the high density of storms with 3-hour durations. As this study is primarily interested in larger intensity storms, we focus on using the lognormal fits for the remainder of the study.

We further note here the quantisation of the aa_H dataset into 3-hour intervals. The underlying physical phenomenon, that of storm duration, is of course continuous. This issue is commonplace in social and medical science for longitudinal studies through discrete-time survival analysis (Allison, 1982; Singer and Willett, 1993; Carlin et al., 2005). These methods will be investigated in future study. However, here we follow the simpler lognormal approach described above. The reason is twofold. Firstly, it would increase the complexity of the models presented here and it is instructive to begin with this simple and well understood approach.

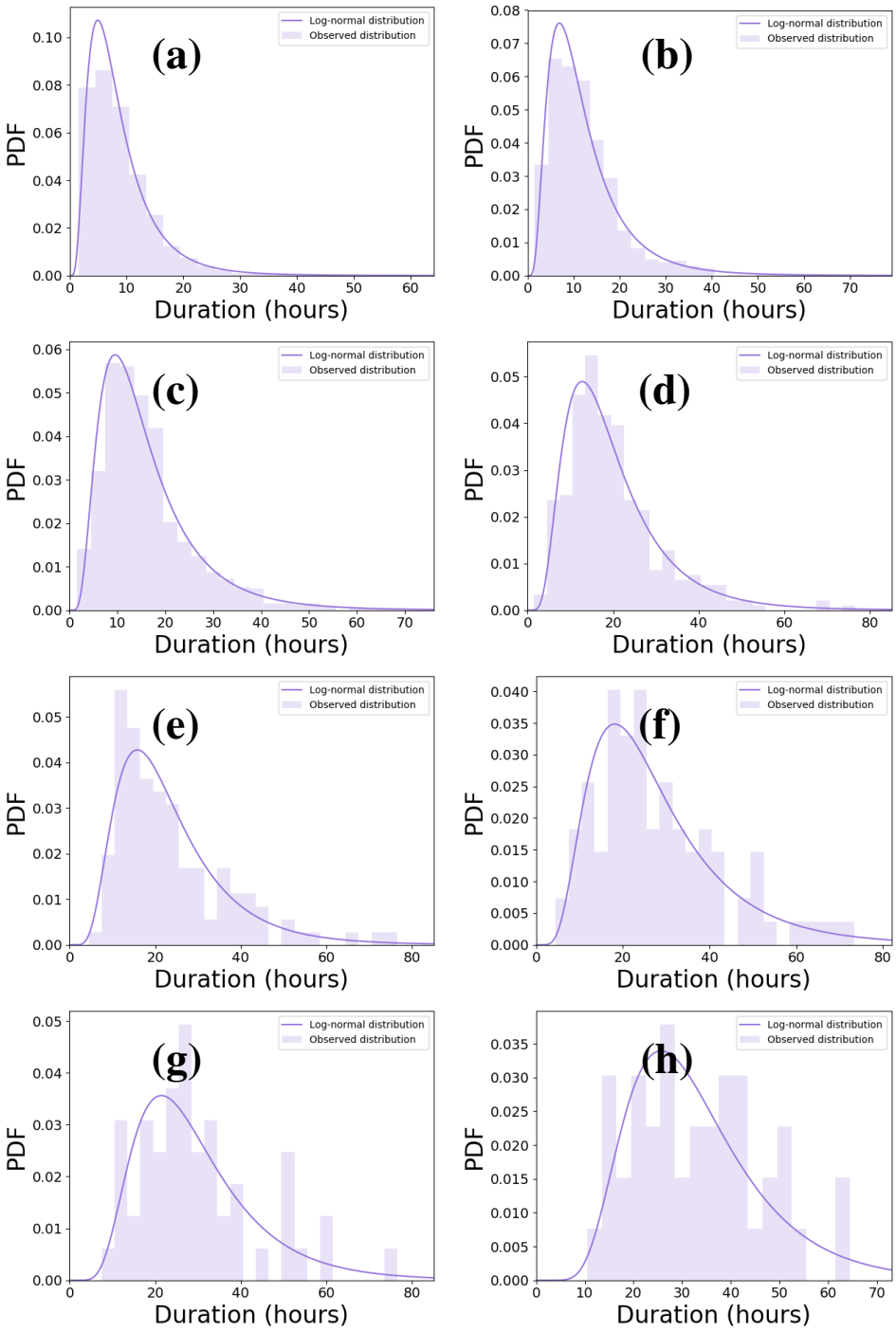


Figure 3.6: Storm durations for each class of peak intensity. The observed probability density function (PDF) of durations is shown in the histogram, while a lognormal with the same mean and width parameters is shown as the purple curve. The intensity classes represented are a) $70-90nT$**, b)** $90-110nT$**, c)** $110-150nT$**, d)** $150-190nT$**, e)** $190-230nT$**, f)** $230-300nT$**, g)** $300-400nT$**, and h) above** $400nT$**, respectively.**

Secondly, for the purposes of the current study, we are not seeking physical understanding through the shape of the distribution function and only require an estimate of the gross properties of the distribution for use as a robust forecast tool.

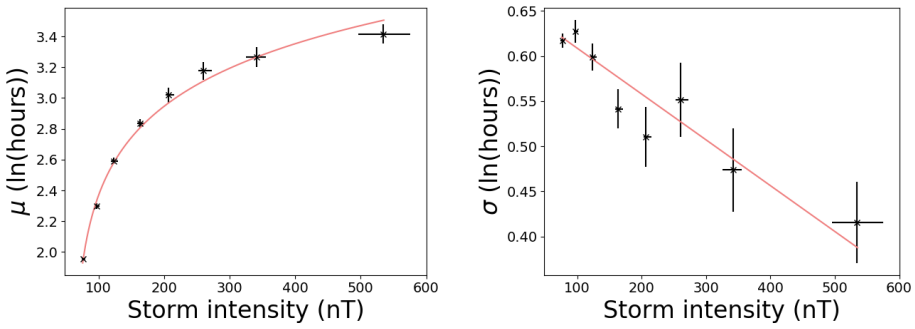


Figure 3.7: Variation of best-fit lognormal parameters to the observed storm duration PDF as a function of storm intensity. Best-fits curves accounting for uncertainties in both variables are shown mean storm duration of the log-space (left) and the standard deviation of the log-space (right). The uncertainty bars for storm intensity are the 33rd and 67th percentile (i.e. the 1-sigma range) around the median, for mean duration they are the standard error of the mean and for standard deviation they are the standard error of the standard deviation.

For each of the peak intensity classes, we have calculated the values of μ and σ for the lognormal fits to the duration distributions shown as the black points in Figure 3.7. The intensity classes are plotted on the x-axis at the median value of the intensity of storms in each class. It is clear from the points in the left panel of Figure 3.7 that μ increases as intensity increases, agreeing with the previous results in Figure 3.4 (i.e., duration increases as intensity increases). The parameter μ can be approximated as a function of storm intensity by

$$\mu(intensity) = A \ \ln(B \ \ intensity - C), \tag{3.1}$$

where A, B and C are free parameters. A least squares fit was implemented, incorporating the 33rd and 67th percentiles (i.e. the 1-sigma range) as the asymmetric uncertainty around the median of each intensity bin and the standard error around μ. The coefficients A, B and C were

found to be 0.455, 4.632, 283.143 respectively and this curve is plotted, along with uncertainty bars, in Figure 3.7 (left). Although the fit is based on weighted bin-centres of storm intensity, the equation can be used to interpolate for a given value of intensity.

Given the error bars, σ can be approximated by a linear fit to give σ as a function of the peak intensity. Figure 3.7 (right) shows the best fit line which has a shallow gradient of -5.08×10^{-4} and y-intercept at 0.659. The plot shows that the fit agrees with the points to within the estimated errors for 5 out the 8 intensity ranges (i.e., 62.5%) which is close to the 65% figure expected for the 1σ errors employed and hence a linear fit is deemed acceptable for our purposes.

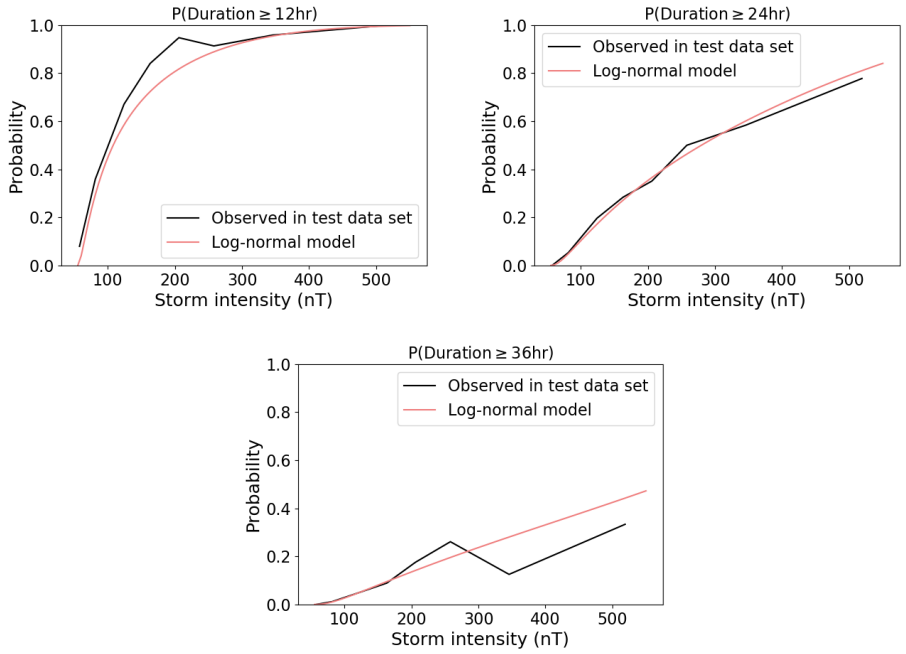


Figure 3.8: Comparison of observed and predicted storm durations. Plots show the probability a storm will exceed a certain duration as a function of peak intensity. The red line computes μ from Equation 3.1 and σ from the linear expression and uses these to compute the relevant probabilities. The black line is the observed probability from the training dataset for comparison. The aa_H data set was split into two equal-sized sets to avoid bias in the comparison.

The relations in Figure 3.7 can be used to produce an ideal log-normal distribution of durations for a given storm peak intensity. This, in turn, can be used to give an estimate of the probability of a storm of a

given intensity exceeding a given duration. To test how well this simple model works, we have made a comparison between the probability given by the model and the observed frequency in the dataset. The aa_H dataset was split up into two equal-sized sets composed of alternating years; a training dataset and a verification dataset. The training set was used to derive coefficients for Equation 3.1 and the linear fit of σ and thus create the model. The verification dataset was used to compute the observed probabilities, thus ensuring an unbiased comparison. Figure 3.8 shows the comparison for the probability of storms exceeding 12, 24, and 36 hours as a function of peak intensity. The black line is the observed probability from the verification aa_H dataset. The red line is the result of implementation of Equation 3.1 and expression for σ. It is seen that the model gives a generally good match to the observed probability but there are differences in detail. The largest storms ($> 400nT$) are likely to last longer than 24 hours and have a probability of approximately 0.4 of lasting longer than 36 hours. The smallest storms ($< 100nT$) have a low probability (0.2) of lasting longer than 12 hours. There is good agreement with the observed occurrence frequencies.

To further quantify the agreement between the model and observations we consider the associated reliability diagrams (Jolliffe and Stephenson, 2003), which compare the model predicted probabilities with the observed occurrence rates. We construct the reliability diagram by binning storms according to the model probability, P_M, of exceeding a given duration. For each model probability bin, we then determine the observed frequential probability, P_O, as a fraction of events which were actually observed to exceed the given duration. The model is reliable if $P_M = P_O$, i.e., follows the $y = x$ line on the reliability diagram. A reliability diagram is shown in Figure 3.9 for storms exceeding durations of 12, 24 and 36 hours. A 5-fold cross validation has been implemented such that independent test and training datasets have been split in five different ways and the analysis carried out on each as shown by the red lines. For all three

of the categories shown it is seen that the curves are generally slightly above the $y = x$ line showing a systematic underestimation of storm duration. However, for the analysis on 24 and 36 hour storms the reliability curves become sporadic for larger values of the model probability. This is due to a small sample size of events for which the model gives a large probability of the storm having these durations and hence very few observations with which to compute the frequential probability of the observations. On the whole, where there is sufficient sample size, the model probability and observed probability are in reasonably good agreement. To make this model more reliable it would require modest calibration. A simple approach to this would be to scale the predicted probability by a constant to reduce the systematic underestimation.

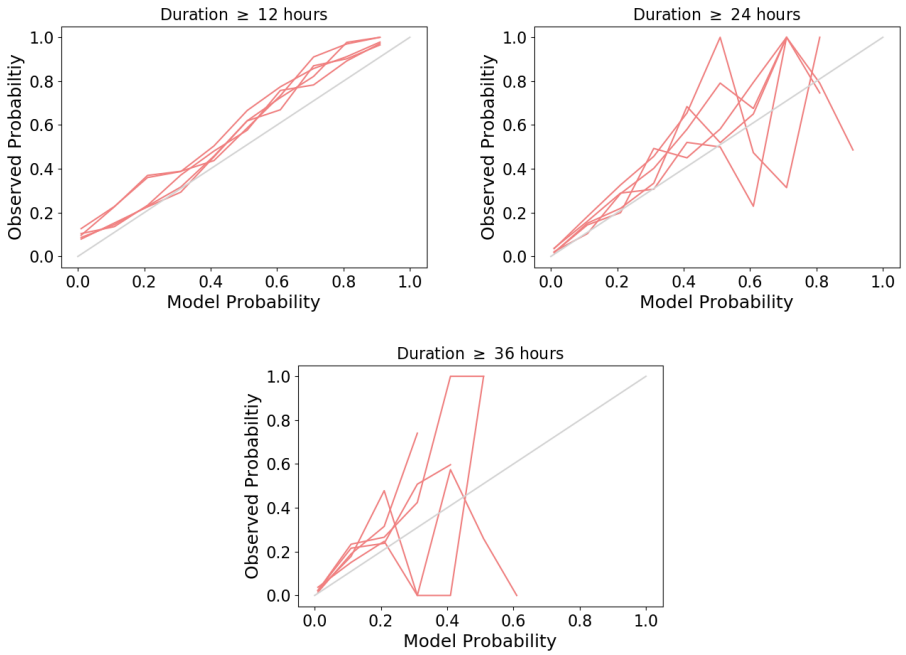


Figure 3.9: Reliability diagrams comparing the model predicted probabilities of storms exceeding 12, 24 and 36 hours with the observed occurrence frequency. A 5-fold cross validation has been carried out and the result of each shown in red. The grey line with a gradient of 1 shows the path of a truly reliable prediction.

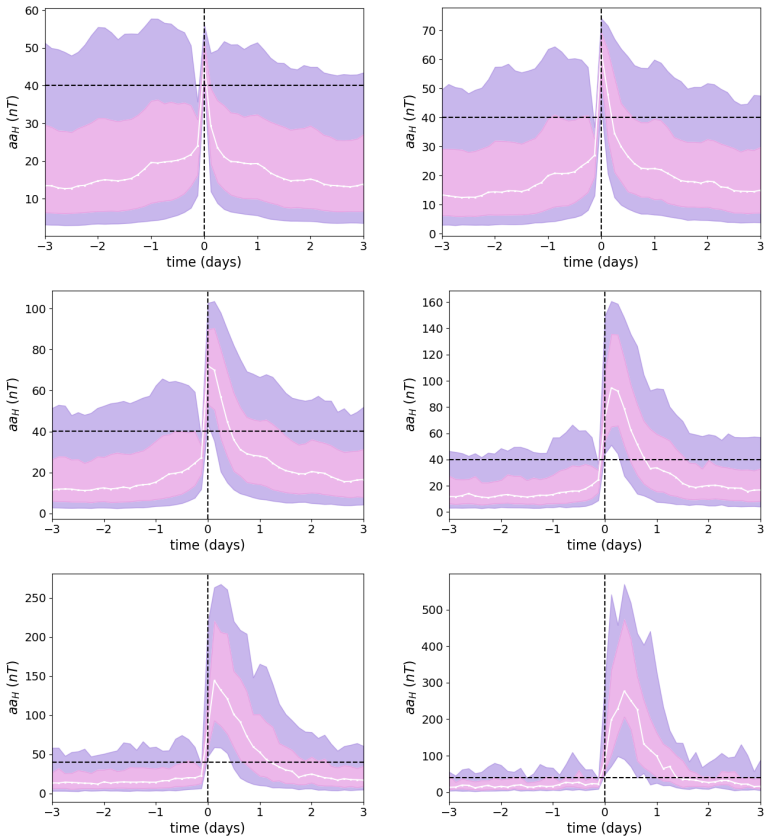


Figure 3.10: A superposed epoch analysis for different storm intensities. The trigger to start an epoch is the first measurement greater than or equal to the lower threshold. The white line represents the median, the pink colour band represents the inter-quartile range and the purple represents the $10th-90th$ percentile range. The horizontal dashed line marks the lower threshold. The plots shown are for storms of peak intensity $40-60nT$ (top left), $60-80nT$ (top right), $80-130nT$ (middle left), $130-200nT$ (middle right), $200-400nT$ (bottom left) and $400+nT$ (bottom right).

3.5 Superposed Epoch Analysis

Though Yokoyama and Kamide (1997) and Hutchinson, Wright and Milan (2011) previously conducted a superposed epoch analysis on geomagnetic storms with the *Dst* dataset, a study using longer term aa_H data is useful to help understand the intensity-duration relations described in the previous sections. Figure 3.10 shows a superposed-epoch analysis of the storms within each of the peak intensity classes. For each group, the $t=0$ epoch time is taken as the first data point which is on or over the lower threshold.

The tendency for more intense storms to have longer durations is once again apparent. The duration of the median for the six bins from least to most intense is 3, 6, 12, 21, 30, and 36 hours respectively. The smallest storms are almost symmetric in their rise/decay, but with increasing peak intensity the shape becomes increasingly "saw tooth" in nature, with a sudden rise and a long decay. Similarly, the time of peak intensity relative to the start of the storm varies with peak intensity. For the lower intensity storms the peak of the median occurs at the same time as the beginning of the storm. As the intensity class increases, the peak of the median storm occurs later and exhibits a more gradual build up towards peak intensity.

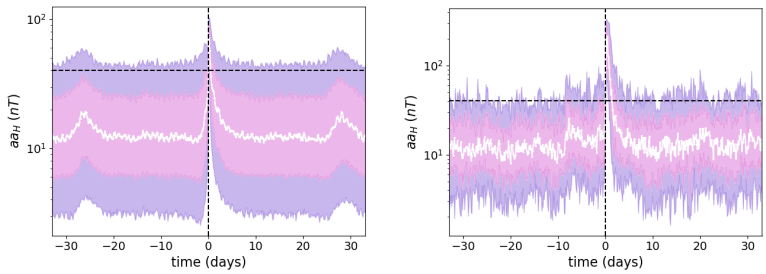


Figure 3.11: Superposed-epoch analysis of small (left) and large (right) storms over a 33-day time window, to look for recurrent signatures. Left: Superposed epoch analysis for storms with peak intensity $80-130nT$ showing signs of 27-day repeat, indicating the presence of recurrent solar wind structures. Right: Superposed epoch analysis for storms with peak intensities $300-400nT$ with no 27-day repeat, indicating a relative absence recurrence.

The 27-day solar rotation period (as observed from the Earth) is visible in the superposed-epoch analysis when the time window is extended, as in Figure 3.11. For a peak intensity bin of less than $200nT$, the 27-day repeating pattern is found albeit with a much lower intensity in the repeat events. This is suggestive of corotating interaction regions (CIRs Gosling and Pizzo, 1999), which are known to drive predictable recurrent geomagnetic activity (e.g. Owens et al., 2013).

3.6 Summary and Conclusions

In this study we have investigated the relationship between geomagnetic storm intensity and duration. We defined storms in the aa_H index by upper and lower thresholds where the upper threshold determines whether an event is classed as a storm and the lower threshold determines its duration.

Using this definition, we found that, on average, storm duration increases with storm intensity, as expected, but in a non-linear fashion. The median durations for the storms with peak intensity in the classes $70-90nT$, $90-110nT$, $110-150nT$, $150-190nT$, $190-230nT$, $230-300nT$, $300-400nT$, and above $400nT$ were found to be 6, 9, 15, 18, 21, 24, 27 and 33 hours respectively. This is similar to Lefevre et al. (2016) who found in the *aa* index that the weakest storms have a median duration of 6 hours and the strongest have a median duration of 30 hours. The longest storm, beginning on 19th September 1951, was 75 hours which is 18 hours shorter that the 93 hour storm analysed by Lefevre et al. (2016).

The distribution of storm duration is approximately lognormal when considering storms with peak intensity above around $150nT$. The lognormal parameters, μ and σ, computed for a number of storm intensity classes revealed that μ increases monotonically with storm intensity. However this is not the case for σ. Expressions for μ and σ as logarithmic and linear functions, respectively, of intensity were found, providing a method to estimate these parameters for a given storm peak intensity. The obtained lognormal distribution can then be used to find the probability of a storm of given intensity lasting longer than a certain duration. This simple model was compared to the observed occurrence probability. Good agreement is found. As expected, more intense storms had a higher probability of lasting longer.

An analysis with reliability diagrams revealed that, while the model

tends to underestimate the probability of a storm exceeding a given duration, the reliability curve follows the gradient of the $y = x$ line reasonably well. If the model was to be used operationally, the predicted probabilities could be easily calibrated to match the observed occurrence frequency.

Finally, a superposed epoch analysis was presented shedding light on the general shape of storms. More intense storms are shown to last longer and have their peak intensity further into the storm. It was observed that 27 day recurrent behaviour becomes less apparent in larger intensity storms, most likely reflecting the sources of the solar wind driving the storm. This is likely the result of co-rotating interaction regions (CIRs) formed by the interaction of fast and slow solar wind (Richardson, 2004). These structures can be long lasting and repeatedly impact the magnetosphere for many solar rotations but do not cause the very largest geomagnetic storms (Borovsky and Denton, 2006; Tsurutani et al., 2006). For storms with a higher intensity, the repeating pattern disappears. This is due to the dominant driver of very large geomagnetic storms being transient coronal mass ejections (CMEs) (Richardson, Cliver and Cane, 2001).

Future work could include a discrete-time survival analysis on storm duration. Although this adds complexity, it would provide a more rigorous structure on which to base the work. Another line of research will be to investigate whether the time history of an event could provide further information such as by using an analogue forecast approach.

Acknowledgements

The authors thank the National Environmental Research Council (NERC) for funding this work under grants NE/L002566/1 and NE/P016928/1.

The aa_H data is available at `https://www.swsc-journal.org/articles/swsc/olm/2018/01/swsc180022/swsc180022-2-olm.txt`

The authors declare no conflicts of interest.

Chapter 4

Forecasting Occurrence and Intensity of Geomagnetic Activity with Pattern-Matching Approaches

This chapter is composed of the second project I worked on during my PhD. I conceived the idea of forecasting a geomagnetic index using data driven methods as a result of my previous work on geomagnetic storm duration. My supervisor suggested the analogue ensemble and support vector machines as viable methods. I gave the main intellectual input into adapting these methods for this application and the evaluation of the approaches. I executed the work and wrote the paper for publication. This Chapter is published in the Journal of Space weather as Haines et al. (2021b).

C. Haines, M.J. Owens, L. Barnard, M. Lockwood, A. Ruffenach, K. Boykin and R. McGranaghan

Key points:

- Pattern-matching techniques are an effective way to forecast geomag-

netic activity.

- The analogue ensemble and support vector machine outperform 27-day recurrence and climatology.
- The best forecast approach for the end user will depend on their need for probabilistic forecast information.

Abstract

Variability in near-Earth solar wind conditions gives rise to space weather which can have adverse effects on space- and ground-based technologies. Enhanced and sustained solar wind coupling with the Earth's magnetosphere can lead to a geomagnetic storm. The resulting effects can interfere with power transmission grids, potentially affecting today's technology-centred society to great cost. It is therefore important to forecast the intensity and duration of geomagnetic storms to improve decision making capabilities of infrastructure operators. The 150-year aa_H geomagnetic index gives a substantial history of observations from which empirical predictive schemes can be built. Here we investigate the forecasting of geomagnetic activity with two pattern-matching forecast techniques, using the long aa_H record. The techniques we investigate are an Analogue Ensemble Forecast (AnEn), and a Support Vector Machine (SVM). AnEn produces a probabilistic forecast by explicitly identifying analogues for recent conditions in the historical data. The SVM produces a deterministic forecast through dependencies identified by an interpretable machine learning approach. As a third comparative forecast, we use the 27-day recurrence model, based on the synodic solar rotation period. The methods are analysed using several forecast metrics and compared. All forecasts outperform climatology on the considered metrics and AnEn and SVM outperform 27-day recurrence. A Cost/Loss analysis reveals the potential economic value is maximised using the AnEn, but the SVM is shown as superior by the true skill score. It

is likely that the best method for a user will depend on their need for probabilistic information and tolerance of false alarms.

Plain Language Summary

Space weather has the potential to disrupt society and the economy on a large scale. One such major impact is on power grids which can be damaged by disturbances in Earth's magnetic field caused by space weather events. As a result, it would be useful to have an accurate forecast of space weather that can help power grid operators make decisions about taking mitigating action. In this work, we test three forecasting techniques which utilise long historical records to exploit patterns in the data and hence predict future disturbances in Earth's magnetic field. We find that all three of the techniques provide valuable information and the best method depends on the individual needs of the forecast user.

4.1 Introduction

Geomagnetic storms present a significant threat to critical infrastructure both in space and on the ground (Oughton et al., 2017a; Cannon et al., 2013). Through solar wind energy input to the magnetosphere and the associated substorm process (e.g. Pulkkinen, 2007; Lockwood, 2019a), Earth's ionospheric current systems can be dramatically enhanced (Buonsanto, 1999). Rapid fluctuations in these enhanced ionospheric currents can generate geomagnetically induced currents (GICs. e.g. Boteler, 1994; Pirjola, 2000) in ground-based conductors, posing a particular risk to power grids and pipelines. To ensure that we minimise service disruption and mitigate economic cost (e.g. Eastwood et al., 2017a), there is a need for forecasting of both the intensity and duration of geomagnetic storms. Reliable forecasts improve the decision-making capabilities of operators of affected systems when taking mitigating action. However, current forecast capabilities are limited (Cannon et al., 2013; Koskinen et al., 2017).

Geomagnetic indices, which combine measurements from multiple ground-based magnetometers, are often used as a convenient measure of global geomagnetic activity because of their ability to reduce large-scale physical processes into a single time-series of observations. Commonly used measures include low latitude indices *Dst* and *SYM-H*, the mid-latitude range index Kp and high latitude auroral index, *AE* (e.g. Lockwood, 2013).

Current approaches to forecasting geomagnetic indices cover a spectrum of techniques from first principal, physics-based attempts (Pulkkinen et al., 2013a) (although even these often incorporate some empirical aspects in practice), through more empirical approaches, which range from those that still rely on domain-specific knowledge for their construction (Burton, McPherron and Russell, 1975), to those that are almost entirely data-driven (e.g. Gu et al., 2019).

Global Magnetohydrodynamic (GMHD) models simulate the magne-

tosphere using solar wind data as the input. These provide a physics-based representation of the magnetosphere which can be run in real time (Eastwood et al., 2017b) and thus can be used operationally. Three of the main GMHD models were tested by Pulkkinen et al. (2013a), with SWMF (Tóth et al., 2005, 2012) found to provide the most accurate reconstruction of $\frac{d\mathbf{B}}{dt}$, which is related to GIC intensity.

Owens, Riley and Horbury (2017a) argued for the use of empirical models for solar wind forecasting in conjunction with numerical physics-based models. Empirical models can add value because they have the advantage of being computationally cheap, meaning that they can be readily run in large ensembles to provide an estimate of uncertainty (Knipp, 2016). Empirical models can also parameterise unknown physics that a first-principles based model does not capture and act as a useful baseline with which to evaluate physics based models.

A range of empirical approaches have been attempted for geomagnetic index forecasting. Recently, Chandorkar, Camporeale and Wing (2017) developed a "one step ahead" forecast of the *Dst* index using an auto-regressive Gaussian process approach. It was tested on a set of 68 storms and concluded that for a 1-hour lead time, it out-preformed persistence on the metrics considered (mean absolute error, root-mean-square error and correlation coefficient). Zhang and Moldwin (2015) produced a probabilistic forecast of *SYM-H* and *AE* using solar wind parameters to construct a cumulative probability distribution that the index would exceed the given intensity thresholds. A non-linear autoregressive with exogenous inputs (NARX) approach was employed by Ayala Solares et al. (2016) for forecasting the Kp index. They found that, in general, the NARX approach gave good results for short and long lead times, however it failed to surpass the neural network models of Wing et al. (2005), to which they were comparing.

Other empirical approaches include: that of O'Brien and McPherron (2000) who employed a differential equation from Burton, McPherron and

Russell (1975) which maps the evolution of the corrected *Dst* index, Dst^*; that by Vassiliadis and Klimas (1995) who used a driven harmonic oscillator circuit analogy; Vassiliadis et al. (1995) who used linear and nonlinear filters to predict the *AL*, *AU* and *AE* indices. Camporeale (2019) summarised efforts using machine learning techniques to forecast geomagnetic indices. The majority of approaches use neural networks, however other machine learning techniques have been proposed. Lu et al. (2016) compared the use of Support Vector Machines (SVMs, Burges, 1998; Cortes and Vapnik, 1995), a machine learning approach that seeks to define a hyper-plane separating two classes, with neural networks for predicting intense storms in the *Dst* index using solar wind data as input parameters. Lu et al. (2016) concluded that SVMs out-perform neural networks for that application and can be improved further through the use of distance correlation learning (Székely, Rizzo and Bakirov, 2007). Liemohn et al. (2018) present an extensive list of works that forecast the behaviour of *Dst*, *SYM-H*, Kp, *AE*, *AL* and *AU* along with the metrics used to evaluate them. They showed no single metric is capable of metering a model for all applications. Therefore we must be rigorous in our application of evaluative metrics. In this work we explore numerous metrics, taking guide from Liemohn et al. (2018).

We here implement two-pattern matching forecasts, both requiring a large dataset for training, and an additional recurrence forecast. The first method is an analogue ensemble (AnEn) forecast, a purely empirical approach. This method assumes that previous observations provide a good analogue for likely future variations (Delle Monache et al., 2013). Thus a historical record that is sufficiently long and covers a large enough range of behaviour of the system can be used to identify previous periods when conditions are similar to the present. A forecast is constructed on the basis of the trajectories of the analogues forward in time. The "best" forecast in a deterministic sense is typically taken to be the median of the chosen analogues, but a large ensemble of analogues can provide prob-

abilistic information. Probabilistic forecasting helps quantify the forecast uncertainty, which benefits decision making, and can also be useful for evaluating a forecasting method (Knipp, Hapgood and Welling, 2018). An implementation of AnEn has been developed for this project in python and is available in Haines (2021a).

AnEn has been used for a variety of parameters in terrestrial weather forecasting (e.g., Van den Dool, 1989; Delle Monache et al., 2013), but has been surpassed by physics-based models. This is largely due to the inherently chaotic nature of the system, which means states can rapidly diverge with a small perturbation to the initial conditions. Recently Owens, Riley and Horbury (2017b) and Riley et al. (2017) used an analogue forecast for solar wind parameters and the *Dst* index with some success, finding that it outperformed benchmarks of climatology and 27-day recurrence.

The second method investigated here is the Support Vector Machine (SVM, Burges, 1998; Cortes and Vapnik, 1995), a supervised machine learning method for two-group classification. The volume and quality of data available, particularly in the aa_H index (see Section 4.2), and capabilities of modern computing means machine learning approaches are ripe for forecasting geomagnetic activity. SVMs seek to define a hyperplane which divides two classes (in this case 'storm' and 'no storm', for a given definition) and optimise it by maximising the distance between it and the closest data-points, called support vectors. To aid linear separability of the classes, the vector space of input parameters is mapped into a higher dimension space using implicit mapping functions with a defined kernel function (Burges, 1998). A brief overview of the SVM and application to space-weather is given by McGranaghan et al. (2018). An implementation of the SVM has been developed for this project in python and is available in Haines (2021b).

Here we will implement an SVM but, unlike many previous works, without the use of solar wind (exogenous) parameters as input and use

solely the time history of observations before the time of forecast. This gives a more direct comparison with AnEn and, importantly, allows the best use of the 150-year aa_H dataset, for most of which we do not have simultaneous solar wind observations. Given that both the solar wind transit time (between the usual point of observation the L1 point, and the magnetosphere), and the magnetospheric response time are small compared the time resolution of the aa_H dataset (3 hourly), this is not expected to reduce forecast capability.

The third forecast type considered is 27-day recurrence, which implicitly assumes the structure of the coronal magnetic field varies slowly compared to the solar rotation period. Thus the same region of the Sun is directed toward the Earth every 27-days. This assumption is generally more valid during solar minimum and the late declining phase of the solar cycle than during solar maximum periods. Near-Earth solar wind conditions and the resulting geomagnetic activity have long been know to exhibit 27-day recurrence (Chree and Stagg, 1928; Bartels, 1932, 1934; Owens et al., 2013). The recurrence pattern is also present in the occurrence of moderate storms in the aa_H index (Haines et al., 2019, see also Section 4.2). Watari (2011) used a 27-day recurrence forecast for the Kp index, concluding that it was a viable forecast method during the declining phase of the solar cycle but not for other parts of the cycle.

Section 4.2 describes the aa_H dataset. Section 4.3 describes our storm definition, the forecast methods considered and benchmarks used in this study. Section 4.4 compares the forecasts using metrics and techniques adopted from terrestrial weather forecasting (Henley and Pope, 2017). Many of the verification techniques are recommended by Liemohn et al. (2018) with the addition of Taylor diagrams and reliability diagrams (described below).

4.2 Data

We use the *aa* index (Mayaud, 1971), with recent recalibrations (aa_H Lockwood et al., 2018a,b). Using a single magnetometer station in each of the UK and Australia, *aa* provides a quasi-global measure of geomagnetic activity with particular sensitivity to the substorm current wedge (Lockwood, 2013; Ganushkina et al., 2015). In order to span 150 years back to 1868, *aa* must be constructed from three different stations in each hemisphere, which introduces issues of calibration. However, this results in the longest available record of geomagnetic activity. The recent recalibrations account for the variation of mean geographic location of the midnight sector auroral oval, due to drifts of the Earth's geomagnetic poles. They also allow for time-of-day/time-of-year response pattern of the stations, thereby reducing uncertainties related to using just two stations. Lockwood et al. (2019) showed aa_H agrees well with *am* (Mayaud, 1981), a similar index but with much greater suppression of longitudinal sampling effects achieved by using multiple stations in each hemisphere. The disadvantage of *am* for the present study, of course, is that the data sequence is much shorter as the greater data requirement means it can only be constructed back to 1959.

Unfortunately, aa_H is limited by its temporal resolution of 3-hours. Space weather impacts such as GICs occur due to magnetic fluctuation on a timescale of seconds and minutes. This means that a 3-hour range index cannot give direct information on potential GICs but it can give an idea of the low frequency variation in the magnetosphere. The 3-hour resolution of aa_H is more coarse than that of the 1-hour resolution of *Dst*. Because of this resolution difference, *Dst* actually has more data points, despite the record only being available for approximately 60 years. Although more data points is useful for training models, aa_H spans around eight more solar cycles than *Dst* and so captures a more complete picture of the space climate.

A further limitation of aa, and hence aa_H, is that it is derived from K indices (Bartels, Heck and Johnston, 1939) which are based on a quasi-logarithmic scale leaving quantisation in the dataset (Bubenik and Fraser-Smith, 1977). The uncertainty created from quantisation is largest in the larger values of aa_H, but still present in the small.

Chapman, Horne and Watkins (2020) investigated the use of aa_H in characterising extreme geomagnetic activity. They made a comparison of extreme aa_H with Dst, finding that there is good correspondence between the two indices and that it is possible to "read across" from extreme aa_H to extreme Dst.

4.3 Methodology

4.3.1 Storm definition

Various definitions of geomagnetic storms have been used, often dependent on the purpose of the study (Riley et al., 2018).The most common method is to set a threshold in a particular geomagnetic index (e.g. Lefevre et al., 2016) with values exceeding the threshold being defined as part of a storm, and a storm ends once the value of the index falls below the threshold. Kilpua et al. (2015) used a slight variation of this in which the last point of a storm is the first point below the threshold. Other approaches, such as that of Hutchinson, Wright and Milan (2011), involve a manual inspection of the data, looking for characteristic storm traces for each event more intense than a chosen threshold. While a more nuanced approach than blindly applying a geomagnetic threshold, it is labour-intensive to apply to a large dataset and is difficult to make truly repeatable.

In this work we are concerned with the full spectrum of geomagnetic activity which can lead to adverse impacts on infrastructure. We

point to the work of Schrijver (2015) and Schrijver et al. (2014a) which examine the impact of moderate space weather using insurance claims data. Schrijver (2015) highlights that high frequency, low impact events may cumulatively be comparable in economic cost to low frequency, high impact events. Congruently, Schrijver et al. (2014a) examines insurance claims on electrical equipment identifying significant rises on both the top 5% and top third of geomagnetically active days. We therefore seek to choose a storm definition that captures moderate geomagntic activity alongside the more rare, extreme events.

With the work of Schrijver (2015) and Schrijver et al. (2014a) in mind, we use the same storm definition as Haines et al. (2019). This approach uses a simple threshold, similar to Lefevre et al. (2016) and Kilpua et al. (2015), but, as in Gonzalez et al. (1994), with a data-informed threshold set at the 90th percentile of the dataset. For aa_H, this is a value of 40.1 nT. The start of the storm is the first point above the threshold and the end of the storm is the last point over the threshold. The effect of threshold on number of events is shown in Figure 2 of Haines et al. (2019).

4.3.2 Analogue Ensemble

To illustrate the methodology for the AnEn forecast, Figure 4.1 shows an event in the aa_H index from 2017-12-05. The observed time series (black) shows a storm, defined by exceeding a threshold of 40.1 nT, with storm onset at t_0. aa_H continues to rise to a peak value of around 100 nT at the next data point (3-hours later), then gradually falls back to non-storm conditions (i.e., below 40.1 nT). We identify the N previous events in the aa_H dataset which most closely match the pattern of the observed time series in the 24-hour time period before t_0, as described in more detail below. The time-series of these analogous periods are then projected forward to provide a probabilistic forecast after t_0. Also shown in Figure

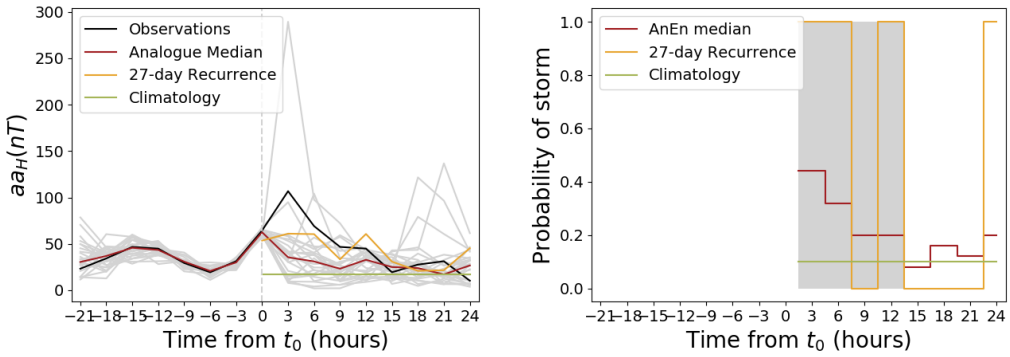


Figure 4.1: Left: An analogue ensemble forecast from 2017-12-05. The median AnEn forecast is shown in red with individual analogues in grey. The observed time-series of aa_H is shown in black. The benchmark forecasts are shown in yellow (27-day recurrence) and green (climatology). Right: The probability of a storm occurrence from each forecast method.

4.1 is a 27-day recurrence forecast which can be used as a deterministic forecast of storm intensity, or, using the storm-definition threshold, to give a dichotomous storm forecast i.e. that there will be a storm or that there will be no storm. The climatological mean value of aa_H is 17.5 nT, shown in green in the left panel, while the climatological probability of exceeding 40.1 nT is, by definition, 10% shown in green in the right-hand panel.

There are a number of aspects of the AnEn that must be tuned for the chosen application. The ensemble size, N, should be large enough to give sufficient resolution as a probabilistic forecast but small enough to ensure that the analogous periods are indeed analogous, particularly for rarer events such as larger storms. Values of N have been varied in the interval of $[10, 50]$ without significant difference in results. Therefore, for clarity, we have selected a single ensemble size of 25 for the presentation of results in the remainder of this study.

While the input data to the AnEn is simply the recent time history of observations (the previous 24 hours was shown in Figure 4.1), it is to be expected that some of these observations will be more relevant than others for predicting future behaviour. For a highly driven system like

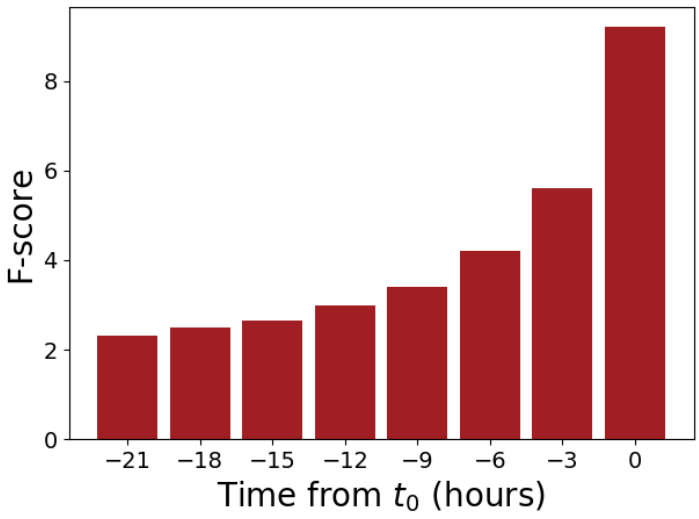


Figure 4.2: F-scores of each 3-hour aa_H data point in the 24-hours leading up to the forecast time t_0. The F-scores show the relevance of each parameter to the data point immediately after t_0

the magnetosphere, the most recent observations are more likely to contain useful information about future evolution than observations from 24 hours ago. We use the univariate F-score to determine the relevance of each input parameter, i.e. each 3-hourly aa_H data point in the previous 24 hours, to the subsequent data point of aa_H when forecasting with a 3-hour lead time (Pedregosa et al., 2011). These F-scores are shown in Figure 4.2 and we see that the most recent observation is the most relevant as expected. The F-score is used as a weighting factor when selecting the best analogues. The total level of agreement is then the inverse of the mean of the weighted squared errors over the 24-hour training window. The N analogues are then those with the lowest mean weighted squared error. These analogues are shown in Figure 4.1 by the grey lines converging as they approach t_0, and diverging significantly after t_0. Thus there is a wide range of possible future behaviour on the basis of previous analogues to recent conditions. In Figure 4.1 the median of these analogues is shown in red. It matches the observations in the "training period", i.e. -21 to 0 hours, but in this particular example, under-predicts the observed intensity in the forecast window, i.e. 0 to 24 hours.

The right-hand panel of Figure 4.1 shows the probabilistic nature of AnEn, 27-day recurrence and climatology. The grey shaded region shows when a storm was actually observed to occur and the coloured lines show the probability of storm conditions from each forecast considered. In this event the AnEn begins by predicting that a storm is likely with a probability of approximately 50%, which then drops over the next 12 hours to around 25%. (i.e., 25% of the ensemble members are predicting $aa_H >$ 40.1 nT at that time). The deterministic 27-day recurrence forecast does reasonably well in this particular example.

For analysing and testing the performance of the forecast methods, we implement them here as hindcasts, predicting past events for which we already have the observations of the predicted period. The whole aa_H dataset (excluding the 1-day prior, and 12.5-days subsequent, to t_0, which excludes the maximum extent of the training and evaluation windows) is made available for computing analogues. We use the full timeseries so that the results give an estimate of the predictive power of the model that would be deployed. The hindcasts have been run with the hindcast start time, t_0, at every point in the aa_H data set.

Due to the class imbalance between quiet and storm times, it is possible for a forecast to be valuable on average but perform poorly during storm times. For this reason we additionally select and validate hindcasts during only the storm subset.

4.3.3 SVM

The SVM is a commonly used classification algorithm which we implement here to classify whether or not a storm will occur. Given a sample of the input and the associated classification labels, the SVM will find a function that separates these input features by their class label. This is simple if the classes are linearly separable, as the function is a hyperplane. The samples lying closest to the hyperplane are called support

vectors and the distance between these samples and the hyperplane is maximised.

Typically, the samples are not linearly separable so we employ Cover's theorem (Cover, 1965) which states that linearly inseparable classification problems are more likely to be linearly separable when cast non-linearly into a higher dimensional spaces. Therefore, a kernel function is used to increase the dimensionality of the space. The Gaussian kernel is used for this purpose. It has a single hyperparameter, γ, which serves as a width parameter, determining the influence of a single data point on training. A sensitivity analysis has shown that an appropriate value for γ is 0.01.

On the basis of the aa_H values in the 24-hour training window, the SVM predicts whether the next 3 hours will be either a storm or not. By comparing this dichotomous hindcast with the observed aa_H, the outcome will be one of True Positive (TP, where a storm is correctly predicted), True Negative (TN, where no storm is correctly predicted), False Positive (FP, where a storm is predicted but not observed), or False Negative (FN, where a storm is not predicted but is observed). This is shown in the form of a contingency table in Figure 4.3 (top left).

For development of the SVM, the aa_H data has been separated into independent training and test intervals. These intervals are chosen to be alternate years. This is longer than the auto-correlation in the data (choosing, e.g. alternate 3-hourly data points, would not generate independent training and test data sets) but short enough that we assume there will not be significant aliasing with solar cycle variations.

Training is an iterative process, whereby a cost function is minimised. The cost function is a combination of the relative proportion of TP, TN, FP and FN. Thus while training itself, an SVM attempts to classify labeled data i.e. data belonging to a known category, in this case "storm" and "no storm" on the basis of the previous 24 hours of aa_H. If the SVM makes an incorrect prediction it is penalised through a cost

function which the SVM minimises. The cost parameter determines the degree to which the SVM is penalised for a mis-classification in training which allows for noise in the data. A sensitivity analysis showed that an appropriate value for cost parameter is 0.1. See also Section 4.4.3.

It is common that data with a class imbalance, that is containing many more samples from one class than the other, causes the classifier to be biased towards the majority class (Longadge, Dongre and Malik, 2013). In this case, there are far more non-storm intervals than storm intervals. Following McGranaghan et al. (2018), we define the cost of mis-classifying each class separately. This is done through the weight ratio $(W_{storm} : W_{no\ \ storm})$. Increasing the W_{storm} increases the frequency at which the SVM predicts a storm and it follows that it predicts "no storm" at a reduced frequency, as seen in the left column of Figure 4.3. The same results have been normalised to reveal more clearly how varying the class weights effects the predictions, shown in the right column of Figure 4.3. In this work we have varied W_{storm} and kept $W_{no\ \ storm}$ constant at 1. A user of the SVM method for forecasting may wish to tune the class weight ratio to give an appropriate ratio of false alarms and hit rate dependent on their needs.

4.3.4 Bench marking

Similar to Owens, Riley and Horbury (2017b) and following the recommendation of Liemohn et al. (2018), we use a benchmark hindcast method to distinguish between times when the studied hindcasts perform poorly and times when conditions are intrinsically more difficult to predict. For this purpose, we use climatology defined by the mean intensity of the entire dataset or, for a probabilistic hindcast of storms, the fraction of measurements in the entire dataset which qualify as storm events. These values are 17.5 nT and 10% respectively.

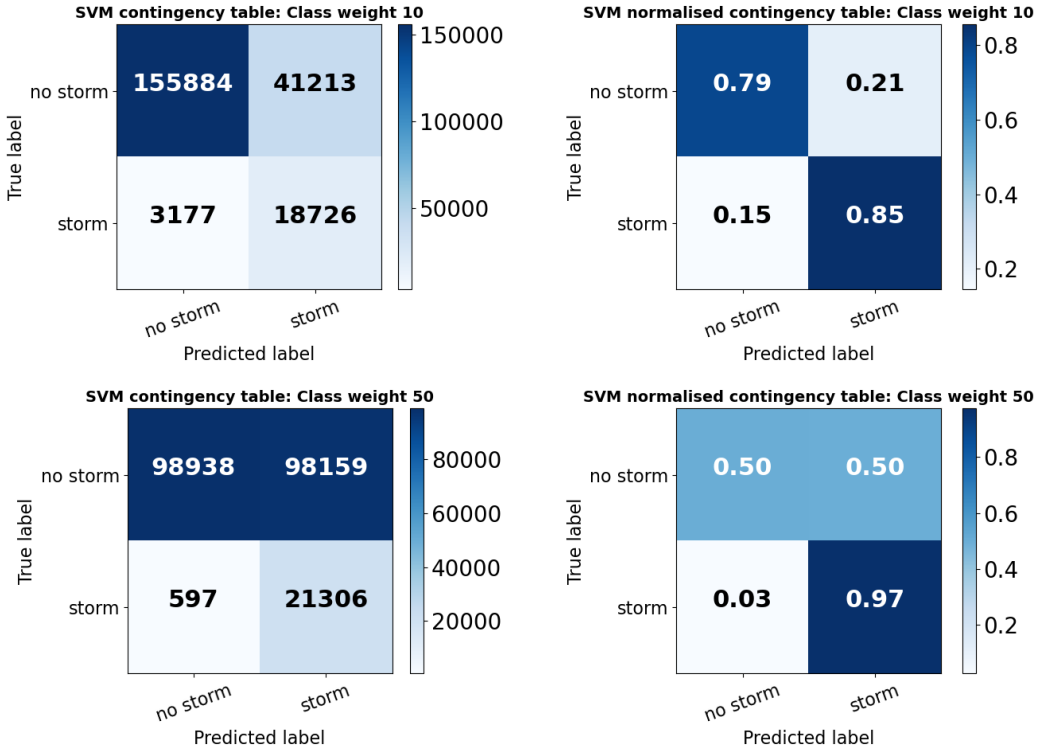


Figure 4.3: Examples of contingency tables, sometimes called a confusion matrix, for the SVM showing the number of occurrences (left) and normalised frequencies (right) of true and false positives and negatives. Top: SVM trained with a class weight of 10. Bottom: SVM trained with a class weight of 50. Changing the class weight affects the ratios of TP, TN, FP and FN.

4.4 Results

4.4.1 AnEn deterministic intensity hindcast

In the present section we consider the deterministic performance of the AnEn by reducing the ensemble to the median value. In this and Section 4.4.2 we present results for three subsets of the aa_H data set. The subsets are: the entire aa_H dataset; the occasions on which a storm was observed at a 3-hour lead time; the occasions on which a storm was predicted by the AnEn median at a 3-hour lead time. The second subset includes only true positives, whereas the third subset includes false negatives and true positives. There are many more non-storm events than storm events in the dataset so a hindcast always predicting no storm would fare well. These

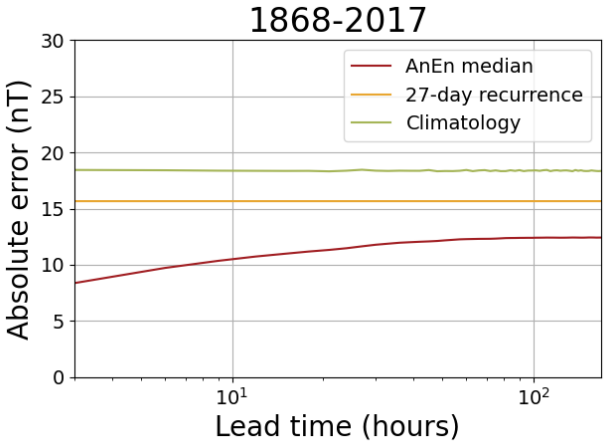
subsets of aa_H help to distinguish whether a hindcast method has any predictive power of storm events.

Figure 4.4 shows the mean absolute error (MAE) of the deterministic hindcasts of aa_H intensity compared to observations for lead times up to 300 hours (12.5 days). Figure 4.4a shows the MAE when the hindcasts are initiated from every time step in the aa_H data-set. The general pattern is for the AnEn median to produce the lowest MAE, followed by 27-day recurrence and climatology with the highest. While the MAE in the 27-day recurrence and climatology are relatively constant with lead time, AnEn clearly displays higher accuracy for shorter lead times until it plateaus at approximately 50 hours lead time. This suggests that the usefulness of information in the preceding 24-hours to t_0 is greatest for short lead times. Figure 4.4b shows the error of hindcasts on which the point immediately after t_0 is classed as a storm, as defined by a threshold of 40.1 nT. All hindcast methods have a high MAE for short lead times which drops off for longer lead times. At long lead times, approximately the same order of accuracy of the hindcasts exists for this storm dataset as for the whole dataset. At shorter lead times, the storms are in progress, and thus the observed aa_H is high, and the same percentage error leads to a higher MAE than non-storm times. This issue, as well as that of storm conditions generally being more difficult to predict than quieter times, can be addressed by computing the skill of the hindcasts relative to a reference forecast. In essence, it allows discrimination between poor forecasts and periods which are inherently difficult to forecast.

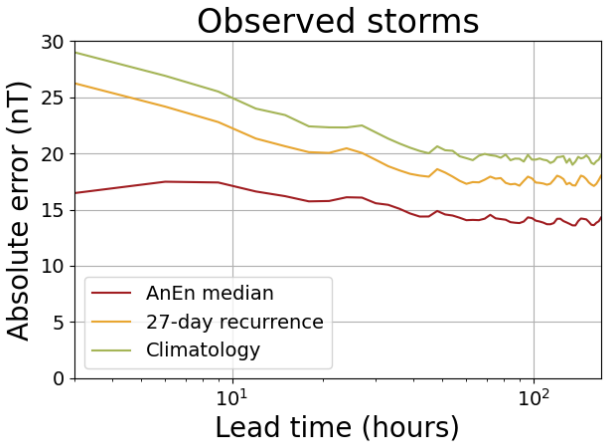
Skill is computed as:

$$\text{skill} = 1 - \frac{\text{forecast error}}{\text{reference error}}, \tag{4.1}$$

Thus skill can vary between $-\infty$ and 1, where a more positive value is more skilful, and zero is identical performance to the reference



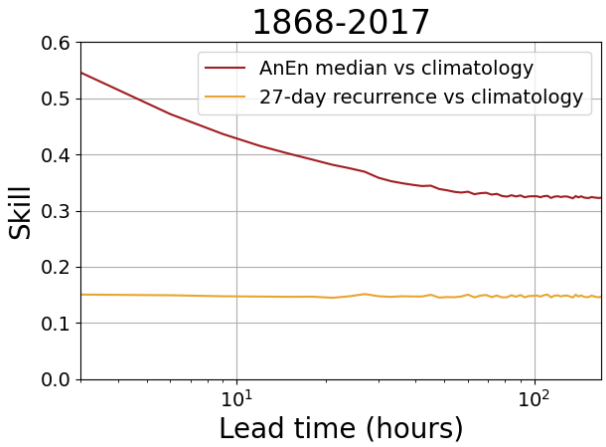
(a) Error on entire dataset



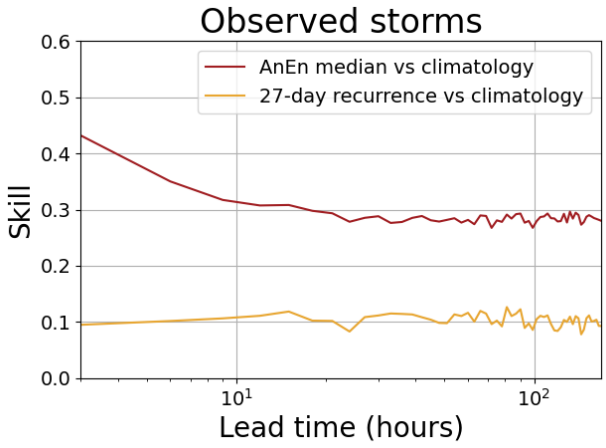
(b) Error on observed storms

Figure 4.4: A comparison of the mean absolute error (MAE) of AnEn median, 27-day recurrence and the intensity climatology for a range of lead times up to 300-hours. a) MAE when the hindcast is run for t_0 at every point in the aa_H data set. b) MAE for t_0 as the start of known storm events only.

hindcast. Figure 4.5 shows the skill of the AnEn median and 27-day recurrence relative to climatology. This is done for the whole data-set in Figure 4.5a. Both AnEn median and 27-day recurrence have positive skill for all lead times. AnEn median achieves substantially higher skill, especially for shorter lead times. Figure 4.5b makes the same comparisons considering only the time periods immediately following observed storm onsets. We again see that AnEn median has positive skill, however skill is reduced by approximately 10% compared to the whole dataset.



(a) Skill on entire dataset



(b) Skill on observed storms

Figure 4.5: A comparison of skill for AnEn median and 27-day recurrence with respect to climatology. a) The entire aa_H dataset. b) Restricted to the period following observed storms.

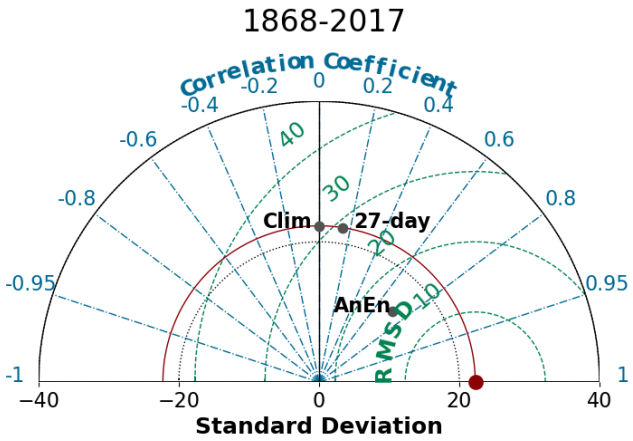
Figure 4.6 shows Taylor diagrams (Taylor, 2001; Owens, 2018) that summarise the performance of a hindcast in terms of three metrics, visualised on a single plot. These metrics are the standard deviation of the hindcast, linear correlation coefficient between hindcast and observed intensities, and the centred root-mean-squared distance (RMSD) between hindcast and observed intensities. These three metrics provide measures of agreement in both statistical terms (standard deviations) and the correspondence on a point-by-point basis (correlation and RMSD). A perfect hindcast would lie on the red dot, having a hindcast standard deviation match-

ing that observed, correlation coefficient of 1, and centred RMSD of 0. Put simply, the further a hindcast lies from the red dot, the worse the forecast is, and the direction of displacement can help diagnose the problem with the forecast.

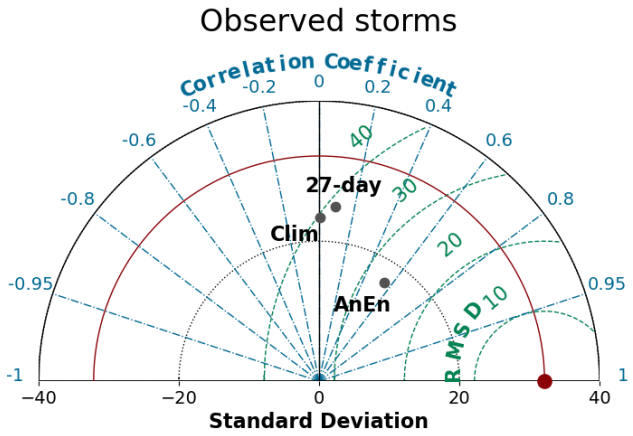
Figure 4.6a shows the three hindcast types for the whole 150 year period of aa_H data for a 3-hour lead time. AnEn median provides the best hindcast by two of the three metrics considered but has a smaller and reduced standard deviation than both climatology and 27-day recurrence compared to the observations (by the construction, this is expected: both benchmarks are direct, unaveraged samples of the observations against which they are tested. Conversely, by taking the median of the AnEn, the variability will be reduced). Figure 4.6b shows the hindcasts run only for observed storm onsets. The general pattern is similar to that of when the hindcast is run on the whole dataset.

4.4.2 AnEn probabilistic dichotomous hindcast

In Section 4.4.1, the AnEn was reduced to a deterministic hindcast of intensity by considering only the ensemble median. But the AnEn can be used as a probabilistic hindcast. We here consider the probabilistic hindcast of (dichotomous) event occurrence, in this case the occurrence/non-occurrence of storms, by considering all the ensemble members together to form a probability distribution of future evolution. While a deterministic intensity hindcast looks to minimise the error of the prediction, a probabilistic dichotomous hindcast aims to predict event occurrence at the observed frequency. That is to say if a hindcast makes a prediction with $x\%$ certainty it is said to be reliable if, on average, an event is subsequently observed $x\%$ of the time. Systematic bias in hindcast probability can be quantified with a reliability diagram (Jolliffe and Stephenson, 2003; Sharpe and Murray, 2017) which compares predicted and observed probabilities. A perfectly reliable hindcast would follow the $y = x$ line, as shown in Figure



(a) Taylor diagram on entire dataset.



(b) Taylor diagram on observed storms

Figure 4.6: Taylor diagrams comparing the AnEn median hindcast to climatology and 27-day recurrence for a 3-hour lead time. The diagrams summarise the RMSD (root mean square deviation) (nT), correlation coefficient and standard deviation (nT) of the hindcast s. a) The entire aa_H dataset. b) Restricted to the period following observed storms.

4.7 by the light grey line. A forecast giving a reliability curve below this line shows overestimation of event likelihood and a reliability curve over the line shows underestimation of the likelihood of an event. Figure 4.7a shows AnEn hindcast of storms from all data points in the aa_H dataset for a hindcast lead time of 3 hours. On the whole, the curve fits well to $y = x$ with a slight overestimate of storm probability for larger values of hindcast probability. When considering only known storm events, the AnEn

is less reliable, as shown Figure 4.7b. While the curve largely follows the $y = x$ line, there is an underestimate of storms for low hindcast probability and an overestimation for high.

This underestimate may be an indicator that there is insufficient information in the observed time-series leading up to t_0 in order to differentiate between storms and not storms, i.e. many of the analogues found for the build up to a storm may be associated with only a small increase in aa_H because there are simply many more instances of smaller variations than larger. This would bias the hindcast towards predicting smaller storms and thus underestimating the probability of an event.

4.4.3 SVM Classification

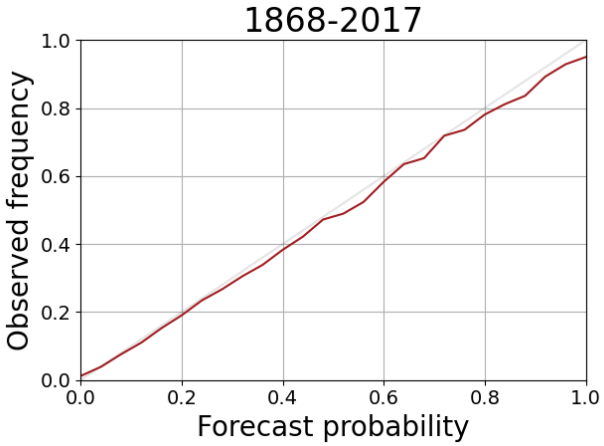
To evaluate the SVM and AnEn for storm classification we need a metric that is robust to class imbalance. This is because using a storm definition of the 90th percentile means we have nine non-storm events for every storm, so a prediction method that always predicts non-storm would do very well under many metrics. The True Skill Score (TSS) is a combination of TP, FP and FN only, meaning that it can handle imbalanced classes, though it neglects a model's ability to correctly predict non events, which can be valuable in its own right. TSS has been recommended and used in the space weather community (e.g. Bloomfield et al., 2012; McGranaghan et al., 2018).

TSS is defined as

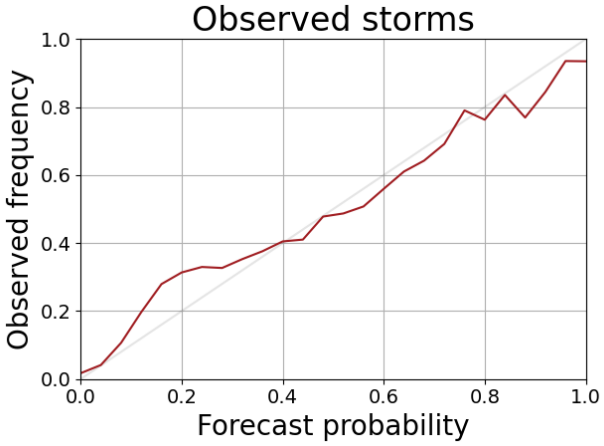
$$TSS = \frac{TP}{TP+FN} - \frac{FP}{FP+TN} \tag{4.2}$$

which gives a score between $-\inf$ and 1 where 0 is a hindcast with no skill and 1 is a perfect hindcast.

TSS has been computed for the SVM, AnEn median and 27-day recurrence in Figure 4.8. Figure 4.8a shows TSS for SVMs with differ-



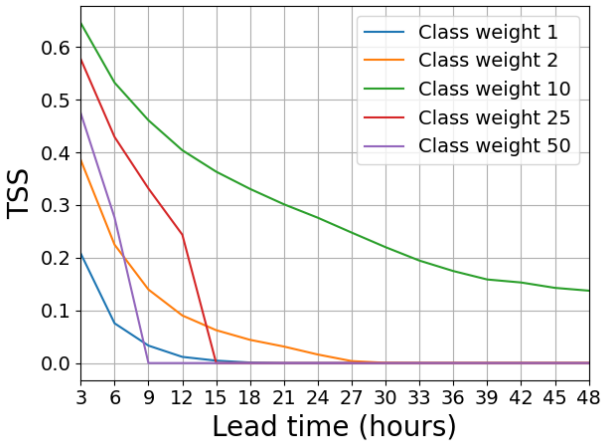
(a) Reliability diagram on entire dataset



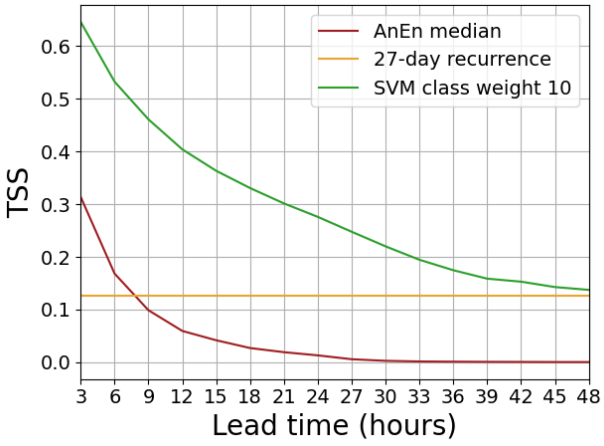
(b) Reliability diagram on observed storms

Figure 4.7: A reliability diagram of the analogue ensemble dichotomous storm hindcast for a 3-hour lead time. The grey line represents the path of a perfectly reliable hindcast, wherein events are forecast with a probability equal to the observed occurrence rate. a) The entire aa_H dataset. b) Restricted to the period following observed storms.

ent class weights. A unique SVM has been trained for each value of lead time. We see that using class weight 10 gives the best performance with positive, reducing skill for the full 48 hours. The other SVMs perform considerably worse, particularly for lead times greater than 3-hours. SVMs with class weight of 1 and 2 end up predicting no storm events will occur at longer lead times and SVMs with class weights 25 and 50 predict storms always occur at longer lead times. SVM with class weight 10 seems to strike a good balance, as it approaches the proportions of



(a) TSS for SVMs of different class weights



(b) TSS for AnEn median, 27-day recurrence and SVM with class weight 10

Figure 4.8: TSS for lead times of 3 to 48 hours. a) shows the TSS for the SVM with a range of class weights. We see that using class weight 10 gives the best skill. b) TSS for AnEn median, 27-day recurrence and SVM with class weight 10. We see that the SVM skill exceeds that of the other hindcasts.

storm and non-storm events in the dataset.

In Figure 4.8b we compare the TSS of SVM class weight 10 to TSS of AnEn median and 27-day recurrence. Both SVM and AnEn median have a similar shape of diminishing skill with lead time, however SVM has a far superior TSS at all lead times considered. The TSS of 27-day recurrence is a flat line since its lead time is, in essence, always 27 days. 27-day recurrence exceeds AnEn median at 9 hours and longer and is approximately equivalent to SVM at 48 hours. It suggests that the

AnEn median does not have predictive power for the storm class at longer lead times and quickly goes back to predicting quiet-time.

Different forecast applications will have different tolerances for false alarms and missed events. A limitation of TSS is that it treats FP and FN the same and does not give useful information for users with an unbalanced tolerance. To accommodate this, and as a further comparison of the hindcasts, a Cost/Loss analysis (Murphy, 1977a; Richardson, 2000; Owens, Riley and Horbury, 2017c) is implemented in Figure 4.9. A space-weather example of how a Cost/Loss analysis is carried out is shown in Figure 7 of Owens, Riley and Horbury (2017c). In short, C is the economic cost associated with taking mitigating action when an event is predicted (whether or not it actually occurs) and L is the economic loss suffered due to damage if no mitigating action is taken when needed. For a deterministic method, such as the SVM, each time a storm is predicted will incur a cost C. Each time a storm is not predicted but a storm occurs a loss L is incurred. If no storm is predicted and no storm occurs then no expense in incurred. By considering some time interval, the total expense can be computed by summing C and L.

A particular forecast application will have a C/L ratio in the domain (0,1). This is because a C/L of 0 would mean it is most cost effective to take constant mitigating action and a C/L of 1 or more means that mitigating action is never cost effective. In either case, no forecast would be helpful. The power of a Cost/Loss analysis is that it allows us to evaluate our methods for the entire range of potential forecast end users without specific knowledge of the forecast application requirements. End users can then easily interpret whether our methods fit their situation.

For a probabilistic forecast, a similar process is applied with the difference that action is taken only when the forecast probability exceeds C/L. See Owens, Lockwood and Barnard (2020) for more information.

Once total costs have been calculated, the potential economic value

(PEV) is given by

$$PEV = \frac{E_C - E}{E_C - E_0}, \tag{4.3}$$

where E_C is the total expense of using a probabilistic climatological forecast, E is the total cost of the forecast under consideration and E_0 is the total cost of a perfect forecast. The PEV of a forecast is therefore equivalent to climatology where PEV=0 and to a perfect forecast where PEV=1. Note that a user's Cost and Loss do not need to be computed in financial terms, only the ratio of the two values is necessary: high C/L suggests that false alarms should be avoided, whereas low C/L suggests missed events would be more problematic.

Figure 4.9 shows the PEV of the SVM with a range of class weights (CW), probabilistic AnEn and 27-day recurrence. Here a deterministic Cost/Loss analysis has been implemented for SVM and 27-day recurrence, and a probabilistic Cost/Loss for AnEn. The shaded regions indicate which hindcast has the highest PEV for that Cost/Loss ratio. The probabilistic AnEn has the highest PEV for the majority of the Cost/Loss domain although SVM has higher PEV for lower Cost/Loss ratios. It is possible that an increased resolution in the scan of class weights would bring the SVM out on top for a larger part of the domain. However certain users may appreciate that the hindcasts generally have a similar PEV for parts of the Cost/Loss domain and will find it more valuable to have the probabilistic hindcast of the AnEn. It also highlights that the 'best' hindcast is dependent on the context in which it is to be employed.

4.5 Future directions

There are a number of possible ways the forecast schemes presented here could be improved in the future. By taking the fraction of ensemble members which result in a storm to be the AnEn hindcast probability of a storm we are implicitly assuming that the analogues form a single distri-

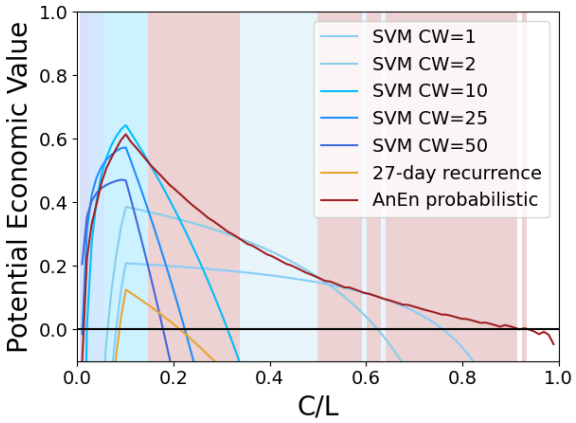


Figure 4.9: A Cost/Loss analysis showing the potential economic value of the hindcasts relative to value of a perfect hindcast (PEV=1) and climatology (PEV=0). The shaded areas indicate which hindcast type has the highest PEV for that C/L ratio. Negative values of potential economic value are shown only down to -0.1.

bution. This potentially throws away information about different modes of behaviour. Clustering ensemble members together using K-means clustering is a way in which we could use the data to extract a number of possible future scenarios. An example is shown in Figure 4.10. The observed storm peaks at $t = 21$ hours, however this behaviour is not captured by the median of the ensemble members or easily visible amongst the grey ensemble member lines. However the scenario in which the storm has a late peak is picked out as a possible mode of behaviour by the red cluster in the right of Figure 4.10, identified by K-means clustering. Here, the clustering algorithm aims to minimise the sum of the square error between the ensemble member and the cluster it is in. The number of clusters has been chosen by using an ”elbow plot” which identifies appropriate K values by minimising both the sum of square errors and the number of clusters.

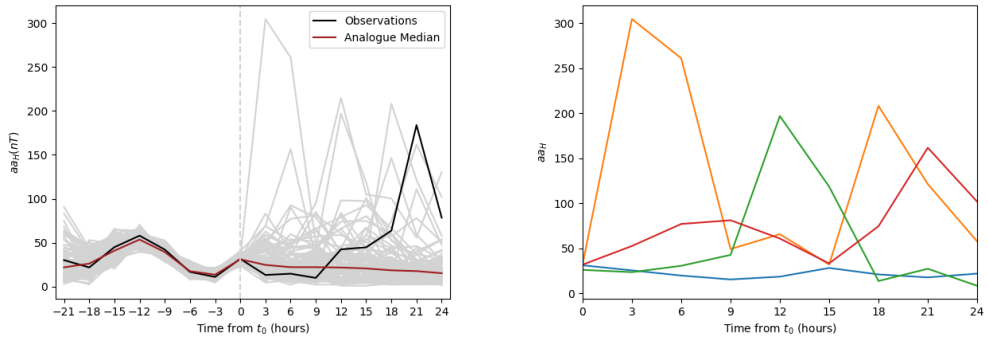


Figure 4.10: Left) An event from 27/2/1997 with an AnEn hindcast. Right) Clusters from the ensemble members using K-means clustering.

4.6 Discussion & Conclusions

This study has considered the effectiveness of two pattern-matching methods in hindcasting the aa_H index. These are an analogue ensemble (AnEn) and a Support Vector Machine (SVM). We have additionally considered the 27-day recurrence hindcast for context. AnEn and 27-day recurrence can be used as intensity hindcasts and AnEn can also give a probability distribution for dichotomous-event hindcast. SVM has only been implemented as a deterministic dichotomous-event hindcast.

Reducing the AnEn to a deterministic intensity hindcast by taking the median value, it outperformed the benchmark of climatology both when applied to the whole aa_H dataset and limited only to observed storm onsets. AnEn clearly outperformed the benchmark for absolute error and skill for lead times up to a week. 27-day recurrence outperformed the benchmark but did not perform as well as AnEn.

When considering the AnEn as a probalistic hindcast of storm occurrence, it was found to be highly reliable when hindcasting each data point in the aa_H dataset, in that the predicted probability closely matches the observed frequency of events. Reliability was found to drop slightly when considering only storm events. In particular, the AnEn underestimates storms when it had a low certainty of a storm and overestimates the prob-

ability of a storm when it was reasonably certain. The underestimation may be an indicator that there is insufficient information in the observed time-series leading up to t_0 in order to differentiate between storms and not storms. I.e. many of the analogues found for the build up to a storm may be associated with only a small increase in aa_H because there are simply many more instances of smaller variations than larger. This would bias the hindcast towards predicting smaller storms and thus underestimating the probability of an event.

Finally, an SVM was implemented for a range of class weights and compared to AnEn and 27-day recurrence using TSS and a Cost/Loss analysis. The SVM was more skilful than AnEn by TSS, though neither hindcast had a conclusively higher potential economic value across the Cost/Loss domain. It is likely that the best method for a user will depend on their individual circumstances.

Acknowledgements

The authors thank the National Environmental Research Council (NERC) for funding this work under grants NE/L002566/1 and NE/P016928/1.

The aa_H data is available at `https://www.swsc-journal.org/articles/swsc/olm/2018/01/swsc180022/swsc180022-2-olm.txt`

Code for AnEn is available in Haines (2021a) at `https://doi.org/10.5281/zenodo.4604487`.

Code for SVM is available in Haines (2021b) at `https://doi.org/10.5281/zenodo.4604485` which includes the code for splitting the data into train and test sets. A data file containing a list of storms in aa_H is also here.

Chapter 5

Towards GIC forecasting: Statistical downscaling of the geomagnetic field to improve geoelectric field forecasts

This Chapter contains my third project from my PhD. The idea of downscaling was discussed as part of the NERC Space Weather Impacts on Ground Systems (SWIGs) highlight topic, and it was determined that I could break initial ground in this area as part of the University of Reading's contribution. While the overarching goal had been envisioned, I was responsible for determining an appropriate proof-of-concept scheme and implementing the methodology. I provided the main intellectual input for the evaluation of the methodology and I wrote the resulting paper. This work is published in the Journal of Space Weather as Haines et al. (2021a).

C. Haines, M.J. Owens, L. Barnard, M. Lockwood, C.D. Beggan, A.W.P. Thomson, N.C. Rogers

Key points:

- Operational global MHD models do not fully capture the ground-level magnetic field variability important for modelling induction hazards

- We provide a proof of concept model to statistically introduce realistic, high-resolution perturbations with which to drive an impacts model
- Our downscaling scheme outperforms a reference linear-interpolation approach under a range of metrics

Abstract

Geomagnetically induced currents (GICs) are an impact of space weather that can occur during periods of enhanced geomagnetic activity. GICs can enter into electrical power grids through earthed conductors, potentially causing network collapse through voltage instability or damaging transformers. It would be beneficial for power grid operators to have a forecast of GICs that could inform decision making on mitigating action. Long lead-time GIC forecasting requires magnetospheric models as drivers of geoelectric field models. However, estimation of the geoelectric field is sensitive to high-frequency geomagnetic field variations which operational global magneto-hydrodynamic models do not fully capture. Furthermore, an assessment of GIC forecast uncertainty would require a large ensemble of magnetospheric runs, which is computationally expensive. One solution that is widely used in climate science is "downscaling", wherein sub-grid variations are added to model outputs on a statistical basis. We present proof-of-concept results for a method that temporally downscales low-resolution magnetic field data on a 1-hour timescale to 1-minute resolution, with the hope of improving subsequent geoelectric field magnitude estimates. An analogue ensemble (AnEn) approach is used to select similar hourly averages in a historical dataset, from which we separate the high-resolution perturbations to add to the hourly average values. We find that AnEn outperforms the benchmark linear-interpolation approach in its ability to accurately drive an impacts model, suggesting GIC forecasting would be im-

proved. We evaluated the ability of AnEn to predict extreme events using the FSS, HSS, cost/loss analysis and BSS, finding that AnEn outperforms the "do-nothing" approach.

Plain Language Summary

Forecasting space weather impacts on ground-based systems, such as power grids, requires the use of computer simulations of the disturbance of the Earth's magnetic field by the solar wind. However, these computer simulations are often too smooth, underestimating small and fast variations in the Earth's magnetic field which are important for modelling induction hazards that may affect power grids. In this paper we present a proof-of-concept scheme that attempts to introduce realistic high-frequency variations using the idea of looking at how the field has previously behaved in historical events. We test the model and find that it allows for better impact forecasting than if our scheme is not used.

5.1 Introduction

Intensification of magnetospheric and ionospheric current systems drives changes in the geomagnetic field measured on the ground ($\frac{d\mathbf{B}}{dt}$) which induces an enhanced geoelectric field, as expressed by the Maxwell-Faraday equation. The induced geoelectric field drives currents within the Earth that can enter grounded conducting networks as geomagnetically induced currents (GICs) (Koskinen et al., 2017; Pulkkinen et al., 2017). GICs can flow into the power grid through earthing points at substations (Oughton et al., 2017b; Cannon et al., 2013), particularly in regions with high ground resistance, as the geoelectric field is larger and the network provides a more favourable path for GICs to flow. The quasi-DC signal introduced into an AC grid system can lead to half cycle saturation in transformers causing

degradation and, in extreme cases, destruction, failure and system collapse. The geomagnetic field can be used as a proxy for potential ground effects and GIC studies commonly use the time derivative $\frac{d\mathbf{B}}{dt}$ to quantify potential effects.

Nowcasting and advanced forecasting of geomagnetic disturbances is generally achieved through global magnetohydrodynamic (MHD) models (Welling, 2019), driven with near-Earth solar wind observations or, for increased lead time, the output of solar-wind simulations (Merkin et al., 2007). The ground-level magnetic field, which is typically extrapolated from much higher in the magnetospheric domain, is used to drive geoelectric field models. Empirical models also exist (Weimer, 2013, 2019).

An example of a global MHD system is the Space Weather Modeling Framework (SWMF Tóth et al., 2005, 2012). Other widely used MHD models include the Lyon-Fedder-Mobarry (LFM) model (Lyon, Fedder and Mobarry, 2004) and the Open Global Generalized Circulation Model (OpenGGCM) (Raeder, Berchem and Ashour-Abdalla, 1998) (see Welling, 2019). The SWMF consists of several numerical modules, such as the ideal MHD solver BATS-R-US (Block Adaptive Tree Solar-wind Roe-type Upwind Scheme) (Powell et al., 1999; De Zeeuw et al., 2000; Gombosi et al., 2002), the Ridley Ionosphere Model (RIM) (Ridley et al., 2002), and the inner magnetosphere Rice Convection Model (RCM) (Toffoletto et al., 2003).

The operational magnetospheric MHD models underestimate the magnitude of the perturbations across a wide frequency range, including the sub-hourly variations important for GICs (Welling, 2019). Pulkkinen et al. (2013b) examined $\frac{d\mathbf{B}}{dt}$ on a 1-minute timescale and found an underestimation of magnitude between a factor of 2 and 10. Without accurate representation of high-frequency variations of the geomagnetic field, resolution of peak amplitudes in the derived surface geoelectric field and GICs may be underestimated (Grawe et al., 2018).

However, a counter example is Raeder et al. (2001) who used an

MHD model to simulate the Bastille day storm and compared their results to observations. Using a power spectral density (PSD) analysis they found that the model worked well for frequencies of 0-3 mHz and actually overestimated the power at higher frequencies. These results are likely due to using a model configuration with a high grid resolution that would currently be prohibitive for operational forecasting, particularly if large ensembles of magnetospheric runs are required to estimate forecast uncertainty.

Figure 5.1 shows an example of SWMF power spectrum at a broad range of frequencies. The observed and modelled (using SWMF) horizontal magnetic field, the magnetic field component most relevant to GICs, is shown for the December 2006 CCMC test case (`https://ccmc.gsfc.nasa.gov/challenges/dBdt/`) at the Newport magnetometer site. The time series are shown in Figure 5.1 a) and the resulting power spectra in Figure 5.1 b). The coloured lines represent different model configurations. The power spectra shows that each configuration of the model underestimates the power spectral density, however the magnitude of underestimation is highly sensitive to model configuration with 12a_SWMF, the current operational configuration, performing best. These models are giving an output at a 1-minute resolution but the timeseries is smoother than that observed, meaning the amplitude of the higher frequency variations is reduced as shown by the power spectra. These simulation results have been provided by the Community Coordinated Modeling Center at Goddard Space Flight Center for the 2013 Space Weather Workshop and and an online interface is available for analysis of the model runs (`https://ccmc.gsfc.nasa.gov/challenges/dBdt/`).

A general underestimation is in agreement with Pulkkinen et al. (2013b), who show in their Figures 3 and 4 that SWMF underestimated $\frac{d\mathbf{B}}{dt}$. Although we here only show that SWMF exhibits this underestimation, we note that this underestimation is a general feature of operational models predicting geomagnetic perturbations (Pulkkinen et al., 2010, 2011,

2013b).

Recent work from Dimmock et al. (2021) tested different spatial resolution configurations of SWMF for the September 2017 event. They found that the high resolution made a significant improvement to the PSD and GIC forecasts. However, they noted that SWMF performs poorly in substorms and increasing the resolution has limited benefit in these periods. They concluded that a skilful GIC forecast can be done with SWMF but that computational power makes this operationally difficult. In contrast, Haiducek et al. (2017) compared the performance of SWMF on an event in 2005 using the resolution of the operational model and a higher resolution. They used these configurations to estimate geomagnetic indices and cross-polar cap potential (CPCP). They found that results were not sensitive to resolution with the exception of predicting AL which may have been improved. The discrepancy is possibly because Haiducek et al. (2017) did not increase the resolution nearly as much as Dimmock et al. (2021). Mukhopadhyay et al. (2020) also used the configurations of Haiducek et al. (2017) finding that the high-resolution configuration performed generally better under the Heidke skill score for predicting $\frac{d\mathbf{B}}{dt}$.

Several further studies have shown that non-standard MHD model configurations can achieve excellent results for small scale phenomena in a statistical sense. Welling et al. (2021) modelled the magnetospheric response to a hypothetical "perfect" coronal mass ejection and successfully resolved high frequency phenomena. Realistic studies of ULF waves have been made by MHD models (Hartinger et al., 2014; Claudepierre et al., 2009) and small spatial and temporal features have been resolved by a new MHD model (Sorathia et al., 2020). These studies show that MHD models have the capability of properly capturing high frequency ground perturbations relevant to GICs, but the model configurations required are currently computationally prohibitive for operational real-time forecasting.

A viable operational alternative to increasing MHD model grid res-

olution is through the use of a method that statistically relates variability across temporal scales, namely a statistical downscaling approach. In addition to improving the geoelectic field reconstruction from a single magnetospheric model run, downscaling also has the potential to allow uncertainty quantification without the need for a magnetospheric model ensemble.

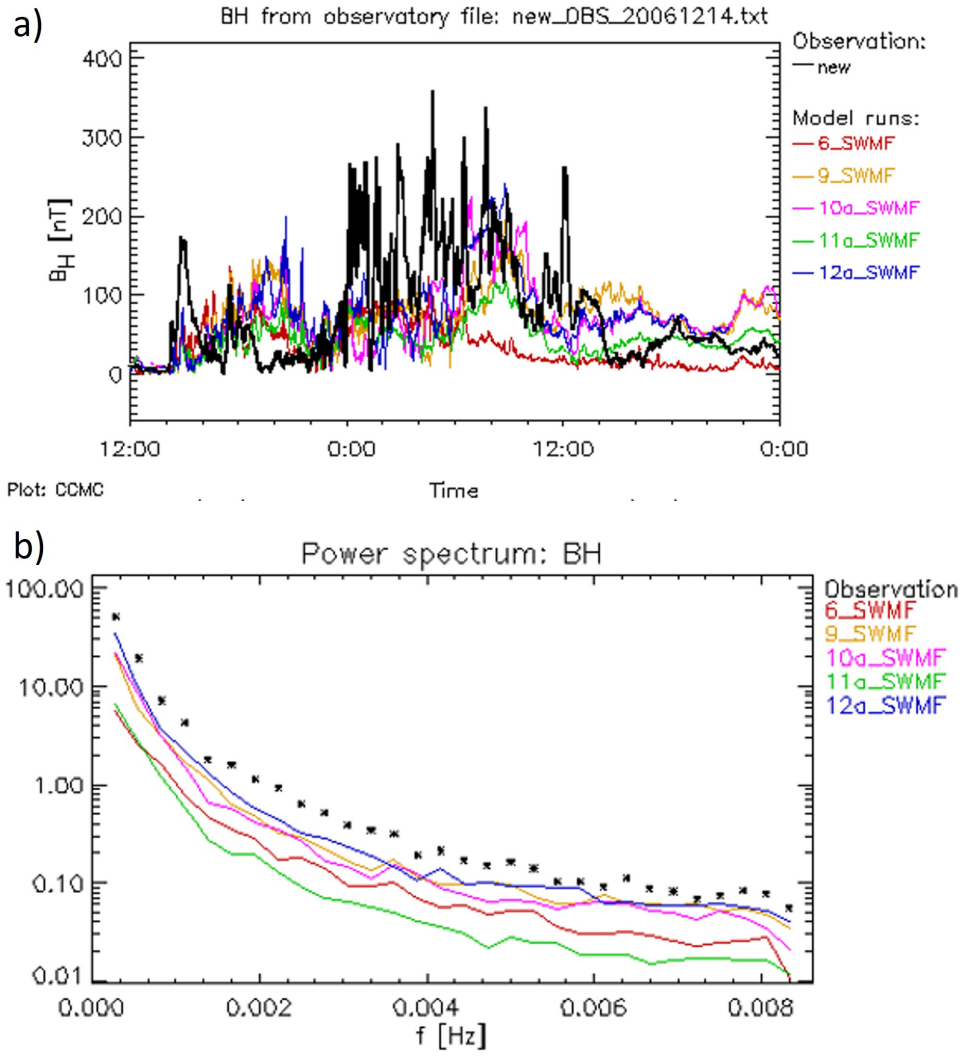


Figure 5.1: Geomagnetic field perturbations at Newport magnetometer station in December 2006. Several configurations of the SWMF (coloured lines) are compared with observations (black). a) shows the time series of the horizontal magnetic field. b) shows the associated power spectra for periods of 2 minutes and less revealing that SWMF underestimated the variability. These plots have been created and downloaded from the Community Coordinated Modelling Centre (CCMC) (`https://ccmc.gsfc.nasa.gov/challenges/dBdt/`.)

This paper addresses the characterisation of high-frequency variability in the magnetic field, **B**, through statistical downscaling. Downscaling has been used in terrestrial weather forecasting to effectively increase the temporal and spatial resolution of global climate models (GCMs)(Maraun et al., 2010; Christensen and Christensen, 2003). For rainfall, this is done because rainfall typically occurs on subgrid scales so cannot be accurately captured with a GCM alone.

Maraun et al. (2010) classifies downscaling into three general categories: perfect prognosis approaches, model output statistics, and weather generators. Perfect prognosis approaches statistically determine relationships between low resolution predictors and the high resolution predictands. This works if the predictors are realistic, such as from a perfect (low resolution) forecast model, i.e. a perfect prognosis. Model output statistics builds a similar statistical relationship but with the aim of also correcting the bias of the forecast model. As such, model output statistics are model-specific. Finally, weather generators generate new high resolution time series that have the same statistical properties as observations, rather than just a probability of a sub grid event. Weather generators can be either perfect prognosis or model output statistics based.

As discussed by Morley (2020), statistical downscaling is relevant to space physics, in particular, to solar wind parameters used as inputs to magnetospheric models. Owens et al. (2014) considered temporal downscaling of solar wind parameters for this purpose. This was done because the magnetospheric models are sensitive to variability at a higher time resolution than is represented in numerical solar wind forecasts. Owens et al. (2014) used a random noise generator that gave high temporal noise with approximately correct statistical properties and added this noise onto the baseline of the solar wind parameters. They found that even relatively simple solar wind downscaling significantly increased the value of the subsequent magnetospheric forecast.

In this work we employ temporal downscaling to increase the variability of magnetic field time series on the ground. By developing a model-independent perfect prognosis scheme, we are assuming that future global MHD models will provide a perfect representation of the low resolution magnetic field variations and/or model biases can be corrected by other means. However, the approach will be applicable to global MHD models that return a skilful and unbiased representation of the low resolution magnetic field. As the high-frequency variations are sampled from an ensemble of observations, an ensemble of geoelectric field estimates can also be reproduced from a single magnetospheric model run.

In the future we hope to apply our downscaling methodology directly to forecasts provided by global MHD models and potentially as a means for uncertainty estimation. However, it is important to develop and test the downscaling scheme in isolation, and not to convolve it with the performance of a specific magnetospheric model. Thus we adopt the widely-used perfect prognostic approach and produce a perfect low-resolution forecast time series by taking 1-hour boxcar means of **B** observed by ground-based magnetometers. This 1-hour series is then linearly interpolated to 1-minute resolution. This represents the undownscaled time series. As will be shown in Section 5.4, this undownscaled series effectively removes all power in variations below 1 hour. Thus it is not a direct proxy for high-resolution magnetospheric model output. However, we start from this 1-hour linearly-interpolated undownscaled series for two reasons. Firstly, we expect magnetospheric models to perform better than this but it can be thought of as a 'worst-case scenario' for low-resolution magnetospheric models such as might be used for real-time forecasting in large ensembles. Secondly, if the downscaling manages to successfully relate the variability at 1-hour resolution to that at 1-minute resolution, it should be more than adequate for use with magnetospheric model output.

The downscaling scheme attempts to reintroduce high-frequency per-

turbations onto the linearly-interpolated 1-hour time series to produce a more realistic (in a statistical sense) **B** time series at the 1-minute resolution. By using observations as the undownscaled time series, rather than model output, we removing model error from the process of developing and testing our methodology. Additionally, this approach allows us to easily create a large database of low-resolution, undownscaled "forecasts" with which to test our model, without requiring decades of magnetospheric model output.

5.2 Data

The ground-based magnetometer measurements we use are provided by SuperMAG (Gjerloev, 2012) (`http://supermag.jhuapl.edu`), an international collaboration bringing together data from over 300 magnetometer stations. The SuperMAG ground-level magnetic field perturbation data has been homogenised in terms of coordinate system, processing technique and file structure.

A ground-based magnetometer measures the magnetic field from all sources in its vicinity. For studies on magnetic perturbations due to ionospheric and magnetospheric current systems, the magnetic baseline needs to be subtracted from the measurements to remove effects from other magnetic sources such as the Earth's intrinsic magnetic field. Gjerloev (2012) describes the SuperMAG data-processing technique for removing the base line, in which knowledge of typical timescales of variations of different magnetic fields is used. These amount to a yearly trend, mainly due to the secular variation in the Earth's main field, and a diurnal trend due to the *Sq* current system, the quiet day daily variation in ionospheric activity due to solar radiation. These are subtracted from the magnetometer measurements, leaving the prime source of variability as space-weather driven activity.

Of course, magnetometer measurements can occasionally have erroneous measurements. These usually take the form of a spike in activity for a single data point during an otherwise quiet period. These errors can sometimes get past the SuperMAG quality control and into the final datasets. The data used for this analysis is a SuperMAG dataset that has been cleaned for occasions where an error has exceeded the $99.97th$ percentile in terms of the change in the magnetic field with time as described in Rogers et al. (2020). The data may still have errors at lower levels of activity.

In this study we primarily use data from the Eskdalemuir (ESK) station located in southern Scotland with geographic coordinates of 55.314°N and 356.794°E, operated by British Geological Survey. In principle, temporal downscaling techniques are applicable to all locations but we first test this one location where we have access to an established model for converting local magnetic field variations to geoelectric field variations, acknowledging the local ground conductivity conditions (Beggan et al., 2021). From the ESK station we have 1-minute **B** measurements for approximately 30 years, from 1983-2016.

5.3 Methodology

5.3.1 Analogue Ensemble

The Analogue Ensemble (AnEn) approach was originally used for terrestrial weather forecasting (e.g. van den Dool, 1989; Delle Monache et al., 2013), but has been far surpassed by physics-based models for that application. However, AnEn has more recently been employed in space and magnetospheric physics where the physical models are less accurate, largely due to the limited availability of observations to completely characterise the necessary boundary conditions. In such a situation, empirical schemes

can be valuable. Haines et al. (2021b), Owens, Riley and Horbury (2017d), Riley et al. (2017) and Barnard et al. (2011) have experimented with AnEn for forecasting the solar wind, geomagnetic activity and changes in space climate. In each case AnEn outperformed the benchmarks considered.

The AnEn methodology exploits an extensive historical dataset for forecasting purposes through analogy to past evolution of a given system. Specifically, an AnEn examines the present state of the predictors, looks in the historical dataset for analogous periods, then takes the predictand from the most analogous period. By selecting multiple analogous periods, an ensemble of predictands can be created, enabling a probabilistic forecast of future evolution.

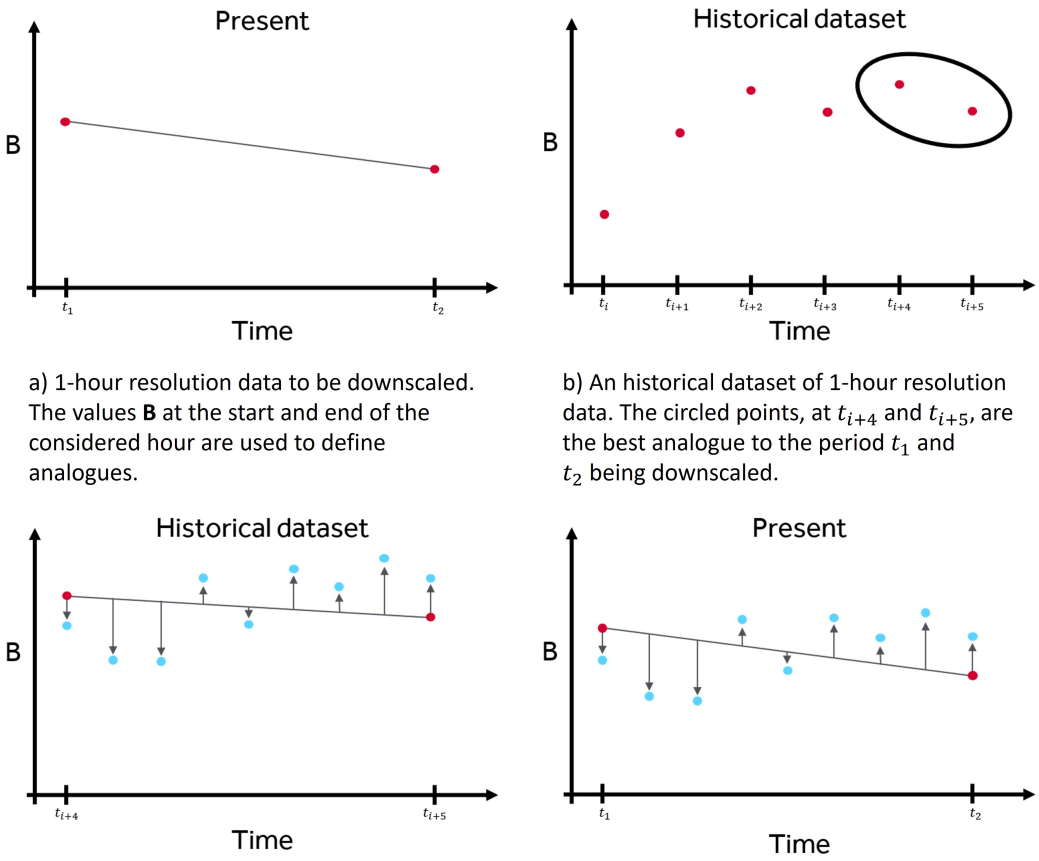
In this work, AnEn is used not for forecasting, but for temporal downscaling to relate variations on long and short timescales. To demonstrate that the downscaling framework works for ground-level **B**, we chose 1-hour and 1-minute for the long and short timescales somewhat arbitrarily, as described in the previous section. They are intended as examples rather than fixed parameters. At the high frequency, 1-minute makes sense as that is the typically available resolution of long-term ground-based **B** series and also the input resolution for many geoelectric field models. At the low frequency, the time scale of interest will depend on the specific model and the situation in which the model is being used. e.g., where real-time forecasting is required and/or ensembles of magnetospheric models are being used, it may be necessary to reduce the model resolution. As said, the low-resolution timescale of 1-hour is a tuneable parameter. If the downscaling is able to successfully relate 1-hour and 1-minute variations, it should perform even better at relating, e.g., 20-minute and 1-minute variations. Due to the perfect prognostic approach we can use the low-resolution time series as predictors. Specifically, the predictors used are the low-resolution values of the horizontal magnetic field at the start and the end of the considered hour. Analogous periods of these are found and

used to predict a 1-minute resolution time series.

The AnEn algorithm is outlined in Figure 5.2 and described in the following points, in which the subscript H stands for 1-hour and M for 1-minute values:

1. Split the 1-minute SuperMAG data into two sets ($D1_M$, $D2_M$). $D1_M$ is the test dataset containing the short period to be downscaled. $D2_M$ is the independent training dataset comprised of the remaining data.
2. Compute low-resolution data using 1-hour box-car means, to give $D1_H$ and $D2_H$.
3. Using $D1_H$, take the values at the start (t_1) and end (t_2) of the hour being considered, as shown in Figure 5.2 a).
4. Search $D2_H$ for the N most similar consecutive values, by mean squared error, to those at t_1 and t_2, as in Figure 5.2 b), where N is the chosen number of analogues.
5. Remove the baseline value from the associated $D2_M$ leaving only the higher frequency structure of the analogue interval, i.e. minute-scale variations with the baseline removed, as in Figure 5.2 c). The baseline is defined as the 60-minute rolling mean.
6. Add each $D2_M$ analogue onto $D1_H$ to produce an ensemble of downscaled values as in Figure 5.2 d).
7. Repeat this process for each hour in $D1_H$.

The data is then repeatedly split into different test and training sets so that the whole 34-year period can be downscaled using an independent training set. Note that this procedure uses data from after the 'forecast' time, so is not strictly a hindcast. However, this approach uses the volume of available historical data available to a forecast made today and thus quantifies the current expected performance of downscaling.



a) 1-hour resolution data to be downscaled. The values **B** at the start and end of the considered hour are used to define analogues.

b) An historical dataset of 1-hour resolution data. The circled points, at t_{i+4} and t_{i+5}, are the best analogue to the period t_1 and t_2 being downscaled.

c) The high resolution, 1-minute, data in the analogous period, between t_{i+4} and t_{i+5}, is used to compute the analogue high-frequency variations by subtracting the baseline.

d) The high resolution, 1-minute, structure from the analogous period is added onto the baseline of the period under consideration, which is taken as the 60-minute rolling mean. The blue dots give the new downscaled time series.

Figure 5.2: A schematic of the AnEn process. This process is repeated with the N best analogous periods to give an ensemble of downscaled time series.

5.3.2 Reference model

We use a reference model, as suggested by Liemohn et al. (2018), as a benchmark of comparison for the AnEn's performance. As this is a proof of concept study, we choose a reference model that represents a "do-nothing" approach to downscaling. For this we downscale the 1-hour time series of the magnetic field using a linear-interpolation, denoted as the linear-interpolation approach. Through this, we end up with 1-minute resolution time series without adding further high resolution structure.

As stated in Section 5.2, this 1-hour linear-interpolation series is not

representative of ground-level **B** produced by typical state-of-the-art magnetospheric models, as can be seen from the power spectra in Figures 5.1 and 5.4. Instead, 1-hour can be seen more as a worst-case scenario – most magnetospheric models would be expected to reasonably reproduce the **B**-field fluctuation power at around 0.00028 Hz, even in real-time ensembles.

5.3.3 MT-transfer function

The goal of this work is not to recreate the high resolution magnetic field on a point-by-point basis, but to add in realistic high-frequency variability in a statistical sense. In particular, we are interested in the higher frequency structure insofar as it improves the subsequent estimate of the induced geoelectric field, which is the driver of GICs.

This can be tested with an "impacts" model. For this purpose we use a magnetotelluric- (MT-) transfer function (Simpson and Bahr, 2020; Beggan et al., 2021) produced for the ESK site by the British Geological Survey (BGS). The MT-transfer function converts a time series of the local magnetic field into a time series of local geoelectric field. The MT-transfer function first makes a Fourier transform of the magnetic field, then multiplies the result by an empirically determined matrix of coefficients which account for the local ground conductivity, and finally makes an inverse Fourier transform to compute the geoelectric field in the time domain. The matrix of coefficients is derived from simultaneous observations of the magnetic and geoelectric fields at ESK.

To quantify the performance of the downscaling scheme, we focus on the magnitude of the estimated **E**-field. Each **B**-field ensemble member was individually transformed with the MT-transfer function to result in an associated **E**-field ensemble member. A 'good' outcome would be that the $|\mathbf{E}|$ from the downscaled series is closer to the $|\mathbf{E}|$ obtained from using the observed series, than the linear-interpolation approach. An ideal out-

come would be that the observed $|\mathbf{E}|$ output falls within the spread of the ensemble of $|\mathbf{E}|$ outputs obtained with the ensemble of downscaled series.

5.4 Evaluation

The AnEn downscaling approach has been applied to the entire 34-year period (1983-2014) of observations using an ensemble of 100 members built hour by hour as described above. Figure 5.3 shows an example spanning six-hours of heightened activity, with the x-component (east-west) in 5.3 a) and the y-component (north-south) in 5.3 b). This period was a geomagnetic storm with a minimum *Dst* of -172 nT. The observed time series is shown in red, the linear-interpolation series is shown in blue, and the median of the AnEn series is shown in black, with colour bands showing the $0th-100th$, $10th-90th$ and $25th-75th$ percentiles. The linear-interpolation approach is shown as a benchmark for the AnEn series to be compared against.

For the interval shown in Figure 5.3, the $10th-90th$ percentile band captures some of the variability seen in the observations, however, it seriously underestimates the variability on several occasions. Notably, towards the middle of the period, when the event is at the peak, the ensemble spread captures less of the variability. This suggests that the AnEn will struggle with the larger events such as this. By the definition of confidence, we would expect the observation to sit within the $0th-100th$ percentile band 100%of the time, in the $10th-90th$ percentile band 80% of the time and in the $25th-75th$ percentile band 50% of the time. In actuality here, the percentage of observations in the $0th-100th$, $10th-90th$ and $25th-75th$ percentile bands for B_x are 83.4%, 40.3% and 20.3%, respectively. For B_y this is 98.9, 51.8% and 21.3%, respectively for this illustrative period.

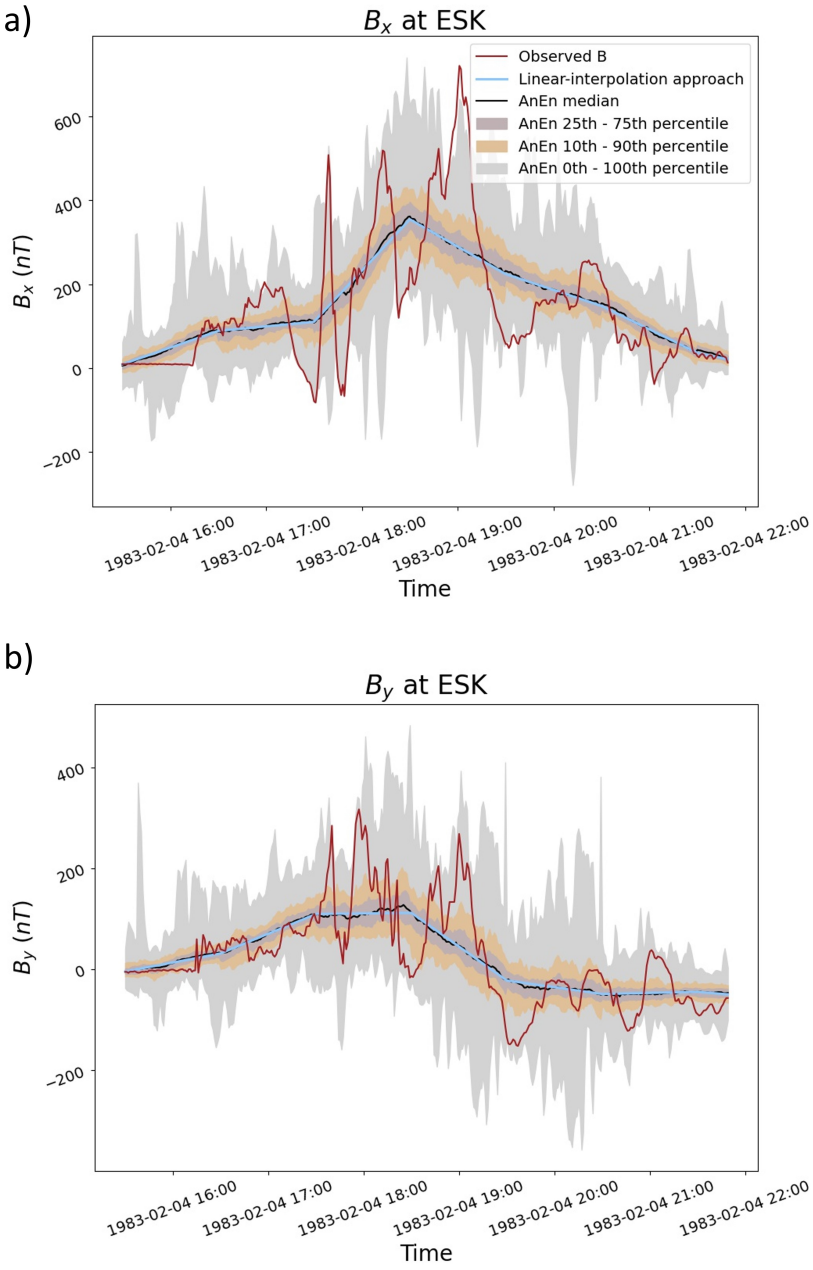


Figure 5.3: A six-hour time series from 1983-02-04 of the magnetic field at ESK in the x (east-west) and y (north-south) directions in the geographic coordinate system. The red line shows the observed 1-minute time series, the colour bands show the spread of the AnEn series (the 10th-90th and 25th-75th percentiles) with the median in black, and the blue line shows the linear-interpolation approach, taken to be the undownscaled magnetic field, as a reference.

Figure 5.4 shows the power spectra of the magnetic field from observations and AnEn. Shown is the median and percentile bands of the PSD's achieved by all the ensemble members computed with Welch's

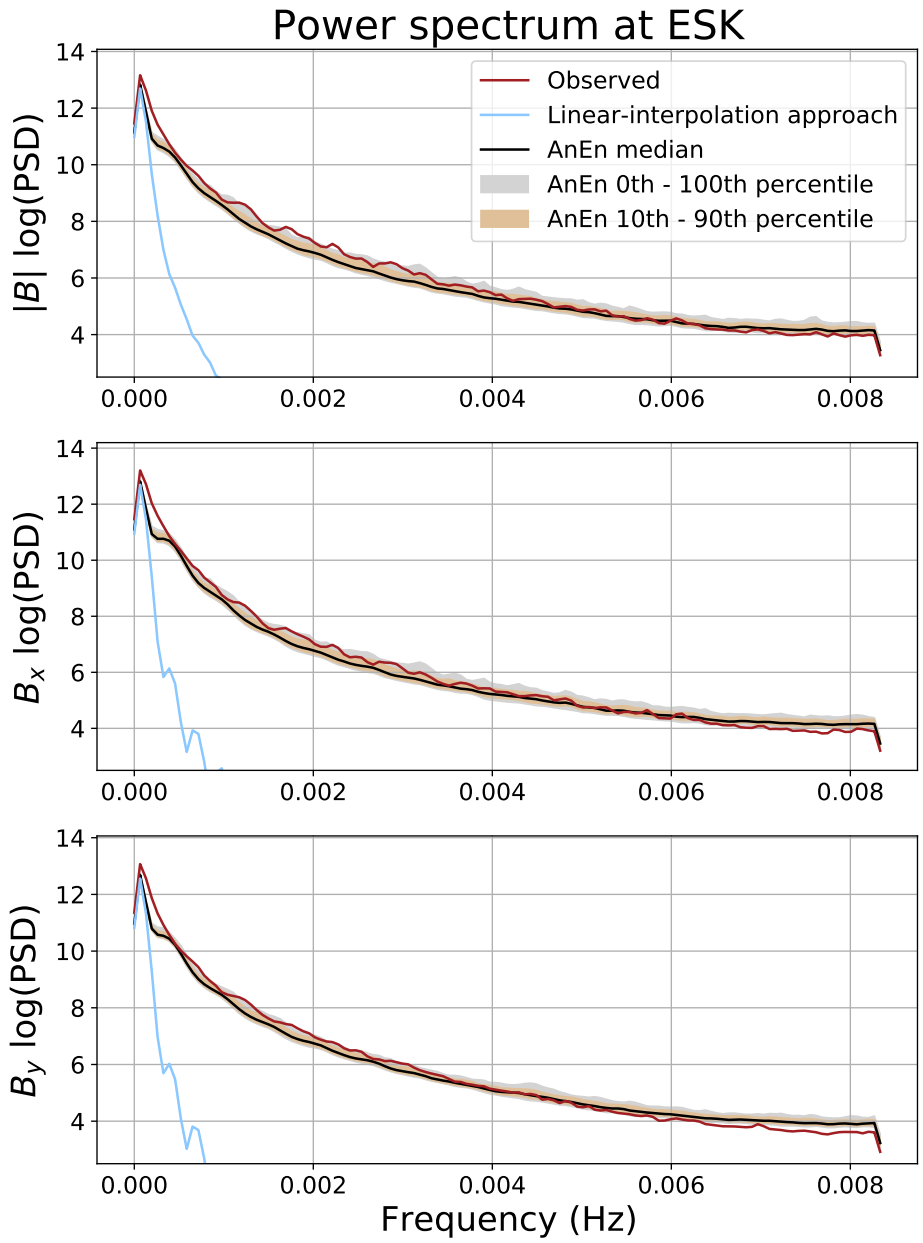


Figure 5.4: The power spectrum of the magnitude, B_x and B_y components of the magnetic field from the whole 34-year period from observations and AnEn. The yellow colour band shows the 10-90% range of the AnEn. The linear interpolation approach is shown in blue, part of which has been cut from the plot due to large differences in scale.

method using the Hanning window without overlap. The AnEn ensemble follows the observations closely with a general trend to slightly underestimate the power at lower frequencies (0-0.003) and slightly overestimate the power for higher frequencies (0.007 and above). The 10-90% range of the AnEn is very narrow at approximately 0.5 at the most, reflecting a consistent performance across the whole ensemble. The linear-interpolation approach is shown in blue but has been cut off because, as expected, the power spectral density is very low and hence makes scaling the y-axis difficult. It is clear that AnEn provides a power spectrum much more similar to that of the observations than the linear-interpolation approach achieves.

To measure the effectiveness of adding higher frequency structure we use the **B** time series magnetic fields from the observations, AnEn and the linear-interpolation approach to drive the MT-transfer function as described in Section 5.3.3. The output of the MT-transfer model is shown in Figure 5.5 for the same six-hour period shown in Figure 5.3. We see that the AnEn captures some of the geoelectric field variability within its spread but the observations lie outside the range of the analogue ensemble on many occasions. The percentage of observations in the $0th-100th$, $10th-90th$ and $25th-75th$ percentile bands for E_x are 97.4%, 59.5% and 31.3%, respectively. For E_y this is 97.4%, 51.8% and 27.4%, respectively for this illustrative period.

Figure 5.5 reveals that, as expected, the linear-interpolation series yields very low geoelectric fields, without any significant variation. With a large ensemble size, the AnEn median will tend toward a smooth line despite variations in individual ensemble members. Therefore, the usefulness of AnEn is not in its median but rather in the spread of its ensemble members for showing possible realisations of the timeseries. Because of this it is not useful to directly compare AnEn median to the linear-interpolation approach values. However we do see that the spread on the analogue ensemble is of a more similar magnitude to that in observations

than the linear-interpolation approach time series. In addition, AnEn provides an idea of the uncertainty in a forecast which is useful for making decisions.

While this example period is illustrative, it is necessary to evaluate AnEn as a downscaling model over the full 34-year period using a set of metrics. In the following evaluation we have taken care to chose metrics that are robust to timing errors, as we make the assumption that the spectral properties of fluctuations and the magnitude of the peaks are generally more important than the phasing for GIC impacts. This is also relevant since operations require a lead time of possible occurrence and an estimate of the severity of that occurrence as they cannot implement system wide mitigation in real-time. When comparing data on a point-by-point basis, timing errors, in which a defined event is correctly predicted to occur but at slightly the wrong time, will incur a double penalty by many common metrics (e.g. see Figure 8 of Owens, 2018). For example, accuracy, which gives a fraction of correct predictions across the whole dataset, will count the forecast as wrong when it predicts an event that doesn't occur at the exact time step and wrong when the forecast does not predict an event that is observed, even if the time step is off by just one step.

The sensitive values of GIC magnitude and timescales are dependent on the set up of individual transformers and the power grid configuration. For example, the size of geoelectric field that will cause a significant GIC is dependent on the ground conductivity in the region around the transformer. We use horizontal geoelectric field as a practical solution to provide a general evaluation of the method (Beamish et al., 2002), however transformers are sensitive to the individual E_x and E_y parameters, depending on grid configuration (Orr, Chapman and Beggan, 2021).

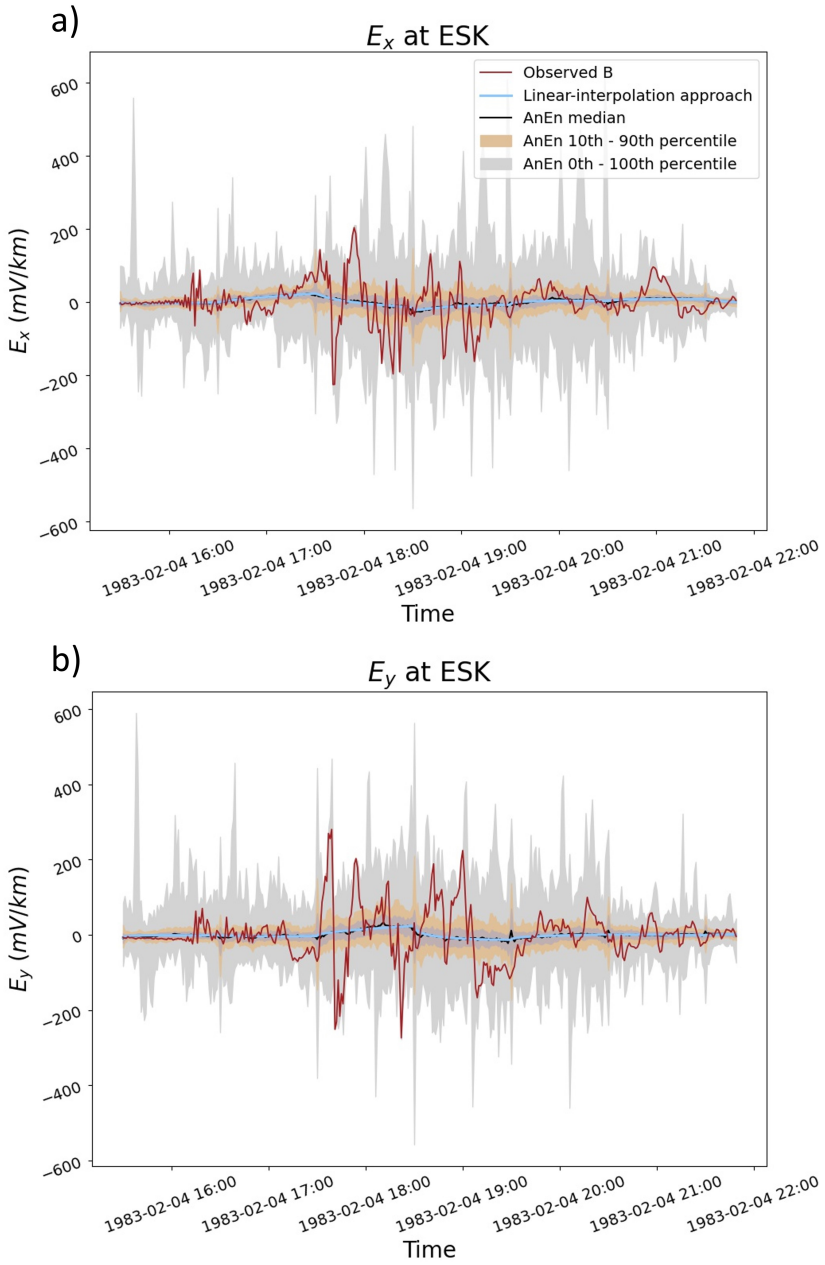


Figure 5.5: A six-hour time series from 1983-02-04 at ESK of the geoelectric field computed from the magnetic field using the MT-transfer function. The data is in the x (east-west) and y (north-south) directions in the geographic coordinate system. The red line shows the time series computed from the 1-minute observed time series, the colour bands show the spread of the geoelectric field computed from the analogue ensemble with the median in black, and the blue line shows geoelectric field computed from the linear-interpolation approach magnetic field.

5.4.1 Threshold-exceedance prediction

In this subsection we evaluate each individual ensemble member within AnEn for its ability to give a binary prediction of an event at individual time steps. We examine three levels of activity for event classification using the magnitude of total horizontal geoelectric field, denoted $|\mathbf{E}|$, from the MT-transfer function. The magnitude of the total horizontal geoelectric field is shown for an illustrative period in Figure 5.6. The chosen thresholds for evaluation are the $99th$, $99.9th$ and $99.99th$ percentiles of the magnitude of the total horizontal geoelectric field from the MT-transfer function driven by observed magnetic field time series over the period 1983 to 2016. These are 22.3, 58.8 and 171.9 mV/km respectively and shown in Figure 5.6 by the horizontal dashed lines. For context, during the March 1989 storm that lead to the Hydro-Quebec network collapse, the peak geoelectric field magnitude for the March 1989 storm at ESK was 411.4 mV/km as computed using the MT-transfer function. It is worth noting that the system collapse experience during this geomagnetic storm occurred before the peak due to the rapid onset of a substorm (Boteler, 2019).

In order to allow for timing errors at the minute scale, we evaluate AnEn using the fraction skill score (FSS) (Roberts and Lean, 2008; Owens, 2018). The FSS is most commonly used to measure the fractional occurrence of events in a given spatial window. Here, we use FSS with a 60-minute temporal window and count the fraction of predicted time points which are classified as events, and the fraction of observed time points which are events, within the same time window. This is repeated for each ensemble member for time windows covering the whole dataset and the mean squared error (MSE) between the observed and predicted fraction time series is computed. This is repeated for a reference forecast, in this case the linear-interpolation series, and the FSS is taken as $1-(MSE_{forecast}/MSE_{reference})$. A perfect forecast would achieve $FSS=1$,

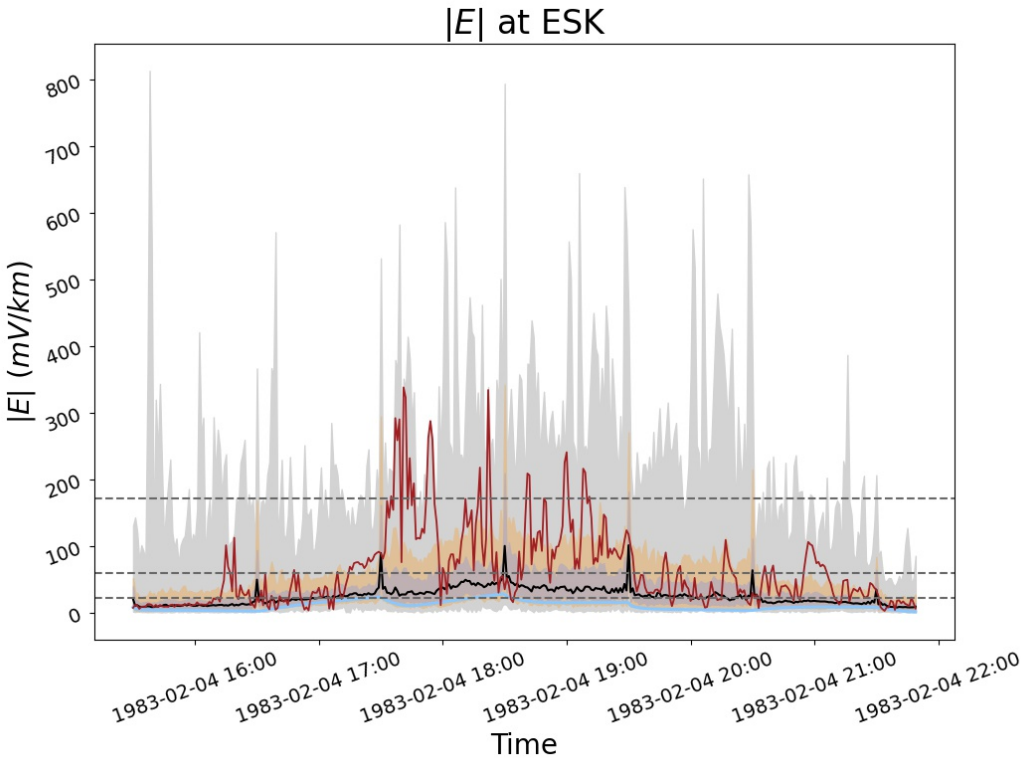


Figure 5.6: A six-hour time series from 1983-02-04 at ESK of the total magnitude of the geoelectric field computed from the magnetic field using the MT-transfer function. The red line shows the magnitude of time series computed from the 1-minute observed time series, the colour bands show the spread of the magnitude of geoelectric field computed from the analogue ensemble with the median in black, and the blue line shows the magnitude of geoelectric field computed from the linear-interpolation approach magnetic field.

a forecast with no skill compared to the reference would achieve $FSS = 0$ and a forecast performing worse than the reference will achieve a negative score. FSS is most useful to end users who need to know if an event will occur within a given time window without the need for exact (in this case, to the minute) knowledge of when it will occur.

Figure 5.7 shows the FSS achieved for each of the 100 ensemble members across the entire dataset for each of the three event thresholds. Ensemble ID is ordered from best to worst analogues considered, where best means the 1-hour values in the analogous periods are most similar to present conditions by RMSE. For the $99th$ percentile threshold (panel a) we see that each ensemble member has a positive FSS, with an av-

erage value across the whole ensemble of 0.095, showing it outperforms the reference method. When considering events over the $99.9th$ percentile, Figure 5.7 b) again shows all ensemble members having a positive FSS with an average across the ensemble of 0.17. We also see a clear trend in which ensemble members based upon better analogues produce better FSS scores. The increased visibility of the trend for the $99.9th$ percentile compared to the $99th$ percentile suggests that at higher thresholds we are inherently considering rarer events, which reduces the number of good analogues available.

For events over the $99.99th$ percentile (panel c) the FSS is mainly positive for the first 50 ensemble members and approximately zero for the second 50. The mean FSS for the whole ensemble is 0.067. There is a very stark decrease in the skill of the ensemble members as the ensemble ID increases suggesting that for such a high threshold there are only around 30 to 50 good analogues for AnEn to work with. This finding can help inform a decision on an appropriate ensemble size for deployment. It also suggests that it would be appropriate to weight ensemble members if they are to be combined in any way.

5.4.2 1-hour mean value prediction

The impact of GICs on transformers can be dependent on time-integrated effects, meaning that problems occur when GICs exceed a certain threshold for a certain duration (Moodley and Gaunt, 2017). With this in mind, we now evaluate the model using events classified using thresholds of the 1-hour mean value of $|\mathbf{E}|$ previously used. The hourly mean of the magnitude of geoelectric field for an illustrative period is shown in Figure 5.8. We again consider thresholds at the $99th$, $99.9th$ and $99.99th$ percentiles of the 1-hour means of the horizontal geoelectric field magnitude from the observed time series. These values are 17.9, 47.0 and 139.0 mV/km respectively. These are shown on Figure 5.8 by the horizontal

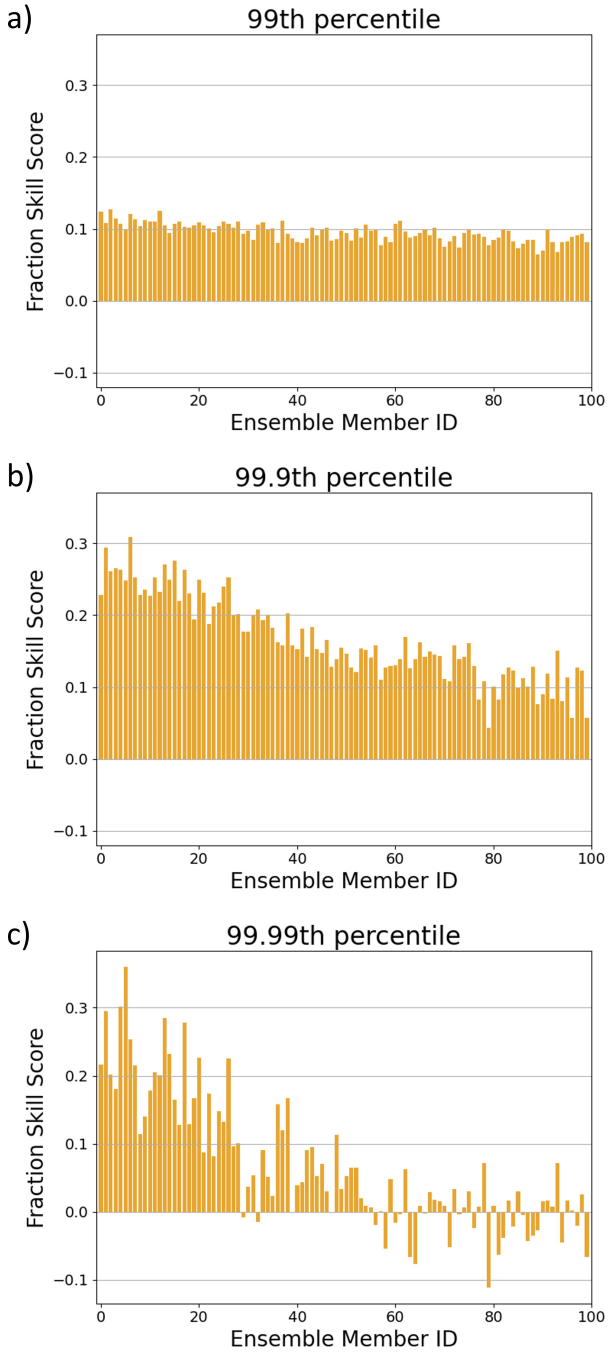


Figure 5.7: The fraction skill score (FSS) for each ensemble member. Ensemble members are ordered from best to worst analogues considered. A FSS of 1 represents a perfect model FSS of 0 represents a model with no skill over the reference. The time window for computing FSS is 60-minutes. a), b) and c) show FSS for events over the $99th$, $99.9th$ and $99.99th$ percentiles of the geoelectric field.

dashed lines. For context, the peak hourly mean observed at ESK during the March 1989 storm was 77.1 mV/km, suggesting that although the

peaks of this storm were large, they were short lived. These metrics are useful as impacts of a heightened geoelectric field are often caused by sustained heightened values on approximately the tens of minutes to 1-hour time scale (Pulkkinen et al., 2017). The metrics in this section are useful to end users who need to know when periods of heightened activity will occur and users who are impacted by time-integrated effects.

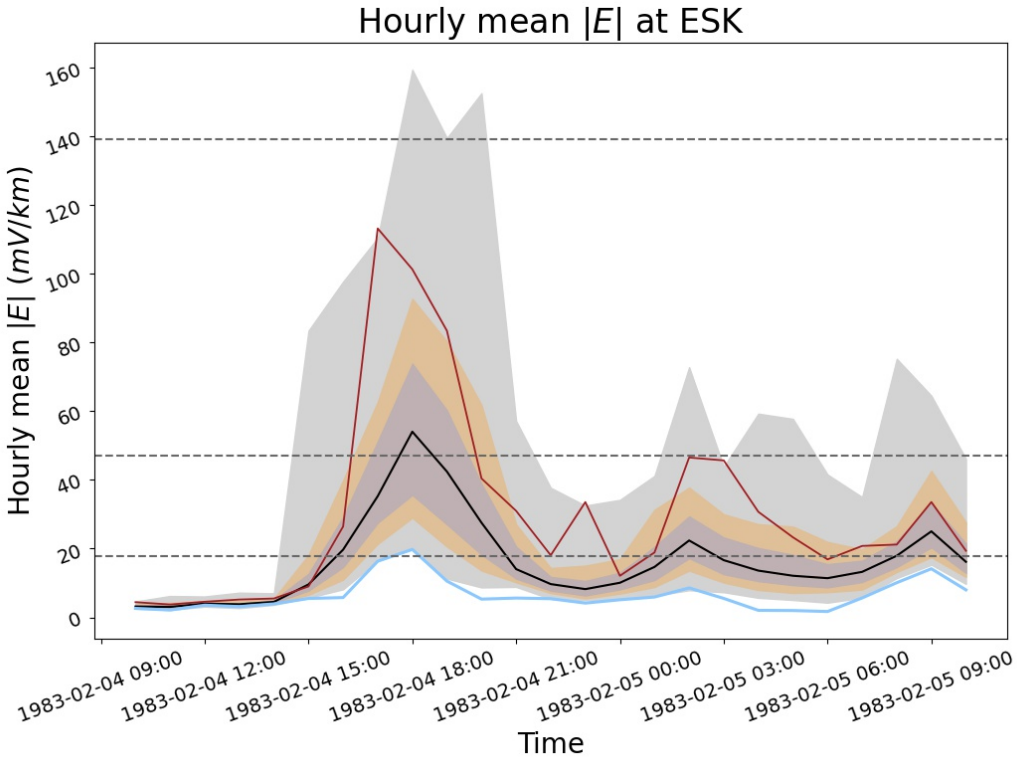


Figure 5.8: A time series from 1983-02-04 to 1982-02-05 at ESK of the 1-hour box-car mean of the magnitude of the geoelectric field computed from the magnetic field using the MT-transfer function. The red line shows the 1-hour mean of the magnitude of time series computed from the 1-minute observed time series, the colour bands show the spread of 1-hour mean magnitude of electric field computed from the analogue ensemble with the median in black, and the blue line shows the 1-hour mean of the magnitude of geoelectric field computed from the linear-interpolation approach magnetic field.

Deterministic prediction

The first metric chosen is the Heidke skill score (HSS) (Jolliffe and Stephenson, 2003). HSS measures the accuracy of a model, taking into account

the number of correct random forecasts. This allows for proper measurement of skill in a situation where an event is rare. In fact, the rarer the event considered, the less HSS takes into account correct predictions of "no event", which becomes the overwhelming majority class. HSS uses the four categories on a standard contingency table: the number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) events. HSS is then given by:

$$HSS = \frac{TP + TN - crf}{TP + TN + FP + FN - crf}, \tag{5.1}$$

where crf, the number of correct random forecasts, is

$$crf = \frac{(TP + FP)(TP + FN) + (FP + TN)(FN + TN)}{n}, \tag{5.2}$$

where n is the total number of predictions.

HSS of AnEn is shown in Figure 5.9 for the three event thresholds considered. HSS is has been computed for each ensemble member shown by the yellow bars and HSS for the linear-interpolation approach is shown by the black dashed horizontal line. AnEn clearly outperforms the linear-interpolation approach and it generally achieves a good positive score with the exception of some of the ensemble members based on weaker analogues for the $99.99th$ percentile threshold. This again suggests that the available dataset is too small for 100 analogues of more extreme events.

Probabilistic prediction

Next we evaluate AnEn in its ability to give a probabilistic prediction of an event by counting how many of the ensemble members predict an event and normalising by the size of the ensemble. This is evaluated using the Cost/Loss analysis (Murphy, 1977b; Richardson, 2000; Owens, Riley and Horbury, 2017d), which allows different end users of a forecast

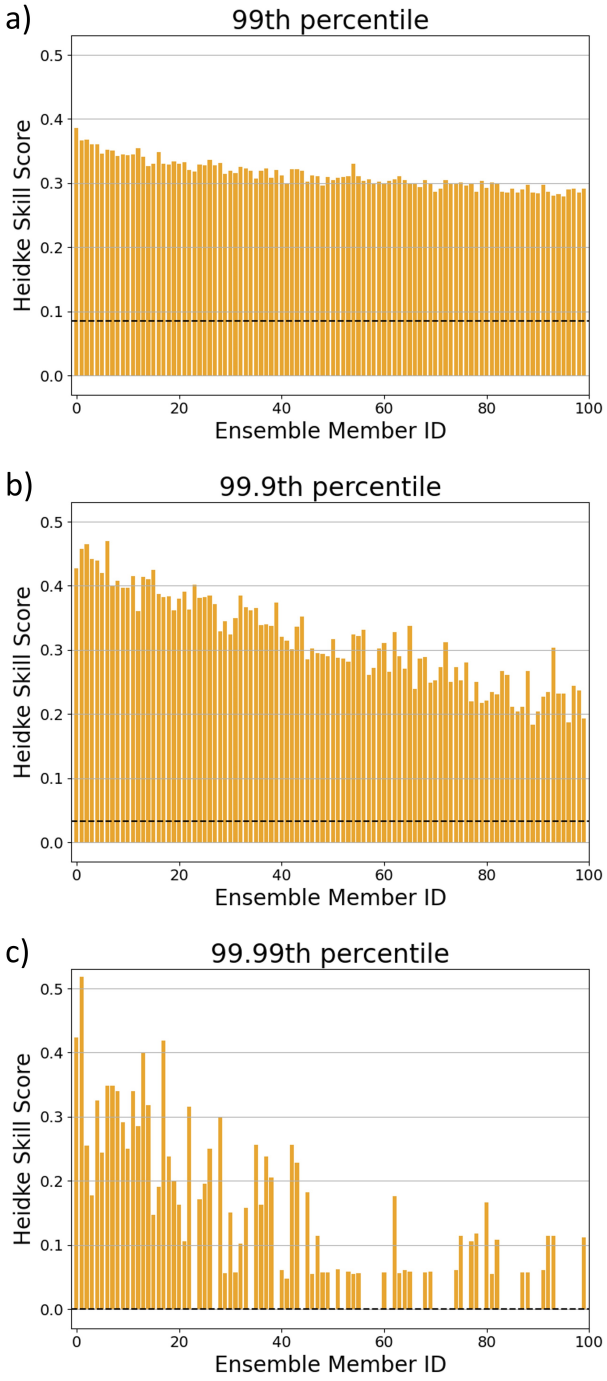


Figure 5.9: The Heidke skill score (HSS) for the three event thresholds on applied to 1-hourly |E| data. Ensemble members are ordered from best to worst analogues considered. A perfect forecast has a score of 1, a forecast with no skill over random prediction has a score of 0, and a forecast with every prediction incorrect has a score of -1. HSS is shown for each ensemble member. The black dashed horizontal line represents the HSS achieved by the linear-interpolation approach for each event threshold

to assess its value for their particular use case. The idea is that taking mitigating action due to a forecast incurs a Cost, C, of fixed value, and experiencing an event without taking mitigating action incurs a Loss, L, of fixed value. The Cost/Loss analysis sums these Costs and Losses for acting on a particular forecast across a long time series and compares the sum to that of a perfect forecast and a climatological forecast method (which, at all times, predicts the probability of an event as the fraction of time in which that event is experienced across the whole dataset). The result is the potential economic value (PEV) which is 1 for a perfect forecast, 0 for a forecast of equal ability to climatology, and negative for a forecast with worse ability than the climatology. PEV is given as a function of the Cost/Loss ratio, C/L, which is between 0 and 1 for all end users that may find a forecast valuable. In a probabilistic Cost/Loss analysis that we employ here, mitigating action is taken if the probability given by AnEn exceeds the Cost/Loss ratio of the end user. For more details see Murphy (1977b); Richardson (2000).

Figure 5.10 shows the PEV for the Cost/Loss domain $(0,1)$ for the probabilistic downscaling from the AnEn and the linear-interpolation approach. We see that for all three event thresholds AnEn outperforms the reference method. We also see that the PEV is highest for the lower end of the Cost/Loss domain which means it will most benefit end users who better tolerate false alarms (false positives) rather than missed events (false negatives). This is because at the low end of the C/L domain the cost of taking mitigating action is very low compared to the loss incurred due to not taking action and an event happening. Therefore, these users would generally prefer to take mitigating action on a false alarm than not take action on a real event.

Finally we look at how AnEn performs under the Brier skill score (BSS) (Jolliffe and Stephenson, 2003). Like Cost/Loss analysis, BSS can compare probabilistic forecasts with deterministic ones, allowing direct com-

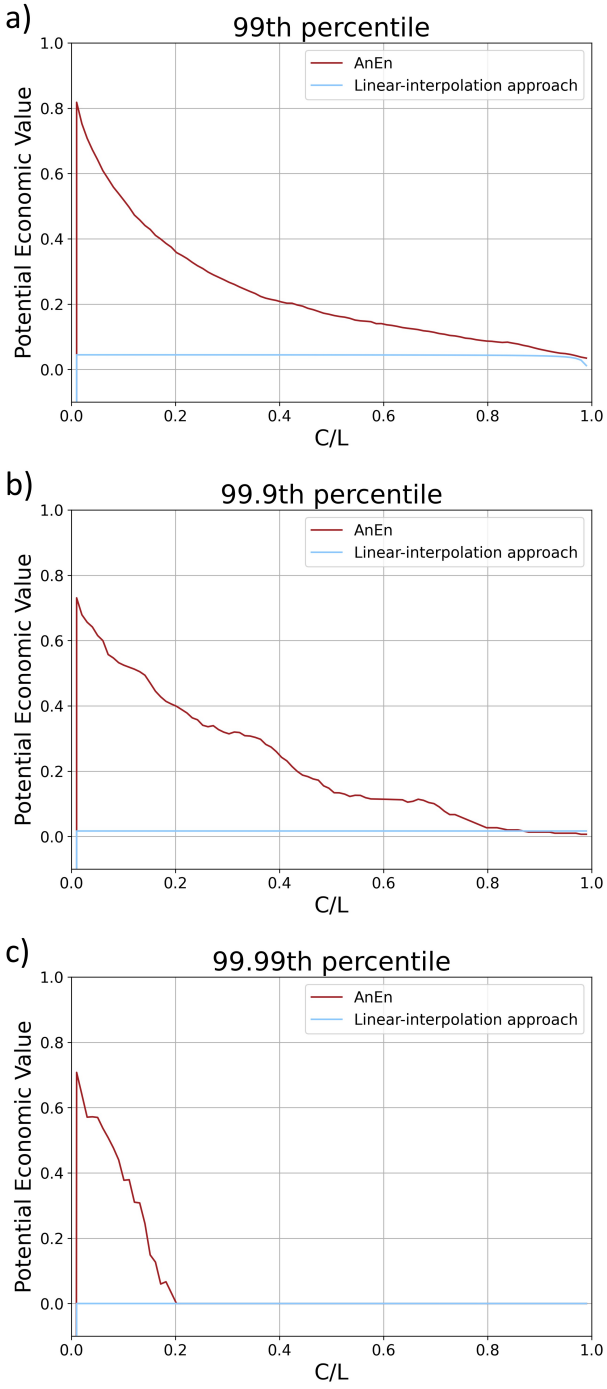


Figure 5.10: A Cost/Loss analysis showing the potential economic value (PEV) of the probabilistic AnEn downscaling method with respect to the undownscaled (linear-interpolation) reference method. A score of $PEV = 1$ represents a perfect forecast and $PEV = 0$ represents no value with respect to the reference method. a), b) and c) show PEV for events over the $99th$, $99.9th$ and $99.99th$ percentiles of the geoelectric field.

parison of the probabilistic AnEn and the deterministic undownscaled series. BSS is useful to end users who wish to use the probabilistic information of AnEn. To compute BSS, the standard Brier score (BS) must first be computed. The BS is the normalised sum of the square error between the probabilistic forecast and the observations over the whole time series, where the observations takes a binary value of 0 or 1 depending on whether an event occurs. Events are again taken to be hours exceeding the $99th$, $99.9th$ and $99.99th$ percentiles of observed $|\mathbf{E}|$. BS is computed for both AnEn and the reference model then combined into BSS by

$$BSS = 1 - \frac{BS_{forecast}}{BS_{reference}}. \tag{5.3}$$

Similarly to the Cost/Loss and FSS, a perfectly skilful forecast receives BSS=1, a forecast with no skill relative to the reference receives BSS=0, and a negative score signals the forecast method performs worse than the reference.

BSS is shown for AnEn for the three event thresholds in Table 5.1. It seems that the 100-member AnEn has skill over the linear-interpolation approach for all considered thresholds but drops in skill for the $99.99th$ percentile events. It is likely that this is the result of the limited span of the dataset and hence number of analogous extreme events. A reduced ensemble size or ensemble-member weighting would likely yield a better BSS, particularly for the $99.99th$ percentile events. This is shown in the third column of the table which gives BSS for a 20 member ensemble. We see that the BSS of the $99.99th$ percentile events increases more in line with the lower thresholds.

Threshold (percentile)	BSS (100 members)	BSS (20 members)
99th	0.30	0.32
99.9th	0.32	0.38
99.99th	0.15	0.31

Table 5.1: Brier skill score (BSS) for AnEn using the linear-interpolation approach as the reference. Three event thresholds are considered.

5.5 Discussion & Conclusions

Statistical downscaling of magnetic field data for the purposes of GIC forecasting has been demonstrated in the form of a perfect prognostic approach. We employed the analogue ensemble (AnEn) methodology, finding that with its spread and higher frequency contributions, a more accurate E-field mapping is obtained than when compared to an E-field derived from undownscaled **B**-field data.

To obtain a "low-resolution" dataset, ground-level magnetic field perturbation data was smoothed from high frequency (1-minute) to low frequency (1-hour) resolution. High frequency structure was then reintroduced into the low-resolution (1-hour) series using the AnEn approach. Both the low frequency and the downscaled time series were then used in a magnetotelluric-transfer function to compute the corresponding horizontal geoelectric fields.

We presented the power spectrum of the observations, showing that AnEn closely resembles the spectral properties of the observations and far outperforms the linear-interpolation approach. Although AnEn has not been applied to the output of a global MHD model, it can be seen that it has the potential to improve the spectral properties of a forecast that has an underestimation of spectral power at the high frequencies.

The method was validated using a range of methods to test different aspects of the downscaling scheme. Specifically, we used the fraction skill score (FSS), Heidke skill score (HSS), Cost/Loss analysis and Brier

skill score (BSS). FSS was used to evaluate AnEn on the occurrence rate of 1-minute events within 1-hour windows. The events were defined using three thresholds, namely, $99th$, $99.9th$ and $99.99th$ percentile of the entire dataset (1983 to 2016). AnEn had a positive FSS for all ensemble members for the $99th$, $99.9th$ percentile thresholds showing that AnEn outperformed the undownscaled approach. For the $99.99th$ percentile threshold, some of the weaker analogues achieved a negative FSS suggesting that the ensemble size of 100 was too large for the current dataset to allow good analogues of the most extreme events to be found. Nevertheless, the overall FSS was still positive.

Since impacts of GICs tend to require an elevated geoelectric field over a sustained period, we also evaluated AnEn for its ability to predict the hourly-mean value of geoelectric field. This was achieved by defining events as the 1-hour mean value exceeding the thresholds of $99th$, $99.9th$ and $99.99th$ percentile of the hourly-means of the entire dataset. With this event definition, HSS revealed that AnEn outperformed the undownscaled series for all ensemble members in the three event thresholds, except for a small number in the $99.99th$ percentile events.

This work has evaluated AnEn with an ensemble size of 100. The ensemble size should be chosen large enough that a wide range of possible outcomes can be included, but small enough to ensure analogues are of a good quality and are in fact analogous. The fraction skill score and Heidke skill score revealed that better quality analogues downscaled more skilfully. The number of good quality analogues available depends both on the size of the historical dataset and on the rarity of event considered. This was particularly evident when considering events above the $99.99th$ percentile suggesting 100 members is too many to ensure all analogues are of a good quality. A more appropriate ensemble size for this threshold would be approximately 20 as shown by the BSS analysis. Future implementations of this method should use these results to inform an ap-

propriate ensemble size for the size of event of interest.

In this work the probabilistic prediction given by AnEn was made by simple ensemble member voting. The impact of analogue quality could be mitigated if, when converting to a probabilistic prediction from an ensemble of predictions, the voting power of each member is dependent on the quality of the analogue, as measure by the inverse of the RMSE between analogue and period under consideration and normalising. This would mean that members expected to have the most insight into the situation have greater sway in the overall prediction.

We implemented a probabilistic Cost/Loss analysis revealing that AnEn has a higher potential economic value than the undownscaled approach and that the value of the forecast was greater for end users who can tolerate false alarms at the lower end of the Cost/Loss domain. Like the previous metrics, AnEn performed better in the $99th$ and $99.9th$ percentile events.

A shortcoming of AnEn is that there is expected to be a lack of good analogues for the most extreme events. To address this AnEn could be improved by expanding the predictors used to include such things as geomagnetic indices and estimates of current systems. This could allow AnEn to be more aware of the drivers of geomagnetic activity and thus allow the use of fewer-but-better-quality analogues in a reduced size ensemble. Although this is a shortcoming, it is important to remember moderate space weather events are problematic as well as the rarer, more extreme events (e.g. Schrijver et al., 2014b; Schrijver, 2015). A further way to increase ensemble member quality would be to create the training dataset, $D2_M$, using a rolling-mean rather than box-car as this would create a more potential analogous periods and hence increase analogue quality overall.

We used a perfect prognostic approach to downscaling which assumes the low time resolution forecast given is a perfect forecast. This allowed us to use historical observations as if they were forecast model

	Eskdalemuir			Lerwick			Hartland		
	99	99.9	99.99	99	99.9	99.99	99	99.9	99.99
Mean FSS	0.10	0.17	0.07	0.41	0.28	0.07	0.11	-0.06	-0.04
Mean HSS	0.31	0.32	0.12	0.54	0.34	0.13	0.27	0.17	0.03
BSS	0.30	0.32	0.15	0.51	0.35	0.14	0.26	0.15	0.04

Table 5.2: The mean FSS, mean HSS and BSS for the three thresholds at Eskdalemuir, Lerwick and Hartland.

outputs. However, this approach is limited because the models are not perfect. It is expected that biases in the forecast model would not be corrected but carried through by the downscaling methodology.

This paper has focused on the results for the Eskdalemuir station, however, an equivalent analysis has been conducted for the Lerwick and Hartland magnetometer stations in the UK. The AnEn downscaling methodology applied to these stations generally perform similarly to ESK, supporting the claim that this methodology could be applied more broadly. The achieved mean FSS, mean HSS and BSS for events above the three thresholds are shown in Table 5.2 for Lerwick and Hartland. The results for ESK are also shown for reference. AnEn is shown to perform to a slightly better standard at Lerwick, particularly for the $99th$ percentile threshold, and slightly worse at Hartland, particularly for the higher thresholds.

In this work, AnEn has been used both to generate a downscaled time series and to estimate the uncertainty of it by using many ensemble members. It would be quite possible to remove the downscaling element and just use the algorithm to provide probabilistic information for a forecast that already has the correct spectral properties.

This work has given proof of concept that downscaling can be implemented to improve a forecast that lacks realistic high-frequency structure. From here, research should be conducted to create downscaling schemes that are optimised to perform better than AnEn when the downscaled data is used to drive an "impacts" model. The optimisation could include find-

ing different model configurations for specific space weather drivers. This would take knowledge of the solar wind driving the magnetosphere and restrict AnEn to choosing analogues from historical periods driven by the same solar wind context. Once downscaling methods have been further investigated, the front runners will need to be manipulated to form a "bolt-on" piece for a global MHD model. We finally note that the methods developed here do not attempt to correct for any biases in the magnetospheric models. Thus it remains to be seen whether the improvements demonstrated here translate directly to a forecasting situation, or where further bias-correction of magnetospheric models is also required.

Acknowledgements

The authors thank the National Environmental Research Council (NERC) for funding this work under grants NE/L002566/1 and NE/P016928/1.

For the ground magnetometer data we gratefully acknowledge: INTERMAGNET, Alan Thomson; CARISMA, PI Ian Mann; CANMOS, Geomagnetism Unit of the Geological Survey of Canada; The S-RAMP Database, PI K. Yumoto and Dr. K. Shiokawa; The SPIDR database; AARI, PI Oleg Troshichev; The MACCS program, PI M. Engebretson; GIMA; MEASURE, UCLA IGPP and Florida Institute of Technology; SAMBA, PI Eftyhia Zesta; 210 Chain, PI K. Yumoto; SAMNET, PI Farideh Honary; IMAGE, PI Liisa Juusola; Finnish Meteorological Institute, PI Liisa Juusola; Sodankylä Geophysical Observatory, PI Tero Raita; UiT the Arctic University of Norway, Tromsø Geophysical Observatory, PI Magnar G. Johnsen; GFZ German Research Centre For Geosciences, PI Jürgen Matzka; Institute of Geophysics, Polish Academy of Sciences, PI Anne Neska and Jan Reda; Polar Geophysical Institute, PI Alexander Yahnin and Yarolav Sakharov; Geological Survey of Sweden, PI Gerhard Schwarz; Swedish Institute of Space Physics, PI Masatoshi Yamauchi; AUTUMN, PI Martin Connors; DTU Space,

Thom Edwards and PI Anna Willer; South Pole and McMurdo Magnetometer, PI's Louis J. Lanzarotti and Alan T. Weatherwax; ICESTAR; RAPID-MAG; British Artarctic Survey; McMac, PI Dr. Peter Chi; BGS, PI Dr. Susan Macmillan; Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation (IZMIRAN); MFGI, PI B. Heilig; Institute of Geophysics, Polish Academy of Sciences, PI Anne Neska and Jan Reda; University of L'Aquila, PI M. Vellante; BCMT, V. Lesur and A. Chambodut; Data obtained in cooperation with Geoscience Australia, PI Andrew Lewis; AALPIP, co-PIs Bob Clauer and Michael Hartinger; MagStar, PI Jennifer Gannon; SuperMAG, PI Jesper W. Gjerloev; Data obtained in cooperation with the Australian Bureau of Meteorology, PI Richard Marshall. The SuperMAG data is available at `https://supermag.jhuapl.edu/`.

Simulation results have been provided by the Community Coordinated Modeling Center at Goddard Space Flight Center through their public Runs on Request system (http://ccmc.gsfc.nasa.gov). This work was carried out using the SWMF and BATS-R-US tools developed at the University of Michigan's Center for Space Environment Modeling (CSEM). The modeling tools described in this publication are available online through the University of Michigan for download and are available for use at the Community Coordinated Modeling Center (CCMC). These simulation results can be found from `https://ccmc.gsfc.nasa.gov/challenges/dBdt/`.

Chapter 6

Discussion and Conclusion

The aim of this thesis was to improve understanding and forecasting of space weather that could lead to geomagnetically induced currents (GICs). This has been achieved by providing new methodologies to the field, establishing statistical relationships and utilising high quality datasets. The three research chapters make advances in understanding duration as a parameter of geomagnetic storms, understanding and forecasting geomagnetic activity and relating low temporal resolution to high resolution parameters to aid GIC forecasting. Taken together, these research results have driven forward our ability to skilfully forecast the physical phenomena related to GIC, as summarised in more detail below

6.1 Summary

Chapter 3 focuses on predicting the duration of geomagnetic storms. Duration is a key parameter for geomagnetic storms as it informs on potential time-integrated effects, such as transformer heating, and is also useful for scheduling the resumption of operation of the affected equipment after a storm. The previous literature contains a few mentions of storm duration and some lists of geomagnetic storms contain duration as a storm param-

eter. This was the first study to establish a statistical relationship between storm duration and intensity by considering a long running measure of geomagnetic activity.

Using a peak-over-threshold approach to defining storms, it was found that more intense storms are expected to last longer, but not in a linear fashion. For a given storm peak intensity, the expected duration followed an approximate log-normal distribution. Based on this I created a method to probabilistically predict the storm duration given peak intensity. This model was evaluated and an analysis using reliability diagrams showed that there is a slight underestimate of the probability of the storm exceeding a given duration. Generally the model follows the gradient of the $y = x$ line well, showing that if the model predicts an $x\%$ chance of a storm then a storm will occur approximately $x\%$ of the time.

To further understand storm duration, a superposed epoch analysis was used. Storms were separated into six classes depending on their peak intensity and overlayed to show the generalised profile of a storm. The results indicated that more intense storms have a longer duration, with the peak occurring longer after the start of the storm. This analysis also revealed evidence of 27-day recurrence activity for the lower intensity events, most likely caused by co-rotating interaction regions (CIRs). This was not present for the more intense storms.

Chapter 4 looks at forecasting geomagnetic activity in the aa_H index, the longest reliable index of global geomagnetic activity on a 3-hourly timescale which extends back to 1868. The key result of this Chapter is bringing a new empirical methodology (the analogue ensemble, or AnEn, discussed further below) to the field of geomagnetic index forecasting and doing so in a way that provides probabilistic estimates. AnEn has previously been used for terrestrial weather forecasting and more recently for solar wind forecasting, but this is its first use for magnetospheric forecasting. It was found that, with a sufficient historical record,

it is possible to use AnEn to provide both a reliable and skilful forecast, particularly for short lead times, of the order of hours.

Previous work in the literature has focused on using solar wind drivers observed at L1 as predictors of geomagnetic indices. This choice limits the amount of data available for model training as in-situ solar wind plasma observations are only available at most for almost 60 years, depending on the parameter of interest. Chapter 4 uses only the geomagnetic index, a much more homogeneous dataset that extends for approximately three times as long, as a predictor meaning that more data-intensive methods can be used.

The two main methods evaluated for this task were the analogue ensemble (AnEn) and support vector machines (SVM) with 27-day recurrence as an additional comparison. AnEn is a "similar day" approach which identifies analogous periods to current conditions in a historical dataset and uses these to provide a forecast. The SVM is a machine-learning approach that seeks a hyper plane in the feature space to separate storm events from non-storm periods. The SVM uses a kernel to increase the dimensionality of the feature space to aid separability. 27-day recurrence simply looks at the observations from 27-days ago and forecasts them to occur again. All three of these methods outperformed climatology with SVM and AnEn outperforming 27-day recurrence. Under the true skill score (TSS, a metric robust to class imbalance which varies from $-\inf$ to 1 with 0 at no skill to 1 at perfect skill), the AnEn, SVM and 27-day recurrence achieved approximately 0.65, 0.31 and 0.12 respectively for a three hour lead time. AnEn became less skilful than 27-day recurrence for lead times of 9 hours or more whereas the SVM performed the best under TSS for lead times up to 48 hours. When considering a range of metrics, the main results were that AnEn and SVM performed similarly and the best method for an end user depends on their need for probabilistic information which AnEn can supply.

Chapter 5 developed a methodology to relate the variability in the geomagnetic (and hence geoelectric) fields across time scales. This was achieved using a long historical data set (approximately 30 years of 1-minute data). This is a necessary step to enable accurate GIC forecasting and to enable efficient estimation of forecast uncertainty. Currently operational magnetospheric MHD models are limited in their ability to produce the observed power spectrum of geomagnetic field variability, commonly giving an underestimation. This is important because it is the high frequency variation in geomagnetic field that induces the geoelectric field relevant to GICs. It is also computationally prohibitive to run large ensembles of MHD models in real time. Chapter 5 provides a proof-of-concept methodology that statistically downscales a low-resolution time series on a 1-hour resolution to a high-resolution forecast on a 1-minute resolution. The methodology takes the 1-minute observed time series and smooths this to a 1-hour resolution using a box-car mean. AnEn is then used to reintroduce perturbations on a 1-minute timescale.

Built on the AnEn methodology, the scheme successfully captures the correct spectral properties of the geomagnetic field. To fully evaluate the methodology, the downscaled time series was run through an MT-transfer function which computes the geoelectric field, as this is the quantity of concern. The same process was carried out for a reference approach to benchmark our method and for the observations. It was found that the AnEn downscaling scheme outperforms the reference model on a range of evaluation metrics providing proof that a downscaling scheme can be effective in this context. On electric field events defined using three thresholds, namely, $99th$, $99.9th$ and $99.99th$ percentile of the entire dataset (1983 to 2016), AnEn broadly achieved a positive fraction skill score and outperformed the reference approach. However, for the highest threshold, some of the weaker ensemble members achieved no or negative skill. On events of the hourly-mean electric field defined using $99th$, $99.9th$ and $99.99th$ percentile of the entire hourly-mean dataset, AnEn per-

formed well under the Hiedke skill score, a Cost/Loss analysis and the Brier skill score outperforming the reference approach.

6.2 Overarching Conclusions

The three research chapters presented in this thesis form a cohesive piece of research and as such there are certain conclusions to be made in the context of them together. These are presented in the current section.

A key theme of this thesis has been the use of long-running homogeneous datasets. These datasets provide value for scientific understanding by allowing robust statistical relationships to be established. They also allow empirical forecasting methods to be fully utilised as the dataset should contain all but the rarest behaviour of the system assuming that it is a time-stationary system, an assumption that would not be true, for example, of annual maximum temperatures in the presence of climate change. The long datasets used in this thesis are the aa_H index, a 150-year record of geomagnetic activity comprised of a combination of observations from the UK and Australia, and 34-year records of the ground-level geomagnetic field at three UK locations provided by SuperMAG. The AnEn methodology relies on the existence of analogous periods in a historical dataset to gain insight into current/future variations. The likelihood of finding an analogous period and the quality of that analogue is dependent on a long-running dataset that has experienced all the events of interest. That is to say that the dataset must have seen it before for the AnEn to be able to succeed. While these two datasets have been running an impressively long time, the number of solar cycles they have recorded is less impressive at approximately 14 and 3, for aa_H and the SuperMAG data, respectively. It is therefore expected that there is rare behaviour that these datasets have not observed which will limit the ability of AnEn to provide accurate prediction for the largest of events. For this reason it is of high importance

to ensure these observations are continued on into the future.

Chapters 4 and 5 both employed the analogue ensemble method. While this method was developed for terrestrial weather forecasting this thesis has shown it useful for magnetospheric activity also. One of the main benefits of this method is that it provides a computationally cheap ensemble of realistic predictions. The size of the ensemble has little impact on computational efficiency but is, however, bounded by the number of similar periods expected to be contained within the dataset. This is a particular limit if the dataset is small in duration or if extreme events are of particular interest. In Chapter 5 rare events were studied in the form of high percentile thresholds. It was found that the 34-year record was able to provide 20-30 good quality analogues for event of above the $99.99th$ percentile. However, sufficient analogues were available to have a 100 member ensemble for events above the $99.9th$, $99th$ and, in Chapter 4, the $90th$ percentile for the aa_H index.

A contribution of Chapters 3 and 4 was the identification of the aa_H index as a suitable dataset for geomagnetic activity studies. The aa_H index is a recent development of the *aa* index and this thesis is some of the first work to use it. Although there are newer indices that have better longitudinal coverage (e.g. *am*), aa_H has an unrivalled length of 150 years. The significance of this is that many of the empirical forecast techniques previously used, such as neural networks, require a large historic database to perform at their best. Having shown the usefulness and appropriateness of aa_H this work acts as a signpost for those in the field to experiment with as a means to improve their own models.

Furthermore, aa_H has allowed this thesis to exploit statistical features of magnetospheric activity. Previous studies looking to do the same would either have compromised on the quality of the dataset by using the *aa* index or sacrificed almost a century of records by using an alternate index. The large and high quality data was instrumental to the work in

Chapter 4 as it enabled both of the forecasting techniques to work to their potential. This was particularly the case for AnEn which needs a large number of good quality analogues. The long aa_H dataset helped the work presented in Chapter 3 by increasing the certainty surrounding the statistical link between storm duration and intensity.

6.3 Future Work

The availability of large data sets lends itself to the use of deep-learning methodologies for predictive tasks such as nowcasting and forecasting, as examined in Chapter 4. With more time I would look to implement a long short-term memory (LSTM) neural network for this time series forecasting task. The LSTM is a type of recurrent neural network that has been shown to work particularly well on timeseries data. I would hope that the LSTM could capture non-linear relationships between the time history and the prediction that would be especially useful in predicting a storm onset. I would also be interested in the ability of the LSTM to take solar wind observations at L1 and provide a forecast for geomagnetic activity in the aa_H index. Further, it could give a forecast of ground-level magnetic field as observed by a SuperMAG magnetometer. This forecast would be of the same resolution as the observations and so could benefit from downscaling.

The evaluation of the work in Chapter 5 was based on the impacts of geomagnetic activity being proportional to the magnitude of the geoelectric field. The evaluation would have benefited from looking at a specific transformer with sufficient GIC observations alongside the set up parameters of the transformer. It would then be possible to evaluate the downscaling methodology on its ability to predict GICs. The main barrier to this is the lack of availability of reliable GIC observations over a sufficient time period.

Chapters 4 and 5 used the AnEn method with 100 ensemble members. As identified in Chapter 5 the appropriate number of analogues is dependent on the rarity of events under consideration. For example, if a very rare event is being considered then, by definition, there will not be many analogous periods for the method to use. It would be useful for future work to provide a model to estimate appropriate ensemble size given event rarity and dataset length. This could be done by either estimating the number of events in the historical dataset using the event amplitude, or by only including analogues that have a normalised RMSE of a determined value or less, rather than including a fixed number of best analogues.

In this thesis, where AnEn has been used to give a deterministic or probabilistic prediction, simple ensemble voting has been used with each ensemble member given equal weight. This approach ignores information on the quality of each analogue, i.e. how analogous the analogue period really is. This quality is measured using RMSE inside the AnEn algorithm. It is particularly evident in Figure 5.7 that those with a better quality analogue (i.e. lower Ensemble Member ID) are producing a better result. In the context of converting the ensemble into a deterministic or probabilistic prediction, it would be beneficial to weight ensemble members as they vote so that better quality members have more of a say.

In Chapter 3 it was established that storms in the aa_H index have statistically related durations and intensities. This relationship could be better understood and characterised by examining other geomagnetic indices. aa_H was chosen because it is the longest running geomagnetic index and would therefore show statistical relationship clearest. However, indices such as Dst and Kp have records for several solar cycles and should exhibit this relationship. For the power industry concerned with GICs, it would be useful to establish the duration-intensity relationship in the geoelectric field. A good proxy for such a record could be computed from the ob-

served geomagnetic field using an MT-transfer function as implemented in Chapter 5.

The model built for predicting storm duration was shown to exhibit a slight underestimation of the probability of storms exceeding certain duration. Future work could focus on calibrating the model to adjust for this underestimation and correct the reliability diagram. Since the reliability diagram shows lines that are predominantly linear, a calibration could be performed using a simple scaling and translation as a function of the model probability. The underestimation may represent a second class of storms that the model has not captured. Should this class exist, it may be due to "compound" storms driven by a pile up of solar wind events.

Chapter 5 used an AnEn approach to provide a proof-of-concept study into downscaling. While AnEn has been successful, there is no reason (*a priori*) to believe it is the best approach available. Further approaches should be researched in future work and evaluated with the same datasets and metrics to allow a like-for-like comparison. One possibility would be to relate low-frequency variations to the parameters of the observed high-frequency power spectrum. Once a relationship is established a coloured noise generator could be used to produce a high frequency time-series with the correct power spectral density.

The end-goal of downscaling is to create a "bolt-on" module for use with global MHD models. Once various downscaling methods have been investigated the front runners should be tested on global MHD models. The downscaling approach here took a perfect prognostic approach, meaning that the forecast is assumed to be perfect. This assumption will not hold for global MHD models, so the effect of model biases will need to be evaluated to test whether downscaling is operationally viable without using model output statistics, which need to be computed on a model-by-model basis.

References

Abda, Z.M.K., Ab Aziz, N.F., Mohd Zainal Abidin Ab Kadir and Rhazali, Z.A., 2020. A review of geomagnetically induced current effects on electrical power system: Principles and theory. *Ieee access*, 8(March 1989), pp.200237–200258. Available from: https://doi.org/10.1109/ACCESS.2020.3034347.

Abunina, M., Papaioannou, A., Gerontidou, M., Paschalis, P., Abunin, A., Gaidash, S., Tsepakina, I., Malimbayev, A., Belov, A., Mavromichalaki, H., Kryakunova, O. and Velinov, P., 2013. Forecasting geomagnetic conditions in near-Earth space. *Journal of physics: Conference series*, 409(1). Available from: https://doi.org/10.1088/1742-6596/409/1/012197.

Allain, D. and Mitchell, C., 2008. Mitigation of delay and scintillation effects using MIDAS. *Ursi 2008 (xxix general assembly of the international union of radio science / union radio scientifique internationale*.

Allison, P.D., 1982. Discrete-Time Methods for the Analysis of Event Histories. *Sociological methodology*, 13(1982), pp.61–98. Available from: https://doi.org/10.2307/270718.

Antiochos, S.K., Mikić, Z., Titov, V.S., Lionello, R. and Linker, J.A., 2011. A model for the sources of the slow solar wind. *Astrophysical journal*, 731(2). 1102.3704, Available from: https://doi.org/10.1088/0004-637X/731/2/112.

Aol, S., Buchert, S. and Jurua, E., 2020. Ionospheric irregularities and scintillations: a direct comparison of in situ density observations with ground-based L-band receivers. *Earth, planets and space*, 72(1). Available from: https://doi.org/10.1186/s40623-020-01294-z.

Arridge, C., 2020. Solar wind interaction with planets. *Oxford research encyclopedia of physics*. Oxford University Press. Available from: https://doi.org/10.1093/acrefore/9780190871994.013.15.

Avdeev, D.B., 2005. *Three-dimensional electromagnetic modelling and inversion from theory to application*, vol. 26. Available from: https://doi.org/10.1007/s10712-005-1836-x.

Ayala Solares, J.R., Wei, H.L., Boynton, R.J., Walker, S.N. and Billings, S.A., 2016. Modeling and prediction of global magnetic disturbance in near-Earth space: A case study for Kp index using NARX models. *Space weather*, 14(10), pp.899–916. Available from: https://doi.org/10.1002/2016SW001463.

Baker, D.N., 1998. What is space weather? *Advances in space research*, 22(1), pp.7–16. Available from: https://doi.org/10.1016/S0273-1177(97)01095-8.

Balan, N., Batista, I.S., Tulasi Ram, S. and Rajesh, P.K., 2016. A new parameter of geomagnetic storms for the severity of space weather. Available from: https://doi.org/10.1186/s40562-016-0036-5.

Barnard, L., Lockwood, M., Hapgood, M.A., Owens, M.J., Davis, C.J. and Steinhilber, F., 2011. Predicting space climate change. *Geophysical research letters*, 38(16), pp.7–12. Available from: https://doi.org/10.1029/2011GL048489.

Bartels, J., 1932. Terrestrial-magnetic activity and its relations to solar phenomena. *Journal of geophysical research*, 37(1), p.1. Available from: https://doi.org/10.1029/te037i001p00001.

Bartels, J., 1934. Twenty-seven day recurrences in terrestrial-magnetic and solar activity, 1923–1933. *Journal of geophysical research*, 39(3), p.201. Available from: https://doi.org/10.1029/te039i003p00201.

Bartels, J., Heck, N.H. and Johnston, H.F., 1939. The three-hour-range index measuring geomagnetic activity. *Terr. magn. atmos. electr.*, 44(4), pp.411–424. Available from: https://doi.org/10.1029/te044i004p00411.

Beamish, D., Clark, T.D., Clarke, E. and Thomson, A.W., 2002. Geomagnetically induced currents in the UK: Geomagnetic variations and surface electric fields. *Journal of atmospheric and solar-terrestrial physics*, 64(16), pp.1779–1792. Available from: https://doi.org/10.1016/S1364-6826(02)00127-X.

Beggan, C.D., Richardson, G.S., Baillie, O., Hubert, J. and Thomson, A.W.P., 2021. Geoelectric field measurement, modelling and validation during geomagnetic storms in the UK. *Journal of space weather and space climate*.

Bloomfield, D.S., Higgins, P.A., McAteer, R.T. and Gallagher, P.T., 2012. Toward reliable benchmarking of solar flare forecasting methods. *Astrophysical journal letters*, 747(2). 1202.5995, Available from: https://doi.org/10.1088/2041-8205/747/2/L41.

Borovsky, J.E. and Denton, M.H., 2006. Differences between CME-driven storms and CIR-driven storms. *Journal of geophysical research: Space physics*, 111(7), pp.1–17. Available from: https://doi.org/10.1029/2005JA011447.

Boteler, D.H., 1994. Geomagnetically induced currents: present knowledge and future research. *Ieee*, 9(1), pp.50–58.

Boteler, D.H., 2001. Space weather effects on power systems. *Space*

weather. AGU publications, pp.347–352. Available from: https://doi.org/10.1029/GM125p0347.

Boteler, D.H., 2019. A 21st Century View of the March 1989 Magnetic Storm. *Space weather*, 17(10), pp.1427–1441. Available from: https://doi.org/10.1029/2019SW002278.

Bubenik, D.M. and Fraser-Smith, A.C., 1977. Evidence for strong artificial components in the equivalent linear amplitude geomagnetic indices. *Journal of geophysical research*, 82(19), pp.2875–2878. Available from: https://doi.org/10.1029/ja082i019p02875.

Buonsanto, M.J., 1999. Ionospheric storms - A review. *Space science reviews*, 88(3-4), pp.563–601. Available from: https://doi.org/10.1023/a:1005107532631.

Burges, C.J., 1998. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2), pp.121–167. Available from: https://doi.org/10.1023/A:1009715923555.

Burton, R.K., McPherron, R.L. and Russell, C.T., 1975. An empirical relationship between interplanetary conditions and Dst. *Journal of geophysical research*, 80(31), pp.4204–4214. Available from: https://doi.org/10.1029/ja080i031p04204.

Camporeale, E., 2019. The Challenge of Machine Learning in Space Weather: Nowcasting and Forecasting. *Space weather*, pp.1–42. arXiv:1411.4389, Available from: https://doi.org/10.1029/2018sw002061.

Cannon, P., Angling, M., Barclay, L., Curry, C., Dyer, C., Edwards, R., Greene, G., Hapgood, M., Horne, R., Jackson, D., Mitchell, C., Owen, J., Richards, A., Rogers, C., Ryden, K., Saunders, S., Sweeting, M., Tanner, R., Thomson, A. and Underwood, C., 2013. Extreme space weather: impacts on engineered systems and infrastruc-

tures. *Royal academy of engineering*, p.70. Available from: https://doi.org/1-903496-96-9.

Carlin, J.B., Wolfe, R., Coffey, C. and Patton, G.C., 2005. Survival Models: Analysis of Binary Outcomes in Longitudinal Studies Using Weighted Estimating Equations and Discrete-Time Survival Methods: Prevalence and Incidence of Smoking in An Adolescent Cohort. *Tutorials in biostatistics, statistical methods in clinical studies*, 1(December 1999), pp.161–185. Available from: https://doi.org/10.1002/0470023678.ch1d(ii).

Carlson, H. and Egeland, A., 1995. The aurora and the auroral ionosphere. *Introduction to space physics*. Cambridge University Press.

Chandorkar, M., Camporeale, E. and Wing, S., 2017. Probabilistic forecasting of the disturbance storm time index: An autoregressive Gaussian process approach. *Space weather*, 15(8), pp.1004–1019. Available from: https://doi.org/10.1002/2017SW001627.

Chapman, S. and Ferraro, V.C.A., 1931. A new theory of magnetic storms. 97, pp.171–186.

Chapman, S.C., Horne, R.B. and Watkins, N.W., 2020. Using the aa Index Over the Last 14 Solar Cycles to Characterize Extreme Geomagnetic Activity. *Geophysical research letters*, 47(3). Available from: https://doi.org/10.1029/2019GL086524.

Charbonneau, P., 2020. *Dynamo models of the solar cycle*, vol. 17. Springer International Publishing. Available from: https://doi.org/10.1007/s41116-020-00025-6.

Chen, J., 2017. Physics of erupting solar flux ropes: Coronal mass ejections (CMEs) - Recent advances in theory and observation. *Physics of plasmas*, 24(9). Available from: https://doi.org/10.1063/1.4993929.

Chree, C. and Stagg, M., 1928. Recurrence phenomena in terrestrial magnetism. *Royal society*, 227(647-658), pp.21–62. Available from: https://doi.org/10.1098/rsta.1928.0002.

Christensen, J.H. and Christensen, O.B., 2003. Severe summertime flooding in Europe. *Nature*, 421(6925), pp.805–806. Available from: https://doi.org/10.1038/421805a.

Claudepierre, S.G., Wiltberger, M., Elkington, S.R., Lotko, W. and Hudson, M.K., 2009. Magnetospheric cavity modes driven by solar wind dynamic pressure fluctuations. *Geophysical research letters*, 36(13), pp.1–5. Available from: https://doi.org/10.1029/2009GL039045.

Cortes, C. and Vapnik, V., 1995. Support-Vector Networks. *Machine leaming*, (20), pp.273–297. Available from: https://doi.org/10.1007/BF00994018.

Courant, Friedrichs and Lewy, 1956. *On the partial difference equations of mathematical physics*. New York: AEC Computing and Applied Mathematics Centre.

Cover, T.M., 1965. Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition. *Ieee transactions on electronic computers*, EC-14(3), pp.326–334. Available from: https://doi.org/10.1109/PGEC.1965.264137.

Cranmer, S.R., 2009. Coronal Holes. *Living reviews in solar physics*, 6, p.66. Available from: http://adsabs.harvard.edu/abs/2009LRSP....6....3C.

Daglis, I.A., Throne, R.M., Baumjohann, W. and Orsini, S., 1999. The terrestrial ring current: origin, formation and decay. *Reviews of geophysics*, pp.407–438. Available from: https://doi.org/10.1029/1999RG900009.

De Zeeuw, D.L., Gombosi, T.I., Groth, C.P., Powell, K.G. and Stout, Q.F., 2000. An adaptive MHD method for global space weather simulations. *Ieee transactions on plasma science*, 28(6), pp.1956–1965. Available from: https://doi.org/10.1109/27.902224.

Delle Monache, L., Eckel, F.A., Rife, D.L., Nagarajan, B. and Searight, K., 2013. Probabilistic Weather Prediction with an Analog Ensemble. *Monthly weather review*, 141(10), pp.3498–3516. Available from: https://doi.org/10.1175/mwr-d-12-00281.1.

Dimmock, A.P., Welling, D.T., Rosenqvist, L., Forsyth, C., Freeman, M.P., Rae, I.J., Viljanen, A., Vandegriff, E., Boynton, R.J., Balikhin, M.A. and Yordanova, E., 2021. Modeling the Geomagnetic Response to the September 2017 Space Weather Event Over Fennoscandia Using the Space Weather Modeling Framework: Studying the Impacts of Spatial Resolution. *Space weather*, 19(5), pp.1–25. Available from: https://doi.org/10.1029/2020sw002683.

Dool, H.M. van den, 1989. A New Look at Weather Forecasting through Analogues. *Monthly weather review*, 117(10), pp.2230–2247. Available from: https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2.

Driel-Gesztelyi, L. van and Owens, M., 2020. Solar Cycle. Available from: https://doi.org/10.1093/acrefore/9780190871994.013.9.

Dungey, J.W., 1961. Interplanetary Magnetic Field and the Auroral Zones. *American physical society*. Available from: https://doi.org/10.1103/PhysRevLett.6.47.

Eastwood, J.P., Biffis, E., Hapgood, M.A., Green, L., Bisi, M.M., Bentley, R.D., Wicks, R., McKinnell, L.A., Gibbs, M. and Burnett, C., 2017a. The Economic Impact of Space Weather: Where Do We Stand? *Risk analysis*, 37(2). Available from: https://doi.org/10.1111/risa.12765.

Eastwood, J.P., Nakamura, R., Turc, L., Mejnertsen, L. and Hesse, M., 2017b. The Scientific Foundations of Forecasting Magnetospheric Space Weather. *Space science reviews*, 212(3-4), pp.1221–1252. Available from: https://doi.org/10.1007/s11214-017-0399-8.

Echer, E., Gonzalez, W.D. and Tsurutani, B.T., 2008. Interplanetary conditions leading to superintense geomagnetic storms (Dst -250 nT) during solar cycle 23. *Geophysical research letters*, 35(6). Available from: https://doi.org/10.1029/2007GL031755.

El-Alaoui, M., Richard, R.L., Ashour-Abdalla, M., Walker, R.J. and Goldstein, M.L., 2012. Turbulence in a global magnetohydrodynamic simulation of the Earth's magnetosphere during northward and southward interplanetary magnetic field. *Nonlinear processes in geophysics*, 19(2), pp.165–175. Available from: https://doi.org/10.5194/npg-19-165-2012.

Elbert, B., 2008. *Introduction to Satellite Communications: Third edition*.

Finch, I., 2007. The Use of Geomagnetic Activity Observations in Studies of Solar Wind – Magnetosphere Coupling and Centennial Solar Change by. (September 2007).

Forbes, K.F. and St. Cyr, O.C., 2004. Space weather and the electricity market: An initial assessment. *Space weather*, 2(10), pp.n/a–n/a. Available from: https://doi.org/10.1029/2003sw000005.

Forbes, K.F. and St.Cyr, O.C., 2017. The Challenge Posed by Geomagnetic Activity to Electric Power Reliability: Evidence From England and Wales. *Space weather*, 15(10), pp.1413–1430. Available from: https://doi.org/10.1002/2017SW001668.

Ganushkina, N.Y., Liemohn, M.W. and Dubyagin, S., 2018. Current Systems in the Earth's Magnetosphere. *Reviews of geophysics*, 56(2), pp.309–332. Available from: https://doi.org/10.1002/2017RG000590.

Ganushkina, N.Y., Liemohn, M.W., Dubyagin, S., Daglis, I.A., Dandouras, I., De Zeeuw, D.L., Ebihara, Y., Ilie, R., Katus, R., Kubyshkina, M., Milan, S.E., Ohtani, S., Ostgaard, N., Reistad, J.P., Tenfjord, P., Toffoletto, F., Zaharia, S. and Amariutei, O., 2015. Defining and resolving current systems in geospace. *Annales geophysicae*, 33(11), pp.1369–1402. Available from: https://doi.org/10.5194/angeo-33-1369-2015.

Gaunt, C.T. and Coetzee, G., 2007. Transformer failures in regions incorrectly considered to have low GIC-risk. *2007 ieee lausanne powertech, proceedings*, pp.807–812. Available from: https://doi.org/10.1109/PCT.2007.4538419.

Gjerloev, J.W., 2012. The SuperMAG data processing technique. *Journal of geophysical research: Space physics*, 117(9), pp.1–19. Available from: https://doi.org/10.1029/2012JA017683.

Gombosi, T.I., Tóth, G., De Zeeuw, D.L., Hansen, K.C., Kabin, K. and Powell, K.G., 2002. Semirelativistic magnetohydrodynamics and physics-based convergence acceleration. *Journal of computational physics*, 177(1), pp.176–205. Available from: https://doi.org/10.1006/jcph.2002.7009.

Gonzalez, W.D., Joselyn, J.A., Kamide, Y., Kroehl, H.W., Rostoker, G., Tsurutani, B.T. and Vasyliunas, V.M., 1994. What is a geomagnetic storm? *Journal of geophysical research*. Available from: https://doi.org/10.1029/93JA02867.

Gosling, J.T. and Pizzo, V.J., 1999. Formation and Evolution of Corotating Interaction Regions and their Three Dimensional Structure. *Space science reviews*. Available from: https://doi.org/10.1023/A:1005291711900.

Grawe, M.A., Makela, J.J., Butala, M.D. and Kamalabadi, F., 2018. The Impact of Magnetic Field Temporal Sampling on Modeled Surface Elec-

tric Fields. *Space weather*, 16(11), pp.1721–1739. Available from: https://doi.org/10.1029/2018SW001896.

Green, L.M., Török, T., Vršnak, B., Manchester, W. and Veronig, A., 2018. The Origin, Early Evolution and Predictability of Solar Eruptions. *Space science reviews*, 214(1). 1801.04608, Available from: https://doi.org/10.1007/s11214-017-0462-5.

Gu, Y., Wei, H.L., Boynton, R.J., Walker, S.N. and Balikhin, M.A., 2019. System Identification and Data-Driven Forecasting of AE Index and Prediction Uncertainty Analysis Using a New Cloud-NARX Model. *Journal of geophysical research: Space physics*, 124(1), pp.248–263. Available from: https://doi.org/10.1029/2018JA025957.

Haiducek, J.D., Welling, D.T., Ganushkina, N.Y., Morley, S.K. and Ozturk, D.S., 2017. SWMF Global Magnetosphere Simulations of January 2005: Geomagnetic Indices and Cross-Polar Cap Potential. *Space weather*, 15(12), pp.1567–1587. Available from: https://doi.org/10.1002/2017SW001695.

Haines, C., 2021a. *Carl-Haines/AnalogueEnsemble: First release of Analogue Ensemble*. Zenodo. Available from: https://doi.org/10.5281/zenodo.4604487.

Haines, C., 2021b. *Carl-Haines/SVM_for_aaH: First release of SVM for aaH*. Zenodo. Available from: https://doi.org/10.5281/zenodo.4604485.

Haines, C., Owens, M., Barnard, L., Lockwood, M., Beggan, C., Thomson, A. and Rogers, N., 2021a. Towards GIC forecasting: Statistical downscaling of the geomagnetic field to improve geoelectric field forecasts. *Space weather*, pp.1–28. Available from: https://doi.org/10.1029/2021sw002903.

Haines, C., Owens, M.J., Barnard, L., Lockwood, M. and Ruffenach, A., 2019. The Variation of Geomagnetic Storm Duration with Intensity.

Haines, C., Owens, M.J., Barnard, L., Lockwood, M. and Ruffenach, A., 2021b. Forecasting Occurrence and Intensity of Geomagnetic Activity with Pattern-Matching Approaches. *Submitted to space weather*, pp.1–23. Available from: https://doi.org/10.1029/2020SW002624.

Haines, G.V., 1985. Spherical cap harmonic analysis. *Journal of geophysical research*, 90(B3), pp.2583–2591. Available from: https://doi.org/10.1029/JB090iB03p02583.

Hapgood, M.A., 2010. Towards a scientific understanding of the risk from extreme space weather. *Advances in space research*, 47(12), pp.2059–2072. Available from: https://doi.org/10.1016/j.asr.2010.02.007.

Hartinger, M.D., Welling, D., Viall, N.M., Moldwin, M.B. and Ridley, A., 2014. The effect of magnetopause motion on fast mode resonance. *Journal of geophysical research: Space physics*, 119(10), pp.8212–8227. Available from: https://doi.org/10.1002/2014JA020401.

Hathaway, D.H., 2015. The solar cycle. *Living reviews in solar physics*, 12(1). 1502.07020, Available from: https://doi.org/10.1007/lrsp-2015-4.

Heindl, M., Beltle, M., Reuter, M., Schneider, D., Tenbohlen, S., Oyedokun, D.T., Gaunt, C.T. and Africa, S., 2011. INVESTIGATION OF GIC RELATED EFFECTS ON POWER DC- Load Source TUT.

Henley, E.M. and Pope, E.C., 2017. Cost-Loss Analysis of Ensemble Solar Wind Forecasting: Space Weather Use of Terrestrial Weather Tools. *Space weather*, 15(12), pp.1562–1566. Available from: https://doi.org/10.1002/2017SW001758.

Hundhausen, A.J., 1995. The Solar Wind. *Introduction to space physics*. Cambridge University Press, chap. The Solar.

Hutchinson, J.A., Wright, D.M. and Milan, S.E., 2011. Geomagnetic storms over the last solar cycle: A superposed epoch analysis. *Journal of geophysical research: Space physics*, 116(9), pp.1–16. Available from: https://doi.org/10.1029/2011JA016463.

Jolliffe, I. and Stephenson, D., 2003. *Forecast verification: a practioners guide in atmospheric science*, vol. 4.

Joselyn, J.A., 1995. Geomagnetic Activity Forecasting: The State OF the Art. *American geophysical union*, (18), pp.383–401. Available from: https://doi.org/10.1029/95RG01304.

Kappenman, J., Ave, S.F. and Suite, E., 2010. Geomagnetic Storms and Their Impacts on the U.S. Power Grid. (805), p.197.

Kappenman, J.G., 1996. Geomagnetic storms and their impact on power systems. *Ieee power engineering review*, 16(5), p.5. Available from: https://doi.org/10.1109/MPER.1996.491910.

Kelbert, A., 2020. *The Role of Global/Regional Earth Conductivity Models in Natural Geomagnetic Hazard Mitigation*, vol. 41. Springer Netherlands. Available from: https://doi.org/10.1007/s10712-019-09579-z.

Kelbert, A., Bedrosian, P. and Murphy, B., 2019. The first 3D conductivity model of the contiguous United States. *Geomagnetically induced currents from the sun to the power grid*.

Kepko, L., Viall, N.M., Antiochos, S.K., Lepri, S.T., Kasper, J.C. and Weberg, M., 2016. Implications of L1 observations for slow solar wind formation by solar reconnection. *Geophysical research letters*, 43(9), pp.4089–4097. Available from: https://doi.org/10.1002/2016GL068607.

Kilpua, E., Koskinen, H.E.J. and Pulkkinen, T.I., 2017. Coronal mass ejections and their sheath regions in interplanetary space. *Living reviews in*

solar physics, 14(1), pp.1–83. Available from: https://doi.org/10.1007/s41116-017-0009-6.

Kilpua, E.K., Olspert, N., Grigorievskiy, A., Käpylä, M.J., Tanskanen, E.I., Miyahara, H., Kataoka, R., Pelt, J. and Liu, Y.D., 2015. Statistical Study of Strong and Extreme Geomagnetic Disturbances and Solar Cycle Characteristics. *Astrophysical journal*, 806(2). Available from: https://doi.org/10.1088/0004-637X/806/2/272.

Knipp, D., 2016. Advances in space weather ensemble forecasting. *Space weather*, 14(2), pp.113–135. Available from: https://doi.org/https://agupubs.onlinelibrary.wiley.com/doi/full/10.1002/2016SW001366.

Knipp, D.J., Hapgood, M.A. and Welling, D., 2018. Communicating Uncertainty and Reliability in Space Weather Data, Models, and Applications. *Space weather*, 16(10), pp.1453–1454. Available from: https://doi.org/10.1029/2018SW002083.

Koskinen, H.E., Baker, D.N., Balogh, A., Gombosi, T., Veronig, A. and Steiger, R. von, 2017. Achievements and Challenges in the Science of Space Weather. *Space science reviews*, 212(3-4), pp.1137–1157. Available from: https://doi.org/10.1007/s11214-017-0390-4.

Krimigis, S.M., Haerendel, G., McEntire, R.W., Paschmann, G. and Bryant, D.A., 1982. The active magnetospheric particle tracer explorers (AMPTE) program. *Eos, transactions american geophysical union*, 63(45), pp.843–850. Available from: https://doi.org/10.1029/EO063i045p00843.

Le, G., Russell, C.T. and Takahashi, K., 2004. Morphology of the ring current derived from magnetic field observations. *Annales geophysicae*, 22(4), pp.1267–1295. Available from: https://doi.org/10.5194/angeo-22-1267-2004.

Lefevre, L., Clette, F., Malandraki, O., Dumbović, M., Crosby, N., Patsou, I., Vršnak, B., Leer, K., Moretto, T., Veronig, A. and Vennerstrom, S., 2016. Extreme Geomagnetic Storms – 1868 – 2010. *Solar physics*, 291(5), pp.1447–1481. Available from: https://doi.org/10.1007/s11207-016-0897-y.

Liemohn, M.W., Kozyra, J.U., Clauer, C.R. and Ridley, A.J., 2001. Computational analysis of the near-Earth magnetospheric current system during two-phase decay storms. *Journal of geophysical research: Space physics*, 106(A12), pp.29531–29542. Available from: https://doi.org/10.1029/2001ja000045.

Liemohn, M.W., McCollough, J.P., Jordanova, V.K., Ngwira, C.M., Morley, S.K., Cid, C., Tobiska, W.K., Wintoft, P., Ganushkina, N.Y., Welling, D.T., Bingham, S., Balikhin, M.A., Opgenoorth, H.J., Engel, M.A., Weigel, R.S., Singer, H.J., Buresova, D., Bruinsma, S., Zhelavskaya, I.S., Shprits, Y.Y. and Vasile, R., 2018. Model Evaluation Guidelines for Geomagnetic Index Predictions. *Space weather*, 16(12), pp.2079–2102. Available from: https://doi.org/10.1029/2018SW002067.

Lockwood, M., 2013. Reconstruction and prediction of variations in the open Solar Magnetic Flux and interplanetary conditions. *Living reviews in solar physics*, 10. Available from: https://doi.org/10.12942/lrsp-2013-4.

Lockwood, M., 2019a. Does Adding Solar Wind Poynting Flux Improve the Optimum Solar Wind-Magnetosphere Coupling Function? *Journal of geophysical research: Space physics*, 124(7), pp.5498–5515. Available from: https://doi.org/10.1029/2019JA026639.

Lockwood, M., 2019b. Lecture notes from Space Weather, University of Reading.

Lockwood, M., Chambodut, A., Barnard, L.A., Owens, M.J. and Mendel,

V., 2018a. A homogeneous aa index : 1 . Secular variation. 2018(April), pp.1–57.

Lockwood, M., Chambodut, A., Finch, I.D., Barnard, L.A., Owens, M.J. and Haines, C., 2019. Time-of-day / time-of-year response functions of planetary geomagnetic indices. *J. space weather and space climate*, 9. Available from: https://doi.org/https://doi.org/10.1051/swsc/2019017.

Lockwood, M., Finch, I.D., Chambodut, A., Barnard, L.A., Owens, M.J. and Clarke, E., 2018b. A homogeneous aa index: 2. Hemispheric asymmetries and the equinoctial variation. *Journal of space weather and space climate*, 8(A58). Available from: https://doi.org/10.1051/swsc/2018044.

Lockwood, M., Owens, M.J., Barnard, L.A., Bentley, S., Scott, C.J. and Watt, C.E., 2016. On the origins and timescales of geoeffective IMF. *Space weather*, 14(6), pp.406–432. Available from: https://doi.org/10.1002/2016SW001375.

Lockwood, M., Owens, M.J., Barnard, L.A., Scott, C.J. and Watt, C.E., 2017. Space climate and space weather over the past 400 years: 1. The power input to the magnetosphere. *Journal of space weather and space climate*, 7, p.A25. 1801.02434, Available from: https://doi.org/10.1051/swsc/2017019.

Longadge, R., Dongre, S. and Malik, L., 2013. Class imbalance problem in data mining: review. *Internation journal of computer science and network*, 2(1).

Lu, J.Y., Peng, Y.X., Wang, M., Gu, S.J. and Zhao, M.X., 2016. Support Vector Machine combined with Distance Correlation learning for Dst forecasting during intense geomagnetic storms. *Planetary and space science*, 120, pp.48–55. Available from: https://doi.org/10.1016/j.pss.2015.11.004.

Lui, A.T.Y., McEntire, R.W. and Krimigis, S.M., 1987. Evolution of the ring current during two geomagnetic storms. *Journal of geophysical research*, 92(A7), p.7459. Available from: https://doi.org/10.1029/ja092ia07p07459.

Lundstedt, H., Gleisner, H. and Wintoft, P., 2003. Operational forecasts of the geomagnetic Dst index. *Geophysical research letters*, 29(24), pp.34–1–34–4. Available from: https://doi.org/10.1029/2002gl016151.

Lyon, J.G., Fedder, J.A. and Mobarry, C.M., 2004. The Lyon-Fedder-Mobarry (LFM) global MHD magnetospheric simulation code. *Journal of atmospheric and solar-terrestrial physics*, 66(15-16 SPEC. ISS.), pp.1333–1350. Available from: https://doi.org/10.1016/j.jastp.2004.03.020.

Malandraki, O. and Crosby, N., 2018. Solar Energetic Particles and Space Weather: Science and Applications. *Solar particle radiation storms forecasting and analysis*. pp.1–26. Available from: https://doi.org/10.1007/978-3-319-60051-2_1.

Maraun, D., Brienen, S., Rust, H.W., Sauter, T., Themeßl, M., Venema, V.K.C. and Chun, K.P., 2010. Precipitation Downscaling under Climate Change. *October*, (2009), pp.1–34. Available from: https://doi.org/10.1029/2009RG000314.

Marti, L., Rezaei-Zare, A. and Narang, A., 2013. Simulation of transformer hotspot heating due to geomagnetically induced currents. *Ieee transactions on power delivery*, 28(1), pp.320–327. Available from: https://doi.org/10.1109/TPWRD.2012.2224674.

Mayaud, P.N., 1971. Une mesure planétaire d'activité magnetique, basée sur deux observatoires antipodaux. *Ann. geophys.*, 27, 67–70. Available from: https://doi.org/10.1002/9781118663837.

Mayaud, P.N., 1981. Planetary Indices Derived From K Indices (Kp, am, and aa). Available from: https://doi.org/https://doi.org/10.1002/9781118663837.ch5.

McGranaghan, R.M., Mannucci, A.J., Wilson, B., Mattmann, C.A. and Chadwick, R., 2018. New Capabilities for Prediction of High-Latitude Ionospheric Scintillation: A Novel Approach With Machine Learning. *Space weather*, 16(11), pp.1817–1846. Available from: https://doi.org/10.1029/2018SW002018.

McPherron, R.L., 1995. Magnetospheric dynamics. *Introduction to space physics*. chap. 13. Available from: https://doi.org/10.1017/9781139878296.014.

McPherron, R.L., Russell, C.T. and Aubry, M.P., 1973. Satellite studies of magnetospheric substorms on August 15, 1968: 9. Phenomenological model for substorms. *Journal of geophysical research*, 78(16), pp.3131–3149. Available from: https://doi.org/10.1029/ja078i016p03131.

Merkin, V.G., Owens, M.J., Spence, H.E., Hughes, W.J. and Quinn, J.M., 2007. Predicting magnetospheric dynamics with a coupled Sun-to-Earth model: Challenges and first results. *Space weather*, 5(12), pp.1–13. Available from: https://doi.org/10.1029/2007SW000335.

Mohd Anuar, N., Mohd Kasran, F.A., Abbas, M., Jusoh, M.H., Enche Ab Rahim, S.A., Abdul Hadi, N., Yoshikawa, A. and Mohd Radzi, Z., 2019. Assessment of the Geomagnetically Induced Current (GIC) at Low Latitude Region based on MAGDAS Data. *Journal of physics: Conference series*, 1152(1). Available from: https://doi.org/10.1088/1742-6596/1152/1/012028.

Molinski, T.S., 2002. Why utilities respect geomagnetically induced currents. *Journal of atmospheric and solar-terrestrial physics*,

64(16), pp.1765–1778. Available from: https://doi.org/10.1016/S1364-6826(02)00126-8.

Moodley, N. and Gaunt, C.T., 2017. Low Energy Degradation Triangle for power transformer health assessment. *Ieee transactions on dielectrics and electrical insulation*, 24(1), pp.639–646. Available from: https://doi.org/10.1109/TDEI.2016.006042.

Morley, S.K., 2020. Challenges and Opportunities in Magnetospheric Space Weather Prediction. *Space weather*, 18(3). Available from: https://doi.org/10.1029/2018SW002108.

Mourenas, D., Artemyev, A.V. and Zhang, X.J., 2018. Statistics of Extreme Time-Integrated Geomagnetic Activity. *Geophysical research letters*. Available from: https://doi.org/10.1002/2017GL076828.

Mukhopadhyay, A., Welling, D.T., Liemohn, M.W., Ridley, A.J., Chakraborty, S. and Anderson, B.J., 2020. Conductance Model for Extreme Events: Impact of Auroral Conductance on Space Weather Forecasts. *Space weather*, 18(11), pp.1–27. 2008.12276, Available from: https://doi.org/10.1029/2020SW002551.

Murphy, A.H., 1977a. The Value of Climatological, Categorical and Probabilistic Forecasts in the Cost-Loss Ratio Situation. Available from: https://doi.org/10.1175/1520-0493(1977)105<0803:tvocca>2.0.co;2.

Murphy, A.H., 1977b. The Value of Climatological, Categorical and Probabilistic Forecasts in the Cost-Loss Ratio Situation. *Monthly weather review*, 105(7), pp.803–816. Available from: https://doi.org/10.1175/1520-0493.

Murray, S.A., 2018. The Importance of Ensemble Techniques for Operational Space Weather Forecasting. *Space weather*, 16(7), pp.777–783. Available from: https://doi.org/10.1029/2018SW001861.

Ngwira, C.M. and Pulkkinen, A.A., 2019. An Introduction to Geomagnetically Induced Currents. *Geomagnetically induced currents from the sun to the power grid*. pp.3–14.

O'Brien, T. and McPherron, R.L., 2000. Forecasting the ring current index Dst in real time. *Journal of atmospheric and solar-terrestrial physics*, 62(14), pp.1295–1299. Available from: https://doi.org/10.1016/s1364-6826(00)00072-9.

Ofman, L., 2004. The origin of the slow solar wind in coronal streamers. *Advances in space research*, 33(5), pp.681–688. Available from: https://doi.org/10.1016/S0273-1177(03)00235-7.

Orr, L., Chapman, S.C. and Beggan, C., 2021. Wavelet and network analysis of magnetic field variation and geomagnetically induced currents during large storms. *Space weather*, pp.1–29. Available from: https://doi.org/10.1029/2021sw002772.

Ossendrijver, M., 2003. The Solar Dynama. *The anstronomy and astrophysics review*, (11), pp.287–367. Available from: https://doi.org/10.1007/s00159-003-0019-3.

Oughton, E.J., Hapgood, M., Richardson, G.S., Beggan, C.D., Thomson, A.W., Gibbs, M., Burnett, C., Gaunt, C.T., Trichas, M., Dada, R. and Horne, R.B., 2019. A Risk Assessment Framework for the Socioeconomic Impacts of Electricity Transmission Infrastructure Failure Due to Space Weather: An Application to the United Kingdom. *Risk analysis*, 39(5), pp.1022–1043. Available from: https://doi.org/10.1111/risa.13229.

Oughton, E.J., Skelton, A., Horne, R.B., Thomson, A.W. and Gaunt, C.T., 2017a. Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure. *Space weather*, 15(1), pp.65–83. Available from: https://doi.org/10.1002/2016SW001491.

Oughton, E.J., Skelton, A., Horne, R.B., Thomson, A.W. and Gaunt, C.T., 2017b. Quantifying the daily economic impact of extreme space weather due to failure in electricity transmission infrastructure. *Space weather*, 15(1), pp.65–83. Available from: https://doi.org/10.1002/2016SW001491.

Owens, M., 2020. Solar wind structure. *Oxford research encyclopedia of physics*. Oxford University Press. Available from: https://doi.org/10.1093/acrefore/9780190871994.013.19.

Owens, M., Lang, M., Barnard, L., Riley, P., Ben-Nun, M., Scott, C.J., Lockwood, M., Reiss, M.A., Arge, C.N. and Gonzi, S., 2020. A Computationally Efficient, Time-Dependent Model of the Solar Wind for Use as a Surrogate to Three-Dimensional Numerical Magnetohydrodynamic Simulations. *Solar physics*, 295(3). Available from: https://doi.org/10.1007/s11207-020-01605-3.

Owens, M.J., 2018. Time-Window Approaches to Space-Weather Forecast Metrics: A Solar Wind Case Study. *Space weather*, 16(11), pp.1847–1861. Available from: https://doi.org/10.1029/2018SW002059.

Owens, M.J., Challen, R., Methven, J., Henley, E. and Jackson, D.R., 2013. A 27 day persistence model of near-Earth solar wind conditions: A long lead-time forecast and a benchmark for dynamical models. *Space weather*, 11(5), pp.225–236. Available from: https://doi.org/10.1002/swe.20040.

Owens, M.J. and Forsyth, R.J., 2013. The heliospheric magnetic field. *Living reviews in solar physics*, 10. Available from: https://doi.org/10.12942/lrsp-2013-5.

Owens, M.J., Horbury, T.S., Wicks, R.T., McGregor, S.L., Savani, N.P. and Xiong, M., 2014. Ensemble downscaling in coupled solar wind-magnetosphere modeling for space weather forecasting. *Space weather*,

12(6), pp.395–405. Available from: https://doi.org/10.1002/2014SW001064.

Owens, M.J., Lockwood, M. and Barnard, L.A., 2018. Ion Charge States and Potential Geoeffectiveness: The Role of Coronal Spectroscopy for Space-Weather Forecasting. *Space weather*. Available from: https://doi.org/10.1029/2018SW001855.

Owens, M.J., Lockwood, M. and Barnard, L.A., 2020. The value of CME arrival-time forecasts for space weather mitigation. *Space weather*. Available from: https://doi.org/10.1029/2020sw002507.

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

Owens, M.J., Riley, P. and Horbury, T., 2017a. The Role of Empirical Space-Weather Models (in a World of Physics-Based Numerical Simulations). *Proceedings of the international astronomical union*, 13(S335), pp.254–257. Available from: https://doi.org/10.1017/S1743921317007128.

Owens, M.J., Riley, P. and Horbury, T.S., 2017b. Probabilistic Solar Wind and Geomagnetic Forecasting Using an Analogue Ensemble or "Similar Day" Approach. *Solar physics*, 292(5), pp.1–16. Available from: https://doi.org/10.1007/s11207-017-1090-7.

Owens, M.J., Riley, P. and Horbury, T.S., 2017c. Probabilistic Solar Wind and Geomagnetic Forecasting Using an Analogue Ensemble or "Similar Day" Approach. *Solar physics*, 292(5), pp.1–16. Available from: https://doi.org/10.1007/s11207-017-1090-7.

Owens, M.J., Riley, P. and Horbury, T.S., 2017d. Probabilistic Solar Wind and Geomagnetic Forecasting Using an Analogue Ensemble or "Similar

Day" Approach. *Solar physics*, 292(5), pp.1–16. Available from: https://doi.org/10.1007/s11207-017-1090-7.

Parker, E., 1959. Extension of the solar corona into interplanetary space. *Journal of geophysical research*, 64(11), pp.1675–1681. Available from: https://doi.org/10.1029/jz064i011p01675.

Patel, K.J., Patel, J.A., Mehta, R.S., Rathod, S.B., Rajput, V.N. and Pandya, K.S., 2016. An analytic review of geomagnetically induced current effects in power system. *International conference on electrical, electronics, and optimization techniques, iceeot 2016*, pp.3906–3908. Available from: https://doi.org/10.1109/ICEEOT.2016.7755445.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E., 2011. Scikit-learn: Machine Learning in Python. *Journal of machine learning research*, 12, pp.2825—-2830.

Pirjola, R., 2000. Geomagnetically induced currents during magnetic storms. *Ieee transactions on plasma science*, 28(6), pp.1867–1873. Available from: https://doi.org/10.1109/27.902215.

Powell, K.G., Roe, P., Linde, T., Gombosi, T. and De Zeeuw, D.L., 1999. A solution-adaptive scheme for ideal magnetohydrodynamics. *Computational physics*, 154, pp.284–309. Available from: https://doi.org/10.1006/jcph.1999.6299.

Preist, E., 2020a. Magnetohydrodynamics: Overview.

Preist, E., 2020b. Solar Physics: Overview. *Oxford research encyclopedia of physics*. Oxford University Press. Available from: https://doi.org/10.1093/acrefore/9780190871994.013.21.

Pritchett, P.L. and Lu, S., 2018. Externally Driven Onset of Localized Magnetic Reconnection and Disruption in a Magnetotail Configuration.

Journal of geophysical research: Space physics, 123(4), pp.2787–2800. Available from: https://doi.org/10.1002/2017JA025094.

Pulkkinen, A., Bernabeu, E., Thomson, A., Viljanen, A., Pirjola, R., Boteler, D., Eichner, J., Cilliers, P.J., Welling, D., Savani, N.P., Weigel, R.S., Love, J.J., Balch, C., Ngwira, C.M., Crowley, G., Schultz, A., Kataoka, R., Anderson, B., Fugate, D., Simpson, J.J. and MacAlester, M., 2017. Geomagnetically induced currents: Science, engineering, and applications readiness. *Space weather*, 15(7), pp.828–856. Available from: https://doi.org/10.1002/2016SW001501.

Pulkkinen, A., Kuznetsova, M., Ridley, A., Raeder, J., Vapirev, A., Weimer, D., Weigel, R.S., Wiltberger, M., Millward, G., Rastätter, L., Hesse, M., Singer, H.J. and Chulaki, A., 2011. Geospace Environment Modeling 2008-2009 Challenge: Ground magnetic field perturbations. *Space weather*, 9(2), pp.1–13. Available from: https://doi.org/10.1029/2010SW000600.

Pulkkinen, A., Lindahl, S., Viljanen, A. and Pirjola, R., 2005. Geomagnetic storm of 29-31 October 2003: Geomagnetically induced currents and their relation to problems in the Swedish high-voltage power transmission system. *Space weather*, 3(8). Available from: https://doi.org/10.1029/2004SW000123.

Pulkkinen, A. and Rastatter, L., 2009. Minimum variance analysis-based propagation of the solar wind observations: Application to real-time global magnetohydrodynamic simulations. *Space weather*, 7(12). Available from: https://doi.org/10.1029/2009SW000468.

Pulkkinen, A., Rastätter, L., Kuznetsova, M., Hesse, M., Ridley, A., Raeder, J., Singer, H.J. and Chulaki, A., 2010. Systematic evaluation of ground and geostationary magnetic field predictions generated by global magnetohydrodynamic models. *Journal of geophysical research:*

Space physics, 115(3), pp.1–12. Available from: https://doi.org/10.1029/2009JA014537.

Pulkkinen, A., Rastätter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., Toth, G., Ridley, A., Gombosi, T., Wiltberger, M., Raeder, J. and Weigel, R., 2013a. Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations. *Space weather*, 11(6), pp.369–385. Available from: https://doi.org/10.1002/swe.20056.

Pulkkinen, A., Rastätter, L., Kuznetsova, M., Singer, H., Balch, C., Weimer, D., Toth, G., Ridley, A., Gombosi, T., Wiltberger, M., Raeder, J. and Weigel, R., 2013b. Community-wide validation of geospace model ground magnetic field perturbation predictions to support model transition to operations. *Space weather*, 11(6), pp.369–385. Available from: https://doi.org/10.1002/swe.20056.

Pulkkinen, T., 2007. Space Weather: Terrestrial Perspective Living Reviews in Solar Physics. *Living rev. solar phys*, 4, p.1. Available from: http://www.livingreviews.org/lrsp-2007-1http://www.ava.fmi.fi/.

Raeder, J., Berchem, J. and Ashour-Abdalla, M., 1998. The Geospace Environment Modeling Grand Challenge: Results from a Global Geospace Circulation Model. *Journal of geophysical research: Space physics*, 103(A7), pp.14787–14797. Available from: https://doi.org/10.1029/98ja00014.

Raeder, J., Wang, Y.L., Fuller-Rowell, T.J. and Singer, H.J., 2001. Global simulation of magnetospheric space weather effects of the Bastille day storm. *Solar physics*, 204(1-2), pp.325–338. Available from: https://doi.org/10.1023/A:1014228230714.

Rajput, V.N., Boteler, D.H., Rana, N., Saiyed, M., Anjana, S. and Shah, M., 2021. Insight into impact of geomagnetically induced currents on

power systems: Overview, challenges and mitigation. *Electric power systems research*, 192(March 2020), p.106927. Available from: https://doi.org/10.1016/j.epsr.2020.106927.

Reynolds, P., Cone, J., Layefsky, M., Goldberg, D.E. and Hurley, S., 2002. Cancer incidence in California flight attendants (United States). *Cancer causes and control*, 13(4), pp.317–324. Available from: https://doi.org/10.1023/A:1015284014563.

Richardson, D.S., 2000. Skill and relative economic value of the ECMWF ensamble prediction system. (October 1998), pp.649–667. Available from: https://doi.org/10.1002/qj.49712656313.

Richardson, I.G., 2004. Energetic particles and corotating interaction regions in the solar wind. *Space science reviews*, 111(3-4), pp.267–376. Available from: https://doi.org/10.1023/B:SPAC.0000032689.52830.3e.

Richardson, I.G., 2018. Solar wind stream interaction regions throughout the heliosphere. *Living reviews in solar physics*, 15(1), pp.1–95. Available from: https://doi.org/10.1007/s41116-017-0011-z.

Richardson, I.G., Cliver, E.W. and Cane, H.V., 2001. Sources of geomagnetic storms for solar minimum and. 28(13), pp.2569–2572. Available from: https://doi.org/https://doi.org/10.1029/2001GL013052.

Richardson, I.G., Webb, D.F., Zhang, J., Berdichevsky, D.B., Biesecker, D.A., Kasper, J.C., Kataoka, R., Steinberg, J.T., Thompson, B.J., Wu, C.C. and Zhukov, A.N., 2006. Major geomagnetic storms (Dst -100 nT) generated by corotating interaction regions. *Journal of geophysical research: Space physics*, 111(7), pp.1–17. Available from: https://doi.org/10.1029/2005JA011476.

Ridley, A.J., Hansen, K.C., Tóth, G., De Zeeuw, D.L., Gombosi, T.I. and Powell, K.G., 2002. University of Michigan MHD results of the geospace global circulation model metrics challenge. *Journal of geophysical research: Space physics*, 107(A10), pp.1–19. Available from: https://doi.org/10.1029/2001JA000253.

Riley, P., 2012. On the probability of occurrence of extreme space weather events. *Space weather*, 10(2), pp.1–12. Available from: https://doi.org/10.1029/2011SW000734.

Riley, P., Baker, D., Liu, Y.D., Verronen, P., Singer, H. and Güdel, M., 2018. Extreme Space Weather Events: From Cradle to Grave. Available from: https://doi.org/10.1007/s11214-017-0456-3.

Riley, P., Ben-Nun, M., Linker, J.A., Owens, M.J. and Horbury, T.S., 2017. Forecasting the properties of the solar wind using simple pattern recognition. *Space weather*, 15(3), pp.526–540. Available from: https://doi.org/10.1002/2016SW001589.

Riley, P. and Love, J.J., 2017. Extreme geomagnetic storms: Probabilistic forecasts and their uncertainties. *Space weather*, 15(1), pp.53–64. Available from: https://doi.org/10.1002/2016SW001470.

Roberts, N.M. and Lean, H.W., 2008. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly weather review*, 136(1), pp.1–19. Available from: https://doi.org/10.1175/2007MWR2123.1.

Rogers, N.C., Wild, J.A., Eastoe, E.F., Gjerloev, J.W. and Thomson, A.W., 2020. A global climatological model of extreme geomagnetic field fluctuations. *Journal of space weather and space climate*, 10, pp.1–19. Available from: https://doi.org/10.1051/swsc/2020008.

Russell, C.T., Luhmann, J.G. and Strangeway, R.J., 2016. Solar wind-

magnetosphere coupling. *Space physics: An introduction*. Cambridge University Press.

Schrijver, C.J., 2015. Socio-Economic Hazards and Impacts of Space Weather: The Important Range between Mild and Extreme. *Space weather*, 13(9), pp.524–528. Available from: https://doi.org/10.1002/2015SW001252.

Schrijver, C.J., Dobbins, R., Murtagh, W. and Petrinec, S.M., 2014a. Assessing the impact of space weather on the electric power grid based on insurance claims for industrial electrical equipment. *Space weather*, 12(7), pp.487–498. 1406.7024, Available from: https://doi.org/10.1002/2014SW001066.

Schrijver, C.J., Dobbins, R., Murtagh, W. and Petrinec, S.M., 2014b. Assessing the impact of space weather on the electric power grid based on insurance claims for industrial electrical equipment. *Space weather*, 12(7), pp.487–498. 1406.7024, Available from: https://doi.org/10.1002/2014SW001066.

Sharpe, M.A. and Murray, S.A., 2017. Verification of Space Weather Forecasts Issued by the Met Office Space Weather Operations Centre. *Space weather*, 15(10), pp.1383–1395. Available from: https://doi.org/10.1002/2017SW001683.

Simpson, F. and Bahr, K., 2020. Nowcasting and validating Earth's electric-field response to extreme space-weather events using magnetotelluric data: application to the September 2017 geomagnetic storm and comparison to observed and modelled fields in Scotland. *Space weather*, (September), pp.0–1. Available from: https://doi.org/10.1029/2019sw002432.

Singer, H.J., Heckman, G.R. and Hirman, J.W., 2001. Space weather forecasting: A grand challenge. *Geophysical monograph series*, 125, pp.23–29. Available from: https://doi.org/10.1029/GM125p0023.

Singer, J.D. and Willett, J.B., 1993. It's about time: using discrete-time survival analysis to study duration and the timing of events. *American educational research association*, 18(2), pp.155–195. Available from: https://doi.org/10.2307/1165085.

Skone, S.H., 2001. The impact of magnetic storms on GPS receiver performance. *Journal of geodesy*, 75(9-10), pp.457–468. Available from: https://doi.org/10.1007/s001900100198.

Sorathia, K.A., Merkin, V.G., Panov, E.V., Zhang, B., Lyon, J.G., Garretson, J., Ukhorskiy, A.Y., Ohtani, S., Sitnov, M. and Wiltberger, M., 2020. Ballooning-Interchange Instability in the Near-Earth Plasma Sheet and Auroral Beads: Global Magnetospheric Modeling at the Limit of the MHD Approximation. *Geophysical research letters*, 47(14). Available from: https://doi.org/10.1029/2020GL088227.

Sreeja, V., 2016. Impact and mitigation of space weather effects on GNSS receiver performance. *Geoscience letters*, 3(1). Available from: https://doi.org/10.1186/s40562-016-0057-0.

Stone, E.C., Frandsen, A.M. and Mewaldt, R.A., 1998. The advanced composition explorer. pp.1–22.

Störmer, C., 1912. Sur les trajectoires des corpuscules électrisés dans l'espace sous l'action du magnétisme terrestre, avec application aux aurores boréales. *Le radium*, 9(11), pp.395–399. Available from: https://doi.org/10.1051/radium:01912009011039501.

Sundberg, T., Burgess, D., Scholer, M., Masters, A. and Sulaiman, A.H., 2017. The Dynamics of Very High Alfvén Mach Number Shocks in Space Plasmas. *The astrophysical journal*, 836(1), p.L4. Available from: https://doi.org/10.3847/2041-8213/836/1/l4.

Székely, G.J., Rizzo, M.L. and Bakirov, N.K., 2007. Measuring and testing dependence by correlation of distances. *Annals of statistics*,

35(6), pp.2769–2794. Available from: https://doi.org/10.1214/009053607000000505.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *Journal of geophysical research: Atmospheres*, 106, pp.7183–7192. Available from: https://doi.org/10.1029/2000JD900719.

Tobiska, W.K., Atwell, W., Beck, P., Benton, E., Copeland, K., Dyer, C., Gersey, B., Getley, I., Hands, A., Holland, M., Hong, S., Hwang, J., Jones, B., Malone, K., Meier, M.M., Mertens, C., Phillips, T., Ryden, K., Schwadron, N., Wender, S.A., Wilkins, R. and Xapsos, M.A., 2015. Advances in atmospheric radiation measurements and modeling needed to improve air safety. *Space weather*, 13(4), pp.202–210. Available from: https://doi.org/10.1002/2015SW001169.

Toffoletto, F., Sazykin, S., Spiro, R. and Wolf, R., 2003. Inner Magnetospheric Modeling with the Rice Convection Model. *Advances in space environment research*. Springer. Available from: https://doi.org/10.1007/978-94-007-1069-6_19.

Tóth, G., Holst, B. van der, Sokolov, I.V., De Zeeuw, D.L., Gombosi, T.I., Fang, F., Manchester, W.B., Meng, X., Najib, D., Powell, K.G., Stout, Q.F., Glocer, A., Ma, Y.J. and Opher, M., 2012. Adaptive numerical algorithms in space weather modeling. *Journal of computational physics*, 231(3), pp.870–903. Available from: https://doi.org/10.1016/j.jcp.2011.02.006.

Tóth, G., Sokolov, I.V., Gombosi, T.I., Chesney, D.R., Clauer, C.R., De Zeeuw, D.L., Hansen, K.C., Kane, K.J., Manchester, W.B., Oehmke, R.C., Powell, K.G., Ridley, A.J., Roussev, I.I., Stout, Q.F., Volberg, O., Wolf, R.A., Sazykin, S., Chan, A., Yu, B. and Kóta, J., 2005. Space weather modeling framework: A new tool for the space science com-

munity. *Journal of geophysical research: Space physics*, 110(A12), pp.1–21. Available from: https://doi.org/10.1029/2005JA011126.

Trichtchenko, L. and Boteler, D.H., 2004. Modeling geomagnetically induced currents using geomagnetic indices and data. *Ieee transactions on plasma science*, 32(4 I), pp.1459–1467. Available from: https://doi.org/10.1109/TPS.2004.830993.

Trichtchenko, L. and Boteler, D.H., 2007. Effects of Recent Geomagnetic Storms on Power Systems. Available from: https://doi.org/10.1109/EMCECO.2007.4371706.

Tsurutani, B.T., Gonzalez, W.D., Gonzalez, A.L., Guarnieri, F.L., Gopalswamy, N., Grande, M., Kamide, Y., Kasahara, Y., Lu, G., Mann, I., McPherron, R., Soraas, F. and Vasyliunas, V., 2006. Corotating solar wind streams and recurrent geomagnetic activity: A review. *Journal of geophysical research: Space physics*, 111(7), pp.1–25. Available from: https://doi.org/10.1029/2005JA011273.

Van den Dool, H.M., 1989. A New Look at Weather Forecasting through Analogues. *Monthly weather review*, 117(10), pp.2230–2247. Available from: https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2.

Vassiliadis, D. and Klimas, A.J., 1995. On the uniqueness of linear moving-average filters for the solar wind-auroral geomagnetic activity coupling. *Journal of geophysical research*, 100(A4), p.5637. Available from: https://doi.org/10.1029/94ja03303.

Vassiliadis, D., Klimas, A.J., Baker, D.N. and Roberts, D.A., 1995. A description of the solar wind-magnetosphere coupling based on nonlinear filters. *Journal of geophysical research: Space physics*, 100(A3), pp.3495–3512. Available from: https://doi.org/10.1029/94ja02725.

Walach, M. and Grocott, A., 2019. SuperDARN observations during geomagnetic storms, geomagnetically active times and enhanced solar wind driving. *Journal of geophysical research: Space physics*, p.2019JA026816. Available from: https://doi.org/10.1029/2019JA026816.

Watari, S., 2011. Forecast of recurrent geomagnetic storms. *Advances in space research*, 47(12), pp.2162–2171. Available from: https://doi.org/10.1016/j.asr.2010.07.029.

Webb, D.F. and Howard, T.A., 2012. Coronal Mass Ejections : Observations. *Living reviews in solar physics*, 9(3).

Weimer, D.R., 2013. An empirical model of ground-level geomagnetic perturbations. *Space weather*, 11(3), pp.107–120. Available from: https://doi.org/10.1002/swe.20030.

Weimer, D.R., 2019. Empirical modeling of the geomagnetic field for GIC prediction. *Geomagnetically induced currents from the sun to the power grid*. pp.67–78. Available from: https://doi.org/10.1002/9781119434412.ch4.

Welling, D., 2019. Magnetohydrodynamic models of B and their use in GIC estimates. *Geomagnetically induced currents from the sun to the power grid*. AGU publications, chap. Magnetohyd, pp.43–65. Available from: https://doi.org/10.1002/9781119434412.ch3.

Welling, D., Love, J., Rigler, J., Oliveira, D., Komar, C. and Morley, S., 2021. Numerical Simulations of the Geospace Response to the Arrival of an Idealized Perfect Interplanetary Coronal Mass Ejection. *Space weather*, 19(2), pp.1–15. Available from: https://doi.org/10.1029/2020sw002489.

Wilson, J.W., Goldhagen, P., Rafnsson, V., Clem, J.M., De Angelis, G. and Friedberg, W., 2003. Overview of Atmospheric Ionizing Radiation

(AIR) research: SST-present. *Advances in space research*, 32(1), pp.3–16. Available from: https://doi.org/10.1016/S0273-1177(03)90364-4.

Wiltberger, M., Rigler, E.J., Merkin, V. and Lyon, J.G., 2017. Structure of High Latitude Currents in Magnetosphere-Ionosphere Models. *Space science reviews*, 206(1-4), pp.575–598. Available from: https://doi.org/10.1007/s11214-016-0271-2.

Wing, S., Johnson, J.R., Jen, J., Meng, C.I., Sibeck, D.G., Bechtold, K., Freeman, J., Costello, K., Balikhin, M. and Takahashi, K., 2005. Kp forecast models. *Journal of geophysical research: Space physics*, 110(A4), pp.1–14. Available from: https://doi.org/10.1029/2004JA010500.

Xie, H., Gopalswamy, N., Cyr, O.C. and Yashiro, S., 2008. Effects of solar wind dynamic pressure and preconditioning on large geomagnetic storms. *Geophysical research letters*, 35(6), pp.6–10. Available from: https://doi.org/10.1029/2007GL032298.

Yokoyama, N. and Kamide, Y., 1997. Statistical nature of geomagnetic storms. *Journal of geophysical research a: Space physics*, 102(A7), pp.14215–14222. Available from: https://doi.org/10.1029/97JA00903.

Zhang, X.Y. and Moldwin, M.B., 2015. Probabilistic forecasting analysis of geomagnetic indices for southward IMF events. *Space weather*, 13(3), pp.130–140. Available from: https://doi.org/10.1002/2014SW001113.